

# Capstone Project - The Battle of Neighborhoods

## 1. Introduction

New York City is the most populated and cosmopolitan city in the United States, as well as the world's leading business center. It has a very favourable environment for business and trade both nationally and internationally. The city is a cluster of headquarters of multinational companies, mainly in the area of banking and finance, both nationally and internationally. For all the above reasons, it is an optimal and very competitive environment for the development of new companies and the expansion of those that wish to reach higher levels. However, prices per square meter are among the highest in the world, so to reduce business investment risks, it is necessary to invest optimally to achieve the objectives and reduce risks in obtaining profitable operating ratios and margins.

The hospitality sector is very broad and culturally diverse in New York City. Its variety of international cuisines due to the high number of immigrants residing and visiting the city on business frequently, such as

- Asian restaurants.
- Italian restaurants.
- Middle Eastern restaurants.
- Central and Eastern European restaurants.
- Jewish restaurants.
- Haute cuisine restaurants.
- Fast food restaurants.

Among many others.

The success of the business implementation depends on the location of the business and the ability to replicate it in new areas using the data. Therefore, the objective would be to locate and recommend the most optimal address to open a restaurant based on the casuistry of the area.

## 2. Data

Location: New York City

Data 1. 2014 New York City Neighborhood Names from NYU:  
[https://geo.nyu.edu/catalog/nyu\\_2451\\_34572](https://geo.nyu.edu/catalog/nyu_2451_34572)

Data 2. DOHMH Farmers Markets from NYC:  
<https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2>

Data 3. New York City from Wikipedia:  
[https://en.wikipedia.org/wiki/New\\_York\\_City](https://en.wikipedia.org/wiki/New_York_City)

Data 4. Economy of New York City from Wikipedia:  
[https://en.wikipedia.org/wiki/Economy\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Economy_of_New_York_City)

Data 5. New York City Portal from Wikipedia:  
[https://en.wikipedia.org/wiki/Portal:New\\_York\\_City](https://en.wikipedia.org/wiki/Portal:New_York_City)

Data 6. Cuisine of New York City from Wikipedia:  
[https://en.wikipedia.org/wiki/Cuisine\\_of\\_New\\_York\\_City](https://en.wikipedia.org/wiki/Cuisine_of_New_York_City)

Data 7. List of Michelin starred restaurants in New York City from Wikipedia:  
[https://en.wikipedia.org/wiki/List\\_of\\_Michelin\\_starred\\_restaurants\\_in\\_New\\_York\\_City](https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City)

Data 8. Foursquare API.

### 3. Methodology

#### Business Understanding

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

#### Analytic Approach

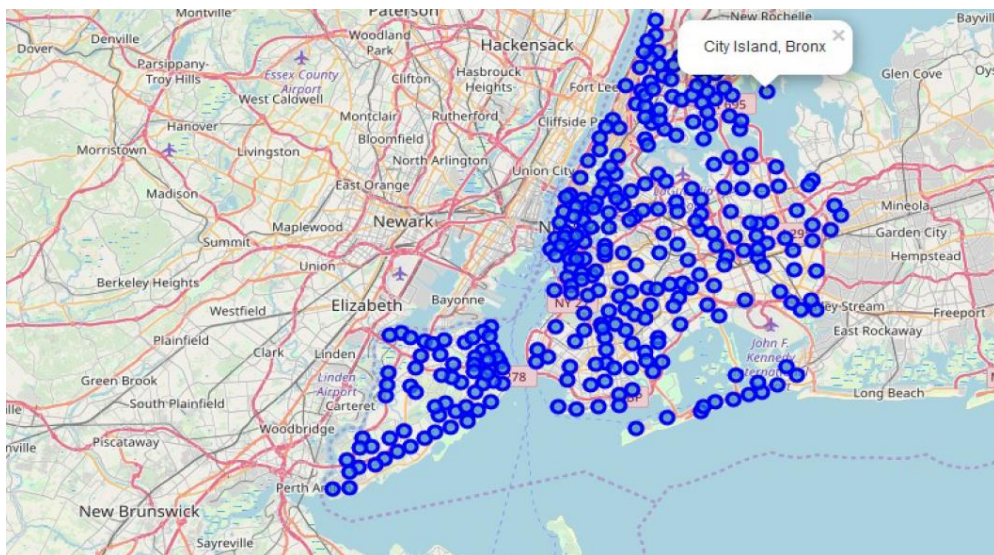
New York city neighbourhood has a total of 5 boroughs and 306 neighbourhoods. In this project first part is clustering of Manhattan and Brooklyn. And second part is clustering of Bronx, Queens and Staten Island. This is done because of the following Exploratory data analysis.

#### Exploratory Data Analysis

Data 1- New york city Geographical Coordinates Data.

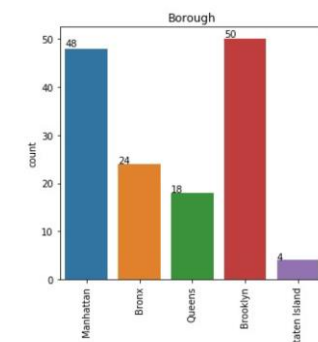
1. In this we load the data and explore data from newyork\_data.json file.
2. Transform the data of nested python dictionaries into a pandas dataframe.
3. This dataframe contains the geographical coordinates of New York city neighborhoods.
4. This data will be used to get Venues data from Foursquare.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

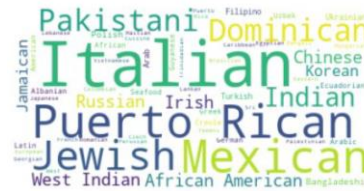
2.New York City Demographics: New York City is the most populous city in the United States, with an estimated record high of 8,622,698 residents as of 2017, incorporating more immigration into the city than outmigration since the 2010 United States Census.

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.

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



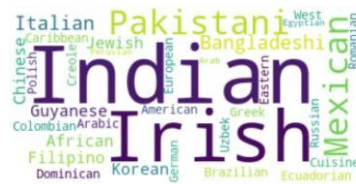
**BROOKLYN CUISINE** -Most Preferred Food in Brooklyn is –Italian, Purto Rican & Mexican



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.

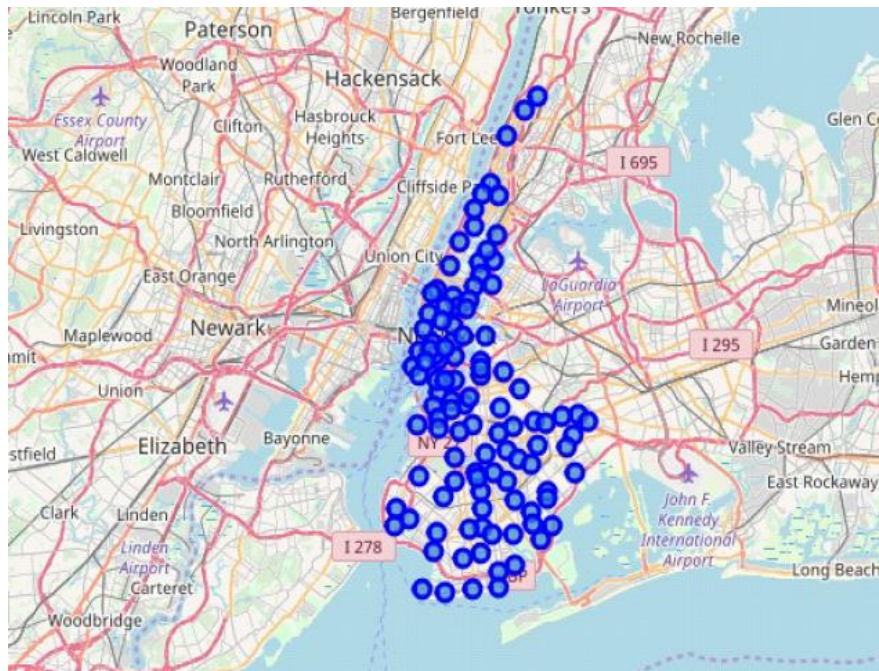


There is very less data of cuisine relating to Staten Island. So could not develop word cloud with it.

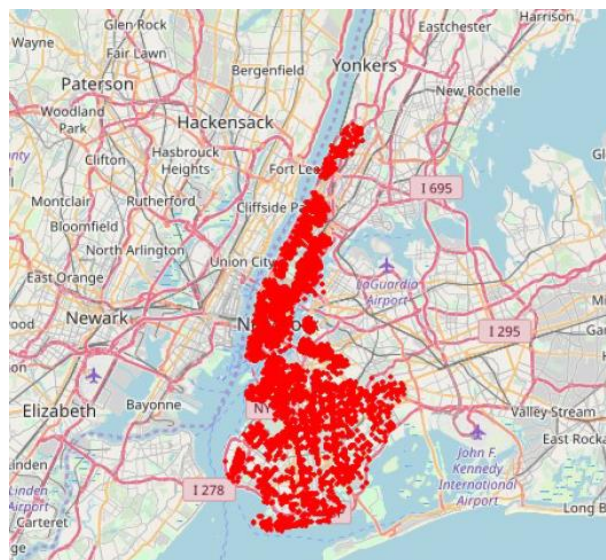


Data 4: New York city geographical coordinates data has been utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighborhood. We used the Foursquare API data to explore neighborhoods in New York City.

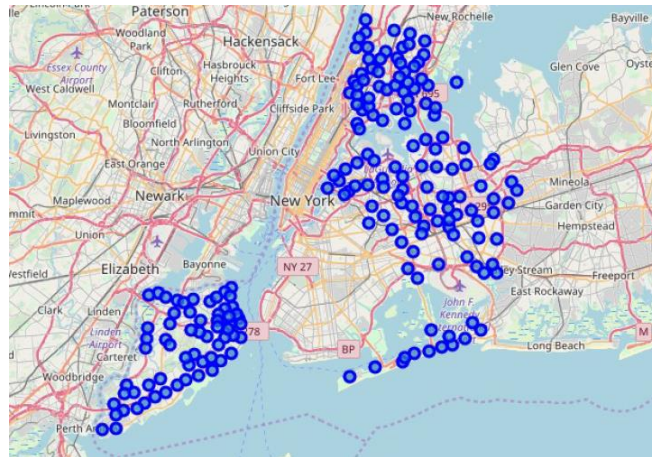
Brooklyn and Manhattan:



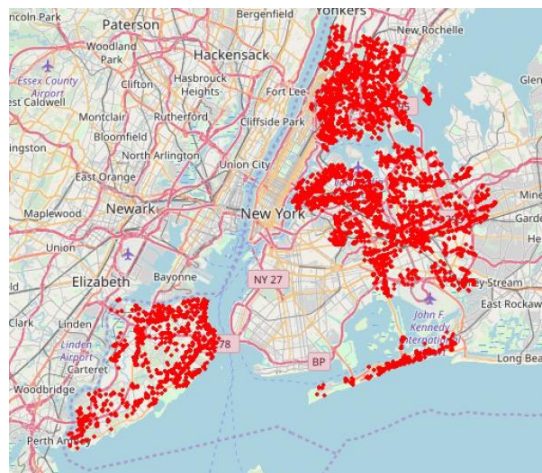
Brooklyn and Manhattan Venues Visualization:



Bronx, Queens and Staten Island:



Bronx, Queens and Staten Island Venues Visualization :



## 4. Results

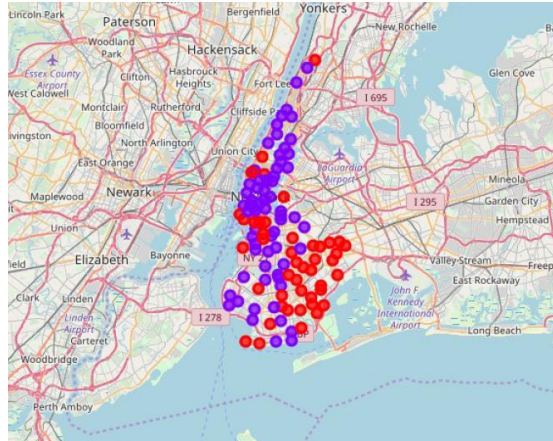
From this venues data we filtered and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens and Staten Island clustering. As we focussed only on restaurants business.

**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.

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 analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available there is scope to come up with better results. If there are lot of restaurants probably there is lot of demand. Brooklyn and Manhattan has high concentration of restaurant business. Very competitive market. Bronx, Queens and Staten Island also has good number of restaurants but not as many as required. So this can be explored.

As per the neighbourhood or restaurant type mentioned like Indian Restaurant analysis can be checked. A venue with lowest risk and competition can be identified.