Devendra - The Battle of the Neighbourhoods - Week 2

1. Introduction & Business Problem:

Problem Background:

The City of New York, is the global hub of business and commerce. 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 good.

Problem Description:

To open new food business **and survive in competitive market it is very important to strategically plan. Various factors need to be studied in order to decide on the Location such as :**

City population, Demographics, Farmers Markets, and Entertainment zones, Parks, competitors in that location and many other factors.

One company want to establish a new branch in particular area of the city so selection of location is very important.

A business in which they 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 excellent cuisine.

Target Audience:

The target audience are any company or individual who wants to start a new restaurant in Newyork city and want correct location for the restaurant with respect to success criteria. It will provide guaranty of success the business.

2. Data :

To do analysis of new york city we required many data point such as near by boroughs and neighbourhoods , Population , geography and Farmers Markets and Food Boxes dataset **etc..**

1. **Neighbourhoods** Data: Total of 5 boroughs and 306 neighbourhoods. we will essentially need a dataset that contains the 5 boroughs and the neighbourhoods that exist in each borough as well as the the latitude and longitude coordinates of each neighbourhood. Below is the link to download data :

   https://geo.nyu.edu/catalog/nyu_2451_34572

2. Farmers Markets and Food Boxes dataset:
   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.

   https://data.cityofnewyork.us/dataset/DOHMH-Farmers-Markets-and-Food-Boxes/8vwk-6iz2
   Website-https://www.grownyc.org/greenmarketco/foodbox

3. **From Wikipedia we will get below information**

   1. New York Population : More population more sales
   2. New York City Demographics
   3. Cuisine of New York city : What is choice if new york cityzen

   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
   https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City

4. **New york city geographical coordinates data:** To plot information on map.
   We will all the information of new york city by below link
   https://cocl.us/new_york_dataset

Postal Code	Latitude	Longitude
M1B	43.8066863	-79.1943534
M1C	43.7845351	-79.1604971
M1E	43.7635726	-79.1887115
M1G	43.7709921	-79.2169174
M1H	43.773136	-79.2394761

5. **FourSquare Data** to search existing venue , nearby restorent to particular loacation etc…
   We get this information through foursquare API.

   ```
   url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'.format(CLIENT_ID, CLIENT_SECRET, latitude, longitude, VERSION, search_query, radius, LIMIT)
   url
   ```

3. Methodology :

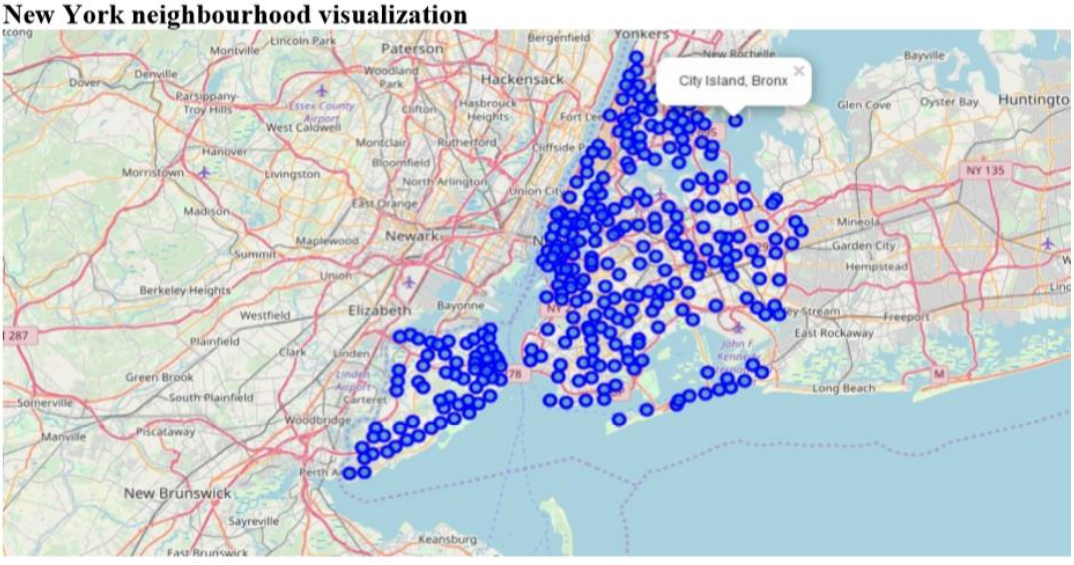
Business Understanding:
Our main goal is to get optimum location for new restaurant business in New York City for any Company or individual.

Analytic Approach:
New York city neighbourhood has a total of 5 boroughs and 306 neighbourhoods. 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 because of the following Exploratory data analysis.

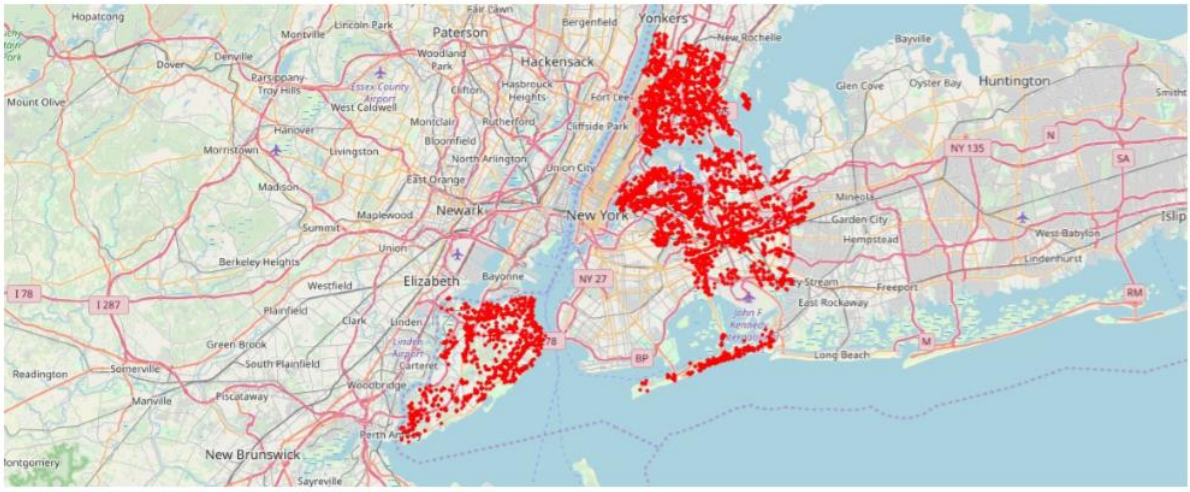
Data Analysis :

- Load New york city Geographical Coordinates Data from newyork_data.json file.
- This data will used to get Venues data from Fouresquare.
- We used geopy and folium libraries to create a map of New York city with neighborhoods superimposed on top.
- 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. Highest number are in Manhattan and Brooklyn. And lowest in Queens, Bronx and Staten Island.
- To analyize New York city Population, Demographics and Cuisine , scrapped the data from Wikipedia pages given above in the data section.
- 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 neighbourhood visualization



- 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.
- **Bronx, Queens and Staten Island Venues Map Visualization :**



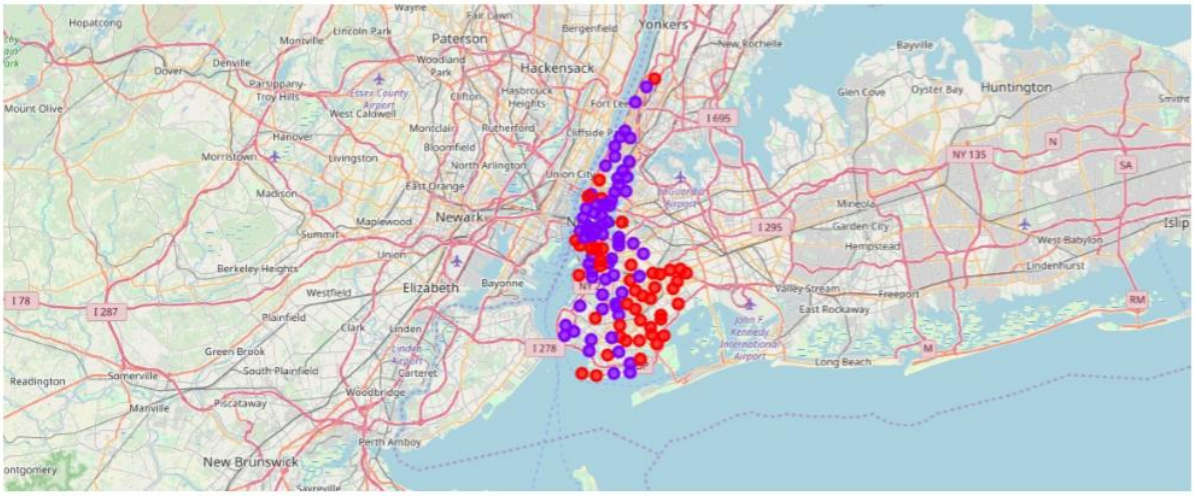
Neighbourhood K-Means clustering based on mean occurrence of venue category :
To cluster the neighbourhoods into two clusters we used the K-Means clustering Algorithm. k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. It uses iterative refinement approach.

4. RESULTS :

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.



Cluster0 : The Total and Total Sum of cluster0 has smallest value. It shows that the market is not saturated.
Cluster1 : 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.

Cluster0 : 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.

	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 : 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. .Scope to increase Farmers markets in Bronx, Queens and Staten Island.
2. .Scope to explore cuisines of various countries in Bronx, Queens and Staten Island.
3. In the 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.

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