Jakarta: Battle of Neighborhood

A. Business Problem

Jakarta is the capital of Indonesia, the largest city in the country, and one of the most populous urban agglomerations on earth. The population of Jakarta metropolitan area, with an area of 6,343 km2 (2,449 sq mi), was 31.6 million according to the Indonesia 2015 Inter-Census, making it the most populous region in Indonesia, as well as the second most populous urban area in the world after Tokyo. The population share of Jakarta metropolitan area to national population increased from 6.1% in 1961 to 11.26% in 2010. Today about 20% of Indonesia's urban population is concentrated in the Jakarta metropolitan area.

Jakarta is now considered a global city and one of the fastest growing economies in the world. Interestingly, Jakarta reported the highest return on investment for luxury real estate in 2014 compared to any other city on earth. Considering the facts, many investors are eager to open their business in Jakarta nowadays, which makes it so important that they choose the right location according to their line of business. However, it is not easy to get the information and it also requires an analytical thinking.

To address the above problem, this project will be focusing on determining the location to open new hotels in Jakarta metropolitan area, with considering the distance to city center, number of existing hotels nearby, and total of population in the area. Those factors contribute in the likelihood of optimal decision.

The result will be recommendation of location candidates according to data analysis method.

B. Data Definition

There are some data needed to solve the business problem as follow.

- Center of Jakarta is defined as ‘Setiabudi’ area and its coordinate along with neighborhoods will be obtained using **Google Maps API**.
- Number of existing hotels in the neighborhood. The data will be generated using **Foursquare API**.
- Number of populations in the neighborhood. For this case, the data will be gathered from **Jakarta Government public data repositories** and converted to csv.

All the required data will be merged into a dataset and will be used for further analysis. Data will be presented in map visualization which allows us to see the location in a quick glance.

C. Methodology

Below is the methodology that will be used for analysis.

1. Data Collection

I define Jakarta city center coordinates using Google Maps API, collected the location of hotels within ~6 km using Foursquare and gathered population within the neighborhoods district from Jakarta public datasets.

2. Exploratory Data Analysis

Calculate the density across neighborhood using heatmaps to identify recommended area with criteria as follow:

- Close to city center
- Low number of hotels
- High number of populations

3. Clustering

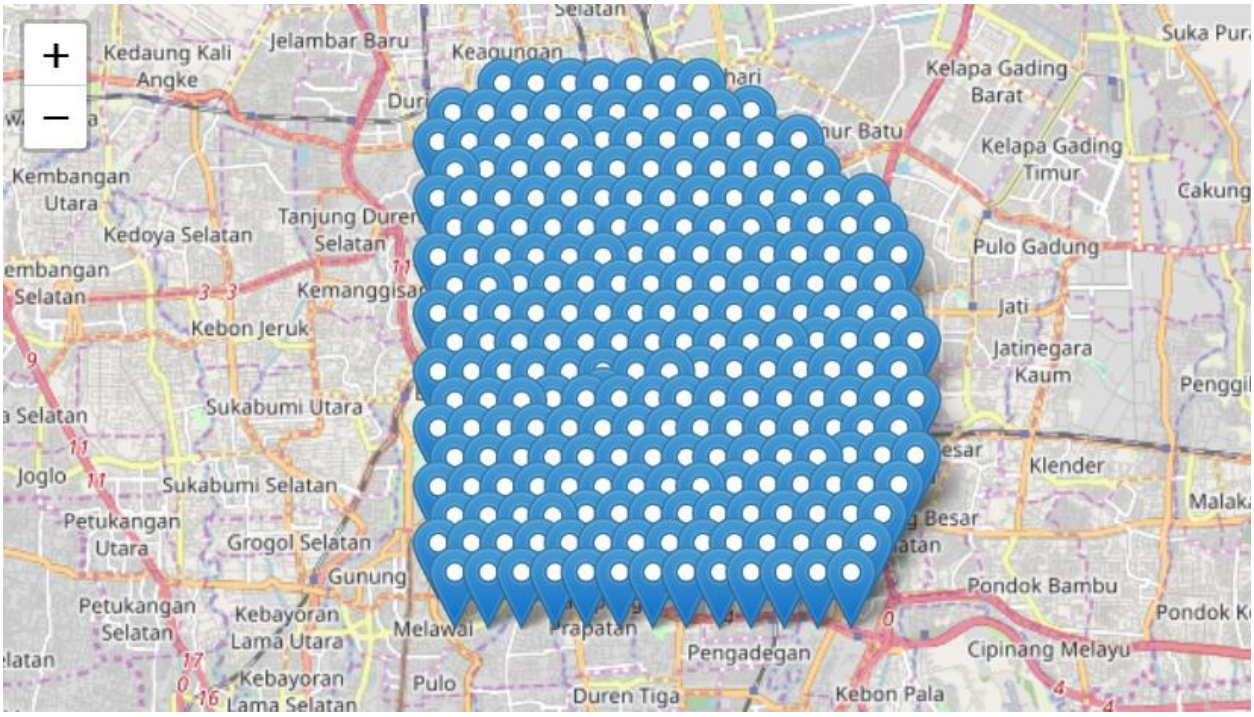
Data analysis will be presented using k-means clustering in Folium map. The recommendation will be defined by focusing on locations close to city center (less than 2.5 km) with no hotels within radius 1 km and more than 150,000 total population.

C.1 Data Collection

A. Get Neighborhood Candidates using Google Maps API

First thing I need to gather is the coordinate of Jakarta city center defined as 'Setiabudi area' using Google Maps API so that we can explore the surrounding neighborhood. The coordinate is found in WGS84 latitude and longitude format [-6.2160764, 106.83001] and converted into UTM X, Y [702479.2103012843, -687440.4328693151] for determining area distances.

The surrounding neighborhood is defined as ~6 km away from city center, as circular areas with a radius of 300 meters and its center will be 600 meters apart from each other. From the analysis, I generated 244 neighborhood candidate centers in a List and display them in a Folium map.



I produce the location address of the coordinates using the Google Maps API and put them into a DataFrame. The data will be used for determining surrounding venues in further analysis. Below is the example of data.

	Address	Latitude	Longitude	X	Y	Distance from center
0	5, Jl. Iskandarsyah I No.18, RT.5/RW.4, Melawa...	-6.245414	106.805719	699779.210301	-690675.815777	4213.988913
1	Jl. Pulo Raya III No.27, RT.1/RW.3, Petogogan,...	-6.245395	106.811140	700379.210301	-690675.815777	3857.162501
2	Jl. Bangka II Gg. V No.21, RT.13/RW.2, Pela Ma...	-6.245376	106.816562	700979.210301	-690675.815777	3566.188800
3	Jl. Pondok Karya VIII Blok I No.72, RT.11/RW.4...	-6.245358	106.821983	701579.210301	-690675.815777	3358.229080
4	Jl. Mampang Prapatan IV No.19, RT.2/RW.5, Mamp...	-6.245339	106.827404	702179.210301	-690675.815777	3249.261848

B. Get Venue in Neighborhood using Foursquare API

I need to collect top 100 venue information within 500 meters from provided address since we will be focusing on areas with no hotel nearby. There were 6025 venues returned by Foursquare. There were couple of things to be processed.

- We only need ‘Hotel’ venue category, so the DataFrame was filtered and returned 304 venues
- We want to merge district name with population data in next analysis, so Neighborhood column value was replaced to standardize district naming convention, e.g. ‘Kec.’, ‘Kecamatan’ and new column was created
- We need venue location in UTM X, Y to measure its distance to each other, so we convert them and create new columns [Venue_x] and [Venue_y]

The resulted datasets can be presented as follow.

Neighborhood	Neighborhood_Latitude	Neighborhood_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category	Kecamatan	Venue_x	Venue_y
5, Jl. Iskandarsyah I No.18, RT.5/RW.4, Melawa...	-6.245414	106.805719	Hotel GranDhika Iskandarsyah Jakarta	-6.244813	106.803989	Hotel	Kby. Baru	699587.931649	-690608.756764
5, Jl. Iskandarsyah I No.18, RT.5/RW.4, Melawa...	-6.245414	106.805719	Hotel Ambhara	-6.243266	106.803678	Hotel	Kby. Baru	699554.124192	-690437.517698
5, Jl. Iskandarsyah I No.18, RT.5/RW.4, Melawa...	-6.245414	106.805719	favehotel Melawai	-6.244321	106.801337	Hotel	Kby. Baru	699294.688423	-690553.296449
Jl. Raya Pasar Minggu No.2 B, RT.2/RW.2, Panco...	-6.245282	106.843668	Cipta Hotel Pancoran	-6.247249	106.843335	Hotel	Pancoran	703941.588910	-690893.251751
Jl. Raya Pasar Minggu No.2 B, RT.2/RW.2, Panco...	-6.245282	106.843668	Amaris Hotel Pancoran	-6.247507	106.842834	Hotel	Pancoran	703886.021918	-690921.551012
Jl. Raya Pasar Minggu No.2 B, RT.2/RW.2, Panco...	-6.245282	106.843668	V Hotel Jakarta	-6.241865	106.844744	Hotel	Pancoran	704099.618926	-690298.267303

C. Get Population Data from Government Public Repositories

Population data was gathered from [Jakarta Government Repositories]. This will be merged into existing dataset and a new column will be created to indicate high population with more than 150,000 in total.

	Nama_Kecamatan	Pria	Wanita	Total_Population
0	Menteng	44475	43628	88103
1	Cemp. Putih	47964	47310	95274
2	Senen	63212	60421	123633
3	Taman Sari	63567	62323	125890
4	Johar Baru	69901	66604	136505

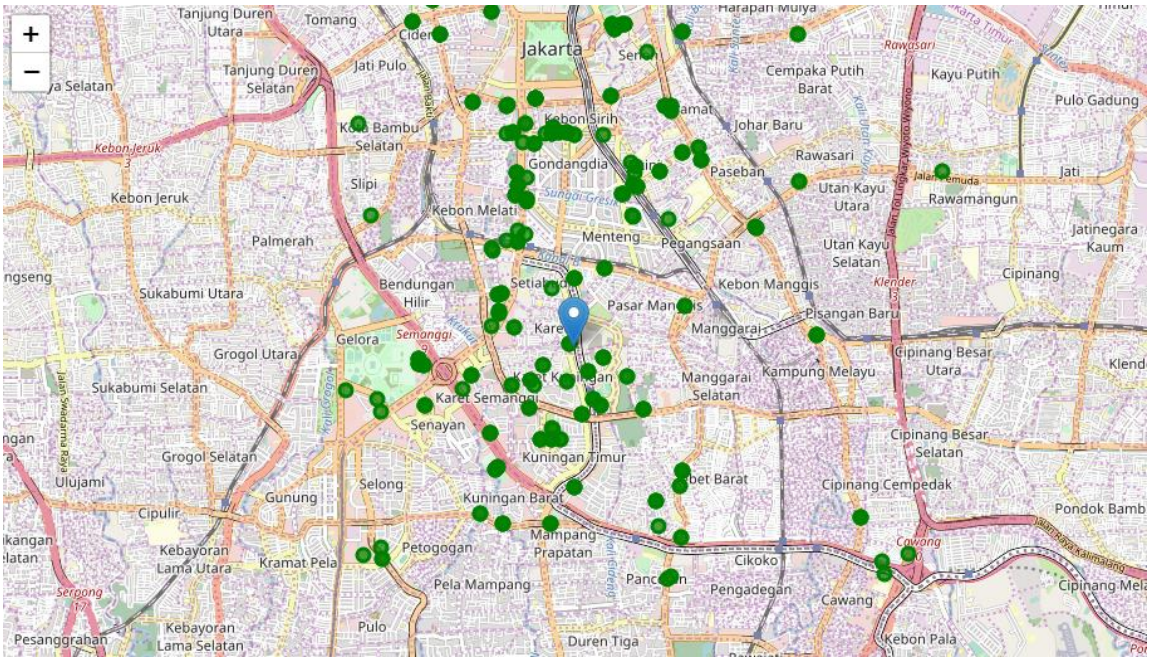
The final datasets as shown below.

Neighborhood	Neighborhood_Latitude	Neighborhood_Longitude	Venue	Venue_Latitude	Venue_Longitude	Venue_Category	Kecamatan	Venue_x	Venue_y	Nama_Kecamatan	Pria	Wanita	Total_Population	Is_PopulationHigh
5, Jl. Iskandarsyah I No.18, RT.5/RW.4, Melawa...	-6.245414	106.805719	Hotel GranDhika Iskandarsyah Jakarta	-6.244813	106.803989	Hotel	Kby. Baru	699587.931649	-690608.756764	Kby. Baru	74982.0	73362.0	148344.0	False
5, Jl. Iskandarsyah I No.18, RT.5/RW.4, Melawa...	-6.245414	106.805719	Hotel Ambhara	-6.243266	106.803678	Hotel	Kby. Baru	699554.124192	-690437.517698	Kby. Baru	74982.0	73362.0	148344.0	False
5, Jl. Iskandarsyah I No.18, RT.5/RW.4, Melawa...	-6.245414	106.805719	favehotel Melawai	-6.244321	106.801337	Hotel	Kby. Baru	699294.688423	-690553.296449	Kby. Baru	74982.0	73362.0	148344.0	False
Jl. Raya Pasar Minggu No.2 B, RT.2/RW.2, Panco...	-6.245282	106.843668	Cipta Hotel Pancoran	-6.247249	106.843335	Hotel	Pancoran	703941.588910	-690893.251751	Pancoran	78803.0	76393.0	155196.0	True
Jl. Raya Pasar Minggu No.2 B, RT.2/RW.2, Panco...	-6.245282	106.843668	Amaris Hotel Pancoran	-6.247507	106.842834	Hotel	Pancoran	703886.021918	-690921.551012	Pancoran	78803.0	76393.0	155196.0	True

By now, we already know the location of hotels in surrounding neighborhoods from Jakarta city center, along with population indicator. The data collection and preparation phase are completed and we're ready to move into analysis phase.

C.2 Exploratory Data Analysis

I try to visualize the datasets in map to see dispersion of hotels in surrounding neighborhoods.



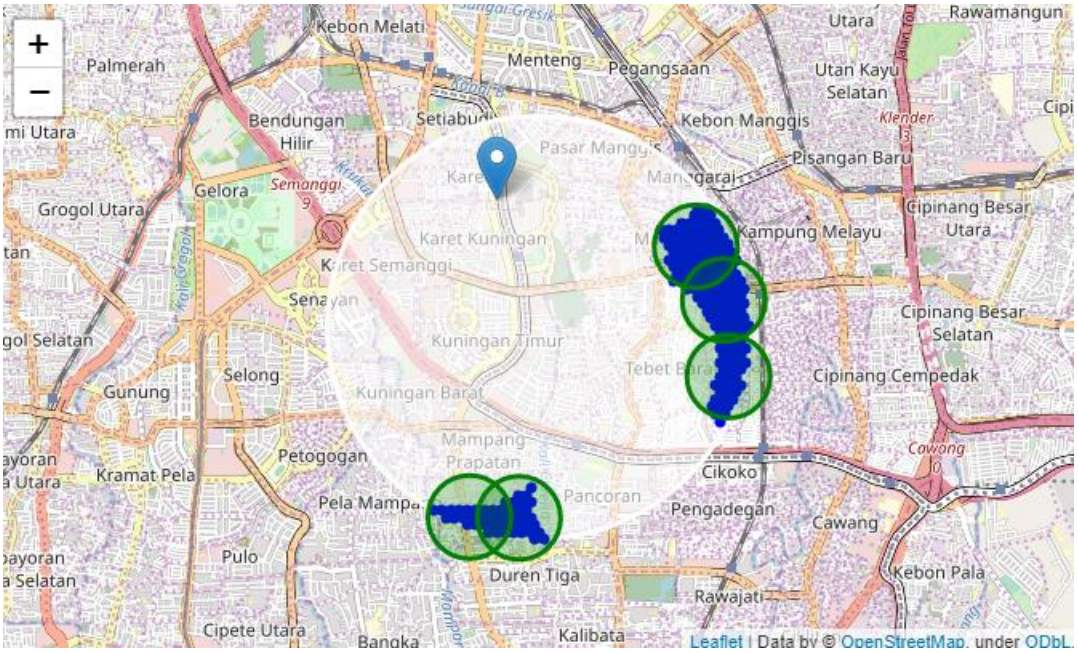
A heatmap was created for hotel density indicator and the marker color was categorized to show black color for high population area.

We can see the low hotel density closest to city center are around south from Setiabudi area, along with high population area. With that, then we can move our focus for area of interest to the south part.

The new center of interest covers the low density of hotels, high population area, and still close enough to city center. I define the new neighborhood within ~2,5 km from center and generate 2261 location candidates with each count the number of hotels in 1 km radius. I want to focus only for location with no hotels nearby and from the analysis, I get 199 ‘good’ candidates. Next step is to cluster them into 5 and find the center of location.

C.3 K-Means Clustering

The result of clustering divides our good candidates into 5 clusters as shown in map visualization below.



The exact address of cluster centers can be generated by reverse geocoding the coordinates using Google Maps API. From that, we have 5 addresses for location candidates:

1. 6, Jl. Tebet Utara III B No.15, RT.7/RW.2, Tebet Tim., Kec. Tebet
=> 2.9km from Setiabudi

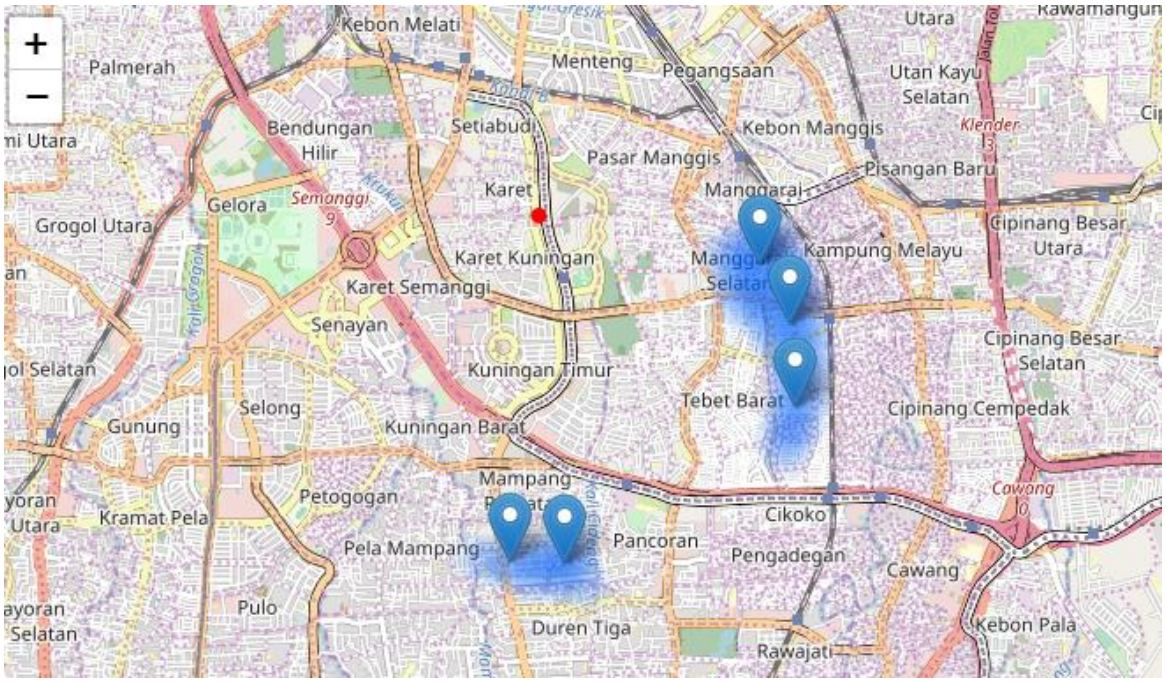
2. Jl. Tegal Parang Selatan V No.47, RT.4/RW.7, Tegal Parang, Kec. Mampang Prpt
=> 3.8km from Setiabudi

3. Jl. Tebet Timur II G No.2, RT.9/RW.5, Tebet Tim., Kec. Tebet
=> 3.5km from Setiabudi

4. Jl. Mampang Prpt. Raya No.3a, RT.7/RW.1, Tegal Parang, Kec. Mampang Prpt
=> 3.7km from Setiabudi

5. 14, Jl. Barkah III No.29, RT.14/RW.2, Manggarai Sel., Kec. Tebet
=> 2.4km from Setiabudi

Below is the final visualization with recommended location as a marker. This will be presented to our investors for their consideration in making decision.



D. Result and Discussion

From analysis that I have done, we can see there's about 144 hotels within 3 km from Jakarta City Center (Setiabudi area) with the closest low-density area detected in south. Additionally, high population density also mainly located in the south so we can focus our interest to the area.

After shifting our focus, I filtered the location candidates to include only area with no hotels within 1 km. Those location were then clustered to create 5 zones of interest which contain greatest number of location candidates. Those are the recommended location for investors to open a new hotel. Finally, I generated the address of the zone centers using reverse geocoding to be used as markers for more detailed analysis based on other factors.

Although we have the recommendation, there are some other factors need to be considered further, such as tourist attraction, socioeconomics and government regulation that are not covered in this analysis. Thus, the result can not be solely used for making decision.

E. Conclusion

Problem addressed in this project is how to get the required information and do analysis for determining the best location candidates in opening new hotel by stakeholder/investors. By generating datasets and clustering the locations, we have identified 5 areas to be considered for further analysis. The recommendation can be used by investors in decision making process.