# **Segmenting and Clustering Neighborhoods of Mumbai City to find Optimal Location for Opening an Indian Restaurant**

# **Maitri A. Prajapati Final project**

# **D:\Maitri Courses\Capstone Project\Mumbai_Arial_Photo.jpg**

1. **Introduction**

1.1 Background Information:

Mumbai, formerly known as Bombay, is the capital city of Maharashtra. The Greater Mumbai area occupies a long, narrow peninsula in the Arabian Sea on the west coast of India. Mumbai is the most populous city in India and in the top five in the world. While the 2011 census estimated population at 12.4 million. Mumbai is the financial center, economic powerhouse, and industrial hub of India. In 2018, a report by New World Wealth ranked the city as the 12th wealthiest global city with a total private wealth of US Dollar 950 billion.

The Mumbai city is one of the world’s top centers of commerce in terms of financial flow. It is also home to important financial institutions, such as the Reserve Bank of India, the Bombay Stock Exchange, the National Stock Exchange of India, and corporate headquarters of many Indian companies and multinational corporations.

1.2 Description of problem:

Mumbai is rich in its Indian cuisine. It has huge variety of Indian foods from different parts of India. Mumbai seems to be a better place for setting up an Indian restaurant. Since there are lots of restaurants in Mumbai we will try to detect locations that are not already crowded with restaurants. We are also particularly interested in areas with no Indian restaurants in vicinity.

1.3 Target Audience:

Companies or Individuals looking into opening a restaurant would be interested in prediction of optimal location in Mumbai City. This project will provide an analysis whether the venture is feasible or not.

1. **Data acquisition and cleaning**

2.1- Data Sources

First, I search the information of Mumbai neighbourhoods from the various web information sources. Finally, I successfully search the information from following link of Wikipedia.

*https://en.wikipedia.org/wiki/List\_of\_neighbourhoods\_in\_Mumbai*

As per the information given on this web page, Mumbai has a total of 31 boroughs and 93 neighbourhoods. This web page contains one table, in which information of Mumbai neighbourhoods, location, latitude and longitude are given. I consider location as a borough of particular neighbourhood. I converted the html table from the website to a pandas dataframe using webscrapping methods of BeautifulSoup package.

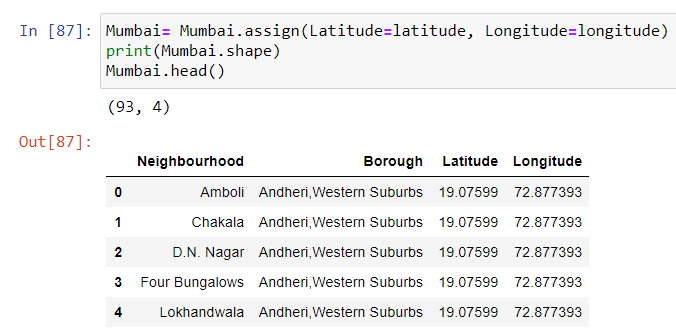
Based on definition of our problem, factors that will influence our decision is the number of existing indian restaurants in the neighbourhood. Number of restaurants and their type and location in every neighbourhood will be obtained using Foursquare API .

2.2- Data Cleaning

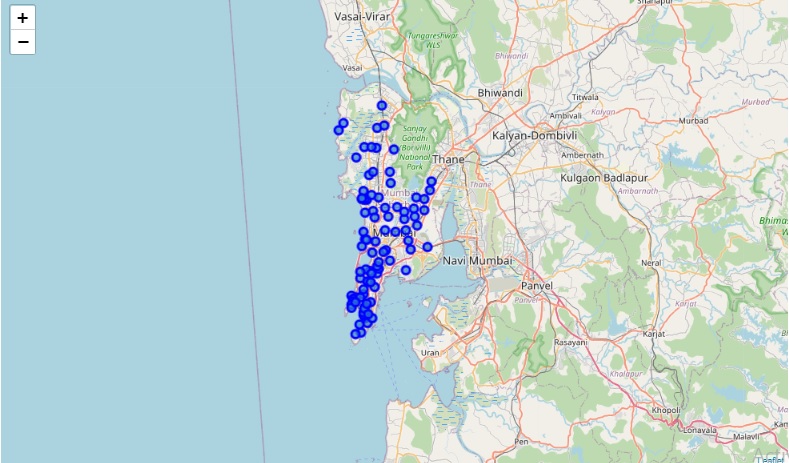
Data downloaded from wikipedia website contains Borough, their Neighbourhoods and latitude and longitude coordinates of each neighbourhood. The details of data cleaning methods are given below.

1. The names of some neighborhoods/borough were found wrong and therefore, such names were corrected.
2. I found some wrong coordinates and therefore I veryfied this information using geopy package of python. Most of the coordinates received using above python package differed from the coordinates given in the website. Therefore, I replaced the coordinates of the dataframe.

Finally, I got the dataframe shown below:



1. **Methodology**

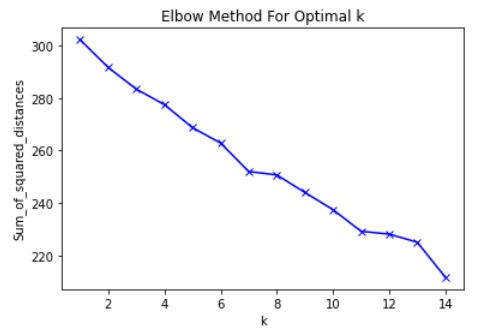
I used python folium library to visualize geographic details of Mumbai and its neighbourhoods and I created a map of Mumbai with neighbourhoods. I used latitude and longitude values to get the visual as shown below:

Then, I utilized the Foursquare API to explore the Neighbourhoods and segment them. I designed the limit as 100 venues and the radius 500 meters for each neighbourhood from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API.



There are 176 unique categories returned by Foursquare API. Then I have made dataset of top venues for each neighbourhood returned by Foursquare API.



I have used unsupervised learning K-means algorithm to cluster the neighbourhoods because We have some common venue categories in neighbourhoods. K-Means algorithm is one of the most common cluster method of unsupervised learning.

First, I Find out optimum k value for k-means clustering. I tried to use elbow method for deciding the number of clusters. But, as shown above the graph is not in clear elbow shape. So, I apporaximately selected number of clusters as 5 and applied K-Means methos to cluster the neighbourhoods into 5 clusters. Then, I created a new dataset that includes the cluster labels as well as the top 5 venues for each neighbourhood.



1. **Result**

Let's examine all the 5 clusters.

## **Cluster 0**

## D:\Maitri Courses\Capstone Project\Maitri Project\new screenshot\cluster 0.jpg

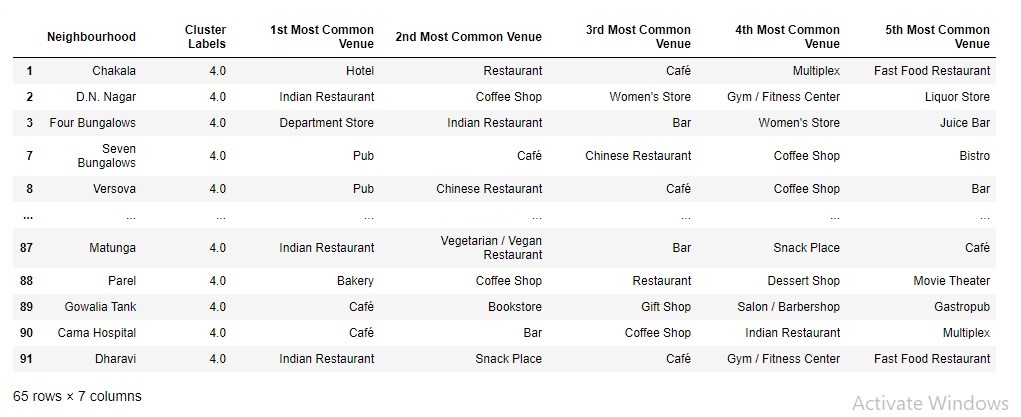
## **Cluster 1**



## **Cluster 2**

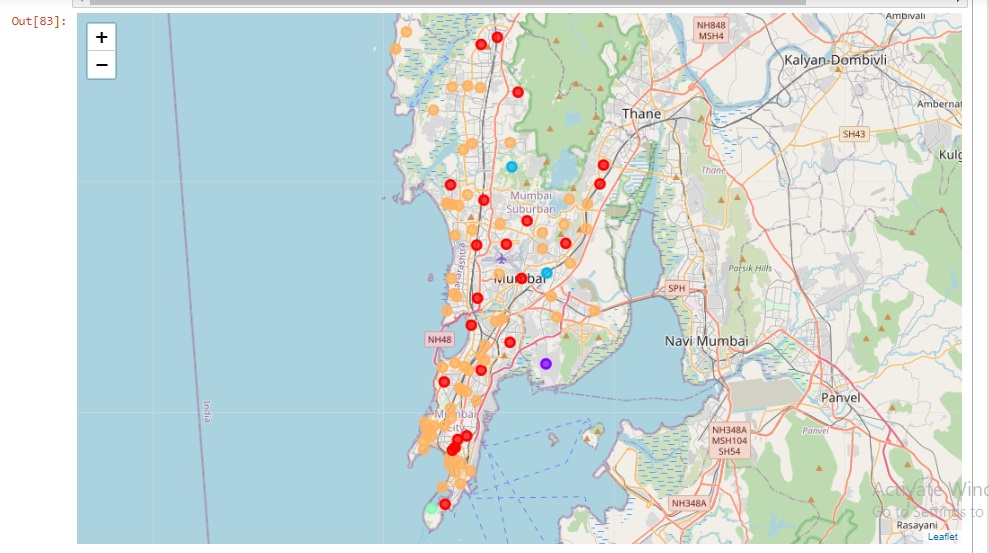
## D:\Maitri Courses\Capstone Project\Maitri Project\new screenshot\cluster 2.jpg

## **Cluster 3**

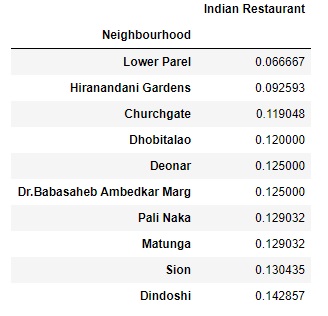


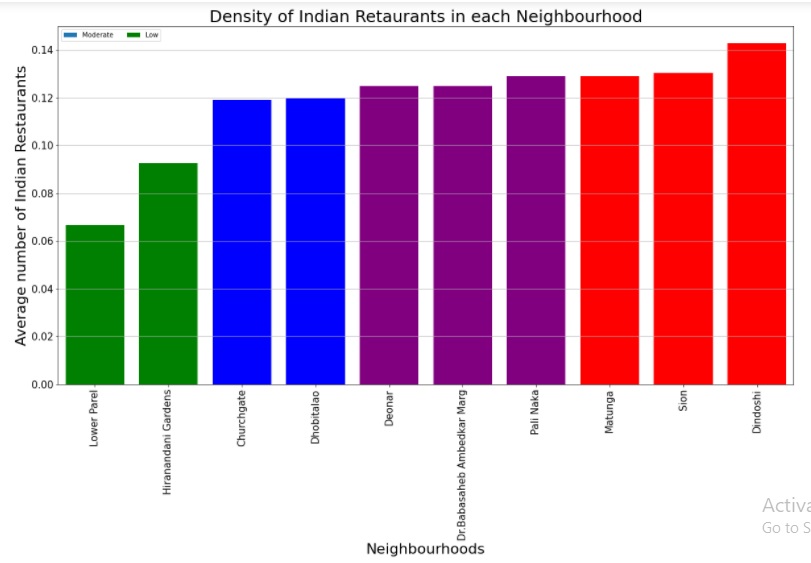
## **Cluster 4**

We can clearly see that Cluster 4 contains highest number of neighbourhoods (65), It contains restaurants in large numbers. In cluster 1, There is only one neighbourhood, but it does not contain indian restaurant top 5 most common venues. Cluster 2 contains only two neighbourhoods but Indian Restaurant is not in top most five common venues. Cluster 3 contains only one neighbourhood but Indian Restaurant is not in top most five common venues. Finally, cluster 0 contains 23 neighborhoods and Indian Restaurant is the first most common venue in most of the neighbourhoods i.e. 20 out of 23 (87%).

The visualized resulting clusters in different colors are as shown below:

For plotting the competition for Indian Restaurant within the each neighbourhood, we framed new dataframe in which we selected only that neighbourhoods where Indian Restaurant is first most common venue. The size of this dataframe is 43 neighbourhoods. We sorted this dataframe by the column average number of Indian Restaurants. Finally, we select first 10 neighbourhoods in which competition for Indian Restaurant is the lowest. Then I have plotted the mean value of Indian restaurants for each selected neighbourhood in the cluster to study the presence of competition in each neighbourhood and visualize them.





1. **Discussion**

The result indicates that among the 10 neighbourhoods, that are selected, i.e., the 10 neighbourhoods that have minimum average number for Indian Restaurants and threfore are most likely to have less competition for Indian Restaurant than the other neighbourhoods in the dataframe, Dindoshi, Sion and Matunga have highest competition for Indian restaurant with average number of restaurant 0.14,0.13 and 0.12 respectively, whereas Hiranandani Gardens and Lower Parel have lowest competition for Indian restaurant with average number of restaurant (0.1 and 0.07 respectively) among the selected 10 neighbourhoods.

From the graph, we can say that the neighbourhoods Dindoshi, Sion and Matunga have the highest competition within the 10 selected neighbourhoods for opening an Indian Restaurants in Mumbai. This indicating that it has the greatest obstacles in opening a new restaurant among the selected neighbourhoods. Dhobitalao and Churchgate; etc. have moderate competition. Following neighbourhoods have lowest competition for opening an Indian Restaurant in Mumbai, which will enable a new business to establish easily:

1. Lower Parel
2. Hiranandani Gardens
3. Churchgate
4. Dhobitalao

Note that these recommendations are based on some assumptions of the analysis, like:

* Radius of the opportunity of each neighbourhood was considered as 500 meters from the location,
* Recommendation opportunities are based on absence of a restaurant which is likely to be appreciated in the top 5 venues.

1. **Conclusion**

This project recommends some of the ideal places to open an Indian restaurant in Mumbai, India. The analysis shows there are better scopes for opening restaurant in Lower Parel and Hiranandani Gardens; Dhobitlao and Churchgate are also good location. This analysis can be helpful for the individuals looking for opening a restaurant or expanding business. This analysis shows the feasible venture and competition landscape of the area.

Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighbourhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location, levels of noise / proximity to major roads, real estate availability, prices, social and economic dynamics of every neighbourhood etc.

**References:**

* Wikipedia: *https://en.wikipedia.org/wiki/List\_of\_neighbourhoods\_in\_Mumbai*
* [Github repository](https://github.com/ashimTarzan/Coursera_Capstone)
* [Foursquare API](https://developer.foursquare.com/)