Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Coffee Shop in Bangalore, India

By: Albert Abraham

October 2019

Introduction

For many people, a coffee shop is a great place to relax and enjoy themselves, hangout with friends, have casual meetings and/or go for a date. Property developers are also taking advantage of this trend to building more coffee shops to cater to the demand. As a result, there are many coffee shops in the city of Bangalore and many more are being built. Opening coffee shops allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new coffee shop requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the coffee shop is one of the most important decisions that will determine whether the shop 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 Bangalore, India to open a new coffee shop. 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 Bangalore, India, if a property developer is looking to open a new coffee shop, 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 coffee shops in the silicon valley of India, Bangalore. This project is timely as the city is currently suffering from oversupply of coffee shops.

Data

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

- List of neighbourhoods in Bangalore.
- 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 coffee shops. We will use this data to perform clustering on the neighbourhoods.

Sources of data and methods to extract them

Government of India provides datasets that are publically available at : data.gov.in.
There is a dataset containing All India Pincode directory with contact details along with Latitude and longitude.
We will filter the dataset to retrieve only data belonging to Bangalore. Latitude & Longitude columns are mostly missing values, so we will compute these using geocoder based on pincode of the location.
After that, we will use the Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest databases 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: data cleaning, data wrangling, working with API (Foursquare), 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.