Exploring Biryani Restaurants in Hyderabad

Tarun Goli

(11th May, 2020)

IBM Data Science Capstone Project

1. Introduction

1.1 Background

Hyderabad is the city of biryanis, kebabs, haleem and chowki dinners. It has been **designated** as a Unesco Creative City of Gastronomy. Food is an integral part of the culture in Hyderabad. The food industry in Hyderabad employs 12% of the city's population either directly or indirectly, according to a study. We can never run out of food stalls in this city.

Hyderabadi Biryani is a popular variety of Biryani. Sometimes the word biryani is synonymous with Hyderabad. There are many number of restaurants in Hyderabad which specialize in cooking the biryani. It is often one of the most common restaurants in any given locality.

With such vibrant food culture, when one is travelling to Hyderabad they wouldn't want to miss this famous dish. But it could be a very daunting task to decide from which restaurant they can experience this famous recipe. The price of biryani ranges from 500 to 3500 for two people.

This project provides a person with all the information he needs regarding the restaurant like it's price, the rating of the restaurant which are the two main factors one consider before making a decision.

1.2 Interested Audience

The majority of the audience who could be benefited by this project are those to plan to visit Hyderabad. This project entails all the details regarding the restaurant on a map which equips them with all the information they need to check for a good restaurant near where they are.

This project could be even beneficial to business official who come for some work and don't really have much time to explore the city. Using this project they can simply glance through the restaurants near him/her and select the one which best suits them the best.

2. Data

To explore the various venues in the city we need the neighborhoods in and near the city. Further we need the location data of each of the neighborhoods to explore the venues in that region.

I retrieved the neighborhoods details of Hyderabad from [Hyderabad Neighbourhood](). This website contains the name of the neighborhood and it's corresponding pincode in a tabular format.

Pgeocode:

Pgeocode is a python library for high performance off-line querying of GPS coordinates, region name and municipality name from postal codes. Distance between postal codes as well as general distance queries are also supported. The used GeoNames database includes postal codes of 83 countries. For India the postal code is('IN').

The result of geo-localization query is **pandas.DataFrame** which contains the following columns

- **Country_code**
- **Postal_code**
- **Place_name**
- **State_name**
- **State_code**
- **Country_name**
- **Country_code**
- **Community_name**
- **Community_code**
- **Latitude**
- **Longitude**
- **Accuracy**: provides the accuracy of the location with 1 being the least and 6 being the highest.

Foursquare API:

With the help of Foursquare API we explore the venues near a particular neighborhood. The foursquare API required certain queries to explore the venues. We need to provide queries:

- **Name of the neighborhood**
- **Latitude and Longitude information**
- **Limit** : Here we set the limit to 100 venues per neighborhood
- **Radius**: After careful trial and error I've decided to enter the radius as 6000 so that Foursquare API returns us maximum venues.
- **Version** : The day, month and year (20191105)
- **Search query** : 'biryani'

Foursquare API provides us with a unique client id and client secret keyword which we have to enter.

Foursquare API provides us 99500 standard API calls per day and upto 500 premium API call per day. Fetching the venues is a standard API call but fetching the venue details is a premium call.

The foursquare API returns information of the venues near the neighborhood in json format. From this information we extract the information we want.

- **Name** : Name of Venue
- **Latitude** : Latitude value of Venue
- **Longitude** : Longitude value of Venue
- **Venue Category** : The category of the venue
- **Venue ID**: The venue id in foursquare database.

Zomato API:

We will use the Zomato API to get the details of each venues. The venue details returned by Foursquare API contained a lot of missing values, so we use the help of Zomato API to resolve this issue.

Zomato API provides up to 1000 free API calls per day. We need to specify the following details

- **Latitude**
- **Longitude**
- **Start :** Fetch results to be returned
- **Count :** No of results to be returned
- **q:** search query
- **sort :** sort restaurants by

The Zomato API returns the information, from which I returned the following data

- **Name**
- **Price for two**
- **Price Range**
- **Latitude**
- **Longitude**
- **Address**
- **Rating**

3. Data Cleaning:

The hyd_Df contains the information regarding the neighborhoods and its corresponding pincode. Each of these neighborhood names contains "S.O" or "B.O" as its suffix. We have to remove this from all the rows in the dataset.

	Location	City	Pincode
0	A.Gs Office S.O	Hyderabad	500004
1	A.Gs. Staff Quarters S.O	Hyderabad	500045
2	Administrative Buildings S.O	Hyderabad	500007
3	Aliabad B.O	Hyderabad	500078
4	Allembylines S.O	Hyderabad	500087

After removing the suffixes the tables look like the below table:

	Location	City	Pincode
0	A.Gs Office	Hyderabad	500004
1	A.Gs. Staff Quarters	Hyderabad	500045
2	Administrative Buildings	Hyderabad	500007
3	Aliabad	Hyderabad	500078
4	Allembylines	Hyderabad	500087

The Pgeocode returns a dataframe containing postal code, country code, place name, state name, state code, community name, community code, latitude, longitude and accuracy.

On exploring the dataset for missing values, I found that row with index no 66 is containing NaN values. So I dropped that row from that dataset. After dropping the dataset I visualized the locations of each neighborhood using **folium** library.
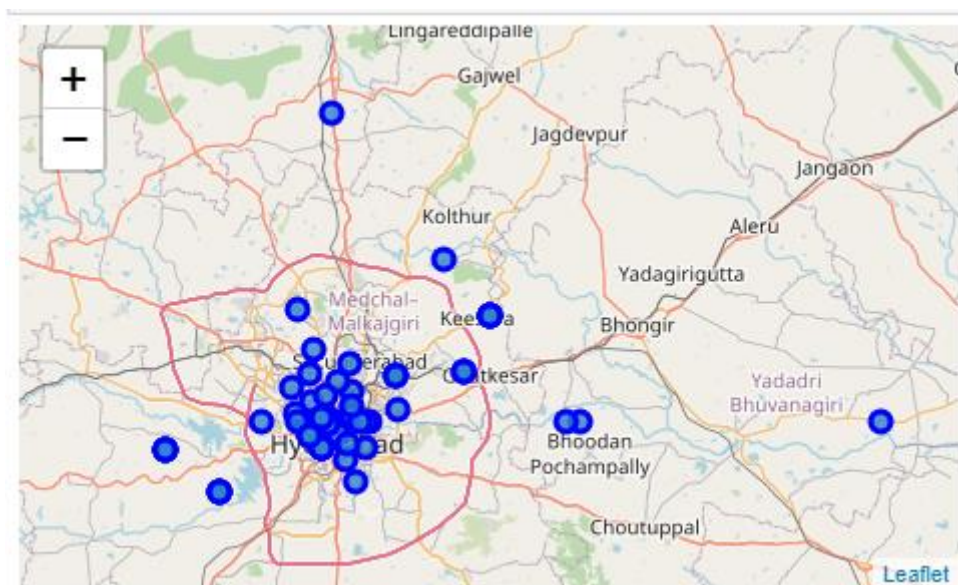

We can see that few neighborhoods are far from the city. So we will find the index numbers of each of those neighborhoods and drop them.

On further exploring the dataset I made an observation that many of the neighborhoods have a location accuracy of 1. With such low accuracy the information can be misleading, so I dropped all the neighborhoods which

contained location accuracy 1. The remaining neighborhoods have a location accuracy of either 3 or 6.

This is the map after dropping all the unnecessary neighborhoods.
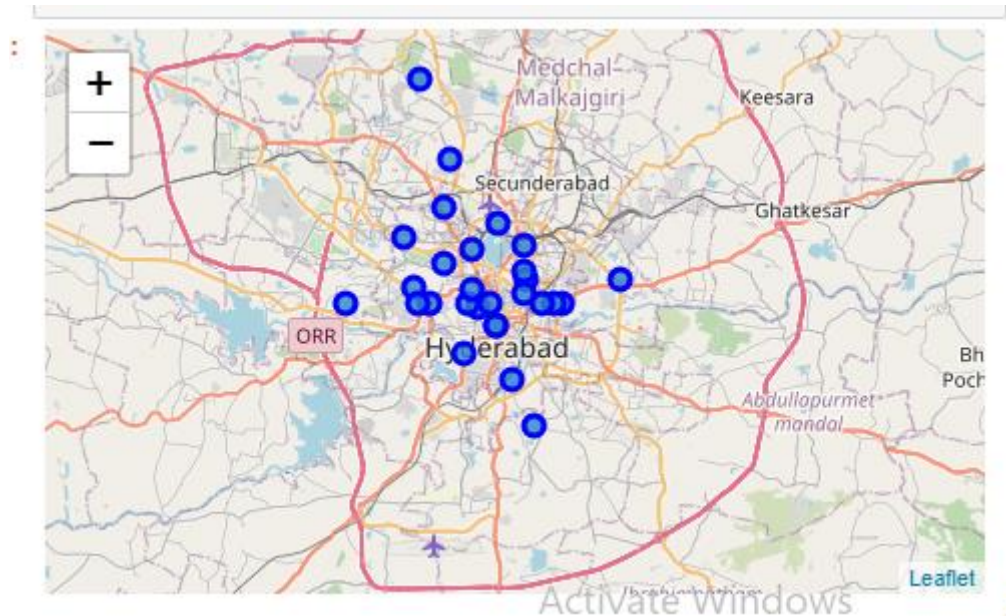
Now using the Foursqaure API we will find out the venues near a given neighborhood. After careful trail and error, I've given the radius as 6000 so as to get maximum venues. The format for api request is :

https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={} &v={}&ll={},{}&radius={}&limit={}&query={}.

Given below is table of the venues returned by the Foursquare API.

	Neighbourhood	Neighbourhood_Latitude	Neighbourhood_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category	Venue_Id
0	Bazarghat (Hyderabad), A.Gs Office, Vidhan Sab...	17.3872	78.4621	Cafe Bahar	17.399595	78.478566	Indian Restaurant	4bdebd29be5120a17e98fe70
1	Bazarghat (Hyderabad), A.Gs Office, Vidhan Sab...	17.3872	78.4621	Shadaab	17.368661	78.475572	Diner	4c642d01eb82d13a1fa809d6
2	Bazarghat (Hyderabad), A.Gs Office, Vidhan Sab...	17.3872	78.4621	Santosh Dhaba	17.388485	78.479509	Indian Restaurant	4d3d4eca14aa8cfaa6d6b15e
3	Bazarghat (Hyderabad), A.Gs Office, Vidhan Sab...	17.3872	78.4621	Sarvi	17.412698	78.449704	Indian Restaurant	4cf0e66088de3704a3c96e2b
4	Bazarghat (Hyderabad), A.Gs Office, Vidhan Sab...	17.3872	78.4621	Paradise Food Court	17.412113	78.465443	Indian Restaurant	5048b6d5e4b08065929b6eb1

Now we have to find the venue details of each of these venues in the biryani_venues dataset. Making an API call using Foursquare to find the venue details is considered as a Premium API call and Foursquare allows us to make upto 500 premium api calls. The Foursquare returns the venue details of all the venues in the dataset. But on exploring the dataset we find more than half of the venues miss information regarding the venue's price or its rating which are both

an important parameter when we cluster the venues. Since there are many missing values we will use Zomato API to get the information regarding the venue details.

Zomato API allows us to make upto 1000 API calls per day.

The format for Zomato API request is:

https://api.foursquare.com/v2/venues/{}?&client_id={}&client_secret={}&v={}

We store the results in zomato_venues dataset.

On exploring the dataset we find that few of the rows in the dataset do not contain the ratings. Row no 119 is completely null so we drop the dataset.

On further exploration we find that few rows have price mentioned in their section but does not contain the rating section empty. So we will try to retrieve the rating from the biryani_venues which might contain the rating for these venues.

	Neighbourhood	Neighbourhood_Latitude	Neighbourhood_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category	Venue_Id	Price	Rating
6	Bazarghat (Hyderabad), A.Gs Office, Vidhan Sab...	17.3872	78.462100	Nizam club	17.403221	78.468729	Lounge	4c08f54f6071a5934901dd32	3	8.4
12	Bazarghat (Hyderabad), A.Gs Office, Vidhan Sab...	17.3872	78.462100	Grand Hotel	17.387760	78.477577	Indian Restaurant	53cf6be6498e3ae4a3273bda	2	7.1
31	Bazarghat (Hyderabad), A.Gs Office, Vidhan Sab...	17.3872	78.462100	Hyderabad House	17.402775	78.456992	Indian Restaurant	4ce39915a3c6721edf4411cf	2	no rating
78	Nallakunta, Ambernagar, Vidyanagar (Hyderabad)...	17.3753	78.474400	Hotel Grand	17.387933	78.476901	Hotel	4c34beaba0ced13ab0ae186e	NaN	NaN
88	Malkajgiri, Anandbagh	17.4284	78.493600	Persis Gold	17.441833	78.487259	Restaurant	510380afe4b026986bf9764e	NaN	NaN
114	Nanakramguda, Lunger House, Sakkubai Nagar, Go...	17.3898	78.370733	Rendezvous	17.414088	78.328712	Bar	554505c3498e1e2d82e6dddf	NaN	NaN
141	Jubilee Hills, Dr.B R Ambedkar O.U	17.4334	78.411100	Lemon n Spice	17.448799	78.357131	Indian Restaurant	4e62636b8877954de82cb2e7	NaN	NaN

These are the following rows in biryani_venues dataset. We can observe that only two rows among them contain the rating section and the rating scale is in 10. We try to divide the rating by 2 to bring it down to a 5 scale and enter there values to the zomato_venues dataset.

	Venue	Latitude	Longitude	Price_for_two	Price_Range	Rating	Address
6	Cosmos99 Club And Lounge	17.4075841690	78.4689759091	800.0	2.0	4.2	Ground Floor, Amtrita Castle, Ambedkar Colony,...
12	Season The Restaurant	17.3880906797	78.4777437150	200.0	1.0	3.5	Hotel Mandakini, Jaya International, Hanuman T...

The zomato dataset does not contain information regarding the neighbourhood and venues category. So we try to include that data from biryani_venues. On including the neighborhood data the final table looks like.

	Venue	Latitude	Longitude	Price_for_two	Price_Range	Rating	Address	Neighbourhood	Venue Category
156	Noodles And More	17.486643	78.393118	150.0	1.0	3.2	LIG 57/1 Flats, KPHB, Near Ramya Grounds, Kuka...	Vivekanandanagar Colony, Kukatpally	Indian Restaurant
157	Hotel Swagath	17.493272	78.403806	700.0	2.0	3.4	1st Floor, Alluri Trade Center, KPHB Main Road...	Vivekanandanagar Colony, Kukatpally	Indian Restaurant
158	Delicacy Family Restaurant	17.498365	78.388470	700.0	2.0	3.3	32/1, Sardar Patel Road, X Road, Near Indian P...	Vivekanandanagar Colony, Kukatpally	Restaurant
159	Coffee Tree Cafe Gourmet	17.490086	78.393225	700.0	2.0	2.7	Manjeera Trinity Mall, 3rd Floor, Food Court, ...	Vivekanandanagar Colony, Kukatpally	North Indian Restaurant
160	Real Bawarchi	17.401434	78.573229	700.0	2.0	3.0	Opposite IBP Petrol Pump, Uppal Bus Depot, Upp...	Peerzadiguda, Uppal (K.V.Rangareddy)	Indian Restaurant

4. Methodology

**4.1 Exploratory Data Analysis**

In this topic we will be analyzing various features of the dataset.

We have few question which need to be answered

1) What is the price range corresponding to the price tier?

We will use a scatter plot to find a relation between price and price range.

```
sns.scatterplot(x=zomato_venues1['Price_Range'],y=zomato_venues1['Price_for_two'])
plt.title('Relationship between Price Range and Price for two');
```
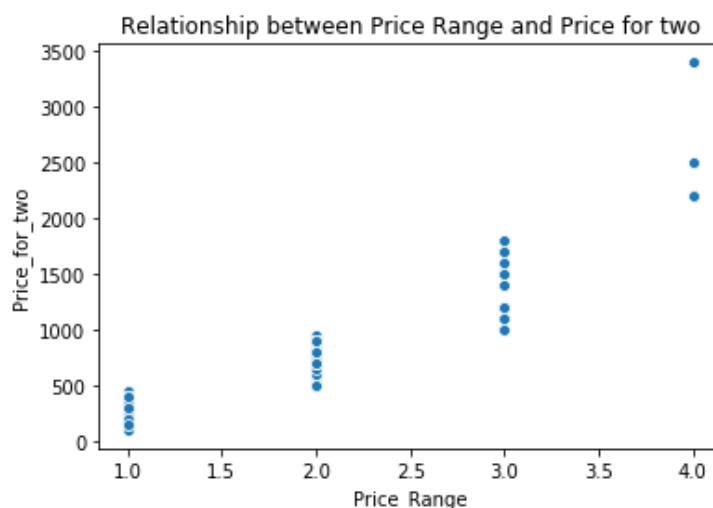

From the above scatter plot we can infer the relationship between Price_for_two and price range.

Price range 1 : Price for two is less than 500 Rs.

Price range 2: Price for two is in between 500 and 1000 Rs.

Price range 3: Price for two is in between 1000 and 2000 Rs.

Price range 4: Price for two is above 2000 Rs.

2) Is there any relationship between Price and Rating?

We want to find if there is any correlation between price and rating. For example is the price increases with increase in rating.

```
sns.scatterplot(x=zomato_venues1['Price_for_two'],y=zomato_venues1['Rating'])
plt.title('Realtionship between Price for two and Rating');
```
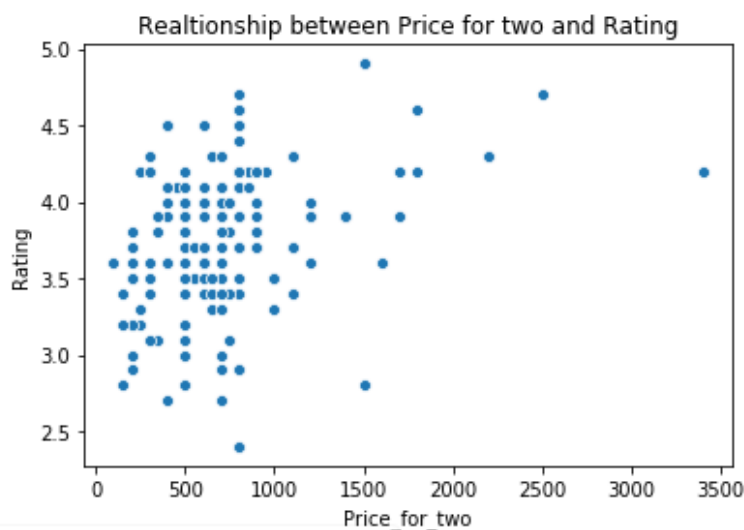

What we can observe from this plot is that there is no significant relationship but from what we can observe there is slight increase in price as the rating increased. Since very few restaurants are there above price 1000. We cannot say if a significant relationship exists. On applying correlation function we find that the correlation between price for two and rating is 0.3246.

3) Which neighborhood has highest average price for two?

By answering this question we can find out the neighborhoods which have their average price for two greater than remaining neighborhoods.

```
sns.barplot(y=neigh.index,x=neigh['Price_for_two'])
plt.title('Relationship between neighborhood and price0');
```
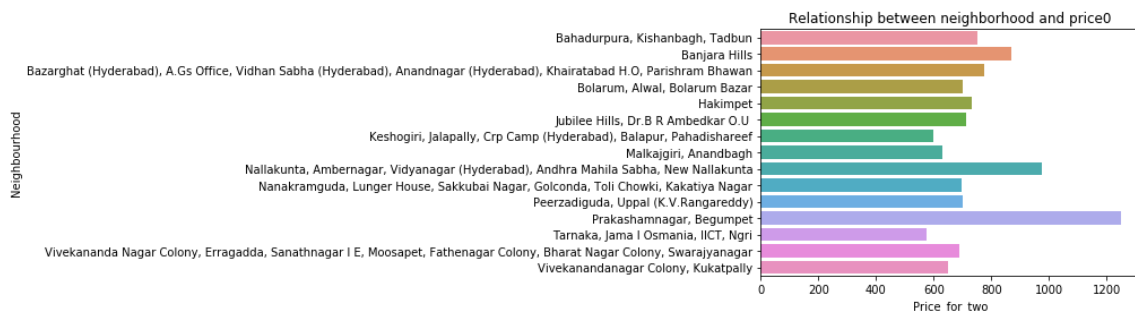

We can observe from the graph that neighborhoods of Begumpet and Prakashamnagar have highest average price for two of 1300 Rs.

4) Which neighborhood has highest average rating?

By answering this question we can find out the neighborhood which have the best rating among other neighborhoods.

```
: sns.barplot(y=neigh.index,x=neigh['Rating'])
  plt.title('Relationship between Average Rating and Neighborhood');
```
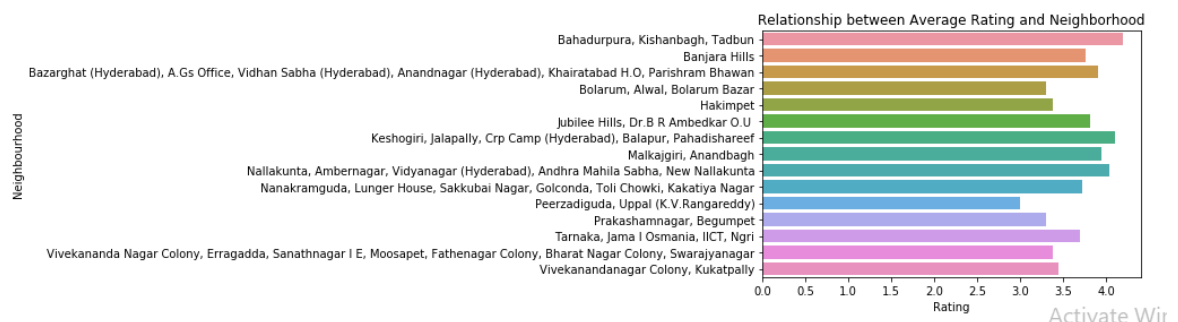

From the above graph we can observe that Bhadurpura has the highest average rating of 4.25.

5) Which Venue Category has the highest average price for two?

By answering this question we can say in general if any of the venues categories have highest average price for two. We can even find out which venue category offer lesser price for two people.
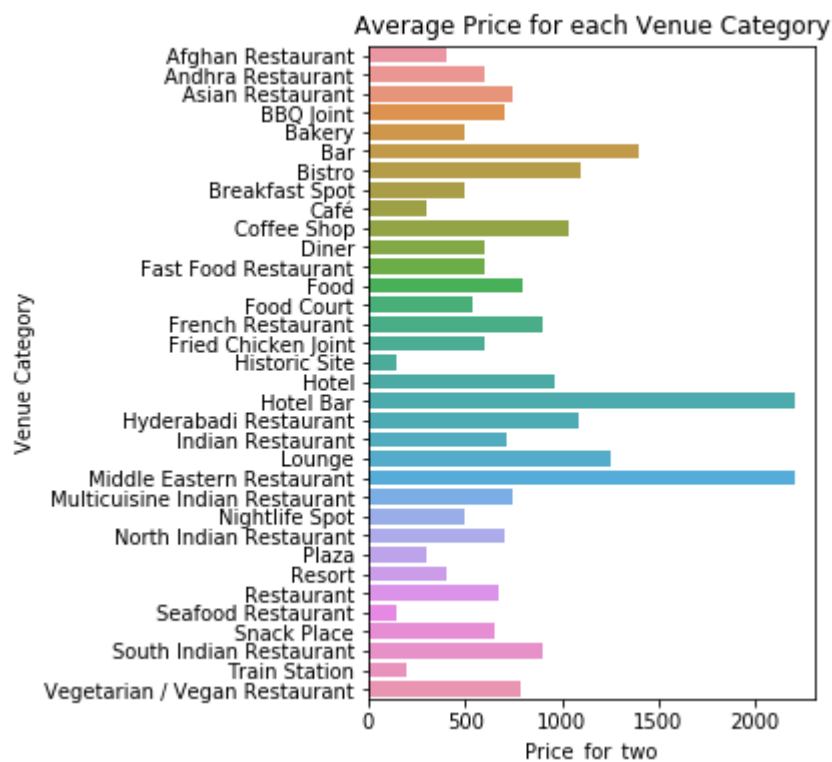
From the above plot we can understand that Hotel bar and Middle Eastern Restaurants have highest average price for two people of 2400 Rs while, Seafood Restaurant has lowest average price for two of 200 Rs. The Indian restaurants venue category which has more number of restaurants in the city has average price for two of 800 Rs.

4.2 Clustering

In this section we will be clustering the dataset based on 3 parameters those are Price for two, Price Range and Rating. Since Price Range is a categorical type we perform One Hot Encoding to the Price Range Column and save it in a dataset called venues_onehot.

From the zomato_venues1 dataset we add the Venues, Price for two, Rating column to the venues_onehot dataset.

The final dataset of venues_one hot looks like:

|	Venue	Price_for_two	Rating	1.0	2.0	3.0	4.0
0	Cafe Bahar	400.0	4.5	1	0	0	0
1	Hotel Shadab	900.0	4.2	0	1	0	0
2	Santosh Dhaba Exclusive	500.0	4.2	0	1	0	0
3	Sarvi Restaurant	850.0	4.2	0	1	0	0
4	Paradise Biryani	950.0	4.2	0	1	0	0

Now we have to find the K value to perform K-Means.

To find the optimal K value we use the **ELBOW METHOD.**

Elbow Method

We plot a graph with different K values and WCC values. Wherever there is a steep decline we can consider that as an optimum K value.
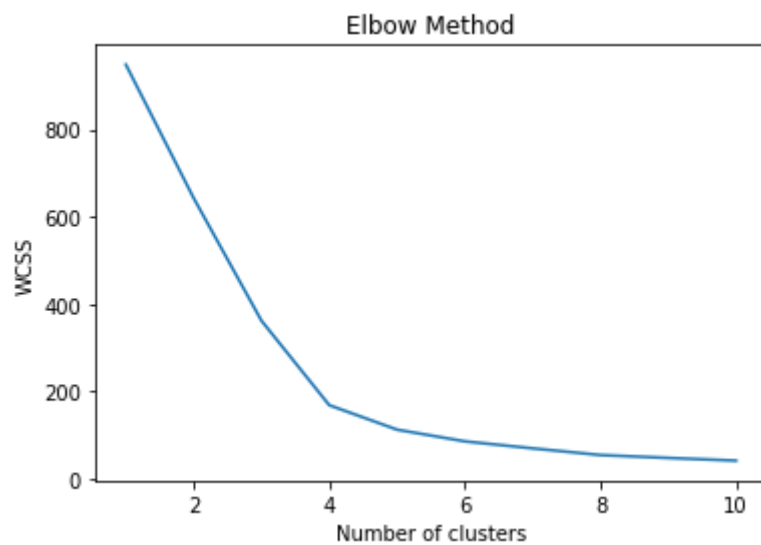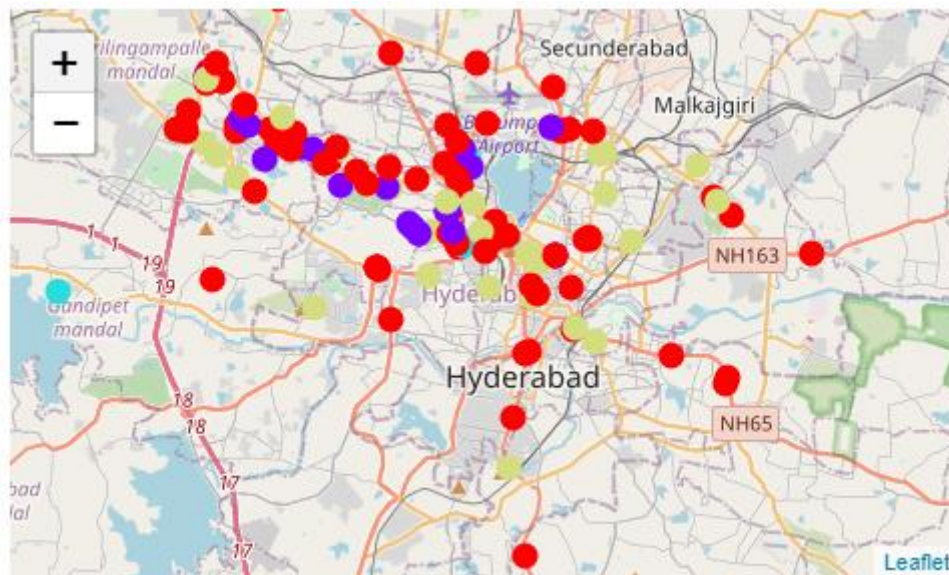

From the above plot we can see that there is a steep decline when the value of K is 4. So our optimum K value is 4.

K Means Clustering

Now that we have the K value we fit and transform the data.

"K_dataSet=StandardScaler().fit_transform(k_means)"

Later we plot the venues on a folium map.

Let us classify the clusters

Cluster 0: (Red Colour)

Venues for Cluster 0 have Average Price of 695.74 and Rating around 3.77 and Price Range 2.00

Cluster 1: (Purple Colour)

Venues for Cluster 1 have Average Price of 1330.43 and Rating around 3.91 and Price Range 3.00

Cluster 2: (Light Blue Colour)

Venues for Cluster 2 have Average Price of 2575.00 and Rating around 4.38 and Price Range 4.00

Cluster 3: (Gold Colour)

Venues for Cluster 3 have Average Price of 289.19 and Rating around 3.64 and Price Range 1.00

5. Result

We have performed a number of analysis on the Biryani Venues in Hyderabad. This project will be definitely helpful to anyone who is visiting Hyderabad.

We have collected the neighborhoods of Hyderabad from a website and using those neighborhoods pincode we were able to find the latitude and longitudes of those neighborhoods using PGeocode Library.

We used the help of Foursqaure API to get the various biryani venues in that neighborhood. We even got the venue details using Foursqaure API but since most of the data is missing value, we used the help of Zomato API. There are a total of 158 venues in our dataset. From what we have obsereved

1) Majority of the venues have their price below 1000 Rs. This is a very reasonble cost and any Indian middle class people can afford to eat it. Further not is the price less for majority of the venues, most of restaurant have an average rating of 3.5 which is very decent for the price.

2) In any given location there are hotels with most of the clusters which gives the visitor more options to choose from his budget.

3) Neighborhoods towards the east and south of city has venues which provide biryani for a much cheaper price, while area towards the northa and west have all kinds of price ranges.

4) The venues are spread pretty much everywhere in the city so it shouldn't be a problem for the visitor.

5) Dwaraka tiffins provides cheapest biryani for just 100 Rs for two people and it has a rating of 3.6 while Kebabs & Kurries - ITC Kakatiya provides that costliest biryani at 3400 for two people with a rating of 4.2.

6) Venues in Begumpet and Prakashamnagar neighgorhood are costliest whicle kukatpally has lowest biryani prices.

6. Conclusion

This project provides in detail information regarding the price and ratings of various venues ranging from the cheapest venue to the costliest venue, which gives the visitor a wide range of option to choose from. It makes the work easier for business people who don't have much time to explore the city, to choose a venue which helps them experience the Hyderabadi Biryani.