Analysis of MFC location in Seoul

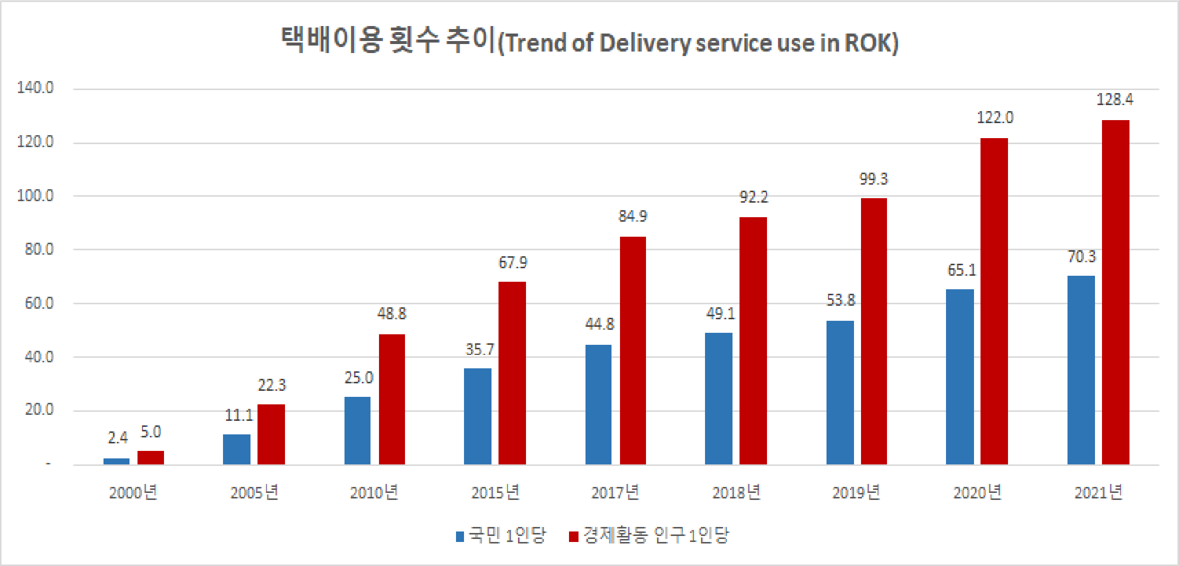
Data Mining Application Team Project

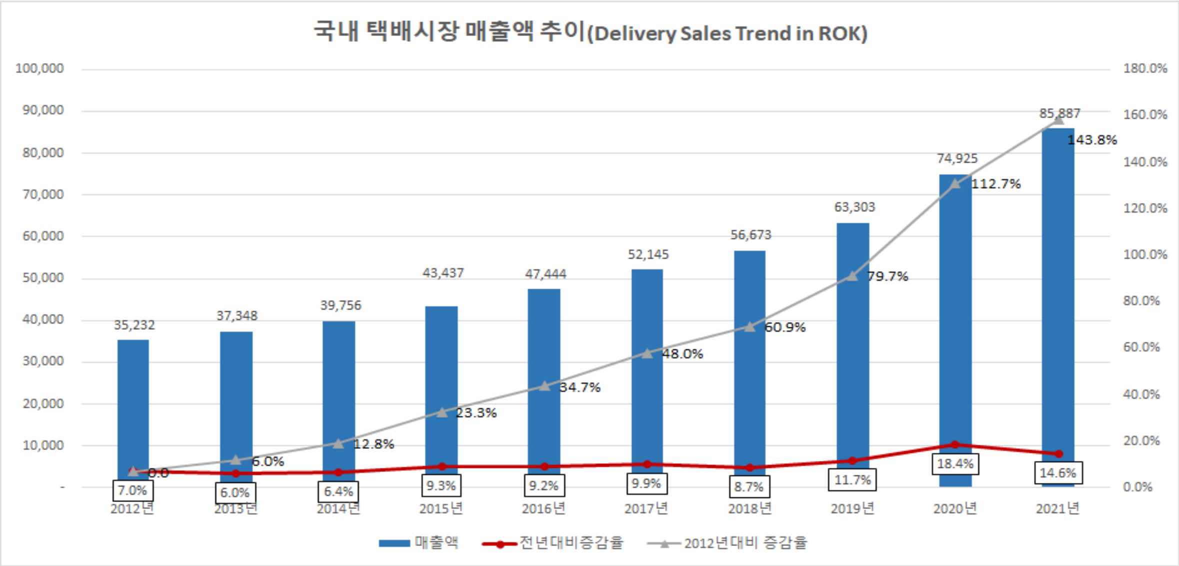
201800302 강대경 201801158 김 찬 201802033 신 용 201904193 김나연

CONTENTS

01. Topic selection background

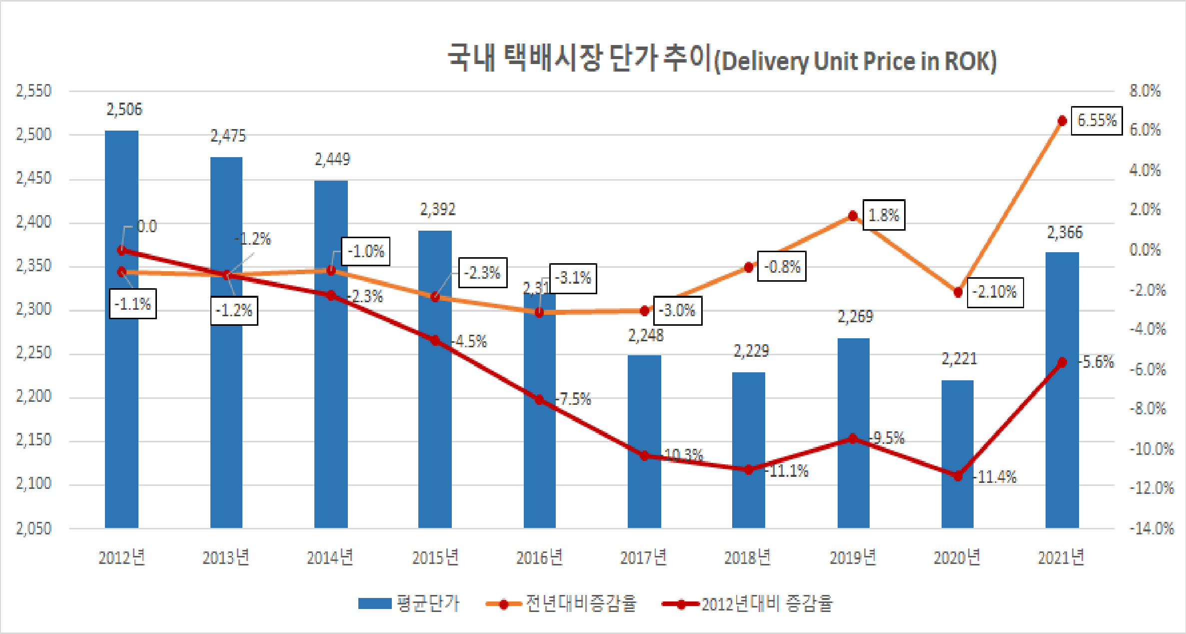
Trend of Delivery service use in ROK & Delivery Sales Trend in ROK

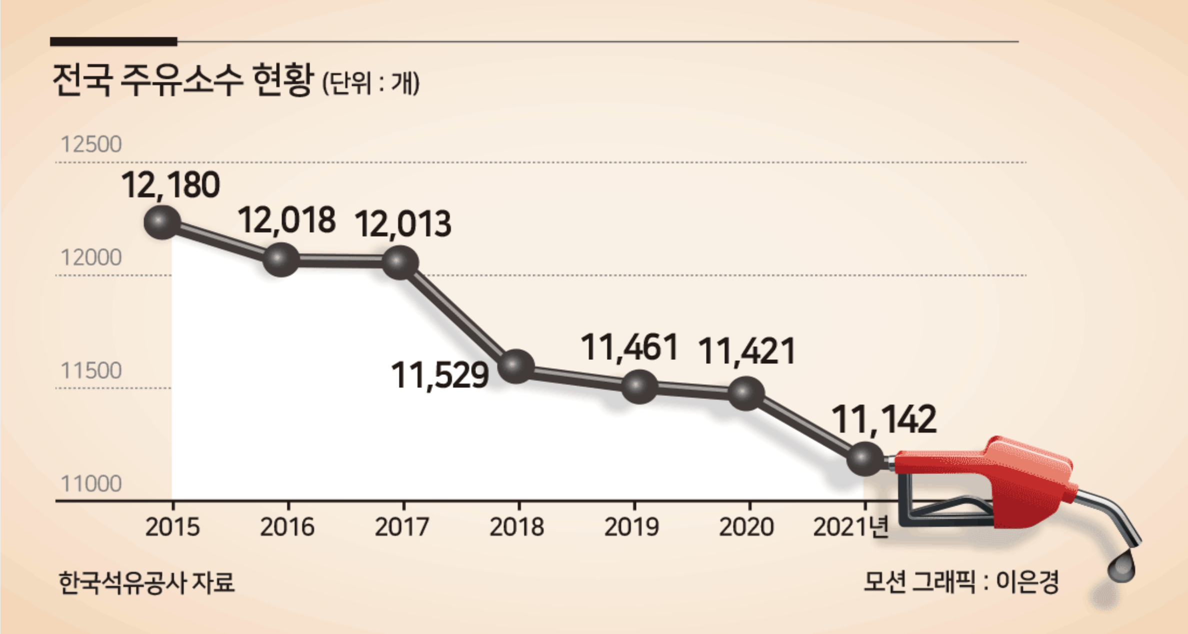




01. Topic selection background

Delivery Unit Price in ROK & Status of the number of gas stations nationwide





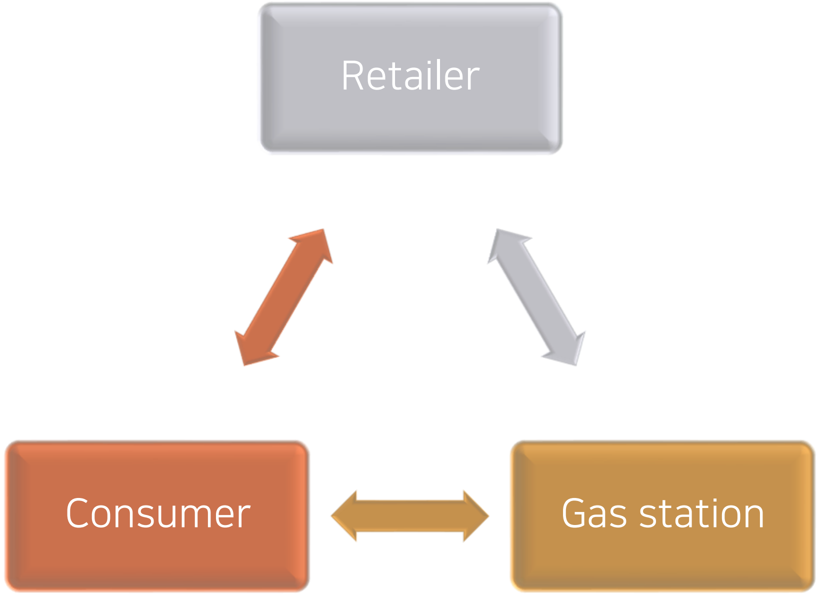
01. Topic selection background

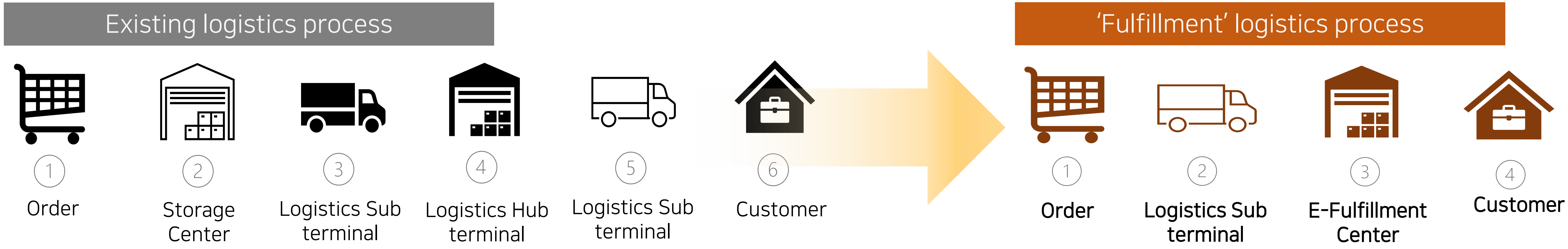
Benefits of MFC Installation

Small Urban distribution center '**Micro Fulfillment Center**'

'**Fulfillment**' refers to a service of all processes of product delivery to customer such as product storage, selection and delivery Existing distribution centers are generally located outside the city

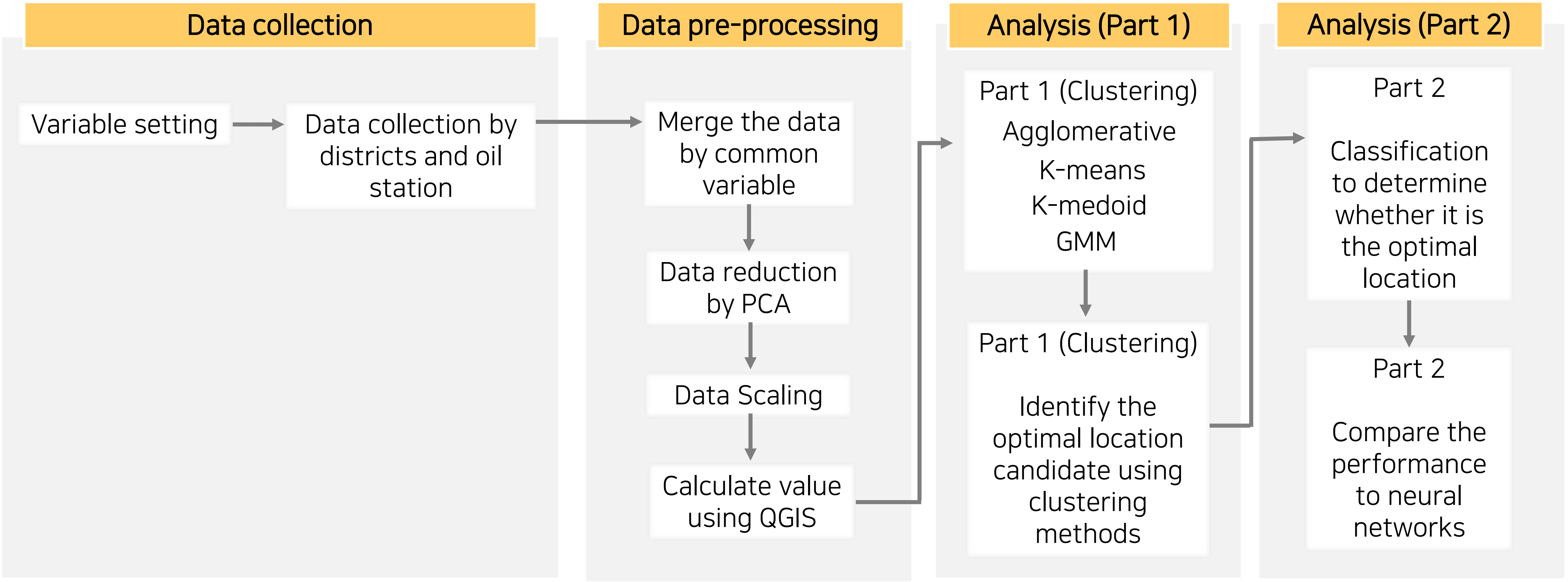
MFCs are located the city and if the product is stored in MFC in advance by predicting consumer preferences
➡ The time between ordering and delivery can be greatly reduced





01. Topic selection background

The Flow of Data Mining approach



Pre-processing

Variables Used for Clustering

Variable	Contents
DRIVEN_EFF	Operational efficiency; 서울특별시 빅데이터 캠퍼스 > CJ택배 운행량 월별 통계.csv
TOTAL_LOGIS	Delivery volume by region; 서울 열린 데이터 광장 > 자치구단위 월별 착지 데이터.csv
POP2040	Number of Producible Populations; 서울 열린 데이터 광장 > 주민등록인구(연령별, 동별).csv
TOTAL_SALES	Off-line commercial sales; 서울 열린 데이터 광장 > 서울시 상권 매출액.csv
NUM_STORE	Size of offline commercial area; 서울 열린 데이터 광장 > 서울시 상권 매출액.csv
ESTATE	Officially assessed reference land price; 서울특별시 빅데이터 캠퍼스 > 국토교통부_표준지공시지가.csv

The following is a preprocessing of collected data for clustering

Preprocessed data

	DISTRICT	DRIVEN_EFF	TOTAL_LOGIS	POP2040	TOTAL_SALES	NUM_STORE	ESTATE
0	동대문구	0.027219	2400374	166805	4.447564e+12	66655	5198000
1	용산구	0.131360	2250955	112398	6.805869e+12	62840	10079000
2	구로구	0.054334	2442377	184583	2.806535e+12	69651	4109000
3	노원구	0.026526	2507172	215172	1.834499e+12	57391	3911000
4	서초구	0.159699	2165093	185381	6.456707e+12	117682	11141000

Generate derivable attribute : Driving Efficiency

Data used to obtain Driving Efficiency

	운행년월 (DRIVEN_YM)	주소(구) (ADDR)	운행거리(평균) (DRIVEN_AVR)	운행거리(총거리) (DRIVEN_SUM)	운행대수 (DRIVEN_CNT)	운행_총시간 (DRIVEN_TIME)	경유지_총건수(VIA_CNT)
0	202011	동대문	10.706256	52579.968458	7122	109180824	402530
1	202006	용산구	12.758485	85058.221296	3671	66316795	375971
2	202007	구로구	12.307220	76255.048379	5280	79544164	299260

$$\text{Driving Efficiency} = \left(\frac{Driven\_Sum}{Driven\_Time}\right) * \left(\frac{VIA\_CNT}{Driven\_CNT}\right)$$

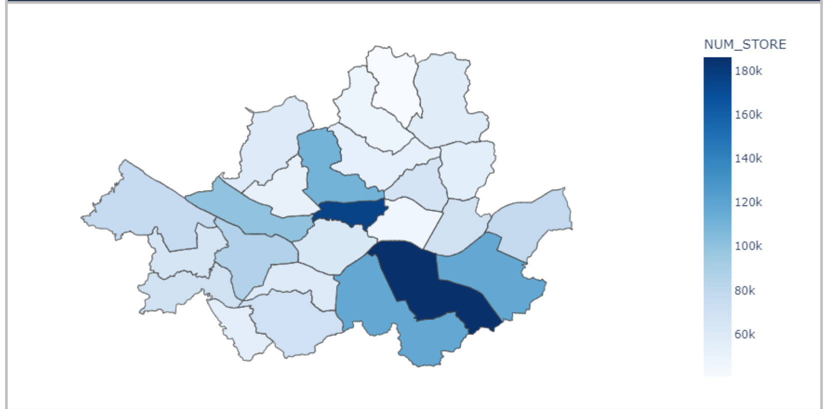
↗	Operation efficiency is proportional to the speed of operation and the number of transit points.
↘	Operational efficiency is inversely proportional to the number of operations.

" Areas with high operational efficiency already have high delivery efficiency "

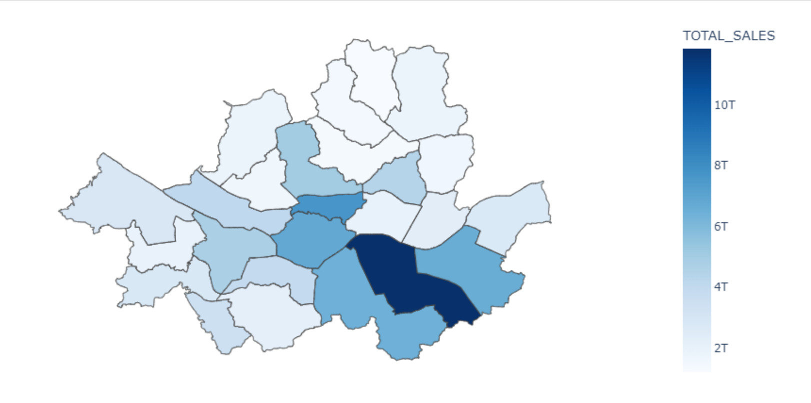
02. Cluster MFC Demand Areas Pre-processing / EDA / Clustering

EDA

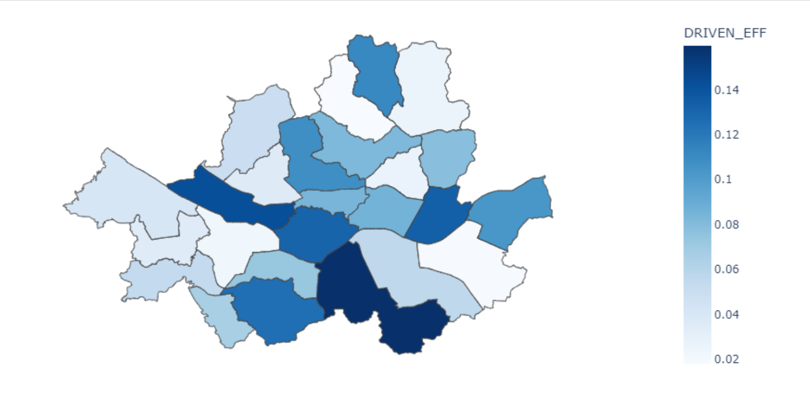
Number of Store In Seoul



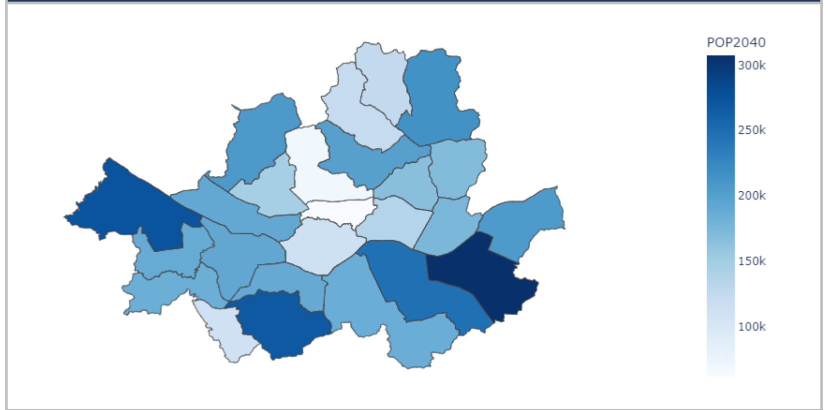
TOTAL_SALES In Seoul



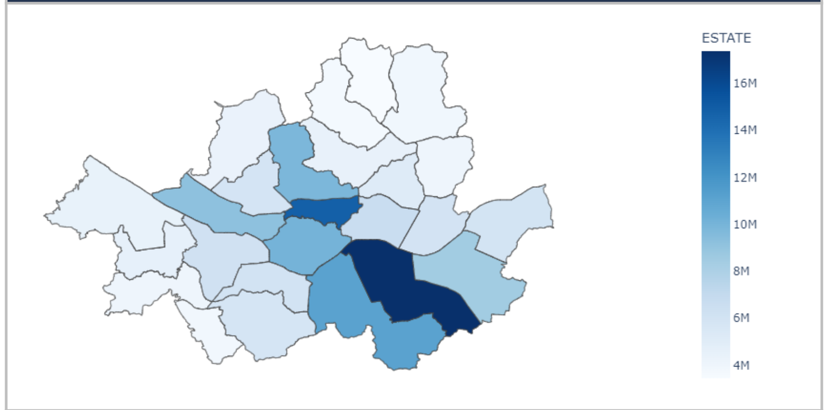
Driven Efficiency In Seoul



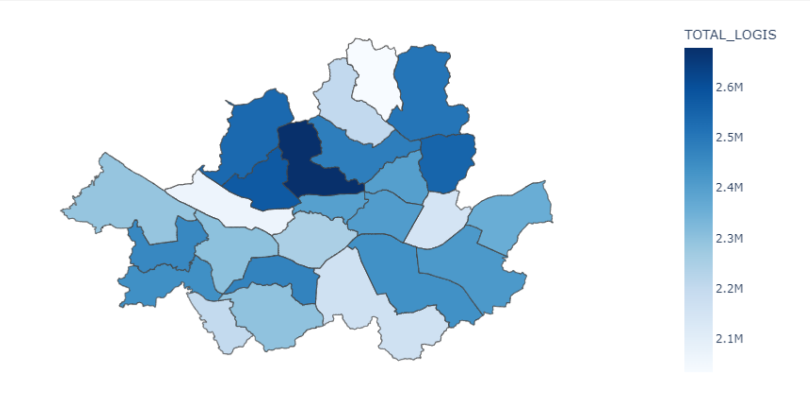
POPULATION(2040) In Seoul



ESTATE PRICE In Seoul



Total Logistics In Seoul



02. Cluster MFC Demand Areas Pre-processing / EDA / Clustering

Standard Scaling

If the data is scaled differently by characteristics, machine learning may not work well, so data scaling ensures that the range of all characteristics is the same

Preprocessed data

	DISTRICT	DRIVEN_EFF	TOTAL_LOGIS	POP2040	TOTAL_SALES	NUM_STORE	ESTATE
0	동대문구	0.027219	2400374	166805	4.447564e+12	66655	5198000
1	용산구	0.131360	2250955	112398	6.805869e+12	62840	10079000
2	구로구	0.054334	2442377	184583	2.806535e+12	69651	4109000
3	노원구	0.026526	2507172	215172	1.834499e+12	57391	3911000
4	서초구	0.159699	2165093	185381	6.456707e+12	117682	11141000

Standard Scaling

	DRIVEN_EFF	TOTAL_LOGIS	POP2040	TOTAL_SALES	NUM_STORE	ESTATE
0	-1.136150	0.227904	-0.197550	0.290781	-0.345691	-0.461560
1	1.346835	-0.703058	-1.124559	1.234272	-0.450148	0.939167
2	-0.489647	0.489605	0.105360	-0.365747	-0.263659	-0.774076
3	-1.152681	0.893314	0.626548	-0.754631	-0.599345	-0.830897

PCA Analysis

PCA Analysis : high-dimensional data → low-dimensional data
There is no need to proceed with PCA because the 6 variables are small enough, but we did. Select the number of PCA with a described variance of 0.7 or greater and a cumulative rate of 80% or greater.

PCA Analysis Results

	Explained variance	Explained variance ratio	Cumulative ratio
pca1	2.960610	0.473698	0.473698
pca2	1.562628	0.250021	0.723718
pca3	1.034192	0.165471	0.889189
pca4	0.526031	0.084165	0.973354
pca5	0.125891	0.020143	0.993496
pca6	0.040648	0.006504	1.000000

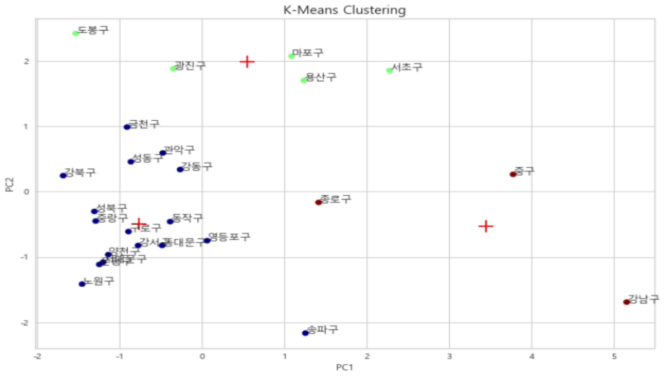
The variance of PCA 3 is greater than 0.7 and the cumulative rate is greater than 80%.

" PCA3 is suitable "

Clustering : Using 4 Techniques

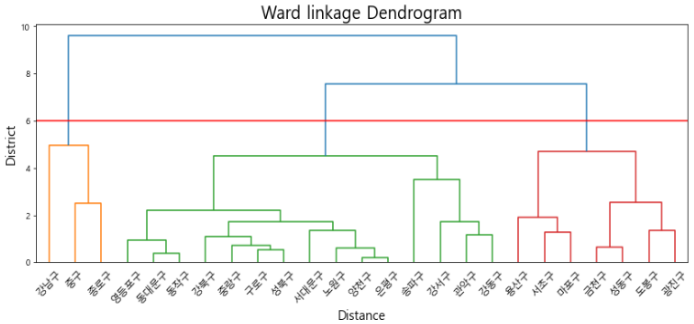
Results of each clustering methods and selected districts are suitable for MFC location

1. K-means



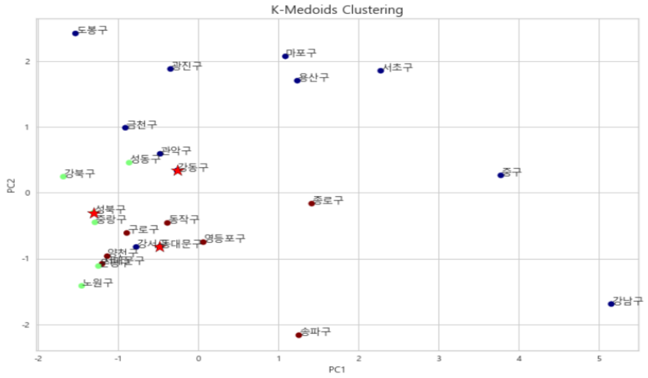
Each cluster is marked as the center of the cluster and must be sensitive to outliers and specify the number of clusters k.

3. Agglomerative



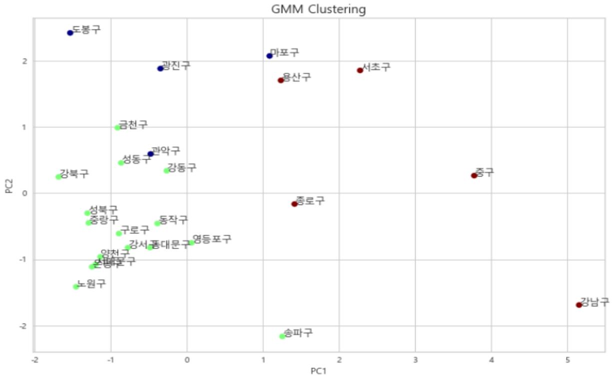
Cluster results can vary depending on distance measurements and do not require setting a value for k.

2. K-medoids



Using midpoints instead of means is less sensitive than k-means clustering, but it takes a long time because all distances are calculated repeatedly.

4. GMM



Even if the variance is not constant, clustering can be well generated, but it has computational complexity and good results cannot be expected if it does not fit the Gaussian distribution assumption.

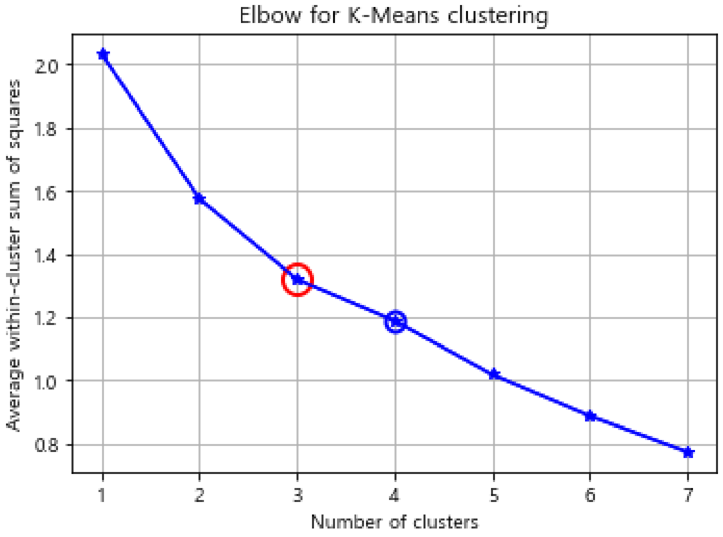
K-means

Step 1

Choose the number of cluster k

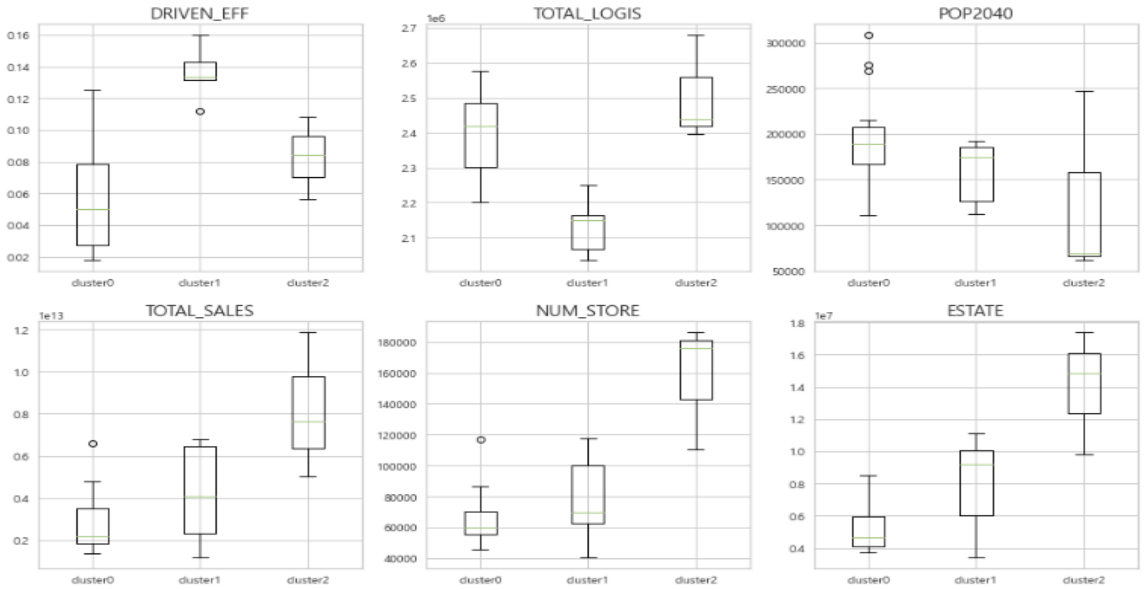
Elbow method

: Use the turning point in the curve of sum of within cluster variance



Step 2

Interpret each cluster and determine suitable or not



Cluster 0

- The logistics and population are high
- The population is large, but the size of the commercial is small, so we can expect untact consumption
- The lowest land estate and driven efficiency

→ Cluster 0 is suitable

Cluster 1

- The lowest logistics
- Since the driven efficiency is the highest, the logistics service rate is already high
- The commercial and consumption are adequate

→ Cluster 1 is not suitable

Cluster 2

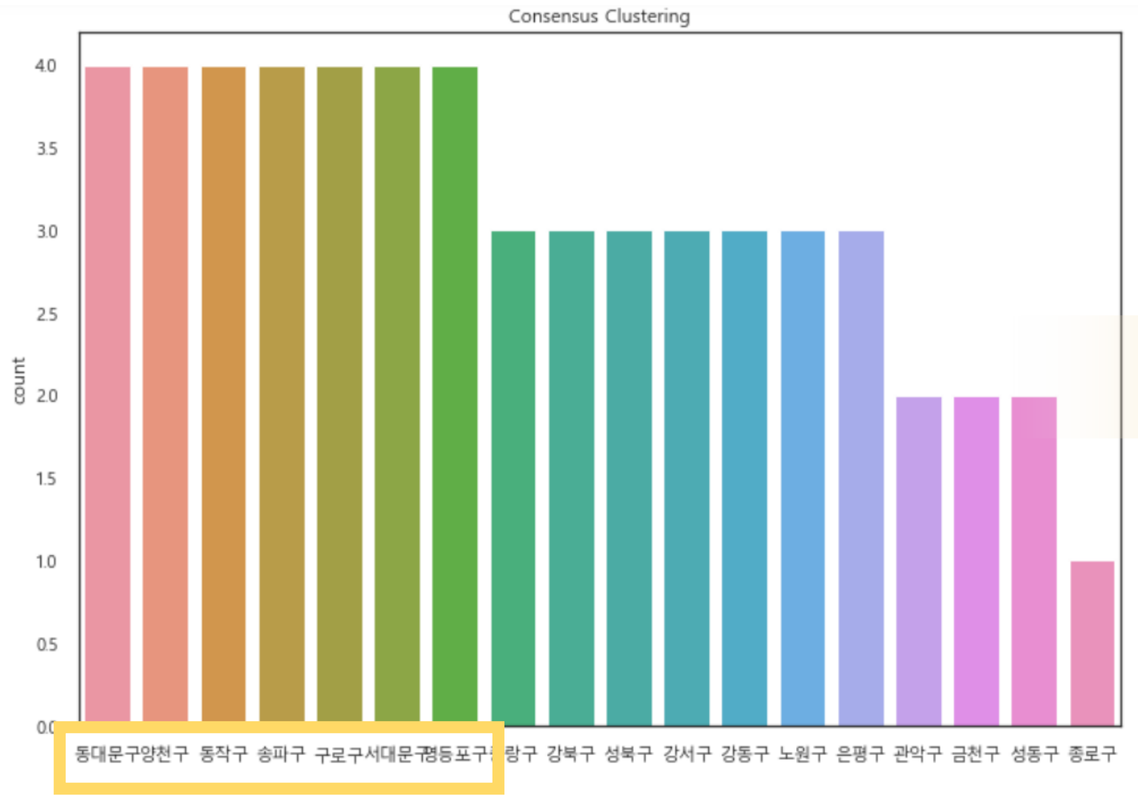
- The highest total logistics
- The highest commercial and consumption power relative to the population
- But the highest land estate

→ Cluster 2 is not suitable

Consensus

By combining all four method results , MFC is most needed has been selected

The areas where the final MFC is most needed : 동대문구, 양천구, 동작구, 송파구, 구로구, 서대문구, 영등포구



03. Classification of suitable gas station for MFC
Pre-processing / Classification / Clustering

Pre-processing

Variables Used for Classification

Variable	Contents
GAS_STATION	Naming of Gas Station 공공 데이터 포털 > 서울시 주유소 현황.csv
AREA	Area of Gas Station; Derived attributes(from QGIS)
ESTATE	Officially assessed individual land price; 서울 열린 데이터 광장 > 서울시 개별공시지가 정보 > 공시지가_2022년.csv
DIST_####DC	Distance from East/West/KoreaDC; Derived attributes(using Haversine)
POP1000	Population within 1km radius of Gas station; Derived attributes(from SGIS)
SCHOOL	The number of school within 1km radius of Gas station;
APART	The number of apartment within 1km radius of Gas station;
MARKET	The number of market within 1km radius of Gas station;

Preprocessed data

	GAS_STATION	DISTRICT	DONG	LATITUDE	LONGITUDE	AREA	ESTATE	DIST_EASTDC	DIST_WESTDC	DIST_KOREADC	POP1000	SCHOOL	APART	MARKET
0	현대오일뱅크(주) 직영소월길주유소	용산구	후암동	37.554409	126.977735	245.0	8280000.0	8.970251	13.202286	11.678968	15660	5	1	58
1	선익상사(주) 동자동주유소	용산구	동자동	37.550201	126.972418	711.0	18850000.0	9.568763	12.593357	11.494239	25467	9	2	6
2	현대오일뱅크㈜ 직영갈월동주유소	용산구	갈월동	37.547029	126.972228	700.0	15050000.0	9.713120	12.454393	11.199948	26924	9	1	22

Independent variables : Estate

We will make dataset grouping by 'Gas station' in the final.

	연번	자치구명	주유소명	주소
0	1	용산구	현대오일뱅크(주) 직영소월길주유소	서울특별시 용산구 소월로66
1	2	용산구	선익상사(주) 동자동주유소	서울특별시 용산구 한강대로 104길 6
2	3	용산구	현대오일뱅크㈜ 직영갈월동주유소	서울특별시 용산구 한강대로 322

Extract information District, Dong, Latitude and Longitude by Gas Station through *Kakao API*.

```
locations = []
for addr in addrs:
    url = 'https://dapi.kakao.com/v2/local/search/address.json?query={}'.format(addr)
    headers = {"Authorization": "KakaoAK [redacted]}
    place = requests.get(url, headers = headers).json()['documents']
    locations.append(place)
```

And then, merge the data 'Officially assessed individual land price'.

	GAS_STATION	LATITUDE	LONGITUDE	DISTRICT	DONG	ESTATE	ADDRESS
0	현대오일뱅크(주) 직영소월길주유소	37.554409	126.977735	용산구	후암동	8280000.0	448-103
1	선익상사(주) 동자동주유소	37.550201	126.972418	용산구	동자동	18850000.0	14-125
2	현대오일뱅크㈜ 직영갈월동주유소	37.547029	126.972228	용산구	갈월동	15050000.0	11-34
3	서계주유소	37.552366	126.968994	용산구	서계동	10330000.0	47-15
4	㈜영원에너지 풍기주유소	37.535589	126.962709	용산구	원효로2가	13300000.0	70-2

The Officially assessed individual land price at the gas stations was selected as an independent variable because it was related to rent incurred when entering MFC.

03. Classification of suitable gas station for MFC Pre-processing / Classification / Clustering

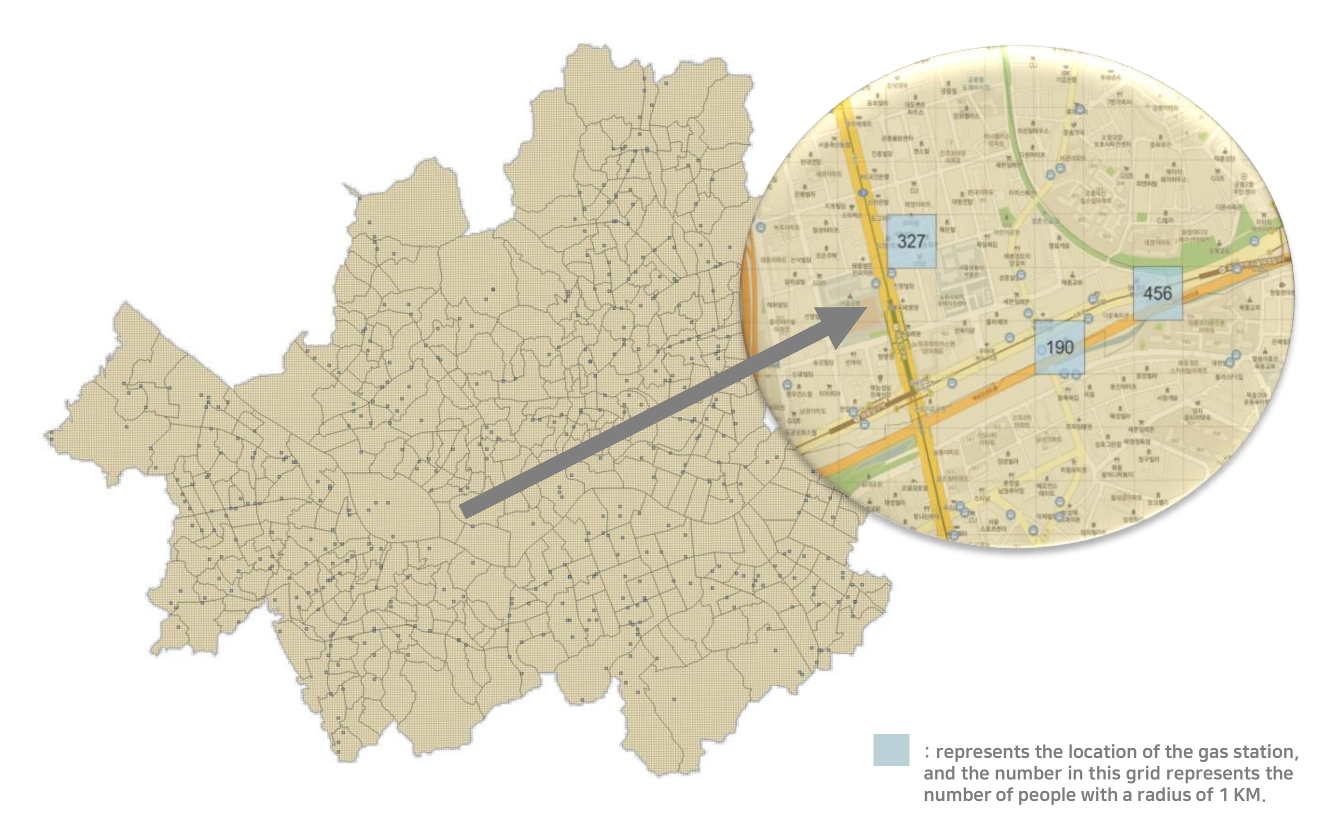
Independent variables : Population

```python
from sklearn.neighbors import KNeighborsClassifier

oilbank = pd.read_excel('서울시주유소_위경도.xlsx')
oilbank.rename(columns = {'LATITUDE': '위도', 'LONGITUDE':'경도'}, inplace = True)
x_train = df[['위도', '경도']]
y_train = df['wkt_geom']

neigh = KNeighborsClassifier(n_neighbors = 1)
neigh.fit(x_train, y_train)

x_test = oilbank.loc[:, ['위도', '경도']]
labels = neigh.predict(x_test)
oilbank['wkt_geom'] = labels
import haversine

df_LEFT_JOIN['points'] = df_LEFT_JOIN[['위도', '경도']].values.tolist()
oilbank['points'] = oilbank[['위도', '경도']].values.tolist()
m = 1000
for i in range(len(oilbank)):
    print(i)
    oil_point = oilbank.loc[i, 'points']
    df_LEFT_JOIN['m'] = df_LEFT_JOIN['points'].apply(lambda x: haversine.haversine(oil_point, x, unit = 'm'))
    population = df_LEFT_JOIN[df_LEFT_JOIN['m'] <= m+1]['인구'].sum()
    oilbank.loc[i, 'population'] = population
result = oilbank.loc[:, ['GAS_STATION', '위도', '경도', 'DISTRICT', 'DONG', 'wkt_geom', 'population']]
result.to_excel('주유소 100m 반경 안 인구수 최종 파일.xlsx',index=False)
result
```

	GAS_STATION	위도	경도	DISTRICT	DONG	wkt_geom	population
0	현대오일뱅크(주) 직영소월길주유소	37.554409	126.977735	용산구	후암동	MultiPolygon (((126.97694474052541125 37.55358...	166.0
1	선익상사(주) 동자동주유소	37.550201	126.972418	용산구	동자동	MultiPolygon (((126.97130965709310146 37.54995...	349.0
2	현대오일뱅크㈜ 직영갈월동주유소	37.547029	126.972228	용산구	갈월동	MultiPolygon (((126.97133512071627592 37.54635...	251.0
3	서계주유소	37.552366	126.968994	용산구	서계동	MultiPolygon (((126.96790067781765288 37.55174...	102.0
4	㈜영원에너지 풍기주유소	37.535589	126.962709	용산구	원효로2가	MultiPolygon (((126.96235679690941822 37.53549...	186.0
...	...	...	...	...	...	...	...
467	현대오일뱅크㈜직영 도봉현대셀프주유소	37.688374	127.045327	도봉구	도봉동	MultiPolygon (((127.04518659989290086 37.68816...	16.0
468	GS칼텍스㈜ 도봉주유소	37.684369	127.045522	도봉구	도봉동	MultiPolygon (((127.04521411792508445 37.68366...	180.0
469	(주)송만에너지 도봉제일주유소	37.674474	127.044067	도봉구	도봉동	MultiPolygon (((127.04300674510970737 37.67373...	465.0
470	노원교주유소	37.679015	127.049751	도봉구	도봉동	MultiPolygon (((127.04864917541478064 37.67826...	55.0
471	오복주유소	37.662280	127.047441	도봉구	방학동	MultiPolygon (((127.0464798867402294 37.662036...	247.0

Data on the total population (personnel) were obtained from the SGIS statistical geographic information service, and the number of factors within a 1km radius of the gas station was extracted. This extracted data is then converted into a SHP file and visualized in QGIS.

Visualize 1KM Population Count with QGIS



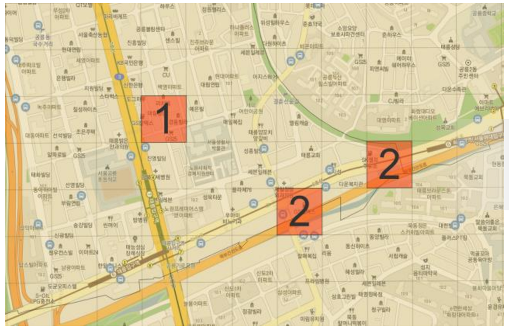
Independent variables : School & Apart & Market

The Ministry of Land, Infrastructure and Transport obtained data on the number of schools, apartments, and markets within the range of Seoul, and obtained numbers within a 1KM radius based on each gas station with the data.

```
import haversine
m = 1000
oilbank_df['point'] = oilbank[['LATITUDE', 'LONGITUDE']].values.tolist()
lst = ['school', 'apart', 'market']
for stat in lst:
    globals()[f'{stat}_df']['point'] = globals()[f'{stat}'][['Latitude', 'Longitude']].values.tolist()
    for i in range(len(oilbank)):
        oil_point = oilbank_df.loc[i, 'point']

        globals()[f'{stat}_df']['stat_per_m'] = globals()[f'{stat}_df']['point'].apply(lambda x: haversine.haversine(x, oil_point, unit = 'm'))
        oilbank_df.loc[i, stat] = len(globals()[f'{stat}_df'][globals()[f'{stat}_df']['stat_per_m'] <= m+1])
```

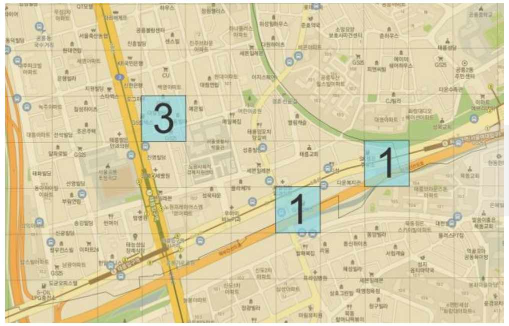
	GAS_STATION	LATITUDE	LONGITUDE	point	school	apart	market
0	현대오일뱅크(주) 직영소월길주유소	37.554409	126.977735	[37.5544085670544, 126.977734582566]	5.0	1.0	58.0
1	선익상사(주) 동자동주유소	37.550201	126.972418	[37.5502005044121, 126.972417738531]	9.0	2.0	6.0
2	현대오일뱅크㈜ 직영갈월동주유소	37.547029	126.972228	[37.5470289447515, 126.972228457829]	9.0	1.0	22.0
3	서계주유소	37.552366	126.968994	[37.5523662854224, 126.968993700509]	5.0	2.0	4.0
4	㈜영원에너지 풍기주유소	37.535589	126.962709	[37.5355890312127, 126.962708973419]	9.0	6.0	20.0
...	...	...	...	...	...	...	...
467	현대오일뱅크㈜직영 도봉현대셀프주유소	37.688374	127.045327	[37.6883740291887, 127.045327229383]	4.0	1.0	0.0
468	GS칼텍스㈜ 도봉주유소	37.684369	127.045522	[37.6843693792253, 127.045522281038]	4.0	1.0	0.0
469	(주)송만에너지 도봉제일주유소	37.674474	127.044067	[37.6744735831616, 127.044066665278]	0.0	0.0	4.0
470	노원교주유소	37.679015	127.049751	[37.679014779402, 127.049750654086]	3.0	1.0	2.0
471	오복주유소	37.662280	127.047441	[37.6622801187124, 127.047441496273]	4.0	3.0	4.0

Visualize 1KM School & Apart & Market Count with QGIS



One **red grid** represents the location of one gas station, and the number in this grid represents the number of schools within a radius of 1 KM.

One **orange grid** represents the location of one gas station, and the number in this grid represents the number of Aparts within a radius of 1 KM.



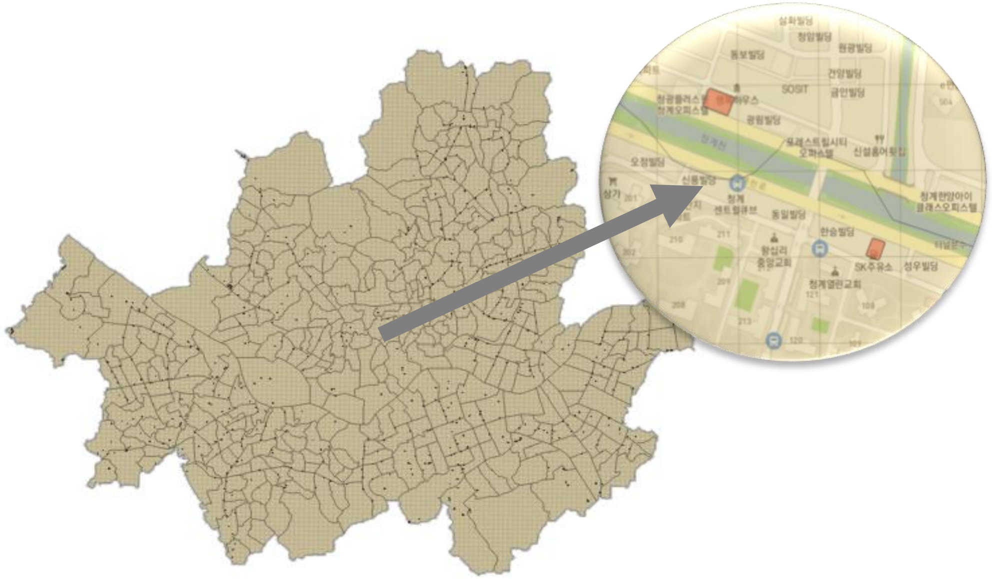
One **blue grid** represents the location of one gas station, and the number in this grid represents the number of markets within a radius of 1 KM.

03. Classification of suitable gas station for MFC

Pre-processing / Classification / Clustering

Independent variables : Area & Land price

Use the '$AREA' function of the QGIS program to find the area of each gas station in the polygon form of the gas station shp file.



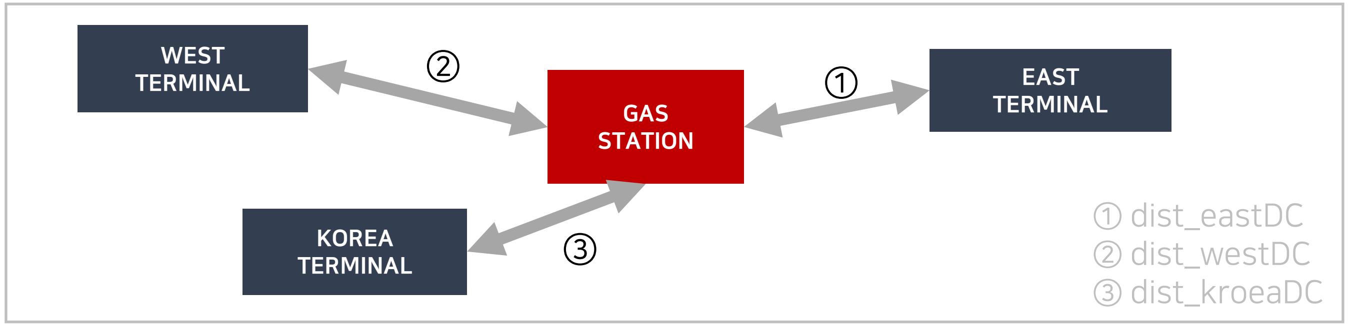
SHP files at the gas station

NAME	KIND	주요소 넓이
SK에너지소월길주유소	GSK001	245.0
SK에너지동자동주유소	GSK001	711.0
SK네트웍스직영갈월동주유소	GSK001	700.0
GS칼텍스서계주유소	GSK001	1010.0
GS칼텍스풍기주유소	GSK001	519.0

Independent variables : Coverable distance

The distance to each terminal was calculated based on each gas station.

Variable	Definition
dist_eastDC	Distance from the east terminal
dist_westDC	Distance from west terminal
dist_koreaDC	Distance to Korea Terminal




```python
from haversine import haversine
from tqdm import notebook

distance_compare_list = []
for i in notebook.tqdm(range(df_DC.shape[0])):
    for j in range(df_gas.shape[0]):
        DC = (df_DC.Latitude[i], df_DC.Longitude[i])
        oil = (df_gas.LATITUDE[j], df_gas.LONGITUDE[j])
        result = haversine(DC, oil, unit='km')
        distance_compare_list.append(result)
```

GAS_STATION	DIST_EASTDC	DIST_WESTDC	DIST_KOREADC
현대오일뱅크(주) 직영소월길주유소	8.970251	13.202286	11.678968
선익상사(주) 동자동주유소	9.568763	12.593357	11.494239
현대오일뱅크㈜ 직영갈월동주유소	9.713120	12.454393	11.199948

Set Dependent variable : Using AHP method

Using the AHP(Analytic Hierarchy Process) to calculate the weights of the independent variables of the location score function

척도	정의
1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong importance
9	Absolute importance
2,4,6,8	Intermediate values of the above values

Pairwise Comparison Matrix

	AREA	ESTATE	DISTANCE	POPULATION	SCHOOL	APARTMENT	MARKET
AREA	1	1/2	3	1/3	9	7	5
ESTATE	2	1	3	3	9	7	6
DISTANCE	1/3	1/3	1	1/3	9	5	3
POPULATION	3	1/3	3	1	9	6	5
SCHOOL	1/9	1/9	1/9	1/9	1	1/3	1/4
APARTMENT	1/7	1/7	1/5	1/6	3	1	1
MARKET	1/5	1/6	1/3	1/5	4	1	1

AREA	0.198
ESTATE	0.327
DISTANCE	0.124
POPULATION	0.244
SCHOOL	0.020
APARTMENT	0.040
MARKET	0.048

$$SCORE = (AREA * 0.198) - (ESTATE * 0.327) - (DIST\_EASTDC * 0.0413) - (DIST\_WESTDC * 0.0413) - (DIST\_KOREADC * 0.0413) + (POP1000 * 0.244) + (SCHOOL * 0.020) + (APART * 0.040) + MARKET * 0.048)$$

" After minmax scaling and calculate location score "

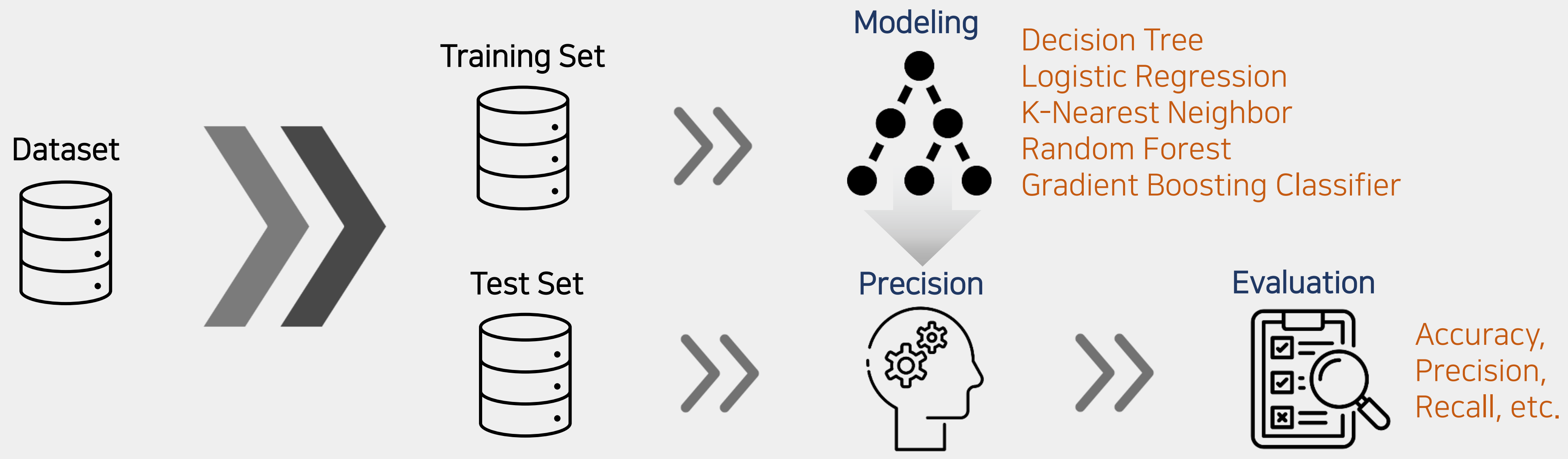
	AREA	ESTATE	DIST_EASTDC	DIST_WESTDC	DIST_KOREADC	POP1000	SCHOOL	APART	MARKET	SCORE
0	0.043427	0.189831	0.318940	0.438800	0.443422	0.253190	0.357143	0.018519	0.983051	0.013764
1	0.169609	0.432974	0.341411	0.418207	0.436269	0.411800	0.642857	0.037037	0.101695	-0.037691
2	0.166631	0.345562	0.346831	0.413507	0.424874	0.435364	0.642857	0.018519	0.372881	0.008770
3	0.250571	0.236987	0.349078	0.411788	0.450253	0.489771	0.357143	0.037037	0.067797	0.053482
4	0.117620	0.305306	0.395409	0.373522	0.403165	0.442966	0.642857	0.111111	0.338983	0.016702

Multi-class Classification

Using qcut(), Generate categorical variable "SCORE_CAT"

	AREA	ESTATE	DIST_EASTDC	DIST_WESTDC	DIST_KOREADC	POP1000	SCHOOL	APART	MARKET	SCORE_CAT
0	245.0	8280000.0	8.970251	13.202286	11.678968	15660	5	1	58	부적합
1	711.0	18850000.0	9.568763	12.593357	11.494239	25467	9	2	6	매우부적합
2	700.0	15050000.0	9.713120	12.454393	11.199948	26924	9	1	22	부적합
3	1010.0	10330000.0	9.772973	12.403577	11.855362	30288	5	2	4	적합
4	519.0	13300000.0	11.006969	11.272070	10.639303	27394	9	6	20	부적합

SCORE	Category
(-0.296, -0.00108]	매우 부적합 (Most Unsuitable)
(-0.00108, 0.0432]	부적합 (Unsuitable)
(0.0432, 0.0825]	적합 (Suitable)
(0.0825, 0.184]	매우 적합 (Most Suitable)



03. Classification of suitable gas station for MFC

Multi-class Classification

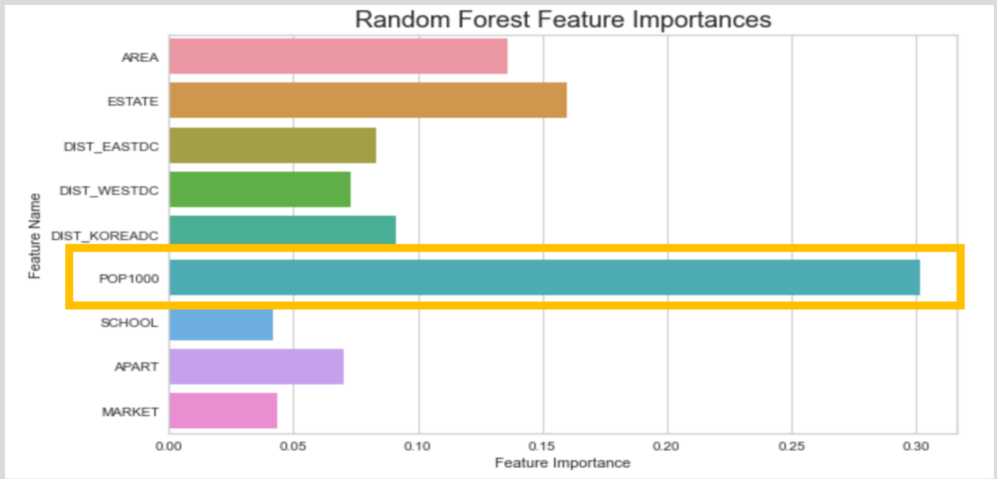
Model Evaluation

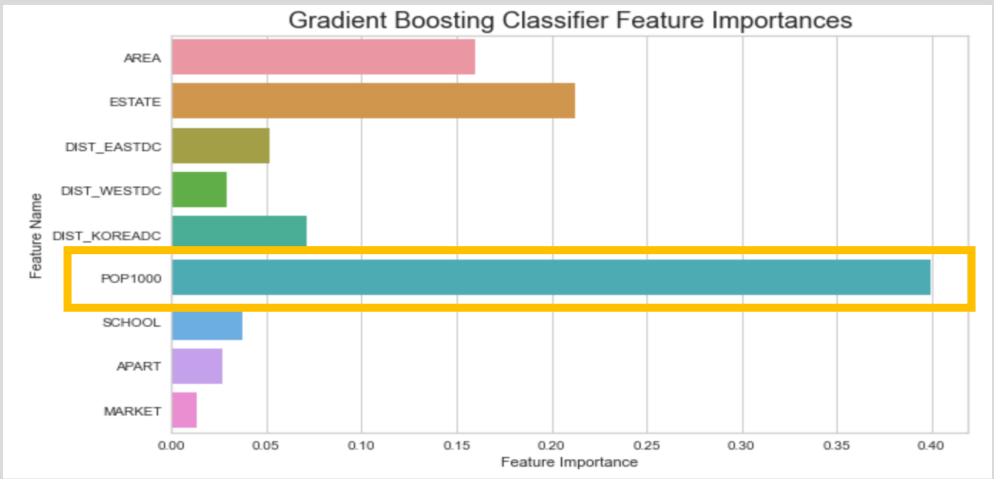
Model	Accuracy
DT	0.7518
LR	0.6667
KNN	0.6667
RF	0.7589
GBC	0.766

Model Evaluation

Model		Precision	Recall	F-1 Score	Accuracy
Random Forest	Most Unsuitable	0.86	0.82	0.84	0.7589
	Unsuitable	0.61	0.63	0.62	
	Suitable	0.69	0.65	0.67	
	Most Suitable	0.82	0.89	0.85	
Gradient Boosting Classifier	Most Unsuitable	0.88	0.84	0.86	0.766
	Unsuitable	0.59	0.63	0.61	
	Suitable	0.67	0.71	0.69	
	Most Suitable	0.88	0.83	0.85	

"Population" is the most important feature in RF & GBC.





03. Classification of suitable gas station for MFC

Pre-processing / Classification / Clustering

Multi-Class-Neural Network : 4 Steps

The method is a technique for constructing a model needed to recognize unique patterns or structures in the data and proceeds to four steps.

1. Label Encoding for Multi-classification ANN

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
le.fit(df['SCORE_CAT'])
label_encoded = le.transform(df['SCORE_CAT'])

label_df = pd.DataFrame(label_encoded,
                        columns = ['Score_label'])

label_df.head()
```

	Score_label
0	2
1	0
2	2
3	3
4	2

- 매우부적합(Most unsuitable) : 0
- 부적합(Unsuitable) : 2
- 적합(Suitable) : 3
- 매우적합(most Suitable) : 1

2. Data processed

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=1113)
```

Train-test data split

```
scaler = MinMaxScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

Data scaling

```
train_scaled, val_scaled, train_target, val_target = train_test_split(X_train_scaled, y_train, test_size=0.3, random_state=1113)
```

Validation

2. Define the Neural Network Model

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 dense (Dense)               (None, 64)                640

 dense_1 (Dense)             (None, 32)                2080

 dense_2 (Dense)             (None, 8)                 264

 dense_3 (Dense)             (None, 4)                 36

=================================================================
Total params: 3,020
Trainable params: 3,020
```

- 9 Indep. Variables -> input shape : 9
- 4 Dep. Variables -> Last Layer: 4
- Multi class -> Activation f(x) : Softmax
- Most popular optimizer -> 'Adam'
- Loss & Metrics -> 'Sparse_categorical'

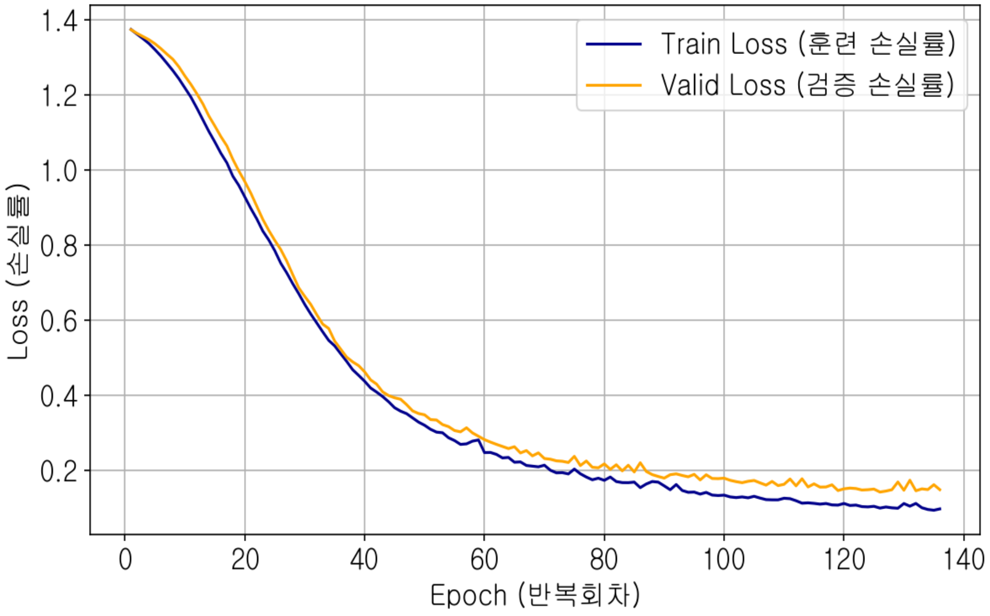
4. Fit model

```
Epoch 144/1000
8/8 [==============================] - 0s 22ms/step - loss: 0.0954 - sparse_categorical_accuracy: 0.9869 - val_loss: 0.1826 - val_sparse_categorical_accuracy: 0.9091 - lr: 0.001

Epoch 145/1000
1/8 [==>...........................] - ETA: 0s - loss: 0.0882 - sparse_categorical_accuracy: 1.0000
Epoch 145: ReduceLROnPlateau reducing learning rate to 0.00010000000474974513.
8/8 [==============================] - 0s 26ms/step - loss: 0.0991 - sparse_categorical_accuracy: 0.9694 - val_loss: 0.1741 - val_sparse_categorical_accuracy: 0.9293 - lr: 0.001
0
Epoch 145: early stopping
```

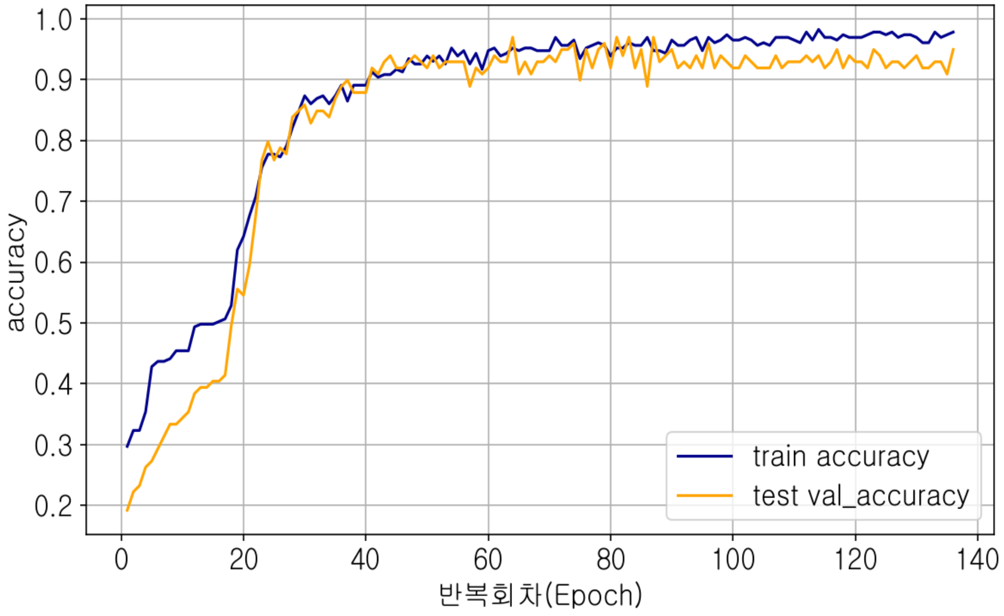
Multi-Class-Neural Network : Evaluation

After completing the 4 steps of the 'Multi-Class-Natural Network',
the training and verification loss rate and training and verification accuracy were evaluated.

Train & Valid Loss (훈련 및 검증 손실률)



Train & Valid Accuracy (훈련 및 검증 정확도)



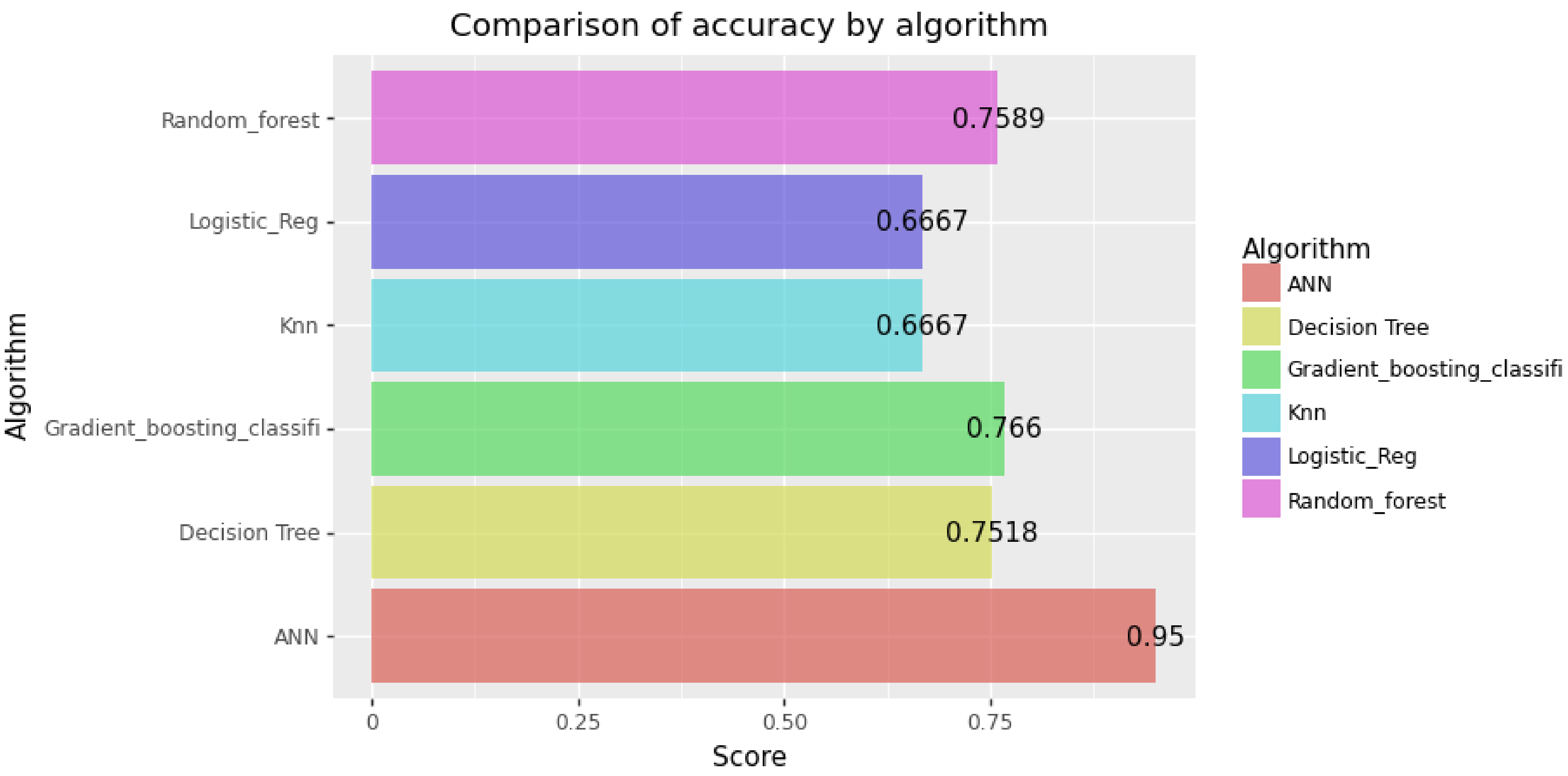
Result

FINAL LOSS : 0.139630 FINAL ACCURACY : 0.957447

As a result of checking the loss rate and accuracy,
it can be confirmed that a high performance of 95% was achieved without overfitting.

Result

The following is a graph comparing the accuracy of the algorithm.



03. Classification Result based on Clustering

The intersection of the most suitable gas station for MFC and Clustering result

1. Predict test-data using Multi-class Neural Network

```
# 예측하기
prediction = neural_model.predict(X_test_scaled) # test data의 예측치
prediction # 예측값이 어레이로 출력됨 따라서 argmax로 최대값 인덱스(위치) 출력

5/5 [==============================] - 0s 5ms/step

array([[9.99999642e-01, 1.39732729e-20, 3.11178837e-07, 1.62431334e-13],
       [1.03021562e-02, 1.08269633e-04, 6.04344785e-01, 3.85244727e-01],
       [9.99689937e-01, 2.16962940e-18, 3.10084433e-04, 6.50333606e-11],
       [1.00000000e+00, 7.40065364e-29, 9.89212712e-10, 2.02480385e-19],
       [9.53878555e-03, 8.00618623e-10, 9.90335226e-01, 1.25935636e-04],
       [7.45588616e-02, 6.21005893e-01, 7.67422989e-02, 2.27692917e-01],
       [7.45588616e-02, 6.21005893e-01, 7.67422989e-02, 2.27692917e-01],
       [7.13217829e-04, 2.73144583e-08, 9.97970760e-01, 1.31596962e-03],
       [1.71531364e-03, 2.99649400e-06, 5.42941034e-01, 4.55340683e-01],
       [9.99520779e-01, 4.20634190e-15, 4.79192851e-04, 2.04642348e-09],
       [9.90311027e-01, 1.34857539e-14, 9.68890637e-03, 4.47825634e-08],

print('Size of Test data : %f, Size of predicted data : %f' % (len(X_test_scaled), len(prediction)))

Size of Test data : 141.000000, Size of predicted data : 141.000000
```

2. Predicted table for test-data

	GAS_STATION	DISTRICT	DONG	SCORE_CAT	Actual_Location_Score	Predicted_value_by_ANN
1	선익상사(주) 동자동주유소	용산구	동자동	매우부적합	0	0
6	현대오일뱅크㈜ 직영강변주유소	용산구	청암동	부적합	2	3
8	(주)중앙에너비스 한남지점	용산구	한남동	매우부적합	0	0
9	한남제3한강주유소 주식회사	용산구	한남동	매우부적합	0	0
11	현대오일뱅크㈜ 직영한남동주유소	용산구	한남동	매우부적합	0	0
...	...	...	...	...	...	...
456	㈜자연에너지 햇살주유소	도봉구	방학동	매우적합	1	1
461	동일석유㈜ 창동주유소	도봉구	창동	매우적합	1	1
462	극동유화㈜ 대안주유소	도봉구	창동	적합	3	3
463	한이에너지㈜ KLP제1주유소	도봉구	창동	적합	3	3
464	현대오일뱅크㈜직영 도봉현대셀프주유소	도봉구	도봉동	매우부적합	0	0

141 rows × 6 columns

2. The most suitable district for MFC ∩ The most suitable gas station for MFC

```
filterling = (
    (df_final['DISTRICT'].isin(['동대문구', '양천구', '동작구', '송파구', '구로구', '서대문구', '영등포구']))&
    (df_final['Actual_Location_Score'] == 1)&
    (df_final['Predicted_value_by_ANN'] == 1)
)

clustering_linkage_ANN = df_final.loc[filterling]
clustering_linkage_ANN
```

	GAS_STATION	DISTRICT	DONG	SCORE_CAT	Actual_Location_Score	Predicted_value_by_ANN	SCORE
49	삼미상사㈜ 장안킹셀프주유소	동대문구	장안동	매우적합	1	1	0.119815
52	삼영주유소	동대문구	장안동	매우적합	1	1	0.108965
60	대성산업㈜청량리주유소	동대문구	청량리동	매우적합	1	1	0.091912
147	에이치지 가로공원주유소	양천구	신월동	매우적합	1	1	0.113783
149	동일석유㈜ 개나리주유소	양천구	신월동	매우적합	1	1	0.143189
165	양천구주유소	양천구	목동	매우적합	1	1	0.118080
235	SK에너지㈜양평주유소	영등포구	양평동3가	매우적합	1	1	0.160287
236	SK에너지㈜ 기린주유소	영등포구	양평동4가	매우적합	1	1	0.145911
243	㈜정수에너지개발	영등포구	신길동	매우적합	1	1	0.106042
247	㈜엠에스주유소	영등포구	대림동	매우적합	1	1	0.088108
249	(주)대청에너지	영등포구	대림동	매우적합	1	1	0.115939
251	현대오일뱅크㈜직영신대방셀프주유소	동작구	신대방동	매우적합	1	1	0.146005

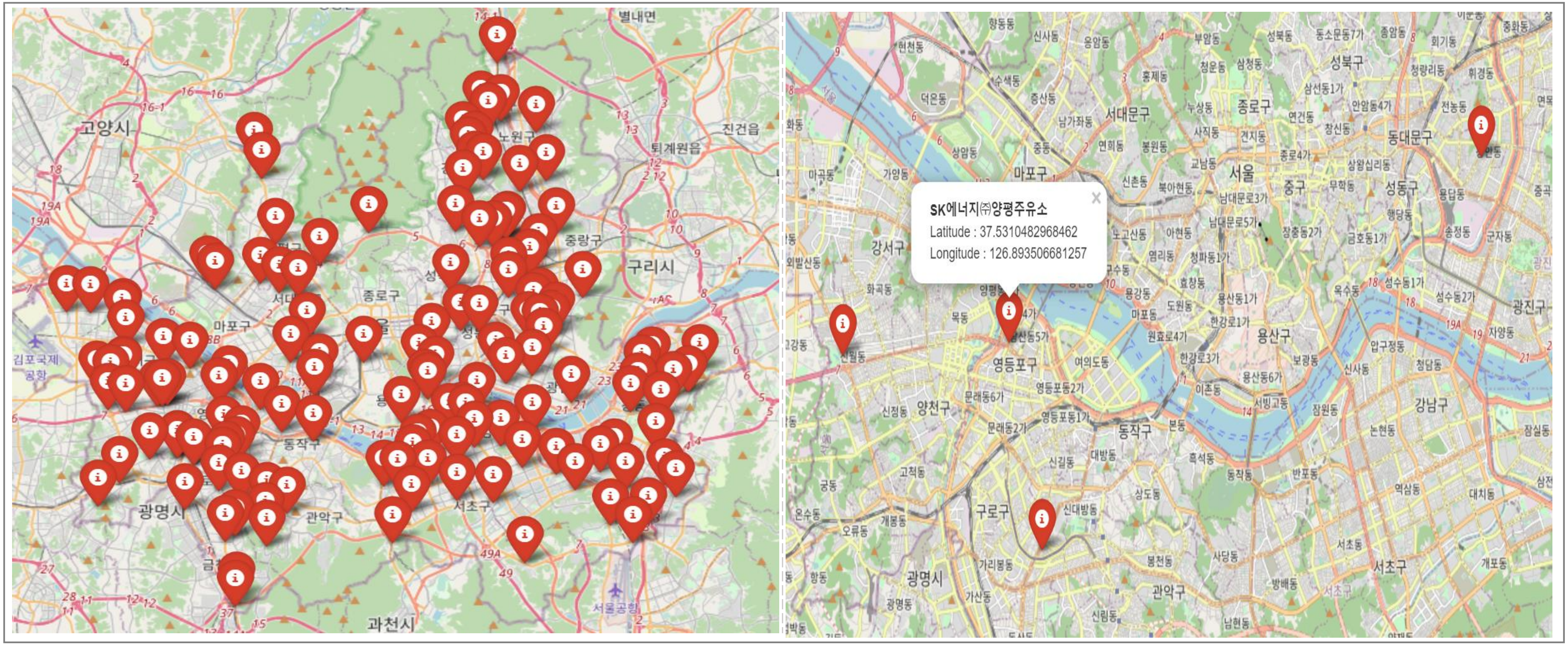
4. The gas station with best location score by district

```
df_tmp2 = clustering_linkage_ANN.groupby('DISTRICT')['SCORE'].max().reset_index()
df_final_selection = pd.merge(left =df_tmp2, right=clustering_linkage_ANN,how='left', left_on=['DISTRICT','SCORE'],right_on=['DISTRICT','SCORE'])
df_final_selection
```

	DISTRICT	SCORE	GAS_STATION	DONG	SCORE_CAT	Actual_Location_Score	Predicted_value_by_ANN
0	동대문구	0.119815	삼미상사㈜ 장안킹셀프주유소	장안동	매우적합	1	1
1	동작구	0.146005	현대오일뱅크㈜직영신대방셀프주유소	신대방동	매우적합	1	1
2	양천구	0.143189	동일석유㈜ 개나리주유소	신월동	매우적합	1	1
3	영등포구	0.160287	SK에너지㈜양평주유소	양평동3가	매우적합	1	1

03. Classification Result based on Clustering

Among the "141" gas station in Seoul, only "4" gas stations that are most suitable for Micro fulfillment center



04. Association Rule

Association Rule

Showing how often events happen together and how much they are related to each other.

Use Binning

Use binning to convert numeric variables into categorical variables, which are divided into binary variables greater than or less than the mean.

	Name	Distribution	Mean
N	AREA		805.6831

- C area_cat := 0 if AREA > 805.6831 else 1
- C estate_cat := 0 if ESTATE > 9526646.06 else 1
- C east_cat := 0 if DIST_EASTDC > 11.4242 else 1
- C west_cat := 0 if DIST_WESTDC > 15.2233 else 1

" Data when AREA is above 805.6831 on average "

GAS_STATION	area_cat	estate_cat	east_cat	west_cat	korea_cat	pop_cat	sch_cat	apart_cat	mkt_cat	SCORE_CAT
대성석유㈜관문...	High	Low	Far	Short	Far	High	High	High	High	적합
㈜타이거오일 ...	Low	Low	Far	Short	Far	High	High	High	High	적합
현대오일뱅크㈜...	Low	Low	Far	Short	Far	High	High	High	High	적합
우장산주유소	Low	Low	Far	Short	Far	High	High	High	High	부적합
주식회사 에스...	High	High	Far	Short	Far	High	High	High	High	부적합

The result of the change between "Minsup" and "Minconf"

Minsup = 10%, Minconf = 75%

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.111	0.800	0.139	1.815	3.180	0.076	estate_cat=High, pop_cat=Low, apart_cat=Low	→	SCORE_CAT=매우부적합
0.136	0.780	0.175	1.439	3.102	0.092	estate_cat=High, pop_cat=Low	→	SCORE_CAT=매우부적합

Minsup = 10%, Minconf = 65%

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.100	0.746	0.134	1.857	2.991	0.067	area_cat=High, estate_cat=Low, pop_cat=High	→	SCORE_CAT=매우적합
0.143	0.670	0.213	1.180	2.663	0.089	pop_cat=Low, sch_cat=Low, apart_cat=Low	→	SCORE_CAT=매우부적합

Minsup = 20%, Minconf = 40%

Supp	Conf	Covr	Strg	Lift	Levr	Antecedent		Consequent
0.213	0.529	0.403	0.619	2.121	0.113	estate_cat=Low, pop_cat=High	→	SCORE_CAT=매우적합
0.232	0.429	0.542	0.461	1.720	0.097	pop_cat=High	→	SCORE_CAT=매우적합

05. Conclusion

Expectation effectiveness

SEOUL

There is a possibility that the introduction of MFC in Seoul on a trial basis will cause innovation in the logistics industry
→Recession caused by COVID-19, can revitalize the economy

CUSTOMER

As the logistics delivery process decreases, various products can be received faster than now

GAS STATION

Even those who have difficulty building their own fulfillment into the city at a low cost
Can increase the competitiveness in the market

LOGISTICS COMPANY

Provides high delivery service at low cost while gaining advantage in the delivery market

Limitation

[1] Logistics data is only available by district, so logistics variables are not considered in the classification model
[2] Considering only the internal conditions of Seoul when analyzing the location
[3] There may be other meaningful variables that we did not think
[4] We don't use MCLP, so we can't find the best place where covers all demands.
[5] Our project can’t consider coverage

Team Member’s role

강대경 : 201800302
Data-collection, Data-preprocessing, Association, presentation material

김 찬 : 201801158
Data-collection, Data-preprocessing, EDA for clustering, ANN, PT, Final report

신 용 : 201802033
Data-collection, Data-preprocessing, Clustering, Classification, EDA for clustering, Appendix

김나연 : 201904193
Data-collection, Data-preprocessing using Qgis, EDA for classification, presentation material, Final report

References

- "CJ대한통운, SK 주유소를 물류거점으로 쓴다."(2021.08.27) 한겨레

- "가구 픽업부터 도심물류거점까지… 주유소 활용 나선 이유는?"(2021-10-06) UPI뉴스

- "CJ대한통운, SK에너지, 주유소를 도심 물류센터로"(2021.08.26) 조선비즈

- 영남권 물류센터 입지 선정 최적화에 관한 연구(2017) 윤관호, 차연두, 여기태

- 마이크로 풀필먼트 센터 도입방향 : 지역 중소 유통업체를 중심으로(2022) 박종헌, 권구포, 김정환

- GIS 및 혼합정수계획법을 이용한 공동물류센터 입지선정 (2016) 유한솔, 김남석

- GIS를 활용한 대전시 주유소 입지와 판매권역 분석 (2004) 김 민


Q & A
Thank you for watching!