Coffee shop Analysis in New York City

1. Introduction

1.1 Background

New York City is a famous destination for people around the world. There are millions of tourists visit this island each year. The popular land marks that attracted people all over the world are such as Liberty Statue, 911 Memorial, Times Square or Empire State Building.

Not only those attractive land marks for tourists but New York is also popular about coffee shops & restaurants. There are many good variety coffee shops & restaurants in this big apple city. Thus, if you are a new entrepreneur who wants to open a new coffee shop in New York it will be very challenging as there are highly competitive as well as high demanding.

There are five boroughs in New York City: The Bronx, Brooklyn, Manhattan, Queens and Staten Island. Each borough is coextensive with a respective county of New York States. Each borough is composed of many neighborhoods in the area which has its own unique characteristics. The location to open a new coffee shop is an important factor to become success or failed in this business.

1.2 Business Problem

Which borough and neighborhood are the best place to open the new coffee shop in New York City?

2. Data

Data used in this analysis are as following:

a. New York City Borough and Neighborhood
   https://geo.nyu.edu/catalog/nyu_2451_34572
b. Foursquare location data
   - Neighborhood Location
   - Venue location
   - Venue Category

3. Methodology and Results

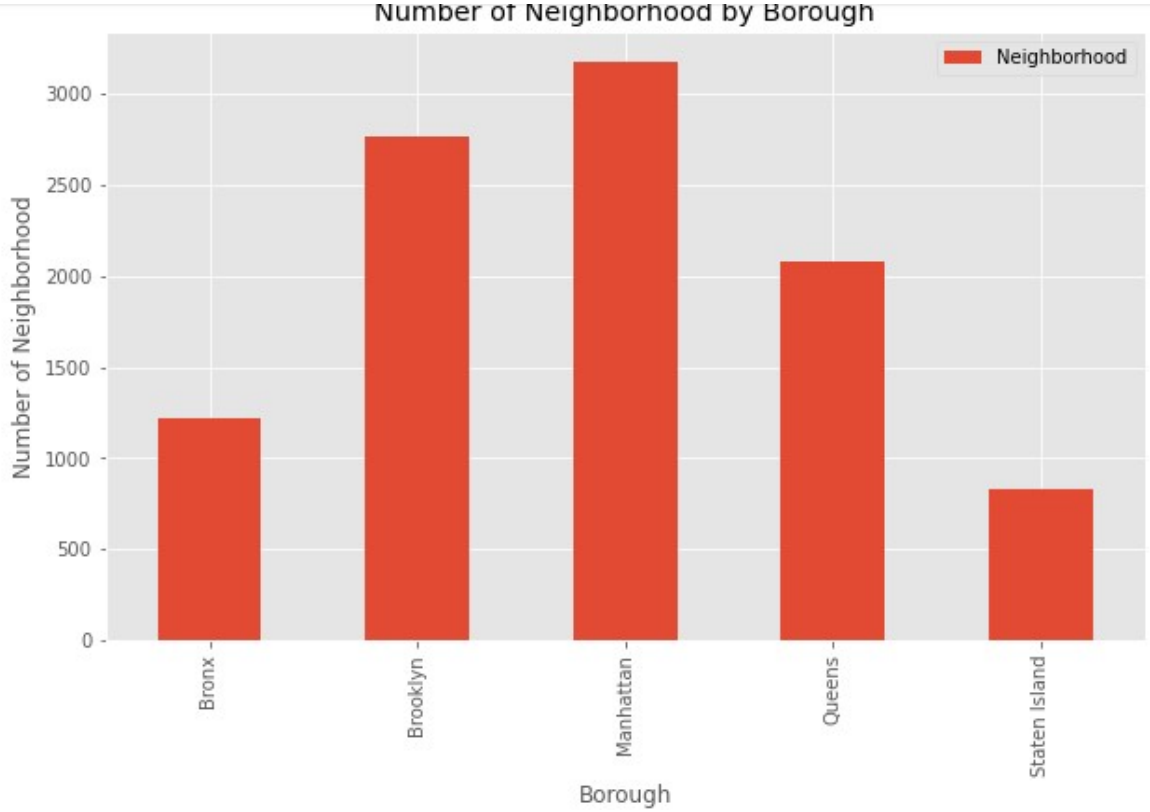
3.1 Methodology

The steps used in this capstone project are described as following

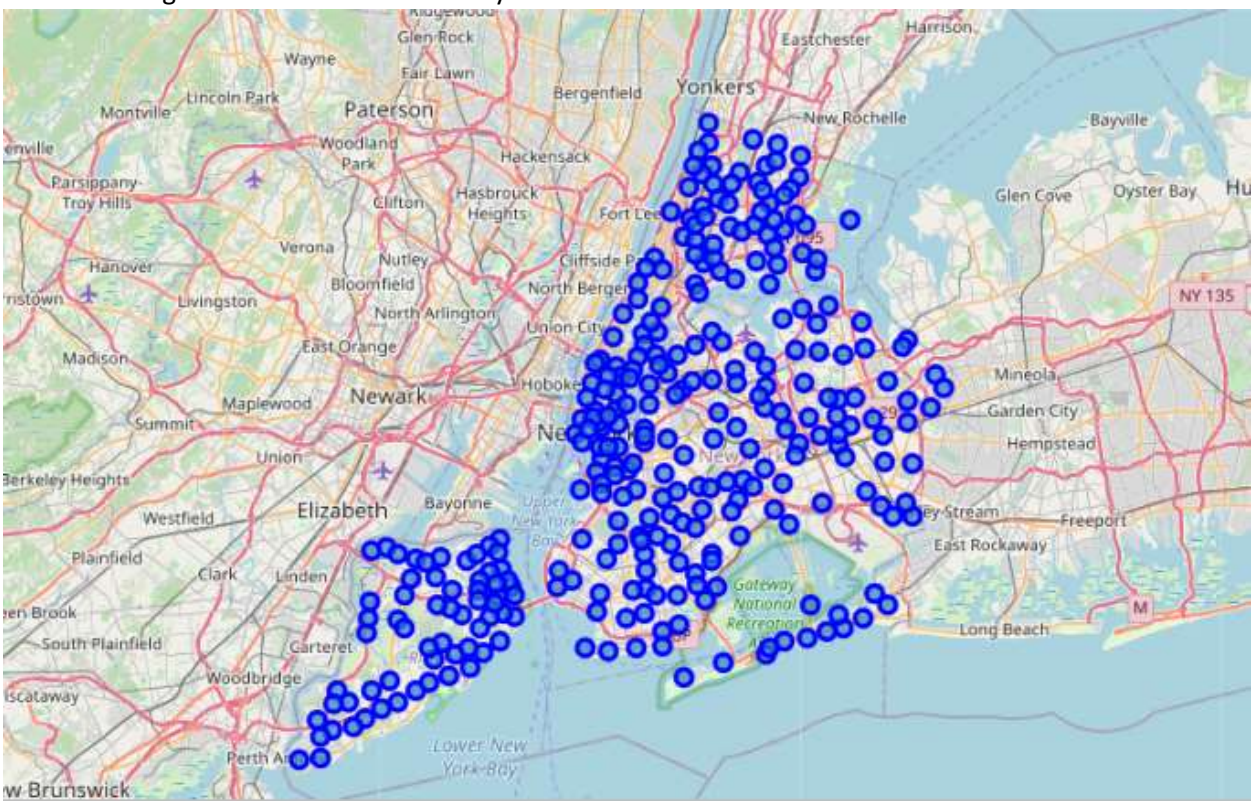
a. Load New York city borough and neighborhoods location from the files to get list of borough and neighborhoods
b. Using Foursquare data to get all venues and locations in each boroughs and neighborhoods
c. Summarize number of venue category by borough and neighborhoods
d. Cluster all venues category in New York City using K-Means Clustering
e. Identify the cluster that contains coffee shop in the area to be the potential area to open a new business

3.2 Data Analysis and Results

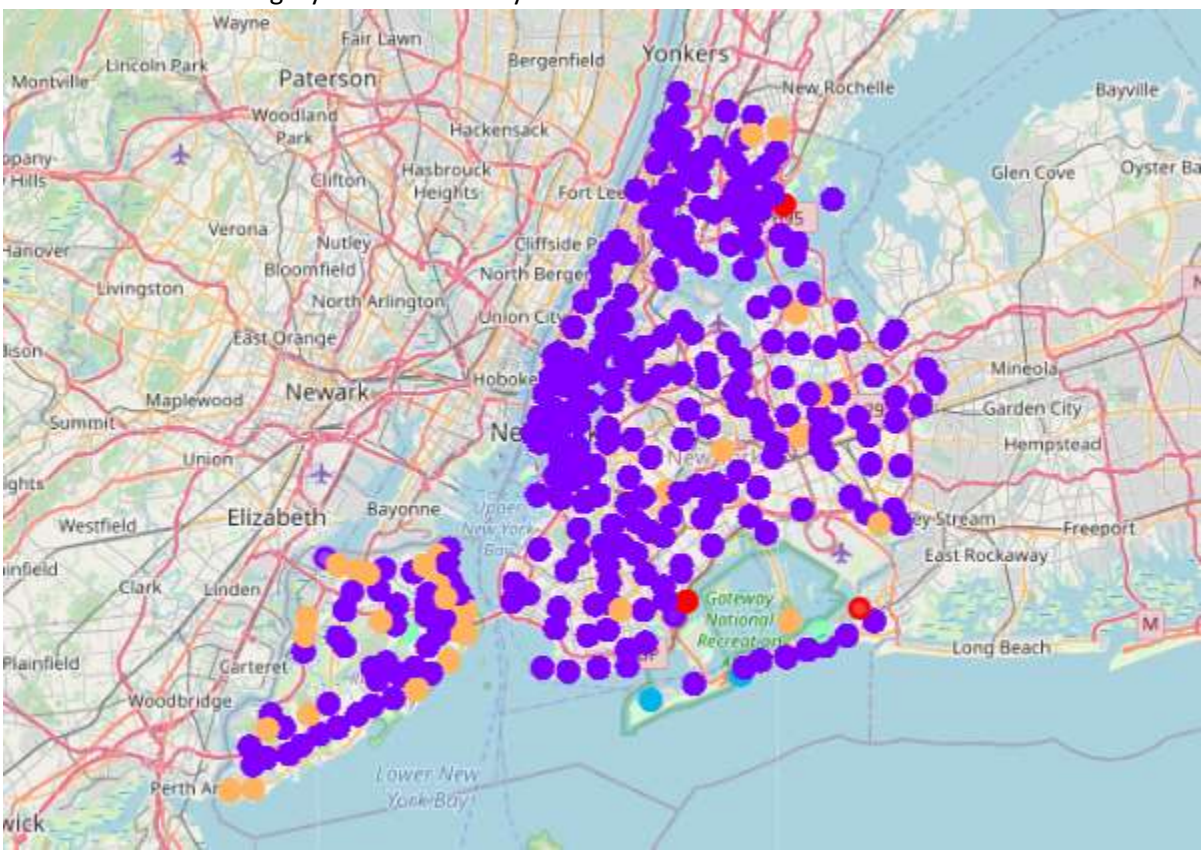
a. Number of borough and neighborhood in New York City



b. Plot of all Neighborhoods in New York City



c. Cluster all venues category in New York City in to 5 clusters as shown below



d. With 5 clusters there are only 3 clusters that contains coffee shops which are cluster 1 and 4

e. Number of coffee shop in cluster 1 and 4

```
newyork_venues_cluster.groupby('Cluster Labels')['Coffee Shop'].mean()
```

```
Cluster Labels
0    0.000000
1    0.021719
2    0.000000
3    0.000000
4    0.022672
Name: Coffee Shop, dtype: float64
```

```
newyork_venues_cluster.groupby('Cluster Labels')['Coffee Shop'].count()
```

```
Cluster Labels
0      3
1    268
2      2
3      2
4     30
Name: Coffee Shop, dtype: int64
```

f. List of Borough and Neighborhood in selected clusters

Cluster Labels	Borough	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	Bronx	Bedford Park	1	1	1	1	1	1
		Belmont	2	2	2	2	2	2
		Concourse Village	1	1	1	1	1	1
		Edgewater Park	1	1	1	1	1	1
		Kingsbridge	1	1	1	1	1	1
...	...	...	...	...	...	...	...	...
4	Brooklyn	Ocean Hill	1	1	1	1	1	1
	Queens	Briarwood	1	1	1	1	1	1
	Staten Island	Arden Heights	1	1	1	1	1	1
		Arlington	1	1	1	1	1	1
		Park Hill	1	1	1	1	1	1

121 rows × 6 columns

4. Discussion and Recommendation

From 5 clusters, the result shown that there are only 2 clusters that has a running coffee shops business while the other 3 clusters do not. Thus, there are 2 options for entrepreneur to consider:

Option 1: open new coffee shop in selected clusters (cluster 1 and 4) as these borough and neighborhoods already have coffee shops business running.
Cluster 1: Borough = Bronx with 116 Neighborhoods
Cluster 4: Borough = Brooklyn (1 Neighborhood), Queens (1 Neighborhood) and Staten Island (3 Neighborhoods)

Even though there will be a lot of competitors but it is proved that these areas are success for coffee shop business. If the new entrepreneur has a very unique in the coffee shop there is a chance to be success.

Considering Cluster 4 would be a good opportunity with less coffee shops in the area.

Option 2: open new coffee shop in cluster 0, 2 and 3 as these borough and neighborhoods do not have any coffee shops opened yet. It is risky to start new business in these areas so that further analysis needs to be take into account if entrepreneur would take a risk to open in these areas. As there are no coffee shop in the neighborhood yet, thus the success rate is high.

5. Conclusion

K-Means is vastly used for clustering in many data science application, in this study we use K-Means to cluster the neighborhood to identify the potential area that is the best to open a new coffee shop business in New York City.

Clustering technique can be used to roughly identify which borough within 5 boroughs in the New York City we should do the next deep dive analysis in the specific areas for further analysis to find the success factors to open a new coffee shop using correlation technique.