Coursera Capstone Project

IBM Applied Data Science Capstone

Opening a New Shopping Mall in Mumbai, India

By: Aditya Andraskar

June 2020

Introduction:

For many shoppers, visiting shopping malls is a great way to relax and enjoy themselves during weekends and holidays. They can do grocery shopping, dine at restaurants, shop at various fashion outlets, watch movies and perform many more activities. Shopping malls are like a one-stop destination for all types of shoppers. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market their products and services. Property developers are also taking advantage of this trend to build more shopping malls according to the demand. As a result, there are many shopping malls in the city of Mumbai and many more are being built. Opening shopping malls allows property developers to earn consistent rental income. 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 failure.

Business Problem:

The objective of this capstone project is to analyse and select the best locations in the city of Mumbai, India to open a new shopping mall. Using Data science methodology and Machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Mumbai, India , if a property developer/ Business man is looking to open a new shopping mall, where would you recommend to open it?

Target Audience of this project:

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in the city of India i.e. Mumbai . This project is timely as the city is currently suffering from oversupply of shopping malls**.** According to MD & CEO of Anarock Retail, Anuj Kejriwal, top seven cities in India will see 69 new malls spread over 35.5 mn sq.ft. area while the remaining 31 malls over 13.5 mn sq.ft. will come up in tier 2 & 3 cities. East India will see a much more modest supply of seven new malls spread over approximately 3.5 mn sq. ft. The local newspaper also released Increased retail investment, expanding cities, growing economies in Tier II cities and a shift in shopping and recreational patterns are driving, and slowly changing, India's mall economy cataloguing the country's continued obsession with building more shopping space despite chronic oversupply.

Data:

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

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

Sources of data and methods to extract them :

This Wikipedia page (https://commons.wikimedia.org/wiki/Category:Suburbs_of_Mumbai) contains a list of neighborhoods in Mumbai, with a total of 42 neighborhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordinates of the neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, 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. 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.