

Coursera Capstone

IBM Applied Data Science Capstone

Opening a Shopping Mall in Delhi, India

By: Sanyam Srivastava
May 2020

In a city of your choice, if someone is looking to open a Shopping Mall, where would you recommend that they open it? Similarly, if a contractor is trying to start their own business, where would you recommend that they setup their office?

Introduction

For many people, visiting Shopping Mall and eating or shopping or exploring is a great way to relax and enjoy themselves during weekends and holidays. They can dine at restaurants, pack it for home, do shopping, etc. Of course, as with any business decision, opening a Shopping Mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of it, the competition in that area are one of the most important decisions that will determine whether the Shopping Mall will be a success or a failure.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of 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: In the city of Delhi, 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 mall in the capital city of India. This project is timely as the city is currently suffering from oversupply of such places.

Data to be used

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

- List of neighbourhoods in Delhi. This defines the scope of this project which is confined to the city of Delhi, the capital city of the country 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 restaurants. 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/Neighbourhoods_of_Delhi) contains a list of neighbourhoods in Delhi. 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).

Eg: If I extract a location say laxminagar from we scraping from wikipedia, I will get it coordinates using python geocoder, then find where the malls are using foursquare API. Similarly, I will extract all location malls, then I will form cluster, 4 clusters and plot them all. The cluster with least number of malls is the place where a mall should be built. This is how I am going to use the data.