Airbnb Visualization Engine

Sagar Jain, Aayush Mandhyan, Akshat Kumar

sj735@rutgers.edu
am2447@rutgers.edu
ak1648@rutgers.edu

May 10, 2019

Abstract

Before AirBnB, it would have been a nerve-wracking prospect to let strangers stay in your home. AirBnB has changed thiswith a mission that connects people with places to stay and things to do around the world. The company has transformed the relationship between the homeowner and the renter. Most of us are familiar with the experience as guests, renting homes to stay in on AirBnB.Keeping this in mind we built, an Interactive Visual Analytics platform which caters to the user needs of finding perfect destination for travelers and helping the business users by mining Airbnb listings and finding the customer pain points and opportunities. Our aim is to present the user with functionality to explore every listing geographically in an interactive clustered fashion, filter down listing for users based on different parameters, analyze listings using various charts, view Demand fluctuation pattern across the year, highlight popular or trending terms based on frequency of use and prominence and finding customer pain points and opportunities for business users to connect.

1 Introduction

Airbnb has seen a meteoric growth since its inception in 2008 with the number of rentals listed on its website growing exponentially each year. Airbnb has successfully disrupted the traditional hospitality industry as more and more travellers, not just the ones who are looking for a bang for their buck but also business travellers resort to Airbnb as their premier accommodation provider. Data plays a vital role for Airbnb. We characterize data as the voice of the users at scale. Thus, data science plays the role of an interpreterwe use data and statistics to understand our users and translate it to a voice that people or machines can understand. We leverage these quantitative insights, paired together with qualitative insights (e.g. in-person user research) to make the best possible decisions for both the business and our community of hosts and guests. (1) Many visualization platform has been developed focussing on various aspects

and gaining insights from Airbnb data but less work has been done in the field of spatial data analysis. Many static map plots have been deployed but interactive cluster maps are still not fully explored which can explain the geographical exploration with the data and make it more appealing. We developed a visual analytic platform to make it more informative and insightful. Following are the features of our platform:

- A rich and extensible set of visualizations including basic charts as well as table structure, world maps, timeseries analysis, word cloud, Topic modelling and more Create and share interactive dashboards as collections of visualizations
- A smooth learning curve: users can be trained in minutes and get value instantaneously
- Flexible data caching, with cascading timeout parameters by report, table and database to relieve your databases from heavy load and to make important dashboards load quickly. (2)
- Customizable and hackable! You can brand and skin AirBnb Analytics with your own bootstrap theme, create CSS templates for your dashboards and modify the controls for specific visualizations

2 Dataset

The dataset comprises of three main tables: listings - Detailed listings data showing 96 attributes for each of the listings. Some of the attributes used in the analysis are price(continuous), longitude (continuous), latitude (continuous), listing_type (categorical), is_superhost (categorical), neighbourhood(categorical), ratings (continuous) among others.
reviews - Detailed reviews given by the guests with 6 attributes. Key attributes include date (datetime), listing_id (discrete), reviewer_id(discrete) and comment (textual).
calendar - Provides details about booking for the next year by listing. Four attributes in total including listing_id (discrete), date(datetime), available (categorical) and price (continuous).
Source: InsideAirBnb.com
DataSet: AirBnb listings
Format Type: CSV format
Number of attributes: 89
Number of sample data: 494,954
Each set of files below for each city
(4)

3 Model Framework

- We collected various categorical data from AirBnb Inside Data platform like review, host and listing data for various states of United States.
- Data wrangling and data preprocessing were performed over this data to make it operable.Data wrangling is the process of gathering, selecting, and transforming data to answer an analytical question. Also known as data cleaning or munging, legend has it that this wrangling costs analytics professionals as much as 80% of their time, leaving only 20% for exploration and modeling.
- Controller: The preprocessed data was passed to controller which takes care of all the logical and conditional attributes of the problem and process results subsequently. We developed various representational cycles inorder to formulate and get the informative insights from the data.
- We used NLP for user review mining in the search of qualitative data that can help and increase app's visibility or quality, word cloud generator inorder to highlight important textual data points and topic modelling which is used for process of identifying topics in a set of documents.
- Clustering listing by geography for spatial data exploration so that distribution of various geographical locations can be visualized and various clustered geographical layers can be explored. (3)
- Visual Analytics Platform consists of a visualization landing page which controls and presents various aspects of analytics for textual analytics, AirBnb trend analysis and listing tab.

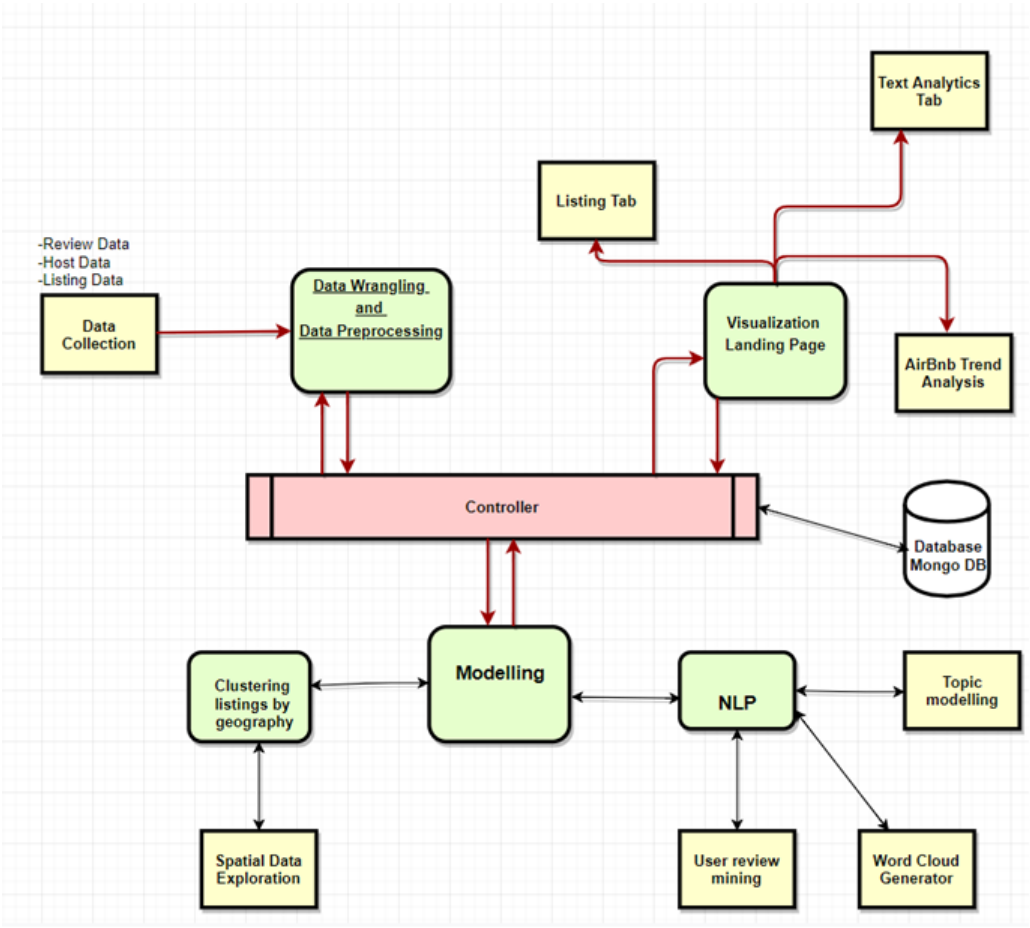

Figure 1: Flow Diagram

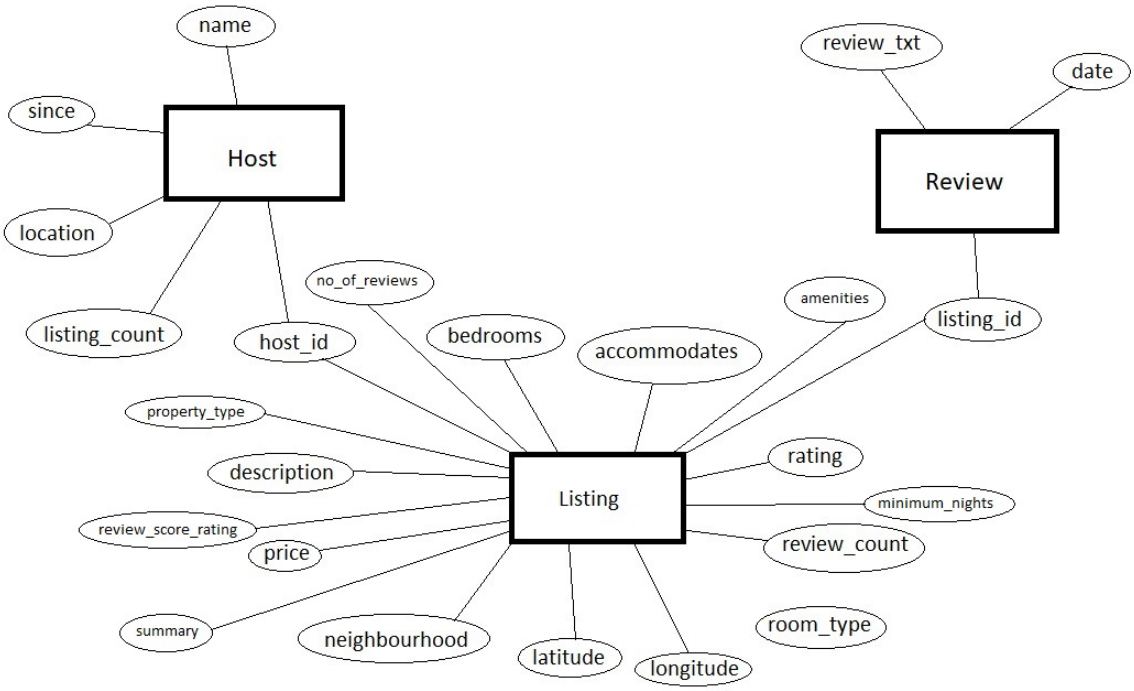

Figure 2: ER Diagram

4 Experiment and Results

In this section, we will detail our analysis to the questions of interest mentioned in the introduction and gain preliminary insights through exploratory data analysis and visualization. We have divided it into four subsections that aim to answer the questions through a variety of different visualization.

4.1 Spatial Data Analysis

- Interactive Map
  This is the basic interactive graph with all the listing in United States appearing in a clustered fashion. You can click on clusters to see the listing it comprises. This gives a zoom-in view. You can further click on each listing to see details like Listing Name, Host Name, Price of the property, Property Type, Room Type. This visualization helps to explore every listing geographically. It gives the overall sense of how the listings are distributed across neighborhood. We can see from the map that maximum

listing are clustered around Manhattan and Brooklyn region, followed by Queens, Bronx and the least number of listing are in Staten Island. (5)

- Listing, Boroughs, and Price Changes
  To explore detailed information in each borough, users can slide the price and rating score bar to check which borough shows more listings that meet their expectations. Last but not least, finding a great time to book your room is another factor that travelers may consider. By choosing the price change per month, you can see that the average price drops a little bit in July.

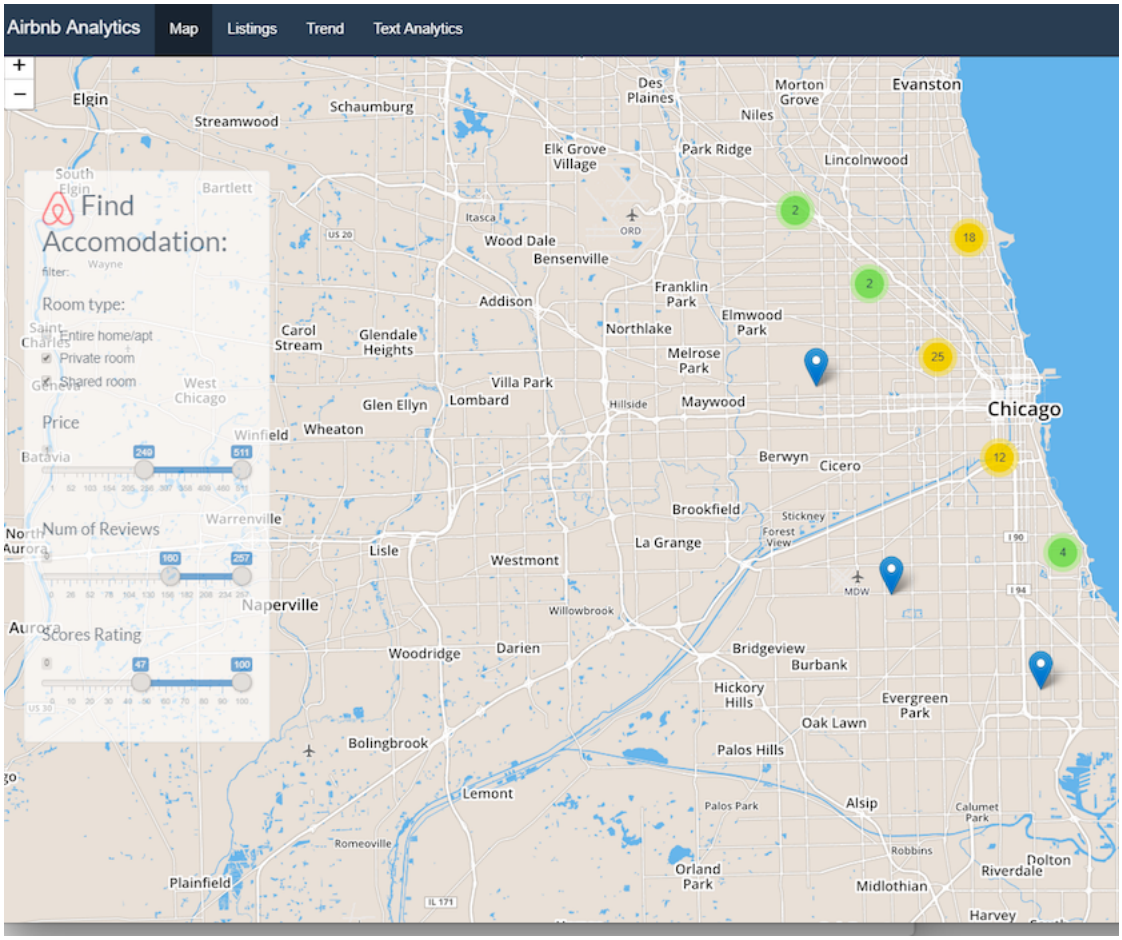

Figure 3: Spatial Airbnb listing Distribution

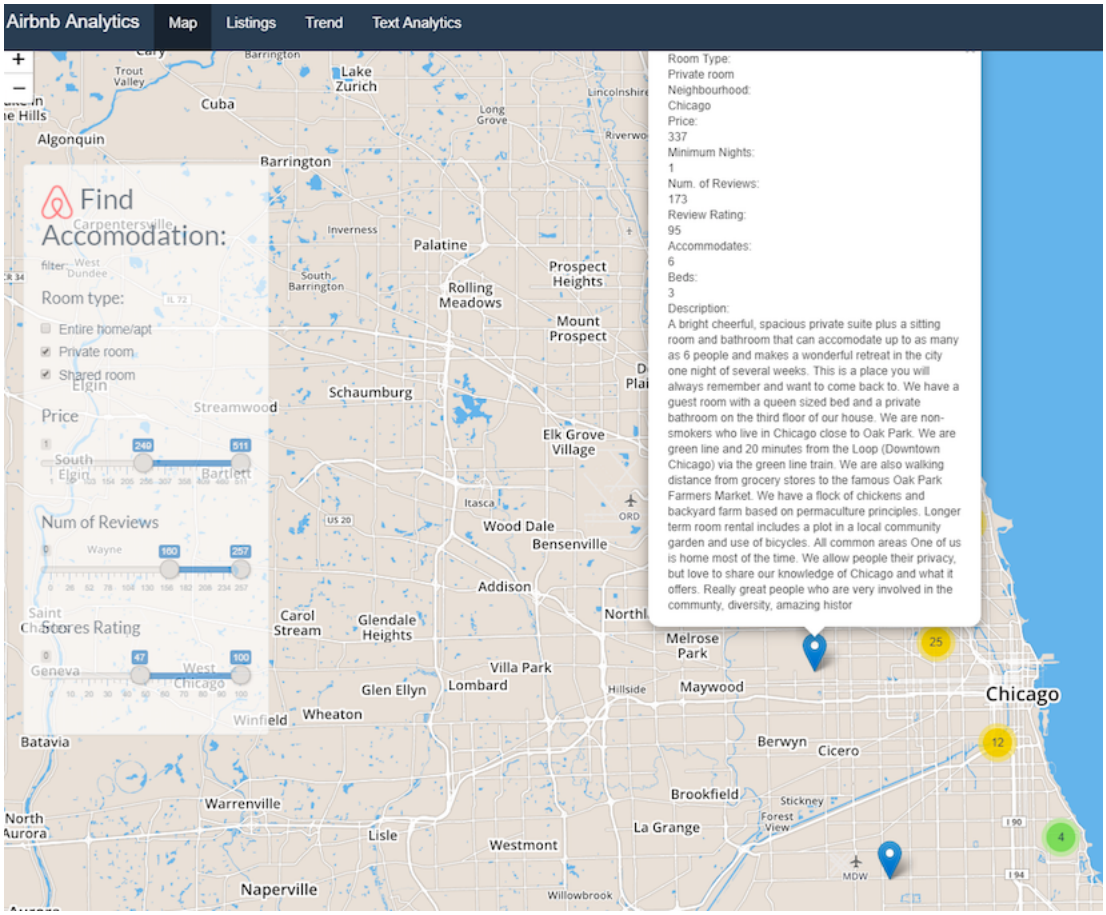

Figure 4: Individual listing information

4.2 Demand and Price Analysis

In this section, we will analyse the demand for Airbnb listings in New York City. We will look at demand over the years since the inception of Airbnb in 2009 and across months of the year to understand seasonlity. We also wish to establish a relation between price and demand. The question we aspire to answer is whether prices of listings fluctuate with demand. We will also conduct a more granular analysis to understand how prices vary by days of the week. To study the demand, since we did not have data on the bookings made over the past year, we will use number of reviews variable as the indicator for demand.

As per Airbnb, about 50% of guests review the hosts/listings, hence studying the number of review will give us a good estimation of the demand.

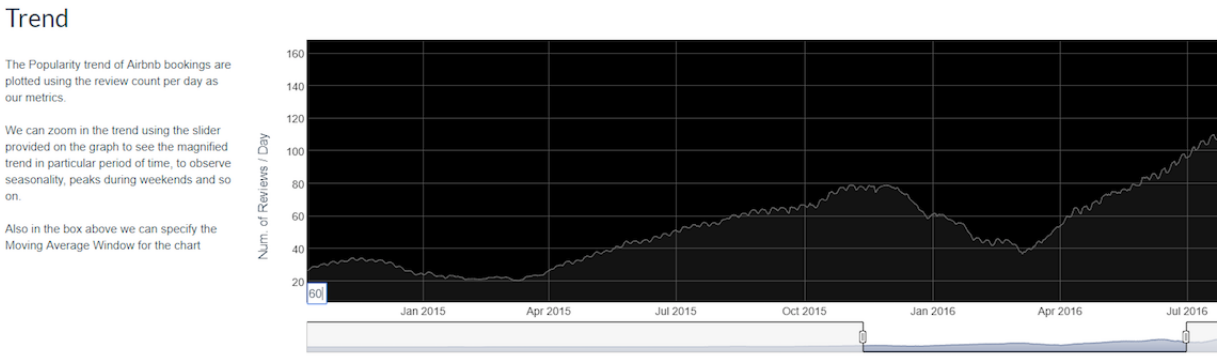

Figure 5: Airbnb Review Distribution over the years

4.3 User Review (Textual Data) Mining

The dataset provides us with a ton of data, but nothing as insightful and close to the customer as their reviews/feedback. If mined properly, they can tell us a lot about the customer mindset, their expectations and how well those were met. There are more than a million reviews, so I take a random sample of this data, in this case 30k reviews.

An analysis of the word cloud shows interesting trends; Location seems to be key, since the words neighbourhood, location, area are featured prominently in the word cloud. Transport options like subway, walk also find frequent mention. Airbnbs are short term rentals, yet people seem to lay stress on the comfort aspect of their stay, words like kitchen tell us that many folks would rather cook than eat out. The word Host finds a lot of mention; indicating the important role that hosts play in shaping the Airbnb experience.

The first word cloud is for the word uncomfortable. Words similar to uncomfortable are usually those that occur in conjunction with it frequently, i.e reasons for the discomfort. The word cloud shows just thatnotice words like cramped, crowded,small, stuffy and cluttered indicating that lack of space is one of the most common complaints. Hot, damp and cold are some of the common temperature issues. Dusty, dirty and unclean surroundings will prompt people to write negative feedback. Many feel nervous,unsafe and stressful; clearly a red flag for future tenants.

Similarly, querying by the keyword comfortable, we hope to see things that led to a positive experience. Prominently featured are words like quiet, walkable,

clean, spotless, etc, again demonstrating the importance of environment, location and cleanliness. Helpful hosts and communication lead to a comfortable. Cleanliness of linens and the bed size leave a decisive impression.

Figure 6: Wordcloud on User Reviews

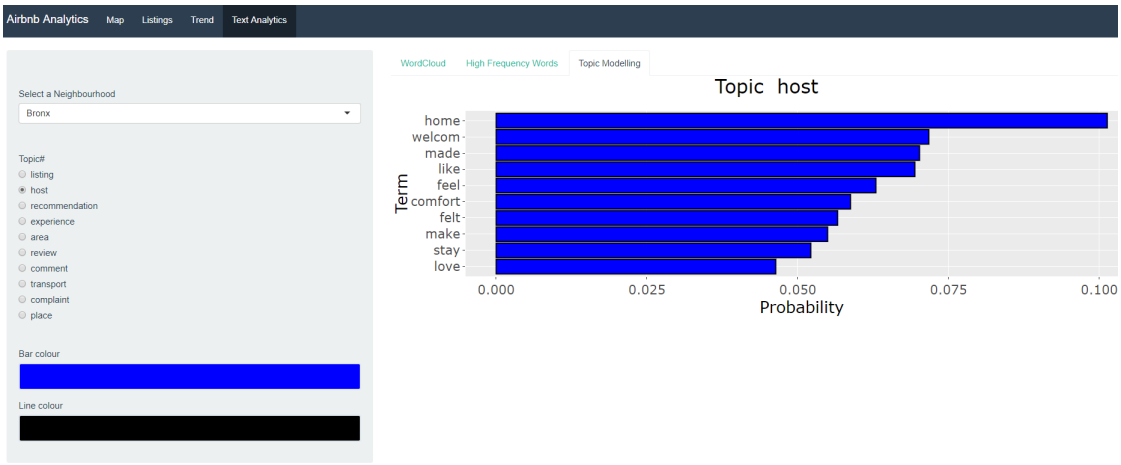

Figure 7: Topic Modelling on User Reviews

4.4 Other Interesting Insights

- More than 80% of Airbnb properties in New York City are outside the mid-Manhattan area especially in East Village and Nolita.
- Brooklyn is a great choice for cheaper cost per night and higher rating score. Therefore, if you are out of time and want to narrow down your options, Brooklyn can be put into your consideration.
- According to the previous booking price (from May 2015) and future listing price (to May 2018), we know that the overall price for Airbnb NYC drops. There are some reasons why. First of all, the New York State Senate passed new laws against short-term rental in 2016, which was a huge blow for Airbnb. On top of this regulation issue, the growth of competition can be another explanation for the price drop. Thirdly, there are more private rooms and shared rooms posted on Airbnb, which are much cheaper than entire apartments. The increase of private rooms and shared rooms contribute to a drop in price.
- There are price peaks in May, June, August, and December. However, during the summertime, there is a dip in July. (6)

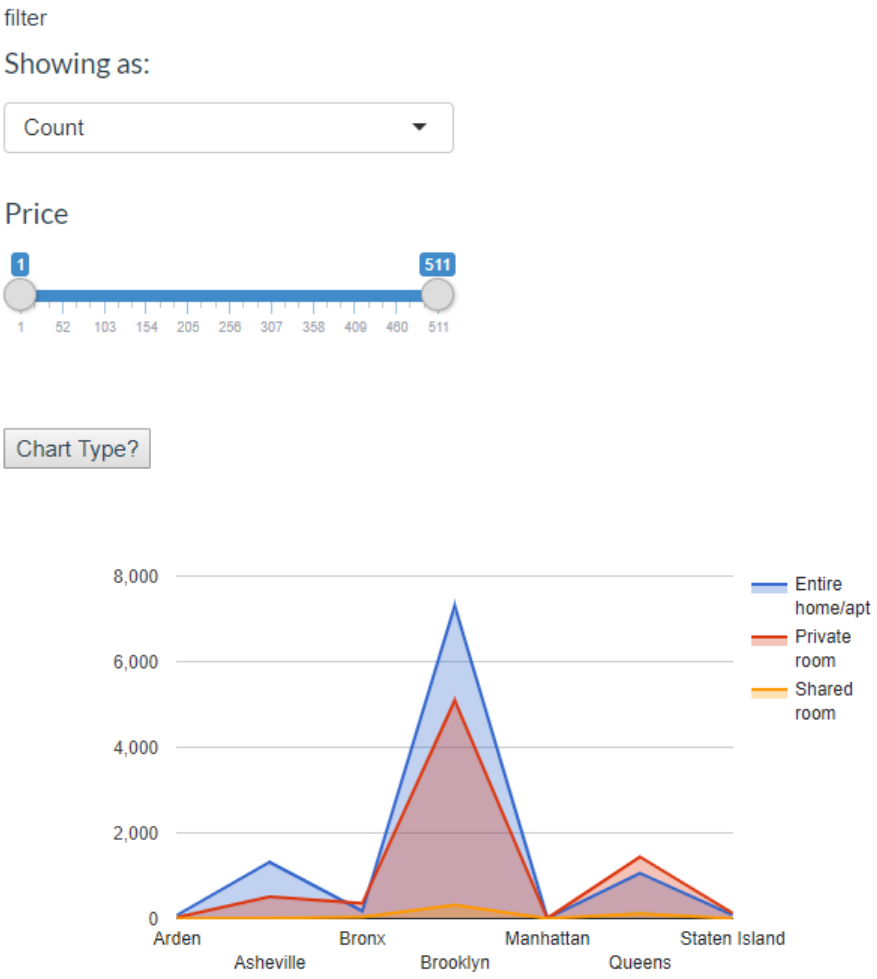

Figure 8: Area Chart of distribution of listings

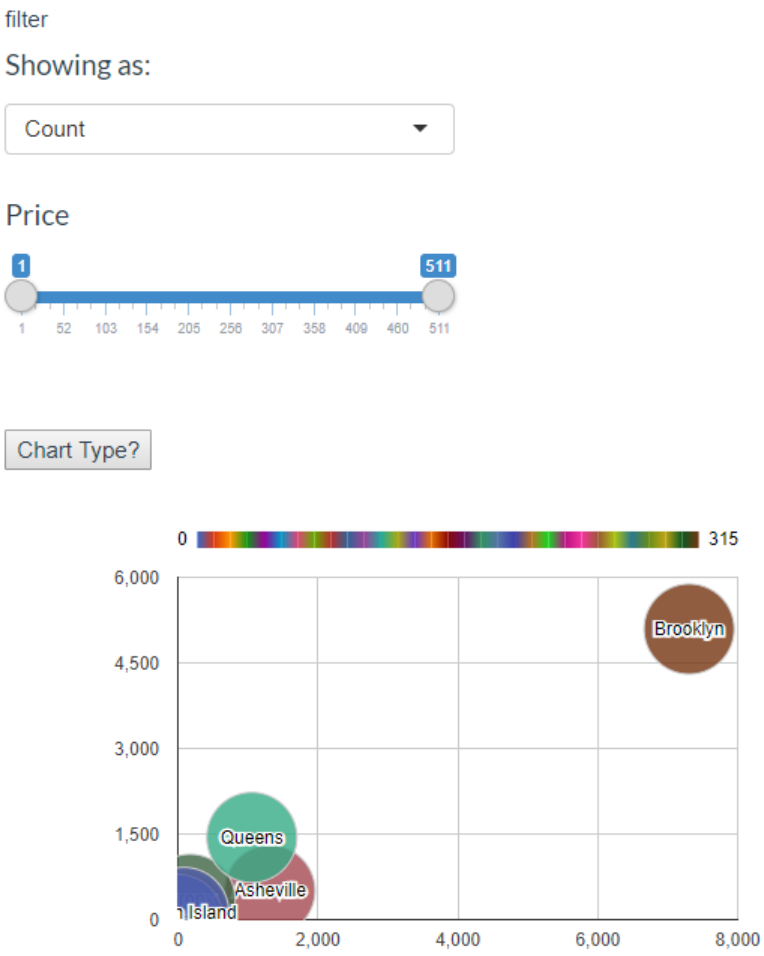

Figure 9: Bubble Chart of distribution of listings

Figure 10: Bar Chart of distribution of listings

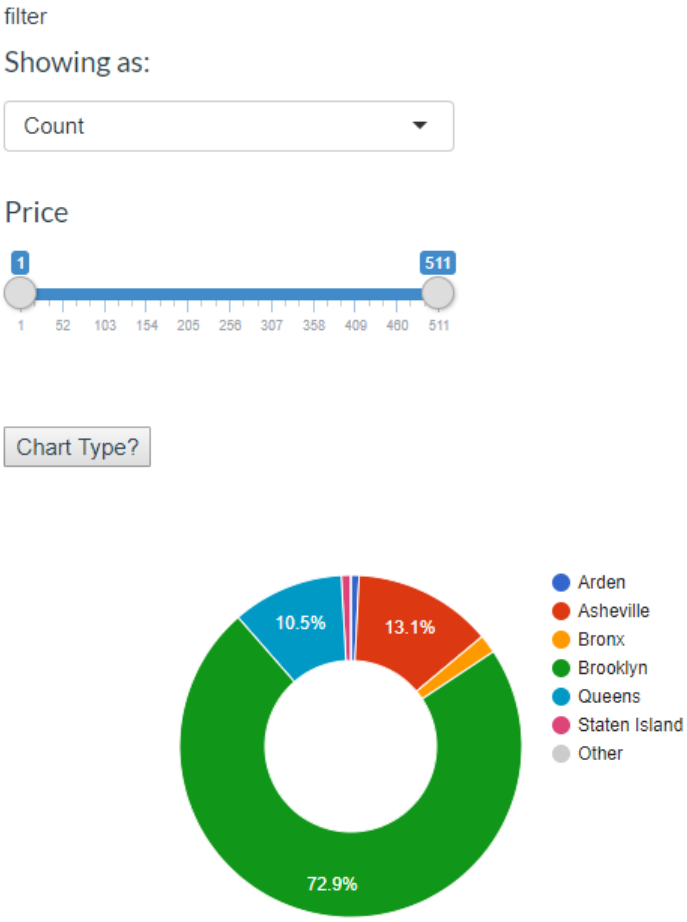

Figure 11: Donut Chart of distribution of listings

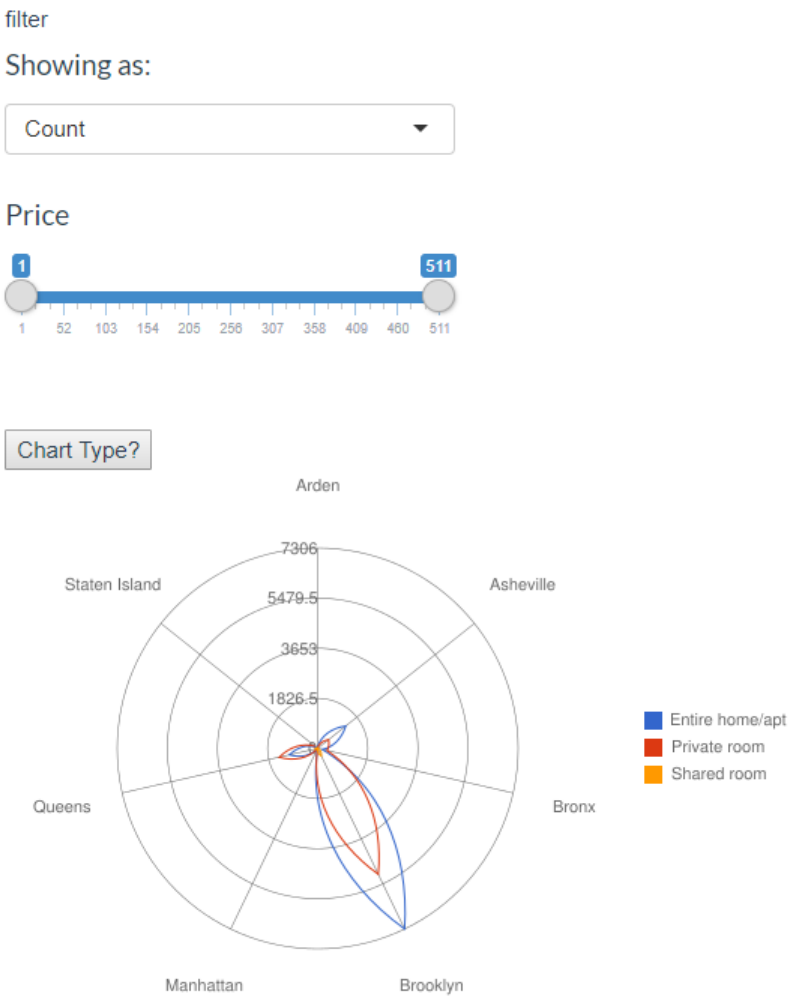

Figure 12: Radar Chart of distribution of listings

5 Conclusions

Through this exploratory data analysis and visualization project, we gained several interesting insights into the Airbnb rental market. Below we will summarise the answers to the questions that we wished to answer at the beginning of the project:

- How do prices of listings vary by location? What localities in United States are rated highly by guests? Manhattan has the most expensive rentals compared to the other boroughs. Prices are higher for rentals closer to city hotspots. Rentals that are rated highly on the location by the host also have higher prices. There are a few outliers in Bronx, Staten Island and Brooklyn that defy the above hypothesis.

- How does the demand for Airbnb rentals fluctuate across the year and over years? The demand (assuming that it can be inferred from the number of reviews) shows a seasonal pattern - demand increases from January to October, then drops slightly in November and December. In general, the demand for Airbnb listings has been steadily increasing over the years.

- Are the demand and prices of the rentals correlated? Average prices of the rentals increase across the year, which correlates with demand. However, the prices show a spike in December as opposed to slowing demand in this month, which is counterintuitive. Prices are higher on average on Fridays and Saturdays, compared to the other days of the week.

- What are the different types of properties in United States? Do they vary by neighborhood? There are more than 20 different types of listings in United States. The ratio of the type of listings to total numbers varies by borough. Staten Island and Brooklyn tend to have property types that are larger and can accommodate more number of people.

- Are there any common themes that can be identified from the free-text section of the reviews? What aspects of the rental experience do people like and what aspects do they abhor? There are certain words such as words such as quiet, walkable, clean, spotless that are associated with the word comfortable demonstrating the importance of environment, location and cleanliness. Words associated with uncomfortable include cramped, crowded,small, stuffy cluttered which indicate that lack of space is one of the most common complaints.

5.1 Suggestion for AirBnb, host and guests

- Airbnb: New York City is the biggest market for Airbnb. According to 2017 data, the estimation of Airbnb NYC's market value is over 27 million. In New York City alone, Airbnbs profitability is about 3.24 million dollars. For expanding the market and retaining users, Airbnb not only has to deal with regulatory problems but also needs to keep improving user experiences.

- Hosts: For hosts who post up their listings in competitive areas, it is always smart to keep looking up others price. They can also consider offering some extra service such as offering breakfast, renting bikes and so forth to attract more guests.

- Guests: Guest should be aware of the differences in timing and location. During summertime, July is a great time to place an order. For highly-rated yet budget-friendly listings, Brooklyn appears to be the best bet

6 Future Scope

We want to expand our analysis to multiple cities and compare patterns and trends amongst these cities. From the insights we have derived, we would also like to build predictive models using different features from the dataset. Lastly, we hope to implement the visualizations and techniques used in this project to many other fields and datasets.

7 Acknowledgement

We would like to express our very great appreciation to Professor James Abello and Alireza Naghizadeh Nohadani for their valuable and constructive guidance during the development of this project.

References

[1] Abello, J., Ham, F.V., & Krishnan, N. (2006). ASK-GraphView: A Large Scale Graph Visualization System. IEEE Transactions on Visualization and Computer Graphics, 12, 669-676.

[2] Abello, James. (2004). Hierarchical Graph Maps. Computers & Graphics. 28. 345-359. 10.1016/j.cag.2004.03.012.

[3] Abello, James & Hohman, Frederick & Bezzam, Varun & Chau, Duen Horng. (2018). Large Graph Exploration via Subgraph Discovery and Decomposition.

[4] http://insideairbnb.com/get-the-data.html

[5] https://shiny.rstudio.com/gallery/

[6] https://cran.r-project.org/web/packages/googleVis/vignettes/googleVisexamples.html