1. A description of the problem and a discussion of the background

Problem statement: Best location to open a coffee shop in Seoul, South Korea.

Coffee shop is not just a grab-and-go shop in South Korea. People in South Korea tend to spend lots of time in cafe for their work, study and meeting with friends. Unlike United States where most of cafe close before 7 PM, coffee shops in South Korea open until midnight. Even, there are more number of Starbucks in Seoul than that of in New York. Therefore, coffee shops are a hot store to open in Seoul. But you do not want to compete with 10 other stores in just one street.

Target Audience:

What type of clients or a group of people would be interested in this project?

   -Business personnel who wants to invest or open a coffee shop.

   -Travelors in Seoul who wants to visit hot Cafes.

   -Freelancer who would like to have their own cafes as a side business.

2. A description of the data and how it will be used to solve the problem.

Data preparation:

2.1 Scaping districts of Seoul from wikipedia. I got a list of districts of Seoul from wikipedia ("https://en.wikipedia.org/wiki/List_of_districts_of_Seoul"). This included the name of districts, population, and area of each district. I scrapped this table using Pandas.

2.2 Getting coordinates of each district using Geopy Client. Here I am choosing top 5 districts on high population density. There should be enough consumers in the neighborhood.

2.3 Using Foursquare API to explore, segment and using KMeans to cluster the neighborhoods in Seoul based on the top 5 venues for each neighborhood district.

2.4 Analyze the clustering result and then propose some suggestion location (district) to open coffee shop in Seoul.

Data acquisition -
   1.   Getting a list of district from Wikipedia
        ("https://en.wikipedia.org/wiki/List_of_districts_of_Seoul").
There are two tables in this page. I am importing and combining both tables:

```
df_seoul=df_borough.merge(df_neighborhood, on='Borough')
df_seoul.head()
```

	Borough	Population	Area	Population density	Neighborhood
0	Dobong-gu	355712	20.70 km²	17184/km²	Dobong-dong, Banghak-dong, Ssangmun-dong, Chang-dong
1	Dongdaemun-gu	376319	14.21 km²	26483/km²	Cheongnyangni-dong, Dapsimni-dong, Hoegi-dong, Hwigyeong-dong, Imun-dong, Jegi-dong, Jangan-dong, Jeonnong-dong, Sinseol-dong, Yongdu-dong
2	Dongjak-gu	419261	16.35 km²	25643/km²	Bon-dong, Daebang-dong, Dongjak-dong, Heukseok-dong, Noryangjin-dong, Sadang-dong, Sangdo-dong, Sindaebang-dong
3	Eunpyeong-gu	503243	29.70 km²	16944/km²	Bulgwang-dong, Daejo-dong, Eungam-dong, Galhyeon-dong, Gusan-dong, Jeungsan-dong, Jingwan-dong, Nokbeon-dong, Sinsa-dong, Susaek-dong, Yeokchon-dong
4	Gangbuk-gu	338410	23.60 km²	14339/km²	Mia-dong, Beon-dong, Suyu-dong, Ui-dong

2.2 Getting coordinates of each district using Geopy Client.

Here I am choosing top 5 borough with highest population density in Seoul.

```
df_seoulsort=df_seoul.sort_values(by='Population density').reset_index(drop=True)
df_seoul5=df_seoulsort[:5]
df_seoul5
```

	Borough	Population	Area	Population density	Neighborhood	latitude	longitude
0	Yongsan-gu	249914	21.87 km²	11427/km²	Cheongpa-dong, Hangangno-dong, Hannam-dong, Huam-dong, Hyochang-dong, Ichon-dong, Itaewon-, Namyoung-dong, Seobinggo-dong, Wonhyoro-dong, Yongmun-dong, Yongsan-dong	37.532300	126.99000
1	Jung-gu	136227	9.96 km²	13677/km²	Not, subdivided, into, dongs[citation, needed]	37.563656	126.99751
2	Gangseo-gu	591653	41.43 km²	14281/km²	Banghwa-dong, Deungchon-dong, Gaehwa-dong, Gayang-dong, Gonghang-dong, Gwahae-dong, Hwagok-dong, Magok-dong, Naebalsan-dong, Oebalsan-dong, Ogok-dong, Osoe-dong, Yeomchang-dong	37.550900	126.84970
3	Gangbuk-gu	338410	23.60 km²	14339/km²	Mia-dong, Beon-dong, Suyu-dong, Ui-dong	37.639500	127.02550
4	Gangnam-gu	583446	39.50 km²	14771/km²	Apgujeong-dong, Cheongdam-dong, Daechi-dong, Dogok-dong, Gaepo-dong, Irwon-dong, Jagok-dong, Nonhyeon-dong, Samseong-dong, Segok-dong, Sinsa-dong, Suseo-dong, Yeoksam-dong, Yulhyeon-dong	37.517700	127.04730

2.3 Using Foursquare API

Here, I am gathering all venues in 1 km on each borough up to 100 venues.

```
print(seoul_venues.shape)
seoul_venues.tail(10)
```

(354, 7)

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
344	Gangnam-gu	37.5177	127.0473	라우라우	37.522060	127.041281	Australian Restaurant
345	Gangnam-gu	37.5177	127.0473	Starbucks Reserve (스타벅스 리저브)	37.525361	127.041525	Coffee Shop
346	Gangnam-gu	37.5177	127.0473	SOIE et ÉPICE (수아에피스)	37.522616	127.040215	Thai Restaurant
347	Gangnam-gu	37.5177	127.0473	아야진 생태찌개	37.511940	127.054732	Korean Restaurant
348	Gangnam-gu	37.5177	127.0473	켄마	37.512172	127.055698	Japanese Restaurant
349	Gangnam-gu	37.5177	127.0473	Kimchee Guesthouse - Gangnam (김치게스트하우스)	37.517804	127.039010	Hostel
350	Gangnam-gu	37.5177	127.0473	외고집설렁탕	37.511137	127.054347	Korean Restaurant
351	Gangnam-gu	37.5177	127.0473	오발탄	37.513607	127.053406	Korean Restaurant
352	Gangnam-gu	37.5177	127.0473	Anderson C (앤더슨씨)	37.523831	127.040189	Café
353	Gangnam-gu	37.5177	127.0473	고향집	37.516796	127.037209	Korean Restaurant

In Venue category, both Café and Coffee Shop are used. To remove any complications, I am going to change 'Café' to coffee shop.

```
seoul_venues.replace("Café","Coffee Shop",inplace=True)
seoul_venues.tail(10)
```

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
344	Gangnam-gu	37.5177	127.0473	라우라우	37.522060	127.041281	Australian Restaurant
345	Gangnam-gu	37.5177	127.0473	Starbucks Reserve (스타벅스 리저브)	37.525361	127.041525	Coffee Shop
346	Gangnam-gu	37.5177	127.0473	SOIE et ÉPICE (수아에피스)	37.522616	127.040215	Thai Restaurant
347	Gangnam-gu	37.5177	127.0473	아야진 생태찌개	37.511940	127.054732	Korean Restaurant
348	Gangnam-gu	37.5177	127.0473	켄마	37.512172	127.055698	Japanese Restaurant
349	Gangnam-gu	37.5177	127.0473	Kimchee Guesthouse - Gangnam (김치게스트하우스)	37.517804	127.039010	Hostel
350	Gangnam-gu	37.5177	127.0473	외고집설렁탕	37.511137	127.054347	Korean Restaurant
351	Gangnam-gu	37.5177	127.0473	오발탄	37.513607	127.053406	Korean Restaurant
352	Gangnam-gu	37.5177	127.0473	Anderson C (앤더슨씨)	37.523831	127.040189	Coffee Shop
353	Gangnam-gu	37.5177	127.0473	고향집	37.516796	127.037209	Korean Restaurant

Let's count how many venues I got from each borough.

```
seoul_venues.groupby('Borough').count()
```

Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Gangbuk-gu	25	25	25	25	25	25
Gangnam-gu	100	100	100	100	100	100
Gangseo-gu	29	29	29	29	29	29
Jung-gu	100	100	100	100	100	100
Yongsan-gu	100	100	100	100	100	100

Analyze each borough:
I would like to see top 5 venues on each borough:

```
----Gangbuk-gu----
                   venue  freq
0            Coffee Shop   0.20
1      Korean Restaurant   0.12
2             Donut Shop   0.12
3               Bus Stop   0.08
4   Fast Food Restaurant   0.08


----Gangnam-gu----
                    venue  freq
0             Coffee Shop   0.13
1       Korean Restaurant   0.10
2               BBQ Joint   0.07
3     Japanese Restaurant   0.07
4            Dessert Shop   0.06


----Gangseo-gu----
                    venue  freq
0       Korean Restaurant   0.21
1             Coffee Shop   0.17
2                  Bakery   0.07
3       Italian Restaurant  0.03
4             Noodle House  0.03


----Jung-gu----
                    venue  freq
0       Korean Restaurant   0.17
1             Coffee Shop   0.16
2                   Hotel   0.10
3             Noodle House  0.08
4                  Market   0.03


----Yongsan-gu----
                    venue  freq
0             Coffee Shop   0.14
1                     Bar   0.05
2       Korean Restaurant   0.05
3               BBQ Joint   0.04
4     Dumpling Restaurant   0.04
```

Also, putting these information to dataframe:

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	
0	Gangbuk-gu	Korean Restaurant	Coffee Shop	Donut Shop	Fast Food Restaurant	Market	G
1	Gangnam-gu	Coffee Shop	Korean Restaurant	BBQ Joint	Japanese Restaurant	Dessert Shop	
2	Gangseo-gu	Korean Restaurant	Coffee Shop	Bakery	Clothing Store	Spa	
3	Jung-gu	Korean Restaurant	Coffee Shop	Hotel	Noodle House	Pub	
4	Yongsan-gu	Coffee Shop	Bar	Korean Restaurant	Cocktail Bar	BBQ Joint	

Here, I can conclude that the most common venue on Gangbuk-gu, Gangnam-gu and Yongsan-gu is 'Coffee Shop'.

**Cluster using K-means.**

Visualization in map:
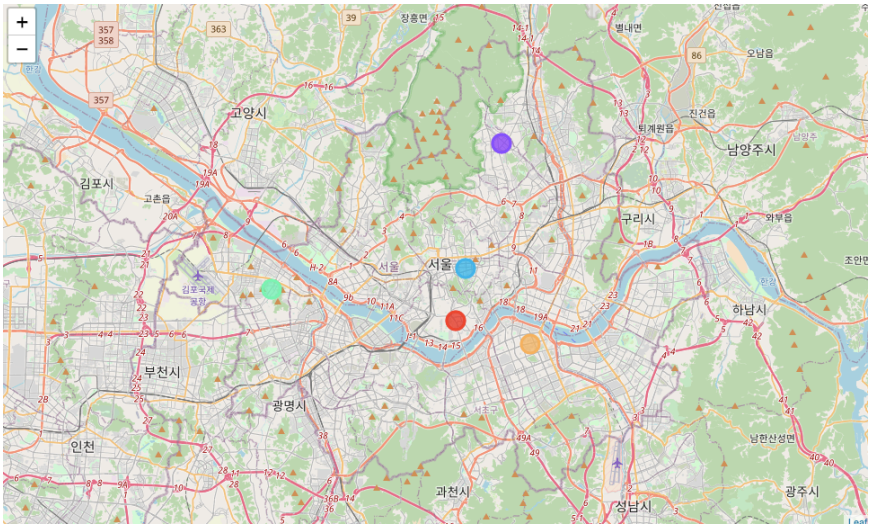
Now, I am curious how many coffee shops in each of these boroughs.

Here, getting dataframe only with 'Coffee shop'.

```
seoul_coffee_count=seoul_coffee.groupby('Borough').count().reset_index()
seoul_coffee_count
```

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Gangbuk-gu	3	3	3	3	3	3
1	Gangnam-gu	13	13	13	13	13	13
2	Gangseo-gu	6	6	6	6	6	6
3	Jung-gu	16	16	16	16	16	16
4	Yongsan-gu	14	14	14	14	14	14

Number of coffee shops in top 5 borough in Seoul
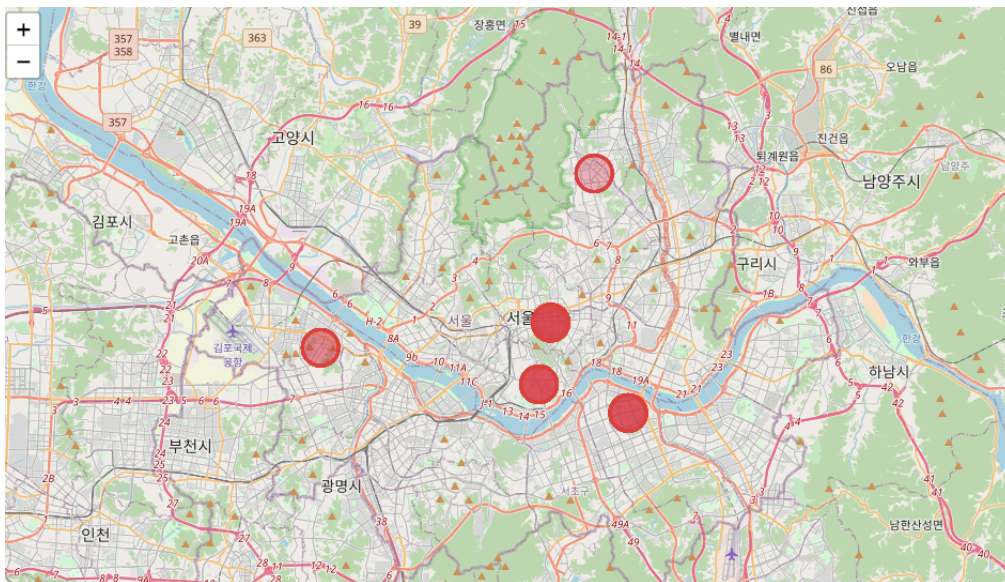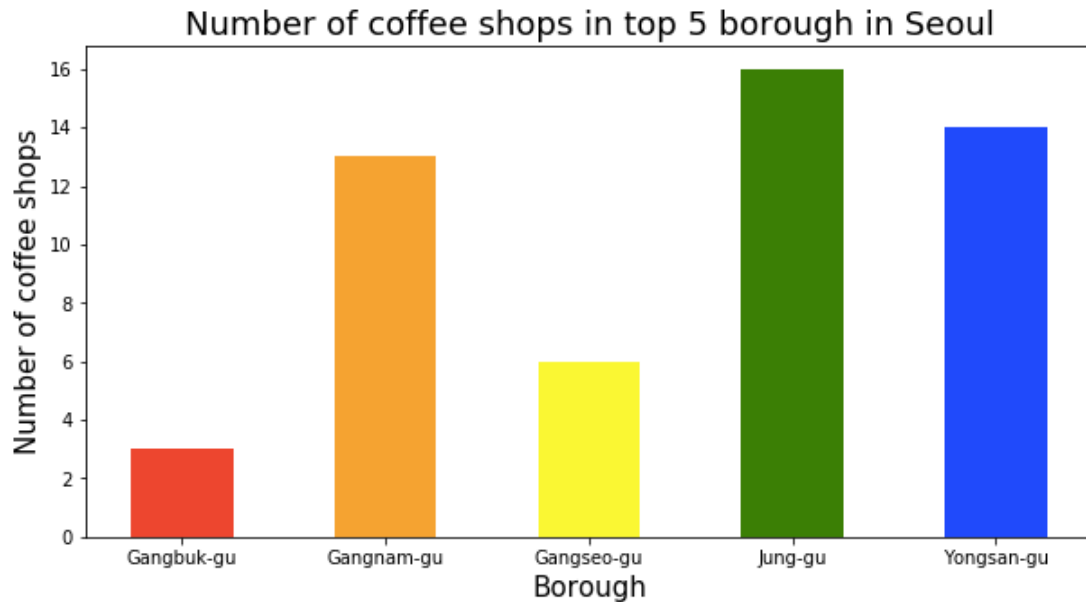
Discussion:

Interestingly, the top 3 absolute number of coffee shops resulted in Gangnam-gu, Jung-gu and Yongsan-gu. We can conclude that even though Gangbuk-gu had the most frequent cafes, the absolute number is still smaller than Gangnam-gu and Yongsan-gu. But, Jung-gu had coffee shop as 2nd most common venue, but its absolute number is pretty high. From these two analyses, my recommendation to open coffee shop is either at Gangbuk-gu or Gangseo-gu. Both boroughs have high population density in Seoul, which will attract many customers, but at the same time, competition with other coffee shops is still lower than other three boroughs.

3. Conclusion

I used python libraries, Geopy client and Foursquare API to explore top 5 boroughs in Seoul. It was very interesting to see the difference between the most frequent venue and the absolute number of coffee shops in Gangbuk-gu. Here, my recommendation to open new cafe is either Gangbuk-gu or Gangseo-gu. However, for those who would like to explore many coffee shops during the visit in Seoul, I definitely

recommend Yongsan-gu and Gangnam-gu. These two boroughs had the most frequent and high number of coffee shops. It should be very convenient to explore cafes in a short, limited time during the visit.