

Capstone Project Report

FINDING BEST AREAS IN CHENNAI

EDWIN ROOSEVELT

PROBLEM STATEMENT

- Find the best areas to live in chennai.
 - The availability of jobs attracts People. This project aims to help people make smart and efficient choices in selecting the right areas to live in Chennai.
-

DATA

Data Sources

- List of Neighborhoods in chenani is obtained from *wikipedia*.
 - Coordinates (latitude and longitude) data was obtained using *geoPy* api.
 - Each neighborhood area was explored using *Foursquare* api.
-

DATA

Data Cleaning

- Spelling mistakes in the names of the areas had to be manually renamed.
 - Areas too close to each other were removed and only one area was left to remain.
-

METHODOLOGY

- Coordinates of 131 areas were obtained using geoPy api.
 - Each area was explored for different venue categories using foursquare api.
 - 1953 venues were obtained.
 - Venues were grouped into each areas.
 - Venue categories were grouped and irrelevant columns were dropped out.
-

METHODOLOGY

- K-means clustering model was instantiated using the venue and location data.
 - The model clustered the areas into five categories.
-

	Latitude	Longitude	Area	School	Hospital	Medical Center	Grocery Store	Convenience Store	Department Store	Hardware Store	...
0	12.982221	80.209121	Adambakkam	4	4	1	0	0	2	0	...
1	13.006450	80.257779	Adyar	0	7	0	0	1	0	0	...
2	13.049901	80.165435	Alapakkam	0	2	0	0	0	0	0	...
3	13.033860	80.254549	Alwarpet	0	0	0	0	0	0	0	...
4	13.047486	80.183669	Alwarthirunagar	0	0	0	0	1	1	0	...
...
126	13.047543	80.192218	Virugambakkam	0	3	0	0	0	1	0	...
127	13.116973	80.257124	Vyasarpadi	2	3	0	0	0	0	0	...
128	13.111656	80.284745	Washermenpet	1	10	0	0	0	0	0	...
129	13.042732	80.227078	West Mambalam	10	8	0	1	0	1	0	...
130	13.014753	80.174494	manapakkam	1	0	0	0	0	0	0	...
131 rows × 22 columns											

RESULT

Cluster 1

- These areas form the heart of chennai.
- They have all the amenities but the cost of living is high.

Cluster 2

- All the necessary venues are present in this cluster.
 - It is spread on the outskirts of the city.
-

RESULT

Cluster 3

- These areas act as a transition between the heart and outskirts of the city.

Cluster 4

- Unable to infer any correlation.

Cluster 5

- Similar to cluster 1 with abundant restaurants (commercial zone)
-

CONCLUSION

Areas in cluster 2 and cluster 3 are the most favourable.

- These areas are widespread.
 - Almost all the amenities are available in the areas of these clusters.
-

The End
