

Coursera Capstone

IBM Applied Data Science Capstone

Opening a restaurant closed Ho Chi Minh City, VietNam



Tran Nhat Huy
HCMC, January 2020

Introduction.

Lately year, almost people go to restaurant is increase. This is a reason opening a restaurant is create. So opening a restaurant is trend of almost people. And there are so many question how to opening a restaurant with compare order such as: amount of investment, food, material, staff,... and a question is important too is where do you open restaurant. So now I will discuss this problem.

Business problem.

This is an important step in restaurant business to opening a restaurant. But that is usually skip to save time and cost. However, that is a first reason lead to a restaurant losses and closures.

A restaurant can't attract all customers. So, a restaurant need decision your market and your object customer. Besides, learning about your competitors which help you create business advantages. So, when you learning market before restaurant business, you need answer some question in bellow:

- Your object business?
- Your competitors?
- Information about your product of competitors?
- What strengths and weaknesses do they have?
- Description about your object customer?
- Habit of customer (product, service)?
- Customer's demand?
- Where do you open restaurant?

When you answer all question, then need ideals for opening a restaurant. Ideals base on object customer and amount of investment. Ideals include style, business model, food,... Your ideals need attention to trend of market. Because maybe you obsolete in market.

Then you need choose restaurant model. Restaurant business market is very large. So you should decision kind of restaurant. You should base on ideals, amount of investment, object customer's demand.

End finally, when decision your all object. you need choose a location to opening restaurant. So how can you choose a suitable position. So now, I will research this problem.

Data.

To definition our problem, we will decision my data are:

- Number of restaurant in the neighborhood
- We will decided location where have a few restaurant in the neighborhood but closed city center.

Data sources:

- Number of restaurant in Ho Chi Minh
- Foursquare API

Sources of data:

Wikipedia page <https://worldpostalcode.com/vietnam/dong-nam-bo/ho-chi-minh/>. This is web have contain about list of neighborhood in Ho Chi Minh City. Which total 24 neighborhoods. We use web craping techniques to get data from the Wikipedia page. With the help of Python request and BeautifulSoup package. Then I will get location (latitude, longitude) from each neighborhood (using Python Geocoder package) and then create a map to show location on map.

Then I use Foursquare API to get the venue data for all neighbourhods. Foursquare has a large database about places, information for each place. Foursquare will provide many categories of the venue data. In this project I will use this data and many data science skills to solve the problem business (web craping, interface with Foursquare API, cleaning data, K-mean clustering and map visualization (folium)).

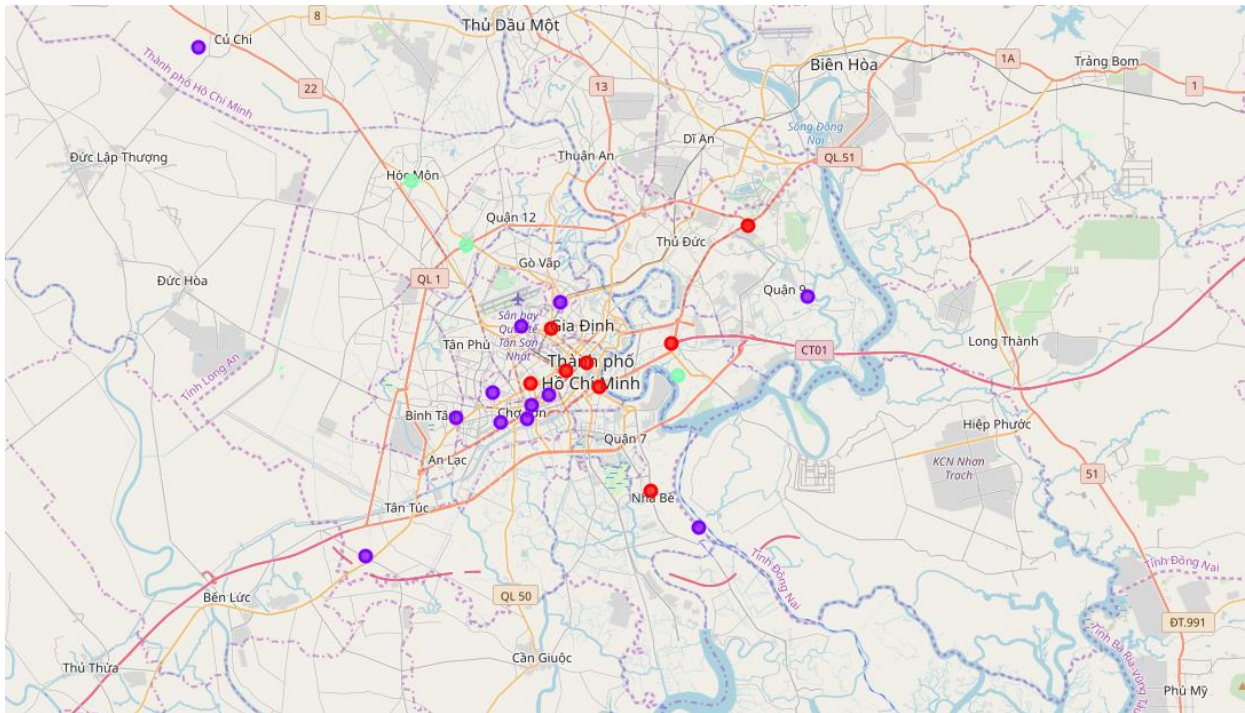
Method.

- Step 1: Get list of neighborhoods of Ho Chi Minh City.
- Step 2: Get location of each neighborhood.
- Step 3: Using Foursquare to get top 200 venues that in a radius of 5000 meters.
- Step 4: Check how many venues for each neighborhood and unique category.
- Step 5: Analysis each neighborhood (Taking the mean of the frequency of occurrence of each venue category).
- Step 6: Using K-mean to clustering data to 3 cluster base on frequency.

Results.

The result from the K-mean clustering show that we can categorize neighborhood into 3 cluster base on the frequency of occurrence for restaurant:

- Cluster 0: Neighborhood with moderate number of restaurants.
- Cluster 1: Neighborhood with low number to no existence of restaurants.
- Cluster 2: Neighborhood with high concentration of restaurants.



Observations.

Most of restaurant are concentrated in the center area of HCMC with the highest number in cluster 2 and moderate number in cluster 0. On the order hand, cluster 1 very low (Don't have any restaurant in there). So this represents a great opportunity and high potential areas to open new restaurant. Meanwhile, restaurant in center likely suffering from intense competition. So maybe I will choose District 5 and District 9 to open a restaurant because that is very low from intense competition and near city center.