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

Maitri A. Prajapati Final project

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 Discription of problem:

Mubai 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.

2. 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:

```
In [87]: Mumbai= Mumbai.assign(Latitude=latitude, Longitude=longitude)
         print(Mumbai.shape)
         Mumbai.head()
```

(93, 4)

Out[87]:

	Neighbourhood	Borough	Latitude	Longitude
0	Amboli	Andheri,Western Suburbs	19.07599	72.877393
1	Chakala	Andheri,Western Suburbs	19.07599	72.877393
2	D.N. Nagar	Andheri,Western Suburbs	19.07599	72.877393
3	Four Bungalows	Andheri,Western Suburbs	19.07599	72.877393
4	Lokhandwala	Andheri,Western Suburbs	19.07599	72.877393

3. 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:
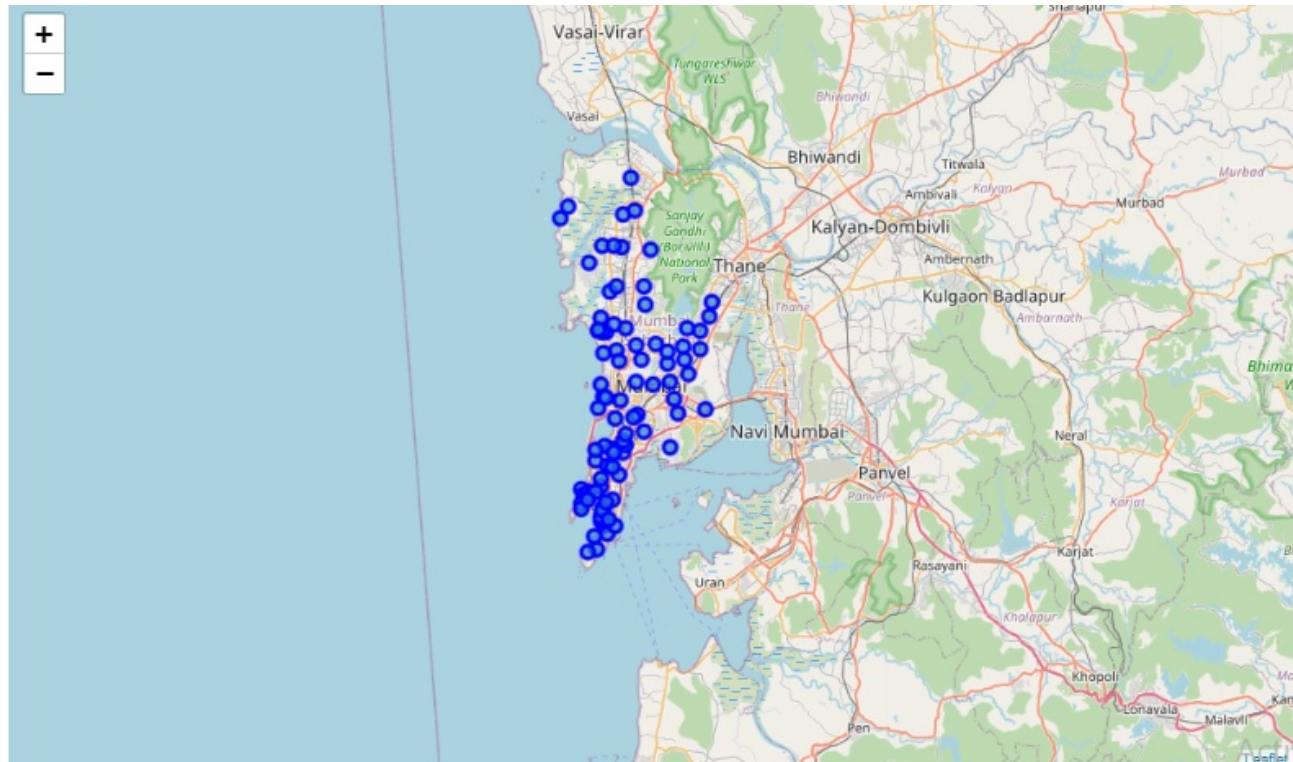
Then, I utilized the Foursquare API to explore the Neighbourhoods and segment them. I designed the limit as 100 venues and the radius 500 meters for each neighbourhood from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API.

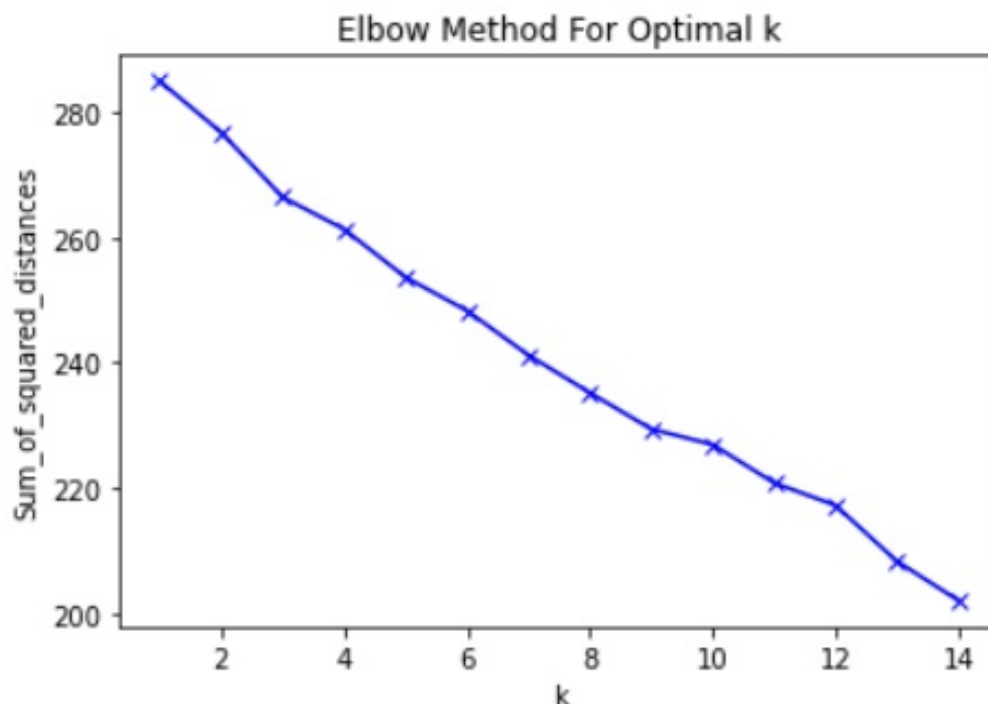|	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Amboli	19.13201	72.849864	5 Spice , Bandra	19.130421	72.847206	Chinese Restaurant
1	Amboli	19.13201	72.849864	Domino's Pizza	19.131000	72.848000	Pizza Place
2	Amboli	19.13201	72.849864	Cafe Arfa	19.128930	72.847140	Indian Restaurant
3	Amboli	19.13201	72.849864	CCD	19.129021	72.849503	Coffee Shop
4	Amboli	19.13201	72.849864	Bostan Restaurant	19.135898	72.847581	Asian Restaurant
...	...	...	...	...	...	...	...
1266	Thane	19.07600	72.878700	Pizza Hut	19.075984	72.877656	Pizza Place
1267	Thane	19.07600	72.878700	Sahara Restaurant	19.079532	72.880152	Mughlai Restaurant
1268	Thane	19.07600	72.878700	Nawab Sheek Corner	19.076933	72.878260	Middle Eastern Restaurant
1269	Thane	19.07600	72.878700	Mexican Olive	19.075981	72.877656	Mexican Restaurant
1270	Thane	19.07600	72.878700	Mughal Flavours	19.079761	72.880302	Indian Restaurant

1271 rows × 7 columns

There are 165 unique categories returned by Foursquare API. Then I have made dataset of top venues for each neighbourhood returned by Foursquare API.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Aarey Milk Colony	Fast Food Restaurant	Yoga Studio	Flower Shop	Fish Market	Field
1	Agripada	Bank	Platform	Coffee Shop	Athletics & Sports	Bakery
2	Altamount Road	Café	Coffee Shop	Concert Hall	Brewery	Cafeteria
3	Amboli	Pizza Place	Indian Restaurant	Asian Restaurant	Coffee Shop	Bakery
4	Amrut Nagar	Indian Restaurant	Electronics Store	Chinese Restaurant	Mediterranean Restaurant	Fast Food Restaurant
...	...	...	...	...	...	...
86	Vidyavihar	Fast Food Restaurant	Cricket Ground	Juice Bar	Comedy Club	Bar
87	Vikhroli	Café	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Seafood Restaurant	Restaurant
88	Vile Parle	Indian Restaurant	Sandwich Place	Fast Food Restaurant	Market	Bakery
89	Walkeshwar	Ice Cream Shop	Food & Drink Shop	Gym	Restaurant	Park
90	Worli	Scenic Lookout	Yoga Studio	Waterfront	Indian Restaurant	Sandwich Place

91 rows × 6 columns

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.

	Neighbourhood	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Amboli	Andheri,Western Suburbs	19.132010	72.849864	0.0	Pizza Place	Indian Restaurant	Asian Restaurant	Coffee Shop	Bakery
1	Chakala	Andheri,Western Suburbs	19.115287	72.861808	0.0	Hotel	Restaurant	Fast Food Restaurant	Multiplex	Indian Restaurant
2	D.N. Nagar	Andheri,Western Suburbs	19.128292	72.830193	0.0	Indian Restaurant	Lounge	Yoga Studio	Chinese Restaurant	Women's Store
3	Four Bungalows	Andheri,Western Suburbs	19.128794	72.825554	0.0	Department Store	Residential Building (Apartment / Condo)	Pub	Fish Market	Market
4	Lokhandwala	Andheri,Western Suburbs	19.143189	72.824081	4.0	Indian Restaurant	Café	Donut Shop	Dessert Shop	Bakery
...	...	...	...	...	...	...	...	...	...	...
88	Parel	South Mumbai	19.009482	72.837661	0.0	Bakery	Food Court	Coffee Shop	Clothing Store	Snack Place
89	Gowalia Tank	Tardeo,South Mumbai	18.962095	72.810098	0.0	Hotel	Gastropub	Deli / Bodega	Concert Hall	Coffee Shop
90	Cama Hospital	South Mumbai	18.942041	72.832259	0.0	Bar	Coffee Shop	Fast Food Restaurant	Café	Multiplex
91	Dharavi	Mumbai	19.044463	72.858618	4.0	Boutique	Snack Place	Indian Restaurant	Café	Yoga Studio
92	Thane	Mumbai	19.076000	72.878700	4.0	Indian Restaurant	Middle Eastern Restaurant	Pizza Place	Mughlai Restaurant	Mexican Restaurant

93 rows × 10 columns

4. Result

Let's examine all the 5 clusters.

Cluster 0

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Amboli	0.0	Pizza Place	Indian Restaurant	Asian Restaurant	Coffee Shop	Bakery
1	Chakala	0.0	Hotel	Restaurant	Fast Food Restaurant	Multiplex	Indian Restaurant
2	D.N. Nagar	0.0	Indian Restaurant	Lounge	Yoga Studio	Chinese Restaurant	Women's Store
3	Four Bungalows	0.0	Department Store	Residential Building (Apartment / Condo)	Pub	Fish Market	Market
7	Seven Bungalows	0.0	Café	Pub	Chinese Restaurant	Coffee Shop	Bar
...	...	...	...	...	...	...	...
80	Dr.Babasaheb Ambedkar Marg	0.0	Plaza	Restaurant	Maharashtrian Restaurant	Lounge	Bar
87	Matunga	0.0	Indian Restaurant	Snack Place	Vegetarian / Vegan Restaurant	Train Station	Bar
88	Parel	0.0	Bakery	Food Court	Coffee Shop	Clothing Store	Snack Place
89	Gowalia Tank	0.0	Hotel	Gastropub	Deli / Bodega	Concert Hall	Coffee Shop
90	Cama Hospital	0.0	Bar	Coffee Shop	Fast Food Restaurant	Café	Multiplex

64 rows × 7 columns

Cluster 1

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
51	Mahul	1.0	ATM	Flower Shop	Fish Market	Field	Fast Food Restaurant

Cluster 2

|	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
31	Naigaon	2.0	Concert Hall	Yoga Studio	Electronics Store	Fish Market	Field

Cluster 3

	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
18	Aarey Milk Colony	3.0	Fast Food Restaurant	Yoga Studio	Flower Shop	Fish Market	Field
46	Vidyavihar	3.0	Fast Food Restaurant	Cricket Ground	Juice Bar	Comedy Club	Bar

Cluster 4

|	Neighbourhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
4	Lokhandwala	4.0	Indian Restaurant	Café	Donut Shop	Dessert Shop	Bakery
5	Marol	4.0	Indian Restaurant	Snack Place	Hotel	Diner	Food
6	Sahar	4.0	Indian Restaurant	Hotel Bar	Hotel	Hockey Arena	Donut Shop
13	Kherwadi	4.0	Indian Restaurant	Café	Pizza Place	Yoga Studio	Electronics Store
30	Kalina	4.0	Food & Drink Shop	Ice Cream Shop	Indian Restaurant	Asian Restaurant	Chinese Restaurant
35	Vile Parle	4.0	Indian Restaurant	Sandwich Place	Fast Food Restaurant	Market	Bakery
36	Bhandup	4.0	Indian Restaurant	Train Station	Track Stadium	Theater	Multiplex
37	Amrut Nagar	4.0	Indian Restaurant	Electronics Store	Chinese Restaurant	Mediterranean Restaurant	Fast Food Restaurant
41	Nehru Nagar	4.0	Ice Cream Shop	Indian Restaurant	Food Court	Seafood Restaurant	Yoga Studio
42	Nahur	4.0	Indian Restaurant	Ice Cream Shop	Restaurant	Yoga Studio	Electronics Store
45	Indian Institute of Technology Bombay campus	4.0	Event Space	Indian Restaurant	Concert Hall	Coffee Shop	Café
48	Chembur	4.0	Indian Restaurant	Seafood Restaurant	Café	Vegetarian / Vegan Restaurant	Restaurant
49	Deonar	4.0	Plaza	Garden	Fast Food Restaurant	Lounge	Indian Restaurant
54	Bhuleshwar	4.0	Indian Restaurant	Arcade	Fast Food Restaurant	Coffee Shop	Food
64	Dongri	4.0	Chinese Restaurant	Indian Restaurant	Vegetarian / Vegan Restaurant	Gym	Donut Shop
68	Mahalaxmi	4.0	Indian Restaurant	Nightclub	Fast Food Restaurant	Gym	Coffee Shop
81	Navy Nagar	4.0	Garden	Yoga Studio	Electronics Store	Fish Market	Field
82	Hindu colony	4.0	Indian Restaurant	Coffee Shop	Café	Gastropub	Juice Bar
83	Ballard Estate	4.0	Indian Restaurant	Irani Cafe	Flea Market	Café	Parsi Restaurant
84	Bhuleshwar	4.0	Indian Restaurant	Arcade	Fast Food Restaurant	Coffee Shop	Food
85	Fanas Wadi	4.0	Yoga Studio	Indian Restaurant	Coffee Shop	Fast Food Restaurant	Food
86	Chor Bazaar	4.0	Indian Restaurant	Dessert Shop	Flea Market	Ice Cream Shop	Antique Shop
91	Dharavi	4.0	Boutique	Snack Place	Indian Restaurant	Café	Yoga Studio
92	Thane	4.0	Indian Restaurant	Middle Eastern Restaurant	Pizza Place	Mughlai Restaurant	Mexican Restaurant

We can clearly see that Cluster 0 contains highest number of neighbourhoods (64), among them only D.N. Nagar and Matunga are the neighbourhoods in which Indian Restaurent is first most common venue exist in this cluster. In cluster 1, There is only one nebourhood in which Indian Restaurant is third most common venue. Cluster 2 contains only one neighbourhood but Indian Restaurant is not in top most five common venues. Cluster 3 contains only two neighbourhoods but Indian Restaurant is not in top most five common venues. Finally, cluster 4 contains 24 neighborhoods and Indian Restaurant is the first most common venue in most of the neighbourhoods i.e. 16 out of 24 (67%).

The visualized resulting clusters in different colors are as shown below:
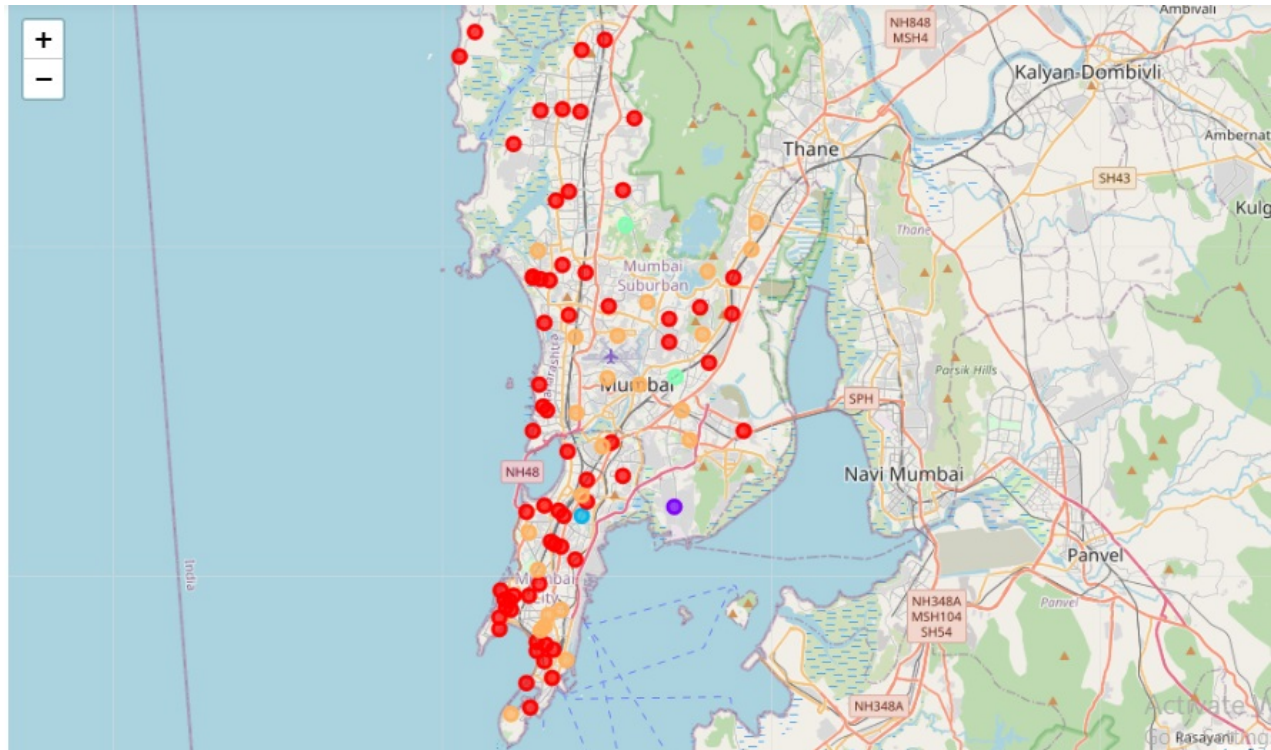
For plotting the competition for Indian Restaurant within the each neighbourhood, we framed new dataframe in which we selected only that neighbourhoods where Indian Restaurant is first most common venue. The size of this dataframe is 38 neighbourhoods. We sorted this dataframe by the column average number 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.

Neighbourhood	Indian Restaurant
Hiranandani Gardens	0.098039
Churchgate	0.111111
D.N. Nagar	0.111111
Dhobitalao	0.148148
Pali Naka	0.156250
Pali Hill	0.156250
Cuffe Parade	0.157895
Matunga	0.160000
Prabhadevi	0.173913
Marine Lines	0.181818

Density of Indian Retaurants in each Neighbourhood
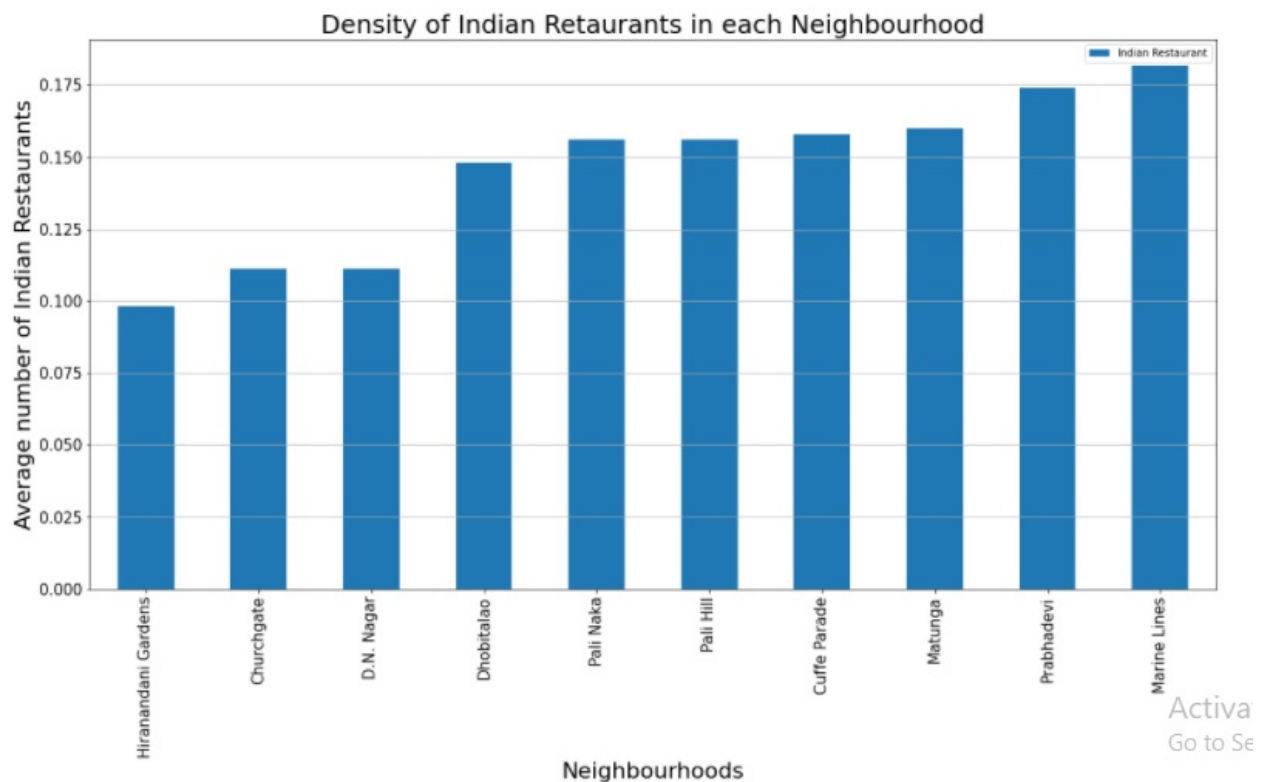
5. Discussion

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, Prabhadevi and Marine Lines have highest competition for Indian restaurant with average number of restaurant 0.36 and 0.43, respectively, whereas Hiranandani Gardens, Church Gate and D.N.Nagar have lowest competition for Indian restaurant with average number of restaurant (approx. 0.1) among the selected 10 neighbourhoods.

From the graph, we can say that the neighbourhoods Prabhadevi and Marine Lines 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. Dhobi Talau and Pali Naka; 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. Hiranandani Gardens
2. Church Gate
3. D.N.Nagar

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.

6. 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 Hiranandani Gardens, Church Gate and D.N.Bagar; Dhobi Talou and Pali Naka 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

- Foursquare API