The Battle of the Neighborhoods - Week 1

Introduction

Problem:

Problem Background:

The most populous city in the United States is The City of New York. It is the financial capital of USA. It provides lot of business opportunities and business friendly environment. It is very attracted into the the field of market. It's a hub of business and commerce. The city is a major centre for banking and finance and also different activities.

The city is highly developed so the cost of expense to do businesses is challenging task. Thus, we should be careful to put a new business venture or expansion which needs a good and careful Analysis. The derived analysis will give good understanding of the business environment which help in strategically targeting the market.

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.

A strategical plan is required to sustain in such a competitive business.
We need to exploit the Various factors 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 etc nearby where floating population is high
5. Identify the competitors in that location?
6. Cuisine served / Menu of the competitors
7. Segmentation of the Borough
8. Available markets
9. Drenched markets etc

The New investor need to choose the correct location to start its first venture. If this is successful they can replicate the same in other locations.

Target Audience:

In Order to identify and recommend the correct location, ABC Company Ltd has appointed me to lead of the Data Science team. The objective is to Identify, locate and recommend the management about which neighborhood of Newyork city will be better choice to open a restaurant.

This would be useful for anyone who wants to start a new restaurant in Newyork city.

Criteria:

The 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 nearby suppliers of ingredients.

Analysis on:

Newyork City: Is analysed.

The below datasets are used for analysing the Newyork city

DataSet 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 Link: https://geo.nyu.edu/catalog/nyu_2451_34572

DataSet 2 : Second data which will be used is the DOHMH Farmers Markets and Food Boxes dataset.
The below link provides data of Farmers Markets.
https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2
Website-https://www.grownyc.org/greenmarketco/foodbox

DataSet 3:

1. To Analyse New York Population we use the link :
   https://en.wikipedia.org/wiki/New_York_City

2. To Analyse New York City Demographics we use the link:
   https://en.wikipedia.org/wiki/Economy_of_New_York_City
   https://en.wikipedia.org/wiki/Portal:New_York_City

3. To Analyse the Cuisine of New York city we use the Links:
   https://en.wikipedia.org/wiki/Cuisine_of_New_York_City
   https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City

DataSet 4: .We will use the Foursquare API to explore neighborhoods in New York City. The below is image of the Foursquare API data.

	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	Rite Aid	40.896521	-73.844680	Pharmacy
2	Wakefield	40.894705	-73.847201	Cooler Runnings Jamaican Restaurant Inc	40.898283	-73.850478	Caribbean Restaurant
3	Wakefield	40.894705	-73.847201	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop
4	Wakefield	40.894705	-73.847201	Dunkin Donuts	40.890631	-73.849027	Donut Shop

Approach :

New York city neighbourhood has a total of 5 boroughs and 306 neighborhoods. The data is analysed as

1) clustering of Manhattan and Brooklyn.

2) clustering of Bronx, Queens and Staten Island.

Exploratory Data Analysis:

Identified New York city Geographical Coordinates Data.

1. Captured data from the file newyork_data.json file.

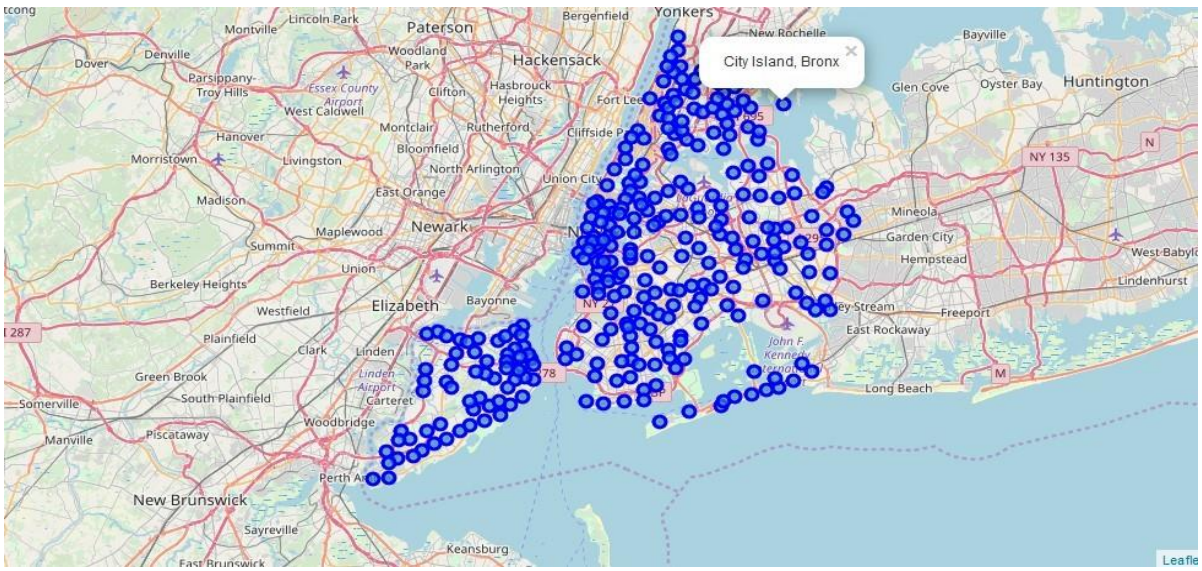
2. From the python dictionaries the data is transformed into a pandas dataframe.

3. The Output in data frame is the geographical coordinates of New York city neighborhoods.

4. Using the above data we got Venues data from Foursquare.

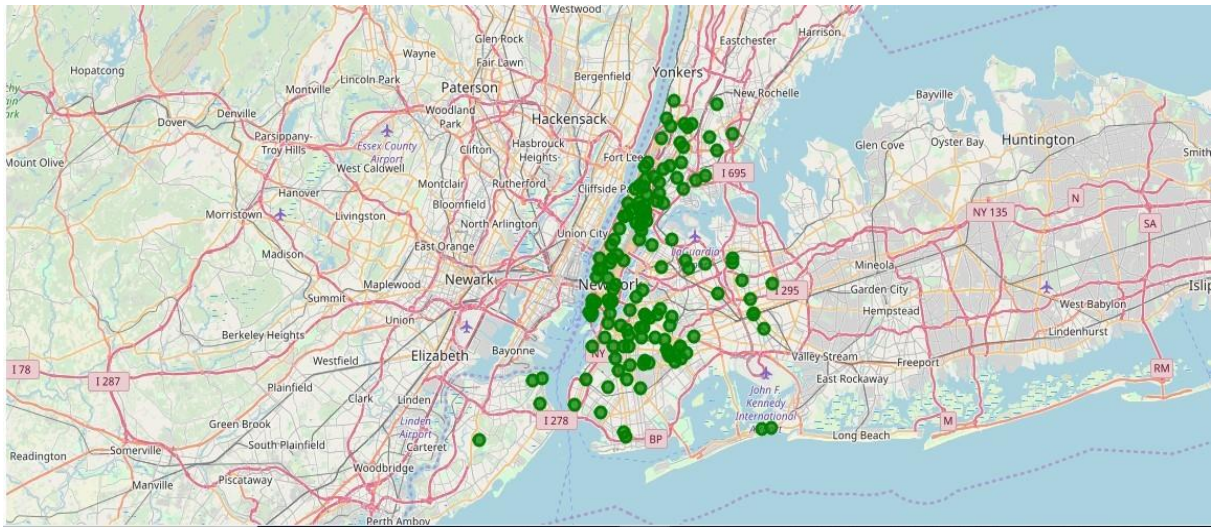
5. geopy and folium libraries are used to create a map of New York city with neighborhoods superimposed on top.

New York neighbourhood visualization:



To get the data of Farmers Markets data the DOHMH Farmers Markets and Food Boxes dataset is used. Outoff 144 Farmers Markets in New York city it was found that in Manhattan and Brooklyn these are Highest in Number whereas, in Queens, Bronx and Staten Island its was found Lowest.

Farmers Market visualisation-New York City:



Next, in order to analysis New York city Population, Demographics and Cuisine, scrapped the data from Wikipedia pages g the Beautiful Soup python library was utilized. The parsing of Xml and Html is done by using Beautiful Soup Library which creates a parse tree for parsed pages that was used to extract data from HTML, and used for web scraping.

1.New York Population : Manhattan (New York County) is the geographically smallest and most densely populated borough.

Manhattan's (New York County'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 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

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. the reason of having scope for restaurants business New York city because it has restaurants serving cuisine from many countries such as Indian, African, Japan etc.

	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%	–

1. 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 . .

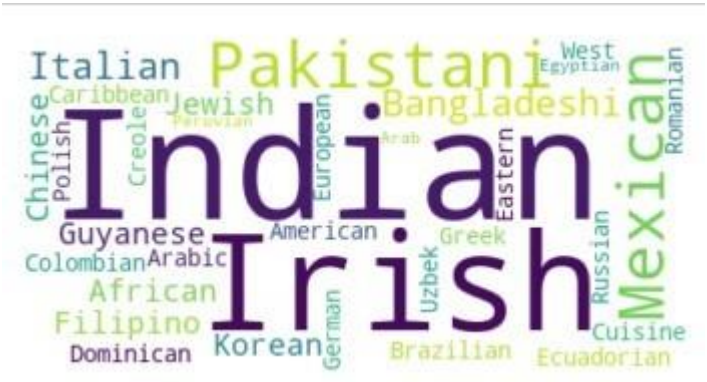
NEW YORK CITY CUISINE : Most Preferred Food in New York City –Italian, Purto Rican, Mexican, Jewish, Indian, Pakistani & Dominican.

Word cloud prepared from the data taken from https://en.wikipedia.org/wiki/Cuisine_of_New_York_City .

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.

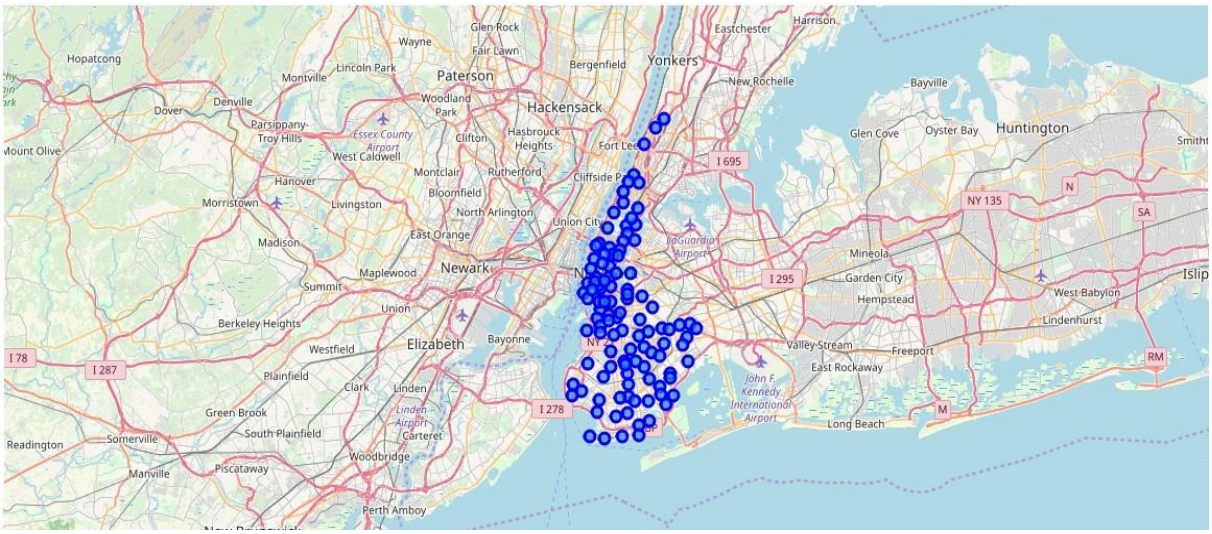
QUEENS CUISINE - Most Preferred Food in Queens is – Indian, Irish, Pakistani and Mexican.

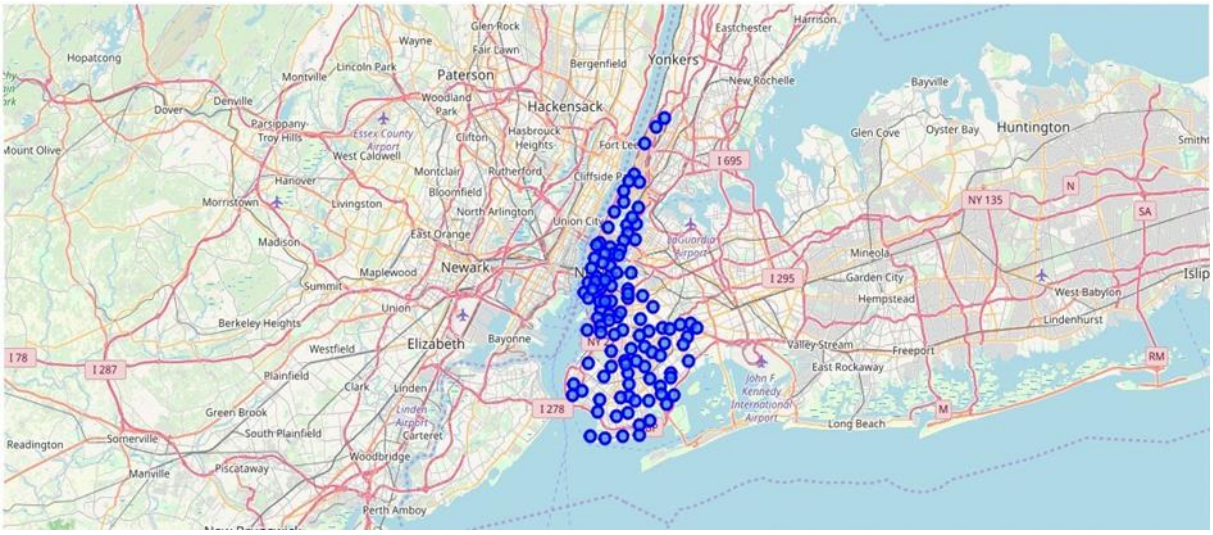


THE BRONX CUISINE - Most Preferred Food in The Bronx is – Italian,Puerto Rican, Albanian and Dominican.

used the Foursquare API data to explore neighborhoods in New York City.

Brooklyn and Manhattan Visualization :

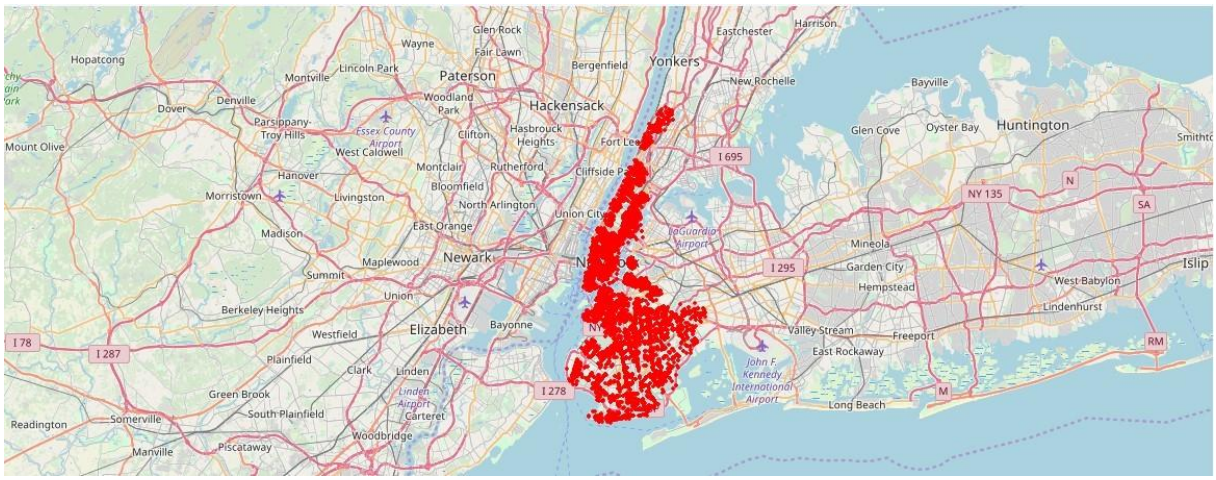




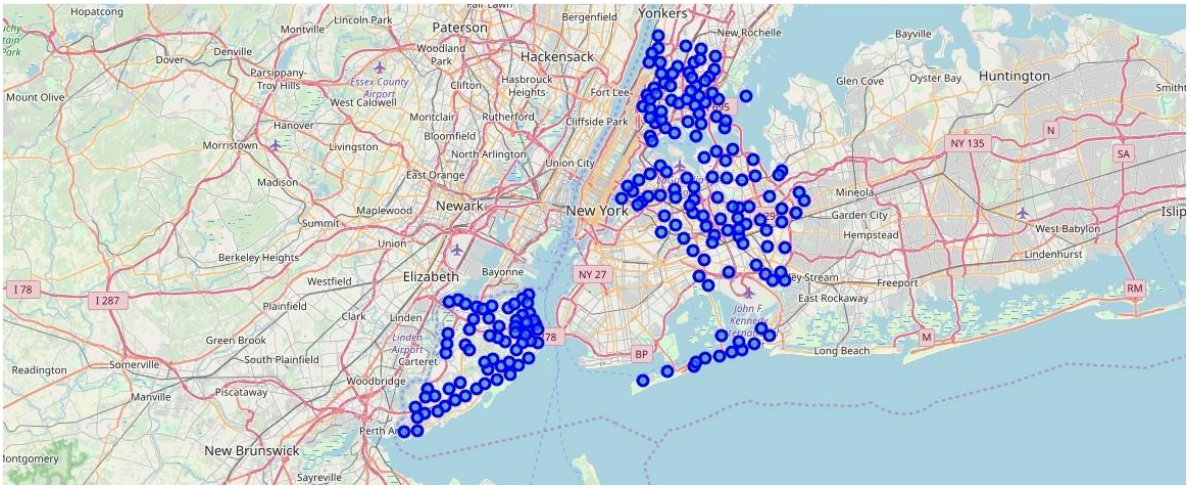
Brooklyn and Manhattan Venues :

	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 :



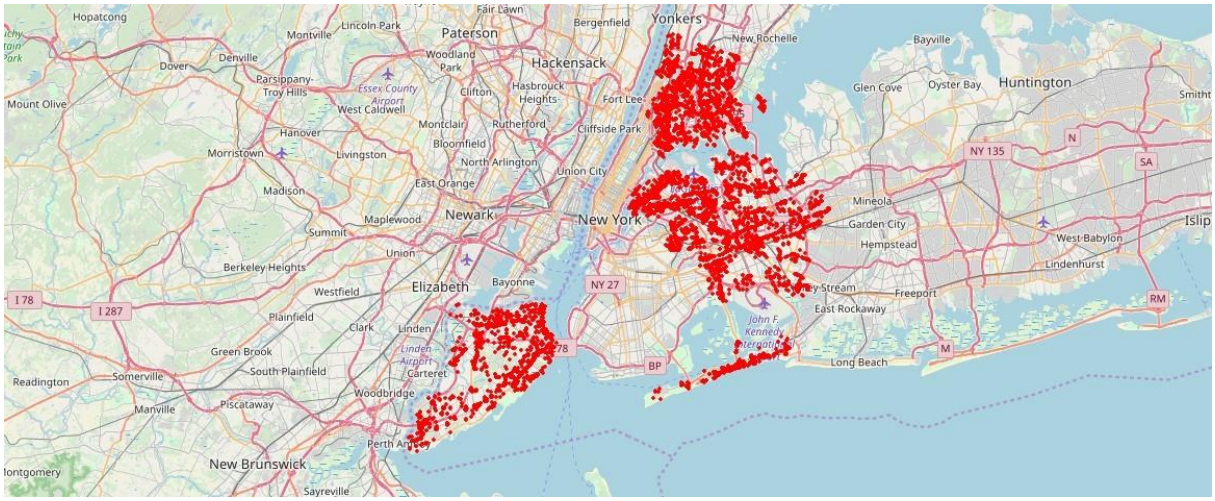
Bronx, Queens and Staten Island Neighborhoods Visualization :



Bronx, Queens and Staten Island Venues Visualization :

	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 :

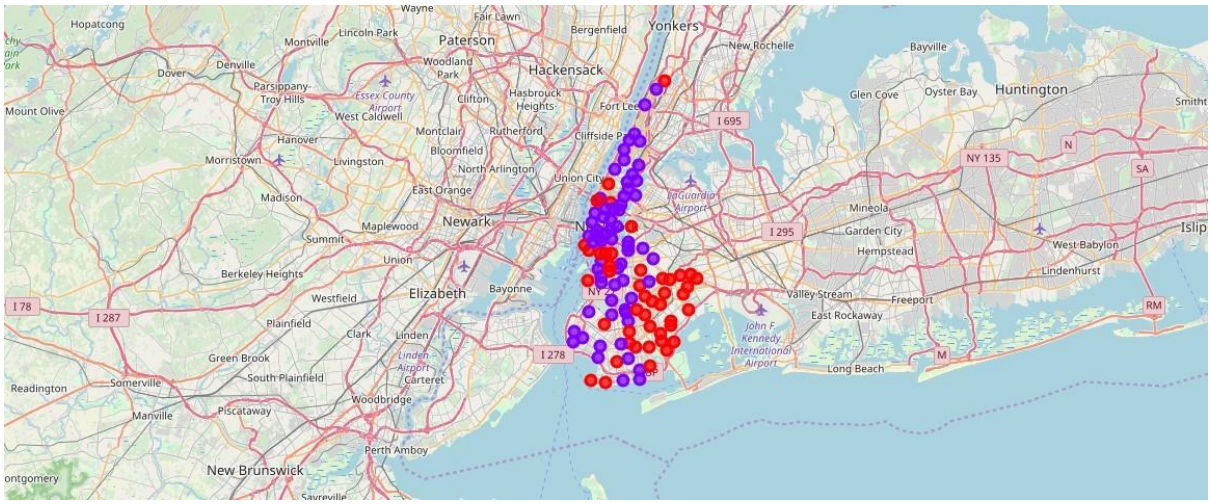


Results:

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

To cluster the neighborhoods into two clusters the K-Means clustering Algorithm is used. *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 :
Different clusters types are created by using *k*-means clustering



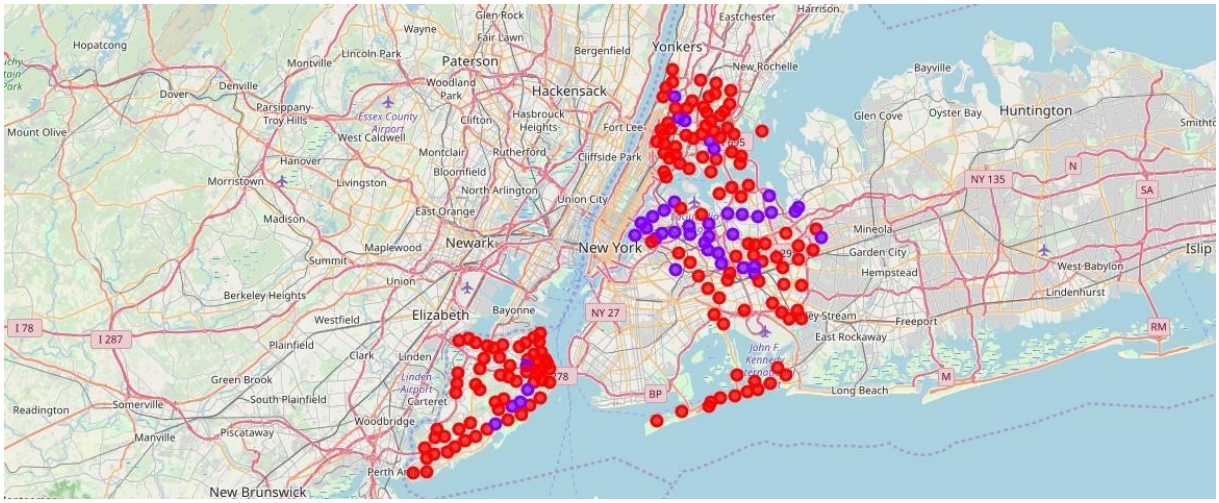
Cluster0 : The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.

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 :

The different clusters visualization



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.

DISCUSSION:

- There is scope to increase Farmers markets in Bronx, Queens and Staten Island.
- There is scope to explore cuisines of various countries in Bronx, Queens and Staten Island.
- 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.

CONCLUSION:

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.