THE BATTLE OF NEIGHBOURHOODS

Segmentation and Clustering

MAY 20, 2019

WALI FAROOQUI

Contents

List of Figures

Introduction

Problem Statement

The New York City, is the most densely populated city in the United States. It is diverse and is the financial capital of the United States. It is known for its multiculturalism and home for business opportunities with supportive business environment. The city has attracted minds across the global with a dream to be successful on global platform which this city has to offer. Indeed, it is a global hub of business and commerce. The city is a major centre for banking and finance, retailing, world trade, transportation, tourism, real estate, new media, traditional media, advertising, legal services, accountancy, insurance, theatre, fashion, and the arts in the United States. Due it's multiculturalism and opportunities, there have been influx of people from across the to come and pursue their dreams. This influx has brought cousins from different parts of the world.

Thus, we should also accept the fact that market in NYC is highly competitive and cost of doing business or cost of living is also very high as compared with other cities. Hence, one must undergo with appropriate survey or analytics with available data to draw a conclusion to take few essential decisions even before going into the market for business. The insights derived from analysis will give good understanding of the demography, purchasing power, business environment etc which will help in strategically targeting the market. This will help in reduction of risk and the Return on Investment will be reasonable.

Problem Description

Our main goal is to find best location for a new restaurant. 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.

- Central and Eastern European immigrants, especially Jewish immigrants - bagels, cheesecake, hot dogs, knishes, and delicatessens
- Italian immigrants - New York-style pizza and Italian cuisine
- Jewish immigrants and Irish immigrants - pastrami and corned beef
- Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city
- mobile food vendors - Some 4,000 licensed by the city
- Middle Eastern foods such as falafel and kebabs examples of modern New York street food
- 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.

So it is evident that 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:

- New York Population

- New York City Demographics
- Are there any Farmers Markets, Wholesale markets etc nearby so that the ingredients can be purchased fresh to maintain quality and cost?
- Are there any venues like Gyms, Entertainment zones, Parks etc nearby where floating population is high etc
- Who are the competitors in that location?
- Cuisine served / Menu of the competitors
- Segmentation of the Borough
- Untapped markets
- Saturated markets etc

Even though well-funded XYZ Company, need to choose the correct location to start its first venture. If this is successful they can replicate the same in other locations. First move is very important, thereby choice of location is very important.

Target Audience:

To recommend the correct location, XYZ Company Ltd has appointed me to lead of the Data Science team. The objective is to locate and recommend to the management which neighbourhood of New York 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 New York city.

Success Criteria:

The success criteria of the project will be a good recommendation of borough/Neighbourhood choice to XYZ Company Ltd based on Lack of such restaurants in that location and nearest suppliers of ingredients.

Data Overview:

The data we will use for this analysis is a combination of different data sources mentioned below, that has been prepared for purpose of the analysis and the location/venue information in foursquare. 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 neighbourhood. We will use the Foursquare API to explore neighbourhoods in New York City.

Neighbourhood Data

Neighbourhood has a total of 5 boroughs and 306 neighbourhoods. In order to segment the neighbourhoods and explore them, we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the latitude and longitude coordinates of each neighbourhood.

This dataset[1] exists for free on the web.

	Borough	Neighbourhood	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

Figure 1 - Neighbourhood Data

Farmer Markets Data

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 [2].

A ***farmers' market*** is often defined as a public site used by two or more local or regional producers for the direct sale of farm products to consumers. In addition to fresh fruits and vegetables, markets may sell dairy products, fish, meat, baked goods, and other minimally processed foods.

	FacilityName	Service Category	Service Type	Address	Address 2	Borough	ZipCode	Latitude	Longitude	AdditionalInfo	StartDate	EndDate	Monday	Tuesday
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
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
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
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
5	175th Street Greenmarket	Farmers Markets and Food	Farmers Markets	175th St bet Wadsworth Ave &	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

Figure 2 - Farmers Markets

Scrape Data from Wikipedia

For the below analysis we will get data from Wikipedia as given below:

1. New York Population [3]
2. New York City Demographics [4]
3. Cuisine of New York city [5]

New York city geographical coordinates data will be rendered as input for the Foursquare API, that will be leveraged to provision venues information for each neighbourhood. We will use the Foursquare API to explore neighbourhoods in New York City. 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

Figure 3 - Neighbourhood Venue from Foursquare

Methodology

Business Understanding:

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

Analytic Approach:

New York city has 306 neighbourhoods within 5 boroughs. In this report, we will discuss clustering of Manhattan and Brooklyn, and clustering of The Bronx, Queens and Staten Island. This is done because of the following Explanatory data analysis.

Data 1: New York City geographical coordinates data

1. We load the data and explore data from network_data.json[1] (New York city geo coordinate data)
2. Transform the data of nested python dictionaries into a pandas Dataframe.
3. Dataframe contains the geographical coordinates of the New York City neighbourhoods and boroughs.
4. We will use this Dataframe to get Venues data from Foursquare.
5. Geopy and folium libraries were used to create a map of New York City with neighbourhoods superimposed on top.

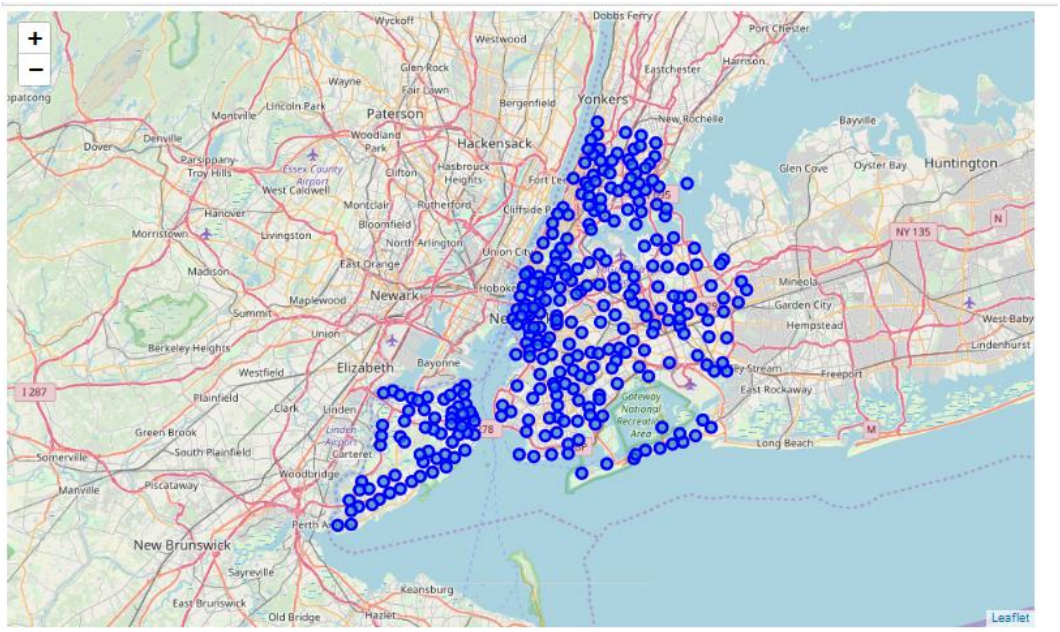

Figure 4 - New York neighbourhood visualization

Data 2: Farmers Markets and Food Boxes dataset

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. Refer graph below for better understanding:

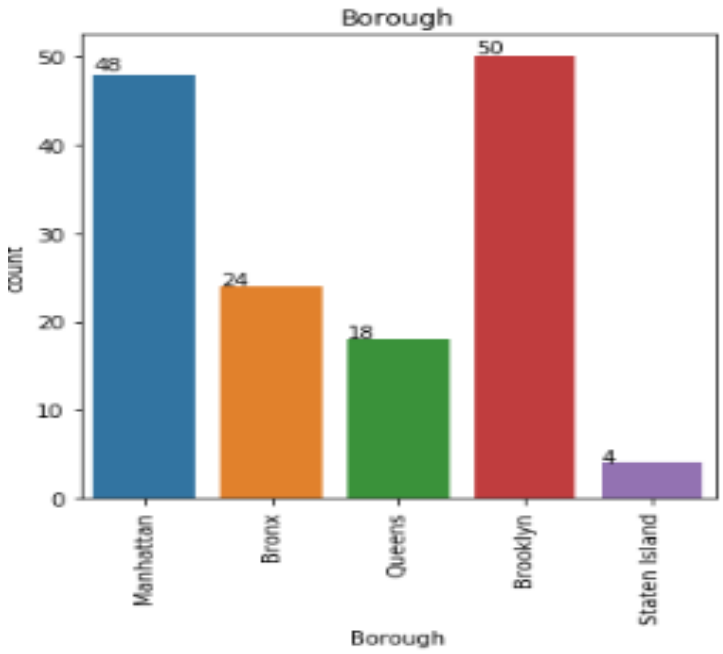

Figure 5 - Farmers Markets in New York

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

Figure 6 - Farmers Market visualisation-New York City

Data 3: New York City data

To analyse New York city population, Demographics and Cuisine, I scraped the data from Wikipedia pages in the data section. We used BeautifulSoup python library which is used for parsing HTML and XML documents. It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping.

New York Population

- Manhattan (New York county) is the geographically smallest and most densely populated borough.
- Manhattan's population density is 72033 people per square mile as 2015 census, and it has highest GDP and per capita income.
- Brooklyn (Kings county) is the most populous borough and has considerate per capita income.
- Queens (Queens county) is geographically largest borough and it is second in New York city for GDP and per capita income.

	Borough	County	Population	GDP	Per Capita Income	Sq Miles	Sq Km	person per sq mi	person per sq km
1	The Bronx	Bronx	1,471,160	28.787	19,570	42.10	109.04	34,653	13,231
2	Brooklyn	Kings	2,648,771	63.303	23,900	70.82	183.42	37,137	14,649
3	Manhattan	New York	1,664,727	629.682	378,250	22.83	59.13	72,033	27,826
4	Queens	Queens	2,358,582	73.842	31,310	108.53	281.09	21,460	8,354
5	Staten Island	Richmond	479,458	11.249	23,460	58.37	151.18	8,112	3,132

Figure 7 - New York City Population

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. This is the reason New York city has restaurants serving cuisine from many countries such as Indian, African, Japan etc. This also increases the scope for restaurants business in New York City.

	Racialcomposition	2010	1990	1970	1940[251]
0	White	44.0%	52.3%	76.6%	93.6%\r
1	—Non-Hispanic	33.3%	43.2%	62.9%	92.0%\r
2	Black or African American	25.5%	28.7%	21.1%	6.1%\r
3	Hispanic or Latino (of any race)	28.6%	24.4%	16.2%	1.6%\r
4	Asian	12.7%	7.0%	1.2%	–\r

Figure 8 - Demography of New York City

Cuisine of New York city

1. **NEW YORK CITY CUISINE:** Most Preferred Food in New York City –Italian, Puerto Rican, Mexican, Jewish, Dominican, Irish and Pakistani.

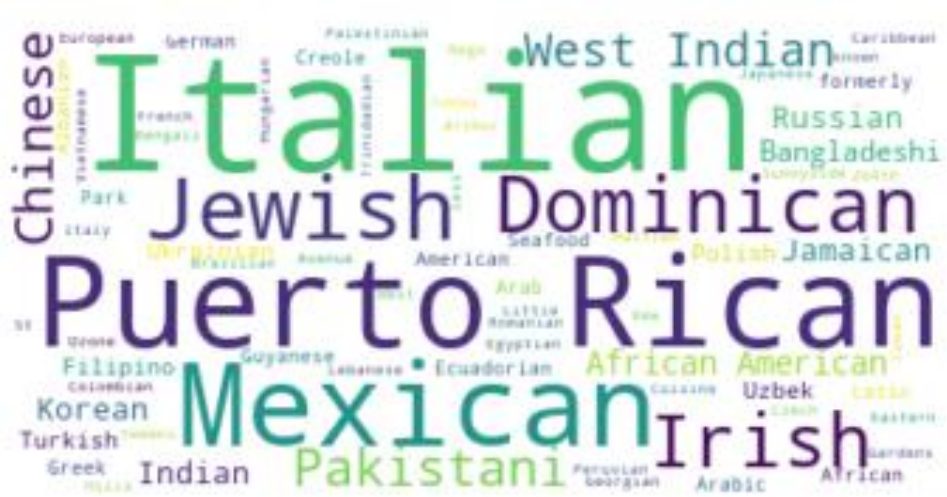

Figure 9 - New York City Cuisine

2. **BROOKLYN CUISINE:** Most Preferred Food in Brooklyn is –Italian, Puerto Rican & Mexican

Figure 10 - Brooklyn Cuisine

3. **MANHATTAN CUISINE:** Most Preferred Food in Manhattan is – Italian, American, Puerto Rican and Indian.

Figure 11 - Manhattan Cuisine

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

Figure 12 - Queens Cuisine

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

Figure 13 - Bronx Cuisine

Data 4: New York City Venues

New York city geographical coordinates data has been utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighbourhood. We used the Foursquare API data to explore neighbourhoods in New York City.

Brooklyn and Manhattan

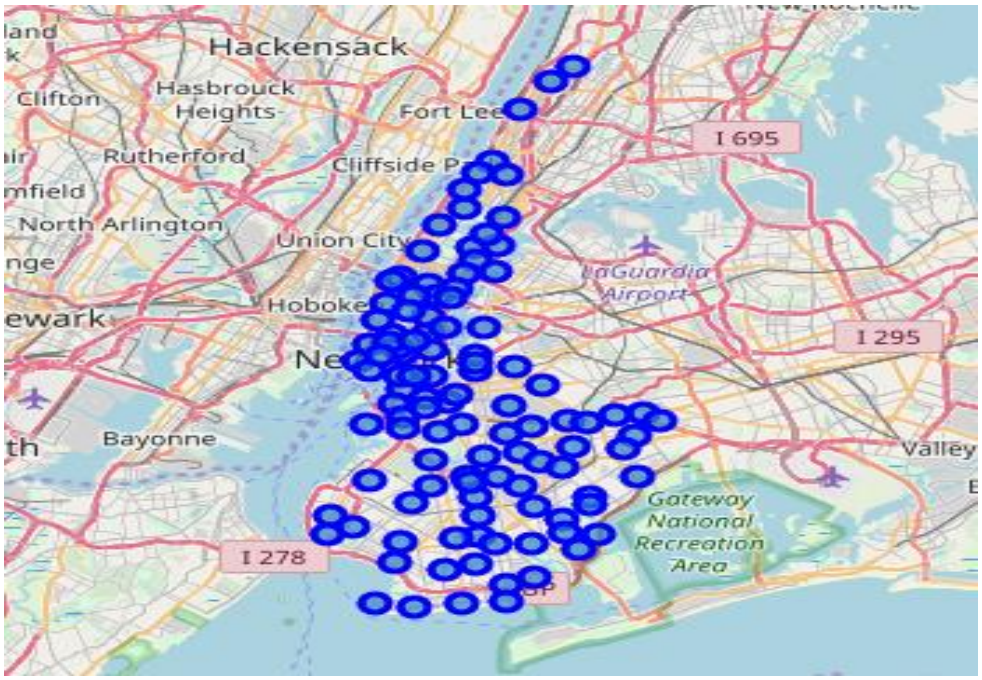

Figure 14 - Brooklyn and Manhattan Neighbourhood

Using the geographical coordinates of each neighbourhood foursquare API calls are made to get top 200 venues in a radius of 1000 meters. The venues data is as given below:

	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

Figure 15 - Brooklyn and Manhattan Venues

Figure 16 - Brooklyn and Manhattan Venues Map

Bronx, Queens and Staten Island

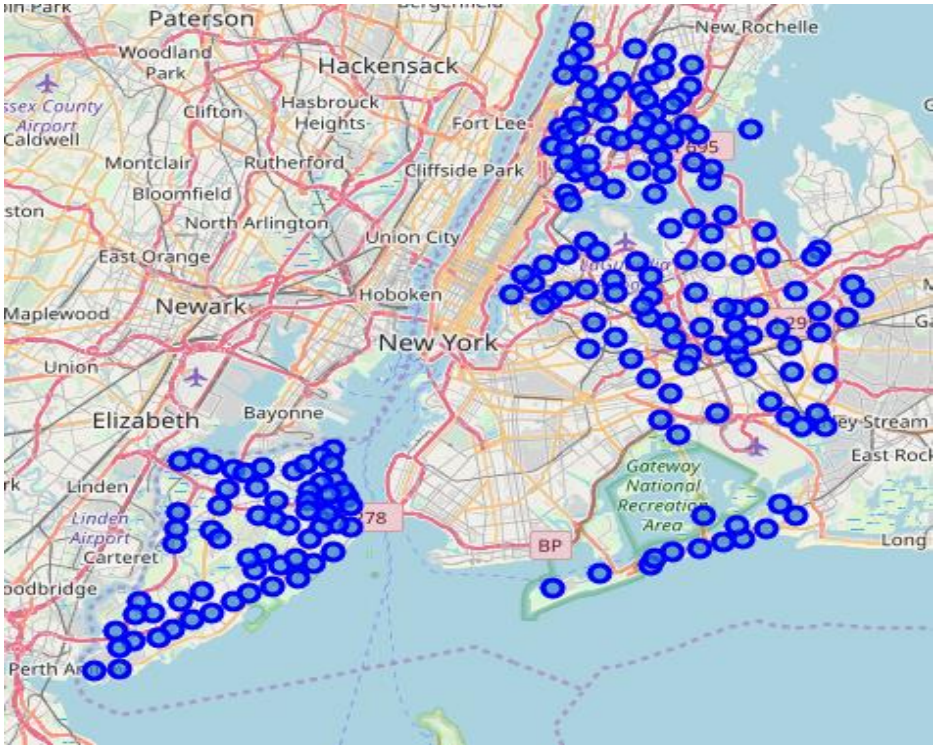

Figure 17 - Bronx, Queens and Staten Island Neighbourhood

	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	Rite Aid	40.896649	-73.844846	Pharmacy

Figure 18 - Bronx, Queens and Staten Island Venues

Bronx, Queens and Staten Island has 10805 venues and 387 unique venue types as data from Foursquare API.

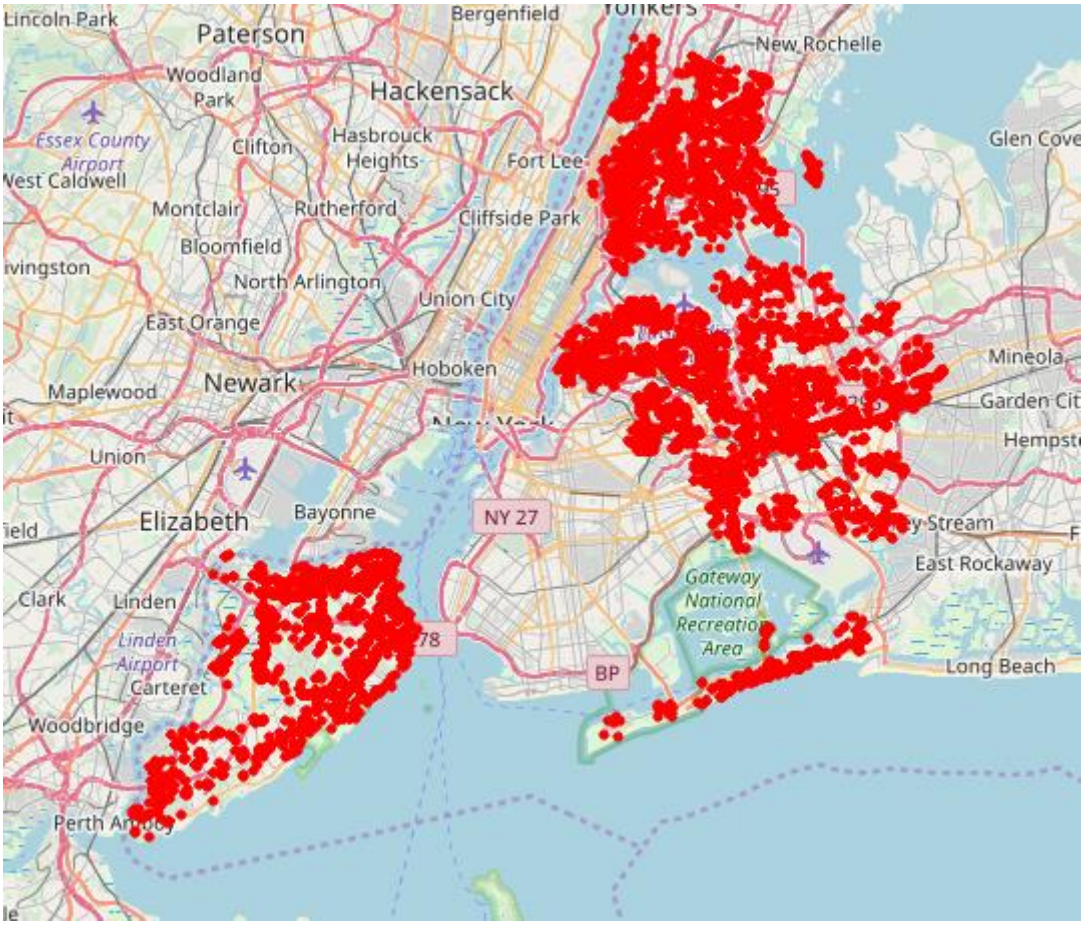

Figure 19 - Bronx, Queens and Staten Island Venues Map

Result

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 focussed only on restaurants business.

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

To cluster the neighbourhoods 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

In the below Map Visualization, we can see the different types of clusters created by using K-Means for Brooklyn & Manhattan.

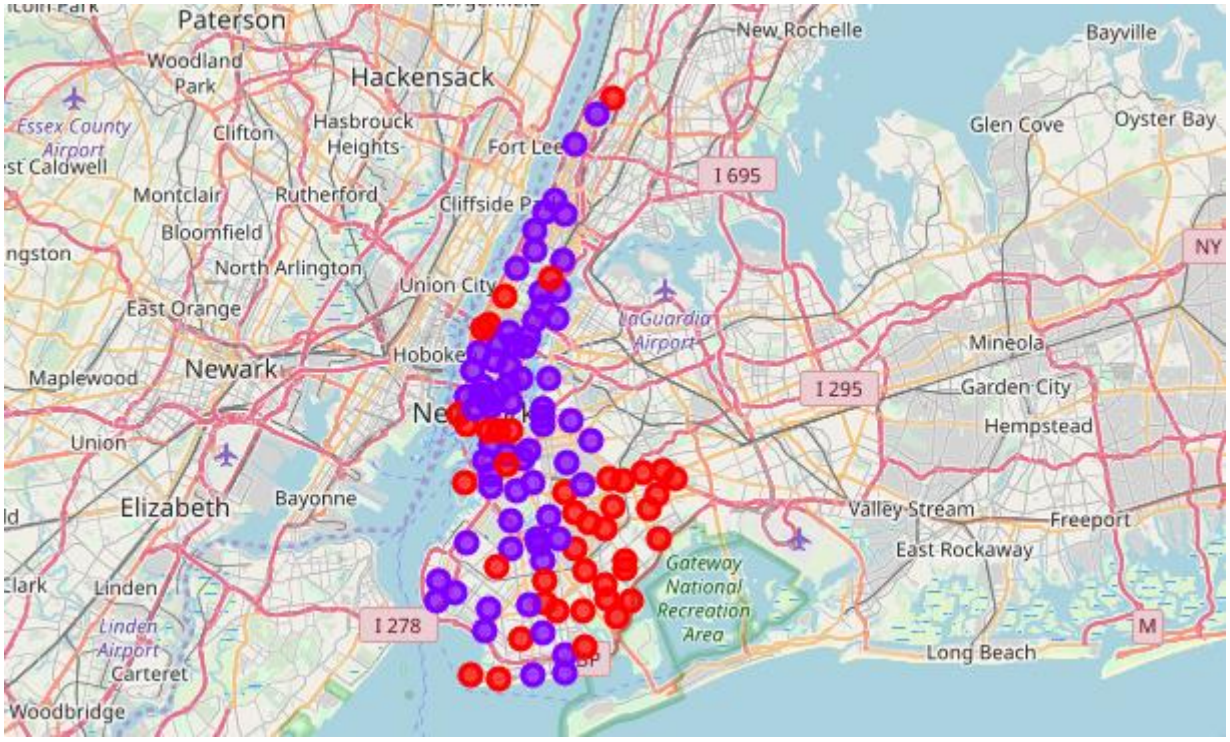

Figure 20 - Restaurant Cluster in Brooklyn and Manhattan

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 markets are saturated. Number of restaurants are very high. There are no untapped neighbourhoods in Brooklyn and Manhattan.

Bronx, Queens and Staten Island

In the below Map Visualization, we can see the different types of clusters created by using K-Means for Bronx, Queens and Staten Island.

Figure 21 - Restaurant Clusters

Cluster 0: The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated. There are untapped neighbourhoods. List is as given below.

	Borough	Neighbourhood	Latitude	Longitude	Total	Cluster_Labels
0	Staten Island	Todt Hill	40.597069	-74.111329	0	1
1	Staten Island	South Beach	40.580247	-74.079553	0	1
2	Staten Island	Port Ivory	40.639683	-74.174645	0	1
3	Staten Island	Butler Manor	40.506082	-74.229504	0	1
4	Staten Island	Bloomfield	40.605779	-74.187256	0	1

Figure 22 - Neighbourhood for New Restaurant

Cluster 1: The Total and Total Sum of cluster 1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

DISCUSSION

1. There is scope to increase Farmers markets in Bronx, Queens and Staten Island.
2. There is scope to explore cuisines of various countries in Bronx, Queens and Staten Island.
3. 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

This analysis is performed on limited data. This may be right or may be wrong. But if good amount of data is available there is scope to come up with better results. If there are lot of restaurants probably there is lot of demand. 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.

Based on per capita and population density, Manhattan and Brooklyn are suitable place for new restaurant but market is almost saturated in these boroughs.

Further, based on farmers' market data, availability of fresh ingredients may have identified best in Brooklyn and lowest in Staten Island. And since Brooklyn has competitive restaurant market where as Staten Island is almost untapped for the restaurant, and not to mention per capital in Brooklyn and Staten Island is almost same but there is huge different in population. We have cuisine data for the Brooklyn and we do not have sufficient data for the Staten Island.

As per the neighbourhood or restaurant type mentioned like Indian Restaurant analysis can be checked. A venue with lowest risk and competition can be identified.

We did not have access to average rent in the neighbourhoods, which could be a key factor while minimizing operational cost for a new business. With additional data we definitely come up with a better result and find best location for the restaurant or any new business.

References

1. New York City Geographical Data: https://geo.nyu.edu/catalog/nyu_2451_34572
2. Farmers' Market data: https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2
3. New York City Population: https://en.wikipedia.org/wiki/New_York_City
4. New York City Demography: https://en.wikipedia.org/wiki/Demographics_of_New_York_City
5. Cuisines of New York City: https://en.wikipedia.org/wiki/Cuisine_of_New_York_City