

BATTLE OF NEIGHBORHOODS

(Capstone Project)

Housing Price In Ho Chi Minh City

Author : Phan Dang Nam Phuong

1. Introduction

1.1 Background

Ho Chi Minh City (HCMC) is the financial centre of Vietnam.

Along with economic development, demand in the real estate market, HCMC housing sale prices now go up and up.

1.2 Problem

Do you ever wonder why housing price is differ in each districts of a city? Are districts that have expensive house similar or dissimilar with districts have cheap houses?

This project will explore the similarities and dissimilarities between certain districts in Ho Chi Minh City, and determine how it make impact on the housing sale price.

This project will help people who want to know where to buy a house at reasonable price.

2. Data:

The house datasets consist of Acreage, number of rooms, districts, price, latitude and longitude.

	District	Acreage	Rooms	Price	Price_m2	lat	long
0	8	45.2	4	4.7	0.103982	10.723333	106.627778
1	4	69.2	7	21.5	0.310694	10.761667	106.702500
2	Phu Nhuan	55.0	2	6.8	0.123636	10.801667	106.677500
3	7	59.0	1	4.8	0.081356	10.738611	106.726389
4	Binh Chanh	243.0	5	13.0	0.053498	10.750278	106.512500

Latitude and longitude coordinates of each district in Ho Chi Minh City is downloaded from Wikipedia.

Other are scraped on website: <https://propzy.vn> , a website sell house in Viet Nam.

Neighborhood venues information are collected from Foursquare API search, which help us to determine how qualities of a district.

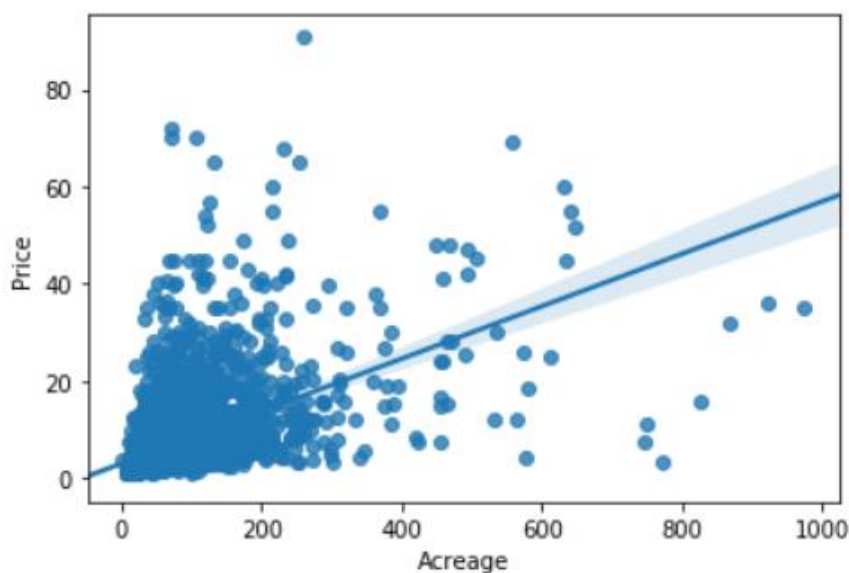
District		District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	1	10.771423	106.698471	Pizza 4P's	10.773301	106.697599	Pizza Place
1	1	10.771423	106.698471	Maison Marou Saigon	10.769667	106.697925	Dessert Shop
2	1	10.771423	106.698471	Quán Lộc Thành	10.772472	106.698441	Vietnamese Restaurant
3	1	10.771423	106.698471	Quynh Nhu Foot Massage	10.771070	106.700277	Massage Studio
4	1	10.771423	106.698471	Bep Me In	10.773810	106.698152	Vietnamese Restaurant

3. Methodology:

3.1 Exploratory Analysis:

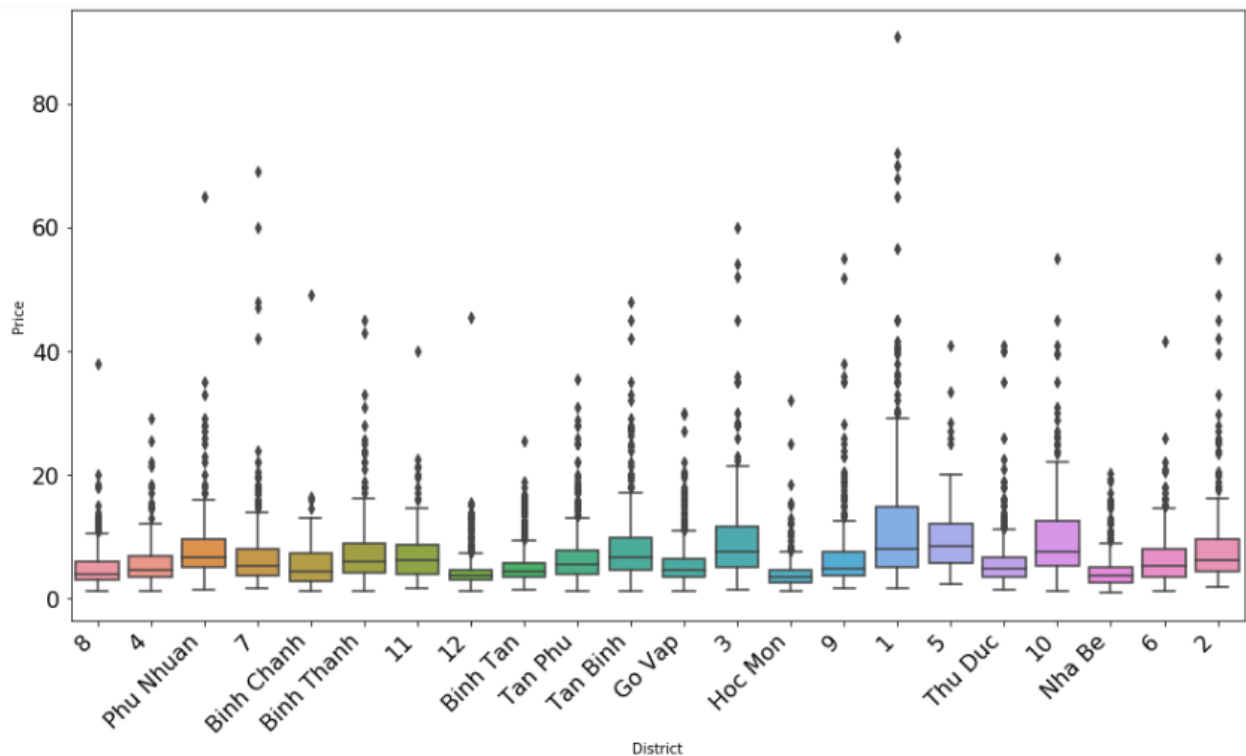
After cleaning datasets, we create some visualizations to interpret the data we have:

We use regplot to see correlation between Acreage with Price.

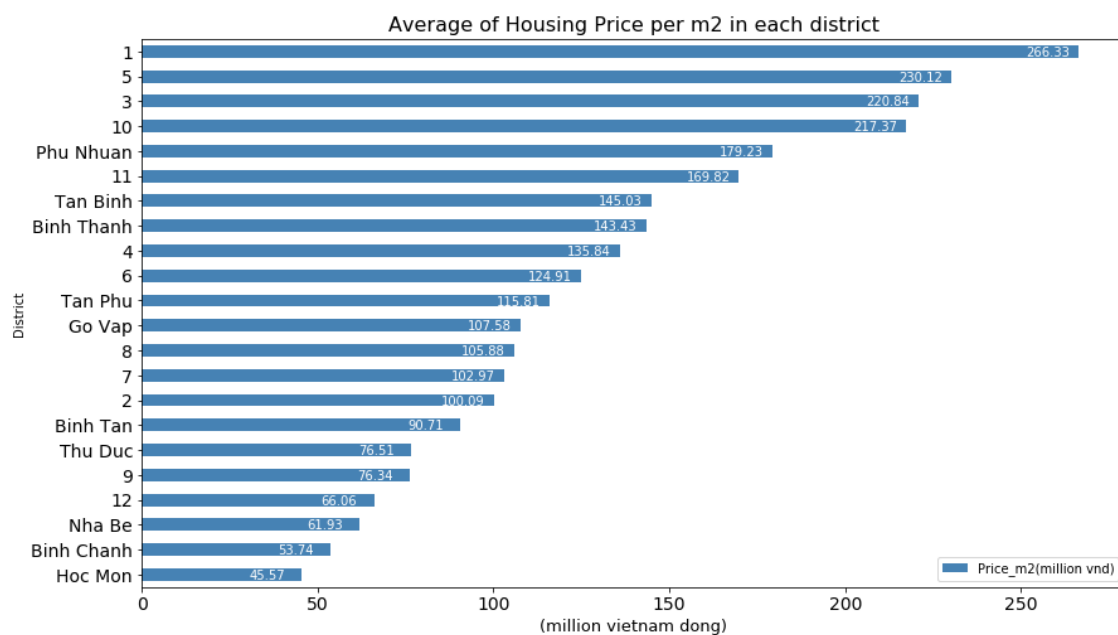


We can easily see the regression line show the higher acreage, the higher price.

In the next plot, we can see boxplot describe the distribution of price in each districts.



Almost house prices has median values appropriate 3-6 billions.



The above is a bar chart displaying districts in descending order of housing price. District 1 has the highest housing price (266 millions vnd/m²) , followed by District 5, 3, 10.

3.2 Cluster Analysis

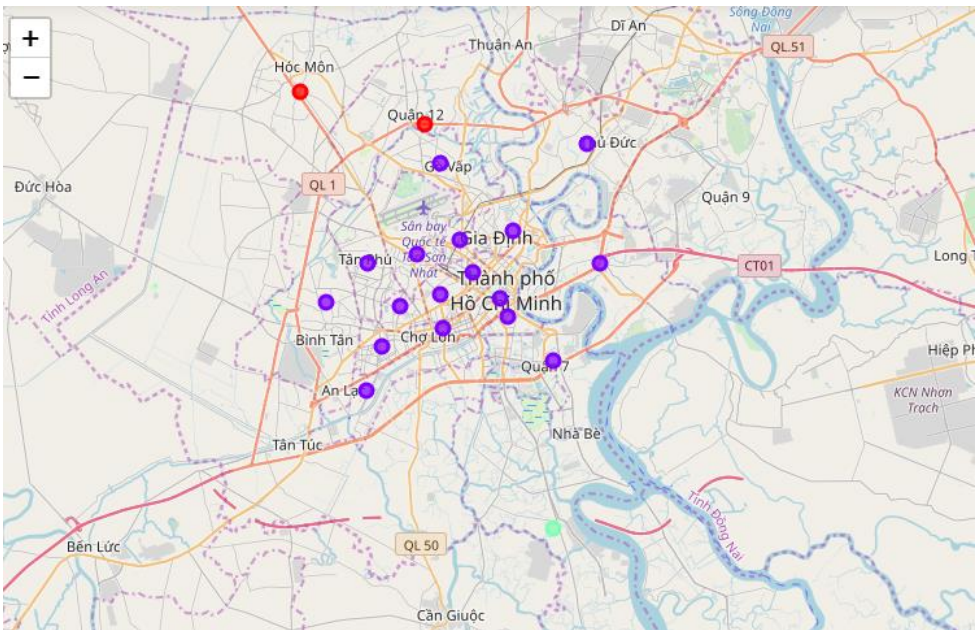
Firstly, with the venues data was gotten from Foursquare, one-hot encoding was conducted to give binary values.

	District	Accessories Store	American Restaurant	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	BBQ Joint	Baby Store	Bagel Shop	Bakery	...	Thai Restaurant	Theater	Theme Park	Travel Agency	Veget / V Restai
0	1	0.000	0.000000	0.000000	0.020000	0.000000	0.020000	0.0	0.000000	0.01	...	0.000000	0.000000	0.000000	0.000000	0.00
1	10	0.000	0.000000	0.015873	0.015873	0.000000	0.000000	0.0	0.000000	0.00	...	0.000000	0.015873	0.000000	0.000000	0.03
2	11	0.000	0.000000	0.000000	0.052632	0.000000	0.000000	0.0	0.000000	0.00	...	0.000000	0.000000	0.052632	0.000000	0.05
3	12	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.00	...	0.000000	0.000000	0.000000	0.000000	0.00
4	2	0.000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	0.00	...	0.000000	0.000000	0.000000	0.000000	0.00
5	3	0.000	0.000000	0.000000	0.060000	0.000000	0.010000	0.0	0.000000	0.00	...	0.000000	0.000000	0.000000	0.000000	0.02

We can create a dataframe of top ten venues in each districts as follows.

District		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	1	Vietnamese Restaurant	Hotel	Café	Coffee Shop	Sandwich Place	Bar	Spa	Food Court	Juice Bar	Restaurant
1	10	Vietnamese Restaurant	Café	Coffee Shop	Seafood Restaurant	Vegetarian / Vegan Restaurant	Fast Food Restaurant	Ice Cream Shop	Burger Joint	Electronics Store	Korean Restaurant
2	11	Café	Vietnamese Restaurant	Seafood Restaurant	Chinese Restaurant	Pizza Place	Vegetarian / Vegan Restaurant	Asian Restaurant	Theme Park	Shopping Mall	Fast Food Restaurant
3	12	Café	Food Truck	Beer Garden	Yoga Studio	French Restaurant	Dessert Shop	Dim Sum Restaurant	Diner	Dumpling Restaurant	Electronics Store
4	2	Café	Coffee Shop	Snack Place	Electronics Store	Convention Center	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Dumpling Restaurant

Next, base on the venue categories, k-means clustering was conducted to group the districts into 3 different clusters base on their similarities. Each clusters is represented by different color points located on HCMC map (blue, red, green).



4. Result

With K –Means algorithm , the districts are divided into 3 clusters:

Cluster 0:

	District	Price_m2	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	1	0.266327	Vietnamese Restaurant	Hotel	Café	Coffee Shop	Sandwich Place	Bar	Spa	Food Court	Juice Bar	Restaurant
1	10	0.217372	Vietnamese Restaurant	Café	Coffee Shop	Seafood Restaurant	Vegetarian / Vegan Restaurant	Fast Food Restaurant	Ice Cream Shop	Burger Joint	Electronics Store	Korean Restaurant
2	11	0.169823	Café	Vietnamese Restaurant	Seafood Restaurant	Chinese Restaurant	Pizza Place	Vegetarian / Vegan Restaurant	Asian Restaurant	Theme Park	Shopping Mall	Fast Food Restaurant
4	2	0.100087	Café	Coffee Shop	Snack Place	Electronics Store	Convention Center	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Dumpling Restaurant

Cluster 1:

	District	Price_m2	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
17	Nha Be	0.061933	Electronics Store	Yoga Studio	French Restaurant	Department Store	Dessert Shop	Dim Sum Restaurant	Diner	Dumpling Restaurant	Fast Food Restaurant	Flea Market

Cluster 2:

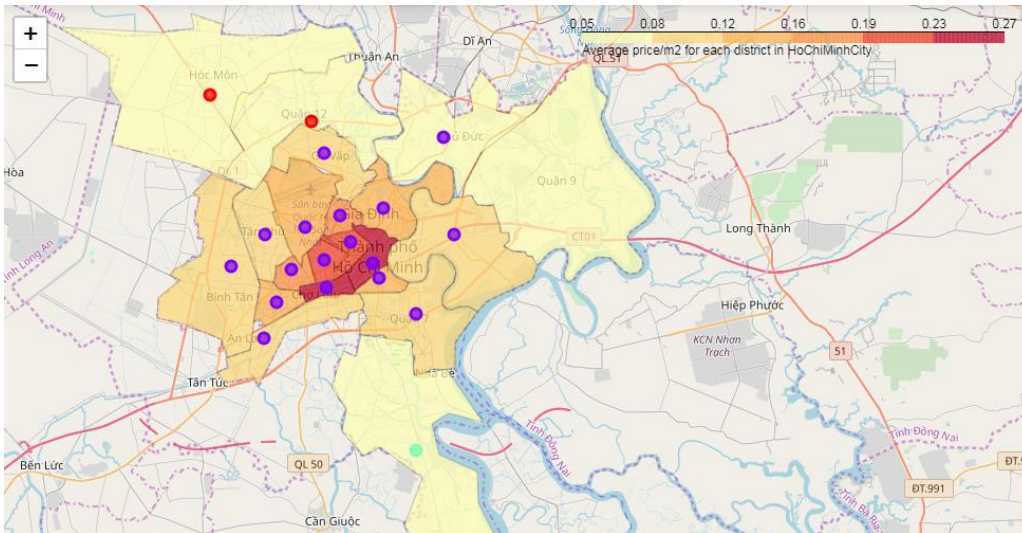
	District	Price_m2	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
3	12	0.066063	Café	Food Truck	Beer Garden	Yoga Studio	French Restaurant	Dessert Shop	Dim Sum Restaurant	Diner	Dumpling Restaurant	Electronics Store
16	Hoc Mon	0.045573	Café	Arts & Crafts Store	Yoga Studio	French Restaurant	Dessert Shop	Dim Sum Restaurant	Diner	Dumpling Restaurant	Electronics Store	Fast Food Restaurant

We use a choropleth map to visualize the gradient of housing price/m², with the color point is represents for each venues cluster.

We can see cluster 0 are bustling area with full of restaurant, hotel, café, store,.....which are represented by blue points.

Cluster 1 are quiet residential area with electronics store, yoga studio, desert shop, market,.....which is represented by green point.

Cluster 2 are good living area with café, food trucks, yoga studio, beer garden, dessert shop,... are presented by red points.



5. Discussion:

From this project, we can see that the more you go into the city center, the higher housing price. District 1 is the most expensive place to live in, followed by District 5...which has much higher housing sale price than others. Although most districts of HCM city has similar venues, except suburban area.

So, should we choose 7/8/Binh Tan/ Tan Phu Districts to buy house, where has cheaper housing price but still guaranteed for living needs?

6. Conclusion

This project help us have overview of housing sale prices in HCMC and partly make right choice in choosing districts to buy house.