COURSERA CAPSTONE

IBM Data Science Professional - Applied Data Science Capstone

Where to Open a Mall in Jakarta

Created By: Leonardo Wijaya

April 2021

1. Problem Description

If we want to go to just one place to do everything like shopping, having dinner, kids recreation, and also watching a movie, we want to go to a place called a Mall. Mall is a good place to relax and refresh yourself from this tiring working world. Not only beneficial for visitors, malls also offer potential new branch places for companies, adding income to property developers, and also a workplace for the unemployed.

In Jakarta, the process of building to opening a new mall is a complicated process. But in that process, there are many things to consider especially from a location perspective. Is the location big enough? Are there many people living there so that the mall is easily approachable? Where are the best possible places to open the mall?

This project is made in order to help property developers who want to build a new mall in Jakarta and also Jakarta landlords who probably and by any means want to build something out of their lands.

2. Data Description

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

- Districts in Jakarta
- Population Density for every district
- District size in how many sub districts there are

The data the writer is collecting comes from wikipedia scraping from this link. The writer will also use geopy to get longitude and latitude to further visualize best possible places to build malls in Jakarta. Also the writer will use population density data to know places and districts in Jakarta.

3. Methodologies

a. Data Collecting and Preprocessing

In the first step, the writer uses BeautifulSoup to scrape https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Daerah_Khusus_Ibukota_Jakarta page to get districts in Jakarta. The writer then converted the webpage to a pandas dataframe containing district code, district name, city part the district is in, and also the number of sub districts in that particular district to determine size of the district. There are a total of 44 districts in Jakarta.

The second data, the population density data is already clean csv data from https://data.go.id/dataset/jumlah-penduduk-wajib-ktp-dki-jakarta.

b. Add Latitude and Longitude Data

After the district data is collected, the writer uses Geopy Geocoding to get longitude and latitude from each district to later ease up the visualization.

c. Add Population Density Data

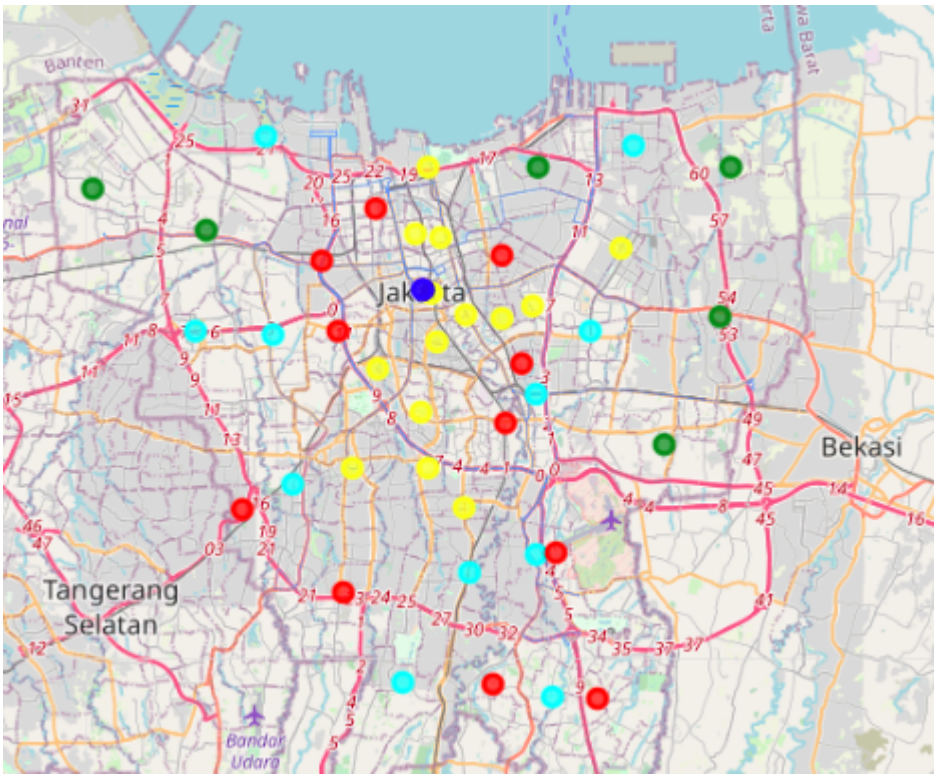
The second data, the population density data, consists of the number of population in every sub districts and every gender, so to get the population number of every district, the writer grouped the data to districts and sum of the number of population for that district.

To combine with district data, the formatting of district names in the population density data needs to be normalized because by default it is different from the district data. After combining, the writer did some validations regarding whether the district number was still the same 44 etc.

d. Clustering

The final step is the clustering process. The writer sets the KMeans to 5 different clusters; smallest, smaller, average, bigger, and biggest. The higher the number of population density and subdistricts, the better the chance the future mall succeeded.

4. Results



- Smallest districts, smallest population density: cluster blue
- Smaller districts, smaller population density: cluster yellow
- Average districts, average population density: cluster cyan
- Bigger districts, bigger population density: cluster red
- Biggest districts, biggest population density: cluster green -- Best places to build a Mall

5. Discussions

Even though nowadays malls in Jakarta are located in the center of the cities, that does not enclose the possibility and the potential of building malls in the places stated above. Because at the end people will tend to visit malls nearby their houses and therefore it is a great opportunity to build malls in large areas with many sub districts and high population.

Note that this segmentation is based on very simple assumptions and not to be taken as critical as the real considerations being made by property developers in the real world before building a mall. Things like average land price and weather also helps determine where to build a new mall in Jakarta. So the writer recommends adding more considerations regarding this subject for further improvements for this project.

6. Conclusions

After going through several processes from identifying business problems to getting districts that have great potentials to build a mall, the writer have the conclusion that it is best to build a mall in this districts: Cilincing, Tanjung Priok, Cakung, Duren Sawit, Cengkareng, Kalideres in Jakarta. Even though it slightly deviates from the current trend to build malls in the center of the city, having to see malls being nearer to people is a very nice sight.