The Battle of Neighborhoods – Report

Introduction

Problem Description

The City of New York, is the most populous city in the United States. With an estimated 2019 population of 8,336,817 distributed over about 302.6 square miles, New York City is also the most densely populated major city in the United States. 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.

Proper market analysis is very important while trying to open any new venture in the city. 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.

The City of New York is well known for its eclectic cuisine in various restaurants. 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:

- Population
- Demographics
- Access to fresh produce
- Locality etc.

The **target audience** is a restaurant chain wanting to start a venture in NYC. 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 **success criteria** of the project will be a good recommendation of borough/neighborhood choice for the restaurant in the location nearest to the suppliers.

Data

The City used for the analysis in this project is New York City (NYC).

We will need the following sets of data:

- New York 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. The data can be obtained in: 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

- New York Population, City Demographics and cuisines can be obtained from Wiki.

  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

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

- DOHMH Farmers Markets and Food Boxes dataset can be obtained from:

 https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2

Boro...	Mark...	Stree...	Latit...	Longi...	Days ...	Hour...	Seas...	Acce...	Open...	Stell...	Food ...	Loca...
Bronx	170 Farm...	E 170th S...	40.839882	-73.916783	Wednesd...	2:30pm-6...	7/15/20-1...	Yes	No			(40.8
Bronx	Bissel Gar...	Baychest...	40.90152	-73.846937	Saturday	10am-5pm	8/1/20-11...	No	No			(40.9
Bronx	Bronx Bo...	Grand Co...	40.826835	-73.922621	Tuesday	8am-4pm	6/2/20 - 1...	Yes	No			(40.8
Bronx	Bronx Fa...	1515 Sou...	40.833734	-73.889934	Thursday	11 am- 2...	Year-Rou...	Yes	Yes			(40.8
Bronx	BronxWo...	1130 Gra...	40.832331	-73.91959	Thursday	10am-4pm	07/09/20-...	Yes	No			(40.8
Bronx	Fordham ...	Park Ave ...	40.86153	-73.891063	Friday	8am-4pm	7/10/20- ...	Yes	No			(40.8
Bronx	Harvest H...	Asch Loo...	40.870249	-73.830963	Wednesd...	8am-4pm	TBD	Yes	No			(40.8
Bronx	Harvest H...	E 163rd S...	40.820986	-73.891613	Wednesd...	8am-4pm	TBD	Yes	No			(40.8
Bronx	Harvest H...	1400 Pelh...	40.857427	-73.847079	Tuesday ...	8am-4pm	TBD	Yes	No			(40.8

- Grow NYC'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 - https://www.grownyc.org/greenmarketco/foodbox

- 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	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	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop

Methodology

Our main goal is to find optimal place for a new restaurant business.

Analytical Approach:

New York has totally 5 boroughs 306 neighborhoods. In this project, the first part in clustering of Manhattan and Brooklyn and second clustering is Bronx, Queens and Staten Island. This can be done as follows.

Exploratory Data Analysis:

New York city Geographical Coordinates Data.

a. In this we load the NYC data and explore the data. This is loaded in pandas data frame.
b. The data frame contains geographical coordinates
c. This data will be used to get venue data from foursquare
d. Geopy and folium is used to create the map.

Cuisines of NYC : Most preferred cuisines in NYC are Italia, Puerto Rican, Mexican, Jewish, Indian Pakistani & Dominican.

Brooklyn Cuisine – Italian, Puerto Rican and Mexican
Manhattan Cuisine - Italian, American, Puerto Rican and Indian
Queens Cuisine – Indian, Irish, Pakistani & Mexican
Bronx Cuisine – Italian, Albanian, Puerto Rican & Dominican

Due to lack of data, Staten Island information could not be determined.

Foursquare was leveraged to get venues information and layer it on top of the NYC map.

1. Brooklyn Manhattan visualization: top 200 venues in 1000m range.
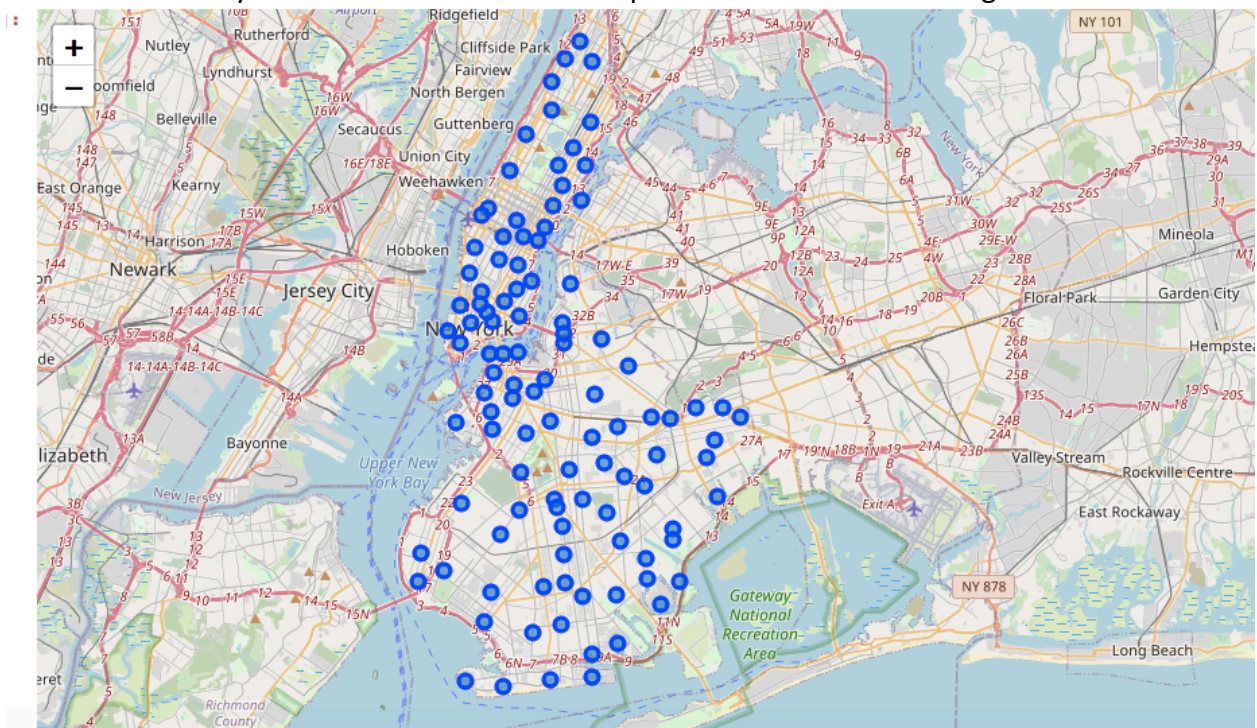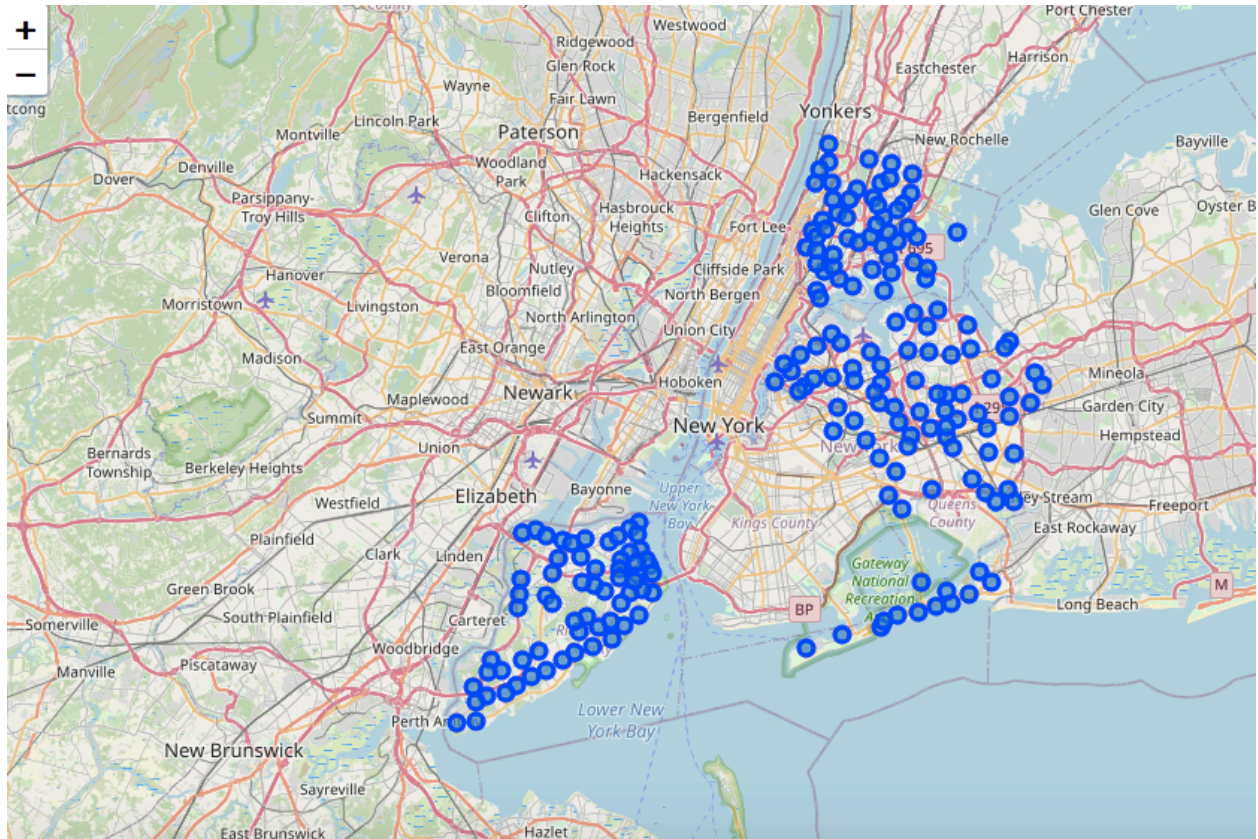

Brooklyn Manhattan venues:

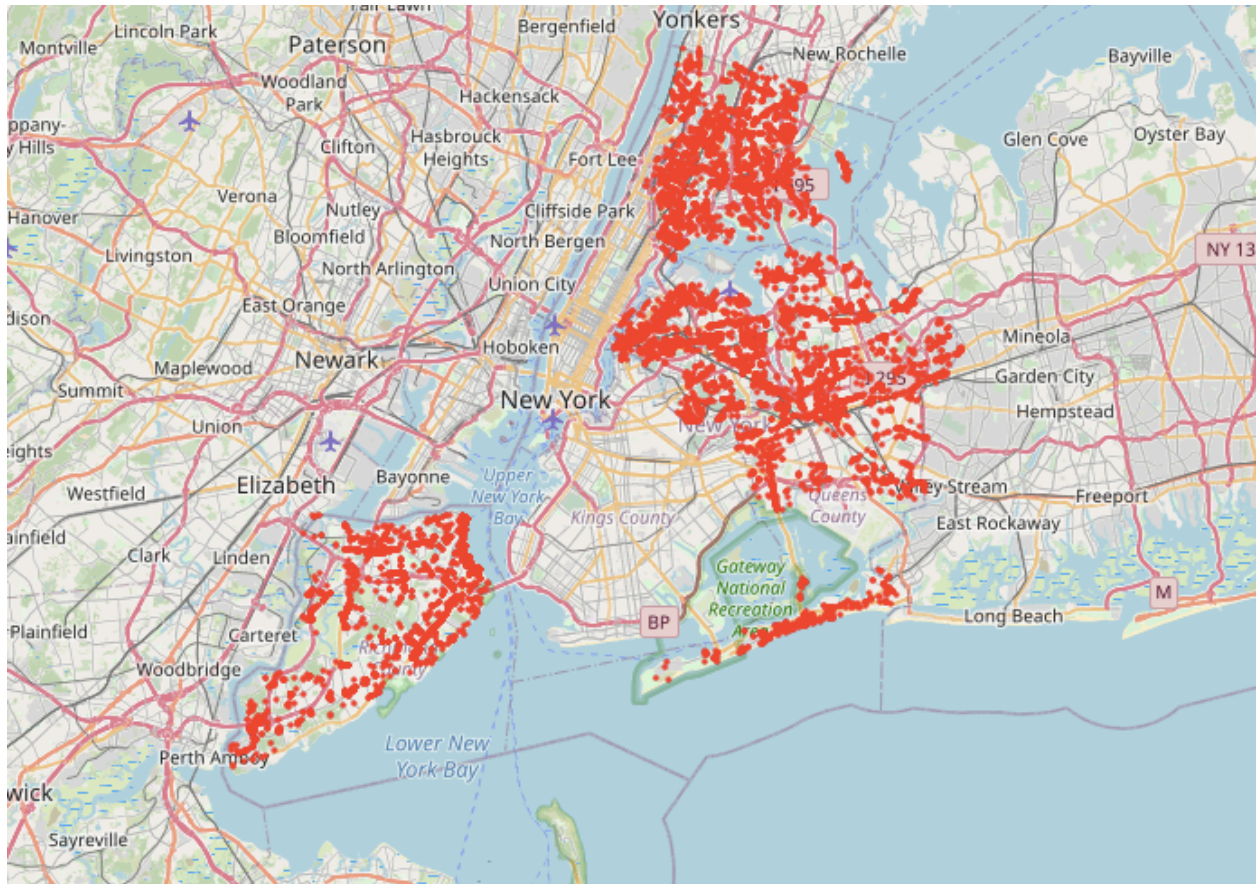|	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Sam's Pizza	40.879435	-73.905859	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett Diner	40.880404	-73.908937	Diner
3	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
4	Marble Hill	40.876551	-73.91066	Starbucks	40.877531	-73.905582	Coffee Shop
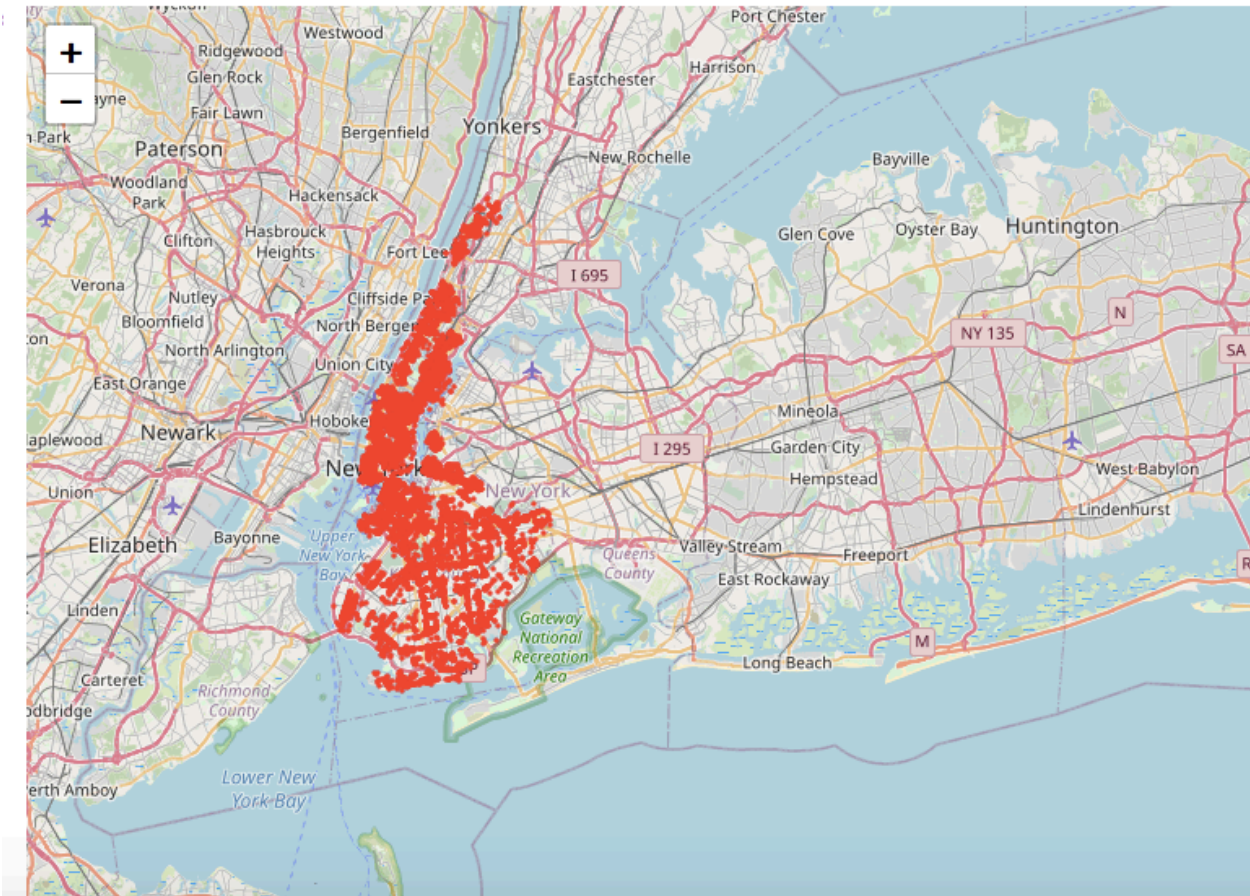
2. Bronx, Queens visualization:

Bronx and Queens Venues:

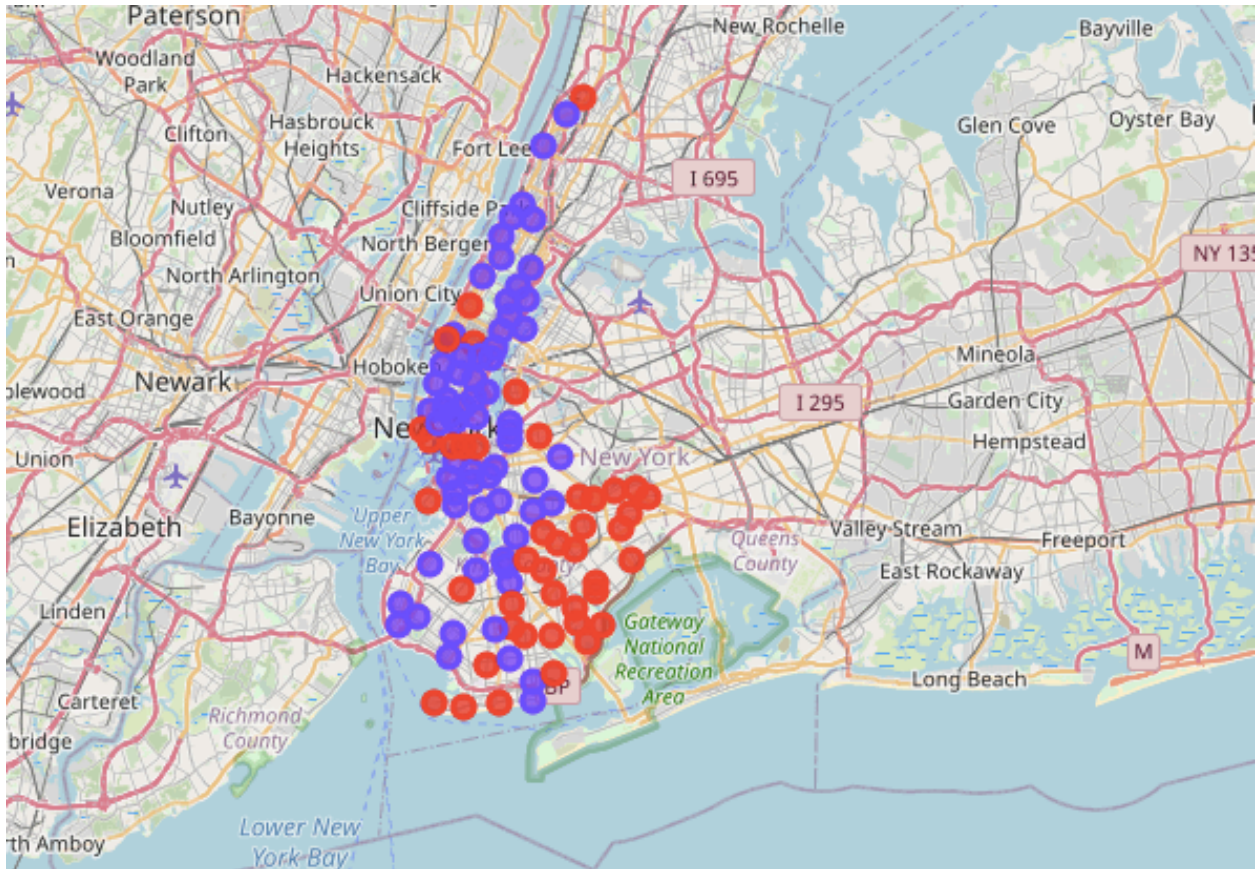|	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Wakefield	40.894705	-73.847201	Lollipops Gelato	40.894123	-73.845892	Dessert Shop
1	Wakefield	40.894705	-73.847201	Ripe Kitchen & Bar	40.898152	-73.838875	Caribbean Restaurant
2	Wakefield	40.894705	-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	Carvel Ice Cream	40.890487	-73.848568	Ice Cream Shop

Results:

From the venues data, we filtered data from the clustering and focused only on restaurants.

K- mean clustering:
K- mean clustering aims to partition *n* observations into *k* clusters in which each observation belongs to a cluster with nearest mean.
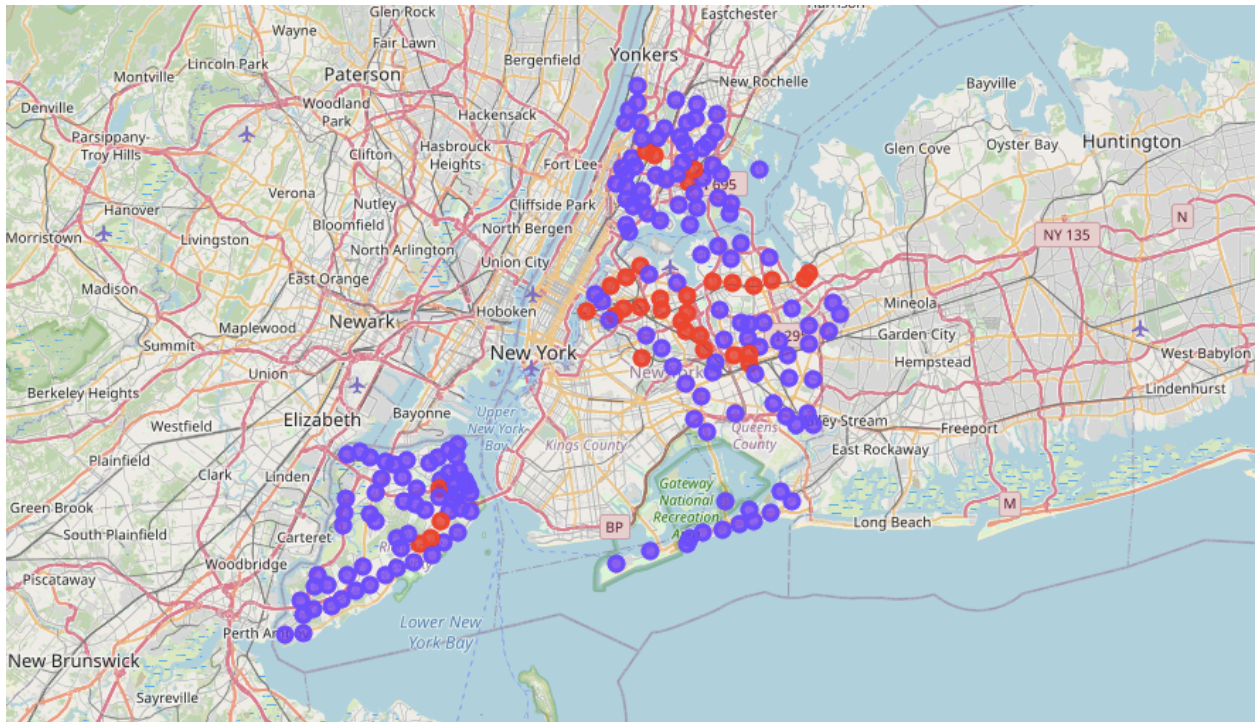
1. Brooklyn and Manhattan:

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 market is saturated.

There are no untapped neighborhoods in Brooklyn and Manhattan.

2. Bronx, queens and Staten Island:

Cluster 0 - The Total and Total Sum of cluster0 has highest value. It shows that the market is saturated.

Cluster 1 - The Total and Total Sum of cluster1 has smallest value. It shows that the market is not saturated. The untapped neighborhoods are:

	Borough	Neighborhood	Latitude	Longitude	Total	Cluster_Labels
0	Bronx	Clason Point	40.806551	-73.854144	0	1
1	Staten Island	Todt Hill	40.597069	-74.111329	0	1
2	Staten Island	South Beach	40.580247	-74.079553	0	1
3	Staten Island	Port Ivory	40.639683	-74.174645	0	1
4	Staten Island	Butler Manor	40.506082	-74.229504	0	1
5	Staten Island	Rossville	40.549404	-74.215729	0	1
6	Staten Island	Bloomfield	40.605779	-74.187256	0	1

Discussion

There is scope to increase farmers market in the Bronx, Queens and Staten Island.
There is scope to explore cuisines in Bronx, Queens and Staten Island.

Manhattan and Brooklyn people like eclectic cuisine styles.

Conclusion

From the limited data exploration, it can be concluded that there is huge demand of restaurants in NYC area. Brooklyn and Manhattan have high number of restaurants and are of various cuisines. Hence the market is extremely competitive. Bronx, Queens and Staten Island has scope of exploration.