The Battle of Neighborhood-Capstone project

Table Of Content:

1.Introduction

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 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 excelllent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history.
So it is evident that to survive in such competitive market it is very important to startegically plan. Various factors need to be studied inorder 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, Entertainmnet 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

Target Audience:

The objective is to locate and recommend to the management which neighborhood of Newyork 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 Newyork city.

2.Data

In this project we will analysed New York city.

We will use following Dataset:
1)Neighborhood has a total of 5 boroughs and 306 neighborhoods. In order to segement 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 the latitude and logitude 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

Neighborhood of Newyork city:

	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

2)We will use Foursquare API to explore neighborhood of New York city.The below image is of 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:

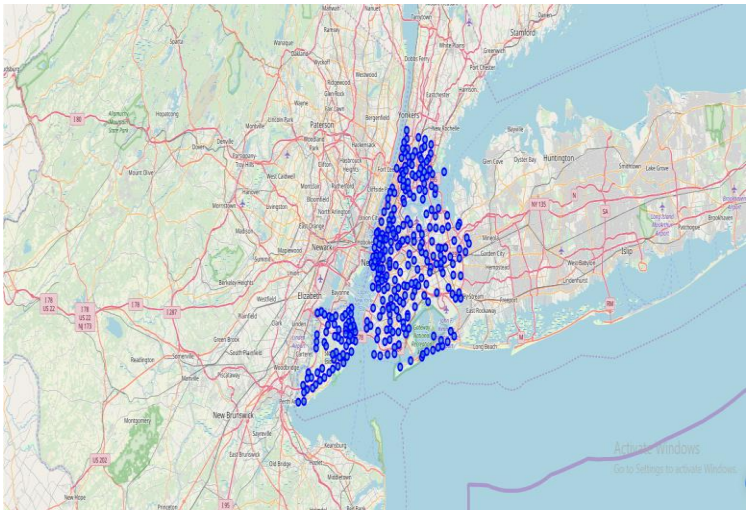
Analytic Approach:

New York city neighborhood has a total of 5 borough and 306 neighborhoods.In this project is clustering of Manhattan and Brooklyn.And second part is clustering of Bronx,Queens and Staten Islands.

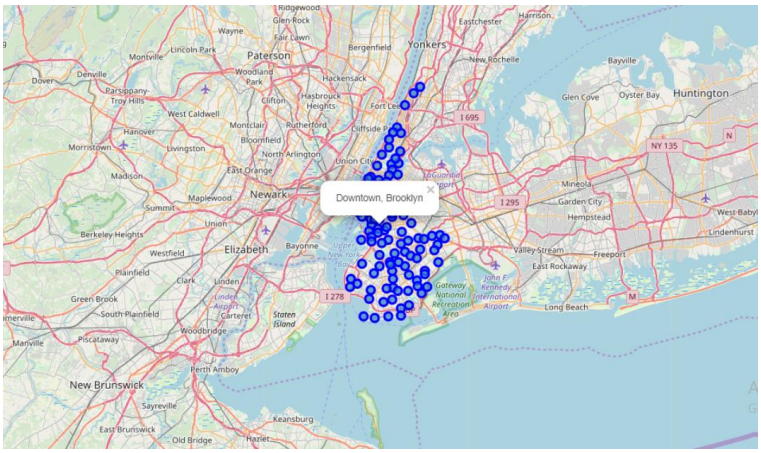
Exploratory Data Analysis:

Data 1:Geographical Data of New York city.

Visualization of Neighborhood of NewYork city:



Brooklyan and Manhattan Visualization:



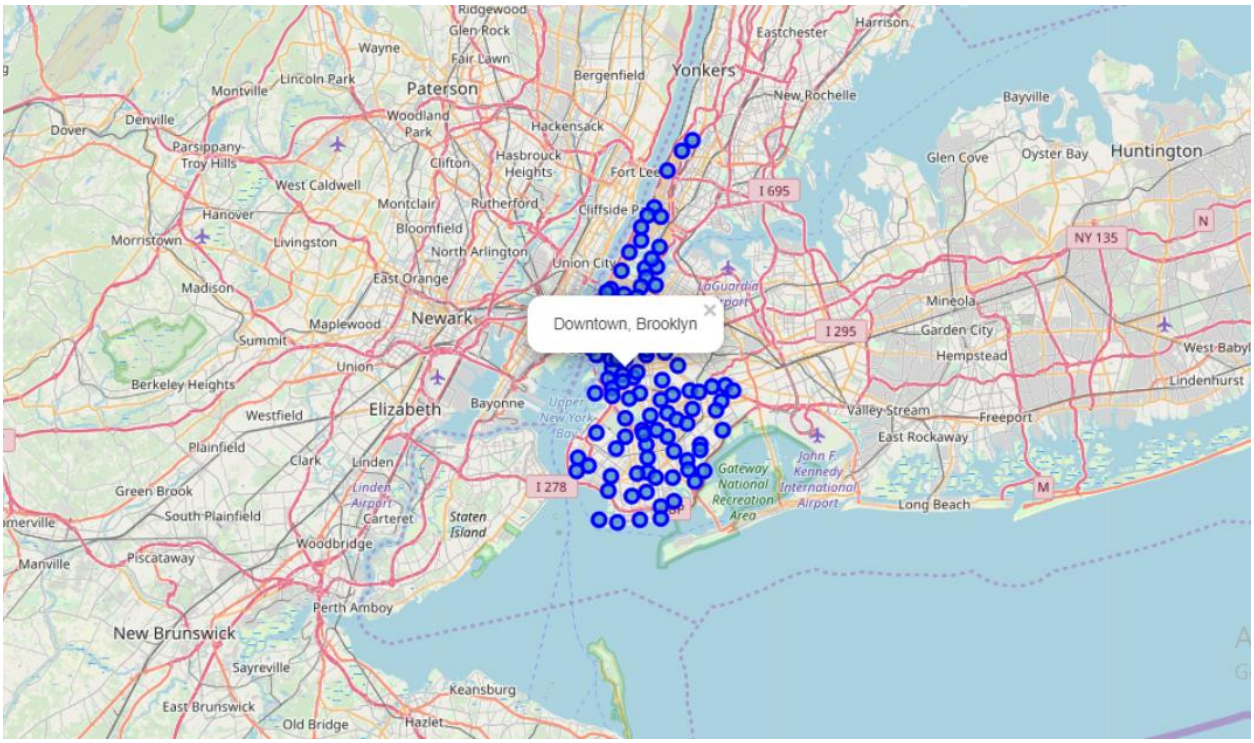
Data 2:We used Foursquare API data to explore neighborhoods in New york city.

Brooklyn and Manhattan:

Brooklyan and Manhattan Venues:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	The Bar Room at Temple Court	40.711448	-74.006802	Hotel Bar
1	Marble Hill	40.876551	-73.91066	The Beekman - A Thompson Hotel	40.711173	-74.006702	Hotel
2	Marble Hill	40.876551	-73.91066	Alba Dry Cleaner & Tailor	40.711434	-74.006272	Laundry Service
3	Marble Hill	40.876551	-73.91066	The Wooly Daily	40.712137	-74.008395	Coffee Shop
4	Marble Hill	40.876551	-73.91066	City Hall Park	40.712415	-74.006724	Park

Brooklyan and Manhattan venues visualization:

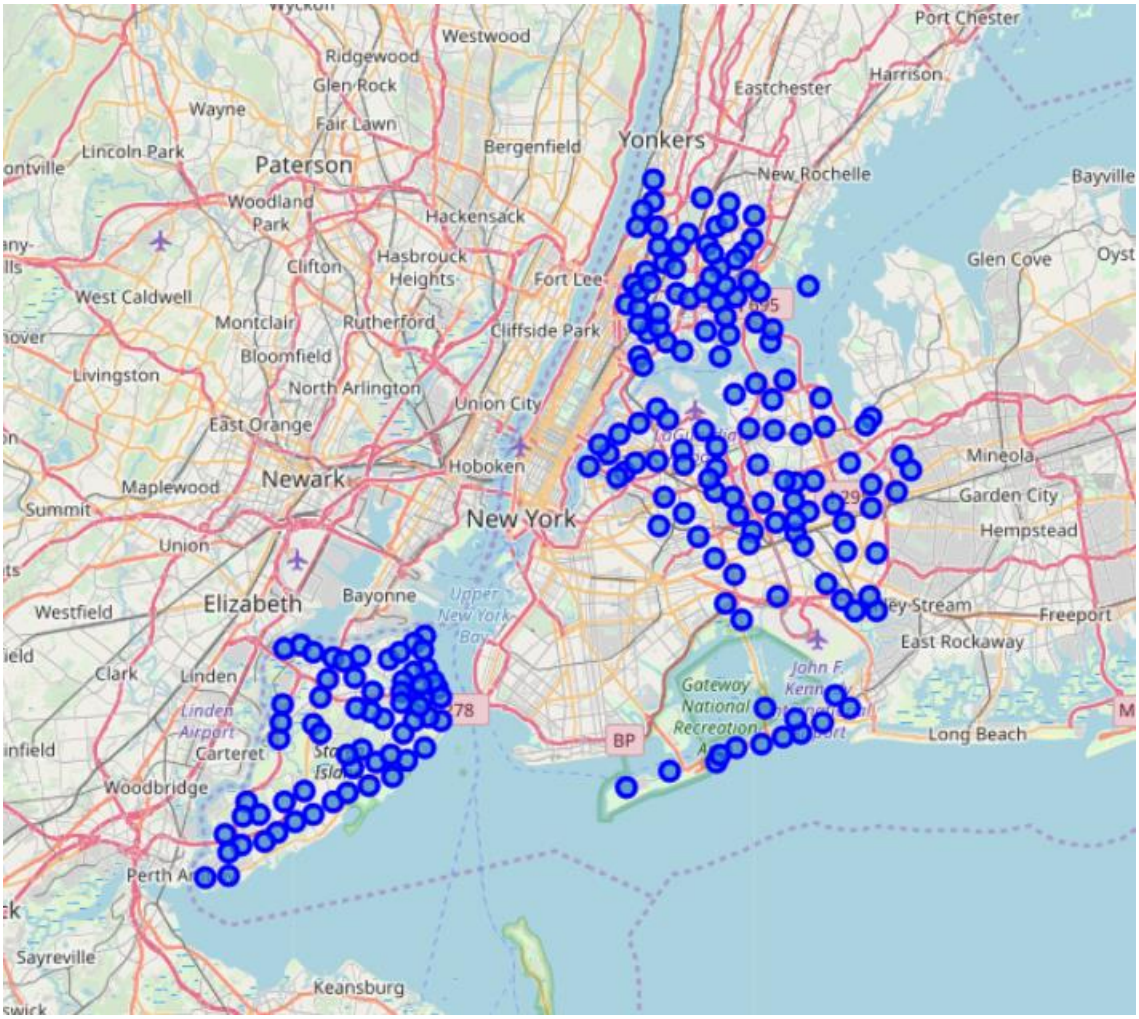


Bronz,Queens and Staten Islands:

Bronz,Queens,Staten Island Neighborhood Venue:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Wakefield	40.894705	-73.847201	The Bar Room at Temple Court	40.711448	-74.006802	Hotel Bar
1	Wakefield	40.894705	-73.847201	The Beekman - A Thompson Hotel	40.711173	-74.006702	Hotel
2	Wakefield	40.894705	-73.847201	Alba Dry Cleaner & Tailor	40.711434	-74.006272	Laundry Service
3	Wakefield	40.894705	-73.847201	The Wooly Daily	40.712137	-74.008395	Coffee Shop
4	Wakefield	40.894705	-73.847201	City Hall Park	40.712415	-74.006724	Park

Bronz,Queens,Staten Island Neighborhood Visualization:

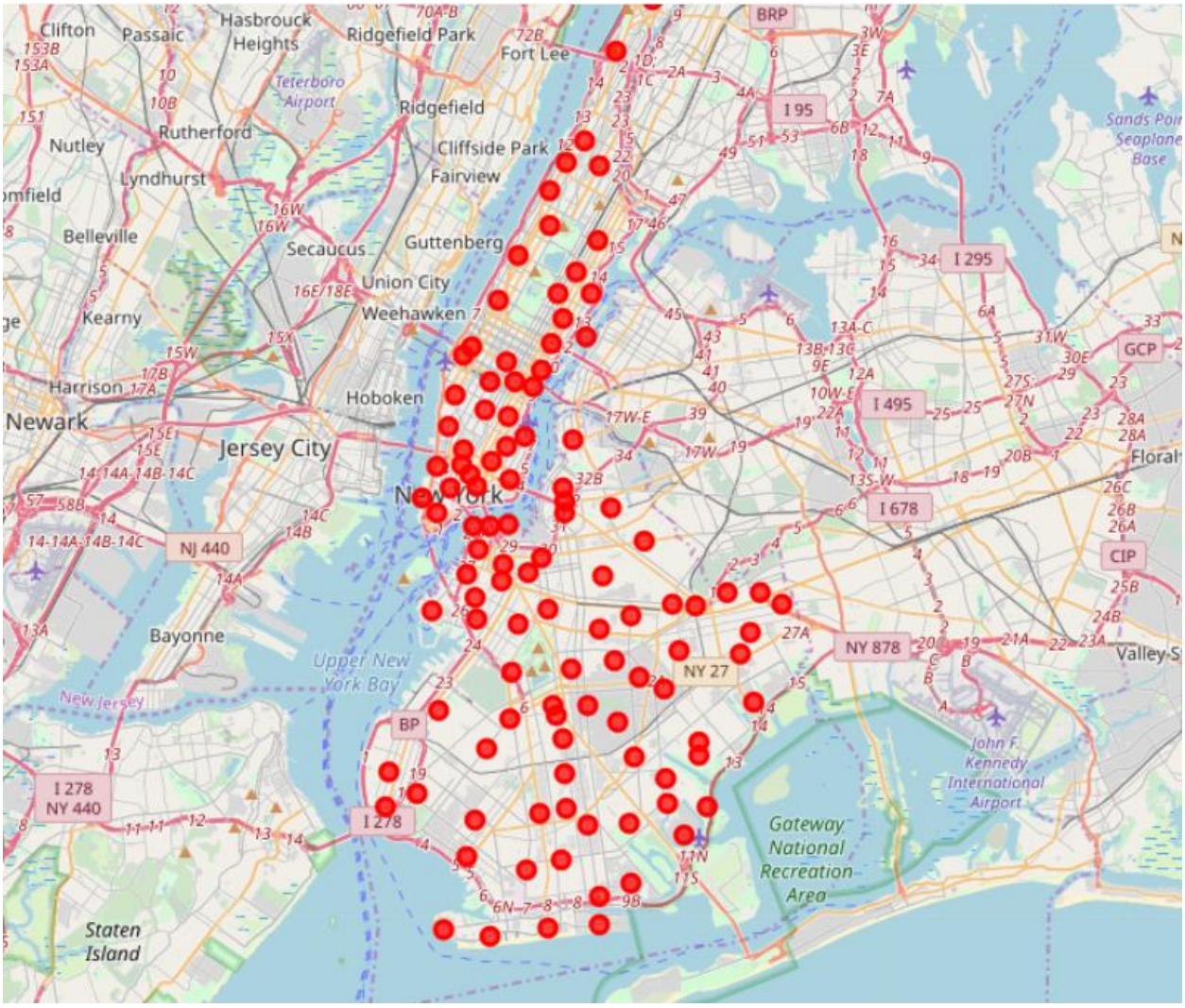


4.Results:

To cluster the neighborhoods into two clusters we used K-Means clustering Algorithm.k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with nearest mean.It uses iterative refinement approach.

Brooklyn and Manhattan:

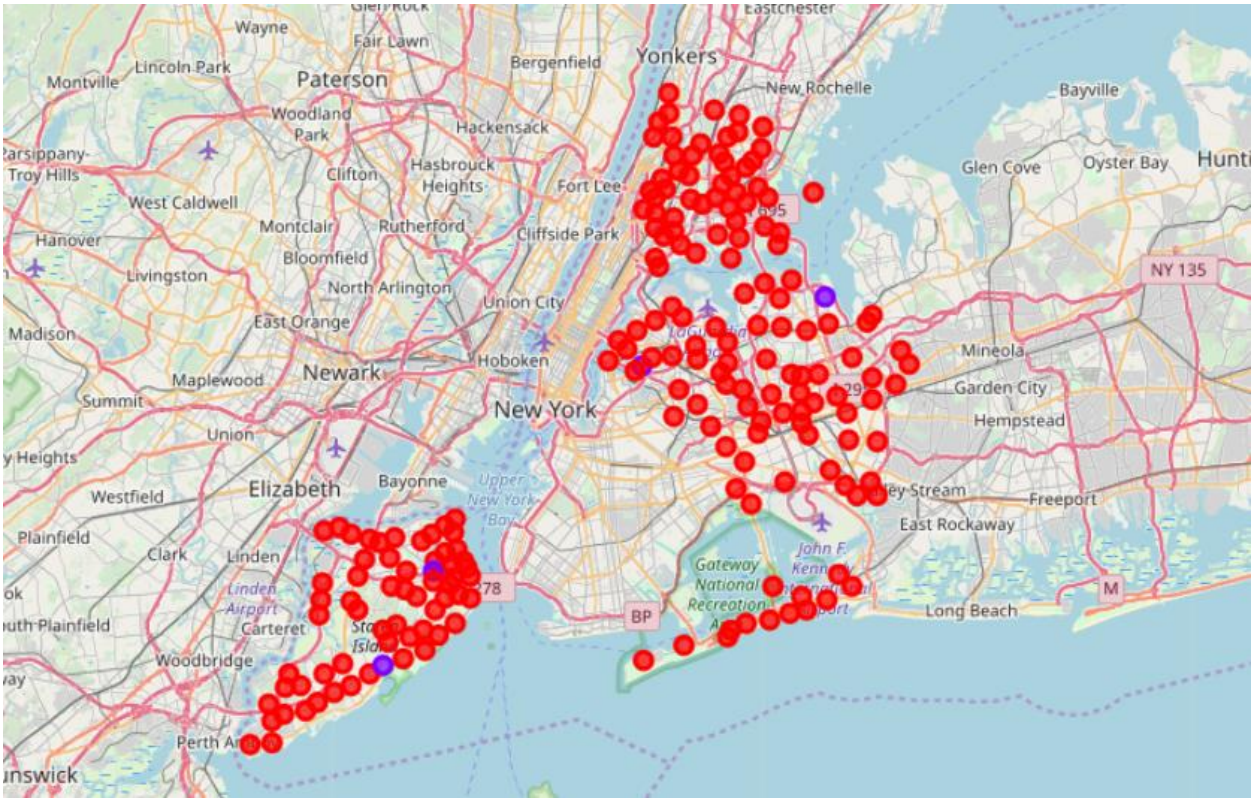
Using k-means algorithm:



Cluster 0:The total and total sum of cluster0 has smallest value.It shows market is not saturated.

Cluster 1: The total and total sum of cluster1 has highest value.It shows markets are saturated.

Bronz,Queens and Staten Island:



Cluster 0:The total and total sum of cluster0 has smallest value.It shows market is not saturated.

Cluster 1: The total and total sum of cluster1 has highest value.It shows markets are saturated.

5.Discussion

Brooklyn and Manhattan has high number of restaurant so very competitive market.

Bronx,Queens and Staten Islans also has good number of restaurants but not as many as required.

6.Conclusion

This analysis is performed on limited data.This may be right or wrong.If there are lot of restaurant,probably there is lot of demand. Brooklyn and Manhattan has high number of restaurant so very competitive market. Bronx,Queens and Staten Islans also has good number of restaurants but not as many as required.So this can be explored.