Insights of the venues in Hyderabad, India Using Foursquare and Zomato API

Y. Kushal Vidya Mohanji

1. Introduction

- **Initial Insight**

  India, the land of tourism and various culinary food from all around the world and the land of unique flavors. Tourism is a gamechanger for Indian food market. Millions of people from all around the world visit India to explore the monuments and as well as food. So, the tourists would like to know the relevance of the choosing which venue to decide among all the restaurants in the city. Using the combination of the relevant price and rating of the place, it makes the decisions easier for the people who visit.

  Hyderabad is a vast region composed of many sectors and having an area of 650 sq.km. Hyderabad is a place where there exist various venues including all the restaurants, hotels, cafes', Pubs, Clubs which have a wide range to explore into. The insight of the project is to explore various venues in Hyderabad and the attributes of the Data based on the average price and rating such that the decision to select the venue becomes easier. To explore a place in India, not only using the Four Square API is required but also the local Restaurant Partner Zomato, its API will be used to fetch the complete information and the ratings of the users which helps in making the task relevant and easier.

- **Target Audience**

  There exist two target audience possibilities, the initial one would be the people who live in that area such that, it gives them the area and new insights which they want to see. People who would like to concentrate on the area where the competition is less and establish their food market in that area.

2. Data

- **Data Sources**

  To retrieve the location and other information from various venues. Two APIs will be used, and the data retrieved will be combined.

  Using the Foursquare's API, the venues are fetched up to a range of 5 kilometers from the heart of the city and collected their names, categories, latitude and longitude.

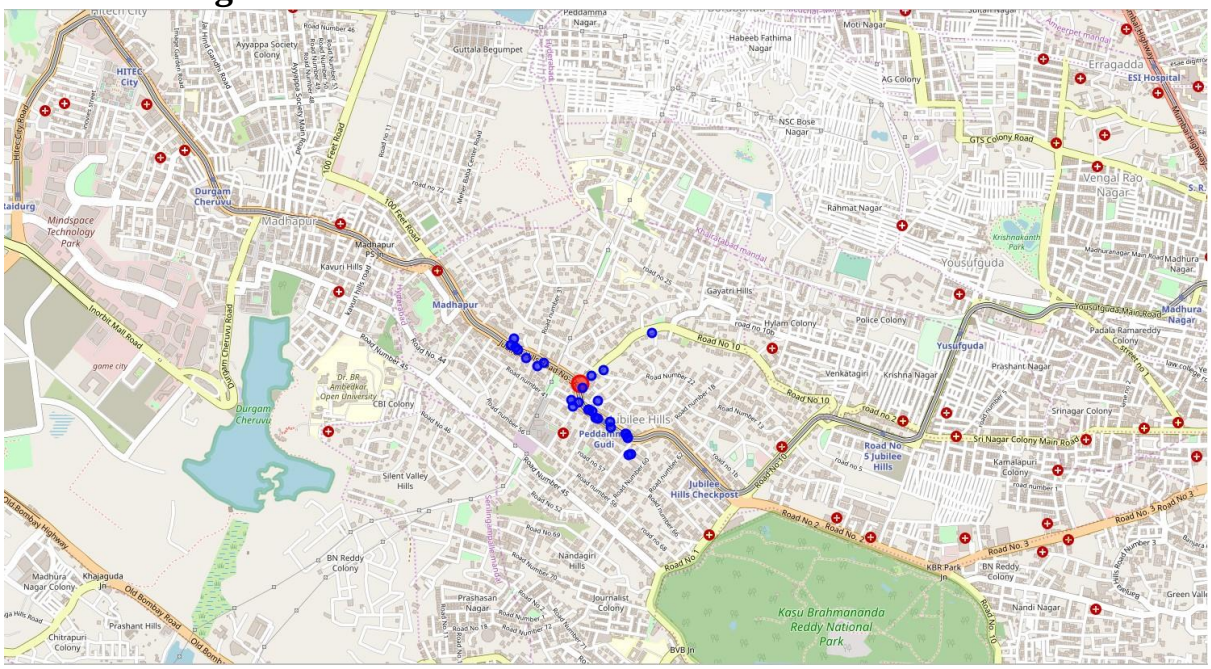
From Foursquare API the following for each venue have been retrieved:

- **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.

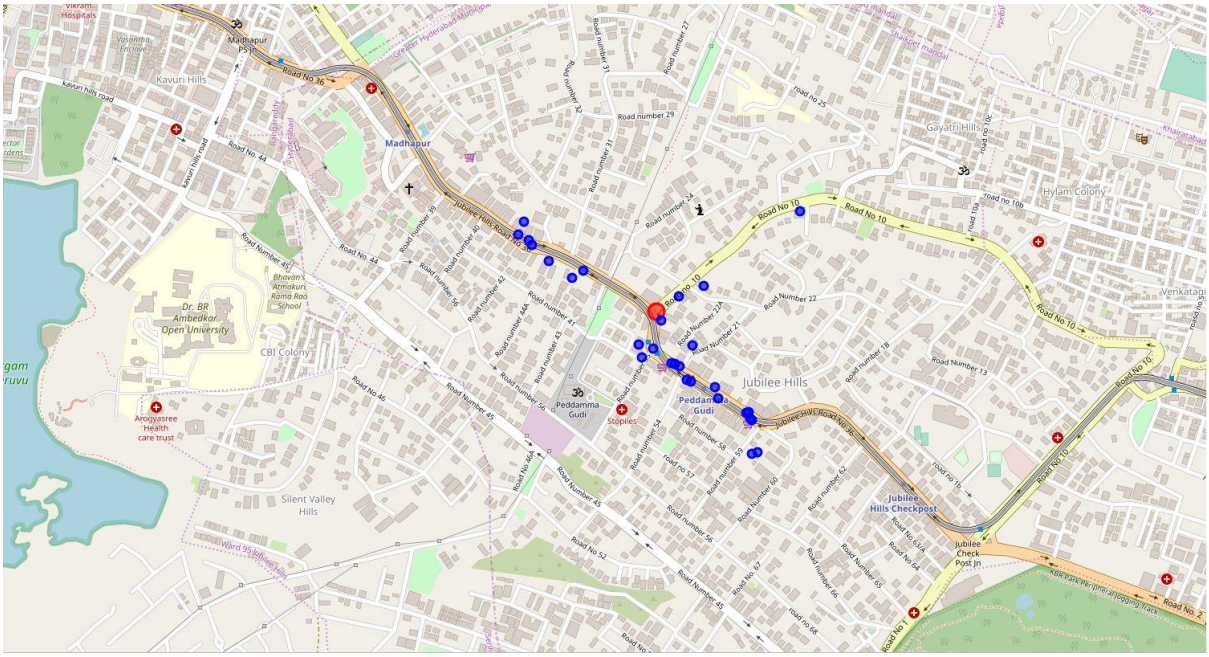
From Zomato API (https://developers.zomato.com/api), I retrieved the following for each venue:

- **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

- ## Data Cleaning



FourSquare API Venue Map



Zomato API Venue Map

From figure 1 and figure 2, we can clearly see that some venues from the two APIs do not align with each other. There exists an offset between the two values.

There exists an offset difference of more than 0.0006 from one another. This has been updated and combined with the two data sets.

They can be categorized 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.

The details of the venue from category 1 and 3 can be added to the data but data from category 2 will be dropped and the final dataset will be constructed based on which the further project will be done.

3. Methodology and Exploratory Data Analysis

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

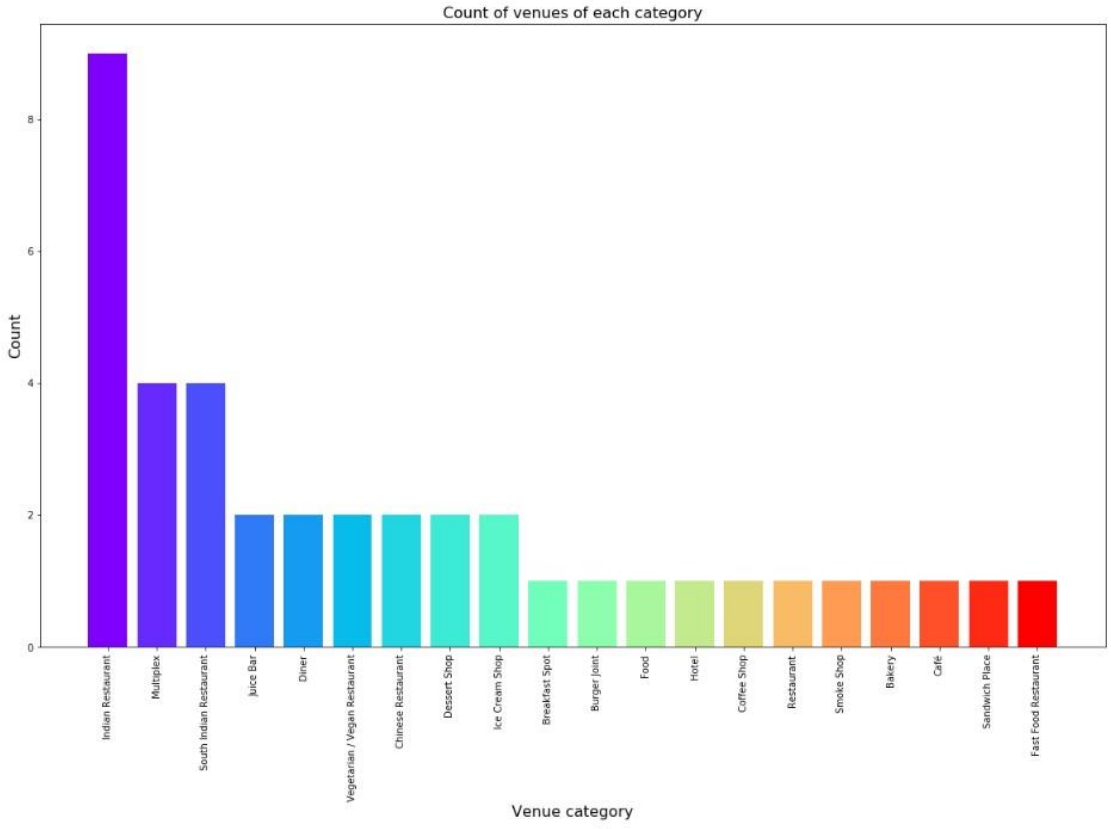
Using data cleaning, the dataset from the two APIs will be combined based on the venue names, latitude, and longitude values. One to one matching and careful data inspection would be used to remove any remaining outliers such as multiple venues at the same location from the two datasets. The final data 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 Hyderabad. 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 Hyderabad.

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.

- **Categories**

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



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

- **Rating**

Next, I'll explore the ratings of various venues in Hyderabad. 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 Hyderabad. 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 Hyderabad. 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. Looking 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 Himayat Nagar, and Abids.

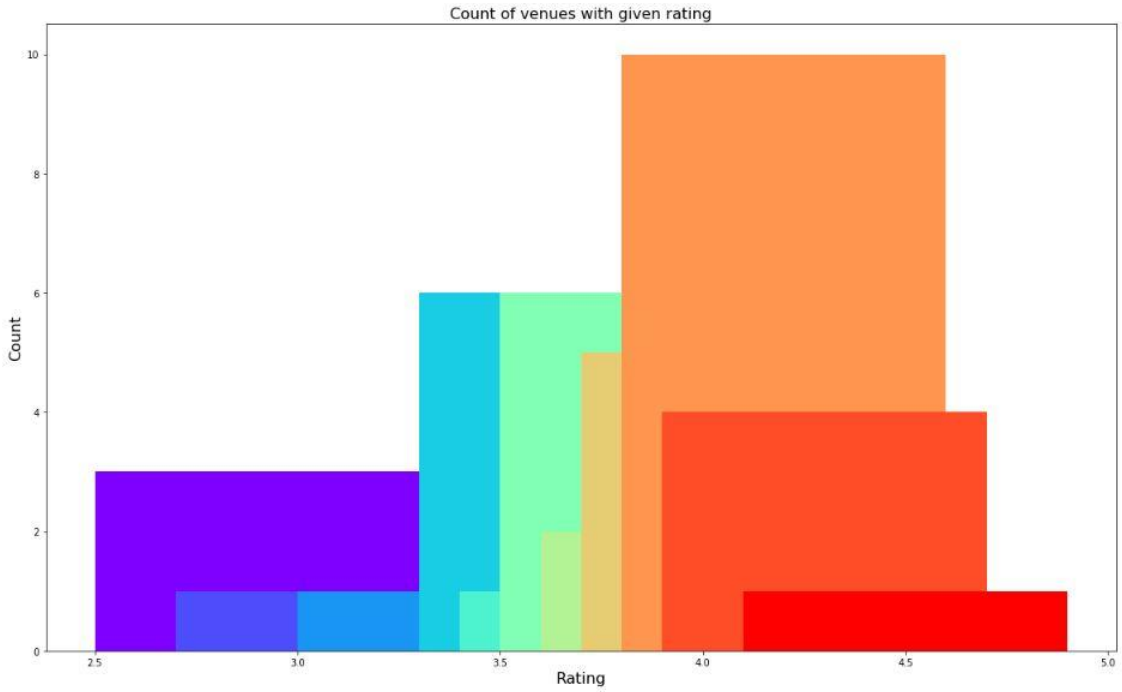


Figure 5: Rating and count of venues with that rating

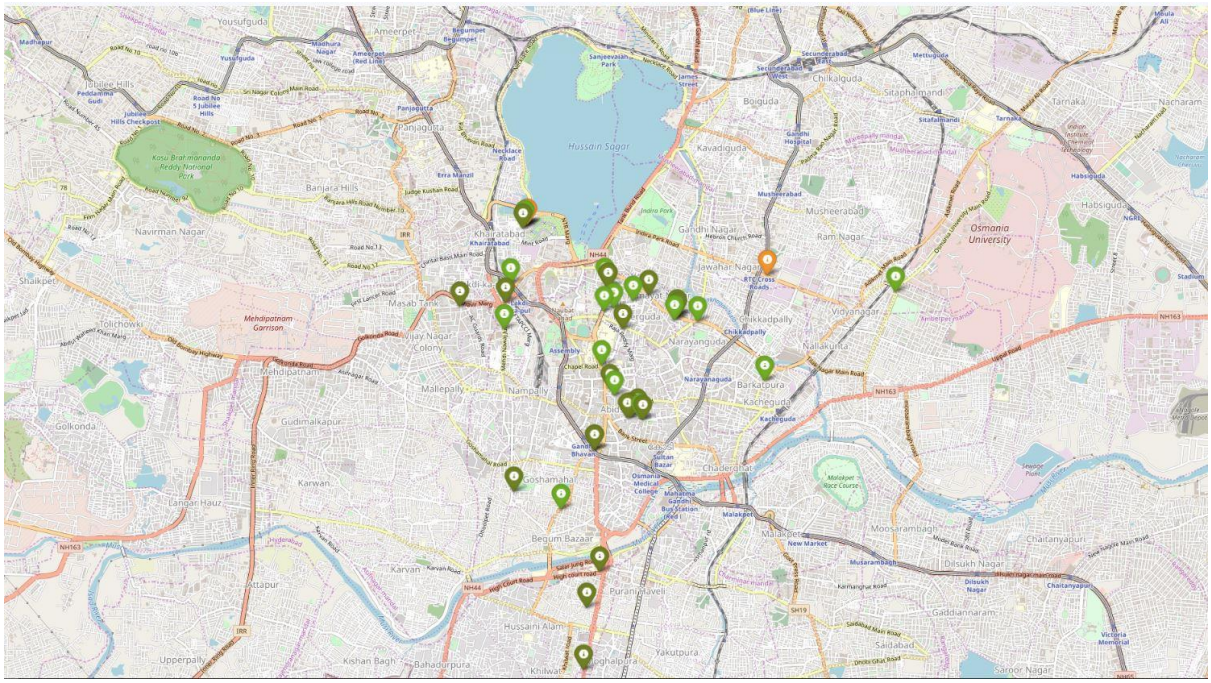


Figure 6: Plot of venues with different ratings

- **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. Looking 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 is not even as high as Rs 800 for one person.

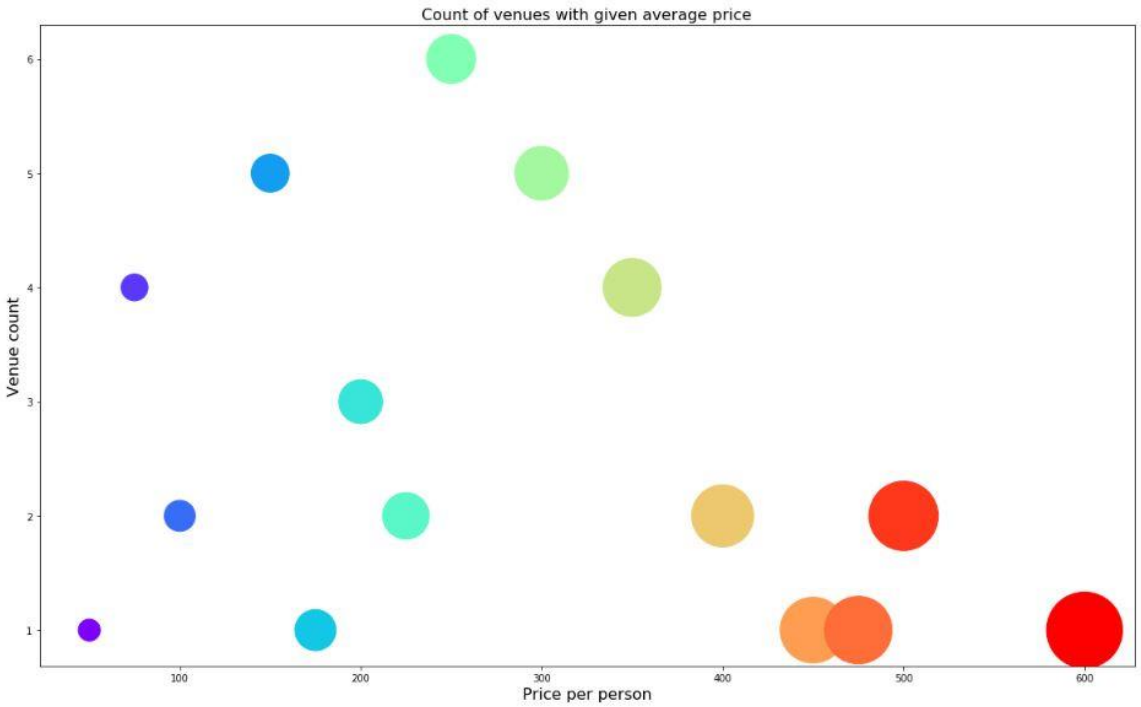


Figure 7: Price per person with count of venues with that price

I also plotted the venues based on their price range.

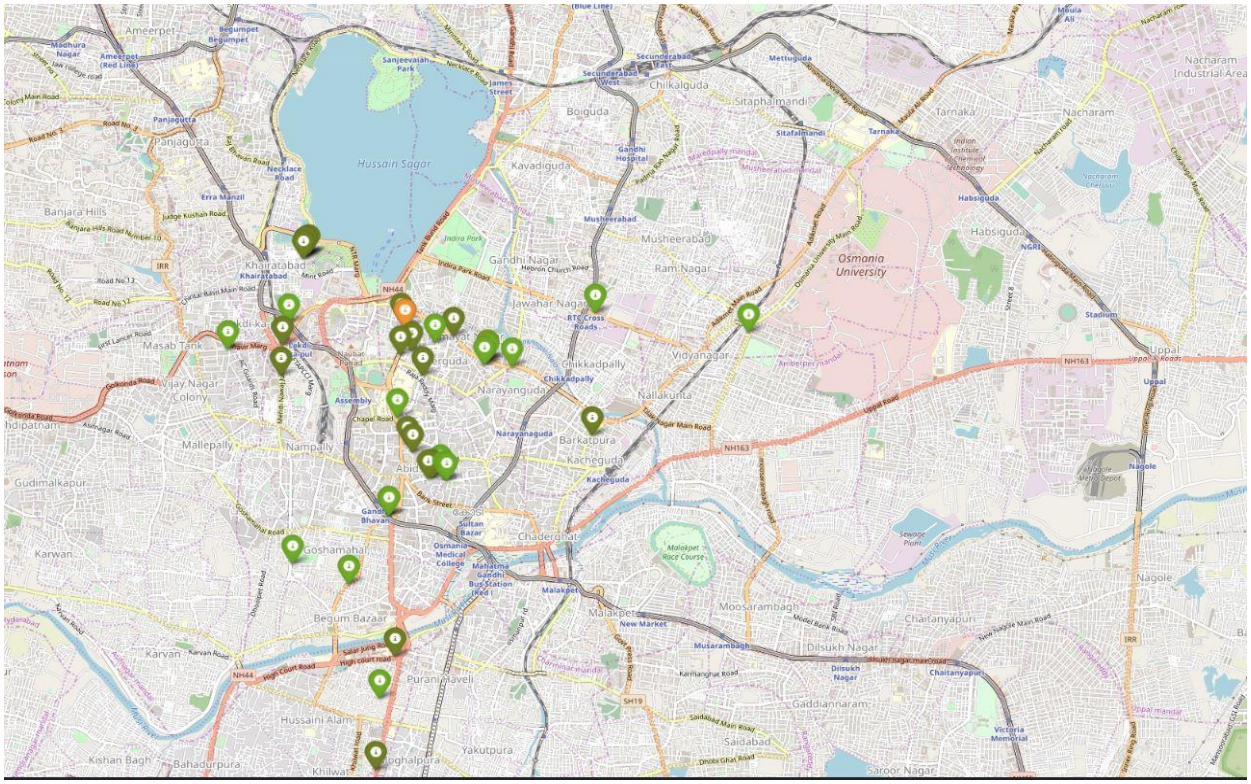


Figure 8: Plot of venues with different prices

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 Abids and Khairtabad are primarily lower priced. The venue near Himayat Nagar have steep prices.

- **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.

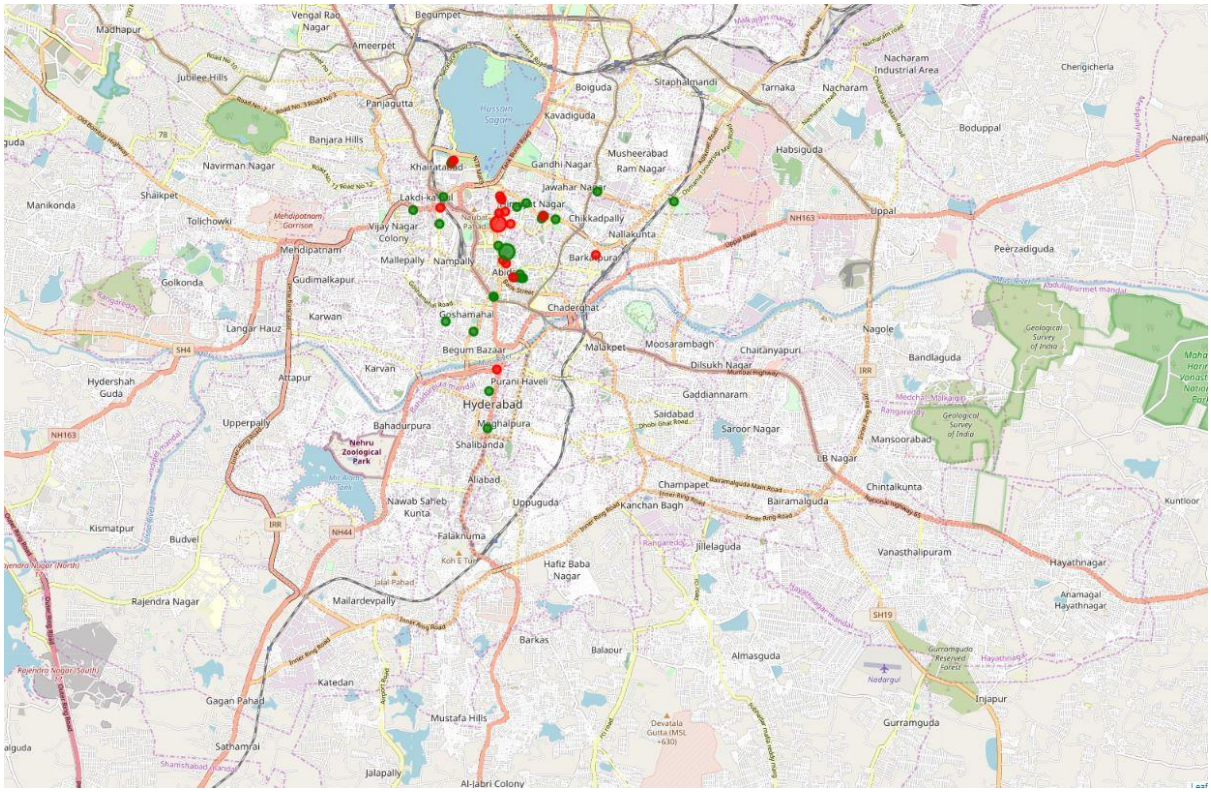


Figure 9: Clusters of venues

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

After collecting data from the **Foursquare** and **Zomato APIs**, we got a list of 107 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 40.

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

While the complete range of ratings range from 1 to 5, the **majority venues have ratings close to 3.8**. 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 **Himayat Nagar and Abids**. 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 200 to Rs 400 for one person**. However, the variation in

prices is not very large, given the complete range starts from Rs 100 and goes up to Rs 600. On plotting the venues based on their price range on the map, we discovered that venues located near **Abids and Khairtabad are relatively priced lower than venues in Himayat Nagar.**
Finally, through clusters we identified that there are **many venues** which are **relatively lower priced but have an average rating of 3.87.** On the other hand, there are few venues which are **high priced and have average rating of 3.95.**

- If you're looking for cheap places with relatively high rating, you should check **Khairtabad.**
- If you're looking for the best places, with the highest rating but might also carry a high price tag, you should visit **Himayat 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 **Abids.**

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).

5. Conclusion

The purpose of this project was to explore the places that a person visiting Hyderabad 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: Khairtabad, Himayat Nagar, Abids. Based on the visitor's venue rating and price requirements, he/she can choose amongst the three places.