

The Battle of the Neighborhoods
New York City
Richard Kalteyer

Introduction: Business Problem

New York City is the largest city in the United States with over 8 million people residing in the metropolitan area. Each year, millions of tourists from around the world visit the area to enjoy what the city has to offer. New York City is divided up into 5 boroughs, including Manhattan, The Bronx, Queens, Brooklyn, and Staten Island. This report focuses specifically on Manhattan. When visiting, tourists have a variety of choices of where to stay. Some of these choices include apartments, hotels, and Airbnb's. Based on their proximity to tourist attractions, restaurants, particular neighborhoods, etc. these options range in price. This report takes a look at the best Airbnb places to stay in relation to the previously stated criterion in the Manhattan borough of New York City.

Target Audience

The target audience for this project is anyone who is looking to visit New York City, particularly in the Manhattan borough for longer than 1 night and needs information on the best available Airbnb based on location, price, quality of area, and proximity to local attractions such as museums, restaurants, food, etc.

Data

The initial data that we analyze comes from the 2019 Airbnb listings in Manhattan. The data is available to the public. Since we are only focused on Manhattan, the master data set was filtered down to our particular area of interest and scraped for other criterion I felt would provide a good representation of what a tourist is looking for. The data set was robust with over 2,000 options.

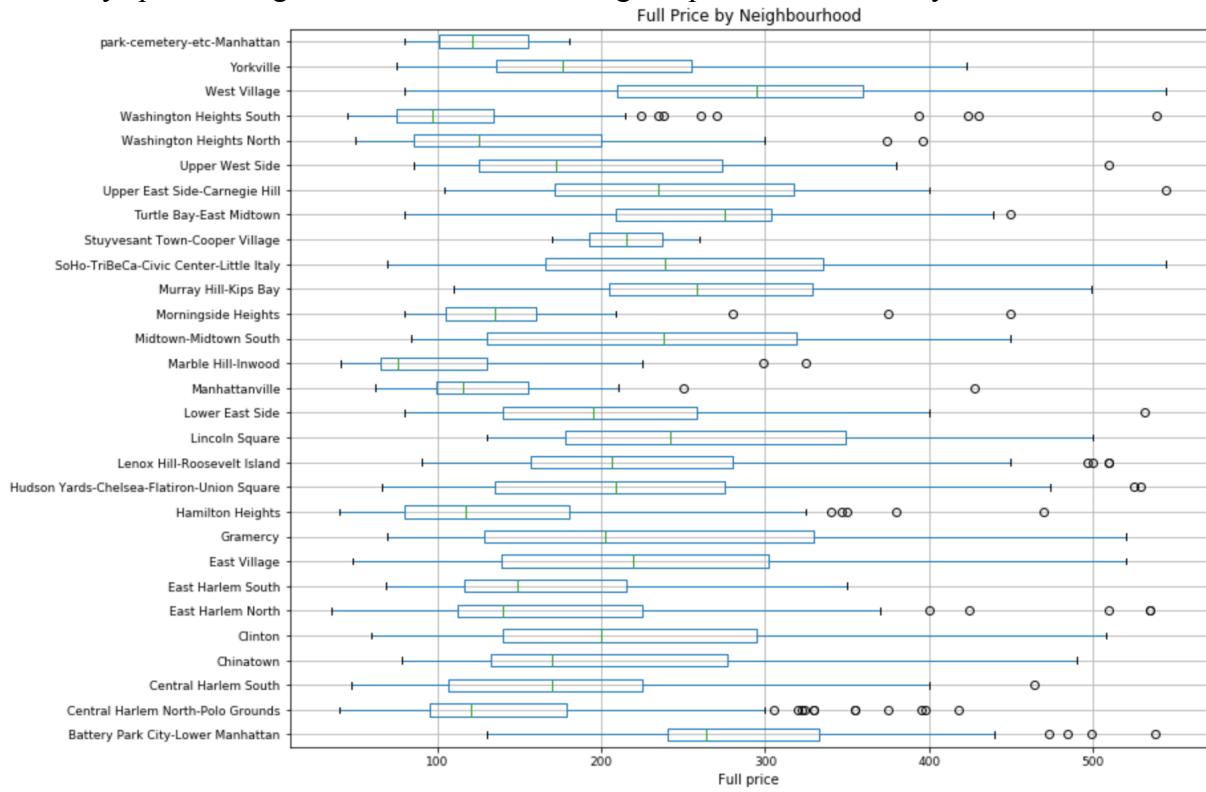
Other datasets used for this analysis include crime records from the NYPD to assess neighborhood safety, neighborhood locations (lat, long), and Foursquare API data with information on different venues tourists will most likely be interested in visiting. The crime reports will classify each offense as either felony or misdemeanor. The Foursquare API data on venues will include museums, food, shopping, sightseeing, concert locations, etc.

Methodology

The purpose of this project is to identify neighborhoods in Manhattan that provide tourists an opportunity to rent an Airbnb with good reviews, a reasonable price, where crime is low, and local attractions are nearby. In order to do so, the following steps were taken.

1. Collect data of Airbnb's for rent with their NYC Tabulation Area, or neighborhood name
2. Collect nearby crime report data
3. Assign a neighborhood to each nearby crime
4. Identify specific metrics through analysis, such as crime numbers in each area, Airbnb rental cost per person, and number of Airbnb listings available
5. Select a top 100 Airbnb list based on the desired criteria of rating and crimes per person
6. Utilize Foursquare API to find highly coveted Airbnb rentals near venues
7. Use k-means clustering to ultimately recommend rental options for tourists

Additional data analysis was performed by producing a boxplot. The boxplot shows the prices of Airbnb's by specific neighborhood to show the range of prices in which they were listed.



Analysis

As previously mentioned, the following was identified.

- Price per person
- Crime rate
- Number of listings

Below is a list of top 10 neighborhoods with highest price per person in 2019:

- West Village (\$112.85), 88 listings
- Lincoln Square (\$112.51), 20 listings
- Stuyvesant Town-Cooper Village (\$107.50), 2 listings
- SoHo-TriBeCa-Civic Center-Little Italy (\$105.38), 81 listings
- Upper East Side-Carnegie Hill (\$96.99), 24 listings
- Lenox Hill-Roosevelt Island (\$96.64), 52 listings
- Turtle Bay-East Midtown (\$94.82), 53 listings
- Gramercy (\$94.36), 48 listings
- Murray Hill-Kips Bay (\$90.41), 62 listings
- Yorkville (\$89.69), 58 listings

Below is a list of top 10 neighborhoods with lowest price per person in 2019:

- Marble Hill-Inwood (\$45.48), 25 listings
- Washington Heights South (\$46.79), 82 listings
- Washington Heights North (\$54.74), 53 listings
- Central Harlem North-Polo Grounds (\$57.04), 132 listings
- Manhattanville (\$59.75), 25 listings

- Hamilton Heights (\$60.07), 97 listings
- park-cemetery-etc.-Manhattan (\$61.40), 7 listings
- East Harlem North (\$62.28), 91 listings
- East Harlem South (\$64.95), 77 listings
- Central Harlem South (\$67.00), 94 listings

Below is a list of top 10 neighborhoods with highest crime rate in 2019:

- Midtown-Midtown South (10397)
- Hudson Yards-Chelsea-Flatiron-Union Square (7788)
- East Harlem North (6221)
- Central Harlem North-Polo Grounds (5186)
- SoHo-TriBeCa-Civic Center-Little Italy (4789)
- West Village (4748)
- Upper West Side (4700)
- Clinton (4364)
- Central Harlem South (4320)
- Washington Heights South (3682)

Below is a list of top 10 neighborhoods with lowest crime rate in 2019:

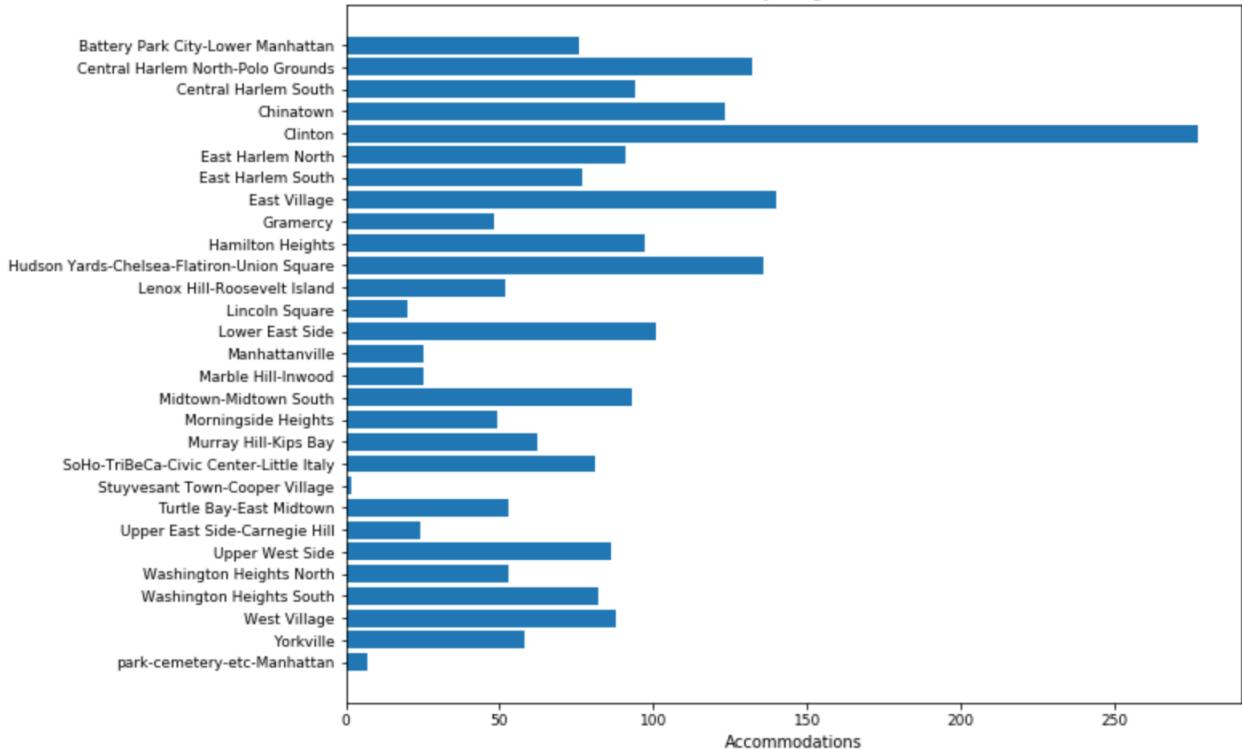
- Stuyvesant Town-Cooper Village (145)
- park-cemetery-etc-Manhattan (1213)
- Lenox Hill-Roosevelt Island (1604)
- Manhattanville (1832)
- Yorkville (1898)
- Marble Hill-Inwood (1900)
- Gramercy (2021)
- Morningside Heights (2321)
- Murray Hill-Kips Bay (2333)
- East Village (2340)

Below you will find some of the data used in the report

	tab_area	crimes_summary
20	Stuyvesant Town-Cooper Village	145
28	park-cemetery-etc-Manhattan	1213
11	Lenox Hill-Roosevelt Island	1604
14	Manhattanville	1832
27	Yorkville	1898
15	Marble Hill-Inwood	1900
8	Gramercy	2021
17	Morningside Heights	2321
18	Murray Hill-Kips Bay	2333
7	East Village	2340
24	Washington Heights North	2440
12	Lincoln Square	2636
21	Turtle Bay-East Midtown	2643
9	Hamilton Heights	2738
0	Battery Park City-Lower Manhattan	2883
3	Chinatown	3137
22	Upper East Side-Carnegie Hill	3403
6	East Harlem South	3629
13	Lower East Side	3650
25	Washington Heights South	3682
2	Central Harlem South	4320
4	Clinton	4364
23	Upper West Side	4700
26	West Village	4748
19	SoHo-TriBeCa-Civic Center-Little Italy	4789
1	Central Harlem North-Polo Grounds	5186
5	East Harlem North	6221
10	Hudson Yards-Chelsea-Flatiron-Union Square	7788
16	Midtown-Midtown South	10397

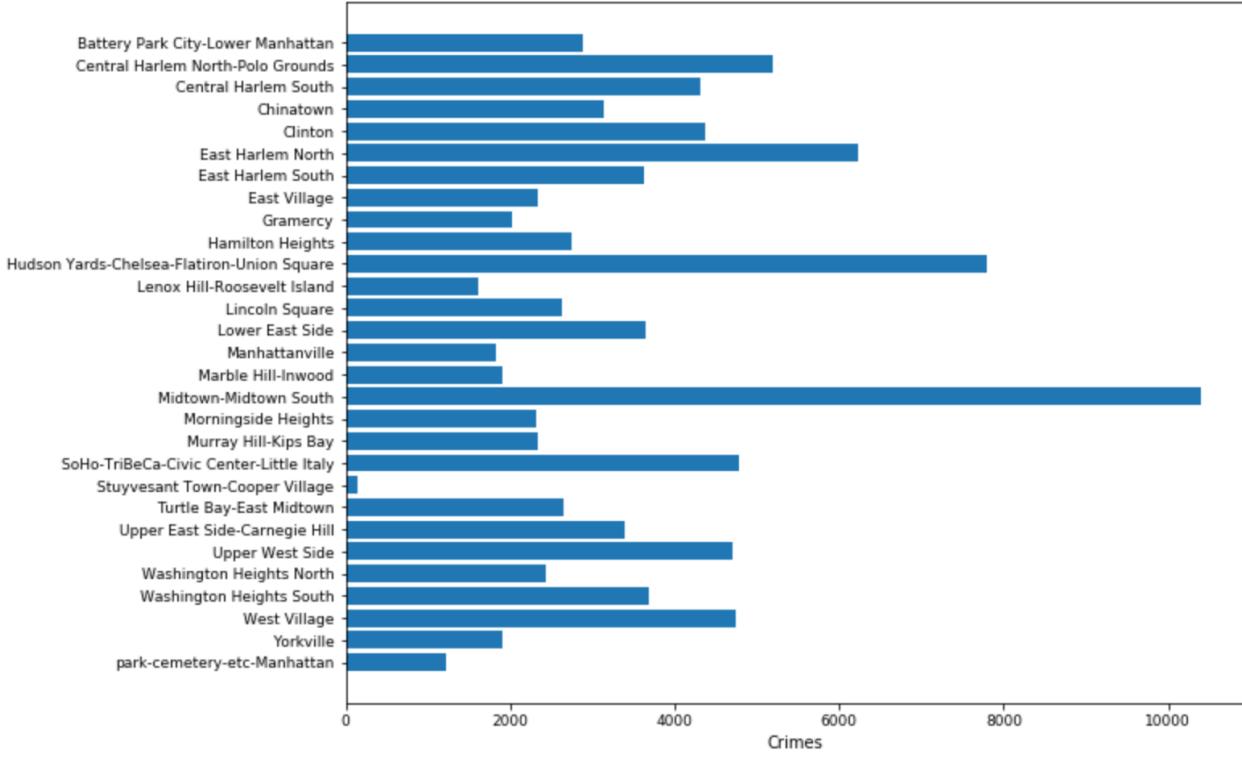
Listings by Neighborhood

Accommodations Number Total by Neighbourhood on Manhattan in 2019

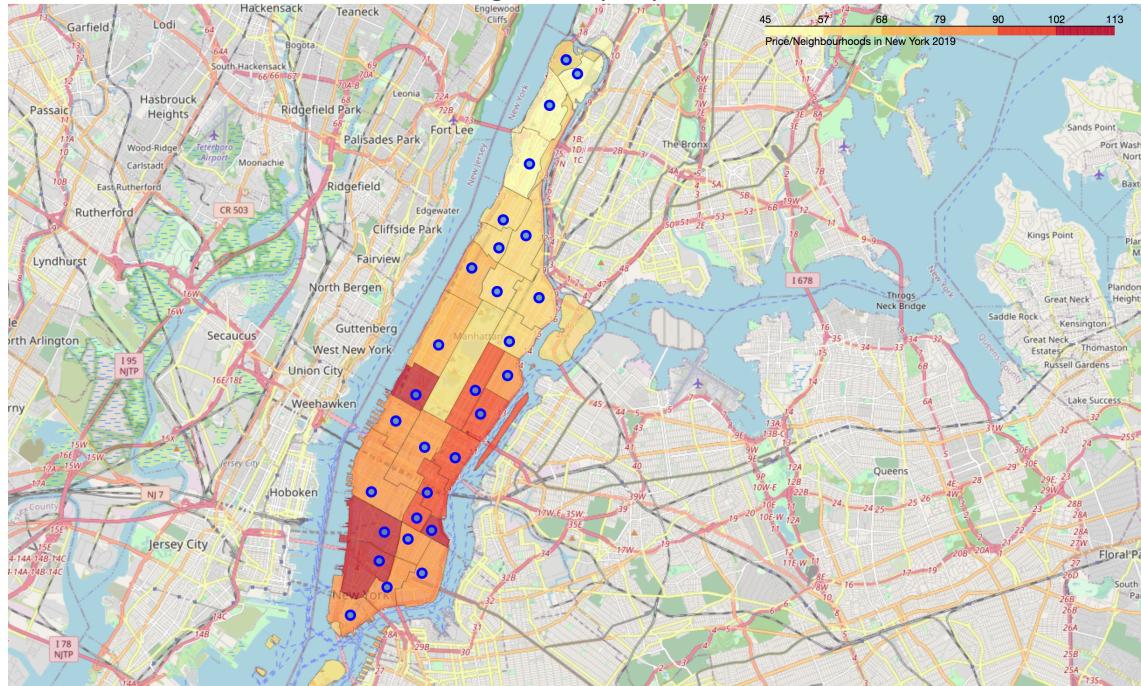


Crimes by Neighborhood

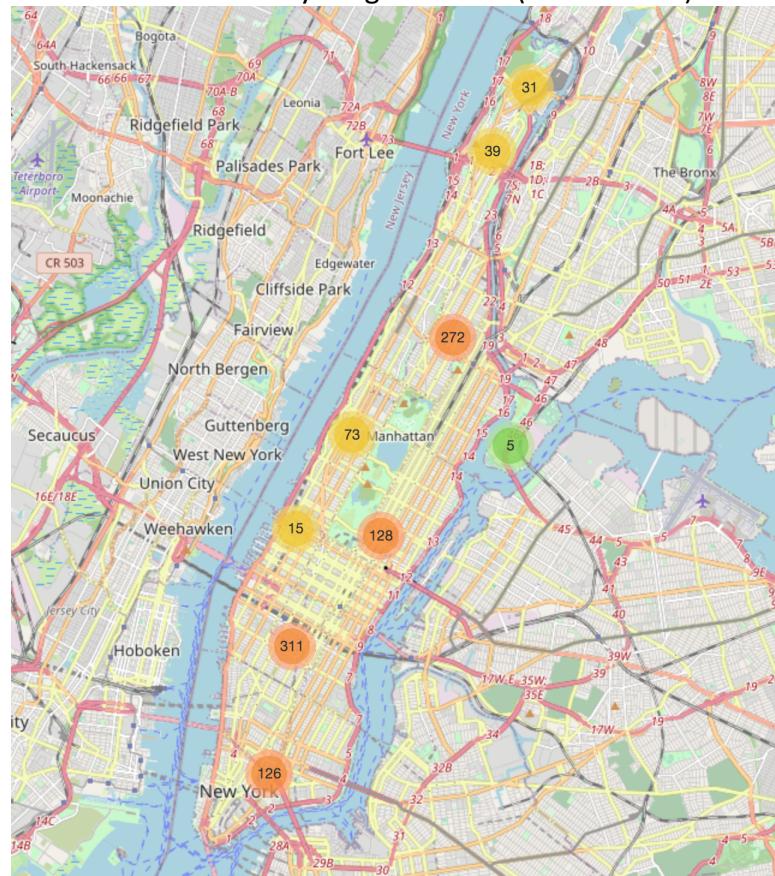
Neighborhoods Total Crimes Records on Manhattan in 2019



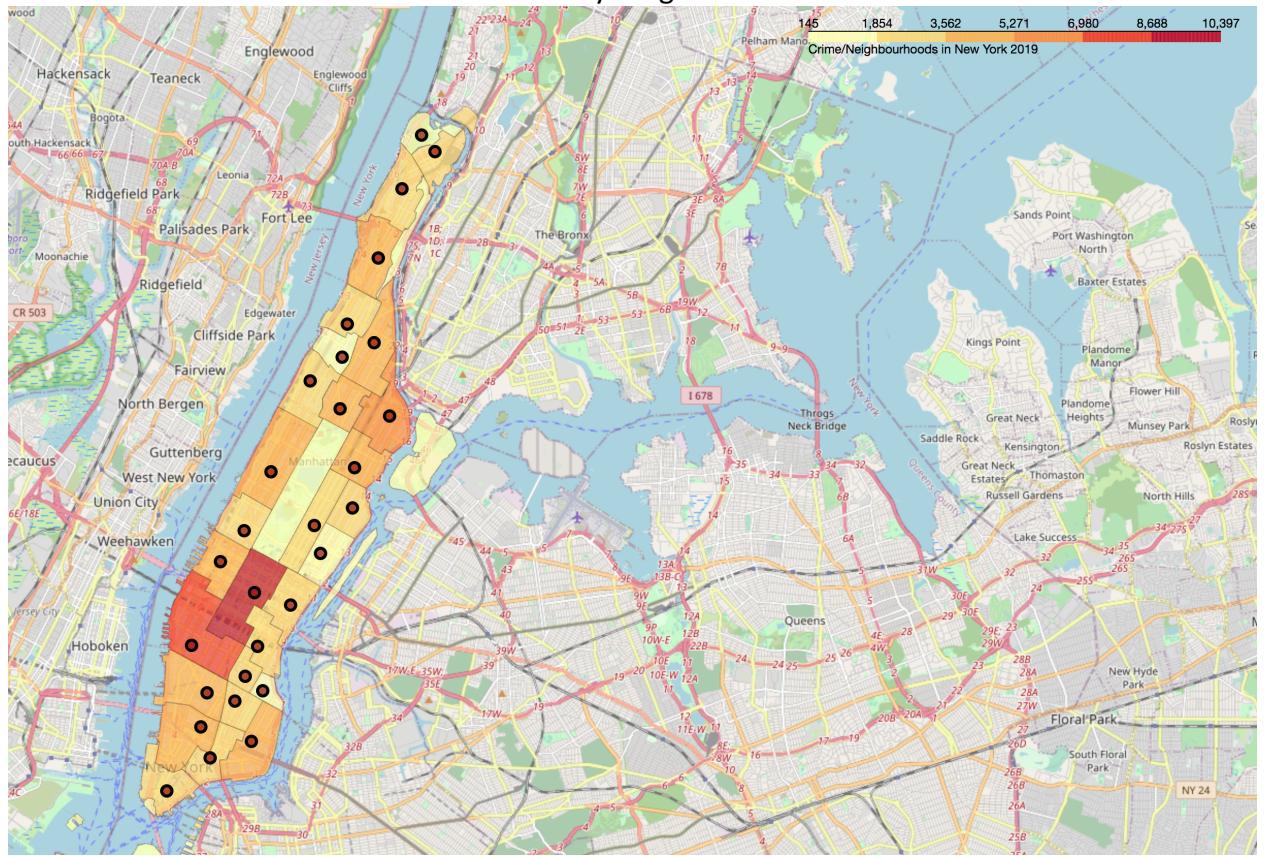
Average Price per person



Crime Cluster by Neighborhood (1000 crimes)



Crime Numbers by Neighborhood



Foursquare API Analysis

We analyzed the top 100 listings from the Airbnb listing dataset, and among those, we identified the top 3 venues within a 1000 m radius. Ultimately, we will create 3 clusters that define these listings to provide recommendations to tourists on where to stay. The top 100 Airbnb listings from the entire dataset were selected based on review score, total price, price/person, and crime numbers. Our venues were categorized into food, entertainment, fine arts, shopping, and outdoor activities.

Run k-means clustering of neighborhoods

Cluster Labels	name	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	1st Most Common Venue Share	2nd Most Common Venue Share	3rd Most Common Venue Share
0	2 **Stylish, Quiet, Centrally Located (9th & 52nd)	Food Place	Entertainment	Sightseeing	0.64	0.36	0
1	1 157-C	Food Place	Sightseeing	Fine Art	0.58	0.25	0.12
2	1 A neat bedroom in a cozy 3-bedroom apartment	Sightseeing	Food Place	Shopping	0.45	0.45	0.03
3	0 Art filled peaceful paradise EV Union Square	Food Place	Entertainment	Sightseeing	0.67	0.17	0.08
4	0 Artsy Parisian Apt in Greenwich Village	Food Place	Entertainment	Sightseeing	0.65	0.25	0.1

Cluster 0-A mixture of everything

- Ranked in the middle for average price per person
- Ranked in the middle for average crimes
- Contains 55% from all listings

	review_scores_rating	accommodates	full_price	price_per_person	crimes	bathrooms	bedrooms	Cluster Labels
count	55.0	55.000000	55.000000	55.000000	55.000000	55.000000	55.000000	55.0
mean	100.0	2.181818	222.109091	106.347879	68.036364	1.018182	1.036364	0.0
std	0.0	0.883824	114.362107	46.475378	59.104103	0.234987	0.428764	0.0
min	100.0	1.000000	67.000000	37.000000	3.000000	0.000000	0.000000	0.0
25%	100.0	2.000000	134.500000	69.750000	38.500000	1.000000	1.000000	0.0
50%	100.0	2.000000	191.000000	97.500000	52.000000	1.000000	1.000000	0.0
75%	100.0	2.500000	290.000000	137.250000	78.000000	1.000000	1.000000	0.0
max	100.0	5.000000	519.000000	259.500000	385.000000	2.000000	2.000000	0.0

Cluster 1-Sightseeing

- Ranked the lowest average price per person
- Ranked the lowest average from crimes
- Sightseeing is 1st and 2nd top venue categories
- Contains 29% from of all listings

	review_scores_rating	accommodates	full_price	price_per_person	crimes	bathrooms	bedrooms	Cluster Labels
count	29.000000	29.000000	29.000000	29.000000	29.000000	29.000000	29.000000	29.0
mean	99.931034	3.103448	202.482759	70.043103	59.586207	1.241379	1.103448	1.0
std	0.257881	1.952010	124.588357	39.698427	37.750608	0.510964	0.488791	0.0
min	99.000000	1.000000	54.000000	27.000000	4.000000	1.000000	0.000000	1.0
25%	100.000000	2.000000	90.000000	47.000000	34.000000	1.000000	1.000000	1.0
50%	100.000000	2.000000	180.000000	62.000000	54.000000	1.000000	1.000000	1.0
75%	100.000000	4.000000	270.000000	70.000000	82.000000	1.000000	1.000000	1.0
max	100.000000	10.000000	470.000000	225.000000	154.000000	3.000000	3.000000	1.0

Cluster 2 –Entertainment

- Ranked the highest average price per person
- Ranked the highest average from crimes
- Entertainment is 1st and 2nd top venue categories
- Contains 15% from of all listings

	review_scores_rating	accommodates	full_price	price_per_person	crimes	bathrooms	bedrooms	Cluster Labels
count	15.0	15.000000	15.000000	15.000000	15.000000	15.0	15.000000	15.0
mean	100.0	2.666667	273.733333	111.403333	102.000000	1.0	0.866667	2.0
std	0.0	1.175139	96.230725	40.353361	64.378346	0.0	0.516398	0.0
min	100.0	1.000000	110.000000	55.000000	23.000000	1.0	0.000000	2.0
25%	100.0	2.000000	210.000000	79.650000	61.000000	1.0	1.000000	2.0
50%	100.0	2.000000	275.000000	105.000000	74.000000	1.0	1.000000	2.0
75%	100.0	4.000000	329.500000	137.500000	130.500000	1.0	1.000000	2.0
max	100.0	5.000000	420.000000	200.000000	257.000000	1.0	2.000000	2.0

Results and Discussion

After running our analysis of Manhattan, we found that all of the three clusters have “Food” in the 1st most common venue, so in order to differentiate the clusters from each other, we move on to the remaining categories. Other influencing factors include the average price per person, crime rates, and location. The following clusters can be predominantly categorized as such:

Cluster 0-A mixture of everything

- Ranked in the middle for average price per person
- Ranked in the middle for average crimes
- Average price per person \$106
- Average crime rate 68
- Contains 55% from all listings
- Distributed among almost all neighborhoods

Cluster 1-Sightseeing

- Ranked the lowest average price per person
- Ranked the lowest average from crimes
- Average price per person \$70
- Average crime rate 59
- Sightseeing is 1st and 2nd top venue categories
- Contains 29% from of all listings
- Distributed among Harlem, East Harlem, Upper West Side, Washington Heights, Roosevelt Island, Morningside Heights, Financial District, West Village, and Inwood

Cluster 2 –Entertainment

- Ranked the highest average price per person
- Ranked the highest average from crimes
- Average price per person \$111
- Average crime rate 102
- Entertainment is 1st and 2nd top venue categories
- Contains 15% from of all listings

- Distributed among Hell's Kitchen, Midtown, and Chelsea

Conclusion

In conclusion, the project was performed to determine which area in the Manhattan borough of New York City a tourist would want to rent an Airbnb based on several criteria. Data was pulled from various public sources and analyzed to solve this problem. After scraping the data and creating tables and other various visualization products, it was determined that the vast amount of choices available to the average tourist could be lumped into 3 distinct clusters. These clusters give the tourist an idea of which Airbnb they would want to rent, based on nearby venues, cost, and neighborhood safety.

- It was determined that if someone was interested in **sightseeing**, cluster 1 would best fit this option. This cluster provides both the lowest average cost and lowest average crime rate. There are many things to see in New York City, including the Empire State Building, the Statue of Liberty, and many other museums.
- If the tourist was primarily interested in only **entertainment**, cluster 2 provides the best fit for that choice. As long as the individual can put up with a higher crime rate and higher prices, this option is for them.
- If the individual does not necessarily have a preference on what they value the most, we recommend cluster 0. This cluster provides a mixture of all the categories. Crime rates are relatively low, but the average cost per person is a little bit on the higher end.