The Battle of the Neighborhoods - Report

1.Introduction & Business Problem:

Problem Background:

The New York city is the most populous city in the United States. It is diverse and is the financial capital of USA. It is multicultural. It provides lot of business oppourtunities and business friendly environment. It has attracted many different players into the market. It is a global hub of business and commerce.

This also means that the market is highly competitive. 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.

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 YorkPopulation
2. New York CityDemographics
3. Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality andcost?
4. Are there any venues like Gyms, Entertainmnet zones, Parks etc nearby where floating population is highetc
5. Who are the competitors in thatlocation?
6. Cuisine served / Menu of thecompetitors
7. Segmentation of theBorough
8. Untappedmarkets
9. Saturated markets
   and so on.

First move is very important, thereby choice of location is very important.

Objective and 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.

This would interest anyone who wants to start a new restaurant in Newyork city.

Success Criteria:

The success criteria of the project will be a good recommendation of Neighborhood based on Lack of such restaurants in that location and nearest suppliers of ingredients.

2. Data:

City of my choice : ***Newyork City***.

The datasets used are :

Data 1 : 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

	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 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

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

1. New YorkPopulation
2. New York CityDemographics
3. Cuisine of New Yorkcity

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

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. The below is image of the 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 :

Business Understanding :

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

Analytic Approach :

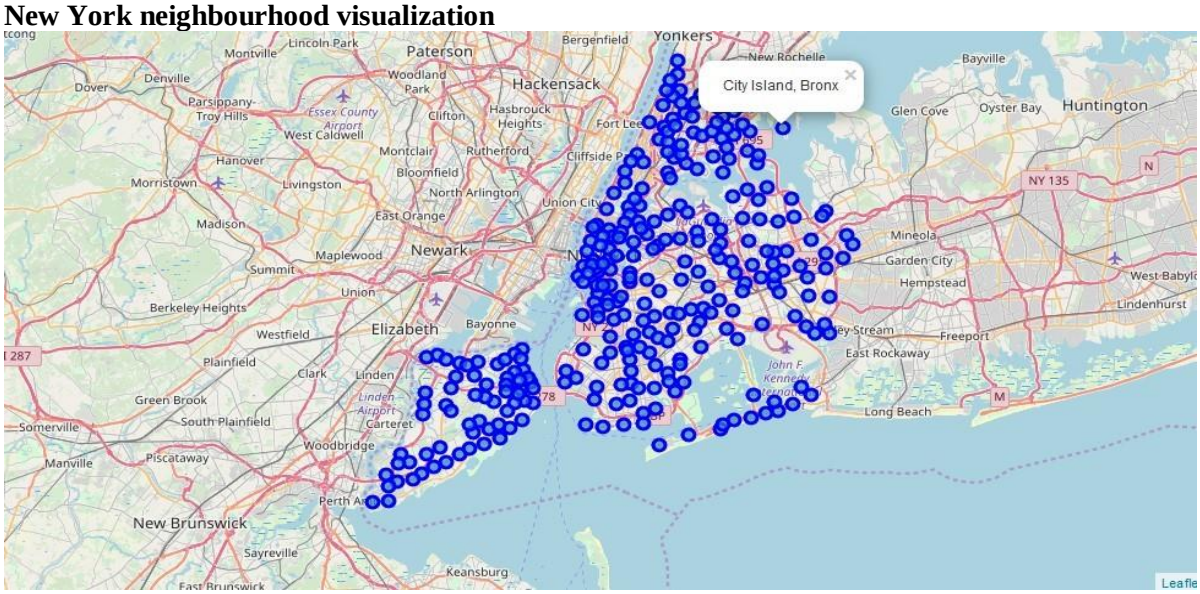
New York city neighbourhood has a total of 5 boroughs and 306 neighborhoods. In this project first part is clustering of Manhattan and Brooklyn . And second part is clustering of Bronx, Queens and Staten Island. This is done due to the following Exploratory data analysis.

Exploratory Data Analysis :

Data 1- New york city Geographical Coordinates Data.

1. In this we load the data and explore data from **newyork_data.json**file.
2. Transform the data into a pandas dataframe.
3. This dataframe contains the geographical coordinates of New York city neighborhoods.
4. This data will used to get Venues data from Fouresquare.
5. We used geopy, folium libraries to create a map of New York city with neighborhoods on top.

New York neighbourhood visualization

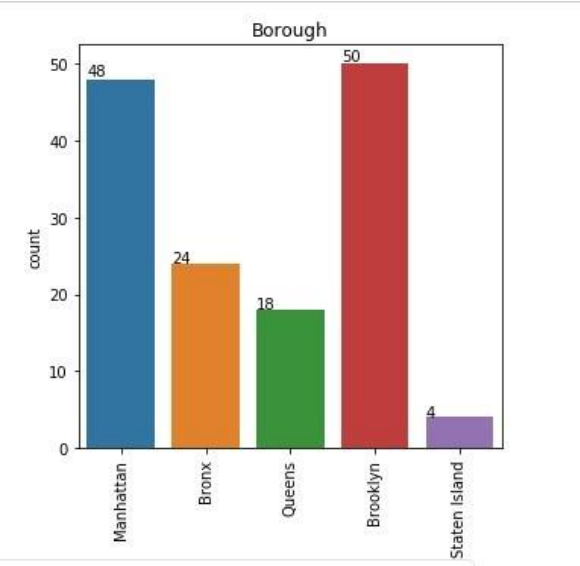


Data 2-

In this we will be using the data of Farmers Markets data.

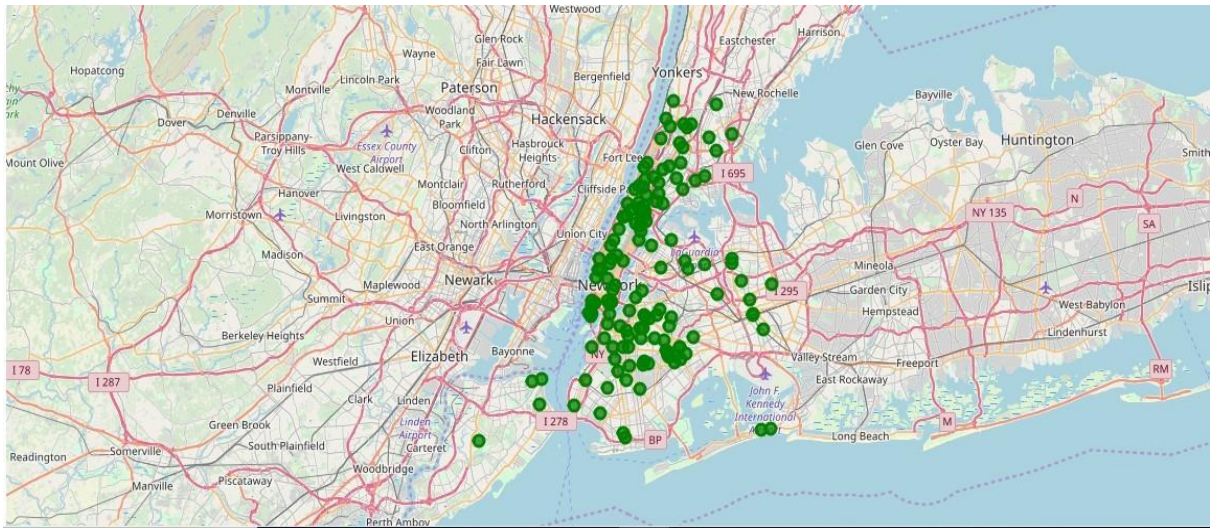
There are totally 144 Farmers Markets in New York city. Highest number are in Manhattan and Brooklyn.
And lowest in Queens, Bronx and Staten Island.

The proof of this is as given below.



We used geopy and folium libraries to create a map to visualise farmers markets of New York city.

Farmers Market visualisation-New York City



Data 3 : To analyize New York city Population, Demographics and Cuisine , scrapped the data from Wikipedia pages given above in the data section. We used BeautifulSoup python library. Beautiful Soup is a Python package for parsing HTML and XML documents (including having malformed markup, i.e. non-closed tags, so named after tag soup).

1. New York Population : Insights from the data:

	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 City Demographics :.

	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 has been manually prepared. Data is taken from Wikipedia page -https://en.wikipedia.org/wiki/Cuisine_of_New_York_City. Using this data we created word cloud.

NEW YORK CITY CUISINE :

BROOKLYN CUISINE

MANHATTAN CUISINE -

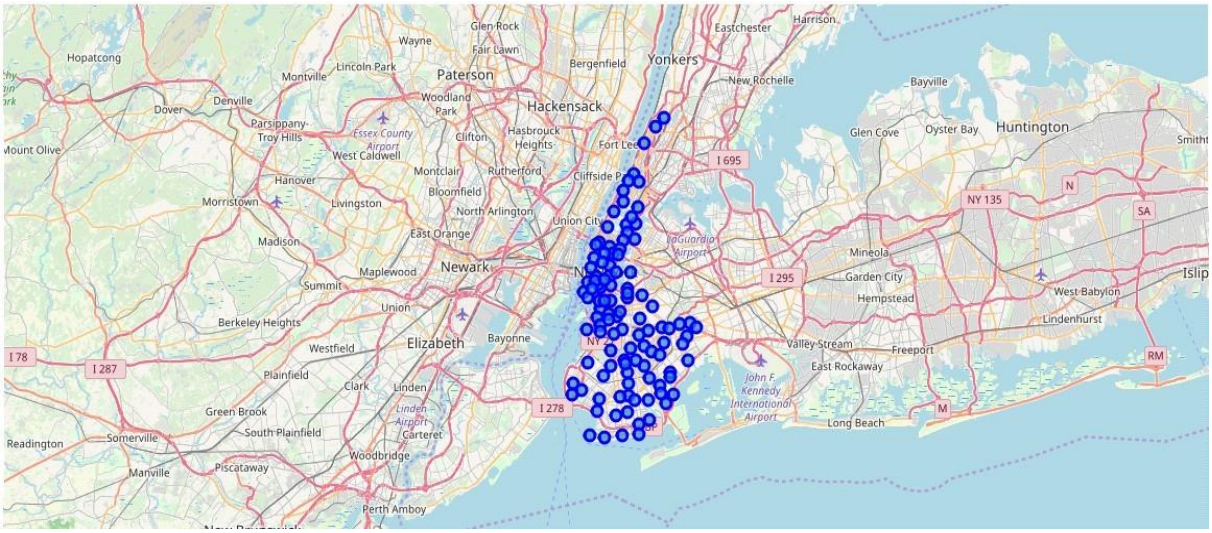
QUEENS CUISINE -

THE BRONX CUISINE -

Data 4 : NewYork city geographical coordinates data has be utilized as input for the Foursquare API. We used the Foursquare API data to explore neighborhoods in New York City.

Brooklyn and Manhattan :

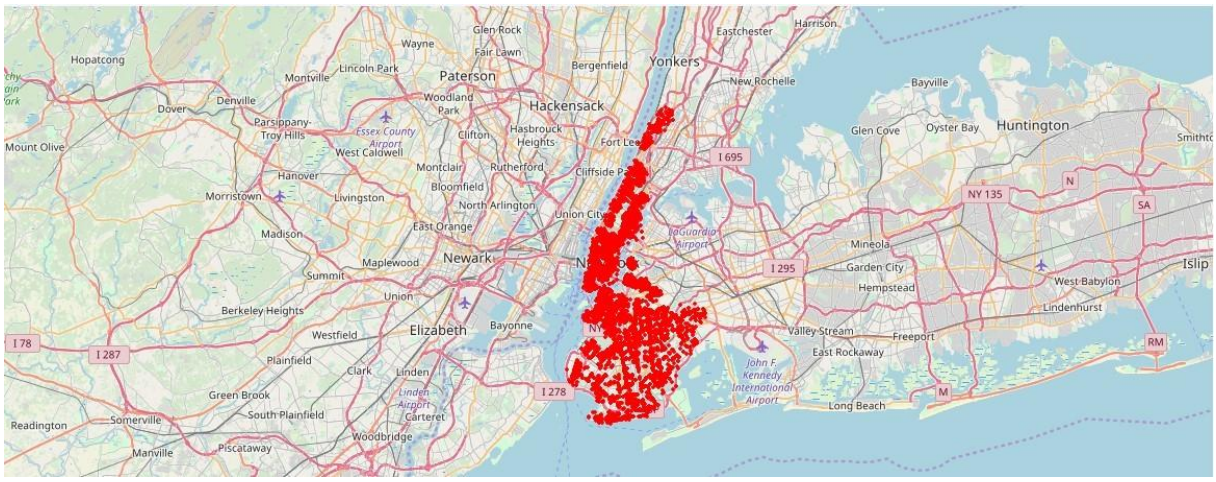
Visualization :



Brooklyn and Manhattan Venues :

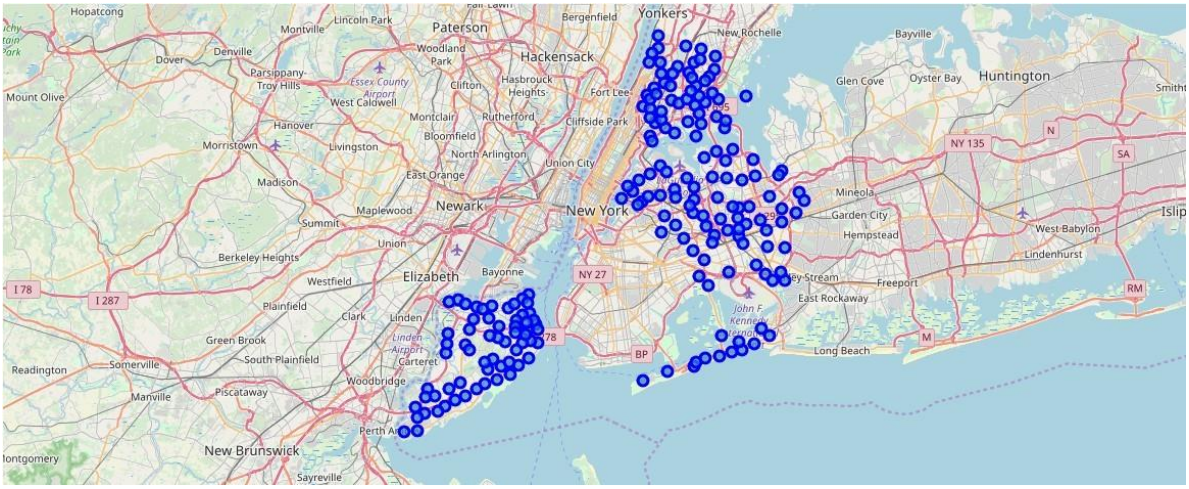
	Neighborhood	NeighborhoodLatitude	NeighborhoodLongitude	Venue	VenueLatitude	VenueLongitude	VenueCategory
0	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
1	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
2	Marble Hill	40.876551	-73.91066	Tibbett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

Visualization :



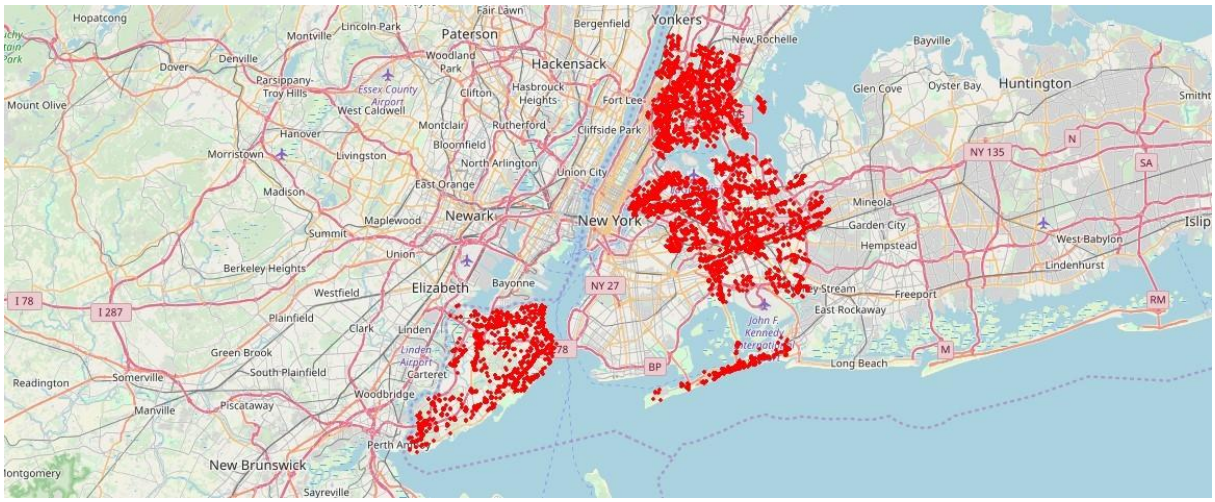
Bronx, Queens and Staten Island :

Visualization :



Bronx, Queens and Staten Island Venues :

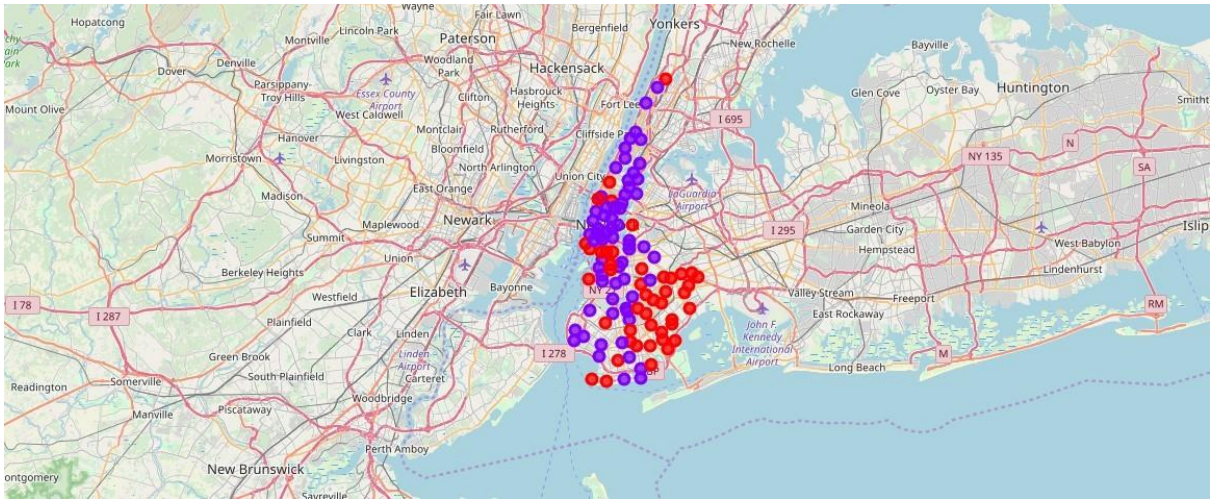
Visualization :



RESULTS:

Neighborhood K-Means clustering:
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.

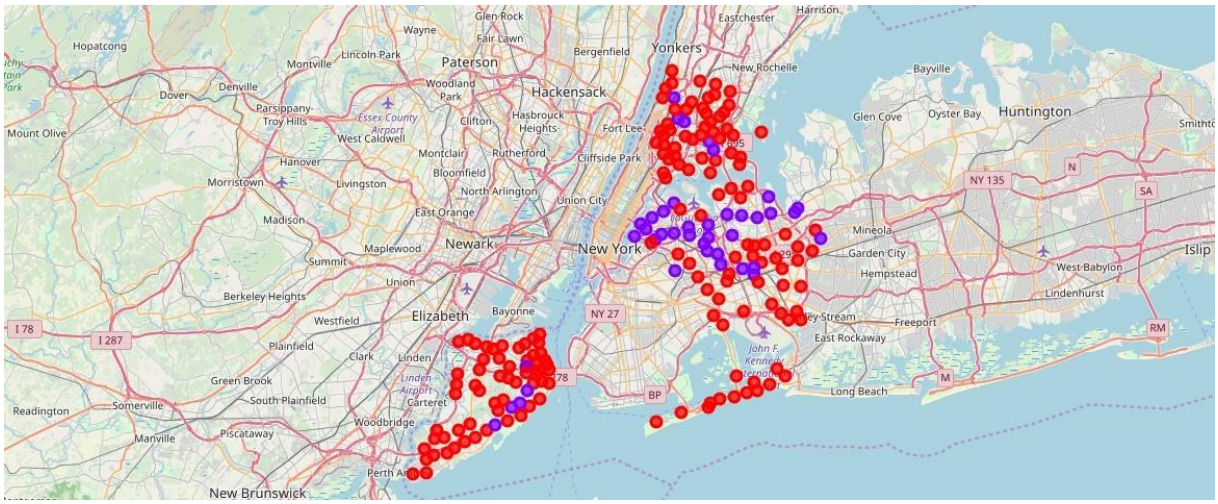
Brooklyn & Manhattan :



Cluster0 : It shows that the market is not saturated.

Cluster1 : It shows that the markets are saturated. Number of restaurants are very high.

Bronx, Queens and Staten Island:



Cluster0 : It shows that the market is not saturated. There are untapped neighborhoods. List is 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 : 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 StatenIsland. There is scope to explore cuisines of various countries in Bronx, Queens and StatenIsland. In Manhattan and Brooklyn restaurants of cuisines of many countries are available. Since people love multi cuisines, risk can be taken.

CONCLUSION:

As the data is less, the insight from the analysis may be right or wrong. If a considerate amount of data is available, it would provide better results.

As of now, a venue with less competition can be easily provided for any company seeking it.

.