Applied Data Science Capstone Project

Analyzing the Neighborhoods in Mumbai for Starting a Restaurant

Table of contents

Introduction :

Problem background :

Mumbai is the capital city of Indian State of Maharashtra and also 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 societies residing here. The multi-cultural environment of the city of Mumbai has brought along with it numerous cuisines from all over the world. Mumbai can also be called a foodie's paradise because of its vast variety of foods. Thus, the aim of this project is to study the neighbourhoods 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 and open a restaurant in Mumbai. The main objective of this project is to analyze appropriate data and find recommendations for the stakeholders

Data :

To find a solution to the questions and build a recommender model, we need data and lots of data. Data can answer question which are unimaginable and non-answerable by humans because humans do not have the tendency to analyze such large dataset and produce analytics to find a solutions. 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 html method. The top 10 rows of the dataframe are shown as below,

	Area	Location	Latitude	Longitude
0	Amboli	Andheri,Western Suburbs	19.1293	72.8434
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	72.82721
4	Lokhandwala	Andheri,Western Suburbs	19.130815	72.82927
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	19.12	72.82
9	Mira Road	Mira-Bhayandar,Western Suburbs	19.284167	72.871111

To access location of a restaurant, it's Latitude and Longitude is to be known so that we can point at its coordinates and create a map displaying all the restaurants with its labels respectively. I have downloaded the commercial property rate data of the neighborhoods as below,

	Area	Location	Latitude	Longitude	Rate	Suburb
0	Amboli	Andheri	19.1293	72.8434	18000.0	Western Suburbs
1	Chakala, Andheri	Western Suburbs	19.111388	72.860833	14750.0	Western Suburbs
2	D.N. Nagar	Andheri	19.124085	72.831373	18000.0	Western Suburbs
3	Four Bungalows	Andheri	19.124714	72.82721	18000.0	Western Suburbs
4	Lokhandwala	Andheri	19.130815	72.82927	18000.0	Western Suburbs
5	Marol	Andheri	19.119219	72.882743	14750.0	Western Suburbs
6	Sahar	Andheri	19.098889	72.867222	14750.0	Western Suburbs
7	Seven Bungalows	Andheri	19.129052	72.817018	18000.0	Western Suburbs
8	Versova	Andheri	19.12	72.82	18000.0	Western Suburbs
9	Mira Road	Mira-Bhayandar	19.284167	72.871111	8500.0	Western Suburbs

Foursquare API:

Use of foursquare is focused to fetch nearest venue locations so that we can use them to form a cluster. Foursquare API leverages the power of finding nearest venues and also corresponding coordinates, venue location and names. After calling, the following data frame is created:

	Area	Location	Latitude	Longitude	Rate	Suburb	Food Joints	All Joints
0	Amboli	Andheri	19.1293	72.8434	18000.0	Western Suburbs	[Ice Cream Shop, Vegetarian / Vegan Restaurant...	[Ice Cream Shop, Residential Building (Apartme...
1	Chakala, Andheri	Western Suburbs	19.111388	72.860833	14750.0	Western Suburbs	[Asian Restaurant, Indian Restaurant, Asian Re...	[Asian Restaurant, Indian Restaurant, Asian Re...
2	D.N. Nagar	Andheri	19.124085	72.831373	18000.0	Western Suburbs	[Seafood Restaurant, Restaurant, Dessert Shop,...	[Seafood Restaurant, Post Office, High School,...
3	Four Bungalows	Andheri	19.124714	72.82721	18000.0	Western Suburbs	[Burger Joint, Snack Place, Cupcake Shop, Bake...	[Residential Building (Apartment / Condo), Res...
4	Lokhandwala	Andheri	19.130815	72.82927	18000.0	Western Suburbs	[Indian Restaurant, Food Truck, Indian Restaur...	[Light Rail Station, Supermarket, Accessories ...
5	Marol	Andheri	19.119219	72.882743	14750.0	Western Suburbs	[Snack Place, Ice Cream Shop, Ice Cream Shop, ...	[Hostel, Housing Development, School, Professi...
6	Sahar	Andheri	19.098889	72.867222	14750.0	Western Suburbs	[Indian Restaurant, Breakfast Spot, Italian Re...	[Church, Indian Restaurant, Airport, Airport, ...
7	Seven Bungalows	Andheri	19.129052	72.817018	18000.0	Western Suburbs	[South Indian Restaurant, Seafood Restaurant, ...	[Pub, South Indian Restaurant, Seafood Restaur...
8	Versova	Andheri	19.12	72.82	18000.0	Western Suburbs	[Fast Food Restaurant, Greek Restaurant, India...	[Beach, Beach, Lounge, Spa, Bank, Wedding Hall...
9	Mira Road	Mira-Bhayandar	19.284167	72.871111	8500.0	Western Suburbs	[Indian Restaurant, Chinese Restaurant, Americ...	[Residential Building (Apartment / Condo), Hou...

Methodology :

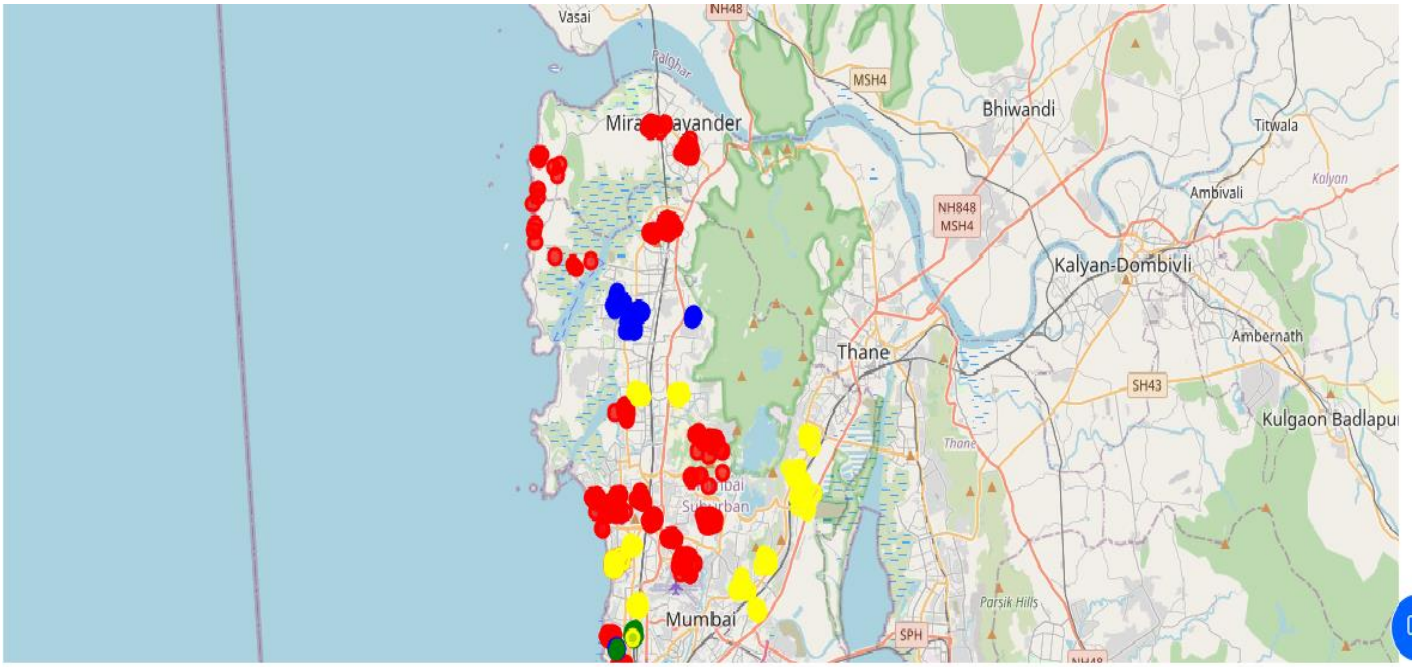
Exploratory analysis:

Exploring the dataset is important because it gives you initial insights and may help you to get partial idea of the answers that you are looking to find out from the data.

Scrapping the data from different sources and then combining it to form a single-ton dataset is a difficult task. To do so, we need to explore the current state of dataset and then list up all the features needed to be fetched. Here I am considering all venues near a neighborhood.

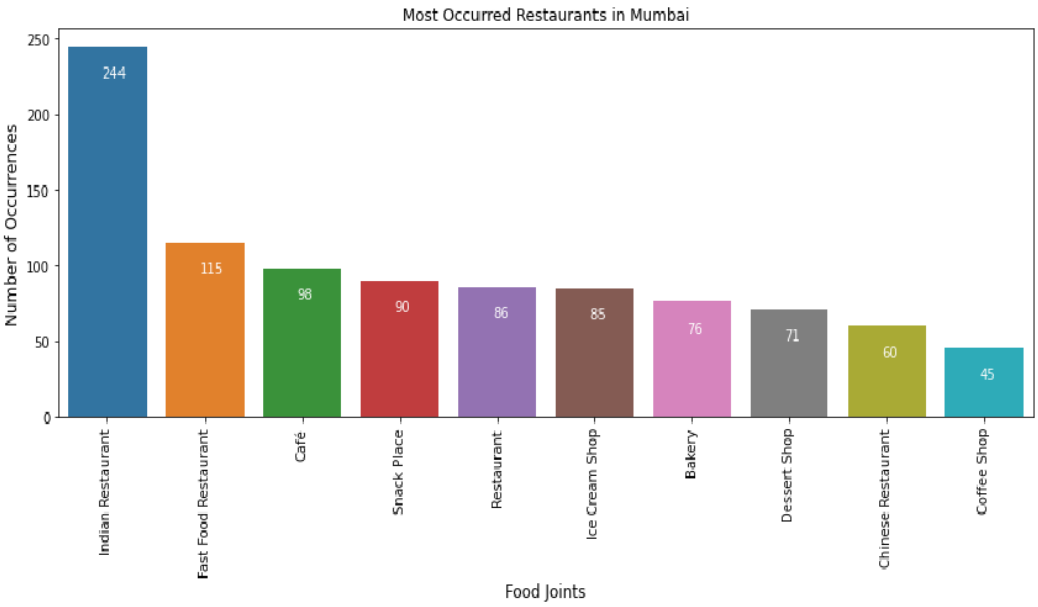
```
----Eastern Suburbs----
                                    venue  freq
0  Residential Building (Apartment / Condo)  0.10
1                          Indian Restaurant  0.05
2                                     Office  0.04
3                       Salon / Barbershop  0.03
4                                       Bank  0.02


----Harbour Suburbs----
                                    venue  freq
0  Residential Building (Apartment / Condo)  0.11
1                                       Bank  0.07
2                          Indian Restaurant  0.04
3                        Housing Development  0.03
4                       Fast Food Restaurant  0.02


----South Mumbai----
                                    venue  freq
0  Residential Building (Apartment / Condo)  0.09
1                                     Office  0.08
2                          Indian Restaurant  0.04
3                                       Bank  0.03
4                                   Building  0.03
```

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

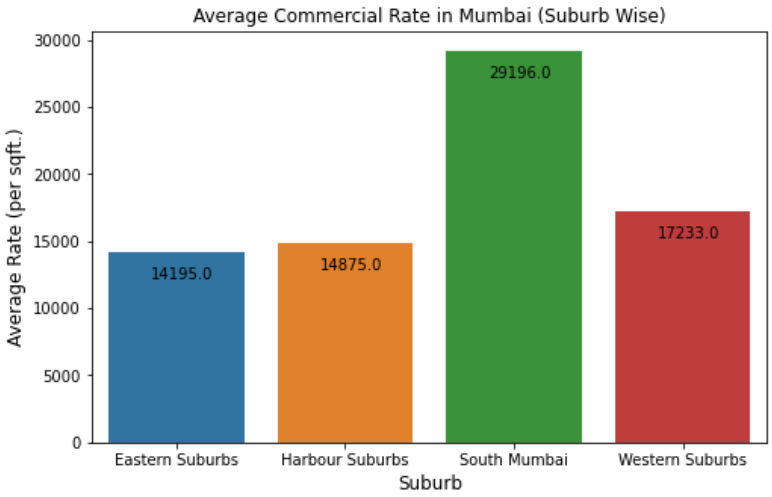


Data Visualization

In order to understand the data obtained for Mumbai neighborhoods, basic visualization was carried out.



Then I have calculated average commercial rate in Mumbai as below,



```
                 venue  freq
0     Indian Restaurant  0.18
1  Fast Food Restaurant  0.08
2            Restaurant  0.07
3       Ice Cream Shop  0.06
4               Bakery  0.06
5          Snack Place  0.05
6      Asian Restaurant  0.05
7                 Café  0.05
8         Dessert Shop  0.04
9    Chinese Restaurant  0.04


----Harbour Suburbs----
                   venue  freq
0       Indian Restaurant  0.25
1    Fast Food Restaurant  0.12
2             Coffee Shop  0.06
3                  Bakery  0.06
4             Snack Place  0.06
5        Asian Restaurant  0.06
6  Comfort Food Restaurant  0.03
7                   Diner  0.03
8     Dumpling Restaurant  0.03
9             Food Truck  0.03


----South Mumbai----
                 venue  freq
0     Indian Restaurant  0.17
1  Fast Food Restaurant  0.08


 ----South Mumbai----
                  venue  freq
 0     Indian Restaurant  0.17
 1  Fast Food Restaurant  0.08
 2                 Café  0.07
 3       Ice Cream Shop  0.05
 4               Bakery  0.05
 5          Snack Place  0.05
 6         Dessert Shop  0.05
 7           Restaurant  0.05
 8    Chinese Restaurant  0.04
 9           Food Truck  0.03


 ----Western Suburbs----
                  venue  freq
 0     Indian Restaurant  0.14
 1                 Café  0.07
 2  Fast Food Restaurant  0.07
 3       Ice Cream Shop  0.06
 4          Snack Place  0.06
 5           Restaurant  0.06
 6               Bakery  0.05
 7         Dessert Shop  0.05
 8          Pizza Place  0.04
 9    Chinese Restaurant  0.04
```

Result :

Feature extraction was carried out to obtain features from the Foursquare API data which was used for building the unsupervised learning model. 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 below,

	Area	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	Rate
0	Aarey Milk Colony	Indian Restaurant	Snack Place	Café	Restaurant	Tea Room	Chinese Restaurant	American Restaurant	Chaat Place	Pizza Place	Fast Food Restaurant	12950.0
1	Altamount Road	Ice Cream Shop	Café	Food Truck	Restaurant	Cupcake Shop	Dessert Shop	Bakery	Fast Food Restaurant	Indian Restaurant	Food	45000.0
2	Amboli	Chinese Restaurant	Snack Place	Fast Food Restaurant	Indian Restaurant	Restaurant	Ice Cream Shop	Middle Eastern Restaurant	Sandwich Place	Coffee Shop	Tea Room	18000.0
3	Amrut Nagar	Indian Restaurant	Fast Food Restaurant	Brewery	Ice Cream Shop	Asian Restaurant	Restaurant	Bagel Shop	Frozen Yogurt Shop	Dumpling Restaurant	Burger Joint	16150.0
4	Asalfa	Indian Restaurant	Bakery	Café	Food Truck	Dessert Shop	Donut Shop	Winery	Falafel Restaurant	Creperie	Cuban Restaurant	16150.0
5	Ballard Estate	Indian Restaurant	Breakfast Spot	Dessert Shop	Snack Place	Fast Food Restaurant	Vegetarian / Vegan Restaurant	Food Truck	Restaurant	Seafood Restaurant	BBQ Joint	29000.0
6	Bandstand Promenade	Coffee Shop	Café	Fast Food Restaurant	Chinese Restaurant	Restaurant	Indian Restaurant	Food Truck	Breakfast Spot	Bakery	Ice Cream Shop	30000.0
7	Bangur Nagar	Indian Restaurant	Food Truck	BBQ Joint	Bakery	Coffee Shop	Pizza Place	Juice Bar	Restaurant	Café	Mexican Restaurant	13500.0
8	Bhandup	Indian Restaurant	Café	Restaurant	Chinese Restaurant	Fast Food Restaurant	Diner	Afghan Restaurant	South Indian Restaurant	Fried Chicken Joint	Pizza Place	15000.0
9	Bhayandar	Fast Food Restaurant	Indian Restaurant	Ice Cream Shop	Bakery	Dessert Shop	Vegetarian / Vegan Restaurant	Pizza Place	Food Truck	Restaurant	Burger Joint	8000.0

Clustering Neighborhoods:

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.

Cluster 1:

	Location	Suburb	Food Joints	All Joints	Cluster Label	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	Rate_y
1	Western Suburbs	Western Suburbs	[Asian Restaurant, Indian Restaurant, Asian Re...	[Asian Restaurant, Indian Restaurant, Asian Re...	1	Indian Restaurant	Fast Food Restaurant	Pizza Place	Café	Restaurant	Coffee Shop	Asian Restaurant	Bakery	Snack Place	Korean Restaurant	14750.0
4	Andheri	Western Suburbs	[Indian Restaurant, Indian Restaurant, Food Tr...	[Supermarket, Light Rail Station, Accessories ...	1	Indian Restaurant	Café	Juice Bar	Food Truck	Coffee Shop	Winery	Dumpling Restaurant	Creperie	Cuban Restaurant	Cupcake Shop	18000.0
15	Borivali (West)	Western Suburbs	[Indian Restaurant, Seafood Restaurant, Pizza ...	[Beach, Resort, Whisky Bar, Indian Restaurant,...	1	Indian Restaurant	Café	Restaurant	Dessert Shop	Breakfast Spot	Coffee Shop	Fast Food Restaurant	Juice Bar	Chinese Restaurant	Food	10550.0
			[Sandwich													

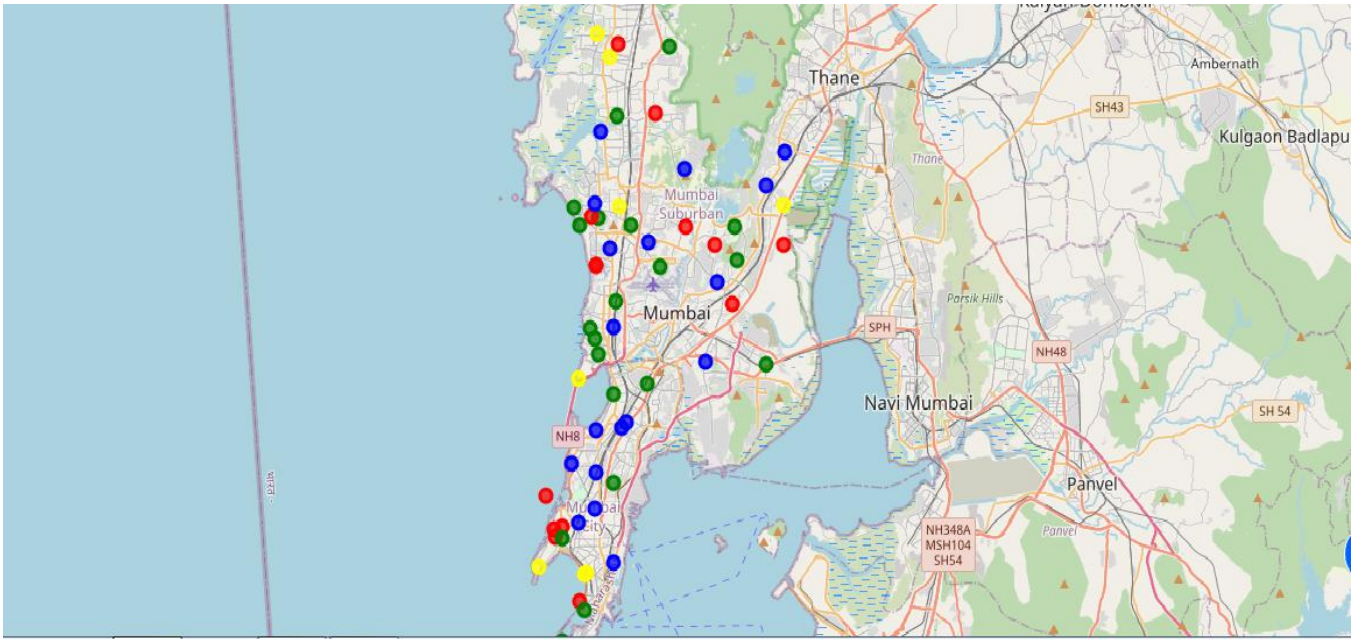
Cluster 2:

	Location	Suburb	Food Joints	All Joints	Cluster Label	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	Rate
2	Andheri	Western Suburbs	[Seafood Restaurant, Restaurant, Restaurant, I...	[Post Office, Seafood Restaurant, High School,...	2	Restaurant	Ice Cream Shop	Indian Restaurant	Café	Pizza Place	Dessert Shop	Breakfast Spot	Seafood Restaurant	Snack Place	Chinese Restaurant	1800(
6	Andheri	Western Suburbs	[Indian Restaurant, Pastry Shop, Breakfast Spo...	[Church, Indian Restaurant, Airport, Airport, ...	2	Restaurant	Café	Indian Restaurant	Breakfast Spot	Coffee Shop	Ice Cream Shop	Fast Food Restaurant	Cupcake Shop	Deli / Bodega	South Indian Restaurant	1475(
7	Andheri	Western Suburbs	[South Indian Restaurant, Seafood Restaurant,	[Pub, South Indian Restaurant, Seafood Restaur...	2	Ice Cream Shop	Indian Restaurant	Café	Chinese Restaurant	Pizza Place	Dessert Shop	Seafood Restaurant	Bistro	Coffee Shop	Mexican Restaurant	1800(

Cluster 3:

	Location	Suburb	Food Joints	All Joints	Cluster Label	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	Rate_
0	Andheri	Western Suburbs	[Ice Cream Shop, Vegetarian / Vegan Restaurant...	[Ice Cream Shop, Playground, Residential Build...	3	Chinese Restaurant	Snack Place	Fast Food Restaurant	Indian Restaurant	Restaurant	Ice Cream Shop	Middle Eastern Restaurant	Sandwich Place	Coffee Shop	Tea Room	18000.
9	Mira-Bhayandar	Western Suburbs	[Indian Restaurant, Chinese Restaurant, Americ...	[Residential Building (Apartment / Condo), Hou...	3	Chinese Restaurant	Indian Restaurant	Asian Restaurant	Fast Food Restaurant	Winery	Pizza Place	American Restaurant	Breakfast Spot	Coffee Shop	Cupcake Shop	8500.
11	Bandra	Western Suburbs	[Café, Italian Restaurant, Indian Restaurant, ...	[Hotel, Performing Arts Venue, Scenic Lookout,...	3	Coffee Shop	Café	Fast Food Restaurant	Chinese Restaurant	Restaurant	Indian Restaurant	Food Truck	Breakfast Spot	Bakery	Ice Cream Shop	30000.
21	Kandivali West	Western Suburbs	[Ice Cream Shop, Fish & Chips Shop, Food, Indi...	[Miscellaneous Shop, Government Building, Ice ...	3	Chinese Restaurant	Indian Restaurant	Pizza Place	Ice Cream Shop	Cupcake Shop	Italian Restaurant	Bakery	Multicuisine Indian Restaurant	Restaurant	Seafood Restaurant	15500.
22	Kandivali West	Western Suburbs	[Pizza Place, BBQ Joint, Ice Cream Shop, Fast ...	[Temple, College Academic Building, General Tr...	3	Fast Food Restaurant	Ice Cream Shop	Indian Restaurant	Coffee Shop	Chinese Restaurant	BBQ Joint	Snack Place	Food Truck	Juice Bar	Café	1550
			[Indian Restaurant,	[Multiplex, Indian												

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



Discussion:

Since there was a nonlinear relationship between income and population, it can be concluded that we must always perform inferential approach to find relationship among different set of features. Also during clustering, similar neighborhoods must be dumped into the right cluster.

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.