The Characteristics of Two Big Cities in Asia, Seoul and Tokyo

Adryan R

12 Januari 2021

1. Introduction

1.1 Background

Asia is one of the biggest continents in the world. South Korea and Japan are the major countries for tourism and a great place for living. But each one has its own characteristics and its own uniqueness. Some of us maybe will prefer only one of them. So it's a good moment to think which one is good for us. There is no good or bad city after all everyone needs is different from each other. So why not make an analysis of these 2 cities to find if we can see the difference between these two.

1.2 Problem

The two countries have their own characteristics. Both of them are great places for living, but you cant live in 2 different places at the same time. So we have to find which one is preferable for us. Maybe you want to find some area that meets your needs. Or you want to know which area to visit if you want to spend your holiday there. So let's make the same cluster of these 2 cities, Seoul and Tokyo, and analyze so you know which one is good for you.

1.3 Interest

You who want to visit or live in Asia but don know the characteristics. Or you want to open a business and find which area is good for your business. Maybe you are local but you live in a different country and you want to know each distinct feature in each district of the cities.

2. Data Acquisition and Cleaning

2.1 Data Source

I scraped the data from Wikipedia. There is a link to a list of districts in Seoul and Tokyo. It shows the name of the district. But we need the coordinate of each district to connect it to Foursquare API later.

2.2 Data Cleaning

As we scraped the data from Wikipedia. The format is in text and we need to clean it. So we only get the table we want to keep and delete the other. Using pandas can help you to take the table only and delete the other.

2.3 Data Processing

We have the table of district in each city. Now we need to find the coordinate using geopy. I provide the code below here:

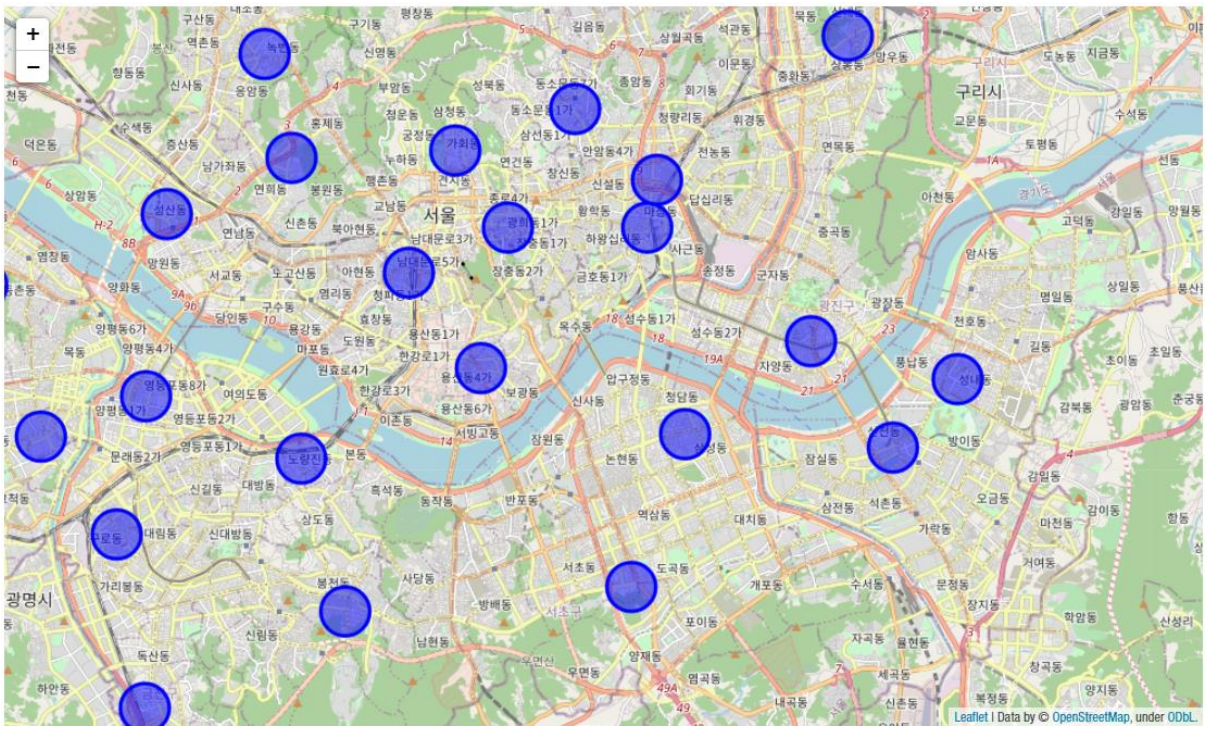
```
from geopy.geocoders import Nominatim
app = Nominatim(user_agent="seoul_explorer")
lat=[]
lng=[]
for name in df['Name']:
    location = app.geocode(name)
    lat.append(location.latitude)
    lng.append(location.longitude)
    print('success')
```

When you run the code and append it to the table we have this table below:

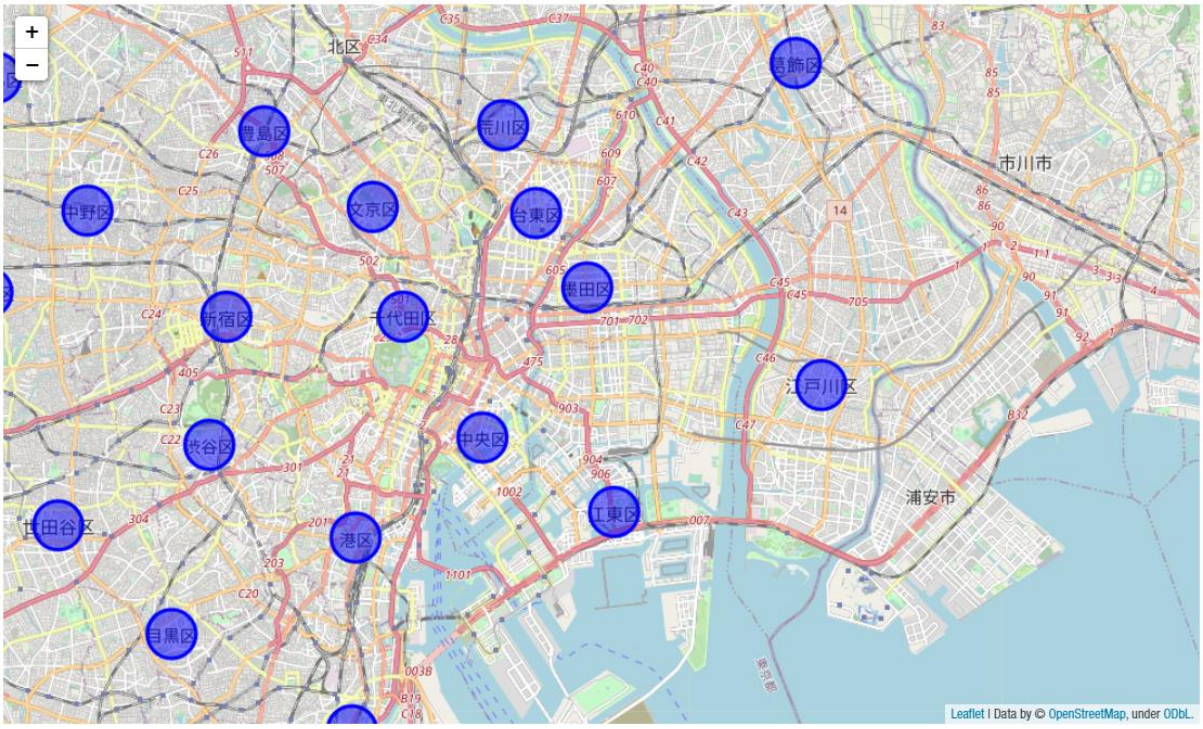
	Name	Latitude	Longitude
0	Dobong-gu (도봉구; 道峰區)	37.668600	127.046600
1	Dongdaemun-gu (동대문구; 東大門區)	37.574200	127.039500
2	Dongjak-gu (동작구; 銅雀區)	37.512100	126.939500
3	Eunpyeong-gu (은평구; 恩平區)	37.602400	126.929300
4	Gangbuk-gu (강북구; 江北區)	37.639500	127.025500
5	Gangdong-gu (강동구; 江東區)	37.530000	127.123700
6	Gangnam-gu (강남구; 江南區)	37.517700	127.047300
7	Gangseo-gu (강서구; 江西區)	37.550900	126.849700
8	Geumcheon-gu (금천구; 衿川區)	37.456500	126.895400
9	Guro-gu (구로구; 九老區)	37.495200	126.887700
10	Gwanak-gu (관악구; 冠岳區)	37.478200	126.951800
11	Gwangjin-gu (광진구; 廣津區)	37.538400	127.082800
12	Jongno-gu (종로구; 鍾路區)	37.580695	126.982799
13	Jung-gu (중구; 中區)	37.563656	126.997510
14	Jungnang-gu (중랑구; 中浪區)	37.606300	127.093000
15	Mapo-gu (마포구; 麻浦區)	37.566571	126.901532

3. Exploratory Data Analysis

We plot it to map to show the location of each district and the map of the city. It will help us to visualize the map of the city:



Map of Seoul



Map of Tokyo

4. Predictive The Model

We will predict the data using K-Means to cluster each district with their own most common venue in their area. Before that, we will find the nearby venue data from

Foursquare API. Foursquare is one of the company that provide dataset of place and venue. After that we will rank top 10 venue in each city and list that to the table below:

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Dobong-gu (도봉구; 道峰區)	Sushi Restaurant	Bakery	Ice Cream Shop	Fast Food Restaurant	Big Box Store	Café	Donut Shop	Farmers Market	Electronics Store	Dumpling Restaurant
1	Dongdaemun-gu (동대문구; 東大門區)	BBQ Joint	Korean Restaurant	Supermarket	Butcher	Metro Station	Donut Shop	Market	Bus Stop	Ice Cream Shop	Department Store
2	Dongjak-gu (동작구; 銅雀區)	Seafood Restaurant	Café	Korean Restaurant	Fast Food Restaurant	Donut Shop	Fried Chicken Joint	Coffee Shop	Ice Cream Shop	Gym	Japanese Restaurant
3	Eunpyeong-gu (은평구; 恩平區)	Korean Restaurant	Sushi Restaurant	Concert Hall	Bus Stop	Coffee Shop	Fried Chicken Joint	Bakery	Clothing Store	Ice Cream Shop	Gukbap Restaurant
4	Gangbuk-gu (강북구; 江北區)	Coffee Shop	Donut Shop	Café	Korean Restaurant	Brewery	Fast Food Restaurant	Bakery	Bookstore	Ice Cream Shop	Gym

Table of Common Venue in Seoul

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Adachi	Convenience Store	Japanese Restaurant	Supermarket	Discount Store	Restaurant	Drugstore	Pharmacy	Park	Sake Bar	Japanese Family Restaurant
1	Arakawa	Grocery Store	Convenience Store	Park	Bus Stop	Intersection	Ramen Restaurant	Chinese Restaurant	Food & Drink Shop	Noodle House	Teishoku Restaurant
2	Bunkyo	Café	Japanese Restaurant	Museum	Convenience Store	Chinese Restaurant	Park	Bus Stop	Steakhouse	Szechuan Restaurant	Grocery Store
3	Chiyoda	Chinese Restaurant	Café	Ramen Restaurant	Historic Site	Art Museum	Sake Bar	Science Museum	Sushi Restaurant	Stadium	Comedy Club
4	Chūō	Sushi Restaurant	Soba Restaurant	Italian Restaurant	Bakery	Tonkatsu Restaurant	Seafood Restaurant	Spanish Restaurant	Café	Park	Tempura Restaurant

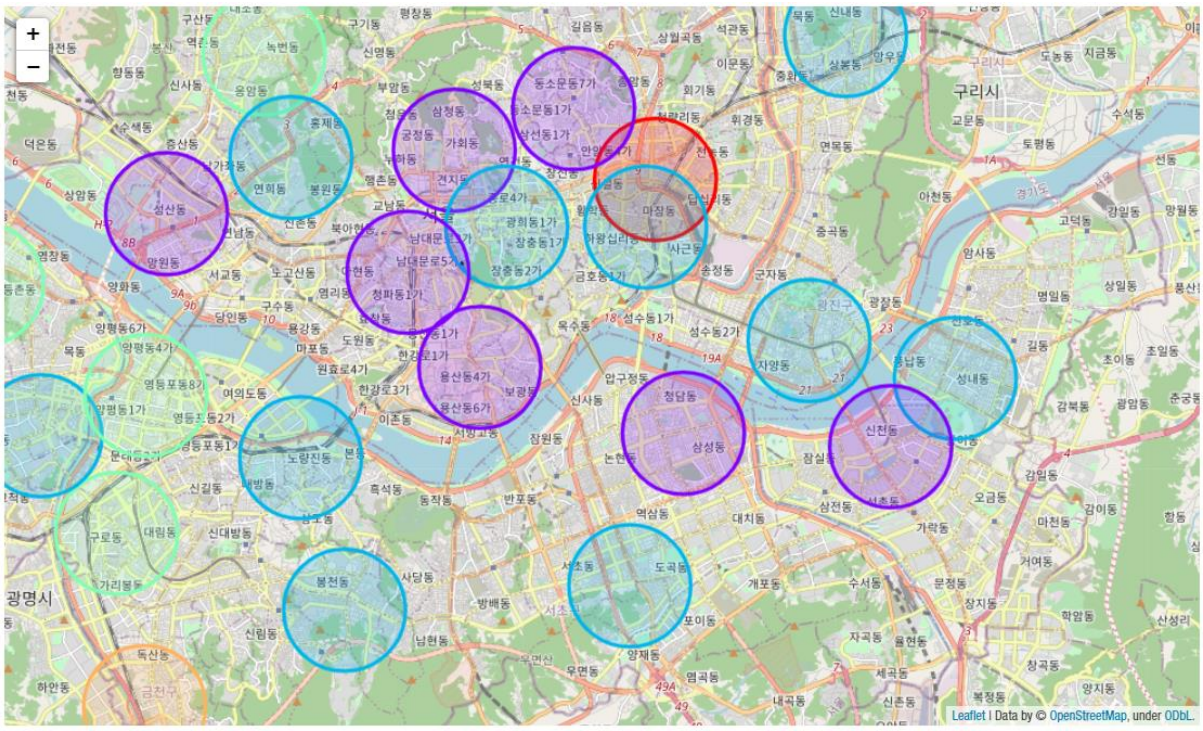
Table of Common Venue in Tokyo

We will predict the cluster using K-Means method and here I provide the code for you to read:

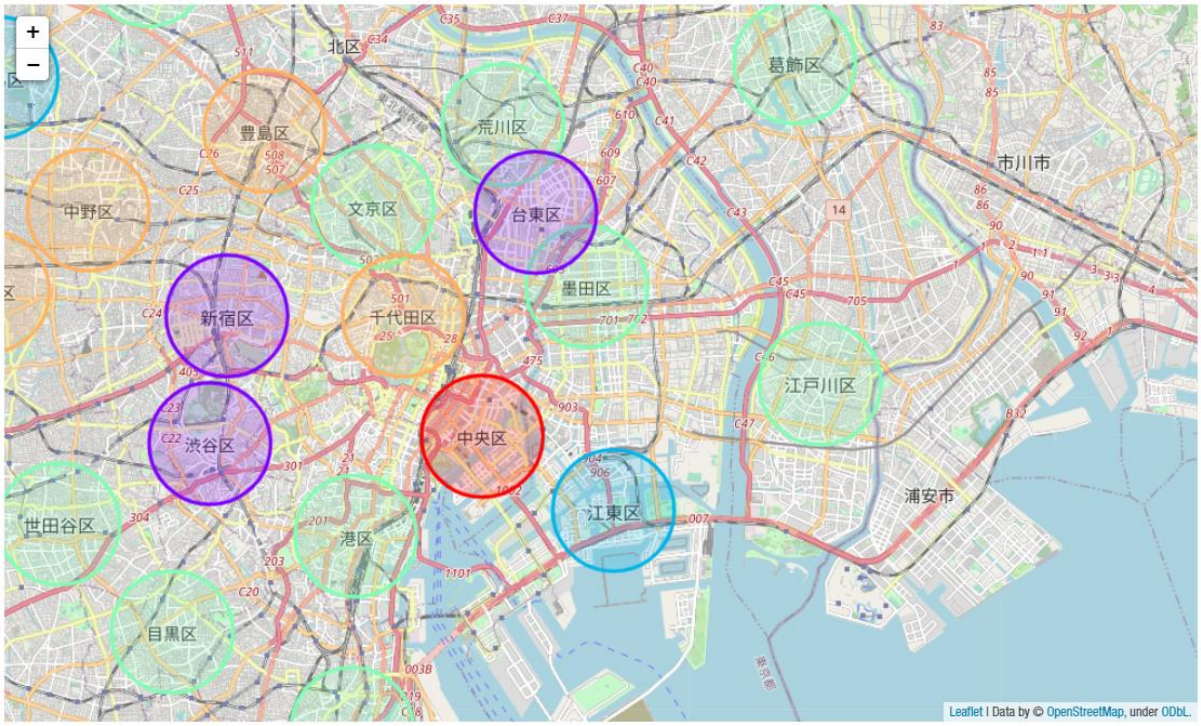
```
from sklearn.cluster import KMeans
# set number of clusters
kclusters = 5

tokyo_grouped_clustering = tokyo_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(tokyo_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:5]
```

5. Result

After we predict the model using K-Means we plot it to map to visualize the cluster on each city.



Cluster Map of Seoul



Cluster Map of Tokyo

Now we have the map of cluster on both cities. The majority of cluster in Seoul is dominate by café and restaurant. While in Tokyo is dominate by convenience store and restaurant.

For the detailed data you can read it in description. The map can be pop out to read the result of each district.

6. Conclusion

Both city are great place for living. Seoul has different café and restaurant vibe and it's a great place to visit. Seoul has lot of place for hangout and shop. While in Tokyo you will find that convenience store is such an interesting place. Every place you visit you will have 5 minute walk range to convenience store. So you will have easy time if you need something to buy. And Tokyo has different vibe of restaurant and modern public transport that can help you go anywhere easily.

7. Future Direction

I am using label and K-Means to validate the data. Next time I will try to use geospatial data so it can map and visualize the area of each district to and give better result using DBscan method.