

BATTLECAPSTONE PROJECT: THE BATTLE OF THE NEIGHBORHOODS

Open A Coffee Shop In Ho Chi Minh City

NGUYEN TAN HUY

March 17, 2020

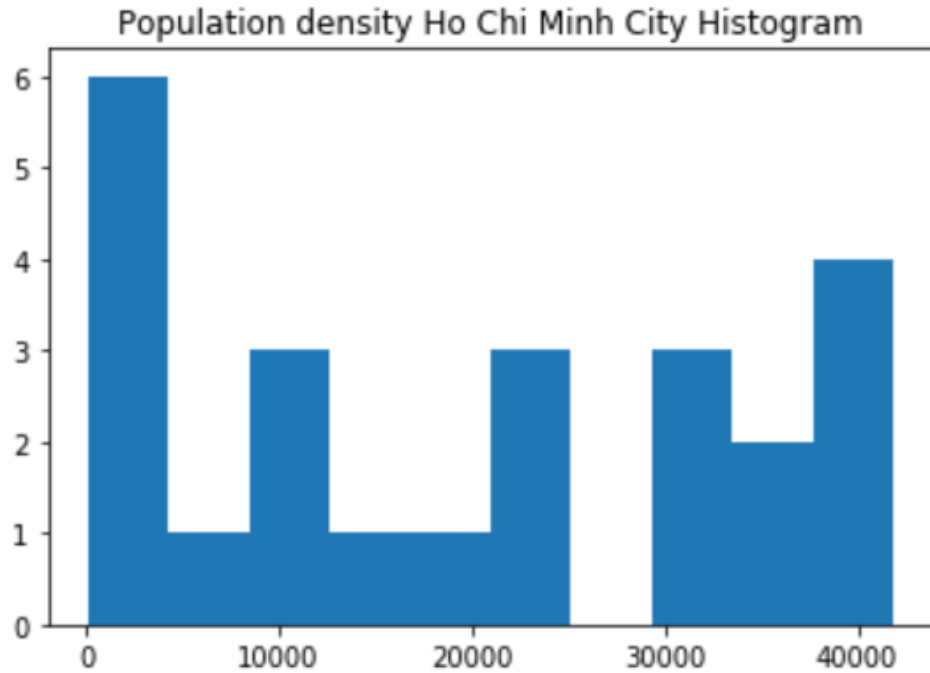
Using data location find out the best place open a new coffee shop in HCMC

- ▶ Problems:
 - ▶ There are 23 districts and 324 wards in HCMC with millions of coffee shop venues.
 - ▶ Are there opportunity for my company to open a new coffee shop in this area?
 - ▶ Where is a good place?

DATA ACQUISITION AND DATA CLEANING

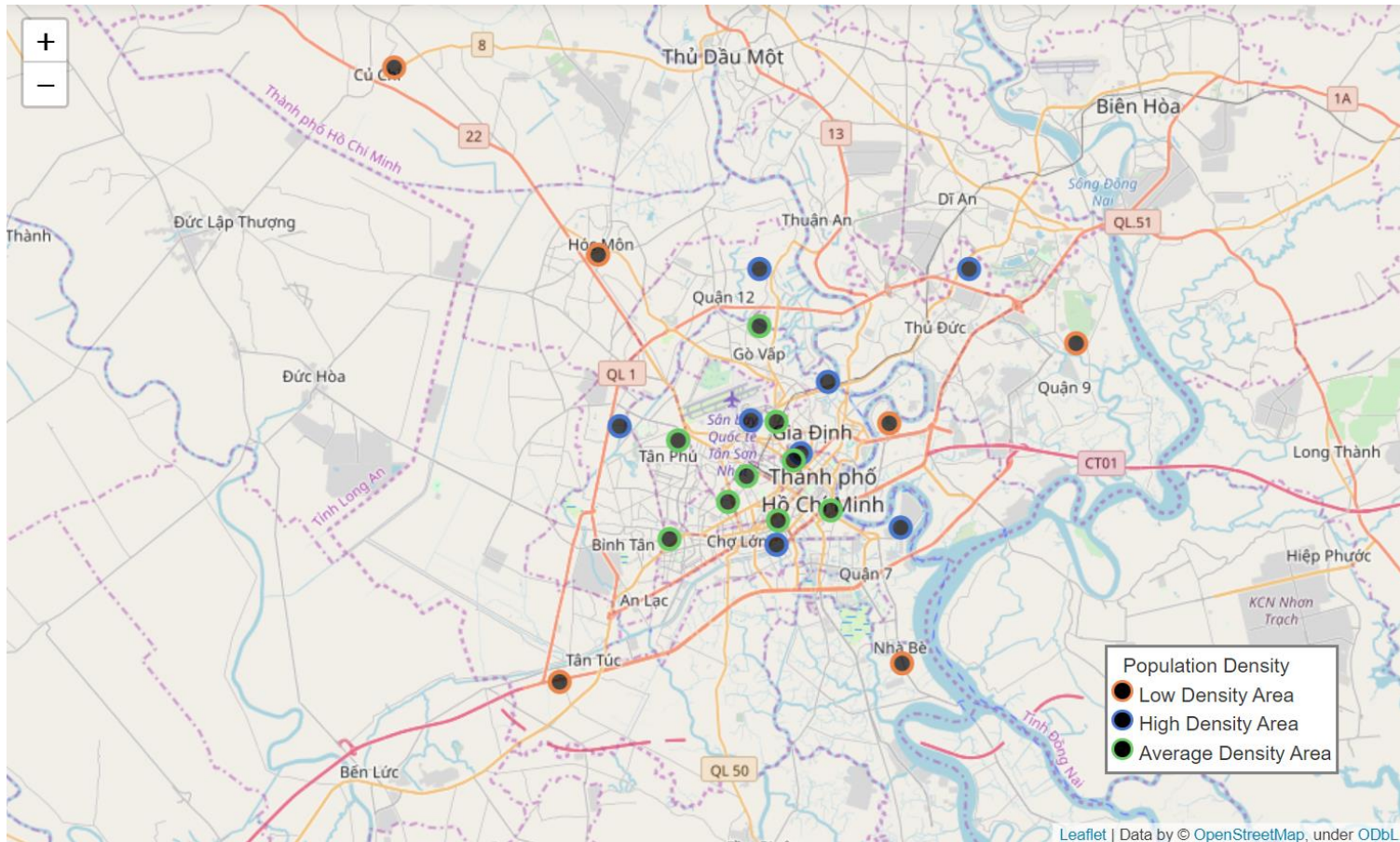
- ▶ Scraped data from [Wikipedia](#). Create tabular data include : District Code, District Name, Acreage, Population for each district.
- ▶ Download data from [official Ho Chi Minh City government website](#). Create tabular data include: Ward Code, Ward Name, District Name, Province/City Name.
- ▶ Getting coordinate from google map each wards and district.
- ▶ Combine two data frames into one data frame.
- ▶ Getting venues using foursquare API with three categories for each ward: coffee shop venues, short term venues, long term venues categories. Getting categories character from [foursquare doc](#).
- ▶ Create one data frame contain : District Code, District Name, Ward Code, Ward Name, Acreage, Population, Longitude and Latitude each ward.

Population Density in HCMC



- Low density : < 10000 persons per acreage (KM²)
- Average density : between 10000 and 30000 persons acreage (KM²)
- High density : > 30000 persons acreage (KM²)

Population Density in HCMC



Open new coffee shop in HCMC downtown which is average and high population density

Find the best place to open a new coffee shop

- Using Kmean to cluster our district with 4 cluster param.

Out[73]:

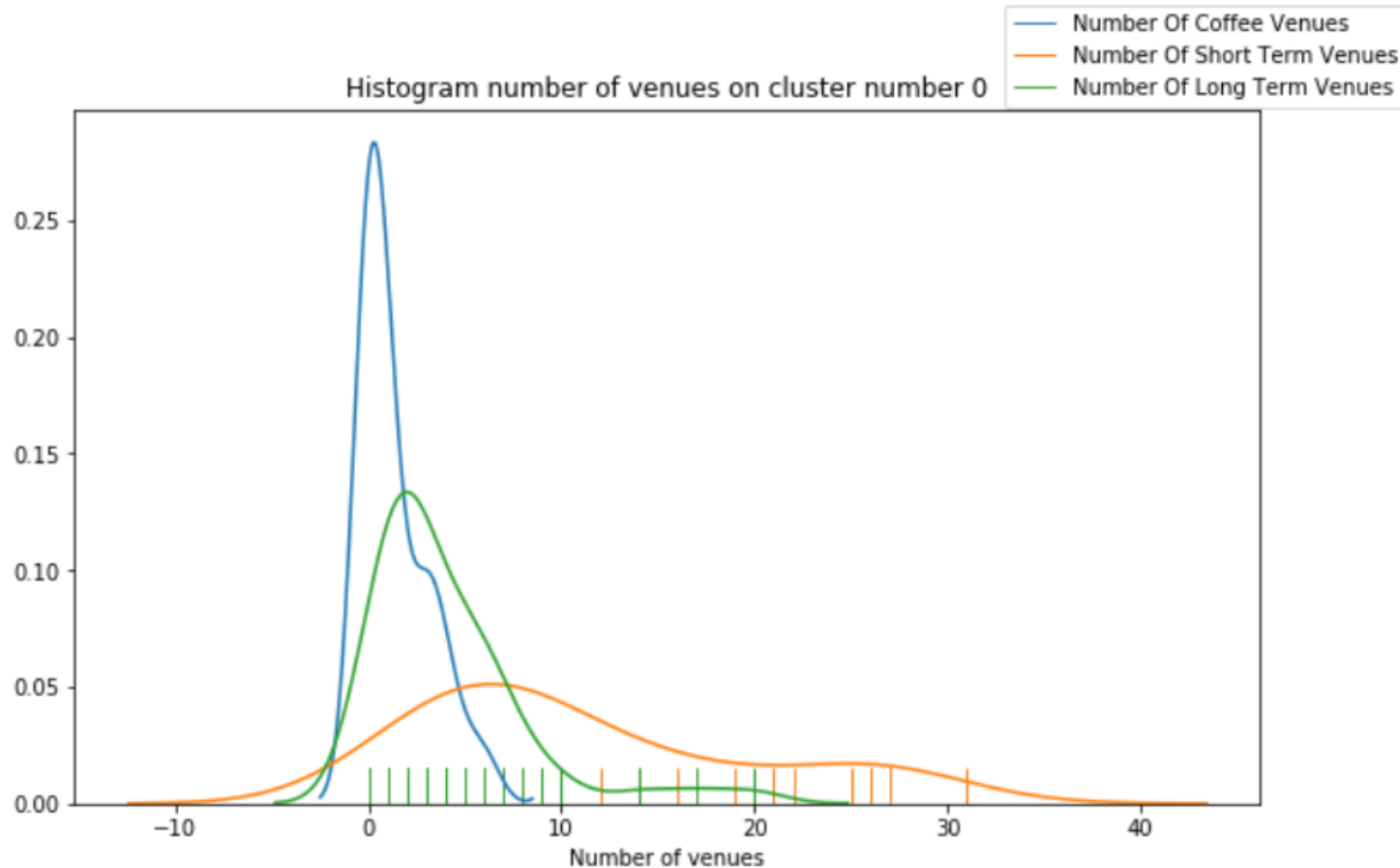
	District Code	District Name	Acreage/Coffee	Acreage/ShortTerm	Acreage/LongTerm	Population/Coffee	Population/ShortTerm	Population/LongTerm	k_Cluster
0	760	Quận 1	17.1556	21.2672	18.2506	315.556	391.185	335.697	0
1	761	Quận 12	224.426	108.967	196.059	2638.3	1280.99	2304.83	3
2	762	Quận Thủ Đức	206.926	95.6	185.271	2562.77	1184	2294.57	3
3	763	Quận 9	525.346	253.898	326.648	1829.49	884.187	1137.54	0
4	764	Quận Gò Vấp	40.5133	25.7908	45.6713	1388.09	883.66	1564.81	0
5	765	Quận Bình Thạnh	24.1067	22.4406	24.3041	578.886	538.877	583.626	0
6	766	Quận Tân Bình	31.2396	36.7705	31.3706	660.167	777.049	662.937	0
7	767	Quận Tân Phú	40.7398	29.1956	42.7005	1237.24	886.654	1296.79	0
8	768	Quận Phú Nhuận	7.58942	6.50667	7.79553	253.499	217.333	260.383	0
9	769	Quận 2	212.778	127.34	146.441	769.231	460.358	529.412	0
10	770	Quận 3	7.43202	7.03863	8.25503	287.009	271.817	318.792	0
11	771	Quận 10	9.0938	7.62667	8.62745	372.019	312	352.941	0
12	772	Quận 11	20.9796	6.9932	17.1333	853.061	284.354	696.667	0
13	773	Quận 4	17.7119	6.211	12.8615	741.525	260.03	538.462	0
14	774	Quận 5	12.0621	5.69333	8.10247	449.153	212	301.708	0
15	775	Quận 6	63.75	12.3958	44.625	2080.36	404.514	1456.25	0
16	776	Quận 8	65	28.9985	43.73	1442.18	643.399	970.252	0
17	777	Quận Bình Tân	230.177	116.637	199.31	3469.03	1757.85	3003.83	3
18	778	Quận 7	97.7808	91.2788	85.7933	986.301	920.716	865.385	0
19	783	Huyện Củ Chi	7246.17	1274.99	3423.39	7700	1354.84	3637.8	2
20	784	Huyện Hóc Môn	565.648	295.854	791.087	2808.29	1468.83	3927.54	3
21	785	Huyện Bình Chánh	2745.22	825.359	1350.59	7663.04	2303.92	3770.05	2
22	786	Huyện Nhà Bè	1304.29	554.862	707.254	2675.32	1138.12	1450.7	3
23	787	Huyện Cần Giờ	78272.2	7915.17	14676	7888.89	797.753	1479.17	1

We can easily observe that cluster number 0 with low number of Acreage per venues and low number of Population per venues. This means we have many venues in this cluster. Removed cluster number 0 from data frame.

Find the best place to open a new coffee shop

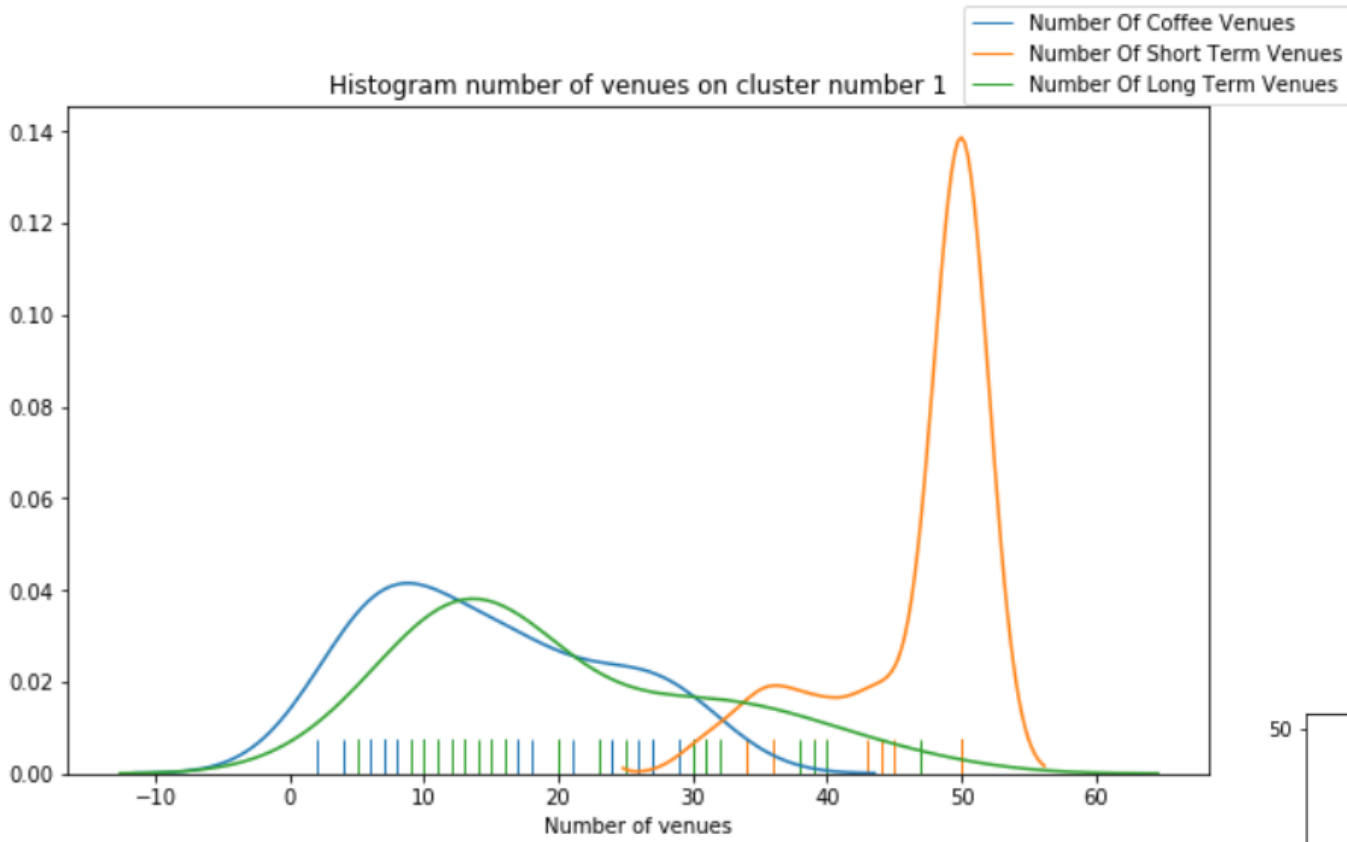
- ▶ Cluster Wards in remaining district.
- ▶ Using Kmean with 3 cluster param.
- ▶ Analyze and find the best place.

Find the best place to open a new coffee shop



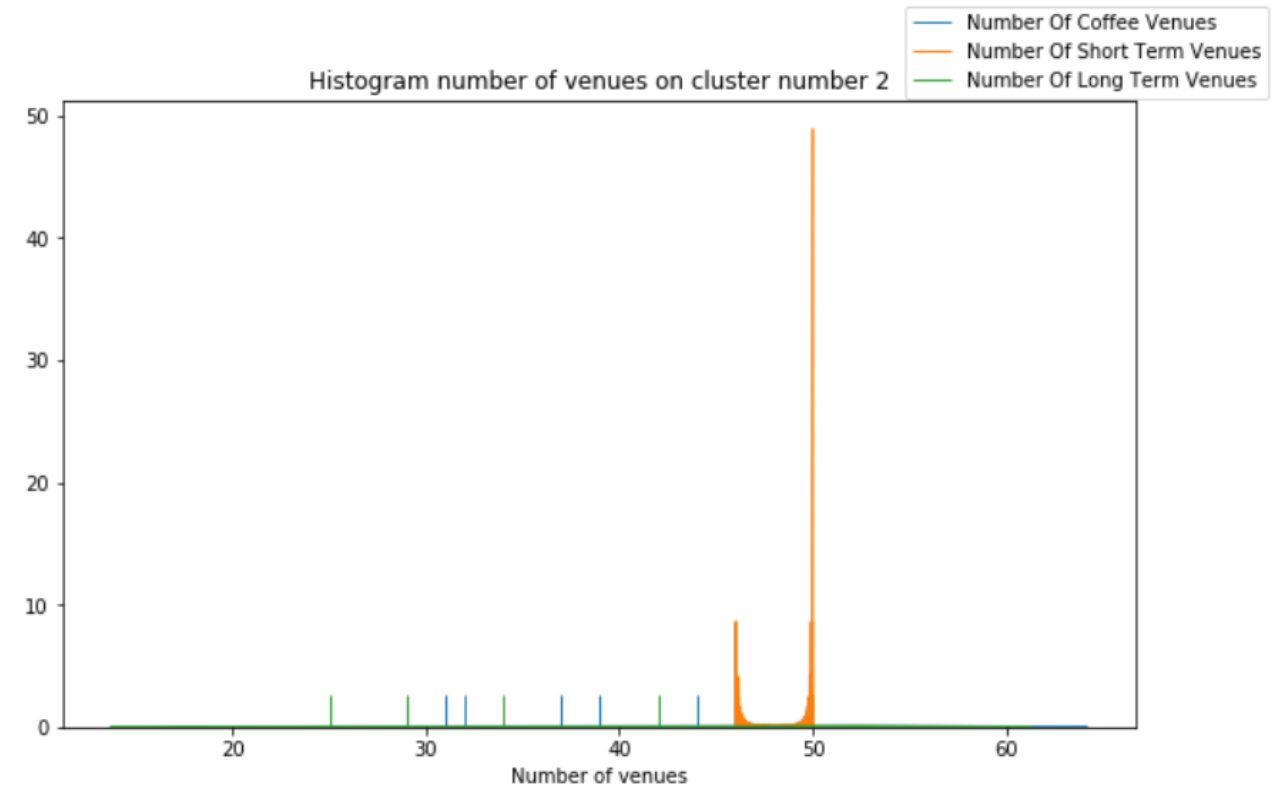
Low number of venues. Removed this ward cluster from dataframe

Histogram number of venues on cluster number 1

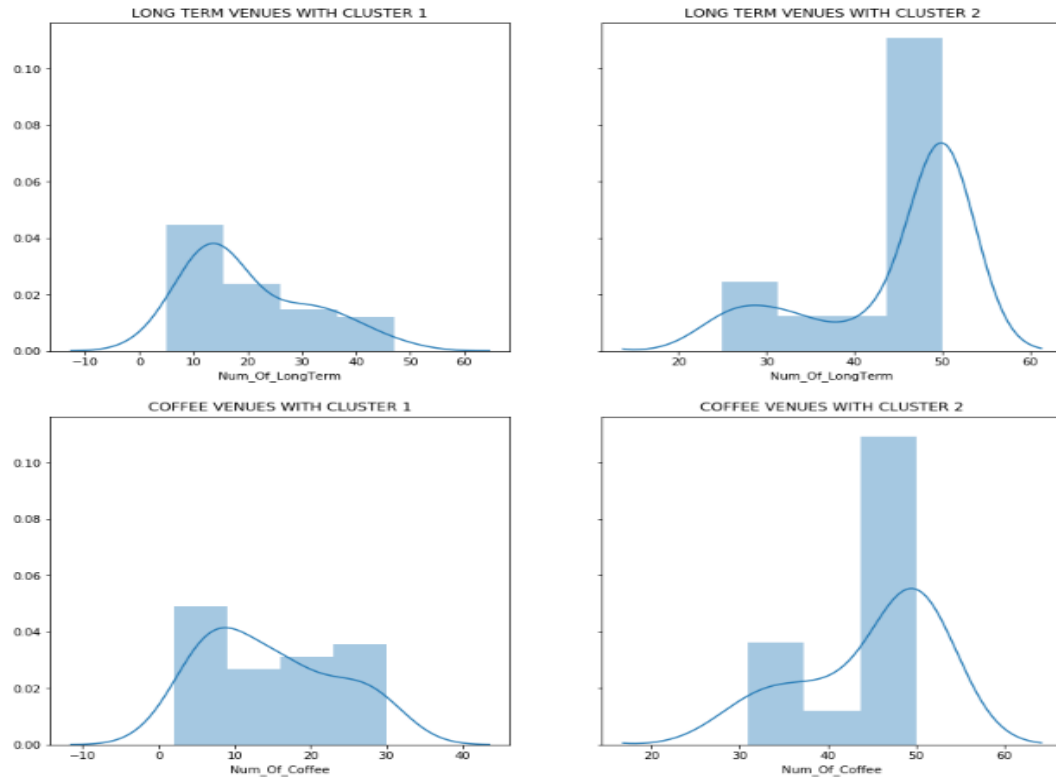


High number of venues. Keep this cluster

Histogram number of venues on cluster number 2



Analyze wards in the remaining cluster.



Conclusion: Cluster number 1 with low number of long term venues than cluster number 2. We also have low number of coffee shops in cluster number 1 than cluster number 2. So we will choose cluster number 1 with low competitors but only choose the wards with high number of long term venues in this cluster (higher than 30 venues).

Discussion and Results

```
In [99]: # Check if wards on average population density
after_visualize_df[after_visualize_df['District Code'].isin(average_density_df['District Code'])]
```

Out[99]:

	District Code	District Name	Ward Code	Ward Name	District Population	District Acreage (Km2)	District Latitude	District Longitude	Num_Of_Coffee	Num_Of_ShortTerm	Num_Of_LongTerm	k_cluster
22	777	Quận Bình Tân	27436	Phường Bình Hưng Hòa	784000	52.02	10.805543	106.603626	26	50	32	1
24	777	Quận Bình Tân	27442	Phường Bình Hưng Hoà B	784000	52.02	10.805374	106.603626	27	50	32	1
30	777	Quận Bình Tân	27460	Phường An Lạc	784000	52.02	10.728195	106.614303	17	44	39	1

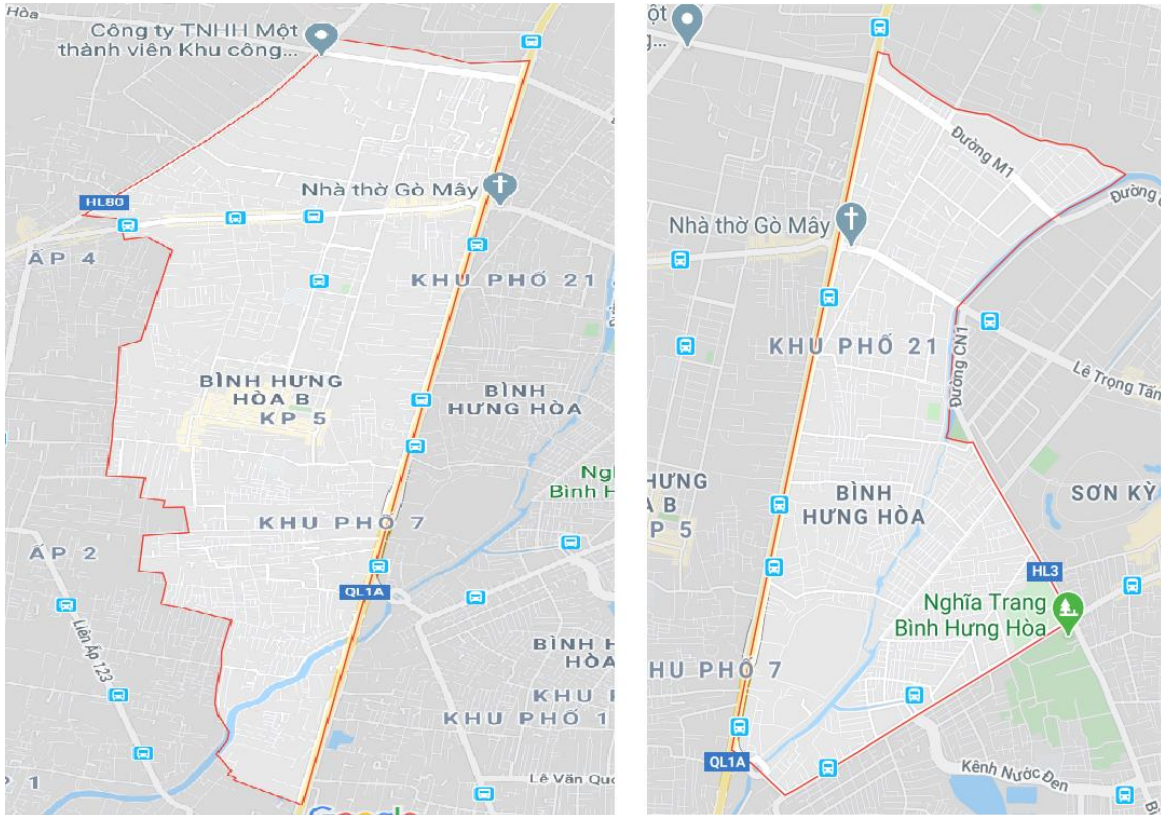
```
In [101]: # Check if wards on high population density
after_visualize_df[after_visualize_df['District Code'].isin(high_density_df['District Code'])]
```

Out[101]:

	District Code	District Name	Ward Code	Ward Name	District Population	District Acreage (Km2)	District Latitude	District Longitude	Num_Of_Coffee	Num_Of_ShortTerm	Num_Of_LongTerm	k_cluster
--	---------------	---------------	-----------	-----------	---------------------	------------------------	-------------------	--------------------	---------------	------------------	-----------------	-----------

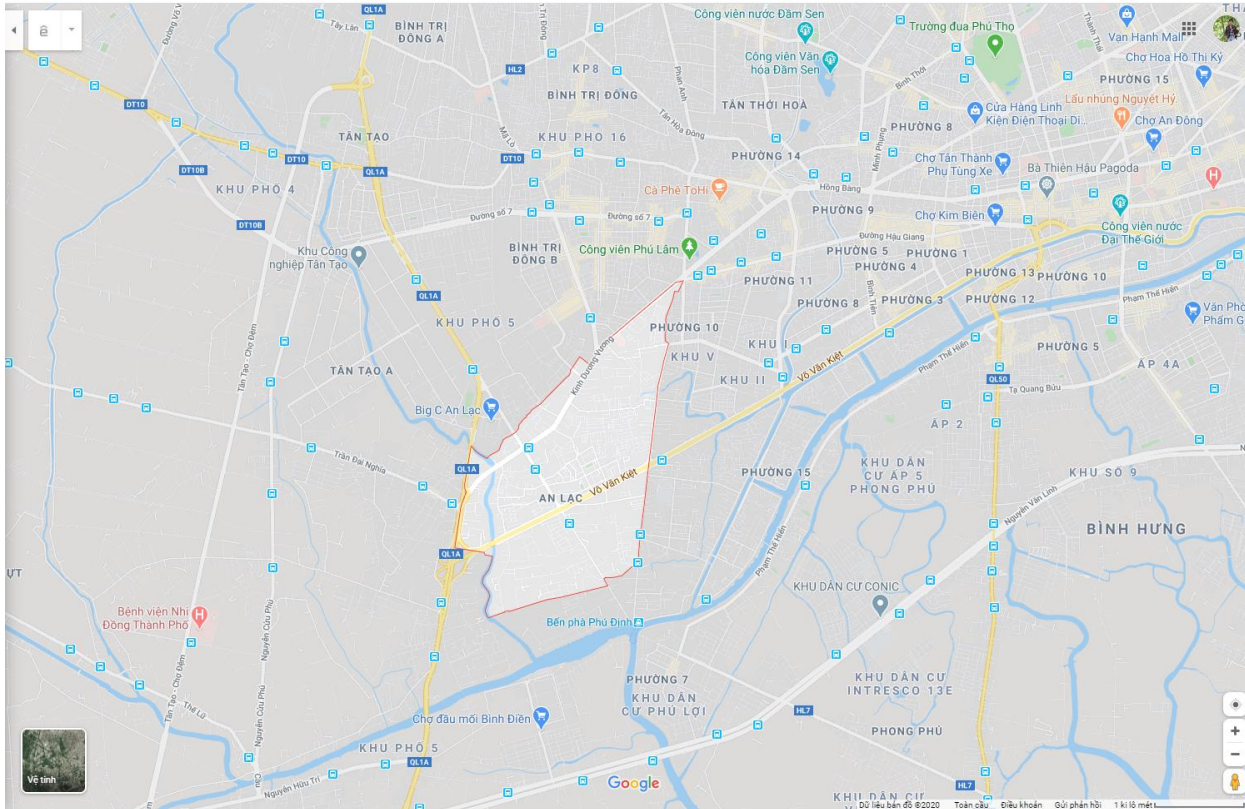
- We have three wards after visualization and filter which are in the average population density

Discussion and Results



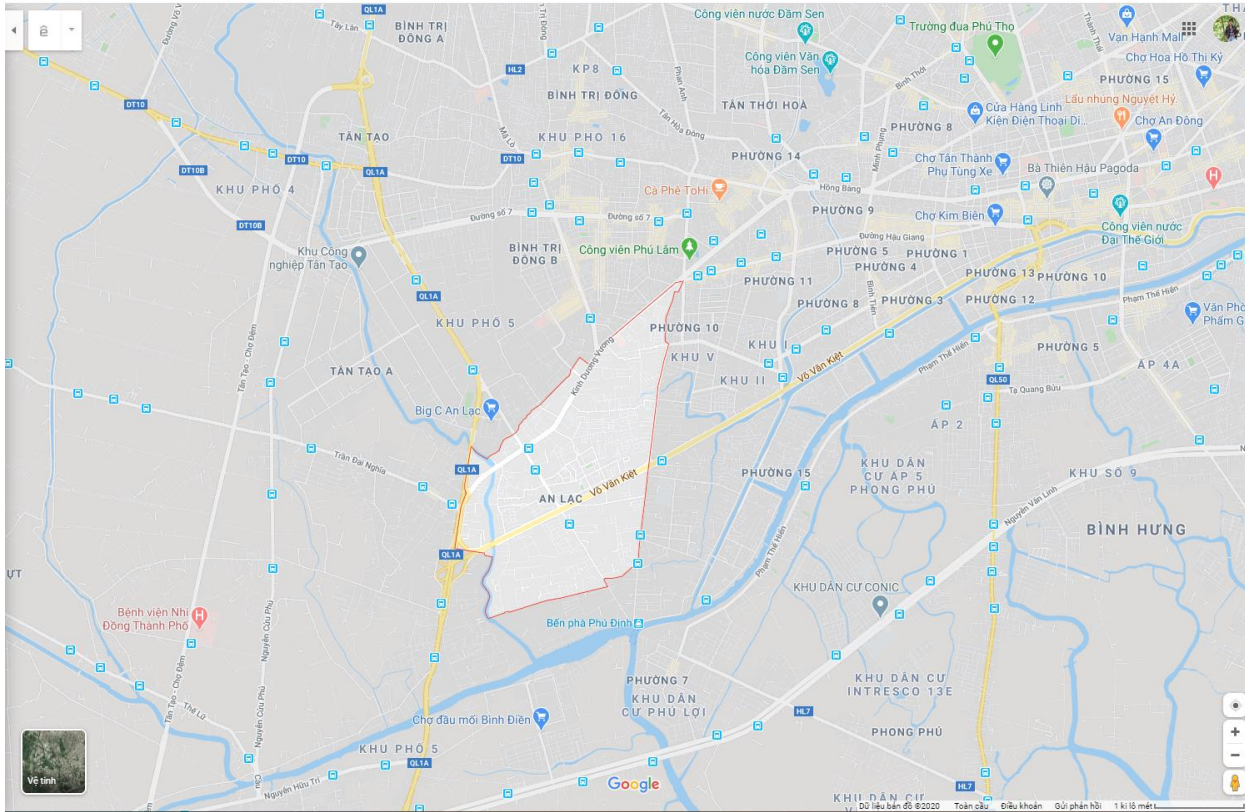
- Two wards are in the same district and same area, separated by a road.
- But they are far from Ho Chi Minh downtown.
- And we have more coffee shop venues on these 2 wards (26 and 27 coffee shop venues).
- So we will not choose these 2 wards to open a new coffee shop.

Discussion and Results



- This ward is near HCMC downtown.
- Border of District 10 which is high population and high density of administrative and career agencies
- Võ Văn Kiệt Street and Kinh Dương Vương Street are arterial roads in Ho Chi Minh City which run through most of this ward's area.
- Lot of department, one industry park, 2 park in around this ward

Final Conclusion



**We will open new coffee shop in AN LAC
WARD with ward code 27460.**