Analysis of Seoul Districts for Foodies

K L Goh

27 June 2020

A. Introduction

Seoul is the capital city of South Korea and it has a population of 9.7 million people. The city is divided into 25 districts.[1] This project will be targeting people who are interested in exploring Seoul, South Korea. And more specifically, this project will target foodies who like to explore a different eatery every week.

B. Data

The data are as follow:

- I have created a list containing the names of districts This list will become the initial data frame for this project. The list of districts of Seoul can be found on Wikipedia [2]
- I used geopy package to generate the geographical coordinates of the districts, and added these coordinates to my initial data
- I used FourSquare API to get the most common venues of given districts of Seoul [3]

My initial data has 4 columns: Korean, District, Latitude and Longitude. The reason for having the Korean names of the districts in the list is because the geopy package can better the identify the locations based on their names in the native language of the country (in this case Korean).

	Korean	District	Latitude	Longitude
0	종로구	Jongno	37.580310	126.983079
1	중구	Jung	37.563656	126.997510
2	용산구	Yongsan	37.532300	126.990000
3	성동구	Seongdong	37.563500	127.036500
4	광진구	Gwangjin	37.538400	127.082800

Fig 1. First 5 rows of the initial data

I used python folium library to visualize geographic details of Seoul and its districts and I created a map of Seoul with its districts superimposed on top. I used the geographical coordinates to get the visual shown in fig 2.

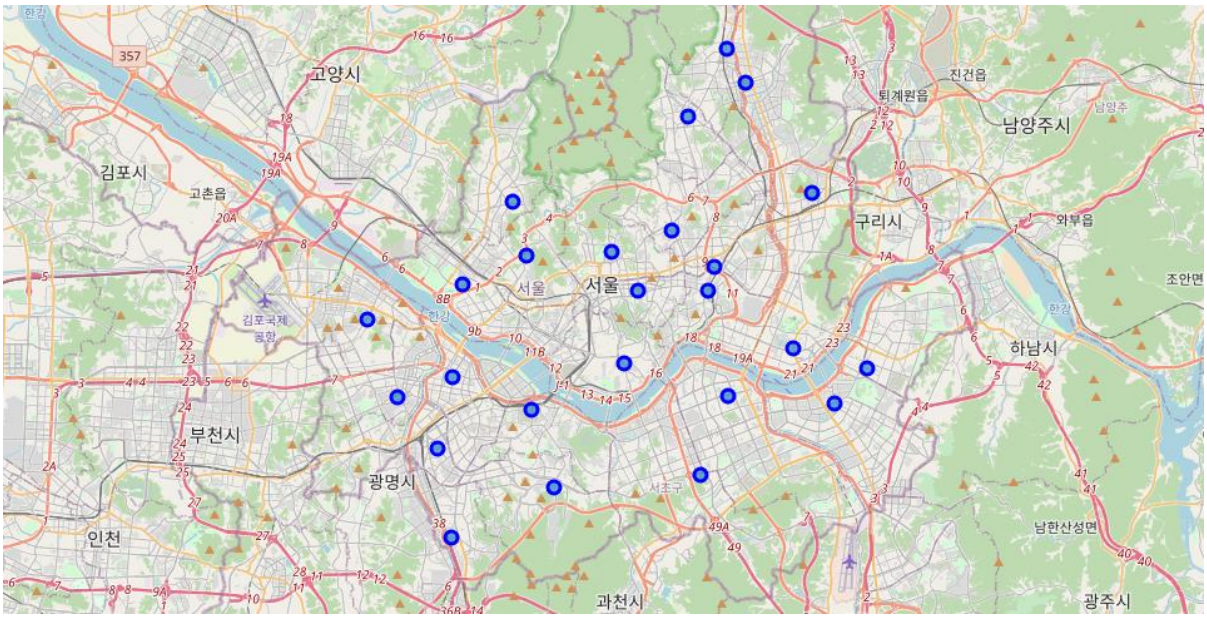

Fig 2. Map of Seoul with the 25 districts marked

I utilized the Foursquare API to explore the districts and find the top venues. I limited the number of venues to 200 and set the radius as 1000 meters for each district from their given geographical coordinates. FourSquare API has managed to detect 722 venues, with 133 unique categories. The first 5 rows of the list of venues is given in fig 3.

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Jongno	37.58031	126.983079	Baek In-Je House Museum (백인제가옥)	37.580508	126.984164	Historic Site
1	Jongno	37.58031	126.983079	KIWA TAPROOM (기와탭룸)	37.578711	126.981770	Brewery
2	Jongno	37.58031	126.983079	Blue Bottle Coffee (블루보틀)	37.580143	126.980845	Coffee Shop
3	Jongno	37.58031	126.983079	Wood & Brick (우드앤브릭)	37.579413	126.984166	Bakery
4	Jongno	37.58031	126.983079	MIRROR ROOM (미러룸)	37.579933	126.981078	Coffee Shop

Fig 3. First 5 rows of the list of venues

C. Methodology

After wrangling with the data, the next thing is to analyse the top venues and cluster the districts based on the top venues. The clustering method used is k-means clustering. The clusters will then be represented on the map. I have created a table which shows list of top 5 venues for each district. The table is shown in fig 4.

	District	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Dobong	Café	Fast Food Restaurant	Big Box Store	Ice Cream Shop	Burger Joint
1	Dongdaemun	BBQ Joint	Korean Restaurant	Supermarket	Donut Shop	Ice Cream Shop
2	Dongjak	Seafood Restaurant	Korean Restaurant	Coffee Shop	Fast Food Restaurant	Ice Cream Shop
3	Eunpyeong	Bakery	Clothing Store	Ice Cream Shop	Coffee Shop	Concert Hall
4	Gangbuk	Coffee Shop	Donut Shop	Multiplex	Fast Food Restaurant	Bus Stop

Fig 4. First 5 rows of top venues

I have decided to use the k-means algorithm to cluster the boroughs. This is because k-means algorithm is one of the most common cluster method of unsupervised learning. In order to do a good clustering, I have to first determine the optimal k value by plotting the “elbow” graph, which is shown in fig 5.

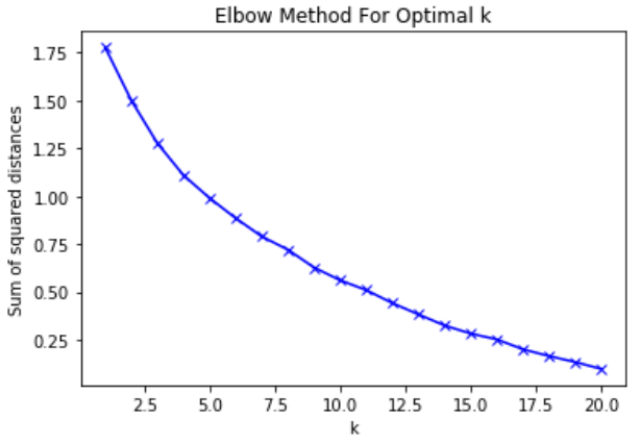

Fig 5. The “elbow” plot

From the plot above, the "elbow" is not very pronounced. However, I observed that the sharpest turn is at k = 4. To confirm that k = 4 is the optimal k value, I had to use the silhouette score. I excluded k = 2 in the silhouette score analysis because I wanted a wider segmentation of the districts. The silhouette score plot is shown in fig 6.

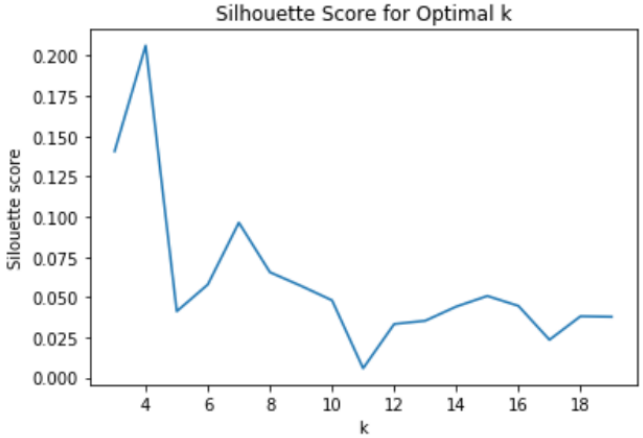

Fig 6. The silhouette score plot

The next step is to run k-means clustering with k = 4 and to visualise the clusters on the map.

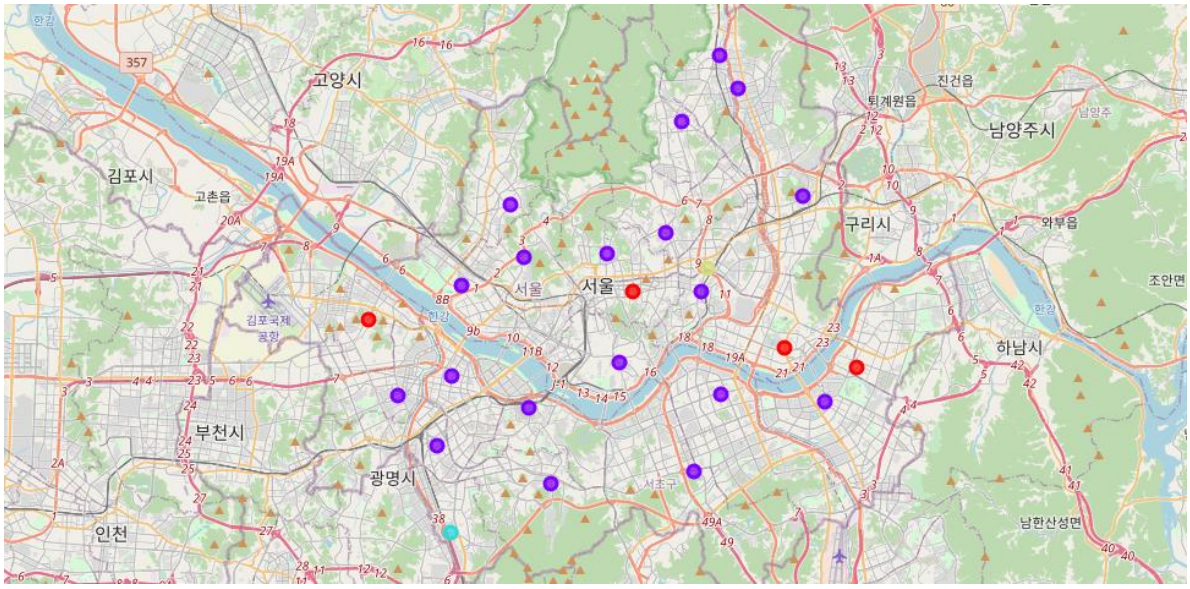

Fig 7. Map with the districts clustered

D. Results and Discussion

The k-means clustering method has resulted in one large cluster, one smaller cluster and two standalone districts.

The majority of Seoul's districts have coffee shops and cafes as the most common venues (the large cluster). The next most common venue is Korean restaurant (the smaller cluster).

In the first standalone district (Geumcheon), the first most common venue is a Chinese restaurant. In the other standalone district (Dongdaemun), the first most common venue is a BBQ Joint.

Across the 25 districts, at least 4 out of 5 of the top venues are eateries, this means approximately 578 out of the 722 venues detected by FourSquare API are eateries. As a foodie, one would be spoilt for choice on which eatery to explore.

If one wants to visit a coffee shop or cafe, I would recommend Gwangjin, Seongdong, Gangbuk, Dobong, Nowon and Seocho districts. These are the districts where the number one most common venue is either a coffee shop or cafe.

If one wants to visit a Korean restaurant, I would recommend Jung, Gangseo, Gangdong, Jongno, Seongbok, Yangcheon, Yeongdeung and Songpa. These are the districts where the number one most common venue is a Korean restaurant.

E. Conclusion

The purpose of this project is to identify the most common venues within each of the 25 districts of Seoul. Clustering of those venues is performed in order to create clusters in which the districts share common top venues. The method of clustering is k-means clustering.

Analysis shows that the majority of the most common venues detected by FourSquare API in this project are eateries. This is a good news for foodies who like to explore a different eatery every week in Seoul. The top three most common venues are coffee shops, cafes and Korean restaurants.

F. References:

[1] https://en.wikipedia.org/wiki/Seoul

[2] https://en.wikipedia.org/wiki/List_of_districts_of_Seoul

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