

CAPSTONE BATTLE OF NEIGHBOURHOOD

Opening New Restaurant in Mumbai City

1.Introduction

1.1 Background

Mumbai is the capital city of the Indian state of Maharashtra. According to United Nations, as of 2018, Mumbai was the second-most populous city in India after Delhi and the seventh-most populous city in the world with a population of roughly 20 million. Mumbai is the financial, commercial, and the entertainment capital of India. The trend that comes to stay, are restaurants. Although there are a lot of them spread in the city, there are new ones opening all the time. Therefore, to analyse locations, types, and the number of these restaurants is a plus for those who want to open a new restaurant in the city.

1.2 Problem

Searching an optimal location to open a restaurant in the city of Mumbai can be challenging. One could think that the better location for it should be at a place where there is no restaurant. But the problem is that perhaps most of the interested customers instead of going to an isolated neighbourhood, prefer to go to a popular neighbourhood, where there are more options and also there is movement of people. At the same time that the concurrence will be big in these regions, the flux of interested customers in this specific region will be relevant as well. Many people, for example, go on the weekends to a specific restaurant and when they arrive, there is a large line waiting for them. This usually happens because it is also a new trend in Mumbai, in some popular restaurants, not to have an option to make a reservation. The good news is that perhaps some of the customers, those who do not want wait too long in line, might want to search for similar options in the neighbourhood.

2.Data Description:

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.

Approach

Collect the Mumbai city data from https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai

- Using Foursquare API we will get all venues for each neighborhood.
- Filter out all venues which have Restaurants.
- Analysing using Clustering (Specially K-Means):
- Find the best value of K
- Visualize the neighbourhood with number of Restaurants.
- Compare the Neighbourhoods to Find the Best Place for Starting up a Restaurant
- Inference From these Results and related Conclusion

Problem Statement

- What is the best location for an restaurant in Mumbai City?
- In what Neighbourhood should I open an restaurant to have the best chance of being successful?

3.METHODLOGY:

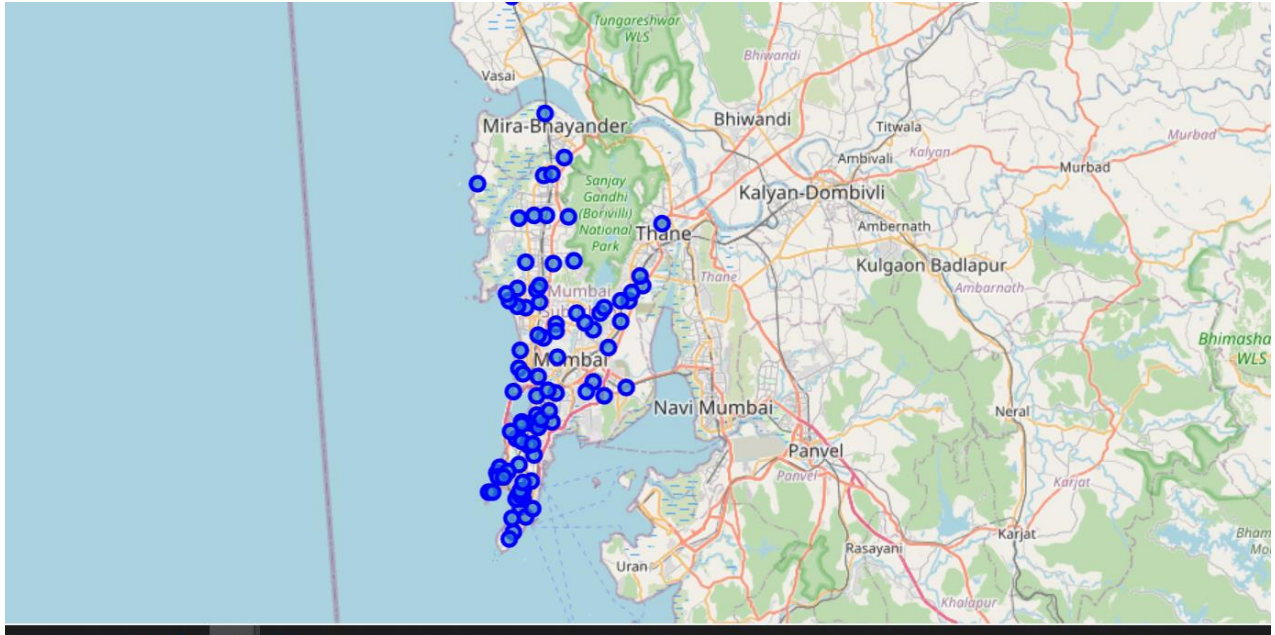
Firstly, we need to get the list of neighbourhoods in the city of Mumbai. Fortunately, the list is available in the Wikipedia page. We will do web scraping using Python requests and beautiful soup packages to extract the list of neighbourhood's data. However, this is just a list of names. We need to get the geographical coordinates in the form of latitude and longitude in order to be able to use Foursquare API. To do so, we will use the wonderful Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude. After gathering the data, we will populate the data into a pandas Data Frame and then visualize the neighbourhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in the city of Mumbai.

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

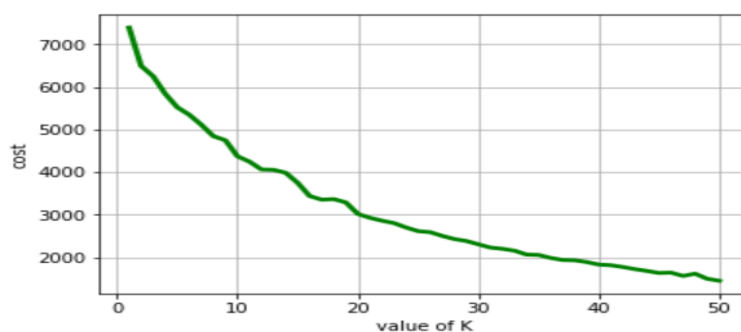
```
Out[21]:
```

	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

Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the "Restaurant" data, we will filter the "Restaurant" as venue category for the neighbourhoods.



Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.



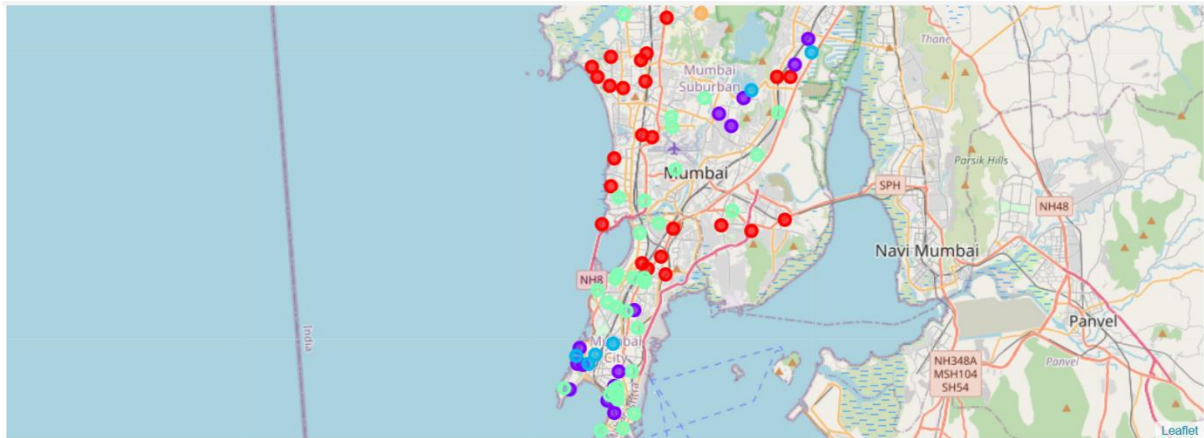
It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 5 clusters based on their frequency of occurrence for “Restaurant”. The results will allow us to identify which neighbourhoods have higher concentration of restaurants while which neighbourhoods have fewer number of restaurants.

4.RESULTS:

The results from the k-means clustering show that we can categorize the neighbourhoods into 5 clusters based on the frequency of occurrence for “Restaurant”:

- Cluster 0: Neighbourhoods with low number to no existence of restaurants
- Cluster 1: Neighbourhoods with moderate number of restaurants

- Cluster 2: Neighbourhoods with moderate number of restaurants
- Cluster 3: Neighbourhoods with moderate number of restaurants
- Cluster 4: Neighbourhoods with high concentration of restaurants

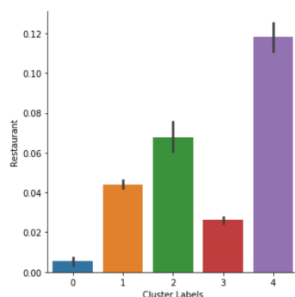


The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, and cluster 2 in mint green colour, cluster 3 in mint green colour, cluster 4 in mint green colour

5.DISSCUSION:

As observations noted from the map in the Results section, Most of the Restaurants are concentrated in the central area of Mumbai city, with the highest number in cluster 4 and moderate number in cluster 1 and cluster 3. other hand, cluster 0 has very low number to totally no restaurants in the neighbourhoods .This represents a great opportunity and high potential areas to open new Restaurants as there is very little to no competition from existing ones. This project recommends property developers to capitalize on these findings to open new Restaurants in neighbourhoods in cluster 0 with little to no competition. 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.

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



Limitations and Suggestions for Future Research In this project, we only consider one factor i.e. frequency of occurrence of restaurants, there are other factors such as population and income of residents that could influence the location decision of a new restaurant. However, to the best knowledge of this researcher such data are not available to the neighbourhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new restaurant. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results.

6.CONCLUSION:

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 5 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers and investors regarding the best locations to open a new restaurant. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 1 are the most preferred locations to open a new restaurant. The 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 restaurant.