Predicting Location in Delhi for a Mall

Satyendra

Introduction

Delhi is the largest commercial centre in northern India. As of 2016 recent estimates of the economy of the Delhi urban area have ranged from $167 to $370 billion (PPP metro GDP) ranking it either the most or second-most productive metro area of India. The nominal GSDP of the NCT of Delhi for 2016–17 was estimated at ₹6,224 billion (US$90 billion), 13% higher than in 2015–16.

Delhi has an attractive real estate market and is a preferred tourist destination. Owing to its location, connectivity and rich cultural history, Delhi has always been a prime tourist attraction of the country. Many shopping malls are present in Delhi and opening a new one requires certain factors in mind.

Problem:

Data that might contribute to determining Shopping mall opening location might include his area population and nearby shopping centres. This project aims to predict whether and how much a feasible will be to open the restaurant in the area.

Target Audience:

Various big brands, which invests in opening malls in Delhi. We can see in the below report that future is bright and companies are coming with the various projects.

Data

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

- List of neighbourhoods in New Delhi. This will elaborate the scope of data.
- For plotting the locations, we need longitudes and latitudes.
- 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/Neighbourhoods_of_Delhi) contains a list of neighbourhoods in New Delhi.
- We will use web-scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautiful soup 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 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 and the machine learning technique that is used.