

## 2. 서울시 범죄 현황 분석

### 2.1 데이터 로드

```
In [ ]: import pandas as pd

df = pd.read_csv("/content/drive/MyDrive/스터디/데이터 주무르기/data/02_경찰청 서울특"
df
```

```
Out [ ]:
```

	구분	죄종	발생검거	건수
0	중부	살인	발생	1
1	중부	살인	검거	1
2	중부	강도	발생	3
3	중부	강도	검거	4
4	중부	강간,추행	발생	113
...	...	...	...	...
305	수서	강간,추행	검거	194
306	수서	절도	발생	1010
307	수서	절도	검거	475
308	수서	폭력	발생	1394
309	수서	폭력	검거	1200

310 rows × 4 columns

```
In [ ]: df["col"] = df["죄종"] + " " + df["발생검거"]
df
```

```
Out [ ]:
```

	구분	죄종	발생검거	건수	col
0	중부	살인	발생	1	살인 발생
1	중부	살인	검거	1	살인 검거
2	중부	강도	발생	3	강도 발생
3	중부	강도	검거	4	강도 검거
4	중부	강간,추행	발생	113	강간,추행 발생
...	...	...	...	...	...
305	수서	강간,추행	검거	194	강간,추행 검거
306	수서	절도	발생	1010	절도 발생
307	수서	절도	검거	475	절도 검거
308	수서	폭력	발생	1394	폭력 발생
309	수서	폭력	검거	1200	폭력 검거

310 rows × 5 columns

In [ ]:

```
# 데이터 프레임 재설정
df_1 = pd.pivot_table(data = df, values = "건수", index = "구분", columns="col")
df_1
```

Out[ ]:

col	강간 검거	강간 발생	강간,추 행 검거	강간,추 행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생
구분												
강남	NaN	NaN	433.0	513.0	10.0	12.0	3.0	5.0	1016.0	1941.0	1911.0	2283.0
강동	NaN	NaN	167.0	188.0	14.0	15.0	3.0	3.0	781.0	1640.0	1697.0	1942.0
강북	NaN	NaN	132.0	159.0	4.0	5.0	8.0	9.0	490.0	780.0	1625.0	1817.0
강서	NaN	NaN	222.0	265.0	5.0	6.0	9.0	10.0	1070.0	1689.0	2087.0	2445.0
관악	NaN	NaN	351.0	403.0	3.0	3.0	12.0	12.0	1137.0	2229.0	2355.0	2614.0
광진	NaN	NaN	224.0	255.0	4.0	4.0	2.0	2.0	874.0	1664.0	1407.0	1676.0
구로	NaN	NaN	142.0	197.0	5.0	5.0	6.0	6.0	1008.0	1763.0	1823.0	2204.0
금천	NaN	NaN	93.0	117.0	6.0	7.0	4.0	4.0	553.0	999.0	1310.0	1471.0
남대문	NaN	NaN	52.0	73.0	2.0	2.0	1.0	1.0	368.0	699.0	494.0	579.0
노원	NaN	NaN	148.0	169.0	3.0	3.0	3.0	4.0	646.0	1404.0	1781.0	2163.0
도봉	87.0	82.0	NaN	NaN	2.0	2.0	0.0	0.0	379.0	861.0	1022.0	1234.0
동대문	NaN	NaN	142.0	161.0	5.0	6.0	8.0	10.0	713.0	1440.0	1597.0	1784.0
동작	NaN	NaN	121.0	226.0	2.0	1.0	1.0	1.0	592.0	1341.0	1370.0	1631.0
마포	NaN	NaN	287.0	348.0	3.0	4.0	8.0	9.0	679.0	1483.0	1557.0	1844.0
방배	50.0	63.0	NaN	NaN	1.0	1.0	1.0	1.0	159.0	348.0	365.0	423.0
서대문	NaN	NaN	149.0	159.0	1.0	2.0	3.0	3.0	524.0	1077.0	1029.0	1292.0
서부	NaN	NaN	71.0	76.0	2.0	3.0	3.0	3.0	322.0	532.0	667.0	748.0
서초	NaN	NaN	253.0	333.0	3.0	5.0	4.0	4.0	801.0	1673.0	1415.0	1750.0

col 구분	강간 검거	강간 발생	강간,추 행 검거	강간,추 행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생
성동	NaN	NaN	98.0	115.0	3.0	3.0	5.0	5.0	565.0	1016.0	1103.0	1223.0
성북	77.0	95.0	NaN	NaN	2.0	2.0	0.0	0.0	233.0	439.0	610.0	672.0
송파	NaN	NaN	235.0	287.0	17.0	13.0	6.0	6.0	1006.0	2429.0	2280.0	2675.0
수서	NaN	NaN	194.0	195.0	1.0	2.0	2.0	1.0	475.0	1010.0	1200.0	1394.0
양천	90.0	109.0	NaN	NaN	3.0	3.0	5.0	5.0	718.0	1517.0	1270.0	1582.0
영등포	NaN	NaN	252.0	315.0	5.0	6.0	4.0	7.0	953.0	2188.0	2157.0	2701.0
용산	NaN	NaN	186.0	230.0	6.0	6.0	3.0	5.0	534.0	1111.0	1327.0	1617.0
은평	NaN	NaN	74.0	95.0	1.0	1.0	1.0	1.0	434.0	839.0	956.0	1092.0
종로	NaN	NaN	93.0	115.0	4.0	4.0	6.0	9.0	284.0	602.0	792.0	964.0
종암	44.0	47.0	NaN	NaN	0.0	1.0	4.0	4.0	347.0	549.0	630.0	758.0
중앙	NaN	NaN	153.0	183.0	6.0	5.0	9.0	10.0	889.0	1506.0	1724.0	2022.0
중부	NaN	NaN	76.0	113.0	4.0	3.0	1.0	1.0	495.0	943.0	830.0	997.0
혜화	NaN	NaN	53.0	77.0	3.0	3.0	0.0	0.0	310.0	581.0	628.0	747.0

In [ ]:

```
# 결측값 0으로 처리
df_1 = df_1.fillna(0)
df_1
```

Out[ ]:

col 구분	강간 검거	강간 발생	강간,추 행 검거	강간,추 행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생
강남	0.0	0.0	433.0	513.0	10.0	12.0	3.0	5.0	1016.0	1941.0	1911.0	2283.0
강동	0.0	0.0	167.0	188.0	14.0	15.0	3.0	3.0	781.0	1640.0	1697.0	1942.0
강북	0.0	0.0	132.0	159.0	4.0	5.0	8.0	9.0	490.0	780.0	1625.0	1817.0
강서	0.0	0.0	222.0	265.0	5.0	6.0	9.0	10.0	1070.0	1689.0	2087.0	2445.0

coi	강간 검거	강간 발생	강간,추 행 검거	강간,추 행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생
구 분												
관악	0.0	0.0	351.0	403.0	3.0	3.0	12.0	12.0	1137.0	2229.0	2355.0	2614.0
광진	0.0	0.0	224.0	255.0	4.0	4.0	2.0	2.0	874.0	1664.0	1407.0	1676.0
구로	0.0	0.0	142.0	197.0	5.0	5.0	6.0	6.0	1008.0	1763.0	1823.0	2204.0
금천	0.0	0.0	93.0	117.0	6.0	7.0	4.0	4.0	553.0	999.0	1310.0	1471.0
남대문	0.0	0.0	52.0	73.0	2.0	2.0	1.0	1.0	368.0	699.0	494.0	579.0
노원	0.0	0.0	148.0	169.0	3.0	3.0	3.0	4.0	646.0	1404.0	1781.0	2163.0
도봉	87.0	82.0	0.0	0.0	2.0	2.0	0.0	0.0	379.0	861.0	1022.0	1234.0
동대문	0.0	0.0	142.0	161.0	5.0	6.0	8.0	10.0	713.0	1440.0	1597.0	1784.0
동작	0.0	0.0	121.0	226.0	2.0	1.0	1.0	1.0	592.0	1341.0	1370.0	1631.0
마포	0.0	0.0	287.0	348.0	3.0	4.0	8.0	9.0	679.0	1483.0	1557.0	1844.0
방배	50.0	63.0	0.0	0.0	1.0	1.0	1.0	1.0	159.0	348.0	365.0	423.0
서대문	0.0	0.0	149.0	159.0	1.0	2.0	3.0	3.0	524.0	1077.0	1029.0	1292.0
서부	0.0	0.0	71.0	76.0	2.0	3.0	3.0	3.0	322.0	532.0	667.0	748.0
서초	0.0	0.0	253.0	333.0	3.0	5.0	4.0	4.0	801.0	1673.0	1415.0	1750.0
성동	0.0	0.0	98.0	115.0	3.0	3.0	5.0	5.0	565.0	1016.0	1103.0	1223.0
성북	77.0	95.0	0.0	0.0	2.0	2.0	0.0	0.0	233.0	439.0	610.0	672.0
송파	0.0	0.0	235.0	287.0	17.0	13.0	6.0	6.0	1006.0	2429.0	2280.0	2675.0
수서	0.0	0.0	194.0	195.0	1.0	2.0	2.0	1.0	475.0	1010.0	1200.0	1394.0
양천	90.0	109.0	0.0	0.0	3.0	3.0	5.0	5.0	718.0	1517.0	1270.0	1582.0
영등포	0.0	0.0	252.0	315.0	5.0	6.0	4.0	7.0	953.0	2188.0	2157.0	2701.0

col	강간 검거	강간 발생	강간,추 행 검거	강간,추 행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생
구분												
용산	0.0	0.0	186.0	230.0	6.0	6.0	3.0	5.0	534.0	1111.0	1327.0	1617.0
은평	0.0	0.0	74.0	95.0	1.0	1.0	1.0	1.0	434.0	839.0	956.0	1092.0
종로	0.0	0.0	93.0	115.0	4.0	4.0	6.0	9.0	284.0	602.0	792.0	964.0
종암	44.0	47.0	0.0	0.0	0.0	1.0	4.0	4.0	347.0	549.0	630.0	758.0
종량	0.0	0.0	153.0	183.0	6.0	5.0	9.0	10.0	889.0	1506.0	1724.0	2022.0
중부	0.0	0.0	76.0	113.0	4.0	3.0	1.0	1.0	495.0	943.0	830.0	997.0
혜화	0.0	0.0	53.0	77.0	3.0	3.0	0.0	0.0	310.0	581.0	628.0	747.0

In [ ]:

```
# 강간추행 통합
df_1["강간,추행 검거"] = df_1["강간,추행 검거"] + df_1["강간 검거"]
df_1["강간,추행 발생"] = df_1["강간,추행 발생"] + df_1["강간 발생"]
df_1 = df_1.drop(["강간 검거","강간 발생"],axis = 1)
df_1
```

Out[ ]:

col	강간,추행 검거	강간,추행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발 생
구분										
강남	433.0	513.0	10.0	12.0	3.0	5.0	1016.0	1941.0	1911.0	2283.0
강동	167.0	188.0	14.0	15.0	3.0	3.0	781.0	1640.0	1697.0	1942.0
강북	132.0	159.0	4.0	5.0	8.0	9.0	490.0	780.0	1625.0	1817.0
강서	222.0	265.0	5.0	6.0	9.0	10.0	1070.0	1689.0	2087.0	2445.0
관악	351.0	403.0	3.0	3.0	12.0	12.0	1137.0	2229.0	2355.0	2614.0
광진	224.0	255.0	4.0	4.0	2.0	2.0	874.0	1664.0	1407.0	1676.0
구로	142.0	197.0	5.0	5.0	6.0	6.0	1008.0	1763.0	1823.0	2204.0
금천	93.0	117.0	6.0	7.0	4.0	4.0	553.0	999.0	1310.0	1471.0
남대문	52.0	73.0	2.0	2.0	1.0	1.0	368.0	699.0	494.0	579.0
노원	148.0	169.0	3.0	3.0	3.0	4.0	646.0	1404.0	1781.0	2163.0
도봉	87.0	82.0	2.0	2.0	0.0	0.0	379.0	861.0	1022.0	1234.0
동대문	142.0	161.0	5.0	6.0	8.0	10.0	713.0	1440.0	1597.0	1784.0
동작	121.0	226.0	2.0	1.0	1.0	1.0	592.0	1341.0	1370.0	1631.0
마포	287.0	348.0	3.0	4.0	8.0	9.0	679.0	1483.0	1557.0	1844.0

col	강간,추행 검거	강간,추행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생
구분										
방배	50.0	63.0	1.0	1.0	1.0	1.0	159.0	348.0	365.0	423.0
서대 문	149.0	159.0	1.0	2.0	3.0	3.0	524.0	1077.0	1029.0	1292.0
서부	71.0	76.0	2.0	3.0	3.0	3.0	322.0	532.0	667.0	748.0
서초	253.0	333.0	3.0	5.0	4.0	4.0	801.0	1673.0	1415.0	1750.0
성동	98.0	115.0	3.0	3.0	5.0	5.0	565.0	1016.0	1103.0	1223.0
성북	77.0	95.0	2.0	2.0	0.0	0.0	233.0	439.0	610.0	672.0
송파	235.0	287.0	17.0	13.0	6.0	6.0	1006.0	2429.0	2280.0	2675.0
수서	194.0	195.0	1.0	2.0	2.0	1.0	475.0	1010.0	1200.0	1394.0
양천	90.0	109.0	3.0	3.0	5.0	5.0	718.0	1517.0	1270.0	1582.0
영등 포	252.0	315.0	5.0	6.0	4.0	7.0	953.0	2188.0	2157.0	2701.0
용산	186.0	230.0	6.0	6.0	3.0	5.0	534.0	1111.0	1327.0	1617.0
은평	74.0	95.0	1.0	1.0	1.0	1.0	434.0	839.0	956.0	1092.0
종로	93.0	115.0	4.0	4.0	6.0	9.0	284.0	602.0	792.0	964.0
종암	44.0	47.0	0.0	1.0	4.0	4.0	347.0	549.0	630.0	758.0
중랑	153.0	183.0	6.0	5.0	9.0	10.0	889.0	1506.0	1724.0	2022.0
중부	76.0	113.0	4.0	3.0	1.0	1.0	495.0	943.0	830.0	997.0
혜화	53.0	77.0	3.0	3.0	0.0	0.0	310.0	581.0	628.0	747.0

```
In [ ]: df_r = df_1.reset_index()
df_r
```

col	구분	강간,추행 검거	강간,추행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생
0	강남	433.0	513.0	10.0	12.0	3.0	5.0	1016.0	1941.0	1911.0	2283.0
1	강동	167.0	188.0	14.0	15.0	3.0	3.0	781.0	1640.0	1697.0	1942.0
2	강북	132.0	159.0	4.0	5.0	8.0	9.0	490.0	780.0	1625.0	1817.0
3	강서	222.0	265.0	5.0	6.0	9.0	10.0	1070.0	1689.0	2087.0	2445.0
4	관악	351.0	403.0	3.0	3.0	12.0	12.0	1137.0	2229.0	2355.0	2614.0
5	광진	224.0	255.0	4.0	4.0	2.0	2.0	874.0	1664.0	1407.0	1676.0
6	구로	142.0	197.0	5.0	5.0	6.0	6.0	1008.0	1763.0	1823.0	2204.0
7	금천	93.0	117.0	6.0	7.0	4.0	4.0	553.0	999.0	1310.0	1471.0
8	남대 문	52.0	73.0	2.0	2.0	1.0	1.0	368.0	699.0	494.0	579.0
9	노원	148.0	169.0	3.0	3.0	3.0	4.0	646.0	1404.0	1781.0	2163.0
10	도봉	87.0	82.0	2.0	2.0	0.0	0.0	379.0	861.0	1022.0	1234.0

col	구분	강간,추행 검거	강간,추행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생
11	동대문	142.0	161.0	5.0	6.0	8.0	10.0	713.0	1440.0	1597.0	1784.0
12	동작	121.0	226.0	2.0	1.0	1.0	1.0	592.0	1341.0	1370.0	1631.0
13	마포	287.0	348.0	3.0	4.0	8.0	9.0	679.0	1483.0	1557.0	1844.0
14	방배	50.0	63.0	1.0	1.0	1.0	1.0	159.0	348.0	365.0	423.0
15	서대문	149.0	159.0	1.0	2.0	3.0	3.0	524.0	1077.0	1029.0	1292.0
16	서부	71.0	76.0	2.0	3.0	3.0	3.0	322.0	532.0	667.0	748.0
17	서초	253.0	333.0	3.0	5.0	4.0	4.0	801.0	1673.0	1415.0	1750.0
18	성동	98.0	115.0	3.0	3.0	5.0	5.0	565.0	1016.0	1103.0	1223.0
19	성북	77.0	95.0	2.0	2.0	0.0	0.0	233.0	439.0	610.0	672.0
20	송파	235.0	287.0	17.0	13.0	6.0	6.0	1006.0	2429.0	2280.0	2675.0
21	수서	194.0	195.0	1.0	2.0	2.0	1.0	475.0	1010.0	1200.0	1394.0
22	양천	90.0	109.0	3.0	3.0	5.0	5.0	718.0	1517.0	1270.0	1582.0
23	영등포	252.0	315.0	5.0	6.0	4.0	7.0	953.0	2188.0	2157.0	2701.0
24	용산	186.0	230.0	6.0	6.0	3.0	5.0	534.0	1111.0	1327.0	1617.0
25	은평	74.0	95.0	1.0	1.0	1.0	1.0	434.0	839.0	956.0	1092.0
26	종로	93.0	115.0	4.0	4.0	6.0	9.0	284.0	602.0	792.0	964.0
27	종암	44.0	47.0	0.0	1.0	4.0	4.0	347.0	549.0	630.0	758.0
28	중랑	153.0	183.0	6.0	5.0	9.0	10.0	889.0	1506.0	1724.0	2022.0
29	중부	76.0	113.0	4.0	3.0	1.0	1.0	495.0	943.0	830.0	997.0
30	혜화	53.0	77.0	3.0	3.0	0.0	0.0	310.0	581.0	628.0	747.0

```
In [ ]: !pip install googlemaps
```

```
Requirement already satisfied: googlemaps in /usr/local/lib/python3.7/dist-packages (4.6.0)
Requirement already satisfied: requests<3.0,>=2.20.0 in /usr/local/lib/python3.7/dist-packages (from googlemaps) (2.23.0)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-packages (from requests<3.0,>=2.20.0->googlemaps) (3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-packages (from requests<3.0,>=2.20.0->googlemaps) (2021.10.8)
Requirement already satisfied: urllib3!=1.25.0,!1.25.1,<1.26,>=1.21.1 in /usr/local/lib/python3.7/dist-packages (from requests<3.0,>=2.20.0->googlemaps) (1.24.3)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages (from requests<3.0,>=2.20.0->googlemaps) (2.10)
```

```
In [ ]: import googlemaps
```

```
gkeys = "*****"
gmaps = googlemaps.Client(key=gkeys)
```

```
In [ ]: # 서울중부경찰서 정보 불러오기
# 구글맵 활용
```

```
gmaps.geocode('서울중부경찰서', language='ko')
```

```
Out[ ]: [{ 'address_components': [{ 'long_name': ' 2 7 ',
    'short_name': ' 2 7 ',
    'types': ['premise']},
  { 'long_name': '수표로',
    'short_name': '수표로',
    'types': ['political', 'sublocality', 'sublocality_level_4']},
  { 'long_name': '을지로동',
    'short_name': '을지로동',
    'types': ['political', 'sublocality', 'sublocality_level_2']},
  { 'long_name': '중구',
    'short_name': '중구',
    'types': ['political', 'sublocality', 'sublocality_level_1']},
  { 'long_name': '서울특별시',
    'short_name': '서울특별시',
    'types': ['administrative_area_level_1', 'political']},
  { 'long_name': '대한민국',
    'short_name': 'KR',
    'types': ['country', 'political']},
  { 'long_name': '100-032',
    'short_name': '100-032',
    'types': ['postal_code']}],
  'formatted_address': '대한민국 서울특별시 중구 을지로동 수표로 27',
  'geometry': { 'location': { 'lat': 37.5636465, 'lng': 126.9895796},
    'location_type': 'R00FTOP',
    'viewport': { 'northeast': { 'lat': 37.56499548029149,
      'lng': 126.9909285802915},
    'southwest': { 'lat': 37.56229751970849, 'lng': 126.9882306197085}}},
  'place_id': 'ChIJc-9q5uSifDURLhQmr5wkXmc',
  'plus_code': { 'compound_code': 'HX7Q+FR 대한민국 서울특별시',
    'global_code': '8Q98HX7Q+FR' },
  'types': ['establishment', 'point_of_interest', 'police']}]
```

```
In [ ]: # 서울 경찰서 불러오기
station_name = []

for name in df_r['구분']:
    station_name.append('서울' + str(name) + '경찰서')

station_name
```

```
Out[ ]: ['서울강남경찰서',
  '서울강동경찰서',
  '서울강북경찰서',
  '서울강서경찰서',
  '서울관악경찰서',
  '서울광진경찰서',
  '서울구로경찰서',
  '서울금천경찰서',
  '서울남대문경찰서',
  '서울노원경찰서',
  '서울도봉경찰서',
  '서울동대문경찰서',
  '서울동작경찰서',
  '서울마포경찰서',
  '서울방배경찰서',
  '서울서대문경찰서',
  '서울서부경찰서',
  '서울서초경찰서',
  '서울성동경찰서',
  '서울성북경찰서',
  '서울송파경찰서',
  '서울수서경찰서',
  '서울양천경찰서',
  '서울영등포경찰서',
  '서울용산경찰서',
  '서울은평경찰서',
```



```
'서울종로경찰서',
'서울중앙경찰서',
'서울중랑경찰서',
'서울중부경찰서',
'서울혜화경찰서']
```

```
In [ ]: station_addreess = []
station_lat = []
station_lng = []
# 경찰서의 이름으로 주소 가져오기
for name in station_name:
    tmp = gmaps.geocode(name, language='ko')
    station_addreess.append(tmp[0].get("formatted_address"))

    tmp_loc = tmp[0].get("geometry")

    station_lat.append(tmp_loc['location']['lat'])
    station_lng.append(tmp_loc['location']['lng'])

    print(name + '-->' + tmp[0].get("formatted_address"))
```

```
서울강남경찰서-->대한민국 서울특별시 강남구 테헤란로114길 11
서울강동경찰서-->대한민국 서울특별시 강동구 성내로 57
서울강북경찰서-->대한민국 서울특별시 강북구 오패산로 406
서울강서경찰서-->대한민국 서울특별시 양천구 신월동 화곡로 73
서울관악경찰서-->대한민국 서울특별시 관악구 관악로5길 33
서울광진경찰서-->대한민국 서울특별시 광진구 구의동 자양로 167
서울구로경찰서-->대한민국 서울특별시 구로구 가마산로 235
서울금천경찰서-->대한민국 서울특별시 금천구 시흥대로73길 50
서울남대문경찰서-->대한민국 서울특별시 중구 한강대로 410
서울노원경찰서-->대한민국 서울특별시 노원구 하계동 노원로 283
서울도봉경찰서-->대한민국 서울특별시 도봉구 노해로 403
서울동대문경찰서-->대한민국 서울특별시 동대문구 약령시로21길 29
서울 동작경찰서-->대한민국 서울특별시 동작구 노량진로 148
서울마포경찰서-->대한민국 서울특별시 마포구 마포대로 183
서울방배경찰서-->대한민국 서울특별시 서초구 동작대로 204
서울서대문경찰서-->대한민국 서울특별시 서대문구 통일로 113
서울서부경찰서-->대한민국 서울특별시 은평구 진흥로 58
서울서초경찰서-->대한민국 서울특별시 서초구 서초3동 반포대로 179
서울성동경찰서-->대한민국 서울특별시 성동구 행당동 왕십리광장로 9
서울성북경찰서-->대한민국 서울특별시 성북구 삼선동 보문로 170
서울송파경찰서-->대한민국 서울특별시 송파구 가락동 중대로 221
서울수서경찰서-->대한민국 서울특별시 강남구 개포로 617
서울양천경찰서-->대한민국 서울특별시 양천구 목동동로 99
서울영등포경찰서-->대한민국 서울특별시 영등포구 영등포동1가 618-7
서울용산경찰서-->대한민국 서울특별시 용산구 백범로 329
서울은평경찰서-->대한민국 서울특별시 은평구 불광동 연서로 365
서울종로경찰서-->대한민국 서울특별시 종로구 종로1.2.3.4가동 율곡로 46
서울중앙경찰서-->대한민국 서울특별시 성북구 중암로 135
서울중랑경찰서-->대한민국 서울특별시 중랑구 신내역로3길 40-10
서울중부경찰서-->대한민국 서울특별시 중구 을지로동 수표로 27
서울혜화경찰서-->대한민국 서울특별시 종로구 인의동 창경궁로 112-16
```

```
In [ ]: # 주소 목록 확인
station_addreess
```

```
Out [ ]: ['대한민국 서울특별시 강남구 테헤란로114길 11',
'대한민국 서울특별시 강동구 성내로 57',
'대한민국 서울특별시 강북구 오패산로 406',
'대한민국 서울특별시 양천구 신월동 화곡로 73',
'대한민국 서울특별시 관악구 관악로5길 33',
'대한민국 서울특별시 광진구 구의동 자양로 167',
'대한민국 서울특별시 구로구 가마산로 235',
'대한민국 서울특별시 금천구 시흥대로73길 50',
'대한민국 서울특별시 중구 한강대로 410',
'대한민국 서울특별시 노원구 하계동 노원로 283',
'대한민국 서울특별시 도봉구 노해로 403',
'대한민국 서울특별시 동대문구 약령시로21길 29',
```

'대한민국 서울특별시 동작구 노량진로 148',  
 '대한민국 서울특별시 마포구 마포대로 183',  
 '대한민국 서울특별시 서초구 동작대로 204',  
 '대한민국 서울특별시 서대문구 통일로 113',  
 '대한민국 서울특별시 은평구 진흥로 58',  
 '대한민국 서울특별시 서초구 서초3동 반포대로 179',  
 '대한민국 서울특별시 성동구 행당동 왕십리광장로 9',  
 '대한민국 서울특별시 성북구 삼선동 보문로 170',  
 '대한민국 서울특별시 송파구 가락동 중대로 221',  
 '대한민국 서울특별시 강남구 개포로 617',  
 '대한민국 서울특별시 양천구 목동동로 99',  
 '대한민국 서울특별시 영등포구 영등포동1가 618-7',  
 '대한민국 서울특별시 용산구 백범로 329',  
 '대한민국 서울특별시 은평구 불광동 연서로 365',  
 '대한민국 서울특별시 종로구 종로1.2.3.4가동 율곡로 46',  
 '대한민국 서울특별시 성북구 종암로 135',  
 '대한민국 서울특별시 중랑구 신내역로3길 40-10',  
 '대한민국 서울특별시 중구 을지로동 수표로 27',  
 '대한민국 서울특별시 종로구 인의동 창경궁로 112-16']

```
In [ ]: # 구별 열 새로 생성하기

gu_name = []

for name in station_addreess:
    tmp = name.split()

    tmp_gu = [gu for gu in tmp if gu[-1] == '구'][0]

    gu_name.append(tmp_gu)

df_r['구별'] = gu_name
df_r
```

col	구 분	강간,추 행 검거	강간,추 행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생	구 별
0	강 남	433.0	513.0	10.0	12.0	3.0	5.0	1016.0	1941.0	1911.0	2283.0	강 남 구
1	강 동	167.0	188.0	14.0	15.0	3.0	3.0	781.0	1640.0	1697.0	1942.0	강 동 구
2	강 북	132.0	159.0	4.0	5.0	8.0	9.0	490.0	780.0	1625.0	1817.0	강 북 구
3	강 서	222.0	265.0	5.0	6.0	9.0	10.0	1070.0	1689.0	2087.0	2445.0	양 천 구
4	관 악	351.0	403.0	3.0	3.0	12.0	12.0	1137.0	2229.0	2355.0	2614.0	관 악 구
5	광 진	224.0	255.0	4.0	4.0	2.0	2.0	874.0	1664.0	1407.0	1676.0	광 진 구
6	구 로	142.0	197.0	5.0	5.0	6.0	6.0	1008.0	1763.0	1823.0	2204.0	구 로 구

col	구 분	강간,추 행 검거	강간,추 행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생	구 별
7	금천	93.0	117.0	6.0	7.0	4.0	4.0	553.0	999.0	1310.0	1471.0	금천구
8	남대문	52.0	73.0	2.0	2.0	1.0	1.0	368.0	699.0	494.0	579.0	종구
9	노원	148.0	169.0	3.0	3.0	3.0	4.0	646.0	1404.0	1781.0	2163.0	노원구
10	도봉	87.0	82.0	2.0	2.0	0.0	0.0	379.0	861.0	1022.0	1234.0	도봉구
11	동대문	142.0	161.0	5.0	6.0	8.0	10.0	713.0	1440.0	1597.0	1784.0	동대문구
12	동작	121.0	226.0	2.0	1.0	1.0	1.0	592.0	1341.0	1370.0	1631.0	동작구
13	마포	287.0	348.0	3.0	4.0	8.0	9.0	679.0	1483.0	1557.0	1844.0	마포구
14	방배	50.0	63.0	1.0	1.0	1.0	1.0	159.0	348.0	365.0	423.0	서초구
15	서대문	149.0	159.0	1.0	2.0	3.0	3.0	524.0	1077.0	1029.0	1292.0	서대문구
16	서부	71.0	76.0	2.0	3.0	3.0	3.0	322.0	532.0	667.0	748.0	은평구
17	서초	253.0	333.0	3.0	5.0	4.0	4.0	801.0	1673.0	1415.0	1750.0	서초구
18	성동	98.0	115.0	3.0	3.0	5.0	5.0	565.0	1016.0	1103.0	1223.0	성동구
19	성북	77.0	95.0	2.0	2.0	0.0	0.0	233.0	439.0	610.0	672.0	성북구
20	송파	235.0	287.0	17.0	13.0	6.0	6.0	1006.0	2429.0	2280.0	2675.0	송파구
21	수서	194.0	195.0	1.0	2.0	2.0	1.0	475.0	1010.0	1200.0	1394.0	강남구
22	양천	90.0	109.0	3.0	3.0	5.0	5.0	718.0	1517.0	1270.0	1582.0	양천구

col	구분	강간,추행 검거	강간,추행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생	구별
23	영등포	252.0	315.0	5.0	6.0	4.0	7.0	953.0	2188.0	2157.0	2701.0	영등포구
24	용산	186.0	230.0	6.0	6.0	3.0	5.0	534.0	1111.0	1327.0	1617.0	용산구
25	은평	74.0	95.0	1.0	1.0	1.0	1.0	434.0	839.0	956.0	1092.0	은평구
26	종로	93.0	115.0	4.0	4.0	6.0	9.0	284.0	602.0	792.0	964.0	종로구
27	중암	44.0	47.0	0.0	1.0	4.0	4.0	347.0	549.0	630.0	758.0	성북구
28	중랑	153.0	183.0	6.0	5.0	9.0	10.0	889.0	1506.0	1724.0	2022.0	중랑구
29	중부	76.0	113.0	4.0	3.0	1.0	1.0	495.0	943.0	830.0	997.0	중구
30	혜화	53.0	77.0	3.0	3.0	0.0	0.0	310.0	581.0	628.0	747.0	종로구

```
In [ ]: # 금천 예외 처리
df_r.loc[df_r["구분"] == "금천", ["구별"]] = "관악구"
df_r
```

col	구분	강간,추행 검거	강간,추행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생	구별
0	강남	433.0	513.0	10.0	12.0	3.0	5.0	1016.0	1941.0	1911.0	2283.0	강남구
1	강동	167.0	188.0	14.0	15.0	3.0	3.0	781.0	1640.0	1697.0	1942.0	강동구
2	강북	132.0	159.0	4.0	5.0	8.0	9.0	490.0	780.0	1625.0	1817.0	강북구
3	강서	222.0	265.0	5.0	6.0	9.0	10.0	1070.0	1689.0	2087.0	2445.0	양천구
4	관악	351.0	403.0	3.0	3.0	12.0	12.0	1137.0	2229.0	2355.0	2614.0	관악구
5	광진	224.0	255.0	4.0	4.0	2.0	2.0	874.0	1664.0	1407.0	1676.0	광진구

col	구 분	강간,추 행 검거	강간,추 행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생	구 별
6	구 로	142.0	197.0	5.0	5.0	6.0	6.0	1008.0	1763.0	1823.0	2204.0	구 로 구
7	금 천	93.0	117.0	6.0	7.0	4.0	4.0	553.0	999.0	1310.0	1471.0	관 악 구
8	남 대 문	52.0	73.0	2.0	2.0	1.0	1.0	368.0	699.0	494.0	579.0	중 구
9	노 원	148.0	169.0	3.0	3.0	3.0	4.0	646.0	1404.0	1781.0	2163.0	노 원 구
10	도 봉	87.0	82.0	2.0	2.0	0.0	0.0	379.0	861.0	1022.0	1234.0	도 봉 구
11	동 대 문	142.0	161.0	5.0	6.0	8.0	10.0	713.0	1440.0	1597.0	1784.0	동 대 문 구
12	동 작	121.0	226.0	2.0	1.0	1.0	1.0	592.0	1341.0	1370.0	1631.0	동 작 구
13	마 포	287.0	348.0	3.0	4.0	8.0	9.0	679.0	1483.0	1557.0	1844.0	마 포 구
14	방 배	50.0	63.0	1.0	1.0	1.0	1.0	159.0	348.0	365.0	423.0	서 초 구
15	서 대 문	149.0	159.0	1.0	2.0	3.0	3.0	524.0	1077.0	1029.0	1292.0	서 대 문 구
16	서 부	71.0	76.0	2.0	3.0	3.0	3.0	322.0	532.0	667.0	748.0	은 평 구
17	서 초	253.0	333.0	3.0	5.0	4.0	4.0	801.0	1673.0	1415.0	1750.0	서 초 구
18	성 동	98.0	115.0	3.0	3.0	5.0	5.0	565.0	1016.0	1103.0	1223.0	성 동 구
19	성 북	77.0	95.0	2.0	2.0	0.0	0.0	233.0	439.0	610.0	672.0	성 북 구
20	송 파	235.0	287.0	17.0	13.0	6.0	6.0	1006.0	2429.0	2280.0	2675.0	송 파 구
21	수 서	194.0	195.0	1.0	2.0	2.0	1.0	475.0	1010.0	1200.0	1394.0	강 남 구

col	구 분	강간,추 행 검거	강간,추 행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생	구 별
22	양 천	90.0	109.0	3.0	3.0	5.0	5.0	718.0	1517.0	1270.0	1582.0	양 천 구
23	영 등 포	252.0	315.0	5.0	6.0	4.0	7.0	953.0	2188.0	2157.0	2701.0	영 등 포 구
24	용 산	186.0	230.0	6.0	6.0	3.0	5.0	534.0	1111.0	1327.0	1617.0	용 산 구
25	이 평	74.0	95.0	1.0	1.0	1.0	1.0	434.0	839.0	956.0	1092.0	은 평 구
26	종 로	93.0	115.0	4.0	4.0	6.0	9.0	284.0	602.0	792.0	964.0	종 로 구
27	종 암	44.0	47.0	0.0	1.0	4.0	4.0	347.0	549.0	630.0	758.0	성 북 구
28	중 랑	153.0	183.0	6.0	5.0	9.0	10.0	889.0	1506.0	1724.0	2022.0	중 랑 구
29	중 부	76.0	113.0	4.0	3.0	1.0	1.0	495.0	943.0	830.0	997.0	중 구
30	혜 화	53.0	77.0	3.0	3.0	0.0	0.0	310.0	581.0	628.0	747.0	종 로 구

```
In [ ]: # 전처리 데이터 저장
df_r.to_csv("/content/drive/MyDrive/스터디/데이터 주무르기/data/02_preprocess.csv", inc
```

## 2.2 피벗 테이블 활용하기

```
In [ ]: import numpy as np
```

```
In [ ]: # 피벗 테이블로 구별로 인덱스
crime_anal = pd.pivot_table(df_r, index='구별', aggfunc=np.sum)
crime_anal.head()
```

```
Out [ ]:
```

col	강간,추 행 검거	강간,추 행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발 생
구별										
강남 구	627.0	708.0	11.0	14.0	5.0	6.0	1491.0	2951.0	3111.0	3677.0
강동 구	167.0	188.0	14.0	15.0	3.0	3.0	781.0	1640.0	1697.0	1942.0
강북 구	132.0	159.0	4.0	5.0	8.0	9.0	490.0	780.0	1625.0	1817.0

col	강간,추행 검거	강간,추행 발생	강도 검거	강도 발생	살인 검거	살인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생
구별										
관악 구	444.0	520.0	9.0	10.0	16.0	16.0	1690.0	3228.0	3665.0	4085.0
광진 구	224.0	255.0	4.0	4.0	2.0	2.0	874.0	1664.0	1407.0	1676.0

In [ ]:

```
# 검거율 열 새로 생성

crime_anal['강간,추행 검거율'] = crime_anal['강간,추행 검거']/crime_anal['강간,추행 발생']
crime_anal['강도검거율'] = crime_anal['강도 검거']/crime_anal['강도 발생']
crime_anal['살인검거율'] = crime_anal['살인 검거']/crime_anal['살인 발생']
crime_anal['절도검거율'] = crime_anal['절도 검거']/crime_anal['절도 발생']
crime_anal['폭력검거율'] = crime_anal['폭력 검거']/crime_anal['폭력 발생']

del crime_anal['강간,추행 검거']
del crime_anal['강도 검거']
del crime_anal['살인 검거']
del crime_anal['절도 검거']
del crime_anal['폭력 검거']

crime_anal
```

Out [ ]:

col	강간,추 행 발생	강도 발생	살인 발생	절도 발생	폭력 발생	강간,추행 검거율	강도검거 율	살인검거 율	절도검거 율	폭력검거 율
구별										
강남 구	708.0	14.0	6.0	2951.0	3677.0	0.885593	0.785714	0.833333	0.505252	0.846070
강동 구	188.0	15.0	3.0	1640.0	1942.0	0.888298	0.933333	1.000000	0.476220	0.873841
강북 구	159.0	5.0	9.0	780.0	1817.0	0.830189	0.800000	0.888889	0.628205	0.894331
관악 구	520.0	10.0	16.0	3228.0	4085.0	0.853846	0.900000	1.000000	0.523544	0.897185
광진 구	255.0	4.0	2.0	1664.0	1676.0	0.878431	1.000000	1.000000	0.525240	0.839499
구로 구	197.0	5.0	6.0	1763.0	2204.0	0.720812	1.000000	1.000000	0.571753	0.827132
노원 구	169.0	3.0	4.0	1404.0	2163.0	0.875740	1.000000	0.750000	0.460114	0.823393
도봉 구	82.0	2.0	0.0	861.0	1234.0	1.060976	1.000000	NaN	0.440186	0.828201
동대 문구	161.0	6.0	10.0	1440.0	1784.0	0.881988	0.833333	0.800000	0.495139	0.895179
동작 구	226.0	1.0	1.0	1341.0	1631.0	0.535398	2.000000	1.000000	0.441462	0.839975
마포 구	348.0	4.0	9.0	1483.0	1844.0	0.824713	0.750000	0.888889	0.457856	0.844360

col	강간,추 행 발생	강도 발생	살인 발생	절도 발생	폭력 발생	강간,추행 검거율	강도검거 율	살인검거 율	절도검거 율	폭력검거 율
구별										
서대 문구	159.0	2.0	3.0	1077.0	1292.0	0.937107	0.500000	1.000000	0.486537	0.796440
서초 구	396.0	6.0	5.0	2021.0	2173.0	0.765152	0.666667	1.000000	0.475012	0.819144
성동 구	115.0	3.0	5.0	1016.0	1223.0	0.852174	1.000000	1.000000	0.556102	0.901881
성북 구	142.0	3.0	4.0	988.0	1430.0	0.852113	0.666667	1.000000	0.587045	0.867133
송파 구	287.0	13.0	6.0	2429.0	2675.0	0.818815	1.307692	1.000000	0.414162	0.852336
양천 구	374.0	9.0	15.0	3206.0	4027.0	0.834225	0.888889	0.933333	0.557704	0.833623
영등 포구	315.0	6.0	7.0	2188.0	2701.0	0.800000	0.833333	0.571429	0.435558	0.798593
용산 구	230.0	6.0	5.0	1111.0	1617.0	0.808696	1.000000	0.600000	0.480648	0.820656
은평 구	171.0	4.0	4.0	1371.0	1840.0	0.847953	0.750000	1.000000	0.551422	0.882065
종로 구	192.0	7.0	9.0	1183.0	1711.0	0.760417	1.000000	0.666667	0.502113	0.829924
중구	186.0	5.0	2.0	1642.0	1576.0	0.688172	1.200000	1.000000	0.525579	0.840102
중랑 구	183.0	5.0	10.0	1506.0	2022.0	0.836066	1.200000	0.900000	0.590305	0.852621

```
In [ ]: con_list = ['강간,추행 검거율', '강도검거율', '살인검거율', '절도검거율', '폭력검거율']
# 검거율이 1이상인 데이터 통일

for column in con_list:
    crime_anal.loc[crime_anal[column] > 1, column] = 1

crime_anal
```

col	강간,추 행 발생	강도 발생	살인 발생	절도 발생	폭력 발생	강간,추행 검거율	강도검거 율	살인검거 율	절도검거 율	폭력검거 율
구별										
강남 구	708.0	14.0	6.0	2951.0	3677.0	0.885593	0.785714	0.833333	0.505252	0.846070
강동 구	188.0	15.0	3.0	1640.0	1942.0	0.888298	0.933333	1.000000	0.476220	0.873841
강북 구	159.0	5.0	9.0	780.0	1817.0	0.830189	0.800000	0.888889	0.628205	0.894331
관악 구	520.0	10.0	16.0	3228.0	4085.0	0.853846	0.900000	1.000000	0.523544	0.897185





crime\_anal

Out[ ]:

col	강간,추 행	강 도	살 인	절도	폭력	강간,추행 검 거율	강도검거 율	살인검거 율	절도검거 율	폭력검거 율
구별										
강남구	708.0	14.0	6.0	2951.0	3677.0	0.885593	0.785714	0.833333	0.505252	0.846070
강동구	188.0	15.0	3.0	1640.0	1942.0	0.888298	0.933333	1.000000	0.476220	0.873841
강북구	159.0	5.0	9.0	780.0	1817.0	0.830189	0.800000	0.888889	0.628205	0.894331
관악구	520.0	10.0	16.0	3228.0	4085.0	0.853846	0.900000	1.000000	0.523544	0.897185
광진구	255.0	4.0	2.0	1664.0	1676.0	0.878431	1.000000	1.000000	0.525240	0.839499
구로구	197.0	5.0	6.0	1763.0	2204.0	0.720812	1.000000	1.000000	0.571753	0.827132
노원구	169.0	3.0	4.0	1404.0	2163.0	0.875740	1.000000	0.750000	0.460114	0.823393
도봉구	82.0	2.0	0.0	861.0	1234.0	1.000000	1.000000	NaN	0.440186	0.828201
동대문 구	161.0	6.0	10.0	1440.0	1784.0	0.881988	0.833333	0.800000	0.495139	0.895179
동작구	226.0	1.0	1.0	1341.0	1631.0	0.535398	1.000000	1.000000	0.441462	0.839975
마포구	348.0	4.0	9.0	1483.0	1844.0	0.824713	0.750000	0.888889	0.457856	0.844360
서대문 구	159.0	2.0	3.0	1077.0	1292.0	0.937107	0.500000	1.000000	0.486537	0.796440
서초구	396.0	6.0	5.0	2021.0	2173.0	0.765152	0.666667	1.000000	0.475012	0.819144
성동구	115.0	3.0	5.0	1016.0	1223.0	0.852174	1.000000	1.000000	0.556102	0.901881
성북구	142.0	3.0	4.0	988.0	1430.0	0.852113	0.666667	1.000000	0.587045	0.867133
송파구	287.0	13.0	6.0	2429.0	2675.0	0.818815	1.000000	1.000000	0.414162	0.852336
양천구	374.0	9.0	15.0	3206.0	4027.0	0.834225	0.888889	0.933333	0.557704	0.833623
영등포 구	315.0	6.0	7.0	2188.0	2701.0	0.800000	0.833333	0.571429	0.435558	0.798593
용산구	230.0	6.0	5.0	1111.0	1617.0	0.808696	1.000000	0.600000	0.480648	0.820656
은평구	171.0	4.0	4.0	1371.0	1840.0	0.847953	0.750000	1.000000	0.551422	0.882065
종로구	192.0	7.0	9.0	1183.0	1711.0	0.760417	1.000000	0.666667	0.502113	0.829924
중구	186.0	5.0	2.0	1642.0	1576.0	0.688172	1.000000	1.000000	0.525579	0.840102
종랑구	183.0	5.0	10.0	1506.0	2022.0	0.836066	1.000000	0.900000	0.590305	0.852621

## 2.3 데이터 표현을 위해 다듬기

In [ ]:

```
from sklearn import preprocessing

col = ['강간,추행', '강도', '살인', '절도', '폭력']

x = crime_anal[col].values
min_max_scaler = preprocessing.MinMaxScaler()

x_scaled = min_max_scaler.fit_transform(x.astype(float))
crime_anal_norm = pd.DataFrame(x_scaled, columns = col, index = crime_anal.index)
```

```
col2 = ['강간,추행 검거율', '강도검거율', '살인검거율', '절도검거율', '폭력검거율']
crime_anal_norm[col2] = crime_anal[col2]
crime_anal_norm
```

Out[ ]:

	강간,추 행	강도	살인	절도	폭력	강간,추 행 검거 율	강도검거 율	살인검거 율	절도검거 율	폭력검거 율
구 별										
강 남 구	1.000000	0.928571	0.3750	0.886846	0.857442	0.885593	0.785714	0.833333	0.505252	0.846070
강 동 구	0.169329	1.000000	0.1875	0.351307	0.251223	0.888298	0.933333	1.000000	0.476220	0.873841
강 북 구	0.123003	0.285714	0.5625	0.000000	0.207547	0.830189	0.800000	0.888889	0.628205	0.894331
관 악 구	0.699681	0.642857	1.0000	1.000000	1.000000	0.853846	0.900000	1.000000	0.523544	0.897185
광 진 구	0.276358	0.214286	0.1250	0.361111	0.158281	0.878431	1.000000	1.000000	0.525240	0.839499
구 로 구	0.183706	0.285714	0.3750	0.401552	0.342767	0.720812	1.000000	1.000000	0.571753	0.827132
노 원 구	0.138978	0.142857	0.2500	0.254902	0.328442	0.875740	1.000000	0.750000	0.460114	0.823393
도 봉 구	0.000000	0.071429	0.0000	0.033088	0.003843	1.000000	1.000000	NaN	0.440186	0.828201
동 대 문 구	0.126198	0.357143	0.6250	0.269608	0.196017	0.881988	0.833333	0.800000	0.495139	0.895179
동 작 구	0.230032	0.000000	0.0625	0.229167	0.142558	0.535398	1.000000	1.000000	0.441462	0.839975
마 포 구	0.424920	0.214286	0.5625	0.287173	0.216981	0.824713	0.750000	0.888889	0.457856	0.844360
서 대 문 구	0.123003	0.071429	0.1875	0.121324	0.024109	0.937107	0.500000	1.000000	0.486537	0.796440
서 초 구	0.501597	0.357143	0.3125	0.506944	0.331936	0.765152	0.666667	1.000000	0.475012	0.819144

	강간,추 행	강도	살인	절도	폭력	강간,추 행 검거 율	강도검거 율	살인검거 율	절도검거 율	폭력검거 율
구 별										
성 동 구	0.052716	0.142857	0.3125	0.096405	0.000000	0.852174	1.000000	1.000000	0.556102	0.901881
성 북 구	0.095847	0.142857	0.2500	0.084967	0.072327	0.852113	0.666667	1.000000	0.587045	0.867133
송 파 구	0.327476	0.857143	0.3750	0.673611	0.507338	0.818815	1.000000	1.000000	0.414162	0.852336
양 천 구	0.466454	0.571429	0.9375	0.991013	0.979734	0.834225	0.888889	0.933333	0.557704	0.833623
영 등 포 구	0.372204	0.357143	0.4375	0.575163	0.516422	0.800000	0.833333	0.571429	0.435558	0.798593
용 산 구	0.236422	0.357143	0.3125	0.135212	0.137666	0.808696	1.000000	0.600000	0.480648	0.820656
은 평 구	0.142173	0.214286	0.2500	0.241422	0.215584	0.847953	0.750000	1.000000	0.551422	0.882065
종 로 구	0.175719	0.428571	0.5625	0.164624	0.170510	0.760417	1.000000	0.666667	0.502113	0.829924
중 구	0.166134	0.285714	0.1250	0.352124	0.123340	0.688172	1.000000	1.000000	0.525579	0.840102
중 랑 구	0.161342	0.285714	0.6250	0.296569	0.279175	0.836066	1.000000	0.900000	0.590305	0.852621

```
In [ ]: # cctv 데이터와 결합
result_CCTV = pd.read_csv('/content/drive/MyDrive/스터디/데이터 주무르기/data/01_prepr
                                index_col='구별')
crime_anal_norm[['인구수', 'CCTV']] = result_CCTV[['인구수', '총계']]
crime_anal_norm.head()
```

	강간,추 행	강도	살인	절도	폭력	강간,추 행 검거 율	강도검거 율	살인검거 율	절도검거 율	폭력검거 율
구 별										
강 남 구	1.000000	0.928571	0.3750	0.886846	0.857442	0.885593	0.785714	0.833333	0.505252	0.846070

	강간,추행	강도	살인	절도	폭력	강간,추행 검거율	강도검거율	살인검거율	절도검거율	폭력검거율
구별										
강동구	0.169329	1.000000	0.1875	0.351307	0.251223	0.888298	0.933333	1.000000	0.476220	0.873841
강북구	0.123003	0.285714	0.5625	0.000000	0.207547	0.830189	0.800000	0.888889	0.628205	0.894331
관악구	0.699681	0.642857	1.0000	1.000000	1.000000	0.853846	0.900000	1.000000	0.523544	0.897185
광진구	0.276358	0.214286	0.1250	0.361111	0.158281	0.878431	1.000000	1.000000	0.525240	0.839499

```
In [ ]: # 범죄 열을 생성 np.sum 활용!!
col = ['강간,추행', '강도', '살인', '절도', '폭력']
crime_anal_norm['범죄'] = np.sum(crime_anal_norm[col], axis=1)
crime_anal_norm.head()
```

	강간,추행	강도	살인	절도	폭력	강간,추행 검거율	강도검거율	살인검거율	절도검거율	폭력검거율
구별										
강남구	1.000000	0.928571	0.3750	0.886846	0.857442	0.885593	0.785714	0.833333	0.505252	0.846070
강동구	0.169329	1.000000	0.1875	0.351307	0.251223	0.888298	0.933333	1.000000	0.476220	0.873841
강북구	0.123003	0.285714	0.5625	0.000000	0.207547	0.830189	0.800000	0.888889	0.628205	0.894331
관악구	0.699681	0.642857	1.0000	1.000000	1.000000	0.853846	0.900000	1.000000	0.523544	0.897185
광진구	0.276358	0.214286	0.1250	0.361111	0.158281	0.878431	1.000000	1.000000	0.525240	0.839499

```
In [ ]: # 검거 열을 새로이 생성
col = ['강간,추행 검거율', '강도검거율', '살인검거율', '절도검거율', '폭력검거율']
crime_anal_norm['검거'] = np.sum(crime_anal_norm[col], axis=1)
crime_anal_norm.head()
```

	강간,추행	강도	살인	절도	폭력	강간,추행 검거율	강도검거율	살인검거율	절도검거율	폭력검거율
--	-------	----	----	----	----	--------------	-------	-------	-------	-------

구 별	강간,추 행	강도	살인	절도	폭력	강간,추 행 검거 율	강도검거 율	살인검거 율	절도검거 율	폭력검거 율
구 별										
강 남 구	1.000000	0.928571	0.3750	0.886846	0.857442	0.885593	0.785714	0.833333	0.505252	0.846070
강 동 구	0.169329	1.000000	0.1875	0.351307	0.251223	0.888298	0.933333	1.000000	0.476220	0.873841
강 북 구	0.123003	0.285714	0.5625	0.000000	0.207547	0.830189	0.800000	0.888889	0.628205	0.894331
관 악 구	0.699681	0.642857	1.0000	1.000000	1.000000	0.853846	0.900000	1.000000	0.523544	0.897185
광 진 구	0.276358	0.214286	0.1250	0.361111	0.158281	0.878431	1.000000	1.000000	0.525240	0.839499

## 2.4 seaborn

```
In [ ]: # # 나눔 글꼴 설치
# !sudo apt-get install -y fonts-nanum
# !sudo fc-cache -fv
# !rm ~/.cache/matplotlib -rf
```

```
In [ ]: import matplotlib.pyplot as plt

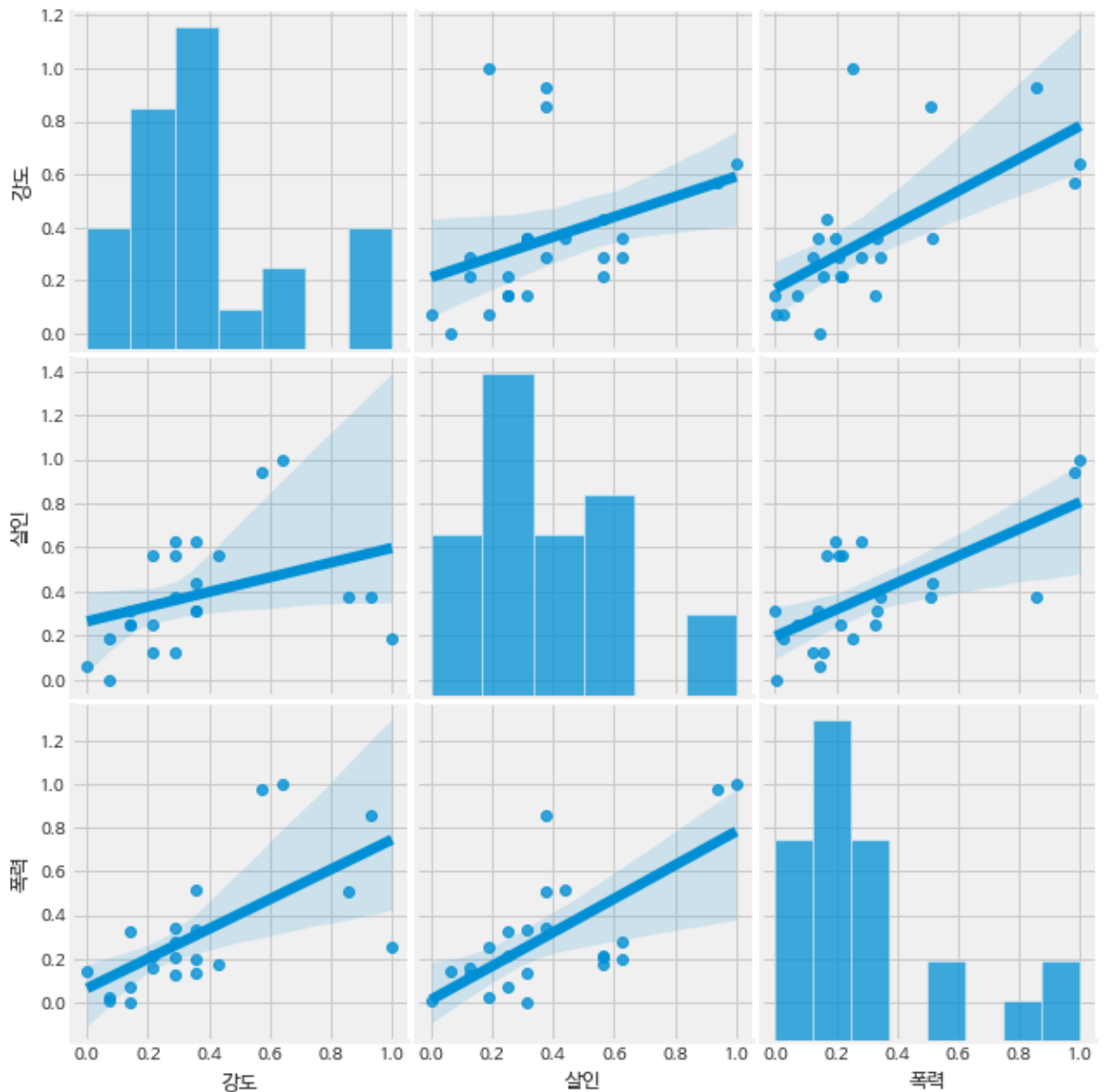
import seaborn as sns
#seaborn 스타일 설정
sns.set_style("whitegrid")

# 플롯 스타일 설정
plt.style.use('fivethirtyeight')

# matplotlib을 사용하는 모든 plot에 나눔 글꼴 적용
plt.rc('font', family='NanumBarunGothic')
%matplotlib inline
```

```
In [ ]: sns.pairplot(crime_anal_norm, vars=["강도", "살인", "폭력"], kind='reg', size=3)
plt.show()
```

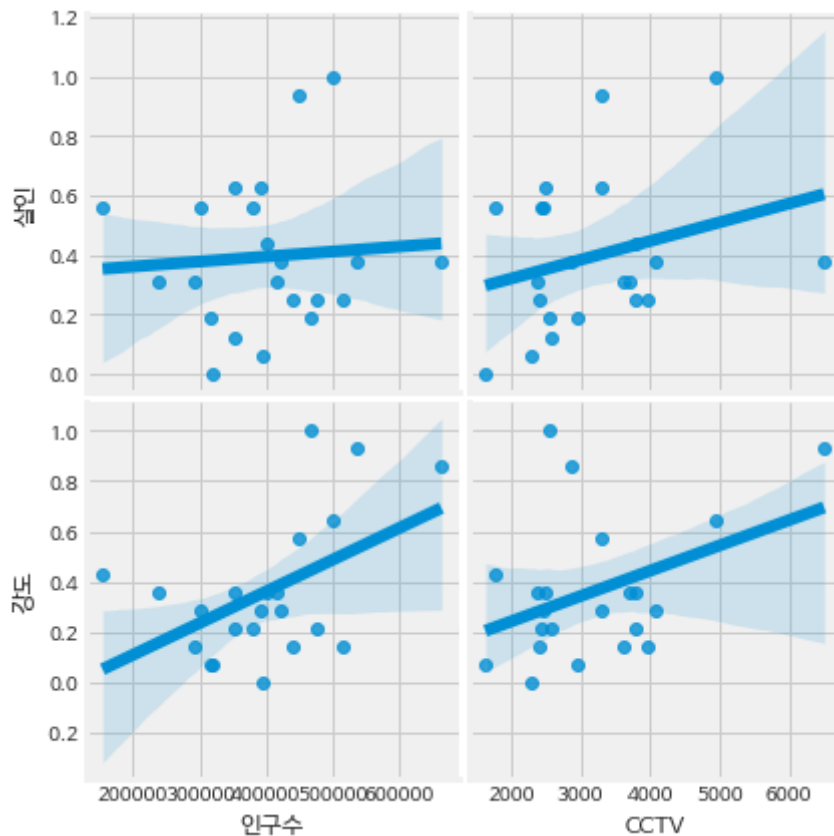
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:2076: UserWarning: The `size` parameter has been renamed to `height`; please update your code.  
warnings.warn(msg, UserWarning)



강도와 폭력, 살인과 폭력, 강도와 살인 모두 양의 상관관계가 있다.

```
In [ ]: sns.pairplot(crime_anal_norm, x_vars=["인구수", "CCTV"],
                  y_vars=["살인", "강도"], kind='reg', size=3)
plt.show()
```

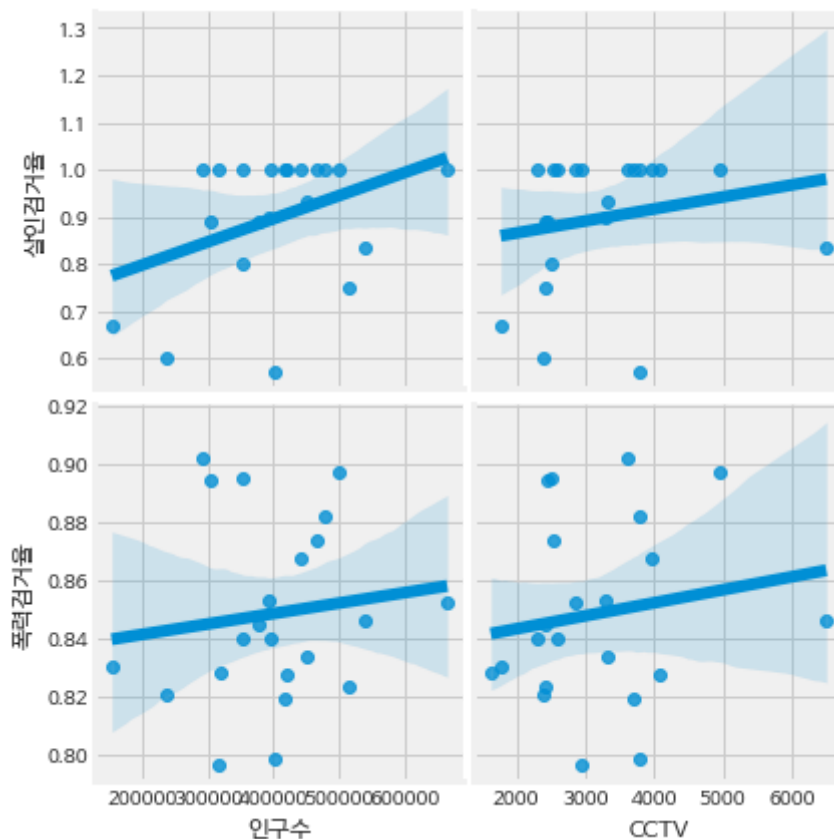
```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:2076: UserWarning: The `size`
parameter has been renamed to `height`; please update your code.
  warnings.warn(msg, UserWarning)
/usr/local/lib/python3.7/dist-packages/matplotlib/backends/backend_agg.py:214: Runtime
Warning: Glyph 8722 missing from current font.
  font.set_text(s, 0.0, flags=flags)
/usr/local/lib/python3.7/dist-packages/matplotlib/backends/backend_agg.py:183: Runtime
Warning: Glyph 8722 missing from current font.
  font.set_text(s, 0, flags=flags)
```



눈에 띄는 인사이트를 발견하지 못하였다. 굳이 찾아보자면 인구수가 증가할 수록 강도가 많이 일어나고 그런 지역일수록 CCTV가 많이 위치해 있는것으로 보인다.

```
In [ ]: sns.pairplot(crime_anal_norm, x_vars=["인구수", "CCTV"],
                  y_vars=["살인검거율", "폭력검거율"], kind='reg', size=3)
plt.show()
```

/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:2076: UserWarning: The `size` parameter has been renamed to `height`; please update your code.  
warnings.warn(msg, UserWarning)





- ```
In [ ]: sns.pairplot(crime_anal_norm, x_vars=["인구수", "CCTV"],
                  y_vars=["절도검거율", "강도검거율"], kind='reg', size=3)
plt.show()
```

Figure 2 consists of four scatter plots arranged in a 2x2 grid, showing the relationship between road network metrics and road network quality. The top row displays '정도점거울' (Degree Point Mirror) on the y-axis, and the bottom row displays '강도점거울' (Strength Point Mirror) on the y-axis. The x-axis for the left column is '인구수' (Population), and for the right column is 'CCTV' (CCTV). Each plot includes a blue regression line and a light blue shaded confidence interval.

- Top Left:** '정도점거울' vs. '인구수'. The y-axis ranges from 0.45 to 0.65, and the x-axis ranges from 200,000 to 700,000. The regression line shows a slight negative correlation.
- Top Right:** '정도점거울' vs. 'CCTV'. The y-axis ranges from 0.45 to 0.65, and the x-axis ranges from 2,000 to 6,000. The regression line shows a positive correlation.
- Bottom Left:** '강도점거울' vs. '인구수'. The y-axis ranges from 0.5 to 1.0, and the x-axis ranges from 200,000 to 700,000. The regression line shows a slight negative correlation.
- Bottom Right:** '강도점거울' vs. 'CCTV'. The y-axis ranges from 0.5 to 1.0, and the x-axis ranges from 2,000 to 6,000. The regression line shows a negative correlation.

- ```
In [ ]: # 검거열 스케일링
tmp_max = crime_anal_norm['검거'].max()
crime_anal_norm['검거'] = crime_anal_norm['검거'] / tmp_max
crime_anal_norm.sort = crime_anal_norm.sort_values(by='검거', ascending=False)
crime_anal_norm.sort.head()
```

강간, 추행      강도      살인      절도      폭력      강간, 추행      강도      강도      살인      절도      폭력      인.

구 별	강간,추 행	강도	살인	절도	폭력	강간,추 행 검거율	강도검거 율	살 인 검거 율	절도검거 율	폭력검거 율	안
성 동 구	0.052716	0.142857	0.3125	0.096405	0.000000	0.852174	1.000000	1.0	0.556102	0.901881	2926
광 진 구	0.276358	0.214286	0.1250	0.361111	0.158281	0.878431	1.000000	1.0	0.525240	0.839499	3526
중 랑 구	0.161342	0.285714	0.6250	0.296569	0.279175	0.836066	1.000000	0.9	0.590305	0.852621	3918
관 악 구	0.699681	0.642857	1.0000	1.000000	1.000000	0.853846	0.900000	1.0	0.523544	0.897185	4994
강 동 구	0.169329	1.000000	0.1875	0.351307	0.251223	0.888298	0.933333	1.0	0.476220	0.873841	4664

In [ ]:

```
# 히트맵 출력
target_col = ['강간,추행 검거율', '강도검거율', '살인검거율', '절도검거율', '폭력검거율']

crime_anal_norm_sort = crime_anal_norm.sort_values(by='검거', ascending=False)

plt.figure(figsize = (10,10))
sns.heatmap(crime_anal_norm_sort[target_col], annot=True, fmt='f',
            linewidths=.5, cmap='RdPu')
plt.title('범죄 검거 비율 (정규화된 검거의 합으로 정렬)')
plt.show()
```

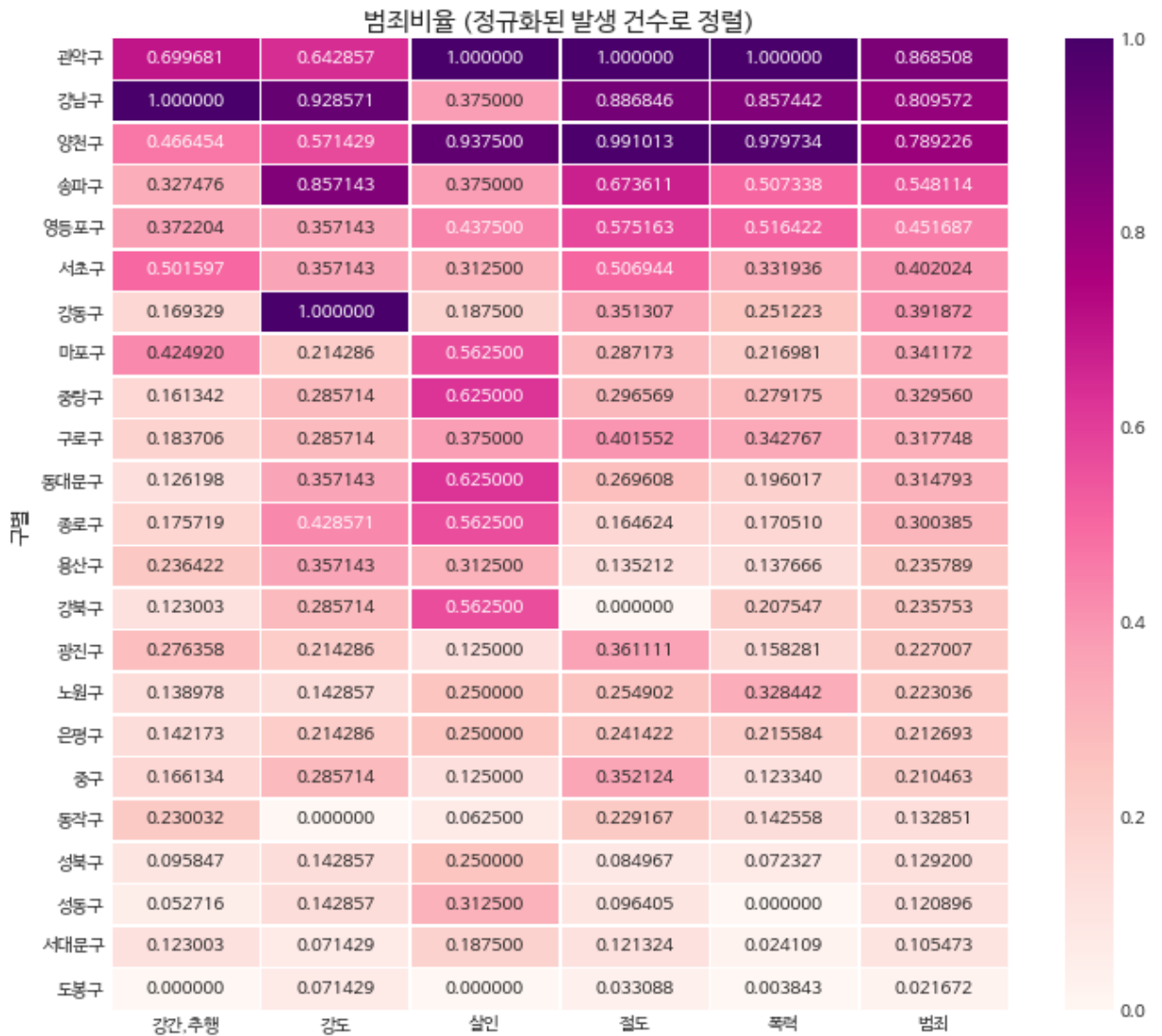


- 절도의 검거율이 다른 범죄의 검거율에 비해 낮은 것을 확인

```
In [ ]: # 히트맵 출력
target_col = ['강간,추행', '강도', '살인', '절도', '폭력', '범죄']

crime_anal_norm['범죄'] = crime_anal_norm['범죄'] / 5
crime_anal_norm_sort = crime_anal_norm.sort_values(by='범죄', ascending=False)

plt.figure(figsize = (10,10))
sns.heatmap(crime_anal_norm_sort[target_col], annot=True, fmt='f', linewidths=.5,
            cmap='RdPu')
plt.title('범죄비율 (정규화된 발생 건수로 정렬)')
plt.show()
```



- 범죄가 가장 많이 일어나는 지역은 관악구, 강남구, 양천구가 높게 관찰된다. 뒤로는 송파, 영등포, 서초가 뒤따르고 있다.

## 2.5 지도 시각화 도구 - Folium

```
In [ ]: #!pip install folium
```

```
In [ ]: import folium
```

```
In [ ]: # json으로 되어있는 서울시 구별 데이터 정보 로딩
import json
geo_path = '/content/drive/MyDrive/스터디/데이터 주무르기/data/02. skorea_municipalities.json'
geo_str = json.load(open(geo_path, encoding='utf-8'))
```

```
In [ ]: #서울시 중심의 위도와 경도 정보를 먼저 입력하고 경계선 설정
map = folium.Map(location=[37.5502, 126.982], zoom_start=11,
                  tiles='Stamen Toner')
# 살인 발생 빈도 추출
map.choropleth(geo_data = geo_str,
               data = crime_anal_norm['살인'],
               columns = [crime_anal_norm.index, crime_anal_norm['살인']],
               fill_color = 'PuRd', #PuRd, YlGnBu
               key_on = 'feature.id')

map
```

/usr/local/lib/python3.7/dist-packages/folium/folium.py:426: FutureWarning: The choropleth method has been deprecated. Instead use the new Choropleth class, which has the same arguments. See the example notebook 'GeoJSON\_and\_choropleth' for how to do this.  
FutureWarning

Out[ ]: Make this Notebook Trusted to load map: File -> Trust Notebook

```
In [ ]: map = folium.Map(location=[37.5502, 126.982], zoom_start=11,
                             tiles='Stamen Toner')
# 강간,추행 발생 빈도 추출
map.choropleth(geo_data = geo_str,
               data = crime_anal_norm['강간,추행'],
               columns = [crime_anal_norm.index, crime_anal_norm['강간,추행']],
               fill_color = 'PuRd', #PuRd, YlGnBu
               key_on = 'feature.id')

map
```

/usr/local/lib/python3.7/dist-packages/folium/folium.py:426: FutureWarning: The choropleth method has been deprecated. Instead use the new Choropleth class, which has the same arguments. See the example notebook 'GeoJSON\_and\_choropleth' for how to do this.  
FutureWarning

Out[ ]: Make this Notebook Trusted to load map: File -> Trust Notebook

```
In [ ]: map = folium.Map(location=[37.5502, 126.982], zoom_start=11,
                        tiles='Stamen Toner')
# 범죄 발생 빈도 추출
map.choropleth(geo_data = geo_str,
               data = crime_anal_norm['범죄'],
               columns = [crime_anal_norm.index, crime_anal_norm['범죄']],
               fill_color = 'PuRd', #PuRd, YlGnBu
               key_on = 'feature.id')

map
```

/usr/local/lib/python3.7/dist-packages/folium/folium.py:426: FutureWarning: The choropleth method has been deprecated. Instead use the new Choropleth class, which has the same arguments. See the example notebook 'GeoJSON\_and\_choropleth' for how to do this.  
FutureWarning

Out[ ]: Make this Notebook Trusted to load map: File -> Trust Notebook

- 강남 지역
- 강서 지역의 범죄 빈도가 높다

**하지만 이는 인구수에 비례해야한다.**

```
In [ ]: tmp_criminal = crime_anal_norm['범죄'] / crime_anal_norm['인구수'] * 1000000

map = folium.Map(location=[37.5502, 126.982], zoom_start=11,
                  tiles='Stamen Toner')

map.choropleth(geo_data = geo_str,
               data = tmp_criminal,
               columns = [crime_anal.index, tmp_criminal],
               fill_color = 'PuRd', #PuRd, YlGnBu
               key_on = 'feature.id')

map
```

/usr/local/lib/python3.7/dist-packages/folium/folium.py:426: FutureWarning: The choropleth method has been deprecated. Instead use the new Choropleth class, which has the same arguments. See the example notebook 'GeoJSON\_and\_choropleth' for how to do this.  
FutureWarning

Out[ ]: Make this Notebook Trusted to load map: File -> Trust Notebook

- 여전히 강서 지역의 범죄율이 높다
- 중구 및 종로구의 범죄율이 높아진게 보인다.

## 2.6 경찰서별 검거현황과 구별 범죄발생 현황을 표현하기

```
In [ ]: df_r['lat'] = station_lat
df_r['lng'] = station_lng

col = ['살인 검거', '강도 검거', '강간,추행 검거', '절도 검거', '폭력 검거']
tmp = df_r[col] / df_r[col].max()

df_r['검거'] = np.sum(tmp, axis=1)

df_r.head()
```

```
Out [ ]:
```

	col	구 분	강간, 추행 검거	강간, 추행 발생	강 도 검거	강 도 발생	살 인 검거	살 인 발생	절도 검거	절도 발생	폭력 검거	폭력 발생	구 별	lat	lng
0		강남	433.0	513.0	10.0	12.0	3.0	5.0	1016.0	1941.0	1911.0	2283.0	강남구	37.509435	127.06695
1		강동	167.0	188.0	14.0	15.0	3.0	3.0	781.0	1640.0	1697.0	1942.0	강동구	37.528511	127.12682
2		강북	132.0	159.0	4.0	5.0	8.0	9.0	490.0	780.0	1625.0	1817.0	강북구	37.637304	127.02734
3		강서	222.0	265.0	5.0	6.0	9.0	10.0	1070.0	1689.0	2087.0	2445.0	양천구	37.539783	126.82995
4		관악	351.0	403.0	3.0	3.0	12.0	12.0	1137.0	2229.0	2355.0	2614.0	관악구	37.474395	126.95134

```
In [ ]: # 경찰서의 위치 확인
map = folium.Map(location=[37.5502, 126.982], zoom_start=11)
```

```
for n in df_r.index:
    folium.Marker([df_r['lat'][n],
                    df_r['lng'][n]]).add_to(map)

map
```

Out[ ]: Make this Notebook Trusted to load map: File -> Trust Notebook

```
In [ ]: # 경찰서의 검거율을 원의 넓이로 표현
map = folium.Map(location=[37.5502, 126.982], zoom_start=11)

for n in df_r.index:
    folium.CircleMarker([df_r['lat'][n], df_r['lng'][n]],
                        radius = df_r['검거'][n]*10,
                        color='#3186cc', fill_color='#3186cc', fill=True).add_to(map)

map
```

Out[ ]: Make this Notebook Trusted to load map: File -> Trust Notebook

```
In [ ]: map = folium.Map(location=[37.5502, 126.982], zoom_start=11)
```



```

map.choropleth(geo_data = geo_str,
               data = crime_anal_norm['범죄'],
               columns = [crime_anal_norm.index, crime_anal_norm['범죄']],
               fill_color = 'PuRd', #PuRd, YlGnBu
               key_on = 'feature.id')

for n in df_r.index:
    folium.CircleMarker([df_r['lat'][n], df_r['lng'][n]],
                        radius = df_r['거리'][n]*10,
                        color='#3186cc', fill_color='#3186cc', fill=True).add_to(map)

map

```

/usr/local/lib/python3.7/dist-packages/folium/folium.py:426: FutureWarning: The choropleth method has been deprecated. Instead use the new Choropleth class, which has the same arguments. See the example notebook 'GeoJSON\_and\_choropleth' for how to do this.  
FutureWarning

Out[ ]: Make this Notebook Trusted to load map: File -> Trust Notebook