Introduction

Let's assume that one has job offers from a company located in Seoul, S.Korea, he stays in Manhattan, New York City at the moment and want to find as similiar environment as possible to that of Manhattan.

- First, he want to segment & cluster the neihborhodds of the Seoul, S.Korea
- and determine which district in Seoul is most similar or dissimilar to Manhattan, NY.

Brief information about both cities: Seoul officially the Seoul Special City, is the capital and largest metropolis of South Korea. With surrounding Incheon metropolis and Gyeonggi province, Seoul forms the heart of the Seoul Capital Area. I expect that Manhattan would be most similar to Gangnam-gu district in Korea, since it is the most busy and popular place for younger generations.

Data

The Districts (Gu) of Seoul are the twenty-five gu ("districts"; 구; 區) comprising Seoul, South Korea. The gu vary greatly in area (from 10 to 47 km²) and population (from less than 140,000 to 630,000). Songpa is the most populated, while Seocho has the largest area. Gu are similar to London's or New York's boroughs or Tokyo's 23 special wards, and a gu's government handles many of the functions that are handled by city governments in other jurisdictions. This city-like standing is underscored by the fact that each gu has its own legislative council, mayor and sister cities. Each gu is further divided into dong or neighborhoods. Some gu have only a few dong while others (like Jongno-gu) have a very large number of distinct neighborhoods(https://en.wikipedia.org/wiki/List_of_districts_of_Seoul)

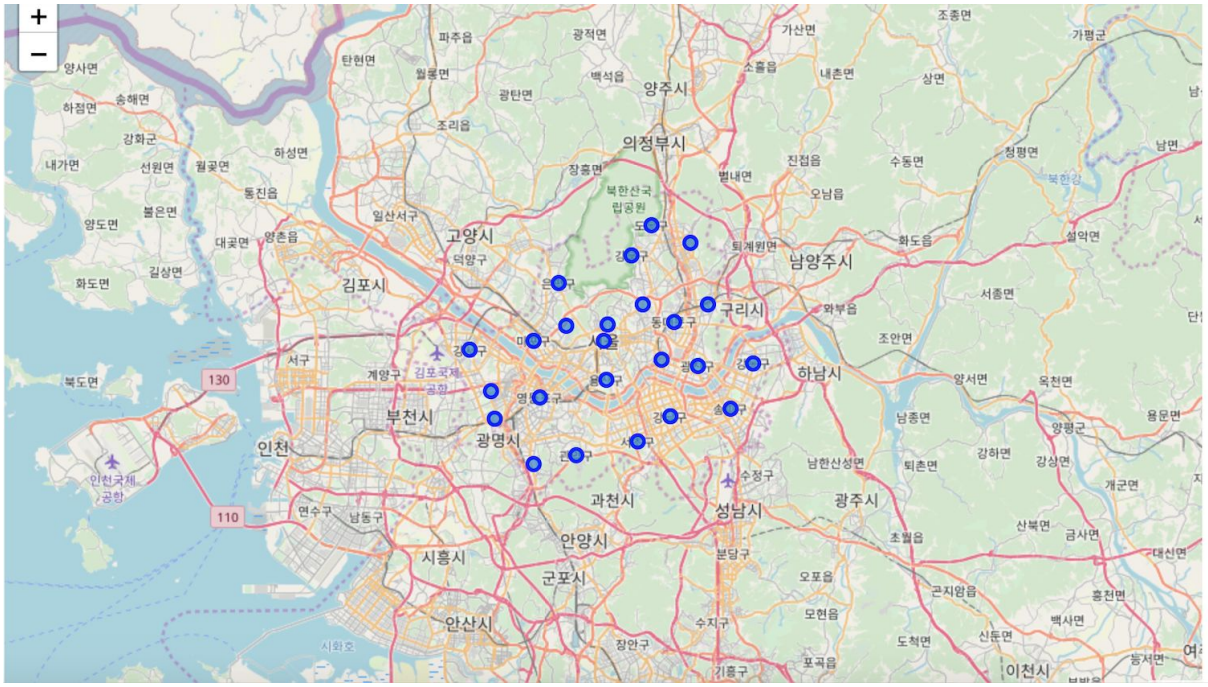
- Seoul7.csv

To get longitude and latitude information for each area, get data from a Korean govenment site(https://www.data.go.kr/dataset/3045281/fileData.do). There are 25 files that include addresses and coordinates of each area of 25 districts in Seoul. Each coordinate of area is calculated by mean() function of every addressable building coordinates' information of the area. For the convenience, I downloaded those 24 files, merged them into 1 file, and placed .csv file on the server

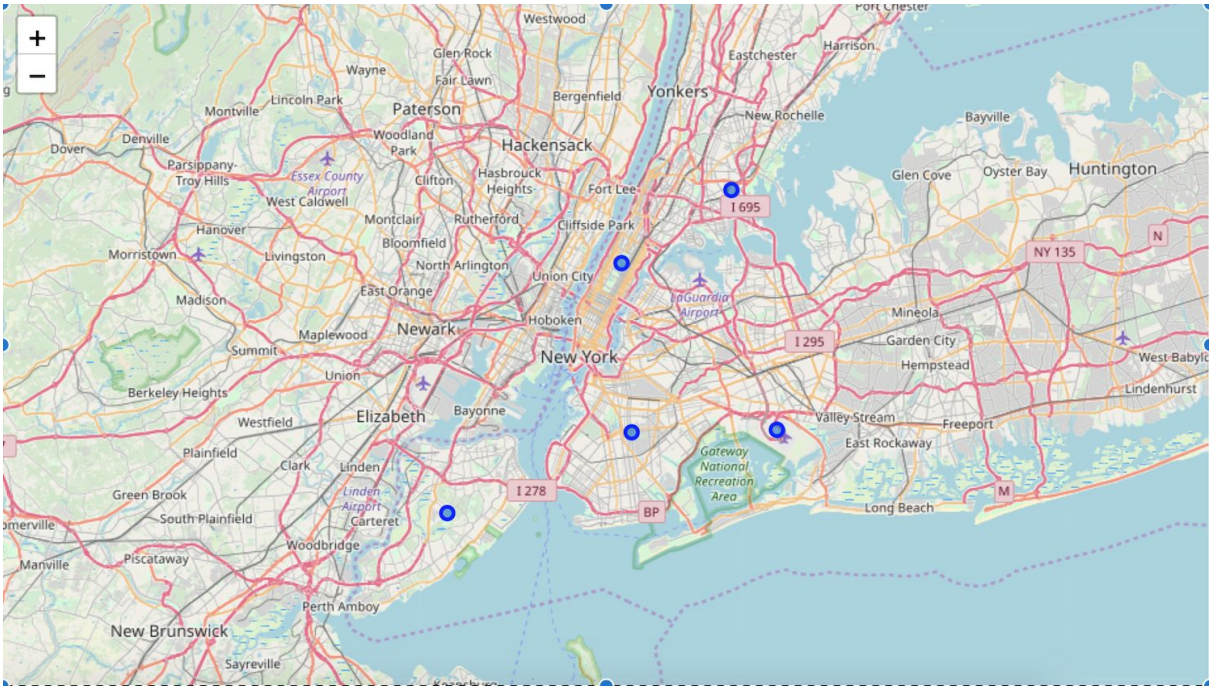
- Seoul_loc2.csv

Note that The coordinates information in the datafile has KR2000 system. We need to change it to wgs84 system. I used pyshp and pyproj packages to convert it. The conversions, however, turned out to be erroneous and in such a case I simply used Nominatim geolocation, which was enough since the comparision was made between Manhattan, New York City and Seoul, S. Korea.

Seoul



New York City



Methodology

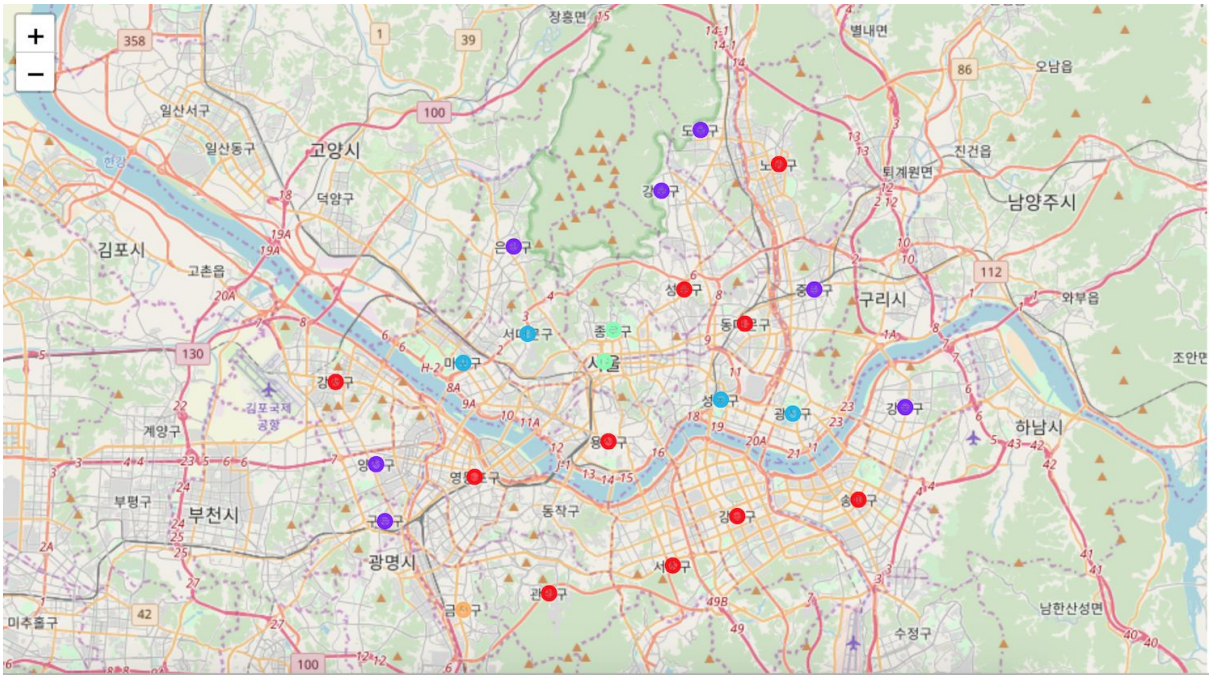
Seoul covers an area of 605/km2, has an estimated population of 9.8M(2018), and has 25 districts, and New York City covers an area of 783/km2(land only), has an estimated population of 8.1M, and has 5 boroughs. Considering size, we first compare 2 cities by grouping boroughs of NY and districts of Seoul.

In order to find similarity and dissimilarity of two cities, Seoul, S.Korea and New York City, I segmented each city into boroughs or districts, and clustered them.

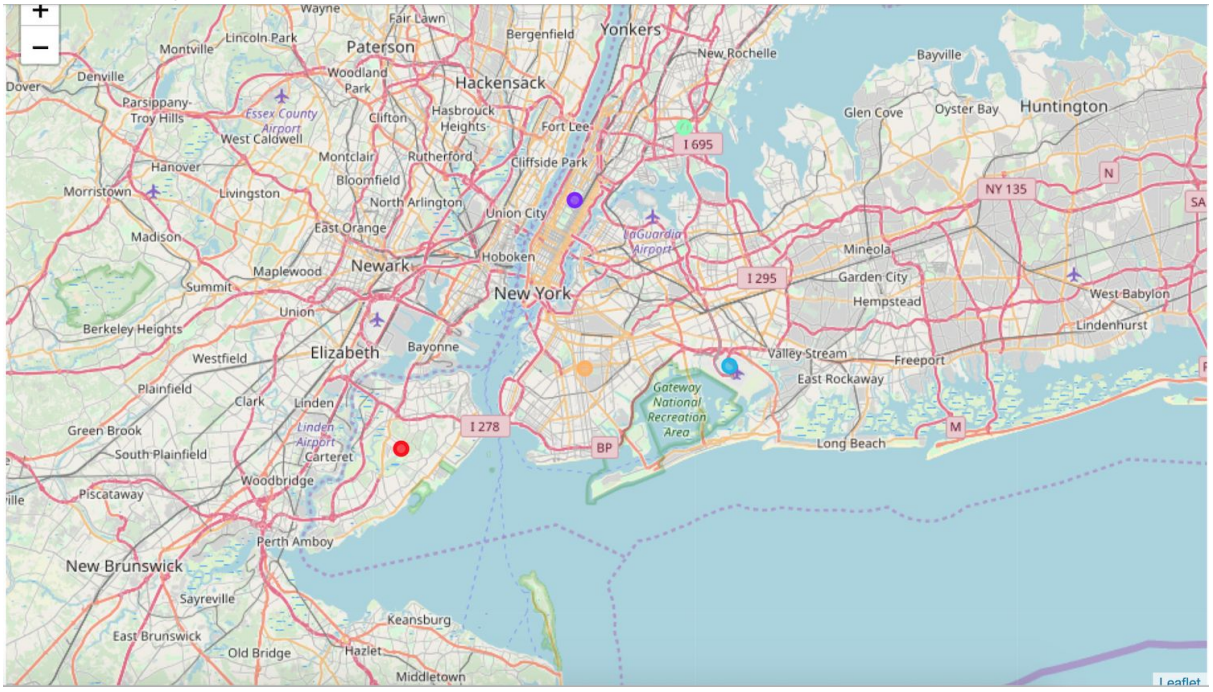
- k-means(k=5)

Manhattan is 59/km2 big and has an estimate population of 1.6M, and Gangnam-gu is 35.55/km2 bia and has an estimated population of 0.65M. For the problem of finding most similar district to Manhattan, we need some distance measurement. Euclidean or Cosine distance measurement will not be adapted naively. we'll leave this problem as a future work.

Seoul



New York City



Results

Since New York City has 5 boroughs, I devided districs of Seoul into 5 clusters. Each cluster seems to be representative as followings.

1. Gangnam-gu,Gangseo-gu,Gwanak-gu,Nowon-gu,Dongdaemun-gu,Seocho-gu,Seongbuk-gu,Songpa-gu,Yeongdeungpo-gu,Yongsan-gu
2. Jungnang-gu,Gangdong-gu,Gangbuk-gu,Guro-gu,Dobong-gu,Yangcheon-gu,Eunpyeong-gu
3. Gwangjin-gu,Mapo-gu,Seodaemun-gu,Seongdong-gu
4. Jongno-gu,Jung-gu
5. Geumcheon-gu

Discussion

- Gangnam-gu,etc., group is a commercial place and center activity
- Jungnang-gu,etc., group is outskirt residential with parks
- Gwangjin-gu,etc., group is residential with coffee shops and cafes but less fast foods
- Jongno-gu,etc., group is Cultural & Tourist area & Hub
- Geumcheon-gu group is residential with fast foods

Manhattan is *Park,**Exhibit,Art Museum**,Yoga Studio,Plaza,Bakery,Garden,Bookstore,American Restaurant,Coffee Shop*. And **Exhibit, Art Museum** seem to be important distinguishing features when Mahattan is compared with districs of Seoul.

Conclusion

The only group with those features is Group0(Jongno-gu, Jung-gu), since it is *, Korean Restaurant **Historic Site,Palace,Art Gallery,Art Museum**,Italian Restaurant,Japanese Restaurant,Bookstore, Coffee Shop,Tea Room.*

Unlike the first though, Manhattan is most similar to Jongro-gu district in Seoul