Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Indian Restaurant in Delhi,India

By: Akshay Rana

April 2020

Introduction

Going to a restaurant is a great way to relax and enjoy with friends and family during weekends and holidays. Apart from going for lunch or dinner various moments can also be celebrated at restaurants such as birthdays, get-together and many more activities. Opening a restaurant provides a great opportunity to serve various cuisines to people who actually haven't got a chance to travel. On the other hand, it proves to be a great business once a name has been made. People are taking advantage of this trend to build more restaurants to cater to the demand. As a result, there are many restaurants in the city of Delhi and many more are being built. Opening restaurants allows people to earn consistent rental income. Of course, as with any business decision, opening a new restaurant requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the restaurant is one of the most important decisions that will determine whether the restaurant 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 Delhi, India to open a new Indian Restaurant. 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 Delhi, India, if a person is looking to open a new indian restraunt, where would you recommend that they open it?

Target Audience of this project

This project is particularly useful to people and investors looking to open or invest in new indian restaurants in the capital city of India i.e. Delhi. This project is timely as the city is currently suffering from oversupply of shopping malls.

Data

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

- List of neighbourhoods in Delhi. This defines the scope of this project which is confined to the city of Delhi, the capital city of the country of India in South East 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 Indian Restaurants. 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 Kuala Lumpur, with a total of 141 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 Indian Restaurants 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.