# The Battle of 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 ia the financial capital of USA. It is multicultural. It provides lot of business opportunities 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 analyzed carefully. The insights derived from analysis will give good understanding of 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 are the examples of modern New York street food
7. It is famous for not just Pizzerias, Café'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 :

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.
5. Who are the competitors in that location?
6. Cuisine served / Menu of the competitors
7. Segmentation of the borough
8. Untapped markets
9. Saturated markets etc

Even though well funded DNB Foodies Company Ltd has appointed me to lead 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 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 DNB Foodies Company Ltd based on lack of such restaurants in that location and the nearest suppliers of ingredients.

# 2.Data :

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

We will bw using the below datasets for analuzing New York City

**Data 1 :** Neighborhoods 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 latitude and longitude coordinates of each Neighborhoods.

This dataset exists for free on the web. Link to the dataset is :

http://geo.nyu.edu/catalog/nyu_2451_34572

|   | Borough | Neighborhood | Latitude | Longitude |
|---|---------|--------------|----------|-----------|
| 0 | Bronx | Wakefield | 40.894705 | -73.847201 |
| 1 | Bronx | Co-op City | 40.874294 | -73.829939 |
| 2 | Bronx | Eastchester | 40.887556 | -73.827806 |
| 3 | Bronx | Fieldston | 40.895437 | -73.905643 |
| 4 | Bronx | Riverdale | 40.890834 | -73.912585 |

**Data 2 :** Second data which will be used is the DOHMH Farmers Markets 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

GrowNYC's Fresh Food Box Program is a food access initiative that enables under-served communities to purchase fresh, healthy, and primarily regionally grown produce well below traditional retail prices.

A farmers' market is often defined as a public site used by two or more local or regional producers for the direct sale of farm products to consumers. In addition to fresh fruits and vegetables, markets may sell dairy products, fish, meat, baked goods, and other minimally processed foods.

| | Borough | Market Name | Street Address | Latitude | Longitude | Days of Operation | Hours of Operations | Season Dates | Accepts EBT | Open Year-Round | Stellar Demor |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Brooklyn | Woodhull Hospital Youthmarket | Broadway & Flushing Ave | 40.700726 | -73.941932 | Wednesday | 9 a.m. - 2 p.m. | 07/10/2019-11/27/2019 | Yes | No | No |
| 1 | Manhattan | Mount Sinai Hospital Greenmarket | E 99th St bet Madison & Park Aves | 40.789169 | -73.952743 | Wednesday | 8 a.m. - 5 p.m. | 06/12/19-11/27/19 | Yes | No | No |
| 2 | Bronx | 170 Farm Stand | E 170th St & Townsend Ave | 40.839882 | -73.916783 | Wednesday | 2:30 - 6:30 p.m. | 07/10/2019-11/27/2019 | Yes | No | No |
| 3 | Manhattan | Greenmarket at Oculus Plaza | Church & Fulton Sts, on Oculus Plaza | 40.711535 | -74.010464 | Tuesday | 7 a.m. - 7 p.m. | 07/09/2019-11/30/19 | Yes | Yes | No |
| 4 | Queens | Ditmars Park Youthmarket | Steinway St bet Ditmars Blvd & 23rd Ave, at Di... | 40.772854 | -73.906061 | Saturday | 8 a.m. - 3 p.m. | 07/13/2019-11/23/2019 | Yes | No | No |

**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/Economy_of_New_York_City

https://en.wikipedia.org/wiki/Portal: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 neighborhood. We will use the Foursquare API to explore neighborhoods in New York City. The below is the image of the Foursquare API data.

| | 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 |

# 3.Methodology :

**Business Understanding :**

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

**Analytic Approach :**

 New York City neighborhood has a total of 5 boroughs and 306 neighborhoods. In this project first part is clustering 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.
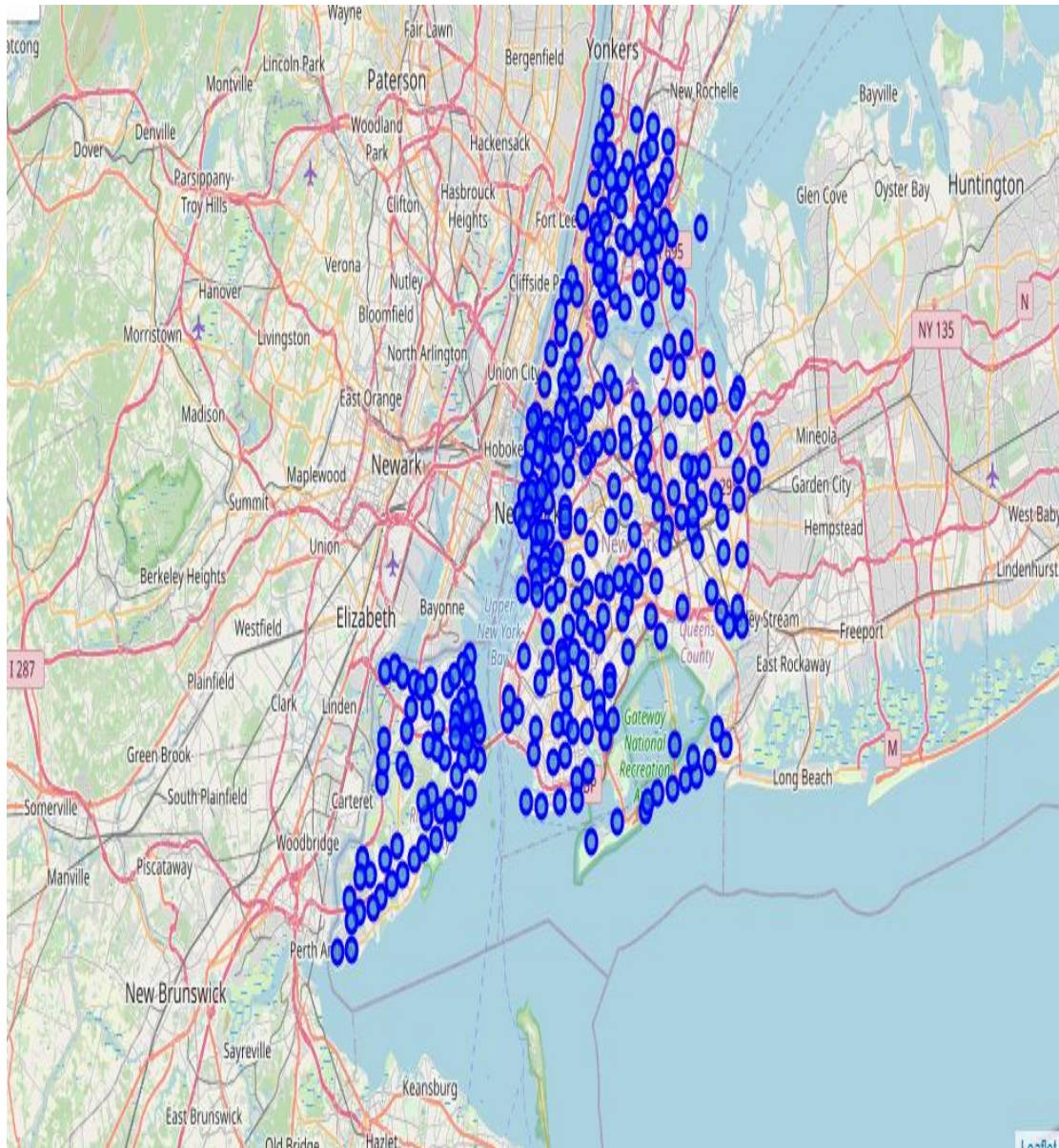
**Data 2 –** Second data which is used is the DOHMH Farmers Markets dataset. In this we will be using the date of Farmers Markets data.

There are total 138 Farmers Markets in New York City. Highest number are in Manhattan and Brooklyn and the lowest in Queens, Bronx and Staten Island. Below is the bar diagram representing the same.
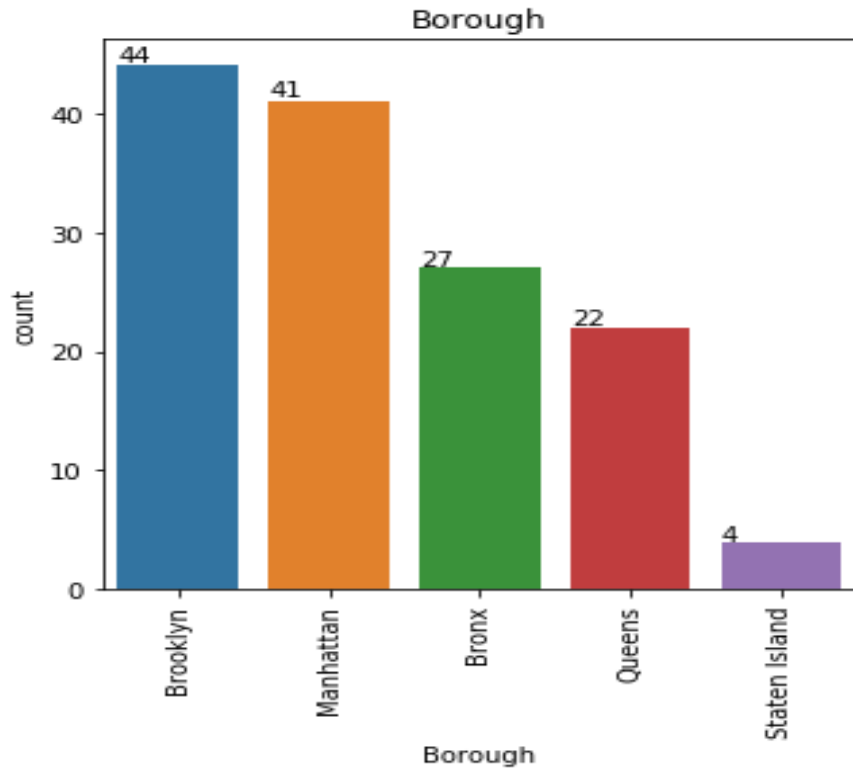
We used geopy and folium libraries to create a map to visualize Farmers Market of New York City.

**Farmers Market Visualization – New York City**
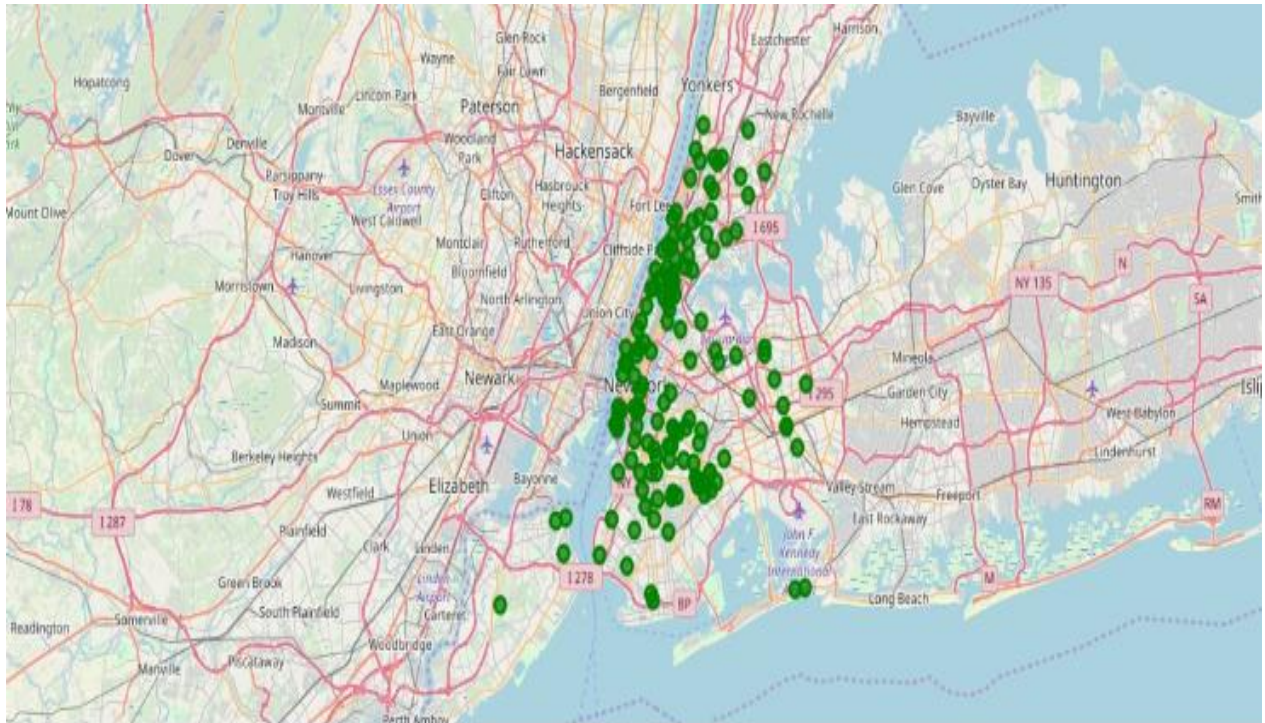
**Data 3 :** To analyze New York City Population, Demographics and Cuisine, data has been scraped from Wikipedia pages given above in the data section. We used Beautiful Soup python library. Beautiful Soup is a python package for parsing HTML and XML documents (including having malformed markup, i.e. 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.
- Manhatten's population density of 72,033 people per square mile (27,812/km2) 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 the 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 business in New York City.

| | Racial composition | 2010 | 1990 | 1970 | 1940 |
|---|---|---|---|---|---|
| 0 | White | 44.00% | 52.30% | 76.60% | 93.60% |
| 1 | Non-Hispanic | 33.30% | 43.20% | 62.90% | 92.00% |
| 2 | Black or African American | 25.50% | 28.70% | 21.10% | 6.10% |
| 3 | Hispanic or Latino | 28.60% | 24.40% | 16.20% | 1.60% |
| 4 | Asian | 12.70% | 7.00% | 1.20% | 0 |

**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. Using this data we did word cloud.

**NEW YORK CITY CUISINE :** Most preferred food in New York City -Italian, Purto Rican, Mexican, Jewish, Indian, Pakistani & Dominican.



**BROOKLYN CUISINE –** Most preferred food in Brooklyn is – Italian, Purto Rican & Mexican
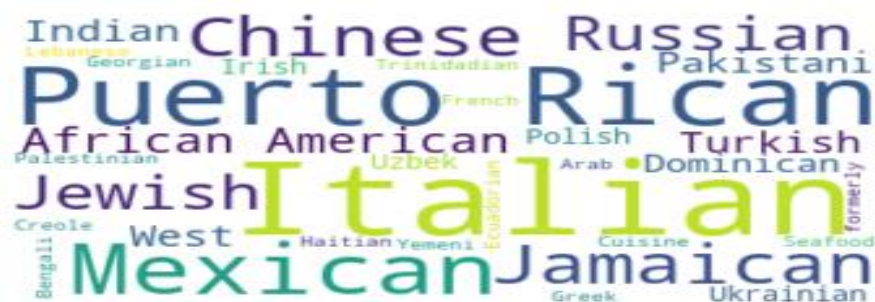
**MANHATTAN CUISINE –** Most preferred food in Manhattan is – Italian, American, Puerto Rican and Indian.



**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 to be utilized as input for 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 :**
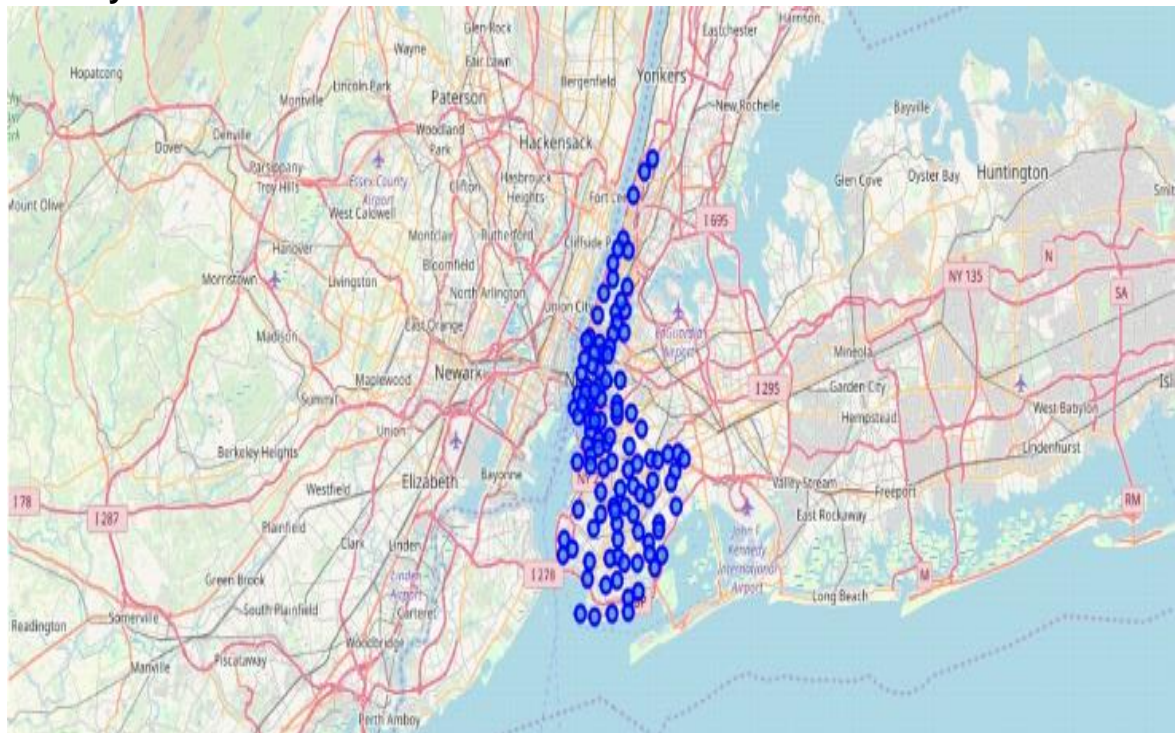
Using the geographical coordinates of each neighborhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters. The venues data is as given below :
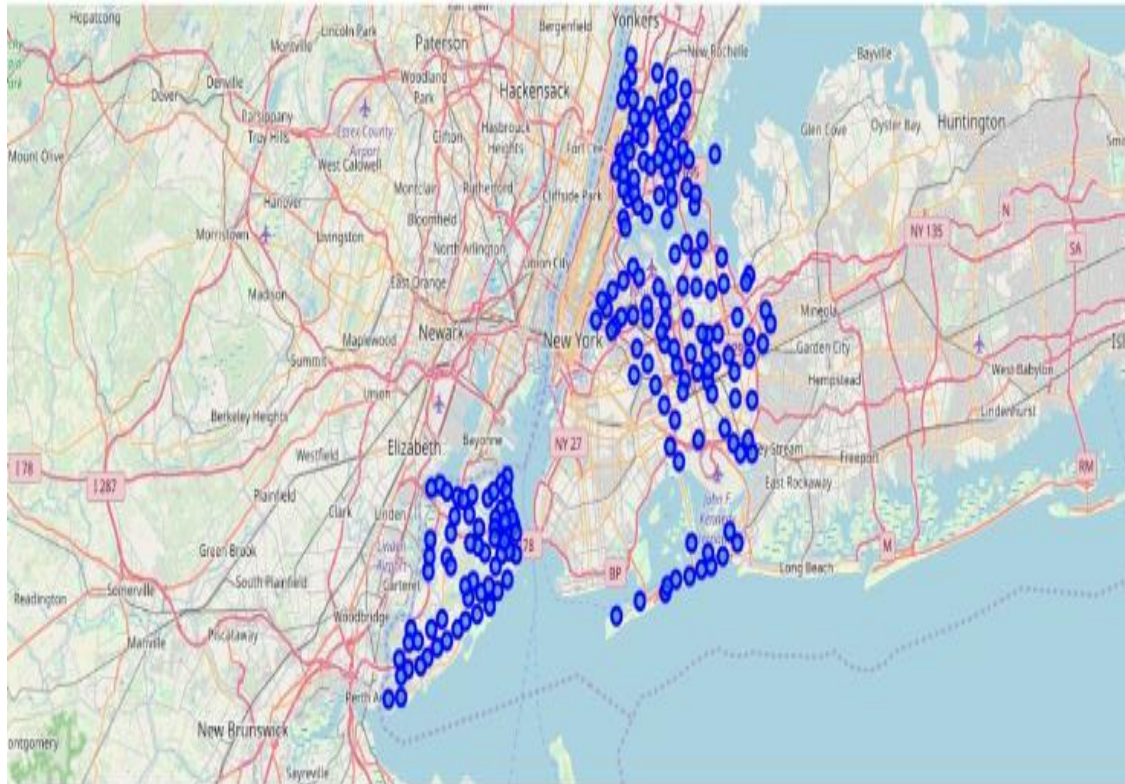
**Brooklyn and Manhattan Venues :**

|  | 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.
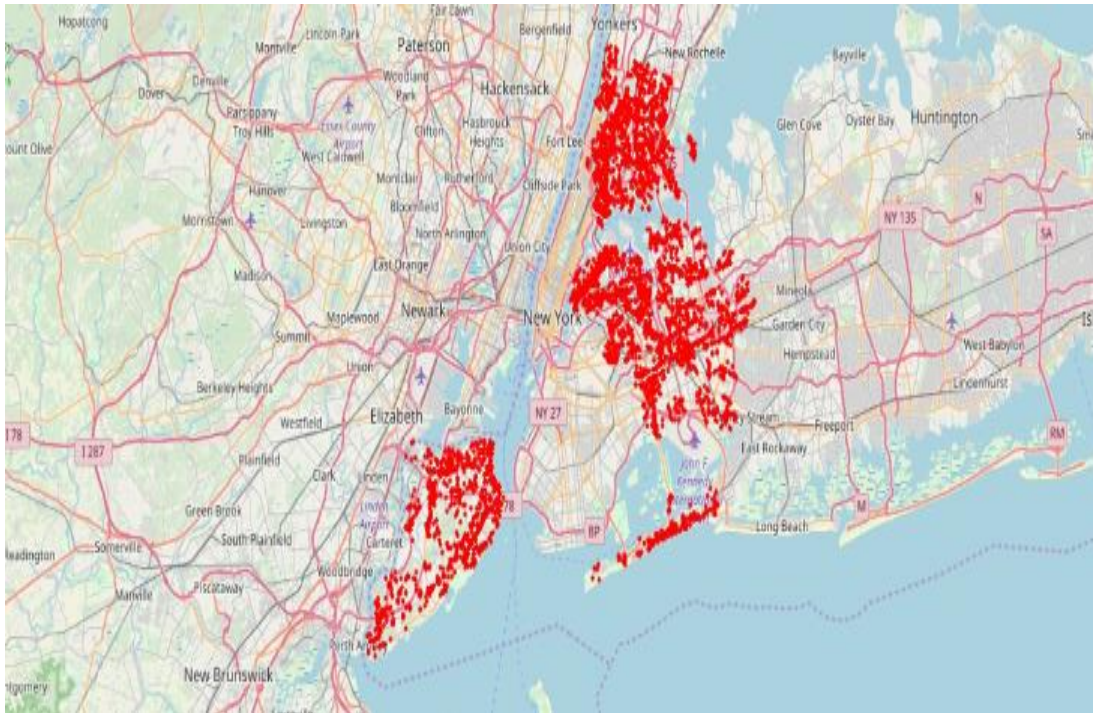
**Bronx, Queens and Staten Island :**

**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 and Manhattan clustering and Bronx, Queens and Staten Island clustering. 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.

**Brooklyn and Manhattan :**

In the below Map Visualization, we can see the different types of clusters created by using K-Means for Brooklyn and Manhattan.
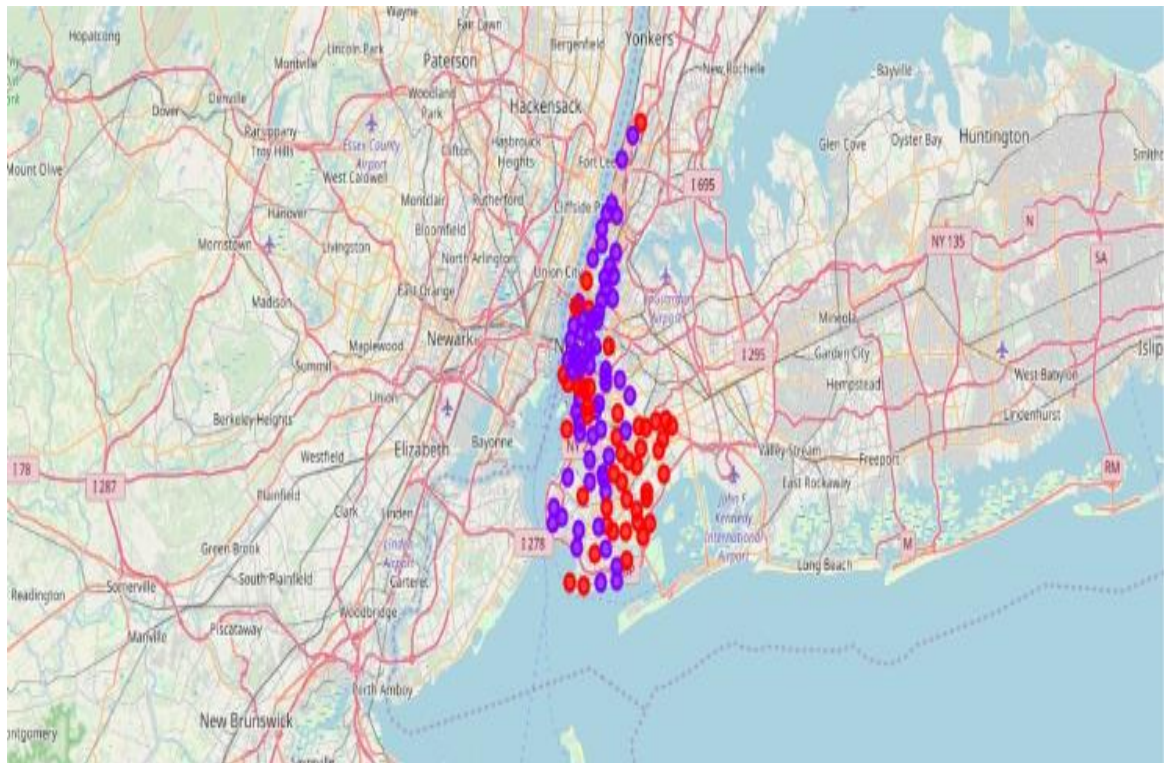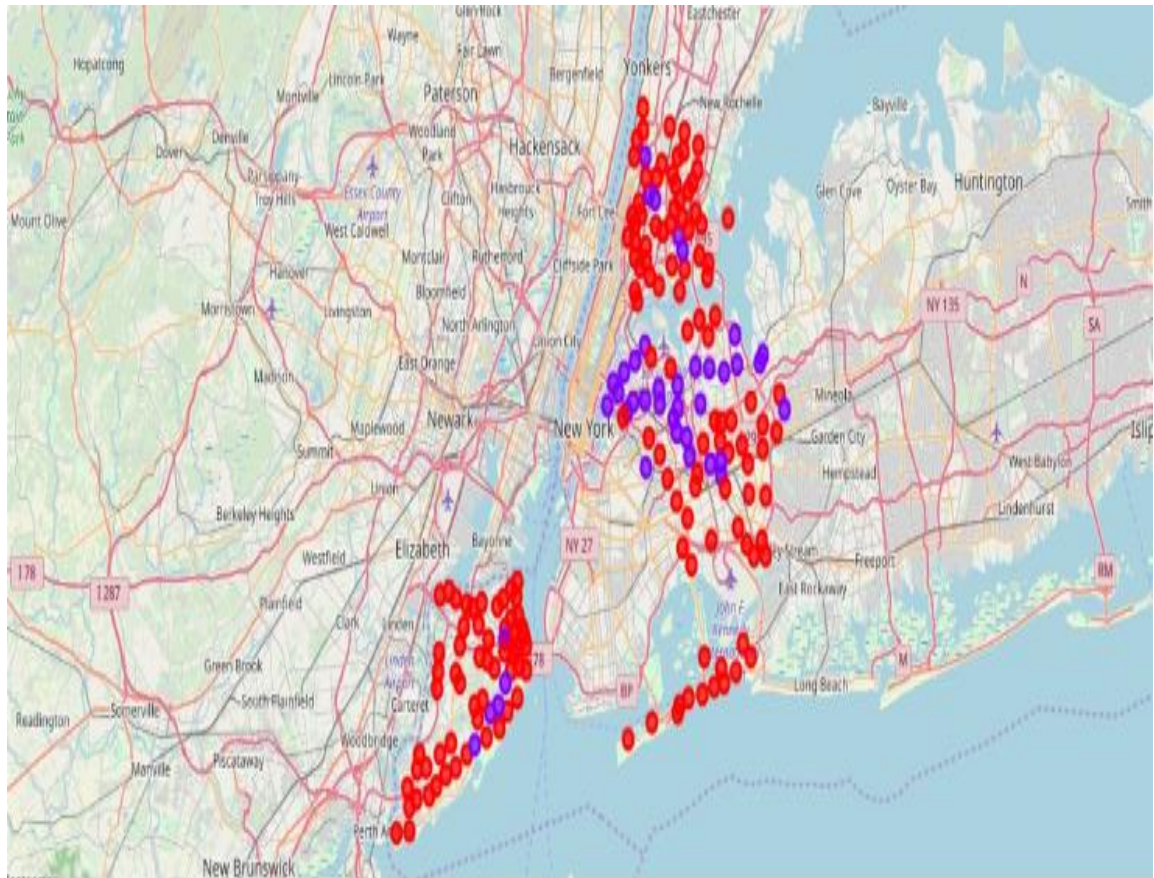
**Cluster 0 :** The Total and Total Sum of cluster 0 has smallest value. It shows that the market is not saturated.

**Cluster 1 :** The Total and Total Sum of cluster 1 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.

**Cluster 0 :** The Total and Total Sum of cluster 0 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 |

**Cluster 1 :** The Total and Total Sum of the cluster 1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

# 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, risk can be taken with great menu on the board. It also shows people love eating cuisines of various countries.

# 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 and have a 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 more.

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