Apartment prices and venues data analysis in Seoul, South Korea

1. Introduction

Seoul is the capital and largest city of Korea as the center of the politics, economy, society and culture of Korea. It is a basin city with the Han River flowing in the center and surrounded by mountains. Seoul has an area of 605.2 km² and a population of about 10 million, accounting for 0.6% of the total area of Korea and 18.7% of the total population.[1] Like other large cities, housing costs in Seoul are difficult to deal with and varies greatly depending on the districts. In this report, we want to find out the correlation between the average housing price and the surrounding venue in Seoul.

2. Data

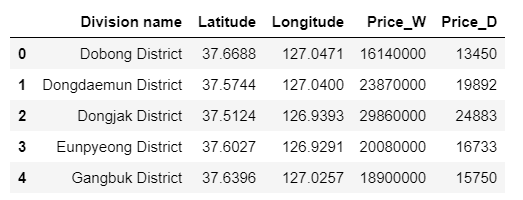
Seoul is comprised of 25 administrative districts, which is similar to Toronto's or New York's boroughs.[2] We will use the apartment price to represent the housing cost because the general type of residence in Seoul is apartments, accounting for 40-50%.[3] The price data include the average apartment price per 3.3m2 area by district for July 2020.[4] We will also use Foursquere API for venues information of given districts in Seoul.[5]

3. Methodology

**Data preprocessing**

I loaded the table including the latitude and the longitude information and average apartment price per 3.3 m2 of 25 districts in Seoul. The table was generated by joining two tables: geographic coordinates and apartment prices. (Table 1)

Table 1 Geographic coordinates and apartment prices of 25 districts in Seoul



It used the Foursquare API to create lists of venues based on the geographic coordinates of Seoul. I have set the parameters. Considering the general area of ​​districts, the limit was set to 100 venues and the radius was set to 3000. 192 unique venue categories are defined in Seoul. The venue list of each district was sorted by venue category frequency in descending order. (Table 2)

Table 2 Most common venues in each district in Seoul



**Analysis**

I performed clustering on most common venues data in each district using k-means clustering algorithm which is one of the popular unsupervised machine learning algorithms. I clustered 25 districts into 5 clusters. Each cluster has given a name representing the characteristics of most common venues. We compared most common venues and apartment prices in 5 clusters.

**Visualization**

I generated a map using a folium library in python for visualizing geographic information in Seoul (Figure 1). Based on this map, k-mean clustering results were presented. In addition, I used a bar plot for analyzing correlation between main venue categories and apartment price in each cluster.

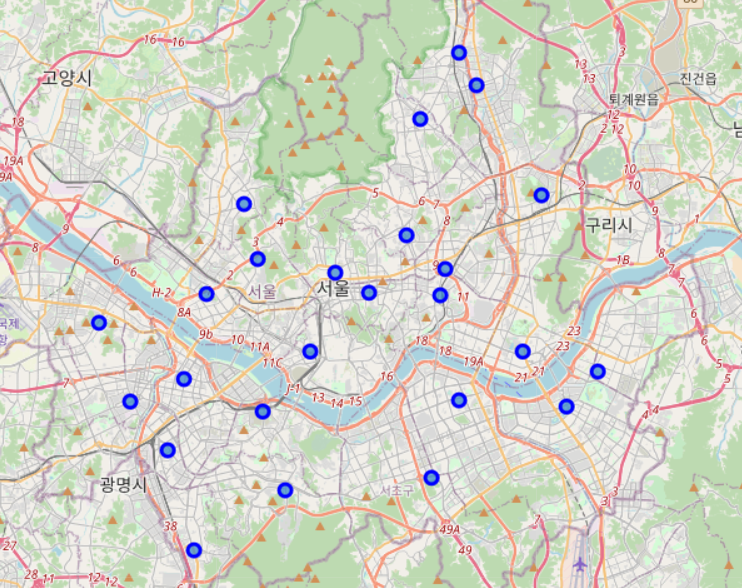


Figure 1 Seoul map with marks of 25 districts

4. Results

**k-means cluster**

Figure 2 shows 5 different color-coded regions clustered by k-means algorithm.

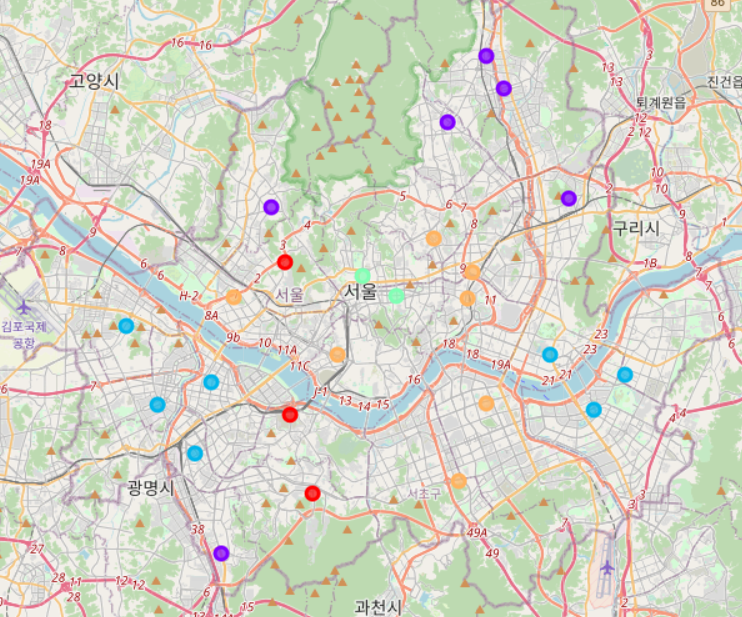


Figure 2 Seoul map clustered into 5 by common venue lists in districts

Table 3 to 7 show the most common venue list of each cluster. We can find that the venues listed in each cluster exhibits different characteristics, which is represented as labels of Tables.

Table 3. Cluster 1 with red dots : Mixed

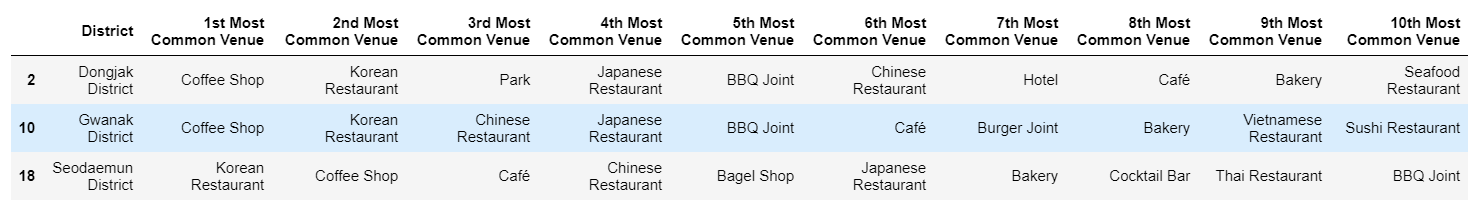


Table 4 Cluster 2 with purple dots : Fast food



Table 5 Cluster 3 with blue dots : Park



Table 6 Cluster 4 with light green dots : Historic Place

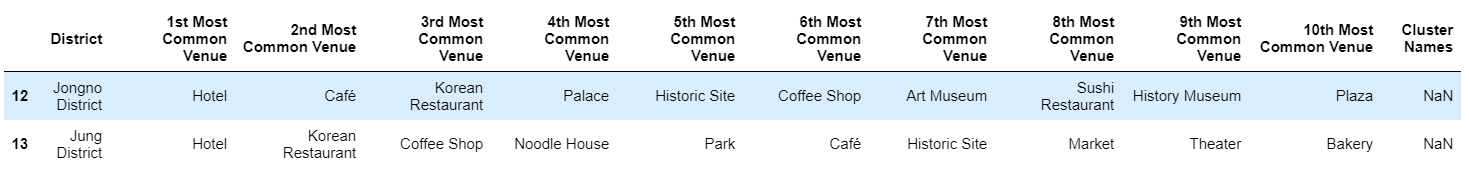


Table 7 Cluster 5 with yellow dots : BBQ



**Box plot**

Figure 3 shows correlation between clusters and apartment prices. As you can see, apartment prices between clusters has significantly different. ‘Fast food’ cluster has the lowest apartment price. On the other hand, ‘BBQ’ cluster has the highest apartment price. ‘Mixed’, ‘Park’, and ‘Historic Place’ clusters has the roughfully moderate apartment price.

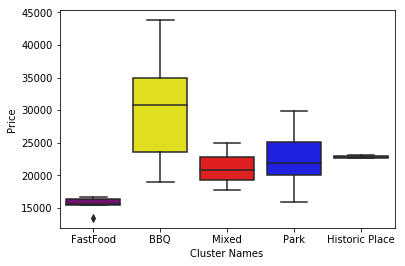


Figure 3 Box plot explaining correlation between clusters and apartment prices

5. Discussion

The map in Figure 2 represents color-coded districts by the clustering results. Districts in the same cluster are located closely each other. Therefore, we can see that different regions in Seoul have different neighborhood characteristics. The center of Seoul except ‘Historic Place’ cluster has lots of high-end restaurants like BBQ. On the other hand, outer regions in the east-west direction have relaxed places like park and café. Outer regions in the north-south direction have inexpensive shop like ice-cream shops and fast food chains. Based on the characteristics of clustered regions, the income level of residents of each cluster may be inferred.

As we assume, the box plot in Figure 3 shows the characteristics of neighborhoods can be used as an indicator of the income level of residents. In this analysis, we use the apartment price instead of the resident income. Apartments in districts in ‘Fast food’ cluster have the cheapest price. Apartments in ‘BBQ’ cluster have the most expensive price.

In addition, we can find out one more interesting point in the box plot in Figure 3. The prices of apartments in the 'Fast food' cluster both are low, and the price deviation is not large. On the other hand, you can see that as the common restaurant in the cluster handles more expensive foods, not only the price of the apartments belonging to the cluster increases, but the deviation in the price of those apartments also increases.

Therefore, we can infer that the prices of expensive apartments are determined by more diverse factors than inexpensive apartments, which means it is hard to be noted by only neighborhood information in a region. More data will be needed if you want to more accurate apartment price estimation.

6. Conclusion

In this report, we analyzed the correlation between the average housing price and the surrounding venue in Seoul. We found out that we can infer the apartment price according to what items dealt with in neighbors.

These results will be a good guideline for not only house buyers but also who want to start a new restaurant.

Reference

[1] https://ko.wikipedia.org/wiki/%EC%84%9C%EC%9A%B8%ED%8A%B9%EB%B3%84%EC%8B%9C

[2] https://en.wikipedia.org/wiki/List\_of\_postal\_codes\_in\_South\_Korea

[3] http://data.si.re.kr/sites/default/files/file/2%EC%9E%A5%20%EC%84%9C%EC%9A%B8%EC%9D%98%20%EC%A3%BC%ED%83%9D%ED%98%84%ED%99%A9.pdf

[4] http://buking.kr/rank.php?m=mg&st=md&gi=1&si=%EC%84%9C%EC%9A%B8 as is 2020/08/15

[5] https://developer.foursquare.com/