**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

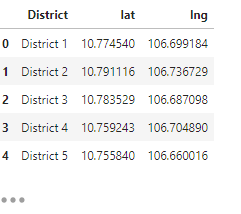
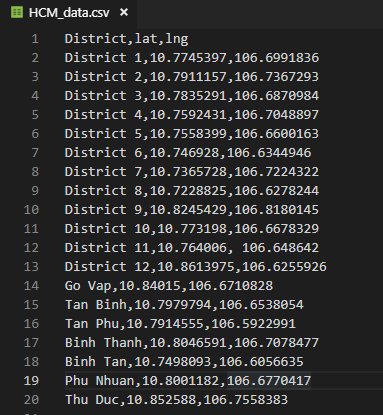
- I used Forsquare API to get the most common venues Ho Chi Minh city and group them by district

**III. Methodology:**

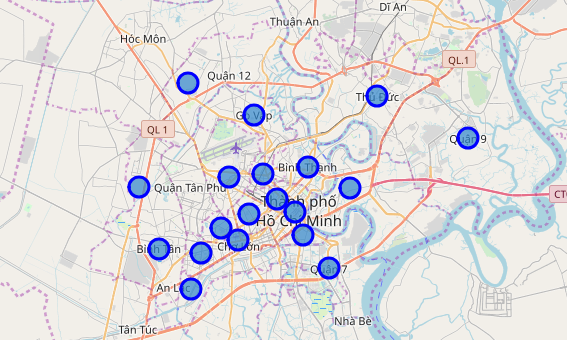
**1. Data preparation:**

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.

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.



Then I merged it into the geological data:

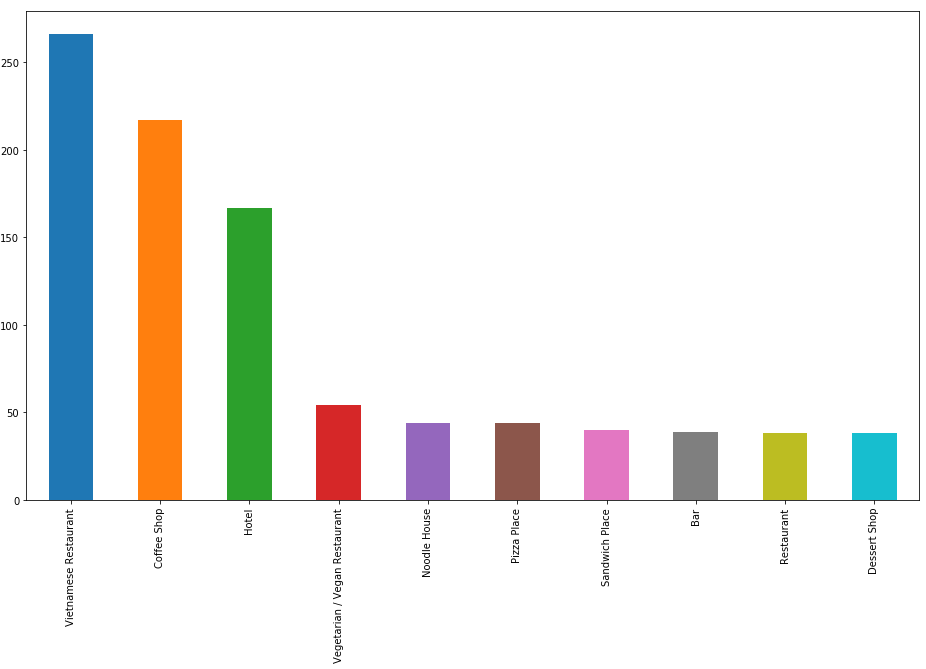


Using above limit and radius, there are total 1854 venues for 18 districts in Ho Chi Minh found from Forsquare API. We can also see that except District 12 and 9, all remain reached the 100 limit of venues

**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

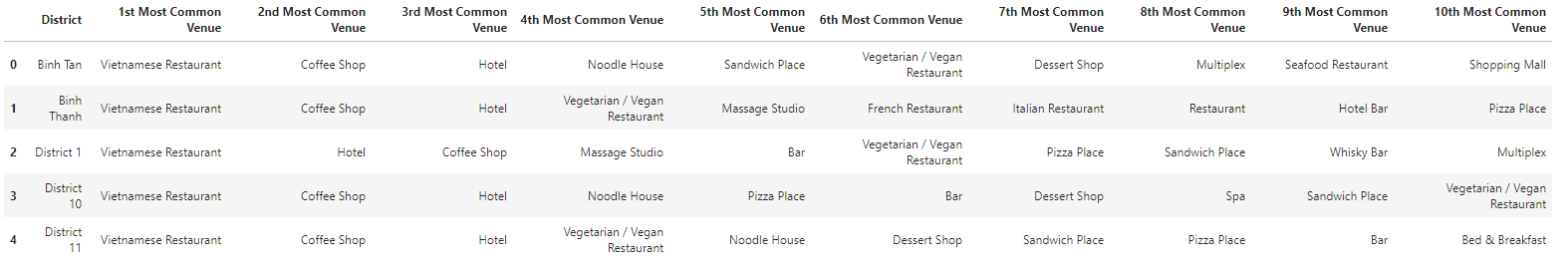
Top 10 venues type Ho Chi Minh city:



Top 10 venues type for each district in Ho Chi Minh

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

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



**3. K-mean Clustering:**

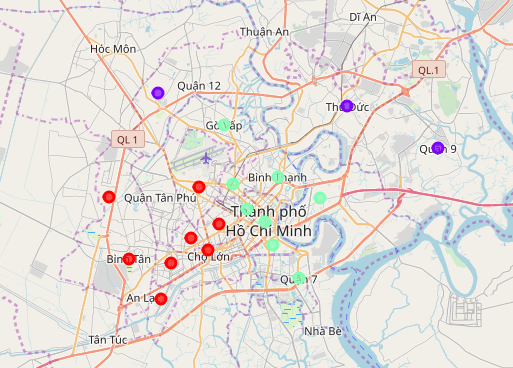
We have some common venue categories in districts. In this reason I used unsupervised learning K-means algorithm to cluster the districts. K-Means algorithm is one of the most common cluster method of unsupervised learning.

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

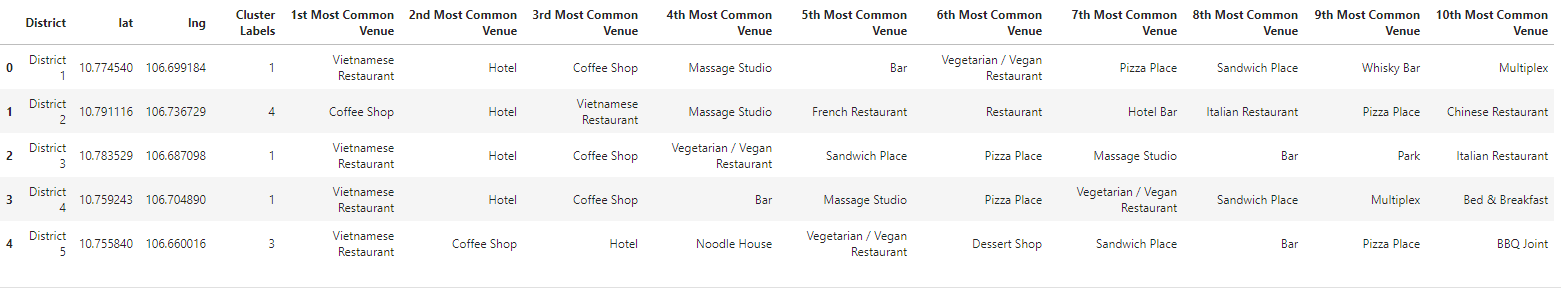
Here is head of my merged table:

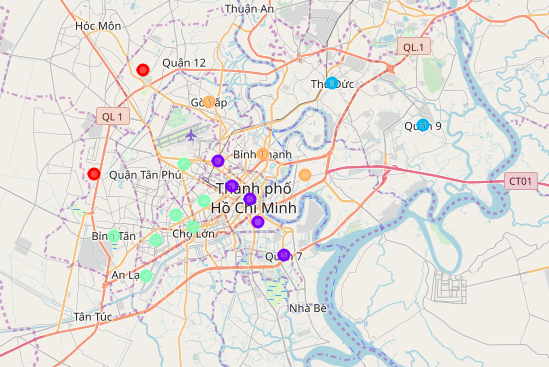


And Map:



Re-run it with 5 clusters.

Result merged and map:



**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