Coursera Capstone
IBM Applied Data Science Capstone
Selecting Area to Open a New Shopping Center in Jakarta, Indonesia
By: Eldo Umboto Gultom
February 2020

Introduction

For many people, visiting shopping center is a great way to relax and enjoy themselves, mainly during weekend and holidays. They can do grocery shopping, dine, fashion shopping, watch movies etc. Shopping centers are like a one-stop destination for all types of people.

For retailers, the usually central location and the crowd at the shopping centers provides a great opportunity to market their products and services. Property developers are also taking advantage of this trend. Opening shopping centers allows possibility for property developers to earn consistent rental income.

As a result, there are many shopping centers in the city of Jakarta, even though they are currently chose more selectively which location to be built for they are requires serious consideration.

Particularly, the location of the shopping center is - in many aspects - will determine whether the center will be a success or a failure.

Business Problem

The objective of this project is to analyze and select the best locations in the city of Jakarta, Indonesia to open a new shopping center. Using data science methodology, this project aims to provide solutions to answer the business question: In Jakarta, Indonesia, if you are looking to open a new shopping center, where should you open it?

Target Audience of this project

This project is particularly useful to property developers or investors looking to open or invest in new shopping centers in Jakarta, Indonesia, currently capital city of Indonesia and soon to be ex capital city but still a business center city of Indonesia in many years to come.

Data

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

- Population per neighborhoods.
- Income per neighborhoods. The first two data (population and income) will be used to assume whether a neighborhood is opt out or not for candidacy.
- List of neighborhoods (Kelurahan) in Jakarta.
- Latitude and longitude coordinates of those neighborhoods. This is required in order to plot the map and also to get the venue data.
- Venue data, particularly data related to shopping centers. We will use this data to perform clustering on the neighborhoods.

Sources of data and methods to extract them

This City-provided web-page (http://data.jakarta.go.id/eu/dataset/jumlahpendudukberdasarkanusiaperkelurahandkijakarta/resource/7a6be211-4e8b-487c-a67a-bc796e793eb0) this data is population per neighborhoods (kelurahan), based on age as per 2017. Next data regarding income per neighborhoods (kelurahan) in Jakarta extracted from several CSV's :

1. http://data.jakarta.go.id/eu/dataset/data-kelahiran-bayi-berdasarkan-tempat-kelahiran-per-kelurahan/resource/cd84e15c-e22b-4b4d-b2f3-1422b74ea298
2. http://data.jakarta.go.id/eu/dataset/data-kelahiran-bayi-tahun-2017-berdasarkan-kategori-panjang-badan-dan-kelurahan/resource/da3b48cb-ef1c-4c2a-b45e-0605ff885072
3. http://data.jakarta.go.id/eu/dataset/data-kelahiran-bayi-berdasarkan-kategori-berat-badan-dan-kelurahan/resource/fc20b30d-bfc9-40ff-8d75-5d22546012fc

In short the data is about:

1. Birth data per neighborhoods (kelurahan).
2. Birth data per neighborhoods (kelurahan), based on the baby's length in cms (is it under or over 45 cm).
3. Birth data per neighborhoods (kelurahan), based on the baby's weight in gramss (is it under or over 2500 grams).

My assumptions are, if a neighborhood (kelurahan) has more than 10-35 percent birth under ideal criteria (would also be depend on population based on age); than that neighborhood (kelurahan) will be opt out for new shopping center location candidacy as I would assume this neighborhood's population income is not good enough for crowding the new shopping center.

For visualization, I will use the provided neighborhood list to get geographical coordinates of the neighborhoods (kelurahan) using Python Geocoder package which will give us the latitude and longitude coordinates of the neighborhoods (kelurahan).

After that, I will use Foursquare API to get the venue data for those sufficient neighborhoods. Foursquare API will provide many categories of the venue data, I particularly interested in the shopping center category in order to help us to solve the business problem.

This project will make use of several data science skills, like working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium).

In the next week, I will present the methodology, the steps taken in this project, the data analysis and the machine learning technique that was used.