

Explore the Impact of Location to Airbnb Price in New York City

Ting Liu
October 2019

1 Introduction

1.1 Background

Airbnb Inc. is an online marketplace for arranging or offering lodging, primarily homestays, or tourism experiences. The company does not own any of the real estate listings, nor does it host events; it acts as a broker, receiving commissions from each booking.

Airbnb provides a platform for hosts to accommodate guests with short-term lodging and tourism-related activities. Guests can search for lodging using filters such as lodging type, dates, location, and price. Guests have the ability to search for specific types of homes, such as bed and breakfasts, unique homes, and vacation homes. Before booking, users must provide personal and payment information. Some hosts also require a scan of government-issued identification before accepting a reservation. Guests have the ability to chat with hosts through a secure messaging system. Hosts provide prices and other details for their rental or event listings, such as the allowed number of guests, home type, rules, and amenities. Pricing is determined by the host, with recommendations from Airbnb. Hosts and guests have the ability to leave reviews about the experience.

In order to choose an appropriate listing price, Airbnb suggests to "search for comparable listings in your city or neighborhood to get an idea of market prices." So, the city or neighborhood, or in other word, the location is one of the determinant component of the price.

1.2 Problem

Tourists might know in general that lodgings in Manhattan are more expensive than Queens, but there are 32 neighbourhoods in Manhattan, the lodging price of some are more expensive than others. By using Foursquare data, I would like to explore the vendors of the neighbourhoods in New York City, in order to find out which neighbourhoods are more expensive and why. By analyzing the New York City Airbnb data and Foursquare data, I am trying to find out the similarities of the neighbourhoods and what features of the location might have impact on the Airbnb price.

1.3 Interests

This study will be a good support of Airbnb pricing strategy. For future hosts, they will get a better idea of how much the location worth based on the surrounding vendors. Airbnb can also use the findings as part of their pricing suggestion. So even no comparable listings in the neighborhood exist, Airbnb will be able to suggest a price using the location data.

2. Data

The New York City Airbnb Open Data is from Kaggle.com. This dataset describes the listing activity and metrics in New York city for 2019. This data file includes information about names, hosts, locations, availability, number of reviews etc. The vendor data will come from Foursquare.

The dataset contains information of 48895 rows and 16 columns. Each row represents an Airbnb listing, and the meanings of the columns are as list:

- Id:listing ID
- name: name of the listing
- host_id: host ID
- host_name: name of the host
- neighbourhood_group: location
- neighbourhood: area
- latitude: latitude coordinates
- longitude: longitude coordinates
- room_type: listing space type
- price: price in dollars
- minimum_nights: amount of nights minimum
- number_of_reviews: number of reviews
- last_review: latest review
- reviews_per_month: number of reviews per month
- calculated_host_listings_count: amount of listing per host
- availability_365: number of days when listing is available for booking

3. Methodology

In order to solve the problem, exploratory analysis will firstly be applied to analyze each variable and the possible relations between some variables. Then, cluster analysis will be used to segment the listings basing on the price and location. Finally, Foursquare location data will be introduced to explore the venue differences among the clusters.

3.1 Exploratory Analysis

3.1.1 Single Variables

In general, there are 48895 Airbnb listings in New York City, listed by 11452 hosts. More than 50% of the hosts list only 1 space, but the host who lists the most spaces has 327 spaces.

The listings are located in 5 neighborhood groups or 221 neighborhoods, the name and the number of listings in each neighborhood groups are:

Neighborhood group	Number of listings
Manhattan	21661
Brooklyn	20104
Queens	5666
Bronx	1091
Staten Island	373

There are 3 room types. The numbers of listing in each room types are:

Room type	Number of listing
Entire home/apt	25409
Private room	22326
Shared room	1160

The average price of all the listings is \$153, with the lowest price of 0 and the highest price of \$10,000.

On average, the listings require 7 nights minimum. However, 50% of them require only 3 days or less. The longest minimum nights requirement is 1250 nights.

Each listing has 23 reviews on average. 25% of listings have only 1 or 0 review. 50% of listings have 5 or less reviews. The listing with the most reviews has 629 reviews. On average, each listing has 1.37 reviews per months.

The dataset contains reviews posted no later than June 23rd, 2019.

On average, the listings are available 112 days in a year. From the variable 'availability_365' we can see , more than 25% of listings in the dataset has 0 available days. But some of them are available all the 365 days.

3.1.2 Price and Room Types

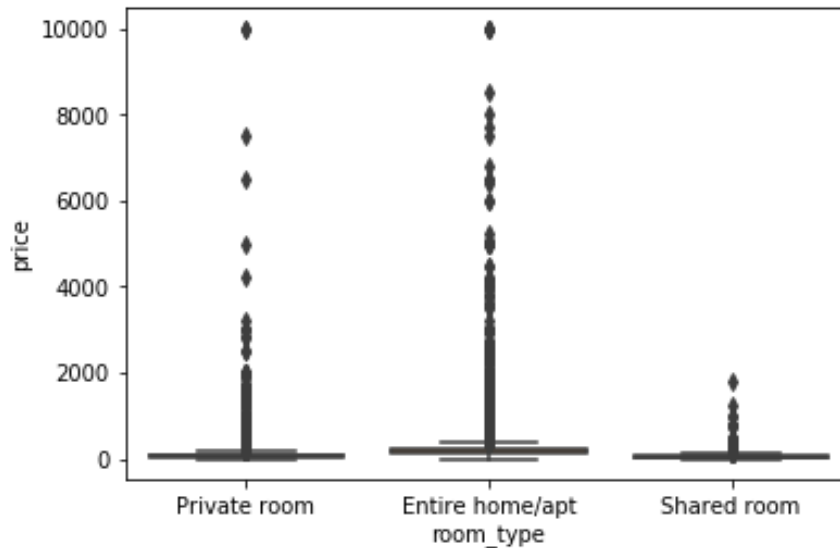
The average listing price for each room type are:

Room type	Average price
Entire home/apt	\$211.79

Private room	\$89.78
Shared room	\$70.13

The average price of entire home/apt is the highest, followed by private rooms, and the average price for shared rooms is the lowest.

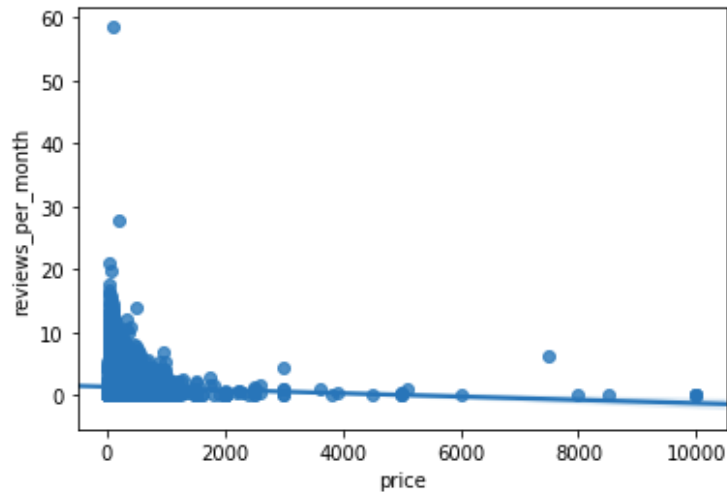
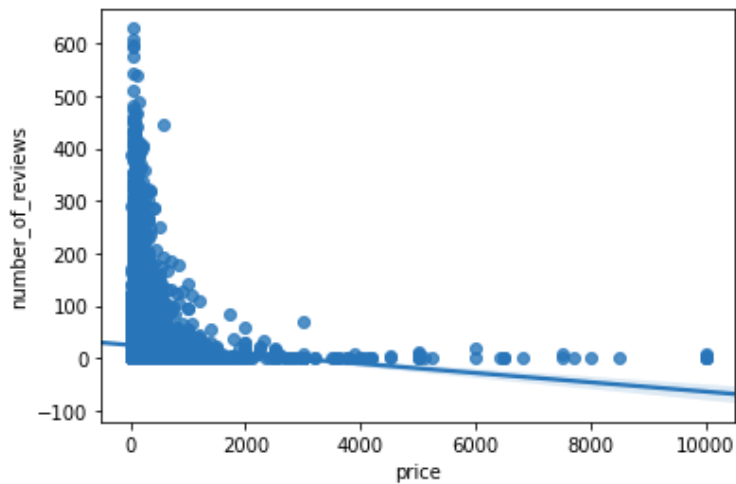
The price variations of the three room types are:



From the chart above, we can observe that the prices of Entire home/apt are more distributed than the other 2 space types.

3.1.3 Price and Reviews

We will use regression plots to find possible correlation between the price and number of reviews that the listing received.



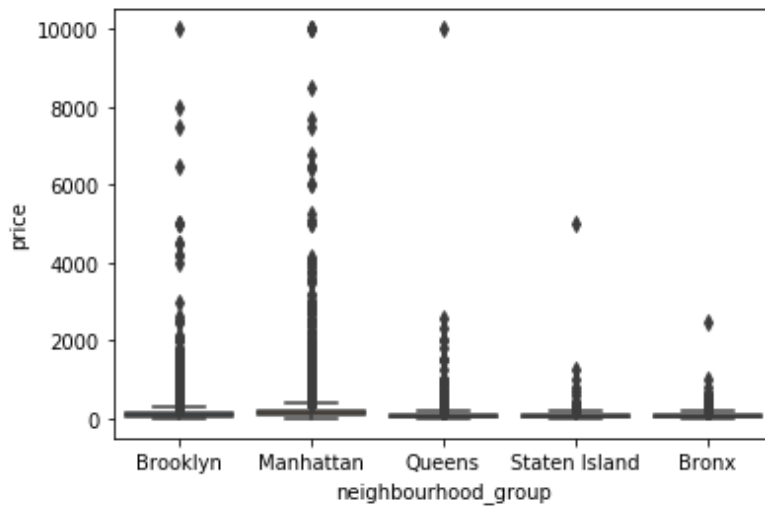
From the charts above, the number of reviews slowly decreases as the listing price increases. However, the correlation is very insignificant.

3.1.4 Price and Neighborhoods

The average listing prices of each neighborhood groups are listed below:

Neighborhood group	Average listing price
Manhattan	\$196.87
Brooklyn	\$124.38
Staten Island	\$114.81
Queens	\$99.52
Bronx	\$87.50

The price distribution are shown here:



The average listing price of Manhattan is the highest, while that of Bronx is the lowest. However, the deviation of listing price of Manhattan is the largest as well. That means, the Airbnb listings in Manhattan can be very expensive and also very affordable.

3.2 Cluster Analysis

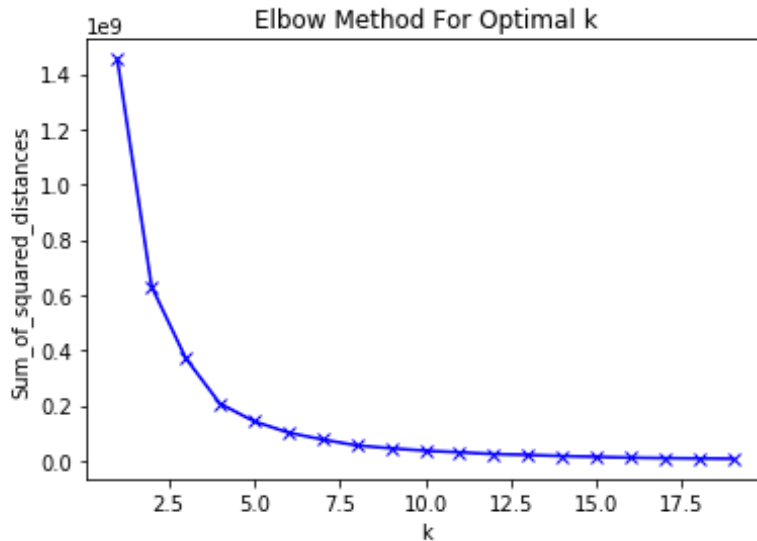
3.2.1 Data cleaning

To simplify the problem, I will focus the study on Entire home/apts in Manhattan. By eliminating other listings, there are only 13199 listings left. The numbers of listings in each Manhattan neighborhood are listed on below:

Upper East Side	1310
Upper West Side	1256
Hell's Kitchen	1187
East Village	1170
Midtown	1100
Harlem	1036
Chelsea	820
West Village	653
Financial District	572
Lower East Side	506
East Harlem	486
Murray Hill	399
Kips Bay	340
Washington Heights	296
Greenwich Village	291
Gramercy	244
SoHo	241
Chinatown	210
Nolita	165
Theater District	164
Morningside Heights	150
Tribeca	143
Inwood	103
Little Italy	76
NoHo	71
Flatiron District	65
Battery Park City	48
Civic Center	35
Two Bridges	26
Roosevelt Island	19
Stuyvesant Town	12
Marble Hill	5

I choose only the variables neighborhood and price for the clustering. The neighborhood variable is converted to dummy variables before model fitting.

In order to find the optimal number of clusters, I used the Elbow Method. The 13199 data points are divided into 1 to 20 clusters using k-means clustering. Then the sum of squared distances of the clusters are calculated and drawn in a chart. Here is the chart:



The elbow point is the optimal k. In this case, the optimal k is 4. So we will cluster the 13199 listings into 4 clusters. The clusters are named Cluster 0,1,2 and 3. The number of listings and average price in each cluster are:

Cluster	Number of Listing	Average price
0	12139	\$195.20
1	64	\$2697.97
2	983	\$663.00
3	13	\$7366.92

So in the 4 clusters, Cluster 0 has the most listings and the lowest average price. Cluster 4 has the least listings and the highest average price.

4 Results

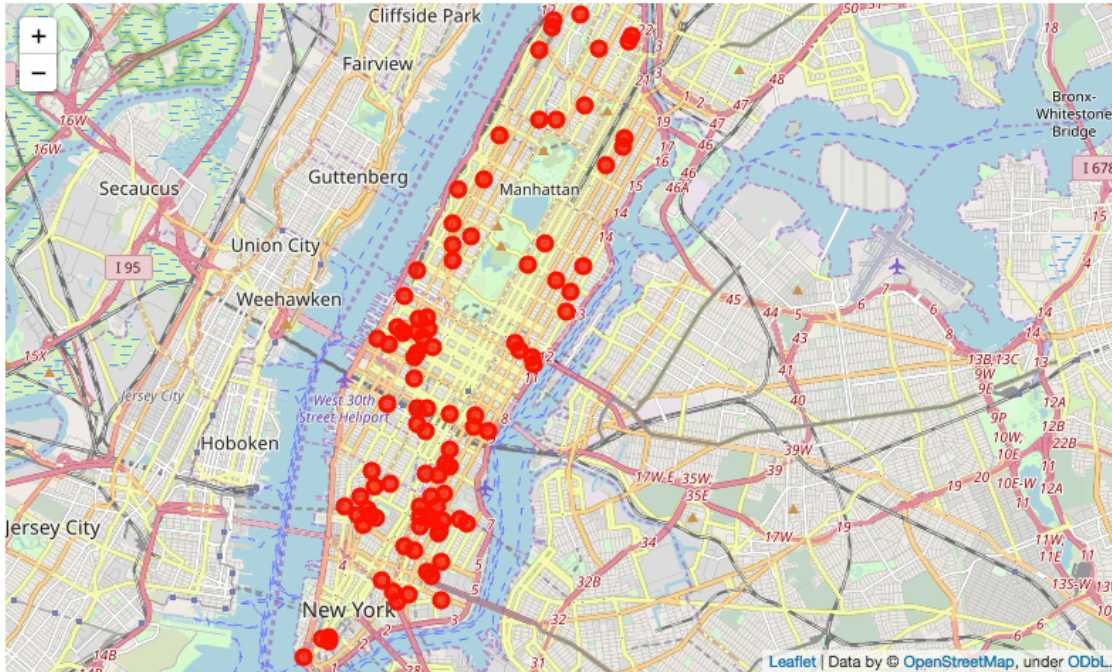
In this session, I will analyze each cluster's price and location distribution, and explore common venues around the listings in each cluster.

4.1 Features of each cluster

4.1.1 Cluster 0

There are 12139 listings in cluster 0. The average listing price in cluster 0 is \$195, with a minimum of \$1, and a maximum of \$1250.

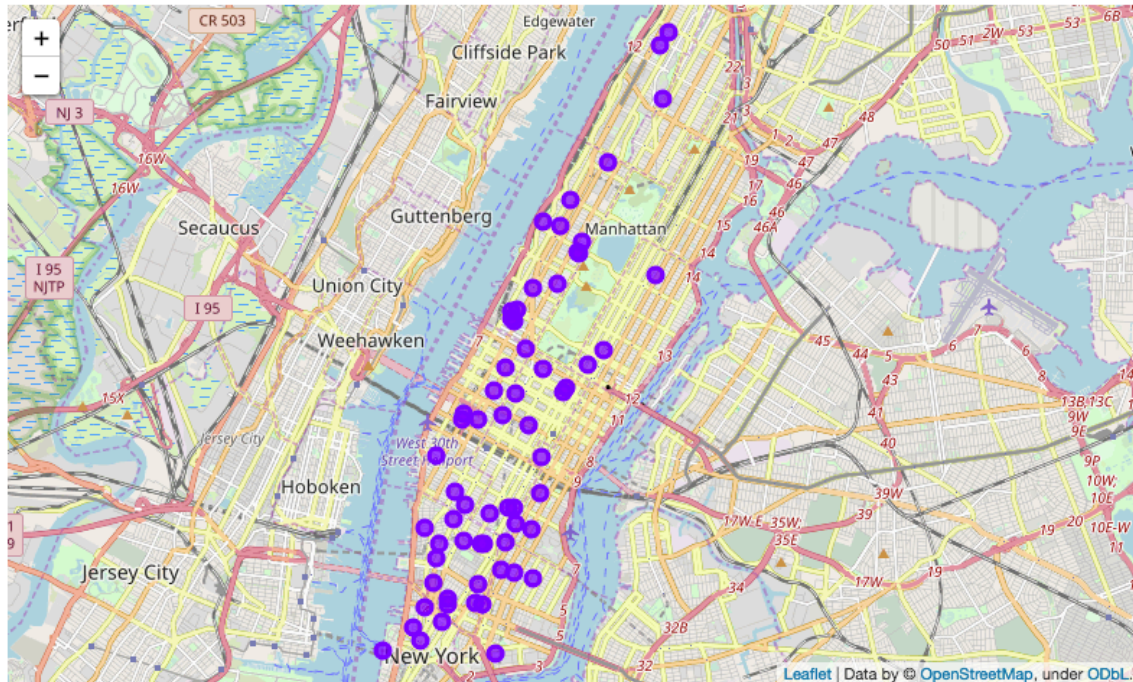
The listings of Cluster 0 are located in every neighborhood in Manhattan. I randomly chose 100 listings in Cluster 0 and plotted them on the map. The map is show on below:



4.1.2 Cluster 1

There are 64 listings in cluster 1. The average listing price in cluster 1 is \$2698, with a minimum of \$1731, and a maximum of \$5000.

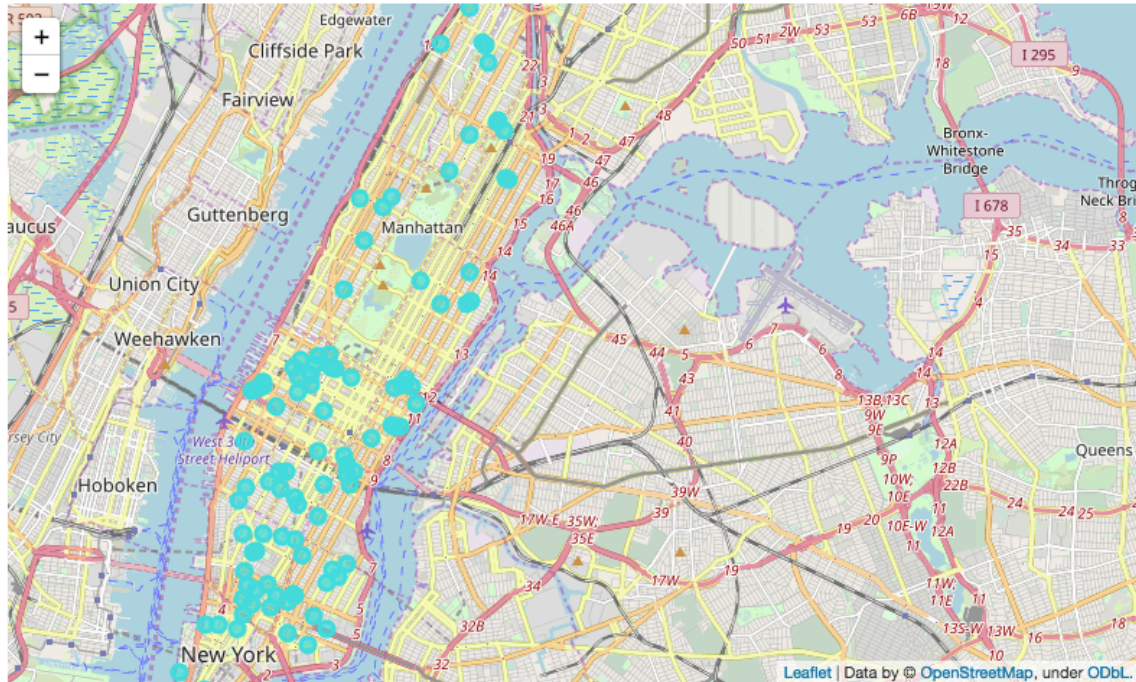
Almost every neighborhood in Manhattan has Cluster 1 listings. The 64 listings are plotted on the map below:



4.1.3 Cluster 2

There are 983 listings in Cluster 2. The average listing price is \$663. The lowest listing price in Cluster 2 is \$430, and the highest one in \$1600.

Similar to Cluster 0 and 1, Cluster 2 listings are located in almost every neighborhood in Manhattan. 100 randomly picked listings in Cluster 2 are plotted on the map below, in order to show the distribution:

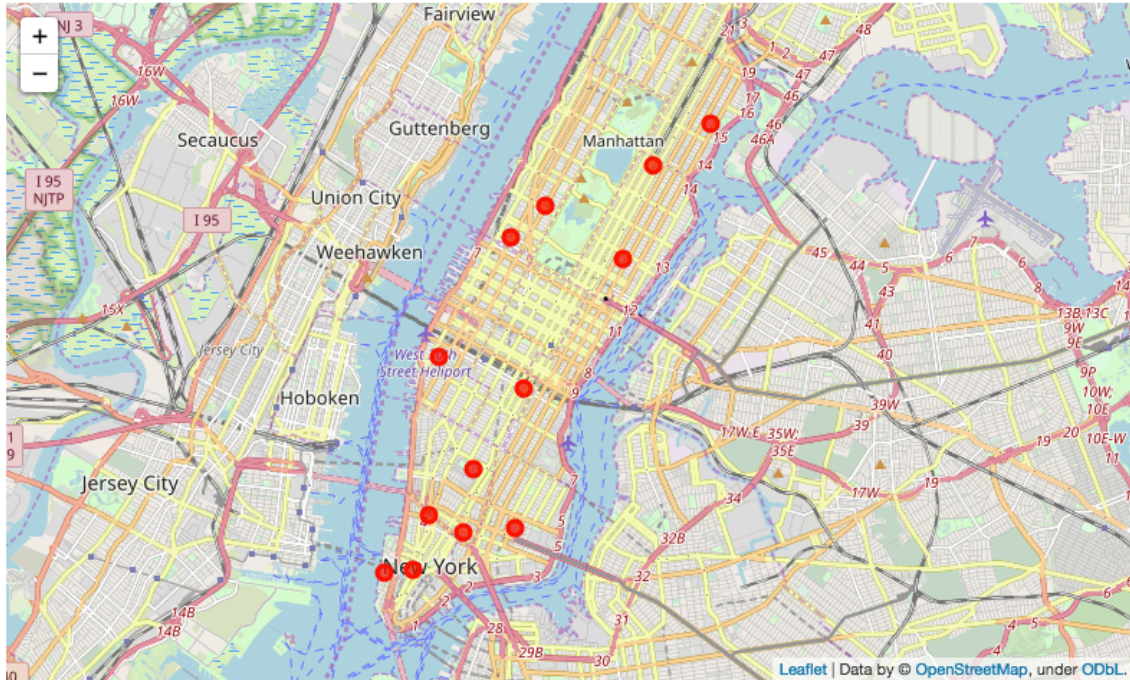


4.1.4 Cluster 3

There are only 13 listings in Cluster 3. The average listing price is \$7367. The lowest listing price in Cluster 3 is \$5100, and the highest one is \$10,000.

Upper East Side, Upper West Side, Midtown, Chelsea, Lower East Side, East Harlem, Greenwich Village, Tribeca, Little Italy, and Battery Park City have listings of Cluster 3. However, there is only 1 or 2 Cluster 3 listings in each of the neighborhood above. So the Cluster 3 are not concentrate as well.

The Cluster 3 listings are plotted on the map below:



4.2 Common Venue Explore

In this section, I will add Foursquare location data and explore the common venues around the listings for each cluster. By doing so, I want to see if the nearby venues of higher-priced Airbnb listings are any different from those of lower-priced listings.

4.2.1 The most expensive cluster: Cluster 3

For there are only 13 listings in Cluster 3, I explored them all. The listings are:

```
UWS 1BR w/backyard + block from CP
SuperBowl Penthouse Loft 3,000 sqft
Spanish Harlem Apt
Luxury townhouse Greenwich Village
1-BR Lincoln Center
East 72nd Townhouse by (Hidden by Airbnb)
Park Avenue Mansion by (Hidden by Airbnb)
Beautiful/Spacious 1 bed luxury flat-TriBeCa/Soho
2br - The Heart of NYC: Mannhattans Lower East Side
70' Luxury MotorYacht on the Hudson
Midtown Manhattan great location (Gramacy park)
3000 sq ft daylight photo studio
Luxury TriBeCa Apartment at an amazing price
```

The top 10 common venue types and the number of such type of venues are:

Coffee Shop	62
Italian Restaurant	51
Art Gallery	44
Hotel	30
Café	30
Park	29
Bakery	29
Gym	28
Wine Shop	24
Gym / Fitness Center	24

4.2.2. Second expensive cluster: Cluster 1

In order to compare the venue density of each clusters, we also randomly picked 13 listings in Cluster 1. They are:

Designer's Beautiful 2BR Apartment in NOLITA/SOHO
 Huge Loft with Private Entrance
 Harrison Green by (Hidden by Airbnb)
 4000 SqFt Luxury Penthouse - Downtown NYC
 Luxury 2Bed/2.5Bath Central Park View
 Times Square Palace
 Ultimate 50th Floor Downtown Penthouse - 4000SqFt
 East 7th Street III by (Hidden by Airbnb)
 Greenwich Village Townhome with Private Garden!
 5 Star Luxury 2 BR Suite Heart of Manhattan
 Lux 2Bed/2.5Bath Central Park Views
 Amazing Chelsea 4BR Loft!
 3 Bedroom Apartment

The top 10 common venue types and the number of such type of venues are:

Italian Restaurant	61
Theater	52
Coffee Shop	39
Hotel	34
American Restaurant	30
Gym / Fitness Center	27
Clothing Store	26
Gym	25
Boutique	25
Café	24

4.2.3. Third expensive cluster: Cluster 2

In order to compare the venue density of each clusters, we also randomly picked 13 listings from Cluster 2. They are:

Spacious and quiet loft in Greenwich village
Designer Downtown Loft
Luxury 2Bed/2Bth Private Loft Best NYC Location
NO GUEST SERVICE FEE Beekman Tower Premium Studio Suite
NO GUEST SERVICE FEE Beekman Tower One Bedroom Suite #1
One-of-a-Kind Luxury NYC EPIC VIEW!
HUGE**Washington Square Park**3BEDROOM SKY-LOFT
Lincoln Center Studio
Sunny, Spacious and Perfect Village Location
Lovely 3 Bedroom, 2 Bath Flatiron Area Loft
Rare Find: SUNNY, LARGE DUPLEX Chelsea 2BR 2BA
Beautiful Times Square Apt. Heart of Manhattan!
West Village 2 Bedroom Apt

The top 10 common venue types and the number of such type of venues are:

Italian Restaurant	56
Coffee Shop	45
Gym / Fitness Center	35
Hotel	34
Bakery	33
American Restaurant	31
Wine Bar	28
Park	28
Café	28
French Restaurant	23

4.2.4 The least expensive cluster: Cluster 0

I also randomly picked 13 listings from Cluster 0. They are:

Luxury Manhattan Loft w Huge Private Back Yard
 Cozy 1 Bedroom Apt in Hamilton Heights
 Beautiful 2 Bedroom Apt Near Times Square
 Cosy private bedroom in Inwood 4 min to train 1,A
 1BR East Village Apt - Entire Apt.
 CUTE PLACE & QUIET DREAMS W BALCONY
 Sunny East Village 1 Bedroom
 Two Bedrooms Tonhouse
 Manhattan Club (1 bedroom/Sleeps 4)
 Beautiful & Cozy One Bedroom Apt
 In the heart of the city!
 MTW- Steffanie
 Cozy garden floor apartment-Brownstone

The top 10 common venue types and the number of such type of venues are:

Hotel	41
Italian Restaurant	38
Pizza Place	38
Coffee Shop	37
Theater	35
Mexican Restaurant	28
Bar	26
Bakery	23
American Restaurant	21
Seafood Restaurant	20

4.2.5 Cluster comparison

In order to compare the common venue category and number, we put all top 10 venue types of each cluster together:

Cluster 3 – Most expensive		Cluster 1 – 2 nd expensive	
Coffee Shop	62	Italian Restaurant	61
Italian Restaurant	51	Theater	52
Art Gallery	44	Coffee Shop	39
Hotel	30	Hotel	34
Café	30	American Restaurant	30
Park	29	Gym / Fitness Center	27
Bakery	29	Clothing Store	26
Gym	28	Gym	25
Wine Shop	24	Boutique	25
Gym / Fitness Center	24	Café	24
Cluster 2 – 3 rd expensive		Cluster 0 – least expensive	
Italian Restaurant	56	Hotel	41
Coffee Shop	45	Italian Restaurant	38
Gym / Fitness Center	35	Pizza Place	38
Hotel	34	Coffee Shop	37
Bakery	33	Theater	35
American Restaurant	31	Mexican Restaurant	28
Wine Bar	28	Bar	26
Park	28	Bakery	23
Café	28	American Restaurant	21
French Restaurant	23	Seafood Restaurant	20

From the table above, we will notice that:

1. As the price decreases, the density of venues decreases as well. For example, for Cluster 3 listings, there are 51 Italian restaurants nearby. So for each listing, there are about 4 Italian restaurants nearby. But for Cluster 0 listings, same as Cluster 3, Italian restaurants are the 2nd most common venues, but the total number is 38. So for every listing, there are 3 Italian restaurants nearby.
2. Art Galleries are very common near Cluster 3 listings. For each Cluster 3 listing, there are more than 3 Art Galleries nearby. Also, Wine shops are common for Cluster 3 listings only. There are about 2 Wine shops near each Cluster 3 listing.
3. Gym/Fitness centers are common for Cluster 1,2 and 3. For the least expensive cluster – Cluster 0, you can find different restaurants around, but not so much fitness centers.
4. Clothing stores and Boutiques are only common near Cluster 1, which is the 2nd expensive cluster.

5. Discussion

From the analysis of common venues of each neighborhood, we can get some rules of the pricing of Manhattan Airbnb listings:

- Rule 1: Concentrated Art Galleries, clothing stores and Boutiques can be an indicator of expensive Airbnb rooms.
- Rule 2: In expensive clusters, the venues are more concentrated.
- Rule 3: In area with a lot of restaurants but no or a few gyms, you may find less expensive Airbnb here.

Other than the location, we also noticed '3,000 sqft', 'Mansion', 'Luxury', '4br' etc. words in expensive listings, while '1bd', '2bd', 'studio', 'loft' in less expensive listings. In this study, we didn't consider the house type of the listing, but that would be another significant impact for the listing price.

6. Conclusion

From the study, we can conclude that:

- Airbnb prices differed from town to town, but in a town, such as Manhattan, the prices are not determined by the neighborhood, as expensive ones and affordable ones are mixed together.
- In areas that Art Galleries, clothing stores and bouquets concentrated, the room prices are high.
- In areas that with a lot of restaurants but no of few gym, the prices are low.

Further study:

- In this study, only explore entire home/apt listings are explored. It would be interesting to learn if the rules also applied to other room types: private room and shared room
- In this study, only Manhattan are studies. It would be interesting to study other towns and see if the same rules applied.
- The study could be more reliable if adding information about the home/apt, such as the sqft, how many bedrooms and how many baths, year of construction, style and so on.