Local Competitive Land Scape for Food Outlet Opening in New York city

K. V. S Sai Sushanth

April 30, 2020

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 New York city. 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 your food outlet because it affects your ability to draw customers. Your 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 New York city. Given the location data of various types of food outlets in New York city, we predict the dominant/common type of food outlets present in different parts of New York city 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 New York city 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 dataframe. There are 5 Boroughs (Manhattan, Brookyln, Queens, Staten Island, and Bronx) and 306 Neighborhoods present in New York city.

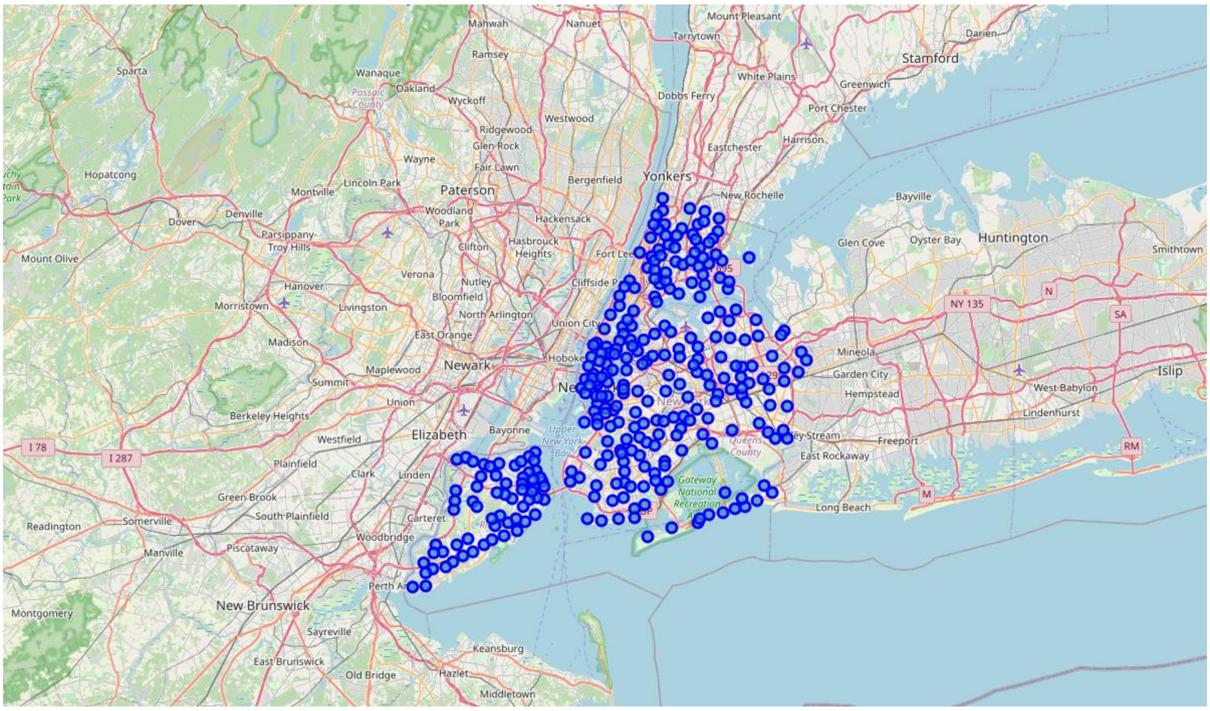

Fig.1: Map showing different neighborhoods present in the New York city

We split the dataframe into 5 dataframes 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
5	Manhattan	Manhattanville	40.816934	-73.957385

Fig.2 First 6 rows of dataframe containing Latitude and Longitudes of neighborhoods in NY

Using the FourSquare API credentials, the data of food outlet/eatery is obtained. For each Borough in NY city, we performed the following:

a) we obtained a maximum of 25 food outlets present inside 500 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 in case of Manhattan borough 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	-73.910660	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.910660	Tibbett Diner	40.880404	-73.908937	Diner
2	Marble Hill	40.876551	-73.910660	Dunkin'	40.877136	-73.906666	Donut Shop
3	Marble Hill	40.876551	-73.910660	Land & Sea Restaurant	40.877885	-73.905873	Seafood Restaurant
4	Marble Hill	40.876551	-73.910660	Parrilla Latina	40.877473	-73.906073	Steakhouse

Fig.3 First 5 elements of a dataframe containing food outlets in different neighborhoods

3.Exploratory Data Analysis:

In this section, we perform the exploratory data analysis on dataframe corresponding to each borough. Note that we will discuss the analysis performed on the data related to Manhattan borough and the same is applied on the remaining four boroughs as well. (Note : Venue represents food outlet/eatery and Venue category represents the type of food outlet/eatery)

In Manhattan, there are 40 neighborhoods. The foursquare API returned a total of 956 food outlets to the request for maximum of 25 food outlets in each neighborhood within a radius of 500 meters. There are 90 unique types of food outlets among them. The type of food outlets and their number of outlets inside the Manhattan borough are as shown in Fig.4

Then we performed one-hot coding based on food outlet types to obtain a new dataframe that contains 956 rows and 91 columns. The first 15 rows of this dataframe are as shown in Fig.5

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. The first 10 rows of the result obtained are as shown in Fig.6

We then prepare a dataframe that contains the top 10 most dominant/common food outlets present in each neighborhood based on their mean frequency. The first 5 rows of this dataframe contains are shown in the Fig.7

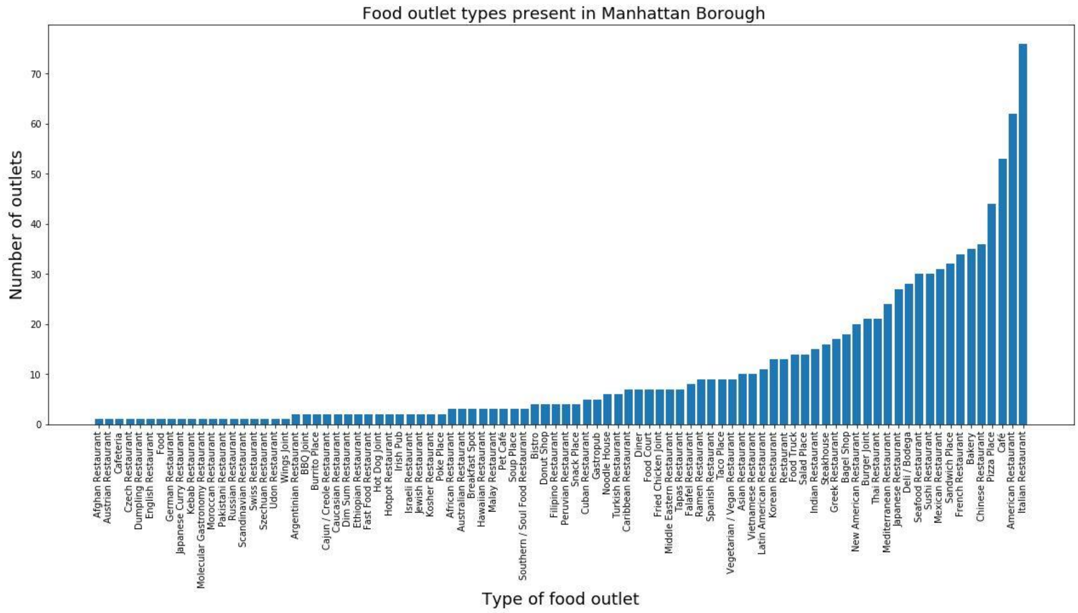

Fig.4 Plot showing different types of food outlets and their number of outlets inside the Manhattan borough

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bistro	Breakfast Spot	Burger Joint
0	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0
5	Marble Hill	0	0	1	0	0	0	0	0	0	0	0	0	0
6	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0
7	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0
8	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0
9	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0
10	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0
11	Marble Hill	0	0	1	0	0	0	0	0	0	0	0	0	0
12	Marble Hill	0	0	0	0	0	0	0	0	0	0	0	0	0
13	Chinatown	0	0	0	0	0	0	0	0	0	0	0	0	0
14	Chinatown	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig.5 First 15 rows of a new dataframe with one-hot coding based on food outlet types

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Argentinian Restaurant	Asian Restaurant	Australian Restaurant	Austrian Restaurant	BBQ Joint	Bagel Shop	Bakery	Bistro	Breakfast Spot	Burger Joint
0	Battery Park City	0.00	0.00	0.040000	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.04	0.00	0.08
1	Carnegie Hill	0.00	0.00	0.040000	0.00	0.00	0.00	0.00	0.00	0.04	0.08	0.00	0.04	0.00
2	Central Harlem	0.00	0.12	0.080000	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.04
3	Chelsea	0.00	0.00	0.120000	0.00	0.04	0.00	0.00	0.00	0.00	0.08	0.00	0.00	0.04
4	Chinatown	0.00	0.00	0.080000	0.00	0.08	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00
5	Civic Center	0.00	0.00	0.080000	0.00	0.04	0.04	0.00	0.00	0.04	0.08	0.00	0.00	0.00
6	Clinton	0.00	0.00	0.120000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	East Harlem	0.00	0.00	0.000000	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00
8	East Village	0.00	0.00	0.040000	0.00	0.00	0.00	0.00	0.00	0.04	0.04	0.00	0.00	0.00
9	Financial District	0.00	0.00	0.120000	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.00	0.00

Fig.6 First 10 rows of a dataframe containing mean frequency of type of food outlets.

	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	Chinese Restaurant	Sandwich Place	Burger Joint	Food Truck	Food Court	Italian Restaurant	BBQ Joint	Mexican Restaurant	Restaurant
1	Carnegie Hill	Café	Pizza Place	Italian Restaurant	French Restaurant	Bakery	Bagel Shop	Mediterranean Restaurant	Ramen Restaurant	Restaurant	Kosher Restaurant
2	Central Harlem	African Restaurant	Fried Chicken Joint	French Restaurant	Chinese Restaurant	American Restaurant	Seafood Restaurant	Cafeteria	Caribbean Restaurant	Burger Joint	Pizza Place
3	Chelsea	American Restaurant	Seafood Restaurant	Japanese Restaurant	Italian Restaurant	Bakery	Chinese Restaurant	Israeli Restaurant	Middle Eastern Restaurant	Mediterranean Restaurant	Sandwich Place
4	Chinatown	Chinese Restaurant	Hotpot Restaurant	American Restaurant	Noodle House	Asian Restaurant	Greek Restaurant	Malay Restaurant	Sandwich Place	Pizza Place	Spanish Restaurant

Fig.7 The first 5 rows of this dataframe containing the top 10 most dominant/common food outlets present in each neighborhood based on their mean frequency

4. Predictive Modeling:

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 as shown in Fig.8. We select the optimum value of K using elbow rule where optimum K=4 in the case of Manhattan data.

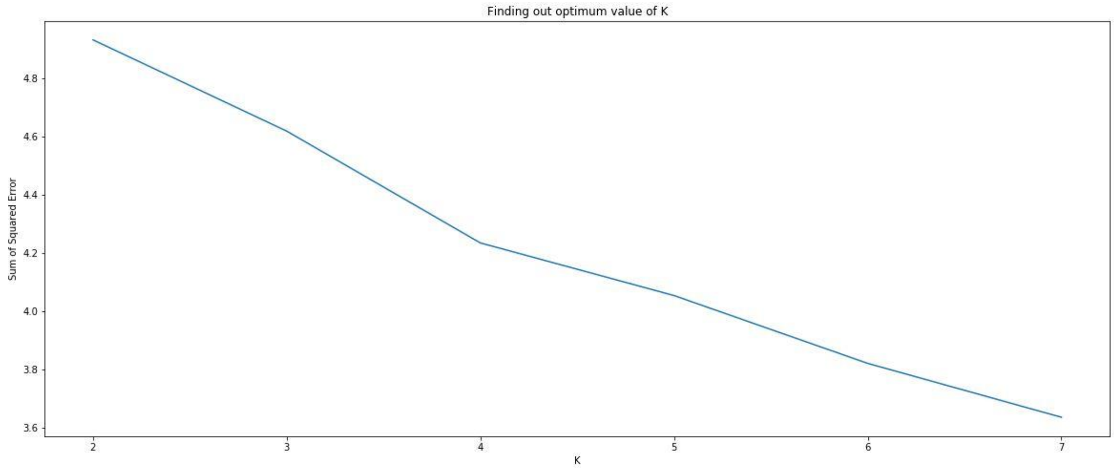

Fig.8 Plot showing sum of squared error for different values of K to determine the optimum value of K.

The same procedure mentioned above is performed for the remaining four boroughs of NY city (Brooklyn, Bronx, Queens, Staten Island). After clustering, the dataframe that contains the top 10 common food outlet types with cluster labels assigned to them will be as shown below:

[94]:

	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	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	Manhattan	Marble Hill	40.876551	-73.910660	0	American Restaurant	Sandwich Place	Deli / Bodega	Chinese Restaurant	Fast Food Restaurant	Pizza Place	Diner	Donut Shop	Bakery	Steakhouse
1	Manhattan	Chinatown	40.715618	-73.994279	0	Chinese Restaurant	Hotpot Restaurant	American Restaurant	Noodle House	Asian Restaurant	Greek Restaurant	Malay Restaurant	Sandwich Place	Pizza Place	Spanish Restaurant
2	Manhattan	Washington Heights	40.851903	-73.936900	0	Deli / Bodega	New American Restaurant	Chinese Restaurant	Tapas Restaurant	Café	Bakery	Breakfast Spot	Latin American Restaurant	Pizza Place	Caribbean Restaurant
3	Manhattan	Inwood	40.867684	-73.921210	0	Mexican Restaurant	Restaurant	Café	Deli / Bodega	Spanish Restaurant	Chinese Restaurant	American Restaurant	Bakery	Latin American Restaurant	Seafood Restaurant
4	Manhattan	Hamilton Heights	40.823604	-73.949688	0	Café	Mexican Restaurant	Caribbean Restaurant	Pizza Place	Bakery	Burger Joint	Sushi Restaurant	Latin American Restaurant	Seafood Restaurant	Mediterranean Restaurant

Fig.9 First 5 rows of dataframe that contains neighborhoods and the top 10 food outlets with their corresponding labels

5.Results and Discussions

5.1 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 Fig.10

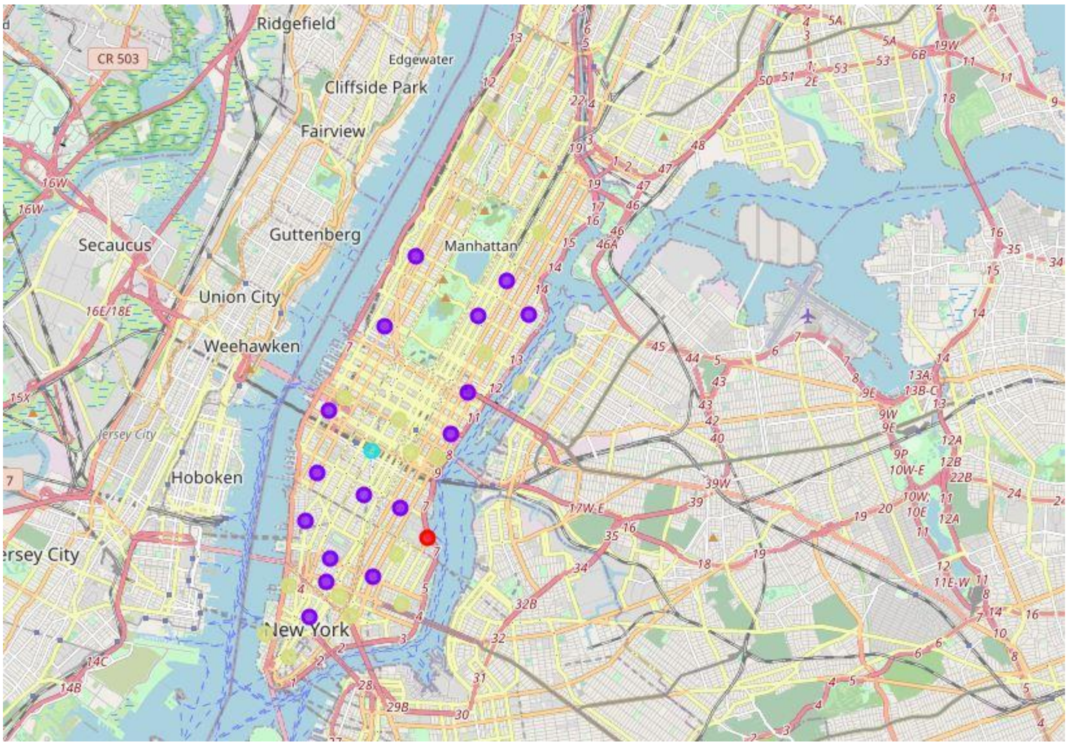

Fig.10 Map showing different clusters of neighborhood present in Manhattan borough

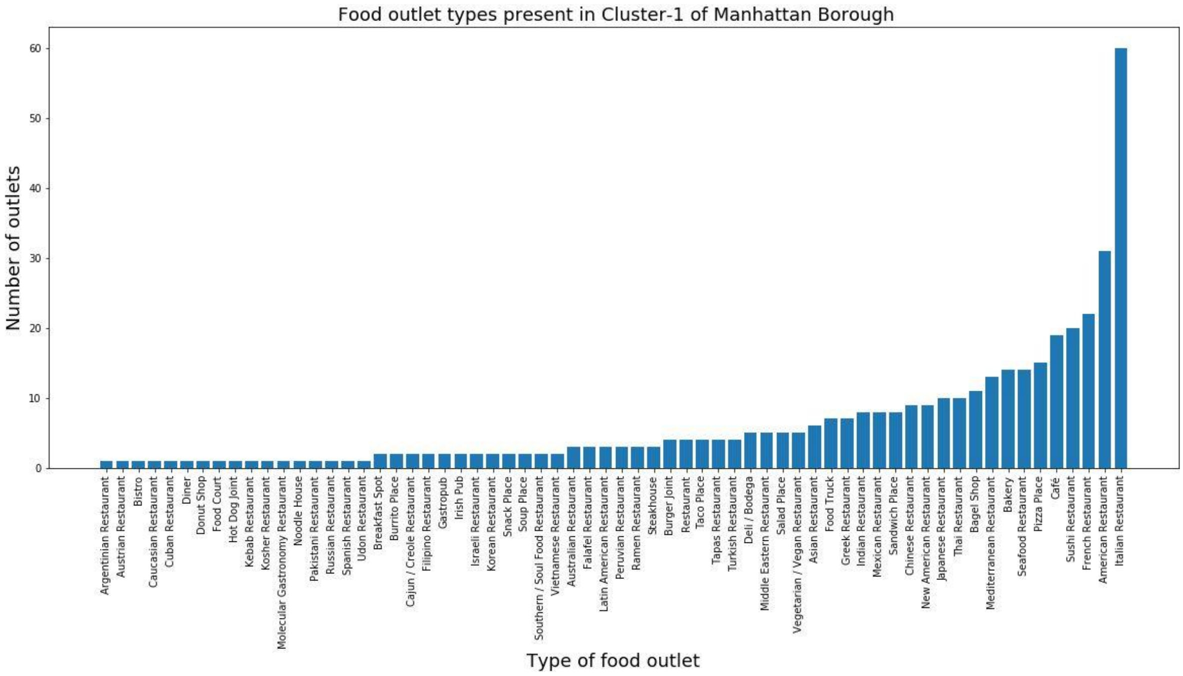

Fig.11 Plot showing Type of food outlets vs Number of outlets in Cluster-1 of Manhattan

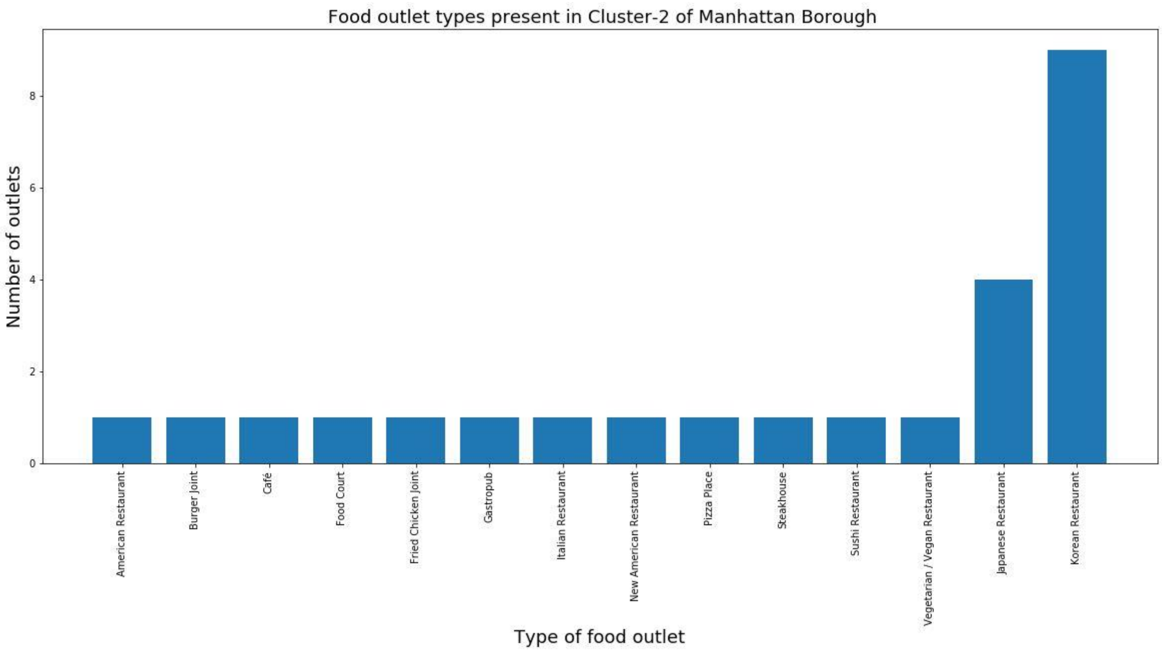

Fig.12 Plot showing Type of food outlets vs Number of outlets in Cluster-2 of Manhattan

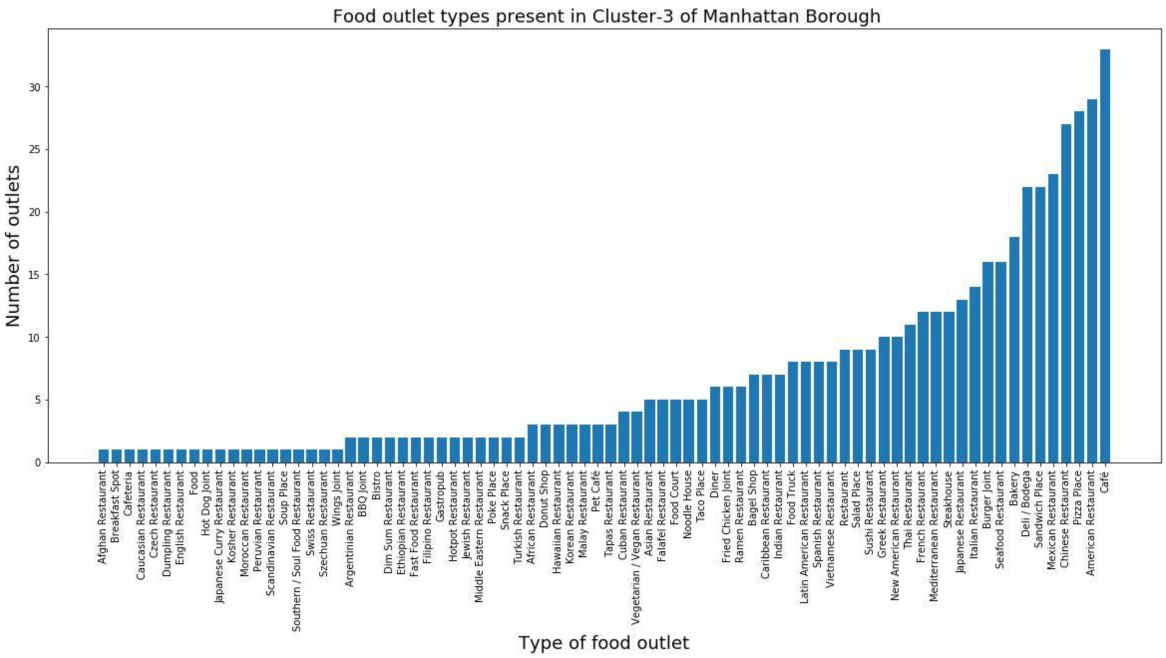

Fig.13 Plot showing Type of food outlets vs Number of outlets in Cluster-3 of Manhattan

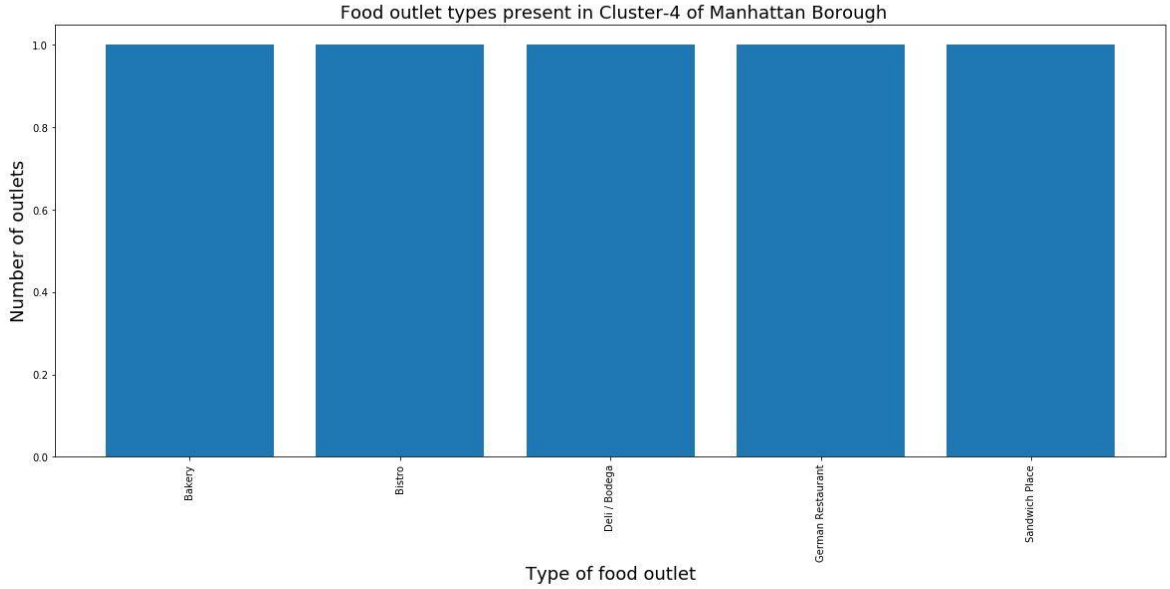

Fig.14 Plot showing Type of food outlets vs Number of outlets in Cluster-4 of Manhattan

The summarized results of K-means clustering on Manhattan borough are shown in Table.1 We can observe that Italian Restaurant, American Restaurant, French Restaurant, Sushi restaurant, Café 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 Italian, American, Sushi, or French food outlets or a Cafe.

Borough	Cluster	Neighbourhoods	Top 5 dominant/common type of restaurants
Manhattan	Cluster-1	Upper East Side, Yorkville, Upper West Side, Lincoln Square, Clinton, Greenwich Village, Chelsea, Soho, West Village, Gramercy, Carnegie Hill, Noho, Civic Center, Sutton Place, Turtle Bay, Flatiron, Hudson Yards	1.Italian Restaurant 2.American Restaurant 3.French Restaurant 4.Sushi Restaurant 5.Cafe
	Cluster-2	Midtown South	1.Korean Restaurant 2.Japanese Restaurant 2.Vegeterian Restaurant 4.Sushi Restaurant 5. Steak house
	Cluster-3	Marble Hill, Chinatown, Washington Heights, Inwood, Hamilton Heights, Manhattanville, Central Harlem, East Harlem, Roosevelt Island East Village, Lower East Side, Little Italy, Financial District, Manhattan Valley, Morningsi de Heights, Tribeca, Battery Park City, Lenox Hi ll, Midtown, Murray Hill, Tudor City	1.Cafe 2.American Restaurant 3.Pizza place 4.Mexican Restaurant 5.Chinese Restaurant
	Cluster-4	Stuyvesant Town	1.Sandwitch place 2.German Restaurant 3.Deli/Bodega 4.Bistro 5.Bakery

Table 1. Details of different clusters present in the Manhattan borough.

We performed similar analysis on the remaining four boroughs of NY city. The summarized results of K-means clustering on five different boroughs of New York city are as follows:

Borough	Cluster	Neighbourhoods	Top 5 dominant/common type of restaurants
Brooklyn	Cluster-1	Bensonhurst, Sunset Park, Gravesend Brighton Beach, Sheepshead Bay Manhattan Terrace, Flatbush, Kensington Windsor Terrace, Brownsville, Cypress Hills Starrett City, Bath Beach, Downtown City Line, Georgetown, Ocean Parkway Fort Hamilton, Ditmas Park, Wingate Rugby, Remsen Village, Mill Basin, Weeksville, Broadway Junction, Homecrest Erasmus	1.Chinese Restaurant 2.Pizza place 3.Deli/Bodega 4.Donut Shop 5.Bakery
	Cluster-2	Crown Heights, Bedford Stuyvesant, East New York, Flatlands, Coney Island Borough Park, Gerritsen Beach, Marine Park Ocean Hill, Midwood Prospect, Park South New Lots, Highland Park, Madison	1.Deli/Bodega 2.Pizza place 3.Fried chicken joint 4.Chinese Restaurant 5.Caribbean Restaurant
	Cluster-3	Bay Ridge, Greenpoint, Prospect Heights Williamsburg , Bushwick, Brooklyn Heights Cobble Hill, Carroll Gardens, Red Hook Gowanus, Fort Greene, Park Slope Manhattan Beach, Dyker Heights Clinton Hill, Boerum Hill, Prospect Lefferts Gar dens, Bergen Beach, East Williamsburg North Side, South Side, Fulton Ferry Vinegar Hill, Dumbo	1.Pizza Place 2.Italian Restaurant 3.Bakery 4.Cafe 5.American Restauarant
	Cluster-4	East Flatbush, Canarsie, Paerdegat Basin	1.General Food 2.Deli/Bodega 3.Chinese Restaurant 4.Caribbean Restaurant 5.Asian Restaurant

Table 2. Details of different clusters present in the Brooklyn borough

Borough	Cluster	Neighbourhoods	Top 5 dominant/common type of restaurants
Staten Island	Cluster-1	Butler Manor	1.BBQ Joint
	Cluster-2	New Brighton, Grymes Hill, South Beach Mariner's Harbor, Arden Heights	1.Deli/Bodega 2.Pizza place 3.Italian Restaurant 4.Chinese Restaurant

			5.American Restaurant
	Cluster-3	St. George, Stapleton, Rosebank, West Brighton Port Richmond, Castleton Corners, New Springville, Travis, New Dorp, Great Kills Eltingville, Annadale, Woodrow, Tompkinsville, Silver Lake, Sunnyside Westerleigh, Graniteville, Arlington Arrochar, Grasmere, Old Town, Dongan Hills Midland Beach, Grant City, Huguenot Pleasant Plains, Charleston, Rossville, Greenridge, Heartland Village, Bulls Head Clifton, Concord, Emerson Hill, Randall Manor Elm Park, Manor Heights, Willowbrook Sandy Ground, Prince's Bay, Richmond Valley Fox Hills	1.Pizza Place 2.Deli/Bodega 3. Italian Restaurant 4.Bagel shop 5. Chinese Restaurant
	Cluster-4	Tottenville, New Dorp Beach, Bay Terrace Chelsea, Richmond Town, Shore Acres Howland Hook, Egbertville, Lighthouse Hill	1. Italian Restaurant 2.Deli/Bodega 3.Cafe 4.Bagel shop 5.Sandwich place

Table 3. Details of different clusters present in the Staten Island borough

Borough	Cluster	Neighbourhoods	Top 5 Most dominant/common type of restaurants
Bronx	Cluster-1	Williamsbridge	1.Caribbean Restaurant 2. Soup place 3.Fast Food place
	Cluster-2	Wakefield, Eastchester, Woodlawn, Norwood, Pelham Parkway, City Island, Bedford Park, West Farms, Melrose, Longwood Clason Point, Throgs Neck, Country Club Van Nest, Morris Park, Belmont, Pelham Bay Schuylerville, Edgewater Park, Castle Hill Concourse, Concourse Village, Mount Hope Bronxdale, Allerton	1.Deli/Bodega 2.Pizza place 3.Chinese Restaurant 4. Italian Restaurant 5. Sandwich place
	Cluster-3	Riverdale	1.Food Truck
	Cluster-4	Co-op City, Kingsbridge, Baychester University Heights, Morris Heights Fordham, East Tremont, High Bridge Mott Haven, Port Morris, Hunts Point Morrisania, Soundview, Parkchester Westchester Square, Spuyten Duyvil North Riverdale, Olinville Pelham Gardens, Unionport, Edenwald Claremont Village, Mount Eden, Kingsbridge Heights	1.Pizza place 2.Chinese Restaurant 3.Common Food 4.Spanish Restaurant 5.Deli/Bodega

Table 4. Details of different clusters present in the Bronx borough

<table>
<tr><td>Borough</td><td>Cluster</td><td>Neighbourhoods</td><td>Top 5 Most dominant/common type of restaurants</td></tr>
<tr><td rowspan="3">Queens</td><td>Cluster-1</td><td>Astoria, Woodside, Jackson Heights, Elmhurst
Howard Beach, Corona, Forest Hills,
Kew Gardens, Richmond Hill, Flushing
Long Island City, Sunnyside, East Elmhurst
Maspeth, Ridgewood, Glendale, Rego Park
Woodhaven, Ozone Park, South Ozone Park
College Point, Bayside, Auburndale
Little Neck, Douglaston, Glen Oaks
Bellerose, Kew Gardens Hills, Fresh Meadows
Briarwood, Jamaica Center, Oakland Gardens
Queens Village, Hollis, South Jamaica
St. Albans, Rochdale, Springfield Gardens
Cambria Heights, Rosedale, Far Rockaway
Steinway, Beechhurst, Bay Terrace
Edgemere, Arverne, Rockaway Beach
Murray Hill, Holliswood, Queensboro Hill
Hillcrest, Ravenswood, Lindenwood
Laurelton, Lefrak City, Belle Harbor
Rockaway Park, Bellaire, North Corona
Forest Hills Gardens, Jamaica Hills, Utopia
Pomonok, Astoria Heights, Hunters Point
Sunnyside Gardens, Blissville, Roxbury
Middle Village, Malba, Hammels, Queensbridge</td><td>1.Deli/Bodega
2.Chinese Restaurant
3.Pizza place
4.Bakery
5.Donut Shop</td></tr>
<tr><td>Cluster-2</td><td>Floral Park, Jamaica Estates</td><td>1.Indian Restaurant
2.Pizza place
3.Dosa Place
4. Chinese Restaurant</td></tr>
<tr><td>Cluster-3</td><td>Whitestone, Broad Channel, Brookville</td><td>1.Deli/Bodega
2.Sandwich Place
3.Pizza Place</td></tr>
</table>

Table 4. Details of different clusters present in the Queens borough

6. Conclusions

In this project, we predicted the dominant/common type of food outlets present in different parts of New York city with help of location data of various types of food outlets in New York city. Using the latitudes and longitudes of neighborhoods in NY city, we obtained the location data of different types of food outlets with the help of Foursquare API.

Neighborhoods of each borough in NY 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 NY city. It can help them in deciding the type of food outlet that provides them the competitive advantage.