The Battle of the Neighborhoods – Report

1.Introduction & Business Problem:

Problem Background:
The City of New York is one of the most populous city in the United States. It is a hub of business and commerce like banking, finance, world trade, retailing, tourism, fashion, real estate etc. It has lot of opportunities. Many fortune 500 companies are headquartered in New York City. High technology startup companies and employment are growing in New York City.

New York city has diversified food culture with a range of international cuisines. It ranges from bagels, cheesecake, hot dogs, knishes, italian pizza, italian cuisine, sandwiches, coffee hourses, kebabs, street food and delicatessens of the city. As the city is highly diversified with pople and culture, location of the restuarent is key to success of the business. So, here the analyses helps in understanding of the business environment, reduction of risk and having a good return on investment.

Problem Description:
A restaurant is a business which prepares and serves food and drinks to customers in exchange for money. Restaurants vary greatly in appearence and offerings, including in a wide variety of cuisine and service models ranging from inexpensive food restaurants and cafeterias to mid-priced family restaurants to high priced luxury establishments.

So, considering all the diversified cultures and food, choosing the right place for restaurant is the key to business success. The key factors to be considered while placing a restarant.
1. New york population distribution.
2. Population demographics like race, ethnicity, religion, wealth and income disparity.
3. Tourism (Places where there will be high floating population).
4. Transportation (Taxis, rails and buses)
5. Who are the other competetors in the location.
6. How far is the place to buy all the materials required to prepare food etc.

So, the objective of the project is to determine the location to start a new restaurent and understand the rationale behind choosing that location. The success of the project will be a recommendation of good neighborhood choice based on lack of restaurants in that location and nearest suppliers of ingredients.

2. Data:

Data: New York city will be analyzed using the following datasets.
Data 1: New York city has 5 boroughs- Brookyln, Queens, Manhattan, The Bronx and StatenIsland and 306 neighborhoods. In order to cluster the neighborhoods we need information of all the 5 boroughs and the neighborhoods present in each borough along with latitude and longtitude coordinates of each neighborhood.
Data is available for free at: https://geo.nyu.edu/catalog/nyu_2451_34572

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.

Data 3: For the below analysis we will get data from wikipedia as given below :

New York Population New York City Demographics Cuisine of New York city https://en.wikipedia.org/wiki/New_York_Cityhttps://en.wikipedia.org/wiki/Economy_of_New_York_City https://en.wikipedia.org/wiki/Portal:New_York_Cityhttps://en.wikipedia.org/wiki/Cuisine_of_New_York_City https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City

Data 4: Newyork 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.

3.Methodology:

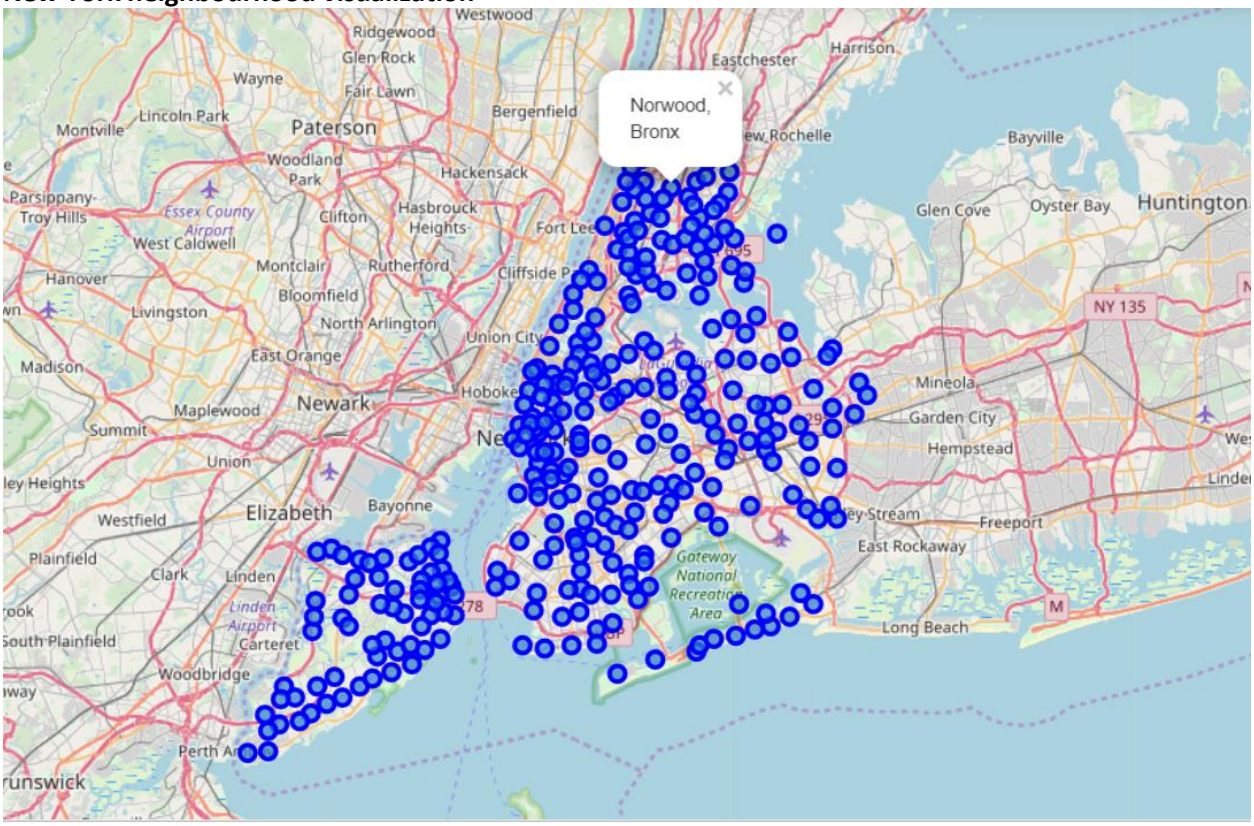
Analytic Approach:

New York city 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.

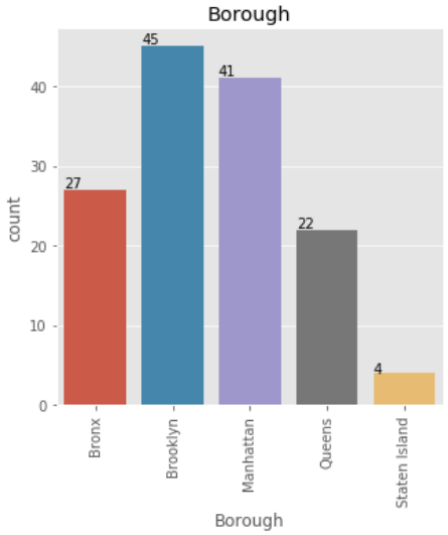
Exploratory Data Analysis:

Data 1- New york city Geographical Coordinates Data.
1. Data is loaded from newyork_data.json file.
2. Transformed the data into a pandas data frame.
3. This data frame contains the geographical coordinates of New York city neighborhoods and it will be used to get Venues data from Fouresquare.
5. We used geopy and folium libraries to create a map of New York city with neighborhoods superimposed on top.

New York neighbourhood visualization

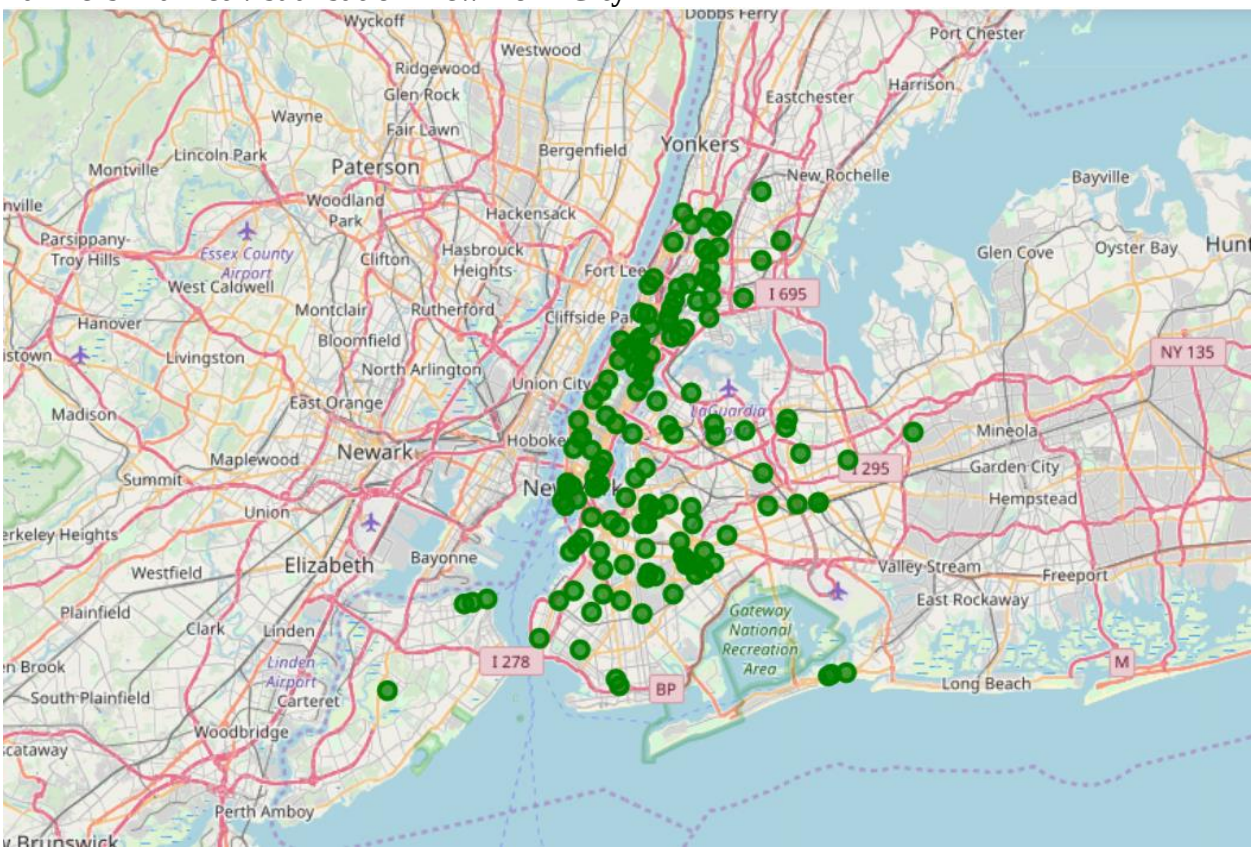


Data 2- Second data which is used is the DOHMH Farmers Markets dataset. This is used to identify the location of Farmers Markets in each neighborhood. We identified a total of 139 Farmers Markets in New York city. Highest number are in Brooklyn and Manhattan followed by Bronx, Queens and Staten Island.



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

Farmers Market visualisation-New York City



Data 3: New York city Population, Demographics and Cuisine data was scrapped from Wikipedia pages given above in the data section. We used BeautifulSoup python library.

1.New York Population: Insights from the data:

	Borough	County	Estimate_2017	square_miles	square_km	persons_sq_mi	persons_sq_km
0	The Bronx	Bronx	1,471,160	42.10	109.04	34,653	13,231
1	Brooklyn	Kings	2,648,771	70.82	183.42	37,137	14,649
2	Manhattan	New York	1,664,727	22.83	59.13	72,033	27,826
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	806.863	783.83	28,188	10,947
6		State of New York	19,849,399	1,547.116	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.

The racial composition is as given below. It has diverse population ranging from White's, African's, Hispanic's, Non-Hispanic's and Asian's. This also increases the scope for restaurants business in New York City.

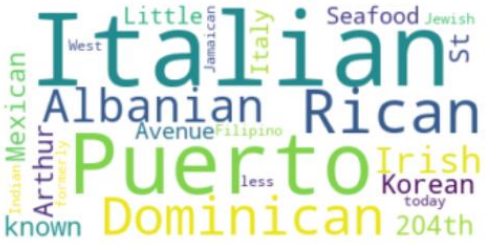
	Racialcomposition	2010	1990	1970	1940
0	White	44.0%	52.3%	76.6%	93.6%
1	—Non-Hispanic	33.3%	43.2%	62.9%	92.0%
2	Black or African American	25.5%	28.7%	21.1%	6.1%
3	Hispanic or Latino (of any race)	28.6%	24.4%	16.2%	1.6%
4	Asian	12.7%	7.0%	1.2%	–

3.Cuisine of New York city: This data is prepared manually. Data source is - https://en.wikipedia.org/wiki/Cuisine_of_New_York_City which served as the basis of word cloud.

NEW YORK CITY CUISINE: Most Preferred Food in New York City –Italian, Purto Rican.

BROOKLYN CUISINE -Most Preferred Food in Brooklyn is –Italian & Purto Rican.
MANHATTAN CUISINE - Most Preferred Food in Manhattan is – Italian.
QUEENS CUISINE - Most Preferred Food in Queens is – Indian.
THE BRONX CUISINE - Most Preferred Food in The Bronx is – Italian.

Borough: The Bronx



<Figure size 2160x3240 with 0 Axes>

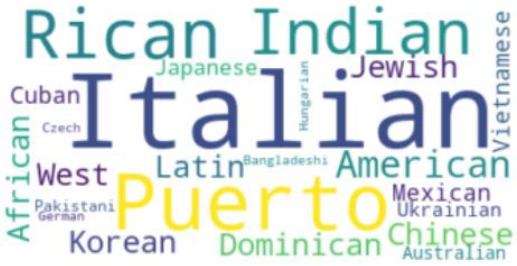
Borough: Brooklyn

<Figure size 2160x3240 with 0 Axes>

Borough: Queens

Borough: Staten Island

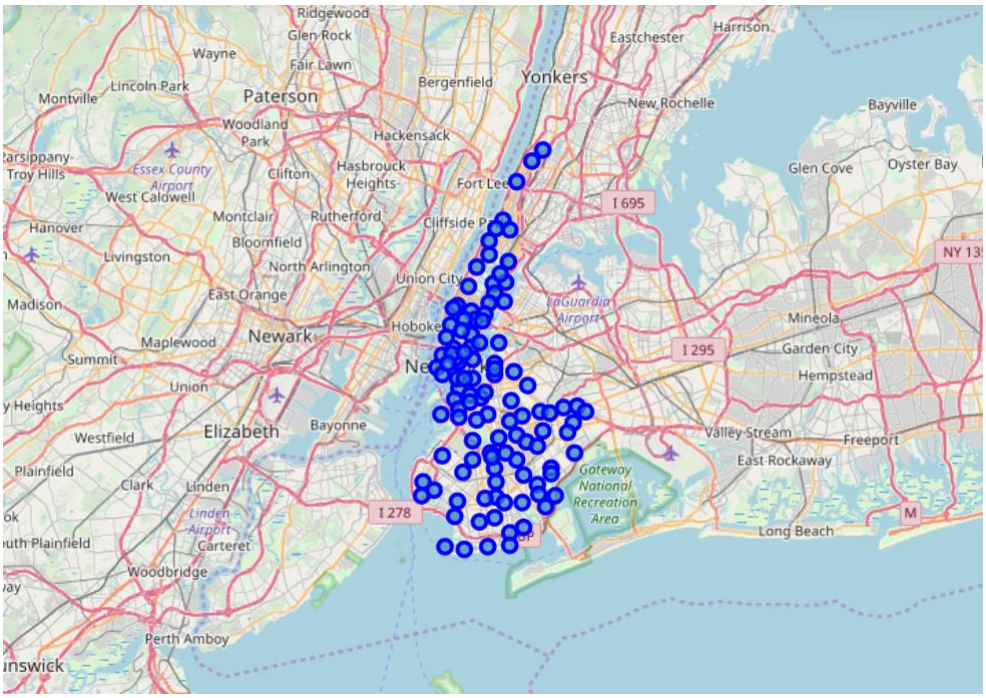
Borough: Manhattan



Data 4: NewYork city geographical coordinates are used with Foursquare API, to get venues information for each neighborhood.

Brooklyn and Manhattan :

Brooklyn and Manhattan Visualization :



Using foursquare API, top 200 venues in 1000 meters radius were identified.

Brooklyn and Manhattan Venues :

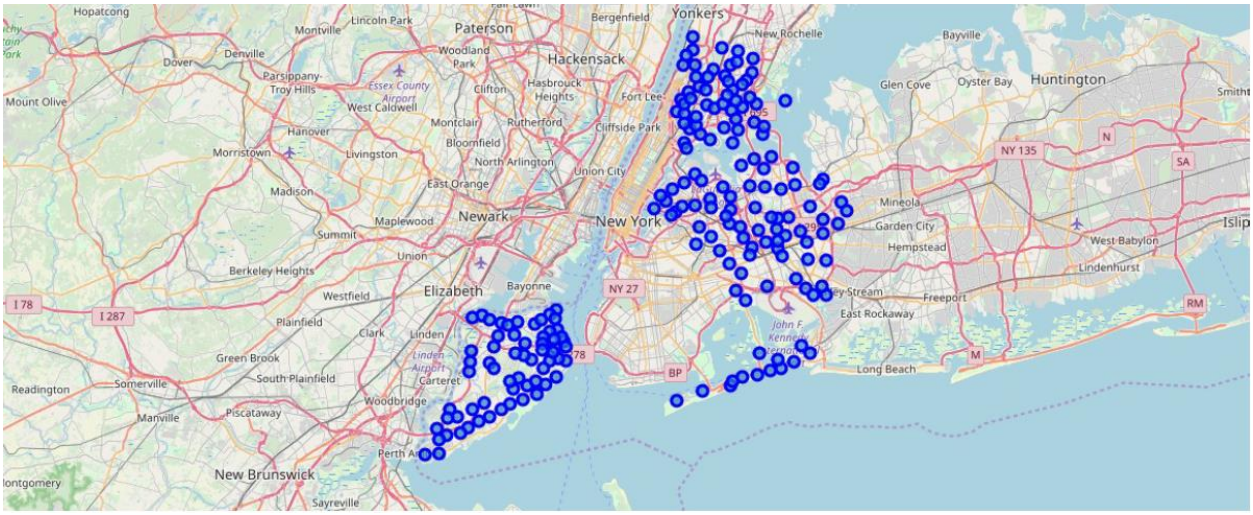
The venues are as below :

	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	Starbucks	40.877531	-73.905582	Coffee Shop

The "BM_venues" dataframe has 9745 venues and 409 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 11086 venues and 390 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	Ali's Roti Shop	40.894036	-73.856935	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Jackie's West Indian Bakery	40.889283	-73.843310	Caribbean Restaurant
4	Wakefield	40.894705	-73.847201	Rite Aid	40.896649	-73.844846	Pharmacy

4. RESULTS:

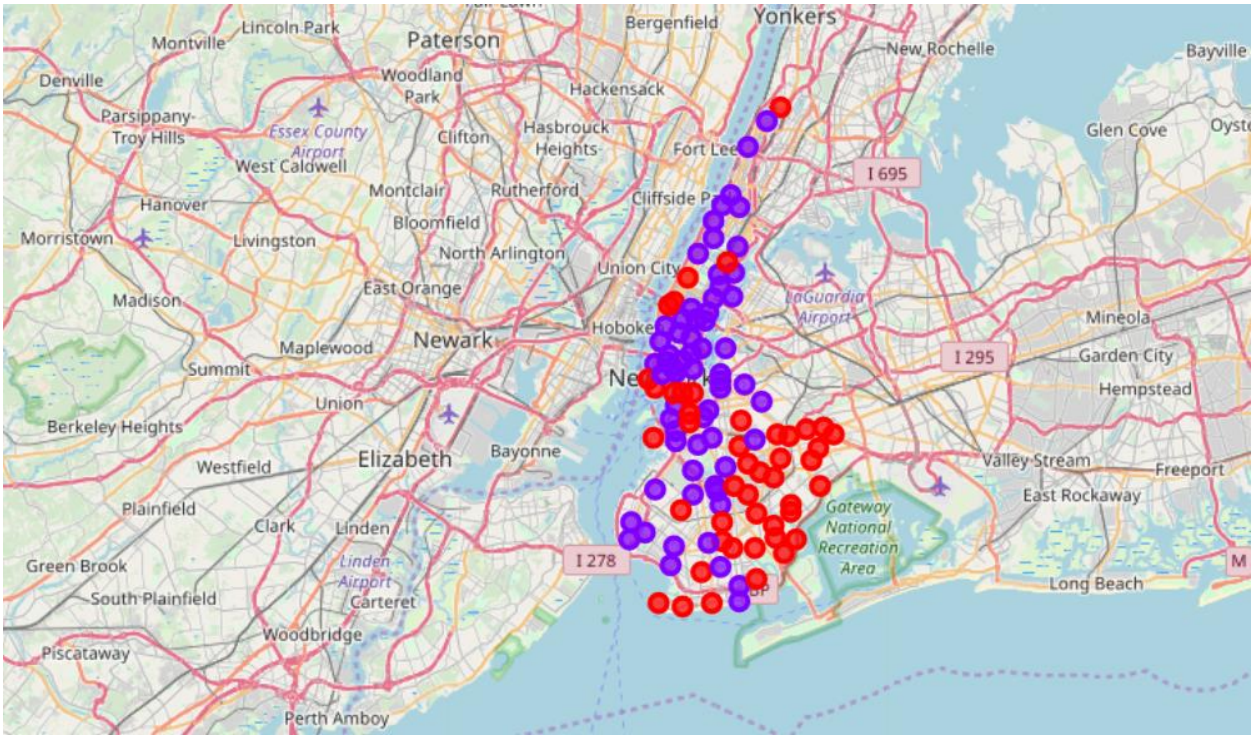
We filtered and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens and Staten Island clustering from this venues data.

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.

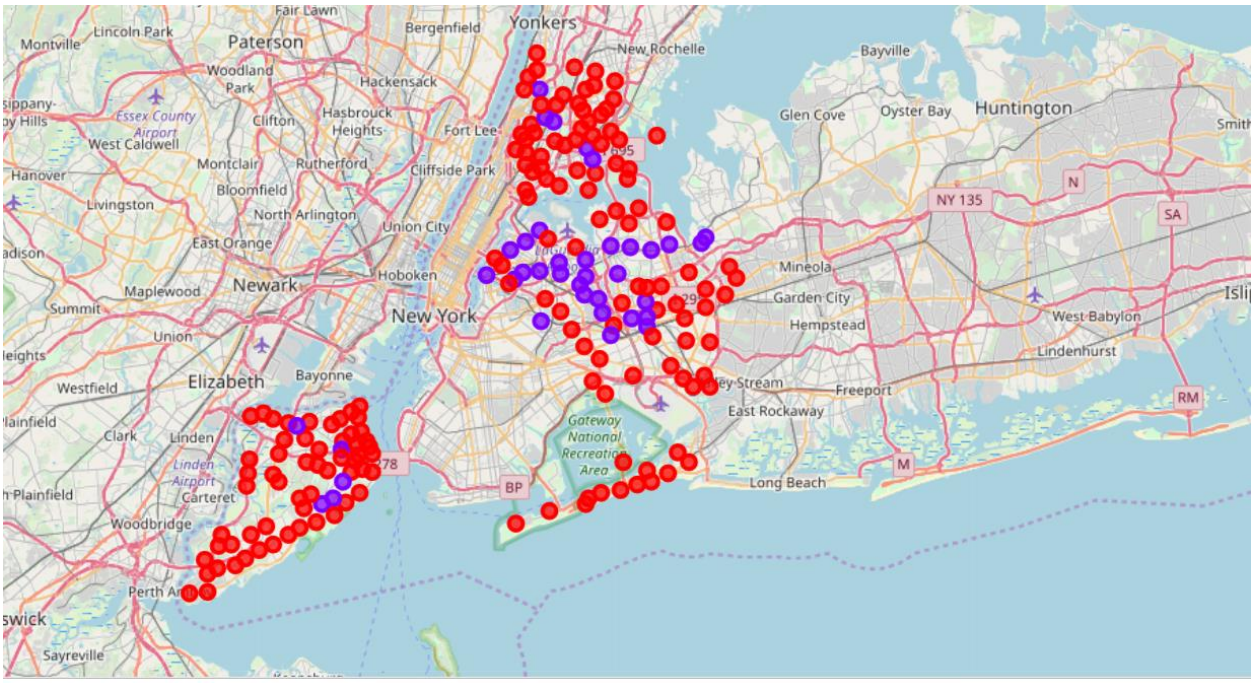
Brooklyn & Manhattan:

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



Cluster 0: The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.
Cluster 1: 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.



Cluster 0: The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated. There are untapped neighborhoods.
Cluster 1: 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. Farmers markets could be increased in Bronx, Queens and Staten Island.
2. Various countries cuisines can be explored in Bronx, Queens and Staten Island.
3. Wide variety of cuisines from different countries are available in Manhattan and Brooklyn. A restaurant with great cuisines would be a good choice.

6.CONCLUSION:

Limited data is available to perform the analysis. Results could be more robust if large amount of data is present. More number of restaurants imply wide business demand.
Brooklyn and Manhattan have very competitive market with high concentration of restaurant business. Bronx, Queens and Staten Island is not very competitive but has good number of restaurants. So, this can be explored. From the neighbourhood data or restaurant types identified like Indian Restaurants can be checked. A venue with low competition can be of low risk.