

Analyzing Venues of Ho Chi Minh City, Vietnam

Nhan Thanh Ngo

ngothanhnhnhan125@gmail.com

Introduction

Ho Chi Minh City is the largest city of Vietnam with more than 10 million in population. This city opens huge opportunity for not only Vietnamese but also foreign investors in launching their own business. In this article, I would like to show you the significant venue analysis including all such as restaurants, foods and beverage stores, clothing shops, fitness clubs, etc. applying for Ho Chi Minh City. Through the article, you could have overview of venue density, distribution, and how they are allocated in the city by map. The similarity of venue distribution among neighborhoods also are discussed.

The result of the study could be used to provide critical information to investors who considering starting-up, expanding their business in Ho Chi Minh City, Vietnam, or related research purpose as well.

Business Problem

Nowadays, the world is flatten, and investors in all over the world are standing in front of big chance to initialize or expand their business at any potential territories. The significant insights in the current business status at the place is definitely crucial. This is one of the main factors for investors in making better decision, rising ability to success of a business.

Among variety things that need to clarify, analyzing venues of a city play a vital role in choosing relevant products or services, and considering potential area to establish the business. Therefore, the study of analyzing venues of Ho Chi Minh City is implemented in order to find answer for the issue.

Data Description

The study needs to use the following data:

- Data of Ho Chi Minh City boroughs and neighborhoods.
- Data of neighborhood coordinate including latitude and longitude.
- Data of neighborhood venues in Ho Chi Minh City.

April 2019

Data of Ho Chi Minh City boroughs and neighborhoods

In Vietnam, the district and ward is used as the data for borough and neighborhood respectively. The Table of Ho Chi Minh City district and ward list could be found at the city government web page [1].

The BeautifulSoup package is used to get the relevant data inside html page content.

Table 1: Boroughs and Neighborhoods in Ho Chi Minh City, Vietnam

Index	District Code	District Name	Neighbor Code	Neighbor Name	Neighbor_dist
1	760	Quận 1	26734	Phường Tân Định	Phường Tân Định, Quận 1
2	760	Quận 1	26737	Phường Đa Kao	Phường Đa Kao, Quận 1
3	760	Quận 1	26740	Phường Bến Nghé	Phường Bến Nghé, Quận 1
4	760	Quận 1	26743	Phường Bến Thành	Phường Bến Thành, Quận 1
5	760	Quận 1	26746	Phường Nguyễn Thái Bình	Phường Nguyễn Thái Bình, Quận 1
6	760	Quận 1	26749	Phường Phạm Ngũ Lão	Phường Phạm Ngũ Lão, Quận 1
7	760	Quận 1	26752	Phường Cầu Ông Lãnh	Phường Cầu Ông Lãnh, Quận 1
8	760	Quận 1	26755	Phường Cô Giang	Phường Cô Giang, Quận 1
9	760	Quận 1	26758	Phường Nguyễn Cư Trinh	Phường Nguyễn Cư Trinh, Quận 1
10	760	Quận 1	26761	Phường Cầu Kho	Phường Cầu Kho, Quận 1
11	761	Quận 12	26764	Phường Thạnh Xuân	Phường Thạnh Xuân, Quận 12
12	761	Quận 12	26767	Phường Thạnh Lộc	Phường Thạnh Lộc, Quận 12

The results indicates that Ho Chi Minh City has 24 districts containing 322 neighborhoods. The number of neighborhoods in each district showed as below.

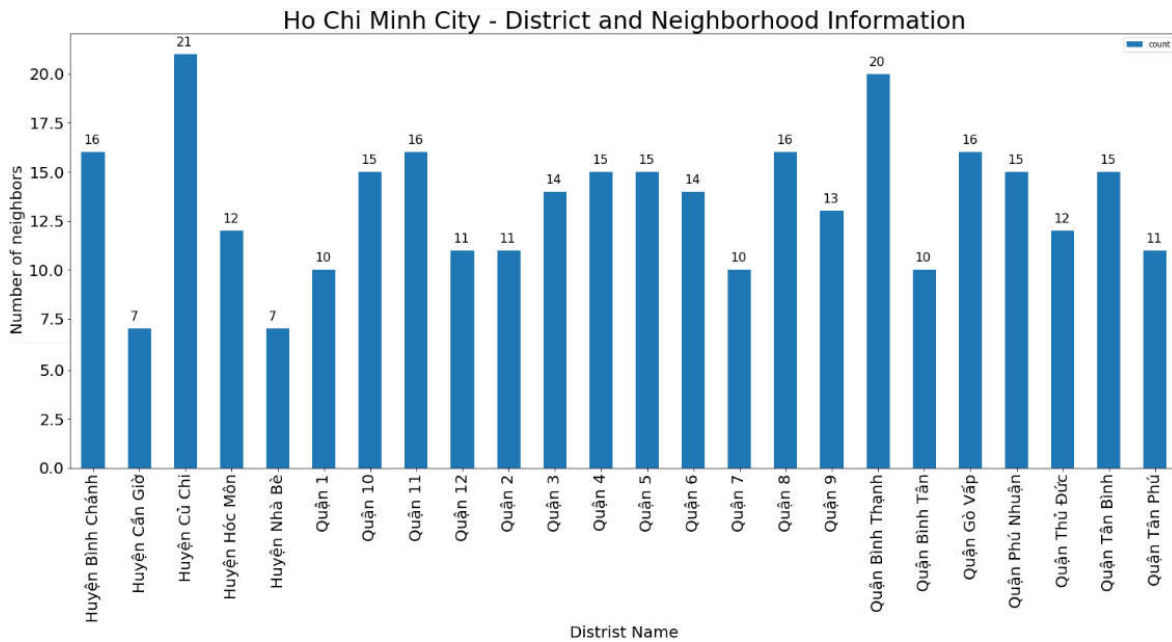


Figure 1: Number of neighborhoods in each District

April 2019

Data of neighborhood coordinate including latitude and longitude

After having the list of boroughs and neighborhoods of Ho Chi Minh City, the neighborhood coordinates including latitude and longitude are needed to get. This data will be used in venues request at the next step and display neighborhoods in the map.

The Google Map API is used to get the neighborhoods latitude and longitude. The Nominatim package also is known as the package used for get latitude and longitude of an address. However, Nominatim is really limited in working address. It is not high precision in comparison with Google Map. Therefore, Google Map API with JSON file reading [2] is used in this study to get higher accuracy of neighborhood coordinates.

Table 2: Coordinate (latitude, longitude) of Neighborhoods

Index	District Code	District Name	Neighbor Code	Neighbor Name	Neighbor_dist	latitude	longitude
1	760	Quận 1	26734	Phường Tân Định	Phường Tân Định, Quận 1	10.793097	106.690295
2	760	Quận 1	26737	Phường Đa Kao	Phường Đa Kao, Quận 1	10.787884	106.698403
3	760	Quận 1	26740	Phường Bến Nghé	Phường Bến Nghé, Quận 1	10.780833	106.702825
4	760	Quận 1	26743	Phường Bến Thành	Phường Bến Thành, Quận 1	10.773599	106.694417
5	760	Quận 1	26746	Phường Nguyễn Thái Bình	Phường Nguyễn Thái Bình, Quận 1	10.769385	106.700614
6	760	Quận 1	26749	Phường Phạm Ngũ Lão	Phường Phạm Ngũ Lão, Quận 1	10.765885	106.690810
7	760	Quận 1	26752	Phường Cầu Ông Lãnh	Phường Cầu Ông Lãnh, Quận 1	10.765545	106.696191
8	760	Quận 1	26755	Phường Cô Giang	Phường Cô Giang, Quận 1	10.761624	106.693243
9	760	Quận 1	26758	Phường Nguyễn Cư Trinh	Phường Nguyễn Cư Trinh, Quận 1	10.764030	106.686610
10	760	Quận 1	26761	Phường Cầu Kho	Phường Cầu Kho, Quận 1	10.757783	106.688821
11	761	Quận 12	26764	Phường Thạnh Xuân	Phường Thạnh Xuân, Quận 12	10.883430	106.670396

Data of neighborhood venues

FourSquare Place API [3] is chosen to get neighborhood venues. The EXPLORE method is used to get in FourSquare request command to get venues around the provided latitude, and longitude of neighborhood.

Table 3: Venue information getting from FourSquare

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Phường Tân Định, Quận 1	10.793097	106.690295	Cục Gạch	10.792957	106.689020	Vietnamese Restaurant
Phường Tân Định, Quận 1	10.793097	106.690295	Bánh canh cua 87	10.794697	106.690917	Vietnamese Restaurant
Phường Tân Định, Quận 1	10.793097	106.690295	Cục Gạch Quan	10.790773	106.691795	Vietnamese Restaurant
Phường Tân Định, Quận 1	10.793097	106.690295	Buddha Chay	10.792731	106.688208	Vegetarian / Vegan Restaurant
Phường Tân Định, Quận 1	10.793097	106.690295	Bánh Canh Hoàng Ty	10.791867	106.691560	Vietnamese Restaurant
Phường Tân Định, Quận 1	10.793097	106.690295	Cơm Tấm Nguyễn Phi Khanh	10.791676	106.692159	Breakfast Spot
Phường Tân Định, Quận 1	10.793097	106.690295	Tib Vegetarian	10.792703	106.689710	Vegetarian / Vegan Restaurant

The request statement for venues information from FourSquare is implemented for 322 neighborhoods, and a list of 3139 venues are returned. Those venues are dramatic concentration at the central areas of city, such as District 1, District 3, or Phu Nhuan district, accounts nearly 50% of total observed venues in Ho Chi Minh city.

April 2019

In the suburb areas, the amount of observed venues is quite low, for example, 14 districts from District 8 to Can Gio, the amount of venues accounts lower than 2 percentage of total ones for each.

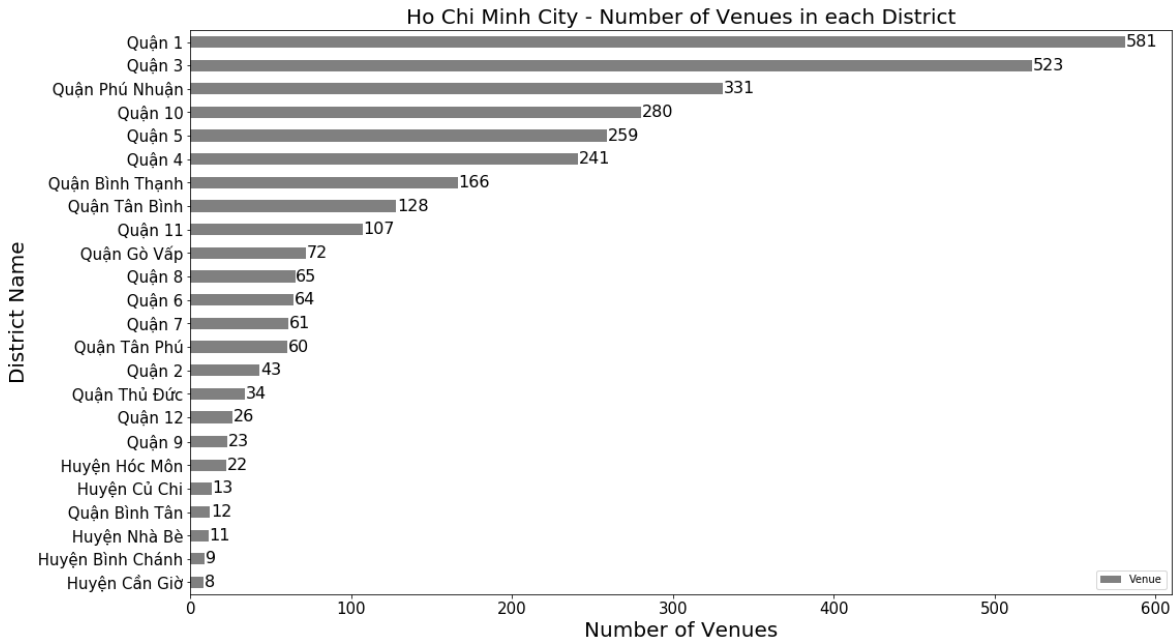


Figure 2: Number of Venues in each District

Exploring the venue category structure, it is straightforward to realize that the restaurant, café and coffee shop holds main stream among 190 different venue categories operating in the city.

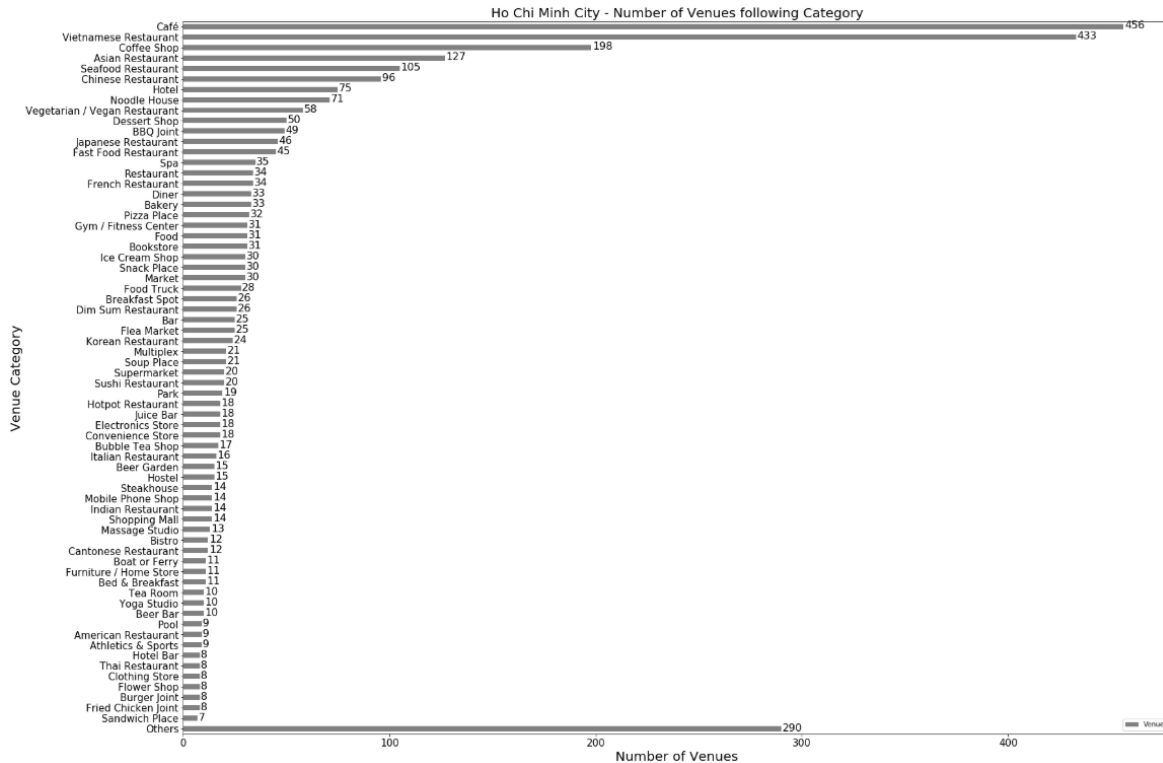


Figure 3: Overview of venue categories and number of venues inside

April 2019

Grouping venue categories, which have same types such as restaurant (Vietnamese Restaurant, Chinese Restaurant, Vegan Restaurant, etc.), coffee shop (café, coffee shop, cafeteria, etc.), other shopping mall, store, market into bigger groups, venue category structure could be shown as in below pie chart.

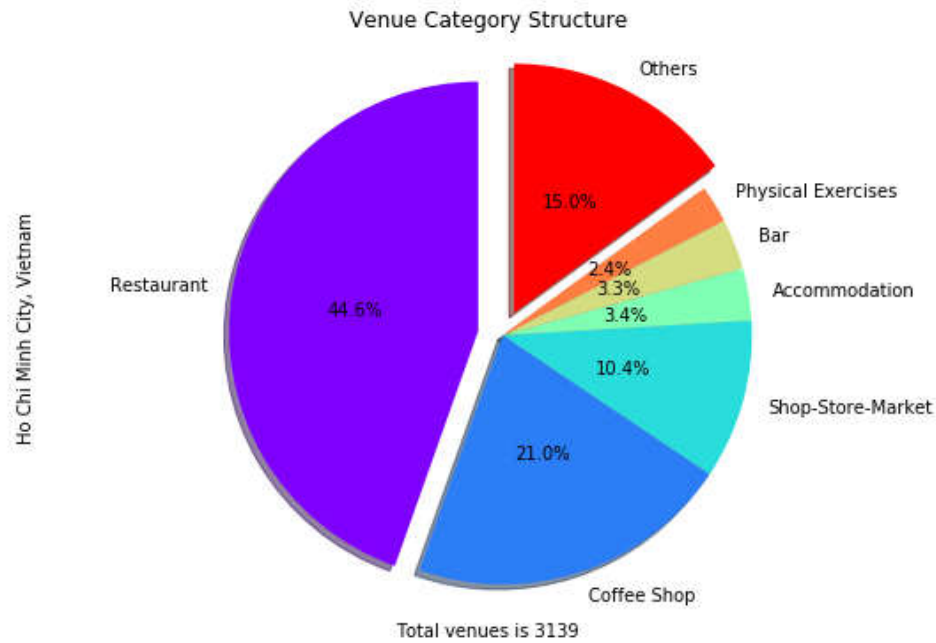


Figure 4: Venue category Structure of all venues in Ho Chi Minh City

Methodology

The venues belonging to a variety of categories could be distributed anywhere in the city. However, the question is if there are any area where is superior that the others in some kind of venues. The clustering method is applied for the response of this question.

Following that, quantity of venues in each category among neighbors are used to analyzed. This data need normalizing by mean method, as shown in *Table 4*, before inputting to Clustering model. This data also have meaning as the proportion of venue category in each neighborhood. The K-Means Clustering algorithm is applied to group neighborhood which are similar to other neighborhoods into cluster. The data need to be the clustering technique is implemented K-Means Clustering.

Table 4: Data Normalization for the appearance of Venue Categories following Neighborhood.

Neighborhood	Afghan Restaurant	Airport Food Court	American Restaurant	Argentinian Restaurant	Art Gallery	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	Athletics & Sports	...	University	Vegetarian / Vegan Restaurant	Video Game Store	Vietnamese Restaurant
Phường 01, Quận 10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.045455	0.000000	...	0.0	0.0	0.0	0.136364
Phường 01, Quận 11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.166667	0.000000	...	0.0	0.0	0.0	0.166667
Phường 01, Quận 3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.055556	0.000000	...	0.0	0.0	0.0	0.166667
Phường 01, Quận 4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.222222	0.111111	...	0.0	0.0	0.0	0.222222
Phường 01, Quận 5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.111111	0.000000	...	0.0	0.0	0.0	0.000000

April 2019

When implementing the K-Means Model, the number of clusters K needs specifying. The value of K which delivers high evaluation result for the model should be chosen. The Elbow method and Silhouette analysis are used to figure out the value of K cluster in this study.

Elbow method based on Sum of Squared Errors (SSE) between data points and the centroids of the cluster. Therefore, the K that provides smaller SSE value is better. However, the chosen K also is traded-off between SSE and number of clusters (K). Among K values which SSE is not so strong different with others, the small value of K is recommended. That value is usually in the elbow of chart shape.

The *Figure 5: K-Means Cluster Evaluation by Elbow Method* shows the relationship between Sum of Squared Errors equivalent with different K values. The chart have shape like an elbow. Following this chart, suitable value of K could be 3.

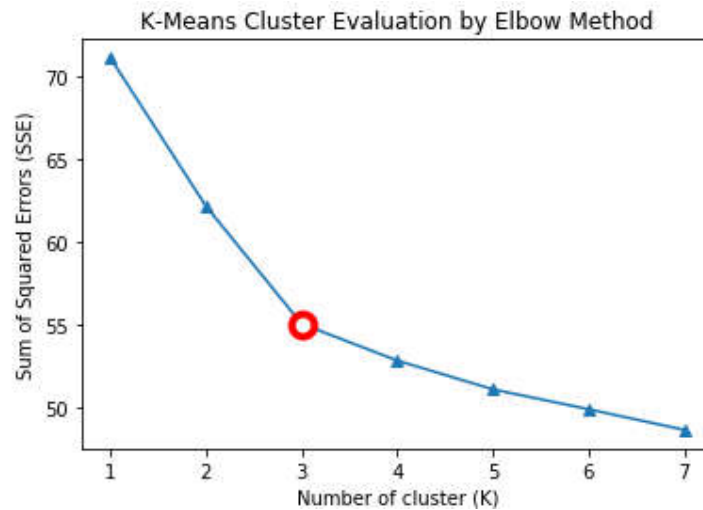


Figure 5: K-Means Cluster Evaluation by Elbow Method

To consolidate decision making for the value of K , the result of Silhouette analysis could be referred.

Silhouette is used to determine the degree of separation between clusters. Following that, Silhouette score closes to 1, this means the average distance from all data points in the same cluster is significant smaller than the average distance from all data points in the closest cluster.

Below, that is the result of Silhouette analysis for different values of K .

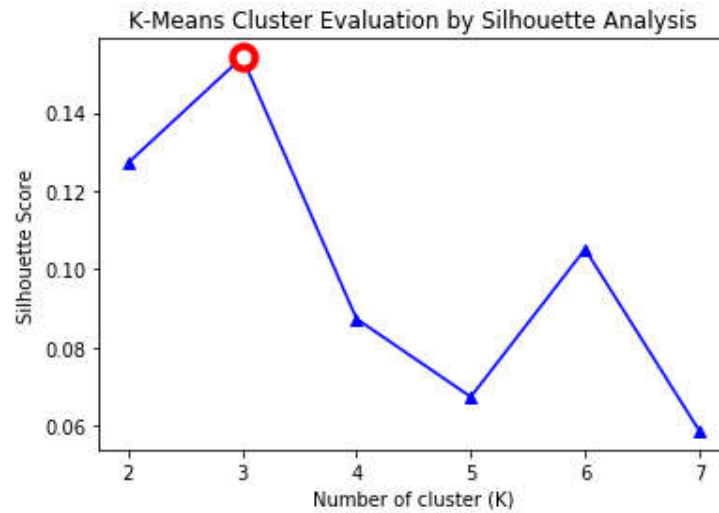


Figure 6: K-Means Cluster Evaluation by Silhouette Analysis

As shown in Figure 6: *K-Means Cluster Evaluation by Silhouette Analysis*, the values of Silhouette score are not high. They around 0.1 or less. This means that clusters, formed by K-Means model, do not dramatically separate. With the value of Silhouette score near 0, data points of clusters are overlapped each other. The result indicates that proportion of venue categories among neighbors are not so different. In another way, it could be said that the generated clusters have just a bit difference with each other.

Though the chart of Silhouette, the value of Silhouette when K equals 3 is really good in comparison with the larger K values.

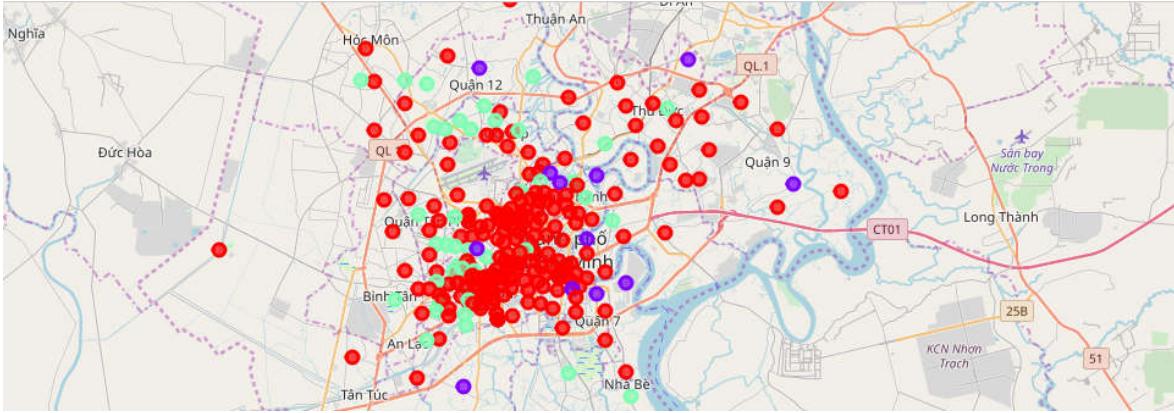
From Elbow model result and Silhouette analysis, the K equal 3 is chosen as the number of clusters in this study.

April 2019

Results

The K-means model (K=3) generates three clusters as shown in Figure 7.

(a) (Zoom In)



(b) (Zoom Out)

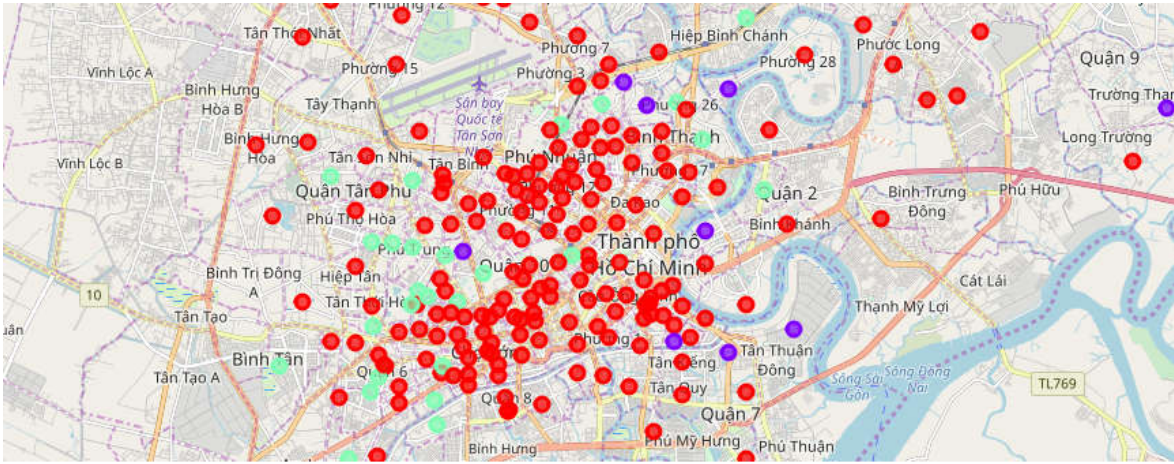


Figure 7: Allocation of Cluster neighborhoods in Map

There cluster performed by Cluster 0 (Red), Cluster 1 (Purple) and Cluster 2 (Aquamarine). The clustering result shows that there are 204 neighborhoods inside cluster 0, the number is 46 for Cluster 2, whereas there are only 16 neighborhoods for Cluster 1.

By computing proportion of each venue category at each neighborhood, the seven most common ones in each neighborhood are founded.

Table 5: Seven most common venue categories of each neighborhood (general)

Neighborhood	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
Phường Tân Định, Quận 1	Vietnamese Restaurant	Café	Coffee Shop	Breakfast Spot	Vegetarian / Vegan Restaurant	Asian Restaurant	Yoga Studio
Phường Đa Kao, Quận 1	Vietnamese Restaurant	Café	French Restaurant	Japanese Restaurant	Coffee Shop	Beer Garden	Vegetarian / Vegan Restaurant
Phường Bến Nghé, Quận 1	Coffee Shop	Hotel	Café	Spa	Vietnamese Restaurant	Cocktail Bar	Massage Studio
Phường Bến Thành, Quận 1	Hotel	Vietnamese Restaurant	Sandwich Place	Asian Restaurant	Food Court	Coffee Shop	Clothing Store
Phường Nguyễn Thái Bình, Quận 1	Vietnamese Restaurant	Café	Coffee Shop	Hotel	Burger Joint	Bistro	Italian Restaurant
Phường Phạm Ngũ Lão, Quận 1	Vietnamese Restaurant	Hotel	Hostel	Café	Coffee Shop	Indian Restaurant	Vegetarian / Vegan Restaurant
Phường Cầu Ông Lãnh, Quận 1	Vietnamese Restaurant	Hostel	Indian Restaurant	Hotel	Japanese Restaurant	Coffee Shop	Juice Bar

April 2019

Exploratory analysis is implemented for three clusters to increase the understanding on them.

- **Cluster 0 analysis**

Getting list of neighborhoods in cluster 0 follow with the data of most common venue categories to understand which kinds of venues have stronger chance in cluster 0.

Table 6: Seven most common venue categories of each neighborhood (Cluster 0)

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	Cluster Labels
Vietnamese Restaurant	Café	Coffee Shop	Breakfast Spot	Vegetarian / Vegan Restaurant	Asian Restaurant	Yoga Studio	0
Vietnamese Restaurant	Café	French Restaurant	Japanese Restaurant	Coffee Shop	Beer Garden	Vegetarian / Vegan Restaurant	0
Supermarket	Restaurant	Yoga Studio	Electronics Store	Food Stand	Food Court	Food & Drink Shop	0
Hotel	Vietnamese Restaurant	Park	Electronics Store	Food Stand	Food Court	Food & Drink Shop	0
Playground	Home Service	Café	Bakery	Yoga Studio	Ethiopian Restaurant	Food Stand	0
Beach	Yoga Studio	Fountain	Food Truck	Food Stand	Food Court	Food & Drink Shop	0
Vietnamese Restaurant	Flower Shop	Flea Market	Bookstore	Yoga Studio	Ethiopian Restaurant	Food Truck	0
Restaurant	Yoga Studio	Electronics Store	Food Stand	Food Court	Food & Drink Shop	Food	0
Clothing Store	Yoga Studio	Ethiopian Restaurant	Food Truck	Food Stand	Food Court	Food & Drink Shop	0
Mobile Phone Shop	Bubble Tea Shop	Seafood Restaurant	Café	Yoga Studio	Fast Food Restaurant	Food Stand	0
Bubble Tea Shop	Health Food Store	Restaurant	Noodle House	Yoga Studio	Ethiopian Restaurant	Food Court	0
Café	Korean Restaurant	Multiplex	Asian Restaurant	Coffee Shop	Restaurant	Shopping Mall	0
Stadium	Convention Center	Food Truck	Food Stand	Food Court	Food & Drink Shop	Food	0
Fountain	Yoga Studio	Ethiopian Restaurant	Food Truck	Food Stand	Food Court	Food & Drink Shop	0
Water Park	Athletics & Sports	Yoga Studio	Ethiopian Restaurant	Food Truck	Food Stand	Food Court	0
Vietnamese Restaurant	Korean Restaurant	Convenience Store	Coffee Shop	BBQ Joint	Electronics Store	Food Stand	0
Food Truck	Restaurant	Electronics Store	Food Stand	Food Court	Food & Drink Shop	Food	0

The result illustrates Cluster 0 is Multiple Social Venues group. A variety kind of categories appear as the most common venues. The further exploratory analysis is executed in just seven most common categories of each neighborhood for more information.

There are 2032 venues when consider seven most common categories of each Cluster 0 neighborhood. They have architecture as shown in the pie chart below.

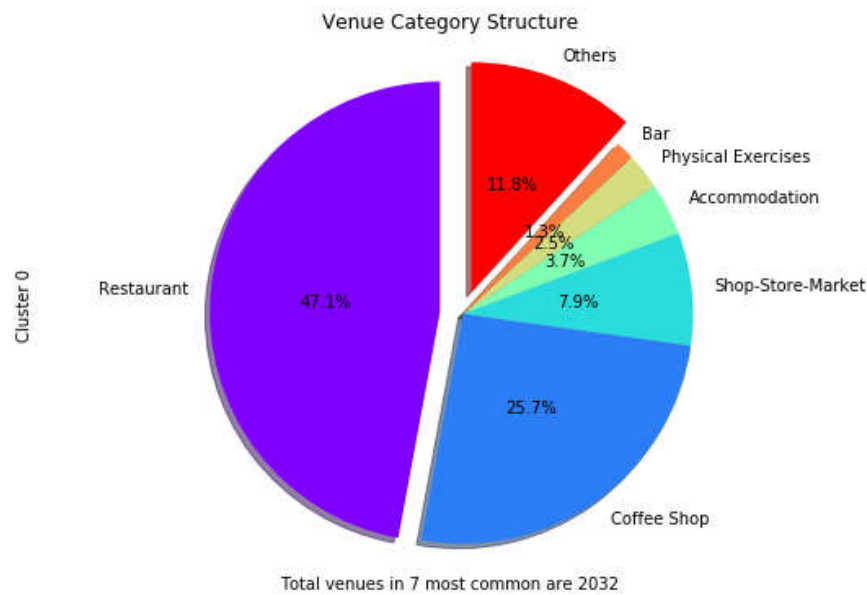


Figure 8: Venue Category Structure (Cluster0)

In the pie chart, the restaurant group performs main appearance with 47.1% amount of observed venues, following by Coffee Shop with 25.7%. The next four groups of Shop-store-Market,

April 2019

Accommodation, Physical Excises, Bar present with lower quantity, account 7.9%, 3.7%, 2.5% and 1.3% respectively.

Analyzing the restaurant categories, there is a diversification in this group with different cuisines. On the top of the table, Vietnamese restaurant hold first position with 40.3% in overall, following by Asian Restaurant, Chinese Restaurant, Seafood Restaurant with occupation of 9.4%, 8.9% and 8%.

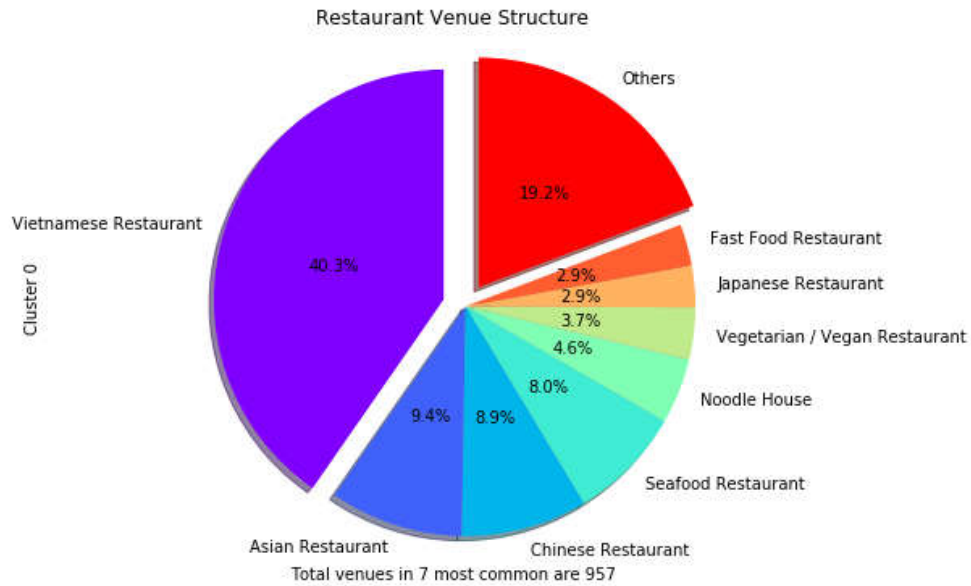


Figure 9: 1st common venue category structure (cluster0)

As shown in the geographic map (Figure 7), the allocation of neighborhoods belonging to Cluster 0 is clarified. The pie chart perform 4 central districts comprising District 3, District 1, Phu Nhuan District and District 5 occupy over a half of venues in cluster 0.

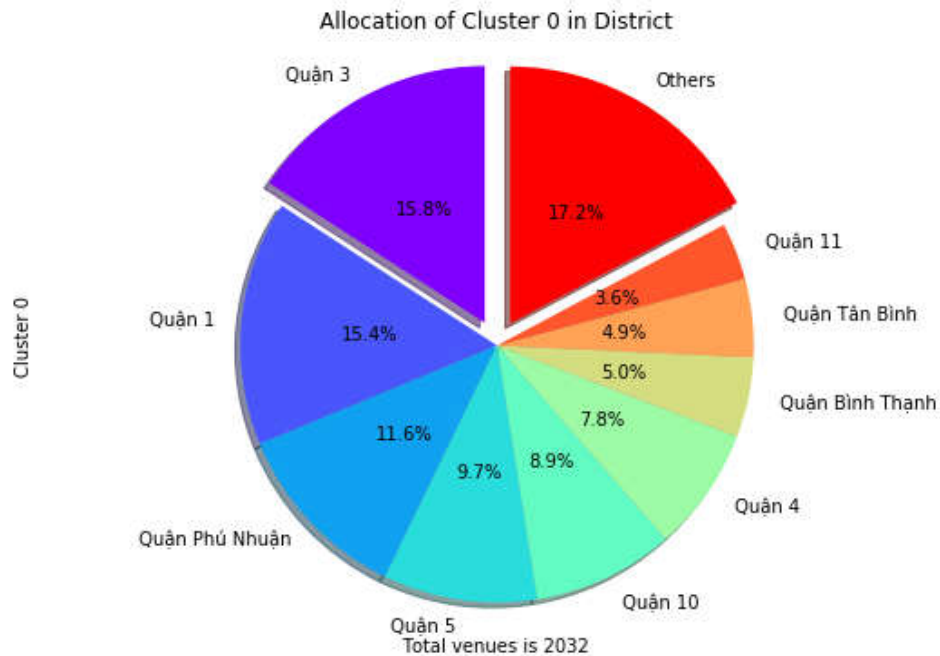


Figure 10: Allocation of Cluster 0 in Districts

April 2019

- **Cluster 1 analysis**

Implementing the same analysis as completed with Cluster 0, the 7 most common categories of each neighborhood is explored and considered. Then making the visualization of Cluster 1 venue category architecture, the first most common venue category structure, and how neighborhoods allocated in districts.

The table of 7 most common venue categories of each neighborhood of Cluster 1 shows the dominated of Restaurant group when in almost neighborhood, Vietnamese restaurant holds the first most common venue.

Table 7: Seven most common venue categories of each neighborhood (Cluster 1)

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	Cluster Labels
Vietnamese Restaurant	Health & Beauty Service	Café	Yoga Studio	Ethiopian Restaurant	Food Stand	Food Court	1
Vietnamese Restaurant	Diner	Yoga Studio	Electronics Store	Food Stand	Food Court	Food & Drink Shop	1
Vietnamese Restaurant	Furniture / Home Store	Yoga Studio	Electronics Store	Food Stand	Food Court	Food & Drink Shop	1
Vietnamese Restaurant	Beer Garden	Snack Place	Yoga Studio	Ethiopian Restaurant	Food Stand	Food Court	1
Vietnamese Restaurant	Seafood Restaurant	Asian Restaurant	Gym	Electronics Store	Food Stand	Food Court	1
Vietnamese Restaurant	Seafood Restaurant	Café	Yoga Studio	Ethiopian Restaurant	Food Stand	Food Court	1
Vietnamese Restaurant	Yoga Studio	Electronics Store	Food Stand	Food Court	Food & Drink Shop	Food	1
Vietnamese Restaurant	Yoga Studio	Electronics Store	Food Stand	Food Court	Food & Drink Shop	Food	1
Vietnamese Restaurant	Snack Place	Seafood Restaurant	Yoga Studio	Electronics Store	Food Court	Food & Drink Shop	1
Vietnamese Restaurant	Asian Restaurant	Yoga Studio	Ethiopian Restaurant	Food Truck	Food Stand	Food Court	1
Vietnamese Restaurant	Yoga Studio	Electronics Store	Food Stand	Food Court	Food & Drink Shop	Food	1

Analyzing by the number of venues, the pie chart of venue category structure indicates the leading role of restaurant group when it filled 83.8% of the circle. However, Cluster 1 is quite small in size with only 37 venues in the observed group.

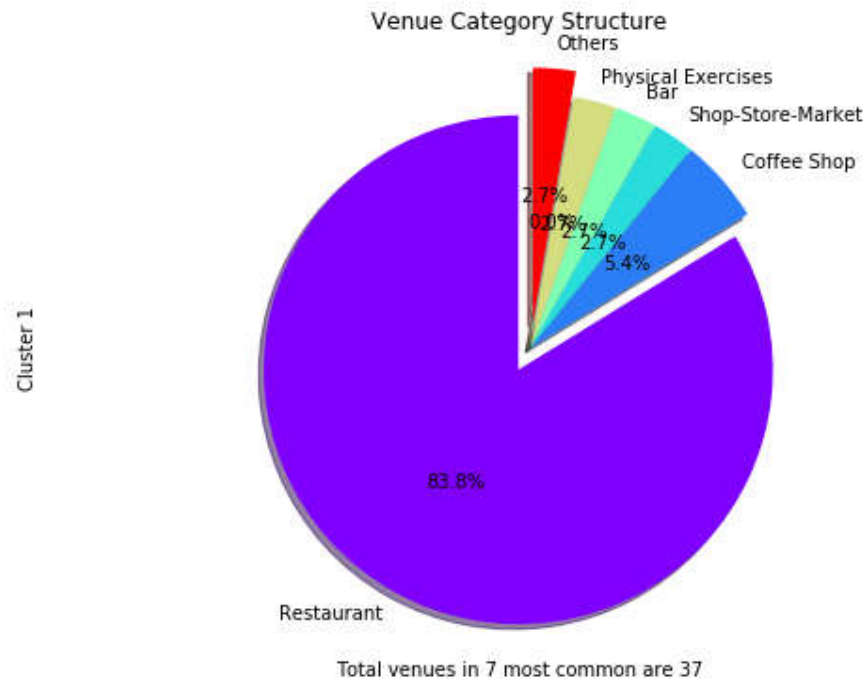


Figure 11: Venue Category Structure (Cluster1)

The restaurant group pie chart performs that Vietnamese Restaurants as role as local restaurants keep 71% overall.

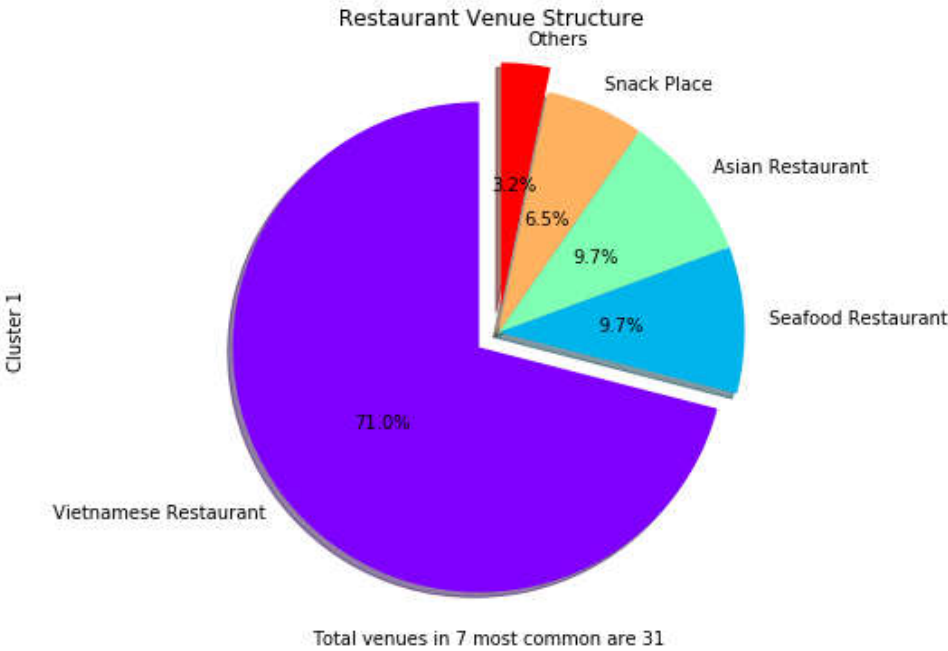


Figure 12: 1st common venue category structure (cluster1)

Different with Cluster 0, venues of Cluster 1 are distributed in special neighborhood, where are low citizen living, therefore low demand such as industrial zone, quiet suburban, and under construction area which low demand in services. Those are Binh Thanh District (Ward 13, Ward 27), District 7 (Tan Thuan Dong, Tuan Thuan Tay), District 4 (Ward 15), District 12 (Thoi An), etc.

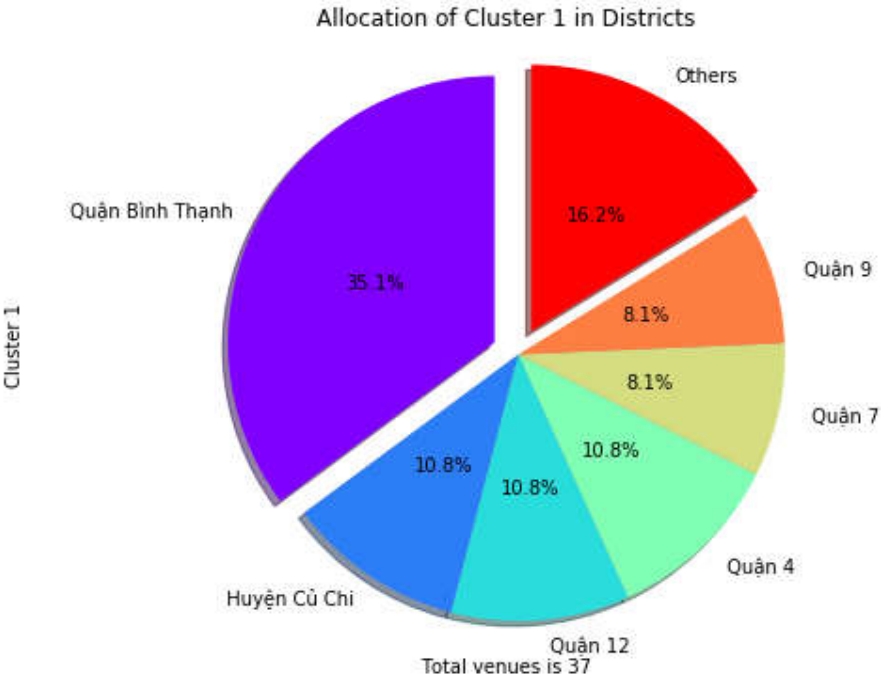


Figure 13: Allocation of Cluster 1 in Districts

April 2019

- **Cluster 2 analysis**

The Table of 7 most common venue categories of each neighborhood provide the first impression that Cluster 2 seem leading by Coffee shop.

Table 8: Seven most common venue categories of each neighborhood (Cluster 2)

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	Cluster Labels
Eastern European Restaurant	Café	Yoga Studio	Fast Food Restaurant	Food Truck	Food Stand	Food Court	2
Construction & Landscaping	Café	Yoga Studio	Fast Food Restaurant	Food Truck	Food Stand	Food Court	2
Café	Noodle House	Yoga Studio	Cruise	Food Stand	Food Court	Food & Drink Shop	2
Café	Vietnamese Restaurant	Asian Restaurant	Yoga Studio	Ethiopian Restaurant	Food Stand	Food Court	2
Café	Ramen Restaurant	Diner	Ethiopian Restaurant	Food Stand	Food Court	Food & Drink Shop	2
Food Truck	Garden	Café	Football Stadium	Food Stand	Food Court	Food & Drink Shop	2
Café	Seafood Restaurant	Yoga Studio	Ethiopian Restaurant	Food Stand	Food Court	Food & Drink Shop	2
Café	Mobile Phone Shop	Vietnamese Restaurant	Restaurant	Diner	Yoga Studio	Ethiopian Restaurant	2
Photography Studio	Café	Yoga Studio	Ethiopian Restaurant	Food Stand	Food Court	Food & Drink Shop	2
Café	Vietnamese Restaurant	Coffee Shop	Yoga Studio	Ethiopian Restaurant	Food Stand	Food Court	2
Café	Vietnamese Restaurant	BBQ Joint	Yoga Studio	Ethiopian Restaurant	Food Stand	Food Court	2
Café	Multiplex	Vietnamese Restaurant	Yoga Studio	Ethiopian Restaurant	Food Stand	Food Court	2

Analyzing to the structure of Cluster 2 venue categories, the result confirms for the role of café in this cluster. Following that, Coffee shop occupies 53.1% of total 213 observed venues. Standing at the second position in this cluster, restaurant group has 25.4%, following by Shop-store-market group with 6.6%.

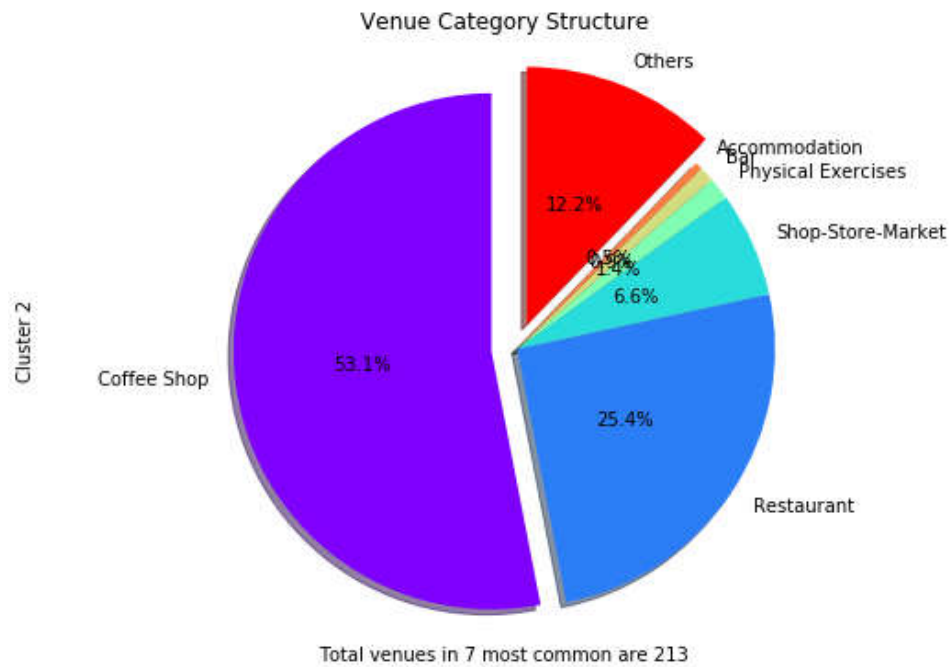


Figure 14: Venue Category Structure (Cluster2)

Below that is further information of which categories the coffee shop group includes.

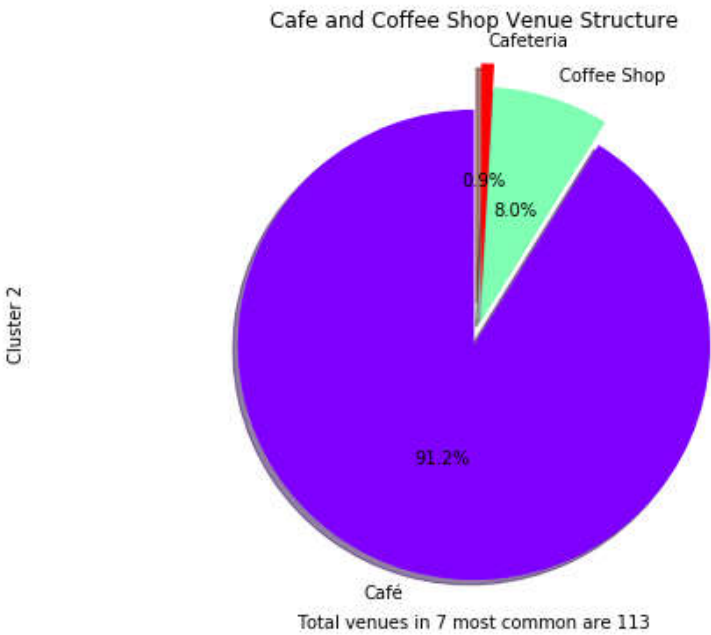


Figure 15: 1st common venue category structure (cluster2)

Venues of Cluster 2 allocates almost in 7 districts comprising District 11, Go Vap District, Tan Phu District, Binh Thanh District, District 3, Tan Binh District and District 6, account nearly three quarters of overall. This area is not in the center of city. Those district are crowded and high density of citizens.

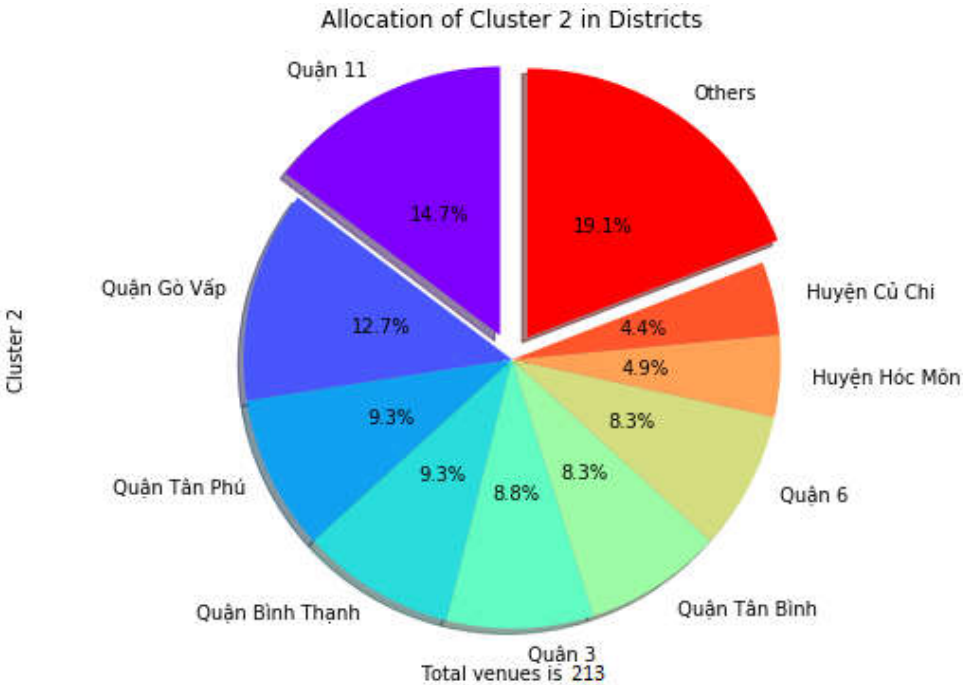


Figure 16: Allocation of Cluster 2 in Districts

Discussion

Based on the analysis result of three clusters, it could be discussed as below.

As illustrated in the folium geographic map, the neighborhoods belong to cluster 0 are located in the center such as District 1, District 3, Phu Nhuan District, and District 5, which hold greater than 50% amount of observed venues. The amount of observed venues in this cluster are also far larger than others. Therefore, it is diversification in types of venue categories including restaurant, coffee shop, store, market, physical exercise services, bar, etc.

The result could be explained by features of city center where has high concentration of office buildings, city government offices, large company representative office, tourism places, and are the busiest business area in the city. Cluster 0 neighborhoods therefore are advantageous for not only restaurant, coffee shop but also the other kinds of business which are appropriate to its market such as bar, club, entertainment activity, physical exercise services, etc.

Cluster 1 including really low number of neighborhoods, following by lower number of venues as well. Looking at the venue category structure pie chart, Vietnamese restaurant holds 83.8% overall. Analyzing the location of neighborhoods, it is discovered that the almost neighborhoods in cluster 1 locate in quiet area such as suburban, industrial zone, under construction area. This could be explained for low number of venues, and significant focusing on local restaurants.

The result shows the main stream venue category of cluster 2 is coffee shop when occupying 53.1% overall. Considering the allocation of neighborhoods, nearly 50% of total venues concentrate on District 11, Go Vap district, Tan Phu district, Binh Thanh district. Those districts have common point, are residential area, where have high density of citizens, high traffic, and young dynamic persons. This could be explained for high demand in coffee shop demand in comparison than others.

Through significant analysis of three clusters and discussion, relevant name for each cluster could be proposed as follows.

- Cluster 0 (Red): Restaurant Dominant in Multiple Social Venue Area– City Central
- Cluster 1 (Purple): Local Restaurant and Foods – Quiet Area
- Cluster 2 (Aquamarine): Intensive Coffee Venues – High density Resident Area

Conclusion

In the defined target and scope, the study provides necessary information and analysis for venues in Ho Chi Minh City, Vietnam. Through this study, audiences could have overview of distribution of different venue categories in the city. Not only investors, entrepreneurs but also business manager could use this information to improve the decision-making.

However, the study still have other directions to expand. The next research could do following directions such as:

- Analyzing venues with data of venues business class such as high class, middle class or reasonable one.
- Analyzing venues with data of venues traffic, amount of customer enter in correlation with its scale.
- Analyzing venues with data of venues rating, combining with venues traffic.
- Analyzing venues with data of price, traffic, rating.

Acknowledgement

By heart, I would like to express my thankful and respect to IBM instructor staff, who support me so much through IBM Data Science Professional Program with critical and absolutely beneficial knowledge, so I could complete the project.

References

- [1] Ho Chi Minh City districts and wards, Ho Chi Minh City Government web page,
<http://www.pso.hochiminhcity.gov.vn/web/guest/danhmucthongke-danhmucphuongxa>
- [2] Matthew Kudija, Instruction for using Google Map API,
<https://matthewkudija.com/blog/2018/11/19/google-maps-api/>
- [3] FourSquare, <https://developer.foursquare.com/places-api>
- [4] Nhan Thanh Ngo, Analyzing Venues of Ho Chi Minh City, Vietnam,
https://github.com/carfirst125/coursera_capstone_ibm_ws/blob/master/Analyzing%20Venues%20of%20Ho%20Chi%20Minh%20City%2C%20Vietnam.ipynb