

Post-secondary Students' Travel Behavior through the Lens of Urban and Rural Contexts

Nishat Naila Meghna

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Civil Engineering

Md. Sami Hasnine, Chair

Kathleen L. Hancock

Gerardo F. Flintsch

Anne R. Driscoll

September 19, 2025

Blacksburg, Virginia

Keywords: Post-secondary student, Travel Behavior, Urban, Rural, Activity type,
Departure time, Mode Choice, Bike Infrastructure, Mental Health

Copyright 2025, Nishat Naila Meghna

Post-secondary Students' Travel Behavior through the Lens of Urban and Rural Contexts

Nishat Naila Meghna

(ABSTRACT)

This four-manuscript research investigates the travel behavior of post-secondary students across urban and rural contexts through four interrelated studies. The research addresses gaps in understanding how activity choices, departure time decisions, and active transportation behaviors are shaped by contextual, demographic, and policy factors. Two studies utilize the StudentMoveTO dataset, a detailed activity-travel diary survey from the Greater Toronto and Hamilton Area (GTHA), to model multi-destination trip-based activity type choices and sequential departure time decisions. These models capture interdependencies across multi-destination trips, enabling a more realistic representation of student travel patterns in dense urban environments. The other two studies draw on a custom-designed revealed-stated preference (RP-SP) survey administered to post-secondary students in rural Virginia. This survey incorporates both actual travel behavior and hypothetical choice experiments to assess rural students' mode preferences and the mental health impacts of active transportation under varying infrastructure and service conditions. The research adopted advanced econometric and machine learning approaches to better understand post-secondary students travel behavior. The urban-focused studies employ the dynamic discrete choice models and deep learning architectures Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM) networks, and Transformers to capture sequential decision-making and nonlinear dependencies in activity type and departure time choices. The rural-focused mode choice analysis estimates both RP-SP multinomial logit (MNL) and RP-SP mixed logit models, enabling the combination of actual revealed preference data with stated preference scenarios while

also capturing unobserved taste heterogeneity across individuals. The MNL model provides a baseline understanding of average mode choice behavior, whereas the mixed logit model relaxes the independence of irrelevant alternatives (IIA) assumption and accounts for random variations in preferences influenced by rural context, demographics, and travel conditions. The mental health and active transportation study uses principal component analysis (PCA) for dimensionality reduction, followed by Random Forest and other interpretable machine learning methods to identify the most influential factors. This combined methodological framework leverages both behavioral realism and predictive accuracy, bridging traditional econometric analysis with modern data-driven approaches. The findings reveal that student travel decisions are strongly influenced by institutional schedules, socio-demographic characteristics, and built environment features, with notable differences between urban and rural contexts. Sequential modeling shows that earlier departure times for initial trips significantly constrain subsequent activity timing, while rural analyses highlight that infrastructure quality and service availability directly affect both mode choice and perceived mental health benefits of active travel. These insights provide valuable evidence for transportation planners and policymakers seeking to design targeted, context-sensitive strategies that enhance mobility options, support student well-being, and promote sustainable transportation in both urban and rural communities.

Post-secondary Students' Travel Behavior through the Lens of Urban and Rural Contexts

Nishat Naila Meghna

(GENERAL AUDIENCE ABSTRACT)

This four-manuscript research examines how college and university students travel, focusing on the choices they make about activities, timing, travel modes, and the role of active transportation in supporting mental health. The research considers students in both large metropolitan areas and rural communities, recognizing that their opportunities and challenges differ greatly depending on where they live and study. Two of the studies draw on the StudentMoveTO survey from the Greater Toronto and Hamilton Area, a detailed record of students' daily trips and activities. These studies explore how students decide the types of activities they do on their trips throughout the day and when they choose to begin each trip. The findings show that earlier departures often limit later activities, and that travel patterns are shaped by a mix of institutional schedules, personal commitments, and the urban environment. The other two studies are based on a custom survey conducted with post-secondary students in rural Virginia. This survey captures both actual travel behavior and responses to "what if" scenarios about new transportation options, changes in infrastructure, or different travel costs. One study examines how walking and cycling in rural areas relate to students' mental health, revealing that better infrastructure and safer routes can encourage active travel and improve well-being. The other investigates how rural students choose their travel mode, whether driving, carpooling, using public transit, or cycling, and how these decisions respond to service availability, infrastructure quality, and travel costs. By combining real-world travel data with hypothetical scenarios, the research highlights the different ways students adapt their travel in urban and rural settings. The results show

that infrastructure, service quality, and travel costs strongly influence mode choice, while active travel can offer important mental health benefits when safe and convenient options are available. These findings provide actionable insights for transportation planners and policymakers, offering strategies to expand mobility, improve student well-being, and promote sustainable travel in both cities and rural areas.

Dedication

To my beloved parents, whose unconditional love, constant support, and countless sacrifices have shaped every step of my journey. From the earliest days of my education, you instilled in me the values of hard work, integrity, and perseverance. Your encouragement during challenges and your joy in my achievements have been my greatest motivation. This dissertation is a reflection of your faith in me and the foundation you built for my dreams.

It is, in every sense, as much yours as it is mine.

Acknowledgments

First and foremost, I am profoundly thankful to Almighty Allah for granting me the strength, patience, and guidance to complete this journey. Without His mercy and blessings, none of this would have been possible.

I am sincerely grateful to my supervisor, Dr. Md Sami Hasnine, whose encouragement, wise counsel, and faith in me have been a constant source of motivation. Before coming to the USA, I had imagined this Ph.D. journey to be filled with relentless challenges and overwhelming obstacles. However, from the very first day, you transformed my expectations entirely. You made this path not only manageable but deeply enjoyable. Your ability to balance high academic standards with genuine care and understanding made all the difference in my experience. You have stood beside me in every situation, celebrating moments of success, guiding me through uncertainty, and offering reassurance during difficult times. Because of your kindness, patience, and trust, I could walk this journey with confidence and joy. You are the best teacher I have met in my life so far, and I know you will continue to inspire and motivate me for many years to come. Along this journey, I also had the pleasure of meeting Dr. Hasnine's daughter, Inara. I still remember the very first day we met, she was so small and cried because she did not want to go home. Watching her grow over the years has been a joy, and from that very first moment, I felt a deep affection for her. I will truly miss seeing her around, and I love her dearly.

I would also like to extend my heartfelt thanks to my dissertation committee members, Dr. Katheleen Hancock, Dr. Gerardo Flintsch, and Dr. Anne Driscoll, for their constructive feedback, encouragement, and valuable contributions to my work.

A special note of appreciation goes to Tanmay Vaia, who joined my journey later but from

whom I learned a great deal in a short time. His knowledge, patience, and willingness to guide me made a meaningful difference in my research.

My deepest gratitude goes to my beloved parents, whose unconditional love, endless sacrifices, and constant encouragement have shaped me into the person I am today. Your prayers have been my greatest shield, and every step of my success is a reflection of your dedication and care. I want to express my heartfelt admiration for my mother-in-law, whose strength and courage as a blood cancer patient have been a constant source of motivation for me. Her resilience has taught me the true meaning of perseverance and hope. I am deeply thankful to my dear sisters, Apumoni and Athoy, for their endless love, encouragement, and constant moral support. You have celebrated my milestones, comforted me during challenges, and stood beside me in every step of my journey. I am also grateful to my sister-in-law, Urmi Apu, for her warmth, kindness, and support. Among my extended family, my cousin Mission Vaia has been a constant motivator, whenever I felt sad or lost, I would turn to him, and his advice and encouragement always lifted my spirits. I have sought his suggestions for almost everything, and his guidance has been invaluable. My cousin brother Mishuk has also supported me in many ways; talking to him always brought me comfort and relief during stressful times. My cousin sister Tonu, who is also a close friend, has been another pillar of support, sharing in both my joys and my struggles with unwavering care. I am also grateful to my chachi, who always took great care of me and was always there to talk to me; her care has meant more than words can express. I miss my kaku and dadi dearly, in my happiest moments, I often think of them and know they would have been so proud and happy if they were alive today.

I will forever be grateful to my ex-labmate, Birat, for his generous help with Gauss coding during the very first days of my PhD life. Those moments of guidance gave me the confidence to move forward. My heartfelt thanks also go to all my other labmates for their support.

To Faria, Samah, Nawsin and Ahmed -I am deeply thankful for the wonderful lab environ-

ment we created together. We became more than colleagues; we became sisters, sharing both our joys and our struggles. Our friendship has been a source of strength and happiness throughout my Ph.D. I also want to thank Logno, Rafia, Nawar, Setu, and Efaz for making my days in Blacksburg so memorable. You guys have been my family away from home. You made me your sister and helped me in every situation. I can never thank you enough for your kindness and care. I am forever grateful to Aboni apu who took great care of me when I was severely sick, cooked food for me, and whose support gave me the strength to recover and continue my journey.

I am very much grateful to Pulok Vaia from Howard University, who has been another brother figure in my life. He helped me the most during my first days in the USA, and I am forever grateful to him. Without his support, I could not have survived those challenging early days. Even now, he continues to support me in different ways, always offering guidance and encouragement whenever I need it, and I look up to him for anything that I need. Rahat Vaia and Dip Vaia, thank you for becoming like brothers to me and for always offering your advice and support.

Lastly, and most importantly, to my husband, the love of my life, Rahad - you are my greatest blessing from Almighty Allah. You have been my unwavering strength, standing by me in every situation, no matter how difficult, and making countless sacrifices without ever expecting anything in return. You traveled from Philadelphia to Blacksburg whenever I needed you, without a moment's hesitation, just to be by my side. You carried my worries as your own, shared in every joy and every tear, and never once let me feel alone in this journey. I cannot thank Almighty enough for writing your name alongside mine in the story of my life. This achievement is as much yours as it is mine, for without your love, patience, and selfless support, I could never have made it here.

Contents

List of Figures	xvi
List of Tables	xviii
Author Contribution Statement	1
1 Introduction	2
1.1 Background and Motivation	2
1.2 The Neglected Demographic: Post-secondary Students	5
1.3 Geographic Inequities: Urban vs. Rural Student Mobility	7
1.4 Gaps in Existing Literature	9
1.4.1 Lack of Post-Secondary Student-Specific Travel Models	9
1.4.2 Urban-Rural Divide in Research Coverage	10
1.4.3 Methodological Limitations	11
1.5 Research Objectives	15
1.6 Methodological Overview	17
1.7 Contributions to Knowledge	18
1.7.1 Centering Post-secondary Students in Urban-Rural Mobility Research	19
1.7.2 Advancing Dynamic and Mixed Econometric Modeling in Student Mo- bility Analysis	19

1.7.3	Advancing Sequential Departure Time Modeling	20
1.7.4	Bridging Transportation Behavior with Mental Health and Well-being	20
1.8	Structure of the Dissertation	21
2	Literature Review	23
2.1	Chapter Overview	23
2.2	Activity type choice modeling within activity-based models	24
2.3	Activity type choice modeling within multi-destination trip-based and tour-based models	26
2.4	Review of Modeling Approaches for Sequential Departure Time Decisions . .	28
2.4.1	Discrete Choice Models (DCMs)	28
2.4.2	Optimization and Game-Theoretic Models	31
2.4.3	Deep Learning Approaches	32
2.5	Active Transportation and Mode Choice in Rural College Towns	34
2.6	Mental Health Impacts of Active Transportation in Rural Communities . . .	40
2.6.1	Transportation and Mental Health	40
2.6.2	Cycling and Mental Health Benefits	41
2.6.3	Cycling and Non-Motorized Travel in Rural Areas	41
3	Investigating Post Secondary Students Activity Type Choice Pattern in a Multi-destination Trip-based Approach	48
3.1	Chapter Overview	48
3.2	Introduction	49

3.3	Data for Empirical Investigation	51
3.4	Econometric Model	53
3.5	Data Analysis	58
3.6	Model Estimation Result	60
3.6.1	Two-destination trip Multinomial Logit Model	61
3.6.2	Two-destination Trip Nested Logit Model	63
3.6.3	Two-destination Trip Mixed Logit Model	64
3.6.4	Three-destination Trip Dynamic Discrete Choice Model	65
3.6.5	Three-destination Trip Dynamic Discrete Choice Mixed Model	66
3.7	Conclusion	82
4	Modeling Departure Time Choices in Multi-destination Travel: A Deep Learning Investigation	84
4.1	Chapter Overview	84
4.2	Introduction	86
4.3	Data Description	90
4.4	Methodology	95
4.4.1	Input Representation and Preprocessing	96
4.4.2	Transformer Model	98
4.4.3	Feature Transformation	98
4.4.4	Model Structure	99
4.4.5	Mathematical Formulation	99

4.4.6	GRU Model Structure and Mathematical Formulation	103
4.4.7	Training Procedure and Loss Function	105
4.4.8	Long Short-Term Memory (LSTM) Model	107
4.4.9	Model Architecture	108
4.4.10	Evaluation and Interpretation	112
4.4.11	Hyperparameter Tuning and Model Configuration	113
4.5	Results and Discussion	115
4.5.1	Baseline Model and Motivation for Deep Learning Approaches	115
4.5.2	Distributional fidelity of predicted departure times	116
4.5.3	Comparative Analysis of Deep Learning Models	121
4.5.4	SHAP-Based Interpretability Analysis	123
4.6	Conclusion and Recommendations	126
5	Investigating the Travel Behavior of Post-Secondary Students' in Rural USA College Towns	129
5.1	Chapter Overview	129
5.2	Introduction	130
5.3	Survey Design	133
5.3.1	Study Area	133
5.3.2	Stated Preference (SP) Survey	134
5.3.3	Revealed Preference (RP) Survey	136
5.3.4	Necessity of Using RP-SP Survey together	137

5.4	Descriptive Statistics	139
5.4.1	Mode Share Analysis	139
5.4.2	Sociodemographic Data Analysis	144
5.5	Econometric Model Specification	144
5.5.1	Combined RP-SP Multinomial Logit (MNL) Model	145
5.5.2	Mathematical Formulation	145
5.5.3	Combined RP-SP Mixed Logit (MxL) Model	147
5.5.4	Mathematical Formulation	147
5.6	Model Results	149
5.6.1	Joint RP-SP MNL	149
5.6.2	Joint SP-RP Mixed Logit Model	157
5.7	Conclusion and Recommendations	164
6	Potential Health Benefits of Choosing Active Mode in Rural College Towns in USA	168
6.1	Chapter Overview	168
6.2	Introduction	169
6.2.1	Survey Design Overview	172
6.2.2	Study Area and Participant Selection	172
6.2.3	Sampling Method	175
6.2.4	Verification Process	175
6.2.5	Data Collection	176

6.2.6	Ethical Considerations	177
6.3	Methodology	177
6.3.1	Principal Component Analysis (PCA)	177
6.3.2	Random Forest Model	181
6.4	Data Analysis	187
6.4.1	Participants Demographics	188
6.4.2	Impact of cycling on Various Demographic Groups and Mental Health	189
6.5	Result and Discussion	192
6.5.1	Result	192
6.5.2	Discussion	198
6.6	Conclusion and Recommendations	199
7	Conclusions	201
7.0.1	Research Summary	201
7.0.2	Research Contributions	204
7.0.3	Limitations and Future Work	206
7.1	Statement of AI Use	208
	Appendices	227
	Appendix A Appendix	228
A.1	RP-SP Survey Design	228
A.2	RP-SP Survey Questionnaire	229

List of Figures

4.1	Histograms of Trip 1 and Trip 2 departure times. Departure times are shown in minutes since midnight, where $400 \approx 6:40$ AM, $600 \approx 10:00$ AM, $800 \approx 1:20$ PM, $1000 \approx 4:40$ PM, $1200 \approx 8:00$ PM, and $1400 \approx 11:20$ PM.	91
4.2	PDF and CDF of actual vs. predicted Trip 1 departure time	117
4.3	PDF and CDF of actual vs. predicted Trip 2 departure time	120
4.4	SHAP summary plots for Trip 1 and Trip 2 Transformer models	124
5.1	Sample Stated Preference Survey Scenario	136
5.2	Sample Structure of RP Survey in QuestionPro	138
5.3	Comparative Analysis of Transportation Mode Share Across Six SP Scenarios	141
5.4	Mode Choice Distribution in RP Survey	143
5.5	Bike lane on Prices Fork Road, Blacksburg, VA	150
5.6	Bike Lane on Peppers Ferry Road, Christiansburg, VA	150
6.1	Geographic Distribution of Surveyed Rural Towns in Virginia	173
6.2	Gender Differences in Self-Reported Mental Health Ratings. <i>Note: The sum of each individual color is 100%. For example, the sum of all dark blue color bars is 100%.</i>	188
6.3	Distribution of Mental Health Ratings by Age Range. <i>Note: The sum of each individual color is 100%. For example, the sum of all brown color bars is 100%.</i>	189

6.4	Comparison of cycling Frequency and Mental Health Ratings Based on Medication Needs	190
6.5	Distribution of Cycling Frequency for Different Mental Health Improvement Ratings by Gender	191
6.6	Cumulative Explained Variance by PCA Components	193
6.7	Confusion Matrix (Percentage Form) for Random Forest Model	194
6.8	Influential Factors in Cycling and Mental Health relationship with their SHAP values	195
6.9	SHAP Value Summary Plot of Feature Importance	196
6.10	SHAP Waterfall Plot of Feature Contributions to Predicted Mental Health Impact of Biking	197

List of Tables

2.1	Summary of Departure Time Choice Modeling Studies	33
2.2	Comparative Overview of Literature	38
2.3	Summary of Key Studies on the Mental Health Benefits of Cycling	43
3.1	Level of service attributes in two and three destination trips	60
3.2	Travel time and duration attributes in two and three destination trips	61
3.4	Summary Table of Dynamic Discrete Choice for Three-destination Trip Model	74
3.5	Summary Table of Dynamic Discrete Mixed Choice Model for Three-destination trip	79
4.1	Summary Statistics for 2-destination trip Dataset	92
4.2	Combined hyperparameter tuning and final configurations for GRU, LSTM, and Transformer models used in Trip 1 and Trip 2 departure time prediction.	114
4.3	Performance comparison of GRU, LSTM, and Transformer models on Trip 1 and Trip 2 using RMSE and MAPE metrics.	122
5.1	Summary of Rural College Towns in Virginia Included in the Study	138
5.2	Mode Share (%) Across Six SP Scenarios	141
5.3	Joint RP–SP MNL Model Outcome	153
5.4	Joint RP–SP MxL Model Outcome	159
6.1	Summary of Rural College Towns in Virginia Included in the Study	174

6.2	Classification Report of Random Forest Model	193
-----	--	-----

List of Abbreviations

- DDCM Dynamic Discrete Choice Model — Framework modeling sequential decision-making over time
- GRU Gated Recurrent Unit — Recurrent neural network variant for sequence modeling
- GTHA Greater Toronto and Hamilton Area — Canadian metropolitan region
- IIA Independence of Irrelevant Alternatives — Assumption that odds between two alternatives are unaffected by other options
- LSTM Long Short-Term Memory — Neural network architecture addressing long-term dependencies
- ML Machine Learning — Computational methods that learn patterns from data
- MNL Multinomial Logit — A basic discrete choice model assuming IIA
- MxL Mixed Logit — A flexible choice model allowing random taste variation
- PCA Principal Component Analysis — Dimensionality reduction technique identifying major variance components
- RF Random Forest — Ensemble ML algorithm based on decision trees
- RP Revealed Preference — Data from observed real-world choices
- RP-SP Revealed-Stated Preference — Combined dataset/model using both RP and SP data
- SP Stated Preference — Data from survey-based hypothetical choices

Author Contribution Statement

Paper 1: Investigating Post Secondary Students Activity Type Choice Pattern in a Multi-destination Trip-based Approach

Nishat Naila Meghna: Conceptualized the study, designed the methodology, and led the writing of the manuscript; Md Sami Hasnine: Supervised the project, provided methodological guidance, and reviewed/edited the manuscript for technical accuracy.

Paper 2: Modeling Departure Time Choices in Multi-destination Travel: A Deep Learning Investigation

Nishat Naila Meghna: Led research design, conducted data preprocessing, developed and applied deep learning models, and drafted the paper; Tanmay Das: Guided methodological direction; Md Sami Hasnine: Supervised the project, guided methodological direction, and reviewed/edited the manuscript for technical rigor.

Paper 3: Investigating the Travel Behavior of Post-Secondary Students in Rural USA College Towns

Nishat Naila Meghna: Designed the research framework, carried out the RP–SP survey and analysis, developed econometric models, and drafted the manuscript; Jungwan Kim: Guided methodological direction; Md Sami Hasnine: Provided supervision, methodological feedback, and manuscript revisions.

Paper 4: Potential Health Benefits of Choosing Active Mode in Rural College Towns in USA

Nishat Naila Meghna: Conceptualized the study, designed and administered the SP and RP survey, performed PCA, Random Forest, and SHAP analyses, and drafted the manuscript; Max Trell: Guided methodological direction; Md Sami Hasnine: Provided supervision, contributed to methodological framework, and reviewed/edited the manuscript.

Chapter 1

Introduction

1.1 Background and Motivation

Transportation is the connective tissue of modern civilization. More than just a functional mechanism for moving people and goods, transportation determines not only the efficiency with which economies operate but also the inclusiveness and fairness of societal structures. Well-designed transportation systems facilitate access to employment, education, health-care, and civic engagement, enabling individuals to participate fully in the opportunities their communities offer. Conversely, inequitable or inadequate transportation systems can create systemic barriers, entrenching patterns of exclusion and limiting life opportunities for disadvantaged populations. Historically, transportation has been understood as both a driver and a product of broader socio-economic and spatial patterns. The form a city takes, along with its economic activity and demographic distribution, has in large part evolved in response to the accessibility that transportation systems provide. These systems shape where people live, where businesses locate, and how communities grow. Transportation is thus not only a technical service, but also a central organizing framework for society. Decisions about its design and operation have far-reaching implications for urban expansion, economic productivity, environmental sustainability, and social equity.

In recent decades, the transportation landscape has been profoundly transformed by a convergence of global trends. Rapid technological innovation has introduced intelligent transportation systems, electric mobility options, shared mobility services, and data-driven opti-

mization of networks. At the same time, demographic changes involving aging populations, migration patterns, and household structures have shifted travel needs and preferences. Climate change has further complicated these technological and demographic shifts and changes by amplifying the urgency of reducing transportation-related greenhouse gas emissions, compelling policymakers to reimagine infrastructure for a lower-carbon future (Creutzig et al. 2015). Meanwhile, socio-economic polarization has heightened awareness of the unequal distribution of mobility benefits and burdens (Lucas and Peter Jones 2012).

A recent United Nations report stated that more than 55% of the global population resides in urban areas, a proportion projected to rise to 68% by 2050 (United Nations, Department of Economic and Social Affairs, Population Division 2018). While urbanization generates agglomeration economies, innovation hubs, and cultural exchange (Glaeser 2011), it also intensifies chronic problems such as traffic congestion, air and noise pollution, road safety concerns, and inequitable access to services. The need to accommodate growing urban populations without exacerbating already existing challenges has elevated the role of sustainable and equitable transportation planning (Banister 2008). In parallel, rural areas face a different set of pressures, including population decline, service centralization, and shrinking public transportation options, leaving residents increasingly isolated (Farrington and Farrington 2005).

Transportation's economic footprint in the United States underscores its importance to the general population. Specifically, transportation is the second-largest household expenditure after housing, accounting for approximately 16% of total household spending (U.S. Bureau of Labor Statistics 2024). For students, low-income households, and other economically vulnerable groups, this burden is especially heavy (Blumenberg 2014). Even small changes in transportation costs or accessibility can have significant effects on educational attainment, employment stability, health status, and overall well-being (Sanchez et al. 2004).

Transportation equity has therefore emerged as a critical policy and research priority. The

concept extends beyond the provision of physical infrastructure to encompass the fairness of how transportation systems distribute benefits, costs, and risks across different population groups (Martens [2016](#)). Equity-oriented transportation planning recognizes that not all users start from the same position. Differences in income, physical ability, location, and institutional constraints mean that identical infrastructure can produce vastly different levels of accessibility for different people (Pereira et al. [2017](#)). Achieving equitable mobility thus requires tailoring transportation solutions to the needs of each vulnerable user group (Lucas and Musso [2014](#)).

Despite the obvious need for equity-based transportation considerations, much of the literature and practice in transportation planning still centers on the “average commuter” archetype: full-time, nine-to-five workers living in metropolitan areas. While this focus addresses an important segment of the population, it continues to marginalize groups whose travel patterns and constraints differ from the norm. Post-secondary students are especially notable among the overlooked groups, as they are a large, mobile, and economically constrained population whose mobility is integral to their academic success and broader life opportunities (Elizabeth M. Delmelle and Delmelle [2012](#)).

The neglect of student-specific transportation needs is concerning, given the scale of the post-secondary population and the pivotal role mobility plays in shaping educational outcomes. With over 19.4 million students enrolled in colleges and universities across the United States (U.S. Department of Education, National Center for Education Statistics [2025](#)), and more than 70% residing off-campus, reliable access to campus and related destinations is not a luxury but a necessity. Students’ ability to access classrooms, libraries, internships, part-time jobs, extracurricular activities, and other destinations directly affects not only their academic performance but also their professional readiness, social integration, and mental health (Elizabeth M. Delmelle and Delmelle [2012](#)).

A transportation system that fails to accommodate these needs risks creating barriers to

achievement. It also reinforces inequalities and undermines the societal benefits of higher education. This research is grounded in the recognition that transportation systems must be evaluated not only in terms of efficiency and environmental impact but also in their capacity to support equitable access for all users, including those whose needs diverge from conventional planning models. In particular, it seeks to illuminate the distinctive mobility challenges faced by post-secondary students, with a focus on how these challenges vary across geographic contexts and how they can be more effectively addressed through targeted modeling and policy interventions.

1.2 The Neglected Demographic: Post-secondary Students

Within the broader discourse on transportation systems, post-secondary students are under-represented in both scholarly research and policymaking agendas. Post-secondary students refer to individuals enrolled in higher education institutions beyond high school. This includes undergraduate and graduate students attending universities and four-year colleges. Students enrolled in junior colleges, community colleges, technical degree programs, or certificate programs were not included in the survey sample. Thus, the focus remains on students pursuing formal degree programs (bachelor's, master's, and doctoral levels) at universities. This is not a small or peripheral group, as over 19.4 million students are currently enrolled in colleges and universities across the United States (U.S. Department of Education, National Center for Education Statistics [2025](#)). They represent a substantial and distinct subset of transportation system users whose mobility patterns and needs diverge sharply from the general commuting population. Notably, more than 70% of students in the U.S. reside off-campus (Kelchen [2018](#)), making daily travel between home and campus an unavoidable necessity rather than a matter of personal choice. Their journeys, however,

seldom end at the classroom door. Post-secondary students also depend on transportation to reach internships, part-time jobs, healthcare facilities, grocery stores, childcare centers, recreational spaces, and other destinations essential to sustaining both academic performance and personal well-being (Blumenberg [2014](#)).

The travel behavior of this population is highly distinct and deeply sensitive to context. Unlike the relatively predictable routines of most 9-to-5 commuters, students must navigate fragmented, irregular, and often unpredictable schedules dictated by a constantly shifting mix of class timetables, laboratory sessions, extracurricular activities, and employment commitments. Their daily mobility usually follows a trip-chaining pattern that links academic, professional, and personal activities in a single travel sequence, such as attending morning lectures, working an afternoon shift, and running errands before returning home (Elizabeth M. Delmelle and Delmelle [2012](#)). This complexity is further compounded by widespread financial constraints: many students cannot afford to purchase or maintain a private vehicle and must instead rely on a patchwork of transportation modes, including public transit, ride-sharing services, cycling, walking, and informal carpooling arrangements (Islam et al. [2022](#)). Decisions about when and how to travel- or even whether to travel at all are shaped by a delicate interplay of factors that range from strict time and financial budgets to institutional scheduling policies and the spatial distribution of campus facilities. Housing affordability and location, along with the availability (or absence) of safe, accessible, and reliable transportation infrastructure are also crucial factors affecting students' travel decisions (Volosin [2014](#)). For this large but atypical group of frequent commuters, the problem is not simply getting from point A to point B and back, but getting from point A to points B, C, D, and E, back to D, and then onto F (and possibly G, if time permits), before finally returning to point A, all in the run of a day.

Despite these clear differences between post-secondary students and other commuting groups, post-secondary students are routinely collapsed into the generic “commuter” category in both large-scale travel surveys and transportation demand models. This statistical aggregation

not only erases the complexity of their travel patterns but also masks the structural, economic, and institutional constraints they navigate daily (Blumenberg 2014). As a result, the transportation strategies and infrastructure investments emerging from these data fail to align with the students' lived realities, in most cases. This misalignment can produce a cascade of negative effects, ranging from underutilized or inaccessible facilities to increased travel stress and diminished opportunities for academic engagement and social participation. In the long term, such oversights risk entrenching accessibility gaps that not only hinder educational attainment and weaken campus-community integration but also undermine student mental health and overall quality of life (Lucas and Peter Jones 2012).

1.3 Geographic Inequities: Urban vs. Rural Student Mobility

Although most post-secondary students face mobility challenges related to transportation, the nature, severity, and consequences of those challenges vary greatly depending on whether students live in urban or rural settings. In urban university environments, students may theoretically benefit from extensive transit systems; however, the practical realities often tell a very different story. Even in cities with seemingly high transit coverage, students frequently contend with service reliability issues, including delays, frequent overcrowding during peak periods, and transit routes that do not align well with off-hour academic schedules (Paudel et al. 2021). The first-mile/last-mile problem is particularly acute for students residing off-campus. Specifically, they may be miles away from a transit stop, face unsafe walking paths, or encounter poorly lit areas at night, forcing them to rely on costlier and less accessible ride-sharing options or walking long distances in sub-optimal conditions (Islam et al. 2022). Meanwhile, urban transportation landscapes are generally characterized by intense competition for space, and pedestrian and bicycle infrastructure is often either unsafe,

poorly maintained, or underprioritized compared to vehicle lanes or parking lots (Pucher and Buehler 2010). Safety concerns, especially for students traveling late at night from campus to off-campus housing or for employment or social reasons, further constrict walkability and deter the use of active modes (Kiyohara and Mondri 2025).

Like their urban peers, many students living in rural or small college-town settings also face mobility obstacles, though these challenges are fundamentally different. Public transportation, where it exists at all, is typically infrequent, poorly integrated, or even entirely lacking, leaving students with virtually no choice but private automobiles, walking, cycling, or, in extreme cases, hitchhiking (Farrington and Farrington 2005). However, dependence on cars can pose significant financial burdens, as upfront purchase costs, maintenance, fuel, insurance, and parking can be prohibitive to students on a limited budget. Furthermore, walking and cycling might be impractical, due to the spatial dispersion of rural campuses, housing, employment, and services, or they might be unsafe, due to a lack of sidewalks, bike lanes, or safe road crossings (Safe Routes to School National Partnership 2015).

These geographic inequities translate into distinct lived experiences. Urban students may face burdens of unreliable service, overcrowded transit, safety concerns, and failing first-mile last-mile connections, which can lead to missed classes, stress, and reduced participation in campus life. Rural students, meanwhile, grapple with almost total car dependency, high mobility costs, and lack of alternatives, resulting in constrained academic schedules, reduced access to off-campus opportunities, and increased exclusion from community and extracurricular engagement (Balsas 2003).

Yet despite these widely divergent realities, research and policy interventions often default to a one-size-fits-all framework, largely informed by urban cases and data. The result is the perpetuation of inequities rather than the correction of them (Pereira et al. 2017). For post-secondary students, who, as we noted above, already face unique mobility constraints, failing to account for the rural–urban divide can deepen spatial inequalities. When this

occurs, equitable access to higher education is hindered and the broader societal value of student mobility is precluded.

1.4 Gaps in Existing Literature

Despite the increasing recognition of the importance of transportation equity and user-centric mobility planning, several critical gaps remain in the existing transportation literature when it comes to understanding and modeling the travel behavior of post-secondary students. These gaps are not only methodological in nature but also reflect broader systemic oversights in transportation research priorities. The present research seeks to address three key dimensions of these gaps: the lack of student-specific travel models, the urban-rural divide in transportation research coverage, and methodological limitations in existing modeling approaches.

1.4.1 Lack of Post-Secondary Student-Specific Travel Models

Most existing transportation research has traditionally focused on work-based commuting patterns, often emphasizing the behavior of full-time workers traveling between fixed home and work locations during peak hours. While this focus is understandable given the significance of work-related travel in shaping transportation demand, it fails to capture the highly irregular, flexible, and context-dependent travel patterns exhibited by post-secondary students. Post-secondary students represent a distinct user group with mobility behaviors that differ fundamentally from standard commuting populations. Their daily travel is rarely confined to a single purpose or destination. Post-secondary students' travel patterns could be characterized as follows:

- Multiple trip purposes within a single day (e.g., academic, work, recreation, social visits,

errands).

- Varying schedules across weekdays and weekends.
- High dependence on multimodal travel that combines walking, cycling, transit, carpooling, ride-hailing, etc.
- Travel decisions that may be constrained by limited financial resources.
- Increased sensitivity to changes in infrastructure, pricing, or availability of modes.

Despite these complexities, current transportation models rarely differentiate students from the general population in their assumptions or parameters. National and regional travel surveys often lack student-specific data, leading to significant knowledge gaps regarding how students make travel decisions, especially regarding how they balance academic commitments, part-time jobs, social obligations, and personal needs within limited time and financial constraints. Without models that explicitly account for student-specific mobility patterns, transportation systems risk failing to meet the actual needs of this large and diverse population. Moreover, ignoring students in transportation planning may lead to underutilized transit services, missed opportunities for sustainable mode promotion, and increased reliance on private vehicles, particularly in resource-constrained settings.

1.4.2 Urban-Rural Divide in Research Coverage

Another critical limitation of existing research is the disproportionate emphasis on urban mobility challenges, often at the expense of understanding transportation dynamics in rural or small-town contexts. While urban areas undeniably face pressing transportation issues including congestion, transit crowding, and last-mile connectivity, the challenges experienced by students in rural college towns are equally or even more severe.

Rural college towns present a unique mobility landscape characterized by:

- Minimal or nonexistent public transportation options.
- Sparse land-use patterns that require longer travel distances.
- Poorly developed active transportation infrastructure (e.g., lack of bike lanes, sidewalks, crosswalks, etc.).
- Heavy dependence on private vehicles for daily travel.
- Financial burdens associated with vehicle ownership.
- Vulnerability to transportation-related isolation for students without access to cars.

However, transportation research has largely overlooked these contexts for rural and small-town students. Much of the existing literature is centered around large metropolitan areas where data is more readily available and transportation agencies are more active. As a result, rural students, particularly those attending institutions in geographically isolated regions, remain an underrepresented and underserved population in transportation research and policy discourse. Because most research focuses on cities, transportation policies are often shaped by urban findings and overlook the needs of rural communities. Infrastructure investment decisions are nearly always guided by findings from urban settings, leading to the adoption of strategies that may be ill-suited to rural environments. Without detailed empirical studies capturing how students in rural college towns navigate mobility constraints, transportation planning risks reinforcing spatial inequalities rather than alleviating them.

1.4.3 Methodological Limitations

A third critical gap lies in the methodological tools. Most transportation studies have historically relied on trip-based modeling frameworks, where individual trips are treated as isolated and independent events. While such approaches offer computational simplicity and a straightforward representation of travel demand, they are inherently limited in their

ability to capture the complexity, interdependencies, and adaptive nature of student mobility patterns. The on-the-ground reality is that students often organize their daily movements as part of an interlinked multi-destination trip of activities. In this multi-destination trip model, decisions made for one trip such as departure time, mode, or destination directly constrain or influence subsequent trips (McGuckin and Murakami 2005). For example, a student's choice to attend a study group before class may lead to an earlier departure, which could in turn influence their mode choice, subsequent activity duration, and the timing of return trips. Trip-based models, by treating trips independently, systematically overlook these temporal and spatial dependencies.

In addition, traditional models often rely on simplified assumptions that can over-simplify real-world travel behavior. For example, they commonly divide departure times into broad categories such as peak or off-peak periods, ignoring the continuous nature of departure decisions. They also frequently assume that each trip decision is made independently, failing to consider how the timing of one trip can influence subsequent trips throughout the day. This gap is magnified when considering departure time choice. Traditional departure time choice models, often based on static discrete choice frameworks, typically assume that travelers respond only to current or observed conditions such as travel times, congestion levels, or service frequencies (Small 1982). Such models do not capture forward-looking, anticipatory decision-making, where students consider expected future conditions such as anticipated congestion after class, parking availability, or weather changes before deciding when to travel (Bhat 1998). The failure to represent this anticipatory behavior leads to underestimation of students' flexibility in adjusting departure times, particularly in response to policy changes such as schedule shifts, parking restrictions, or transit service modifications. Our research addresses these limitations by applying advanced AI models that predict departure times on a continuous scale and capture the sequential and interdependent nature of multiple trips within a day.

Another persistent limitation is the assumption of behavioral homogeneity within the traveler

population. Many conventional models overlook the substantial heterogeneity in preferences, constraints, and behavioral elasticities within student populations. Critical differentiators such as enrollment status (full-time vs. part-time), employment commitments, residential location (on-campus vs. off-campus), income levels, vehicle ownership, gender, age, and access to alternative transport modes can significantly influence both mode and departure-time choices (Blumenberg 2014). For instance, students without private vehicles may be more sensitive to transit headways and schedule coordination, whereas car commuters may adapt departure times strategically to avoid peak congestion or compete for limited parking spaces.

From a data perspective, much of the existing literature on student travel behavior relies heavily on revealed preference (RP) data (Small 1982). These type of data capture actual choices made by individuals under prevailing real-world conditions, offering valuable insights into behavior within existing transportation systems and policy environments. However, this reliance poses limitations when the research objective involves understanding how students might adapt to new, hypothetical, or emerging scenarios, such as the introduction of micromobility services, expansion of bicycle infrastructure, widespread adoption of electric vehicles, or implementation of dynamic transit scheduling. In such cases, RP data cannot reveal preferences for options that students have not yet experienced or that are absent from their current choice set. Stated preference (SP) data, in contrast, allow researchers to systematically explore decision-making under controlled hypothetical conditions (Louviere et al. 2000). This flexibility enables the examination of scenarios that are rare, novel, or yet to be implemented, making SP particularly useful for assessing the potential impacts of future transportation policies or infrastructure investments. Nonetheless, SP methods are not without their shortcomings. For example, they are susceptible to hypothetical bias, where respondents' stated intentions may deviate from their actual behavior when faced with real-world constraints. In addition, they may suffer from limited external validity. To address these limitations, a hybrid RP-SP approach has emerged as a robust alternative,

combining the strengths of both data types. By grounding hypothetical choice experiments in observed behavioral patterns, hybrid datasets offer both the realism of RP and the flexibility of SP, allowing for richer modeling of student travel decisions in both current and prospective contexts (Rose et al. 2008). This approach is especially valuable for evaluating the likely responses of post-secondary students to emerging mobility trends, targeted policy interventions, and infrastructure enhancements, particularly when aimed at improving accessibility, equity, and sustainability in campus-adjacent transportation networks.

Finally, traditional modeling frameworks for student mobility have yet to fully embrace advanced approaches that capture complex decision-making under uncertainty and over time. Emerging methods, such as dynamic discrete choice models (DDCMs) and machine learning-based predictive frameworks offer promising avenues for integrating behavioral heterogeneity. They are also useful for combining temporal interdependencies and anticipatory decision-making into a unified analytical structure (Moshe E. Ben-Akiva and Lerman 2014; K. E. Train 2009; Vovsha et al. 2004). By explicitly modeling sequential dependencies and feedback mechanisms, these approaches can more accurately predict how students adapt to policy interventions such as congestion pricing, parking management, transit schedule redesign, or the introduction of time-shared transportation modes like bicycles, scooters, or cars.

In sum, failing to incorporate behavioral heterogeneity, activity-tour interdependence, anticipatory decision-making, and joint decision frameworks leaves policymakers with predictive tools that may be systematically biased or incomplete. Addressing these methodological limitations is essential for designing targeted, context-sensitive transportation policies and infrastructure investments that enhance accessibility, affordability, equity, and sustainability for post-secondary students.

1.5 Research Objectives

Post-secondary students represent a substantial yet underexamined segment of the traveling public whose daily mobility patterns are shaped by unique institutional, economic, employment, and social constraints. Their travel decisions are rarely the outcome of isolated choices but are instead made based on complex interdependent processes that link activity type, mode, departure time, and destination within the framework of multi-destination daily trips. Traditional transportation research and planning tools often fail to capture these nuanced dynamics, leading to policies and infrastructure investments that inadequately address post-secondary students' specific transportation needs. This research seeks to fill the existing research and planning gaps through a series of empirical investigations. The overarching objective is to develop and apply modeling frameworks that account for the sequential, context-sensitive nature of student travel decisions across diverse geographic settings. Specifically, the research pursues the following objectives:

- **To model multi destination trip-based activity type choices in the urban context using dynamic discrete choice models.** Examines how students structure their daily tours by selecting different activities across multiple destinations. Using dynamic discrete choice models, the study captures the interdependence between successive activity choices, showing how earlier commitments influence subsequent decisions. This provides a richer and more realistic framework compared to traditional trip-based models.
- **To analyze sequential departure time choices in multi-destination trips through deep learning approaches (GRU, LSTM, Transformer).** Departure time is treated as a primary decision in multi-destination trips, capturing how one trip influences the next. Instead of broad peak/off-peak categories, it predicts departure times on a continuous scale. Using GRU, LSTM, and Transformer models, it captures

complex sequential patterns, aiming for behavioral realism that reflects real-world student travel dynamics.

- **To evaluate the influence of bike infrastructure and active transportation on student mobility using a joint RP-SP survey in rural settings.** Active modes such as cycling and walking play a crucial role in rural college towns where transit options are limited. This objective evaluates how infrastructure availability, perceived safety, cost etc. affect students' adoption of active modes. Using a joint RP-SP survey framework, the study highlights how infrastructure improvements enhance mobility for underserved post-secondary student populations.
- **Examine the intersection of transportation systems and post-secondary students' well-being.** The research extends the analysis of travel behavior to consider health and mental health outcomes associated with different modes, particularly active transportation. By linking transportation choices to well-being, it provides a more holistic understanding of the implications of planning and policy decisions for student populations.
- **To investigate the mental health benefits of active transportation using data-driven modeling approaches.** This objective extends the analysis of travel behavior to include outcomes related to student well-being. Data-driven modeling approaches are used to examine how active transportation modes (e.g., walking and cycling) are linked to reduced stress.

Through these interconnected objectives, the research not only enhances theoretical understanding of post-secondary students' travel behavior but also produce actionable insights for transportation planners, campus administrators, and policymakers. The expected outcomes directly support the design of equitable, efficient, and sustainable transportation systems that are responsive to the realities of student life across both urban and rural settings.

1.6 Methodological Overview

The present research employs a multi-method research design integrating econometric modeling, machine learning, and combined revealed–stated preference analysis to examine the sequential and context-sensitive nature of post-secondary students’ travel behavior. The methods are applied across four complementary research components, as listed below:

- Utilizes the StudentMoveTO survey from the Greater Toronto and Hamilton Area (GTHA) to implement a multi-destination trip-based Dynamic Discrete Choice Model (DDCM) and Dynamic Discrete Mixed Choice Model (DDMCM). Some of the key features include:
 - Captures interdependencies among activities in two- and three-destination trips.
 - Models how early-day activity decisions reshape subsequent travel patterns.
 - The model accounts for socio-demographics and a range of other considerations.
- Applies deep learning architectures including Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM) networks, and Transformer-based models to predict departure times for each trip stage in a multi-destination trip. The main features are:
 - Models temporal interdependencies between earlier and later trips in a day.
 - Predicts departure times on a continuous scale for greater behavioral realism.
- Designs and administers a revealed and stated preference survey in Virginia college towns to evaluate mode choice under different infrastructure scenarios.
 - Uses RP-SP MNL and RP-SP mixed logit models to assess the role of infrastructure, safety perceptions, weather, etc.
 - Focuses on the adoption of cycling, and walking options where public transit is limited.

- Produces policy-relevant insights for expanding non-motorized travel in rural contexts.
- Investigates the relationship between cycling and student well-being using machine learning and statistical techniques.
 - Employs Random Forest (RF) models, Principal Component Analysis (PCA), and SHAP values to identify influential factors.
 - Examines how factors such as cycling frequency, duration of cycling, socio-demographics, etc., affect mental health.
 - Provides a holistic view of the societal and health implications of active travel.

This methodological approach evaluates both the interpretability of econometric models and the predictive capabilities of machine learning. It also spans both urban and rural contexts and links travel behavior to broader well-being outcomes. The resulting insights offer robust policy-relevant guidance for designing sustainable transportation systems for student populations in both urban and rural settings.

1.7 Contributions to Knowledge

This research makes several significant and novel contributions to the field of transportation research and mobility behavior modeling. Along with bridging crucial knowledge gaps at the intersection of transportation planning, behavioral analysis, public health, and geographic equity, it focuses on a population that has historically been underrepresented in transportation literature: post-secondary students.

1.7.1 Centering Post-secondary Students in Urban-Rural Mobility Research

One of the most important contributions of this research is its unique focus on post-secondary students as a distinct and critical population within the transportation ecosystem. While previous research has extensively examined the mobility behavior of 9-to-5 commuters, workers, urban residents, and even tourists, very few studies have sought to systematically analyze and compare the travel patterns of college and university students across urban and rural settings. This research is among the first to position post-secondary students at the center of a comparative urban-rural mobility study. It moves beyond treating students as a homogeneous or peripheral group within broader transportation datasets. Instead, it highlights students' unique mobility constraints, behavioral patterns, and decision-making processes, which are shaped by multi-factorial considerations such as academic schedules, financial limitations, modal preferences, and residential locations. By bringing students to the forefront of transportation analysis and explicitly contrasting their mobility behavior in urban versus rural contexts, this research offers a much-needed contribution to the discourse on spatial equity that is rooted in educational accessibility and sustainable mobility planning.

1.7.2 Advancing Dynamic and Mixed Econometric Modeling in Student Mobility Analysis

Methodologically, this research advances the application of sophisticated econometric models to capture the behavioral complexity of student travel choices. Traditional static models often fail to reflect the dynamic, interconnected, multi-activity, and multi-moded nature of student travel. To overcome this, the research integrates advanced modeling techniques such as Dynamic Discrete Choice Models (DDCM), Mixed Logit Models (MxL), Joint Revealed and Stated Preference Frameworks, and multi-destination trip-based Activity Choice Models.

These approaches not only allow for the modeling of interdependent trips and behavioral heterogeneity but also look at long-term decision-making processes. In this way, they can more aptly capture how students' choices evolve not only within a single day (across multiple trips) but also over time in response to changing contexts, costs, policies, and infrastructure availability. This research therefore contributes to the methodological advancement of travel behavior analysis, offering robust tools that are sensitive to both the temporal dynamics and individual-specific variations inherent in student mobility.

1.7.3 Advancing Sequential Departure Time Modeling

Most post-secondary students schedule their various trips within multi-destination daily trips. By framing departure time as a sequential decision based on two-destination trips, cascading effects are captured. These show where one departure influences the timing of subsequent activities, thereby offering a more realistic representation of daily travel shaped by flexible schedules, interdependent commitments, competing priorities, travel mode availability, and cost. Methodologically, it integrates advanced deep learning techniques, including Gated Recurrent Units (GRU), Long Short-Term Memory networks (LSTM), and Transformer models, with behavioral modeling to incorporate factors such as socio-demographics, trip characteristics, dynamic pricing, and temporal context. This approach enhances predictive accuracy within the various models' complex nonlinear relationships, bridging the gap between traditional econometric frameworks and modern data-driven methods.

1.7.4 Bridging Transportation Behavior with Mental Health and Well-being

A particularly innovative contribution of this research lies in its interdisciplinary approach linking transportation behavior with health outcomes, especially mental health and well-

being. While previous studies have examined the environmental or physical health benefits of active transportation, very limited research has explicitly investigated the psychological or emotional impacts of travel behavior, particularly among students in rural college towns. Through the application of machine learning techniques such as RF models and SHAP analysis, this research empirically investigates how active travel modes like cycling contribute to mental health improvements, including stress reduction, better sleep quality, anxiety management, improved interrelationships, and overall emotional well-being. This integration of transportation analysis with public health perspectives offers critical insights not only for transportation planners but also for campus administrators, mental health professionals, policymakers, and any other stakeholders who are seeking holistic strategies to improve student well-being through mobility interventions.

1.8 Structure of the Dissertation

The remainder of the dissertation is organized as follows:

- Chapter 2 presents a critical review of the literature on travel behavior modeling, student mobility, urban-rural transportation equity, and health-related outcomes of active transportation.
- Chapter 3 details objective 1, focusing on multi-destination trip-based activity type choices in the urban context, using dynamic discrete choice models.
- Chapter 4 looks at objective 2, discussing the sequential departure time choice modeling within multi-destination daily trips. The chapter applies econometric and deep learning approaches (GRU, LSTM, Transformer) to capture temporal decision-making patterns.
- Chapter 5 introduces objective 3, examining the role of bike infrastructure and active

transportation in enhancing student mobility. The research involves the design and modeling of a joint RP-SP survey in rural settings.

- Chapter 6 presents objective 4, investigating the mental health benefits of active transportation using data-driven models.
- Chapter 7 synthesizes findings across the four studies, suggests potential applications, presents policy implications, and outlines promising directions for future research.

Chapter 2

Literature Review

2.1 Chapter Overview

This section synthesizes the research across three critical areas of transportation behavior modeling: activity type choice within activity-based frameworks, mode choice behavior in rural college towns, and the mental health implications of cycling in rural settings. By integrating econometric modeling with behavioral insights, the reviewed studies advance our understanding of how individual decisions are influenced by infrastructure, personal preferences, and contextual constraints.

The first set of literature emphasizes the importance of activity type modeling within multi-destination trip-based systems. Prior research reveals that simplifications in activity representation (e.g., only considering the main activity in a multi-destination trip) can distort forecasts and lead to misguided policy decisions. Dynamic Discrete Choice Models (DDCM), Mixed Logit, and Nested Logit models have been increasingly used to better capture the sequence, interdependency, and behavioral complexity of activity engagement, especially among student populations. In this section, we categorize the literature review into three segments: activity type choice modelling within activity-based models, activity type choice modelling within multi-destination trip-based models, and activity type choice modelling within trip-based models.

The second group of studies addresses mode choice in rural college towns, highlighting the challenges students face due to inadequate infrastructure and car dependency. Existing lit-

erature points to the strong influence of cost, distance, and safety on students' travel mode selection. Notably, the presence of protected bicycle infrastructure is shown to significantly increase the likelihood of cycling, whereas standard or shared lanes offer limited encouragement. Demographic factors such as gender, income, and vehicle ownership are also critical in shaping travel decisions.

The third body of literature explores the psychological and emotional benefits of biking, with a specific focus on rural communities. Although cycling is known to reduce stress, improve sleep, and foster social connection, rural residents often lack the infrastructure to safely realize these benefits. Studies employing machine learning techniques—such as Random Forests, PCA, and SHAP—reveal that mental health gains from cycling are strongest when high-quality infrastructure is present and when cycling is embedded in daily routines.

Taken together, these studies establish a foundation for the development of a comprehensive modeling framework that integrates multi-destination trip-based activity choices, multi-modal travel behavior, and well-being outcomes, particularly in rural and under-resourced post-secondary settings. These insights guide the formulation of the empirical models and conceptual framework in the subsequent chapters.

2.2 Activity type choice modeling within activity-based models

A substantial number of scholars have engaged in the study of activity type choice modeling within Activity-Based Models and demonstrated how daily routines influence travel decisions. For example, Yuchen et al. used the dynamic discrete choice model (DDCM) to understand daily activity-travel planning. They developed a version of the DDCM that integrates unobserved heterogeneity that enables the model to reveal critical details about the characteristics of the travellers (Song et al. [2022](#)). One study implemented an activity-based

model that integrates individual behavior simulation with demand and supply dynamics across different time schedules (Luo et al. 2024). The objective of this model is to capture the complex interdependencies between different activities and travel decisions when forecasting travel demand. One study introduced an advanced modeling framework (BayABM) that integrates activity-based models with Bayesian networks and population synthesis to enhance the simulation of individual mobility dynamics (Toaza and Esztergár-Kiss 2023). It particularly focuses on the sequence of activities and their associated travel attributes, which are crucial for understanding individual travel behaviors within an activity-based modelling framework. Bladimir Toraza looked into optimizing activity scheduling using a novel approach known as Activity Chain Optimization (ACO), which is designed to optimize travel time during daily activity schedules (Hesam Hafezi et al. 2022). The study demonstrates that this optimization results in shorter travel times compared to non-optimized scenarios. He also suggests that travelers can achieve significant time savings and emission reductions by adopting the ACO method in their daily activities. One researcher introduced the SALT model, which is designed to forecast and replicate individuals travel behavior by modeling their activity choices and sequences (Pougala et al. 2023). The SALT model effectively constructs 24-hour activity schedules that reflect realistic travel patterns. It was validated with a high accuracy against real-world data from the Halifax Space Time Activity Research (STAR) household survey. Another researcher focused on how individual behaviors and decisions in activity scheduling can extend beyond a single day (Oh et al. 2020). Result shows that incorporating behavioral dynamics over multiple days allows for a more realistic simulation of individual travel and activity patterns. Another study adopted an activity-based model system to analyze the demand side of transportation and integrated it with agent-based simulation techniques to evaluate the impacts of Automated Mobility-on-Demand (AMOD) in urban settings like Singapore (Scheffer et al. 2021). It finds that while Automated Mobility-on-Demand (AMOD) can potentially reduce congestion and car ownership, it could also increase total vehicle kilometers traveled due to induced demand.

One study explored how the optimization of daily activity choices within an activity-based modeling framework using a genetic algorithm can impact the efficiency of travel times (Vyas et al. 2015). It includes the successful application of a genetic algorithm to optimize activity scheduling because it results in significant travel time reductions for daily activities. Vyas et al. focused on the allocation of individual nonmandatory activities into various day segments as part of multi-destination trip formation by considering the constraints and priorities that influence these decisions within the structure of activity-based models (Hoang et al. 2022). One researcher also focused on the latest advancements in activity-based models by highlighting how these models are used to simulate the complex travel behavior based on the activities scheduled for the day (Doherty and Mohammadian 2011).

2.3 Activity type choice modeling within multi-destination trip-based and tour-based models

Some researchers have engaged in the study of activity type choice modeling within multi-destination trip-based Models. One study examined the hierarchical structure of activity scheduling within multi-destination trip-based models, particularly focusing on non-work activities and their scheduling in motorcycle-based cities (MBCs) (Doherty and Mohammadian 2011). The study reveals that due to mixed land use and high amenity density, intrazonal multi-destination trips are prevalent, influencing the scheduling and structuring of daily activities significantly. Another study investigated the decision-making process of individuals choosing their travel locations dynamically and considered both spatial and temporal constraints within a multi-destination trip-based framework (Garikapati et al. 2014). This involved modeling the types of activities and their locations as part of an integrated approach to understanding travel behavior. One study discussed the interrelated decisions of mode choice and timing within the framework of multi-destination trip-based models and

highlighted how the timing and mode of transport can influence the sequence and choice of activities during a multi-destination trip (Daisy et al. 2018). The study found that adjustments in either the timing of activities or the selection of transportation mode can substantially affect the other. Another study conducted research on how individuals allocate their time across various activities within a multi-destination trip, integrating the choices of activity type and duration using a multiple discrete-continuous extreme value (MDCEV) modeling approach (Moshe E Ben-Akiva and Lerman 1985). The model is robust in accounting for different socio-economic, demographic, and built environment influences on activity participation and time allocation. Yagi and his team members presented a unique approach by incorporating behavioral rules into the microsimulation of activity and travel tours and examined how individuals schedule and participate in daily activities based on observed behavior probabilities (McFadden 1974). They included the development and estimation of a nested logit model structure that evaluates the daily activity-travel patterns of individuals. One team adopted a cluster-based approach to model the complexities of trip chaining and mode choice decisions among non-worker populations based on their daily activities (K. E. Train 2009). It showed that specific clusters of activities lead to distinct patterns in mode selection and trip chaining. However, this study stands out due to its unique focus on comprehensively understanding the daily activity type choice patterns of post-secondary students. Unlike previous research, which has largely overlooked this population cohort, our study our study investigates the specific behaviors and preferences of these students and provides valuable insights that have not been previously explored by other researchers.

2.4 Review of Modeling Approaches for Sequential Departure Time Decisions

Researchers have studied departure time choice using many different methods. Each method provides a different way of understanding how people decide when to begin their trips (Bhat and Steed 2002; Habib 2013; Jou et al. 2008; Thorhauge 2015). This section categorizes the existing research on departure time choice into three main themes based on the modeling approach: (1) Discrete Choice Models (DCMs), (2) Optimization and Game-Theoretic Models and (3) Deep Learning Approaches.

2.4.1 Discrete Choice Models (DCMs)

A substantial body of research has modeled departure time choice using traditional econometric frameworks such as multinomial logit (MNL), mixed logit (ML), and latent class models (LCM). For example, Cheng et al. (2020) investigated metro passengers' departure time choice in Nanjing, China, using a latent class logit model with elasticity analysis to understand the effect of various parameters (Q. Cheng et al. 2020). The latent class logit model was estimated using the collected data, and an elasticity analysis was incorporated to learn about the impact of all the parameters (Q. Cheng et al. 2020). The strength of this study is its ability to capture group-level behavioral differences and provide useful elasticity insights. However, a key limitation is that it relies on stated preference data and does not capture continuous heterogeneity within classes. Ben-Elia et al. (2010) discretized time into 15 minute intervals and analyzed departure time choices using data from the Dutch Spitsmijden experiment, which tracked 341 commuters over 13 weeks to evaluate how monetary rewards and smartphone credits influence peak-hour avoidance (Ben-Elia et al. 2010). Using multinomial logit, mixed logit, and latent class models, the authors found that both incentive types effectively reduced peak-hour driving, with elasticities close to -1 . The la-

tent class model revealed that individuals with fixed schedules such as women with childcare responsibilities and no flexible work hours, were more likely to have a preferred arrival time and less likely to adjust behavior. While the multinomial logit model was easy to interpret, it lacked flexibility, the mixed logit model improved fit by capturing heterogeneity but was computationally intensive, and the latent class model allowed for behavioral segmentation, though at the cost of increased estimation complexity. Cheng et al. (2020) developed a mixed logit model to estimate the departure time choice behavior of individuals who travel by rail transit on a regular basis (Y. Cheng et al. 2020). This study examined how perceived crowding affects commuters' departure time choices using stated preference survey data from 406 metro users in Nanjing, China. The authors applied a mixed logit model to capture individual heterogeneity and found that crowding significantly influences departure time choices, with many commuters willing to shift to earlier or later times to avoid crowded trains. The data included socio-demographics, trip attributes, and crowding levels across different departure time alternatives. The mixed logit model was effective in capturing random taste variation and offered better model fit, but it was computationally intensive and required complex estimation. Thorhauge et al. (2021) adopted a Mixed Logit Model to understand the departure time preference of car commuters in Denmark (Thorhauge, Vij, and Cherchi 2021). This study investigates how time-of-day pricing affects departure time choices using GPS trajectory data from 1,161 car commuters in the Netherlands. A mixed logit model reveals that higher tolls during peak hours significantly deter peak-period travel, with notable heterogeneity in price sensitivity across individuals. The model accounts for random taste variation and provides flexibility, but at the cost of complex estimation. Another research adopted a mixed MNL to get insight into the car commuters' departure time choice (Zannat et al. 2022). This study models the departure time choice of car commuters in Dhaka, Bangladesh using revealed preference data from 260 individuals. A multinomial logit model was used to analyze how socio-demographic and travel-related factors influence peak, shoulder, and off-peak departure choices. The model revealed that variables like trip pur-

pose, travel time, and vehicle ownership significantly affect departure time decisions. While the multinomial logit model is simple and easy to estimate, it assumes independence of irrelevant alternatives (IIA) and cannot account for unobserved heterogeneity across individuals. Afandizadeh et al. (2020) used the MNL model incorporating clustering methods and compared the models (Afandizadeh Zargari and Safari 2020). This study analyzes departure time choice behavior of car commuters in Tehran using revealed preference data from 588 individuals. The authors used a clustered multinomial logit (MNL) model, where k-means clustering was applied first to segment commuters into homogeneous groups before estimating separate MNL models. The approach improved model fit and interpretability compared to a single MNL model. While the clustered MNL is computationally straightforward and enhances behavioral insights, it does not capture continuous heterogeneity like mixed logit models and may oversimplify class boundaries. Many studies investigated departure time choice to see how individuals change their departure time choice to avoid traffic congestion, road toll, crowding, or infectious disease. Singh et al. (2023) used the latent class model along with two marginal rates of substitution (Singh et al. 2023) to learn about the departure time behavior of travelers during the COVID-19 pandemic. His study examines departure time choice behavior of train commuters in the Netherlands using a large-scale stated preference survey of 1,025 respondents. The authors applied an error component latent class choice model to capture both observed and unobserved heterogeneity in preferences related to crowding, reliability, and schedule delays. The model revealed that commuters value reliability and crowding reduction differently across latent classes and that policies targeting specific groups can better influence travel behavior. While the model offers rich behavioral understanding and flexibility, it is complex to estimate and interpret compared to standard models.

2.4.2 Optimization and Game-Theoretic Models

A few studies have adopted optimization and game-theoretic approaches to investigate departure time choices, particularly under competitive or constrained environments. Xue et al. (2019) examined how parking space shortages and parking costs influence commuter departure time decisions using a game-theoretic approach combined with simulation (Xue et al. 2019). This study investigates how parking space shortages and parking costs influence commuter departure time decisions using a game-theoretic approach combined with simulation. Based on a hypothetical network and simulation experiments, the model captures how drivers adjust departure times in response to limited parking and pricing strategies. It helps explain how people change their travel behavior under parking constraints. However, it lacks empirical validation and may oversimplify real-world behavioral complexities. A few studies tackled the departure time choice using unconventional theories. One study estimated commuters' departure time choice behavior empirically, where undecided congestion charge was an important scenario. This study used Cumulative Prospect Theory (CPT) to determine the departure time choice behavior of people (Geng et al. 2023). The strength of this study is that it uses behavioral factors like loss aversion and risk perception. It provides a more realistic understanding of commuter decisions. However, a key limitation is its reliance on stated preference data with hypothetical scenarios, which may limit the accuracy of real-world behavior prediction. Ameli et al. (2022) proposed a framework to estimate the departure time behavior of players, and Mean Field Game (MFG) theory, along with the bathtub model was incorporated into the framework (Ameli et al. 2022). The main advantage of this approach is that it can handle interactions among many commuters and account for different traveler preferences. It also provides mathematically stable solutions and produces faster results compared to traditional models. However, it relies on strong theoretical assumptions about traveler homogeneity and lacks direct empirical validation using real-world travel data.

2.4.3 Deep Learning Approaches

In contrast to traditional models, very few studies have used deep learning techniques to model departure time choices. Gu et al. (2024) proposed a deep reinforcement learning framework using a multi-agent actor-critic approach to jointly model travelers' mode and departure time choices in dynamic multimodal transportation networks (Gu et al. 2024). It outperforms traditional models by adapting to traffic changes and capturing complex behaviors, though it is computationally intensive and less interpretable. Another researcher proposes a deep learning approach using Long Short-Term Memory (LSTM) networks to predict flight departure delays based on U.S. flight data (H. Zhou et al. 2022). The LSTM model outperforms traditional machine learning models like Random Forest and XGBoost by capturing temporal dependencies in flight operations. The model improves prediction accuracy and handles complex time-series data well. But it requires large datasets, is computationally intensive, and offers limited interpretability. After carefully reviewing the existing literature, it is clear that several important gaps remain unaddressed. First, many studies treat departure time as a secondary outcome rather than a primary decision, specially in multi-destination trips where the timing of one trip affects the next. Second, much of the literature relies on simple assumptions like broad time categories and independent trip decisions and fails to capture the complexity of real-world travel. Third, few studies use data-driven approaches like deep learning that can automatically identify complex patterns without relying on fixed formulas, which are the main contributions of this study.

Table 2.1: Summary of Departure Time Choice Modeling Studies

Author	Data Used	Model Used	Pros	Cons
(Q. Cheng et al. 2020)	406 metro users (SP survey, Nanjing, China)	Latent Class Logit	Captures class-level behavior and performs elasticity analysis	Cannot capture continuous heterogeneity; class segmentation may oversimplify
(Ben-Elia et al. 2010)	341 commuters (Dutch Spitsmijden experiment)	MNL, Mixed Logit, Latent Class	Captures heterogeneity and behavioral segmentation	MNL assumes IIA; Mixed Logit is computationally intensive; LCM complex to estimate
(Y. Cheng et al. 2020)	406 metro users (SP survey)	Mixed Logit	Captures random taste variation, better model fit	Computationally intensive, complex estimation
(Thorhauge, Vij, and Cherchi 2021)	1,161 car commuters (GPS trajectory data, Netherlands)	Mixed Logit	Captures heterogeneity in toll sensitivity	High computational complexity
(Zannat et al. 2022)	260 car commuters (RP data, Dhaka)	Multinomial Logit	Simple and easy to estimate	Assumes IIA, can't capture unobserved heterogeneity
(Afandizadeh Zargari and Safari 2020)	588 car commuters (RP data, Tehran)	Clustered MNL (MNL + K-means)	Improves model fit and behavioral insight	Lacks continuous heterogeneity, oversimplifies segments
(Singh et al. 2023)	1,025 train commuters (SP survey, Netherlands)	Error Component Latent Class Choice	Captures observed and unobserved heterogeneity, rich insights	Complex to estimate and interpret
(Xue et al. 2019)	Simulation on hypothetical network	Game-Theoretic + Simulation	Provides theoretical insights into parking behavior	Lacks empirical validation, may oversimplify real-world behavior
(Geng et al. 2023)	SP data on congestion charge scenario	Cumulative Prospect Theory (CPT)	Models behavior under uncertainty and loss aversion	Less conventional, more difficult to validate empirically
(Ameli et al. 2022)	Simulated players	Mean Field Game (MFG) + Bathtub Model	Captures strategic interactions in large populations	Highly theoretical, lacks real-world calibration

2.5 Active Transportation and Mode Choice in Rural College Towns

Mode choice is a complex decision-making process influenced by a combination of infrastructure availability, travel distance, safety concerns, socio-economic status, personal preferences, and institutional policies. While urban areas have seen significant shifts toward sustainable travel modes, students in rural college towns continue to rely heavily on private automobiles due to the absence of viable alternatives. Zhou et al. conducted a comprehensive study on student travel behavior, emphasizing the relationship between housing location and mode choice. One of the most influential determinants of student mode choice is residential proximity to campus. It was found that students living closer to their university are significantly more likely to walk or bike, whereas those residing in off-campus housing or suburban areas overwhelmingly rely on private vehicles (J. Zhou et al. [2018](#)). Similarly, Delmelle and Delmelle (2012) examined commuting behaviors at a university and found that students preferred walking and driving over cycling due to safety concerns and poor cycling infrastructure (Eric M Delmelle and Delmelle [2012](#)). The availability and quality of public transportation also influence student travel decisions. A study found that students were more likely to use public transit when bus services were frequent, reliable, and well-integrated with walking and cycling networks. However, in rural college towns, where bus networks are often infrequent and poorly maintained, private vehicle ownership becomes a necessity (Shaaban and Kim [2016](#)).

Socioeconomic factors, particularly housing affordability, significantly impact student transportation choices. Research indicates that students who can afford on-campus or near-campus housing are more likely to use active transportation modes, while those living in lower-cost housing further from campus are generally forced to rely on cars (J. Zhou [2014](#)). This pattern was confirmed in an Iowa State University study, which found that students

living in affordable off-campus housing tended to prioritize public transit proximity when making housing decisions (J. Zhou et al. 2018). Personal perceptions of safety and comfort further influence students' willingness to bike. Páez and Whalen (2010) found that students who perceive cycling as unsafe or inconvenient are significantly less likely to consider it as a viable travel option (Páez and Whalen 2010). Gender differences also play a role, with female students being more likely to drive due to safety concerns related to cycling infrastructure and road conditions (Eric M Delmelle and Delmelle 2012). Despite these challenges, research suggests that investment in high-quality bicycle infrastructure has the potential to shift student travel behavior.

Previous studies have demonstrated that the presence of designated bicycle infrastructure significantly enhances cyclist safety. A research team in Melbourne, Australia, conducted a study revealing that motorists maintain a greater lateral passing distance when overtaking cyclists on roads with clearly marked bike lanes (Nolan et al. 2021). This conclusion was derived from an analysis of data collected through sensors and cameras mounted on participants' bicycles. The findings were further substantiated by a systematic review of related studies examining the impact of road markings on vehicle passing behavior concerning cyclists (Rubie et al. 2020). Moreover, the Melbourne study indicated that driver behavior became more predictable when bicycle lanes were present, as the variability in passing distances decreased significantly compared to roads that lacked marked or dedicated cycling infrastructure. A study conducted by researchers at the University of Massachusetts investigated how protected bike lanes influence driver behavior and cyclist safety using eye-tracking technology (Deliali et al. 2021). The findings revealed that, compared to standard painted bike lanes, drivers tend to reduce their speed when traveling alongside protected cycling infrastructure, contributing to improved safety conditions. However, the study also noted that drivers paid less visual attention to cyclists within protected lanes, suggesting a potential drawback in awareness levels. Interestingly, at intersections with protected bicycle crossings, driver behavior shifted—motorists not only reduced their speed but also increased

their visual engagement with cyclists, indicating that protected infrastructure at intersections enhances cyclist visibility and overall safety. These findings were further supported by an experiment conducted at the University of Iowa, which analyzed driver-cyclist interactions at both protected and unprotected intersections (Subramanian et al. 2020). The study concluded that infrastructure designed to physically separate cyclists from vehicles significantly reduces potential conflict points by improving drivers' ability to detect and respond to cyclists more effectively.

A research team conducted a comprehensive safety evaluation of different bicycle lane configurations using previously published crash data to assess their effectiveness in reducing cyclist-vehicle collisions (Morrison et al. 2019). The study analyzed multiple bike lane designs, including exclusive bike lanes positioned between the travel lane and the parking lane, shared parking and bicycle lanes, curbside bicycle lanes reserved solely for cyclists, and marked curbside lanes allowing for shared use by both vehicles and bicycles. The findings indicated that bicycle lanes providing the greatest physical separation from vehicular traffic, such as designated curbside lanes, were the most effective at reducing collision rates, particularly on roads with higher speed limits. However, shared parking and bicycle lanes introduced a significant risk of “dooring” incidents, where cyclists collided with opening vehicle doors—a hazard that has been widely documented in prior studies (Lawrence et al. 2018; Schimek 2018). Data from multiple cities indicate that dooring collisions account for 12 % to 27 % of all reported urban bicycle-vehicle accidents, highlighting the dangers of poorly designed shared lanes. To mitigate this risk, many transportation design manuals recommend incorporating a buffer zone of several feet between the parking lane and the bike lane, creating a safer environment for cyclists (Chand n.d.). However, the study ultimately concluded that bicycle lane configurations featuring a physical barrier separating cyclists from vehicle lanes provide the highest level of safety, reinforcing the importance of dedicated and protected bike infrastructure in urban and high-speed traffic environments.

Researchers at Portland State University conducted a comprehensive study examining the

impact of protected bicycle infrastructure on ridership levels across five major U.S. cities: Austin, Chicago, Portland, San Francisco, and Washington D.C.. Their findings revealed a notable increase in bicycle mode share, with ridership surging between 21% and 171% following the installation of protected bike lanes (Monsere et al. 2014). The study employed a multi-method approach, utilizing official city bicycle count data, field observations, and structured interviews with both cyclists and drivers to assess the effectiveness of these facilities. Surveys conducted across all five cities indicated that residents were significantly more inclined to use bicycles when adequate cycling infrastructure, particularly protected lanes, was available. These results were further validated by a follow-up survey that confirmed similar trends in cycling behavior (McNeil et al. 2015). Similarly, a study conducted at the University of California, Berkeley and UCLA found that bicycle usage on and around college campuses has been steadily increasing, particularly as universities implement policies that encourage multimodal transportation accessibility (Loukaitou-Sideris et al. 2014). However, alongside this rise in ridership, the study also documented an increased risk of cyclist-involved crashes, with a significant proportion of these incidents occurring on multi-use paths shared by pedestrians, cyclists, and other small, non-motorized transportation modes. Follow-up research at California State University – Sacramento and Clemson University supported these findings, suggesting that many cycling-related accidents on college campuses go unreported, further underscoring the need for dedicated cycling infrastructure within university settings (Dobbs 2009; Medury et al. 2019). Beyond university settings, a study analyzing injury and fatality data in New South Wales, Australia, examined the risks associated with bicycle-pedestrian shared pathways versus shared vehicle-bicycle roadways (Chong et al. 2010). The findings revealed that cyclist fatalities were highest when bicycles and motor vehicles shared the same road space, while shared pedestrian pathways also presented a significant risk of serious injuries, particularly among vulnerable populations such as elderly pedestrians and cyclists. A separate study further supported these findings, concluding that the implementation of protected cycling infrastructure would significantly enhance safety for both elderly

cyclists and pedestrians, reducing conflict points and accident severity (Velasco et al. 2015).

Table 2.2: Comparative Overview of Literature

Authors	Findings	Model Used	Pros	Cons
(Eric M Delmelle and Delmelle 2012)	Commuting patterns and modal choices based on spatial and temporal factors.	GIS and spatial analysis	Deep spatial insights into commuting trends.	Limited by spatial data resolution.
(J. Zhou 2014)	College town students prefer non-motorized modes due to compact form.	MNL and NLM	Captures mode choice well.	Needs extensive data.
(J. Zhou et al. 2018)	Housing and commuting are linked; pass subsidies affect mode.	Confirmatory framework	Accounts for housing-commute link.	May miss external factors.
(Shaaban and Kim 2016)	Bus quality (stops, drivers) impacts student choices.	SEM	Connects quality with usage.	Survey bias risk.
(Iamtrakul and Raungratanaamporn 2015)	University affects city transport development.	Factor and cluster analysis	Reveals urban mobility patterns.	Case-specific.
(Balsas 2003)	Promotes non-motorized travel and TDM on campuses.	Case study analysis	Good for campus planning.	Hard to generalize.
(Nie 2020)	SAVs could reduce car reliance in college towns.	Agent-based + social media	Forecasts mobility shift.	Relies on AV adoption.
(Monsere et al. 2014)	Protected lanes boost cycling and safety perception.	Survey + ridership data	Strong real-world impact.	Limited to 5 U.S. cities.

Table 2.2: Comparative Overview of Literature (continued)

Authors	Findings	Model Used	Pros	Cons
(Oldread 2011)	Transit ridership tied to density and income.	Regression analysis	Links urban form with transit use.	Misses behavior dynamics.
(McNeil et al. 2015)	Buffers (e.g., flex-posts) raise cycling comfort.	Survey + experiment	Compares multiple designs.	Short-term focus.
(Aultman-Hall et al. 2023)	Travel mode and destination are interdependent.	Binary logistic regression	Captures joint decision-making.	Limited to inter-city context.
(Müller et al. 2008)	School location, car access, and weather affect mode.	MNL	Education-specific insights.	Location-specific.
(Deliali et al. 2021)	Protected intersections improve driver caution.	Simulator experiment	Shows driver adaptation.	Less focus on cyclists.

Across these studies, a recurring theme emerges: the critical role of transportation infrastructure in shaping mode choice behavior and promoting safer, more sustainable travel options. However, despite extensive research on urban cycling and transit systems, there remains a significant gap in understanding how mode choice behavior and infrastructure demand differ between urban and rural college towns. This study bridges the gap investigating the mode choice behavior of rural college town students and the impact of safe bicycle infrastructure on their mode choice.

2.6 Mental Health Impacts of Active Transportation in Rural Communities

2.6.1 Transportation and Mental Health

Individuals' everyday transportation choice has an intrinsic impact on shaping overall well-being, extending far beyond the basic need to travel from one place to another. A well-functioning transportation system not only provides physical mobility but also significantly influences mental and emotional health. Easy and reliable access to transportation allows individuals to connect with important aspects of daily life, including employment opportunities, educational institutions, healthcare services, and social activities, all of which contribute to improved quality of life and greater life satisfaction (Giles-Corti et al. 2016). On the other hand, when transportation systems are inadequate, characterized by long and stressful commutes, a lack of public transit options, or over-reliance on private vehicles, people are more likely to experience negative mental health outcomes. These transportation barriers can lead to increased stress levels, feelings of social isolation, and lower overall well-being, especially among those who have limited mobility choices (Clark et al. 2020; Lucas, Mattioli, et al. 2016; J. Zhang, Hayashi, and Frank 2021). In contrast, people who regularly use active transportation modes such as walking or cycling, tend to report better mental health and greater overall well-being. Active travel provides built-in opportunities for physical activity, which is well-known to reduce stress, improve mood, and enhance cognitive functioning. Moreover, active transportation allows individuals to experience outdoor environments, which can further promote psychological resilience and life satisfaction (Celis-Morales et al. 2017). By reducing the dependency on motorized vehicles and encouraging more movement in daily life, active transportation serves as a valuable tool for supporting both physical and mental health.

2.6.2 Cycling and Mental Health Benefits

The relationship between active transportation, particularly cycling, and mental well-being has sparked growing interest among researchers due to its potential benefits for both physical health and psychological resilience. Across various global contexts, studies have highlighted how cycling impacts mental health, uncovering patterns that extend beyond simple physical activity. Scholars have explored how factors like commute frequency, sensory experiences, and access to infrastructure contribute to mental health outcomes, such as reduced stress, improved mood, and enhanced life satisfaction. One research showed that participation in mountain cycling has shown significant mental health benefits, such as reduced stress and enhanced well-being, with mountain bikers often using it as a proactive coping strategy for emotional management (Roberts et al. 2018). Regular cycling for transportation is associated with a marked reduction in psychological distress and improved life satisfaction, supporting its role as a mental health intervention even beyond physical exercise benefits (Ma et al. 2021). Cycling activates multiple senses and facilitates outdoor exposure, both of which can reduce rumination and boost mental well-being by fostering a heightened state of environmental attunement (Wild and Woodward 2019). Moreover, Physical activities in natural environments, including cycling, have been linked to substantial mental health improvements, with benefits in managing anxiety and improving mood and self-esteem across various demographics (Roberts et al. 2018). Researchers also mentioned that cycling as an active transport, positively influences self-perceived health and mental health, with frequent cyclists reporting higher vitality and lower feelings of loneliness compared to other transport modes (Avila-Palencia, De Nazelle, et al. 2017).

2.6.3 Cycling and Non-Motorized Travel in Rural Areas

While the mental and physical health benefits of cycling have been widely recognized in urban settings, there is comparatively limited research addressing the rural context. Rural

areas face distinct transportation challenges, including a lack of dedicated bike lanes, greater distances between destinations, and a stronger dependence on personal vehicles as the primary mode of transportation. Data from the National Household Travel Survey (2022) highlights this disparity: only 5.7% of rural respondents reported cycling at least once in the past 30 days, compared to 11.6% in urban settings (Bricka et al. 2024). These statistics reflect how infrastructure availability and accessibility directly influence cycling behaviors. In rural environments, the absence of safe, accessible, and affordable cycling infrastructure can limit opportunities for residents to benefit from the positive mental health effects of active transportation (Starnes et al. 2011). This gap is particularly important in rural college towns, where students and young adults could significantly gain from improved active travel options (Fan et al. 2017). Addressing these shortcomings through targeted infrastructure improvements such as protected bike lanes, traffic-calmed streets, and bike-sharing programs has been shown to encourage cycling participation and improve both physical and mental well-being (Aldred et al. 2019; Mueller et al. 2015).

These studies encompass urban and rural settings, as well as employ a range of methodologies, from cross-sectional surveys to instrumental variable analyses, to isolate cycling's unique effects on mental health. While some studies focus on the physiological benefits of exercise, others delve into qualitative aspects, capturing the sensory and emotional experiences of cycling in diverse environments. The following table synthesizes these findings, providing a comprehensive overview of each study's main conclusions, settings, models, strengths, and limitations.

Table 2.3: Summary of Key Studies on the Mental Health Benefits of Cycling

Authors	Key Findings	Setting	Model	Strengths	Limitations
(Berrie et al. 2024)	This study reveals that cycle commuting can significantly reduce the need for mental health prescriptions, particularly for commuters traveling shorter distances. The reduction in antidepressant use among cyclists highlights cycling’s potential as an accessible mental health intervention.	Urban (Edinburgh & Glasgow)	Instrumental Variable Analysis	Effectively reduces omitted-variable bias; relies on objective NHS data on mental health prescriptions, strengthening causal claims.	Limited geographic scope may challenge the broader applicability of the findings.
(Tripathy and Sahu 2021)	The study found that urban students generally report good mental health, likely due to factors such as access to improved infrastructure, resources, and a supportive educational environment.	Urban (India)	Cross-sectional Survey	The paper’s in-depth analysis of demographic factors, like gender, within urban environments offers a nuanced understanding of mental health patterns.	Cross-sectional design limits causal inference, and self-reported data may affect reliability.

Table 2.3: Summary of Key Studies on the Mental Health Benefits of Cycling (continued)

Authors	Key Findings	Setting	Model	Strengths	Limitations
(Chekroud et al. 2018)	Exercise, particularly team sports and cycling, reduces mental health burdens. The study suggests that moderate exercise, ideally 3-5 times a week, is optimal for mental health improvement, shedding light on the exercise patterns most beneficial for psychological well-being.	Urban (USA)	Regression with Non-parametric Matching	Large sample size enhances the study's representativeness; variety of exercise types provides detailed insights.	Reliance on self-reported exercise data; causality remains unclear due to cross-sectional design.
(Ohta et al. 2007)	Among male participants, engaging in leisure-time exercise and active commuting by walking or cycling showed a reduction in mental distress, while the effect was not significant in women. This study underscores the mental health benefits of physical activity, particularly for men under job stress.	Urban (Japan)	ANCOVA	Controls for job stress, providing a nuanced understanding of mental health in a work-stressed population.	Gender-specific findings may limit generalizability; focuses only on workers with job stress.

Table 2.3: Summary of Key Studies on the Mental Health Benefits of Cycling (continued)

Authors	Key Findings	Setting	Model	Strengths	Limitations
(Phil Jones 2012)	Jones explores the sensory richness of urban cycling, portraying it as a physical and emotional challenge yet rewarding for those who engage. This study offers a unique perspective on the sensory experiences of cycling, contrasting it with mainstream commuter culture.	Urban (UK)	Ethno-graphic Approach (Audio Diaries)	Provides rich qualitative insights into the sensory dimensions of cycling, fostering a deeper understanding of the commuter experience.	Limited sample size and scope, which may hinder generalizability beyond the UK's cultural context.
(Wild and Woodward 2019)	Cyclists exhibit high levels of commute satisfaction, attributed to sensory engagement, moderate exercise, and control over commute timing. These factors create a fulfilling commute experience, suggesting that the joys of cycling extend beyond physical health benefits.	Urban (New Zealand)	Ethno-graphic Interviews	Highlights psychological and social benefits of cycling, emphasizing its holistic contributions to well-being.	Small sample size; findings are specific to New Zealand, which may limit broader applicability.

Table 2.3: Summary of Key Studies on the Mental Health Benefits of Cycling (continued)

Authors	Key Findings	Setting	Model	Strengths	Limitations
(Avila-Palencia, Panis, et al. 2018)	Bicycling is associated with positive self-perceived health, reduced loneliness, and improved mental health, while car usage is linked with increased loneliness. Walking is linked with greater vitality and social connection, suggesting varied health impacts by mode of transport.	Urban (European Cities)	Mixed-Effects Logistic Regression	Large, diverse international sample strengthens the study's generalizability across urban European contexts.	Findings may not fully translate to non-European settings due to cultural and infrastructural differences.
(Avila-Palencia, De Nazelle, et al. 2017)	Regular bicycle commuters report lower levels of stress compared to non-cyclists, with the stress-reducing benefits increasing with frequency. This study highlights the mental health advantages of routine cycling in an urban Mediterranean environment.	Urban (Barcelona, Spain)	Cross-sectional Survey with Poisson Regression	Considers environmental factors in analysis; sheds light on the value of cycling in stress reduction.	Cross-sectional nature limits causal claims; self-reported data introduces potential biases.

Table 2.3: Summary of Key Studies on the Mental Health Benefits of Cycling (continued)

Authors	Key Findings	Setting	Model	Strengths	Limitations
(Ma et al. 2021)	Regular cycling is linked to lower psychological distress and enhanced life satisfaction. These mental health benefits appear independent of physical environment and individual health conditions, emphasizing the unique impact of transportation cycling on well-being.	Urban (Victoria, Australia)	Propensity Score Matching	Reduces selection bias, offering robust causal insights into the relationship between transportation cycling and mental health.	Specific to Victoria, limiting generalizability; focuses mainly on utilitarian cycling, which may not reflect recreational contexts.

Taken together, the reviewed literature shows that (i) activity type choice is primarily trip-based and does not depend on previous trips; (ii) mode choice in rural college towns is strongly influenced by distance, cost, safety, and most critically, the availability of high-quality bicycle infrastructure; and (iii) cycling is consistently associated with improved mental health, yet these benefits are often limited in rural contexts due to inadequate infrastructure. Despite advances in DDCM, Mixed Logit, and related frameworks, many studies still simplify multi-destination trips, give limited attention to departure time, or rely primarily on urban samples. Few integrate RP–SP data with modern machine learning approaches to jointly model activity participation, mode choice, and well-being. These gaps motivate the empirical strategy in the following chapters of this dissertation.

Chapter 3

Investigating Post Secondary Students Activity Type Choice Pattern in a Multi-destination Trip-based Approach

3.1 Chapter Overview

This chapter explores the nuanced patterns of daily activity type selection among post-secondary students through the lens of multi-destination trip-based modeling and dynamic discrete choice analysis. Building on the limitations of existing activity-based models—which often oversimplify activity types within a multi-destination trip or overlook the specific needs of student populations—this study introduces a comprehensive econometric framework to examine how students in the Greater Toronto and Hamilton Area (GTHA) choose their activity types across multiple daily trips.

The chapter is structured around a comparative model-based approach. For multi-destination trips involving two activities, Multinomial Logit (MNL), Nested Logit (NL), and Mixed Logit models are employed to account for student preferences and the influence of demographic, temporal, and travel-related variables. For three activity-based multi-destination trips, the study applies a Dynamic Discrete Choice Model (DDCM) and a Dynamic Discrete Mixed

Choice Model (DDCMM) to capture sequential decision-making processes and preference heterogeneity over time.

Key findings suggest that students' activity choices are significantly influenced by institutional setting (urban vs. suburban), student status (full-time vs. part-time), employment hours, and household attributes such as vehicle ownership and dependent children. The chapter also reveals distinct behavioral patterns between male and female students and between domestic and international student groups. The modeling approach demonstrates the importance of considering multi-destination trip structure, forward-looking behavior, and unobserved heterogeneity when modeling student travel decisions.

This work represents one of the first empirical attempts to incorporate DDCMM into the analysis of student activity type choice within a multi-destination trip-based framework, offering valuable insights for transportation planning, campus accessibility strategies, and broader policy interventions targeting student mobility.

3.2 Introduction

Activity-type choice is a critical component in activity-based modeling systems. The type of activity one chooses to engage in can influence other components of the activity-based modeling system, such as its duration, the time of the trip, its destination, and the mode of transportation used. For instance, the choices made by someone heading to their workplace differ in terms of transportation mode, activity duration, destination, and timing compared to someone visiting a park for leisure. The selection of activities for trips is a crucial factor in shaping and modeling the dynamics of transportation systems (Khan et al. 2022). When activity-type choices are modeled within an activity-based modeling system, especially within the multi-destination trip, activity-type decisions are simplified for the entire multi-destination trip. Instead of considering the combination of activities within a multi-

destination trip (e.g., home-work-grocery-home), operational activity-based models rely more on simplified activity-type choice (e.g., work activity instead of home-work-grocery-home). When this type of simplification is used, the forecasting results are incorrect, eventually impacting the corresponding policy recommendations and infrastructure development decisions. One study showed that when activity chains are not modeled, the resulting simplification leads to inaccurate forecasting of travel behavior, impacting the effectiveness of transportation policies and infrastructure development decisions (Loa et al. 2021; Moeckel et al. 2024). Also, traditional static discrete choice models fail to capture the interconnectivity between various activity types by ignoring state-dependence and forward-looking behavior. Another limitation of the existing research is the lack of models for the specific population cohorts, such as post-secondary students' population. Previous research has concentrated on modeling activity type choices for the general population and using student status as an explanatory variable. This modeling approach is unable to capture in-depth travel behavior of various population cohorts. This study contributes to the existing literature in several ways:

- It presents a multi-destination trip-based activity type choice model, focusing specifically on patterns of student travel and their impacts on transportation systems.
- Unlike traditional multinomial logit (MNL) models, this research employs a mixed dynamic model to better capture the complexities of student travel behavior over time.
- By selecting a sample that includes students from multiple campuses across both urban and suburban settings in the GTHA, the study captures a wide range of transportation needs and preferences. This varied demographic representation enhances the applicability of the findings, ensuring that the results reflect the diverse realities of student travel experiences, thus supporting more effective and inclusive transportation policy planning.

This study focuses on understanding the daily activity choices of students in the Greater Toronto and Hamilton Area (GTHA) and the impact of these choices on their travel pat-

terns. As of 2021, GTHA had a population of 7,281,694, accounting for 19.7% of Canada's total population. Students constitute a substantial segment of the GTHA's overall population, with over 60,000 students regularly commuting across the region's expanse of 8,296.8 square kilometers. The infrastructure of GTHA often falls short in adequately supporting various modes of travel, such as cars, bikes, or public transport, posing significant challenges for students. The transportation issues faced by students not only affect them but can also provide valuable insights for the broader population. Incorporating the transportation needs and challenges of students into urban transportation policies can be beneficial. Since students' choices in activities and modes of transport are intertwined with other transportation-related factors, understanding these patterns can contribute to enhancing and integrating transportation infrastructure. This study looked at multi-destination trips with two or three trips, but the same method can be used for multi-destination trips with four, five, or more trips. These longer multi-destination trips can be efficiently represented by using sub-tours based on home and work locations. Even without any sub-tours, the number of possible combinations for these longer trip tours is still very low and manageable using the approach followed in this chapter.

3.3 Data for Empirical Investigation

The data for this study was sourced from a web-based student transportation survey. In the Fall of 2019, a survey was conducted in the densely populated Greater Toronto and Hamilton Area in Canada. 18,500 post-secondary students from the area participated in the survey, with the objective of gaining insights into the travel behaviors and preferences of students residing in both urban and suburban settings. A subset of data corresponding to 2-trip and 3-multi-destination trips was used in this research. The survey was divided into four key sections, each focusing on different aspects of the students' experiences and backgrounds. The first section gathered personal information about the students, including

their gender, age, the educational institution they attended, and whether they were international students. It also included their academic status, differentiating between full-time and part-time students, and undergraduate versus graduate students. Additional details such as the year of study, faculty affiliation, the number of hours worked per day, whether they held a driver's license, their average monthly transportation expenses, and how frequently they utilized trip planning applications were also collected. The second category focused on household information. This part of the survey investigated to understand the students' living situations, e.g., whether they lived with family, partners, roommates, alone, or at a friend's house. It also included the composition of their families, including scenarios such as being a single parent, part of a couple, or a couple with children, along with the total number of household members and dependent children. This section also inquired about the family's gross income and the type of dwelling they resided in, whether it was a single-detached house, semi-detached house, apartment, condo, or townhouse, and whether these were owned or rented properties. The third category focused on the students' attitudes towards commuting. This part of the survey was designed to know how commuting affected various aspects of their academic life. Questions were asked to understand whether the commute discouraged them from attending classes, influenced their choice of courses, played a role in their decision on which school to attend, and whether it was a barrier to their academic success. Finally, the fourth category gathered information on the students' travel patterns. This included the number of trips made by the students, the Planning Districts and Census Subdivisions of the origins and destinations of these trips, the modes of transportation used (such as transit, driving, or cycling), the purpose of these trips, and the specific days and times these trips were made.

3.4 Econometric Model

This study adopted a series of advanced discrete choice model to analyze the activity type choices of post-secondary students. The discrete choice model operates on the principle of random utility maximization, where an individual plans to make the optimal decision from a range of available options within a specific timeframe based on the utility derived from each option. The model efficiently summarizes the factors that regularly influence students selections regarding the type of activity they engage in and their mode of transportation. A Multinomial Logit Model (MNL), Nested Logit (NL) Model and Mixed Logit Model are estimated for 2-destination trips, and a dynamic discrete choice model as well as dynamic discrete mixed choice model are adopted for estimating 3-destination trips in this study. There are 13 choices which are discrete and they are Teaching assistant (TA)/ research assistant (RA) work, Usual work location, Shopping and errands, Visiting friends and family, Business Meeting, Restaurants/Bar/Coffee, Daycare, Recreation, Drop Someone Off, Health & Personal Care, Pick Someone off, Services and Other. The utility associated with activity (i) can be written as:

$$V_i = W_i + \delta_i \quad (3.1)$$

The systematic utility:

$$W_i = \alpha Y_i \quad (3.2)$$

Where:

- V_i = total utility of activity i ,
- W_i = systematic utility component,
- δ_i = random utility component with zero mean and scale parameter λ ,

- α = parameter vector,
- Y_i = attribute vector of activity i .

The MNL model operates under the assumption that the random components of the utility associated with the available alternatives are independently and identically distributed and follow the Type I extreme value distribution. Thus, the MNL probability for choosing a specific activity i is defined as,

$$Q_i = \frac{e^{\lambda W_i}}{\sum_{n=1}^N e^{\lambda W_n}} \quad (3.3)$$

However, this implies a scenario of perfect substitution, where all the alternatives are viewed as exact equivalents of each other. However, such an assumption may not hold true in practical scenarios, particularly for specific groups like post-secondary students, who may perceive these alternatives as distinct rather than perfect substitutes. The NL model is an extension of the traditional MNL model, used in the analysis of discrete choices. It is designed to address some of the limitations of the MNL model, particularly the assumption of Independence from Irrelevant Alternatives (IIA). It introduces the concept of nests for alternatives, allowing for varied substitution patterns within each nest compared to alternatives that fall outside these nests. Building upon the NL model allows for even greater flexibility by permitting a single alternative to be a part of multiple nests. This advancement enables a clearer understanding of how different alternatives relate to each other. If the scale parameter at the root level is denoted as R and the scale parameter for the non-work nest is represented by T , then in the framework of the Nested Logit model, the probability of selecting an activity n within the non-work nest T can be expressed in the following way,

$$Q(n \mid \text{non-work nest } T) = \frac{e^{\alpha W_n}}{\sum_{j=1}^4 e^{\alpha W_j^k} \cdot e^{\frac{\alpha R}{\alpha T} \ln \left(\sum_{j=1}^6 e^{\alpha W_j^T} \right)} + \sum_{l=1}^6 e^{\alpha W_l}} \quad (3.4)$$

Probability of choosing an activity l not in a nest:

$$G_l = \frac{e^{\alpha_R W_n}}{e^{\frac{\alpha_R}{\alpha_T} \ln \left(\sum_{j=1}^6 e^{\alpha W_j^T} \right)} + \sum_{l=1}^6 e^{\alpha W_l}} \quad (3.5)$$

The Mixed Logit model is used to analyze situations where individuals make choices from a set of alternatives. Each choice has a utility associated with it, which the model tries to quantify. The utility (X_{in}) of choosing alternative n by individual i is modeled as a function of observable utility (Y_{in}), a random component (Z_{in}) that accounts for mixed effects or heterogeneity in preferences across the population, and an error term (ι_{in}) that captures unobserved influences. Y_{in} includes variables that are observable and affect the utility directly, such as the cost of a choice, the time it takes, or other attributes relevant to the decision-making scenario. Z_{in} represents the variation in preferences among individuals that is not captured by the observable variables. It allows for individual-specific taste variations and can include random parameters that follow a certain distribution. ι_{in} is an error term that captures all other unmodeled influences on the utility.

$$X_{in} = Y_{in} + Z_{in} + \iota_{in} \quad (3.6)$$

The probability of individual i choosing alternative n is:

$$Q_{in} = \frac{1}{D} \sum_{d=1}^D \frac{e^{\mu(\gamma A_{in} + \tau^d B_{in})}}{\sum_j e^{\mu(\gamma A_{ij} + \tau^d B_{jn})}} \quad (3.7)$$

Here, the probability is modelled as an average of exponentiated utilities over D simulations, which accounts for the randomness in the parameter estimates. This is necessary because the integral involved in the computation of probabilities in a Mixed Logit model does not usually have a closed form and must be approximated by simulation. γA_{in} and $\tau^d B_{in}$ terms represent the effects of observed variables (Y_{in}) and random parameters (dB_{in}) on the choice, scaled

by their respective coefficients (γ and d). μ is a scale parameter that adjusts the influence of utility components on the probability of choice.

Discrete choice models suggest that when making decisions, individuals plan to maximize their utility, essentially trying to make the best possible choices from a variety of options available at any moment **reference22**. They want to maximize what they expect to gain in the future, discounted to its present value **reference23**. This concept is used in this study to create a model that explains how postsecondary students in the Greater Toronto and Hamilton Area (GTHA) choose their mode of transportation based on different tour activities.

The description provided focuses on a mathematical model used to analyze the utility or benefit derived from choosing a specific activity type over time within a defined horizon. The utility of choosing a trip activity at any time t , denoted as W , depends on specific variables that describe the state of the individual at that time. M_t represents the attribute vector for an individual at time t . This vector could include various factors such as the individual's location, the time of day, the cost of the trip, and other relevant attributes that influence the decision-making process. ν_t is a random utility component associated with time t , representing the unpredictable elements affecting the utility. This component has a variance of $\sigma_{M,t}^2$ which indicates the degree of uncertainty or variability in the utility from these random factors. The state variables M_t and ν_t , combined as $W(M_t, \nu_t)$, follow a controlled Markov process. A Markov process is a stochastic model used to describe a sequence of possible events in which the probability of each event depends only on the state attained in the previous event. The transition from one trip state M_t to the next state M_{t+1} within a journey is determined by the decision variable a_t . This variable represents the choices or actions taken by the individual. $\lambda(M_{t+1}, \nu_{t+1} \mid M_t, \nu_t, z_t)$ is the Markov transition probability, which calculates the likelihood of moving from one state to the next based on the current state and the decision made.

The formula for anticipated discounted future utility $W(M_t, \nu_t)$ is defined as the maximum expected sum of discounted utilities over an infinite horizon. Here, θ^{k-t} represents the discount factor applied to future utilities, decreasing their contribution to the total utility as k (future time steps) moves further away from t . The function $f(M_k, \nu_k, z_k; v)$ within the summation calculates the utility at each future state M_k , with ν_k being the random component at that future state, and z_k representing other relevant variables at time k . The expectation $E(M_t, \nu_t)$ is conditioned on the current state, emphasizing that the future utility depends on the present state and choices.

The formula for the anticipated discounted future utility value function is presented as follows:

$$W(M_t, \nu_t) = \max E \left[\sum_{k=t}^{\infty} \theta^{(k-t)} f(M_k, \nu_k, z_k; v) \middle| M_t, \nu_t \right] \quad (3.8)$$

Here, θ^{k-t} represents the discount factor applied to future utilities, decreasing their contribution to the total utility as K (future time steps) moves further away from t . The function $f(M_k, \nu_k, z_k; v)$ within the summation calculates the utility at each future state M_k , with ν_k being the random component at that future state, and Z_k representing other relevant variables at time k . The expectation $E(M_t, \nu_t)$ is conditioned on the current state, emphasizing that the future utility depends on the present state and choices.

The value function in equation (7) can be calculated for individual i using the equation from (24) for individual i which is,

$$H(N_t, \delta_t) = \max_{z \in D(N_t)} \{g(N_t, z) + \delta(z)\} + \lambda \sum_{N_{t+1} \in N} \int H(N_{t+1}, \delta_{t+1}) \phi(N_{t+1}, \delta_{t+1} \mid N_t, \delta_t, z) \quad (3.9)$$

The likelihood function can be simplified by leveraging the principle of conditional independence, which takes into account two specific conditions. At any time, t , the unobserved component of the utility, denoted by ξ_t , does not impact the transition from the state U_t to U_{t+1} . The probability density for moving to the next state U_{t+1} does not depend on the un-

observed component ξ_t , instead, it is solely dependent on the observed state U_t . Given these conditions, the transition probability within this controlled Markov process, incorporating both the observed and unobserved components, can be articulated as follows:

The probability of transitioning to the states U_{t+1} and ξ_{t+1} , given the current states U_t and ξ_t along with other variables w , can be described as the product of the transition probability of U_{t+1} dependent on U_t and w , and the conditional probability of ξ_{t+1} given U_{t+1} . This relationship is mathematically expressed as:

$$\psi(U_{t+1}, \xi_{t+1} \mid U_t, \xi_t, w) = \psi(U_{t+1} \mid U_t, w) \cdot \psi(\xi_{t+1} \mid U_{t+1}) \quad (3.10)$$

By assuming a multivariate extreme value distribution, a straightforward equation for the conditional probability can be derived. However, when the goal is to capture the varied effects or heterogeneity in preferences across the population, the equation becomes more complex. In such cases, a closed-form solution is not possible, and simulation techniques must be employed to accurately represent these diverse preferences. In the Dynamic Discrete Choice Model (DDCM), a finite horizon was assumed to define the time frame for decision-making. Furthermore, the backward propagation method was used to compute the expectation of future utilities. This approach ensures that the model accurately considers the anticipated outcomes of future decisions, providing a thorough and forward-looking analysis. Gauss software was used to estimate the models (Aptech Systems, Inc. [2024](#)).

3.5 Data Analysis

An analysis of the dataset revealed that over 65% of students engaged in either two or three multi-destination trips on a given day, while the remaining students engaged in either a single trip or more than three trips. This study primarily focused on those students who

were engaged in two or three multi-destination trips during weekdays; this subgroup was used as a basis for modeling activity type choice within a multi-destination trip. Specifically, the study focused on 2,408 students who undertook two destination trips and an additional 787 students who participated in three destination trips on weekdays. Table 3.1 in the study presents a comparison of the level-of-service variables between two and three-destination trips and the household attributes of the people making these trips, including the mean and standard deviation for driving costs, transit costs, in-vehicle transit time, household size, number of dependent children, and the number of vehicles per household. This comparison offers insights into the travel behaviors of post-secondary students. In three destination trips, students undertook more trips within a single destination trip, resulting in shorter average trip lengths compared to those in two destination trips. As a result, three-destination trips have shorter travel times and lower average trip costs than two-destination trips. Table 3.2 illustrates that for a two-destination trips, the average travel time to reach the first activity is 33.64 minutes, and for the second activity, it is slightly longer (34.13 minutes). The average duration of the first activity is considerably long, which is 613.21 minutes. It suggests that this activity might be the main purpose of the trip, such as a work or school day. In contrast, the duration of the second activity is shorter, which is 356.92 minutes, which could indicate a secondary activity like shopping, entertainment, or others. Table 2 illustrates that for three-destination trips, the average travel times are generally shorter than those of the two-destination trips, with 28.63 minutes to reach the first activity, 19.09 minutes to reach the second, and 23.10 minutes to reach the third. The average durations of activities also differ from those on the two-destination trips. The first activity still has a long duration of 606.74 minutes. Again, it indicates that it might be a primary activity like work or education. However, the average durations for the second and third activities are shorter, such as 268.41 minutes and 180.18 minutes, respectively. These shorter durations could represent activities like errands, meetings, or leisure activities.

Table 3.1: Level of service attributes in two and three destination trips

Two-destination trip	Mean	Standard Deviation
Driving cost (in Dollars)	5.58	6.89
Transit cost (in Dollars)	1.73	1.65
In vehicle transit time (min)	0.48	0.61
Household size	2.52	2.22
No of dependent children	0.37	0.93
No of vehicles in the household	0.82	1.11
Three-destination trip		
Driving cost	3.73	5.86
Transit cost	1.45	1.64
In vehicle transit time	0.33	0.51
Household size	2.18	2.16
No of dependent children	0.33	0.87
No of vehicles in the household	0.76	1.03

3.6 Model Estimation Result

In this research, a comprehensive analysis was conducted including the activity type choice for both two and three destination trips. For the two-destination trip analysis, a total of 56 parameters were estimated for Multinomial Logit Model, Nested Logit Model and Mixed Logit Model. For the three-destination trip analysis, a total of 60 parameters were estimated for Dynamic Discrete Choice Model and 61 parameters were estimated for Dynamic Discrete Mixed Choice Model. The significance of each parameter was determined using the t-statistic and most parameters included in the models were statistically significant. The Rho squared value against a null model was calculated for both the two and three-trip models. Typically, the adjusted Rho squared values range from 0 to 1, with higher values indicating a better model fit. In this study, the adjusted Rho squared for the two-destination trip model was approximately 0.3780, 0.3781, and 0.3784 for the MNL, NL, and Mixed Logit models, respectively. The t-stat of the error parameter 1.524 for the mixed multinomial logit model. The adjusted Rho squared for the three-destination trip model was approximately 0.234 and 0.256 for the Dynamic Discrete Choice Model and Dynamic Discrete Mixed Choice

Table 3.2: Travel time and duration attributes in two and three destination trips

Two-destination trip	
Average travel time of activity 1 (min)	33.64
Average travel time of activity 2 (min)	34.13
Average duration of activity 1 (min)	613.21
Average duration of activity 2 (min)	356.92
Three-destination trip	
Average travel time of activity 1 (min)	28.63
Average travel time of activity 2 (min)	19.09
Average travel time of activity 3 (min)	23.10
Average duration of activity 1 (min)	606.74
Average duration of activity 2 (min)	268.41
Average duration of activity 3 (min)	180.18

Model respectively. The t-stat of the error parameter is 3.322 for the Dynamic Discrete Mixed Choice Model. The increase in adjusted Rho squared values across the MNL, NL, and Mixed Logit models shows that as these models become more sophisticated, they better capture the complexities of how students make transportation choices. Each model adds more detail to the analysis, improving the fit of them by considering how different factors influence students differently.

3.6.1 Two-destination trip Multinomial Logit Model

Table 3 shows the estimates and t-statistics for different activity types and related variables for the MNL model in two-destination trip. Alternative Specific Constants (ASCs) for all activity types are included in the table. The negative coefficients for distance and cost across several activities indicate that increases in these variables decrease the likelihood of choosing those activities. This suggests that students are sensitive to both the travel distance and costs associated with their activities. The negative coefficient for time across various activities implies that longer durations negatively impact the likelihood of engaging in these activities. Part-time students show a negative preference for TA/RA work and daycare activities but

a positive preference for shopping, errands, and recreation. Full-time students have a slight positive preference for TA/RA work but negative preferences for restaurant/bar/coffee and other activities. International students have mixed preferences, with a slight negative preference for TA/RA work but a positive one for restaurant/bar/coffee. Students working more than 20 hours a week show a range of preferences, with a positive inclination towards the usual work location and business meetings but a negative one towards dropping someone off. Female students aged 18 to 45 show a positive preference for TA/RA work and usual work location. Male students over 45 show a strong positive preference for TA/RA work in their first trip. Students attending urban institutes have varying preferences, with a positive inclination towards usual work location and visiting friends but a negative inclination towards dropping/picking someone off. Students in suburban institutes show a positive preference for TA/RA work but negative for activities like daycare and services. Households with 1 to 5 vehicles and more than 2 members show varied preferences for different activities, with negative preferences for certain activities like shopping and errands, but positive for others like dropping someone off. The MNL model offers a straightforward way to look at how students choose their means of transportation based on different activities. This model highlights how distance, cost, and the time involved in an activity greatly influence the decisions of the students. For students, traveling longer distances or incurring higher travel costs makes an activity less appealing. This reflects a general trend where students, often working with limited financial resources and tight schedules, tend to prefer activities that are cost-effective and geographically accessible. The MNL model also notes a negative correlation with the time duration of activities. This suggests that students favor activities that are brief and do not consume too much of their day, likely due to their need to balance school responsibilities, work, and personal life. Part-time students, as depicted in the model, show distinct preferences: they steer clear of time-intensive activities like TA/RA work and daycare. This could be because such activities conflict with their flexible needs, where they must juggle educational commitments with part-time jobs. However, these students show a

positive inclination towards engaging in shopping, errands, and recreational activities. These activities not only fulfill essential needs but also provide a break from their otherwise hectic schedules, offering short bursts of leisure that fit well within their time constraints.

3.6.2 Two-destination Trip Nested Logit Model

Table 3 also shows the estimates and t-statistics for different activity types and related variables for Nested Logit Model in a two-destination trip. The estimated value for the coefficient of Expected Maximum Utility (EMU) is 0.86. Activities like shopping and errands, usual work location, visiting friends and family, business meeting, restaurant/bar/coffee, recreation, and others have positive ASCs whereas activities such as daycare, drop someone off, health and personal care, pick someone off, and services show negative ASCs. Distance, cost, and time have negative coefficients, indicating that increases in these factors decrease the likelihood of choosing certain activities. This suggests sensitivity to travel distance, costs, and time among students. Part-time students show a negative preference for TA/RA work, daycare, and recreation, but a positive preference for shopping, errands, restaurant/bar/coffee, and services whereas full-time students have mixed preferences, with slight positive inclination towards TA/RA work but negative for restaurant/bar/coffee, shopping, errands, recreation, and health and personal care. International students show different preferences, with negative values for TA/RA work and restaurant/bar/coffee but positive for visiting friends and family. Students working more than 20 hours a week exhibit diverse preferences, with a positive inclination towards usual work location, business meetings, shopping and errands, restaurant/bar/coffee, and recreation, but negative for dropping someone off and picking someone off. Female students aged 18-45 show a strong positive preference for TA/RA work and usual work location. Male students over 45 on their first trip have a strong positive preference for TA/RA work. Students attending institutions in urban areas have mixed preferences, e.g., negative preference for TA/RA work but positive preferences for

usual work location, visiting friends and family, and recreation. Suburban institute students show negative preferences for shopping, errands, daycare, and services. Households with 1 to 5 vehicles and more than 2 members exhibit varied preferences for different activities, with negative preferences for some activities like TA/RA work and shopping and errands, but positive for others like dropping someone off. The NL model builds on the MNL model by considering the relationships between different activity type choices. This model is particularly useful for understanding complex decision-making processes where students may see certain activities as complementary or interchangeable based on their specific needs and preferences. For example, international students might prefer dining out over engaging in TA/RA work, possibly because eating out offers them a way to socialize and experience the local culture, an important aspect of their stay in a foreign country. This model suggests that the choices of students are influenced not just by the direct costs and benefits of activities but also by how these activities fit together in their overall schedule and lifestyle. The NL model captures nuances such as why students working more than 20 hours a week prefer commuting directly to work-related activities but avoid personal errands that could disrupt their work-life balance. This indicates a strategic approach where they prioritize efficiency and minimize travel time, reflecting a need to manage limited time effectively across their multiple roles.

3.6.3 Two-destination Trip Mixed Logit Model

The Mixed Logit Model introduces an element of individual variability, making it a sophisticated tool for analyzing transportation choices among students. This model recognizes that each student might respond differently to the same set of circumstances, influenced by personal preferences, individual schedules, and the specific context of their daily lives. For students in urban areas, the model suggests a positive inclination towards activities linked to their work or social life, but a negative inclination towards tasks that involve logistical

challenges like picking or dropping someone off. This could be due to the complexity and congestion typical of urban transport systems, which might deter them from engaging in activities that require frequent car use. For students residing in suburban areas, the model shows a preference for activities that align with their academic or professional commitments but a reluctance towards engagements like daycare, which might necessitate more complicated and less frequent transportation options. This reflects how environmental factors, such as the availability and convenience of public transportation in suburban versus urban settings, can significantly influence student preferences.

3.6.4 Three-destination Trip Dynamic Discrete Choice Model

In the Dynamic Discrete Choice Model, demographic factors significantly influence the transportation choices of students throughout their day, with various parameters highlighting how personal and situational contexts shape these decisions. Full-time students demonstrate distinct preferences, possibly due to the demands of their academic schedules. They show a negative inclination toward recreational activities on their first trip, likely due to prioritizing educational commitments during the early parts of the day. On the contrary, they exhibit a positive inclination towards attending business meetings, suggesting an alignment of their travel patterns with professional or academic obligations rather than leisure. The choices of international students reflect a mix of integration and academic focus. They are more likely to start their day with work-related activities such as TA/RA work and usual work locations, possibly due to the structure of their academic programs or a greater emphasis on utilizing campus resources. However, they show a slight aversion to leisure activities, such as recreation, which might indicate a prioritization of academic and professional commitments over leisure pursuits. The preferences of working students are strongly inclined towards managing their professional and academic responsibilities efficiently. They display a significant preference for business meetings, suggesting that they prioritize professional engagements

highly, likely due to the need to balance these with their academic schedules. On the other hand, their preference is lower for activities like recreation that reflects their limited availability for leisure due to work commitments. Male students aged between 18 and 45 show a mixed set of preferences. They are less likely to choose TA/RA work as a first activity, perhaps indicating a preference for starting the day with less demanding or more flexible activities. However, they show a positive inclination toward usual work locations that suggests a prioritization of work-related activities once their day gets going. Students attending urban institutes show a slight positive tendency for engaging in TA/RA work and a negative inclination towards shopping and errands which is possibly due to the accessibility and convenience of urban settings that favor quick and essential activities over extended shopping trips. Conversely, students from institutions located in suburban areas show a preference for engaging in TA/RA work and a negative preference for recreation. This might reflect the transportation options available in suburban areas, where commuting might be less frequent or more time-consuming, influencing students to prioritize essential activities over leisure. The number of vehicles and the presence of children in the household also influence student choices, particularly in their capacity to undertake certain activities. For example, owning a vehicle and having children negatively impact the likelihood of choosing TA/RA work and shopping and errands as initial activities. This could be due to the logistical challenges and time constraints imposed by family responsibilities, which make starting the day with work or shopping less feasible. These demographic results provide a comprehensive view of how various factors act to shape the daily transportation decisions of students.

3.6.5 Three-destination Trip Dynamic Discrete Choice Mixed Model

The Dynamic Discrete Choice Mixed Model (DDCMM) for the three-destination trip offers a complex view of how students make activity type choices on regular basis. The ASC values for the first trip reflect significant aversion to starting the day with certain activities. Ac-

tivities such as shopping and errands, business meetings, and restaurant/bar/coffee outings have highly negative values that indicates a strong reluctance to engage in these activities early in the day. A highly negative coefficient for distance indicates that students prefer to engage in activities closer to their starting locations. The cost parameter also significantly influences the choice of activities, with higher costs deterring students from selecting certain activities. Time spent on activities is another critical factor, with a negative and statistically significant parameter. Students appear to prefer shorter activities that do not consume much of their time that allows for more flexibility during the rest of their day. Students also show a preference for business meetings and a disinterest in leisure activities like recreation, indicating a prioritization of professional or academic activities. International students are more likely to engage in activities such as TA/RA work and usual work locations. The second trip choices are significantly influenced by the activity corresponding to the first trip. Students who start their day with TA/RA work tend to continue with similar activities, potentially indicating a preference to maintain a professional or academic focus. In contrast, if the first activity involves shopping and errands, there is a greater probability of a transition towards more social interactions such as meeting friends and family and going to restaurants/bars/coffee shops. The preferences regarding distance, time and cost remain consistent in the second trip, with students continuing to favor activities that are logistically simpler and economically more feasible. Similar patterns observed in the first trip persist, with international students and students working more than 20 hours a week continuing to show preferences that align with their initial choices, emphasizing consistency in their daily routines. Ownership of a vehicle and family responsibilities continue to influence choices, with negative coefficients indicating that these factors may restrict the ability to engage freely in preferred activities due to logistical and time constraints.

Table 3.3: Summary Table for MNL, NL, and Mixed Multinomial Logit Model for Two-destination trip (with p-values)

Category	Activity/Group	MNL	t-stat	p-val	NL	t-stat	p-val	Mixed	t-stat	p-val
Model Characteristics										
Number of observations		2408								
Rho squared (null)		0.3870			0.3871			0.3876		
Adjusted Rho squared (null)		0.378032			0.378071			0.37842		
AIC		7683.04			7682.56			7676.20		
BIC		8007.09			8006.61			8000.25		
Coefficient of EMU					0.86					
t-stat of EMU						-0.88	0.38	0.97	1.78	0.075
Category Characteristics										
ASC	TA/RA work	0.00	-		0.00	0.00	1.00	0.00	0.00	1.00
	Shopping and errands	1.28	0.96	0.34	1.17	1.34	0.18	0.61	0.48	0.63
	Usual work location	1.09	1.28	0.20	0.81	0.81	0.42	1.78	2.22	0.026
	Visiting friends and family	0.19	0.14	0.89	0.37	0.37	0.71	-0.07	-0.08	0.93
	Business meeting	0.55	0.41	0.68	0.26	0.32	0.75	0.51	0.62	0.53
	Restaurant/bar/coffee	0.61	0.45	0.65	0.47	0.49	0.62	0.34	0.41	0.68
	Daycare	-1.93	-1.15	0.25	-2.15	-1.66	0.09	-2.06	-1.59	0.11
	Recreation	0.51	0.37	0.71	0.55	0.57	0.57	0.25	0.29	0.77
	Drop someone off	-0.84	-0.57	0.57	-1.05	-1.06	0.29	-0.92	-0.94	0.35

Continued on next page

Category	Activity/Group	MNL	t-stat	p-val	NL	t-stat	p-val	Mixed	t-stat	p-val
	Health and personal care	-1.02	-0.73	0.47	-0.73	-0.63	0.53	-1.31	-1.45	0.15
	Pick someone off	-0.96	-0.56	0.58	-1.00	-1.02	0.31	-1.022	-0.77	0.44
	Services	-1.18	-0.86	0.39	-1.44	-1.66	0.10	-1.18	-1.37	0.17
	Others	0.76	0.56	0.58	0.71	0.80	0.42	0.50	0.61	0.54
Distance	TA/RA work, Shopping and errands, Usual work location, Business meeting, Restaurant/bar/coffee, Daycare, Recreation, Drop someone off, Health and personal care, Pick someone off, Services	-0.03	-3.82	0.01	-0.03	-2.83	0.47	-0.92	-1.38	0.17

Continued on next page

Category	Activity/Group	MNL	t-stat	p-val	NL	t-stat	p-val	Mixed	t-stat	p-val
Cost	Shopping and errands, Usual work location, Business meeting, visit- ing friends and family, Restaurant/bar/coffee, Daycare, Recreation, drop someone off, Pick someone off, Services, others	-0.0300	-2.22	0.03	-0.03	-2.66	0.78	-0.03	-2.44	0.01
Time	Shopping and errands, Usual work location, Business meeting, visit- ing friends and family, Restaurant/bar/coffee, Daycare, Recreation, Drop someone off, Health and personal care, Pick some- one off, Services	-0.83	-2.53	0.01	-0.68	-2.18	0.03	-0.78	-2.34	0.02
Part-time student	TA/RA work, Daycare	-0.6200	-0.68	0.49	-0.29	-0.34	0.73	-0.67	-0.75	0.46

Continued on next page

Category	Activity/Group	MNL	t-stat	p-val	NL	t-stat	p-val	Mixed	t-stat	p-val
Full-time student	Shopping, bar/coffee, recreation, services	1.2700	1.47	0.14	1.05	1.22	0.22	1.48	1.63	0.10
	TA/RA work	0.13	0.91	0.36	0.11	0.81	0.42	0.17	1.19	0.23
	Restaurant/bar/coffee	-1.36	-2.16	0.03	-1.25	-1.99	0.05	-1.43	-2.27	0.023
	Shopping and errands	-0.20	-0.72	0.47	-0.18	-0.72	0.47	-0.19	-0.66	0.51
	Recreation, health/personal care	-0.15	-0.49	0.62	-0.16	-0.53	0.59	-0.16	-0.49	0.62
International student	TA/RA work	-0.19	-1.15	0.25	-0.24	-0.43	0.67	-0.22	-1.23	0.22
Students working >20 hrs	Usual work location	0.05	0.02	0.98	-0.21	-1.22	0.22	-0.02	-0.07	0.95
	Restaurant/bar/coffee	0.31	0.82	0.41	-0.04	-0.16	0.87	-0.26	-0.64	0.52
	Visiting friends/family	-0.27	-0.67	0.50	-0.27	-0.74	0.45	0.30	0.77	0.44
	TA/RA work	-0.12	-0.71	0.48	0.34	1.00	0.32	-0.12	-0.71	0.48
	Shopping and errands	0.44	1.94	0.05	-0.16	-0.98	0.33	0.39	1.71	0.09
	Usual work location	0.95	5.31	0.00001	0.32	1.46	0.14	0.95	5.16	0.00001
	Business meeting	0.57	2.31	0.02	0.89	4.99	0.00001	0.59	2.35	0.02
	Restaurant/bar/coffee	0.33	1.08	0.28	0.56	2.24	0.03	0.31	1.02	0.31
	Recreation	0.42	1.63	0.10	0.28	1.03	0.30	0.42	1.62	0.11
	Drop someone off	-0.68	-1.30	0.19	0.36	1.47	0.14	-0.66	-1.26	0.21

Continued on next page

Category	Activity/Group	MNL	t-stat	p-val	NL	t-stat	p-val	Mixed	t-stat	p-val
Female (age 18–45)	Health/personal care	0.81	1.87	0.06	-0.66	-1.26	0.21	0.77	1.78	0.075
	Pick someone off	-0.85	-0.73	0.47	0.59	1.49	0.14	-0.93	-0.80	0.43
	TA/RA work	0.73	1.13	0.26	-0.83	-0.71	0.48	1.24	2.05	0.04
	Usual work location	0.75	0.89	0.37	1.24	1.98	0.05	0.43	0.39	0.70
Male (age >45)	TA/RA work	1.54	3.09	0.002	0.74	0.88	0.38	2.00	4.02	0.0001
Urban institute	TA/RA work	0.75	0.89	0.37	1.79	3.65	0.0003	-0.25	-1.28	0.19
	Usual work location	0.19	0.82	0.41	-0.29	-1.53	0.13	0.21	0.91	0.36
	Visiting friends/family	0.39	1.12	0.26	0.19	0.90	0.37	0.35	0.99	0.32
	Recreation	0.31	0.91	0.36	0.22	0.70	0.48	0.32	0.94	0.35
	Drop someone off	-0.25	-0.49	0.62	0.26	0.84	0.40	-0.22	-0.44	0.66
	Pick someone off	-1.12	-1.33	0.18	-0.27	-0.54	0.59	-1.16	-1.38	0.17
	Others	-0.32	-1.16	0.25	-1.00	-1.33	0.18	-0.33	-1.17	0.24
	TA/RA work	1.02	1.18	0.24	-0.32	-1.18	0.24	-0.16	-0.56	0.57
Suburban institute	Shopping and errands	-0.06	-0.21	0.83	-0.09	-0.38	0.70	-0.58	-0.52	0.60
	Daycare	-0.50	-0.45	0.65	-0.54	-0.49	0.62	-0.35	-0.45	0.65
	Services	-0.32	-0.41	0.69	-0.36	-0.46	0.65	-0.26	-1.47	0.14
	Shopping and errands	-0.24	-1.39	0.16	-0.23	-1.39	0.16	-0.09	-0.81	0.42
HH>2, more than 1 and less than 4 vehicles	Usual work location	-0.09	-0.84	0.40	-0.82	-3.25	0.001	0.24	1.11	0.27

Continued on next page

Category	Activity/Group	MNL	t-stat	p-val	NL	t-stat	p-val	Mixed	t-stat	p-val
	Business meeting	0.24	1.14	0.25	-0.12	-0.84	0.40	-0.41	-1.53	0.12
	Restaurant/bar/coffee	-0.43	-1.62	0.11	0.21	1.14	0.25	0.81	0.74	0.46
	Daycare	0.65	0.59	0.55	-0.18	-1.62	0.11	-0.27	-1.22	0.22
	Recreation	-0.21	-0.98	0.33	0.55	0.59	0.56	0.98	2.06	0.04
	Drop someone off	0.95	2.00	0.05	-0.28	-0.98	0.33	0.19	0.46	0.64
	Health/personal care	0.23	0.54	0.59	0.85	2.00	0.05	0.63	0.69	0.49

Table 3.4: Summary Table of Dynamic Discrete Choice for Three-destination Trip Model

Category	Activity Type	Estimate	t-stat	p-val
Model Characteristics				
Number of observations		787		
Rho squared (null)		0.25		
Adjusted Rho squared (null)		0.23		
AIC		4414.29		
BIC		4680.39		
Category Characteristics				
First trip: ASC	TA/RA work	0.0000	N/A	1.000
	Shopping and errands	-0.89	-0.76	0.45
	Usual work location	-1.65	-2.28	0.02
	Business meeting	-3.98	-3.48	< 0.001
	Restaurant/bar/coffee	-3.43	-2.53	0.01
	Recreation	-2.71	-2.33	0.02
First trip: Distance	TA/RA work, Shopping and errands, Usual work location, Business meeting, Restaurant/bar/coffee, Daycare, Recreation, Drop someone off, Health and personal care, Pick someone off, Services	-0.98	-0.02	0.99

Category	Activity Type	Estimate	t-stat	p-val
First trip: Cost	Shopping and errands,	-2.56	-0.06	0.95
	Usual work location, Business meeting, visiting friends and family, Restaurant/bar/coffee, Daycare, Recreation, Drop someone off, Pick someone off, Services, others			
First trip: Time	Shopping and errands,	-2.25	-1.55	0.12
	Usual work location, Business meeting, visiting friends and family, Restaurant/bar/coffee, Daycare, Recreation, Drop someone off, Health and personal care, Pick someone off, Services			
First trip: Full time student	Usual work location	-1.39	-0.46	0.65
	Business meeting	1.62	0.89	0.37
	Recreation	-1.59	-1.75	0.08

Category	Activity Type	Estimate	t-stat	p-val
First trip: Num. of vehicles & children	TA/RA work, Shopping and errands, Business meeting, visiting friends and family, Restaurant/bar/coffee, Daycare, Recreation, drop someone off, Health and personal care, Pick someone off, Services	-0.26	-0.05	0.96
	TA/RA work	-2.05	-0.43	0.67
First trip: Number of vehicles	Shopping and errands	-1.90	-0.63	0.53
	Restaurant/bar/coffee	-2.93	-2.54	0.01
First trip: International student	TA/RA work	1.02	0.41	0.68
	Usual work location	1.52	0.67	0.50
	Business meeting, Recreation	-0.22	-0.15	0.88
First trip: Students working > 20 hrs	TA/RA work	-0.16	-0.05	0.97
	Usual work location	1.21	0.39	0.70
	Business meeting	3.02	1.70	0.09
	Recreation	-1.26	-0.75	0.45
Male, age 18–45	TA/RA work	-0.95	-0.66	0.51
	Usual work location	0.94	1.01	0.35
Urban institute	TA/RA work, Restaurant/bar/coffee	0.06	0.03	0.98
	Shopping and errands	-1.04	-0.41	0.68

Category	Activity Type	Estimate	t-stat	p-val
Suburban institute	TA/RA work, Business meeting	0.29	0.14	0.89
	Recreation	-1.01	-0.86	0.39
Second trip ASC: 1st activity = TA/RA work	Recreation	0.00	N/A	1.00
	Usual work location	1.34	0.31	0.75
	Business meeting	-1.90	-0.51	0.61
	Shopping & errands	4.65	0.82	0.38
	Friends & family	-1.42	-0.39	0.70
	Restaurants/bar/coffee	0.39	0.09	0.93
	Health & personal care	-2.56	-1.48	0.14
	TA/RA work	4.98	0.92	0.36
Second trip ASC: 1st activity = Shopping & errands	TA/RA work	0.00	N/A	1.00
	Shopping & errands	2.13	0.76	0.45
	Friends & family	-1.22	-0.59	0.55
	Restaurants/bar/coffee	0.29	0.16	0.88
Second trip ASC: 1st activity = Usual work location	Restaurants/bar/coffee	0.00	N/A	1.00
	Recreation	0.27	0.11	0.92
	Friends & family	0.18	0.07	0.95
	TA/RA work	3.98	1.37	0.17
	Shopping & errands	2.83	0.93	0.35
	Usual work location	-0.09	-0.04	0.97
	Business meeting	-1.14	-0.48	0.63
Second trip ASC: 1st activity = Business meeting	Shopping & errands	0.00	N/A	1.00
	TA/RA work	2.01	0.91	0.36
Second trip ASC: 1st activity = Restaurants/bar/coffee	TA/RA work	0.00	N/A	1.00

Category	Activity Type	Estimate	t-stat	p-val
Second trip ASC: 1st activity =	Shopping & errands	0.00	N/A	1.00
Recreation	TA/RA work	0.88	0.62	0.54
Second trip: Distance	Multiple activities	-1.06	-0.009	0.99
Second trip: Cost	Multiple activities	-0.86	-0.06	0.54
Second trip: Time	Multiple activities	-1.84	-0.86	0.39
Second trip: Num. of vehicles & chil- dren	Restaurants/bar/cof- fee, Shopping & errands, TA/RA work	-0.49	-0.08	0.60
Second trip: International student	Multiple activities	-0.47	-0.12	0.91
Second trip: Students working > 20 hrs	Recreation, visiting family	-0.39	-2.39	0.02
Second trip: Full time student	Usual work location, Friends & family	0.49	1.54	0.13
Second trip: Num. of dependent chil- dren	Friends & family, TA/RA work, Shopping & errands	-0.79	-0.51	0.61
Second trip: Urban institute	Health & personal care, TA/RA work	-0.66	-0.35	0.73

Table 3.5: Summary Table of Dynamic Discrete Mixed Choice Model for Three-destination trip

Category	Activity Type	Estimate	t-stat	p-val
Model Characteristics				
Number of observations		787		
Rho squared (null)		0.2548		
Number of simulations		10000		
Rho squared (null)		0.2557		
Adjusted Rho squared (null)		0.2349		
AIC		4412.846		
BIC		4678.935		
Error Parameter		5.2671	3.322	< 0.001
Category Characteristics				
First trip: ASC	TA/RA work	0.0000	N/A	1.000
	TA/RA work	N/A	0.00	1.000
	Shopping and errands	-6.337	-3.558	< 0.001
	Usual work location	-0.1352	-0.113	0.910
	Business meeting	-3.2421	-3.778	< 0.001
	Restaurant/bar/coffee	-8.0035	-4.463	< 0.001
	Recreation	-7.8406	-4.453	< 0.001
First trip: Distance	Multiple activities	-0.018	-1.129	0.259
First trip: Cost	Multiple activities	-0.0452	-2.066	0.039
First trip: Time	Multiple activities	-1.2057	-1.763	0.078
First trip: Full time student	Usual work location	-0.3465	-1.059	0.289
	Business meeting	0.9662	1.758	0.079
	Recreation	-1.2137	-1.104	0.270
First trip: Num. of vehicles & children	Multiple activities	-0.0421	-0.210	0.833

Category	Activity Type	Estimate	t-stat	p-val
First trip: Number of vehicles	TA/RA work	-0.4314	-2.053	0.040
	Shopping and errands	-0.6945	-2.107	0.035
	Restaurant/bar/coffee	-2.3059	-2.658	0.008
First trip: International student	TA/RA work	0.2941	0.732	0.464
	Usual work location	0.5494	1.236	0.216
	Business meeting,	-0.2757	-0.406	0.685
	Recreation			
Students working > 20 hrs	TA/RA work	-0.0362	-0.125	0.901
	Usual work location	0.3711	1.176	0.240
	Business meeting	1.6939	3.009	0.003
	Recreation	-0.6549	-1.107	0.268
Male, age 18–45	TA/RA work	-0.6550	-0.955	0.340
	Usual work location	1.0119	0.943	0.346
Urban institute	TA/RA work,	-0.1316	-0.285	0.775
	Restaurant/bar/coffee			
	Shopping and errands	-0.3650	-0.936	0.349
Suburban institute	TA/RA work, Business	-0.0432	-0.091	0.928
	meeting			
	Recreation	-0.9515	-1.085	0.278
Second trip ASC: 1st activity = TA/RA work	Recreation	N/A	0.00	1.000
	Usual work location	0.3068	1.319	0.187
	Business meeting	-0.5603	-2.075	0.038
	Shopping & errands	0.8857	4.680	< 0.001
	Friends & family	-0.3700	-1.370	0.171
	Restaurants/bar/coffee	0.0941	0.412	0.680
	Health & personal care	-1.382	-2.391	0.017

Category	Activity Type	Estimate	t-stat	p-val
	TA/RA work	0.9306	5.038	< 0.001
Second trip ASC: 1st activity = Shopping & errands	TA/RA work	N/A	0.00	1.000
	Shopping & errands	0.7590	0.0332	0.973
	Friends & family	-0.5488	0.2755	0.783
	Restaurants/bar/coffee	0.1246	0.8222	0.411
Second trip ASC: 1st activity = Usual work location	Restaurants/bar/coffee	N/A	0.00	1.000
	Recreation	0.3509	0.881	0.378
	Friends & family	0.2643	0.702	0.482
	TA/RA work	1.5591	4.548	< 0.001
	Shopping & errands	1.1019	3.361	< 0.001
	Usual work location	0.1858	0.455	0.649
	Business meeting	-0.3155	-0.746	0.456
Second trip ASC: 1st activity = Business meeting	Shopping & errands	N/A	0.00	1.000
	TA/RA work	0.8741	1.943	0.052
Second trip ASC: 1st activity = Restaurants/bar/coffee	TA/RA work	N/A	0.00	1.000
Second trip ASC: 1st activity = Recreation	Shopping & errands	N/A	0.00	1.000
	TA/RA work	0.5661	0.798	0.425
Second trip: Distance	Multiple activities	-0.0085	-1.052	0.293
Second trip: Cost	Multiple activities	-0.0519	-0.729	0.466
Second trip: Time	Multiple activities	-0.7287	-1.567	0.117
Second trip: Num. of vehicles & children	Restaurants/bar/coffee,	-0.0801	-0.521	0.602
	Shopping & errands			
Second trip: International student	Shopping & errands	-0.0877	-0.362	0.717
Second trip: Students working > 20 hrs	Recreation	-0.4083	-2.456	0.014

Category	Activity Type	Estimate	t-stat	p-val
Second trip: Full time student	Usual work location, Friends & family	0.5317	1.637	0.102
Second trip: Num. of dependent children	Friends & family, TA/RA work, Shopping & errands	-0.5296	-0.831	0.406
Second trip: Urban institute	Health & personal care, TA/RA work	-0.2971	-0.558	0.577

3.7 Conclusion

This paper represents an advancement in understanding the daily activity type choice of students in the GTHA. Information corresponding to two and three-destination trips collected through a web-based survey was used in this study. 56 parameters were estimated for Multinomial Logit, Nested Logit Model and Mixed Multinomial Model. 61 parameters were estimated for Dynamic Discrete Choice Model and Dynamic Discrete Mixed Choice Model. This research highlights several important trends and variations in preferences among students based on various attributes such as their status, age, employment, and institutional location. The study also reveals distinct patterns in the preferences of activities among part-time and full-time students, international students, and those working more than 20 hours a week. This study also distinguishes the differences of activity choices for female students between the ages of 18 to 45 and male students between the ages of 18 to 45. Moreover, the study highlights the impact of institutional location on student activity choices. Students attending urban and suburban institutes exhibit different preferences, with those who attend institutions located in urban areas showing a positive inclination towards usual work location and visiting friends, while those attending institutions in suburban areas lean towards TA/RA work but show less interest in activities like daycare and services. This paper

represents an advancement in understanding the daily activity type choice of students in the GTHA. Unlike previous research, which often overlooks the unique activity patterns of post-secondary students, this study exclusively focuses on this population cohort. A unique subset of two and three-destination trips was analyzed, providing an in-depth examination that had not been previously explored in existing research. A comprehensive approach was used, estimating MNL, NL, and MxL, as well as Dynamic Discrete Choice Model and Dynamic Discrete Mixed Choice Model. A notable and unique contribution of this research is the incorporation of the Mixed Logit Model within the Dynamic Discrete Choice Model, which allows for capturing the heterogeneity in the preferences of students and provides deeper insights into their decision-making processes. The thoroughness of this research establishes it as a substantial contribution to the field of activity-type choice modeling. Several critical trends and variations in preferences among students based on various attributes such as status, age, employment, and institutional location were identified through this analysis. Distinct patterns in the activity preferences of part-time and full-time students, international students, and those working more than 20 hours a week were revealed. Moreover, differences in activity choices between female students aged 18 to 45 and male students of the same age range were distinguished. The impact of institutional location on student activity choices was highlighted, with students attending institutions in urban areas showing a positive inclination towards usual work locations and visiting friends, while those who attend institutions in suburban areas leaned more towards TA/RA work but showed less interest in activities like daycare and services. While valuable insights have been offered by this research, it is not without limitations. A specific subset of students in the GTHA was the focus of this study; however, this subset might not capture the full diversity of student activity patterns. Moreover, only certain variables were considered in this study, and future research should incorporate a broader range of explanatory variables to gain a more comprehensive understanding of activity type choices. Future research should consider a broader demographic and geographic scope to validate and expand upon these results.

Chapter 4

Modeling Departure Time Choices in Multi-destination Travel: A Deep Learning Investigation

4.1 Chapter Overview

This paper presents a comprehensive investigation into modeling sequential departure time decisions within multi-destination trips, with a specific focus on post-secondary students in the Greater Toronto and Hamilton Area (GTHA). The research begins by identifying key gaps in the literature—namely, the tendency to treat departure time as a secondary decision in Activity-Based Models, the oversimplification of continuous departure time into discrete categories, and the limited application of data-driven approaches for capturing complex, interdependent travel behaviors.

The Introduction outlines the study’s objectives: (1) to directly model departure time as a primary behavioral decision within multi-destination trips, (2) to predict departure times on a continuous scale while accounting for sequential dependencies between trips, and (3) to leverage deep learning models to capture nuanced behavioral patterns without restrictive functional assumptions.

The Literature Review categorizes prior studies into three main methodological streams—Discrete Choice Models (DCMs), Optimization/Game-Theoretic Models, and Deep Learning

Approaches, highlighting their respective strengths and limitations. While DCMs and optimization models offer interpretability and theoretical grounding, they often rely on simplifying assumptions that may overlook behavioral complexity. Deep learning methods, though less explored in this domain, provide a promising alternative for modeling high-dimensional and sequential travel data.

The Data Description details the use of the StudentMoveTO 2.0 survey dataset, consisting of 2,408 two-destination trips with rich socio-demographic and trip-specific attributes. Descriptive statistics and departure time distributions reveal distinct temporal patterns—tight morning clustering for Trip 1 and greater temporal flexibility for Trip 2.

The Methodology section explains the implementation of three deep learning architectures: Gated Recurrent Units (GRU), Long Short-Term Memory networks (LSTM), and Transformer encoders. Separate models are trained for Trip 1 and Trip 2 using socio-demographic variables, trip-specific attributes, and, for Trip 2, the predicted or actual Trip 1 departure time. The section also details input preprocessing, feature representation, hyperparameter tuning, and evaluation metrics (RMSE, MAPE).

The Results and Discussion compare the models' predictive performance and behavioral fidelity. The Transformer model consistently outperforms GRU and LSTM, especially for Trip 2, due to its self-attention mechanism that effectively captures dependencies between trips. Probability Density Functions (PDF) and Cumulative Distribution Functions (CDF) illustrate the models' ability to replicate actual departure time distributions, with the Transformer showing the highest alignment. SHAP-based interpretability analysis reveals that for Trip 1, travel distance, travel time, cost, and socio-demographic characteristics (e.g., international student status, housing type) are key predictors. For Trip 2, the strongest predictor is Trip 1 departure time, underscoring the sequential nature of multi-destination trip-based travel behavior.

The Conclusion and Recommendations emphasize the study's contribution to advancing se-

quential departure time modeling by treating it as a primary decision and applying deep learning to capture complex behavioral patterns. Future work should address sample size limitations, incorporate additional behavioral and attitudinal variables, and explore more robust hyperparameter optimization techniques. The findings have practical implications for designing student-responsive transit schedules, enhancing campus accessibility, and managing peak-hour congestion.

4.2 Introduction

Understanding and accurately predicting individuals' departure time choices is a critical component of travel behavior modeling, which has far-reaching implications for urban mobility, transportation planning, and policy development (Jin 2020). Every individual, consciously or unconsciously, chooses their departure time based on an intricate mix of personal, social, economic, and infrastructural constraints (J. Zhang, Jiang, and Zhang 2025). Departure time choice affects not only when individuals start their trips but also influences their travel mode, route, and even activity participation throughout the day (H. Cheng and Yang 2015). With the growing complexity of urban transportation networks and the increasing emphasis on demand-responsive transport strategies, the need for sophisticated modeling of temporal travel behavior has become more pressing than ever specially in the context of multi-destination trips.

Previous studies have explored departure time choice extensively using a variety of modeling approaches, from discrete choice models to more sophisticated dynamic frameworks (Bhat and Steed 2002; H. Cheng and Yang 2015; Habib 2013; Hossain et al. 2021; Jou et al. 2008). However, two major limitations persist in the literature. First, within the domain of Activity-Based Models (ABMs), departure time is typically not modeled as a primary behavioral decision. Instead, it is often treated as a secondary outcome that emerges from the

broader activity scheduling process (Arentze and Timmermans 2004; Bhat and Koppelman 1999; Castiglione et al. 2015). In most ABMs, core choices such as activity type, duration, location, and sequencing are modeled explicitly. Departure time, on the other hand, is subsequently determined through rule based scheduling mechanisms or simulation based allocation routines (Arentze and Timmermans 2004; Moshe E Ben-Akiva and Bowman 1998; Bhat and Koppelman 1999). This indirect treatment limits the model’s ability to capture flexible, individual-level decisions about departure time. This limitation is more noticeable in multi-destination trips, where the timing of one trip can influence when the next trip starts. Second, Discrete Choice Models (DCMs) have advanced to capture interdependencies between trips and include flexible structures like random parameters, latent classes, and non-linear utility forms (K. E. Train 2009; Walker and Ben-Akiva 2002). However, in practice, these models often need simplifying assumptions to keep them interpretable and computationally manageable (Hess and Daly 2024; Hess and K. Train 2017). DCMs usually rely on pre-specified explanatory variables and fixed utility structures, which can make it difficult to capture complex relationships or latent behavioral patterns in high-dimensional datasets (Greene and Hensher 2007; Hensher and Greene 2003). In contrast, deep learning models use a data-driven approach. They can automatically find important features and complex patterns from the data without the requirement to define specific formulas in advance (Chughtai et al. 2022; Gu et al. 2024; Yu et al. 2017; H. Zhang et al. 2018; H. Zhou et al. 2022). While both approaches have unique strengths, deep learning is particularly useful in settings with large sample sizes, diverse input variables (Liu et al. 2017; C. Sun et al. 2017; Yu et al. 2017), and sequential travel behaviors that are difficult to model with traditional DCM methods (Esteva et al. 2017; Goodfellow et al. 2016; Gu et al. 2024; LeCun et al. 2015; H. Zhang et al. 2018; H. Zhou et al. 2022).

Despite these methodological advancements, much of the existing literature continues to focus on departure time decisions in isolation or under simplified assumptions, particularly in the context of single-trip travel (Badiola et al. 2019; Bajwa et al. 2008; Habib et al. 2009;

Steed and Bhat [2000](#); Thorhauge, Cherchi, et al. [2019](#); Yang et al. [2013](#)). This approach does not capture the more complicated ways that departure time is connected to a person’s full day of activities, especially when they make multiple trips. In chained trips, the time someone leaves for their first trip can strongly affect when and how they make their next trips, including the order and length of those trips (Hasnine and Habib [2020](#); Hossain et al. [2021](#)).

Additionally, predicting when students leave for each trip in a day add another layer of challenge because every trip depends on the timing and duration of earlier ones. Class schedules, part-time jobs, and personal errands vary a lot, making patterns hard to pin down. Unpredictable factors like traffic or transit delays add to the challenge(Jou et al. [2008](#)). On top of that, differences in housing, vehicle access, and lifestyles shape when students choose to travel. Capturing all these moving parts requires models that can handle complex, interconnected behaviors. To address these gaps and better capture the complexity of sequential departure time decisions, this study makes the following key contributions:

- First, this study treats departure time as a main behavioral decision, which is missing in many previous studies. Our approach focuses on departure time directly, specially in multi-destination trips where the timing of one trip influences the next.
- Second, traditional models often rely on simplified assumptions that can oversimplify real-world travel behavior. For example, they commonly divide departure times into broad categories like peak, or off-peak periods and ignores the continuous nature of departure decisions. Additionally, they frequently assume that each trip decision is made independently. It fails to consider how the timing of one trip can influence subsequent trips throughout the day. Our research addresses these limitations by applying advanced AI models that predict departure times on a continuous scale and capture the sequential and interdependent nature of multiple trips within a day.

- Third, unlike traditional Discrete Choice Models (DCMs), which require researchers to pre-define how variables influence choices using fixed functional forms (such as linear, quadratic, or log-transform relationships), this study uses deep learning architectures (e.g., GRU, LSTM, or Transformer) to capture sequential dependencies in departure time choices that can automatically learn complex patterns directly from data without these restrictive assumptions.
- Fourth, our research contributes to sustainable transportation by examining the departure time choices of post-secondary students in the Greater Toronto and Hamilton Area (GTHA), a unique travel segment shaped by institutional schedules, part-time jobs, limited vehicle access, and varying transit dependencies, to uncover their distinct temporal preferences and constraints.

Building on these contributions, our research models sequential departure time behavior, focusing on two and three-destination trips common among commuters, students, and individuals with flexible schedules. These multi-destination trips typically follow patterns such as home-to-work, work-to-restaurant, and restaurant-to-home or home-to-campus, campus-to-part-time job, and job-to-home. Such sequences are rich in behavioral complexity. Modeling departure time at each stage helps reveal how flexible people are with their travel times and activity durations. This is especially relevant today, with hybrid work, flexible class schedules, and evolving responsibilities leading to more varied travel patterns. Capturing these dynamics can inform policies to reduce peak-hour congestion and improve transit scheduling.

The objective of this research is thus twofold:

- To develop deep learning models that accurately predict sequential trip departure times using survey data enriched with socio-demographics, travel distance, time, cost, and trip details.

- To gain a deeper understanding of the factors that influence individuals' decisions to depart from home within multi-destination trips. This includes identifying key factors influencing departure time decisions, including how earlier trips affect subsequent ones during the day.

This understanding can help transportation planners and policymakers design more student-responsive transit schedules, improve first-mile/last-mile connectivity, and enhance the overall accessibility of campuses. Also, studying when students choose to start their trips can help us test ideas like flexible class start times or cheaper transit fares during less busy hours. By seeing how students might change their travel times in response, we can better understand how useful and fair these strategies would be.

4.3 Data Description

This study uses data from the StudentMoveTO 2.0 online travel survey that was conducted in Fall 2019 across the Greater Toronto and Hamilton Area (GTHA). The survey included post-secondary students from ten institutions. From this survey, a subset of 2,408 records of students was selected who completed two trips within a single day. The dataset provides exact departure times (in minutes since midnight), along with detailed trip characteristics such as travel cost, distance, and duration for each trip. It also includes socio-demographic details like age, gender, employment status, and enrollment type, offering a comprehensive view of student travel behavior. In this analysis, trip 1 refers to the first trip made by a student on a given day (e.g., from home to campus or work), while trip 2 represents the second trip that follows (e.g., from campus to another activity location, to work, or back home). Defining these explicitly is important, as the timing and purpose of the first trip often set the stage for the flexibility or constraints observed in the second trip.

Figure 1 shows the histograms of trip 1 and trip 2 departure times. As illustrated, most

first trips begin between 7:00 AM and 9:00 AM (i.e., 420 to 540 minutes after midnight), which is consistent with the typical timing of morning classes or commute-to-work schedules. The distribution is sharply peaked and right-skewed, with fewer individuals departing later in the day. This strong morning clustering suggests that the first trip in a student's day is often constrained by fixed schedules. Trip 2 departure time, on the other side, displays the distribution of Trip 2 departure times from the same 2 destination trips. The pattern here is much broader, with most second trips occurring between 1:00 PM and 4:00 PM (approximately 780 to 960 minutes). This dispersion suggests that students' second trips are not as tightly bound to fixed time commitments. Instead, they may reflect a mix of return trips, transitions between campus locations, errands, or part-time work shifts. The shape of the distribution supports the idea that second trip timing is influenced by the duration and flexibility of the first activity.

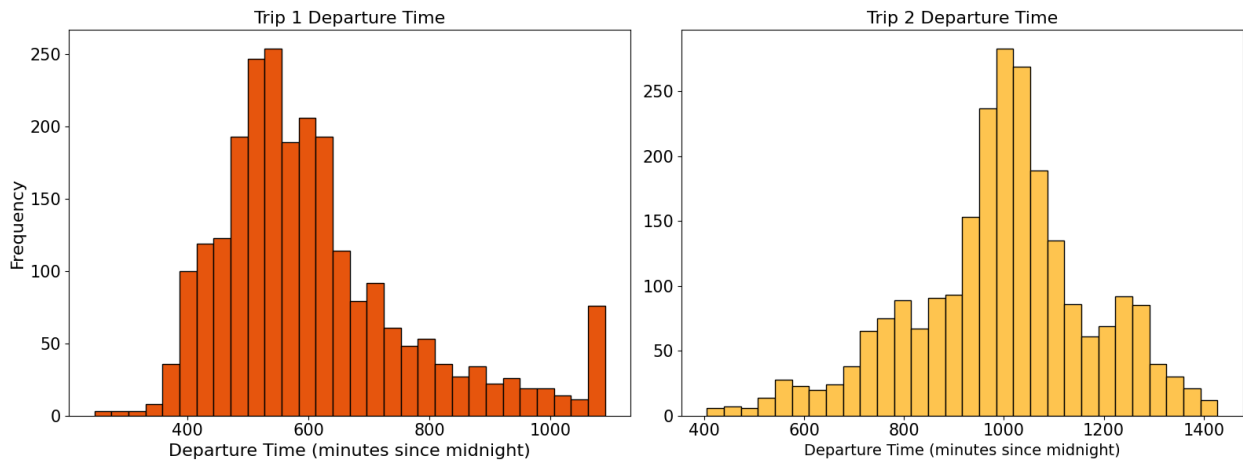


Figure 4.1: Histograms of Trip 1 and Trip 2 departure times. Departure times are shown in minutes since midnight, where 400 $\approx$ 6:40 AM, 600 $\approx$ 10:00 AM, 800 $\approx$ 1:20 PM, 1000 $\approx$ 4:40 PM, 1200 $\approx$ 8:00 PM, and 1400 $\approx$ 11:20 PM.

The datasets include rich socio-demographic profiles. Most individuals are between the ages of 18 and 30. They are either full-time or part-time students, and a significant portion also balance part-time jobs. These personal characteristics play a critical role in shaping travel schedules. For instance, working students may be more likely to depart earlier or later than peers with only academic responsibilities. Similarly, students living off-campus may begin

their first trip earlier due to longer travel distances.

Table 4.1: Summary Statistics for 2-destination trip Dataset

Variable	Mean	Standard Deviation	Minimum	Maximum
Age	23.12	6.52	17	97.00
Monthly Travel Cost	205.93	201.04	0	1500.00
Driving distance (Trip 1) in miles	9.57	11.79	0	85.99
Transit Distance (Trip 1) in miles	9.02	14.21	0	155.97
Park & Ride Distance (Trip 1) in miles	1.71	6.08	0	106.27
Ridehailing distance (Trip 1) in miles	9.57	11.79	0	85.99
Driving time (hr) (Trip 1)	0.25	0.25	0	1.48
Transit time (hr) (Trip 1)	0.76	1.01	0	10.66
Park & Ride time (Trip 1)	0.43	0.59	0	4.33
Ridehailing time (hr) (Trip 1)	0.25	0.25	0	1.48
Driving cost (Trip 1) \$	5.55	6.84	0	49.87
Transit cost (Trip 1) \$	1.75	1.65	0	3.30
Park & Ride cost (Trip 1) \$	4.38	5.23	0	51.54
Ridehailing cost (Trip 1) \$	33.76	39.08	0	275.02
Driving distance (Trip 2) in miles	9.63	11.88	0	90.52

Continued on next page

Table 4.1: Summary Statistics for 2-destination trip Dataset (continued)

Variable	Mean	Standard Deviation	Minimum	Maximum
Transit Distance (Trip 2) in miles	8.50	12.86	0	155.97
Park and Ride (Trip 2) in miles	8.50	12.86	0	155.97
Ridehailing distance (Trip 2) in miles	9.63	11.88	0	90.52
Driving time (hr) (Trip 2)	0.26	0.27	0	1.78
Transit time (hr) (Trip 2)	0.72	0.90	0	10.66
Park & Ride Time (Trip 2)	0.49	0.69	0	9.58
Ridehailing time (hr) (Trip 2)	0.26	0.27	0	1.78
Driving cost (Trip 2) \$	5.58	6.89	0	52.50
Transit cost (Trip 2) \$	1.73	1.65	0	3.30
Park & Ride cost (Trip 2) \$	4.16	4.71	0	51.18
Ridehailing cost (Trip 2) \$	34.16	39.60	0	294.91
No of dependent children	0.37	0.92	0	12.00
No of hr work/wk	1.40	1.67	0	5.00
Household Member	2.52	2.22	0	20.00
Start Time (Trip 1) in minutes	608.39 (10:08 AM)	162.68	246	1090 (6:10 PM)
Travel time (Trip 1) in minutes	33.64	42.59	0	368.96

Continued on next page

Table 4.1: Summary Statistics for 2-destination trip Dataset (continued)

Variable	Mean	Standard Deviation	Minimum	Maximum
Start Time (Trip 2) in minutes	993.70 (4:34 PM)	179.95	405	1428 (11:48 PM)
Travel time (Trip 2) in minutes	34.13	42.24	0	368.97

The descriptive statistics presented in Table 4.1 provide a comprehensive overview of the dataset used for modeling departure time decisions in two-destination trips. The average age of respondents is 23.12 years, with a moderate variation (SD is 6.52), indicating a relatively young and diverse student population. On average, students start their first trip at 608 minutes (approximately 10:08 AM) and second trip at 993 minutes (around 4:33 PM), with standard deviations of 162.68 and 179.95 minutes, respectively. This highlights the wide variability in daily activity scheduling among students.

In terms of travel modes, the mean driving distance for both trips is around 9.6 miles, while transit distances average between 8.5 and 9 miles. Park & Ride distances show high variability, with means around 1.7 miles for Trip 1 and higher for Trip 2, suggesting varied access points to transit facilities. The ridehailing distances closely mirror driving patterns, reflecting modal overlaps in travel behavior.

Travel time and cost metrics follow expected patterns. Driving time for both trips averages 0.25 to 0.26 hours (15 minutes), while transit time is significantly longer at around 0.72–0.76 hours (43–46 minutes). Ridehailing time is consistent with driving, reinforcing the assumption of similar service levels. Transit cost is modest (\$1.73–\$1.75), while ridehailing cost is much higher (mean \$34), highlighting stark cost differentials that may influence mode and timing choices. Park & Ride costs fall between the two, suggesting an intermediate trade-off option.

The travel time in minutes for each trip aligns with the hour-based statistics, showing averages around 33–34 minutes per trip, with large standard deviations, indicative of route and congestion variability. The dataset also includes socio-demographic variables such as number of dependent children (mean 0.37), weekly working hours (mean 1.40), and household size (mean 2.52), which add further contextual depth to departure time choices. These factors reflect differences in responsibility levels, time flexibility, and potential household constraints that may influence when students choose to travel.

4.4 Methodology

To capture the complex temporal dynamics inherent in multi-destination trip-based departure time decisions, we used and compared three state-of-the-art deep learning architectures: Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM) networks, and Transformers. These models were carefully selected for their ability to model sequential dependencies, nonlinear interactions, and high-dimensional feature spaces that are typical in travel behavior data. GRU and LSTM, as recurrent neural networks (RNNs), are well-known for their effectiveness in learning patterns over time, especially when the sequence of past activities affects future decisions (Chung et al. 2014). While GRU offers computational efficiency with fewer parameters, LSTM is capable of retaining long-term contextual information (Bai et al. 2018). On the other hand, the Transformer model introduces a non-recurrent architecture based on self-attention, which allows it to capture complex inter-feature relationships without assuming a fixed sequence order (Lundberg and Lee 2017). This diversity in modeling capabilities enables a robust comparison of how well each architecture captures the temporal and contextual nuances of multi-trip travel behavior among students in the Greater Toronto and Hamilton Area (GTHA).

4.4.1 Input Representation and Preprocessing

The dataset used in this study comprises 2,408 records of student travel behavior collected through a survey. Let $N = 2408$ denote the number of student travel records. Each record corresponds to a two-destination daily trip and is represented by a set of socio-demographic and trip-specific attributes. Separate models are developed for predicting the departure times of Trip 1 and Trip 2. While the socio-demographic variables remain constant for both trips, the trip-specific features differ across the two predictions.

Trip 1 Input Representation: For Trip 1, the input feature vector comprises a total of $d_{\text{Trip1}} = 98$ variables. Each input vector for the i^{th} student is constructed by concatenating the following components:

- $\mathbf{x}_{i,\text{socio}} \in \mathbb{R}^{88}$: One-hot encoded socio-demographic variables such as gender, age group, international status, enrollment type, vehicle ownership, license status, marital status, and housing characteristics.
- $\mathbf{x}_{i,\text{trip1}} \in \mathbb{R}^{10}$: Standardized continuous variables specific to Trip 1, including travel distances, travel times, and costs for various modes (driving, transit, park-and-ride, and ridehailing).

The complete input vector for Trip 1 is:

$$\mathbf{x}_i^{(1)} = [\mathbf{x}_{i,\text{socio}}, \mathbf{x}_{i,\text{trip1}}], \quad \mathbf{x}_i^{(1)} \in \mathbb{R}^{1 \times 98} \quad (4.1)$$

This input is reshaped into a 3D tensor $\mathbf{X}^{(1)} \in \mathbb{R}^{2408 \times 1 \times 98}$ to comply with the expected input format of all deep learning models used in this study. A linear projection layer is applied to map this input to an embedding space of dimension $\mathbb{R}^{1 \times 32}$, enabling the models to capture higher-order interactions across features.

Trip 2 Input Representation: For Trip 2, the input dimensionality increases to $d_{\text{Trip2}} = 109$, reflecting the inclusion of additional features from both Trip 1 and Trip 2. Each input vector for the i^{th} student is composed of:

- $\mathbf{x}_{i,\text{socio}} \in \mathbb{R}^{88}$: The same socio-demographic features used for Trip 1.
- $\mathbf{x}_{i,\text{trip12}} \in \mathbb{R}^{21}$: A combined set of standardized continuous features from both trips, including travel distances, times, and costs, as well as the duration of the home activity between the two trips and the predicted or actual departure time of Trip 1.

The complete input vector for Trip 2 is:

$$\mathbf{x}_i^{(2)} = [\mathbf{x}_{i,\text{socio}}, \mathbf{x}_{i,\text{trip12}}], \quad \mathbf{x}_i^{(2)} \in \mathbb{R}^{1 \times 109} \quad (4.2)$$

This input is also reshaped into a 3D tensor $\mathbf{X}^{(2)} \in \mathbb{R}^{2408 \times 1 \times 109}$ to serve as input to the model. Similar to Trip 1, a linear projection layer transforms the input into an embedding of dimension $\mathbb{R}^{1 \times 32}$.

The socio-demographic features are not standardized due to their categorical nature and are handled through one-hot encoding. In contrast, all continuous trip-specific features are standardized using z-score normalization as follows:

$$x_{ij}^{\text{norm}} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (4.3)$$

where x_{ij} is the raw value of feature j for individual i , and μ_j and σ_j are the mean and standard deviation of feature j calculated from the training set.

The final input feature matrix for each individual and trip is constructed by concatenating the socio-demographic vector and the normalized trip-specific feature vector:

$$X^{(t)} = [X_{\text{socio}} \mid X_{\text{trip}t}^{\text{norm}}] \quad (4.4)$$

for $t = 1, 2$, denoting Trip 1 and Trip 2, respectively.

Prediction Targets: The output variable for each model is the predicted departure time for the corresponding trip, measured in minutes after midnight:

- $y_{\text{Trip}1,i}^{\text{actual}} \in \mathbb{R}$: The observed departure time of Trip 1.
- $y_{\text{Trip}2,i}^{\text{actual}} \in \mathbb{R}$: The observed departure time of Trip 2.

Each model is trained to minimize the mean squared error (MSE) between predicted and actual values, enabling continuous output suitable for modeling time-of-day. The final outputs are expressed in minute-level resolution, maintaining temporal precision required for downstream transit scheduling and congestion mitigation applications.

4.4.2 Transformer Model

This section outlines the methodology used to predict the departure time of the first trip in a two-destination trip using a Transformer-based deep learning model. The modeling pipeline consists of five main components: (1) data preparation, (2) feature engineering, (3) model architecture design, (4) training strategy including hyperparameter tuning, and (5) performance evaluation.

4.4.3 Feature Transformation

We one-hot encoded 11 categorical socio-demographic variables into 88 dummy features using `pandas.get_dummies()` and standardized the continuous trip-specific variables with

`StandardScaler()`. The same preprocessing was applied to Trip 1 and Trip 2. All features were converted to `float32` PyTorch tensors and reshaped into sequences of length 1, yielding final dimensions $\mathbb{R}^{1 \times 98}$ for Trip 1 and $\mathbb{R}^{1 \times 109}$ for Trip 2.

4.4.4 Model Structure

To model the relationship between student characteristics and departure time, we implement the Transformer-based deep learning model separately for Trip 1 and Trip 2 of the multi-destination trip.

- **Input Projection Layer:** A linear layer maps the 98-dimensional input vector for Trip 1 and 109-dimensional input vector for Trip 2 into a $\mathbb{R}^{1 \times 32}$ embedding space.
- **Transformer Encoder:** Each model includes a two-layer Transformer encoder with 4 self-attention heads, feedforward layers of 64 units, and a dropout rate of 0.1 for Trip 1 and 0.2 for Trip 2.
- **Output Layer:** The output token is extracted from the final encoder layer and passed through a dense layer to predict a scalar value representing the departure time in minutes from midnight.

The Transformer allows the model to learn complex interactions among high-dimensional inputs without relying on spatial or temporal ordering. Early stopping is used with a patience of 20 epochs for Trip 1 (max 800 epochs) and 15 epochs for Trip 2 (max 1200 epochs).

4.4.5 Mathematical Formulation

Let $\mathbf{X}^{\text{Trip1}} \in \mathbb{R}^{2408 \times 1 \times 98}$ and $\mathbf{X}^{\text{Trip2}} \in \mathbb{R}^{2408 \times 1 \times 109}$ represent the input tensors for Trip 1 and Trip 2, respectively, where:

- 2408 is the number of valid student records after preprocessing,
- 1 is the artificial sequence length to align with Transformer input format,
- 98 and 109 are the input feature sizes for Trip 1 and Trip 2, including one-hot encoded socio-demographics and z-score normalized trip-specific attributes.

Transformer Encoder: For both Trip 1 and Trip 2 departure time prediction models, the Transformer encoder comprises $L = 2$ stacked layers. Each layer contains a multi-head self-attention mechanism followed by a feedforward network, residual connections, and layer normalization. The input tensor for each sample has the shape $X_{\text{student}} \in \mathbb{R}^{1 \times 98}$, where:

- 1 represents an artificial sequence length used for compatibility with the Transformer model,
- 98 is the total number of input features, consisting of 88 one-hot encoded socio-demographic variables and 10 trip-specific variables for Trip 1, or 21 trip-specific variables for Trip 2 (in that case, total dimension becomes 109).

Each input vector is projected into a 32-dimensional embedding space using a fully connected linear transformation:

$$E_{\text{student}} = X_{\text{student}} \cdot W_{\text{input}} + b_{\text{input}}, \quad E_{\text{student}} \in \mathbb{R}^{1 \times 32}$$

Where X_{student} denotes the input feature vector of a student. For Trip 1, $X_{\text{student}} \in \mathbb{R}^{1 \times 98}$, and for Trip 2, $X_{\text{student}} \in \mathbb{R}^{1 \times 109}$. It includes socio-demographic and travel-related features, W_{input} represents the learnable weight matrix that projects the input features into a lower-dimensional embedding space. For Trip 1, $W_{\text{input}} \in \mathbb{R}^{98 \times 32}$, and for Trip 2, $W_{\text{input}} \in \mathbb{R}^{109 \times 32}$,

b_{input} denotes the learnable bias vector added during the projection, $b_{\text{input}} \in \mathbb{R}^{1 \times 32}$, E_{student} is the final projected embedding vector, $E_{\text{student}} \in \mathbb{R}^{1 \times 32}$, which is used as the input to the self-attention mechanism.

Multi-Head Self-Attention: In each of the 4 attention heads, the queries, keys, and values are computed using distinct learned projection matrices applied to the embedded input:

$$Q_{\text{trip}} = E_{\text{student}} \cdot W_Q^{\text{trip}}, \quad K_{\text{trip}} = E_{\text{student}} \cdot W_K^{\text{trip}}, \quad V_{\text{trip}} = E_{\text{student}} \cdot W_V^{\text{trip}}$$

where:

$$W_Q^{\text{trip}}, \quad W_K^{\text{trip}}, \quad W_V^{\text{trip}} \in \mathbb{R}^{32 \times 8}$$

The scaled dot-product attention for each head is computed as:

$$\text{Attention}(Q_{\text{trip}}, K_{\text{trip}}, V_{\text{trip}}) = \text{softmax} \left(\frac{Q_{\text{trip}} \cdot K_{\text{trip}}^\top}{\sqrt{8}} \right) \cdot V_{\text{trip}}$$

Outputs from all four heads are concatenated to form a 1×32 tensor and passed through a linear transformation to return to the original embedding dimension of 32.

Feedforward Sub-Layer: Following the self-attention block, a feedforward neural network is applied identically to each student representation. For a projected student embedding $z_{\text{student}} \in \mathbb{R}^{1 \times 32}$, the feedforward transformation is:

$$\text{FFN}(z_{\text{student}}) = \text{ReLU}(z_{\text{student}} \cdot W_1^{\text{trip}} + b_1^{\text{trip}}) \cdot W_2^{\text{trip}} + b_2^{\text{trip}}$$

where:

$$W_1^{\text{trip}} \in \mathbb{R}^{32 \times 64}, \quad b_1^{\text{trip}} \in \mathbb{R}^{64}, \quad W_2^{\text{trip}} \in \mathbb{R}^{64 \times 32}, \quad b_2^{\text{trip}} \in \mathbb{R}^{32}$$

This operation is followed by a residual connection and layer normalization:

$$Z_\ell = \text{LayerNorm}(\text{FFN}(\text{Attention output}) + \text{Attention output}), \quad \ell = 1, 2$$

Trip-Specific Notes: Trip 1 uses 88 one-hot socio-demographic attributes and 10 standardized variables for features such as driving time, transit cost, and ridehailing cost. Trip 2 uses the same socio-demographics plus 21 trip-specific variables including both Trip 1 and Trip 2 driving time, transit time, park and ride cost, and ridehailing cost. The shared encoder structure enables the model to effectively learn cross-feature dependencies from heterogeneous tabular inputs.

Output Layer: For each trip, the final encoder output $z_{\text{last}}^{\text{Trip1}}, z_{\text{last}}^{\text{Trip2}} \in \mathbb{R}^{32}$ is transformed into predicted departure times using:

$$y_{\text{Trip1},i}^{\text{predicted}} = z_{\text{last},i}^{\text{Trip1}} W_{\text{out}}^{\text{Trip1}} + b_{\text{out}}^{\text{Trip1}}, \quad y_{\text{Trip2},i}^{\text{predicted}} = z_{\text{last},i}^{\text{Trip2}} W_{\text{out}}^{\text{Trip2}} + b_{\text{out}}^{\text{Trip2}} \quad (4.5)$$

Where:

- $z_{\text{last},i}^{\text{Trip1}}, z_{\text{last},i}^{\text{Trip2}}$ are final encoder outputs for the i^{th} student for Trip 1 and 2,
- $W_{\text{out}}^{\text{Trip1}}, W_{\text{out}}^{\text{Trip2}} \in \mathbb{R}^{32 \times 1}$ are learned weights,

- $b_{\text{out}}^{\text{Trip1}}, b_{\text{out}}^{\text{Trip2}} \in \mathbb{R}$ are bias terms,
- $y_{\text{Trip1},i}^{\text{predicted}}, y_{\text{Trip2},i}^{\text{predicted}} \in \mathbb{R}$ are the predicted departure times in minutes after midnight.

4.4.6 GRU Model Structure and Mathematical Formulation

Each student's input feature vector is represented as $x_i \in \mathbb{R}^{1 \times 98}$, where the 98 input dimensions consist of standardized trip-specific attributes and one-hot encoded socio-demographic variables. The GRU model maps this input to a predicted departure time using a series of gating operations and hidden state updates.

For the i -th student, the GRU computes the following components:

$$z_i = \sigma \left(x_i \cdot W_{\text{input_to_update}}^{98 \times 64} + h_i^{(0)} \cdot W_{\text{hidden_to_update}}^{64 \times 64} + b_{\text{update}}^{1 \times 64} \right) \quad (4.6)$$

$$r_i = \sigma \left(x_i \cdot W_{\text{input_to_reset}}^{98 \times 64} + h_i^{(0)} \cdot W_{\text{hidden_to_reset}}^{64 \times 64} + b_{\text{reset}}^{1 \times 64} \right) \quad (4.7)$$

$$\tilde{h}_i = \tanh \left(x_i \cdot W_{\text{input_to_candidate}}^{98 \times 64} + (r_i \odot h_i^{(0)}) \cdot W_{\text{hidden_to_candidate}}^{64 \times 64} + b_{\text{candidate}}^{1 \times 64} \right) \quad (4.8)$$

$$h_i = (1 - z_i) \odot h_i^{(0)} + z_i \odot \tilde{h}_i \quad (4.9)$$

where:

$x_i \in \mathbb{R}^{1 \times 98}$ is the input feature vector for student i , $h_i^{(0)} \in \mathbb{R}^{1 \times 64}$ is the initial hidden state (initialized to zero), $z_i \in \mathbb{R}^{1 \times 64}$ is the update gate, $r_i \in \mathbb{R}^{1 \times 64}$ is the reset gate, $\tilde{h}_i \in \mathbb{R}^{1 \times 64}$ is the candidate hidden state, $h_i \in \mathbb{R}^{1 \times 64}$ is the updated hidden state. The weight matrices $W_{\text{input_to_}*} \in \mathbb{R}^{98 \times 64}$ and $W_{\text{hidden_to_}*} \in \mathbb{R}^{64 \times 64}$, and biases $b_* \in \mathbb{R}^{1 \times 64}$ are learnable parameters.

The final prediction for Trip 1 is:

$$y_{\text{predicted},i} = h_i \cdot W_{\text{output}}^{64 \times 1} + b_{\text{output}}^{1 \times 1} \quad (4.10)$$

Trip 2 GRU Formulation:

For Trip 2, the input vector is extended to $x_i^{(2)} \in \mathbb{R}^{1 \times 109}$, which includes all socio-demographic variables and 21 trip-level features for both trips. The GRU components are calculated as:

$$z_i^{(2)} = \sigma \left(x_i^{(2)} \cdot W_{\text{input_to_update}}^{109 \times 64} + h_i^{(1)} \cdot W_{\text{hidden_to_update}}^{64 \times 64} + b_{\text{update}}^{1 \times 64} \right) \quad (4.11)$$

$$r_i^{(2)} = \sigma \left(x_i^{(2)} \cdot W_{\text{input_to_reset}}^{109 \times 64} + h_i^{(1)} \cdot W_{\text{hidden_to_reset}}^{64 \times 64} + b_{\text{reset}}^{1 \times 64} \right) \quad (4.12)$$

$$\tilde{h}_i^{(2)} = \tanh \left(x_i^{(2)} \cdot W_{\text{input_to_candidate}}^{109 \times 64} + (r_i^{(2)} \odot h_i^{(1)}) \cdot W_{\text{hidden_to_candidate}}^{64 \times 64} + b_{\text{candidate}}^{1 \times 64} \right) \quad (4.13)$$

$$h_i^{(2)} = (1 - z_i^{(2)}) \odot h_i^{(1)} + z_i^{(2)} \odot \tilde{h}_i^{(2)} \quad (4.14)$$

The predicted departure time for Trip 2 is:

$$y_{\text{predicted},i}^{(2)} = h_i^{(2)} \cdot W_{\text{output}}^{64 \times 1} + b_{\text{output}}^{1 \times 1} \quad (4.15)$$

Note that the GRU structure is identical for both trips, but the input dimensions differ due to the additional features for Trip 2.

4.4.7 Training Procedure and Loss Function

Trip 1 Model: The GRU model for Trip 1 is trained to predict the departure time in minutes after midnight. The training objective is to minimize the Mean Squared Error (MSE) between the predicted values $y_{\text{predicted},i}^{(1)}$ and the actual observed values $y_{\text{actual},i}^{(1)}$:

$$\mathcal{L}_{\text{MSE}}^{(1)} = \frac{1}{2408} \sum_{i=1}^{2408} \left(y_{\text{predicted},i}^{(1)} - y_{\text{actual},i}^{(1)} \right)^2 \quad (4.16)$$

Here, $y_{\text{predicted},i}^{(1)} \in \mathbb{R}$ is the predicted Trip 1 departure time for student i , and $y_{\text{actual},i}^{(1)} \in \mathbb{R}$ is the ground truth from the column 'Trip 1 start time' in the dataset.

The total dataset of 2,408 students is split into training (80%, 1,926 students) and testing (20%, 482 students) sets using a fixed random seed. Training is conducted using the Adam optimizer with a learning rate of 0.001 and the following hyperparameters:

- Batch size: 64

- Hidden units in GRU: 64
- Number of GRU layers: 1
- Number of epochs: 800

Early stopping is applied with a patience of 15 epochs, meaning if the validation loss does not improve for 15 consecutive epochs, training is terminated. The model is trained using Backpropagation Through Time (BPTT) to update all weights, with all computations performed on a CPU. Only the 10 trip-level numerical features are standardized, while the 2,498 socio-demographic variables are one-hot encoded, yielding a total input size of 2,508.

Trip 2 Model: The GRU model for Trip 2 is trained in a similar fashion but uses a larger input feature space to capture information from both Trip 1 and Trip 2. The predicted output $y_{\text{predicted},i}^{(2)}$ is compared against the actual departure time $y_{\text{actual},i}^{(2)}$ using MSE:

$$\mathcal{L}_{\text{MSE}}^{(2)} = \frac{1}{2408} \sum_{i=1}^{2408} \left(y_{\text{predicted},i}^{(2)} - y_{\text{actual},i}^{(2)} \right)^2 \quad (4.17)$$

Here, $y_{\text{actual},i}^{(2)} \in \mathbb{R}$ refers to the true Trip 2 departure time obtained from the Trip 2 start time column. The loss function is same for each model of this study.

The input vector for each student includes the same 88 one-hot encoded socio-demographic features plus 21 trip-level features (10 from Trip 1, 10 from Trip 2, and 1 representing the duration of the home activity between trips), resulting in an input dimension of 109. All trip-level features are standardized prior to training.

Training setup and hyperparameters remain consistent with the Trip 1 model:

- Batch size: 64
- Hidden units in GRU: 64

- Number of GRU layers: 1
- Number of epochs: 800

Early stopping and BPTT are applied as before. Like Trip 1, the model is trained on CPU and uses the Adam optimizer. The same 80:20 train-test split is applied. Although the GRU architecture remains identical, the richer feature representation in Trip 2 requires the model to generalize over more complex input patterns that include interdependencies between both trips.

4.4.8 Long Short-Term Memory (LSTM) Model

We used LSTM model to predict students' departure times based on socio-demographic and trip-specific attributes. Although each data sample is non-temporal, the input is reshaped into a sequence of length one to fit the LSTM structure in PyTorch.

Let $\mathbf{x}_i \in \mathbb{R}^{98}$ be the input feature vector for student i , consisting of 88 one-hot encoded socio-demographic variables and 10 standardized trip attributes (distance, time, and cost across multiple modes). The objective is to predict $y_i \in \mathbb{R}$, the departure time (in minutes from midnight) for Trip 1. The model learns a mapping $f_\theta(\cdot)$, parameterized by weights θ , defined as:

$$\hat{y}_i = f_\theta(\mathbf{x}_i) = \text{LSTM}_\theta(\text{reshape}(\mathbf{x}_i) \in \mathbb{R}^{1 \times 98}) \quad (4.18)$$

where:

- $\mathbf{x}_i \in \mathbb{R}^{98}$ is the raw feature vector for student i ,
- $\text{reshape}(\cdot) \in \mathbb{R}^{1 \times 98}$ converts the input into a single-step sequence,

- $\hat{y}_i \in \mathbb{R}$ is the predicted departure time in minutes.

To conform to the input requirements of PyTorch’s LSTM module, the vector $\mathbf{x}_i \in \mathbb{R}^{98}$ is reshaped into a single-timestep sequence:

$$\mathbf{X}_i = \text{reshape}(\mathbf{x}_i) \in \mathbb{R}^{1 \times 98} \quad (4.19)$$

The entire dataset is then a tensor of shape $N \times 1 \times 98$, where N is the number of students. Even though the sequence has only one time step, the LSTM still learns transformations via its internal memory and hidden-state update equations.

4.4.9 Model Architecture

To predict departure times for both Trip 1 and Trip 2, we implemented two separate single-layer LSTM networks followed by regression layers. Each input sample consists of one-hot encoded socio-demographic variables (88 dimensions) and trip-specific standardized variables (10 for Trip 1; 20 for Trip 2, including both trip legs and home activity duration).

The LSTM input $\mathbf{x}_i$ for each student is reshaped to a sequence of length 1 to conform with PyTorch’s input format. Below are the computations for both trips:

Trip 1 LSTM Computation

Input gate:

$$\mathbf{i}_i^{(1)} = \sigma \left(\mathbf{x}_i^{1 \times 98} \cdot \mathbf{W}_{\text{input_to_inputgate}}^{98 \times 64} + \mathbf{h}_i^{(0)} \cdot \mathbf{W}_{\text{hidden_to_inputgate}}^{64 \times 64} + \mathbf{b}_{\text{inputgate}}^{1 \times 64} \right) \quad (4.20)$$

Forget gate:

$$\mathbf{f}_i^{(1)} = \sigma \left(\mathbf{x}_i^{1 \times 98} \cdot \mathbf{W}_{\text{input_to_forgetgate}}^{98 \times 64} + \mathbf{h}_i^{(0)} \cdot \mathbf{W}_{\text{hidden_to_forgetgate}}^{64 \times 64} + \mathbf{b}_{\text{forgetgate}}^{1 \times 64} \right) \quad (4.21)$$

Output gate:

$$\mathbf{o}_i^{(1)} = \sigma \left(\mathbf{x}_i^{1 \times 98} \cdot \mathbf{W}_{\text{input_to_outputgate}}^{98 \times 64} + \mathbf{h}_i^{(0)} \cdot \mathbf{W}_{\text{hidden_to_outputgate}}^{64 \times 64} + \mathbf{b}_{\text{outputgate}}^{1 \times 64} \right) \quad (4.22)$$

Candidate cell state:

$$\tilde{\mathbf{c}}_i^{(1)} = \tanh \left(\mathbf{x}_i^{1 \times 98} \cdot \mathbf{W}_{\text{input_to_candidate}}^{98 \times 64} + \mathbf{h}_i^{(0)} \cdot \mathbf{W}_{\text{hidden_to_candidate}}^{64 \times 64} + \mathbf{b}_{\text{candidate}}^{1 \times 64} \right) \quad (4.23)$$

Cell state update:

$$\mathbf{c}_i^{(1)} = \mathbf{f}_i^{(1)} \odot \mathbf{c}_i^{(0)} + \mathbf{i}_i^{(1)} \odot \tilde{\mathbf{c}}_i^{(1)} \quad (4.24)$$

Hidden state update:

$$\mathbf{h}_i^{(1)} = \mathbf{o}_i^{(1)} \odot \tanh(\mathbf{c}_i^{(1)}) \quad (4.25)$$

Where:

- $\mathbf{x}_i^{1 \times 98}$: Input vector of 88 socio-demographic one-hot features + 10 Trip 1 variables,
- $\mathbf{h}_i^{(0)}, \mathbf{c}_i^{(0)}$: Initial hidden and cell states, both zeros of shape 1×64 ,
- $\mathbf{W}_{\text{input_to_inputgate}}, \mathbf{W}_{\text{input_to_forgetgate}}, \mathbf{W}_{\text{input_to_outputgate}}, \mathbf{W}_{\text{input_to_candidate}}$: Weights from input to gates/candidate,
- $\mathbf{W}_{\text{hidden_to_inputgate}}, \mathbf{W}_{\text{hidden_to_forgetgate}}, \mathbf{W}_{\text{hidden_to_outputgate}}, \mathbf{W}_{\text{hidden_to_candidate}}$: Recurrent weights from hidden state to respective gates,
- $\mathbf{b}_*$: Bias vectors of shape 1×64 ,
- $\sigma(\cdot)$: Sigmoid activation, $\tanh(\cdot)$: Hyperbolic tangent, $\odot$: Element-wise multiplication.

The final output for Trip 1 is:

$$\hat{y}_i^{(1)} = \mathbf{h}_i^{(1)} \cdot \mathbf{W}_{\text{output_regression}}^{64 \times 1} + \mathbf{b}_{\text{output_regression}}^{1 \times 1} \quad (4.26)$$

Trip 2 LSTM Computation

Input gate:

$$\mathbf{i}_i^{(2)} = \sigma \left(\mathbf{x}_i^{1 \times 108} \cdot \mathbf{W}_{\text{input_to_inputgate}}^{108 \times 64} + \mathbf{h}_i'^{(1)} \cdot \mathbf{W}_{\text{hidden_to_inputgate}}^{64 \times 64} + \mathbf{b}_{\text{inputgate}}'^{1 \times 64} \right) \quad (4.27)$$

Forget gate:

$$\mathbf{f}_i^{(2)} = \sigma \left(\mathbf{x}_i^{1 \times 108} \cdot \mathbf{W}'_{\text{input_to_forgetgate}}^{108 \times 64} + \mathbf{h}_i'^{(1)} \cdot \mathbf{W}'_{\text{hidden_to_forgetgate}}^{64 \times 64} + \mathbf{b}'_{\text{forgetgate}}^{1 \times 64} \right) \quad (4.28)$$

Output gate:

$$\mathbf{o}_i^{(2)} = \sigma \left(\mathbf{x}_i^{1 \times 108} \cdot \mathbf{W}'_{\text{input_to_outputgate}}^{108 \times 64} + \mathbf{h}_i'^{(1)} \cdot \mathbf{W}'_{\text{hidden_to_outputgate}}^{64 \times 64} + \mathbf{b}'_{\text{outputgate}}^{1 \times 64} \right) \quad (4.29)$$

Candidate cell state:

$$\tilde{\mathbf{c}}_i^{(2)} = \tanh \left(\mathbf{x}_i^{1 \times 108} \cdot \mathbf{W}'_{\text{input_to_candidate}}^{108 \times 64} + \mathbf{h}_i'^{(1)} \cdot \mathbf{W}'_{\text{hidden_to_candidate}}^{64 \times 64} + \mathbf{b}'_{\text{candidate}}^{1 \times 64} \right) \quad (4.30)$$

Cell state update:

$$\mathbf{c}_i^{(2)} = \mathbf{f}_i^{(2)} \odot \mathbf{c}_i'^{(1)} + \mathbf{i}_i^{(2)} \odot \tilde{\mathbf{c}}_i^{(2)} \quad (4.31)$$

Hidden state update:

$$\mathbf{h}_i^{(2)} = \mathbf{o}_i^{(2)} \odot \tanh(\mathbf{c}_i^{(2)}) \quad (4.32)$$

Where:

- $\mathbf{x}_i^{1 \times 108}$: Input vector with 88 socio-demographic + 20 trip variables (Trip 1 and Trip 2),
- $\mathbf{W}'_{\text{input_to_outputgate}}, \mathbf{W}'_{\text{hidden_to_outputgate}}, \mathbf{b}'_{\text{outputgate}}$: Corresponding output gate weights/biases,
- All other terms follow the same dimensionality and semantics as Trip 1, but with separate parameters.

The predicted Trip 2 departure time is computed as:

$$\hat{y}_i^{(2)} = \mathbf{h}_i^{(2)} \cdot \mathbf{W}'_{\text{output_regression}} + \mathbf{b}'_{\text{output_regression}} \quad (4.33)$$

4.4.10 Evaluation and Interpretation

After training, the model predictions based on test data, for Trip 1 and Trip 2 were evaluated against the actual departure times using three accuracy metrics: Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). These metrics are defined as follows:

- **Root Mean Squared Error (RMSE):**

$$\text{RMSE}^{(t)} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(y_{\text{predicted},i}^{(t)} - y_{\text{actual},i}^{(t)} \right)^2}$$

RMSE penalizes larger errors more heavily and provides an interpretable metric in the same units as the target variable (minutes).

- **Mean Absolute Percentage Error (MAPE):**

$$\text{MAPE}^{(t)} = \frac{100\%}{N} \sum_{i=1}^N \left| \frac{y_{\text{predicted},i}^{(t)} - y_{\text{actual},i}^{(t)}}{y_{\text{actual},i}^{(t)}} \right|$$

MAPE expresses the error as a percentage of the actual value and is most meaningful when all actual departure times $y_{\text{actual},i}^{(t)}$ are non-zero and not too small.

Here, $t \in \{1, 2\}$ denotes the trip number, $y_{\text{actual},i}^{(t)}$ is the true departure time, $y_{\text{predicted},i}^{(t)}$ is the model's prediction, and N is the number of records (2408 for both Trip 1 and Trip 2).

4.4.11 Hyperparameter Tuning and Model Configuration

To ensure a fair and consistent comparison among the three deep learning models, GRU, LSTM, and Transformer, we carefully tuned their hyperparameters for both Trip 1 and Trip 2 departure time prediction tasks. Table 4.2 shows the ranges we explored and the final values we selected after testing different settings. For each model, we adjusted architecture-related settings like the size of hidden layers, number of layers, and dropout rates, as well as training-related settings such as optimizer type, learning rate, batch size, and early stopping. This tuning process helped make sure each model had the best chance to learn the patterns in the student travel data.

The GRU and LSTM models had similar structures. Both used a hidden layer size of 64 and had only one recurrent layer, making them simple but effective. We did not use dropout in these models because early testing showed it actually made performance worse. The GRU model stopped training at 553 epochs for Trip 1 and 874 epochs for Trip 2. In contrast, the LSTM model finished training earlier, after 316 epochs for Trip 1 and 274 for Trip 2, even though it used a higher learning rate of 0.01. This suggests that LSTM benefited from faster learning, especially with fewer epochs.

Table 4.2: Combined hyperparameter tuning and final configurations for GRU, LSTM, and Transformer models used in Trip 1 and Trip 2 departure time prediction.

Parameter	Range	Trip 1	Trip 2
GRU Model Architecture			
GRU hidden size	[32, 128]	64	64
Number of GRU layers	[1, 3]	1	1
Dropout rate	[0.0, 0.5]	0.0	0.0
GRU Training Settings			
Optimizer	Adam, RMSprop, SGD	Adam	Adam
Learning rate	{0.01, 0.001, 0.0005}	0.001	0.001
Batch size	{32, 64, 128}	64	64
Epochs	[100, 1200]	800 (early stop: 553)	1200 (early stop: 874)
Patience (early stopping)	[10, 30]	15	15
Loss function	MSE, MAE	MSE	MSE
Device	CPU, GPU	CPU	CPU
LSTM Model Architecture			
LSTM hidden size	[32, 128]	64	64
Number of LSTM layers	[1, 3]	1	1
Dropout rate	[0.0, 0.5]	0.0	0.0
LSTM Training Settings			
Optimizer	Adam, RMSprop, SGD	Adam	Adam
Learning rate	{0.01, 0.001, 0.0005}	0.01	0.01
Batch size	{32, 64, 128}	64	64
Epochs	[100, 1000]	500 (early stop: 316)	500 (early stop: 274)
Patience (early stopping)	[10, 30]	15	15
Loss function	MSE, MAE	MSE	MSE
Device	CPU, GPU	CPU	CPU
Transformer Encoder Architecture			
Embedding dimension (d_{model})	[16, 64]	32	64
Number of encoder layers	[1, 4]	2	2
Number of attention heads	[2, 8]	4	4
Feedforward hidden size	[32, 128]	64	64
Dropout rate	[0.0, 0.5]	0.1	0.2
Transformer Training Settings			
Optimizer	Adam, RMSprop, SGD	Adam	Adam
Learning rate	{0.01, 0.001, 0.0005}	0.001	0.001
Batch size	{32, 64, 128}	64	64
Epochs	[100, 1200]	800 (early stop: 613)	1200 (early stop: 1032)
Patience (early stopping)	[10, 30]	20	15
Loss function	MSE, MAE	MSE	MSE
Device	CPU, GPU	CPU	CPU

The Transformer model had a more complex structure, using self-attention layers instead of recurrence. It had two encoder layers and four attention heads, which allowed it to learn more detailed relationships between features. The model used a 32-dimensional embedding space for Trip 1 and a 64-dimensional one for Trip 2, to account for the additional features in the second trip. The feedforward layer had 64 units, and we applied dropout to prevent overfitting, 0.1 for Trip 1 and 0.2 for Trip 2. Although it was the most complex model, the Transformer performed best overall, especially in capturing how the timing of Trip 1 influenced Trip 2. It stopped training at 613 epochs for Trip 1 and 1032 epochs for Trip 2, showing its ability to learn deeper patterns over more training time.

All models used the Adam optimizer because of its ability to adjust learning rates automatically. We kept the batch size at 64 for every model, and applied early stopping based on the validation loss, with patience values between 15 and 20 epochs. All models were trained on CPU.

By using the same tuning strategy and clearly reporting our final model settings in Table 4.2, this study makes the modeling process transparent and fair. The careful design choices we made for each model also show how differences in architecture, like using attention versus recurrence can affect how well the models learn and predict departure time decisions.

4.5 Results and Discussion

4.5.1 Baseline Model and Motivation for Deep Learning Approaches

Before turning to deep learning approaches, we initially developed a baseline model using the Random Forest (RF) regression algorithm to predict departure times for both Trip 1 and Trip 2 in the multi-destination trip. While Random Forest is known for its strong performance in many regression tasks due to its ensemble learning structure and resistance

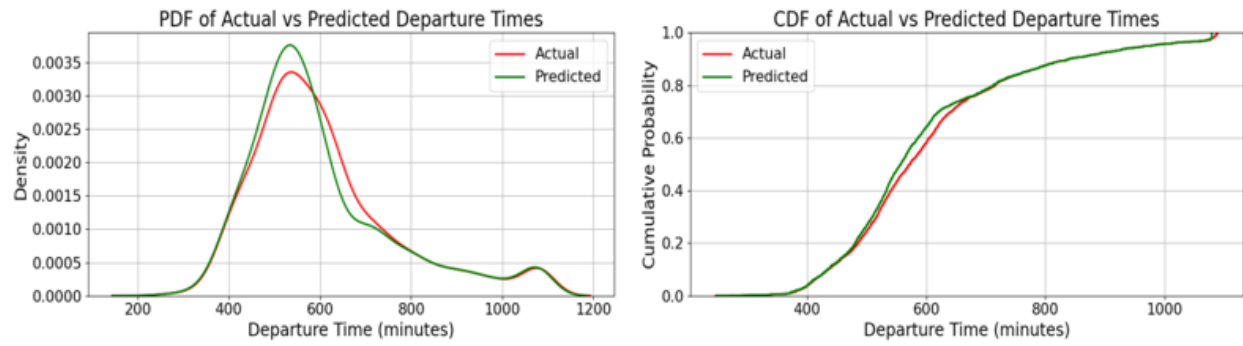
to overfitting (Biau 2012; Breiman 2001; Géron 2022; Hastie et al. 2009), it didn't perform well when it came to predicting when students would leave for their trips. It's predictions were off by more than an hour for exxample, for Trip 1, the RF model produced a root mean squared error (RMSE) of 1.2 hr and a mean absolute percentage error (MAPE) of 10.36%. For Trip 2, the performance further declined, with an RMSE of 1.5 hr and a MAPE of 7.00%. Because of these limitations, we turned to three deep learning models i.e., GRU, LSTM, and Transformer that are better at learning how things change over time (Cho et al. 2014; Goodfellow et al. 2016; Hochreiter and Schmidhuber 1997; Vaswani et al. 2017). In this section, we test all three models to find out which one makes the most accurate predictions. We check how close their predicted times are to the actual departure times, and we also look at whether they reflect realistic student travel patterns.

4.5.2 Distributional fidelity of predicted departure times

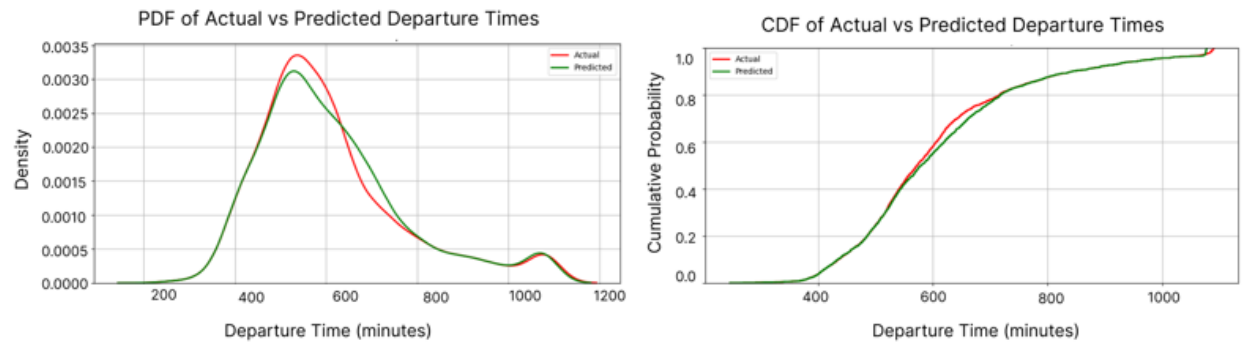
First Trip of the Day

To better understand how well the deep learning models predicted when students leave for their first trip of the day, we compared the actual and predicted departure times using Probability Density Functions (PDF) and Cumulative Distribution Functions (CDF) for each model. Figure 2 shows the CDF and PDF of trip 1 or a multi-destination trip for all three models.

The actual departure time distribution for Trip 1 shows a sharp peak at approximately 500 minutes (around 8:20 a.m.). It indicates that the majority of students begin their day with an early morning departure. This is consistent with previous research highlighting that post-secondary students tend to cluster their first trips around morning class start times, particularly between 7:30 and 9:00 a.m. It may be due to fixed institutional schedules and the need to arrive before class begins (Allen and Farber 2018; Wu et al. 2021). The distribution



(a) PDF and CDF of actual vs. predicted Trip 1 departure time (GRU)



(b) PDF and CDF of actual vs. predicted Trip 1 departure time (LSTM)

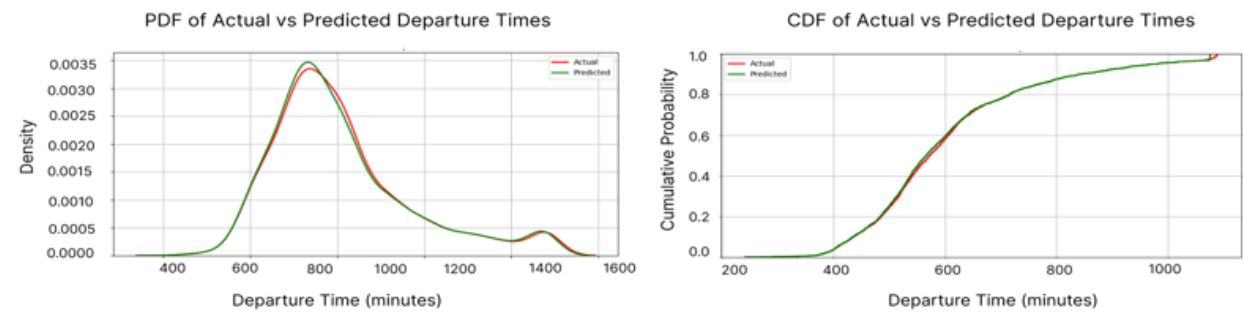


Figure 4.2: PDF and CDF of actual vs. predicted Trip 1 departure time

rapidly declines after 600 minutes (10:00 a.m.), with relatively few students departing later in the day. Both the GRU and Transformer models closely reflect this morning peak in their predicted PDFs. Their predicted distributions also peak near 500 minutes and it aligns well with the actual data. This suggests that both models are capable of learning not only the correct average departure time but also the real-world behavioral concentration of morning departures. The Transformer model, in particular, demonstrates a smooth and consistent density curve that matches the actual distribution across the entire time span. It captures both the early peak and the gradual decline in departures toward the afternoon better than both GRU and LSTM model. The GRU model also correctly identifies the timing of the early peak and captures the steep morning increase in departures. However, its decline after the peak is sharper than observed in the actual distribution. It may indicate a tendency to underpredict departures that occur after 9:00 a.m. As a result, it struggles to accurately reflect the behavior of students with late classes or more flexible schedules. This shows a weakness in how well the model captures the variety in student travel behavior. Cho et al. (2014), who introduced the GRU architecture, demonstrated that GRUs are efficient for sequence modeling and offer a simpler alternative to traditional RNNs (Cho et al. 2014). Later studies have shown that GRUs may underperform compared to LSTM models in capturing long-range and behaviorally complex temporal patterns (Cahuantzi et al. 2023; Chung et al. 2014). Moreover, Transformers have been shown to outperform recurrent models like GRU and LSTM in various sequence modeling tasks involving long-term dependencies, though direct comparisons in behavioral data contexts remain limited (Bai et al. 2018; Vaswani et al. 2017).

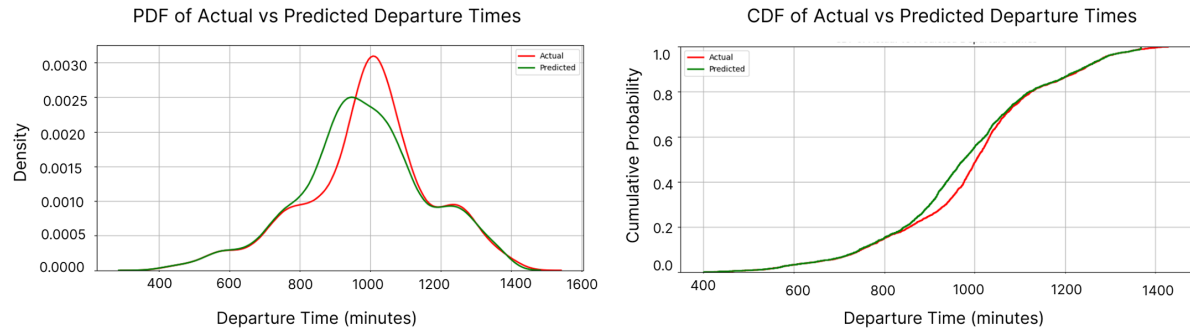
The CDF plots provide additional information about how well each model replicates the actual distribution of Trip 1 departure times across the day. Both the GRU and LSTM models show steep increases in their predicted CDFs between 400 and 600 minutes (approximately 6:40 a.m. to 10:00 a.m.), closely following the shape of the actual CDF during this morning period when most students begin their first trip. This indicates that both models effectively

learn the early-morning departure behavior, which is typical for post-secondary students constrained by class start times (Eluru et al. 2012). However, after 600 minutes, the models diverge. The LSTM model continues to closely follow the actual CDF throughout the mid-morning and early afternoon and maintains alignment with the true cumulative progression up to around 1000 minutes (4:40 p.m.). This suggests that it captures the overall spread of departures reasonably well. The Transformer model follows a somewhat different pattern. Its predicted PDF aligns closely with the actual distribution and captures both the timing and shape of the morning peak around 500 minutes (approximately 8:20 a.m.). The Transformer model demonstrates strong alignment with the actual CDF throughout the entire day, including the early morning period before 600 minutes (10:00 a.m.). Its cumulative curve rises at nearly the same pace as the actual curve. It indicates that the model accurately captures not only the timing of early departures but also their overall distribution.

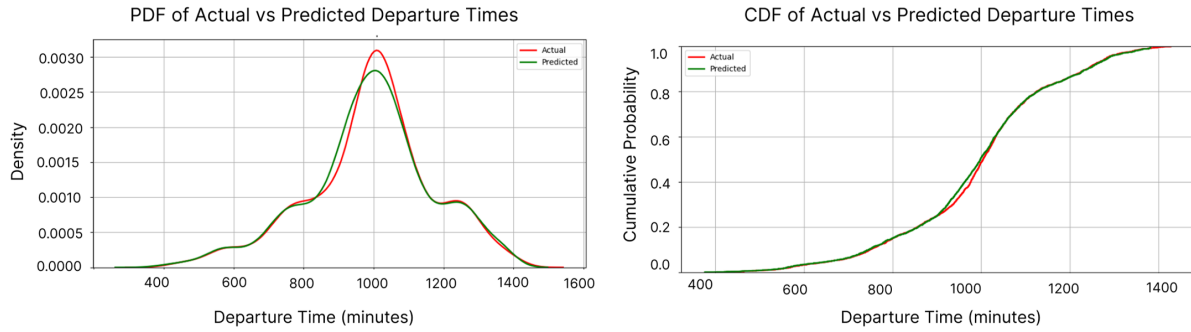
Second Trip of the Day

To better understand how well the deep learning models predicted when students leave for their second trip of the day, we again compared the actual and predicted departure times using PDF and CDF for each model. Figure 3 shows the PDF and CDF plots for Trip 2 of a multi-destination trip for all three models.

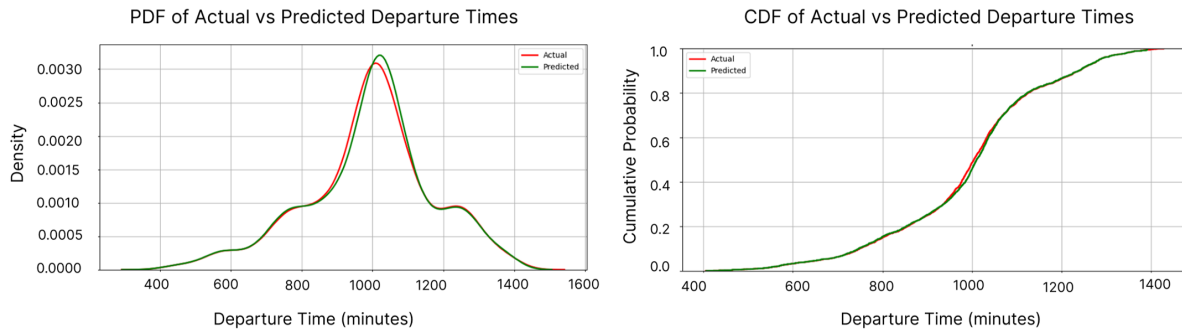
The actual departure time distribution for Trip 2 shows a sharp peak around 1000 minutes (approximately 4:40 p.m.). It shows that a majority of students tend to undertake their second trip in the late afternoon maybe because of likely corresponding to the end of classes, part-time jobs, or afternoon activities. The Transformer model shows the most accurate prediction for Trip 2. Its predicted PDF aligns remarkably well with the actual distribution in terms of both timing and peak intensity. It successfully captures the gradual rise in departures leading up to the peak and the smooth decline afterward. In addition, the Transformer's predicted CDF closely follows the actual cumulative curve throughout



(a) PDF and CDF of actual vs. predicted Trip 2 departure time (GRU)



(b) PDF and CDF of actual vs. predicted Trip 2 departure time (LSTM)



(c) PDF and CDF of actual vs. predicted Trip 2 departure time (Transformer)

Figure 4.3: PDF and CDF of actual vs. predicted Trip 2 departure time

the whole day, including the early afternoon and evening hours. The LSTM model also performs reasonably well. Its predicted PDF closely matches the peak timing around 1000 minutes and moderately approximates the shape of the actual distribution. However, the LSTM prediction slightly overconcentrates around the peak and underestimates both early and late departures. This over sharpness is evident in its CDF as well, which shows mild divergence from the actual curve in the middle. In contrast, the GRU model shows weaker performance relative to the other two. While it correctly identifies the afternoon peak’s approximate timing, it underestimates the peak’s height and overestimates departure rates in the early afternoon (between 800 and 900 minutes). Its PDF appears less concentrated than the actual distribution.

4.5.3 Comparative Analysis of Deep Learning Models

This section presents a comparative evaluation of the GRU, LSTM, and Transformer models to assess their effectiveness in predicting departure times for Trip 1 and Trip 2. We compare their performance based on quantitative metrics and behavioral fidelity.

Prediction Accuracy: This section compares the prediction performance of the three deep learning architectures such as GRU, LSTM, and Transformer for forecasting departure times in a multi-trip travel. The evaluation focuses on two key metrics: Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE), measured separately for Trip 1 and Trip 2. These metrics offer a clear understanding of each model’s accuracy in predicting actual departure times and their proportional error in relation to true values. Table 4 shows the evaluation metrics of each model.

Among the three, the Transformer model showed the strongest overall performance. It achieved the lowest RMSE for both trips, 0.296 hours for Trip 1 and 0.304 hours for Trip 2, alongside the smallest MAPE values, approximately 1.27% and 1.29% respectively. These

results reflect the model’s ability to make precise predictions with minimal deviation from actual values, both in absolute terms and relative to the true departure times. One major factor contributing to this strong performance is the Transformer’s self-attention mechanism, which helps the model to evaluate the importance of each input feature regardless of its order (Vaswani et al. 2017). This is particularly helpful when modeling sequential trips, as the timing of Trip 2 may be heavily influenced by the outcome of Trip 1 and various contextual factors across the entire input space.

The LSTM model performed well, particularly for Trip 1, with an RMSE of 0.326 hours and a MAPE of 1.37%. However, for Trip 2, its performance declined slightly, indicating challenges in modeling dependencies that are more contextually or behaviorally driven, such as activities between trips.

The GRU model showed stable and reasonable accuracy across both trips, with RMSE values of 0.411 hours and 0.430 hours for Trip 1 and Trip 2 respectively, and slightly higher MAPE values compared to the other two models. While not as accurate as the Transformer or LSTM in terms of RMSE, GRU performed well in capturing routine-based trip timing and showed moderate consistency, particularly for the second trip.

Table 4.3: Performance comparison of GRU, LSTM, and Transformer models on Trip 1 and Trip 2 using RMSE and MAPE metrics.

Metric	GRU		LSTM		Transformer	
	Trip 1	Trip 2	Trip 1	Trip 2	Trip 1	Trip 2
Root Mean Squared Error (RMSE)	0.411 hr	0.430 hr	0.326 hr	0.423 hr	0.296 hr	0.304 hr
Mean Absolute Percentage Error (MAPE)	1.67%	1.92%	1.37%	1.86%	1.27%	1.29%

Behavioral Realism: In terms of replicating the real-world departure time distributions, all three models showed good alignment between predicted and actual values as reflected in CDF and PDF plots. However, the Transformer model best captured the distributional

fidelity for both trips. The GRU model slightly underestimated departures in the mid-day peak for Trip 2, while the LSTM model showed smoother but less precise peak matching.

4.5.4 SHAP-Based Interpretability Analysis

Since the Transformer model yielded the best predictive performance among the three models, we focused our interpretability analysis on this model. To understand the internal decision logic of the Transformer model and how individual input features influence predicted departure times, we used SHAP (SHapley Additive exPlanations) values as SHAP values quantify the marginal contribution of each input feature to the model’s output, enabling both global and local interpretability (Lundberg and Lee 2017).

A SHAP summary plot visualizes the impact of each feature on the model’s predictions across all samples in the dataset. In the plot, each point represents a SHAP value for a particular feature and data instance. The color indicates the actual feature value (e.g., high or low), while the position on the x-axis shows the extent and direction of that feature’s impact on the prediction. Features are ranked by overall importance, with the most influential ones appearing at the top.

Trip 1 Interpretation: The left panel of Figure 4.4 presents the SHAP summary plot for the Transformer model trained to predict Trip 1 departure times. The most influential predictors include a combination of travel-related features and socio-demographic characteristics that reflect the underlying heterogeneity in students’ daily mobility constraints and preferences.

- *Transit Distance* and *Driving Time (hr)* are the most impactful features. This is behaviorally intuitive: students who face longer travel distances or longer driving durations are more likely to leave earlier in the day to ensure they arrive on time for

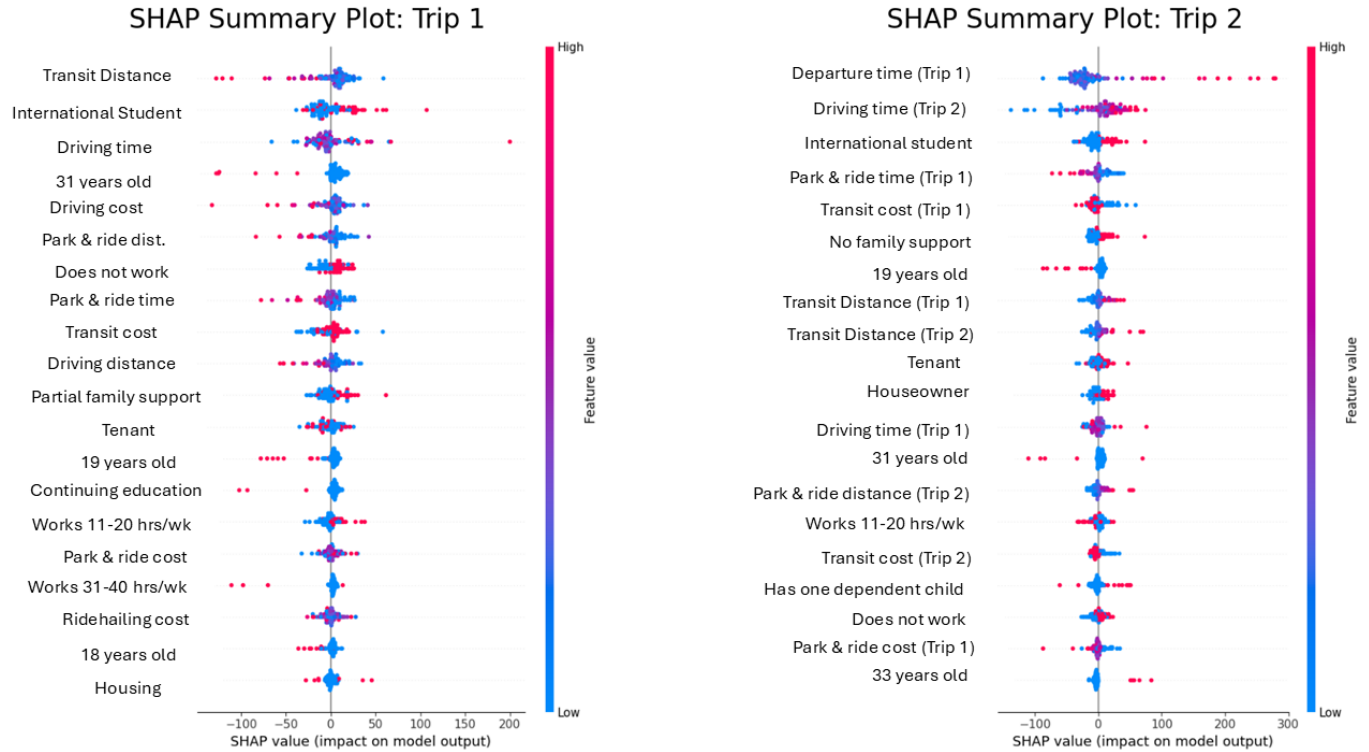


Figure 4.4: SHAP summary plots for Trip 1 and Trip 2 Transformer models

fixed activities such as classes, work shifts, or academic appointments. In particular, a long driving time may signal traffic congestion or route complexity, prompting earlier departure. Likewise, high transit distance may indicate reliance on multiple transfer points or infrequent service, which naturally leads students to depart earlier as a buffer against uncertainty.

- *International Student* status strongly influences departure time predictions. International students often live farther from campus or rely more heavily on public transportation due to lower car ownership. These conditions often push them to depart earlier or structure their day around limited mobility options. This variable's importance highlights how access constraints may shape trip timing behavior in a student population.
- *Tenant* status (as opposed to being a homeowner or living in a dormitory) is another

key contributor. Students who rent housing off-campus often live in more dispersed, lower-density areas not directly served by transit. This typically introduces longer first-mile connections, additional transfer time, and potential delays, factors that encourage earlier departure to maintain reliability. The model correctly learns this implicit association, assigning high importance to this housing-related variable.

- *Driving Cost* and *Park & Ride Time* also exhibit strong contributions to departure time. High driving costs (fuel, tolls, parking fees) can act as a deterrent for early or discretionary trips, resulting in compressed travel schedules or modal shifts. Conversely, longer Park & Ride durations indicate students are integrating transit into their commute, possibly facing delays related to coordination, transfer timing, or parking availability. In both cases, these features influence the degree of flexibility students have in selecting their departure window.

Trip 2 Interpretation: The right panel of Figure 4.4 shows the SHAP summary plot for the Trip 2 Transformer model. While it is expected that the *Trip 1 departure time* influences Trip 2 departure time, the analysis allows us to quantify this effect. The SHAP values indicate that Trip 1 departure time is not just relevant but by far the most influential predictor compared to all other features. This quantification provides empirical evidence of the strength of temporal carryover across trips, which is often assumed but rarely measured in multi-destination travel studies. The model’s ability to capture this temporal interdependence between trips not only enhances its behavioral realism but also supports the core premise of our multi-destination trip-based modeling framework. Other influential features include:

- *Driving Time (Trip 2)* and *Transit Cost (Trip 1)*: Longer travel durations and higher earlier-trip costs continue to impact departure decisions, indicating consistency in time and cost sensitivity across trips.

- *Housing status* (e.g., *Tenant*, *Houseowner*) and *Family Support*: These socio-demographic factors influence the flexibility and timing of travel. For instance, students who rent housing or rely on family support may face external constraints, such as household responsibilities or shared resources that affect when they are able to depart for their second trip of the day.
- *Transit Distance (Trip 2)* and *Park & Ride Cost (Trip 1)*: These features show that conditions faced during Trip 1 (such as costs and transfer burdens) shape the expectations, choices, and departure behavior for Trip 2, emphasizing the sequential dependencies embedded in daily travel.

Taken together, the SHAP analysis validates the behavioral realism of the model and provides interpretable insight into how travel and socio-demographic variables jointly shape departure time decisions in both trips of the multi-destination trip.

4.6 Conclusion and Recommendations

This study addresses a notable gap in the literature by directly modeling departure time as a primary behavioral decision within multi-destination trip. It used three advanced deep learning models (GRU, LSTM, and Transformer) that can capture complex behavioral patterns without relying on restrictive modeling assumptions. Additionally, the study combines predictive accuracy with behavioral interpretability by analyzing both model performance and key influencing factors through SHAP values and distributional comparisons. We used a dataset of 2408 post-secondary students from the Greater Toronto and Hamilton Area (GTHA), with key input features including socio-demographics and travel information.

This study began with a Random Forest (RF) model as a baseline for predicting sequential departure times, but its performance was poor. To improve accuracy, we applied deep

learning models; GRU, LSTM, and Transformer. Among them, GRU performed the worst (RMSE: 0.411 hr, 0.430 hr; MAPE: 1.67%, 1.92%), showing limited ability to capture complex travel patterns. LSTM offered better predictions for Trip 1 (RMSE: 0.326 hr, MAPE: 1.37%) but struggled with Trip 2. The Transformer model outperformed both, achieving the highest accuracy (RMSE: 0.296 hr, 0.304 hr; MAPE: 1.27%, 1.29%). Additionally, the Transformer model closely matched the actual data for both trips, accurately replicating morning and afternoon peaks as well as the rise and fall patterns. In contrast, the GRU tended to underpredict mid-day departures, while the LSTM produced smoother outputs that failed to reflect sharp behavioral peaks.

Since the Transformer model achieved the best predictive performance, we focused our explainability analysis on this model using SHAP, which revealed how different input features influenced its predictions. For Trip 1, the most influential features were transit distance, driving time, and cost-related variables. These factors reflect practical travel constraints, as students with longer or more expensive commutes were more likely to depart earlier to ensure on-time arrival. Socio-demographic factors like international student status and tenant housing also significantly influenced departure times. For Trip 2, Trip 1 departure time emerged as the top predictor that confirms the sequential nature of daily student travel. This is particularly important for this study as it validates the multi-destination trip-based modeling approach, where departure decisions are interlinked and influenced by earlier trips within the same day.

While this study offers valuable insights into students' departure time choices using advanced deep learning models, several steps can improve the robustness and generalizability of future work. The dataset, though rich in features, was relatively small with 2,408 students, limiting the diversity of travel behavior patterns the models could learn, particularly across socio-demographic and regional variations. We recommend expanding data collection to include a larger, more diverse sample of students from multiple colleges and universities across different cities and states. Such broader coverage would better capture variations

in urban form, transit availability, and lifestyle constraints, leading to more generalizable models and deeper behavioral insights. Second, the observed influence of socio-demographic factors such as housing status, family support, and international student identity highlights the importance of understanding contextual and cultural influences on travel behavior. Future surveys should aim to incorporate additional behavioral and attitudinal variables (e.g., flexibility in daily schedule, preferences for sustainable travel modes, perceived safety) that may further explain variability in departure time choices. Third, for hyperparameter tuning, the study performed only grid search; therefore, the parameters that are reported may not be the best optimal parameters. Future research should use a more robust hyperparameter tuning strategy (e.g., Optuna).

This study can help shape better travel behavior policies by showing when students are likely to leave for their trips. Knowing departure times for trips in a day allows transit agencies and university planners to spot busy travel periods and adjust bus or shuttle schedules to meet demand. It also shows how things like class times or earlier trips affect when students travel later in the day. These findings can guide policies that spread out travel to avoid crowding, improve shuttle timing, and connect different travel options more smoothly.

Chapter 5

Investigating the Travel Behavior of Post-Secondary Students' in Rural USA College Towns

5.1 Chapter Overview

This chapter investigates the travel mode choices of post-secondary students in rural college towns, with a focus on cycling as a sustainable alternative to private vehicle use. Traditional travel behavior studies often prioritize urban contexts or overlook the infrastructural and perceptual barriers faced by rural students. In contrast, this study introduces a joint econometric modeling framework that combines Revealed Preference (RP) and Stated Preference (SP) data to explore how students respond to variations in bike infrastructure, travel costs, distances, and facility availability.

The chapter is structured around a comparative application of Multinomial Logit (MNL) and Mixed Logit (MXL) models to estimate both observed behavior and hypothetical shifts in mode choice. The modeling framework captures preference heterogeneity across individuals and accounts for infrastructure improvements such as protected bike lanes, designated bike lanes, and supportive facilities like maintenance stations and showers.

Key findings reveal that students are significantly more likely to choose cycling when protected bike lanes are available, while designated lanes without physical separation fail to

generate a comparable effect. Factors such as cost sensitivity, trip length, car ownership, and gender identity emerge as influential predictors of mode choice, with female students showing a greater propensity to avoid cycling due to perceived safety concerns. The analysis also emphasizes the role of institutional characteristics, student status (full-time vs. part-time), and access to travel alternatives in shaping modal decisions.

This work contributes one of the first empirically grounded studies to model rural student mobility using a combined RP-SP framework. The findings offer actionable insights for rural transportation planning and policy design, particularly in identifying infrastructure investments and supportive amenities that encourage bicycle use among post-secondary students in non-urban settings.

5.2 Introduction

Transportation is a fundamental part of student life, influencing access to education, employment, and social activities. However, students in rural college towns face unique mobility challenges due to limited public transportation, long travel distances, and a heavy reliance on private vehicles. Unlike urban areas where public transit, biking, and walking are more accessible, students in rural settings often have no choice but to drive, making sustainable travel options underutilized. This study seeks to understand the mode choice behavior of post-secondary students in rural college towns and, more importantly, examine how the introduction of new and safe bicycle infrastructure can influence their travel decisions.

As travel by personal vehicle continues to dominate the roadway network in the USA, planners are seeking to implement safer transportation options, such as cycling, that can ease the congestion stemming from this vehicular demand. At its current stage, much of this focused implementation is targeted towards development in urban regions while the necessity for alternative modes of transportation in rural areas goes unnoticed and untouched. This

results in rural locations with high concentrations of bicycle users, such as university towns, not having sufficient, safe bicycle facilities, leading to an increase in vehicular mode choice. The addition of single-occupant vehicles to the roadway network due to a lack of alternative modes of transportation is often the culprit of increased congestion and vehicular-related injuries in these areas.

A study at the University of Rhode Island observed that only 10% of daily commuters use means of transportation other than single-occupant vehicles (Flesser 2020). This approximate mode split is consistent with other universities and colleges. As a result, the roadways surrounding the campus were calculated as greatly exceeding their capacity at each peak hour. However, through advanced simulation, decreasing vehicular traffic by just 10% provided “significant relief” to the roadways. This study is one of many that indicate vehicular congestion can be decreased by increasing accessibility to bicycle facilities surrounding universities, while simultaneously enabling safe, efficient travel. Research has shown that bicycle ridership is rapidly increasing as a mode choice in both rural and college towns (Loukaitou-Sideris et al. 2014; Woodson et al. 2024). However, a lack of safe facilities in rural areas has caused this trend to lead to many serious and fatal crashes between cyclists and other modes of transportation, with many instances going unreported (Dobbs 2009; Medury et al. 2019). These types of statistics, in addition to a lack of proper infrastructure, lead to commuters opting for different modes of transportation. There is limited literature available on what specific types of infrastructure could be developed to foster the use of active transportation in rural environments, while simultaneously ensuring that the users are safe and unbothered when traveling. Using analyses of previously implemented bicycle infrastructure policies and conducting new surveys to determine what cities can do to incentivize bicycle ridership, planners can bridge this gap between urban and rural bicycle facility development.

This study investigates the factors influencing students’ mode choice behavior in rural college towns, shedding light on the barriers that prevent them from adopting more sustainable transportation options. More importantly, it examines how the introduction of new, well-

designed, and safer bicycle infrastructure such as protected bike lanes, dedicated cycling paths, and improved support facilities could encourage students to shift from private vehicle use to active transportation. Three critical bike-related issues are addressed in this study:

- **First**, there are no current, commonly used means of identifying which policies are effective in rural towns. This leads to municipalities or towns becoming hesitant about investment decisions regarding cycling infrastructure. These municipalities or towns do not have a mechanism to identify which bike policies or infrastructure would work best.
- **Second**, planning agencies across North America have been proactive in conducting before-and-after surveys to evaluate the effectiveness of different cycling infrastructure projects and programs. A limitation of the before-and-after approach is that implementation must occur first, which can be costly if the intervention proves ineffective. In such cases, the benefits can only be assessed after construction, making this a resource-intensive and time-consuming way to test new solutions.
- **Finally**, it is difficult to measure how cycling infrastructure increases local accessibility. Many existing cycling infrastructure projects do not measure accessibility as a benchmark of the infrastructure's development. According to the Federal Sustainability Plan, the USA aims to achieve net-zero emissions by 2050 (6). Some key actions include carbon-free electricity, zero-emission motor vehicles, etc. This mandate also aligns with many Vision Zero Plans, which aim to significantly decrease traffic fatalities and injuries by 2025. In the last 20 years, some cities (e.g., Washington, DC) built more than 94 miles of bike lanes. Among these 94 miles, around 17 miles are protected bike lanes. However, an adverse community reaction after building protected bike lanes in a few cases displays that protected bike lanes reduce parking capacity. Sometimes, communities are not aware of the environmental and equity benefits of protected bike lanes. In addition, protected bike lanes can help create calming traffic zones.

To address the above mentioned gap, this study designed a stated choice experiment to understand which factors encourage rural residents to use bikes as a mode choice for their daily activities. The study area includes rural college towns in Virginia State, USA. The survey comprises a wide range of socio-demographic information and personal, household, and trip-related attributes. Based on the collected data, an econometric model has been estimated, revealing the factors affecting individuals' mode choices if they are provided with sustainable bike infrastructure in rural Virginia. Ideally, showing the communities the environmental, accessibility, and equity benefits of cycling policies (e.g., protected bike lanes) can help them realize these policies' long-term benefits. This also helps college town officials pick appropriate policies to increase sustainable mode shares.

5.3 Survey Design

5.3.1 Study Area

This study focuses on post-secondary students' transportation behavior in 14 rural college towns across Virginia. These towns - Blacksburg, Christiansburg, Radford, Dublin, Pulaski, Roanoke, Salem, Lynchburg, Charlottesville, Harrisonburg, Williamsburg, Bridgewater, Hampden Sydney, and Longwood - were selected based on their classification as rural communities with significant student populations. Each of these towns is close to at least one higher education institution, making them ideal for studying transportation behavior among college students. Their population sizes, student percentages, and higher education institutions are summarized in Table 5.1.

The geographic and infrastructural characteristics of these towns vary, influencing the availability and use of different transportation modes. Some towns, such as Blacksburg and Charlottesville, have relatively well-developed public transit systems, while others rely more heavily on personal vehicles and active transportation modes like walking and biking. The

presence of universities also contributes to unique travel demand patterns, with students often relying on a mix of personal vehicles, ridesharing, public transit, and non-motorized modes to navigate their daily activities.

Additionally, these towns experience seasonal fluctuations in transportation demand, particularly during academic breaks when many students leave, affecting transit ridership and traffic patterns. The focus of the study on these towns provides valuable insights into how transportation behaviors shift in rural college environments, where infrastructure and transit options may differ from those in urban settings.

5.3.2 Stated Preference (SP) Survey

The Stated Preference (SP) survey was designed to evaluate how students in rural college towns would respond to hypothetical transportation scenarios, particularly in the context of new or improved infrastructure. This approach allowed for an in-depth analysis of potential shifts in student travel behavior under varying conditions. By incorporating hypothetical choices, the survey was designed to capture decision-making processes that may not be observable in real-world settings.

To ensure statistical rigor, the SP survey was developed using NGENE software, which facilitated the creation of D-efficient experimental designs. These designs maximize the amount of meaningful data obtained from respondents while minimizing the number of scenarios needed to achieve reliable and robust results. The design process followed a structured methodology to refine the survey parameters systematically:

Literature Review: The initial set of parameters was derived from existing research on student transportation behavior and cycling infrastructure preferences.

Pilot Survey: A preliminary survey was conducted to collect initial responses and test the effectiveness of the experimental design.

Design Optimization: Based on the pilot results, parameters were adjusted to reduce D-error and improve the S-estimate, ensuring that the final design was both statistically efficient and capable of capturing nuanced behavioral insights.

Participants were presented with various hypothetical transportation scenarios, each designed to measure their sensitivity to different attributes of cycling and broader mobility options. The scenarios covered multiple dimensions, including:

Types of Bike Lanes: Respondents evaluated four distinct bike lane configurations - designated lanes, buffered lanes, protected lanes, and two-way protected lanes. This helped assess how the presence and type of cycling infrastructure influence students' willingness to bike.

Supportive Facilities: The survey examined the impact of sheltered bike parking, showers, and changing rooms on cycling adoption, recognizing that access to these amenities can significantly enhance the feasibility of cycling as a primary mode of transportation.

Perceptions of Safety: Participants provided insights into their perceived risk of accidents when using different transportation modes and infrastructure, helping to gauge the psychological barriers to cycling and other active transportation options.

Operational Factors: Additional attributes were included to understand the broader factors influencing travel choices, such as maintenance requirements, parking availability, lane gradients, travel distance, travel time, congestion levels, CO2 emissions, and weather conditions.

Through these hypothetical scenarios, the SP survey offered an idea of how students might modify their travel behavior in response to infrastructure improvements or new policy interventions. By analyzing the responses, researchers can better predict the effectiveness of potential transportation investments and design strategies that align with student needs and preferences. Figure 5.1 shows a sample SP scenario.

Attribute / Travel mode	Driving	Auto Passenger	Uber	Carpool	Public Transit	Bike
These remain consistent between scenarios						
Total Travel Distance (miles)	5	5	5	5	5	5
Cost (Dollars)	0.70	0.35	13.50	0.35	Free	Free
Traffic Congestion Level	High	High	High	High	High	High
Carbon Di-Oxide Emissions	High	High	High	High	High	Low
Availability During Peak Hours to you	Always	Sometimes	Mostly all the time	Sometimes	For fixed hours	Always
Dependence on Weather	Independent	Independent	Dependent	Independent	dependent	Dependent
These changes between scenarios						
Shared bicycle lane	-	-	-	-	-	Yes
Sheltered bike parking, showers and changing room available at workplace	-	-	-	-	-	Yes
Risk of Accidents	-	-	-	-	Low	High
Maintenance Needs	Low	-	-	-	-	Low
Parking availability during peak hours	Available	-	-	-	-	Unavailable
Uphill Road	-	-	-	-	Yes	No
Downhill road	-	-	-	-	Yes	No

Figure 5.1: Sample Stated Preference Survey Scenario

5.3.3 Revealed Preference (RP) Survey

The RP survey was designed to collect information on participants' current travel behavior. Respondents were asked about the modes of transportation they used, trip purposes, travel times, and associated costs. This data provided a snapshot of how students navigate transportation systems under existing infrastructure and policies. The RP survey was implemented using QuestionPro, which facilitated efficient survey administration and screening. The survey was structured with multiple sections to guide participants smoothly through the process, as illustrated in Figure 2. The survey included:

Survey Introduction and Consent (Segment a): Participants were provided with detailed information about the study's purpose, confidentiality, and voluntary nature. They

were assured that their responses would be anonymized and accessible only to the research team.

Screening Questions (Segment b): These questions verified participants' eligibility. They were asked to confirm:

- Their consent to participate in the study
 - Their age eligibility (at least 18 years old)
 - Whether their home or work location was in one of the target towns
- Participants who did not meet these criteria were unable to continue the survey, ensuring that the dataset reflected responses from relevant individuals.

Survey Flow Customization (Segment c): The survey incorporated dynamic question logic to tailor the experience to each participant. For instance, participants who did not bike were automatically allowed to skip questions about cycling. This approach minimized respondent fatigue and improved the accuracy of responses by focusing only on relevant questions.

5.3.4 Necessity of Using RP-SP Survey together

Using both RP and SP surveys provided a more comprehensive understanding of transportation behavior than either approach alone. The RP survey captured actual behavior based on existing conditions, reflecting real-world constraints like infrastructure availability, travel costs, and time. However, it did not provide insights into how students would respond to potential changes in transportation systems. The SP survey filled this gap by allowing researchers to test various hypothetical scenarios. For example, it examined how different types of bike lanes, safety features, and supportive facilities might influence students' decisions to switch from driving to cycling. By combining RP and SP data, models can be developed that account for both current behaviors and future possibilities, providing robust

The screenshot shows a survey interface with the following sections and callouts:

- Study Description:** A text box explaining the survey's purpose, confidentiality, and voluntary participation details. Callout 'a' points to this section.
- Consent and Eligibility:** Two radio button questions: "Do you consent to participate?" and "Are you of at least 18 years old?". Callout 'b' points to these questions.
- Location:** A question asking if the respondent's home or work location is in any of the listed Virginia areas (Blacksburg, Christiansburg, Radford, Dublin, Pulaski, Roanoke, Salem, Lynchburg, Charlottesville, Harrisonburg, Williamsburg, Bridgewater, Hampden Sydney, Longwood). Callout 'b' points to this question.
- Cycling Safety:** A question asking "How safe do you feel while cycling on main roads in your area?" with a scale from "Extremely Unsafe" to "Extremely Safe" using smiley face icons. Callout 'c' points to this question.

Callout 'a' points to the Study Description section, explaining the survey's purpose, confidentiality, and voluntary participation details.

Callout 'b' points to the consent and eligibility questions, explaining that these questions verify consent, eligibility (age 18+), and location, with the survey halting if criteria are not met.

Callout 'c' points to the cycling safety question, explaining that participants were allowed to answer questions, with the option to skip if they do not cycle.

Figure 5.2: Sample Structure of RP Survey in QuestionPro

insights for transportation planning.

Table 5.1: Summary of Rural College Towns in Virginia Included in the Study

Town	Population*	% Students†	Higher Education Institutions
Blacksburg	~45,000	~55%	Virginia Tech
Christiansburg	~22,000	~15%	Near Virginia Tech, New River Valley Community College
Radford	~18,000	~45%	Radford University
Dublin	~2,500	~10%	New River Community College (main campus nearby)
Pulaski	~9,000	~10%	Near New River Community College

Continued on next page

Town	Population	% Students	Higher Education Institutions
Roanoke	~97,000	~12%	Hollins University, Roanoke College, Virginia Tech Carilion School of Medicine
Salem	~25,000	~25%	Roanoke College
Lynchburg	~80,000	~30%	Liberty University, University of Lynchburg, Randolph College
Charlottesville	~47,000	~45%	University of Virginia
Harrisonburg	~52,000	~40%	James Madison University, Eastern Mennonite University
Williamsburg	~15,000	~50%	College of William & Mary
Bridgewater	~6,000	~40%	Bridgewater College
Hampden Sydney	~1,500	~85%	Hampden–Sydney College
Farmville (Longwood)	~8,500	~60%	Longwood University

5.4 Descriptive Statistics

5.4.1 Mode Share Analysis

The Stated Preference (SP) survey dataset captures how individuals hypothetically prefer to travel under six different scenarios. Each scenario varied in travel distance, cost, congestion, emissions, and the presence of cycling infrastructure or supportive facilities, while keeping

*Population values are approximate based on U.S. Census Bureau 2020 estimates.

†Percent students refers to approximate share of post-secondary students in the local population

the available modes consistent: driving alone, being a passenger in a car, carpooling, using ridehailing, taking public transit, and cycling. In Scenario 1 (2 miles), motorized modes faced low costs but high congestion and emissions, while cycling was free with a designated lane and available parking. Scenario 2 (5 miles) introduced higher costs across modes, with cycling supported by a shared lane and workplace facilities, though accident risk was elevated. Scenario 3 (7 miles) featured lower congestion and emissions overall, with cycling benefiting from a two-way protected lane and workplace facilities, yet challenged by high maintenance needs and limited parking. Scenario 4 (10 miles) further increased costs, with public transit and cycling becoming less competitive due to higher accident risks and reduced parking availability. In Scenario 5 (15 miles), longer commutes brought high congestion and emissions for motorized modes, while cycling was supported by a protected lane and parking access but faced greater maintenance demands and safety concerns. Finally, Scenario 6 (20 miles) reflected the longest-distance trip, with buffered bicycle lanes and hilly terrain making cycling the least practical choice despite dedicated facilities, while personal vehicles remained reliable.

Across all scenarios, driving alone is the most preferred mode of transportation. In Scenario 1, 64% of individuals chose to drive. This trend continues across the remaining scenarios with similar high shares: 59% in Scenario 2, 59% in Scenario 3, 60% in Scenario 4, 66% in Scenario 5, and 65% in Scenario 6. The consistently high preference for driving suggests that individuals value the flexibility, independence, and convenience of personal vehicles over shared or public transportation options.

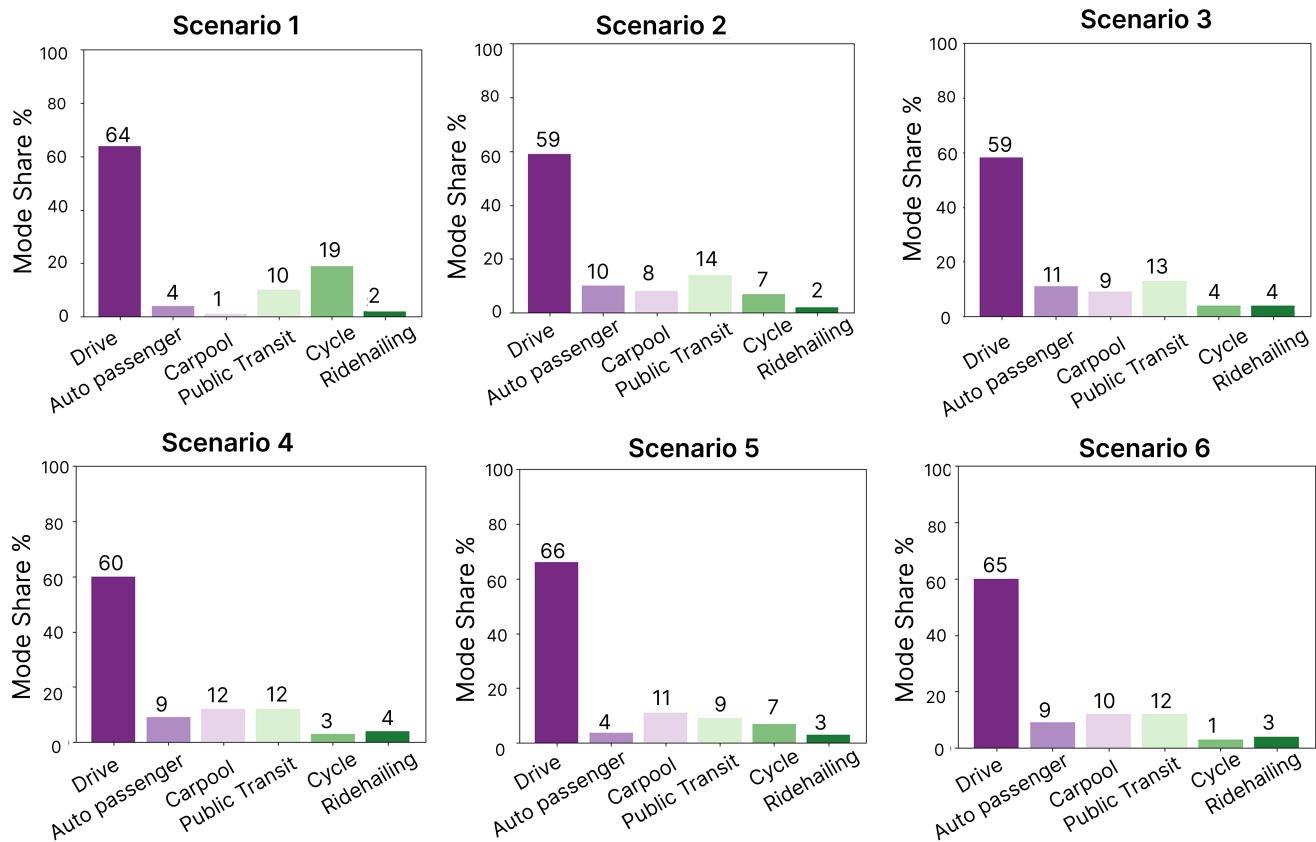


Figure 5.3: Comparative Analysis of Transportation Mode Share Across Six SP Scenarios

Table 5.2: Mode Share (%) Across Six SP Scenarios

Mode	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
Drive	64	59	59	60	66	65
Auto Passenger	4	10	11	9	4	9
Carpool	1	8	9	12	11	10
Public Transit	10	14	13	12	9	12
Cycle	19	7	4	3	7	1
Ridehailing	2	2	4	4	3	3

Being a passenger in a car is significantly less common than driving alone. In Scenario 1, 4% of respondents traveled as passengers. The percentages increased slightly in Scenario 2 (10%), Scenario 3 (11%), Scenario 4 (9%), Scenario 5 (4%), and Scenario 6 (9%). The lower numbers indicated that fewer people relied on others to drive them, likely preferring the control of driving themselves.

Carpooling, an option that promotes shared travel and sustainability, has the lowest preference among all modes. In Scenario 1, only 1% of participants opted for carpooling. Other scenarios showed slightly higher values, with 8% in Scenario 2, 9% in Scenario 3, 12% in Scenario 4, 11% in Scenario 5, and 10% in Scenario 6. The difficulty of coordinating with others and the lack of carpooling programs might have contributed to these low numbers.

Public transit usage varies across scenarios, reflecting differences in availability, reliability, and user preferences. Scenario 1 had the transit share of (10%), indicating an inclination towards public transport when it is accessible. However, the numbers dropped significantly in other scenarios: 14% in Scenario 2, 13% in Scenario 3, 12% in Scenario 4, 9% in Scenario 5, and 12% in Scenario 6. The cost-effectiveness and environmental benefits of public transit might make it a viable option, especially in areas with good transit systems.

Cycling, a mode of transportation that promotes sustainability and active mobility, was one of the least preferred options in most scenarios. In Scenario 1, 19% of individuals chose to bike. The other scenarios exhibited even lower rates: 7% in Scenario 2, 4% in Scenario 3, 3% in Scenario 4, 7% in Scenario 5, and 1% in Scenario 6. The low percentage of bikers suggests that safety concerns, lack of dedicated cycling infrastructure, and longer travel distances discourage individuals from choosing this mode.

Ridehailing usage remained relatively low but stable across scenarios. In Scenario 1, 2% of respondents chose ridehailing, with similar results in other scenarios: 2% in Scenario 2, 4% in Scenario 3, 4% in Scenario 4, 3% in Scenario 5, and 3% in Scenario 6. The moderate usage of ridehailing suggests that while it serves as a convenient alternative, cost and availability

might limit its popularity compared to other transportation options.

The Revealed Preference (RP) survey shows individuals real life usage of different modes. It highlights the dominance of private vehicle use, with 63.33% of respondents choosing to drive alone. This strong preference suggests that convenience, flexibility, and reliability play key roles in mode choice decisions, reinforcing a car-dependent travel culture.

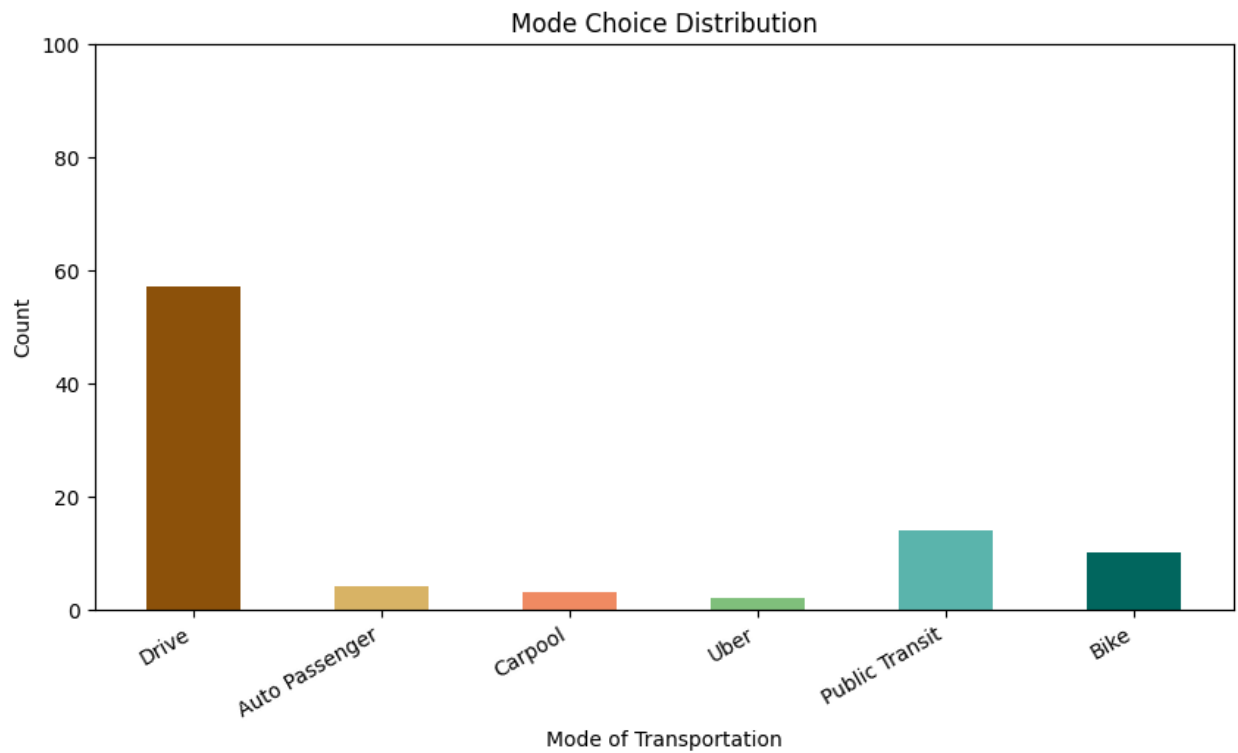


Figure 5.4: Mode Choice Distribution in RP Survey

Public transit emerges as the second most preferred option, with 15.56% of respondents opting for it. This indicates that while some individuals find transit viable, accessibility, service reliability, and travel time constraints may limit broader adoption. Similarly, cycling accounts for 11.11% of mode share, showing that a segment of the population is willing to cycle, but infrastructure quality and safety concerns could be barriers to wider usage.

Shared mobility options, such as auto passenger travel and carpooling, remain relatively low, with 4.44% and 3.33% participation, respectively. These low figures suggest that lack of carpool programs, difficulty in coordinating shared rides, and a preference for independent

travel may discourage their use. Ridehailing, the least chosen mode at 2.22%, indicates that factors such as cost and availability might limit its appeal compared to other travel options.

5.4.2 Sociodemographic Data Analysis

A total of 102 people participated in the survey, of whom 48 were male and 54 were female. When percentages were calculated, males were found to make up approximately 47.06% of the group, while females accounted for about 52.94%. This means that nearly half of the people are male, and just over half are female. The survey also reveals a notable age distribution. The largest group is aged 61-70 with 23 participants (about 22.55%) followed closely by the 31-40 age group which has 22 individuals (approximately 21.57%). The 41-50 range includes 18 people (around 17.65%) while the 21-30 group consists of 19 individuals (about 18.63%). In contrast, the younger 18-20 age group has only 7 participants (around 6.86%) and the 51-60 group contains 8 people (approximately 7.84%). The 71-80 age range is the least represented with just 5 individuals (about 4.90%). The average household size is approximately 2.94 individuals. Each household has an average of 1.91 functioning motor vehicles and about 1.97 functioning bicycles. This shows that people in the community use both cars and bikes for transportation.

5.5 Econometric Model Specification

In our study, two widely used discrete choice models were used: the Multinomial Logit (MNL) model and the Mixed Logit (MxL) model to analyze mode choice behavior using Revealed Preference (RP) and Stated Preference (SP) data. The RP data represented actual observed travel choices made by individuals, whereas the SP data captured hypothetical travel scenarios designed to understand how people might react to changes in the transportation system, such as the introduction of new travel modes or policy interventions.

By combining RP and SP data, the realism of RP choices was leveraged while incorporating the flexibility of SP choices, allowing us to estimate parameters that reflected actual decision-making while also capturing choices in potential future scenarios. The use of both MNL and MxL models in our study enabled us to examine how different assumptions about preference heterogeneity, correlation among alternatives, and error structures impacted the estimated mode choice probabilities.

5.5.1 Combined RP-SP Multinomial Logit (MNL) Model

The MNL model served as the foundation for our mode choice analysis due to its computational efficiency and well-established theoretical properties. It assumed that an individual's choice was determined by a systematic utility component (based on measurable attributes such as travel time and cost), with an unobserved error term that followed an independent and identically distributed (IID) Gumbel distribution.

This model was used in this study to provide a benchmark estimation, allowing us to analyze mode choice probabilities under the assumption that:

- Preferences were homogeneous across individuals (i.e., all travelers weighed travel time, cost, and other attributes in the same way).
- The error terms were IID across alternatives, leading to the Independence of Irrelevant Alternatives (IIA) property, which implied that the relative probabilities of choosing between two alternatives remained unchanged even if another alternative was introduced or removed.

5.5.2 Mathematical Formulation

For an individual i choosing an alternative j , the utility function in the RP dataset was given by:

$$U_{ij}^{RP} = \theta X_{ij}^{RP} + \alpha Z_{ij}^{RP} + \eta_{ij}^{RP} \quad (5.1)$$

For the SP dataset, the utility function followed a similar form:

$$U_{ij}^{SP} = \theta X_{ij}^{SP} + \gamma Z_{ij}^{SP} + \eta_{ij}^{SP} \quad (5.2)$$

where θ represents the vector of parameters to be estimated, X_{ij}^{RP}, X_{ij}^{SP} are the common attributes across all alternatives (e.g., travel time, travel cost), Z_{ij}^{RP}, Z_{ij}^{SP} are the alternative-specific attributes, α, γ are alternative-specific parameters, $\eta_{ij}^{RP}, \eta_{ij}^{SP}$ are the IID Gumbel-distributed error terms.

The choice probability for individual i selecting alternative j followed the standard logit function:

$$LL(\beta) = \sum_{i=1}^M y_{ij} \ln \left(\frac{e^{V_{ij}}}{\sum_{j' \in J} e^{V_{ij'}}} \right) + \sum_{i=1}^M \sum_{s=1}^Z y_{ijs} \ln \left(\frac{e^{\delta^{SP} V_j^{SP}}}{\sum_{j' \in n} e^{\delta^{SP} V_{j'}^{SP}}} \right) \quad (5.3)$$

Where, $LL(\beta)$: Log-likelihood function to be maximized, M : Total number of individuals in the sample, J : Set of available alternatives in the RP dataset, n : Set of available alternatives in the SP dataset, Z : Number of SP choice scenarios y_{ij} : Indicator variable:

- $y_{ij} = 1$, if individual i selects alternative j in the RP dataset.
- $y_{ij} = 0$, otherwise.
- y_{ijs} : Indicator variable:
 - $y_{ijs} = 1$, if individual i selects alternative j in SP scenario s .
 - $y_{ijs} = 0$, otherwise.

V_{ij} is the systematic utility component for the RP dataset, V_j^{SP} is the systematic utility component for the SP dataset, δ^{SP} : Scale parameter adjusting the variability of SP data relative to RP data, $e^{V_{ij}}$ is the exponential transformation of systematic utility for RP alternatives, $e^{\delta^{SP} V_j^{SP}}$ is exponential transformation of systematic utility for SP alternatives.

5.5.3 Combined RP-SP Mixed Logit (MxL) Model

Although the RP-SP MNL model effectively integrates real-world choices and stated preferences in a structured and interpretable framework, it has certain limitations in capturing the full complexity of travel behavior. Specifically, it assumes homogeneous preferences across individuals and adheres to the Independence of Irrelevant Alternatives (IIA) assumption, which may not always reflect real-world decision-making. The RP-SP Mixed Logit (MxL) model addresses these limitations by allowing preference heterogeneity, relaxing the IIA assumption, and modeling correlations between alternatives. By introducing random coefficients, the MxL model provides a more flexible and realistic representation of individual travel behavior.

5.5.4 Mathematical Formulation

For an individual i choosing an alternative j in the SP dataset, the utility function consists of:

$$U_{ij}^{SP} = V_{ij}^{SP} + \eta_{ij}^{SP} \quad (5.4)$$

where, U_{ij}^{SP} is the total utility of individual i for alternative j in the SP dataset, V_{ij}^{SP} is the systematic (deterministic) utility component based on observed attributes, η_{ij}^{SP} is the random error component capturing unobserved influences.

Expanding the systematic utility component V_{ij}^{SP} to include a random term, we get:

$$U_{ij}^{SP} = \theta x + \xi_l W_{jl} d_{jl} + \eta_{ij}^{SP} \quad (5.5)$$

where, θx is the effect of observed attributes (e.g., travel time, travel cost) on utility, $\xi_l W_{jl} d_{jl}$ is the random component allowing for variation in preferences across individuals, ξ_l represents unobserved preference heterogeneity, W_{jl} is a normally distributed variable that captures randomness in preferences, d_{jl} is a dummy variable, equal to 1 if alternative j belongs to a specific group (or "nest"), otherwise 0, η_{ij}^{SP} is the random error component, assumed to follow a Gumbel distribution.

Unlike the MNL model, the probability of choosing an alternative j in the SP dataset is obtained by integrating over all possible values of the random component:

$$P_{is}^{SP}(j) = \int \frac{\exp(\delta^{SP} V_{ij})}{\sum_{j'} \exp(\delta^{SP} V_{ij'})} g(\psi) d\psi \quad (5.6)$$

Where, $P_{is}^{SP}(j)$ is the probability that individual i selects alternative j in the SP dataset, δ^{SP} is the scale parameter, which adjusts for differences in variability between RP and SP choices, V_{ij} is the deterministic utility component, $g(\psi)$ is the probability density function (PDF) of the random parameters, which accounts for preference heterogeneity.

Since individuals in the SP dataset make repeated choices across different hypothetical scenarios, the joint probability of making a series of choices is given by:

$$P_i^{SP}(j_1, j_2, j_3, j_4, j_5, j_6) = \prod_{s=1}^6 P_{is}^{SP}(j_s) \quad (5.7)$$

Where, $P_i^{SP}(j_1, j_2, j_3, j_4, j_5, j_6)$ is that probability that individual i selects a particular sequence of choices in six SP choice scenarios. The product notation $\prod_{s=1}^6$ represents the joint

probability across all SP scenarios.

The log-likelihood function combines RP and SP choices into a single estimation framework:

$$LL(\theta, x, \xi, \psi, y) = \prod_{i=1}^N P_i^{RP}(j)^{y_{ij}^{RP}} \times P_{1i}^{SP}(q)^{y_{qi}^{SP}} \times P_{2i}^{SP}(q)^{y_{qi}^{SP}} \times P_{3i}^{SP}(q)^{y_{qi}^{SP}} \times P_{4i}^{SP}(q)^{y_{qi}^{SP}} \times P_{5i}^{SP}(q)^{y_{qi}^{SP}} \times P_{6i}^{SP}(q)^{y_{qi}^{SP}} \quad (5.8)$$

By including δ^{SP} in the model, it was ensured that the parameter estimates are not biased due to differences in how choices were made in RP vs. SP data. The RP-SP Mixed Logit (MxL) model improves on the RP-SP MNL model by:

- Incorporating random parameters for preference heterogeneity.
- Allowing for correlated choices across alternatives.
- Integrating both RP and SP data into a unified framework.

5.6 Model Results

5.6.1 Joint RP-SP MNL

This section discusses the econometric model results. Table 1 shows the parameters and t-stat values of the joint RP-SP MNL estimated model. The Rho squared values indicate that the model fits well, with a value of 0.33 and an adjusted value of 0.30, suggesting that the explanatory variables effectively capture travel preferences.

The results from the joint RP-SP MNL model provide important understandings of student transportation behavior in rural college towns. The findings highlight the challenges associated with biking, public transit, and shared mobility, while reinforcing the dominant role of private vehicles in these areas. One of the main findings from the joint RP-SP MNL model is that driving is still the most preferred mode of transportation in both the revealed and

stated preference data. This highlights the strong dependence on personal vehicles, likely due to their convenience and flexibility. While some respondents show interest in using public transit or carpooling, these options were not as popular as driving. On the other hand, ride-hailing services like ridehailing and cycling are less preferred, suggesting that there are still challenges in making these alternatives more appealing to travelers. Distance and cost are both significant deterrents for many travel modes. Longer distances reduce the likelihood of choosing options such as public transit, ridehailing, and cycling, especially when comfort, time, or physical effort is a concern. Cost also discourages the use of modes like driving, public transit, and shared travel options, suggesting that financial considerations weigh heavily in travel decisions.



Figure 5.5: Bike lane on Prices Fork Road, Blacksburg, VA



Figure 5.6: Bike Lane on Peppers Ferry Road, Christiansburg, VA

It has been found that individuals are less likely to bike even when designated bike lanes are provided. This result is reasonable, as designated bike lanes often run alongside mixed traffic, creating significant safety risks for cyclists, particularly in rural areas. For example, Prices Fork Road in Blacksburg has a designated bike lane, yet the traffic moves at speeds of up to 40 mph, making cycling unsafe. Additionally, several highway ramps near the intersection further increase the risk for cyclists (figure 5). Similar concerns are observed in Christiansburg, where traffic speeds vary between 35 and 55 mph, further limiting the effectiveness of designated bike lanes in encouraging cycling (Figure 6). Given these safety

concerns, it is not surprising that the SP model results show that respondents are more likely to choose a two-way protected bike lane over a designated lane in rural Virginia. A two-way protected bike lane provides a physical barrier between cyclists and motor vehicles, significantly reducing safety risks and making cycling a more attractive option. However, despite this preference, rural areas still lack investment in high-quality bike infrastructure. While cities like Washington, DC, have taken initiatives to build over 20 miles of protected bike lanes, similar investments are absent in small towns, raising concerns about the equity of transportation infrastructure across urban and rural areas Buehler et al. 2021. However, this type of bike infrastructure-related investment is absent in small towns in the USA, which brings the equity aspect. Another key finding is that students in rural areas overwhelmingly prefer driving or being auto passengers for long-distance trips and are less likely to use ride-hailing, public transit, or biking. This is largely due to the longer wait times for ride-hailing services in rural areas compared to urban settings like Atlanta or Chicago. Additionally, many small towns, including Christiansburg and Radford, VA, do not have a public transit agency, making both transit and ride-hailing unreliable for commuting, medical appointments, or airport travel. This eventually pushes people to rely on privately owned motorvehicle. As a result, it is not surprising at all that car ownership is much higher in rural areas than in urban areas in the USA (Moody et al. 2021).

The availability of bike maintenance facilities also plays a significant role in cycling decisions. The model results indicate that individuals are less likely to bike if the maintenance needs are high. This is particularly relevant in rural areas, where bike repair shops and maintenance stations are extremely rare. In addition, many small towns in Southwest Virginia are part of the Appalachian Mountains, making them naturally hilly. These steep terrains pose physical challenges to biking, further discouraging students from choosing cycling as a daily mode of transportation.

The estimated model also accounts for various socio-demographic factors, including gender, age, and household characteristics. The results indicate that Females are more likely to

choose driving, as shown by the positive coefficients in both the SP and RP models. This preference may reflect concerns about personal safety, time management, or comfort, which are often better addressed by private vehicle use. At the same time, women are less likely to choose carpooling, possibly due to reduced control over travel schedules or discomfort with sharing rides. Age also influences transportation decisions. Individuals aged 31 to 40 are less likely to choose driving or ridehailing, which may be due to lifestyle factors, financial considerations, or a greater openness to alternative modes of transportation. Their slight preference for being auto passengers suggests that this group may still rely on cars but not necessarily as drivers. Employment status has a significant impact. Full-time working students are more likely to drive, likely because of tight schedules and the need for flexibility in commuting. However, they are less likely to travel as auto passengers, indicating a preference for independence. In contrast, part-time students are more inclined to use public transit, which may reflect more flexible schedules or cost-saving behavior. They are notably less likely to choose auto passenger travel, possibly because they do not share rides frequently or lack consistent ride options. Individuals with more motor vehicles in their household are more likely to use both auto passenger and public transit modes. The positive association with public transit may reflect a greater openness to multimodal travel or simply more mobility options within the household. Travel frequency also matters. Respondents who take two or more trips per day are more likely to drive, which underscores the value of convenience and time savings in their daily routines. The flexibility of private vehicles makes them especially attractive for people with multiple destinations throughout the day.

Table 5.3: Joint RP–SP MNL Model Outcome

	SP				RP			
Parameters	Param.	Est.	t-stat	p-value	Param.	Est.	t-stat	p-value
Rho squared against a null model	0.33							
Adjusted Rho squared against a null model	0.30							
AIC	1435.95							
BIC	1513.81							
Alternative Specific Constant (ASC)	Drive	1.1525	1.1304	0.258	Drive	1.3220	2.1304	0.033
	Auto	0	0	1.000	Auto	0	0	1.000
	Passenger				Passenger			
	Carpool	0.7815	0.9490	0.343	Public	1.0434	1.6830	0.092
	Ridehail- ing	-0.962	-1.855	0.064	Transit			
	Public	-0.778	-1.758	0.079	N/A	N/A	N/A	N/A
	Transit				N/A	N/A	N/A	N/A

	SP				RP			
Parameters	Param.	Est.	t-stat	p-value	Param.	Est.	t-stat	p-value
	Cycle	-0.565	-1.589	0.112	N/A	N/A	N/A	N/A
Distance	ridehail- ing, Public Transit, Cycle	-0.1520	-1.458	0.145	Drive, Public Transit	-1.2570	-1.875	0.061
Cost	Drive, Auto Passenger, Carpool, Public Transit, Cycle	-1.2835	-1.567	0.117	Auto Passenger	-0.1653	-1.258	0.208
Time	N/A	N/A	N/A	N/A	Drive	-0.4672	-1.587	0.113
Designated bicycle lane	Bike	-1.1798	-2.547	0.011	N/A	N/A	N/A	N/A
Two-way protected Bicycle Lane	Bike	1.1627	2.578	0.010	N/A	N/A	N/A	N/A

	SP				RP			
Parameters	Param.	Est.	t-stat	p-value	Param.	Est.	t-stat	p-value
Parking availability during peak hour	Drive, Bike	-0.1826	-0.162	0.871	N/A	N/A	N/A	N/A
Maintenance Needs	Drive, Bike	-0.3467	-1.687	0.092	N/A	N/A	N/A	N/A
Uphill	Public Transit, Bike	-0.4043	-0.867	0.386	N/A	N/A	N/A	N/A
Downhill	Public Transit	-0.7387	-0.841	0.400	N/A	N/A	N/A	N/A
Female	Drive	0.7224	1.825	0.068	Drive	0.6637	1.647	0.100
	Carpool	-1.533	-0.832	0.405	Auto	2.5069	1.368	0.171
	Ridehail- ing	0.6858	0.657	0.511	Passenger N/A	N/A	N/A	N/A
Age between 31 to 40	Drive	-0.6782	-1.500	0.134	Drive	-0.4044	-1.426	0.154
	Ridehail- ing	-0.2530	-0.223	0.824	Auto Passenger	0.2072	0.069	0.945
Full time working student	Drive	1.3590	1.758	0.079	Drive	1.2347	1.689	0.091

	SP				RP			
Parameters	Param.	Est.	t-stat	p-value	Param.	Est.	t-stat	p-value
	Auto Passenger	-0.9580	-1.256	0.209	Auto Passenger	-0.7890	-1.360	0.174
Part-Time college/university student	Auto	-1.8570	-0.195	0.845	Auto	-2.6850	-0.195	0.845
	Passenger				Passenger			
	Public	1.0018	2.336	0.020	Public	1.0010	2.236	0.025
	Transit				Transit			
Number of motor vehicles	Auto	0.5640	1.256	0.209	Auto	0.3490	1.112	0.266
	passenger				passenger			
	Public	2.0230	1.563	0.118	Public	2.3210	1.664	0.096
	Transit				Transit			
Who take two trips per day	Drive	1.6850	1.786	0.074	Drive	2.4678	2.120	0.034
Scale Parameter	Household size (2 members)	-0.2780	-1.657	0.098	N/A	N/A	N/A	N/A

5.6.2 Joint SP-RP Mixed Logit Model

Table 3 presents the results of the joint Stated Preference (SP) and Revealed Preference (RP) mixed logit model, which analyzes travel behavior by combining real-world travel choices (RP) with hypothetical scenarios (SP). The mixed logit model allows for variations in individual preferences, making it more flexible than standard logit models. The Rho squared value (0.57) and the adjusted Rho squared value (0.48) indicate a good model fit, meaning the model effectively explains variations in travel choices. Higher values of Rho squared suggest that the model performs well in predicting individuals' mode choices.

In the SP model, driving continues to have a negative alternative-specific constant, suggesting that in hypothetical scenarios where more sustainable or shared alternatives are available, people are more open to reducing their reliance on personal vehicles. However, the RP model shows a positive constant for driving, confirming that in real-world situations, individuals still largely prefer using their own vehicles. Public transit appears positively in the RP data, while carpooling, ridehailing, and cycling remain less preferred, especially in the SP model. These contrasts reveal the behavioral gap between people's stated intentions and their actual commuting choices.

Distance and cost continue to significantly shape travel behavior. Longer distances reduce the appeal of ridehailing, cycling, and public transit, reflecting users' sensitivity to travel effort and time. Cost plays an important role in how people choose their travel mode. In both the stated and revealed preference data, higher costs make people less likely to choose certain options. This shows that affordability is a key factor when people decide how to travel. Additional situational factors also affect travel decisions. Parking availability and maintenance needs influence the appeal of driving and biking. The presence of parking during peak hours has a slightly positive effect on choosing these modes, while increased maintenance needs tend to discourage them.

Demographic variables remain highly influential. Women tend to prefer driving and ride-

hailing, while being less inclined to carpool or use public transit. Public transit also shows a notably negative coefficient among female respondents in the SP model, which may indicate safety concerns or discomfort with this mode. Individuals aged 31 to 40 are more likely to be auto passengers and less likely to drive, suggesting a preference for shared or passive travel among this age group. Employment status also plays a role. For example, full-time students are less likely to drive, while part-time students are more likely to use public transit and shared modes.

The model results provide a more nuanced understanding of how different types of bicycle infrastructure influence mode choice, particularly the decision to bike. While designated bicycle lanes are associated with a negative effect on biking, suggesting that their presence alone does not significantly encourage cycling, the model shows a positive effect for two-way protected bicycle lanes. This indicates that individuals are more likely to consider cycling when infrastructure offers a higher level of physical separation from motor vehicles which is safe. The negative coefficient for designated bicycle lanes may reflect ongoing concerns about safety, as these lanes are often only separated by paint and may still expose cyclists to fast-moving traffic, opening car doors, or intersection conflicts. In contrast, the positive association with two-way protected bicycle lanes suggests that infrastructure that physically separates cyclists from vehicle lanes significantly enhance the attractiveness of biking. These protected facilities likely improve riders' perceptions of safety and comfort, making cycling a more viable and appealing option for everyday travel. This finding reinforces the importance of not just providing bike lanes, but ensuring they are designed to meet users' needs for safety, continuity, and ease of use.

Table 5.4: Joint RP–SP MxL Model Outcome

	SP				RP			
Parameters	Param.	Est.	t-stat	p-val	Param.	Est.	t-stat	p-val
Rho squared against a null model	0.57							
Adjusted Rho squared against a null model	0.48							
AIC	1433.95							
BIC	1509.21							
No of simulations	400							
Alternative Specific Constant (ASC)	Drive	-0.8028	-1.156	0.248	Drive	0.322	1.1304	0.258
	Auto Passenger	0	0	1.000	Auto Passenger	0	0	1.000
	Carpool	-0.6302	-1.254	0.210	Public Transit	1.215	1.723	0.085
	Ridehailing	-1.8028	-1.156	0.248	N/A	N/A	N/A	N/A
	Public Transit	-1.1244	-1.328	0.184	N/A	N/A	N/A	N/A
	Cycle	-0.5650	-1.589	0.112	N/A	N/A	N/A	N/A

Table 5.4: Joint RP–SP MxL Model Outcome (continued)

	SP				RP			
Parameters	Param.	Est.	t-stat	p-val	Param.	Est.	t-stat	p-val
Distance	Ridehailing, Public Transit, Cycle	-0.1244	-1.350	0.177	Drive, Public Transit	-0.1156	-1.292	0.196
Cost	Drive, Auto Passenger, Carpool, Public Transit	-0.6302	-1.218	0.223	Auto Passenger	-0.4065	-1.345	0.179
Designated bicycle lane	Bike	-0.1798	-0.472	0.637	Drive	-0.3218	-1.256	0.209
Two-way protected Bicycle Lane	Bike	0.1432	1.752	0.080	N/A	N/A	N/A	N/A
Parking availability during peak hour	Drive, Cycle	0.2564	0.835	0.404	N/A	N/A	N/A	N/A
Maintenance Needs	Bike	-0.3467	-1.342	0.180	N/A	N/A	N/A	N/A
	Drive	0.5322	1.102	0.270	N/A	N/A	N/A	N/A
Female	Carpool	-1.5330	-1.832	0.067	Drive	0.7159	0.683	0.495

Table 5.4: Joint RP–SP MxL Model Outcome (continued)

	SP				RP			
Parameters	Param.	Est.	t-stat	p-val	Param.	Est.	t-stat	p-val
	Ridehailing	0.4858	0.657	0.511	Auto Passenger	2.3181	1.360	0.174
	Public Transit	-0.9734	-1.725	0.085	N/A	N/A	N/A	N/A
Age between 31 and 40	Auto Passenger	1.012	0.725	0.469	Auto Passenger	0.5403	0.891	0.373
	Drive	-0.3290	-1.346	0.180	Drive	-0.8239	-1.597	0.110
Full time working student	Drive	-0.6782	-1.500	0.134	Drive	-1.6548	-1.989	0.047
Part-Time college/university student	Ridehailing	-0.2530	-0.223	0.824	Auto Passenger	0.7419	1.442	0.149
	Drive	0.8425	-1.634	0.102	Public Transit	1.0018	2.336	0.019
More than two motor vehicles	Drive	0.2202	1.658	0.097	Drive	0.7288	0.883	0.377
	Auto passenger	0.4378	1.687	0.092	Auto passenger	-0.4174	-0.357	0.721
Who take two trips per day	Drive	0.5870	1.638	0.101	Drive	0.4310	1.456	0.145

Table 5.4: Joint RP–SP MxL Model Outcome (continued)

	SP				RP			
Parameters	Param.	Est.	t-stat	p-val	Param.	Est.	t-stat	p-val
Error parameter	Public transit	0.6880	1.524	0.128	N/A	N/A	N/A	N/A
Scale parameter	Household size (2 members)	-0.3450	-1.589	0.112	N/A	N/A	N/A	N/A

Frequent travelers tend to prefer private vehicles over public transport, primarily due to time efficiency and reliability concerns. Individuals who take two trips per day are less likely to use public transit, possibly because public transportation schedules and routes are not always convenient for multiple daily trips. Public transit may involve longer wait times, transfers, or service limitations, making it an inefficient option for those with tight schedules. Similarly, students who do not order online are more likely to drive, indicating that in-person shopping or errands increase car dependency. This suggests that people who frequently shop in physical stores or travel for errands require the flexibility and convenience that private vehicles offer, compared to the rigid schedules and carrying limitations of public transit. One key observation from the SP model is the error parameter for public transit, which indicates that people's perceptions of public transit vary significantly in hypothetical scenarios. This means that while some respondents might express an interest in using public transit in a survey, their real-world behavior may differ due to factors like delays, reliability, or personal habits.

Travel attributes such as distance and cost have strong effects in the SP model. The model shows that longer travel distances significantly reduce the likelihood of using ridehailing, public transit, or cycling. This suggests that respondents prefer private vehicles for longer trips, reinforcing findings from the RP model. Similarly, higher travel costs discourage the use of drive, auto-passenger, carpool, and public transit, meaning affordability is a critical factor in transportation decision-making.

One of the key areas explored in the SP model is bicycle infrastructure and its influence on cycling behavior. The results show that having a designated bicycle lane does not significantly encourage cycling. This may be because designated lanes still share space with vehicle traffic, making them feel unsafe. A two-way protected bicycle lane is slightly better than designated lanes, but it still does not strongly increase cycling. This suggests that improving infrastructure alone may not be enough to encourage more people to cycle, as other barriers such as safety perceptions, weather conditions, and travel distance play a role.

The availability of parking and maintenance needs also affect transportation preferences. Lack of parking discourages both driving and cycling, highlighting the importance of accessible parking facilities. In contrast, better parking availability during peak hours makes driving and cycling more attractive, showing that infrastructure availability can influence mode choice. It reinforces the idea that well-maintained roads and bike paths are crucial for encouraging sustainable transportation.

5.7 Conclusion and Recommendations

This study explored the transportation preferences of students in rural college towns, with a focus on understanding how the availability and quality of bicycle infrastructure influence mode choice. By employing a joint RP-SP modeling framework, first with the Multinomial Logit (MNL) model and then with the more flexible Mixed Logit model, both real-world and hypothetical decision-making behaviors were captured. The results confirmed that private vehicle use remains dominant, primarily due to its convenience and flexibility and not having the safe active transportation infrastructure in college towns. However, the analysis also revealed a latent openness to alternative modes, such as public transit and biking, particularly when safety, affordability, and infrastructure quality are improved. While most bike-related case studies highlight the success and adoption of micromobility in large metro areas such as Washington, DC, Boston, MA, and Chicago, IL, there are only a few that delve into rural transportation in the USA. This study, therefore, designed a joint RP-SP survey to understand the need for safer bike infrastructure in rural Virginia. The collected data was used to estimate econometric models, which capture individual mode choice preference under bike-related hypothetical scenarios.

The results show that individuals are deeply concerned about safe bike infrastructure in rural Virginia. They are particularly worried about the lack of transit integration and adequate

bike maintenance shops. Furthermore, when more people are in the household, there may be better options than bike riding. Based on the econometric model and description statistics, the following recommendations are provided:

- Investing in "safe" bike infrastructure is recommended, not just in "any" bike infrastructure. This study shows that people are particularly concerned about safe bike infrastructure and are more comfortable with protected bike lanes.
- Many town or city offices could open a small-scale bike repair and maintenance workshop where they could provide some tools and tips so individuals can fix their bikes themselves. This recommendation is also based on the econometric model, and it offers a practical and cost-effective solution to the lack of bike maintenance shops in rural areas.
- Another important aspect is providing enough bike racks or lockers in office spaces. A limited number of lockers are found in office spaces, which increases bike theft in rural areas.
- It is also recommended that a small-scale shower and locker facility be provided in the workplace so that individuals who bike for 5 to 15 miles can take some time and freshen up before starting work.
- It is also critical to integrate bike and transit. The buses in small towns should have bike racks in front or back of them, so individuals can take their bikes everywhere, which increases overall bike mobility manifold. Moreover, small towns should provide some sort of subsidy or discounted transit pass to families with a high number of members. This motivates families to choose more sustainable transportation modes. It is worth noting that some small towns (e.g., Blacksburg Transit) are already providing free transit services to the residents.

- The strong negative association between travel cost and mode choice indicates that affordability is a critical barrier. Institutions and local governments should consider offering student transit passes, subsidized bike-share programs, and financial incentives for non-driving modes.
- Women were more likely to drive and less likely to use public transit or carpool, possibly due to safety concerns. Planners should consider gender-sensitive design strategies such as better lighting, surveillance, and secure waiting areas to build trust in alternative modes.

The primary limitation concerns the relatively small sample size. Because data collection occurred during the summer semester, the pool of available participants was smaller than what would have been possible in the Fall or Spring, when campus populations are at their peak. A limited sample restricts the generalizability of findings and reduces statistical power. A larger and more representative sample would allow for richer comparisons between institutions, more reliable estimation of behavioral heterogeneity, and greater confidence in the policy implications drawn from the study.

Future research should place greater emphasis on collecting data at a larger scale and during academic terms when student presence is highest, such as the Fall and Spring semesters. Doing so would ensure that the sample more accurately reflects the true diversity of post-secondary students, capturing not only a larger number of respondents but also a broader spectrum of demographic, institutional, and geographic characteristics. A larger, more representative dataset would allow for deeper exploration of subgroup differences, such as variations across full-time versus part-time students, domestic versus international students, and those living on versus off campus. Such richness in data would enhance the generalizability of findings and strengthen the evidence base for transportation planning in college towns. In addition to expanding the scale and timing of data collection, future surveys could be re-

financed to capture a wider range of behavioral and attitudinal factors. Beyond recording travel times, costs, and distances, surveys could integrate questions about safety perceptions, environmental attitudes, social norms, and the role of academic or work schedules in shaping travel choices. Collecting this type of information would help researchers better understand the motivations and constraints that underlie observed travel patterns, particularly in the context of active modes such as cycling.

Chapter 6

Potential Health Benefits of Choosing Active Mode in Rural College Towns in USA

6.1 Chapter Overview

This chapter investigates the role of cycling as a catalyst for improving mental health among residents of rural communities in Virginia. While the physical benefits of cycling are well-documented, this study focuses on its mental health advantages—particularly in resource-limited rural settings where barriers such as poor infrastructure and limited access to alternative transportation modes are prevalent. The chapter fills a critical gap by centering its analysis on how cycling contributes to mental well-being across demographic and behavioral dimensions.

Using a combination of Principal Component Analysis (PCA), Random Forest modeling, and SHAP (SHapley Additive exPlanations), the study offers a multi-layered, data-driven understanding of how and why cycling impacts rural mental health. PCA distills key dimensions of mental health benefits, such as stress reduction and sleep improvement, while Random Forest quantifies the predictive power of various cycling behaviors and demographic features. SHAP enhances interpretability by identifying the most influential predictors driving mental health outcomes.

The analysis reveals that frequent biking, use of cycling for stress and anxiety management, and perceived improvements in sleep are among the strongest contributors to mental health enhancement. These effects are magnified among specific subpopulations, including older adults, individuals with years of cycling experience, and those who rely on cycling as a coping mechanism. Gender, age, cycling infrastructure quality, and medication needs further influence outcomes.

By combining survey-based behavioral data with machine learning methods, the chapter not only reinforces the mental health benefits of cycling but also emphasizes the importance of tailored, infrastructure-informed interventions. It concludes with practical recommendations for rural transportation planners and health policymakers aimed at maximizing biking's positive impact on community well-being.

6.2 Introduction

Transportation is a key factor influencing overall well-being, with far-reaching effects on both physical and mental health. Reliable transportation provides access to education, employment, healthcare, and social engagement, which are essential for maintaining a good quality of life and reducing health disparities (Moodie et al. [2013](#)). Conversely, transportation challenges such as long commute times, traffic congestion, and lack of mobility options can negatively impact mental health, contributing to increased stress levels, social isolation, and reduced life satisfaction (Clark et al. [2020](#)). The mode of transportation individuals use plays a crucial role in shaping these outcomes. Among various transportation options, active transportation modes like cycling and walking have been widely recognized for their health-promoting potential. Physical activity, including cycling, has consistently been associated with reduced psychological stress and improved mental health outcomes (Deslandes et al. [2009](#)). Despite these benefits, many individuals find it difficult to allocate dedicated time for

physical activity due to busy schedules (Manaf [2013](#)). Integrating cycling into daily travel routines such as commuting or running errands, provides an accessible way to achieve regular physical activity, thereby contributing to better mental well-being (Wojan and Hamrick [2015](#)). Of all types of physical exercise examined, cycling had the second largest positive impact on reducing poor mental health (Deslandes et al. [2009](#); Taylor et al. [1985](#)). This correlation between physical and aerobic exercise, such as biking, and positive mental well-being has been verified for decades, as shown in the results of studies that examine physical activity's effect on neuropsychiatric disorders due to aging, substance abuse, self-image, and cognitive functioning (Deslandes et al. [2009](#); Taylor et al. [1985](#)). For other bike riders who may not engage in cycling as a form of high-strength physical exercise, research congruently shows that low to moderately intense activity can be a key agent in reducing depressive symptoms (Weyerer and Kupfer [1994](#)). This was concluded by observing the symptoms of patients who frequently engage in low-intensity exercise and receive psychiatric counseling compared to those patients who solely receive counseling without exercising. The findings of these medical studies draw parallels to results of other research that examine the positive effects of aerobic exercise, including biking, on both physical and mental health (Kandola, Hendrikse, et al. [2016](#); Morres et al. [2019](#); Yao et al. [2021](#)).

One critical obstacle to gaining the full potential of cycling is that rural America lacks safe, affordable, equitable, and efficient bike infrastructure. Rural areas cover 68% of the total road miles in the United States. However, the car is the most dominant mode of transportation for individuals for working, shopping, and traveling long distances travels. According to the National Household Travel Survey 2022, 5.7% of respondents reported using a bike at least once in the last 30 days. This number is much higher for the urban areas (11.6%) (Bricka et al. [2024](#)).

cycling has been widely acknowledged for its physical health benefits, yet its role in enhancing mental health, particularly in rural settings, remains underexplored. While urban areas boast extensive research and infrastructure supporting the mental health benefits of cycling,

rural communities face unique barriers like limited cycling paths, safety concerns, and sparse recreational options. These limitations underscore the need for understanding how cycling impacts mental health differently in rural versus urban environments. Motivated by this gap, this study explores the mental health benefits of cycling in rural towns across Virginia, addressing specific factors that contribute to or limit these benefits.

The study is structured around three key research questions:

Research Question #1: What factors of cycling influence mental health outcomes for residents in rural America, where infrastructure and accessibility challenges are unique?

Research Question #2: Which demographic and behavioral characteristics amplify the mental health benefits of cycling for rural populations?

Research Question #3: How can advanced machine learning techniques like PCA, Random Forest, and SHAP analysis provide a data-driven perspective on the factors that influence biking's mental health benefits in rural settings?

To address these questions, methods such as Principal Component Analysis (PCA) were applied to simplify the dataset by reducing its dimensionality while enabling us to focus on the core variables that drive mental health outcomes, and a Random Forest model to predict mental health outcomes based on various demographic and behavioral factors. SHAP analysis was further used to offer interpretability, pinpointing the most influential variables. Each of these models contributed a distinct layer: PCA provided data clarity, Random Forest delivered predictive insights, and SHAP added interpretability, together offering a comprehensive understanding of biking's impact on mental health.

This paper follows a structure that includes the Literature Review, Survey Design, Methodology, Data Analysis and Result Analysis, and ends with Conclusion and Future Recommendations to guide the reader through the study's background, methods, findings, and implications.

6.2.1 Survey Design Overview

The survey was conducted via a multi-section, web-based questionnaire. Respondents were presented with six travel scenarios involving different distances and transportation modes, including driving, carpooling, auto passenger, ridehailing, public transit, and cycling with specific infrastructure features (e.g., designated bike lanes, shared lanes). Each scenario was designed to understand preferences influenced by factors like cost, environmental impact, and available infrastructure. The Stated Preference (SP) survey design was developed using NGENE software to ensure a D-efficient experimental design. Initial parameters were sourced from the existing literature to set up reasonable attribute levels and utility functions. To validate and refine the survey, a pilot survey was conducted to collect preliminary data. Based on the pilot survey outcomes, parameters were refined to reduce D-error and improve the S-estimate, which allowed for more efficient model estimation and better predictive performance. The approximately 20-minute survey gathered data on participants' preferences and behaviors related to cycling and other transportation modalities by altering factors influencing their choices. The study maintained anonymity, ensuring the confidentiality of responses and aiming to inform future bicycle infrastructure improvements in specific areas of Virginia.

6.2.2 Study Area and Participant Selection

Participants who live and/or work in different rural towns in Virginia, e.g., Blacksburg, Christiansburg, Radford, Pulaski, Dublin, Roanoke, Salem, Lynchburg, Charlottesville, Harrisonburg, Williamsburg, Bridgewater, Hampden Sydney, Longwood were preliminarily acquired through personal connections of the research team via email, and later through Dynata, a market research company. The towns were selected based on key rural characteristics. Specifically, rural college towns within Virginia with a population size generally less than 40,000 people were prioritized, in alignment with federal rural definitions and transporta-

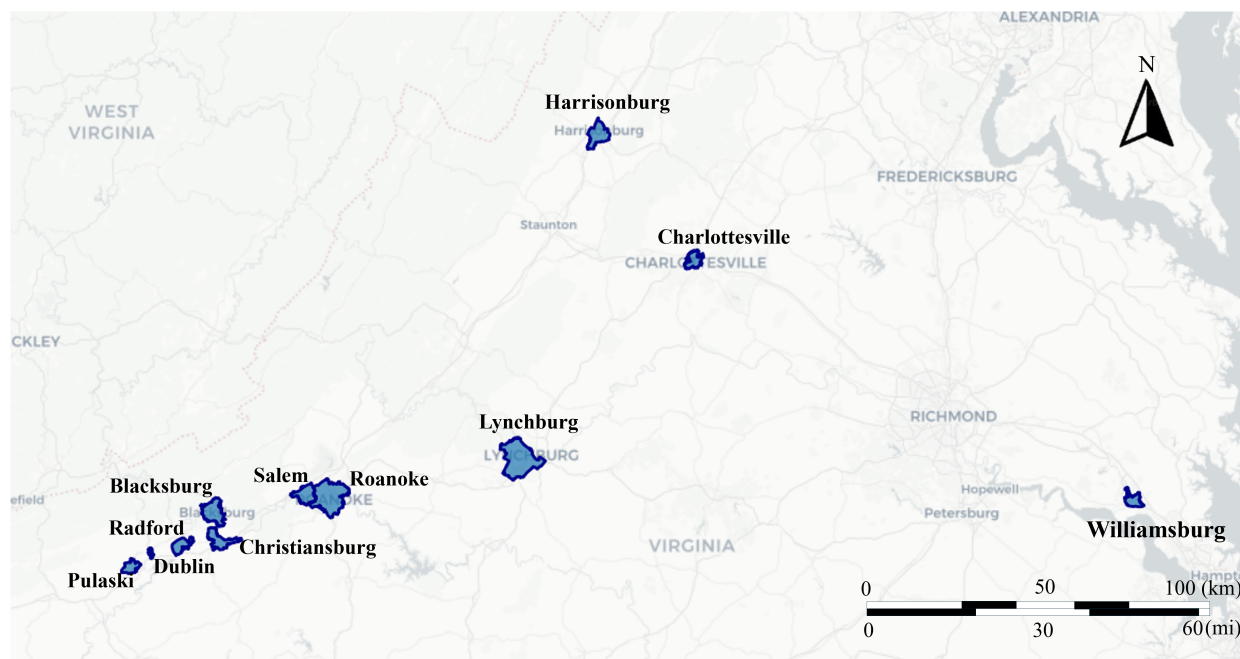


Figure 6.1: Geographic Distribution of Surveyed Rural Towns in Virginia

tion research standards (U.S. Census Bureau [2016](#)). Additionally, geographic diversity was ensured by selecting towns located primarily in the southwestern and central regions of Virginia. This selection approach was meant to capture the experiences of college towns with limited transportation options, smaller populations, and similar economic and social conditions common in rural areas. These towns are classified as rural, as they generally feature low population density, limited infrastructure, economic reliance on agriculture or small-scale industries, and are located away from major metropolitan hubs (Deavers [1992](#)). Participants confirmed their eligibility based on age, residence, or employment in these specific areas. This preliminary filter ensured that the data collected were relevant to the local infrastructure.

Table 6.1: Summary of Rural College Towns in Virginia Included in the Study

Town	Population*	% Students†	Higher Education Institutions
Blacksburg	~45,000	~55%	Virginia Tech
Christiansburg	~22,000	~15%	Near Virginia Tech, New River Valley Community College
Radford	~18,000	~45%	Radford University
Dublin	~2,500	~10%	New River Community College (main campus nearby)
Pulaski	~9,000	~10%	Near New River Community College
Roanoke	~97,000	~12%	Hollins University, Roanoke College, Virginia Tech Carilion School of Medicine
Salem	~25,000	~25%	Roanoke College
Lynchburg	~80,000	~30%	Liberty University, University of Lynchburg, Randolph College
Charlottesville	~47,000	~45%	University of Virginia
Harrisonburg	~52,000	~40%	James Madison University, Eastern Mennonite University
Williamsburg	~15,000	~50%	College of William & Mary
Bridgewater	~6,000	~40%	Bridgewater College
Hampden Sydney	~1,500	~85%	Hampden–Sydney College

Continued on next page

Town	Population	% Students	Higher Education Institutions
Farmville (Longwood)	~8,500	~60%	Longwood University

6.2.3 Sampling Method

Participants were recruited in two phases. Initially, participants were recruited through personal and professional contacts of the research team via email. To diversify the sample and enhance representativeness, recruitment was later expanded using Dynata. This study used random sampling through the Dynata panel to recruit participants and ensured a broader and more representative sample of rural college town residents. Random sampling was chosen because it reduces selection bias and increases the likelihood of capturing diverse demographic and behavioral profiles, making the findings more generalizable to the wider population of rural college towns.

6.2.4 Verification Process

The preliminary questionnaire of the survey ensures that respondents are appropriate sources of data based on specific criteria. It confirms that all respondents consent to participate in the survey, are at least 18 years of age, and either reside or work in the aforementioned areas in Virginia. The survey proceeds to other sections if all criteria are met and is terminated if any are not.

*Population values are approximate based on U.S. Census Bureau 2020 estimates.

†Percent students refers to approximate share of post-secondary students in the local population

6.2.5 Data Collection

Travel Mode Selection and Scenario Questions

The SP instrument, developed with NGENE for a D-efficient design and refined through a pilot, presented six hypothetical scenarios varying modes (drive alone, auto passenger, carpool, ridehailing, public transit, cycle), trip distances (2, 5, 7, 10, 15, 20 miles), and attributes such as cost, environmental impact, and bicycle-infrastructure details (e.g., designated or shared bike lanes). These items were designed to capture respondents' hypothetical choices under controlled changes in costs, time, and infrastructure quality. The RP instrument, administered through QuestionPro, recorded respondents' current travel behavior under existing conditions, for example, modes used, trip purposes, travel times, and costs, providing a real-world behavior.

Mental Health and Cycling Questions

Participants were asked about various aspects related to mental health, including whether they had any mental health issues, if they used cycling as a means to manage these problems, the duration of their cycling activities aimed at improving mental health, their sleep patterns, and the number of trips taken by cycling.

Building on these self-reported indicators, the RP dataset was analyzed to understand the connection between travel behavior and mental health. This dataset contained students' actual mode choices (e.g., cycling, driving, transit) along with associated trip attributes such as travel times, distances, and costs, in addition to socio-demographic information. Because the RP data reflected real-world choices under existing infrastructure conditions, it provided an opportunity to directly link observed cycling participation with mental health outcomes. Specifically, the responses of students who reported cycling as part of their daily travel were examined to explore whether their travel patterns were associated with indicators

such as lower stress levels, greater life satisfaction, and fewer self-reported mental health burdens. This integration of behavioral and health variables offered a grounded view of how active travel, particularly cycling relates to student well-being in rural college towns.

6.2.6 Ethical Considerations

The study was conducted under the guidelines of an approved Institutional Review Board (IRB) of Virginia Tech. Participants were informed about the purpose of the study, their voluntary participation, and the confidentiality of their responses. They were also given contact information for the research team and the IRB for any inquiries or concerns regarding their rights and the procedures of the study.

6.3 Methodology

This methodology section covers the different models that was used to determine the impacts of cycling on mental health. A dataset consisting of hundred individuals was analyzed who provided responses on cycling behaviors and mental health outcomes. The Random Forest (RF) model was selected for its ability to reduce overfitting through the ensemble learning technique and effectively capture complex, non-linear patterns in the data without the need for extensive hyperparameter tuning. Moreover, RF provides easy-to-understand feature importance scores, which helped us explain our model's predictions using SHAP values. These characteristics made RF a practical and effective choice for our study.

6.3.1 Principal Component Analysis (PCA)

The dataset contained many variables, and Principal Component Analysis (PCA) was applied to simplify it. This method changes the original variables into new, uncorrelated

variables called principal components (Bro and Smilde 2014). The first component shows the most variation, followed by the second, and so on. It changed our data into a simpler form with fewer dimensions while keeping most of the important information. The PCA process involves several mathematical steps:

Data Standardization

Let $D \in \mathbb{R}^{100 \times 43}$ denote the feature matrix, where each of the 100 rows represents a participant, which is denoted by i and each of the 43 columns corresponds to a biking-related behavioral or demographic feature, for example positive impact on sleep patten from biking, cyclinf distance per week, years of cycling experience, cycling to manage mental health, gender, race, etc, which is denoted by j . To ensure that all features contributed equally during model training, standardization was applied using the z-score formula:

$$D_{\text{standardized}}^{(i=100, j=43)} = \frac{D^{(i,j)} - \mu_j}{\sigma_j}, \quad \text{where } i \in \{1, 2, \dots, 100\}, \quad j \in \{1, 2, \dots, 43\}$$

Here, $D^{(i,j)}$ is the value of the j -th feature for the i -th participant, μ_j is the mean of the j -th feature, and σ_j is the corresponding standard deviation. The standardized dataset $D_{\text{standardized}}$ thus has zero mean and unit variance across each feature column:

$$\mu_j = \frac{1}{100} \sum_{i=1}^{100} D^{(i,j)}, \quad \sigma_j = \sqrt{\frac{1}{100} \sum_{i=1}^{100} (D^{(i,j)} - \mu_j)^2}, \quad \text{where } j \in \{1, 2, \dots, 43\}$$

100 is the number of participants here. This preprocessing step ensures that no single feature disproportionately influences the PCA and machine learning models used in the subsequent analysis.

Covariance Matrix and Eigen Decomposition

After standardizing the data, the covariance matrix $C \in \mathbb{R}^{43 \times 43}$ was computed to capture the linear relationships between biking-related features:

$$C_j = \frac{1}{n-1} \sum_{i=1}^n D_{\text{standardized}}^{(i,j)} \cdot \left(D_{\text{standardized}}^{(i,j)} \right)^\top, \quad \text{where } i \in \{1, \dots, 100\}, j \in \{1, \dots, 43\}$$

Here, $D_{\text{standardized}} \in \mathbb{R}^{100 \times 43}$ is the standardized data matrix, where each row i represents a participant and each column j a biking-related feature. The transpose $D_{\text{standardized}}^\top \in \mathbb{R}^{43 \times 100}$ reverses the dimensions, such that each row corresponds to a feature and each column to a participant, enabling the calculation of the covariance matrix across features. Each $C_j \in \mathbb{R}^{43 \times 43}$ captures the covariance structure associated with feature j . Eigen decomposition was then performed on the covariance matrix C to obtain eigenvalues λ_j and corresponding eigenvectors u_j :

$$Cu_j = \lambda_j u_j, \quad \text{for } j \in \{1, 2, \dots, 43\}$$

Each eigenvector $u_j \in \mathbb{R}^{43}$ represents the j -th principal component (i.e., a direction in the feature space), and each eigenvalue λ_j quantifies the amount of variance explained by that component. The eigenvalues were sorted in descending order, and the top components with the highest λ_j values were retained to reduce dimensionality while preserving the most significant variance in the data.

Principal Component Projection

Let $D_{\text{standardized}} \in \mathbb{R}^{100 \times 43}$ denote the standardized data matrix, where each row $i \in \{1, 2, \dots, 100\}$ represents a participant, and each column $j \in \{1, 2, \dots, 43\}$ corresponds to a biking-related

behavioral or demographic feature. To reduce dimensionality while preserving the most significant variance in the data, $D_{\text{standardized}}$ was projected onto a lower-dimensional subspace defined by the top k principal components:

$$D_{\text{PCA}}^{(i,k)} = \sum_{j=1}^{43} D_{\text{standardized}}^{(i,j)} \cdot u_k^{(j)} \quad \text{for } i \in \{1, \dots, 100\}, k \in \{1, \dots, 10\}$$

Here, $u_k \in \mathbb{R}^{43}$ is the k -th eigenvector (principal component), and $D_{\text{PCA}}^{(i,k)}$ represents the projection of the i -th participant's standardized feature vector onto the k -th principal component.

In matrix form, the projection is written as:

$$D_{\text{PCA}} = D_{\text{standardized}} \cdot U_k$$

where $U_k \in \mathbb{R}^{43 \times 10}$ is the matrix containing the top $k = 10$ eigenvectors as columns, and $D_{\text{PCA}} \in \mathbb{R}^{100 \times 10}$ is the transformed data matrix in the reduced subspace. These top 10 components were selected based on their eigenvalues, as they capture the majority of variance across the original 43 features, enabling more interpretable and computationally efficient analysis in subsequent machine learning models.

Explained Variance

To evaluate the significance of each principal component, the proportion of variance explained was calculated by the i -th principal component using its corresponding eigenvalue λ_i :

$$\text{Explained Variance Ratio}_i = \frac{\lambda_i}{\sum_{j=1}^{43} \lambda_j}, \quad \text{for } i \in \{1, 2, \dots, 43\}$$

Here, λ_i denotes the eigenvalue associated with the i -th principal component, which quantifies the amount of variance captured along that component's direction. The denominator is the sum of all 43 eigenvalues, representing the total variance in the standardized dataset. The cumulative explained variance ratio was used to determine how many principal components were needed to retain a substantial proportion of the dataset's total variance. In our analysis, the top $k = 10$ components were selected, as they preserved the majority of the variance while reducing the original 43-dimensional space to a more tractable 10-dimensional representation.

6.3.2 Random Forest Model

In this study, a Random Forest (RF) classification model was developed to predict the likelihood of experiencing positive mental health impacts from biking. The model was trained on a processed dataset containing $n = 100$ participants and $p = 14$ features, including both categorical and numerical variables.

Data Preprocessing

To prepare the raw dataset for machine learning modeling, a two-step preprocessing pipeline was applied to handle both categorical and numerical features derived from the 100 survey participants:

- **Encoding Categorical Features:** Demographic variables such as gender, race, and employment status were categorical in nature and required conversion to numerical format. These variables were encoded using One-Hot Encoding, producing a binary indicator matrix for each category level. The transformation is defined as:

$$X_{\text{categorical}}^{(i,j)} \rightarrow X_{\text{encoded}}^{(i,k)} = \text{OneHotEncode}(X_{\text{categorical}}^{(i,j)}), \quad \text{where } i \in \{1, \dots, 100\}$$

Here, $X_{\text{categorical}} \in \mathbb{R}^{100 \times m}$ is the original matrix of m categorical variables, and $X_{\text{encoded}} \in \mathbb{R}^{100 \times p}$ is the resulting matrix after One-Hot Encoding, with $p \geq m$ indicating the expanded feature space due to multiple category levels.

- Combining Numerical and Encoded Features: Numerical variables such as age, cycling frequency, cycling duration, and self-reported stress levels, etc. were concatenated with the encoded categorical variables to form the final machine learning input matrix:

$$X \in \mathbb{R}^{100 \times q} = [X_{\text{encoded}} \mid X_{\text{numerical}}]$$

where q denotes the total number of columns after combining both types of features. This unified feature matrix X served as input for the subsequent Random Forest and SHAP-based classification and interpretation.

Model Structure

The Random Forest classifier used in this study consists of $B = 200$ decision trees. The model incorporated the following key hyperparameters:

- Number of estimators (trees): $n_{\text{estimators}} = 200$
- Maximum depth of each tree: $\text{max_depth} = 10$
- Minimum samples required to split a node: $\text{min_samples_split} = 5$
- Class weights: $\text{class_weight} = \{0 : 1, 1 : 4\}$ to address residual imbalance

Decision Rule: Gini Impurity in Random Forest Splitting

In our Random Forest model, each decision tree uses the Gini impurity criterion to determine the best feature-based splits for classifying whether a participant experienced mental health

improvement from biking.

Let a node m contain a subset of training data $\{(x^{(i)}, y^{(i)})\}_{i=1}^{N_m}$, where:

- $x^{(i)} \in \mathbb{R}^{43}$ denotes the feature vector for participant $i \in \{1, \dots, 100\}$, consisting of 43 biking-related and demographic attributes.
- $y^{(i)} \in \{0, 1\}$ is the binary class label, where $y^{(i)} = 1$ indicates reported mental health improvement from biking, and $y^{(i)} = 0$ indicates no improvement.
- N_m is the number of participants in node m .

The Gini impurity G_m for node m is defined as:

$$G_m = 1 - \sum_{k=0}^1 \left(p_k^{(m)}\right)^2$$

Here, $p_k^{(m)}$ is the proportion of participants in node m that belong to class $k \in \{0, 1\}$, computed as:

$$p_k^{(m)} = \frac{1}{N_m} \sum_{\ell=1}^{N_m} \mathbb{I}(y^{(\ell)} = k)$$

where $\mathbb{I}(y^{(\ell)} = k)$ is the indicator function that equals 1 if participant ℓ in node m belongs to class k , and 0 otherwise. Here, $k \in \{0, 1\}$ represents the binary class label indicating mental health improvement due to biking:

- $k = 1$: Participant reported mental health improvement from biking.
- $k = 0$: Participant did not report mental health improvement from biking.

To determine the best split, the algorithm evaluates candidate thresholds for each feature $j \in \{1, \dots, 43\}$ and computes the resulting left and right child nodes m_L and m_R . The optimal split minimizes the weighted Gini impurity:

$$\min_{\text{split}} \left(\frac{N_{m_L}}{N_m} G_{m_L} + \frac{N_{m_R}}{N_m} G_{m_R} \right)$$

where:

- N_{m_L} and N_{m_R} are the number of participants assigned to the left and right child nodes based on the chosen threshold. For example, if the selected feature is weekly cycling frequency (i.e., $j = 12$ for weekly cycling frequency), and the threshold is 3, then participants with a frequency ≤ 3 are assigned to the left child node m_L , while those with a frequency > 3 are assigned to the right child node m_R .
- G_{m_L} and G_{m_R} are the Gini impurities of those child nodes.

This recursive splitting process continues until stopping criteria are met, enabling the model to learn complex patterns in cycling behavior and demographic profiles that predict mental health benefits.

Feature Importance

To quantify the influence of each feature on the classification outcome, feature importance scores were computed by averaging the Gini impurity reductions caused by splits on each feature across all decision trees in the Random Forest.

Let $q = 43$ be the total number of biking-related and demographic features, and let $B = 200$ be the total number of decision trees in the model. For a given feature $j \in \{1, 2, \dots, 43\}$, the importance score is computed as:

$$\text{Importance}(j) = \frac{1}{B} \sum_{b=1}^B \sum_{m \in T_b(j)} \Delta G_m(j)$$

Here:

- $T_b(j)$ is the set of nodes in tree T_b where feature j was used for splitting.
- $\Delta G_m(j) = G_m - \left(\frac{N_{m_L}}{N_m} G_{m_L} + \frac{N_{m_R}}{N_m} G_{m_R} \right)$ is the decrease in Gini impurity at node m due to splitting on feature j .
- N_m , N_{m_L} , and N_{m_R} are the number of participants in the parent node and its left and right child nodes, respectively.
- G_m , G_{m_L} , and G_{m_R} are the Gini impurities of the parent and child nodes.

This formulation reflects how much each feature j contributes to improving classification purity across all trees, allowing us to rank the features by their predictive power in identifying participants who report mental health improvement from biking.

Performance Evaluation

The trained model was evaluated on the test set using standard classification metrics:

- **Accuracy:** Proportion of correctly classified instances.
- **Confusion Matrix:** Capturing true positives, false positives, true negatives, and false negatives.
- **Classification Report:** Including precision, recall, and F1-score for both classes.

The final accuracy of the Random Forest classifier was approximately 84%, with superior recall and precision scores for the 'Positive Impact' class.

SHAP Value Computation

To interpret the predictions made by the Random Forest model for each participant, SHAP (SHapley Additive exPlanations) values were computed. These values quantify the contribution of each input feature to the model's output for a given participant.

Let $n = 100$ denote the number of participants and $q = 43$ the total number of biking-related and demographic features used in the model. For each participant $i \in \{1, 2, \dots, 100\}$ and each feature $f \in \{1, 2, \dots, 43\}$, the SHAP value $\phi_f^{(i)}$ is computed as:

$$\phi_f^{(i)} = \sum_{S \subseteq F \setminus \{f\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{f\}}(x^{(i)}) - f_S(x^{(i)})]$$

where:

- F is the full set of features used in the model, i.e., $F = \{1, 2, \dots, 43\}$,
- $S \subseteq F \setminus \{f\}$ represents a subset of all features excluding feature f ,
- $f_{S \cup \{f\}}(x^{(i)})$ is the model's predicted probability for participant i when feature f is included,
- $f_S(x^{(i)})$ is the prediction when feature f is withheld.

Each feature f corresponds to a specific biking-related or demographic attribute such as weekly cycling frequency, stress level, access to cycling infrastructure, or encoded variables for gender and race. For example:

- $f = 12$: weekly cycling frequency,
- $f = 7$: distance biked per week,
- $f = 4$: stress level before biking,
- $f = 2$: encoded gender,
- $f = 39$: perceived cycling safety.

These SHAP values allow us to decompose the model output $f(x^{(i)})$ into a sum of contributions from each feature:

$$f(x^{(i)}) = \phi_0 + \sum_{f=1}^{43} \phi_f^{(i)}$$

where ϕ_0 is the model's expected output (the base value), and each $\phi_f^{(i)}$ measures how much feature f shifts the prediction for participant i above or below this base value.

In this study, SHAP values were computed using TreeExplainer from the SHAP library, which is optimized for Random Forest models. This approach ensures that explanations are consistent with the ensemble decision paths and computationally efficient for the full dataset.

Global Feature Importance Ranking

To provide a global interpretation, SHAP summary plots were generated using two formats:

- **Bar Plot:** Aggregated mean absolute SHAP values across all samples were plotted to rank features based on their overall contribution to the model predictions.
- **Beeswarm (Dot) Plot:** Individual SHAP values for each sample were visualized, showing the distribution and direction (positive or negative) of feature contributions.

For our analysis, the top eight most influential features were visualized to highlight which biking-related behaviors most strongly predict positive mental health outcomes.

6.4 Data Analysis

This section explores the demographic composition of the participants and the statistical analyses performed to extract meaningful findings from the data.

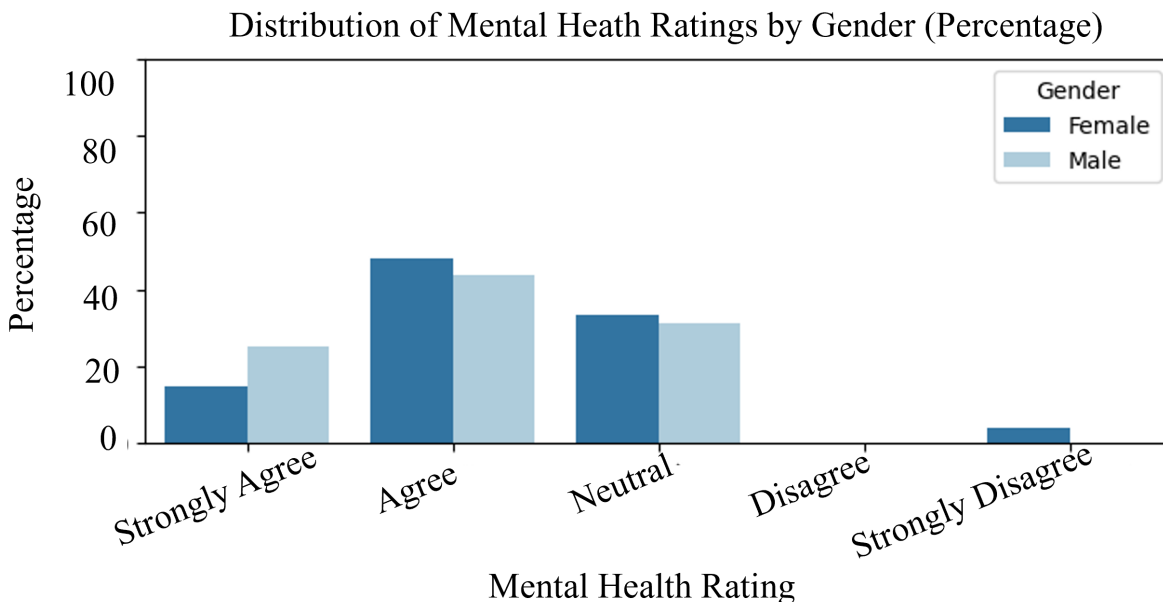


Figure 6.2: Gender Differences in Self-Reported Mental Health Ratings. *Note: The sum of each individual color is 100%. For example, the sum of all dark blue color bars is 100%.*

6.4.1 Participants Demographics

The survey had a total of 138 participants, with a completion rate of 72.46 %. The demographic breakdown included a balanced representation across different age groups, genders, and educational backgrounds. Notably, the percentage of female and male participants were 60 % and 40 % , respectively. The age distribution showed a higher representation of individuals between 41-50 years (28.75 %) and 61-70 years (26.25 %). Regarding educational attainment, the survey predominantly included participants with a Bachelor's degree (36.25 %) and those with a Graduate/Professional degree (22.5 %). The survey looked at what transportation options participants preferred in different situations. Driving was the top choice in all cases, with 62.07 % to 72.41 % of people choosing it. In contrast, fewer people preferred cycling, and this varied based on the type of bike infrastructure available. For example, 11.49 % chose cycling for short trips of 2 miles, but this number decreased for longer distances. Other options like public transit and carpooling were chosen even less frequently

than driving and cycling.

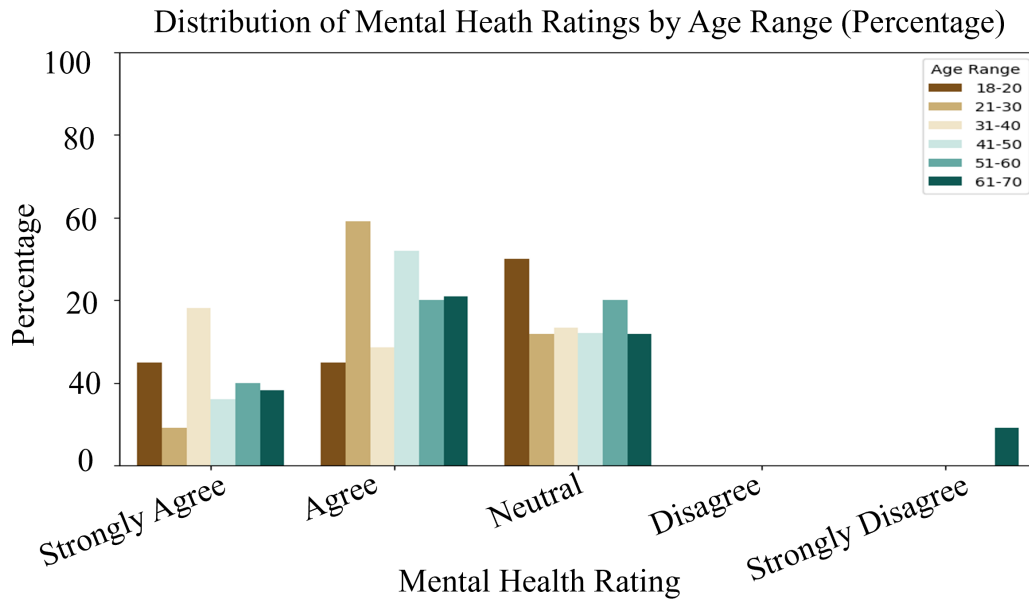


Figure 6.3: Distribution of Mental Health Ratings by Age Range. *Note: The sum of each individual color is 100%. For example, the sum of all brown color bars is 100%.*

6.4.2 Impact of cycling on Various Demographic Groups and Mental Health

Participants were asked about how they use and feel about different types of cycle lanes. The results showed that people preferred protected bike lanes over shared ones, highlighting their concern for safety. Designated bike lanes and two-way protected bike lanes were also used, but to a lesser extent. However, participants identified several problems, such as traffic congestion and safety concerns, as major obstacles to cycling.

Poor road conditions and the lack of bike lanes also discouraged many from cycling. The survey gathered feedback on the safety and condition of bike paths. Most people felt either neutral (61.76 %) or unsafe (20.59 %) when cycling on main roads, showing that safety is a big concern. Interestingly, 5 % of participants had close calls while cycling in the past year,

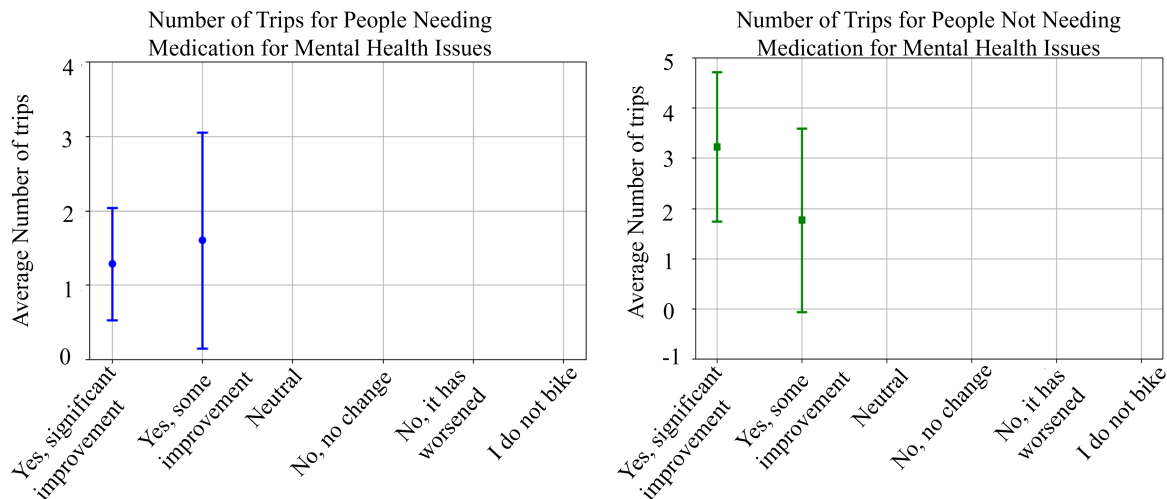


Figure 6.4: Comparison of cycling Frequency and Mental Health Ratings Based on Medication Needs

and 42.5 % did not bike at all. Regarding the maintenance of bike paths, most rated it as neutral (51.25 %) or well-maintained (27.5 %). Figure 3 visually represents the mental health ratings of male and female respondents. The x-axis categorizes the mental health ratings into five groups: Strongly Agree, Agree, Neutral, Disagree, and Strongly Disagree. These ratings, referred to as mental health ratings, were derived from a five-point Likert scale. Respondents were asked whether they agreed with the statement that cycling contributes positively to their mental health. The scale ranged from Strongly Agree (5), Agree (4), Neutral (3), Disagree (2), to Strongly Disagree (1). The y-axis shows the percentage of respondents in each category. The chart reveals that a higher percentage of females (represented by dark blue bars) strongly agree and agree with positive mental health ratings compared to males (represented by light blue bars). Specifically, nearly 50 % of females agree, while about 45 % of males fall into the same category. A notable portion of both genders feel neutral about their mental health, with females slightly higher than males. Few respondents, regardless of gender, disagree or strongly disagree with positive mental health ratings, with a slightly higher percentage of females strongly disagreeing compared to males.

Figure 4 summarizes the two error bar charts. These charts show how often people bike

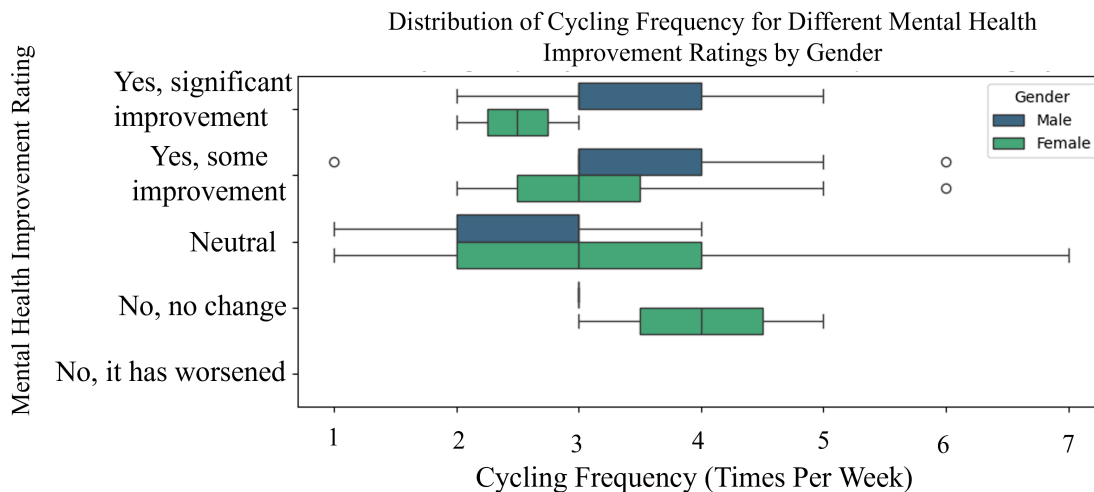


Figure 6.5: Distribution of Cycling Frequency for Different Mental Health Improvement Ratings by Gender

each week and their mental health ratings, comparing those who need medication with those who do not. In the left chart (blue), people who need medication have low cycling frequency when they see significant mental health improvement, and higher variability in cycling for those with neutral ratings. Very few people report no change or worsened mental health. The right chart (green) shows people who do not need medication generally bike moderately, with some variability, regardless of their mental health ratings.

Figure 5 shows how often men and women bike each week and their self-reported mental health improvements. The X-axis represents the number of cycling sessions per week, while the Y-axis shows the different mental health improvement ratings, ranging from "Yes, significant improvement" to "No, it has worsened." The points outside the whiskers are considered outliers and represent data points that fall significantly outside the typical range. The data is divided by gender, with male respondents in blue and female respondents in green. The box plots reveal that both men and women who report significant mental health improvements typically bike 3 to 5 times a week. Those who see some improvement usually bike around 2 to 4 times a week. Neutral ratings show a wider range, with both genders cycling from 1 to 4 times a week. Respondents who report no change or worsened mental health

have lower cycling frequencies, around 1 to 3 times a week. The graph indicates that higher cycling frequency is generally associated with better mental health improvement ratings for both men and women.

6.5 Result and Discussion

6.5.1 Result

The analysis of the model results covers findings from PCA, the Random Forest model, and SHAP value analysis.

In this research, Principal Component Analysis (PCA) was used to make our high-dimensional data easier to understand by focusing on a few main components that capture the mental health benefits of biking. PCA is a helpful tool that reduces data complexity, allowing us to concentrate on the most important factors while removing any unnecessary details. This approach is crucial in our study because it allows us to clearly identify the key areas where cycling has the greatest positive impact on mental health.

The PCA results in figure 6, reveal that five main components together explain all of the data's variation. This means these components capture the essential patterns in our dataset. The first component, 'Reduction in Mental Health Symptoms,' is the most influential, accounting for 60% of the variance. This suggests that a large part of biking's mental health benefits is related to reducing symptoms of stress and anxiety. The second component, 'Improvement in Sleep Patterns,' explains another 17% of the variance, showing biking's positive effect on sleep quality. The remaining components, 'Stress and Anxiety Management' (12%), 'Group cycling for Social Support' (7%), and 'Overall Quality of Life Enhancement' (4%), highlight other mental health benefits of biking.

In Figure 6, cumulative explained variance is plotted against the number of components

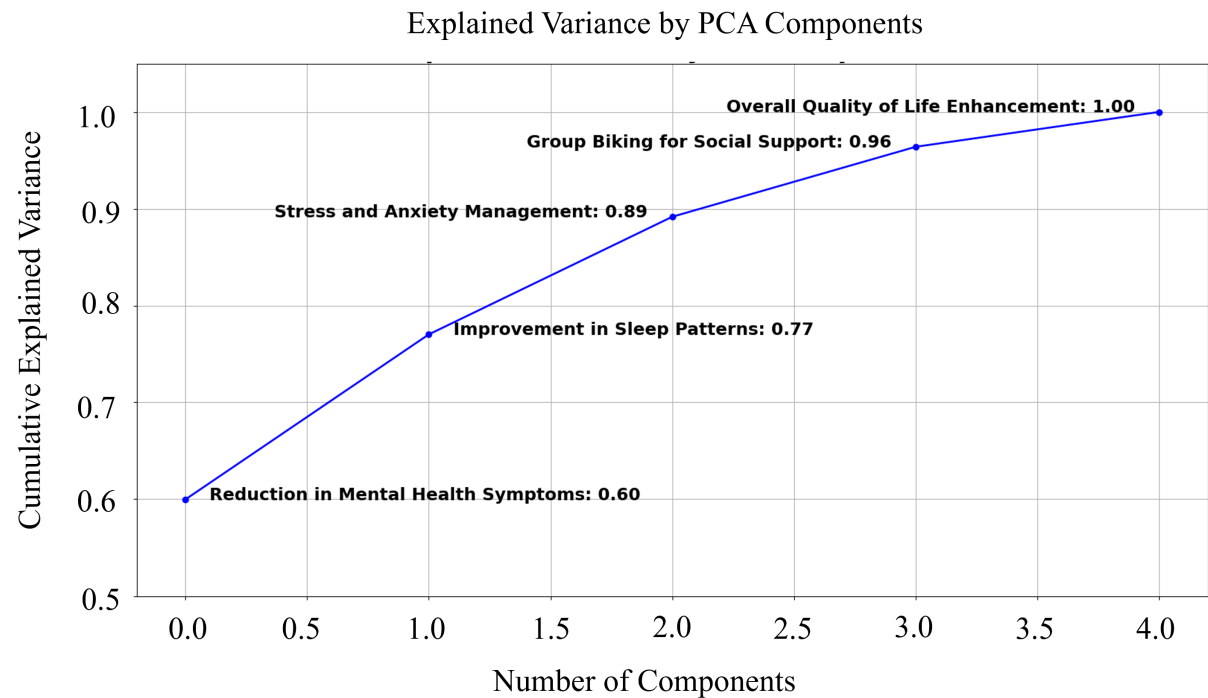


Figure 6.6: Cumulative Explained Variance by PCA Components

to illustrate each component’s contribution to the overall pattern. Each labeled point on the curve represents a component, with its position showing how much it contributes. For instance, the first component has a steep increase, capturing most of the variance, while the subsequent components add smaller but still meaningful contributions. This visual confirms that the primary mental health benefits are captured within the first few components, highlighting PCA’s effectiveness in isolating the most impactful aspects of the data.

Class	Precision	Recall	F1-Score
Negative Impact	0.73	0.62	0.67
Positive Impact	0.88	0.92	0.90
Accuracy		0.84	
Macro Avg	0.80	0.77	0.78
Weighted Avg	0.84	0.84	0.84

Table 6.2: Classification Report of Random Forest Model

The results of Random Forest model play a significant role in understanding the relationship between cycling and mental health benefits. With an accuracy of 84%, this model effectively

differentiates between individuals who experience a 'Positive Impact' on mental health from cycling and those who report 'Less/No Impact'. As shown in the confusion matrix in figure 7, the model accurately identifies 92.11% of individuals in the 'Positive Impact' group, indicating strong performance in classifying those who derive mental health benefits from biking. However, the model is somewhat less effective at identifying cases with 'Less/No Impact,' with an accuracy of 61.54% in this category. This variation is also reflected in the classification report, where high precision (0.88) and recall (0.92) were obtained for the 'Positive Impact' group, resulting in a high F1-score of 0.90. These metrics underscore the model's reliability in recognizing individuals who benefit mentally from biking, while also highlighting areas for improvement in identifying those who don't report the same level of impact.

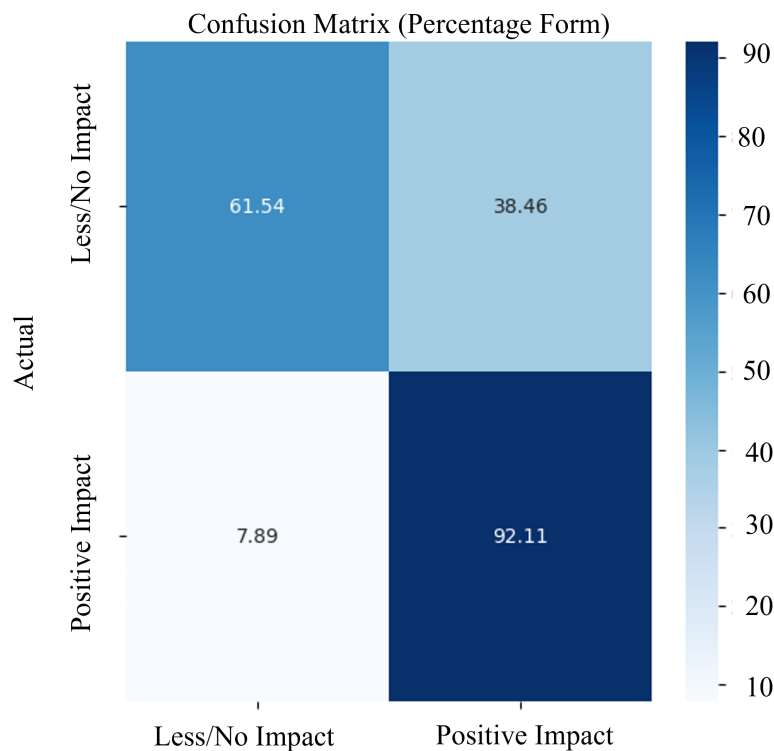


Figure 6.7: Confusion Matrix (Percentage Form) for Random Forest Model

This classification model provides a foundation for identifying specific cycling behaviors or personal characteristics that are associated with positive mental health outcomes. By

accurately categorizing individuals based on their mental health response to biking, the model allows us to pinpoint the key factors or patterns that correlate with enhanced mental well-being. For instance, it may help reveal whether certain frequencies, durations, or types of cycling are more likely to yield mental health benefits.

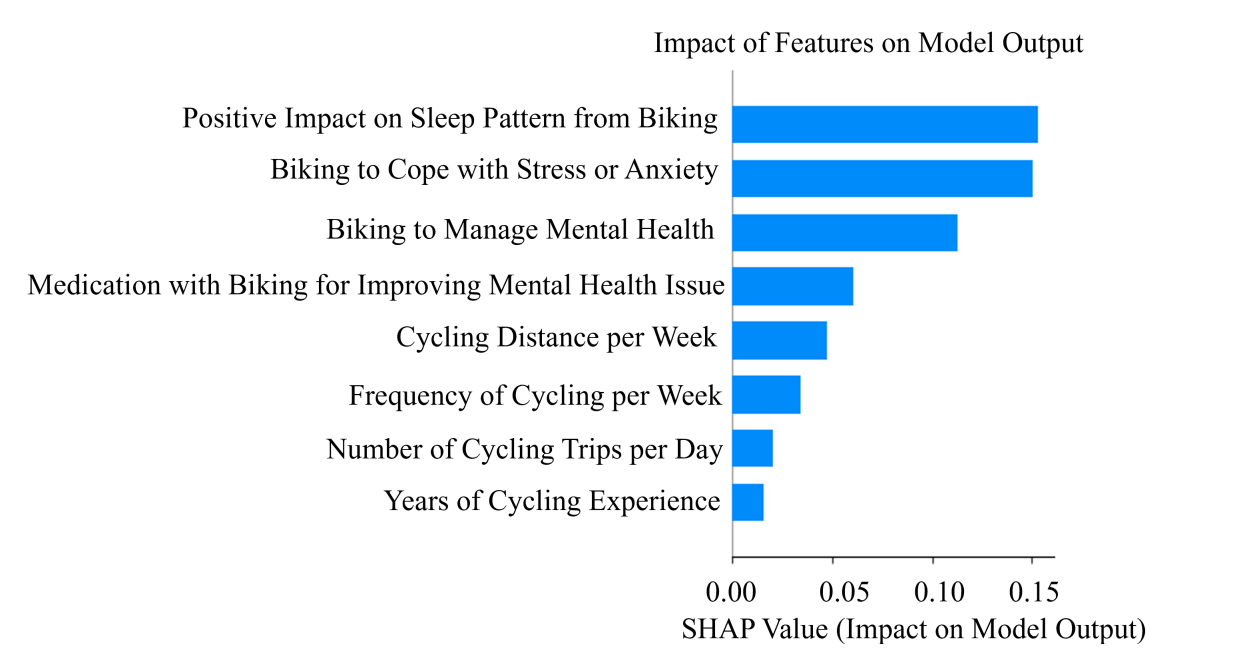


Figure 6.8: Influential Factors in Cycling and Mental Health relationship with their SHAP values

The SHAP analysis shows that various factors might have a positive impact on mental health through cycling. Figure 6.8, a bar plot, highlights the biking-related features with the most significant positive impacts on mental health outcomes, as determined by SHAP values. Each horizontal bar represents a feature, with the length of the bar indicating the magnitude of its effect on the model’s predictions. The feature with the strongest impact is Positive Impact on Sleep Pattern from Biking, indicating that sleep improvements are closely associated with mental health benefits derived from biking. This is good for mental health because it is a physical activity and physical activities release endorphins, reduce stress hormones, and give a sense of accomplishment (Al Sudani 2015). Cycling to Cope with Stress or Anxiety and cycling to Manage Mental Health follow, suggesting that using

cycling as a tool for managing stress and maintaining mental health has substantial positive effects. Features such as Cycling Distance per Week and Years of Cycling Experience have smaller, yet still relevant, impacts, reflecting that these factors contribute to mental health but to a lesser degree.

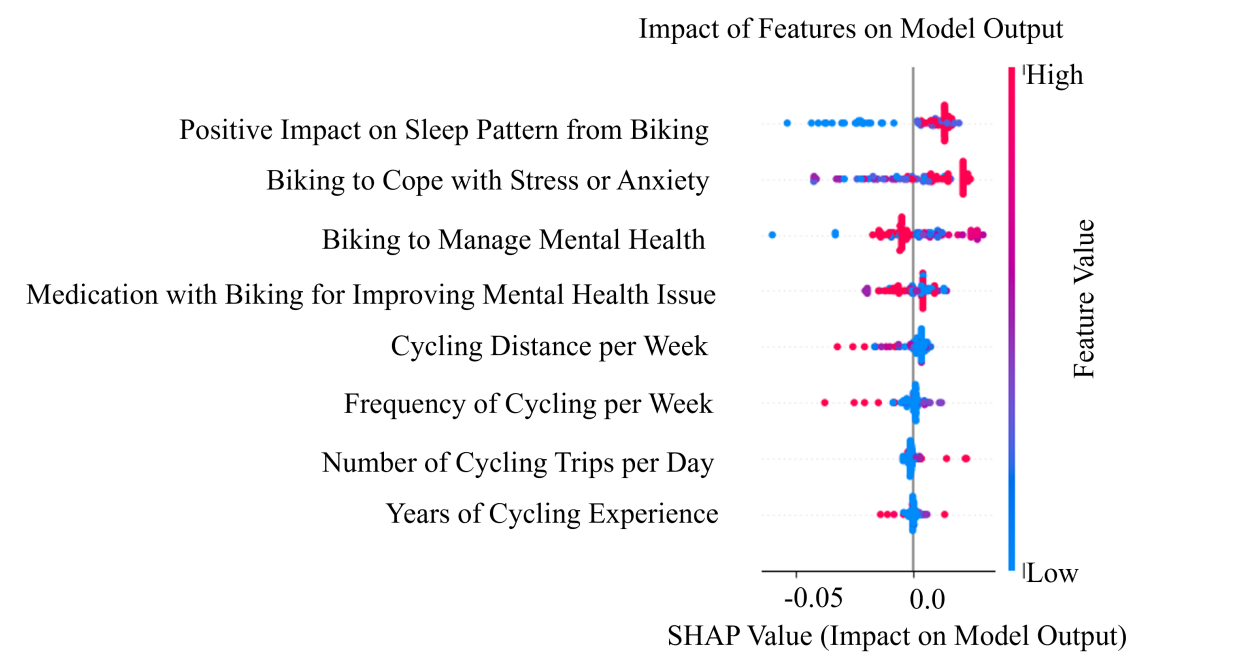


Figure 6.9: SHAP Value Summary Plot of Feature Importance

Figure 6.9, a SHAP summary plot, provides a detailed view of how each feature influences mental health outcomes on an individual basis. Each dot represents an individual observation, with the color of the dot reflecting the feature value (blue for lower values, red for higher values). For example, Positive Impact on Sleep Pattern from cycling consistently shows positive SHAP values for higher feature values (red dots to the right), reinforcing its strong association with improved mental health outcomes. Similarly, cycling to Cope with Stress or Anxiety has higher values clustering on the right, suggesting that frequent use of cycling to manage stress generally aligns with better mental health. Other features, like Cycling Distance per Week and Years of Cycling Experience, exhibit more variability, indicating that these aspects of cycling affect mental health in nuanced ways depending on the individual.

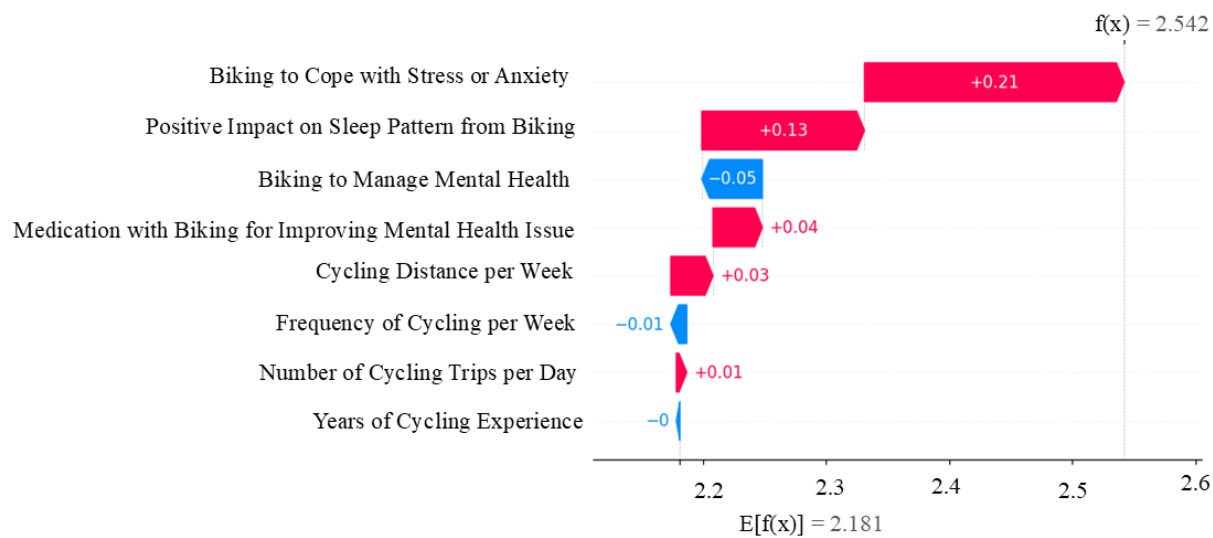


Figure 6.10: SHAP Waterfall Plot of Feature Contributions to Predicted Mental Health Impact of Biking

Additionally, Figure 10 presents a SHAP waterfall plot, which illustrates how individual features contribute to the predicted mental health outcome for a specific person. In this plot, the base value represents the average model prediction across the sample, while each bar shows how much a particular feature increases or decreases the predicted outcome for the selected individual. Features with positive SHAP values (red bars) increase the likelihood of experiencing positive mental health impacts from biking, whereas features with negative SHAP values (blue bars) reduce it. For example, 'cycling to Cope with Stress or Anxiety' and 'Positive Impact on Sleep Pattern from Biking' are the most influential positive contributors, significantly raising the individual's predicted mental health score. On the other hand, 'cycling to Manage Mental Health' and 'Frequency of Cycling per Week' show slight negative contribution and suggest these factors may not always have a positive effect for the selected individual.

6.5.2 Discussion

Our study confirms and builds on existing literature about the mental health benefits of physical activity, particularly biking, by applying advanced analysis methods like PCA, Random Forest modeling, and SHAP analysis. Previous research highlights that physical activity can reduce stress, anxiety, and depressive symptoms (Kandola, Ashdown-Franks, et al. 2019). Our study supports these findings but goes further by identifying specific benefits of biking, such as stress reduction, sleep improvement, and overall life quality enhancement, using advanced statistical and machine learning methods to provide robust evidence for these relationships. Using PCA, key components such as Reduction in Mental Health Symptoms and Improvement in Sleep Patterns were isolated, which explain a large part of the mental health benefits. These findings corroborate studies such as that by Deslandes et al. (Deslandes et al. 2009), which show significant positive effects of aerobic exercise, including biking, on mood regulation and anxiety reduction. Additionally, improved sleep quality has been a well-documented outcome of regular cycling, as highlighted by Youngstedt (Youngstedt 2005), supporting the consistency of our findings with established research.

The Random Forest model pinpointed cycling behaviors such as frequency and purpose for stress management that most effectively predict mental health improvements. This aligns with studies like Ma et al. (Martin et al. 2014), which emphasize the role of active commuting in lowering psychological distress, and Glanzener et al. (Glazener et al. 2021), whose systematic review confirmed that cycling for transport is positively associated with improved mental health outcomes, including reductions in depression and stress levels. Our findings further extend this evidence by using predictive modeling to highlight behavioral patterns most associated with mental health benefits. SHAP analysis helped identify sleep improvement and stress reduction as the most important factors, supporting broader evidence that regular aerobic exercise improves mental well-being by enhancing cognitive function and reducing depressive symptoms (Kandola, Ashdown-Franks, et al. 2019). Together, our study offers a

more targeted view of biking's mental health benefits and provides a valuable foundation for mental health interventions that incorporate biking.

6.6 Conclusion and Recommendations

This study contributes to a growing body of research on active transportation and public health by demonstrating how cycling supports mental well-being in rural American communities, where residents often face unique infrastructure and accessibility challenges. Using data from 14 rural college towns, PCA, Random Forest, and SHAP analysis were applied to identify the key behavioral and demographic factors that shape individuals' mental health outcomes from cycling. Key findings from this study include:

- Individuals who cycle frequently, have years of cycling experience, or actively use cycling as a means of managing mental health report the most positive outcomes.
- Age and experience further amplify these effects, with older and more experienced cyclists benefiting notably in terms of stress reduction and sleep improvement.
- cyclingproves to be an effective strategy for reducing stress and anxiety in rural areas where mental health services are often limited.
- Principal Component Analysis (PCA) highlighted stress reduction and sleep improvement as the dominant mental health benefits of biking.
- The Random Forest model achieved a high predictive accuracy of 84%. It identified specific cyclingbehaviors as important predictors of positive mental health outcomes.
- SHAP analysis revealed that sleep improvement and stress management are the most influential individual factors contributing to mental health benefits from biking.

To effectively translate the findings of this study into practice, several targeted programs can serve as models for promoting mental well-being through cycling in rural areas. For example, the Madison BCycle Community Pass Program (Madison, WI) enables public library cardholders to borrow e-bike access passes that is valid for up to a week, from any library branch, granting free access to over 350 electric bikes without requiring credit cards or membership fees (Madison Public Library [2022](#)). Similarly, the Colorado E-Bike Rebate Program provides income-based rebates to make e-bike purchases more affordable, helping low to moderate income residents adopt active transportation alternatives and reduce dependence on personal vehicles (Colorado Energy Office [2024](#)). These initiatives exemplify how subsidized and community-based cycling programs can increase access to cycling, particularly in underserved rural and semi-rural areas, thereby supporting both mobility and mental well-being (Virginia Tech Transportation Services [2024](#)).

Despite the strengths of our analysis, this study has some limitations. First, the cross-sectional data prevents us from making conclusions about long term effects or causality. Second, our sample may not fully represent all rural populations, which could limit generalizability. Lastly, only Random Forest model was used , and other machine learning methods could strengthen the result.

Future research can include a more diverse and representative sample of rural populations to improve generalizability. Longitudinal studies are recommended to better understand how regular cycling affects mental health over time. Exploring additional machine learning techniques, such as deep learning and hybrid models could further enhance predictive accuracy. These methods may also help uncover deeper behavioral patterns related to cycling and mental health.

Chapter 7

Conclusions

7.0.1 Research Summary

This research presents a comprehensive, multi-faceted investigation into the travel behavior of post-secondary students. It examines how students make interconnected travel decisions covering activity type selection, sequential departure time choices, and mode preferences while incorporating considerations of mental well-being and geographic context. The work employs econometric approaches with modern deep learning models, producing insights that are both theoretically grounded and computationally advanced.

Objective 1 - Multi-destination trip-Based Activity Type Choice

The first study analyzed multi-destination trips in an urban university setting using dynamic discrete choice models. It examined how one activity choice shapes subsequent activities within the same multi-destination trips, highlighting the structured interplay between fixed commitments (e.g., class schedules) and flexible engagements (e.g., leisure or social activities). The results underscored the importance of capturing sequence effects for more accurate modeling of student travel behavior, with implications for campus transit planning and urban mobility services.

Objective 2 - Sequential Departure Time Choice

The second study focused on the temporal dimension of student travel, modeling departure time as a sequential decision within two-destination trips. It applied Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM) networks, and Transformer architectures to capture the nonlinear and context-dependent relationships between trips. By incorporating socio-demographics, trip characteristics, and other factors, the study demonstrated that deep learning methods can improve predictive performance while retaining interpretability regarding the factors that influence departure time flexibility.

Objective 3 - Rural Student Mobility

The third study investigated mode choice behavior in rural college towns through a joint revealed–stated preference survey. Combined RP-SP MNL and mixed logit models were used to estimate current preferences and simulate responses to hypothetical infrastructure improvements. While the findings highlighted the constraints of limited transit services, they also revealed that targeted investments, e.g., enhanced cycling infrastructure, could shift mode shares toward more sustainable options.

Objective 4 - Mental Health and Active Transportation

The fourth study bridged transportation planning and public health by examining the relationship between cycling and mental well-being among post-secondary students. Using a Random Forest classifier with class balancing techniques and SHAP value analysis, the study identified conditions under which cycling is most strongly associated with improved mental health. The findings positioned cycling as a potential mental health intervention within university contexts.

Across the four studies, a number of important factors emerge that shape post-secondary

students' travel-related decisions. In the activity type choice analysis, travel related variables such as time, distance, and cost consistently reduced the likelihood of selecting additional activities, while socio-demographic characteristics (e.g., full versus part-time status, employment load, gender, age, and international student status) and household context (e.g., vehicle ownership, children, household size) further conditioned choices; institutional setting also played a role, with urban campuses encouraging broader activity participation compared to suburban ones. These factors directly influenced students' activity-type choices within multi-destination trips, thereby affecting subsequent destinations, timing, and mode. In the departure time modeling, mode-specific attributes such as transit distance, driving time, and Park & Ride time significantly affected trip timing, alongside socio-demographic influences including international student and off-campus tenant status, both of which were linked to earlier departures; critically, the Trip 1 departure time strongly predicted Trip 2 timing, underscoring temporal dependence across the day. These factors shaped departure time decisions for consecutive trips. In the mode choice study focused on rural college towns, the availability of safe and high-quality cycling infrastructure emerged as the most important factor influencing travel behavior. Simply providing painted or shared lanes did not substantially motivate cycling, as these still exposed riders to high crash risks and perceptions of danger. By contrast, protected or physically separated facilities—including two-way protected bike lanes, buffered lanes that reduce dooring hazards, and protected intersections that improve visibility—proved far more effective in encouraging students to cycle. These designs not only enhance safety by reducing conflicts with motor vehicles but also make cycling a more viable and attractive option. Evidence from U.S. cities reinforces this finding, showing that protected lanes are associated with significant increases in ridership, with growth ranging from 21% to 171%. Together, these results demonstrate that it is not the mere presence of infrastructure, but rather the quality and safety of that infrastructure, that motivates cycling in rural college towns. These factors influenced students' mode choice decisions, particularly the decision to cycle versus driving or using transit. Finally, the men-

tal health analysis revealed that cycling frequency and experience were strongly associated with positive outcomes, with age and cycling experience amplifying benefits; the dominant reported outcomes were stress reduction and sleep improvement, reflecting how students use cycling as a coping or health-support strategy. These factors influenced students' decisions to engage in cycling for transport or well-being, linking travel behavior to mental health. Collectively, these findings highlight that post-secondary student decisions around activity type, departure time, mode choice, and mental health are shaped by a complex interplay of trip attributes, socio-demographic profiles, household circumstances, institutional settings, infrastructure availability, and psychosocial outcomes.

7.0.2 Research Contributions

This research brings together four interconnected studies to construct a comprehensive understanding of post-secondary students' travel behavior, addressing it from the perspectives of activity participation, departure time choice, mode selection under different infrastructure scenarios, and the health implications of active travel. The studies collectively fill a significant gap in the transportation literature by recognizing post-secondary students' as a distinct traveler group whose daily mobility decisions are shaped by unique temporal patterns, institutional contexts, resource constraints, lifestyle priorities, and other factors. Through a series of advanced econometric frameworks, deep learning models, joint revealed-stated preference methods, and interpretable machine learning approaches, the research highlights how employing a range of methodological approaches helps capture the nuanced aspects of travel behavior.

The first major insight comes from the analysis of multi-destination trip-based activity type choices. This perspective emphasizes the need to view post-secondary student travel as a linked set of interdependent decisions rather than a series of isolated trips. Such a broad viewpoint captures how early activity participation shapes the feasibility and desirability

of subsequent trips. The findings challenge the limitations of traditional static models and illustrate the value of dynamic discrete choice formulations in capturing the interconnectedness of daily travel patterns, particularly for students whose schedules often involve academic commitments, part-time employment, social engagements, and other activities with varying degrees of flexibility.

Building on this behavioral foundation, the sequential departure time modeling extends the concept of interdependence to temporal decision-making. By situating departure time choices within the broader structure of a two-destination trip, the research reveals the cascading effects of initial departure decisions on subsequent trips. The application of advanced sequence-learning architectures, namely, GRU, LSTM, and Transformer models, proves instrumental in identifying temporal patterns that are often invisible to traditional models. These models successfully capture the influence of factors such as socio-demographic characteristics, trip purpose, prior departure times, etc., enabling a richer predictive capacity and advancing the integration of deep learning into behavioral travel modeling. These models balance interpretability with predictive strength, offering a pathway for future temporal choice modeling.

The third component of the research shifts the focus from temporal sequencing to the spatial and infrastructural enablers of mode choice, particularly in rural college-town contexts where transportation options are often limited. The joint revealed-stated preference survey design and subsequent mixed logit modeling provide compelling evidence that improvements in active transportation infrastructure, e.g., two-way protected bicycle lane can significantly alter mode choice probabilities. The findings reveal that, for students, the perception of safety, trip convenience, low cost, and integration with existing infrastructure all play decisive roles in shifting away from car dependency. This has important implications for rural mobility planning, as it highlights cost-effective strategies that can increase mode diversity even in areas with low baseline transit service. The fourth and final study brings a public health perspective into the mobility discussion by examining the mental health benefits of active travel,

with a focus on cycling. Using machine learning models such as RF and SHAP-based interpretability techniques, the study identifies strong associations between cycling frequency and indicators of improved mental well-being. This link underscores the dual societal benefits of promoting active transportation, not only as a means to reduce congestion, emissions, and costs but also as a policy lever to support mental health and overall quality of life among post-secondary students. In bridging transportation and health domains, the research encourages interdisciplinary strategies for mobility planning, where student well-being is recognized as both a goal and an outcome of transportation investment. Taken together, the four studies converge on a central insight: post-secondary student travel behavior is a unique, dynamic, and context-sensitive process that cannot be fully understood through any single methodological lens or disciplinary framework. The investigation of activity type choice, departure time choice, infrastructure sensitivity, and health outcome produces a more holistic picture of how and why students travel the way they do. The broader societal benefits that arise from sustainable and health-promoting travel can only be unlocked by understanding how and why post-secondary students travel as they do, and then using that information to change transportation networks. Moreover, by positioning post-secondary students as a central focus in mobility research and employing innovative analytical techniques, this research not only advances theoretical and methodological understanding but also delivers actionable insights for policymakers, planners, campus administrators, and other stakeholders. The overall objective is to design transportation systems that are more user-friendly, equitable, sustainable, and supportive of student life.

7.0.3 Limitations and Future Work

While this research provides valuable insights into post-secondary students' travel behavior, several limitations remain across the four studies that should be acknowledged when interpreting the results.

First, dataset constraints represent a central limitation. The activity type choice study was restricted to students in the Greater Toronto and Hamilton Area (GTHA), while the mode choice and mental health studies focused on rural college towns in Virginia. These region-specific datasets limit the broader generalizability of the findings. Moreover, some datasets were relatively small, such as the 2,408-student sample used for departure time modeling and the summer-only data collection in the rural cycling study, which underrepresents the full student population present during fall and spring semesters. Future research should prioritize collecting larger and more diverse samples across multiple institutions, geographic contexts, and academic terms to ensure robust representation of different socio-demographic and infrastructural conditions.

Second, survey design limitations are evident. Much of the data relied on cross-sectional surveys, which capture behavior at a single point in time and cannot fully account for temporal variation or adaptation to new infrastructure and policies. In the cycling and mental health analysis, the cross-sectional design further limited the ability to make causal claims about long-term effects. Future studies should adopt repeated cross-sectional or panel survey designs, enabling the estimation of temporal dynamics, learning effects, and long-term behavioral changes. These approaches would also allow researchers to better isolate causal mechanisms, particularly in evaluating the health benefits of cycling.

Third, there are notable modeling scope limitations. The studies addressed activity type choice, departure time choice, and mode choice as separate processes, but in reality these decisions are highly interconnected: when students travel, what activities they pursue, and what modes they use are all interdependent. While the models applied here (MNL, NL, Mixed Logit, and Random Forest) captured important behavioral heterogeneity within each dimension, they did not jointly represent these choices in an integrated framework. Future research should develop unified modeling approaches that simultaneously account for activity type, timing, and mode choice, aligning more closely with the reality of interconnected multi-destination trip-based travel behavior. This would also strengthen the predictive and policy

relevance of the models.

Finally, methodological opportunities remain. While advanced models such as GRU, LSTM, and Transformer architectures were implemented in this research, the analyses relied on relatively standard hyperparameter settings (e.g., grid search). Future work should incorporate more systematic approaches to optimization, such as Bayesian search techniques (e.g., Optuna), to fine-tune parameters more efficiently and improve both predictive accuracy and model stability. Beyond optimization, extending hybrid frameworks that combine econometric interpretability with machine learning flexibility could further deepen insights into behavioral drivers, particularly when paired with attitudinal and perceptual variables, such as safety concerns, schedule flexibility, or sustainability preferences.

To conclude, this research establishes an important foundation for understanding post-secondary student travel behavior in both urban and rural contexts, but future research must move toward larger, more representative datasets, longitudinal survey designs, integrated modeling frameworks, and advanced methodological tools. These enhancements will enable richer behavioral realism, improved generalizability, and stronger evidence to inform transportation planning and policies that support student mobility and well-being.

7.1 Statement of AI Use

Artificial intelligence tools were used exclusively for editing purposes in this research. Specifically, Grammarly was employed to check grammar, spelling, and writing clarity. No AI tools were used for data analysis, modeling, or generating research results. All substantive content, analysis, and interpretation are the original work of the author.

Bibliography

- Afandizadeh Zargari, S., & Safari, F. (2020). Using clustering methods in multinomial logit model for departure time choice. *Journal of Advanced Transportation*, 2020(1), 7382569.
- Al Sudani, A. A. D. (2015). Physical activity, mental health and positive emotions. *Lase J Sports Sci*, 6, 94–100.
- Aldred, R., Croft, J., & Goodman, A. (2019). Impacts of an active travel intervention with a cycling focus in a suburban context: One-year findings from an evaluation of london’s in-progress mini-hollands programme. *Transportation research part A: policy and practice*, 123, 147–169.
- Allen, J., & Farber, S. (2018). How time-use and transportation barriers limit on-campus participation of university students. *Travel behaviour and society*, 13, 174–182.
- Ameli, M., Faradonbeh, M. S. S., Lebacque, J.-P., Abouee-Mehrizi, H., & Leclercq, L. (2022). Departure time choice models in urban transportation systems based on mean field games. *Transportation Science*, 56(6), 1483–1504.
- Aptech Systems, Inc. (2024). *Gauss: Matrix programming language* [Accessed: 2025-08-27]. Version 24. Aptech Systems, Inc. <https://www.aptech.com>
- Arentze, T. A., & Timmermans, H. J. (2004). A learning-based transportation oriented simulation system. *Transportation Research Part B: Methodological*, 38(7), 613–633.
- Aultman-Hall, L., Onayev, A., Dowds, J., & LaMondia, J. J. (2023). Understanding the relationship between mode and destination choices for personal out-of-town travel. *Travel Behaviour and Society*, 31, 399–409.
- Avila-Palencia, I., De Nazelle, A., Cole-Hunter, T., Donaire-Gonzalez, D., Jerrett, M., Rodriguez, D. A., & Nieuwenhuijsen, M. J. (2017). The relationship between bicycle commuting and perceived stress: A cross-sectional study. *BMJ open*, 7(6), e013542.

- Avila-Palencia, I., Panis, L. I., Dons, E., Gaupp-Berghausen, M., Raser, E., Götschi, T., Gerike, R., Brand, C., De Nazelle, A., Orjuela, J. P., et al. (2018). The effects of transport mode use on self-perceived health, mental health, and social contact measures: A cross-sectional and longitudinal study. *Environment international*, 120, 199–206.
- Badiola, N., Raveau, S., & Galilea, P. (2019). Modelling preferences towards activities and their effect on departure time choices. *Transportation Research Part A: Policy and Practice*, 129, 39–51.
- Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv preprint arXiv:1803.01271*.
- Bajwa, S., Bekhor, S., Kuwahara, M., & Chung, E. (2008). Discrete choice modeling of combined mode and departure time. *Transportmetrica*, 4(2), 155–177.
- Balsas, C. J. (2003). Sustainable transportation planning on college campuses. *Transport Policy*, 10(1), 35–49.
- Banister, D. (2008). The sustainable mobility paradigm. *Transport Policy*, 15(2), 73–80. <https://doi.org/10.1016/j.tranpol.2007.10.005>
- Ben-Akiva, M. E. [Moshe E], & Bowman, J. L. (1998). Activity based travel demand model systems. In *Equilibrium and advanced transportation modelling* (pp. 27–46). Springer.
- Ben-Akiva, M. E. [Moshe E], & Lerman, S. R. (1985). *Discrete choice analysis: Theory and application to travel demand* (Vol. 9). MIT press.
- Ben-Akiva, M. E. [Moshe E.], & Lerman, S. R. (2014). *Discrete choice analysis: Theory and application to travel demand*. MIT Press.
- Ben-Elia, E., Bierlaire, M., & Ettema, D. (2010). A behavioural departure time choice model with latent arrival time preference and rewards for peak-hour avoidance. *European Transport Conference*.
- Berrie, L., Feng, Z., Rice, D., Clemens, T., Williamson, L., & Dibben, C. (2024). Does cycle commuting reduce the risk of mental ill-health? an instrumental variable analysis

- using distance to nearest cycle path. *International journal of epidemiology*, 53(1), dyad153.
- Bhat, C. R. (1998). Analysis of travel mode and departure time choice for urban shopping trips. *Transportation Research Part B: Methodological*, 32(6), 361–371. [https://doi.org/10.1016/S0191-2615\(98\)00018-4](https://doi.org/10.1016/S0191-2615(98)00018-4)
- Bhat, C. R., & Koppelman, F. S. (1999). Activity-based modeling of travel demand. In *Handbook of transportation science* (pp. 35–61). Springer.
- Bhat, C. R., & Steed, J. L. (2002). A continuous-time model of departure time choice for urban shopping trips. *Transportation Research Part B: Methodological*, 36(3), 207–224.
- Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, 13(1), 1063–1095.
- Blumenberg, E. (2014). Transportation survival strategies of the poor. *Journal of Planning Literature*, 29(2), 3–18. <https://doi.org/10.1177/0885412213504853>
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32.
- Bricka, S., Reuscher, T., Schroeder, P., Fisher, M., Beard, J., Sun, X. L., et al. (2024). *Summary of travel trends: 2022 national household travel survey* (tech. rep.). United States. Federal Highway Administration. Office of Policy and ...
- Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical methods*, 6(9), 2812–2831.
- Buehler, R., Teoman, D., & Shelton, B. (2021). Promoting bicycling in car-oriented cities: Lessons from washington, dc and frankfurt am main, germany. *Urban Science*, 5(3), 58.
- Cahuantzi, R., Chen, X., & Güttel, S. (2023). A comparison of lstm and gru networks for learning symbolic sequences. *Science and information conference*, 771–785.
- Castiglione, J., Bradley, M., & Gliebe, J. (2015). *Activity-based travel demand models: A primer*.

- Celis-Morales, C. A., Lyall, D. M., Welsh, P., Anderson, J., Steell, L., Guo, Y., Maldonado, R., Mackay, D. F., Pell, J. P., Sattar, N., et al. (2017). Association between active commuting and incident cardiovascular disease, cancer, and mortality: Prospective cohort study. *bmj*, 357.
- Chand, C. (n.d.). An evaluation of door zone collisions between bicycles and vehicles.
- Chekroud, S. R., Gueorguieva, R., Zheutlin, A. B., Paulus, M., Krumholz, H. M., Krystal, J. H., & Chekroud, A. M. (2018). Association between physical exercise and mental health in 1 · 2 million individuals in the usa between 2011 and 2015: A cross-sectional study. *The lancet psychiatry*, 5(9), 739–746.
- Cheng, H., & Yang, X. (2015). Random parameter nested logit model for combined departure time and route choice. *International Journal of Transportation Science and Technology*, 4(1), 93–105.
- Cheng, Q., Deng, W., & Ammar Raza, M. (2020). Analysis of the departure time choices of metro passengers during peak hours. *IET Intelligent Transport Systems*, 14(8), 866–872.
- Cheng, Y., Ye, X., & Fujiyama, T. (2020). Identifying crowding impact on departure time choice of commuters in urban rail transit. *Journal of advanced transportation*, 2020(1), 8850565.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.
- Chong, S., Poulos, R., Olivier, J., Watson, W. L., & Grzebieta, R. (2010). Relative injury severity among vulnerable non-motorised road users: Comparative analysis of injury arising from bicycle–motor vehicle and bicycle–pedestrian collisions. *Accident Analysis & Prevention*, 42(1), 290–296.
- Chughtai, J.-R., Haq, I. U., & Muneeb, M. (2022). An attention-based recurrent learning model for short-term travel time prediction. *Plos one*, 17(12), e0278064.

- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- Clark, B., Chatterjee, K., Martin, A., & Davis, A. (2020). How commuting affects subjective wellbeing. *Transportation*, 47(6), 2777–2805.
- Colorado Energy Office. (2024). Colorado e-bike rebate program [Accessed: 2025-07-25]. <https://energyoffice.colorado.gov/ebike-rebates>
- Creutzig, F., Jochem, P., Edelenbosch, O. Y., Mattauch, L., Vuuren, D. P. v., McCollum, D., & Minx, J. (2015). Transport: A roadblock to climate change mitigation? *Science*, 350(6263), 911–912.
- Daisy, N. S., Millward, H., & Liu, L. (2018). Trip chaining and tour mode choice of non-workers grouped by daily activity patterns. *Journal of Transport Geography*, 69, 150–162.
- Deavers, K. (1992). What is rural? *Policy Studies Journal*, 20(2).
- Deliali, K., Christofa, E., & Knodler Jr, M. (2021). The role of protected intersections in improving bicycle safety and driver right-turning behavior. *Accident Analysis & Prevention*, 159, 106295.
- Delmelle, E. M. [Elizabeth M.], & Delmelle, E. C. (2012). Exploring spatio-temporal commuting patterns in a university environment. *Transport Policy*, 21, 1–9. <https://doi.org/10.1016/j.tranpol.2011.12.007>
- Delmelle, E. M. [Eric M], & Delmelle, E. C. (2012). Exploring spatio-temporal commuting patterns in a university environment. *Transport Policy*, 21, 1–9.
- Deslandes, A., Moraes, H., Ferreira, C., Veiga, H., Silveira, H., Mouta, R., Pompeu, F. A., Coutinho, E. S. F., & Laks, J. (2009). Exercise and mental health: Many reasons to move. *Neuropsychobiology*, 59(4), 191–198.
- Dobbs, G. L. (2009). *Pedestrian and bicycle safety on a college campus: Crash and conflict analyses with recommended design alternatives for clemson university* [Master's thesis, Clemson University].

- Doherty, S. T., & Mohammadian, A. (2011). The validity of using activity type to structure tour-based scheduling models. *Transportation*, 38(1), 45–63.
- Eluru, N., Chakour, V., & El-Geneidy, A. M. (2012). Travel mode choice and transit route choice behavior in montreal: Insights from mcgill university members commute patterns. *Public Transport*, 4(2), 129–149.
- Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *nature*, 542(7639), 115–118.
- Fan, J. X., Wen, M., & Wan, N. (2017). Built environment and active commuting: Rural-urban differences in the us. *SSM-population health*, 3, 435–441.
- Farrington, J., & Farrington, C. (2005). Rural accessibility, social inclusion and social justice: Towards conceptualisation. *Journal of Transport Geography*, 13(1), 1–12. <https://doi.org/10.1016/j.jtrangeo.2004.10.002>
- Flessner, M. (2020). *Analysis of the traffic flows caused by the kingston campus of the university of rhode island* [Master’s thesis, University of Rhode Island].
- Garikapati, V. M., You, D., Pendyala, R. M., Vovsha, P. S., Livshits, V., & Jeon, K. (2014). Multiple discrete-continuous model of activity participation and time allocation for home-based work tours. *Transportation Research Record*, 2429(1), 90–98.
- Geng, K., Wang, Y., Cherchi, E., & Guarda, P. (2023). Commuter departure time choice behavior under congestion charge: Analysis based on cumulative prospect theory. *Transportation research part A: policy and practice*, 168, 103564.
- Géron, A. (2022). *Hands-on machine learning with scikit-learn, keras, and tensorflow*. ” O’Reilly Media, Inc.”
- Giles-Corti, B., Vernez-Moudon, A., Reis, R., Turrell, G., Dannenberg, A. L., Badland, H., Foster, S., Lowe, M., Sallis, J. F., Stevenson, M., et al. (2016). City planning and population health: A global challenge. *The lancet*, 388(10062), 2912–2924.

- Glaeser, E. L. (2011). *Triumph of the city: How our greatest invention makes us richer, smarter, greener, healthier, and happier*. Penguin Press.
- Glazener, A., Sanchez, K., Ramani, T., Zietsman, J., Nieuwenhuijsen, M. J., Mindell, J. S., Fox, M., & Khreis, H. (2021). Fourteen pathways between urban transportation and health: A conceptual model and literature review. *Journal of transport & health*, 21, 101070.
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.
- Greene, W. H., & Hensher, D. A. (2007). Heteroscedastic control for random coefficients and error components in mixed logit. *Transportation Research Part E: Logistics and Transportation Review*, 43(5), 610–623.
- Gu, Z., Wang, Y., Ma, W., & Liu, Z. (2024). A joint travel mode and departure time choice model in dynamic multimodal transportation networks based on deep reinforcement learning. *Multimodal Transportation*, 3(3), 100137.
- Habib, K. M. N. (2013). A joint discrete-continuous model considering budget constraint for the continuous part: Application in joint mode and departure time choice modelling. *Transportmetrica A: Transport Science*, 9(2), 149–177.
- Habib, K. M. N., Day, N., & Miller, E. J. (2009). An investigation of commuting trip timing and mode choice in the greater toronto area: Application of a joint discrete-continuous model. *Transportation Research Part A: Policy and Practice*, 43(7), 639–653.
- Hasnine, M. S., & Habib, K. N. (2020). Modelling the dynamics between tour-based mode choices and tour-timing choices in daily activity scheduling. *Transportation*, 47(5), 2635–2669.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). An introduction to statistical learning.
- Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: The state of practice. *Transportation*, 30(2), 133–176.

- Hesam Hafezi, M., Sultana Daisy, N., Millward, H., & Liu, L. (2022). Framework for development of the scheduler for activities, locations, and travel (salt) model. *Transport-metrica A: Transport Science*, 18(2), 248–280.
- Hess, S., & Daly, A. (2024). *Handbook of choice modelling*. Edward Elgar Publishing.
- Hess, S., & Train, K. (2017). Correlation and scale in mixed logit models. *Journal of choice modelling*, 23, 1–8.
- Hoang, T. L., Adnan, M., Vu, A. T., Hoang-Tung, N., Kochan, B., & Bellemans, T. (2022). Modeling and structuring of activity scheduling choices with consideration of intra-zonal tours: A case study of motorcycle-based cities. *Sustainability*, 14(10), 6367.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735–1780.
- Hossain, S., Hasnine, M. S., & Habib, K. N. (2021). A latent class joint mode and departure time choice model for the greater toronto and hamilton area. *Transportation*, 48, 1217–1239.
- Iamtrakul, P., & Raungratanaamporn, I.-s. (2015). Sustainable campus town development in suburban area of bangkok, thailand. *International Journal of Building, Urban, Interior and Landscape Technology (BUILT)*, 6, 39–52.
- Islam, R., Rafique, S., Bredikhina, O. A., & Jones, S. (2022). Getting around in a college town: A case study of transportation barriers faced by international students at ua [Accessed on 2025-08-20].
- Jin, W.-L. (2020). Stable day-to-day dynamics for departure time choice. *Transportation Science*, 54(1), 42–61.
- Jones, P. [Phil]. (2012). Sensory indiscipline and affect: A study of commuter cycling. *Social & Cultural Geography*, 13(6), 645–658.
- Jou, R.-C., Kitamura, R., Weng, M.-C., & Chen, C.-C. (2008). Dynamic commuter departure time choice under uncertainty. *Transportation Research Part A: Policy and Practice*, 42(5), 774–783.

- Kandola, A., Ashdown-Franks, G., Hendrikse, J., Sabiston, C. M., & Stubbs, B. (2019). Physical activity and depression: Towards understanding the antidepressant mechanisms of physical activity. *Neuroscience & Biobehavioral Reviews*, *107*, 525–539.
- Kandola, A., Hendrikse, J., Lucassen, P. J., & Yücel, M. (2016). Aerobic exercise as a tool to improve hippocampal plasticity and function in humans: Practical implications for mental health treatment. *Frontiers in human neuroscience*, *10*, 373.
- Kelchen, R. (2018). A look at college students' living arrangements [Accessed on 2025-08-20].
- Khan, N. A., Shahrier, H., & Habib, M. A. (2022). Validation of an activity-based travel demand modeling system. *Transportation Letters*, *14*(9), 1004–1018.
- Kiyohara, M., & Mondri, E. (2025). Coolight: Enhancing nighttime safety for urban student commuters [Preprint discussing design-informed solutions for improving nighttime walking safety among urban university students].
- Lawrence, B. M., Oxley, J. A., Logan, D. B., & Stevenson, M. R. (2018). Cyclist exposure to the risk of car door collisions in mixed function activity centers: A study in melbourne, australia. *Traffic injury prevention*, *19*(sup1), S164–S168.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.
- Liu, B., Wei, Y., Zhang, Y., & Yang, Q. (2017). Deep neural networks for high dimension, low sample size data. *IJCAI, 2017*, 2287–2293.
- Loa, P., Hossain, S., Liu, Y., & Nurul Habib, K. (2021). How have ride-sourcing users adapted to the first wave of the covid-19 pandemic? evidence from a survey-based study of the greater toronto area. *Transportation Letters*, *13*(5-6), 404–413.
- Loukaitou-Sideris, A., Medury, A., Fink, C., Grembek, O., Shafizadeh, K., Wong, N., & Orrick, P. (2014). Crashes on and near college campuses: A comparative analysis of pedestrian and bicyclist safety. *Journal of the American Planning Association*, *80*(3), 198–217.
- Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: Analysis and applications*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511753831>

- Lucas, K., & Jones, P. [Peter]. (2012). Social impacts and equity issues in transport: An introduction. *Journal of Transport Geography*, 21, 1–3.
- Lucas, K., Mattioli, G., Verlinghieri, E., & Guzman, A. (2016). Transport poverty and its adverse social consequences. *Proceedings of the institution of civil engineers-transport*, 169(6), 353–365.
- Lucas, K., & Musso, A. (2014). Policies for social inclusion in transportation: An introduction to the special issue. *Case Studies on Transport Policy*, 2(2), 37–40. <https://doi.org/10.1016/j.cstp.2014.06.002>
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.
- Luo, N., Nara, A., Khoo, H. L., & Chen, M. (2024). An integration modeling framework for individual-scale daily mobility estimation. *Travel Behaviour and Society*, 34, 100650.
- Ma, L., Ye, R., & Wang, H. (2021). Exploring the causal effects of bicycling for transportation on mental health. *Transportation research part D: transport and environment*, 93, 102773.
- Madison Public Library. (2022). Madison bicycle community pass program [Accessed: 2025-07-25]. <https://www.madisonpubliclibrary.org/resources/services/bicycle>
- Manaf, H. (2013). Barriers to participation in physical activity and exercise among middle-aged and elderly individuals. *Singapore Med J*, 54(10), 581–586.
- Martens, K. (2016). *Transport justice: Designing fair transportation systems*. Routledge.
- Martin, A., Goryakin, Y., & Suhrcke, M. (2014). Does active commuting improve psychological wellbeing? longitudinal evidence from eighteen waves of the british household panel survey. *Preventive medicine*, 69, 296–303.
- McFadden, D. (1974). The measurement of urban travel demand. *Journal of public economics*, 3(4), 303–328.
- McGuckin, N., & Murakami, E. (2005). *Trip-chaining trends in the united states: Understanding travel behavior for policy making* (tech. rep.). U.S. Department of Transportation,

- Federal Highway Administration. <https://nhts.ornl.gov/2001/articles/TripChaining.pdf>
- McNeil, N., Monsere, C. M., & Dill, J. (2015). Influence of bike lane buffer types on perceived comfort and safety of bicyclists and potential bicyclists. *Transportation research record*, 2520(1), 132–142.
- Medury, A., Grembek, O., Loukaitou-Sideris, A., & Shafizadeh, K. (2019). Investigating the underreporting of pedestrian and bicycle crashes in and around university campuses—a crowdsourcing approach. *Accident Analysis & Prevention*, 130, 99–107.
- Moeckel, R., Huang, W.-C., Ji, J., Llorca, C., Moreno, A. T., Staves, C., Zhang, Q., & Erhardt, G. D. (2024). The activity-based model abit: Modeling 24 hours, 7 days a week. *Transportation Research Procedia*, 78, 499–506.
- Monsere, C., Dill, J., McNeil, N., Clifton, K. J., Foster, N., Goddard, T., Berkow, M., Gilpin, J., Voros, K., van Hengel, D., et al. (2014). Lessons from the green lanes: Evaluating protected bike lanes in the us.
- Moodie, R., Stuckler, D., Monteiro, C., Sheron, N., Neal, B., Thamarangsi, T., Lincoln, P., & Casswell, S. (2013). Profits and pandemics: Prevention of harmful effects of tobacco, alcohol, and ultra-processed food and drink industries. *The lancet*, 381(9867), 670–679.
- Moody, J., Farr, E., Papagelis, M., & Keith, D. R. (2021). The value of car ownership and use in the united states. *Nature Sustainability*, 4(9), 769–774.
- Morres, I. D., Hatzigeorgiadis, A., Stathi, A., Comoutos, N., Arpin-Cribbie, C., Krommidas, C., & Theodorakis, Y. (2019). Aerobic exercise for adult patients with major depressive disorder in mental health services: A systematic review and meta-analysis. *Depression and anxiety*, 36(1), 39–53.
- Morrison, C. N., Thompson, J., Kondo, M. C., & Beck, B. (2019). On-road bicycle lane types, roadway characteristics, and risks for bicycle crashes. *Accident Analysis & Prevention*, 123, 123–131.

- Mueller, N., Rojas-Rueda, D., Cole-Hunter, T., De Nazelle, A., Dons, E., Gerike, R., Götschi, T., Panis, L. I., Kahlmeier, S., & Nieuwenhuijsen, M. (2015). Health impact assessment of active transportation: A systematic review. *Preventive medicine*, 76, 103–114.
- Müller, S., Tscharaktschiew, S., & Haase, K. (2008). Travel-to-school mode choice modelling and patterns of school choice in urban areas. *Journal of transport Geography*, 16(5), 342–357.
- Nie, Q. (2020). *Preparing small urban areas for shared mobility with autonomous vehicles: A case study of college town*. The University of Alabama.
- Nolan, J., Sinclair, J., & Savage, J. (2021). Are bicycle lanes effective? the relationship between passing distance and road characteristics. *Accident Analysis & Prevention*, 159, 106184.
- Oh, S., Seshadri, R., Azevedo, C. L., Kumar, N., Basak, K., & Ben-Akiva, M. (2020). Assessing the impacts of automated mobility-on-demand through agent-based simulation: A study of singapore. *Transportation Research Part A: Policy and Practice*, 138, 367–388.
- Ohta, M., Mizoue, T., Mishima, N., & Ikeda, M. (2007). Effect of the physical activities in leisure time and commuting to work on mental health. *Journal of occupational health*, 49(1), 46–52.
- Oldread, K. M. (2011). How well do neighborhood characteristics predict transit ridership in a college town?
- Páez, A., & Whalen, K. (2010). Enjoyment of commute: A comparison of different transportation modes. *Transportation Research Part A: Policy and Practice*, 44(7), 537–549.
- Paudel, J., et al. (2021). Bus ridership and service reliability: The case of public transportation to campuses [Observed that service reliability improves during summer breaks but

- declines significantly during academic terms when student demand is high.]. *Transportation Research Record*.
- Pereira, R. H. M., Schwanen, T., & Banister, D. (2017). Distributive justice and equity in transportation. *Transport Reviews*, 37(2), 170–191. <https://doi.org/10.1080/01441647.2016.1257660>
- Pougala, J., Hillel, T., & Bierlaire, M. (2023). From one-day to multiday activity scheduling: Extending the oasis framework.
- Pucher, J., & Buehler, R. (2010). Walking and cycling for healthy cities. *Built Environment*, 36(4), 391–414.
- Roberts, L., Jones, G., & Brooks, R. (2018). Why do you ride?: A characterization of mountain bikers, their engagement methods, and perceived links to mental health and well-being. *Frontiers in psychology*, 9, 1642.
- Rose, J. M., Bliemer, M. C., & Hensher, D. A. (2008). Estimating the costs of fuel-efficient vehicle technologies in the presence of rp and sp data. *Transportation Research Part B: Methodological*, 42(5), 447–463. <https://doi.org/10.1016/j.trb.2007.09.004>
- Rubie, E., Haworth, N., Twisk, D., & Yamamoto, N. (2020). Influences on lateral passing distance when motor vehicles overtake bicycles: A systematic literature review. *Transport reviews*, 40(6), 754–773.
- Safe Routes to School National Partnership. (2015). Rural communities: A two-pronged approach for improving walking and bicycling [Accessed 2025-08-20].
- Sanchez, T. W., Shen, Q., & Peng, Z.-R. (2004). Transit mobility, jobs access and low-income labour participation in us metropolitan areas. *Urban Studies*, 41(7), 1313–1331. <https://doi.org/10.1080/0042098042000214815>
- Scheffer, A., Connors, R., & Viti, F. (2021). Trip chaining impact on within-day mode choice dynamics: Evidences from a multi-day travel survey. *Transportation Research Procedia*, 52, 684–691.

- Schimek, P. (2018). Bike lanes next to on-street parallel parking. *Accident Analysis & Prevention*, 120, 74–82.
- Shaaban, K., & Kim, I. (2016). The influence of bus service satisfaction on university students' mode choice. *Journal of Advanced Transportation*, 50(6), 935–948.
- Singh, J., de Almeida Correia, G. H., van Wee, B., & Barbour, N. (2023). Change in departure time for a train trip to avoid crowding during the covid-19 pandemic: A latent class study in the netherlands. *Transportation Research Part A: Policy and Practice*, 170, 103628.
- Small, K. A. (1982). The scheduling of consumer activities: Work trips. *The American Economic Review*, 72(3), 467–479. <https://www.jstor.org/stable/1804013>
- Song, Y., Li, D., Liu, D., Cao, Q., Chen, J., Ren, G., & Tang, X. (2022). Modeling activity-travel behavior under a dynamic discrete choice framework with unobserved heterogeneity. *Transportation research part E: logistics and transportation review*, 167, 102914.
- Starnes, H. A., Troped, P. J., Klenosky, D. B., & Doehring, A. M. (2011). Trails and physical activity: A review. *Journal of Physical Activity and Health*, 8(8), 1160–1174.
- Steed, J. L., & Bhat, C. R. (2000). On modeling departure-time choice for home-based social/recreational and shopping trips. *Transportation Research Record*, 1706(1), 152–159.
- Subramanian, L. D., Plumert, J., & Kearney, J. (2020). Using simulation to assess right-hook conflicts between bicycles and cars at protected and unprotected intersections. *Proceedings of the DSC*.
- Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. *Proceedings of the IEEE international conference on computer vision*, 843–852.
- Taylor, C. B., Sallis, J. F., & Needle, R. (1985). The relation of physical activity and exercise to mental health. *Public health reports*, 100(2), 195.

- Thorhauge, M. (2015). Departure time choice: Modelling individual preferences, intention and constraints.
- Thorhauge, M., Cherchi, E., Walker, J. L., & Rich, J. (2019). The role of intention as mediator between latent effects and behavior: Application of a hybrid choice model to study departure time choices. *Transportation*, 46(4), 1421–1445.
- Thorhauge, M., Vij, A., & Cherchi, E. (2021). Heterogeneity in departure time preferences, flexibility and schedule constraints. *Transportation*, 48(4), 1865–1893.
- Toaza, B., & Esztergár-Kiss, D. (2023). A preliminary evaluation of activity scheduling optimization using real data from a household travel survey. *2023 8th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 1–6.
- Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.
- Tripathy, M., & Sahu, B. (2021). A comparative study on mental health of rural and urban students. *Int. Educ. Sci. Res. J*, 7, 67–77.
- United Nations, Department of Economic and Social Affairs, Population Division. (2018). World urbanization prospects: The 2018 revision—key facts [Accessed on 2025-08-20].
- U.S. Bureau of Labor Statistics. (2024). Consumer expenditures—2023 [Transportation accounts for 17.0% of total expenditures, behind housing at 32.9% (accessed August 20, 2025)].
- U.S. Census Bureau. (2016). *New census data show differences between urban and rural populations* [Accessed: July 13, 2025]. <https://www.census.gov/newsroom/press-releases/2016/cb16-210.html>
- U.S. Department of Education, National Center for Education Statistics. (2025). Postsecondary enrollment, fall 2023 (ipeds trend generator) [Accessed on 2025-08-20].
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

- Velasco, L., Rojo, M., Gonzalo-Orden, H., & Diez, J. M. (2015). Safety issues with elderly cyclists and barriers to cycling. *Proceedings of the institution of civil engineers-municipal engineer*, 168(2), 87–95.
- Virginia Tech Transportation Services. (2024). Virginia tech e-bike share program [Accessed: 2025-07-25]. <https://parking.vt.edu/alternative/bikeshare.html?>
- Volosin, S. E. (2014). *A study of university student travel behavior* (Technical Report) (Survey-based analysis of student travel modes, campus facility access, housing location, and budget constraints). Arizona State University.
- Vovsha, P., Petersen, E., & Donnelly, R. (2004). Modeling joint travel destination and departure time choices for urban trips. *Transportation Research Record*, 1898(1), 30–39.
- Vyas, G., Vovsha, P., Paul, B., Givon, D., & Livshits, V. (2015). Allocation of individual nonmandatory activities to day segments in tour formation procedure: Application to activity-based models for jerusalem, israel, and phoenix, arizona. *Transportation Research Record*, 2493(1), 88–98.
- Walker, J., & Ben-Akiva, M. (2002). Generalized random utility model. *Mathematical social sciences*, 43(3), 303–343.
- Weyerer, S., & Kupfer, B. (1994). Physical exercise and psychological health. *Sports Medicine*, 17, 108–116.
- Wild, K., & Woodward, A. (2019). Why are cyclists the happiest commuters? health, pleasure and the e-bike. *Journal of Transport & Health*, 14, 100569.
- Wojan, T. R., & Hamrick, K. S. (2015). Can walking or biking to work really make a difference? compact development, observed commuter choice and body mass index. *PLoS One*, 10(7), e0130903.
- Woodson, C., Ashqar, H. I., Almannaa, M., Elhenawy, M., & Buehler, R. (2024). Factors influencing bikeshare service and usage in a rural college town: A case study of mont-

- gomery county, va. *International Journal of Sustainable Transportation*, 18(4), 291–300.
- Wu, C., Fritz, H., Miller, M., Craddock, C., Kinney, K., Castelli, D., & Schnyer, D. (2021). Exploring post covid-19 outbreak intradaily mobility pattern change in college students: A gps-focused smartphone sensing study. *Frontiers in Digital Health*, 3, 765972.
- Xue, Y., Fan, H., & Guan, H. (2019). Commuter departure time choice considering parking space shortage and commuter's bounded rationality. *Journal of Advanced Transportation*, 2019(1), 4864287.
- Yang, L., Zheng, G., & Zhu, X. (2013). Cross-nested logit model for the joint choice of residential location, travel mode, and departure time. *Habitat International*, 38, 157–166.
- Yao, L., Fang, H., Leng, W., Li, J., & Chang, J. (2021). Effect of aerobic exercise on mental health in older adults: A meta-analysis of randomized controlled trials. *Frontiers in psychiatry*, 12, 748257.
- Youngstedt, S. D. (2005). Effects of exercise on sleep. *Clinics in sports medicine*, 24(2), 355–365.
- Yu, B., Yin, H., & Zhu, Z. (2017). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. *arXiv preprint arXiv:1709.04875*.
- Zannat, K., Choudhury, C. F., & Hess, S. (2022). Modeling departure time choice of car commuters in dhaka, bangladesh. *Transportation research record*, 2676(2), 247–262.
- Zhang, H., Wu, H., Sun, W., & Zheng, B. (2018). Deeptrip: A neural network based travel time estimation model with auxiliary supervision. *arXiv preprint arXiv:1802.02147*.
- Zhang, J., Hayashi, Y., & Frank, L. D. (2021). Covid-19 and transport: Findings from a world-wide expert survey. *Transport policy*, 103, 68–85.
- Zhang, J., Jiang, Y., & Zhang, Y. (2025). Transport-in-all-policies approach. In *Research handbook on transport and covid-19* (pp. 469–490). Edward Elgar Publishing.

- Zhou, H., Li, W., Jiang, Z., Cai, F., & Xue, Y. (2022). Flight departure time prediction based on deep learning. *Aerospace*, 9(7), 394.
- Zhou, J. (2014). From better understandings to proactive actions: Housing location and commuting mode choices among university students. *Transport Policy*, 33, 166–175.
- Zhou, J., Wang, Y., & Wu, J. (2018). Mode choice of commuter students in a college town: An exploratory study from the united states. *Sustainability*, 10(9), 3316.

Appendices

Appendix A

Appendix

A.1 RP-SP Survey Design

This appendix provides the complete survey questionnaire that was administered as part of the data collection for Chapters 5 and 6 of this dissertation. The survey was designed to capture both revealed preference (RP) and stated preference (SP) data from students in 14 rural college towns across Virginia. The stated preference (SP) section consisted of six hypothetical scenarios, each carefully constructed using D-efficient experimental design in NGENE. These scenarios evaluated how students might respond to changes in transportation options under different conditions, particularly focusing on the role of bicycle infrastructure, travel costs, and time. Each respondent faced six SP scenarios in total, which allowed us to capture variations in preferences under diverse policy and infrastructure setups .

The revealed preference (RP) section of the survey captured participants' actual travel behavior, including their current mode choices, trip purposes, travel times, and travel costs. This section provided real-world insights into existing travel patterns under current infrastructure and policies .

Together, the RP and SP sections of the survey provided a comprehensive understanding of how students in rural college towns make transportation decisions, both under current conditions and in response to potential improvements in bicycle infrastructure.

The questions attached in this appendix are the actual questionnaire pages as presented to

respondents. These materials document the instrument used in data collection and demonstrate the structure of the experiment.

A.2 RP-SP Survey Questionnaire

The following section contains the survey instrument that was used to collect both stated preference (SP) and revealed preference (RP) data for chapter 5 and 6. The SP portion of the questionnaire presented respondents with six hypothetical transportation scenarios, each designed to evaluate how students would adjust their mode choices under different conditions of cost, time, and bicycle infrastructure. These scenarios provided insights into how policy or infrastructure changes could influence behavior beyond existing conditions.

In contrast, the RP portion of the questionnaire asked respondents about their actual, day-to-day travel choices, including mode, purpose, time, and cost. This combination of SP and RP data allowed for the development of joint econometric models in Chapters 5 and 6, where real-world travel behavior could be compared against hypothetical choices.

The following figures reproduce the exact survey pages shown to respondents, with displaying the six SP scenarios and RP questions:

* Consent to Participate in Research Study

Title of Research Study: IRB #24-358 A Policy Analysis Tool for Measuring the Effectiveness of Cycling Infrastructure in College Towns

Student Investigator: Nishat Naila Meghna

You are invited to participate in a research study aiming to evaluate the effectiveness of cycling infrastructure in the Blacksburg, Virginia area. This study, conducted by Virginia Tech, seeks to understand how well the current cycling infrastructure meets the needs of its users and to gather insights that could lead to improvements. Your participation will involve completing an online survey, expected to take about 20-35 minutes.

Study Description:

Upon agreeing to participate, you will be asked to complete the survey in a single session. The survey will query your preferences and behaviors regarding cycling and other modes of transportation, focusing on factors that influence your choices. This is a voluntary study, and you may withdraw at any time without any consequences.

Privacy and Confidentiality:

We assure you that your responses will be treated with the highest confidentiality. All data collected will be anonymized, ensuring that personal identifiers are not linked to your responses. Access to this data will be restricted to the research team and approved personnel only.

Potential Risks and Benefits:

Participating in this study involves minimal risk, comparable to everyday use of the internet. While there is no direct compensation for participating, your input will contribute significantly to our understanding of transportation infrastructure effectiveness and may influence future improvements in your community.

Consent Acknowledgment:

By participating in this study, you acknowledge that you understand the purpose and procedures of the study and agree to partake voluntarily. A copy of this consent form will be provided to you for your records.

Contact Information:

Should you have any questions about the study or your participation, feel free to contact the research team via email at nishatmeghna@vt.edu. If you have concerns about your rights as a participant, or if you wish to discuss your participation with someone other than the research team, you may contact the Institutional Review Board at Virginia Tech at 540-231-3732 or via email at irb@vt.edu.

Do you consent to participate?

☐ Yes

☐ No

Are you of at least 18 years old?

☒ Yes

☐ No

Is your home or work location in any of the following area in Virginia: Blacksburg, Christiansburg, Radford, Dublin, Pulaski, Roanoke, Salem, Lynchburg, Charlottesville, Harrisonburg, Williamsburg, Bridgewater, Hampden Sydney, Longwood?

☐ Yes

☐ No

* Below you will find a scenario that describes different travel modes along with their associated conditions such as cost, environmental impact, and so on. We assumed the same travel distance of 2 miles for each mode. An auto passenger is someone riding in a car with a driver and carpooling is when several people share a car ride to save money. A designated bicycle lane is a portion of the roadway exclusively reserved for bicycles, marked by lines and signage.

Please select the option that best fits your preferences from the scenario.

Attribute / Travel mode	Driving	Auto Passenger	Uber	Carpool	Public Transit	Bike
These remain consistent between scenarios						
Total Travel Distance (miles)	2	2	2	2	2	2
Cost (Dollars)	0.28	0.14	7.50	0.14	Free	Free
Traffic Congestion Level	High	High	High	High	Low	Low
Carbon Di-Oxide Emissions	High	High	High	High	Low	Low
Availability During Peak Hours to you	Always	Sometimes	Mostly all the time	Sometimes	For fixed hours	Always
Dependence on Weather	Independent	Independent	Dependent	Independent	dependent	Dependent
These changes between scenarios						
Designated Bicycle Lane	-	-	-	-	-	Yes
Sheltered bike parking, showers and changing room available at workplace	-	-	-	-	-	No
Risk of Accidents	-	-	-	-	High	Low
Maintenance Needs	Low	-	-	-	-	High
Parking availability during peak hours	Available	-	-	-	-	Available
Uphill Road	-	-	-	-	No	No
Downhill road	-	-	-	-	No	No

- ☐ Driving
- ☐ Auto Passenger
- ☐ Carpool
- ☐ Public Transit
- ☐ Bike
- ☐ Uber

* Below you will find a scenario that describes different travel modes along with their associated conditions such as cost, environmental impact, and so on. We assumed the same travel distance of 5 miles for each mode. An auto passenger is someone riding in a car with a driver and carpooling is when several people share a car ride to save money. A shared bike lane is a roadway space marked for both bicycles and other vehicles, typically indicated by shared lane markings or "sharrows".

Please select the option that best fits your preferences from the scenario.

Attribute / Travel mode	Driving	Auto Passenger	Uber	Carpool	Public Transit	Bike
These remain consistent between scenarios						
Total Travel Distance (miles)	5	5	5	5	5	5
Cost (Dollars)	0.70	0.35	13.50	0.35	Free	Free
Traffic Congestion Level	High	High	High	High	High	High
Carbon Di-Oxide Emissions	High	High	High	High	High	Low
Availability During Peak Hours to you	Always	Sometimes	Mostly all the time	Sometimes	For fixed hours	Always
Dependence on Weather	Independent	Independent	Dependent	Independent	dependent	Dependent
These changes between scenarios						
Shared bicycle lane	-	-	-	-	-	Yes
Sheltered bike parking, showers and changing room available at workplace	-	-	-	-	-	Yes
Risk of Accidents	-	-	-	-	Low	High
Maintenance Needs	Low	-	-	-	-	Low
Parking availability during peak hours	Available	-	-	-	-	Unavailable
Uphill Road	-	-	-	-	Yes	No
Downhill road	-	-	-	-	Yes	No

- ☐ Driving
- ☐ Auto Passenger
- ☐ Carpool
- ☐ Public Transit
- ☐ Bike
- ☐ Uber

* Below you will find a scenario that describes different travel modes along with their associated conditions such as cost, environmental impact, and so on. We assumed the same travel distance of 7 miles for each mode. An auto passenger is someone riding in a car with a driver and carpooling is when several people share a car ride to save money. A two-way protected bicycle lane allows bike traffic in both directions and is separated from vehicle lanes by physical barriers for enhanced safety.

Please select the option that best fits your preferences from the scenario.

Attribute / Travel mode	Driving	Auto Passenger	Uber	Carpool	Public Transit	Bike
These remain consistent between scenarios						
Total Travel Distance (miles)	7	7	7	7	7	7
Cost (Dollars)	0.98	0.49	17.50	0.49	Free	Free
Traffic Congestion Level	Low	Low	Low	Low	Low	Low
Carbon Di-Oxide Emissions	Low	Low	Low	Low	Low	High
Availability During Peak Hours to you	Always	Sometimes	Mostly all the time	Sometimes	For fixed hours	Always
Dependence on Weather	Independent	Independent	Dependent	Independent	dependent	Dependent
These changes between scenarios						
Two-Way Protected Bicycle Lane	-	-	-	-	-	Yes
Sheltered bike parking, showers and changing room available at workplace	-	-	-	-	-	Yes
Risk of Accidents	-	-	-	-	High	Low
Maintenance Needs	High	-	-	-	-	High
Parking availability during peak hours	Unavailable	-	-	-	-	Unavailable
Uphill Road	-	-	-	-	Yes	Yes
Downhill road	-	-	-	-	No	Yes

- ☐ Driving
- ☐ Auto Passenger
- ☐ Carpool
- ☐ Public Transit
- ☐ Bike
- ☐ Uber

* Below you will find a scenario that describes different travel modes along with their associated conditions such as cost, environmental impact, and so on. An auto passenger is someone riding in a car with a driver and carpooling is when several people share a car ride to save money. We assumed the same travel distance of 10 miles for each mode.

Please select the option that best fits your preferences from the scenario.

Attribute / Travel mode	Driving	Auto Passenger	Uber	Carpool	Public Transit	Bike
These remain consistent between scenarios						
Total Travel Distance (miles)	10	10	10	10	10	10
Cost (Dollars)	1.40	0.70	23.50	0.70	Free	Free
Traffic Congestion Level	Low	Low	Low	Low	High	High
Carbon Di-Oxide Emissions	Low	Low	Low	Low	High	High
Availability During Peak Hours to you	Always	Sometimes	Mostly all the time	Sometimes	For fixed hours	Always
Dependence on Weather	Independent	Independent	Dependent	Independent	dependent	Dependent
These changes between scenarios						
Sheltered bike parking, showers and changing room available at workplace	-	-	-	-	-	Yes
Risk of Accidents	-	-	-	-	Low	High
Maintenance Needs	High	-	-	-	-	Low
Parking availability during peak hours	Unavailable	-	-	-	-	Available
Uphill Road	-	-	-	-	Yes	Yes
Downhill road	-	-	-	-	No	Yes

- ☐ Driving
- ☐ Auto Passenger
- ☐ Carpool
- ☐ Public Transit
- ☐ Bike
- ☐ Uber

* Below you will find a scenario that describes different travel modes along with their associated conditions such as cost, environmental impact, and so on. We assumed the same travel distance of 15 miles for each mode. An auto passenger is someone riding in a car with a driver and carpooling is when several people share a car ride to save money. A protected bicycle lane is separated from vehicle traffic by physical barriers to ensure cyclist safety. Please select the option that best fits your preferences from the scenario.

Attribute / Travel mode	Driving	Auto Passenger	Uber	Carpool	Public Transit	Bike
These remain consistent between scenarios						
Total Travel Distance (miles)	15	15	15	15	15	15
Cost (Dollars)	2.10	1.05	33	1.05	Free	Free
Traffic Congestion Level	High	High	High	High	High	Low
Carbon Di-Oxide Emissions	High	High	High	High	High	High
Availability During Peak Hours to you	Always	Sometimes	Mostly all the time	Sometimes	For fixed hours	Always
Dependence on Weather	Independent	Independent	Dependent	Independent	dependent	Dependent
These changes between scenarios						
Protected Bicycle Lane	-	-	-	-	-	Yes
Sheltered bike parking, showers and changing room available at workplace	-	-	-	-	-	No
Risk of Accidents	-	-	-	-	High	Low
Maintenance Needs	Low	-	-	-	-	High
Parking availability during peak hours	Available	-	-	-	-	Available
Uphill Road	-	-	-	-	No	No
Downhill road	-	-	-	-	Yes	No

- ☐ Driving
☐ Auto Passenger
☐ Carpool
☐ Public Transit
☐ Bike
☐ Uber

* Below you will find a scenario that describes different travel modes along with their associated conditions such as cost, environmental impact, and so on. We assumed the same travel distance of 20 miles for each mode. An auto passenger is someone riding in a car with a driver and carpooling is when several people share a car ride to save money. A buffered bicycle lane features extra space between the bike lane and vehicle traffic for increased cyclist safety.

Please select the option that best fits your preferences from the scenario.

Attribute / Travel mode	Driving	Auto Passenger	Uber	Carpool	Public Transit	Bike
These remain consistent between scenarios						
Total Travel Distance (miles)	20	20	20	20	20	20
Cost (Dollars)	2.80	1.40	43.50	1.40	Free	Free
Traffic Congestion Level	Low	Low	Low	Low	Low	High
Carbon Di-Oxide Emissions	Low	Low	Low	Low	Low	Low
Availability During Peak Hours to you	Always	Sometimes	Mostly all the time	Sometimes	For fixed hours	Always
Dependence on Weather	Independent	Independent	Dependent	Independent	dependent	Dependent
These changes between scenarios						
Buffered Bicycle Lane	-	-	-	-	-	Yes
Sheltered bike parking, showers and changing room available at workplace	-	-	-	-	-	No
Risk of Accidents	-	-	-	-	Low	High
Maintenance Needs	High	-	-	-	-	Low
Parking availability during peak hours	Unavailable	-	-	-	-	Unavailable
Uphill Road	-	-	-	-	No	Yes
Downhill road	-	-	-	-	Yes	Yes

- ☐ Driving
- ☐ Auto Passenger
- ☐ Carpool
- ☐ Public Transit
- ☐ Bike
- ☐ Uber

* What is your sex assigned at birth?

- ☐ Male
- ☐ Female
- ☐ Other

* What is your age range in years?

-- Select --



* Are you of Hispanic or Latino origin?

- ☐ Yes
- ☐ No

* Select your race from dropdown:

-- Select -- 

* What is the highest grade or level of education you have earned?

- ☐ Less than a high school graduate
- ☐ High school graduate or GED
- ☐ Some college or associates degree
- ☐ Bachelor's degree
- ☐ Graduate degree or professional degree
- ☐ Others

* What is your current employment status (if you are working and a student please select the working option closest to you)?

-- Select -- 

* Are you a student?

- ☐ Yes, in K-12th Grade
- ☐ Yes, Vocation/Technical/Trade School
- ☐ Yes – Part-Time college/University
- ☐ Yes – Full Time college/University
- ☐ Yes – Other
- ☐ No

- * What is your home zip code? (If you do not know please visit this website: <https://tools.usps.com/zip-code-lookup.htm?byaddress> or another website that provides your zip code)?

- * What is the Zip Code of your work location(If you do not know please visit this website: <https://tools.usps.com/zip-code-lookup.htm?byaddress> or another website that provides your zip code)?

- * Which dwelling type best describes your current home?

- ☐ Single Family House (detached)
- ☐ Semi-Detached House
- ☐ Town/Row House
- ☐ Apartment/Flat in detached duplex
- ☐ Apartment/Condo with less than 5 stories
- ☐ Apartment/Condo with more than 5 stories
- ☐ Other

- * How many individuals (including yourself) live in your household?

-- Select -- 

* How many individuals (including yourself) live in your household over the age of 12?

-- Select -- 

* Do you possess a driver's license?

☐ Yes

☐ No

* How many functioning motor vehicles does your household have?

-- Select -- 

* How mow many functioning bikes does your household have?

-- Select -- 

* Do you own an easy toll pass?

☐ Yes

☐ No

* Do you use the easy toll pass on your average trip?

☐ Yes

☐ No

* Do you use the Smart Trip or other pass for public transportation on your average trip?

☐ Yes

☐ No

* What is your regular mode choice?

☐ Drive

☐ Auto Passenger

☐ Carpool

☐ Bike

☐ Uber/Lyft

☐ Public Transit

☐ Other

* How many trips do you have in a typical day?

-- Select -- 

* How many time do you cycle in a typical week?

-- Select -- 

* How many trips do you have in a single day where your mode of transportation is cycle?

-- Select -- 

* What is your primary reason for cycling?

- ☐ Commuting to work/school
- ☐ Exercise
- ☐ Leisure/recreation
- ☐ Running errands
- ☐ I do not bike to commute

* How far (in mile) do you typically cycle in a typical week?

-- Select --



* Which type of route do you commonly use for cycling? If you don't bike, Please select another and write N/A in that.

- ☐ Residential streets
- ☐ Main roads with bike lanes
- ☐ Off-road trails
- ☐ Mixed-use paths (shared with pedestrians)
- ☐ I do not use bike for commuting

* What type of cycling infrastructure do you use most often?

- ☐ Designated bike lanes
- ☐ Shared bike lanes (sharrows)
- ☐ Protected bike lanes
- ☐ I do not use bike for commuting
- ☐ Buffered Bicycle Lane
- ☐ Two-way protected bicycle lane

* What are the biggest challenges you face while cycling in Montgomery Area?

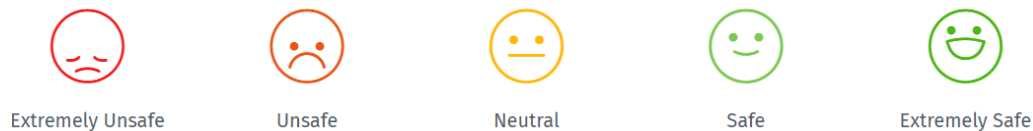
- ☐ Traffic congestion
- ☐ Lack of bike lanes
- ☐ Poor road conditions
- ☐ Safety concerns
- ☐ Weather conditions
- ☐ I do not use bike for commuting

* How often do you encounter: (Irrespective of what mode you use to commute)



(If you do not bike to commute, you can ignore this question and move to the next question)

How safe do you feel while cycling on main roads in your area?



* Have you experienced any accidents or near misses while cycling in the past year?

- ☐ Yes, an accident
- ☐ Yes, a near miss
- ☐ No
- ☐ I do not bike

* Do you feel that the current cycling infrastructure meets your needs?

Yes, completely Yes, Mostly Neutral No, Somewhat No, not at all

Select Your Opinion



(If you do not bike to commute, you can ignore this question and move to next question)

How often do you use:

	Never	Once in a while	Most of the time	Always
Protected Bike Lanes				
Designated Bike Lanes				
Two-way protected bicycle lane				
Shared bike lane				
Buffered Bicycle Lane				
Bike Racks (A bike rack is a device to which bicycles can be securely attached for parking purposes)				

* Would you prefer wider bike lanes or more bike lanes in general?

- ☐ Wider bike lanes
- ☐ More bike lanes
- ☐ No preference
- ☐ I do not use bike to commute

* How accessible are bike lanes from your residence?

- ☐ Very accessible
- ☐ Accessible
- ☐ Neutral
- ☐ Inaccessible
- ☐ Very inaccessible

* Do you feel that cycling infrastructure is well-connected to key destinations (e.g., work, shops, parks)?



Yes



No

* If you are a biker, how do you do in-store shopping i.e. which mode do you use for in-store shopping?

- ☐ Drive
- ☐ Public transit
- ☐ Carpool
- ☐ Auto passenger
- ☐ Uber
- ☐ Bike
- ☐ I do not use bike to commute

* If you are a biker, do you order more on-line?

- ☐ Yes
- ☐ No
- ☐ I am not a biker

* How important is the environmental impact of cycling to you?

- ☐ Not at all important
- ☐ Slightly important
- ☐ Moderately important
- ☐ Very important
- ☐ Extremely important

* How often do you use cycling infrastructure for:

Commuting Purpose

Never Once in a while Most of the time Always



(If you do not bike to commute, you can ignore this question and move to the next question)

How often do you cycle with:

Children or other dependents

Never Once in a while Most of the time Always



* Do you feel that there is adequate lighting on cycling paths?

- ☐ Yes
- ☐ No
- ☐ I did not notice

* Have you ever been a victim of bike theft?

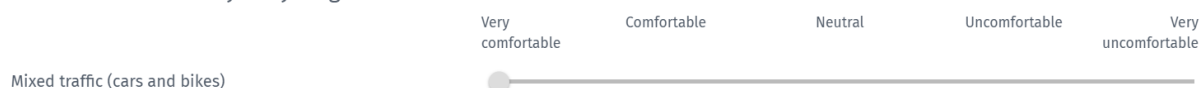
- ☐ Yes
- ☐ No
- ☐ I never had a bike

* Do you think there are enough bike repair stations in Blacksburg?

- ☐ Yes
- ☐ No
- ☐ I do not know

(If you do not bike, you can ignore this question and move to the next question)

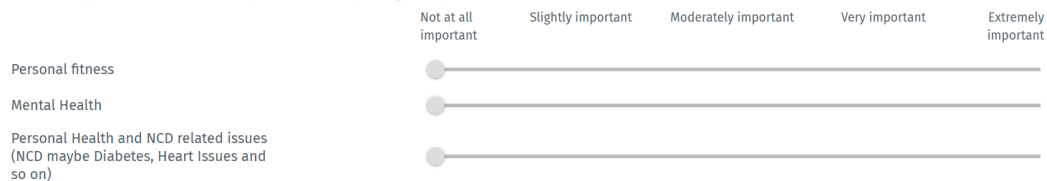
How comfortable are you cycling in:



* Do you believe that cycling infrastructure has improved in the past five years in your area?

- ☐ Yes, significantly
- ☐ Yes, somewhat
- ☐ Neutral
- ☐ No, it has stayed the same
- ☐ No, it has worsened

* What do you think? How important is cycling for your:



* What time do you usually depart from home for your first trip of the day?

-- Select -- 

* How many trips do you make in a single day?

-- Select -- 

* What factor influence your decision most on the time of departure for your trips?

- ☐ Work schedule
- ☐ School schedule
- ☐ Weather conditions
- ☐ Traffic conditions
- ☐ Personal preference
- ☐ Availability of public transportation
- ☐ Other

* How flexible are you with your departure time for daily trips?

- ☐ Very flexible
- ☐ Somewhat flexible
- ☐ Neutral
- ☐ Somewhat inflexible
- ☐ Very inflexible

* Give your opinion on these:

You plan your departure times to avoid peak traffic hours

Unexpected events (e.g., delays, emergencies) affect your planned departure times

You use tools or apps to plan your trip departures

You coordinate your departure times with other people (e.g., family, friends, colleagues) for group trips

You change your planned departure time due to safety concerns (e.g., low light, high traffic)

Always Often Sometimes Rarely Never

Five horizontal sliders are provided for rating the following behaviors: 'You plan your departure times to avoid peak traffic hours', 'Unexpected events (e.g., delays, emergencies) affect your planned departure times', 'You use tools or apps to plan your trip departures', 'You coordinate your departure times with other people (e.g., family, friends, colleagues) for group trips', and 'You change your planned departure time due to safety concerns (e.g., low light, high traffic)'. Each slider has a grey dot at the 'Always' end and a horizontal line extending to the 'Never' end.

(If you do not bike, you can ignore this question and move to the next question)

How does weather affect your decision on:

Not at all A little Neutral A lot Completely

Departure time for cycling trips

Decision to cycle

Two horizontal sliders are provided for rating the impact of weather on cycling decisions: 'Departure time for cycling trips' and 'Decision to cycle'. Each slider has a grey dot at the 'Not at all' end and a horizontal line extending to the 'Completely' end.

* Do you adjust your departure time based on the availability of biking infrastructure (e.g., bike lanes, bike-sharing stations)?

☐ Yes

☐ No

* How important is it for you to have real-time updates on traffic and road conditions before departing?

☐ Very important

☐ Important

☐ Neutral

☐ Not important

☐ Not important at all

* Do you prefer departing for trips during certain times of the day (e.g., morning, afternoon, evening)?

- ☐ Yes, morning
- ☐ Yes, afternoon
- ☐ Yes, evening

* How do you prioritize your activities when planning departure times for multiple trips in a day?

- ☐ Based on urgency of the activities
- ☐ Based on convenience
- ☐ Based on distance
- ☐ Based on other people's schedules
- ☐ Other

* Do you have any chronic diseases (e.g., diabetes, hypertension, heart disease)?

- ☐ Yes
- ☐ No

* Do you take medications for any chronic diseases (e.g., diabetes, hypertension, heart disease)?

- ☐ Yes
- ☐ No

* Do you have any diagnosed chronic conditions? (Select all that apply)

- ☐ Diabetes
- ☐ Hypertension
- ☐ Heart Disease
- ☐ Asthma
- ☐ Chronic Obstructive Pulmonary Disease (COPD)
- ☐ Cancer
- ☐ Kidney Disease
- ☐ Arthritis
- ☐ Other
- ☐ None

* What do you engage yourself in to manage your chronic condition(s)? (Select all that apply)

- ☐ Walking
- ☐ Jogging/Running
- ☐ Cycling
- ☐ Swimming
- ☐ Strength Training
- ☐ Yoga/Pilates
- ☐ Other

* How often do you visit a healthcare professional for your chronic condition?

-- Select --



* Do you engage in regular physical activity to manage your chronic condition?

- ☐ Yes, daily
- ☐ Yes, several times a week
- ☐ Yes, occasionally
- ☐ No
- ☐ I do not have any chronic disease

* How important do you believe physical activity is in managing your chronic condition?

- ☐ Very important
- ☐ Important
- ☐ Neutral
- ☐ Not important
- ☐ Not important at all

* Have you received any professional guidance on physical activity for managing your chronic condition?

- ☐ Yes
- ☐ No
- ☐ I do not have any chronic disease

* Do you think cycling could help in managing your chronic condition?

- ☐ Yes
- ☐ No
- ☐ I do not have any chronic disease

How many times do you cycle as part of your physical activity routine in a week?

-- Select --



* Do you have any mental health issues?

☐ Yes

☐ No

* Do you take medication for your mental health issues?

☐ Yes

☐ No

☐ I do not have mental health issue

* How often do you engage in physical activities like biking to manage your mental health?

-- Select --



* Do you feel that biking has a positive impact on mental health?

☒ Strongly agree

☐ Agree

☐ Neutral

☐ Disagree

☐ Strongly disagree

* In which year did you start biking regularly solely to manage mental health issues? (If you do not have any mental health issues, write N/A in the answer)

* How many months or years have you been biking regularly solely to only manage mental health issues if you have any? (If no, then select 0)

* Have you noticed any changes in your mood or stress levels after biking?

- ☐ Yes, significant improvement
- ☐ Yes, some improvement
- ☐ Neutral
- ☐ No, no change
- ☐ No, it has worsened
- ☐ I do not bike

* Do you use biking as a method to cope with stress or anxiety?

- ☐ Yes, always
- ☐ Yes, often
- ☐ Sometimes
- ☐ Rarely
- ☐ Never
- ☐ I do not bike

(If you do not bike, you can ignore this question and move to the next question)

How important is biking to your:

	Very important	Important	Neutral	Not important	Not important at all
Overall mental well-being	☐				
Overall quality of life	☐				

* Have you ever been advised by a healthcare professional to use physical activity, such as biking, to manage your mental health?

- ☐ Yes
- ☐ No
- ☐ I do not have mental health issue

* Do you participate in group biking activities to improve your social interactions and mental health?

- ☐ Yes, always
- ☐ Yes, often
- ☐ Sometimes
- ☐ Rarely
- ☐ Never

* How does biking affect your sleep patterns?

- ☐ Improves significantly
- ☐ Improves somewhat
- ☐ Neutral
- ☐ Worsens somewhat
- ☐ Worsens significantly
- ☐ I do not bike

(If you do not bike, you can ignore this question and move to the next question)

Have you experienced any reduction in symptoms of depression or anxiety due to regular biking?

- ☐ Yes, significant reduction
- ☐ Yes, some reduction
- ☐ Neutral
- ☐ No, no reduction
- ☐ No, symptoms worsened

(If you do not bike, you can ignore this question and move to the next question)

How often do you combine biking with other mindfulness activities (e.g., meditation, deep breathing) to enhance mental well-being?

- ☐ Daily
- ☐ Several times a week
- ☐ Weekly
- ☐ Occasionally
- ☐ Never

* Do you track your cycling activity (e.g., using apps or devices)?

- ☐ Yes
- ☐ No
- ☐ I do not bike

* Do you cycle year-round or only in specific seasons?

- ☐ Year-round
- ☐ Spring and summer
- ☐ Summer only
- ☐ Spring, summer, and fall
- ☐ I do not bike

* Would you use electric bikes if they were more accessible?

- ☐ Definitely
- ☐ Probably
- ☐ Neutral
- ☐ Probably not
- ☐ Definitely not

* What is your primary mode of transportation for commuting?

- ☐ Cycling
- ☐ Walking
- ☐ Driving
- ☐ Public transportation
- ☐ Auto passenger
- ☐ Carpool
- ☐ Other

(If you do not bike, you can ignore this question and move to the next question)

How often do you combine cycling with other modes of transportation (e.g., bus, train)?

- ☐ Always
 - ☐ Often
 - ☐ Sometimes
 - ☐ Rarely
 - ☐ Never
-

* How likely are you to support policies that restrict car traffic to promote cycling?

- ☐ Very likely
- ☐ Likely
- ☐ Neutral
- ☐ Unlikely
- ☐ Very unlikely

(If you do not bike, you can ignore this question and move to the next question)

How often do you cycle with a group?

- ☐ Always
- ☐ Often
- ☐ Sometimes
- ☐ Rarely
- ☐ Never

* How will you prioritize the following aspects when choosing a cycling route? (Rank from 1 to 5)

Safety	-- Select --	▼
Scenic beauty	-- Select --	▼
Directness	-- Select --	▼
Availability of bike lanes	-- Select --	▼
Smoothness of the path	-- Select --	▼

* How does the presence of cycling infrastructure influence your choice of residence?

- ☐ Significantly influences
- ☐ Somewhat influences
- ☐ Neutral
- ☐ Rarely influences
- ☐ Does not influence

* Which of the following technological innovations would most enhance your cycling experience?

- ☐ Real-time traffic updates
- ☐ Smart bike lights that adjust to conditions
- ☐ Integrated bike lock and GPS tracking (integrated bike lock with GPS tracking is a security device designed to protect bicycles from theft and help owners locate their bikes if they are stolen)
- ☐ Electric bike charging stations
- ☐ Others

* If cycling infrastructure improvements were to be made, what would you prefer to see first?

- ☐ Expanding the network of bike lanes
- ☐ Enhancing the safety of existing lanes
- ☐ Increasing bike parking facilities
- ☐ Introducing bike-sharing programs
- ☐ Others

* How do you think cycling infrastructure can contribute to local economic development?

- ☐ Significantly contribute
- ☐ Somewhat contribute
- ☐ Neutral
- ☐ Rarely contribute
- ☐ Does not contribute

* How do you feel about cycling infrastructure being part of mixed-use developments (e.g., shopping centers, residential complexes)?

- ☐ Very favorable
- ☐ Favorable
- ☐ Neutral
- ☐ Unfavorable
- ☐ Very unfavorable

* What type of community support would enhance your cycling experience most?

- ☐ More cycling clubs
- ☐ Local business discounts for cyclists
- ☐ Educational programs for drivers and cyclists
- ☐ Increased police presence for safety
- ☐ Other

* How important it is:



* How important it is:

