Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Shopping Mall in Mumbai,India

By: Ayush Falor
April 2020

INTRODUCTION

Shopping Malls are now an integral part of our lives, apart from being a great place to hang out, it serves as a multiple activity zone with activities ranging from Places to have food, shop, have coffee, or watch a movie. Malls cater to multiple needs at once and hence are so popular in nature, apart form they make great places of business with special attention paid to improving their profitability, hence when opening a mall the most important attribute is the location, as the right location and can make huge difference to the malls fortunes.

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?

Data

To solve the problem, we will need 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 Maharasthra, the 2nd largest stae 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 methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur) contains a list of neighbourhoods in Mumbai , with a total of 42 neighbourhoods. 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 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.