

범죄 발생에 영향을 미치는 사회적 요인

<서울시 지역구 단위를 중심으로>

과 목 명 컴퓨팅 기초

지도교수 변해선 교수님

팀 명 유민석, 김민우 팀

학 과 / 이 름 자연과학대학 화학부 김민우, 유민석



연구배경 및 필요성

- 최근 신림동 칼부림 사건, 부산 돌러차기 사건 등 도심에서 일어나는 다양한 범죄가 이슈가 되고 있다.

- 이러한 상황에 경찰 및 치안 인력은 한정되어 있으므로 더욱 효율적인 인원 배분이 필요하고, 더 나아가 범죄 발생 자체를 줄이려는 노력이 필요할 것이다.

- 따라서 본 연구에서는 범죄발생에 영향을 미칠 수 있는 사회적 요인과 범죄 발생 사이의 관계에 대해 분석함으로써 도시 범죄 발생을 줄이기 위한 방법을 제안하고자 한다.



출처 : 조선일보, "부산 돌러차기 가해자, 다른 주거침입 사건 벌금액 줄어" 2023.10.27
https://www.chosun.com/national/regional/yeongnam/2023/10/27/HZGF MFAXRBBRTE7YGPJ6INHFGQ/?utm_source=naver&utm_medium=referral&utm_campaign=naver-news

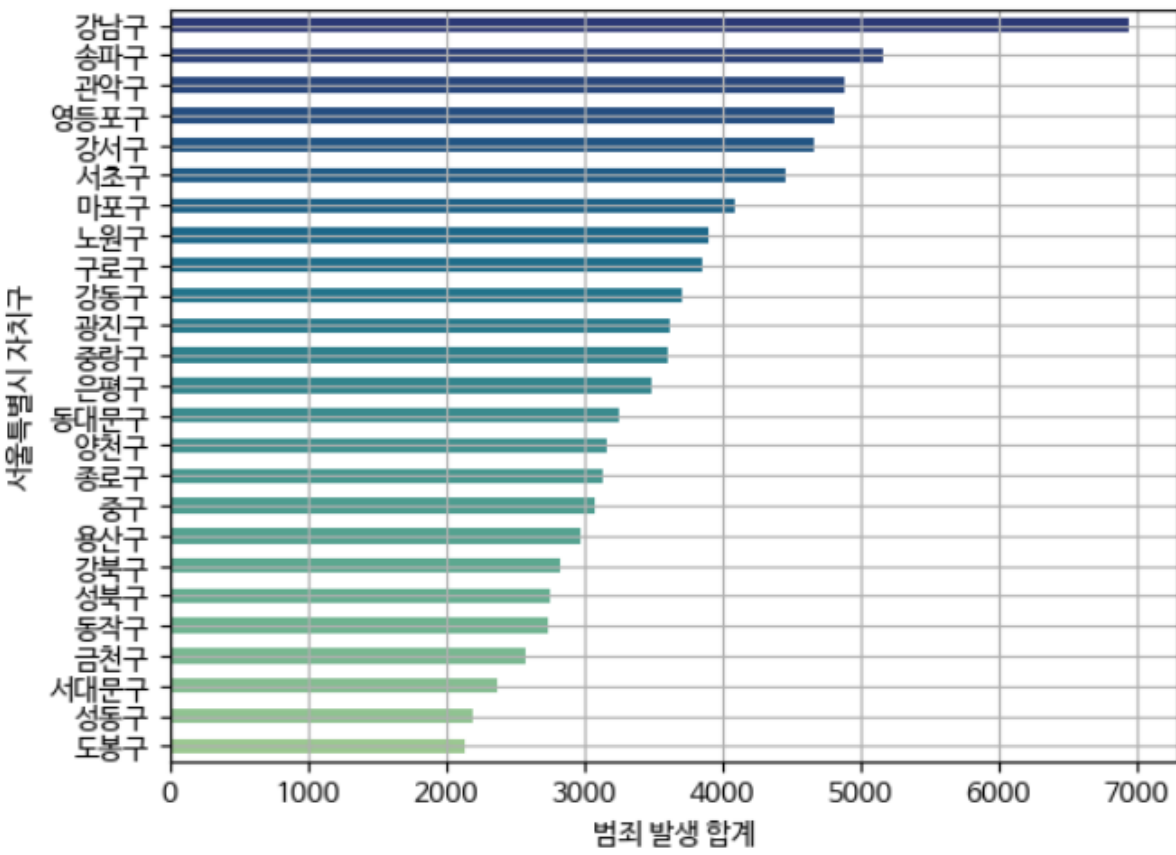


프로젝트 진행과정

① 가설 설정

- 범죄 발생에 영향을 미칠 것으로 생각되는 요인들을 선정, 알맞은 가설을 수립

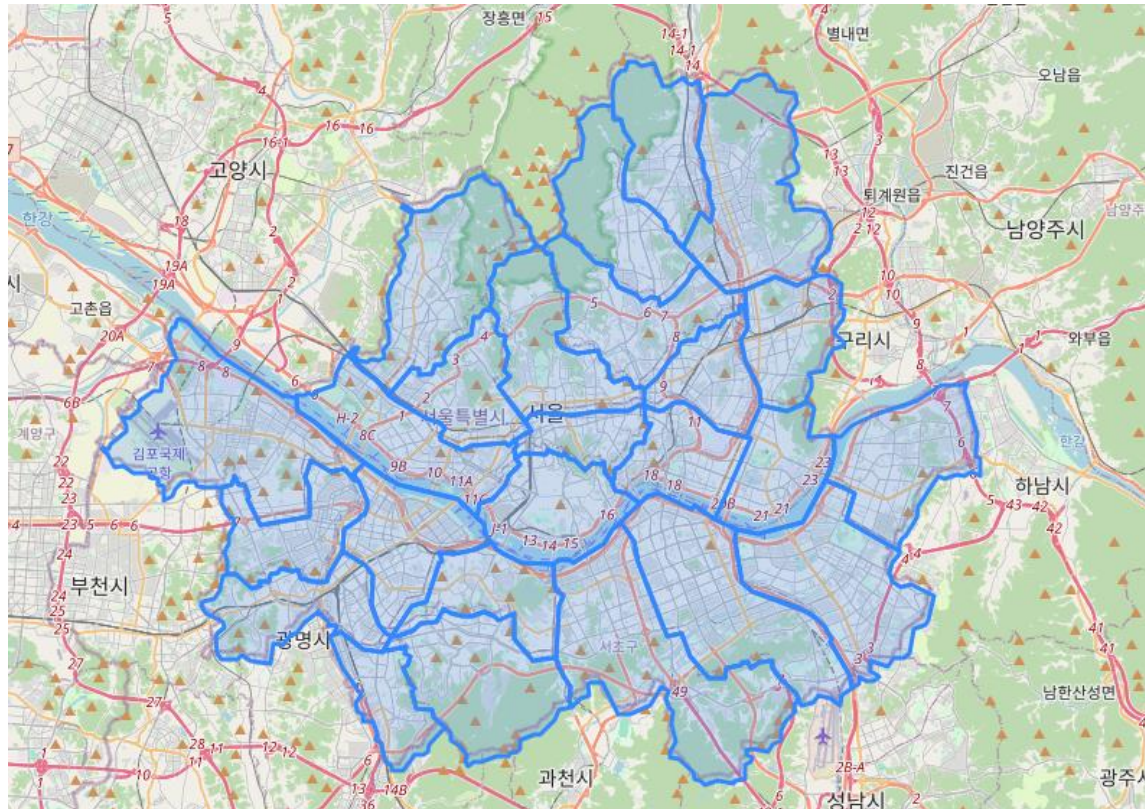
ex) 사람들의 소득 수준이 낮을 수록 범죄발생 수가 많을 것이다.



② 분석 범위 설정

- 가설을 뒷받침해줄 데이터의 수집 범위와 구분 단위를 설정

ex) (전국 / 시도), (서울 / 자치구)



③ 데이터 수집

- 각 요인들을 대표할 수 있는 지표 선정, 공공데이터 사이트를 이용한 데이터 수집

- 서울시 5대 범죄 발생현황 통계 : 서울 열린데이터 광장 (2022년 통계)
- 서울시 인구밀도 (구별) 통계 : 서울 열린데이터 광장 (2022년 통계)
- 서울시 외국인주민 총괄(구별) 통계 : 서울 열린데이터 광장 (2022년 통계)
- 서울시 자치구 연도별 방범용 CCTV 운영 현황 : 서울 열린데이터 광장 (2015년 ~ 2023년 통계)
- 서울시 시내주요기관 (경찰·소방관서) 통계 : 서울 열린데이터 광장 (1988년 ~ 2022년 통계)
- 서울시 주택가격지수 (매매) 통계 : 서울 열린데이터 광장 (1986년 ~ 2022년 통계)

④ 데이터 전처리

- Pandas 등의 모듈을 이용한 데이터 전처리

```
1 매매_df.head(5)
```

	시점	자치구별(1)	자치구별(2)	종합	아파트
0	1986	서울시	소계	34.0	20.1
1	1986	서울시	강북구	47.6	28.9
2	1986	서울시	강남구	25.6	17.1
3	1987	서울시	소계	34.7	21.1
4	1987	서울시	강북구	48.2	28.8

```
1 crime_price.head()
```

	자치구별(2)	주택가격지수	범죄발생합계
	강남구	101.2	6947
	강동구	99.0	3711
	강북구	98.8	2832
	강서구	98.6	4663
	관악구	98.7	4879

```
1 crime_df.head()
```

	자치구별(2)	살인 발생	강도 발생	강간/강제추행 발생	절도 발생	폭력 발생
0	강남구	12	11	667	2495	3762
1	강동구	5	5	157	1672	1872
2	강북구	5	5	177	872	1773
3	강서구	3	7	261	1991	2401
4	관악구	10	9	327	1966	2567

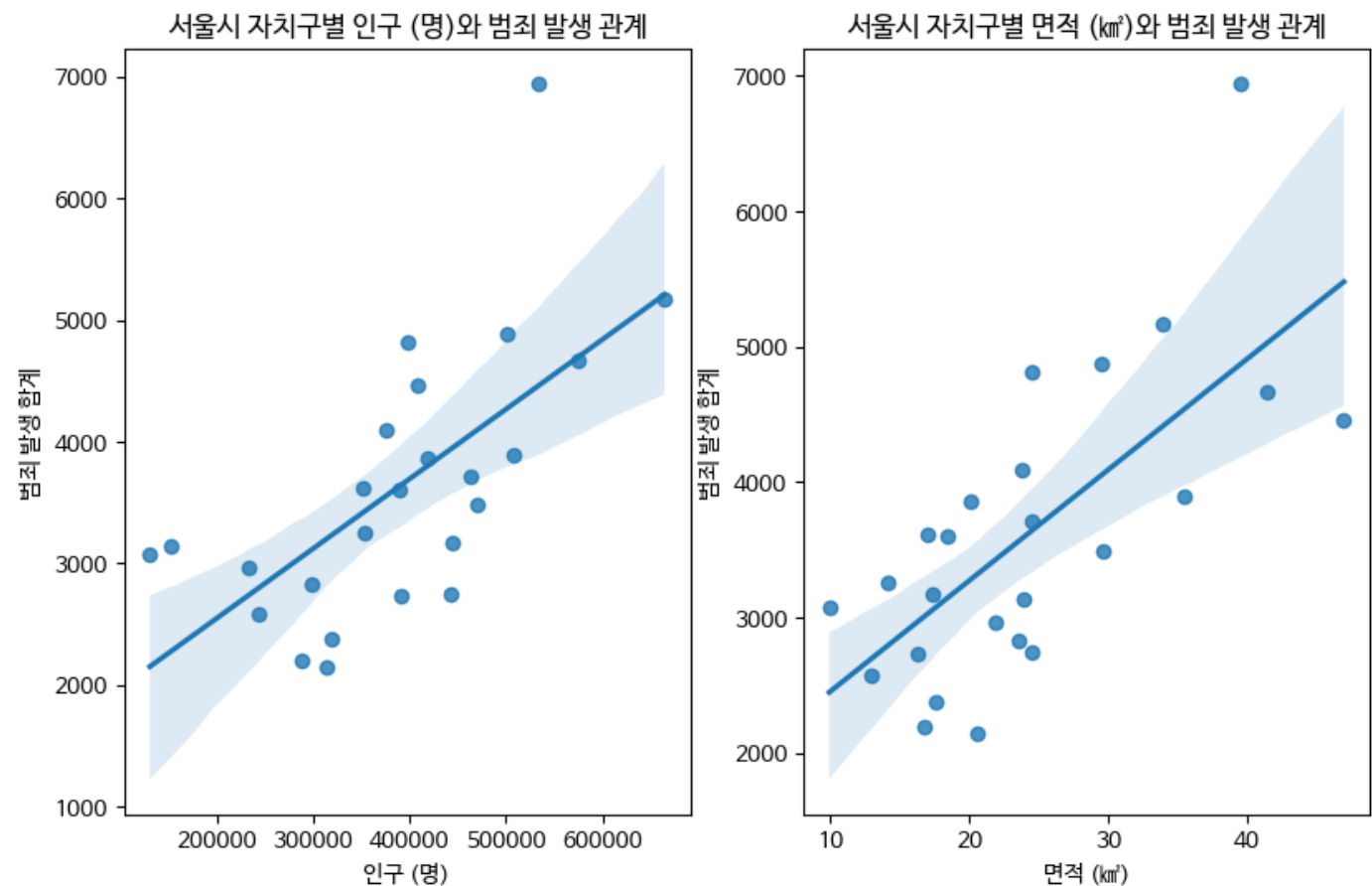
⑤ 시각화 및 상관관계분석

- Seaborn, matplotlib 등의 모듈을 통해 시각화, numpy 등의 모듈을 이용한 상관관계 분석



프로젝트 성과

가설 1. 인구 수가 많을 수록 범죄 발생수가 많을 것이다.



```
1 pop_seoul_corr()
```

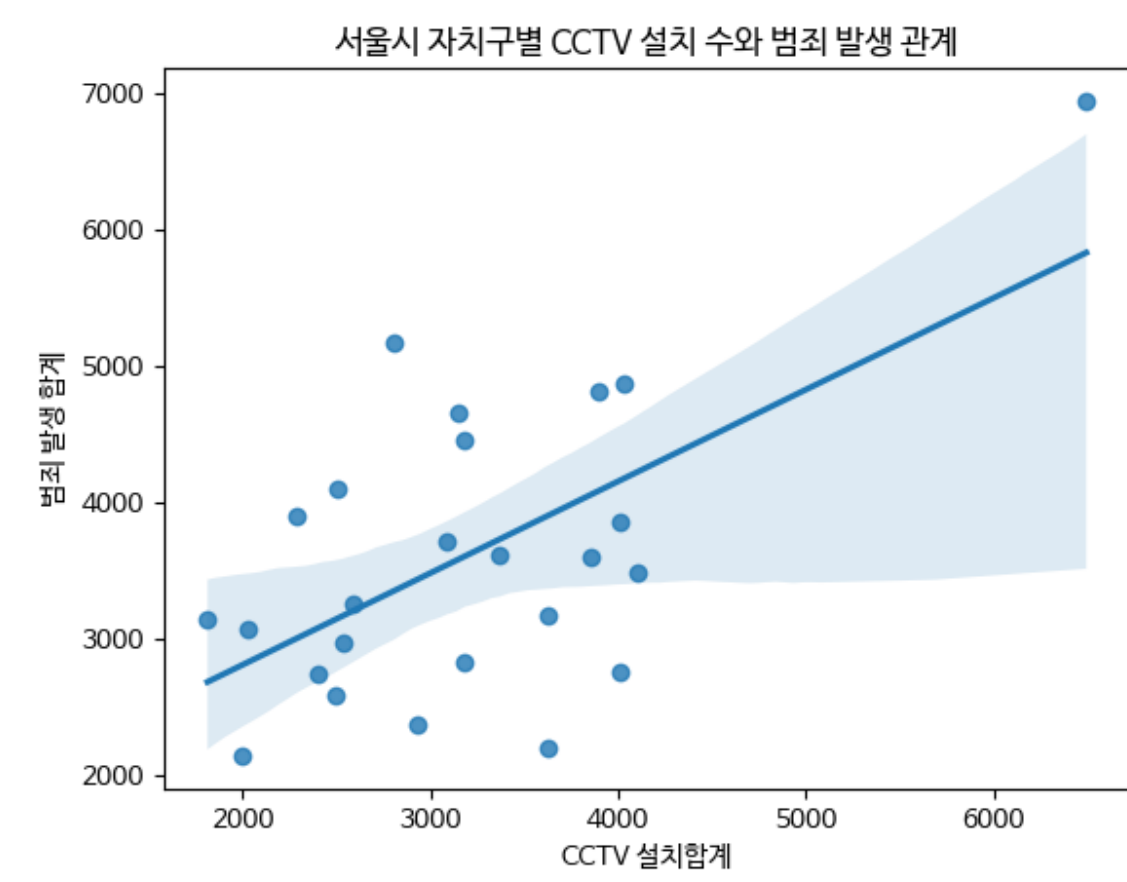
	인구 (명)	면적 (㎢)	인구밀도 (명/㎢)	범죄 발생 합계
인구 (명)	1.000000	0.654737	0.291402	0.654139
면적 (㎢)	0.654737	1.000000	-0.480772	0.694063
인구밀도 (명/㎢)	0.291402	-0.480772	1.000000	-0.148738
범죄 발생 합계	0.654139	0.694063	-0.148738	1.000000

상관관계 with 인구 수 = **0.65**

상관관계 with 면적 = **0.69**

→ 가설 성립

가설 2. CCTV 수가 적을 수록 범죄 발생수가 많을 것이다.



```
1 seoul_cctv_corr()
```

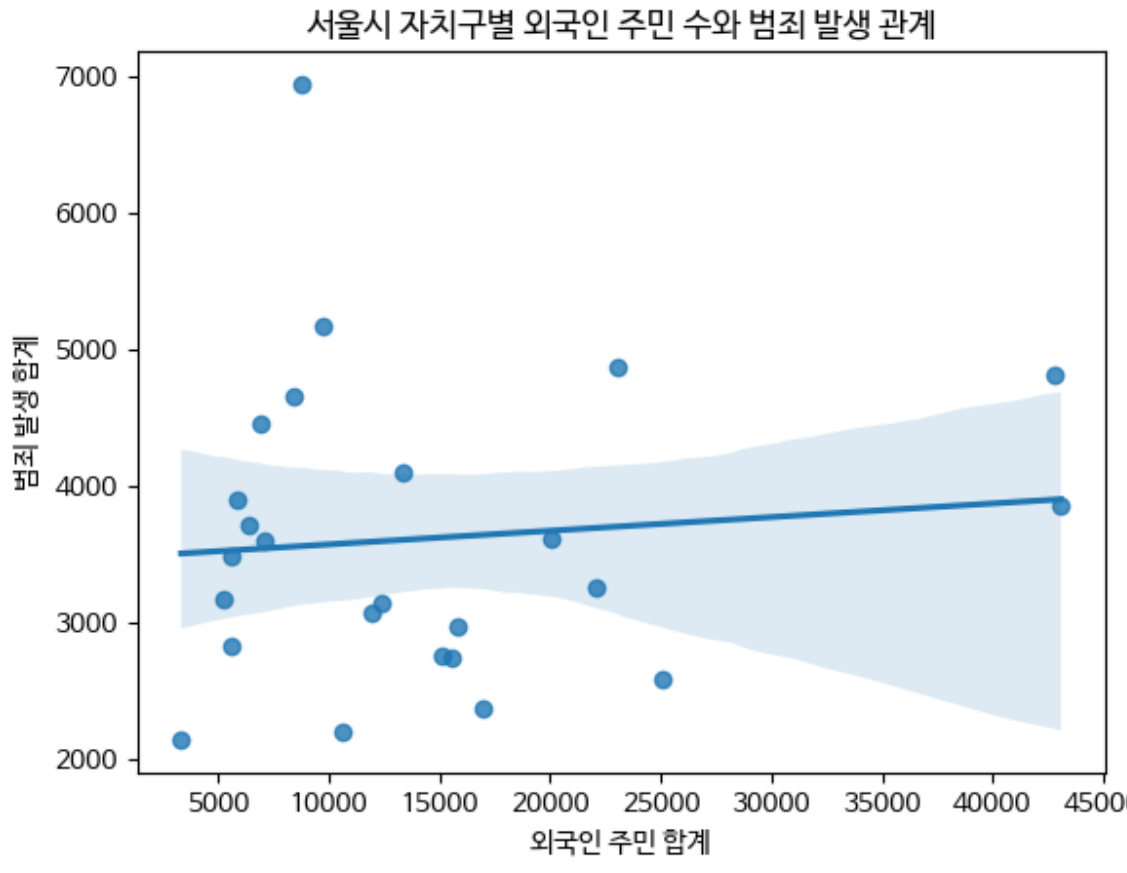
	CCTV 설치합계	범죄 발생 합계
CCTV 설치합계	1.000000	0.602961
범죄 발생 합계	0.602961	1.000000

상관관계 with CCTV 수 = **0.60**

→ 가설 미성립

→ 범죄 다발 지역에 CCTV 설치가 많이 이루어진 것.

가설 3. 외국인 수가 많을 수록 범죄 발생수가 많을 것이다.



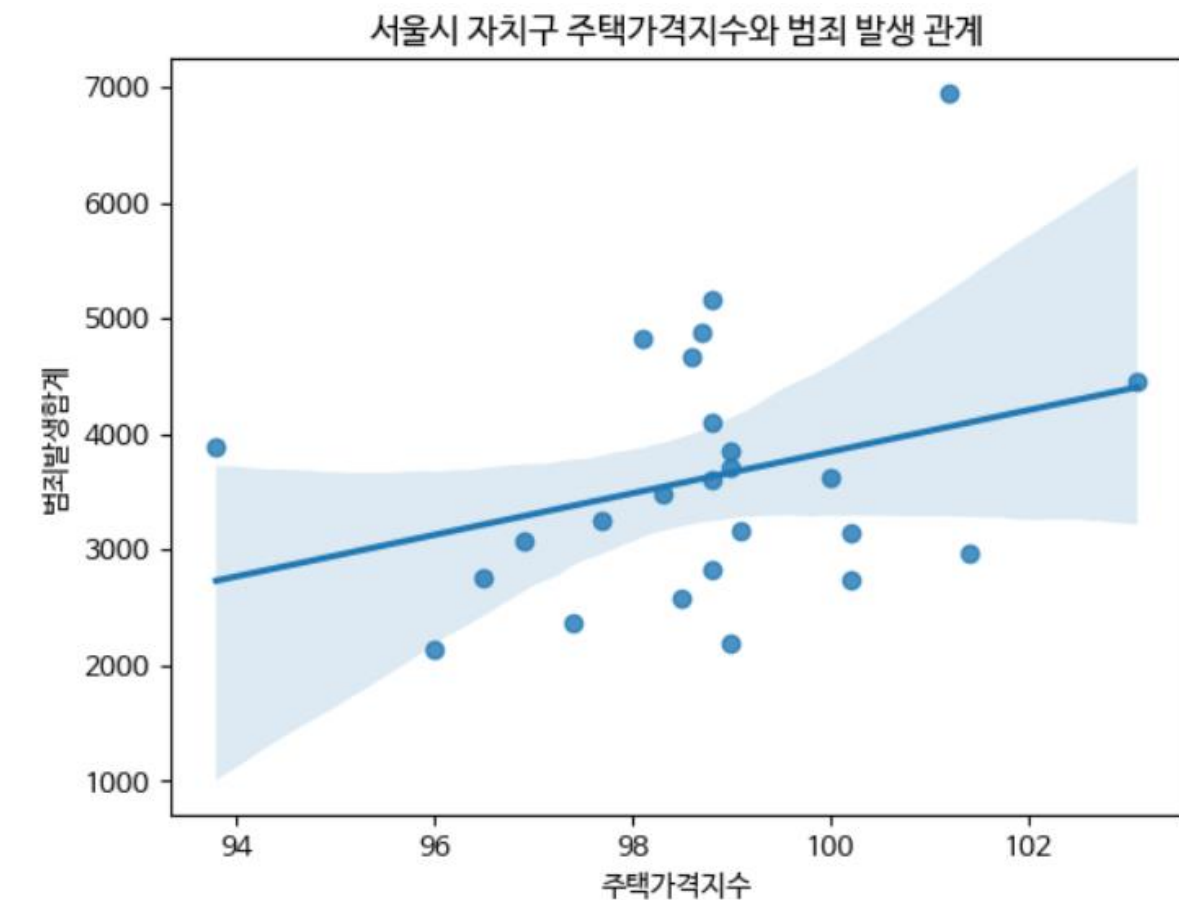
	외국인 주민 합계	범죄 발생 합계
외국인 주민 합계	1.000000	0.095505
범죄 발생 합계	0.095505	1.000000

상관관계 with 외국인 수 = **0.10**

→ 가설 미성립

→ 외국인 수와 관련한 유의미한 결과를 위해서는 더 세부적인 분류가 필요

가설 4. 소득 수준이 낮을 수록 범죄 발생수가 많을 것이다.



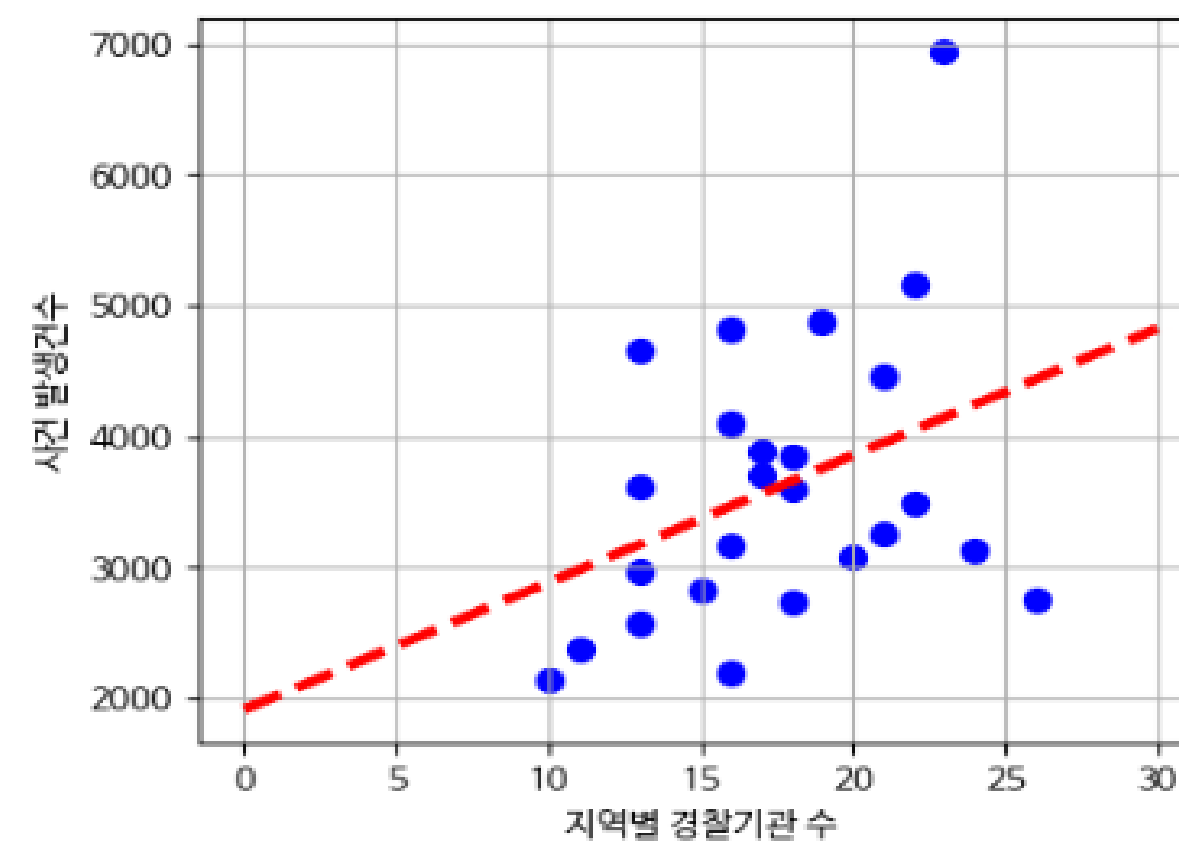
	주택가격지수	범죄발생합계
주택가격지수	1.000000	0.302766
범죄발생합계	0.302766	1.000000

상관관계 with 주택가격지수 = **0.30**

→ 가설 미성립

→ 소득수준이 지역별로 명확히 나뉘지 않는 것으로 사료. 더 정확한 분석을 위한 소득수준 데이터의 최적화 필요.

가설 5. 경찰서의 수가 적을수록 범죄 발생수가 많을 것이다.



	경찰 기관 합계	범죄 발생 합계
경찰 기관 합계	[1.0, 0.36673712]	
범죄 발생 합계	[0.36673712, 1.0]	

상관관계 with 경찰서 수 = **0.37**

→ 가설 미성립

→ 지역별 면적당 경찰서 수와의 상관관계 분석이 필요해 보임.



프로젝트 기대효과

- 프로젝트 결과로 몇몇 유의미한 상관관계를 찾아낼 수 있었지만, 연구의 목표에 부합하는 결론을 내리기에는 부족한 점이 있다.

- 데이터의 수집 범위가 너무 제한적이라는 점, 범죄 발생과 상관관계가 낮은 지표 선정, 변수에 대한 더 세부적인 처리와 최적화 부재 등이 원인으로 생각된다.

- 서울 이외의 수도권 지역 등으로 데이터 수집 범위를 확장하고, 유동인구 수, 지리적 요건 등의 지표를 추가한다면 더 유의미한 결론을 얻을 수 있을 것으로 사료된다.