Capstone Project - The Battle of Neighborhoods

Opening Coffee Shop in New York City

Background

New York City is one of the busiest cities in the world and one of the world's major commercial, financial and cultural centers.. Being the most populous as well as the most densely populated city in the US, lots of people find great business opportunities here.

An entrepreneur is looking to open a coffee shop in one of New York's neighborhoods. He does not understand the area and the availability of coffee shops in each neighborhood. So he wants to investigate the area and identify the best place to open a new coffee shop.

Data Requirements and Sources

1. Dataset 1:

The below link contains information about the boroughs and the different neighbourhoods under each borough of NYC along with its latitude and longitude coordinates.

https://geo.nyu.edu/catalog/nyu_2451_34572
https://cocl.us/new_york_dataset

2. Dataset 2:

The below link contains information about the demographics of NYC and segmentation of ethnic races across different boroughs.

https://en.wikipedia.org/wiki/Demographics_of_New_York_City

3. Dataset 3:

The Foursquare API will be used to obtain the geographical location data for the NYC Area. These will be used to explore the venues in the neighbourhoods of NYC. The venues will provide the categories needed for the analysis, such as the coffee shop.

Methodology and Data Analysis

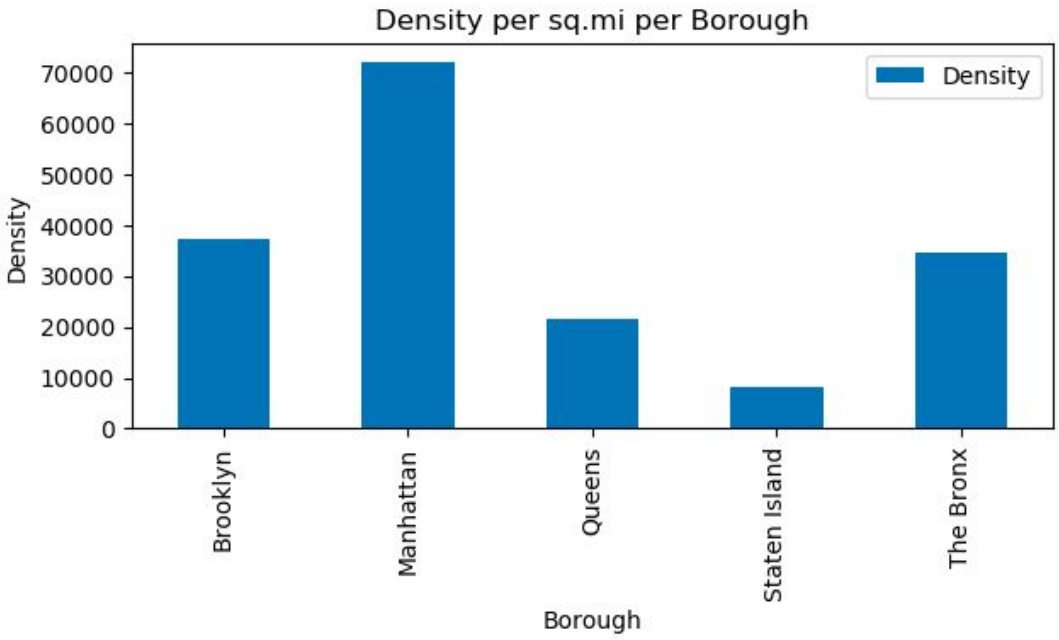
1. Density of People per Borough

The first two datasets are analysed by considering the venues within the neighbourhood of NYC postal code areas. The combined data will be used to determine the highest density of people

in each Borough. The density of each borough in New York was achieved through the extraction of data. With the various acrapping of data from the source the following data is extracted:

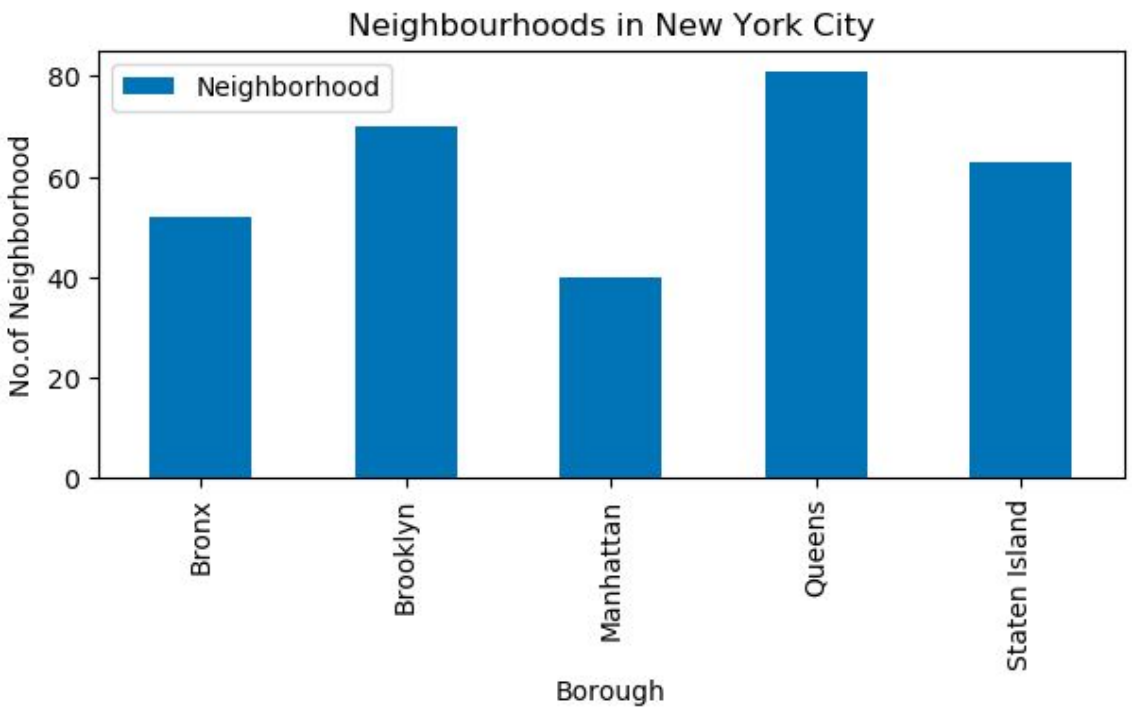
	Borough	County	Population	Density
0	The Bronx	Bronx	1432132	34653
1	Brooklyn	Kings	2582830	37137
2	Manhattan	New York	1628701	72033
3	Queens	Queens	2278906	21460
4	Staten Island	Richmond	476179	8112

Then, I use the bar chart to visualize the density of people per sq. mi for each borough. The below graph shows that the Manhattan area is the most crowded area based on density.

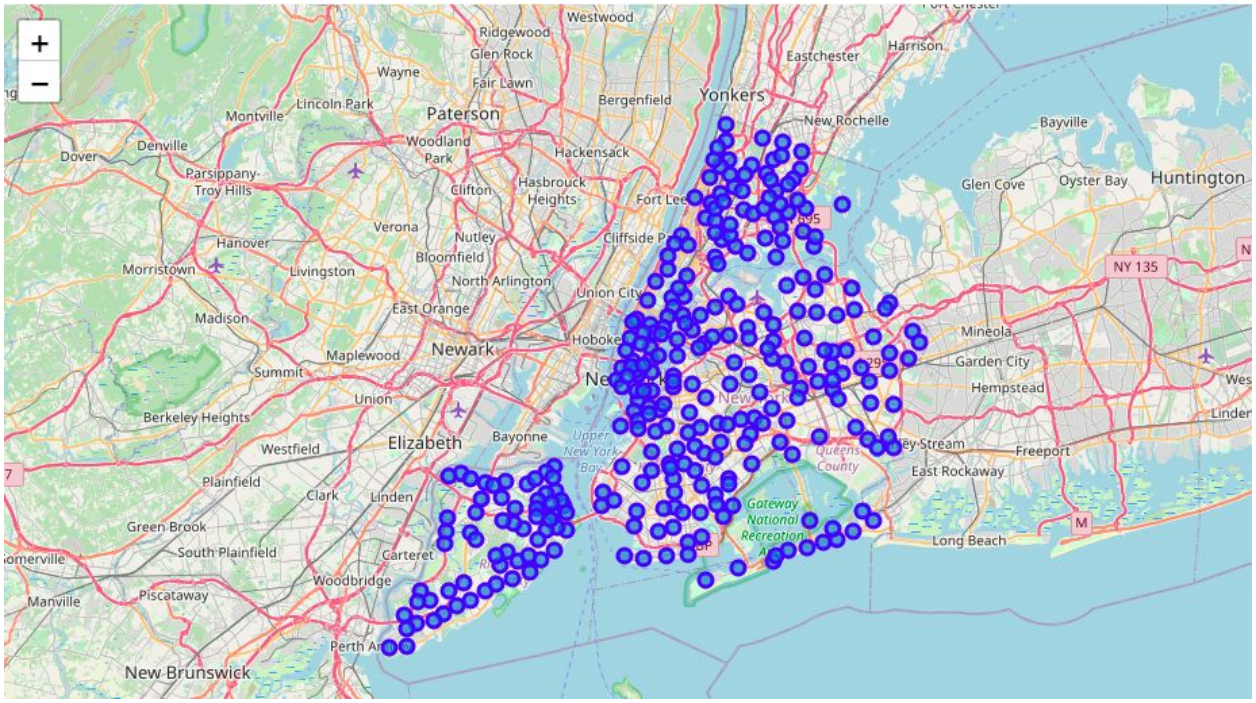


2. Neighbourhoods in New York City

The next step is to understand the number of neighborhoods in each borough and how they are spreaded in the area. The below graph was obtained through scraping the New York city data and visualized by plotting a bar chart. It is noted that Manhattan has the fewest number of neighbourhoods and potentially has the highest population density in each neighbourhoods.



To better understand how neighborhoods are spreaded in New York city, I used the geopy library to get the latitude and longitude values of New York City and created a map of New York with neighborhoods superimposed on top, as shown below.

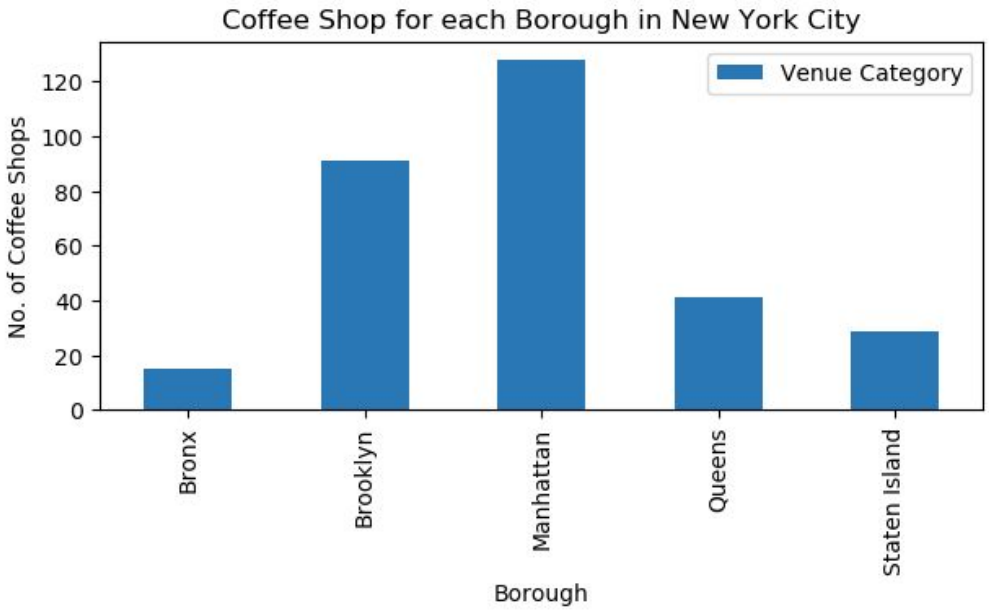


3.3. Number of gyms in each Borough

The next metric was to validate and source the number of gyms per borough. This will give valuable insight of competitors and the location of the competitors. When extracting data from Foursquare the follow number of gyms were noted in each area.

3. Coffee Shops in New York City

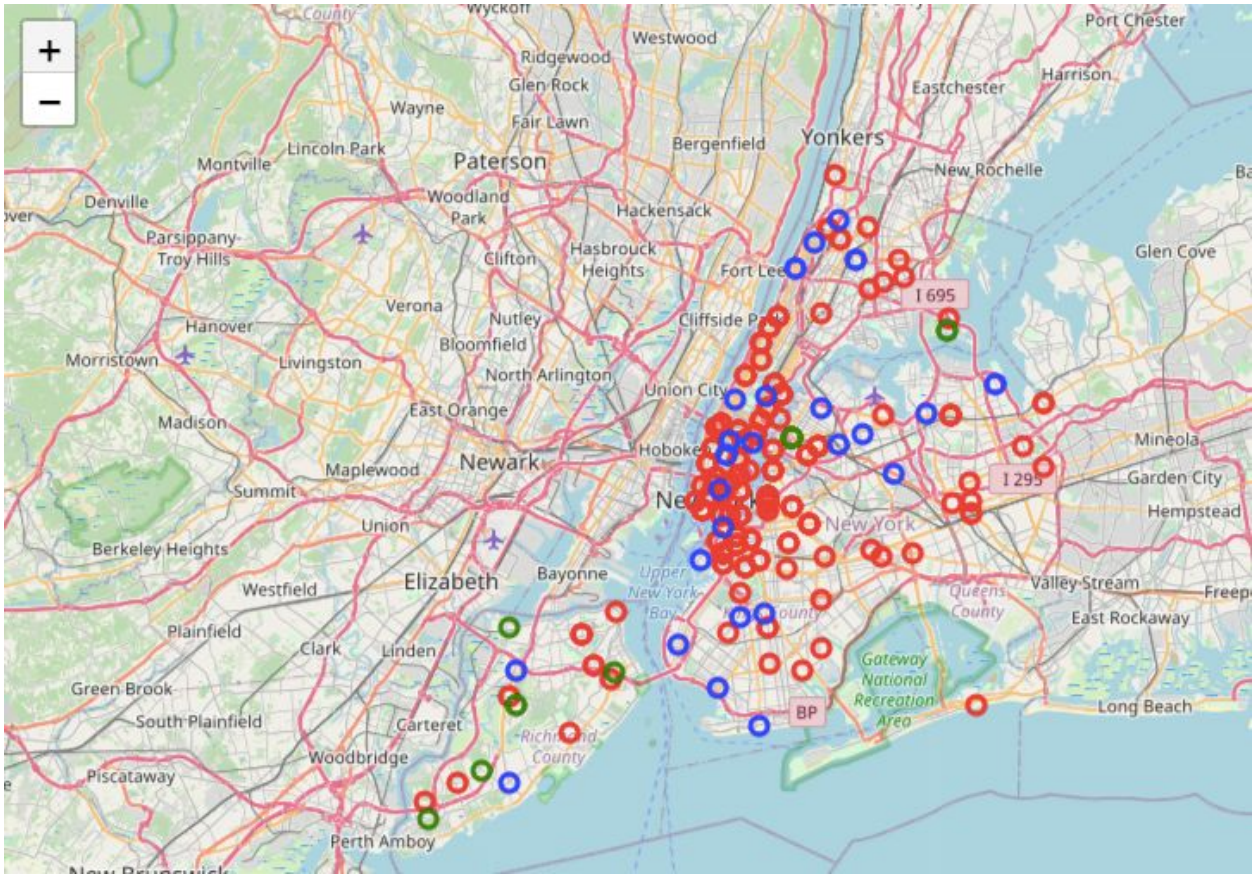
Then I decided to study the number of coffee shop in New York city. This will provide valuable insights of the current competitors and the location of them. I used the Foursquare API to obtain all venues in the neighborhoods and then selected the coffee shop from the venue category. Due to Foursquare restrictions, the number of venues will be limited to 100 venues. The coffee shops in each borough in New York city are shown as below. It shows that Manhattan has the most number of coffee shops.



4. Clusters

By using k-means clustering, I have clustered the coffee shops in New York into 3 clusters based on their venue category for “coffee shop”. The results are visualized in the below map

with 3 clusters in red, blue and green. It can also be noted that the red cluster areas are mostly shown in Manhattan, while the green cluster has the least coffee shops shown on the map.



Map: Coffee Shops in New York Clustered into 3 Categories

Conclusion and Future Decision

With all the above taken into consideration, the best potential area to open a new coffee shop is to evaluate the neighborhoods in the Manhattan area based on the density and people per coffee shop as well as the coffee shop spread in New York city. It is noted that the area in the red cluster, which is neighborhood Chelsea, Civic Center, Carnegie Hill etc., is the most popular area for coffee shops.

Future decisions should be made based on other factors such as cost of location, logistics, rental fees and so on for the exact neighborhood to choose in Manhattan within the red cluster area.