Explore Restaurant in Seoul

Background and Project Objective

Music and film industries from South Korea nowadays has spread widely all around the world. Many people interested in South Korea as a country since K-Pop and K-Drama become more famous. Following that fact, no wonder that South Korea now become one of favorite destination for travelling from tourist all around the world. Seoul as a capital city of South Korea become one of interesting city and can attract more visitors than the other city. This phenomenon can be an advantage from business point of view. Seoul is a potential city for those who are looking for a good city to start the business. However, in this project, we are focused on restaurant business.

Choosing a location is very important especially for business restaurant. Seoul has 25 districts. The objective of this project is to find out what district is more potential if we would like to open new restaurant, and also what kind of restaurant is popular in that district.

Result of this project will be useful for these people:

1. Business owner who are looking for a potential location to open a new restaurant.
2. Business owner who want to know what kind of restaurant is popular in a certain district.
3. Tourist who want to know what venue that they can visit if they go to certain district.
4. Data scientist beginner that looking for some references.

Data Collection

1. Web scrapping from Wikipedia website
   I did web scrapping for Wikipedia website using BeautifulSoup to obtain data about district in South Korea. Here is the link : https://en.wikipedia.org/wiki/List_of_districts_in_South_Korea

   Data about district in South Korea was filtered so it only contains data about district in Seoul. After I got the data about district in Seoul, I was using Foursquare API to get the list of nearby venues in each of district in Seoul. When list of nearby venues obtained, the data was filtered so it only contains list of Restaurant in each of district in Seoul.

	District	City	Population	Area
1	Dobong District	Seoul	366,879	20.70
2	Dongdaemun District	Seoul	366,633	14.20
3	Dongjak District	Seoul	402,567	16.35
4	Eunpyeong District	Seoul	491,741	29.71
5	Gangbuk District	Seoul	345,502	23.61

2. Get location data
   I upload another csv file to notebook that contain location latitude and longitude for each of district in Seoul. After that, previous data frame merged with location data frame. Here are final data frame about district in Seoul.

	District	City	Population	Area	Latitude	Longitude
0	Dobong District	Seoul	366,879	20.70	37.670255	127.032688
1	Dongdaemun District	Seoul	366,633	14.20	37.582540	127.055219
2	Dongjak District	Seoul	402,567	16.35	37.499687	126.951246
3	Eunpyeong District	Seoul	491,741	29.71	37.619948	126.925604
4	Gangbuk District	Seoul	345,502	23.61	37.645173	127.009270

3. Foursquare API to get nearby venues in each district
   I used Foursquare API to get nearby venues in each of district using district location. After obtained list of nearby venues in each of district, I filtered it, so it showed only restaurant and drop other venues. From Foursquare API, I got details about nearby restaurant like restaurant name, location of the restaurant and the category (Korean restaurant, American restaurant, etc.)

	District	Dist_Latitude	Dist_Longitude	Venue	Venue_Lat	Venue_Long	Venue_Category
1	Dobong District	37.670255	127.032688	수정궁	37.662404	127.032934	Chinese Restaurant
2	Dobong District	37.670255	127.032688	대문	37.663411	127.027870	Korean Restaurant
3	Dobong District	37.670255	127.032688	맥도날드 (McDonald's) (맥도날드)	37.670196	127.043726	Fast Food Restaurant
4	Dobong District	37.670255	127.032688	강촌칼국수	37.662150	127.030537	Korean Restaurant
5	Dobong District	37.670255	127.032688	일심해장국	37.661906	127.029542	Korean Restaurant

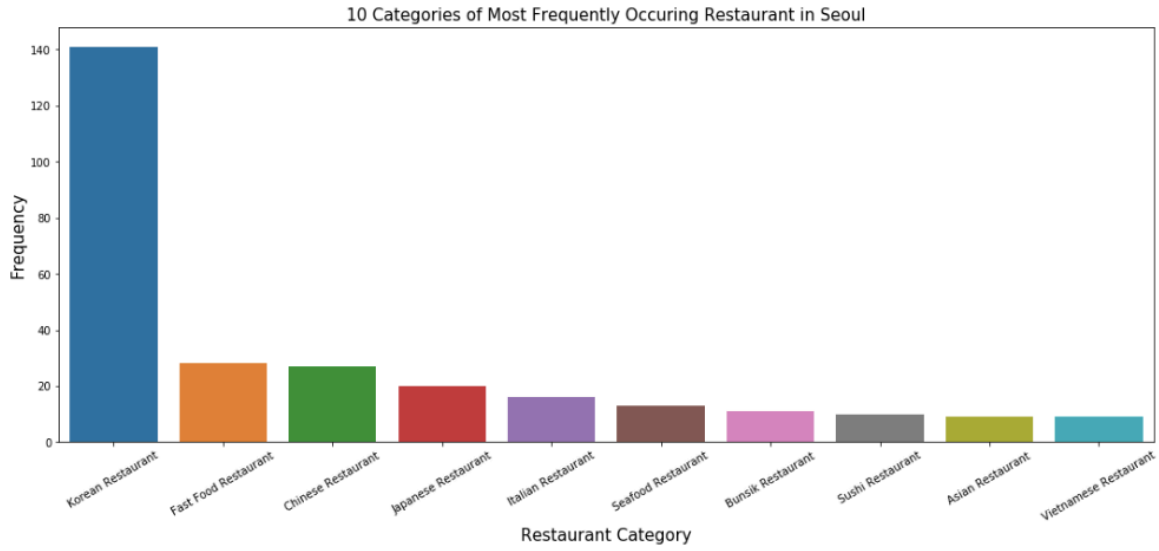
The next question is how to use this data to achieve the objective of this project? We can use data about nearby restaurant in each of district to find out what kind of restaurant is popular in that district since we can see what restaurant category is frequently occurring. We also can analyze what district is potential if we want to open new restaurant by looking for number of restaurants in that district. The details will be explained in result and discussion section.

Methodology

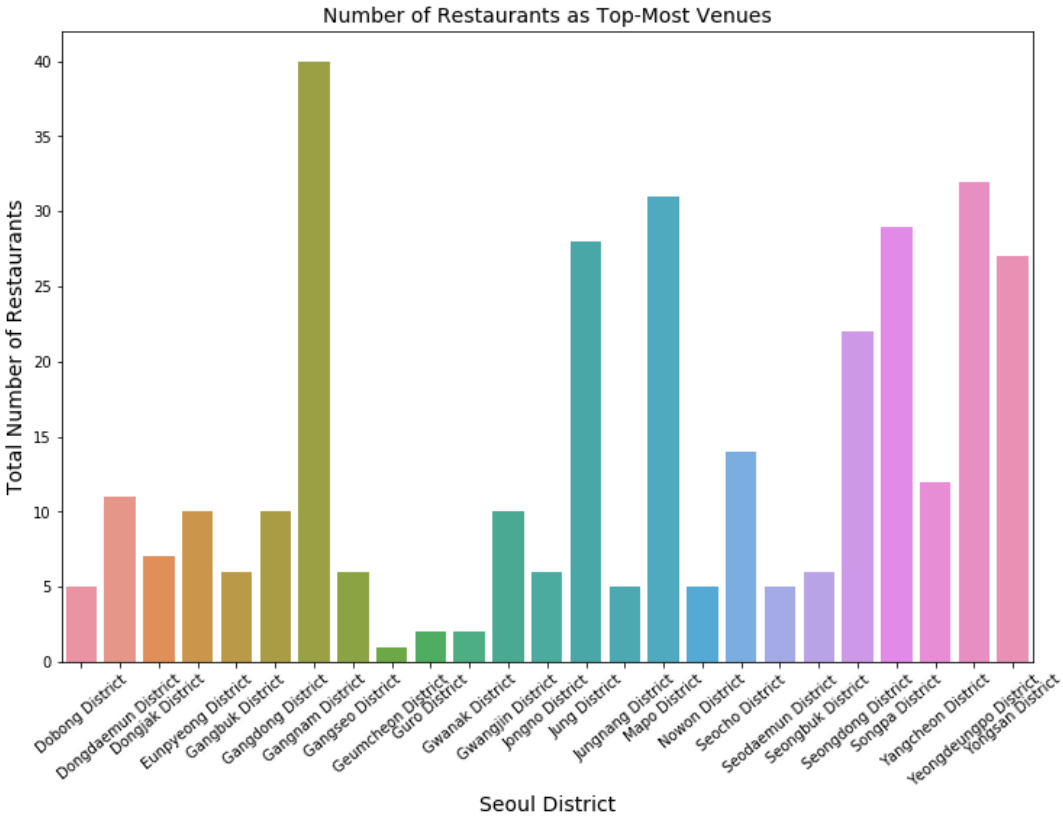
We already had data about restaurant in each of district in Seoul. From that data, we are looking for unique categories to find out how many restaurant categories in Seoul. We are focused on top 10 restaurant category that frequently occurred in Foursquare data. We got this data that contain top 10 restaurant category along with the number it is occurred in Foursquare:

	Venue_Category	Frequency
1	Korean Restaurant	141
2	Fast Food Restaurant	28
3	Chinese Restaurant	27
4	Japanese Restaurant	20
5	Italian Restaurant	16
6	Seafood Restaurant	13
7	Bunsik Restaurant	11
8	Sushi Restaurant	10
9	Asian Restaurant	9
10	Vietnamese Restaurant	9

We visualize previous data frame to make it easier to compare between restaurant category



We also find out number of restaurants in each of district to see what district that has more option for restaurant. From data that we obtained from Foursquare, we grouped it by District and Venue Category and count number of restaurants in that district. Here is what we got:

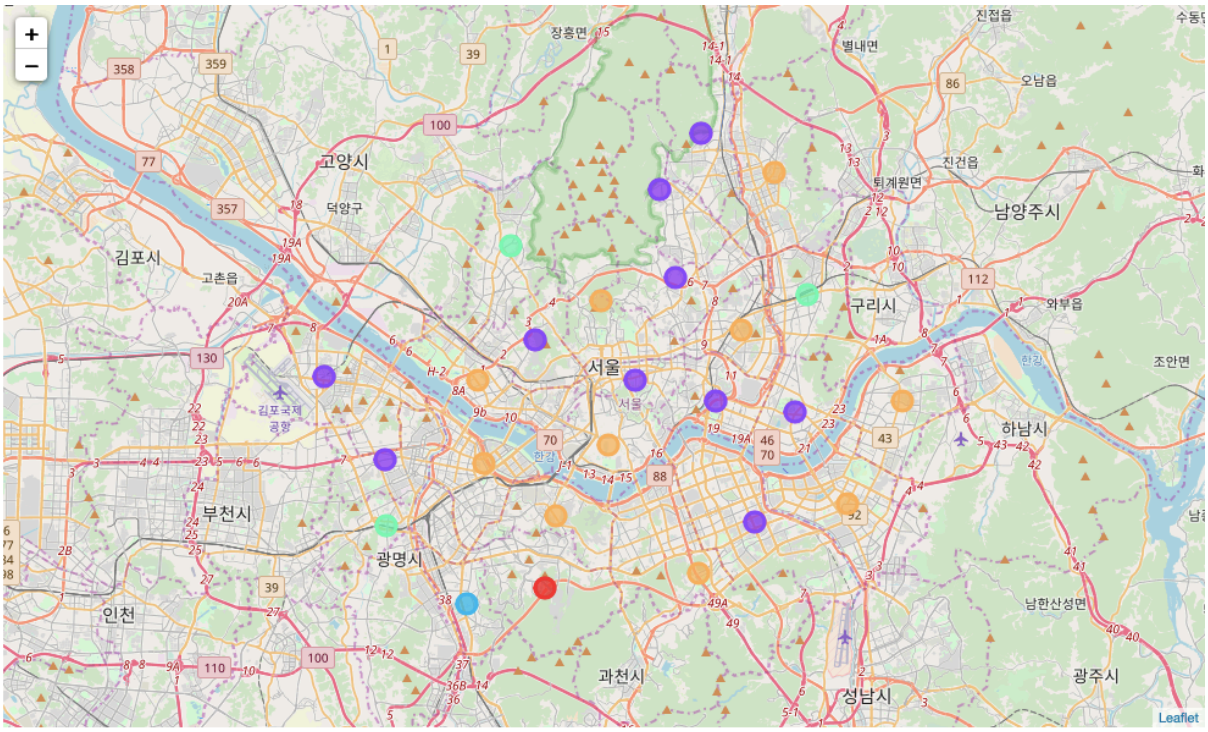


Next I'm going to see top 5 restaurant categories in each of district. To see that, use one hot encoding to get more information about restaurant categories. After that, using group by District and transpose the data. We got order of top restaurant categories in each of district along with the frequency (frequency obtained from one hot encoding). Here is small part of what we got (because there are 25 districts, so I can't show it all)

```
Dobong District
                  Venue  Freq
0     Korean Restaurant   0.6
1    Chinese Restaurant   0.2
2  Fast Food Restaurant   0.2
3       Soba Restaurant   0.0
4      Ramen Restaurant   0.0


Dongdaemun District
                        Venue  Freq
0          Korean Restaurant  0.27
1        Japanese Restaurant  0.18
2         Italian Restaurant  0.18
3  Japanese Curry Restaurant  0.09
4         Chinese Restaurant  0.09
```

For the final part, I'm doing clustering for all districts in Seoul using K Means clustering. The objective for this clustering is to find out similarities between the districts based on restaurant categories. I used k=5 for this clustering. After visualizing it using map, here what we got:



Results

From the previous section, here are some results:

1. Among all of districts in Seoul, the most frequently occurred restaurant category is Korean Restaurant. And the difference is quite high that you can see in the bar graph.
2. If we compare it between the district, almost all of district has Korean Restaurant as the most frequently occurred restaurant category.
3. Top 3 categories for restaurant category in Seoul is Korean Restaurant, Fast Food Restaurant and Chinese Restaurant.
4. Gangdong District is a district that has most restaurant among all of districts in Seoul.
5. From the clustering, here are the district that have similarities for restaurant categories:

   Cluster 0: Gwanak District
   Cluster 1: Gangseo District, Seodamun District, Jung District, Yangcheon District, Gangnam District, Gwangjin District, Seongdong District, Seongbuk District, Gangbuk District, Dobong District
   Cluster 2: Geumcheon District
   Cluster 3: Guro District, Eunpyeong District, Jungnang District

Cluster 4: Yeongdeungpo District, Mapo District, Yongsan District, Dongjak District, Seocho District, Songpa District, Gangdong District, Dongdaemun District, Nowon District, Jongno District

Discussion

Based on previous section, the district that has most frequently restaurant among all of district is Gangdong District. Korean restaurant is the most frequently restaurant in Gangdong District and also the most popular category in overall district in Seoul. We can see that number of Korean Restaurant in Seoul is big and the difference from another restaurant category is quite high. So Korean cuisine is still number one for people in Seoul.

In top 5 restaurant categories, we can see that 3 of them are Korean Restaurant, Chinese Restaurant and Japanese Restaurant. China and Japan are located near Korea so maybe it has some similarities in the taste of cuisine. That's why it is quite popular in Korea, especially in Seoul.

For the clustering, I only use restaurant category in each of district to determine the similarities between all of district. It would be better if there is another factor to determine such as average price of the food in each restaurant, user rating in restaurant, etc. Those data can be obtained from Zomato by using Zomato API. So, I advise to combine data from Foursquare API and Zomato API to get more info about the restaurant in each of district.

Conclusion

- If you are business owner that considering what kind of restaurant is popular, then Korean Restaurant is the best choice.
- If you are choosing for the right location to build new restaurant, the answer can be different depend on what you are looking for. If you are looking for the district that has many restaurants so probably there are much tourist there, then consider these districts:
  1. Gangdong District
  2. Yangcheon District
  3. Jungnang District
  4. Seongdong District
  5. Jongno District

  Otherwise, if you are looking for the district that has least restaurant so it can reduce your competitor, then consider these districts:
  1. Gangseo District
  2. Geumcheon District
  3. Guro District

- If you are tourist who looking for district that has many options for restaurant, then Gangdong District is the best choice.