Location and Price Analysis for Housing in Indonesia

Jonathan Tanone

April 2021

2. Data acquisition and processing

2.1.Data Sources

I scraped the data for this analysis is procured through Lamudi[2]. There are several options of other property listings websites, but only Lamudi has the most location variety. To complement this data, I requested the location coordinates using Gecoders and the list of venues near the locations using the foursquare API[3]. I used Geojson data from this Github repository for the map [4].

2.2.Data processing

Through the scraping process from Lamudi, I acquired the location address and the price of the property. Below is the table preview.

	Location	Price	Area
0	JlCisarantenArcamanikCisarantenKulon,Bandung	Rp550.000.000	Bandung
1	TanahKusir,JakartaSelatan	Rp5.500.000.000	JakartaSelatan
2	Cilandak,JakartaSelatan	Rp3.900.000.000	JakartaSelatan
3	Indonesia,JakartaSelatan,JalanBangkaIIIBangka,...	Rp5.500.000.000	JakartaSelatan
4	Serpong,TangerangSelatan	Rp700.000.000	TangerangSelatan

Unfortunately, the location address format is very inconsistent. Some of the listings provide the detailed location, and some of them only display the city name. I decided to group the listings based on the city and average the listings price for that city.

After grouping them, I realized that there are cities that have only one or two listings. In my opinion, the number is not sufficient to create valuable information. I decided to drop the areas that have fewer than three listings. After appending and visualizing the data points using Folium, I noticed that a few of the coordinates are false. I manually corrected it. Below is the finished group table.

	Location	Listings	Latitude	Longitude	Mean_Price
0	Tangerang	402	-6.176031	106.638447	4.545191e+09
1	Jakarta Selatan	330	-6.283818	106.804863	9.785526e+09
2	Bekasi	259	-6.234986	106.994544	5.438207e+09
3	Bandung	240	-6.934469	107.604954	2.390222e+09
4	Bogor	229	-6.596299	106.797242	4.291715e+09
...	...	...	...	...	...
62	Ciamis	3	-7.326661	108.353095	3.006500e+10
63	Madiun	3	-7.611888	111.673193	6.186667e+08
64	Lamongan	3	-7.122912	112.328216	2.275000e+08
65	Cilegon	3	-6.017389	106.053769	3.258333e+08
66	Maros	3	-4.965502	119.692843	1.700000e+08

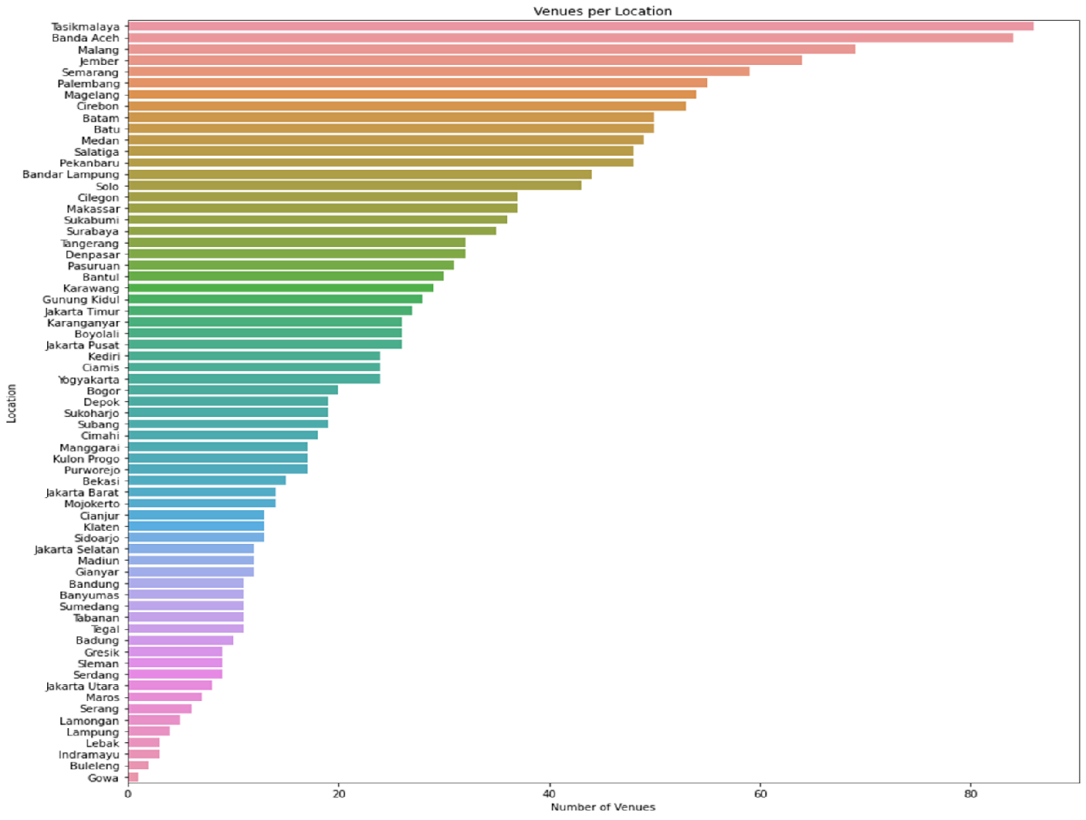
67 rows × 5 columns

The next step is requesting the coordinates using geocoders and the venues list using the foursquare API. For the venues, I plan to list 100 venues near the location coordinates. I collected only the category and the number of venues. The table above transformed into the table below.

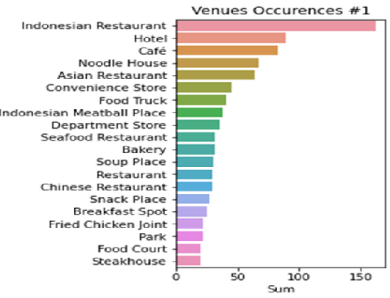
	Unnamed: 0	Unnamed: 0.1	key_0	Location	Listings	Latitude	Longitude	Mean_Price	Fast Food Restaurant	Coffee Shop	Café	Noodle House	Donut Shop	Indonesian Meatball Place	Bookstore	Sou Plac
0	0	0	Tangerang	Tangerang	402	-6.176031	106.638447	4.545191e+09	2.0	2.0	3.0	1.0	1.0	1.0	1.0	1
1	1	1	Jakarta Selatan	Jakarta Selatan	330	-6.283818	106.804863	9.785526e+09	0.0	1.0	1.0	1.0	0.0	0.0	0.0	1
2	2	2	Bekasi	Bekasi	259	-6.234986	106.994544	5.438207e+09	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0
3	3	3	Bandung	Bandung	240	-6.934469	107.604954	2.390222e+09	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0
4	4	4	Bogor	Bogor	229	-6.596299	106.797242	4.291715e+09	1.0	1.0	1.0	1.0	0.0	0.0	0.0	0

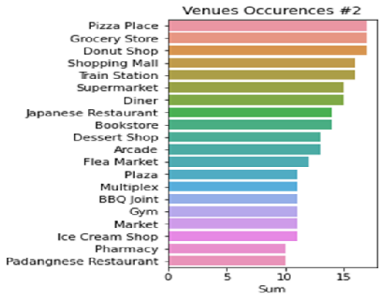
5 rows × 218 columns

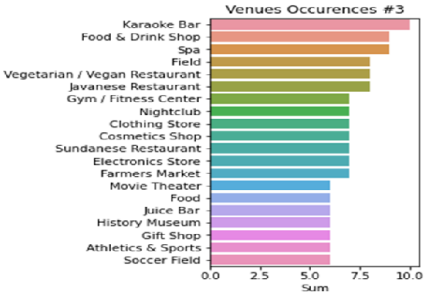
To make it easier to understand, I created a figure to visualize the table. In the figure below, the total number of venues per location the API requested is not uniform. The number ranges from two to ninety. I believe that clustering the data points first will yield the best information quality. Furthermore, there is a lot of venue categories collected in this data.

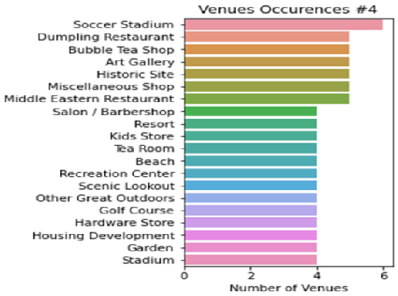


Ranking the venue occurrence will provide the most informative result. In the venue category occurrence graph below, I divided the data into bins with 20 members each. There are around 200 venue categories recorded, but venues in “Venue Occurrences #4" have small numbers. Thus, venues ranked lower than those in "Venue Occurrences #4" will not contribute much to the cluster profile. On the other hand, those in Figures #1 to #3 will contribute to the cluster profile significantly.


Venues Occurences #1
Indonesian Restaurant
Hotel
Café
Noodle House
Asian Restaurant
Convenience Store
Food Truck
Indonesian Meatball Place
Department Store
Seafood Restaurant
Bakery
Soup Place
Restaurant
Chinese Restaurant
Snack Place
Breakfast Spot
Fried Chicken Joint
Park
Food Court
Steakhouse
0
50
100
150
Sum


Venues Occurences #2
Pizza Place
Grocery Store
Donut Shop
Shopping Mall
Train Station
Supermarket
Diner
Japanese Restaurant
Bookstore
Dessert Shop
Arcade
Flea Market
Plaza
Multiplex
BBQ Joint
Gym
Market
Ice Cream Shop
Pharmacy
Padangnese Restaurant
0
5
10
15
Sum


Venues Occurences #3
Karaoke Bar
Food & Drink Shop
Spa
Field
Vegetarian / Vegan Restaurant
Javanese Restaurant
Gym / Fitness Center
Nightclub
Clothing Store
Cosmetics Shop
Sundanese Restaurant
Electronics Store
Farmers Market
Movie Theater
Food
Juice Bar
History Museum
Gift Shop
Athletics & Sports
Soccer Field
0.0
2.5
5.0
7.5
10.0
Sum


Venues Occurences #4
Soccer Stadium
Dumpling Restaurant
Bubble Tea Shop
Art Gallery
Historic Site
Miscellaneous Shop
Middle Eastern Restaurant
Salon / Barbershop
Resort
Kids Store
Tea Room
Beach
Recreation Center
Scenic Lookout
Other Great Outdoors
Golf Course
Hardware Store
Housing Development
Garden
Stadium
0
2
4
6
Number of Venues