

빅데이터 처리

교통사고 데이터 분석

3-C 202044092 김 산

목차

- 데이터 소개 및 전처리
- 연령별 교통사고 중상자 비율 분석
- 가장 많은 교통사고 발생 요일 및 시간대
- 사상자 수 지도표현
- 도로 유형별 사고 심각도
- 결론
- ETC

데이터 소개 및 전처리

- 영국 경찰 제공
- 영국 내 모든 차량 충돌에 대한 데이터

출처 : [Road Safety Data](#)



```
import kagglehub

path = kagglehub.dataset_download("silicon99/dft-accident-data")
print("Path to dataset files:", path)
```

데이터 소개 및 전처리

- 데이터 구성
 - Accidents0514.csv (사고 데이터)
 - Casualties0514.csv (피해자 데이터)
 - Vehicles0514.csv (차량 데이터)



1,780,653의 사고 데이터

데이터 소개 및 전처리

- 전처리 주요 과정
 - 결측값 제거
 - 데이터 병합 ('Accident_Index')
 - 분석에 필요한 컬럼 선택

```
mergeData = pd.merge(pd.merge(accidents, casualites, on='Accident_Index'), vehicles,  
on='Accident_Index')
```

```
df[['Accident_Index', 'Age_of_Driver', 'Accident_Severity', 'Day_of_Week', 'Time']] =  
mergeData[['Accident_Index', 'Age_of_Driver', 'Accident_Severity', 'Day_of_Week', 'Time']]
```

```
#인덱스, 운전자 연령, 사고 심각도, 날씨, 요일, 시간
```

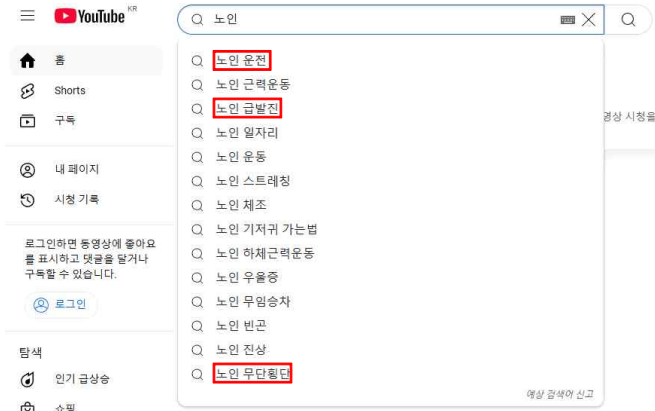
데이터 소개 및 전처리

Accident_Index	Age_of_Driver	Accident_Severity	''''
0	43	3	''''
1	26	2	''''
2	84	3	''''
3	34	3	''''
4	44	1	''''
5	24	3	''''

Drive.csv

연령별 교통사고 중상자 비율 분석

유튜브에 [노인] 검색 시 교통사고 관련 검색어만 3개가 존재한다.

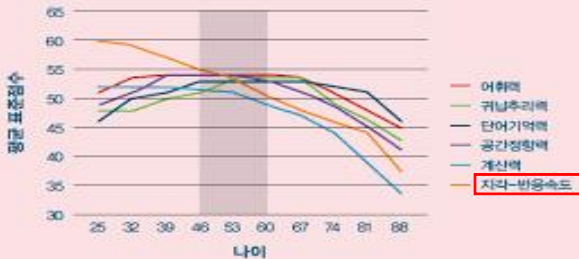


연령별 교통사고 중상자 비율 분석

시애틀 종단연구는 40년동안 6,000명의 정신적 기량을 연구

- 결과에서 운전 시 중요한
지각-반응속도가 나이가
들수록 급격하게 줄

시애틀 종단 연구에서 연령별 인지능력 결과



연령별 교통사고 중상자 비율 분석

목표 : 연령대별 사고 중상자 비율을 분석하여 연령대가 사고의 심각도에 미치는 영향을 파악

• 분석방법

- 연령대 그룹화 (e.g., 20 under, 21-25, 26-30 등)
- 중상자 비율 계산: 경상 제외, 사고 심각도(1: 경상, 2: 중상, 3: 사망) 기준

연령별 교통사고 중상자 비율 분석

주요 코드 (그룹화 및 비율 계산)

```
bins = [17, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 100]
labels = ['20 under', '21-25', '26-30', '31-35', '36-40', '41-45', '46-50', '51-55', '56-60',
'61-65', '66-70', '70 over']

df['Age_Group'] = pd.cut(df['Age_of_Driver'], bins=bins, labels=labels, right=True)

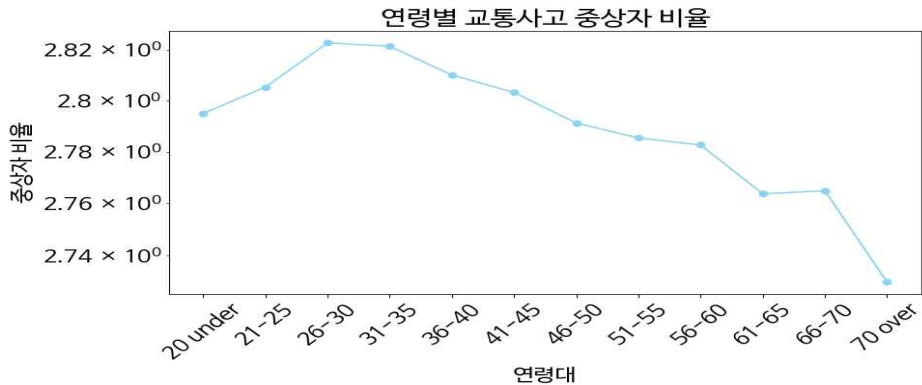
def count(x):
    return x[x!=1].sum()/x.count() #경상제외

age_severity = df.groupby('Age_Group')
```

연령별 교통사고 중상자 비율 분석

Index	Age	Severity	****	Group
0	43	3	****	41-45
1	26	2	****	26-30
2	84	3	****	70 over
3	34	3	****	31-35
4	44	1	****	41-45
5	24	3	****	21-25

연령대별 중상자 비율 그래프



연령대가 증가함에 따라 중상자 비율은 감소하는 추세

연령대별 교통사고 중상자 비율 분석 결론

• 결론

- 연령대가 증가할수록 운전 경험이 쌓이고, 면허 갱신 주기가 짧아지며, 의사 소견서 제출이 요구되는 등의 요인으로 중상자 비율이 감소
- 노년층 운전자는 교통사고 발생 시 중상자 비율이 낮아짐.

참고 링크: [영국 노년 운전면허 갱신방법](#)

교통사고 발생 요일 및 시간대 분석

목표 : 교통사고가 주로 발생하는 요일과 시간대 분석

- 분석방법

- 사고 발생 시간을 1시간 단위로 그룹화
- 사고 발생 요일에 대한 매핑 (e.g., 1: 일요일, 2: 월요일 등)

- 주요 코드

- `heatmap_data = df.groupby(['Day_of_Week','Time']).size().unstack(fill_value=0)`

교통사고 발생 요일 및 시간대 분석

주요 코드 (형식 변환)

```
df['Time'] = pd.to_datetime(df['Time'], format='%H:%M')
```

```
df['Time'] = df.groupby(pd.Grouper(key='Time',  
freq='1H'))['Time'].transform('first').dt.strftime('%H:%M')
```

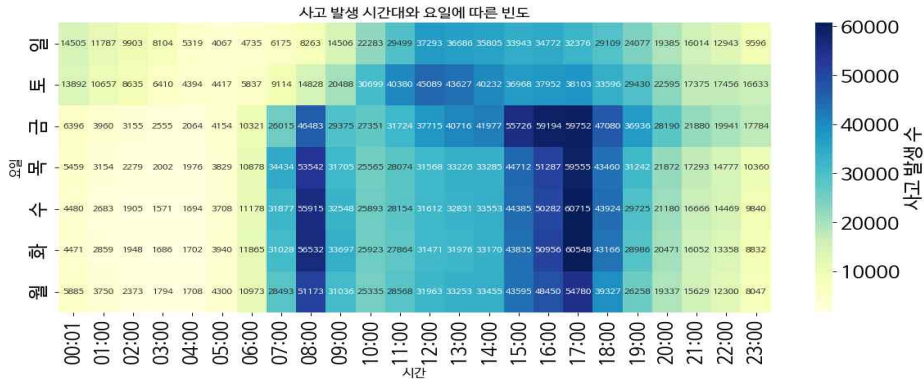
```
day_mapping = { 1: '일', 2: '월', 3: '화', 4: '수', 5: '목', 6: '금', 7: '토', }
```

```
df['Day_of_Week'] = df['Day_of_Week'].map(day_mapping)
```

```
# 사고 발생 빈도를 계산 (시간대와 요일 기준)
```

```
heatmap_data = df.groupby(['Day_of_Week', 'Time']).size().unstack(fill_value=0)
```

교통사고 발생 요일 및 시간대



평일 출퇴근 시간대와 주말 11:00 ~ 14:00 시간대가 사고 발생이 가장 많음

교통사고 발생 요일 및 시간대 분석 결론

- 결론

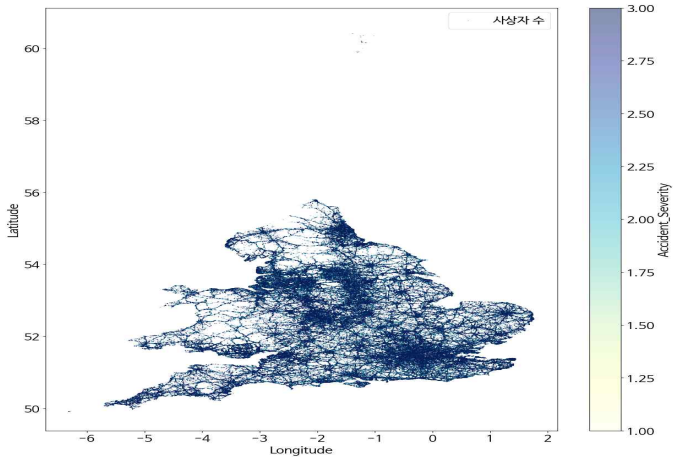
- 교통사고는 평일 출퇴근 시간대 (오전 8-9시, 오후 5-6시)와 주말 11:00 ~ 14:00 시간대에 집중됨

사상자 수 지도표현

주요 코드 (데이터 시각화)

```
df.plot(kind = "scatter", x = "Longitude", y = "Latitude", alpha = 0.5,  
      s = df["Number_of_Casualties"]/100, label = "사상자 수", figsize=(15,15),  
      c = "Accident_Severity", cmap = plt.get_cmap("YlGnBu"), colorbar= True  
      )  
plt.legend()
```

사상자 수 지도표현

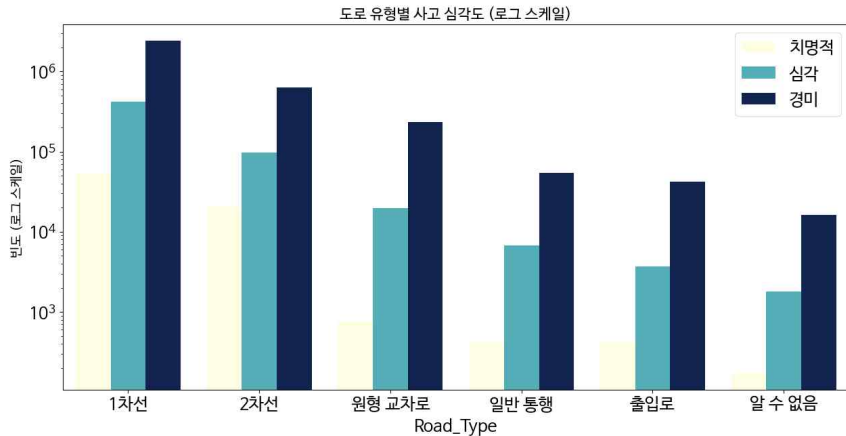


도로 구조별 사고 심각도

주요 코드 (데이터 시각화)

```
plt.figure(figsize=(15, 7))  
ax = sns.countplot(  
    x='Road_Type',  
    hue='Accident_Severity',  
    data=df,  
    order=df["Road_Type"].value_counts().index,  
    palette="YlGnBu"
```

도로 구조별 사고 심각도



결론

- **연령별 교통사고 중상자 비율**

연령대가 증가할수록 중상자 비율 감소

- **가장 많은 교통사고가 발생하는 요일과 시간대**

평일 출퇴근 시간대와 주말 11:00 ~ 14:00 시간대가 사고 발생이 가장 많음

- **사상자 수 지도표현**

인구집중 구역 사고빈도 증가

- **도로 구조별 사고 심각도**

1차선 도로 시 사고발생 빈도가 높음

ETC.

- 데이터

- UK Car Accidents

- **GitHub**

- https://github.com/RealSan1/Car_accident