IBM Applied Data Science
Capstone Project

Opening a new restaurant in Kuala Lumpur, Malaysia

Introduction

Business Problem:

Kuala Lumpur is the cultural, financial and economic centre of Malaysia. It is among the fastest growing metropolitan regions in Southeast Asia, in both population and economic development. Tourism plays an important role in the city's service driven economy. It is driven by the city's cultural diversity, relatively low costs, and wide gastronomic and shopping variety. Businessmen are looking for opening new restaurants in Kuala Lumpur to take maximum benefit out of city's tourism.

The objective of this report is to analyse and select the best locations in Kuala Lumpur, Malaysia to open a new restaurant. This report is intended for investors and businessmen looking to open restaurants in a strategic location. In this project, we will use foursquare location data and regional clustering of venue information to determine what must be the best neighborhood in Kuala Lumpur to open a restaurant. The objective of performing analysis is to select a cluster in such a way that the investor will face very little competition from others.

Data

We considered following data to build the analytical model:

- List of neighborhoods in Kuala Lumpur. This also defines the scope of project, which is confined to the city of Kuala Lumpur, capital of Malaysia.
- Latitude and Longitude coordinates of those neighborhoods. This will be required to plot the map and get the venue data.
- The data obtained will be used to perform clustering on the neighborhoods.

The mentioned Wikipedia page: https://en.wikipedia.org/wiki/Category:Suburbs_in_Kuala_Lumpur contains a list of neighborhoods in Kuala Lumpur, with total of 70 neighborhoods. Web scrapping technique was used to extract the data from Wikipedia page with the help of Python requests and

BeautifulSoup packages. Geographical coordinates (latitudes and longitudes) of the neighborhoods were obtained using Python Geocoder package. Next, Foursquare API was used to obtain venue data for those neighborhoods. It provided many categories of the venue data but only restaurant category was used in this model to solve the business problem.

This project puts into implementation combination of skills learnt in data science course ranging from web scrapping (Wikipedia), working with API Foursquare, data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).