# Opening a Coffee shop in Ho Chi Minh City

*Trinh Quoc Quang*

*July 14, 2020*

## 1. Introduction

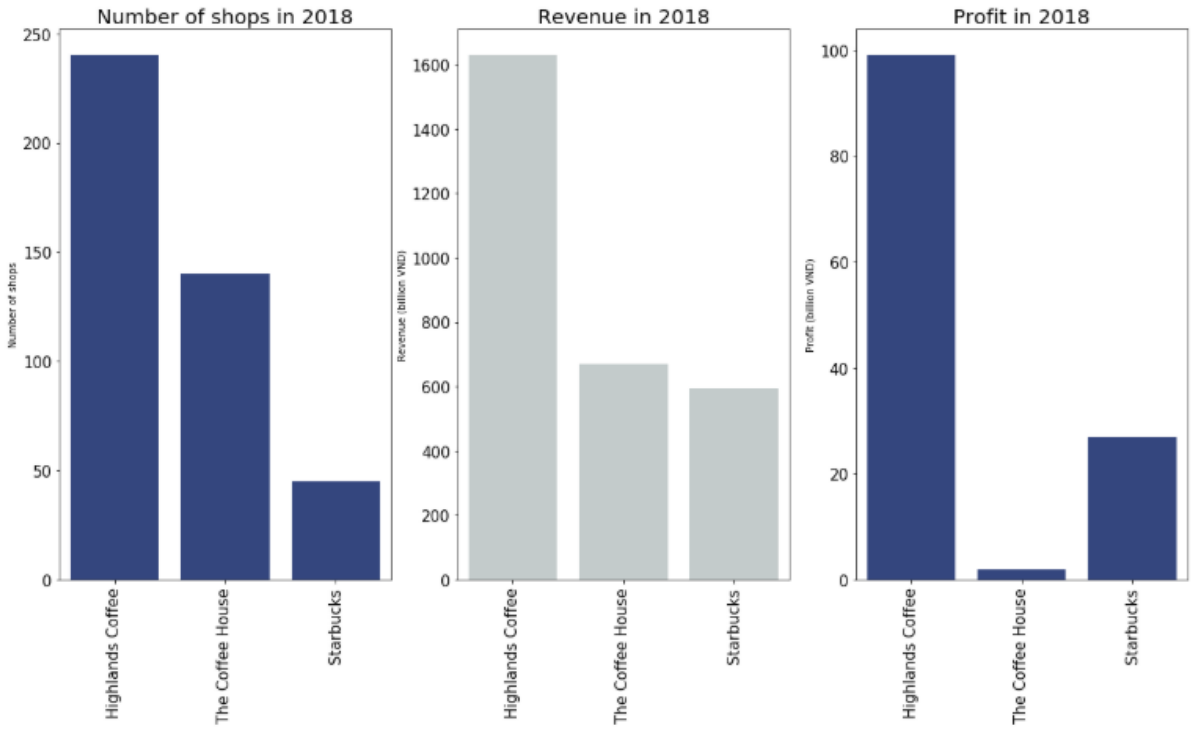
### 1.1. Background

Ho Chi Minh City, also known as Saigon, is the largest city in Vietnam. According to the 2019 census, Ho Chi Minh City has a population of over 8.9 million within city proper and over 21 million within the metropolitan area.

The catering services, therefore, has shown strong growth in the last decade, especially the coffee shop market. Thus, for many entrepreneurs, there has never been a better time to start their own coffee shops, one of the best business ideas in Vietnam.

According to VIRAC, a market analysis firm, Highlands Coffee led the market in 2018 in number of shops as 240 shops in Vietnam. The second position belonged to The Coffee House with 140 shops followed by Starbucks with 45 shops.

Regarding revenue and profit, Highlands Coffee made turnover of VND473 billion and post-tax profit of VND99 billion, The Coffee House VND669 billion and VND2 billion, Starbucks VND593 billion and VND27 billion.




Firstly, we should clarify the differences between a coffee shop and a café. Coffee shop and cafe are two different phrases that tend to be confused by many.

In simple terms, the line between cafe and coffee shop is actually coffee itself. Generally in a coffee shop, coffee is the main focus. In cafes, the main focus is on food rather than coffee, though most cafes will offer coffee pairings on their menus. On the other hand, because a coffee shop mostly deals with coffee, it does not have the qualities – like the main focus on food – that make it possible to be called a restaurant.

In this project, we suppose to work only on the coffee shops. Although there are already a lot of coffee shops in Ho Chi Minh City, their density between districts is not uniform. There are some districts containing too many coffee shops while there are less in some others.

### 1.2. Interests

If we have some knowledge about the population density, the housing price in each district and an overview of the number of coffee shops, we can have a better idea to set up a new business there.

If we think of it by an investor, we expect to choose a place where the population density is high but less competitors. If the housing price in that place is low, it is more attractive to us.

By using Data Science and some geometric factors about the relation between districts in HCMC, we can give good answers of following questions to the investors so that they can have a better vision about not only the coffee shops but also about other venues in HCMC.

i. *How many venues in each district?*
   Answering this question gives us a better understanding about the dynamic level of a district.

ii. *How many categories in each district?*
   Answering this question helps us know about the diversity in business of a district.

iii. *How many venues in each category?*
   This question shows the magnitude of a category in a district.

iv. *What are the most popular categories in each district?*
   If investors change their mind to focus on other commercial fields instead of opening a coffee shop.

v. *How many clusters we can use to categorize the districts based on the popularity of coffee shops?*

vi. *In which districts, the average housing price is low and the number of coffee shops is low also*?

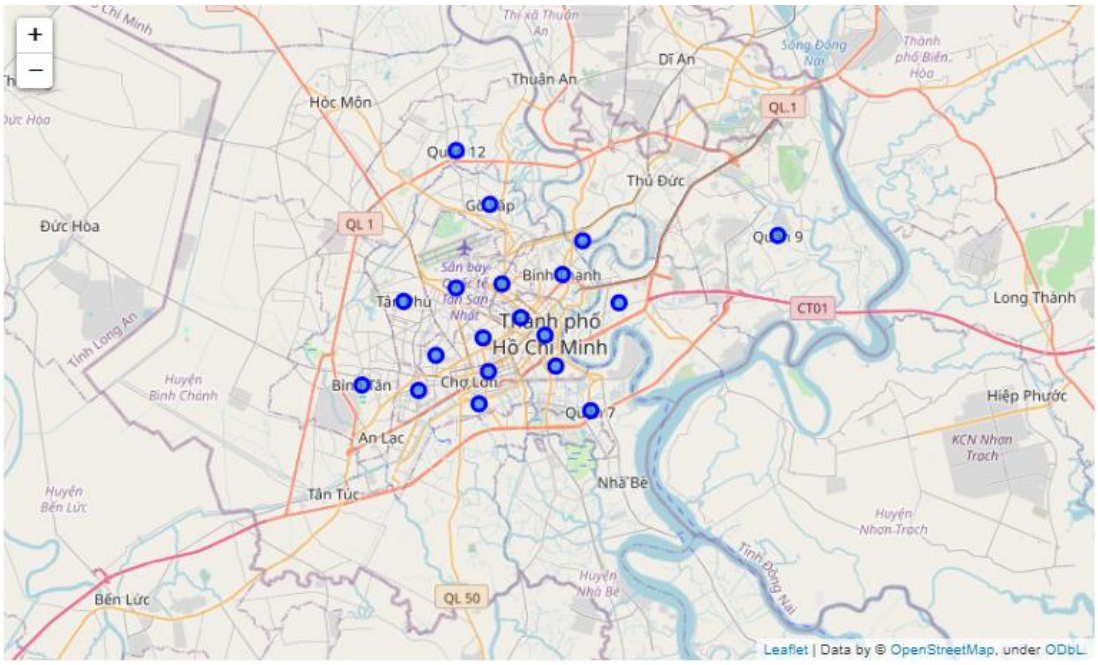
vii. *Where has high population but less coffee shop?*

viii. *Visualize all information on the map so that we can have a better look on what we want to find answer*

## 2. Data acquisition and cleaning

In order to explore these questions, we need to use following data in the research.

i. List of Ho Chi Minh City administrative units from Wikipedia. It gives us a list of all urban districts of HCMC with their area (in Km2), population (in 2015) and the density of each district (people/Km2). The list is given in https://en.wikipedia.org/wiki/Ho_Chi_Minh_City




ii. List of the coordinates (latitude, longitude) of all urban districts in HCMC. This list can be generated based on the name of each district and package geopy.geocoders.Nominatim.

iii. List of average housing prices per m2 in HCMC. The list is frequently updated in https://mogi.vn/gia-nha-dat

iv. A .json file contains all coordinates where we use it to create a choropleth map of Housing Sales Price Index of HCMC.

## 3. Methodology

i. First, we need to collect the data by scraping the table of HCMC units on the Wikipedia page and the average housing price (AHP) on a website. The BeautifulSoup package is very useful in this case.
ii. The column Density is calculated later based on columns Population and Area of each district.
iii. Throughout the project, we use numpy and pandas packages to manipulate dataframes.
iv. We use geopy.geocoders.Nominatim to get the coordinates of districts and add them to the main data frame.
v. We use folium package to visualize the HCMC map with its districts. The central coordinate of each district will be represented as a small circle on top of the city map.
vi. We use Foursquare API to explore the venues in each district and segment the districts based on them.
vii. For clustering the Coffee Shop venues between districts, we use K-Means Clustering method and the package scikit-learn will help us 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.
viii. In order to visualize the charts, we use package matplotlib.
ix. We use again the package folium to visualize the clusters on the main map and the choropleth map of AHP.

## 4. Results

We will answer all questions in the Section 1.2

### 4.1. The main data frame

After scraping all information from the internet, we have a table like in Figure 1

| | District | Subdistrict | Area (km2) | Population 2015 | Density (pop/m2) | Average Housing Price (1M VND/m2) | Latitude | Longitude |
|---|---|---|---|---|---|---|---|---|
| **0** | 1 | 10 wards | 7.73 | 193632 | 25049.418 | 427 | 10.774540 | 106.699184 |
| **1** | 2 | 11 wards | 49.74 | 147168 | 2958.745 | 71.9 | 10.791116 | 106.736729 |
| **2** | 3 | 14 wards | 4.92 | 196333 | 39905.081 | 268 | 10.783529 | 106.687098 |
| **3** | 4 | 15 wards | 4.18 | 186727 | 44671.531 | 107 | 10.759243 | 106.704890 |
| **4** | 5 | 15 wards | 4.27 | 178615 | 41830.211 | 254 | 10.756129 | 106.670376 |
| **5** | 6 | 14 wards | 7.19 | 258945 | 36014.604 | 113 | 10.746928 | 106.634495 |
| **6** | 7 | 10 wards | 35.69 | 310178 | 8690.894 | 89.1 | 10.736573 | 106.722432 |
| **7** | 8 | 16 wards | 19.18 | 431969 | 22521.846 | 70.9 | 10.740400 | 106.665843 |
| **8** | 9 | 13 wards | 114 | 290620 | 2549.298 | 48.5 | 10.824543 | 106.818015 |
| **9** | 10 | 15 wards | 5.72 | 238558 | 41705.944 | 211 | 10.773198 | 106.667833 |
| **10** | 11 | 16 wards | 5.14 | 230596 | 44863.035 | 167 | 10.764208 | 106.643282 |
| **11** | 12 | 11 wards | 52.78 | 510326 | 9668.928 | 46.9 | 10.867233 | 106.653930 |
| **12** | Go Vap | 16 wards | 19.74 | 634146 | 32124.924 | 97.4 | 10.840150 | 106.671083 |
| **13** | Tan Binh | 15 wards | 22.38 | 459029 | 20510.679 | 146 | 10.797979 | 106.653805 |
| **14** | Tan Phu | 11 wards | 16.06 | 464493 | 28922.354 | 104 | 10.791640 | 106.627302 |
| **15** | Binh Thanh | 20 wards | 20.76 | 487985 | 23506.021 | 133 | 10.804659 | 106.707848 |
| **16** | Phu Nhuan | 15 wards | 4.88 | 182477 | 37392.828 | 180 | 10.800118 | 106.677042 |
| **17** | Thu Duc | 12 wards | 49.76 | 528413 | 10619.232 | 64.4 | 10.822023 | 106.718302 |
| **18** | Binh Tan | 10 wards | 51.89 | 686474 | 13229.408 | 60.8 | 10.749809 | 106.605664 |

*Figure 1: The main data frame*

## 4.2. Venues per District

We plot a chart to compare visually the difference of number of venues between districts. This chart is shown in Figure 2.

From this chart, we see that the districts 1, 3, 5, 10, Phu Nhuan are the most dynamic ones. For the districts 1, 3 or 5, they are three center districts of HCMC, thus the high number of venues in these districts are as expected. Besides, districts 6, 9, 12, Binh Tan have poor number of venues due to their far distances from city center.

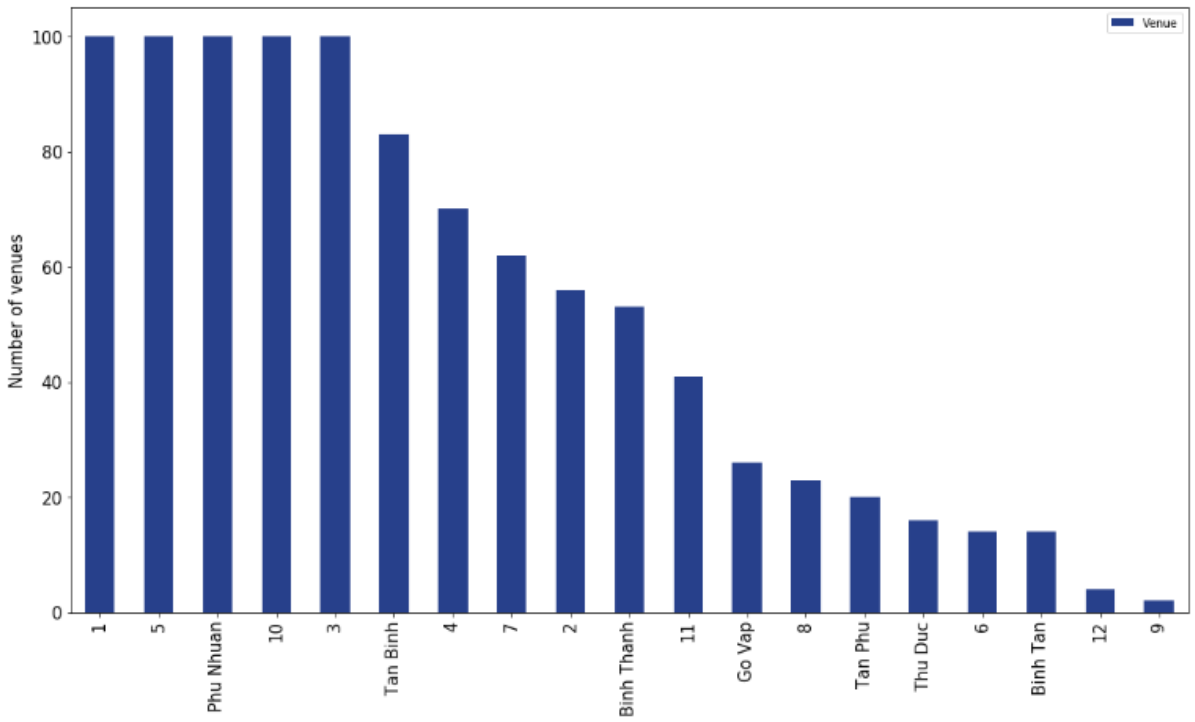



*Figure 2: Venues per district*

### 4.3. Categories per District

The chart in Figure 3 gives us an overview of the number of categories in each district.

Again, the district 1 wins the top. However, in this time, Binh Thanh district runs the second instead of district 3 or 5 like in Figure 2. The District 4 is still very diversity. The reason for that there are many venues but les categories in some districts is maybe there are some principle categories in these districts. Those principle categories play the major role in the commercial activities of these districts.

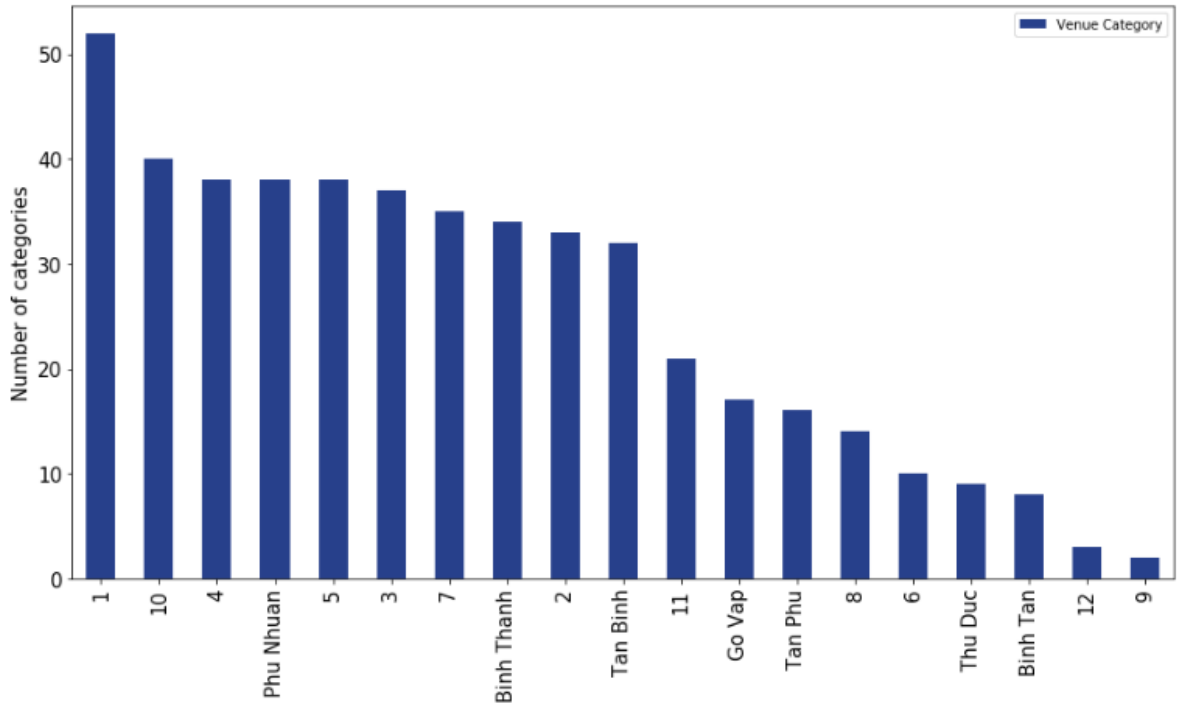



*Figure 3: Venue Categories per district*

| Venue Category | Venue |
|---|---|
| Vietnamese Restaurant | 137 |
| Café | 129 |
| Coffee Shop | 73 |
| Seafood Restaurant | 40 |
| Hotel | 30 |

### 4.4. Venues per Category

Look at the most 5 categories, we have Vietnamese Restaurant (137), Café (129), Coffee Shop (73), Seafood Restaurant (40), Hotel (30). The cafe and coffee shop are the main category in the beverage business with total 202 different venues.

### 4.5. Top 10 venue categories in each district

Figure 4 shows us the most 10 categories in each district. For less competition, we can choose districts whose fits most common venue is not coffee shop. For examples, districts 1, 10, 2, 3, 4, 5.

| | 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 | Hotel | Vietnamese Restaurant | Café | Coffee Shop | Vegetarian / Vegan Restaurant | Pizza Place | Hotel Bar | Massage Studio | Brewery | Hotpot Restaurant |
| 1 | 10 | Vietnamese Restaurant | Café | Coffee Shop | Dessert Shop | Hotpot Restaurant | Vegetarian / Vegan Restaurant | Seafood Restaurant | Spa | Hotel | Market |
| 2 | 11 | Café | Chinese Restaurant | Seafood Restaurant | Vietnamese Restaurant | Cantonese Restaurant | Fast Food Restaurant | Mobile Phone Shop | Theme Park | Dessert Shop | Shopping Mall |
| 3 | 12 | Café | Department Store | Vietnamese Restaurant | Film Studio | Food Truck | Food Court | Food | Flower Shop | Flea Market | Fast Food Restaurant |
| 4 | 2 | Vietnamese Restaurant | Restaurant | BBQ Joint | Café | Coffee Shop | Tea Room | Multiplex | Burger Joint | Bakery | Sandwich Place |
| 5 | 3 | Vietnamese Restaurant | Café | Coffee Shop | Vegetarian / Vegan Restaurant | Hotel | Asian Restaurant | Pizza Place | Noodle House | Spa | Breakfast Spot |
| 6 | 4 | Vietnamese Restaurant | Seafood Restaurant | Coffee Shop | Café | Snack Place | Food | Diner | Mexican Restaurant | Asian Restaurant | Hostel |
| 7 | 5 | Vietnamese Restaurant | Chinese Restaurant | Coffee Shop | Dim Sum Restaurant | Café | Dessert Shop | Noodle House | Asian Restaurant | Seafood Restaurant | BBQ Joint |
| 8 | 6 | Café | Supermarket | Dessert Shop | Fast Food Restaurant | Movie Theater | Breakfast Spot | Pizza Place | Bagel Shop | Vietnamese Restaurant | Department Store |
| 9 | 7 | Café | Vietnamese Restaurant | Coffee Shop | Seafood Restaurant | Gym / Fitness Center | Sushi Restaurant | Flea Market | Hotel | Multiplex | Pizza Place |
| 10 | 8 | Vietnamese Restaurant | Dim Sum Restaurant | Dessert Shop | Chinese Restaurant | Coffee Shop | Food Truck | Grocery Store | Café | Badminton Court | Plaza |
| 11 | 9 | Seafood Restaurant | Racetrack | Yoga Studio | Fast Food Restaurant | Food Court | Food | Flower Shop | Flea Market | Film Studio | Eastern European Restaurant |
| 12 | Binh Tan | Café | Coffee Shop | Shopping Mall | Multiplex | Pizza Place | Fast Food Restaurant | Japanese Restaurant | Asian Restaurant | Electronics Store | French Restaurant |
| 13 | Binh Thanh | Café | Coffee Shop | Soup Place | Vietnamese Restaurant | Asian Restaurant | Seafood Restaurant | Multiplex | Fast Food Restaurant | Comfort Food Restaurant | Residential Building (Apartment / Condo) |
| 14 | Go Vap | Café | Vietnamese Restaurant | Coffee Shop | Shopping Mall | Food Truck | Bookstore | Market | Pizza Place | Park | Electronics Store |
| 15 | Phu Nhuan | Café | Vietnamese Restaurant | Coffee Shop | Hotel | Vegetarian / Vegan Restaurant | Snack Place | Park | Spa | Gym / Fitness Center | Diner |
| 16 | Tan Binh | Vietnamese Restaurant | Café | Coffee Shop | Noodle House | Seafood Restaurant | Flea Market | Korean Restaurant | Karaoke Bar | Multiplex | Pizza Place |
| 17 | Tan Phu | Japanese Restaurant | Vietnamese Restaurant | Fast Food Restaurant | BBQ Joint | Supermarket | Ramen Restaurant | Diner | Coffee Shop | Shopping Mall | Shopping Plaza |
| 18 | Thu Duc | Vietnamese Restaurant | Café | Seafood Restaurant | Supermarket | Asian Restaurant | Venezuelan Restaurant | Coffee Shop | Public Art | Multiplex | Diner |

*Figure 4: Top 10 venue categories for each district*

### 4.6. How many clusters?

We consider the data relating to category "coffee shop" only. We want to cluster them into several groups. 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 5 shows that 4 is the best choice.

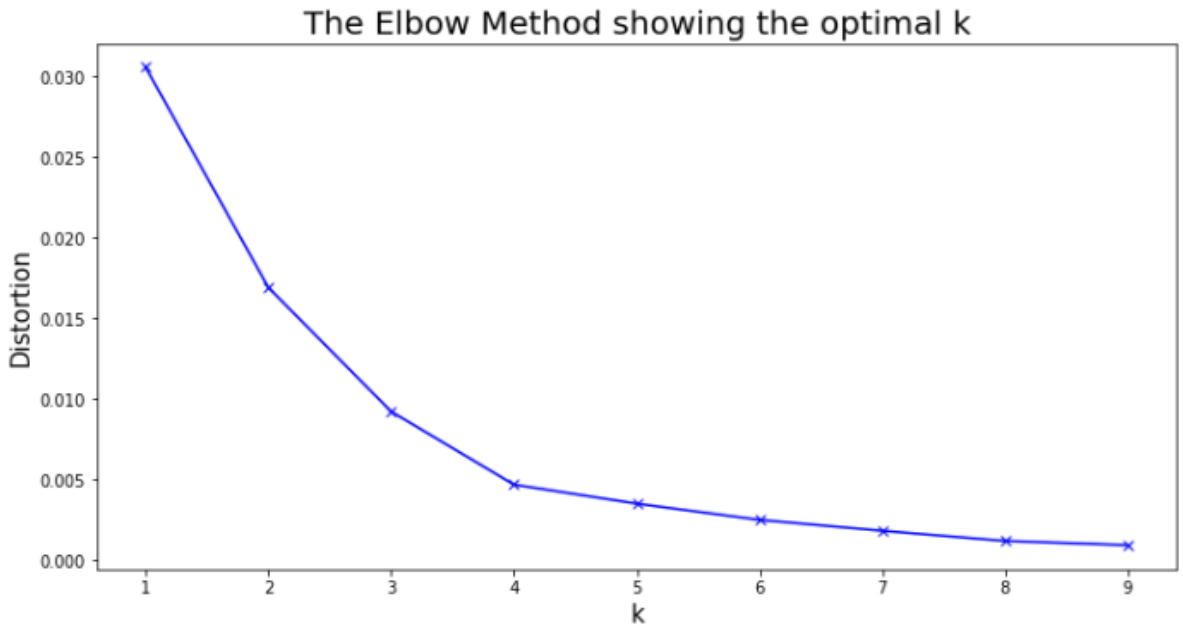



*Figure 5: The optimum number of groups/clusters*

We can define the clusters as below

- Cluster 0: There are a lot of coffee shops in these districts.
- Cluster 1: The number of coffee shops in these districts is medium..
- Cluster 2: There are no coffee shops in these districts
- Cluster 3: There are a few coffee shops in these districts

Figure 6 illustrates the clusters of all urban districts in HCMC. With this map, we can easily distinguish the clusters between districts.

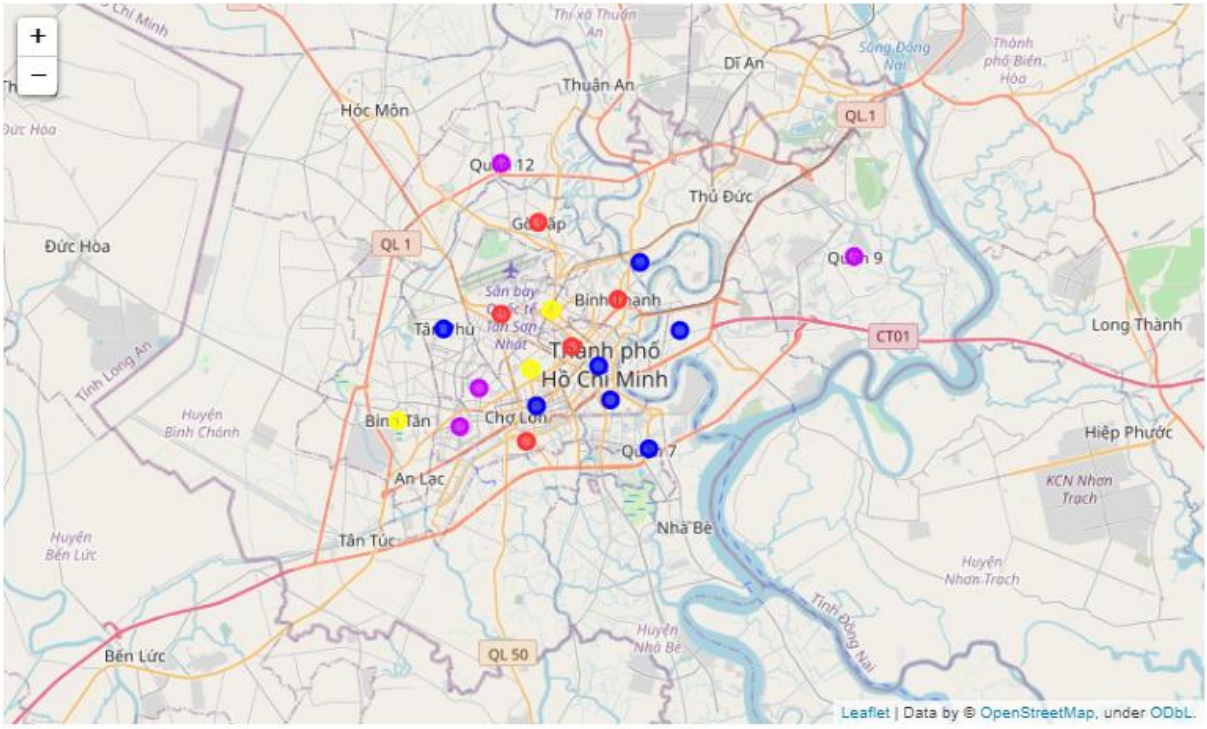



*Figure 6: The maps of clusters. Cluster 0 (Yellow), Cluster 1 (Red), Cluster 2 (Purple), Cluster 3 (Blue)*

### 4.7. AHP vs the number of coffee shops

Look back to the Average Housing Price (AVH) table, we categorize them into 4 groups (unit: million VND). Figure 7 indicates that the low housing price take the majority. We need to focus on the Low and Medium housing price to set up our business.

- Low: $30 < AHP \leq 100$.
- Medium: $100 < AHP \leq 200$.
- High: $200 < AHP \leq 300$.
- Very high: $300 \leq AHP$.

Look at Figure 8, we focus on:

- Low AHP & no coffee shops (cluster 2): district 9 and district 12.
- Low AHP & not many coffee shops (cluster 3): district 2, district 7, Tan Phu district and Thu Duc district.

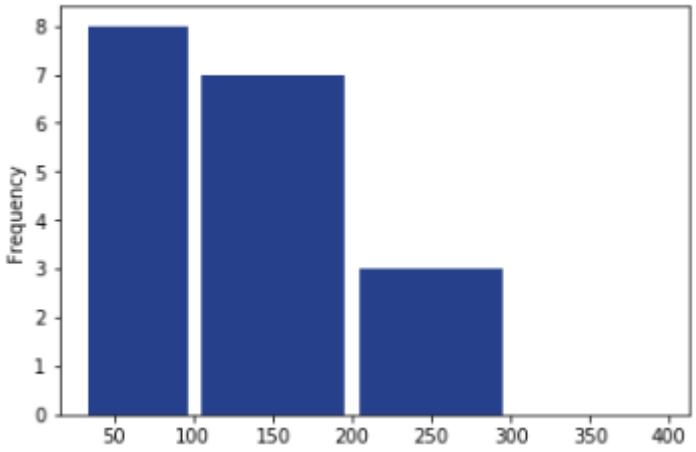



*Figure 7: The distribution of AHP*

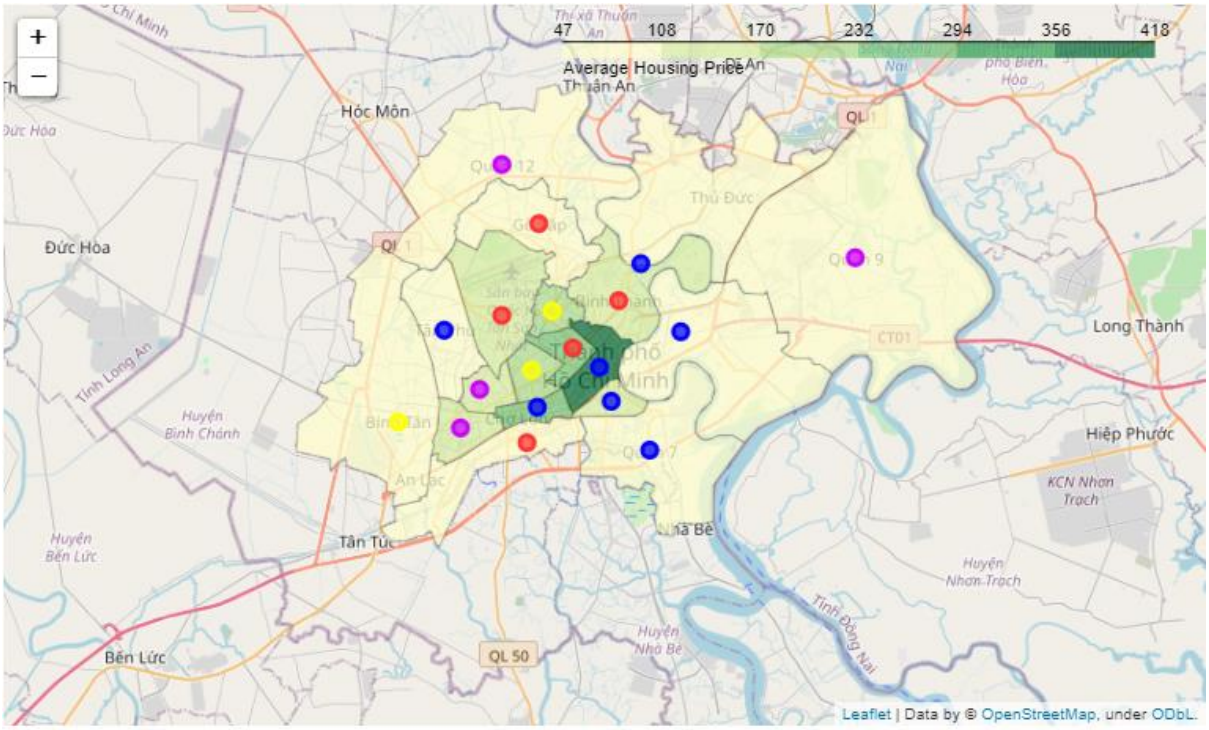



*Figure 8: The couple maps of AHP and the clusters given in Section 4.6*

### 4.8. Population density vs the number of coffee shop

We should not rely only on the relationship between AHP and clusters. For example, district 9 has almost no coffee shop and it has also very low AHP but actually this district contains many industry zones and there are not many people living around here. That is why we need to also consider the density of each district. Just think if there are not enough people to come to our coffee shop, how can we make a profit?

Figure 9 gives us a full picture about the relation between population density and the clusters.

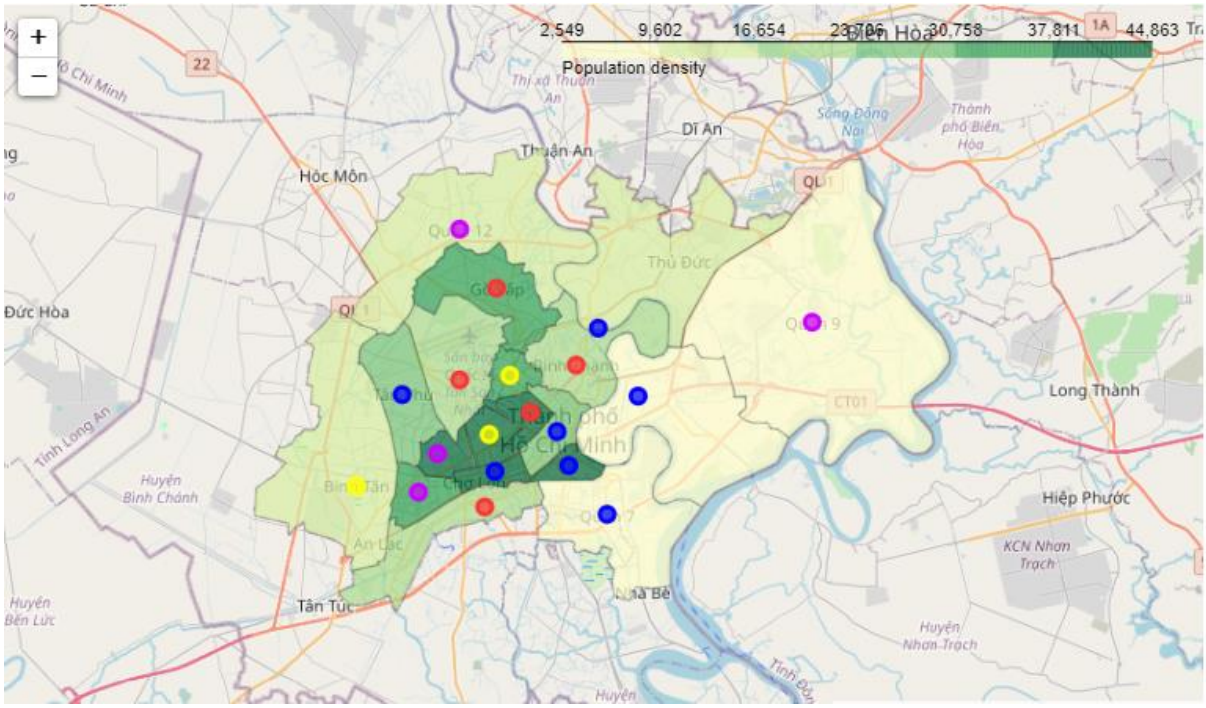



*Figure 9: The couple maps of population density and the clusters given in Section 4.6*

We focus on:

- High density + no coffee shops (cluster 2): no districts
- High density + not many coffee shops (cluster 3): district 1, district 4, district 5, district 6, district 11 and Tan Phu district

## 5. Conclusion

From all above results, we conclude that, the best place for us to set up a new coffee shop is in Tan Phu district because there are a lot of people living there (high density as 28,922.35 pop/m2), there are not many already-working coffee shop (cluster 3) and the average housing price is low (as 102 million VND/m2).