

Jakarta Land Price and Venues

Sofyan Assauri

May 2020

1. Introduction

1.1 Background

Jakarta is one of the largest metropolises in the world where over 10,7 million people live and it has a population density of 4.282 people per square kilometer. As a resident of this city, I decided to use Jakarta in my project. The city is divided into 42 districts in total.

1.2 Problem

As you can see from the figures, Jakarta is a city with a high population and population density. Being such a crowded city leads the owners of shops and social sharing places in the city where the population is dense. When we think of it by the investor, we expect from them to prefer the districts where there is a lower real estate cost and the type of business they want to install is less intense. If we think of the city residents, they may want to choose the regions where real estate values are lower, too. At the same time, they may want to choose the district according to the social places density. However, it is difficult to obtain information that will guide investors in this direction, nowadays.

When we consider all these problems, we can create a map and information chart where the real estate index is placed on Istanbul and each district is clustered according to the venue density.

1.3 Interest

I believe with the methodology, data, and tools used in this project is relevant for an investor who want to expand their business. The use of Foursquare data, mapping techniques, and data analysis will help resolve the questions. This project is also a good practical case for a person developing Data Science skill.

2. Data acquisition and cleaning

2.1 Data sources

To consider the problem we can list the datas as below:

- I used Foursquare API to get the most common venues of given Borough of Jakarta.
- 2018 per square meter Nilai Jual Objek Pajak (NJOP) or land tax valuation averages for each Borough of Jakarta from Jakarta Tax Office.
- Google Map, 'Search Nearby' option to get the center coordinates of the each Borough.

2.2 Data cleaning

There are several obstacles in data management mainly due to data limitations in names and coordinates (latitude and longitude) and land value data in Jakarta. While venue data from Foursquare is good enough and can be used immediately.

For coordinate data of each Borough in Jakarta, done through the Google Map for 42 Boroughs in Jakarta, the recording is done manually and needs to be converted into coordinates that can be read by the Foursquare API.

As for the land tax value data, it is taken from the Jakarta Tax Office data. The data is in PDF format and contains land values per square meter for each road in Jakarta. then the data is converted into CSV form and then analyzed by finding the average value of land prices for each Neighborhood through the median of the lowest and highest values, then the value will be averaged with all neighborhoods in a Borough.

3. Exploratory Data Analysis

The Analysis began with the preparation of master data which has the main components Borough, Neighborhood, ZIP, Average Land Price, Latitude and Longitude informations of the city which I downloaded from my Github repository.

	Borough	Neighborhood	ZIP	Latitude	Longitude	Avg. Land Price
0	Cengkareng	Cengkareng Barat, Cengkareng Timur, Duri Kosam...	11730, 11750, 11720, 11710, 11740	-6.1487	106.7353	7,734
1	Grogol Petamburan	Grogol, Jelambar, Jelambar Baru, Tanjung Duren...	11450, 11460, 11470, 11440	-6.1617	106.7846	14,706
2	Kalideres	Kalideres, Kamal, Pegadungan, Semanan, Tegal Alur	11840, 11810, 11830, 11850, 11820	-6.1343	106.7058	6,809
3	Kebon Jeruk	Duri Kepa, Kebon Jeruk, Kedoya Selatan, Kedoya...	11510, 11530, 11520, 11550, 11560, 11540	-6.1835	106.7647	9,922
4	Kembangan	Joglo, Kembangan Selatan, Kembangan Utara, Mer...	11640, 11610, 11650, 11620, 11630	-6.1930	106.7426	12,675

Table 1. Jakarta Borough and average of its land price.

Next, using Python folium library I try to visualize geographic details of Jakarta and its Boroughs and I created a map of Jakarta with boroughs superimposed on top. I used latitude and longitude values from table before to get the visual as below:

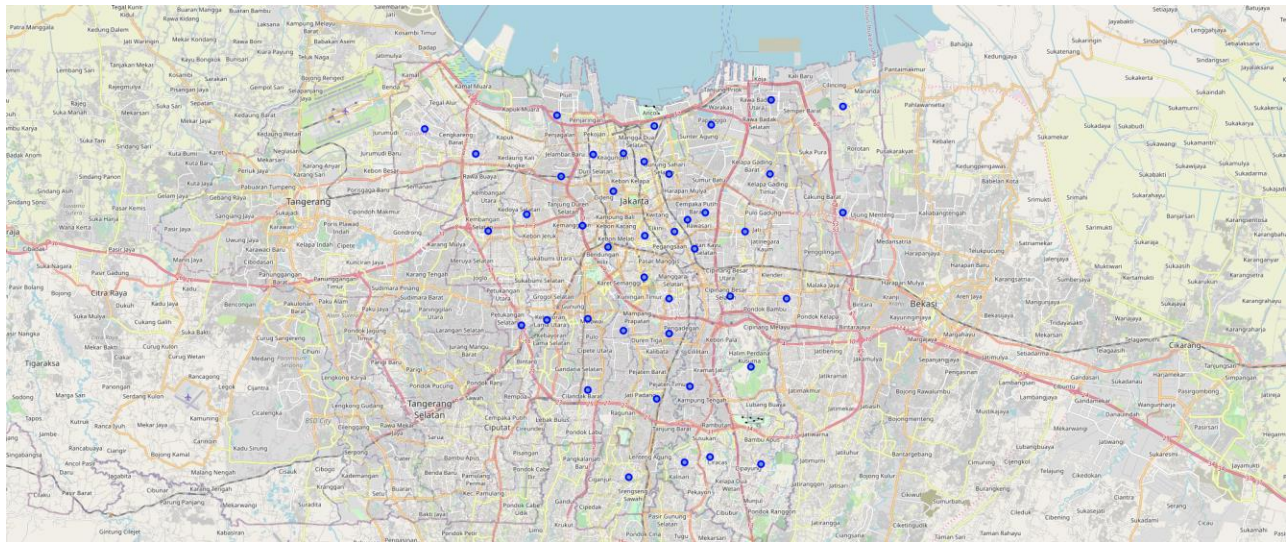


Figure 1. Jakarta Map generated from Folium

I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venue and the radius 1000 meter for each borough from their given latitude and longitude informations. Here is a head of the list Venues name, category, latitude and longitude informations from Forsquare API.

	name	categories	lat	Ing
0	Rumah Makan Medan Baru	Indonesian Restaurant	-6.158865	106.831763
1	Gado Gado Tamansari Raya	Salad Place	-6.152726	106.827043
2	Masterpiece KTV	Karaoke Bar	-6.148706	106.833402
3	Starbucks - Rs Husada	Coffee Shop	-6.147282	106.828670
4	Swiss-Belhotel Mangga Besar	Hotel	-6.149521	106.833898
5	Es Campur Ko Acia	Dessert Shop	-6.148447	106.829944

Tab;e 2. Venues generated from Foursquare API

In summary of this data 92 venues were returned by Foursquare. The result doesn't mean that inquiry run all the possible results in boroughs. Actually, it depends on given Latitude and Longitude informations and here is we just run single Latitude and Longitude pair for each borough. We can increase the possibilities with Neighborhood informations with more Latitude and Longitude informations.

Then I created a table which shows list of top 10 venue category for each borough in Table below.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ancol, Pademangan Barat, Pademangan Timur	Seafood Restaurant	Coffee Shop	Hotel	Theme Park	Fast Food Restaurant	Sundanese Restaurant	Fried Chicken Joint	Chinese Restaurant	Arts & Crafts Store	General Entertainment
1	Angke, Duri Selatan, Duri Utara, Jembatan Besi...	Chinese Restaurant	Indonesian Restaurant	Asian Restaurant	Convenience Store	Seafood Restaurant	Train Station	Fast Food Restaurant	Coffee Shop	Food Truck	Beer Garden
2	Balekambang, Batampar, Cawang, Cililitan, Duk...	Restaurant	Coffee Shop	Middle Eastern Restaurant	Department Store	Mediterranean Restaurant	Bakery	Pizza Place	Food Truck	Gym	Food Court
3	Bali Mester, Bidaracna, Cipinang Besar Selata...	Coffee Shop	Asian Restaurant	Food Truck	Indonesian Restaurant	Middle Eastern Restaurant	Cafe	Soccer Stadium	Mosque	Food Court	Shopping Mall
4	Bambu Apus, Ceger, Cilangip, Cipayung, Lubang...	Garden	Pizza Place	Farmers Market	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food Stand	Food Court	Food & Drink Shop
5	Bangka, Kuningan Barat, Mampang Prapatan, Pela...	Restaurant	Coffee Shop	Asian Restaurant	Noodle House	Food Truck	Bakery	Snack Place	Bar	Convenience Store	Fast Food Restaurant
6	Baru, Cijantung, Gedung, Kalikari, Pekayon	Indonesian Restaurant	Food Truck	Grocery Store	Basketball Court	Soccer Stadium	Convenience Store	Pizza Place	Stadium	Bakery	Food Court
7	Bendungan Hilir, Gelaora, Kampung Bali, Karet T...	Indonesian Restaurant	Coffee Shop	Pizza Place	Convenience Store	Japanese Restaurant	Food Truck	Restaurant	Lounge	Snack Place	Pool
8	Bintaro, Pesanggrahan, Petukangan Selatan, Pet...	Noodle House	Indonesian Restaurant	Convenience Store	Pizza Place	Fla Market	Wings Joint	Fast Food Restaurant	Fried Chicken Joint	French Restaurant	Food Truck
9	Bukit Duri, Kebon Baru, Manggarai, Manggarai S...	Indonesian Restaurant	Coffee Shop	Cafe	Seafood Restaurant	Convenience Store	Noodle House	Restaurant	Steakhouse	Park	Asian Restaurant

Table 3. top 10 venue category for each borough

We have some common venue categories in Boroughs. In this reason I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning. Then, I run K-Means to cluster the boroughs into 5 clusters.

Borough	Neighborhood	ZIP	Latitude	Longitude	Avg. Land Price	Cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Cempingang	Cempingang Barat, Cempingang Timur, Duri Kiri...	11730, 11750, 11720, 11710, 11740	-6.1487	106.7353	7.734	1	Department Store	Pet Store	Movie Theater	Music Venue	Cafe	Farmers Market	Fast Food Restaurant	Donut Shop	Dumping Restaurant
1	Orong Palatunuran	Orong, Jembatan, Jembatan Baru, Tanjung Duran...	11450, 11460, 11470, 11440	-6.1617	106.7546	14.706	1	Noodle House	Chinese Restaurant	Indonesian Restaurant	Asian Restaurant	Food Truck	Hotel	Coffee Shop	Cafe	Fast Food Restaurant
2	Kalideres	Kalideres, Kamal, Pegatungan, Semanan, Tegal Alur	11840, 11810, 11830, 11850, 11820	-6.1343	106.7056	6.809	0	Noodle House	Asian Restaurant	Fast Food Restaurant	Pizza Place	Convenience Store	Food & Drink Shop	Cafe	Supermarket	Fried Chicken Joint
3	Kebon Jenuk	Duri Kapa, Kebon Jenuk, Kedyra Selatan, Kedyra...	11510, 11530, 11520, 11550, 11560, 11540	-6.1635	106.7647	9.922	1	Food Truck	Indonesian Restaurant	Asian Restaurant	Convenience Store	Fast Food Restaurant	Restaurant	Coffee Shop	Noodle House	Food Court
4	Kembangan	Jepin, Kembangan Selatan, Kembangan Utara, Sbr...	11640, 11610, 11650, 11620, 11630	-6.1930	106.7426	12.675	1	Asian Restaurant	Coffee Shop	Chinese Restaurant	Japanese Restaurant	Gym / Fitness Center	Noodle House	Cafe	Bubble Tea Shop	Food Court

Table 4. Cluster of the Borough based on common venues

After that, we analyze the common vanues for each Borough. We can also estimate the number of 1st Most Common Venue in each cluster by understand the Borough through categorization. to help readers understand the nuances of each Borough, each cluster will be named according to the common venues, as follows

cluster 0: Middle East Feast

cluster 1: Kitchen of Indonesia

cluster 2: Social Venues

cluster 3: Coffee Shop Galore

cluster 4: Snackbites Joint

We can also examine that what is the frequency of average land prices in different ranges. Thus, histogram can help to visualization:

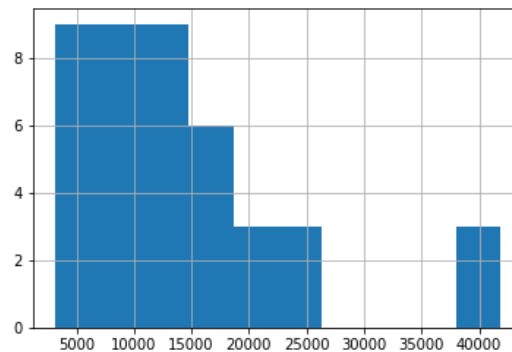


Figure 2. Jakarta land price histogram

As it seems in above histogram, we can define the ranges as below (in thousand IDR)

Avg. Land Price (in Thousand IDR)	Level
0-7000	Low 1
7000 - 14000	Low 2
14000 – 21000	Med 1
21000 – 28000	Med 2
28000 – 35000	Med 3
35000 – 42000	High 1
<42000	High 2

Table 5. Average land price level

Then we combine the level of the average land price with the master table for enriching the master data as follows:

	Borough	Neighborhood	ZIP	Latitude	Longitude	Price m2/sq	Land Price Level
0	Cengkareng	Cengkareng Barat, Cengkareng Timur, Duri Kosam...	11730, 11750, 11720, 11710, 11740	-6.1487	106.7353	7734	Low 2
1	Grogol Petamburan	Grogol, Jelambar, Jelambar Baru, Tanjung Duren...	11450, 11460, 11470, 11440	-6.1617	106.7846	14706	Mid 1
2	Kalideres	Kalideres, Kamal, Pegadungan, Semanan, Tegal Alur	11840, 11810, 11830, 11850, 11820	-6.1343	106.7058	6809	Low 1
3	Kebon Jeruk	Duri Kepa, Kebon Jeruk, Kedoya Selatan, Kedoya...	11510, 11530, 11520, 11550, 11560, 11540	-6.1835	106.7647	9922	Low 2
4	Kembangan	Joglo, Kembangan Selatan, Kembangan Utara, Mer...	11640, 11610, 11650, 11620, 11630	-6.1930	106.7426	12675	Low 2

Table6. Average land price level in Master Data

4. Result

After merging those new variables with related cluster informations in our main master table, we can now see Join, Labels and Level_labels columns as the last three ones in above table. we can also see a clustered map boroughs of Jakarta in the below.

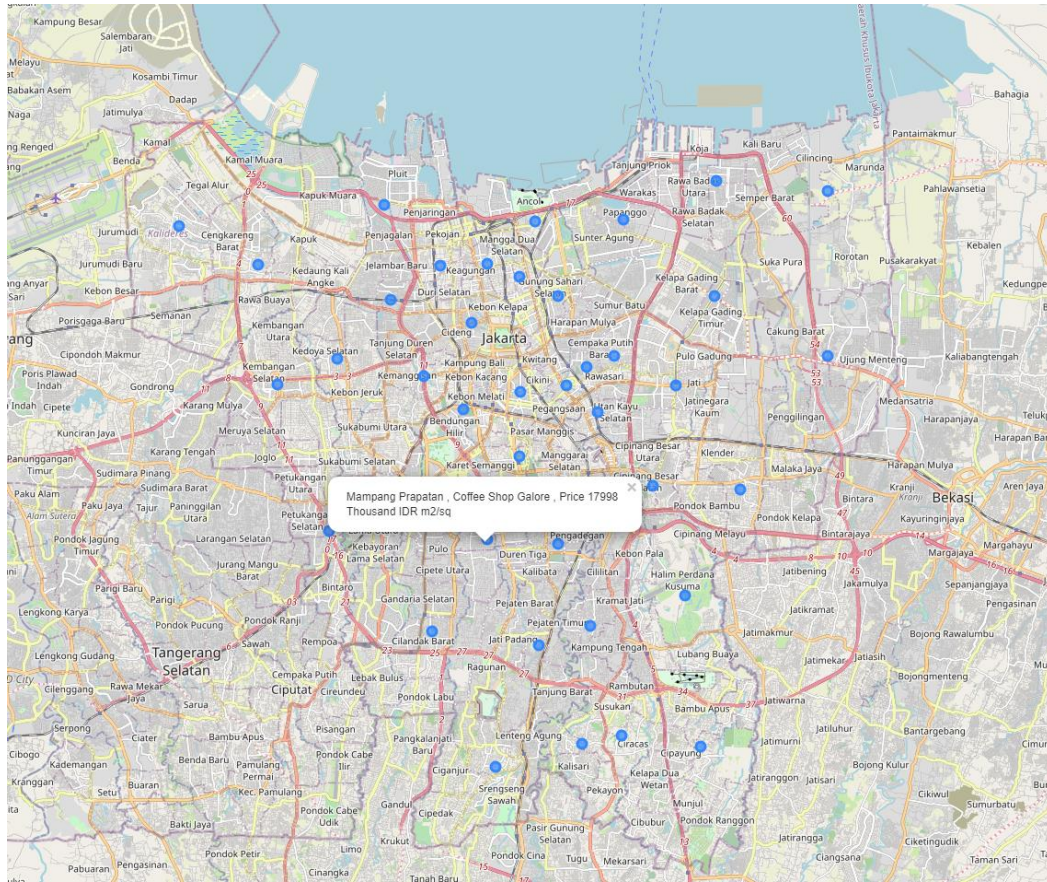


Figure 3. Integrated Map of Jakarta

5. Discussion

As we mentioned before, Jakarta is a big city with a high population density in a narrow area. The total number of measurements and population densities of the 42 districts in total can vary. As there is such a complexity, very different approaches can be tried in clustering and classification studies. Moreover, it is obvious that not every classification method can yield the same high quality results for this metropol.

We used the Kmeans algorithm as part of this clustering study, we set the optimum k value to 5. I also performed data analysis through this information by adding the coordinates of districts and Average land price averages as static data on GitHub. In future studies, these data can also be accessed dynamically from specific platforms or packages.

I ended the study by visualizing the data and clustering information on the Jakarta map. In future studies, web or telephone applications can be carried out to direct investors.

5. Conclusions

As a result, people are turning to big cities to start a business or work. For this reason, people can achieve better outcomes through their access to the platforms where such information is provided. Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.