

# Coursera Data Science Capstone IBM Professional Data Science Course

## OPENING A RESTAURANT IN HO CHI MINH CITY



Hoang Long Tran  
HoChiMinh city, 25/4/2020

## **Introduction**

I have always wanted to have a restaurant of my own because I love cuisine. However, we must have a strategy prior to doing everything, including opening a restaurant. Ho Chi Minh City is a huge urban area with a population of more than 10 millions. Therefore, the opportunity is also significant. However, the key to be successful in every business is the location. As a result, this report is conducted in order to help me to make a location decision in a data driven approach instead of randomly picking the location.

## **Business Problem**

The purpose of this report is to help me minimize risks caused by location malpicking, such as high competitiveness from other restaurants and lack of customers. Therefore, the question needs to be answered : Which Area is the best place to open a restaurant in Ho Chi Minh City.

## **Target Audience**

Since this report is made in order to help me make my decision for my restaurant, I am the one that this report aims to. However, it can also be used as a reference for those interested in opening a restaurant in HCMC, or maybe for my co-owners, or investors.

## Methodology

In order to make this report , I retrieved the data from Wikipedia : District of Ho Chi Minh city to get the name of 24 districts.

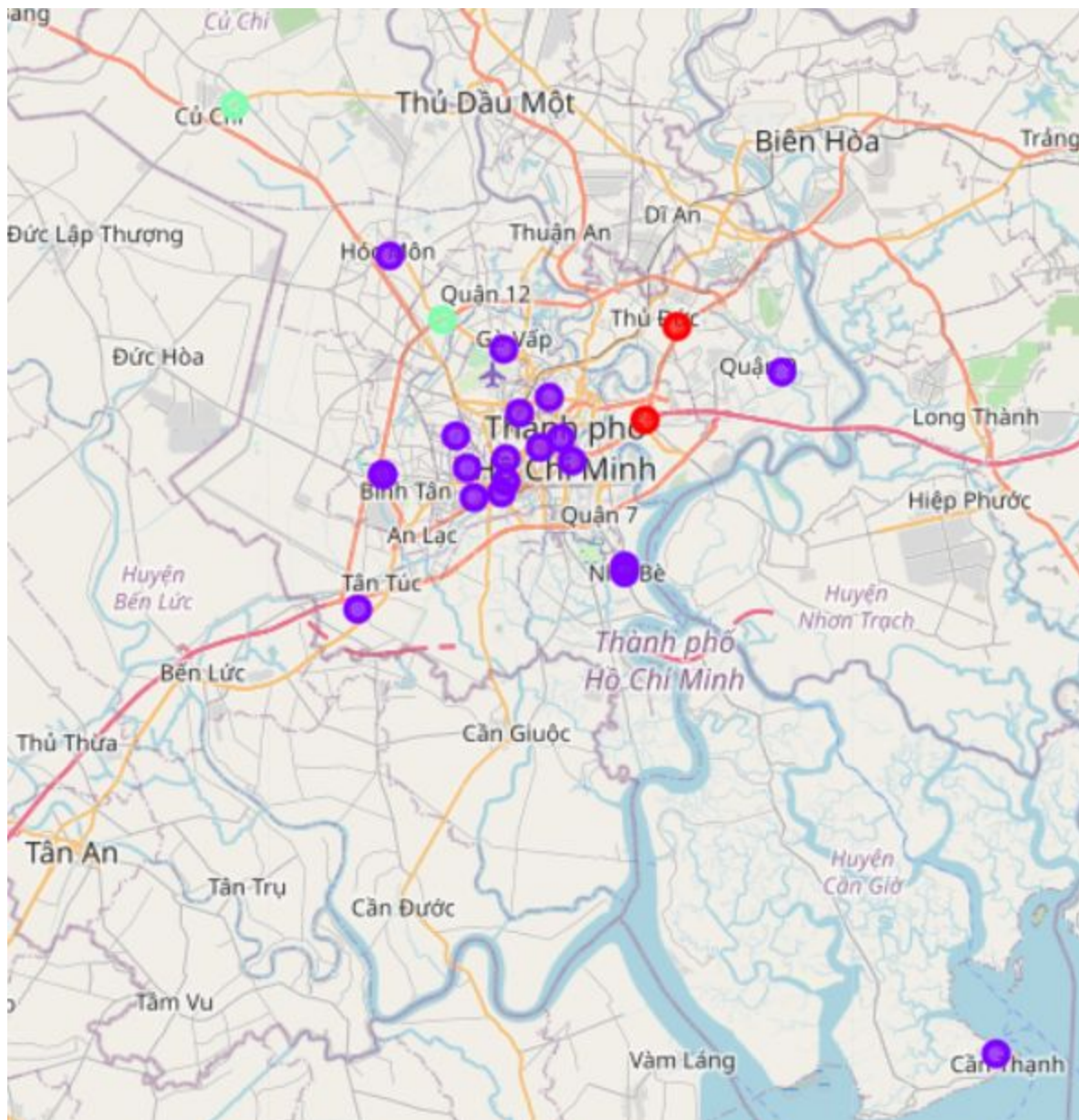
[https://en.wikipedia.org/wiki/Category:Districts\\_of\\_Ho\\_Chi\\_Minh\\_City](https://en.wikipedia.org/wiki/Category:Districts_of_Ho_Chi_Minh_City).

After that, I will use the Python request and get the Geographical coordinates of each district from Python Geocoder. Then, I will use FourSquare API to explore the venues in these districts. Lastly, I will focus on these Districts based on their number of restaurants to clarify which district has an acceptable degree of competitiveness for my restaurant. The details will be given in the final report.

## Results

I used K Means for conducting a clustering analysis based on the data retrieved. There are three clusters based on the data.

- The first cluster is the one with almost no recognized restaurants (Purple)
- The second one is the one with a moderate number of restaurants (Red)
- The third one is the one with a high density of restaurants. (Green)



The results are shown in the map above.



	District	Restaurant	Cluster Labels	Latitude	Longitude
0	Bình Chánh District	0.00	1	10.67922	106.57654
20	Phú Nhuận District	0.00	1	10.79565	106.67464
19	Nhà Bè District	0.00	1	10.70153	106.73818
18	Hóc Môn District	0.00	1	10.88839	106.59642
17	Gò Vấp District	0.00	1	10.83379	106.66557
16	District 9	0.00	1	10.82004	106.83185
15	District 8	0.00	1	10.74771	106.66334
14	District 7	0.00	1	10.70515	106.73748
13	District 6	0.00	1	10.74597	106.64769
12	District 5	0.00	1	10.75569	106.66637
11	District 4	0.01	1	10.76670	106.70650
10	District 3	0.01	1	10.77567	106.68670
7	District 11	0.00	1	10.76316	106.64314
6	District 10	0.00	1	10.76867	106.66564
5	District 1	0.01	1	10.78095	106.69911
3	Cần Giờ District	0.00	1	10.41566	106.96130
2	Bình Tân District	0.00	1	10.75909	106.59150
1	Bình Thạnh District	0.00	1	10.80518	106.69280
22	Tân Bình District	0.00	1	10.75909	106.59150
23	Tân Phú District	0.00	1	10.78232	106.63667

#### Custer Purple

	District	Restaurant	Cluster Labels	Latitude	Longitude
8	District 12	0.222222	2	10.85043	106.62732
4	Củ Chi District	0.333333	2	10.97734	106.50223

#### ClusterGreen

	District	Restaurant	Cluster Labels	Latitude	Longitude
21	Thủ Đức District	0.052632	0	10.84626	106.76992
9	District 2	0.105263	0	10.79199	106.74985

#### Cluster Red

## Discussion

A glance at the generated map reveals that the majority of the restaurants in HCMC are situated up North and East of the city. Cluster 1 has almost no recognized restaurants while cluster 2 has the highest number of recognized restaurants. Last not least, cluster 0 has a moderate number of recognized restaurants. As a result, Opening a restaurant in districts in cluster 1 and 0 are more challenging due to the higher competitiveness. More importantly, it is significant that although most districts in Cluster 1 are closer to the city center, they have very few recognized restaurants. Therefore, it is better to open a restaurant in Cluster 1.

## **Limitations and suggestions for future research**

Since Vietnam is still a developing country and the data provided by Foursquare is still inferior. There are definitely more restaurants in downtown districts but they are not recognized by Foursquare. As a result, the outcome of the report is considered a reference only. In future research, we must do category mapping after retrieving categories from API to generalize the numbers of the restaurant categories. For Example, Steakhouse is also a restaurant but FourSquare does not list them in the restaurant category..

## **Conclusion**

In conclusion, districts in the southern area and proximity to downtown are ideal for opening a restaurant due to the minimal density of restaurants there. Especially District 4.