Coursera Capstone

IBM Applied Data Science Capstone

Finding ideal hotels in Singapore

By Do Thi Kim Uyen

January, 2020

Introduction

For many travelers, to find out an ideal hotel for their trip to Singapore is quite important. As a heart of Asia, which is easy to reach by many transportation types, Singapore is now growing incredibly in hostibility industry. Many entertainment parks, hotels, man-made beautiful views, sightseeings.. were being built to maximize exploit the strength of Singapore in travelling.

Think as a travelers, we should find a hotel which is easy to reach to shopping mall also entertainment area, sightseeings… With a small country like Singapore and the population is quite big enough, living standard is high, and in downtown Singapore, the price is extremely high within real estate for rent, not except for hotel price. We also can rent accommodation in suburbs but in my point of view, it must be close to MRT and convenience stores. If your pocket has its limitation, tourists do not want to put much of their expenses on just accommodation. Our choices focus on entertainment and traveling experience. That is why we do not have many choices on hotel.

As a result, I come to a research on finding the ideal hotel which is close to utilities and help to match with your expected hotel expense. Particularly, the location of the hotel is the most important decision which is ideally match to our expectations on a trip to Singapore.

Busniness problem

The objective of this capstone is to analyse the best hotel location in Singapore. Using data science methodology and machine learning techniques like clustering, the porject aims to provide solutions to answer question: In Singapore, if travelers are looking for an ideal hotel to stay, which hotel would you recommend them?

Target Audience of this Project

This project is obviously useful for travelers, who want to help great travelling experience to Singapore with a suitable amount of money. This project is always timely.

Because Singapore is always a hot destination of travelers all year around. Data from Budget Direct Insurance about Singapore Tourism Statistics 2019 released recently showed that there are 7.8 millions international tourists arrived in Singapore, which is up 1.49% compared to 2018. Anyone, who desire to visit Singapore, may be really interest in this project.

Data

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

- List of neighborhood in Singapore.
- Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and also to get the venue data.
- Venue data, particularly data related to hotels. We will use this data to perform clustering on the neighborhoods.

2. Sources of data and methods to extract them

This wikipedia page (https://en.wikipedia.org/wiki/List_of_hotels_in_Singapore) contains a list of hotels in Singapore, with a total 29 hotels. We will use web scraping techniques to extract data from Wikipedia page, with the help of Python Beautifulsoup and requests packages. Then we can get cooridinates of neighborhood using Geocorder package. This package give us the latitude and longitude coordinates of those neighborhood.

The next step, we will use Foursquare API to get the venue data of the neighborhood. Foursquared API can provide may of categories of utilities data, which is surround a hotel. We are particularly interest in entertainment area, shopping mall, sightseeing categories in order to find out which hotel is a good choice. To finish this project, I must use Data science skills and knowledges, from web scraping, working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-Means Clustering) and visualizing use Folium and then analyse to give comments on hotel choices depend on types of target audience.