

IBM Applied Data Science Capstone Project

Opening Bar in Delhi, India

By

Kumar Sanu

August '20



Introduction

For many people, going at a place with their family and friends to have a good time and have good food and drinks are a good way to relax during weekends and holidays. These drinks also help people in reducing their depression and frustration occurred due to both professional and personal life. With this said, bars are becoming a common place for all sorts of people across the city. Whether you stay in your home or you are visiting a place, you should consider visiting a bar at least once a week or more. Many stressful things are happening in the world around us, so sometimes all one needs is to go to a place that serves good food and drinks. Hence the availability of bars became the requirement in the life of people.

And where there is demand of services, there is business. Hence business people can think to start their business. But to open any business there is many things to be taken in consideration.

Chances are we've come across the term 'location, location, location' and when it comes to opening a new bar, nothing is more important. Sure, we'll need a great concept, a first-rate product, and excellent customer service, but our bar's location will be the foundation of your business and have a major impact on its success. So for a project developer, location is the key factor to consider in case of opening the bar so that they can target the desired number of peoples.

Business Problem

The aim of this project is to help the project developers in selecting the best locations in Delhi, India to open bar. We will use Data Science Methodology and Machine Learning Techniques to solve this problem.

Basically this project will provide answer to this Business Problem:-

If someone wants to open a new bar in Delhi, the capital of India, where they should open?

Data

We need following data to solve this problem:-

1. List of neighborhoods in Delhi. The scope of the project will be confined to the city of Delhi.
2. Latitude and longitude of each neighborhoods. This is required in plotting the map and getting venues information.
3. Venue data especially related to bar, which is the area of interest for us. This will be used in clustering those neighborhoods.

Source Information and Methods for extraction

We will use the Wikipedia page for getting data which consists of 9 Boroughs comprising of 117 Neighborhoods of Delhi. (https://en.wikipedia.org/wiki/Neighbourhoods_of_Delhi)

1. We will use web scraping techniques to extract the list of neighborhoods from wiki page with help of python packages.
2. We will use geopy package to find latitude and longitude of each neighborhoods.
3. We will use Foursquare API to get all venues related information of neighborhoods using their latitude and longitude values.

Foursquare API will provide many categories of data in those neighborhoods. Out of those categories, we are highly interested in data related to bar in order to solve this problem.

This project will make use of many data science skills from data extracting (web scraping), Working with API(Foursquare),data cleaning, data wrangling, data visualization(Folium) to Machine Learning(K-means Clustering). We will discuss these all methodologies and the steps taken in detail while implementing it.