Coursera Capstone
IBM Applied Data Science Capstone

BY: JAGADEESH NADIMPALLI

JUNE ,2020

Business Problem

- Location of restaurant is one of the most important decisions that will determine whether that will be success or a failure.
- Objective : To analyse and select the best locations in the city of Mumbai to open a new restaurant.
- This project is timely as the city is currently suffering from oversupply of restaurants.
- Business Question
- In what Neighbourhood should I open an restaurant to have the best chance of being successful?

Data Section:

```
In [21]: # check the neighborhoods and the coordinates
         #Mumbai dataset
         Mdf
```

Out[21]:

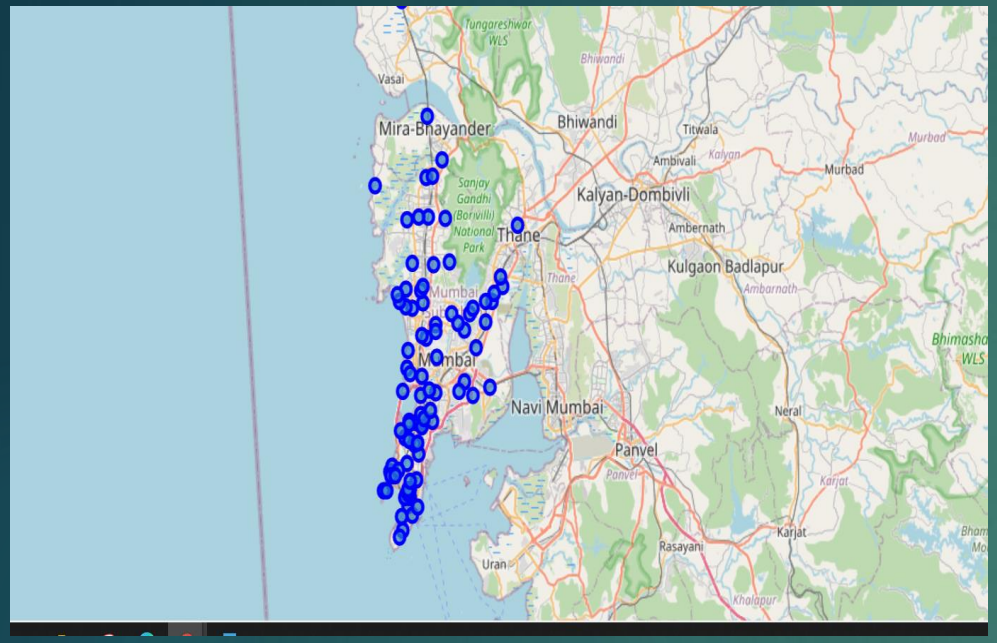
	Neighbourhood	Latitude	Longitude
0	Amboli	19.129060	72.846440
1	Chakala, Andheri	19.108360	72.862330
2	D.N. Nagar	19.125050	72.832480
3	Four Bungalows	19.126320	72.824330
4	Lokhandwala	19.143160	72.824870
5	Marol	19.119050	72.882810
6	Sahar	19.102654	72.862580
7	Seven Bungalows	19.131460	72.816460
8	Versova	19.137690	72.813480
9	Mira Road	19.265705	72.870693
10	Bhayandar	19.307430	72.851840

1. DataSource:https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai
   - Data Description: This data set contains the required information. And we will use this data set to explore various neighbourhoods of Mumbai city.

2.Restaurants in neighbourhood of Mumbai city.

- Data Source: Foursquare API
- Description: By using this API we will get all the venues in neighbourhood. We can filter these venues to get only restaurants.

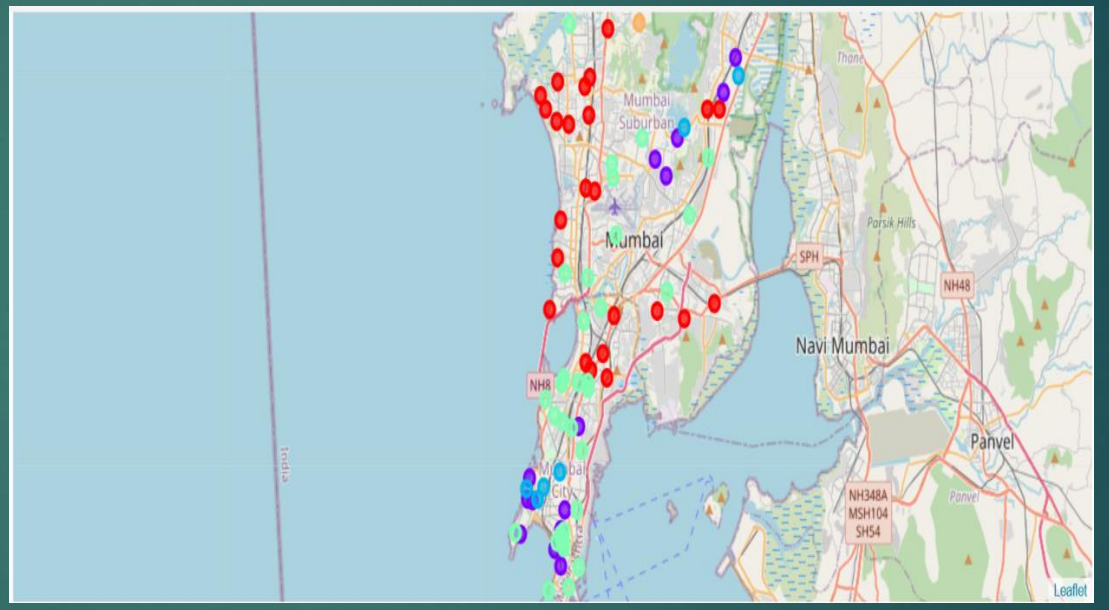
Methodology



- Web scraping Wikipedia page for neighborhoods list.
- Get latitude and longitude coordinates using Geocoder
- Use Foursquare API to get venue data.
- Group data by neighborhood and taking the mean of the frequency of occurrence of each venue category .
- Filter venue category by Restaurant.
- Perform clustering on the data by using k-means clustering .
- Visualize the clusters in a map using Folium.

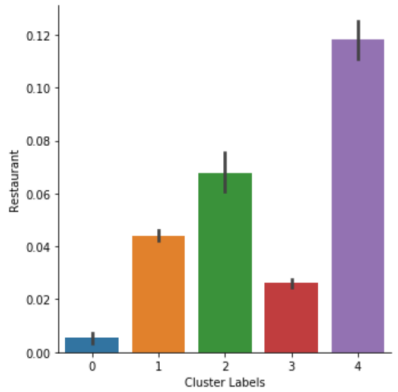
Results

- Categorized the neighbourhoods into 3 clusters :
  - Cluster 0: Neighborhoods with low number to no existence of restaurants.
  - Cluster 1: Neighborhoods with moderate number of restaurants.
  - Cluster 2: Neighborhoods with moderate no of restaurants.
  - Cluster 3: Neighborhoods with moderate no of restaurants
  - Cluster 4: Neighborhoods with high concentration of restaurants



Discussion

```
In [79]: import seaborn as sns
         sns.catplot(x='Cluster Labels', y='Restaurant', data=M_merged, kind='bar');
```



- Most of the restaurants are concentrated in the central area of the Mumbai city.
- Highest number in cluster 4 and moderate number in cluster 2.
- Cluster 0 has very low number to no restaurants in the neighborhoods .
- Oversupply of restaurants mostly happened in the central area of the Mumbai city.
- Property developers with unique selling propositions to stand out from the competition can also open new Restaurants in neighbourhoods in cluster 2 with moderate competition.
- Lastly, property developers are advised to avoid neighbourhoods in cluster 4 which already have high concentration of shopping malls and suffering from intense competition.

Recommendations

- Open new shopping malls in neighborhoods in cluster 0 with little to no competition .
- Can also open in neighborhoods in cluster 1 with moderate competition if have unique selling propositions to stand out from the competition
- Avoid neighborhoods in cluster 4, already high concentration of shopping malls and intense competition

Conclusion:

- Answer to business question: The neighborhoods in cluster 0 are the most preferred locations to open a new restaurant.
- Findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new restaurants.

THANK YOU!