Business Research in Hanoi City

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

Nov 09, 2020

Contents

1 Problem's description

Hanoi is the Capital of Vietnam, and is a leading city that has a significant role in the growth of the country. Because of its dense population and openness to new opportunities, Hanoi is an ideal place for investors, for entrepreneurs to start their businesses or make valuable investments.

However, from an investor's perspective, it will be hard to figure out which type of business to open and in which area that business could be attractive to customers and give optimal profit to owner.

By using Data Science and exploring some geometric data of Hanoi, you can have an overview of the area and understand the general characteristics of different districts in Hanoi. In order to provide investors with this information, it is necessary to answer to some of the questions below

1. *How many venues in each district?* Answering this question gives us a better understanding of *the dynamic level* of a district.
2. *How many categories in each district?* Answering this question helps us know which services the districts offer and how diverse the business operations in these districts are.
3. *What are the top categories in each district?* Based on this question, investors can get a hold of the districts' characteristics. (E.g: Is it an entertainment center? Is it a food center?)
4. *If we divide the clusters into 3 groups, what can we see about the characteristics of these groups?*
5. *If we visualize all information on the map, how does it look like?*

2 Data presentation

In order to answer the questions, there is some data that we need to collect.

1. Information of Hanoi City (HNC)'s districts from Wikipedia. It gives us a list of all main districts of HNC with their area (in Km^2), population (in 2017) and the density of each district (people/Km^2). Wiki page: https://en.wikipedia.org/wiki/Hanoi
2. List of the coordinates (latitude, longitude) of all districts in HNC. This list can be generated from the package *geopy.geocoders.Nominatim*.
3. List of average housing prices per m^2 in HNC. The list is updated through https: //mogi.vn/gia-nha-dat.

3 Methodology

1. First, we need to collect the data by crawling the table of Hanoi's districts' information on the wikipedia page and the average housing price on a website. The *BeautifulSoup* package is used for the scrapping.

2. The column *Density* is calculated by dividing *Population by area* of each district.

3. *Numpy* and *pandas are the most frequent used* packages to manipulate dataframes in this project.

4. *geopy.geocoders.Nominatim is used* to get the coordinates of districts and add them to the main data frame.

5. *Folium* package is used to visualize the HNC map with its districts.

6. *Foursquare API is used* to explore the venues in each district

7. *K-Means Clustering* method is implemented to segment the districts. We then examine the characteristics of each district cluster based on venue observation.

8. In order to visualize the charts, seaborn is used.

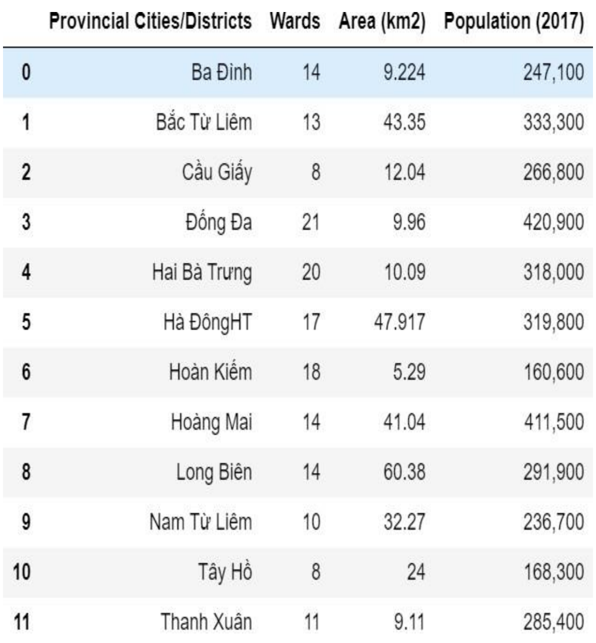
9. Package *folium* is used to visualize the clusters on the map.

4 Results

The main data frame

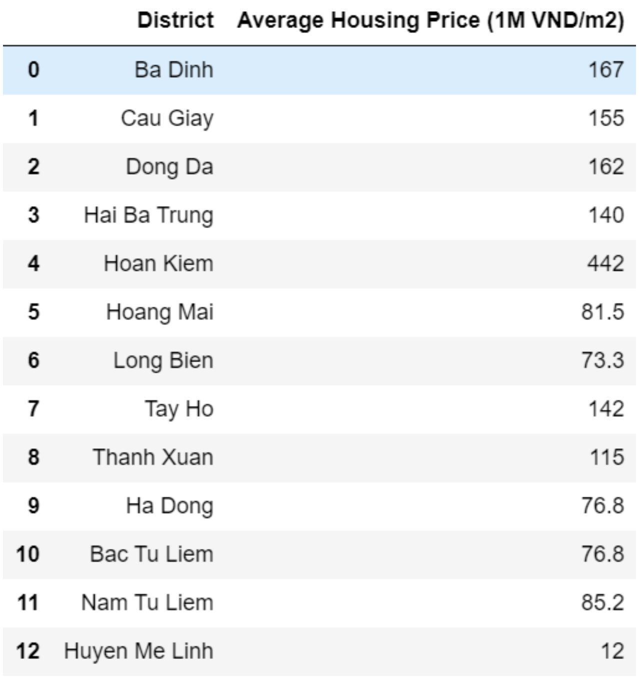
After crawling data from the Internet, we obtain 2 tables as below:

Figure 1



	Provincial Cities/Districts	Wards	Area (km2)	Population (2017)
0	Ba Đình	14	9.224	247,100
1	Bắc Từ Liêm	13	43.35	333,300
2	Cầu Giấy	8	12.04	266,800
3	Đống Đa	21	9.96	420,900
4	Hai Bà Trưng	20	10.09	318,000
5	Hà ĐôngHT	17	47.917	319,800
6	Hoàn Kiếm	18	5.29	160,600
7	Hoàng Mai	14	41.04	411,500
8	Long Biên	14	60.38	291,900
9	Nam Từ Liêm	10	32.27	236,700
10	Tây Hồ	8	24	168,300
11	Thanh Xuân	11	9.11	285,400

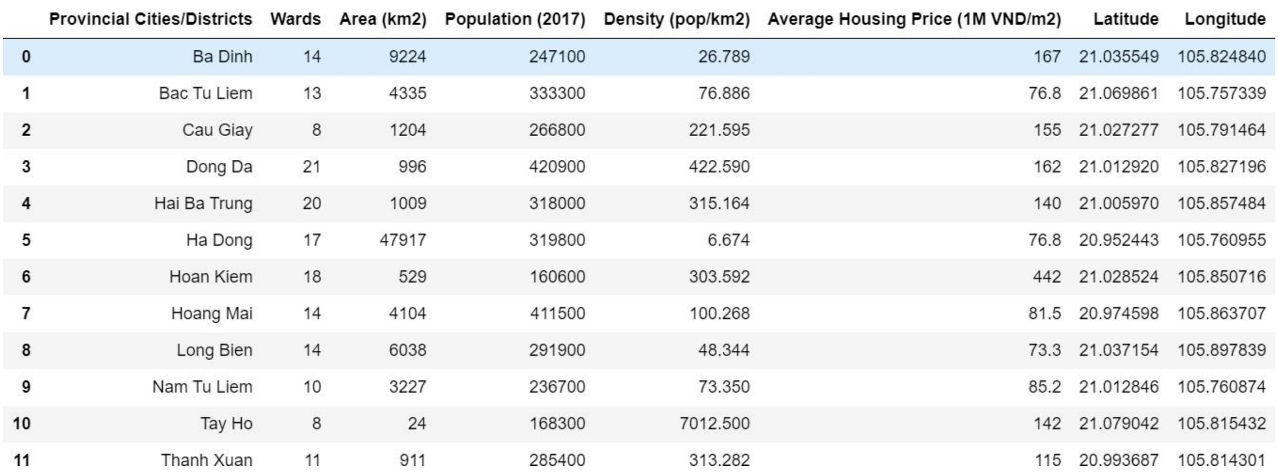
Figure 2



	District	Average Housing Price (1M VND/m2)
0	Ba Dinh	167
1	Cau Giay	155
2	Dong Da	162
3	Hai Ba Trung	140
4	Hoan Kiem	442
5	Hoang Mai	81.5
6	Long Bien	73.3
7	Tay Ho	142
8	Thanh Xuan	115
9	Ha Dong	76.8
10	Bac Tu Liem	76.8
11	Nam Tu Liem	85.2
12	Huyen Me Linh	12

After combining 2 tables and getting lat, long of these districts, we have the main data frame as follow:

Figure 3: Data frame



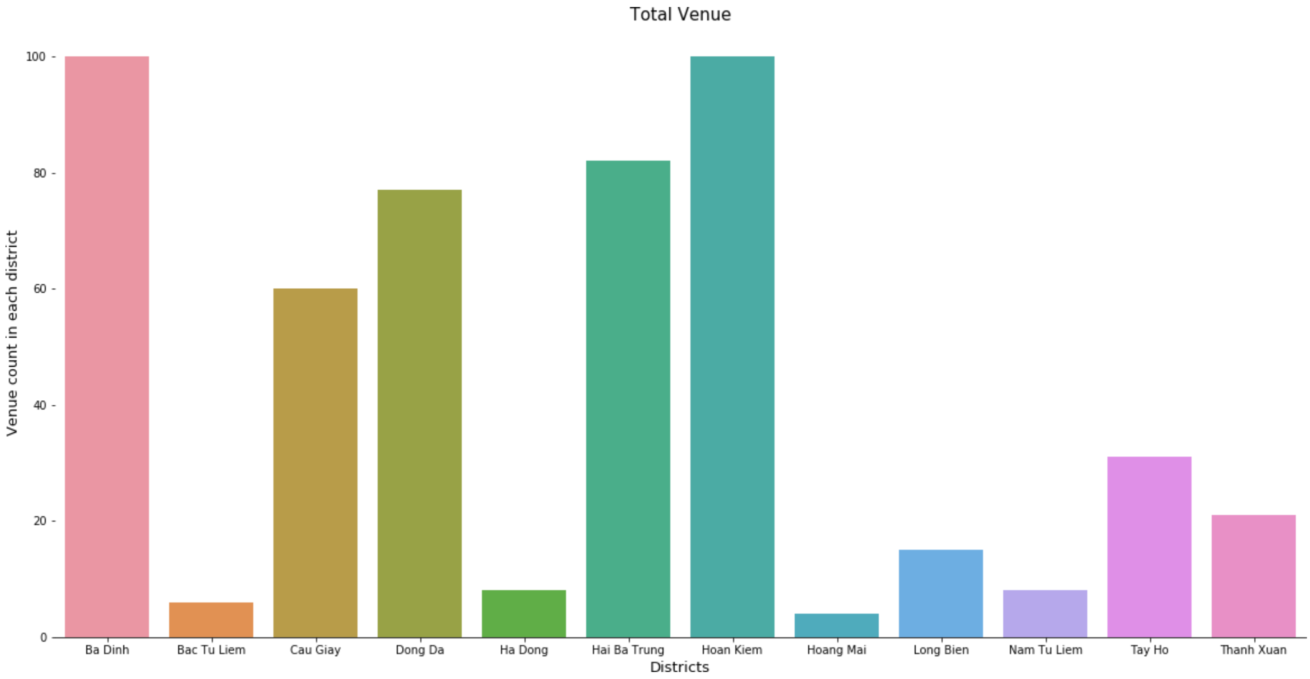
	Provincial Cities/Districts	Wards	Area (km2)	Population (2017)	Density (pop/km2)	Average Housing Price (1M VND/m2)	Latitude	Longitude
0	Ba Dinh	14	9224	247100	26.789	167	21.035549	105.824840
1	Bac Tu Liem	13	4335	333300	76.886	76.8	21.069861	105.757339
2	Cau Giay	8	1204	266800	221.595	155	21.027277	105.791464
3	Dong Da	21	996	420900	422.590	162	21.012920	105.827196
4	Hai Ba Trung	20	1009	318000	315.164	140	21.005970	105.857484
5	Ha Dong	17	47917	319800	6.674	76.8	20.952443	105.760955
6	Hoan Kiem	18	529	160600	303.592	442	21.028524	105.850716
7	Hoang Mai	14	4104	411500	100.268	81.5	20.974598	105.863707
8	Long Bien	14	6038	291900	48.344	73.3	21.037154	105.897839
9	Nam Tu Liem	10	3227	236700	73.350	85.2	21.012846	105.760874
10	Tay Ho	8	24	168300	7012.500	142	21.079042	105.815432
11	Thanh Xuan	11	911	285400	313.282	115	20.993687	105.814301

Venues per District

The bar chart in figure 4 below compares total number of venues per district, to see the level of dynamic in each district.

As can be seen from the chart, Ba Dinh and Hoan Kiem are the districts with most venues located. It can be inferred that these areas are the most dynamic spots in the city. Investors may pay attention to these areas as they are quite crowded with many businesses opened. Meanwhile, Bac Tu Liem, Ha Dong, Nam Tu Liem and Hoang Mai are not preferred by many owners as there are very few venues in these areas.

Figure 4: Total venue per districts

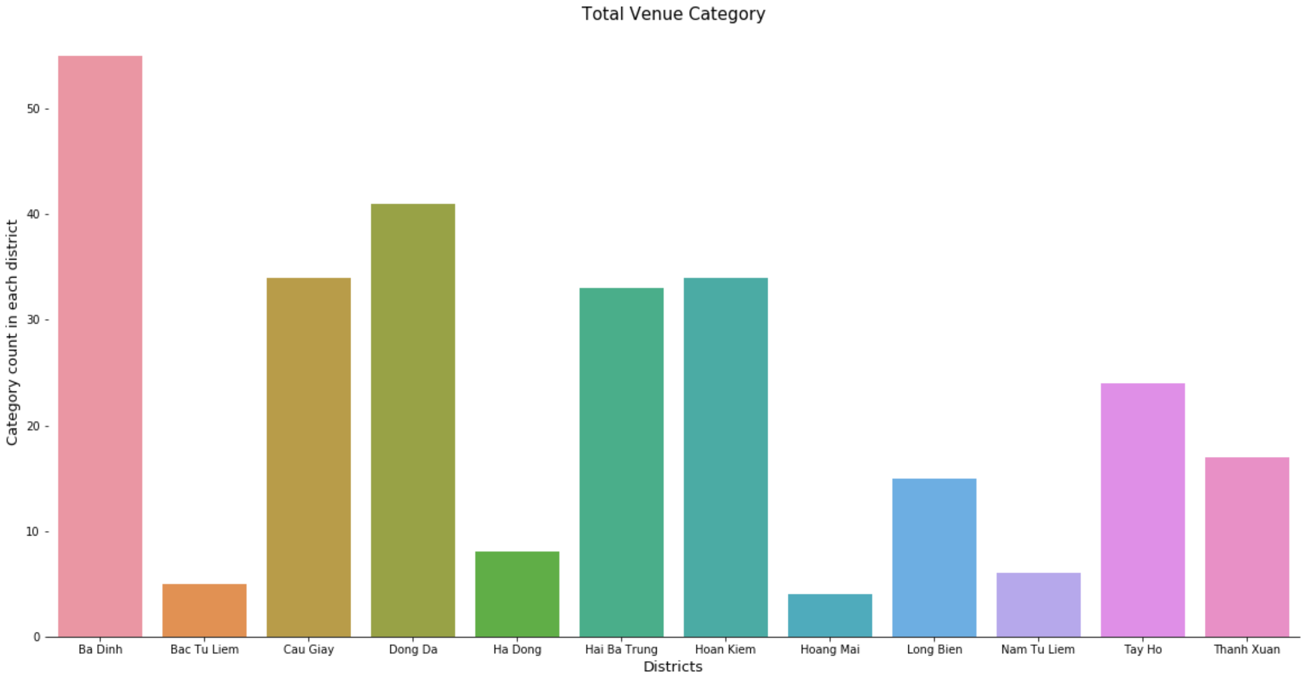


Categories per District

The bar chart in figure 5 below shows the total number of categories per district to see the level of business's diversification in each district.

Although Hoan Kiem is regarded as one of the most crowded areas, its unique business categories are not as many as others'. In other word, the businesses in this district are not really diversified. The reason behind this trend may be due to a fact that the district is commercially popular with some categories, therefore people tend to open those categories only. In contrast, Dong Da district is not a destination with comparable number of venues to Hoan Kiem, but the former is filled with different kinds of business. On top of that, Ba Dinh still remains as the most popular spot for opening businesses.

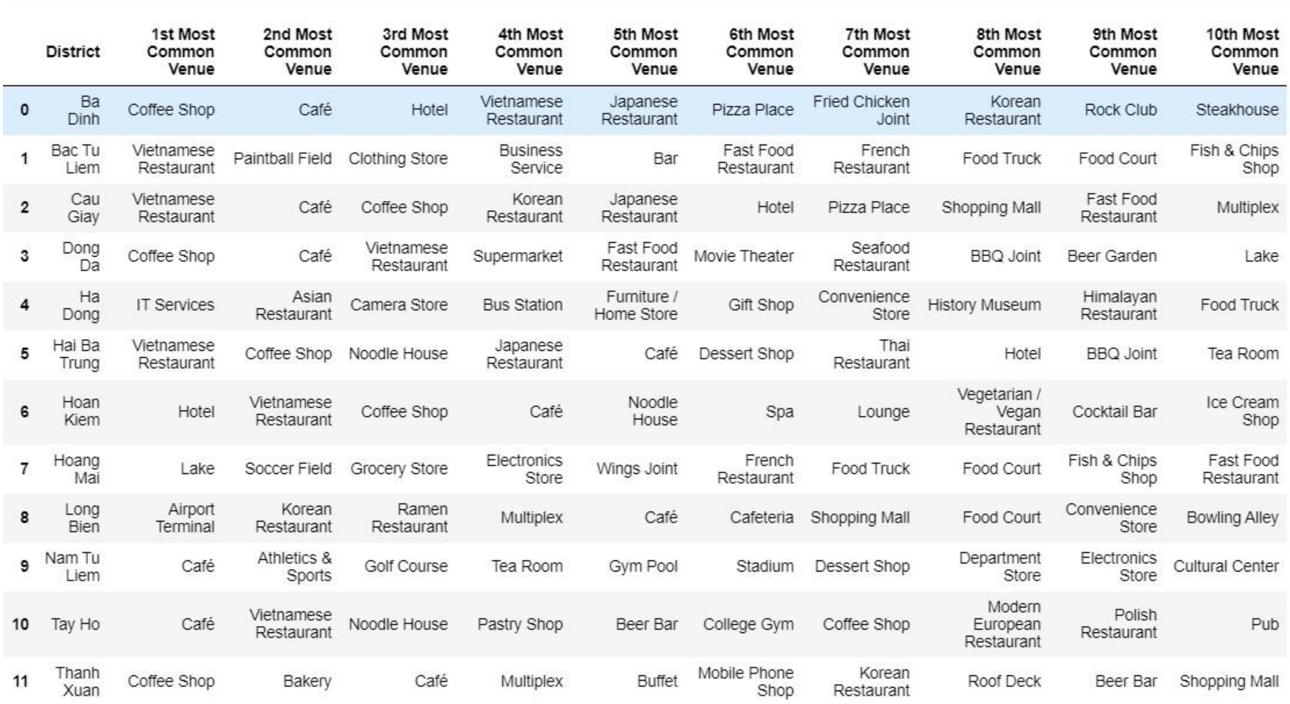
Figure 5: Total category per district



Top 10 venue categories in each district

Figure 6 demonstrates the most common 10 categories in each district. Later ~~on~~, we can use these categories to interpret some characteristics of different district clusters.

Figure 6: Top 10 categories in each district



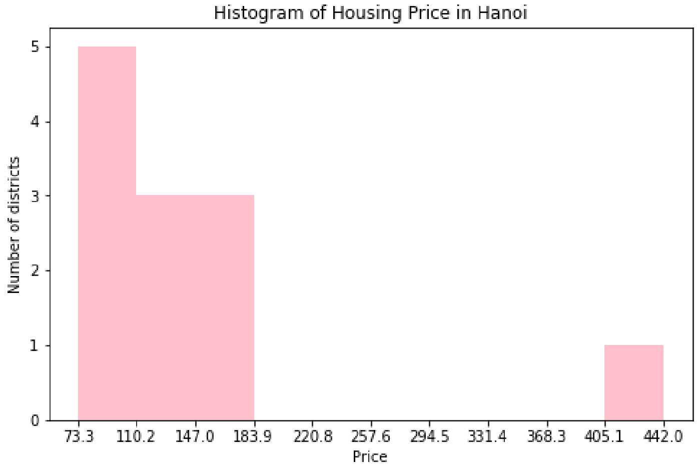
	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ba Dinh	Coffee Shop	Café	Hotel	Vietnamese Restaurant	Japanese Restaurant	Pizza Place	Fried Chicken Joint	Korean Restaurant	Rock Club	Steakhouse
1	Bac Tu Liem	Vietnamese Restaurant	Paintball Field	Clothing Store	Business Service	Bar	Fast Food Restaurant	French Restaurant	Food Truck	Food Court	Fish & Chips Shop
2	Cau Giay	Vietnamese Restaurant	Café	Coffee Shop	Korean Restaurant	Japanese Restaurant	Hotel	Pizza Place	Shopping Mall	Fast Food Restaurant	Multiplex
3	Dong Da	Coffee Shop	Café	Vietnamese Restaurant	Supermarket	Fast Food Restaurant	Movie Theater	Seafood Restaurant	BBQ Joint	Beer Garden	Lake
4	Ha Dong	IT Services	Asian Restaurant	Camera Store	Bus Station	Furniture / Home Store	Gift Shop	Convenience Store	History Museum	Himalayan Restaurant	Food Truck
5	Hai Ba Trung	Vietnamese Restaurant	Coffee Shop	Noodle House	Japanese Restaurant	Café	Dessert Shop	Thai Restaurant	Hotel	BBQ Joint	Tea Room
6	Hoan Kiem	Hotel	Vietnamese Restaurant	Coffee Shop	Café	Noodle House	Spa	Lounge	Vegetarian / Vegan Restaurant	Cocktail Bar	Ice Cream Shop
7	Hoang Mai	Lake	Soccer Field	Grocery Store	Electronics Store	Wings Joint	French Restaurant	Food Truck	Food Court	Fish & Chips Shop	Fast Food Restaurant
8	Long Bien	Airport Terminal	Korean Restaurant	Ramen Restaurant	Multiplex	Café	Cafeteria	Shopping Mall	Food Court	Convenience Store	Bowling Alley
9	Nam Tu Liem	Café	Athletics & Sports	Golf Course	Tea Room	Gym Pool	Stadium	Dessert Shop	Department Store	Electronics Store	Cultural Center
10	Tay Ho	Café	Vietnamese Restaurant	Noodle House	Pastry Shop	Beer Bar	College Gym	Coffee Shop	Modern European Restaurant	Polish Restaurant	Pub
11	Thanh Xuan	Coffee Shop	Bakery	Café	Multiplex	Buffet	Mobile Phone Shop	Korean Restaurant	Roof Deck	Beer Bar	Shopping Mall

Average price in each area

Looking at the range of housing price in Hanoi (Figure 7), we evaluate the values into 3 groups:

- **0 (Low)** : $Price \leq 110$.
- **1 (Medium)** : $110 < Price < 186$.
- **2 (High)** : $Price \geq 186$

Figure 7: Histogram of housing price in Hanoi



Density level in each area

Similarly, we also evaluate the density level of each district in Hanoi. We use density quantiles to classify these areas.

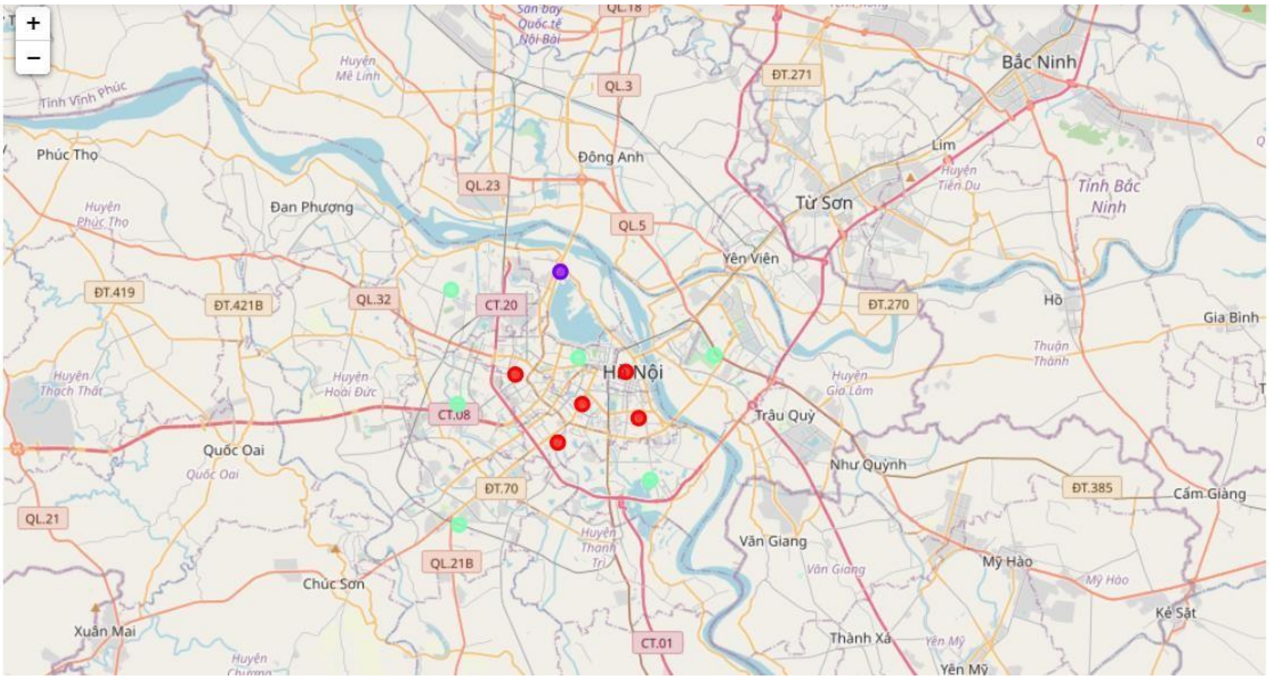
Therefore, if:

- **0 (Low):** Density ≤ Quantile(0.4):
- **1 (Medium):** Quantile(0.4) < Density ≤ Quantile(0.8)
- **2 (High):** Density > Quantile(0.8)

Choosing number of clusters

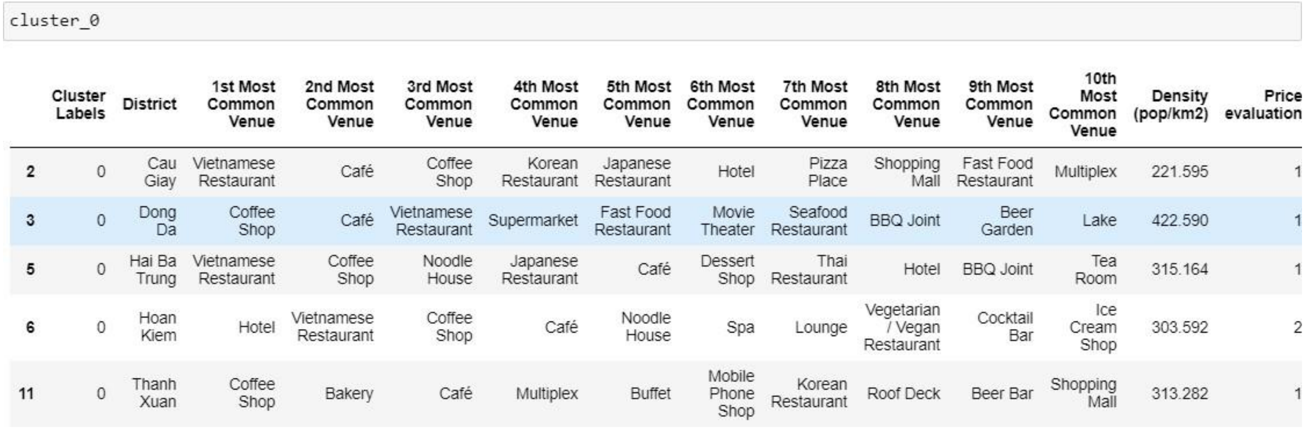
With 12 districts, we decide to cluster the districts into 3 groups (K for the K-means method) to observe its characteristics of these areas.

Figure 8: The maps of clusters. Cluster 0 (Red), Cluster 1 (Violet), Cluster 2 (Cyan).



Examine these clusters

- **Cluster 0**

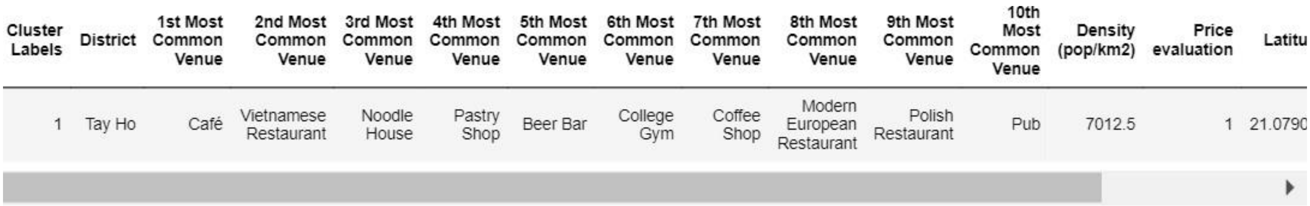


```
cluster_0
```

	Cluster Labels	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	Density (pop/km2)	Price evaluation
2	0	Cau Giay	Vietnamese Restaurant	Café	Coffee Shop	Korean Restaurant	Japanese Restaurant	Hotel	Pizza Place	Shopping Mall	Fast Food Restaurant	Multiplex	221.595	1
3	0	Dong Da	Coffee Shop	Café	Vietnamese Restaurant	Supermarket	Fast Food Restaurant	Movie Theater	Seafood Restaurant	BBQ Joint	Beer Garden	Lake	422.590	1
5	0	Hai Ba Trung	Vietnamese Restaurant	Coffee Shop	Noodle House	Japanese Restaurant	Café	Dessert Shop	Thai Restaurant	Hotel	BBQ Joint	Tea Room	315.164	1
6	0	Hoan Kiem	Hotel	Vietnamese Restaurant	Coffee Shop	Café	Noodle House	Spa	Lounge	Vegetarian / Vegan Restaurant	Cocktail Bar	Ice Cream Shop	303.592	2
11	0	Thanh Xuan	Coffee Shop	Bakery	Café	Multiplex	Buffet	Mobile Phone Shop	Korean Restaurant	Roof Deck	Beer Bar	Shopping Mall	313.282	1

- These districts are the central of entertainment in the city, with a lot of Coffee, Restaurants, or Supermarket, Shopping Mall and Multiplex
- The housing prices are from medium to high in these areas quite high population density
- Competitiveness in Cafe shop and restaurant is high

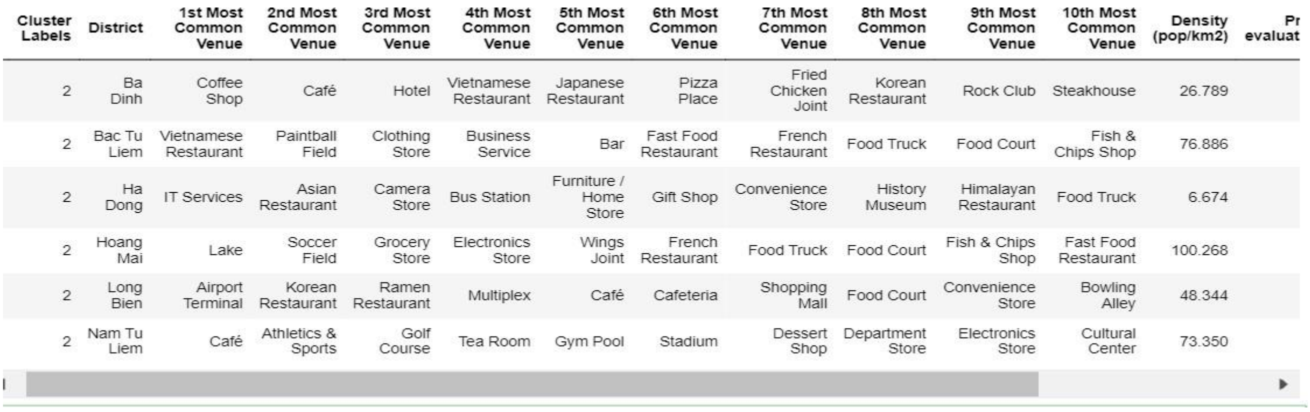
- **Cluster 1**



Cluster Labels	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	Density (pop/km2)	Price evaluation	Latitu
1	Tay Ho	Café	Vietnamese Restaurant	Noodle House	Pastry Shop	Beer Bar	College Gym	Coffee Shop	Modern European Restaurant	Polish Restaurant	Pub	7012.5	1	21.0790

- Although the density level is quite high in this area, it seems to be a green district as it has garden as its 10th most common venue.
- It is obviously the go-to place for foreigners as it has a lot of foreign restaurants and ~~also~~ pastry shops

- **Cluster 2**



Cluster Labels	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	Density (pop/km2)	Pr evaluat
2	Ba Dinh	Coffee Shop	Café	Hotel	Vietnamese Restaurant	Japanese Restaurant	Pizza Place	Fried Chicken Joint	Korean Restaurant	Rock Club	Steakhouse	26.789	
2	Bac Tu Liem	Vietnamese Restaurant	Paintball Field	Clothing Store	Business Service	Bar	Fast Food Restaurant	French Restaurant	Food Truck	Food Court	Fish & Chips Shop	76.886	
2	Ha Dong	IT Services	Asian Restaurant	Camera Store	Bus Station	Furniture / Home Store	Gift Shop	Convenience Store	History Museum	Himalayan Restaurant	Food Truck	6.674	
2	Hoang Mai	Lake	Soccer Field	Grocery Store	Electronics Store	Wings Joint	French Restaurant	Food Truck	Food Court	Fish & Chips Shop	Fast Food Restaurant	100.268	
2	Long Bien	Airport Terminal	Korean Restaurant	Ramen Restaurant	Multiplex	Café	Cafeteria	Shopping Mall	Food Court	Convenience Store	Bowling Alley	48.344	
2	Nam Tu Liem	Café	Athletics & Sports	Golf Course	Tea Room	Gym Pool	Stadium	Dessert Shop	Department Store	Electronics Store	Cultural Center	73.350	

- The density level and the prices in these areas are low, and they seem to be not the central of the city.
- As the density levels are low and the areas are quite big, these districts are the ideal spots for many outdoor sport activities such as Paintball Field, Soccer Stadium, Golf Course. Long Bien area is even cultivated for Airport.

5 Conclude

By consulting the overview of each district and its business picture, we can get the idea of what kind of business to invest in each area