

# The Battle of the Neighborhoods

## Introduction

The City of New York, is the one of the most famous cities in the world. It is diverse in culture. It provides lot of business oppourtunities and business friendly environment. It has attracted many different players into the market. The city is a major center for banking and finance, retailing, world trade, transportation,tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theater, fashion, and the arts in the United States. Apart from that, the city is famous for the myriad of cuisines it offers.

They say 'if you make it here, you can make it anywhere'. Our aim for this project would be to predict what borough of New York must one choose if they were to open their own restaurant.

## Problem

The city of New York offers a variety of cuisines and restaurants. Hence, in order to survive in such a competitive market it is very important to startegically plan. Various factors need to be taken into consideration in order to decide on the location such as :

1. Population
2. Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased?
3. Is there any landmark nearby i.e. Timesquare, Central Park
4. Competitors
5. Cuisines
6. Segmentation of the Borough
7. Unexplored/Sparse areas

## Data

We will be using Dataset the below datasets for analysing NYC.

- Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement the neighborhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighborhoods that exist in each borough as well as the the latitude and logitude coordinates of each neighborhood. This dataset exists for free on the web. Link to the dataset is : [https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)
- Second data which will be used is the DOHMH Farmers Markets and Food Boxes dataset. In this we will be using the data of Farmers Markets. <https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2> Website-<https://www.grownyc.org/greenmarketco/foodbox>
- New York Population New York City Demographics 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)
- Foursquare API will also be leveraged

## **2.Methodology :**

### **Business Understanding :**

Our main goal is to get optimum location for new restaurant business in New York City for XYZ Company.

## Analytic Approach :

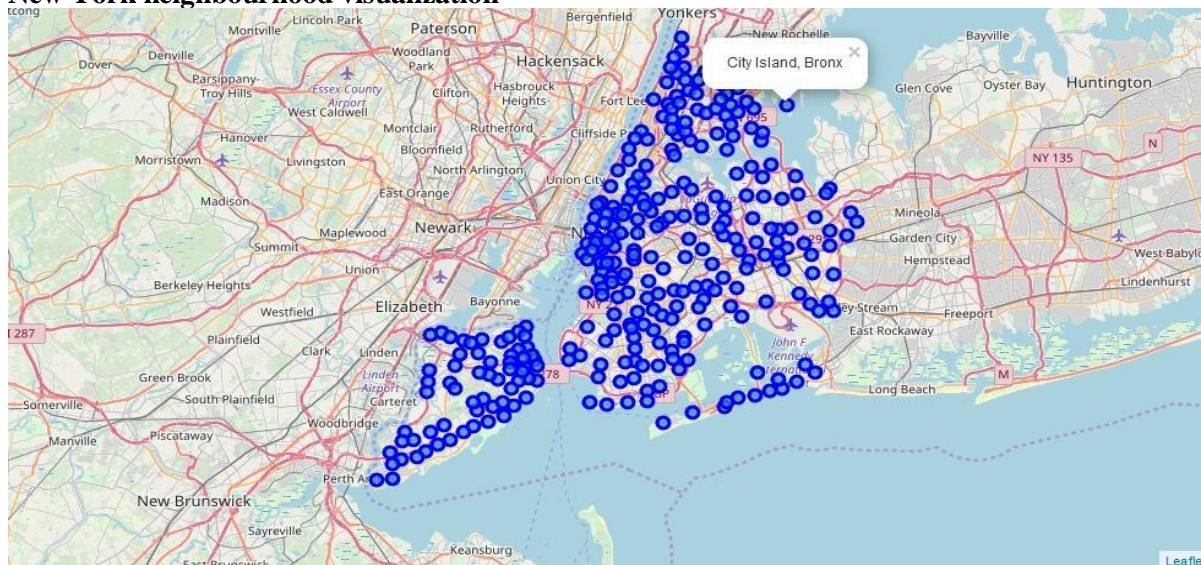
New York city neighbourhood has a total of 5 boroughs and 306 neighborhoods. In this project first part is clustering of Manhattan and Brooklyn . And second part is clustering of Bronx, Queens and Staten Island. This is done because of the following Exploratory data analysis.

## Exploratory Data Analysis :

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

1. In this we load the from **newyork\_data.json** file.
2. Convert the data to a pandas dataframe.
3. This dataframe contains the geographical coordinates of New York city neighborhoods.
4. This data will used to get Venues data from Foursquare.
5. Using folium library to create a map of New York city with neighborhoods superimposed on top.

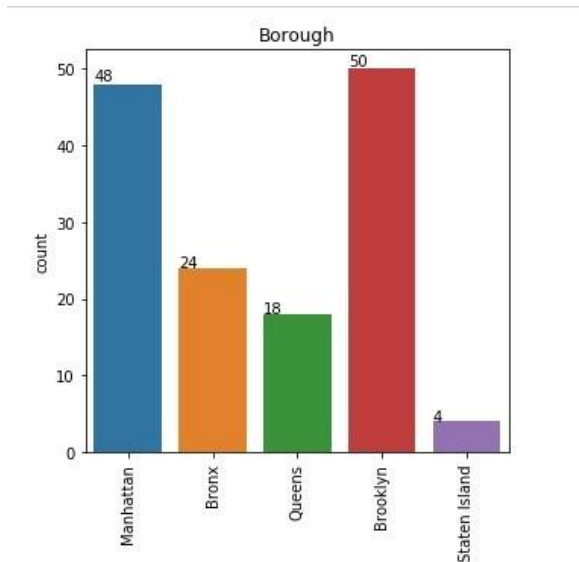
### New York neighbourhood visualization



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

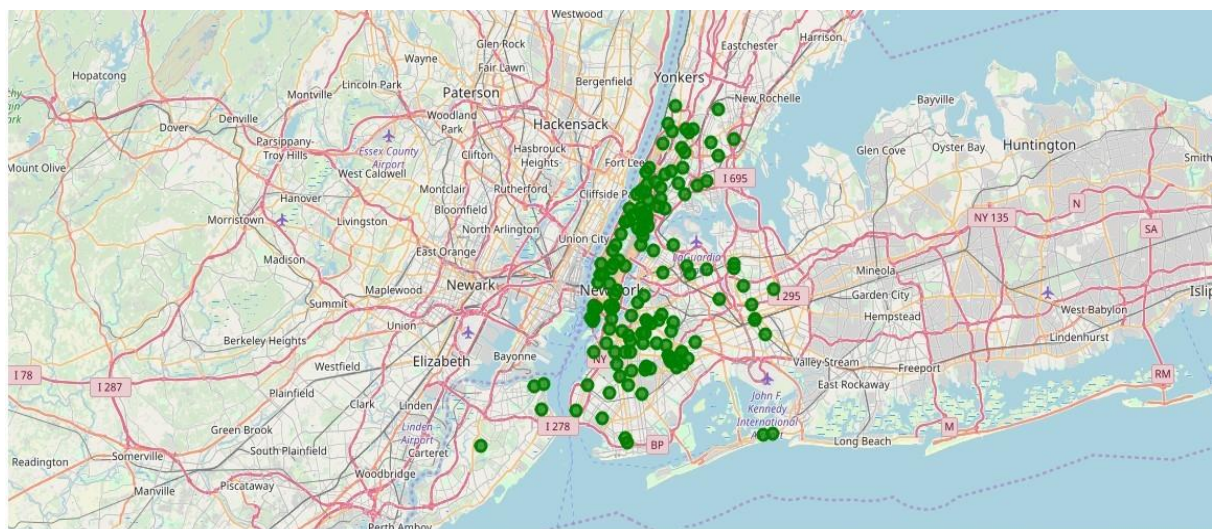
There are totally 144 Farmers Markets in New York city. Highest number are in Manhattan and Brooklyn.  
And lowest in Queens, Bronx and Staten Island.

The proof of this is as given below.



We used geopy and folium libraries 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 , scrapped the data from Wikipedia pages given above in the data section. We used BeautifulSoup python library. BeautifulSoup is a Python package for parsing HTML and XML documents (including having malformed markup, i.e. non-closed tags, so named after tag soup). It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping

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

- Manhattan (New York County) is the geographically smallest and most densely populated borough.
- Manhattan's (New York County's) population density of 72,033 people per square mile (27,812/km<sup>2</sup>) in 2015 makes it the highest of any county in the United States and higher than the density of any individual American city.



- Brooklyn (Kings County), on the western tip of Long Island, is the city's most populous borough.
- Queens (Queens County), on Long Island north and east of Brooklyn, is geographically the largest borough.

	Borough	County	Estimate_2017	square_miles	square_km	persons_sq_mi	persons_sq_km
0	Manhattan	New York	1,664,727	22.83	59.13	72,033	27,826
1	The Bronx	Bronx	1,471,160	42.10	109.04	34,653	13,231
2	Brooklyn	Kings	2,648,771	70.82	183.42	37,137	14,649
3	Queens	Queens	2,358,582	108.53	281.09	21,460	8,354
4	Staten Island	Richmond	479,458	58.37	151.18	8,112	3,132
5		City of New York	8,622,698	302.64	783.83	28,188	10,947
6		State of New York	19,849,399	47,214	122,284	416.4	159

**2. New York City Demographics :** New York City is the most populous city in the United States,<sup>[9]</sup> with an estimated record high of 8,622,698 residents as of 2017,<sup>[7]</sup> incorporating more immigration into the city than outmigration since the 2010 United States Census.

The racial composition is as given below. This is the reason New York city has restaurants serving cuisine from many countries such as Indian, African, Japan etc. This also increases the scope for restaurants business in New York City.

	Racialcomposition	2010	1990	1970	1940
0	White	44.0%	52.3%	76.6%	93.6%
1	—Non-Hispanic	33.3%	43.2%	62.9%	92.0%
2	Black or African American	25.5%	28.7%	21.1%	6.1%
3	Hispanic or Latino (of any race)	28.6%	24.4%	16.2%	1.6%
4	Asian	12.7%	7.0%	1.2%	–

**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 we did word cloud.

**NEW YORK CITY CUISINE :** Most Preferred Food in New York City –Italian, Puerto Rican, Mexican, Jewish, Indian, Pakistani & Dominican.



**BROOKLYN CUISINE** -Most Preferred Food in Brooklyn is –Italian, Purto Rican & Mexican



**MANHATTAN CUISINE** - Most Preferred Food in Manhattan is – Italian, American, Puerto Rican and Indian.



**QUEENS CUISINE** - Most Preferred Food in Queens is – Indian, Irish, Pakistani and Mexican.



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

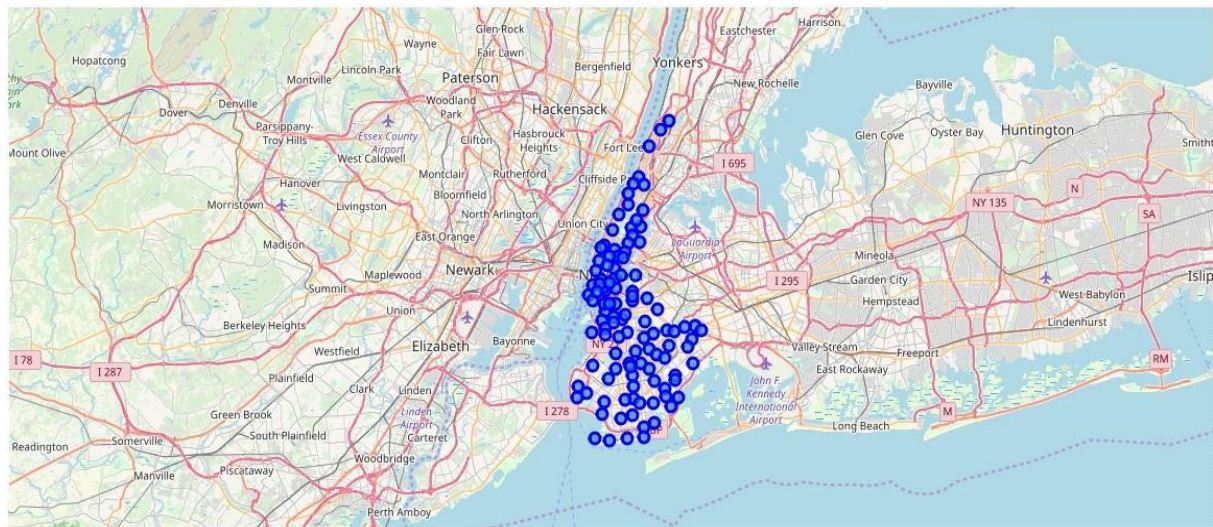


There is very less data of cuisine relating to Staten Island. So could not develop word cloud with it.

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

**Brooklyn and Manhattan :**

**Brooklyn and Manhattan Visualization :**



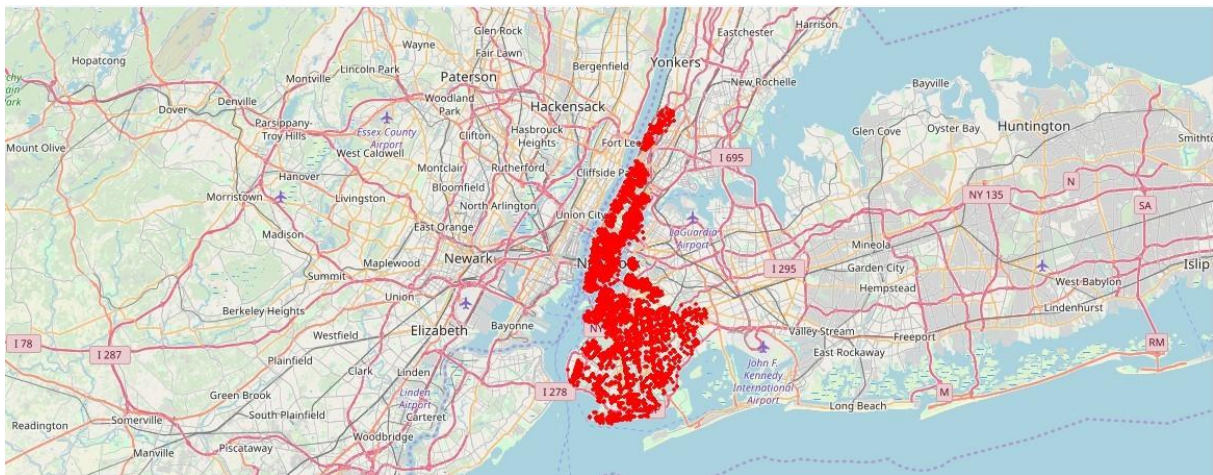
Using the geographical coordinates of each neighbourhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters. The venues data is as given below :



## Brooklyn and Manhattan Venues :

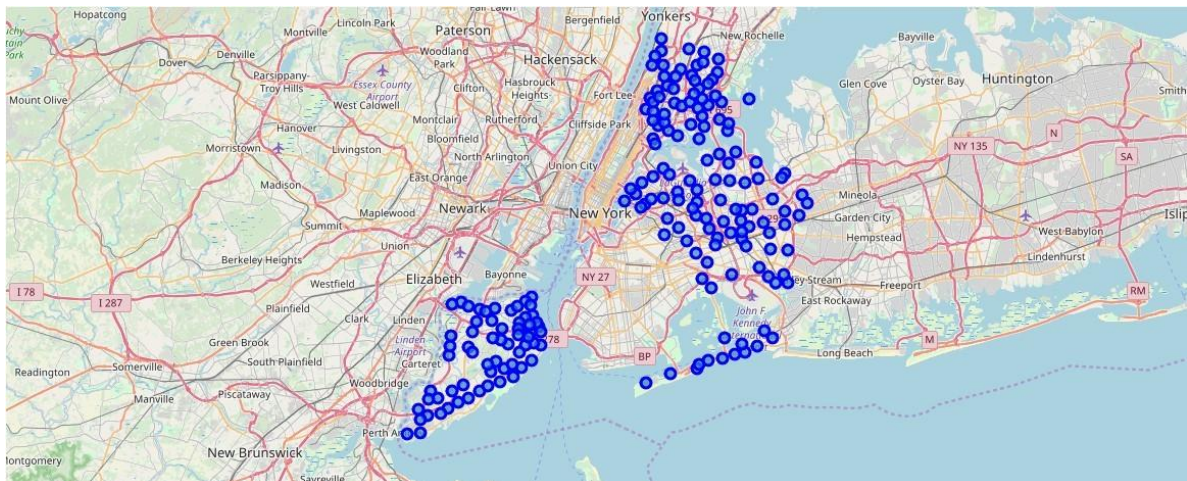
	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett 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	Loeser's Delicatessen	40.879242	-73.905471	Sandwich Place

**Brooklyn and Manhattan Venues Visualization :** Generated the below Brooklyn and Manhattan Venues Visualization. The "BM\_venues" dataframe has 9708 venues and 397 unique venue types.



## Bronx, Queens and Staten Island :

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

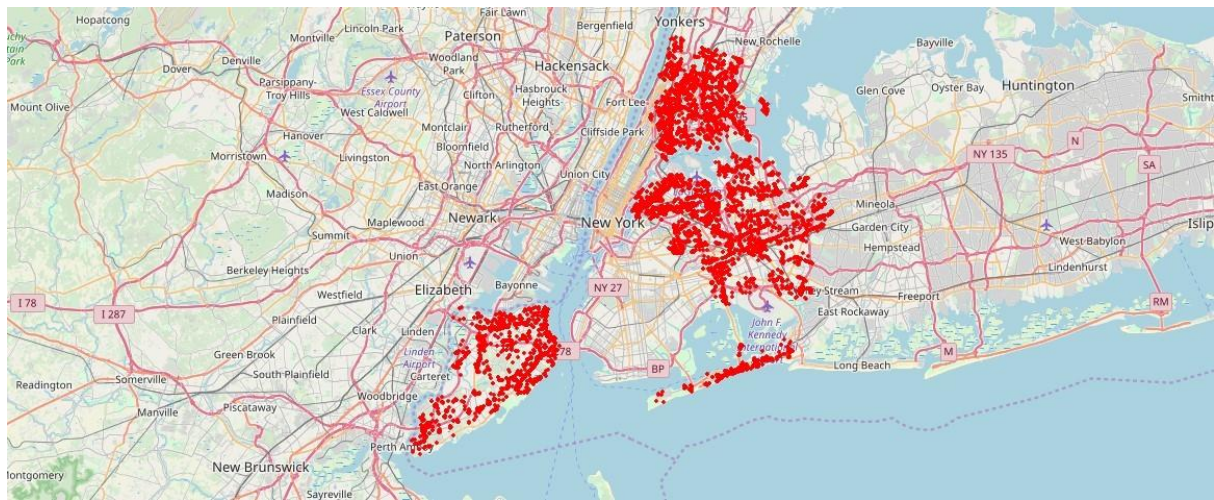




**Bronx, Queens and Staten Island Venues Visualization :** The "BQS\_venues" dataframe has 10805 venues and 387 unique venue types.

	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
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	40.896521	-73.844680	Pharmacy

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



## 4.RESULTS :

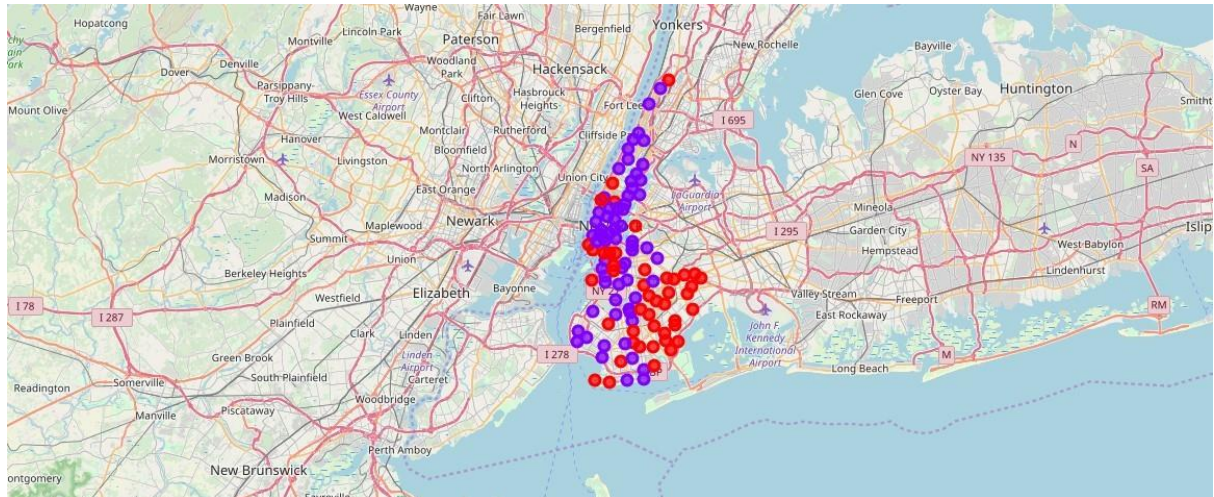
From this venues data we filtered and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens and Staten Island clustering. As we focussed only on restaurants business.

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

To cluster the neighborhoods 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.



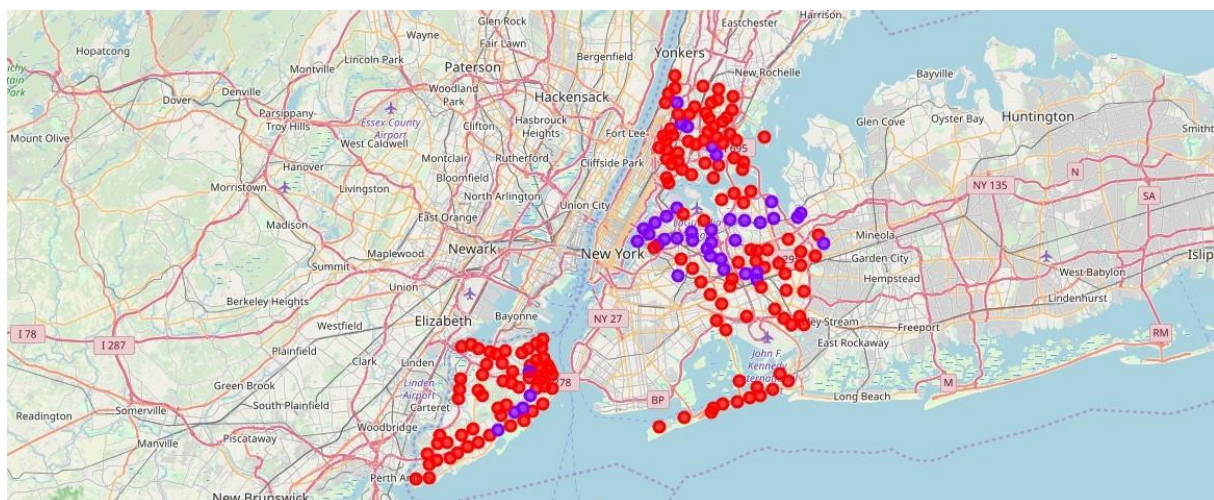
**Cluster0** : The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.

**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 neighborhoods 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 neighborhoods. List is as given below.

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Staten Island	Todt Hill	40.597069	-74.111329	0	0
1	Staten Island	Port Ivory	40.639683	-74.174645	0	0
2	Staten Island	Bloomfield	40.605779	-74.187256	0	0

**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 increase Farmers markets in Bronx, Queens and Staten Island.
2. There is scope to explore cuisines of various countries in Bronx, Queens and Staten Island.
3. In Manhattan and Brooklyn restaurants of cuisines of many countries are available. So if risk can be taken with great menu on board. It also shows people love eating cuisines of various countries.

## **6.CONCLUSION:**

This project is performed using limited data and hence may not fetch optimum results. But if good amount of data is available, better results can be attained. Bronx, Queens and Staten Island have fewer number of restaurants and hence it is a good market for exploration. Brooklyn and Manhattan have many restaurants, which can be a good market if one wants to get into the peak market. The cuisine can be chosen as per each neighborhood.