ANALYSIS OF RESTAURANT DISTRIBUTION IN THE CITY OF CHENNAI

By S Karthik Krishnan

BUSINESS UNDERSTANDING

- To analyse the restaurant distribution in the city of Chennai.
- To cluster them into meaningful groups that provide useful insights while choosing a location to open a restaurant.
- The target stakeholders are individuals that are involved in or are interested in the restaurant industry and is especially useful for stakeholders interested in opening up restaurants in different venues in the city.
- The reason for choosing Chennai is because it is one of the major metropolitan cities in the country of India and considering the fact that India is a developing nation with ever-growing middle class, it makes sense that a large number of stakeholders could benefit from this analysis.

DATA PREPARATION – Location information

- The latitude and longitude data of the different neighbourhoods in the city of Chennai:- https://chennaiiq.com/chennai/latitude_longitude_areas.asp
- Beautiful soup module in python is used to extract the information from the website and put it into a pandas dataframe.
- The latitude and longitude data is obtained in DMS system and is converted to decimals system for convenience and ease of usage.

	Neighborhood	latitude	longitude
1	Adyar Bus Debot.	12.9972	80.2569
2	Adyar Signal	13.0064	80.2575
3	Alandur	13.0078	80.2097
4	Ambattur	13.11	80.17
5	Anna Arch	13.0744	80.2183
...	...	...	...
101	Velachery Bus Terminus	12.9758	80.2208
102	Villivakkam	13.11	80.2033
103	Vyasar Padi	13.1092	80.2625
104	Washermanpet	13.1086	80.2811
105	Woodlands Drive In	13.05	80.2511

105 rows × 3 columns

DATA PREPARATION – Venue information

- The venue information is obtained from the foursquare website by making API calls.
- The url for the API calls involve the Client ID, client SECRET, the version date (upto date) , the latitude and longitude of the location, the radius about which the information is to be returned, the limit on number of instances to be returned and the endpoint(explore? Or categories? Or search? Etc)
- The API call looks as shown below:-

```
https://api.foursquare.com/v2/venues/categories?&client_id=GIL5KFTBUZBWMUCEWGMC1NISL2E1BFTEWROCHKPEJDBQVU2M&client_secret=TBEJZ
QQAZ4SCSEZ2WZIJ2GPTOMO4C1ZQ4FI5OC5YBSJKT450&v=20200404&ll=13.0801721,80.2838331&radius=1000&limit=100
```

- This API call is sent for each neighbourhood and the venue data returned is stored into a dataframe by using a for loop to extract the required informationfrom the venues dictionary

METHODOLOGY – ONE HOT ENCODING

The data is in the form of categorical information and cannot be fed into a clustering or machine learning algorithm easily. As a result, one hot encoding is performed to convert the information into n columns, where n corresponds to the number of categories and each column either has a 0 or 1.

When we find the mean of this column grouped about the neighbourhood data,we get the dataframe ready for clustering.

Neighborhood	Asian Restaurant	BBQ Joint	Bistro	Breakfast Spot	Buffet	Burger Joint	Chinese Restaurant	Coffee Shop	Dessert Shop	Fast Food Restaurant	...	Juice Bar	Kebab Restaurant	Middle Eastern Restaurant
AVM Studio	0.095238	0.047619	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.095238	...	0.000000	0.0	0.000000
Adyar Bus Debot.	0.090909	0.045455	0.0	0.045455	0.0	0.0	0.045455	0.000000	0.000000	0.090909	...	0.045455	0.0	0.045455
Adyar Signal	0.029412	0.000000	0.0	0.000000	0.0	0.0	0.029412	0.029412	0.058824	0.058824	...	0.058824	0.0	0.000000
Alandur	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.0	0.000000
Ambattur	0.000000	0.000000	0.0	0.000000	0.0	0.0	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.0	0.000000

METHODOLOGY – KMEANS CLUSTERING

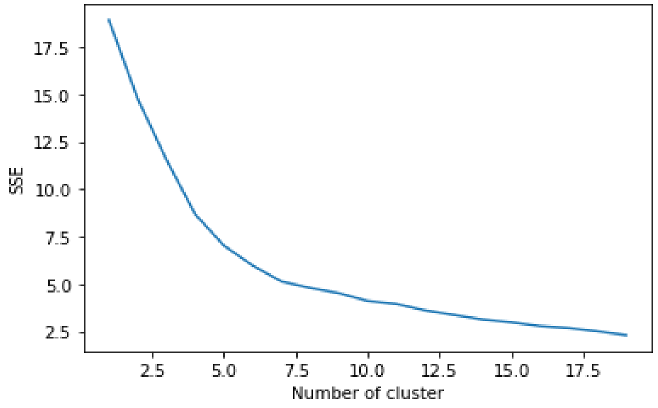
- The data is now clustered using an unsupervised learning algorithm called kmeans clustering. The data splits the information into k clusters based on the minimum sum of squares distance within clusters.

To find the ideal k value, we loop it for different values of k and choose the k at the elbow point of the graph of K vs SSE

```
sse_hf={}
sse_lf={}
for k in range(1,20):
    kmeans_hf=KMeans(n_clusters=k).fit(hf_clustering)
    kmeans_lf=KMeans(n_clusters=k).fit(lf_clustering)
    sse_hf[k]=kmeans_hf.inertia_
    sse_lf[k]=kmeans_lf.inertia_
```

k=7 is the elbow point for the high frequency data

```
plt.figure()
plt.plot(list(sse_hf.keys()), list(sse_hf.values()))
plt.xlabel("Number of cluster")
plt.ylabel("SSE")
plt.show()
```



METHODOLOGY – DISTRIBUTION ANALYSIS

- The restaurant distribution has been analysed and the following information is of great importance

- Restaurant categories

```
56

array(['Middle Eastern Restaurant', 'Asian Restaurant', 'Pizza Place',
       'Fast Food Restaurant', 'Indian Restaurant', 'Chinese Restaurant',
       'Diner', 'Sandwich Place', 'BBQ Joint', 'Breakfast Spot',
       'Juice Bar', 'North Indian Restaurant',
       'Vegetarian / Vegan Restaurant', 'Dessert Shop', 'Snack Place',
       'Mediterranean Restaurant', 'Italian Restaurant', 'Coffee Shop',
       'Comfort Food Restaurant', 'Cafeteria', 'South Indian Restaurant',
       'Food', 'Restaurant', 'Burger Joint', 'Bistro', 'Food Court',
       'Multicuisine Indian Restaurant', 'Korean Restaurant',
       'Rajasthani Restaurant', 'Steakhouse', 'Mexican Restaurant',
       'American Restaurant', 'Halal Restaurant', 'Bed & Breakfast',
       'African Restaurant', 'Kerala Restaurant', 'Thai Restaurant',
       'Buffet', 'Kebab Restaurant', 'Bengali Restaurant',
       'Japanese Curry Restaurant', 'Portuguese Restaurant',
       'Seafood Restaurant', 'Japanese Restaurant',
       'Molecular Gastronomy Restaurant', 'Modern European Restaurant',
       'Russian Restaurant', 'Hyderabadi Restaurant', 'Food Truck',
       'Andhra Restaurant', 'Food & Drink Shop', 'Fried Chicken Joint',
       'Burrito Place', 'Malay Restaurant', 'New American Restaurant',
       'Falafel Restaurant'], dtype=object)
```

Quantitative distribution

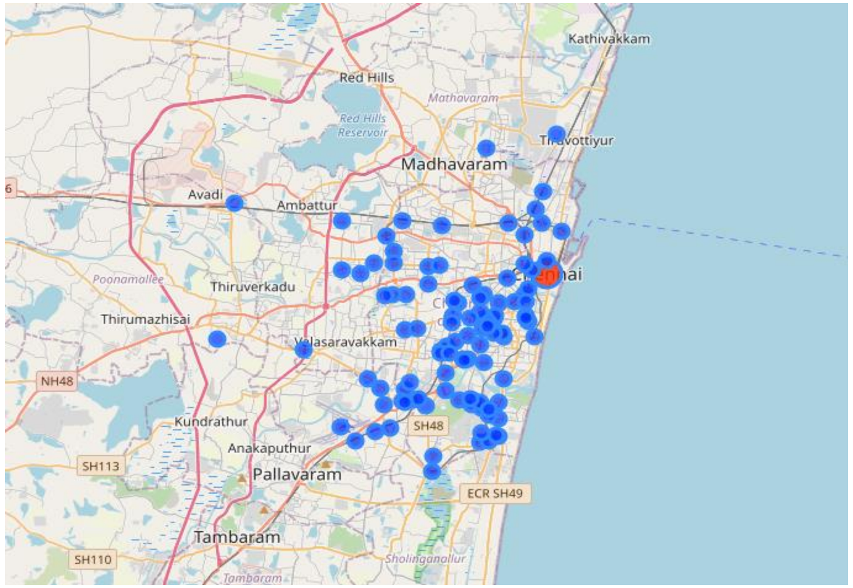
Venue Category	
Indian Restaurant	625
Chinese Restaurant	115
Pizza Place	92
Fast Food Restaurant	84
Coffee Shop	83
Restaurant	78
Sandwich Place	70
Vegetarian / Vegan Restaurant	68
Italian Restaurant	64
Asian Restaurant	63
Juice Bar	63
Dessert Shop	50
South Indian Restaurant	47
Middle Eastern Restaurant	46
BBQ Joint	33
Snack Place	28
Thai Restaurant	20
Breakfast Spot	15
Food Court	14

METHODOLOGY – CLUSTER REPRESENTATION

- The neighbourhoods with the respective clusters are shown on a folium map

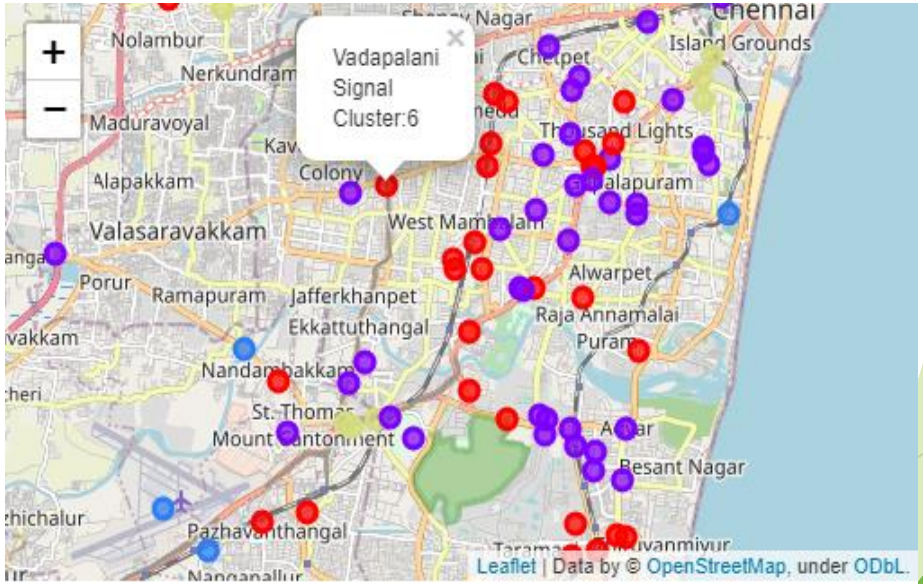
- Neighbourhoods

(city centre in red)



Cluster representation

(cluster labels shown in popup)



METHODOLOGY – CLUSTER ANALYSIS

- For example, cluster 6 has information displayed as shown below.

Cluster 6 - A very high incidence of **Vegetarian and Thai restaurants** and a high incidence of **BBQ and bistro joints**

Low incidence of **Pizza places**, **snacking places**, and **exotic cuisines(apart from thai)**

```
hf_rest_sorted.loc[hf_rest_sorted['Cluster Labels']==6,:]
```

	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	11th Most Common Venue
42	Korukkupet	6	Vegetarian / Vegan Restaurant	Thai Restaurant	BBQ Joint	Bistro	Breakfast Spot	Buffet	Burger Joint	Chinese Restaurant	Coffee Shop	Dessert Shop	Fast Food Restaurant
68	Poonamallee	6	Vegetarian / Vegan Restaurant	Thai Restaurant	BBQ Joint	Bistro	Breakfast Spot	Buffet	Burger Joint	Chinese Restaurant	Coffee Shop	Dessert Shop	Fast Food Restaurant
78	TVS Signal	6	Vegetarian / Vegan Restaurant	Thai Restaurant	BBQ Joint	Bistro	Breakfast Spot	Buffet	Burger Joint	Chinese Restaurant	Coffee Shop	Dessert Shop	Fast Food Restaurant
90	Tondiarpet	6	Vegetarian / Vegan Restaurant	Thai Restaurant	BBQ Joint	Bistro	Breakfast Spot	Buffet	Burger Joint	Chinese Restaurant	Coffee Shop	Dessert Shop	Fast Food Restaurant

The above table shows the restaurant distribution for neighbourhoods in cluster 6 and it shows a clear excess of vegan and thai restaurants and a clear lack of pizza places and snacking places. (The end columns aren't present in the screenshot but have been taken into consideration for the analysis).

METHODOLOGY – CLUSTER ANALYSIS

- Similarly, all the clusters were analysed and the information has been combined into a table as shown below.

	Cluster Labels	High Frequency locations	Low Frequency Locations
0	0	Indian followed by Pizza & Chinese	BBQ, asian, Cafe & coffe shops
1	1	Fast food restaurants followed by vegan restaurants	Indian restaurants followed by exotic cuisines
2	2	Indian followed by Dessert & Breakfast shops	South Indian, Snacking & Sandwich places, Pizza places, fast food restaurants & Italian & thai restaurants
3	3	Pizza place followed by vegetarian & BBQ joints	Exotic restaurants
4	4	Not statistically significant	Not statistically significant
5	5	Indian followed by Chinese and BBQ joints	Non chinese asian, burger, buffet, snacking
6	6	Vegetarian & Thai followed by Vegan & BBQ joints	Pizza places, snacking places & exotic cuisines

RESULT, DISCUSSION AND CONCLUSION

- The analysis focuses on retrieving all the venues in Chennai, accurately filtering out the restaurants and clustering them to analyse the neighbourhoods
- The analysis showed an exceptionally high number of Indian restaurants and a reasonably high number of Chinese restaurants
- the few major areas that had a major lack of Indian restaurants were identified which could be crucial for stakeholders interested in entering the Indian restaurant game.
- For a future design, the data could be fed into an app or platform where based on the type of restaurant or choice of area, the cluster data, similar neighbourhoods, distribution of restaurants etc can be easily obtained.
- Final decision on optimal restaurant location will be made by stakeholders based on specific characteristics of neighbourhoods and locations in every recommended zone, taking into consideration additional factors like attractiveness of each location.