COURSERA CAPSTONE

IBM Applied Data Science Capstone

Opening a New Coffee Shop in Jakarta, Indonesia

By : Muhammad Sabastian Riva'i

June 2020

Introduction

Coffee shop is the most visited place to hang out with friends and release stress. Beside of that, coffee is Indonesian's most favourite beverage. The citizen of Indonesia have a huge desire to visit coffee shop to chit chat with friends and enjoy coffee. For those who wants to make a coffee shop, location is an important aspects. We have to know if we have any competitor in the area, the crowd of the area, rent cost of the area etc. Of course, as with many business decision, opening a new coffee shop requires serious consideration and a lot more complicated than it seems. Particularly, the location of the coffee shop is one of the most important decisions that will determine whether the coffee shop will be a success or a failure.

Business Problem

The objective of this capstone project is to analyze and select the best location in Jakarta, Indonesia to open a new coffee shop. Using data scince methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Kuala Lumpur, Malaysia, if an investor is looking to open a new coffee shop, where would you recommend that they open it?

Target Audience of This Project

This project is particularly useful to property developers and investors looking to open or invest in new coffee shop in Jakarta, Indonesia. This project is timely as the city is currently suffering from oversupply of coffee shops. Recently, there are a lot of coffee shops are opened in Jakarta. Chairman Specialty Coffee Association of Indonesia predict the number of coffee shops in Jakarta will increase to 20% by the end of the year. So this project will be useful for anyone who want to open coffee shop in Jakarta.

Data

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

- List of districts in Jakarta. This defines the scope of this project which is to confined to the city of Jakarta, the capital city of Indonesia.
- Latitude and longitude coordinates of those districts. This is required in order to plot the map and get the venue data from foursquare.
- Venue data, particularly data related to coffee shop. We will use this data to perform clustering on the districts.

Sources of data and methods to extract them

This Wikipedia page (https://en.wikipedia.org/wiki/Category:Districts_of_Jakarta) contains a list of districts in Jakarta, with a total of 44 districts. We will use web scraping techniques to extract the data from the Wikipedia page, with the help of Python requests and beautifulsoup packages. Then we will get the geographical coordiinates of the districts using Python Geocoder package which will give us the latitude and longitude coordinates of the districts.

After that, we will use Foursquare API to get the venue data for those districts. Foursquare has one of the largest database of 105+ million places and is used by over 125000 developers. Foursquare API will provide many categories of the venue data, we are particularly interested in the Coffee Shop category in order to help us solve the business problem described above. This is a project that will make use of many data science skills, from web scraping (Wikipedia), working with API (Foursquare), machine learning (K-means clustering) and map visualization (Folium).