THE BATTLE OF NEIGHBORHOODS

IBM DATA SCIENCE CAPSTONE

COURSERA

ABOUT THE CITY:

- Manchester Of South India
- Lots of Colleges and Hospitals
- Best Weather

TARGET AUDIENCE:

Those who are looking for the perfect location to open their restaurants, cafes, etc.

Localities with a greater number of restaurants present currently, have the greatest demand thereby earning high profits.

Question to be answered…

Which is the best location to open a restaurant (food place) at Coimbatore?

DATASET:

Coimbatore Data

- Link: https://news.abplive.com/pincode/tamil-nadu/coimbatore.html
- Consists of a table that includes the name of the Post Offices (Similar to neighborhoods), Taluks (Similar to Boroughs), District, State, and Postal Code.

Geospace Data

- To fetch the coordinates of the Post Offices

Foursquare Data

- Localities with a large number of restaurants indicate the high demand.
- This API is employed to fetch the nearby venues

WEB SCRAPPING AND PRE-PROCESSING:

- The text scrapped from the link is then parsed using Beautiful Soup thereby creating a beautiful soup object.
- The table from the soup object is converted into a pandas data frame.
- There are 587 unique post offices and 20 unique taluks.
- The objective of the project is to find a suitable location to start any food places. Therefore, the main part of the city is concentrated which includes Coimbatore, Coimbatore North and Coimbatore South Taluks.
- The resulting data frame consists of 183 instances.

	PostOffice	Taluk	Pincode
6	Agraharasamakulam	Coimbatore North	641110
13	Alandurai	Coimbatore North	641101
22	Anaikatti	Coimbatore North	641108
38	Athipalayam	Coimbatore North	641110
45	Bharathiyar University	Coimbatore North	641046

Processed Data Frame

FETCHING COORDINATES:

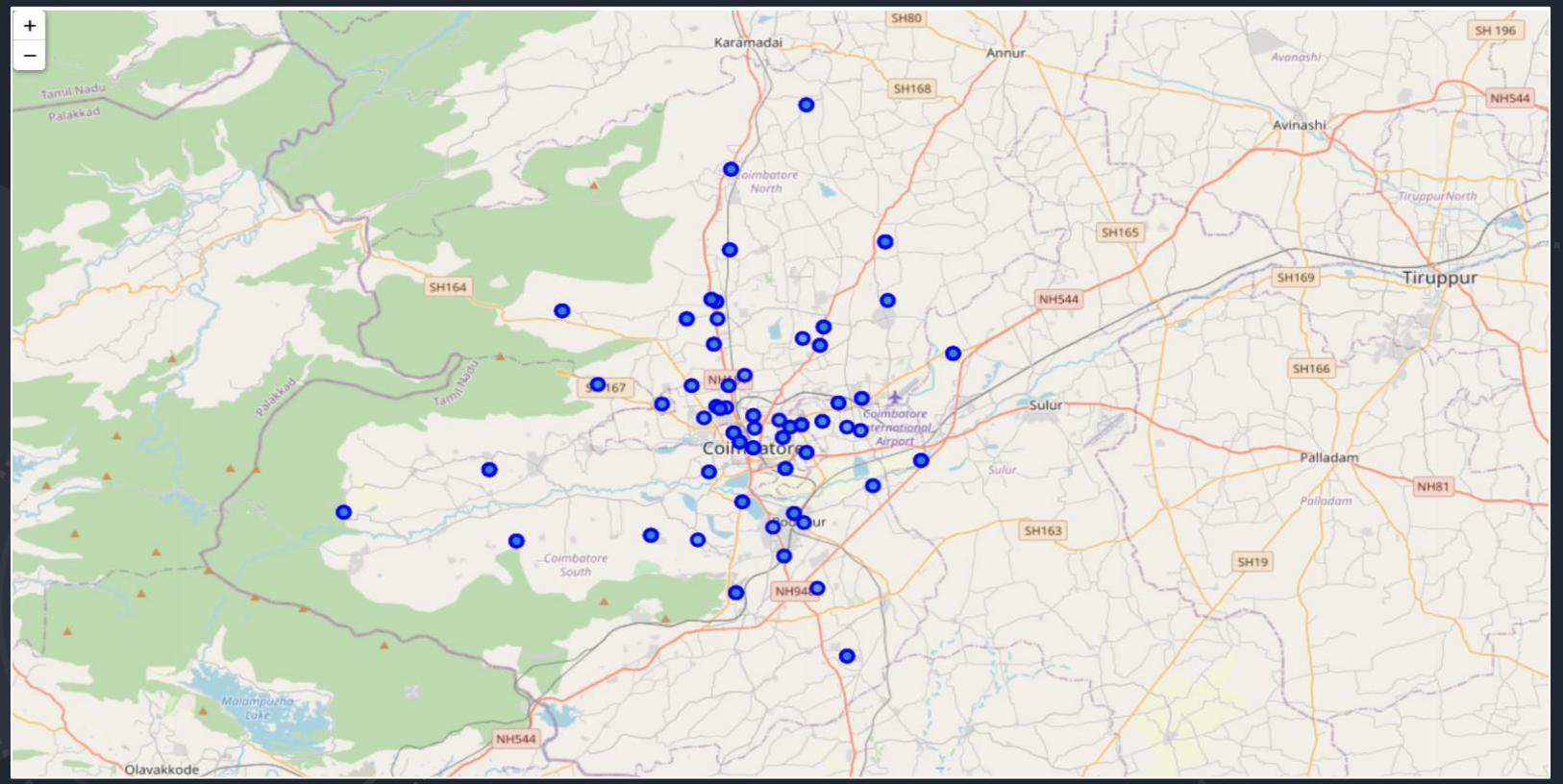
- To plot the localities and to fetch the nearby venues, the coordinates of the localities are essential.
- The indices of localities whose coordinates are not fetched are stored in a separate list.
- All the coordinates are perfectly obtained without any glitches except for index 113.
- This index is dropped from the data frame and the resulting data frame with shape 182 x 5 is obtained.

	PostOffice	Taluk	Pincode	Latitude	Longitude
0	Agraharasamakulam	Coimbatore North	641110	11.0776	76.9253
1	Alandurai	Coimbatore North	641101	10.9472	76.8305
2	Anaikatti	Coimbatore North	641108	11.0821	76.8566
3	Athipalayam	Coimbatore North	641110	11.0776	76.9253
4	Bharathiyar University	Coimbatore North	641046	11.039	76.8764
...	...	...	...	...	...
177	Kuttagam	Coimbatore	638462	11.0081	76.9795
178	Malumichampatti	Coimbatore	641050	10.9196	76.9985
179	Merkupathi	Coimbatore	638103	11.0081	76.9795
180	Vadavalli	Coimbatore	641041	11.0273	76.9116
181	Vallipuram	Coimbatore	638103	11.0081	76.9795

182 rows × 5 columns

Data frame incorporating Latitudes and Longitudes

PLOTTING LOCALITIES:



An interactive map is generated using Folium with all the Post Offices mapped.

FETCHING VENUES USING FOURSQUARE:

- Using the foursquare API, the venues that are around 500 mts from the localities are fetched.
- The resulting data frame consisted of 508 venues with 61 unique categories.
- Out of all the venue categories, all the categories related to food are alone considered.
- The data frame is then encoded and the total number of eating places are calculated for all the localities by summing up the columns.
- The latitudes and longitudes are then merged with the resultant data frame.

```
merged
```

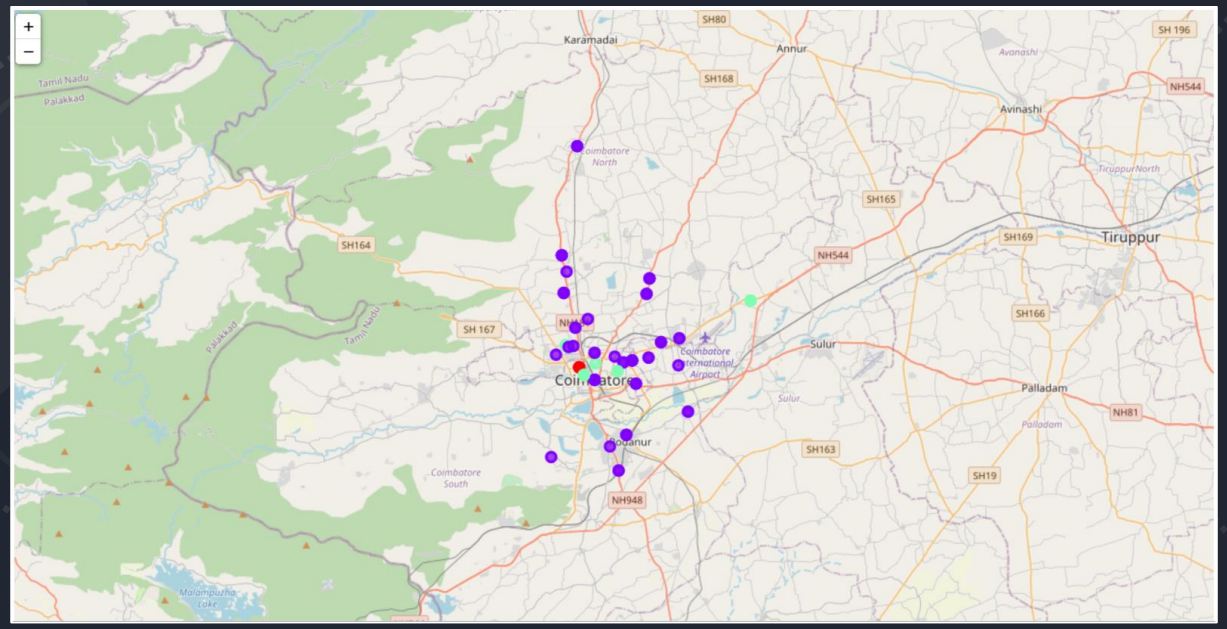
	PostOffice	Total Eating Places	Latitude	Longitude
0	Amritanagar	3	11.001812	76.962842
3	CBE Mpl Central Busstand	3	11.015528	76.989695
6	Cherannagar	3	11.062781	76.940771
9	Chettipalayam	3	11.001812	76.962842
12	Coimbatore Aerodrome	2	11.030835	77.023088
...	...	...	...	...
307	Vellakinar	4	11.062781	76.940771
311	Vellalapalayam Podanur	1	10.979933	77.029073
312	Vellalore	1	10.979933	77.029073
313	Venkitapuram	6	11.056904	77.073897
319	Vilankurichi	4	11.072893	77.001949

83 rows × 4 columns

Final Data frame

CLUSTERING:

- K-Means clustering is performed on the resulting data frame with the aim of clustering the whole dataset into 3 clusters thereby we can visualize the areas with high/moderate/low demands.
- After performing clustering operation, the post offices are visualized using folium, with different colours for each cluster.



Map after clustering

CONCLUSION:

- When the localities of each clusters are analyzed, post offices with cluster label 0 have the highest demand (red), those with cluster label 1 have the least demand (violet), and those with cluster label 2 have a moderate demand.

- Thus, the areas with red markers have the highest demand for food places and hence, they are the regions to be considered while starting a restaurant.

- Thus, the main part of Coimbatore is analysed using various parameters and the resulting localities are grouped into 3 clusters according to the total number of venues nearby.

THANK YOU!