IBM Data Science Professional Certificate

Capstone Project: The Battle of Neighbourhoods

Analyzing the Neighborhoods in Mumbai to Start a new Restaurant

Author: Eric Kwasi Gadosey

Table of Contents

Introduction

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 with people from various ethnicities residing here. The multi-cultural nature of the city of Mumbai has brought alongwith it numerous cuisines from all over the world. The people of India generallylove food and experience different flavors. Thus, the aim of this project is to study the neighborhoods in Mumbai to determine possible locations for starting 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 appropriatedata and find recommendations for the stakeholders.

Data Collection

The following data is required for the project:

1) Neighborhood data of Mumbai
2) Geographical coordinates of Mumbai and all neighborhoods in Mumbai
3) Venue data for neighborhoods in Mumbai

Neighborhoods Data

The data of the neighborhoods in Mumbai was scraped from 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 a comprehensive and detailed table of the data which can easily be scraped using the read_html() method of pandas. The top 10 rows of the dataframe are shown in Figure 1.

	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

Figure 1: Top 10 rows of Mumbai neighborhoods data scraped from Wikipedia.

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 shown in Figure 2.

```
address = 'Mumbai, IN'
geolocator = Nominatim()
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.
```

Figure 2: Obtaining geographical coordinates of Mumbai.

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. Once again only the top 10 rows are shown.

	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	19.1263	72.8243	19.1263	72.8243	0.001606	0.00288
4	Lokhandwala	Western Suburbs	19.1432	72.8249	19.1432	72.8249	0.012345	0.0044
5	Marol	Western Suburbs	19.1192	72.8827	19.1191	72.8828	0.000169	6.7e-05
6	Sahar	Western Suburbs	19.1027	72.8626	19.1027	72.8626	0.00376476	0.00464166
7	Seven Bungalows	Western Suburbs	19.1315	72.817	19.1315	72.8165	0.00240802	0.000558001
8	Versova	Western Suburbs	19.1377	72.8135	19.1377	72.8135	0.01769	0.00652
9	Mira Road	Western Suburbs	19.2657	72.8711	19.2657	72.8707	0.0184624	0.000418149

Figure 3: Absolute difference between latitude and longitude values obtained from Wikipedia and Geocoder.

Figure 4 shows the top 10 rows of the final Mumbai neighborhoods dataframe after replacing the latitude and longitude values as mentioned before and dropping unnecessary columns.

	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	19.1263	72.8243
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	19.1315	72.817
8	Versova	Western Suburbs	19.1377	72.8135
9	Mira Road	Western Suburbs	19.2657	72.8711

Figure 4: Final Mumbai neighborhoods dataframe.

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. Figure 5 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	Shawarma Factory	19.124591	72.840398	Falafel Restaurant
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	Domino's Pizza	19.131000	72.848000	Pizza Place
7	Amboli	19.1293	72.84644	Garden Court	19.127188	72.837478	Indian Restaurant
8	Amboli	19.1293	72.84644	Subway	19.127860	72.844461	Sandwich Place
9	Amboli	19.1293	72.84644	Sarvodaya Veg. Restaurant	19.123760	72.850893	Indian Restaurant

Figure 5: Data obtained from Foursquare API after cleaning.

Methodology

This section provides details for the methodology used in the project.

Data Visualization

In order to understand the data obtained for Mumbai neighborhoods, basic visualization was carried out. Figure 6 shows a bar plot depicting the number of neighborhoods in each location in Mumbai.

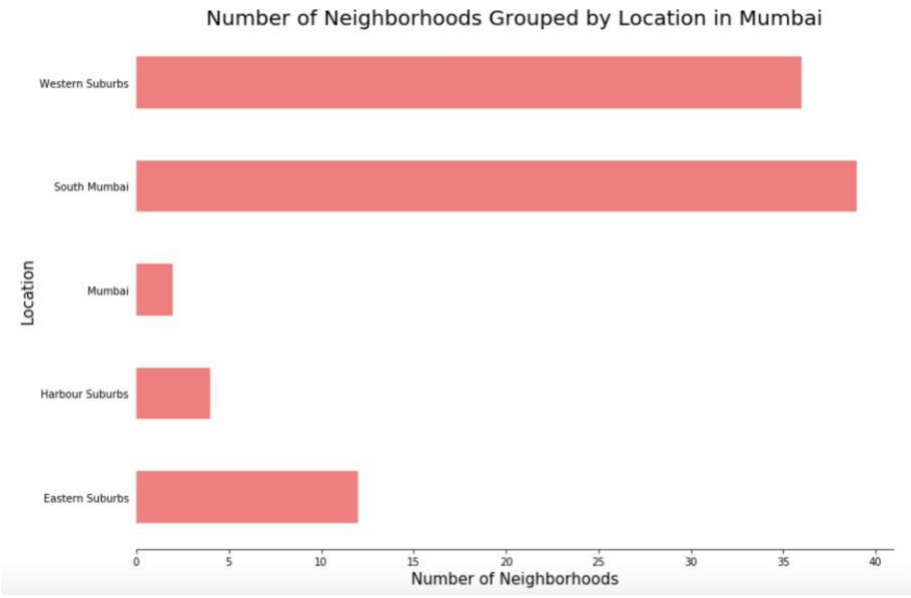

Figure 6: Number of neighborhoods grouped by location.

It is evident from Figure 6 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. This is shown in Figure 7.

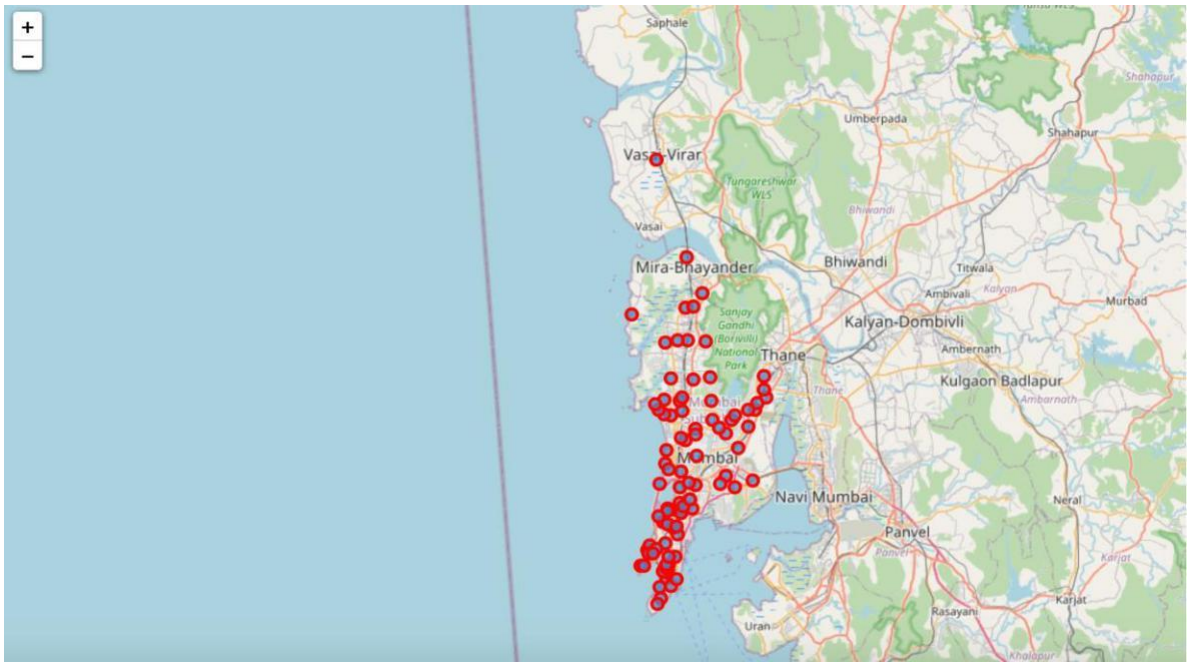

Figure 7: Depicting the neighborhood spread across Mumbai.

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 Figure 8 which shows only the top 10 rows.

	Neighborhood	ATM	Accessories Store	Airport Terminal	American Restaurant	Antique Shop	Aquarium	Arcade	Art Gallery	Arts & Crafts Store	...	Trail	Train	Train Station	Vegetarian / Vegan Restaurant	Whisky Bar	Wine Bar	Wine Shop	Women's Store	Yoga Studio	Zoo
0	Amboli	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000	0.000000	0.0	0.0
1	Chakala, Andheri	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.047619	0.0	0.0	0.000	0.000000	0.0	0.0
2	D.N. Nagar	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.043478	0.0	0.0	0.000	0.021739	0.0	0.0
3	Four Bungalows	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.030303	0.0	0.0	0.000	0.015152	0.0	0.0
4	Lokhandwala	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.010753	0.0	0.0	0.000	0.010753	0.0	0.0
5	Marol	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000	0.000000	0.0	0.0
6	Sahar	0.0	0.0	0.033333	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000	0.000000	0.0	0.0
7	Seven Bungalows	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.014925	...	0.0	0.0	0.0	0.029851	0.0	0.0	0.000	0.000000	0.0	0.0
8	Versova	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.025000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.025	0.000000	0.0	0.0
9	Mira Road	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.000	0.066667	0.0	0.0

10 rows × 221 columns

Figure 8: One-hot Encoding resulting dataframe.

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 9.

	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	Coffee Shop	Bakery	Bar	Asian Restaurant	Pizza Place	Sandwich Place	Bowling Alley	Bus Station	Bike Rental / Bike Share
1	Chakala, Andheri	Hotel	Indian Restaurant	Café	Fast Food Restaurant	Pizza Place	Asian Restaurant	Hotel Bar	Vegetarian / Vegan Restaurant	Restaurant	Gym
2	D.N. Nagar	Bar	Indian Restaurant	Pub	Gym / Fitness Center	Pizza Place	Lounge	Coffee Shop	Vegetarian / Vegan Restaurant	Snack Place	Gym
3	Four Bungalows	Pub	Café	Indian Restaurant	Gym / Fitness Center	Chinese Restaurant	Bar	Seafood Restaurant	Lounge	Vegetarian / Vegan Restaurant	Coffee Shop
4	Lokhandwala	Indian Restaurant	Chinese Restaurant	Café	Pub	Bakery	Bar	Italian Restaurant	Gym / Fitness Center	Coffee Shop	Asian Restaurant
5	Marol	Indian Restaurant	Hotel	Diner	Bakery	Dance Studio	Ice Cream Shop	Chinese Restaurant	Fast Food Restaurant	Restaurant	Lounge
6	Sahar	Hotel	Café	Indian Restaurant	Lounge	Gym	Asian Restaurant	Pizza Place	Seafood Restaurant	Restaurant	Falafel Restaurant
7	Seven Bungalows	Café	Pub	Seafood Restaurant	Chinese Restaurant	Pizza Place	Coffee Shop	Bar	Ice Cream Shop	Asian Restaurant	Bistro
8	Versova	Café	Ice Cream Shop	Beach	Pizza Place	Coffee Shop	Chinese Restaurant	Salon / Barbershop	Frozen Yogurt Shop	Bistro	Sandwich Place
9	Mira Road	Indian Restaurant	Convenience Store	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	Food Truck	Motorcycle Shop	Movie Theater	Basketball Court	Bar

Figure 9: Top 10 most common venues for neighborhoods.

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 10.

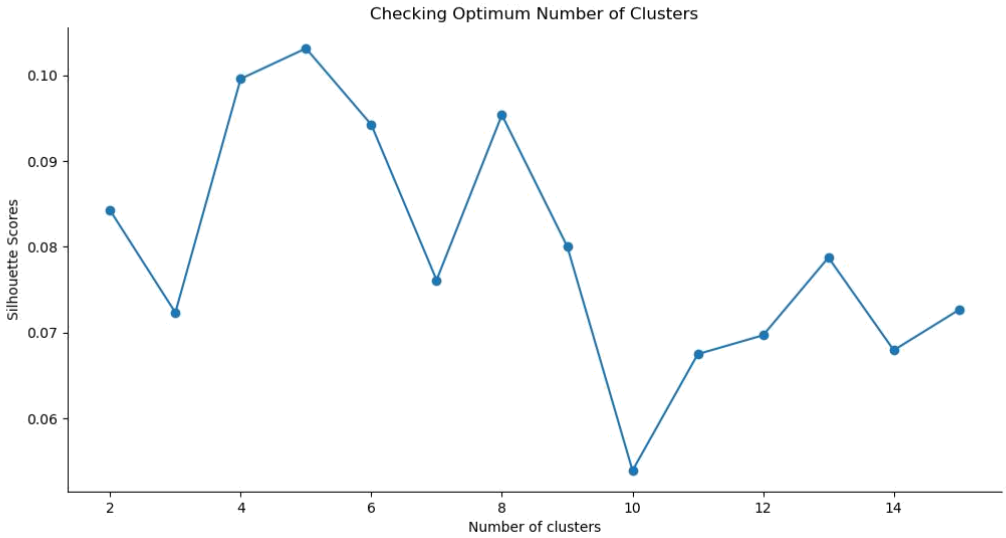

Figure 10: Silhouette scores for different number of clusters.

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 as seen in Figure 10.

Results

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 9 along with the

Location, Latitude, and Longitude columns to provide a complete summary of the clustering. The top 10 rows are shown in Figure 11.

	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 Most Common Venue	10th Most Common Venue
0	Amboli	Western Suburbs	19.1293	72.8464	1	Indian Restaurant	Coffee Shop	Bakery	Bar	Asian Restaurant	Pizza Place	Sandwich Place	Bowling Alley	Bus Station	Bike Rental / Bike Share
1	Chakala, Andheri	Western Suburbs	19.1084	72.8623	1	Hotel	Indian Restaurant	Café	Fast Food Restaurant	Pizza Place	Asian Restaurant	Hotel Bar	Vegetarian / Vegan Restaurant	Restaurant	Gym
2	D.N. Nagar	Western Suburbs	19.1241	72.8325	0	Bar	Indian Restaurant	Pub	Gym / Fitness Center	Pizza Place	Lounge	Coffee Shop	Vegetarian / Vegan Restaurant	Snack Place	Gym
3	Four Bungalows	Western Suburbs	19.1263	72.8243	0	Pub	Café	Indian Restaurant	Gym / Fitness Center	Chinese Restaurant	Bar	Seafood Restaurant	Lounge	Vegetarian / Vegan Restaurant	Coffee Shop
4	Lokhandwala	Western Suburbs	19.1432	72.8249	0	Indian Restaurant	Chinese Restaurant	Café	Pub	Bakery	Bar	Italian Restaurant	Gym / Fitness Center	Coffee Shop	Asian Restaurant
5	Marol	Western Suburbs	19.1192	72.8827	1	Indian Restaurant	Hotel	Diner	Bakery	Dance Studio	Ice Cream Shop	Chinese Restaurant	Fast Food Restaurant	Restaurant	Lounge
6	Sahar	Western Suburbs	19.1027	72.8626	0	Hotel	Café	Indian Restaurant	Lounge	Gym	Asian Restaurant	Pizza Place	Seafood Restaurant	Restaurant	Falafel Restaurant
7	Seven Bungalows	Western Suburbs	19.1315	72.817	0	Café	Pub	Seafood Restaurant	Chinese Restaurant	Pizza Place	Coffee Shop	Bar	Ice Cream Shop	Asian Restaurant	Bistro
8	Versova	Western Suburbs	19.1377	72.8135	0	Café	Ice Cream Shop	Beach	Pizza Place	Coffee Shop	Chinese Restaurant	Salon / Barbershop	Frozen Yogurt Shop	Bistro	Sandwich Place
9	Mira Road	Western Suburbs	19.2657	72.8711	1	Indian Restaurant	Convenience Store	Coffee Shop	Mexican Restaurant	Fast Food Restaurant	Food Truck	Motorcycle Shop	Movie Theater	Basketball Court	Bar

Figure 11: Clustering neighborhoods in Mumbai.

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.

	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	Hotel Bar	Asian Restaurant	Pizza Place	Vegetarian / Vegan Restaurant	Restaurant	Burger Joint	Multiplex
6	Sahar	Western Suburbs	Hotel	Indian Restaurant	Restaurant	Gym	Asian Restaurant	Bar	Coffee Shop	Café	Italian Restaurant	Pub
27	Khar Danda	Western Suburbs	Hotel	Clothing Store	Park	Coffee Shop	Dessert Shop	Bookstore	Bistro	French Restaurant	Boutique	Pool
40	Kanjurmarg	Eastern Suburbs	Train Station	Gym	Hotel	Gift Shop	Chinese Restaurant	French Restaurant	Asian Restaurant	Multiplex	Donut Shop	Electronics Store
70	Malabar Hill	South Mumbai	Gym	Hotel	Park	Convenience Store	Lighthouse	Coffee Shop	Dessert Shop	Indian Restaurant	Cupcake Shop	Cosmetics Shop
77	Walkeshwar	South Mumbai	Gym	Park	Hotel	Convenience Store	Food & Drink Shop	Food Truck	Lighthouse	Restaurant	Dessert Shop	Coffee Shop

Figure 12: 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
0	Amboli	Western Suburbs	Indian Restaurant	Pizza Place	Sandwich Place	Bar	Bakery	Athletics & Sports	Metro Station	Burger Joint	Bike Rental / Bike Share	Mughlai Restaurant
2	D.N. Nagar	Western Suburbs	Pub	Bar	Indian Restaurant	Pizza Place	Gym / Fitness Center	Women's Store	Vegetarian / Vegan Restaurant	Juice Bar	Snack Place	Hotel Bar
3	Four Bungalows	Western Suburbs	Pub	Lounge	Chinese Restaurant	Indian Restaurant	Seafood Restaurant	Vegetarian / Vegan Restaurant	Pizza Place	Café	Coffee Shop	Ice Cream Shop
4	Lokhandwala	Western Suburbs	Indian Restaurant	Café	Coffee Shop	Chinese Restaurant	Bar	Pub	Italian Restaurant	Fast Food Restaurant	Asian Restaurant	Seafood Restaurant
7	Seven Bungalows	Western Suburbs	Café	Pub	Bar	Chinese Restaurant	Indian Restaurant	Seafood Restaurant	Pizza Place	Ice Cream Shop	Vegetarian / Vegan Restaurant	Coffee Shop
...	...	...	...	...	...	...	...	...	...	...	...	...
83	Ballard Estate	South Mumbai	Indian Restaurant	Café	Coffee Shop	Seafood Restaurant	Bar	Parsi Restaurant	Lounge	Train Station	Irani Cafe	Dessert Shop
88	Parel	South Mumbai	Coffee Shop	Playground	Indian Restaurant	Chinese Restaurant	Plaza	Restaurant	Rest Area	Maharashtrian Restaurant	Bar	Vegetarian / Vegan Restaurant
89	Gowalia Tank	South Mumbai	Indian Restaurant	Fast Food Restaurant	Coffee Shop	Bakery	Electronics Store	Café	Sandwich Place	Vegetarian / Vegan Restaurant	Snack Place	Ice Cream Shop
90	Dava Bazaar	South Mumbai	Train Station	Indian Restaurant	Fish Market	Hotel	French Restaurant	Fast Food Restaurant	Café	Coffee Shop	Clothing Store	Asian Restaurant
91	Dharavi	Mumbai	Indian Restaurant	Paper / Office Supplies Store	Café	Fast Food Restaurant	Shoe Store	Seafood Restaurant	Sandwich Place	Luggage Store	Diner	Bus Station

Figure 13: 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
92	Thane	Mumbai	Platform	Pizza Place	Performing Arts Venue	Dessert Shop	Fish & Chips Shop	Field	Fast Food Restaurant	Falafel Restaurant	Event Space	Electronics Store

Figure 14: 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
5	Marol	Western Suburbs	Indian Restaurant	Chinese Restaurant	Hotel	Coffee Shop	Bakery	Restaurant	Lounge	Department Store	Dance Studio	Diner
9	Mira Road	Western Suburbs	Indian Restaurant	Mexican Restaurant	Bar	Coffee Shop	Fast Food Restaurant	Movie Theater	Basketball Court	Convenience Store	Gym / Fitness Center	Dhaba
11	Uttan	Western Suburbs	Convenience Store	Indian Restaurant	Whisky Bar	Restaurant	Zoo	Dhaba	Fish Market	Fish & Chips Shop	Field	Fast Food Restaurant
20	Jogeshwari West	Western Suburbs	Indian Restaurant	Restaurant	Gym	Café	Business Service	Bus Station	Men's Store	Mughlai Restaurant	Ice Cream Shop	Chinese Restaurant
36	Bhandup	Eastern Suburbs	Train Station	Indian Restaurant	Fast Food Restaurant	Asian Restaurant	Zoo	Dim Sum Restaurant	Flea Market	Fish Market	Fish & Chips Shop	Field
37	Amrut Nagar	Eastern Suburbs	Indian Restaurant	Ice Cream Shop	Pharmacy	Train Station	Café	Multiplex	Bus Station	Mughlai Restaurant	Restaurant	Plaza
39	Pant Nagar	Eastern Suburbs	Indian Restaurant	Coffee Shop	Shopping Mall	Café	Vegetarian / Vegan Restaurant	Bakery	Platform	Fast Food Restaurant	Bus Station	Diner
43	Chandivali	Eastern Suburbs	Indian Restaurant	Fast Food Restaurant	Restaurant	Seafood Restaurant	Brewery	Café	Food	Lounge	Chinese Restaurant	Pub
46	Vidyavihar	Eastern Suburbs	Indian Restaurant	Fast Food Restaurant	Juice Bar	Playground	Chinese Restaurant	Dessert Shop	Ice Cream Shop	Vegetarian / Vegan Restaurant	Train Station	Bar
48	Chembur	Harbour Suburbs	Indian Restaurant	Pizza Place	Café	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Dessert Shop	Asian Restaurant	Bar	Seafood Restaurant	Garden
51	Mahul	Harbour Suburbs	Indian Restaurant	Fast Food Restaurant	Smoke Shop	Hotel	Pizza Place	Cosmetics Shop	Ice Cream Shop	Vegetarian / Vegan Restaurant	Pool	Bar
54	Bhuleshwar	South Mumbai	Indian Restaurant	Bakery	Chinese Restaurant	Market	Café	Ice Cream Shop	Dessert Shop	Multiplex	Fast Food Restaurant	Rest Area
57	Cavel	South Mumbai	Indian Restaurant	Bakery	Café	Chinese Restaurant	Market	Restaurant	Ice Cream Shop	Fast Food Restaurant	Multiplex	Arcade
63	Dhobitalao	South Mumbai	Indian Restaurant	Coffee Shop	Café	Bar	Multiplex	Flea Market	Chinese Restaurant	Train Station	Bakery	Market
64	Dongri	South Mumbai	Indian Restaurant	Dessert Shop	Chinese Restaurant	Convenience Store	Garden	Restaurant	Market	Café	BBQ Joint	Boat or Ferry
68	Mahalaxmi	South Mumbai	Indian Restaurant	Café	Chinese Restaurant	Multiplex	Bakery	Fast Food Restaurant	Ice Cream Shop	Bar	Snack Place	Jewelry Store
69	Mahim	South Mumbai	Indian Restaurant	Fast Food Restaurant	Bakery	Chinese Restaurant	Dessert Shop	Café	Clothing Store	Breakfast Spot	Coffee Shop	Asian Restaurant
71	Marine Drive	South Mumbai	Indian Restaurant	Bakery	Multiplex	Ice Cream Shop	Chinese Restaurant	Fast Food Restaurant	Café	Train Station	Cheese Shop	Market
78	Worli	South Mumbai	Indian Restaurant	Scenic Lookout	Playground	Café	Garden	Diner	Thai Restaurant	Sandwich Place	Food	Gourmet Shop
84	Chira Bazaar	South Mumbai	Indian Restaurant	Café	Bakery	Multiplex	Ice Cream Shop	Chinese Restaurant	Fast Food Restaurant	Gym	Juice Bar	Flea Market
85	Fanas Wadi	South Mumbai	Indian Restaurant	Café	Fast Food Restaurant	Chinese Restaurant	Market	Ice Cream Shop	Juice Bar	Dessert Shop	Train Station	Coffee Shop
86	Chor Bazaar	South Mumbai	Indian Restaurant	Dessert Shop	Middle Eastern Restaurant	Arts & Crafts Store	Rest Area	Café	BBQ Joint	Market	Chinese Restaurant	Fast Food Restaurant
87	Matunga	South Mumbai	Indian Restaurant	Fast Food Restaurant	Café	Snack Place	Ice Cream Shop	Gym	Miscellaneous Shop	Chinese Restaurant	Vegetarian / Vegan Restaurant	Train Station

Figure 15: 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
59	Cotton Green	South Mumbai	Plaza	Pizza Place	Fast Food Restaurant	Train Station	Bakery	Dhaba	Fish Market	Fish & Chips Shop	Field	Falafel Restaurant

Figure 16: Cluster 5.

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.

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 3, 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 not well suited for opening a new restaurant. On the other hand, neighborhoods in clusters 1 and 2 contain a much higher degree of restaurants, hotels, multiplex, cafes, bars and other food joints. Thus, the neighborhoods in these clusters would be well suited for opening a new restaurant.

Comparing clusters 1 and 2, neighborhoods in cluster 1 seem to be more suited for starting a restaurant since they contains a larger percentage of food joints in the top 10 most common venues than cluster 2. The neighborhoods in cluster 1 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 1. This neighborhood can be further plotted on a map as shown below in Figure 18.

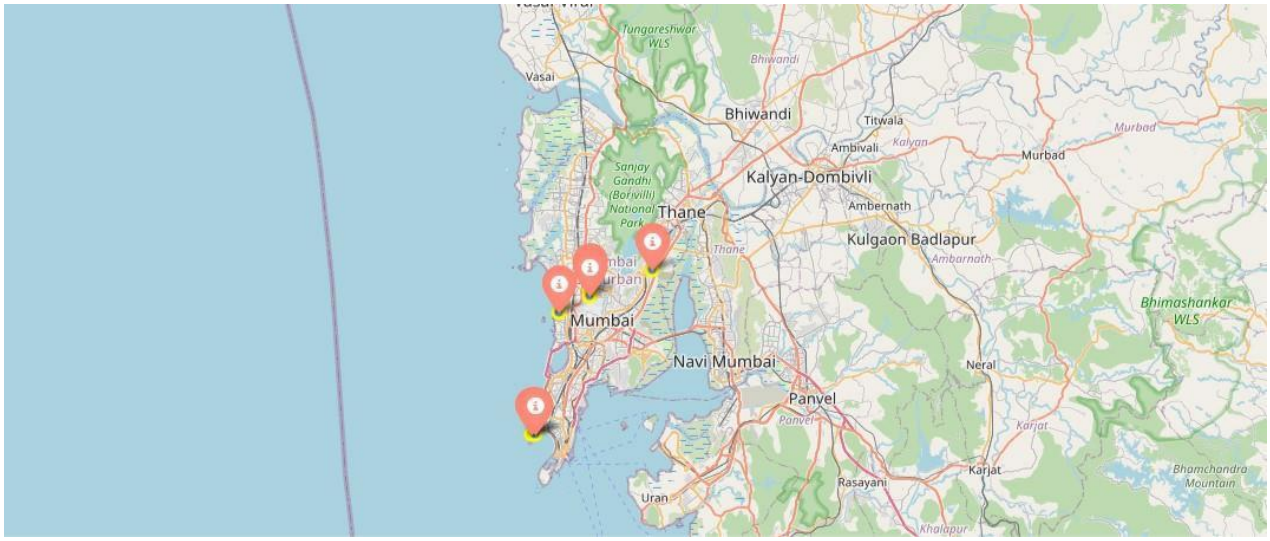

Figure 17: Neighborhoods most suited for starting a new restaurant.

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. 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.

References

1. *"Mumbai Suburban" (PDF). National Informatics Centre (Maharashtra State Centre). Archived from the original (PDF) on 13 May 2012. Retrieved 17 July 2013.*
2. *"Mumbai Urban Infrastructure Project". Mumbai Metropolitan Region Development Authority (MMRDA). Archived from the original on 26 February 2009. Retrieved 17 July 2013.*
3. *"Area and Density – Metropolitan Cities" (PDF). Ministry of Urban Development (Government of India). p. 33. Archived from the original (PDF) on 29 April 2009. Retrieved 17 July 2013.*
4. Mumbai Plan, 1.2 Area and Divisions
5. Greater Bombay District Gazetteer 1960, p. 2
6. Mumbai Plan, 1.1 Location
7. Bapat 2005, pp. 111–112