CAPSTONE PROJECT - COURSERA

EXPLORING HOTELS IN HYDERABAD, INDIA

* INTRODUCTION / BUSINESS PROBLEM

The main moto of this project is to identify the hotels in Hyderabad and cluster them on the basic of their ratings and price-per-head extracted using Zomato API.

Whenever a new person visits a city, they search for good place and affordable place to stay and eat. In this project I have clustered hotels using their ratings (to measure how good they are) and price per head (to measure how affordable it is). This will give you map with different mark colors to clustered hotels.

* DATA

Here we have fetched data from two different API’s

**Foursquare API:** Here we have fetched data of hotels in Hyderabad.

From foursquare API we got following columns of data which we have to clean

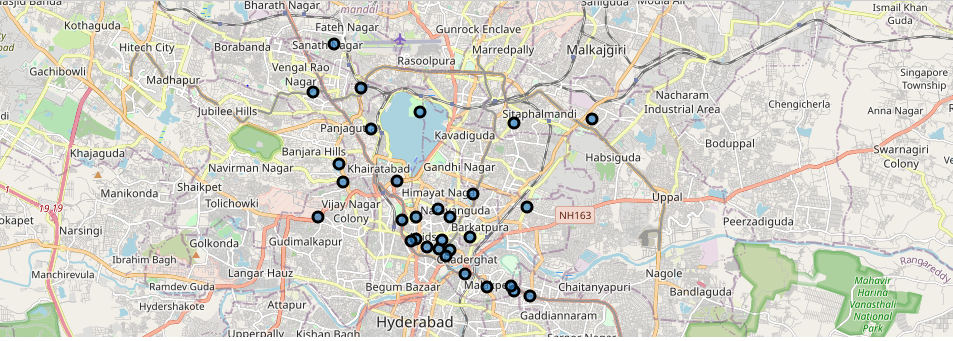
'categories', 'hasPerk', 'id', 'location.address', 'location.cc','location.city', 'location.country', 'location.crossStreet','location.distance', 'location.formattedAddress','location.labeledLatLngs', 'location.lat', 'location.lng','location.postalCode', 'location.state', 'name', 'referralId'

**Zomato API:** From this we fetch useful information like ratings, price etc.

* DATA CLEANING:
  + **FOURSQUARE API DATA:**

We don’t require every data they provide so I dropped some columns I used the below columns

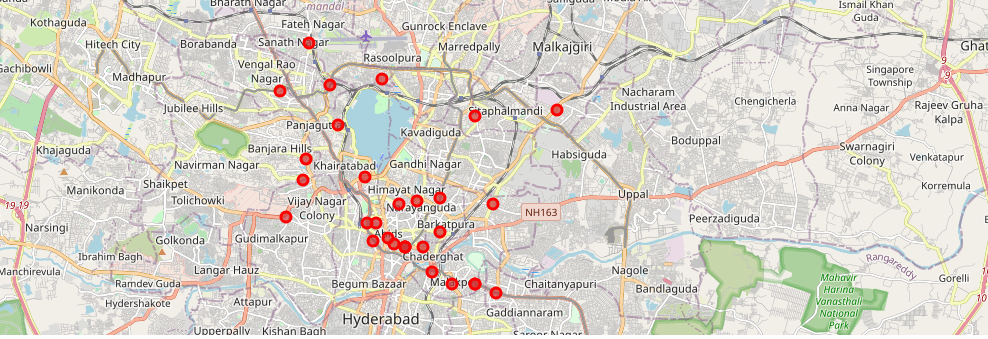
| **name** | **categories** | **lat** | **lng** | **categories** |
| --- | --- | --- | --- | --- |

The above image gives you the data of the foursquare API plotted in Hyderabad map

* + **Zomato API data:**

We don’t require whole data they provide, so I took only some particular columns only, they are

| **venue** | **latitude** | **longitude** | **price\_for\_two** | **price\_range** | **rating** | **address** |
| --- | --- | --- | --- | --- | --- | --- |



The above image gives you the data of Zomato API plotted in Hyderabad map

* + Now we merge the both datasets to get a new dataset with more information

We also calculate 2 new columns that is difference of latitudes and longitudes of foursquare data to Zomato dataset to check how accurate the both venues are.

* + Now we only keep the venues which has less difference of latitudes (of foursquare and zomato ) and longitudes so that the venues we get from foursquare API matches with Zomato API
  + Now looking into data, we can make some analyses that some venues aren’t matched in foursquare and zomato
  + So, we manually delete that type of data
  + Now after cleaning the data , now we have 22 venues with all proper details
* METHODOLOGY AND EDA:
* Now from merged and cleaned data we select some particular columns for our further work, we select these columns

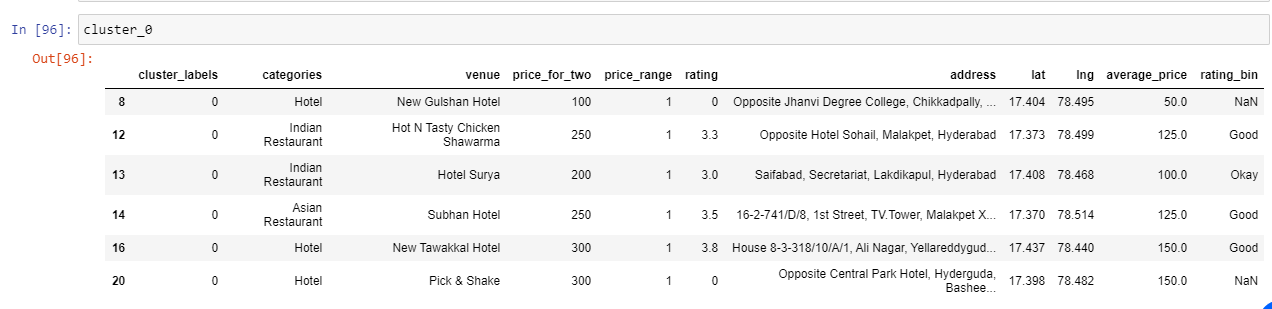
"categories","venue","price\_for\_two","price\_range","rating","address",'lat','lng'

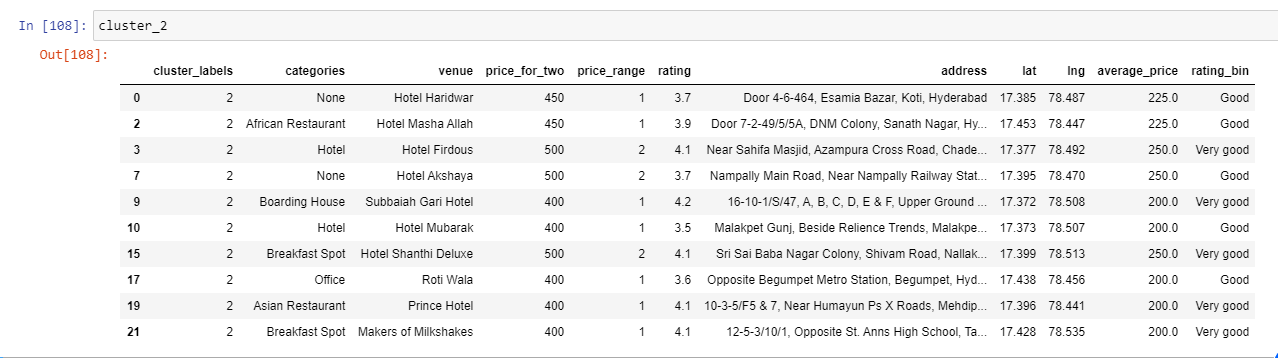
* To this data we add one more column that is average price that is price per head
* Also we add one more column that tells that the hotel is low or okay or good or very good using the ratings column using bins
* Now we have our required dataset with which we can do so many things

The new, merged , cleaned data contains :

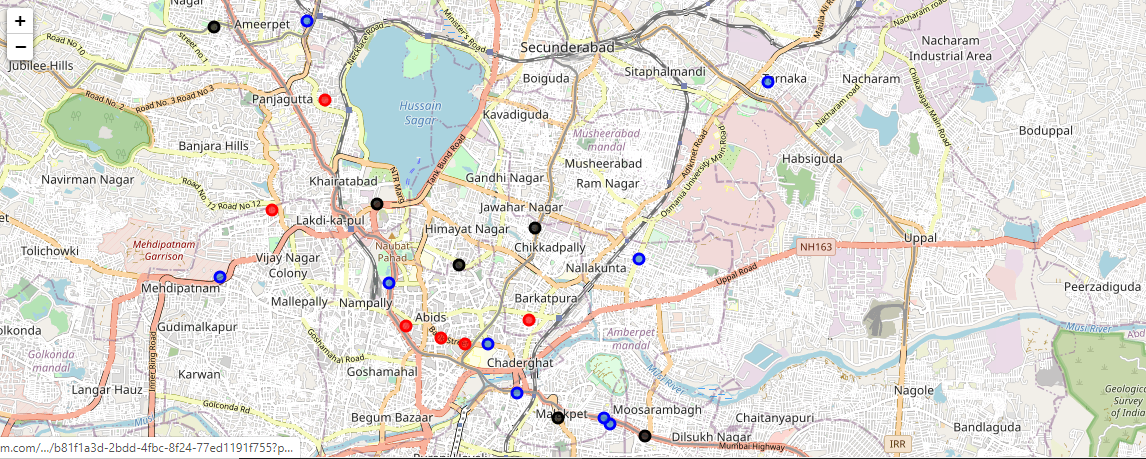
| **categories** | **venue** | **price\_for\_two** | **price\_range** | **rating** | **address** | **lat** | **lng** | **average\_price** | **rating\_bin** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |

* MACHINE LEARNING:
* I have implemented k means clustering using ratings and average price so that I can cluster venues on the basis of ratings and price which takes care of venue being good and affordable
* All the venues are divided into 3 clusters:
* The 3 clusters are :





* RESULTS AND DISCUSSION:



This is the graph of hospitals we have clustered on the basic of their ratings and average price.

Black color marked venues are with low average price

Red color marked venues are with high average price

Blue color marked venues are with medium average price.

**Further Improvisations:** Instead of giving the Hyderabad location details, we can use geocoder to find the live location and calculate the hospitals nearby, this was this application is not restricted to Hyderabad and can be used anywhere in this world

* CONCLUSION

So the whole process was to collect data from both API’s, merge them and clean data just by keeping the required data. After cleaning, clustering the data such that we cluster the similar hotels on the basis of their ratings and average price. This application will help people who are in Hyderabad and want to go to a hotel and who wants to check whether the hotel is good and affordable or not.