



# Airbnb Scraping



Vinh Pham, Alex Nikolov, Alan Huang, Sabrina Wang, Han Liu

CS 4624 Multimedia, Hypertext, and Information Access

Instructor: Fox, Edward A.

Virginia Tech, Blacksburg, VA 24061, May 9, 2020

---

# Outline

- Background
- Work Completed
- Data
- Timeline
- Visualization
- Website
- Testing and Assessment
- The Future
- Acknowledgements/References



---

# Background

- Impact of Airbnb on local residents
- Lower the price of short rentals
- Transient occupancy tax



---

## Where We Come In

- Our job is to scrape data from Airbnb
- Starting with open-source code
- Visualize data for future research



---

# Completed

- Tableau -> Python with pyechart.
  - Static maps and dynamic graphs.
- Data collection.
  - Expanded script.
  - Docker.
- Websites done. Some designs have been changed.
  - Can select different regions.
- Collected data for Austria, now collecting data for rural Virginia.





# Review Data

listing_id	id	date	reviewer_id	reviewer_name	rating	comments
25909468	618105219	03-14-2020	338835038	Korey	5	Renee was such an awesome person to boom with! She treated my brother and I
25909468	614878841	03-06-2020	109025473	Brandon	5	Renee is an excellent host who goes above and beyond to make your stay comfort
25909468	611348914	02-26-2020	5627278	'Julianna	5	Very easy to check-in, super clean, all I needed for a quick overnight. I would stay
25909468	609858279	02-23-2020	106496974	Sarah	5	Superb places to stay. Renee was a great host
25909468	607697254	02-18-2020	130387777	越	5	Renee 是一位特别好的主人。她的房间很温馨,对于每个旅行者都是另一个家。她有
25909468	595557578	01-20-2020	173240950	Betsy	5	Renee was an awesome hostess, very friendly and welcoming. The space was w
25909468	594380660	01-18-2020	153097207	Bryan	5	Great space for what we needed. Thanks so much!!
25909468	591148054	01-10-2020	313828791	Stacey	5	Renee is kind, friendly, and a great communicator. Her home is clean, comfy, spa
25909468	590790555	01-09-2020	133995502	Crystal	5	Description and reviews are spot on. The bathroom is really for Airbnb guests but
25909468	588469966	01-04-2020	302818760	David	5	Clean, cozy room with a private entrance. Quiet. It was a nice space. We enjoyed



## Listing Data

	A	B	C	D	E	F	G	H	I
1	id	name	host_id	host_name	neighbourh	neighbourh	latitude	longitude	room_type
2	22560448	Luxury next	76104209	Rated		Centrum-O	52.36248	4.88658	Entire hom
3	22992570	3 BEDROOM	76104209	Rated		Buitenvelde	52.3266	4.87082	Entire hom
4	23136800	DELIGHTFU	76104209	Rated		De Baarsjes	52.36953	4.87382	Entire hom
5	23246529	SUPERBLY I	76104209	Rated		Zuid	52.35408	4.88435	Entire hom
6	23657826	VERY TREN	76104209	Rated		De Pijp - Ri	52.35346	4.90233	Entire hom
7	25477536	LUXURY IN	76104209	Rated		Centrum-W	52.37276	4.88016	Entire hom
8	9249719	DutchStay I	41042925	Dario		Centrum-W	52.3722	4.89953	Entire hom
9	1508944	AAA: Amaz	5113662	Stefan		De Baarsjes	52.37359	4.85833	Entire hom
10	5341871	Crane Hote	21831151	Faralda		Oud-Noord	52.4005	4.89558	Private roo



# Calendar Data

Montgomery\_calendar

listing_id	date	price	minimum_nights	maximum_nights	available	survey_id
25909468	2020-03-01	38	1	48	f	1
25909468	2020-03-02	48	1	48	f	1
25909468	2020-03-03	48	1	48	f	1
25909468	2020-03-04	48	1	48	f	1
25909468	2020-03-05	48	1	48	f	1
25909468	2020-03-06	48	1	48	f	1
25909468	2020-03-07	48	1	48	f	1
25909468	2020-03-08	48	1	48	f	1
25909468	2020-03-09	48	1	48	f	1



# TIMELINE



2/1

Old Code Up

2/18

Code mostly working



3/10

Think of automation

4/1

Data Collecting



5/5

Visualization on Website



2/10

Old Code Understood

3/1

Website Up

3/18

Map & Charts

4/18

Data Collecting

<- Currently



---

# Visualizations

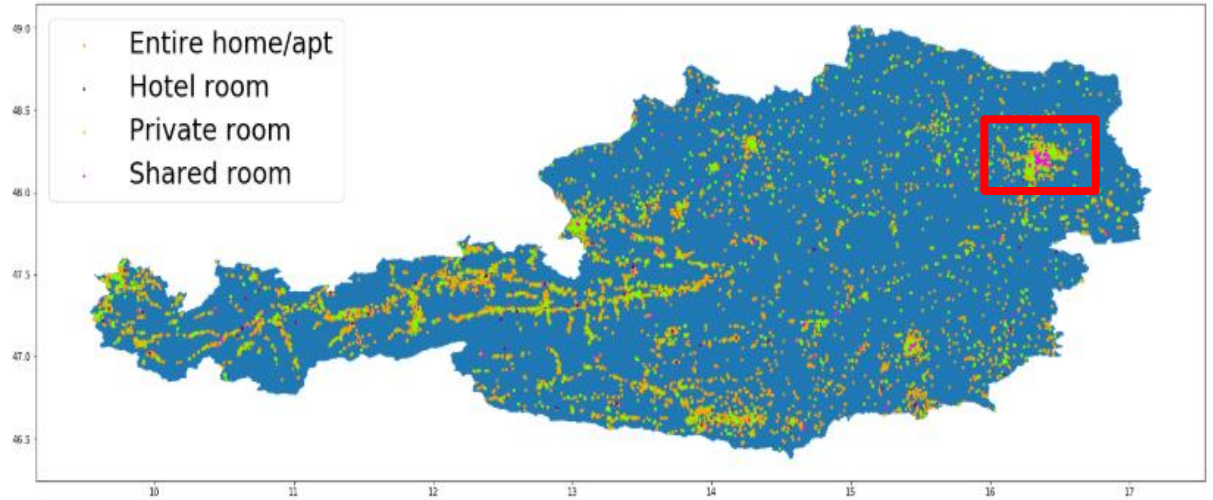


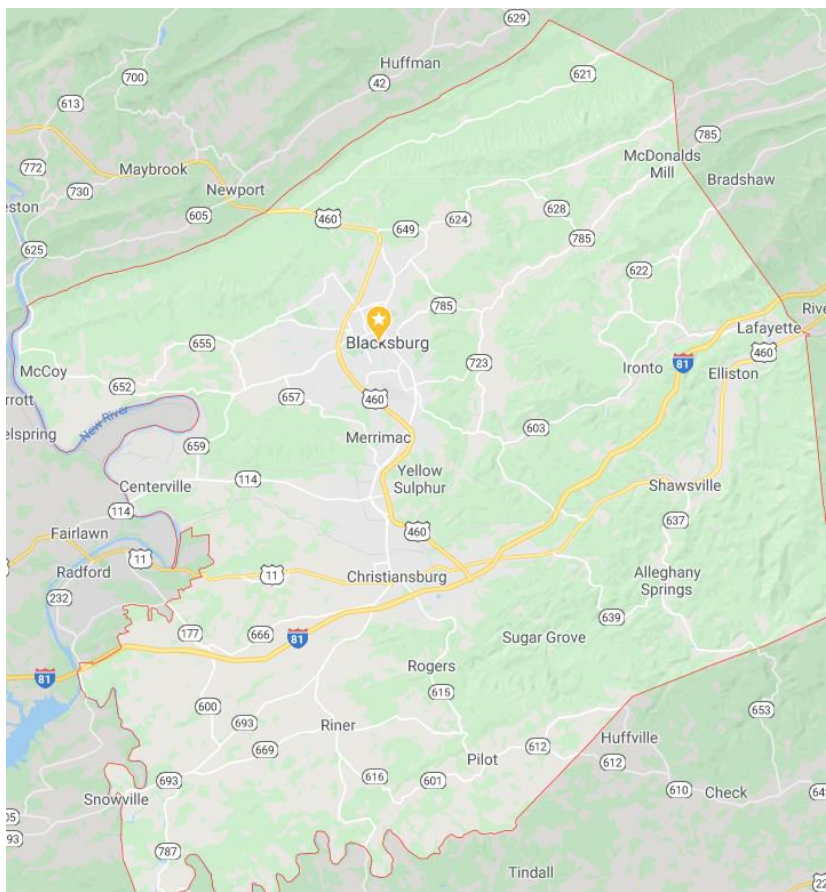


# Geomap

A geographic graph for the room types distribution in Austria

Austria

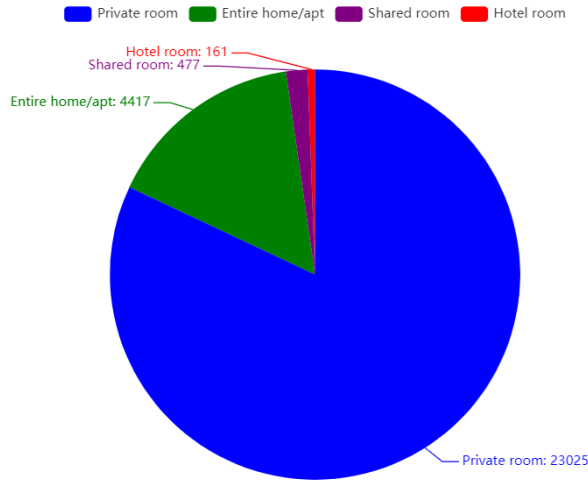




# Geomap – Montgomery Co.

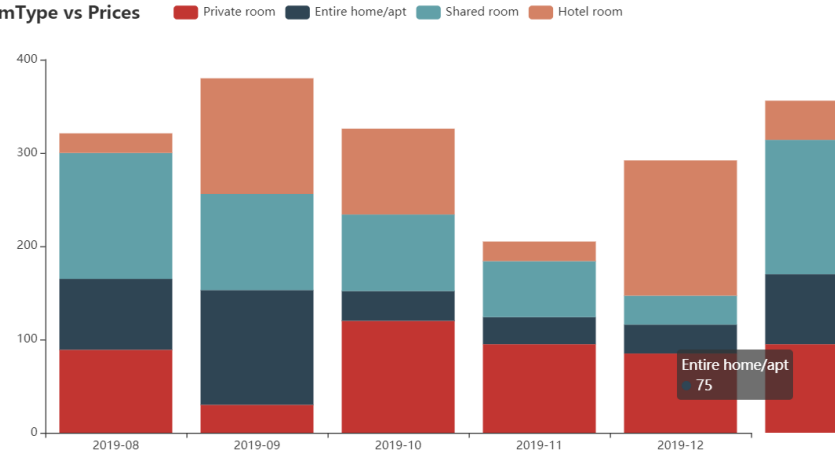


Pie-RoomType

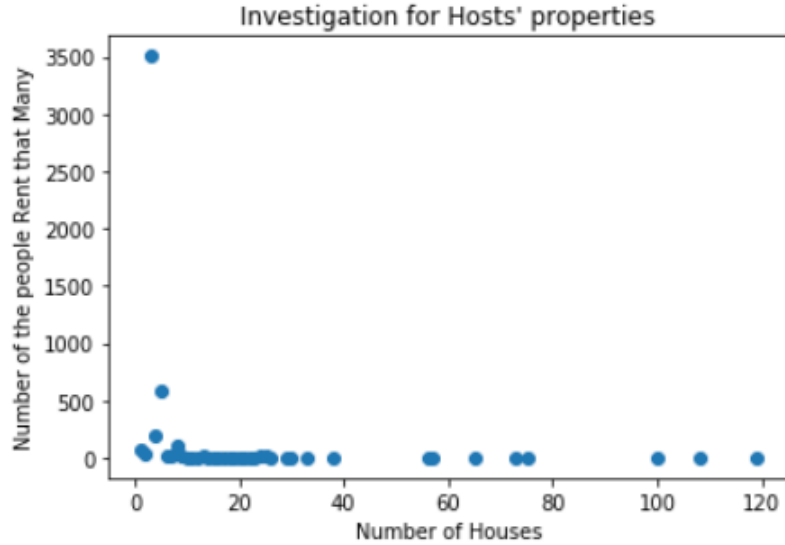


An interactive pie chart of room types

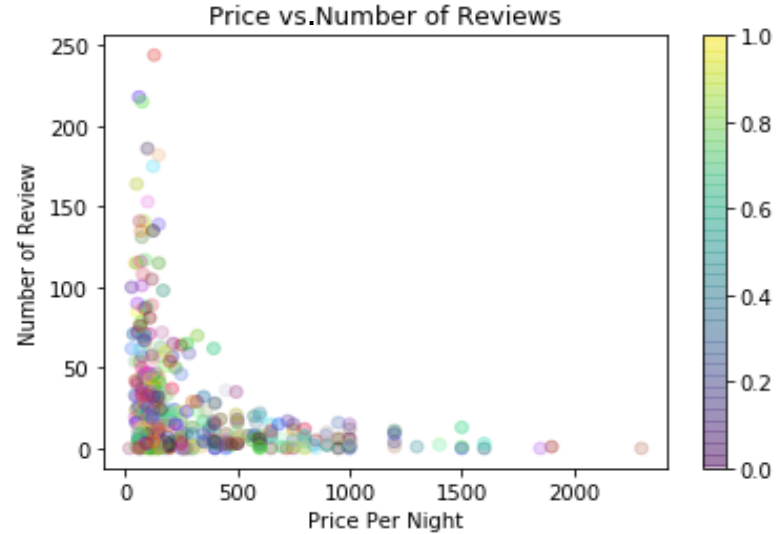
RoomType vs Prices



An interactive stacked bar chart for the representation of room types, prices, and timeline



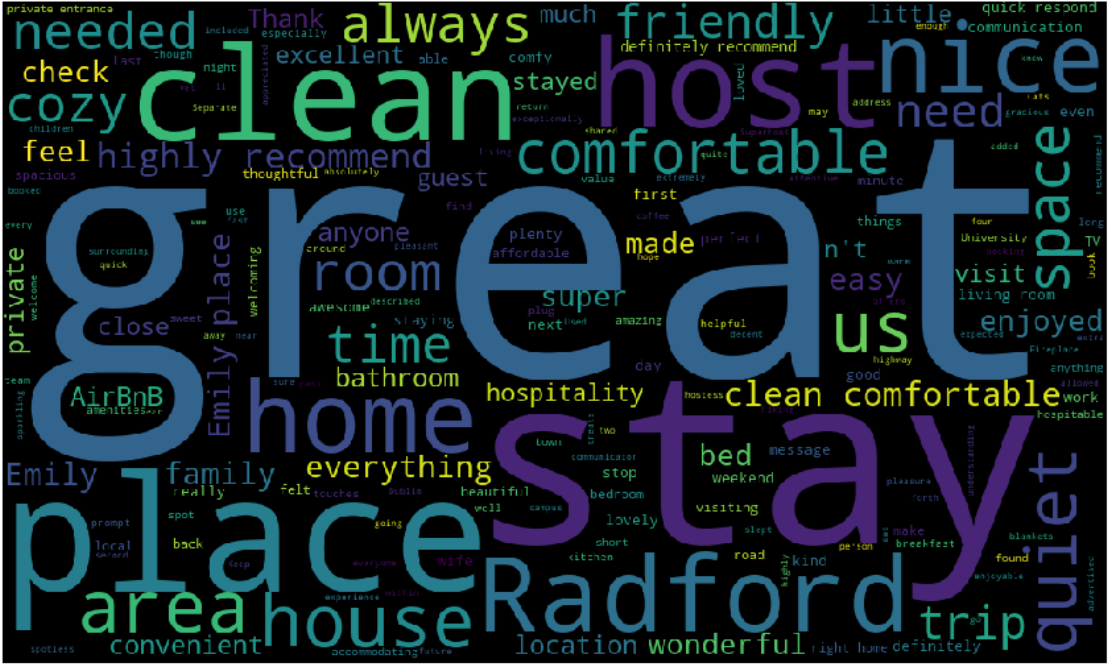
People vs. How many properties they own



The relationship between reviews and prices

A WordCloud map for the reviews from the users for Montgomery County

NLTK, Stop Words, Word Frequency





# Website

- List
- Switch between different county and countries
- Display visualizations



# Website

- Old design



Countries	About Us.	Where does our data come from?
Austria		
United State	<input type="text"/>	<input type="button" value="Search!"/>



State	About Us.	Where does our data come from?
Virginia		
California	<input type="text"/>	<input type="button" value="Search!"/>
...		



# Website

- New Design



Geographic map





# Website

Make a selection ▼

Make a selection

Austria

United States

Make a selection ▼

Make a selection ▼

United States ▼

Virginia ▼

Make a selection

Virginia

California

Illinois

Make a selection ▼

United States ▼

Virginia ▼

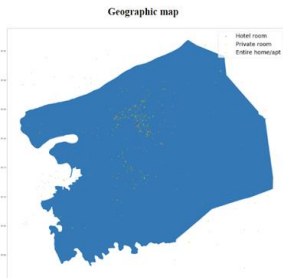
Make a selection ▼

Make a selection

Montgomery

Nelson

# Website



---

# Testing and Assessment

- User interviews
- Benchmark testing
- Surveys



# matplotlib

## Lessons Learned



---

## Ideas for the Future

- Make the website more visually appealing
- Dynamic and interactive maps
- More visualizations
- Collect data on other regions



---

# Acknowledgements

Clients:

Dr. Florian Zach





---

Questions?