PROJECT

BEST LOCALITY

By:
Prince

INTRODUCTION/ BUSINESS PROBLEM

The objective of this project is to find the best locality to live in Faridabad district which is similar to Janakpuri Locality.

As many people migrate from one place to another because of jobs , climate and personal reasons , it is difficult for them to choose the locality which is similar to their Hometown. So, this project helps to find that perfect place for them easily without wasting so much time and energy.

This project is particularly useful to :-

- Businessman ,
- People who have to migrate to another areas,
- People who want to know about the venues in a particular areas

as it helps them to know about the differences in the environment of many localities.

DATA SECTION

For this project, I used three data sets to execute my idea :-

- Faridabad's places data
- Foursquare Location Data
- Hometown location

Faridabad's places data - This data is collected from the the internet and then changes to a .CSV file. It contains all Neighborhoods' name of Faridabad.

I cleaned and processed the data , and display the map of Faridabad containing all Neighborhoods in data.

Foursquare Location Data - For getting the venues information nearby each Neighborhood I used Foursquare API , to get Venue's name , Venue's category and other things.

This data helps me to get all Venues' data near each Neighborhood in Faridabad.

Hometown location - Coordinates of a place where person's live i.e Janakpuri ,Delhi.

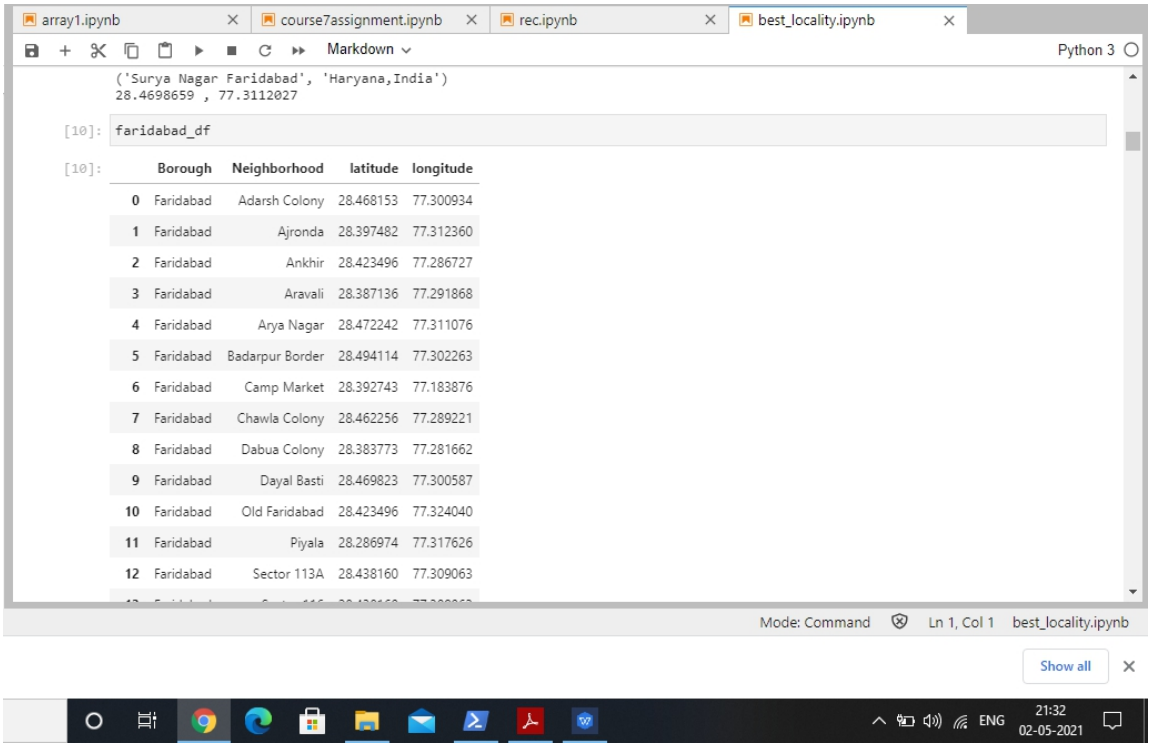
METHODS USED TO SOLVE PROBLEMS

For, the Project's Problem to be solved I have used many techniques.

- First, to get the list of Neighborhoods data of Delhi I convert the .csv file to Pandas Dataframe.

- Then, I use Geocoder Library to get the Latitudes and Longitudes of each Neighborhood in Faridabad for further analysis.

- After, the cleaning of data, our Dataframe look like this : -

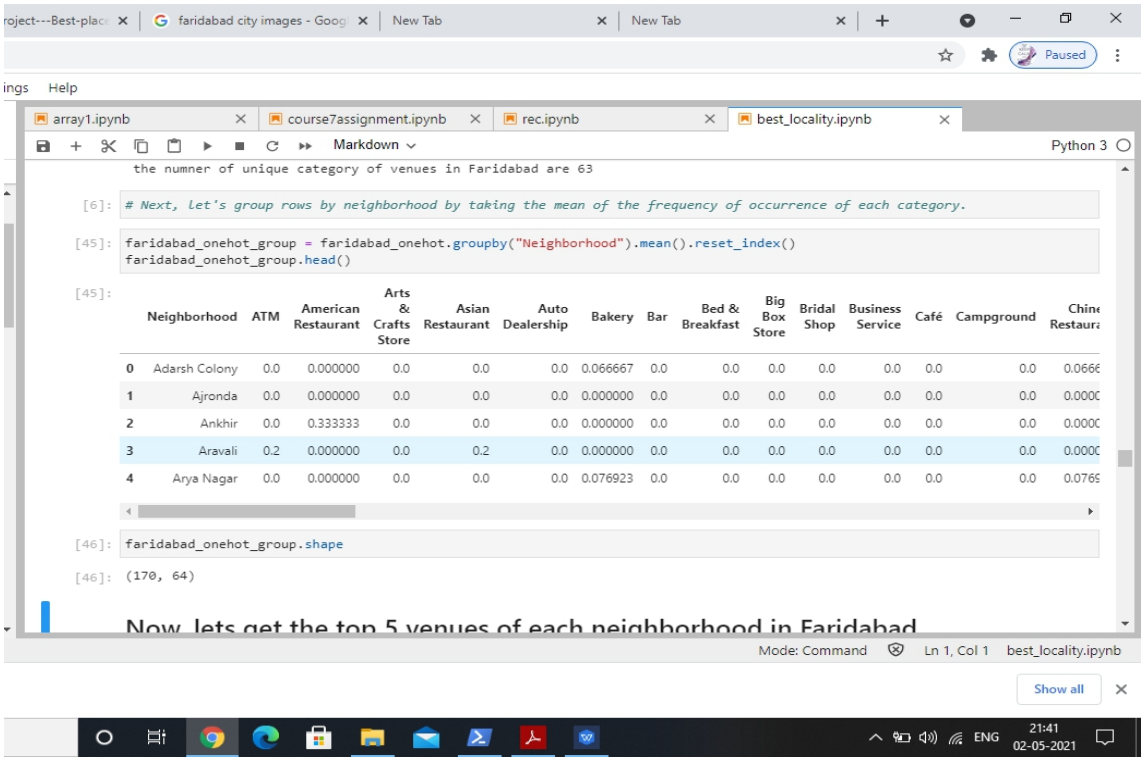


- Then , we will visualize the map of Faridabd with its Neighborhood using Folium Library as this library is very effective and easy to use.

- To get the all nearby venues of each Neighborhood I call the Foursquare API .

I , get the venues of each Neighborhood which present under 1000 meter of range and the maximum venue can be get is limit to 100.

- Then , I analyse each neighborhood by grouping the rows by neighborhood and taking the mean of the frequency of each occurrence of venue .



- Lastly, we will perform clustering on the data by using k-means clustering . K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

We will cluster the neighborhoods into clusters based on their most common frequency in Venue's Category.

The results will allow us to identify which neighborhoods are similar to each other and which are not.

The code is like :-

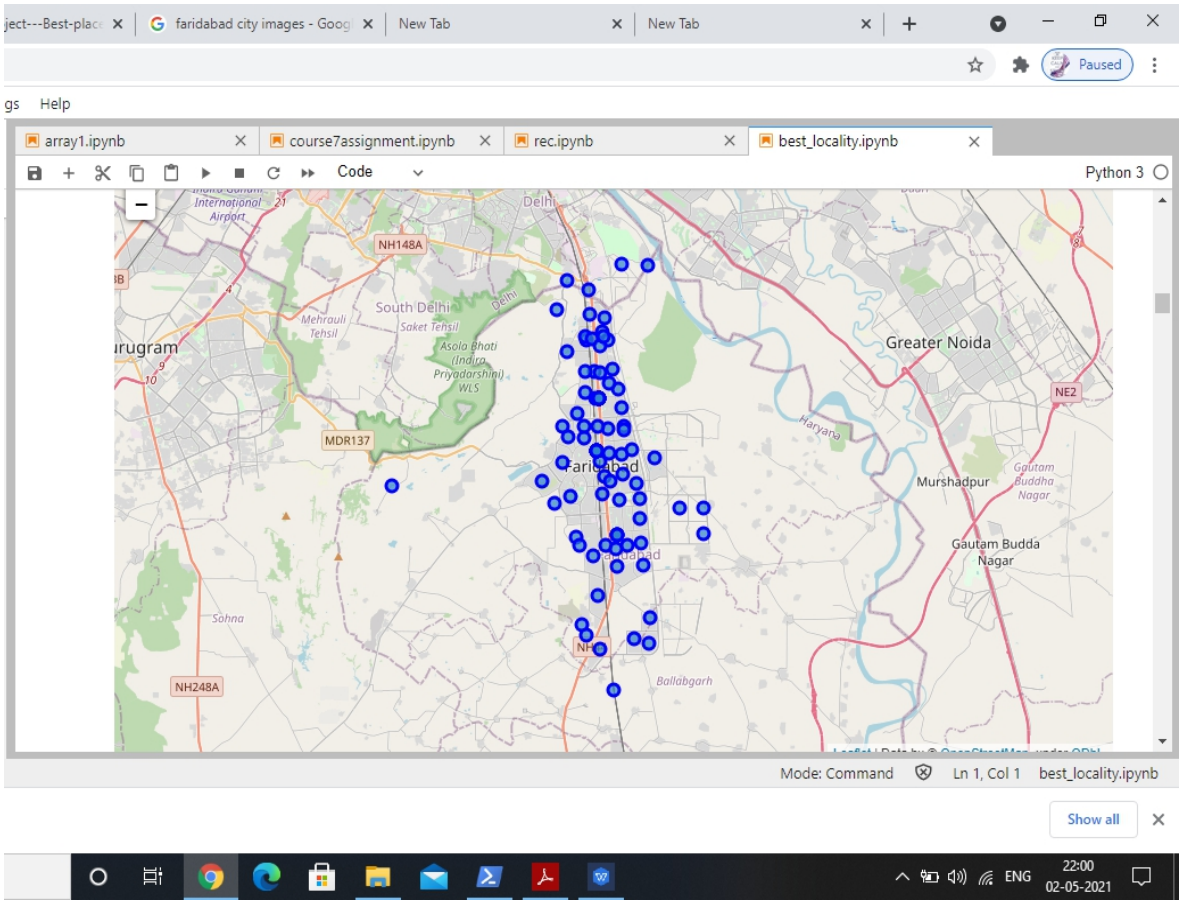
```
number_of_clusters = 6
faridabad_group_clustering =
faridabad_onehot_group.drop('Neighborhood' ,
axis =1)

kmeans = KMeans(n_clusters =
number_of_clusters ,random_state =
0).fit(faridabad_group_clustering)

kmeans.labels_
```

- After this , I made a map to show the neighborhoods with their labels(its cluster label) to easily know which Neighborhood are in same clusters I.e which are similar to each other.

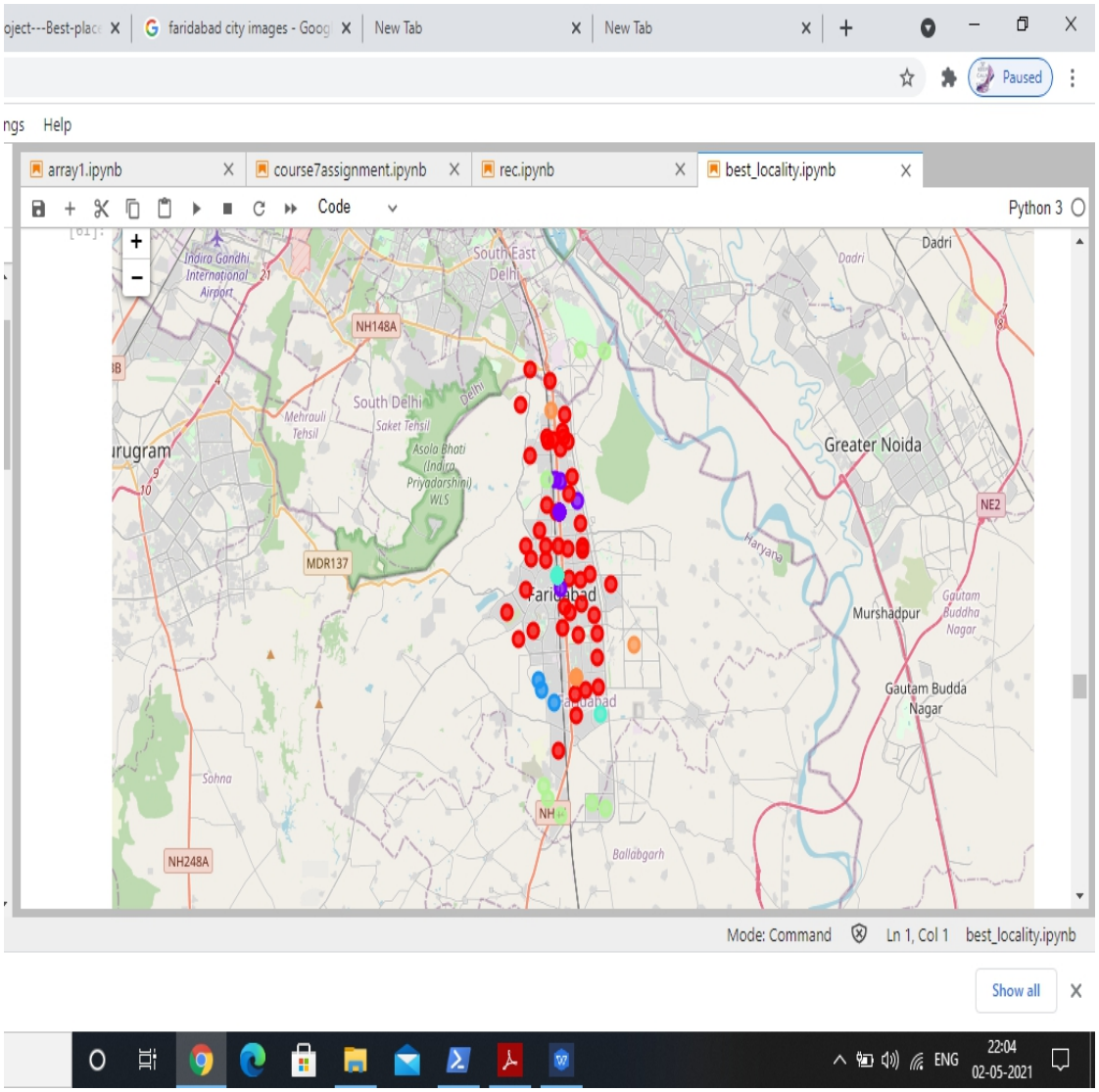
ANALYSIS SECTION

In this project , we analyzed many things.

1. Visualizing all Neighborhood in Faridabad using Folium map.

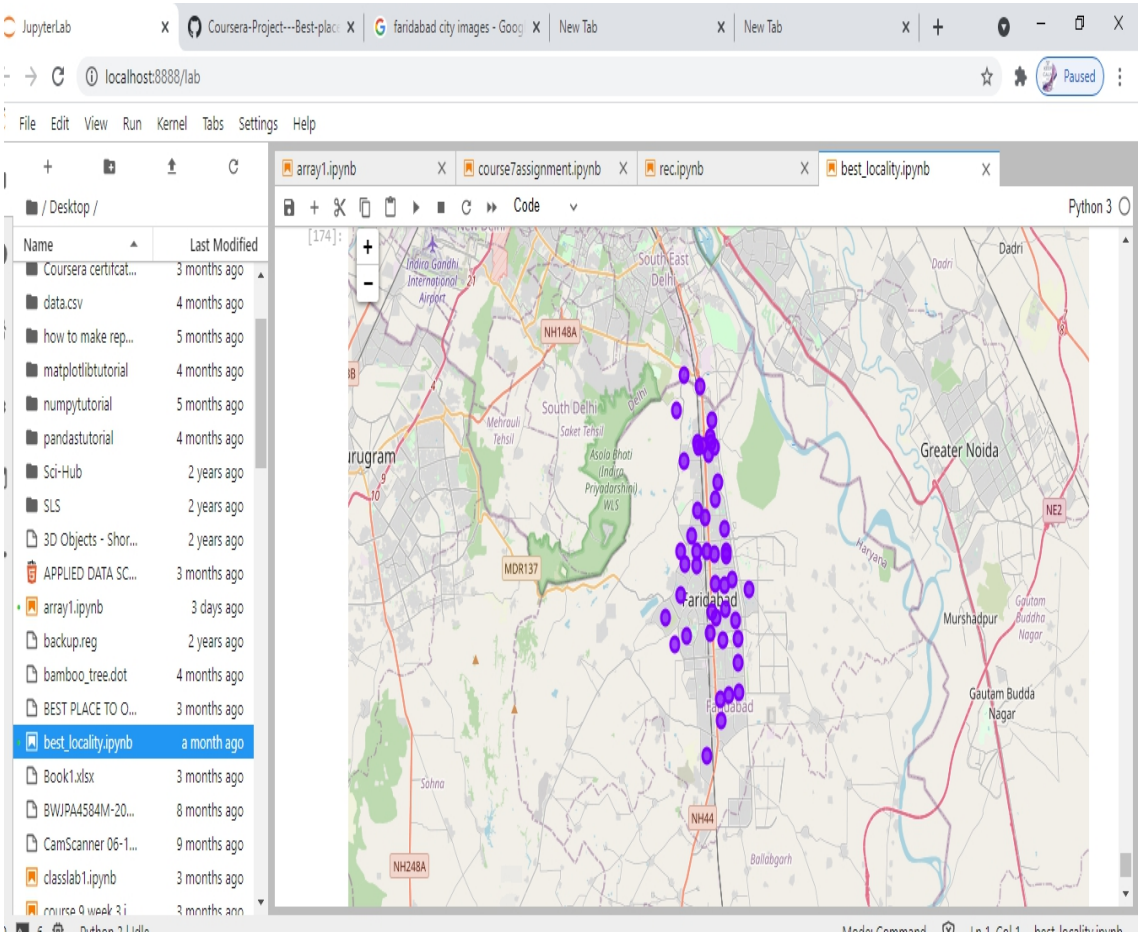


2. After analyzing and making cluster , making map of the Neighborhoods in Faridabad with its cluster label.



Here , different colors denotes the each 6 clusters.

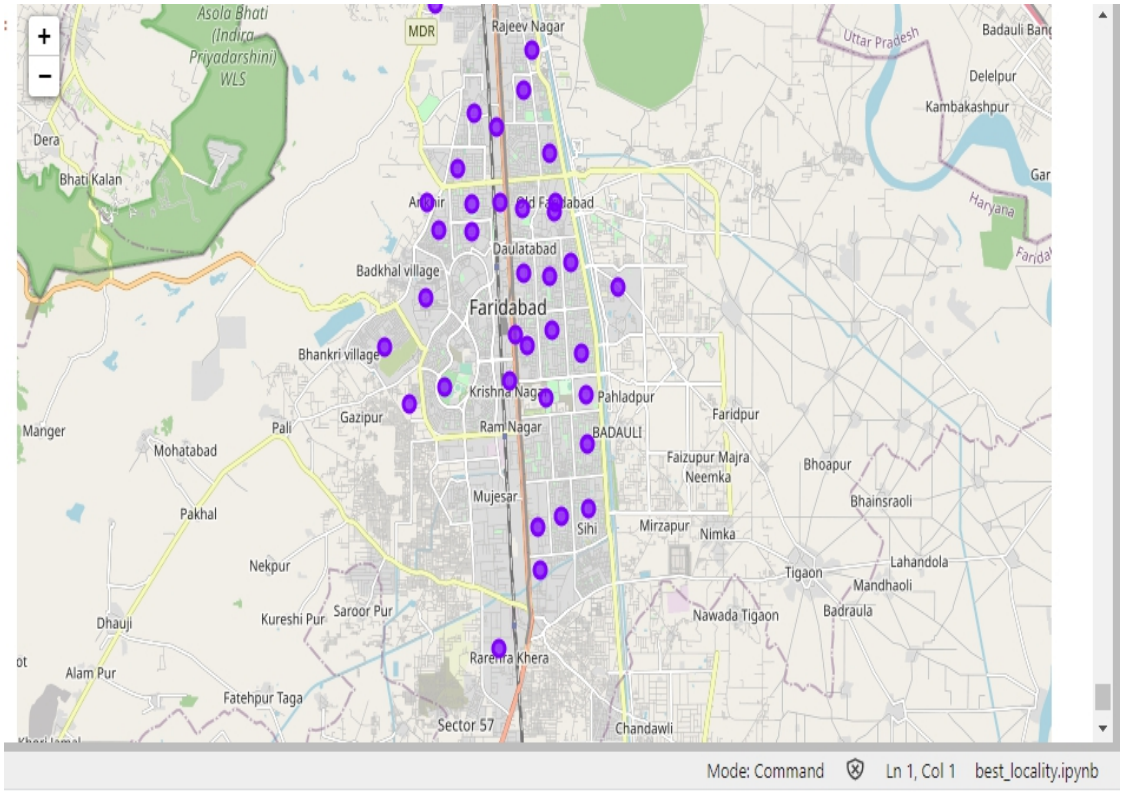
3. Visualizing the Neighborhoods which are alike Janakpuri, Delhi.



Here, Purple color denotes that Neighborhood in Faridabad which are like Janakpuri locality.

RESULT SECTION

The results from the k-means clustering shows that we can have the neighborhoods into a clusters which have similar Venues like in Janakpuri.



So, the all Neighborhood in cluster 0 (which are in Purple) are those which have same nearby venues like Janakpuri locality.

DISCUSSION SECTION

In this project we only consider one factor i.e only nearby venues present in each neighborhood , there are also other factors like population of people in each neighborhood which is similar to hometown , rate of land , crime rates , rate

of pollution , transport availability , and other factors which can directly affect our results.

However , this type of data is couldn't obtained easily.So, it is out of scope of this project.

Future research may can bypass these limitations by getting paid-information and obtain more and accurate results.

CONCLUSION

In this project, we have gone through the process of identifying the business problem ,specifying the data required, extracting and preparing the data, performing machine learning by clustering the data and then making 6 clusters
based on their similarities , and lastly providing recommendations to the relevant stakeholders i.e.
businessman and people regarding the best locations to live as similar to their Hometown.

The answer proposed by this project is: -->

The Neighborhood in cluster 0 (in purple) are those places which are most similar o hometown Jankapuri , Delhi .