

201944034  
컴퓨터정보 A  
정성민

빅데이터 프로젝트 - 교통사고 원인과 상관관계 분석 및 시각화

# 교통사고 원인분석

- 음주운전 재범율과 교통사고 비율의 상관관계
- 연령대별 교통사고 추이
- 시도별 전년대비 교통사고 건수 및 증감률 분석
- 2022년 시도별 교통사고 수치 표시

# 개발환경

- **1. 프로그래밍 언어:**
  - **Python:** 데이터 분석 및 시각화에 주로 사용되는 언어.
- **2. 주요 라이브러리 및 프레임워크:**
  - **Pandas:** 데이터 조작 및 분석을 위한 라이브러리.
  - **Matplotlib:** 데이터 시각화를 위한 라이브러리.
  - **Folium:** 지도 시각화를 위한 라이브러리.
  - **Seaborn:** Matplotlib을 기반으로 한 통계 데이터 시각화 라이브러리.
- **3. 데이터 가공 및 시각화 툴:**
  - **Google Colab:** 온라인 기반의 Jupyter Notebook 플랫폼.
- **4. 데이터 수집:**
  - **도로교통공단 API:** 대한민국 도로교통공단에서 제공하는 API를 활용해 교통사고 데이터를 수집.
- **5. 버전 관리:**
  - **Git:** 소스 코드의 버전을 관리하는 데 사용.

## 공공데이터 포털 및 도로교통공단 자료수집

# 데이터 수집

### 파일데이터 (359건)

교통물류

공공기관

CSV

JSON + XML

한국도로공사\_교통사고통계

한국도로공사 고속도로 교통사고 현황 통계에 대한 데이터입니다. (연도, 사고, 사망, 부상) 등의 형

제공기관 한국도로공사 수정일 2023-07-31 조회수 28665 다운로드 10229 주:

공공행정

공공기관

CSV

JSON + XML

도로교통공단\_사고유형별 교통사고 통계

- 경찰에서 조사, 처리한 교통사고에 대한 통계 정보로 인적 피해가 있는 사고만 집계 됨 - 사고유

제공기관 도로교통공단 수정일 2023-06-20 조회수 65574 다운로드 12213 주:

재난안전

국가행정기관

국가행정

CSV

JSON + XML

경찰청\_졸음운전 교통사고 현황

졸음운전, 교통사고, 졸음, 교통사고 현황, 교통사고 사건 사망 부상자 현황2017년, 2018년, 2019년

제공기관 경찰청 수정일 2021-09-29 조회수 18594 다운로드 11598 주기성 데

공공행정

공공기관

CSV

JSON + XML

도로교통공단\_시도 시군구별 교통사고 통계

- 경찰에서 조사, 처리한 교통사고에 대한 통계 정보로 인적 피해가 있는 사고만 집계 됨 - 시도 및

제공기관 도로교통공단 수정일 2023-06-20 조회수 46614 다운로드 5890 주기

# 데이터 전처리



# 결론

## 1. 연도별 사고 비율:

1. 코로나 전염병 발생 이전인 2019년에는 교통사고 비율이 가장 높았습니다.
2. 연초에는 사고 비율이 낮았으며, 휴가철과 연말에 비율이 상승했습니다.

## 2. 요일 및 시간대별 사고 현황:

1. 요일 중에서는 금요일에 교통사고가 가장 많았으며, 특히 출퇴근 시간대에 비율이 높았습니다.

## 3. 음주운전과 교통사고 상관관계:

1. 음주운전과 교통사고 간 상관관계 분석에서는 -0.9의 강한 음의 상관성이 나타났습니다.
2. 이를 통해 대부분의 교통사고가 음주운전으로 인해 발생되었다고 해석할 수 있습니다.

## 4. 연령대별 교통사고 비율:

1. 연령대별로 보면, 4-30대가 교통사고 비율이 가장 높았습니다.
2. 반면에 시니어층과 20대는 다른 연령대에 비해 상대적으로 적은 교통사고가 보고되었습니다.

## 5. 시도별 전년 대비 교통사고 건수:

1. 시도별로 비교하면, 유동인구가 많은 경기와 서울 지역에서 교통사고가 가장 많았습니다.
2. 세종은 전년 대비 교통사고 건수에서 가장 높은 증감률을 기록했습니다.

