Coursera Capstone Project: Applied Data Science

Abhishek Kumar
abhiavisekkr@gmail.com

July, 2020

Opening a new Shopping Mall in Delhi, India

---

1. Introduction

The importance of shopping malls as retailing formats has become increasingly remarkable, and today malls plays a significant role in consumers' lifestyle. But nowadays shopping malls have become not just a place to shop it has become a place where social factors get deeper. Shopping malls are like one-stop destination for not only various types of shoppers but also for dine at restaurants, watch movies, celebrate etc. This gives retailers a central location and large crowd at the shopping mall which in turn provides a great distribution channel to market their products and services. Real Estate, builders are also taking interest to build shopping malls to cater the demand. This also becomes a consistent rental income for the owners. As a result, there are many shopping malls in the Delhi and more and more malls are being built. But to open a shopping and such that the shopping mall succeeds lot of factors come into play. Among them one of the major factors is the location of the shopping mall.

2. Business Problem

Since, shopping malls are main interest to various groups therefore they are built and real estate investors invest in these projects. But for shopping malls to attract large crowd there are few major factors, one of those is location. The objective of this capstone project is to analyse and select the best locations in Delhi, India to open a new shopping mall with high chances of success using data science methodology and machine learning techniques. We will make this decision in this project. This will specially benefit the real estate builders since Indian retail sector has metamorphosed significantly over last few decades. Rapid urbanization and digitization, rising disposable incomes and lifestyle changes of particularly the middle-class has led to a major revolution in the retail sector, projected to grow from US $672 billion in 2017 to US $1.3 trillion in 2020. Evolving rapidly from usual 'kirana shops' to large multi-format stores offering global experience to the e-commerce model that is highly technology-driven, the Indian retail sector has evolved.

3. Data

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 Delhi, the capital of India.
- ➢ Latitude and Longitude coordinates of the neighbourhoods. This required to plot the map and get the venue data
- ➢ Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighbourhoods.

Data Sources:

- ➢ The data of the neighbourhoods in Delhi can be extracted from Wikipedia page. (https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi)
- ➢ Then the latitude and longitude data can be retrieved from Python geocoder package.
- ➢ Then using latitude and longitude data venues can be fetched from Foursquare API.

Methods to extract data:

- ➢ We will do web scraping using BeautifulSoup a library of python to get the neighbourhoods from the Wikipedia page.
- ➢ Then we will pass the scraped data to Geocoder to get the latitude and longitude.
- ➢ Then pass the latitude and longitude data along with required parameters to the FourSquare API and get the venues data i.e. the shopping malls.

In the next section, we will present the detail methodology where we will discuss about data cleaning, data wrangling to using machine learning algorithm and visualising every step on the map using folium. Also, we will do a complete data analysis and reach to some result.