

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-we will focus on Manhattan.
- The target audience for this project is any tourist who is looking to visit New York City, particularly in the Manhattan borough for longer than one night and needs information on the best available Airbnb based on:
 - Cost
 - Quality of area (safety)
 - Proximity to venues of interest such as museums, shops, landmarks, restaurants, etc.

DATA

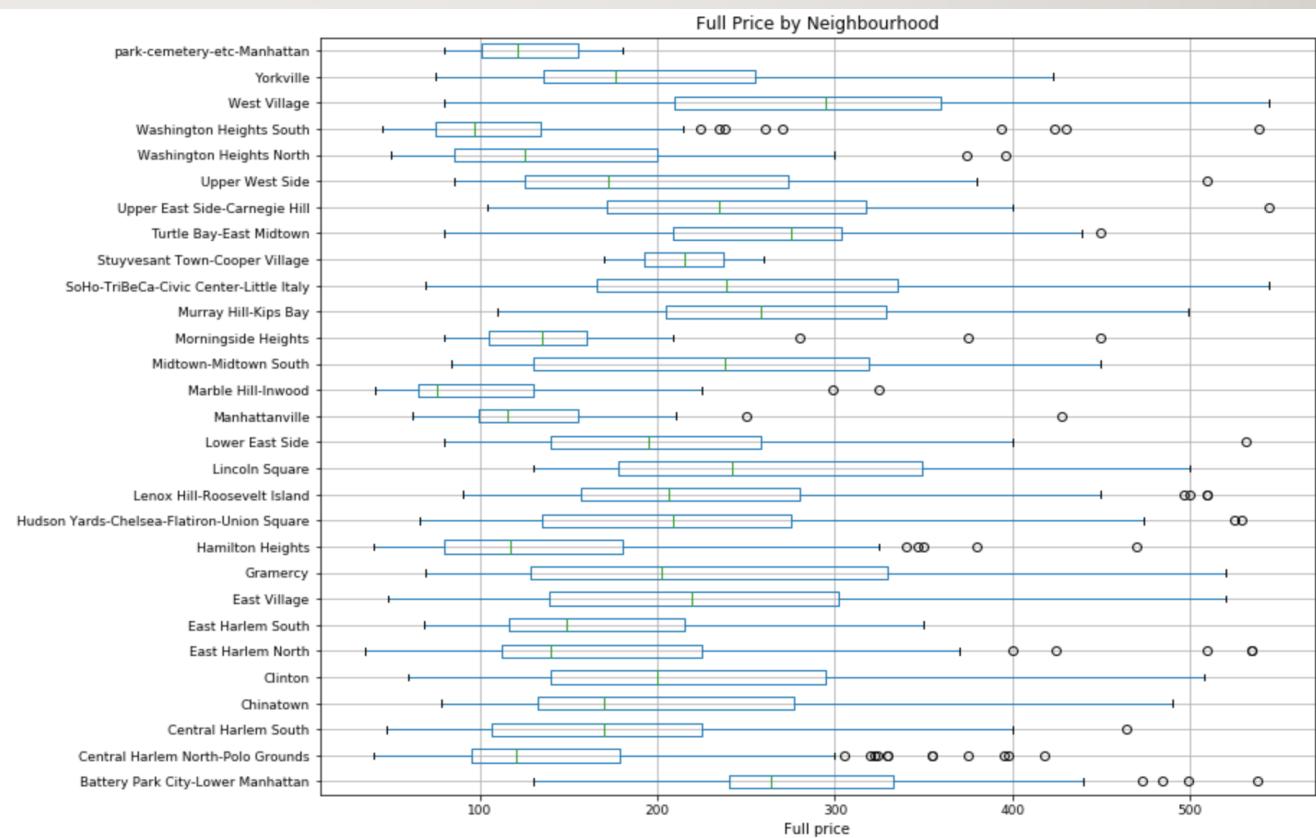
- Initial data to analyze comes from the 2019 Airbnb listings in Manhattan (public)
- Dataset was scraped and broken down into essential criteria
- Other data to be used includes:
 - Crime records (2019) from the NYPD to asses neighborhood safety
 - Neighborhood locations (latitude, longitude)
 - Foursquare API for info on venues like museums, food, shopping, sightseeing, etc.

METHODOLOGY

- The following steps were taken to answer the problem at hand
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

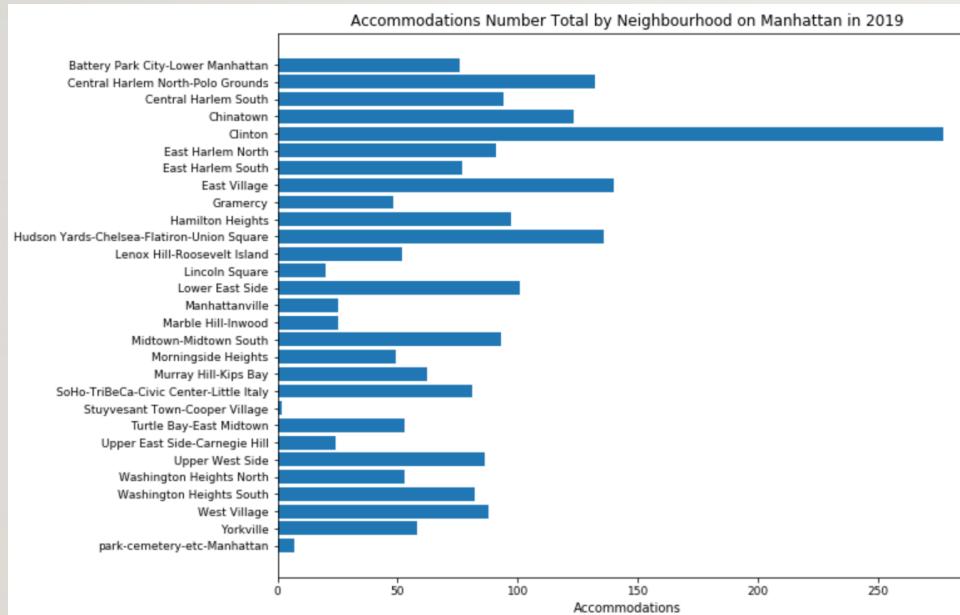
METHODOLOGY

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

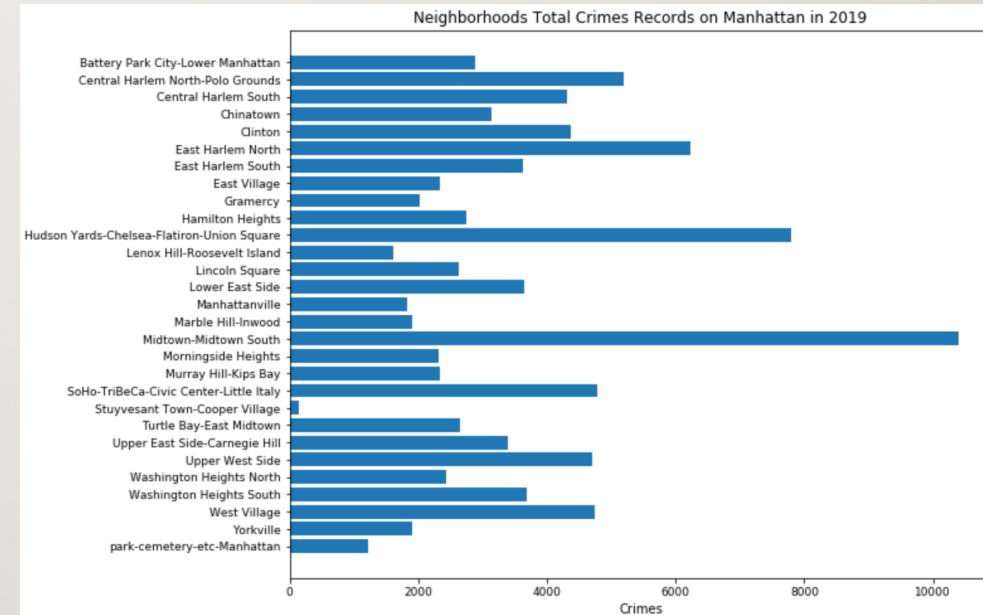


ANALYSIS

Listings by Neighborhood



Crimes by Neighborhood



ANALYSIS

Top 10 neighborhoods with highest crime (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)

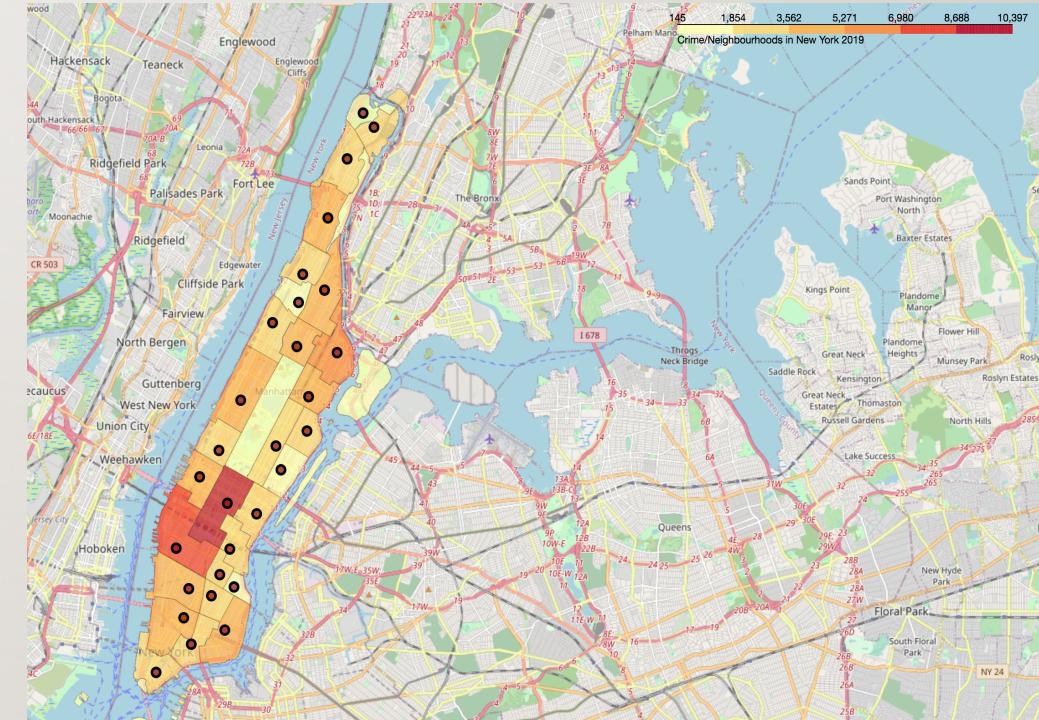
Top 10 neighborhoods with lowest crime rate (2019)

- 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)

Crime Cluster by Neighborhood



Crime Numbers by Neighborhood



ANALYSIS

- **Foursquare API**
- We analyzed the top 100 listings from the Airbnb listing dataset
- We identified the top 3 venues within a 1000 m radius.
- The top 100 Airbnb listings from the entire dataset were selected based on
 - review score
 - total price
 - price/person
 - crime numbers.
- Venues were categorized into food, entertainment, fine arts, shopping, and outdoor activities.

ANALYSIS

- **K-means clustering of neighborhoods**
- We create 3 clusters that define these listings to provide recommendations to tourists on where to stay.

- Cluster 0

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
• Cluster 1	1	157-C	Food Place	Sightseeing	Fine Art	0.58	0.25	0.12
• Cluster 2	2	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

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.
- In order to differentiate the clusters from each other, we move on to the remaining categories including:
 - average price per person
 - crime rates
 - location

RESULTS AND DISCUSSION

- **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

	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

RESULTS AND DISCUSSION

- **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

	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

RESULTS AND DISCUSSION

- **Cluster 2- Entertainment**

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

CONCLUSION

- 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.