IBM Applied Data Science Capstone

Opening a new hotel in Singapore

Introduction

The tourism industry is an integral part for any country where hotels play an important role in ensuring that tourists have a proper and safe place to rest. Many developers are taking advantage of this trend to see where they can build more hotels to generate more revenue. Furthermore, opening hotels offer developers a somewhat steady stream of income especially in a country such as Singapore which is reliant on its tourism industry. However, it is important to weigh the pros and cons before making a business decision considering the high costs. Therefore, the importance of choosing an ideal location to set up a hotel cannot be understated.

Business Problem

The objective of this project is to select the best locations in the city of Singapore where it is ideal to open a new shopping mall. This can be achieved by using data science and machine learning techniques such as clustering. This is especially prevalent in a country such as Singapore where land is scarce

Data

To tackle this issue, the data that is required is:

- A list of districts in Singapore, where the capital is Singapore
- The latitude and longitude coordinates of the districts to plot the map
- Proximity data, to see where the hotels are in each district

Sources of data

The district data (https://en.wikipedia.org/wiki/Planning_Areas_of_Singapore) is obtained from Wikipedia where we can use web scrapping packages and techniques from beautifulsoup to extract data from this page.

Following that, we can use the geocoder package to obtain the exact latitude and longitude coordinates before visualising it on a map using folium. We will also be using Foursquare's API to get the venue data to check where the hotels are in each district.