Capstone Project - The Battle of Neighborhoods (Week 2)

Introduction: Business Problem

This project is for who is planning to open a Coffee House in Seoul , Korea. This project suggests the best locations for Coffee Houses in Seoul. Seoul is the capital of Korea with a population of 10M.

Korea's coffee culture has developed rapidly over the past 20 years. The number of coffee shops has increased dramatically and is gaining huge popularity. Annual coffee consumption is also steadily increasing. According to a survey by the Hyundai Economic Research Institute, the number of coffee an adult drinks over a year continued to rise to 291 in 2015, 317 in 2016, 336 in 2017, and 353 in 2018.

This report explores which neighborhoods of Seoul have the most as well as the best Coffee Houses. Also, this project answers the questions “Where should I open an Coffee House?” and “Where should I stay If I want a tasty coffee?”

Data

- District of Seoul are obtained from https://en.wikipedia.org/wiki/List_of_districts_of_Seoul
- Latitude and Longitude values are obtained by using "geocoder".
- All data related to locations will be obtaine by using FourSquare API and Python Libraries.

	Name	Population	Area	Population density
0	[illegible]	355,712	20.70km²	17184/km²
1	Dongdaemun-gu (동대문구; 東大門區)	376,319	14.21km²	26483/km²
2	Dongjak-gu (동작구; 銅雀區)	419,261	16.35km²	25643/km²
3	Eunpyeong-gu (은평구; 恩平區)	503,243	29.70km²	16944/km²
4	Gangbuk-gu (강북구; 江北區)	338,410	23.60km²	14339/km²

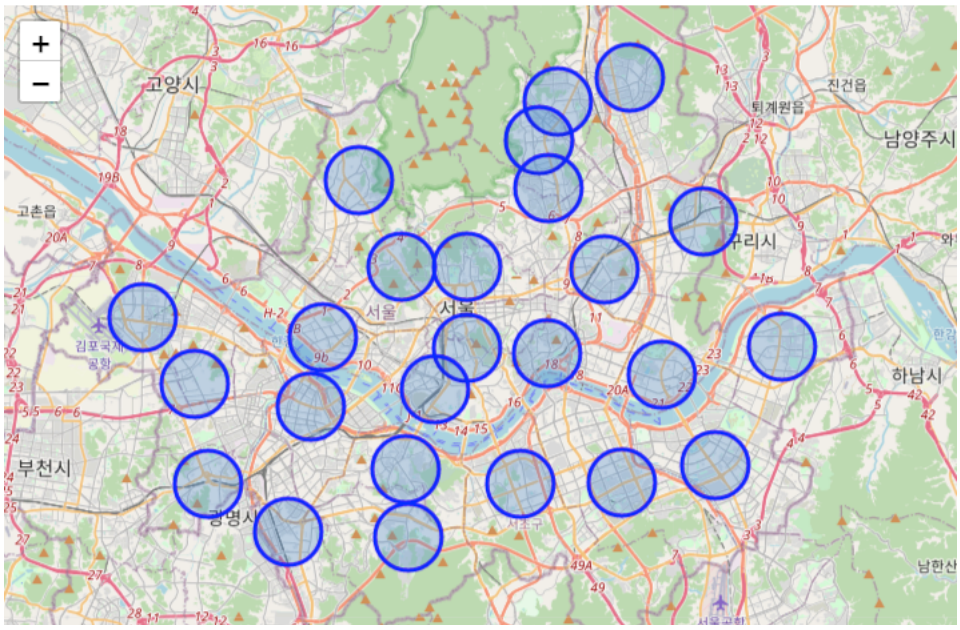
Methodology

Master data which includes “District”, “Latitude” and “Longitude” information of Seoul.

Python Folium library was used to visualize map of Seoul and its districts. In order to get Latitude and Longitudes of Seoul, geopy and geocoder are used. The map below was obtained.

In order to explore and categorize places Foursquare API was used. Limit was set to 100 and Radius was set to 500. Here is a head of the list Venues name, category, latitude and longitude information from Forsquare API. 25 venues are returned.

I will also use one-hot encoding to replace categorical data with numbers and perform clustering.



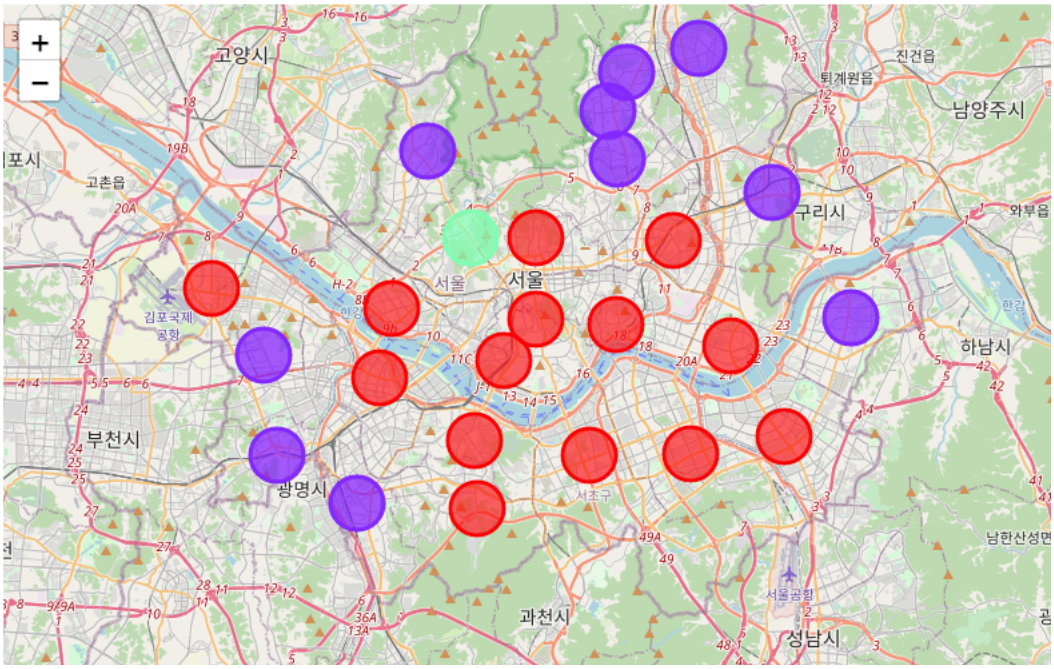
	District	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Dobong-gu	37.65066	127.03011	Dooly Museum (둘리뮤지엄)	37.652075	127.027749	Museum
1	Dobong-gu	37.65066	127.03011	Starbucks (스타벅스)	37.648425	127.034681	Coffee Shop
2	Dobong-gu	37.65066	127.03011	스시혼	37.646351	127.034020	Sushi Restaurant
3	Dobong-gu	37.65066	127.03011	MEGABOX Changdong (메가박스 창동)	37.654821	127.038507	Multiplex
4	Dobong-gu	37.65066	127.03011	수정궁	37.662404	127.032934	Chinese Restaurant
...	...	...	...	...	...	...	...
245	Gangbuk-gu	37.63490	127.02015	PARIS BAGUETTE	37.637766	127.025640	Bakery
246	Gangbuk-gu	37.63490	127.02015	PENELOPE	37.637768	127.023048	Brewery
247	Gangbuk-gu	37.63490	127.02015	어니언	37.623902	127.028078	Coffee Shop
248	Gangbuk-gu	37.63490	127.02015	Burger King (버거킹)	37.625123	127.026403	Fast Food Restaurant
249	Gangbuk-gu	37.63490	127.02015	Baskin-Robbins (배스킨라빈스)	37.626042	127.026330	Ice Cream Shop

Analysis

After processing with one-hot encoding, the data were analyzed using K-means clustering.

Results and Discussion

We set the k value as 3, and the result is shown in the picture below.



Conclusion

Kmeans algorithm was used as part of this clustering study. K value was set to 3. For more detailed and accurate guidance, the data set can be expanded, and the details of the neighbourhood or street can also be drilled.

The areas of Cluster 0 and Cluster 1 seem to be suitable places to open new cafes.

The area with Cluster 0 is close to the center of Seoul, while the area with Cluster 1 is judged to be the periphery of Seoul. In addition, data such as real estate or floating population could be obtained to determine which areas are Cluster 0 and Cluster 1. In general, Cluster 0, which is the center of Seoul, has a high store rent and a large floating population, while Cluster 1 has a low store rent and a small floating population.