Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Restaurants in Delhi,India

Author: Souvik Ghosh Roy Chowdhury

October, 2019

Introduction

Delhi is city, growing exponentially in Economical sector. With its population near about 1.9 crores it becomes one of the top market for investors. This notebook gives investors brief idea for opening new restaurants. For many visitors, food lovers, visiting Restaurants is a great way to relax and enjoy verities of foods themselves during weekends and holidays, even on weekdays after hectic office hours. As a market Delhi is great place where crowd is key for running a Restaurant Business. 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 Delhi, Capital of India, 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 Restaurants requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the Restaurants 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 analyze and select the best locations in the city of Delhi, India 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 Delhi, India, if a property developer is looking to open a new restaurant , 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 restaurants in the capital city of India i.e. Delhi. This project is timely as the demand of different types restaurants in city is increasing day by day. For example Japan curry chain to open 1st India restaurant in New Delhi suburb and many more to come in near future.

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.

> • 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:Neighbourhoods_in_Delhi) contains a list of neighbourhoods in Delhi, with a total of 135 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 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.