

Semi-project

2팀 : 서태원, 김예린, 김의중, 장윤식, 조희정

CONTENTS



분석 주제 및 목표



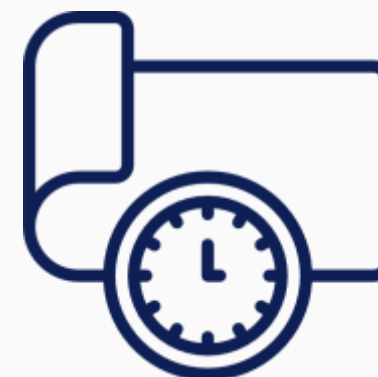
데이터 수집 방법



분석 프로세스

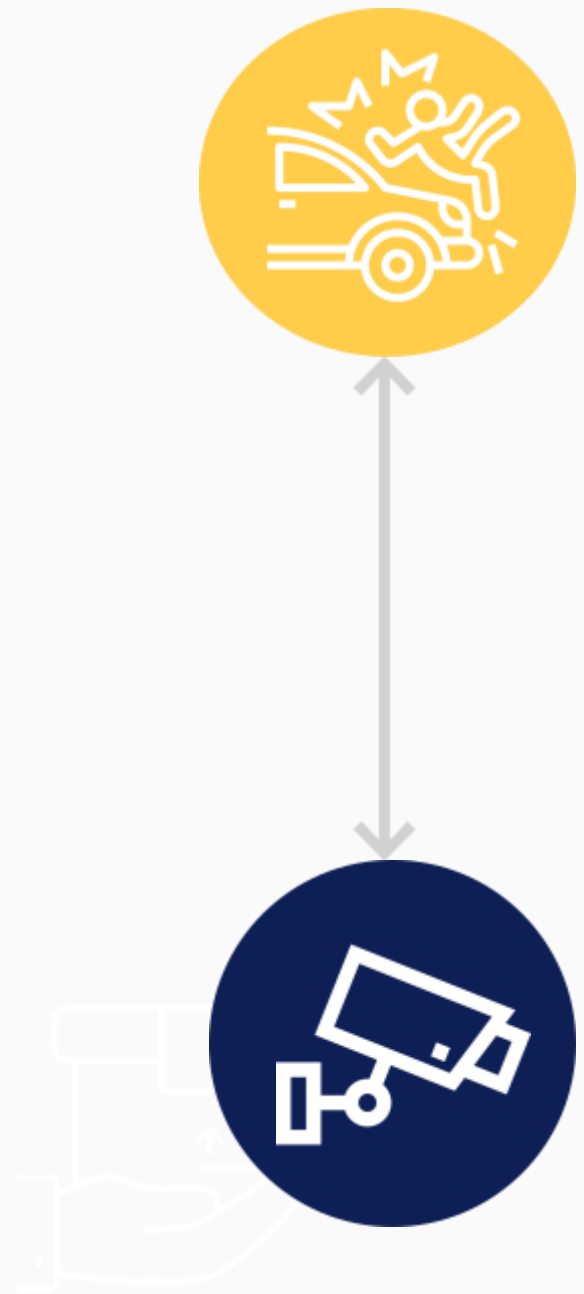


분석 결과



후속 과제

분석 주제 및 목표



서울시 각 구별 뺑소니 건수와
CCTV 설치 개수와의 연관성

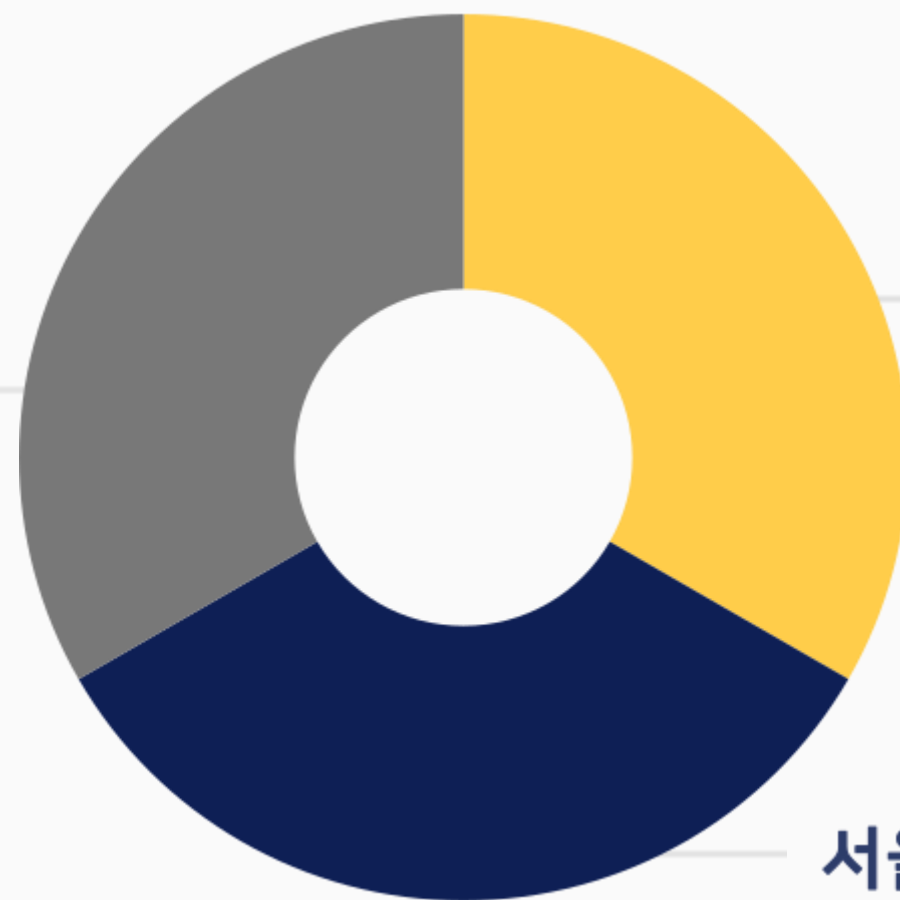


CCTV가 많을수록 뺑소니 건수가 적을것이다



데이터 수집 방법

서울시 자치구 년도별
CCTV 설치 현황
(서울 열린데이터 광장)



지자체별
대상사고통계 API
(교통사고정보 개방 시스템)

서울시 주민등록인구
(구별) 통계
(서울 열린데이터 광장)

데이터 수집 방법

서울시 자치구 년도별 CCTV 설치 현황

구분	총계	2012년 이	2012년	2013년	2014년	2015년	2016년
계	77,032	7,667	2,200	3,491	4,439	6,582	8,129
종로구	1,772	813	0	0	210	150	1
중 구	2,333	16	114	87	77	236	240
용산구	2,383	34	71	234	125	221	298
성동구	3,602	448	125	212	105	339	310
광진구	2,588	35	57	100	187	98	52
동대문구	2,497	1,090	146	60	29	111	233
중랑구	3,296	302	24	253	88	141	161
성북구	3,958	83	78	170	230	323	594
강북구	2,462	0	0	24	65	105	243
도봉구	1,629	39	22	96	181	79	159
노원구	2,415	0	97	193	77	516	331
은평구	3,791	14	3	44	332	329	555
서대문구	2,940	730	253	155	108	137	223
마포구	2,421	58	93	77	61	162	454
양천구	3,312	1,413	167	181	143	180	350
강서구	2,560	55	129	106	213	188	168
구로구	4,075	852	216	349	187	268	326
금천구	2,374	0	0	178	80	361	133
영등포구	3,778	572	136	238	123	209	248

서울시 주민등록인구 (구별) 통계

기간	자치구	세대	합계		
			계	남자	여자
2016	합계	4,189,839	10,204,057	5,007,379	5,196,678
	종로구	72,118	161,922	79,322	82,600
	중구	59,651	134,409	66,452	67,957
	용산구	106,607	245,102	119,636	125,466
	성동구	128,727	307,161	152,006	155,155
	광진구	158,390	372,104	181,251	190,853
	동대문구	159,426	370,312	183,444	186,868
	중랑구	177,400	415,677	206,772	208,905
	성북구	188,304	461,617	224,504	237,113
	강북구	141,229	330,704	162,211	168,493
	도봉구	136,898	350,272	172,006	178,266
	노원구	219,736	571,212	277,866	293,346
	은평구	201,891	495,937	241,048	254,889
	서대문구	136,766	325,871	156,597	169,274
	마포구	169,432	390,887	186,668	204,219
	양천구	177,434	481,845	238,163	243,682
	강서구	245,045	602,104	293,734	308,370
	구로구	172,220	449,600	225,411	224,189
	금천구	104,483	254,654	130,429	124,225
	영등포구	165,729	406,779	204,712	202,067

분석 프로세스



데이터 추출

뺑소니 사건현황 데이터

- ▶ 반복문으로 연도별(2016~2021)
서울시의 데이터만 추출



데이터 가공 및 분석 변수 처리

뺑소니 사건 현황 데이터

- ▶ 필요없는 열 데이터삭제
- ### 서울시 CCTV 데이터
- ▶ 신규 CCTV 발생 건수를
누적 데이터로 변환



데이터 결합

인구수 대비 뺑소니 사고 건수
연도별 데이터 열방향으로 결합



데이터시각화

- heatmap(seaborn)
- jointplot(seaborn)
- plot(matplotlib)
- rcParams(matplotlib)
- choropleth map(folium)
- pie(matplotlib)



분석 프로세스

```
arr_col = df_year.columns.values

df_year[arr_col] = df_year[arr_col].cumsum()

df_year[arr_col]
```

- ▶ cumsum함수 사용해서 cctv갯수 구한부분

```
cctv_columns = df_cctv.columns.values
for column in cctv_columns:
    df_cctv[column] = df_cctv[column].str.replace(',', '')
    df_cctv[column] = df_cctv[column].apply(pd.to_numeric)
```

- ▶ 모든 데이터 정수(int64)변환,
, '콤마 제거(ex)1,000→ 1000)

```
#df_cctv를 df_year와 합치기, df_year - df_year

list_new_cctv = []
for f_yyyy in ['2016', '2017', '2018', '2019', '2020']:
    for idx, row in df_cctv.iterrows():

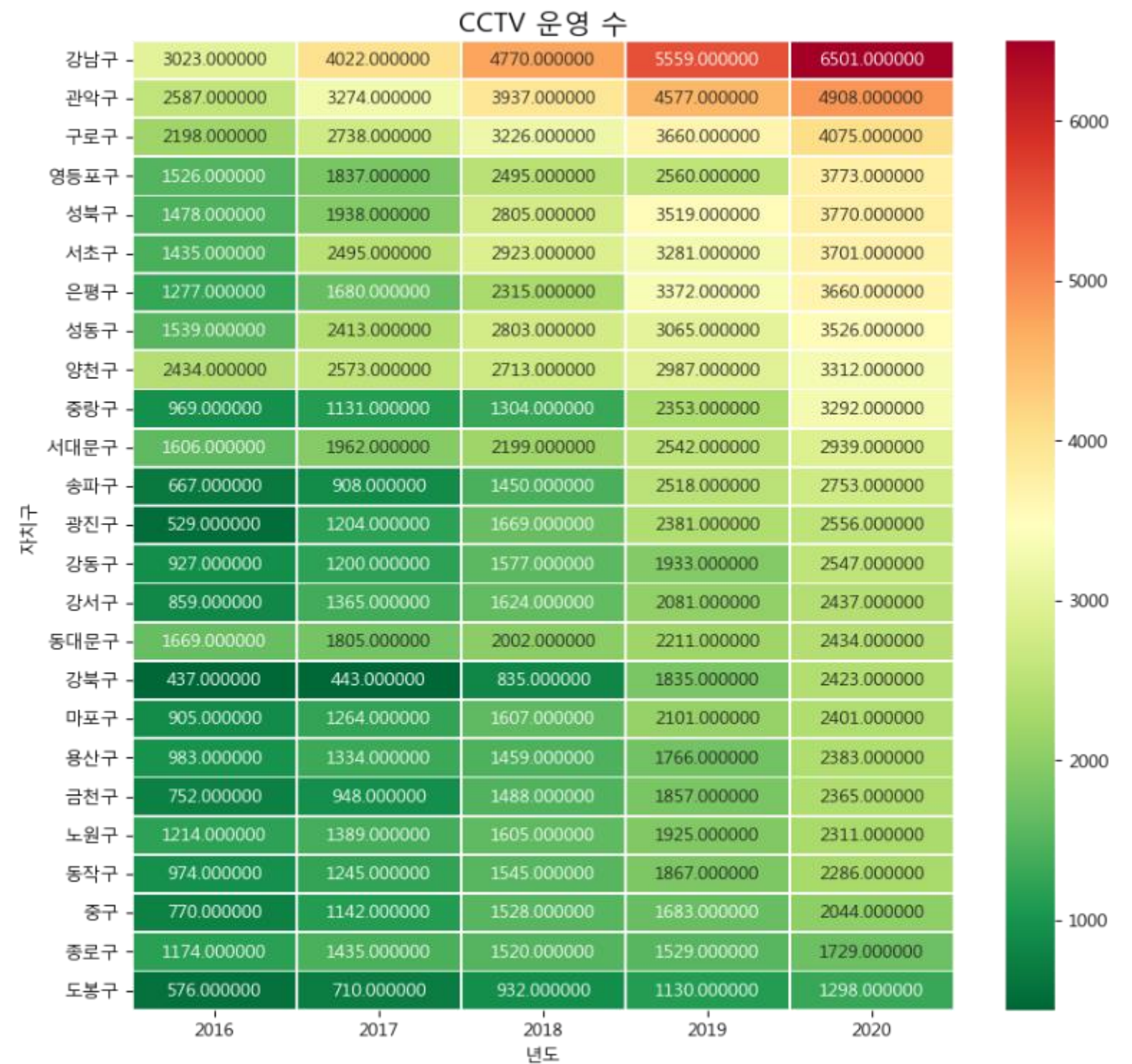
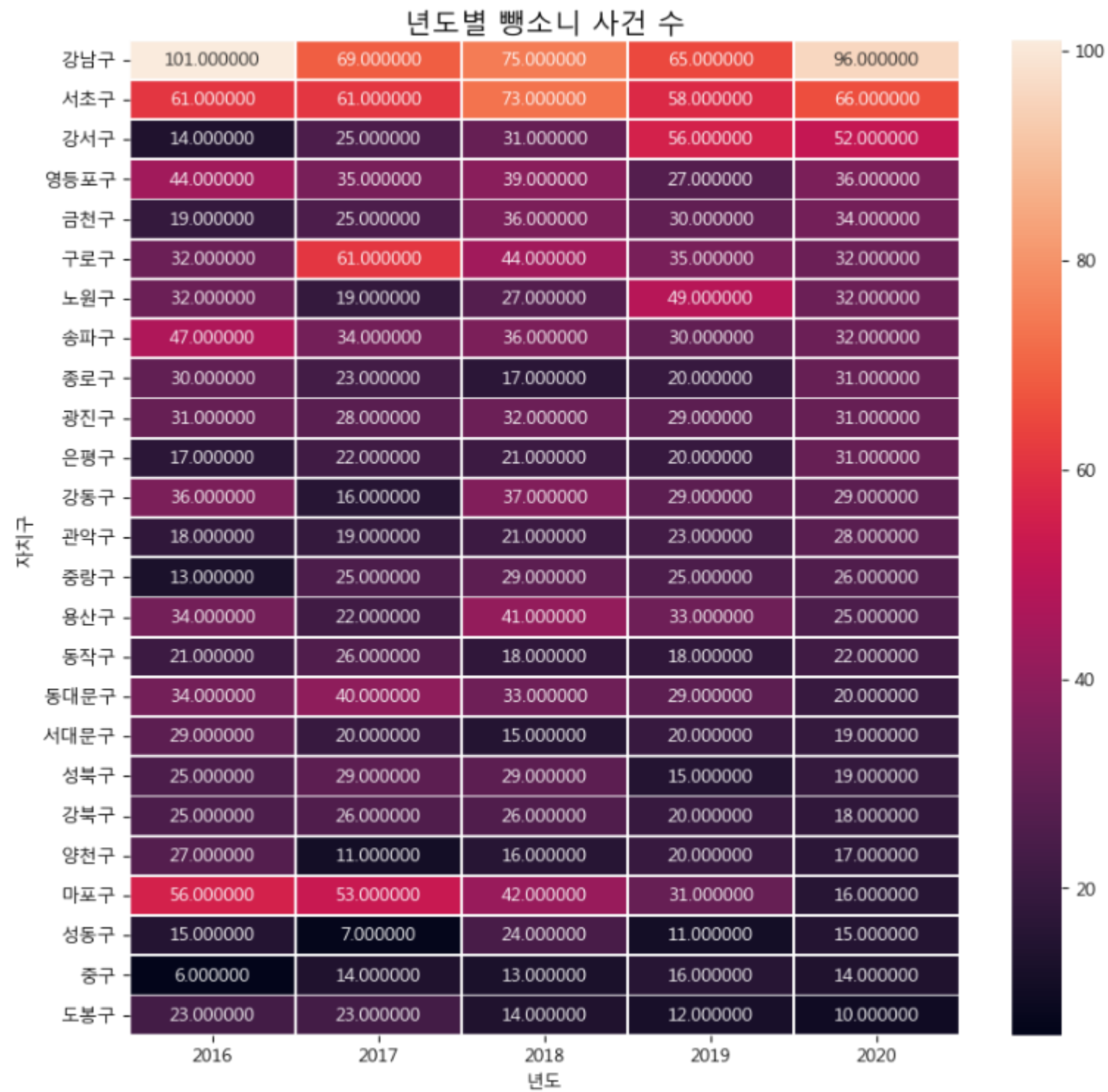
        dict_cctv = {
            '년' : f_yyyy,
            '구' : idx,
            'CCTV 갯수' : row[f_yyyy]
        }
        list_new_cctv.append(dict_cctv)

df_new_cctv = pd.DataFrame(list_new_cctv)
df_new_cctv
```

- ▶ 기존 데이터프레임을 iterrows함수를 통해
새로운 데이터 프레임 만들기

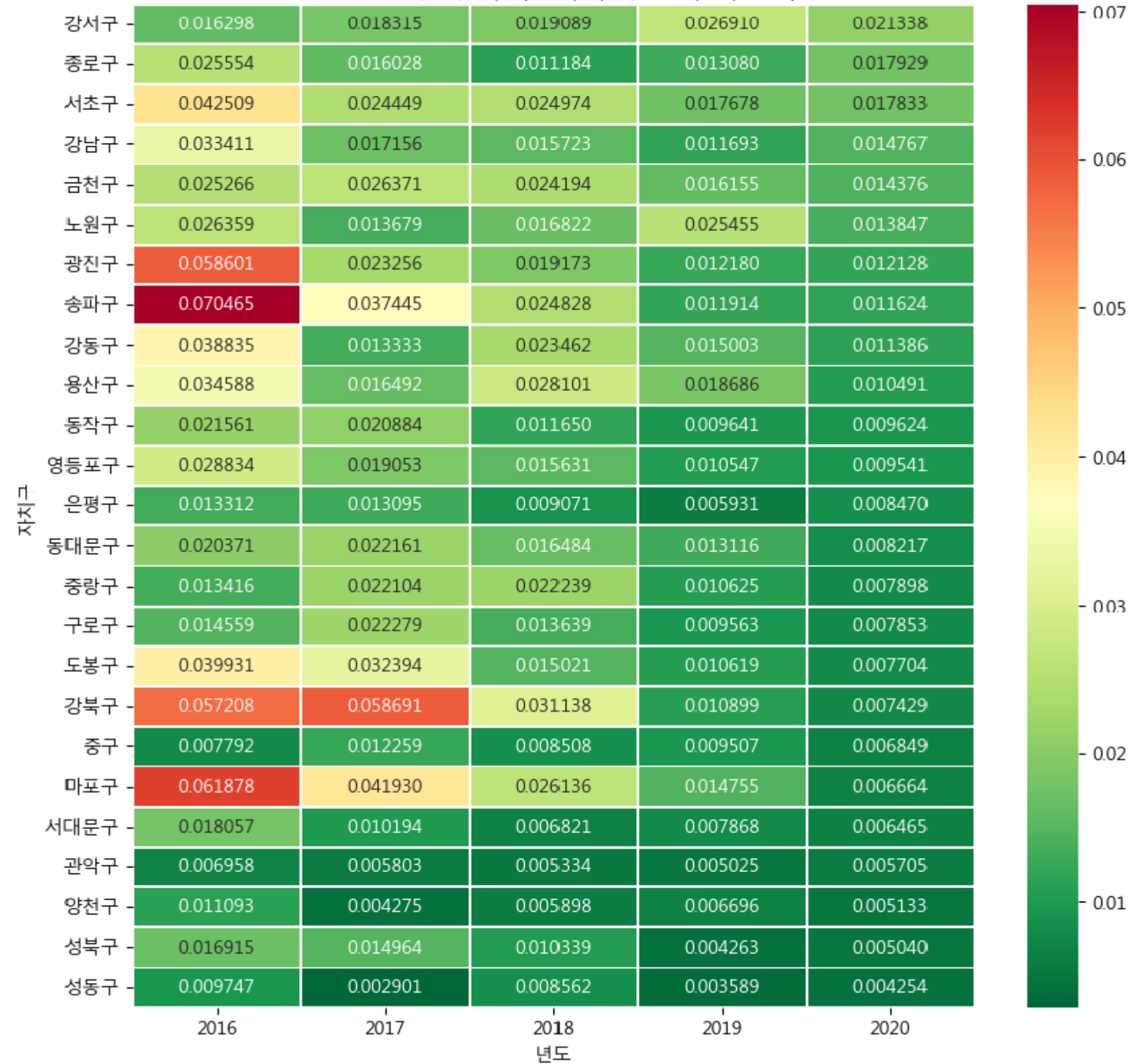
분석 결과

시각화



분석 결과

CCTV 운영 장비 수 대비 뺑소니 사건 수



분석 결과

산점도 분석

```
#독립표본 2016년 행사나 & 2016년 cctv
```

```
stats.ttest_ind(df_tot_2016["횟수"], df_tot_2016["CCTV 갯수"])
```

```
Ttest_indResult(statistic=-9.461060880318874, pvalue=1.5044704755337304e-12)
```

```
#독립표본 2017년 행사나 & 2017년 cctv
```

```
stats.ttest_ind(df_tot_2017["횟수"], df_tot_2017["CCTV 갯수"])
```

```
Ttest_indResult(statistic=-10.001797294846918, pvalue=2.518048059917295e-13)
```

```
#독립표본 2018년 행사나 & 2018년 cctv
```

```
stats.ttest_ind(df_tot_2018["횟수"], df_tot_2018["CCTV 갯수"])
```

```
Ttest_indResult(statistic=-11.0217905739632, pvalue=9.526862941424921e-15)
```

```
#독립표본 2019년 행사나 & 2019년 cctv
```

```
stats.ttest_ind(df_tot_2019["횟수"], df_tot_2019["CCTV 갯수"])
```

```
Ttest_indResult(statistic=-12.647776812981668, pvalue=6.78313604865569e-17)
```

```
#독립표본 2020년 행사나 & 2020년 cctv
```

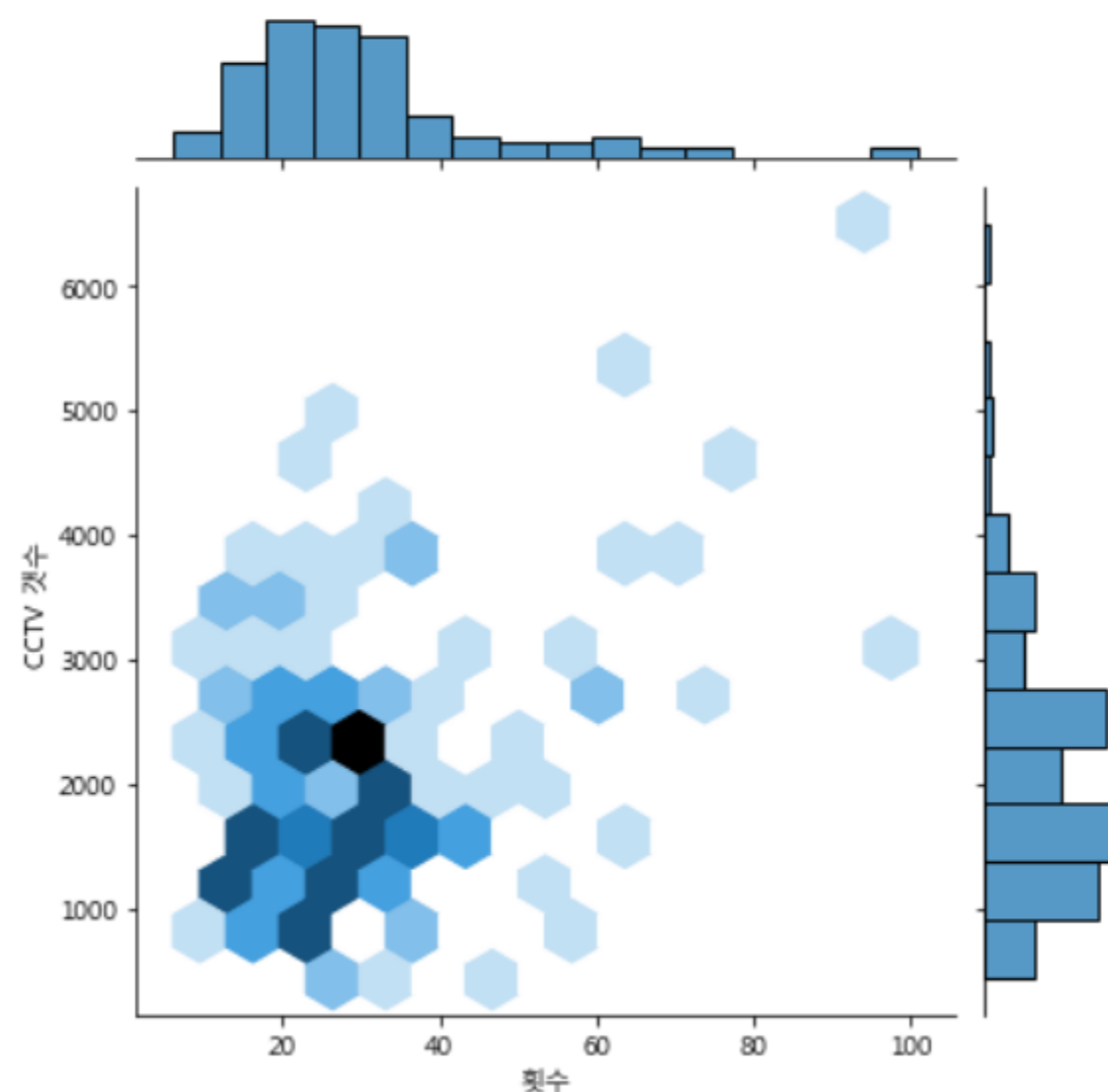
```
stats.ttest_ind(df_tot_2020["횟수"], df_tot_2020["CCTV 갯수"])
```

```
Ttest_indResult(statistic=-13.675832722920964, pvalue=3.542932962562021e-18)
```

```
import scipy as sp #과학계산을 위해 사용# 통계학기법사용시 활용
from scipy import stats # statistics
import seaborn as sns
```

```
sns.jointplot(x=df_tot['횟수'], y=df_tot['CCTV 갯수'], kind='hex')
```

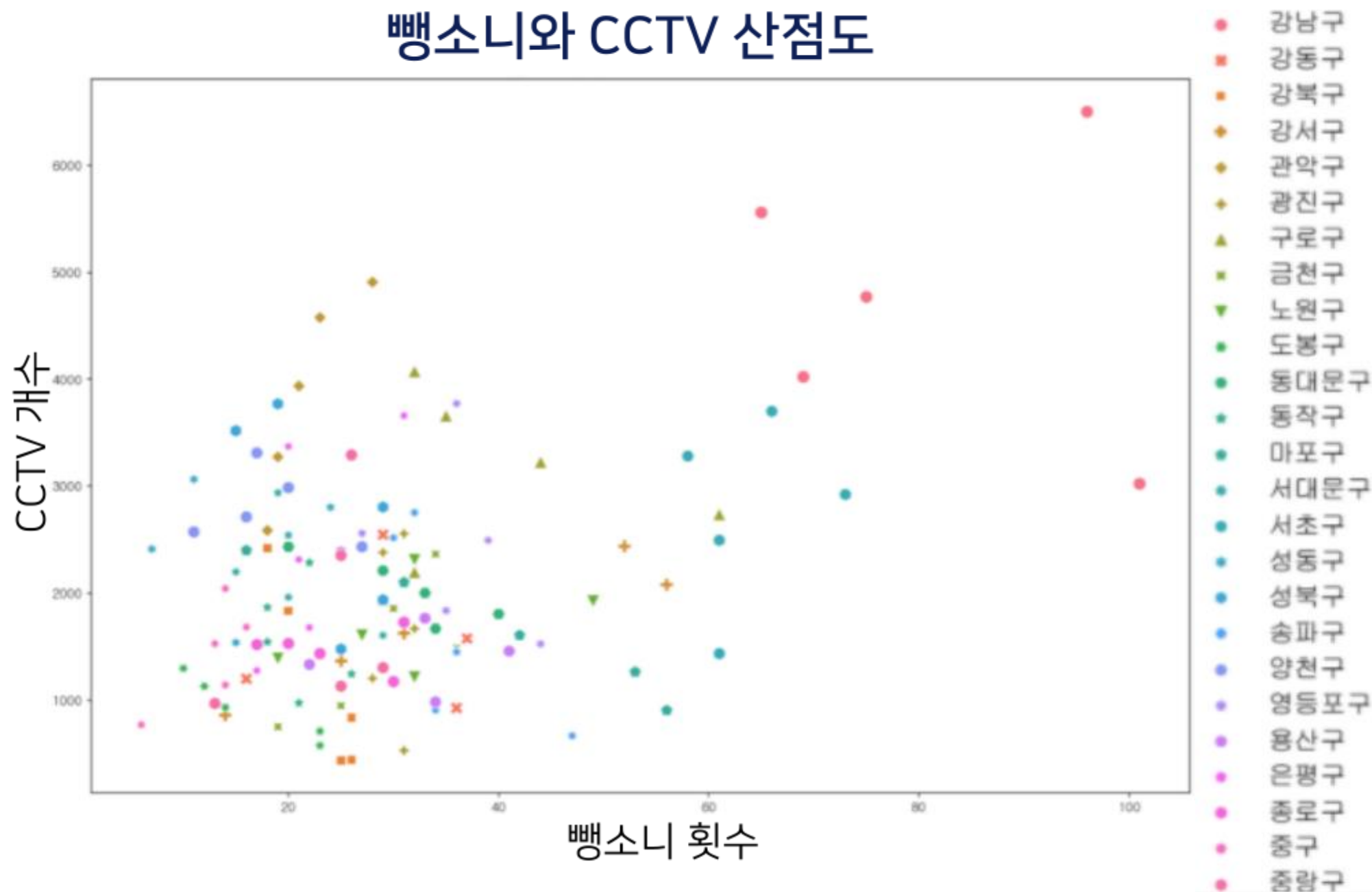
<seaborn.axisgrid.JointGrid at 0xbf8b908>



분석 결과

```
for name, group in groups :  
    sns.scatterplot(x=group['횃수'],  
                   y=group['CCTV 갯수'],  
                   hue='자치구', # different colors by group  
                   style='자치구', # different shapes by group  
                   s=150, # marker size  
                   data=df_acc_cctv,  
                   legend=False)  
  
plt.legend(fontsize=15, loc='upper left', bbox_to_anchor=(1.05, 1))  
plt.title('뽕소니와 CCTV 산점도', fontsize=20)  
plt.xlabel('뽕소니 횃수', fontsize=14)  
plt.ylabel('CCTV 갯수', fontsize=14)  
  
plt.show()
```

뽕소니와 CCTV 산점도



분석 결과



2016



2017



2018

분석 결과



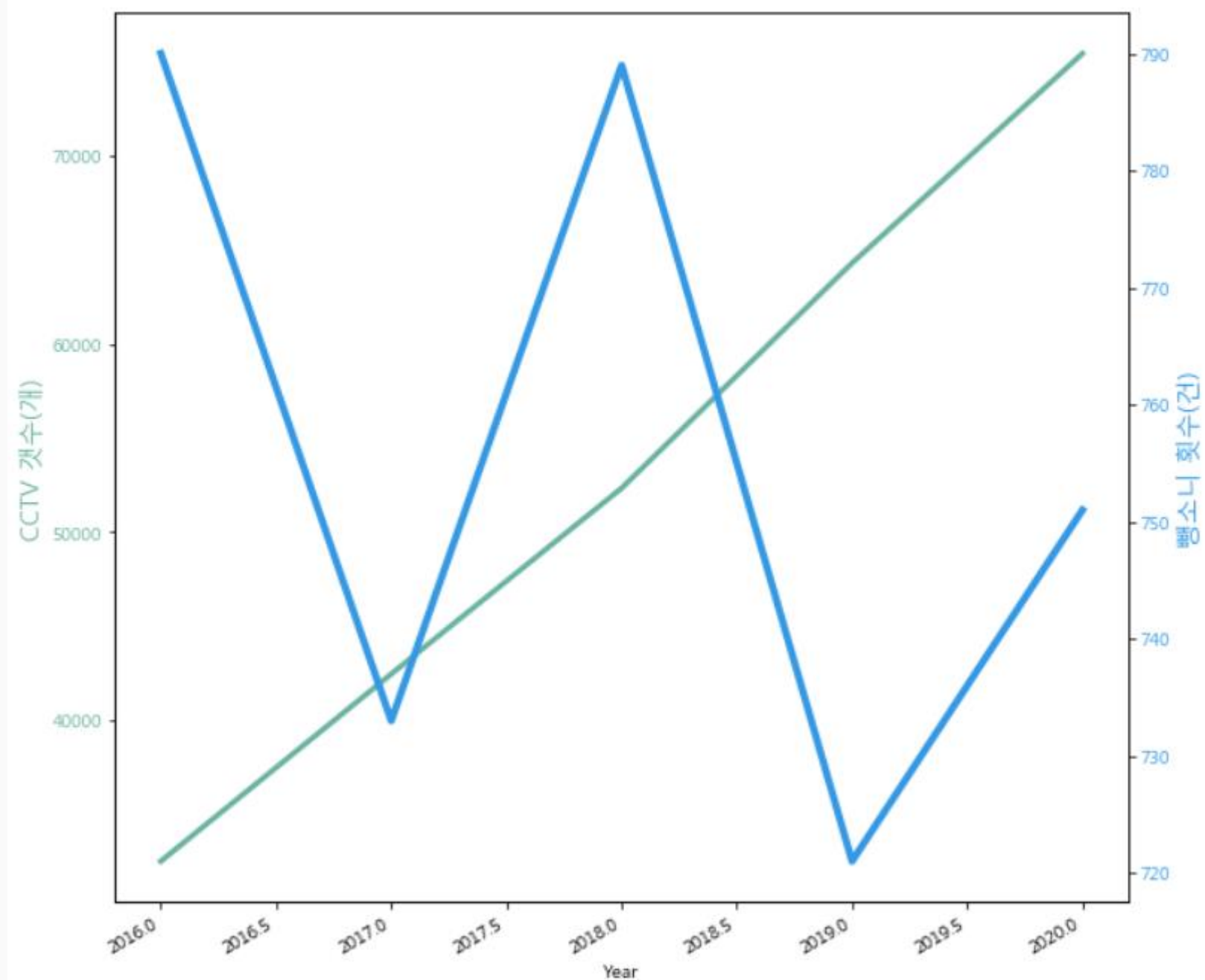
2019



2020

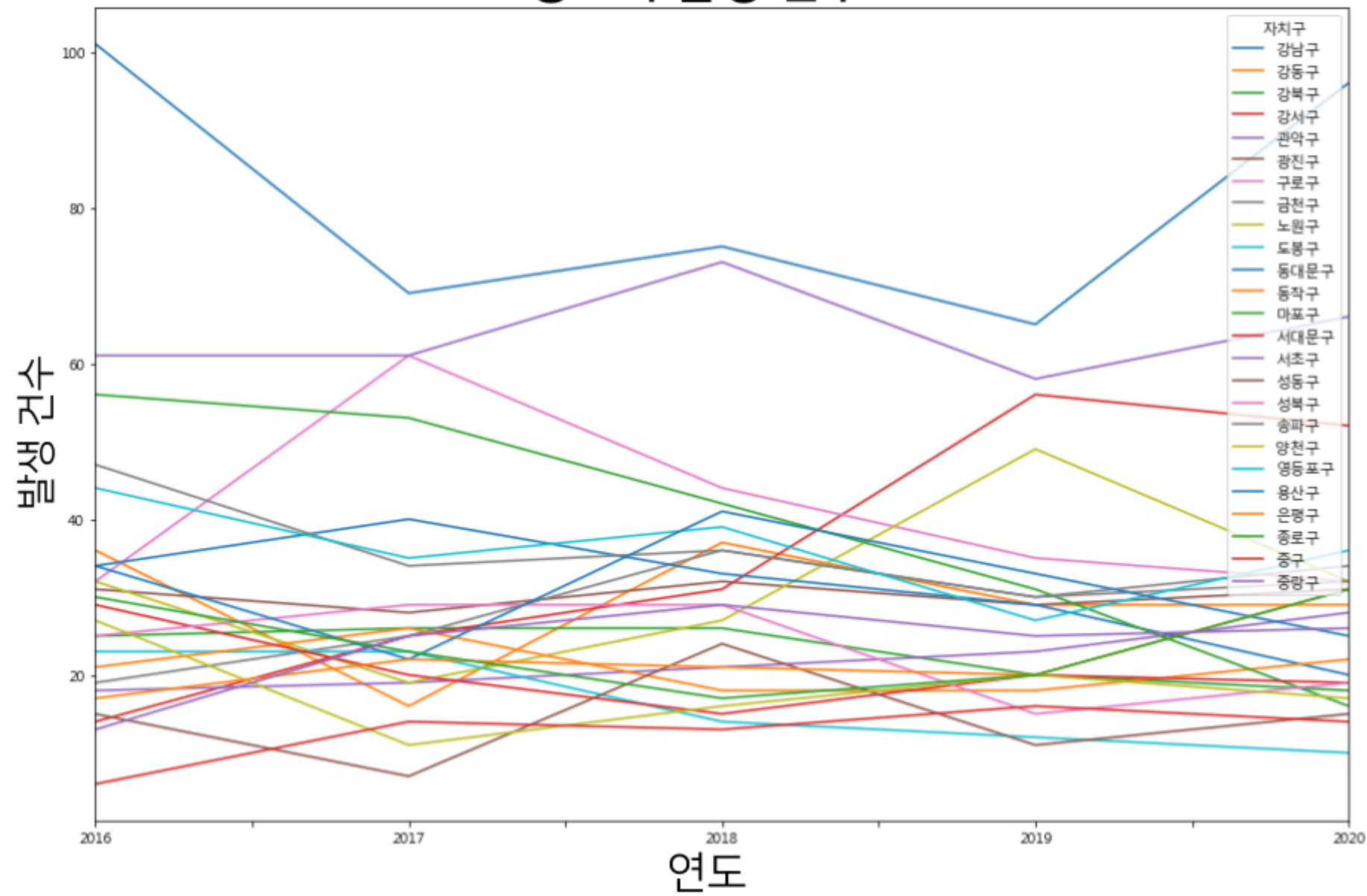
분석 결과

5개년 CCTV와 뺑소니자료의 변화추이

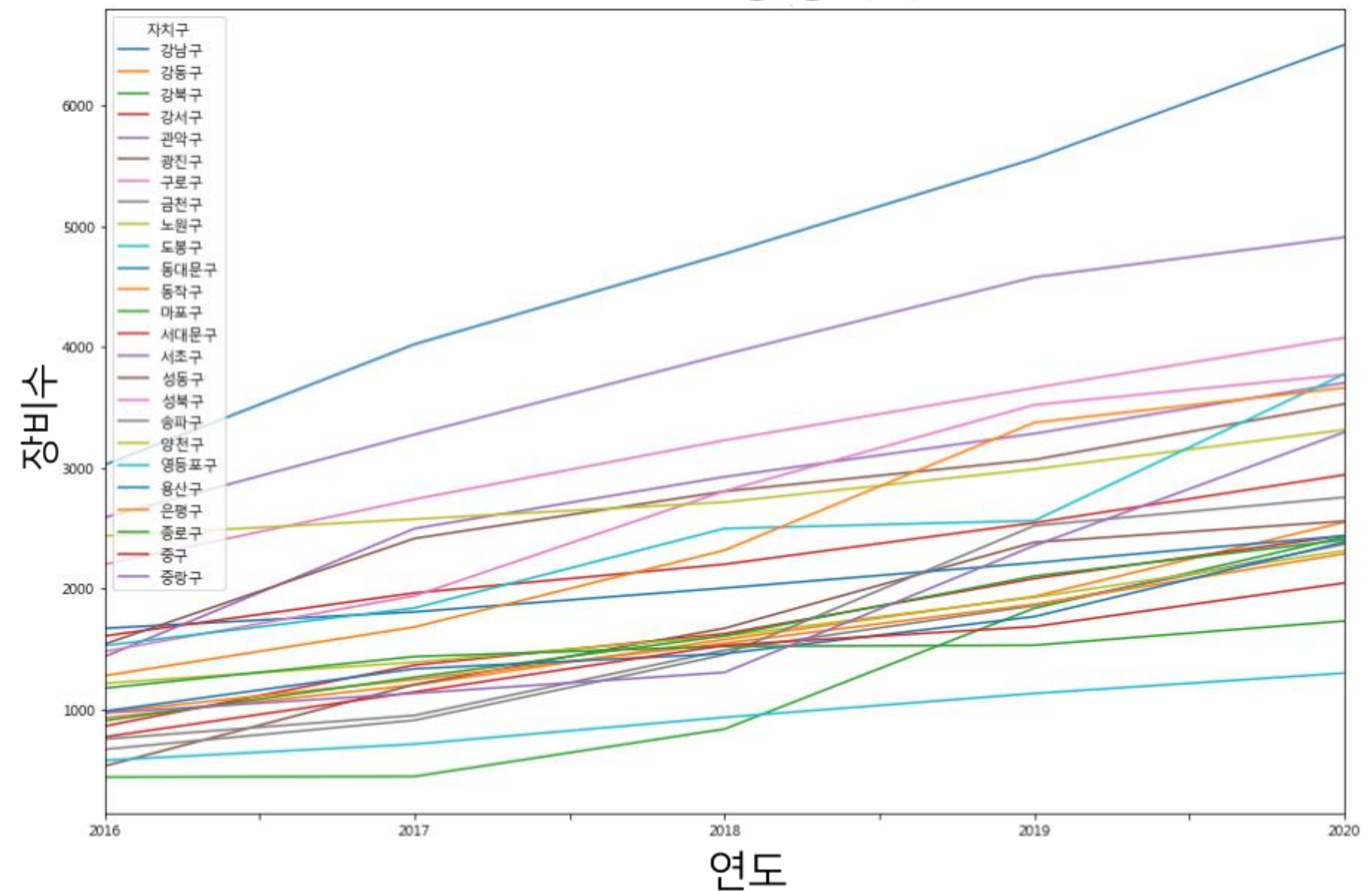


분석 결과

뺑소니 발생 건수



CCTV 운영 장비 수

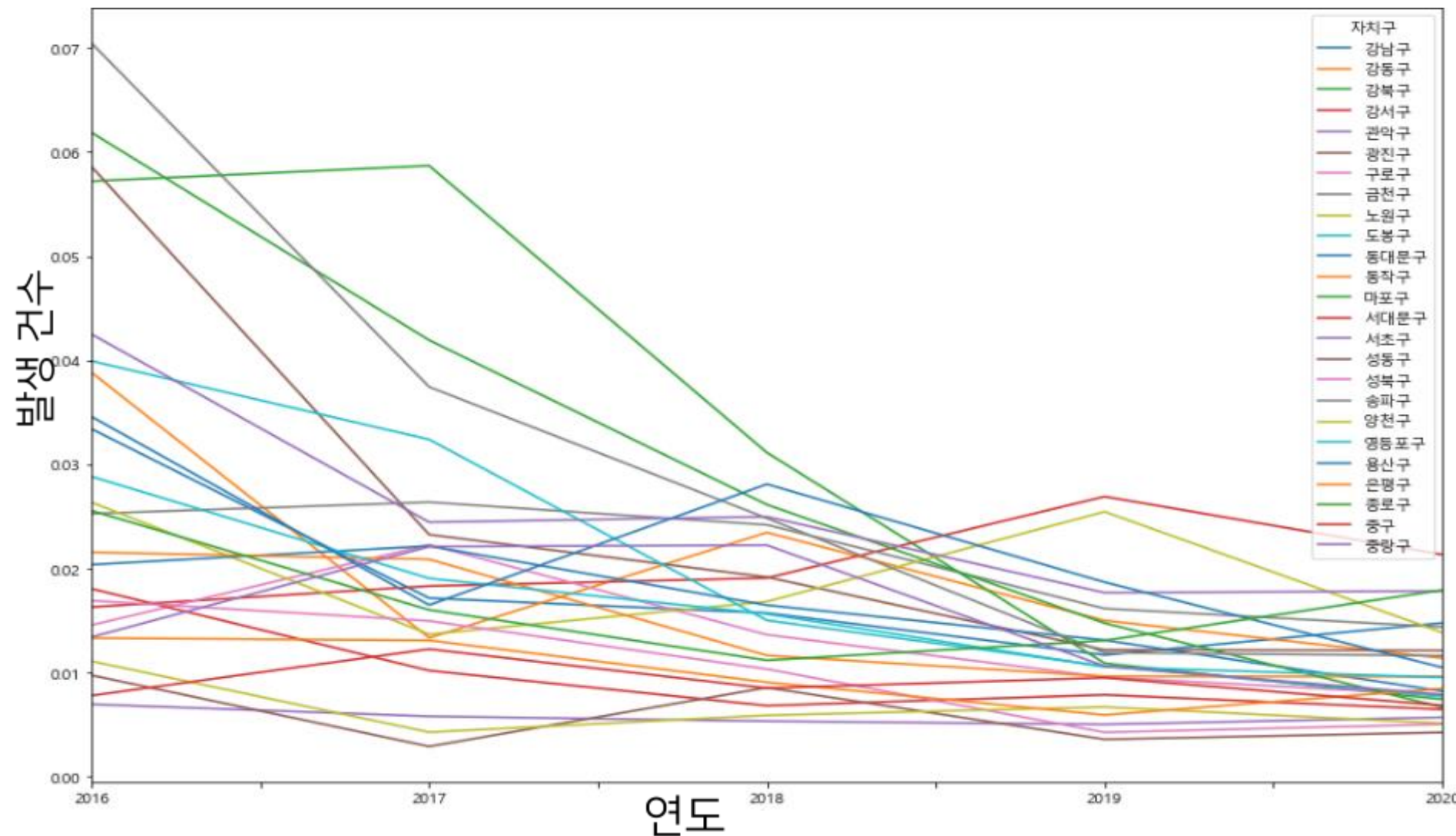


분석 결과

두 데이터간의 피어슨 상관관계분석

```
year_data={'년도':year_index, 'CCTV':df_tot_CCTV, '뺑소니':df_tot_run}  
year_data  
df_year_data=pd.DataFrame(year_data)  
pivot_year = pd.pivot_table(df_year_data, index='년도', aggfunc=np.sum)  
pivot_year
```

CCTV 운영 장비 수 대비 뺑소니 사건 수



corr_df.head()

	년도	자치구	횟수	CCTV 갯수
0	2016	강남구	101	3023
1	2016	강동구	36	927
2	2016	강북구	25	437
3	2016	강서구	14	859
4	2016	관악구	18	2587

corr_df.corr()

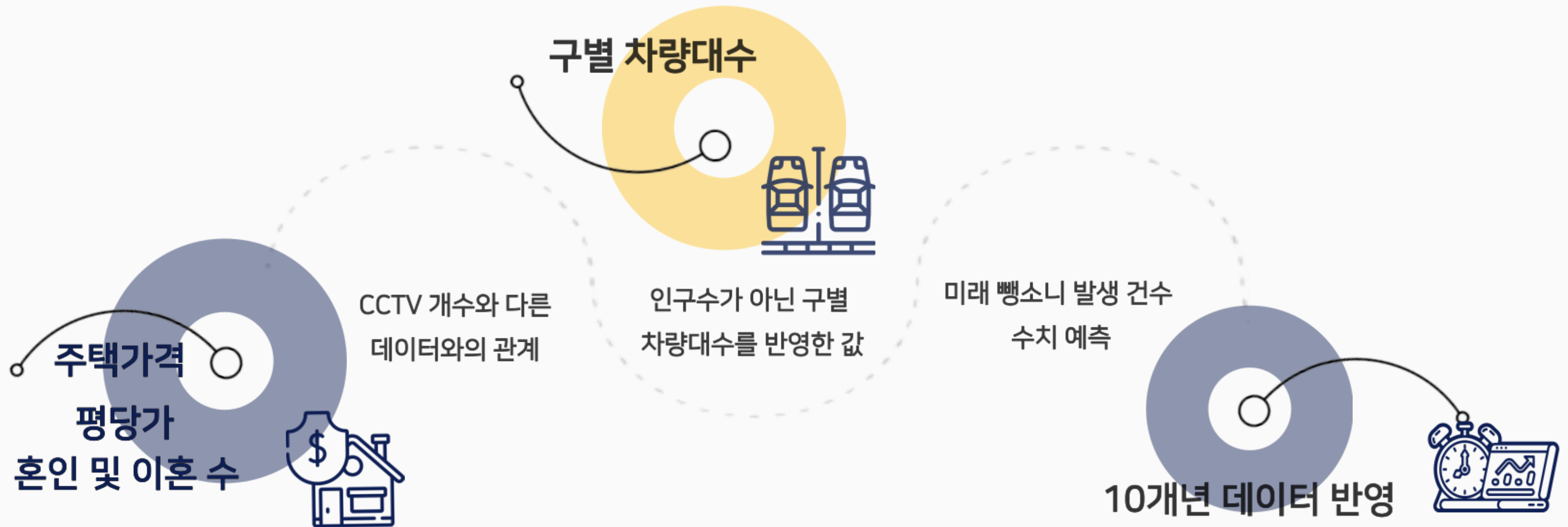
	횟수	CCTV 갯수
횟수	1.000000	0.356749
CCTV 갯수	0.356749	1.000000

CCTV 뺑소니

년도	CCTV	뺑소니
2016	32508	790
2017	42455	733
2018	52331	789
2019	64292	721
2020	75424	751

CCTV와 뺑소니 발생건수는 약한 양의 상관관계

후속 과제



THANK YOU!