IBM Applied Data Science Capstone

Title: Opening a New Shopping Mall in Mumbai, India

By: Swapnil Kulkarni

Introduction

Shopping malls are getting increasing traction in the 21st century. With time becoming more and more important day by day, shopping malls provide one stop solution for the needs of people. It provides varieties of products across ranges. It also has other leisure things like game rooms, restaurants and movie halls.

Mumbai is a city in the West part of India. It is known as a financial capital of the country. Due to a lot of job availabilities, it is a very densely populated. There are people range from all the classes in the city and hence shopping malls get a huge footfall every day. They provide items from jewelry to cloths, food items, electronics etc.

As a result of growing popularity of malls in Mumbai, property developers are building more and more malls to get the advantage of growing population of the city. There are currently 250+ malls and more malls are in the development stage to open in a year or two. Of course, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure.

Business Problem

The objective of this project is to analyze and select the best location for opening a shopping mall in Mumbai, India. The project will use data science methodology, machine learning techniques such as cluster analysis to answer a question- where would you recommend opening a new shopping mall?

Who is the end consumer for this project:

The end stakeholder is property developers and investors looking to open or invest in new shopping mall in Mumbai.

Currently there are lot of factors taken into consideration while deciding a place such as local restrictions/government approvals, locality, commute time from important places, property rates etc.

Data requirement

To solve the problem, we will need the following data:

- List of neighborhoods in Mumbai, India. This defines the scope of this project which is confined to the city of Mumbai, India.

- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.

• Venue data, particularly data related to shopping malls. We will use this data to perform

clustering on the neighborhoods.

Data Sources:

a. Wikipedia Page- https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Mumbai

b. Foursquare API to get venue data

We will use Foursquare API to get the venue data for those neighborhoods. Foursquare has one of the largest database of 105+ million places and is used by over 125,000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Shopping Mall category in order to help us to solve the business problem put forward.

Methodology:

This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used.

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

This is based with the use K means clustering and setting K=3. It shows the neighborhoods having high number of malls and should ideally be avoided.

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:

- 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.