

# 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 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 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
5. ubiquitous throughout the city
6. mobile food vendors - Some 4,000 licensed by the city
7. Middle Eastern foods such as falafel and kebabs examples of modern New York street food
8. It is famous for not just Pizzerias, Cafe's but also for fine dining Michelin starred
9. restaurants. The city is home to "nearly one thousand of the finest and most diverse haute
10. 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 etc.
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 XYZ 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, XYZ 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 XYZ Company Ltd based on Lack of such restaurants in that location and nearest suppliers of ingredients.

## 2. Data:

One city will be analyzed 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 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)

	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 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>

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	Service_Type	Street Address	Latitude	Longitude	Days of Operation	Hours of Operations	Season Dates
0	Brooklyn	Woodhull Hospital Youthmarket	Farmers Markets	Broadway & Flushing Ave	40.700726	-73.941932	Wednesday	9 a.m. - 2 p.m.	07/10/2019-11/27/2019
1	Manhattan	Mount Sinai Hospital Greenmarket	Farmers Markets	E 99th St bet Madison & Park Aves	40.789169	-73.952743	Wednesday	8 a.m. - 5 p.m.	06/12/19-11/27/19
2	Bronx	170 Farm Stand	Farmers Markets	E 170th St & Townsend Ave	40.839882	-73.916783	Wednesday	2:30 - 6:30 p.m.	07/10/2019-11/27/2019
3	Manhattan	Greenmarket at Oculus Plaza	Farmers Markets	Church & Fulton Sts, on Oculus Plaza	40.711535	-74.010464	Tuesday	7 a.m. - 7 p.m.	07/09/2019-11/30/19
4	Queens	Ditmars Park Youthmarket	Farmers Markets	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

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 neighborhood. We will use the foursquare API to explore neighborhoods in New York City. The below is 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 XYZ Company.

### Analytic Approach:

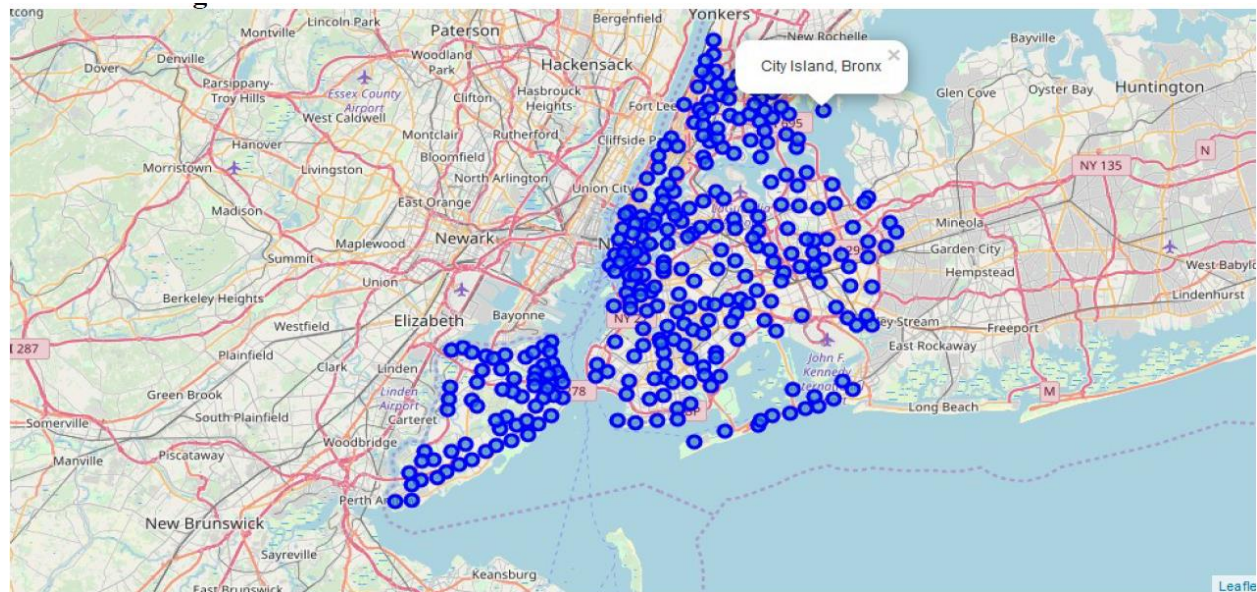
New York city has a total of 5 boroughs and 306 neighborhoods. In this project the first part is clustering of Manhattan and Brooklyn. And the second part is clustering of Bronx, Queens and Staten Island. This is done because of the following Exploratory data analysis.

### 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 data frame.
3. This data frame 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.

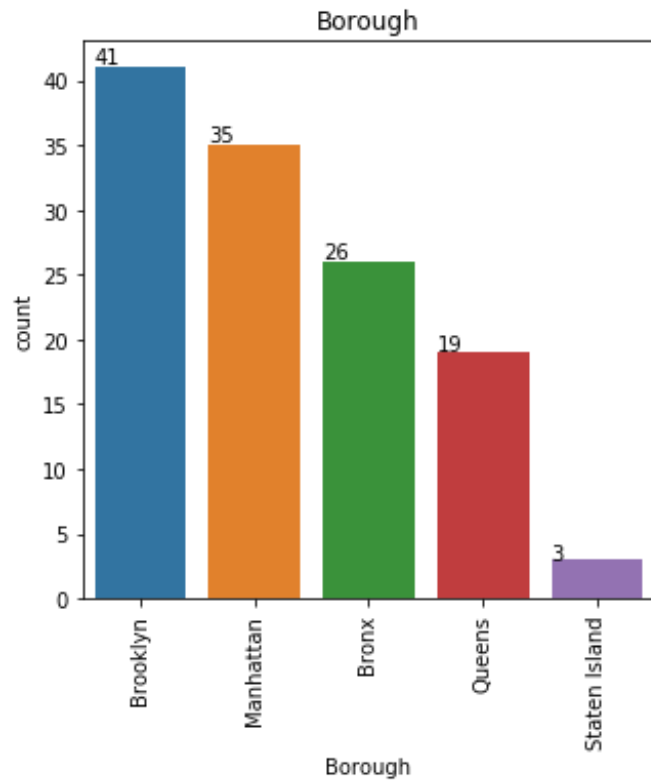
### New York neighborhoods visualization



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.

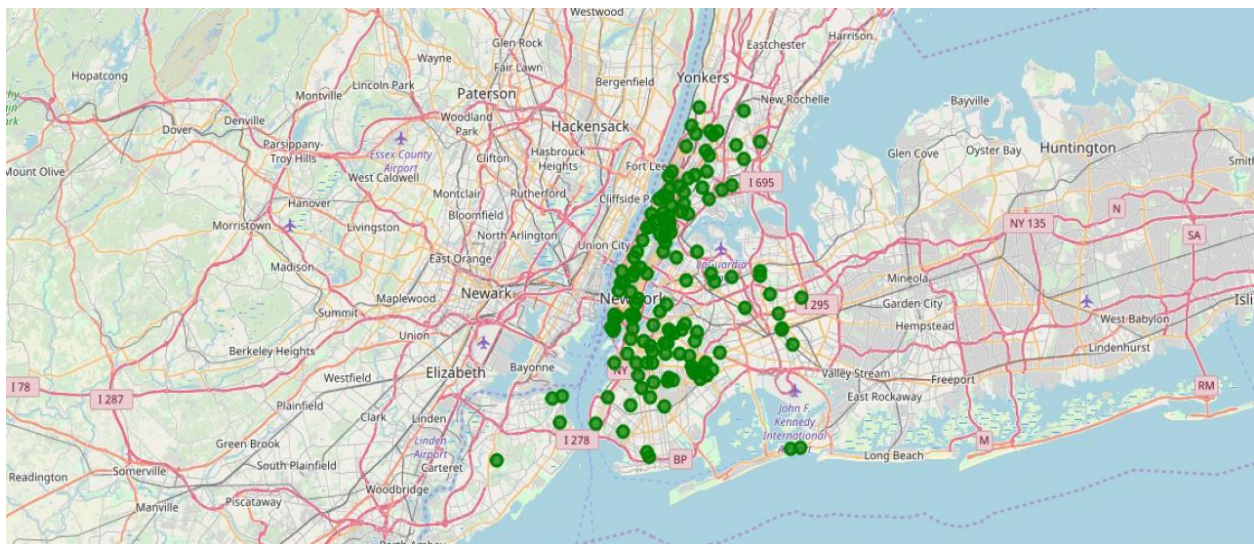
There are totally 124 Farmers Markets in New York city. Highest number are in Manhattan and Brooklyn. And lowest in Queens, Bronx and Staten Island. The proof of this is as given below.





We used geopy and folium libraries to create a map to visualize farmers markets of New York city.

### Farmers Market visualization - 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.
- 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	GrossDomesticProduct	square_miles	square_km	persons_sq_mi
0	The Bronx	Bronx	1,418,207	42.695	30,100	42.10	109.04
1	Brooklyn	Kings	2,559,903	91.559	35,800	70.82	183.42
2	Manhattan	New York	1,628,706	600.244	368,500	22.83	59.13
3	Queens	Queens	2,253,858	93.310	41,400	108.53	281.09
4	Staten Island	Richmond	476,143	14.514	30,500	58.37	151.18
5	City of New York		8,336,817	101,000	842.343	302.64	783.83
6	State of New York		19,453,561	89,000	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.

	Racialcomposition	2010	1990	1970	1940
0	None	None	None	None	None
1	White	44.0%	52.3%	76.6%	93.6%
2	—Non-Hispanic	33.3%	43.2%	62.9%	92.0%
3	Black or African American	25.5%	28.7%	21.1%	6.1%
4	Hispanic or Latino (of any race)	28.6%	24.4%	16.2%	1.6%
5	Asian	12.7%	7.0%	1.2%	—

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.



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



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



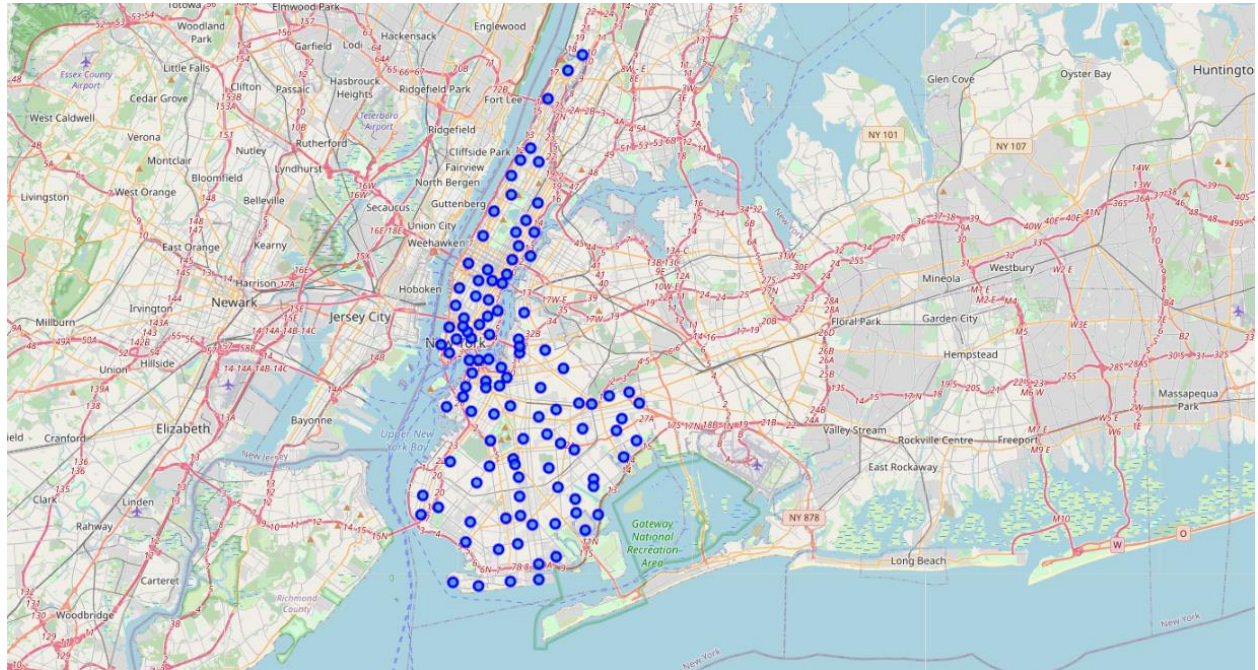
There is very little data of cuisine relating to Staten Island. So we could not develop word cloud with it.

Data 4: New York city geographical coordinates data had to 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 Venues:

### Brooklyn and Manhattan Venues 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:

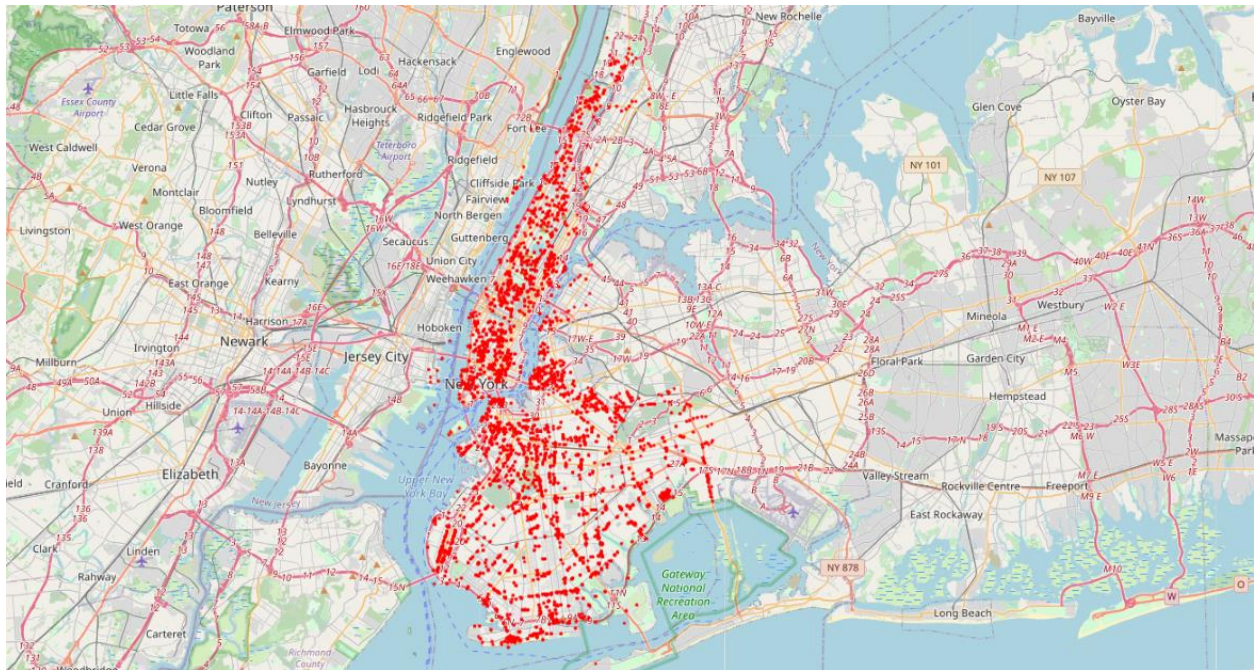
### Brooklyn & Manhattan venues

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Garden Gourmet Market	40.881350	-73.903389	Gourmet Shop
4	Marble Hill	40.876551	-73.91066	Indian Road Café	40.872922	-73.918459	Café

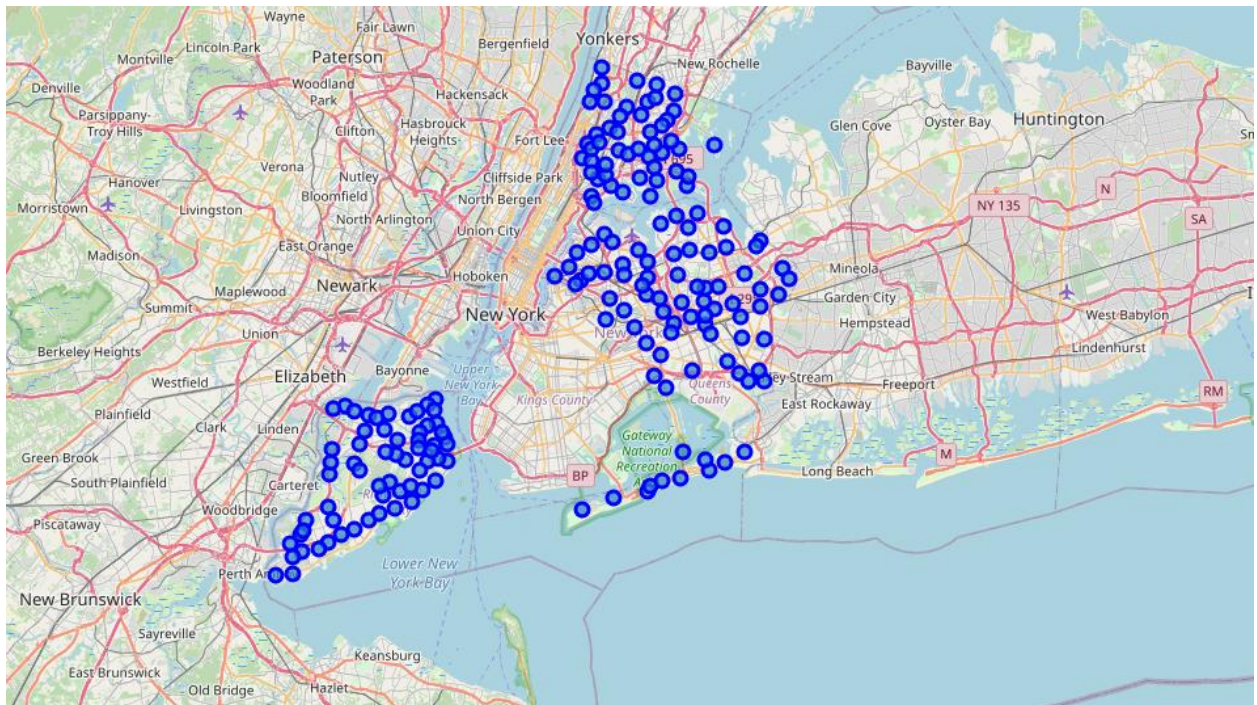
### Brooklyn and Manhattan Venues Visualization:

The "BM\_venues" dataframe has 10997 venues and 300 unique venue types.





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



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

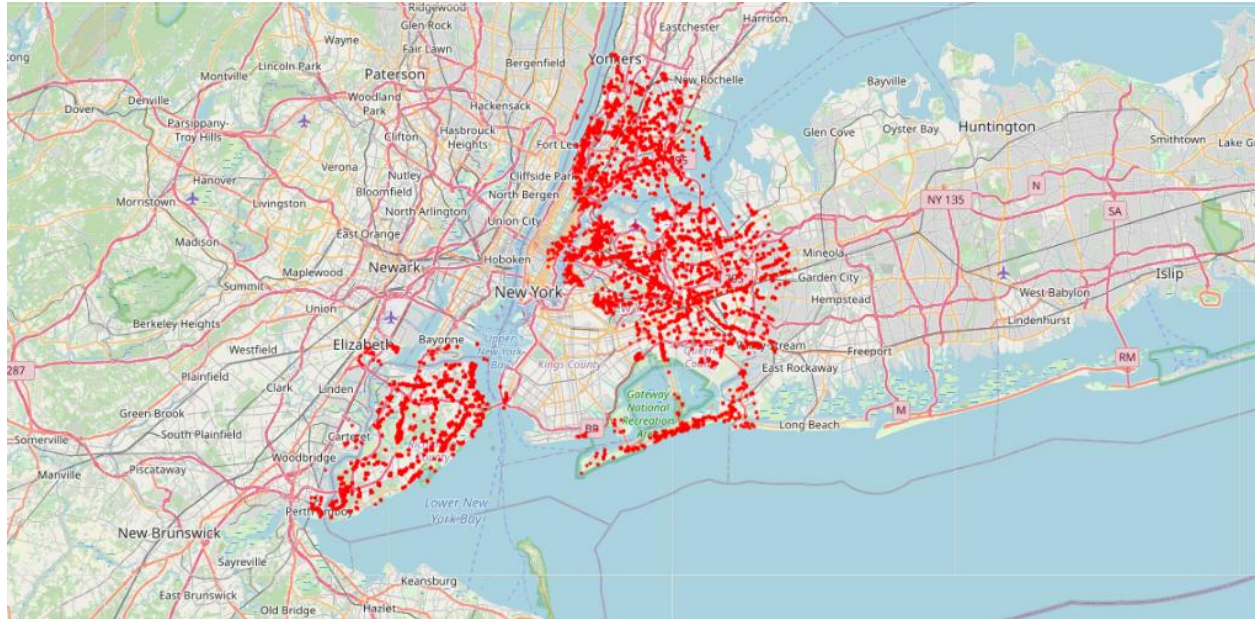
The "BQS\_venues" data frame has 18173 venues and 345 unique venue types.



:[42]:

	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	Alli's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Jimbo's	40.891740	-73.858226	Burger Joint
4	Wakefield	40.894705	-73.847201	Kingston Tropical Bakery	40.888568	-73.859885	Bakery

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



## 4. RESULTS

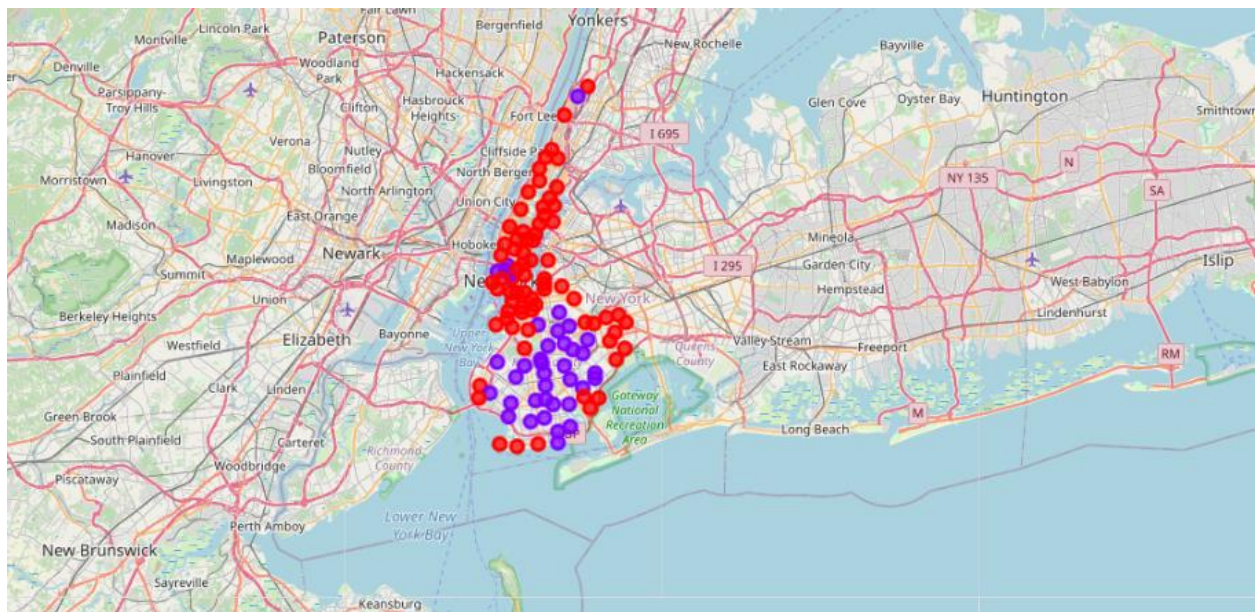
From the venues data we filtered and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens and Staten Island clustering. As we focused only on restaurants business.

### 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.

#### Brooklyn & Manhattan:

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

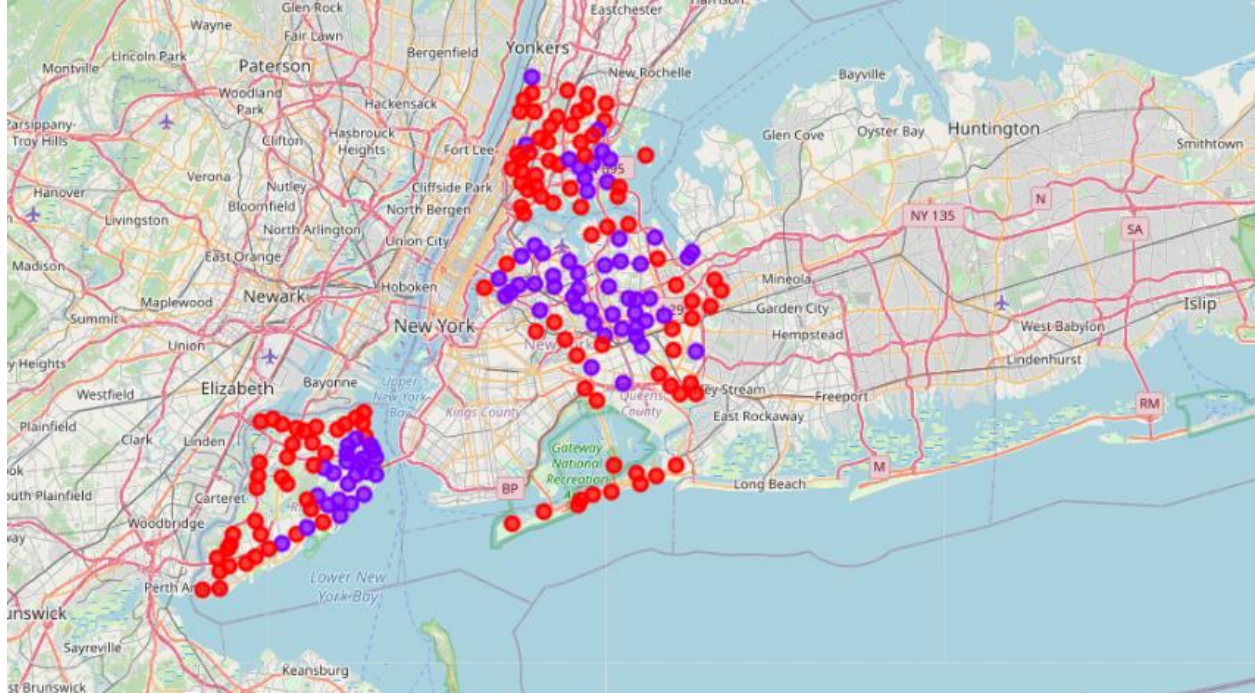


**Cluster0:** The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.

**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.

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 analysis is performed on limited data. This may be right or may be wrong. But if a 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 a high concentration of restaurant business. It's a very competitive market. Bronx, Queens and Staten Island also have a good number of restaurants but not as many as required. So this can still be explored.

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.