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

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.