Restaurants in lucknow, India
using Foursquare and Zomato API

shashank vikram singh

20/05/2020

1.Introduction:

1.1 Background: Whenever a person searches for a venue in a new city, they're highly interested in the best places that the city has to offer. The person might want to know how good a given restaurant is or the price range it falls under. This extra information would help decide which venue to choose amongst the many venues in the city. Combining the location of the venues in the city with their price and rating information would surely help visitors in a city make better informed decisions about the places they should visit. . This project explores various venues in Chandigarh and attributes the data based on user ratings and average price. To explore this information, this project involves the juxtaposition of both the Foursquare API and the Zomato API to fetch complete information of various venues (including name, address, category, rating, and price). Further, a map of the venues with specific color attributes will be plotted to highlight their position, and information about these venues. Such plots imbibe bountiful information in the form of their colored representations and location on the map. This enables any visitor to take a quick glance and decide what place to visit.

1.2 Interested audience: The target audience for such a project is twofold. Firstly, any person who is visiting 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

2.1 Data sources: To get location and other information about various venues .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 4 kilometers from the center of lucknow 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.

2.2Data cleaning:

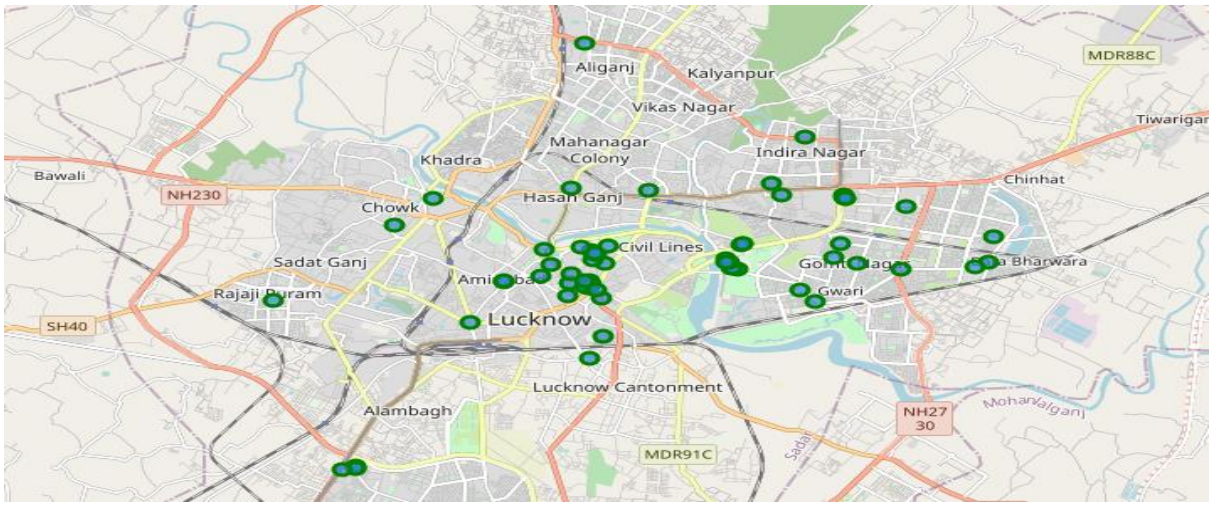


Fig 1: eateries retrieved from foursquare API

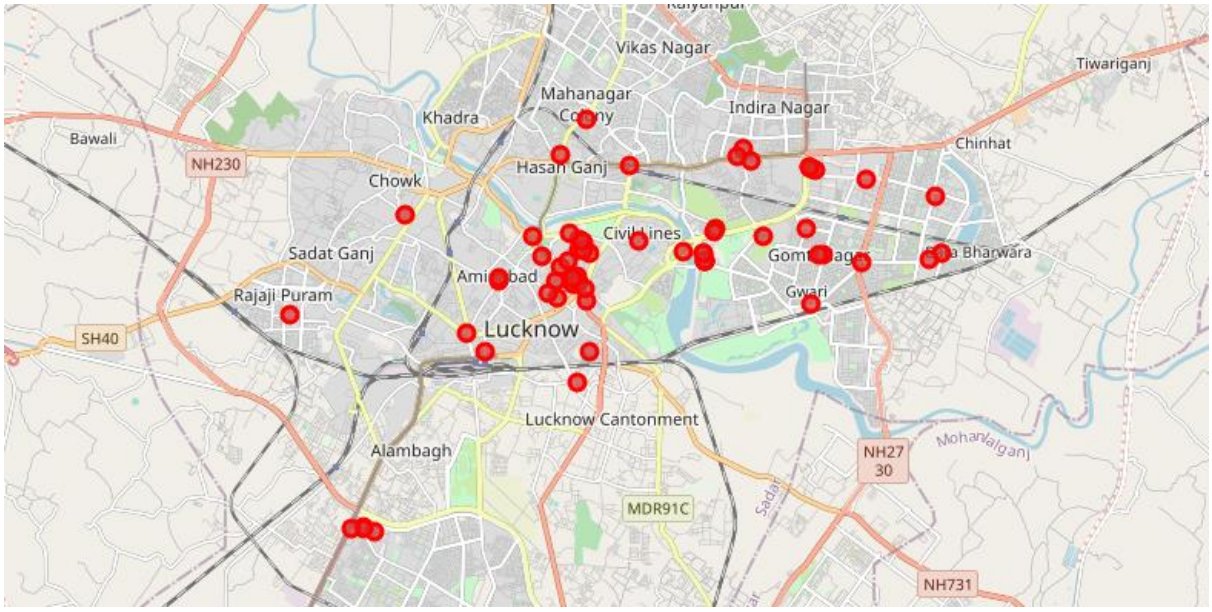


Fig 2: venues retrieved from Zomato API

From fig1 and 2 we can clearly see that some venues from the two API do not align with each other.Thus I combined them wsing their longitude and latitude. 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

So the data we got looks like this.

	categories	venue	latitude	longitude	price_range	rating	address	average_price
0	Indian Restaurant	The Mughal's Dastarkhwan	26.8442	80.9403	3	4.5	29, BN Road, Near Royal Hotel Crossing, Lalbag...	400.0
1	Indian Restaurant	Royal Sky	26.8504	80.9411	3	4.4	31/039, GF - 3 (31/37), Mahatma Gandhi Marg, H...	550.0
2	Neighborhood	Madras Restaurant	26.8484	80.9448	1	3.6	Near Gandhi Ashram, Behind Amar Fax, Hazratgan...	100.0
3	Café	Cappuccino Blast	26.8337	80.9478	2	4.3	12, Mall Avenue, Near, Hazratganj, Lucknow	350.0
4	Indian Restaurant	Dastarkhwan	26.8526	80.9368	2	4.5	20, Wala Qadar Road, DM Compound Colony, Kaise...	300.0
5	Bakery	La Reine	26.8726	80.9412	2	3.3	Shop No 26 , Close to HDFC Bank ATM, Sinha Mar...	250.0
6	Hotel	EOS Bar & Bistro- Hotel Levana	26.8474	80.9445	4	4.2	Hotel Levana, 3rd Floor, 72, MG Marg, Hazratga...	750.0
7	Shopping Mall	Quality Like Five Star	26.8469	80.9438	1	3.0	A 21, Janpath Market, Off MG Marg, Lalbagh, ...	150.0

Fig 3 : combined dataset

3.Methodology and analysis: . I extract the location data from the Foursquare API for all venues up to a distance of 4 kilometers from the center of Chandigarh. 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.

As a final step, I will analyse 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.

3.1 Categories

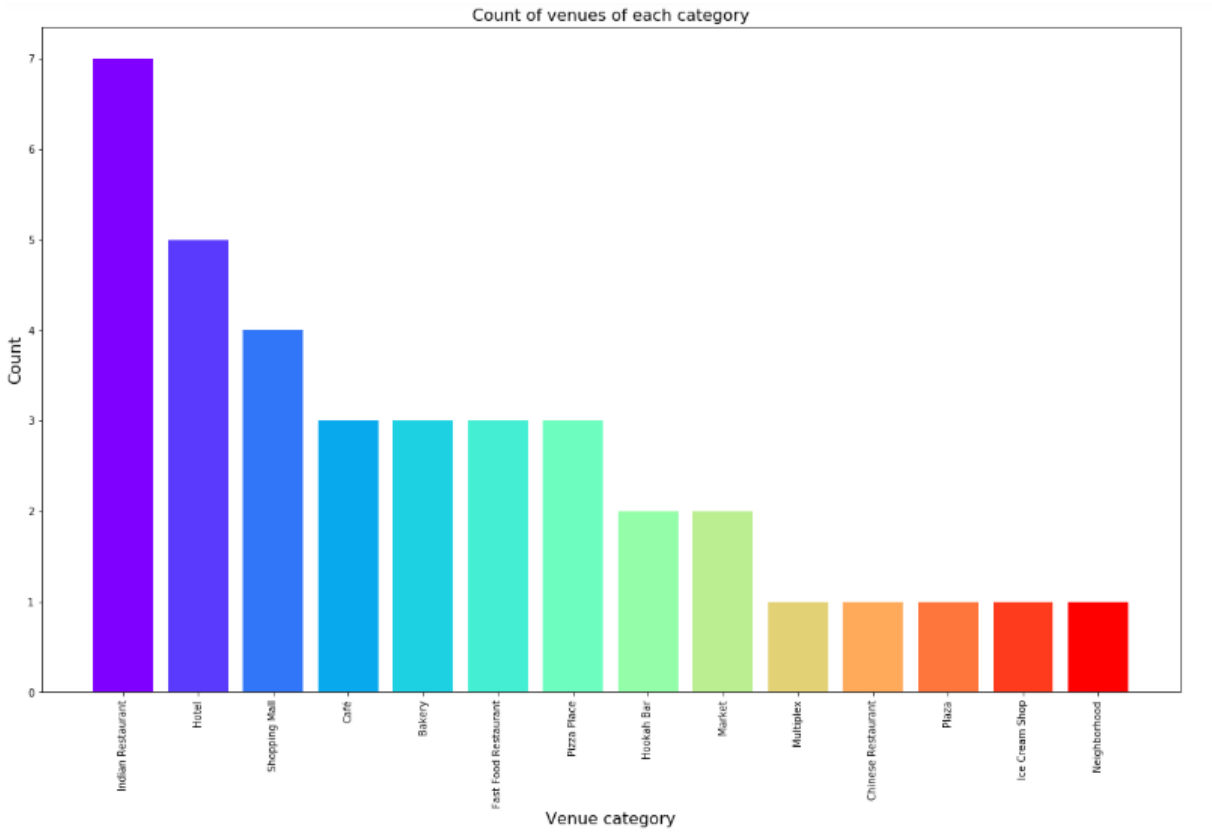


Fig 4:count of various type of restaurants

From figure 4, we see that the majority venues are Indian Restaurants . This is closely followed by Hotels

3.2 Rating : 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

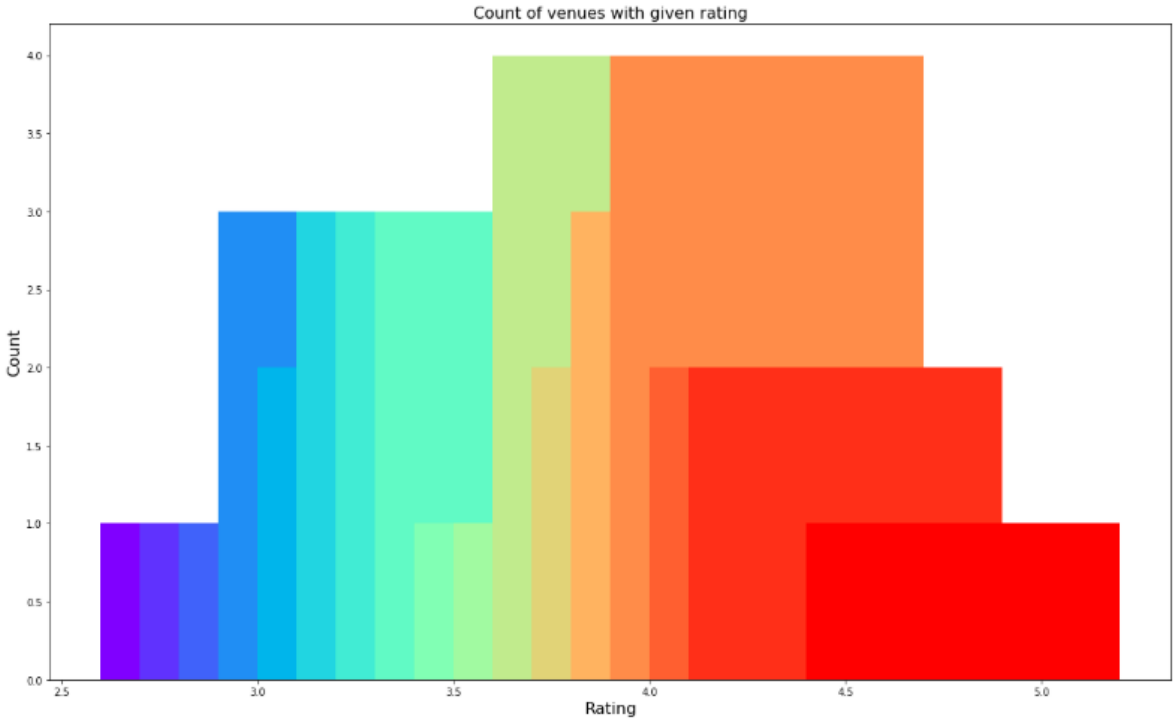


Fig 5:count of venue with the rating

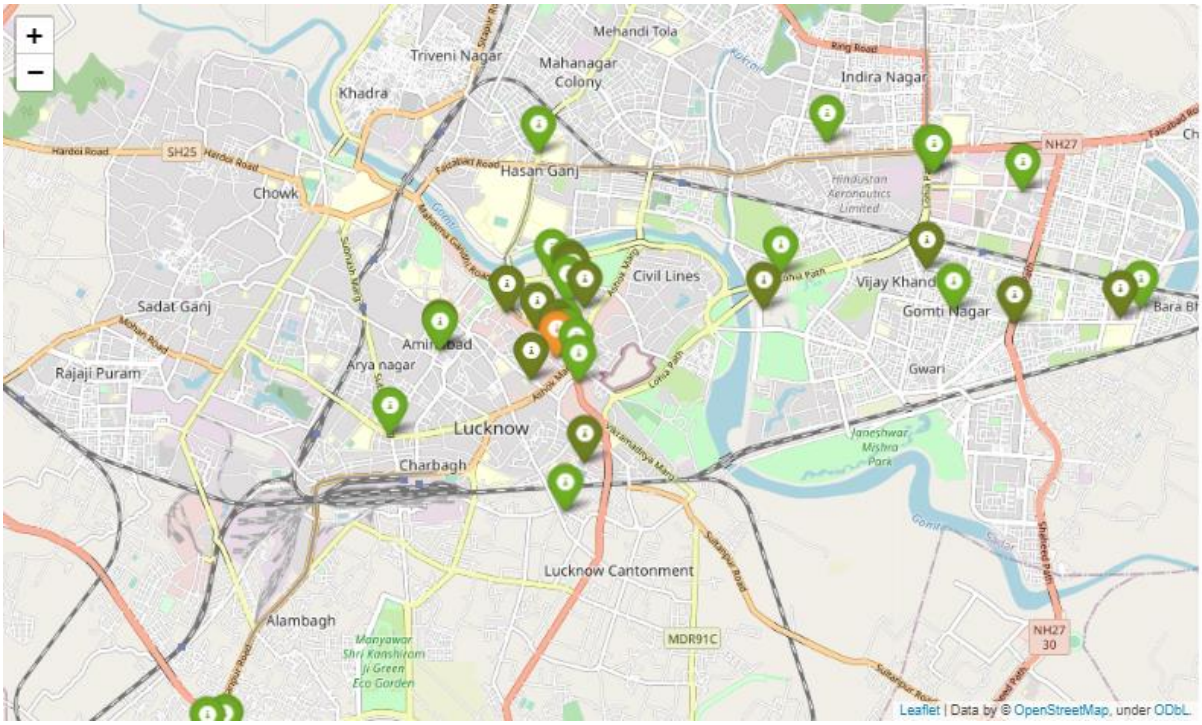


Fig: plot with different ratings

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. 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.

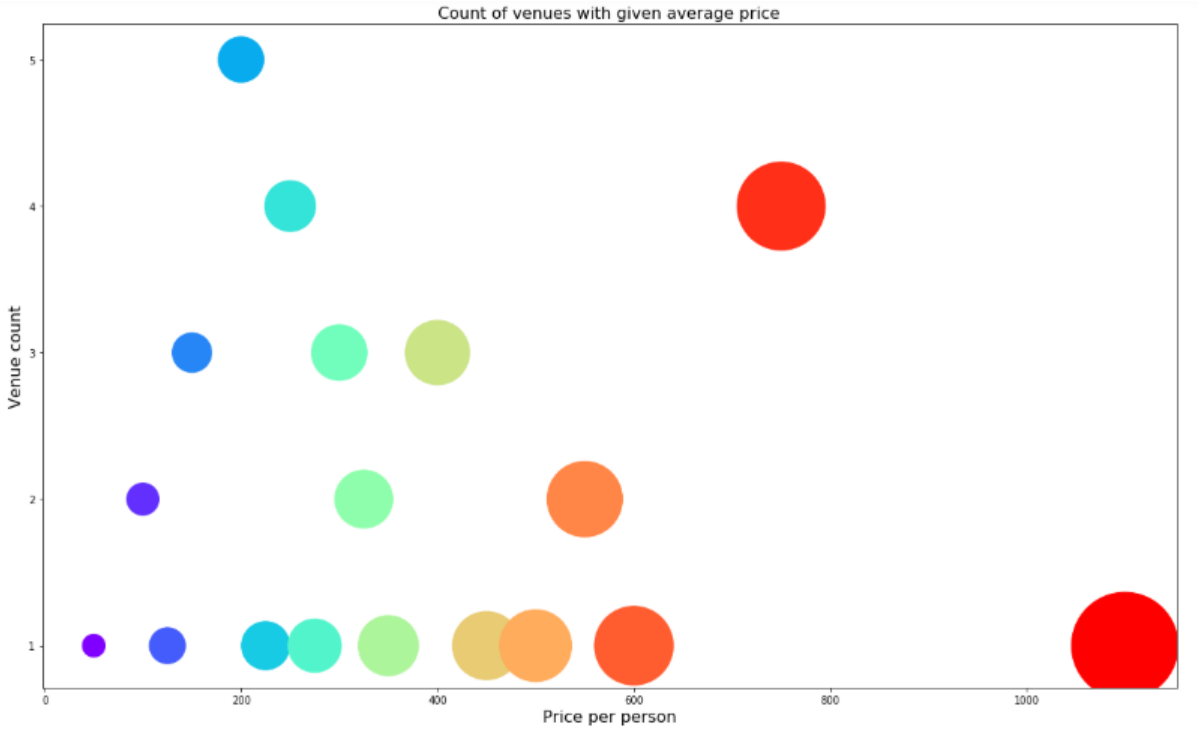
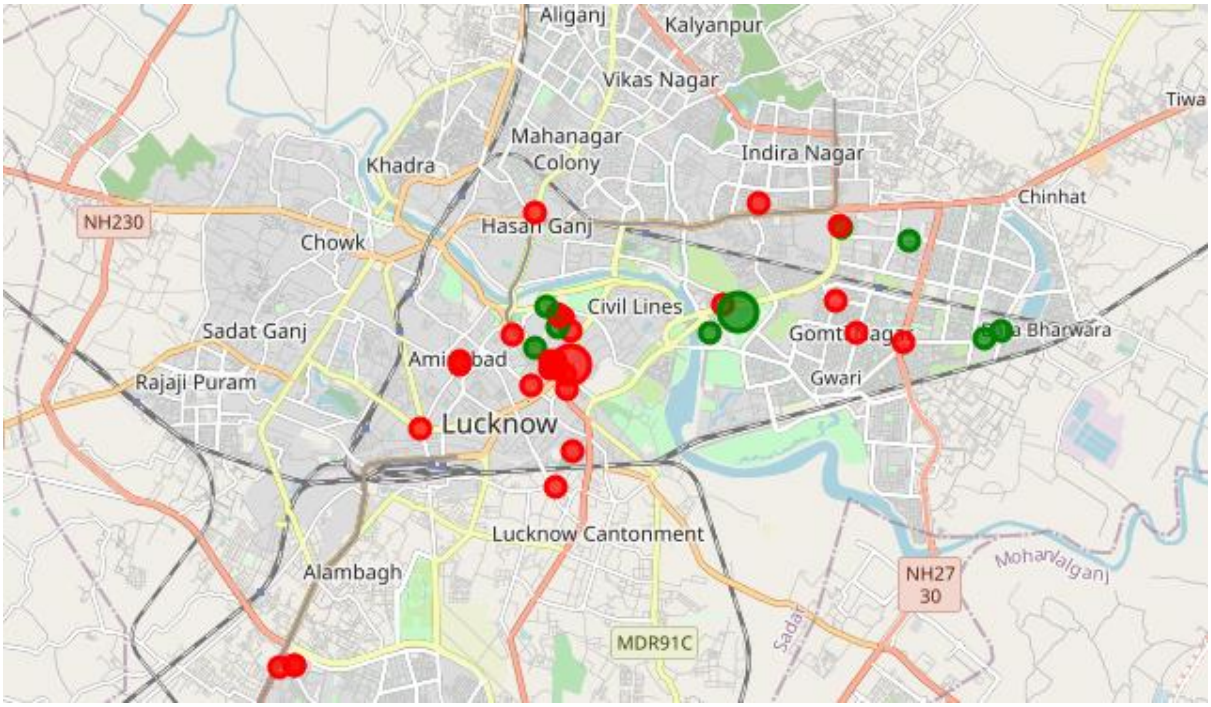


Fig 7 price per person with count of venues with that price

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.



we see the two clusters:

1. 1. The first cluster (green) is spread across the whole city and includes the majority venues. These venues have mean price range of 1.71 and rating spread around 3.57.
2. The second cluster (red) is very sparsely spread and has very limited venues. These venues have mean price range of 3.21 and rating spread around 4.03

Results: After collecting data from the Foursquare and Zomato APIs, we got a list of 120 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 9. We identified that from the total set of venues, majority of them were Indian Restaurants. A visitor who loves hotels/Indian Restaurants would surely benefit from coming to Lucknow.

Finally, through clusters we identified that there are many venues which are relatively lower priced but have an average rating of 3.57. On the other hand, there are few venues which are high priced and have average rating of 4.03.

1. If you're looking for cheap places with relatively high rating, you should check Hazratganj.
2. If you're looking for the best places, with the highest rating but might also carry a high price tag, you should visit gomti nagar.

5. Conclusion: The purpose of this project was to explore the places that a person visiting Chandigarh 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:Gomti Nagar or Hazratganj. Based on the visitor’s venue rating and price preferences, he/she can choose amongst the three places.