

Applied Data Science Capstone project:

Restaurant Battle of Neighborhoods

By

Shweta Avinash Ijardar

Table of contents :

1.Introduction/Business Problem

2.Data Collection from APIs

3.Data Cleaning

4.Methodology

5.Analysis

6.Results

7.Conclusion

Exploring venues in Pune, India using Foursquare and Zomato API

1. Introduction/Business Problem:

Pune has been ranked as "the most livable city in India" several times. It is second largest city in the Indian State of Maharashtra. It is also known as cultural capital of Maharashtra. Pune boasts several historical monuments and museums and hence attracts lots of tourists. Because of the best education system in Pune, it is also known as Oxford of the east and have students across the world here. Whenever a person searches for a venue in a new city, they're highly interested in the best and affordable places that the city has to offer. The person might want to know how good a given restaurant is with food options provided

Pune is lovely city of 331.3 km² area with many venues with variety of foods. The aim of project is to provide information to visitor about places with best restaurants with help of rating, price per person criteria which would surely help visitors in a city make better informed decisions about the places they should visit. For tourists, finding the right place to eat can be a challenge. But we can recommend visitors a good overview about what to eat where.

Target audience:

Tourist as well as students basically the once who are visiting Pune city. Any company or Entrepreneurs interested in investing in the business of Food can use this information like maps, plots from project to choose best place according to rating and price 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 from APIs

Description of the data:

For assignment we are asked to use foursquare data. Here to collect data of restaurants in Pune we are using two APIS one is Foursquare and other is Zomato data and will combine data together. We will provide latitude longitude directly.

Using the Foursquare's explore API which gives venues recommendations, I fetched venues up to a range of 4 kilometers from the center of Pune 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.

Retrieving data-

From Foursquare API (<https://api.foursquare.com/v2/venues/>), 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 Cleaning

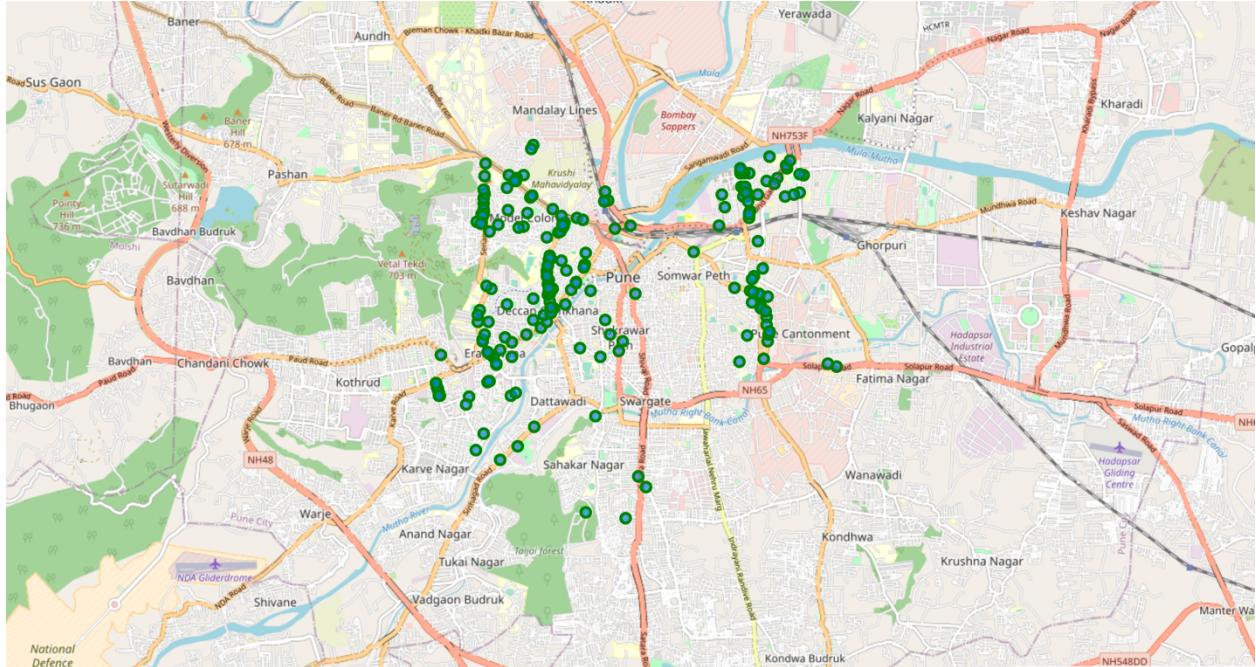


Figure 1: Venues of Pune retrieved from Foursquare API

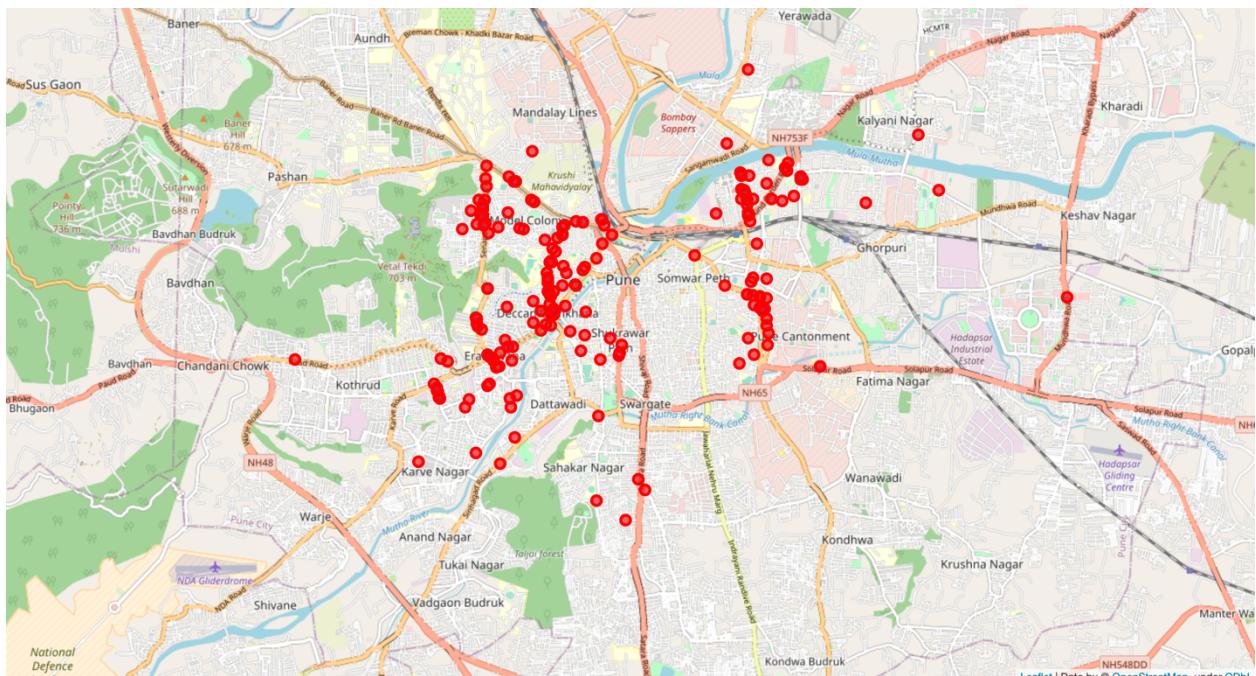


Figure 2: Venue retrieved from Zomato API

1. To combine the two datasets, I had to check that the latitude and longitude values of each corresponding venue match. After analysis, I decided to drop all corresponding venues from the two datasets that had their latitude and longitude values different by less 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 like There are some venues that have specific restaurants/cafes inside them as provided by Zomato API (Restaurant in shopping Mall).
2. Two locations are so close by that they have practically same latitude and longitude values.
3. Some venues have been replaced with new venues

We need to clean this data so that we should not use redundant, data with ill information. We retrieved venues within 4km from center of Pune in each direction which generated total venues as 212 and after careful inspection and removal of data total number of venues is 109 rows and 8 columns.

4. Methodology:

We will do the data analysis and how this data is used for retrieving the results. This project aims is identifying the venues in Pune based restaurant on their rating and average costs. This would enable any visitor to identify the venues they want to visit based on their interest.

We identify the top category types of venues using dataset. We identify there are lot of places where many venues are located so that any visitor can go to one place and enjoy the option to choose amongst many venue options. We also explore areas that are high rated and those that are low rated while also plotting the map of high- and low-priced venues. Lastly, we cluster the venues based on the available information of each venue. 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. Lastly, I'll also cluster the venues and see if we can draw meaningful information out of what kind of venues exist in Pune.

5. Analysis:

During the analysis phase, I explored the venue categories, the rating distribution of the venues and the price range across the map of Pune.

Category: As we extracted categories from the Foursquare API, identifying what type of venues are most popular in the city would really be helpful. We plot a bar chart for the same

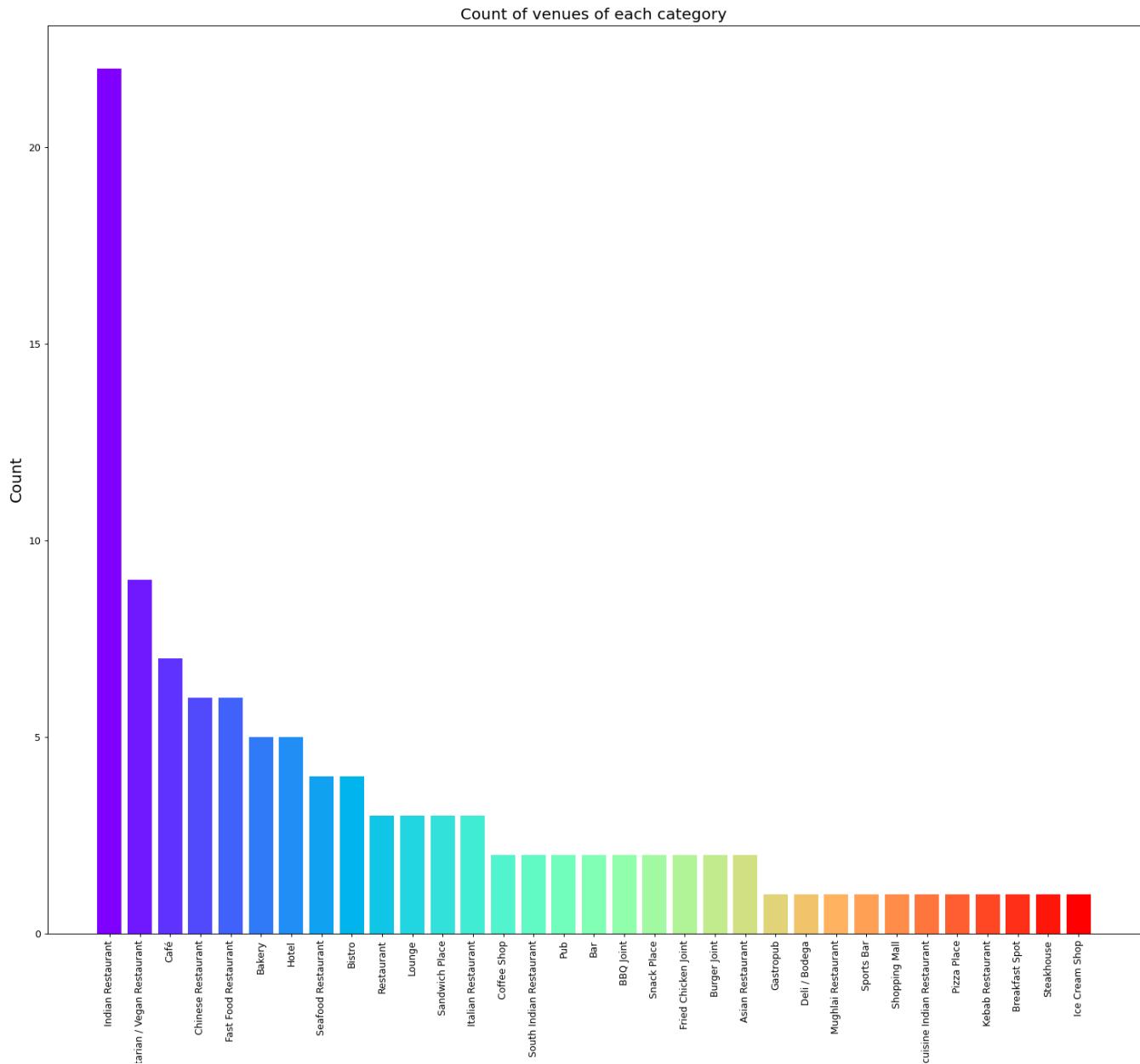


Figure 3: Count of venues of each category

From bar chart, we can see the majority venues are Indian Restaurant and vegan/vegetarian restaurants.

So, if you are tourist and want to try authentic food culture of Pune you can try these.

Rating vs count of venue

Here we will show bar diagram for ratings of the venues. As a visitor, you'd like to know the places that have good-rated venues. We can plot a bar chart of the ratings of all venues and the count of each rating to see what the average rating across all venues is.

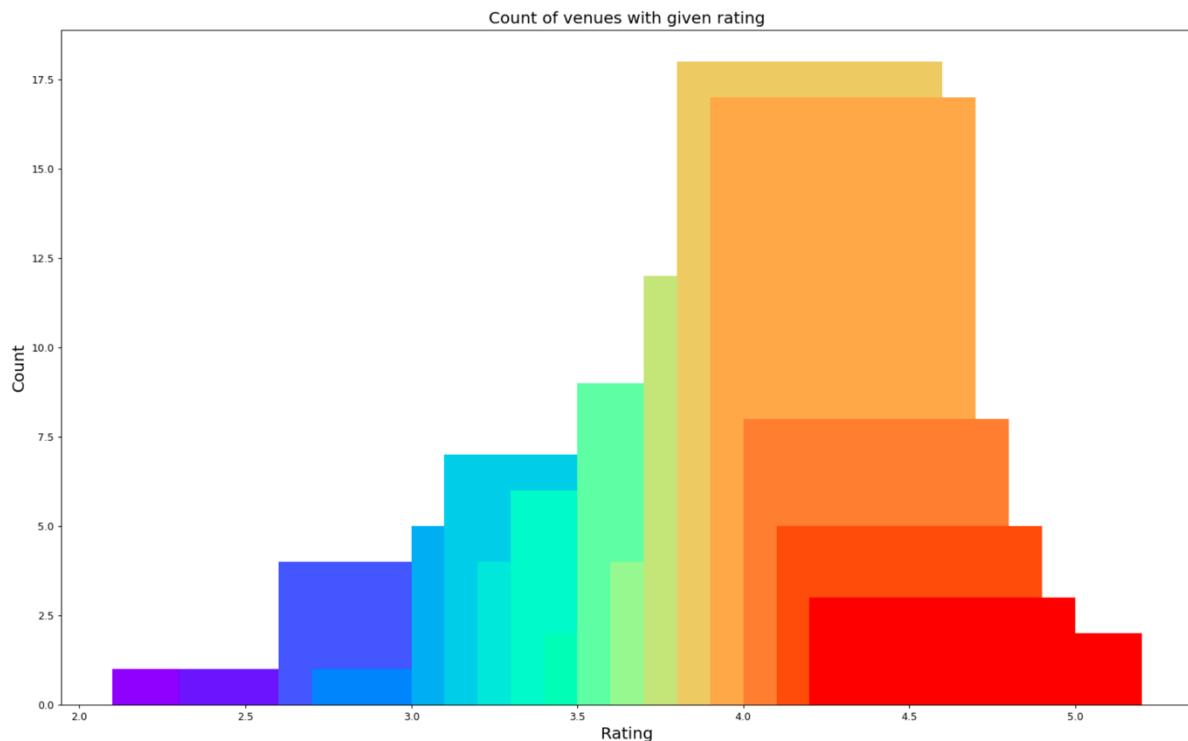


Figure 4: Rating and count of venues with that rating

From the plot above, it is clear that majority venues have their rating close to 4. We see that the ratings range from 1.0 to 5.0

Now we will check price vs venue map:

Plot venues on map-

We will first plot venues on map. The map has the location of all the venues. It appears that many venues are located near Deccan, Bund garden camp area with rating above 3. If someone wants to explore new venues, they should definitely check out here.

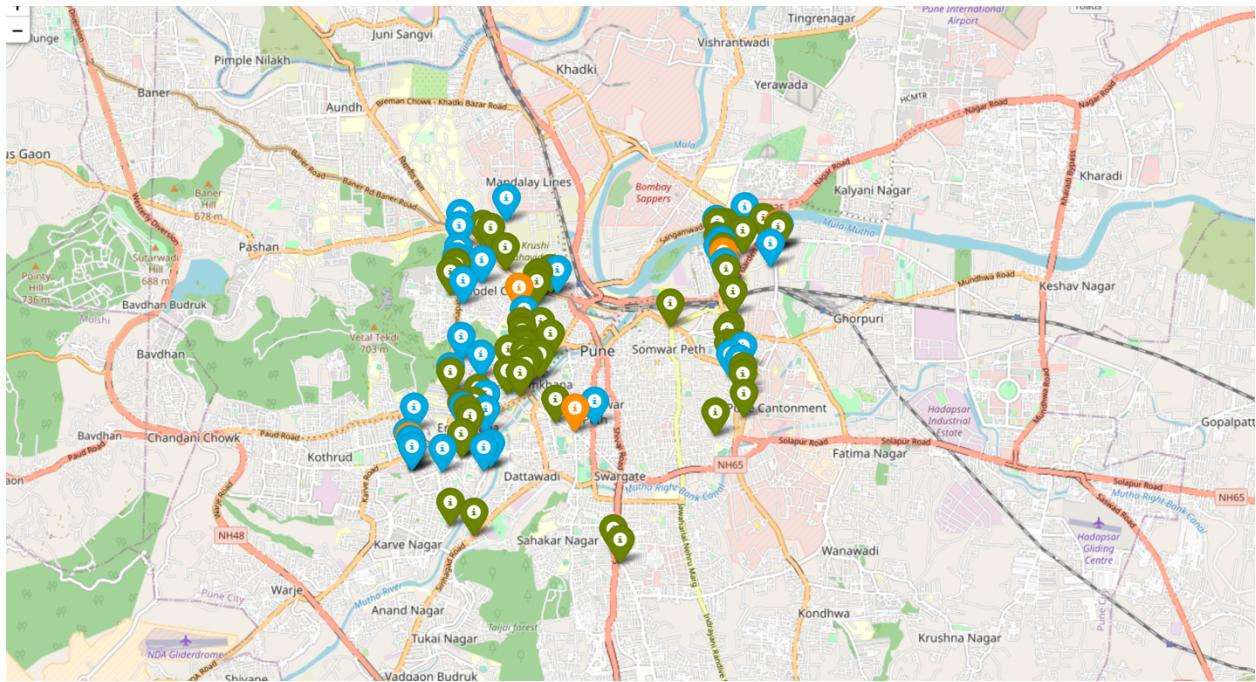


Figure 5: Plot of venues with different rating

Price Vs Venue -

We can plot the price per person. Price range is defined by Zomato and other way is to calculate average cost per person.

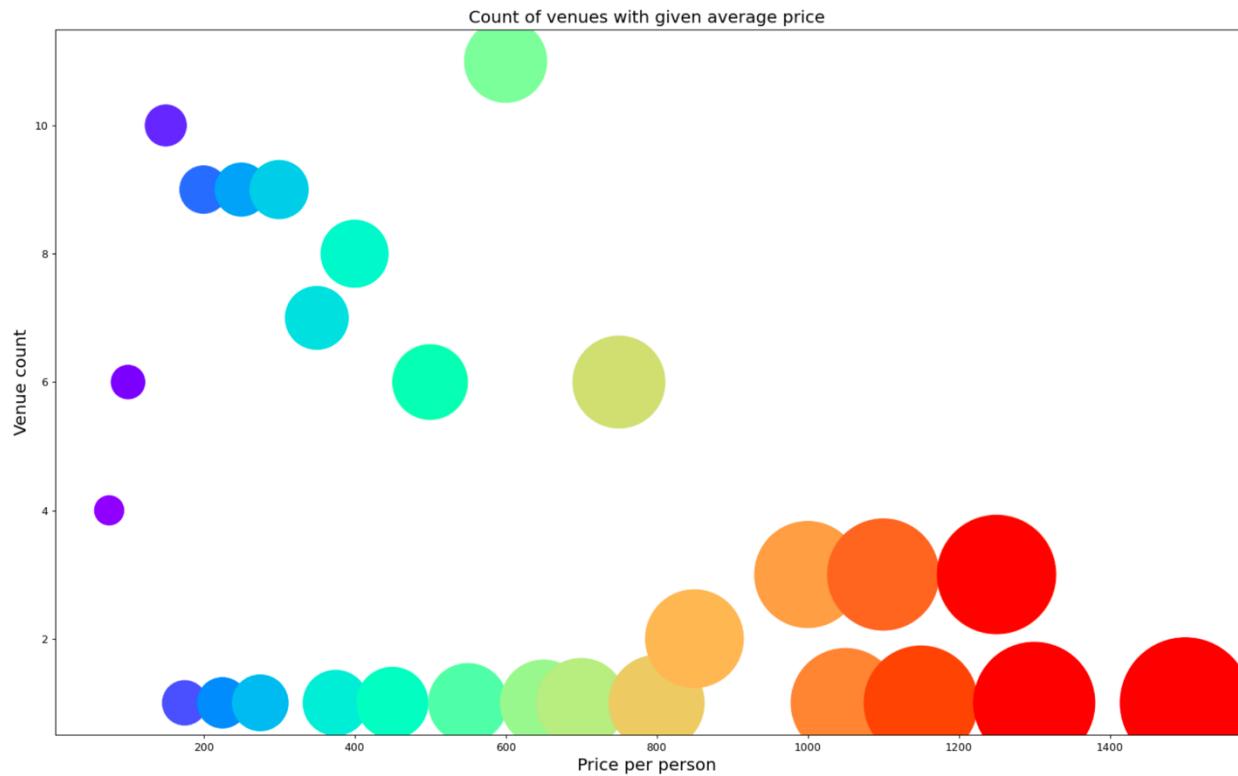


Figure 6: Average price range of venues

From map we will get to know that per person average price range of large number of venues have an average price between Rs 200 and Rs 400.

Zomato defined the price range and we will plot the price for venue on map. Here we will use color scheme like venues with lower price we will use green and move towards red as the price increases-

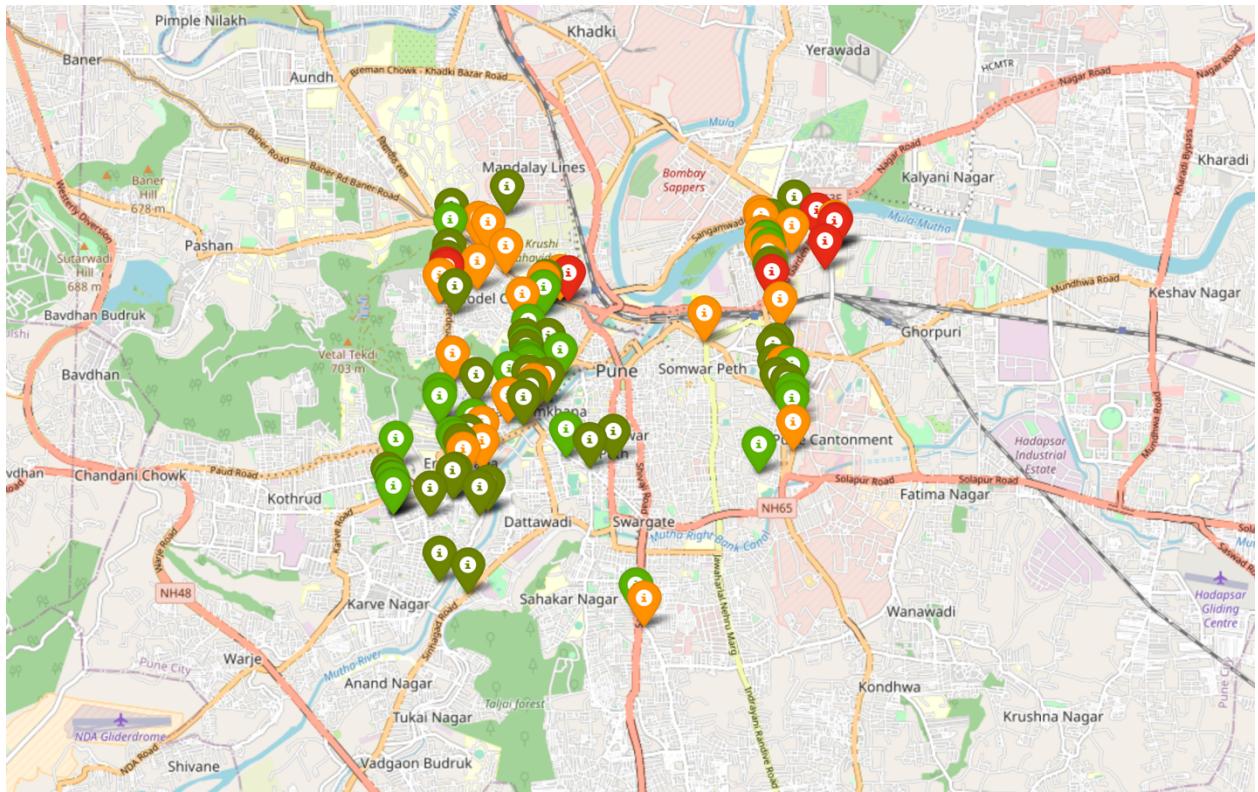


Figure 7: Price range Vs venue

From we come to know Bund Garden camp area of Pune is very expensive while venues near Erandwane area is much less price range. Deccan area venues are mix of both low as well little expensive price.

Clustering

We will now cluster all these venues based on their price range, location and more to identify similar venues and the relationship amongst them. We'll cluster the venues into two separate groups.

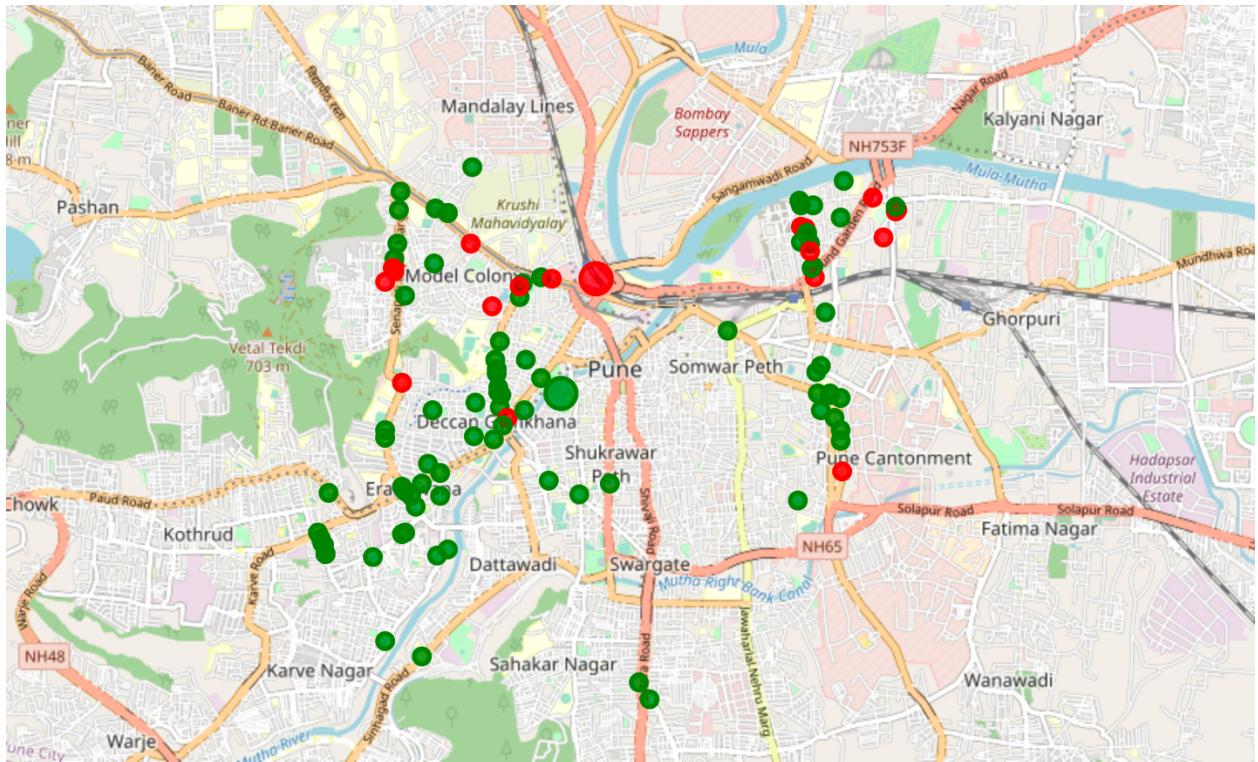


Figure 8: Clusters of venue

The first cluster (green) is spread across the whole city and includes the majority venues. These venues for cluster 0 have mean price range of 1.86 and rating spread around 3.97. The second cluster (red) is very sparsely spread and has very limited venues. These venues for cluster 1 have mean price range of 3.57 and rating spread around 4.12

6. Result:

Based on our analysis above, we can draw a number of conclusions that will be useful to aid any visitor visiting the city of Pune, India. After collecting data from the Foursquare and Zomato APIs, we got a list of 212 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 109. We identified that from the total set of venues, majority of them were Indian restaurants and vegetarian/vegan restaurants.

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. Finally, through clusters we identified that there are many venues which are relatively lower priced but have an average rating of 3.97. On the other hand, there are few venues which are high priced and have average rating of around 4.12.

A visitor can find affordable restaurant of his/her interest with help of rating, price as we found range variation in prices is very large, given the complete range starts from Rs 100 and goes until Rs 1200 but we can find venues providing good quality food in reasonable price as well. 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).

7. Conclusion:

The purpose of this project was to explore the places that a person visiting to Pune. The venues have been identified using Foursquare and Zomato API and have been plotted on the map. The map reveals that there are so many places where one can visit during his/her stay in Pune. Based on the visitor's venue rating and price requirements, he/she can choose the places.