

Applied Data Science Capstone Assignment Report

Exploring Venues in Bhopal, India using

FourSquare and Zomato APIs

Shivam Mishra

Dec 30, 2019

1. Introduction

1.1 Background

When a person comes to a city, he/she is interested in good places the city has to offer. The person might want to know about the restaurants- category, price and rating. This could in turn decide which venues the visitor must go to in the city. Combining the location of venues in the city with their price and rating information can surely help visitors to make a better and informed decision.

Bhopal is a city located in Central India. It is close to UNESCO World Heritage Sites of Bhimbetka and Stupas of Sanchi. Thus a number of visitors come to the city. Apart from these, visitors can explore number of restaurants, cafes in the city. This project explores various venues in Bhopal and filters the data based on user ratings and prices. To explore this information, we shall make use of FourSquare and Zomato APIs to fetch complete information of various venues (including name, address, category, rating and price for two people). We shall plot a map of the venues with color attributes to highlight their location and other useful information. This will enable a visitor to have a glance at the map and decide where to visit.

1.2 Interested Audience

The target audience for this project can be the visitors looking for venues in the city. They can make use of plots and maps from this project to select places that can suit their category, budget and rating preferences. Also as a add on, companies can build applications using this information and provide users/visitors with further information.

2. Data acquisition and cleaning

2.1 Data sources

In order to get location and venues data for Bhopal city, APIs provided by FourSquare and Zomato have been made use of. Further data from the sources has been combined for analysis and clustering in later stages.

Using FourSquare API, I gathered venues in the city using a radius of 10 kilometers. The query gave a list of venue names, categories, latitude and longitude. The total number of such venues obtained was 47.

Similarly using the Zomato API, I gathered data using the name, latitude and longitude of venues obtained through the FourSquare API(47 in number). The Zomato API gave a list of venue names, address, rating, price range, price for two, latitude and longitude information.

On observing the data from the two sources and mapping them, it was found that there were some mismatches found in venues. Consequently, data cleaning was performed to get a filtered and improved dataset.

2.2 Data cleaning

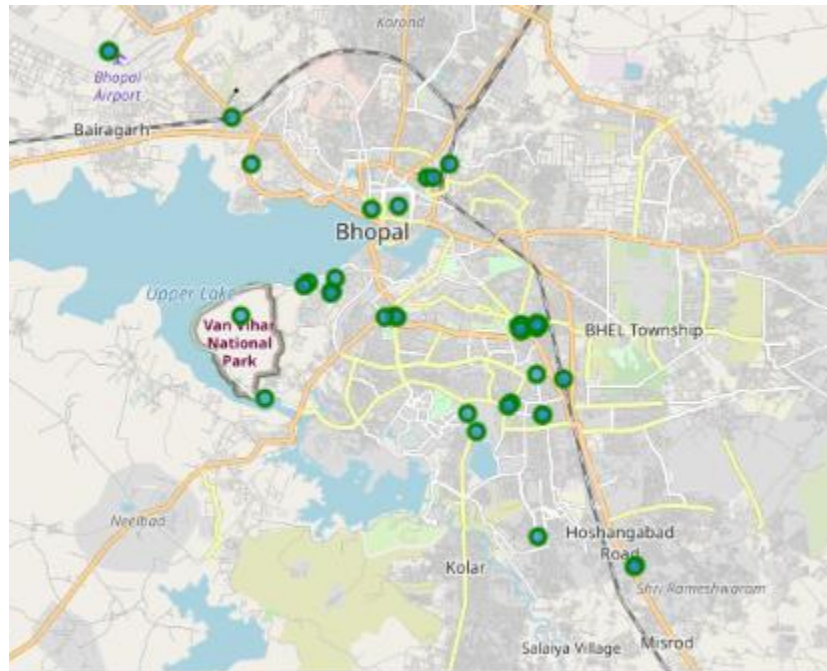


Figure 1: Venues retrieved from FourSquare API

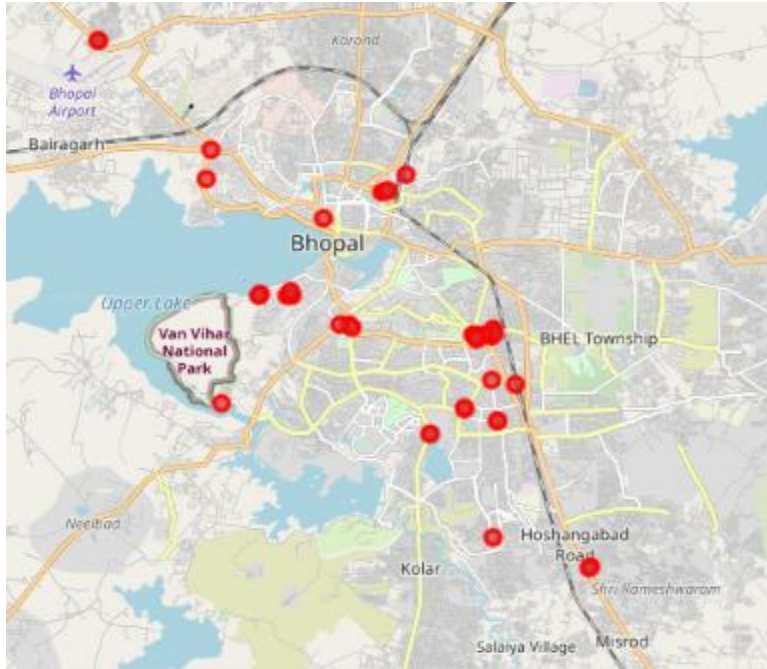


Figure 2: Venues retrieved from Zomato API

From figure 1 and 2, we can clearly see that some venues from the two APIs do not align with each other(notice the green marker in the Van Vihar National Park and red, green markers near the airport). Therefore, I decided to combine the two using similar latitude and longitude vales.

In order to combine the two dataframes, I checked for latitude and longitude values for each venue. After rounding off the latitude and longitude values to four decimal places(a reasonable approximation), I decided to keep venues having a difference of less than 0.0005 in their latitude and longitude values. This helped in removing the outliers from the dataframe. This resulted in a dataframe with 26 venues

After doing the latitude and longitude filtering as above, I observed there were still certain rows in the dataframe which had anomalies. Consequently, I dropped such rows after careful inspection for venue category and names.

The final dataframe thus obtained had 21 rows and 8 columns(See figure 3).

	categories	venue	latitude	longitude	price_for_two	price_range	rating	address
0	Indian Restaurant	Manohar Dairy And Restaurant	23.2642	77.4085	300.0	1.0	4.5	6, Hamidia Road, Opposite Alpana Cineplex, Pee...
1	Asian Restaurant	Coriander Leaf	23.2183	77.3718	1500.0	4.0	3.8	Jehan Numa Retreat, Dr. Sallem Ali Road, Near ...
2	Bakery	Amer Bakery Hut (ABH)	23.2171	77.4275	400.0	2.0	4.2	Bittan Market, Arera Colony, Bhopal
3	Asian Restaurant	Milan Sweets And Namkeen	23.2328	77.4336	200.0	1.0	4.1	Plot 189, Zone 1, Maharana Pratap Nagar, Bhopal
4	Department Store	Burger King	23.2329	77.4304	500.0	2.0	3.9	11 & 12, Food Court, 3rd Floor, DB City Mall, ...

Figure 3: Final dataframe after data cleaning

3. Methodology and exploratory data analysis

I have obtained data from the two APIs- FourSquare and Zomato for venues in Bhopal city. The venues data obtained from the FourSqaure API is then fed into the query for Zomato API in order to get detailed information about the venues.

Data cleaning is thereafter performed on the resulting dataframes- one each from FourSquare and Zomato. The dataframes are merged first of all, based on the similarity in latitude and longitude values. Further inspection to remove any outliers leads to a much more filtered and improved dataframe consisting of 21 venues.

Using this filtered dataframe(a snapshot of which is present in figure 3), I begin my exploratory data analysis. The venue types are analysed through maps which give us a better understanding of certain spots in the city where many venues may coexist. Thereafter, I analyse the venues based on their prices and ratings data using various plots that will follow. My aim is to identify places/venues which can be recommended to visitors using the pricing and rating information. Finally, I will cluster the venues using kmeans algorithm to draw out some meaningful information about the types of venues in Bhopal city.

3.1 Categories

I begin my data analysis by having a look at the various categories of venues that exist in the city. A bar graph is plotted to have a look at the categories and the number of such venues.

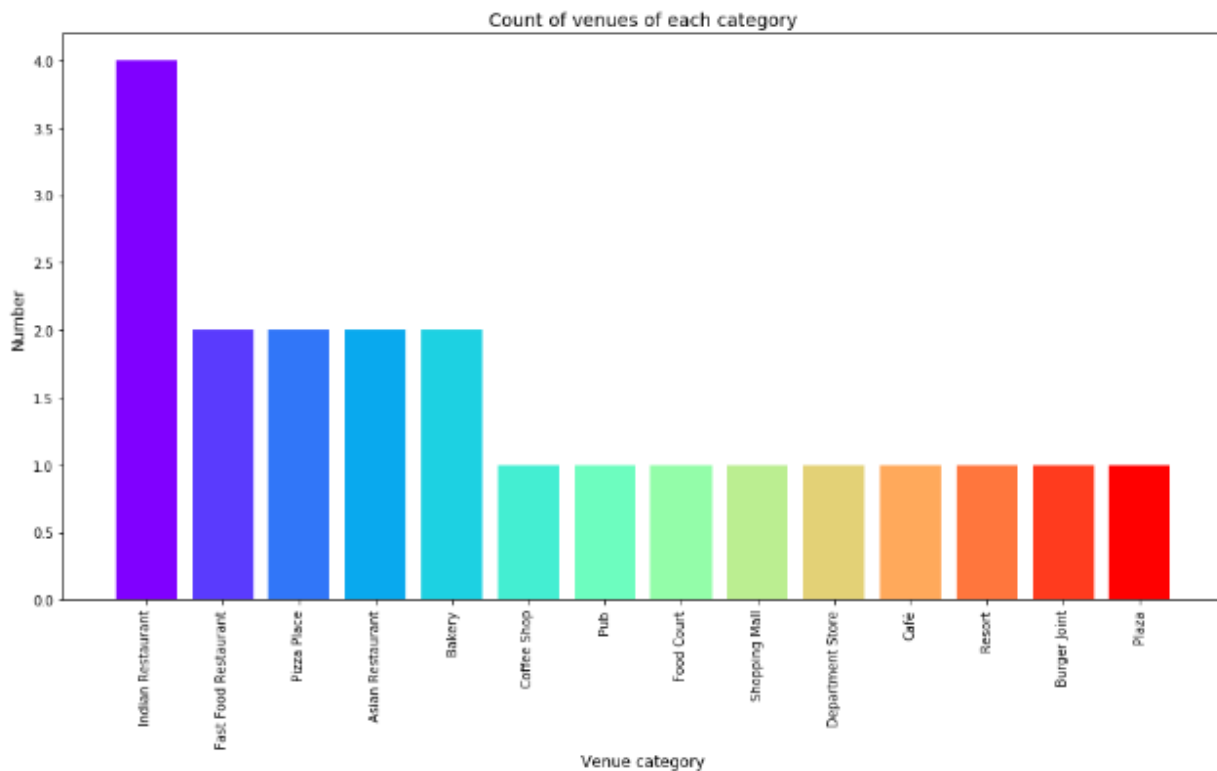


Figure 4: Number of various types of venues in Bhopal

From figure 4, it is clear that the most common type of venues in Bhopal are Indian restaurant, Fast food restaurant, Pizza place, Asian restaurant and Bakery.

3.2 Rating

Next, I will explore the venues in Bhopal based on their rating. I will make a bar graph having venue rating on the x-axis and number of venues on the y-axis. The resulting graph can be seen in figure 5.

The range of ratings for the venues range from 2.5 to 5 in figure 5. The average rating is somewhere between 3.5 and 4. The maximum number of venues also fall into this range bracket.

I follow this graph by a map of Bhopal having the venues. The venues that were rated below 3 were marked by red color, while the venues having ratings between 3 to 4 and those with more than 4 were marked with orange and green respectively. Figure 6 represents this map; it shows that moderately rated venues(orange) are present nearby upper lake and down south in the city while high rated venues are spread out across the city.

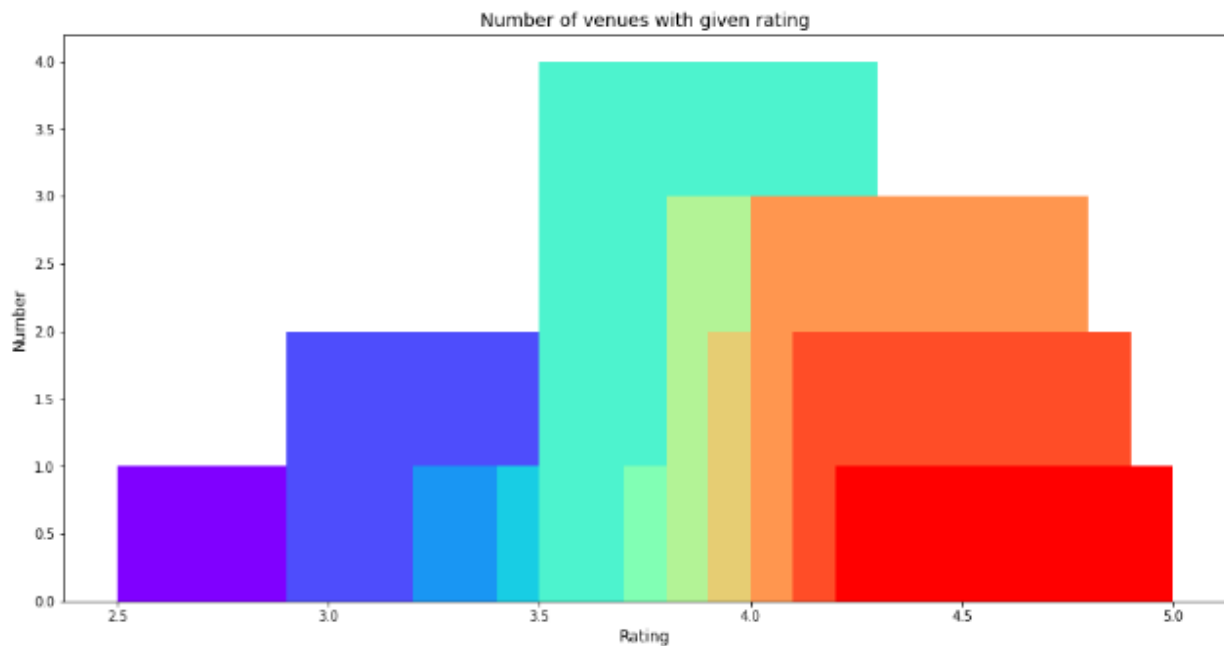


Figure 5: Rating and number of venues with such rating

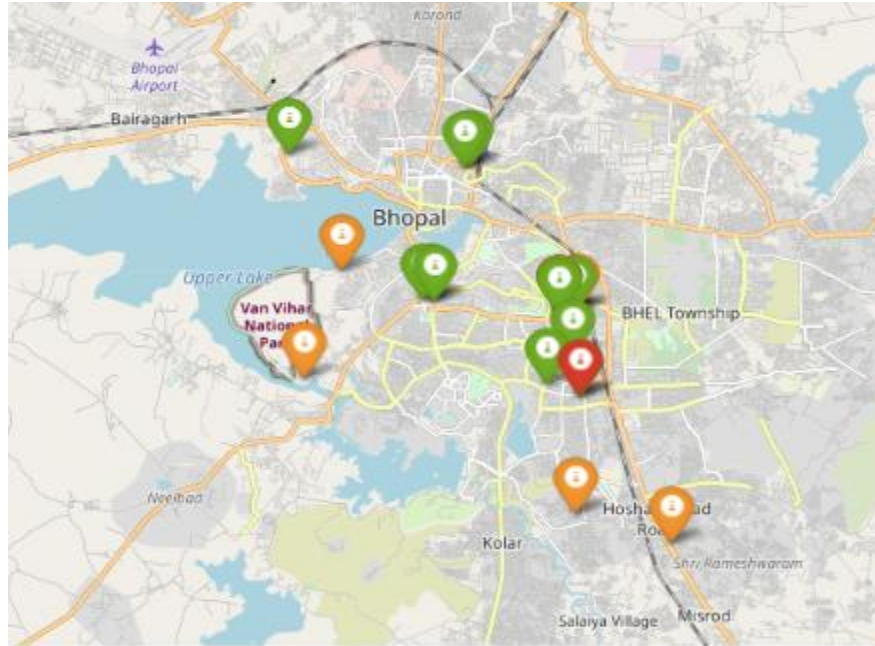


Figure 6: Map of venues with different ratings

It is evident that Bhopal on an average has good ratings for its venues.

3.3 Price

Next, I will explore the prices for two persons of all venues using a scatter plot. The y-axis of this scatter plot will consist of number of venues having that much price for two persons.



Figure 7: Price for two persons with count of venues for that price

Taking a look at figure 7, we can see that majority of venues have prices less than 600 Indian Rupees. We can also note that there are places as expensive as ones that charge more than 1400 Indian Rupees for two persons.

Next, I will plot the venues based on their price range. The lower price range venues in figure 8 are represented by green color markers and higher price range venues with red.

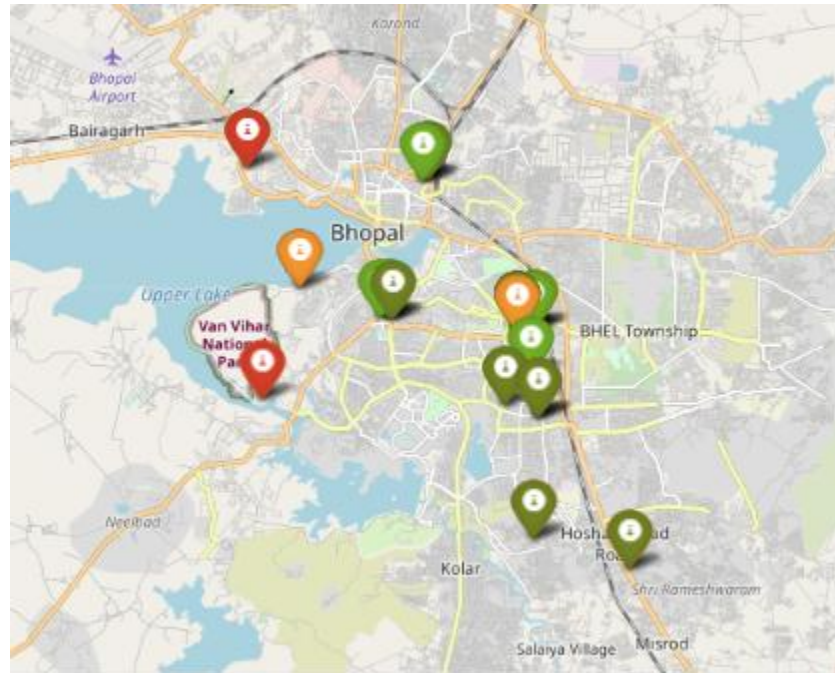


Figure 8: Map of venues in Bhopal with different price ranges

From figure 8 we can observe that venues besides upper lake are in higher price range while those in lower price range are present mostly in south, south-east part of the city.

3.4 Clustering

Finally, I will cluster all the venues using kmeans clustering based on their address, rating to identify similar venues and the relationship amongst them. I decided to cluster the venues into two separate groups or clusters.

In figure 9 below, we see the following clusters:

1. The first cluster(green) is spread throughout the city(north, centre, south) and has a majority of venues.
2. The second cluster(red) is spread around areas just near the upper lake and has a limited number of venues.

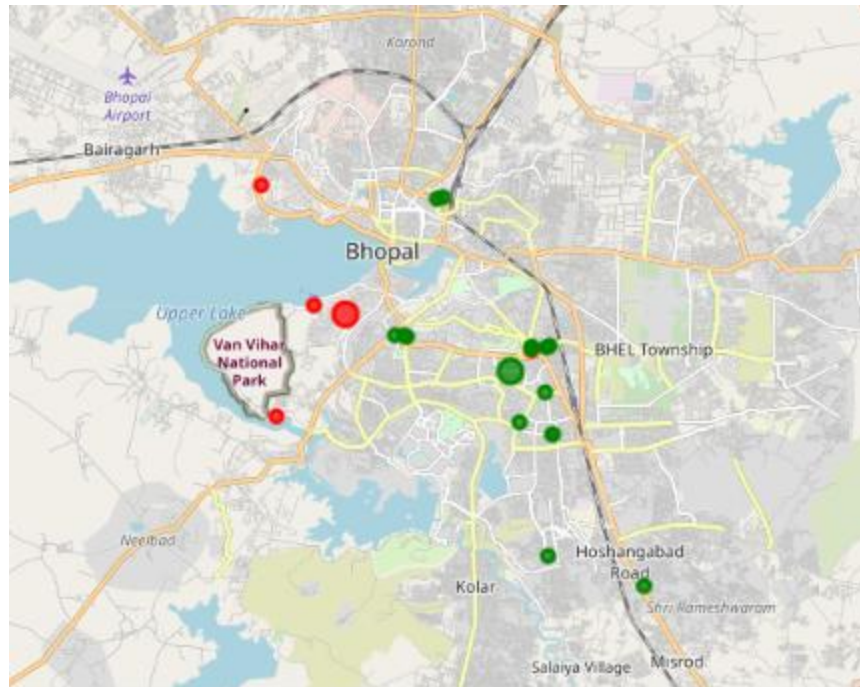


Figure 9: Cluster of venues

4. Results and Discussion

After collecting data from FourSquare and Zomato APIs, we got a list of 47 venues. However not all the 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 dataframe of 21 venues.

We identified from the total number of venues, majority of them were Indian restaurant, Fast food restaurant, Asian restaurant and bakery.

The ratings of the venues ranged from 2.5 from 5 but the majority of them had a rating close to 4. This means that most restaurants provide good quality food that is liked by most people in the city. On plotting these venues on a map, we saw that there are a cluster of venues in and around the upper lake, down south which are moderately rated(3 to 4), while venues in the other cluster have a high rating(above 4).

When we look at the price of these venues, we found that most of the venues offered prices of less than 600 Indian rupees for two people. A map of the price range of these venues showed that venues nearby lake fall in higher price range while other venues of the city fall in lower price range.

Finally through clusters(kmeans clustering) we identified that there are many venues which are relatively low on prices and have an average rating of 4.06. On the other hand, there were a handful of venues which are higher priced and have an average rating of 3.88.

A visitor to Bhopal can use this information to help him decide an outdoor venue for a meal.

5. Conclusion

The project looked at venues in the city of Bhopal, India. The venues were identified using the FourSquare and Zomato APIs. Maps were used to plot the venues. The maps reveal that there are two major areas one can visit- nearby upper lake and rest of the city. Based on the visitor's choice of venue rating and budget, he/she can choose accordingly.