Exploring venues in Jamshedpur – The Steel City of India Using Foursquare and Zomato API

SUPRIYO SINHA

June 17, 2020

1. Introduction

1.1 Background

Jamshedpur is one of the first industrial planned cities of India and most populous urban agglomeration in the Indian state of Jharkhand. It was founded by Jamsetji Tata (Founder of Tata Groups) and was also named after him. Jamshedpur is also known as Tatanagar.

Jamshedpur is ranked as 2nd in India in terms of 'Quality of Life'. The city is located in nature's den in Chota Nagpur plateau, surrounded by the picturesque Dalma Hills and bordered by the rivers Subarnarekha and Kharkai.
Tata Steel, the largest iron and steel producing plant in India and also the oldest is situated almost at the center of the city.

1.2 Problem statement

The objective of this capstone project is to explore the different venues in Jamshedpur. Here we will be using different Machine Learning algorithms like k-means-clustering, etc.

Questions that would be answered:
- List of Cities in Jharkhand
- Plotting the cities on the map of Jharkhand
- Plotting the map of Jamshedpur with the venues superimposed on it
- Analyzing the venues on basis of category, ratings and price
- Clustering the venues

2. Data

2.1 Data Sources

We have collected the list of cities of Jharkhand from the Wikipedia link: -
https://en.wikipedia.org/wiki/List_of_cities_in_Jharkhand_by_population
To get the coordinates of the cities I used Python **Geocoders** library.

To get location and other information about various venues in Jamshedpur, I used two APIs and combined 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 from the center of Jamshedpur 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://developer.foursquare.com/), 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.
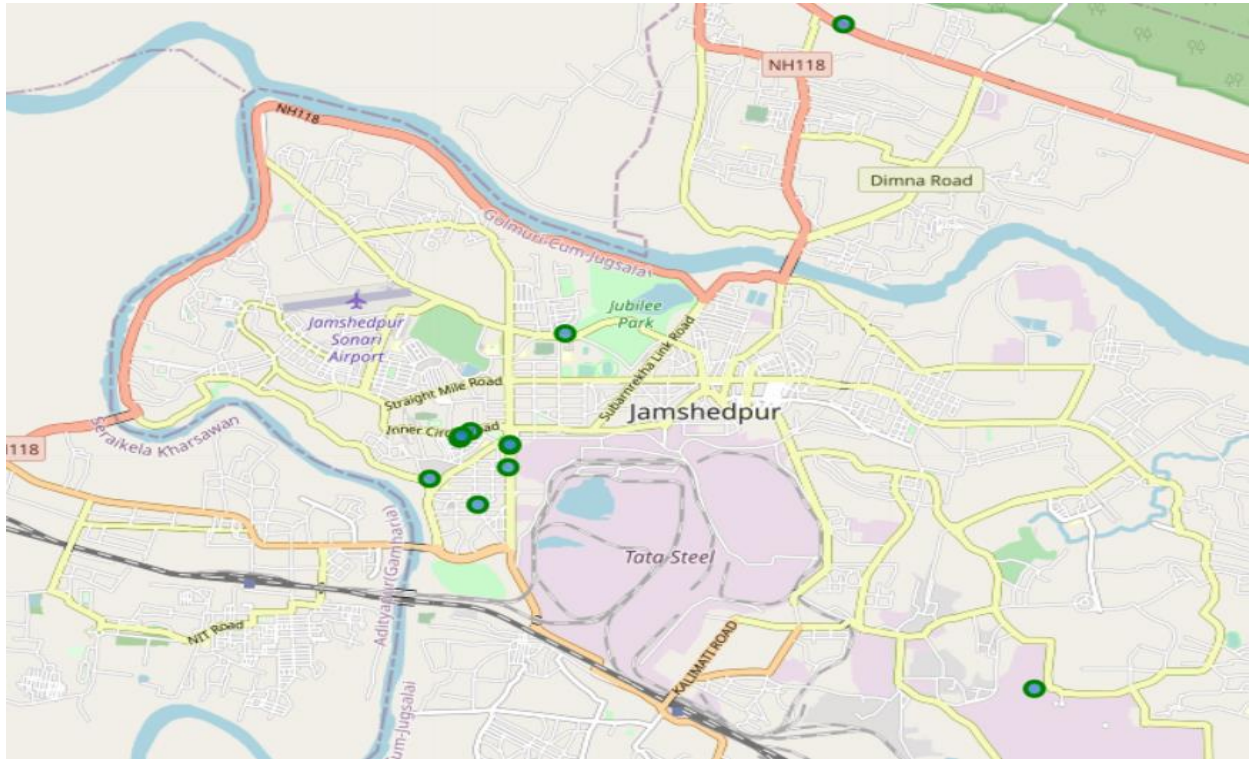
2.2 Data Cleaning

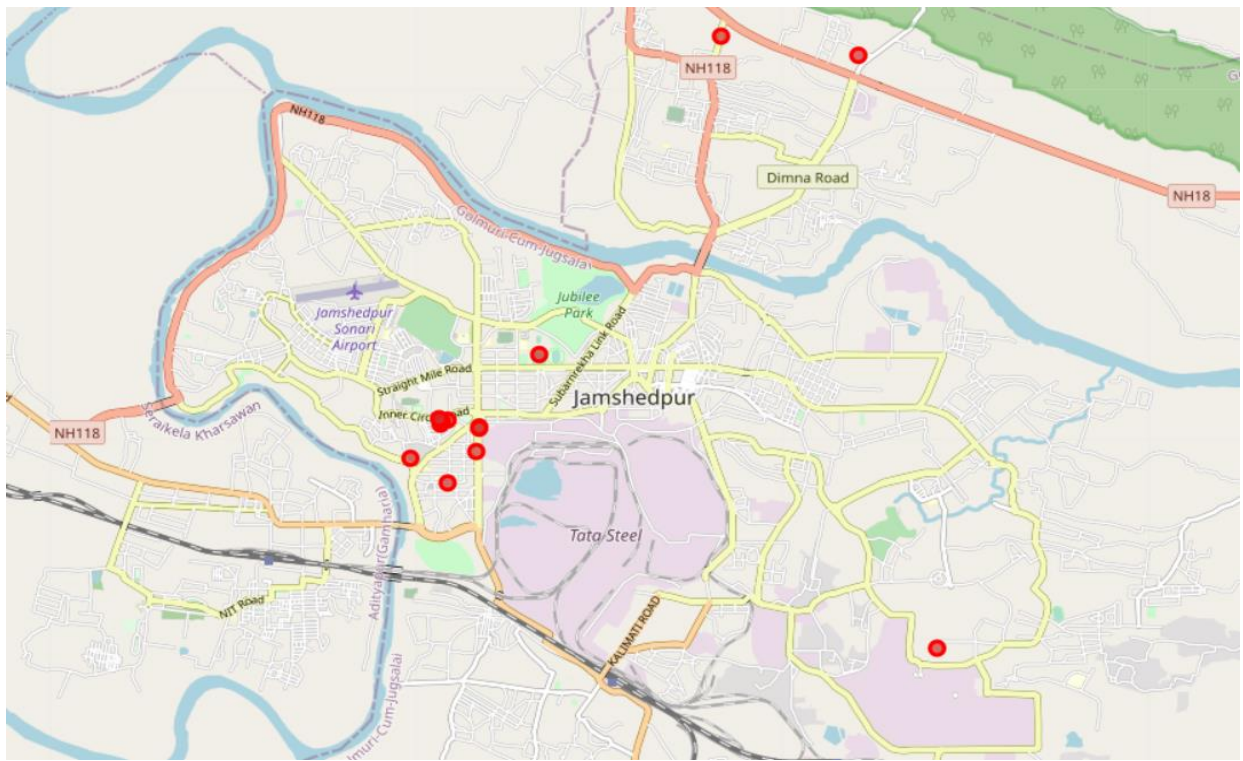*Figure 1: Venues retrieved from Foursquare API*

Figure 2: Venues retrieved from Zomato API

Next, I combined the two data of the Foursquare and Zomato using their latitudes and longitudes 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.

As a final dataset, we're left with 08 venues with 8 columns as described in figure 3.

	categories	venue	latitude	longitude	price_range	rating	address	average_price
0	Bakery	Brubeck Bakery	22.7978	86.1843	2	4.4	Dcosta Mansion, Main Road, Bistupur, Jamshedpur	225.0
1	Café	Cafe Coffee Day	22.7976	86.1843	2	3.3	D'costa Mansion, 669 Main Road, Bistupur, Jams…	325.0
2	Pizza Place	Domino's Pizza	22.7914	86.1806	2	4.4	Roshan Tower, Beside Sachdev Cycle Mart, Bistu…	200.0
3	Hotel	Jampot	22.7942	86.1761	4	3.7	Hotel Ramada, Holding 3, Bistupur	800.0
4	Hotel	The Chocolate Room	22.7985	86.1797	2	4.4	Contractors Area, Near The Sonnet Hotel, Bistu…	300.0
5	Indian Restaurant	Dosa King	22.7982	86.1797	2	3.9	Contractors Area, Near The Sonnet Hotel, Bistu…	250.0
6	Market	Spice Zone	22.7951	86.1839	3	3.9	Natraj Mansion Main Road Bistupur	500.0
7	Hotel	Sugar Works	22.7988	86.1795	2	3.9	Inner Circle Road, Contractors Area, Bistupur,…	250.0

Figure 3: Final data aggregated from both APIs

3. Methodology and Exploratory Data Analysis

As a first step, I retrieve the venues in Jamshedpur from Foursquare and Zomato APIs. I extract the location data from the Foursquare API for all venues up to a distance of 10 kilometers from the center of Jamshedpur. Using this, I fetch the venue information including price and rating data from Zomato API.

Using data cleaning, the dataset from the two APIs is combined on the basis of 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 Jamshedpur. 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 Jamshedpur. As a final step, I will analyze these plots and try to draw conclusions on what places can be recommended to visitors.

3.1 Categories

I begin my analysis by taking a look at the various categories of venues that exist in Jamshedpur.

Count of venues of each category
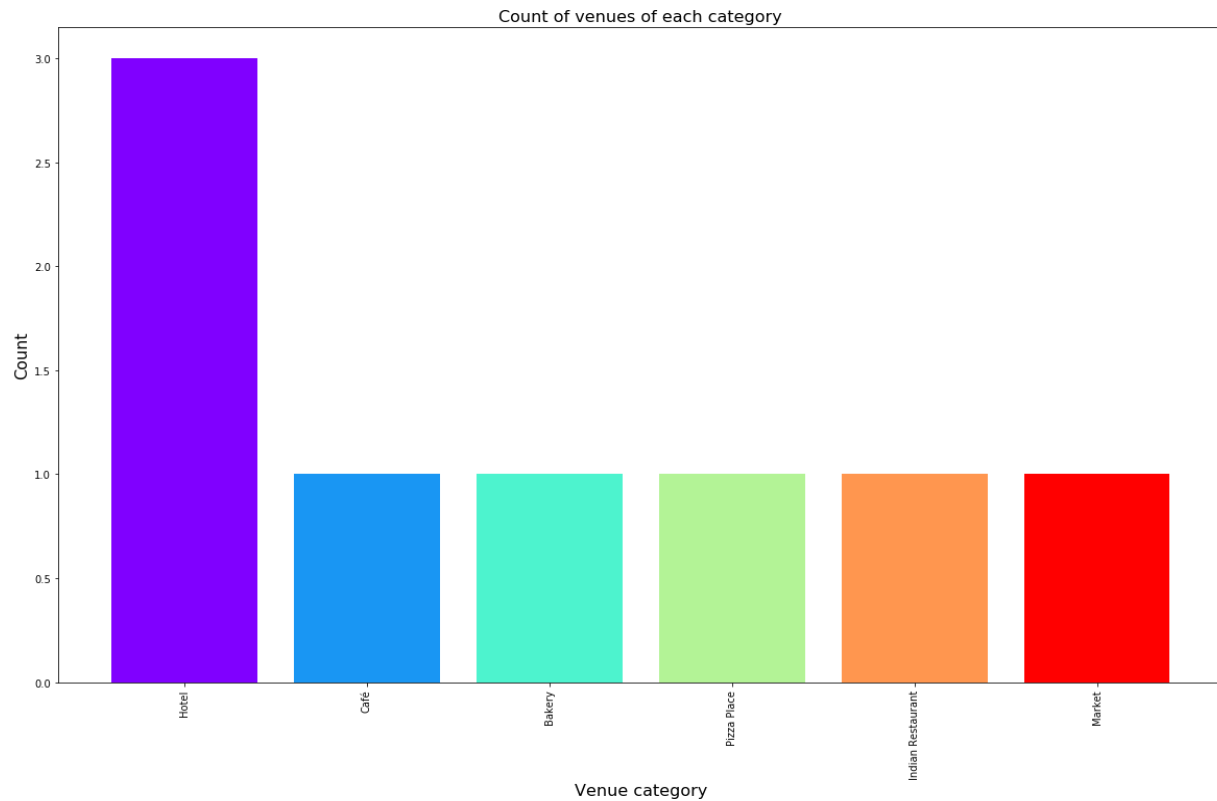
From figure 4, we *Figure 4: Count of various types of venues in Jamshedpur* closely followed by Café, Bakery, etc. For someone who is visiting Jamshedpur would have a variety of choices of hotels, café, bakery, markets, etc. to visit.

3.2 Rating

Next, I'll explore the ratings of various venues in Jamshedpur. 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 Jamshedpur. This can be seen in figure 5.
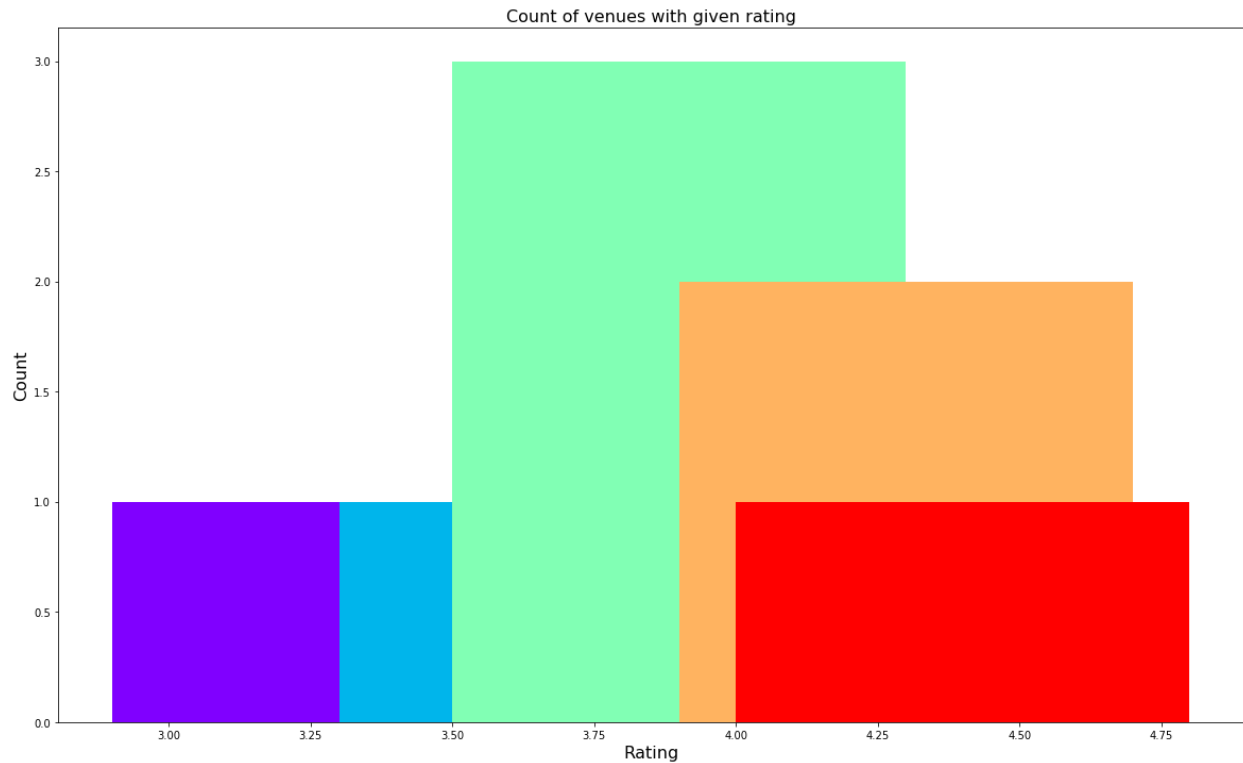

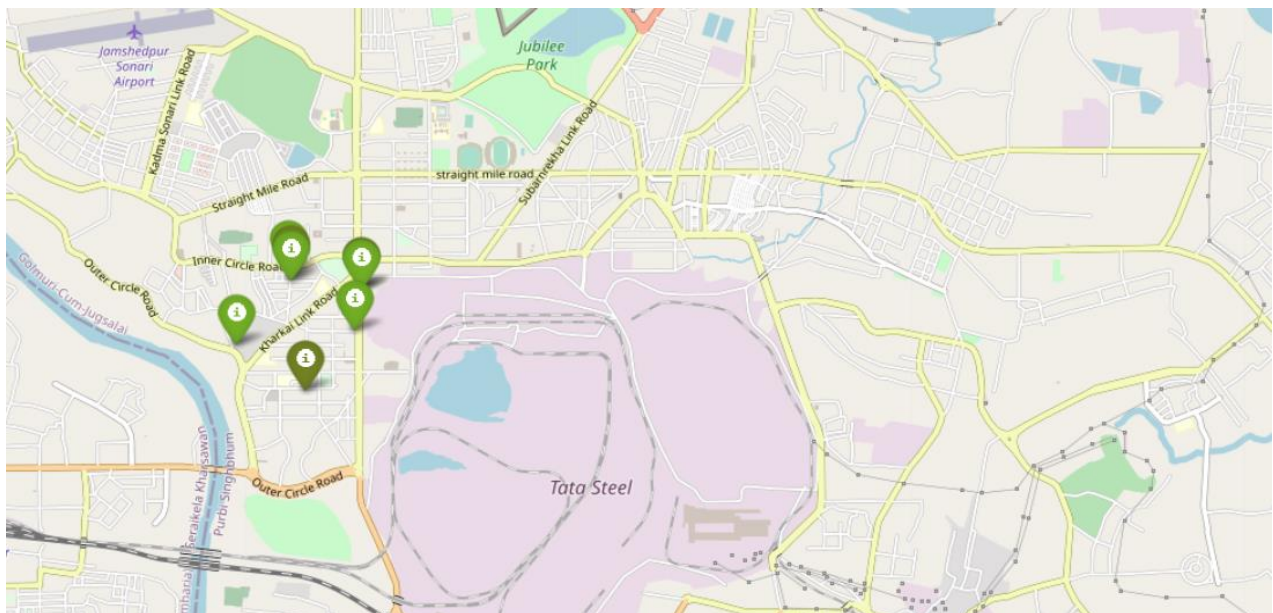*Figure 5: Rating and count of venues with that rating*

Figure 6: Plot of venues with different ratings

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.
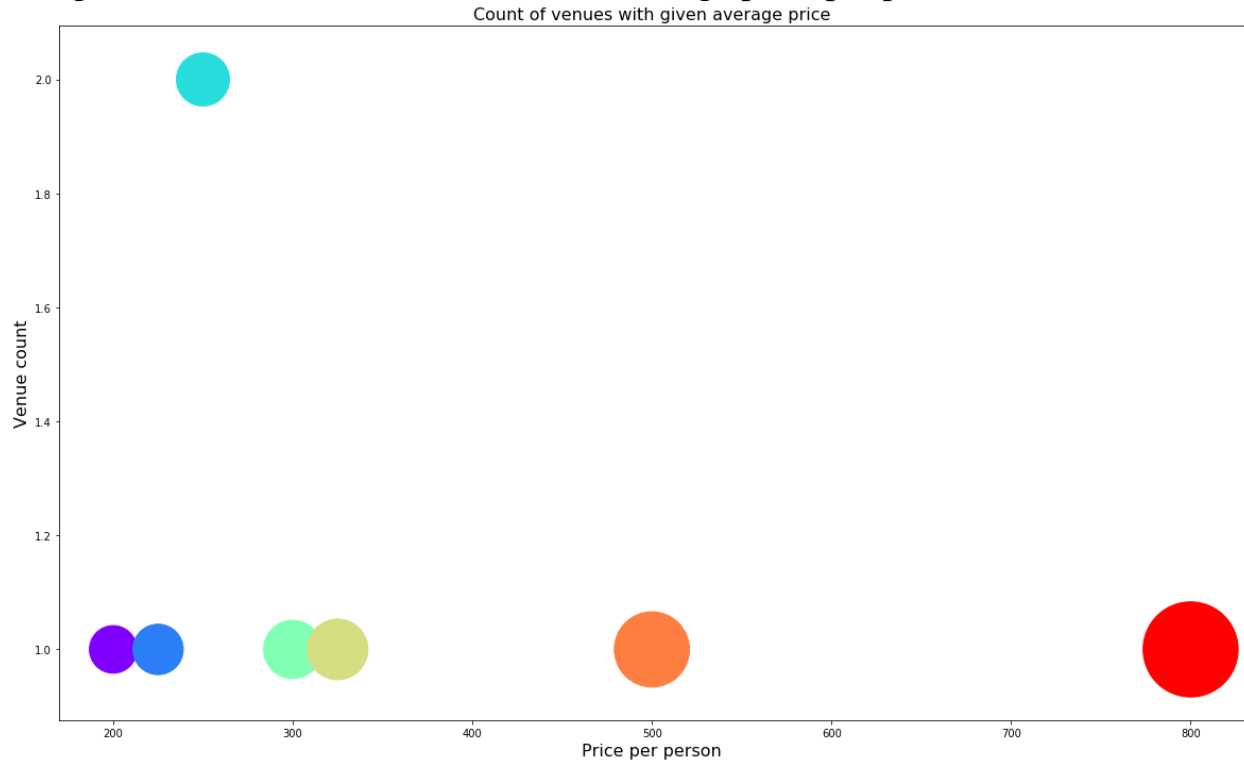

Figure 7: Price per person with count of venues with that price

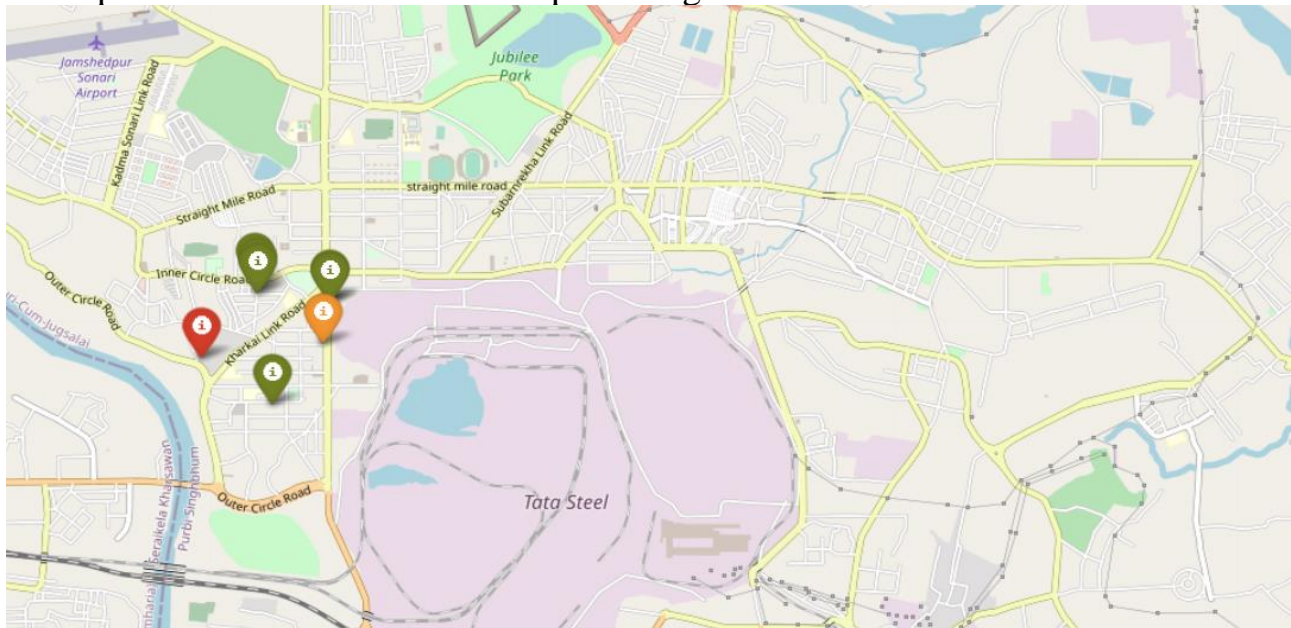I also plot the venues based on their price range.

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.

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.
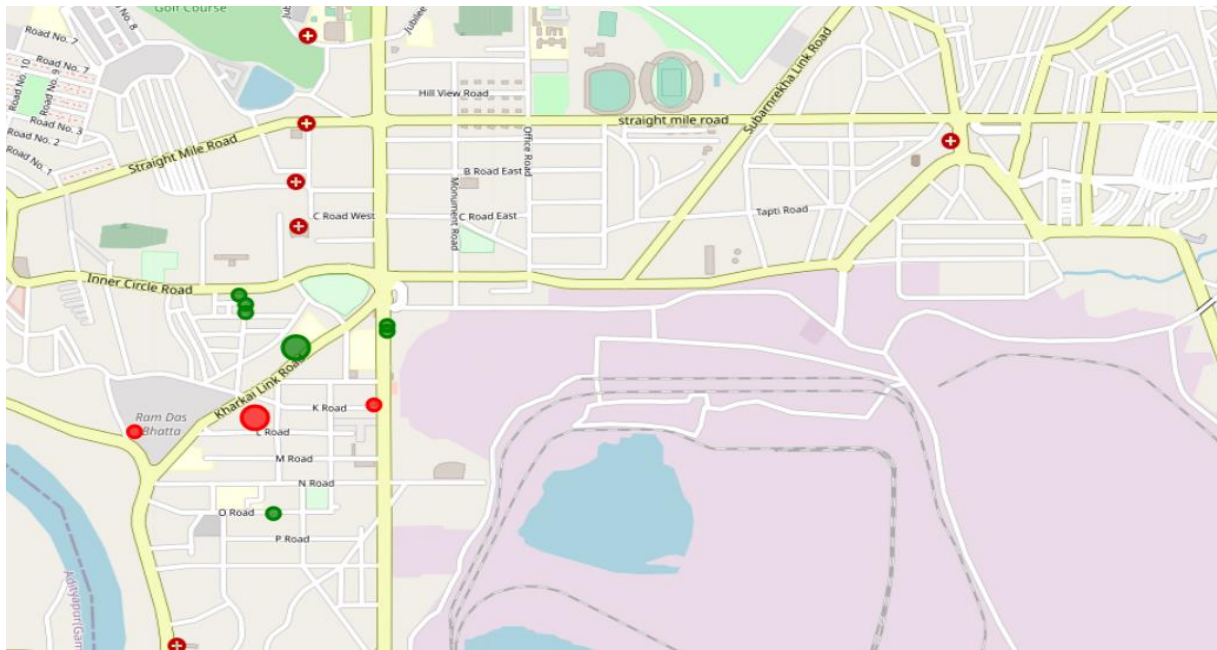

Figure 9: Clusters of venues

In figure 9, we see the two clusters:

1. The first cluster (green) is includes the majority venues. These venues have mean price range of 2.00 and rating spread around 4.07.
2. The second cluster (red) is very sparsely spread and has very limited venues. These venues have mean price range of 3.50 and rating spread around 3.80.

4. Results and Discussion

After collecting data from the Foursquare and Zomato APIs, we got a list of 14 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 their names to combine them and remove all the outliers. This resulted in a total venue count of 7.

We identified that from the total set of venues, majority of them were Hotels and Cafes and Markets, etc. A visitor who loves Hotels/Cafes would surely benefit from coming to Jamshedpur.

While the ratings range from 1 to 5, 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 Bistupur.

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 200 to Rs 400 for one person. However, the variation in prices is very large, given the complete range starts from Rs 200 and goes up till Rs 800.

Finally, through clusters we identified that there are many venues which are relatively lower priced that have an average rating of 4.07. On the other hand, there are few venues which are high priced and have average rating of 3.80. Thus, there are a number of places with low price but good rating which are pocket friendly.

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 Jamshedpur could explore. The venues have been identified using Foursquare and Zomato API and have been plotted on the map. The map reveals that most the venues are located around Bistupur. Based on the visitor's venue rating and price preferences, he/she can choose amongst the three places.