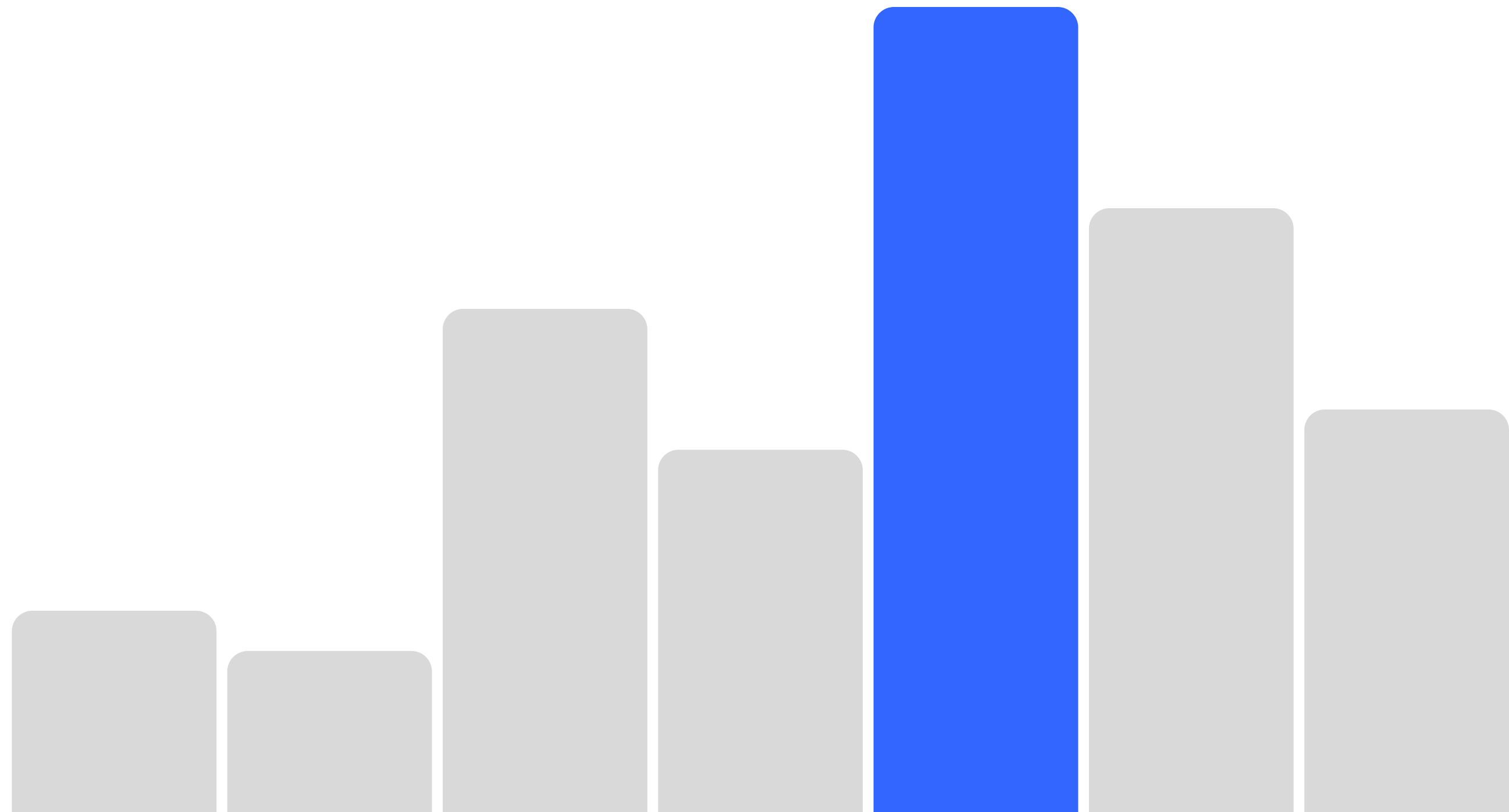


# 빅데이터 분석 프로젝트

## 사망자 교통사고 분석

건설환경공학과  
20003244  
오주흠



# 목차

---

1. 주제 선정 배경

---

2. 데이터 개요

---

3. 데이터 가공

---


4. 시각화 및 분석

---


5. 결론

---

# 1.주제 선정 배경



 연합뉴스TV

[자막뉴스] 새벽예배 다녀오다 참변...춘천서 교통사고로 3명 사망

 강원도민일보

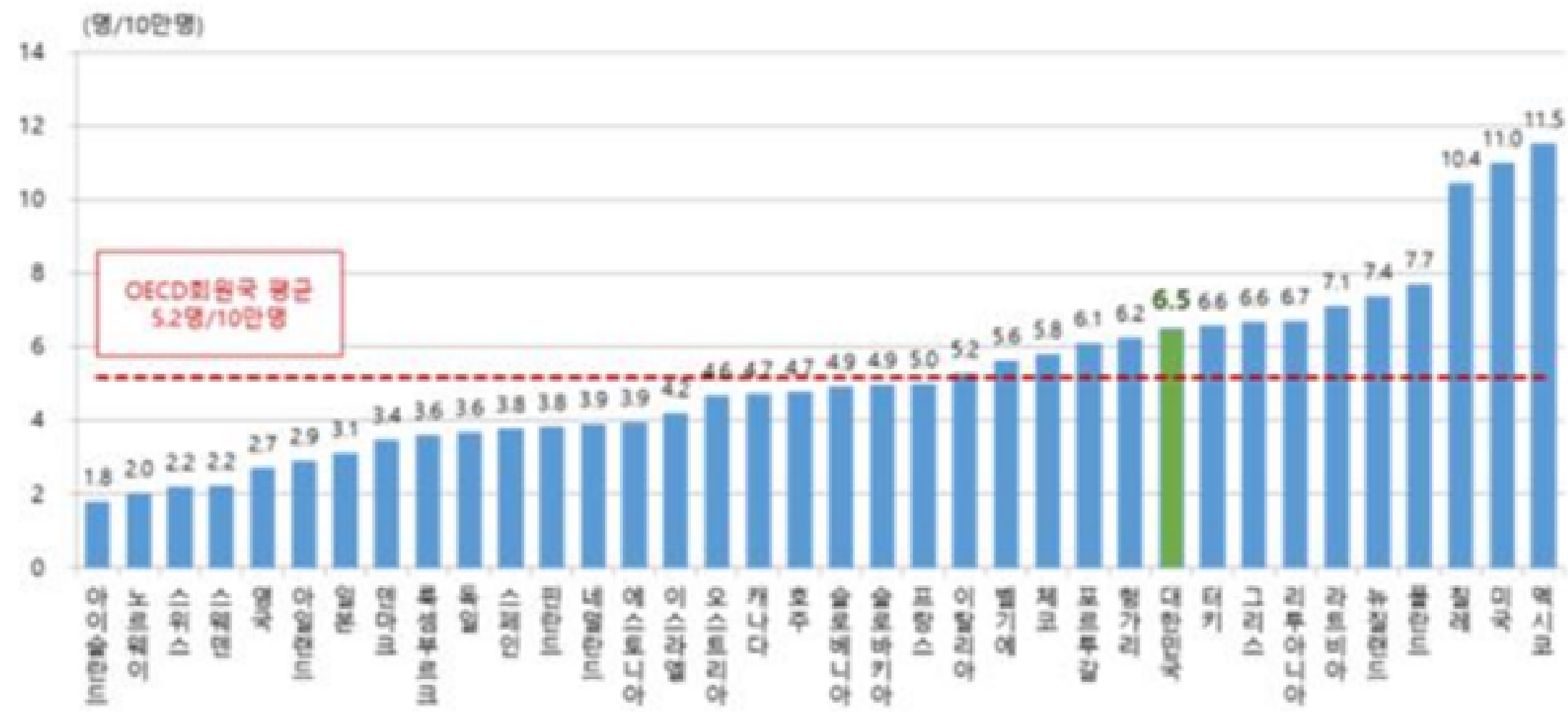
엄마 택배일 돕던 중학생 사망... '신호위반' 가해 운전자 재판

지난 6월 재량휴업일에 모친의 택배 배달일을 돕기 위해 이동 중 신호위반 차량으로 인해 중학생 아들이 숨진 교통사고와 관련, 가해 차량 운전자가 재판에 넘겨졌다.



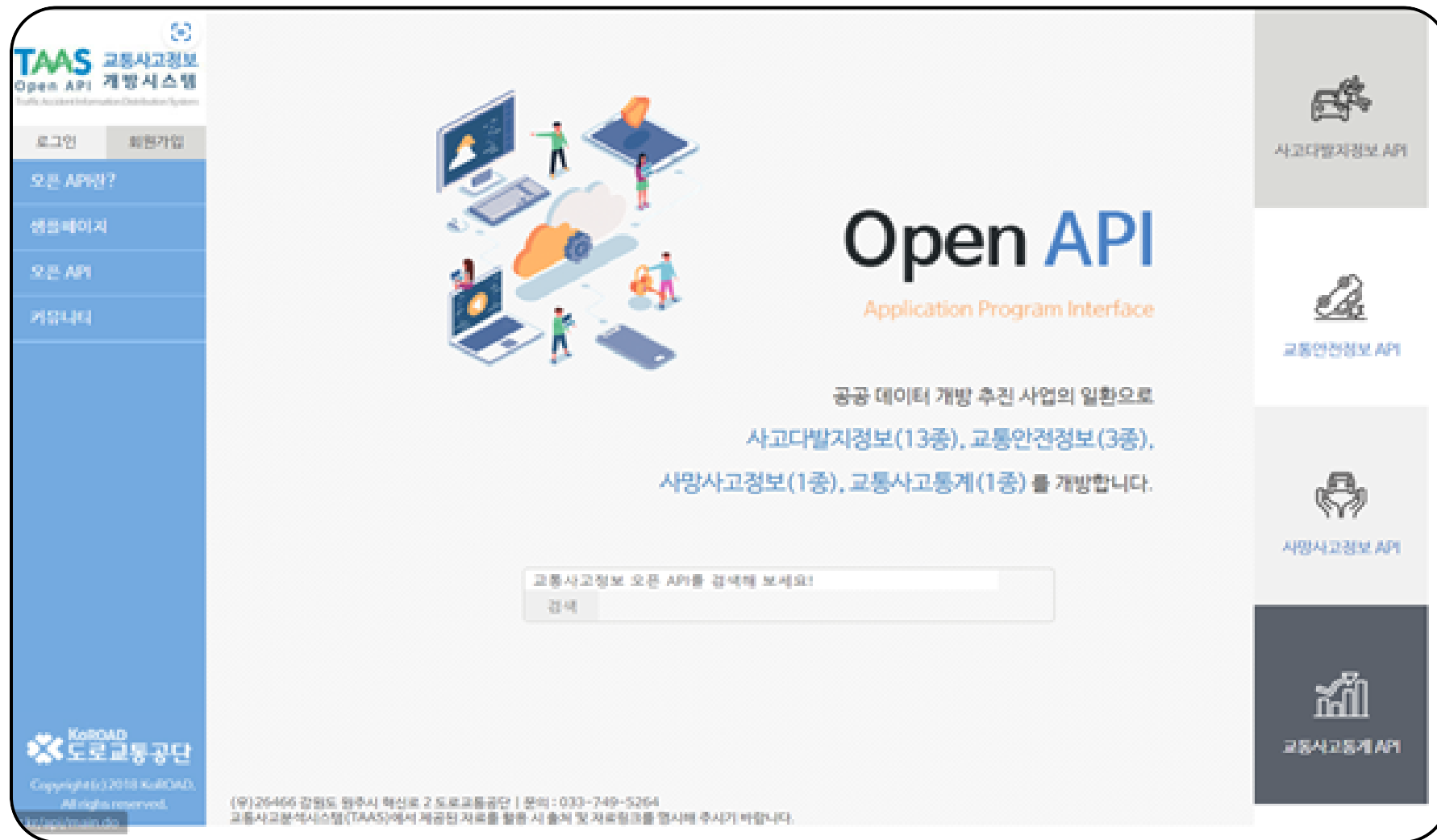
최근 포털사이트의 메인 뉴스 기사에서는 “새벽예배 다녀오다 참변...춘천서 교통사고로 3명 사망”, “엄마 택배일 돕던 중학생 사망” 등 안타까운 교통사고 사례들이 자주 보도되고 있고, 사회적인 관심을 불러일으키고 있다

# 1.주제 선정 배경



인구 10만명 당 교통사고 사망자 27위, 자동차 1만대 당 사망자 31위, 특히 교통사고 사망자 중 보행자 비율이 38.9%로 OECD 회원국 가운데 최하위를 기록

# 2.데이터 개요



## API 레퍼런스

### 기본정보

API 인증키 발급			
신청API	사망사고정보 API <span>JSON</span> <span>XML</span> <span>CSV</span>		
요청 URL	<a href="https://opendata.koroad.or.kr/data/rest/accident/death?authKey=인증키&amp;요청변수">https://opendata.koroad.or.kr/data/rest/accident/death?authKey=인증키&amp;요청변수</a>		
데이터 제공기간	2012~2022		
데이터포맷	JSON, XML	서비스유형	REST
관리부서	데이터융합처	담당자 연락처	033-749-5264
요청변수 코드	<a href="#">코드 다운로드</a>		
사망사고정보 CSV	<a href="#">CSV 다운로드</a> * 해당파일은 openAPI를 활용하지 못하는 분들이 데이터를 활용하실 수 있도록 CSV로 제공합니다.		

데이터 출처 : 교통사고정보개방시스템

<https://opendata.koroad.or.kr/api/selectDeathDataSet.do>

## 2.데이터 개요

---

### 데이터 소개

교통사고 일시 부터 30일이내 사망한경우를 사망교통사고라 정의한다.

### 데이터 대상

전국 약 42,602건의 교통사고를 대상

### 데이터 기간

2012년 1월 1일 ~ 2022년 12월 31일

### 데이터 항목

기본항목(발생년, 발생년월일시, 주야, 요일 등)을 포함하여 인명 피해 통계, 장소 및 사고 유형, 법규 위반 및 도로 형태

# 3.데이터 가공

발생 년	발생년월일 시	주 야	요 일	사 망 자 수	사 상 자 수	중 상 자 수	경 상 자 수	부 상 신 고 자 수	발 생 지 시 도	...	사 고 유 형	법 규 위 반	도 로 형 태 _ 대 분 류	도 로 형 태	당 사 자 종 별_1 _ 대 분 류	당 사 자 종 별_2 _ 대 분 류	발 생 위 치 X_UTMK	발 생 위 치 Y_UTMK	경 도	위 도	
0	2012	2012010101	야 간	일	1	1	0	0	0	서 울	...	차 도 통 행 중	안 전 운 전 의 무 불 이 행	단 일 로	기 타 단 일 로	승 용 차	보 행 자	949860	1957179	126.931890	37.612680



- 1. 발생년월일시를 datetime형식으로 변환
- 2. 필요없는 컬럼 삭제
- 3. 컬럼 이름 변경
- 4. 발생월, 발생일 컬럼 생성
- 5. 컬럼 순서 재정렬

	발생년월일시	발생 년	발생 월	발생시 간	발생지시 도	사 망 자 수	사 상 자 수	중 상 자 수	경 상 자 수	부 상 신 고 자 수	사 고 유 형	법 규 위 반	도 로 형 태	경 도	위 도
0	2012-01-01 01:00:00	2012	1	1	서울	1	1	0	0	0	차대사 람	안전운전 의무 불 이행	단일로	126.931890	37.612680
1	2012-01-01 01:00:00	2012	1	1	전북	1	6	5	0	0	차대차	중앙선 침범	단일로	126.909523	35.633956

### 3.데이터 가공

## 대물피해 환산법 이란?

비중 '1'로 주어진 물피사고를 기준으로  
사망, 중상, 경상, 부상 사고에 상대적인 비중을  
두어 가중치를 곱해 합산한 심각도지수로,  
교통사고의 심각도를 평가하는 방법

$$\text{EPDO} = \text{사망} \times 12 + \text{중상} \times 10 + \text{경상} \times 3 + \text{부상} \times 2 + \text{물피} \times 1$$

```
df = df.assign(심각도 = df['사망자수'] * 12 + df['중상자수'] * 10 + df['경상자수'] * 3 + df['부상신고자수'] * 2 )
df.head()
```

	발생년월일시	발생 년	발생 월	발생시 간	발생지 시도	사망자 수	사상자 수	중상자 수	경상자 수	부상신고 자수	사고유 형	법규위반	도로형 태	경도	위도	심각 도
0	2012-01-01 01:00:00	2012	1	1	서울	1	1	0	0	0	차대사 람	안전운전 의무 불이행	단일로	126.931890	37.612680	12
1	2012-01-01 01:00:00	2012	1	1	전북	1	6	5	0	0	차대차	중앙선 침범	단일로	126.909523	35.633958	62



# 3.데이터 가공

## 심각도 분석을 위해 새로운 데이터 프레임 생성

발생년	2012	2013	2014	2015	2016	2017	2018	2019	20
사망자수	5392	5092	4762	4621	4292	4185	3781	3349	30
사상자수	8665	8068	7222	7282	6493	6390	5696	5121	46
심각도	86573	81188	73245	73158	65279	63322	56665	50459	457

## 사고건수 컬럼 추가를 위한 새로운 데이터프레임 생성

	발생년	사망자수	사상자수	심각도	사고건수
0	2012	5392	8665	86573	5165
1	2013	5092	8068	81188	4876
2	2014	4762	7222	73245	4583
3	2015	4621	7282	73158	4447
4	2016	4292	6493	65279	4119
5	2017	4185	6390	63322	4065
6	2018	3781	5696	56665	3657
7	2019	3349	5121	50459	3233

## 발생년 컬럼 기준으로 병합

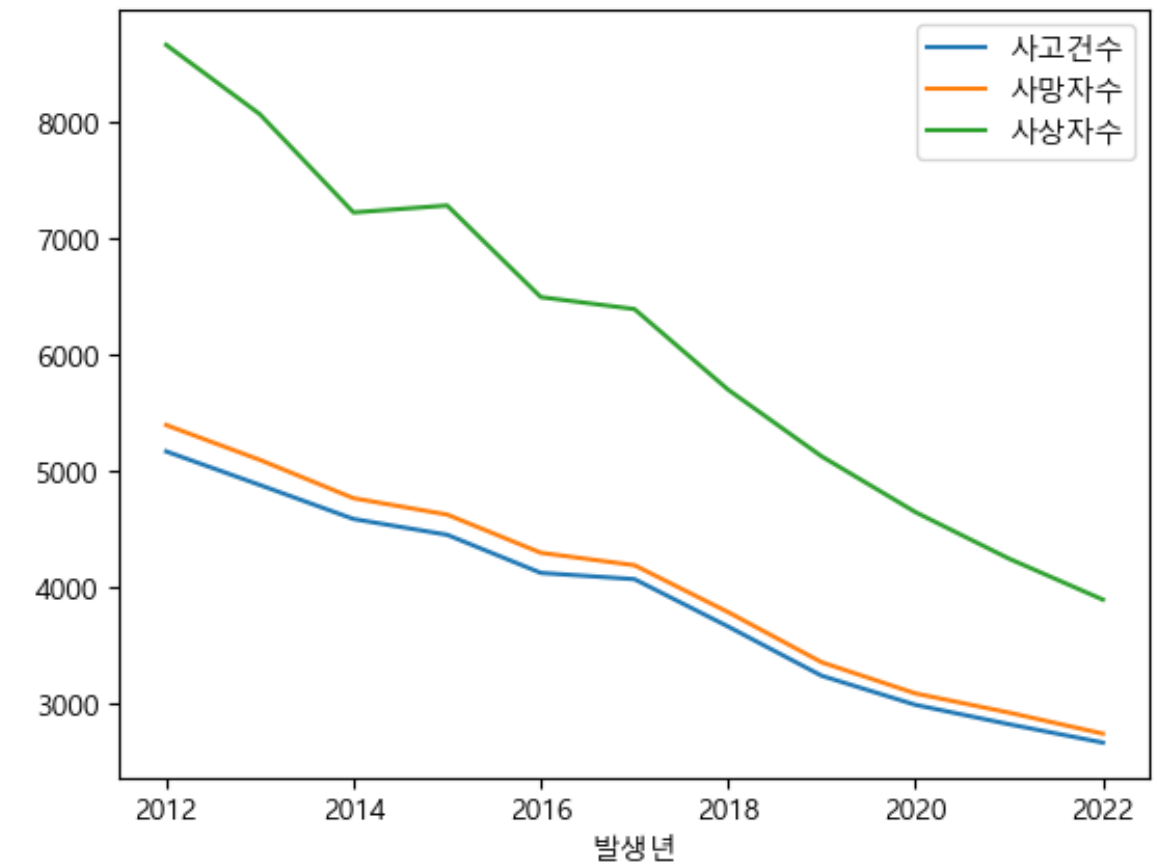
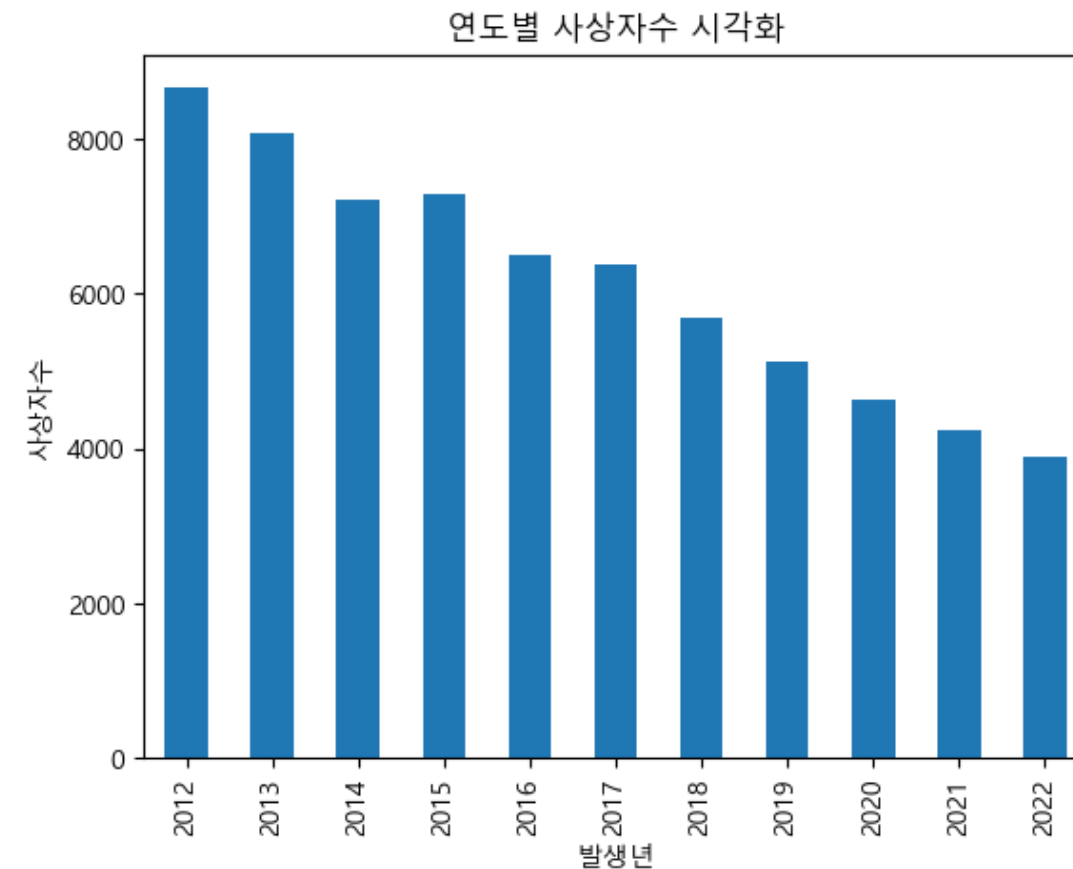
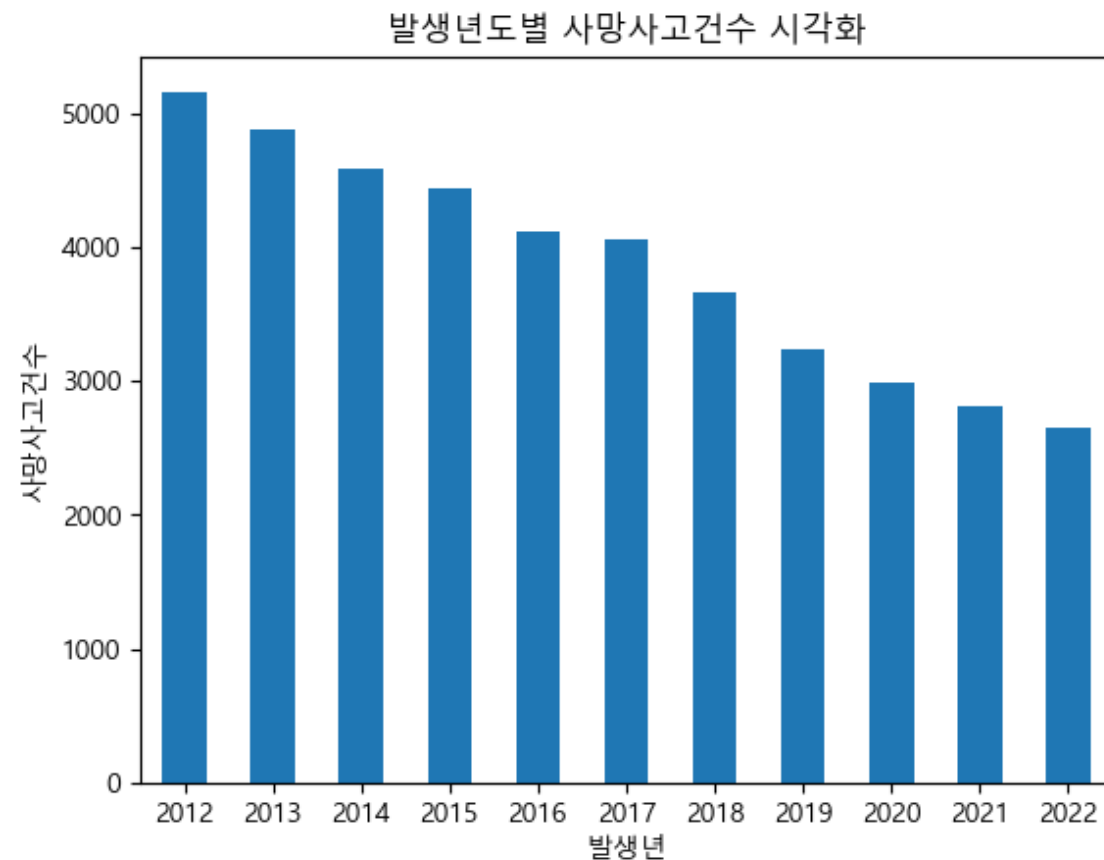
	발생년	사망자수	사상자수	심각도	사고건수
0	2012	5392	8665	86573	5165
1	2013	5092	8068	81188	4876
2	2014	4762	7222	73245	4583
3	2015	4621	7282	73158	4447
4	2016	4292	6493	65279	4119
5	2017	4185	6390	63322	4065
6	2018	3781	5696	56665	3657
7	2019	3349	5121	50459	3233
8	2020	3081	4642	45771	2983
9	2021	2916	4241	42730	2816
10	2022	2735	3889	39240	2658

# 3.데이터 가공

## 지수별 증감률 확인을 위한 컬럼 추가

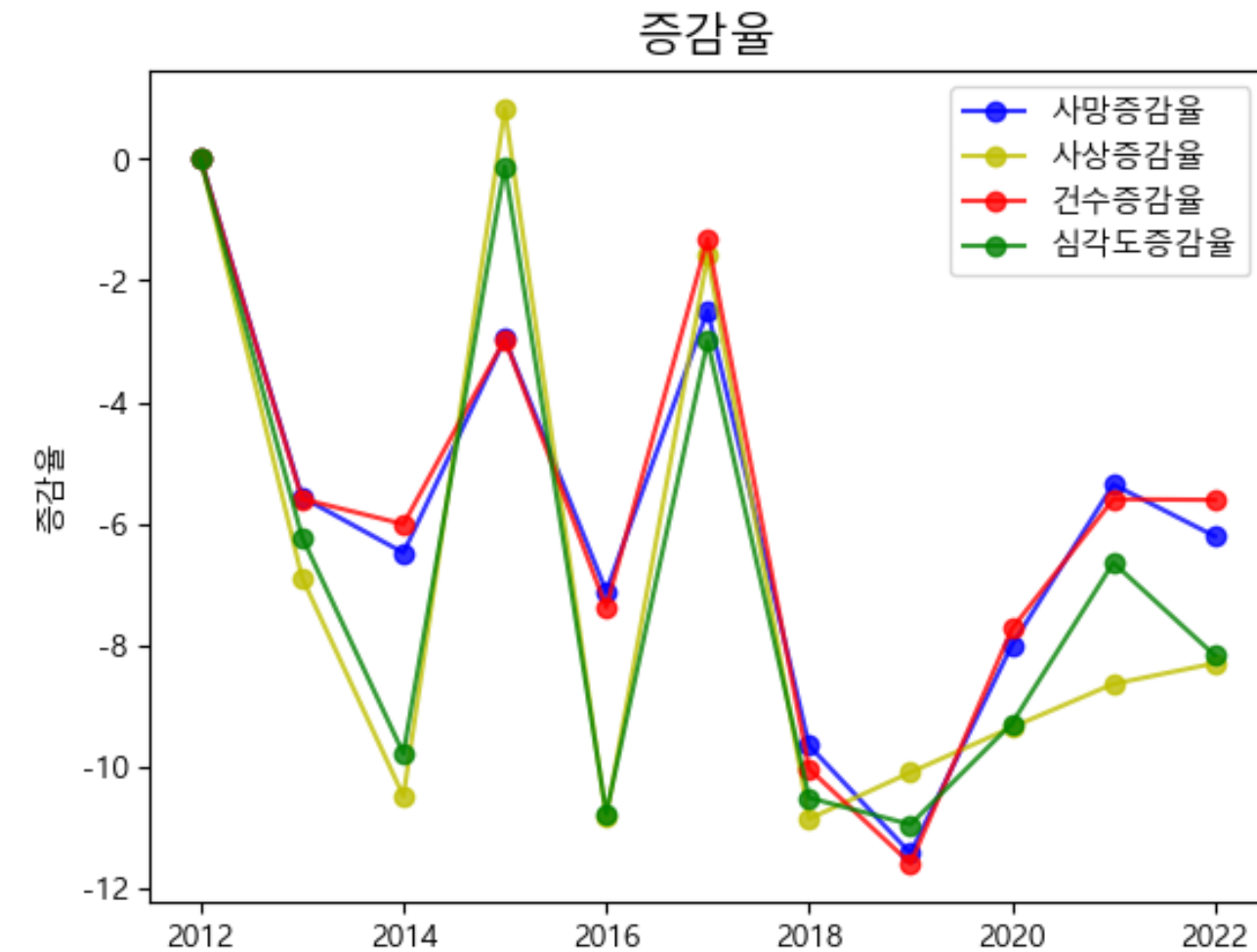
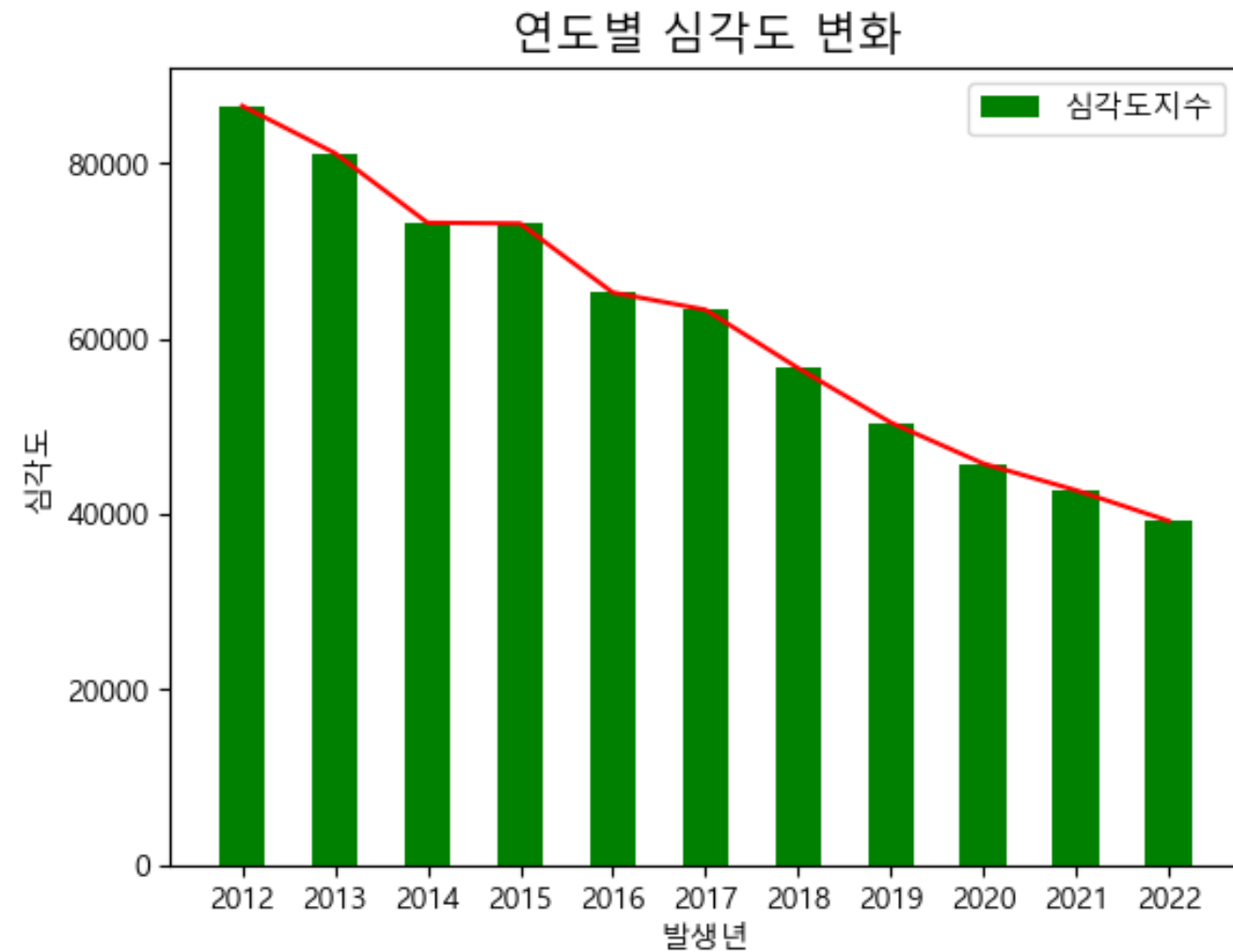
	발생년	사망자수	사상자수	사고건수	심각도	사망증감	사상증감	건수증감	심각도증감	사망증감율	사상증감율	건수증감율	심각도증감율
0	2012	5392	8665	5165	86573	0.0	0.0	0.0	0.0	0.000000	0.000000	0.000000	0.000000
1	2013	5092	8068	4876	81188	-300.0	-597.0	-289.0	-5385.0	-5.563798	-6.889786	-5.595353	-6.220184
2	2014	4762	7222	4583	73245	-330.0	-846.0	-293.0	-7943.0	-6.480754	-10.485870	-6.009024	-9.783466
3	2015	4621	7282	4447	73158	-141.0	60.0	-136.0	-87.0	-2.960941	0.830795	-2.967489	-0.118779
4	2016	4292	6493	4119	65279	-329.0	-789.0	-328.0	-7879.0	-7.119671	-10.834935	-7.375759	-10.769841
5	2017	4185	6390	4065	63322	-107.0	-103.0	-54.0	-1957.0	-2.493010	-1.586324	-1.310998	-2.997901
6	2018	3781	5696	3657	56665	-404.0	-694.0	-408.0	-6657.0	-9.653524	-10.860720	-10.036900	-10.512934
7	2019	3349	5121	3233	50459	-432.0	-575.0	-424.0	-6206.0	-11.425549	-10.094803	-11.594203	-10.952087
8	2020	3081	4642	2983	45771	-268.0	-479.0	-250.0	-4688.0	-8.002389	-9.353642	-7.732756	-9.290711
9	2021	2916	4241	2816	42730	-165.0	-401.0	-167.0	-3041.0	-5.355404	-8.638518	-5.598391	-6.643945
10	2022	2735	3889	2658	39240	-181.0	-352.0	-158.0	-3490.0	-6.207133	-8.299929	-5.610795	-8.167564

### 3. 시각화/ 분석



- 사망 사고 건수와 사상자수는 2012년부터 2022년까지 감소하는 추세를 보인다.
- 사고 건수가 줄어들수록 사망자 수 사상자 수가 줄어드는 것을 확인할 수 있다.
- 사상자 수가 상대적으로 크게 나타나는 이유는 한 건의 교통사고에서 사상자가 많이 발생할 수 있기 때문이라고 추측된다.

## 4. 시각화/ 분석



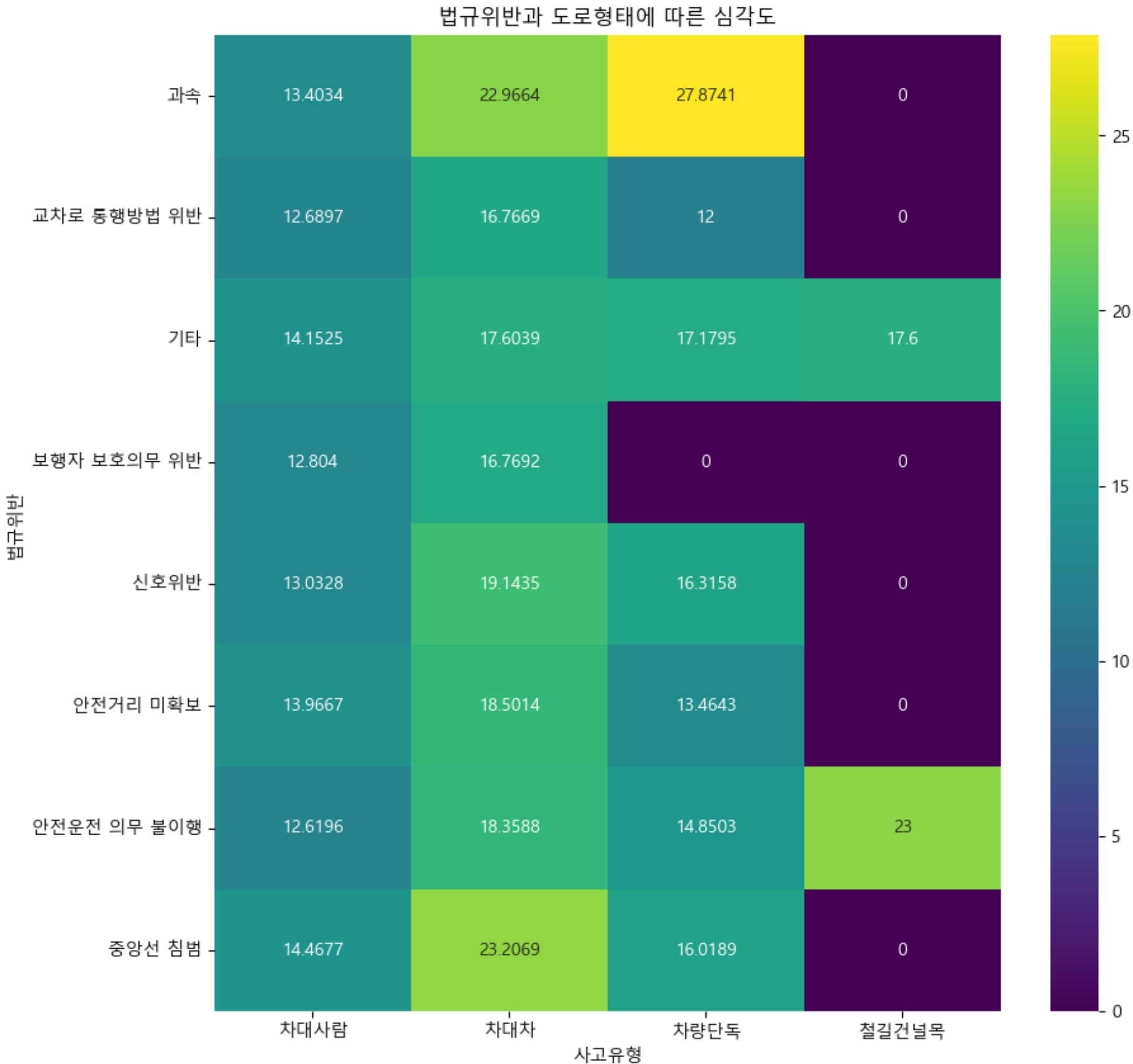
- 사망사고건수와 사상자수 그래프와 마찬가지로 심각도 또한 감소하는 추세
- 2018년 부터 2019년 대부분의 지표가 큰 폭으로 감소
- 교통사고 예방 정책이 효과를 보인다고 판단

# 4. 시각화/ 분석

법규위반과 사고유형에 따른 심각도  
피벗 테이블 생성

사고유형	차대사람	차대차	차량단독	철길건널목
법규위반				
과속	13.403433	22.966355	27.874126	0.0
교차로 통행방법 위반	12.689655	16.766871	12.000000	0.0
기타	14.152542	17.603908	17.179487	17.6
보행자 보호의무 위반	12.804023	16.769231	0.000000	0.0
신호위반	13.032825	19.143485	16.315789	0.0
안전거리 미확보	13.966667	18.501445	13.464286	0.0
안전운전 의무 불이행	12.619649	18.358778	14.850278	23.0

중앙선 침범 14.467742 23.206920 16.018868 0.0



# 4. 시각화/ 분석

과속

"차대사람", "차대차", "차량단독"에서 모두 높은 심각도를 보이고 있다. 특히 "차량단독"에서 가장 높은 심각도를 나타낸다.

보행자 보호의무 위반

보행자 보호의무 위반"은 "차대사람"과 "차대차"에서 높은 심각도를 보인다.

안전의무 불이행

"차대사람"과 "차량단독"에서 높은 심각도를 보이며, 특히 "철길건널목"에서 가장 높은 심각도를 나타낸다.

안전거리 미확보

"차대사람"과 "차대차"에서 높은 심각도를 보인다.

신호 위반

모든 사고유형에서 상대적으로 높은 심각도를 나타낸다.

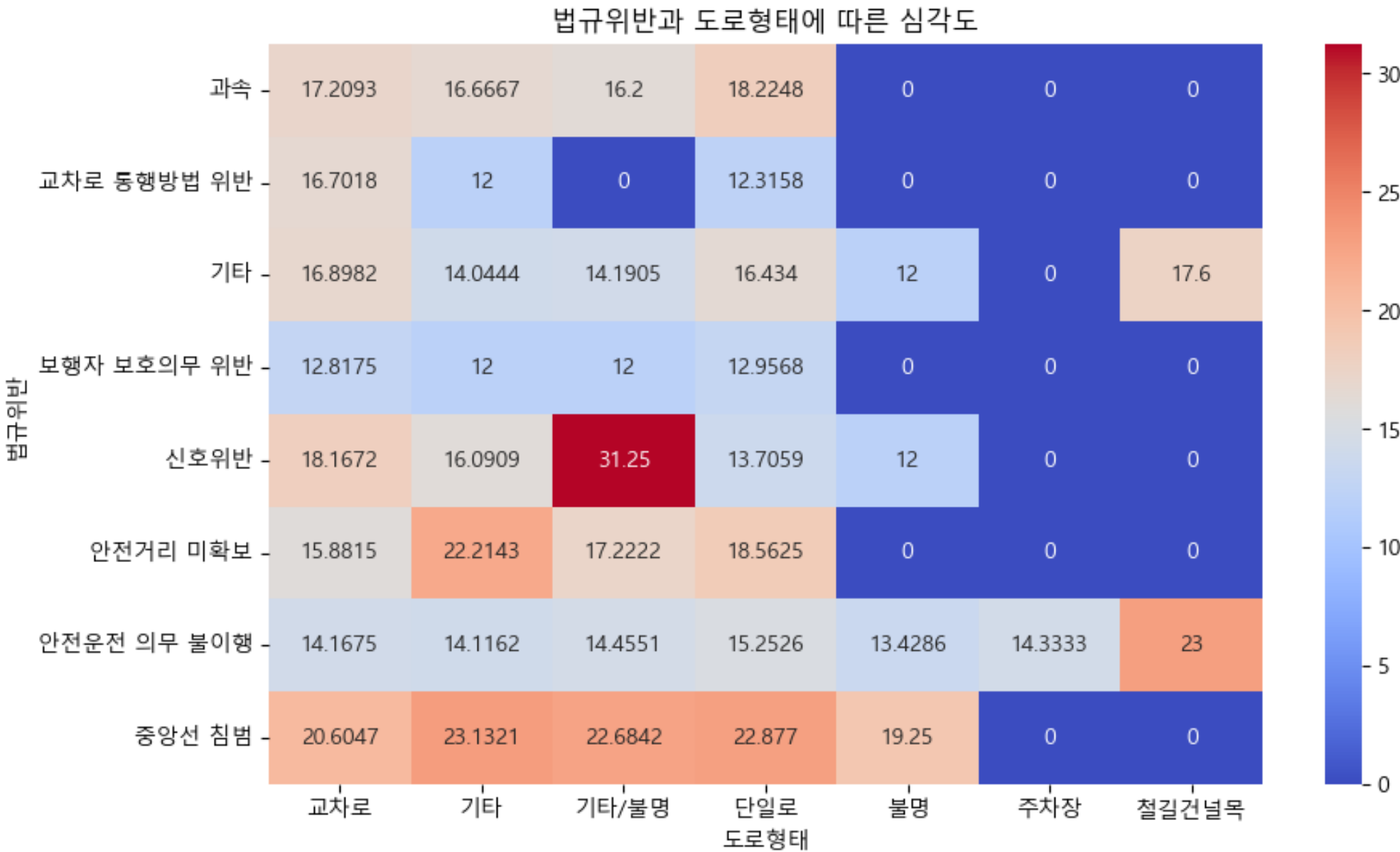
중앙선 침범

"중앙선 침범"은 "차대사람"과 "차대차"에서 높은 심각도를 보인다

# 4. 시각화/ 분석

## 법규위반과 도로형태에 따른 심각도 피벗 테이블 생성

도로형태	교차로	기타	기타/불명	단일로	불명	주차장	철길건널목
법규위반							
과속	17.209302	16.666667	16.200000	18.224835	0.000000	0.000000	0.0
교차로 통행방법 위반	16.701807	12.000000	0.000000	12.315789	0.000000	0.000000	0.0
기타	16.898190	14.044444	14.190476	16.433962	12.000000	0.000000	17.6
보행자 보호의무 위반	12.817544	12.000000	12.000000	12.956772	0.000000	0.000000	0.0
신호위반	18.167185	16.090909	31.250000	13.705882	12.000000	0.000000	0.0
안전거리 미확보	15.881481	22.214286	17.222222	18.562500	0.000000	0.000000	0.0
안전운전 의무 불이행	14.167545	14.116239	14.455115	15.252569	13.428571	14.333333	23.0
중앙선 침범	20.604706	23.132075	22.684211	22.876964	19.250000	0.000000	0.0



## 4. 시각화/ 분석

### 과속

“단일로”에서 과속에 의한 교통사고의 심각도가 가장 높았다. 이는 주로 고속도로나 국도 등의 단일 도로에서의 과속이 교통사고의 심각성을 증가시킨다는 것을 시사

### 교차로 통행방법 위반

“보행자 보호의무 위반”은 모든 도로에서 비슷한 수치를 나타낸다. 이는 도로형태에 상관없이 발생할 수 있는 교통사고의 원인임을 알 수 있다.

### 안전의무 불이행

“안전운전 의무 불이”행은 철길건널목을 제외한 모든 도로에서 비슷한 수치를 나타낸다. 이는 도로에 상관없이 발생할 수 있는 교통사고의 원인임을 알 수 있다.

### 안전거리 미확보

기타/불명”, “단일로”에서 안전거리 미확보에 의한 교통사고의 심각도가 높다. 고속도로 등에서의 차간 거리 유지에 대한 중요성 강조를 시사

### 신호 위반

“기타/불명”에서 신호위반에 따른 교통사고의 심각도가 가장 높았지만, “단일로”와 “교차로”에서도 높은 수치를 보이며, 신호위반이 심각한 교통사고를 초래하는 것으로 해석할 수 있다.

### 중앙선 침범

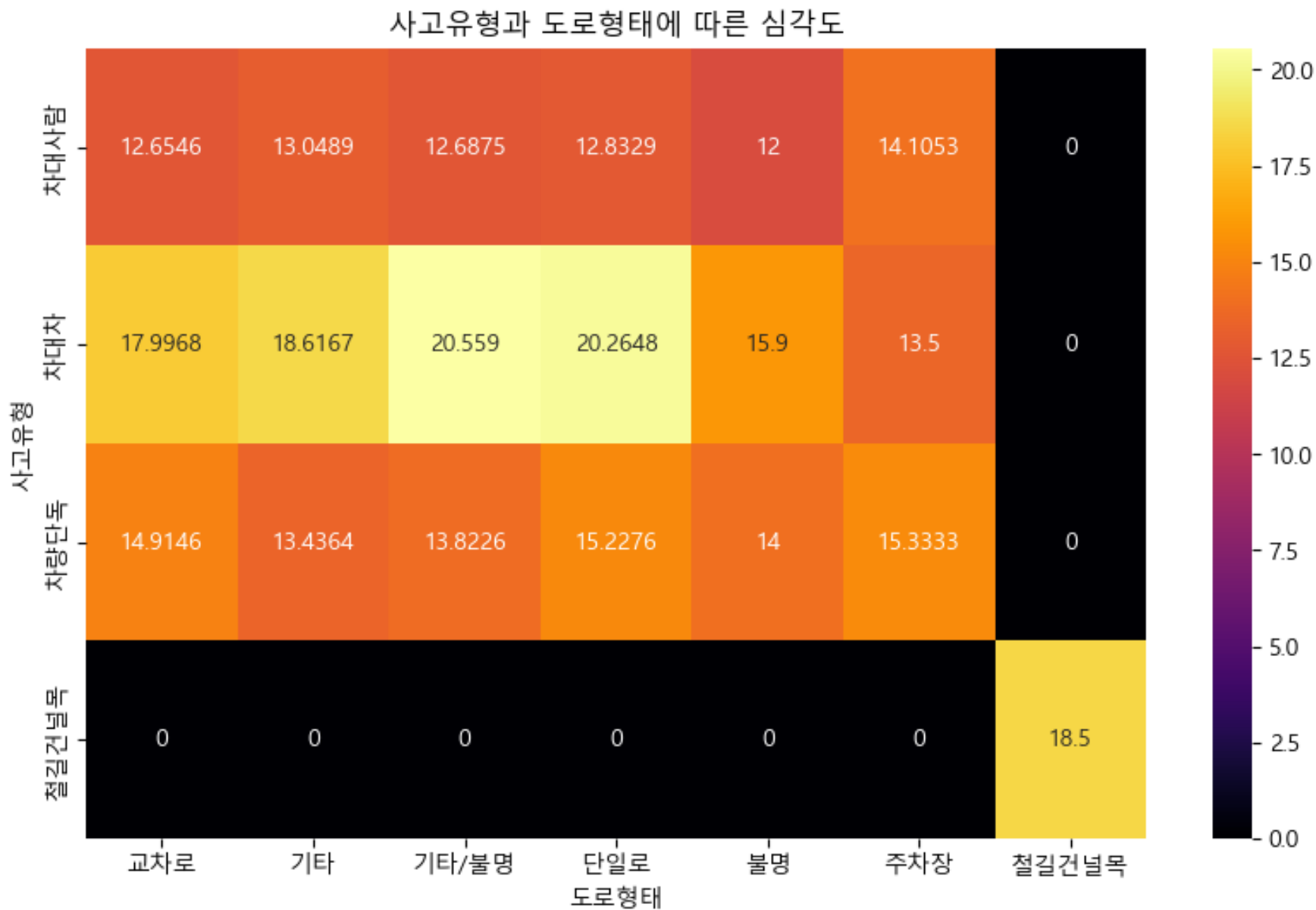
모든 도로에서 중앙선 침범에 의한 교통사고의 심각도가 매우 높다. 모든 도로에서 중앙선 침범이 교통사고의 가장 큰 원인 중 하나로 작용하는 것으로 보인다.



# 4. 시각화/ 분석

## 사고유형과 도로형태에 따른 심각도 피벗 테이블 생성

도로형태	교차로	기타	기타/불명	단일로	불명	주차장	철길건널목
사고유형							
차대사람	12.654630	13.048872	12.687500	12.832902	12.0	14.105263	0.0
차대차	17.996769	18.616667	20.559006	20.264811	15.9	13.500000	0.0
차량단독	14.914626	13.436441	13.822581	15.227641	14.0	15.333333	0.0
철길건널목	0.000000	0.000000	0.000000	0.000000	0.0	0.000000	18.5



## 4. 시각화/ 분석

---

### 차대사람

“주차장”에서의 차대사람 사고의 심각도가 높는데, 주차장에서의 차량 속도가 낮음에도 불구하고, 운전자들이 주의를 기울이지 않아 발생한 것으로 추측된다.

### 차대차

단일로와 기타/불명 도로형태에서 차대차 사고의 심각도가 높게 나타났다. 특히 단일로에서의 사고가 두드러지며, 이는 차량이 서로 마주치는 구간에서의 충돌이 더 위험하다는 것을 시사한다.

### 차량단독

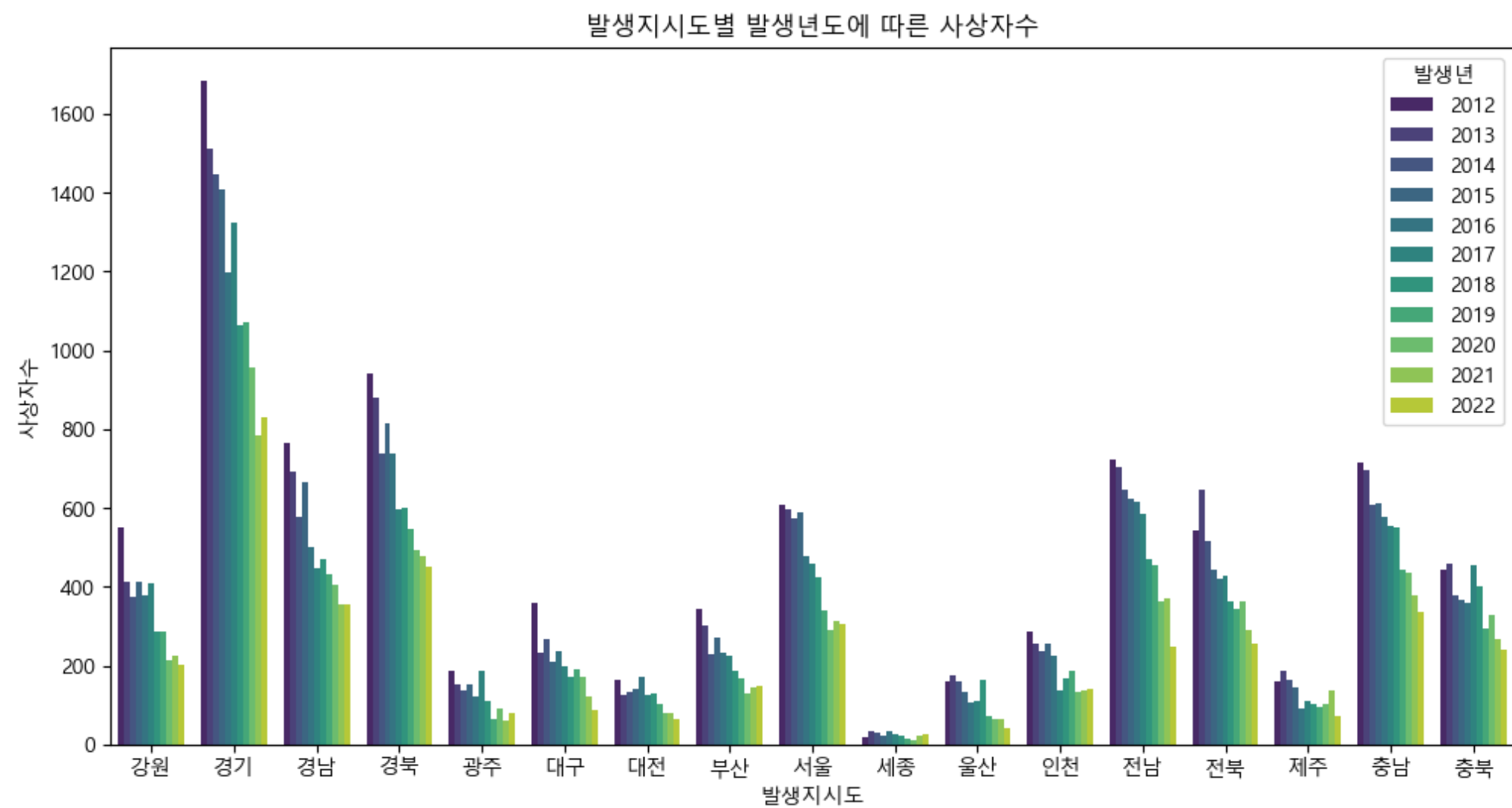
주로 주차장과 단일로에서 높은 심각도를 보이고 있으며, 차량 단독 사고의 원인은 주로 과속, 도로 이탈, 운전자의 주의 부족 등이 될 수 있다

## 4. 시각화/ 분석

### 발생지시도별 발생년, 월별 사상자 발생 교통사고

```
year_month_casualty = df.groupby(['발생지시도', '발생년', '발생월'])['사상자수'].sum().reset_index()
year_month_casualty
```

	발생지시도	발생년	발생월	사상자수
0	강원	2012	1	20
1	강원	2012	2	121
2	강원	2012	3	38
3	강원	2012	4	30
4	강원	2012	5	45
...	...	...	...	...
2198	충북	2022	8	16
2199	충북	2022	9	32
2200	충북	2022	10	21
2201	충북	2022	11	20
2202	충북	2022	12	15



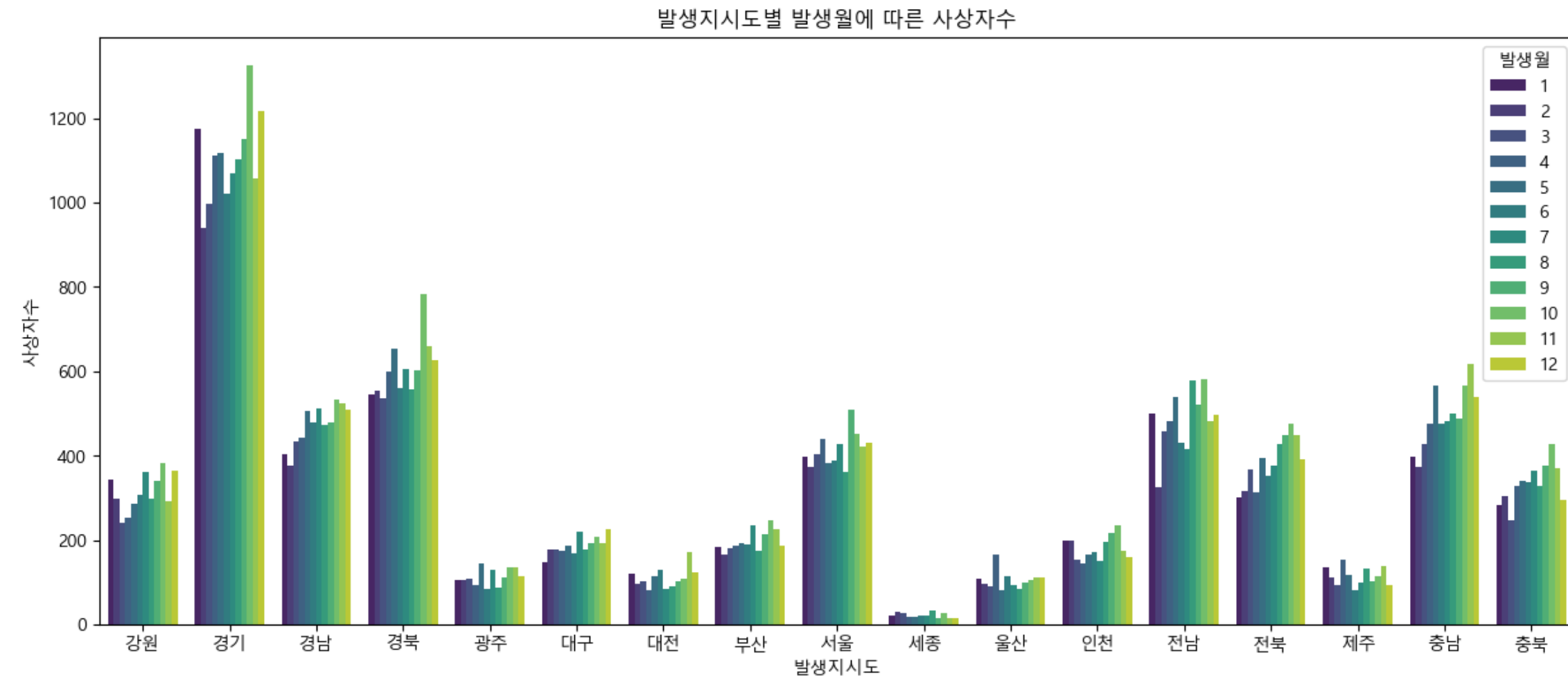
- 대부분의 지역에서 2022년으로 갈수록 사상자수가 줄어드는 것을 확인할 수 있다

# 4. 시각화/ 분석

## 발생월에 따른 지역별 사상자수

```
month_casualty = year_month_casualty.groupby(['발생월', '발생지시도'])['사상자수'].sum().reset_index()
month_casualty
```

	발생월	발생지시도	사상자수
0	1	강원	342
1	1	경기	1173
2	1	경남	405
3	1	경북	546
4	1	광주	104
...	...	...	...
199	12	전남	498
200	12	전북	390
201	12	제주	92
202	12	충남	540
203	12	충북	296



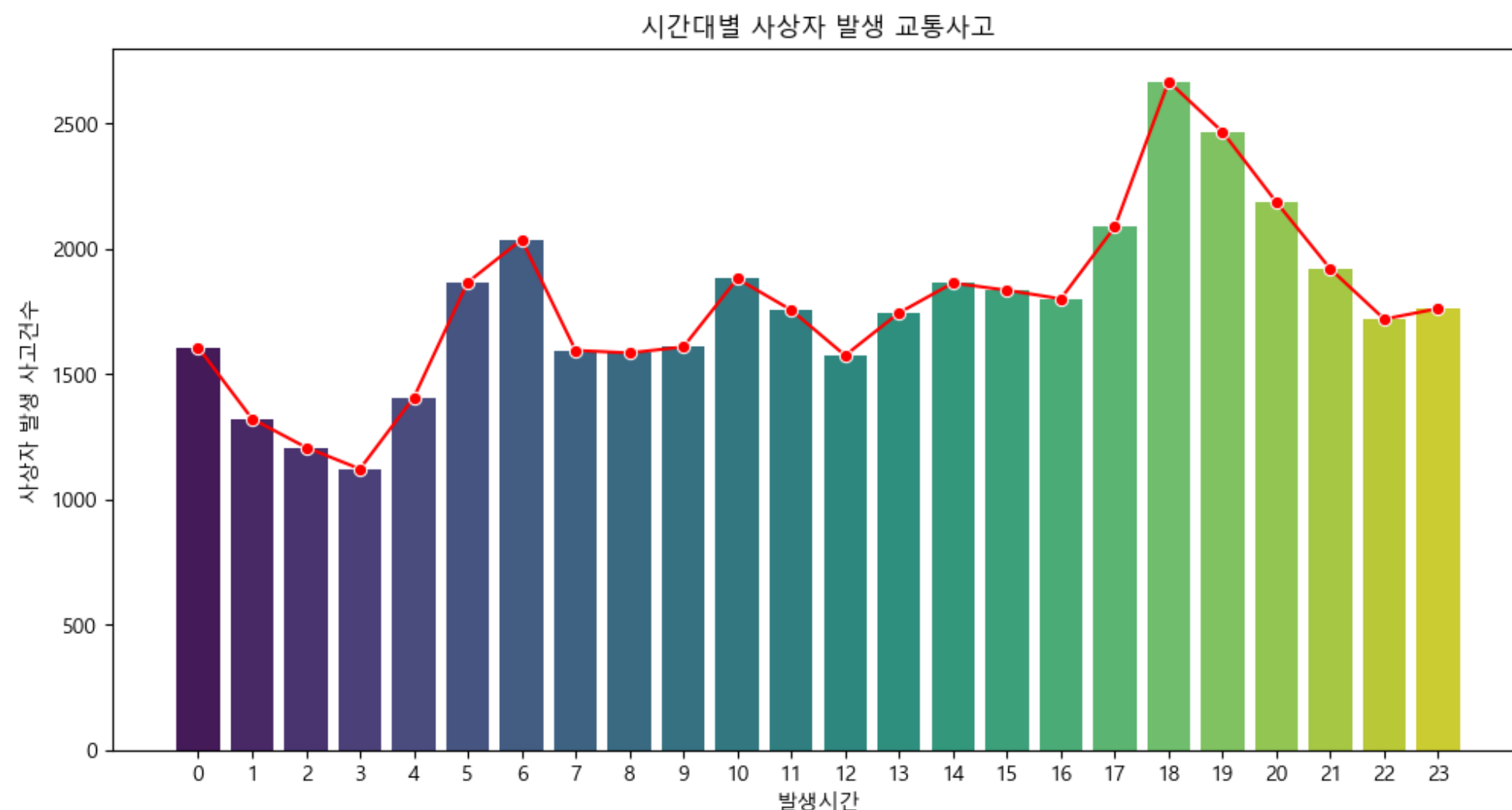
- 대부분의 지역에서 10~11월에 교통사고 사망자수가 많다는 것을 확인 할 수 있다
- 대표적인 가을 행락철로 이동인구가 늘어나는 것이 원인
- 운전자들에게 교통사고 위험성에 대한 경각심을 상기시키고, 과속, 졸음운전 등 교통사고 유발 위험 운전에 대한 특별 단속과 예방 홍보를 강화할 필요가 있다

# 4. 시각화/ 분석

## 시간대별 사상자 발생 교통사고 건수

```
hour_casualties = df.groupby('발생시간')['사상자수'].count().reset_index() # count로 사상자수의 합이 아닌 사상자수 발생 건수
hour_casualties
```

발생시간	사상자수
0	1606
1	1321
2	1207
3	1121
4	1406
5	1866
6	2036
7	1594
8	1584
9	1608
10	1880



- 교통량이 많은 출근시간대와 퇴근시간대에 주로 발생하는 것을 확인 할 수 있다

# 4. 시각화/ 분석

## 서울지역에서의 기온과 교통사고의 관계를 알아보기 위해 데이터 준비



### 기온분석 - 그래프

그래프	분포도
-----	-----

#### 자료설명

oracle.sql.CLOB@208de57a

\* '지역/지점'의 '지역'은 전국 및 광역 단위의 평균 제공(1973년~)

- 전국 및 광역별 평균 산출에 사용되는 지점은 62개 지점이며 제주도 4개 지점은 제주지역 산출에만 사용됩니다.

[[전국/지역별 통계 산출 지점 정보\(더보기\)](#)]

#### 검색조건

자료구분	일	자료형태	기본	기간	20120101	~	20221231
지역/지점	서울	선택					
> 검색							

CSV	Excel
-----	-------

데이터 출처 : 기상정보개방포털

<https://opendata.koroad.or.kr/api/selectDeathDataSet.do>

## 2.기온데이터 전처리 후

		날짜	평균기온
0	2012-01-01		-3.0
1	2012-01-02		-4.8
2	2012-01-03		-4.5
3	2012-01-04		-7.4
4	2012-01-05		-5.7

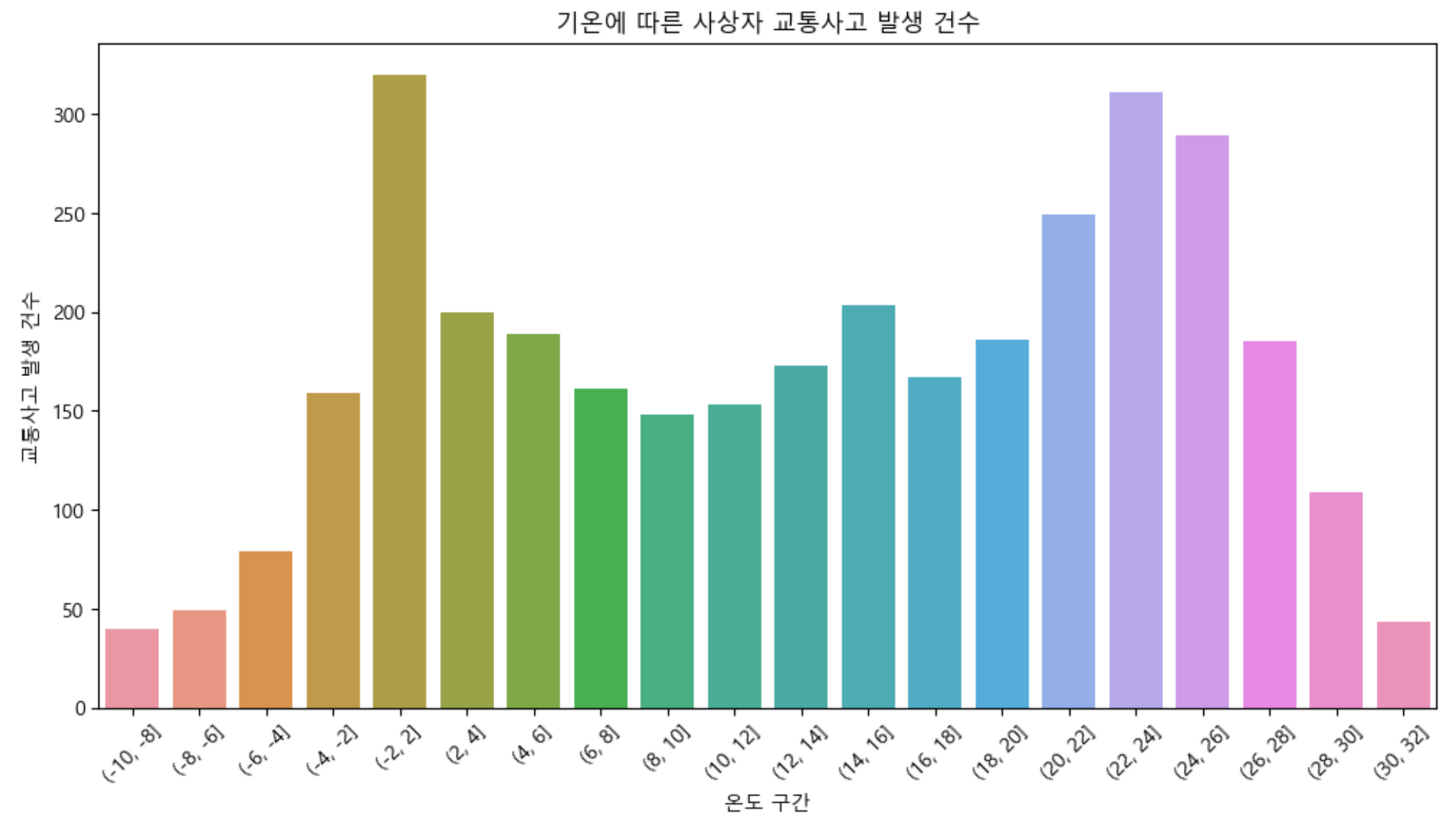
[illegible]



# 4. 시각화/ 분석

필요한 컬럼만 추출하고 온도구간을 나누어 그에 따른 사상자수 집계

발생년월일시				평균기온	사상자수	온도구간			사상자수
0	2012-01-01	01:00:00		-3.0	1	0	(-10, -8]		40
1	2012-01-01	02:00:00		-3.0	4	1	(-8, -6]		49
2	2012-01-01	02:00:00		-3.0	1	2	(-6, -4]		79
3	2012-01-01	18:00:00		-3.0	1	3	(-4, -2]		159
4	2012-01-03	07:00:00		-4.5	2	4	(-2, 2]		320
...	...			...	...	5	(2, 4]		200
3430	2022-12-26	11:00:00		-3.9	1	6	(4, 6]		189
3403	2022-12-27	13:00:00		-2.6	1	7	(6, 8]		161
3433	2022-12-28	09:00:00		-3.3	1	8	(8, 10]		148
3437	2022-12-29	15:00:00		-2.9	5	9	(10, 12]		153
3412	2022-12-30	23:00:00		-1.8	2	10	(12, 14]		173
						11	(14, 16]		203
						12	(16, 18]		167
						13	(18, 20]		186
						14	(20, 22]		249
						15	(22, 24]		311

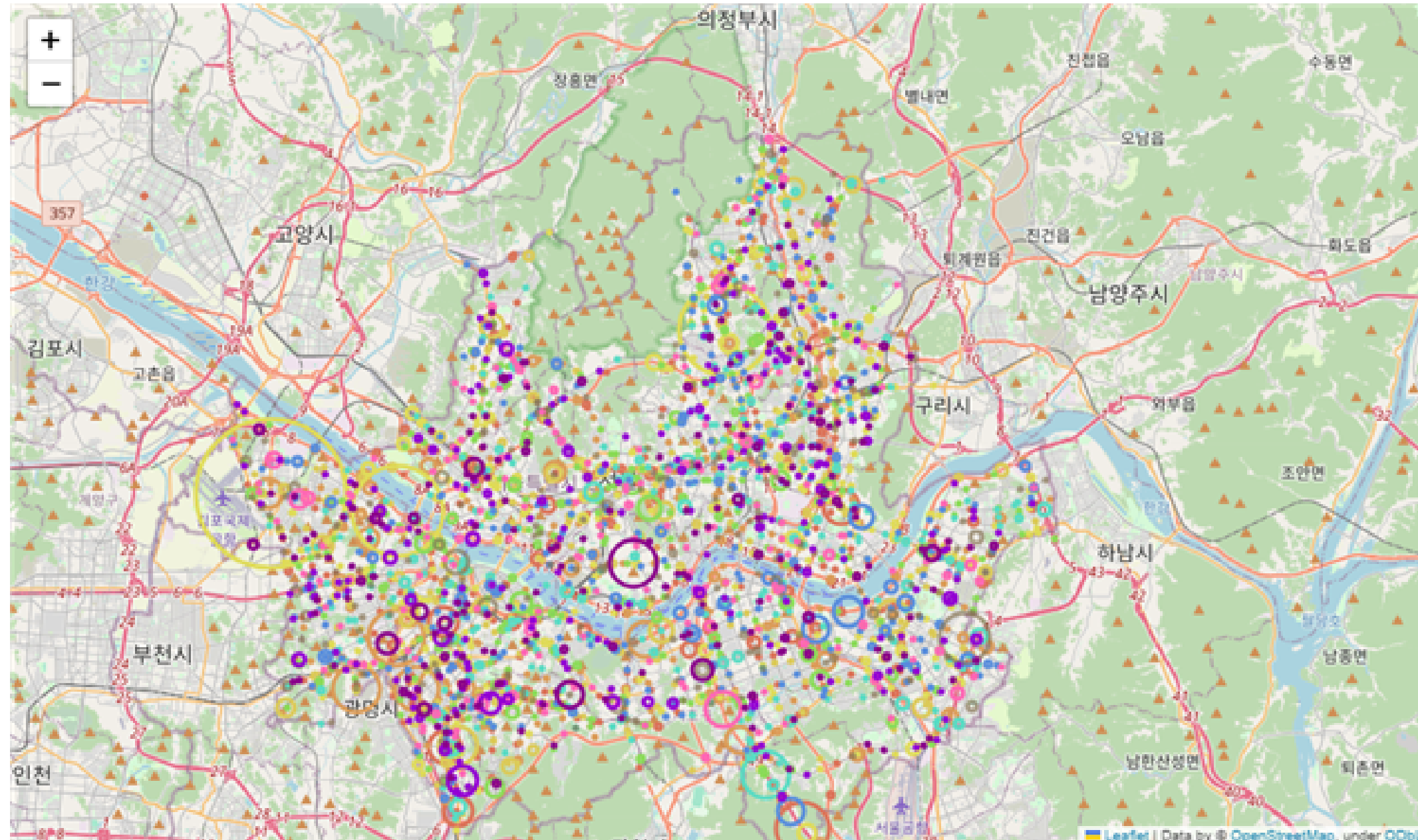


- 낮은 온도와 높은 온도에서 교통사고가 각각 많이 발생함을 확인할 수 있다. 도로교통공단에 따르면 계절별 발생한 교통사고 현황을 분석한 결과, 사고의 심각성을 말해주는 치사율(사고 100건당 사망자 수)의 경우에는 블랙아이스로 인해 겨울철이 다른 계절에 비해 높은 것으로 나타났다.



# 4. 시각화/ 분석

## 지도 라이브러리를 활용한 시각화



주로 대로와 고속도로에서 사상자가 많이 발생하는 것을 확인 할 수 있으며, 특히 경부고속도로에서 사상자가 많이 발생했다.

# 5. 결론

---

- 결과적으로 교통사고의 수와 심각도는 전반적으로 감소하는 추세를 보이고 있으며, 이는 정부의交通安全 정책과 법규 개정 등이 교통사고 감축에 기여한 것으로 판단
- 사고유형은 도로형태에 따라 심각도가 달라짐
- 날씨와 계절, 시간대와 지역 등의 환경적 요인도 교통사고의 발생과 심각도에 영향을 미침
- 과속, 교차로 통행방법 위반, 안전운전 의무 불이행 등의 법규위반에 대한 교육과 단속 강화 필요
- 통량이 많은 시간대와 계절, 그리고 특정 도로에서의 사고 예방에 초점을 맞추는 것이 중요
- 보다 정확하고 신뢰할 수 있는 분석을 위해서는 더 많은 데이터와 다양한 분석이 필요하다고 느낌
- 교통사고를 줄이기 위한 노력은 우리나라 국민 모두의 협력이 필요

# 감사합니다!

