Coursera Capstone

IBM Applied Data Science

Capstone Opening a New Shopping Mall in Mumbai,India

By: Abhay Kotal

January 2019

Introduction:

Shopping mall is a great place where people can find all sort of things at one place from clothing ,grocery , daily needs, movies etc. many people spend huge amount of money in shopping mall. But opening a shopping mall at a right location is very important. There are already many shopping malls in the city opening the best one at the right place and attracting more crowd. Attracting different crowd group based on their age financial status, etc into the mall and making profit out of those crowd is also an important factor .So opening a shopping mall with all features at a right place is very important and challenging. Opening a shopping mall in Mumbai is very challenging because finding suitable property at a given place is very challenging and also costly .Most important factor is the shopping mall location that will determine the success o the shopping mall or failure of a shopping mall

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 is looking to open a new shopping mall, where would you recommend that they open it?

Target Audience of this project:

This project is for the investor and for the developers those who want to build a shopping mall at a right location and invest for this project to make this project successful. This project will show the no of clustered shopping mall at a place due to which the city is suffering from over supply of shopping mall and a large no of shopping mall are built at close by place. Due to which there is a clustered of shopping mall at one place. In the city shopping malls always target for the people with financial status so most of the shopping mall end up at the costly area of the city.

Data:

To solve the problem we require the following data:

- List of neighbourhoods in Mumbai. This defines the scope of this project which is confined to the city of Mumbai, the capital city of the country of India .
- Latitude and longitude coordinates of those neighbourhoods. 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 neighbourhoods

Sources of data and method to extract them:

Sources of data and methods to extract them This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai) contains a list of neighbourhoods in Mumbai. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautiful-soup packages. Then we will get the geographical coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods. After that, we will use Foursquare API to get the venue data for those neighbourhoods. 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.

References :

[Wikipedia] for the data .

[Chia Hooi Lim , neighbourhood project] for the report .