Coursera Capstone Project – IBM Data Science Certification

Sofyan Assauri

2020, May

Report Content

- **Introduction** – The business problem to be solved and who may be interested

- **Data** – Data requirements and source needed to solve the business problem

- **Methodology** – Data processing, explanatory data analysis, statistical testing performed

- **Result** – The result and finding of answer

- **Discussion** – Discussion of observation noted and any recommendation

- **Conclusion** – Answer choosen and conclusions

Introduction

I. Scenario and 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.

II. Problem to be resolved

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.

III. Interested Audience

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.

Data

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

- I used **Forsquare 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.
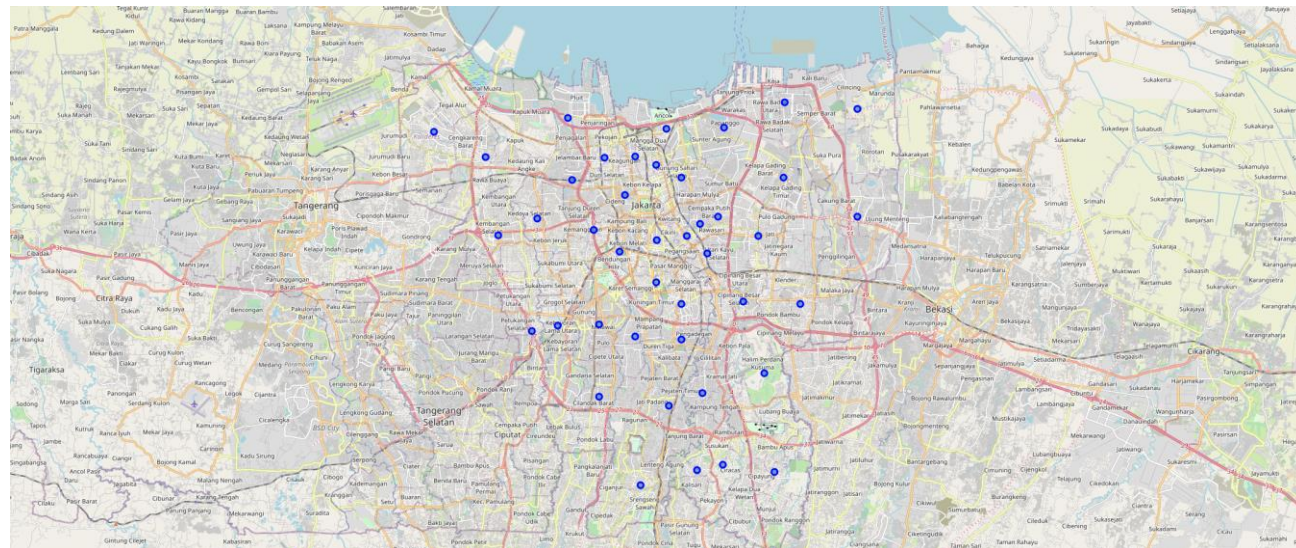
Methodology 1/6

- My master data which has the main components *Borough, Neighborhood, ZIP, Average Land Price, Latitude* and *Longitude* informations of the city.

	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

- Python **folium** library 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 to get the visual as below:

Methodology 2/6

- 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	lng
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

- 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.

Methodology 3/6

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

	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, Batuampar, Cawang, Cililitan, Duk...	Restaurant	Coffee Shop	Middle Eastern Restaurant	Department Store	Mediterranean Restaurant	Bakery	Pizza Place	Food Truck	Gym	Food Court
3	Bali Mester, Bidaracina, Cipinang Besar Selata...	Coffee Shop	Asian Restaurant	Food Truck	Indonesian Restaurant	Middle Eastern Restaurant	Café	Soccer Stadium	Mosque	Food Court	Shopping Mall
4	Bambu Apus, Ceger, Cilangkap, 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, Gedong, Kalisari, Pekayon	Indonesian Restaurant	Food Truck	Grocery Store	Basketball Court	Soccer Stadium	Convenience Store	Pizza Place	Stadium	Bakery	Food Court
7	Bendungan Hilir, Gelora, Kampung Bali, Karet T...	Indonesian Restaurant	Coffee Shop	Pizza Place	Convenience Store	Japanese Restaurant	Food Truck	Restaurant	Lounge	Soup Place	Pool
8	Bintaro, Pesanggrahan, Petukangan Selatan, Pet...	Noodle House	Indonesian Restaurant	Convenience Store	Pizza Place	Flea 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	Café	Seafood Restaurant	Convenience Store	Noodle House	Restaurant	Steakhouse	Park	Asian Restaurant

- 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.

Methodology 4/6

- Here is my merged table with cluster labels for each borough.

	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	Cengkareng	Cengkareng Barat, Cengkareng Timur, Duri Kosam...	11730, 11750, 11720, 11710, 11740	-6.1487	106.7353	7,734	1	Department Store	Pet Store	Movie Theater	Music Venue	Café	Farmers Market	Fast Food Restaurant	Donut Shop	Dumpling Restaurant	Food
1	Grogol Petamburan	Grogol, Jelambar, Jelambar Baru, Tanjung Duren...	11450, 11460, 11470, 11440	-6.1617	106.7846	14,706	1	Noodle House	Chinese Restaurant	Indonesian Restaurant	Asian Restaurant	Food Truck	Hotel	Coffee Shop	Café	Fast Food Restaurant	Korean Restaurant
2	Kalideres	Kalideres, Kamal, Pegadungan, Semanan, Tegal Alur	11840, 11810, 11830, 11850, 11820	-6.1343	106.7058	6,809	0	Noodle House	Asian Restaurant	Fast Food Restaurant	Pizza Place	Convenience Store	Food & Drink Shop	Café	Supermarket	Fried Chicken Joint	Food Court
3	Kebon Jeruk	Duri Kepa, Kebon Jeruk, Kedoya Selatan, Kedoya...	11510, 11530, 11520, 11550, 11560, 11540	-6.1835	106.7647	9,922	1	Food Truck	Indonesian Restaurant	Asian Restaurant	Convenience Store	Fast Food Restaurant	Restaurant	Coffee Shop	Noodle House	Food Court	Salon / Barbershop
4	Kembangan	Joglo, Kembangan Selatan, Kembangan Utara, Mer...	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	Café	Bubble Tea Shop	Food Court	Supermarket

- We can also estimate the number of **1st Most Common Venue** in each cluster by understand the Borough through categorization.

Cluster 0

	Borough	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
2	Kalideres	6,809	0	Noodle House	Asian Restaurant	Fast Food Restaurant	Pizza Place	Convenience Store	Food & Drink Shop	Café	Supermarket	Fried Chicken Joint	Food Court
8	Cempaka Putih	12,871	0	Pizza Place	Indonesian Restaurant	Indonesian Meatball Place	Café	Coffee Shop	Japanese Restaurant	Seafood Restaurant	Fast Food Restaurant	Asian Restaurant	Restaurant
10	Johar Baru	10,581	0	Convenience Store	Asian Restaurant	Seafood Restaurant	Restaurant	Indonesian Restaurant	Hospital	Gastropub	Noodle House	Nightclub	Stadium
13	Sawah Besar	19,069	0	Noodle House	Chinese Restaurant	Hotel	Indonesian Restaurant	Karaoke Bar	Asian Restaurant	Restaurant	Convenience Store	Hotel Bar	Dessert Shop
14	Senen	16,919	0	Indonesian Restaurant	Hotel	Coffee Shop	Fast Food Restaurant	Asian Restaurant	Convenience Store	Pharmacy	Food Truck	Javanese Restaurant	Café
15	Tanah Abang	41,861	0	Indonesian Restaurant	Coffee Shop	Pizza Place	Convenience Store	Japanese Restaurant	Food Truck	Restaurant	Lounge	Soup Place	Pool
17	Jagakarsa	6,862	0	Indonesian Restaurant	Asian Restaurant	Food Court	Other Great Outdoors	BBQ Joint	Noodle House	Pharmacy	Convenience Store	Department Store	Arcade
23	Pesanggrahan	7,539	0	Noodle House	Indonesian Restaurant	Convenience Store	Pizza Place	Flea Market	Wings Joint	Fast Food Restaurant	Fried Chicken Joint	French Restaurant	Food Truck

Cluster 1

	Borough	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	Cengkareng	7,734	1	Department Store	Pet Store	Movie Theater	Music Venue	Café	Farmers Market	Fast Food Restaurant	Donut Shop	Dumpling Restaurant	Food
1	Grogol Petamburan	14,706	1	Noodle House	Chinese Restaurant	Indonesian Restaurant	Asian Restaurant	Food Truck	Hotel	Coffee Shop	Café	Fast Food Restaurant	Korean Restaurant
3	Kebon Jeruk	9,922	1	Food Truck	Indonesian Restaurant	Asian Restaurant	Convenience Store	Fast Food Restaurant	Restaurant	Coffee Shop	Noodle House	Food Court	Salon / Barbershop
4	Kembangan	12,675	1	Asian Restaurant	Coffee Shop	Chinese Restaurant	Japanese Restaurant	Gym / Fitness Center	Noodle House	Café	Bubble Tea Shop	Food Court	Supermarket
5	Palmerah	19,605	1	Coffee Shop	Hotel	Pizza Place	Noodle House	Gym	Asian Restaurant	Convenience Store	Park	Indonesian Meatball Place	Basketball Court
6	Taman Sari	20,699	1	Chinese Restaurant	Noodle House	Asian Restaurant	Coffee Shop	Hotel	Nightclub	Fast Food Restaurant	Food Truck	Restaurant	Bakery

Methodology
5/6

Cluster 2

	Borough	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
36	Cilincing	4,703	2	Seafood Restaurant	Wings Joint	Farmers Market	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food Stand	Food Court	Food & Drink Shop

Cluster 3

	Borough	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
32	Makasar	6,878	3	Golf Course	Soccer Field	Restaurant	Arcade	Track	Wings Joint	Farmers Market	French Restaurant	Food Truck	Food Stand

Cluster 4

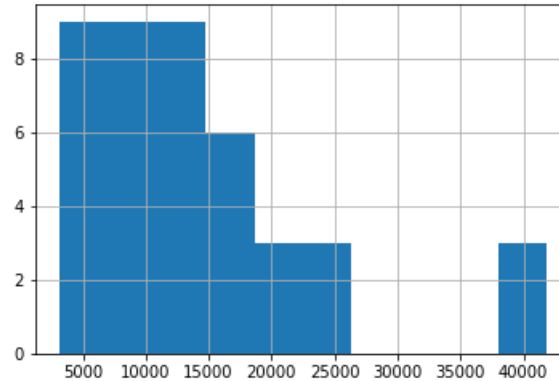|	Borough	Avg. Land Price	cluster	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
27	Cipayung	3,125	4	Garden	Pizza Place	Farmers Market	Frozen Yogurt Shop	Fried Chicken Joint	French Restaurant	Food Truck	Food Stand	Food Court	Food & Drink Shop

From the categotization above, then I rename the cluster based on common venues available for each Borough, as follows

	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	Cengkareng	Cengkareng Barat, Cengkareng Timur, Duri Kosam...	11730, 11750, 11720, 11710, 11740	-6.1487	106.7353	7,734	Coffee Shop Galore	Farmers Market	Movie Theater	Department Store	Music Venue	Donut Shop	Pet Store	Fast Food Restaurant	Café	Wings Joint	French Restaurant
1	Grogol Petamburan	Grogol, Jelambar, Jelambar Baru, Tanjung Duren...	11450, 11460, 11470, 11440	-6.1617	106.7846	14,706	Coffee Shop Galore	Noodle House	Chinese Restaurant	Indonesian Restaurant	Asian Restaurant	Coffee Shop	Hotel	Café	Fast Food Restaurant	Food Truck	Korean Restaurant
2	Kalideres	Kalideres, Kamal, Pegadungan, Semanan, Tegal Alur	11840, 11810, 11830, 11850, 11820	-6.1343	106.7058	6,809	Kitchen of Indonesia	Noodle House	Indonesian Restaurant	Asian Restaurant	Pizza Place	Food & Drink Shop	Fast Food Restaurant	Convenience Store	Chinese Restaurant	Food Court	Dumpling Restaurant
3	Kebon Jeruk	Duri Kepa, Kebon Jeruk, Kedoya Selatan, Kedoya...	11510, 11530, 11520, 11550, 11560, 11540	-6.1835	106.7647	9,922	Kitchen of Indonesia	Indonesian Restaurant	Food Truck	Noodle House	Fast Food Restaurant	Asian Restaurant	Convenience Store	Coffee Shop	Restaurant	Juice Bar	Cafeteria
4	Kembangan	Joglo, Kembangan Selatan, Kembangan Utara, Mer...	11640, 11610, 11650, 11620, 11630	-6.1930	106.7426	12,675	Coffee Shop Galore	Asian Restaurant	Coffee Shop	Chinese Restaurant	Japanese Restaurant	Noodle House	Gym / Fitness Center	Café	Supermarket	Bubble Tea Shop	Food Court
5	Palmerah	Jatipulo, Kemanggisan, Kota Bambu Selatan, Kot...	11430, 11480, 11420, 11410	-6.1901	106.7972	19,605	Coffee Shop Galore	Coffee Shop	Pizza Place	Hotel	Indonesian Meatball Place	Convenience Store	Noodle House	Asian Restaurant	Gym	Food Truck	Multiplex

Methodology 6/6

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

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

    7000 : "Low 1", 7000–14000 : "Low 2", 14000–21000 : "Mid 1", 21000–280000 : "Mid 2", 28000-35000 : "Mid 3", 35000-42000 : "High 1", 42000 > : "High 2"
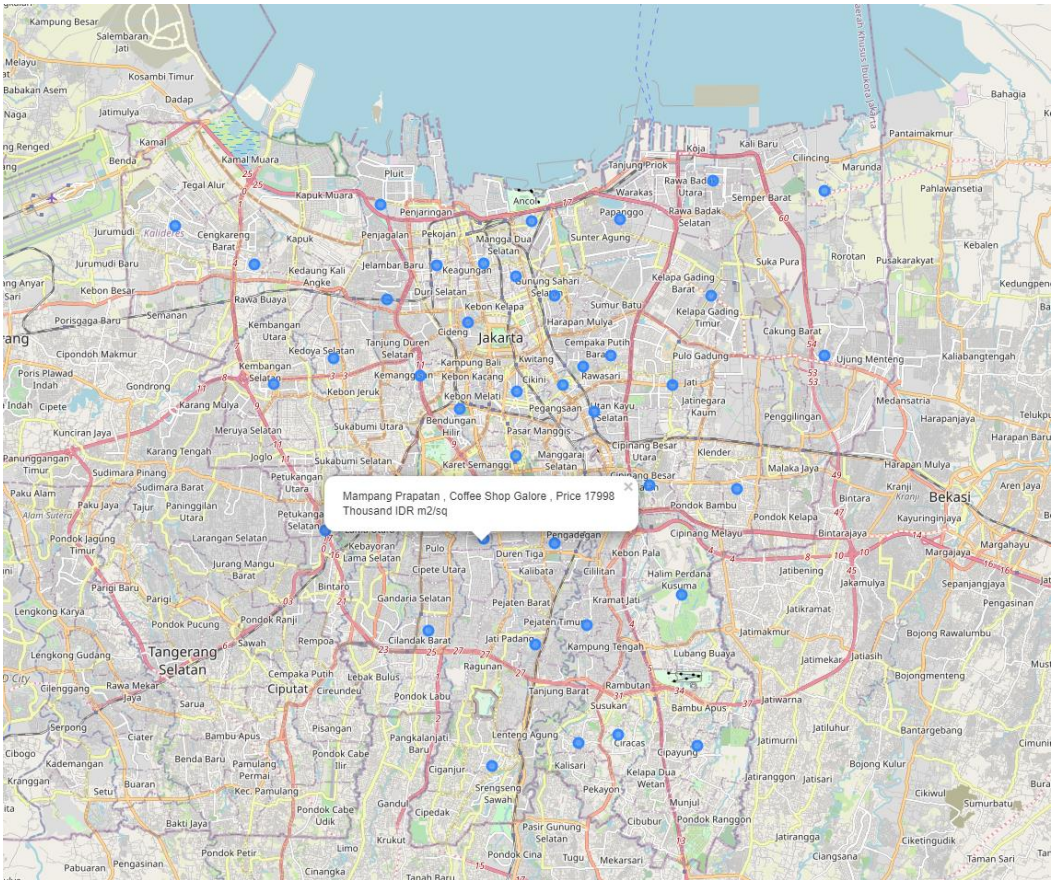
-

	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

Result

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

	Borough	Neighborhood	ZIP	Latitude	Longitude	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	Price m2/sq	Land Price Level
0	Cengkareng	Cengkareng Barat, Cengkareng Timur, Duri Kosam...	11730, 11750, 11720, 11710, 11740	-6.1487	106.7353	Coffee Shop Galore	Farmers Market	Movie Theater	Department Store	Music Venue	Donut Shop	Pet Store	Fast Food Restaurant	Café	Wings Joint	French Restaurant	7734	Low 2
1	Grogol Petamburan	Grogol, Jelambar, Jelambar Baru, Tanjung Duren...	11450, 11460, 11470, 11440	-6.1617	106.7846	Coffee Shop Galore	Noodle House	Chinese Restaurant	Indonesian Restaurant	Asian Restaurant	Coffee Shop	Hotel	Café	Fast Food Restaurant	Food Truck	Korean Restaurant	14706	Mid 1
2	Kalideres	Kalideres, Kamal, Pegadungan, Semanan, Tegal Alur	11840, 11810, 11830, 11850, 11820	-6.1343	106.7058	Kitchen of Indonesia	Noodle House	Indonesian Restaurant	Asian Restaurant	Pizza Place	Food & Drink Shop	Fast Food Restaurant	Convenience Store	Chinese Restaurant	Food Court	Dumpling Restaurant	6809	Low 1
3	Kebon Jeruk	Duri Kepa, Kebon Jeruk, Kedoya Selatan, Kedoya...	11510, 11530, 11520, 11550, 11560, 11540	-6.1835	106.7647	Kitchen of Indonesia	Indonesian Restaurant	Food Truck	Noodle House	Fast Food Restaurant	Asian Restaurant	Convenience Store	Coffee Shop	Restaurant	Juice Bar	Cafeteria	9922	Low 2
4	Kembangan	Joglo, Kembangan Selatan, Kembangan Utara, Mer...	11640, 11610, 11650, 11620, 11630	-6.1930	106.7426	Coffee Shop Galore	Asian Restaurant	Coffee Shop	Chinese Restaurant	Japanese Restaurant	Noodle House	Gym / Fitness Center	Café	Supermarket	Bubble Tea Shop	Food Court	12675	Low 2

- You can now see Join, Labels and Level_labels columns as the last three ones in above table. You can also see a clustered map boroughs of Istanbul in the below.

-

Discussions

- As I 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.

- I used the Kmeans algorithm as part of this clustering study, I set the optimum k value to 3. 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.

Discussions

- 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.