

**An analysis of Ho Chi Minh City Neighborhoods**

**Anh Ngo Hai**

**December 2019**

## 1. Introduction and Objective

Covering over 2000 square kilometers, with a population of 8.4 million, Ho Chi Minh City is Vietnam's largest city.

Considering its diverse and fast-paced economic development, HCMC is without a doubt one of the best place in Vietnam for budding entrepreneurs to carve out a niche of their own. On the other hand, it is also highly competitive environment.

As a side business, I am looking into opening a restaurant or a café, and would like to find a spot with fewer competitions. Hopefully I can get some insight with this data analysis.

## 2. Data acquisition and cleaning

### Data acquisition

#### HCMC Neighborhoods Data

- List of districts and wards

<https://www.gso.gov.vn/dmhc2015/Default.aspx>.

- Data file

[https://github.com/anhnhvn/first-github-repository/blob/master/HCMC\\_wards.csv](https://github.com/anhnhvn/first-github-repository/blob/master/HCMC_wards.csv)

#### Google map API

Google Map API Geocoder to get the Latitude and Longitude of each area.

#### Foursquare API

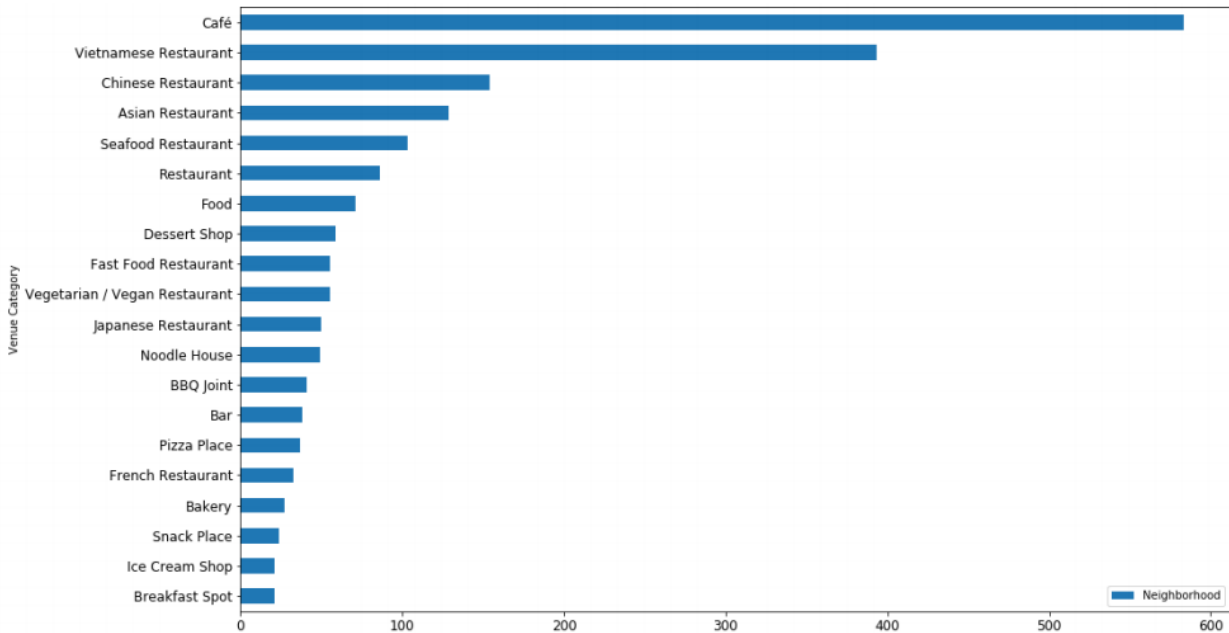
Four-square API is the project's main data gathering source. This API enables users to perform location search, location sharing and details about a business

### Cleaning Downloaded data

As the excel file includes all districts and wards of Vietnam, I used pandas function to narrow the data frame to contain only the districts and wards of Ho Chi Minh City, and only the inner city area.

Next I used Google API to get the longitude and latitude of each ward, to serve as input to the nearby Foursquare API search. Afterwards, combined the longitude and latitude obtained with the districts and wards data. I then used the Foursquare API to search and continued to merge it into district and ward data. The Foursquare API returns all the

venues, so to get only food and beverage service venues (bar, coffee, restaurant, food, wine, ice cream ...). I used the Pandas function to filter the result. I used Pandas to make horizontal bar of top 20 Venues:



### 3. Methodology

#### 3.1. K-Means Clustering

This project applies K-mean clustering unsupervised machine learning algorithm to cluster the venues based on a list of locations for different types of food and beverage service points such as bars, cafes, Chinese restaurants, Vietnamese restaurants, Seafood restaurants etc. This would give a better understanding of the similarities and dissimilarities between the chosen neighborhoods to retrieve more insights and to conclude with ease which neighborhood wins over other.

Analyze Each Neighborhood, group rows by neighborhood and by taking the mean of the frequency of occurrence of each category. Next create the new dataframe and display the top 8 venues for each neighborhood.

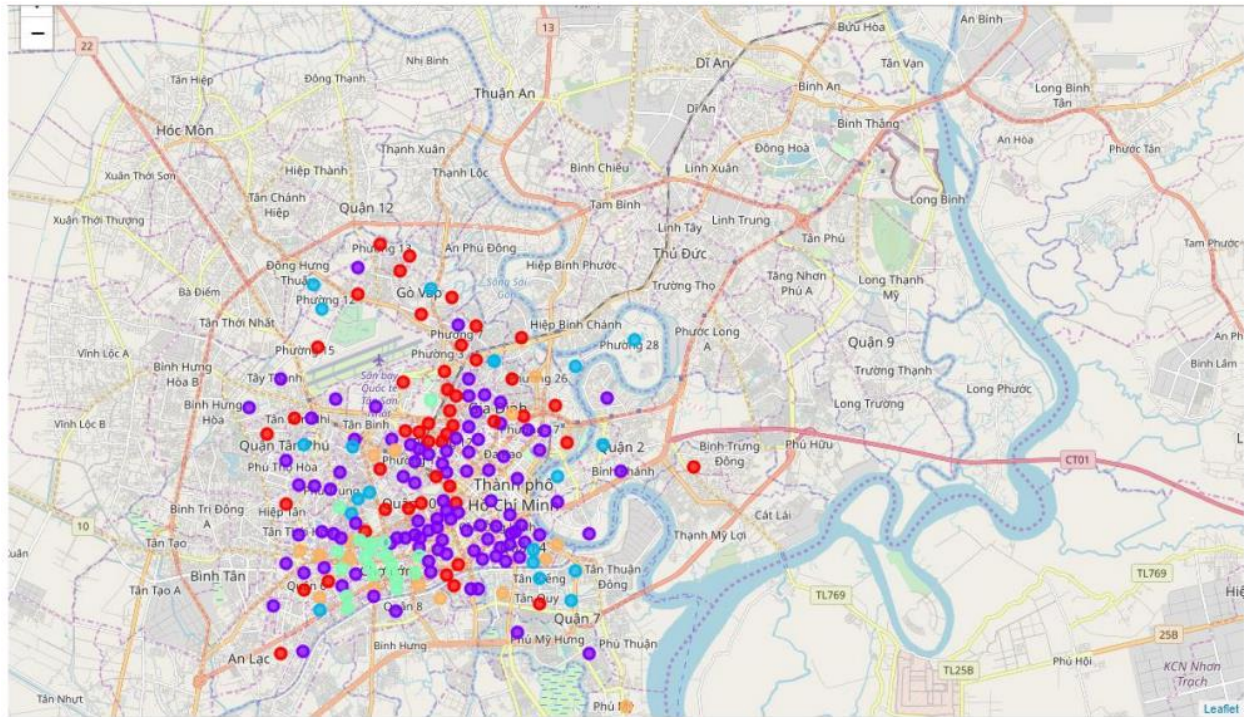
I used the Silhouette Analysis method to find the optimal cluster number for the Kmean algorithm.

Then use the Kmean algorithm from the sklearn library to divide it into 5 groups with similar properties. Next, assign labels from Kmean result to each neighborhood using the Pandas merge function.

The next step is to display and view a list of each cluster, here to see what the clusters have in common, from which to draw useful information.

### 3.2. Use folium and matplotlib to draw on the map

This project use folium to draw map and folium to draw circle marker to visualize the resulting clusters.



## 4. Results

After dividing into 5 clusters with similar characteristics, I had the following results:

### Cluster 1

```
hcm_merged.loc[hcm_merged['Cluster Labels'] == 0, hcm_merged.columns[[2] + list(range(5, hcm_merged.shape[1]))]]
```

	Neighbourhood	Cluster Labels	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
9214	Phường Tân Định, Quận 1	0.0	Café	Vietnamese Restaurant	Vegetarian / Vegan Restaurant	Asian Restaurant	Breakfast Spot	Food	Japanese Restaurant	Bar
9215	Phường Đa Kao, Quận 1	0.0	Vietnamese Restaurant	Café	French Restaurant	Japanese Restaurant	Vegetarian / Vegan Restaurant	Dessert Shop	Beer Garden	Bubble Tea Shop
9216	Phường Bến Nghé, Quận 1	0.0	Café	Bar	Vietnamese Restaurant	Thai Restaurant	Asian Restaurant	French Restaurant	Restaurant	Ramen Restaurant
9217	Phường Bến Thành, Quận 1	0.0	Vietnamese Restaurant	Café	Sandwich Place	Food	Beer Garden	National Park	Cajun / Creole Restaurant	Pizza Place
9218	Phường Nguyễn Thái Bình, Quận 1	0.0	Vietnamese Restaurant	Café	Bar	Restaurant	Fast Food Restaurant	Sports Bar	Beer Bar	Bubble Tea Shop
9219	Phường Phạm Ngũ Lão, Quận 1	0.0	Vietnamese Restaurant	Café	Bar	Vegetarian / Vegan Restaurant	Restaurant	Japanese Restaurant	Bakery	Seafood Restaurant
9220	Phường Cầu Ông Lãnh, Quận 1	0.0	Vietnamese Restaurant	Café	Restaurant	Juice Bar	Japanese Restaurant	Indian Restaurant	Whisky Bar	Ice Cream Shop
9221	Phường Cô Giang, Quận 1	0.0	Vegetarian / Vegan Restaurant	Restaurant	Steakhouse	Vietnamese Restaurant	Japanese Restaurant	Middle Eastern Restaurant	Seafood Restaurant	Food
9222	Phường Nguyễn Cư Trinh, Quận 1	0.0	Café	Vietnamese Restaurant	Seafood Restaurant	Dessert Shop	Indian Restaurant	Food	Sandwich Place	Noodle House
9223	Phường Cầu Kho, Quận 1	0.0	Café	Vietnamese Restaurant	Asian Restaurant	American Restaurant	Hotpot Restaurant	Dessert Shop	Food	Seafood Restaurant
9269	Phường 7, Quận Gò Vấp	0.0	Restaurant	Vietnamese Restaurant	Asian Restaurant	Café	Hotpot Restaurant	Fast Food Restaurant	Food	French Restaurant
9272	Phường 9, Quận Gò Vấp	0.0	BBQ Joint	Vietnamese Restaurant	Café	Whisky Bar	Hotpot Restaurant	Food	French Restaurant	Frozen Yogurt Shop

## Cluster 2

```
hcm_merged.loc[hcm_merged['Cluster Labels'] == 1, hcm_merged.columns[[2] + list(range(5, hcm_merged.shape[1]))]]
```

	Neighbourhood	Cluster Labels	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
9262	Phường 17, Quận Gò Vấp	1.0	Vietnamese Restaurant	Warehouse Store	Ice Cream Shop	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop	Gastropub
9265	Phường 12, Quận Gò Vấp	1.0	Vietnamese Restaurant	Japanese Restaurant	Whisky Bar	Hotpot Restaurant	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop
9266	Phường 14, Quận Gò Vấp	1.0	Vietnamese Restaurant	Café	Whisky Bar	Hotpot Restaurant	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop
9274	Phường 11, Quận Gò Vấp	1.0	Vietnamese Restaurant	Whisky Bar	Hotpot Restaurant	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop	Gastropub
9277	Phường 11, Quận Bình Thạnh	1.0	Vietnamese Restaurant	Whisky Bar	Hotpot Restaurant	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop	Gastropub
9278	Phường 27, Quận Bình Thạnh	1.0	Vietnamese Restaurant	Asian Restaurant	Whisky Bar	Hotpot Restaurant	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop
9295	Phường 28, Quận Bình Thạnh	1.0	Vietnamese Restaurant	BBQ Joint	Whisky Bar	Hotpot Restaurant	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop
9302	Phường 11, Quận Tân Bình	1.0	Vietnamese Restaurant	Noodle House	Fast Food Restaurant	Whisky Bar	Food	French Restaurant	Frozen Yogurt Shop	Gastropub
9308	Phường 9, Quận Tân Bình	1.0	Vietnamese Restaurant	Café	Seafood Restaurant	Vegetarian / Vegan Restaurant	Chinese Restaurant	Greek Restaurant	Fast Food Restaurant	Food
9315	Phường Tân Thành, Quận Tân Phú	1.0	Vietnamese Restaurant	Snack Place	Steakhouse	Whisky Bar	German Restaurant	Eastern European Restaurant	Fast Food Restaurant	Food
9339	Phường Bình An, Quận 2	1.0	Vietnamese Restaurant	Café	Whisky Bar	Hotpot Restaurant	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop
9343	Phường An Khánh, Quận 2	1.0	Vietnamese Restaurant	Whisky Bar	Hotpot Restaurant	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop	Gastropub

## Cluster 3



### Cluster 3

```
hcm_merged.loc[hcm_merged['Cluster Labels'] == 2, hcm_merged.columns[[2] + list(range(5, hcm_merged.shape[1]))]]
```

	Neighbourhood	Cluster Labels	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
9324	Phường 9, Quận Phú Nhuận	2.0	Chinese Restaurant	Whisky Bar	Eastern European Restaurant	Food	French Restaurant	Frozen Yogurt Shop	Gastropub	German Restaurant
9378	Phường 5, Quận 11	2.0	Ice Cream Shop	Chinese Restaurant	Café	Hotpot Restaurant	Food	French Restaurant	Frozen Yogurt Shop	Gastropub
9386	Phường 12, Quận 11	2.0	Chinese Restaurant	Seafood Restaurant	Asian Restaurant	Basketball Stadium	Café	Hotpot Restaurant	Food	French Restaurant
9387	Phường 7, Quận 11	2.0	Chinese Restaurant	Café	Asian Restaurant	BBQ Joint	Whisky Bar	Ice Cream Shop	French Restaurant	Frozen Yogurt Shop
9388	Phường 6, Quận 11	2.0	Chinese Restaurant	Café	Seafood Restaurant	Asian Restaurant	BBQ Joint	Hotpot Restaurant	Food	French Restaurant
9389	Phường 4, Quận 11	2.0	Chinese Restaurant	Seafood Restaurant	Asian Restaurant	Hotpot Restaurant	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop
9391	Phường 2, Quận 11	2.0	Chinese Restaurant	Asian Restaurant	Whisky Bar	Ice Cream Shop	Food	French Restaurant	Frozen Yogurt Shop	Gastropub
9411	Phường 12, Quận 5	2.0	Chinese Restaurant	Café	Seafood Restaurant	Fast Food Restaurant	Dessert Shop	Asian Restaurant	Vietnamese Restaurant	Juice Bar
9413	Phường 8, Quận 5	2.0	Chinese Restaurant	Café	Vietnamese Restaurant	Noodle House	Dessert Shop	Asian Restaurant	BBQ Joint	Seafood Restaurant
9414	Phường 15, Quận 5	2.0	Chinese Restaurant	Seafood Restaurant	Café	Restaurant	Juice Bar	Korean Restaurant	Fast Food Restaurant	Food
9417	Phường 11, Quận 5	2.0	Chinese Restaurant	Dessert Shop	Vietnamese Restaurant	Café	BBQ Joint	Fast Food Restaurant	Greek Restaurant	Food
9418	Phường 14, Quận 5	2.0	Chinese Restaurant	Dessert Shop	Seafood Restaurant	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop	Gastropub

### Cluster 4

```
hcm_merged.loc[hcm_merged['Cluster Labels'] == 3, hcm_merged.columns[[2] + list(range(5, hcm_merged.shape[1]))]]
```

	Neighbourhood	Cluster Labels	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
9279	Phường 26, Quận Bình Thạnh	3.0	Food	Asian Restaurant	Bakery	Café	Ice Cream Shop	French Restaurant	Frozen Yogurt Shop	Gastropub
9288	Phường 2, Quận Bình Thạnh	3.0	Seafood Restaurant	Asian Restaurant	Food	Café	Hotpot Restaurant	Fast Food Restaurant	French Restaurant	Frozen Yogurt Shop
9303	Phường 7, Quận Tân Bình	3.0	Restaurant	Asian Restaurant	Food	Café	Whisky Bar	Hotpot Restaurant	French Restaurant	Frozen Yogurt Shop
9304	Phường 5, Quận Tân Bình	3.0	Middle Eastern Restaurant	Food	Asian Restaurant	Juice Bar	Café	Hotpot Restaurant	French Restaurant	Frozen Yogurt Shop
9384	Phường 8, Quận 11	3.0	Asian Restaurant	Basketball Stadium	Café	Whisky Bar	Ice Cream Shop	French Restaurant	Frozen Yogurt Shop	Gastropub
9390	Phường 1, Quận 11	3.0	Asian Restaurant	Vegetarian / Vegan Restaurant	Food	Whisky Bar	Hotpot Restaurant	Fast Food Restaurant	French Restaurant	Frozen Yogurt Shop
9392	Phường 16, Quận 11	3.0	Seafood Restaurant	Asian Restaurant	Food	Chinese Restaurant	Hotpot Restaurant	Fast Food Restaurant	French Restaurant	Frozen Yogurt Shop
9400	Phường 18, Quận 4	3.0	Restaurant	Food	Whisky Bar	Mediterranean Restaurant	Fast Food Restaurant	French Restaurant	Frozen Yogurt Shop	Gastropub
9404	Phường 16, Quận 4	3.0	Seafood Restaurant	Snack Place	Food	Greek Restaurant	Eastern European Restaurant	Fast Food Restaurant	French Restaurant	Frozen Yogurt Shop
9423	Phường 14, Quận 6	3.0	Asian Restaurant	Whisky Bar	Ice Cream Shop	Food	French Restaurant	Frozen Yogurt Shop	Gastropub	German Restaurant
9431	Phường 1, Quận 6	3.0	Asian Restaurant	Vegetarian / Vegan Restaurant	Food	Whisky Bar	Hotpot Restaurant	Fast Food Restaurant	French Restaurant	Frozen Yogurt Shop
9433	Phường 8, Quận 6	3.0	Asian Restaurant	Food	Whisky Bar	Ice Cream Shop	French Restaurant	Frozen Yogurt Shop	Gastropub	German Restaurant

### Cluster 5

```
hcm_merged.loc[hcm_merged['Cluster Labels'] == 4, hcm_merged.columns[[2] + list(range(5, hcm_merged.shape[1]))]]
```

	Neighbourhood	Cluster Labels	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
9260	Phường 15, Quận Gò Vấp	4.0	Café	Whisky Bar	Eastern European Restaurant	Food	French Restaurant	Frozen Yogurt Shop	Gastropub	German Restaurant
9261	Phường 13, Quận Gò Vấp	4.0	Café	Food	Whisky Bar	Eastern European Restaurant	French Restaurant	Frozen Yogurt Shop	Gastropub	German Restaurant
9263	Phường 6, Quận Gò Vấp	4.0	Café	Fast Food Restaurant	Noodle House	Hotpot Restaurant	Food	French Restaurant	Frozen Yogurt Shop	Gastropub
9264	Phường 16, Quận Gò Vấp	4.0	Café	Vietnamese Restaurant	Japanese Restaurant	Whisky Bar	Greek Restaurant	Fast Food Restaurant	Food	French Restaurant
9267	Phường 10, Quận Gò Vấp	4.0	Café	Vietnamese Restaurant	Pizza Place	Bubble Tea Shop	Whisky Bar	Greek Restaurant	Food	French Restaurant
9268	Phường 5, Quận Gò Vấp	4.0	Café	Asian Restaurant	Whisky Bar	Ice Cream Shop	Food	French Restaurant	Frozen Yogurt Shop	Gastropub
9270	Phường 4, Quận Gò Vấp	4.0	Café	Vietnamese Restaurant	Whisky Bar	Hotpot Restaurant	Fast Food Restaurant	Food	French Restaurant	Frozen Yogurt Shop
9271	Phường 1, Quận Gò Vấp	4.0	Café	Vietnamese Restaurant	Fast Food Restaurant	Whisky Bar	Hotpot Restaurant	Food	French Restaurant	Frozen Yogurt Shop
9273	Phường 8, Quận Gò Vấp	4.0	Café	Vietnamese Restaurant	BBQ Joint	Whisky Bar	Hotpot Restaurant	Food	French Restaurant	Frozen Yogurt Shop
9275	Phường 3, Quận Gò Vấp	4.0	Café	Seafood Restaurant	Noodle House	Korean Restaurant	Juice Bar	Fast Food Restaurant	Food	French Restaurant
9276	Phường 13, Quận Bình Thạnh	4.0	Asian Restaurant	Café	Whisky Bar	Ice Cream Shop	Food	French Restaurant	Frozen Yogurt Shop	Gastropub
9280	Phường 12, Quận Bình Thạnh	4.0	Café	Food	Whisky Bar	Eastern European Restaurant	French Restaurant	Frozen Yogurt Shop	Gastropub	German Restaurant

## 5. Discussion

After reviewing the data of each cluster, the following were observed:

- Cluster 1,2 has a lot of Vietnamese restaurants. Opening Vietnamese restaurants in these areas will face competition
- Cluster 1,5 has a lot of cafes, so it would be difficult to have successful a cafe here.
- Cluster 3 has a lot of Chinese restaurants

## 6. Conclusion

The project gives me a hands on experience with running a data-science project.

I used various types of APIs to collect data, used the Pandas library to eliminate redundant data, used it, and used Python libraries to draw graphs, using unsupervised machine learning algorithms to group data into similar characteristics.

From that it is possible to discover the information that is hidden in it, making it easier to make decisions such as where to open a restaurant and which restaurant is suitable.