

Housing Sales Prices & Venues in Jabodetabek

Havez Vazirani Al Kautsar

This study is inspired by Sercan Yildiz [House Sales Prices and Venues in Istanbul \(https://towardsdatascience.com/housing-sales-prices-venues-in-istanbul-dbdefb8b06f0\)](https://towardsdatascience.com/housing-sales-prices-venues-in-istanbul-dbdefb8b06f0) article, thank you so much Sercan !!!

A. Introduction

Description & Discussion of the Background

Jakarta is the capital and largest city of Indonesia. Situated on the northwest coast of the world's most populous island of Java, it is the centre of economy, culture and politics of Indonesia with a population of more than ten million as of 2014. Albeit only covers 699.5 square kilometres, the smallest among any Indonesian provinces, its metropolitan area covers 6,392 square kilometres; it is the world's second most populous urban area after Tokyo, with a population of about 30 million as of 2010. Jakarta's business opportunities, as well as its potential to offer a higher standard of living, have attracted migrants from across the Indonesian archipelago, making it a melting pot of numerous cultures. The city proper has a very high population density of 14,464 people per square kilometer, while the metro area has a density of 4,383 people/sq km. [1]

Not only Jakarta residents, Jakarta is filled with workers from outside Jakarta every day, so it is often called Jabodetabek (Jakarta, Bogor, Depok, Tangerang, Bekasi) or Greater Jakarta area. Lots of public transportation makes it easy for these workers to go to Jakarta even though the distance is quite far adrift. Therefore, there are a lot of investors who have started investing in the Jabodetabek area, and this study will try to explore parts of Jabodetabek related to relatively low property prices with low intensity of business competition.

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

Data Description

To consider the problem we can list the datas as belows

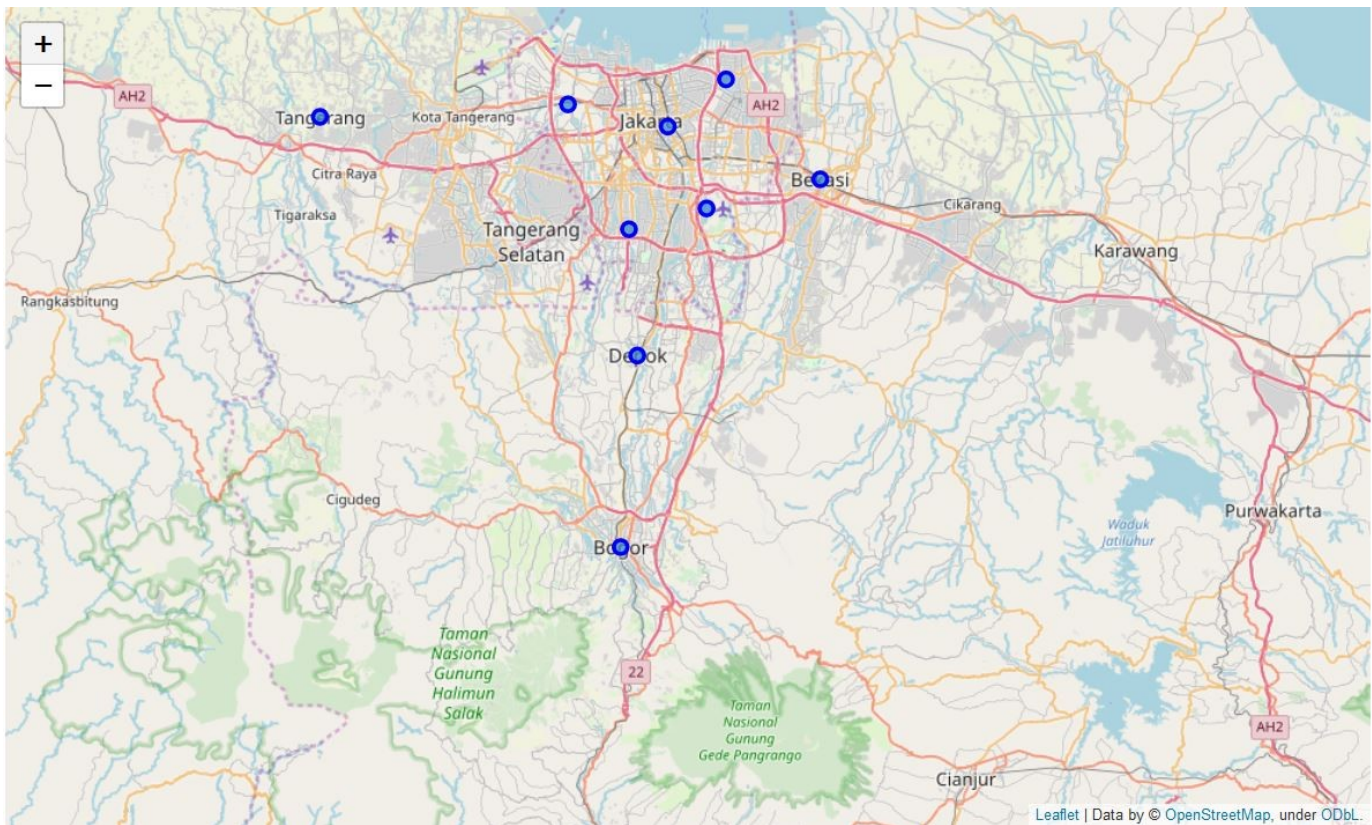
- I used **Forsquare API** to get the most common venues of given cities of Jabodetabek [2].
- There are not too many public datas related to demographic and social parameters for the cities of Jabodetabek. Therefore you must set-up your own data tables in most cases. In this case, I collected latest House Price Averages for each Cities of Jabodetabek from housing retail web page [3].

B. Methodology

My master data **df** which has the main components *Cities*, and *Average House Price* informations of the cities.

	city	price (rp/m2)	latitude	longitude
0	Central Jakarta	26997450	-6.18233995	106.84287153600738
1	North Jakarta	25199166	-6.1362053	106.90069097114528
2	South Jakarta	22044101	-6.28381815	106.80486349194814
3	West Jakarta	17971174	-6.16156235	106.74389124027667
4	East Jakarta	13004566	-6.26289085	106.88222894692834
5	Tangerang	10498210	-6.1734808	106.4992042
6	Bekasi	5953004	-6.2349858	106.9945444
7	Depok	6584794	-6.4074657	106.8138131
8	Bogor	5811302	-6.5962986	106.7972421

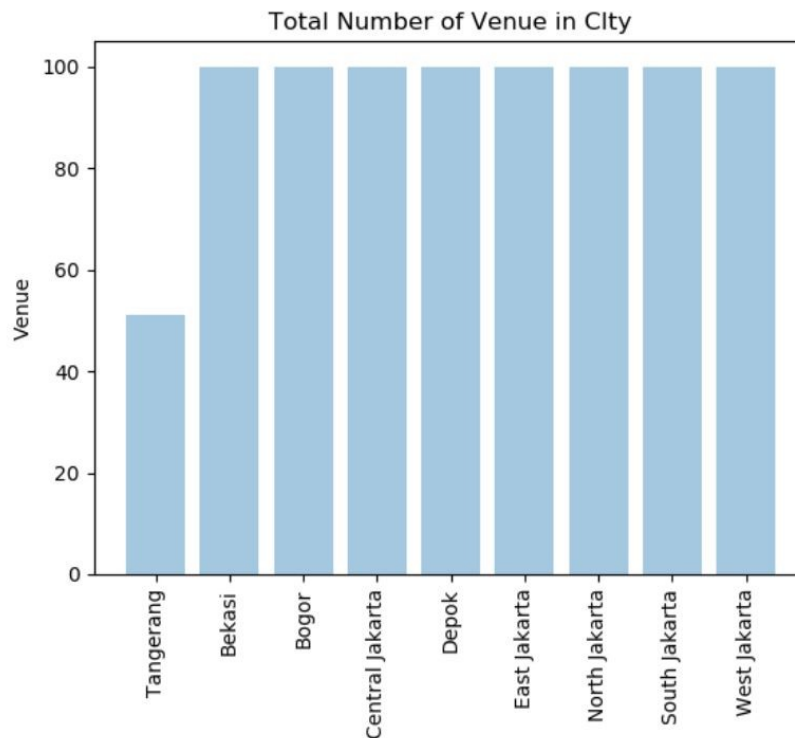
I used python folium library to visualize geographic details of Jabodetabek and it's cities and I created a map of Jabodetabek with cities superimposed on top. I used latitude and longitude values to get the visual as below:



With foursquare API's, i tried to explore what and how many venues that exist in the cities of Jabodetabek. I limit the radius of 10 km for each city in Jabodetabek, Foursquare API also only limits 100 venues for each coordinate.

	City	City Latitude	City Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Central Jakarta	-6.18234	106.842872	Bunga Pepaya	-6.189087	106.835859	Manadonese Restaurant
1	Central Jakarta	-6.18234	106.842872	Starbucks	-6.187875	106.836585	Coffee Shop
2	Central Jakarta	-6.18234	106.842872	Mercure Jakarta Cikini	-6.194325	106.839137	Hotel
3	Central Jakarta	-6.18234	106.842872	DoubleTree by Hilton	-6.198763	106.843126	Hotel
4	Central Jakarta	-6.18234	106.842872	The Hermitage	-6.198318	106.838120	Hotel

Then we can create a bar chart and analyze the number of venues of each cities.



Here is we can see that 8 of 9 cities reached the **100** limit of venues. On the other hand; Tangerang is below **60** venues in our given coordinates with Latitude and Longitude.

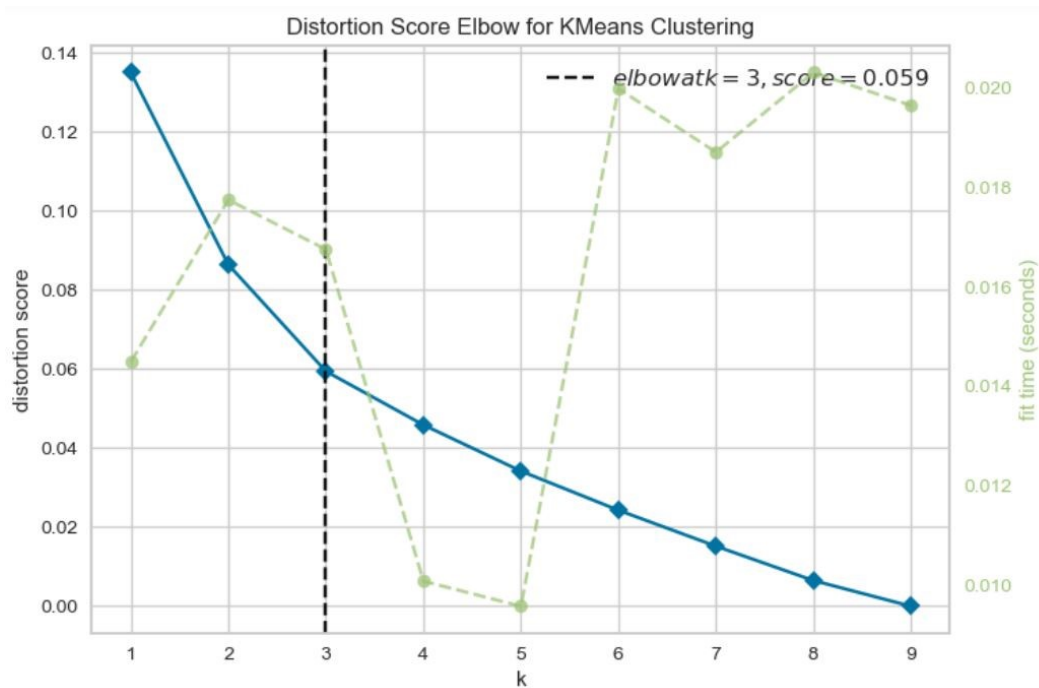
The result doesn't mean that inquiry run all the possible results in cities. Actually, it depends on given Latitude and Longitude informations and here is we just run single Latitude and Longitude pair for each city. We can increase the possibilities with Neighborhood informations with more Latitude and Longitude informations.

There are 144 unques categories of venue on Jabodetabek by Foursquare, then I created a table which shows list of top 10 venue category for each borough in below table.

	City	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	Bekasi	Indonesian Restaurant	Coffee Shop	Fast Food Restaurant	Asian Restaurant	Restaurant	Multiplex	Seafood Restaurant	Shopping Mall	Department Store	Movie Theater
1	Bogor	Coffee Shop	Hotel	Indonesian Restaurant	Bakery	Café	Noodle House	Restaurant	Fast Food Restaurant	Asian Restaurant	Seafood Restaurant
2	Central Jakarta	Coffee Shop	Hotel	Sushi Restaurant	Restaurant	Shopping Mall	Fast Food Restaurant	Clothing Store	Bakery	Javanese Restaurant	Dessert Shop
3	Depok	Indonesian Restaurant	Coffee Shop	Food Truck	Fast Food Restaurant	Café	Bakery	Snack Place	Golf Course	Pizza Place	Seafood Restaurant
4	East Jakarta	Coffee Shop	Hotel	Fast Food Restaurant	Restaurant	Indonesian Restaurant	Padangnese Restaurant	Bakery	Sushi Restaurant	Steakhouse	Ice Cream Shop
5	North Jakarta	Coffee Shop	Indonesian Restaurant	Noodle House	Asian Restaurant	Chinese Restaurant	Shopping Mall	Multiplex	Beach	Restaurant	Bakery
6	South Jakarta	Coffee Shop	Hotel	Indonesian Restaurant	Bakery	Shabu-Shabu Restaurant	Restaurant	Steakhouse	Padangnese Restaurant	Asian Restaurant	Motorcycle Shop
7	Tangerang	Indonesian Restaurant	Asian Restaurant	Convenience Store	Fast Food Restaurant	Coffee Shop	Furniture / Home Store	Pool	Japanese Restaurant	Nightclub	Seafood Restaurant
8	West Jakarta	Coffee Shop	Seafood Restaurant	Indonesian Restaurant	Asian Restaurant	Korean Restaurant	Sushi Restaurant	Chinese Restaurant	Café	Pizza Place	Shopping Mall

We have some common venue categories in cities. 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**. I will use K-Means algorithm for my study in this project.

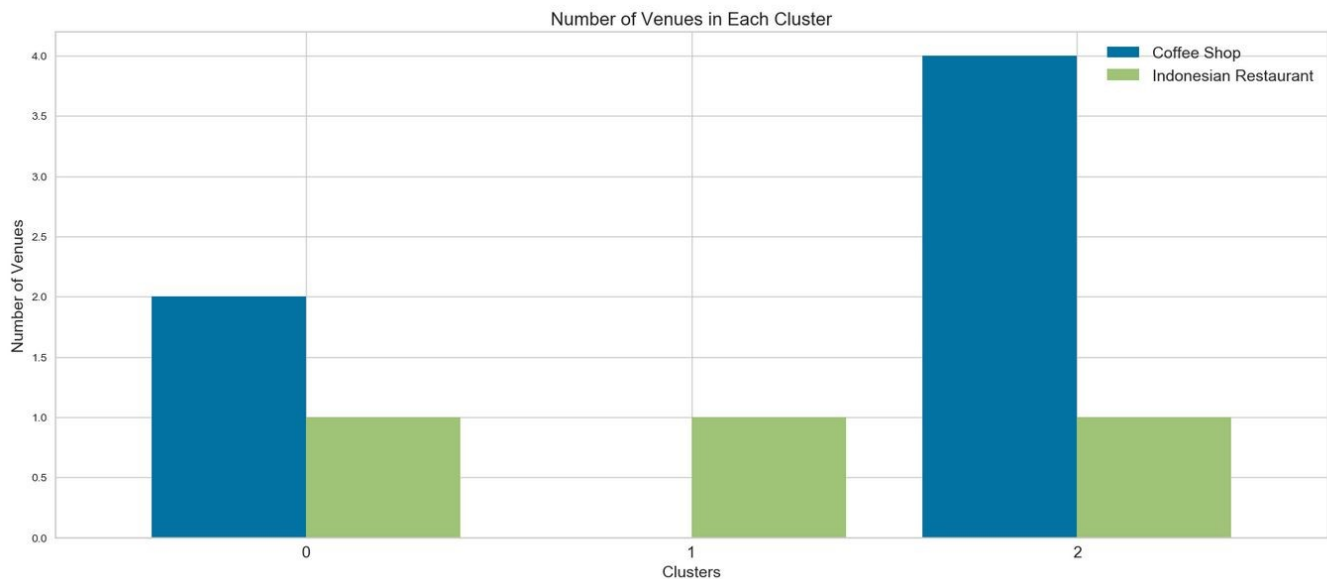
I use the elbow method to gain the proper k for clustering.



Based on the figure above, I run K-Means to cluster the cities into 3 clusters and create a new dataframe that includes the cluster as well as the top 10 venues for each neighborhood.

City	price (rp/m2)	latitude	longitude	Cluster Labels	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
Central Jakarta	26997450	-6.18233995	106.84287153600738	2	Coffee Shop	Hotel	Sushi Restaurant	Restaurant	Shopping Mall	Fast Food Restaurant	Clothing Store	Bakery
North Jakarta	25199166	-6.1362053	106.90069097114528	2	Coffee Shop	Indonesian Restaurant	Noodle House	Asian Restaurant	Chinese Restaurant	Shopping Mall	Multiplex	Beach
South Jakarta	22044101	-6.28381815	106.80486349194814	0	Coffee Shop	Hotel	Indonesian Restaurant	Bakery	Shabu-Shabu Restaurant	Restaurant	Steakhouse	Padangnese Restaurant
West Jakarta	17971174	-6.16156235	106.74389124027667	2	Coffee Shop	Seafood Restaurant	Indonesian Restaurant	Asian Restaurant	Korean Restaurant	Sushi Restaurant	Chinese Restaurant	Café
East Jakarta	13004566	-6.26289085	106.88222894692834	0	Coffee Shop	Hotel	Fast Food Restaurant	Restaurant	Indonesian Restaurant	Padangnese Restaurant	Bakery	Sushi Restaurant
Tangerang	10498210	-6.1734808	106.4992042	2	Indonesian Restaurant	Asian Restaurant	Convenience Store	Fast Food Restaurant	Coffee Shop	Furniture / Home Store	Pool	Japanese Restaurant
Bekasi	5953004	-6.2349858	106.9945444	0	Indonesian Restaurant	Coffee Shop	Fast Food Restaurant	Asian Restaurant	Restaurant	Multiplex	Seafood Restaurant	Shopping Mall
Depok	6584794	-6.4074657	106.8138131	1	Indonesian Restaurant	Coffee Shop	Food Truck	Fast Food Restaurant	Café	Bakery	Snack Place	Golf Course
Bogor	5811302	-6.5962986	106.7972421	2	Coffee Shop	Hotel	Indonesian Restaurant	Bakery	Café	Noodle House	Restaurant	Fast Food Restaurant

We can also estimate the number of **1st Most Common Venue** in each cluster. Thus, we can create a bar chart which may help us to find proper label names for each cluster.



From the result above, coffee shops are very common in Jabodetabek even though it does not always in the 1st place of the most common venue but it makes at least 2nd place on Jabodetabek (coffee shop in cluster 1 is in the 2nd place of the most common venue).

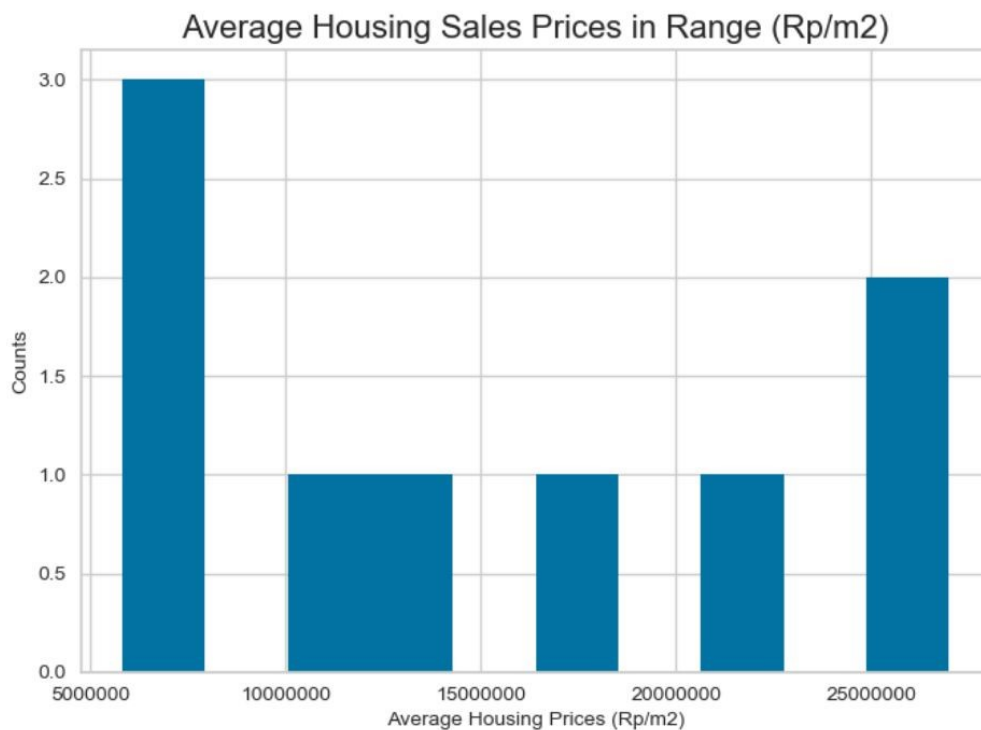
When we examine above graph we can label each cluster as follows:

- Cluster 0 : "Coffee Shop and Indonesian Restaurant Venues"
- Cluster 1 : "Indonesian Restaurant Venues"
- Cluster 2 : "Coffee Shop Venues"

Then I analyze the housing sales prices for per square meter in specific range. Thus we can create new labels which involve pricing features, as well.

	City	price (rp/m2)	Cluster Labels
0	Bogor	5811302	2
1	Bekasi	5953004	0
2	Depok	6584794	1
3	Tangerang	10498210	2
4	East Jakarta	13004566	0
5	West Jakarta	17971174	2
6	South Jakarta	22044101	0
7	North Jakarta	25199166	2
8	Central Jakarta	26997450	2

We can examine that what is the frequency of housing sales prices in different ranges. Thus, histogram can help to visualization



As it seems in above histogram, we can define the ranges as below:

- '<' 10,000,000 AHP : "Low Level HSP"
- 10,000,000 - 20,000,000 AHP : "Mid Level HSP"
- '>' 20,000,000 AHP : "High Level HSP"

Then I can add a label to my previous dataframe.

	City	price (rp/m2)	Cluster Labels	Level_labels
0	Bogor	5811302	2	Low Level HSP
1	Bekasi	5953004	0	Low Level HSP
2	Depok	6584794	1	Low Level HSP
3	Tangerang	10498210	2	Mid Level HSP
4	East Jakarta	13004566	0	Mid Level HSP
5	West Jakarta	17971174	2	Mid Level HSP
6	South Jakarta	22044101	0	High Level HSP
7	North Jakarta	25199166	2	High Level HSP
8	Central Jakarta	26997450	2	High Level HSP

One of my aim was also show the number of top 3 venues information for each city on the map. Thus, I grouped each city by the number of top 3 venues and I combined those informations in **Join** column.

	City	Join
0	Bekasi	8 Indonesian Restaurant, 7 Coffee Shop, 6 Fast...
1	Bogor	11 Coffee Shop, 8 Hotel, 8 Indonesian Restaurant
2	Central Jakarta	15 Coffee Shop, 15 Hotel, 4 Sushi Restaurant
3	Depok	11 Indonesian Restaurant, 7 Coffee Shop, 5 Café
4	East Jakarta	11 Coffee Shop, 9 Hotel, 6 Fast Food Restaurant
5	North Jakarta	13 Coffee Shop, 7 Indonesian Restaurant, 5 Asi...
6	South Jakarta	7 Coffee Shop, 6 Hotel, 5 Indonesian Restaurant
7	Tangerang	12 Indonesian Restaurant, 4 Asian Restaurant, ...
8	West Jakarta	13 Coffee Shop, 6 Seafood Restaurant, 5 Indone...

C. Results

Let's merge those new variables with related cluster informations in our main **Jabodetabek_merged** table.

	City	price (rp/m2)	latitude	longitude	Cluster Labels	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
0	Central Jakarta	26997450	-6.18233995	106.84287153600738	2	Coffee Shop	Hotel	Sushi Restaurant	Restaurant	Shopping Mall	Fast Food Restaurant	Clothing Store	Bak
1	North Jakarta	25199166	-6.1362053	106.90069097114528	2	Coffee Shop	Indonesian Restaurant	Noodle House	Asian Restaurant	Chinese Restaurant	Shopping Mall	Multiplex	Bee
2	South Jakarta	22044101	-6.28381815	106.80486349194814	0	Coffee Shop	Hotel	Indonesian Restaurant	Bakery	Shabu- Shabu Restaurant	Restaurant	Steakhouse	Padangne Restaur
3	West Jakarta	17971174	-6.16156235	106.74389124027667	2	Coffee Shop	Seafood Restaurant	Indonesian Restaurant	Asian Restaurant	Korean Restaurant	Sushi Restaurant	Chinese Restaurant	C
4	East Jakarta	13004566	-6.26289085	106.88222894692834	0	Coffee Shop	Hotel	Fast Food Restaurant	Restaurant	Indonesian Restaurant	Padangnese Restaurant	Bakery	Su Restaur
5	Tangerang	10498210	-6.1734808	106.4992042	2	Indonesian Restaurant	Asian Restaurant	Convenience Store	Fast Food Restaurant	Coffee Shop	Furniture / Home Store	Pool	Japan Restaur

You can now see **Join**, **Labels** and **Level_labels** columns as the last three ones in above table.

Finally, let's visualize the resulting clusters



D. Conclusion

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.

Depok is considered as one of the cities with the lowest average housing prices, and Depok is in a cluster of venues that are different from other cities in Jabodetabek. So there are still good opportunities for investors to invest in Depok by bringing trends from other Jabodetabek areas. As an example of investing in the coffee shop area in Depok, considering that the venue is most common in Jabodetabek.

However, this can be studied more deeply considering the limitations of the Foursquare API which is only 100 venues per city, so investors and stakeholders will get a bigger picture related to the potential of Jabodetabek.

Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.

E. References:

- [1] [Jabodetabek - Wikipedia \(https://en.wikipedia.org/wiki/Jabodetabek\)](https://en.wikipedia.org/wiki/Jabodetabek)
- [2] [Forsquare API \(https://developer.foursquare.com/\)](https://developer.foursquare.com/)
- [3] [Housing Sales Prices of Each City from "Hurriyet Retail Index for 2018" \(https://www.99.co/id/hpi\)](https://www.99.co/id/hpi)