Finding a Good Location to Open Thailand Restaurant in Jakarta, Indonesia.

A final report for the course "Applied Data Science Capstone" given by IBM on Coursera

imamjabar

November, 2020

Contents

1 Problem's description

The city of Jakarta is well known to be a cosmopolitan city where you can find people from all around the world with over 11 million people live and it has a population density of over 16 thousand people per square kilometer. The city is divided into 44 districts in total (including Thousand Island districts).

An investor is looking to open a new **Thailand Restaurant** in Jakarta, but he is not sure about the best location for his new venue and needs input for making the decision. Although there are a lot of districts in Jakarta, their density between them is not uniform. Some districts are containing too many restaurants while there are less in some others.

If we have some knowledge about the population density, the housing price in each district coupling with an overview of the number of restaurants, we can have a better idea to set up a new business there. We expect to choose a place where the population density is high but fewer competitors. If the housing price in that place is low, it's more attractive to us.

The project aims to find a good location for a Thailand Restaurant in Jakarta. This will be determined by analyzing the number of restaurants, the population density, and the average housing price in each district.

2 Data Description

1. List of Jakarta City administrative units from **official Jakarta annual publication** and **OpenStreetMap mapping project**. It gives us a list of all urban districts of Jakarta with their area (in km^2), population (in 2020), and the density of each district (people/km^2). The list is given in `https://bit.ly/32NpEPQ`.

2. List of the coordinates (latitude, longitude) of 42 urban districts in Jakarta (without Seribu Island). This list can be generated based on the name of each district using **Nominatim** package. *geopy.geocoders.Nominatim.*

3. List of average housing prices per m^2 from **real estate marketplace web page** `https://www.lamudi.co.id/trends/`.

4. A modified *.json* file that contains all coordinates where we use it to create a choropleth map of the Housing Sales Price Index of Jakarta. From **Jakarta Geospatial Information site** `http://gis.bpbd.jakarta.go.id/layers/geonode%3Adki_kecamatan`.

5. **Foursquare API** to select the number of restaurants and their location in some neighborhoods of Jakarta `https://developer.foursquare.com/`.

3 Methodology

1. First, we need to collect all urban districts of Jakarta data from Jakarta annual publication to get **District Name**, **Area**, and **Population**. Also, collect **Average Housing Price** from the real estate marketplace. Then save it to *.csv* file.

2. The column **Density** is calculated later based on columns **Population** and **Area** of each district.

3. Throughout the project, we use **numpy** and **pandas** packages to manipulate data frames

4. We use *geopy.geocoders.Nominatim* package to get the coordinates of districts and add them to the main data frame.

5. We use **Foursquare API** to explore Thailand Restaurant venues in each district.

6. For clustering the "Thailand Restaurant" venues between districts, we use **K-Means Clustering** method and the package **scikit-learn** to implement the algorithm on our data. In order to indicate how many K for the method, we try with 10 different values of K from 1 to 10 and use the **elbow method** to choose the most appropriate one.

7. In order to visualize the charts, we use package **matplotlib** and **seaborn**.

8. We use the package **folium** to visualize the Jakarta map with its districts.

3.1 Data Preparation

I read the data that I collected from the Jakarta annual report and real estate marketplace that I store in GitHub repository, which contains *City name*, *District*, *Area*, *Population*, and *Average Housing Price*. Then I calculate the *Density* based on columns population and area of each district. I use Nominator to get the *Latitude* and *Longitude* of all districts.

	Cities	District	Area (km2)	Population 2020	Average Housing Price (1M IDR)	Density (pop/km2)	Latitude	Longitude
0	Jakarta Selatan	Cilandak	17.767	203573	22.4	11457.928	-6.289798	106.796926
1	Jakarta Selatan	Jagakarsa	25.459	413252	12.1	16232.059	-6.330101	106.822237
2	Jakarta Selatan	Kebayoran Baru	12.847	144140	62.5	11219.740	-6.243164	106.799850
3	Jakarta Selatan	Kebayoran Lama	19.503	309463	26.0	15867.456	-6.249128	106.777782
4	Jakarta Selatan	Mampang Prapatan	7.978	147909	22.1	18539.609	-6.250878	106.823021

Figure 1: The main data frame.

I used python Folium library to visualize geographic details of Jakarta and its districts and I created a map of Jakarta with districts superimposed on top. I used latitude and longitude values to get the visual as below:

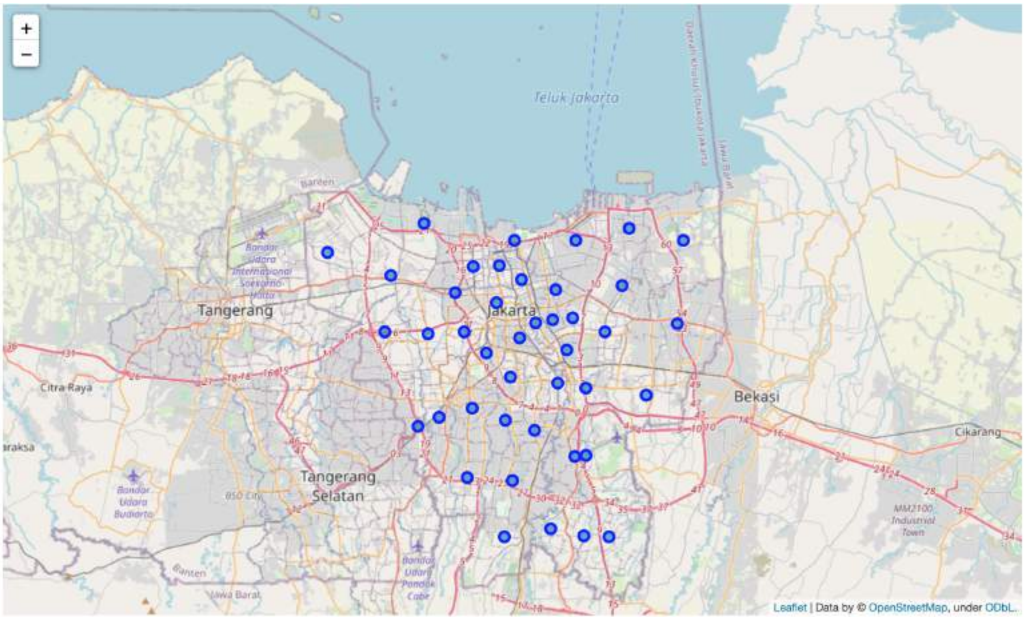

Figure 2: District of Jakarta

I utilized the Foursquare API to explore all **Thai Restaurants** venues in each District using the category parameter of Thai Restaurant from Foursquare. I designed the limit as 100 venue and the radius 5000 meter for each district from their given latitude and longitude information, and . Here is a head of the list *id*, *Venues name*, *latitude and longitude*, *distance* and *category* from Foursquare API.

	District	District Latitude	District Longitude	id	Venue	Venue Latitude	Venue Longitude	Distance	Venue Category
0	Cilandak	-6.289798	106.796926	4decd76145dd3993a8c7678d	CJ Tomyum	-6.297734	106.776382	2438	Thai Restaurant
1	Cilandak	-6.289798	106.796926	53d65133498efa2a3544ef6f	OTTERHOUND	-6.291680	106.800365	434	Thai Restaurant
2	Cilandak	-6.289798	106.796926	4db027b2f7b1bd003ac41f20	Opa Suki & Seafood	-6.266598	106.783293	2990	Thai Restaurant
3	Cilandak	-6.289798	106.796926	4bd67d3f7b1876b06ead8c86	Jittlada Thai Cuisine	-6.265335	106.782796	3140	Thai Restaurant
4	Cilandak	-6.289798	106.796926	4c46956b0f5aa59327527b76	White Elephant	-6.291671	106.800508	447	Thai Restaurant

Figure 3: List of Venues

2021 venues were returned by Foursquare, but the data contain other venues than Thai Restaurant categories also there are duplicated data.

	Venue Category	District	District Latitude	District Longitude	id	Venue	Venue Latitude	Venue Longitude	Distance
0	American Restaurant	17	17	17	17	17	17	17	17
1	Asian Restaurant	176	176	176	176	176	176	176	176
2	Bakery	6	6	6	6	6	6	6	6
3	Bar	7	7	7	7	7	7	7	7
4	Café	101	101	101	101	101	101	101	101
5	Chinese Restaurant	49	49	49	49	49	49	49	49
6	Coffee Shop	96	96	96	96	96	96	96	96
7	Convenience Store	4	4	4	4	4	4	4	4
8	Dim Sum Restaurant	18	18	18	18	18	18	18	18
9	Donut Shop	56	56	56	56	56	56	56	56
10	Eastern European Restaurant	24	24	24	24	24	24	24	24
11	Fast Food Restaurant	7	7	7	7	7	7	7	7
12	Food Court	71	71	71	71	71	71	71	71
13	Hotel	1	1	1	1	1	1	1	1
14	Indonesian Restaurant	24	24	24	24	24	24	24	24
15	Japanese Restaurant	33	33	33	33	33	33	33	33
16	Malay Restaurant	13	13	13	13	13	13	13	13
17	Martial Arts School	1	1	1	1	1	1	1	1
18	Nightclub	4	4	4	4	4	4	4	4
19	Noodle House	27	27	27	27	27	27	27	27
20	Photography Lab	3	3	3	3	3	3	3	3
21	Restaurant	45	45	45	45	45	45	45	45
22	Sandwich Place	15	15	15	15	15	15	15	15
23	Seafood Restaurant	10	10	10	10	10	10	10	10
24	Steakhouse	8	8	8	8	8	8	8	8
25	Sushi Restaurant	9	9	9	9	9	9	9	9
26	Thai Restaurant	1185	1185	1185	1185	1185	1185	1185	1185
27	Vietnamese Restaurant	11	11	11	11	11	11	11	11

Figure 4: List of Categories

After cleaning the data there are only **183** Thai Restaurants. From 42 Districts, only 27 districts that have at least 1 Thai Restaurant with the highest number is 21 Thai Restaurants.

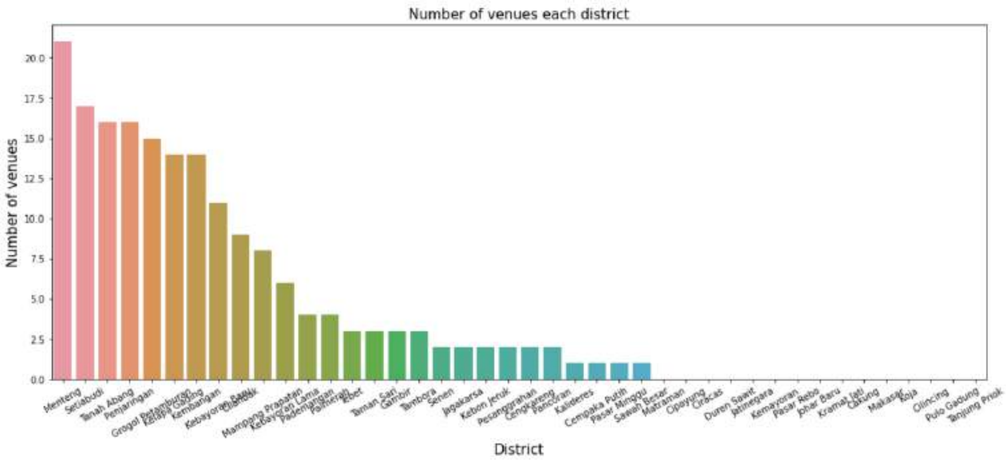

Figure 5: Number of Venue in District

3.2 Clustering

We want to cluster them into several groups based on **Average Housing Price**, **Density**, and **number of Venues**

	District	Average Housing Price (1M IDR)	Density (pop/km2)	Venue_Count
0	Cilandak	22.4	11457.928	9
1	Jagakarsa	12.1	16232.059	2
2	Kebayoran Baru	62.5	11219.740	11
3	Kebayoran Lama	26.0	15867.456	6
4	Mampang Prapatan	22.1	18539.609	8

Figure 6: Cluster Data

First, we need to determine the number of groups (or K for the K-means method). Using the elbow method with different values of K, Figure 7 shows that **4** is the best choice.

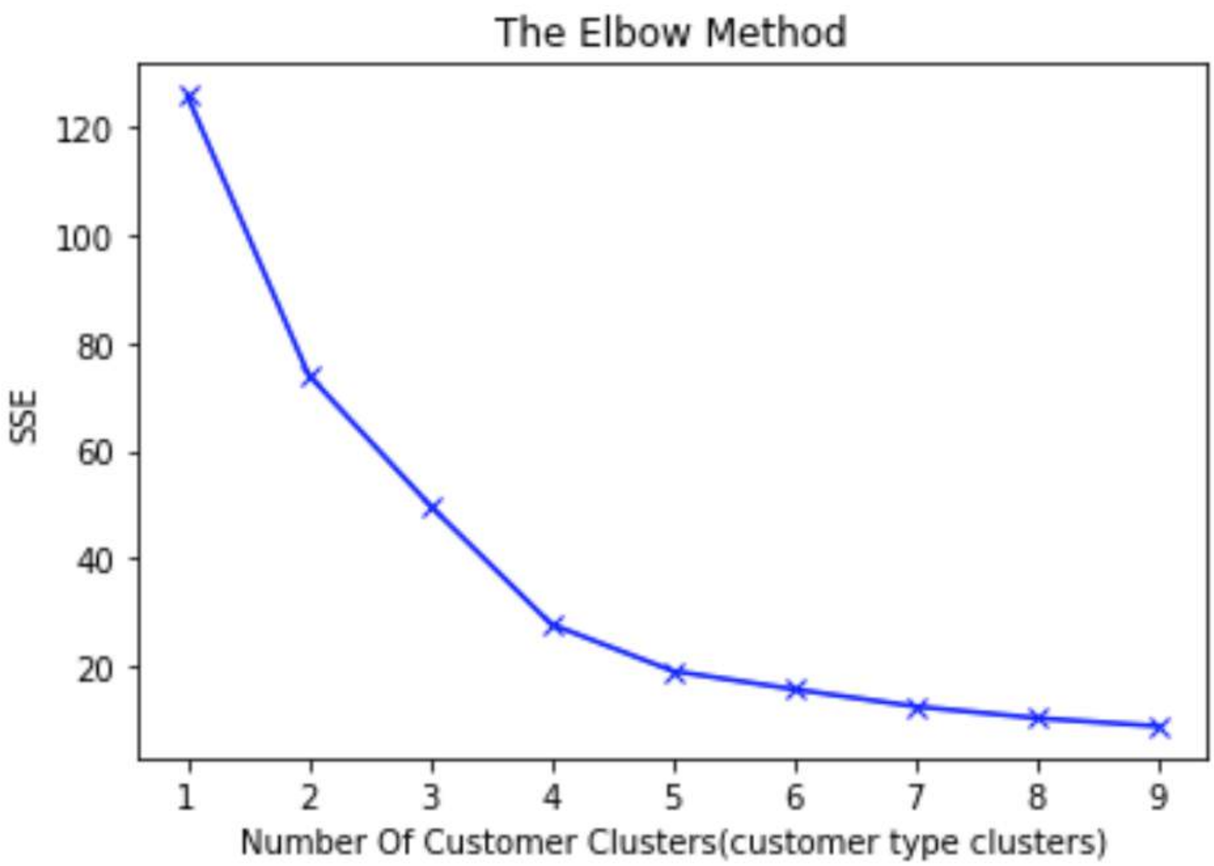

Figure 7: The optimal number of clusters

4 Results

Here is my merged table with cluster labels for each district.

	index	District	Venue_Count	Cluster Labels	Cities	Average Housing Price (1M IDR)	Density (pop/km2)	Latitude	Longitude	AHP Level
0	0	Cilandak	9	0	Jakarta Selatan	22.4	11457.928	-6.289798	106.796926	Medium
1	38	Kembangan	14	0	Jakarta Barat	22.1	13172.285	-6.191395	106.740586	Medium
2	21	Kelapa Gading	14	0	Jakarta Utara	20.0	9835.448	-6.159938	106.902483	Low
3	4	Mampang Prapatan	8	0	Jakarta Selatan	22.1	18539.609	-6.250878	106.823021	Medium
4	35	Grogol Petamburan	15	0	Jakarta Barat	16.1	22247.559	-6.164188	106.788317	Low
5	24	Penjaringan	16	0	Jakarta Utara	19.9	9771.436	-6.117265	106.767433	Low
6	33	Tanah Abang	16	0	Jakarta Pusat	26.5	14474.829	-6.205258	106.809500	Medium
7	8	Setiabudi	17	0	Jakarta Selatan	34.5	15897.419	-6.221706	106.826308	Medium
8	13	Duren Sawit	0	1	Jakarta Timur	13.3	17933.534	-6.234138	106.919247	Low
9	12	Ciracas	0	1	Jakarta Timur	12.5	16859.326	-6.329635	106.876604	Low
10	11	Cipayung	0	1	Jakarta Timur	12.3	10395.438	-6.330046	106.893782	Low
11	3	Kebayoran Lama	6	1	Jakarta Selatan	26.0	15867.456	-6.249128	106.777782	Medium
12	18	Pasar Rebo	0	1	Jakarta Timur	12.2	17290.126	-6.324973	106.853376	Low
13	6	Pasar Minggu	1	1	Jakarta Selatan	18.0	14169.928	-6.291950	106.827835	Low
14	9	Tebet	3	1	Jakarta Selatan	23.8	22392.476	-6.226016	106.858396	Medium
15	7	Pesanggrahan	2	1	Jakarta Selatan	15.9	16501.805	-6.255458	106.763112	Low
16	1	Jagakarsa	2	1	Jakarta Selatan	12.1	16232.059	-6.330101	106.822237	Low
17	5	Pancoran	2	1	Jakarta Selatan	24.1	17458.250	-6.258085	106.842733	Medium
18	10	Cakung	0	1	Jakarta Timur	13.5	13146.140	-6.185562	106.940109	Low
19	15	Kramat Jati	0	1	Jakarta Timur	12.7	22593.482	-6.275477	106.870376	Low
20	16	Makasar	0	1	Jakarta Timur	13.1	9501.463	-6.275002	106.877415	Low
21	20	Cilincing	0	1	Jakarta Utara	19.9	10144.721	-6.129015	106.944454	Low
22	23	Pademangan	4	1	Jakarta Utara	19.9	12553.598	-6.129052	106.828972	Low
23	26	Cempaka Putih	1	1	Jakarta Pusat	20.0	17970.841	-6.181214	106.868548	Low
24	32	Senen	2	1	Jakarta Pusat	26.5	22141.593	-6.184971	106.843235	Medium
25	31	Sawah Besar	1	1	Jakarta Pusat	26.5	18476.940	-6.155891	106.833580	Medium
26	40	Taman Sari	3	1	Jakarta Barat	16.1	24188.679	-6.146142	106.818499	Low
27	25	Tanjung Priok	0	1	Jakarta Utara	19.9	17156.921	-6.128858	106.870793	Low

Figure 8: Merged Table with clusters

We can label the clusters like these and put it to the main data frame:

- **Cluster 0** : There are a lot of Thai Restaurants in these districts, and the AHP price is medium.
- **Cluster 1** : There are not many Thai Restaurants in these districts, and the AHP price is low.
- **Cluster 2** : There are a lot of Thai Restaurants in these districts, and the AHP price is high.
- **Cluster 3** : The number of Thai Restaurants in these districts are low, the AHP price is low, but the density is high

Figure 9 illustrates the clusters of all districts in Jakarta. With this map, we can easily distinguish the clusters between districts.

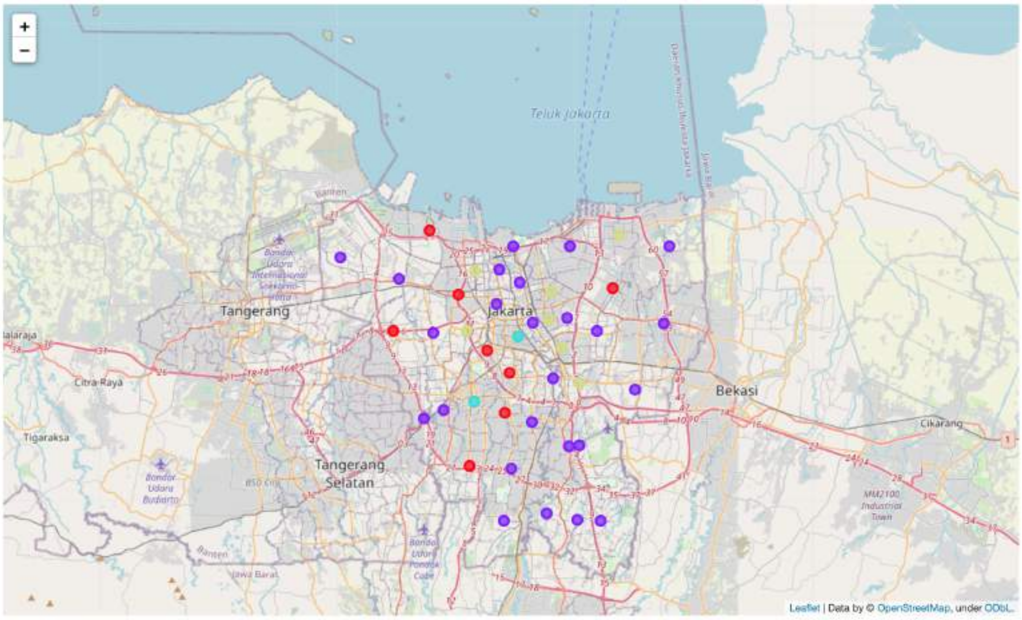

Figure 9: The maps of clusters

If we look back to the average housing price table (AHP), we can define them into 3 groups (unit: million IDR). Figure 10 indicates that the low price housing take the majority. We need to focus on the **Low** housing price to set up our business.

- **Low** : $10 < AHP \leq 20$.
- **Medium** : $20 < AHP \leq 50$.
- **High** : $50 \leq AHP$.

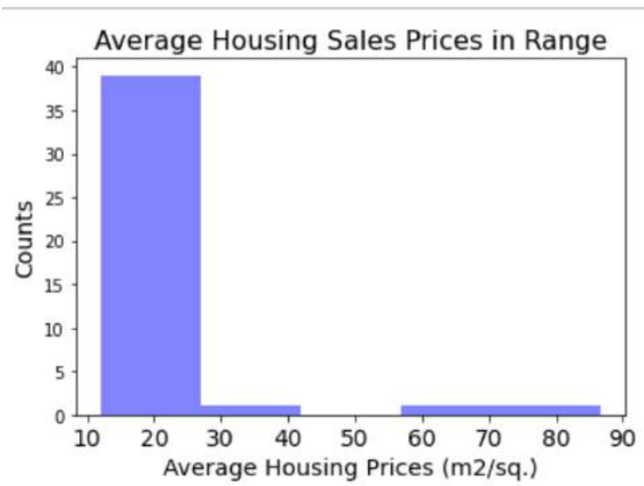

Figure 10: The distribution of AHP.

Figure 11 shows us the choropleth map of AHP.

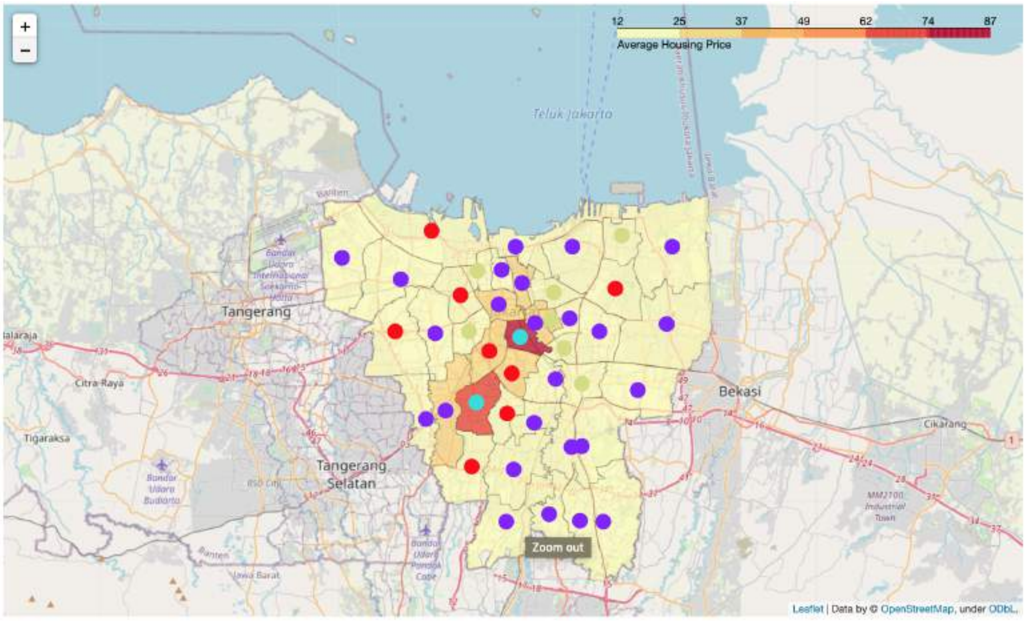

Figure 11: The couple maps of AHP and the clusters

We also can see the relationship between Density and clusters. Figure 12 gives us a full picture about the relation between population density and the clusters.

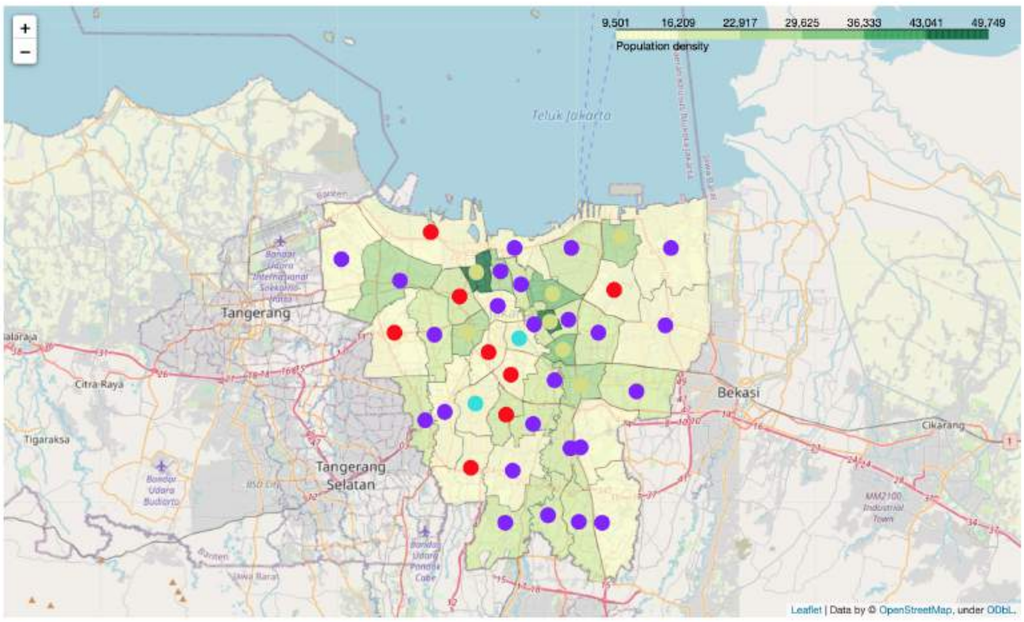

Figure 12: The couple maps of clusters and the population density of each district.

We focus on Cluster 3:

- Where there are **not many Thai Restaurants + Low AHP Price + High density**: **Matraman**, **Jatinegara**, **Palmerah**, **Tambora**, **Kemayoran**, **Koja**, and **Johar Baru**

	index	District	Venue_Count	Cluster Labels	Cities	Average Housing Price (1M IDR)	Density (pop/km2)	Latitude	Longitude	AHP Level	Labels
0	17	Matraman	0	3	Jakarta Timur	13.1	31035.773	-6.203624	106.864579	Low	Low number of Thai Restaurants, Low AHP Price,...
1	14	Jatinegara	0	3	Jakarta Timur	16.0	26394.624	-6.229147	106.877417	Low	Low number of Thai Restaurants, Low AHP Price,...
2	39	Palmerah	4	3	Jakarta Barat	16.1	27672.388	-6.191002	106.794363	Low	Low number of Thai Restaurants, Low AHP Price,...
3	41	Tambora	3	3	Jakarta Barat	16.1	44027.849	-6.146614	106.801046	Low	Low number of Thai Restaurants, Low AHP Price,...
4	29	Kemayoran	0	3	Jakarta Pusat	19.3	31597.270	-6.162546	106.856890	Low	Low number of Thai Restaurants, Low AHP Price,...
5	22	Koja	0	3	Jakarta Utara	19.9	27654.907	-6.120750	106.907362	Low	Low number of Thai Restaurants, Low AHP Price,...
6	28	Johar Baru	0	3	Jakarta Pusat	26.5	49748.756	-6.183125	106.855332	Medium	Low number of Thai Restaurants, Low AHP Price,...

Figure 13: Cluster 3

5 Conclusion

From all above results, we conclude that the best place for us to set up a new Thai Restaurant is in **cluster 3** because there are a lot of people living there (high density), there are not many already-working Thai Restaurant and the average housing price is low.

There are 7 districts in cluster 3, but if we have to choose one District, **Matraman** is the best district over cluster 3, because it has the lowest Average Housing Price and the density is high.