Capstone Project

Housing Sales Rate & Neighborhoods Analysis of Chennai, India

BY JITENDRA K SINGH

Introduction and Problem Statement

Chennai – A major metropolitan city of India.

Area – 426 square km and Population – 8.6 million

Many neighbourhoods and each of these neighbourhoods offer vide variety of amenities to people.

It is a tough choice for buyers and investors to make a decision where do they want to put their money as a huge number of neighbourhoods make it a daunting task.

As a solution we will cluster these neighbourhoods based on *available amenities, number of housing societies, average rate of housing.*

Data Acquisition

For the project the data was captured and used for various tasks as

I will utilize BeautifulSoup library to extract relevant information for housing societies rate from [here](#)

The Foursquare API will be utilized to get the top venues for these neighborhoods

K-Means library and algorithm will be utilized to cluster these neighborhoods based on top venues and average rate of housing

Folium library will be utilized to create a map of Chennai with marker of clustered neighborhoods

Web Scraping and Data Cleaning

The data on website was captured and cleaned to get the desired dataframe

Society Name ⬍	Buy Rates			
Filter societies	Price Range (Rs. per sq. ft.)	Q/Q ⬍ ⓘ	Trends	Properties
Chennai West				
Akshaya Republic, Kovur	Rs. 4,972 - 5,398/sq. ft.	6.9% ↑	See Trends	Not Available
Altis Ashraya, Mangadu	Rs. 4,378 - 4,462/sq. ft.	5% ↑	See Trends	Not Available
Arihant Tiara, Nandambakkam	Rs. 6,375 - 6,375/sq. ft.	-	Not Available	Not Available
Bharathi Brikhouse, Vanagaram	Rs. 5,185 - 5,610/sq. ft.	6.5% ↑	See Trends	Not Available
Dugar Sky City, Vanagaram	Rs. 5,185 - 6,078/sq. ft.	-	Not Available	Not Available
Fomra Hues, Porur	Rs. 5,185 - 6,162/sq. ft.	0%	See Trends	Not Available
Golden Opulence, Poonamallee	Rs. 4,208 - 5,058/sq. ft.	-	Not Available	Not Available
Jains Westminster, Saligramam	Rs. 7,778 - 9,138/sq. ft.	-	Not Available	Not Available

Data from web converted to data frame using BeautifulSoup and python codes

	Neighbourhood	Number_of_societies	Average_min	Average_max	Mean_rate
0	Adambakkam	1	7820.00	9392.00	8606.0
1	Adyar	1	13515.00	13898.00	13706.5
2	Agraharam	1	4420.00	5015.00	4717.5
3	Alandur	1	6672.00	6672.00	6672.0
4	Ambattur	4	4441.25	4908.75	4675.0

Geo-coordinates and Map

Nominatim was used to get geo-coordinates for each of neighborhood and plot map using folium for each of the neighborhood

	Neighbourhood	Number_of_societies	Average_min	Average_max	Mean_rate	geo_loc	Latitude	Longitude
0	Adambakkam	1	7820.00	9392.00	8606.0	(Adambakkam, Ward 177, Zone 13 Adyar, Chennai,...	12.982221	80.209121
1	Adyar	1	13515.00	13898.00	13706.5	(Adyar, Ward 176, Zone 13 Adyar, Chennai, Chen...	13.006450	80.257779
2	Agraharam	1	4420.00	5015.00	4717.5	(Agraharam, Yelamanchili, Visakhapatnam, Andhr...	17.568600	82.850391
3	Alandur	1	6672.00	6672.00	6672.0	(Alandur, Tamil Nadu, India, (13.00282155, 80....	13.002822	80.171919
4	Ambattur	4	4441.25	4908.75	4675.0	(Ambattur, Thiruvallur District, Tamil Nadu, I...	13.112886	80.159862

Neighborhood data with geo-coordinates
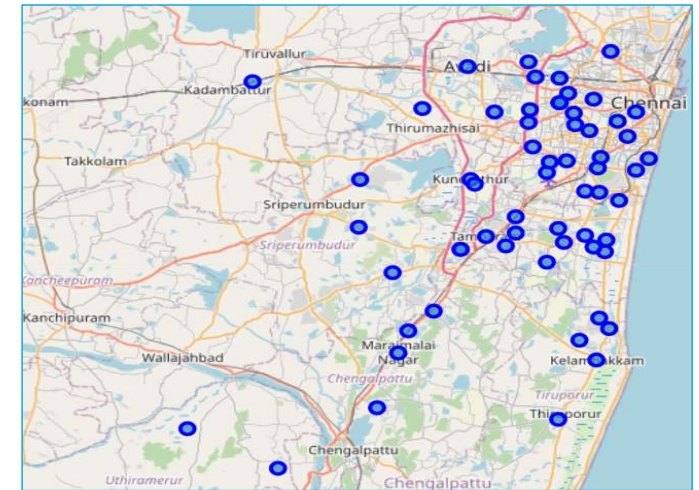
Top Venues

Foursquare API was used to extract nearby venues of each neighbourhood and then top 10 venues were listed for each of these neighborhood

Number of venues in Chennai is 362

	Neighborhood	Neighborhood latitude	Neighborhood longitude	Venue name	Venue latitude	Venue longitude	Venue category
0	Adambakkam	12.982221	80.209121	arun icecream	12.983447	80.207847	Dessert Shop
1	Adambakkam	12.982221	80.209121	MedPlus	12.980940	80.207356	Pharmacy
2	Adambakkam	12.982221	80.209121	Sutherland	12.981002	80.205200	IT Services
3	Adambakkam	12.982221	80.209121	Bistro	12.983193	80.205020	Indian Restaurant
4	Adyar	13.006450	80.257779	Bombay Brassiere	13.006961	80.256419	North Indian Restaurant

Neighbourhood venues from Foursquare API converted to top 10 venues for each neighborhood

	Neighbourhood	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
0	Adambakkam	IT Services	Pharmacy	Indian Restaurant	Dessert Shop	Food Service	Convenience Store	Cupcake Shop	Department Store	Electronics Store	Farmers Market
1	Adyar	Indian Restaurant	Electronics Store	North Indian Restaurant	Ice Cream Shop	Rock Club	Grocery Store	Vegetarian / Vegan Restaurant	Italian Restaurant	Juice Bar	Lounge
2	Alandur	Airport Service	Whisky Bar	Gastropub	Convenience Store	Cupcake Shop	Department Store	Dessert Shop	Electronics Store	Farmers Market	Fast Food Restaurant
3	Ambattur	Multiplex	Movie Theater	River	Whisky Bar	Coffee Shop	Convenience Store	Cupcake Shop	Department Store	Dessert Shop	Electronics Store
4	Anna Nagar	Indian Restaurant	Fast Food Restaurant	Electronics Store	Chinese Restaurant	Café	Park	Bistro	Vegetarian / Vegan Restaurant	Middle Eastern Restaurant	Department Store

Clustering

Top ten venues, number of societies and average rate in each neighborhood were used to cluster neighborhoods using K-Means clustering method. Elbow method was used to define the value of optimum K(number of clusters)
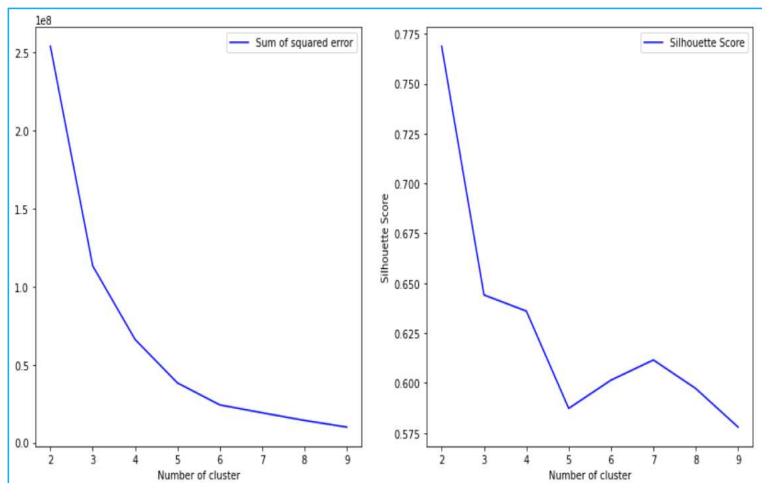
Elbow method outcome

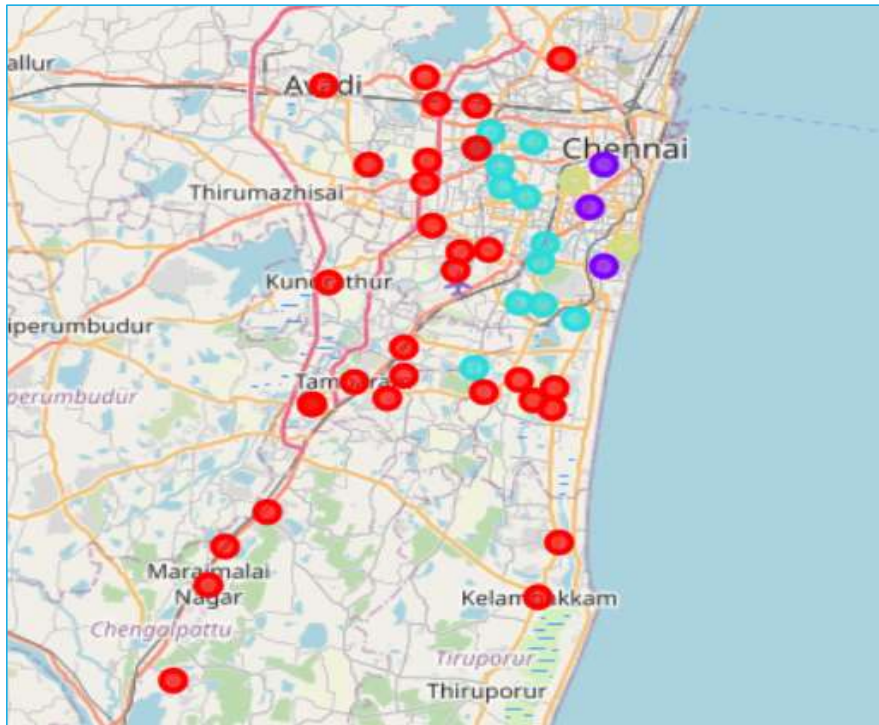|	Neighbourhood	Number_of_societies	Mean_rate	ATM	Afghan Restaurant	Airport Service	American Restaurant	Antique Shop	Art Gallery	Asian Restaurant	...	Spa	Sports Club	Tattoo Parlor	Tea Room	Thai Restaurant
0	Adambakkam	1	8606.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0
1	Adyar	1	13706.5	0.0	0.0	0.0	0.0	0.0	0.0	0.034483	...	0.0	0.0	0.0	0.0	0.0
2	Alandur	1	6672.0	0.0	0.0	1.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0
3	Ambattur	4	4675.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0
4	Anna Nagar	1	9010.0	0.0	0.0	0.0	0.0	0.0	0.0	0.043478	...	0.0	0.0	0.0	0.0	0.0

5 rows × 110 columns

Data for Clustering

Visualise on Map

Neighbourhoods were plotted in map with colours identifying each cluster.

Distance from Center

Geodesic library was used to create a data frame of each neighbourhood with distance from center.

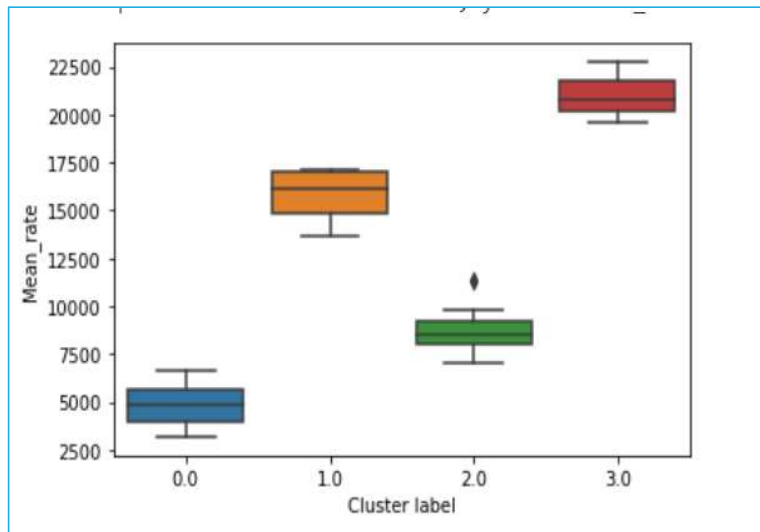|	Neighbourhood	Number_of_societies	Mean_rate	Latitude	Longitude	Cluster label	Distance_from_center
0	Adambakkam	1	8606.0	12.982221	80.209121	2.0	13.034431
1	Adyar	1	13706.5	13.006450	80.257779	1.0	8.650870
2	Alandur	1	6672.0	13.002822	80.171919	0.0	13.916017
3	Ambattur	4	4675.0	13.112886	80.159862	0.0	12.391949
4	Anna Nagar	1	9010.0	13.087200	80.216442	2.0	5.841288

Result

Number of neighbourhoods in each cluster

Cluster 1 – 33 neighborhoods ; Cluster 2 – 4 neighborhoods

Cluster 3 – 12 neighborhoods ; Cluster 4 – 3 neighborhoods

Scatter plot Average rate vs Distance from centre

Box plot of average rate in each cluster

Correlation between average rates and distance from centre
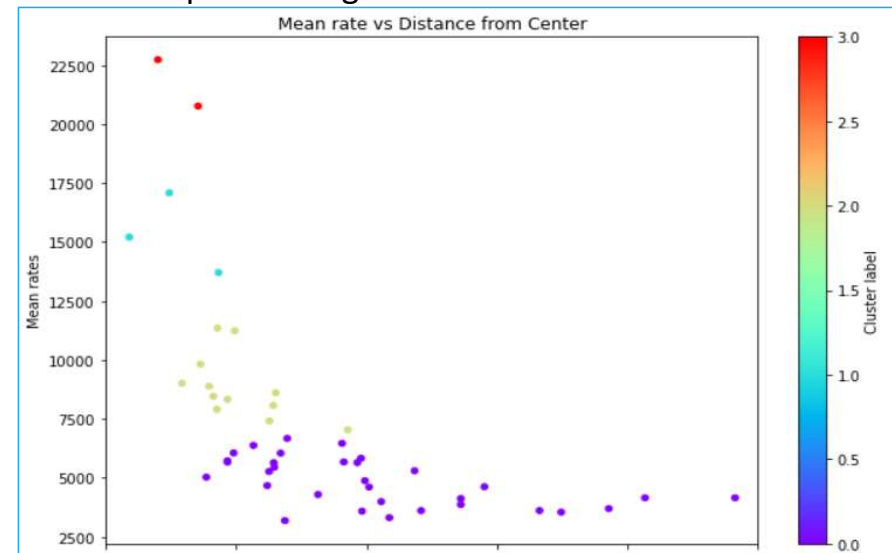
	Distance_from_center	Mean_rate
Distance_from_center	1.000000	0.357453
Mean_rate	0.357453	1.000000

Discussion

Clusters are ranked based on average rates and distance from center.

Rank 1 is highest and rank 4 is lowest

Cluster Label	Mean_Rate Rank	Distance_from_Center rank	No. of Neighbourhoods
Cluster 4	1	4	3
Cluster 2	2	3	4
Cluster 3	3	2	12
Cluster 1	4	1	33

Conclusion

The clusters are now defined based on their features which will help the buyer and investor to have a quick glimpse and make better decisions

Cluster Label	Definition
Cluster 1	Lowest average rates, farthest from center with highest neighborhoods
Cluster 2	High average rates, closer to center with low neighborhoods
Cluster 3	Low average rates, far from center with higher neighborhoods
Cluster 4	Highest average rates, closest to center with least neighborhoods

Availability of more data can improve results and estimate the housing prices in a given neighbourhoods so that one can make better investment decision.

Future Direction

With better data that has more information of neighbourhood venues, type of housing complex, etc, clustering can be further improved.

Average rates of housing can be estimated with availability of more data and deploying regression models.

It can be explored why there are more neighbourhoods in the far distance from city centre, could it be because of lower rates in those neighbourhoods.

Can data of average income of residents in city centre neighbourhood and outskirts be correlated to the number of housing complex in their neighbourhoods.

Thank you