CAPSTONE PROJECT-
THE BATTLE OF THE NEIGHBORHOODS

Applied Data Science Capstone by IBM/Coursera

Business Problem:

"Prospects of starting a Restaurant-cum-Catering service by inspecting the Zones of Chennai"

- Chennai being one of the metropolitan areas, is one of the growing IT hubs of India.
- Select the busiest zones in Chennai where a constant crowd is guaranteed
- Analyzing the office areas of the zones

The Business Problem can be stated as:

“What is the best place to open a Restaurant-cum-Catering Service in Chennai?”

Data (Requirement and collection)

- Zones Data (along with Coordinates): Collected by web scraping and got Coordinates using Geocoders
- Professional Venue Data: Collected using Foursquare by providing a unique category ID
- Nearby Venues Data: Collected using Foursquare by exploring near the selected zone coordinates
- Pricing Data: Collected manually from web-site for the selected zones

Exploratory Data Analysis

- All the 15 zones of Chennai were analysed to find the most frequent venue and the top professional zones.

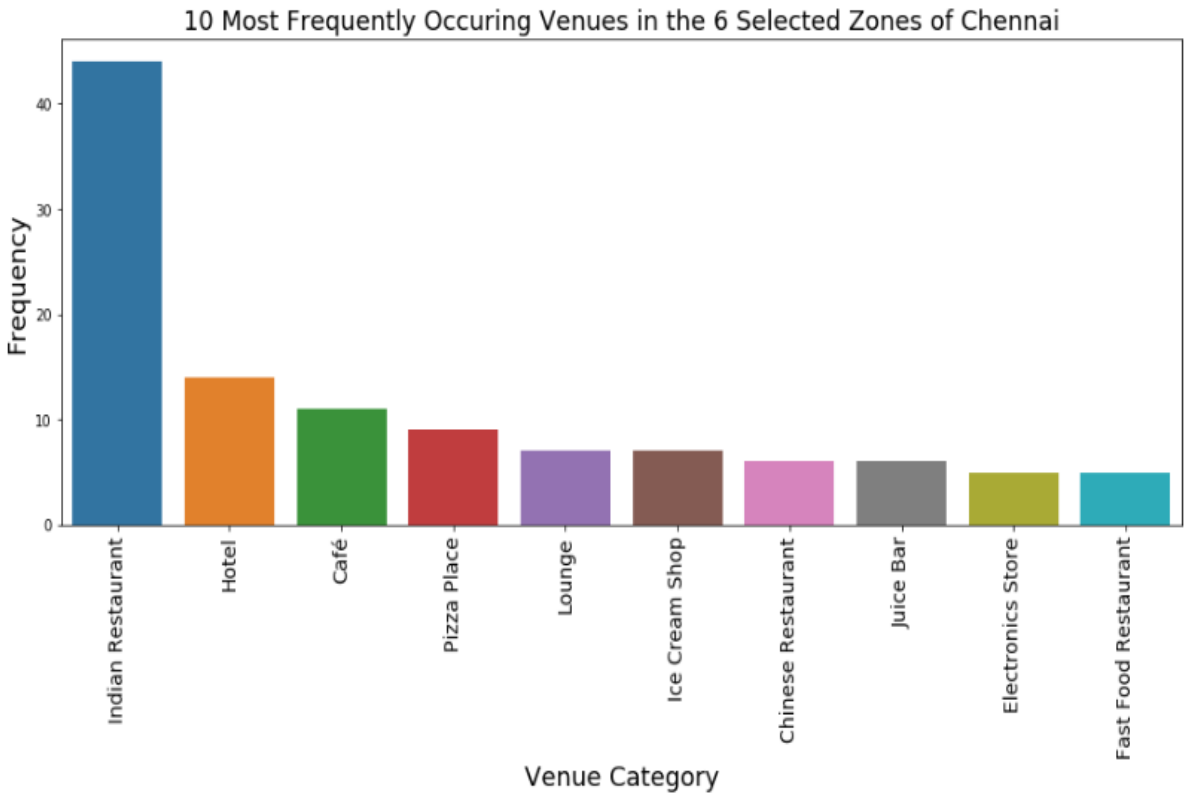
	Venue_Category	Count
0	Office	93
1	Hospital	49
2	Building	32
3	Event Space	30
4	Temple	28
5	Factory	24
6	Church	20

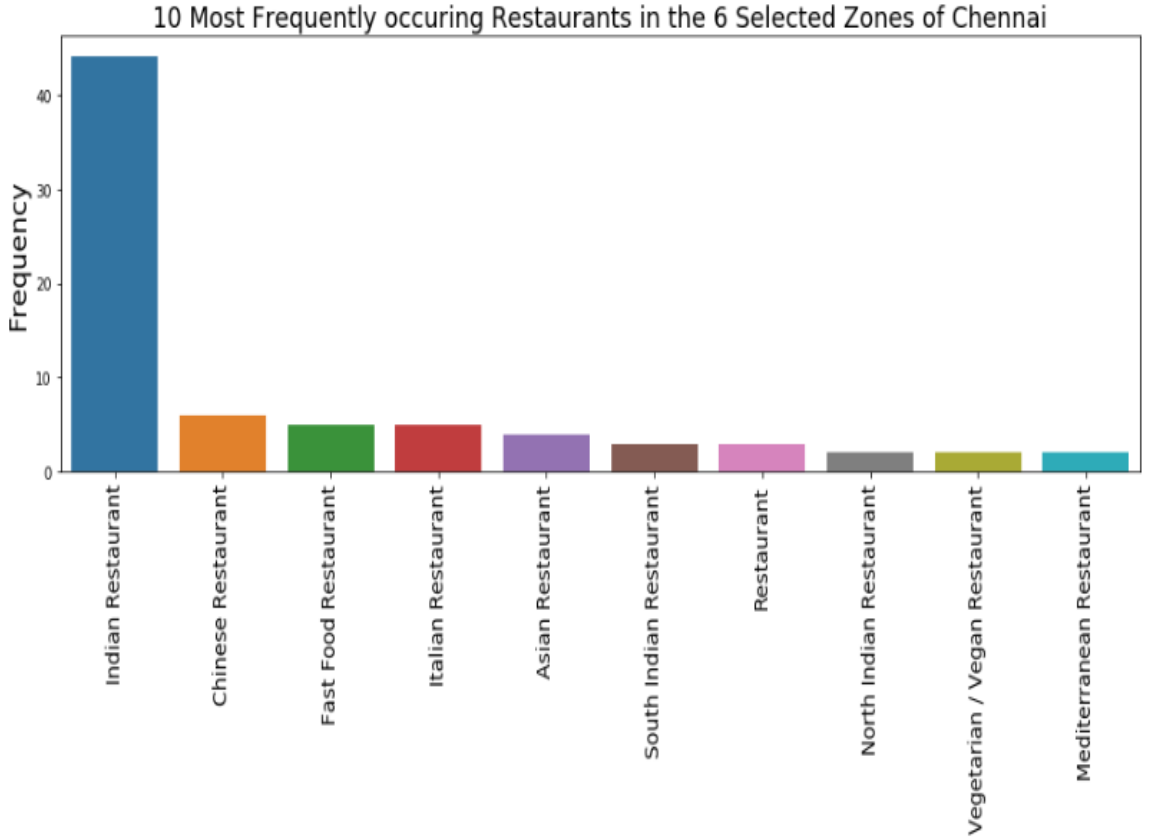
	Zone(Location)	Count
0	Teynampet	50
1	Ambattur	48
2	Perungudi	48
3	Adyar	47
4	Alandur	46
5	Kodambakkam	46
6	Valasaravakkam	44

- Based on this data 6 top busiest zones of Chennai are selected for further analysis

	Location	Latitude	Longitude
0	Adyar	13.006450	80.257779
1	Alandur	12.994373	80.194284
2	Ambattur	13.119375	80.150765
3	Kodambakkam	13.049207	80.224283
4	Perungudi	12.971024	80.241805
5	Teynampet	13.044324	80.249846

Exploratory Data Analysis



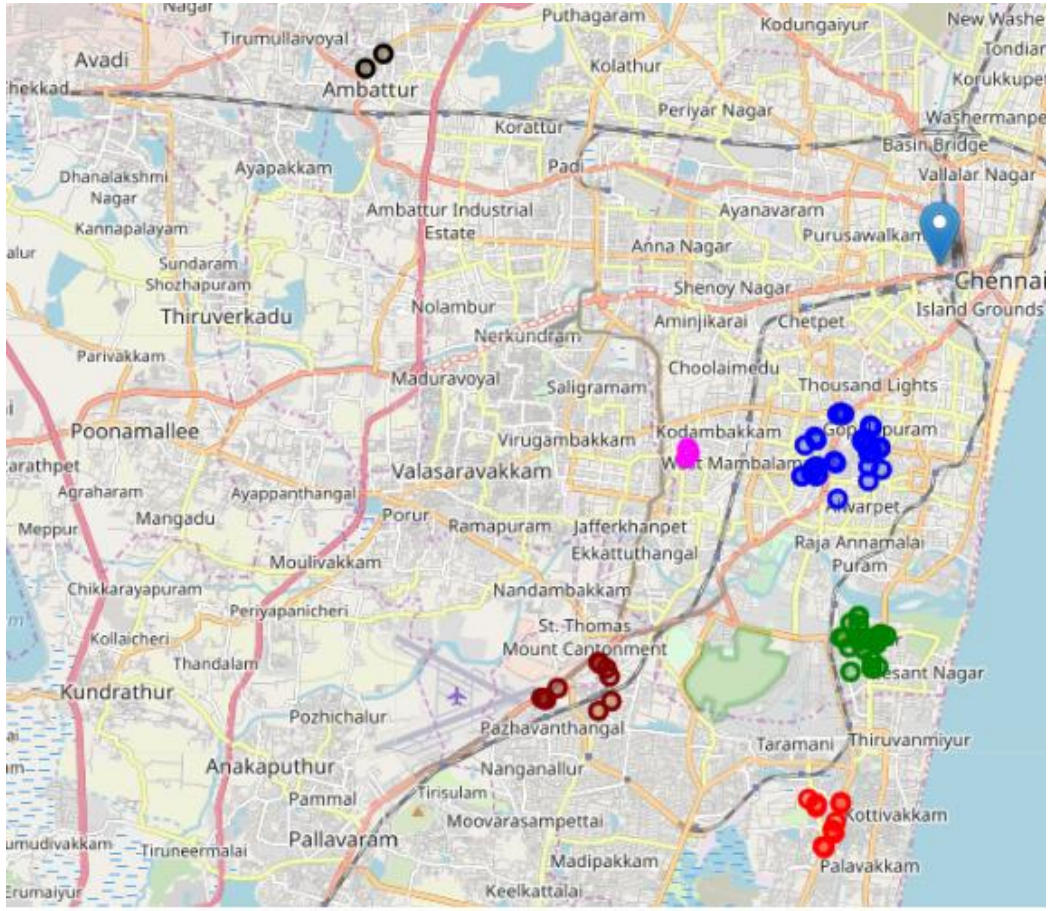


We can see that the most frequent place in these selected zones is “Indian Restaurant”.
The most demanded cuisine is Indian.

Exploratory Data Analysis

This Folum map shows how the restaurants are distributed and their count in the selected zones.

Location	Count
Teynampet	32
Adyar	28
Alandur	11
Perungudi	10
Kodambakkam	3
Ambattur	2



Analysis

- To better under the venue categories, one hot encoding is used.
- The top 6 frequent venues of each zone are found along with the frequency.

=========Adyar=========

	Venue	Freq
0	Office	0.21
1	Building	0.09
2	Temple	0.09
3	Hospital	0.09
4	Medical Center	0.06

=========Kodambakkam=========

	Venue	Freq
0	Office	0.22
1	Event Space	0.17
2	Building	0.11
3	Tech Startup	0.07
4	Hospital	0.04

=========Alandur=========

	Venue	Freq
0	Temple	0.13
1	Event Space	0.09
2	Post Office	0.09
3	Hospital	0.09
4	Spiritual Center	0.09

=========Perungudi=========

	Venue	Freq
0	Office	0.38
1	Conference Room	0.19
2	Meeting Room	0.10
3	Tech Startup	0.08
4	Building	0.06

=========Ambattur=========

	Venue	Freq
0	Hospital	0.19
1	Office	0.15
2	Event Space	0.08
3	Building	0.08
4	Factory	0.08

=========Teynampet=========

	Venue	Freq
0	Office	0.30
1	Building	0.12
2	Event Space	0.08
3	Hospital	0.06
4	Tech Startup	0.06

=========Adyar=========

	Venue	Freq
0	Indian Restaurant	0.29
1	Café	0.07
2	Pizza Place	0.05
3	North Indian Restaurant	0.03
4	Electronics Store	0.03

=========Kodambakkam=========

	Venue	Freq
0	Indian Restaurant	0.15
1	Juice Bar	0.15
2	Electronics Store	0.15
3	Bakery	0.08
4	Jewelry Store	0.08

=========Alandur=========

	Venue	Freq
0	Indian Restaurant	0.24
1	Hotel	0.08
2	Breakfast Spot	0.08
3	Train Station	0.08
4	Pizza Place	0.08

=========Perungudi=========

	Venue	Freq
0	Indian Restaurant	0.21
1	Boutique	0.11
2	Chinese Restaurant	0.11
3	Platform	0.05
4	Pizza Place	0.05

=========Ambattur=========

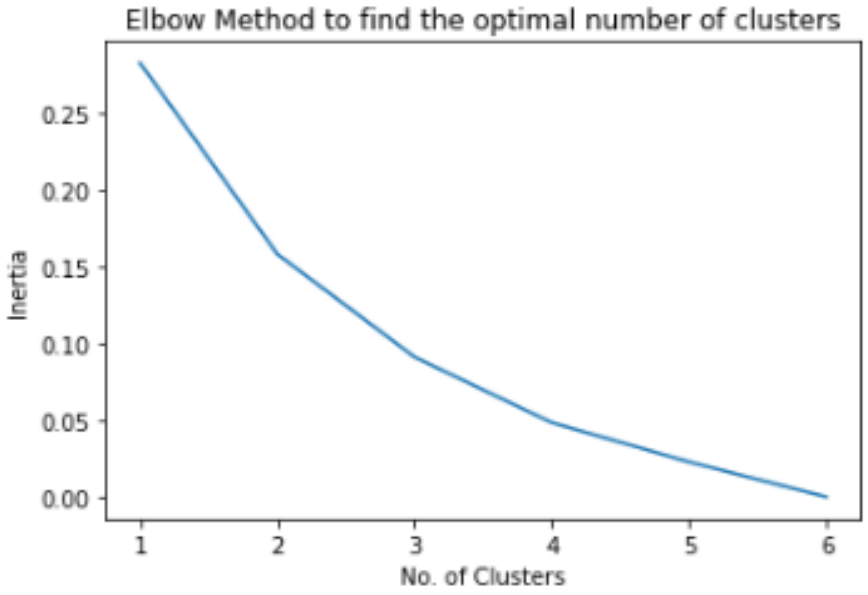
	Venue	Freq
0	Flea Market	0.18
1	Ice Cream Shop	0.18
2	Movie Theater	0.18
3	Indian Restaurant	0.09
4	Clothing Store	0.09

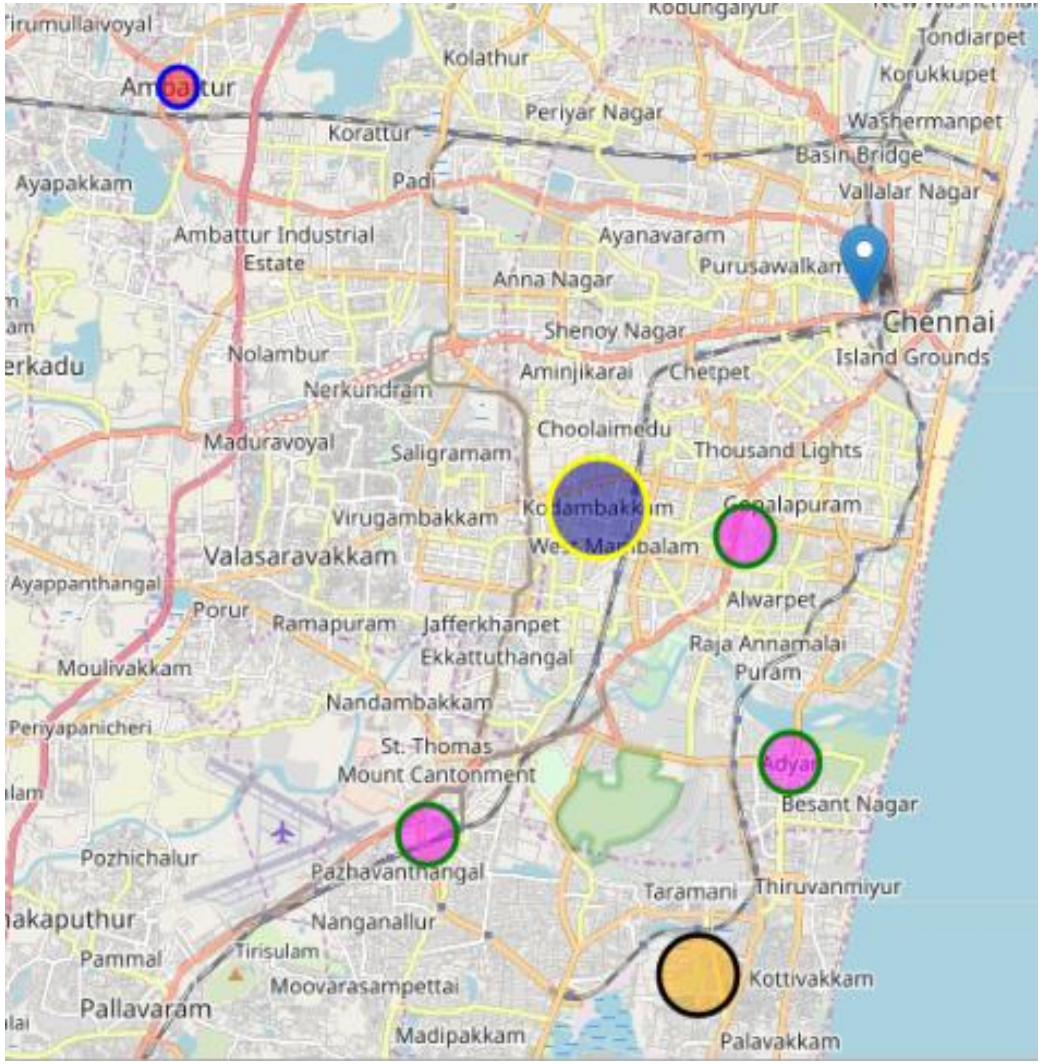
=========Teynampet=========

	Venue	Freq
0	Indian Restaurant	0.16
1	Hotel	0.14
2	Lounge	0.06
3	Café	0.06
4	Italian Restaurant	0.05

K-Means Clustering

- Using the elbow method, the number of clusters are selected (clustering with least inertia is preferred)
- The zones are clustered into 4 clusters





K-Means Clustering

	Location	Cluster Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	Avg Price per sqft
2	Ambattur	1	Flea Market	Movie Theater	Ice Cream Shop	4420
0	Adyar	2	Indian Restaurant	Café	Pizza Place	12325
1	Alandur	2	Indian Restaurant	Hotel	Train Station	7607
5	Teynampet	2	Indian Restaurant	Hotel	Café	12792
4	Perungudi	3	Indian Restaurant	Chinese Restaurant	Boutique	6885
3	Kodambakkam	4	Juice Bar	Electronics Store	Indian Restaurant	7607

- Cluster 1 contains zones whose common venues are not restaurants: Entertainment and shopping areas
- Cluster 2 contains zones whose 1st most frequent venue is a restaurant (with the highest frequencies)
- Cluster 3 contains zones with top 2 most common venues being restaurants
- Cluster 4 contains zones whose 1st common venue is a restaurant but with the same frequency as the 2nd and 3rd common venues.

Results and Discussion

- ‘Adyar’, ’Kodambakkam’, ’Perungudi’ and ’Teynampet’ zones have ‘Office’ as the 1st frequent professional venue with ‘Perungudi’ having the highest frequency among them.
- This tells us that these 4 zones, out of the 6 selected zones, will be good for our business as our potential customers are employees.
- Also, Indian Restaurants are the most frequent venues near the selected zones, suggesting the type of cuisine customers in that area prefer.
- Teynampet has the most restaurants, out of all the selected zones, based on the venues explored.
- 'Ambattur' has the least average price per sqft, followed by 'Perungudi', among the selected zones.
- **With maximum frequency of offices and moderate restaurants in the area ‘Perungudi’ seems like a potential zone to open up our Restaurant-cum-Catering service.**
- **The pricing data also seems favorable to this. Clustering also shows these venues in cluster 3 which represents the cluster with restaurants as the frequent venues.**

Conclusion

- Understood how to deal with real life data science projects using some of the popular Python packages such as seaborn, folium, BeautifulSoup and geocoders.
- I have also got a glimpse of how web scraping is done and how FourSquare can be used to acquire data of frequent venues in a selected area.
- The idea of opening a “Restaurant-cum-Catering service in an area which has a huge pool of office workers (‘Perungudi’) is an interesting and a potential idea to try in Chennai where Catering Services are not very well established
- Although the analysis is very preliminary and requires a lot of refining based on the data used (refined ward data per each zone, pricing data) it will be of great help in the beginning stages of the business plan.