The Battle of Neighborhoods

Finding the best location to set up a new boutique café in Manhattan,NY

Applied Data Science Capstone Project

Danielle O'Sullivan

1. Introduction

1.1 Background

New York City is one of the most diverse cities in the world bursting with attractions. With over 3000 coffee shops in New York it can be said that New Yorkers love their coffee. Recently, MassiveHealth found that New Yorkers drink 6.7 times the amount of coffee consumed by the average denizen of any other US city. -https://medium.com/topos-ai/the-next-wave-predicting-the-future-of-coffee-in-new-york-city-23a0c5d62000.

1.2 Business Problem

Selecting a location in New York to open a new boutique cafe is crucial to a businesses' success.

There are many factors to consider when choosing the location of this new business:

1. Nearby Venues: Other cafes, restaurants, gyms, etc.
2. Crime: Will employees be safe coming to/leaving work, will there be people visiting this area
3. Nearby subway stations: a coffee shop near a subway station might attract more visitors

In this project we will analyse the above factors to determine which area would be the best to open this new café. We will be looking at neighbourhoods in the Manhattan borough only.

1.3 Target Audience:

The target audience for this analysis would be those looking to open a new café/small business in Manhattan.

2 Data

For this analysis, the following datasets were used:

2.1 New York City Data

New York City Data which was downloaded as a json file from https://cocl.us/new_york_dataset. This includes each neighborhood, borough, latitude and longitude.

	Borough	Neighborhood	Latitude	Longitude	tab_area
0	Manhattan	Marble Hill	40.876551	-73.910660	Marble Hill-Inwood
1	Manhattan	Chinatown	40.715618	-73.994279	Chinatown
2	Manhattan	Washington Heights	40.851903	-73.936900	Washington Heights North
3	Manhattan	Inwood	40.867684	-73.921210	Marble Hill-Inwood
4	Manhattan	Hamilton Heights	40.823604	-73.949688	Hamilton Heights

Figure 1 - New York City Data

2.2 Neighborhood Tabulation Areas

This data was downloaded as a json file from NYC Open Data https://data.cityofnewyork.us/City-Government/Neighborhood-Tabulation-Areas-NTA-/cpf4-rkhq. This data was used to add the tabulation area to each of the datasets. This was also used on the chloropleth Folium maps to clearly show the tabulation borders between neighborhoods.

2.3 Crime Data

The crime data was downloaded as csv file from NYC Open Data https://data.cityofnewyork.us/Public-Safety/Crime-Map-/5jvd-shfj. The data was filtered to Manhattan borough. Violation was removed from the law category as a lot of these crimes will not be relevant. We will only be looking at felony and misdemeanour crimes. The dataset was also filtered to only include crimes reported in 2019. After filtering there is 98,693 records.

	CMPLNT_FR_DT	LAW_CAT_CD	BORO_NM	Latitude	Longitude	tab_area
0	01/02/2019	FELONY	MANHATTAN	40.773332	-73.961074	Upper East Side-Carnegie Hill
1	01/03/2019	FELONY	MANHATTAN	40.787567	-73.943132	East Harlem South
2	01/09/2019	MISDEMEANOR	MANHATTAN	40.815732	-73.945420	Central Harlem North-Polo Grounds
3	01/14/2019	FELONY	MANHATTAN	40.794515	-73.966324	Upper West Side
4	01/18/2019	MISDEMEANOR	MANHATTAN	40.723659	-73.991022	East Village
5	01/27/2019	MISDEMEANOR	MANHATTAN	40.732356	-73.984941	Gramercy
6	01/31/2019	MISDEMEANOR	MANHATTAN	40.749780	-73.987781	Midtown-Midtown South
7	01/29/2019	MISDEMEANOR	MANHATTAN	40.823575	-73.937675	Central Harlem North-Polo Grounds
8	02/04/2019	MISDEMEANOR	MANHATTAN	40.710783	-73.996632	Chinatown
9	02/09/2019	FELONY	MANHATTAN	40.813164	-73.941644	Central Harlem North-Polo Grounds

Figure 2- New York City Crime Data

2.4 Foursquare API

Foursquare data will be used to explore each neighborhood in Manhattan and venues nearby each of the coordinates in the New York City data.

2.5 Nearby Subway Stations

Subway Entrances dataset was downloaded as a geojson file from NYC open Data https://data.cityofnewyork.us/Transportation/Subway-Entrances/drex-xx56.

The json file was loaded and read into a new Dataframe.

	Object_ID	URL	Latitude	Longitude	tab_area
0	1763	http://web.mta.info/nyct/service/	40.824069	-73.936981	Central Harlem North-Polo Grounds
1	1764	http://web.mta.info/nyct/service/	40.820343	-73.936508	Central Harlem North-Polo Grounds
2	1765	http://web.mta.info/nyct/service/	40.820662	-73.936275	Central Harlem North-Polo Grounds
3	1766	http://web.mta.info/nyct/service/	40.814024	-73.941117	Central Harlem North-Polo Grounds
4	1767	http://web.mta.info/nyct/service/	40.814316	-73.940910	Central Harlem North-Polo Grounds

Figure 3 - New York City Subway Data

3 Methodology

3.1 Exploratory Data Analysis

3.1.1 Crimes per Neighborhood Tabulation Area

Exploratory data analysis was carried out on the Manhattan crime dataset to identify which Tabulation Areas had the most crime. Crimes were broken down by 'Felony' and 'Misdemeanor'. We can see the top five areas with the highest and lowest crimes below.

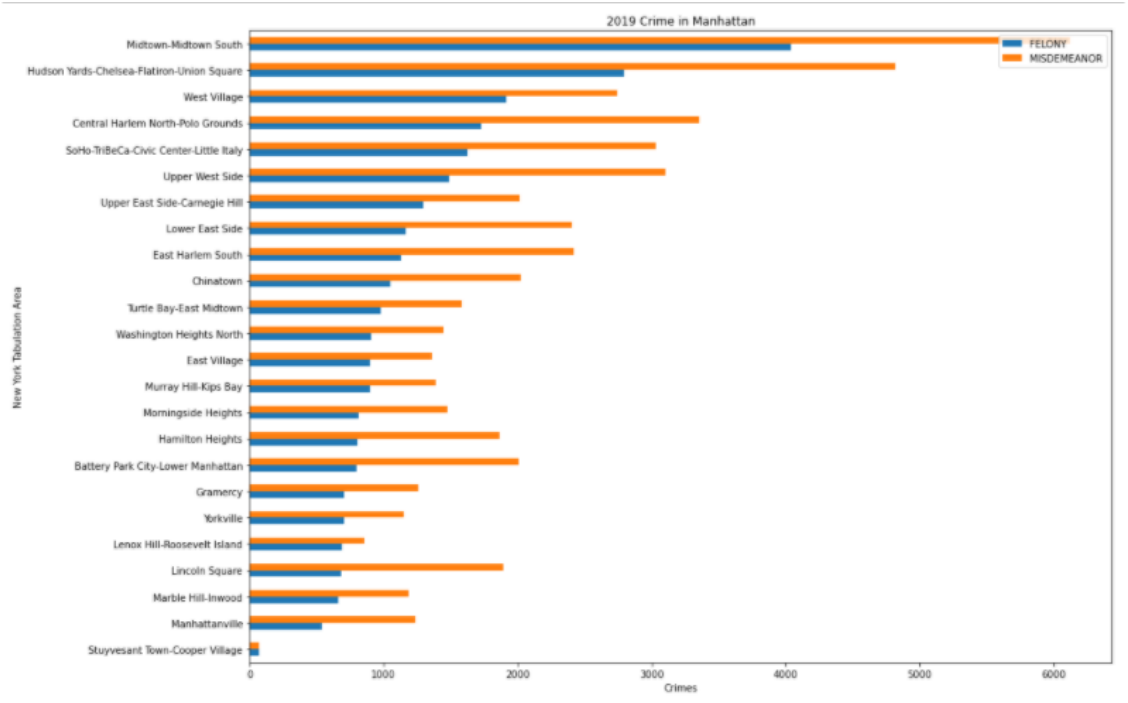

Figure 4 - Crime Bar Chart

	Highest Crime Areas	Lowest Crime Areas
1	Midtown-Midtown South	Stuyvesant Town-Cooper Village
2	Hudson Yards-Chelsea-Flatiron-Union Square	Manhattanville
3	West Village	Marble Hill-Inwood
4	Central Harlem North-Polo Grounds	Lincoln Square
5	SoHo-TriBeCa-Civic Center-Little Italy	Lenox Hill-Roosevelt Island

3.1.2 Subway Entrances per Neighborhood Tabulation Area

A similar analysis is carried out on the Subway Entrances data. A bar chart was created to visualize the number of subway entrances per each neighborhood tabulation area.

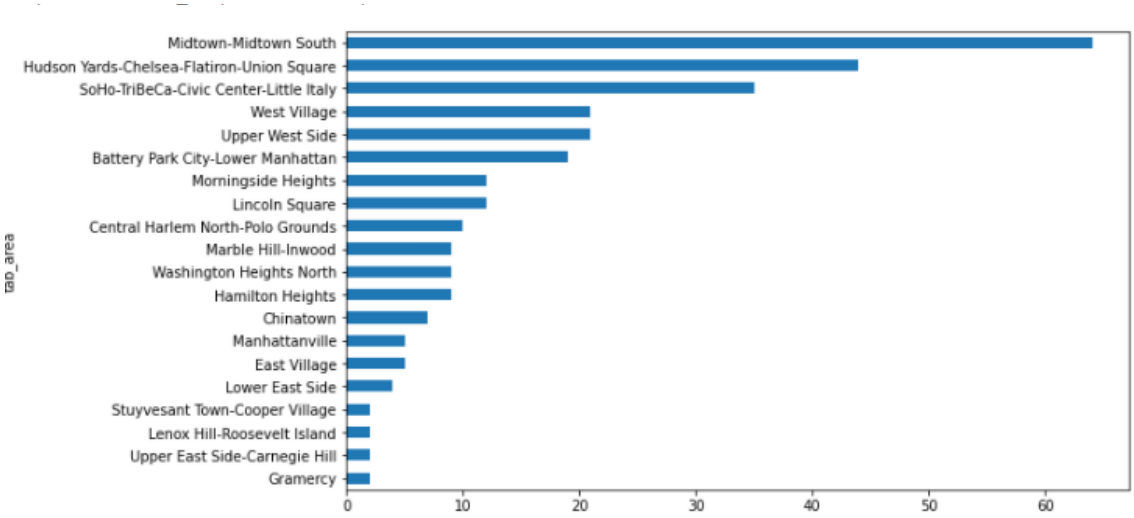

Figure 5- subway data bar chart

The top five areas with highest and lowest number of subway entrances are shown in the below table.

	Most Subway Entrances	Least Subway Entrances
1	Midtown-Midtown South	Gramercy
2	Hudson Yards-Chelsea-Flatiron-Union Square	Upper East Side-Carnegie Hill
3	SoHo-TriBeCa-Civic Center-Little Italy	Lenox Hill-Roosevelt Island
4	West Village	Stuyvesant Town-Cooper Village
5	Upper West Side	Lower East Side

A map is then created using the Folium library to display the subway entrances on a map of Manhattan.

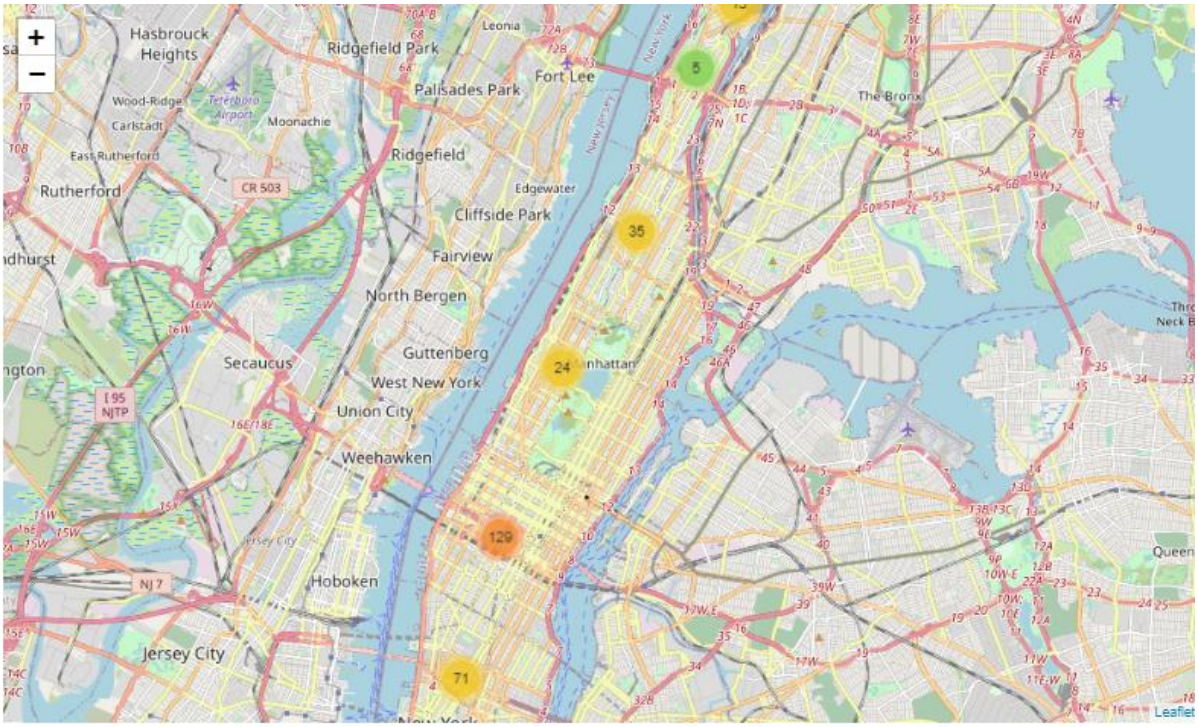

Figure 6 - Subway Map

3.1.3 Foursquare API Data Analysis

To explore the New York data a function is created to get all the nearby venues for each neighborhood. The foursquare API is used to get this data – Venue, Venue Category, Venue Latitude and Venue Longitude.

3.1.3.1 Coffee Shops Analysis

There are 326 unique venue categories in the Foursquare data. As the aim of this analysis is to determine the best location to set up a new coffee shop, a new dataset is created to filter on the below venue categories.

'Donut Shop',
'Coffee Shop',
'Sandwich Place'
'Deli / Bodega'
'Bakery'
'Tea Room'
'Bubble Tea Shop'
'Café'
'Breakfast Spot'
'Bagel Shop

These ten categories were used as coffee would be the main reason someone may visit a business in any of these categories. A new view is created sorted by neighborhood tabulation area and neighborhood name to show many are at each.

tab_area	Neighborhood	Count
Battery Park City-Lower Manhattan	Battery Park City	6
	Financial District	20
Central Harlem North-Polo Grounds	Central Harlem	2
Chinatown	Chinatown	13
East Harlem South	East Harlem	10
East Village	East Village	7
	Noho	11
Gramercy	Gramercy	15
Hamilton Heights	Hamilton Heights	17
Hudson Yards-Chelsea-Flatiron-Union Square	Chelsea	15
	Clinton	12
	Flatiron	8
	Hudson Yards	4
Lenox Hill-Roosevelt Island	Lenox Hill	15
	Roosevelt Island	4
Lincoln Square	Lincoln Square	8
Lower East Side	Lower East Side	9
Manhattanville	Manhattanville	7
Marble Hill-Inwood	Inwood	9
	Marble Hill	6
Midtown-Midtown South	Midtown	19
	Midtown South	10
Morningside Heights	Morningside Heights	9
Murray Hill-Kips Bay	Murray Hill	17
SoHo-TriBeCa-Civic Center-Little Italy	Civic Center	14
	Greenwich Village	13
	Little Italy	26
	Soho	14
	Tribeca	8
Stuyvesant Town-Cooper Village	Stuyvesant Town	1
Turtle Bay-East Midtown	Sutton Place	9
	Tudor City	12
	Turtle Bay	14
Upper East Side-Carnegie Hill	Carnegie Hill	14
	Upper East Side	11
Upper West Side	Manhattan Valley	6
	Upper West Side	12
Washington Heights North	Washington Heights	18
West Village	West Village	11
Yorkville	Yorkville	15

3.1.3.2 Top Venues per Neighborhood

Once we have all the venues a new function is created to get the top five most popular venues per neighborhood. We create a new dataframe as shown below to hold this data. This will be used for clustering the data later.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Battery Park City	Park	Hotel	Gym	Coffee Shop	Shopping Mall
1	Carnegie Hill	Coffee Shop	Yoga Studio	Pizza Place	Bakery	Bookstore
2	Central Harlem	African Restaurant	Gym / Fitness Center	French Restaurant	Bar	Cosmetics Shop
3	Chelsea	Coffee Shop	Art Gallery	American Restaurant	Italian Restaurant	Hotel
4	Chinatown	Chinese Restaurant	Bakery	Cocktail Bar	Vietnamese Restaurant	Dessert Shop

Figure 7 - Top 5 venues Dataframe

3.2 K-Means Clustering

K- means clustering, an unsupervised machine learning algorithm will be applied to the data to cluster difference venue categories and nighborhoods. There will be five clusters created which we can then analyse to determine which neighborhood best suits to set up a new boutique coffee shop.

4 Results and Discussion

4.1 Clustering Results

From the five clusters created, they were broken down into the following:

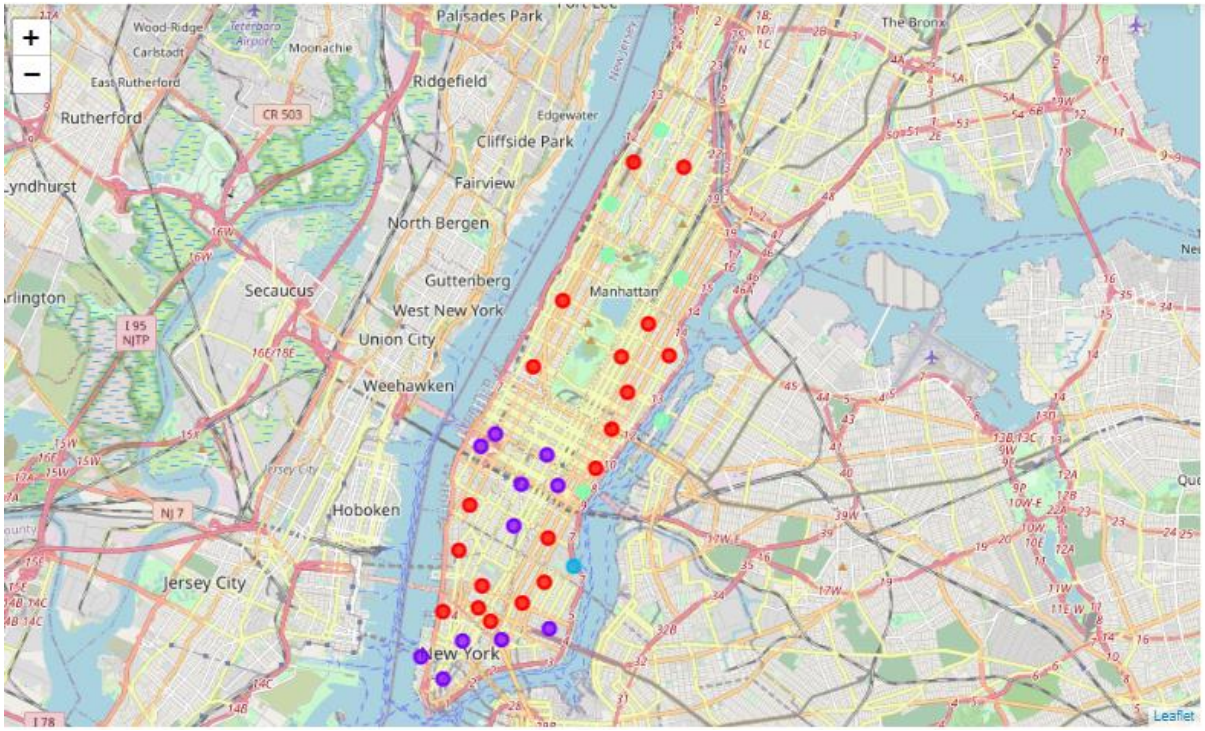

Figure 8- Cluster Map

Cluster 1 – Restaurants/Coffee Shops – Cluster 1 consisted of mainly Restaurants and Coffee Shops

Neighborhoods- 'Manhattanville', 'Central Harlem', 'Upper East Side', 'Yorkville', 'Lenox Hill', 'Upper West Side', 'Lincoln Square', 'Chelsea','Greenwich Village', 'East Village', 'Tribeca', 'Little Italy', 'Soho', 'West Village', 'Gramercy', 'Carnegie Hill', 'Noho', 'Sutton Place', 'Turtle Bay'

Tabulation Areas- Manhattanville', 'Central Harlem North-Polo Grounds', 'Upper East Side-Carnegie Hill', 'Yorkville', 'Lenox Hill-Roosevelt Island', 'Upper West Side', 'Lincoln Square', 'Hudson Yards-Chelsea-Flatiron-Union Square', 'SoHo-TriBeCa-Civic Center-Little Italy', 'East Village', 'West Village', 'Gramercy', 'Turtle Bay-East Midtown

Cluster 2 – Hotel cluster – Cluster 2 consisted of hotels and restaurants/coffee shops again. A lot of these locations around the financial district which could attract businesspeople that would be staying in hotels and tourists.

Neighborhoods- 'Chinatown', 'Clinton', 'Midtown', 'Murray Hill', 'Lower East Side', 'Battery Park City' , 'Financial District','Civic Center', 'Midtown South', 'Flatiron', 'Hudson Yards'

Tabulation Areas- 'Chinatown','Hudson Yards-Chelsea-Flatiron-Union Square''Midtown-Midtown South', 'Murray Hill-Kips Bay', 'Lower East Side','Battery Park City-Lower Manhattan','SoHo-TriBeCa-Civic Center-Little Italy'

Cluster 3 – Tourism cluster – There is only one neighborhood listed here which has a park and a boat/ferry which would indicate tourism. There is currently only one coffee shop listed in this neighborhood which may indicate there is not a lot of business here.

Neighborhoods-Stuyvesant Town
Tabulation Areas- Stuyvesant Town-Cooper Village

Cluster 4 – international – There is a diverse range of food options in this cluster. There are also a lot of parks.

Cluster 5 – There is only one neighborhood here again so it would be too difficult to decide on a name for this cluster.

4.2 Discussion

After analysing each cluster, Cluster 1 and Cluster 2 appear to have the best mix of venues and neighborhoods where a new coffee shop should be set up.

From these clusters and the above analysis on crime and subway data the best locations to set up a new boutique coffee shop are:

Tabulation Area	Neighborhood	Count of coffee venues already here
Battery Park City-Lower Manhattan	Battery Park City	6
	Financial District	20
Hudson Yards-Chelsea-Flatiron-Union Square	Chelsea	15
	Clinton	12
	Flatiron	8
	Hudson Yards	4
SoHo-TriBeCa-Civic Center-Little Italy	Civic Center	14
	Greenwich Village	13
	Little Italy	26
	Soho	14
	Tribeca	8

Hudson Yards-Chelsea-Flatiron-Union Square and Soho-TriBeCa-Civi Center-Little Italy both have high crime rates. They also have the most subway entrances so they would be convenient locations for employees.

Battery Park City- Lower Manhattan has a much lower crime rate and has a lot of subway entrances which makes it an ideal location. Coffee shops place in the top five venues for both Battery Park City and Financial District.

5 Conclusion

In this analysis, New York city neighborhood data, neighborhood tabulation data, crime data, subway entrances data and Foursquare API Venue data was analysed to determine what the best Manhattan neighborhood would be to open up a new boutique coffee shop.

It was determined that Battery Park City or Financial District Neighborhoods would be the best locations due to popular venues nearby such as hotels, coffee shops, gyms, low crime rates in the area and a lot of subway entrances.

5.1 Future Improvements

- All New York Data could be used instead of just Manhattan
- Extra venue information such as apartment buildings/businesses could be included