IBM Applied Data Science Capstone

Finding Best Location to Open New Shopping Mall in New Delhi, India

By: Aditya Uniyal

Introduction

For many people, 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 the 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 to cater to the demand. As a result, there are many shopping malls in the city of New Delhi and many more are being built. Opening shopping malls allow 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 a failure.

Business Problem

The objective of this capstone project is to analyze and select the best location in the city of New Delhi, 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:
Considering existing competitors in the market, what could be the best location in the city of New Delhi, India to open a new shopping mall?

Data

To solve the problem, we need the following data:

- List of neighborhoods in New Delhi. This defines the scope of this project which is confined to the city of New Delhi, that serves as a capital of India.
- 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://en.wikipedia.org/wiki/Category:Neighbourhoods_in_Delhi) contains a list of neighborhoods in New Delhi, with a total of 141 neighborhoods. Web scraping techniques such as requests module and beautiful-soup packages can be used to extract the data from the Wikipedia page. Then I will be using the Python Geocoder package which will return the latitude and longitude coordinates of the neighborhoods.

After that, I will be using 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.
This 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).