

Choosing the best location for a restaurant in Manhattan: A Use Case of Crime Data and Foursquare [Capstone project]

Mari Takashima

* This is a capstone project and may not be used as an advice.

Introduction¶

In this report, I will explore and analyse a dataset collected about Manhattan restaurant businesses inspections. The target audience for this project is the foreigners, who are not currently residing in Manhattan, that are looking into opening a restaurant business. For a foreigner starting a business in an unfamiliar location, the first concern will be the safety. Is the location safe enough for a foreigner to open a business? Is it close to the neighbourhood with high crime rates? This report will analyse the crime data in Manhattan by using the techniques learnt in this IBM course and select a few locations that have low crime rates. The second concern will be to know about the popular venues in the locations. The popular venues in the location can tell us about the people and the needs of the neighbourhood. For example, if we see cafes and pards as most popular venues in the location, the needs will be a food that can be taken away or food trucks. The report will analyse the foursquare venue data in Manhattan to see the trend and potential needs of the locations that were selected previously.

Data

1) The crime data that was used in the first analysis is from City of New York Open data. <https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc>

This is a breakdown of every arrest effected in NYC by the NYPD during the current year. This data is manually extracted every quarter and reviewed by the Office of Management Analysis and Planning. Each record represents an arrest effected in NYC by the NYPD and includes information about the type of crime, the location and time of enforcement. In addition, information related to suspect demographics is also included. This data can be used by the public to explore the nature of police enforcement activity.

2) The neighbourhood data for Manhattan was extracted from the NYU Spatial data repository. The address is https://geo.nyu.edu/catalog/nyu_2451_34572. This New York City Neighborhood Names point file was created as a guide to New York City's neighbourhoods that appear on the web resource, "New York: A City of Neighborhoods." The data was created in 2014.

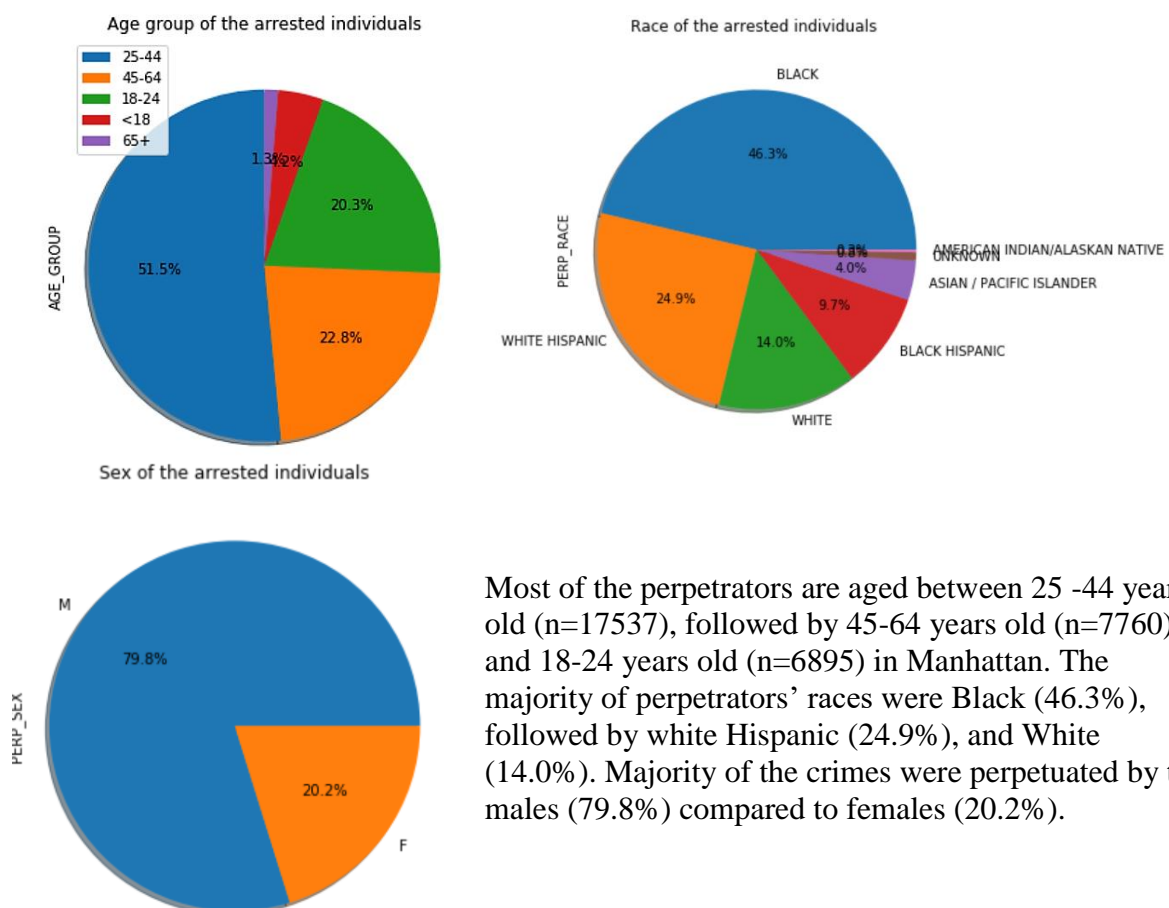
3) The Foursquare API data was also be used to explore venues around the location.

Methodology

The crime data of New York was downloaded here: <https://data.cityofnewyork.us/Public-Safety/NYPD-Arrest-Data-Year-to-Date-/uip8-fykc>. Only Manhattan data was extracted by using the filtering the "ARREST_BORO" that had "M". The data dimension was 12 columns and 34033 rows.

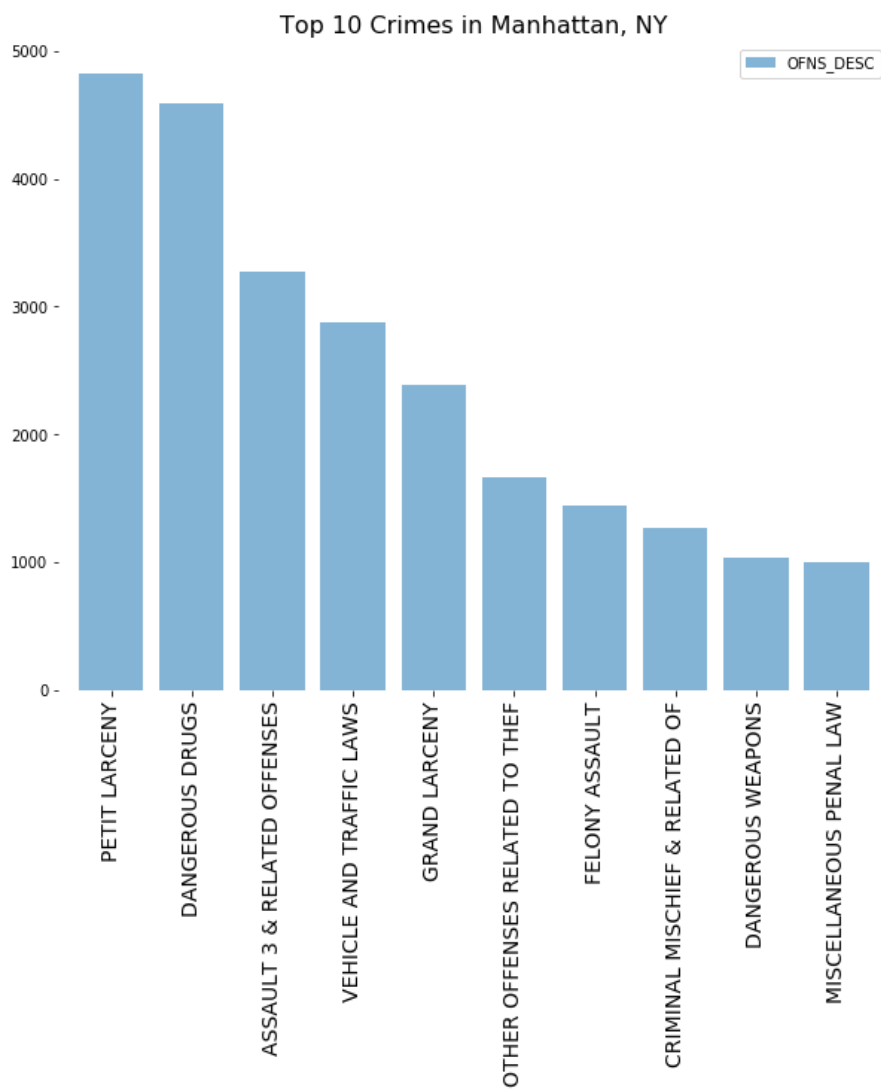
| | ARREST_KEY | ARREST_DATE | PD_DESC | OFNS_DESC | ARREST_BORO | AGE_GROUP | PERP_SEX | PERP_RACE | X_COORD_CD | Y_COORD_CD | Latitude | Longitude |
|---|------------|---------------------------|---|-----------------------------|-------------|-----------|----------|-------------------|------------|------------|-----------|------------|
| 0 | 184611183 | 06/30/2018 12:00:00 AM | BURGLARY,UNCLASSIFIED,UNKNOWN TIME | BURGLARY | M | 25-44 | M | WHITE HISPANIC | 983190 | 208988 | 40.740309 | -74.003827 |
| 1 | 184619323 | 06/30/2018 12:00:00 AM | ASSAULT 2,1,UNCLASSIFIED | FELONY ASSAULT | M | 45-64 | F | BLACK | 987633 | 203880 | 40.726288 | -73.987796 |
| 2 | 184621759 | 06/30/2018 12:00:00 AM | TRAFFIC,UNCLASSIFIED MISDEMEAN | VEHICLE AND TRAFFIC LAWS | M | 25-44 | M | BLACK | 1002892 | 232227 | 40.804074 | -73.932665 |
| 3 | 184626775 | 06/30/2018 12:00:00 AM | LARCENY,PETIT FROM OPEN AREAS,UNCLASSIFIED | PETIT LARCENY | M | 25-44 | M | WHITE HISPANIC | 983312 | 228266 | 40.793217 | -73.967274 |
| 4 | 184616966 | 06/30/2018 12:00:00 AM | ROBBERY,UNCLASSIFIED,OPEN AREAS | ROBBERY | M | 25-44 | M | BLACK HISPANIC | 993373 | 229301 | 40.796058 | -73.967052 |

There are some interesting demographic details in this table like perpetrator's sex description (PERP_SEX), perpetrator's race description (PERP_RACE), and perpetrator's age group (AGE_GROUP) to see the picture of crimes in Manhattan before dropping them from the dataframe.



Most of the perpetrators are aged between 25 -44 years old (n=17537), followed by 45-64 years old (n=7760), and 18-24 years old (n=6895) in Manhattan. The majority of perpetrators' races were Black (46.3%), followed by white Hispanic (24.9%), and White (14.0%). Majority of the crimes were perpetuated by the males (79.8%) compared to females (20.2%).

The crime offence records were then sorted and grouped by the offence code. The top 10 offence codes of 2018 were examined. The top was petit larceny followed by dangerous drugs and assault. The random 1000 incidents were sampled and superimposed on the Manhattan map to visualise the location with fewer incidents.



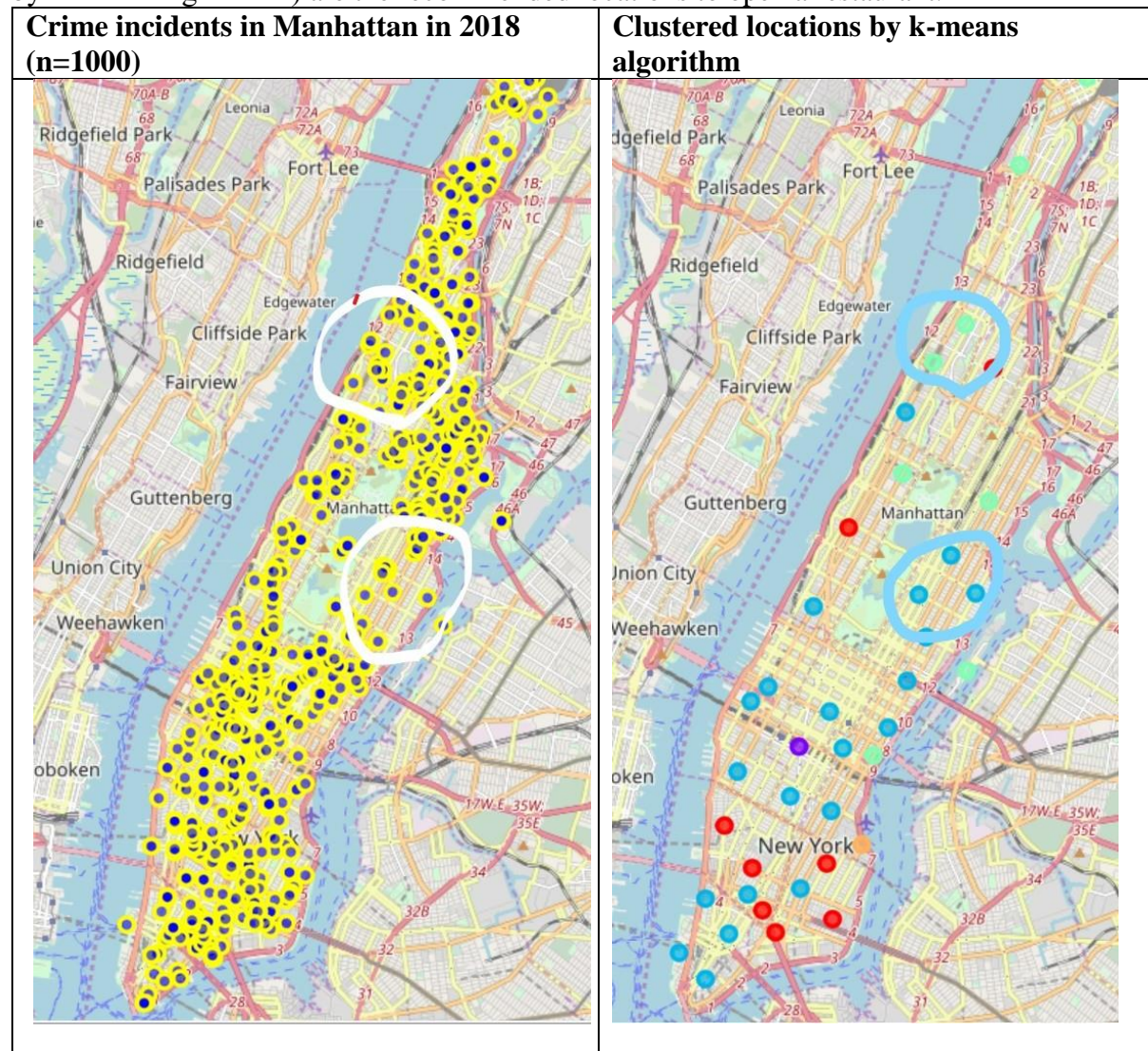
```
k: 1 cost: 0.815492061168
k: 2 cost: 0.721634692463
k: 3 cost: 0.705673665837
k: 4 cost: 0.617725592767
k: 5 cost: 0.580047652606
k: 6 cost: 0.54890272922
k: 7 cost: 0.535785310084
k: 8 cost: 0.487206933403
k: 9 cost: 0.461631915032
k: 10 cost: 0.434352632628
```

The Foursquare API data were used to explore trending venues around the location. The unsupervised learning K-means algorithm was used to cluster the boroughs. K-Means algorithm is one of the most common cluster methods of unsupervised learning. Using the SKLearn (scikits-learn), the value was calculated to choose the best K. The difference in values plateaus after k=5, so 5 was chosen for the number of clusters.

K-means algorithm is used for a customer segmentation. The five clusters were examined for its characteristics of popular trending venues to recommend what kind of restaurants will be suitable for that location.

Results

This is the Manhattan map with the random 1000 incidents superimposed (“Crime incidents in Manhattan in 2018 (n=1000)”). As marked with the white pen, two locations had lesser incidents than other places. These clusters in these marked circles (see “Clustered locations by k-means algorithm”) are the recommended locations to open a restaurant.



This is the list of five clusters derived from the k-means algorithm. The locations that had less crime incidents were:

- 1) Hamilton heights: Cluster #3
- 2) Manhattanville: Cluster #3
- 3) Carnegie hill: Cluster #2
- 4) Yorkville: Cluster #2
- 5) Upper East Side: Cluster #2

The result of the top 10 trending venues around the locations are below:

Cluster #2 (Only few rows that included the recommended locations are shown.)

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|----|---------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|-----------------------|-----------------------|-----------------------|------------------------|-------------------------|
| 0 | Marble Hill | Coffee Shop | Discount Store | Yoga Studio | Kids Store | Steakhouse | Supplement Shop | Shopping Mall | Shoe Store | Seafood Restaurant | Tennis Stadium |
| 8 | Upper East Side | Italian Restaurant | Exhibit | Art Gallery | Juice Bar | Bakery | Coffee Shop | French Restaurant | Hotel | Gym / Fitness Center | Boutique |
| 9 | Yorkville | Coffee Shop | Gym | Bar | Italian Restaurant | Pizza Place | Mexican Restaurant | Deli / Bodega | Sushi Restaurant | Japanese Restaurant | Sandwich Place |
| 10 | Lenox Hill | Italian Restaurant | Coffee Shop | Sushi Restaurant | Gym / Fitness Center | Pizza Place | Gym | Sporting Goods Shop | Burger Joint | Thai Restaurant | Bakery |
| 13 | Lincoln Square | Theater | Gym / Fitness Center | Italian Restaurant | Concert Hall | Plaza | Café | French Restaurant | Performing Arts Venue | Park | Indie Movie Theater |
| 14 | Ciinton | Theater | Coffee Shop | American Restaurant | Italian Restaurant | Gym | Gym / Fitness Center | Hotel | Spa | Wine Shop | New American Restaurant |
| 15 | Midtown | Hotel | Clothing Store | Theater | Steakhouse | Coffee Shop | American Restaurant | Cocktail Bar | Bookstore | Sporting Goods Shop | Spa |
| 26 | Morningside Heights | Coffee Shop | Food Truck | American Restaurant | Bookstore | Park | Deli / Bodega | Tennis Court | Burger Joint | Sandwich Place | New American Restaurant |
| 27 | Gramercy | Italian Restaurant | Cocktail Bar | American Restaurant | Wine Shop | Bagel Shop | Coffee Shop | Mexican Restaurant | Grocery Store | Thrift / Vintage Store | Restaurant |
| 28 | Battery Park City | Coffee Shop | Park | Hotel | Wine Shop | Italian Restaurant | Ice Cream Shop | Fountain | Shopping Mall | Pizza Place | Food Truck |
| 29 | Financial District | Coffee Shop | Hotel | Wine Shop | Steakhouse | Bar | Gym | Pizza Place | Italian Restaurant | Food Truck | Falafel Restaurant |
| 30 | Carnegie Hill | Pizza Place | Coffee Shop | Cosmetics Shop | Café | Yoga Studio | Spa | Bar | Bookstore | French Restaurant | Gym |
| 31 | Noho | Italian Restaurant | French Restaurant | Cocktail Bar | Coffee Shop | Grocery Store | Gift Shop | Sushi Restaurant | Hotel | Bookstore | American Restaurant |
| 32 | Civic Center | Gym / Fitness Center | Italian Restaurant | Bakery | French Restaurant | Cocktail Bar | Yoga Studio | Gym | Coffee Shop | Sandwich Place | Park |
| 33 | Chelsea | Gym / Fitness Center | Italian Restaurant | Bar | French Restaurant | Furniture / Home Decor | American Restaurant | Spa | Bookstore | French Restaurant | Gym |

From the observation, there are many Italian restaurants, coffee shops, and entertainment facilities (gym, theatre, park, theatre) in cluster #2. The Upper East Side is located around the art gallery and exhibit. Yorkville has gym as the second most common venue, but there are no healthy food restaurants that attract the gym-goers. The cosmetics shop, yoga and spa are in the top 10 venues in Carnegie Hill. It might be a popular location for young females.

Cluster 3 (Only few rows that included the recommended locations are shown.)

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | 6th Most Common Venue | 7th Most Common Venue | 8th Most Common Venue | 9th Most Common Venue | 10th Most Common Venue |
|---|--------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|------------------------|
| 2 | Washington Heights | Café | Bakery | Mobile Phone Shop | Tapas Restaurant | Caribbean Restaurant | Sandwich Place | Chinese Restaurant | Supermarket | Grocery Store | Gym |
| 3 | Inwood | Lounge | Café | Mexican Restaurant | Pizza Place | Wine Bar | Park | Bakery | Frozen Yogurt Shop | Deli / Bodega | Restaurant |
| 4 | Hamilton Heights | Mexican Restaurant | Coffee Shop | Café | Pizza Place | Deli / Bodega | Cocktail Bar | Indian Restaurant | Liquor Store | Sandwich Place | Yoga Studio |
| 5 | Manhattanville | Deli / Bodega | Seafood Restaurant | Italian Restaurant | Chinese Restaurant | Mexican Restaurant | Supermarket | Music School | Beer Garden | Falafel Restaurant | Grocery Store |

The cluster 3 has many cafes, Mexican restaurants and Pizza place. Hamilton heights looks like a multicultural area consisting many Mexican and Indian restaurants. Manhattanville is also a multicultural area with many Italian, Mexican and Chinese restaurants.

Discussion

Five potential locations that had low crime incidences were selected and the trending venues were observed. In Upper Eastside, the upper-class restaurant that matches the surrounding Italian and French restaurants will be suitable for people that visit the exhibition and museum. In Yorkville and Carnegie Hill, a healthy food restaurant that can also sell a To-Go healthy takeaway food for the health-conscious population may be recommended.

In Hamilton Heights and Manhattanville, multicultural cuisines are popular. It probably means that they have a multicultural population in these locations. A franchise Asian restaurant like Ramen or bubble-tea shops will be suitable as they are low-cost and high return.

It will be preferable if the data of housing price was also analysed in the future to assess the cost and return after paying the rent. It may also highlight the socioeconomic differences between the locations in Manhattan to set the price of the menu.

Conclusion

This report concludes that Hamilton Heights and Manhattanville will be realistic for a foreigner to start the restaurant business in Manhattan as it has low crime incidence and can be started from relatively lower cost than opening a store in Yorkville, Upper Eastside and Carnegie Hill.

Thank you for reviewing!