

# 파이썬으로 범죄 발생 장소별 통계의 데이터 분석

202144107\_박인수

## 1. 배경

현대 사회에서 범죄는 공공 안전과 더불어 주요 사회 문제 중 하나로 여겨집니다. 범죄의 발생과 분포를 이해하고 분석함으로써, 효과적인 예방과 대응 정책을 수립하는 데 기여할 수 있습니다. 특히, 범죄 발생 장소에 초점을 맞추어 이러한 현상을 탐구함으로써, 우리는 범죄의 특성과 동향을 더욱 명확하게 파악할 수 있습니다.

### 1.1 연구의 필요성

이 연구에서는 파이썬을 활용하여 범죄 발생 장소별 통계를 분석함으로써, 특정 지역 또는 환경에서의 범죄 패턴을 도출하고자 합니다. 이는 정부 기관, 경찰, 그리고 지역 사회에게 범죄 예방 및 대응에 필요한 정보를 제공할 수 있는 중요한 자료로 활용될 것입니다.

### 1.2 목적 및 범위

본 보고서의 목적은 주어진 시간 동안의 범죄 발생 장소별 통계 데이터를 파이썬을 사용하여 분석하고, 그 결과를 통해 특정 지역 또는 환경에서 범죄의 특징을 도출하는 것입니다. 또한, 이 연구를 통해 얻은 인사이트는 지역 사회의 안전성 강화와 범죄 예방을 위한 논의에 기여할 것으로 기대됩니다.

### 1.3 연구의 중요성

본 연구는 범죄 관련 정책 및 대응 전략의 개발에 적극적으로 기여함으로써, 지역 사회의 안전을 높이고 공공 안전에 도움을 줄 수 있습니다. 또한 파이썬과 같은 데이터 분석 도구의 활용은 향후 범죄 관련 연구 및 대응에서의 유용성을 제고할 것으로 기대됩니다.

## 2. 데이터 수집: 경찰청 범죄 통계 데이터 소개

연구에서 사용된 데이터는 경찰청이 제공하는 "경찰청\_범죄 발생 장소별 통계\_20221231"입니다. 이 데이터는 공공데이터포털에서 제공되며, 경찰청이 수집한 국내 범죄 발생 장소에 대한 통계 정보를 포함하고 있습니다.

### 2.1 데이터 수집 기간

데이터 수집은 2022년 12월 31일을 기점으로 하며, 현재까지의 최신 데이터를 확보하기 위해 2023년 9월 13일까지의 정보를 포함하였습니다. 이 데이터는 매년 9월 13일마다 업데이트가 이루어지므로, 연간 데이터 갱신을 통해 최신 동향을 파악할 수 있습니다.

## 2.2 데이터의 신뢰성과 정확성에 대한 고려 사항

이 데이터는 경찰청이 수집하고 관리하는 공식 통계 자료이므로 신뢰성과 정확성이 높다고 판단됩니다. 경찰청은 범죄 발생과 관련된 데이터를 철저하게 수집하고 정리하여 제공하고 있으며, 이러한 프로세스에서 품질 관리 및 검증 절차를 통해 데이터의 신뢰성을 유지하고 있습니다.

## 3. 데이터 전처리

### 데이터 필터링

`violent_crimes = crime[crime['범죄대분류'] == '강력범죄']`: 이 코드는 '범죄대분류' 열이 '강력범죄'인 행들만을 선택하여 `violent_crimes`에 저장합니다. 이는 주어진 데이터에서 '강력범죄'에 해당하는 부분만을 추출하는 과정으로, 필요한 정보에 집중할 수 있도록 데이터를 걸러내는 역할을 합니다.

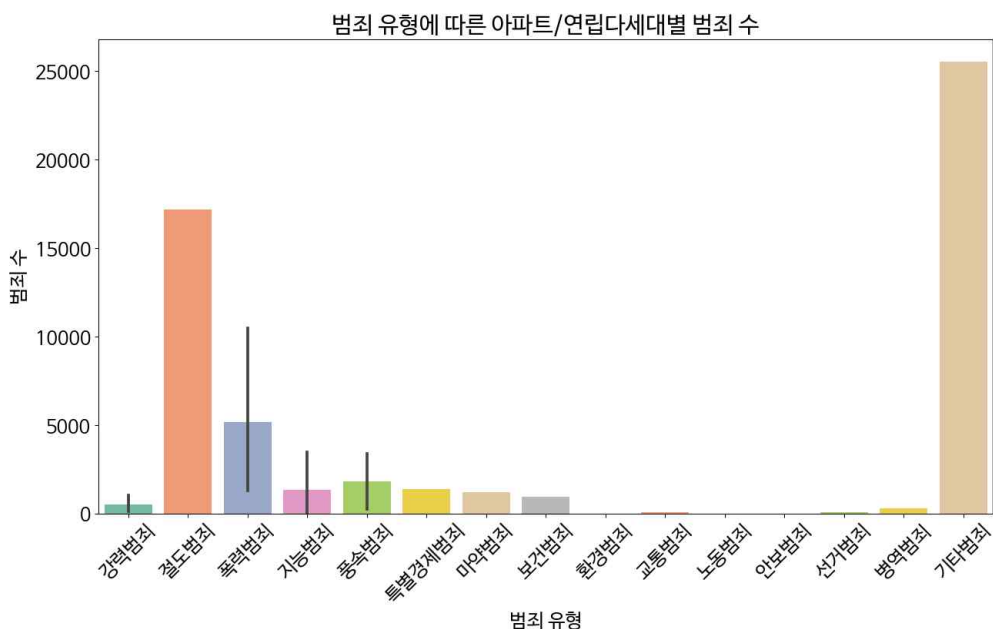
### 데이터 집계

`location_crime_counts = violent_crimes.sum()[2:]`: 이 코드는 '강력범죄'에 해당하는 데이터를 사용하여 각 장소(컬럼)별로 범죄 수를 집계합니다. `sum()` 함수를 사용하여 각 열의 값을 모두 더한 후, `[2:]`를 사용하여 불필요한 첫 두 행(범죄대분류, 범죄소분류)을 제외하고 나머지 데이터를 선택합니다.

이 두 단계를 통해 코드는 '강력범죄'에 해당하는 데이터를 추출하고, 해당 데이터에서 장소별 범죄 수를 계산하여 시각화할 수 있도록 데이터를 준비하고 있습니다.

## 4. 데이터 시각화

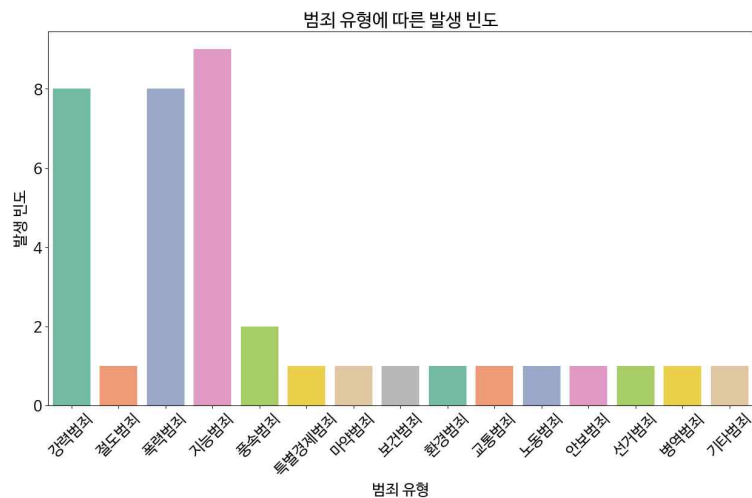
### 4.1 범죄 유형에 따른 장소별 범죄 수 시각화



```
# 1. 범죄 유형에 따른 장소별 범죄 수 시각화
plt.figure(figsize=(15, 8))
sns.barplot(x='범죄대분류', y='아파트_연립다세대', data=crime, palette='Set2')
plt.title('범죄 유형에 따른 아파트/연립다세대별 범죄 수')
plt.xlabel('범죄 유형')
plt.ylabel('범죄 수')
plt.xticks(rotation=45)
plt.show()
```

위 시각화에서 아파트/연립다세대 별 범죄유형에 따른 범죄수를 볼 수 있으며, y값을 병원, 편의점, 노상 등으로 변경하면 해당 위치에 대한 범죄유형별 범죄수를 알 수 있습니다.

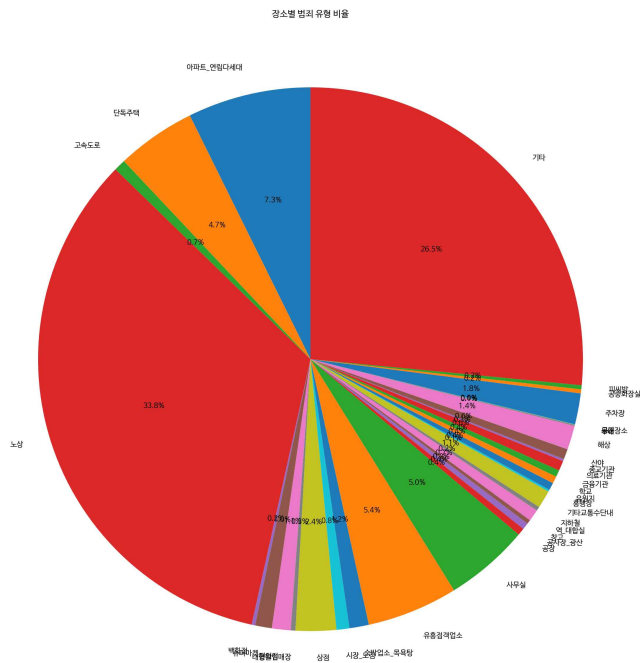
#### 4.2 범죄 유형에 따른 범죄 발생 빈도 시각화



위는 범죄 대분류에 있는 범죄별 발생빈도를 알기 위해 제작된 시각화 자료입니다. 지금까지 데이터를 수집하면서 쌓인 내역을 바탕으로 어떤 범죄의 발생빈도가 가장 높은지 알 수 있습니다.

```
# 2. 범죄 유형에 따른 범죄 발생 빈도 시각화
plt.figure(figsize=(15, 8))
sns.countplot(x='범죄대분류', data=crime, palette='Set2')
plt.title('범죄 유형에 따른 발생 빈도')
plt.xlabel('범죄 유형')
plt.ylabel('발생 빈도')
plt.xticks(rotation=45)
plt.show()
```

### 4.3 장소별 범죄 유형 비율 시각화



```
# 3. 장소별 범죄 유형 비율 시각화
plt.figure(figsize=(40, 40))
crime_location_counts = crime.iloc[:, 2:].sum()
crime_location_counts.plot.pie(autopct='%1.1f%%', startangle=90)
plt.title('장소별 범죄 유형 비율')
plt.ylabel('')
plt.show()
```

위 장소별 범죄 유형 비율 시각화는 각 장소에서 발생한 범죄유형의 총 빈도를 얻고, 퍼센트(#.##)와 파이차트로 표현합니다. 현재 파이차트를 보면 노상이 범죄 유형이 1등이고 기타가 2등인 것을 쉽게 알 수 있습니다.

## 5. 통계 분석

### 5.1 주요 통계 분석

연립다세대의 평균, 중앙값, 표준편차를 알 수 있습니다.

- 아파트\_연립다세대의 평균: 2850.0263157894738
- 아파트\_연립다세대의 중앙값: 251.0
- 아파트\_연립다세대의 표준편차: 6051.394230331257

연립다세대의 연평균 범죄율은 2850회, 중앙값은 251회 표준편차는 6051임을 알 수 있습니다.

### 5.2 주요 통계 분석 - 2

강력범죄의 평균, 중앙값, 표준편차를 알 수 있습니다.

- 강력범죄의 평균: 91.74264705882354
- 강력범죄의 중앙값: 2.0
- 강력범죄의 표준편차: 147.6807153826601

이를 통해 강력범죄의 평균은 91회, 중앙값은 2회, 표준편차는 147임을 알 수 있습니다.

## 6. 결과 해석

프로젝트 진행으로 범죄대분류별, 범죄중분류별, 장소별에 대한 범죄 횟수를 알 수 있으며, 위 시각화에서 특정 지역, 특정 시기에 따른 범죄율이 높은지 낮은지는 알 수가 없지만. 특정 장소에 따라 어떤 범죄율이 높은지 알 수 있기 때문에 그런 장소에 방문할 때 조심할 수 있습니다.

## 7. 프로젝트의 한계

정확히 말하면 프로젝트의 한계라기보다 데이터 값의 한계라고도 볼 수 있습니다. 데이터의 양이 적으므로 5.1과 5.2에 나온 평균, 중앙값, 표준편차처럼 0회와 100회가 넘어가는 범죄율이 같이 섞여 저렇게 높은 표준편차와 중앙값, 평균을 만들어내기 때문에 통계를 분석하기는 어렵습니다. 하지만 총 범죄율은 확인이 가능합니다.

## 8. 결론

프로젝트를 통해 범죄 정보를 시각적으로 전달하여 사람들이 위험한 지역을 쉽게 파악하고 예방할 수 있습니다. 이는 안전을 높이는 데 도움이 될 뿐만 아니라, 큰 규모의 정보를 효과적으로 다루는 능력을 키우게 됩니다. 앞으로 취업 시에도 이러한 기술과 경험을 활용하여 흥미로운 분야에서 일할 수 있을 것으로 기대됩니다.

## 9. 참고문헌

경찰청 – 경찰청 범죄 발생 장소별 통계

라이브러리 – pandas, numpy, matplotlib.pyplot

인하공업전문대학 민정혜 교수님 강의 자료