

# **COVID-19 Cases & Venues Analysis of South-Jakarta**

M. Abdurrahman Shidiq

August 10, 2021

## **1. Introduction**

### **1.1. Problem Statement**

As a national capital of Indonesia, Jakarta considered as one of the most largest capital & busiest city in the world with over 10 million populations. Jakarta is the heart of economic & business activities of the Indonesian archipelago. With these characteristics, Jakarta has many destination locations, e.g. office centers, business centers, shopping centers, residential areas, public space etc.

During the COVID-19 pandemic, Jakarta has become the area with the most positive cases in Indonesia. The government requires to restrict or even close some places e.g. office, market, store, restaurant, etc. in order to reduce the transmission rate of the virus.

Restricting or even closing all the places, building, public places, etc. can suppress the infection of COVID-19 virus, but in other hand it will caused economic issue. The government should take the equilibrium between those two option. Before taking decision which venues should be restricted or closed, the questions are, Is there any relationship between type of venues/places and the number of positive case ?. if so, which type of venues/places should be restricted / closed ?

### **1.2. Scoop of Project**

This project will be focus on South-Jakarta City instead of examine & analyzing all the Jakarta Province.

### **1.3 Task Description**

Providing insight / information for Government or Decision maker to answer the business problem:

1.3.1 Clustering all venues in South-Jakarta into certain number of clusters then identifying and grouping regencies in South-Jakarta based on clustering results. Each regency will have certain characteristic cluster based on its venue and then visualize it on map

1.3.2 Visualizing Covid-19 cases on map

1.3.3 Describing insight from the map visualization to answer business problem

## **2. Data Acquisition**

### **2.1. Data Source**

2.1.1. Wikipedia, List of districts – regencies in South-Jakarta

2.1.2. Foursquare API, List data of venues in South-Jakarta

2.1.3. Jakarta Covid-19 data, COVID-19 cases in South-Jakarta in November 2020

2.1.4. GeoJSON South-Jakarta district border

## **3. Methodology**

### 3.1. Data Preprocessing

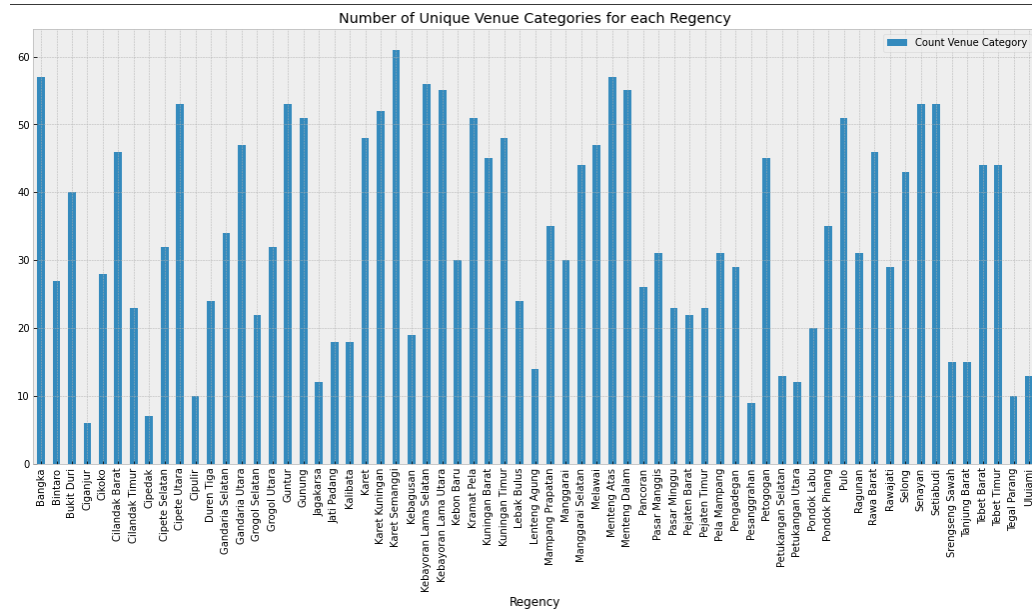
First thing for this step is getting the data. To get list of districts-regencies in South-Jakarta, I used beautiful soup package for web scrapping Wikipedia page. At this stage we'll get a data frame that contain information about districts-regencies in South-Jakarta. After that I utilize geopy package for getting coordinate location for each regency. The final product of this step is a data frame contain information about district name, regency name and its coordinate location.

	District	Regency	Latitude	Longitude
0	Cilandak	Cilandak Barat	-6.290266	106.796215
1	Cilandak	Cipete Selatan	-6.272540	106.805768
2	Cilandak	Gandaria Selatan	-6.271562	106.795172
3	Cilandak	Lebak Bulus	-6.302568	106.780354
4	Cilandak	Pondok Labu	-6.309144	106.797938
...	...	...	...	...
60	Tebet	Manggarai Selatan	-6.226016	106.858396
61	Tebet	Manggarai	-6.210134	106.850058
62	Tebet	Menteng Dalam	-6.232051	106.840997
63	Tebet	Tebet Barat	-6.226016	106.858396
64	Tebet	Tebet Timur	-6.226016	106.858396
65 rows x 4 columns				

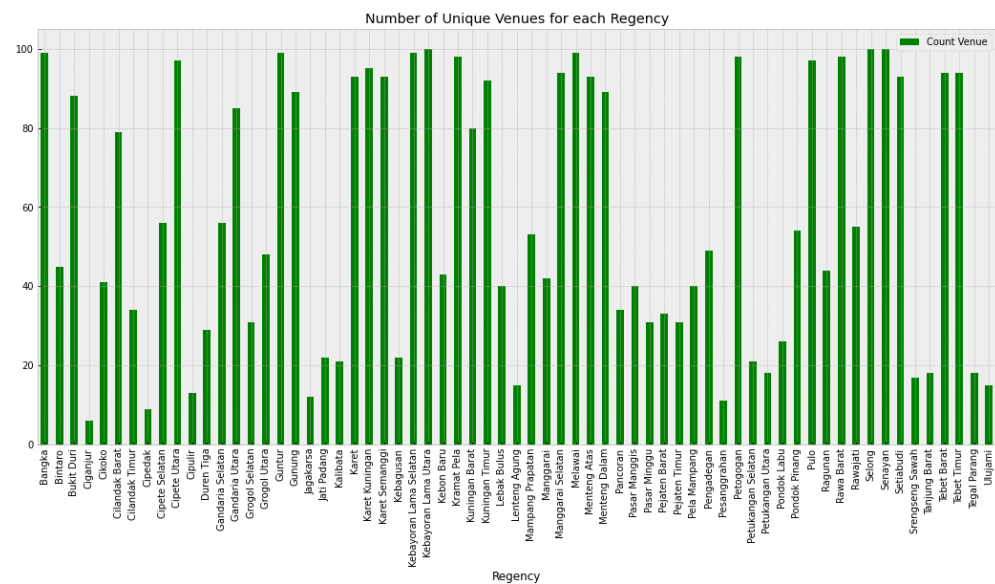
Pic 1 : South-Jakarta Dataframe

Visualizing each regency using folium package





Pic 4 : Number of venues for each regency



Pic 5 : Number of venue categories for each regency

The information we care about in this dataframe is Venue Category column, thus we will extract this information by using get dummies method. This data will be our base data for making clustering. The process will give us an output down below

	Regency	Acehnese Restaurant	Airport	American Restaurant	Arcade	Art Gallery	Art Museum	Books & Crafts Store	Entertainment	Arts & Restaurant	Asian Restaurant	Athletics & Sports	Australian Restaurant	Auto Dealership	Automotive Shop	BBQ Joint	Bakery	Balinese Restaurant	Bar	Basketball Court
0	Cilandak Barat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Cilandak Barat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
2	Cilandak Barat	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
3	Cilandak Barat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Cilandak Barat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3916	Tebet Timur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3917	Tebet Timur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3918	Tebet Timur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3919	Tebet Timur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3920	Tebet Timur	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

3921 rows x 231 columns

Pic 6 : On hot encoding venue categories

To summary the dataframe above, I created a new dataframe which shows list of 10 most common venue categories for each regencies.

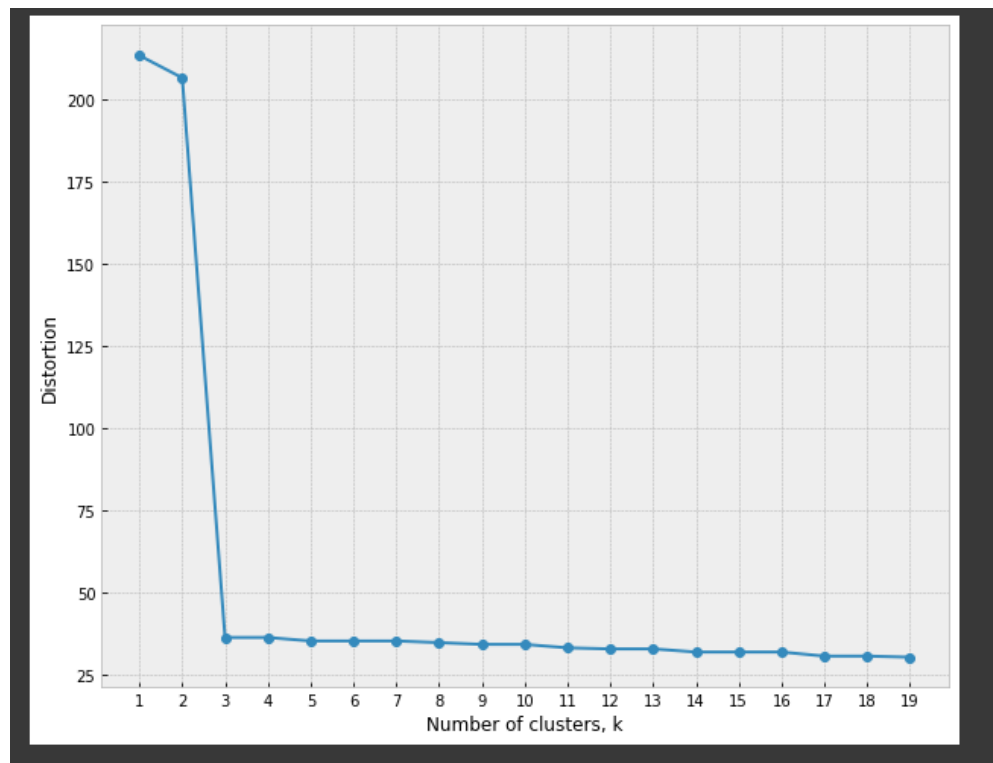
	Regency	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Bangka	Coffee Shop	Restaurant	Pizza Place	American Restaurant	Italian Restaurant	Café	Chinese Restaurant	Plaza	Playground	Bistro
1	Bintaro	Coffee Shop	Convenience Store	Indonesian Restaurant	Restaurant	Food Truck	Fast Food Restaurant	Gym	Salon / Barbershop	Asian Restaurant	Clothing Store
2	Bukit Duri	Indonesian Restaurant	Convenience Store	Asian Restaurant	Cafe	Coffee Shop	Bakery	Karaoke Bar	Indonesian Meatball Place	Restaurant	Boutique
3	Ciganjur	Arcade	Convenience Store	Soccer Stadium	Soccer Field	Café	Park	Yoga Studio	Flea Market	Fruit & Vegetable Store	Frozen Yogurt Shop
4	Cikoko	Convenience Store	Asian Restaurant	Coffee Shop	Supermarket	Park	Fast Food Restaurant	Café	Food Truck	Noodle House	Salon / Barbershop
...	...	...	...	...	...	...	...	...	...	...	...
60	Tanjong Barat	Coffee Shop	Javanese Restaurant	Indonesian Restaurant	Fast Food Restaurant	Gourmet Shop	Japanese Restaurant	Playground	Donut Shop	Food Court	Noodle House
61	Tebet Barat	Indonesian Restaurant	Convenience Store	Coffee Shop	Asian Restaurant	Café	Bakery	Restaurant	Seafood Restaurant	Karaoke Bar	Art Gallery
62	Tebet Timur	Indonesian Restaurant	Convenience Store	Coffee Shop	Asian Restaurant	Café	Bakery	Restaurant	Seafood Restaurant	Karaoke Bar	Art Gallery
63	Tegal Parang	Asian Restaurant	Food Truck	Fast Food Restaurant	Coffee Shop	Bakery	Vape Store	Middle Eastern Restaurant	Salon / Barbershop	Indonesian Restaurant	Market
64	Uluwatu	Department Store	Convenience Store	Music Venue	Burger Joint	Noodle House	Café	Clothing Store	Hobby Shop	Electronics Store	Sporting Goods Shop

65 rows x 11 columns

Pic 7 : On hot encoding venue categories

### 3.2 Modelling

For Clustering venue, I used unsupervised learning K-Means algorithm to cluster the regency based on its venue categories. Before feed the data into K-Means algorithm, we need to determine the number of cluster we want to create. With the help of elbow method the optimum number of cluster for this data is 3 cluster.



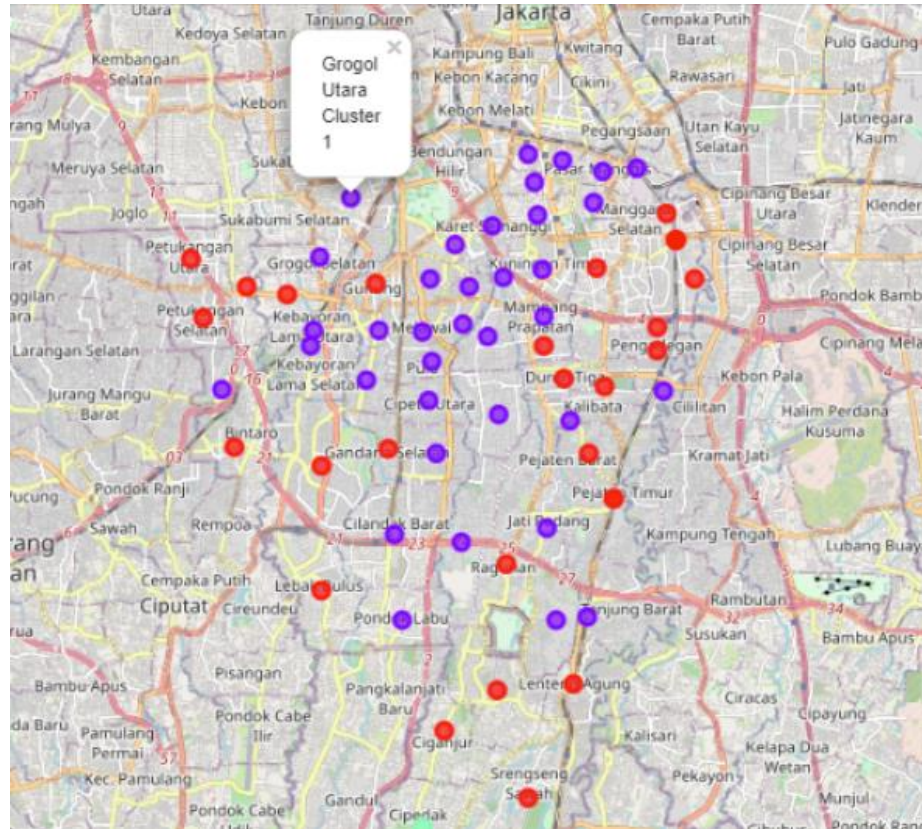
Pic 8 : Elbow method

District	Regency	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	9th Most Common Venue	10th Most Common Venue
0	Cilandak Barat	-6.290266	106.796215	1	Indonesian Restaurant	Coffee Shop	Asian Restaurant	Café	Convenience Store	Gym	Food Truck	Chinese Restaurant	Soccer Stadium	Steakhouse
1	Cilandak Cipte Selatan	-6.272840	106.806768	1	Coffee Shop	Food Truck	Café	Indonesian Restaurant	Convenience Store	Noodle House	Motorcycle Shop	Bakery	Padangnese Restaurant	Restaurant
2	Cilandak Gandaria Selatan	-6.271662	106.795172	0	Food Truck	Convenience Store	Indonesian Restaurant	Coffee Shop	Bakery	Seafood Restaurant	Furniture / Home Store	Motorcycle Shop	Fast Food Restaurant	Music Venue
3	Cilandak Lebak Bulus	-6.302568	106.780354	0	Asian Restaurant	Indonesian Restaurant	Convenience Store	Restaurant	Seafood Restaurant	Café	Soup Place	Paper / Office Supplies Store	Coffee Shop	Noodle House
4	Cilandak Pondok Labu	-6.309144	106.797938	1	Coffee Shop	Grocery Store	Indonesian Restaurant	Movie Theater	Café	Dumpling Restaurant	Burger Joint	Shopping Mall	Department Store	Diner
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
60	Tebet Manggarai Selatan	-6.226016	106.858396	0	Indonesian Restaurant	Convenience Store	Coffee Shop	Asian Restaurant	Café	Bakery	Restaurant	Seafood Restaurant	Karaoke Bar	Art Gallery
61	Tebet Manggarai	-6.216134	106.850068	1	Coffee Shop	Fast Food Restaurant	Food Truck	Park	Indonesian Restaurant	Pizza Place	Arcade	Asian Restaurant	Music Venue	Bookstore
62	Tebet Menteng Dalam	-6.232051	106.840997	0	Convenience Store	Indonesian Restaurant	Coffee Shop	Café	Hotel	Noodle House	Indonesian Meatball Place	Restaurant	Japanese Restaurant	Sushi Restaurant
63	Tebet Tebet Barat	-6.226016	106.858396	0	Indonesian Restaurant	Convenience Store	Coffee Shop	Asian Restaurant	Café	Bakery	Restaurant	Seafood Restaurant	Karaoke Bar	Art Gallery
64	Tebet Tebet Timur	-6.226016	106.858396	0	Indonesian Restaurant	Convenience Store	Coffee Shop	Asian Restaurant	Café	Bakery	Restaurant	Seafood Restaurant	Karaoke Bar	Art Gallery

Pic 9 : Cluster dataframe

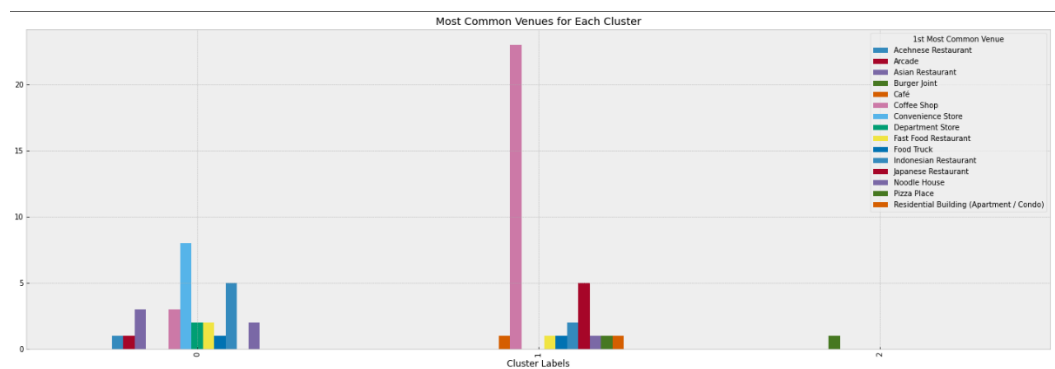
The dataframe above shows us that each regency will signed into specific cluster either Cluster 0, 1 or 2 based on its venue lists. Now lets analyze characteristic for each cluster.





Pic 10 : Regency with its cluster label

Since all regencies already assign into certain cluster (0,1,2), the next step is to examine the characteristic of each cluster and then assign cluster 0,1,2 into popper label name. We can estimate label names by referring into the most common venues for each cluster. To make it easy to understand, lets visualize 1<sup>st</sup> most common venue for each cluster.



Pic 11 : Common venues for each cluster label

Now we can assign the labels into clusters:

- Cluster 0 : Shopping Center
- Cluster 1 : Culinary Center & Residential

- Cluster 2 : Borderland

Before going into final result, the table down below is the table summary for clustering step. The dataframe contain information about most common venue & cluster labels for each regency.

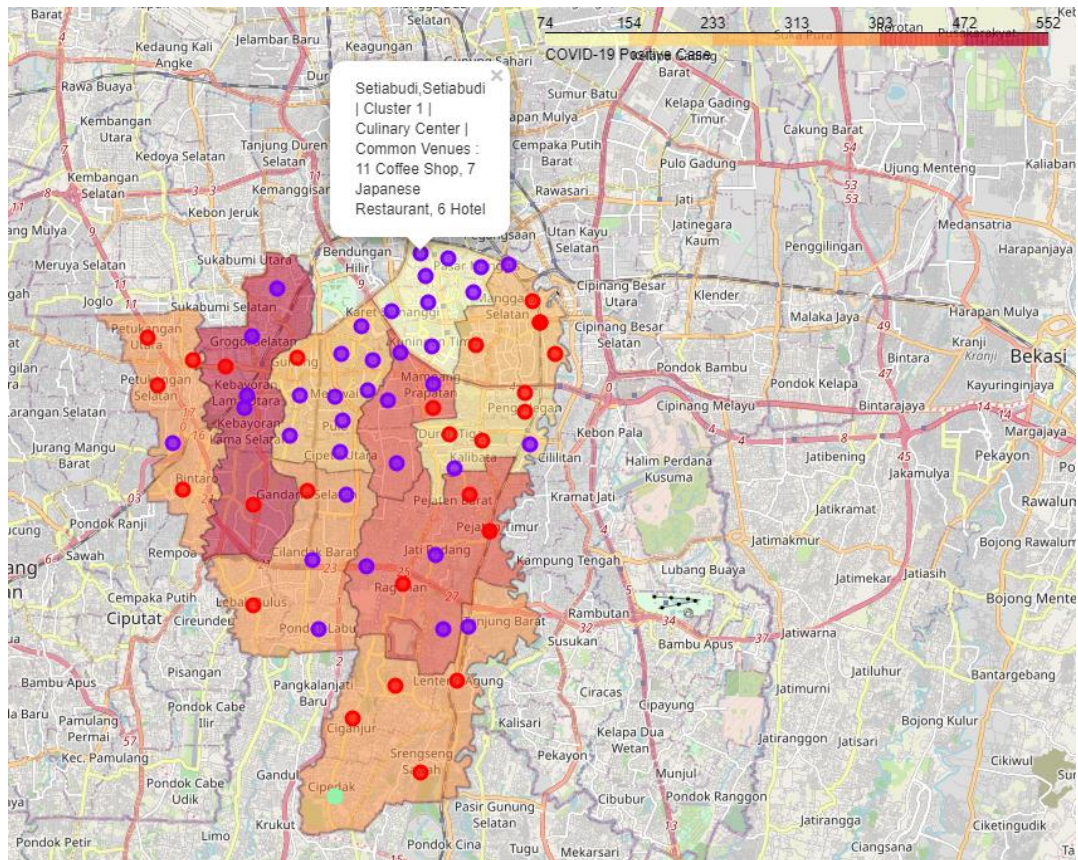
	Regency	Common Venue	District	Latitude	Longitude	Cluster Labels	Label
0	Bangka	14 Coffee Shop, 6 Restaurant, 5 Pizza Place	Mampang Prapatan	-6.264007	106.819292	1	Culinary Center
1	Bintaro	5 Coffee Shop, 5 Convenience Store, 5 Indonesi...	Pesanggrahan	-6.271393	106.760862	0	Shopping Center
2	Bukit Duri	15 Indonesian Restaurant, 9 Convenience Store,...	Tebet	-6.220104	106.856599	0	Shopping Center
3	Ciganjur	1 Arcade, 1 Café, 1 Convenience Store	Jagakarsa	-6.333430	106.807537	0	Shopping Center
4	Cikoko	7 Convenience Store, 3 Asian Restaurant, 3 Cof...	Pancoran	-6.245102	106.854456	0	Shopping Center
...	...	...	...	...	...	...	...
60	Tanjung Barat	3 Coffee Shop, 2 Fast Food Restaurant, 2 Indon...	Jagakarsa	-6.308478	106.838856	1	Culinary Center
61	Tebet Barat	17 Indonesian Restaurant, 8 Convenience Store,...	Tebet	-6.226016	106.858396	0	Shopping Center
62	Tebet Timur	17 Indonesian Restaurant, 8 Convenience Store,...	Tebet	-6.226016	106.858396	0	Shopping Center
63	Tegal Parang	3 Asian Restaurant, 3 Fast Food Restaurant, 3 ...	Mampang Prapatan	-6.249144	106.829488	0	Shopping Center
64	Ulujami	2 Convenience Store, 2 Department Store, 1 Asi...	Pesanggrahan	-6.236150	106.763881	0	Shopping Center

65 rows x 7 columns

Pic 12 : Summary dataframe

Last step is visualizing map South-Jakarta map using choropleth. The map will show us number of COVID-19 cases for each District and also the cluster label that already assign for each regency. Before we need to import COVID-19 cases data in Jakarta and GeoJSON data for South-Jakarta map. Now here it is the final result





Pic 13 : South-Jakarta map with COVID-19 case & Cluster

Based on the map above we got Setiabudi regency which located in Setiabudi District is identified as a Cluster Culinary Center. This particular area has relatively low number of Positive Case of COVID-19.

#### 4. Discussion

In this section we will attempt to answer the business problem based on the analysis & result we already got. Remember that in order to control the number of COVID-19 cases, the government should restrict the public places but the consequence, the economic activity will be suppressed. So which place should be restricted ?

##### 4.1. Is there any relationship between type of venues/places and the number of positive case?

On the visualization result we cannot be very sure if there is a relationship on COVID-19 cases and the Cluster area. For examples regencies with cluster Culinary Center (purple) have randomly COVID-19 cases from low (territory with light yellow color) until high case (territory with dark red color). If we have more data such as number of visitors per venue for particular range of time, may be we will have better chance to find relationship between cluster area and COVID-19 cases.

#### 4.2. Which type of venues / places should be restricted / closed

Because we cannot find relationship between number of COVID-19 case and Cluster area, it will be difficult to decide which area should be restricted or not.

#### 5. Conclusion

Unfortunately we cannot answer the business problem based on our analysis due to lack of data. Because we cannot found any pattern or relationship between COVID-19 cases and area cluster based on result in picture 13. For further analysis on the future, if we can get number of visitors per venue may be it will get better result to answer the business problem.

#### 6. Data Source

- [https://id.wikipedia.org/wiki/Daftar\\_kecamatan\\_dan\\_kelurahan\\_di\\_Kota\\_Administrasi\\_Jakarta\\_Selatan](https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Administrasi_Jakarta_Selatan)
- <https://developer.foursquare.com/docs/api-reference/venues/explore/>
- <https://data.jakarta.go.id/dataset/rekap-data-harian-covid-19-per-kelurahan-provinsi-dki-jakarta-bulan-november-2020>
- <https://github.com/thetrisatria/geojson-indonesia/blob/master/city-regency/id-jk-jaksel.geojson>