Choosing the most appropriate district for starting a Multiplex in The Most Densely Populated District of Tamilnadu, India

Meenakshi S Arya

Background

TamilNadu is one of the most popular states in the South India as it is the homeground of the Tamil movie industry, popularly known as "Tollywood".

People in Tamilnadu are extremely fond of watching movies and dining out.

Thus, it is an opportune candidate for setting up a Multiplex with movie theatres and eating joints.

These multiplexes are great business ideas as they provide a one-stop solution to all the needs of customers, ranging from household groceries, to beauty products to dining options to options for entertainment.

Problem Statement

Tamilnadu majorly comprises of 37 districts however, the population of a district will determine the feasibility of starting a new multiplex at a given location.

The project aims to predict which district will be most suitable for the same considering the overall population and the number of eating joints and multiplexes in a given district.

The battle of these neighborhoods will be unrevealed in a detailed manner in the subsequent parts of the implementation.
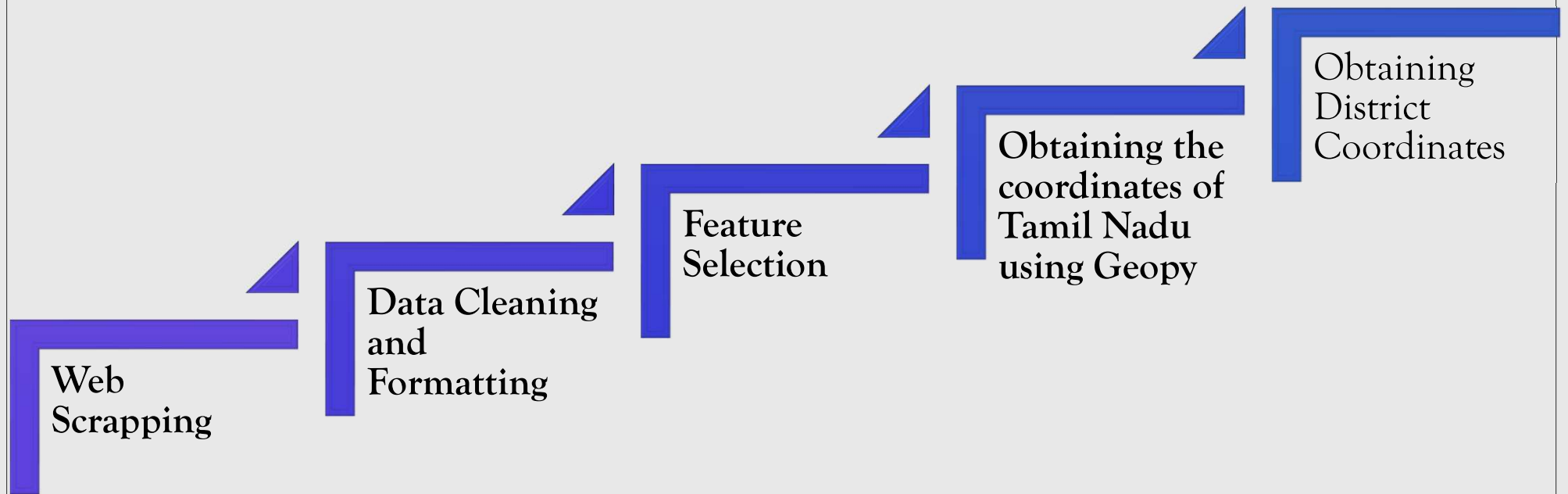
Dataset Description

The dataset available in wikipedia for <u>districts of Tamilnadu</u> will be used for obtaining the relevant information about the various districts in Tamilnadu, their area, population and talukas within each district.

This data along with Foursquare location data will be used for obtaining the information about the number of multiplexes in the most populous districts of Tamilnadu.
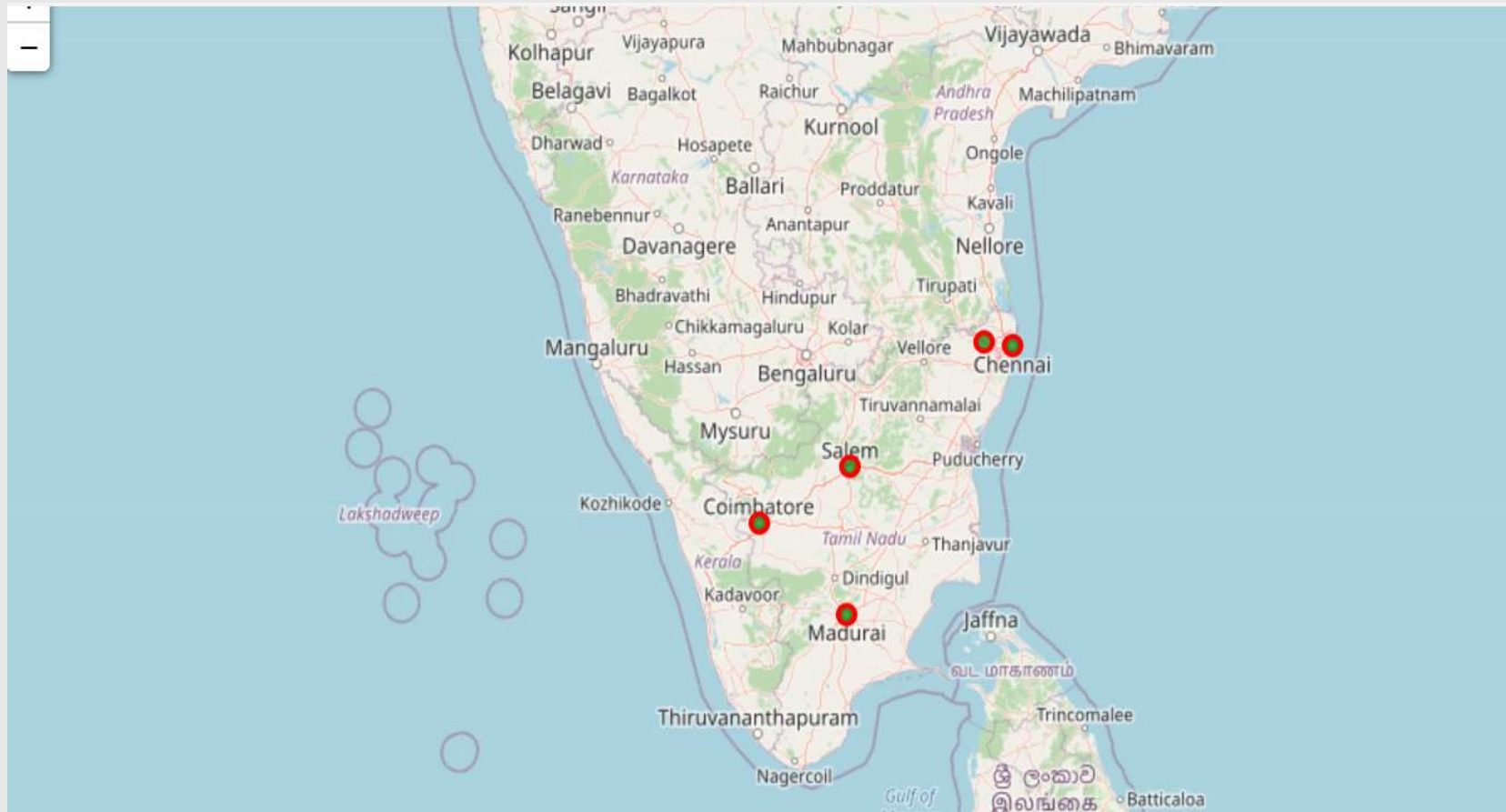
Flow of the process

Web Scrapping

Data Cleaning and Formatting

Feature Selection

Obtaining the coordinates of Tamil Nadu using Geopy

Obtaining District Coordinates

The final look of the dataset

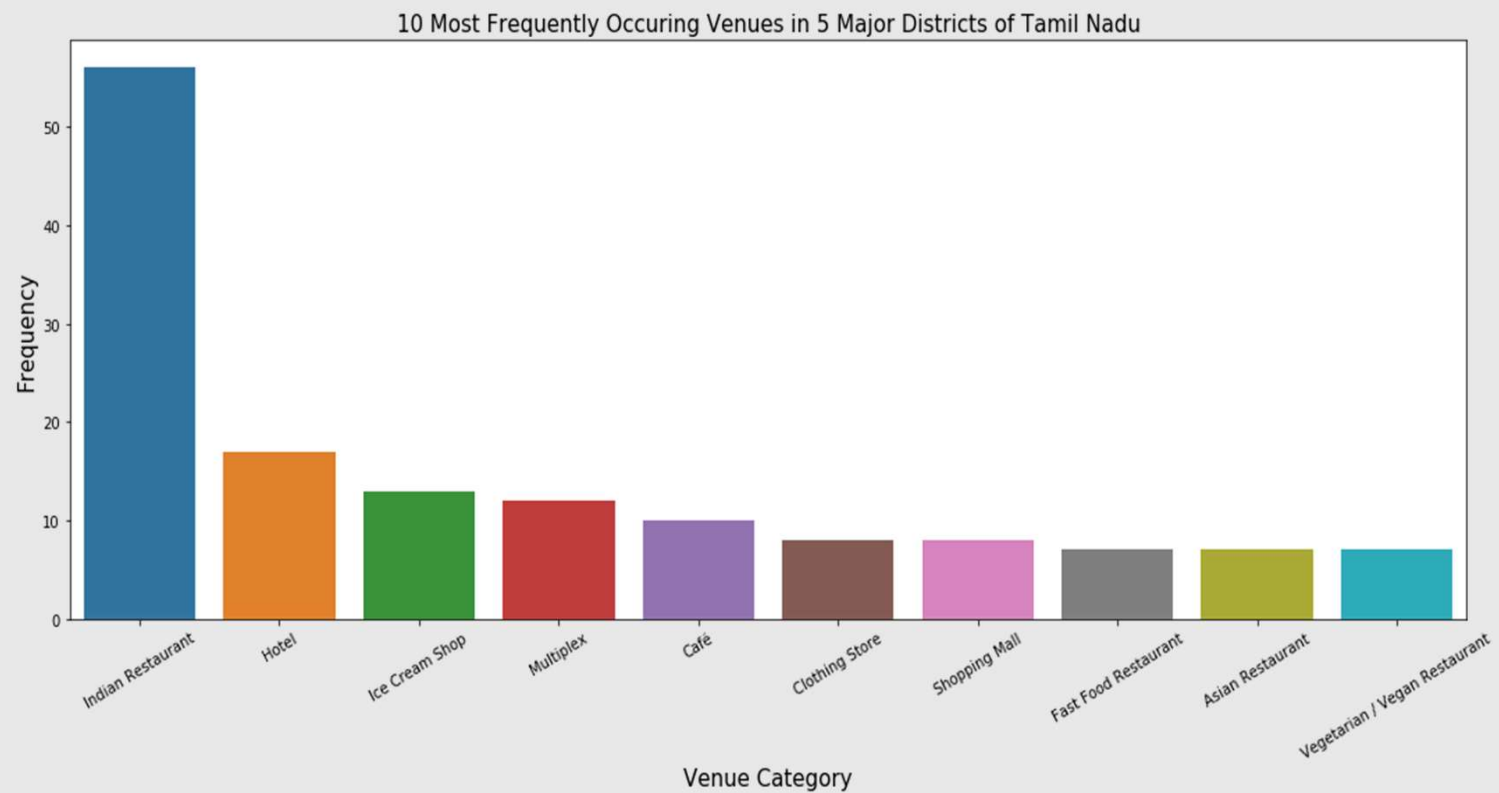|	District	Population	Talukas	Dist_Latitude	Dist_Longitude	Area
2	Ariyalur	754894.0	Andimadam, Ariyalur, Udayarpalayam, Sendurai	11.076036	79.117455	1949.31
3	Chengalpattu	2556244.0	Chengalpattu, Cheyyur, Madurantakam, Pallavara...	12.766574	79.999319	2944.96
4	Chennai	4646732.0	Alandur, Ambattur, Aminjikarai, Ayanavaram, Eg...	13.080172	80.283833	426.00
5	Coimbatore	3458045.0	Anaimalai, Annur, Coimbatore-North, Coimbatore...	11.001812	76.962842	NaN
6	Cuddalore	2605914.0	Bhuvanagiri, Chidambaram, Cuddalore, Kattumann...	11.742694	79.750306	3678.00
7	Dharmapuri	1506843.0	Dharmapuri, Palacode, Pennagaram, Harur, Pappi...	12.096805	78.193043	4497.77
8	Dindigul	2159775.0	Athoor, Dindigul-West, Dindigul-East, Gujiliam...	10.330330	78.067398	6266.64
9	Erode	2251744.0	Anthiyur, Bhavani, Erode, Gobichettipalayam, K...	11.369204	77.676627	NaN
10	Kallakurichi	1370281.0	Kallakkurichi, Thirukoilur, Kalvarayan Hills, ...	11.794685	79.038821	3520.37
11	Kanchipuram	1166401.0	Kanchipuram, Sriperumbudur, Uthiramerur, Walaj...	12.964716	79.983969	1655.94
12	Kanyakumari	1870374.0	Agastheeswaram, Kalkulam, Vilavancode, Thovala...	8.079252	77.549934	1672.00
13	Karur	1064493.0	Aravakurichi, Kadavur, Karur, Krishnarayapuram...	10.930152	78.084855	2895.57
14	Krishnagiri	1879809.0	Krishnagiri, Hosur, Pochampalli, Uthangarai, D...	12.518883	78.220654	5143.00
15	Madurai	3038252.0	Madurai-North, Madurai-South, Madurai-West, Ma...	9.926115	78.114098	3741.73
16	Mayiladuthurai	918356.0	Mayiladuthurai, Sirkazhi, Tharangambadi, Kuthalam	11.155182	79.627394	1172.00
17	Nagapattinam	697069.0	Kilvelur, Nagapattinam, Thirukkuvalai, Vedaranyam	10.805628	79.824660	1397.00
18	Namakkal	1726601.0	Kollimalli, Mohanur, Namakkal, Paramathi-Velur...	11.219132	78.237398	3363.00
19	Nilgiris	735394.0	Coonoor, Gudalur, Kotagiri, Kundah, Pandalur, ...	11.400000	76.700000	2452.50
20	Perambalur	565223.0	Alathur, Kunnam, Perambalur, Veppanthattai	11.228772	78.818256	1752.00
21	Pudukkottai	1618345.0	Alangudi, Aranthangi, Avadaiyarkoil, Gandarvak...	10.500000	78.833333	4663.00
22	Ramanathapuram	1353445.0	Kadaladi, Kamuthi, Kilakarai, Manamelkudi, Mud...	9.389552	78.859071	4089.57
23	Ranipet	1210277.0	Arakkonam, Arcot, Nemili, Walajapet, Sholingur...	12.927264	79.333008	2234.32
24	Salem	3482056.0	Attur, Idappadi, Gangavalli, Kadyampatti, Mett...	11.664300	78.146000	5205.00
25	Sivagangai	1339101.0	Devakottai, Ilayangudi, Kalayarkoil, Karaikudi...	9.965060	78.720428	4086.00
26	Tenkasi	1407627.0	Alankulam, Kadayam, Kadayanallur, Keelapavoor,...	9.094075	77.475837	2916.13
27	Thanjavur	2405890.0	Budalur, Kumbakonam, Orathanadu, Papanasam, Pa...	10.786027	79.138150	3396.57
28	Theni	1245899.0	Theni, Periyakulam, Andipatti, Bodniayakkanur,...	9.969664	77.474200	3066.00
29	Thoothukudi	1750176.0	Thoothukudi, Tiruvaikuntam, Kovilpatti, Ottapi...	8.805260	78.145274	4621.00
30	Tiruchirappalli	2722290.0	Lalgudi, Manachanallur, Manapparai, Marungapur...	10.804973	78.687030	4407.00
31	Tirunelveli	1665253.0	Ambasamudram, Nanguneri, Palayamkottai, Manur,...	8.808234	77.811484	3842.37
32	Tirupattur	1111812.0	Natrampalli, Tirupattur, Vaniyambadi, Ambur	12.498356	78.561817	1797.92
33	Tiruppur	2479052.0	Avinashi, Palladam, Dharapuram, Kangeyam, Mada...	10.783227	77.526048	5186.34
34	Tiruvallur	3728104.0	Avadi, Gummidipoondi, Pallipattu, Ponneri, Poo...	13.139436	79.907304	3424.00
35	Tiruvannamalai	2464875.0	Aarani, Chengam, Chetpet, Cheyyar, Jamunamarat...	12.227213	79.070156	6191.00
36	Tiruvarur	1264277.0	Kudavasal, Koothanallur, Mannargudi, Nannilam,...	10.774598	79.627972	2161.00
37	Vellore	1614242.0	Anaicut, Gudiyatham, Pernambut, Katpadi, Vello...	12.794811	79.000641	2030.11
38	Viluppuram	2093003.0	Gingee, Marakkanam, Melmalaianur, Tindivanam, ...	11.913787	79.507893	3725.54
39	Virudhunagar	1942288.0	Aruppukkottai, Kariapatti, Rajapalayam, Sattur...	9.520894	77.878456	4288.00

Using Foursquare API to explore Tamil Nadu for the number of multiplexes in the top 5 most populated districts

The five most populous districts visualized
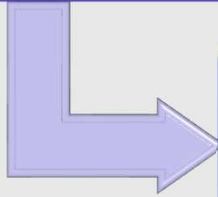
Exploring the most popular venues in 5 topmost districts

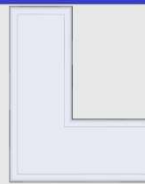10 Most Frequently Occuring Venues in 5 Major Districts of Tamil Nadu

The winner is Indian Restaurants

A typical south Indian thali

Obtaining information about the top 5 venues of each district

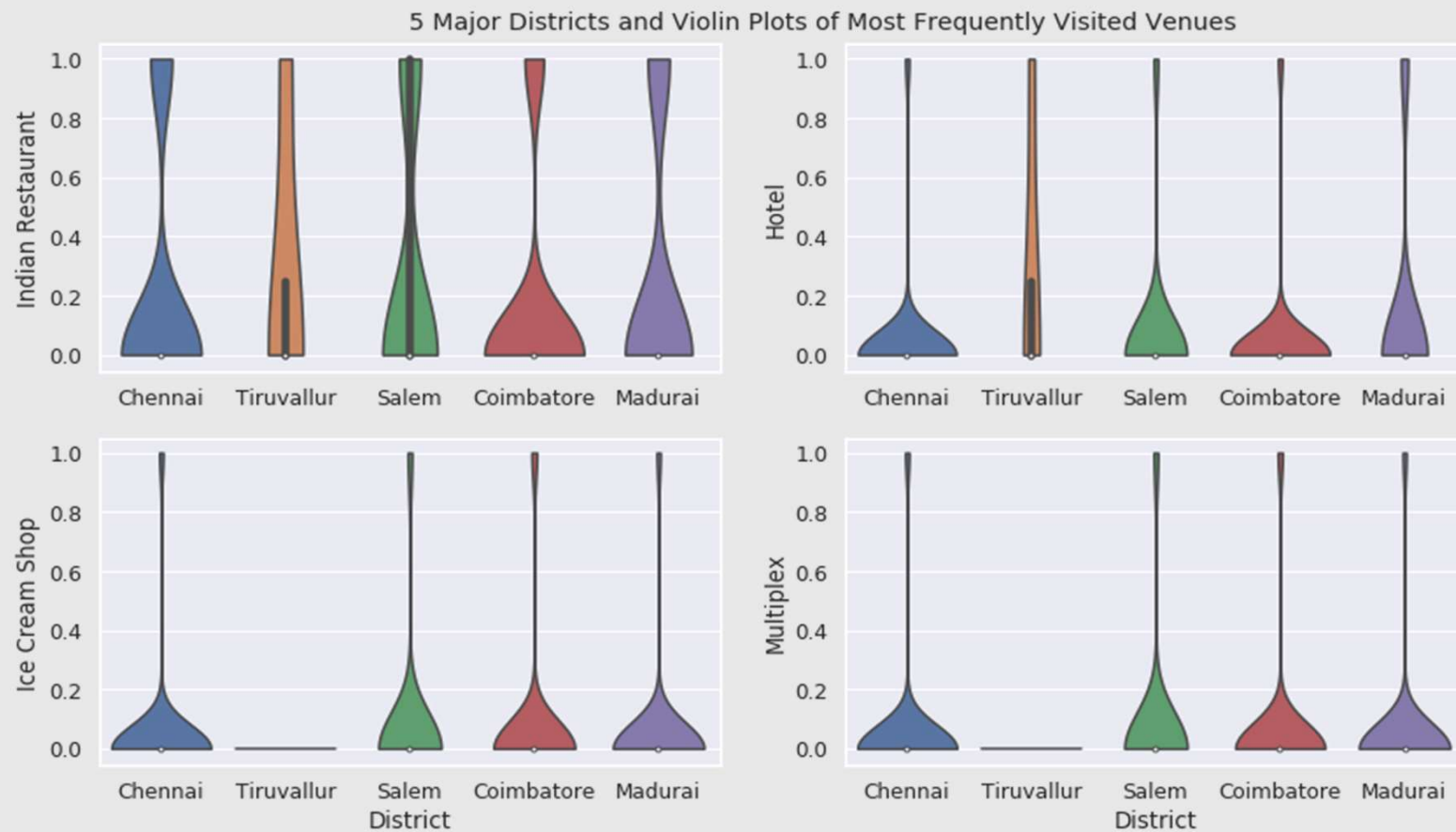Create a data-frame with pandas one hot encoding for the venue categories

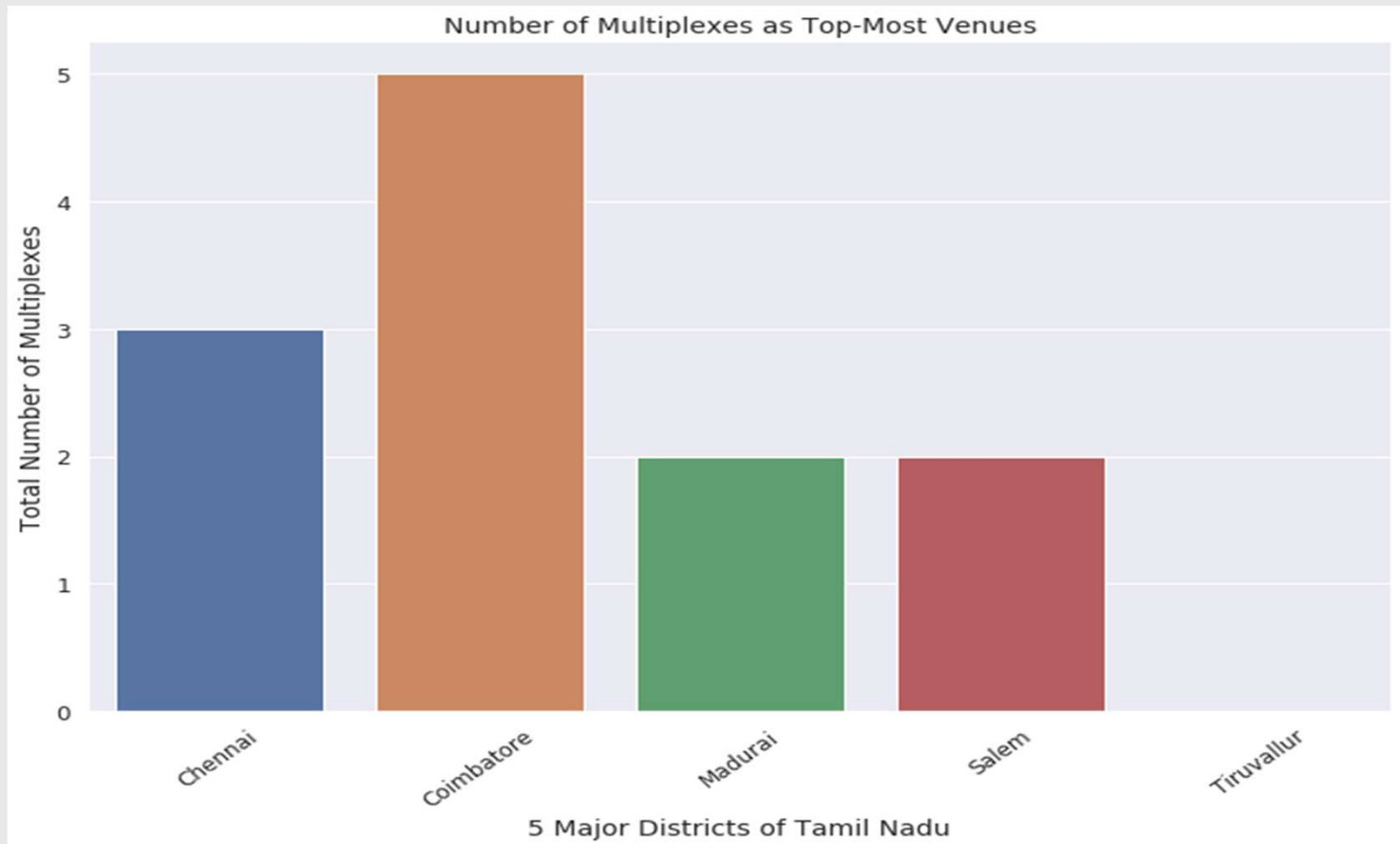Use pandas groupby on District column and obtain the mean.

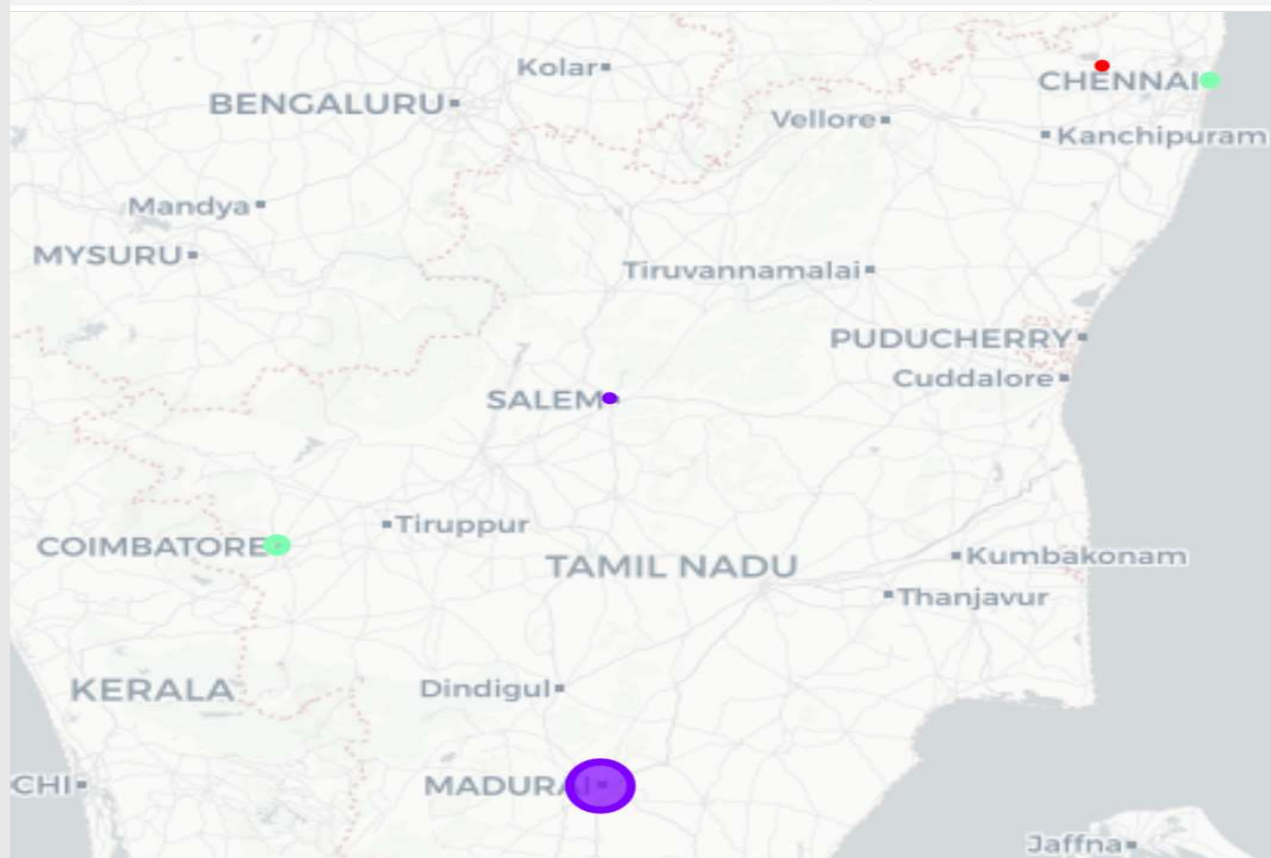Transpose the data-frame at and arrange in descending order.

Plot of most frequently Visited Venues in Top 5 Districts

5 Major Districts and Violin Plots of Most Frequently Visited Venues

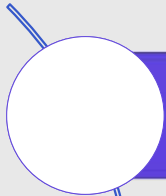# *Number of Multiplexes as Top Venues in Topmost 5 Districts*

Using K-Means Clustering for clustering the Tamil Nadu Districts (Neighborhoods)
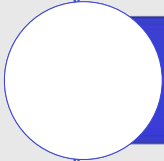
Coimbatore maximum number of multiplexes (green cluster), Chennai (red cluster) followed by Madurai and Salem having same number of multiplexes (purple cluster).
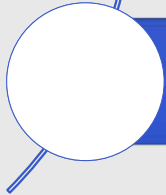
Results and Discussion

Indian restaurants top the charts of most common venues in the 5 districts.

Coimbatore district has the maximum number of multiplexes as the most popular venues.

Despite being thickly populated, Tiruvallur does not have any multiplex whereas this district has the maximum number of Indian restaurants.

Tiruvallur district is the most appropriate district for opening a new multiplex to provide people with a weekend getaway and an avenue to watch movies as well as eat out.

Conclusion

The project was a real hands-on experience for applying the concepts of software engineering.

Right from the conceptualizing of the idea to executing it and then finally drawing conclusions from it for a major business decision has really been informative.

Scrapping data from the web, cleaning it and then applying various visualization tools to it so that it can make sense and communicate the information required was really a great learning experience.

Also it helped in understanding the usage of some frequently used python libraries to scrap web-data, Foursquare API to explore the major districts of Tamil Nadu and visualizing the results of segmentation of districts using Folium leaflet map.

Potential for this kind of analysis in a real-life business problem is discussed in great detail.

Since the entire analysis was mostly concentrated on the possibilities of opening a multiplex targeting the fact that people are fond of going out and eating out, some of the results obtained are contrary to what the expectations were.

Finally to conclude this project provided a true glimpse of how real life data-science projects look like

Thank you!!!!!