Coursera Capstone

IBM Applied Data Science Capstone

Exploring venues in Shimla, India

By: Deepesh Yadav

November, 2019

Introduction

Whenever a person is visiting or planning to visit a new city, he/she always wants to know what the city has in store for them, And a city's restaurants and cafe's are a major part of the experience when someone visits a city, Most of the people who visit these restaurants, might want to know about the food and ratings of the given place. This information can help people decide about the places the want to eat and help them get reservations if needed.

Shimla city is spread over an area of 35.34 km^2 and has various number of places spread over the city, as in most these places vehicles are not allowed (near the mall road) so it would be very helpful for the people to efficiently find their kind of place to eat. By using the foursquare API we'll be able to provide people with sufficient knowledge about the place they are visiting and what to expect from the place.

Business Problem

The objective of this capstone project is to analyse and provide the user with attributes like ratings, price, reviews and other necessary details for them to decide at what place to eat in Shimla. Using data science methodology and machine learning techniques which I've learned over the period of past few months, this project aims to provide solution to the dilemma of where to eat in a new city ?.

Target Audience of this project

The target audience for such a project bifold. First anyone who is visiting the city for the first time can use the plots and maps from this project to decide quickly at what place to eat that suits their budget, schedule and rating preferences and the second part of the audience can be companies that want to use this information to create a service(website/mobile app). Which has all this data, to allow visitors to city's places and even provide the same details for some other city.

Data

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

List of venues in Shimla. This defines the scope of this capstone project which is confined to the city of Shimla, the capital and the largest city of the Indian state of Himachal Pradesh.

Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the maps and also to get the venue data.

Venue data, particularly data related to eating places. We will utilise this data to extract necessary details from it.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Shimla) contains a lot of details about the city of shimla that I've used in this project (e.g. Coordinates, coverage, etc.).

The data about the place and venues was extracted by using the Foursquare API. 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, I am only interested in Restaurants and cafe category in order to help us to solve the business problem stated. This is project will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning 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(s) that was used.