Capstone Project: The Battle of Neighborhood
Data analysis for dertemining locations for coffee shops in three Vietnam major cities

1. Introduction

1.1. Background

   Vietnam is the developing country and income of young people have been increasing. Therefore, they tend to live in the big cities with modern facilities and entertainment. Coffee shops are considered as a new culture in this country where people can work, study or meet their clients, etc.

   Taking it as an opportunity, the report helps to provide data visualization for businesses to find suitable locations for coffee shops.

1.2. Objective

 Find suitable location for openning new coffee shops in three major cities of Vietnam.

1.3. Methods

 • Find suitable districts by K-mean clustering.

 • Using data of venues and population density in each district to determine the districts having related business and attracting people to come.

2. Data collection and preparation
   In this project, data sets will be used:
- Data of districts in 3 cities (Hanoi, Ho Chi Minh, Danang) from the source Wikipedia: Name,
Latitude/Longitude and population.
- Foursquare API

2.1 Import data of districts in cities

	District	City	Area	Population	Latitude	Longitude
0	Ba Dinh	Ha Noi	9.224	247100	21.036667	105.836111
1	Bac Tu Liem	Ha Noi	43.350	333300	21.070351	105.762098
2	Cau Giay	Ha Noi	12.040	266800	21.028307	105.790783
3	Dong Da	Ha Noi	9.960	420900	21.013825	105.828677
4	Hai Ba Trung	Ha Noi	10.090	318000	21.011667	105.847778

Figure 1 Data from Wikipedia show information of districts in 3 major cities of Vietnam

Calculate population destiny

	District	City	Area	Population	Latitude	Longitude	Density
0	Ba Dinh	Ha Noi	9.224	247100	21.036667	105.836111	26788.811795
1	Bac Tu Liem	Ha Noi	43.350	333300	21.070351	105.762098	7688.581315
2	Cau Giay	Ha Noi	12.040	266800	21.028307	105.790783	22159.468439
3	Dong Da	Ha Noi	9.960	420900	21.013825	105.828677	42259.036145
4	Hai Ba Trung	Ha Noi	10.090	318000	21.011667	105.847778	31516.352825

Figure 2: Data includes district names, cities, area, population, latitude and longgitude, density
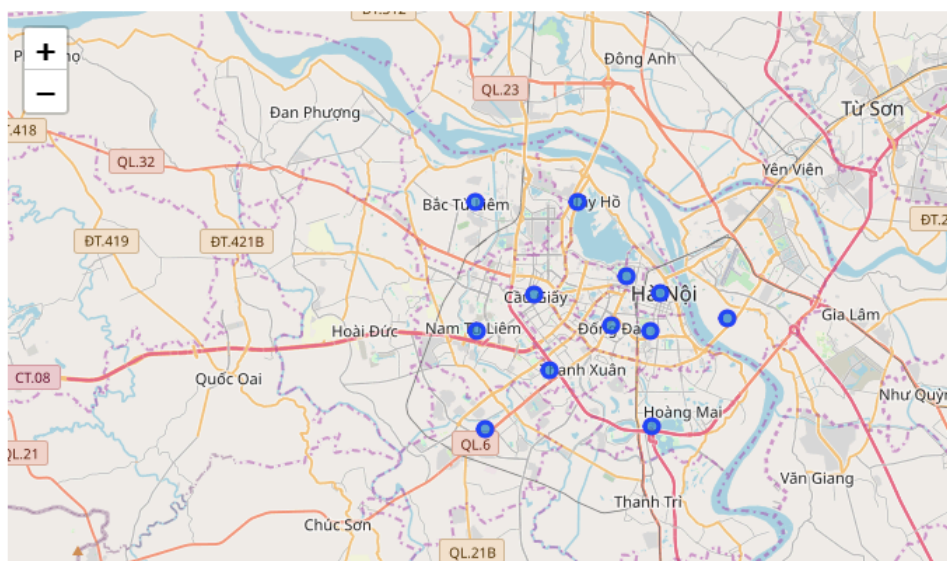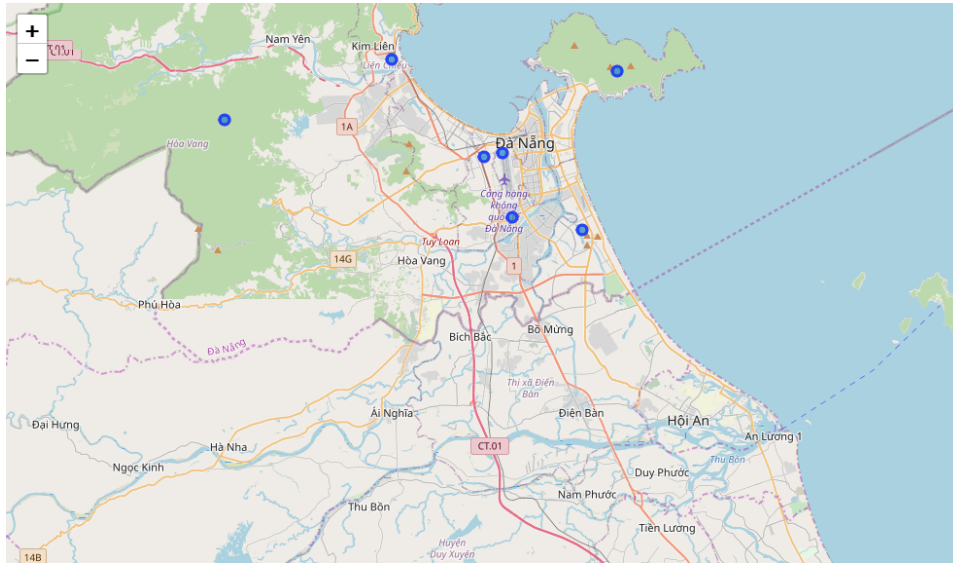
Plot districts
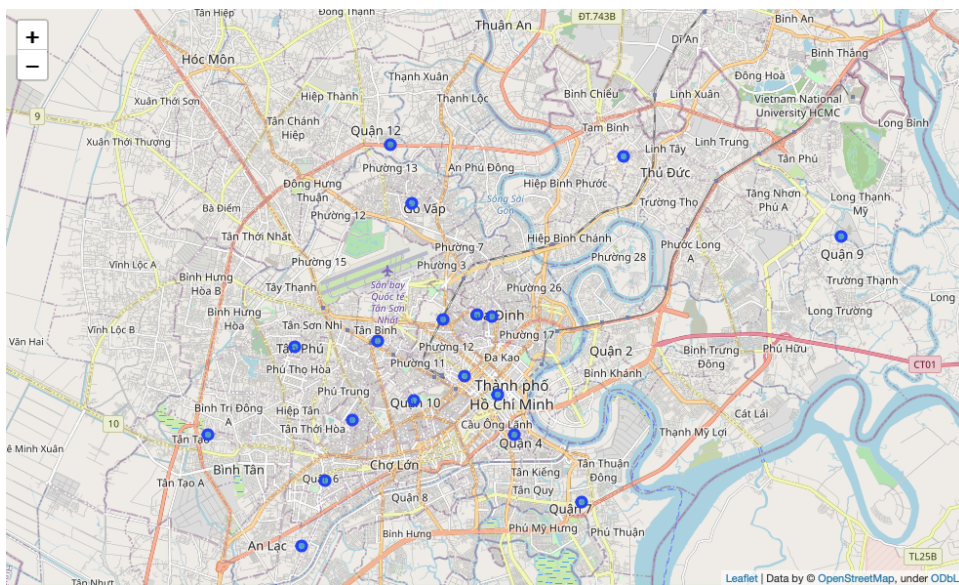

Figure 3 Hanoi districts

Figure 4 Danang districts

Figure 5 Ho Chi Minh district

2.2 Explorer districts in Cities using Foursquare API
Popular venues

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
376	Cam Le	16.016667	108.202778	Mỳ quảng Dì vân	16.018154	108.204996	Asian Restaurant
377	Cam Le	16.016667	108.202778	A4 Cafe	16.014535	108.199816	Café
378	Cam Le	16.016667	108.202778	100K Studio	16.016044	108.198364	Art Gallery
379	Hai Chau	16.060556	108.196111	Đầm Sen Quán	16.062066	108.191925	Vietnamese Restaurant
380	Lien Chieu	16.123889	108.117778	Ga Liên Chiểu	16.121328	108.120861	Train Station

Figure 6 Popular venues using foursquare API

District	Venue Category
Ba Dinh	9
Bac Tu Liem	1
Binh Tan	1
Binh Thanh	14
Cam Le	5
Cau Giay	3
Dong Da	16
Go Vap	3
Ha Dong	4
Hai Ba Trung	32
Hai Chau	1
Hoan Kiem	95
Hoang Mai	3
Lien Chieu	1
Long Bien	2
Nam Tu Liem	2
Phu Nhuan	10
Quan 1	2
Quan 10	26
Quan 11	4
Quan 12	3
Quan 3	47
Quan 4	22
Quan 5	37
Quan 6	2
Quan 7	4
Quan 8	4
Tan Binh	7
Tan Phu	9
Tay Ho	4
Thanh Xuan	2
Thu Duc	6

Figure 7 Summary number of venues for districts

	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	Ba Dinh	Coffee Shop	Monument / Landmark	Palace	History Museum	Historic Site	Temple	Castle	Plaza	Zoo Exhibit	Diner
1	Bac Tu Liem	Vietnamese Restaurant	Zoo Exhibit	Food Truck	Dim Sum Restaurant	Diner	Dog Run	Electronics Store	Event Space	Farm	Fast Food Restaurant
2	Binh Tan	Business Service	Zoo Exhibit	Dessert Shop	Diner	Dog Run	Electronics Store	Event Space	Farm	Fast Food Restaurant	Flea Market
3	Binh Thanh	Vegetarian / Vegan Restaurant	Café	French Restaurant	Food	Food Truck	Fast Food Restaurant	Massage Studio	Gym	Spa	Coffee Shop
4	Cam Le	Café	Art Gallery	Asian Restaurant	Zoo Exhibit	French Restaurant	Dog Run	Electronics Store	Event Space	Farm	Fast Food Restaurant

Figure 8 Top most common venues per district

3. **Cluster Analysis**

K-means clustering was conducted to group districts according to their popular venues using Foursquare data

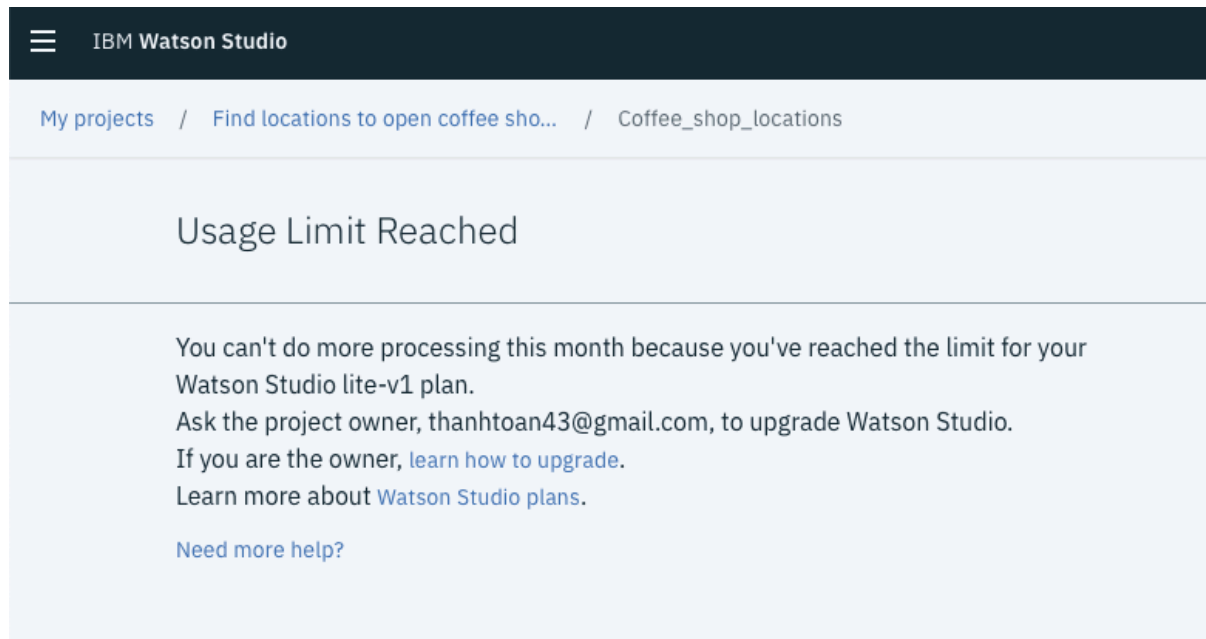|	District	Arepa Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	BBQ Joint	Bakery	Bar	Baseball Stadium	Basketball Court	Basketball Stadium	Beer Garden	Bistro	Booksto
0	Ba Dinh	0.0000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000
1	Bac Tu Liem	0.0000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000
2	Binh Tan	0.0000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000
3	Binh Thanh	0.0000	0.0	0.000000	0.071429	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000
4	Cam Le	0.0000	0.2	0.000000	0.200000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.000
5	Cau Giay	0.0000	0.0	0.000000	0.000000	0.333333	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000
6	Dong Da	0.0625	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000
7	Go Vap	0.0000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000
8	Ha Dong	0.0000	0.0	0.000000	0.000000	0.000000	0.250000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000
9	Hai Ba Trung	0.0000	0.0	0.000000	0.000000	0.062500	0.000000	0.000000	0.00	0.000000	0.00	0.031250	0.000000	0.0000
10	Hai Chau	0.0000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000
11	Hoan Kiem	0.0000	0.0	0.010526	0.000000	0.000000	0.010526	0.010526	0.00	0.000000	0.00	0.000000	0.010526	0.0000
12	Hoang Mai	0.0000	0.0	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.00	0.000000	0.000000	0.0000

Figure 9 Result of one hot coding

	District	City	Area	Population	Latitude	Longitude	Density	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Ba Dinh	Ha Noi	9.224	247100	21.036667	105.836111	26788.811795	1.0	Coffee Shop	Monument / Landmark	Palace	History Museum	Historic Site
1	Bac Tu Liem	Ha Noi	43.350	333300	21.070351	105.762098	7688.581315	2.0	Vietnamese Restaurant	Zoo Exhibit	Food Truck	Dim Sum Restaurant	Diner
2	Cau Giay	Ha Noi	12.040	266800	21.028307	105.790783	22159.468439	1.0	Japanese Restaurant	BBQ Joint	Park	Zoo Exhibit	Food Truck
3	Dong Da	Ha Noi	9.960	420900	21.013825	105.828677	42259.036145	1.0	Café	Arepa Restaurant	Park	Notary	Coffee Shop
4	Hai Ba Trung	Ha Noi	10.090	318000	21.011667	105.847778	31516.352825	1.0	Japanese Restaurant	Coffee Shop	Vietnamese Restaurant	BBQ Joint	Café
5	Ha Dong	Ha Noi	47.917	319800	20.966667	105.766667	6674.040528	0.0	Café	Bakery	Zoo Exhibit	French Restaurant	Diner
6	Hoan Kiem	Ha Noi	5.290	160600	21.028889	105.8525	30359.168242	1.0	Hotel	Coffee Shop	Café	Vietnamese Restaurant	Noodle House
7	Hoang Mai	Ha Noi	41.040	411500	20.968056	105.848333	10026.803119	1.0	Zoo Exhibit	Park	Women's Store	Arts & Crafts Store	French Restaurant
8	Long Bien	Ha Noi	60.380	291900	21.0175	105.8858	4834.382246	1.0	Soccer Field	Farm	Zoo Exhibit	Cupcake Shop	Dim Sum Restaurant
9	Nam Tu Liem	Ha Noi	32.270	236700	21.011526	105.762449	7334.986055	0.0	Café	Mattress Store	Zoo Exhibit	Food Truck	Diner

Figure 10 the top 5 most common venue category in each district

Plot district grouped by K-mean

Sorry I can't contine because of this dump limitation and no teaching assitants would help me!!!

IBM Watson Studio

My projects / Find locations to open coffee sho… / Coffee_shop_locations

Usage Limit Reached

You can't do more processing this month because you've reached the limit for your Watson Studio lite-v1 plan.
Ask the project owner, thanhtoan43@gmail.com, to upgrade Watson Studio.
If you are the owner, learn how to upgrade.
Learn more about Watson Studio plans.

Need more help?

4. Conclusion

From cluster analysis, 2 out of 5 cluster are suitable for open coffee shops including cluster 1 and 2. Common venues found in these cluster, such as coffee shops, cafe, and restaurants, indicated high change of more target customers. Finally, five districts that have higher population in cluster were selected, including Hoan Kiem (Hanoi), Hai Ba Trung (Hanoi), Quan 3 (Ho Chi Minh), Quan 5 (Ho Chi Minh), Quan 10 (Ho Chi Minh).