

Data Analysis of Housing Prices, Prices for Lease and Neighborhoods in Ho Chi Minh City, Vietnam

By Pham Hoang Viet
May 2020

Contents

Introduction/ Business Problem	2
Data Description	2
Methodology	3
Average Housing Prices in Ho Chi Minh City	3
Map of Districts in Ho Chi Minh City	3
Calculating the Average Price for Leasing in Ho Chi Minh City	4
Analysis of Leasing demand in Ho Chi Minh City	4
Exploring Neighborhood in Ho Chi Minh City	6
Clustering Districts of Ho Chi Minh City	7
Result	9
Average Housing Price and Price for Lease of each District	9
Leasing Demand	10
4 Clusters of Districts	10
Discussion	11
Conclusion	12

Introduction/ Business Problem

Ho Chi Minh City, abbreviated as HCMC, is the most populous city in Vietnam with a population of 9 million as of 2019. Located in southeastern Vietnam, the metropolis surrounds the Saigon River and covers about 2,061 square kilometers (796 square miles).

Being the financial center of Vietnam, HCMC has the headquarters of many national and international banks and companies. As a major gateway to Vietnam, the city received over 8.6 million international visitors in 2019, of which 4.1 million were overnight visitors.

That is the reason why Ho Chi Minh city is the most developed city in Vietnam opening many opportunities for Small and Medium Enterprises (SMEs) to start a new business. Apart from online business, companies also need stores or shops for displaying and selling their products. Therefore, in current years, the real estate market in HCMC is hot with high prices and high demand.

As an investor or an owner of an enterprise, we expect to buy or rent a place with a low price and easy to approach the customers. Especially in the Covid-19 period, the price for lease is falling but what is the price offered by the owners? I will try to answer the question in this project by analyzing the HCMC neighborhood and the average prices for housing and lease in each district. The purpose is to give the SMEs an overview of the status of each area to help them have a better decision to choose a place depending on their budget.

HCMC consist of 24 districts including 19 urban districts and 5 rural districts. Because there is not much data for rural districts, this project will focus only on 19 urban districts of HCMC. We will create a map and information chart to illustrate the current situation of the real estate market in HCMC and cluster districts according to the venue density and housing and leasing price.

Data Description

To analyze the Business Problem above, I will collect data from sources as below:

- Data of Housing Average Price of each District per square meter scraped from a real estate broker website (<https://mogi.vn/gia-nha-dat>). It has 2 columns: District name and Average price.
- Also, in that website (<https://mogi.vn/ho-chi-minh/thue-mat-bang-cua-hang-shop>), I will take the data of 60 latest articles of each district and 1200 latest articles of office, store, shop leasing with the information of address, area and price of monthly rental and convert it to a data frame to calculate the average leasing price per square meters in each District and other analysis.
- Foursquare API to explore the most common venues in each District.
- To create the map, I need to use Geocoder to get the latitude and longitude. However, some districts are large, and the coordinates generated by Geocoder are not the good ones (not many venues surrounding the point), I would use Google Map to adjust the coordinates of 3 districts (District 7, District 9 and District Binh Tan).
- I created the GeoJSON file of HCMC on <http://geojson.io> to draw choropleth maps.

After getting the data from the website, we need to do some preprocessing: extract only District from the full address, delete the unit label of each value, convert the price from numbers and words into numbers.

Methodology

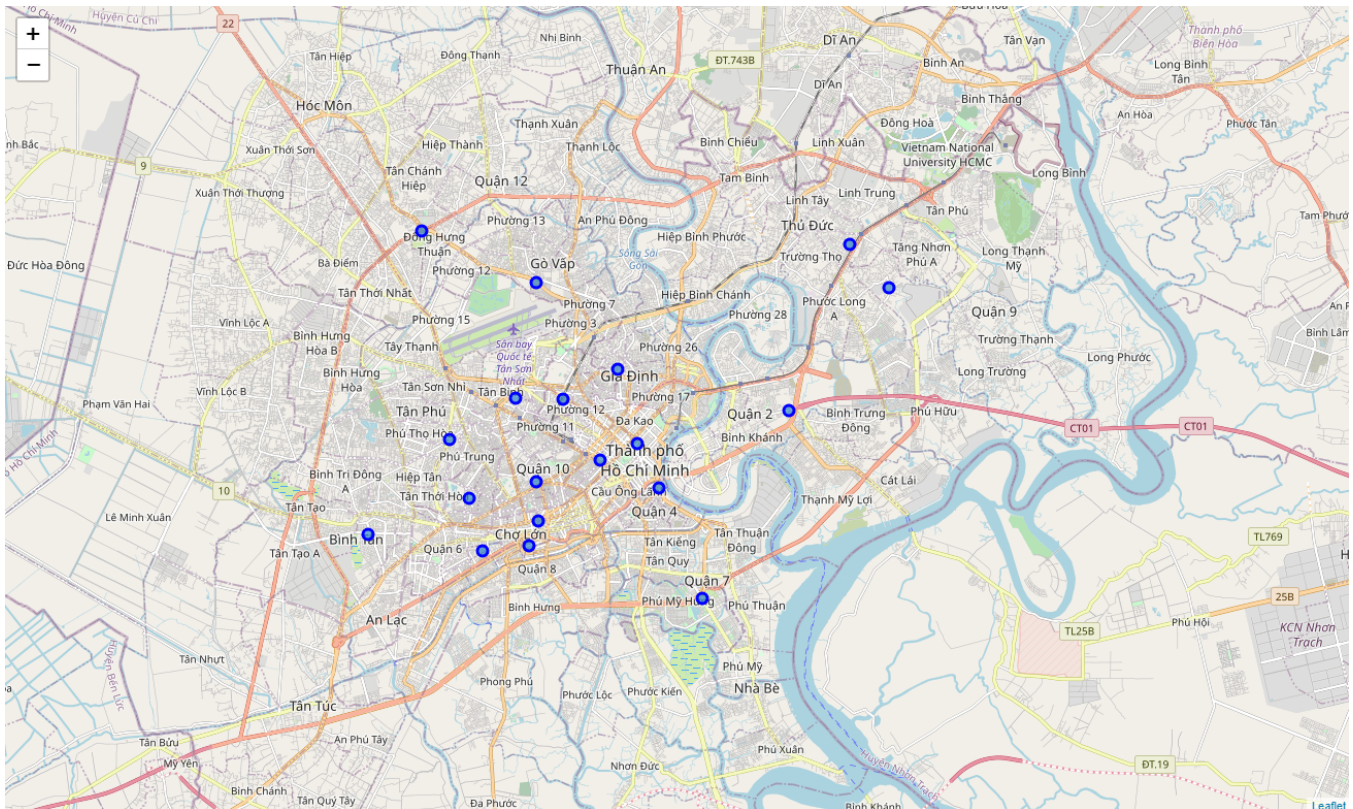
Average Housing Prices in Ho Chi Minh City

I scraped from the website mentioned above and got the list of 19 urban districts in HCMC with the average housing prices. The table here is the data with the average housing price calculated in million VND per square meter.

	District	Avg Housing Price
0	District 1	414.0
1	District 2	72.6
2	District 3	266.0
3	District 4	100.0
4	District 5	254.0
5	District 6	113.0
6	District 7	88.4
7	District 8	68.8
8	District 9	49.7
9	District 10	209.0
10	District 11	166.0
11	District 12	46.9
12	District Binh Tan	62.8
13	District Binh Thanh	132.0
14	District Go Vap	98.0
15	District Phu Nhuan	180.0
16	District Tan Binh	146.0
17	District Tan Phu	103.0
18	District Thu Duc	64.7

Map of Districts in Ho Chi Minh City

I used Geocoder to get coordinates of each district then used the python folium library to create a map of HCMC with geographic details and markers of the districts. However, I saw that the given coordinates of District 7, District 9, and District Binh Tan were not the good ones because the 3 areas are too big, and the points were not at the crowded living neighborhoods. Therefore, I got better coordinates from Google Maps. Finally, I create a map as below:



Calculating the Average Price for Leasing in Ho Chi Minh City

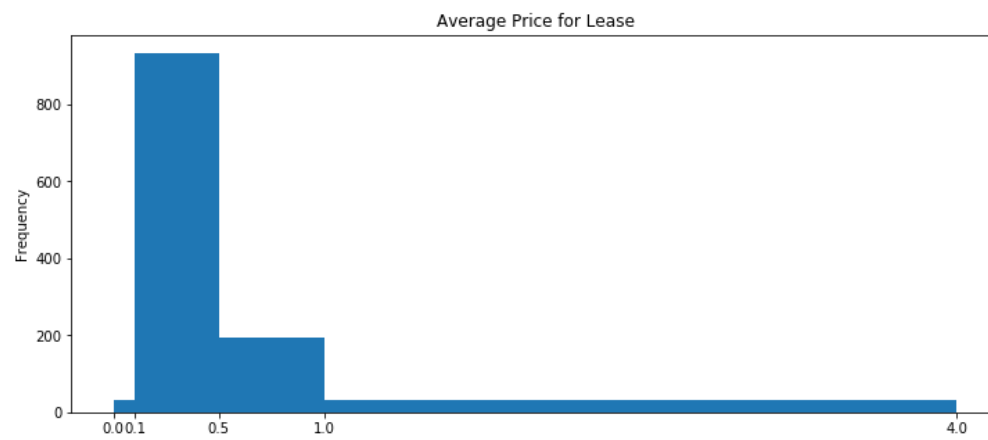
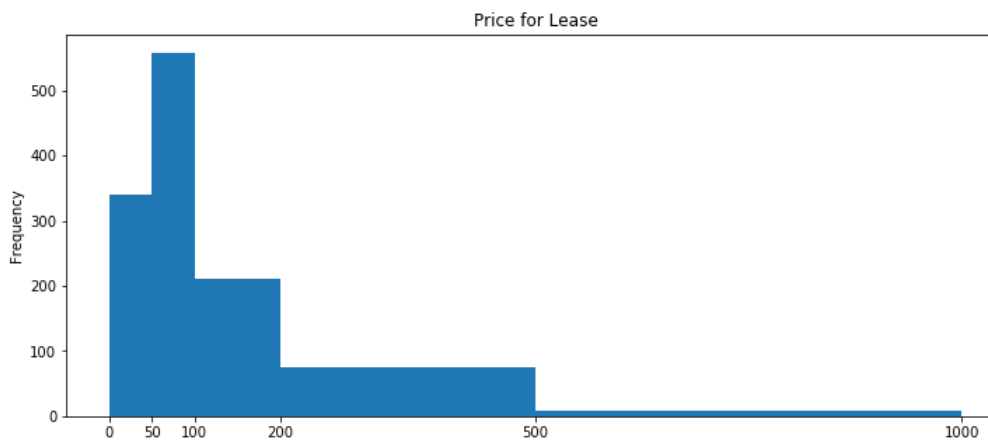
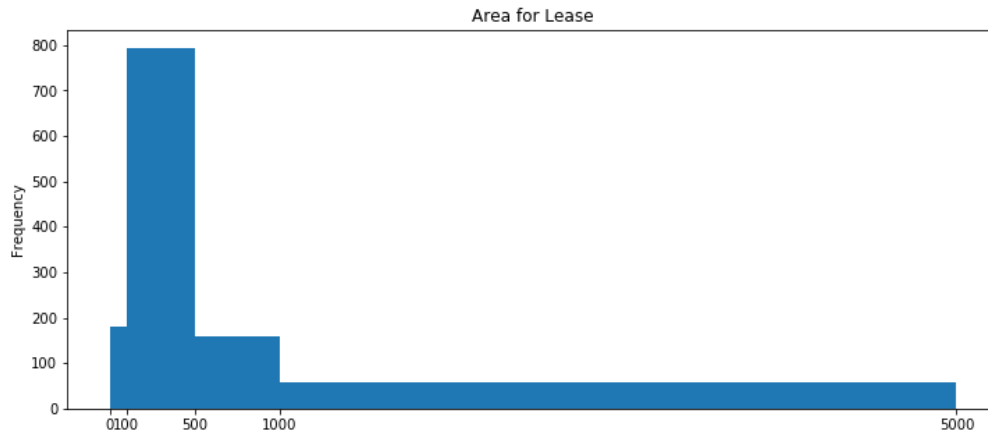
There were no available average prices for lease on the internet, so I calculated them with the data scraped from the real estate broker website mentioned above. I scraped 60 latest articles of each district to find out the average price for lease (million VND per square meter per month) in each district of HCMC. The data was preprocessed by dropping articles that had the area for lease less than 5 square meters and the average price more than 3 million VND per square meter per month. We have the table of both average housing price and average price for lease for all the districts:

	District	Avg Housing Price	Avg Price for Lease
0	District 1	414.0	0.406495
1	District 2	72.6	0.392694
2	District 3	266.0	0.263403
3	District 4	100.0	0.266698
4	District 5	254.0	0.375345
5	District 6	113.0	0.238760
6	District 7	88.4	0.452621
7	District 8	68.8	0.275344
8	District 9	49.7	0.238519
9	District 10	209.0	0.223050
10	District 11	166.0	0.267159
11	District 12	46.9	0.227379
12	District Binh Tan	62.8	0.178719
13	District Binh Thanh	132.0	0.225084
14	District Go Vap	98.0	0.244059
15	District Phu Nhuan	180.0	0.234923
16	District Tan Binh	146.0	0.211840
17	District Tan Phu	103.0	0.227169
18	District Thu Duc	64.7	0.303811

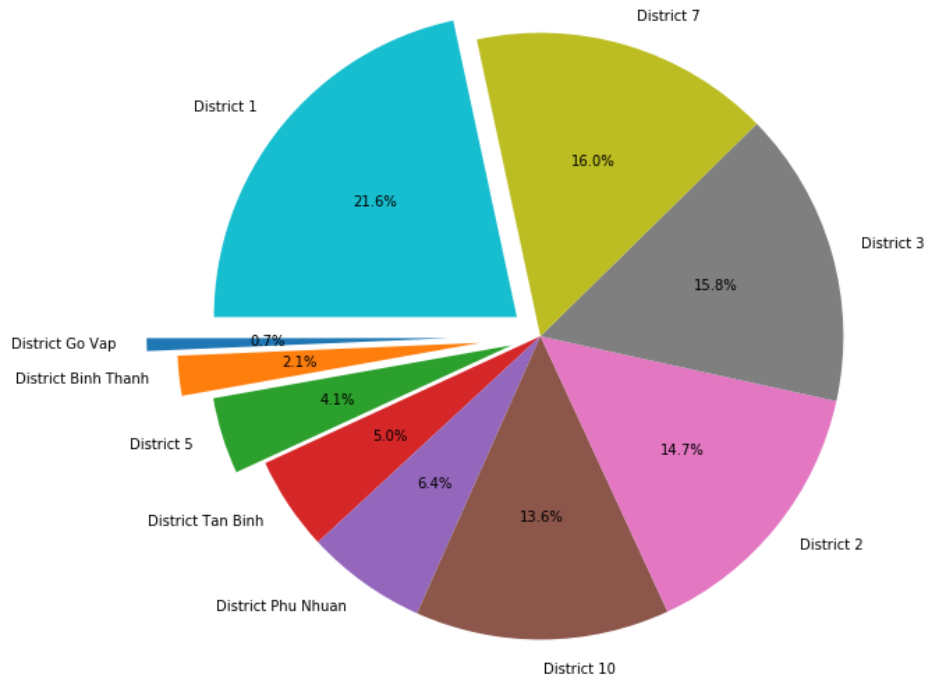
However, this calculation is not very accurate because some articles collected were posted more than 1 year ago. There were not many places for lease in some districts (such as Thu Duc, District 9) posted on that website recently.

Analysis of Leasing demand in Ho Chi Minh City

Moreover, I would like to know the demand for leasing place in HCMC. I used 1200 latest articles of HCMC scraped from the same website to create a data frame and then made histogram charts and a pie chart to analyze the leasing demand. We can find the current situation of renting a place, how much is the area and price that most of the owners offer, and what districts have the most available places for lease.



Percentage of Place for Lease in Ho Chi Minh City (in 1200 latest articles)



Exploring Neighborhood in Ho Chi Minh City

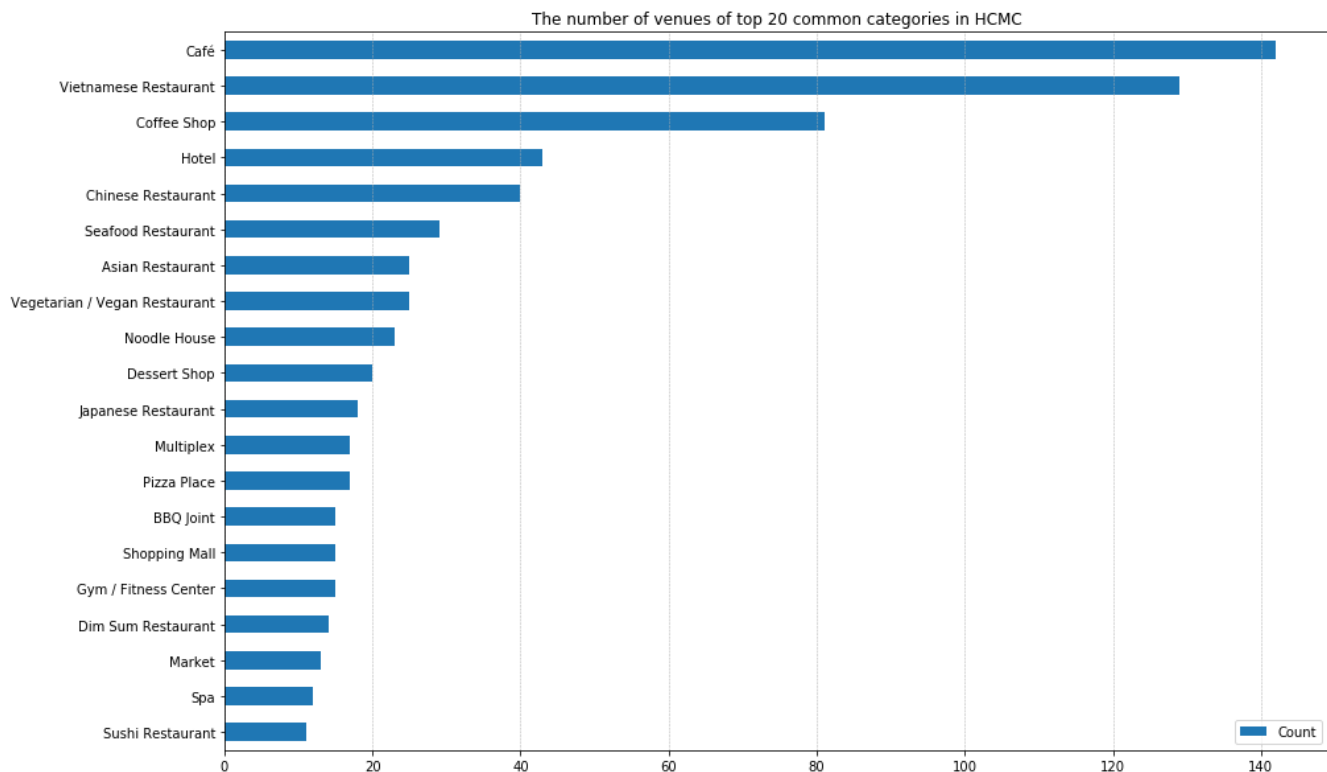
I explored the districts with Foursquare API with the limit of 100 venues and a radius of 1,500 meters for each district from the given coordinates. We got the data frame of the list of all the venues found in the area:

	District	Latitude	Longitude	VenueName	VenueLatitude	VenueLongitude	VenueCategory
0	Quan 1	10.78095	106.69911	Pacey Cupcakes	10.780202	106.700773	Cupcake Shop
1	Quan 1	10.78095	106.69911	Pizza 4P's Hai Bà Trưng	10.783017	106.697017	Pizza Place
2	Quan 1	10.78095	106.69911	MGallery Hotel des Art	10.782382	106.697103	Hotel
3	Quan 1	10.78095	106.69911	Club Lounge @ InterContinental Saigon	10.780975	106.700968	Hotel Bar
4	Quan 1	10.78095	106.69911	Social Club @ Hotel Des Arts	10.782372	106.697303	Hotel Bar

District	
District 1	100
District 10	100
District 11	36
District 12	10
District 2	28
District 3	100
District 4	100
District 5	59
District 6	25
District 7	100
District 8	65
District 9	4
District Binh Tan	13
District Binh Thanh	75
District Go Vap	18
District Phu Nhuan	100
District Tan Binh	91
District Tan Phu	17
District Thu Duc	28

1069 venues and 134 unique categories were found. Some districts reached the limit of 100 venues, but some only showed under 20 venues including District 9, District 12, Binh Tan, Tan Phu, Go Vap. The reason for this is because the areas of the districts are big and most of the venues there are small and not registered on Foursquare.

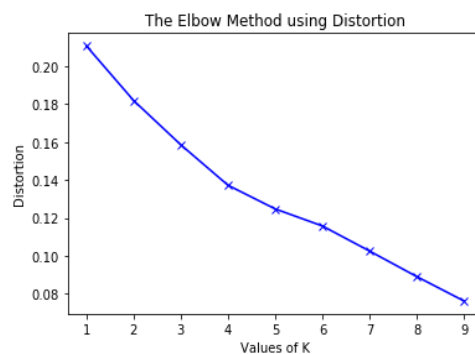
We can have a glance at the top 20 unique categories in HCMC:



As can be seen from the graph, the most popular venue is Café, the second is the Vietnamese Restaurant and the third is the Coffee Shop.

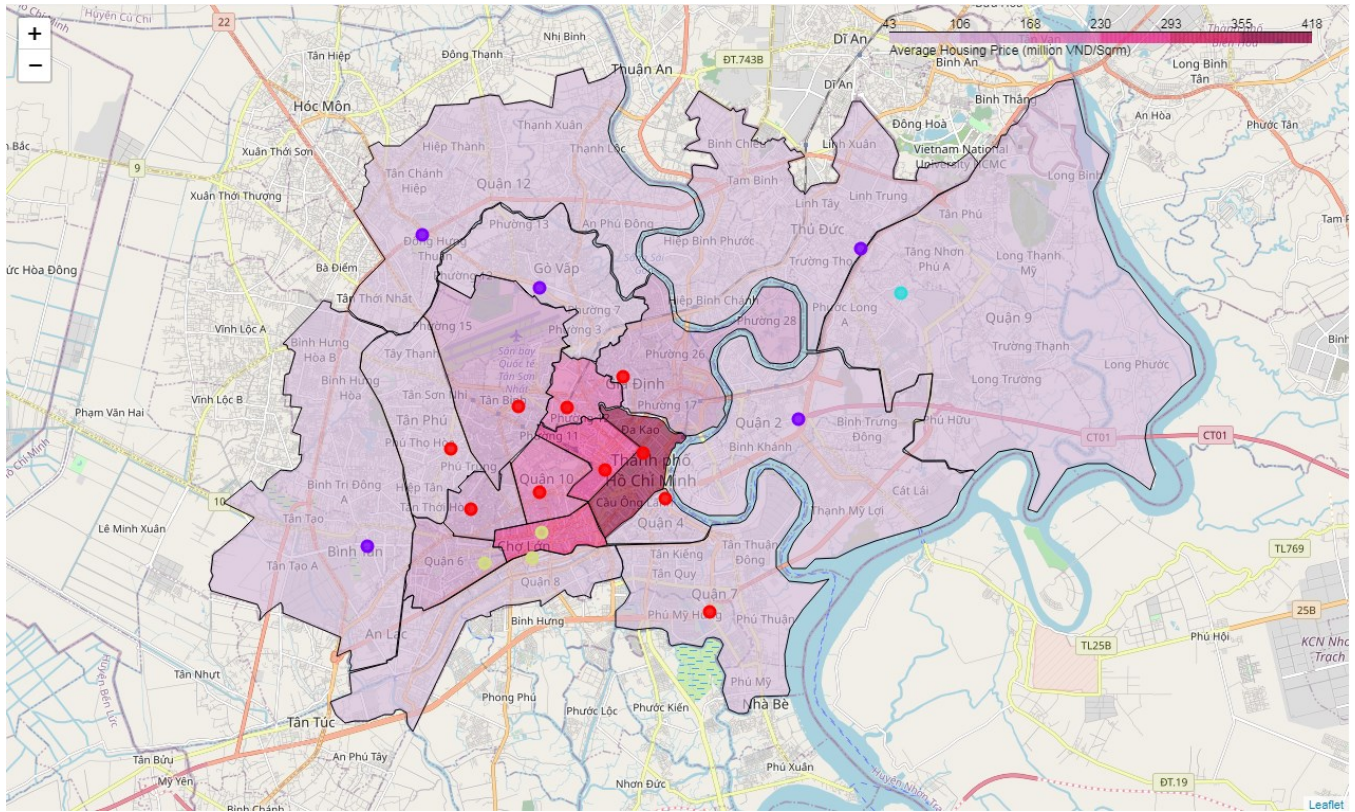
Clustering Districts of Ho Chi Minh City

From the data, I used K-means Clustering to divide the districts into groups. To find the best K, I used the elbow method to see how the distortion changed when we increased the K.



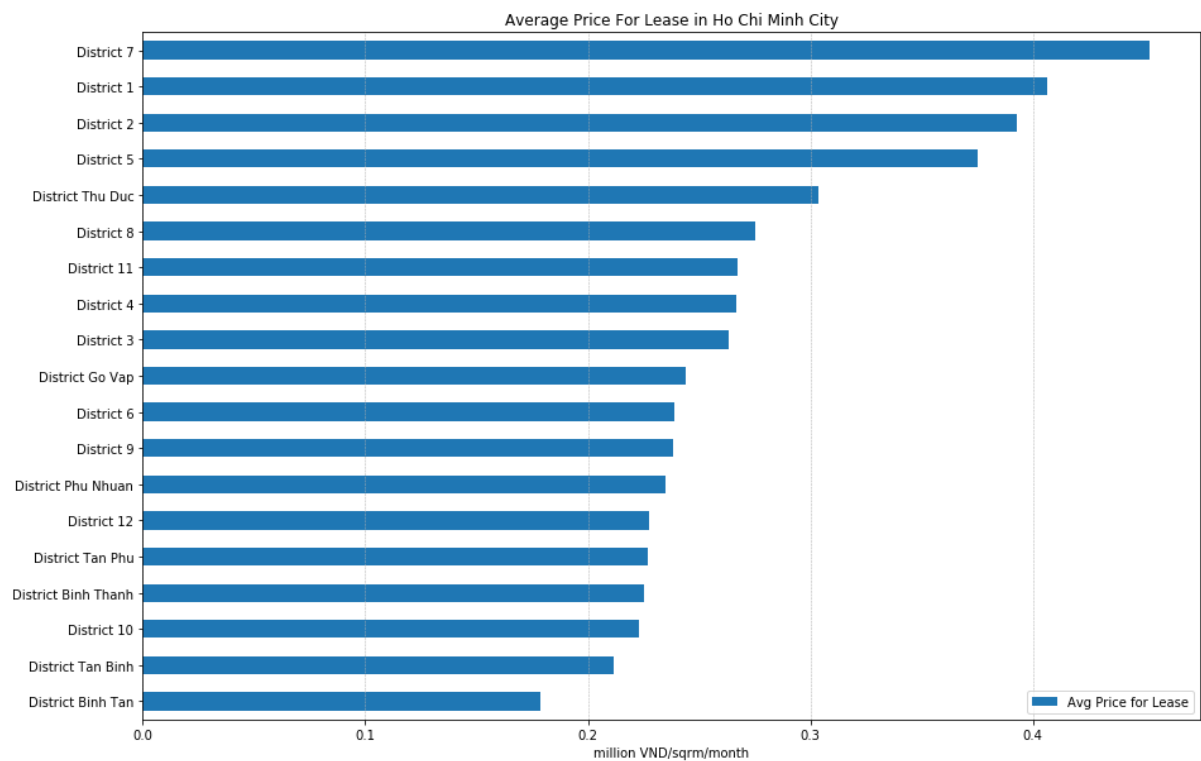
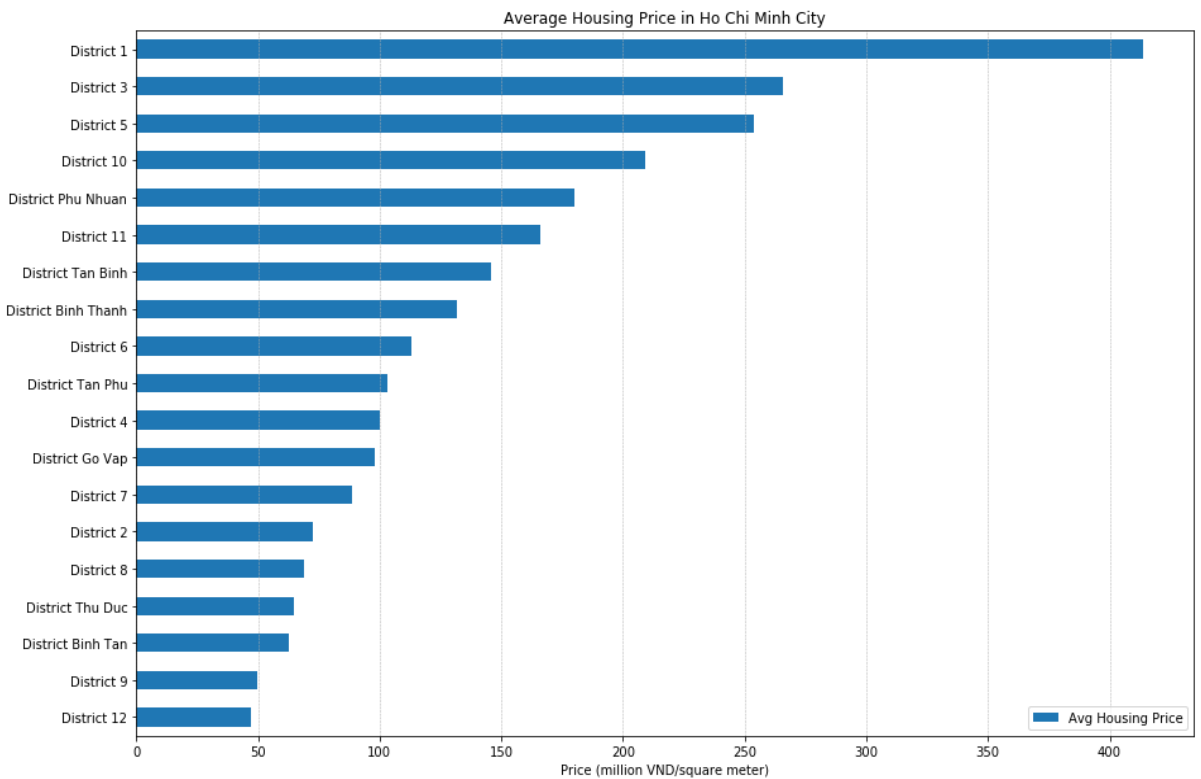
It is not a chart that is easy to see the elbow, so we can understand that the neighborhoods of districts are not very typical to differentiate them into groups. However, there is a little bit change when $k=4$, so, I divided them into 4 groups.

And we had the map of HCMC after clustered with the average housing price of each district. I displayed the average housing prices and the clusters of districts with a choropleth map as below:



Result

Average Housing Price and Price for Lease of each District



As I said above, the Average Price for Lease (APL) is not very accurate because there is not much information about some districts and I only take around 60 latest articles. However, we can see that in general, the districts having higher AHP have higher APL. District 7 has not very high AHP but have the highest APL, this is because District 7 is big, it has an area of modern buildings, shopping malls with high AHP and APL, but it also has an industrial area low AHP and not many places for lease.

Moreover, the gaps of APL among the districts are not as significant as that of AHP. It can be seen from the first graph that District 1 has the highest AHP in HCMC which is more than 8 times higher than the lowest price (of District 12), and far surpass the second-highest number (of District 3). But, about the APL, the highest one (of District 7) is just 2.5 times higher than the lowest one (of District Binh Tan).

Leasing Demand

At the moment, 2/3 of available places for lease have a medium area of 100-500 square meters. Small area (<100 sqm) acquires 1/6 and the rest are big and very big area (>500 sqm and >1000 sqm)

About the price for lease per month, near 50% of available places for lease cost around 50-100 MVND, more than 25% cost under 50 MVND, only 16% cost 100-200 MVND, and the remaining cost more than 200 MVND.

80% of the average prices are 0.1-0.5 MVND/square meter/month.

Therefore, SMEs with low budgets can still have a place to sell their product, and the chance to achieve that is not very difficult.

District 1, District 7, District 3 offers the most places for lease currently. The districts have high population density and in the center of the city, but the APLs are higher than other districts'. So, this is a good opportunity for SMEs if they have a high budget and excellent developing plan.

4 Clusters of Districts

We have 4 groups of Districts as below:

Group 1:

District	Avg Housing Price	Avg Price for Lease	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
District 1	414.0	0.406495	Café	Hotel	Vietnamese Restaurant	Coffee Shop	Massage Studio
District 3	266.0	0.263403	Hotel	Vietnamese Restaurant	Café	Coffee Shop	Hostel
District 4	100.0	0.266698	Hotel	Vietnamese Restaurant	Café	Coffee Shop	Massage Studio
District 7	88.4	0.452621	Vietnamese Restaurant	Café	Coffee Shop	Fried Chicken Joint	Japanese Restaurant
District 10	209.0	0.223050	Café	Vietnamese Restaurant	Coffee Shop	Chinese Restaurant	Seafood Restaurant
District 11	166.0	0.267159	Café	Vietnamese Restaurant	Chinese Restaurant	Seafood Restaurant	Cantonese Restaurant
District Binh Thanh	132.0	0.225084	Café	Vietnamese Restaurant	Coffee Shop	Vegetarian / Vegan Restaurant	Asian Restaurant
District Phu Nhuan	180.0	0.234923	Vietnamese Restaurant	Café	Coffee Shop	Hotel	Asian Restaurant
District Tan Binh	146.0	0.211840	Vietnamese Restaurant	Café	Coffee Shop	Hotel	Seafood Restaurant
District Tan Phu	103.0	0.227169	Vietnamese Restaurant	Café	Supermarket	Fast Food Restaurant	Movie Theater

This group has many Café, Vietnamese Restaurant and Coffee Shop (which are also top 3 common venues in HCMC). The districts are highly populated, have high AHP and APL.

Group 2:

District	Avg Housing Price	Avg Price for Lease	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
District 2	72.6	0.392694	Café	Shopping Mall	Vietnamese Restaurant	Coffee Shop	Department Store
District 12	46.9	0.227379	Café	Park	Noodle House	Coffee Shop	Gym
District Binh Tan	62.8	0.178719	Café	Shopping Mall	Coffee Shop	Fast Food Restaurant	Multiplex
District Go Vap	98.0	0.244059	Café	Pizza Place	Shopping Mall	Lounge	Coffee Shop
District Thu Duc	64.7	0.303811	Café	Shopping Mall	Vietnamese Restaurant	Coffee Shop	Multiplex

This group also has many Café but it also has a shopping mall as the second most common venue. These districts are very big, the AHP and APL are not very high.

Group 3:

District	Avg Housing Price	Avg Price for Lease	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
District 9	49.7	0.238519	Lake	Restaurant	Shopping Mall	Market	Eastern European Restaurant

The Foursquare API found only 4 venues in this district. It has big factories and industrial zones, so there are not many venues.

Group 4:

District	Avg Housing Price	Avg Price for Lease	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
District 5	254.0	0.375345	Chinese Restaurant	Vietnamese Restaurant	Coffee Shop	Noodle House	Dim Sum Restaurant
District 6	113.0	0.238760	Chinese Restaurant	Dessert Shop	Cantonese Restaurant	Market	Dim Sum Restaurant
District 8	68.8	0.275344	Chinese Restaurant	Dim Sum Restaurant	Café	Vietnamese Restaurant	Coffee Shop

This group has Chinese Restaurants as the most common venue. It is easy to understand because, the area has China Town, the biggest market with many traders and Chinese-origin people living. That is why the AHP and APL of this group are quite high.

Discussion

In this study, I used a real estate website to research the leasing situation in Ho Chi Minh City. However, a website cannot show all accurate information. Some of the articles are fake where the owners want to attract customers with low prices and big areas so that people would contact them and maybe they will have an effect on the final result of my project.

I used the Kmeans algorithm as part of this clustering study. When I tested the Elbow method, there is not a clear elbow, so the higher k, the more accurate the clustering result is.

In general, the project gives a review of the housing price, the price for leasing, and the neighborhood of each district in Ho Chi Minh City. SMEs can take these results for reference to buy or rent a place to build a store and set up a financial plan for their business. For example, if an entrepreneur wants to open a café, he/she can choose a place in group 1 where there are many people (potential customers) if they have enough budget. If you want to open a restaurant, group 1 or group 4 is reasonable. Moreover, this project also helps people who want to buy a house in HCMC, if they want to live in a neighborhood with many shopping malls and café, the districts in group 2 are suitable.

Besides that, currently, District 1, District 7 and District 3 have the most articles of places for lease. I think one of the reasons for this is COVID-19 because in the period of this virus breakout, some companies cannot afford to maintain the activities, many entrepreneurs return the place as mentioned in a lot of newspapers. I think this will be a chance for companies who have a good financial plan to open a store in these developed districts.

Conclusion

As a result of urbanization, people are moving from provinces to big cities to start a business or to live. Therefore, to have the access to the accurate informative resources will help people to live in the place with the neighborhood they want and give an opportunity for a start-up to approach the right market where they can easily attract their potential customers. Although this project is just at a basic level, I believe it will somehow help people to give a good decision on choosing a place for business or living.