COURSERA
CAPSTONE PROJECT

Introduction

South Jakarta (Indonesian: Jakarta Selatan) is one of the five administrative cities (kota administrasi) which form Special Capital Region of Jakarta, Indonesia. South Jakarta is not self-governed and does not have a city council, hence it is not classified as a proper municipality. It had a population of 2,057,080 at the 2010 Census, and is the third most populous among the five administrative cities of Jakarta, after East Jakarta and West Jakarta.

,Many shopping malls in the city of South Jakarta are being built. Opening shopping malls allows property developers to earn consistent rental income. Of course, as with any business decision, opening a new shopping mall requires serious consideration and is a lot more complicated than it seems. Particularly, the location of the shopping mall is one of the most important decisions that will determine whether the mall will be a success or a failure

Bussiness Problem and Target Audience

The objective of this capstone project is to analyse and select the best locations in the city of South Jkarta 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 the city of South Jakarta, Indonesia, if a property developer is looking to open a new shopping mall, where would you recommend that they open it?

This project is particularly useful to property developers and investors looking to open or invest in new shopping malls in the South Jakarta specifically.

Data

To solve the problem, we will need the following data:•Categories of Siuth Jakarta. This defines the scope of this project which is confined to the city of South Jakart, Indonesia.

Latitude and longitude coordinates of those neighbourhoods. This is required in order to plot the map and also to get the venue data.

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 coordinates of the neighbourhoods using Python Geocoder package which will give us the latitude and longitude coordinates of the neighbourhoods.

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 Shopping Mall 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 web scraping (Wikipedia), working with API (Foursquare), data cleaning, data wrangling, to machine learning (K-means clustering) and map visualization (Folium). In the next section, we will present the Methodology section where we will discuss the steps taken in this project, the data analysis that we did and the machine learning technique that was used

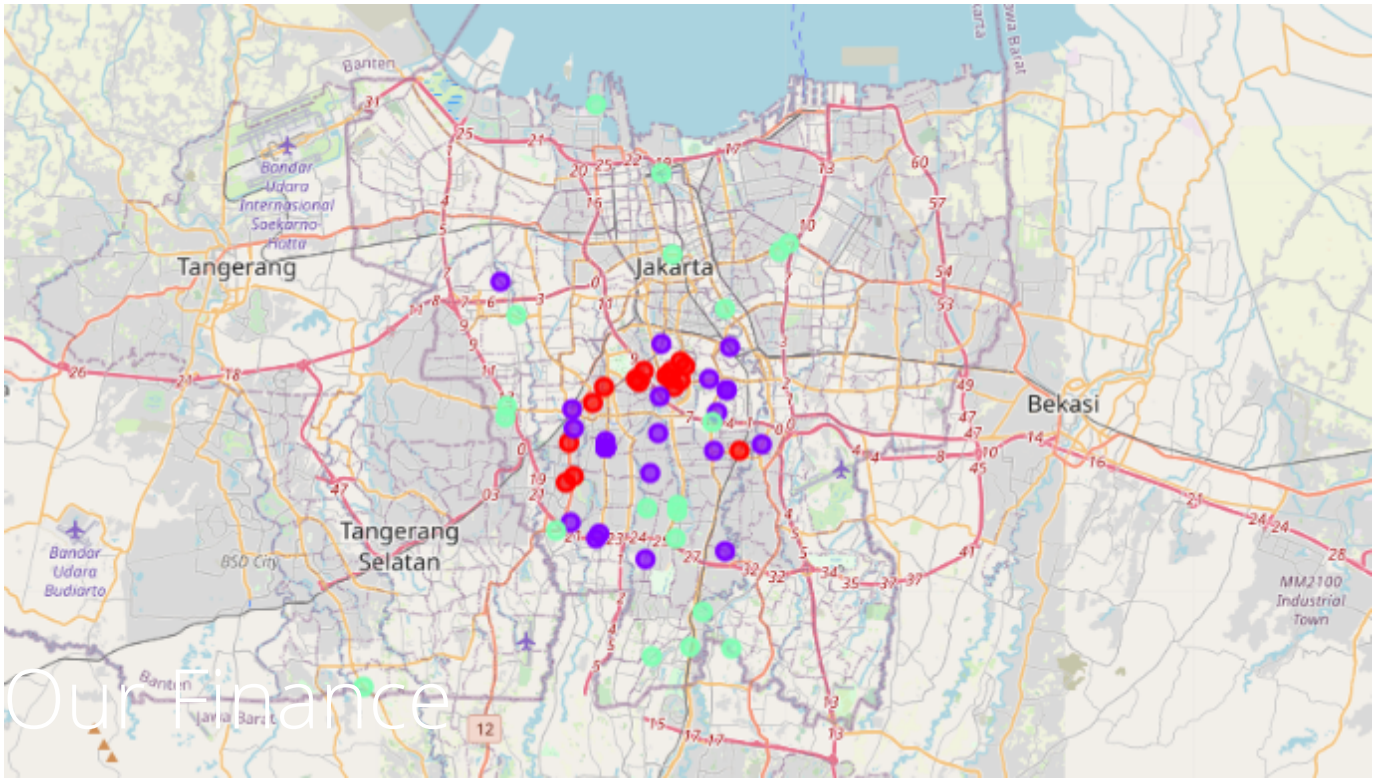
Methodology

1. Web scrapping wikipedia page for neighborhoohds list
2. Get latitude and longtitude coordinates using Geocoder
3. Use Forsquare API to get venue data
4. Group data by neighborhood d taking the mean of the frequency of occurence of each venue category
5. Filter venue category by Shopping Mall
6. Perform clustering on the data by using k-means clustering
7. Visualize the clusters in a mpa using Folium

Results

1. Categorized the neighborhoods with moderate number of shopping malls

a. Cluster 0: Neighborhoods with low number of shopping malls
b. Cluster 1: Neighborhoods with low number to no exstence of shopping malls
c. Cluster 3: Neighborhoods with high concentration of shopping malls



Discussion

1. Most of the shopping malls are concentrated in the central of area the city
2. Highest number in cluster 2 and moderate number in cluster 0
3. Cluster 1 has very low number to no shopping mall in the neighborhoods
4. Oversupply of shopping malls mostly happened in the central area of the city with the suburb area still have very few shopping malls

Recommendations

1. Open new shopping malls in neighborhoods in cluster 1 with little to no competition
2. Can also open in neighborhoods in cluster 0 with moderae competition if it has unique selling propositions to stand out from the competition
3. Avoid neighborhoods in cluster 2, already high concentration of shopping malls and intense competition

Conclusion

1. Answer to business question: The neighborhoods in cluster 1 are the most preferred locations to open a new shopping malss
2. Findings od this prohject will helep relevant stakeholders to capitalize on the opportunities on high potential locations while avoiding overcrowded areas in their decision to open a new shopping mall.