HO CHI MINH CITY VENUES ANALYSIS

AND RECOMMENDATION FOR STARTING FOOD BUSINESS

I. Introduction and Disscusion of the problem:

Ho Chi Minh city is the largest in Viet Nam where over 8 million people live. As a resident of this city, I decided to use Ho Chi Minh in my project. The city is divided into 18 districts in total.

As you can see from the figures, Ho Chi Minh is a city with a high population and population density. Being such a crowded city leads the owners of shops and social sharing places in the city where the population is dense. When we think of starting food business, we expect looking for the districts where the type of business they want to install is less intense but high demand in overall

When we consider this problem, we can create a map and information chart describe venues Ho Chi Minh city and each district is clustered according to the venue density and make a decision based on that info

II. Data Description:

To consider the problem we can list the data as below:

- There are not too many public geographical data related to Ho Chi Minh so I have to used Google Map and manually create geographical location for each district in Ho Chi Minh

- I used Forsquare API to get the most common venues Ho Chi Minh city and group them by district

III. Methodology:

1. Data preparation:

First, I created a location data using Google Map and put all of them in to a csv file then read it in my notebook.

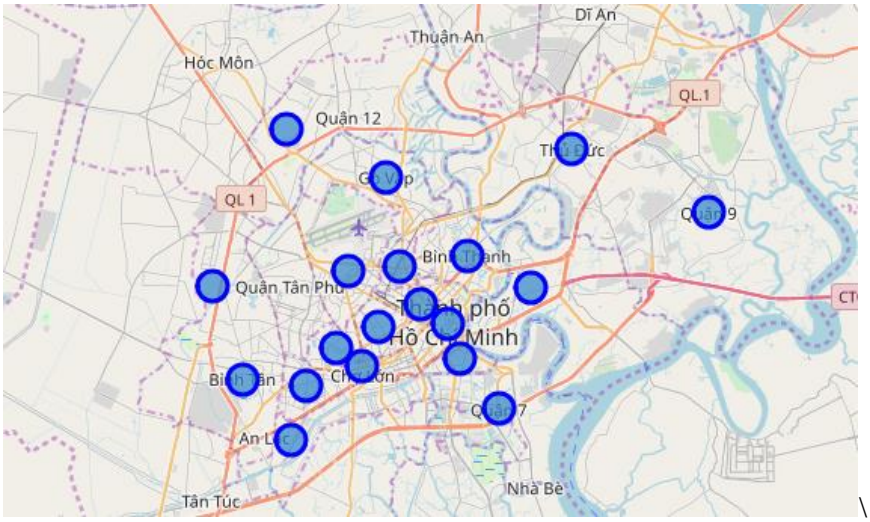
My Ho Chi Minh geographical location data only has three main attributes: District, Latitude("lat"), and Longitude("lng") as shown below.

HCM_data.csv

```
District,lat,lng
District 1,10.7745397,106.6991836
District 2,10.7911157,106.7367293
District 3,10.7835291,106.6870984
District 4,10.7592431,106.7048897
District 5,10.7558399,106.6600163
District 6,10.746928,106.6344946
District 7,10.7365728,106.7224322
District 8,10.7228825,106.6278244
District 9,10.8245429,106.8180145
District 10,10.773198,106.6678329
District 11,10.764006, 106.648642
District 12,10.8613975,106.6255926
Go Vap,10.84015,106.6710828
Tan Binh,10.7979794,106.6538054
Tan Phu,10.7914555,106.5922991
Binh Thanh,10.8046591,106.7078477
Binh Tan,10.7498093,106.6056635
Phu Nhuan,10.8001182,106.6770417
Thu Duc,10.852588,106.7558383
```

	District	lat	lng
0	District 1	10.774540	106.699184
1	District 2	10.791116	106.736729
2	District 3	10.783529	106.687098
3	District 4	10.759243	106.704890
4	District 5	10.755840	106.660016

Then I used python folium library to visualize geographic details of Ho Chi Minh and its district. Each label which represents for each district is marked in the center of the district



I utilized the Foursquare API to explore the district and segment them. I designed the limit as 100 venue and the radius 10 kilometers for each district from their given latitude and longitude information because the mean areas of them is about 20km2. Here is a head of the list Venues name, category, latitude and longitude informations from Forsquare API.

	name	categories	.	lat	lng
0	Seventeen Coffee	Lounge		10.825413	106.629618
1	Sân Golf Tân Sơn Nhất	Golf Course		10.829959	106.649974
2	AEON Supermarket	Supermarket		10.801783	106.618443
3	Celadon City	Park		10.802461	106.618018
4	Lẩu Dê Đức Dưỡng	Vietnamese Restaurant		10.833790	106.644277
5	CGV Cinemas Hoàng Văn Thụ	Multiplex		10.798997	106.659965
6	Hoa Viên Phổ Quang	Beer Garden		10.803630	106.666228
7	Starbucks	Coffee Shop		10.813034	106.662770

Then I merged it into the geological data:

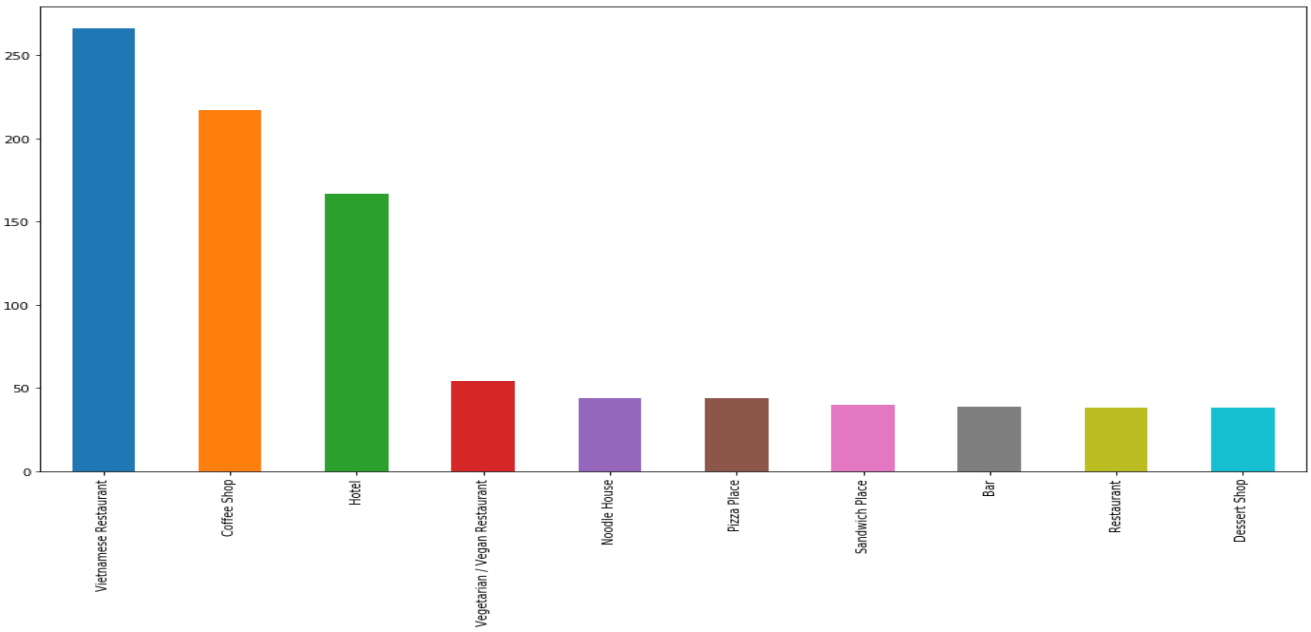
	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	District 1	10.77454	106.699184	Pizza 4P's	10.773301	106.697599	Pizza Place
1	District 1	10.77454	106.699184	Pasteur Street Brewing Company	10.775220	106.700894	Brewery
2	District 1	10.77454	106.699184	Park Hyatt Saigon	10.777574	106.703609	Hotel
3	District 1	10.77454	106.699184	Silverland Yen Hotel	10.774850	106.696160	Hotel
4	District 1	10.77454	106.699184	Takashimaya	10.773194	106.701075	Department Store

Using above limit and radius, there are total 1854 venues for 18 districts in Ho Chi Minh found from Forsquare API. We can also see that except District 12 and 9, all remain reached the 100 limit of venues

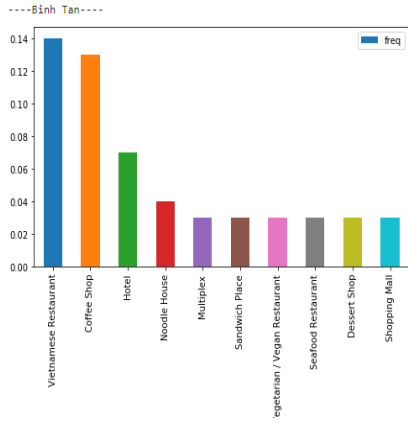
2. Data Exploration:

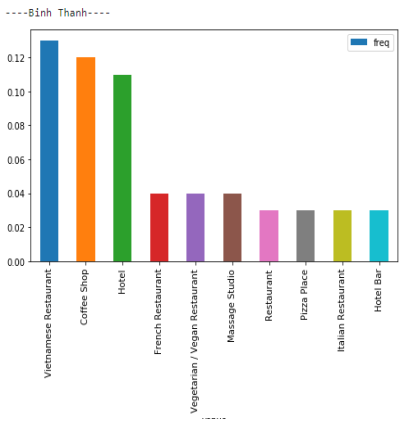
- In Viet Nam, coffee shop and café are the same type of venue so I replace all “café” venue type by “coffee shop” and use bar chart to describe top 10 venues type for Ho Chi Minh city and its districts

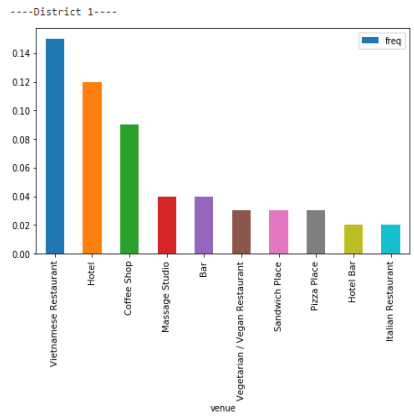
Top 10 venues type Ho Chi Minh city:

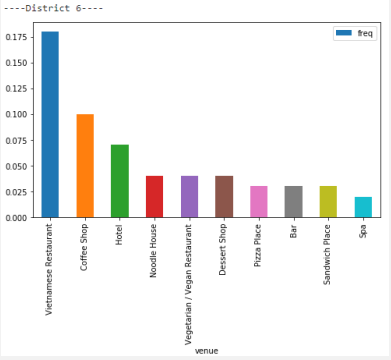


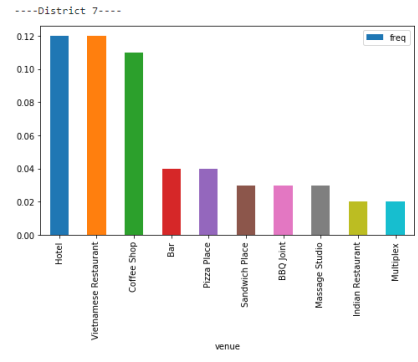
Top 10 venues type for each district in Ho Chi Minh

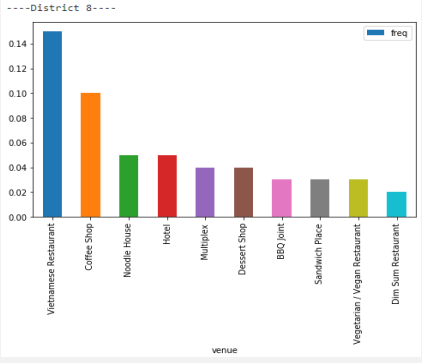


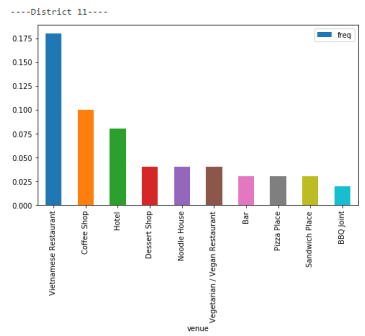


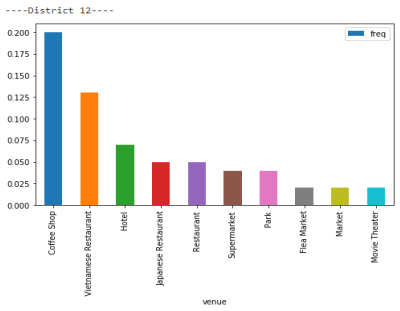


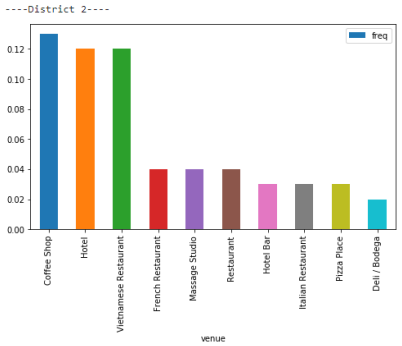


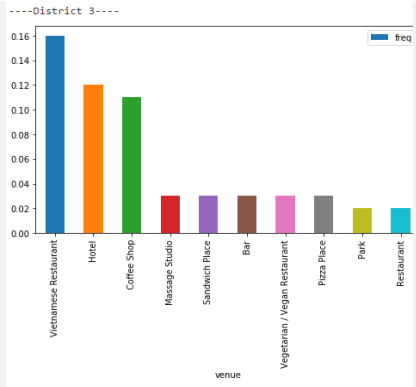


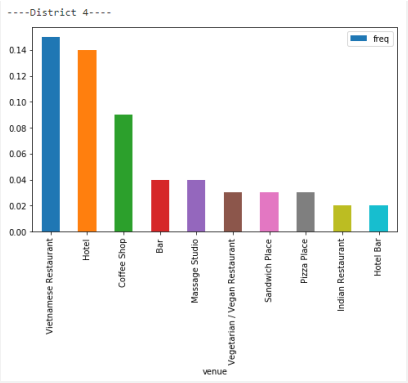


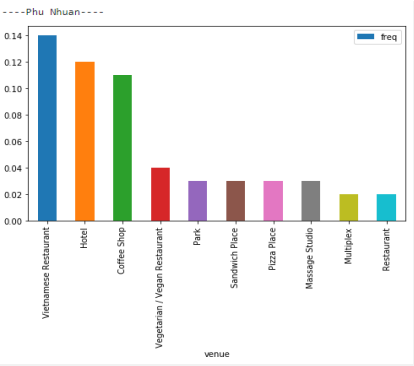

----District 11----
freq
Vietnamese Restaurant
Coffee Shop
Hotel
Dessert Shop
Noodle House
Vegetarian / Vegan Restaurant
Bar
Pizza Place
Sandwich Place
BBQ Joint
venue

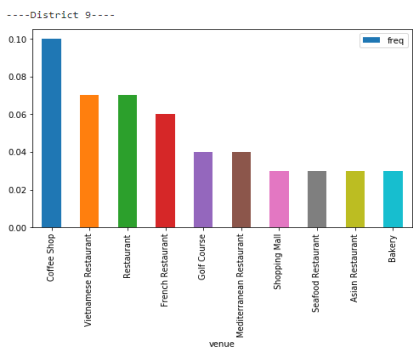

----District 12----
freq
Coffee Shop
Vietnamese Restaurant
Hotel
Japanese Restaurant
Restaurant
Supermarket
Park
Flea Market
Market
Movie Theater
venue

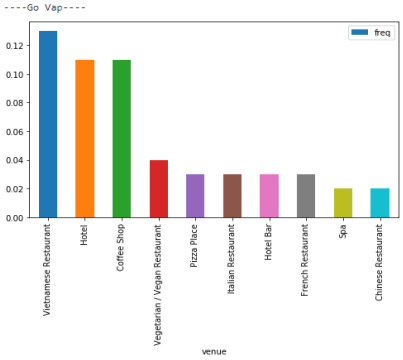

----District 2----
freq
Coffee Shop
Hotel
Vietnamese Restaurant
French Restaurant
Massage Studio
Restaurant
Hotel Bar
Italian Restaurant
Pizza Place
Deli / Bodega
venue

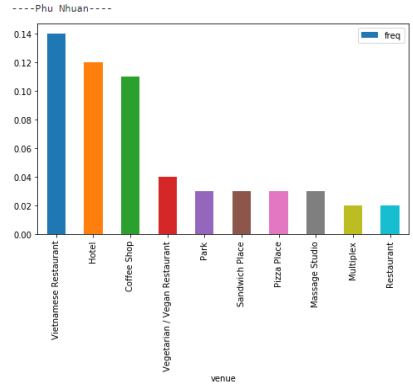

----District 3----
freq
Vietnamese Restaurant
Hotel
Coffee Shop
Massage Studio
Sandwich Place
Bar
Vegetarian / Vegan Restaurant
Pizza Place
Park
Restaurant
venue

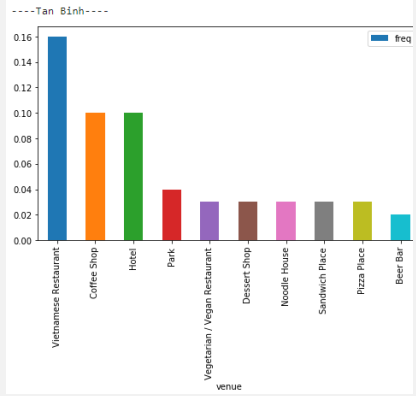

----District 4----
freq
Vietnamese Restaurant
Hotel
Coffee Shop
Bar
Massage Studio
Vegetarian / Vegan Restaurant
Sandwich Place
Pizza Place
Indian Restaurant
Hotel Bar
venue

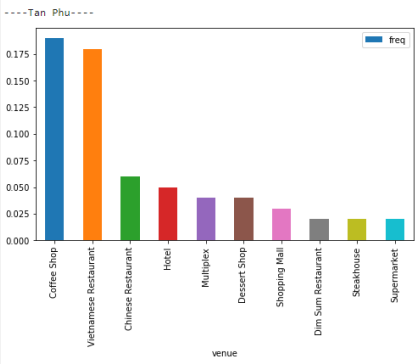

----Phu Nhuan----
freq
Vietnamese Restaurant
Hotel
Coffee Shop
Vegetarian / Vegan Restaurant
Park
Sandwich Place
Pizza Place
Massage Studio
Multiplex
Restaurant
venue

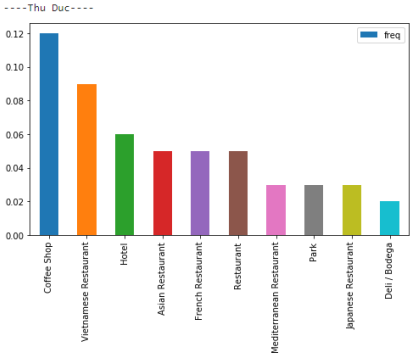

----District 9----
freq
Coffee Shop
Vietnamese Restaurant
Restaurant
French Restaurant
Golf Course
Mediterranean Restaurant
Shopping Mall
Seafood Restaurant
Asian Restaurant
Bakery
venue


----Go Vap----
freq
Vietnamese Restaurant
Hotel
Coffee Shop
Vegetarian / Vegan Restaurant
Pizza Place
Italian Restaurant
Hotel Bar
French Restaurant
Spa
Chinese Restaurant
venue


----Phu Nhuan----
freq
Vietnamese Restaurant
Hotel
Coffee Shop
Vegetarian / Vegan Restaurant
Park
Sandwich Place
Pizza Place
Massage Studio
Multiplex
Restaurant
venue


----Tan Binh----
freq
Vietnamese Restaurant
Coffee Shop
Hotel
Park
Vegetarian / Vegan Restaurant
Dessert Shop
Noodle House
Sandwich Place
Pizza Place
Beer Bar
venue


----Tan Phu----
freq
Coffee Shop
Vietnamese Restaurant
Chinese Restaurant
Hotel
Multiplex
Dessert Shop
Shopping Mall
Dim Sum Restaurant
Steakhouse
Supermarket
venue


----Thu Duc----
freq
Coffee Shop
Vietnamese Restaurant
Hotel
Asian Restaurant
French Restaurant
Restaurant
Mediterranean Restaurant
Park
Japanese Restaurant
Deli / Bodega

In summary of these graphs, 10 largest unique categories were returned by Foursquare, then I created a table which shows list of them in below table for processing later

	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	Binh Tan	Vietnamese Restaurant	Coffee Shop	Hotel	Noodle House	Sandwich Place	Vegetarian / Vegan Restaurant	Dessert Shop	Multiplex	Seafood Restaurant	Shopping Mall
1	Binh Thanh	Vietnamese Restaurant	Coffee Shop	Hotel	Vegetarian / Vegan Restaurant	Massage Studio	French Restaurant	Italian Restaurant	Restaurant	Hotel Bar	Pizza Place
2	District 1	Vietnamese Restaurant	Hotel	Coffee Shop	Massage Studio	Bar	Vegetarian / Vegan Restaurant	Pizza Place	Sandwich Place	Whisky Bar	Multiplex
3	District 10	Vietnamese Restaurant	Coffee Shop	Hotel	Noodle House	Pizza Place	Bar	Dessert Shop	Spa	Sandwich Place	Vegetarian / Vegan Restaurant
4	District 11	Vietnamese Restaurant	Coffee Shop	Hotel	Vegetarian / Vegan Restaurant	Noodle House	Dessert Shop	Sandwich Place	Pizza Place	Bar	Bed & Breakfast

3. K-mean Clustering:

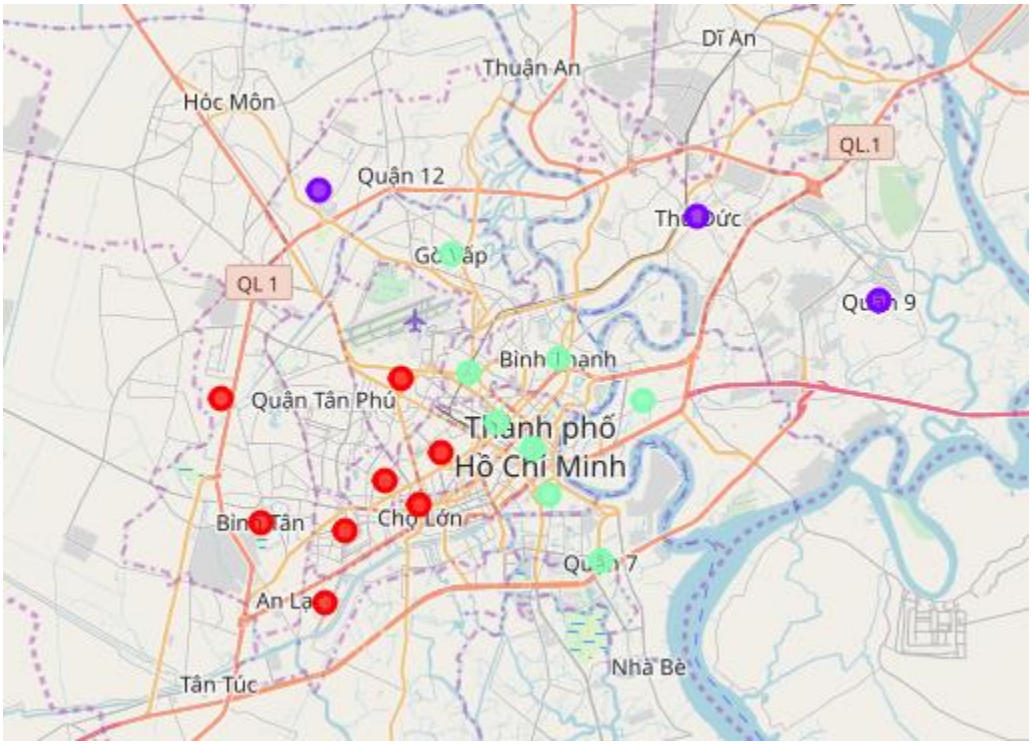
We have some common venue categories in districts. In this reason I used unsupervised learning K-means algorithm to cluster the districts. K-Means algorithm is one of the most common cluster method of unsupervised learning.

First, I will run K-Means to cluster the boroughs into 3 clusters

Here is head of my merged table:

	District	lat	lng	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	9th Most Common Venue	10th Most Common Venue
0	District 1	10.774540	106.699184	1	Vietnamese Restaurant	Hotel	Coffee Shop	Massage Studio	Bar	Vegetarian / Vegan Restaurant	Pizza Place	Sandwich Place	Whisky Bar	Multiplex
1	District 2	10.791116	106.736729	4	Coffee Shop	Hotel	Vietnamese Restaurant	Massage Studio	French Restaurant	Restaurant	Hotel Bar	Italian Restaurant	Pizza Place	Chinese Restaurant
2	District 3	10.783529	106.687098	1	Vietnamese Restaurant	Hotel	Coffee Shop	Vegetarian / Vegan Restaurant	Sandwich Place	Pizza Place	Massage Studio	Bar	Park	Italian Restaurant
3	District 4	10.759243	106.704890	1	Vietnamese Restaurant	Hotel	Coffee Shop	Bar	Massage Studio	Pizza Place	Vegetarian / Vegan Restaurant	Sandwich Place	Multiplex	Bed & Breakfast
4	District 5	10.755840	106.660016	3	Vietnamese Restaurant	Coffee Shop	Hotel	Noodle House	Vegetarian / Vegan Restaurant	Dessert Shop	Sandwich Place	Bar	Pizza Place	BBQ Joint

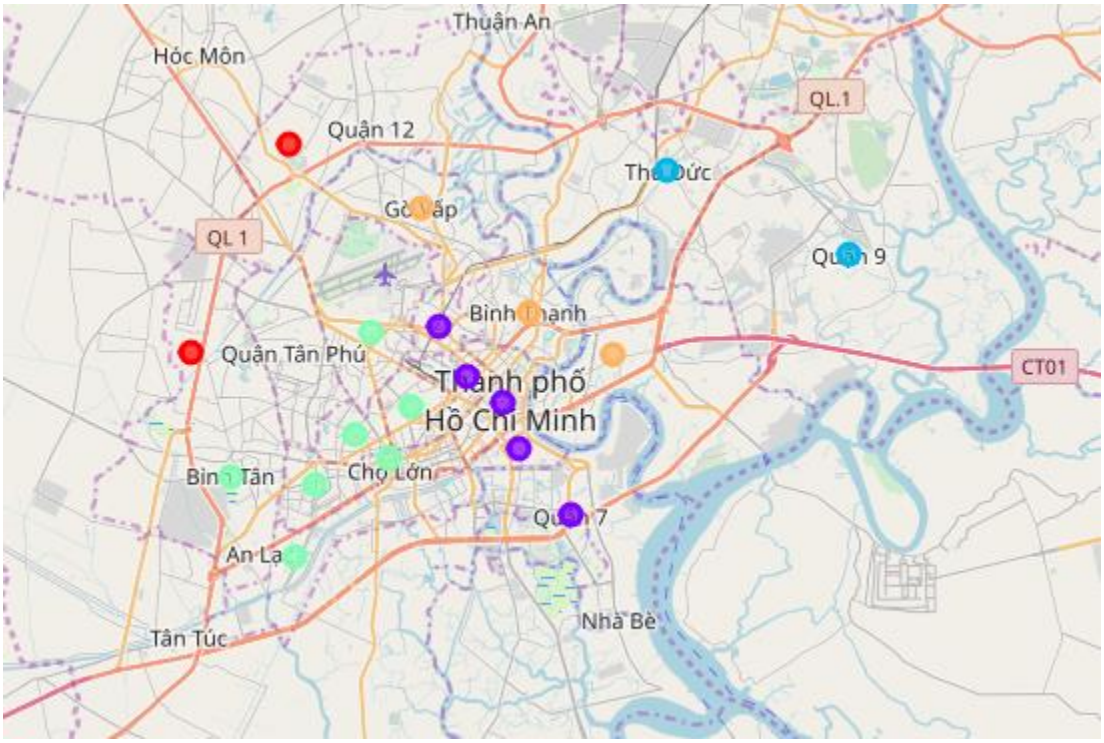
And Map:



Re-run it with 5 clusters.

Result merged and map:

	District	lat	lng	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	9th Most Common Venue	10th Most Common Venue
0	District 1	10.774540	106.699184	1	Vietnamese Restaurant	Hotel	Coffee Shop	Massage Studio	Bar	Vegetarian / Vegan Restaurant	Pizza Place	Sandwich Place	Whisky Bar	Multiplex
1	District 2	10.791116	106.736729	4	Coffee Shop	Hotel	Vietnamese Restaurant	Massage Studio	French Restaurant	Restaurant	Hotel Bar	Italian Restaurant	Pizza Place	Chinese Restaurant
2	District 3	10.783529	106.687098	1	Vietnamese Restaurant	Hotel	Coffee Shop	Vegetarian / Vegan Restaurant	Sandwich Place	Pizza Place	Massage Studio	Bar	Park	Italian Restaurant
3	District 4	10.759243	106.704890	1	Vietnamese Restaurant	Hotel	Coffee Shop	Bar	Massage Studio	Pizza Place	Vegetarian / Vegan Restaurant	Sandwich Place	Multiplex	Bed & Breakfast
4	District 5	10.755840	106.660016	3	Vietnamese Restaurant	Coffee Shop	Hotel	Noodle House	Vegetarian / Vegan Restaurant	Dessert Shop	Sandwich Place	Bar	Pizza Place	BBQ Joint



IV. Discussion and Result/Recommendation:

Whether running K-mean Algorithm with number of cluster is 3 or 5, there are always 2 separated clusters of Western and Eastern, and 2 separated clusters of center and suburb of Ho Chi Minh city.

The most common venue type is “Vietnamese Restaurant” for most of district in center of Ho Chi Minh city. However, for urban district like District 9 and Thu Duc, the most common one is coffee shop.

Government is now investing more on the infrastructure for suburb. There are more and more company and factory there. So we can consider starting a Vietnamese Restaurant.

When running K-mean with k = 5, the Eastern of city is separated to 2 different clusters. We can see that the number of venue type Vietnamese restaurant, Coffee Shop and Hotel is quite similar. So we can hope the distribution of venue type in the future of those district will be effected by center district and start opening Vietnamese Restaurant in District 2 or Binh Thanh, opening coffee shop in District 10