
빅데이터 분석가 양성과정

미니 프로젝트(1)

2015년 범죄 발생 지역별 통계 분석 및 시각화

목차



STEP
01

Purpose of analysis

분석 목적 및 사용한 라이브러리 제시

STEP
02

Project Timeline

프로젝트 진행 기간을 표현

STEP
03

Preprocessing

시각화하기 전 적합한 데이터를 만들기 위한 전처리 과정

STEP
04

Visualization

결과를 눈에 보기 쉽도록 다양한 그래프를 이용하여 표현

Purpose of analysis

01

분석 목적

강력범죄와 지능범죄가 가장 많이 발생하는 지역에서의 강력범죄와 지능범죄 비율

노동범죄 발생 건수가 200이상이거나, 선거범죄 건수가 100이상인 지역 상위 3곳의 범죄 발생 건수 비교

강원, 경북 각 지역에서의 범죄 발생 건수 비교

서울특별시, 세종특별자치시 및 6개 광역시 구분하여 각 지역에서의 폭력범죄 별 발생 건수 비교

02

사용한 라이브러리

Pandas

Numpy

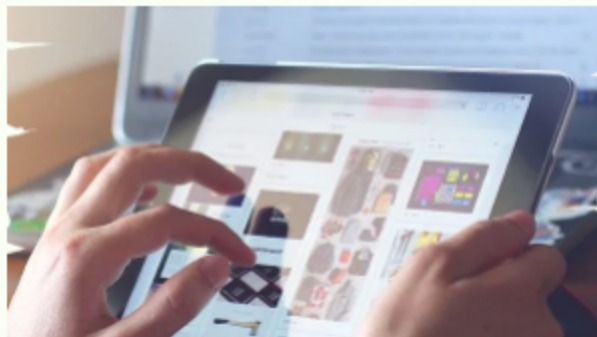
Matplotlib

PROJECT TIMELINE

0803

주제 및 데이터 선정

공공데이터포털에서 "2015 범죄 발생 지역별 통계" 데이터로 선정



0804

전처리 및 시각화

본격적인 분석에 앞서 경찰청에 접속하여 데이터 설명을 보며 잘 알지 못하는 용어를 이해 그 후 시각화하기에 적합한 형태의 데이터로 만들어 주기 위해 전처리 실시



0805

시각화 및 PPT 제작

어제 진행한 부분을 이어서 부족한 부분을 보완하기 위해 추가적으로 시각화할 부분 찾아 시각화 실시
다른 사람이 봐도 이해하기 쉽도록 코드 정리
모든 결과를 PPT로 제작



Preprocessing



데이터 불러오기

다운로드 한 데이터를
불러오고 데이터 유형확인



전치 및 불필요한 행 제거

불러온 데이터의 행과 열
전치시킨 후 불필요한 행 삭제



형 변환

숫자 형식이 문자로 되어있어
숫자형으로 변경



인덱스 초기화 및 컬럼명 변경

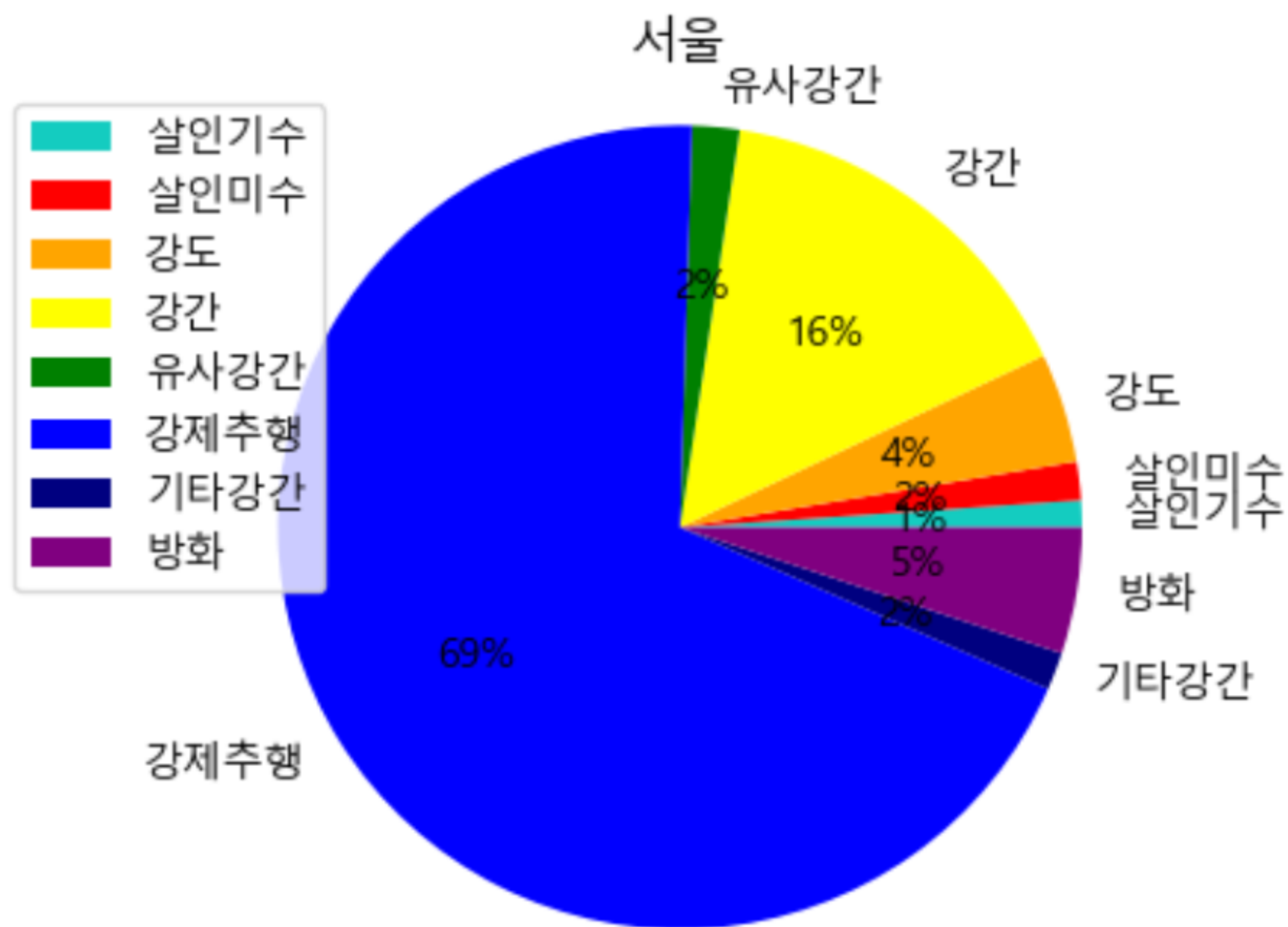
인덱스 초기화 한 후
특정 컬럼명 변경



최종 데이터프레임

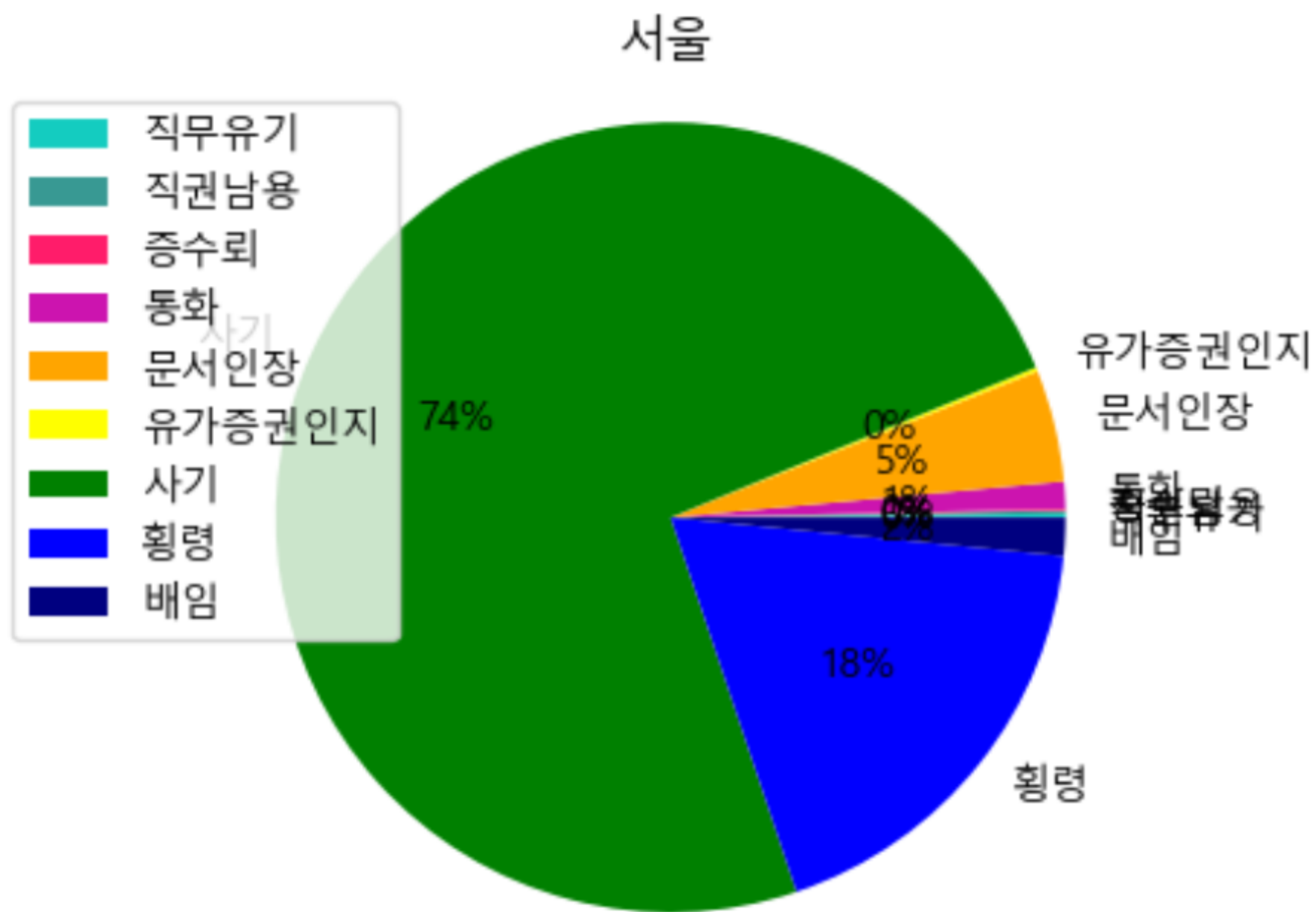
시각화에 적합한
최종 데이터프레임 생성

- 강력 범죄가 가장 많이 발생하는 지역에서의 강력 범죄 비율



**강력 범죄 중 가장 많은 비율을 차지
하는 범죄명은 강제추행
유사강간, 살인미수, 살인기수, 기타
강간은 매우 드물게 발생**

01 지능 범죄가 가장 많이 발생하는 지역에서의 지능 범죄 비율



**지능 범죄 중 가장 많은 비율을 차지
하는 범죄명은 사기
유가증권인지, 증수뢰, 직권남용은
다른 범죄에 비해 매우 드물게 발생**

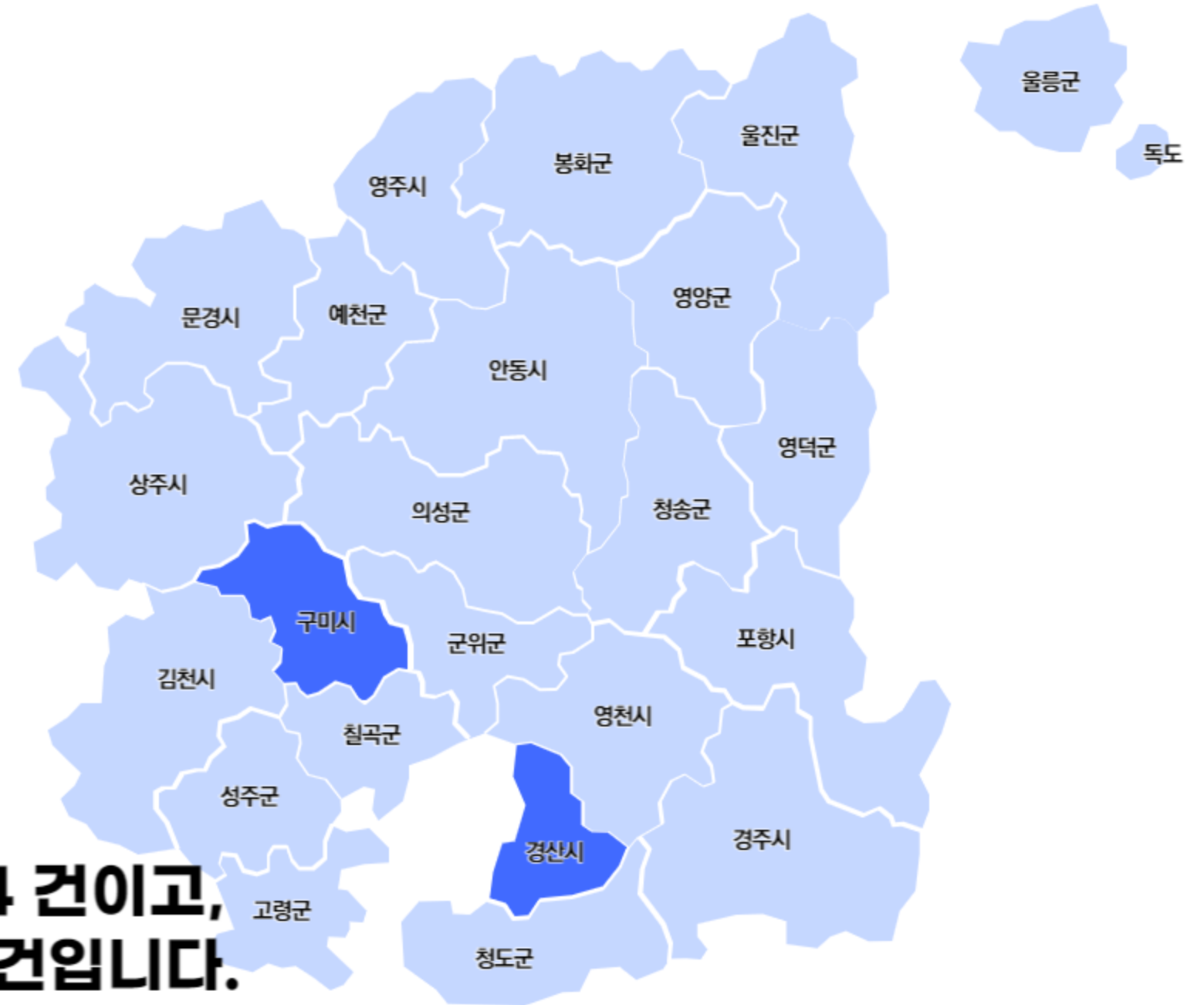
- 환경범죄가 가장 많이 발생하는 지역과 가장 적게 발생하는 지역의 이름과 건수

**환경범죄가 가장 많이 발생하는 지역명은 도시이외 503 건이고,
환경범죄가 가장 적게 발생하는 지역명은 경기 군포 0 건입니다.**



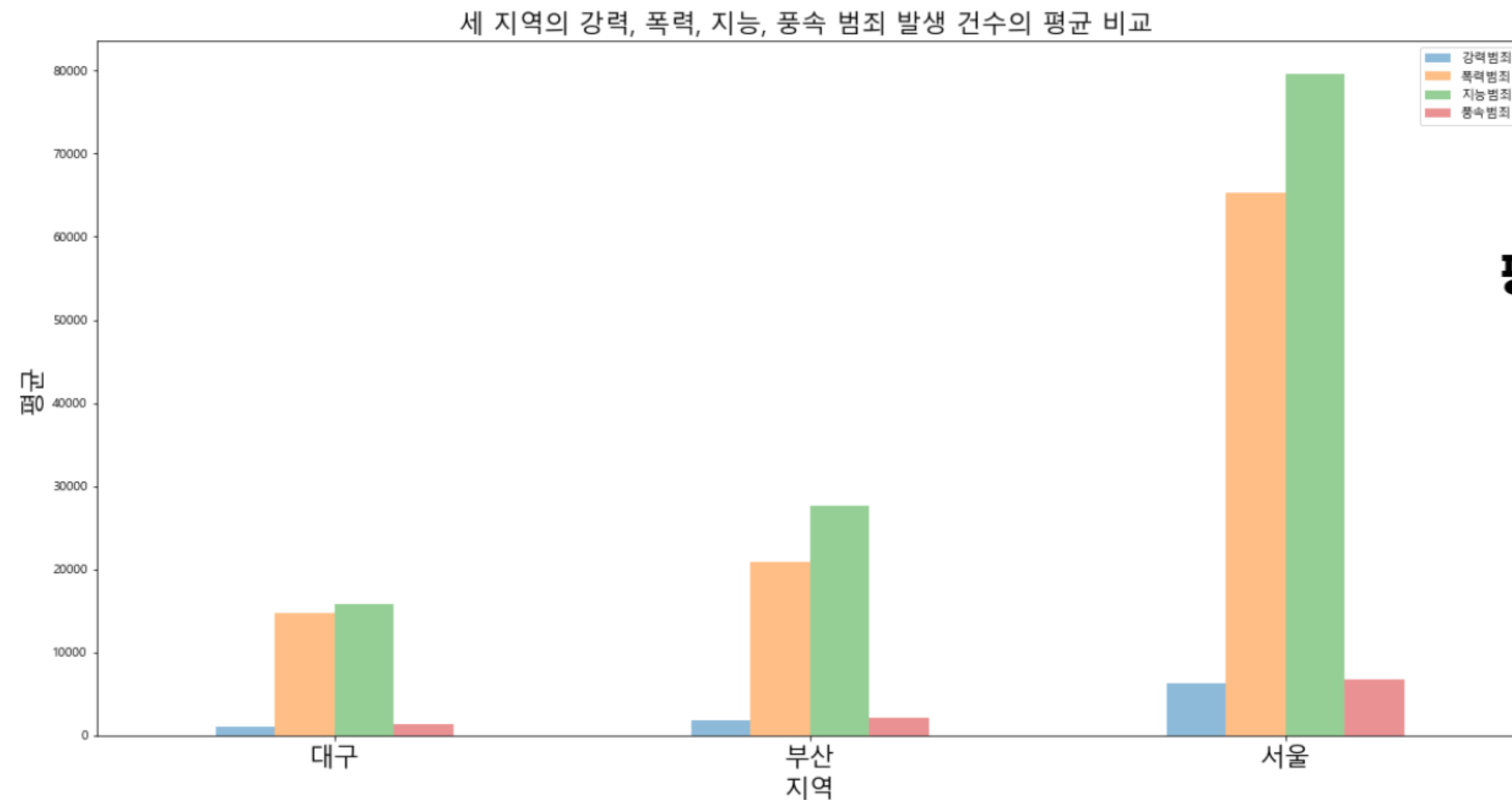
*도시이외 : 지역명이 ~시로 끝나지 않은 지역

- 경북 지역에서 마약범죄가 가장 많이 발생하는 지역과 마약범죄가 가장 적게 발생하는 지역의 이름과 건수



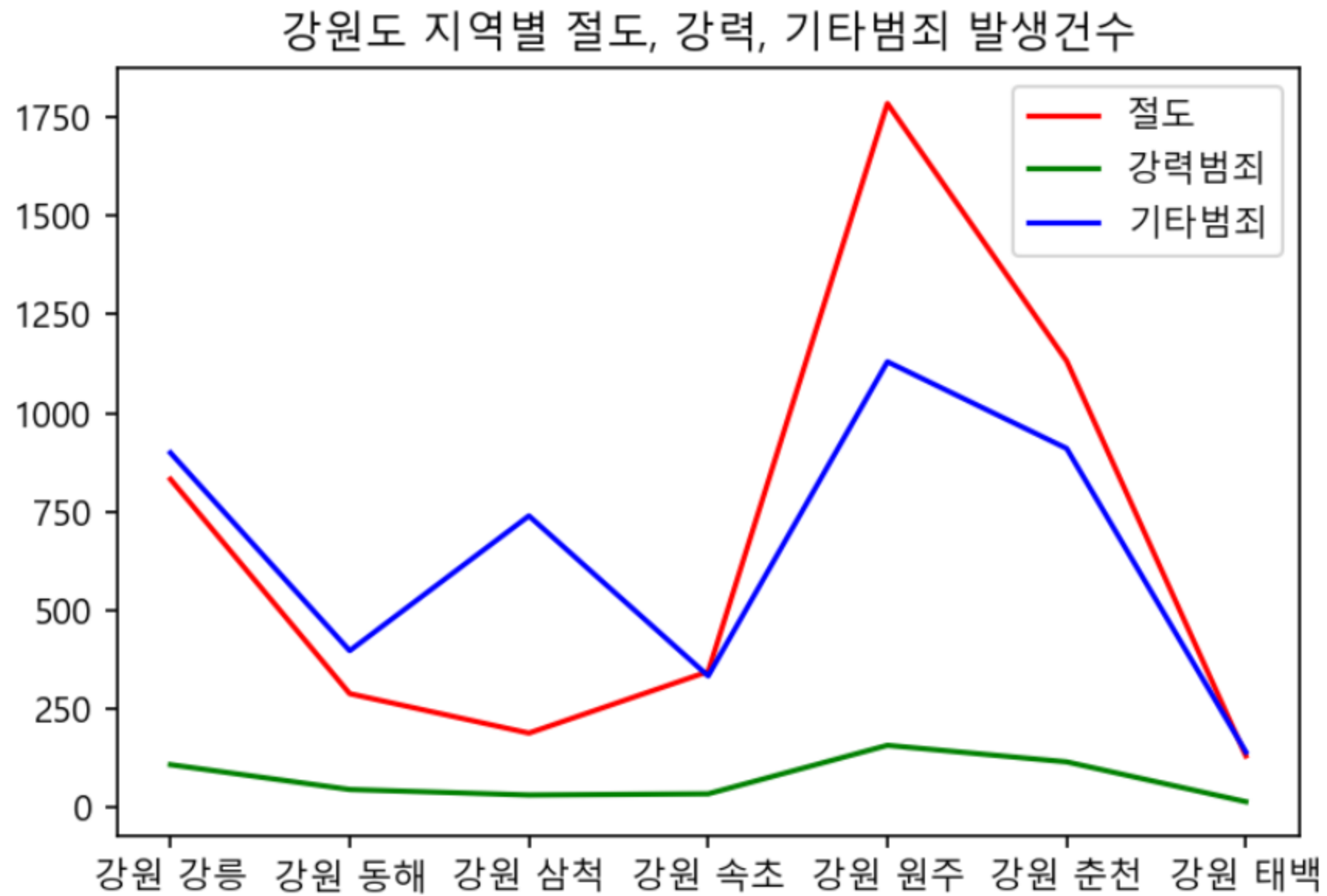
**마약범죄가 가장 많이 발생하는 지역명은 경북 구미 44 건이고,
마약범죄가 가장 적게 발생하는 지역명은 경북 경산 11 건입니다.**

- 세 지역의 강력, 폭력, 지능, 풍속범죄 발생 건수의 평균 비교

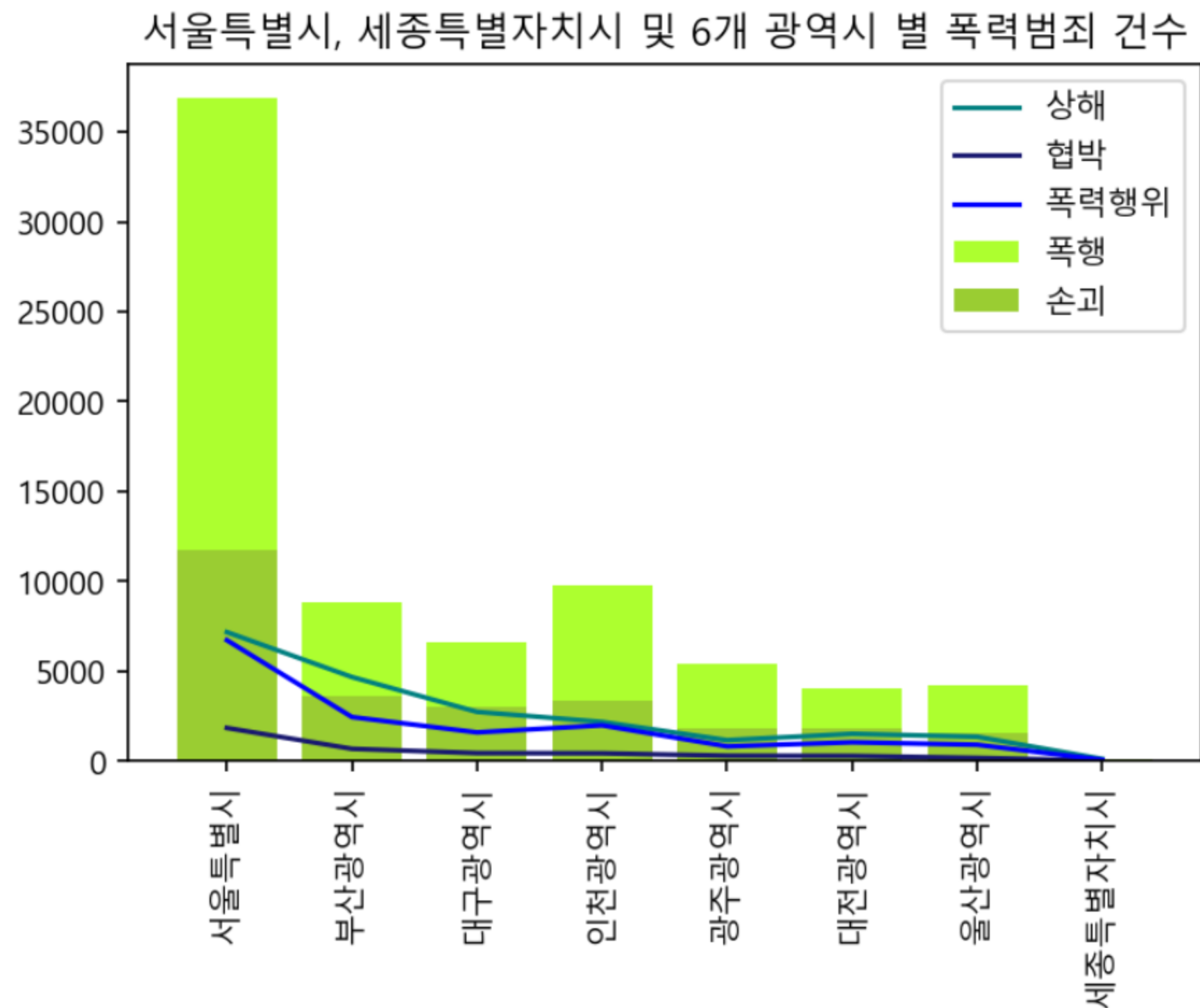


**평균적으로 세 지역 모두 지능
범죄가 많이 발생
강력범죄나 풍속범죄는 거의
비슷하게 발생**

- 강원도 지역별 절도, 강력, 기타범죄 발생건수



- 서울특별시, 세종특별자치시 및 6개 광역시 별 폭력범죄 건수



8개 지역 중 서울특별시가 가장 많이 발생
그 중 폭행이 가장 많이 많은 빈도를 차지함
협박은 폭력범죄 중 매우 적게 발생

THANK YOU

감사합니다

2022.08.05

정보통계학과 20190709 이자영
