EXPLORING VENUES IN CHENNAI USING FOURSQUARE API AND ZOMATO API

ANAND KUMAR E

29-06-2020

TABLE OF CONTENTS

1.Introduction

The aim of this project is to identify venues in Chennai, India based on their rating and average prices. In this notebook, we will identify various venues in the city of Chennai, 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.

1.1 Interested audience

The target audience for such a project is twofold. Firstly, any person who is visiting Chennai, 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 Collection

2.1 Data Sources

To get location and other information about various venues in Chennai, 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 15 kilometers from the center of Chennai 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.
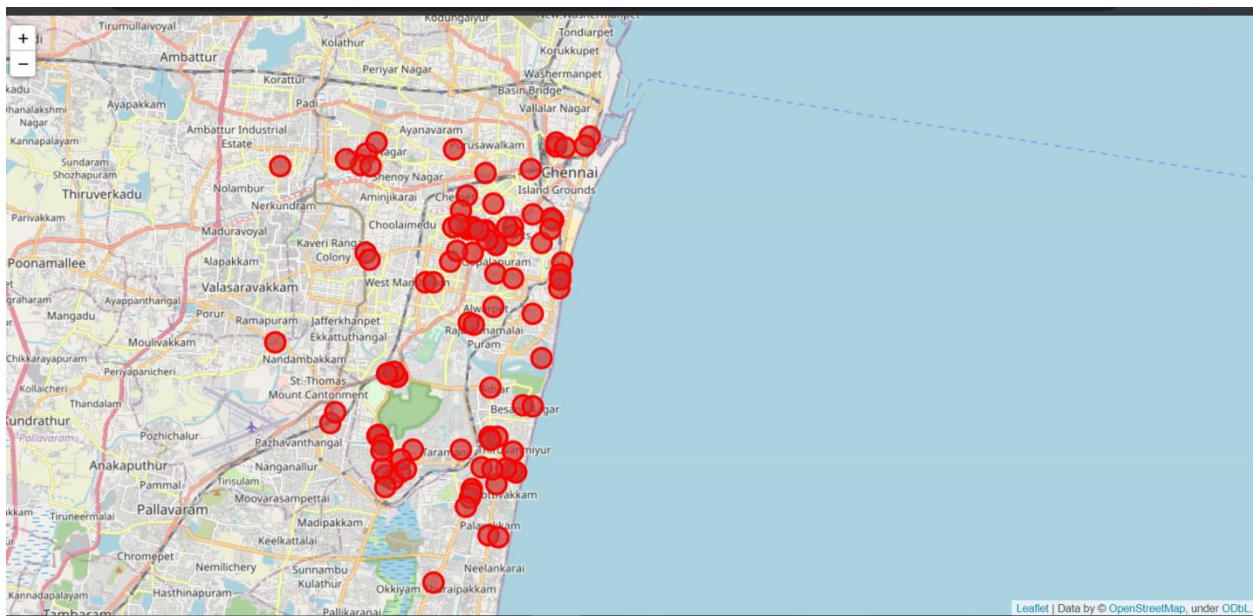
- **Longitude:** The longitude value of the venue.

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.

3. Data Preparation
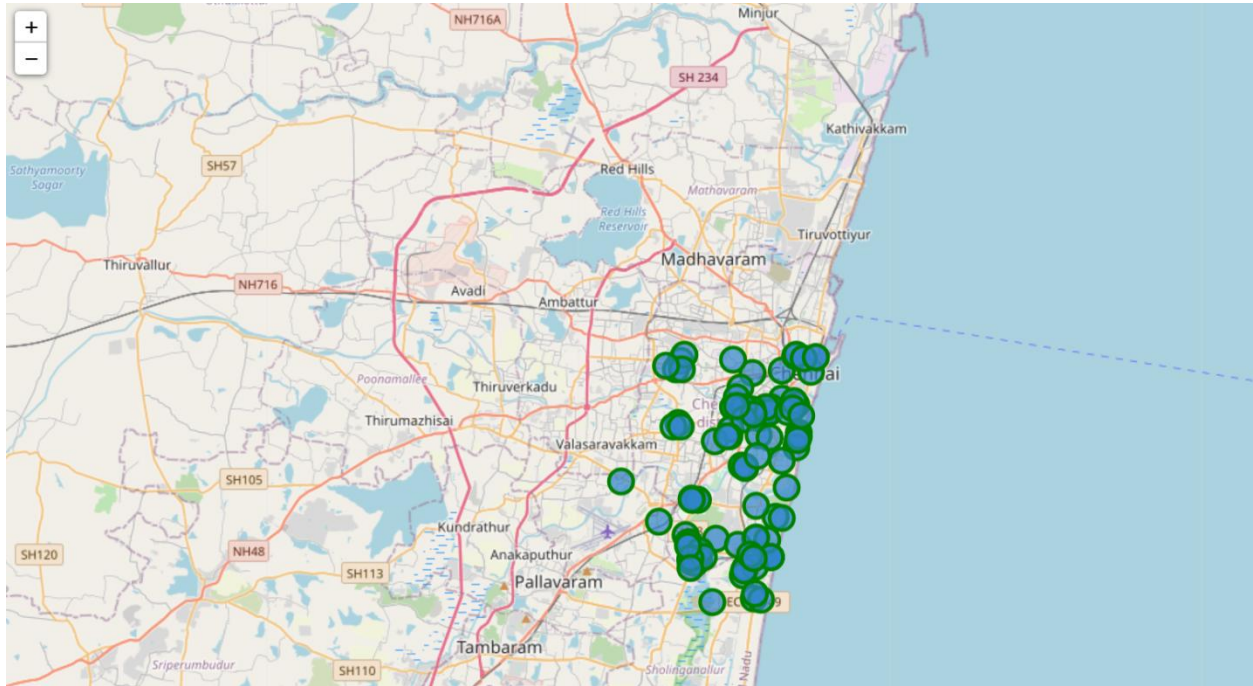
1.Venues by Foursquare API

2.Venues by 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 categorized as follows:

1. There are venues that have specific restaurants/cafes inside them as provided by Zomato.

2. Some 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 55 venues with 8 columns as described in figure 3.

	categories	venue	latitude	longitude	price_range	rating	address	average_price
0	Indian Restaurant	Hotel Greens Gate	13.0812	80.2703	2.0	0	12 & 13, V V Koil Street, Opposite Rippon Buil...	275.0
1	Snack Place	Links	13.0799	80.2545	1.0	4.0	22, Raja Annamalai Road, Purasavakkam, Chennai	125.0
2	Indian Restaurant	Shree Mithai	13.0722	80.2479	1.0	4.6	18, Dr TV Road, Chetpet, Chennai	150.0
3	Movie Theater	ID	13.0554	80.2581	1.0	4.5	8, 1st Floor, Sathyam Cinemas Complex, Thiruvi...	150.0
4	Juice Bar	The Cake World	13.0604	80.2546	2.0	3.6	11/2, Greams Road, Near Police Station, Thousa...	250.0

3. Aggregated Final Data from both API's

4. Methodology and Exploratory Data Analysis

As a first step, I retrieve the venues in Chennai from Foursquare and Zomato APIs. I extract the location data from the Foursquare API for all venues up to a distance of 15 kilometers from the center of Chennai. 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 Chennai. 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 Chennai.

As a final step, I will analyze 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.

4.1 Categories.

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

Count of venues of each category

tear container object of 50 artists?
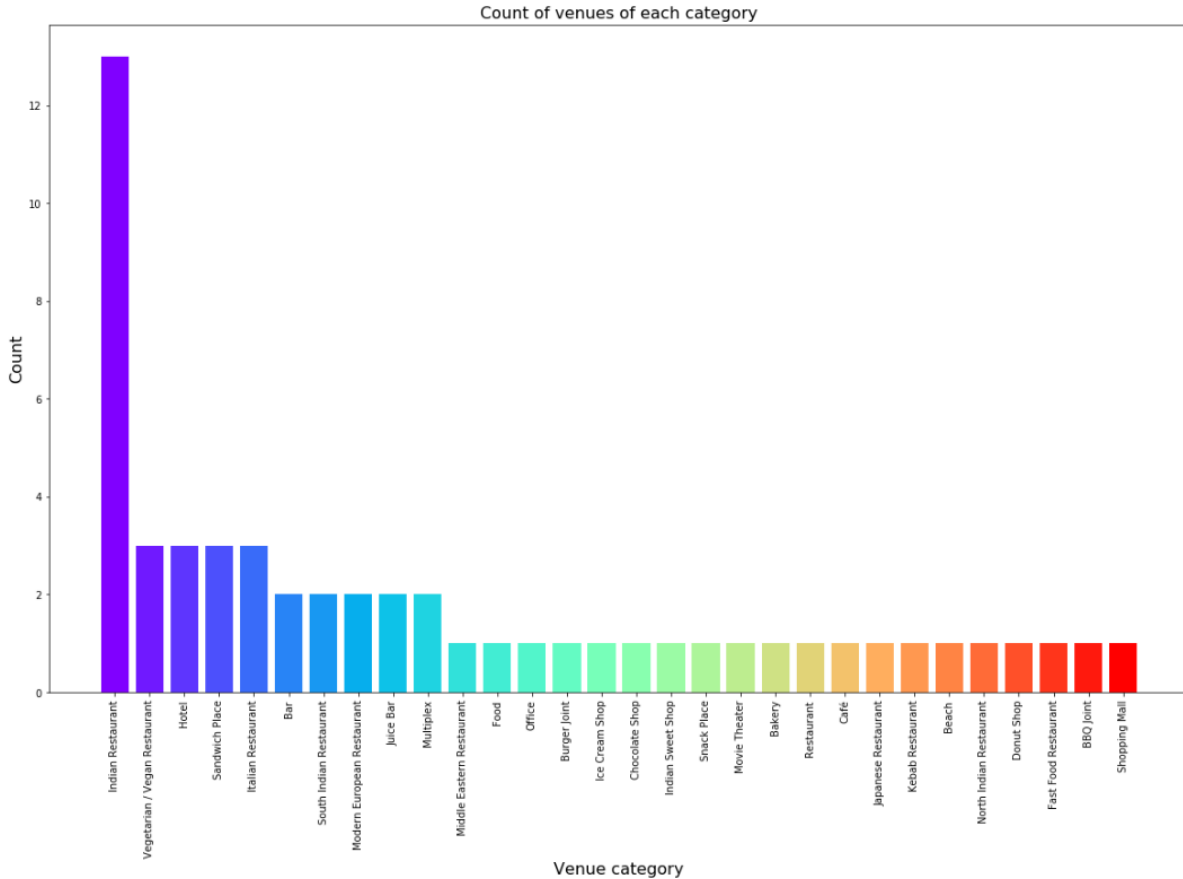
4.Venue Categories

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 Chennai and loves either Cafes or Indian Restaurants, they'd surely love their stay.
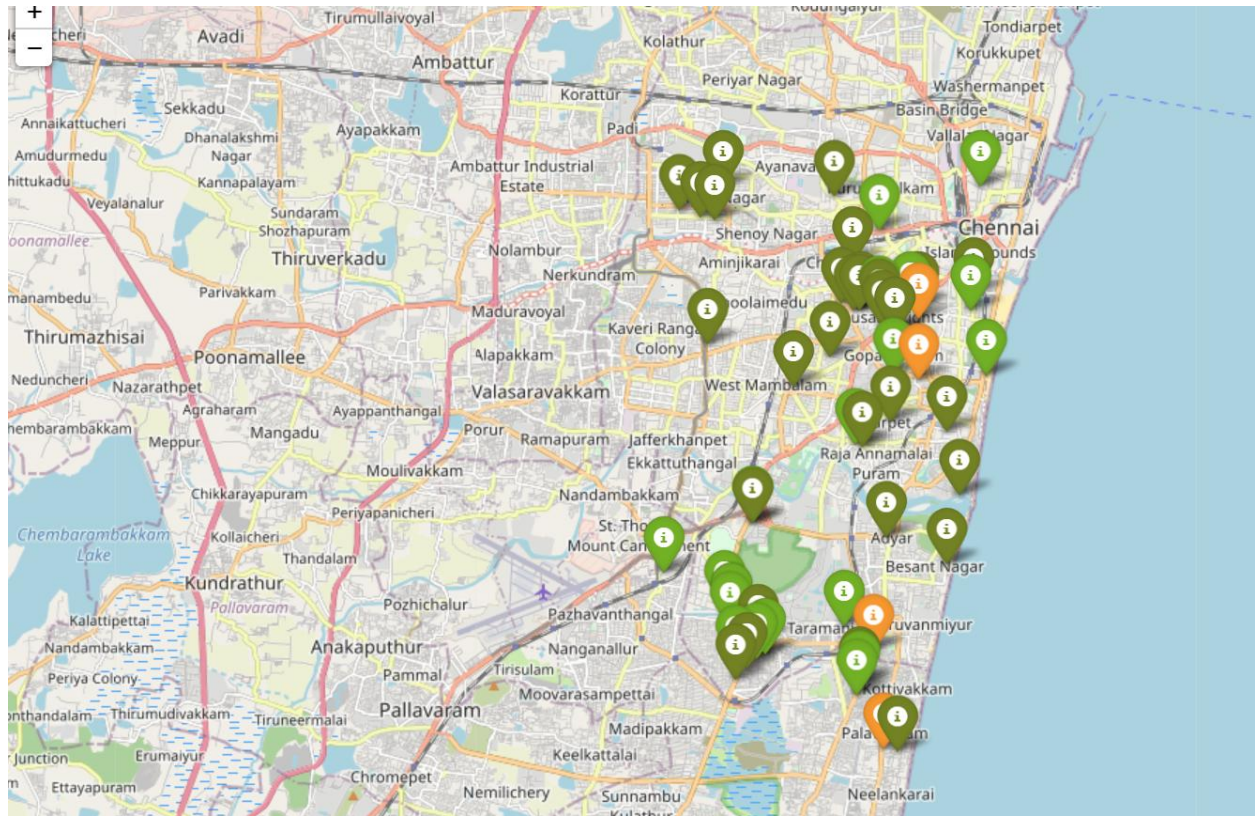
4.2 Rating

Next, I'll explore the ratings of various venues in Chennai. 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 Chennai. 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 Chennai. 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. 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.
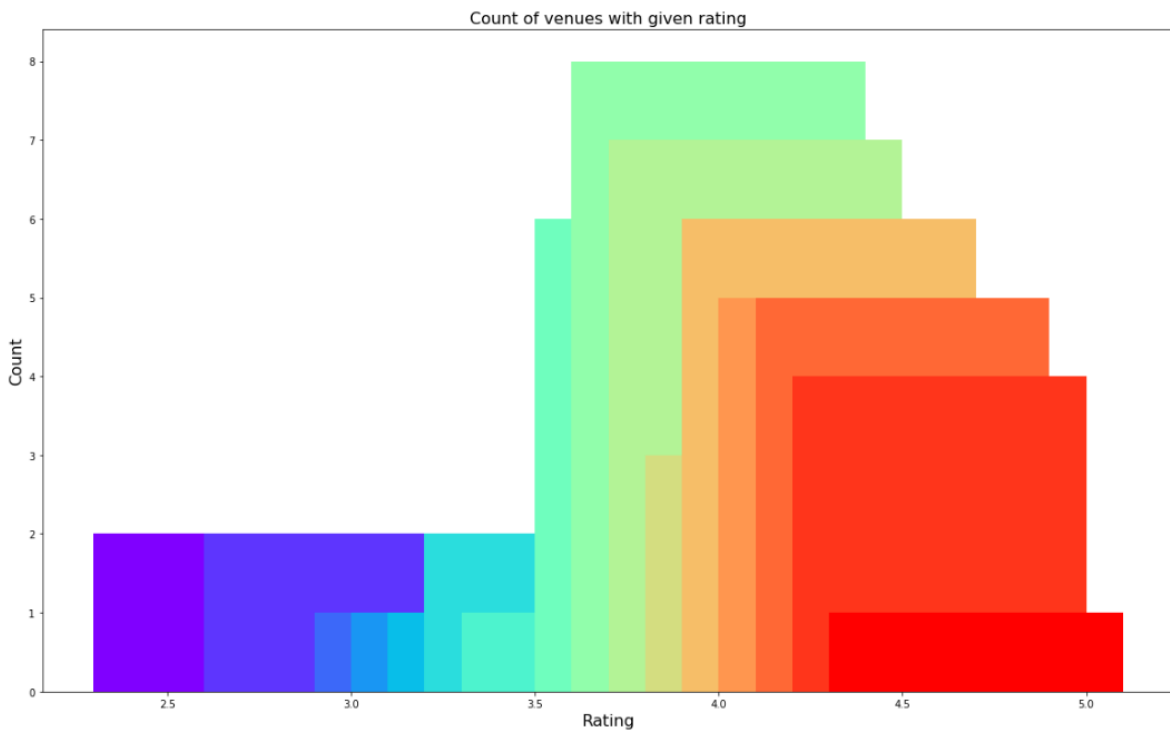

5.Rating and Count of venues

6.Venues with different ratings

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

4.3 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 200 to Rs 400 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.

Count of venues with given average price
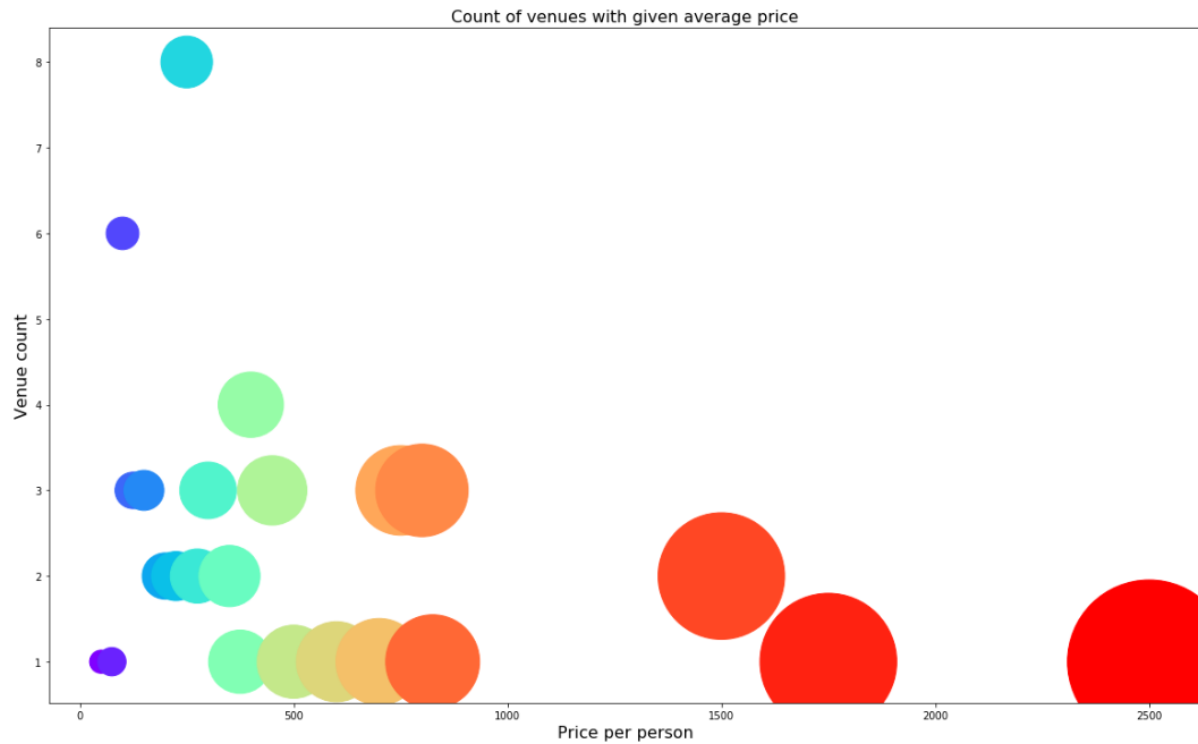
7. Count of venues with average price

I also plot the venues based on their price range.
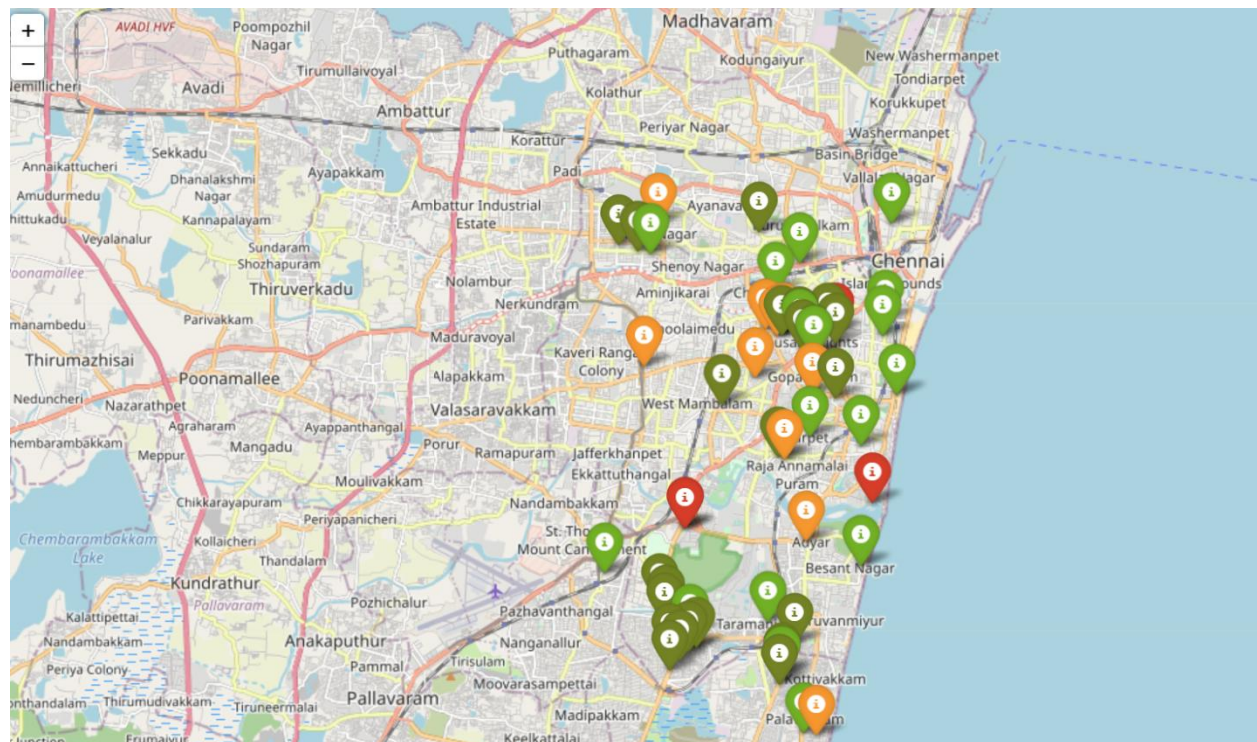
8. Price range of venues

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 Tambaram and Royapettah are primarily lower priced. The venues near Guindy and MRC Nagar have steep prices. Thousand Lights seems to have a mix of both high priced and low priced venues.

4.4 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 K Means clustering and decided to cluster the venues into two separate groups.
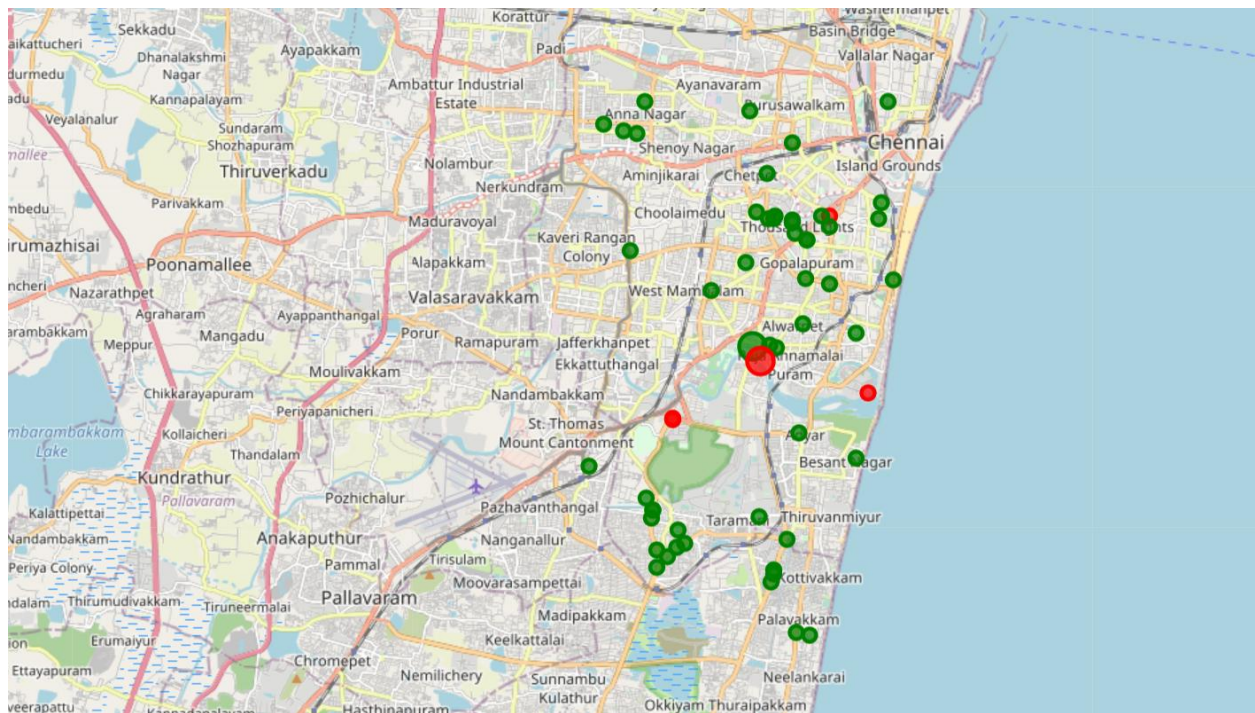
9. Cluster of Venues

In figure 9, we see the two clusters:

1. The first cluster (green) is spread across the whole city and includes the majority venues. These venues have mean price range of mean price range of 1.84 and rating spread around 4.04.

2. The second cluster (red) is very sparsely spread and has very limited venues. These venues have mean price range of 4.00 and rating spread around 4.12.

5.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 Chennai, India.

After collecting data from the Foursquare and Zomato APIs, we got a list of 102 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 55.

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

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. When we plot these venues on the map, we discover that there are clusters of venues around Guindy and Royapettah. These clusters also have very high ratings (more than 3).

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 100 to Rs 900 for one person. However, the variation in prices is very large, 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 that venues located near Anna Nagar and are relatively priced lower than venues in Guindy and MRC Nagar . A mix of low price and high price exist in Thousand Lights.

Finally, through clusters we identified that there are many venues which are relatively lower priced but have an average rating of 4.04. On the other hand, there are few venues which are high priced and have average rating of 4.12.

If you're looking for cheap places with relatively high rating, you should check Tambaram. If you're looking for the best places, with the highest rating but might also carry a high price tag, you should visit Guindy and MRC Nagar If you're looking to explore the city and have no specific criteria to decide upon the places you want to visit, you should try Thousand Lights. 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).

6. Conclusion

The purpose of this project was to explore the places that a person visiting Chennai 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: Tambaram, Guindy and Thousand Lights based on the visitor's venue rating and price preferences, he/she can choose amongst the three places.