[Year]

Coursera Capstone – IBM Applied Data Science Capstone

FINDING A NEAREST FINE DINNING RESTAURANTS IN SINGAPORE

Introduction

The objective of this capstone project is to analyse and select the best Fine Dining restaurant in the Singapore to have a beautiful dinner date. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the country of Singapore, if a couple is finding to have a decent fine dining restaurant, where would you recommend where they go to?

Business Problem

Since Singapore is the most religiously diverse country where the different cultures, religions and foods meet, the food taste of the people are also difference. And Singapore is the country where the love of those difference people is born. When those couples who are from the different culture and having the different taste of food taste find to have a dinner date, that will take a quite long process to do a research to know the best place for the dinner date based on the food, locations, rating and price. So, the project will be helpful for all those people who want to go on the date with their beloved one to the Fine Dining restaurant and they can easily know which restaurant is located in their nearest place, serving what kind of food, has a good or bad rating and pricing.

Target Audience

This project is particularly useful to those couple who are finding a restaurant which has a decent pricing, food and nearby their place.

Data API

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

- List of MRT station in Kuala Lumpur. This defines the scope of this project which is confined to the Singapore in South East Asia.
- Latitude and longitude coordinates of those MRT. This is required in order to plot the map and also to get the venue data.
- Rating, particularly data related to restaurant. We will use this data to perform clustering on the MRT.

Sources of data and methods to extract them

This dov.gov.sg page (https://data.gov.sg/dataset/lta-mrt-station-exit) contains a list of MRT station in Singapore, with a total of 70 MRT. We will use web scraping techniques to extract the data from the dov.gov.sg page, with the help of Python requests. 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 databases 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 (DOv.gov.sg), 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.