Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Shopping Mall in Tamil Nadu, India

By: Rooban Sappani

June 2020

Introduction

The importance of a shopping mall is mainly for people to get out of the house for a while and do something entertaining. Shopping malls can provide the best shopping experiences such as social gatherings, entertainment, performances, product launches, promotions and festivals. The events list at shopping malls goes on and on for any, particular, person to be entertained for a number of hours. Any shopping mall can be a great place to hang out with friends, eat, shop, and more. Malls can be very helpful because all the needs are in one building. Shopping malls tend to be a major tourist attraction. The mall can be more convenient, for a tourist, to have one central location to do all their shopping; rather than to have to drive many miles just to buy different types of products for their personal needs. There are many shopping malls in the and many more are being built in the city of Chennai. 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 city of Chennai, Tamil Nadu 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 Chennai, Tamil Nadu, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

Data

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

- List of neighbourhoods in Chennai. This defines the scope of this project which is confined to the city of Chennai, the capital city of the state of Tamil Nadu in 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 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:Suburbs_of_Chennai) contains a list of neighbourhoods in Tamil Nadu, with a total of 68 neighbourhoods. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and bs4 package. 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.