CAPSTONE PROJECT - THE BATTLE OF NEIGHBORHOODS IN GURUGRAM

Ashish Kamble

INTRODUCTION

Gurugram, formerly known as Gurgaon, is a city located in the northern Indian state of Haryana. It is situated near the Delhi-Haryana border, about 30 kilometres (19 mi) southwest of the national capital New Delhi and 268 km (167 mi) south of Chandigarh, the state capital. It is one of the major satellite cities of Delhi and is part of the National Capital Region of India as of 2011, Gurgaon had a population of 1,153,000.

Gurugram is India's second largest information technology hub and third largest financial and banking hub. Gurugram is also home to India's largest medical tourism industry. Despite being India's 56th largest city in terms of population, Gurugram is the 8th largest city in the country in terms of total wealth. Gurugram serves as the headquarters of many of India's largest companies, is home to thousands of startup companies and has local offices for more than 250 Fortune 500 companies. It accounts for almost 70% of the total annual economic investments in Haryana state, which has helped it become a leading hub for high-tech industry in northern India. Gurugram is categorised as very high on the Human Development Index, with an HDI of 0.889 (2017).

Gurgaon is full of young people working in IT & Manufacturing sector, thousands of professionals' visits hundreds of restaurants in Gurugram. Almost all the restaurants are full on weekends as well as on weekdays. Our idea behind this project is to open a North Indian restaurant near any IT sector area where we can attract lots of professional's lunch, dinner as well as for hangout.

BUSINESS PROBLEM

As part of north India, North Indian cuisine is famous among people in Gurgaon like roti, dal makhani, paneer tikka, chicken tandoori, etc. as because it is local cuisine, readily available & pocket friendly majority of people prefer north Indian food. Our main objective of this project is to finalize a location in Gurgaon where we can successfully open and run our restaurant. We have to finalize our location in the busiest neighbourhood so as to generate good amount revenue as well as excellent star rating to attract more visitors.

TARGET AUDIENCE

- Business professionals who want to open north Indian restaurant in locality
- People who love eating north Indian food

DATASETS

- For data exploration we use Kaggle Zomato dataset on kaggel.com link: https://www.kaggle.com/shrutimehta/zomato-restaurants-data
- Foursquare database: https://Foursquare.com to be used in order to explore the desired neighbourhood data for various restaurant details and access the JSON files. This data shall be utilized to map the North Indian restaurants in various locations.

METHODOLOGY

- Using Zomato Kaggle dataset we retrieve table of restaurant & locality with latitude & longitude.
- Then we pre-process the dataset and map the neighbourhoods using folium library.
- Using Foursquare API, we explore venues in datasets
- After exploring venues, we use K-Means Clustering algorithm & map the clusters on map
- We examine all clusters and this clustering results will help us to finalize a neighbourhood where we can start our business.

DATA CLEANING/PREPROCESSING

After pre-processing our dataset will look like this

```
df_votes = df_Rev.groupby(['Locality'])['Votes'].sum().to_frame()
df_lat = df_Rev.groupby('Locality').mean()['Latitude'].to_frame()
df_lng = df_Rev.groupby('Locality').mean()['Longitude'].to_frame()
df_final = pd.merge(df_lat,df_lng,on='Locality').merge(df_res_loc,on='Locality').merge(df_cuisines,on='Locality').merge(df_res_r
```

[93]:
```
df_final.columns = ['Locality','Lat','Lng','No_of_Restaurant','Cuisines','Agg_rating','Comments','No_of_Votes']
df_final.head()
```

[93]:

	Locality	Lat	Lng	No_of_Restaurant	Cuisines	Agg_rating	Comments	No_of_Votes
0	ILD Trade Centre Mall, Sohna Road	28.424872	77.039265	2	Cafe, Beverages,North Indian, Mughlai	3.05	Average,Average	96
1	Ambience Mall, Gurgaon	28.502845	77.097152	2	North Indian, Mughlai,South Indian, Seafood, K...	4.15	Very Good,Very Good	2689
2	Ansal Plaza Mall, Palam Vihar	28.511460	77.041994	3	Bakery,Ice Cream,Chinese	3.00	Average,Average,Average	19
3	Ardee City	28.441660	77.083050	3	Japanese, Thai, Chinese,Pizza, Desserts,Raw Me...	3.00	Average,Average,Average	65
4	Best Western Skycity Hotel, Sector 15, Gurgaon	28.458165	77.034839	2	Japanese,North Indian, Chinese, Continental	2.85	Average,Average	28

[94]:
```
df_final.shape
```

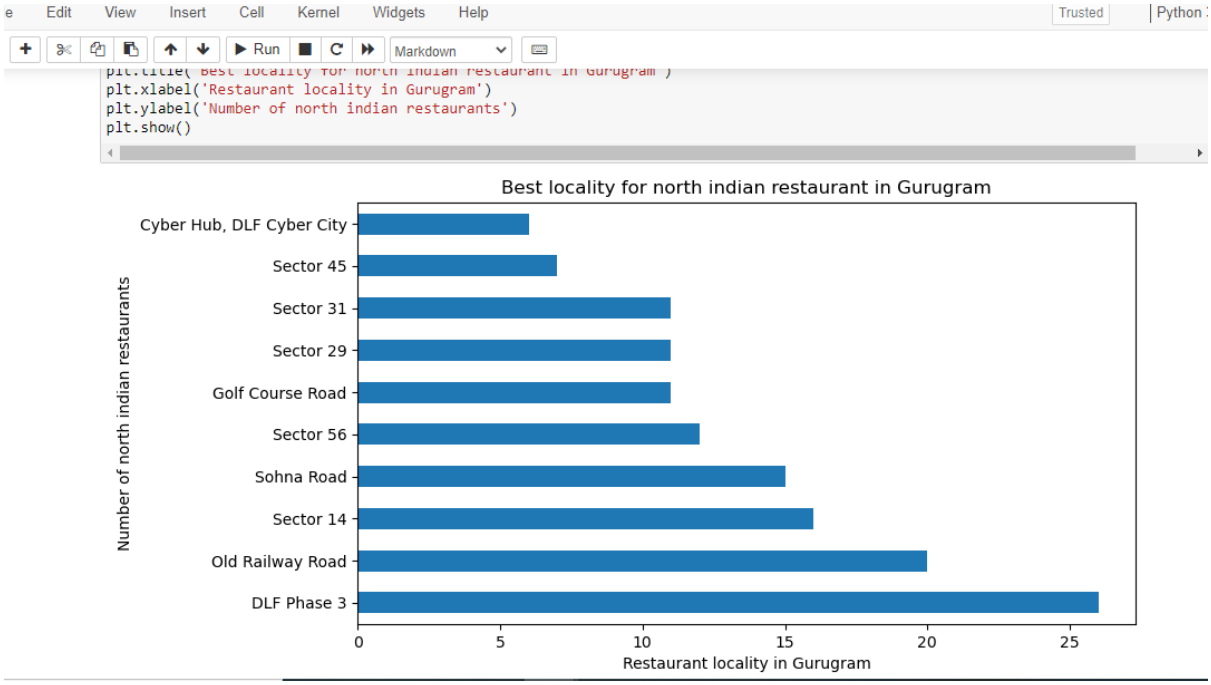
[94]:
```
(110, 8)
```

[92]:
```
df_final.isnull().sum()
```

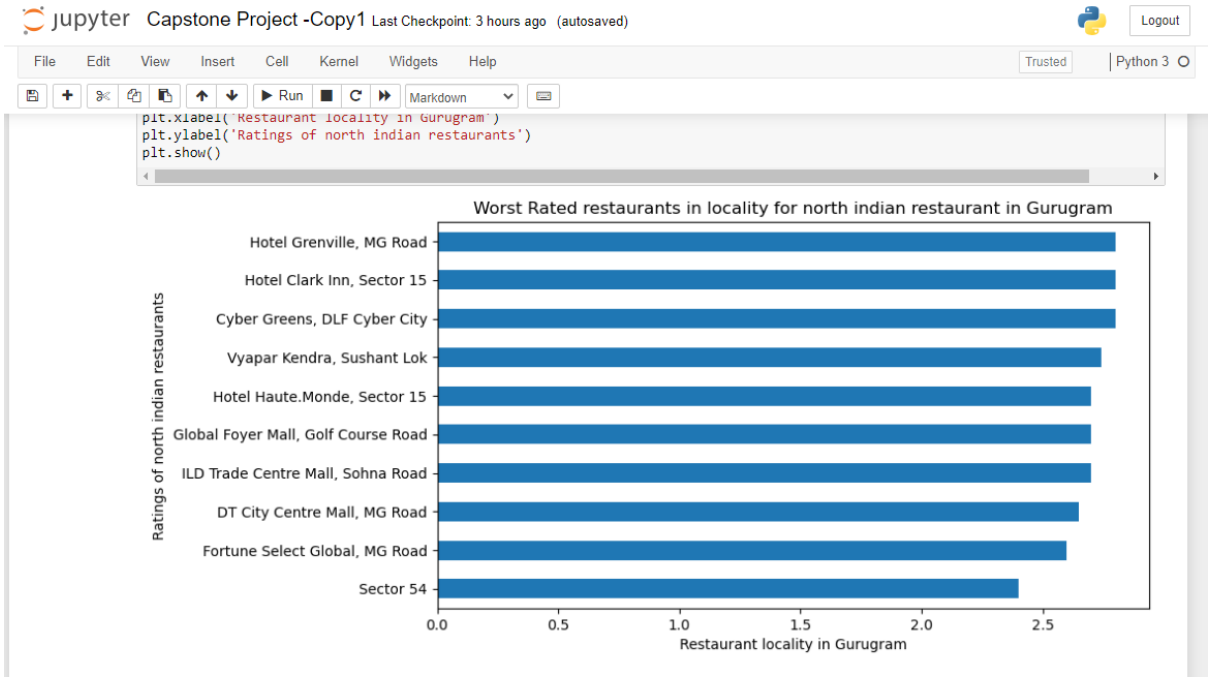
[92]:
```
Locality          0
Lat               0
```

To procced further, we first plot the bar for best locality & worst locality in Gurugram Neighbourhood.

Best Locality for north Indian restaurant in Gurugram

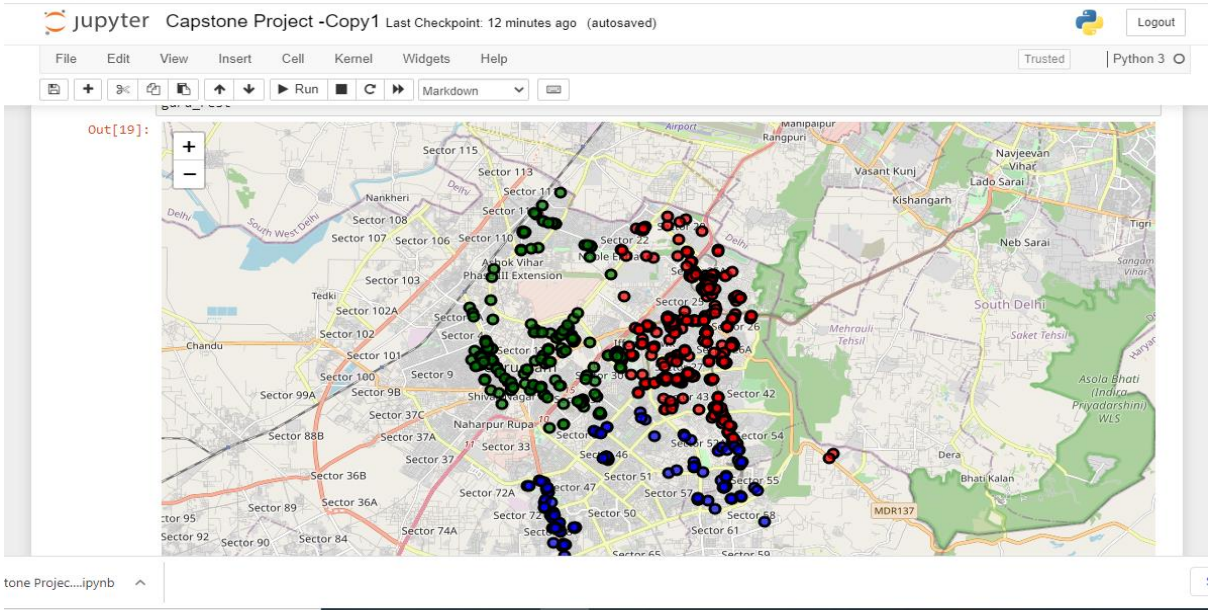


Worst Locality for North Indian Restaurant in Gurugram



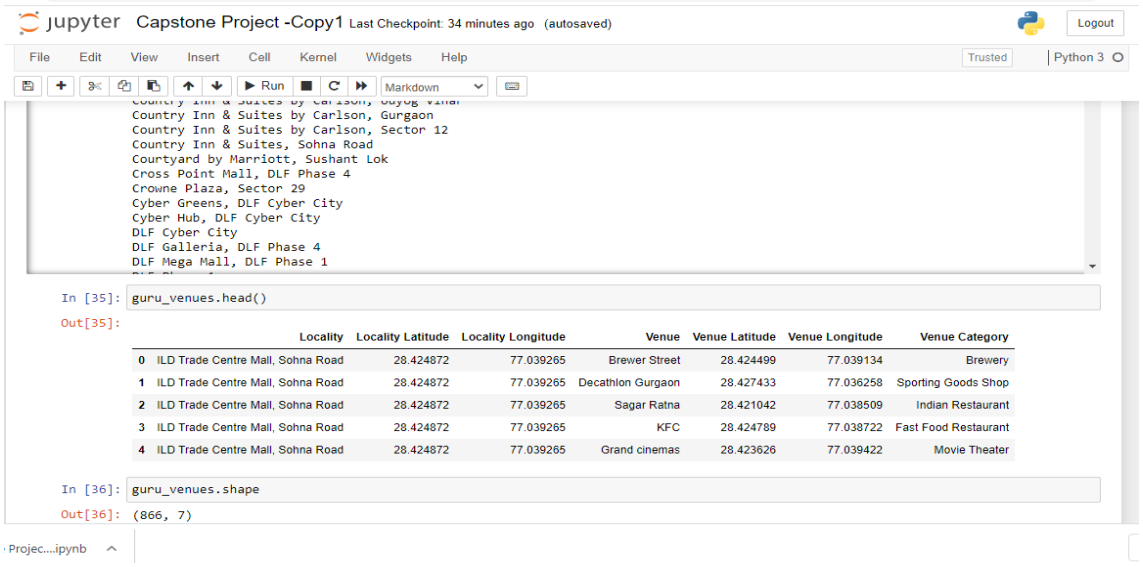
USING FOLIUM LIBRARY & FOURSQURE API

The pre-process dataset we plot on map using folium library



Then comes the process of Data Transformation followed by using foursquare API to explore venues in datasets, it consists of using our ‘Client Name’, ‘Client Secret’ & ‘Version’ all these parameters are called “Foursquare Credentials”.

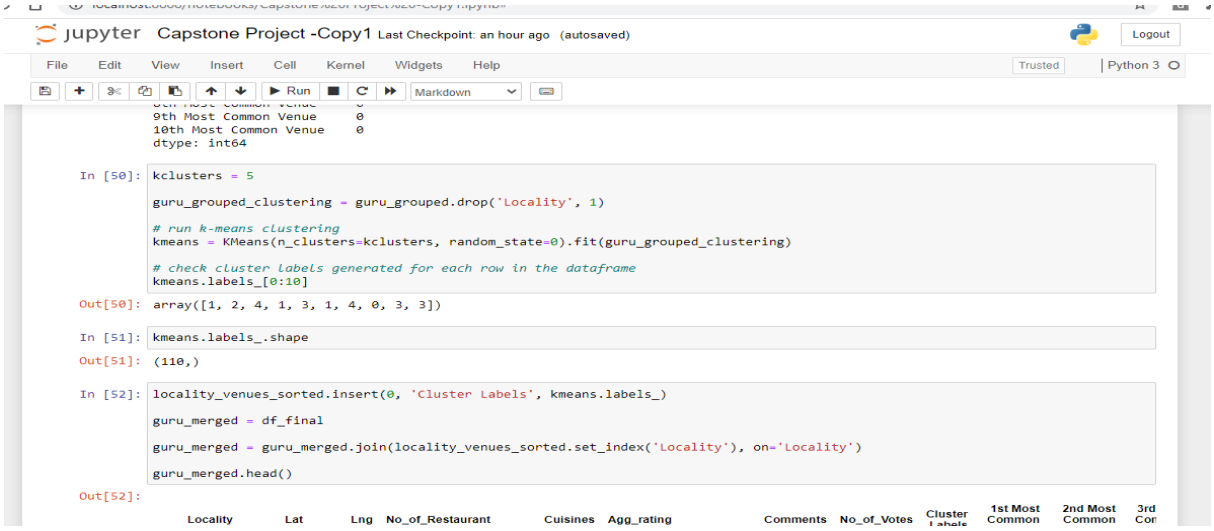
After successfully retrieving foursquare credentials we print result of ‘foursquare URL’ by using pandas.io.json library , which helps us to get the results of nearby_venues, & venue category in pandas data frame.



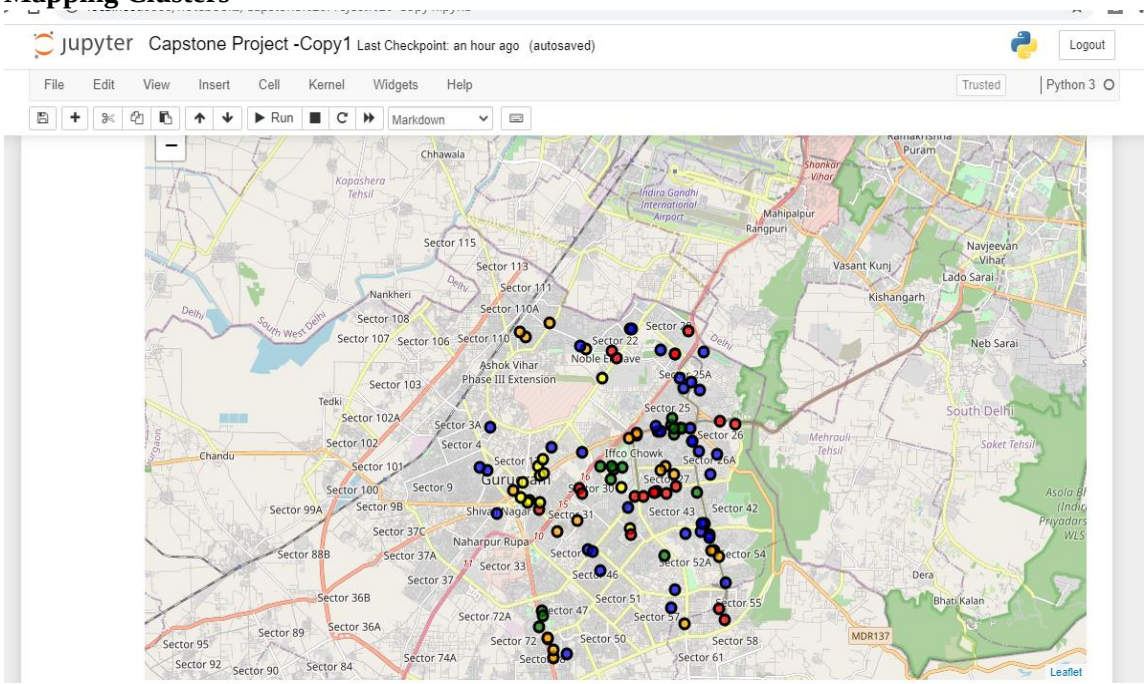
K-MEANS CLUSTERING ALGORITHM

k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean (cluster centres or cluster centroid), serving as a prototype of the cluster. This results in a partitioning of the data space into Voronoi cells. k-means clustering minimizes within-cluster variances (squared Euclidean distances), but not regular Euclidean distances, which would be the more difficult Weber problem: the mean optimizes squared errors, whereas only the geometric median minimizes Euclidean distances. For instance, better Euclidean solutions can be found using k-medians and k-medoids.

Then comes the step of sorting venues by locality getting results of all grouped venues (most common venues) and sorted venues. The crucial step of KMeans clustering comes into picture this time first we import clustering algorithm and name it kclusters = 5 as shown below, and plot the clusters on map We name it ‘guru_merged’

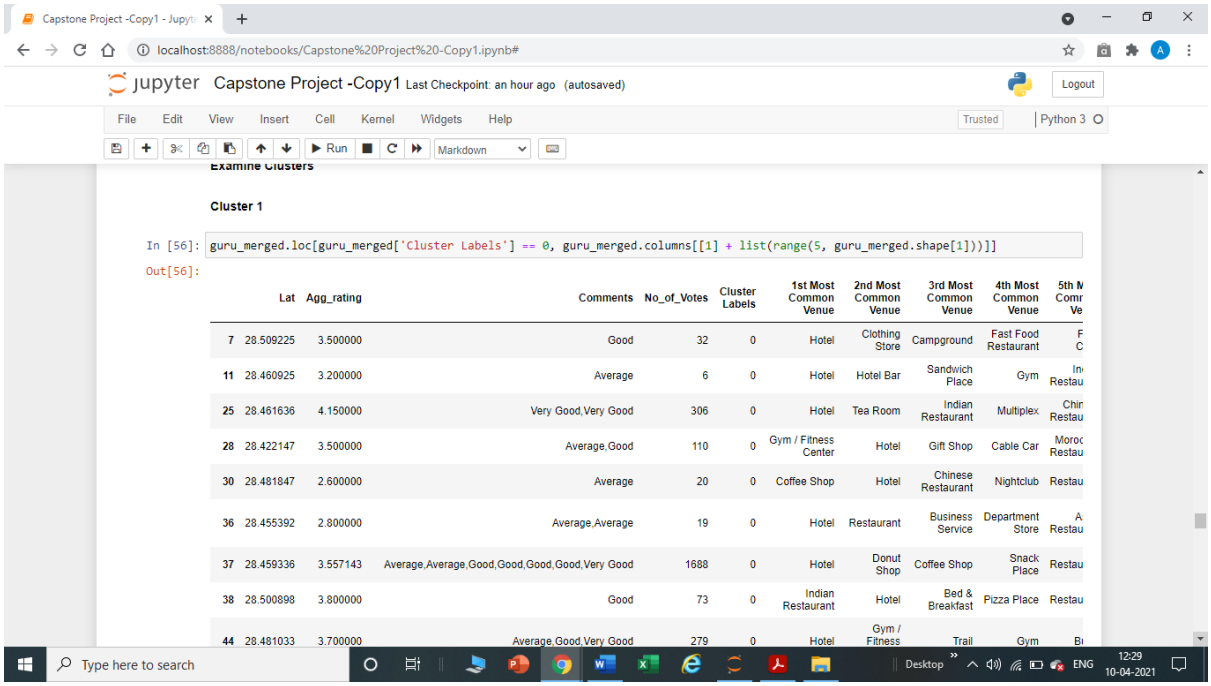


Mapping Clusters



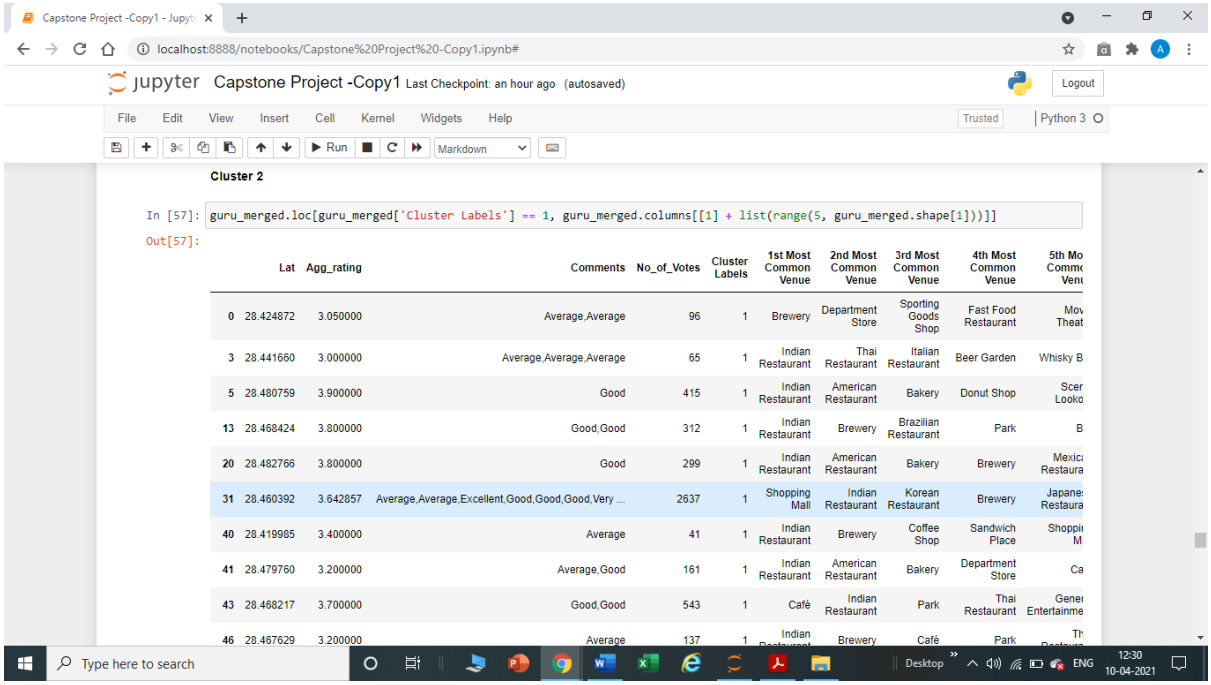
EXAMINING CLUSTERS

CLUSTER 1



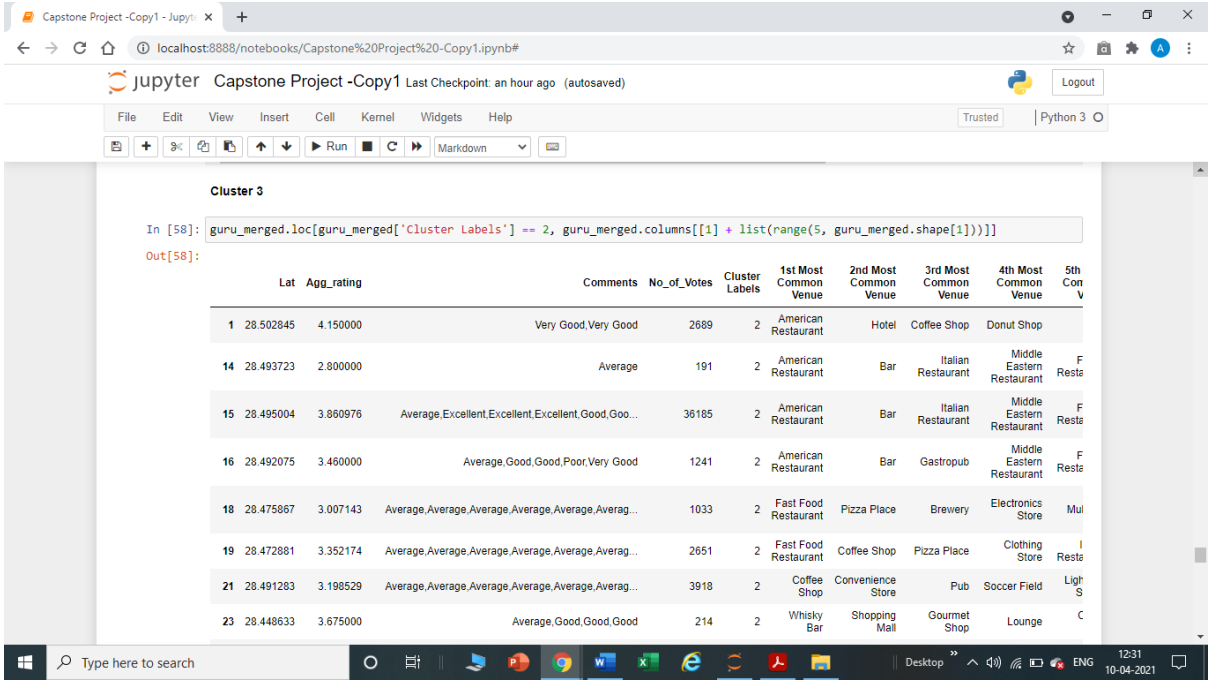
	Lat	Agg_rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th M Comm Ve
7	28.509225	3.500000	Good	32	0	Hotel	Clothing Store	Campground	Fast Food Restaurant	F C
11	28.460925	3.200000	Average	6	0	Hotel	Hotel Bar	Sandwich Place	Gym	In Restau
25	28.461636	4.150000	Very Good,Very Good	306	0	Hotel	Tea Room	Indian Restaurant	Multiplex	Chir Restau
28	28.422147	3.500000	Average,Good	110	0	Gym / Fitness Center	Hotel	Gift Shop	Cable Car	Moroc Restau
30	28.481847	2.600000	Average	20	0	Coffee Shop	Hotel	Chinese Restaurant	Nightclub	Restau
36	28.455392	2.800000	Average,Average	19	0	Hotel	Restaurant	Business Service	Department Store	A Restau
37	28.459336	3.557143	Average,Average,Good,Good,Good,Good,Very Good	1688	0	Hotel	Donut Shop	Coffee Shop	Snack Place	Restau
38	28.500898	3.800000	Good	73	0	Indian Restaurant	Hotel	Bed & Breakfast	Pizza Place	Restau
44	28.481033	3.700000	Average,Good,Very Good	279	0	Hotel	Gym / Fitness	Trail	Gym	Bi

CLUSTER 2



	Lat	Agg_rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Mo Comm Ven
0	28.424872	3.050000	Average,Average	96	1	Brewery	Department Store	Sporting Goods Shop	Fast Food Restaurant	Mov Theat
3	28.441660	3.000000	Average,Average,Average	65	1	Indian Restaurant	Thai Restaurant	Italian Restaurant	Beer Garden	Whisky B
5	28.480759	3.900000	Good	415	1	Indian Restaurant	American Restaurant	Bakery	Donut Shop	Scer Looko
13	28.468424	3.800000	Good,Good	312	1	Indian Restaurant	Brewery	Brazilian Restaurant	Park	B
20	28.482766	3.800000	Good	299	1	Indian Restaurant	American Restaurant	Bakery	Brewery	Mexici Restaura
31	28.460392	3.642857	Average,Average,Excellent,Good,Good,Good,Very ...	2637	1	Shopping Mall	Indian Restaurant	Korean Restaurant	Brewery	Japanes Restaura
40	28.419985	3.400000	Average	41	1	Indian Restaurant	Brewery	Coffee Shop	Sandwich Place	Shoppi M
41	28.479760	3.200000	Average,Good	161	1	Indian Restaurant	American Restaurant	Bakery	Department Store	Ca
43	28.468217	3.700000	Good,Good	543	1	Café	Indian Restaurant	Park	Thai Restaurant	Gener Entertainme
46	28.467629	3.200000	Average	137	1	Indian	Brewery	Café	Park	Th

CLUSTER 3



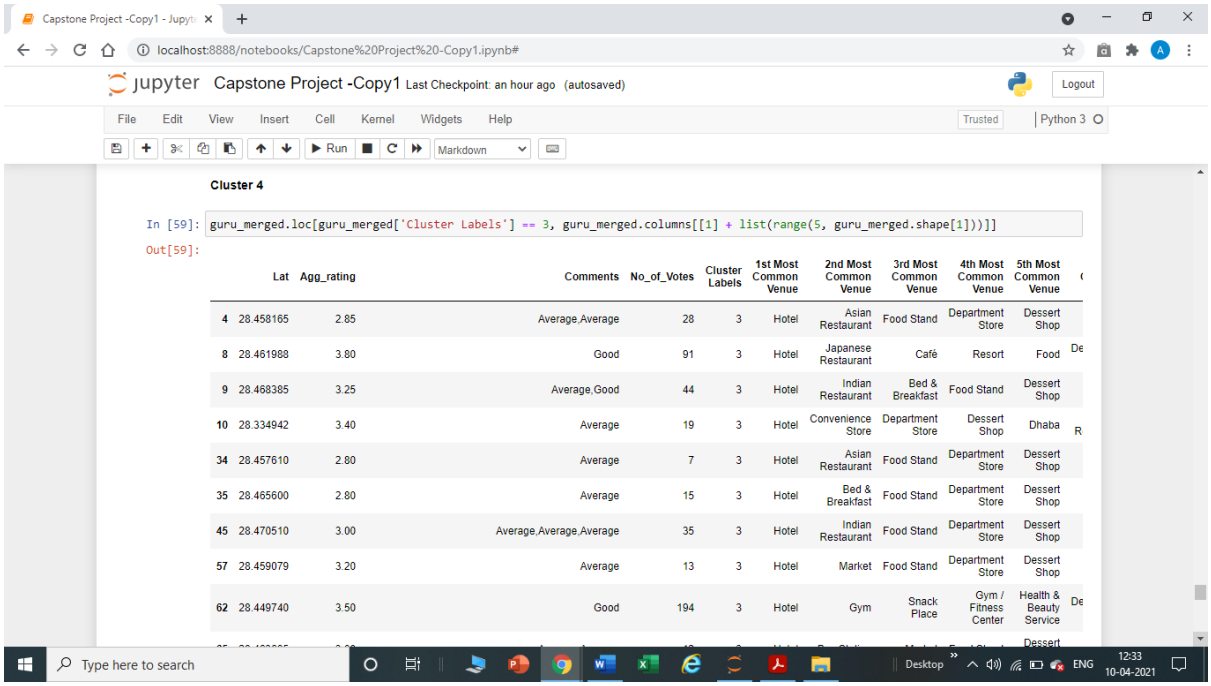
Cluster 3

```
In [58]: guru_merged.loc[guru_merged['Cluster Labels'] == 2, guru_merged.columns[[1] + list(range(5, guru_merged.shape[1]))]]
```

Out[58]:

	Lat	Agg_rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Com
1	28.502845	4.150000	Very Good,Very Good	2689	2	American Restaurant	Hotel	Coffee Shop	Donut Shop	
14	28.493723	2.800000	Average	191	2	American Restaurant	Bar	Italian Restaurant	Middle Eastern Restaurant	F Resta
15	28.495004	3.860976	Average,Excellent,Excellent,Excellent,Good,Goo...	36185	2	American Restaurant	Bar	Italian Restaurant	Middle Eastern Restaurant	F Resta
16	28.492075	3.460000	Average,Good,Good,Poor,Very Good	1241	2	American Restaurant	Bar	Gastropub	Middle Eastern Restaurant	F Resta
18	28.475867	3.007143	Average,Average,Average,Average,Average,Averag...	1033	2	Fast Food Restaurant	Pizza Place	Brewery	Electronics Store	Mul
19	28.472881	3.352174	Average,Average,Average,Average,Average,Averag...	2651	2	Fast Food Restaurant	Coffee Shop	Pizza Place	Clothing Store	I Resta
21	28.491283	3.198529	Average,Average,Average,Average,Average,Averag...	3918	2	Coffee Shop	Convenience Store	Pub	Soccer Field	Ligh S
23	28.448633	3.675000	Average,Good,Good,Good	214	2	Whisky Bar	Shopping Mall	Gourmet Shop	Lounge	C

CLUSTER 4



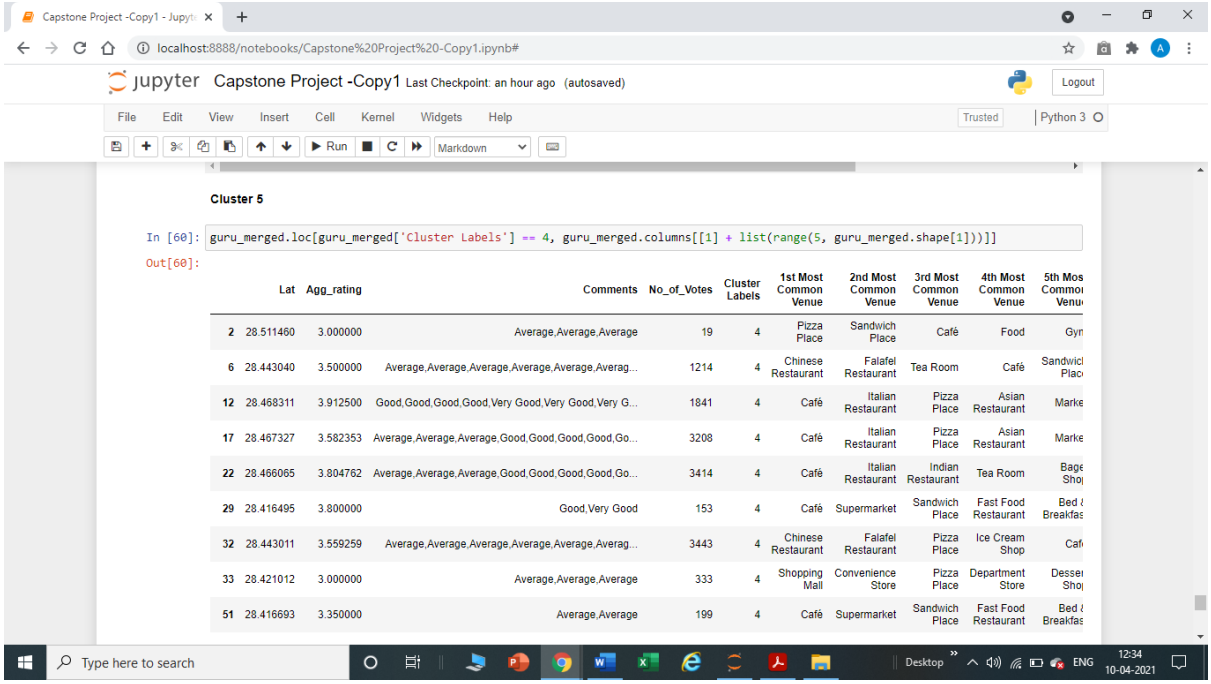
Cluster 4

```
In [59]: guru_merged.loc[guru_merged['Cluster Labels'] == 3, guru_merged.columns[[1] + list(range(5, guru_merged.shape[1]))]]
```

Out[59]:

	Lat	Agg_rating	Comments	No_of_Votes	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
4	28.458165	2.85	Average,Average	28	3	Hotel	Asian Restaurant	Food Stand	Department Store	Dessert Shop	
8	28.461988	3.80	Good	91	3	Hotel	Japanese Restaurant	Café	Resort	Food	De
9	28.468385	3.25	Average,Good	44	3	Hotel	Indian Restaurant	Bed & Breakfast	Food Stand	Dessert Shop	
10	28.334942	3.40	Average	19	3	Hotel	Convenience Store	Department Store	Dessert Shop	Dhaba	R
34	28.457610	2.80	Average	7	3	Hotel	Asian Restaurant	Food Stand	Department Store	Dessert Shop	
35	28.465600	2.80	Average	15	3	Hotel	Bed & Breakfast	Food Stand	Department Store	Dessert Shop	
45	28.470510	3.00	Average,Average,Average	35	3	Hotel	Indian Restaurant	Food Stand	Department Store	Dessert Shop	
57	28.459079	3.20	Average	13	3	Hotel	Market	Food Stand	Department Store	Dessert Shop	
62	28.449740	3.50	Good	194	3	Hotel	Gym	Snack Place	Gym / Fitness Center	Health & Beauty Service	De

CLUSTER 5



DISCUSSION

- Cluster 1 is good for hotels, bars, fast food & Indian restaurants, Cluster 2 has good number of Indian & Asian restaurants, with good user ratings, Cluster 3 is good for having Indian food & brewery, Cluster 4 famous for Chinese, Italian restaurants & coffee shops.
- DLF Phase3, Ambience Mall & golf course road are the neighbourhoods contain maximum number of North Indian restaurants.
- Ambience Mall, DLF star Mall sec30 & MG Road are the neighbourhoods contain best North Indian restaurants with good user ratings.

CONCLUSION

- DLF Cyber city and Sector 29 would be ideal neighbourhoods to start north Indian restaurant as it contains fair amount of north Indian restaurants with average user rating which results into less competition as compared to other neighbourhoods.
- So, a Restaurant with nice menu & ambience would be perfect in this neighbourhood to attract crowd as well as for running successful business.