

In [248]:

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_samples, silhouette_score
import matplotlib.pyplot as plt
import seaborn as sns
import matplotlib.cm as cm
import math
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import RobustScaler
from sklearn.mixture import GaussianMixture
from sklearn.cluster import DBSCAN
from sklearn.cluster import AgglomerativeClustering

import matplotlib as mpl
# 한글 깨짐 방지
from matplotlib import font_manager, rc
font_name =
font_manager.FontProperties(fname="C:/Windows/Fonts/malgun.ttf").get_name(

rc('font', family=font_name)
# 마이너스 수식 깨짐 방지
import matplotlib
matplotlib.rcParams['axes.unicode_minus'] = False

matplotlib.rc('xtick', labelsiz=20)
matplotlib.rc('ytick', labelsiz=20)
font= {'family' : font_name,
       'size' : 20}
matplotlib.rc('font', **font)
paramter = {'font.size': 20}

import warnings
warnings.filterwarnings(action = 'ignore')
%matplotlib inline
```

In [249]:

```
df = pd.read_csv('E:/dataset3.csv', encoding = 'cp949', parse_dates =
['날짜'])
```

In [250]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 610 entries, 0 to 609
Data columns (total 20 columns):
날짜                610 non-null datetime64[ns]
방역수칙            596 non-null object
확진자수            570 non-null float64
확진자수증감률      610 non-null float64
전일 대비 확진자수증감률  570 non-null float64
감염지수            560 non-null float64
폐업수             549 non-null float64
폐업증감률          610 non-null float64
카드매출액          518 non-null float64
전일 대비 매출증감률  610 non-null object
매출증감률          610 non-null object
18시 전 카드매출액  518 non-null float64
18시 후 카드매출액  518 non-null float64
18시 후 카드매출액증감률  610 non-null float64
카드매출건수        518 non-null float64
카드매출건수증감률  610 non-null float64
유동인구            518 non-null float64
유동인구증감률      610 non-null float64
이동거리            518 non-null float64
백신도입전후        362 non-null float64
dtypes: datetime64[ns](1), float64(16), object(3)
memory usage: 95.4+ KB
```

## 0. scaler호출 및 실루엣 점수 시각화 함수 생성

In [251]:

```
ss = StandardScaler()
```

In [252]:

```
def visualize_silhouette(cluster_lists, X_features):

    n_cols = len(cluster_lists)
    fig, axs = plt.subplots(figsize = (4*n_cols, 4), nrows = 1, ncols =
n_cols)

    for ind, n_cluster in enumerate(cluster_lists):

        clusterer = KMeans(n_clusters = n_cluster, max_iter = 500,
random_state = 777)
        cluster_labels = clusterer.fit_predict(X_features)

        sil_avg = silhouette_score(X_features, cluster_labels)
        sil_values = silhouette_samples(X_features, cluster_labels)
        y_lower = 10

        axs[ind].set_title('Number of Clusetr : ' + str(n_cluster)+'\n'
W
                           'Silhouette Score : ' + str(round(sil_avg,3)))
        axs[ind].set_xlabel('The silhouette coefficient values')
        axs[ind].set_ylabel('Cluster label')
        axs[ind].set_xlim([0, 1])
        axs[ind].set_ylim([0, len(X_features) + (n_cluster + 1) * 10])
```

```

    axs[ind].set_yticks([])
    axs[ind].set_xticks([0, 0.2, 0.4, 0.6, 0.8, 1])

    for i in range(n_cluster):
        ith_cluster_sil_values = sil_values[cluster_labels == i]
        ith_cluster_sil_values.sort()

        size_cluster_i = ith_cluster_sil_values.shape[0]
        y_upper = y_lower + size_cluster_i

        color = cm.nipy_spectral(float(i) / n_cluster)
        axs[ind].fill_betweenx(np.arange(y_lower, y_upper), 0,
                               ith_cluster_sil_values, facecolor = color, edgecolor = color,
                               alpha = 0.7)
        axs[ind].text(-0.05, y_lower + 0.5 * size_cluster_i, str(i))
        y_lower = y_upper + 10

    axs[ind].axvline(x = sil_avg, color = 'red', linestyle = '--')

```

In [253]:

```

def GMM_silhouette(cluster_lists, X_features):

    n_cols = len(cluster_lists)
    fig, axs = plt.subplots(figsize = (4*n_cols, 4), nrows = 1, ncols =
n_cols)

    for ind, n_cluster in enumerate(cluster_lists):

        clusterer = GaussianMixture(n_components = n_cluster,
random_state = 777)
        cluster_labels = clusterer.fit_predict(X_features)

        sil_avg = silhouette_score(X_features, cluster_labels)
        sil_values = silhouette_samples(X_features, cluster_labels)
        y_lower = 10

        axs[ind].set_title('Number of Clusetr : ' + str(n_cluster)+'\n'

W                               'Silhouette Score : ' + str(round(sil_avg,3)))
        axs[ind].set_xlabel('The silhouette coefficient values')
        axs[ind].set_ylabel('Cluster label')
        axs[ind].set_xlim([0, 1])
        axs[ind].set_ylim([0, len(X_features) + (n_cluster + 1) * 10])
        axs[ind].set_yticks([])

```

```

    axs[ind].set_xticks([0, 0.2, 0.4, 0.6, 0.8, 1])

    for i in range(n_cluster):
        ith_cluster_sil_values = sil_values[cluster_labels == i]
        ith_cluster_sil_values.sort()

        size_cluster_i = ith_cluster_sil_values.shape[0]
        y_upper = y_lower + size_cluster_i

        color = cm.nipy_spectral(float(i) / n_cluster)
        axs[ind].fill_betweenx(np.arange(y_lower, y_upper), 0,
                               ith_cluster_sil_values, facecolor = color, edgecolor = color,
                               alpha = 0.7)
        axs[ind].text(-0.05, y_lower + 0.5 * size_cluster_i, str(i))
        y_lower = y_upper + 10

    axs[ind].axvline(x = sil_avg, color = 'red', linestyle = '--')

```

In [254]:

```

def Agg_silhouette(cluster_lists, X_features):

    n_cols = len(cluster_lists)
    fig, axs = plt.subplots(figsize = (4*n_cols, 4), nrows = 1, ncols =
n_cols)

    for ind, n_cluster in enumerate(cluster_lists):

        clusterer = AgglomerativeClustering(n_clusters = n_cluster,
linkage = 'ward')
        cluster_labels = clusterer.fit_predict(X_features)

        sil_avg = silhouette_score(X_features, cluster_labels)
        sil_values = silhouette_samples(X_features, cluster_labels)
        y_lower = 10

        axs[ind].set_title('Number of Clusetr : ' + str(n_cluster)+'\n'

                               'Silhouette Score : ' + str(round(sil_avg,3)))
        axs[ind].set_xlabel('The silhouette coefficient values')
        axs[ind].set_ylabel('Cluster label')
        axs[ind].set_xlim([0, 1])
        axs[ind].set_ylim([0, len(X_features) + (n_cluster + 1) * 10])
        axs[ind].set_yticks([])
        axs[ind].set_xticks([0, 0.2, 0.4, 0.6, 0.8, 1])

```

```

for i in range(n_cluster):
    ith_cluster_sil_values = sil_values[cluster_labels == i]
    ith_cluster_sil_values.sort()

    size_cluster_i = ith_cluster_sil_values.shape[0]
    y_upper = y_lower + size_cluster_i

    color = cm.nipy_spectral(float(i) / n_cluster)
    axs[ind].fill_betweenx(np.arange(y_lower, y_upper), 0,
                           ith_cluster_sil_values, facecolor = color, edgecolor = color,
                           alpha = 0.7)
    axs[ind].text(-0.05, y_lower + 0.5 * size_cluster_i, str(i))
    y_lower = y_upper + 10

    axs[ind].axvline(x = sil_avg, color = 'red', linestyle = '--')

```

## 1. 전체 변수에 따른 군집화

In [255]:

```

col = ['확진자수', '확진자수증감률', '감염지수', '폐업수', '폐업증감률',
       '카드매출액', '매출증감률', '카드매출건수', '카드매출건수증감률',
       '18시 전 카드매출액', '18시 후 카드매출액', '18시 후 카드매출액증
       감률', '유동인구', '유동인구증감률', '이동거리']
df_all = df[col]

```

In [256]:

```

df_all = df_all[:518]
df_all.corr()

```

Out[256]:

	확진자수	확진자수 증감률	감염지수	폐업수	폐업증감 률	카드매출 액	카드매출 건수	카드매출 건수증감 률	18시 전 카드매출 액
확 진 자 수	1.000000	-0.074700	0.049248	0.111540	-0.011854	0.014611	-0.153726	-0.087188	0.174220
확 진 자 수 증 감 률	-0.074700	1.000000	0.350348	-0.036038	0.041347	-0.033377	0.004657	-0.040749	-0.066905
감 염 지 수	0.049248	0.350348	1.000000	0.026174	0.157910	0.049547	0.067257	-0.033687	0.030133

	확진자수	확진자수 증감률	감염지수	폐업수	폐업증감 률	카드매출 액	카드매출 건수	카드매출 건수증감 률	18시 전 카드매출 액
폐업수	0.111540	-0.036038	0.026174	1.000000	0.102512	0.507017	0.483466	0.069489	0.557086
폐업증감률	-0.011854	0.041347	0.157910	0.102512	1.000000	0.051905	0.054672	0.460779	0.049515
카드매출액	0.014611	-0.033377	0.049547	0.507017	0.051905	1.000000	0.911543	0.261472	0.969557
카드매출건수	-0.153726	0.004657	0.067257	0.483466	0.054672	0.911543	1.000000	0.250343	0.819687
카드매출건수증감률	-0.087188	-0.040749	-0.033687	0.069489	0.460779	0.261472	0.250343	1.000000	0.248681
18시 전 카드매출액	0.174220	-0.066905	0.030133	0.557086	0.049515	0.969557	0.819687	0.248681	1.000000
18시 후 카드매출액	-0.326205	0.042928	0.078574	0.295540	0.045792	0.855965	0.916267	0.235653	0.703448
18시 후 카드매출액증감률	-0.119954	-0.053671	-0.102595	0.069350	0.411233	0.261037	0.210195	0.936924	0.242940

	확진자수	확진자수 증감률	감염지수	폐업수	폐업증감 률	카드매출 액	카드매출 건수	카드매출 건수증감 률	18시 전 카드매출 액
유동인구	-0.009231	-0.034119	0.026343	0.579384	0.063093	0.848364	0.897026	0.205040	0.792702
유동인구 증감률	-0.161766	-0.049093	-0.103384	0.070498	0.380802	0.230230	0.235564	0.867426	0.212036
이동거리	-0.035869	0.007045	0.071185	0.186004	0.034311	0.687477	0.711508	0.134546	0.584539

In [257]:

```
df_all_scale = pd.DataFrame(ss.fit_transform(df_all))
df_all_scale.columns = col

df_all_scale
```

Out[257]:

	확진자수	확진자수 증감률	감염지수	폐업수	폐업증감 률	카드매출 액	매출증감 률	카드매출 건수	카드매출 건수증감 률
0	-0.894051	0.048805	1.582144	-1.278597	-0.231800	-2.393347	-0.303886	-2.325570	-0.746495
1	-0.886599	-0.011548	1.582144	3.088007	-0.010413	-0.681780	-0.481203	-0.356203	-0.670421
2	-0.931311	-0.544075	1.582144	0.479637	-0.157724	-0.641508	-0.303886	-0.404988	-0.061829
3	-0.879147	-0.295562	1.582144	0.151176	-0.113952	-0.842868	-0.702850	-0.585614	-0.594347
4	-0.923859	-0.505023	1.582144	0.209140	-0.152673	-0.722052	-0.569862	-0.457359	-0.290051
5	-0.879147	-0.338165	1.582144	-0.157964	-0.167825	-0.218650	-0.481203	0.053735	0.546764
6	-0.879147	-0.047050	1.582144	-1.278597	-0.231800	-0.581099	0.050748	-0.661062	0.166394
7	-0.856790	0.165960	-0.834149	-1.259276	-0.062603	-2.453755	-0.170899	-2.088770	0.394616
8	-0.737557	0.993152	-0.834149	0.267103	-0.202338	-0.742188	-0.126569	-0.092212	0.242468
9	-0.588515	16.305063	-0.834149	0.131855	-0.164458	-0.842868	-0.303886	-0.678034	-0.366125
10	-0.804626	0.318618	-0.834149	0.015927	-0.155198	-0.621372	0.139406	-0.117607	0.546764
11	-0.841886	3.716137	-0.834149	0.189818	-0.148464	-0.762324	-0.126569	-0.155004	0.318542
12	-0.864243	-0.086102	-0.834149	-0.235249	-0.153515	-0.178378	-0.037911	0.210436	0.090320
13	-0.871695	-0.139355	-0.834149	-1.278597	-0.147623	-0.440147	0.050748	-0.334658	0.394616
14	-0.886599	-0.331064	-1.187299	-1.278597	-0.231800	-2.252394	0.228065	-2.102586	-0.061829
15	-0.849338	-0.394967	-1.187299	0.576243	-0.130787	-0.802596	-0.126569	-0.257586	-0.290051
16	-0.886599	-0.497923	-1.187299	0.518280	-0.124895	-0.621372	0.139406	-0.258313	0.470690

	확진자수	확진자수 증감률	감염지수	폐업수	폐업증감 률	카드매출 액	매출증감 률	카드매출 건수	카드매출 건수증감 률
17	-0.849338	-0.313313	-1.187299	0.440995	-0.119844	-0.560963	-0.037911	-0.012481	0.090320
18	-0.804626	-0.039950	-1.187299	0.344388	-0.138363	-0.782460	-0.126569	-0.433510	-0.366125
19	-0.804626	0.126909	-1.187299	0.943347	-0.052502	-0.017289	0.050748	0.396293	0.166394
20	-0.871695	-0.189057	-1.187299	-1.278597	-0.147623	-0.359603	0.006418	-0.173421	0.166394
21	-0.886599	-0.189057	-0.611106	-1.278597	-0.147623	-2.312802	-0.170899	-1.866517	0.318542
22	-0.886599	-0.348815	-0.611106	0.576243	-0.147623	-0.560963	0.139406	0.005866	0.242468
23	-0.834434	0.226313	-0.611106	0.344388	-0.156040	-0.500555	0.050748	0.019772	0.242468
24	-0.834434	-0.125154	-0.611106	0.189818	-0.160249	-0.339466	0.139406	0.201507	0.166394
25	-0.826982	-0.252960	-0.611106	0.383031	-0.145939	-0.379739	0.316723	-0.021854	0.470690
26	-0.819530	-0.231659	-0.611106	0.093212	-0.179610	0.365296	0.228065	0.510555	0.090320
27	-0.767365	0.432224	-0.611106	-1.278597	-0.147623	-0.279058	0.006418	-0.162045	-0.061829
28	-0.767365	0.758840	-0.443824	-1.259276	-0.062603	-2.051033	0.316723	-1.674367	0.242468
29	-0.789722	0.581331	-0.443824	0.672850	-0.143414	-0.158242	0.272394	0.091133	0.014246
...	...	...	...	...	...	...	...	...	...
488	1.743994	-0.057701	0.206716	0.924026	-0.102167	1.331828	-0.348216	1.407020	-0.137903
489	1.222346	-0.121604	0.206716	-1.220634	0.020732	0.244480	-0.481203	-0.266779	-0.974718
490	1.356484	0.045254	0.281064	-1.239955	-0.198129	-1.567767	-0.170899	-1.599027	-0.670421
491	1.453362	0.009753	0.281064	0.402352	-0.154357	0.747882	-0.392545	0.759547	-0.061829
492	3.405813	0.006203	0.281064	0.498958	-0.153515	0.667337	-0.614191	0.828605	0.014246
493	3.167345	0.041704	0.281064	0.170497	-0.192237	0.405568	-0.968825	0.504112	-0.518273
494	2.817096	-0.029299	0.281064	0.479637	-0.201496	0.284752	-0.614191	0.457851	-0.594347
495	2.861809	-0.039950	0.281064	0.035248	-0.181293	0.647201	-0.525533	0.832856	-0.594347
496	2.861809	0.080756	0.281064	-1.220634	-0.147623	0.103527	-0.215228	-0.031817	0.242468
497	2.071886	-0.079002	0.429759	-1.220634	-0.105534	-2.091305	-0.702850	-1.889170	-0.518273
498	2.176216	-0.082552	0.429759	0.440995	-0.145939	-0.218650	-0.791508	0.183669	-0.670421
499	3.815678	-0.157106	0.429759	-0.100001	-0.176243	-0.097834	-0.658520	0.191197	-0.746495
500	2.936330	-0.210358	0.429759	0.035248	-0.155198	-0.097834	-0.481203	0.200777	-0.366125
501	3.308935	-0.142905	0.429759	-0.042037	-0.172876	-0.138106	-0.436874	0.026893	-0.518273
502	3.234414	-0.153555	0.429759	-0.196607	-0.162774	0.445841	-0.215228	0.440825	-0.442199
503	2.958686	-0.178407	0.429759	-1.239955	-0.175401	-0.198514	-0.348216	-0.222625	-0.290051
504	2.191120	-0.174856	0.002261	-1.278597	-0.231800	-1.910081	0.139406	-1.692195	0.242468
505	2.004817	-0.210358	0.002261	0.189818	-0.160249	-0.117970	0.006418	-0.040306	-0.290051
506	3.569759	-0.206808	0.002261	-0.022716	-0.141730	0.264616	0.228065	0.237987	0.014246
507	2.787288	-0.203258	0.002261	-0.409141	-0.176243	0.244480	0.228065	0.303183	0.090320
508	2.936330	-0.221009	0.002261	-0.003394	-0.145097	0.143799	0.183735	0.201664	0.166394

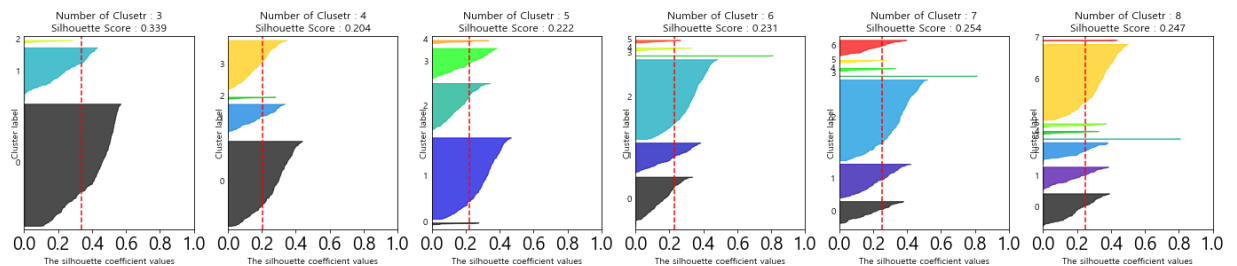


	확진자수	확진자수 증감률	감염지수	폐업수	폐업증감 률	카드매출 액	매출증감 률	카드매출 건수	카드매출 건수증감 률
<b>509</b>	2.533916	-0.249410	0.002261	0.286425	-0.109743	0.788154	0.183735	0.650824	0.166394
<b>510</b>	2.712767	-0.210358	0.002261	-1.259276	-0.189711	-0.138106	-0.037911	-0.188402	0.014246
<b>511</b>	1.609856	-0.256511	-0.090674	-1.220634	0.107435	-2.191986	-0.481203	-1.903596	-0.366125
<b>512</b>	1.684377	-0.228109	-0.090674	0.073891	-0.154357	0.163936	0.183735	0.345525	0.394616
<b>513</b>	3.338744	-0.206808	-0.090674	3.493753	0.088074	0.586793	0.183735	0.286545	0.014246
<b>514</b>	2.899070	-0.178407	-0.090674	-0.235249	-0.130787	0.365296	0.006418	0.228376	-0.137903
<b>515</b>	2.705315	-0.210358	-0.090674	0.325067	-0.125736	0.465977	0.183735	0.248219	0.014246
<b>516</b>	2.608438	-0.181957	-0.090674	0.189818	-0.152673	1.090195	0.139406	0.605825	-0.137903
<b>517</b>	2.593533	-0.199708	-0.090674	-0.872851	1.535926	-0.178378	-0.126569	-0.347440	-0.213977

518 rows × 15 columns

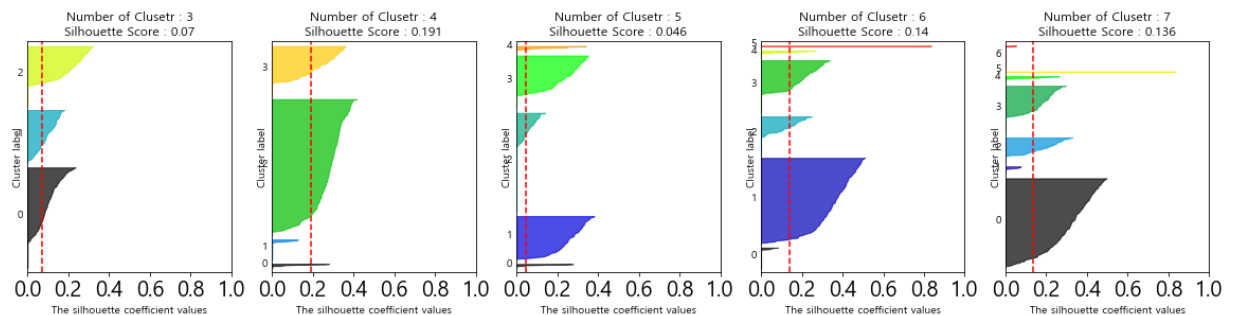
In [258]:

```
visualize_silhouette([3,4,5,6,7,8], df_all_scale)
```



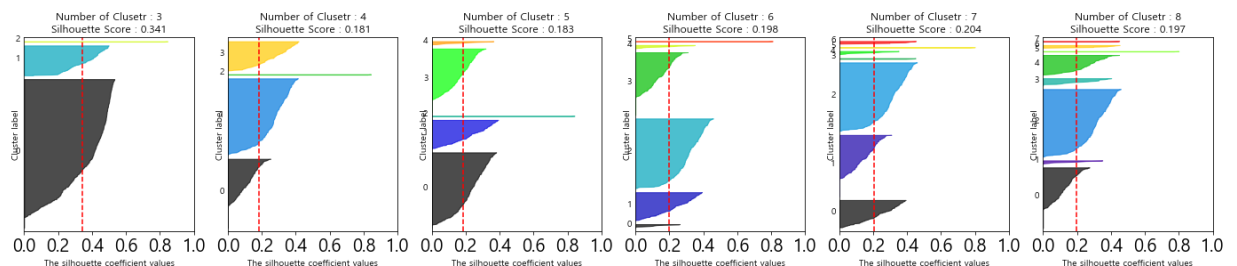
In [259]:

```
GMM_silhouette([3,4,5,6,7], df_all_scale)
```



In [260]:

```
Agg_silhouette([3,4,5,6,7,8], df_all_scale)
```



k-means가 평균적 실루엣 점수가 높음

## 2. 파생변수 없이 군집화

In [261]:

```
col = ['확진자수', '감염지수', '폐업수', '카드매출액', '카드매출건수',
       '18시 전 카드매출액', '18시 후 카드매출액', '유동인구', '이동거리']
df_nonmute = df[col]
df_nonmute.head()
```

Out[261]:

	확 진 자 수	감 염 지 수	폐 업 수	카드매출액	카드매출 건수	18시 전 카드 매출액	18시 후 카드 매출액	유동인구	이동거리
0	5.0	1.94	0.0	1.420000e+11	5774284.0	9.130702e+10	5.099040e+10	6877708.0	4.042818e+10
1	6.0	1.94	226.0	2.270000e+11	8450794.0	1.570000e+11	7.000068e+10	11615759.0	5.695667e+10
2	0.0	1.94	91.0	2.290000e+11	8384492.0	1.550000e+11	7.399556e+10	11655130.0	5.612155e+10
3	7.0	1.94	74.0	2.190000e+11	8139008.0	1.490000e+11	6.965669e+10	11270053.0	5.588185e+10
4	1.0	1.94	77.0	2.250000e+11	8313315.0	1.510000e+11	7.437857e+10	11604949.0	5.753296e+10

In [262]:

```
df_nonmute = df_nonmute[:518]
df_nonmute.corr()
```

Out[262]:

	확진자수	감염지수	폐업수	카드매출 액	카드매출 건수	18시 전 카드매출 액	18시 후 카드매출 액	유동인구	이동거리
확 진 자 수	1.000000	0.049248	0.111540	0.014611	-0.153726	0.174220	-0.326205	-0.009231	-0.035869
감 염 지 수	0.049248	1.000000	0.026174	0.049547	0.067257	0.030133	0.078574	0.026343	0.071185
폐 업 수	0.111540	0.026174	1.000000	0.507017	0.483466	0.557086	0.295540	0.579384	0.186004
카 드 매 출 액	0.014611	0.049547	0.507017	1.000000	0.911543	0.969557	0.855965	0.848364	0.687477
카 드 매 출 건 수	-0.153726	0.067257	0.483466	0.911543	1.000000	0.819687	0.916267	0.897026	0.711508

	확진자수	감염지수	폐업수	카드매출 액	카드매출 건수	18시 전 카드매출 액	18시 후 카드매출 액	유동인구	이동거리
18 시 전 카 드 매 출 액	0.174220	0.030133	0.557086	0.969557	0.819687	1.000000	0.703448	0.792702	0.584539
18 시 후 카 드 매 출 액	-0.326205	0.078574	0.295540	0.855965	0.916267	0.703448	1.000000	0.789023	0.761065
유 동 인 구	-0.009231	0.026343	0.579384	0.848364	0.897026	0.792702	0.789023	1.000000	0.786713
이 동 거 리	-0.035869	0.071185	0.186004	0.687477	0.711508	0.584539	0.761065	0.786713	1.000000

In [263]:

```
df_nonmute_scale = pd.DataFrame(ss.fit_transform(df_nonmute))
df_nonmute_scale.columns = col

df_nonmute_scale
```

Out [263]:

	확진자수	감염지수	폐업수	카드매출 액	카드매출 건수	18시 전 카드매출 액	18시 후 카드매출 액	유동인구	이동거리
0	-0.894051	1.582144	-1.278597	-2.393347	-2.325570	-2.477514	-1.699593	-2.723383	-3.201881
1	-0.886599	1.582144	3.088007	-0.681780	-0.356203	-0.661122	-0.584453	-0.691159	-1.476383
2	-0.931311	1.582144	0.479637	-0.641508	-0.404988	-0.716422	-0.350114	-0.674272	-1.563565
3	-0.879147	1.582144	0.151176	-0.842868	-0.585614	-0.882320	-0.604631	-0.839437	-1.588589
4	-0.923859	1.582144	0.209140	-0.722052	-0.457359	-0.827020	-0.327646	-0.695795	-1.416220
5	-0.879147	1.582144	-0.157964	-0.218650	0.053735	-0.467574	0.356991	-0.459137	-0.774068
6	-0.879147	1.582144	-1.278597	-0.581099	-0.661062	-0.744071	-0.106128	-1.411448	-1.827342
7	-0.856790	-0.834149	-1.259276	-2.453755	-2.088770	-2.645610	-1.537348	-2.268955	-2.517430
8	-0.737557	-0.834149	0.267103	-0.742188	-0.092212	-0.744071	-0.541342	-0.426327	-1.100277
9	-0.588515	-0.834149	0.131855	-0.842868	-0.678034	-0.909970	-0.554394	-1.002048	-1.597335
10	-0.804626	-0.834149	0.015927	-0.621372	-0.117607	-0.716422	-0.263714	-0.515285	-1.166553
11	-0.841886	-0.834149	0.189818	-0.762324	-0.155004	-0.909970	-0.269263	-0.552757	-1.114059

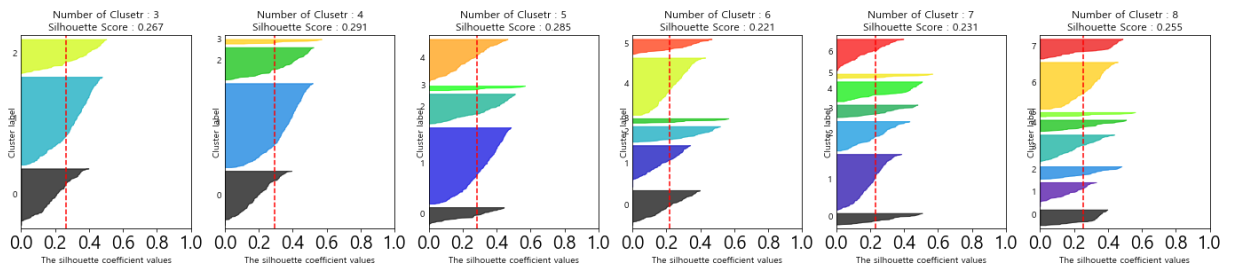
	확진자수	감염지수	폐업수	카드매출 액	카드매출 건수	18시 전 카드매출 액	18시 후 카드매출 액	유동인구	이동거리
12	-0.864243	-0.834149	-0.235249	-0.178378	0.210436	-0.439925	0.398580	-0.464928	-0.544988
13	-0.871695	-0.834149	-1.278597	-0.440147	-0.334658	-0.605823	-0.021184	-1.328184	-1.476672
14	-0.886599	-1.187299	-1.278597	-2.252394	-2.102586	-2.395240	-1.484310	-2.422617	-2.507333
15	-0.849338	-1.187299	0.576243	-0.802596	-0.257586	-0.771721	-0.661255	-0.511394	-1.105887
16	-0.886599	-1.187299	0.518280	-0.621372	-0.258313	-0.688772	-0.349948	-0.549703	-1.236393
17	-0.849338	-1.187299	0.440995	-0.560963	-0.012481	-0.716422	-0.145953	-0.284221	-1.024655
18	-0.804626	-1.187299	0.344388	-0.782460	-0.433510	-0.909970	-0.334309	-0.842587	-1.302102
19	-0.804626	-1.187299	0.943347	-0.017289	0.396293	-0.274026	0.531027	-0.141432	-0.258468
20	-0.871695	-1.187299	-1.278597	-0.359603	-0.173421	-0.522874	0.076933	-1.037741	-1.203264
21	-0.886599	-0.611106	-1.278597	-2.312802	-1.866517	-2.476570	-1.468671	-2.028874	-2.007976
22	-0.886599	-0.611106	0.576243	-0.560963	0.005866	-0.495224	-0.592003	-0.295957	-0.892122
23	-0.834434	-0.611106	0.344388	-0.500555	0.019772	-0.522874	-0.320212	-0.239927	-0.973246
24	-0.834434	-0.611106	0.189818	-0.339466	0.201507	-0.412275	-0.122420	-0.155260	-0.871539
25	-0.826982	-0.611106	0.383031	-0.379739	-0.021854	-0.384625	-0.254748	-0.392231	-1.005775
26	-0.819530	-0.611106	0.093212	0.365296	0.510555	0.196019	0.647127	-0.180903	-0.300708
27	-0.767365	-0.611106	-1.278597	-0.279058	-0.162045	-0.412275	0.044069	-1.011720	-1.127548
28	-0.767365	-0.443824	-1.259276	-2.051033	-1.674367	-2.237156	-1.235069	-1.880289	-1.939445
29	-0.789722	-0.443824	0.672850	-0.158242	0.091133	-0.025179	-0.403916	-0.207890	-0.830485
...	...	...	...	...	...	...	...	...	...
488	1.743994	0.206716	0.924026	1.331828	1.407020	1.108459	1.527262	1.565404	1.891970
489	1.222346	0.206716	-1.220634	0.244480	-0.266779	0.555465	-0.455026	-0.189327	0.688269
490	1.356484	0.281064	-1.239955	-1.567767	-1.599027	-1.545913	-1.313337	-1.333490	-0.530197
491	1.453362	0.281064	0.402352	0.747882	0.759547	0.887261	0.288000	1.321163	0.775094
492	3.405813	0.281064	0.498958	0.667337	0.828605	0.721363	0.424364	1.376369	0.627708
493	3.167345	0.281064	0.170497	0.405568	0.504112	0.527815	0.032592	1.121108	0.425805
494	2.817096	0.281064	0.479637	0.284752	0.457851	0.334267	0.123826	0.909343	0.369486
495	2.861809	0.281064	0.035248	0.647201	0.832856	0.555465	0.675989	1.048622	1.194453
496	2.861809	0.281064	-1.220634	0.103527	-0.031817	0.278968	-0.280342	-0.330011	0.125644
497	2.071886	0.429759	-1.220634	-2.091305	-1.889170	-2.015958	-1.796605	-1.508500	-0.793160
498	2.176216	0.429759	0.440995	-0.218650	0.183669	0.168369	-0.989972	0.374470	0.000278
499	3.815678	0.429759	-0.100001	-0.097834	0.191197	0.251318	-0.830332	0.347679	-0.191092
500	2.936330	0.429759	0.035248	-0.097834	0.200777	0.196019	-0.719759	0.203236	-0.167967
501	3.308935	0.429759	-0.042037	-0.138106	0.026893	0.196019	-0.848134	0.092271	-0.136517
502	3.234414	0.429759	-0.196607	0.445841	0.440825	0.776663	-0.379438	0.263252	0.643496
503	2.958686	0.429759	-1.239955	-0.198514	-0.222625	0.113070	-0.831873	-0.739233	-0.355175

	확진자수	감염지수	폐업수	카드매출 액	카드매출 건수	18시 전 카드매출 액	18시 후 카드매출 액	유동인구	이동거리
504	2.191120	0.002261	-1.278597	-1.910081	-1.692195	-1.822410	-1.730915	-1.812349	-1.044457
505	2.004817	0.002261	0.189818	-0.117970	-0.040306	0.306618	-1.008430	0.172608	-0.072505
506	3.569759	0.002261	-0.022716	0.264616	0.237987	0.693713	-0.755515	0.333431	-0.136599
507	2.787288	0.002261	-0.409141	0.244480	0.303183	0.610764	-0.613105	0.259261	-0.094832
508	2.936330	0.002261	-0.003394	0.143799	0.201664	0.555465	-0.727908	0.153115	-0.016845
509	2.533916	0.002261	0.286425	0.788154	0.650824	1.136109	-0.077214	0.317857	0.743321
510	2.712767	0.002261	-1.259276	-0.138106	-0.188402	0.140719	-0.665545	-0.814211	-0.298980
511	1.609856	-0.090674	-1.220634	-2.191986	-1.903596	-2.071257	-2.008757	-2.002721	-1.076431
512	1.684377	-0.090674	0.073891	0.163936	0.345525	0.583115	-0.745606	0.073591	-0.093127
513	3.338744	-0.090674	3.493753	0.586793	0.286545	1.136109	-0.650981	-0.005923	-0.374328
514	2.899070	-0.090674	-0.235249	0.365296	0.228376	0.887261	-0.800168	0.176837	0.075075
515	2.705315	-0.090674	0.325067	0.465977	0.248219	0.942561	-0.679237	0.156095	0.231357
516	2.608438	-0.090674	0.189818	1.090195	0.605825	1.523205	-0.012587	0.260614	1.025854
517	2.593533	-0.090674	-0.872851	-0.178378	-0.347440	0.140719	-0.802671	-1.004317	-0.363758

518 rows × 9 columns

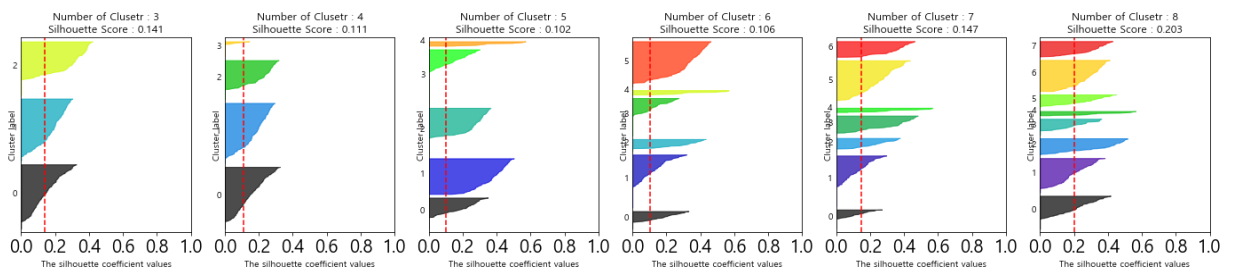
In [264]:

```
visualize_silhouette([3,4,5,6,7,8], df_nonmute_scale)
```



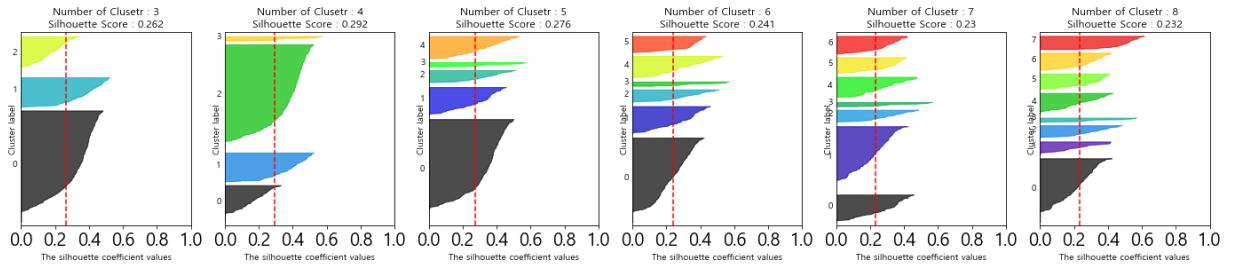
In [265]:

```
GMM_silhouette([3,4,5,6,7,8], df_nonmute_scale)
```



In [266]:

```
Agg_silhouette([3,4,5,6,7,8], df_nonmute_scale)
```



K-means 평균적 실루엣 점수가 높음

### 3. 파생변수만 가지고 군집화

```
In [267]: col = ['확진자수증감률', '폐업증감률', '매출증감률', '카드매출건수증감률',
               '18시 후 카드매출액증감률', '유동인구증감률']
df_mute = df[col]
df_mute.head()
```

```
Out[267]:
```

	확진자수증감률	폐업증감률	매출증감률	카드매출건수증감률	18시 후 카드매출액증감률	유동인구증감률
0	0.67	-1.00	-0.05	-0.09	-0.09	-0.22
1	0.50	1.63	-0.09	-0.08	-0.07	-0.08
2	-1.00	-0.12	-0.05	0.00	-0.02	0.04
3	-0.30	0.40	-0.14	-0.07	-0.07	-0.05
4	-0.89	-0.06	-0.11	-0.03	-0.03	-0.02

```
In [268]: df_mute = df_mute[:518]
df_mute.corr()
```

```
Out[268]:
```

	확진자수증감률	폐업증감률	카드매출건수증감률	18시 후 카드매출액증감률	유동인구증감률
확진자수증감률	1.000000	0.041347	-0.040749	-0.053671	-0.049093
폐업증감률	0.041347	1.000000	0.460779	0.411233	0.380802
카드매출건수증감률	-0.040749	0.460779	1.000000	0.936924	0.867426
18시 후 카드매출액증감률	-0.053671	0.411233	0.936924	1.000000	0.791082
유동인구증감률	-0.049093	0.380802	0.867426	0.791082	1.000000

```
In [269]: df_mute_scale = pd.DataFrame(ss.fit_transform(df_mute))
df_mute_scale.columns = col

df_mute_scale
```

```
Out[269]:
```

	확진자수증감률	폐업증감률	매출증감률	카드매출건수증감률	18시 후 카드매출액증감률	유동인구증감률
--	---------	-------	-------	-----------	----------------	---------

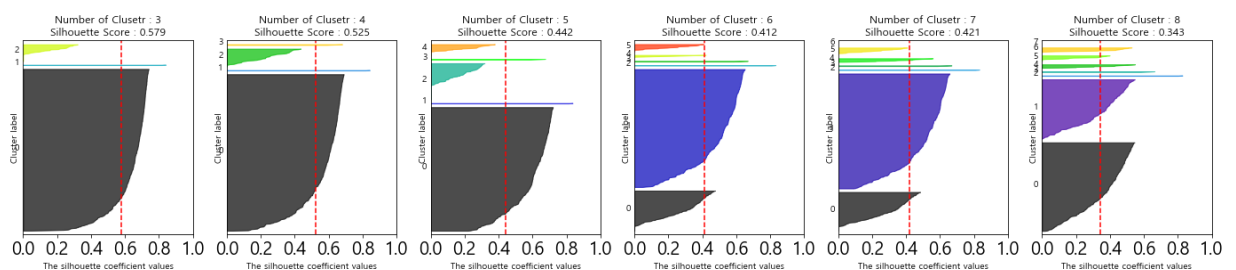
	확진자수증감 률	폐업증감 률	매출증감 률	카드매출건수증감 률	18시 후 카드매출액증 감률	유동인구증감 률
0	0.048805	-0.231800	-0.303886	-0.746495	-0.548756	-1.903780
1	-0.011548	-0.010413	-0.481203	-0.670421	-0.440736	-0.736532
2	-0.544075	-0.157724	-0.303886	-0.061829	-0.170685	0.263967
3	-0.295562	-0.113952	-0.702850	-0.594347	-0.440736	-0.486407
4	-0.505023	-0.152673	-0.569862	-0.290051	-0.224695	-0.236282
5	-0.338165	-0.167825	-0.481203	0.546764	0.369417	0.930966
6	-0.047050	-0.231800	0.050748	0.166394	0.207387	0.514091
7	0.165960	-0.062603	-0.170899	0.394616	0.207387	1.181090
8	0.993152	-0.202338	-0.126569	0.242468	-0.008654	0.347342
9	16.305063	-0.164458	-0.303886	-0.366125	-0.332715	-0.653157
10	0.318618	-0.155198	0.139406	0.546764	0.369417	0.514091
11	3.716137	-0.148464	-0.126569	0.318542	-0.008654	0.180592
12	-0.086102	-0.153515	-0.037911	0.090320	-0.008654	-0.069533
13	-0.139355	-0.147623	0.050748	0.394616	0.045356	0.097217
14	-0.331064	-0.231800	0.228065	-0.061829	0.045356	-0.486407
15	-0.394967	-0.130787	-0.126569	-0.290051	-0.224695	-0.236282
16	-0.497923	-0.124895	0.139406	0.470690	0.207387	0.764216
17	-0.313313	-0.119844	-0.037911	0.090320	0.099366	0.263967
18	-0.039950	-0.138363	-0.126569	-0.366125	-0.116674	-0.569782
19	0.126909	-0.052502	0.050748	0.166394	0.099366	0.430716
20	-0.189057	-0.147623	0.006418	0.166394	0.045356	0.514091
21	-0.189057	-0.147623	-0.170899	0.318542	-0.062664	0.930966
22	-0.348815	-0.147623	0.139406	0.242468	0.045356	0.263967
23	0.226313	-0.156040	0.050748	0.242468	-0.008654	0.430716
24	-0.125154	-0.160249	0.139406	0.166394	-0.008654	0.097217
25	-0.252960	-0.145939	0.316723	0.470690	0.045356	0.680841
26	-0.231659	-0.179610	0.228065	0.090320	0.045356	-0.152908
27	0.432224	-0.147623	0.006418	-0.061829	-0.116674	0.013842
28	0.758840	-0.062603	0.316723	0.242468	0.315407	0.263967
29	0.581331	-0.143414	0.272394	0.014246	0.207387	0.097217
...	...	...	...	...	...	...
488	-0.057701	-0.102167	-0.348216	-0.137903	-0.116674	-0.069533
489	-0.121604	0.020732	-0.481203	-0.974718	-1.304898	-0.819906
490	0.045254	-0.198129	-0.170899	-0.670421	-0.224695	-1.236781
491	0.009753	-0.154357	-0.392545	-0.061829	-0.170685	0.180592

	확진자수증감 률	폐업증감 률	매출증감 률	카드매출건수증감 률	18시 후 카드매출액증 감률	유동인구증감 률
492	0.006203	-0.153515	-0.614191	0.014246	-0.278705	0.764216
493	0.041704	-0.192237	-0.968825	-0.518273	-1.088858	-0.152908
494	-0.029299	-0.201496	-0.614191	-0.594347	-0.872817	-0.653157
495	-0.039950	-0.181293	-0.525533	-0.594347	-0.764797	-0.653157
496	0.080756	-0.147623	-0.215228	0.242468	0.153376	-0.319657
497	-0.079002	-0.105534	-0.702850	-0.518273	-0.818807	-0.403032
498	-0.082552	-0.145939	-0.791508	-0.670421	-1.466929	-1.236781
499	-0.157106	-0.176243	-0.658520	-0.746495	-1.412919	-1.320156
500	-0.210358	-0.155198	-0.481203	-0.366125	-0.926827	-1.236781
501	-0.142905	-0.172876	-0.436874	-0.518273	-1.142868	-1.070031
502	-0.153555	-0.162774	-0.215228	-0.442199	-1.142868	-1.070031
503	-0.178407	-0.175401	-0.348216	-0.290051	-0.764797	-0.736532
504	-0.174856	-0.231800	0.139406	0.242468	0.045356	-0.653157
505	-0.210358	-0.160249	0.006418	-0.290051	-0.062664	-0.319657
506	-0.206808	-0.141730	0.228065	0.014246	0.045356	-0.069533
507	-0.203258	-0.176243	0.228065	0.090320	0.099366	0.013842
508	-0.221009	-0.145097	0.183735	0.166394	0.099366	0.013842
509	-0.249410	-0.109743	0.183735	0.166394	0.315407	0.013842
510	-0.210358	-0.189711	-0.037911	0.014246	0.153376	-0.236282
511	-0.256511	0.107435	-0.481203	-0.366125	-0.548756	-0.486407
512	-0.228109	-0.154357	0.183735	0.394616	0.315407	-0.236282
513	-0.206808	0.088074	0.183735	0.014246	0.099366	-0.569782
514	-0.178407	-0.130787	0.006418	-0.137903	-0.332715	-0.152908
515	-0.210358	-0.125736	0.183735	0.014246	-0.008654	-0.069533
516	-0.181957	-0.152673	0.139406	-0.137903	-0.008654	-0.152908
517	-0.199708	1.535926	-0.126569	-0.213977	-0.224695	-0.403032

518 rows × 6 columns

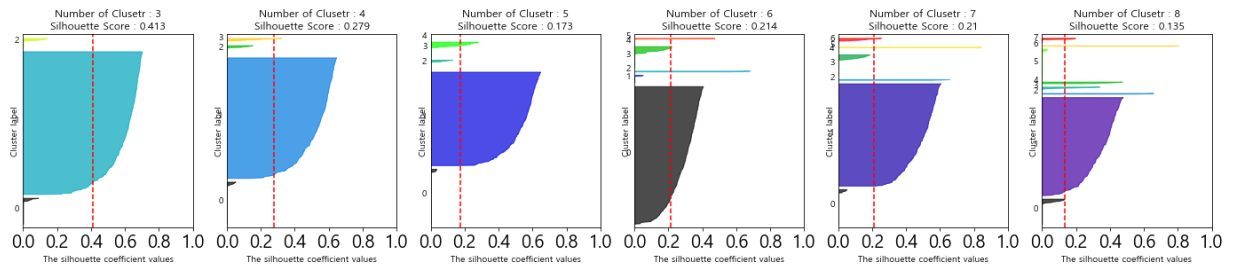
In [270]:

visualize\_silhouette([3,4,5,6,7,8], df\_mute\_scale)

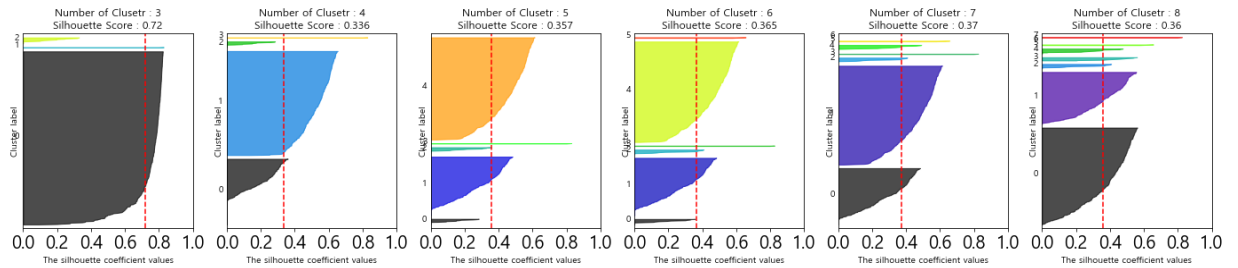




```
In [271]: GMM_silhouette([3,4,5,6,7,8], df_mute_scale)
```



```
In [272]: Agg_silhouette([3,4,5,6,7,8], df_mute_scale)
```



## 4. 적절히 섞어 군집화

```
In [273]: col = ['확진자수', '확진자수증감률', '감염지수', '카드매출액', '매출증감률',  
                '유동인구', '유동인구증감률']  
df_ano = df[col]  
df_ano.head()
```

	확진자수	확진자수증감률	감염지수	카드매출액	매출증감률	유동인구	유동인구증감률
0	5.0	0.67	1.94	1.420000e+11	-0.05	6877708.0	-0.22
1	6.0	0.50	1.94	2.270000e+11	-0.09	11615759.0	-0.08
2	0.0	-1.00	1.94	2.290000e+11	-0.05	11655130.0	0.04
3	7.0	-0.30	1.94	2.190000e+11	-0.14	11270053.0	-0.05
4	1.0	-0.89	1.94	2.250000e+11	-0.11	11604949.0	-0.02

```
In [274]: df_ano = df_ano[:,5:18]  
df_ano.corr()
```

	확진자수	확진자수증감률	감염지수	카드매출액	유동인구	유동인구증감률
확진자수	1.000000	-0.074700	0.049248	0.014611	-0.009231	-0.161766
확진자수증감률	-0.074700	1.000000	0.350348	-0.033377	-0.034119	-0.049093
감염지수	0.049248	0.350348	1.000000	0.049547	0.026343	-0.103384
카드매출액	0.014611	-0.033377	0.049547	1.000000	0.848364	0.230230
유동인구	-0.009231	-0.034119	0.026343	0.848364	1.000000	0.234593
유동인구증감률	-0.161766	-0.049093	-0.103384	0.230230	0.234593	1.000000

```
In [275]: df_ano_scale = pd.DataFrame(ss.fit_transform(df_ano))
df_ano_scale.columns = col

df_ano_scale
```

```
Out[275]:
```

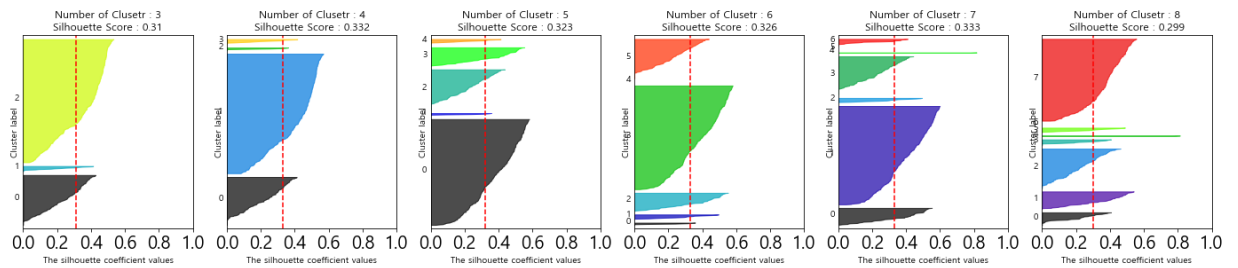
	확진자수	확진자수증감률	감염지수	카드매출액	매출증감률	유동인구	유동인구증감률
0	-0.894051	0.048805	1.582144	-2.393347	-0.303886	-2.723383	-1.903780
1	-0.886599	-0.011548	1.582144	-0.681780	-0.481203	-0.691159	-0.736532
2	-0.931311	-0.544075	1.582144	-0.641508	-0.303886	-0.674272	0.263967
3	-0.879147	-0.295562	1.582144	-0.842868	-0.702850	-0.839437	-0.486407
4	-0.923859	-0.505023	1.582144	-0.722052	-0.569862	-0.695795	-0.236282
5	-0.879147	-0.338165	1.582144	-0.218650	-0.481203	-0.459137	0.930966
6	-0.879147	-0.047050	1.582144	-0.581099	0.050748	-1.411448	0.514091
7	-0.856790	0.165960	-0.834149	-2.453755	-0.170899	-2.268955	1.181090
8	-0.737557	0.993152	-0.834149	-0.742188	-0.126569	-0.426327	0.347342
9	-0.588515	16.305063	-0.834149	-0.842868	-0.303886	-1.002048	-0.653157
10	-0.804626	0.318618	-0.834149	-0.621372	0.139406	-0.515285	0.514091
11	-0.841886	3.716137	-0.834149	-0.762324	-0.126569	-0.552757	0.180592
12	-0.864243	-0.086102	-0.834149	-0.178378	-0.037911	-0.464928	-0.069533
13	-0.871695	-0.139355	-0.834149	-0.440147	0.050748	-1.328184	0.097217
14	-0.886599	-0.331064	-1.187299	-2.252394	0.228065	-2.422617	-0.486407
15	-0.849338	-0.394967	-1.187299	-0.802596	-0.126569	-0.511394	-0.236282
16	-0.886599	-0.497923	-1.187299	-0.621372	0.139406	-0.549703	0.764216
17	-0.849338	-0.313313	-1.187299	-0.560963	-0.037911	-0.284221	0.263967
18	-0.804626	-0.039950	-1.187299	-0.782460	-0.126569	-0.842587	-0.569782
19	-0.804626	0.126909	-1.187299	-0.017289	0.050748	-0.141432	0.430716
20	-0.871695	-0.189057	-1.187299	-0.359603	0.006418	-1.037741	0.514091
21	-0.886599	-0.189057	-0.611106	-2.312802	-0.170899	-2.028874	0.930966
22	-0.886599	-0.348815	-0.611106	-0.560963	0.139406	-0.295957	0.263967
23	-0.834434	0.226313	-0.611106	-0.500555	0.050748	-0.239927	0.430716
24	-0.834434	-0.125154	-0.611106	-0.339466	0.139406	-0.155260	0.097217
25	-0.826982	-0.252960	-0.611106	-0.379739	0.316723	-0.392231	0.680841
26	-0.819530	-0.231659	-0.611106	0.365296	0.228065	-0.180903	-0.152908
27	-0.767365	0.432224	-0.611106	-0.279058	0.006418	-1.011720	0.013842
28	-0.767365	0.758840	-0.443824	-2.051033	0.316723	-1.880289	0.263967
29	-0.789722	0.581331	-0.443824	-0.158242	0.272394	-0.207890	0.097217
...	...	...	...	...	...	...	...
488	1.743994	-0.057701	0.206716	1.331828	-0.348216	1.565404	-0.069533

	확진자수	확진자수증감률	감염지수	카드매출액	매출증감률	유동인구	유동인구증감률
489	1.222346	-0.121604	0.206716	0.244480	-0.481203	-0.189327	-0.819906
490	1.356484	0.045254	0.281064	-1.567767	-0.170899	-1.333490	-1.236781
491	1.453362	0.009753	0.281064	0.747882	-0.392545	1.321163	0.180592
492	3.405813	0.006203	0.281064	0.667337	-0.614191	1.376369	0.764216
493	3.167345	0.041704	0.281064	0.405568	-0.968825	1.121108	-0.152908
494	2.817096	-0.029299	0.281064	0.284752	-0.614191	0.909343	-0.653157
495	2.861809	-0.039950	0.281064	0.647201	-0.525533	1.048622	-0.653157
496	2.861809	0.080756	0.281064	0.103527	-0.215228	-0.330011	-0.319657
497	2.071886	-0.079002	0.429759	-2.091305	-0.702850	-1.508500	-0.403032
498	2.176216	-0.082552	0.429759	-0.218650	-0.791508	0.374470	-1.236781
499	3.815678	-0.157106	0.429759	-0.097834	-0.658520	0.347679	-1.320156
500	2.936330	-0.210358	0.429759	-0.097834	-0.481203	0.203236	-1.236781
501	3.308935	-0.142905	0.429759	-0.138106	-0.436874	0.092271	-1.070031
502	3.234414	-0.153555	0.429759	0.445841	-0.215228	0.263252	-1.070031
503	2.958686	-0.178407	0.429759	-0.198514	-0.348216	-0.739233	-0.736532
504	2.191120	-0.174856	0.002261	-1.910081	0.139406	-1.812349	-0.653157
505	2.004817	-0.210358	0.002261	-0.117970	0.006418	0.172608	-0.319657
506	3.569759	-0.206808	0.002261	0.264616	0.228065	0.333431	-0.069533
507	2.787288	-0.203258	0.002261	0.244480	0.228065	0.259261	0.013842
508	2.936330	-0.221009	0.002261	0.143799	0.183735	0.153115	0.013842
509	2.533916	-0.249410	0.002261	0.788154	0.183735	0.317857	0.013842
510	2.712767	-0.210358	0.002261	-0.138106	-0.037911	-0.814211	-0.236282
511	1.609856	-0.256511	-0.090674	-2.191986	-0.481203	-2.002721	-0.486407
512	1.684377	-0.228109	-0.090674	0.163936	0.183735	0.073591	-0.236282
513	3.338744	-0.206808	-0.090674	0.586793	0.183735	-0.005923	-0.569782
514	2.899070	-0.178407	-0.090674	0.365296	0.006418	0.176837	-0.152908
515	2.705315	-0.210358	-0.090674	0.465977	0.183735	0.156095	-0.069533
516	2.608438	-0.181957	-0.090674	1.090195	0.139406	0.260614	-0.152908
517	2.593533	-0.199708	-0.090674	-0.178378	-0.126569	-1.004317	-0.403032

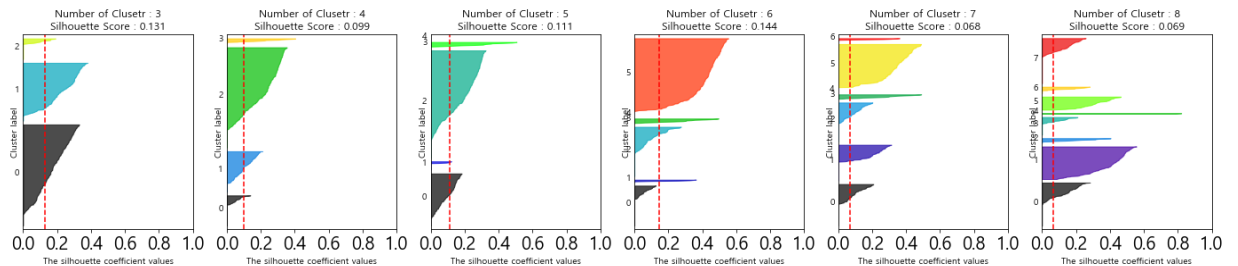
518 rows × 7 columns

In [276]:

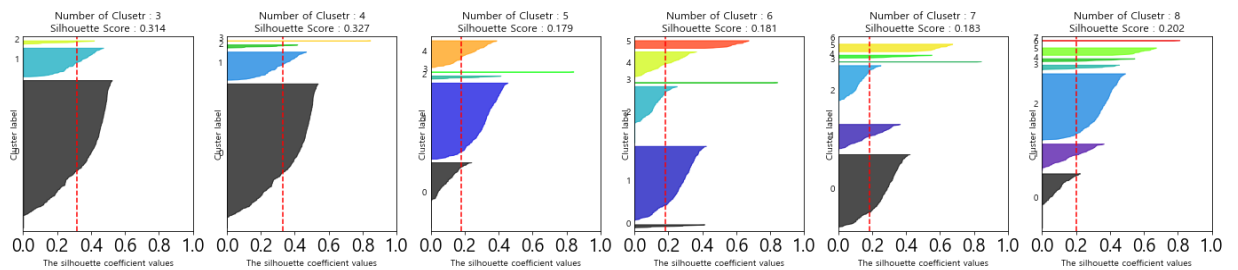
```
visualize_silhouette([3,4,5,6,7,8], df_ano_scale)
```



```
In [277]: GMM_silhouette([3,4,5,6,7,8], df_ano_scale)
```



```
In [278]: Agg_silhouette([3,4,5,6,7,8], df_ano_scale)
```



## 5. 감염지표 군집화

```
In [279]: # col = ['확진자수', '확진자수증감률', '감염지수', '폐업수', '폐업증감률', '카드매출액', '매출증감률', '카드매출건수', '카드매출건수증감률', '18시 전 카드매출액', '18시 후 카드매출액', '18시 후 카드매출액증감률', '유동인구', '유동인구증감률', '이동거리']

col = ['확진자수', '확진자수증감률', '감염지수']
df_infection = df[col]

df_infection = df_infection[:518]

df_infection_scale = pd.DataFrame(ss.fit_transform(df_infection))
df_infection_scale.columns = col

df_infection_scale
```

```
Out [279]:
```

	확진자수	확진자수증감률	감염지수
0	-0.894051	0.048805	1.582144

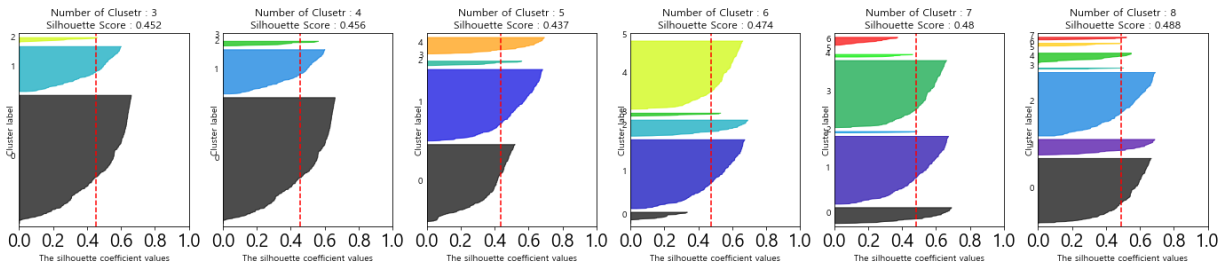
	확진자수	확진자수증감률	감염지수
<b>1</b>	-0.886599	-0.011548	1.582144
<b>2</b>	-0.931311	-0.544075	1.582144
<b>3</b>	-0.879147	-0.295562	1.582144
<b>4</b>	-0.923859	-0.505023	1.582144
<b>5</b>	-0.879147	-0.338165	1.582144
<b>6</b>	-0.879147	-0.047050	1.582144
<b>7</b>	-0.856790	0.165960	-0.834149
<b>8</b>	-0.737557	0.993152	-0.834149
<b>9</b>	-0.588515	16.305063	-0.834149
<b>10</b>	-0.804626	0.318618	-0.834149
<b>11</b>	-0.841886	3.716137	-0.834149
<b>12</b>	-0.864243	-0.086102	-0.834149
<b>13</b>	-0.871695	-0.139355	-0.834149
<b>14</b>	-0.886599	-0.331064	-1.187299
<b>15</b>	-0.849338	-0.394967	-1.187299
<b>16</b>	-0.886599	-0.497923	-1.187299
<b>17</b>	-0.849338	-0.313313	-1.187299
<b>18</b>	-0.804626	-0.039950	-1.187299
<b>19</b>	-0.804626	0.126909	-1.187299
<b>20</b>	-0.871695	-0.189057	-1.187299
<b>21</b>	-0.886599	-0.189057	-0.611106
<b>22</b>	-0.886599	-0.348815	-0.611106
<b>23</b>	-0.834434	0.226313	-0.611106
<b>24</b>	-0.834434	-0.125154	-0.611106
<b>25</b>	-0.826982	-0.252960	-0.611106
<b>26</b>	-0.819530	-0.231659	-0.611106
<b>27</b>	-0.767365	0.432224	-0.611106
<b>28</b>	-0.767365	0.758840	-0.443824
<b>29</b>	-0.789722	0.581331	-0.443824
...	...	...	...
<b>488</b>	1.743994	-0.057701	0.206716
<b>489</b>	1.222346	-0.121604	0.206716
<b>490</b>	1.356484	0.045254	0.281064
<b>491</b>	1.453362	0.009753	0.281064
<b>492</b>	3.405813	0.006203	0.281064
<b>493</b>	3.167345	0.041704	0.281064

	확진자수	확진자수증감률	감염지수
494	2.817096	-0.029299	0.281064
495	2.861809	-0.039950	0.281064
496	2.861809	0.080756	0.281064
497	2.071886	-0.079002	0.429759
498	2.176216	-0.082552	0.429759
499	3.815678	-0.157106	0.429759
500	2.936330	-0.210358	0.429759
501	3.308935	-0.142905	0.429759
502	3.234414	-0.153555	0.429759
503	2.958686	-0.178407	0.429759
504	2.191120	-0.174856	0.002261
505	2.004817	-0.210358	0.002261
506	3.569759	-0.206808	0.002261
507	2.787288	-0.203258	0.002261
508	2.936330	-0.221009	0.002261
509	2.533916	-0.249410	0.002261
510	2.712767	-0.210358	0.002261
511	1.609856	-0.256511	-0.090674
512	1.684377	-0.228109	-0.090674
513	3.338744	-0.206808	-0.090674
514	2.899070	-0.178407	-0.090674
515	2.705315	-0.210358	-0.090674
516	2.608438	-0.181957	-0.090674
517	2.593533	-0.199708	-0.090674

518 rows × 3 columns

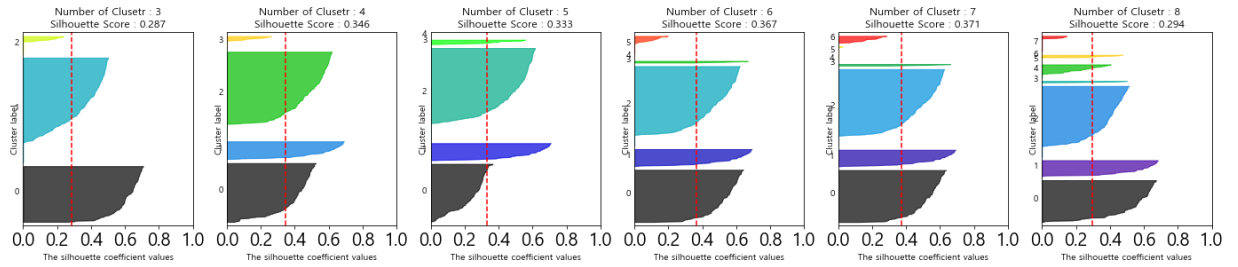
In [280]:

visualize\_silhouette([3,4,5,6,7,8], df\_infection\_scale)

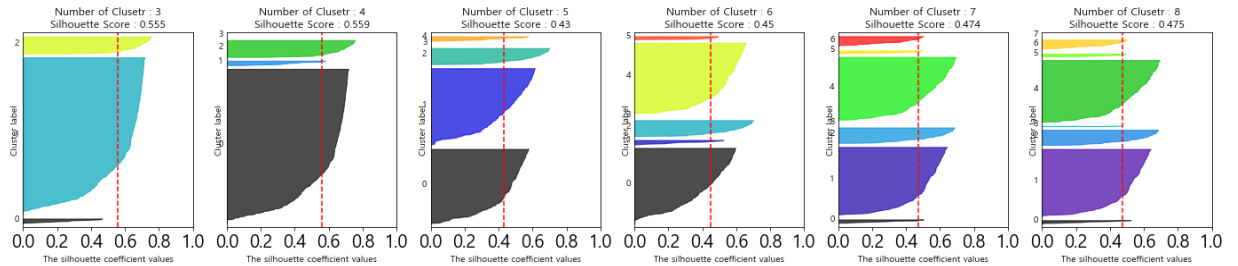


In [281]:

GMM\_silhouette([3,4,5,6,7,8], df\_infection\_scale)



```
In [282]: Agg_silhouette([3,4,5,6,7,8], df_infection_scale)
```



## 6. 소상공인지표 군집화 (+ 요인 기준)

```
In [283]: # col = ['확진자수', '확진자수증감률', '감염지수', '폐업수', '폐업증감률', '카드매출액', '매출증감률', '카드매출건수', '카드매출건수증감률', '18시 전 카드매출액', '18시 후 카드매출액', '18시 후 카드매출액증감률', '유동인구', '유동인구증감률', '이동거리']

col = ['카드매출액', '매출증감률', '카드매출건수', '카드매출건수증감률', '18시 전 카드매출액', '18시 후 카드매출액', '18시 후 카드매출액증감률']
df_plus = df[col]

df_plus = df_plus[:518]

df_plus_scale = pd.DataFrame(ss.fit_transform(df_plus))
df_plus_scale.columns = col

df_plus_scale
```

Out [283]:

	카드매출액	매출증감률	카드매출건수	카드매출건수증감률	18시 전 카드매출액	18시 후 카드매출액	18시 후 카드매출액증감률
0	-2.393347	-0.303886	-2.325570	-0.746495	-2.477514	-1.699593	-0.548756
1	-0.681780	-0.481203	-0.356203	-0.670421	-0.661122	-0.584453	-0.440736
2	-0.641508	-0.303886	-0.404988	-0.061829	-0.716422	-0.350114	-0.170685
3	-0.842868	-0.702850	-0.585614	-0.594347	-0.882320	-0.604631	-0.440736
4	-0.722052	-0.569862	-0.457359	-0.290051	-0.827020	-0.327646	-0.224695
5	-0.218650	-0.481203	0.053735	0.546764	-0.467574	0.356991	0.369417
6	-0.581099	0.050748	-0.661062	0.166394	-0.744071	-0.106128	0.207387

	카드매출 액	매출증감 률	카드매출 건수	카드매출건수 증감률	18시 전 카드 매출액	18시 후 카드 매출액	18시 후 카드매출 액증감률
7	-2.453755	-0.170899	-2.088770	0.394616	-2.645610	-1.537348	0.207387
8	-0.742188	-0.126569	-0.092212	0.242468	-0.744071	-0.541342	-0.008654
9	-0.842868	-0.303886	-0.678034	-0.366125	-0.909970	-0.554394	-0.332715
10	-0.621372	0.139406	-0.117607	0.546764	-0.716422	-0.263714	0.369417
11	-0.762324	-0.126569	-0.155004	0.318542	-0.909970	-0.269263	-0.008654
12	-0.178378	-0.037911	0.210436	0.090320	-0.439925	0.398580	-0.008654
13	-0.440147	0.050748	-0.334658	0.394616	-0.605823	-0.021184	0.045356
14	-2.252394	0.228065	-2.102586	-0.061829	-2.395240	-1.484310	0.045356
15	-0.802596	-0.126569	-0.257586	-0.290051	-0.771721	-0.661255	-0.224695
16	-0.621372	0.139406	-0.258313	0.470690	-0.688772	-0.349948	0.207387
17	-0.560963	-0.037911	-0.012481	0.090320	-0.716422	-0.145953	0.099366
18	-0.782460	-0.126569	-0.433510	-0.366125	-0.909970	-0.334309	-0.116674
19	-0.017289	0.050748	0.396293	0.166394	-0.274026	0.531027	0.099366
20	-0.359603	0.006418	-0.173421	0.166394	-0.522874	0.076933	0.045356
21	-2.312802	-0.170899	-1.866517	0.318542	-2.476570	-1.468671	-0.062664
22	-0.560963	0.139406	0.005866	0.242468	-0.495224	-0.592003	0.045356
23	-0.500555	0.050748	0.019772	0.242468	-0.522874	-0.320212	-0.008654
24	-0.339466	0.139406	0.201507	0.166394	-0.412275	-0.122420	-0.008654
25	-0.379739	0.316723	-0.021854	0.470690	-0.384625	-0.254748	0.045356
26	0.365296	0.228065	0.510555	0.090320	0.196019	0.647127	0.045356
27	-0.279058	0.006418	-0.162045	-0.061829	-0.412275	0.044069	-0.116674
28	-2.051033	0.316723	-1.674367	0.242468	-2.237156	-1.235069	0.315407
29	-0.158242	0.272394	0.091133	0.014246	-0.025179	-0.403916	0.207387
...	...	...	...	...	...	...	...
488	1.331828	-0.348216	1.407020	-0.137903	1.108459	1.527262	-0.116674
489	0.244480	-0.481203	-0.266779	-0.974718	0.555465	-0.455026	-1.304898
490	-1.567767	-0.170899	-1.599027	-0.670421	-1.545913	-1.313337	-0.224695
491	0.747882	-0.392545	0.759547	-0.061829	0.887261	0.288000	-0.170685
492	0.667337	-0.614191	0.828605	0.014246	0.721363	0.424364	-0.278705
493	0.405568	-0.968825	0.504112	-0.518273	0.527815	0.032592	-1.088858
494	0.284752	-0.614191	0.457851	-0.594347	0.334267	0.123826	-0.872817
495	0.647201	-0.525533	0.832856	-0.594347	0.555465	0.675989	-0.764797
496	0.103527	-0.215228	-0.031817	0.242468	0.278968	-0.280342	0.153376
497	-2.091305	-0.702850	-1.889170	-0.518273	-2.015958	-1.796605	-0.818807
498	-0.218650	-0.791508	0.183669	-0.670421	0.168369	-0.989972	-1.466929

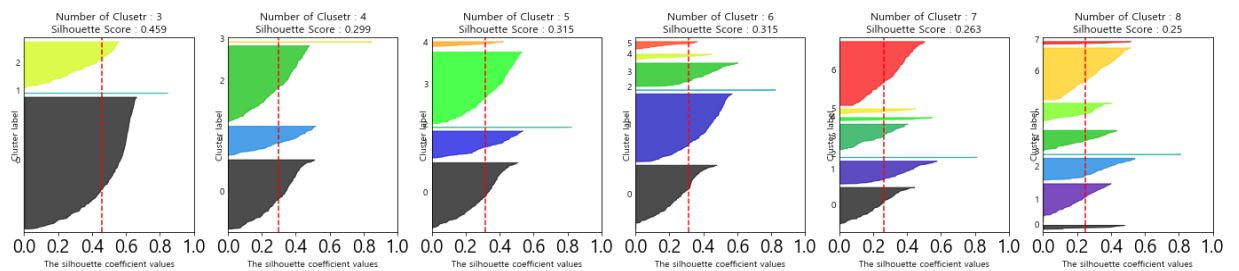


	카드매출 액	매출증감 률	카드매출 건수	카드매출건수 증감률	18시 전 카드 매출액	18시 후 카드 매출액	18시 후 카드매출 액증감률
499	-0.097834	-0.658520	0.191197	-0.746495	0.251318	-0.830332	-1.412919
500	-0.097834	-0.481203	0.200777	-0.366125	0.196019	-0.719759	-0.926827
501	-0.138106	-0.436874	0.026893	-0.518273	0.196019	-0.848134	-1.142868
502	0.445841	-0.215228	0.440825	-0.442199	0.776663	-0.379438	-1.142868
503	-0.198514	-0.348216	-0.222625	-0.290051	0.113070	-0.831873	-0.764797
504	-1.910081	0.139406	-1.692195	0.242468	-1.822410	-1.730915	0.045356
505	-0.117970	0.006418	-0.040306	-0.290051	0.306618	-1.008430	-0.062664
506	0.264616	0.228065	0.237987	0.014246	0.693713	-0.755515	0.045356
507	0.244480	0.228065	0.303183	0.090320	0.610764	-0.613105	0.099366
508	0.143799	0.183735	0.201664	0.166394	0.555465	-0.727908	0.099366
509	0.788154	0.183735	0.650824	0.166394	1.136109	-0.077214	0.315407
510	-0.138106	-0.037911	-0.188402	0.014246	0.140719	-0.665545	0.153376
511	-2.191986	-0.481203	-1.903596	-0.366125	-2.071257	-2.008757	-0.548756
512	0.163936	0.183735	0.345525	0.394616	0.583115	-0.745606	0.315407
513	0.586793	0.183735	0.286545	0.014246	1.136109	-0.650981	0.099366
514	0.365296	0.006418	0.228376	-0.137903	0.887261	-0.800168	-0.332715
515	0.465977	0.183735	0.248219	0.014246	0.942561	-0.679237	-0.008654
516	1.090195	0.139406	0.605825	-0.137903	1.523205	-0.012587	-0.008654
517	-0.178378	-0.126569	-0.347440	-0.213977	0.140719	-0.802671	-0.224695

518 rows × 7 columns

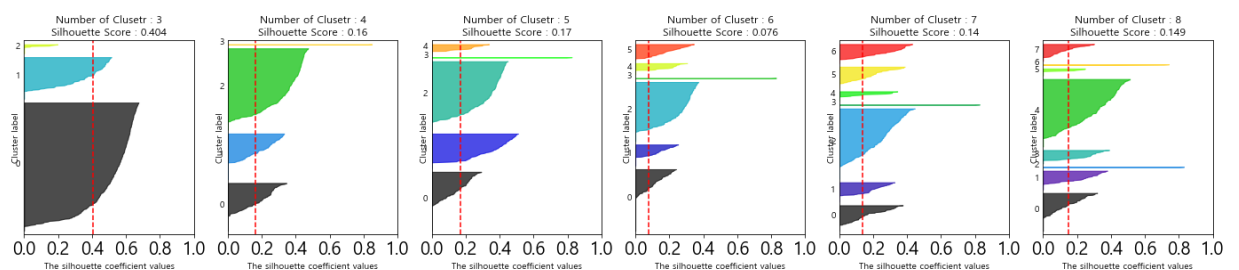
In [284]:

visualize\_silhouette([3,4,5,6,7,8], df\_plus\_scale)

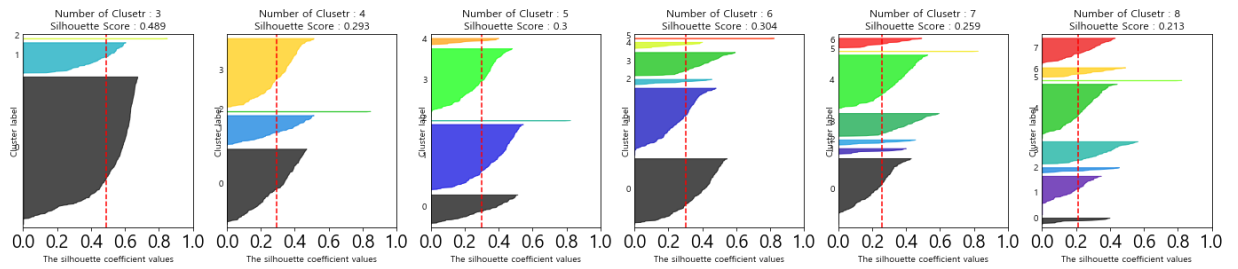


In [285]:

GMM\_silhouette([3,4,5,6,7,8], df\_plus\_scale)



```
In [286]: Agg_silhouette([3,4,5,6,7,8], df_plus_scale)
```



## 7. 인구가동량 지표 군집화

```
In [287]: # col = ['확진자수', '확진자수증감률', '감염지수', '폐업수', '폐업증감률', '카드매출액', '매출증감률', '카드매출건수', '카드매출건수증감률', '18시 전 카드매출액', '18시 후 카드매출액', '18시 후 카드매출액증감률', '유동인구', '유동인구증감률', '이동거리']

col = ['유동인구', '유동인구증감률', '이동거리']
df_pop = df[col]

df_pop = df_pop[:518]

df_pop_scale = pd.DataFrame(ss.fit_transform(df_pop))
df_pop_scale.columns = col

df_pop_scale
```

```
Out[287]:
```

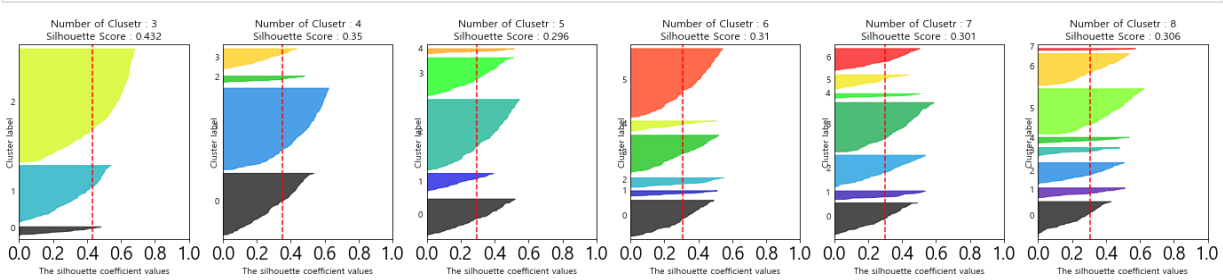
	유동인구	유동인구증감률	이동거리
0	-2.723383	-1.903780	-3.201881
1	-0.691159	-0.736532	-1.476383
2	-0.674272	0.263967	-1.563565
3	-0.839437	-0.486407	-1.588589
4	-0.695795	-0.236282	-1.416220
5	-0.459137	0.930966	-0.774068
6	-1.411448	0.514091	-1.827342
7	-2.268955	1.181090	-2.517430
8	-0.426327	0.347342	-1.100277
9	-1.002048	-0.653157	-1.597335
10	-0.515285	0.514091	-1.166553
11	-0.552757	0.180592	-1.114059
12	-0.464928	-0.069533	-0.544988
13	-1.328184	0.097217	-1.476672

	유동인구	유동인구증감률	이동거리
<b>14</b>	-2.422617	-0.486407	-2.507333
<b>15</b>	-0.511394	-0.236282	-1.105887
<b>16</b>	-0.549703	0.764216	-1.236393
<b>17</b>	-0.284221	0.263967	-1.024655
<b>18</b>	-0.842587	-0.569782	-1.302102
<b>19</b>	-0.141432	0.430716	-0.258468
<b>20</b>	-1.037741	0.514091	-1.203264
<b>21</b>	-2.028874	0.930966	-2.007976
<b>22</b>	-0.295957	0.263967	-0.892122
<b>23</b>	-0.239927	0.430716	-0.973246
<b>24</b>	-0.155260	0.097217	-0.871539
<b>25</b>	-0.392231	0.680841	-1.005775
<b>26</b>	-0.180903	-0.152908	-0.300708
<b>27</b>	-1.011720	0.013842	-1.127548
<b>28</b>	-1.880289	0.263967	-1.939445
<b>29</b>	-0.207890	0.097217	-0.830485
...	...	...	...
<b>488</b>	1.565404	-0.069533	1.891970
<b>489</b>	-0.189327	-0.819906	0.688269
<b>490</b>	-1.333490	-1.236781	-0.530197
<b>491</b>	1.321163	0.180592	0.775094
<b>492</b>	1.376369	0.764216	0.627708
<b>493</b>	1.121108	-0.152908	0.425805
<b>494</b>	0.909343	-0.653157	0.369486
<b>495</b>	1.048622	-0.653157	1.194453
<b>496</b>	-0.330011	-0.319657	0.125644
<b>497</b>	-1.508500	-0.403032	-0.793160
<b>498</b>	0.374470	-1.236781	0.000278
<b>499</b>	0.347679	-1.320156	-0.191092
<b>500</b>	0.203236	-1.236781	-0.167967
<b>501</b>	0.092271	-1.070031	-0.136517
<b>502</b>	0.263252	-1.070031	0.643496
<b>503</b>	-0.739233	-0.736532	-0.355175
<b>504</b>	-1.812349	-0.653157	-1.044457
<b>505</b>	0.172608	-0.319657	-0.072505
<b>506</b>	0.333431	-0.069533	-0.136599

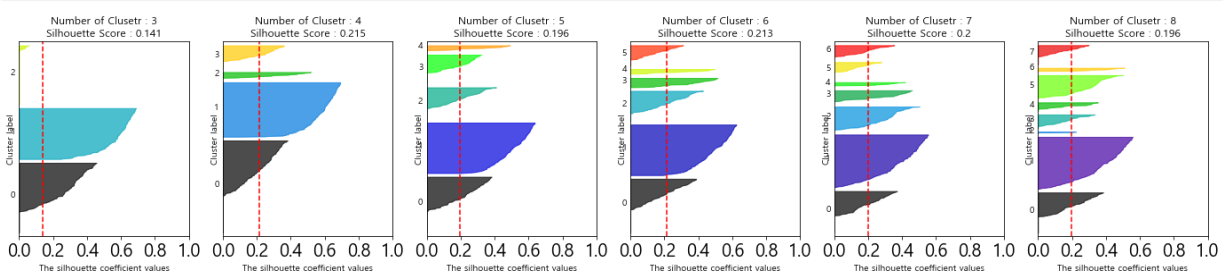
	유동인구	유동인구증감률	이동거리
507	0.259261	0.013842	-0.094832
508	0.153115	0.013842	-0.016845
509	0.317857	0.013842	0.743321
510	-0.814211	-0.236282	-0.298980
511	-2.002721	-0.486407	-1.076431
512	0.073591	-0.236282	-0.093127
513	-0.005923	-0.569782	-0.374328
514	0.176837	-0.152908	0.075075
515	0.156095	-0.069533	0.231357
516	0.260614	-0.152908	1.025854
517	-1.004317	-0.403032	-0.363758

518 rows × 3 columns

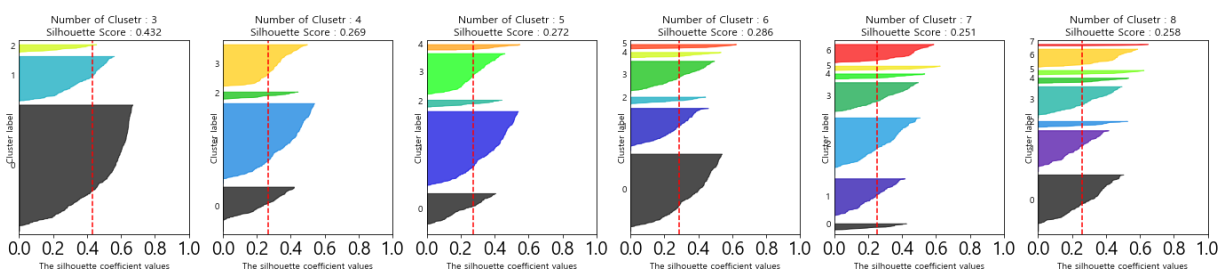
```
In [288]: visualize_silhouette([3,4,5,6,7,8], df_pop_scale)
```



```
In [289]: GMM_silhouette([3,4,5,6,7,8], df_pop_scale)
```



```
In [290]: Agg_silhouette([3,4,5,6,7,8], df_pop_scale)
```



```
In [ ]:
```

In [ ]:

## k-means 가 평균적 실루엣 점수가 높음

최종적으로 7, 5, 7 선택

In [291]:

```
df_1 = df.copy()

df_1 = df_1[:518]

KM = KMeans(n_clusters = 7, random_state = 777)

df_1['cluster'] = KM.fit_predict(df_all_scale)

df_1.to_csv('E:WKmeans_1.csv', encoding = 'UTF-8-SIG')
```

In [292]:

```
df_2 = df.copy()

df_2 = df_2[:518]

KM = KMeans(n_clusters = 5, random_state = 777)

df_2['cluster'] = KM.fit_predict(df_nonmute_scale)

df_2.to_csv('E:WKmeans_2.csv', encoding = 'UTF-8-SIG')
```

In [293]:

```
df_3 = df.copy()

df_3 = df_3[:518]

KM = KMeans(n_clusters = 7, random_state = 777)

df_3['cluster'] = KM.fit_predict(df_mute_scale)

df_3.to_csv('E:WKmeans_3.csv', encoding = 'UTF-8-SIG')
```

In [294]:

```
df_4 = df.copy()

df_4 = df_4[:518]

Agg = AgglomerativeClustering(n_clusters = 7, linkage = 'ward')

df_4['cluster'] = Agg.fit_predict(df_infection_scale)
```

```
df_4.to_csv('E:W감염지표군집화.csv', encoding = 'UTF-8-SIG')
```

In [295]:

```
df_5 = df.copy()

df_5 = df_5[:518]

KM = KMeans(n_clusters = 6, random_state = 777)

df_5['cluster'] = KM.fit_predict(df_plus_scale)

df_5.to_csv('E:W소상공인지표군집화.csv', encoding = 'UTF-8-SIG')
```

In [296]:

```
df_6 = df.copy()

df_6 = df_6[:518]

KM = KMeans(n_clusters = 7, random_state = 777)

df_6['cluster'] = KM.fit_predict(df_pop_scale)

df_6.to_csv('E:W인구이동량지표군집화.csv', encoding = 'UTF-8-SIG')
```

In [ ]:

In [ ]:

In [297]:

```
df_1.groupby('cluster').mean()[['확진자수', '폐업수', '카드매출액', '18  
시 후 카드매출액', '카드매출건수', '유동인구']]
```

Out[297]:

	확진자수	폐업수	카드매출액	18시 후 카드매출액	카드매출건수	유동인구
cluster						
0	324.192308	88.525641	2.626667e+11	6.909094e+10	8.591269e+06	1.282768e+07
1	60.292453	37.094340	2.253208e+11	7.056211e+10	8.125613e+06	1.165303e+07
2	93.227092	83.490040	2.947371e+11	9.262950e+10	9.905892e+06	1.494143e+07
3	75.500000	82.500000	3.035000e+11	9.540487e+10	9.796030e+06	1.499050e+07
4	72.428571	56.285714	2.587857e+11	8.660383e+10	9.282100e+06	1.282607e+07
5	86.400000	76.666667	2.752667e+11	8.479200e+10	9.316797e+06	1.394100e+07
6	138.403846	7.365385	1.618224e+11	5.053038e+10	6.176374e+06	8.594884e+06

In [298]:

```
df_2.groupby('cluster').mean()[['확진자수', '폐업수', '카드매출액', '18
```

```
시 후 카드매출액', '카드매출건수', '유동인구']]
```

```
Out[298]:
```

	확진자수	폐업수	카드매출액	18시 후 카드매출액	카드매출건수	유동인구
cluster						
0	399.907407	91.759259	2.647593e+11	6.741185e+10	8.582854e+06	1.276433e+07
1	99.724891	86.305677	2.976812e+11	9.351376e+10	9.992795e+06	1.518939e+07
2	109.659341	5.516484	1.754480e+11	5.700823e+10	6.693772e+06	9.644762e+06
3	69.500000	62.000000	2.626429e+11	8.626783e+10	9.365254e+06	1.297141e+07
4	71.938462	63.000000	2.539692e+11	7.670012e+10	8.740051e+06	1.249813e+07

```
In [299]: df_3.groupby('cluster').mean()[['확진자수', '폐업수', '카드매출액', '18시 후 카드매출액', '카드매출건수', '유동인구']]
```

```
Out[299]:
```

	확진자수	폐업수	카드매출액	18시 후 카드매출액	카드매출건수	유동인구
cluster						
0	169.173554	59.826446	2.330476e+11	6.990695e+10	8.154070e+06	1.203529e+07
1	116.615169	69.500000	2.701601e+11	8.283290e+10	9.185576e+06	1.363859e+07
2	75.500000	82.500000	3.035000e+11	9.540487e+10	9.796030e+06	1.499050e+07
3	125.666667	129.000000	2.743333e+11	8.341487e+10	9.738942e+06	1.462217e+07
4	46.250000	43.000000	2.505000e+11	8.665107e+10	9.060956e+06	1.273657e+07
5	70.521739	50.304348	2.649565e+11	8.359868e+10	8.957245e+06	1.315169e+07
6	46.000000	73.000000	2.190000e+11	7.051311e+10	8.013402e+06	1.089093e+07

```
In [ ]:
```

```
In [ ]:
```

```
In [300]: df_new = pd.read_csv('E:\Wcd.csv', parse_dates = ['날짜'], encoding = 'cp949')
```

```
In [301]: df_new.head()
```

```
Out[301]:
```

	날짜	전체변수	파생변수만	파생변수없이	감염지표	소상공인	인구이동
0	2020-03-01	6	0	2	6	3	1
1	2020-03-02	0	0	4	6	0	5
2	2020-03-03	1	1	4	6	0	2
3	2020-03-04	1	0	4	6	0	5
4	2020-03-05	1	1	4	6	0	2

```
In [302]: import copy
```

```
In [303]: df_agg = df[:518].merge(df_new, how = 'inner', on = df_new.index)
```

## 전체변수 군집분포

```
In [304]: plt.figure(figsize =(25, 10))

plt.scatter(df_new['날짜'], df_new['전체변수'])
```

Out[304]: <matplotlib.collections.PathCollection at 0xe826759518>

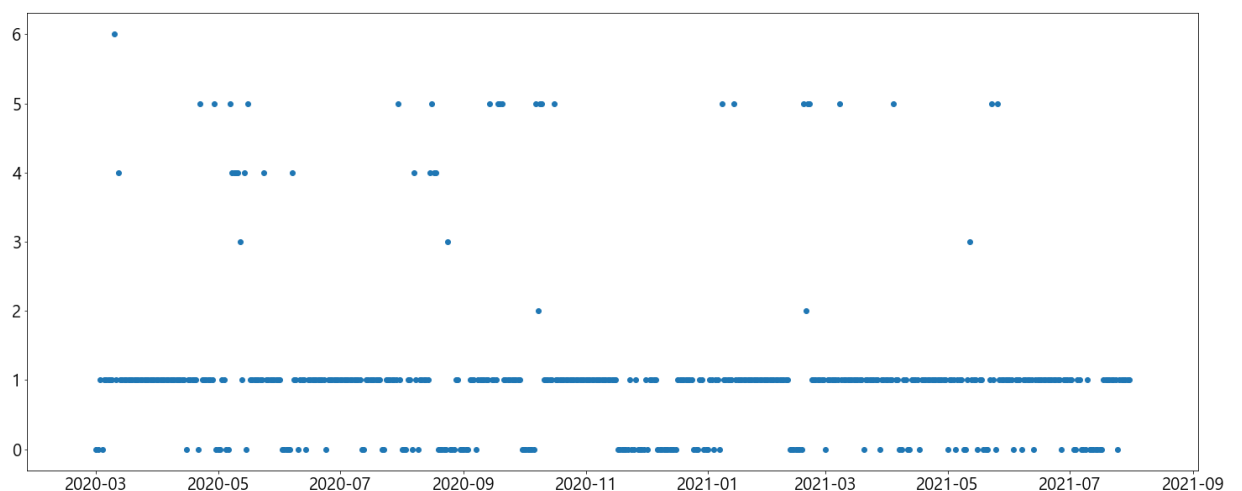


## 파생변수only 군집분포

```
In [305]: plt.figure(figsize =(25, 10))

plt.scatter(df_new['날짜'], df_new['파생변수만'])
```

Out[305]: <matplotlib.collections.PathCollection at 0xe826727278>



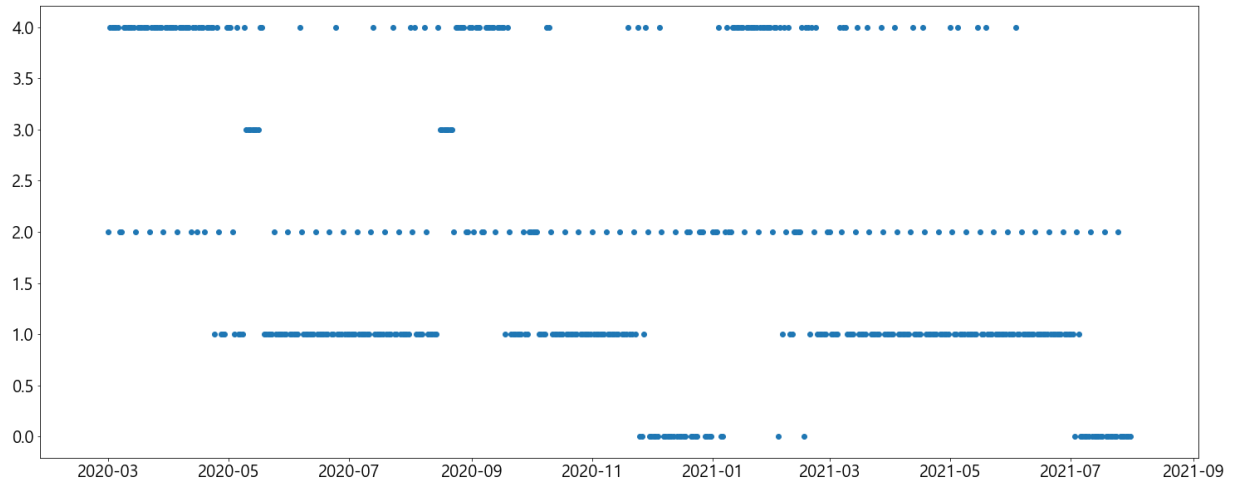
## 파생변수 없이 군집분포

```
In [306]: plt.figure(figsize =(25, 10))
```



```
plt.scatter(df_new['날짜'], df_new['파생변수없이'])
```

Out[306]: <matplotlib.collections.PathCollection at 0xe8318af2b0>

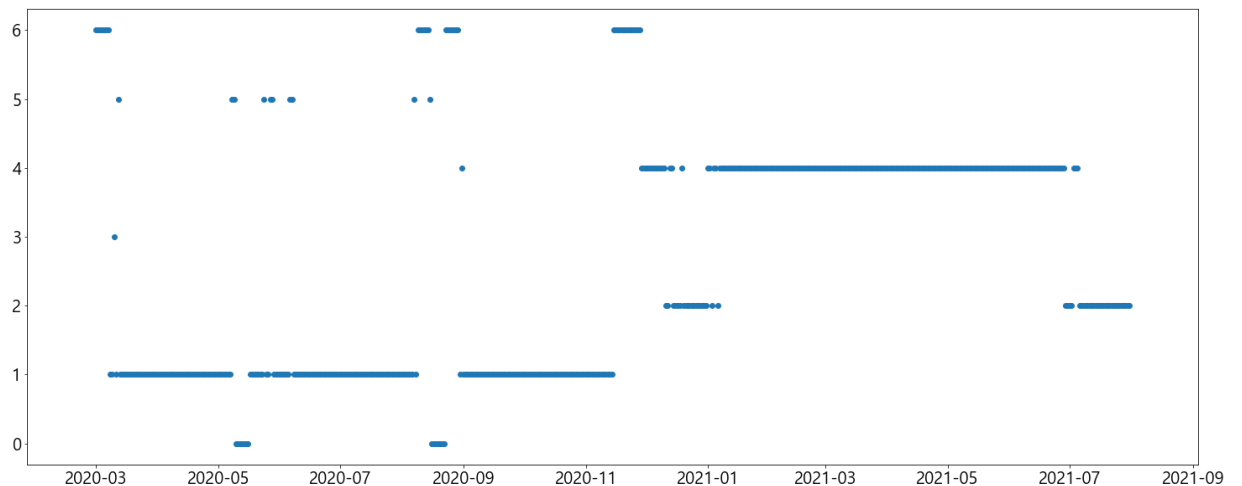


## 감염지표 군집분포

```
In [307]: plt.figure(figsize=(25, 10))

plt.scatter(df_new['날짜'], df_new['감염지표'])
```

Out[307]: <matplotlib.collections.PathCollection at 0xe8253c5860>

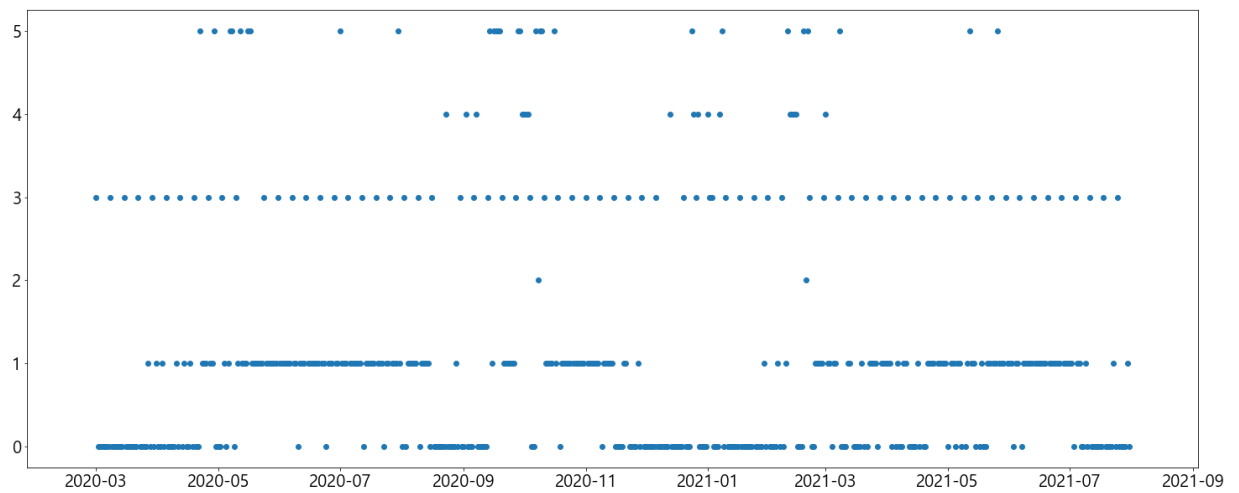


## 소상공인 + 군집분포

```
In [308]: plt.figure(figsize=(25, 10))

plt.scatter(df_new['날짜'], df_new['소상공인'])
```

Out[308]: <matplotlib.collections.PathCollection at 0xe831623358>

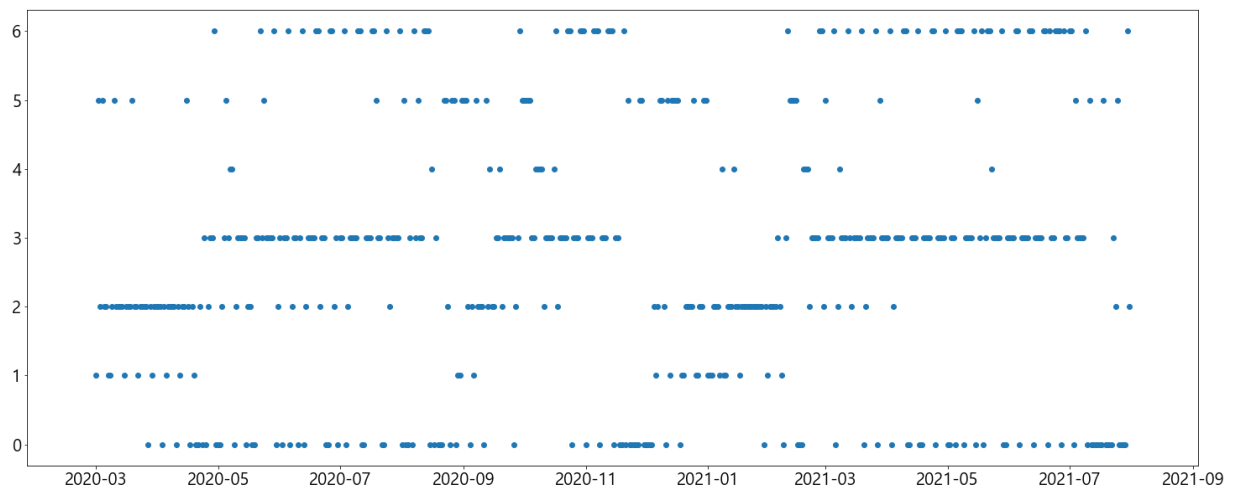


## 유동인구량 군집분포

```
In [309]: plt.figure(figsize=(25, 10))

plt.scatter(df_new['날짜'], df_new['인구이동'])
```

Out[309]: <matplotlib.collections.PathCollection at 0xe831282208>



```
In [310]: df_agg.groupby('인구이동').mean()[['확진자수', '폐업수', '카드매출액',
'18시 후 카드매출액', '카드매출건수', '유동인구']]
```

```
Out[310]:
```

	확진자수	폐업수	카드매출액	18시 후 카드매출액	카드매출건수	유동인구
인구이동						
0	163.466019	56.825243	2.594951e+11	7.839397e+10	8.851409e+06	1.294451e+07
1	137.666667	4.074074	1.654074e+11	4.987609e+10	6.117860e+06	8.009284e+06
2	96.180000	58.610000	2.364000e+11	7.050614e+10	8.298727e+06	1.184754e+07
3	108.421053	95.223684	2.895921e+11	8.910681e+10	9.856737e+06	1.524356e+07
4	72.647059	54.176471	2.677059e+11	8.435619e+10	9.055180e+06	1.345825e+07
5	160.500000	42.083333	1.987034e+11	5.898833e+10	7.123642e+06	1.012773e+07
6	128.802817	70.985915	3.124507e+11	1.005610e+11	1.024550e+07	1.528791e+07

## 종합

In [311]:

```
df_new2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 518 entries, 0 to 517
Data columns (total 7 columns):
날짜          518 non-null datetime64[ns]
전체변수      518 non-null int64
파생변수만    518 non-null int64
파생변수없이  518 non-null float64
감염지표      518 non-null float64
소상공인      518 non-null float64
인구이동      518 non-null float64
dtypes: datetime64[ns](1), float64(4), int64(2)
memory usage: 28.4 KB
```

In [312]:

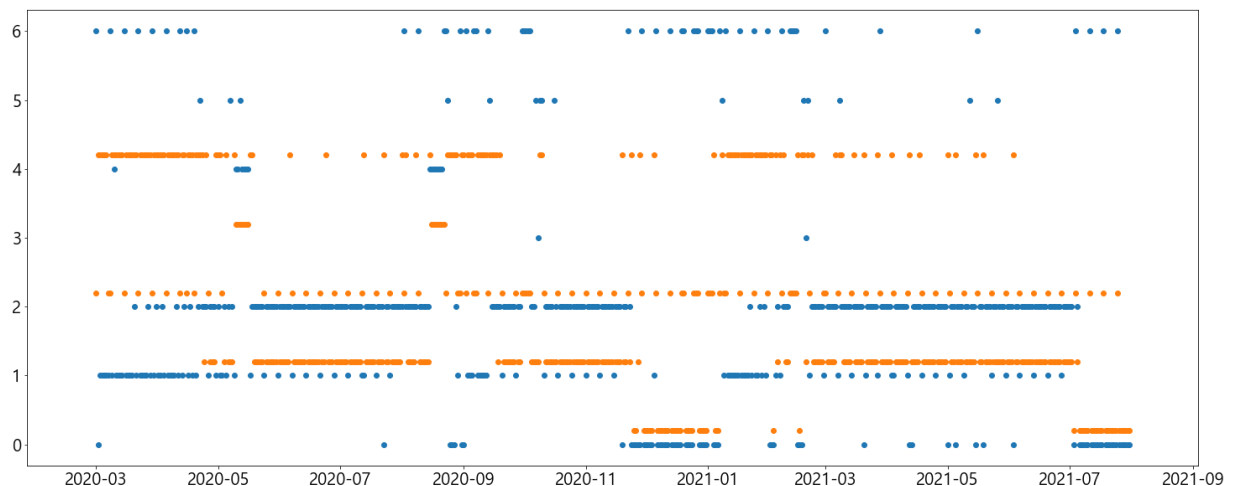
```
df_new2 = df_new
```

```
df_new2['파생변수없이'] = df_new2['파생변수없이'] + 0.2
df_new2['감염지표'] = df_new2['감염지표'] + 0.4
df_new2['소상공인'] = df_new2['소상공인'] + 0.6
df_new2['인구이동'] = df_new2['인구이동'] + 0.8
```

In [313]:

```
plt.figure(figsize=(25, 10))
plt.scatter(df_new['날짜'], df_new['전체변수'])
plt.scatter(df_new['날짜'], df_new['파생변수없이'])
```

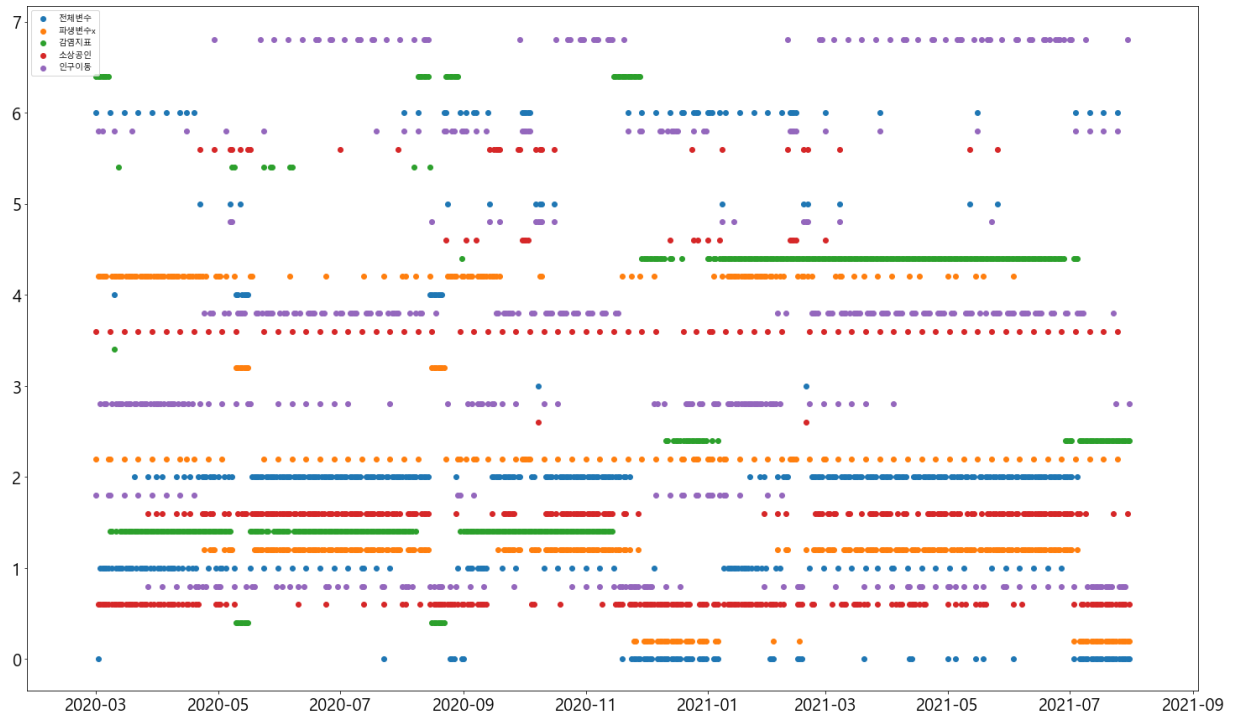
Out[313]: &lt;matplotlib.collections.PathCollection at 0xe827e6c2b0&gt;



In [314]:

```
plt.figure(figsize=(25, 15))
plt.scatter(df_new['날짜'], df_new['전체변수'], label='전체변수')
plt.scatter(df_new['날짜'], df_new['파생변수없이'], label='파생변수x')
plt.scatter(df_new['날짜'], df_new['감염지표'], label='감염지표')
plt.scatter(df_new['날짜'], df_new['소상공인'], label='소상공인')
plt.scatter(df_new['날짜'], df_new['인구이동'], label='인구이동')
plt.legend()
```

Out[314]: &lt;matplotlib.legend.Legend at 0xe829efa358&gt;



```
In [315]: df_new[df_new['파생변수없이'] == 2.2][['파생변수없이']].index
```

```
Out[315]: Int64Index([ 0,  6,  7, 14, 21, 28, 35, 42, 45, 49, 56, 63, 84,
 91, 98, 105, 112, 119, 126, 133, 140, 147, 154, 161, 175, 181,
182, 185, 189, 190, 196, 203, 210, 213, 214, 215, 216, 217, 224,
231, 238, 245, 252, 259, 266, 273, 280, 287, 293, 294, 299, 300,
301, 306, 307, 308, 312, 314, 315, 322, 329, 336, 343, 347, 348,
349, 350, 357, 364, 365, 371, 378, 385, 392, 399, 406, 413, 420,
427, 434, 441, 448, 455, 462, 469, 476, 483, 490, 497, 504, 511],
dtype='int64')
```

```
In [316]: for i in df_new[df_new['파생변수없이'] == 2.2][['파생변수없이']].index:
df_new['파생변수없이'][i] = -1

for i in df_new[df_new['소상공인'] == 3.6][['소상공인']].index:
df_new['소상공인'][i] = -1

df_new[df_new['파생변수없이'] == 2.2]['파생변수없이'] = -1
df_new[df_new['소상공인'] == 3.6]['소상공인'] = -1

for i in df_new[df_new['전체변수'] == 3][['전체변수']].index:
df_new['전체변수'][i] = -1

for i in df_new[df_new['소상공인'] == 2.6][['소상공인']].index:
df_new['소상공인'][i] = -1

for i in df_new[df_new['소상공인'] == 5.6][['소상공인']].index:
df_new['소상공인'][i] = -1

for i in df_new[df_new['소상공인'] == 4.6][['소상공인']].index:
```

```

df_new['소상공인'][i] = -1

for i in df_new[df_new['감염지표'] == 5.4][['감염지표']].index:
    df_new['감염지표'][i] = -1

for i in df_new[df_new['인구이동'] == 4.8][['인구이동']].index:
    df_new['인구이동'][i] = -1

for i in df_new[df_new['인구이동'] == 6.8][['인구이동']].index:
    df_new['인구이동'][i] = 3.8

for i in df_new[df_new['인구이동'] == 5.8][['인구이동']].index:
    df_new['인구이동'][i] = 2.8

for i in df_new[df_new['전체변수'] == 5][['전체변수']].index:
    df_new['전체변수'][i] = 6

for i in df_new[df_new['전체변수'] == 6][['전체변수']].index:
    df_new['전체변수'][i] = 5

for i in df_new[df_new['감염지표'] == 6.4][['감염지표']].index:
    df_new['감염지표'][i] = 3.4

```

In [317]:

```

plt.figure(figsize =(28, 15))
plt.scatter(df_new['날짜'], df_new['전체변수'], label = '전체변수')
plt.scatter(df_new['날짜'], df_new['파생변수없이'], label = '파생변수x')
plt.scatter(df_new['날짜'], df_new['감염지표'], label = '감염지표')
plt.scatter(df_new['날짜'], df_new['소상공인'], label = '소상공인')
plt.scatter(df_new['날짜'], df_new['인구이동'], label = '인구이동')
plt.axvline(x = '2020-05-07')
plt.axvline(x = '2020-05-19')
plt.axvline(x = '2020-08-08')
plt.axvline(x = '2020-09-15')
plt.axvline(x = '2020-11-15')
plt.axvline(x = '2021-02-21')
plt.axvline(x = '2021-07-06')

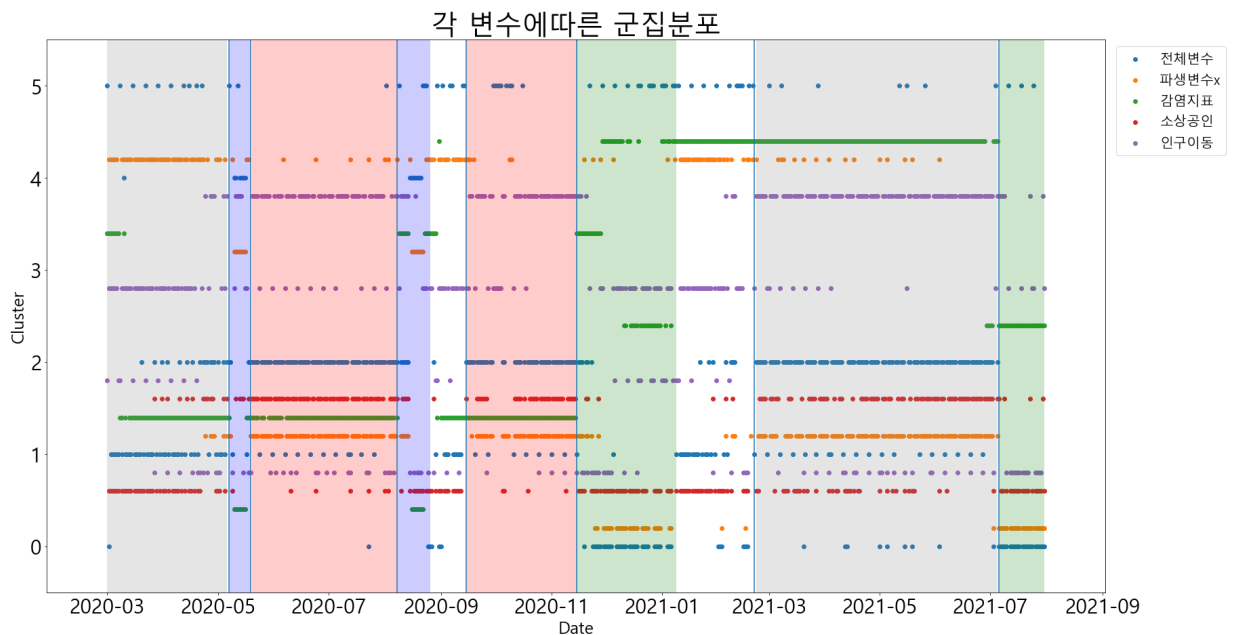
plt.fill_between(df_new['날짜'][67:80], y1 = -0.5, y2 = 5.5, facecolor =
'blue', alpha = 0.2)
plt.fill_between(df_new['날짜'][160:179], y1 = -0.5, y2 = 5.5, facecolor
= 'blue', alpha = 0.2)
plt.fill_between(df_new['날짜'][198:260], y1 = -0.5, y2 = 5.5, facecolor
= 'red', alpha = 0.2)

```

```
plt.fill_between(df_new['날짜'][80:161], y1 = -0.5, y2 = 5.5, facecolor = 'red', alpha = 0.2)
plt.fill_between(df_new['날짜'][259:315], y1 = -0.5, y2 = 5.5, facecolor = 'green', alpha = 0.2)
plt.fill_between(df_new['날짜'][492:], y1 = -0.5, y2 = 5.5, facecolor = 'green', alpha = 0.2)
plt.fill_between(df_new['날짜'][:67], y1 = -0.5, y2 = 5.5, facecolor = 'gray', alpha = 0.2)
plt.fill_between(df_new['날짜'][358:492], y1 = -0.5, y2 = 5.5, facecolor = 'gray', alpha = 0.2)

plt.ylim([-0.5,5.5])
plt.xticks(fontsize = 30)
plt.yticks(fontsize = 30)
plt.title('각 변수에 따른 군집분포', fontsize = 40)
plt.ylabel('Cluster', fontsize = 25)
plt.xlabel('Date', fontsize = 25)
plt.legend(fontsize = 20, bbox_to_anchor = (1.12, 1))
```

Out[317]: <matplotlib.legend.Legend at 0xe82652eba8>



In [ ]:

In [338]:

```
df_ind = pd.read_csv('E:\전체지표1.csv', encoding = 'cp949', parse_dates = ['날짜'])
```

In [339]:

```
df_ind
```

Out[339]:

	날짜	유통인구	이동거리	폐업수	카드매출 액	카드매출 건수	확진자수	감염지수	감염지표	지
0	2020-03-01	-2.723383	-3.201881	-1.278597	-2.393347	-2.325570	-0.894051	1.582144	0.344047	-C
1	2020-03-02	-0.691159	-1.476383	3.088007	-0.681780	-0.356203	-0.886599	1.582144	0.347773	-1
2	2020-03-03	-0.674272	-1.563565	0.479637	-0.641508	-0.404988	-0.931311	1.582144	0.325416	-C
3	2020-03-04	-0.839437	-1.588589	0.151176	-0.842868	-0.585614	-0.879147	1.582144	0.351499	-C
4	2020-03-05	-0.695795	-1.416220	0.209140	-0.722052	-0.457359	-0.923859	1.582144	0.329142	-C
5	2020-03-06	-0.459137	-0.774068	-0.157964	-0.218650	0.053735	-0.879147	1.582144	0.351499	C
6	2020-03-07	-1.411448	-1.827342	-1.278597	-0.581099	-0.661062	-0.879147	1.582144	0.351499	C
7	2020-03-08	-2.268955	-2.517430	-1.259276	-2.453755	-2.088770	-0.856790	-0.834149	-0.845469	-C
8	2020-03-09	-0.426327	-1.100277	0.267103	-0.742188	-0.092212	-0.737557	-0.834149	-0.785853	-C
9	2020-03-10	-1.002048	-1.597335	0.131855	-0.842868	-0.678034	-0.588515	-0.834149	-0.711332	-C
10	2020-03-11	-0.515285	-1.166553	0.015927	-0.621372	-0.117607	-0.804626	-0.834149	-0.819387	-C
11	2020-03-12	-0.552757	-1.114059	0.189818	-0.762324	-0.155004	-0.841886	-0.834149	-0.838017	-C
12	2020-03-13	-0.464928	-0.544988	-0.235249	-0.178378	0.210436	-0.864243	-0.834149	-0.849196	C
13	2020-03-14	-1.328184	-1.476672	-1.278597	-0.440147	-0.334658	-0.871695	-0.834149	-0.852922	C
14	2020-03-15	-2.422617	-2.507333	-1.278597	-2.252394	-2.102586	-0.886599	-1.187299	-1.036949	-C
15	2020-03-16	-0.511394	-1.105887	0.576243	-0.802596	-0.257586	-0.849338	-1.187299	-1.018319	-C
16	2020-03-17	-0.549703	-1.236393	0.518280	-0.621372	-0.258313	-0.886599	-1.187299	-1.036949	-C
17	2020-03-18	-0.284221	-1.024655	0.440995	-0.560963	-0.012481	-0.849338	-1.187299	-1.018319	-C
18	2020-03-19	-0.842587	-1.302102	0.344388	-0.782460	-0.433510	-0.804626	-1.187299	-0.995962	-C
19	2020-03-20	-0.141432	-0.258468	0.943347	-0.017289	0.396293	-0.804626	-1.187299	-0.995962	-C
20	2020-03-21	-1.037741	-1.203264	-1.278597	-0.359603	-0.173421	-0.871695	-1.187299	-1.029497	C

	날짜	유동인구	이동거리	폐업수	카드매출 액	카드매출 건수	확진자수	감염지수	감염지표	지
<b>21</b>	2020-03-22	-2.028874	-2.007976	-1.278597	-2.312802	-1.866517	-0.886599	-0.611106	-0.748852	-C
<b>22</b>	2020-03-23	-0.295957	-0.892122	0.576243	-0.560963	0.005866	-0.886599	-0.611106	-0.748852	-C
<b>23</b>	2020-03-24	-0.239927	-0.973246	0.344388	-0.500555	0.019772	-0.834434	-0.611106	-0.722770	-C
<b>24</b>	2020-03-25	-0.155260	-0.871539	0.189818	-0.339466	0.201507	-0.834434	-0.611106	-0.722770	-C
<b>25</b>	2020-03-26	-0.392231	-1.005775	0.383031	-0.379739	-0.021854	-0.826982	-0.611106	-0.719044	-C
<b>26</b>	2020-03-27	-0.180903	-0.300708	0.093212	0.365296	0.510555	-0.819530	-0.611106	-0.715318	C
<b>27</b>	2020-03-28	-1.011720	-1.127548	-1.278597	-0.279058	-0.162045	-0.767365	-0.611106	-0.689236	C
<b>28</b>	2020-03-29	-1.880289	-1.939445	-1.259276	-2.051033	-1.674367	-0.767365	-0.443824	-0.605595	-C
<b>29</b>	2020-03-30	-0.207890	-0.830485	0.672850	-0.158242	0.091133	-0.789722	-0.443824	-0.616773	-C
...	...	...	...	...	...	...	...	...	...	
<b>488</b>	2021-07-02	1.565404	1.891970	0.924026	1.331828	1.407020	1.743994	0.206716	0.975355	C
<b>489</b>	2021-07-03	-0.189327	0.688269	-1.220634	0.244480	-0.266779	1.222346	0.206716	0.714531	C
<b>490</b>	2021-07-04	-1.333490	-0.530197	-1.239955	-1.567767	-1.599027	1.356484	0.281064	0.818774	-C
<b>491</b>	2021-07-05	1.321163	0.775094	0.402352	0.747882	0.759547	1.453362	0.281064	0.867213	C
<b>492</b>	2021-07-06	1.376369	0.627708	0.498958	0.667337	0.828605	3.405813	0.281064	1.843438	C
<b>493</b>	2021-07-07	1.121108	0.425805	0.170497	0.405568	0.504112	3.167345	0.281064	1.724204	C
<b>494</b>	2021-07-08	0.909343	0.369486	0.479637	0.284752	0.457851	2.817096	0.281064	1.549080	-C
<b>495</b>	2021-07-09	1.048622	1.194453	0.035248	0.647201	0.832856	2.861809	0.281064	1.571436	C
<b>496</b>	2021-07-10	-0.330011	0.125644	-1.220634	0.103527	-0.031817	2.861809	0.281064	1.571436	C
<b>497</b>	2021-07-11	-1.508500	-0.793160	-1.220634	-2.091305	-1.889170	2.071886	0.429759	1.250822	-C
<b>498</b>	2021-07-12	0.374470	0.000278	0.440995	-0.218650	0.183669	2.176216	0.429759	1.302987	-C



	날짜	유동인구	이동거리	폐업수	카드매출 액	카드매출 건수	확진자수	감염지수	감염지표	cluster
499	2021-07-13	0.347679	-0.191092	-0.100001	-0.097834	0.191197	3.815678	0.429759	2.122718	C
500	2021-07-14	0.203236	-0.167967	0.035248	-0.097834	0.200777	2.936330	0.429759	1.683044	C
501	2021-07-15	0.092271	-0.136517	-0.042037	-0.138106	0.026893	3.308935	0.429759	1.869347	-C
502	2021-07-16	0.263252	0.643496	-0.196607	0.445841	0.440825	3.234414	0.429759	1.832086	C
503	2021-07-17	-0.739233	-0.355175	-1.239955	-0.198514	-0.222625	2.958686	0.429759	1.694222	C
504	2021-07-18	-1.812349	-1.044457	-1.278597	-1.910081	-1.692195	2.191120	0.002261	1.096690	-C
505	2021-07-19	0.172608	-0.072505	0.189818	-0.117970	-0.040306	2.004817	0.002261	1.003539	-C
506	2021-07-20	0.333431	-0.136599	-0.022716	0.264616	0.237987	3.569759	0.002261	1.786010	C
507	2021-07-21	0.259261	-0.094832	-0.409141	0.244480	0.303183	2.787288	0.002261	1.394774	C
508	2021-07-22	0.153115	-0.016845	-0.003394	0.143799	0.201664	2.936330	0.002261	1.469295	C
509	2021-07-23	0.317857	0.743321	0.286425	0.788154	0.650824	2.533916	0.002261	1.268089	C
510	2021-07-24	-0.814211	-0.298980	-1.259276	-0.138106	-0.188402	2.712767	0.002261	1.357514	C
511	2021-07-25	-2.002721	-1.076431	-1.220634	-2.191986	-1.903596	1.609856	-0.090674	0.759591	-C
512	2021-07-26	0.073591	-0.093127	0.073891	0.163936	0.345525	1.684377	-0.090674	0.796851	C
513	2021-07-27	-0.005923	-0.374328	3.493753	0.586793	0.286545	3.338744	-0.090674	1.624035	-1
514	2021-07-28	0.176837	0.075075	-0.235249	0.365296	0.228376	2.899070	-0.090674	1.404198	C
515	2021-07-29	0.156095	0.231357	0.325067	0.465977	0.248219	2.705315	-0.090674	1.307321	C
516	2021-07-30	0.260614	1.025854	0.189818	1.090195	0.605825	2.608438	-0.090674	1.258882	C
517	2021-07-31	-1.004317	-0.363758	-0.872851	-0.178378	-0.347440	2.593533	-0.090674	1.251430	C

518 rows × 20 columns

```
In [340]: col = ['확진자수', '감염지수', '감염지표', '폐업수', '카드매출액', '카드
```

```
매출건수', '소상공인지표', '유동인구', '이동거리', '인구이동량지표', '전체지표', '최종클러스터']
df_ind = df_ind[col]
```

```
In [341]: coll = ['확진자수', '감염지수', '감염지표', '폐업수', '카드매출액', '카드매출건수', '소상공인지표', '유동인구', '이동거리', '인구이동량지표', '전체지표']
```

```
In [342]: a = df_ind.groupby('최종클러스터').mean()
```

```
In [343]: a
```

```
Out[343]:
```

	확진자수	감염지수	감염지표	폐업수	카드매출액	카드매출건수	소상공인지표	유동인구	이동거리
최종클러스터									
1	0.007909	-0.277636	-0.134863	0.039416	0.056839	0.074143	0.013038	0.248660	0.163664
2	-0.497847	2.680009	1.091081	-0.163117	0.022312	0.294241	0.160697	-0.112581	0.167886
3	-0.777301	-0.145380	-0.461341	-0.030526	0.316094	0.529391	0.226634	0.333047	0.480019
4	1.595517	0.190717	0.893117	0.127208	-0.334369	-0.640476	-0.307315	-0.607234	-0.655800
5	-0.048819	-0.286417	-0.167618	-0.138945	-0.475292	-0.749985	-0.236847	-0.724021	-0.848416

```
In [ ]:
```