Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Shopping Mall in Pune, India

Simran Kaur Ahluwalia
March 2021

Introduction

A shopping centre is defined as an aggregate of trade enterprises and enterprises providing services located in a certain territory, planned, built and managed as a whole and providing parking for the vehicles within its territory. For many shoppers, visiting malls are a great way for relaxation. They can do grocery shopping,dine at resturants, shop at various fashion outlets, watch movies and perform many more activities. For retailers, the central location and the large crowd at the shopping malls provides a great distribution channel to market tgeir products and services. As a result, there are many shopping malls in Pune and many more are being built. In particular, the location of shopping mall is one of the most important decisions that will determine whether the mall will be success or a failure.

Business Problem

The objective of this capstone project is to analyse and select the best locations in Pune, India to open a new shopping mall.Placement of shopping centers is the most important aspect of developing the concept of designing a modern shopping center.Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In Pune, India, if a property developer is looking to open a new shopping mall, where would you recommend that they will open it?

Target Audience of this project

This project would particularly be useful for investors & property developers looking out for investment and developing opportunities in Pune, India. 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 neighbourhoods in Pune. This defines the dcope of this project which is confined to the Pune city.
- Latitude & Longitude coordinates of those neighbourhoods. This 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:Villages_in_Pune_district) contains a list of neighbourhoods in pune, with a total 137 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 neighbourhood 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 API has one of the largest database of 105+ million places and 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 the machine learning technique that was used.