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 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.

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()
```

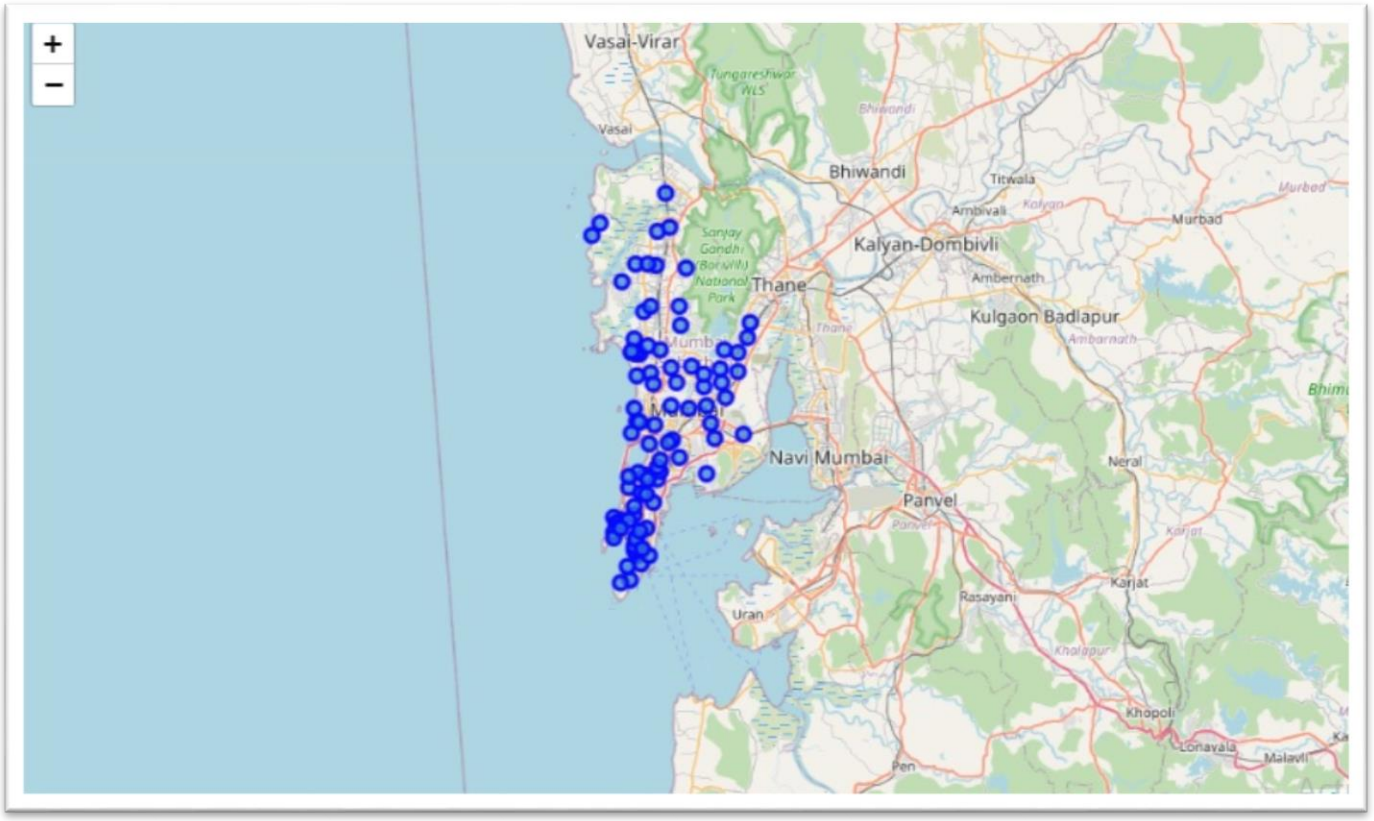
```
(93, 4)
```

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	19.132010	72.849864	5 Spice , Bandra	19.130421	72.847206	Chinese Restaurant
1	Amboli	19.132010	72.849864	Domino's Pizza	19.131000	72.848000	Pizza Place
2	Amboli	19.132010	72.849864	Cafe Arfa	19.128930	72.847140	Indian Restaurant
3	Amboli	19.132010	72.849864	Bostan Restaurant	19.135898	72.847581	Asian Restaurant
4	Chakala	19.115287	72.861808	Courtyard Mumbai International Airport	19.114167	72.864131	Hotel
...	...	...	...	...	...	...	...
1301	Thane	19.076000	72.878700	Sahara Restaurant	19.079532	72.880152	Mughlai Restaurant
1302	Thane	19.076000	72.878700	Nawab Sheek Corner	19.076933	72.878260	Middle Eastern Restaurant
1303	Thane	19.076000	72.878700	Vigour gym	19.075981	72.877656	Gym
1304	Thane	19.076000	72.878700	Mithi Nadi	19.076005	72.874680	River
1305	Thane	19.076000	72.878700	Sahara	19.079696	72.880148	Diner

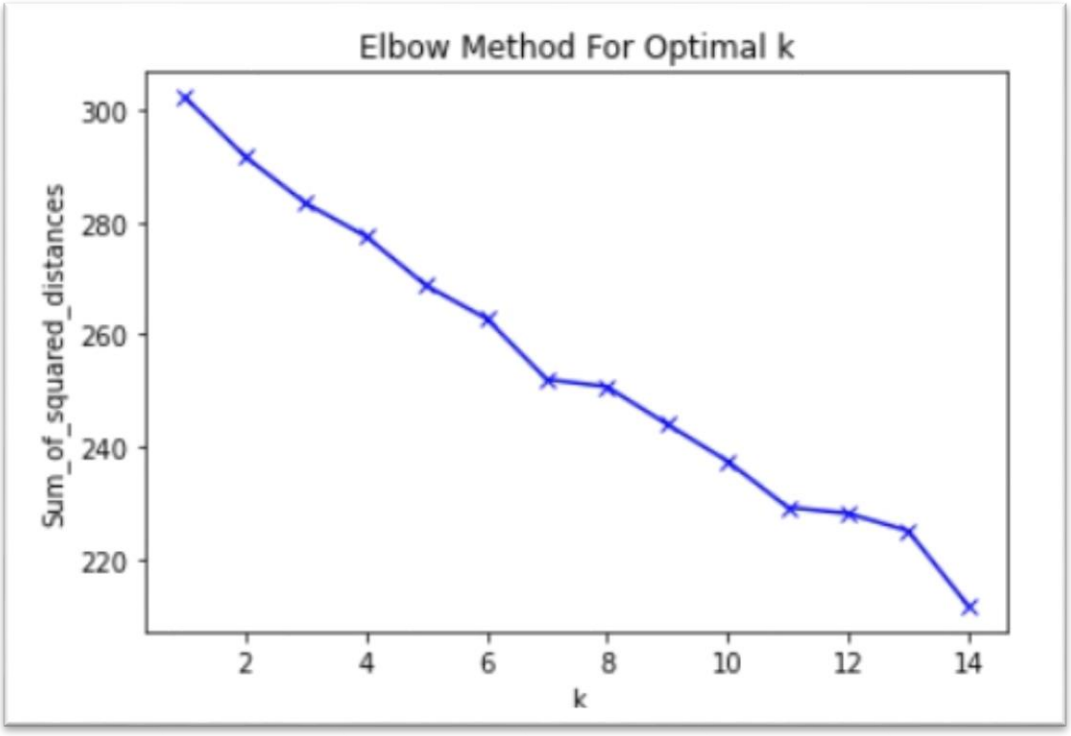
306 rows × 7 columns

There are 176 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	Women's Store	French Restaurant	Food Truck	Food Court
1	Agripada	Coffee Shop	Soccer Field	Platform	Athletics & Sports	Bakery
2	Altamount Road	Café	Coffee Shop	Restaurant	Salon / Barbershop	Bookstore
3	Amboli	Indian Restaurant	Asian Restaurant	Pizza Place	Chinese Restaurant	Field
4	Amrut Nagar	Indian Restaurant	Electronics Store	Steakhouse	Restaurant	Fast Food Restaurant
...	...	...	...	...	...	...
86	Vidyavihar	Fast Food Restaurant	Restaurant	Cricket Ground	Juice Bar	Comedy Club
87	Vikhroli	Café	Restaurant	Racetrack	Vegetarian / Vegan Restaurant	Seafood Restaurant
88	Vile Parle	Indian Restaurant	Fast Food Restaurant	Sandwich Place	Café	Tea Room
89	Walkeshwar	Ice Cream Shop	Restaurant	Food & Drink Shop	Park	Food Truck
90	Worli	Scenic Lookout	Yoga Studio	Paper / Office Supplies Store	Garden	Sports Club

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	Indian Restaurant	Asian Restaurant	Pizza Place	Chinese Restaurant	Field
1	Chakala	Andheri,Western Suburbs	19.115287	72.861808	4.0	Hotel	Restaurant	Café	Multiplex	Fast Food Restaurant
2	D.N. Nagar	Andheri,Western Suburbs	19.128292	72.830193	4.0	Indian Restaurant	Coffee Shop	Women's Store	Gym / Fitness Center	Liquor Store
3	Four Bungalows	Andheri,Western Suburbs	19.128794	72.825554	4.0	Department Store	Indian Restaurant	Bar	Women's Store	Juice Bar
4	Lokhandwala	Andheri,Western Suburbs	19.143189	72.824081	0.0	Indian Restaurant	Café	Bakery	Pizza Place	Cupcake Shop
...	...	...	...	...	...	...	...	...	...	...
88	Parel	South Mumbai	19.009482	72.837661	4.0	Bakery	Coffee Shop	Restaurant	Dessert Shop	Movie Theater
89	Gowalia Tank	Tardeo,South Mumbai	18.962095	72.810098	4.0	Café	Bookstore	Gift Shop	Salon / Barbershop	Gastropub
90	Cama Hospital	South Mumbai	18.942041	72.832259	4.0	Café	Bar	Coffee Shop	Indian Restaurant	Multiplex
91	Dharavi	Mumbai	19.044463	72.858618	4.0	Indian Restaurant	Snack Place	Café	Gym / Fitness Center	Fast Food Restaurant
92	Thane	Mumbai	19.076000	72.878700	0.0	Indian Restaurant	Diner	Middle Eastern Restaurant	River	Pizza Place

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	Indian Restaurant	Asian Restaurant	Pizza Place	Chinese Restaurant	Field
4	Lokhandwala	0.0	Indian Restaurant	Café	Bakery	Pizza Place	Cupcake Shop
5	Marol	0.0	Indian Restaurant	Hotel	Diner	Bus Station	Snack Place
6	Sahar	0.0	Indian Restaurant	Hotel Bar	Hotel	Gym	Farmers Market
13	Kherwadi	0.0	Indian Restaurant	Pizza Place	Café	Tunnel	Gym / Fitness Center
15	I.C. Colony	0.0	Indian Restaurant	Chinese Restaurant	Bar	Liquor Store	Food Truck
17	Dahisar	0.0	Indian Restaurant	Diner	Restaurant	Pizza Place	Train Station
25	Thakur village	0.0	Ice Cream Shop	Indian Restaurant	Pizza Place	Gym	General Entertainment
31	Naigaon	0.0	Indian Restaurant	Gym	Historic Site	Grocery Store	Food Court
35	Vile Parle	0.0	Indian Restaurant	Fast Food Restaurant	Sandwich Place	Café	Tea Room
36	Bhandup	0.0	Indian Restaurant	Train Station	Jewelry Store	Track Stadium	Dessert Shop
37	Amrut Nagar	0.0	Indian Restaurant	Electronics Store	Steakhouse	Restaurant	Fast Food Restaurant
41	Nehru Nagar	0.0	Ice Cream Shop	Indian Restaurant	Playground	Fast Food Restaurant	Seafood Restaurant
42	Nahur	0.0	Indian Restaurant	Bus Station	Restaurant	Ice Cream Shop	Antique Shop
54	Bhuleshwar	0.0	Indian Restaurant	Fast Food Restaurant	Hotel	Arcade	Vegetarian / Vegan Restaurant
60	Cuffe Parade	0.0	Indian Restaurant	Spa	Italian Restaurant	Thai Restaurant	Hotel
64	Dongri	0.0	Indian Restaurant	Smoke Shop	Bar	Field	Food Truck
69	Mahim	0.0	Indian Restaurant	Dessert Shop	Fast Food Restaurant	Intersection	Arcade
79	CGS Colony	0.0	Pizza Place	Yoga Studio	Fast Food Restaurant	Food Truck	Food Court
84	Bhuleshwar	0.0	Indian Restaurant	Fast Food Restaurant	Hotel	Arcade	Vegetarian / Vegan Restaurant
85	Fanas Wadi	0.0	Indian Restaurant	Fast Food Restaurant	Food	Vegetarian / Vegan Restaurant	Jewelry Store
86	Chor Bazaar	0.0	Indian Restaurant	Dessert Shop	BBQ Joint	Market	Restaurant
92	Thane	0.0	Indian Restaurant	Diner	Middle Eastern Restaurant	River	Pizza Place

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	Women's Store	French Restaurant	Food Truck	Food Court

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
18	Aarey Milk Colony	2.0	Fast Food Restaurant	Women's Store	French Restaurant	Food Truck	Food Court
46	Vidyavihar	2.0	Fast Food Restaurant	Restaurant	Cricket Ground	Juice Bar	Comedy Club

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
81	Navy Nagar	3.0	Garden	Women's Store	French Restaurant	Food Truck	Food Court

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
1	Chakala	4.0	Hotel	Restaurant	Café	Multiplex	Fast Food Restaurant
2	D.N. Nagar	4.0	Indian Restaurant	Coffee Shop	Women's Store	Gym / Fitness Center	Liquor Store
3	Four Bungalows	4.0	Department Store	Indian Restaurant	Bar	Women's Store	Juice Bar
7	Seven Bungalows	4.0	Pub	Café	Chinese Restaurant	Coffee Shop	Bistro
8	Versova	4.0	Pub	Chinese Restaurant	Café	Coffee Shop	Bar
...	...	...	...	...	...	...	...
87	Matunga	4.0	Indian Restaurant	Vegetarian / Vegan Restaurant	Bar	Snack Place	Café
88	Parel	4.0	Bakery	Coffee Shop	Restaurant	Dessert Shop	Movie Theater
89	Gowalia Tank	4.0	Café	Bookstore	Gift Shop	Salon / Barbershop	Gastropub
90	Cama Hospital	4.0	Café	Bar	Coffee Shop	Indian Restaurant	Multiplex
91	Dharavi	4.0	Indian Restaurant	Snack Place	Café	Gym / Fitness Center	Fast Food Restaurant

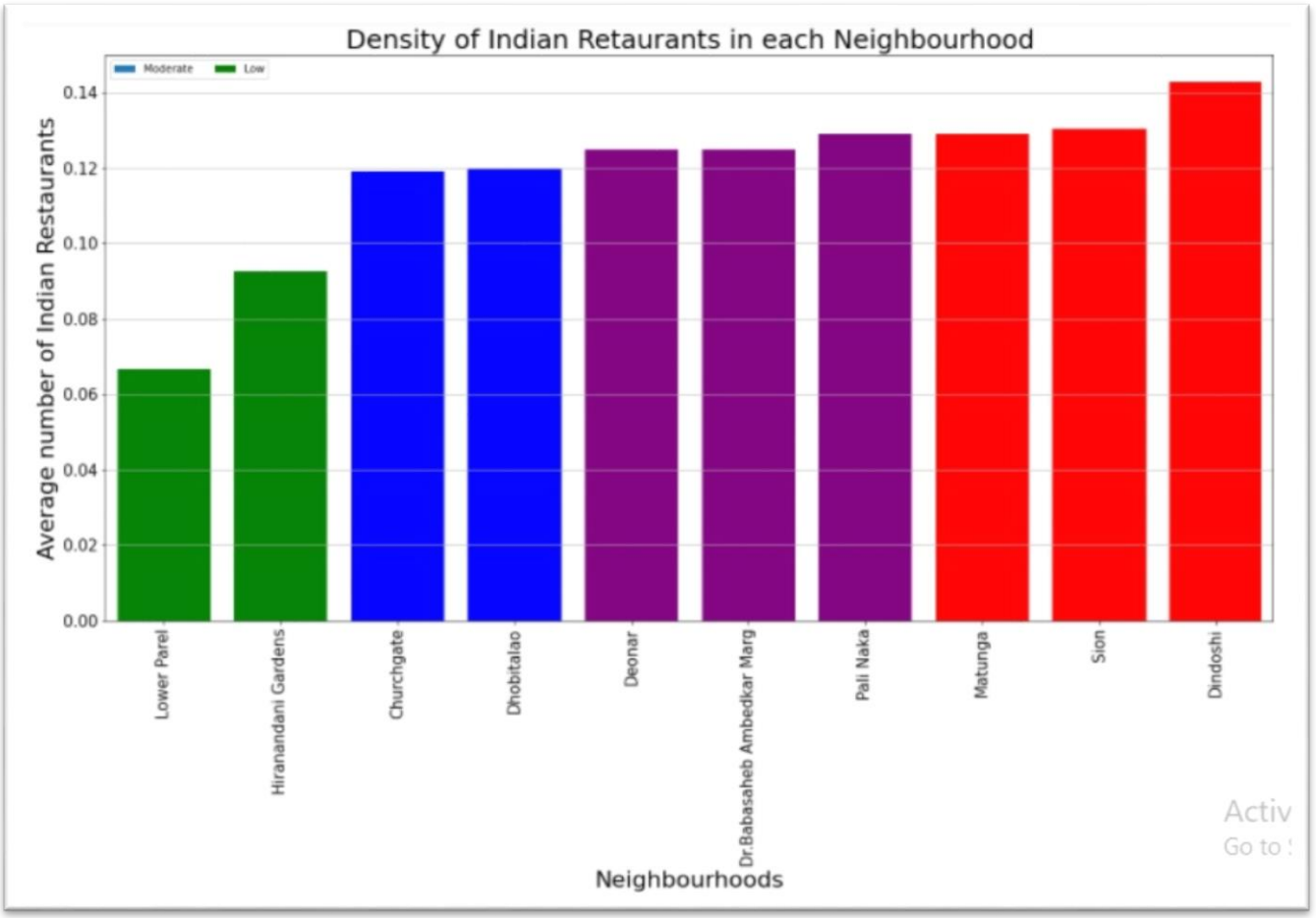
65 rows × 7 columns

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.

Neighbourhood	Indian Restaurant
Lower Parel	0.066667
Hiranandani Gardens	0.092593
Churchgate	0.119048
Dhobitalao	0.120000
Deonar	0.125000
Dr.Babasaheb Ambedkar Marg	0.125000
Pali Naka	0.129032
Matunga	0.129032
Sion	0.130435
Dindoshi	0.142857



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, 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.

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 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
- Foursquare API