Developing a Framework for Prioritizing Bicycle Safety Improvement Projects using Crowdsourced and Image-Based Data

August 2023 | Final Report

SAFE-D
SAFETY THROUGH DISRUPTION

Texas A&M Transportation Institute

San Diego State University
Leadership Starts Here

VIRGINIA TECH TRANSPORTATION INSTITUTE
VIRGINIA TECH

Disclaimer

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. 06-010	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Developing a Framework for Prioritizing Bicycle Safety Improvement Projects using Crowdsourced and Image-Based Data		5. Report Date August 2023 6. Performing Organization Code:
7. Author(s) Amir Reza Sadeghi [1] Arash Jahangiri [1] Sahar Ghanipoor Machiani [1] Steve Hankey [2] Seyed Sajjad Abdollahpour [2]		8. Performing Organization Report No. Report 06-010
9. Performing Organization Name and Address: Safe-D National UTC [1] San Diego State University [2] Virginia Tech School of Public and International Affairs		10. Work Unit No. 11. Contract or Grant No. 69A3551747115/Project 09-010
12. Sponsoring Agency Name and Address Office of the Secretary of Transportation (OST) U.S. Department of Transportation (US DOT)		13. Type of Report and Period Final Research Report 01/2022-12/2023 14. Sponsoring Agency Code

15. Supplementary Notes

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.

16. Abstract

Active transportation, including walking and cycling, has gained popularity due to the economic, environmental, and energy-efficient benefits. However, the rise of active transportation has also led to an increase in fatalities, particularly for bicyclists. A crash-risk scoring method was proposed to prioritize bicycle safety improvement projects for 50 bridges located in San Diego County. This study employs surrogate safety measures to estimate crash risk, addressing the limitations of traditional data collection methods, and incorporates transportation equity factors into the safety measure scoring method. To identify significant factors contributing to the likelihood of bicyclists exceeding 10 mph on bridges, binomial logistic regression models were employed, with three models focusing on different predictor variables. The results showed that factors such as race, home-to-work travel patterns, education levels, and crime rates influenced bicyclists' speeds on bridges. This study provides a foundation for understanding the factors associated with bicyclists' speeds on bridges and can inform future safety improvement projects in San Diego County and beyond. The findings highlight the importance of considering a range of factors to improve bicyclist safety and can ultimately lead to safer and more equitable transportation for all.

17. Key Words Bicycle safety, risk analysis, surrogate safety measure, crowdsourced data, equity	18. Distribution Statement No restrictions. This document is available to the public through the Safe-D National UTC website, as well as the following repositories: VTechWorks, The National Transportation Library, The Transportation Library, Volpe National Transportation Systems Center, Federal Highway Administration Research Library, and the National Technical Reports Library.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	66	$0

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

Abstract

Active transportation, including walking and cycling, has gained popularity due to the economic, environmental, and energy-efficient benefits. However, the rise of active transportation has also led to an increase in fatalities, particularly for bicyclists. A crash-risk scoring method was proposed to prioritize bicycle safety improvement projects for 50 bridges located in San Diego County. This study employs surrogate safety measures to estimate crash risk, addressing the limitations of traditional data collection methods, and incorporates transportation equity factors into the safety measure scoring method. To identify significant factors contributing to the likelihood of bicyclists exceeding 10 mph on bridges, binomial logistic regression models were employed, with three models focusing on different predictor variables. The results showed that factors such as race, home-to-work travel patterns, education levels, and crime rates influenced bicyclists' speeds on bridges. This study provides a foundation for understanding the factors associated with bicyclists' speeds on bridges and can inform future safety improvement projects in San Diego County and beyond. The findings highlight the importance of considering a range of factors to improve bicyclist safety and can ultimately lead to safer and more equitable transportation for all.

Acknowledgements

This project was funded by the Safety through Disruption (Safe-D) National University Transportation Center, a grant from the U.S. Department of Transportation – Office of the Assistant Secretary for Research and Technology, University Transportation Centers Program.

Table of Contents

List of Figures

List of Tables

Introduction

Over the last decade, non-motorized modes of transportation, including cycling and walking, have grown, as they are considered economical, environment-friendly, and energy-efficient (*1–3*). Active transportation enhancement is aligned with the U.S. Department of Transportation's objective of moving towards green cities and improving public health. Yet less than 2% of federal transportation funding is allocated to active transportation (*4*). However, with the expansion of active transportation, such as walking, cycling, in-line skating, skateboarding, etc., statistics and studies indicate a significant increase in the number of fatalities. According to the National Highway Traffic Safety Administration (NHTSA), between 2010 and 2019, a 36% increase in bicycle deaths occurred in the U.S. Furthermore, in 2021, 961 bicyclists were killed in crashes with vehicles (*5*). According to the Fatality Analysis Reporting System (FARS), pedestrians and bicyclists account for an increasing portion of overall traffic deaths in the U.S., with 19.5% of fatalities involving these vulnerable road users in 2018, compared to 12.6% in 2003 (*6*). Moreover, despite a 41% reduction in traffic volume in response to spring lockdowns caused by the Covid-19 pandemic, 697 bicyclists lost their lives in crashes in 2020; California was the deadliest state for bicyclists with 118 fatalities (*7*). Clearly, studies and statistics demonstrate an urgent need for bicycle safety improvements. Additionally, previous studies show a disproportionate bicycle crash distribution among people with different socio-demographic characteristics. According to prior research, Black and Hispanic bicyclists, as well as people who reside in areas with higher populations of non-White residents, lower and middle incomes, and higher poverty levels, are more likely to be involved in bicycle crashes (*7–9*).

Several approaches to identifying high-risk locations for pedestrians and cyclists rely solely on historical crash data, disregarding how different factors contribute to high-risk locations (*10*). They also either do not account for exposure data or only use rough estimates to measure pedestrian or bicyclist activities (*4*). The need for accurate exposure data is even more pronounced for pedestrians and bicyclists because, in some cities, volume data is not collected for these modes in the same way as for vehicular traffic (i.e., through loop detectors) (*11*). Thanks to advances in technology, acquiring travel information is no longer limited to traditional methods such as site counting and travel surveying (*12*).

Bike/pedestrian data can be categorized into two main groups: traditional data and emerging data (*13–17*). Manual site counting has been the primary method of collecting pedestrian/bike data, but with advances in traffic sensing technologies, automatic counting has been increasing (*18*). Various sensors such as inductive-loop detectors, pneumatic tubes, passive infrared sensors (e.g., PYRO from Ecocounter), and remote traffic microwave sensors have been used for automatic counting (*19*, *20*). Other data collection methods such as surveys can reveal additional information, including travelers' gender and age, but these are costly, time-consuming, and may not be reliable enough (*14*). In this project, a framework is proposed to prioritize bicycle safety improvement projects in the county of San Diego by developing crash risk scoring models to

study the relationships between different factors and crash risk. The incorporation of factors obtained from image-based data into crash risk scoring models and how they contribute to the risk scores was also investigated. A logistic regression model was utilized to obtain the p-values of each variable by fitting all potential predictors, and variables with a high p-value, indicating a lack of statistical significance in predicting the outcome variable, were subsequently removed. Based on their overall contribution to the model, each variable was assigned a component, and the odds ratios were utilized to compare the impact of each category of each variable on the target variable.

Literature Review

A number of factors, such as crash frequency, crash severity, and exposure, can be taken into account when determining high crash risk locations for bicyclists through scoring methods (*10*, *14*, *21–23*). Pedestrian and Bicyclist Intersection Safety Indices (Ped ISI and Bike ISI), for example, are a set of models developed by the Federal Highway Administration that are designed to help users prioritize intersections that need to be improved in terms of safety (*24*, *25*). A score index is allocated to each individual approach leg according to safety ratings, expert opinions, and observed behaviors, and higher scores indicate a higher priority for bicycle safety improvement (*10*, *24*, *25*).

In order to inform drivers about road safety in North America, the U.S. Road Assessment Program conducts systematic road safety assessments (*26*). As part of the program, roads are classified according to factors including fatal crash rates and a road protection score, based upon the successful European Road Assessment Program. By examining the potential for severe outcomes in different crash types, the protection score takes into account critical factors. The relative risk score for a road segment is calculated by weighing the risk-associated sub-factors together and determining the overall score (*10*, *26*).

The National Cooperative Highway Research Program sponsored a research initiative to develop ActiveTrans Priority Tool (APT) in recognition of the multitude of competing project priorities and selection criteria for pedestrian and bicycle projects (*27*). There was also a user-adaptable spreadsheet developed as part of this research effort. APT is a systematic process for improving pedestrian and bicycle facilities along existing roads using a systematic methodology (*28*, *29*). In order to develop the tool, a comprehensive literature review was conducted on pedestrian and bicycle prioritization methods, as well as different jurisdictional approaches. Approximately 450 agencies throughout North America were surveyed, interviewed, and subjected to case studies. According to the background research findings, there is a need for prioritization methodologies that balance the need for various projects and locations with their feasibility. By developing a proactive approach to prioritizing improvements for pedestrians and bicycles, the APT is able to meet the needs of a variety of transportation agencies (*10*, *27–29*). Using the Level of Traffic Stress (LTS) method, a new methodology has been developed to assess bicycle networks,

evaluating cyclist risk factors. By assessing the network's ability to provide travelers with seamless connections between their origins and destinations without exposing them to highly stressful links, this method is designed to determine low-stress connectivity (*30–32*). A street or intersection can be categorized as LTS 1 (suitable for children) up to LTS 4 (suitable for experienced cyclists who can share the road with vehicles traveling up to 35 mph). The Dutch standards for bicycle facilities serve as a benchmark for all levels, and they serve as a guide for assessing the level of cycling risk (*30–32*).

A logit model was used in some studies to determine the significance and influence of bicycle crash severity and frequency (*33–39*). Based on the results of a mixed generalized ordered response logit analysis of injury severity, findings revealed that a combination of factors such as age, speed limit on the roadway, location of crashes, and time of day were most influential. In a similar study, factors contributing to the injury severity in bicycle-motor vehicle accidents were explored using a multinomial logit model (*35*). Also, a logistic regression model was employed to determine how cyclist characteristics, road characteristics, and crash severity were related to crash severity and fatality (*33*). According to a recent study (*10*), segments and intersections were prioritized in terms of crash risk by developing logistic regression models and using odds ratios to quantify the risk by assigning scores to the attributes of segments and intersections. Previous research indicates that apps such as STRAVA are biased towards males and more athletic individuals (*40*, *41*). To our knowledge, previous studies did not include transportation equity factors (e.g., age, income level, race) and image-based data in their models to prioritize bicycle safety improvements. Statistics indicate that bicycle accidents are disproportionately more frequent among people with different socio-demographic profiles (*42–45*). Since fewer people own cars in low-income neighborhoods, a study indicates that people are more likely to walk and cycle in these areas, even when conditions are unsafe (*46*). Thus, they are more likely to be involved in bicycle accidents compared to other neighborhoods. According to (*7*, *47*), Hispanics and Blacks were more likely to be involved in bicycle crashes between 1999 and 2003. NHTSA also reported that in 2010, 38% of Hispanic bicyclists were involved in bicycle fatalities despite the fact that only 16% of U.S. citizens are Hispanic. The Governors Highway Safety Association concluded, from the analysis FARS data from 2015–2019, that it is clear that Black, Indigenous, and People of Color are disproportionately represented in fatal accidents. The bicyclist death per capita and nighttime bicyclist death rate was highest among American Indian/Alaskan Native people during this period. Among nighttime bicyclist deaths, Blacks were second in line, followed by Hispanics and Whites (*48*).

Historical bicycle accidents are infrequent on bridges, which makes obtaining sufficient data to develop a meaningful model challenging. Therefore, this study uses surrogate safety measures (SSMs) to predict crash risk in order to address the gap in historic crash data for bicyclists. An SSM is a quantitative indicator used to assess the potential for traffic conflicts or collisions without relying on actual crash data (*49*, *50*). Proactive safety assessment methods are needed to lower the likelihood of crashes. A foundation of accident prevention design can be established by

conducting an analysis of existing conditions (*51*). SSMs are frequently used to assess a transportation system's or a specific roadway's safety performance. They are useful because they can provide insights into potential safety problems before crashes occur. Speed is a significant factor in the occurrence and severity of crashes. Research indicates that higher speed variance is associated with higher crash frequency and severity (*52*, *53*). Speed-based measures include average speed, speed variance, and speed limit compliance. The SSM selected for this study was the percentage of bicyclists whose average speed on the bridge was greater than 10 mph. The threshold of 10 mph was used based on the literature and the data range (*54*).

Data Collection

StreetLight Data

StreetLight Data is a transportation analytics company that specializes in providing data and analytics to support transportation planning and decision-making (*55*). Their products are used by a variety of clients, including public agencies, private companies, and academic researchers. A variety of data sources are employed by the company, including mobile devices, connected vehicles, and public data sets. In addition to providing data on travel behavior, StreetLight Data can also provide demographic information, including information on age, income, and education level, as well as trip duration and origins (*12*, *14*, *56*). In Appendix B, supplementary discussion is presented regarding StreetLight data. In addition, Appendix C shows a few graphs visualizing some of the StreetLight variables.

EJScreen Data

ESRI GIS Layers and SANDAG shapefiles

Street-Level Metrics Based on Google Street View (GSV) Images

Most data sources that characterize the urban environment (e.g., Census Block) are aggregated at a relatively coarse spatial resolution. In this study, we employed an approach to extract street-level information on the built and natural environment from Google Street View (GSV) imagery. Briefly, we created a 100 m x 100 m grid for San Diego County. Where available, we then downloaded panoramic GSV images at the centroid of each grid cell (250,000 images). We then processed these images using a previously published deep learning model called the Pyramid Scene Parsing Network (PSPNet). PSPNet classifies each pixel in an image into one of 150 features that capture aspects of the built environment (e.g., road, building, sidewalk, etc.) and the natural environment (e.g., tree, grass, sky, water, etc.). We then tabulated metrics for each grid cell based on the processed imagery. The resulting dataset includes a raster with 150 bands (for each feature from PSPNet) at 100 m spatial resolution for San Diego County. For each pixel, the PSPNet predicts the category with the highest probability. The final step was selecting objects relevant to the aim of studies from a pool of 150 categories, and then aggregating them to the new categories of variables. To ensure that the selected categories were consistent with the aim of this study, we excluded items such as benches, signboards, or other indoor objects—overall, selected objects that

are a part of the human-controlled environment. The objects that we considered for this study can be categorized broadly into five main categories: (1) built environment, (2) transport network, (3) transport vehicles, (4) nature, and (5) vegetation. Additional information can be found in Appendix D.

Methodology

Model Selection

Logistic regression, initially introduced by the statistician David Cox in 1958, represents a type of generalized linear model designed for modeling the relationship between a binary outcome variable and various predictor variables (*61*). This class of models is capable of accommodating non-normal distributions of the response variable and non-linear relationships between the predictor and response variables (*62*). The logistic regression algorithm, a popular machine learning algorithm, is employed to classify data into two categories based on the linear combination of predictor variables transformed into a probability value between 0 and 1 using the sigmoid function, also known as the logistic function (*53*). The coefficients (β), representing the relationship between the predictor variables and the outcome variable, are crucial in logistic regression. The coefficients denote the change in the log-odds of the outcome variable for a one-unit increase in the corresponding predictor variable, holding all other predictor variables constant (*63*). The odds ratio, used to interpret the coefficients, is the multiplicative change in the odds of the outcome variable for one level of the categorical variable relative to the reference level (*64*). The logistic regression algorithm can handle interactions between predictor variables and continuous and categorical predictor variables, while techniques such as one-vs-all regression or SoftMax regression can be utilized to address multiclass classification problems (*65*).

Model Development

The objective of this study was to develop a model that could identify the risk of bicycle accidents on bridges by using an SSM. The model aimed to enable the comparison of each category of each variable with other categories of the same predictor and also with categories of the other predictors. The goal was to identify the San Diego County bridges with the highest need for improvement.

Surrogate Safety Measure (SSM)

Historical bicycle accidents are infrequent on bridges, which makes obtaining sufficient data to develop a meaningful model challenging. Therefore, an SSM was used instead of the actual number of accidents. An SSM refers to a measurable parameter that is correlated with the occurrence or severity of crashes. It is used to assess the safety of a road segment or intersection without relying on crash data. These measures are often used to identify high-risk locations for safety improvements before crashes occur, allowing for proactive and cost-effective safety planning (*66*). The SSM selected for this study was the percentage of bicyclists whose average

speed on the bridge was greater than 10 mph. This decision was based on the existing literature, which consistently points to a concerning correlation between increased speed and elevated crash rates, as well as the severity of those crashes. In this context, we considered the percentage of bicyclists whose average speed on the bridge exceeded 10 mph as our SSM, acknowledging that it serves as a practical indicator of potential safety concerns associated with higher speeds (*54*, *67*).

Categorization

To develop a meaningful model, the target variable and the predictors had to be categorized. The target variable was categorized into two groups based on the 50th percentile. In order to categorize the continuous variables, we considered several methods, including equal interval, equal frequency, and k-means. The equal interval method divides the range of values into equal-sized intervals, while the equal frequency method ensures that each interval has the same number of observations. After comparing the results of each method, we decided to use the equal frequency method to categorize the variables. This method ensured that each category had a sufficient number of observations, while allowing for more flexibility in the size of each bin.

Model Selection

After categorizing the variables, a logistic regression model was fitted with all potential predictors to obtain the *p*-values of each variable. The threshold of 0.05 has traditionally served as a widely accepted criterion for determining whether predictor variables exert a significant influence on the target variable. In model fitting, efforts were made to ensure that the p-values were less than 0.05 for at least one category within categorical variables. Three logistic regression models were then fitted with the remaining significant variables to obtain the final results.

Odds Ratio Calculation

Odds ratio represents the multiplicative change in the odds of the outcome variable for one level of the categorical variable relative to the reference level. For instance, consider a categorical variable with three distinct levels: A, B, and C. The odds ratio concerning level A in comparison to the reference level can be calculated by exponentiating the coefficient β_1.

$$Odds\ Ratio\ for\ Level\ A = \exp\left(\beta_1\right)$$

If the odds ratio for level A equals 2, this implies that when the predictor variable is at level A, the odds of the outcome variable are twice as high as they are when the predictor variable is at the reference level, while maintaining the constancy of all other predictor variables (*65*). The odds ratios for each category of the predictors were then extracted from the logistic regression model. The odds ratios were used to compare the impact of each category of each variable on the target variable. A higher odds ratio indicated a higher probability of accidents in that category. For additional information, please see Appendix E.

Risk Score Calculation

Each variable was assigned a component based on its overall contribution to the model. The odds ratios were then used to compare the impact of each category of each variable on the target variable (*65*). The significance level was used to compare the categories of different predictors and calculate the ratio between them. The odds ratios of all the categories of all the variables were scaled so that their total was equal to 100. Scores were assigned to each bridge out of 100, based on the data and variables related to each bridge. To be more specific, it should be determined which categories each variable falls into for each bridge. For Model 1, refer to Table 1; for Model 2, refer to Table 4; and for Model 3, refer to Table 7. Then, using Table 10 to Table 12, the summation of the scores for the variables of each bridge will provide a score ranging from 0 to 100 based on the following formula.

$$Risk\ Score\ by\ level = Max\ score\ \times \frac{Odds\ ratio - 1}{Highest\ Odds\ Ratio\ within\ the\ variable\ categories - 1}$$

$$Bridge\ Score = \sum_{i} Risk\ Score\ by\ level_i$$

Where: the term "Risk Score by Level" represents the score assigned to each level (category) of the categorical predictor. The "Max Score" is set at 33.33 when there are 3 predictors in the model and 25 when there are 4 predictors in the model. This adjustment ensures that the final score of each model falls within the range of 0 to 100. The "Odds Ratio" is calculated by exponentiating the coefficient associated with each predictor. The "Bridge Score" is the cumulative score assigned to each predictor of the bridge, calculated based on their respective odds ratios.

Prioritizing Locations for Improvement

To identify and prioritize the areas in greatest need of improvement, our approach involved the utilization of three distinct models to calculate a score for each bridge. While these models offer valuable insights, it is essential to also consider the actual usage of these bridges by cyclists in comparison to one another. A bridge may receive a high score from the models, yet if it is underutilized by cyclists, it may not warrant immediate attention. Conversely, a bridge with a lower model-assigned score may be heavily frequented by cyclists, indicating a substantial need for improvements. To strike a balance between these factors, we employ a weighted approach to prioritize locations. This entailed multiplying the average score of each bridge by the normalized bicycle volume specific to that bridge, which was calculated as the number of bicycles crossing that bridge divided by the total number of bicycles. This methodology highlighted bridges with both significant bicycle usage and a pressing need for improvements, allowing us to pinpoint the top priority locations for intervention.

$$Average\ Score = \frac{1st\ Model\ Score + 2nd\ Model\ Score + 3rd\ Model\ Score}{3}$$

$$Adjusted\ Score\ for\ each\ bridge = Average\ Score \times \frac{Bicycle\ Vol.}{Total\ bicycle\ Vol.}$$

Results and Discussion

First Model

Summary

Table 1. First Model Variable Categories

Variable	Source	Level 1	Level 2	Level 3	Level 4
2020 Households with One or More Persons with Disabilities (ACS 5-Year) – Percentage	Esri – Census Data	7.26–16.23	16.23–20.73	20.73–25.39	25.39–36.85
Individuals with Disabilities – Percentage	StreetLight Data	8.00 – 9.00	9.00 – 10.00	10.00 – 11.00	11.00 – 12.00
2022 Hispanic Population – Percentage	Esri – Census Data	15.05–20.81	20.81–26.08	26.08–40.83	40.83–85.84

Table 2. First Model Summary

Variable	Coefficients	Estimate Std. Error	z value	Pr(>|z|)
(Intercept)	-5.87	2.28	-2.57	0.0100 *
`SD2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent`1	3.62	1.77	2.05	0.0403 *
`SD2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent`2	4.71	1.84	2.56	0.0105 *
`SD2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent`4	0.67	1.19	0.57	0.5714
`With a disability`2	1.95	1.18	1.65	0.0985
`With a disability`3	2.38	1.46	1.62	0.1046
`With a disability`4	5.21	1.95	2.68	0.0074 *
`SD2022 Hispanic Population: Percent`2	1.44	1.19	1.21	0.2252
`SD2022 Hispanic Population: Percent`3	0.71	1.15	0.62	0.5365
`SD2022 Hispanic Population: Percent`4	3.91	1.74	2.24	0.0248 *

*variables with p-value less than 0.05
Dispersion parameter for binomial family taken to be 1
Null deviance: 69.315 on 49 degrees of freedom
Residual deviance: 44.716 on 40 degrees of freedom
AIC: 64.716
Number of Fisher Scoring iterations: 6

Table 3. First Model Results

Variable	Coefficient	Odds Ratio	Lower CI	Upper CI	P_Value
`SD2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent`1	3.62	37.30	1.79	2309.80	0.0403 *
`SD2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent`2	4.71	111.32	4.96	8140.76	0.0105 *
`SD2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent`4	0.68	1.96	0.19	23.95	0.5714

Variable	Coefficient	Odds Ratio	Lower CI	Upper CI	P_Value
`With a disability`2	1.95	7.04	0.81	98.67	0.0985
`With a disability`3	2.38	10.76	0.71	254.23	0.1046
`With a disability`4	5.21	183.54	6.80	17960.06	0.0074 *
`SD2022 Hispanic Population: Percent`2	1.44	4.22	0.48	60.37	0.2252
`SD2022 Hispanic Population: Percent`3	0.71	2.04	0.21	22.96	0.5365
`SD2022 Hispanic Population: Percent`4	3.91	49.83	2.59	2984.44	0.0248 *

Interpretation

An increase in the percentage of the Hispanic population residing in the vicinity of the bridge, which may be attributable to these communities being located in areas that receive less governmental attention.

- A decrease in the percentage of households with at least one person with a disability residing near the bridge. This variable, derived from census block groups, is distinct from the following variable, which focuses on travelers utilizing the bridge.
- An increase in the percentage of bicyclists with disabilities. This data, derived from StreetLight data, differs significantly from the previous variable. The discrepancy may be explained by the possibility that individuals with disabilities utilize specialized bicycles that allow them to travel at faster speeds on bridges.

Second Model

Summary

This section presents a logistic regression model (shown in Table 4) with a binary target variable and three predictor variables.

Table 4. Second Model Variable Categories

Variable	Source	Level 1	Level 2	Level 3
2022 Hispanic Population – Percentage	Esri – Cnsus	15.05–23.03	23.03–32.67	32.67–85.84
Home to Work – Percentage	StreetLight Data	0–21.67	21.67–28.07	28.07–34.5
Individuals with Disabilities – Percentage	StreetLight Data	7.80–9.63	9.63–11.17	11.17–12.90

As shown in Table 5, in this model, the null deviance is 69.315, and the residual deviance is 48.968. The difference between the two deviances is 20.347, indicating that the predictors explain a significant amount of variation in the response variable. The AIC value of 62.968 indicates a relatively good fit of the model, and the residual deviance of 48.968 on 43 degrees of freedom suggests that the model explains a significant portion of the variance in the data.

Table 5. Second Model Summary

Variable	Coefficients	Estimate Std. Error	z value	Pr(>|z|)
(Intercept)	-3.51	1.38	-2.55	0.0107 *
`SD2022 Hispanic Population: Percent`2	0.41	0.91	0.44	0.6562
`SD2022 Hispanic Population: Percent`3	2.30	1.10	2.10	0.0360 *

Variable	Coefficients	Estimate Std. Error	z value	Pr(>|z|)
`Home to Work`1	3.36	1.23	2.72	0.0065 *
`Home to Work`2	2.09	1.08	1.94	0.0521
`With a disability`2	3.34	0.85	0.40	0.6901
`With a disability`3	2.32	1.10	2.10	0.0356 *

*variables with p-value less than 0.05
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 69.315 on 49 degrees of freedom
Residual deviance: 48.968 on 43 degrees of freedom
AIC: 62.968
Number of Fisher Scoring iterations: 5

Table 6. Second Model Results

Variable	Coefficient	Odds Ratio	Lower CI	Upper CI	P_Value
`SD2022 Hispanic Population: Percent`2	0.41	1.50	0.25	10.00	0.6562
`SD2022 Hispanic Population: Percent`3	2.30	9.97	1.39	114.79	0.0360 *
`Home to Work`1	3.36	28.65	3.43	481.07	0.0065 *
`Home to Work`2	2.09	8.09	1.20	94.21	0.0522
` With a disability `2	0.34	1.40	0.27	7.89	0.6901
` With a disability `3	2.32	10.13	1.37	117.20	0.0356 *

*variables with p-value less than 0.05

Interpretation

Based on the output of the model, which is shown in Table 6, the odds ratios in this model indicate that the probability of exceeding 10 mph on bridges is higher under the following conditions:

- An increase in the percentage of the Hispanic population in level 3 areas, suggesting that bicyclists are more likely to exceed 10 mph in areas with a higher percentage of Hispanic population.
- A decrease in the percentage of Home-to-Work trips, indicating that individuals commuting to work routinely may not prioritize arriving at their destinations more quickly.
- An increase in the percentage of people with disabilities, suggesting that bridges where a higher percentage of people with disabilities travel are associated with faster bicyclists.

Third Model

Summary

This section presents the third logistic regression model with a binary target variable and four predictor variables, as seen in Table 7.

Table 7. Third Model Variable Categories

Variable	Source	Level 1	Level 2	Level 3
Home to Work – Percentage	StreetLight Data	0–21.32	21.32–27.93	27.93–34.50

Less Than High School Education – Percentage	EPA	0.00 – 5.00	5.00–12.36	12.36 – 37.00
2022 Total Crime Index	EPA	30–74	74–113	113–279
Standard Deviation building	GSV Images	0.0005–0.0327	0.0327–0.0571	0.0571–0.1174

As shown in Table 8, in this model, the null deviance is 69.315, and the residual deviance is 43.419. The difference between the two deviances is 25.896, indicating that the predictors explain a significant amount of variation in the response variable. The AIC value of 61.419 indicates a relatively good fit of the model, and the residual deviance of 43.419 on 41 degrees of freedom suggests that the model explains a significant portion of the variance in the data.

Table 8. Third Model Summary

Variable	Coefficients	Estimate Std. Error	z value	Pr(>| z|)
(Intercept)	-7.04	2.19	-3.22	0.0013*
`Home to Work`1	3.78	1.46	2.59	0.0095*
`Home to Work`2	1.97	1.18	1.67	0.0955
`Less Than High School Education (%)`2	3.01	1.38	2.19	0.0285*
`Less Than High School Education (%)`3	4.22	1.50	2.80	0.0051*
`SD2022 Total Crime Index`2	1.94	1.04	1.86	0.0627
`SD2022 Total Crime Index`3	2.36	1.18	2.01	0.0445*
`Standard Deviation building`1	3.20	1.18	2.71	0.0067*
`Standard Deviation building`3	1.09	0.94	1.16	0.2468

*variables with p-value less than 0.05
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 69.315 on 49 degrees of freedom
Residual deviance: 43.419 on 41 degrees of freedom
AIC: 61.419
Number of Fisher Scoring iterations: 5

Table 9. Third Model Results

Variable	Coefficient	Odds Ratio	Lower CI	Upper CI	P_Value
`Home to Work`1	3.78	43.62	3.71	1327.15	0.0095*
`Home to Work`2	1.97	7.15	0.81	95.90	0.0955
`Less Than High School Education (%)`2	3.01	20.36	1.90	470.09	0.0285*
`Less Than High School Education (%)`3	4.22	67.87	5.35	2367.50	0.0050*
`SD2022 Total Crime Index`2	1.94	6.99	1.03	70.27	0.0627
`SD2022 Total Crime Index`3	2.36	10.62	1.26	141.81	0.0445*
`Standard Deviation building`1	3.20	24.56	3.13	358.63	0.0067*
`Standard Deviation building`3	1.09	2.96	0.50	20.93	0.2468

*variables with p-value less than 0.05

Interpretation

Based on the output of the model, which is shown in Table 9, the odds ratios in this model indicate that the probability of exceeding 10 mph on bridges is higher under the following conditions:

- A decrease in the percentage of Home-to-Work trips (Level 1 and Level 2) compared to Level 3. This suggests that individuals commuting to work routinely may not prioritize arriving at their destinations more quickly.
- An increase in the percentage of the population with less than a high school education (Level 2 and Level 3) compared to Level 1. This may indicate that individuals with lower educational attainment are more likely to exceed 10 mph on bridges, potentially due to factors such as a lack of safety awareness or different commuting patterns.
- An increase in the total crime index (Level 2 and Level 3) compared to Level 1. This suggests that areas with higher crime rates are associated with a higher probability of bicyclists exceeding 10 mph on bridges. The relationship between crime rates and bicyclist behavior could be influenced by a variety of factors, such as perceptions of safety or the built environment.
- A decrease in the standard deviation of building characteristics (level 1) compared to other levels. This indicates that bridges located in areas with a lower standard deviation in building characteristics are associated with significantly higher odds of the positive class compared to reference level 2.

An in-depth comparison and discussion of models are presented in Appendix G.

Risk-Scoring Method

Table 10. Risk Score by Level (First Model)

Variable	Coefficient	Levels	Internal Weight (Odds Ratio)	Risk Score by Level
`SD2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent`1	3.62	7.26 %–16.23 %	37.30	10.97
`SD2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent`2	4.71	16.23 %–20.73 %	111.32	33.33
`SD2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent`3 (ref)	0	20.73 %–25.39 %	1	0
`SD2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent`4	0.68	25.39 %–36.85 %	1.96	0.29
`With a disability`1 (ref)	0	8 %–9 %	1	0
`With a disability`2	1.95	9 %–10 %	7.04	1.10
`With a disability`3	2.38	10 %–11 %	10.76	1.78
`With a disability`4	5.21	11 %–12 %	183.54	33.33
`SD2022 Hispanic Population: Percent`1 (ref)	0	15.05 %–20.81 %	1	0
`SD2022 Hispanic Population: Percent`2	1.44	20.81 %–26.08 %	4.22	2.20
`SD2022 Hispanic Population: Percent`3	0.71	26.08 %–40.83 %	2.04	0.71
`SD2022 Hispanic Population: Percent`4	3.91	40.83 %–85.84 %	49.83	33.33

Table 11. Risk Score by Level (Second Model)

Variable	Coefficient	Level	Internal Weight (OR)	Risk Score by Level
`SD2022 Hispanic Population: Percent`1 (ref)	0	15.05 %–23.03 %	1	0
`SD2022 Hispanic Population: Percent`2	0.41	23.03 %–32.67 %	1.50	1.86

`SD2022 Hispanic Population: Percent`3	2.30	32.67 %–85.84 %	9.97	33.33
`Home to Work`1	3.36	0 %–21.67 %	28.65	33.33
`Home to Work`2	2.09	21.67 %–28.07 %	8.09	8.55
`Home to Work`3 (ref)	0	28.07 %–34.5 %	1	0
`With a disability`1 (ref)	0	7.80 %–9.63 %	1	0
`With a disability`2	0.34	9.63 %–11.17 %	1.40	1.47
`With a disability`3	2.32	11.17 %–12.9 %	10.13	33.33

Table 12. Risk Score by Level (Third Model)

Variable	Coefficient	Levels	Internal Weight (OR)	Risk Score by Level
`Home to Work`1	3.78	0 %–21.32 %	43.62	25
`Home to Work`2	1.97	21.32 %–27.93 %	7.15	3.61
`Home to Work`3 (ref)	0	27.93 %–34.5 %	1	0
`Less Than High School Education (%)`1 (ref)	0	0 %–5 %	1	0
`Less Than High School Education (%)`2	3.01	5 %–12.36 %	20.36	7
`Less Than High School Education (%)`3	4.22	12.36 %–37 %	67.87	25
`SD2022 Total Crime Index`1 (ref)	0	30–74	1	0
`SD2022 Total Crime Index`2	1.94	74–113	6.99	15.57
`SD2022 Total Crime Index`3	2.36	113–279	10.62	25
`Standard Deviation building`1	3.20	0.0005–0.0327	24.56	25
`Standard Deviation building`2 (ref)	0	0.0327–0.0571	1	0
`Standard Deviation building`3	1.09	0.0571–0.1174	2.96	2.08

Total Risk Score for Bridges

BRIDGE	Score (Model 1)	Score (Model 2)	Score (Model 3)	Average Score	The percentage of bicyclists with a speed over 10 mph	Average Daily Zone Traffic (StL Volume) - Number of Bicyclists	(Avg. Score) * (Normalized Bicycle Vol.)
57 0112	35.11	66.66	42.65	48.14	0.27	160	0.82
57 0180	68.44	1.47	65.57	45.16	0.23	248	1.21
57 0219	12.07	66.66	52.08	43.60	0.20	677	3.27
57 0220	10.97	68.52	30.69	36.73	0.28	298	1.31
57 0260	66.95	34.8	50	50.58	0.25	178	1.12
57 0261	66.66	1.86	69.18	45.90	0.23	166	0.97
57 0269	35.11	35.19	53.61	41.30	0.23	243	1.30
57 0270	35.11	33.33	75	47.81	0.23	1	0.01
57 0272	35.11	10.41	51.18	32.23	0.19	224	0.97
57 0298	66.95	1.86	69.18	46.00	0.25	156	0.99
57 0341	2.1	11.88	26.18	13.39	0.26	229	0.43
57 0354	68.44	0	90.57	53.00	0.24	159	1.23
57 0367	10.97	43.35	22.57	25.63	0.40	72	0.28
57 0374	13.17	66.66	52.08	43.97	0.31	254	1.69
57 0376	34.43	34.8	50	39.74	0.33	53	0.33
57 0401	1	41.88	32	24.96	0.38	56	0.22
57 0429	12.07	33.33	50	31.80	0.37	954	4.85
57 0441	34.43	66.66	52.08	51.06	0.15	249	2.40
57 0548	68.86	34.8	50	51.22	0.22	88	0.89
57 0561	34.33	66.66	27.08	42.69	0.29	35	0.30
57 0570	34.04	35.19	44.18	37.80	0.39	26	0.20
57 0575	34.33	10.41	35.61	26.78	0.24	273	1.49
57 0598	33.62	41.88	32	35.83	0.22	108	0.83
57 0599	35.53	10.41	35.61	27.18	0.23	82	0.49
57 0606	1.1	34.8	50	28.63	0.43	87	0.56
57 0607	35.53	66.66	42.65	48.28	0.22	8	0.09
57 0614	35.14	43.74	28.26	35.71	0.19	356	2.92
57 0615	12.78	43.74	28.26	28.26	0.24	348	2.46
57 0619	11.68	99.99	27.08	46.25	0.26	461	5.85
57 0624	14.27	36.66	19.18	23.37	0.39	190	1.39
57 0626	13.17	34.8	40.57	29.51	0.22	2	0.02
57 0648	35.4	33.33	75	47.91	0.23	230	3.68
57 0673	34.04	43.74	12.69	30.16	0.22	328	3.58
57 0675	4.27	75.21	9.08	29.52	0.31	243	2.94
57 0678	3.3	75.21	9.08	29.20	0.32	138	1.84
57 0681	4.27	43.35	7	18.21	0.28	167	1.48
57 0683	4.27	11.88	10.61	8.92	0.28	204	0.96
57 0684	3.3	11.88	26.18	13.79	0.30	149	1.22
57 0709	35.4	33.33	65.57	44.77	0.21	109	3.18
57 0711	99.99	1.47	50	50.49	0.23	188	6.66
57 0758	10.97	68.13	25	34.70	0.55	100	2.80
57 0762	10.97	41.88	59.08	37.31	0.33	45	1.48
57 0764	68.44	1.47	65.57	45.16	0.24	343	14.17
57 0772	34.04	66.66	27.08	42.59	0.26	212	12.04
57 0848	34.43	34.8	50	39.74	0.26	387	28.59
57 0872	13.17	8.55	57	26.24	0.00	0	0.00
57 0922	1.81	33.33	77.08	37.41	0.15	13	3.22
57 0927	13.17	33.33	50	32.17	0.75	10	2.33
57 1000	33.33	35.19	28.61	32.38	0.51	6	1.52
57 1001	34.04	3.33	28.61	21.99	0.29	122	21.99

Figure 1. Bridges total risk scores.

Conclusions and Recommendations

- **Implementation of alternative surrogate safety measures.** By employing various surrogate safety measures, it is anticipated that the model development process can be enhanced, allowing a more comprehensive understanding of the factors influencing bicyclist speeds.
- **Examination of a larger bridge sample.** Utilizing a more extensive sample of bridges in the analysis is expected to yield more interpretable and robust results. The inclusion of a larger dataset will enable us to better identify trends, patterns, and relationships between variables, consequently improving the validity and reliability of the developed models.
- **Exploration of alternative modeling techniques.** To increase the predictive accuracy and generalizability of the findings, it is recommended that future studies investigate other statistical methodologies. Techniques such as machine learning algorithms or generalized linear mixed models may provide a more sophisticated approach to understanding the complex interactions between predictor variables and bicyclist speeds on bridges.

Additional Products

Education and Workforce Development Products

Technology Transfer Products

- The team is currently working on a journal paper.
- The project team is planning to present the work to Caltrans

Data Products

- Link to Dataset – https://doi.org/10.15787/VTT1/2PQA0O
- Project Description – The goal of this project was to identify the locations withs the greatest requirements for bicycle enhancement initiatives.
- Data Scope – The data from multiple sources were compiled to create a data table in CSV format. Total number of observations for this table is 50, with a total of 239 variables (i.e., columns).
- Data Specification – A detailed description of each variable in the dataset can be found in Appendix I.
- Citation Metadata:
  - Title of datasets: “SafeD-SDSU-06-010-Data.csv”
  - Author list with researcher ORCIDs

References

1. Bamwesigye, D., and P. Hlavackova. Analysis of Sustainable Transport for Smart Cities. *Sustainability*, Vol. 11, No. 7, 2019, p. 2140. https://doi.org/10.3390/su11072140.

2. Hafizyar, R., K. M. Shinwaray, and M. Ali Mosaberpanah. Sustainable and Green Transportation for Best Quality of Life: A Case Study in Kabul, Afghanistan. 2022, pp. 225–238. https://doi.org/10.1061/9780784484340.021.

3. Nematchoua, M., C. Deuse, M. Cools, and S. Reiter. Evaluation of the Potential of Classic and Electric Bicycle Commuting as an Impetus for the Transition towards Environmentally Sustainable Cities: A Case Study of the University Campuses in Liege, Belgium. *Renewable and Sustainable Energy Reviews*, Vol. 119, 2020, p. 109544. https://doi.org/10.1016/j.rser.2019.109544.

4. Saad, M., M. Abdel-Aty, J. Lee, and Q. Cai. Bicycle Safety Analysis at Intersections from Crowdsourced Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2673, No. 4, 2019, pp. 1–14. https://doi.org/10.1177/0361198119836764.

5. Bicycle Safety. *NHTSA*. https://www.nhtsa.gov/book/countermeasures/countermeasures-work/bicycle-safety. Accessed Mar. 14, 2023.

6. FARS Encyclopedia. https://www-fars.nhtsa.dot.gov/Main/index.aspx. Accessed Apr. 14, 2023.

7. Barajas, J. M. Not All Crashes Are Created Equal: Associations between the Built Environment and Disparities in Bicycle Collisions. *Journal of Transport and Land Use*, Vol. 11, No. 1, 2018. https://doi.org/10.5198/jtlu.2018.1145.

8. Durkin, M. S., L. L. Davidson, L. Kuhn, P. O'Connor, and B. Barlow. Low-Income Neighborhoods and the Risk of Severe Pediatric Injury: A Small-Area Analysis in Northern Manhattan. *American Journal of Public Health*, Vol. 84, No. 4, 1994, pp. 587–592. https://doi.org/10.2105/AJPH.84.4.587.

9. Lindsey, G., T. Tao, J. Wang, and J. Cao. *Pedestrian and Bicycle Crash Risk and Equity: Implications for Street Improvement Projects*. 2019.

10. Wang, Y., C. M. Monsere, C. Chen, and H. Wang. Development of a Crash Risk-Scoring Tool for Pedestrian and Bicycle Projects in Oregon. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2672, No. 32, 2018, pp. 30–39. https://doi.org/10.1177/0361198118794285.

11. Hasani, M., A. Jahangiri, I. N. Sener, S. Munira, J. M. Owens, B. Appleyard, S. Ryan, S. M. Turner, and S. Ghanipoor Machiani. Identifying High-Risk Intersections for Walking and Bicycling Using Multiple Data Sources in the City of San Diego. *Journal of Advanced Transportation*, Vol. 2019, 2019, p. e9072358. https://doi.org/10.1155/2019/9072358.

12. Lee, K., and I. N. Sener. Strava Metro Data for Bicycle Monitoring: A Literature Review. *Transport Reviews*, Vol. 41, No. 1, 2021, pp. 27–47. https://doi.org/10.1080/01441647.2020.1798558.

13. Beck, B., M. Winters, T. Nelson, C. Pettit, S. Z. Leao, M. Saberi, J. Thompson, S. Seneviratne, K. Nice, and M. Stevenson. Developing Urban Biking Typologies: Quantifying the Complex Interactions of Bicycle Ridership, Bicycle Network and Built Environment Characteristics. *Environment and Planning B: Urban Analytics and City Science*, Vol. 50, No. 1, 2023, pp. 7–23. https://doi.org/10.1177/23998083221100827.

14. Lee, K., and I. N. Sener. Emerging Data for Pedestrian and Bicycle Monitoring: Sources and Applications. *Transportation Research Interdisciplinary Perspectives*, Vol. 4, 2020, p. 100095. https://doi.org/10.1016/j.trip.2020.100095.

15. Nelson, T., A. Roy, C. Ferster, J. Fischer, V. Brum-Bastos, K. Laberee, H. Yu, and M. Winters. Generalized Model for Mapping Bicycle Ridership with Crowdsourced Data. *Transportation Research Part C: Emerging Technologies*, Vol. 125, 2021, p. 102981. https://doi.org/10.1016/j.trc.2021.102981.

16. Nelson, T. A., C. Ferster, A. Roy, and M. Winters. Bicycle Streetscapes: A Data Driven Approach to Mapping Streets Based on Bicycle Usage. *International Journal of Sustainable Transportation*, Vol. 0, No. 0, 2022, pp. 1–11. https://doi.org/10.1080/15568318.2022.2121670.

17. Roy, A., T. A. Nelson, A. S. Fotheringham, and M. Winters. Correcting Bias in Crowdsourced Data to Map Bicycle Ridership of All Bicyclists. *Urban Science*, Vol. 3, No. 2, 2019, p. 62. https://doi.org/10.3390/urbansci3020062.

18. Nordback, K., S. Kothuri, T. Petritsch, P. McLeod, E. Rose, H. Twaddell, ICF International (Firm), Portland State University, and Inc. Sprinkle Consulting. *Exploring Pedestrian Counting Procedures*. Publication FHWA-HPL-16-026. 2016.

19. Appiah, O., E. Quayson, and E. Opoku. Ultrasonic Sensor Based Traffic Information Acquisition System; a Cheaper Alternative for ITS Application in Developing Countries. *Scientific African*, Vol. 9, 2020, p. e00487. https://doi.org/10.1016/j.sciaf.2020.e00487.

20. Divatankar, S., U. N. Hivarkar, and A. D. Shaligram. Survey and Comparative Study of Various Approaches to Monitor the Road Traffic. https://papers.ssrn.com/abstract=3856100. Accessed Mar. 23, 2023.

21. Dill, J. Bicycling for Transportation and Health: The Role of Infrastructure. *Journal of Public Health Policy*, Vol. 30, No. S1, 2009, pp. S95–S110. https://doi.org/10.1057/jphp.2008.56.

22. Ryus, P., E. Ferguson, K. M. Laustsen, R. J. Schneider, F. R. Proulx, T. Hull, L. Miranda-Moreno, National Cooperative Highway Research Program, Transportation Research Board, and National Academies of Sciences, Engineering, and Medicine. *Guidebook on Pedestrian and Bicycle Volume Data Collection*. Transportation Research Board, Washington, D.C., 2014.

23. Berdica, K. An Introduction to Road Vulnerability: What Has Been Done, Is Done and Should Be Done. *Transport Policy*, Vol. 9, No. 2, 2002, pp. 117–127. https://doi.org/10.1016/S0967-070X(02)00011-2.

24. Carter, D. L., W. W. Hunter, C. V. Zegeer, J. R. Stewart, H. F. Huang, and University of North Carolina (System). Highway Safety Research Center. Pedestrian and Bicycle Information Center. *Pedestrian and Bicyclist Intersection Safety Indices: Final Report*. Publication FHWA-HRT-06-125. 2006.

25. Carter, D. L., W. W. Hunter, C. V. Zegeer, J. R. Stewart, and H. Huang. Bicyclist Intersection Safety Index. *Transportation Research Record*, Vol. 2031, No. 1, 2007, pp. 18–24. https://doi.org/10.3141/2031-03.

26. Harwood, D. W., K. M. Bauer, D. K. Gilmore, R. Souleyrette, and Z. N. Hans. Validation of U.S. Road Assessment Program Star Rating Protocol: Application to Safety Management of U.S. Roads. *Transportation Research Record*, Vol. 2147, No. 1, 2010, pp. 33–41. https://doi.org/10.3141/2147-05.

27. Lagerwey, P. A., M. J. Hintze, J. B. Elliott, J. L. Toole, and R. J. Schneider. Pedestrian and Bicycle Transportation Along Existing Roads—ActiveTrans Priority Tool Guidebook. *NCHRP Report*, No. 803, 2015.

28. Hintze, M., and J. Elliott. A Model Methodology for Prioritizing Pedestrian and Bicycle Improvements Along Existing Roads. Presented at the Transportation Research Board 94th Annual MeetingTransportation Research Board, 2015.

29. Xaykongsa, A. *AADT Estimation Models and Analytical Comparison of Pedestrian Safety Risk Evaluation Methods for Signalized Intersections*. Master Thesis. University of Waterloo, 2021.

30. Furth, P. G., M. C. Mekuria, and H. Nixon. Network Connectivity for Low-Stress Bicycling. *Transportation Research Record*, Vol. 2587, No. 1, 2016, pp. 41–49. https://doi.org/10.3141/2587-06.

31. Mekuria, M., P. Furth, and H. Nixon. Low-Stress Bicycling and Network Connectivity. *Mineta Transportation Institute Publications*, 2012.

32. Mekuria, M. C. Bicycle Connectivity and Safety Model. Presented at the Transportation Research Board 93rd Annual MeetingTransportation Research Board, 2014.

33. Boufous, S., L. de Rome, T. Senserrick, and R. Ivers. Risk Factors for Severe Injury in Cyclists Involved in Traffic Crashes in Victoria, Australia. *Accident; Analysis and Prevention*, Vol. 49, 2012, pp. 404–409. https://doi.org/10.1016/j.aap.2012.03.011.

34. Eluru, N., C. R. Bhat, and D. A. Hensher. A Mixed Generalized Ordered Response Model for Examining Pedestrian and Bicyclist Injury Severity Level in Traffic Crashes. *Accident Analysis & Prevention*, Vol. 40, No. 3, 2008, pp. 1033–1054. https://doi.org/10.1016/j.aap.2007.11.010.

35. Kim, J.-K., S. Kim, G. F. Ulfarsson, and L. A. Porrello. Bicyclist Injury Severities in Bicycle-Motor Vehicle Accidents. *Accident; Analysis and Prevention*, Vol. 39, No. 2, 2007, pp. 238–251. https://doi.org/10.1016/j.aap.2006.07.002.

36. Lenguerrand, E., J. L. Martin, and B. Laumon. Modelling the Hierarchical Structure of Road Crash Data—Application to Severity Analysis. *Accident Analysis & Prevention*, Vol. 38, No. 1, 2006, pp. 43–53. https://doi.org/10.1016/j.aap.2005.06.021.

37. Pai, C.-W. Overtaking, Rear-End, and Door Crashes Involving Bicycles: An Empirical Investigation. *Accident; Analysis and Prevention*, Vol. 43, No. 3, 2011, pp. 1228–1235. https://doi.org/10.1016/j.aap.2011.01.004.

38. Parkin, J., M. Wardman, and M. Page. Models of Perceived Cycling Risk and Route Acceptability. *Accident; Analysis and Prevention*, Vol. 39, No. 2, 2007, pp. 364–371. https://doi.org/10.1016/j.aap.2006.08.007.

39. Schepers, P., and B. den Brinker. What Do Cyclists Need to See to Avoid Single-Bicycle Crashes? *Ergonomics*, Vol. 54, No. 4, 2011, pp. 315–327. https://doi.org/10.1080/00140139.2011.558633.

40. Franken, R., H. Bekhuis, and J. Tolsma. Kudos Make You Run! How Runners Influence Each Other on the Online Social Network Strava. *Social Networks*, Vol. 72, 2023, pp. 151–164. https://doi.org/10.1016/j.socnet.2022.10.001.

41. Hochmair, H. H., E. Bardin, and A. Ahmouda. Estimating Bicycle Trip Volume for Miami-Dade County from Strava Tracking Data. *Journal of Transport Geography*, Vol. 75, 2019, pp. 58–69. https://doi.org/10.1016/j.jtrangeo.2019.01.013.

42. Abegaz, T., and S. Gebremedhin. Magnitude of Road Traffic Accident Related Injuries and Fatalities in Ethiopia. *PLOS ONE*, Vol. 14, No. 1, 2019, p. e0202240. https://doi.org/10.1371/journal.pone.0202240.

43. Bhatia, R., and M. Wier. "Safety in Numbers" Re-Examined: Can We Make Valid or Practical Inferences from Available Evidence? *Accident Analysis & Prevention*, Vol. 43, No. 1, 2011, pp. 235–240. https://doi.org/10.1016/j.aap.2010.08.015.

44. Ferenchak, N. N., and W. E. Marshall. Equity Analysis of Proactively- vs. Reactively-Identified Traffic Safety Issues. *Transportation Research Record*, Vol. 2673, No. 7, 2019, pp. 596–606. https://doi.org/10.1177/0361198119841296.

45. Ugan, J., M. Abdel-Aty, Q. Cai, N. Mahmoud, and M. Al-Omari. Effect of Various Speed Management Strategies on Bicycle Crashes for Urban Roads in Central Florida. *Transportation Research Record*, Vol. 2676, No. 1, 2022, pp. 544–555. https://doi.org/10.1177/03611981211036681.

46. Hosford, K., and M. Winters. Quantifying the Bicycle Share Gender Gap. *Findings*, 2019. https://doi.org/10.32866/10802.

47. Knoblauch, R. L., R. F. Seifert, N. B. Murphy, Center for Applied Research, and Inc. The Media Network. *The Pedestrian and Bicyclist Highway Safety Problem as It Relates to the Hispanic Population in the United States*. Publication DTFH61-03-P-00324. 2004.

48. Governors Highway Safety Association. An Analysis of Traffic Fatalities by Race and Ethnicity. , 2021.

49. Johnsson, C., A. Laureshyn, and C. Dágostino. Validation of Surrogate Measures of Safety with a Focus on Bicyclist–Motor Vehicle Interactions. *Accident Analysis & Prevention*, Vol. 153, 2021, p. 106037. https://doi.org/10.1016/j.aap.2021.106037.

50. Li, P., M. Abdel-Aty, and J. Yuan. Using Bus Critical Driving Events as Surrogate Safety Measures for Pedestrian and Bicycle Crashes Based on GPS Trajectory Data. *Accident Analysis & Prevention*, Vol. 150, 2021, p. 105924. https://doi.org/10.1016/j.aap.2020.105924.

51. Akhavian, R., A. Jahangiri, F. Shahnavaz, and S. Salehipour. *A Holistic Work Zone Safety Alert System through Automated Video and Smartphone Sensor Data Analysis*. Safe-D National UTC, 2022.

52. Aarts, L., and I. van Schagen. Driving Speed and the Risk of Road Crashes: A Review. *Accident Analysis & Prevention*, Vol. 38, No. 2, 2006, pp. 215–224. https://doi.org/10.1016/j.aap.2005.07.004.

53. Choudhary, P., M. Imprialou, N. R. Velaga, and A. Choudhary. Impacts of Speed Variations on Freeway Crashes by Severity and Vehicle Type. *Accident Analysis & Prevention*, Vol. 121, 2018, pp. 213–222. https://doi.org/10.1016/j.aap.2018.09.015.

54. Thompson, S. R., C. M. Monsere, M. Figliozzi, P. Koonce, and G. Obery. Bicycle-Specific Traffic Signals: Results from a State-of-the-Practice Review. *Transportation Research Record*, Vol. 2387, No. 1, 2013, pp. 1–9. https://doi.org/10.3141/2387-01.

55. Sanguinetti, A., E. Alston-Stepnitz, M. Ruhl, N. Dessouky, and A. Broaddus. Equipping Active Travel Advocates with Digital Mobility Data and Tools: An Evaluation of a US Trial Program. *Active Travel Studies*, Vol. 3, No. 1, 2023. https://doi.org/10.16997/ats.1198.

56. Yang, H., M. Cetin, Q. Ma, and Virginia Transportation Research Council (VTRC). *Guidelines for Using StreetLight Data for Planning Tasks*. Publication FHWA/VTRC 20-R23. 2020.

57. Lee, C. A Game Changer in the Making? Lessons from States Advancing Environmental Justice through Mapping and Cumulative Impact Strategies. *Environmental Law Reporter*, Vol. 50, 2020, p. 10203.

58. Mullen, H., K. Whyte, and R. Holifield. Indigenous Peoples and the Justice40 Screening Tool: Lessons from EJSCREEN. *Environmental Justice*, 2023. https://doi.org/10.1089/env.2022.0045.

59. Driver, A., C. Mehdizadeh, S. Bara-Garcia, C. Bodenreider, J. Lewis, and S. Wilson. Utilization of the Maryland Environmental Justice Screening Tool: A Bladensburg,

Maryland Case Study. *International Journal of Environmental Research and Public Health*, Vol. 16, No. 3, 2019, p. 348. https://doi.org/10.3390/ijerph16030348.

60. Mullen, H. *Indigenous Environmental Justice and Screening Tools: Lessons Learned from EJSCREEN and Paths Forward for the Climate and Economic Justice Screening Tool*. Thesis. 2022.

61. Cox, D. R. The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 20, No. 2, 1958, pp. 215–232. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x.

62. Domínguez-Almendros, S., N. Benítez-Parejo, and A. R. Gonzalez-Ramirez. Logistic Regression Models. *Allergologia et Immunopathologia*, Vol. 39, No. 5, 2011, pp. 295–305. https://doi.org/10.1016/j.aller.2011.05.002.

63. Al-Ghamdi, A. S. Using Logistic Regression to Estimate the Influence of Accident Factors on Accident Severity. *Accident Analysis & Prevention*, Vol. 34, No. 6, 2002, pp. 729–741. https://doi.org/10.1016/S0001-4575(01)00073-2.

64. Bland, J. M., and D. G. Altman. The Odds Ratio. *BMJ*, Vol. 320, No. 7247, 2000, p. 1468. https://doi.org/10.1136/bmj.320.7247.1468.

65. Monsere, C., H. Wang, Y. Wang, C. Chen, Portland State University. Dept. of Civil & Environmental Engineering, and Oregon State University. School of Civil and Construction Engineering. *Risk Factors for Pedestrian and Bicycle Crashes*. Publication FHWA-OR-RD-17-13. 2017.

66. Lin, Y., and R. Li. Real-Time Traffic Accidents Post-Impact Prediction: Based on Crowdsourcing Data. *Accident Analysis & Prevention*, Vol. 145, 2020, p. 105696. https://doi.org/10.1016/j.aap.2020.105696.

67. Thompson, M. J., and F. P. Rivara. Bicycle-Related Injuries. *American Family Physician*, Vol. 63, No. 10, 2001, pp. 2007–2015.

68. Baffour, B., T. King, and P. Valente. The Modern Census: Evolution, Examples and Evaluation. *International Statistical Review*, Vol. 81, No. 3, 2013, pp. 407–425. https://doi.org/10.1111/insr.12036.

69. Wang, H., Y. Wang, M. B. Lowry, C. Chen, and Z. Pu. Bicycle Safety Analysis: Crowdsourcing Bicycle Travel Data to Estimate Risk Exposure and Create Safety Performance Functions. 2016.

70. Wang, X., X. Zheng, Q. Zhang, T. Wang, and D. Shen. Crowdsourcing in ITS: The State of the Work and the Networking. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17, No. 6, 2016, pp. 1596–1605. https://doi.org/10.1109/TITS.2015.2513086.

71. Misra, A., A. Gooze, K. Watkins, M. Asad, and C. A. Le Dantec. Crowdsourcing and Its Application to Transportation Data Collection and Management. *Transportation Research Record*, Vol. 2414, No. 1, 2014, pp. 1–8. https://doi.org/10.3141/2414-01.

72. Ali, A., N. Ayub, M. Shiraz, N. Ullah, A. Gani, and M. A. Qureshi. Traffic Efficiency Models for Urban Traffic Management Using Mobile Crowd Sensing: A Survey. *Sustainability*, Vol. 13, No. 23, 2021, p. 13068. https://doi.org/10.3390/su132313068.

73. Wan, X., H. Ghazzai, and Y. Massoud. Mobile Crowdsourcing for Intelligent Transportation Systems: Real-Time Navigation in Urban Areas. *IEEE Access*, Vol. 7, 2019, pp. 136995–137009. https://doi.org/10.1109/ACCESS.2019.2942282.

74. Nelson, T., C. Ferster, K. Laberee, D. Fuller, and M. Winters. Crowdsourced Data for Bicycling Research and Practice. *Transport Reviews*, Vol. 41, No. 1, 2021, pp. 97–114. https://doi.org/10.1080/01441647.2020.1806943.

75. Seidl, D. E., P. Jankowski, and A. Nara. An Empirical Test of Household Identification Risk in Geomasked Maps. *Cartography and Geographic Information Science*, Vol. 46, No. 6, 2019, pp. 475–488. https://doi.org/10.1080/15230406.2018.1544932.

76. Seidl, D. E., P. Jankowski, K. C. Clarke, and A. Nara. Please Enter Your Home Location: Geoprivacy Attitudes and Personal Location Masking Strategies of Internet Users. *Annals of the American Association of Geographers*, Vol. 110, No. 3, 2020, pp. 586–605. https://doi.org/10.1080/24694452.2019.1654843.

77. Schewel, L., S. Co, C. Willoughby, L. Yan, N. Clarke, J. Wergin, and StreetLight Data. *Non-Traditional Methods to Obtain Annual Average Daily Traffic (AADT)*. Publication FHWA-PL-21-030. 2021.

78. Goodspeed, R., M. Yuan, A. Krusniak, and T. Bills. Assessing the Value of New Big Data Sources for Transportation Planning: Benton Harbor, Michigan Case Study. In *Urban Informatics and Future Cities* (S. C. M. Geertman, C. Pettit, R. Goodspeed, and A. Staffans, eds.), Springer International Publishing, Cham, pp. 127–150.

79. Fish, J. K., S. E. Young, A. Wilson, B. Borlaug, and National Renewable Energy Laboratory (NREL) (U.S.). *Validation of Non-Traditional Approaches to Annual Average Daily Traffic (AADT) Volume Estimation*. Publication FHWA-PL-21-033. 2021.

80. Goel, R., L. M. T. Garcia, A. Goodman, R. Johnson, R. Aldred, M. Murugesan, S. Brage, K. Bhalla, and J. Woodcock. Estimating City-Level Travel Patterns Using Street Imagery: A Case Study of Using Google Street View in Britain. *PloS One*, Vol. 13, No. 5, 2018, p. e0196521. https://doi.org/10.1371/journal.pone.0196521.

81. Koo, B. W., S. Guhathakurta, and N. Botchwey. How Are Neighborhood and Street-Level Walkability Factors Associated with Walking Behaviors? A Big Data Approach Using Street View Images. *Environment and Behavior*, Vol. 54, No. 1, 2022, pp. 211–241. https://doi.org/10.1177/00139165211014609.

82. Nguyen, Q. C., S. Khanna, P. Dwivedi, D. Huang, Y. Huang, T. Tasdizen, K. D. Brunisholz, F. Li, W. Gorman, T. T. Nguyen, and C. Jiang. Using Google Street View to Examine Associations between Built Environment Characteristics and U.S. Health Outcomes. *Preventive Medicine Reports*, Vol. 14, 2019, p. 100859. https://doi.org/10.1016/j.pmedr.2019.100859.

83. Ki, D., and S. Lee. Analyzing the Effects of Green View Index of Neighborhood Streets on Walking Time Using Google Street View and Deep Learning. *Landscape and Urban Planning*, Vol. 205, 2021, p. 103920. https://doi.org/10.1016/j.landurbplan.2020.103920.

84. Yin, L., Q. Cheng, Z. Wang, and Z. Shao. 'Big Data' for Pedestrian Volume: Exploring the Use of Google Street View Images for Pedestrian Counts. *Applied Geography*, Vol. 63, 2015, pp. 337–345. https://doi.org/10.1016/j.apgeog.2015.07.010.

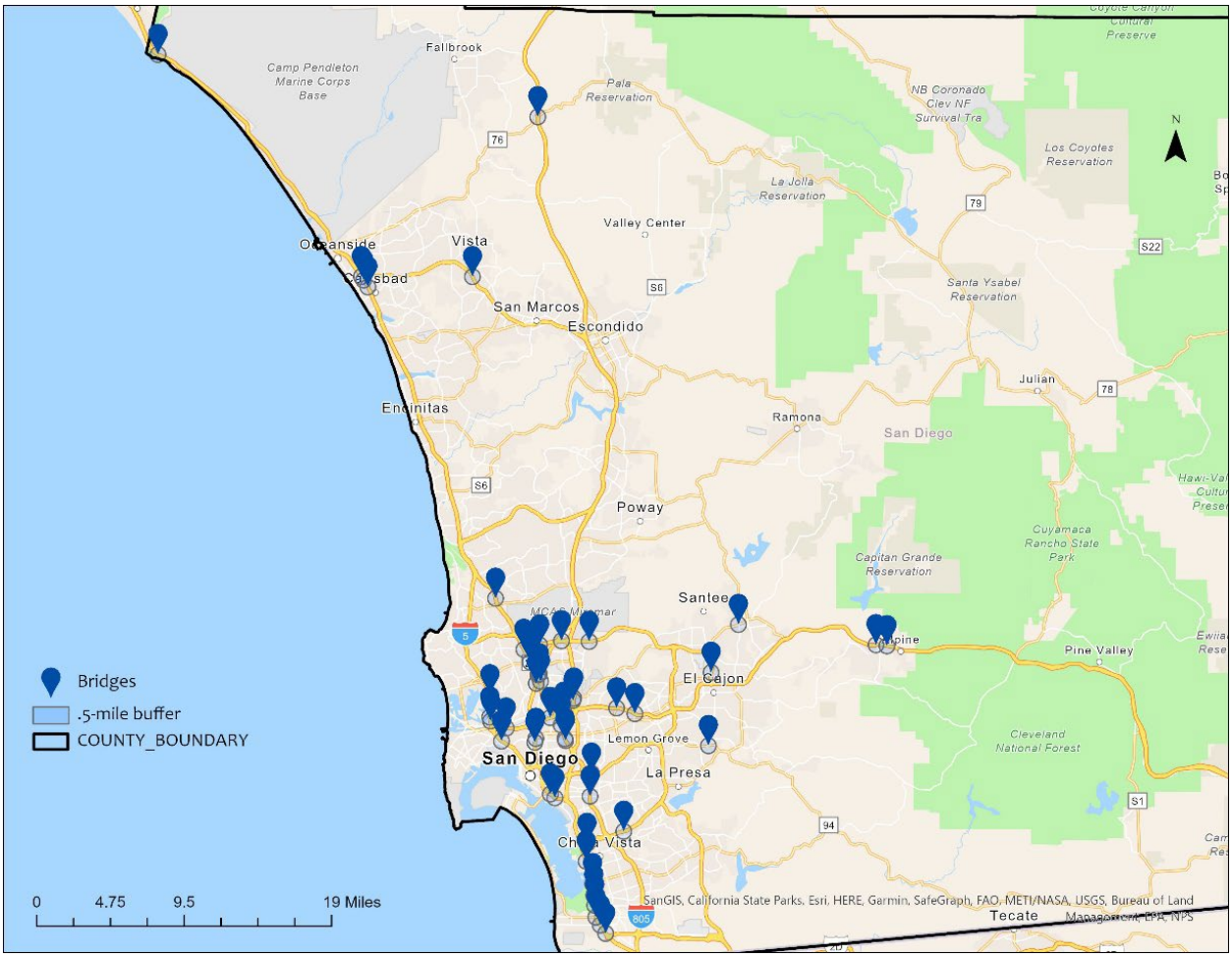
85. Qi, M., and S. Hankey. Using Street View Imagery to Predict Street-Level Particulate Air Pollution. *Environmental Science & Technology*, Vol. 55, No. 4, 2021, pp. 2695–2704. https://doi.org/10.1021/acs.est.0c05572.

86. Cox, D. R. The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, Vol. 20, No. 2, 1958, pp. 215–232. https://doi.org/10.1111/j.2517-6161.1958.tb00292.x.

87. Hoehler, F. K. Logistic Equations in the Analysis of S-Shaped Curves. *Computers in Biology and Medicine*, Vol. 25, No. 3, 1995, pp. 367–371. https://doi.org/10.1016/0010-4825(95)00013-T.

Appendices

Appendix A – Bridge Locations



Appendix B – Discussion on Non-Crowdsourced and Crowdsourced Data

There are several methods of collecting data in transportation engineering that are used to inform planning, design, and operations decisions. There are two common methods of gathering data: Crowdsourcing data and non-crowdsourced data, such as the U.S. Census Bureau, which is a federal agency within the United States responsible for accumulating demographic and economic data, conducting surveys, and collecting other official information (*68*, *69*).

The concept of crowdsourcing data refers to the collection of data using the devices and technologies of individuals or groups (*70*, *71*). Due to the widespread use of smartphones, GPS devices, and other mobile technologies, this type of data collection has become increasingly popular in recent years (*72*). It is possible to gain valuable insights into transportation behavior through crowdsourced data, such as travel patterns, route choices, and mode preferences.

On the other hand, non-crowdsourced data sources are those derived from traditional methods, including the United States Census Bureau, surveys, and official sources. Although more planning and coordination is required for crowdsourced data sources, more comprehensive and detailed information can be obtained from these sources. For instance, census data can assist in planning and investment in transportation by providing information about the demographic characteristics of the population, including age, income, and employment.

There are advantages and disadvantages to crowdsourced data as well as non-crowdsourced data, and they can both be utilized together to provide a more comprehensive view of transportation behavior.

The following are some of the advantages of crowdsourced data:

- Real-time data: Data collected by crowdsourcing can be analyzed in real-time, allowing transportation professionals to respond rapidly to changing road conditions.
- Large sample size: Through crowdsourcing, data can be collected from a large number of people, allowing more comprehensive analyses of travel behaviors. However, it is important to note that the sample size may be limited depending on the crowdsourcing platform and the number of users who participate.
- Cost-effective: Since crowdsourced data eliminates the need for expensive equipment and hardware, it can be a cost-effective way of gathering data (*66*, *70*, *73*).

The following are some disadvantages of crowdsourced data (*14*, *74*):

- Bias: Crowdsourced data may only reflect the behavior of a specific subset of users and may not reflect the entire population.

- Limited data quality control: Since crowdsourced data relies on users to provide accurate information, it is susceptible to errors and inaccuracies.
- Limited information: Certain aspects of travel behavior may not be captured by crowdsourced data as thoroughly as those captured by manual data sources.
- Privacy concerns: Users may publish their cycling data without considering the potential risks to their privacy. As a result, confidential information about individuals and their travel behavior may be inadvertently shared (*75*, *76*).

StreetLight Data Insight

A number of tools are available on the Streetlight Data platform for analyzing data, including:

- Trip analysis: Streetlight Data's trip analysis tool allows users to analyze information such as trip origins and destinations, trip duration, and trip mode to identify patterns and trends in travel behavior.
- Demographic analysis: Using Streetlight Data's demographic analysis tool, users can analyze characteristics of a specific population, such as age, income, and education, in order to gain a greater understanding of its characteristics. Using this information can assist in the planning and investment of transportation infrastructure.
- Traffic analysis: Users can analyze traffic volume and speed data using Streetlight Data's traffic analysis tool in order to understand traffic patterns and identify congestion hotspots (*14*, *56*).

StreetLight Data Advantages

The following are some of the advantages of crowdsourced data:

- Comprehensive data: Streetlight Data provides comprehensive information about travel behavior and demographics, which can be used to inform decision-making regarding transportation investment and planning.
- Large sample size: Streetlight Data can provide data on a large number of people, allowing a more comprehensive understanding of travel patterns.
- Customizable analysis**:** Streetlight Data allows users to customize their analysis of the data in order to answer their specific research questions (*14*, *56*, *77*, *78*)

StreetLight Data Reliability and Validity

In order to determine the reliability and validity of StreetLight Data's location-based data, several factors must be considered, including the quality of the data sources and the accuracy of the machine learning algorithms used to analyze the data (*79*). In StreetLight Data's platform, anonymized location data including GPS and Wi-Fi signals is collected from mobile devices. Data is collected from a number of sources by the company, and data is cleaned and analyzed using a combination of machine learning algorithms and statistical modeling. Furthermore,

StreetLight Data has published several case studies demonstrating the effectiveness of its analytics and data for urban planning and transportation planning.

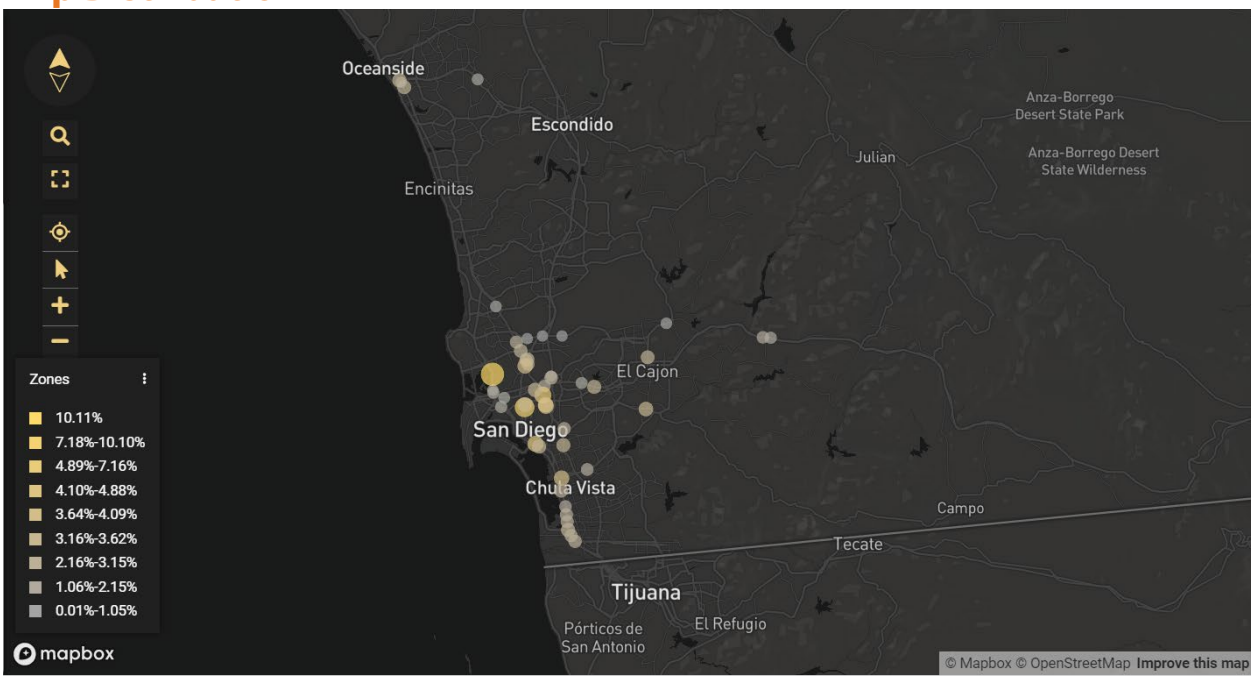
StreetLight Data Limitations

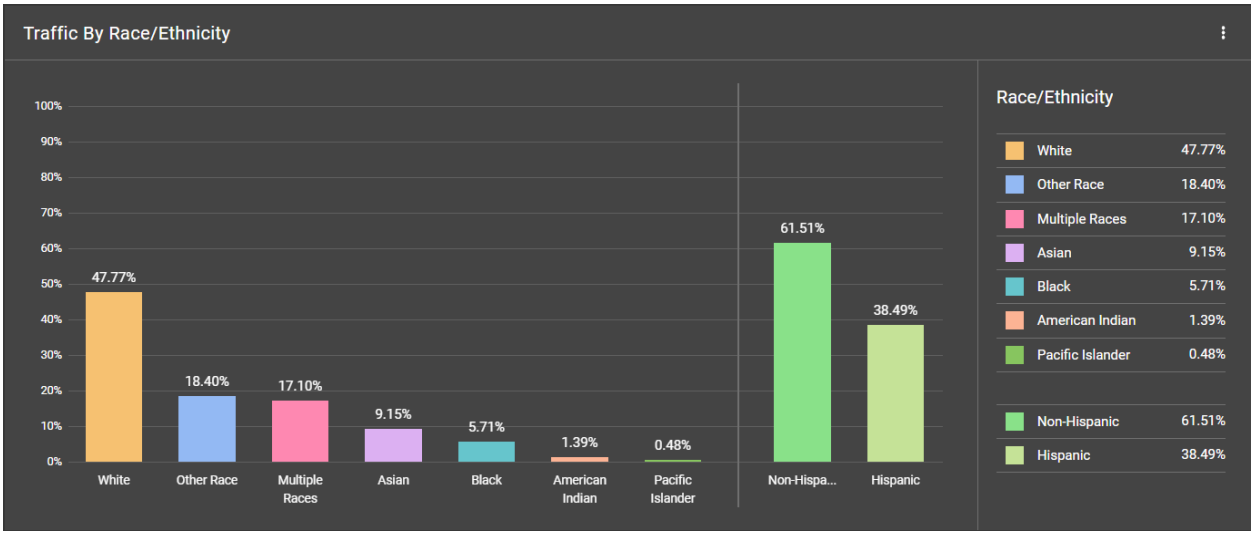
There are some limitations to consider when using Streetlight Data despite its value as a transportation data source (*14*):

1. Limited coverage: Streetlight Data may not be comprehensively covered in each area. It may be limited to areas in which mobile devices and other sources of data are available, resulting in data gaps in certain areas.
2. Data quality: It is important to note that Streetlight Data relies on data from various sources, which can affect its quality and accuracy. The accuracy of the data shall also vary based on the data source and the quality of the underlying data.
3. Privacy concerns: In addition to relying on mobile device and other sources of sensitive data, Streetlight Data may raise privacy concerns. Although Streetlight Data takes steps to ensure the privacy and security of the data, certain data sets may be unavailable due to concerns related to privacy and security.
4. Limited granularity: As Streetlight Data may only capture certain aspects of travel behavior, it may not provide as much detail as manual data sources. There may be limitations to the data's usefulness for certain types of analysis, such as the fact that it may not capture detailed information about trip purposes.
5. Propriety data: As Streetlight Data is a proprietary data source, access may be restricted and a fee may be required. This may limit the availability of the data for certain research projects or planning initiatives.

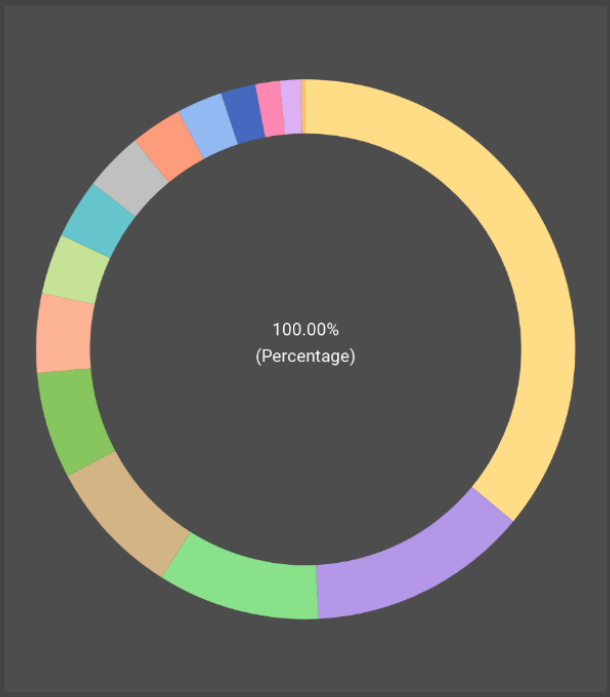
Appendix C – Visualizing Aggregate Variable Values for All Bridges: A Data Display

Trip Distribution



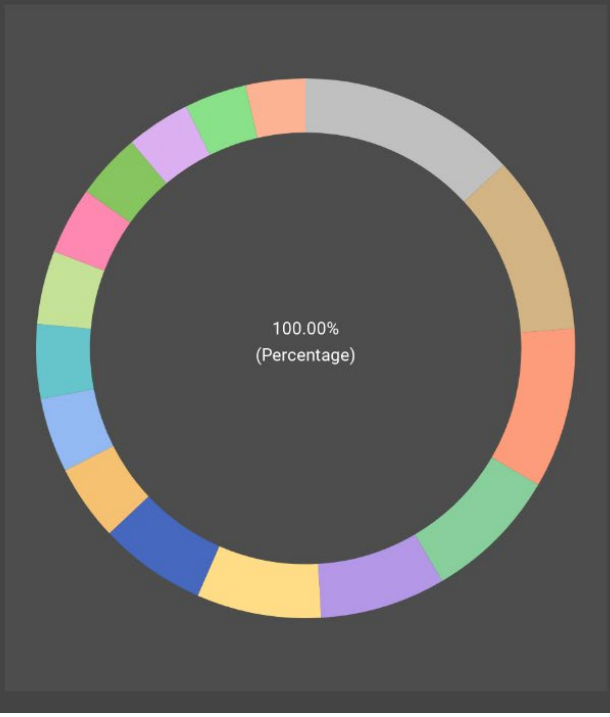


Traffic By Employment Industry & Occupation



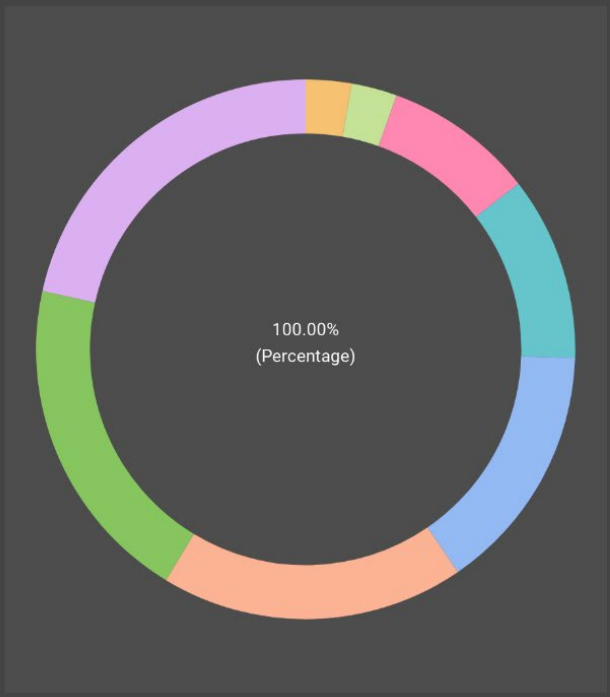
Not employed (Employment Industry)	36.00%
Educational services, and health care and social assistance	13.25%
Professional, scientific, management, and administrative and waste management services	9.72%
Arts, entertainment, and recreation, and accommodation and food services	8.26%
Retail trade	6.37%
Manufacturing	4.74%
Finance and insurance, and real estate and rental and leasing	3.62%
Construction	3.58%
Other services (except public administration)	3.55%
Public administration	3.10%
Transportation, warehousing, utilities	2.75%
Military (Employment Industry)	2.08%
Information	1.46%
Wholesale trade	1.26%
Agriculture, forestry, fishing, hunting, mining	0.27%

Traffic By Household Income



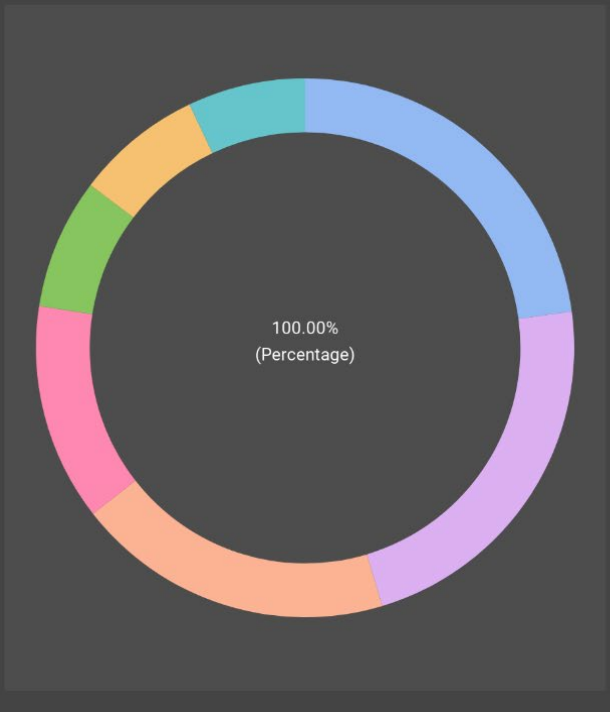
Income 75K to 100K	13.13%
Income 60K to 75K	10.70%
Income 100K to 125K	9.61%
Income More than 200K	8.11%
Income 50K to 60K	7.53%
Income 150K to 200K	7.47%
Income 125K to 150K	6.42%
Income Less than 10K	4.55%
Income 30K to 35K	4.46%
Income 10K to 15K	4.46%
Income 40K to 45K	4.40%
Income 35K to 40K	4.03%
Income 25K to 30K	3.96%
Income 20K to 25K	3.84%
Income 45K to 50K	3.73%
Income 15K to 20K	3.59%

Traffic By Trip Speed (Mph)

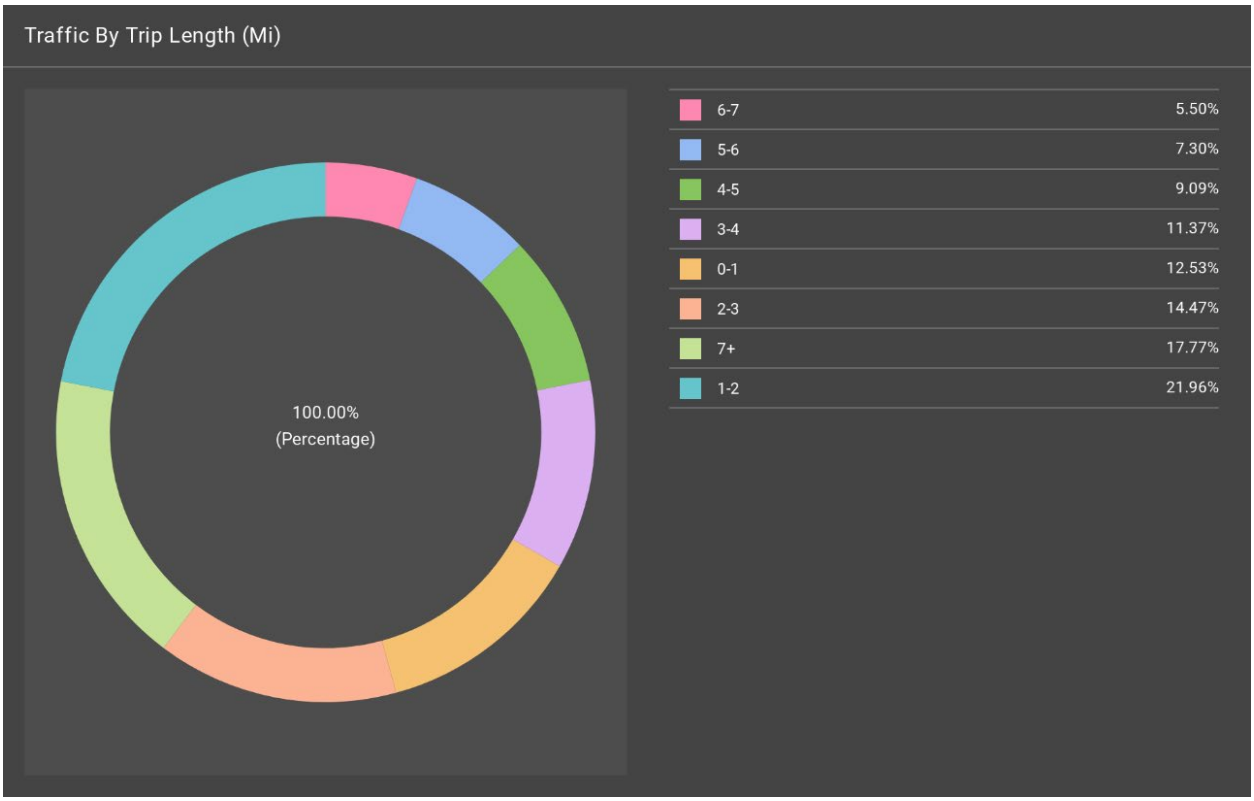


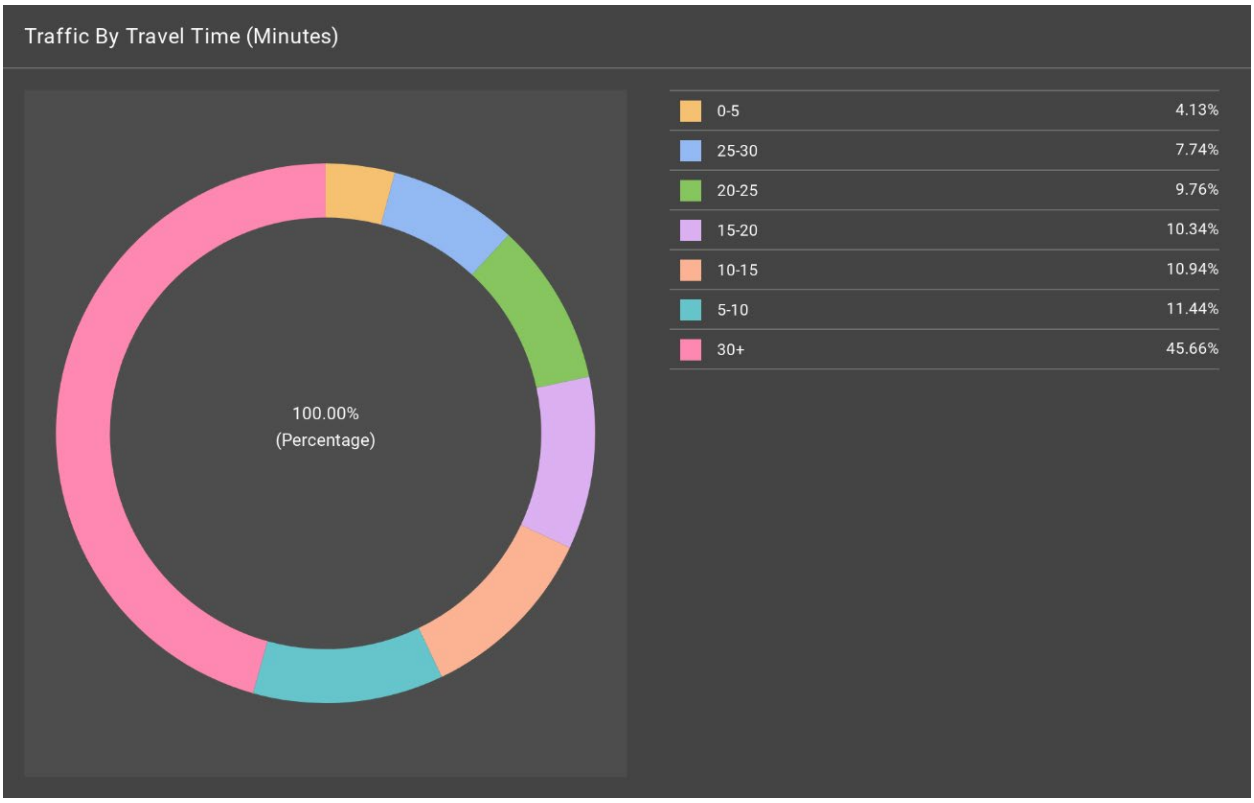
0-2	2.74%
14+	2.74%
12-14	9.01%
2-4	11.02%
10-12	14.94%
4-6	18.21%
8-10	19.81%
6-8	21.53%

Traffic By Education Status



Bachelor's degree	22.85%
Some college, no degree	22.51%
High school graduate	19.07%
Graduate or professional degree	13.04%
Associate's degree	7.84%
Less than 9th grade	7.61%
9th to 12th grade, no diploma	7.07%





Appendix D – Image-Based Data Extraction Example

GSV imagery is produced by seamlessly stitching together a continuous sequence of 360-degree panoramas from numerous overlapping images. Each panorama corresponds to a specific location and the time at which the images were taken. The process of obtaining an image for a given location and direction is fully automated via an Application Programming Interface (API). To use the GSV API, input is required in the form of either geographic coordinates or a unique panorama ID. Additionally, the GSV metadata API can utilize geographic coordinates to provide information such as the year and month of the corresponding panorama, as well as its unique ID (*80*). By specifying a geographic location, the API can retrieve the most recent image captured at that location along with its corresponding metadata.

Previous studies have utilized two primary methods for sampling images, which can be broadly divided into two categories (*81*). The first approach is centered on capturing the built environment at intersections (*82*), while the second approach utilizes multiple images along street segments with fixed distance intervals, such as 20, 50, or 100 meters (*83*). In both methods, 360° panoramic images are commonly used for each image location, although some exceptions have relied on specific portions of images or directions (*84*). Focusing on intersections allows for the measurement of streetscapes at critical nodes of street networks with a smaller number of images covering a given area, but it is limited in capturing the streetscape characteristics that pedestrians experience as they move from one intersection to another. On the other hand, collecting images for the entire length of segments provides a comprehensive view of the streetscapes but requires a larger number of images and more computational resources for image processing. Overall, in this study, GSV images were identified and downloaded using a methodology that combines the two approaches found in literature.

In this study, we employed an approach to extract street-level information on the built and natural environment from Google Street View (GSV) imagery. Briefly, we created a 100m x 100m grid for San Diego county. Where available, we then downloaded panoramic GSV images at the centroid of each grid cell (more than 250000 images). We then processed these images using a previously published deep learning model called the Pyramid Scene Parsing Network (PSPNet). The PSPNet is built on the fully convolutional network architecture and is pre-trained on ADE20K, a database that provides annotated images with 150 categories. For each pixel, the PSPNet predicts the category with the highest probability. We then tabulated metrics for each grid cell based on the processed imagery. The resulting dataset includes a raster with 150 bands (for each feature from PSPNet) at 100m spatial resolution for San Diego county.

The final step is selecting objects that are relevant to the aim of studies from a pool of 150 categories, and then aggregate them to the new categories of variables. To ensure that the selected categories are consistent with the aim of this study, we excluded items such as benches, signboards, or other indoor objects. Overall, we can select objects that are a part of the human-controlled environment. However, landscape features such as trees and grass are included. The objects that we considered for this study can be categorized broadly into seven main categories, presented in Table 1 along with descriptive statistics. Also, Fig1, 2, and 3 are presented the GSV

images location distribution, spatial built environment features, and transport network distribution across the county based on the availability of data, and images.

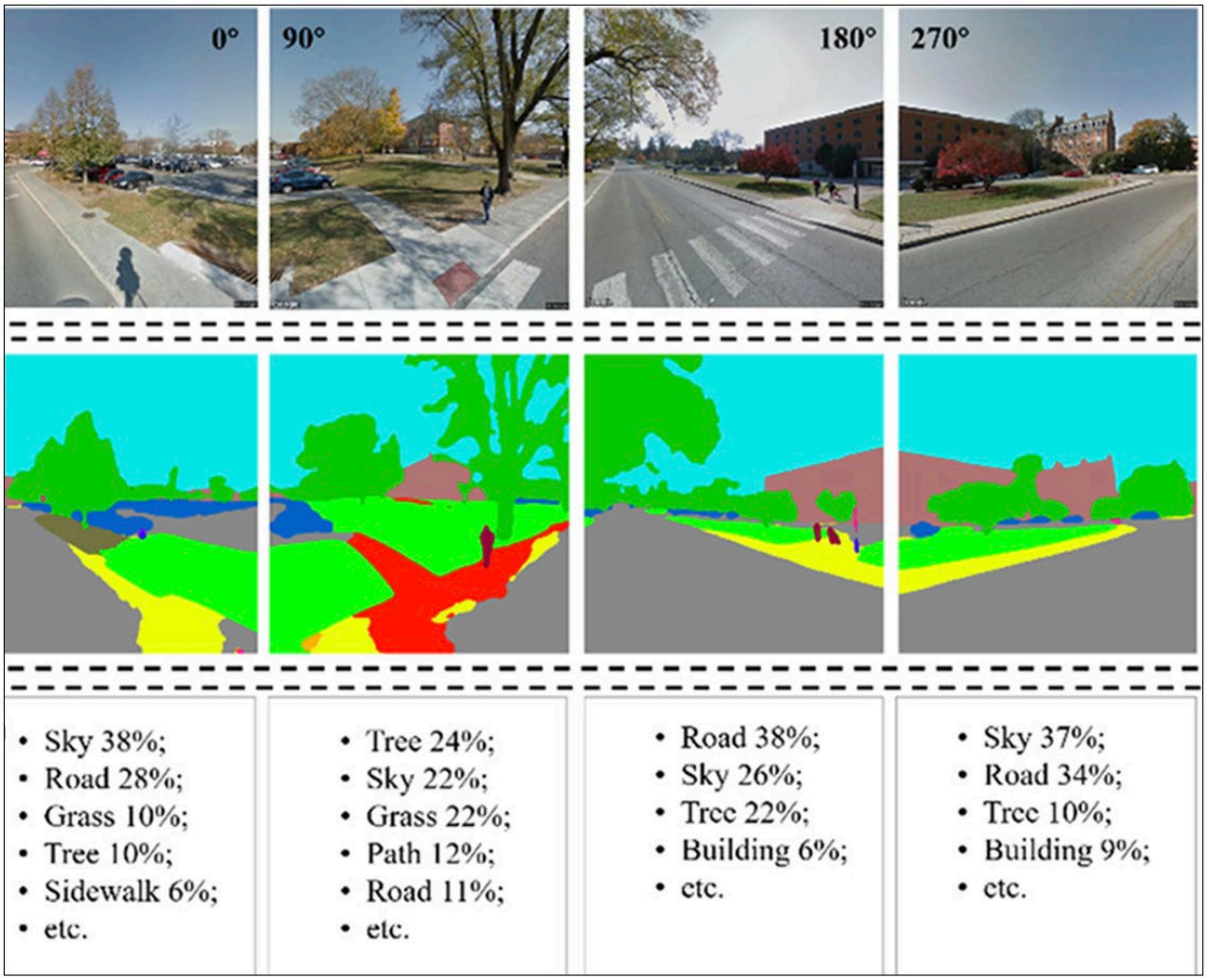


Figure 2. Examples of Google Street View images and their output from PSPNet process. (*85*)

Table 13. Aggregation of GSV Images Features into Seven Main Categories

Categories	Description	Mean	St. dv	Min	Max
Built environment	Building, canopy, house, skyscraper, wall	0.0429	0.000267	0.00	0.6990
Transport network	Bridge, path, road, sidewalk, streetlight, traffic light	0.3210	0.00040	0.00	0.4790
Transport vehicles	Bus, car, minibike, truck, van	0.0161	0.00011	0.00	0.270
Nature	Earth, hill, land, mountain	0.0791	0.08550	0.00	0.6210

Categories	Description	Mean	St. dv	Min	Max
Vegetations	Field, grass, palm, plant, tree	0.1490	0.1060	0.00	0.7160
Water	Lake, river, sea, water, waterfall	0.00072	0.00503	0.00	0.2270
Human	Person, bicycle	0.000094	0.00049	0.00	0.0186

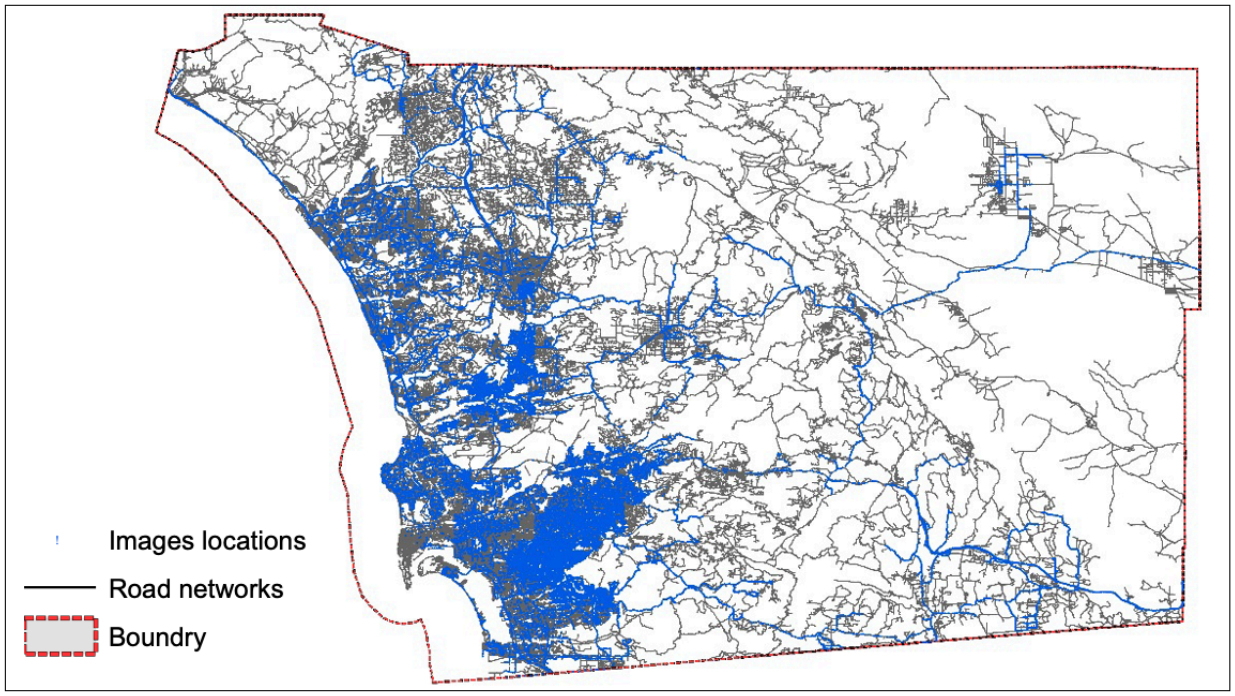


Figure 2. Spatial distribution of images location.

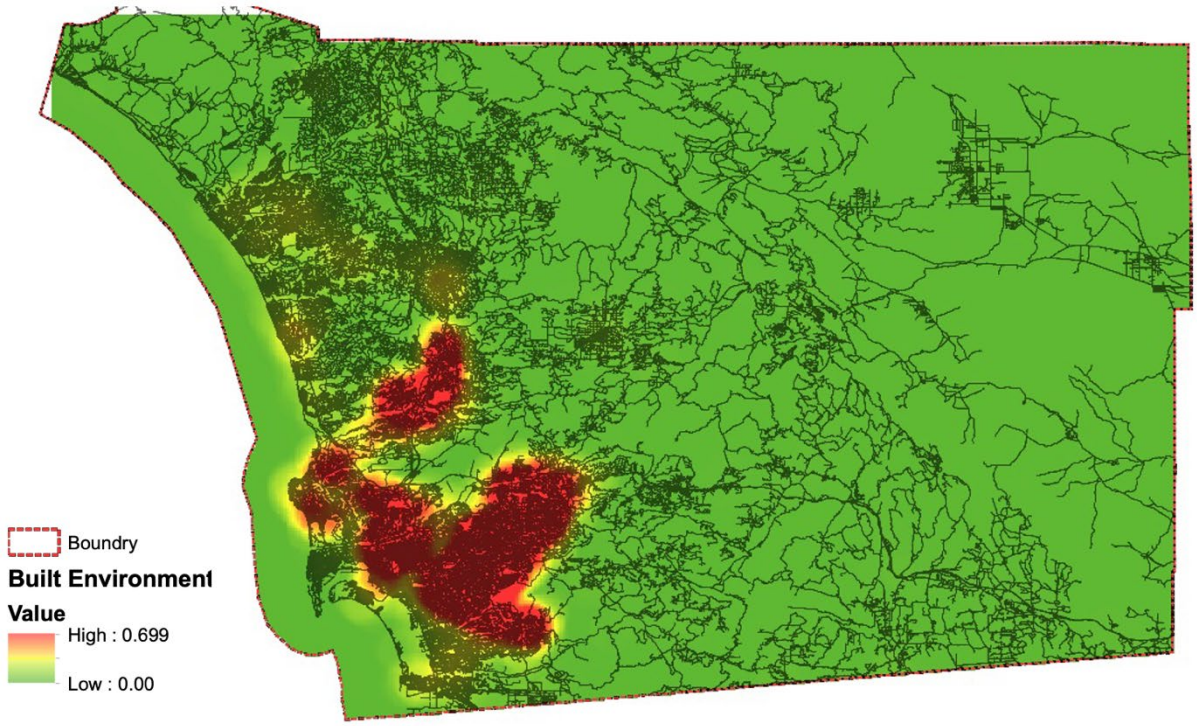


Figure 3. Spatial distribution of built environment feature

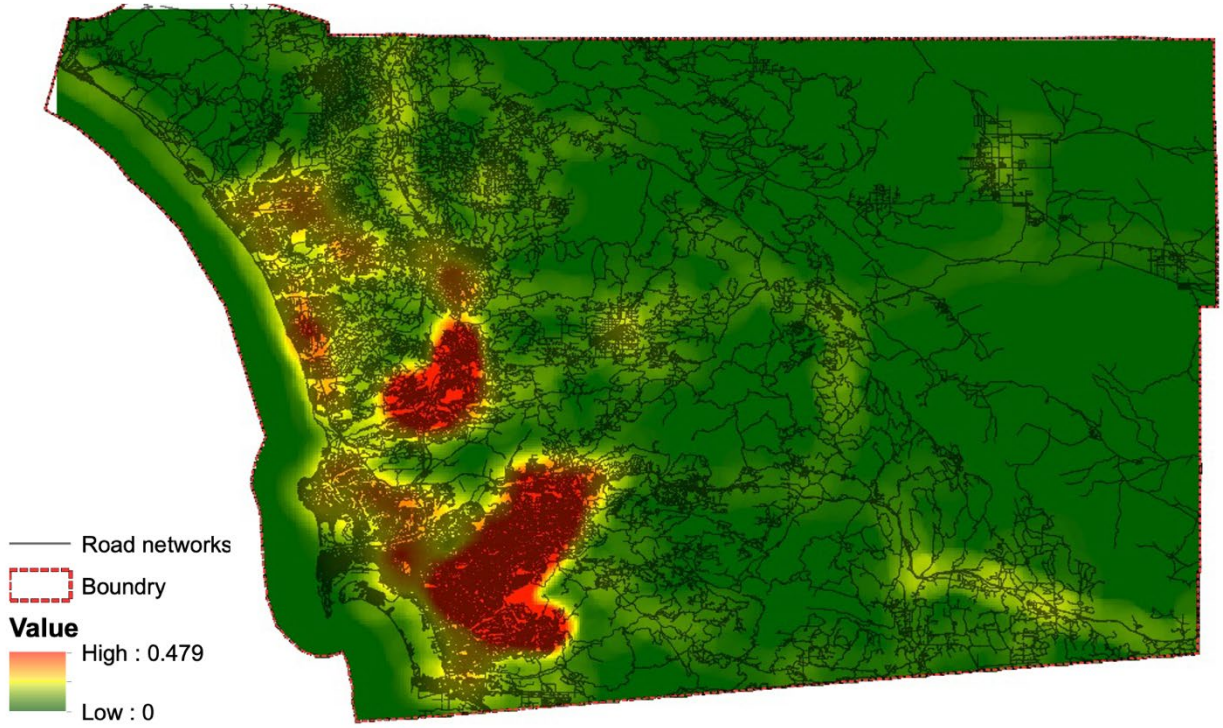


Figure 4. Spatial distribution of transport network

Appendix E – Preliminaries

Logistic Regression

In 1958, the method of logistic regression was introduced by statistician David Cox as a method for modeling the relationship between a binary outcome variable and several predictor variables (*86*). A logistic regression model represents a type of generalized linear model (GLM), which refers to a class of models that are capable of accommodating non-normal distributions of the response variable and non-linear relationships between the predictor and response variables (*62*).

Developing logistic regression was motivated by the need for a method for modeling the probability of an event occurring, such as whether an accident occurs or not. In contrast to linear regression, which is used to model continuous outcome variables, logistic regression is used to model binary outcome variables that can be expressed in one of two ways, namely as zero or one, yes or no, or true or false. In binary classification problems, logistic regression is a popular machine learning algorithm that can be utilized to classify data into two categories (e.g., zero or one, yes or no, true or false).

As part of the logistic regression model, the linear combination of predictor variables are transformed into a probability value between 0 and 1 using a logistic function, also referred to as the sigmoid function. The sigmoid function is an S-shaped curve that has an output value ranging from 0 to 1 (*87*). The Sigmoid function is defined as follows:

$$P(x) = \frac{1}{1+e^{-z}} \tag{1}$$

Where:

P(x): the probability of the outcome variable

z: the linear combination of predictor variables

e: the base of the natural logarithm.

It is possible for the probability of the outcome variable to increase or decrease depending on the values of the predictor variables in the logistic function due to the shape of the logistic function.

Based on the logit scale, logistic regression assumes that the relationship between predictor variables and outcome variables is linear. Logit functions are the natural logarithms of odds ratios, which represent the ratio of the probability of an event occurring to the probability of it not occurring.

It is capable of handling interactions between predictor variables and continuous and categorical predictor variables in logistic regression (*65*). Additionally, the model can be extended to deal with multiclass classification problems using techniques such as one-vs-all regression or SoftMax regression.

$$P(x) = \frac{L}{1+e^{-k(x-x_0)}} \tag{2}$$

Where:

e = the natural logarithm base (2.71828)

x_0 = the x-value of the sigmoid function's midpoint

L = the curve's maximum value, L, equals 1 in the binomial model

k = the steepness of the curve

P(x) = the probability of the dependent variable

Coefficients

In logistic regression, the coefficients (often denoted as "β") represent the relationship between the predictor variables and the outcome variable. Specifically, the coefficients represent the change in the log-odds of the outcome variable for a one-unit increase in the corresponding predictor variable, while holding all other predictor variables constant (*63*).

The log-odds of the outcome variable can be expressed as the linear combination of the predictor variables, plus an intercept term:

$$log(odds) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \tag{3}$$

Where:

β_0 : the intercept term, which represents the log-odds of the outcome variable when all the predictor variables are equal to zero

$\beta_1, \beta_2, \ldots, \beta_k$: the coefficients of the predictor variables x_1, x_2, ..., x_k, which represent the change in the log-odds of the outcome variable for a one-unit increase in the corresponding predictor variable, while holding all other predictor variables constant.

To interpret the coefficients in logistic regression, it is common to exponentiate them to obtain the odds ratio. The odds ratio represents the multiplicative change in the odds of the outcome variable for a one-unit increase in the corresponding predictor variable, while holding all other predictor variables constant.

Odds Ratio

When all the predictor variables in logistic regression are categorical, the coefficients represent the log odds of the outcome variable for each level of the categorical variable relative to a reference level. The odds ratio is used to interpret the coefficients in this scenario, and it represents the multiplicative change in the odds of the outcome variable for one level of the categorical variable relative to the reference level (*64*).

To calculate the odds ratio for a particular level of a categorical variable, we can exponentiate the coefficient for that level and compare it to the reference level. For example, suppose we have a categorical variable with three levels (A, B, and C), and we have estimated the following coefficients in logistic regression:

- β_0 is the intercept term
- β_1 represents the log odds of the outcome variable for level A relative to the reference level
- β_2 represents the log odds of the outcome variable for level B relative to the reference level

To interpret the odds ratio for level A relative to the reference level, we can exponentiate the coefficient β_1 to obtain the odds ratio:

$$Odds\ Ratio\ for\ Level\ A = \exp(\beta_1) \tag{4}$$

This odds ratio represents the multiplicative change in the odds of the outcome variable when the predictor variable is at level A compared to the reference level. For example, if the odds ratio for level A is 2, then the odds of the outcome variable are twice as high when the predictor variable is at level A compared to the reference level, while holding all other predictor variables constant.

Similarly, we can calculate the odds ratio for level B relative to the reference level:

$$Odds\ Ratio\ for\ Level\ B = \exp(\beta_2) \tag{5}$$

This odds ratio represents the multiplicative change in the odds of the outcome variable when the predictor variable is at level B compared to the reference level.

Appendix F – Subset of Potential Predictor Variables

Variable Name	Source
SD2022 Population Density	ESRI – Census
SD2022 Median Household Income	ESRI – Census
SD2022 Average Household Income	ESRI – Census
SD2022 Median Age	ESRI – Census
SD2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent	ESRI – Census
SD2022 Male Population: Percent	ESRI – Census
SD2022 Total Crime Index	ESRI – Census
SD2022_Per_Capita_Income_Cat	ESRI – Census
SD2020 HHs: Inc Below Poverty Level (ACS 5-Yr): Percent	ESRI – Census
SD2022 Diversity Index	ESRI – Census
SD2022 Black Population: Percent	ESRI – Census
Average Daily Zone Traffic (StL Volume)	StreetLight Data
Avg Travel Time (sec)	StreetLight Data
Black	StreetLight Data
With a disability	StreetLight Data
Home to Work	StreetLight Data
Home to Other	StreetLight Data
Non-Home Based Trip	StreetLight Data
Avg All Travel Time	StreetLight Data

Variable Name	Source
Avg Trip Length (mi)	StreetLight Data
Avg Trip Speed (mph)	StreetLight Data
Speed Limit	Google Street View
Number of Lane	Google Street View
Demographic Index (%)	EPA
People of Color (%)	EPA
Low Income (%)	EPA
Unemployment Rate (%)	EPA
Limited English Speaking households (%)	EPA
Less Than High School Education (%)	EPA
Under Age 5 (%)	EPA
Over Age 64 (%)	EPA
SD2022 Hispanic Population: Percent	ESRI – Census
Standard Deviation building	GSV Image Processing
Limited English Speaking households (%)	EPA

Appendix G – Comparison and Discussion of Models

Model 1, Model 2, and Model 3 each have their unique set of predictors and categorization levels. A comparison of these models can help identify the most important factors affecting the percentage of bicyclists exceeding 10 mph on bridges.

Model 1 uses predictors such as the percentage of households with one or more persons with disabilities, the percentage of people with disabilities, and the percentage of Hispanic population. This model has a residual deviance of 44.716 and an AIC value of 64.716, indicating a relatively good fit.

Model 2 uses predictors such as the percentage of Hispanic population, the percentage of Home-to-Work trips, and the percentage of people with disabilities. This model has a residual deviance of 48.968 and an AIC value of 62.968, indicating a relatively good fit.

Model 3 uses predictors such as the percentage of Home-to-Work trips, the percentage of the population with less than a high school education, the total crime index, and the standard deviation of building characteristics. This model has a residual deviance of 43.419 and an AIC value of 61.419, suggesting the strongest fit among these three models.

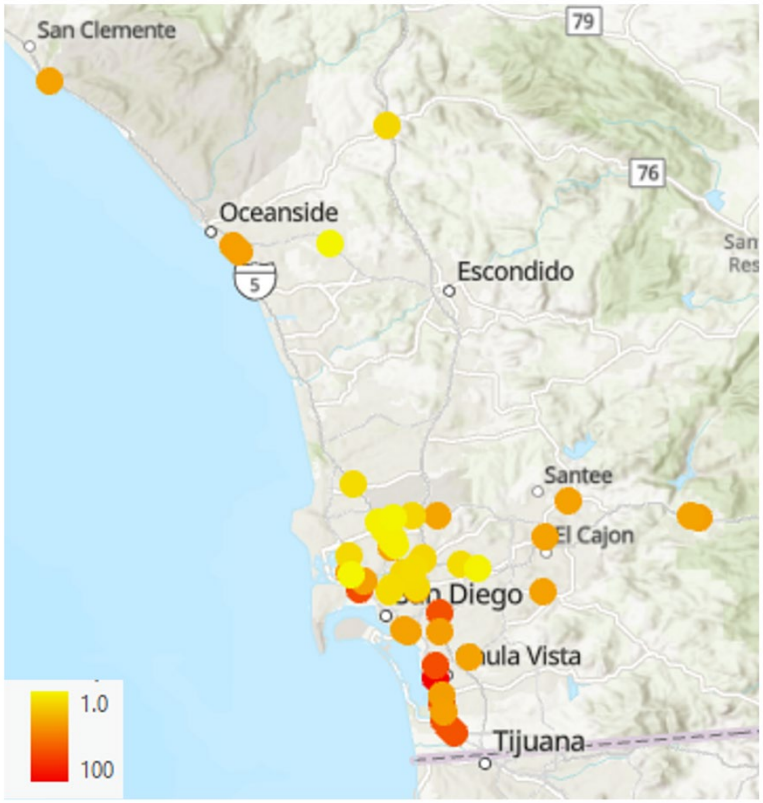
The residual deviance measures the deviance after fitting the logistic regression model with predictor variables. It quantifies how well the model accounts for the observed variations in the data. A smaller residual deviance indicates a better fit of the model compared to the null model. The difference between the null deviance and the residual deviance represents the deviance explained by the predictors. A significant reduction in deviance (i.e., a large difference) indicates that the predictors in the model provide valuable information for explaining the outcome variable, which is a sign of a better model fit.

Comparing the goodness of fit statistics, Model 3 appears to have the strongest explanatory power. Model 3 also has the lowest AIC value, a better fit for the data than the other two models. It is essential to consider both the goodness of fit and parsimony when selecting the best model.

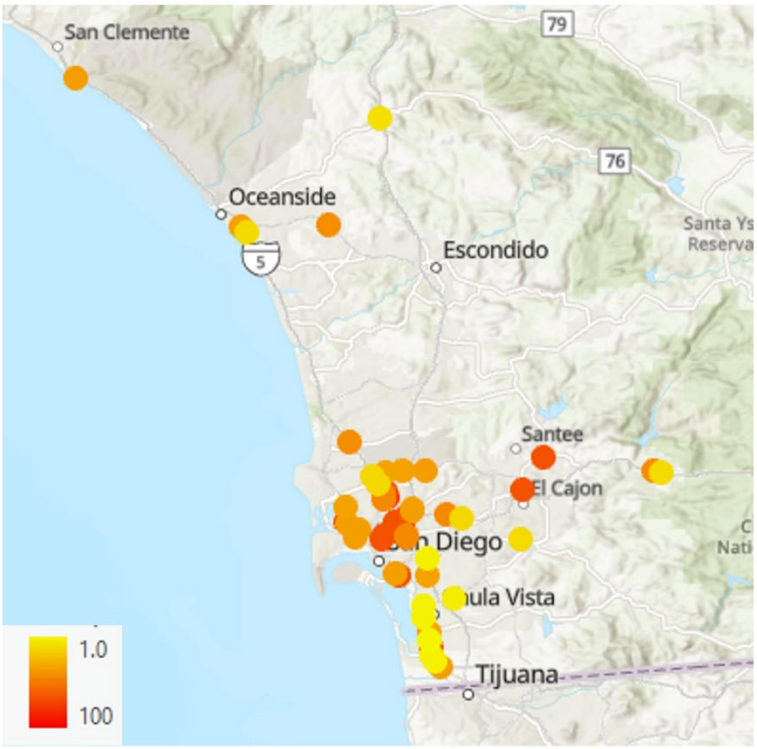
Appendix H – Bridges Total Score Symbolization

In order to gain a more comprehensive understanding of the locations of bridges with the highest need for bicycle safety improvements, additional figures have been created, accompanied by the clarification that darker bubbles represent bridges with higher scores, thus indicating increased risk.

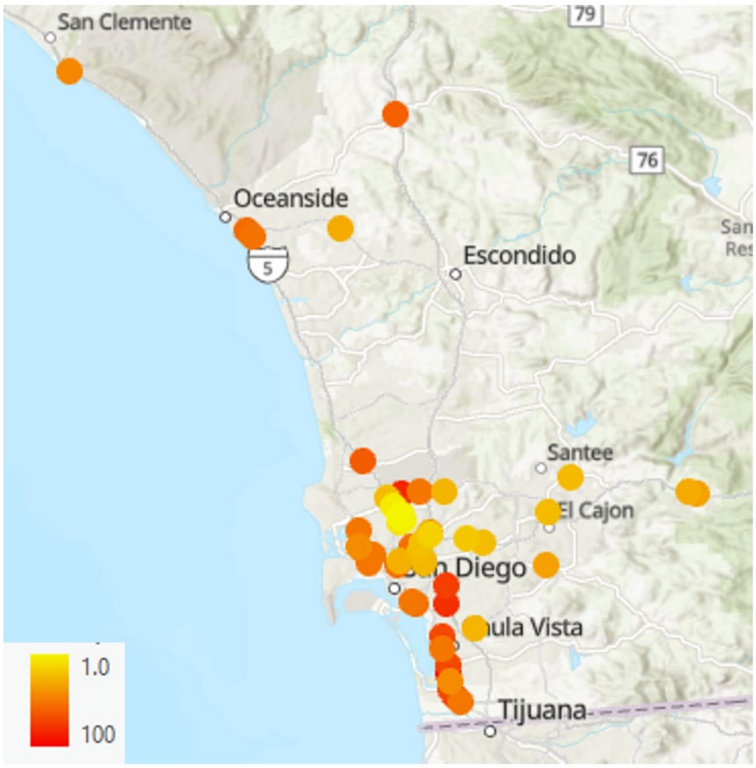
- **First Model's Total Score Symbolization**



- **Second Model's Total Score Symbolization**



- **Third Model's Total Score Symbolization**



Appendix I – List of Variables

Variable Name	Description
2022 Population Density	The estimated number of individuals per square mile of land area in the given geographic unit for the year 2022.
2022 Total Population	The estimated total number of individuals living in the given geographic unit for the year 2022.
2022 Median Household Income	The estimated median income earned by households in the given geographic unit for the year 2022.
2022 Median Household Income: Index	An index that compares the estimated median household income of the given geographic unit to the national median household income. A value greater than 100 indicates an above-average median household income, while a value less than 100 indicates a below-average median household income.
2022 Median Disposable Income	The estimated median income after taxes and other deductions have been taken out of the gross income for the given geographic unit for the year 2022.
2022 Median Disposable Income: Index	An index that compares the estimated median disposable income of the given geographic unit to the national median disposable income. A value greater than 100 indicates an above-average median disposable income, while a value less than 100 indicates a below-average median disposable income.
2022 HH Income <$15000: Percent	The estimated percentage of households in the given geographic unit with an annual income less than $15,000 for the year 2022.
2022 HH Income $15000-24999: Percent	The estimated percentage of households in the given geographic unit with an annual income between $15,000 and $24,999 for the year 2022.
2022 HH Income $25000-34999: Percent	The estimated percentage of households in the given geographic unit with an annual income between $25,000 and $34,999 for the year 2022.
2022 HH Income $35000-49999: Percent	The estimated percentage of households in the given geographic unit with an annual income between $35,000 and $49,999 for the year 2022.
2022 HH Income $50000-74999: Percent	The estimated percentage of households in the given geographic unit with an annual income between $50,000 and $74,999 for the year 2022.

Variable Name	Description
2022 HH Income $75000-99999: Percent	The estimated percentage of households in the given geographic unit with an annual income between $75,000 and $99,999 for the year 2022.
2022 HH Income $100000-149999: Percent	The estimated percentage of households in the given geographic unit with an annual income between $100,000 and $149,999 for the year 2022.
2022 HH Income $150000-199999: Percent	The estimated percentage of households in the given geographic unit with an annual income between $150,000 and $199,999 for the year 2022.
2022 HH Income $200000+: Percent	The estimated percentage of households in the given geographic unit with an annual income of $200,000 or more for the year 2022.
2022 Average Household Income	The estimated average income earned by households in the given geographic unit for the year 2022.
2022 Average Household Income: Index	An index that compares the estimated average household income of the given geographic unit to the national average household income. A value greater than 100 indicates an above-average income, while a value less than 100 indicates a below-average income.
2022 Median Age	The estimated median age of individuals living in the given geographic unit for the year 2022.
2022 Median Age: Index	An index that compares the estimated median age of individuals living in the given geographic unit to the national median age. A value greater than 100 indicates an above-average median age, while a value less than 100 indicates a below-average median age.
2022 Population Age 0-4: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 0 and 4 years old for the year 2022.
2022 Population Age 5-9: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 5 and 9 years old for the year 2022.
2022 Population Age 10-14: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 10 and 14 years old for the year 2022.
2022 Population Age 15-19: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 15 and 19 years old for the year 2022.
2022 Population Age 20-24: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 20 and 24 years old for the year 2022.

Variable Name	Description
2022 Population Age 25-29: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 25 and 29 years old for the year 2022.
2022 Population Age 30-34: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 30 and 34 years old for the year 2022.
2022 Population Age 35-39: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 35 and 39 years old for the year 2022.
2022 Population Age 40-44: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 40 and 44 years old for the year 2022.
2022 Population Age 45-49: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 45 and 49 years old for the year 2022.
2022 Population Age 50-54: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 50 and 54 years old for the year 2022.
2022 Population Age 55-59: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 55 and 59 years old for the year 2022.
2022 Population Age 60-64: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 60 and 64 years old for the year 2022.
2022 Population Age 65-69: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 65 and 69 years old for the year 2022.
2022 Population Age 70-74: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 70 and 74 years old for the year 2022.
2022 Population Age 75-79: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 75 and 79 years old for the year 2022.
2022 Population Age 80-84: Percent	The estimated percentage of the population in the given geographic unit that is between the ages of 80 and 84 years old for the year 2022.
2022 Population Age 85+: Percent	The estimated percentage of the population in the given geographic unit that is age 85 or older for the year 2022.
2022 Housing Affordability Index	A measure of the affordability of housing in the given geographic unit for the year 2022. A higher index indicates more affordable housing.

Variable Name	Description
2022 Total Households	The estimated total number of households in the given geographic unit for the year 2022.
2020 HHs w/1+ Persons w/Disability (ACS 5-Yr): Percent	The estimated percentage of households in the given geographic unit that have at least one person with a disability for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.
2020 HHs w/No Persons w/Disability (ACS 5-Yr): Percent	The estimated percentage of households in the given geographic unit with no persons with disabilities for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.
2022 Male Population: Percent	The estimated percentage of the population in the given geographic unit that is male for the year 2022.
2022 Female Population: Percent	The estimated percentage of the population in the given geographic unit that is female for the year 2022.
2022 Health Care	An estimation of the availability and quality of healthcare services in the given geographic unit for the year 2022. This includes the availability of hospitals, doctors, and other medical facilities in the area.
2022 Health Care: Average	The estimated average quality of healthcare services in the given geographic unit for the year 2022, based on various indicators such as the number of medical professionals per capita, the availability of hospital beds, and other factors.
2022 Health Care: Index	An index that compares the estimated quality of healthcare services in the given geographic unit to the national average quality of healthcare services. A value greater than 100 indicates above-average quality, while a value less than 100 indicates below-average quality.
2022 Pop Age 25+: < 9th Grade: Percent	The estimated percentage of the population in the given geographic unit age 25 or older with less than a 9th grade education level for the year 2022.
2022 Pop Age 25+: High School/No Diploma: Percent	The estimated percentage of the population in the given geographic unit age 25 or older with a high school diploma or less for the year 2022.
2022 Pop Age 25+: High School Diploma: Percent	The estimated percentage of the population in the given geographic unit age 25 or older with a high school diploma for the year 2022.

Variable Name	Description
2022 Pop Age 25+: GED: Percent	The estimated percentage of the population in the given geographic unit age 25 or older with a General Educational Development (GED) certificate for the year 2022.
2022 Pop Age 25+: Some College/No Degree: Percent	The estimated percentage of the population in the given geographic unit age 25 or older with some college education, but no degree for the year 2022.
2022 Pop Age 25+: Associate's Degree: Percent	The estimated percentage of the population in the given geographic unit age 25 or older with an associate's degree for the year 2022.
2022 Pop Age 25+: Bachelor's Degree: Percent	The estimated percentage of the population in the given geographic unit age 25 or older with a bachelor's degree for the year 2022.
2022 Pop Age 25+: Grad/Professional Degree: Percent	The estimated percentage of the population in the given geographic unit age 25 or older with a graduate or professional degree for the year 2022.
2022 Educational Attainment Base	A measure of the level of education attained by residents in the given geographic unit for the year 2022. This includes the percentage of the population with different levels of educational attainment.
2022 Total Crime Index	An index that measures the overall level of crime in the given geographic unit for the year 2022. A higher index indicates higher crime rates.
2022 Per Capita Income	The estimated total income earned by residents of the given geographic unit divided by the total population for the year 2022.
2022 Per Capita Income: Index	An index that compares the estimated per capita income of the given geographic unit to the national per capita income. A value greater than 100 indicates an above-average income, while a value less than 100 indicates a below-average income.
2022 Median Home Value	The estimated median value of owner-occupied homes in the given geographic unit for the year 2022.
2022 Median Home Value: Index	An index that compares the estimated median home value of the given geographic unit to the national median home value. A value greater than 100 indicates above-average home value, while a value less than 100 indicates below-average home value.

Variable Name	Description
2020 HHs: Inc Below Poverty Level (ACS 5-Yr): Percent	The estimated percentage of households in the given geographic unit with income below the poverty level for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.
2022 Diversity Index	An index that measures the diversity of the population in the given geographic unit for the year 2022. The index takes into account the distribution of different racial and ethnic groups in the population. A higher index indicates greater diversity.
2022 Hispanic Population: Percent	The estimated percentage of the population in the given geographic unit that identifies as Hispanic for the year 2022.
2022 Hispanic White Pop: Percent	The estimated percentage of the population in the given geographic unit that identifies as both Hispanic and White for the year 2022.
2022 Non-Hispanic White Pop: Percent	The estimated percentage of the population in the given geographic unit that identifies as White but not Hispanic for the year 2022.
2022 White Population: Percent	The estimated percentage of the population in the given geographic unit that identifies as White for the year 2022.
2022 American Indian Population: Percent	The estimated percentage of the population in the given geographic unit that identifies as American Indian or Alaska Native for the year 2022.
2022 Black Population: Percent	The estimated percentage of the population in the given geographic unit that identifies as Black or African American for the year 2022.
2022 Asian Population: Percent	The estimated percentage of the population in the given geographic unit that identifies as Asian for the year 2022.
2022 Pacific Islander Population: Percent	The estimated percentage of the population in the given geographic unit that identifies as Native Hawaiian or other Pacific Islander for the year 2022.
2022 Other Race Population: Percent	The estimated percentage of the population in the given geographic unit that identifies as a race other than White, Black, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, or two or more races for the year 2022.
2022 Population of 2+ Races: Percent	The estimated percentage of the population in the given geographic unit that identifies as two or more races for the year 2022.

Variable Name	Description
2022 Population by Race Base	A measure of the racial distribution of the population in the given geographic unit for the year 2022. This includes the percentage of the population in different racial categories.
2020 Pop 5-17 Speak Only English (ACS 5-Yr): Percent	The estimated percentage of the population aged 5 to 17 in the given geographic unit who speak only English at home for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.
2020 Pop 5-17 Speak Spanish (ACS 5-Yr): Percent	The estimated percentage of the population aged 5 to 17 in the given geographic unit who speak Spanish at home for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.
2020 Pop 5-17 Speak Span/English VW/W (ACS 5-Yr): Percent	The estimated percentage of the population aged 5 to 17 in the given geographic unit who speak both Spanish and English at home, or who speak Spanish well enough to have a conversation in English at home for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.
2020 Pop 18-64 Speak Only English (ACS 5-Yr): Percent	The estimated percentage of the population aged 18 to 64 in the given geographic unit who speak only English at home for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.
2020 Pop 18-64 Speak Spanish (ACS 5-Yr): Percent	The estimated percentage of the population aged 18 to 64 in the given geographic unit who speak Spanish at home for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.
2020 Pop 18-64 Speak Span/English VW/W (ACS 5-Yr): Percent	The estimated percentage of the population aged 18 to 64 in the given geographic unit who speak both Spanish and English at home, or who speak Spanish well enough to have a conversation in English at home for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.
2020 Pop 65+ Speak Only English (ACS 5-Yr): Percent	The estimated percentage of the population aged 65 and older in the given geographic unit who speak only English at home for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.

Variable Name	Description
2020 Pop 65+ Speak Spanish (ACS 5-Yr): Percent	The estimated percentage of the population aged 65 and older in the given geographic unit who speak Spanish at home for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.
2020 Pop 65+ Speak Span/English VW/W (ACS 5-Yr): Percent	The estimated percentage of the population aged 65 and older in the given geographic unit who speak both Spanish and English at home, or who speak Spanish well enough to have a conversation in English at home for the year 2020. This is based on data from the American Community Survey (ACS) 5-year estimates.
2022 Education	A measure of the education level of residents in the given geographic unit for the year 2022. This includes the percentage of the population with different levels of educational attainment.
2022 Education: Average	The estimated average education level of residents in the given geographic unit for the year 2022. This is calculated by assigning a numerical value to each level of educational attainment and computing the average value for the population.
2022 Education: Index	An index that compares the estimated education level of the given geographic unit to the national education level. A value greater than 100 indicates above-average education, while a value less than 100 indicates below-average education.
2022 only speak english	The estimated percentage of the population in the given geographic unit that speaks only English at home for the year 2022.
2022 only speak spanish	The estimated percentage of the population in the given geographic unit that speaks only Spanish at home for the year 2022.
2022 Spanish vw / English w	The estimated percentage of the population in the given geographic unit that speaks both Spanish and English at home, or who speaks Spanish well enough to have a conversation in English at home for the year 2022.
Average Daily Zone Traffic (StL Volume)	The estimated average daily bicycle volume in the given streetlight zone. This variable is measured in bicycle per day.
Avg Travel Time (sec)	The estimated average travel time for bicyclists in the given streetlight zone. This variable is measured in seconds.

Variable Name	Description
Avg All Travel Time (sec)	The estimated average travel time for all modes of transportation in the given streetlight zone. This variable is measured in seconds.
Avg Trip Length (mi)	The estimated average trip length for bicyclists in the given streetlight zone. This variable is measured in miles.
Avg All Trip Length (mi)	The estimated average trip length for all modes of transportation in the given streetlight zone. This variable is measured in miles.
Income Less than 10K	The estimated percentage of households in the given streetlight zone with income less than $10,000.
Income 10K to 15K	The estimated percentage of households in the given streetlight zone with income between $10,000 and $15,000.
Income 15K to 20K	The estimated percentage of households in the given streetlight zone with income between $15,000 and $20,000.
Income 20K to 25K	The estimated percentage of households in the given streetlight zone with income between $20,000 and $25,000.
Income 25K to 30K	The estimated percentage of households in the given streetlight zone with income between $25,000 and $30,000.
Income 30K to 35K	The estimated percentage of households in the given streetlight zone with income between $30,000 and $35,000.
Income 35K to 40K	The estimated percentage of households in the given streetlight zone with income between $35,000 and $40,000.
Income 40K to 45K	The estimated percentage of households in the given streetlight zone with income between $40,000 and $45,000.
Income 45K to 50K	The estimated percentage of households in the given streetlight zone with income between $45,000 and $50,000.
Income 50K to 60K	The estimated percentage of households in the given streetlight zone with income between $50,000 and $60,000.
Income 60K to 75K	The estimated percentage of households in the given streetlight zone with income between $60,000 and $75,000.

Variable Name	Description
Income 75K to 100K	The estimated percentage of households in the given streetlight zone with income between $75,000 and $100,000.
Income 100K to 125K	The estimated percentage of households in the given streetlight zone with income between $100,000 and $125,000.
Income 125K to 150K	The estimated percentage of households in the given streetlight zone with income between $125,000 and $150,000.
Income 150K to 200K	The estimated percentage of households in the given streetlight zone with income between $150,000 and $200,000.
Income More than 200K	The estimated percentage of households in the given streetlight zone with income greater than $200,000. This may include households with incomes in the millions of dollars, depending on the income distribution in the given geographic area.
Less than 9th grade	The estimated percentage of the population aged 25 years or older in the given streetlight zone who have not completed 9th grade. This includes individuals who have not completed any formal education or have completed only kindergarten or some elementary school.
9th to 12th grade, no diploma	The estimated percentage of the population aged 25 years or older in the given streetlight zone who have completed some high school but have not obtained a high school diploma. This includes individuals who have completed 9th grade, 10th grade, 11th grade, or 12th grade but have not graduated.
High school graduate	The estimated percentage of the population aged 25 years or older in the given streetlight zone who have obtained a high school diploma or equivalent (e.g., GED).
Some college, no degree	The estimated percentage of the population aged 25 years or older in the given streetlight zone who have completed some college coursework but have not obtained a degree. This includes individuals who have completed trade or vocational school programs.
Associate's degree	The estimated percentage of the population aged 25 years or older in the given streetlight zone who have obtained an associate's degree.

Variable Name	Description
Bachelor's degree	The estimated percentage of the population aged 25 years or older in the given streetlight zone who have obtained a bachelor's degree.
Graduate or professional degree	The estimated percentage of the population aged 25 years or older in the given streetlight zone who have obtained a graduate or professional degree (e.g., master's, doctoral, or professional degree such as law or medicine).
Agriculture, forestry, fishing, hunting, mining	The estimated percentage of workers in the given streetlight zone employed in the agriculture, forestry, fishing, hunting, or mining industries.
Construction	The estimated percentage of workers in the given streetlight zone employed in the construction industry.
Manufacturing	The estimated percentage of workers in the given streetlight zone employed in the manufacturing industry.
Wholesale trade	The estimated percentage of workers in the given streetlight zone employed in the wholesale trade industry.
Retail trade	The estimated percentage of workers in the given streetlight zone employed in the retail trade industry.
Transportation, warehousing, utilities	The estimated percentage of workers in the given streetlight zone employed in the transportation, warehousing, or utilities industries.
Information	The estimated percentage of workers in the given streetlight zone employed in the information industry, which includes publishing, motion picture and sound recording industries, broadcasting, telecommunications, data processing, and other information services.
Finance, insurance, real estate rental and leasing	The estimated percentage of workers in the given streetlight zone employed in the finance, insurance, or real estate rental and leasing industries.
Professional, scientific, management, etc. services	The estimated percentage of workers in the given streetlight zone employed in professional, scientific, management, administrative, and waste management services. This industry category includes occupations in computer systems design, research and development, and management consulting.

Variable Name	Description
Educational services, health care, social assistance	The estimated percentage of workers in the given streetlight zone employed in educational services, health care, or social assistance industries.
Arts, entertainment, recreation, etc. services	The estimated percentage of workers in the given streetlight zone employed in arts, entertainment, recreation, accommodation, or food services industries.
Other services (except public administration)	The estimated percentage of workers in the given streetlight zone employed in industries other than those listed above, such as repair and maintenance, personal care services, or membership organizations.
Public administration	The estimated percentage of workers in the given streetlight zone employed in public administration, including federal, state, and local government.
Military (Employment Industry)	The estimated percentage of workers in the given streetlight zone employed in the military industry.
Not employed (Employment Industry)	The estimated percentage of individuals aged 16 years or older in the given streetlight zone who are not employed. This may include individuals who are retired, students, or not actively seeking employment for other reasons.
Private wage and salary workers	The estimated percentage of workers in the given streetlight zone employed in private wage and salary jobs.
Government workers	The estimated percentage of workers in the given streetlight zone employed in government jobs.
Self-employed workers	The estimated percentage of workers in the given streetlight zone who are self-employed. This may include individuals who own their own businesses or work as independent contractors.
Unpaid family workers	The estimated percentage of workers in the given streetlight zone who are unpaid family workers.
Military (Employment Class)	The estimated percentage of individuals aged 16 years or older in the given streetlight zone who are employed in the military.
Not employed (Employment Class)	The estimated percentage of individuals aged 16 years or older in the given streetlight zone who are not employed. This may include individuals who are retired, students, or not actively seeking employment for other reasons.

Variable Name	Description
White	The estimated percentage of individuals in the given streetlight zone who identify as White alone, regardless of Hispanic or Latino origin. This is a racial category used by the US Census Bureau.
Black	The estimated percentage of individuals in the given streetlight zone who identify as Black or African American alone, regardless of Hispanic or Latino origin. This is a racial category used by the US Census Bureau.
American Indian	The estimated percentage of individuals in the given streetlight zone who identify as American Indian or Alaska Native alone, regardless of Hispanic or Latino origin. This is a racial category used by the US Census Bureau.
Asian	The estimated percentage of individuals in the given streetlight zone who identify as Asian alone, regardless of Hispanic or Latino origin. This is a racial category used by the US Census Bureau.
Pacific Islander	The estimated percentage of individuals in the given streetlight zone who identify as Native Hawaiian or Other Pacific Islander alone, regardless of Hispanic or Latino origin. This is a racial category used by the US Census Bureau.
Other Race	The estimated percentage of individuals in the given streetlight zone who identify as a race other than White, Black, American Indian or Alaska Native, Asian, or Native Hawaiian or Other Pacific Islander, regardless of Hispanic or Latino origin. This is a racial category used by the US Census Bureau.
Multiple Races	The estimated percentage of individuals in the given streetlight zone who identify as two or more races, regardless of Hispanic or Latino origin. This is a racial category used by the US Census Bureau.
Hispanic	The estimated percentage of individuals in the given streetlight zone who identify as Hispanic or Latino of any race.
Non-Hispanic	The estimated percentage of individuals in the given streetlight zone who do not identify as Hispanic or Latino.
Foreign Born	The estimated percentage of individuals in the given streetlight zone who were born outside the United States, Puerto Rico, or US Island Areas.

Variable Name	Description
Non-foreign Born	The estimated percentage of individuals in the given streetlight zone who were born in the United States, Puerto Rico, or US Island Areas.
Speak English less than 'very well'	The estimated percentage of individuals aged 5 years or older in the given streetlight zone who speak a language other than English at home and who reported speaking English less than "very well." This is an indicator of limited English proficiency. It is based on responses to the US Census Bureau's American Community Survey.
With a disability	The estimated percentage of individuals aged 16 years or older in the given streetlight zone who report having a disability, as defined by the Americans with Disabilities Act (ADA). This includes individuals who have difficulty performing activities of daily living or who have a physical or mental impairment that limits one or more major life activities.
Without a disability	The estimated percentage of individuals aged 16 years or older in the given streetlight zone who do not report having a disability, as defined by the Americans with Disabilities Act (ADA).
With Kids	The estimated percentage of households in the given streetlight zone with at least one child under the age of 18 years old.
With No Kids	The estimated percentage of households in the given streetlight zone with no children under the age of 18 years old.
With Kids under 6 years	The estimated percentage of households in the given streetlight zone with at least one child under the age of 6 years old.
With Kids between 6-17 years	The estimated percentage of households in the given streetlight zone with at least one child between the ages of 6 and 17 years old.
Owner occupied	The estimated percentage of occupied housing units in the given streetlight zone that are owned by the occupant.
Renter occupied	The estimated percentage of occupied housing units in the given streetlight zone that are rented by the occupant.
No vehicle available	The estimated percentage of households in the given streetlight zone with no vehicle available.
1 vehicle available	The estimated percentage of households in the given streetlight zone with one vehicle available.

Variable Name	Description
2 vehicles available	The estimated percentage of households in the given streetlight zone with two vehicles available.
3 or more vehicles available	The estimated percentage of households in the given streetlight zone with three or more vehicles available.
1 Unit Structure	The estimated percentage of occupied housing units in the given streetlight zone that are single-unit structures, such as detached houses, rowhouses, or townhouses.
2 Unit Structure	The estimated percentage of occupied housing units in the given streetlight zone that are two-unit structures, such as duplexes or side-by-side townhouses.
3-4 Unit Structure	The estimated percentage of occupied housing units in the given streetlight zone that are three- or four-unit structures, such as triplexes or fourplexes.
5-9 Unit Structure	The estimated percentage of occupied housing units in the given streetlight zone that are structures containing five to nine units.
10-19 Unit Structure	The estimated percentage of occupied housing units in the given streetlight zone that are structures containing ten to nineteen units.
20-49 Unit Structure	The estimated percentage of occupied housing units in the given streetlight zone that are structures containing twenty to forty-nine units.
50+ Unit Structure	The estimated percentage of occupied housing units in the given streetlight zone that are structures containing fifty or more units.
Mobile homes, RV, boat, van, other	The estimated percentage of occupied housing units in the given streetlight zone that are not classified as single-unit structures, two-unit structures, or multi-unit structures. This includes mobile homes, RVs, boats, vans, and other types of non-traditional housing.
Home to Work	The estimated percentage of trips originating from the given streetlight zone that are commutes to work.
Home to Other	The estimated percentage of trips originating from the given streetlight zone that are not commutes to work or school, and are not classified as non-home based trips.

Variable Name	Description
Non-Home Based Trip	The estimated percentage of trips originating from the given streetlight zone that are not associated with a residence or workplace, and are not classified as home to other trips.
Avg Travel Time (sec)	The estimated average travel time in seconds for all trips originating from the given streetlight zone. This includes all modes of transportation, such as walking, biking, driving, or public transit.
Avg All Travel Time (sec)	The estimated average travel time in seconds for all trips ending or passing through the given streetlight zone. This includes all modes of transportation, such as walking, biking, driving, or public transit.
Avg Trip Length (mi)	The estimated average length in miles for all trips originating from the given streetlight zone. This includes all modes of transportation, such as walking, biking, driving, or public transit.
Avg All Trip Length (mi)	The estimated average length in miles for all trips ending or passing through the given streetlight zone. This includes all modes of transportation, such as walking, biking, driving, or public transit.
Travel Time 0-5 min (percent)	The estimated percentage of all trips originating from the given streetlight zone with a travel time of 0 to 5 minutes.
Travel Time 5-10 min (percent)	The estimated percentage of all trips originating from the given streetlight zone with a travel time of 5 to 10 minutes.
Travel Time 10-15 min (percent)	The estimated percentage of all trips originating from the given streetlight zone with a travel time of 10 to 15 minutes.
Travel Time 15-20 min (percent)	The estimated percentage of all trips originating from the given streetlight zone with a travel time of 15 to 20 minutes.
Travel Time 20-25 min (percent)	The estimated percentage of all trips originating from the given streetlight zone with a travel time of 20 to 25 minutes.
Travel Time 25-30 min (percent)	The estimated percentage of all trips originating from the given streetlight zone with a travel time of 25 to 30 minutes.
Travel Time 30+ min (percent)	The estimated percentage of all trips originating from the given streetlight zone with a travel time of more than 30 minutes.

Variable Name	Description
Trip Length 0-1 mi (percent)	The estimated percentage of all trips originating from the given streetlight zone with a length of 0 to 1 mile.
Trip Length 1-2 mi (percent)	The estimated percentage of all trips originating from the given streetlight zone with a length of 1 to 2 miles.
Trip Length 2-3 mi (percent)	The estimated percentage of all trips originating from the given streetlight zone with a length of 2 to 3 miles.
Trip Length 3-4 mi (percent)	The estimated percentage of all trips originating from the given streetlight zone with a length of 3 to 4 miles.
Trip Length 4-5 mi (percent)	The estimated percentage of all trips originating from the given streetlight zone with a length of 4 to 5 miles.
Trip Length 5-6 mi (percent)	The estimated percentage of all trips originating from the given streetlight zone with a length of 5 to 6 miles.
Trip Length 6-7 mi (percent)	The estimated percentage of all trips originating from the given streetlight zone with a length of 6 to 7 miles.
Trip Length 7+ mi (percent)	The estimated percentage of all trips originating from the given streetlight zone with a length of more than 7 miles.
Trip Speed 0-2 mph (percent)	The estimated percentage of all trips originating from the given streetlight zone with a speed of 0 to 2 miles per hour. This typically includes walking trips or very slow cycling trips.
Trip Speed 2-4 mph (percent)	The estimated percentage of all trips originating from the given streetlight zone with a speed of 2 to 4 miles per hour. This typically includes slower cycling trips or trips in heavy traffic conditions.
Trip Speed 4-6 mph (percent)	The estimated percentage of all trips originating from the given streetlight zone with a speed of 4 to 6 miles per hour. This typically includes moderate cycling speeds or driving in very slow traffic conditions.
Trip Speed 6-8 mph (percent)	The estimated percentage of all trips originating from the given streetlight zone with a speed of 6 to 8 miles per hour. This typically includes faster cycling speeds or driving in moderate traffic conditions.

Variable Name	Description
Trip Speed 8-10 mph (percent)	The estimated percentage of all trips originating from the given streetlight zone with a speed of 8 to 10 miles per hour. This typically includes driving in uncongested urban areas or on suburban roads.
Trip Speed 10-12 mph (percent)	The estimated percentage of all trips originating from the given streetlight zone with a speed of 10 to 12 miles per hour. This typically includes driving on arterial roads or highways in uncongested conditions.
Trip Speed 12-14 mph (percent)	The estimated percentage of all trips originating from the given streetlight zone with a speed of 12 to 14 miles per hour. This typically includes driving on highways in moderate traffic conditions.
Trip Speed 14+ mph (percent)	The estimated percentage of all trips originating from the given streetlight zone with a speed of more than 14 miles per hour. This typically includes driving on highways in uncongested conditions or on high-speed arterial roads.
Circuity 1-2 (percent)	The estimated percentage of all trips originating from the given streetlight zone with a circuity ratio of 1 to 2. Circuity ratio is defined as the ratio between the length of the actual route taken by the traveler and the shortest possible direct route between the origin and destination.
Circuity 2-3 (percent)	The estimated percentage of all trips originating from the given streetlight zone with a circuity ratio of 2 to 3.
Circuity 3-4 (percent)	The estimated percentage of all trips originating from the given streetlight zone with a circuity ratio of 3 to 4.
Circuity 4-5 (percent)	The estimated percentage of all trips originating from the given streetlight zone with a circuity ratio of 4 to 5.
Circuity 5-6 (percent)	The estimated percentage of all trips originating from the given streetlight zone with a circuity ratio of 5 to 6.
Circuity 6+ (percent)	The estimated percentage of all trips originating from the given streetlight zone with a circuity ratio greater than 6.
Speed Limit	The posted speed limit on the roadway associated with the given streetlight zone.

Variable Name	Description
Bicycle Lane	A binary variable indicating whether the roadway associated with the given streetlight zone has a dedicated bicycle lane or not. A value of 1 indicates the presence of a bicycle lane, while a value of 0 indicates the absence of a bicycle lane.
Street Median	A binary variable indicating whether the roadway associated with the given streetlight zone has a median or not. A value of 1 indicates the presence of a median, while a value of 0 indicates the absence of a median.
Number of Lane	The number of lanes on the roadway associated with the given streetlight zone. This includes both the number of travel lanes and the number of turning lanes.
Bound Type	The type of geographic boundary associated with the given streetlight zone. This can include census tracts, neighborhoods, or other custom-defined boundaries.
Demographic Index	A composite index that summarizes various demographic factors associated with the given streetlight zone, such as income, education, race, and age. This index is often used as a proxy for measuring overall community health and well-being.
People of Color	The estimated percentage of the population within the given streetlight zone that identifies as a racial or ethnic minority, including Hispanic or Latino populations.
Low Income	The estimated percentage of households within the given streetlight zone with an annual income below the poverty line.
Unemployment Rate	The estimated percentage of the labor force within the given streetlight zone that is currently unemployed and actively seeking work.
Limited English Speaking households	The estimated percentage of households within the given streetlight zone in which no one over the age of 14 speaks English "very well".
Less Than High School Education	The estimated percentage of the population over the age of 25 within the given streetlight zone who have not completed a high school education or equivalent.
Under Age 5	The estimated percentage of the population within the given streetlight zone that is under the age of 5.

Variable Name	Description
Over Age 64	The estimated percentage of the population within the given streetlight zone that is over the age of 64.
Mean building	The average percentage of the area covered by buildings in the Google Street View (GSV) images within a 0.3-mile buffer around each bridge.
Standard Deviation building	The standard deviation of the percentage of the area covered by buildings in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in building coverage across the different GSV images.
Mean car	The average percentage of the area covered by cars in the GSV images within a 0.3-mile buffer around each bridge.
Standard Deviation car	The standard deviation of the percentage of the area covered by cars in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in car coverage across the different GSV images.
Mean earth	The average percentage of the area covered by earth (e.g., bare ground, gravel, dirt) in the GSV images within a 0.3-mile buffer around each bridge.
Standard Deviation earth	The standard deviation of the percentage of the area covered by earth in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in earth coverage across the different GSV images.
Mean fence	The average percentage of the area covered by fences in the GSV images within a 0.3-mile buffer around each bridge.
Standard Deviation fence	The standard deviation of the percentage of the area covered by fences in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in fence coverage across the different GSV images.
Mean grass	The average percentage of the area covered by grass (or other vegetation) in the GSV images within a 0.3-mile buffer around each bridge.
Standard Deviation grass	The standard deviation of the percentage of the area covered by grass in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in grass coverage across the different GSV images.

Variable Name	Description
Mean plant	The average percentage of the area covered by plants (e.g., trees, shrubs) in the Google Street View (GSV) images within a 0.3-mile buffer around each bridge.
Standard Deviation plant	The standard deviation of the percentage of the area covered by plants in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in plant coverage across the different GSV images.
Mean road	The average percentage of the area covered by roads (e.g., pavement, asphalt) in the GSV images within a 0.3-mile buffer around each bridge.
Standard Deviation road	The standard deviation of the percentage of the area covered by roads in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in road coverage across the different GSV images.
Mean sidewalk	The average percentage of the area covered by sidewalks in the GSV images within a 0.3-mile buffer around each bridge.
Standard Deviation sidewalk	The standard deviation of the percentage of the area covered by sidewalks in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in sidewalk coverage across the different GSV images.
Mean signboard	The average percentage of the area covered by signboards (e.g., billboards, street signs) in the GSV images within a 0.3-mile buffer around each bridge.
Standard Deviation signboard	The standard deviation of the percentage of the area covered by signboards in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in signboard coverage across the different GSV images.
Mean sky	The average percentage of the area covered by the sky in the GSV images within a 0.3-mile buffer around each bridge.
Standard Deviation sky	The standard deviation of the percentage of the area covered by the sky in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in sky coverage across the different GSV images.

Variable Name	Description
Mean streetlight	The average percentage of the area covered by streetlights in the GSV images within a 0.3-mile buffer around each bridge.
Standard Deviation streetlight	The standard deviation of the percentage of the area covered by streetlights in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in streetlight coverage across the different GSV images.
Mean tree	The average percentage of the area covered by trees in the GSV images within a 0.3-mile buffer around each bridge.
Standard Deviation tree	The standard deviation of the percentage of the area covered by trees in the GSV images within a 0.3-mile buffer around each bridge. This indicates the degree of variation in tree coverage across the different GSV images.
Speed Limit	The posted speed limit for the road segment
Bicycle Lane	Whether there is a designated lane for bicycles on the road segment
Street Median	Whether there is a median dividing the road segment
Number of Lane	The number of lanes in the road segment
Demographic Index (%)	A composite index that combines several demographic factors such as race, ethnicity, age, education, and income, to describe the overall diversity and socioeconomic status of the area
People of Color (%)	The percentage of the population that identifies as a race or ethnicity other than non-Hispanic White
Low Income (%)	The percentage of households with an income below the poverty level
Unemployment Rate (%)	The percentage of the labor force that is unemployed and actively seeking employment
Limited English Speaking households (%)	The percentage of households where no one over the age of 14 speaks English "very well"
Less Than High School Education (%)	The percentage of the population aged 25 and over with less than a high school education
Under Age 5 (%)	The percentage of the population that is under the age of 5
Over Age 64 (%)	The percentage of the population that is over the age of 64