**Realized by: EL GHOUZAL SAAD**
**January 2021**

# *The Battle of the Neighborhoods - Report -*

## 1. Introduction & Business Problem:

**Problem Background:**

The City of New York, is the most populous city in the United States. It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business oppourtunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce. 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. This also means that the market is highly competitive. As it is highly developed city so cost of doing business is also one of the highest. Thus, any new business venture or expansion needs to be analysed carefully. The insights derived from analysis will give good understanding of the business environment which help in strategically targeting the market. This will help in reduction of risk. And the Return on Investment will be reasonable.

**Problem Description:**

A restaurant is a business which prepares and serves food and drink to customers in return for money, either paid before the meal, after the meal, or with an open account. The City of New York is famous for its excellent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history.

1. Central and Eastern European immigrants, especially Jewish immigrants - bagels, cheesecake, hot dogs, knishes, and delicatessens
2. Italian immigrants - New York-style pizza and Italian cuisine
3. Jewish immigrants and Irish immigrants - pastrami and corned beef
4. Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city
5. mobile food vendors - Some 4,000 licensed by the city
6. Middle Eastern foods such as falafel and kebabs examples of modern New York street food
7. It is famous for not just Pizzerias, Cafe's but also for fine dining Michelin starred restaurants.The city is home to "nearly one thousand of the finest and most diverse haute cuisine restaurants in the world", according to Michelin. So it is evident that to survive in such competitive market it is very important to strategically plan. Various factors need to be studied in order to decide on the Location such as :

- New York Population
- New York City Demographics
- Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost?
- Are there any venues like Gyms, Entertainment zones, Parks etc nearby where floating population is high etc
- Who are the competitors in that location?

- Cuisine served / Menu of the competitors
- Segmentation of the Borough

1. Untapped markets
2. Saturated markets etc...

The list can go on...

Even though well funded ABC Company Ltd. need to choose the correct location to start its first venture.If this is successful they can replicate the same in other locations. First move is very important, thereby choice of location is very important.

**Target Audience:**

To recommend the correct location, ABC Company Ltd has appointed me to lead of the Data Science team. The objective is to locate and recommend to the management which neighborhood of New York city will be best choice to start a restaurant. The Management also expects to understand the rationale of the recommendations made. This would interest anyone who wants to start a new restaurant in New York city.

**Success Criteria:**

The success criteria of the project will be a good recommendation of borough/Neighborhood choice to ABC Company Ltd based on Lack of such restaurants in that location and nearest suppliers of ingredients.

## 2. Data :

One city will be analyze in this project : **New York City.**

We will be using the below datasets for analyzing New York city

**Data 1** : Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segment 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 longitude 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)

**Data 2** : 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) https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j/data#Export (https://data.cityofnewyork.us/Health/DOHMH-New-York-City-Restaurant-Inspection-Results/43nn-pn8j/data#Export)

**Data 3** : 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 neighborhood.We will use the Foursquare API to explore neighborhoods in New York City. The below is image of the Foursquare API data.

## 3. Methodology :

**Business Understanding :**

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

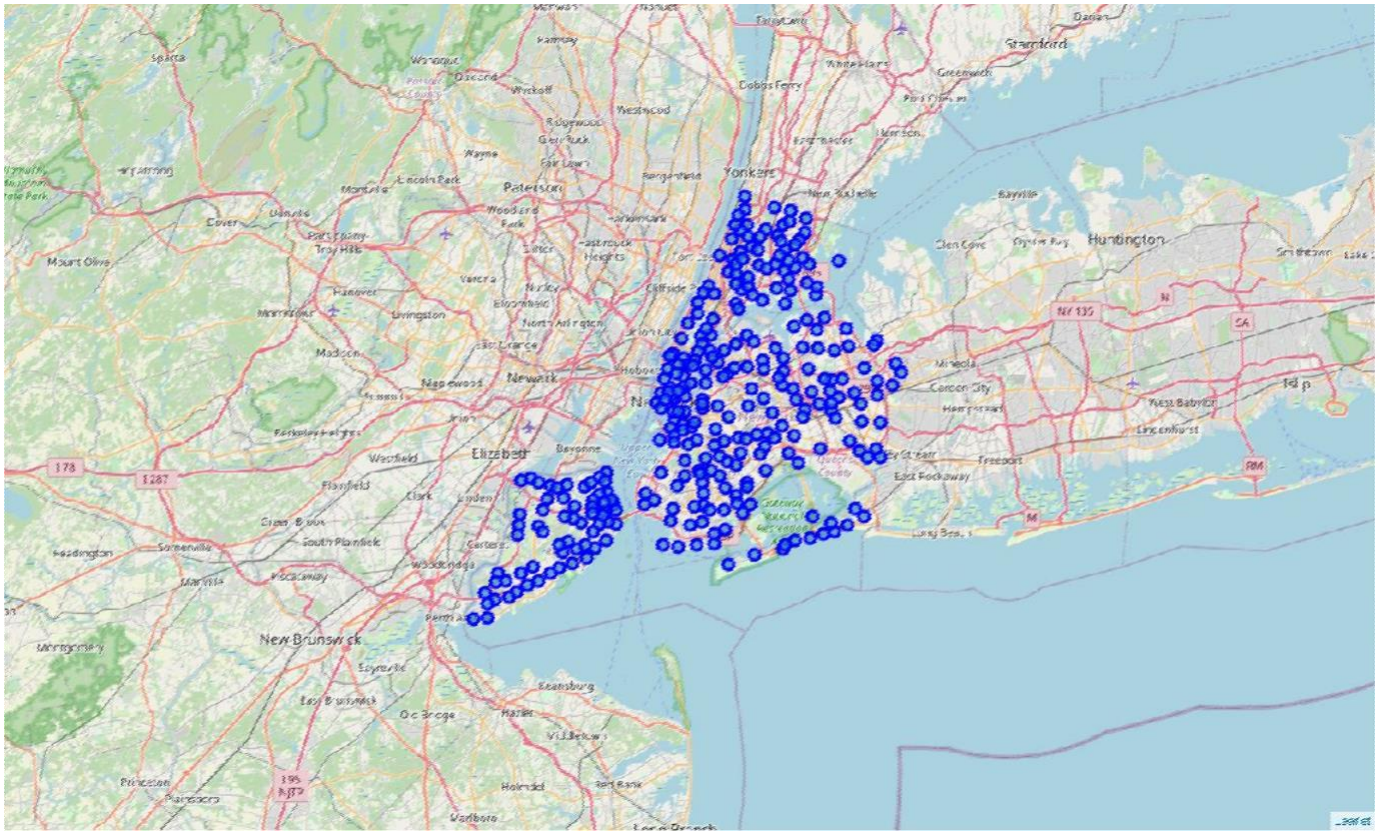
**Exploratory Data Analysis :**

**Data 1 : New York city Geographical Coordinates Data.**

4. In this we load the data and explore data from newyork_data.json file.
5. Transform the data of nested python dictionaries into a pandas dataframe.
6. This dataframe contains the geographical coordinates of New York city neighborhoods.
7. This data will used to get Venues data from Fouresquare.
8. We used geopy and folium libraries to create a map of New York city with neighborhoods superimposed on top.

**Realized by: EL GHOUZAL SAAD**
**January 2021**

**New York neighbourhood visualization.**




**Data 2 :** 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. Beautiful Soup 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 368,500 people per square mile (22.83/km$^2$) in 2017 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 |
|---|---|---|---|---|---|---|
| 0 | The Bronx | Bronx | 1,418,207 | 30,100 | 42.10 | 109.04 |
| 1 | Brooklyn | Kings | 2,559,903 | 35,800 | 70.82 | 183.42 |
| 2 | Manhattan | New York | 1,628,706 | 368,500 | 22.83 | 59.13 |
| 3 | Queens | Queens | 2,253,858 | 41,400 | 108.53 | 281.09 |
| 4 | Staten Island | Richmond | 476,143 | 30,500 | 58.37 | 151.18 |
| 5 | | City of New York | 8,336,817 | 842.343 | 302.64 | 783.83 |
| 6 | | State of New York | 19,453,561 | 1,731.910 | 47,214 | 122,284 |

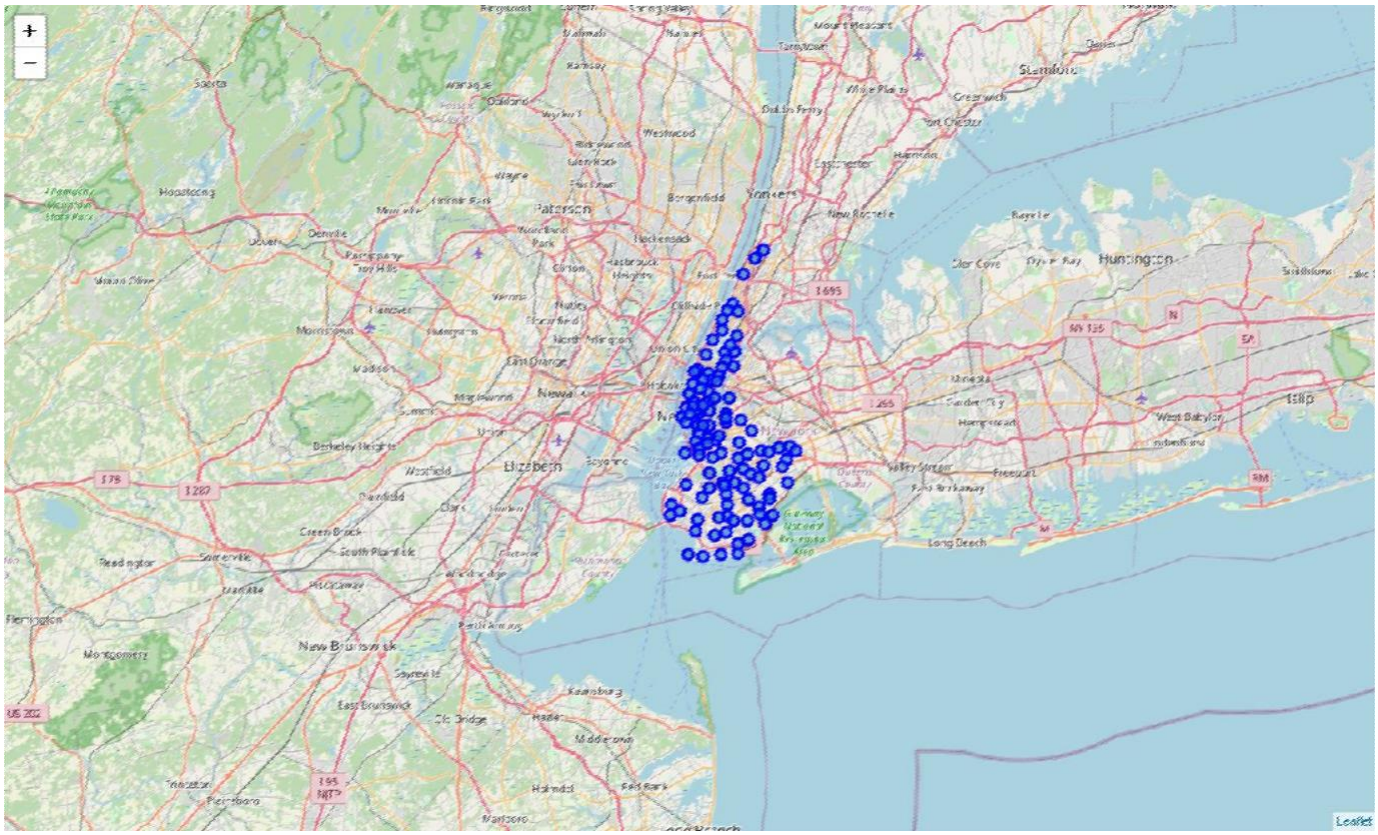
**2. New York City Demographics** : New York City is the most populous city in the United States, with an estimated record high of 8,336,817 residents as of 2017, incorporating more immigration into the city than outmigration since the 2010 United States Census. The racial composition is as given below. 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.

| | Year | Population | White(includes/WhiteHispanics) | %W | Non HispanicWhites | %NHG | Black | %B | Asian | %A | OtherorMixed | %O/M | Hispanic/Latino | %H/L | Foreignborn | %FB/n |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1900 | 3,437,202 | 3,380,898 | 98.04 | | | 50,666 | 1.76 | 6,607 | 0.19 | 31 | 0.00 | | | 1,270,080 | 36.95 |
| 1 | 1910 | 4,766,883 | 4,666,162 | 97.85 | | | 91,709 | 1.92 | 5,669 | 0.12 | 343 | 0.01 | | | 1,941,257 | 40.79 |
| 2 | 1920 | 5,620,048 | 5,459,463 | 97.14 | | | 152,467 | 2.71 | 7,969 | 0.14 | 149 | 0.00 | | | 2,028,160 | 36.09 |
| 3 | 1930 | 6,930,446 | 6,536,377 | 95.03 | | | 327,706 | 4.73 | 12,972 | 0.19 | 391 | 0.01 | | | 2,358,686 | 34.03 |
| 4 | 1940 | 7,454,995 | 6,977,501 | 93.53 | 6,456,586 | 91.97 | 458,444 | 6.15 | 17,986 | 0.24 | 1,064 | 0.01 | 178,915 | 1.67 | 2,118,657 | 28.69 |
| 5 | 1950 | 7,891,957 | 7,116,441 | 90.17 | | | 747,608 | 9.47 | 21,441 | 0.27 | 6,467 | 0.08 | | | 1,734,206 | 22.61 |
| 6 | 1960 | 7,781,984 | 6,940,662 | 85.33 | | | 1,087,931 | 13.38 | 43,103 | 0.53 | 10,288 | 0.13 | | | 1,558,890 | 20.03 |
| 7 | 1970 | 7,894,862 | 6,048,641 | 76.62 | 4,968,749 | 62.95 | 1,668,115 | 21.13 | 54,499 | 1.40 | 63,407 | 1.08 | 1,278,830 | 16.2 | 1,457,058 | 18.20 |
| 8 | 1990 | 7,071,639 | 4,234,075 | 60.72 | 3,666,945 | 51.88 | 1,784,337 | 25.23 | 231,501 | 3.27 | 761,762 | 10.77 | 1,426,024 | 19.88 | 1,670,199 | 23.62 |
| 9 | 1990 | 7,322,564 | 3,827,088 | 52.25 | 3,168,115 | 43.2 | 2,102,512 | 28.71 | 512,719 | 7.30 | 360,245 | 12.02 | 1,735,511 | 24.36 | 2,036,931 | 28.45 |
| 10 | 2000 | 8,008,278 | 3,576,385 | 44.65 | 2,801,267 | 34.98 | 2,129,762 | 26.59 | 792,477 | 9.90 | 1,509,654 | 18.85 | 2,160,554 | 26.98 | 2,871,032 | 35.85 |
| 11 | 2010 | 8,175,133 | 3,597,341 | 44.03 | 2,722,984 | 33.31 | 2,056,218 | 25.55 | 1,343,535 | 12.77 | 1,445,747 | 17.68 | 2,336,076 | 28.58 | 3,056,199 | 37.51 |

**Data 3 :** 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 neighborhood. We used the Foursquare API data to explore neighborhoods in New York City.

**Brooklyn and Manhattan Visualization :**

**Realized by: EL GHOUZAL SAAD**
**January 2021**



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 :**

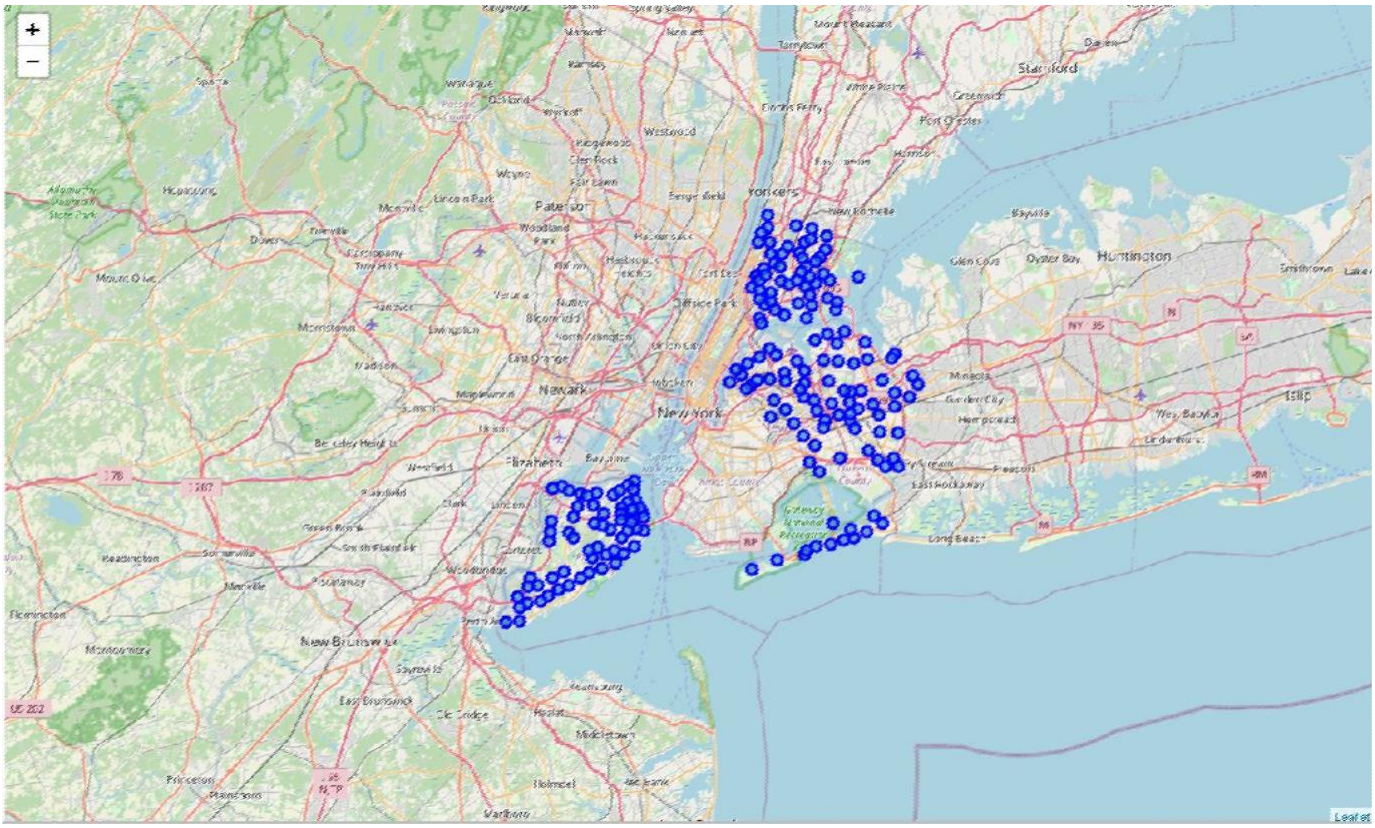
| | Neighborhood | NeighborhoodLatitude | NeighborhoodLongitude | Venue | VenueLatitude | VenueLongitude | VenueCategory |
|---|---|---|---|---|---|---|---|
| 0 | Marble Hill | 40.876551 | -73.91066 | Bikram Yoga | 40.876844 | -73.906204 | Yoga Studio |
| 1 | Marble Hill | 40.876551 | -73.91066 | Arturo's | 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 Visualization : Generated the below Brooklyn and Manhattan Venues Visualization. The "BM_venues" dataframe has 407 uniques categories.

**Realized by: EL GHOUZAL SAAD**
**January 2021**

| Neighborhood | NeighborhoodLatitude | NeighborhoodLongitude | Venue | VenueLatitude | VenueLongitude | VenueCategory |
|---|---|---|---|---|---|---|
| Bath Beach | 94 | 94 | 94 | 94 | 94 | 94 |
| Battery Park City | 100 | 100 | 100 | 100 | 100 | 100 |
| Bay Ridge | 100 | 100 | 100 | 100 | 100 | 100 |
| Bedford Stuyvesant | 100 | 100 | 100 | 100 | 100 | 100 |
| Bensonhurst | 100 | 100 | 100 | 100 | 100 | 100 |
| Bergen Beach | 10 | 10 | 10 | 10 | 10 | 10 |
| Boerum Hill | 100 | 100 | 100 | 100 | 100 | 100 |

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




Bronx, Queens and Staten Island Venues Visualization : The "BQS_venues" dataframe has 386 uniques categories.

| Neighborhood | Neighborhood Latitude | Neighborhood Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|
| Allerton | 66 | 66 | 66 | 66 | 66 | 66 |
| Annadale | 19 | 19 | 19 | 19 | 19 | 19 |
| Arden Heights | 22 | 22 | 22 | 22 | 22 | 22 |
| Arlington | 22 | 22 | 22 | 22 | 22 | 22 |
| Arrochar | 23 | 23 | 23 | 23 | 23 | 23 |
| Arverne | 34 | 34 | 34 | 34 | 34 | 34 |
| Astoria | 100 | 100 | 100 | 100 | 100 | 100 |
| Astoria Heights | 76 | 76 | 76 | 76 | 76 | 76 |

## 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 focused only on restaurants business.

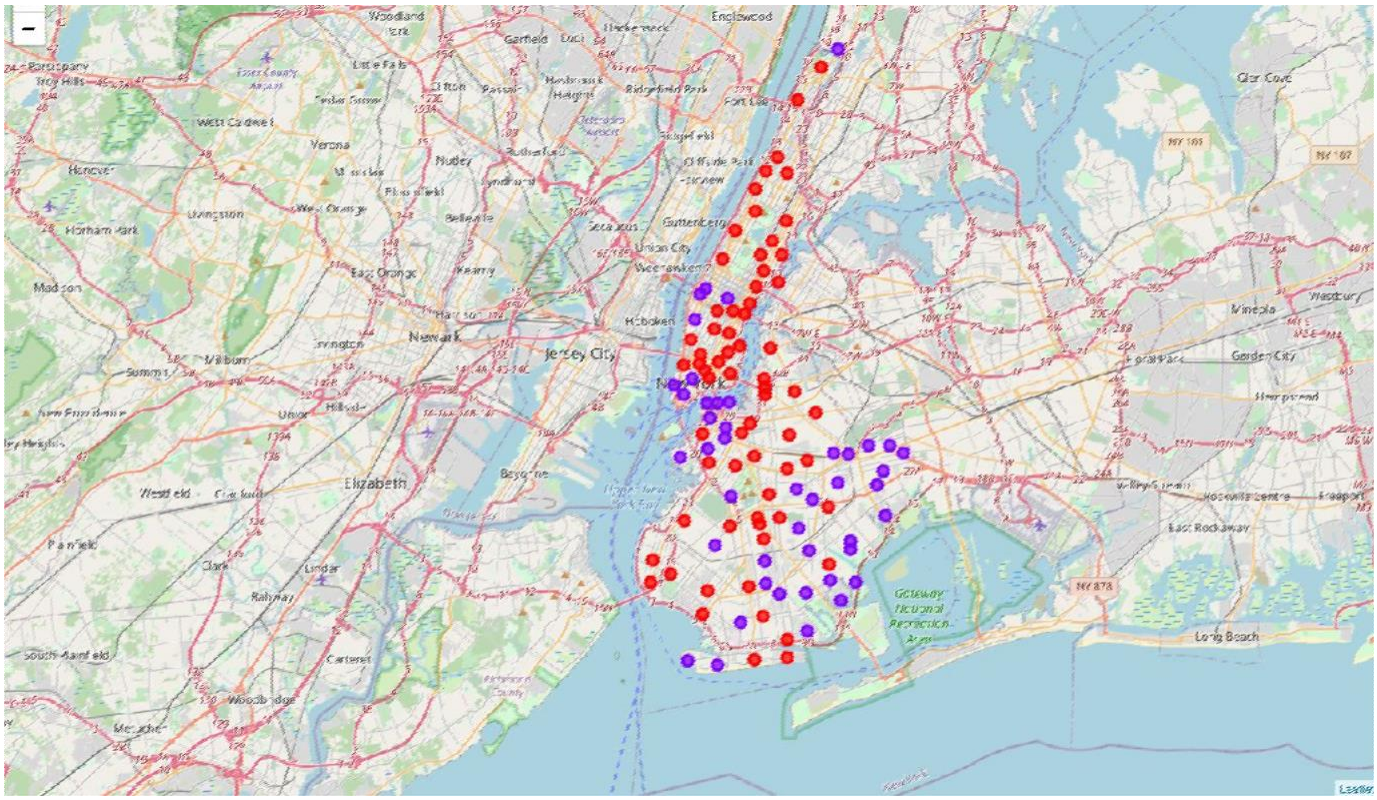
**Neighborhood K-Means clustering based on mean occurrence of venue category :**

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.

**Realized by: EL GHOUZAL SAAD**
**January 2021**

**Brooklyn & Manhattan :**

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



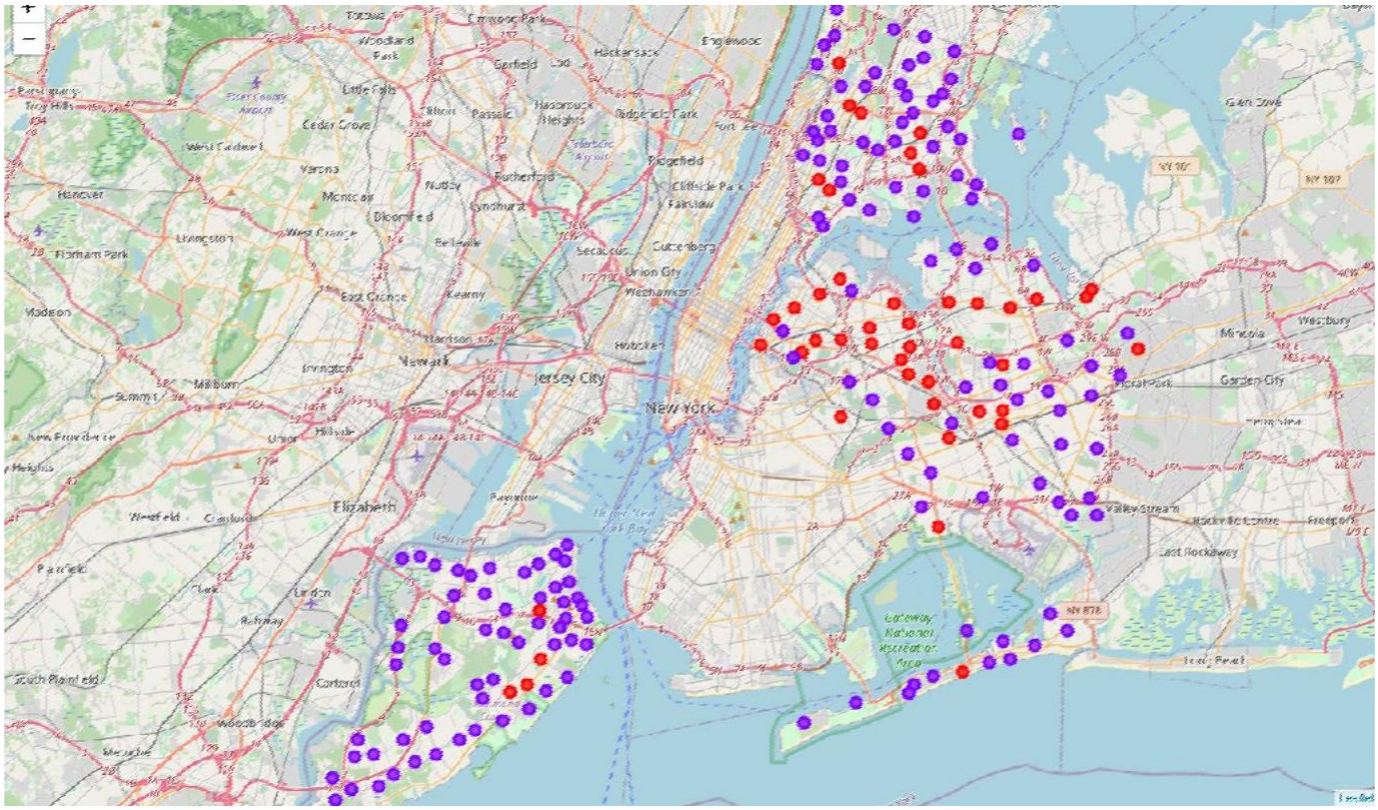

**Cluster0 :** The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.

**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 | 1 |
| 1 | Staten Island | Port Ivory | 40.639683 | -74.174645 | 0 | 1 |
| 2 | Staten Island | Butler Manor | 40.506082 | -74.229504 | 0 | 1 |
| 3 | 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 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.