

Exploring venues in Chandigarh, India using Foursquare and Zomato API

Karan Bhanot

June 15, 2018

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 to 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.

Chandigarh is composed of a number of sectors spread across a total area of 114 sq Km. There are many venues (especially restaurants, hotels and cafes) which can be explored. 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 Chandigarh, India 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 in Chandigarh, 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 Chandigarh 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.

2.2 Data Cleaning

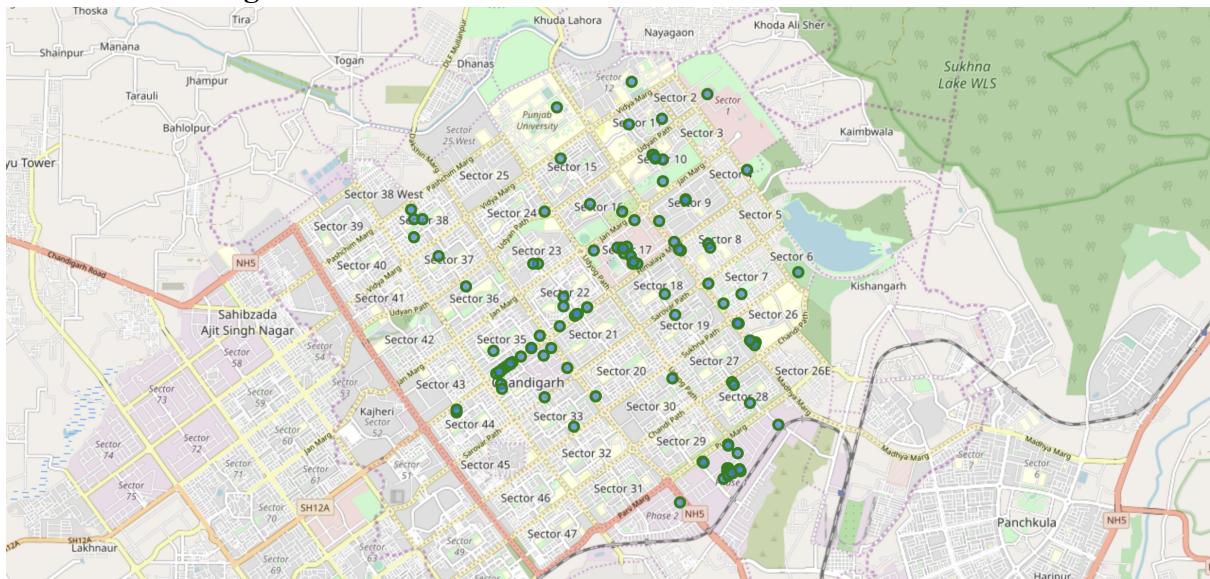


Figure 1: Venues retrieved from Foursquare API

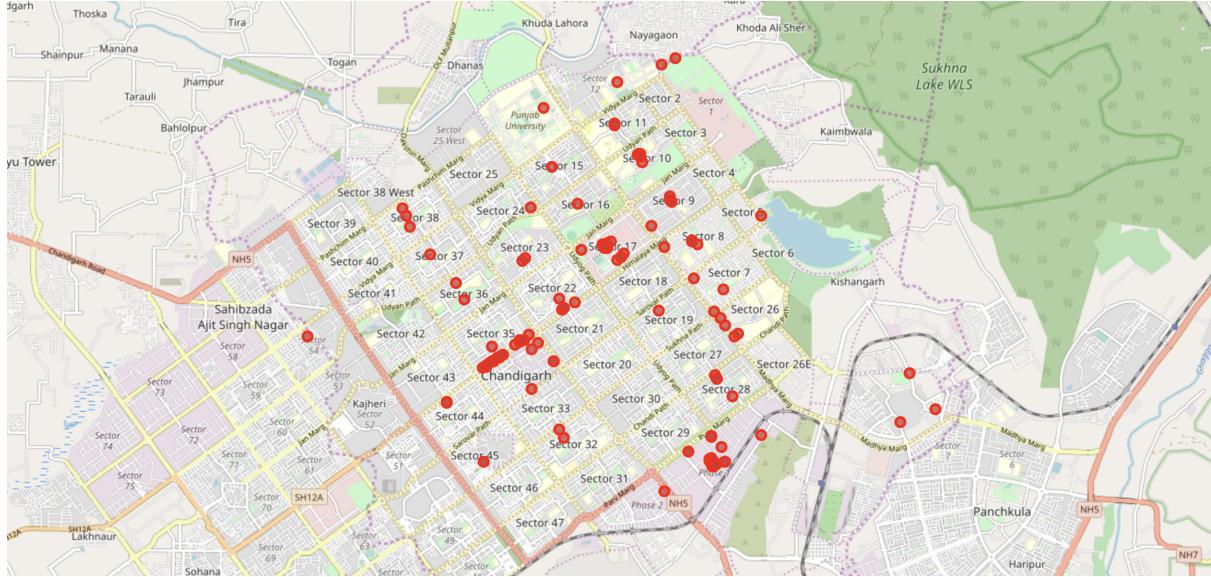


Figure 2: Venues retrieved from Zomato API

From figure 1 and figure 2, we can clearly see that while some venues from the two APIs do not align with each other. Thus, I decided to combine them using their latitude and longitude 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.

They can be categorised as follows:

1. There are venues that have specific restaurants/cafes inside them as provided by Zomato API (Pizza Hut inside Elante Mall).
2. Two locations are so close that they have practically same latitude and longitude values (The Pizza Kitchen and Zara).
3. Some venues have been replaced with new venues (Underdoggs has now been replaced by The Brew Estate).

Venues belonging to category 1 and 3 are perfect to keep. However, the venues that belong to category 2 should be dropped. After careful inspection and removal, the final dataset had a total of 49 venues with which we can work.

	categories	venue	latitude	longitude	price_range	rating	address	average_price
0	Hotel Bar	Sundarams	30.7302	76.7735	1.0	3.4	Hotel Aroma Complex, Himalaya Marg, Sector 22 ...	175.0
1	Sandwich Place	Backpackers Cafe	30.7475	76.7933	3.0	4.3	SCF 16, Inner Market, Sector 9 D, Sector 9, Ch...	600.0
2	Ice Cream Shop	Softy Corner	30.7405	76.7816	1.0	4.6	SCO 87, Sector 17, Chandigarh	150.0
3	Fast Food Restaurant	Hot Millions	30.7408	76.7822	2.0	2.2	SCO 73 & 74, Sector 17 D, Sector 17, Chandigarh	400.0
4	Bakery	Nik Baker's	30.7216	76.7601	2.0	4.5	SCO 441 - 442, Sector 35 C, Sector 35, Chandigarh	350.0

Figure 3: Final data aggregated from both APIs

2.3 Data Cleaning

The complete dataset for this project includes the location information of various venues in Chandigarh, India and the venue information (rating and others) for each venue. For fetching the location information, the project uses the Foursquare location data API and for fetching the venue information it uses the Zomato API. Once the information is aggregated, the dataset would then be used to plot information on the map as well as filter results based on ratings and price.

The dataset would include the following fields:

1. Name: The name of the venue.
2. Category: The category to which the venue belongs to. As there are two APIs, which have different set of category division, I'll keep only one category information.
3. Latitude: Describes the latitude of the venue.
4. Longitude: Describes the longitude of the venue.
5. Address: The complete address of the venue.
6. Average cost: It defines the estimated cost for a person visiting the venue.
7. Price range: Each venue is divided into separate price bands by Zomato and this range defines it.
8. Rating: The overall rating of the venue.

Steps of data aggregation:

1. The latitude and longitude values are fetched for Chandigarh.
2. Using the Foursquare API, nearby venues are fetched.
3. For each of those venues, the Zomato's search API helps to extract the closest match (the same venue itself) as the parameters are set to fetch only one result with the given latitude and longitude of the venue.
4. The data including address, average cost, price range and rating are extracted and added to the data from the Foursquare API.

Data Source APIs:

1. Foursquare API: <https://developer.foursquare.com>
2. Zomato API: <https://developers.zomato.com/api>

3. Methodology

As a first step, we retrieved the data from two APIs (Foursquare and Zomato). We extract venue information from the center of Chandigarh, upto a distance of 4 Km. The latitude and longitude values are then used to fetch venue rating and price from Zomato.

The data from the two sources is carefully combined based on the name, latitude and longitude values from the two sources. The final dataset would include the rating and price values for each venue

Next, we analyse the data that we created based on the ratings and price of each venue. We identify the top category types. We identify 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. This will allow us to clearly identify which venues can be recommended and with what characteristics.

Finally, we'll discuss and conclude which venues to be explored based on visitor requirement of rating and cost.

4. Analysis and results

5. Discussion

6. Conclusion

1 Basketball Assoc