THE BATTLE OF THE NEIGHBORHOOD

1. INTRODUCTION & BUSINESS PROBLEM

Problem Background:

The City of New York, is the most populous city in the United States. It is the financial capital of USA and provides lot of business opportunities and business friendly environment. It is a global hub of business and commerce which means that the market is highly competitive. The high competitiveness leads to higher cost of doing business. 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.

Problem Description:

A restaurant is a place where you can eat a meal and pay for it. In restaurants your food is usually served to you at your table by a waiter or waitress. The City of New York is famous for its excellent cuisine. The food culture includes an array of international cuisines influenced by the city's immigrant history.

So, it is evident that to survive in such a competitive market, it is very important to strategically implement the entire business. One of the factors which plays a crucial plan in the restaurant business is the location. Considering New York as the city to start the business, few of the parameters to analysis are:

1. New York Population
2. New York City Demographics
3. Nearby suppliers like Farmers Markets, Wholesale markets, etc
4. Nearby venues where moving population is high
5. Location Competitiveness
6. Targeted Audience

The objective is to analyse the locality and recommend which neighborhood of New York City will be best choice to start a restaurant.

2. DATA INFORMATION

The city to be analysed in this project is New York.

The following datasets would be used for the project:

Data 1: New York City Neighbourhood: https://geo.nyu.edu/catalog/nyu_2451_34572

This dataset will give a general idea of the neighborhood. In order to segment the neighborhoods and explore them, we will essentially need a dataset that contains the number of boroughs in the neighborhoods, along with the co-ordinates of each neighborhood.

Data 2: Farmers Markets and Food Boxes: https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2

In order to have a better insight of the availability of the raw materials, an analysis of the major Farmers Markets and Food Boxes in the neighborhood is necessary, which aide the business. 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.

Data 3: GrowNYC's Fresh Food Box: https://www.grownyc.org/greenmarketco/foodbox

As an example, we would take GrowNYC's Fresh Food Box Program, which is a food access initiative that enables under-served communities to purchase fresh, healthy, and primarily regionally grown produce well below traditional retail prices.

Data 4: New York Population: https://en.wikipedia.org/wiki/New_York_City

The overall New York population has to be analysed to make a better prediction for the business, which will include various key factors like age, gender, area population, etc.

Data 5: New York City Demographics:
https://en.wikipedia.org/wiki/Economy_of_New_York_City ,
https://en.wikipedia.org/wiki/Portal:New_York_City

The City Demographic Analysis would help in narrowing the best fit locality of the restaurant.

Data 6: Cuisine of 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

To run a restaurant business, a better understanding of the food preferences is required to make an informed decision.

3. METHODOLOGY

- **Business Understanding**

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

- **Analytic Approach**

New York city neighbourhood has numerous boroughs and neighborhoods. In order to analyze it, we have used clustering of Manhattan and Brooklyn in the first part and clustering of Bronx, Queens and Staten Island in the second part. This is done using various Exploratory data analysis.

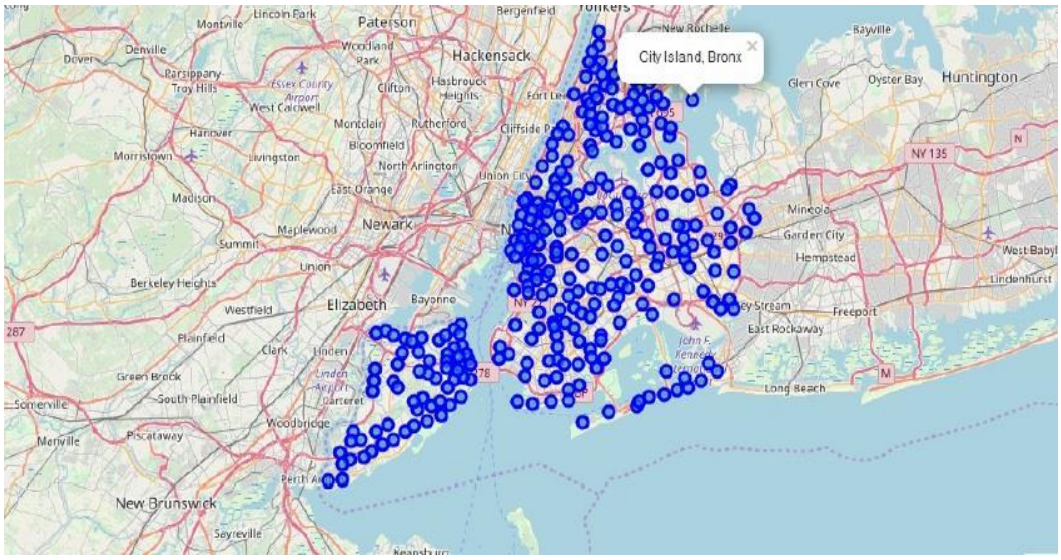
- **Exploratory Data Analysis**

➢ Data 1- New York city Geographical Coordinates Data.

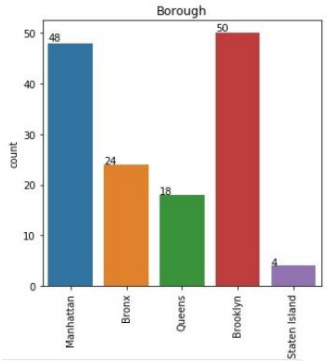
a. We load the data and explore data from a json file.
b. Transform the data of nested python dictionaries into a pandas data frame.

c. The data frame will now contain the geographical coordinates of New York City neighborhoods.
d. This data will used to get venues data from Fouresquare.
e. We used geopy and folium libraries to create a map of New York city with neighborhoods superimposed on top.

New York neighbourhood visualization



➢ Data 2- Second data used is the Farmers Markets and Food Boxes dataset. In this we will be using the data of Farmers Markets data.



a. There are totally 144 Farmers Markets in New York City, highest being in Manhattan and Brooklyn. The lowest are in Queens, Bronx and Staten Island.

The proof of this is as shown in the picture.

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

➢ Data 3: To analyze New York city Population, Demographics and Cuisine, we use BeautifulSoup python library to scrap the data.

a. New York Population Analysis:

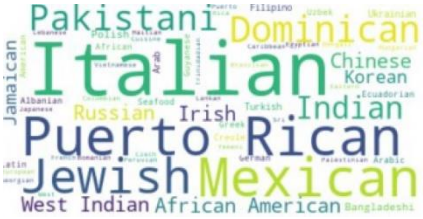
	Borough	County	Estimate_2017	square_miles	square_km	persons_sq_mi	persons_sq_km
0	Manhattan	New York	1,664,727	22.83	59.13	72,033	27,826
1	The Bronx	Bronx	1,471,160	42.10	109.04	34,653	13,231
2	Brooklyn	Kings	2,648,771	70.82	183.42	37,137	14,649
3	Queens	Queens	2,358,582	108.53	281.09	21,460	8,354
4	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

b. New York City Demographics Analysis:

	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%	–

c. Cuisine of New York city: Using this data we worked with word cloud to do the analysis.

- NEW YORK CITY CUISINE - Most preferred food in New York City – Italian, Purto Rican, Mexican, Jewish, Indian, Pakistani & Dominican.



- 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.

There is very less data of cuisine relating to Staten Island. So could not develop word cloud with it.

➢ Data 4: New York City geographical coordinates data has be utilized as input for the Foursquare API, that has been leveraged to provision venues information for each neighborhood. We used the Foursquare API data to explore neighborhoods in New York City.

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

Bronx, Queens and Staten Island Venues Visualization: The "BQS_venues" dataframe has 10805 venues and 387 unique venue types.

	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

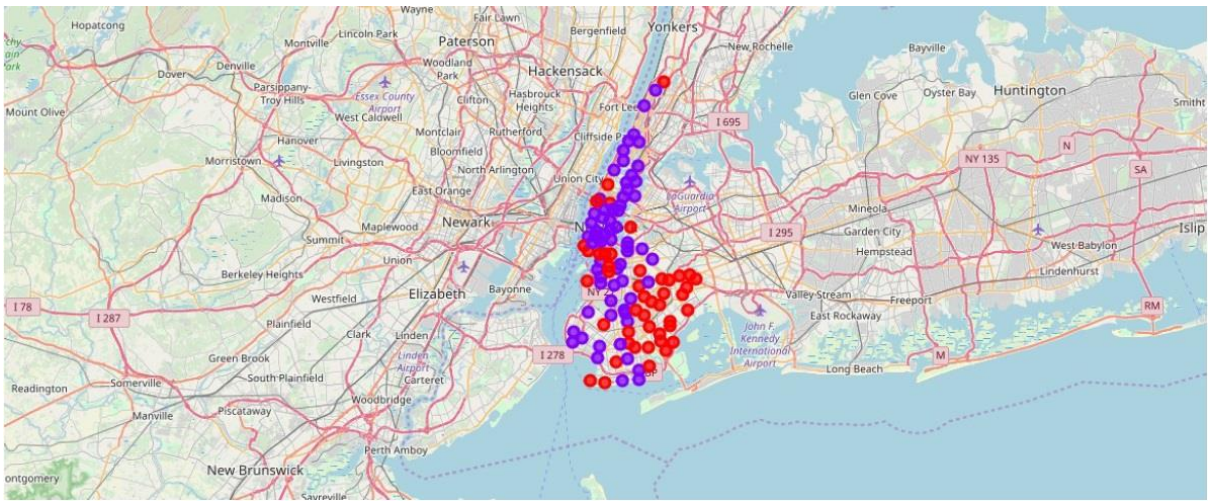
From this venues data we filtered and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens and Staten Island clustering.

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

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.

1. Brooklyn & Manhattan

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



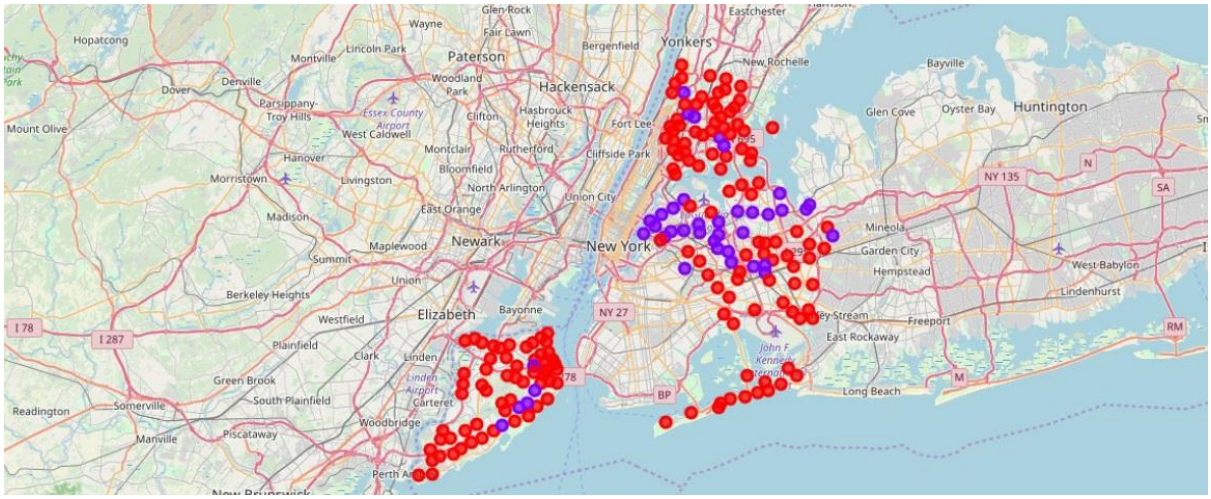
Cluster 1: The Total and Total Sum of cluster 1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

Cluster 0: The Total and Total Sum of cluster 0 has smallest value. It shows that the market is not saturated.

There are no untapped neighborhood in Brooklyn and Manhattan.

2. 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.



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.

Cluster 0: The Total and Total Sum of cluster 0 has smallest value. It shows that the market is not saturated. There are untapped neighborhood. 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

4. CONCLUSION

This analysis is performed on limited data. But if good amount of data is available there is scope to come up with better results. Brooklyn and Manhattan have high concentration of restaurant business. Bronx, Queens and Staten Island have some opportunity to tap in the restaurant business. A venue with lowest risk and competition can be identified. There is scope to increase Farmers markets in Bronx, Queens and Staten Island. In Manhattan and Brooklyn restaurants of cuisines of many countries are available. Risk can be taken with great menu on board.