The Battle of the Neighborhoods

By: J.Carrasquillo

1.Introduction :

Problem Background:

New York City, is the most populous cities in the United States. It is diverse and it is the financial capital of USA. It is multicultural place that provides a lot of business oppourtunities and a business friendly environment. It has attracted many different players into the market. 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.

Situation:

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. New York City is famous for its excelllent cuisine. It's food culture includes an array of international cuisines influenced by the city's immigrant history.
1. Central and Eastern European immigrants, - bagels, cheesecake, hot dogs, knishes, and delicatessens
2. Italian immigrants - New York-style pizza and Italian cuisine
3. Jewish immigrants and Irish immigrants - pastrami and corned beef
4. Chinese and other Asian restaurants, sandwich joints, trattorias, diners, and coffeehouses are ubiquitous throughout the city
5. mobile food vendors - Some 4,000 licensed by the city
6. Middle Eastern foods such as falafel and kebabs examples of modern New York street food
7. It is famous for not just Pizzerias, Cafe's but also for fine dining Michelin star restaurants.The city is home to "nearly one thousand of the finest and most diverse cuisine restaurants in the world", according to Michelin.
In order to survive in such competitive market it is very important to startegically plan.
Various factors need to be studied in order to decide on the location :
1. New York Population
2. New York City Demographics
3. Are there any Farmers Markets, Wholesale markets etc.

The One Company Ltd. needs to choose the correct location to start its first restaurant . If they are successful they can replicate the same in other locations. It is very important to select the right location for the right venue.

Target Audience:

To select the correct location, ONE Company Ltd has appointed a team of Data Scientists for this project. The objective is to locate and recommend to the management which New York neighborhood will be best choice to start a restaurant. The Management also expects to understand the recommendations made.
This would be very important for anyone who wants to start a new restaurant in New York city.

Criteria:

The success of the project will be determined by the neighborhood of choice to ONE Company Ltd based on lack of specific type of restaurants in that location and nearest suppliers of ingredients.

2. Data :

One city will be analysed in this project : New York City.
We will be using the below datasets for analyzing New York City

Data 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 is available on the web. Link to the dataset is :
https://geo.nyu.edu/catalog/nyu_2451_34572

Data 2 : Second data which will be used is the DOHMH Farmers Markets and Food Boxes dataset. 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
GrowNYC'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.

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

1. New York Population
2. New York City Demographics
3. Cuisine of New York city

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 : New York City geographical coordinates data will be used for Foursquare API, And will be leveraged to provide venues 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.

3.Methodology :

Business Understanding
Our main goal is to get a high traffic location for ONE Company Ltd. New restaurant venue.

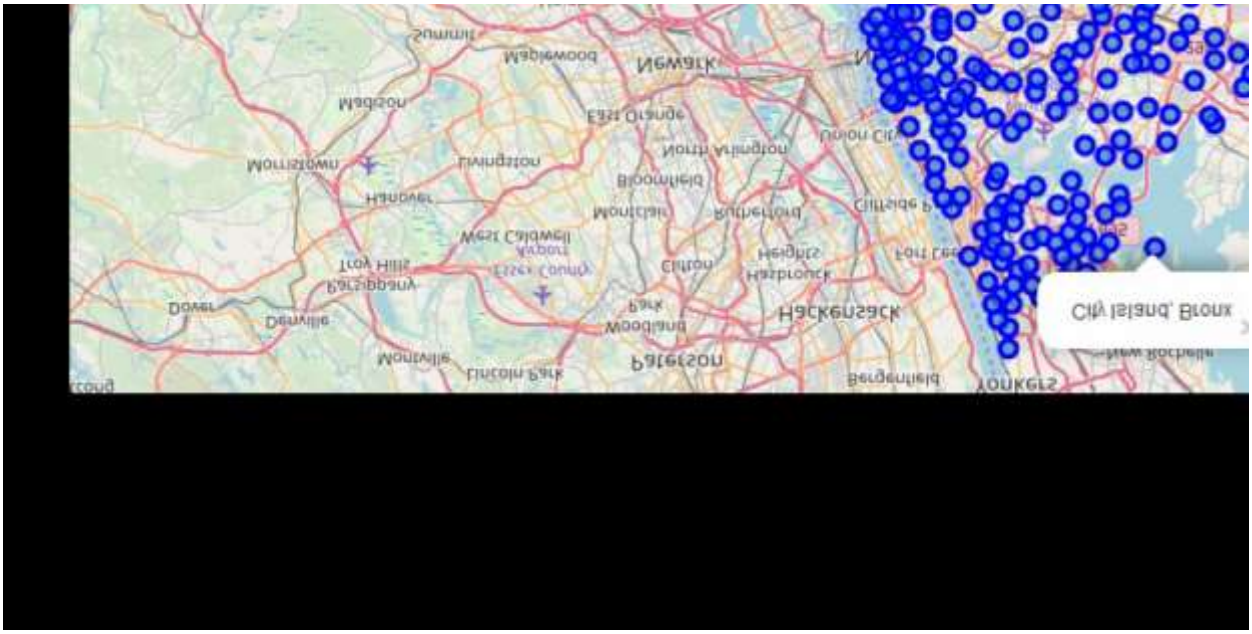
Analysis Approach :

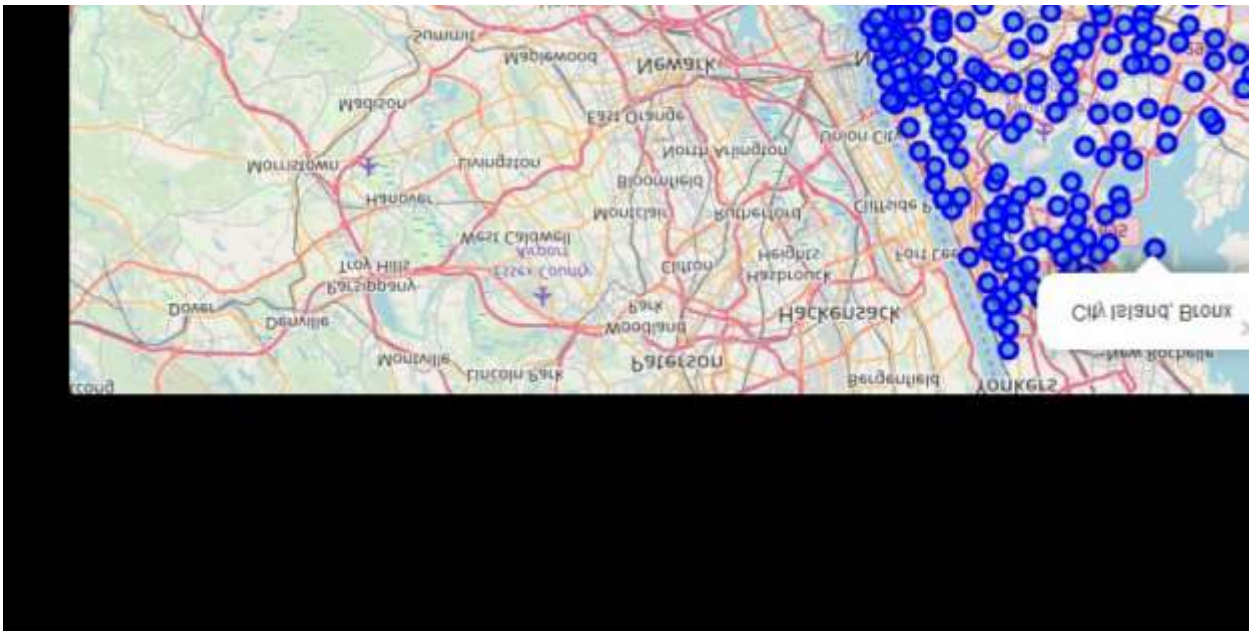
New York City neighborhood has a total of 5 boroughs and 306 neighborhoods. The first part is the clustering of Manhattan and Brooklyn . And second part is the clustering of Bronx, Queens and Staten Island. This is done because of the following Exploratory data analysis.

Data Analysis :

Data 1- New York Cty Geographical Coordinates Data.
1. In this we load the data and explore data from newyork_data.json file.
2. Transform the data of nested python dictionaries 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 Foursquare.
5. We used geopy and folium libraries to create a map of New York City with neighborhoods superimposed on top.

New York neighborhood visualization





Data 2- Second data which is used is the DOHMH 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. The highest number are located 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 visualization - New York City

Data 3 : To analyize New York City Population, Demographics and Cuisine , download 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). It creates a parse tree for parsed pages that can be used to extract data from HTML, which is useful for web scraping

1.New York Population : Insights from the data :
 Manhattan (New York County) is the geographically smallest and most densely populated borough.
 Manhattan's (New York County's) population density of 72,033 people per square mile (27,812/km²) in 2015 makes it the highest of any county in the United States and higher than the density of any individual American city.

 Brooklyn (Kings County), on the western tip of Long Island, is the city's most populous borough.
 Queens (Queens County), on Long Island north and east of Brooklyn, is geographically the largest borough.

2. New York City Demographics : New York City is the most populous city in the United States,[9] with an estimated record high of 8,622,698 residents as of 2017,[7] 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.

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 did word cloud.

NEW YORK CITY CUISINE: Most Preferred Food in New York City –Italian, Puerto Rican, Mexican, Jewish, Indian, Pakistani & Dominican.

BROOKLYN CUISINE -Most Preferred Food in Brooklyn is –Italian, Puerto 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 : NewYork 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.

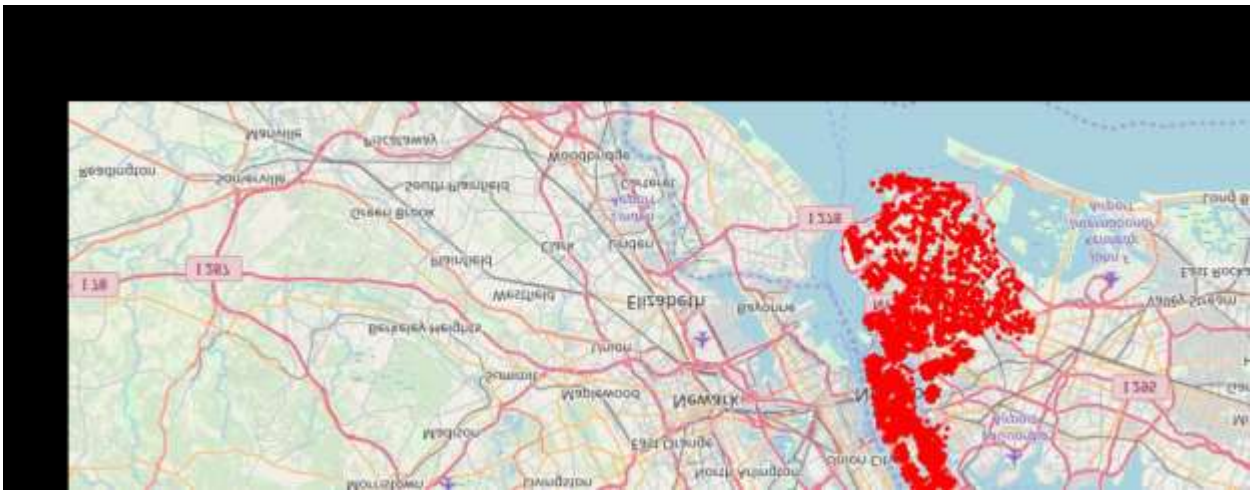
Brooklyn and Manhattan :

Brooklyn and Manhattan Visualization :

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

Brooklyn and Manhattan Venues :

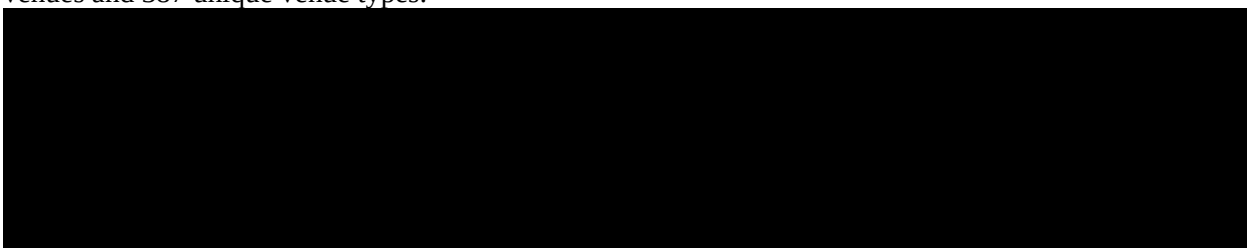
Brooklyn and Manhattan Venues Visualization : Generated the below Brooklyn and Manhattan Venues Visualization. The "BM_venues" dataframe has 9708 venues and 397 unique venue types.



Bronx, Queens and Staten Island :

Bronx, Queens and Staten Island Neighborhoods Visualization :

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



Bronx, Queens and Staten Island Venues Map Visualization :

4.RESULTS

We filtered the venues data and used only the restaurant data for Brooklyn & Manhattan clustering and Bronx, Queens and Staten Island clustering. As we focused only on restaurants business.

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.

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

Cluster 0 : The Total and Total Sum of cluster 0 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 neighborhoods 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.

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

Cluster1 : The Total and Total Sum of cluster1 has highest value. It shows that the markets are saturated. Number of restaurants are very high.

5.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. The risk can be taken with great menu on board. It also shows people love eating cuisines of various countries.

6.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.

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