

# IBM\_Coursera Capstone Project

## -The Battle of Neighborhoods-

Written by: JinSeob Shim

### The 2nd-week assignment

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

Seoul, officially known as the Seoul Special City, is the largest metropolis and capital of South Korea. This megacity is the largest city proper in the developed world and the Seoul Capital Area is the second largest metropolitan in the world with more than 25.6 million people, which is half of all the residents in the country. The estimated population within the city limits for 2016 is 10.29 million. Interestingly, Seoul's population density is almost twice that of New York City, four times higher than Los Angeles and eight times higher than the density of Rome.

People from around the world come to Seoul, Korea with various reasons to make the city more diverse and dynamic. The number of people coming to Seoul has been growing due to Korea's remarkable economic growth and its cultural power so-called K-wave which includes k-drama, tv show, k-pop, and etc., The more interesting thing is people have more likely tended to stay in the city for more than three months and they have been willing to experience a variety of things not only trying related to Korean pieces of stuff but also trying diverse and mixed vibe by visiting something new places in the city.

Plus, the number of Korean people who have been to abroad for various purposes has incredibly risen so especially young generation -under 40 years old- they are open to experience other culture. In particular, when it comes to the size of the food industry for foreign foods in Seoul recorded approximately 35 billion dollars as of 2016 according to the Statistics Bureau of Korea.

Therefore, there are a growing number of potential investors who want to open or operate a unique or trending venue like a restaurant in the city. They might wonder what kind of venue they should choose to invest or open in order to avoid unnecessary risks like getting into red ocean competition. With geospatial data and data science techniques, we can provide information about which type of venue is the most popular and clustered distribution of venues within Seoul to a potential investor. Based on this information, we can suggest a better business investment to the investor. And it helps he or she make an informed decision to invest in Seoul particularly by getting them to know about the geospatial distribution of data by districts related to the economic activity of people in Seoul.

**-Target clients: potential clients looking for a business opportunity in Seoul but not sure because of insufficient knowledge about geographical economic features of the city.**

**-Stakeholders**

- 1) **Seoul city government**
- 2) **Sellers**
- 3) **Buyers**
- 4) **Real estate agents**
- 5) **Tourists**
- 6) **gourmet**

## 2. Data Section

Here are data sources with description used for doing the capstone project below:

- 1) List of districts of Seoul with population and its density as of 2017:  
[https://en.wikipedia.org/wiki/List\\_of\\_districts\\_of\\_Seoul](https://en.wikipedia.org/wiki/List_of_districts_of_Seoul)
- 2) Average Monthly income level by districts in Seoul from Seoul Statistics Bureau:

=> applied usage of data in pdf format in Korean as of 2017

- 3) Average Monthly consumption level by districts in Seoul from Shinhan Bank in Korea

=>price paid data in pdf format in Korean as of 2017

**I converted three data above into CSV format**

- 4) Location coordinates (latitude, longitude) of district names are fetched by API calls:

- Data item: Google Maps Geocoding API

- Type of data: JSON

- 5) The recommended districts and types of venue will be determined with data from Foursquare API calls which include venue and its category data

- Data item: Foursquare location data

- Type of data: JSON

- Source: <https://foursquare.com/>

- 6) Important facilities such as Hospitals and Universities will be searched and provided in a vicinity of each location and then reported in a tabular form to the client.

### 3. Methodology

With this data, I will be able to suggest the best category of venue (probably kind of restaurants) by districts to clients according to the resident's economic strength and activity tendency with rental cost estimation.

- 1) Suppose income level, consumption level, population, population density, and monthly rent cost are considered economic strength factors.
- 2) Venue and its category from Foursquare API call are defined as venue factors.
- 3) Cluster districts with economic strength factors by using K-Mean algorithm
- 4) Cluster districts with venue factors by using K-Mean algorithm
- 5) Create an economic cluster map and venue cluster map respectively
- 6) Do exploratory analysis through visualization of each cluster data with pivot tables, graph, and scatter plots to find insights
- 7) Analyze the combinations of the economic clusters and venue clusters then find 5 recommendable combinations based on economic possibility and potential factors.
- 8) Recategorize combination of economic clusters and venue clusters by the mixed feature analysis and draw it on the map
- 9) Make suggestions for each recommendable recategorized clusters

## 4.Results

Seoul is comprised of 25 districts in which its top 10 venue categories people frequently visit can be divided into 3 clusters respectively.

Each venue cluster has its own features as follows:

- (1) **Venue 0(Red)**: Venue 0 is the most populated but least spend money although they earn money than any other venue clusters. However, the rental cost is the cheapest among Venue clusters.

**Coffee shop, Park, BBQ, Chinese restaurant, Market, Fast Food restaurant, Bakery, Noodle House, Multiplex, Hotel**

People probably tend to spend a small amount of money at venue 0 since products cost of the venues usually cheap and people want to go to other places without staying long. It also can infer the number of floating population and resident many as people frequently visit Fast food restaurant and market to buy groceries

- (2) **Venue 1(purple)**: It spends the largest money than any other venue cluster although it shows the least populated and most expensive rental cost. Venue 1 cluster spend monthly about 210 dollars more than venue 0 and even spend approximately 100 dollars more than venue 2. Interestingly, their monthly income level is the lowest even though it shows a slight difference compared to other venues about 92 dollars, 65 dollars for venue0 and venue2 respectively.

**korean restaurant, cafe, BBQ Joint, sushi&Japanese restaurant,**

**Historic site,palace, hotel, plaza, art gallery, cocktail bar, French restaurant**

Venue 1 has cultural, touristic, and expensive places.

It is the old central part of the city as it has palaces and historic sites. Therefore, the level of willingness to pay is more likely high and presumably many foreigners are likely to visit places at venue cluster 1 with group or guide

- (3) **venue 2 (light green)**: it is in the middle among clusters in every aspect. The majority of district belong to venue2 cluster.

Venue2 has a variety type of venue due to many districts belong to cluster 2

**Coffee shop&cafe, Korean Restaurant BBQ, Bakery, Park, Chinese**

**restaurant, Japanese restaurant, Hotel, Historic site**

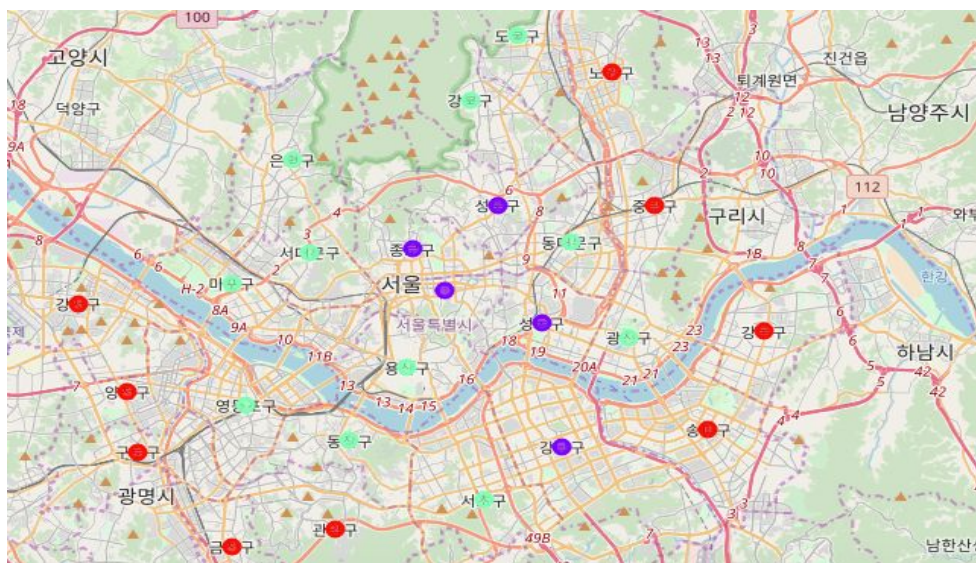
-At the top5 places such as cafe and Bakery, people tend not to spend a lot of money even though they can stay longer. However, they may visit those places often as it is rather cheap or let people feel comfy and refreshing.

-People may enjoy looking around Venue cluster 2 since it has also historic sites including palace, museum, bookstore. That can produce a well-mixed vibe which comes from a good combination of modern and the past architecture with leisure facility like a book store, a museum at a lower cost. It leads people to come and go often.

-People also visit often Japanese restaurant and an Italian restaurant where its cost is pricey.

-Tourist info Center and Hostel may suggest there are backpackers who want to see around the city and stay at a low cost.

### Venue clusters map

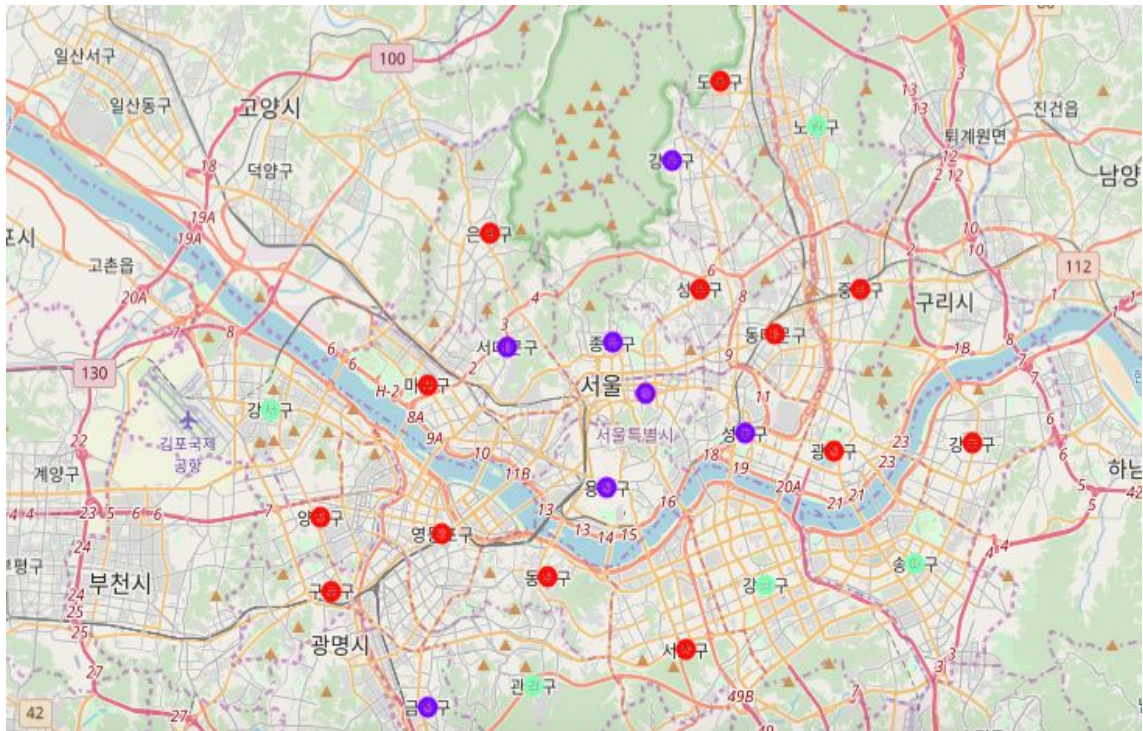




## Economic Strength Clusters

- label 0 (red) region is in the middle between other labels=> In terms of consumption, however, they tend to spend less money given that their income level compared to the consumption level of label 1 region. In other words, they earn more but less spending compared to that of label 1. Plus, label 0 shows the highest level of its population density.
- label 1(Purple) region is the least populated and less likely to spend money than any other labels. However, their consumption tendency is relatively high despite its income level.
- label 2(light green) region is the most populated and wealthy (income and spending level is the highest among neighbors)

## Economic Strength clusters map



## 5.Discussion

Factors mostly should think about when Investing assets in the city are perhaps how many people live with buying power and popular venues people often visit nearby places. Through putting significant factors together into consideration, five recommendable districts can be provided as follows.

- Districts with venue cluster 1 & economic label 2 combinations show the strongest buying power and need the most expensive investment in renting places=> "1"
- Districts with venue cluster 2 & economic label 0 combinations show the second strongest buying power and need third most expensive in renting places=> "2"
- Districts with venue cluster 1 & economic label 0 combinations show the third strongest buying power but only need 680 dollars for monthly rent cost which is the second cheapest among rental cost =>"3"
- Districts with venue cluster 0 & economic label 0 combinations show the second weakest buying power but it is the very attractive since not only cheapest region it is but also have the fourth highest income level: potential growth in buying power is high =>"4"
- The district with venue cluster 0 & economic label 2 combination have the largest population and it has second strongest growth potential in buying power=>"5"

In sum, recommendable venue(hereafter 'v') and eco cluster(hereinafter 'eco') combinations are as follows:

1) v1,eco1="1" (purple)

2) v2,eco0="2" (light blue)

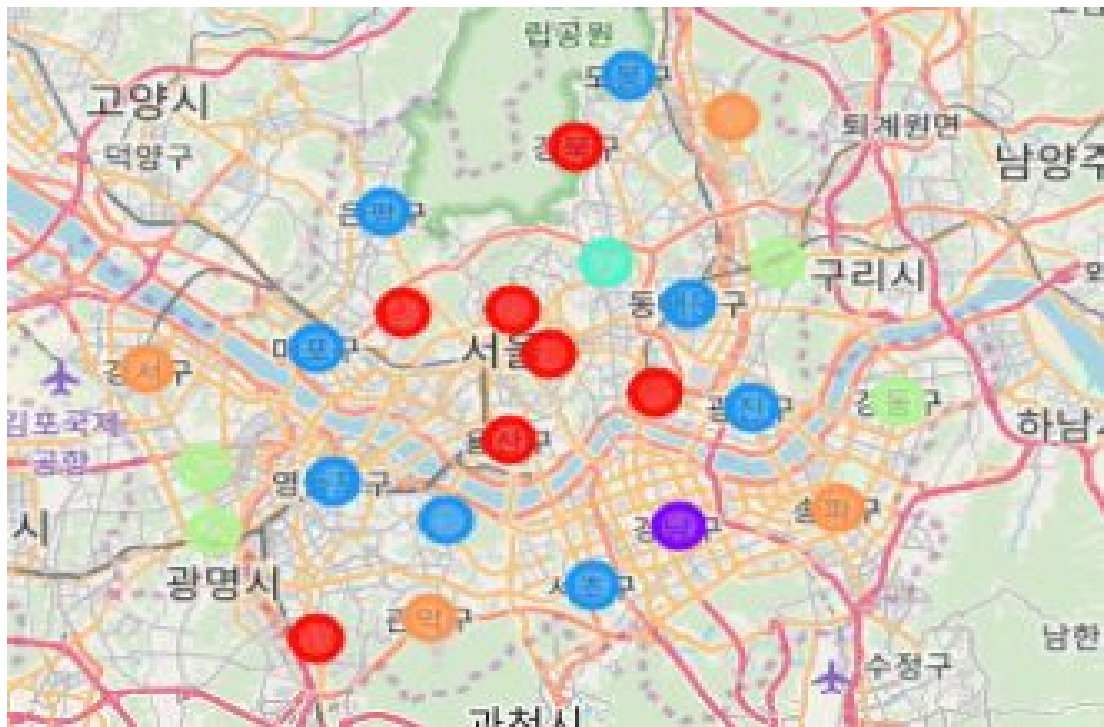
3) v1,eco0="3"(light mint)

4) v0,eco0="4"(light green)

5) v0,eco2="5"(light ochroid)

6) remainders="6" => no invest

## Final suggestion districts clusters map



## 6.Conclusion

Suggest venues or business investment according to the recommendable districts clusters as follows.

-Cluster 1:

Rental cost high (monthly average rental cost \$880) but most lucrative district (monthly average consumption \$1950). The level of willingness to pay is the highest. It is recommended to invest in opening a high-end restaurant such as Japanese or luxurious western style restaurant. Interior design with a mixed Korean-western style would be much attractive to tourists as well.

-Cluster 2:

It has the second strongest buying power and needs the third most expensive in renting places. It has places like historic sites, museum, and park which lead people to enjoy looking around the venues in cluster 2 as a group or family. The residents in cluster 2 tend to not only spend money and their time on cheap and comfy products



like visiting cafes but also spend on a kind of expensive restaurant. Given that they have the second strongest buying power Cluster 2 districts also do have many business offices as it functions subcenters of the city, investors should be careful whether invest in expensive and fancy western or Chinese and Japanese cuisine places or in a cheap and fast food chain as well as franchise cafe.

It is recommended investing in the mid-high price range of western and Chinese restaurant which can serve meals relatively fast with cafeteria or buffet style to offset its expensive renting cost and make the use of its buying power of busy people.

#### -Cluster 3:

Even if the residents make the lowest level of money in the city, it shows the third strongest buying power but only needs 680 dollars for monthly rent cost which is the second cheapest among clusters. Investors can think about doing lodging business with Korean style architecture and guest house with foods services as the district close to the old center of the city. Plus, there are relatively many traditional Korean houses than any other districts. You can meet the demands of tourists who want to experience Korean culture by providing places, foods, and opportunities to meet new people or casually meet local people.

#### -Cluster 4:

It shows the second weakest buying power but it is very attractive since not only the cheapest region it is in terms of rent cost but also have the fourth highest income level. In addition, they represent the highest level of population density in the city it is probably because Cluster 4 districts are located in the outskirts of the city to function as bed towns.

Therefore, the local residents visit the markets often for grocery shopping and pass by fast food restaurant or Korean restaurants. They tend not to stay and not to spend money in the districts. In that regards, opening a cafe which starts early in the morning with healthy dessert or fast food restaurant with delivery service and running a family restaurant at a mid-row range cost would be considerable. A venue can easily make regular customers with low cost and large sales. It is essential to provide

quick and cheap but reliable diet or service in the districts. One thing should be reminded that its consumption potential in buying power is the highest of all districts.

-Cluster 5:

Cluster 5 has the largest population and indicates the second strongest consumption potential in buying power. Interestingly it shows the residents' income level is the second highest(\$4312) but they spend a lot less (they uses only \$1390 which is the third lowest of all districts).

It can infer there are not many places let people open their wallets. Even though its rent cost is somewhat expensive, it has pretty good opportunities there with the largest population and economic strength. It is recommendable to invest in a luxurious cafe and restaurant with fancy European interior. It needs more research about the trend since the residents are wealthy enough but spend less. It suggests referring to trend venue of cluster 1 as it has a similar income and rent cost level to boost its consumption of cluster 5.

### **Limitations**

Although the project allows recognizing Seoul's distinctive features by its districts through machine learning algorithms, it needs more research and factors such as consumption information or trending places and hours at street level should be put into in order to provide more reliable and specific information to customers.

Unfortunately, it was not able to collect and be put into the algorithm due to lack of budget and skills.

However, It was a great pleasure to apply knowledge and skill that I have learned and try to find insights as if a real data analyst or scientist.

Thanks.

Sincerely,

Jinseob Shim

