

## 1) Introduction & Business Problem:

### Problem Background:

New York is one of the most famous cities in the world and it is also the financial capital of USA. It is multicultural and it provides lot of business opportunities and business friendly environment. The city is a major center for several sectors like banking, finance, tourism.

However, the cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully.

After the analysis an insight will be obtained with regards to the business environment that will help in the process for targeting the market.

The aim is to reduce the risk and to increase the potential return on the investment.

### Problem Description:

During the pandemic there is an increased demand for food delivery and due to New York's immigrant history there is a wide availability and need of different cuisines.

For example there were several Central Europeans, Italians, Jewish, Chinese immigrants in the city and hence the below needs can be identified respectively for each immigrants group:

1. Bagels, cheesecake, hot dogs, knishes, and delicatessens
2. Pizza and Italian cuisine
3. Pastrami and corned beef
4. Sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city

Also, people are more interested in higher quality of ingredients like organic food and locally produced products.

Various factors need to be studied in order to decide on the Location of a potential restaurant offering also delivery services such as:

1. New York Population
2. New York City Demographics
3. Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost?
4. Are there any venues like Gyms, Entertainment zones, Parks etc nearby where floating population is high etc
5. Who are the competitors in that location?
6. Cuisine served / Menu of the competitors
7. Segmentation of the Borough

## 2) Data :

**New York City** will be analysed.

The below datasets will be used for analysing New York city

**Data 1 :** New York has several boroughs which are further divided into several neighbourhoods. In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighborhood.

The above mentioned can be obtained from the below link:

[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

**Data 2 :** Second data which will be used is the DOHMH Farmers Markets. In this we will be using the data of Farmers Markets.

A **farmers' market** is a public site used several local producers for the direct sale of farm products to consumers.

<https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2>

Website-<https://www.grownyc.org/greenmarketco/foodbox>

**Data 3 :** For the below analysis we will get data from wikipedia as given below :

1. New York Population
2. New York City Demographics
3. Cuisine of New York city

[https://en.wikipedia.org/wiki/New\\_York\\_City](https://en.wikipedia.org/wiki/New_York_City)

[https://en.wikipedia.org/wiki/Economy\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Economy_of_New_York_City)

[https://en.wikipedia.org/wiki/Portal:New\\_York\\_City](https://en.wikipedia.org/wiki/Portal:New_York_City)

[https://en.wikipedia.org/wiki/Cuisine\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Cuisine_of_New_York_City)

**Data 4 :** New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighbourhood. We will use the Foursquare API to explore neighbourhoods in New York City.

### 3) Methodology:

**Business Understanding:** Our main goal is to get optimum location for new delivery offering restaurant business in New York City.

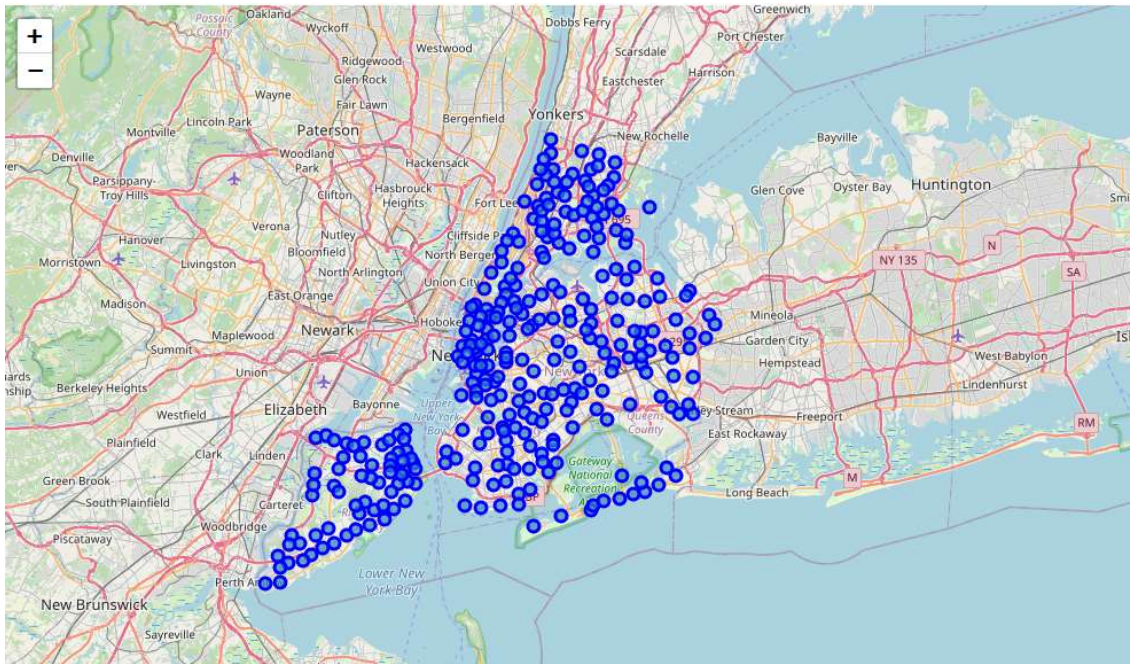
**Analytic Approach:** New York city is divided into 5 boroughs and several neighbourhoods. In this project first part is clustering of Manhattan and Brooklyn . And second part is clustering of Bronx, Queens and Staten Island. This is done because of the following Exploratory data analysis.

Exploratory Data Analysis :

**Data 1-** New york city Geographical Coordinates Data.

- 1.In this we load the data and explore data from newyork\_data.json file.
- 2.Transform the data of nested python dictionaries into a pandas dataframe.
- 3.This dataframe contains the geographical coordinates of New York city neighbourhoods.
- 4.This data will used to get Venues data from Foursquare.
- 5.We used geopy and folium libraries to create a map of New York city with neighbourhoods superimposed on top.

## New York neighbourhood visualization

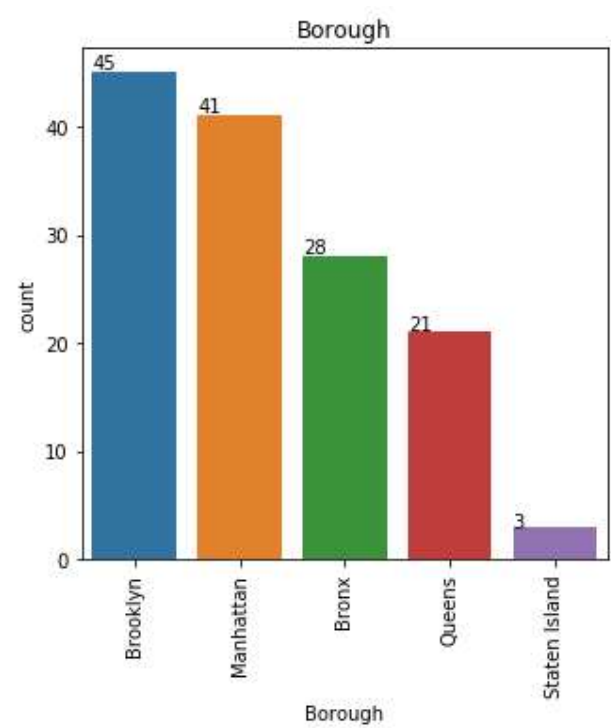


**Data 2-** Second data which is used is the DOHMH Farmers Markets dataset. In this we will be using the data of Farmers Markets data.

There are totally 138 Farmers Markets in New York city. Highest number are in Brooklyn(45) and Manhattan(41).

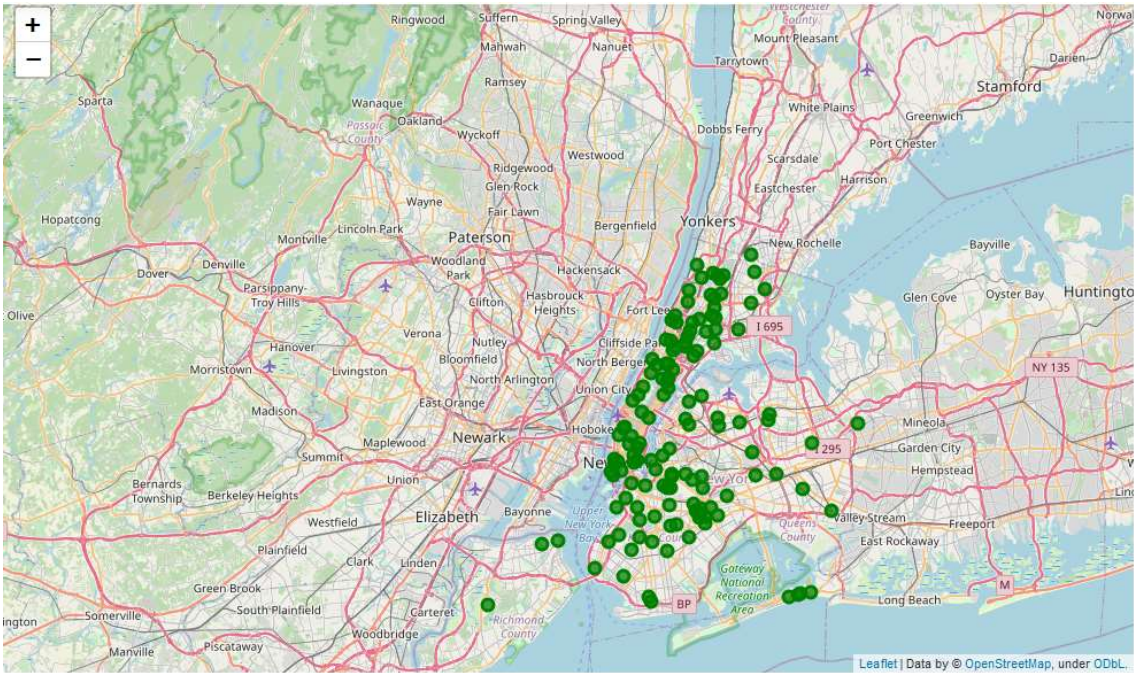
And lowest in Bronx(28), Queens(21), and Staten Island(3).

The allocation of the farmer markets are set in the below histogram:



Geopy and folium libraries were used to create a map to visualise farmers markets of New York city.

**Farmers Market visualisation-New York City**





**Data 3 :** To analyze New York city Population, Demographics and Cuisine, data from Wikipedia pages given above in the data section were used. BeautifulSoup python library was used in some occasions and in some other the data was gathered manually.

#### 1.New York Population : Insights from the data:

	Borough	County	Estimate_2019	GrossDomesticProduct	GDP/capita	square miles	square km	squarekm	persons/mi2	persons/km2
0	The Bronx	Bronx	1,418,207	42,695	30,100	42.10	109.04	NaN	33867	13006
1	Brooklyn	Kings	2,559,903	91,559	35,800	70.82	183.42	NaN	36147	13957
2	Manhattan	New York	1,628,706	600,244	368,500	22.83	59.13	NaN	71341	27544
3	Queens	Queens	2,253,858	93,310	41,400	108.53	281.09	NaN	20767	8018
4	Staten Island	Richmond	476,143	14,514	30,500	58.37	151.18	NaN	8157	3150
5	City of New York		8,336,817	842,343	101,000	302.64	783.83	NaN	27547	10636
6	State of New York		19,453,561	1,731,910	89,000	47,126.40	122,056.82	NaN	412	159

- Manhattan (New York County) is the geographically smallest but the borough with highest population density.
- Manhattan's (New York County's) population density is 71,341 people per square mile (27,544/km<sup>2</sup>) in 2019.
- Brooklyn is the New York's most populous borough with 2,56m people.
- Queens is geographically the largest borough.

#### 2.New York City Demographics: New York City has 8,3m residents as of 2019.

The racial composition of the above population is set below. For this reason there is a wide variety of restaurants offering different cuisines (Italian, Greek, Indian etc).

Racialcomposition	2010	1990	1970	1940
White	0.440	0.523	0.766	0.936
—Non-Hispanic	0.333	0.432	0.629	0.920
Black or African American	0.255	0.287	0.211	0.061
Hispanic or Latino (of any race)	0.286	0.244	0.162	0.016
Asian	0.127	0.070	0.012	0.000

**3.Cuisine of New York city:** This data has been manually prepared. Data is taken from Wikipedia page - [https://en.wikipedia.org/wiki/Cuisine\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Cuisine_of_New_York_City) .

Using this data word cloud was prepared so as to find the most popular cuisine per area.



QUEENS CUISINE - Most Preferred Food in Queens is – Indian, Italian, Pakistani.



<Figure size 2160x3240 with 0 Axes>

THE BRONX CUISINE - Most Preferred Food in The Bronx is – Italian, Puerto Rican, Dominican and Albanian.



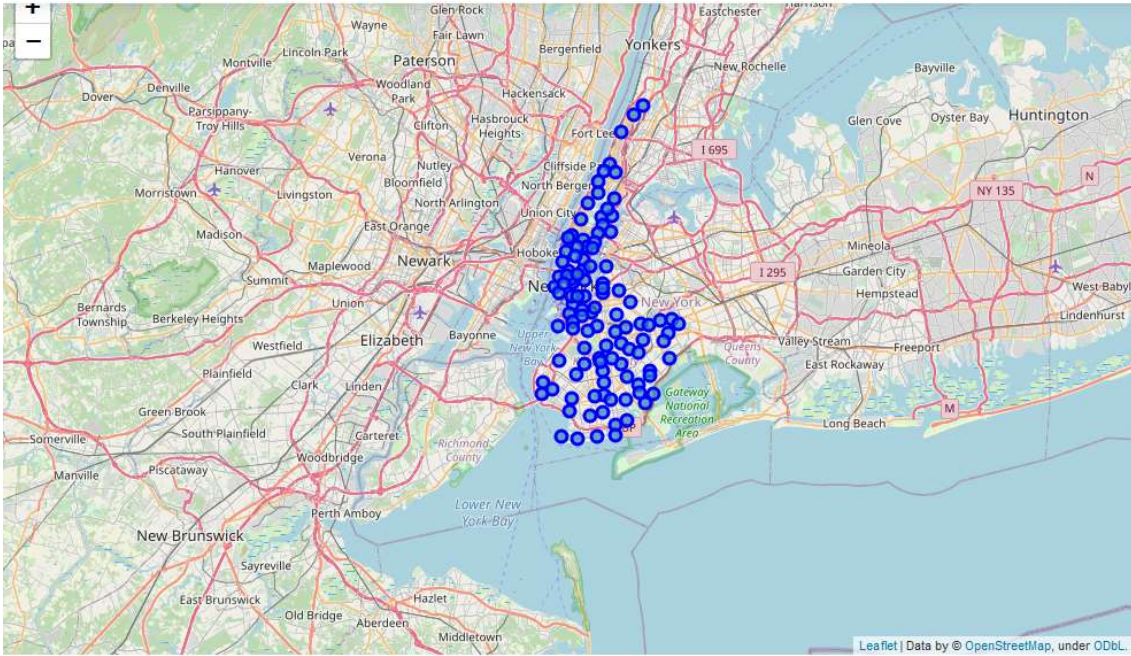
<Figure size 2160x3240 with 0 Axes>

There were no enough data for Staten Island so the same analysis could not have been developed.

**Data 4:** New York city geographical coordinates data has be utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighbourhood. We used the Foursquare API data to explore neighbourhoods in New York City.

**Brooklyn and Manhattan:**

Brooklyn and Manhattan Visualization:



Using the geographical coordinates of each neighbourhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters.

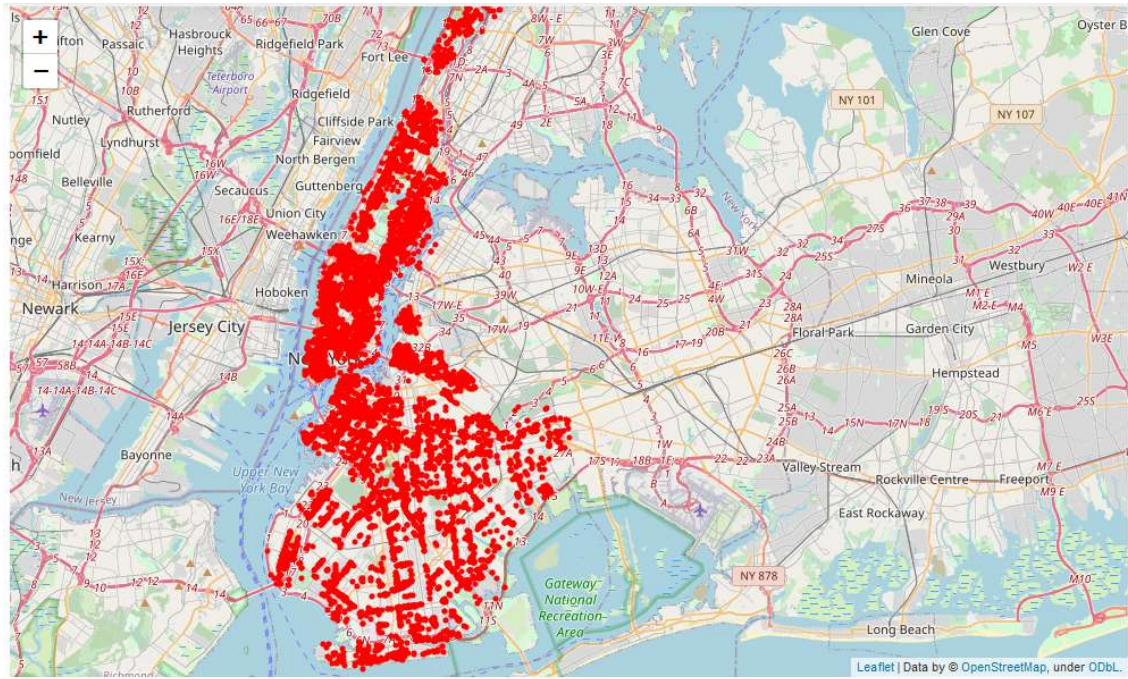
The venues data is as given below:

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop



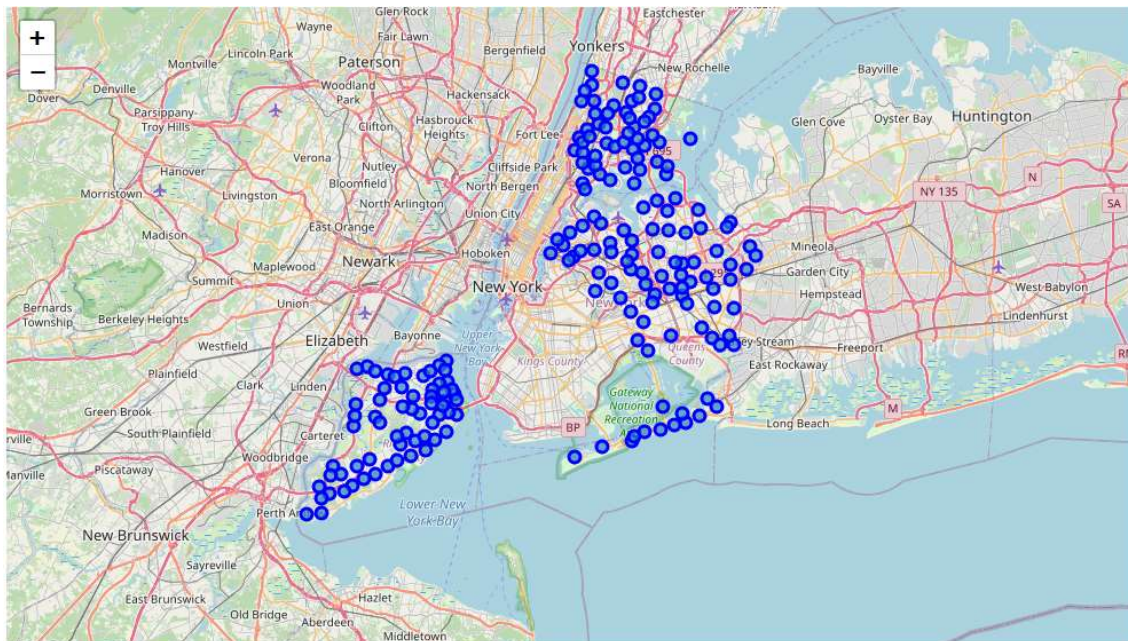
### Brooklyn and Manhattan Venues:

Brooklyn and Manhattan Venues Visualization : Generated the below Brooklyn and Manhattan Venues Visualization. The "BM\_venues" dataframe has 9581 venues and 409 unique venue types.



### Bronx, Queens and Staten Island:

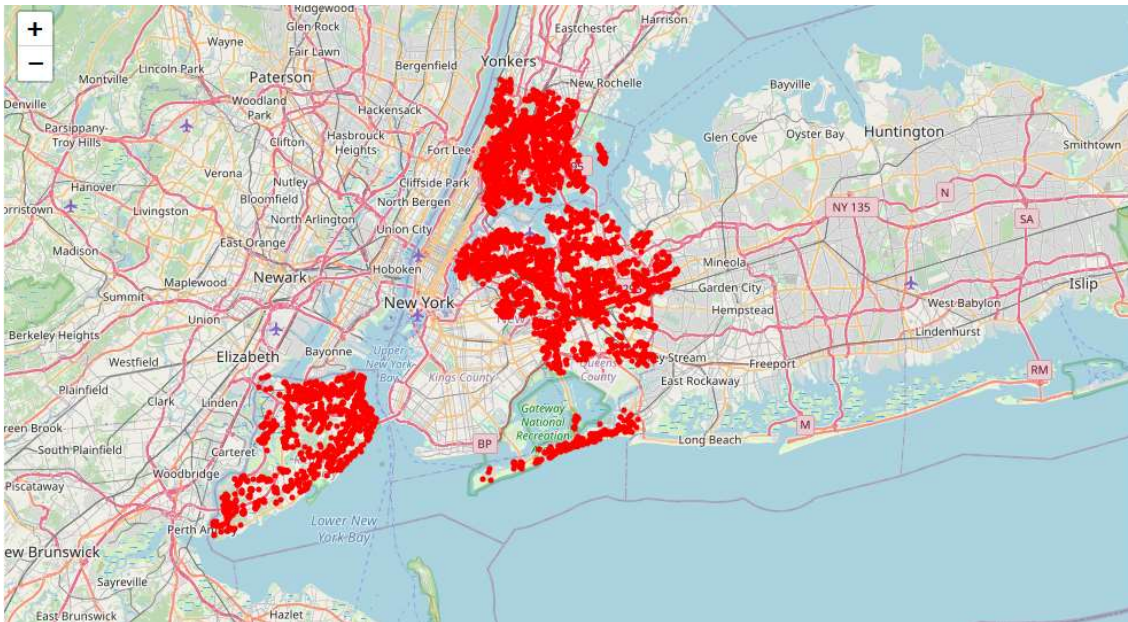
#### Bronx, Queens and Staten Island Neighborhoods Visualization:



**Bronx, Queens and Staten Island Venues Visualization :** The "BQS\_venues" dataframe has 10899 venues and 394 unique venue types.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy

**Bronx, Queens and Staten Island Venues Map Visualization:**



#### 4.RESULTS:

From this venues data we filtered and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens and Staten Island clustering. As we focussed only on restaurants business.

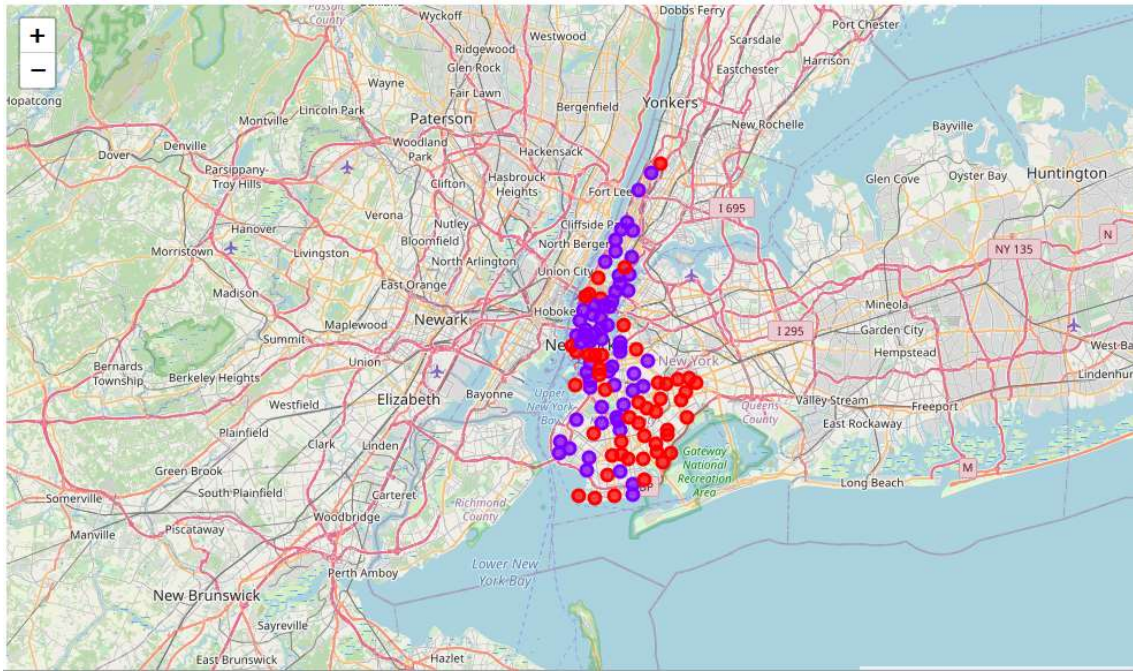
**Neighborhood K-Means clustering based on mean occurrence of venue category:**

To cluster the neighbourhoods into two clusters we used the K-Means clustering Algorithm.k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach.



### Brooklyn & Manhattan:

In the below Map Visualization, we can see the different types of clusters created by using K-Means for Brooklyn & Manhattan.



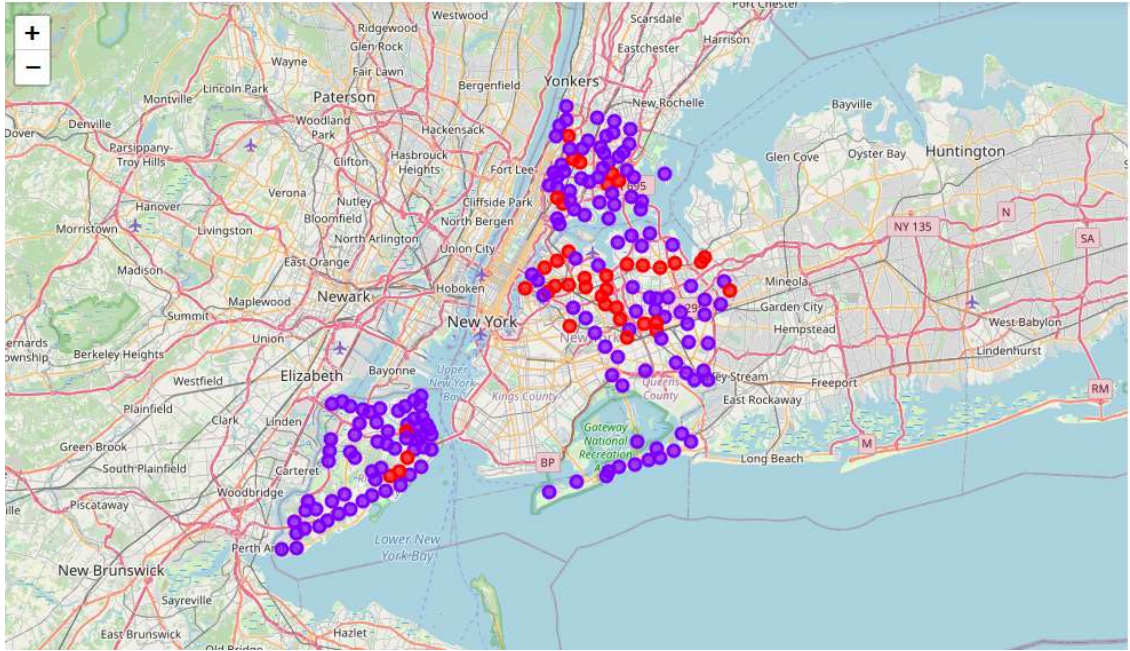
**Cluster0:** The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.

**Cluster1:** The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

There are no untapped neighbourhoods in Brooklyn and Manhattan.

### Bronx, Queens and Staten Island:

In the below Map Visualization, we can see the different types of clusters created by using K-Means for Bronx, Queens and Staten Island.



**Cluster0:** The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated. There are untapped neighbourhoods. List is as given below.

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Bronx	Clason Point	40.806551	-73.854144	0	1
1	Staten Island	Todt Hill	40.597069	-74.111329	0	1
2	Staten Island	Port Ivory	40.639683	-74.174645	0	1
3	Staten Island	Butler Manor	40.506082	-74.229504	0	1
4	Staten Island	Bloomfield	40.605779	-74.187256	0	1

**Cluster1:** The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

**5.DISCUSSION:**

- 1.There is scope to explore cuisines of various countries in Bronx, Queens and Staten Island.
- 2.In Manhattan and Brooklyn restaurants of cuisines of many countries are available. This is also an indication that people love eating cuisines of various countries.

**6.CONCLUSION:**

Brooklyn and Manhattan has high concentration of restaurant business and this also indicates a highly competitive market. The market in Bronx, Queens and Staten Island is not saturated and although there are several restaurants in the area and therefore this area could be further examined for a business opportunity.