

INVESTMENT CANDIDATE IN SEOUL

Ki Hoon Kwon

1. Introduction: Business Problem

In this project, I will try to find a candidate for real property investment in Seoul, the capital of South Korea. The target of this project is investors who want to find underestimated areas in Seoul.

Seoul has **25 districts('gu')** and **467 neighborhoods('dong')**. To characterize each neighborhood, I will use the Foursquare location data. I will try to find a few neighborhoods that have similar characteristics of areas whose public land prices (KRW/m²) are expensive, but the land's value is relatively cheap. There would be reasons why the land price of the neighborhood is cheap, such as bad public transportation, no schools nearby, etc. But I assume that those kinds of weak points would be overcome in the future.

I will use data science methodology to find the best investment candidates. For example, I may be able to find some undervalued neighborhood which has very similar to the Gangnam area.

2. Data

To solve the problem, I need the following information.

- Geospatial Coordinates of each neighborhood in Seoul
- The average land price of each neighborhood in Seoul

I can obtain the information above from the website Seoul Metropolitan Government.

- Geospatial coordination of neighborhoods in Seoul

(<http://data.seoul.go.kr/dataList/OA-13221/S/1/datasetView.do>)

The CSV file contains **neighborhood's code**(읍면동코드) and **name**(읍면동명),

latitudes(위도) and **longitudes**(경도).

- The annual standard land price of Seoul (<http://data.seoul.go.kr/dataList/OA-1180/F/1/datasetView.do>)

The XLS file contains **district's**

name(SIGUNGU_NM) and **code**(SGG_CD) and **neighborhood's**

name(BJDONG_NAME) and **code**(BJDONG_CD) and **standard price**(JIGA) of each address.

- Venues of each neighborhood of Seoul obtained via Foursquare API

3. Methodology

In this project, I will characterize neighborhoods in Seoul based on the proportions of venue categories obtained via the Foursquare API. Then, I will cluster neighborhoods by the K-means algorithm. The number of clusters is a key parameter of the K-means algorithm. I will choose the appropriate size using a heuristic approach.

After clustering neighborhoods, I can find the most expensive clusters. I assume that neighborhoods in a cluster have the same characteristic. So, a low priced neighborhood is under-evaluated than other neighborhoods in that cluster.

I will select 3 most under estimated neighborhoods and examine 3 candidates on the map.

4. Result and Discussion

I downloaded a CSV file that contains a geospatial location of 467 neighborhoods in Seoul from the [Seoul Metropolitan Government website](#). The file contains neighborhood code(Neighbor_CD), neighborhood name in English and Korean(Neighbor_NM, Neighbor_NM_KOR), latitude, longitude, district code(SGG_CD), neighborhood code(BJDONG_CD). Neighbor_CD is a concatenation of SGG_CD and the first 3 digits of BJDONG_CD.

Table 1. Geospatial CSV File

Number	Neighbor_CD	Neighbor_NM_KOR	Neighbor_NM	Latitude	Longitude	SGG_CD	BJDONG_CD
76	11140108	수하동	Suha-dong	37.567417	126.984642	11140	10800
84	11140102	다동	Da-dong	37.567883	126.981194	11140	10200
85	11230107	청량리동	Cheongnyangni-dong	37.588315	127.046090	11230	10700
86	11230110	이문동	Imun-dong	37.599401	127.062471	11230	11000
87	11260106	신내동	Sinnae-dong	37.610252	127.101155	11260	10600

The land price data is also downloadable from the [Seoul Metropolitan Government website](#). The XLSX file contains the public land price of 916,873 addresses in 467 neighborhoods. Land_CD is the code for each address. It consists of 8 digits Neighbor_CD and 00 and 9 digits address code. JIGA is the land price of each address.

Table 2. Land Price of Each Address

	SIDO_NM	SIGUNGU_NM	BJDONG_NM	LAND_CD	JIGA	SGG_CD	BJDONG_CD
0	서울특별시	종로구	청운동	1111010100100010000	3961000	11110	10100
1	서울특별시	종로구	청운동	1111010100100010001	1296000	11110	10100

2	서울특별시	종로구	청운동	1111010100100010002	3267000	11110	10100
3	서울특별시	종로구	청운동	1111010100100010003	4255000	11110	10100
4	서울특별시	종로구	청운동	1111010100100010004	1524000	11110	10100

I will analyze data in neighborhood units. I calculate the average land price of each neighborhood by using the **groupby** method. And I merge neighborhood data and average land prices.

Table 3. Merged Dataframe

	Number	Neighbor_CD	Neighbor_NM_KOR	Neighbor_NM	Latitude	Longitude	JIGA
0	76	11140108	수하동	Suha-dong	37.567417	126.984642	21067642
1	84	11140102	다동	Da-dong	37.567883	126.981194	21260062
2	85	11230107	청량리동	Cheongnyangni-dong	37.588315	127.046090	3111159
3	86	11230110	이문동	Imun-dong	37.599401	127.062471	2890170
4	87	11260106	신내동	Sinnae-dong	37.610252	127.101155	1394616

Let's examine the land price of neighborhoods. There are several outliers whose land price is very expensive. Among the top 50 expensive neighborhoods, 29 neighborhoods are in Jung-gu and 19 are in Jongno-gu. These 2 districts are the downtown and center of Seoul. The land price is high mainly because of the geographical location. In this analysis, I exclude these districts (161 neighborhoods) because it is meaningless to find alternative investment candidates for these areas. Among 306 neighborhoods, the expensive areas are Gangnam-gu and Yeouido-dong areas.

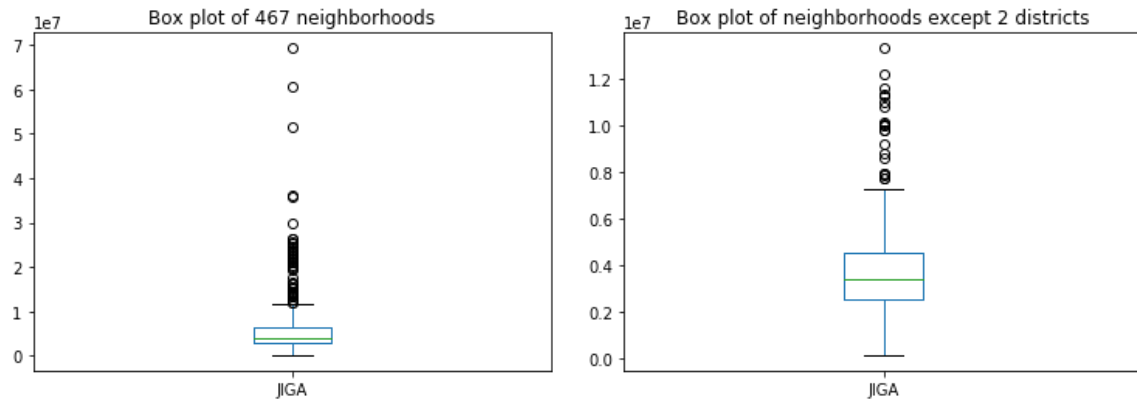


Figure 1. Box Plot of Neighborhoods

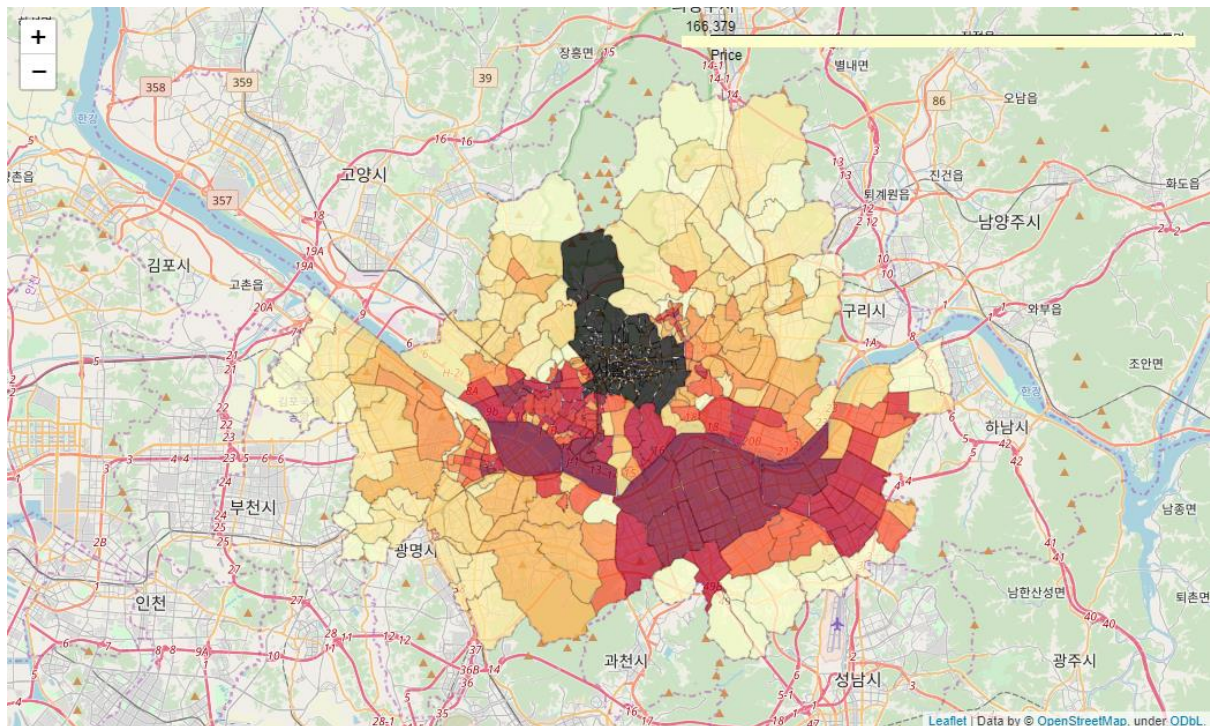


Figure 2. Land Price of Neighborhoods except for Jung-gu and Jongno-gu

I obtain the venue information of 306 neighborhoods via Foursquare API. I calculate the proportions of venue categories of neighborhoods. Then, I use the K-means algorithm to cluster neighborhoods. For deciding the number of clusters, I use the elbow test. The result shows the optimal number of clusters is 14.

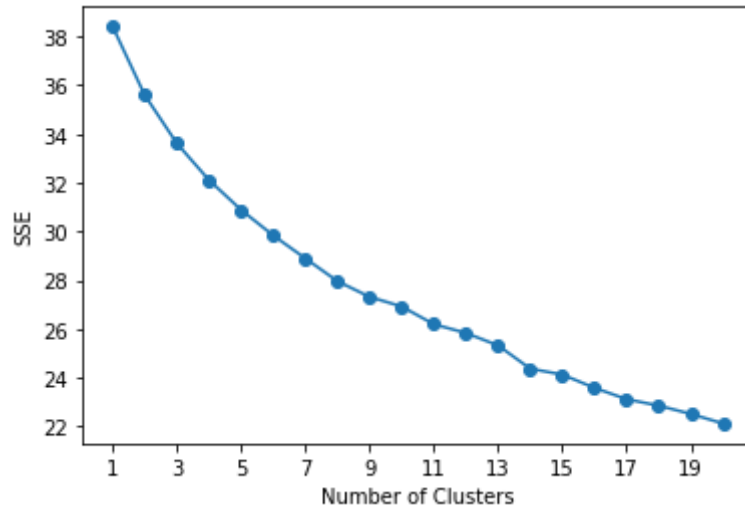


Figure 3. Elbow Test for Cluster Size

Figure 4 and Table 4 show the results of clustering. Cluster Label 6 is the most expensive cluster. Among the 20 cheapest neighborhoods, 17 clusters are in Cluster Label 6.

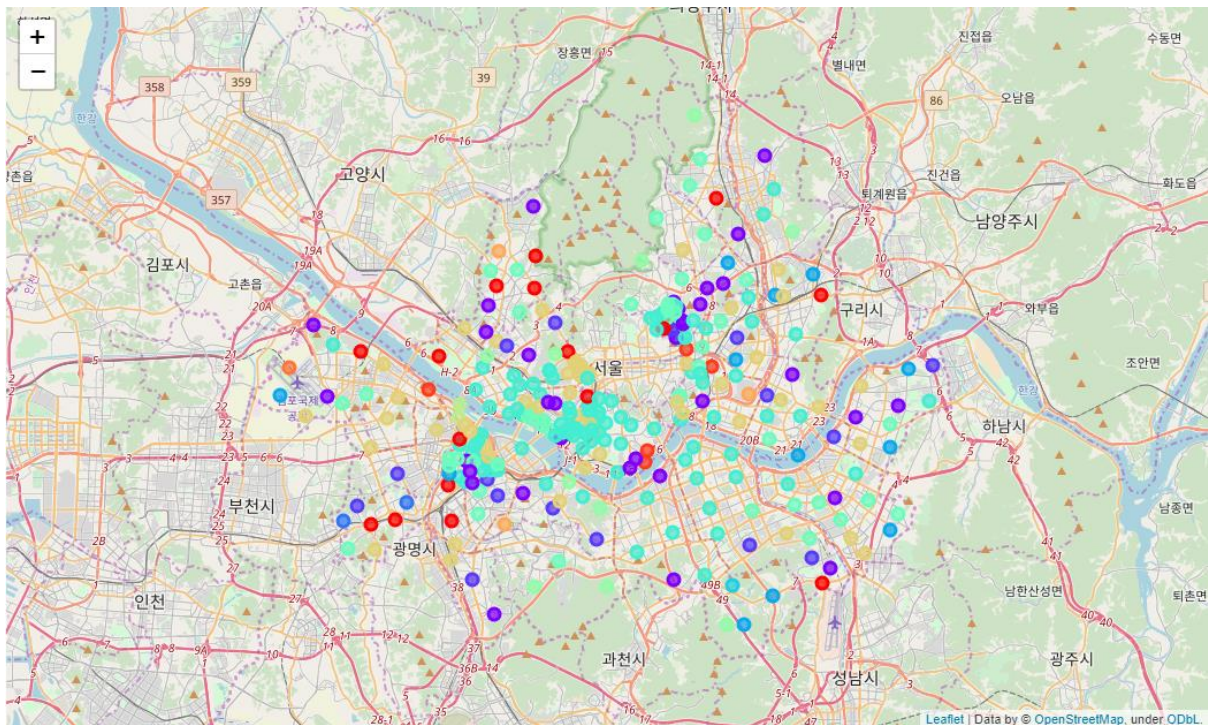


Figure 4. Result of K-means Algorithm

Table 4. Land Price of 14 Clusters

	JIGA							
Cluster Labels	count	mean	std	min	25%	50%	75%	max
0	17	2472961	901089	350000	2252250	2528762	2783310	4577571
1	36	3189269	1445453	379601	2370522	3060330	3926405	7944870
2	17	2866357	874169	679742	2550468	2948846	3183300	4480626
3	3	2882156	1410147	1756245	2091326	2426408	3445112	4463815
4	12	2477687	1788740	166379	1439857	2337414	2691486	7258678
5	1	1477851	nan	1477851	1477851	1477851	1477851	1477851
6	99	5315552	2636837	1659649	3529316	4474408	6401429	13290394
7	50	3655728	1515785	1152576	2829696	3397082	4130405	10052447
8	10	2180593	919114	731268	1635328	2060103	2695624	3854158
9	2	1739062	1724585	519597	1129330	1739062	2348795	2958528
10	49	3514322	1293392	204200	2724537	3417662	4184848	8611433
11	3	2837250	310484	2580177	2664775	2749373	2965787	3182201
12	1	168448	nan	168448	168448	168448	168448	168448
13	5	4141624	995100	3220410	3284748	3869196	4849478	5484289

Figure 5 and Table 5 show the 10 cheapest neighborhoods in Cluster Labels 6.

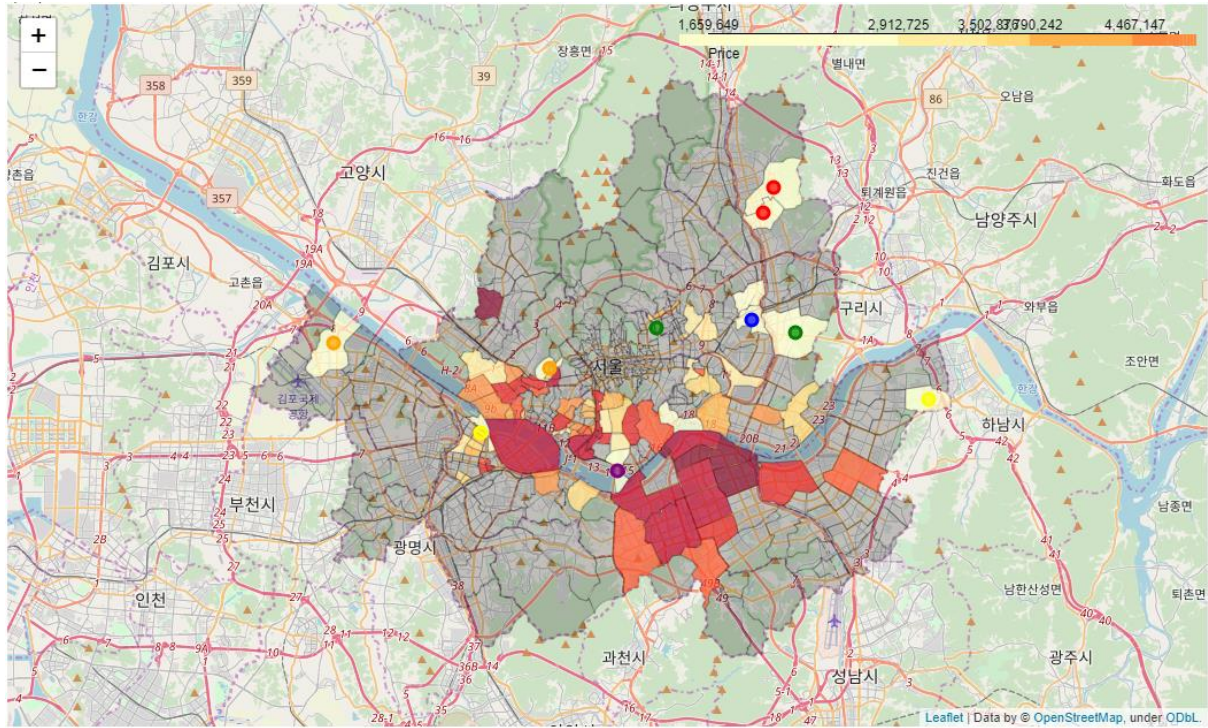


Figure 5. 10 Cheapest Neighborhood of Cluster Label 6

Table 5. 10 Cheapest Neighborhoods in Cluster Label 6

No.	Neighbor_CD	Neighbor_NM_KOR	Neighbor_NM	JIGA	Latitude	Longitude
328	11350104	하계동	Hagye-dong	1659649	37.638495	127.071507
271	11350106	중계동	Junggye-dong	1870211	37.650877	127.077508
371	11410114	신촌동	Sinchon-dong	1927879	37.565156	126.943345
289	11500109	방화동	Banghwa-dong	2196740	37.577367	126.814752
232	11560117	당산동	Dangsan-dong	2231074	37.534892	126.902468
369	11740103	상일동	Sangil-dong	2355359	37.550801	127.169869
56	11260101	면목동	Myeonmok-dong	2504948	37.581925	127.090210
439	11290111	삼선동 1 가	Samseon-dong 1(il)-ga	2577980	37.584384	127.007517
21	11230109	휘경동	Hwigyeong-dong	2618373	37.588185	127.064327
150	11170135	용산동 6 가	Yongsan-dong 6(yuk)-ga	2655125	37.516861	126.984213

Hagye-dong and Junggye-dong are the cheapest neighborhoods. They are located in the northeastern part of Seoul. The reason why these neighborhoods are cheap is the geographical

location. But these neighborhoods are a well-balanced area for living there. They have residential areas, parks, commercial areas.

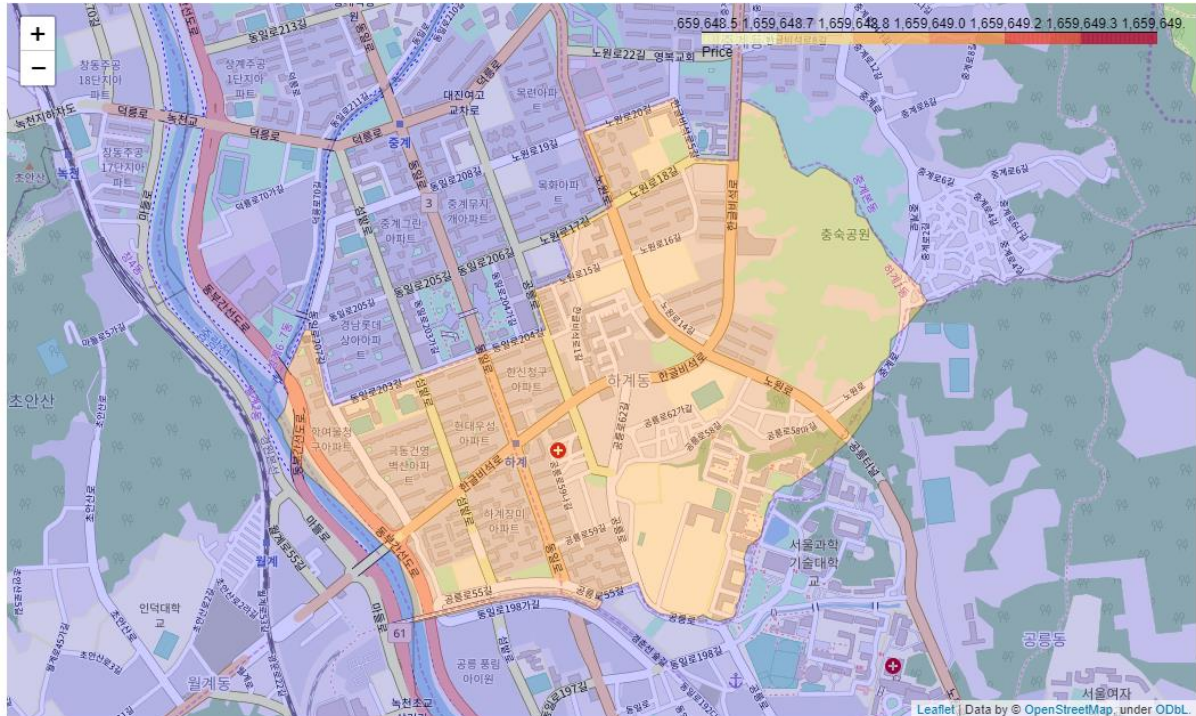


Figure 6. Map of Hagye-dong

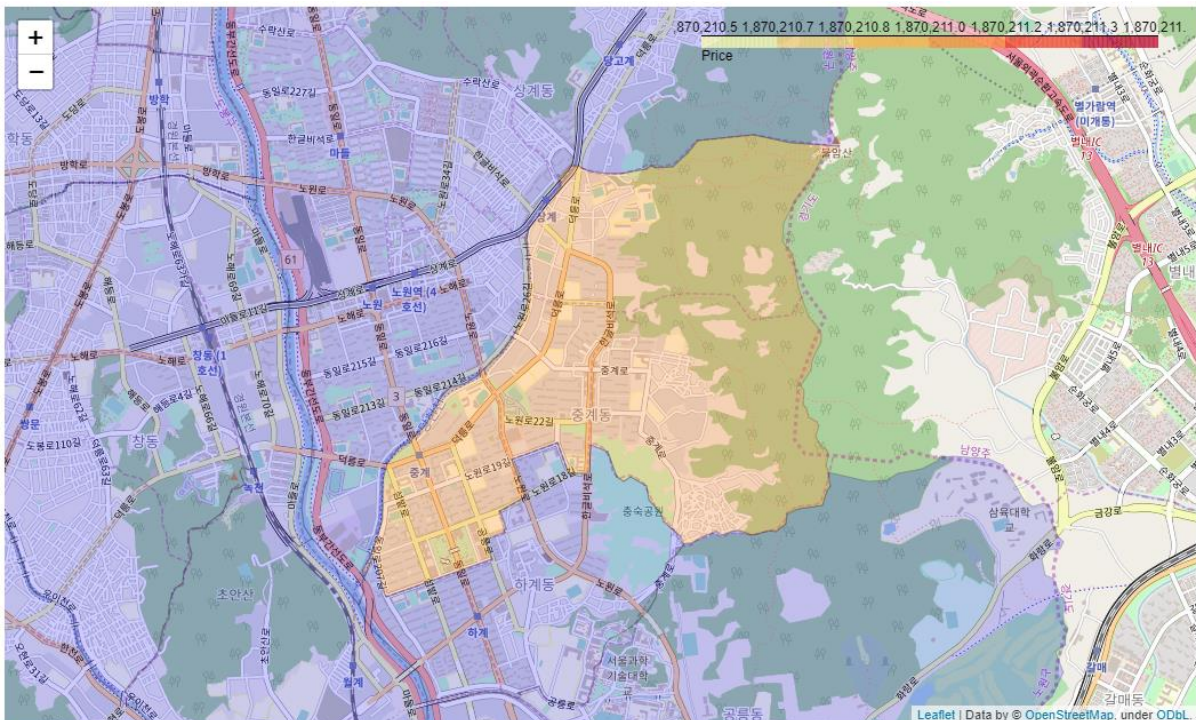


Figure 7. Map of Junggye-dong

Sinchon-dong is not too far from downtown. But Yonsei University owns this neighborhood. So it is nearly impossible to invest here.

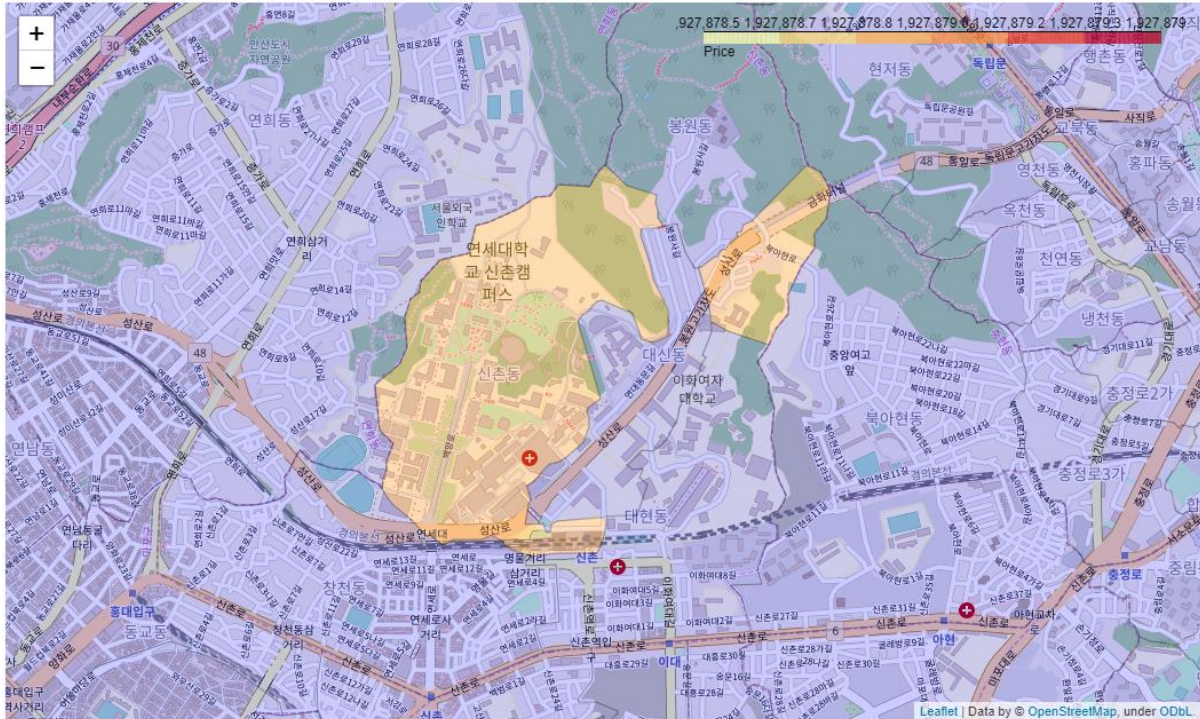


Figure 8. Map of Sinchon-dong

5. Conclusion

In this project, I collected location data to characterize neighborhoods in Seoul by using the Foursquare API. And I obtained the geospatial location and land price information of neighborhoods from the Seoul Metropolitan Government website. I used the K-means algorithm to cluster neighborhoods. I found the most expensive cluster. I examined 3 cheapest neighborhoods from that cluster. These neighborhoods are a good investment candidate in Seoul because they have similar characteristics with expensive areas, but the prices are relatively cheap now.