IBM Applied Data Science Capstone

Predicting the Best Location to Open a New Indian Restaurant in New Delhi, India

By: Dinesh Ogirala

July 2020

Introduction

For many food lovers, Indian Restaurants are a great way to relax and enjoy not only during weekends and holidays but also during weekdays. Indian Restaurants are like a one-stop destination for family gatherings. Most of the people in Delhi prefer Dumdar Indian Food to Western Foods. For retailers, the central location and the large crowd at the Restaurant provides a great distribution channel to market their products and services. Property developers are also 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 New Delhi and many more are being built. Opening Restaurants allows property developers 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 it will be a success or a failure.

Business Problem

The objective of this capstone project is to analyze and select the best locations in the city of New Delhi, India to open a new 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 New Delhi, India, if a property developer is looking to open a new Restaurant, where would you recommend that they open it?

Data

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

- List of neighborhoods in New Delhi. This defines the scope of this project, which is confined to the city of New Delhi, the capital city of India.
- Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and to get the venue data.
- Venue data, particularly data related to shopping malls. We will use this data to perform clustering on the neighborhoods.

Data Extraction

This Wikipedia page (https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi) contains a list of neighborhoods in New Delhi, with a total of 9 neighborhoods. 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 neighborhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods.

After that, we will use Foursquare API to get the venue data for those neighborhoods. 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 Indian Restaurant 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).