

Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Shopping Mall in Delhi, India



By: Ravi Sharma

May 2020

Introduction

For most of the people living in metro cities, visiting shopping malls is a great way to relax and enjoy themselves during weekends, holidays and even at regular days. They can do grocery shopping, dine at restaurants, shop at the various fashion outlets, watch movies and indulge in many more activities. Shopping malls are like a one-stop destination for all types of shoppers for all activities. 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 huge number of shopping malls in the capital city of India i.e. Delhi and many more are being built. Opening shopping malls allows 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 analyse and select the best locations in the capital city 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 capital city Delhi, 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 capital city of India i.e. Delhi. Also, it could suggest businessmen if they should open a shop in particular malls as opening a shop at place which is already crowded with shops may not lead to much business. This project is timely as the city is currently suffering from oversupply of shopping malls in particular areas, however due to increasing migration of people to capital city for job opportunities, the places which used to be less crowded are becoming populated. People living at these places have to travel to existing shopping malls, if they could have new shopping malls nearby, they will surely visit nearest shopping malls in their own areas.

Data

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

- List of neighbourhoods Delhi. This defines the scope of this project which is confined to the city of Delhi, the capital city of India in Asia.

- 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 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:Neighbourhoods_in_Delhi) contains a list of neighbourhoods in Delhi, with a total of 142 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 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). 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.