The Battle of the Neighborhoods - Week 1
Introduction & Business Problem:
Problem Background:

The City of Modern York is the foremost swarmed city within the United States. It is distinctive and is the cash related capital of USA. It is multicultural. It gives allocate of commerce openings and commerce pleasant condition. It has pulled in a wide extend of players into the advertise. It may be an around the world center point of trade and exchange. The city could be a vital put for keeping money and money, retailing, world trade, transportation, the travel industry, arrive, unused media, standard media, advancing, genuine organizations, bookkeeping, security, theater, fashion, and expressions of the human involvement within the Joined together States.

This moreover implies that the showcase is exceedingly competitive. Because it is profoundly created city so cost of doing commerce is additionally one of the highest. Thus, any modern trade wander or extension must be dissected carefully. The bits of knowledge determined from examination will deliver great understanding of the trade environment which offers assistance in deliberately focusing on the showcase. This will offer assistance in diminishment of chance. And the Return on Speculation will be sensible.

Problem Description:

A restaurant is a business that helps prepare and provides drinks and snacks to clients in exchange for cash, either paid until meal, after plate of food, or with an accessible account. New York City is renowned for its excellent cuisine. Its cuisine includes a variety of worldwide cuisines that are impacted by the immigrant history of the place.

1. Foreigners from Central and Eastern Europe, particularly Jewish immigrants-bagels, cheesecake, hot dogs, knives and delicacies
2. Italian immigrants-pizza in New York and Spanish cuisine
3. Italian pizza for immigrants in New York and Spanish cuisine Chinese and other Asian restaurants, sandwich joints, tropical gardens, dinners and coffee shops all over the city
4. Mobile food suppliers-Approximately 4,000 city licensed Examples of modern New York street food such as shawarma and kebabs
5. It is popular not only for Pizzerias, Cafe's, but also for Michelin starred restaurants. According to Michelin, its city is home to nearly one thousand of the world's finest and most diverse gourmet dinners.
6. So it is clear that to outlive in such competitive advertise it is very imperative to strategically arrange.

Different components have to be examined in order to choose on the Area such as:
1. New York Population New York City Demographics
2. Are there any Farmers Markets, Wholesale markets and more nearby so that the ingredients can be purchased fresh to maintain quality and cost?
3. Are there any venues like Gyms, Entertainment zones, Parks etc. nearby where floating population is high Who are the competitors in that location?
4. Cuisine served / Menu of the competitors
5. Segmentation of the Borough
6. Untapped markets saturated markets etc.
7. The list can go on. Even though well supported Company Ltd.
8. Got to choose the proper area to begin it's to begin with venture.

If this can be fruitful they can duplicate the same in other areas. To begin with move is exceptionally vital; in this manner choice of area is exceptionally imperative

Target Audience:

To suggest the right area, Company Ltd has designated me to lead of the Information Science group. The objective is to find and prescribe to the administration which neighborhood of Modern York city will be best choice to begin an eatery. The Administration moreover anticipates getting it the method of reasoning of the suggestions made.

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: First 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

A farmer's market is often defined as a public site used to sell agricultural products directly to the public by two or even more local and regional distributors. Other than fresh fruits and vegetables, marketplaces can sell milk products, fish, meat, baked goods, as well as other food products which are minimally processed.

Data set is available with the link below: https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2

	FacilityName	Service Category	Service_Type	Address	Address 2	Borough	ZipCode	Latitude	Longitude	AdditionalInfo	StartDate	EndDate	Monday	Tuesday	Wednesday	Thursday
0	Inwood Park Greenmarket	Farmers Markets and Food Boxes	Farmers Markets	Isham St bet Seaman & Cooper	NaN	Manhattan	10034	40.869009	-73.920320	Open year-round	NaN	NaN	NaN	NaN	NaN	NaN
1	82nd Street Greenmarket	Farmers Markets and Food Boxes	Farmers Markets	82nd St bet 1st & York Aves	NaN	Manhattan	10028	40.773448	-73.948954	Open year-round	NaN	NaN	NaN	NaN	NaN	NaN
3	125th Street Farmers Market	Farmers Markets and Food Boxes	Farmers Markets	125th St & Adam Clayton Powell Jr Blvd	NaN	Manhattan	10027	40.808981	-73.948327	Market open dates: 6/13/2017 to 11/21/2017	06/13/2017	11/21/2017	NaN	10am-7pm	NaN	NaN
4	170 Farm Stand	Farmers Markets and Food Boxes	Farmers Markets	170th St & Townsend Ave	NaN	Bronx	10452	40.840095	-73.916827	Market open dates: 7/5/2017 to 11/22/2017	07/05/2017	11/22/2017	NaN	NaN	2:30pm-6:30pm	NaN
5	175th Street Greenmarket	Farmers Markets and Food Boxes	Farmers Markets	175th St bet Wadsworth Ave & Broadway	NaN	Manhattan	10033	40.845956	-73.937813	Market open dates: 6/29/2017 to 11/30/2017	06/29/2017	11/30/2017	NaN	NaN	NaN	8am-5pm

Data 2: **The district has a total of 5 boroughs and 306 neighborhoods to section and examine the neighborhoods,** We will fundamentally need a set of data containing the 5 boroughs and neighborhood that exist in each borough as well as the longitude and latitude coordinate of each neighborhood.

Data set is available with the link below:

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 3: Wikipedia details discussing:

- New York Population

- New York City Demographics

- Cuisine of New york

Data set is available with the link below:

https://en.wikipedia.org/wiki/Cuisine_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/New_York_City

Data 4: Coordinated data from Foursquare.com Modern New York City Geographical data will be used as input for the Foursquare API, which will be used to provide data on places for each neighborhood. We're going to use the Foursquare API to discover New York City neighborhoods. Under the Foursquare API data picture.

|	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

Methodology :

Business Understanding:

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

Analytic Approach:

- First Step: Clustering of Manhattan and Brooklyn

- Second Step: Clustering of Bronx, Queens and Staten Island

- Only Restaurant data is filtered from foursquare.com venues data

- New York city District has a total of 5 boroughs and 306 neighborhood

- and utilized for this project

- This is done because of the following exploratory data analysis

Exploratory Data Analysis:

- Data 1- Data 1- DOHMH Farmers Markets and Food Boxes data set

- We will be using the data of Farmers Markets

- 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
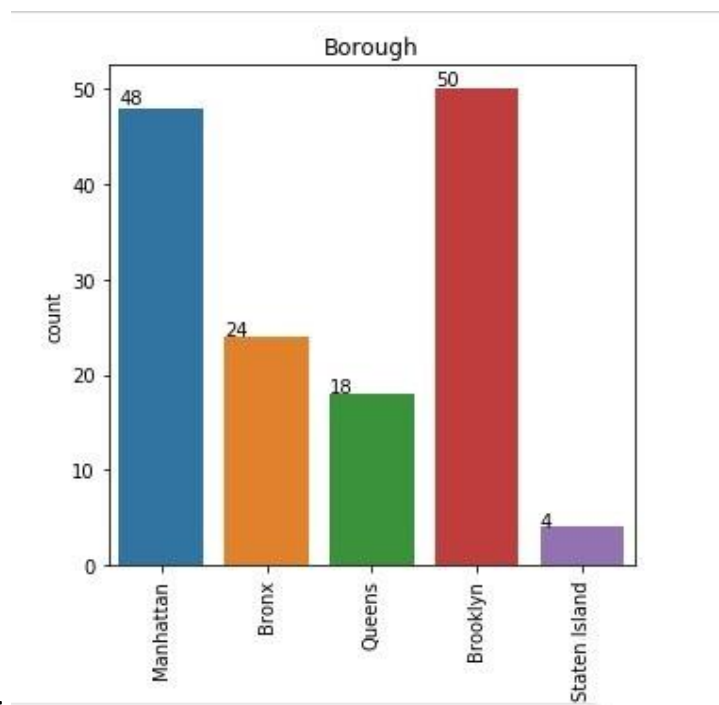

Artifacts:

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 2- **Data 2: New York City Geographical Coordinates Data.**

- Load the data and examine data- newyork_data.json

- Transform the data of nested python dictionaries into a pandas data frame

- Data frame contains the geographical coordinates of New York city neighborhoods.

- Data will used to get Venues data from Foursquare

- Geopy and folium libraries used to create map of New York city with neighborhoods superimposed on top

  New York neighborhood visualization image Below:
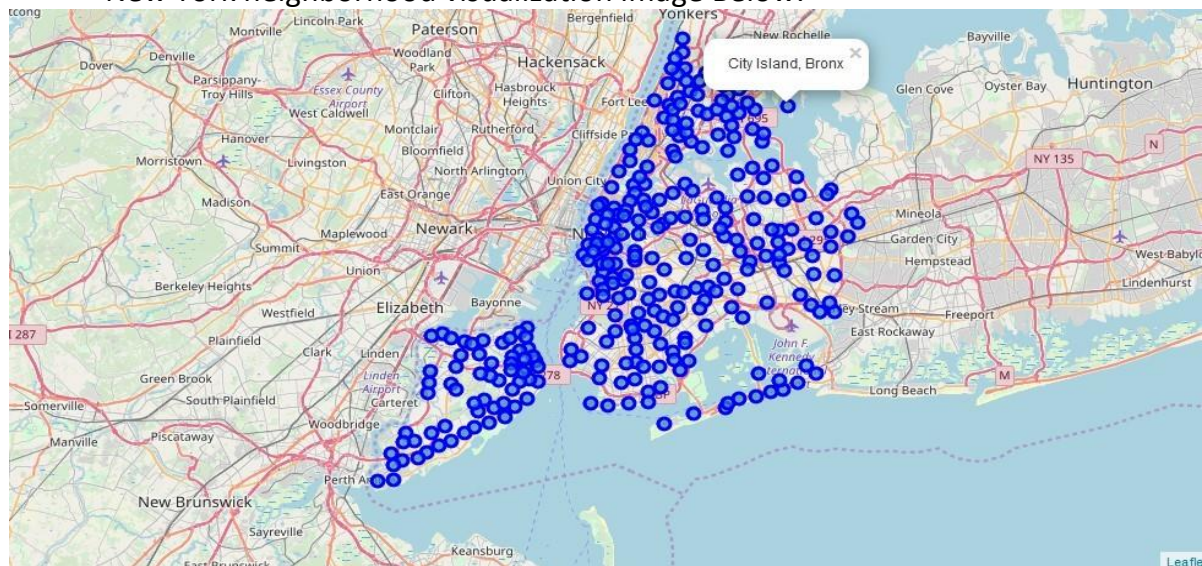
Data 3 : Data 3: The data from Wikipedia pages actually given above in the data section was completely scrapped to collect and analyze new York City entire populations, changing demographics, different cuisines.

- We used BeautifulSoup python library BeautifulSoup is a python package for parsing HTML and XML documents (including the malformed huge markups, tags that are not closed, so name after tag soup).

- It creates a parsing tree for parsing pages to extract data

1.New York Population

- Manhattan (New York County) is the smallest and most densely populated

- Manhattan (New York County) with a population density of 72,033 people per square mile (27,812/Km2) in 2015 continuing to make it the highest in any county in the United States and greater than any single American city density.

- Brooklyn (kings County), on the western tip of long island, is the 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 Demographics

- New York City is the most populous city in the United States, with an estimated high of 8,622,698 residents as of 2017,
- incorporating more immigration into the city than emigration since the 2010 U.S. census
- The racial composition shows below.
- This is why New York City has restaurants serving cooking from many countries like India, Africa, Japan, China, and so on. 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- New York Cuisines

- Data has been manually prepared. Data is taken in Wikipedia page

  https://en.wikipedia.org/wiki/Cuisine_of_New_York_City

- Data used to create Word Cloud

New York City Cuisine: Most preferred Food in New York City-Italian, Puerto Rican, Mexican, Jewish, Indian, Pakistani & Dominican
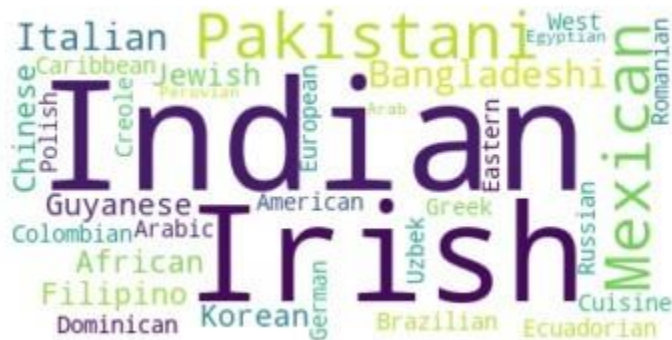
Brooklyn, Cuisine- Most Preferred food in Brooklyn are Italian, Puerto Rican and Mexican

Manhattan Cuisine- Most proffered food are 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.
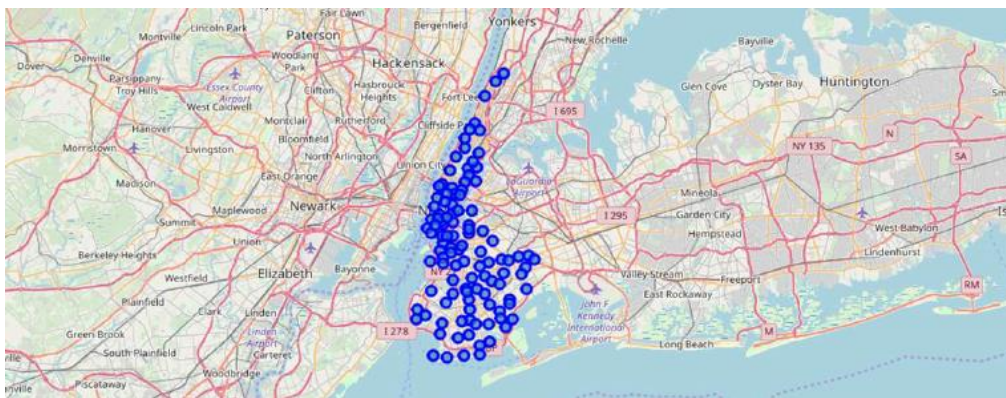

There is very less data of cuisine relating to Staten Island. So could not develop word cloud with it.

Data 4: Foursquare.com

- New York city Geographical coordinates data has been utilized as input of Foursquare API, that has been leveraged to provision venues information for each neighborhood.

- We used the Foursquare API data to examine neighborhood in New York City

- Using the geographical coordinates of each neighborhood foursquare API calls are made to get top 200 venues in radius of 1000 meters

- Part1: Brooklyn and Manhattan

- Part2: Bronx, Queens and Staten Island

Part 1: Brooklyn and Manhattan Venues data
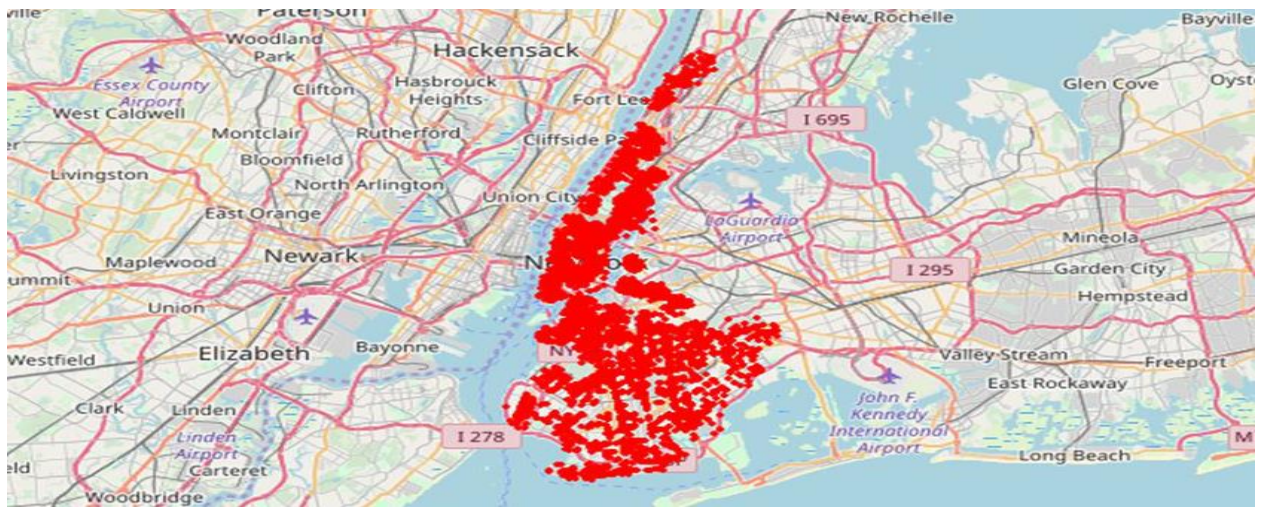
Brooklyn and Manhattan Visualization

Brooklyn and Manhattan Venues data

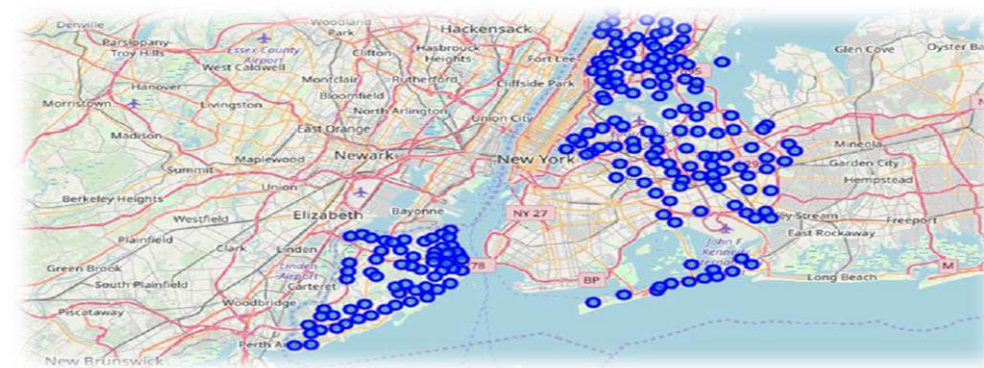|	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett 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

Brooklyn and Manhattan Venues Visualization: Generated below Brooklyn and Manhattan Venues Visualization,

The"BM_venues" data frame has 9708 venues and 397 unique venues types
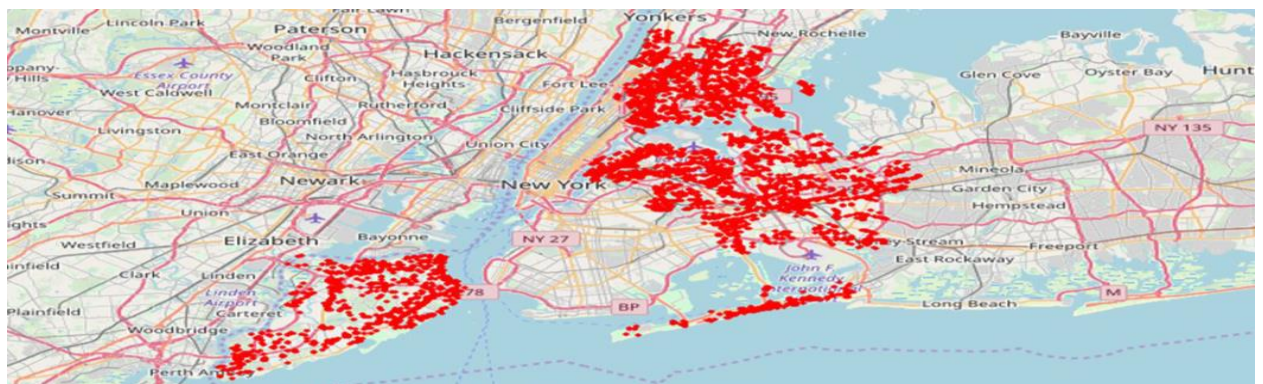
Part 2: Bronx, Queens and Staten Island

Bronx, Queens and Staten Island Visualization

Bronx, Queens and Staten Island Venues data

	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 Visualization: The "BQS_venues" datafram has 10805 venues and 387 unique venue types.

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.

Part1- Brooklyn and Manhattan

- Segmenting and Clustering Neighborhoods

- Cluster 0: the Total and Total Sum of Cluster 0 has smallest value. It shows that the Market is not Saturated

- Cluster 1: Total and Total Sum is very High. Lot of Competition. Saturated Neighborhoods
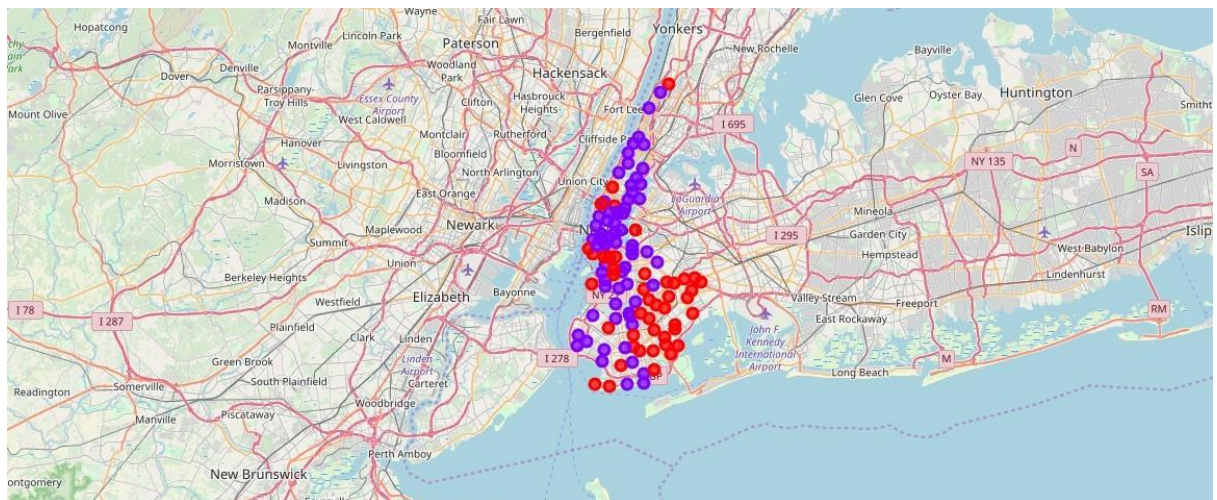
Part2- Bronx, Queens and Staten Island

- Segmenting and Clustering Neighborhoods

- Cluster 0: the Total and Total Sum of Cluster 0 has smallest value. It shows that the Market is not saturated. There are potentially untapped Markets
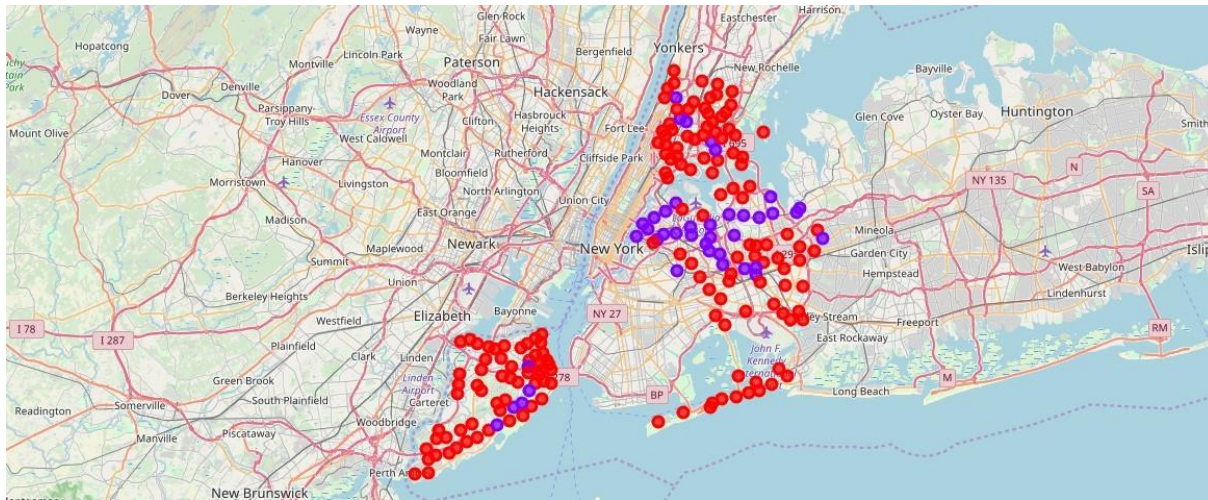
- Cluster 1: Total and Total Sum is very High. Lot of Competition. Saturated Neighborhoods

	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

DISCUSSION:

- Scope to examine cuisines of various countries in Bronx, Queens and Staten Island

- In Manhattan and Brooklyn restaurants, cuisines of many countries are part of their Menu. Risk can be taken with great Menu on board. It also shows people love to explore cuisines of various countries

- Scope to increase Farmers market in Bronx, Queens and Staten Island

- Region –Cluster0

- Neighborhood-Staten Island-Tod Hill, Port Ivory, Bloomfield

CONCLUSION:

- Analysis of limited data re-run program with updated information Brooklyn and Manhattan has high restaurant business concentration.

- Bronx, Queens and Staten Island also have a very competitive environment with a good variety of restaurants, but not as much as needed.

- As per the neighborhood or restaurant type mentioned as Indian results, this can be a enormous potential to start a food place.

- Analysis can be checked according to the type of neighborhood or restaurant mentioned as Indian results. It is possible to identify a place with the lowest risk and competition