THE BATTLE OF NEIGHBORHOODS

MUMBAI

APPLIED DATA SCIENCE CAPSTONE PROJECT

SUBMITTED BY

ANSHU RAJ

Table of Contents

1. **Introduction** ……

2. **Data Collection** ……
   - 2.1. Neighborhoods Data ……
   - 2.2. Geographical Coordinates ……
   - 2.3. Venue Data ……

3. **Methodology** ……
   - 3.1. Data Visulization ……
   - 3.2. Feature Extraction ……

4. **Unsupervised learning** ……

5. **Results and Discussion** ……

6. **Conclusion** ……

7. **Final Comments** ……

INTRODUCTION

The purpose of this project is to study the neighborhoods in Mumbai to determine possible locations for opening a restaurant. This project can be useful for business owners and entrepreneurs who are looking to invest in a restaurant in Mumbai. The main objective of this project is to carefully analyze appropriate data and find recommendations for the stakeholders.

The multi-cultural nature of the city of Mumbai has brought along with it numerous cuisines from all over the world. The people of India generally love food and try different cuisines and experience different flavors.

This Project aim to create an analysis of features for business owners and entrepreneurs who are looking to invest in a restaurant in Mumbai to search a best neighborhood as a comparative analysis between neighborhoods.

Problem which tried to solve:

The major purpose of this project, is to suggest a better neighborhood in Mumbai for entrepreneurs who are looking forward to open a restaurant.This project will help to determine the best neighborhood for a particular type of cusine.

The Location:

Mumbai is the financial capital of India and is one of the most densely populated cities in the world. It lies on the west coast of India and attracts heavy tourism from all over the globe every year.It is one of the major hubs of the world and is extremely diverse.

Data Collection

The following data is required for the project:

1. Neighborhood Data.
2. Geographical coordinates of Mumbai and all Neighborhoods of Mumbai.
3. Venue Data for neighborhoods in Mumbai.

Neighborhood Data:

The data of the neighborhoods in Mumbai was scraped from Wikipedia. Link to the Wikipedia page is https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Mumbai.

The data is read into a pandas data frame using the read_html() method. The main reason for doing so is that the Wikipedia page provides comprehensive and detailed table of the data which can be easily scraped using the read_html() method of pandas library.

The top 10 rows of the dataframe are:

	Neighborhood	Location	Latitude	Longitude
0	Amboli	Andheri,Western Suburbs	19.129300	72.843400
1	Chakala, Andheri	Western Suburbs	19.111388	72.860833
2	D.N. Nagar	Andheri,Western Suburbs	19.124085	72.831373
3	Four Bungalows	Andheri,Western Suburbs	19.124714	72.827210
4	Lokhandwala	Andheri,Western Suburbs	19.130815	72.829270
5	Marol	Andheri,Western Suburbs	19.119219	72.882743
6	Sahar	Andheri,Western Suburbs	19.098889	72.867222
7	Seven Bungalows	Andheri,Western Suburbs	19.129052	72.817018
8	Versova	Andheri,Western Suburbs	19.120000	72.820000
9	Mira Road	Mira-Bhayandar,Western Suburbs	19.284167	72.871111

Geographical Coordinates:

The geographical coordinates for Mumbai has been obtained from the GeoPy library in python. This data is relevant for plotting the map of Mumbai using the Folium library in python. The code for getting the geographical coordinates of Mumbai is :

```
address = 'Mumbai, IN'
geolocator = Nominatim(user_agent="http")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print('The geograpical coordinates of Mumbai are {}, {}.'.format(latitude, longitude))
```

```
The geograpical coordinates of Mumbai are 19.0759899, 72.8773928.
```

The geocoder library in python has been used to obtain latitude and longitude data for various neighborhoods in Mumbai. The coordinates of all neighborhoods in Mumbai are used to check the accuracy of coordinates given on Wikipedia and replace them in our data frame if the absolute difference is more than 0.001. These
refined coordinates are then further used for plotting neighborhoods using the Folium library in python. Figure 3 shows the coordinates of neighborhoods in Mumbai obtained from Wikipedia as 'Latitude' and 'Longitude' and those obtained from geocoder as 'Latitude1' and 'Longitude1'. Furthermore, it also shows the absolute difference between the two latitude columns and the two longitude columns as 'Latdiff' and 'Longdiff', respectively. The dataframe looks like this:

	Neighborhood	Location	Latitude	Longitude	Latitude1	Longitude1	Latdiff	Longdiff
0	Amboli	Western Suburbs	19.1293	72.8464	19.1291	72.8464	0.00024	0.00304
1	Chakala, Andheri	Western Suburbs	19.1084	72.8623	19.1084	72.8623	0.003028	0.001497
2	D.N. Nagar	Western Suburbs	19.1241	72.8325	19.1251	72.8325	0.000965	0.001107
3	Four Bungalows	Western Suburbs	19.1264	72.8242	19.1264	72.8242	0.001666	0.00301
4	Lokhandwala	Western Suburbs	19.1432	72.8249	19.1432	72.8249	0.012345	0.0044
...	...	...	...	...	...	...	...	...
88	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
89	Gowalia Tank	South Mumbai	18.9645	72.8112	18.9645	72.8112	0.00201	0.001467
90	Dava Bazaar	South Mumbai	19.1314	72.927	19.1314	72.927	0.184518	0.095598
91	Dharavi	Mumbai	19.0467	72.8546	19.0467	72.8546	0.006532	0.00376
92	Thane	Mumbai	19.1409	72.8826	19.1409	72.8826	0.0590678	0.0873957

Final Mumbai neighborhoods dataframe after replacing the latitude and longitude values as mentioned before and dropping unnecessary columns looks like this:

	Neighborhood	Location	Latitude	Longitude
0	Amboli	Western Suburbs	19.1293	72.8464
1	Chakala, Andheri	Western Suburbs	19.1084	72.8623
2	D.N. Nagar	Western Suburbs	19.1241	72.8325
3	Four Bungalows	Western Suburbs	19.1264	72.8242
4	Lokhandwala	Western Suburbs	19.1432	72.8249
5	Marol	Western Suburbs	19.1192	72.8827
6	Sahar	Western Suburbs	19.1027	72.8626
7	Seven Bungalows	Western Suburbs	19.1291	72.8212
8	Versova	Western Suburbs	19.1377	72.8135
9	Mira Road	Western Suburbs	19.2656	72.8711

Venue Data:

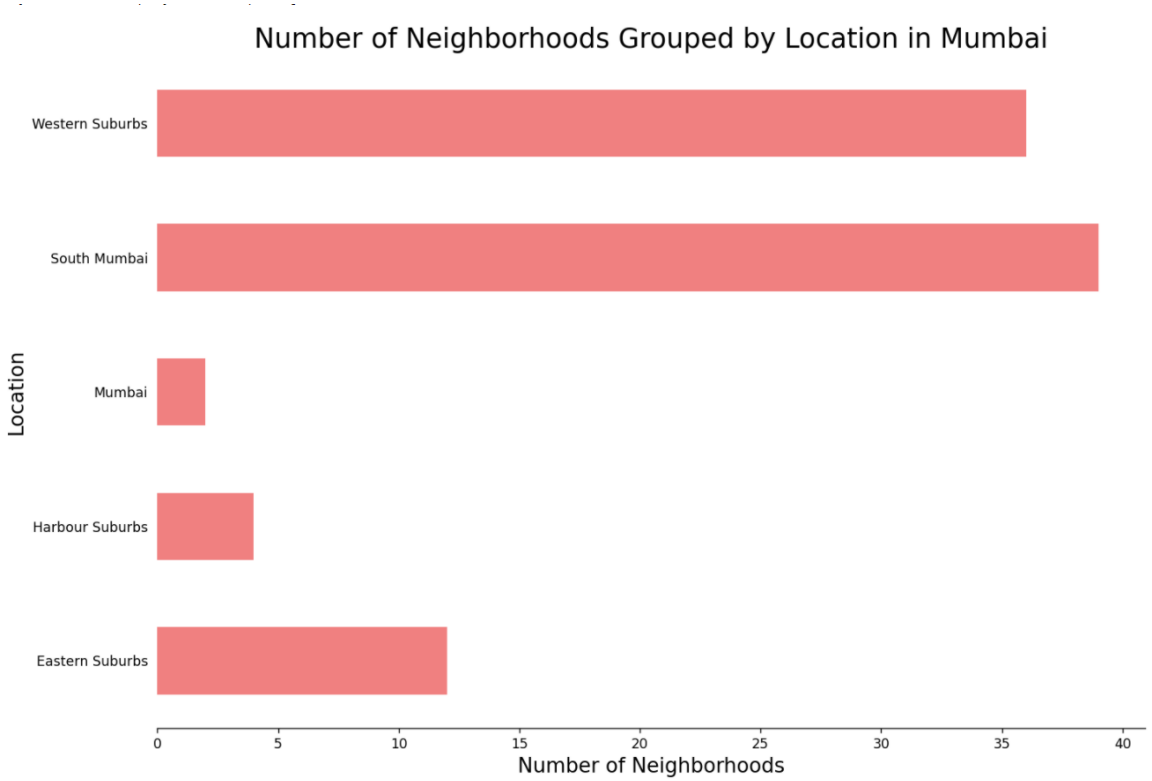
The venue data has been extracted using the Foursquare API. This data contains venue recommendations for all neighborhoods in Mumbai and is used to study the popular venues of different neighborhoods as well as build the unsupervised learning model to cluster neighborhoods. The venue recommendations of all neighborhoods were obtained with a limit of 200, that is, maximum of 200 venue recommendations per neighborhood and a radius of 1 km around the neighborhood's geographical coordinates. Following figure shows the top 10 rows depicting the results obtained after cleaning the data from Foursquare API.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Amboli	19.1293	72.84644	Cafe Arfa	19.128930	72.847140	Indian Restaurant
1	Amboli	19.1293	72.84644	5 Spice , Bandra	19.130421	72.847206	Chinese Restaurant
2	Amboli	19.1293	72.84644	Domino's Pizza	19.131000	72.848000	Pizza Place
3	Amboli	19.1293	72.84644	Jaffer Bhai's Delhi Darbar	19.137714	72.845909	Mughlai Restaurant
4	Amboli	19.1293	72.84644	Narayan Sandwich	19.121398	72.850270	Sandwich Place
5	Amboli	19.1293	72.84644	Persia Darbar	19.136952	72.846822	Indian Restaurant
6	Amboli	19.1293	72.84644	Garden Court	19.127188	72.837478	Indian Restaurant
7	Amboli	19.1293	72.84644	Subway	19.127860	72.844461	Sandwich Place
8	Amboli	19.1293	72.84644	Kamal Chhaya Bar	19.128245	72.837610	Bar
9	Amboli	19.1293	72.84644	Domino's Pizza	19.130000	72.837000	Pizza Place

METHODOLOGY

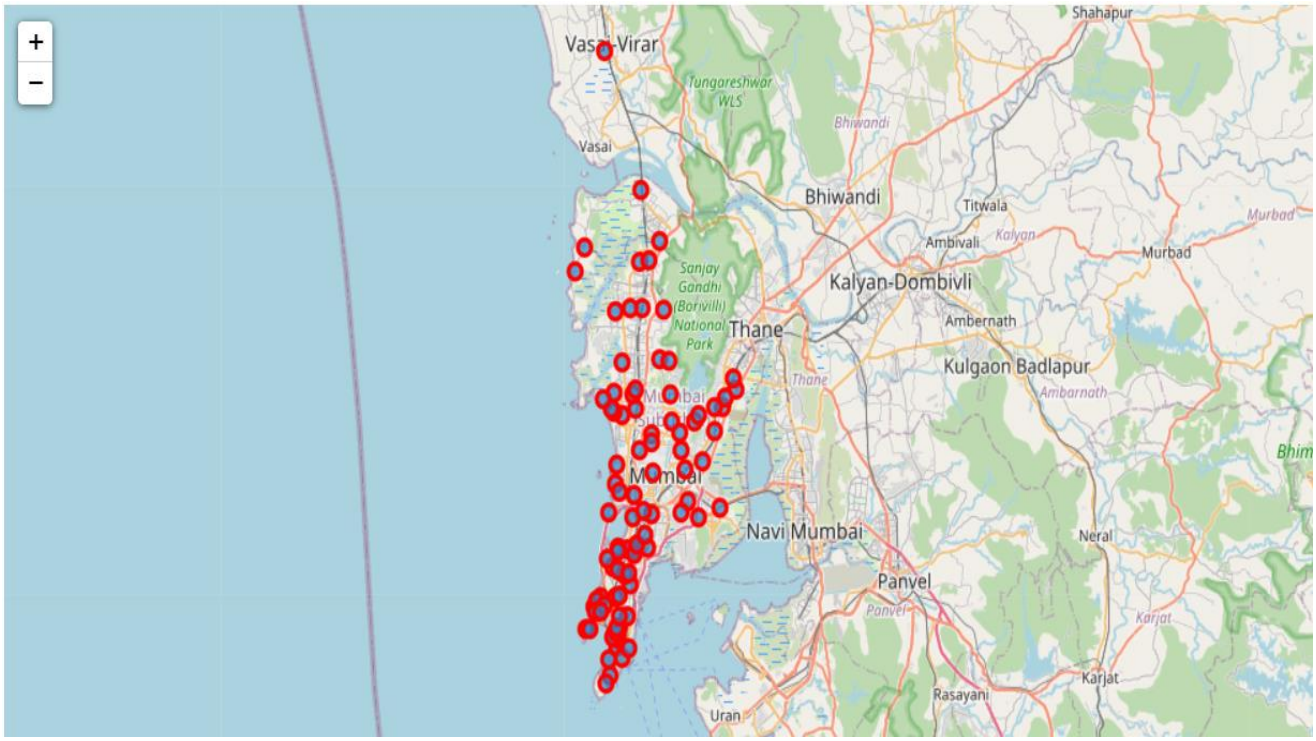
Data Visualization:

In order to understand the data obtained for Mumbai neighborhoods, basic visualization was carried out. Following figure shows a bar plot depicting the number of neighborhoods in each location in Mumbai.



It is evident from above figure that South Mumbai and Western Suburbs have the most number of neighborhoods. Notice how we see one of the locations as Mumbai itself? This is because the neighborhoods contained in this location are located at the outskirts of the city and thus have been termed as just Mumbai.

Using folium, a map was plotted to show how the different neighborhoods are spread all across Mumbai.



Feature Extraction:

Feature extraction was carried out to obtain features from the Foursquare API data (as shown in Figure 5) which was used for building the unsupervised learning model. In order to achieve this, the “Venue Category” column had to be converted to some form of numeric value to be used for building the model. This was achieved by the One-hot Encoding method which takes all the unique categories and creates a column for each category. Then, if a neighborhood venue belongs to that category, it would get a value of 1 for that row in that specific category column and if a neighborhood venue does not belong to the particular category, the value would be 0.

This process was repeated for all venues in all neighborhoods and the result was a sparse matrix containing the neighborhood name and all unique category columns with either 1 or 0 based on whether the neighborhood venue belonged to that category or not. This dataframe was then grouped by the neighborhood name and the average value was taken for all categories. The result is shown in following figure which shows only the top 10 rows.

	Neighborhood	Accessories Store	Airport Lounge	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Arcade	Art Gallery	Arts & Crafts Store	...	Trail	Train	Train Station	Vegetarian / Vegan Restaurant	Waterfront	V E
0	Amboli	0.000000	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0
1	Chakala, Andheri	0.000000	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.058824	0.0	0
2	D.N. Nagar	0.023256	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.046512	0.0	0
3	Four Bungalows	0.019608	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.039216	0.0	0
4	Lokhandwala	0.000000	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.011628	0.0	0
5	Marol	0.000000	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0
6	Sahar	0.000000	0.0	0.05	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0
7	Seven Bungalows	0.019231	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.038462	0.0	0
8	Versova	0.000000	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.027778	...	0.0	0.0	0.0	0.000000	0.0	0
9	Mira Road	0.000000	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0

10 rows x 213 columns

Notice that most of the values are 0 since there were a large number of unique categories and not all neighborhoods had venues belonging to each category.

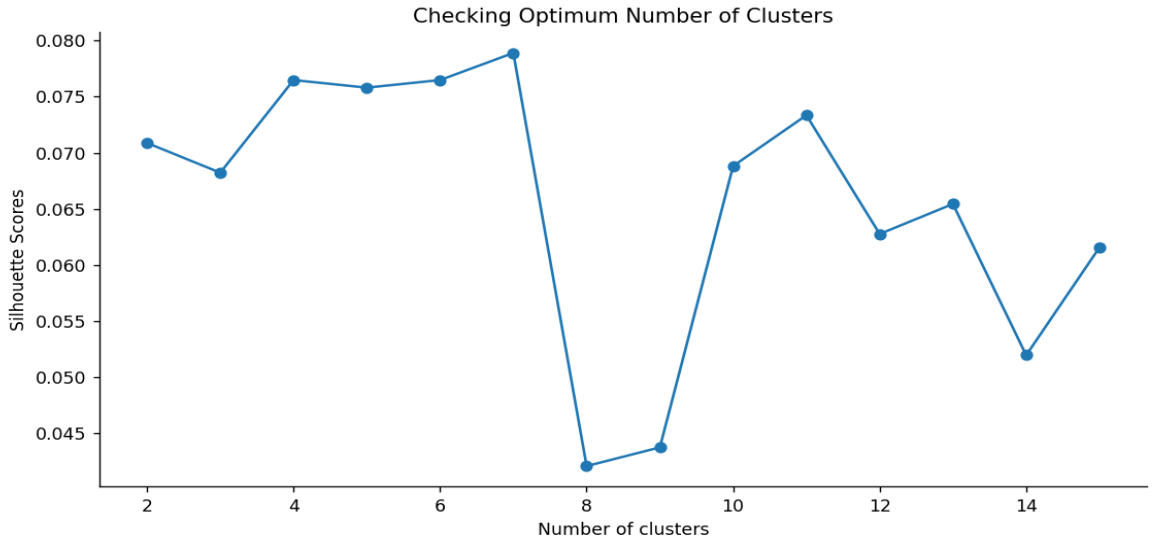
This data was used for the unsupervised learning model with the neighborhood name dropped. The unsupervised learning model is explained in the next section.

A dataframe was also created which contained the top 10 most common venues of all neighborhoods. Though this is not a part of Feature Extraction, it is important to provide a glimpse into what this dataframe looks like as it will be used later to combine the results from the unsupervised learning model. The top 10 rows of this dataframe are shown in Figure

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Amboli	Indian Restaurant	Asian Restaurant	Bar	Sandwich Place	Coffee Shop	Pizza Place	Athletics & Sports	Metro Station	Chinese Restaurant	Mughlai Restaurant
1	Chakala, Andheri	Hotel	Indian Restaurant	Café	Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	Asian Restaurant	Multiplex	Burger Joint	Spa
2	D.N. Nagar	Indian Restaurant	Pizza Place	Bar	Pub	Coffee Shop	Vegetarian / Vegan Restaurant	Snack Place	Falafel Restaurant	Sports Club	Department Store
3	Four Bungalows	Pub	Chinese Restaurant	Indian Restaurant	Seafood Restaurant	Ice Cream Shop	Gym / Fitness Center	Vegetarian / Vegan Restaurant	Bar	Lounge	Pizza Place
4	Lokhandwala	Indian Restaurant	Café	Coffee Shop	Chinese Restaurant	Fast Food Restaurant	Bar	Gym / Fitness Center	Pub	Lounge	Italian Restaurant
...	...	...	...	...	...	...	...	...	...	...	...
88	Parel	Indian Restaurant	Coffee Shop	Maharashtrian Restaurant	Cafeteria	Sporting Goods Shop	Restaurant	Plaza	Playground	Chinese Restaurant	Bar
89	Gowalia Tank	Indian Restaurant	Fast Food Restaurant	Bakery	Café	Coffee Shop	Sandwich Place	Snack Place	Dessert Shop	Vegetarian / Vegan Restaurant	Pizza Place
90	Dava Bazaar	Train Station	Indian Restaurant	Café	Fish Market	Asian Restaurant	Beer Garden	Multiplex	Coffee Shop	Hotel	Falafel Restaurant
91	Dharavi	Indian Restaurant	Café	Coffee Shop	Music Venue	Sandwich Place	Seafood Restaurant	Shoe Store	Fast Food Restaurant	Lake	Juice Bar
92	Thane	Performing Arts Venue	Snack Place	Pizza Place	Fast Food Restaurant	Café	Department Store	Farmers Market	Falafel Restaurant	Event Space	Electronics Store

UNSUPERVISED LEARNING

K-means unsupervised learning technique was used to cluster the neighborhoods based on the category of venues near the neighborhoods. One important aspect of the k-means model is to determine the number of clusters to use in model development. This was determined by the Silhouette score which was calculated for a range of clusters from 2 to 15. The resulting number of clusters and their respective Silhouette scores are shown in Figure



It is evident that the Silhouette scores are not very high even as the number of clusters increases. This means that the inter-cluster distance is not very high over the range of k-values. Despite this, the data will be clustered to the best possible extent. For this, 5 clusters will be used for the k-means clustering model since it provides the highest silhouette score.

RESULTS AND DISCUSSION

The clustering model then clusters the neighborhoods in Mumbai and provides a label for each neighborhood which is representative of the cluster it belongs to. The cluster labels were then added to the dataframe in Figure along with the Location, Latitude, and Longitude columns to provide a complete summary of the clustering.

	Neighborhood	Location	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Co Ve
0	Amboli	Western Suburbs	19.1293	72.8464	2	Indian Restaurant	Asian Restaurant	Bar	Sandwich Place	Coffee Shop	Pizza Place	Athletics & Sports	Metro Station	Ch Re
1	Chakala, Andheri	Western Suburbs	19.1084	72.8623	1	Hotel	Indian Restaurant	Café	Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	Asian Restaurant	Multiplex	Bu Joi
2	D.N. Nagar	Western Suburbs	19.1241	72.8325	2	Indian Restaurant	Pizza Place	Bar	Pub	Coffee Shop	Vegetarian / Vegan Restaurant	Snack Place	Falafel Restaurant	Sp Cl
3	Four Bungalows	Western Suburbs	19.1264	72.8242	2	Pub	Chinese Restaurant	Indian Restaurant	Seafood Restaurant	Ice Cream Shop	Gym / Fitness Center	Vegetarian / Vegan Restaurant	Bar	Lo
4	Lokhandwala	Western Suburbs	19.1432	72.8249	2	Indian Restaurant	Café	Coffee Shop	Chinese Restaurant	Fast Food Restaurant	Bar	Gym / Fitness Center	Pub	Lo
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
88	Parel	South Mumbai	18.9957	72.84	1	Indian Restaurant	Coffee Shop	Maharashtrian Restaurant	Cafeteria	Sporting Goods Shop	Restaurant	Plaza	Playground	Ch Re
89	Gowalia Tank	South Mumbai	18.9645	72.8112	2	Indian Restaurant	Fast Food Restaurant	Bakery	Café	Coffee Shop	Sandwich Place	Snack Place	Dessert Shop	Ve / V Re
90	Dava Bazaar	South Mumbai	19.1314	72.927	0	Train Station	Indian Restaurant	Café	Fish Market	Asian Restaurant	Beer Garden	Multiplex	Coffee Shop	Ho
91	Dharavi	Mumbai	19.0467	72.8546	2	Indian Restaurant	Café	Coffee Shop	Music Venue	Sandwich Place	Seafood Restaurant	Shoe Store	Fast Food Restaurant	Lal
92	Thane	Mumbai	19.1409	72.8826	2	Performing Arts Venue	Snack Place	Pizza Place	Fast Food Restaurant	Café	Department Store	Farmers Market	Falafel Restaurant	Ev Sp

Furthermore, neighborhoods in each individual cluster can be extracted using cluster labels and thus the details of specific clusters can be seen. This is done below for all clusters with only 10 rows for clusters that contain a high number of neighborhoods.

Cluster 1

	Neighborhood	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
36	Bhandup	Eastern Suburbs	Train Station	Fast Food Restaurant	Indian Restaurant	Asian Restaurant	Bakery	Zoo	Dhaba	Fish & Chips Shop	Field	Farmers Market
40	Kanjurmarg	Eastern Suburbs	Train Station	Multiplex	Hotel	Gym	Chinese Restaurant	Asian Restaurant	Gift Shop	Zoo	Donut Shop	Dim Sum Restaurant
42	Nahur	Eastern Suburbs	Coffee Shop	Train Station	Bus Station	Restaurant	Pub	Indian Restaurant	Fast Food Restaurant	Pizza Place	Convention Center	Community Center
50	Mankhurd	Harbour Suburbs	Coffee Shop	Train Station	Sports Bar	Bus Station	Zoo	Dhaba	Field	Fast Food Restaurant	Farmers Market	Falafel Restaurant
59	Cotton Green	South Mumbai	Plaza	Fast Food Restaurant	Pizza Place	Snack Place	Vegetarian / Vegan Restaurant	Train Station	Zoo	Donut Shop	Dhaba	Dim Sum Restaurant
90	Dava Bazaar	South Mumbai	Train Station	Indian Restaurant	Café	Fish Market	Asian Restaurant	Beer Garden	Multiplex	Coffee Shop	Hotel	Falafel Restaurant

Cluster 2

	Neighborhood	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
1	Chakala, Andheri	Western Suburbs	Hotel	Indian Restaurant	Café	Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	Asian Restaurant	Multiplex	Burger Joint	Spa
5	Marol	Western Suburbs	Indian Restaurant	Chinese Restaurant	Restaurant	Diner	Hotel	Coffee Shop	Farmers Market	Ice Cream Shop	Dance Studio	Electronics Store
6	Sahar	Western Suburbs	Hotel	Indian Restaurant	Café	Bakery	Asian Restaurant	Airport Terminal	Hotel Pool	Pizza Place	Italian Restaurant	Coffee Shop
9	Mira Road	Western Suburbs	Indian Restaurant	Mexican Restaurant	Fast Food Restaurant	Garden	Movie Theater	General College & University	Restaurant	Bar	Shopping Mall	Coffee Shop
20	Jogeshwari West	Western Suburbs	Indian Restaurant	Dessert Shop	Chinese Restaurant	Gym	Café	Men's Store	Mughlai Restaurant	Ice Cream Shop	Seafood Restaurant	Clothing Store
21	Juhu	Western Suburbs	Indian Restaurant	Movie Theater	Coffee Shop	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Café	Convention Center	Plaza	Breakfast Spot	Restaurant
23	Poisar	Western Suburbs	Indian Restaurant	Gym / Fitness Center	Train Station	Men's Store	Mexican Restaurant	Fast Food Restaurant	Mobile Phone Shop	Food	Electronics Store	Snack Place
29	Sunder Nagar	Western Suburbs	Indian Restaurant	Coffee Shop	Movie Theater	Fast Food Restaurant	Café	Pizza Place	Breakfast Spot	Bakery	Restaurant	Train Station

Cluster 3

	Neighborhood	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Amboli	Western Suburbs	Indian Restaurant	Asian Restaurant	Bar	Sandwich Place	Coffee Shop	Pizza Place	Athletics & Sports	Metro Station	Chinese Restaurant	Mughlai Restaurant
2	D.N. Nagar	Western Suburbs	Indian Restaurant	Pizza Place	Bar	Pub	Coffee Shop	Vegetarian / Vegan Restaurant	Snack Place	Falafel Restaurant	Sports Club	Department Store
3	Four Bungalows	Western Suburbs	Pub	Chinese Restaurant	Indian Restaurant	Seafood Restaurant	Ice Cream Shop	Gym / Fitness Center	Vegetarian / Vegan Restaurant	Bar	Lounge	Pizza Place
4	Lokhandwala	Western Suburbs	Indian Restaurant	Café	Coffee Shop	Chinese Restaurant	Fast Food Restaurant	Bar	Gym / Fitness Center	Pub	Lounge	Italian Restaurant
7	Seven Bungalows	Western Suburbs	Bar	Pub	Café	Seafood Restaurant	Pizza Place	Chinese Restaurant	Vegetarian / Vegan Restaurant	Indian Restaurant	Ice Cream Shop	Coffee Shop
8	Versova	Western Suburbs	Ice Cream Shop	Café	Beach	Restaurant	Coffee Shop	Chinese Restaurant	Sandwich Place	Clothing Store	Dim Sum Restaurant	Pub
10	Bhayandar	Western Suburbs	Ice Cream Shop	Soccer Field	Diner	Lake	Restaurant	Pizza Place	Food Truck	Train Station	Bakery	Indian Restaurant
12	Bandstand Promenade	Western Suburbs	Coffee Shop	Performing Arts Venue	Scenic Lookout	Café	Tea Room	Deli / Bodega	Indian Restaurant	Food Truck	Lounge	Fast Food Restaurant

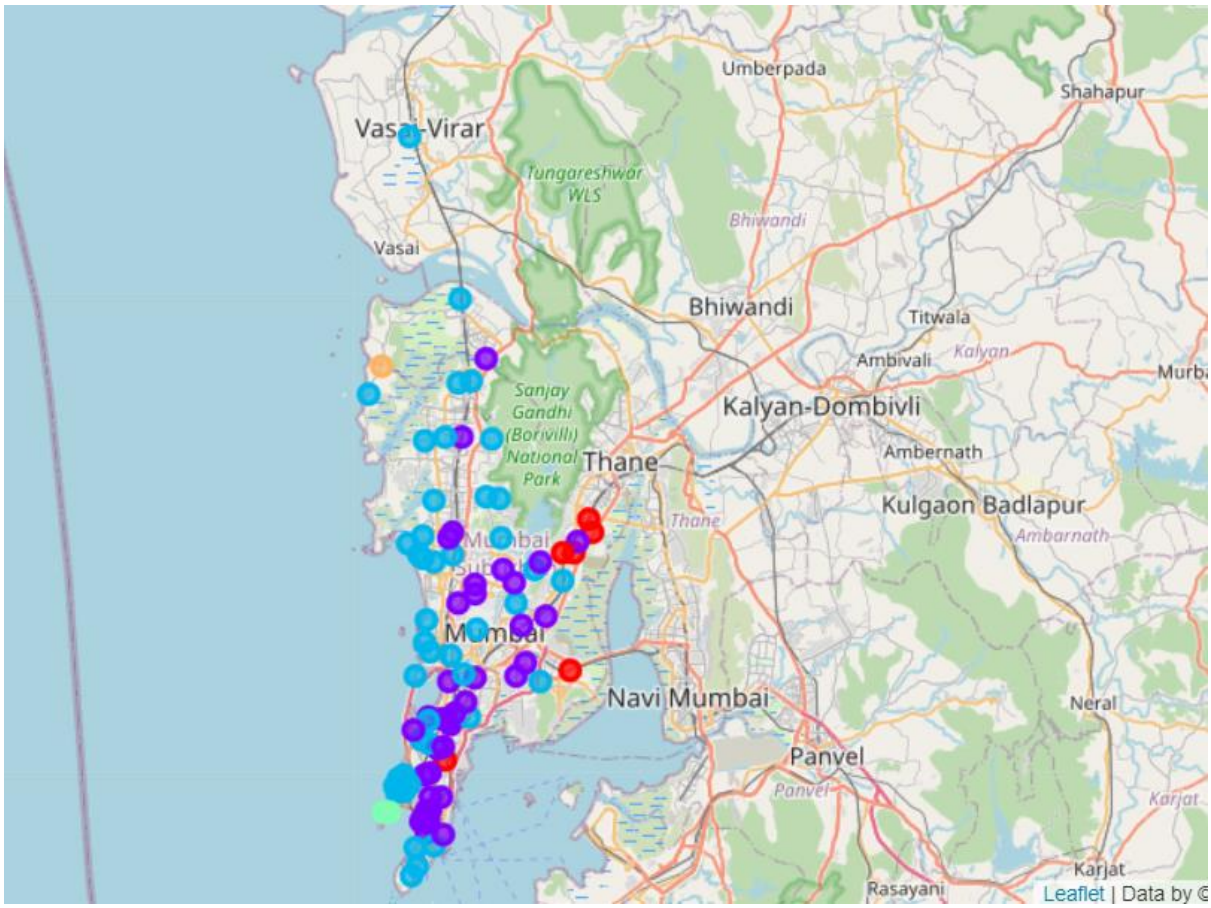
Cluster 4

	Neighborhood	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
70	Malabar Hill	South Mumbai	Coffee Shop	Park	Indian Restaurant	Lighthouse	Dessert Shop	Zoo	Dhaba	Field	Fast Food Restaurant	Farmers Market
77	Walkeshwar	South Mumbai	Park	Indian Restaurant	Food & Drink Shop	Restaurant	Lighthouse	Food Truck	Dessert Shop	Coffee Shop	Event Space	Dhaba

Cluster 5

	Neighborhood	Location	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
11	Uttan	Western Suburbs	Skate Park	Whisky Bar	Convenience Store	Indian Restaurant	Beach	Restaurant	Zoo	Dhaba	Fast Food Restaurant	Farmers Market

Based on the clusters shown above, the neighborhoods can once again be plotted on a map of Mumbai, however, this time with different color markers to distinguish between different clusters. This is shown in Figure



Discussion:

By analyzing the five clusters obtained we can see that some of the clusters are more suited for restaurants and hotels, whereas, other clusters are less suited. Neighborhoods in clusters 1, 4, and 5 contain a small percentage of restaurants, hotels, cafe and pubs in their top 10 common venues. These clusters contain a higher degree of other venues like train station, bus station, fish market, gym, performing arts venue and smoke shop, to name a few. Thus, they are well suited for opening a new restaurant as there are less number of competitors. On the other hand, neighborhoods in clusters 2 and 3 contain a much higher degree of restaurants, hotels, multiplex, cafes, bars and other food joints. Thus, the neighborhoods in these clusters would not be well suited for opening a new restaurant.

Comparing clusters 2 and 3, neighborhoods in cluster 2 seem to be more suited for starting a restaurant since they contains a smaller percentage of food joints in the top 10 most common venues than cluster 3. The neighborhoods in cluster 3 contain a variety of food joints like restaurants, tea rooms, bakery, cafe, steakhouse and pubs and also contain very diverse cuisines like Japanese, Indian, Chinese, Italian and seafood restaurants. Most neighborhoods in cluster 2 seem to have Indian Restaurant as their top most common venue; however, on careful analysis we can see that neighborhoods in cluster 2 also contain other venues like soccer field, flea market, smoke shop, gym, train station, dance studio, music store, cosmetics shop and so on. Thus, it is recommended that the new restaurant can be opened in the neighborhoods belonging to cluster 2.

CONCLUSION

In this project, the neighborhoods in Mumbai, India have been successfully analyzed for determining which would be the best neighborhoods for opening a new restaurant. Based on the analysis carried out, neighborhoods in cluster 1 are recommended as locations for the new restaurant. This has also been plotted in the map in Figure 18. The stakeholders and investors can further tune this by considering various other factors like transport, legal requirements, and costs associated. These were out of the scope for this project and thus were not considered.

FINAL COMMENTS

Note: Please find the code on github at https://github.com/anshu03654/Coursera_Capastone/blob/master/Code_Notebook.ipynb