Recommending Location to Open a Hotel in Chennai, India

by

Gregory Okechukwu Offodum

February, 2021

INTRODUCTION

Have you ever wondered about how to take the best decision in siting a business venture ?

This project investigates and studies the tourism industry; the considerations to take when going into the hotel business so as to maximize profit and satisfy customers. The case study deals with Chennai area, India.

Problem Statement

- This project aims to analyze, infer and recommend the best suited location to open a new hotel in Chennai, India.

DATA DESCRIPTION

- The project utilizes dataset which we scrapped from wikipedia, with a total of 67 neighborhoods, comprising of latitude and longitude, zip codes.
- Link to the data used: https://en.wikipedia.org/wiki/Category:Suburbs_of_Chennai

- Foursquare API was utilized as the data source. Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined.

- Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

LIBRARIES

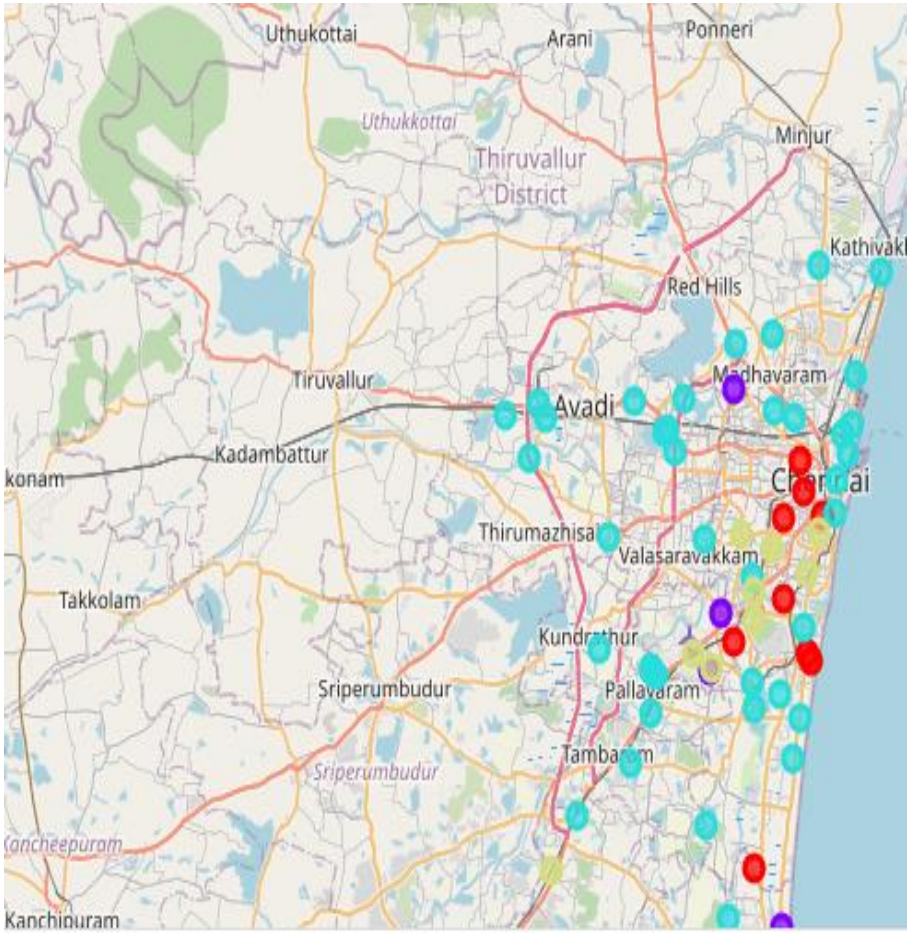
The following libraries were used in the project:

- Pandas: For creating and manipulating dataframes.
- Scikit Learn: For importing k-means clustering.
- Matplotlib: Python Plotting Module.
- JSON: Library to handle JSON files.
- XML: To separate data from presentation and XML stores data in plain text format.
- Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.
- Geocoder: To retrieve Location Data.

Techniques Used

- Exploratory data analysis to understand dataset
- Preprocessing and cleaning of dataset
- Best feature selection
- Machine learning and model development
- K-Means Clustering technique
- Cluster Analysis and selection of best suited location

Results



- The areas were clustered into four based on frequency.
  - First Cluster, 0: Red
  - Second Cluster, 1: Purple
  - Third Cluster, 2: Light Blue
  - Fourth Cluster, 3: Yellow
- The second and third clusters had the highest and lowest concentration respectively

Observations

Cluster no	Cluster	Suitability	Description
0	First	Maybe	Moderate availability of hotels
1	Second	No	Highest number of hotels around
2	Third	Yes	Lowest dominance of hotels
3	Fourth	Maybe	Moderate availability of hotels

Recommendations for a better project

- Larger dataset
- Multiple model deployment and interpretation
- Consulting expertise from real estate professionals

Conclusion

It can be inferred that the third Cluster is the best suited location to open a new hotel by virtue of the relatively low number of hotels in these areas.

THANK YOU