CAPSTONE PROJECT - COURSERA

EXPLORING HOSPITALS IN HYDERABAD , INDIA

* Introduction /Business Problem :

Nowadays hospitals have become one of the important parts of our society , we never know when we get a medical emergency and need to attend a hospital. The aim of my project is to give the basic idea of hospitals in Hyderabad,India. I have used the foursquare API to get data of hospitals in hyderabad . You can choose the best hospitals in your surrounding and with required ratings. When a person searches for a hospital in their locality they search for a best hospital.This project will help you to detect the best hospital in your surroundings using the ratings given by many people which visited that hospital in google. It further provides the contact details,address of the particular hospital you want to access.

You can use the output data from this whole project for further analysis .

* Data :

Firstly , we’ll look at Hyderabad on a map using the folium library. After that I have used the **foursquare API** to extract basic data of hospitals in hyderabad . **Foursquare API :** we’ll get hospital data within range from the center of hyderabad( with longitude and latitude given )I have used ‘search’ keyword for the purpose of getting data , **Dataset :** I couldn't find a appropriate data I require for my project on the internet , so I made my own dataset from scratch . As I made my own dataset I got freedom to collect only required data for my project so no need of data cleaning again.I have formatted the data required in excel sheet and then imported the sheet into notebook. I have collected the data for hospitals that are in API data only. I have collected ratings ,contact numbers , and addresses of hospitals from google . Ratings are given by many people who had experience in that hospital .

* **Data from FourSquare API includes :**

'categories', 'hasPerk', 'id', 'location.address', 'location.cc',

'location.city', 'location.country', 'location.crossStreet',

'location.distance', 'location.formattedAddress',

'location.labeledLatLngs', 'location.lat', 'location.lng',

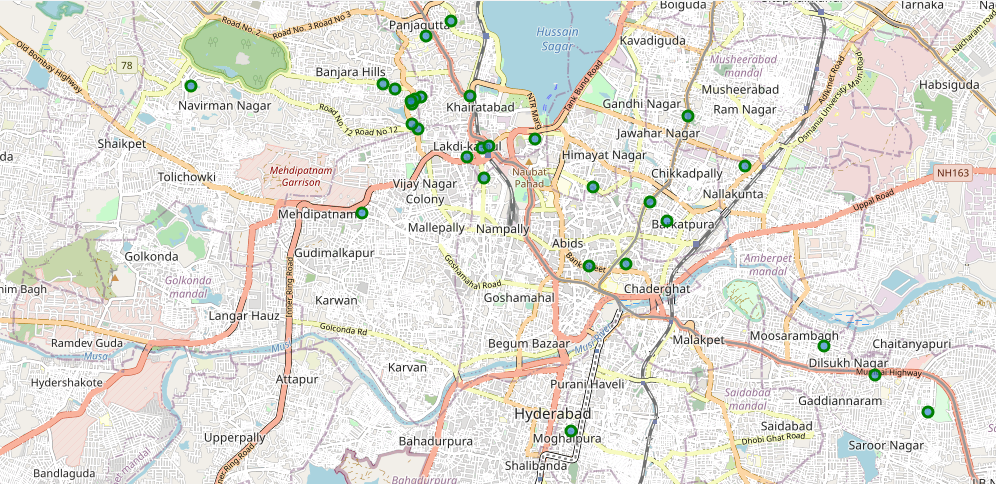
'location.postalCode', 'location.state', 'name', 'referralId'

* **Dataset includes ( I collected them from scratch ):**

‘Name’ , ’Address’ , ’ratings’ , ’contact details’

* Data Cleaning :

### Plotting Venues retrieved from foursquare API using Folium :



I got 30 venues in total from APIs .

There are some unnecessary columns in the data , so dropped some columns in foursquare API data .

* Columns I have dropped :

'hasPerk','location.labeledLatLngs','location.distance','location.cc','location.city','location.state','location.country','location.postalCode','location.crossStreet','referralId','categories',’id’

I had to rename some column names for further easy to work on data

* Columns renaming:

' Location.address ' : ' address ' , ' location.formattedAddress ' : ' formattedaddress '

' location.lat ' : ' lat ' , ' location.lng ' : ' lng '

* Dealing with Address column ‘NaN’ values :

There are some NaN values in the Address column which is an important key factor for our data .So replace all the NaN valued address columns with values in ‘formatted\_address’ column.

* Deleting some rows in venues dataset :

I observed that there are some extra rows which do not have proper information and some hospitals are repeated as their departments which aren't required .

so deleted rows from the dataset.

After all this cleaning ,got a dataset which can be further used .I collected and created the data set for hospitals I got after removing and cleaning data

* Merging the API data and Data set

I have merged the data using ‘ Name of the hospital ‘ as keyword

* This new data set includes :

‘id', 'name', 'lat', 'lng', 'address\_x', 'formattedaddress',

'address\_y', 'ratings', 'contact details'

* I have dropped some columns in this data asI don’t require them and they are

'Address\_y', ' formattedaddress '

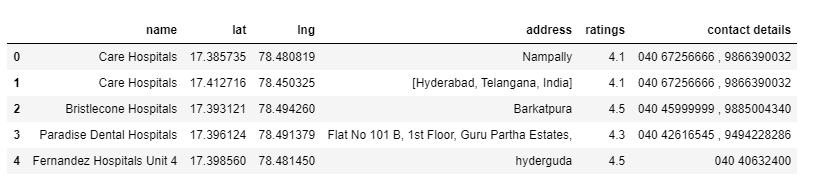
Renamed ‘Address\_x ‘ as ‘address’

* Final data aggregated :

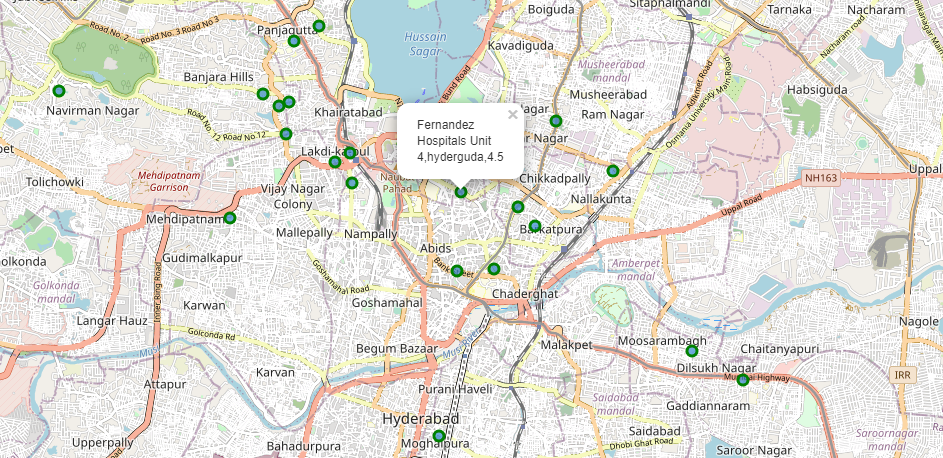
Final data contains 22 rows and 7 columns

It means data of 22 hospitals .

* The final data looks like this (first five) :



Plotting of final data :



If you put a cursor on a particular point you can see the hospital name , address , rating as shown in above plot.

* Methodology and Exploitary Analysis :

We know that the dataset we received after data cleaning is so small ,it has only 22 rows and 7 columns

Ratings is the only keyword we can work on

* After analysing the data , I got results as :

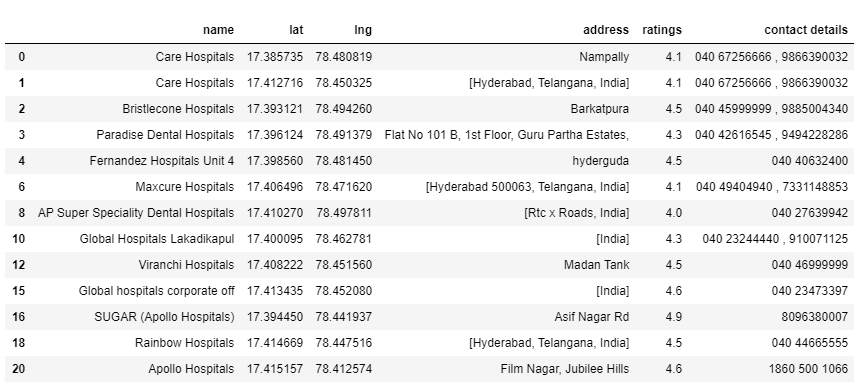
There are 13 hospitals with ratings >= 4

8 hospitals with ratings >=3 and < 4

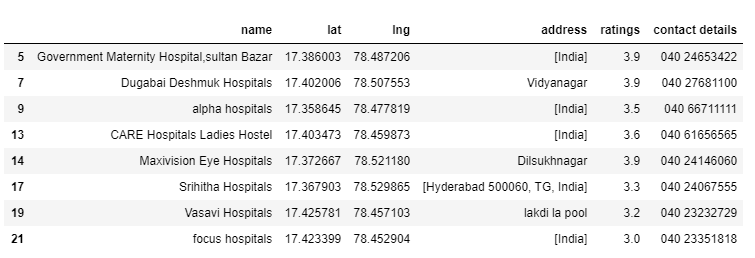
1 hospital with ratings < 2

I have three more tables with above division for further use

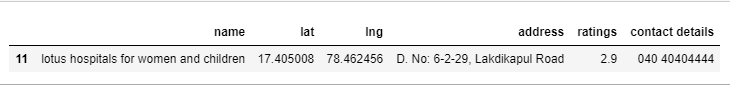
* Hospitals with ratings >= 4 :



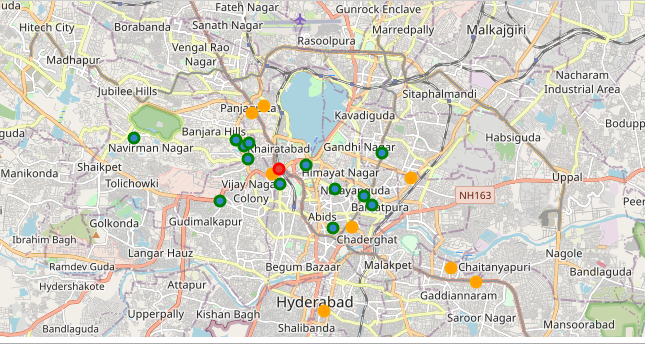
* Hospitals with ratings >=3 and < 4 :



* Hospitals with rating < 3:



* Visualizing this analysed Data :

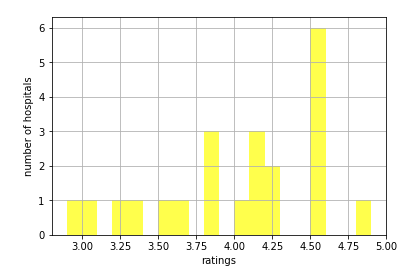


\* green plotted dots are of hospitals with rating >= 4

\* orange plotted dots are of hospitals with rating >= 3 and < 4

\* red plotted dots are of hospitals with rating < 3

**Plotting histogram of ratings :**



We can observe number of hospitals with a given range of ratings

* Results and Discussion :

After collecting the data from API we can observe we have also extracted some unwanted data which you had to eliminate by keeping only the required data . we also noticed that all the hospitals mentioned from the data we got from API are also not totally appropriate for analysis as some are repeated and some are having incomplete data.

We can eliminate that hospital from the data.

As I couldn’t find the appropriate data set I wanted anywhere , so I had created my own data set which reduced the step of data cleaning for the data set . As the total number of hospitals after cleaning the data were just 22 . I have collected ratings and contact details of them from google.

After Analysing we can see that many hospitals were having ratings greater than 4 .but only few are extremely good ( rating > 4.5 ) and some are extremely bad ( rating <3 ).

The map plot we did after analysing the data can be used by the user to look which hospitals are good for them in their nearby surroundings .

**Further use of the data :**

We can make an app in which the customer gives their location and we find the best hospitals in his surroundings along with rating ,address and contact details ,so that he can choose whatever best he wants.

* Conclusion :

The purpose of this project is to help a person in hyderabad who wants to search for a best hospital in his surroundings depending on the ratings of other people and get the contact details of the hospital he wants to visit .

The maps mentioned above will help users to choose the best hospital for him.