IBM CAPSTONE PROJECT - DS

Entering the coffee shop market in Jakarta, Indonesia

By: Ega Kurnia Yazid

INTRODUCTION

Background

Indonesia is one of the world largest coffee exporters and drinking coffee has been Indonesian morning mantra for centuries. Therefore, considering Indonesia potential as the home of many coffee varieties; then, bringing Indonesia coffee culture a step further will be a great movement for the industry. Moreover, on-the-go delivery coffee or beverages is becoming a new hype these days. Utilizing these opportunities, then opening a new coffee shop at the right time will be really challenging. Then, it is important to be careful to pick a strategic location to open a new coffee shop.

Business Problem

This report aims to figure out the best location to enter the coffee market in Jakarta, Indonesia. Using data science and machine learning capacities such as clustering, this project will answer the less coffee shop district around Jakarta, so that it will be less competitive to enter the coffee shop market.

The smell of fresh-made coffee is one of the world's greatest inventions. - Hugh Jackman

Jakarta, March 14, 2020

Ega Kurnia Yazid
Project's Data Scientist

DATA

To respond to the problem, hence, this project will use multiple datasets, containing:

- List of neighborhoods (districts) in Jakarta, Indonesian capital city;
- Latitude and longitude data that showing the coordinate of its neighborhood;
- Venue data which informs the nearby venue around the neighborhoods.

Sources of data and method to extract them

The main neighborhood data is retrieved from the Wikipedia page which contains information about Jakarta's districts (https://en.wikipedia.org/wiki/Category:Districts_of_Jakarta), with total of 44 districts. This data retrieved using Python, particularly by scraping with BeautifulSoup package. Furthermore, geographical coordinates dataset is retrieved by using Geocoder package which allow me to get the geolocation of Jakarta's districts. Along with it, I used Foursquare API to get the venue data around the districts and, finally, map the results using Folium.

METHODOLOGY

Clustering

Clustering is an unsupervised machine learning that aims to cluster a data based on similar characteristics. To address the business problem, then clustering is a relevant method to cluster the coffeeshop density around Jakarta's districts.

K-Means Clustering

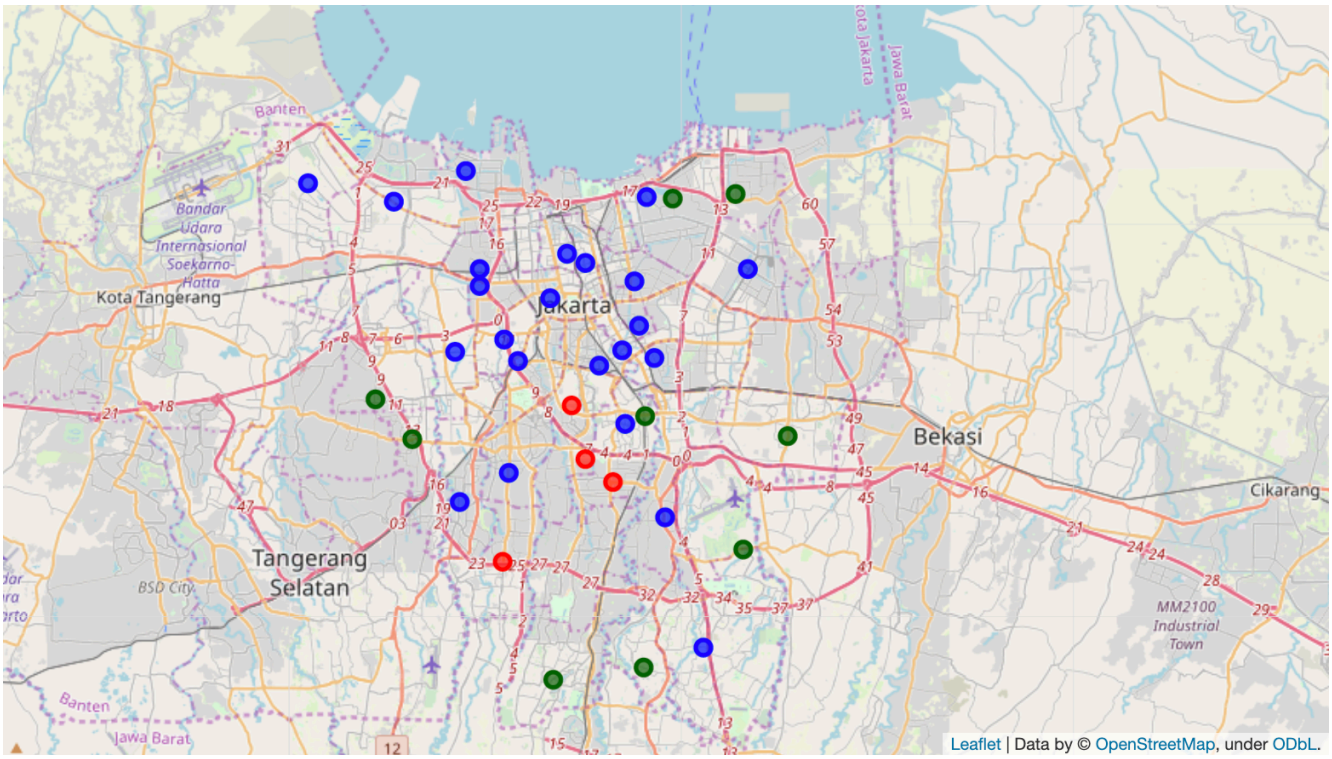
There are different types of clustering with specific condition and usage. Particularly, for this business problem, K-means clustering become the best fit to cluster the data. K-means, specifically, will calculate the similarity between the coffeeshop density inter-districts. Then, it will divide the districts into three main clusters. The first one is a cluster with the least amount of coffeeshop, the second one is a cluster with a medium amount of coffee shop, and finally, the third cluster is a cluster with the highest density of coffee shop.

Analysis

After mapping the cluster upon the Jakarta map, then it will be analyzed so that we will know the best location(s) to open a new coffeeshop. The best location(s) will be selected according to the least coffeeshop density. This is important, to minimize the competition to entry the coffee shop market.

RESULTS

Cluster around DKI Jakarta



Discussion

This research attempts to investigate the distribution of coffeeshop around DKI Jakarta; also, it figures out the best location to open a new coffeeshop considering the its density in Jakarta. Simply, the best location to open new coffeeshop is determined by finding out the locations that is relatively containing less coffee shop. Therefore, by applying a clustering method, I try to cluster jakarta districts by three regions. Cluster 0 is a cluster that contain some coffee shops. Secondly, Cluster 1 is a cluster that contain less coffee shops, where relatively located at uptown. Finally, Cluster 2 is a cluster that contain the highest density coffee shop, where relatively located at downtown Jakarta. Consequently, according to the data, the best location to open new coffeeshop is around the uptown Jakarta, where it relatively populous because it islocated at residential. Moreover, the concept of the coffee shop may be fit with the visitors' profile and its main market will be good if it focused on weekend.

CONCLUSION

Uptown Coffeeshop

To conclude, after applying coffeeshop clustering, then the data has shown that the least coffeeshop density is located around outer ring of Jakarta, where mostly a residential place. Therefore, building a coffeeshop around it will be an advantage because it will be less competitive. Along with it, the coffee shop will be interesting if they can focus on residential market, and they can boost their promo at night or on the weekend.