Exploring venues in Mumbai, India using Foursquare and Zomato API

Shubham Patil

May 2, 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.

The aim of the project is to identify venues in Mumbai, India based on their rating and average prices. In this notebook, we will identify various venues in the city of Mumbai, India, using Foursquare API and Zomato API, to help visitors select the restaurants that suit them the best.

Whenever a user is visiting a city they start looking for places to visit during their stay. They primarily look for places based on the venue ratings across all venues and the average prices such that the locations fits in their budget.

Here, we'll identify places that are fit for various individuals based on the information collected from the two APIs and Data Science. Once we have the plot with the venues, any company can launch an application using the same data and suggest users such information.
.

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 10 kilometers radius 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.zomato.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.
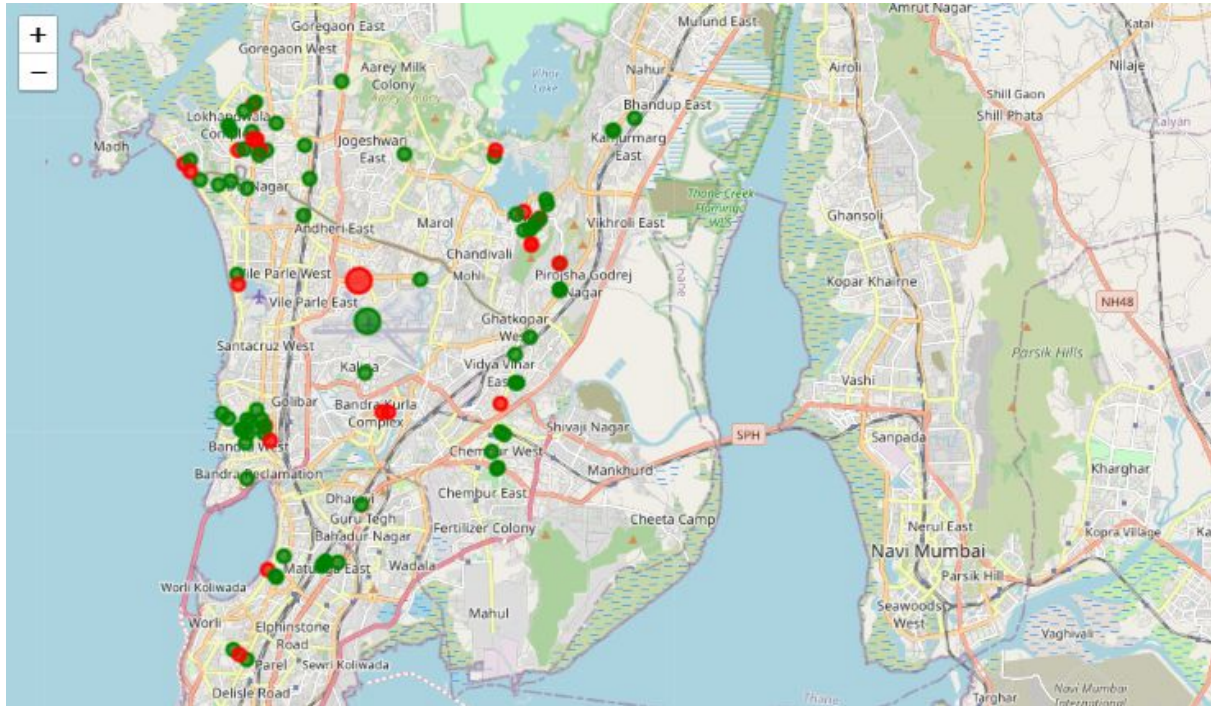 ● **Longitude:** The longitude value of the venue.

2.2 Data Cleaning

Figure 1: Venues retrieved from Foursquare API

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.
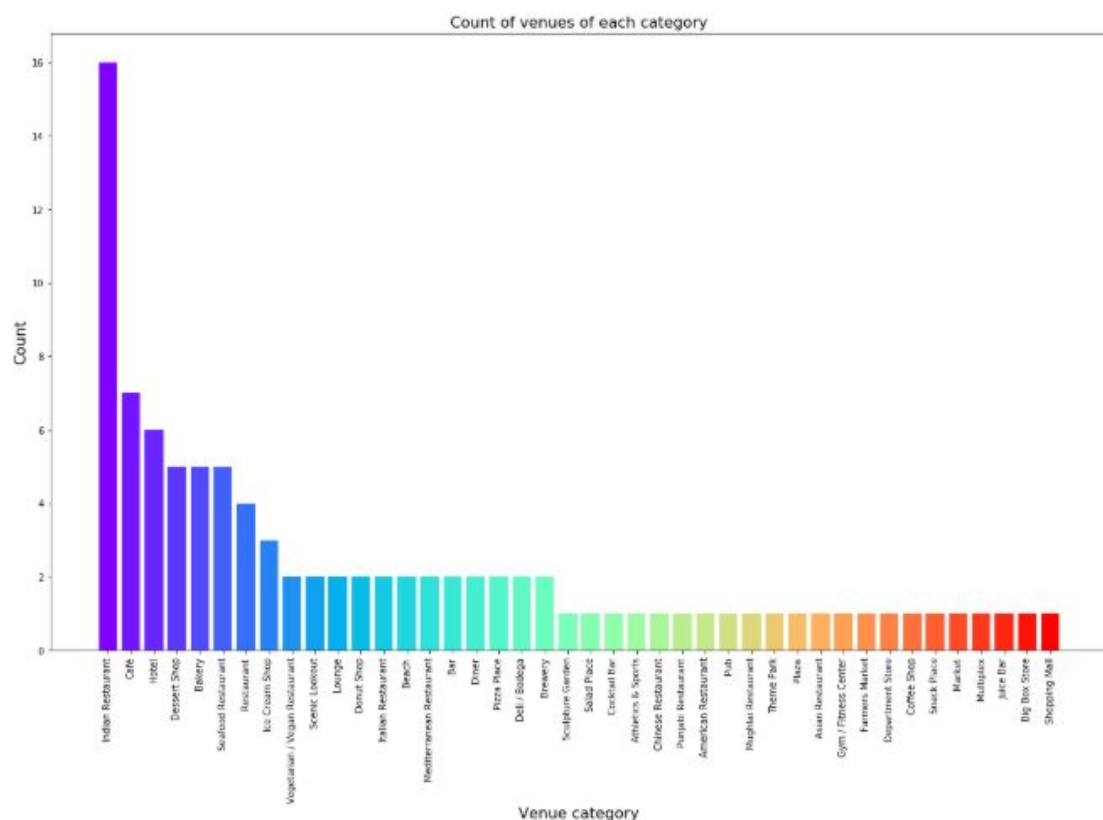

Figure 4: Count of various types of venues in Mumbai

From figure 4, we see that the majority venues are actually Cafes. This is closely followed by Indian Restaurants. For someone who is visiting Mumbai and loves either Cafes or 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 4 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. The map has the location of all the venues. It appears that many venues are located around the area of Kurla, Bandra and Chembur with rating above 3
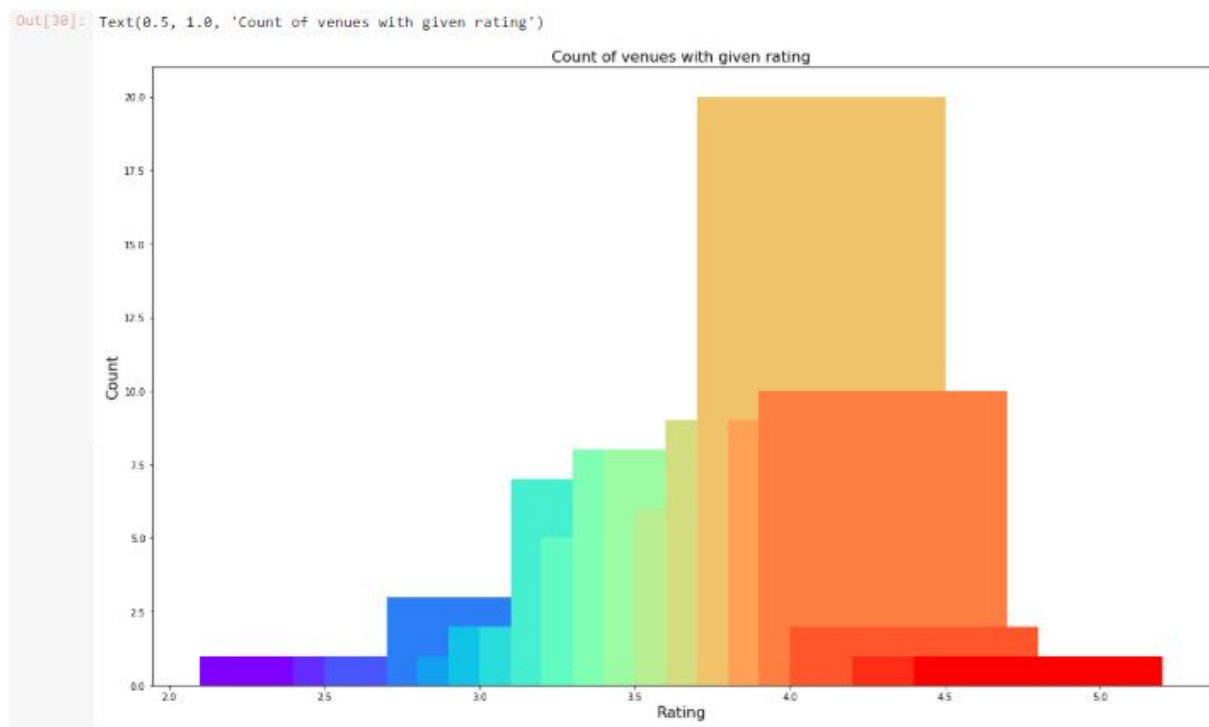

Out[30]: Text(0.5, 1.0, 'Count of venues with given rating')

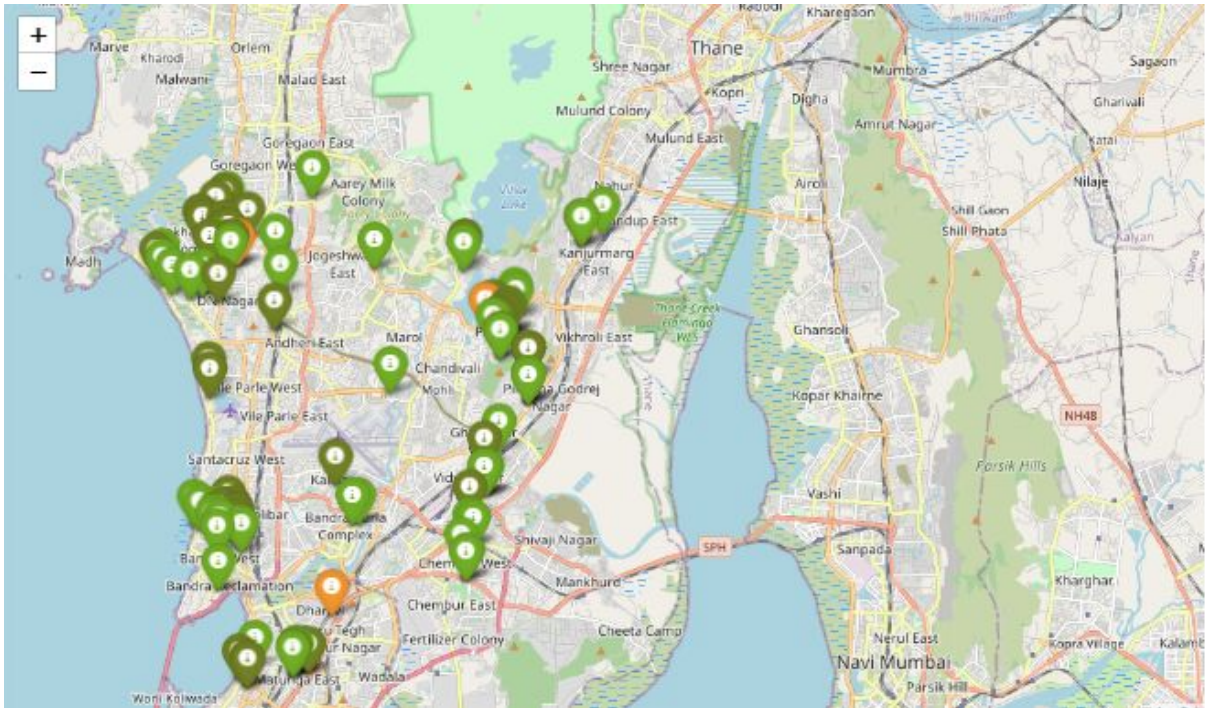*Figure 5: Rating and count of venues with that rating*

Figure 6: Plot of venues with different ratings

The venues in sectors 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

From the plot below we can see that the price range is well scattered, it means we have enough number of restaurants & cafes for every price range.

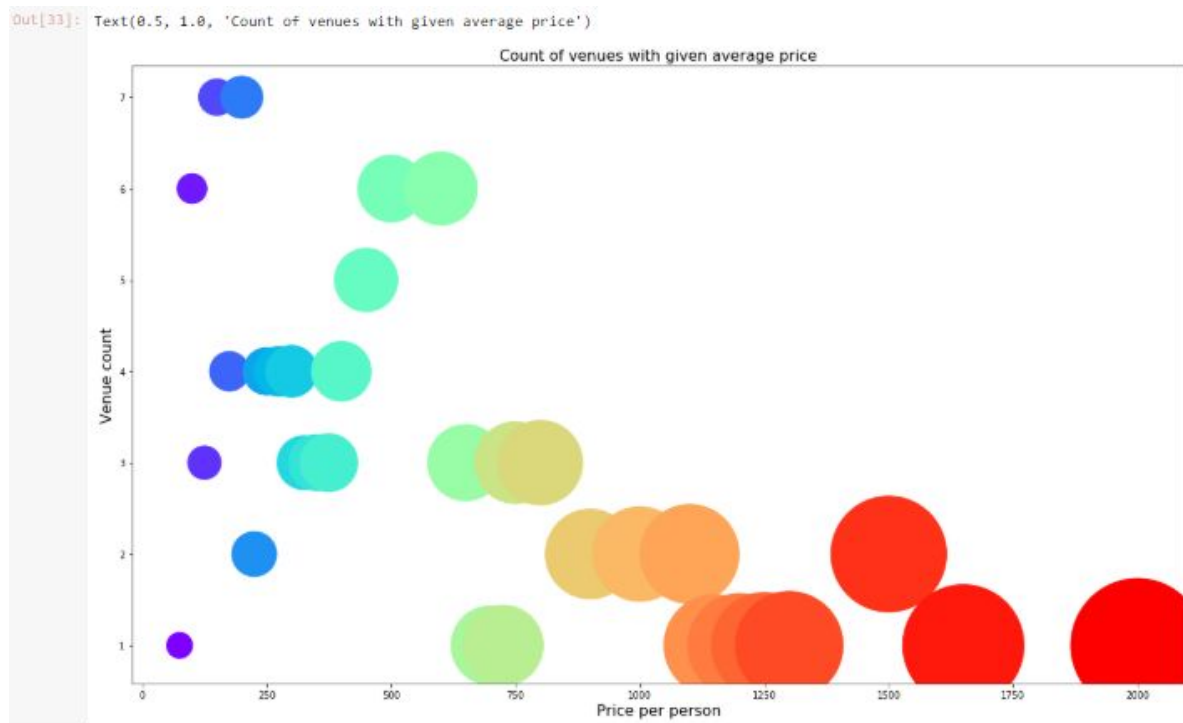Users might also be interested in going to a place that fits in their budget.

Text(0.5, 1.0, 'Count of venues with given average price')

Figure 7: Price per person with count of venues with that price

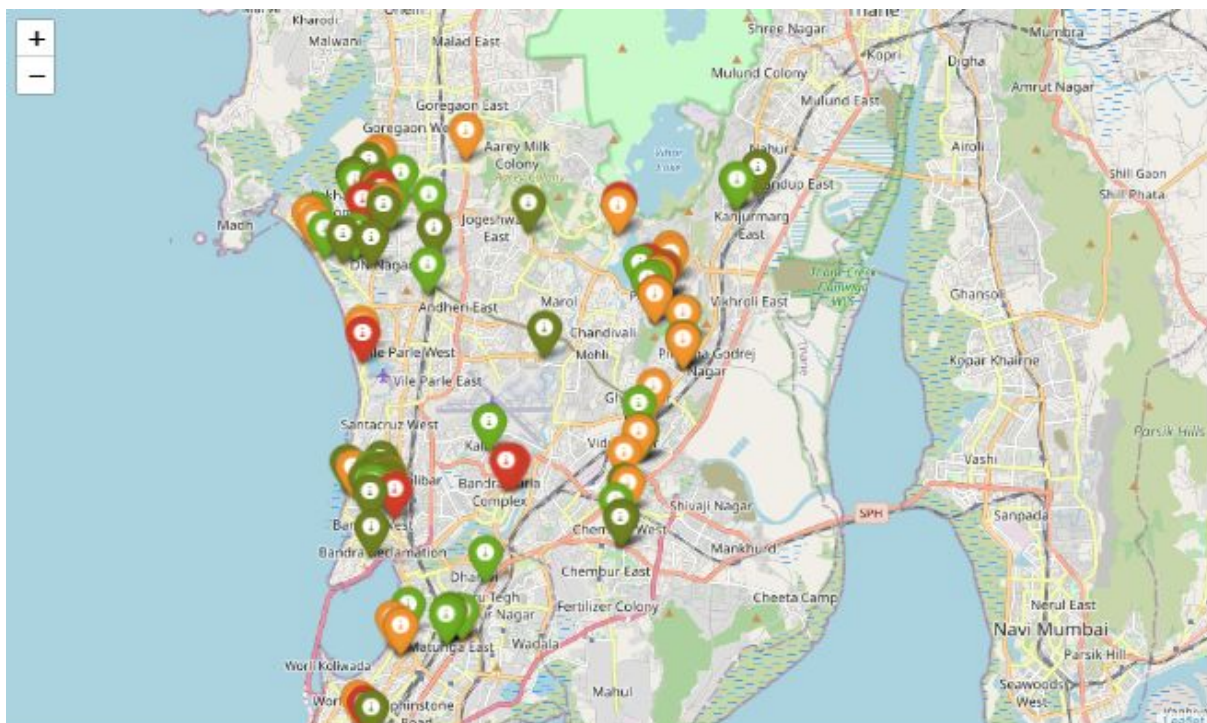I also plot the venues based on their price range.

Figure 8: Plot of venues with different prices

The venues around Kondhwa & Deccan Gymkhana Road are primarily lower priced. The venues around Kurla and Bandra have medium to steep prices. Places near Thane & Kalyan have low price restaurants.

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.
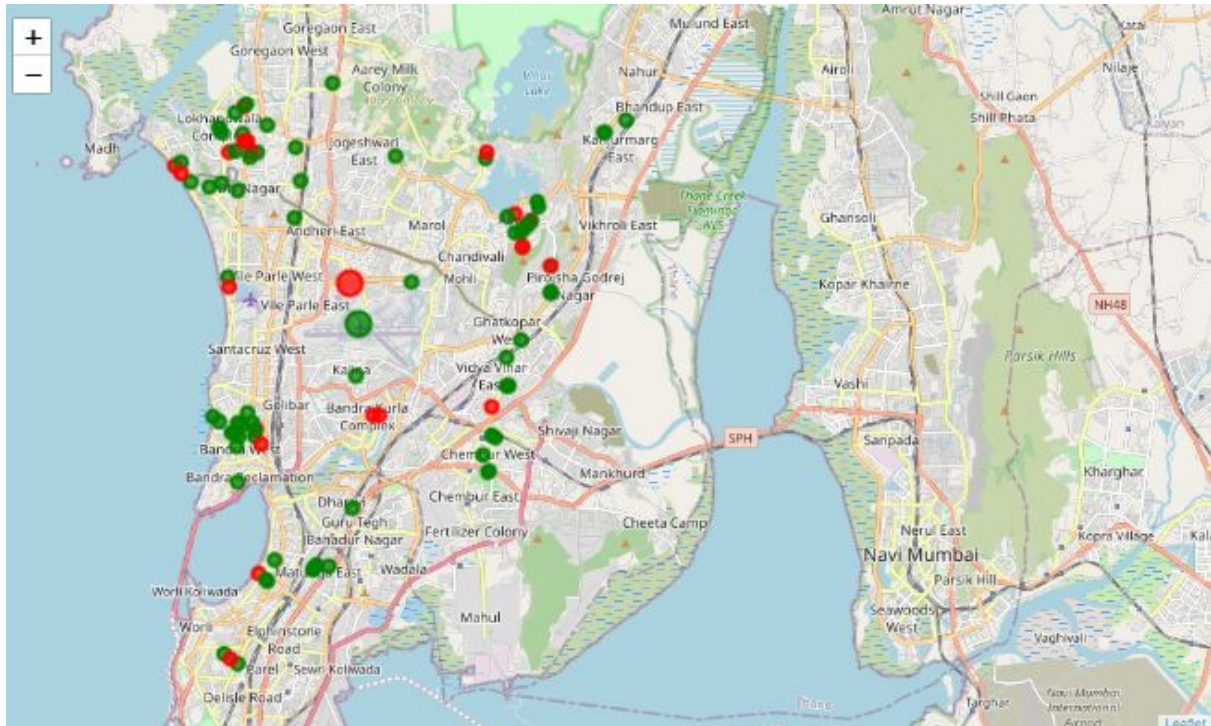

Figure 9: Clusters of venues

From the map, we see the two clusters:

- The first cluster is spread across the whole city and includes the majority venues.
- The second cluster is very sparsely spread and has very limited venues.

4. Results and Discussion

Based on our analysis above, we can draw a number of conclusions that will be useful to aid any visitor visiting the city of Mumbai, India.

We identified that from the total set of venues, majority of them were Cafes and Indian Restaurants. A visitor who loves Cafes/Indian Restaurants would surely benefit from coming

to Mumbai. Many venues are located around the area of Kurla , Chembur, Bandra and Sion with rating above 3.

While the complete range of ratings range from 1 to 5, the majority venues have ratings close to 4. This means that most restaurants provide good quality food which is liked by the people of the city, thus indicating the high rating.

However, the variation in prices is very well scattered in Mumbai, given the complete range starts from Rs 100 and goes upto Rs 1200. On plotting the venues based on their price range on the map, we discovered. The venues around Kurla & Kalyan are primarily lower priced.The venues around Chembur & Bandra have medium to steep prices.Places near Sion and Thane have low price restaurants.

Finally, through clusters we identified that there are many venues which are relatively lower priced but have an average rating of 3.57. On the other hand, there are few venues which are high priced and have an average rating of 4.03.
.
A company can use this information to build 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).

5. Conclusion

The purpose of this project was to explore the places that a person visiting Mumbai could explore. 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:Kurla, Bandra, Sion and Chembur.