Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Shopping Mall in Mumbai, India

Naman Bhandari

Introduction

We all will accept that Shopping Malls are fun. Not only do they offer great discounts and a full shopping experience but they also allow people to relax and enjoy, especially on weekends . Due to these reasons, they have now become a quite important part of our live and thus a profitable business.
Therefore, many new shopping malls are now opening, but they all face a common challenge, that is to remain profitable. They may be the biggest, most beautiful, lavish or may be offering great discounts but then also they may not be profiting accordingly.
One of the main reasons for loss of shopping malls is the sever competition they face. Not only from other retail stores, but also from other shopping malls.

Therefore, this project will help new shopping malls to make profit by suggesting hem better locations to avoid competition.

Business Problem

The objective of this capstone project is to analyze 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 particularly useful to property developers and investors looking to open or invest in new shopping malls in the financial capital of India, Mumbai. This project is timely as the city is currently suffering from oversupply of shopping malls.

Data

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

- List of neighborhoods in Mumbai.
- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and 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://en.wikipedia.org/wiki/Category:Suburbs_of_Mumbai) contains a list of neighborhoods in Mumbai, with a total of 40 neighborhoods. 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 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. 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.