Local Competitive Land Scape for Food Outlet Opening in Manhattan

1.Introduction

1.1 Background

Consider a scenario where there are some businessmen who want to invest their capital in eatery business by launching new food outlets in Manhattan. It requires a right business strategy in order to have a successful eatery business. Majority of the strategy should be confined to the process of food outlet opening. There are a lot of things to consider while opening a food outlet. Among them, competitor analysis plays a crucial role.

The competitor analysis serves many purposes. By understanding the competitors, we position ourselves to truly understand our market and create value propositions, differentiators, and a marketing strategy that goes above and beyond the competition. One important step in competitor analysis is understanding the geo-distribution of competitors. The old real-estate dictum that location should be a key factor in deciding on a property is just as true for food outlets. Where our food outlet is located can be just as important to your success as our menu, marketing, or customer reviews.

Moreover, it is also quite important to perform the location wise competitor analysis before we decide about the type of food outlet we want to start. When we are choosing our location and type of outlet, you need to map out the local competitive landscape to know exactly who we are up against. Doing this will also help you determine whether a particular neighborhood is ripe for the picking or is over-saturated with competitors. This is very critical for our food outlet because it affects your ability to draw customers. Our location and concept (type of food outlet) must complement each other. A good location itself draws a greater customer.

1.2 Objective

In this capstone project, we aim to provide valuable insight on local competitive landscape to those who want to establish new food outlet/eatery in Manhattan. Given the location data of various types of food outlets in Manhattan, we predict the dominant/ common type of food outlets present in different parts of Manhattan which will be quite useful while in deciding about the location and type of new food outlet.

1.3 Stake Holders

Mainly, the businessmen who are interested in setting up a new food outlet/eatery in Manhattan would be interested very interested to have the details of local competitive landscape as a input for their business strategy making. Besides, any already existing eatery owners also would be interested to have the details of this analysis as it would help them understand the geo-distribution of competitors in that locality in order to refine their business strategy.

2. Data acquisition and processing

In this project, first, we obtain the data that contains the boroughs and the neighborhoods of the New York City. It also contains the latitude and longitude of each neighborhood of every borough in the city. Later, using them, we obtain the data of food outlets/ eateries (along with their type) inside each Borough.
The details of Boroughs and their neighborhoods of New York city along with their latitudes and longitudes are obtained from https://geo.nyu.edu/catalog/nyu_2451_34572 in a '.json'file. We load the '. json' file and extract the 'features' data which has four columns 'Borough', 'Neighborhood', 'Latitude', 'Longitude'. Now, this data is transformed into a Pandas data frame. There are 5 Boroughs (Manhattan, Brooklyn, Queens, Staten Island, and Bronx) and 306 Neighborhoods present in New York city.

We split the dataframe into a dataframe based on boroughs, each containing the neighborhood, latitude, and longitude data of the borough. For instance, the first 6 rows of the dataframe containing Manhattan data is as below:

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

Using the FourSquare API credentials, the data of food outlet/eatery is obtained. we performed the following:
a) we obtained a maximum of 1,000 food outlets present inside 15,000 meters radius along with their type using the latitude and longitude of each neighborhood present inside the given borough.
b) This data is merged with the corresponding data frame that contains the neighborhood, latitudes, and longitude values. The first 5 rows of resultant data frame is as shown below: (Note: Venue represents food outlet/eatery and Venue category represents the type of food outlet/eatery)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
2	Marble Hill	40.876551	-73.91066	Dunkin'	40.877136	-73.906666	Donut Shop
3	Marble Hill	40.876551	-73.91066	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant
4	Marble Hill	40.876551	-73.91066	Parrilla Latina	40.877473	-73.906073	Steakhouse

3. Methodology

In this section, we perform the exploratory data analysis on dataframe. (Note : Venue represents food outlet/eatery and Venue category represents the type of food outlet/eatery)

In Manhattan, there are 122 neighborhoods. The foursquare API returned a total of 2,634 food outlets to the request for maximum of 1,000 food outlets in each neighborhood within a radius of 15,000 meters. There are 121 unique types of food outlets among them. The type of food outlets and their number of outlets inside the Manhattan borough.

Then we performed one-hot coding based on food outlet types to obtain a new dataframe that contains 2,634 rows and 112 columns. The first 15 rows of this dataframe are as shown in figure below.

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant
0	Battery Park City	0.0	0.000000	0.040000	0.000000	0.000000	0.000000	0.00
1	Carnegie Hill	0.0	0.000000	0.014706	0.000000	0.014706	0.000000	0.00
2	Central Harlem	0.0	0.065217	0.043478	0.000000	0.000000	0.000000	0.00
3	Chelsea	0.0	0.000000	0.060606	0.015152	0.000000	0.015152	0.00
4	Chinatown	0.0	0.000000	0.030000	0.000000	0.000000	0.020000	0.01

Next, we grouped the rows of above dataframe by neighborhood and took the mean of the frequency of occurrence of each type of food outlet.

Clustering

In this section, we will perform predictive modeling of the data by using K-means clustering. This clustering technique partitions the data into K distinct non-overlapping clusters such that intra cluster data samples will be as similar as possible while keeping the clusters as dissimilar as possible. It assigns data points to a cluster such that the sum of the squared distance between the data samples and the cluster's centroid (arithmetic mean of all the data samples that belong to that cluster) is at the minimum. The less variation we have within clusters, the more homogeneous (similar) the data samples are within the same cluster.

In our problem, clustering helps to divide the neighborhoods of a given borough into clusters so that each neighborhood in a given cluster will have similarity in the types of food outlets with respect to the other neighborhoods within the same cluster as well as dissimilarity with respect to the neighborhoods present in the different cluster. With this, we can get to know the common/dominant type of food outlets in each cluster of a borough which is essentially the landscape of type of food outlets located inside a borough.

Here, we discuss the K-mean clustering of neighborhoods in the Manhattan borough We use the dataframe that contains the frequency of occurrence of food outlets type as shown in Fig. by dropping the neighborhood column. We run the K-means clustering algorithm Using Scikit Learn library for different values of K i.e from K=2 to K=7.

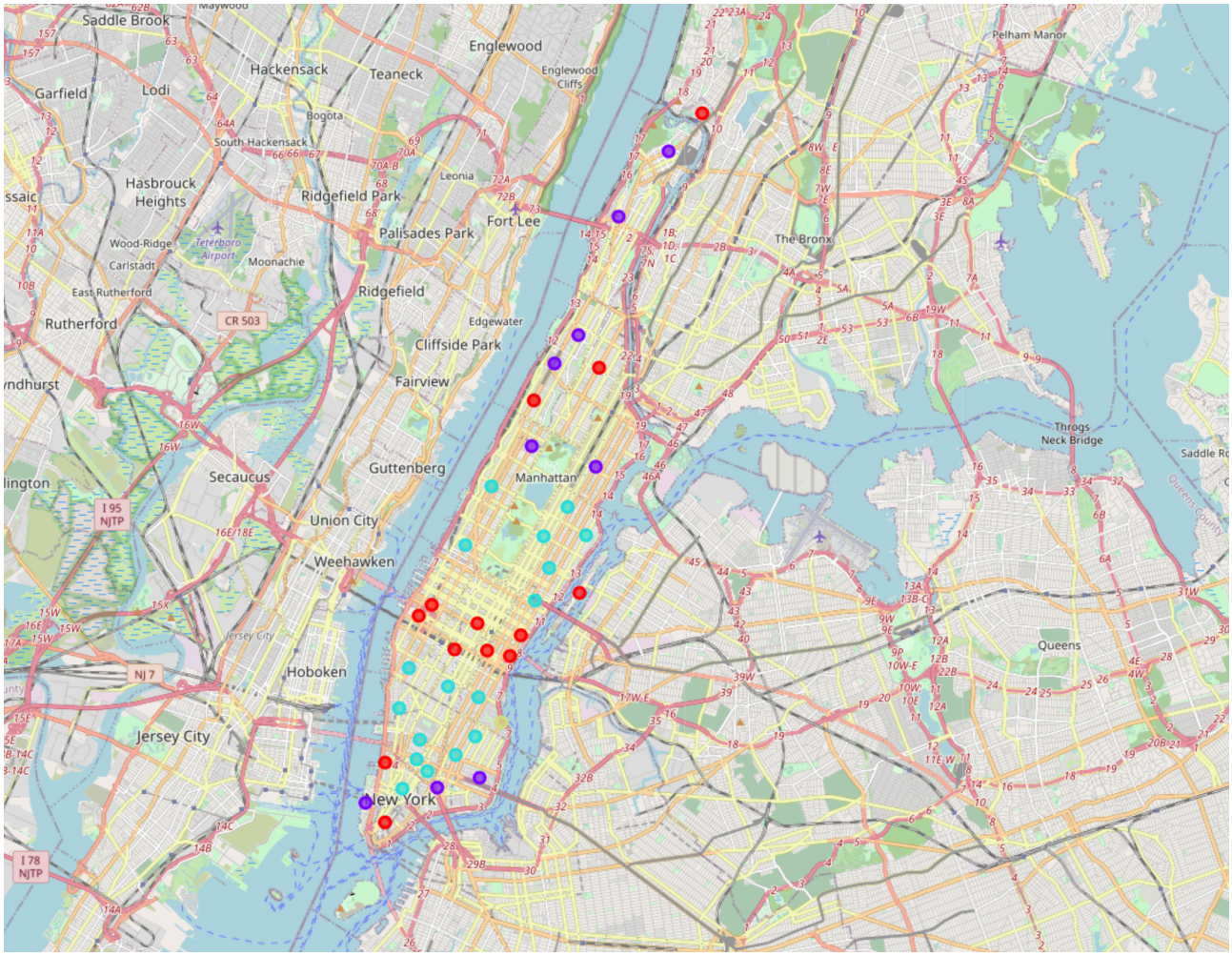
We plot the sum of squared distance value vs K. We select the optimum value of K using elbow rule where optimum K=4 in the case of Manhattan data.

After clustering, the dataframe that contains the top 10 common food outlet types with cluster labels assigned to them will be as shown below:

	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	Battery Park City	Pizza Place	Sandwich Place	Burger Joint	Food Court	Chinese Restaurant	Food Truck	Steakhouse	Donut Shop	American Restaurant	Café
1	Carnegie Hill	Pizza Place	Café	Italian Restaurant	Bakery	Sushi Restaurant	Japanese Restaurant	French Restaurant	Restaurant	Diner	New American Restaurant
2	Central Harlem	Fried Chicken Joint	Deli / Bodega	Southern / Soul Food Restaurant	Chinese Restaurant	Caribbean Restaurant	African Restaurant	Seafood Restaurant	French Restaurant	Sandwich Place	Pizza Place
3	Chelsea	Café	French Restaurant	Bakery	American Restaurant	Italian Restaurant	Japanese Restaurant	Mexican Restaurant	Sandwich Place	Seafood Restaurant	Pizza Place
4	Chinatown	Chinese Restaurant	Bakery	Dumpling Restaurant	Vietnamese Restaurant	Malay Restaurant	Dim Sum Restaurant	Noodle House	Mexican Restaurant	Hotpot Restaurant	Pizza Place

4. Results

In case of Manhattan borough, the K-mean clustering with K=4 partitioned the neighborhoods of Manhattan into 4 clusters as shown in Fig. Each dot represents a neighborhood and similar color dots represents a cluster. The type of food outlets and their number of outlets present in neighborhoods of each of the four clusters in Manhattan are as shown in figure.

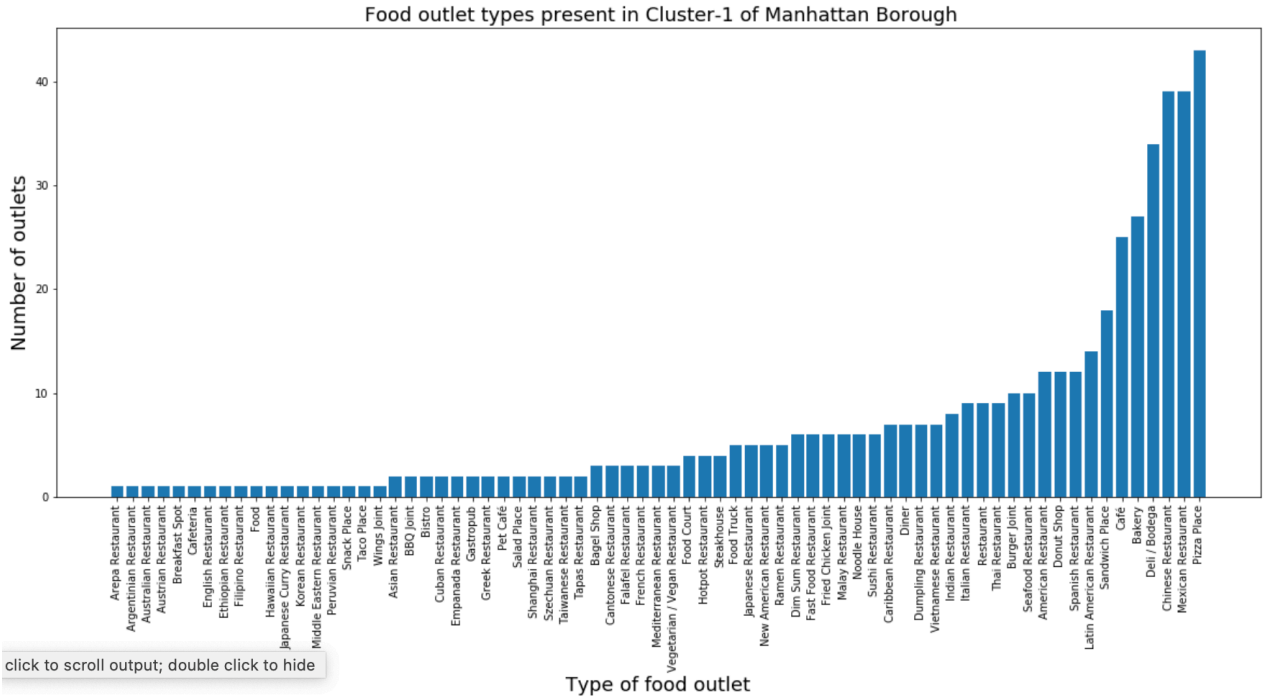


5. Discussion

Now, you can examine each cluster and determine the discriminating venue categories that distinguish each cluster.

Cluster 1

We can observe that Bakery, Deli / Bodega, Chinese Restaurant, Mexican Restaurant, Pizza Place are the top 5 common/dominant types of food outlets present in Cluster-1 of Manhattan. This suggests that any new food outlet to be launched in Cluster-1 make sure that type of food offered is different to those 5 types of food outlets in order to draw more customers. It also suggests that Cluster-1 of Manhattan is not the ideal place to open a new Bakery, Deli / Bodega, Chinese Restaurant, Mexican Restaurant, Pizza Place.

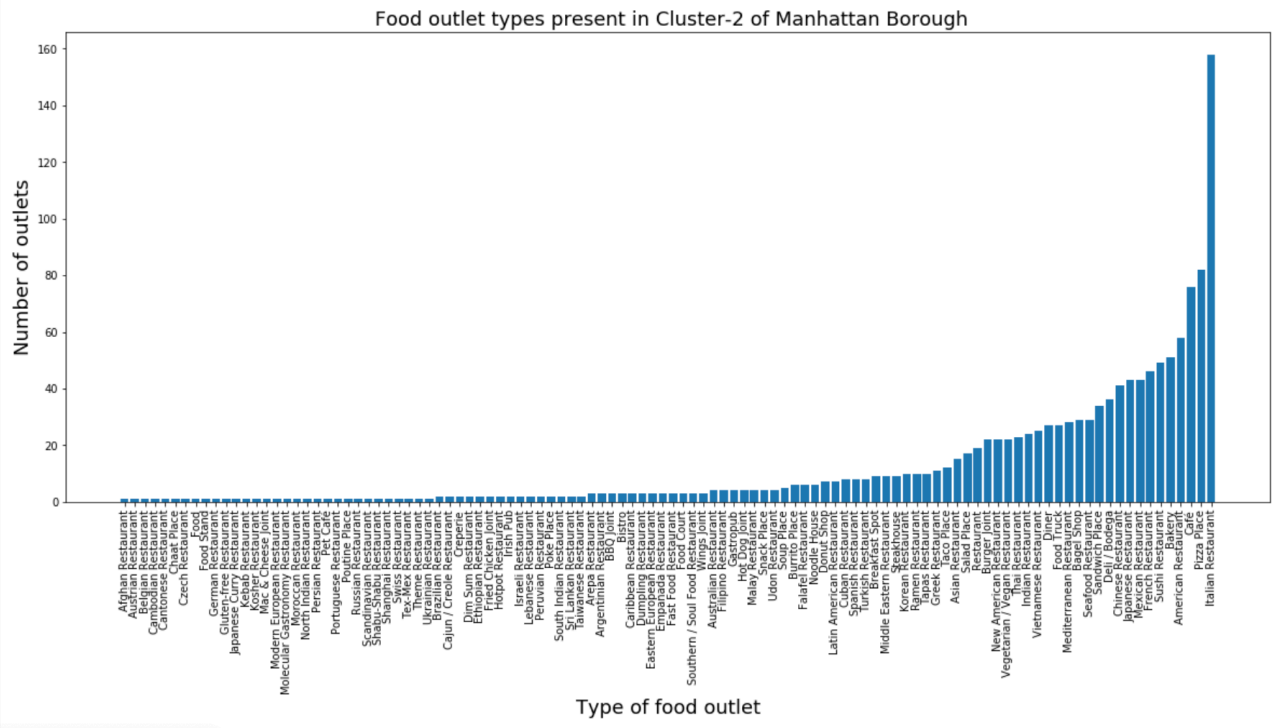


Neighborhoods present inside Cluster-1 are

- Chinatown
- Washington Heights
- Inwood
- Hamilton Heights
- Manhattanville
- East Harlem
- Lower East Side
- Manhattan Valley
- Battery Park City

Cluster 2

We can observe that Bakery, American Restaurant, Cafe, Pizza Place, Italian Restaurant are the top 5 common/dominant types of food outlets present in Cluster-2 of Manhattan. This suggests that any new food outlet to be launched in Cluster-2 make sure that type of food offered is different to those 5 types of food outlets in order to draw more customers. It also suggests that Cluster-2 of Manhattan is not the ideal place to open a new Bakery, American Restaurant, Cafe, Pizza Place, Italian Restaurant.

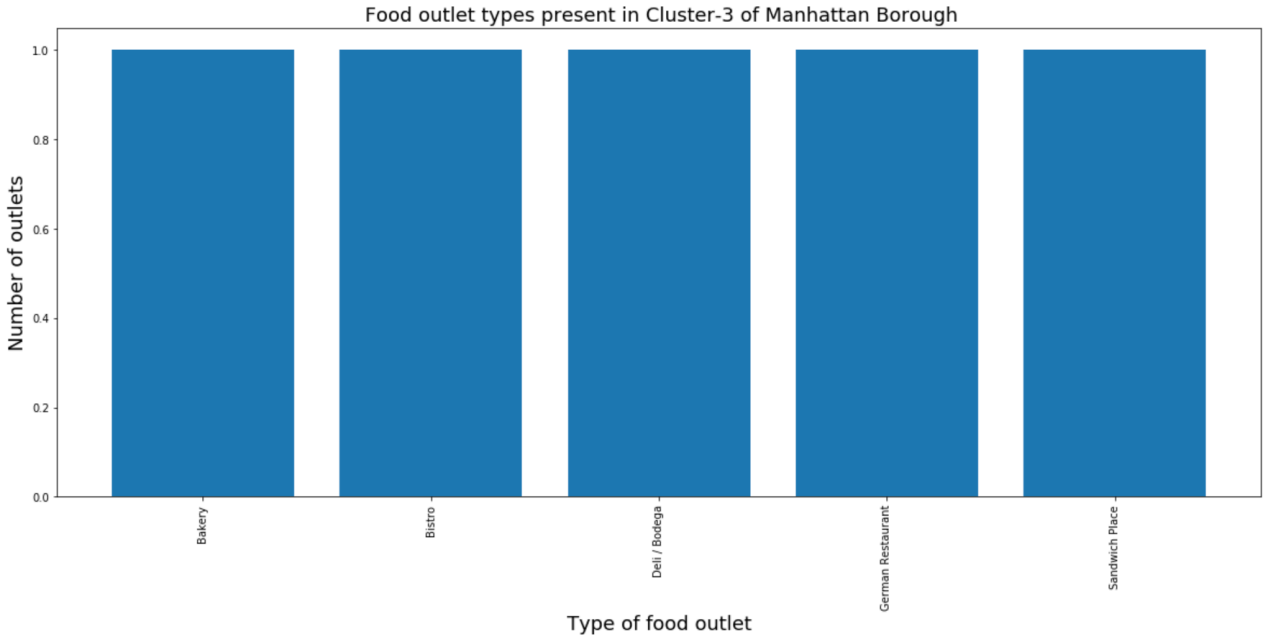


Neighborhoods present inside Cluster-2 are

- Upper East Side
- Yorkville
- Lenox Hill
- Upper West Side
- Lincoln Square
- Chelsea
- Greenwich Village
- East Village
- Little Italy
- Soho
- West Village
- Gramercy
- Carnegie Hill
- Noho
- Civic Center
- Sutton Place
- Flatiron

Cluster 3

We can observe that Bakery, Bistro, Deli / Bodega, German Restaurant, Sandwich Place are the top 5 common/dominant types of food outlets present in Cluster-3 of Manhattan. This suggests that any new food outlet to be launched in Cluster-3 make sure that type of food offered is different to those 5 types of food outlets in order to draw more customers. It also suggests that Cluster-3 of Manhattan is not the ideal place to open a new Bakery, Bistro, Deli / Bodega, German Restaurant, Sandwich Place are.

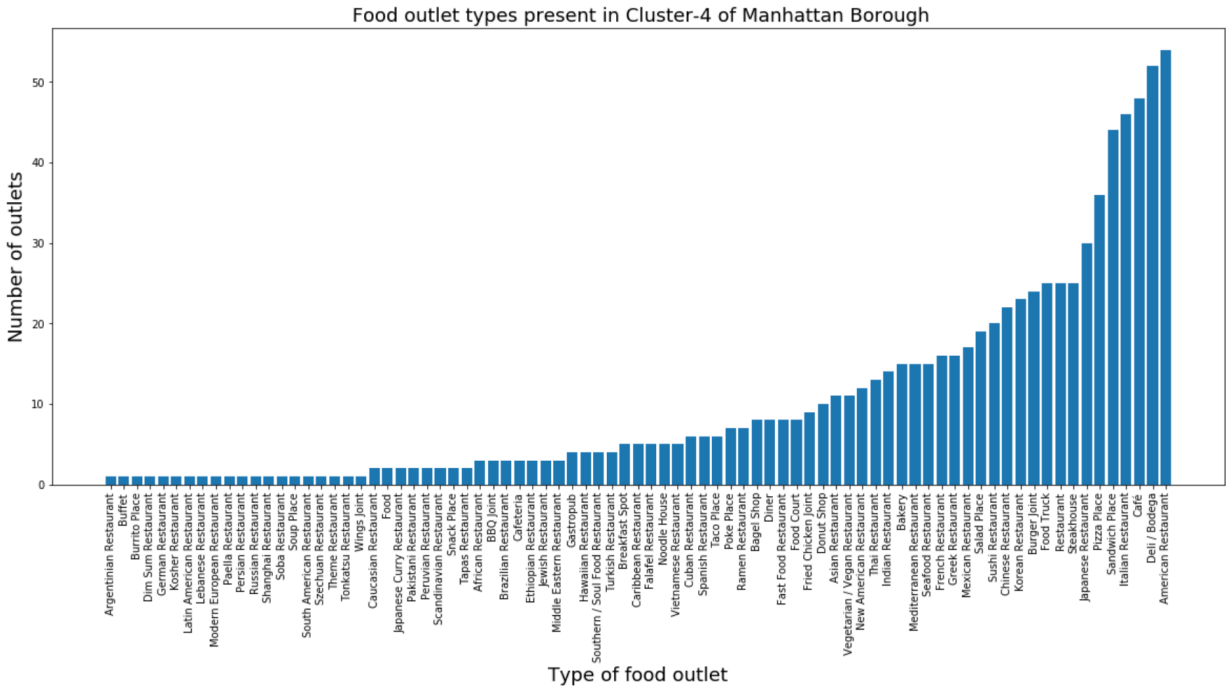


Neighborhoods present inside Cluster-3 are

- Stuyvesant Town

Cluster 4

We can observe that Sandwich Place, Italian Restaurant, Cafe, Deli / Bodega, American Restaurant are the top 5 common/dominant types of food outlets present in Cluster-4 of Manhattan. This suggests that any new food outlet to be launched in Cluster-4 make sure that type of food offered is different to those 5 types of food outlets in order to draw more customers. It also suggests that Cluster-4 of Manhattan is not the ideal place to open a new Sandwich Place, Italian Restaurant, Cafe, Deli / Bodega, American Restaurant.



Neighborhoods present inside Cluster-4 are

- Marble Hill
- Central Harlem
- Roosevelt Island
- Clinton
- Midtown
- Murray Hill
- Tribeca
- Morningside Heights
- Financial District
- Midtown South
- Turtle Bay
- Tudor City
- Hudson Yards

6. Conclusion

In this project, we predicted the dominant/common type of food outlets present in Manhattan with help of location data of various types of food outlets in Manhattan. Using the latitudes and longitudes of neighborhoods in New York city, we obtained the location data of different types of food outlets with the help of Foursquare API. Neighborhoods of Manhattan borough in New York city has been partitioned into different clusters using K-means clustering algorithm by taking top 10 common food outlets of each neighborhood as a data sample. This analysis can be helpful for those businessmen who are planning to open a new food outlet in Manhattan. It can help them in deciding the type of food outlet that provides them the competitive advantage