**Exploring venues in Jalandhar, India**

**Using Foursquare and Zomato API**

Sakshi Aggarwal

30 May, 2020

1. **Introduction**

* 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.

Jalandhar has many venues (especially restaurants, hotels and cafes) which can be explored. This project explores various venues in Jalandhar 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 colour 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. **Interested audience**

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

1. **Data**

* 1. **Data Sources**

To get location and other information about various venues in Jalandhar, 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 20 kilometres from the centre of Jalandhar 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.

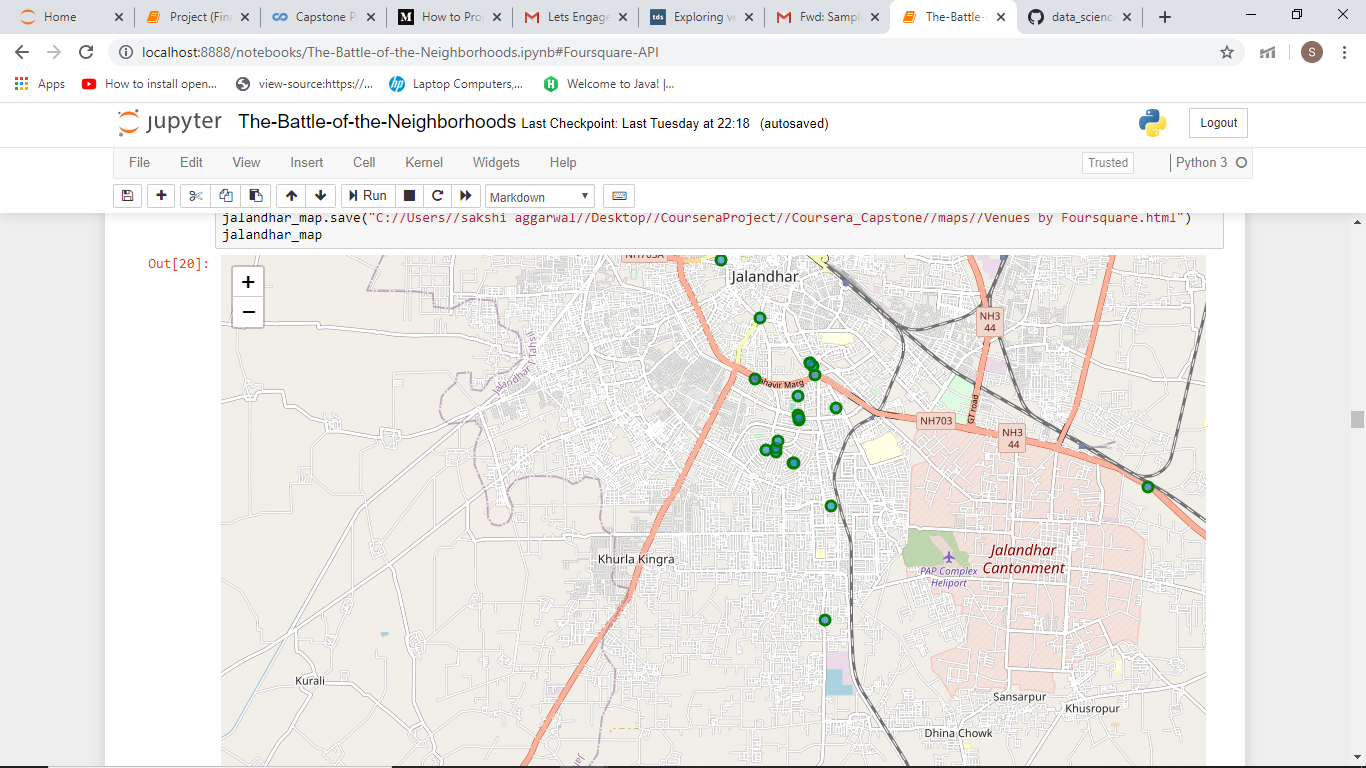
From Foursquare API (<https://developer.foursquare.com/>), I retrieved the following for each venue:

* **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.

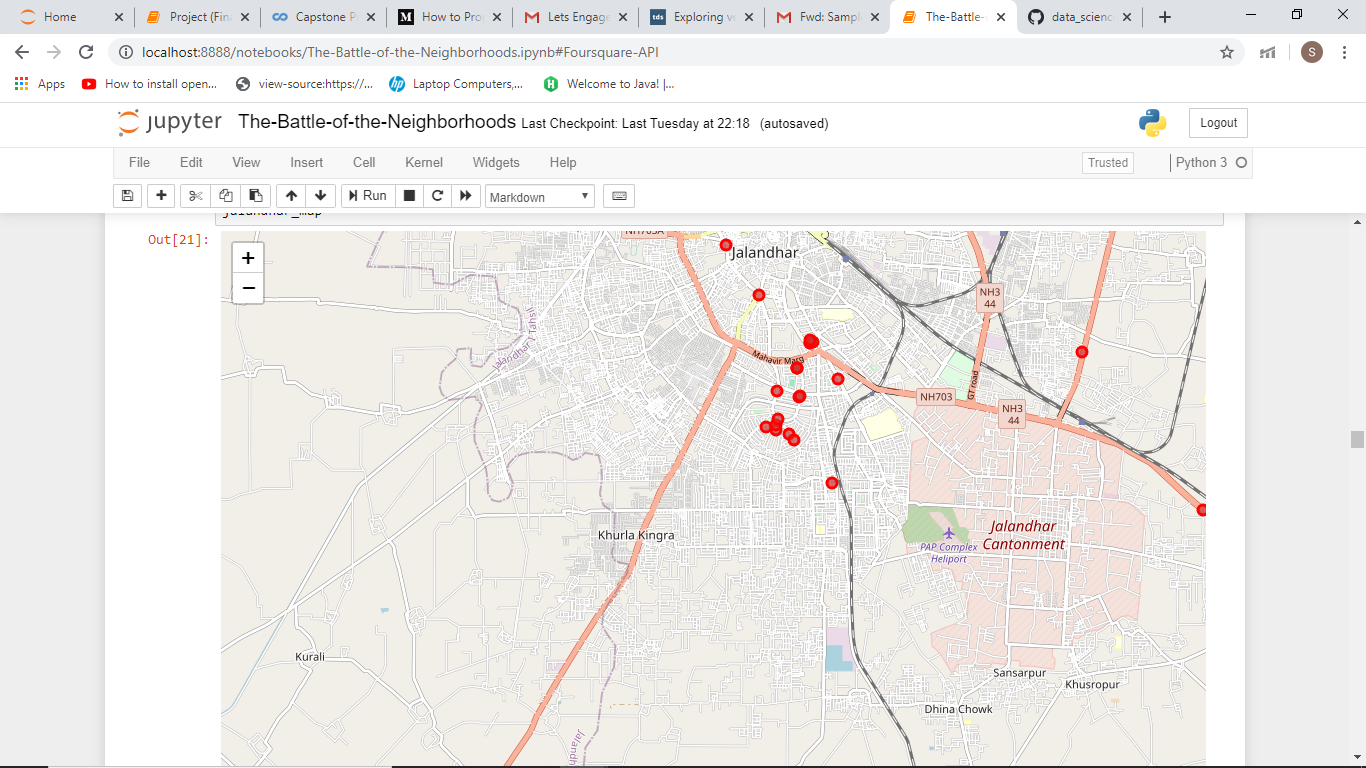
From Zomato API (<https://developers.zomato.com/api>), I retrieved the following for each venue:

* **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.

* 1. **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 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.

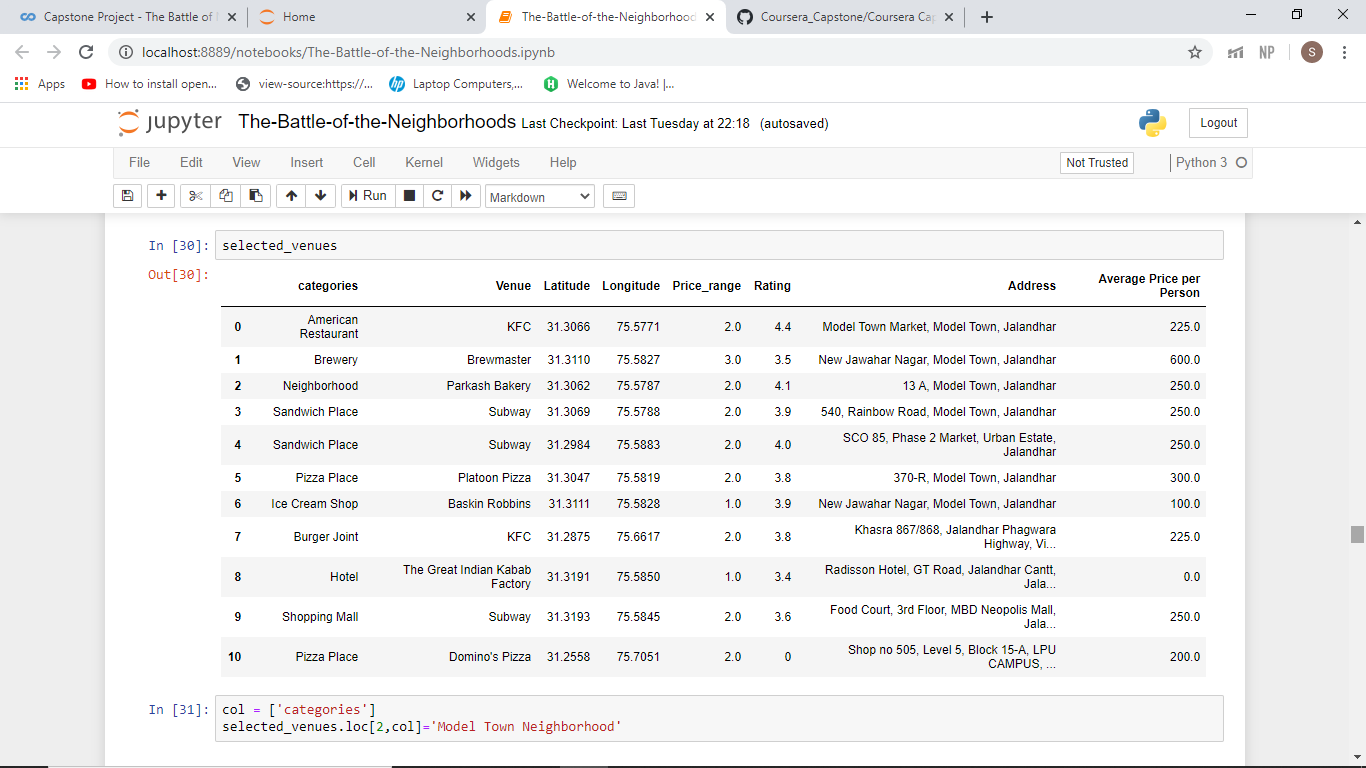
They can be categorised as follows:

1. Two locations are so close by that they have practically same latitude and longitude values.
2. There are venues that have specific restaurants/cafes inside them as provided by Zomato API.

The venues which belong to ***category 2*** are alright to keep, the venues that fall in ***category 1*** should be removed.

After careful inspection and removal, the final dataset had a total of 10 venues with which we can work.

As a final dataset, we’re left with 10 venues with 8 columns as described in figure 3.



*Figure 3: Final data aggregated from both APIs*

1. **Methodology and Exploratory Data Analysis**

As a first step, I retrieve the venues in Jalandhar from Foursquare and Zomato APIs. I extract the location data from the Foursquare API for all venues up to a distance of 20 kilometres from the centre of Jalandhar. Using this, I fetch the venue information including price and rating data from Zomato API.

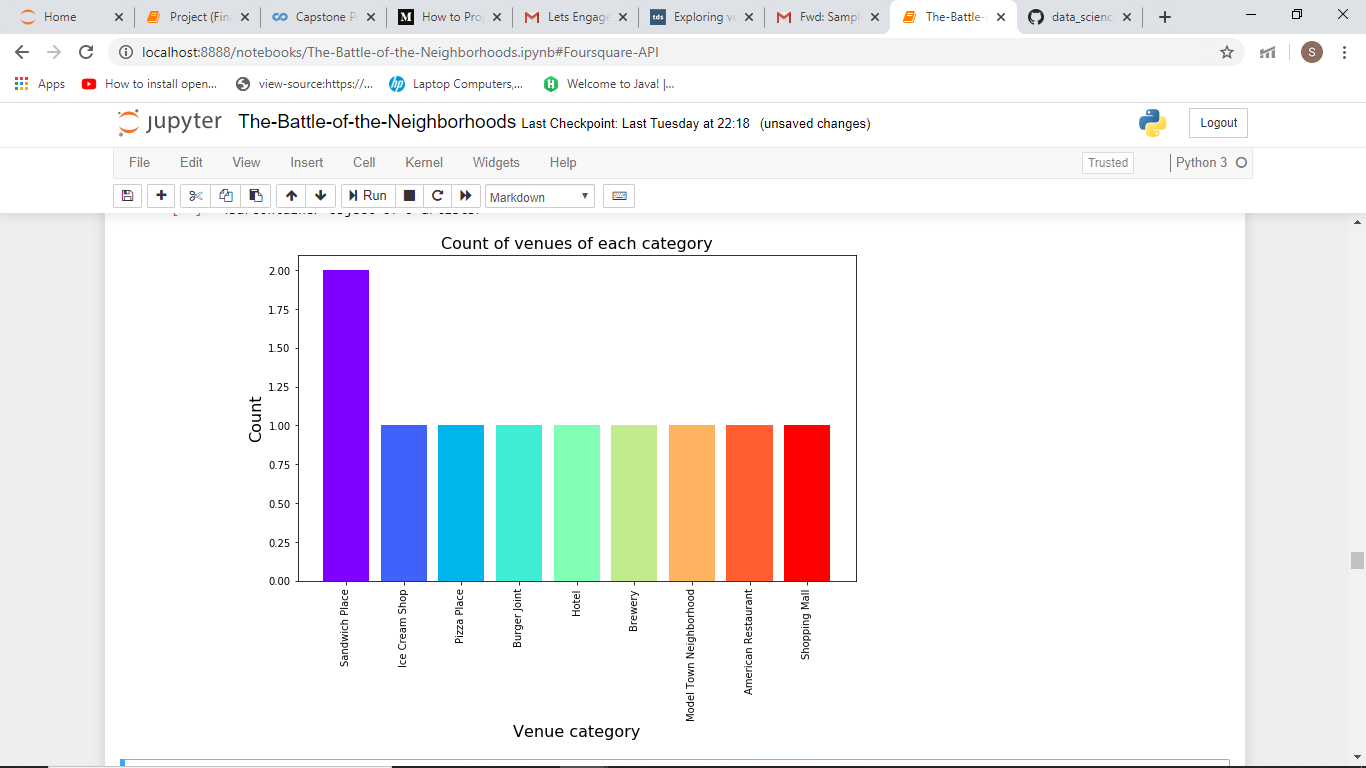
Using data cleaning, the dataset from the two APIs will be combined based on the venue names, latitude, and longitude values. One to one matching and careful data inspection would be used to remove any remaining outliers such as multiple venues at the same location from the two datasets. The final data will include the venue name, category, address, latitude, longitude, rating, price range, and average cost per person.

Using this dataset, I begin by analysing the top venue types that exist in Jalandhar. 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 Jalandhar.

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.

* 1. **Categories**

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



*Figure 4: Count of various types of venues in Jalandhar*

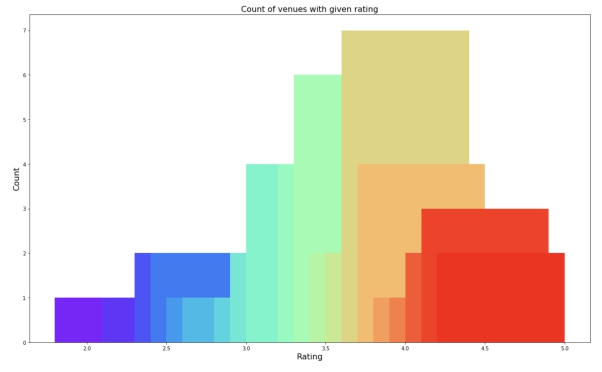
From figure 4, we see as the data retrieved for venues of Jalandhar was not very large, It can be observed from bar-chart that except **Sandwich Place Category**, all other **Venue Categories** have only one venue each.

* 1. **Rating**

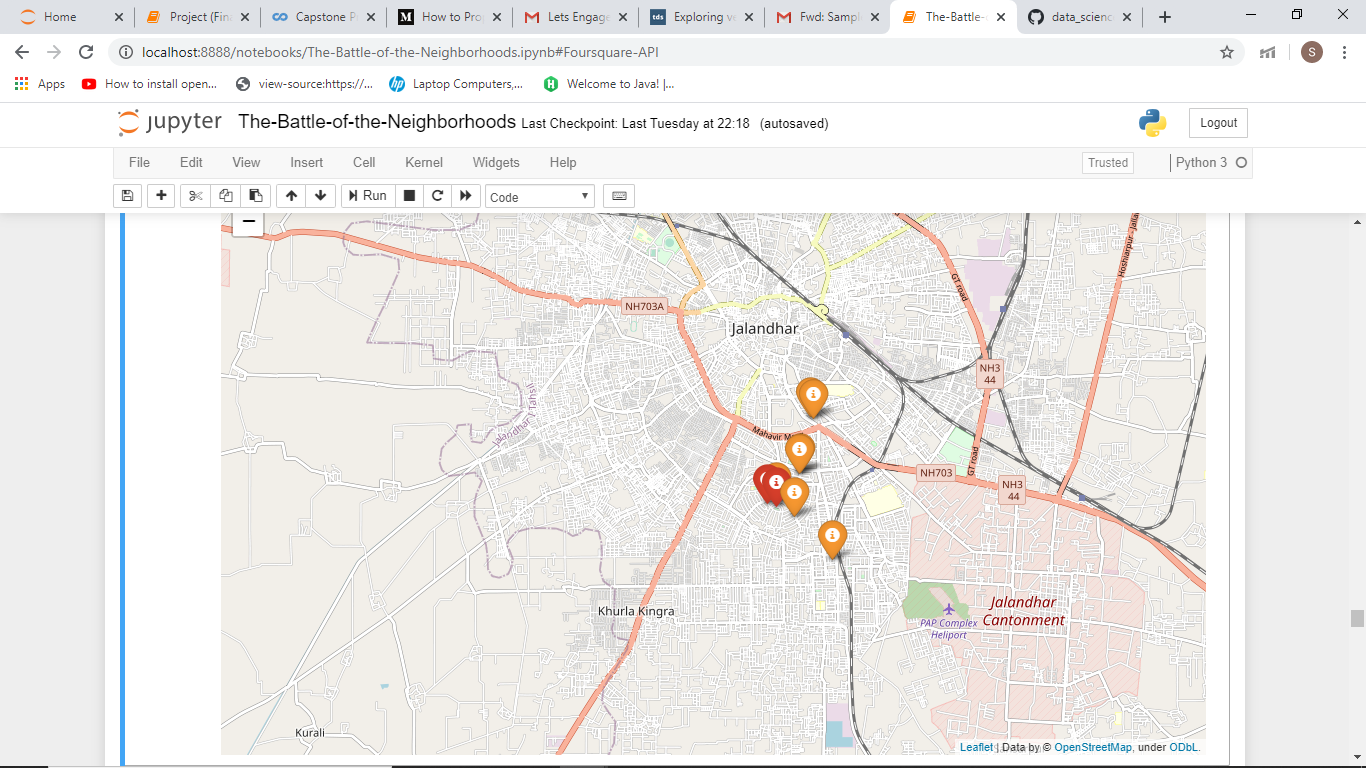
Next, I’ll explore the ratings of various venues in Jalandhar. I decided to plot a bar chart with x-axis as the rating from 1 to 5 and the y-axis as the count of venues with that rating. I decided to plot the bar chart to see what average rating venues get in Jalandhar. This can be seen in figure 5.

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 Jalandhar. The venues that were rated below 3 were marked by green and dark green while the venues that were rated more than or equal to 3 were plot as orange and red. Taking a look at figure 6 reveals the same results as the bar plot. From the plot it is clear that majority venues have a rating between 3.5 to 4.From map, It appears that many venues are located in **Model Town** with rating above 3.5.



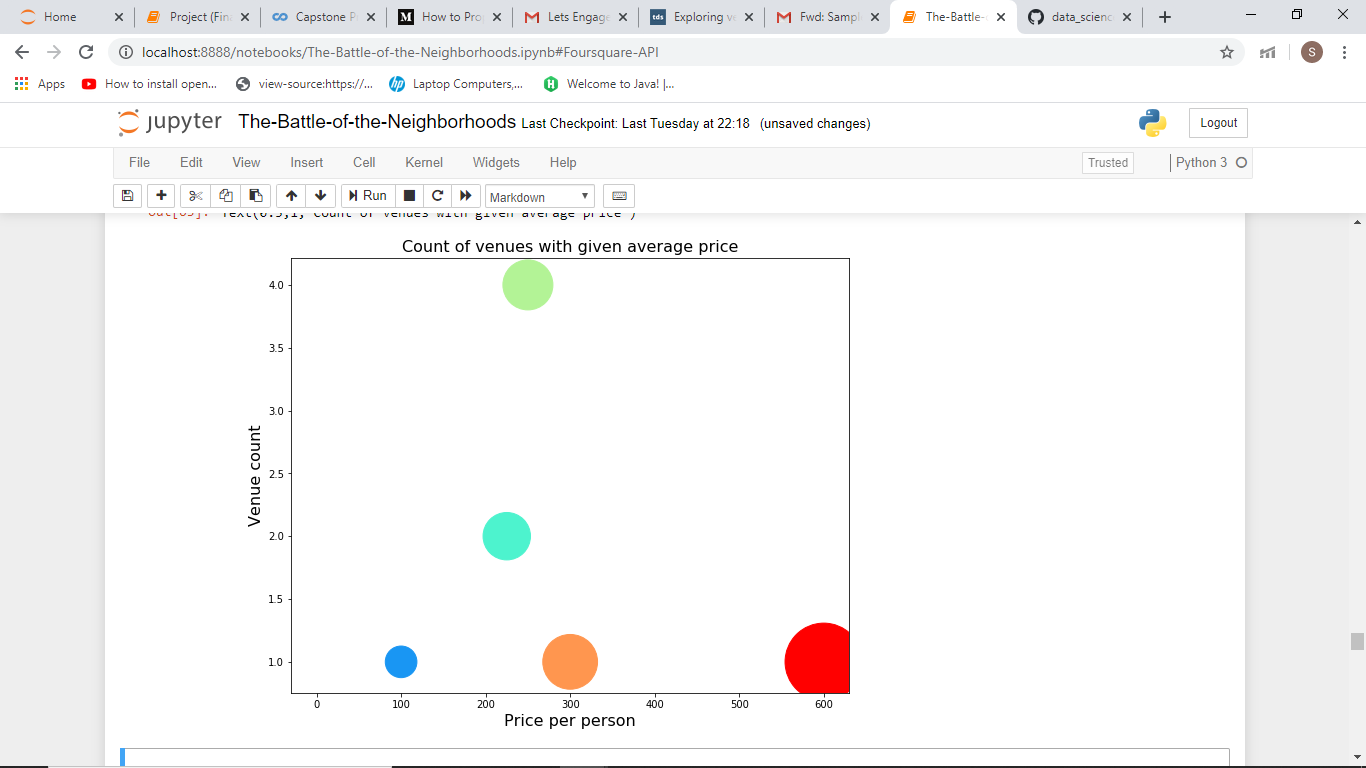
*Figure 5: Rating and count of venues with that rating*



*Figure 6: Plot of venues with different ratings*

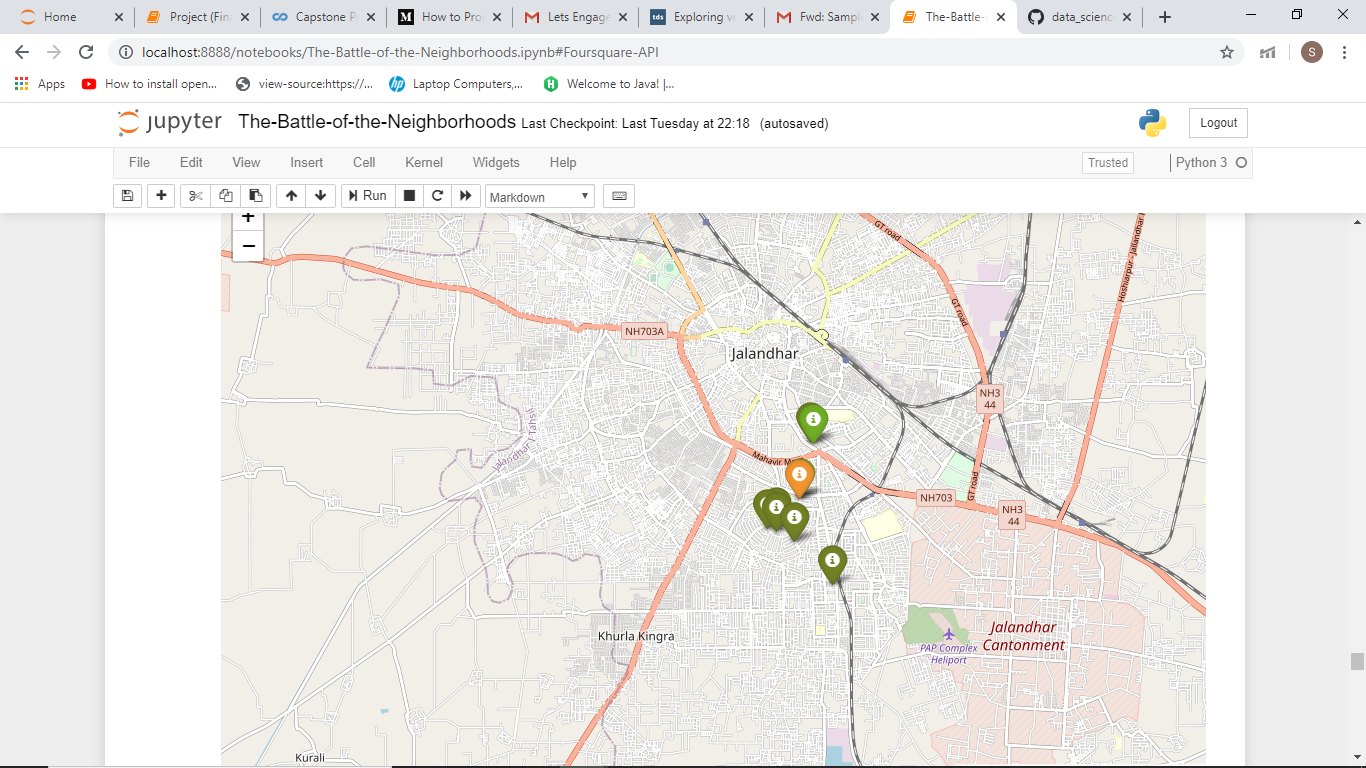
* 1. **Price**

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 300 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 600 for one person.



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

I also plot the venues based on their price range.



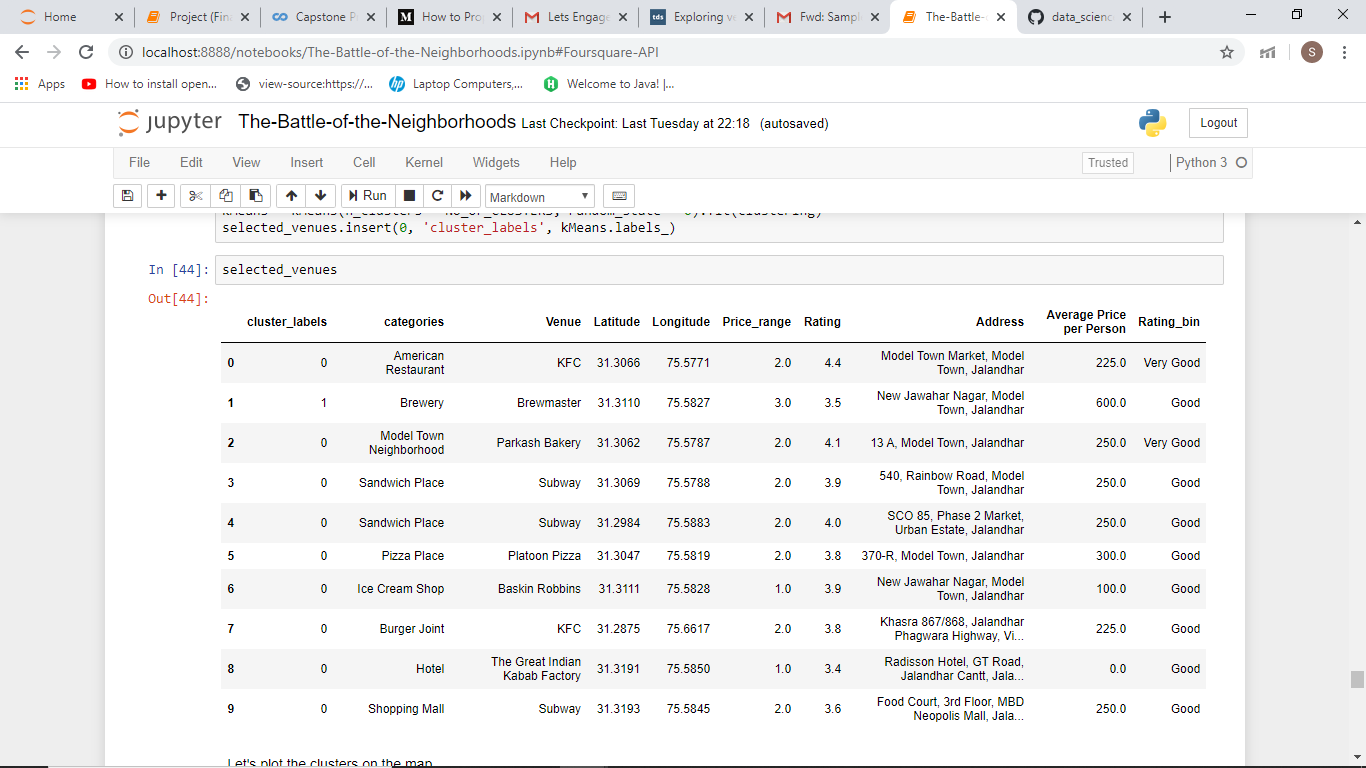
*Figure 8: Plot of venues with different prices*

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 with good ratings are located in Model Town.

Model Town seems to have average to high priced venues.

**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*

In figure 9, we see the two clusters:

1. The first cluster (green): The venues for cluster 0 have Average price per person of 205.56 and rating spread around 3.88.
2. The second cluster (red) is very sparsely.
3. **Results and Discussion**

Based on our analysis, we can draw a number of conclusions that will be useful to aid any visitor visiting **Jalandhar, India**.

After collecting data from the **Foursquare and Zomato APIs**, we got a list of **29 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 10.

We identified that from the total set of venues, majority of them were **Sandwich and Pizza Places**.

While the complete range of ratings range from 1 to 5, the majority venues have ratings close to 4. 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 is cluster of venues around **Model Town**. The cluster also have high rating (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 300** for one person. Also, it's been observed that Model Town seems to have average to high priced venues.

Finally, through clusters we identified that there are many venues which are relatively lower priced but have an average rating of **3.88**.

**So, 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 Model Town**.

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).

1. **Conclusion**

The purpose of this project was to explore the places that a person visiting Jalandhar could visit. The venues have been identified using Foursquare and Zomato API and have been plotted on the map. The map reveals that the major area in Jalandhar to visit is Model Town. Based on the visitor's venue rating and price requirements, he/she can choose places.