Choosing the Best Neighborhood to Rent an Airbnb in NYC

1. **Introduction**

1.1. Background

More and more people choose to stay in Airbnbs when they decide to stay in a foreign city, be it for tourism or work. When they do, there are a number of factors to consider when choosing where to stay. Airbnb provides users with maps of the different Airbnbs at their disposal, but especially when entering a new city, there is little a person knows about a given neighborhood. Of course, distance to the city center and price are important, but it is just as important to many people that the type of neighborhood they stay in fits their requirements. That type of knowledge, however, is much harder to retrieve on Airbnb's website. This is where this project enters the stage.

1.2 Problem

With this project, I want to add a layer to the Airbnb search that is typically harder to include: the types of venues and amenities surrounding the neighborhood the Airbnb is located in.

1.3 Interest

The project is interesting due to the ubiquity of Airbnb in travel and the clear benefit it can provide to travelers.

2. Data acquisition and cleaning

2.1 Data sources

I use two different data sources in my project. On the one hand, I use the Foursquare.com Places API to identify venues around NYC. This data encompasses the types of venues in the New York City area, the name of the establishment, the location of the venue, and other aspects that we will not use for this particular project. I also use a public data set provided by Airbnb and posted on Kaggle.com for Airbnb listings in New York City in 2019. This dataset provides the name of the listing, the location, the neighborhood it is located in, the price, number of reviews and other aspects to an Airbnb listing.[1]

---

[1] https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data

2.2 Data cleaning

I first create a dataset of venues in New York City based on the neighborhood they are located in. To do this, I parse through each neighborhood and find recommended venues in that neighborhood. The airbnb data does not have the exact same classification of neighborhoods as the neighborhood data I use to identify the longitude and latitude of each neighborhood. I therefore merge the two neighborhood lists to account for this.

I will use this data in the following way:

- Use Foursquare Data to cluster neighborhoods based on venue.
- Use airbnb data in New York City to look at these clusters and compare the airbnbs based on quality and price.

3. Methodology

3.1 Exploratory Data Analysis

Figure 1 shows the Airbnbs as listed on the dataset. We can see that there are quite a couple of listings. Overall, there are almost 50,000 listings. Figure 2 shows the descriptive statistics for the Airbnb data. We can see that prices average at around 150 dollars a night, with the lowest quartile ending at around 70 dollars and the highest quartile starting at 170 dollars. We can also see that the average number of reviews is at around 23.

Figure 1: Airbnbs in New York City

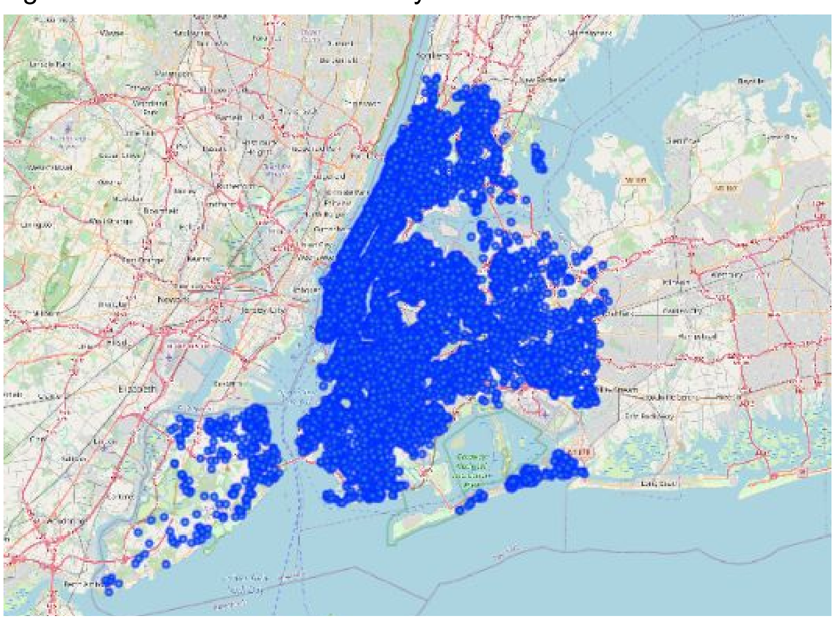
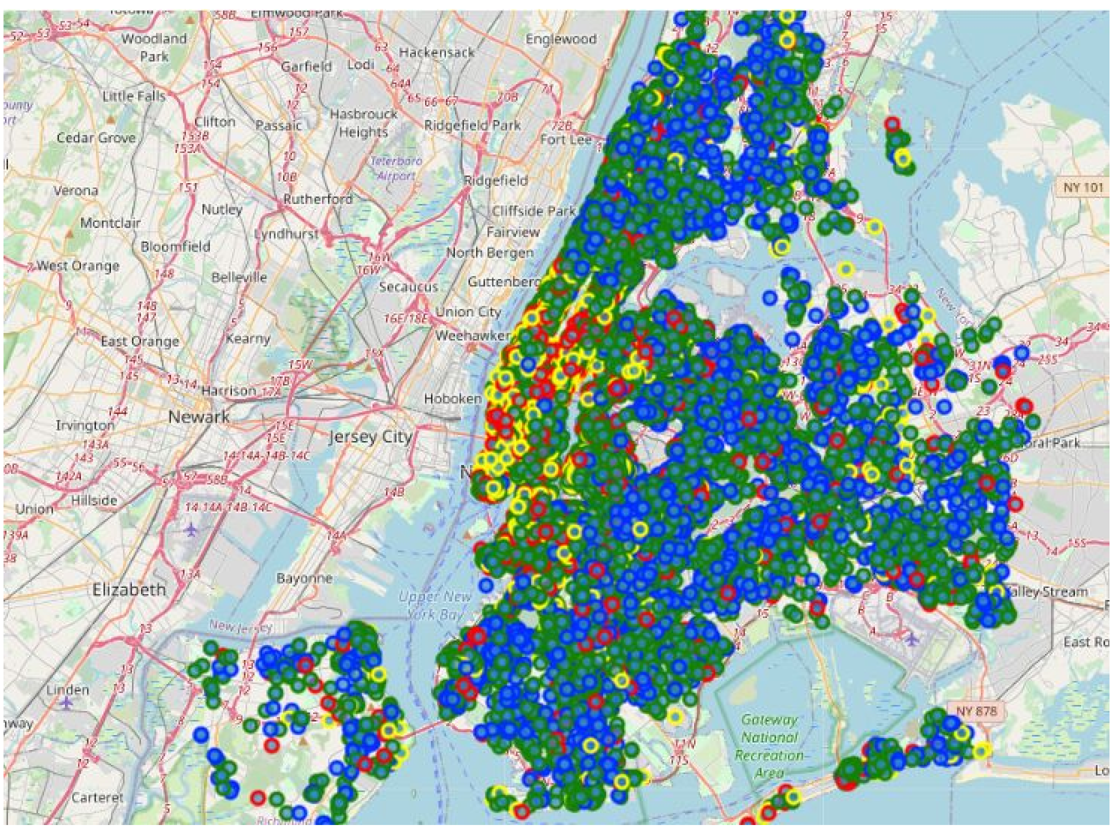


Figure 2: Descriptive Statistics

	price	minimum_nights	number_of_reviews	availability_365
count	48895.000000	48895.000000	48895.000000	48895.000000
mean	152.720687	7.029962	23.274466	112.781327
std	240.154170	20.510550	44.550582	131.622289
min	0.000000	1.000000	0.000000	0.000000
25%	69.000000	1.000000	1.000000	0.000000
50%	106.000000	3.000000	5.000000	45.000000
75%	175.000000	5.000000	24.000000	227.000000
max	10000.000000	1250.000000	629.000000	365.000000

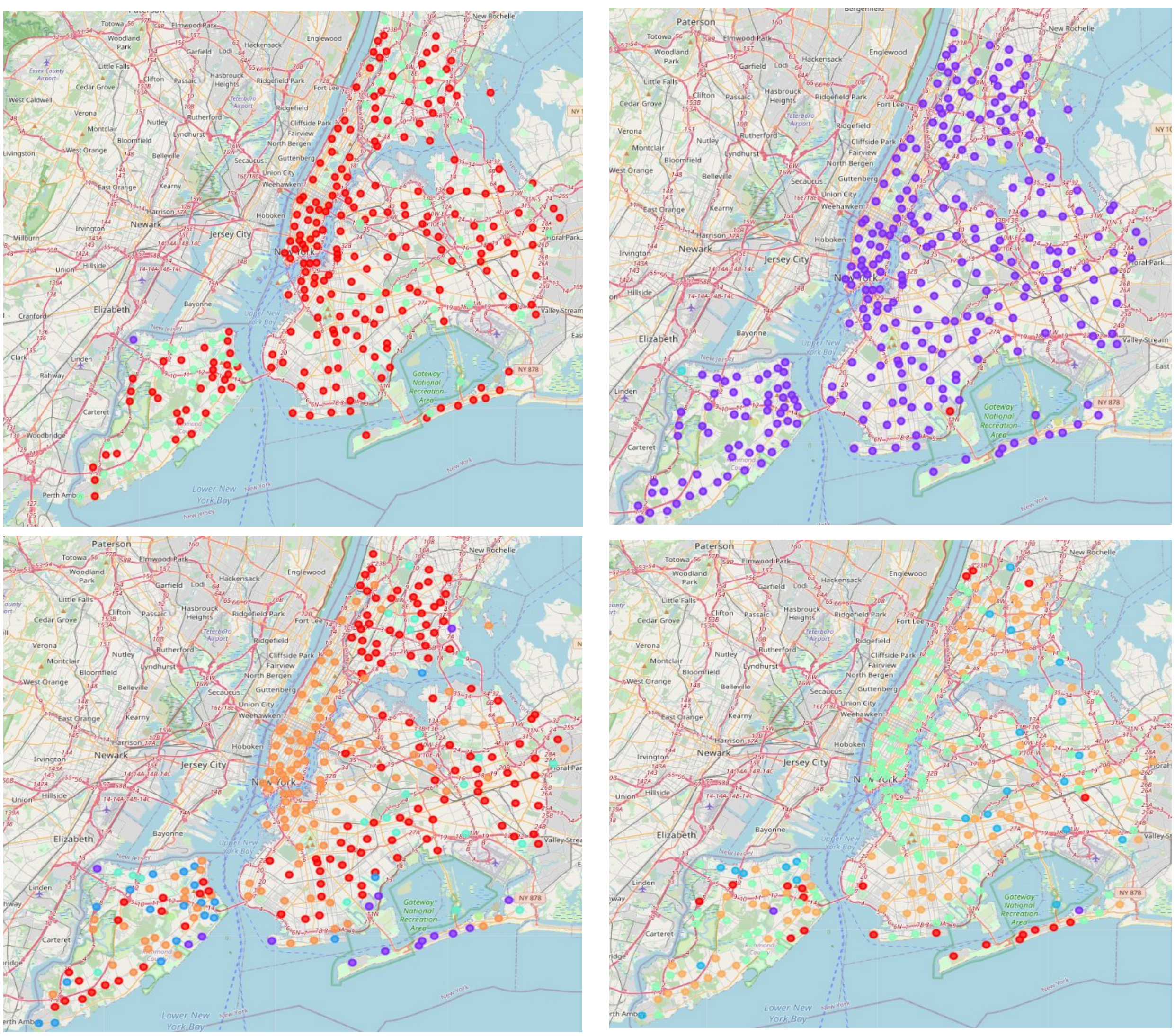
We can also explore the listings based on price geographically. The yellow points represent listings that are priced at over 250 dollars a night. Red points represent listings over 170 dollars, the highest quartile, green points represent values between 70 and 170 dollars and blue points represent the lowest quartile, listings under 70 dollars. It is clear that as we move away from the center, the listings become cheaper. But there are also clusters outside of the center that vary in price.

Figure 3: Listings Based on Price



3.2 Modeling

In order to decide in which neighborhoods on should search for Airbnbs, we have to first cluster the neighborhoods based on venue type. I choose to use k-means modelling to cluster my data. In that respect it is important to determine the best amount of clusters to pick in order to find meaningful categories. Below, we can see a number of different amounts of clusters to show the effect different clusters have on the categorization. The images are aligned based on cluster sized, starting with a cluster of three on the left upper corner and ending with a cluster of six in the right lower corner. After analyzing the different clusters, I determine that five clusters is an appropriate number to pick. We can see this in the images. Clusters of three and four don't categorize enough and the cluster of six seems to over-categorize.



4. Results

Clustering the dataset of the different venues based on neighborhood gives us the following categories: parks, beaches, a cluster centered around delis and bodegas and the likes, a cluster with a versatile range of venues including restaurants, shops and amenities and a last cluster that seems to be more residential with restaurants and amenities such as gyms. I decide to focus on the third category, since it appears to be the most versatile cluster with a range of venues that might attract tourists. Examining the map above, that would be the orange spots on the lower left image. We can see that the dots are spread around the map, but a majority of them are clustered in the center and downtown Manhattan as well as downtown Brooklynn. Going with this category of clusters, we now look at the best option for housing. I decide to limit the analysis based on the amount of reviews there are on average for the neighborhood and what the average minimum stay is. Then, I identify the neighborhood that has the cheapest housing options on average.

5. Discussion

There are a number of nteresting insight into the market of short-term hosuing options we can make from this study. Firstly, Ishowed that airbnb prices are not somewhat dependent on distance to the center of Manhattan. However, there are different areas where prices hike outside of this typciallu tourist heavy center. I also showed that the types of surroundings a tourist might find attractive, a wide variety of venues including stores, restaurants and other amenities, can be found outside of said center, possible allowing for cheaper housing with similar benfits.

There are a number of ways in which my analysis could be improved. As of now, I am not factoring in the distance to certain attractions or specific neighborhoods. Another aspect I am not taking into account with this analysis is the proximity to public transportation.

Of course, I could have chosen different requirements to make my choice. The cheapest option is not necessarily the one a tourist wants to go with and there might be different needs and standards for length of stay and how many people have reviewed the

6. Conclusion

In conclusion, this report introduces a different angle in choosing short term housing for stays in foreign cities by analysing the types of venues that can be found in different neighborhoods of the city. By categorizing the different types of neighborhoods and picking the category that fits the needs of tourists best, we can identify the best housing options within that type of neighborhood.