

Coursera Capstone Project – “The Battle of Neighborhoods”



Opening a new Hookah Bar in Colombo, Sri Lanka

By Zufar Haseeb

July 2019

Introduction

For many tourist and youngsters, Hookah Bars are an addiction and are increasingly becoming a preferred location to hang out with friends. Subsequent to the 30 year long civil war in Sri Lanka, the tourism industry has boomed and tourist from many destinations are visiting the country. As Colombo is the capital city of the country, a new Hookah Bar may generate consistent revenue for any investor while the location of the Hookah Bar being the key determinant of consistent revenue through the success of the Hookah Bar.

Business Problem

Until now, neither the residents of Colombo nor the increasing number of tourist arriving in Sri Lanka do not have a international level Hookah Bar in Colombo that would provide authentic Italian food with the same quality as it is seen in other developed countries. Objective of this project is to analyze and select the best location in Colombo, Sri Lanka to open a new Hookah Bar. Using data science methodologies such as clustering, this project aims to provide suggestions for locations in Colombo city for a new Hookah Bar to be opened.

Target Audience

- Investors
- Tourists
- Franchises, etc.

Data

We will require following data to solve this problem

- List of neighborhoods in Colombo. This defines the scope of the project which will be confined to the city of Colombo, the capital city of the country of Sri Lanka in South Asia
- Latitude and Longitude coordinates of above neighborhoods in order to plot the maps and obtain venue data
- Venue data, particularly data of shopping malls. This can be used to perform clustering of neighborhoods.

Sources of Data

- The Wikipedia page (https://en.wikipedia.org/wiki/Category:Suburbs_of_Colombo) contains a list of neighborhoods in Colombo with a total of 65 neighborhoods. We will be using web scraping techniques to extract data from the Wikipedia page with the help of Python and beautifulsoup packages. Then we will get the geographical coordinates of the above neighborhoods using Python's Geocoder package.

- Foursquare API to get venue data for above neighborhoods. Foursquare has one of the largest databases of over 100 million places and is used by over 150,000 developers across the world. Foursquare API will provide many categories of venue data and we will be focusing on Hookah Bar category in order for us to solve the problem that has been put forward.