Applied Data Science Capstone

Final Report

OPENING A COFFEE SHOP IN HO CHI MINH CITY (VIETNAM)

A Battle of Neighborhood

Name: VO VAN VIET

CONTENTS

1/ INTRODUCTION:

Ho Chi Minh City (Vietnamese: Thành phố Hồ Chí Minh), also commonly referred to as Saigon, is the largest city of Vietnam. Located in southeastern Vietnam, the city surrounds the Saigon River and covers about 2,061 square kilometers (796 square miles). Ho Chi Minh City is the economic and financial center of Vietnam, and plays an important role in the country's culture and scientific developments.

2/ BUSINESS PROBLEM:

A Local Coffee Shop wants to open their first store in Ho Chi Minh City. The new store would be located in an ideal place with crowded customers, acceptable housing prices, and low density of other coffee shops.

By using Data Science, they can give good answers to the following questions:

- Where should we locate the shop?
- Which districts have the lowest housing price?
- How many other coffee shops in each district?
- Which districts have low number of coffee shops, and big populations?

3/ DATA:

To find the ideal place for the new coffee shop, we need the following data:

- List of Districts in Ho Chi Minh City contains district names, and their population, area, density from wikipedia (https://en.wikipedia.org/wiki/Ho_Chi_Minh_City)
- List of Housing Price in Ho Chi Minh City from Mogi (https://mogi.vn/gia-nha-dat)
- Foursquare API to get the venues.

4/ METHODOLOGY:

- First, by using the library BeautifulSoup to scrape the data from wikipedia, the list of districts had been collected.
- Then, the density had been calculated by the subtraction of population and area.

- geopy.geocoders.Nominatim had been used to get the coordinates of districts and add them to the main data frame.
- The library Folium had been used to sketch the map of the districts.
- Foursquare API to explore the venues in each district and segment the districts based on them.
- Clustering the venues "Coffee Shop" by the algorithm K-Mean Clustering from the library Scikit-Learn.

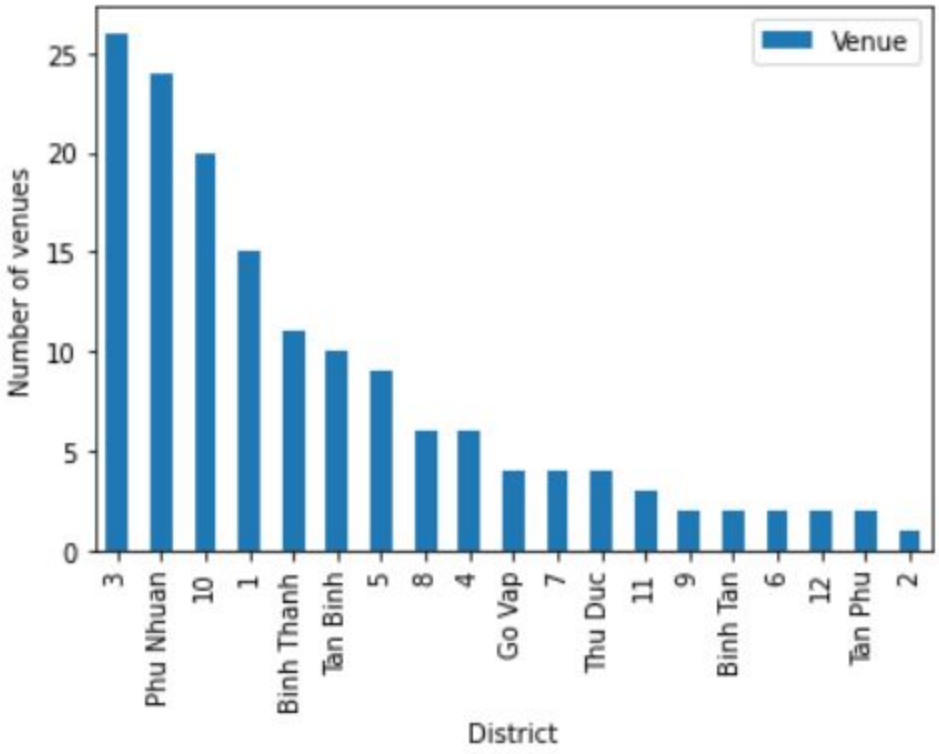
5/ RESULTS:

5.1/ The Main DataFrame:

By manipulating the data from the previous sources, the final dataset that mainly used in the project is below:

	District	Coffee Shop	Cluster Labels	Population	Density (pop/m2)	Average Housing Price (Million VND)	Latitude	Longitude
2	11	0.000000	0	230596	44863.035	164.0	10.764208	106.643282
3	12	0.000000	0	510326	9668.928	46.5	10.867233	106.653930
4	2	0.000000	0	147168	2958.745	81.4	10.791116	106.736729
6	4	0.038462	0	186727	44671.531	109.0	10.759243	106.704890
8	6	0.000000	0	258945	36014.604	114.0	10.746928	106.634495
11	9	0.000000	0	290620	2549.298	48.8	10.849307	106.802055
12	Binh Tan	0.000000	0	686474	13229.408	60.0	10.749809	106.605664
10	8	0.300000	1	431969	22521.846	70.9	10.740400	106.665843
0	1	0.060000	2	193632	25049.418	418.0	10.774540	106.699184
16	Tan Binh	0.108108	2	459029	20510.679	145.0	10.797979	106.653805
15	Phu Nhuan	0.142857	2	182477	37392.828	182.0	10.800118	106.677042
14	Go Vap	0.083333	2	634146	32124.924	96.3	10.840150	106.671083
13	Binh Thanh	0.083333	2	487985	23506.021	134.0	10.804659	106.707848
9	7	0.057143	2	310178	8690.894	90.6	10.736573	106.722432
7	5	0.071429	2	178615	41830.211	254.0	10.756129	106.670376
5	3	0.080000	2	196333	39905.081	265.0	10.783529	106.687098
1	10	0.128571	2	238558	41705.944	211.0	10.773198	106.667833
17	Tan Phu	0.142857	2	464493	28922.354	102.0	10.791640	106.627302
18	Thu Duc	0.083333	2	528413	10619.232	64.8	10.822023	106.718302

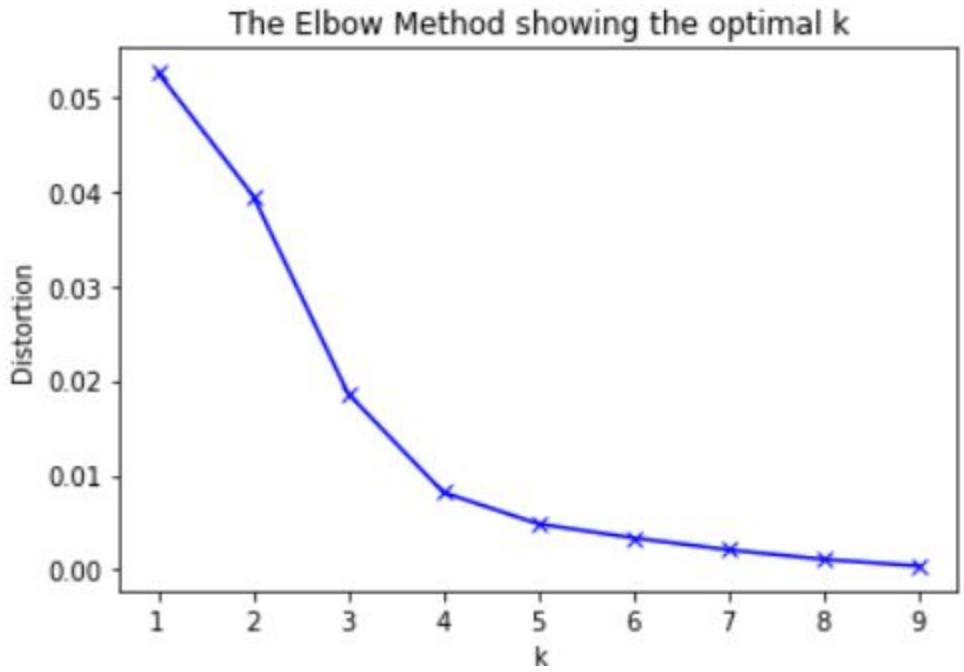
5.2/ Coffee Shop per District:



The following chart illustrates the number of coffee shops per district. District 3, 10, 1, Phu Nhuan had the biggest number of coffee shops. District 2, 12, 6, 9, Tan Phu, Binh Tan had the lowest number of shops.

5.3/ Clustering:

Number of K:

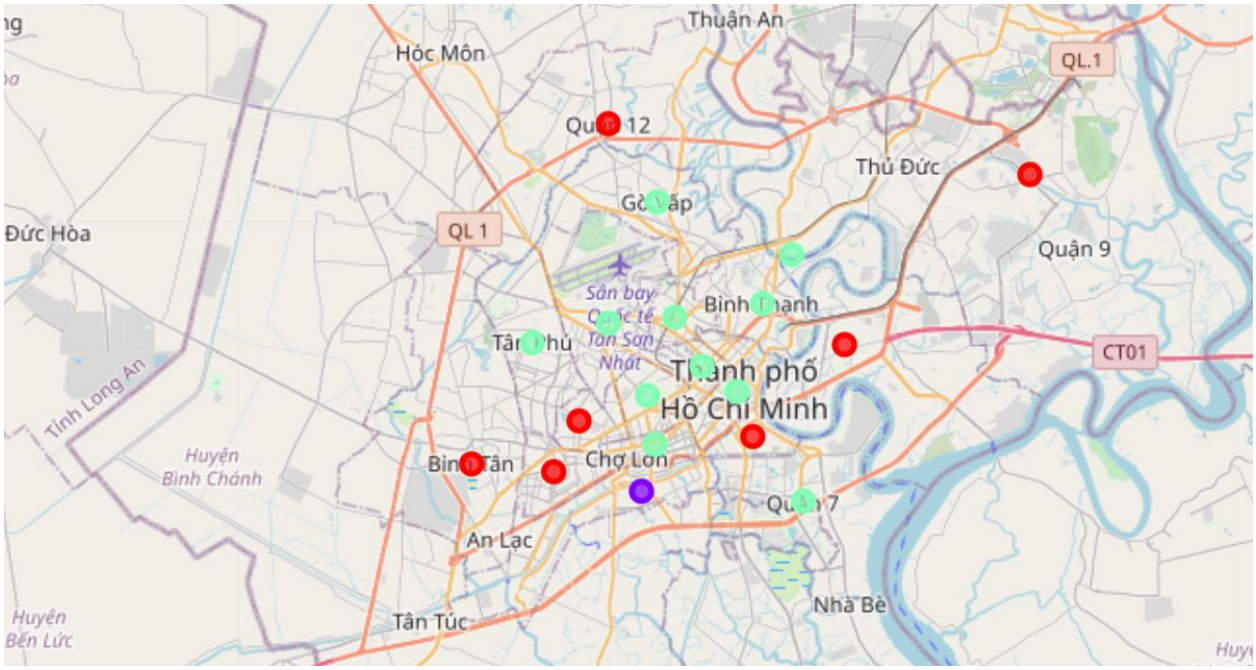


K=3 is the best

After the K-Means algorithm has been run, there are 3 clusters:

- Cluster 0: There are not many coffee shops in these districts. (red)
- Cluster 1: The number of coffee shops in these districts is high. (blue)
- Cluster 2: The number of coffee shops in these districts is medium. (green)

Plotting the clusters on the map:



5.4/ The Housing Price:

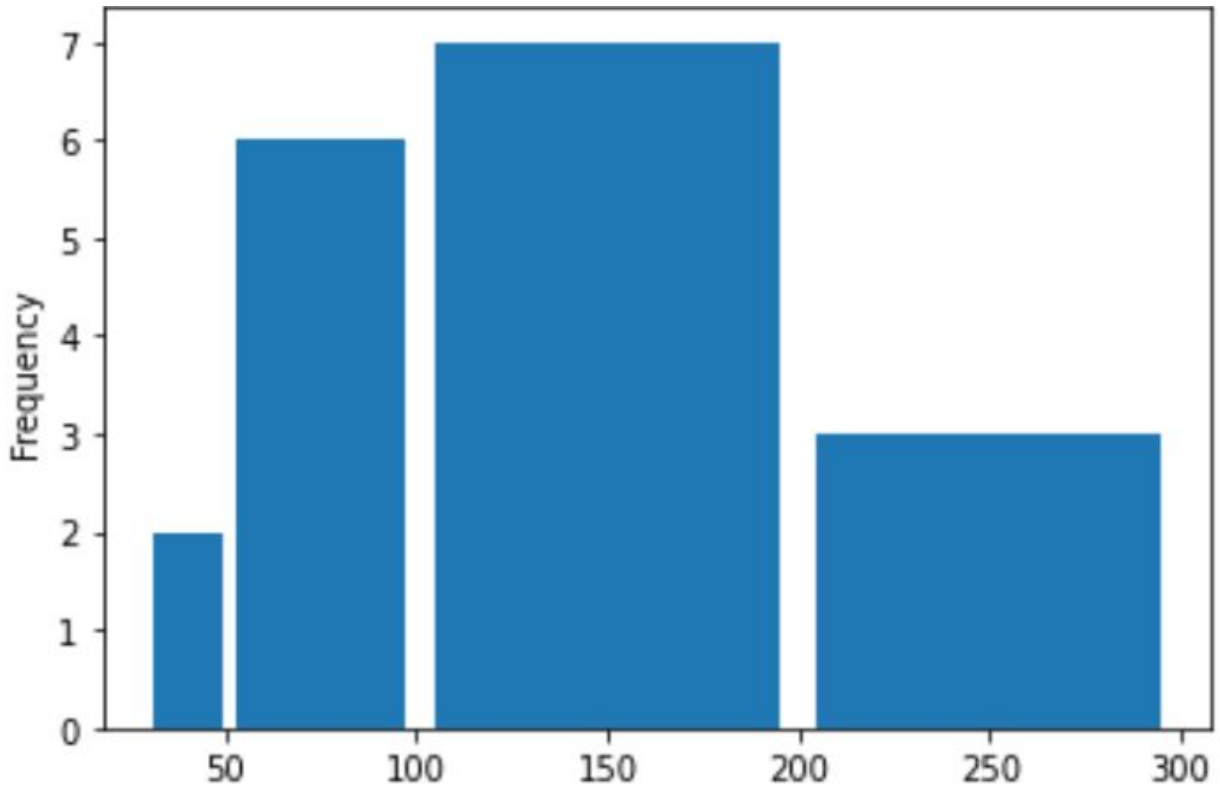
Range of AHP:

Low: 30mVND < AHP < 50mVND.

Medium: 50mVND < AHP < 100mVND.

High: 100mVND < AHP < 200mVND.

Very High: AHP > 200mVND.



5.5/ Final Result:

	District	Coffee Shop	Cluster Labels	Population	Density (pop/m2)	Average Housing Price (Million VND)	Latitude	Longitude	AHP Level
2	11	0.000000	0	230596	44863.035	164.0	10.764208	106.643282	High
3	12	0.000000	0	510326	9668.928	46.5	10.867233	106.653930	Low
4	2	0.000000	0	147168	2958.745	81.4	10.791116	106.736729	Medium
6	4	0.038462	0	186727	44671.531	109.0	10.759243	106.704890	High
8	6	0.000000	0	258945	36014.604	114.0	10.746928	106.634495	High
11	9	0.000000	0	290620	2549.298	48.8	10.849307	106.802055	Low
12	Binh Tan	0.000000	0	686474	13229.408	60.0	10.749809	106.605664	Medium
10	8	0.300000	1	431969	22521.846	70.9	10.740400	106.665843	Medium
0	1	0.060000	2	193632	25049.418	418.0	10.774540	106.699184	Very High
16	Tan Binh	0.108108	2	459029	20510.679	145.0	10.797979	106.653805	High
15	Phu Nhuan	0.142857	2	182477	37392.828	182.0	10.800118	106.677042	High
14	Go Vap	0.083333	2	634146	32124.924	96.3	10.840150	106.671083	Medium
13	Binh Thanh	0.083333	2	487985	23506.021	134.0	10.804659	106.707848	High
9	7	0.057143	2	310178	8690.894	90.6	10.736573	106.722432	Medium
7	5	0.071429	2	178615	41830.211	254.0	10.756129	106.670376	Very High
5	3	0.080000	2	196333	39905.081	265.0	10.783529	106.687098	Very High
1	10	0.128571	2	238558	41705.944	211.0	10.773198	106.667833	Very High
17	Tan Phu	0.142857	2	464493	28922.354	102.0	10.791640	106.627302	High
18	Thu Duc	0.083333	2	528413	10619.232	64.8	10.822023	106.718302	Medium

Group 1 (Not many coffee shops, low density): 12, 2, 9, Binh Tan.

Group 2 (Not many coffee shops, medium or high density, Medium or High AHP): 4, 6.

Group 3 (Medium coffee shops, low density): 7, Thu Duc.

Group 4 (Medium coffee shops, medium or high density, High or Very High AHP): Tan Phu, 10, 3, 5, Binh Thanh, Tan Binh, Phu Nhuan, 1.

Group 5 (Medium coffee shops, high density, medium AHP): Go Vap.

6/ Conclusion:

From all above results, we conclude that, the best place for us to set up a new coffee shop is in Go Vap district because there are a lot of people living there (high density), the number of already-working coffee shop is medium (cluster 2) and the average housing price is medium.