

Coursera Capstone IBM

Applied Data Science Capstone

“Opening a New Coffee Shop in Jakarta, Indonesia”

By: Aries Fitriawan

August 2019



Introduction

For many coffee lovers, visiting Coffee Shop is a best way to relax and enjoy themselves during their free time. Nowadays, Coffee Shop can be a best place for working, meeting client, or just to enjoy the ambience with a taste of signature local coffee. Coffee Shop are like a multipurpose place. For the business owner, the central location and the potential of the coffee shop provides a great distribution channel to market their products and services. The barista also more creative to provide the latest taste of coffee, combined with the other ingredients to attract the customer. Shopping Mall developers are also taking advantage of this trend to build more Coffee Shop tenant in their shopping mall to cater to the demand. As a result, there are many Coffee Shop in the city of Jakarta and many more are being built. Opening Coffee Shop also can allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new Coffee Shop requires serious consideration and is 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 store will be a success or not.

Business Problem

The objective of this capstone project is to analyse and select the best locations in the city of Jakarta, Indonesia to open a new Coffee Shop. Using data science methodology and machine learning techniques for clustering, this project aims to provide solutions to answer the business question:

In the city of Jakarta, Indonesia, if a business owner is looking to open a new Coffee Shop, where would you recommend that they open it?

Target Audience

This project is particularly useful to business owner and/or investors looking to open or invest in new Coffee Shop in the capital city of like Jakarta. This project is timely as the city is currently suffering from oversupply of Coffee Shop. Data from the Indonesia Jajak Pendapat (JAKPAT - <https://blog.jakpat.net/indonesian-coffee-drinking-habit-survey-report/>) released last year showed that an additional 15.84% percent of people who drink Coffee will choose to get their coffee from the Coffee Shop and it will be grow in 2019 and 5.94% considering the location of Coffee Shop as the most important factor to buy a coffee. And many of business owner are continued obsession with building more Coffee Shop.

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 confined to the city of Jakarta, the capital city of the country of Indonesia in South East Asia.
- Latitude and longitude coordinates of those districts based on Indonesian Census and Goggle Maps (manual) Calibration. This is required in order to plot the map and also to get the venue data.
- 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

Data from census in Jakarta were scraped and longitude-latitude data were manually calibrated using Google Maps. Kepulauan Seribu were excluded from the data, because it is not related to the business problem. A total of 42 districts (Kecamatan) were collected. The cleaned data was extracted in *jkt_district.csv* file

<http://data.jakarta.go.id/dataset/jumlahkecamatankelurahanrtwdankkdkiakarta/resource/1d5b0bb0-3aa7-482a-9e65-fc03d466efac>

After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest database 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 Coffee Shop 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 working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).