**Exploring Hotels in Gurgaon, India**

**using Foursquare and Goibibo API**

Swati Nagpal

June 28, 2019

Table of Contents

[1. Introduction 3](#_Toc12623907)

[1.1 Background 3](#_Toc12623908)

[1.2 Interested audience 3](#_Toc12623909)

[2. Data 3](#_Toc12623910)

[2.1 Data Sources 3](#_Toc12623911)

[2.2 Data Cleaning 5](#_Toc12623912)

[3. Methodology and Exploratory Data Analysis 6](#_Toc12623913)

[3.1 Categories 7](#_Toc12623914)

[3.2 Rating 7](#_Toc12623915)

[4. Results and Discussion 9](#_Toc12623916)

[5. Conclusion 10](#_Toc12623917)

# **Introduction**

## **Background**

Whenever a person plan to visit a new city, their most important task is to book a nice hotel to enjoy their stay. The person might want to know how good a given hotel is or the price range it falls under. This extra information would help decide which hotel to choose amongst the many available in the city. Combining the location of the hotels in the city with their rating information would surely help visitors in a city make their stay comfortable.

Gurgaon, officially named Gurugram, is a city located in the northern Indian state of Haryana. It is situated near the Delhi-Haryana border, about 30 kilometres (19 mi) southwest of the national capital New Delhi, 153 km (95 mi) south of Karnal and 268 km (167 mi) south of Chandigarh, the state capital. It is one of the major satellite cities of Delhi and is part of the National Capital Region of India.

This project explores various hotels in Gurgaon and attributes the data based on user ratings. To explore this information, I have used Foursquare API and the Goibibo API to fetch complete information of various hotels (including name, address, pincode, category and rating). Further, a map of the hotels with specific color attributes will be plotted to highlight their position, and information about these hotels.

## **Interested audience**

Any person who is visiting Gurgaon, India can use the plots and maps from this project to quickly select places that suit their location and rating preferences. Any 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.

# **Data**

## **Data Sources**

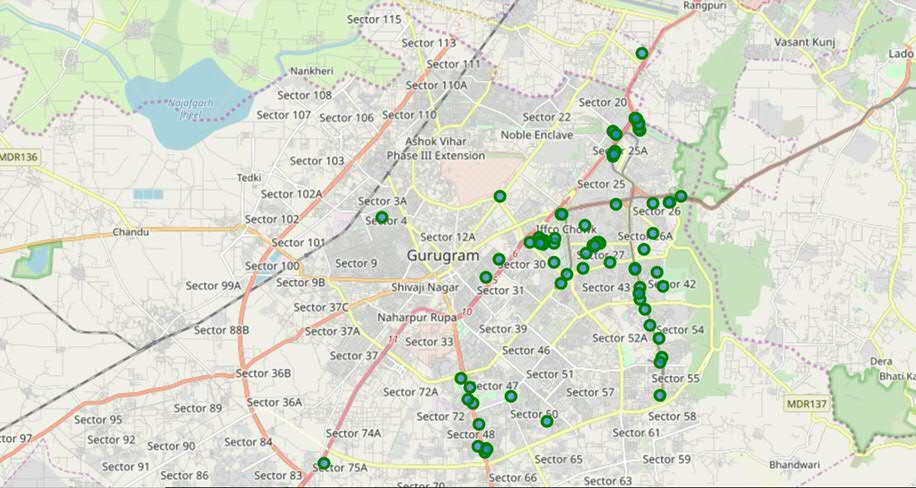
To get location and other information about various venues/hotels in Gurgaon, I used two APIs and decided to combine the data from both together.

Using the Foursquare’s explore API (which gives venues recommendations), I fetched venues up to a range of 10 kilometers from the center of Gurgaon and collected their names, categories and locations (latitude and longitude). Then I filter the data based on categories and extracted only hotel records.

Using the city ID for Gurgaon, I used the Goibibo search API to fetch hotel details from its database. This API allows to find hotels based on search criteria (usually the city id), 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://developer.foursquare.com/docs/api/venues/explore>), 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.



*Figure 1: Venues retrieved from FourSquare*

From Goibibo API (<https://developer.goibibo.com/docs>), I retrieved the following for each Hotel:

* **Name:** The name of the hotel.
* **Pincode:** The pincode of the hotel.
* **Rating:** The ratings as provided by many users.
* **Latitude:** The latitude value of the venue.
* **Longitude:** The longitude value of the venue.

## **Data Cleaning**



*Table 1: Data extracted for Hotel Category – FourSquare API*



*Table 2: Hotel Rating extracted from Goibibo API*

In table 1, I have extracted data for ‘Hotel’ category and you can see I only got 11 hotels. Then I extracted rating for these hotels using Goibibo API which you can see in table 2.

As it’s clearly visible that, we couldn’t get rating for all the foursquare API hotels, also rating not diverse for listed hotels. This doesn’t make good case for exploring the hotel.

So I decided to extract some more hotel data from Goibibo API and merge with foursquare API data so that we can have good analysis. I extracted data based on pincode, also dropped hotels with rating less than 2.0. All tourist prefer to have nice and comfortable stay and would mostly prefer hotels with rating more than 3.0.

As a final dataset, we’re left with 198 hotels with 5 columns as described in figure 3.



*Table 3: Final data combined from both APIs*

# **Methodology and Exploratory Data Analysis**

As a first step, I retrieve the venues in Gurgaon from Foursquare API. I extract the location data from the Foursquare API for all venues up to distance of 10 kilometers from the center of Gurgaon. Then I filter venue based on ‘Hotel’ category. Next step, I retrieve hotel rating from Goibibo API using city id as ‘Gurgaon.

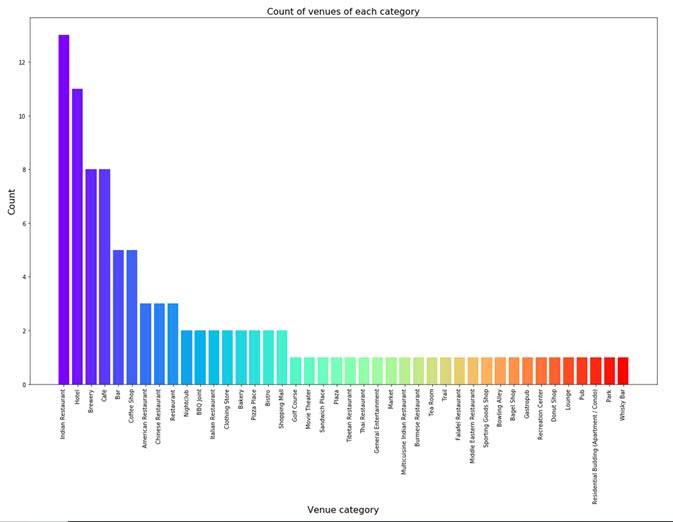
Using data cleaning, the dataset from the two APIs will be combined based on the hotel names, latitude, and longitude values as explained in Data Cleaning section. The final data will include the Hotel name, pincode, latitude, longitude and rating.

Using this dataset, I begin by analyzing the top hotel that exist in Gurgaon. I will then explore the venues on maps. This will allow us to better understand the location of various hotels. I’ll also explore the hotels based on the ratings. The hotel will be plot using proper color coding such that a simple glance at the map would reveal the location of the hotel as well as give information about them. I aim to identify places which can be recommended to visitors based on their rating preferences. I’ll also cluster the hotels and see if we can draw meaningful information.

As a final step, I will analyze these plots and try to draw conclusions on which places can be recommended to tourist to find good hotel. 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 Gurgaon. As there are many restaurants, I believe that the majority venues shall include restaurants.

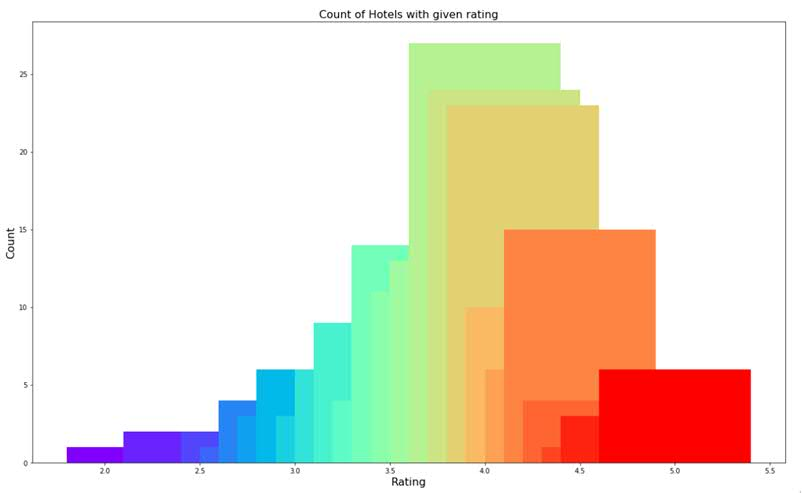


*Figure 2: Count of various types of venues in Gurgaon*

From figure 2, we see that the majority venues are Indian Restaurants. This is closely followed by Hotel. Gurgaon is clearly hub for great food and comfortable stay.

## **Rating**

Next, I’ll explore the ratings of various hotels in Gurgaon. 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 hotels with that rating. I decided to plot the bar chart to see what average rating hotels get in Gurgaon. This can be seen in figure 3.



*Figure 3: Rating and count of Hotel with that rating*

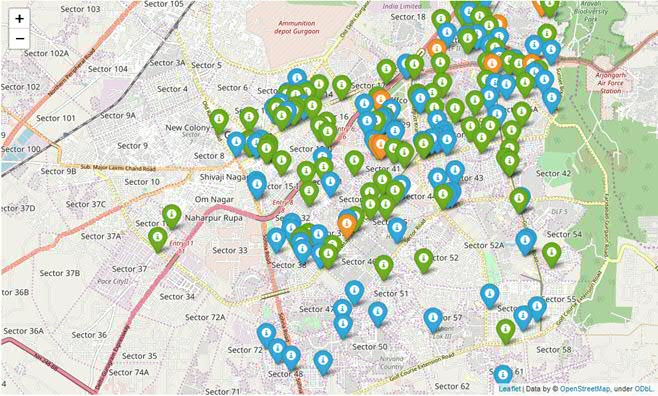
While the whole range of rating of venues might stretch from 1 to 5, the average rating is spread across 4 with maximum number of hotels scoring between 3 and 5.

I followed this information by plotting the venues on the map of Gurgaon. You can see color mapping with Label in table 4.

|  |  |  |
| --- | --- | --- |
| Bin | Label | Color |
| 1-2 | Low | Red |
| 2-3 | Okay | Orange |
| 3-4 | Good | Green |
| 4-5 | Very Good | Blue |

*Table 4: Label and color mapping*

Looking at figure 4 reveals the same results as the bar plot(figure 3). However, it is interesting to note that many high rated hotels are located near Sector 29, Iffco Chowk and Golf Course Road. Few are present near Sector 48 and Sector 50.



*Figure 4: Plot of hotels with different ratings*

Overall, Gurgaon on an average has good rating for its hotel.

# **Results and Discussion**

After collecting data from the Foursquare APIs, we got a list of 100 different venues and we extracted data for Hotel. We got 11 hotels from Foursquare API. Then we used Goibibo API to extract rating for those11 hotel. However, rating was available only for 7 hotels, so we extracted more data from Goibibo database based on location. Final dataset results in total 198 hotels.

We identified that from the total set of venues, majority of them were Indian Restaurants and Hotels. Then we further continued analysis based on Hotel.

While the ratings range from 1 to 5, majority hotels have ratings close to 4. This means that most hotel provide nice and comfortable stay to the visitors, thus indicating the high rating. When we plot these hotels on the map, we discover that there are clusters of hotels around Sector29, Iffco Chowk and Golf Course Road. These clusters also have very high ratings (more than 3).

A company can use this information to build an online website/mobile application, to provide users with up to date information about various hotels in the city based on the search criteria (name, location, rating).

# **Conclusion**

The purpose of this project was to explore the venues and hotels that a person visiting Gurgaon can visit. The venues have been identified using Foursquare API and hotel rating is taken from Goibibo API and have been plotted on the map.