

CAPSTONE PROJECT

IDENTIFYING LOCATION FOR THE NEW MALL IN MUMBAI



BUSINESS PROBLEM

- Shopping malls are a key source of entertainment in India. Most of the families in India visit a mall during weekends or holidays to spend some quality time by indulging in offerings of a typical mall which includes movie screens, restaurants, shopping, game zone etc.
- As income level are rising in India, the number of people who are going to the mall and being able to spend money in the mall is increasing rapidly.
- Therefore, a lot of real estate companies like DLF have invested money in the commercial malls as it generates great income for them. Key factor determining the success or failure of this investment is the location of the mall.



TARGET AUDIENCE

- Property developers/investors: Companies/individuals looking to build a new commercial mall in Mumbai would find this work useful as it would help them in identifying the potential location for their investment.
- Commercial Entities: This could be used by commercial entities like retail brands, restaurants to identify attractive locations to open an outlet in the city

DATA SOURCES AND ACQUISITION METHODOLOGY

Data Sources

- 1.1. List of neighborhoods in Mumbai
- 1.2. Venue data: Details on the venue using Foursquare API

Data Acquisition

- 1.1. Wikipedia data would be extracted using web scraping. Data includes name of the neighborhoods and latitude and longitude positioning.
- 1.2. Foursquare data would be extracted using python libraries and user credentials created earlier in the project.



METHODOLOGY USED FOR CODING

1. Download all the required libraries
2. Get the list of all areas in the city of Mumbai, India from the Wikipedia page through web scraping (https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai). Since latitude and longitude data is also available within the table we do not need to use any other external geospatial data.
3. Create a map of Mumbai using using Folium package.
4. Use Foursquare API to get the top 100 venues that are within a radius of 1 km using the credentials created on Foursquare developer website. We then make API calls to Foursquare passing in the geographical coordinates of the area in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, latitude and longitude.
5. In the next step, we review the number of unique categories and the number of venues by each area in Mumbai.
6. We then filter the data on 'Shopping Malls' to create clusters using k-means clustering
7. We then review the results to identify the concentration of malls by clusters.

RESULTS

We can see three clusters in the data:

- Cluster 0: Area with medium concentration of shopping malls
- Cluster 1: Area with low concentration of shopping malls
- Cluster 2: Area with high concentration of shopping malls

Our suggestion would be to consider either areas in cluster 1 or cluster 0 for the new malls

CONCLUSIONS

- Cluster 2 has only one area Bhandup. This area should be avoided as there are already too many malls in this locality.
- Our suggestion would be to consider either areas in cluster 1 or cluster 0 for the new malls