The Battle of the Neighbourhoods - Week 1

1. Introduction & Business Problem

Problem Background:

The City of New York, is the most populous city in USA. It is multicultural. It has attracted many different players into the market 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. 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. And the Return on Investment will be reasonable.

Problem Description:

A restaurant is a booming business in New York. The city is famous for its excellent cuisine.

So, to survive in such competitive market it is very important plan and analyse various factors in

order to decide on the Location such as

1. Population
2. Demographics
3. Are there any Farmers Markets, Wholesale markets etc nearby so that the
4. ingredients can be purchased fresh to maintain quality and cost?
5. Are there any venues like Gyms, Entertainment zones, Parks etc nearby where
6. floating population is high etc?
7. Who are the competitors in that location?
8. Cuisine served / Menu of the competitors
9. Saturated markets etc
10. The list can go on...

Even though well-funded ABC restaurants chain needs 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, ABC restaurants chain 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 the city.

Success Criteria: The success criteria of the project will be a good recommendation of Neighbourhood choice to ABC restaurants chain based on Lack of such 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 analysing New York city

Data 1: In order to segment the neighbourhoods and explore them, we will essentially need a dataset neighbourhoods data with the latitude and longitude coordinates of each neighbourhood.

This dataset exists for free on the web. Link to the dataset is:
https://geo.nyu.edu/catalog/nyu_2451_34572

Data 2: 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
https://en.wikipedia.org/wiki/List_of_Michelin_starred_restaurants_in_New_York_City

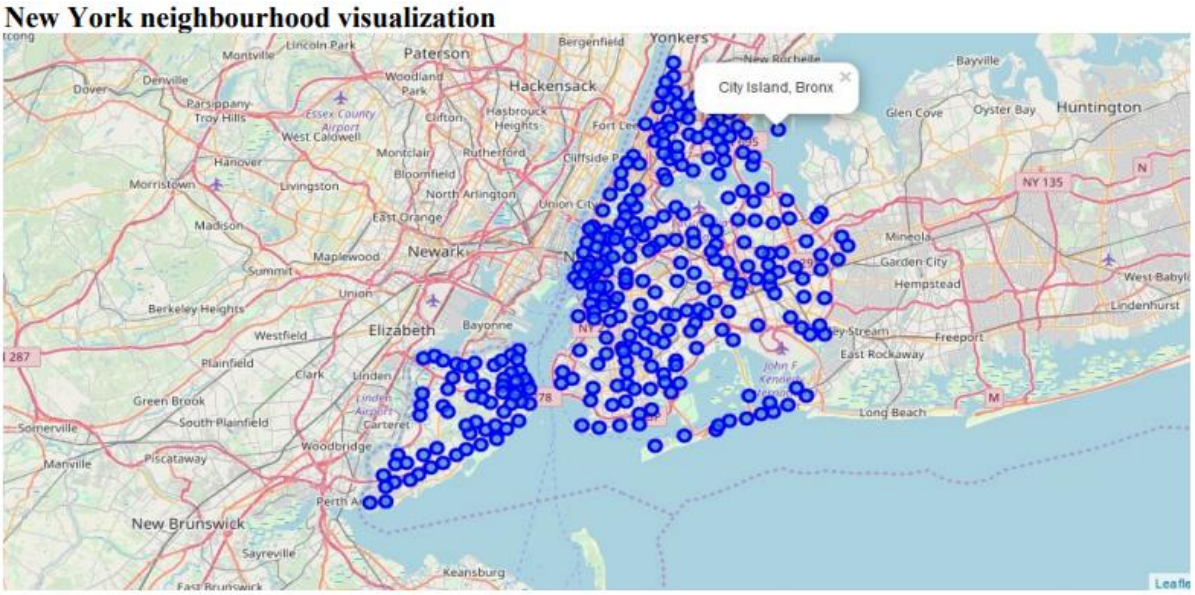
Data 3 : New York city geographical coordinates data will be utilized as input for the Foursquare API, that will be leveraged to provision venues information for each neighbourhood will use the Foursquare API to explore neighbourhoods in New York City. The below is image of the Foursquare API data.

3. Approach:

New York city neighbourhood has a total of 5 boroughs and 306 neighbourhoods. In this project we will focus on clustering Brooklyn and Manhattan areas only.

Exploratory Data Analysis:

Data 1- New york city Geographical Coordinates Data



Data 3 : To analysis of New York city Population, Demographics and Cuisine , scrapped the data from Wikipedia pages given above in the data section. We used BeautifulSoup python library.

Below insights were found from population data.

1. New York city population

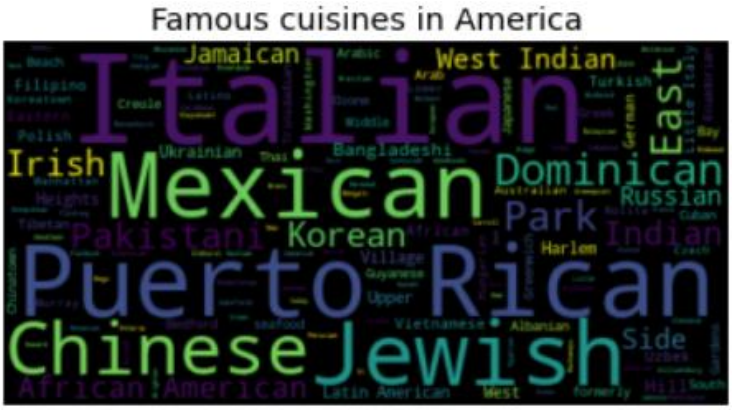
- Manhattan is the geographically smallest and most densely populated borough.
- Brooklyn on the western tip of Long Island, is the city's most populous borough.
- Queens on Long Island north and east of Brooklyn, is geographically the largest borough.

	NY_Borough	Jurisdiction	Population	GDP	square_miles	square_km	Borough	County
0	The Bronx	Bronx	1,418,207	42.695	42.10	109.04	33,867	13,006
1	Brooklyn	Kings	2,559,903	91.559	70.82	183.42	36,147	13,957
2	Manhattan	New York	1,628,706	600.244	22.83	59.13	71,341	27,544
3	Queens	Queens	2,253,858	93.310	108.53	281.09	20,767	8,018
4	Staten Island	Richmond	476,143	14.514	58.37	151.18	8,157	3,150
5	City of New York	8,336,817	842.343	302.64	783.83	27,547	10,636	NaN
6	State of New York	19,453,561	1,731.910	47,126.40	122,056.82	412	159	NaN

2. New York City Demographics :
   New York City is the most populous city in the United States. 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
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. Cuisines of New York:

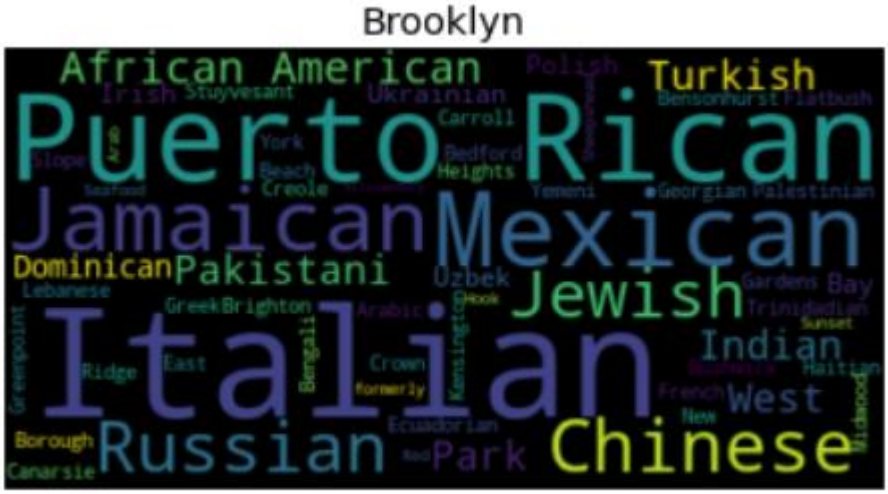



```
<Figure size 2160x3240 with 0 Axes>
```

Most Preferred Food in New York City -

Italian >> Purto Rican >> Mexican >> Jewish >> Indian >> Pakistani >> Dominican

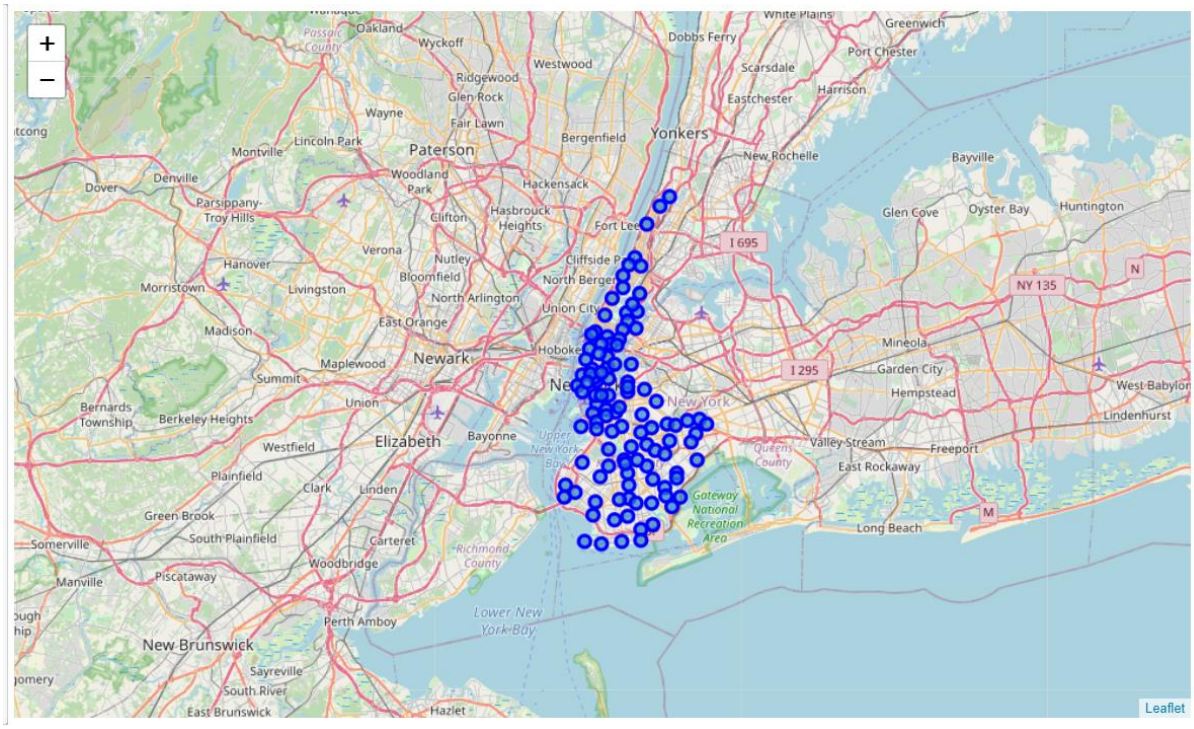
Brooklyn Cuisines: Most famous cuisines in Brooklyn are Italian, Puerto Rican, Mexican



Manhattan Cuisines: Most famous cuisines in Manhattan are Italian, American, Puerto Rican, Indian

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 neighbourhood. We used the Foursquare API data to explore neighbourhoods in New York City.

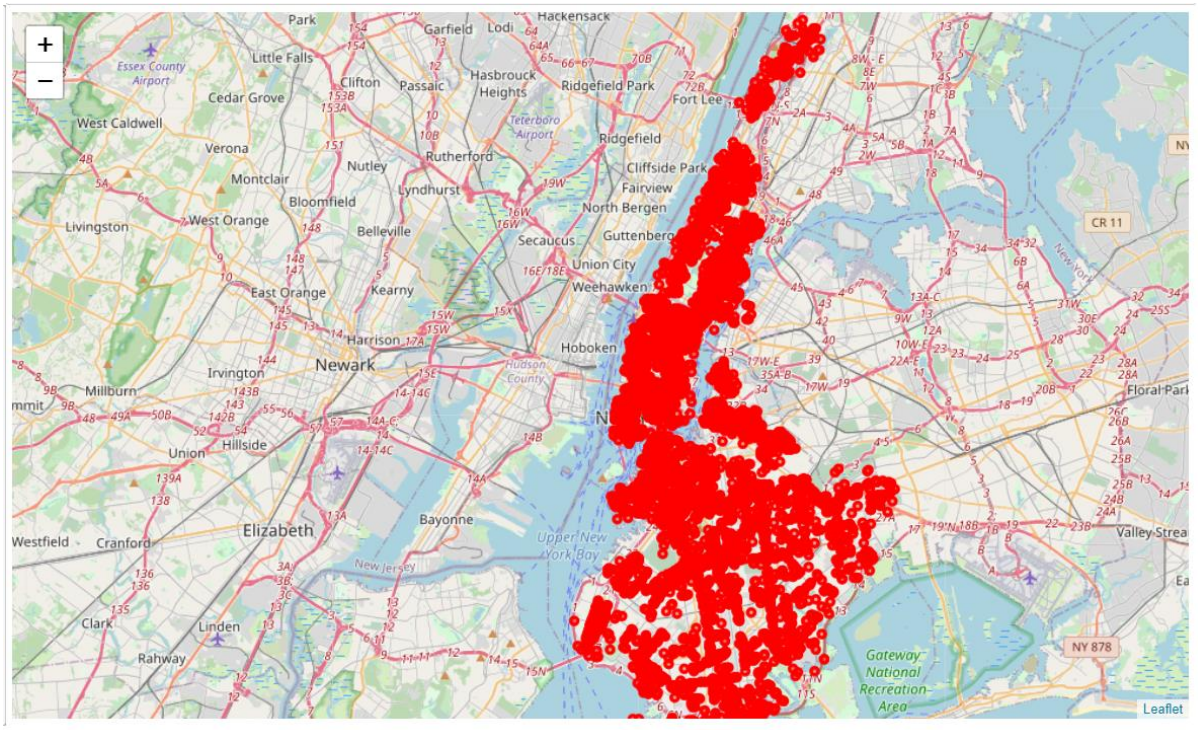
Below is the Brooklyn and Manhattan visualisation of boroughs



Brooklyn and Manhattan Venues:

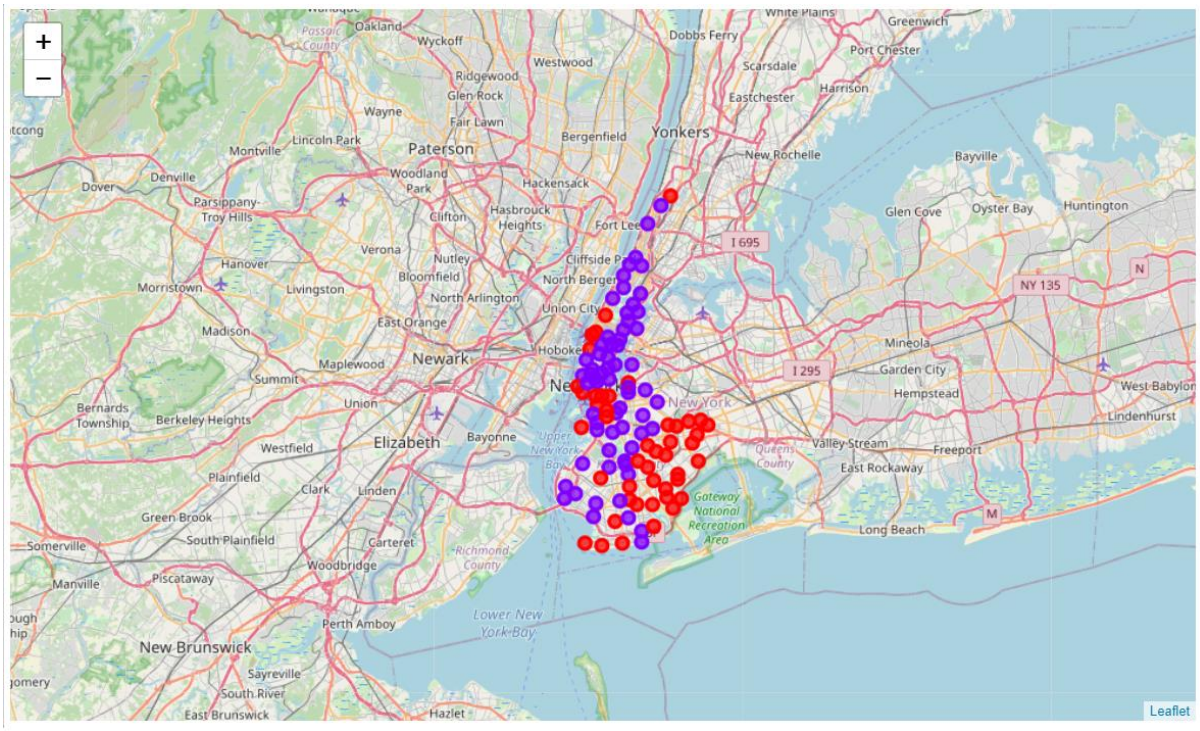
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Marble Hill	40.876551	-73.91066	Bikram Yoga	40.876844	-73.906204	Yoga Studio
1	Marble Hill	40.876551	-73.91066	Arturo's	40.874412	-73.910271	Pizza Place
2	Marble Hill	40.876551	-73.91066	Tibbett 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	The Bronx Public	40.878377	-73.903481	Pub

Brooklyn and Manhattan Venues Visualisation: Generated the below Brooklyn and Manhattan Venues Visualization. It has 9611 venues and 413 unique venue types.



4. **RESULTS** :

From this venues data we filtered and used only the restaurant data for Brooklyn & Manhattan
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.

5. 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 have high concentration of restaurant business. Very competitive market