Coursera Capstone IBM

Applied Data Science Capstone

Opening a New Shopping Mall in Jakarta, Indonesia

By: Gani Priambodo

April 2020

Image from : http://infopromodiskon.com/mall/profile/mal-taman-anggrek/

Introduction

Shopping mall is one of the bulding that forming a complex of shops with interconnecting walkways, usually indoors. They are the heart and soul of communities, the foundation of retail economies, and a social sanctuary for teenagers everywhere.

In Indonesia, amid rise of e-commerce shopping mall continue to lure costumers with their attractive design and promotion as well as culinary spots. Visitors no longer come to shopping malls primarily to buy next season's fashions; they go to socialize, to engage and be entertained. This is a new trend for malls to use as leisure venues. In fact, around a quarter of mall floor space is now taken up by food and entertainment and we expect this to rise to up to half, especially in newly built malls.

Head of the Indonesian Shopping Center Association (APPBI), Alexander Stefanus Ridwan said that Shopping Mall boast interaction and shopping experience that could not be satisfied by e-commerce.

Indonesia is home to 321 shopping centers of various class and size. Some of the malls are categorized as upscale, such as Pacific Place, Plaza Senayan, Plaza Indonesia and Senayan City, which all are located in Jakarta. However, it is Surabaya in East Java that actually hosts the biggest mall in Indonesia.

Therefore, Opening a shopping malls in Indonesia is one of the challenges which must be overcome by property developer with serious consideration, such as location of the mall. The location will drive the business in the mall will be success or failure.

Business problem

The objective of this capstone project is to analyze and select the best locations in Jakarta, Indonesia to open a new shopping mall. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In Jakarta, Indonesia, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

Target Audience

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in the capital city of Indonesia i.e. Jakarta. This project is timely as the city is currently suffering from oversupply of shopping malls. Colliers International said in its latest report that the retail space occupancy rate in the city stagnated at 83.6 percent throughout 2018 and is expected to fall to between 82 percent and 82.5 percent in 2019. Colliers senior associate director Ferry Salanto said that, as long as landlords and retailers were willing to adjust to the trends, the growth of the occupancy rate would not be affected by online shopping. In order to thrive, Ferry suggested that mall construction projects take into account infrastructure and transportation accessibility in addition to catering to millennials.

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.

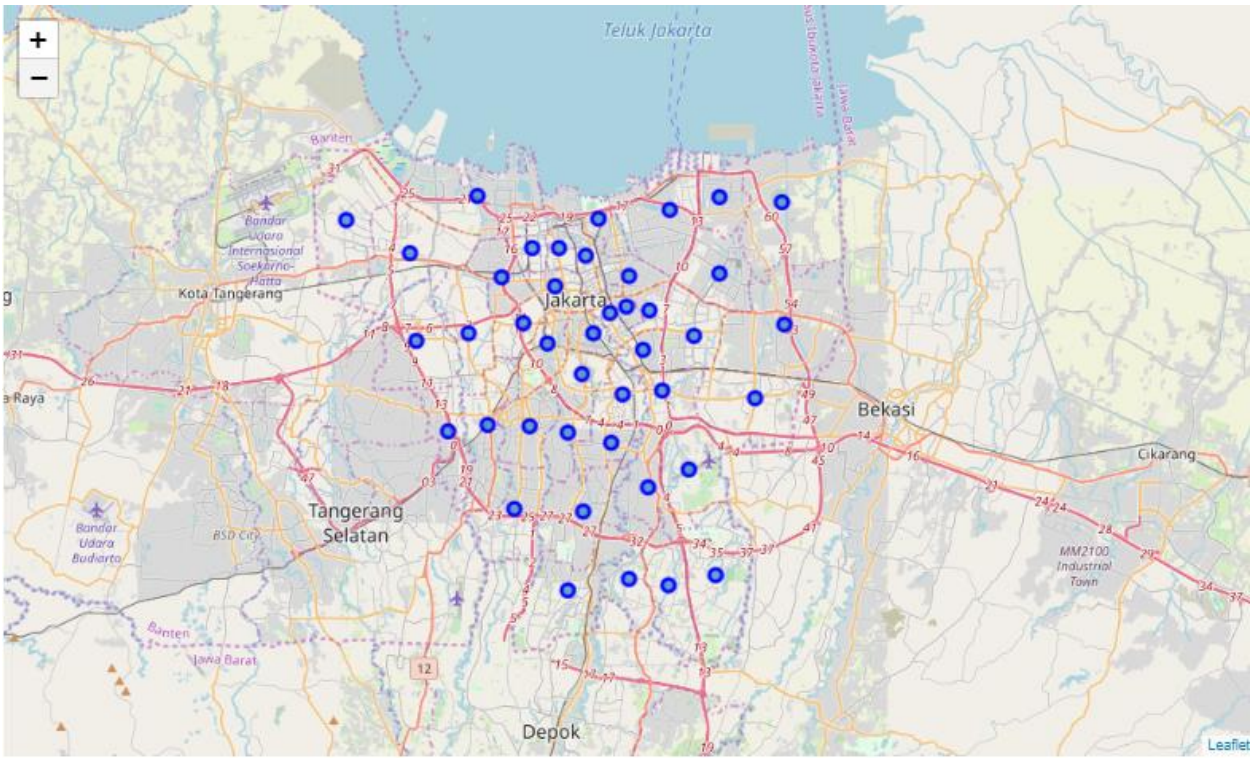
• Latitude and longitude coordinates of those districts based on Indonesian Census and Goggle Maps (manual) Calibration. This is required in order to plot the map and also to get the venue data.

• Venue data, particularly data related to Shopping Mall. 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_2.csv.
(http://data.jakarta.go.id/dataset/jumlahkecamatankelurahanrtrwdankkdkijakarta/resource/1d5b0b b0-3aa7-482a-9e65-fc03d466efac) After that, we will use Foursquare API to get the venue data for those neighbourhoods. Foursquare has one of the largest databases 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 Shopping Malls 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).

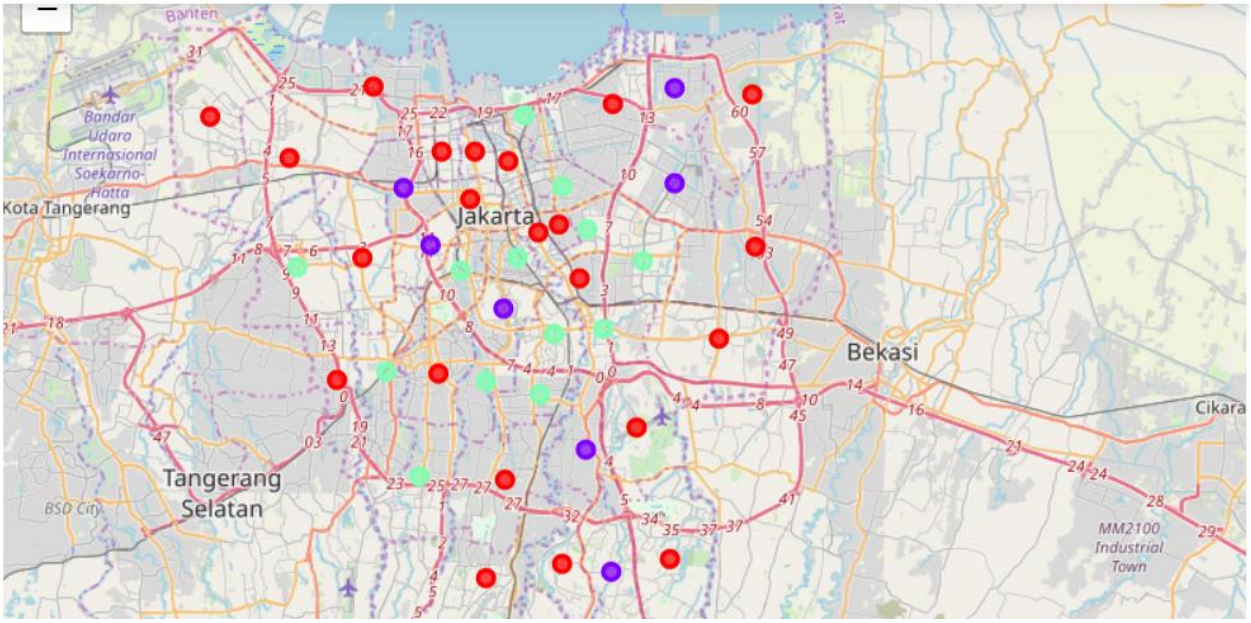
Firstly, we need to get the list of districts in Jakarta. Fortunately, the list of districts data have been collected from
http://data.jakarta.go.id/dataset/jumlahkecamatankelurahanrtrwdankkdkijakarta/resource/1d5b0b b0-3aa7-482a-9e65-fc03d466efac 4 and manually search centre of districts using Google Maps. Data can be seen in in jkt_district.csv file. After gathering the data, we will populate the data into a pandas DataFrame and then visualize the neighbourhoods in a map using Folium package. This allows us to perform a sanity check to make sure that the geographical coordinates data returned by Geocoder are correctly plotted in Jakarta City as shown in image 1



Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. We need to register a Foursquare Developer Account in order to obtain the Foursquare ID and Foursquare secret key. We then make API calls to Foursquare passing in the geographical coordinates of the neighbourhoods in a Python loop. Foursquare will return the venue data in JSON format and we will extract the venue name, venue category, venue latitude and longitude. With the data, we can check how many venues were returned for each neighbourhood and examine how many unique categories can be curated from all the returned venues. Then, we will analyse each neighbourhood by grouping the rows by neighbourhood and taking the mean of the frequency of occurrence of each venue category. By doing so, we are also preparing the data for use in clustering. Since we are analysing the “Shopping Mall” data, we will filter the “Shopping Mall” as venue category for the neighbourhoods. Lastly, we will perform clustering on the data by using k-means clustering. K-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is one of the simplest and popular unsupervised machine learning algorithms and is particularly suited to solve the problem for this project. We will cluster the neighbourhoods into 3 clusters based on their frequency of occurrence for “Shopping Mall”. The results will allow us to identify which neighbourhoods have higher concentration of shopping malls while which neighbourhoods have fewer number of shopping malls. Based on the occurrence of shopping malls in different neighbourhoods, it will help us to answer the question as to which neighbourhoods are most suitable to open new shopping malls.

Results

The results from the k-means clustering show that we can categorize the neighbourhoods into 3 clusters based on the frequency of occurrence for “Shopping Mall”:



• Cluster 0: Neighbourhoods with low number to no existence of shopping malls

• Cluster 1: Neighbourhoods with high concentration of shopping malls

• Cluster 2: Neighbourhoods with moderate number of shopping malls

The results of the clustering are visualized in the map below with cluster 0 in red colour, cluster 1 in purple colour, and cluster 2 in mint green colour. 6 Discussion As observations noted from the map in the Results section, most of the shopping malls are concentrated in the south area of Jakarta city, with the highest number in cluster 1 and moderate number in cluster 2. On the other hand, cluster 0 has very low number to no shopping mall in the neighbourhoods. This represents a great opportunity and high potential areas to open new shopping malls as there is very little to no competition from existing malls. Meanwhile, shopping malls in cluster 1 are likely suffering from intense competition due to oversupply and high concentration of shopping malls. From another perspective, the results also show that the oversupply of shopping malls mostly happened in the south area of the city. Therefore, this project recommends property developers to capitalize on these findings to open new shopping malls in neighbourhoods in cluster 0 with little to no competition. Property developers with unique selling propositions to stand out from the competition can also open new shopping malls in neighbourhoods in cluster 2 with moderate competition. Lastly, property developers are advised to avoid neighbourhoods in cluster 1 which already have high concentration of shopping malls and suffering from intense competition.

Limitations and Suggestions for Future Research

In this project, we only consider one factor i.e. frequency of occurrence of shopping malls, there are other factors such as population and income of residents that could influence the location decision of a new shopping mall. However, to the best knowledge of this researcher such data are not available to the neighbourhood level required by this project. Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new shopping mall. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that

came with limitations as to the number of API calls and results returned. Future research could make use of paid account to bypass these limitations and obtain more results.

Conclusion

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing machine learning by clustering the data into 3 clusters based on their similarities, and lastly providing recommendations to the relevant stakeholders i.e. property developers and investors regarding the best locations to open a new shopping mall. To answer the business question that was raised in the introduction section, the answer proposed by this project is: The neighbourhoods in cluster 0 are the most preferred locations to open a new shopping mall. The findings of this project will help the relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decisions to open a new shopping mall.