

# 2021 전국 교통사고 통계

정보통계학과/20190699/김지은

# CONTENTS

## 2021 전국 교통사고 통계

### Chapter 01 데이터 전처리 및 확인

- 데이터 불러오기
- 결측치 확인 후 새로운 데이터 생성
- 각 열을 기준으로 행 출력

### Chapter 02 전국 통계

- 전국 사상자 통계 그래프
- 전국 교통사고율

### Chapter 03 시도별 통계

- 시도별 교통사고 추이 그래프

### Chapter 04 지역별 통계

- 지역별 교통사고 추이 그래프
- 각 지역의 월 별 교통사고 사상자 그래프
  - \* 사고건수가 가장 많은 월과 지역 출력
- 원하는 지역과 사상자율 추이가 가장 비슷한 지역 그래프



# Chapter 01

## 데이터 전처리 및 확인

- 데이터 불러오기
- 결측치 확인 후 새로운 데이터 생성
- 각 열을 기준으로 행 출력

## 1. 데이터 불러오기

```
df = pd.read_csv('accident.csv', encoding = 'cp949') #데이터 불러오기
print(df.info())
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2737 entries, 0 to 2736
Data columns (total 9 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   시도        2737 non-null   object  
 1   시군구       2737 non-null   object  
 2   지역        2737 non-null   object  
 3   발생월      2737 non-null   int64   
 4   사고건수    2737 non-null   int64   
 5   사망자수    2737 non-null   int64   
 6   중상자수    2737 non-null   int64   
 7   경상자수    2737 non-null   int64   
 8   부상신고자수 2737 non-null   int64   
dtypes: int64(6), object(3)
memory usage: 192.6+ KB
None
```

## 2. 결측치 확인 후 새로운 데이터 생성

```
df.isnull().values.any() #결측치 확인
```

```
False
```

```
df2=copy.deepcopy(df)

df2['발생월']=df2['발생월'].astype('str')
del df2['사고건수']
ax = df2.sum(axis=1)
df2.insert(8, '전체사상자수', ax, True)
df2
df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2737 entries, 0 to 2736
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype  
---  -
0   시도        2737 non-null   object 
1   시군구      2737 non-null   object 
2   지역        2737 non-null   object 
3   발생월      2737 non-null   object 
4   사망자수    2737 non-null   int64  
5   중상자수    2737 non-null   int64  
6   경상자수    2737 non-null   int64  
7   부상신고자수 2737 non-null   int64  
8   전체사상자수 2737 non-null   int64  
dtypes: int64(5), object(4)
memory usage: 192.6+ KB
```

### 3. 열의 최대값을 기준으로 행 출력

```
df[df['사고건수'] == df.iloc[:, [4]].max()[0]] # 사고건수가 제일 높은 행 출력
df[df['사망자수'] == df.iloc[:, [4, 5]].max()[1]] # 사망자수가 제일 높은 행 출력
df[df['중상자수'] == df.iloc[:, [4, 5, 6]].max()[2]] # 중상자수가 제일 높은 행 출력
df[df['경상자수'] == df.iloc[:, [4, 5, 6, 7]].max()[3]] # 경상자수가 제일 높은 행 출력
df[df['부상신고자수'] == df.iloc[:, [4, 5, 6, 7, 8]].max()[4]] # 부상신고자수가 제일 높은 행 출력
```

사고건수 ->

	시도	시군구	지역	발생월	사고건수	사망자수	중상자수	경상자수	부상신고자수
502	경기	수원시	경기수원시	11	459	2	93	515	58

사망자수 ->

	시도	시군구	지역	발생월	사고건수	사망자수	중상자수	경상자수	부상신고자수
1083	충북	청주시	충북청주시	4	355	8	113	395	11
2300	경남	창원시(통합)	경남창원시	5	208	8	70	214	10
2726	세종	세종시	세종시	2	50	8	20	55	1

중상자수 ->

	시도	시군구	지역	발생월	사고건수	사망자수	중상자수	경상자수	부상신고자수
1084	충북	청주시	충북청주시	5	353	4	114	372	14

경상자수 ->

	시도	시군구	지역	발생월	사고건수	사망자수	중상자수	경상자수	부상신고자수
502	경기	수원시	경기수원시	11	459	2	93	515	58

부상신고자수 ->

	시도	시군구	지역	발생월	사고건수	사망자수	중상자수	경상자수	부상신고자수
748	경기	용인시	경기용인시	5	299	7	60	343	62

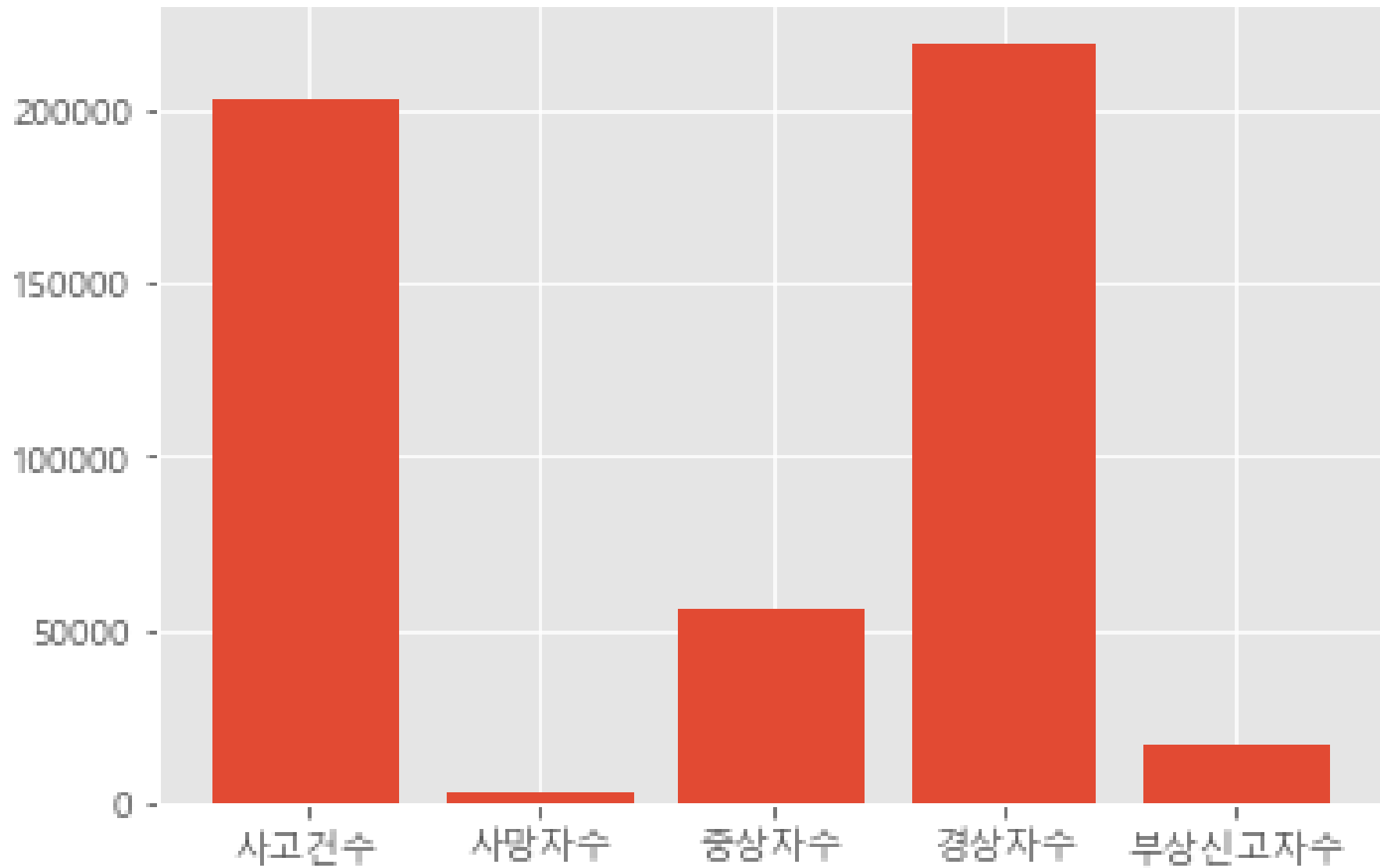
# Chapter 02

## 전국통계

- 전국 사상자 통계 그래프

## 1. 전국 사상자 통계 그래프

2021년 전국 사상자 통계





## 2. 전국 교통사고율

```
df_sum2 = np.sum(df2.iloc[:, [4, 5, 6, 7, 8]], axis=0)
a = round((df_sum2[0]/df_sum2[4]),3).astype(str)
b = round((df_sum2[1]/df_sum2[4]),3).astype(str)
c = round((df_sum2[2]/df_sum2[4]),3).astype(str)
d = round((df_sum2[3]/df_sum2[4]),3).astype(str)

print('전국 교통사고 사망률은 ' + a + '% 입니다.')
print('전국 교통사고 중상률은 ' + b + '% 입니다.')
print('전국 교통사고 경상률은 ' + c + '% 입니다.')
print('전국 교통사고 부상신고율은 ' + d + '% 입니다.')
```

전국 교통사고 사망률은 0.01% 입니다.

전국 교통사고 중상률은 0.19% 입니다.

전국 교통사고 경상률은 0.745% 입니다.

전국 교통사고 부상신고율은 0.056% 입니다.

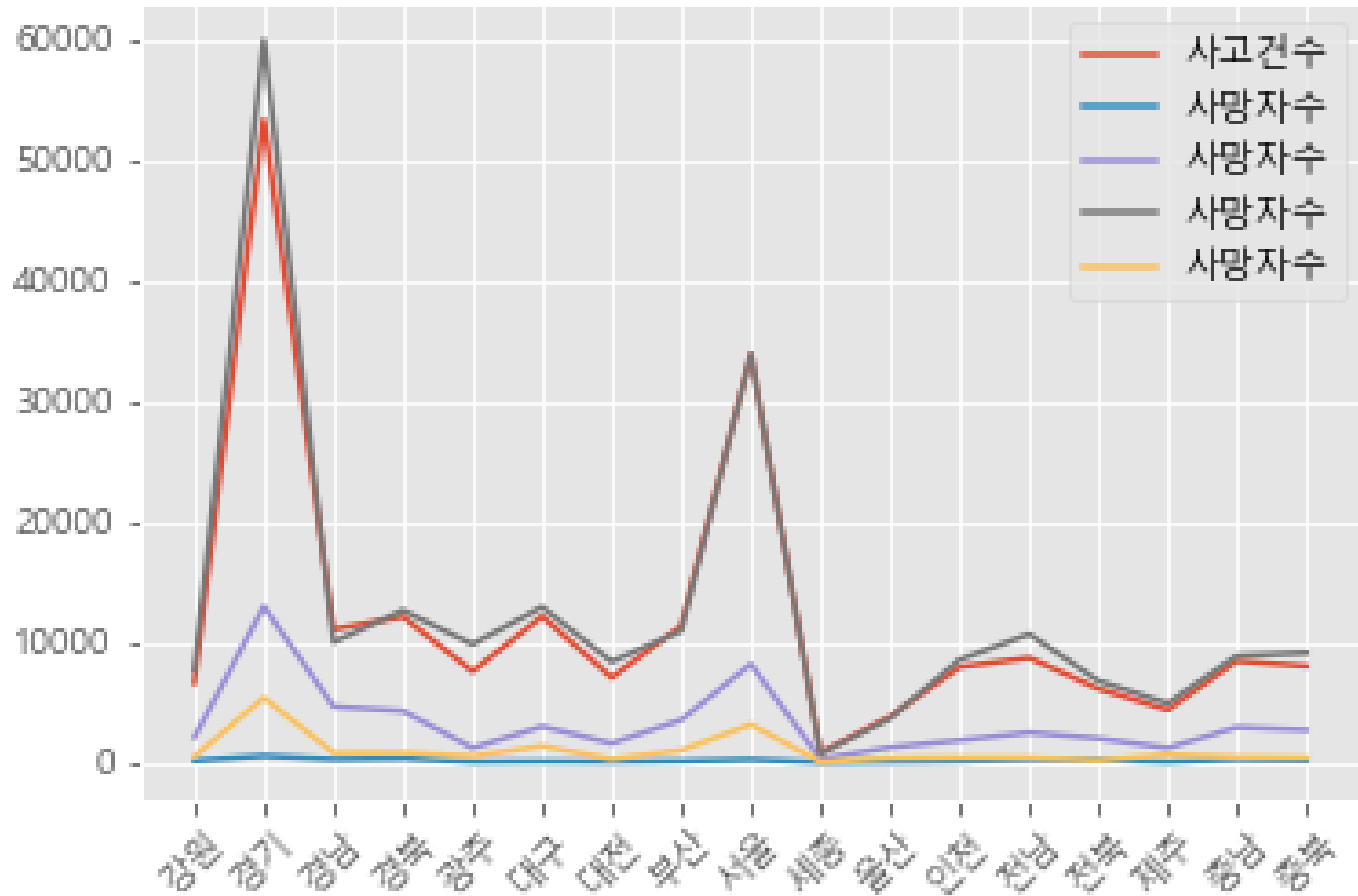
# Chapter 03

## 시도별 통계

- 시도별 교통사고 추이 그래프

## 1. 시도별 교통사고 추이 그래프

시도별 교통사고 추이 그래프



## Chapter 04

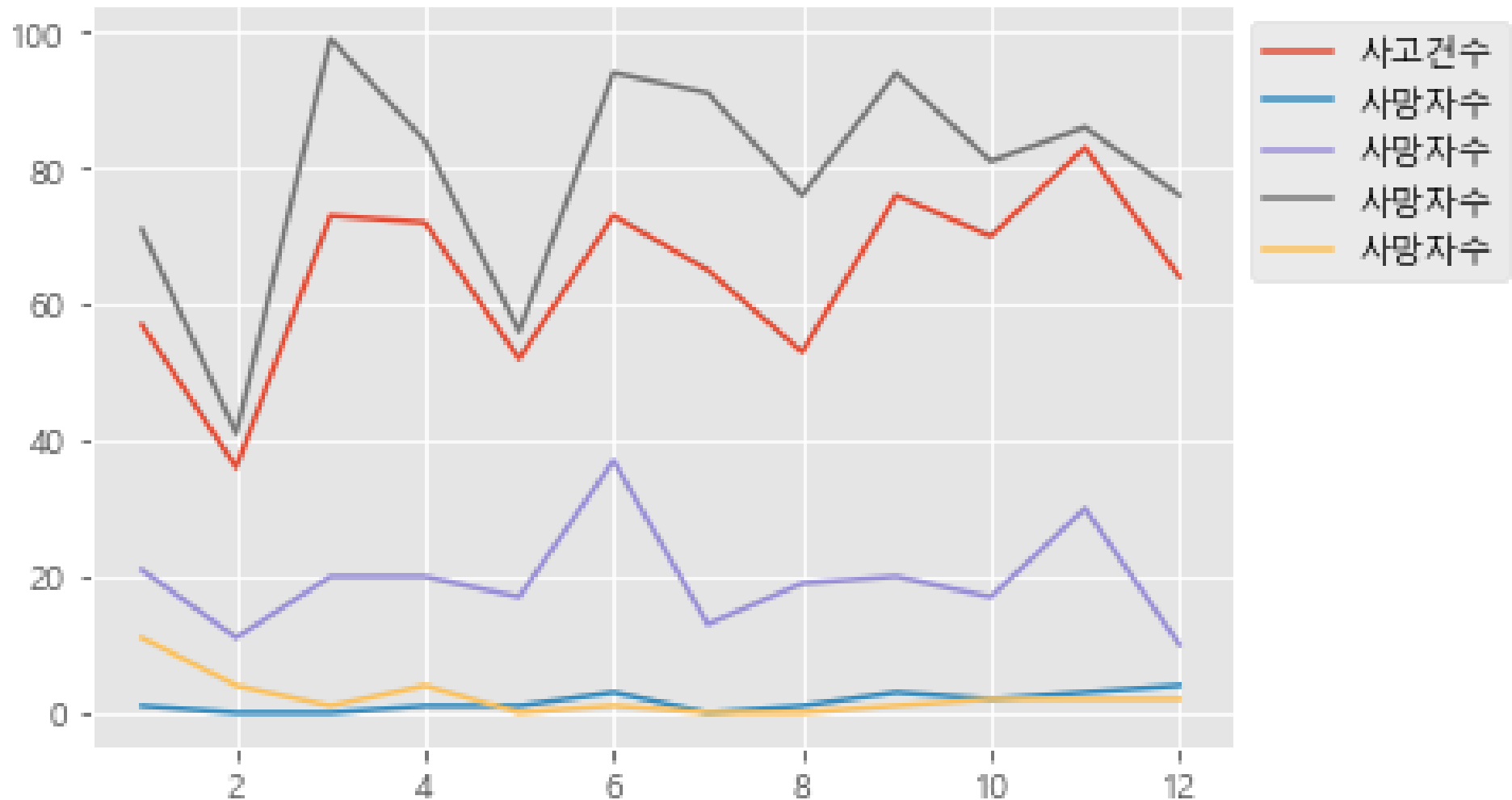
# 지역별 통계

- 지역별 교통사고 추이 그래프
- 각 지역의 월 별 교통사고 사상자 그래프
  - \* 사고건수가 가장 많은 월과 지역 출력
- 원하는 지역과 사상자율 추이가 가장 비슷한 지역 그래프

## 1. 지역별 교통사고 추이 그래프

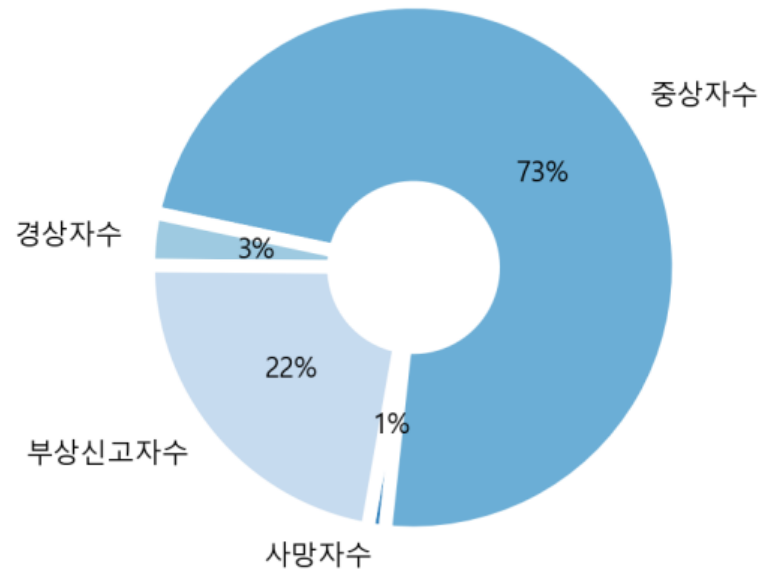
원하는 지역의 이름을 입력해주세요: 경북안동시

경북안동시 지역의 교통사고 추이 그래프

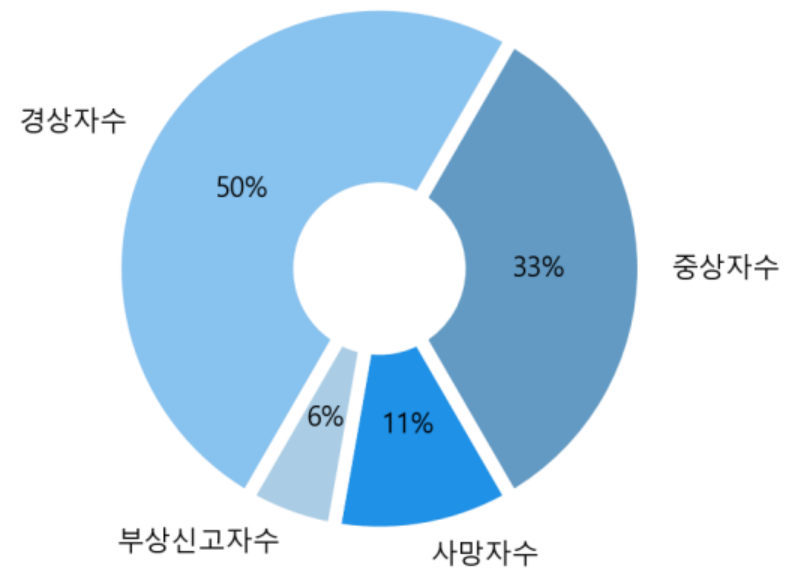


## 2. 각 지역의 월 별 교통사고 사상자 그래프

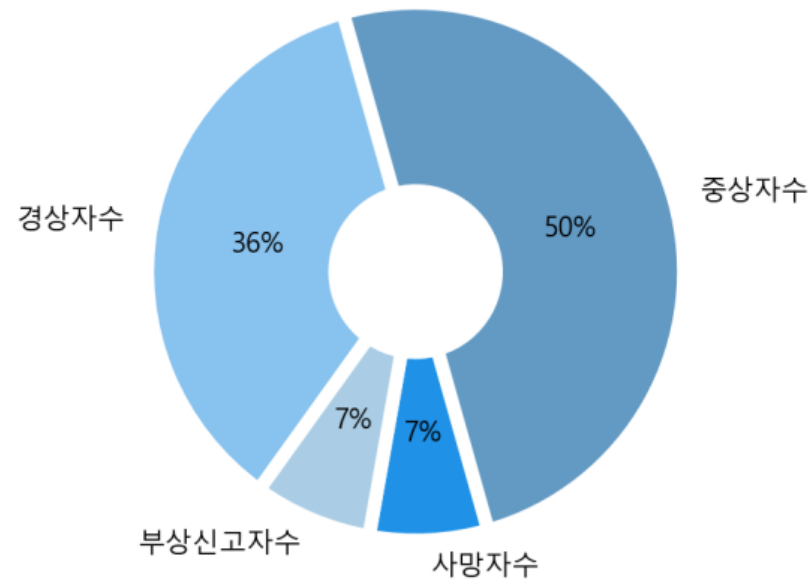
서울종로구 1월 교통사고 사상자 통계



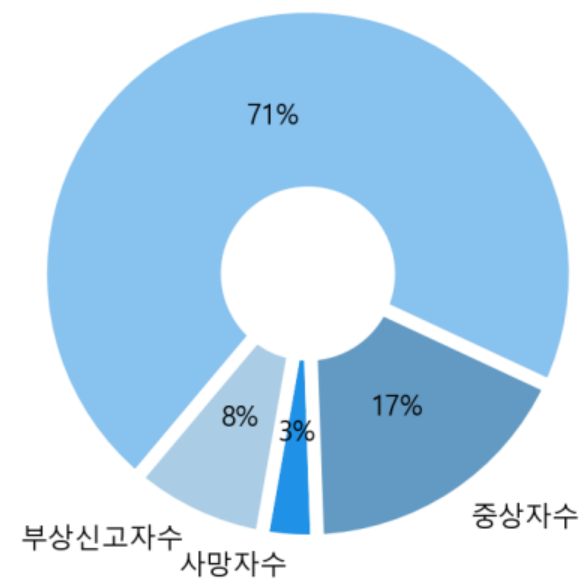
경남밀양시 6월 교통사고 사상자 통계



전북완주군 9월 교통사고 사상자 통계

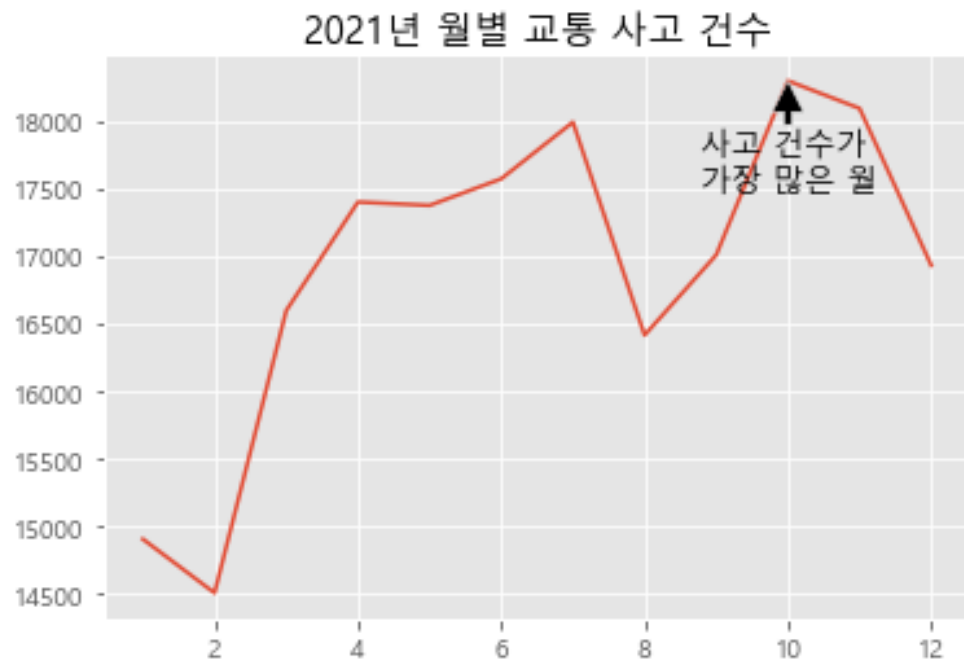


인천미추홀구 10월 교통사고 사상자 통계



## 2-1 사고건수가 가장 많은 월과 지역 출력

사고건수가 가장 많은 월은 10월 입니다.



#사고건수가 제일 많은 지역 출력

```
gb2=df.groupby('지역').sum()
```

```
#del gb2['발생월']
```

```
gb2.reset_index(drop=False, inplace=True)
```

```
i = np.argmax(gb2['사고건수'])
```

```
region = gb2.iloc[i]['지역']
```

```
#np.max(gb2['사고건수'])
```

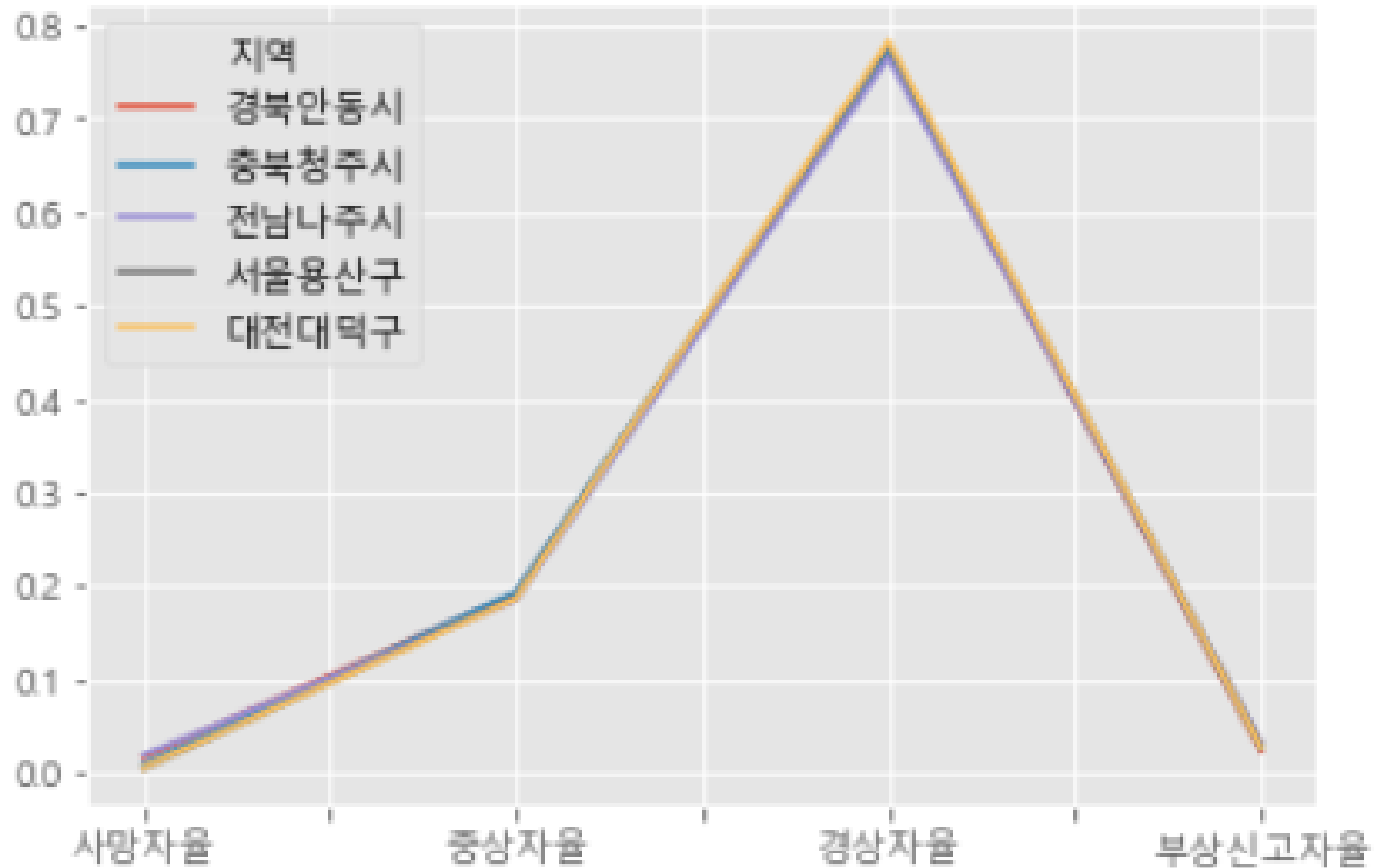
```
print('사고건수가 가장 많은 지역은 ' + region + ' 입니다.')
```

사고건수가 가장 많은 지역은 경기수원시 입니다.

### 3. 원하는 지역과 사상자율 추이가 가장 비슷한 지역 그래프

원하는 지역의 이름을 입력해주세요 : 경북안동시

경북안동시지역과 사상자율 추이가 가장 비슷한 5개 지역







감사합니다. ———