Venues in Dehradun, India

Contents

Introduction

The aim of the project is to identify venues in Dehradun, India based on their rating and average prices. In this notebook, we will identify various venues in the city of Dehradun using Foursquare API and Zomato API, to help visitors select the restaurants that suit them the best.

Target Audience: Whenever a user is visiting a city they start searching for places to visit during their stay. They primarily look for places based on the venue ratings across all venues and the average prices such that the locations fits in their budget.

Here, we'll identify places that are fit for various individuals based on the information collected from the two APIs topped with Data Science. Once we have the plot with the venues, any firm/individual can launch an application using the same data and suggest users the necessary information

Data

Collection: To begin with, let's look at Dehradun on the Map using the folium library. Latitude & Longitutes where the map is centered are supplied *manually*.

The data is fetched from two different APIs.

- **Foursquare API**: The Foursquare API to fetch venues in Dehradun starting from the centre upto 10 Kilometers in each direction.
- **Zomato API**: The Zomato API provides information about various venues including the complete address, user ratings, price for two people, price range and a lot more.

Cleaning: The data from multiple resources might not always align. Thus, it is important to combine the data retrieved from multiple resources properly.

First plot the two data points on the map. Then try to combine data points that have their latitude and longitude values very close to one another. From the remaining selected venues, inspect the venues to ensure that any remaining mismatched venues are also removed from the final dataset of venues before any analysis.

- To combine the two (Foursquare & Zomato) datasets, check that the latitude and longitude values of each corresponding venue match. Thus, round both the latitude and longitude values up to 4 decimal places. Then, calculate the difference between the corresponding latitude and longitude values and check if the difference is less than 0.0004 which should ideally mean that the two locations are same.
- Drop the venues which have 0.0 rating as it's significant or not been rated yet.

In [11]:
```
foursquare_venues['lat'] = foursquare_venues['lat'].apply(lambda lat: round(float(lat), 4))
foursquare_venues['lng'] = foursquare_venues['lng'].apply(lambda lng: round(float(lng), 4))
zomato_venues['latitude'] = zomato_venues['latitude'].apply(lambda lat: round(float(lat), 4))
zomato_venues['longitude'] = zomato_venues['longitude'].apply(lambda lng: round(float(lng), 4))
```

In [12]:
```
dataset = pd.concat([foursquare_venues, zomato_venues], axis = 1)
#dataset
#dataset['lat_diff'] = dataset['latitude'] - dataset['lat']
#dataset['lng_diff'] = dataset['longitude'] - dataset['lng']
```

In [13]:
```
#selected_venues = dataset[(abs(dataset['lat_diff']) <= 0.0004) & (abs(dataset['lng_diff']) <= 0.0004)].reset_index(drop = True)
selected_venues=dataset
selected_venues
```

Out[13]:

	name	categories	lat	lng	venue	latitude	longitude	price_for_two	price_range	rating	address
0	The Buffet	Fast Food Restaurant	30.3270	78.0452	Ambrosia	30.3279	78.0453	1200.0	4.0	3.6	21, Centre Point, Near Astley Hall, Rajpur Roa...
1	Anandam	Indian Restaurant	30.3353	78.0532	Anandam	30.3351	78.0533	650.0	3.0	4.3	69, Krishna Tower, Rajpur Road, Hathibarkala S...
2	KFC	Fast Food Restaurant	30.3338	78.0515	KFC	30.3336	78.0516	450.0	2.0	4.3	65A, Rajpur Road, Hathibarkala Salwala, Dehradun

Drop the venues which have 0.0 rating as it's significant or not been rated yet.

In [15]:
```
selected_venues = selected_venues[selected_venues['rating'] != 0.0]
print("Total venues available: {}".format(selected_venues.shape[0]))
```

```
Total venues available: 29
```

Methodology

This project aims at identifying the venues in Dehradun based on their rating and average costs. This would enable any visitor to identify the venues he/she wants to visit based on their rating and cost preference.

As a first step, **the data from two APIs (Foursquare and Zomato)is retrieved**. Extract venue information from the center of Dehradun, upto a distance of 10 Km. The latitude and longitude values are then used to fetch venue rating and price from Zomato.
Secondly, **the data retrieved from the two APIs on the map is explored** and the top category types are identified. **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, **Analyze the data** that is created based on the ratings and price of each venue. **Identify places where cluster of venues are located** so that any visitor can go to one place and enjoy the option to choose amongst many venue options. 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, **cluster the venues using partitioning i.e. *K-means*** based on the available information of each venue. This will allow one to clearly identify which venues can be recommended and with what characteristics.

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

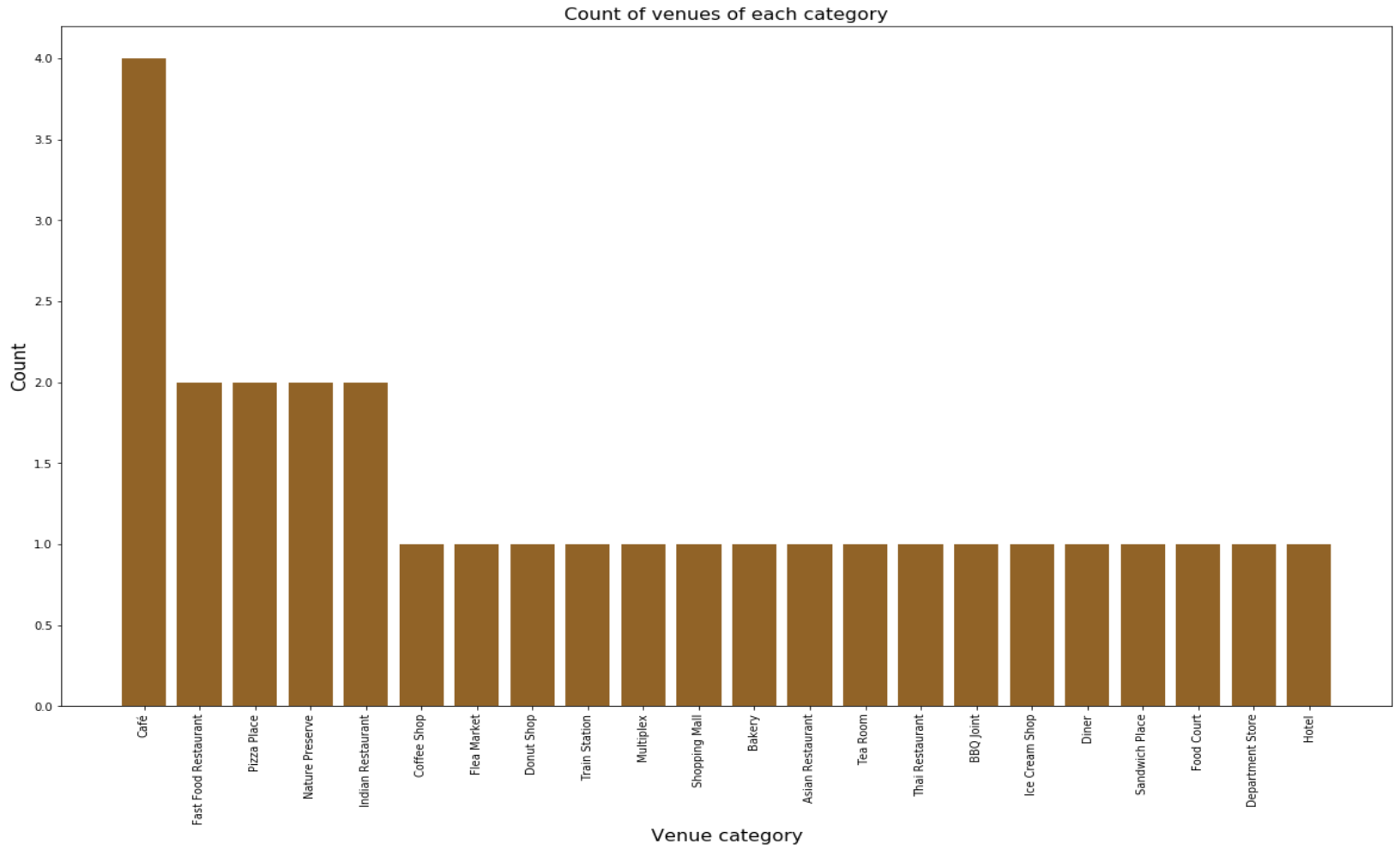
Analysis:

Out[14]:

	categories	venue	latitude	longitude	price_range	rating	address	average_price
0	Fast Food Restaurant	Ambrosia	30.3279	78.0453	4.0	3.6	21, Centre Point, Near Astley Hall, Rajpur Roa...	600.0
1	Indian Restaurant	Anandam	30.3351	78.0533	3.0	4.3	69, Krishna Tower, Rajpur Road, Hathibarkala S...	325.0
2	Fast Food Restaurant	KFC	30.3336	78.0516	2.0	4.3	65A, Rajpur Road, Hathibarkala Salwala, Dehradun	225.0
3	Bakery	Ellora Homeaids	30.3288	78.0463	2.0	3.7	25, Rajpur Road, Chukkuwala, Dehradun	150.0
4	Café	Barista	30.3269	78.0449	3.0	3.7	15 A, English Book Depot, Rajpur Road, Chukkuw...	250.0

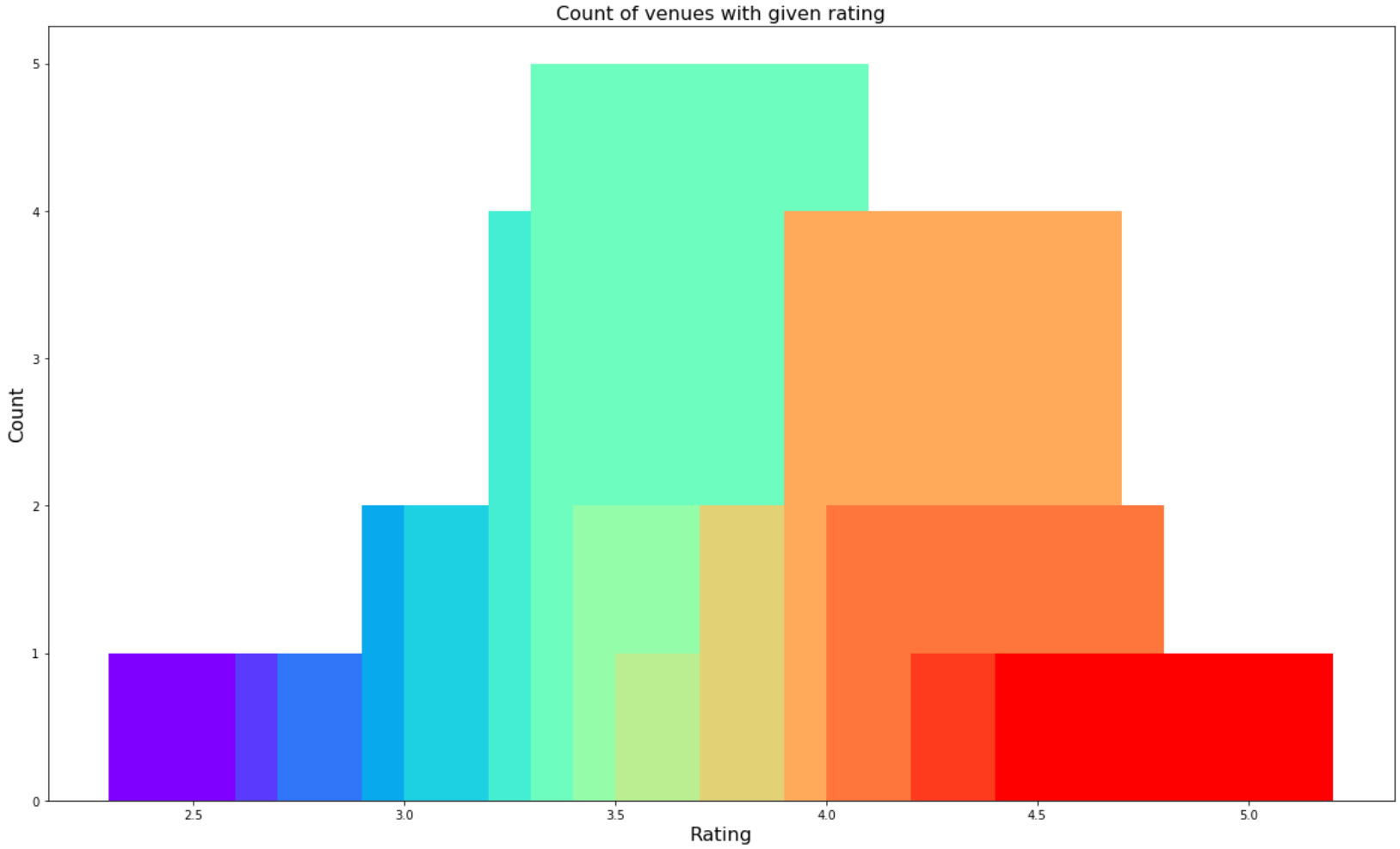
The complete dataset is now in its final form. **Inspect these venues based on their rating**. The rating of a venue are based on user reviews and belongs to a range from 1 to 5. Also **analyse the venues based on their price per person as well as the price range**.

Categories: There are various types of venues in the final dataset. Take a look at the venues and check which are the majority venue categories in the list.

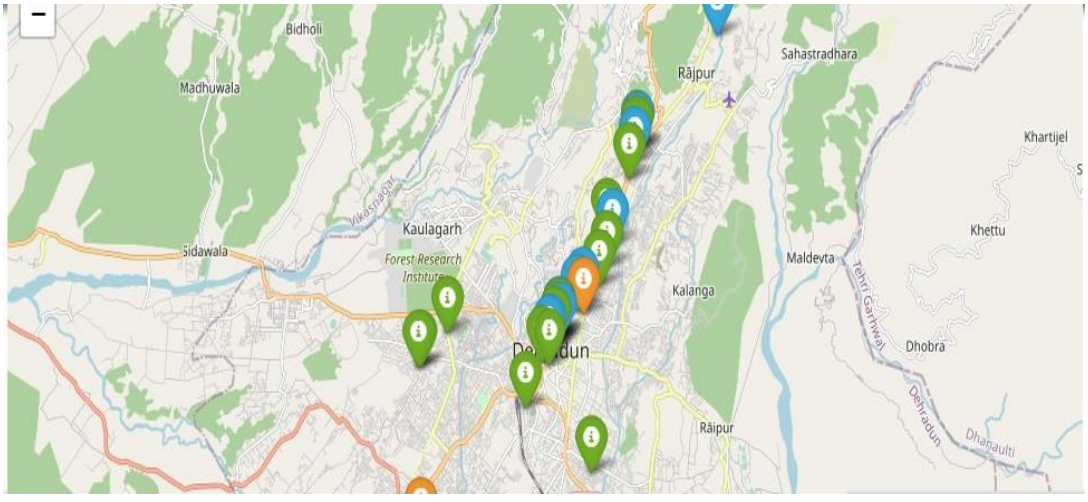


The majority venues are **Cafe, Fast Food Restaurants and Indian Restaurant**.

Rating: Rating of a venue is an important factor on which a visitor decides whether it is worth visiting the place. To cater to this, see what is the average rating for all the venues in the city. Next, plot the venues on the map and color code them.



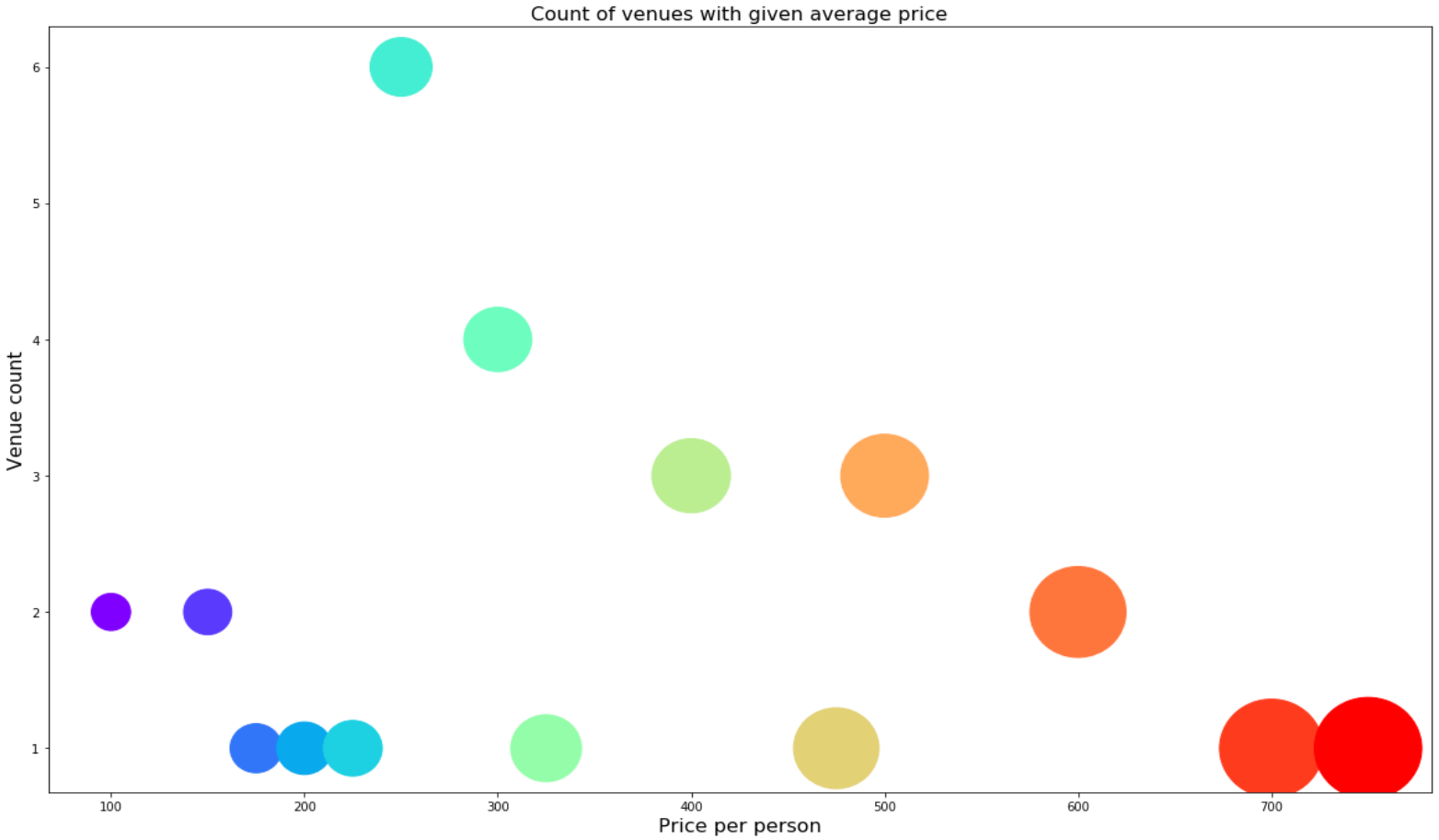
From the plot above, majority venues have their rating close to 4.



The map has the location of all the venues. It appears that many venues are located around **Rajpur Road, Chakrata Road and ISBT road** with rating above *3(green)*. A visitor should definitely check out these areas.

Price:
Take a look at the venues based on the price values. Having two price features for our venues, one is average_price which defines the average cost for one person and the other is price_range which determines the price range as defined by Zomato.

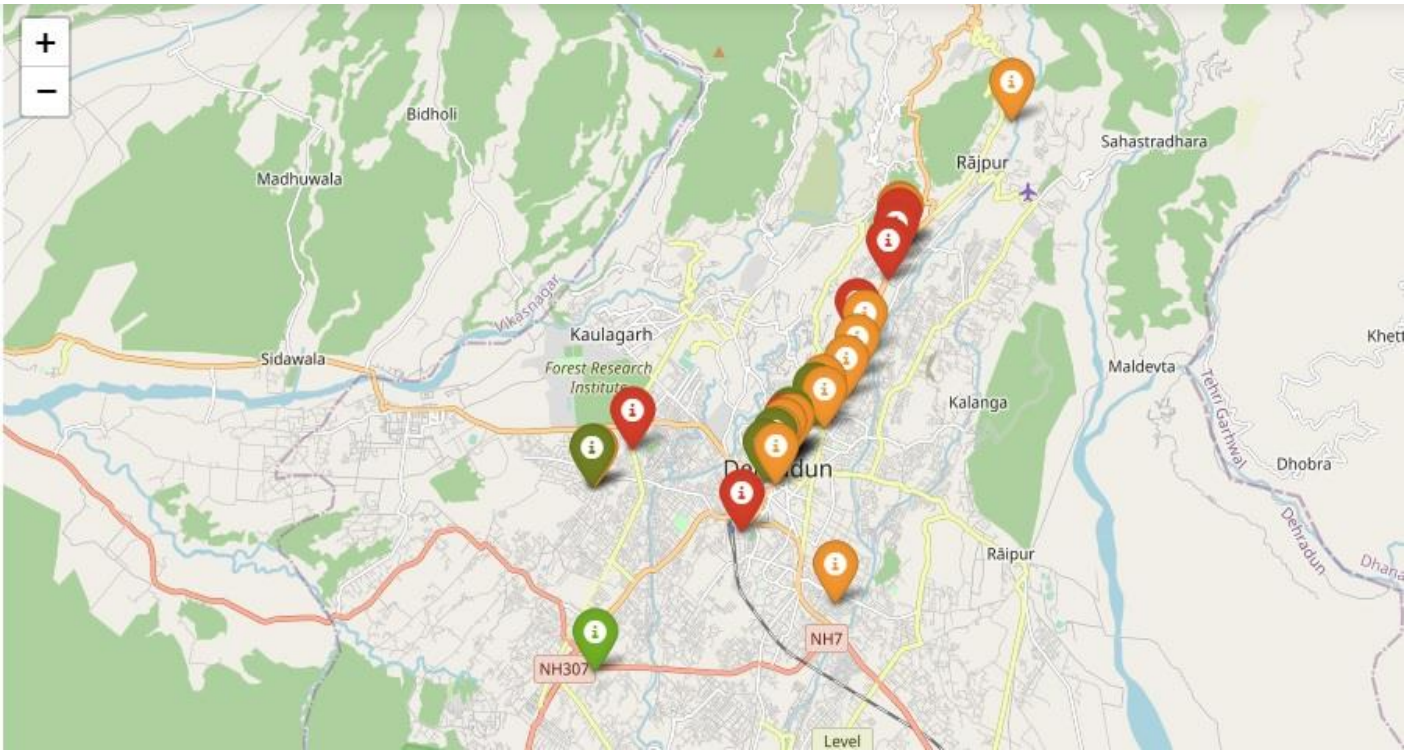
First explore the average_price using a scatter plot between the price and the count of venues with that average price. Size the points based on the price to highlight their price.



From the plot above, a large number of venues have an average price between **Rs 250 and Rs 500**.

Users might also be interested in going to a place that fits in their budget. Let's use the price_range column to plot the venues on a map and represent the venues with lower price in green and move towards red as the price increases.

```
In [21]: color_map = {'1.0': 'green', '2.0': 'darkgreen', '3.0': 'orange', '4.0': 'red'} #ordered as low to high price range
```



The venues near **Clock Tower, Chakrata Road and ISBT Road** are primarily lower priced. The venues at **upper Rajpur Road** has steep prices.

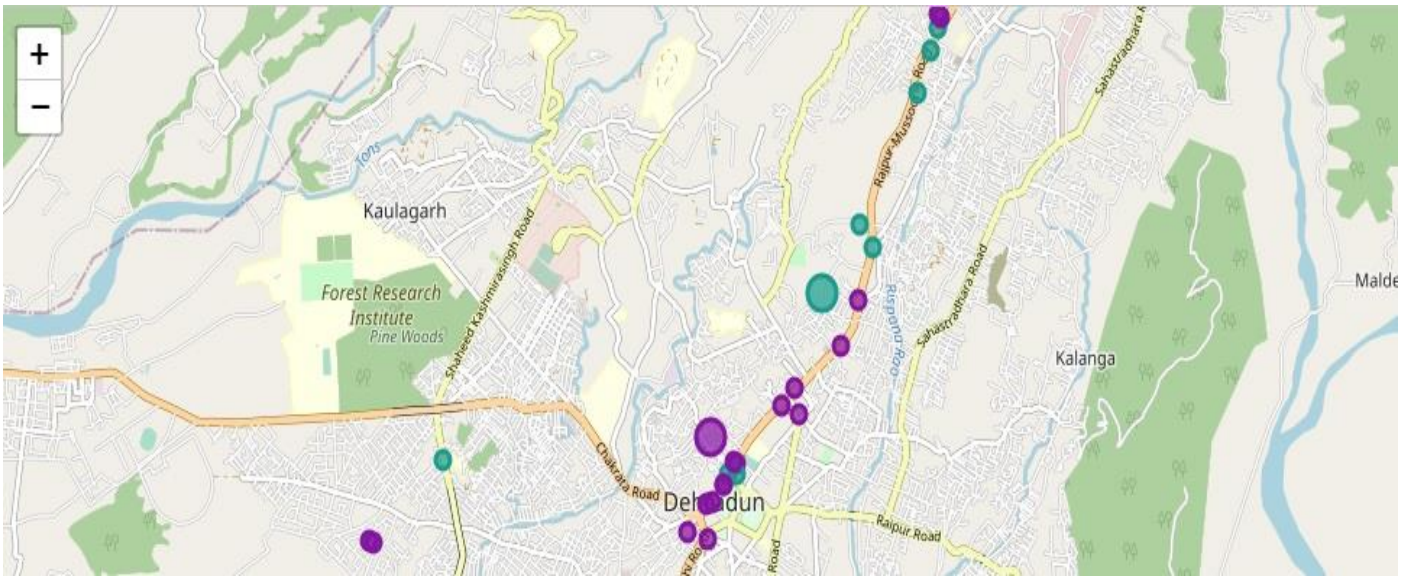
Clustering: Cluster all these venues based on their price range, location and more to identify similar venues and the relationship amongst them. Let's cluster the venues into two separate groups.

In [24]:
```
NO_OF_CLUSTERS = 2

clustering = selected_venues.drop(['venue', 'address', 'rating_bin', 'categories'], 1)
kMeans = KMeans(n_clusters = NO_OF_CLUSTERS, random_state = 0).fit(clustering)
selected_venues.insert(0, 'cluster_labels', kMeans.labels_)
selected_venues.head(36)
```

Out[24]:

	cluster_labels	categories	venue	latitude	longitude	price_range	rating	address	average_price	rating_bin
0	0	Fast Food Restaurant	Ambrosia	30.3279	78.0453	4.0	3.6	21, Centre Point, Near Astley Hall, Rajpur Roa...	600.0	Good
1	1	Indian Restaurant	Anandam	30.3351	78.0533	3.0	4.3	69, Krishna Tower, Rajpur Road, Hathibarkala S...	325.0	Very good
2	1	Fast Food Restaurant	KFC	30.3336	78.0516	2.0	4.3	65A, Rajpur Road, Hathibarkala Salwala, Dehradun	225.0	Very good



The two clusters from the map:

The first(0) cluster is very sparsely spread and has very less venues(mostly at **upper Rajpur Road**). The second(1) cluster is at the central part of the city(Supplied lat, long **@Clock tower**) and includes the majority venues.

Results & Discussions

Based on our analysis above, one can draw a number of conclusions that will be useful to aid any visitor visiting the city of Dehradun, India.

After collecting data from the Foursquare and Zomato APIs, a list of 37 different venues is acquired. However, not all venues from the two APIs were identical. Hence, they were to be inspected, 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 29.
It was infered that from the total set of venues, majority of them were Cafes, Fast Food joints and Indian Restaurants. A visitor who loves Cafes/Indian Restaurants would surely benefit from coming to Dehradun.

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 plotted on the map, it's implied that there are clusters of venues around Clock Tower and Rajpur Road. These clusters also have very high ratings (more than 3).

When the price values of each venue is analyzed, it's implied that many venues have prices which are in the range of Rs 250 to Rs 500 for one person. However, the variation in prices is

very large, given the complete range starts from Rs 75 and goes uptil Rs 750. On plotting the venues based on their price range on the map, it was discovered that venues located near Clock Tower are relatively priced lower than venues in Rajpur Road.

Finally, through clusters it is evident that there are many venues which are relatively lower priced but have an average rating of 3.81 i.e. Cluster1. On the other hand, there are few venues which are high priced and have average rating of 3.79.

If a visitor is looking for cheap places with relatively high rating, you should check near Clock Tower, Chakrata Road, ISBT. If a visitor is looking for the best places, with the highest rating but might also carry a high price tag, you should visit Rajpur Road. Any firm/individual 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).

Conclusions

The purpose of this project was to explore the places that a person visiting Dehradun could visit. 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: Clock Tower, Rajpur Road and Chakrata Road. Based on the visitor's venue rating and price requirements, he/she can choose amongst the three places.