

Coursera Capstone

Opening a shopping mall in Mumbai, India

Prepared by- Swapnil Kulkarni

Date- 5/24/2020

Background and Objectives

Background

- Mumbai is one of the highest growing and most populous city in India
- In recent years, with ever growing population, shopping malls have been increasing to serve the people of the city
- Shopping malls provide one stop solution to all income group people and save on time along with providing variety
- Property prices have been increasing tremendously in Mumbai and selecting a right site has become crucial
- Location of a shopping mall is one of the important decisions which is a significant factor of the success

Objectives

- This project will help understand the current presence of shopping malls in Mumbai – neighborhoods
- This project is timely as the city is suffering from oversupply of shopping malls
- The ultimate objective is to recommend the best location to the property developer/investors who want to open a new shopping complex in Mumbai

Data Sources and Caveats

Data

- **Data required:**
 - List of neighborhoods in Mumbai
 - Latitude and Longitude of neighborhoods in Mumbai
 - Venue data, particularly data related to shopping malls
- **Data Source:**
 - Wikipedia page for Mumbai neighborhoods-
https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai
 - Latitudes and longitudes (from Google.com) for neighborhoods
 - Foursquare API for Venue data

Caveat

- This project only looks at current presence of malls and does not **consider other factors** like demographics property rates etc. Final decision may require these factors to be considered in the model.

Methodology

- Web scraping of Wikipedia data to get the neighborhoods in Python DataFrame
- Cleaning the neighborhoods data to remove anything that does not make sense
- Getting Latitudes and Longitudes from Geocoder package (since Lat and long for Mumbai weren't available on Geocoder, google.com is used)
- Use Foursquare API to get venue data
- Group data by neighborhood and taking the mean of the frequency of occurrence of each venue category
- Filter the venue for Shopping mall
- Perform K means clustering and select the cluster having minimum shopping malls

Results

Neighborhoods have been classified into three clusters-

- Cluster 0- Neighborhoods having very high density of shopping malls
- Cluster 1: Neighborhoods having relatively lower density than Cluster 0
- Cluster 2: Neighborhoods which are ideal for opening new shopping mall due to their lower density

Discussion

- Most of the shopping malls are concentrated in the central area of the city
- Highest number in cluster 0 and moderate number in cluster 1
- Cluster 2 has very low number to no shopping mall in the neighborhoods
- Oversupply of shopping malls mostly happened in the central area of the city, with the suburb area still have very few shopping malls

Recommendation & Conclusion

Recommendation:

- Open new shopping malls in neighborhoods in cluster 2 with little to no competition
- Second best alternate is cluster 1 where moderate number of malls are present
- Additional factors can be considered to decide where to open out of Cluster 1 or 2.
- Avoid neighborhoods in cluster 0 as already there are lot of malls already present

Conclusion:

- Answer to business question- Neighborhood in cluster 1 and 2 are better to choose for a opening a new shopping mall
- This project suggests us to do further study on other factors to choose between cluster 1 and 2.

Thank You