Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Café in New Delhi, India

By : Rishabh Maheshwari

June 2020

Introduction

The Café business in India, nay, the entire world has been a happy hunting ground for many entrepreneurs, big or small. The sector is still seeing high growth and a huge amount of investments, with new chains opening at a rapid pace. For some of the older players though, the time has come to consolidate their businesses and/or expand. This requires systematic management and oversight to ensure a smooth transition so that the basic fundamentals on which their businesses have been built is not scarified at the altar of growth. Opening café allows property owners to earn consistent rental income. Of course, as with any business decision, opening a new café requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the café is one of the most important decisions that will determine whether the café 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 New Delhi, India to open a new Café. 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 New Delhi, India, if a property developer is looking to open a new Café, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to property owner and investors looking to open or invest in new café in the capital city of India i.e. New Delhi. It also identify the best demographic location i.e. where to build your Café for your target audience.

Data

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

• List of neighbourhoods in New Delhi. This defines the scope of this project which is confined to the city of New Delhi, the capital city of the country of 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 café 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 New Delhi, with a total of 138 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 Café 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.