Exploring places in Mumbai, India using Foursquare and Zomato API

Nirmal Kirola

May 16, 2020

1. Introduction

1.1 Background

Whenever a person searches for a venue in a new city, they're highly interested in the best places that the city has to offer. The person might want to know how good a given restaurant is or the price range it falls under. This extra information would help decide which venue to choose amongst the many venues in the city. Combining the location of the venues in the city with their price and rating information would surely help visitors in a city make better informed decisions about the places they should visit.

Mumbai is composed of a number of sectors spread across a total area of 603 sq Km. There are many venues (especially restaurants, hotels and cafes) which can be explored. This project explores various venues in Mumbai and attributes the data based on user ratings and average price. To explore this information, this project involves the juxtaposition of both the Foursquare API and the Zomato API to fetch complete information of various venues (including name, address, category, rating, and price). Further, a map of the venues with specific color attributes will be plotted to highlight their position, and information about these venues. Such plots imbibe bountiful information in the form of their colored representations and location on the map. This enables any visitor to take a quick glance and decide what place to visit.

1.2 Interested audience

The target audience for such a project is twofold. Firstly, any person who is visiting Mumbai, India can use the plots and maps from this project to quickly select places that suit their budget and rating preferences. Secondly, a company can use this information to create a website or a mobile application, which is updated on a regular basis, to allow individuals to the city or even expand same functionality to other places.

2. Data

2.1 Data Sources

To get location and other information about various venues in Mumbai, I used two APIs and decided to combine the data from both of them together.

Using the Foursquare's explore API (which gives venues recommendations), I fetched venues up to a range of 4 kilometers from the center of Mumbai and collected their names, categories and locations (latitude and longitude).

Using the name, latitude and longitude values, I used the Zomato search API to fetch venues from its database. This API allows to find venues based on search criteria (usually the name), latitude and longitude values and more. Given that the data from the two APIs did not align completely, I had to use data cleaning to combine the two datasets properly.

From Foursquare API (https://developers.foursquare.com/api), I retrieved the following for each venue:
- **Name:** The name of the venue.
- **Category:** The category type as defined by the API.
- **Latitude:** The latitude value of the venue.
- **Longitude:** The longitude value of the venue.

From Zomato API (https://developers.zomato.com/api), I retrieved the following for each venue:

- **Name:** The name of the venue.
- **Address:** The complete address of the venue.
- **Rating:** The ratings as provided by many users.
- **Price range:** The price range the venue belongs to as defined by Zomato.
- **Price for two:** The average cost for two people dining at the place. I later convert the same to average price per person by dividing by 2.
- **Latitude:** The latitude value of the venue.
- **Longitude:** The longitude value of the venue.
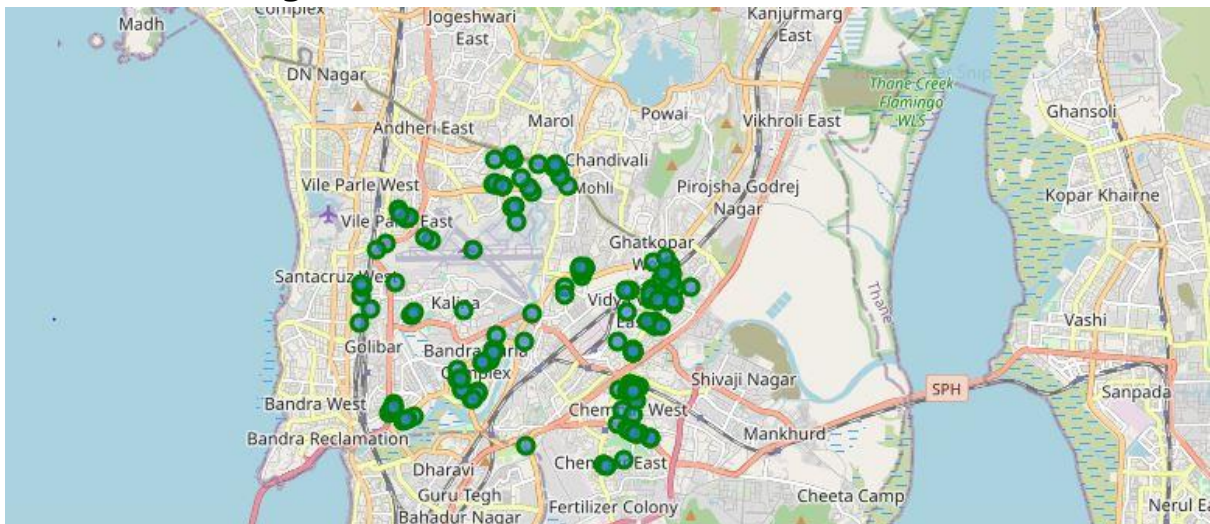
2.2 Data Cleaning

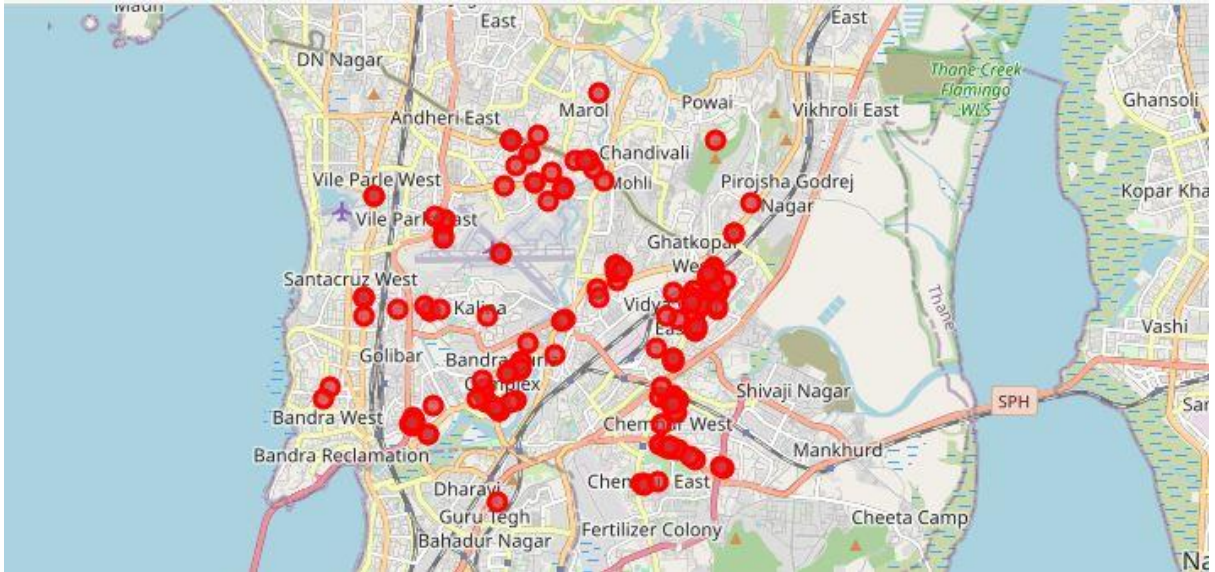*Figure 1: Venues retrieved from Foursquare API*

Figure 2: Venues retrieved from Zomato API

From figure 1 and figure 2, we can clearly see that some venues from the two APIs do not align with each other. Thus, I decided to combine them using their latitude and longitude values.

To combine the two datasets, I had to check that the latitude and longitude values of each corresponding venue match. After careful analysis, I decided to drop all corresponding venues from the two datasets that had their latitude and longitude values different by more than 0.0004 from one another. Thus, I rounded both the latitude and longitude values up to 4 decimal places. Then, I calculated the difference between the corresponding latitude and longitude values and saw if the difference was less than 0.0004 which should ideally mean that the two locations are the same. This removed many outliers from the two datasets. Once this was done, I observed that there were still some venues which were not correctly aligned.

They can be categorised as follows:
1. There are venues that have specific restaurants/cafes inside them as provided by Zomato API .
2. Two locations are so close that they have practically same latitude and longitude values .

3. Some venues have been replaced with new venues .

Venues belonging to category 1 and 3 are perfect to keep. However, the venues that belong to category 2 should be dropped. After careful inspection and removal, the final dataset had a total of 49 venues with which we can work.

As a final dataset, we're left with 75 venues with 8 columns as described in figure 3.

	categories	venue	latitude	longitude	price_range	rating	address	average_price
0	Hotel	Tuskers - Sofitel	19.0673	72.8692	4.0	4.0	Sofitel Hotel, C 57, Bandra Kurla Complex, Mumbai	1650.0
1	Hotel	O22 - Trident	19.0672	72.8675	4.0	4.0	Trident Hotel, C 56, G Block, Bandra Kurla Com...	2000.0
2	Indian Restaurant	Masala Library	19.0690	72.8696	4.0	4.4	Ground Floor, First International Financial Ce...	2500.0
3	Ice Cream Shop	Natural Ice Cream	19.0776	72.8628	1.0	4.1	3, Gokul Harmony, Kalina Market, Sunder Nagar,...	150.0
4	Deli / Bodega	Smoke House Deli	19.0688	72.8695	4.0	4.3	3A, Ground Floor, 1st International Financial ...	1250.0

Figure 3: Final data aggregated from both APIs

3. Methodology and Exploratory Data Analysis

As a first step, I retrieve the venues in Mumbai from Foursquare and Zomato APIs. I extract the location data from the Foursquare API for all venues up to a distance of 4 kilometers from the center of Mumbai. Using this, I fetch the venue information including price and rating data from Zomato API.

Using data cleaning, the dataset from the two APIs will be combined based on the venue names, latitude, and longitude values. One to one matching and careful data inspection would be used to remove any remaining outliers such as multiple venues at the same location from the two datasets. The final data will include the venue name, category, address, latitude, longitude, rating, price range, and average cost per person.

Using this dataset, I begin by analyzing the top venue types that exist in Mumbai. I will then explore the venues on maps. This will allow us to better understand the location of various venues and the places where many venues co-exist and create place worth visiting. I'll also explore the venues based on the ratings and price range of various venues. The venues will be plot using proper color coding such that a simple glance at the map would reveal the location of the venues as well as give information about them. I aim to identify places which can be recommended to visitors based on their price and rating preferences. I'll also cluster the venues and see if we can draw meaningful information out of what kind of venues exist in Mumbai.

As a final step, I will analyse these plots and try to draw conclusions on what places can be recommended to visitors. I'll discuss my findings and any inferences I can draw.

3.1 Categories

I begin my analysis by taking a look at the various categories of venues that exist in Mumbai. As there are many restaurants, I believe that the majority venues shall include restaurants.
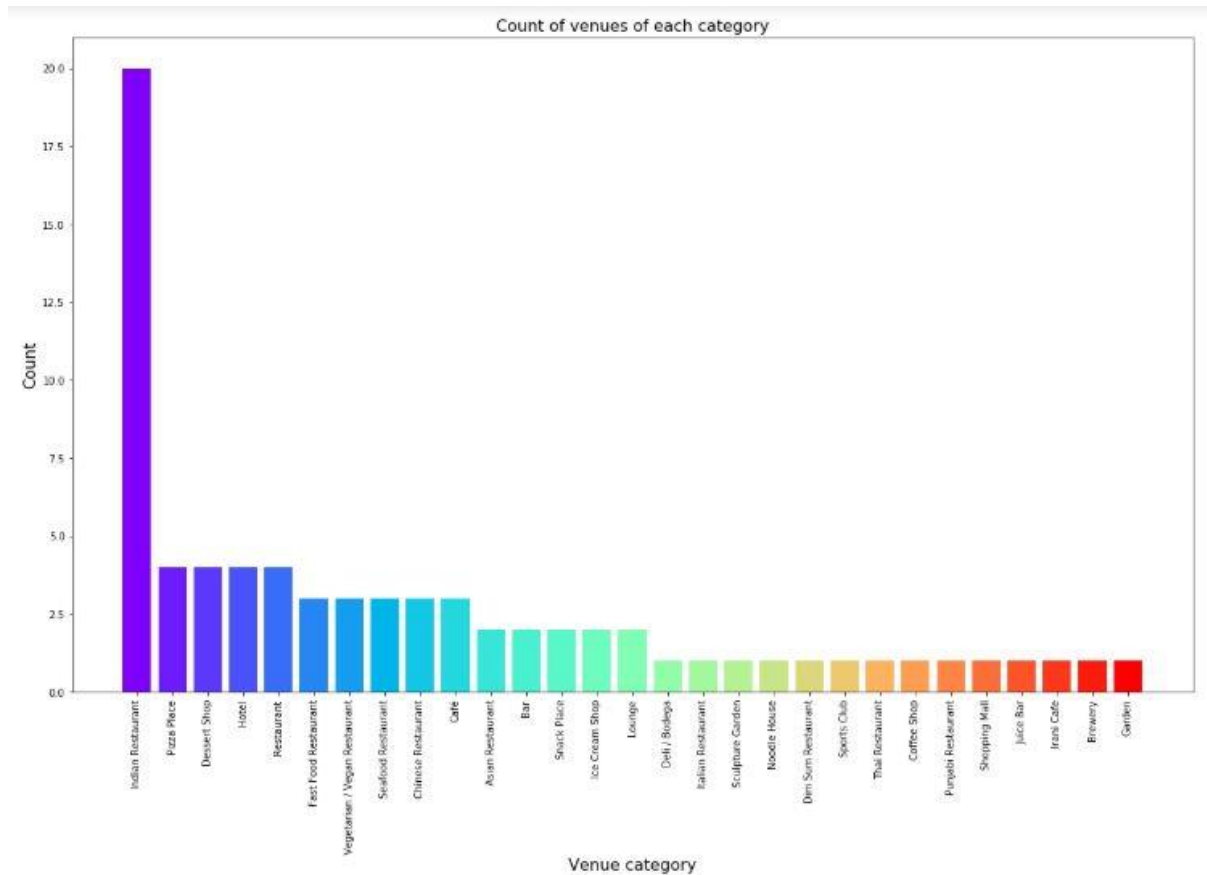
Figure 4: Count of various types of venues in Mumbai

From figure 4, we see that the majority venues are actually Indian Restaurants. This is closely followed by Indian Restaurants. For someone who is visiting Mumbai and loves Indian Restaurants, they'd surely love their stay.

3.2 Rating

Next, I'll explore the ratings of various venues in Mumbai. I decided to plot a bar chart with x-axis as the rating from 1 to 5 and the y-axis as the count of venues with that rating. I decided to plot the bar chart to see what average rating venues get in Mumbai. This can be seen in figure 5.

While the whole range of rating of venues might stretch from 1 to 5, the average rating is spread across 3.75 with maximum number of venues scoring between 3 and 5.

I followed this information by plotting the venues on the map of Mumbai. The venues that were rated below 3 were marked by red and orange while the venues that were rated more than or equal to 3 were plot as green and dark green. Taking a look at figure 6 reveals the same results as the bar plot. However, it is interesting to note that many high rated venues are located near Bandra Complex, and Marol Naka. Ghatkopar has venues with rating in the complete range from low to high. Also, the belt of venues in Chembur have high rated venues.
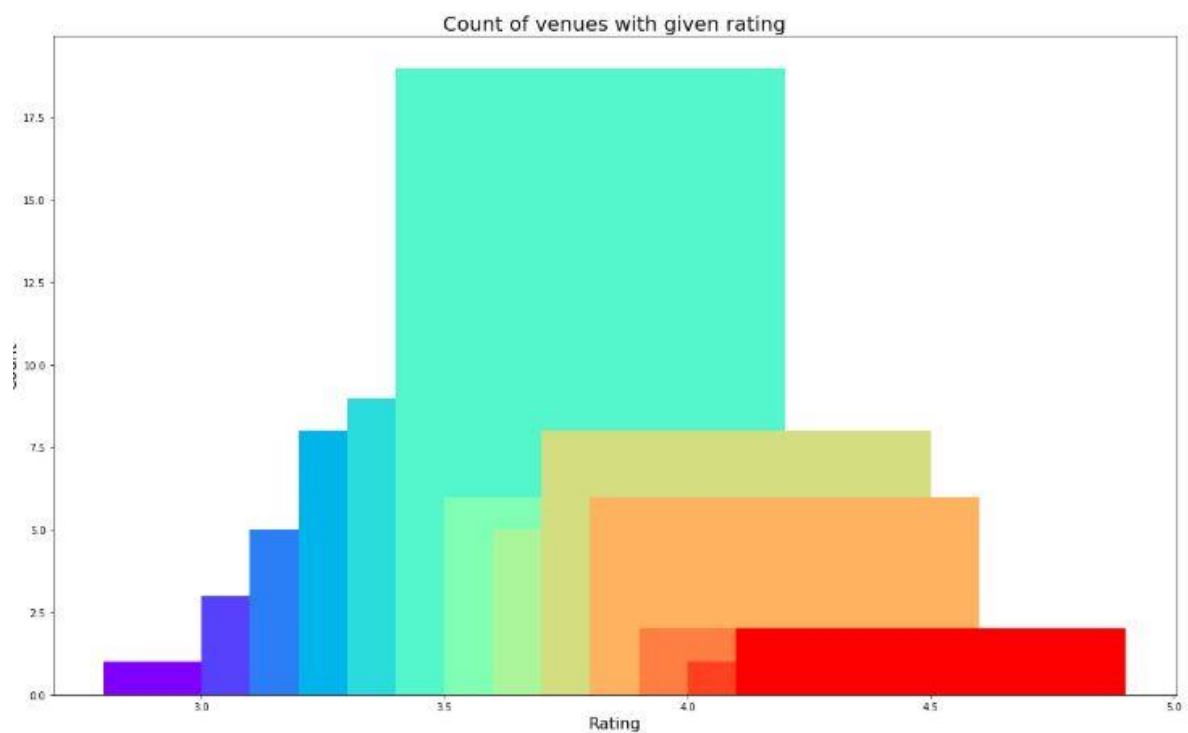

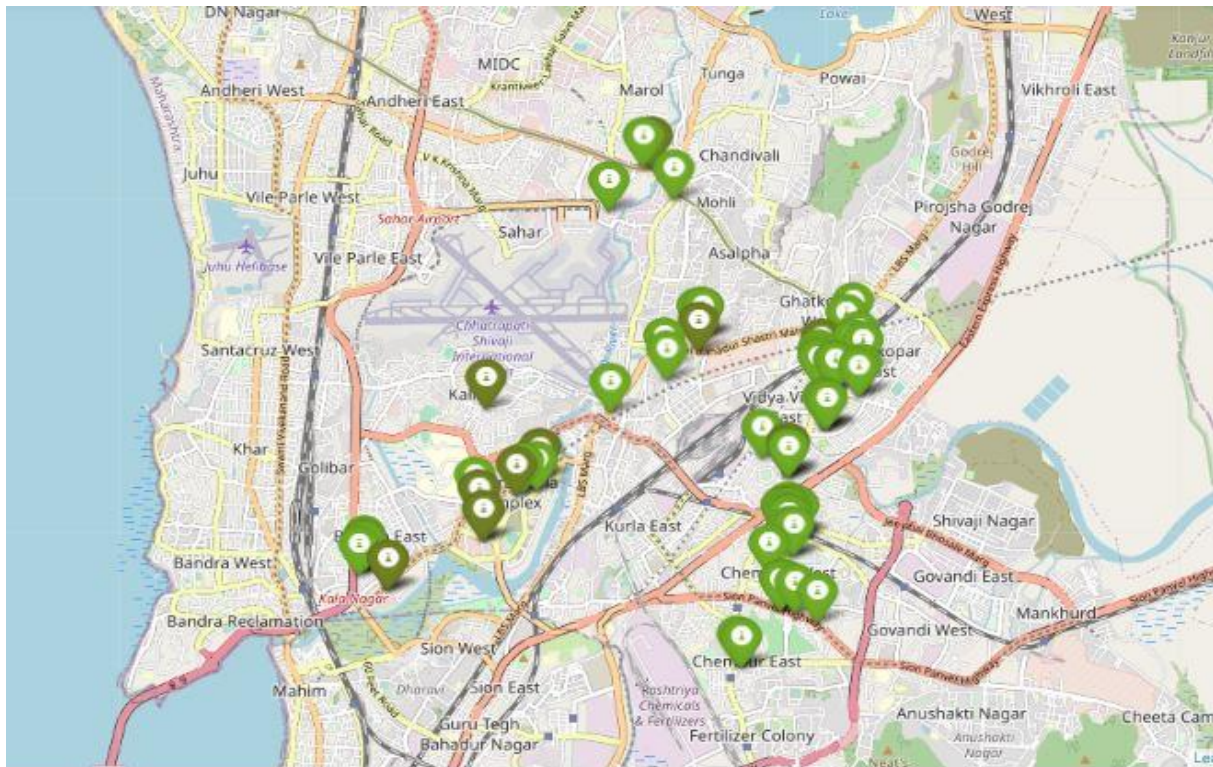*Figure 5: Rating and count of venues with that rating*

Figure 6: Plot of venues with different ratings

The venues in Mumbai that do not have many venues have rating more than 3. Overall, Mumbai on an average has good rating for its venues.

3.3 Price

Next, I explore the average prices of all venues for one person using a scatter plot along with the count of venues with that average price per person. Taking a look at figure 7, reveals that the majority venues have an average cost of Rs 500 to Rs 1000 for one person. Even though the maximum venues lie in that range, the actual range of prices is very different. There are places with average price even as high as Rs 1000+ for one person.
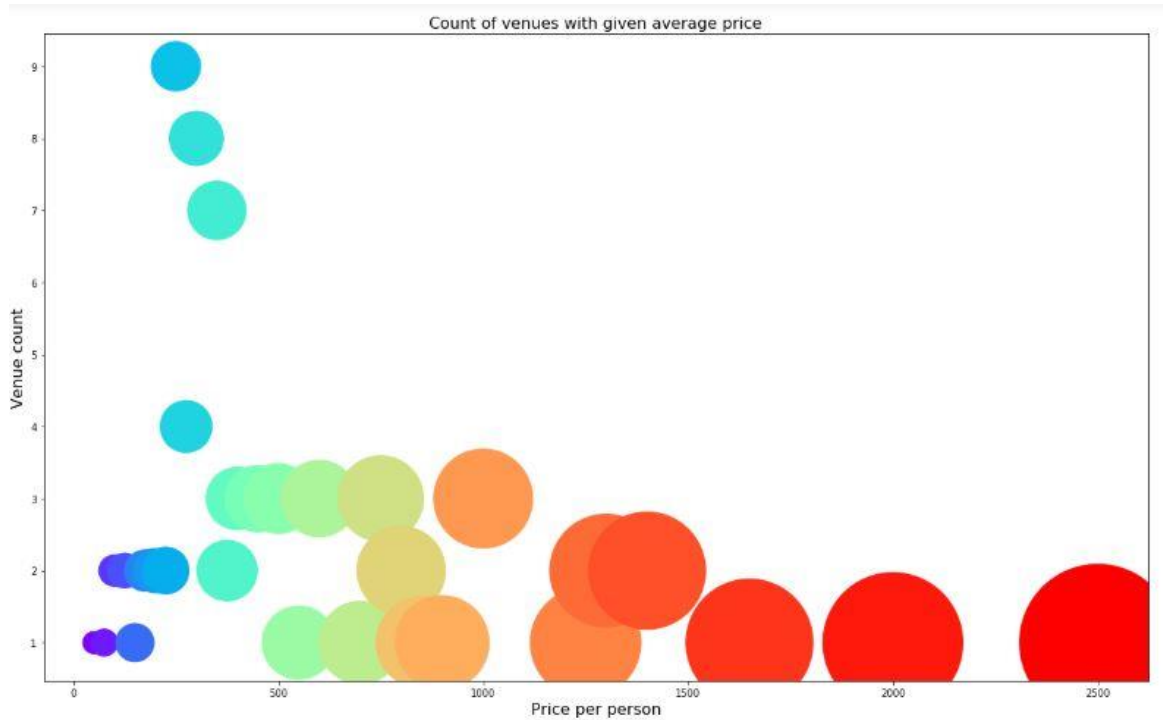
Figure 7: Price per person with count of venues with that price

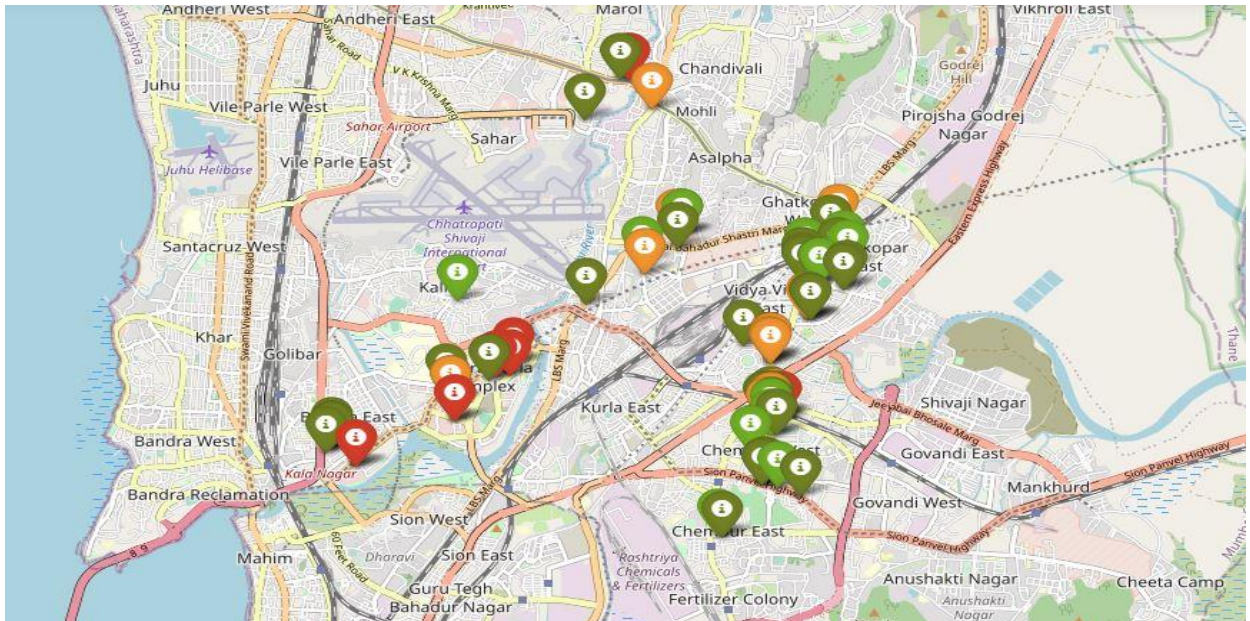I also plot the venues based on their price range.

Figure 8: Plot of venues with different prices

Figure 8 includes all the venues where high priced venues are marked by orange and red while the low priced venues are marked with green and dark green. From the plot, we observe that venues near Ghatkopar and Chembur are primarily lower priced. The venues near Bandra Complex

have steep prices. Marol Naka seems to have a mix of both high priced and low priced venues.

3.4 Clustering

Finally, I cluster all the venues based on their price range, location and more to identify similar venues and the relationship amongst them. I used KMeans clustering and decided to cluster the venues into two separate groups.
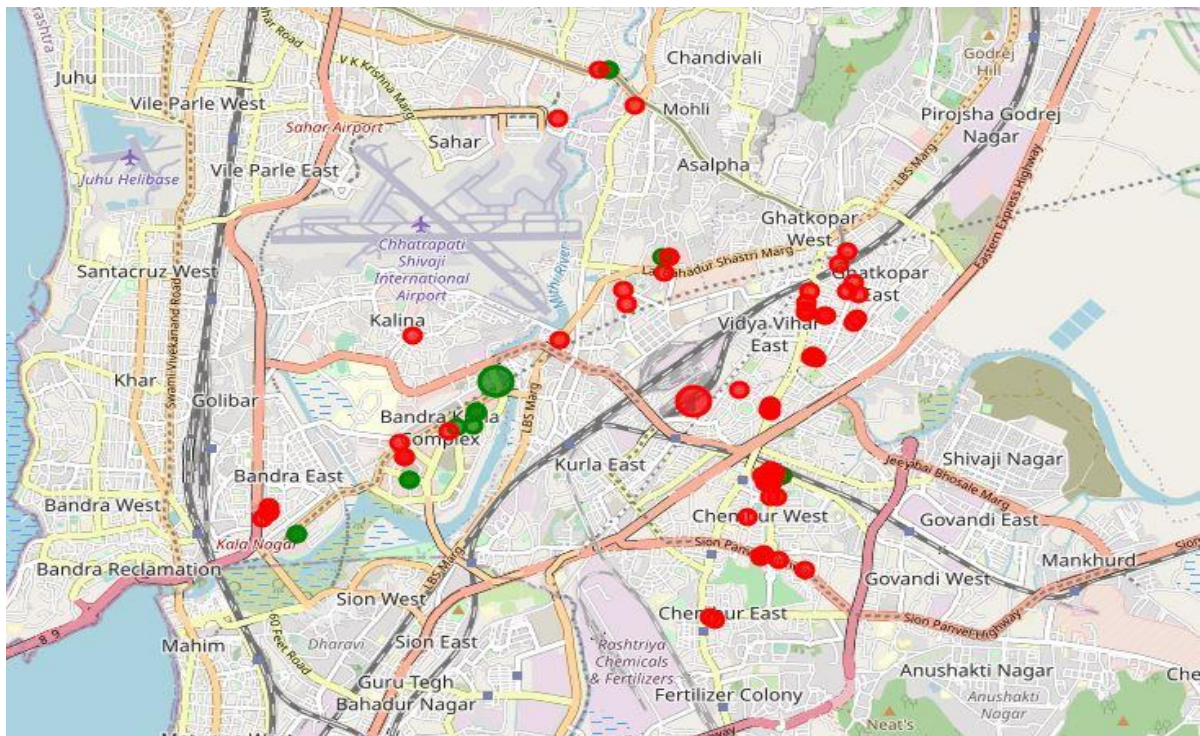

Figure 9: Clusters of venues

In figure 9, we see the two clusters:

1. The first cluster (red) is spread across the whole city and includes the majority venues. These venues have mean price range of 2.02 and rating spread around 3.79.

2. The second cluster (green) is very sparsely spread and has very limited venues. These venues have mean price range of 3.92 and rating spread around 4.17.

4. Results and Discussion

After collecting data from the Foursquare and Zomato APIs, we got a list of 181 different venues. However, not all venues from the two APIs were identical. Hence, we had to inspect their latitude and longitude values as well as names to combine them and remove all the outliers. This resulted in a total venue count of 75.

We identified that from the total set of venues, majority of them were Indian Restaurants. A visitor who loves Indian Restaurants would surely like to visit Mumbai.

While the complete range of ratings range from 1 to 5, the majority venues have ratings close to 3.75. This means that most restaurants provide good quality food which is liked by the people of the city, thus indicating the high rating.

When we take a look at the price values of each venue, we explore that many venues have prices which are in the range of Rs 500 to Rs 1000 for one person. However, the variation in prices is very large, given the complete range starts from Rs 100 and goes uptil Rs 2500. On plotting the venues based on their price range on the map, we discovered that venues located near Ghatkopar and Chembur are relatively priced lower than venues in Bandra Complex and Marol Naka.

Finally, through clusters we identified that there are many venues which are relatively lower priced but have an average rating of 3.79. On the other hand, there are few venues which are high priced and have average rating of 4.17.

If you're looking for cheap places with relatively high rating, you should check Ghatkopar and Chembur.

If you're looking for the best places, with the highest rating but might also carry a high price tag, you should visit Bandra Complex.

A company can use this information to build up an online website/mobile application, to provide users with up to date information about various venues in the city based on the search criteria (name, rating and price) or to start a food delivery business.

5. Conclusion

The purpose of this project was to explore the places that a person visiting Mumbai could visit. The venues have been identified using Foursquare and Zomato API and have been plotted on the map. The map reveals that there are three major areas a person can visit: Ghatkopar, Bandra Complex and Chembur. Based on the visitor's venue rating and price requirements, he/she can choose amongst the three places.