

2023년 지역 치안 안전 데이터 분석 공모전

# 이륜차 교통사고 위험도 분석 및 대전광역시 위험지역 선정

폴보아즈(Pol-Boaz)

신재욱 (jwshin0908@naver.com)

김혜연 (sally879@naver.com)

박규연 (20203065@kookmin.ac.kr)

최유진 (youjin0100450@sookmyung.ac.kr)



# INDEX

1

## 분석 개요

- 주제 소개
- 분석 배경 및 목적
- 분석 순서도

2

## 데이터 수집 및 전처리

- 데이터 수집
- 데이터 전처리

3

## 위험도 분석

- 모델링 & 튜닝
- 최적 모델 선정
- 변수 중요도 해석

4

## 이륜차 사고 위험지역

- 대전 위험지역 선정
- 위험지역 산출 공식
- 상위 위험지역 선정
- 대전 위험지역 분석

5

## 결론

- 기대효과
- 의의 및 한계

6

## 툴· 데이터 및 참고문헌



# Part 1

## 분석 개요



## 이륜차 교통사고 위험도 분석 및 위험지역 선정

1



날짜, 시간대, 날씨 등  
다양한 변수에 따른  
사고 위험도 모델링

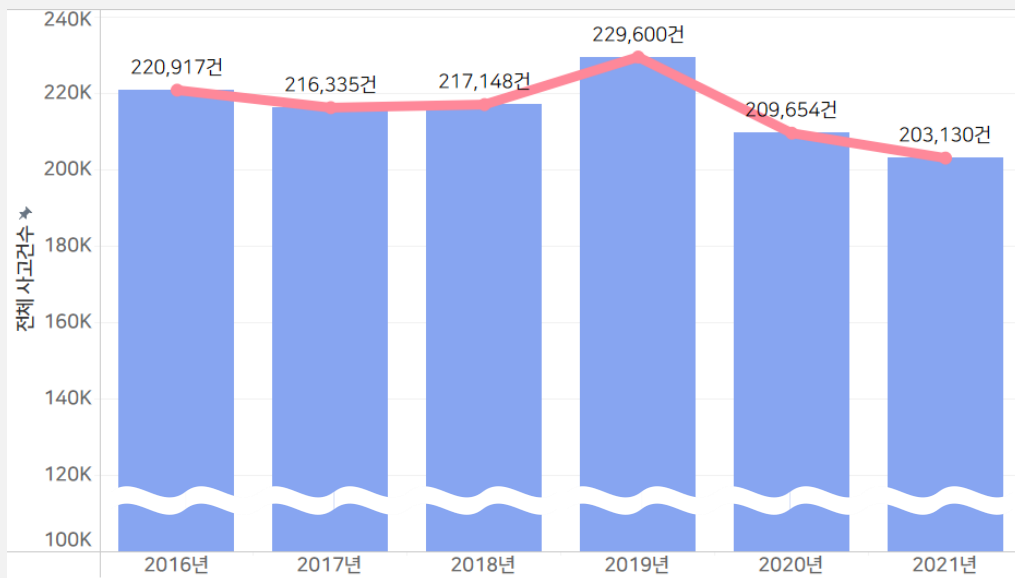
2



이륜차 사고 위험지역 선정 후  
위험지역별 개선사항 제언

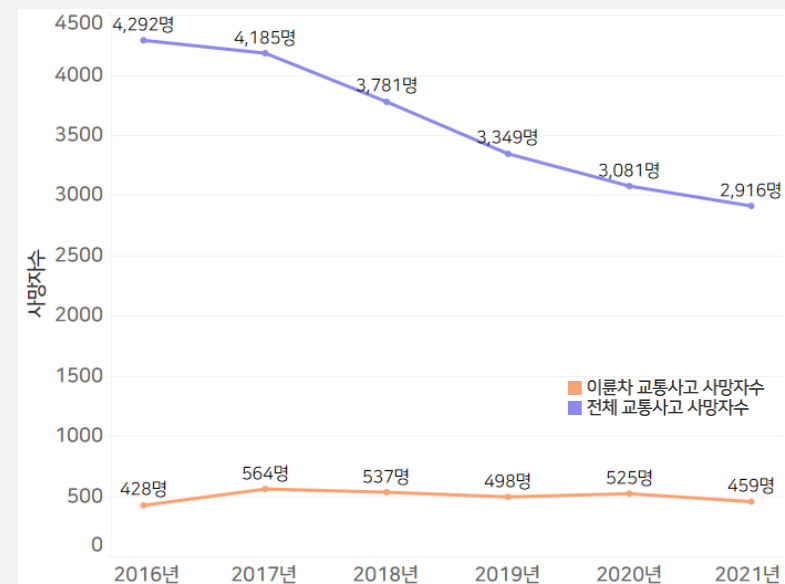
## 연도별 교통사고

### 교통사고 건수는 감소 추세



교통사고 건수는 2019년에 급증했지만 전반적으로 감소하여  
2021년에는 약 20만 건에 이르렀음

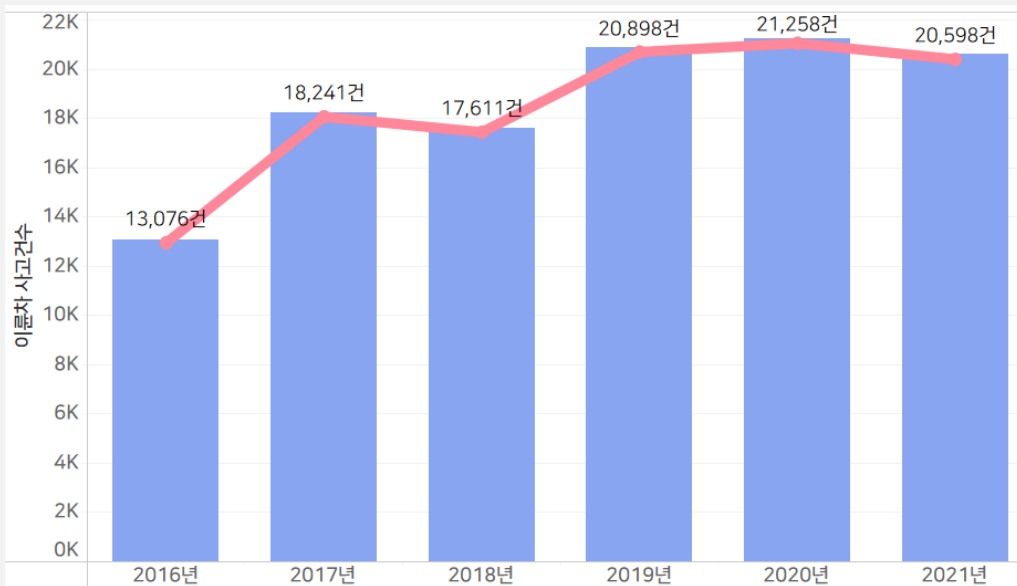
### 교통사고 전체 사망자수는 감소 추세



2016~2021년 교통사고 전체 사망자수는 감소하지만,  
이륜차 사망자수는 뚜렷한 변동을 보이지 않았음

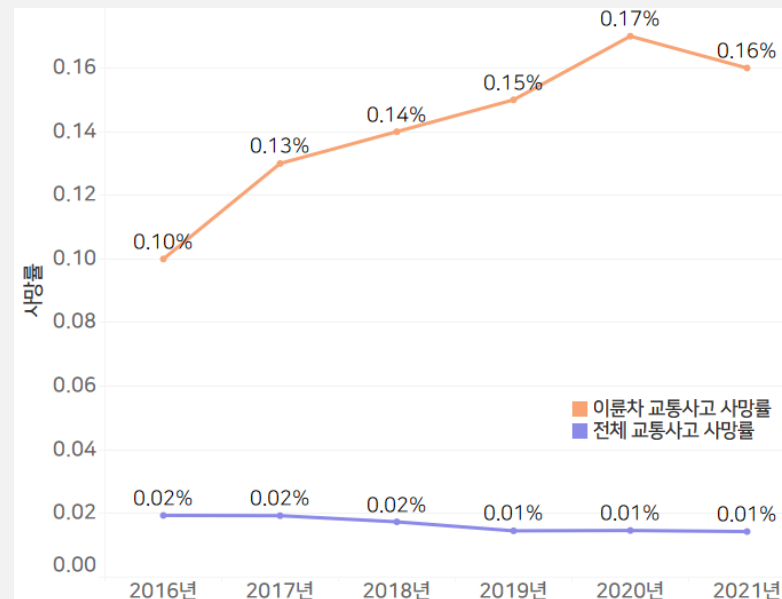
## 이륜차 교통사고 위험성

### 이륜차 교통사고는 증가 추세



교통사고 건수는 전반적으로 증가 추세를 보이며  
2021년의 경우 5년 전에 비해 사고 건수 50% 이상 상승

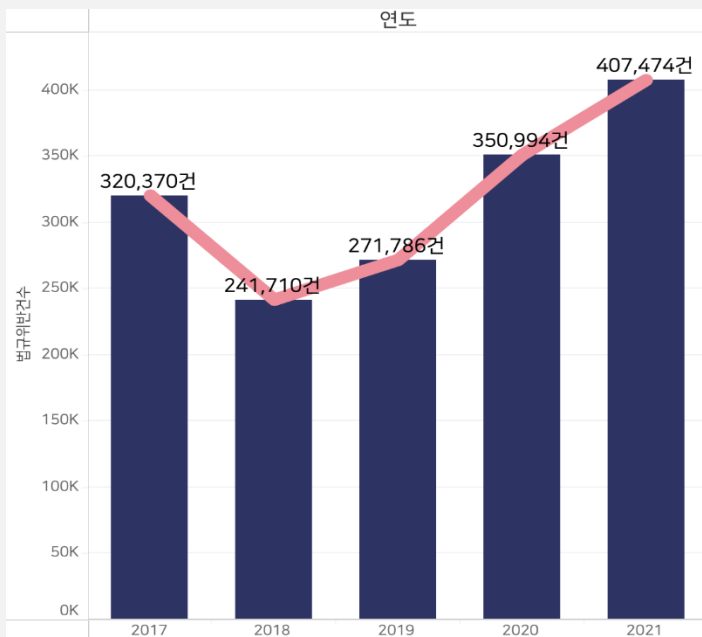
### 이륜차 교통사고 사망률은 증가 추세



2016~2021년 교통사고 전체 사망률은 감소하지만,  
이륜차 교통사고 사망률은 점차 증가하는 추세

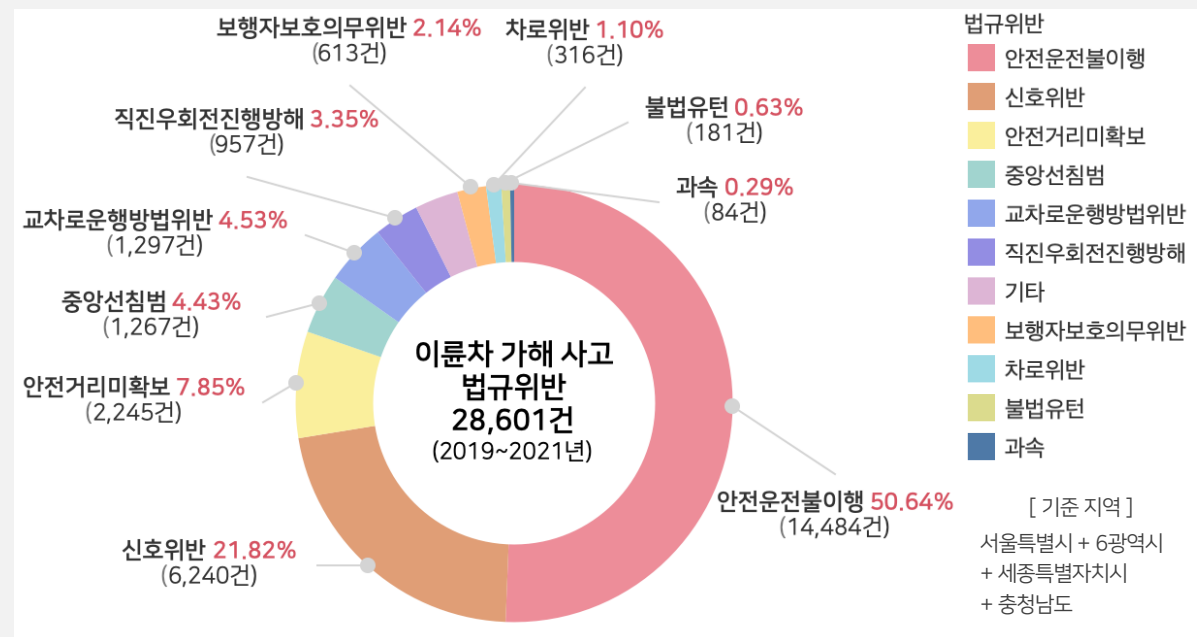
## 이륜차 법규위반

### 이륜차 법규위반은 증가 추세



이륜차 법규위반 건수는 증가 추세를 보이며  
2018년 24만 건, 2021년에는 40만 건에 이르렀음

### 사고로 이어지는 법규위반



2019~2021년 이륜차 가해사고로 이어진 법규위반은 28,601건  
그중 안전운전불이행이 가장 많았고, 신호 위반이 그 뒤를 이음

# 분석 배경 및 목적 - (3) 이륜차 안전시설 문제

## 이륜차 법규위반

이륜차 법규위반은 증가 추세

하지만,

사고로 이어지는 법규위반

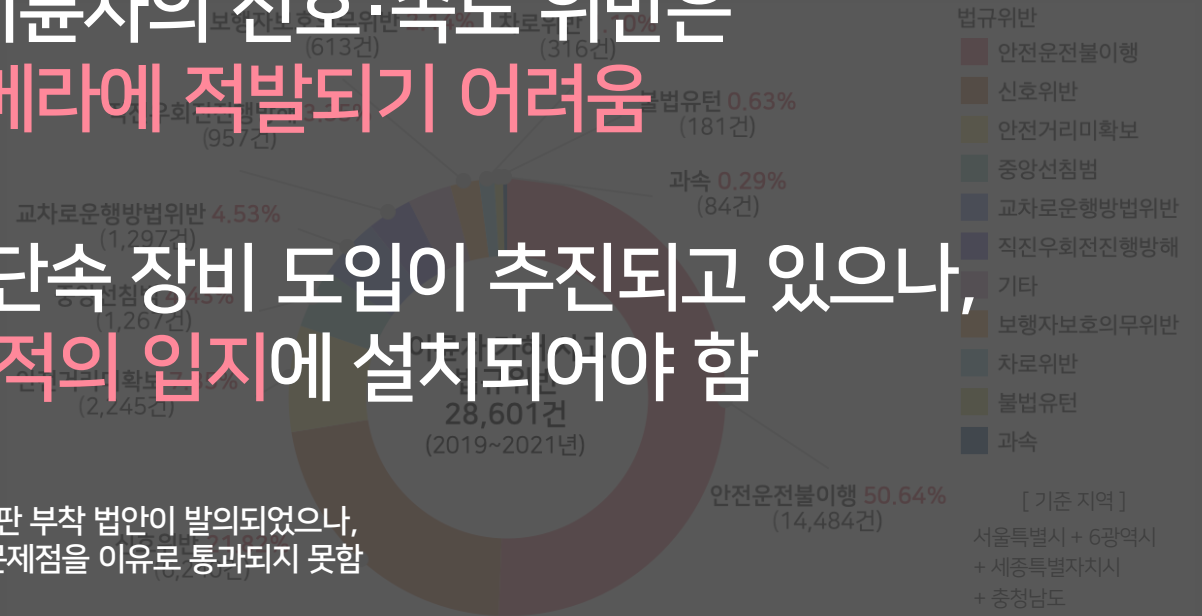
후면 번호판만 있는 이륜차의 신호·속도 위반은  
일반 무인 단속 카메라에 적발되기 어려움

영상데이터 기반 이륜차 무인 단속 장비 도입이 추진되고 있으나,  
비용이 비싸 필요한 **최적의 입지**에 설치되어야 함

이륜차의 전면 번호판 부착 법안이 발의되었으나,  
비용, 현실성 등의 문제점을 이유로 통과되지 못함



이륜차 법규위반 건수는 증가 추세를 보이며  
2018년 24만 건, 2021년에는 40만 건에 이르렀다.



2019~2021년 이륜차 가해사고로 이어진 법규위반은 28,601건  
그중 안전운전불이행이 가장 많았고, 신호 위반이 그 뒤를 이었다.

[기준 지역]  
서울특별시 + 6광역시  
+ 세종특별자치시  
+ 충청남도



## 노면상태



### 노면상태에 민감한 이륜차 교통사고

- 자동차보다 노면 접지 면적이 적은 이륜차는 노면상태에 따라 쉽게 위험에 노출됨
- 빙판길에서는 제동거리 20% 증가  
→ 이륜차 조향 능력과 균형을 잃기가 쉬움
- 특히 블랙아이스는 이륜차에 취약  
(블랙아이스: 도로 위에 얇은 얼음 막이 생기는 현상)

**주기적인 노면상태 점검이 필수적임**

포장 유무 & 기상 상태 (건조, 습기, 서리/결빙, 적설)

## 1 사고 위험도 분석

“어떤 상황에서 위험한  
이륜차 사고가 일어날까?”

심각한 인명 피해를 예방하는 것이 중요  
→ 이륜차 교통사고 위험도 분석

사고 피해 계수인 EPDO에 대한 회귀 모델을 통해  
유의해야 할 환경을 알아보고자 함

사고 상황/요인 데이터 증량 (지역 범위 확장)  
→ 대전, 충남, 세종 + 서울 + 6 광역시

## 2 위험지역 선정

“위험한 이륜차 사고가  
발생할 것 같은 곳은?”

근본적으로는 이륜차 사고 발생을 줄이기가 목표  
→ 이륜차 교통사고 위험지역 선정

예측 위험도 > 실제 위험도인 경우,  
위험한 이륜차 사고가 발생할 가능성 ↑

면적에 비해 사고 데이터가 많은  
대전광역시 내 그리드에 한해 위험지역 선정

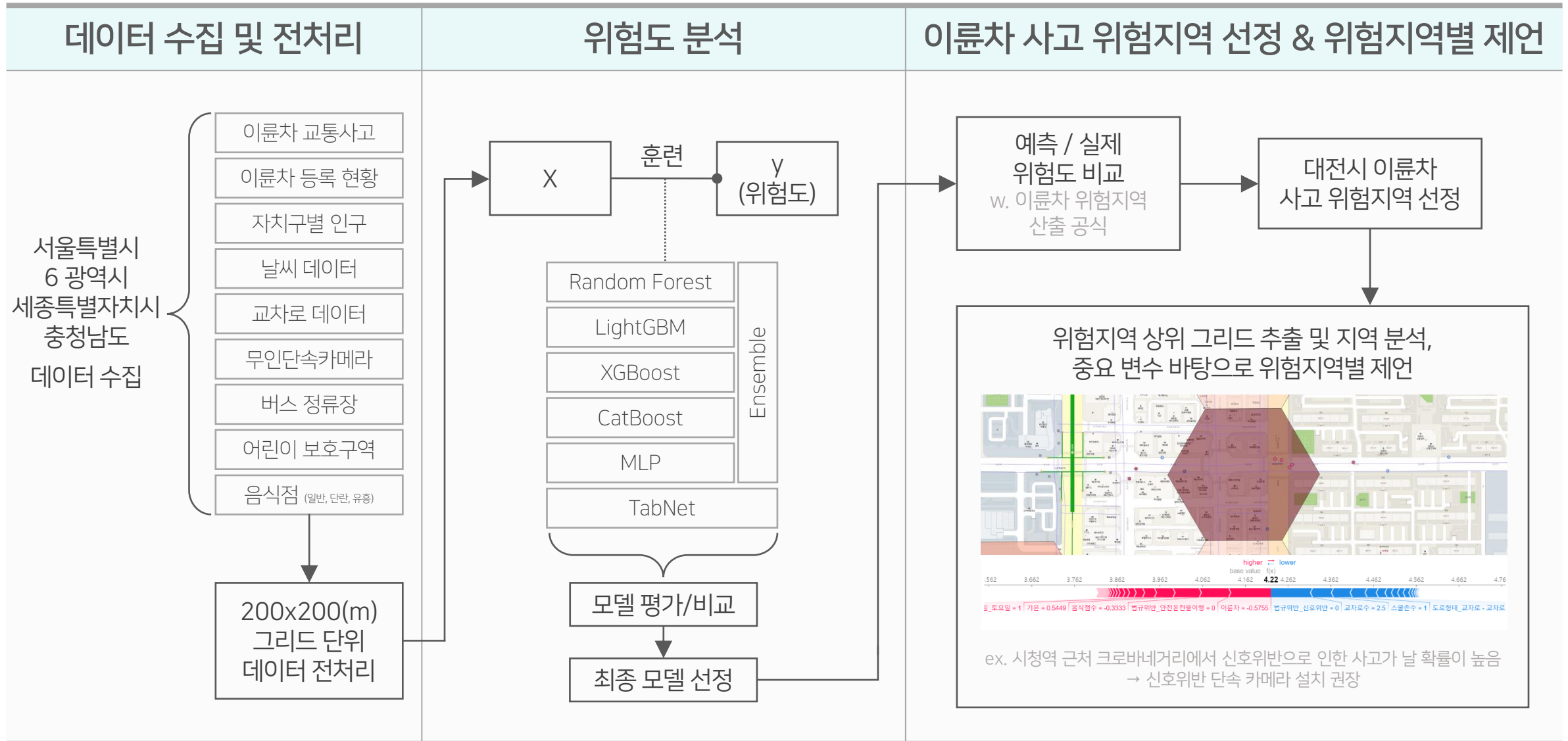
## 3 위험지역별 제언

“이륜차 사고가 발생하지 않도록  
위험지역을 개선하자!”

상위 위험지역 그리드에 대해  
사고 위험도에 중요한 영향을 미치는 요인들 분석

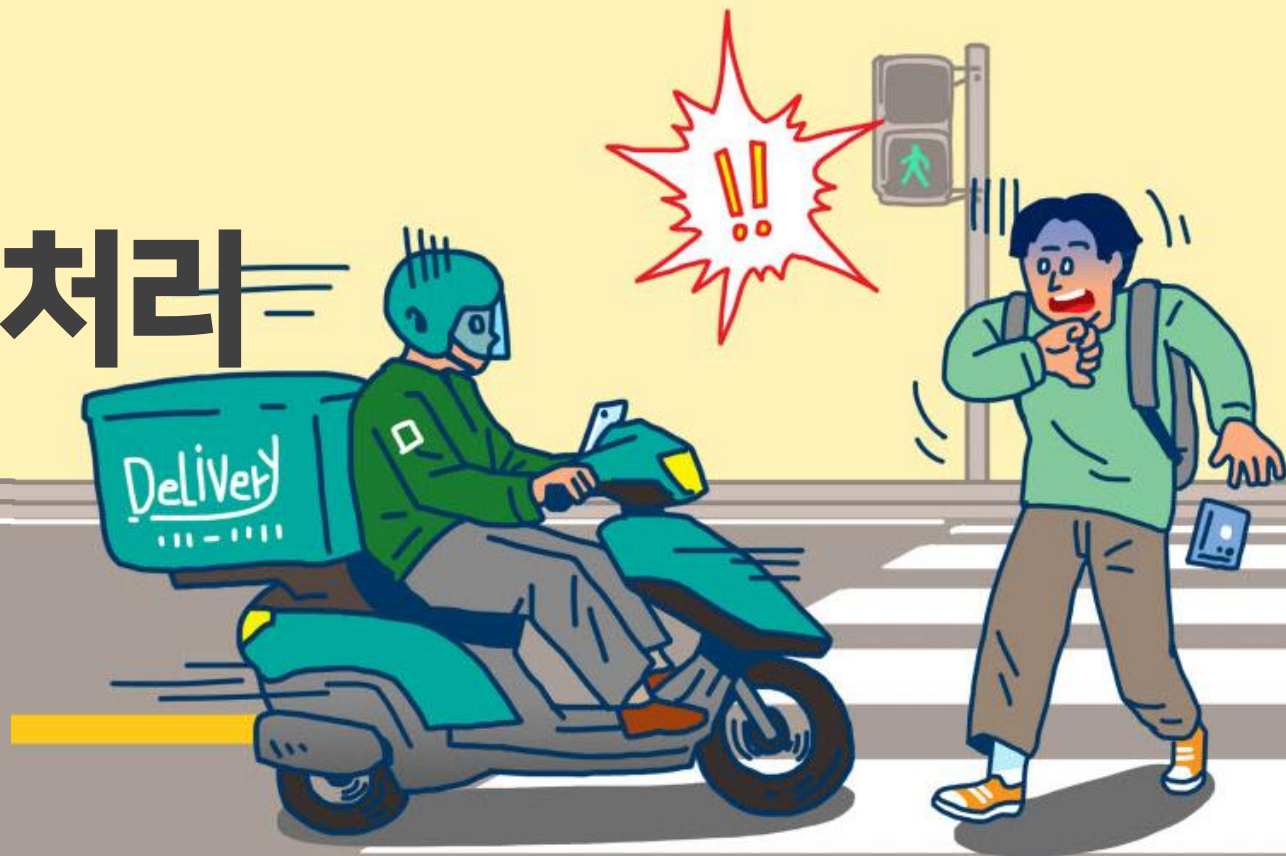
해당 그리드가 어떤 요인 때문에  
위험지역으로 선정되었는지 파악

위험지역에서 사고가 발생하지 않도록  
안전시설 개선 등 제언



Part 2

# 데이터 수집 및 전처리



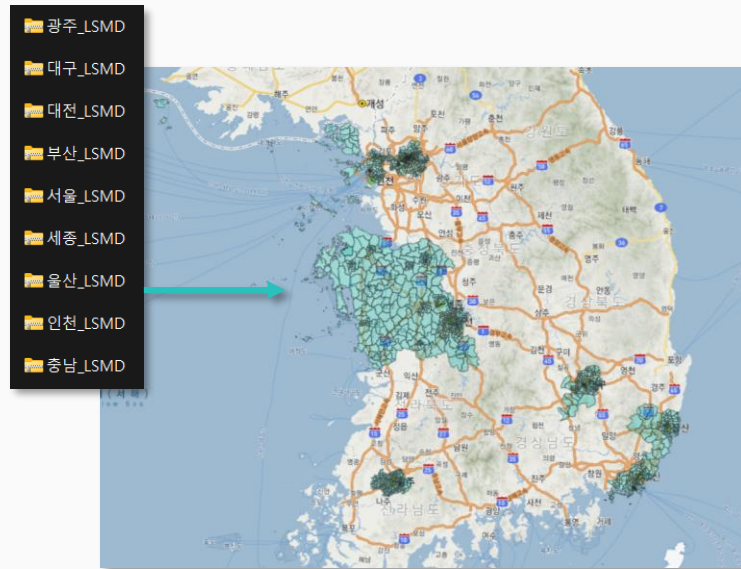
 <b>스마트 치안 빅데이터 플랫폼</b> Smart Policing Big Data Platform	 <b>TAAS</b> 교통사고분석시스템 Traffic Accident Analysis System			
제공 데이터	이륜차 교통사고	이륜차 신고 현황	행정구역별 인구	날씨(기온, 강수량, 풍속)
				
교차로	무인단속카메라	버스 정류장	어린이 보호구역	음식점(일반, 단란, 유흥)



# 데이터 전처리 - (1) 그리드 데이터 생성(QGIS)

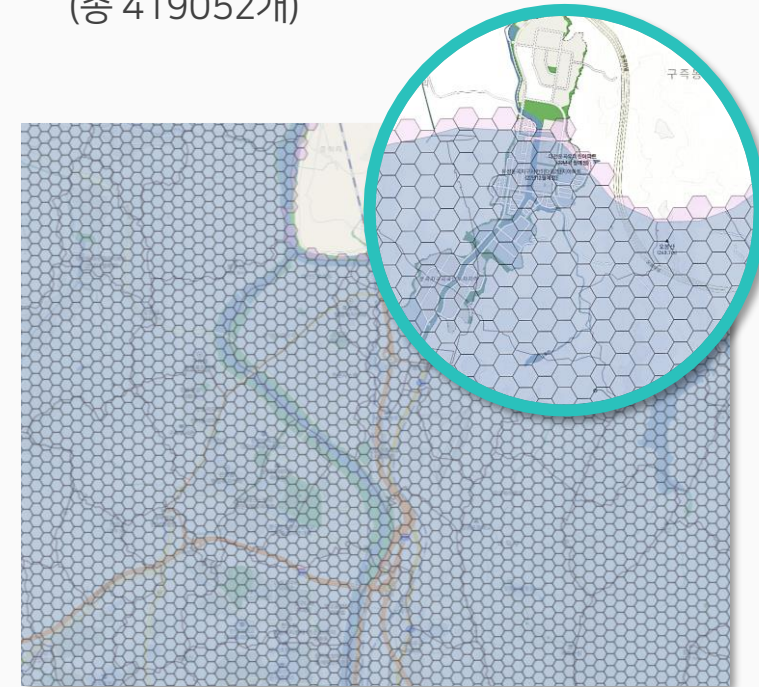
## 그리드 생성을 위한 지도 생성

서울특별시 +  
 6 광역시 +  
 세종특별자치시 +  
 충청남도를 결합한 **행정구역 지도** 생성



## 데이터 병합을 위한 그리드 생성

행정구역 지도와 겹치는 모든 영역에  
**200 x 200(m) hexagon 그리드** 생성  
 (총 419052개)



# 데이터 전처리 - (2) TAAS 사고 데이터

지역: 서울특별시, 6 광역시, 세종특별자치시, 충청남도

활용 데이터: 가해차종 이론차 or 피해차종 이론차인 데이터 → **이론차가 사고에 포함된 데이터** 활용

## 사고일시: 시간대로 변경

%Y년 %m월 %d일 %H시에서  
시간대 데이터로 변경

2019년 1월 1일 00시



23, 0, 1, 2, 3, 4	→ 새벽
5, 6, 7, 8, 9, 10	→ 출근
11, 12, 13	→ 점심
14, 15, 16	→ 오후
17, 18, 19	→ 저녁
20, 21, 22	→ 퇴근

## 사고유형, 노면상태: 세부 범주 통합

사고유형: 세부 범주를 통합하여 3가지 범주로 변경  
노면상태: 세부 범주를 통합하여 7가지 범주로 변경

### [사고유형]

차대차 - 측면충돌	→	<b>차대차</b>
차대차 - 기타		
⋮		
차량단독 - 주/정차차량 충돌		
철길건널목 - 철길건널목		<b>차량단독</b>

### [노면상태]

포장 - 건조	→	<b>건조</b>
포장 - 젖음/습기		
⋮		
포장 - 해빙		
비포장 - 적설		<b>침수</b>
		<b>해빙</b>

## 가해운전자: 연령대로 변경

구체적인 가해운전자 연령을  
생애주기별 연령대 데이터로 변경

13~18세	→	<b>청소년</b>
19~29세	→	<b>청년</b>
30~49세	→	<b>중년</b>
50~64세	→	<b>장년</b>
65세 이상	→	<b>노년</b>

## EPDO 추가

### EPDO(대물피해환산계수)

= 사망자수 x 12 +  
중상자수 x 5 +  
경상자수 x 3 +  
부상신고자수

# 데이터 전처리 - (2) TAAS 사고 데이터

## 제외한 컬럼

- 사고내용(경상사고/중상사고/부상신고사고/사망사고), 사망자수, 중상자수, 경상자수, 부상신고자수: 위험도를 예측하는 모델의 x값으로 불필요
- 기상상태: 기온, 강수량, 풍속과 같은 연속형 변수만 유지
- 가해운전자 차종, 가해운전자 성별, 가해운전자 상해정도
- 피해운전자 차종, 피해운전자 성별, 피해운전자 연령, 피해운전자 상해정도

## 최종적으로 사용한 컬럼 총 15개

	사고번호	사고일시	요일	시군구	사고유형	법규위반	노면상태	도로형태	가해운전자	연령	x	y	EPD0
0	2019010100100002	201901010000	화요일	서울특별시 구로구 고척동	차대차	안전운전불이행	건조	단일로 - 기타	청년	126.867286	37.499889	3	
1	2019010100100141	201901011100	화요일	서울특별시 서초구 서초동	차대차	안전거리미확보	건조	단일로 - 기타	중년	127.017439	37.481949	3	
2	2019010100100170	201901011300	화요일	서울특별시 서대문구 북아현동	차대사람	보행자보호의무위반	건조	단일로 - 기타	청년	126.956285	37.558028	3	
3	2019010100100236	201901011700	화요일	서울특별시 중랑구 상봉동	차대차	신호위반	건조	교차로 - 교차로안	청년	127.085745	37.596356	3	
4	2019010100100249	201901011800	화요일	서울특별시 구로구 천왕동	차대차	안전운전불이행	젖음/습기	교차로 - 교차로안	청년	126.842479	37.484115	3	

⋮



# 데이터 전처리 - (3) 이륜차, 인구 데이터

지역: 서울특별시, 6 광역시, 세종특별자치시, 충청남도의 시도, 시군구

활용 데이터: 이륜차 신고 현황, 인구 데이터

## • 이륜차 데이터 (시/군, 일시별 데이터)

	시도명	시군구	2019년 1월	2019년 2월	2019년 3월	2019년 4월
0	서울특별시	강남구	16456	16439	16463	164
1	서울특별시	강동구	16122	16122	16129	161
2	서울특별시	강북구	21689	21670	21635	216

## • 인구 데이터 (시/군, 일시별 데이터)

	시도명	시군구	2019년 1월	2019년 2월	2019년 3월	2019년 4월
0	서울특별시	종로구	152866	152880	152778	1524
1	서울특별시	중구	125872	125995	125942	1259
2	서울특별시	용산구	229167	229279	229168	2289

## • TAAS 사고 데이터

	사고번호	사고일시	요일	시군구	사고유
0	2019010100100002	201901010000	화요일	서울특별시 구로구 고척동	차대
1	2019010100100141	201901011100	화요일	서울특별시 서초구 서초동	차대
2	2019010100100170	201901011300	화요일	서울특별시 서대문구 북아현동	차대
3	2019010100100236	201901011700	화요일	서울특별시 중랑구 상봉동	차대

이륜차, 인구 데이터의 시도/시군구와 일시(년/월)를 활용해 행정구역과 일시가 같은 사고번호에 이륜차, 인구 데이터 추가

	사고번호	이륜차	인구
0	2019010100100159	16847	361565
1	2019010100100237	16847	361565
2	2019010100100305	16847	361565
3	2019010100100321	16847	361565
4	2019010200100206	16847	361565

# 데이터 전처리 - (4) 날씨 데이터

지역: 서울특별시, 6 광역시, 세종특별자치시, 충청남도의 시/군

활용 데이터: 기온, 강수량, 풍속 데이터

- 날씨 데이터(시/군, 일시별 데이터)

지점	지점 이름	일시	기온(°C)	강수량(mm)	풍속(m/s)	지점 일시
0	156 광주광역시	2021-01-01 00:00	-1.8	0.0	1.3	202101010000
1	156 광주광역시	2021-01-01 01:00	-2.9	NaN	1.1	202101010100
2	156 광주광역시	2021-01-01 02:00	-2.9	NaN	1.0	202101010200
3	156 광주광역시	2021-01-01 03:00	-2.8	0.0	0.1	202101010300
4	156 광주광역시	2021-01-01 04:00	-2.9	NaN	0.7	202101010400

- TAAS 사고 데이터

	사고번호	사고일시	요일	시군구	사고유형
0	2019010100100002	201901010000	화요일	서울특별시 구로구 고척동	차대차
1	2019010100100141	201901011100	화요일	서울특별시 서초구 서초동	차대차
2	2019010100100170	201901011300	화요일	서울특별시 서대문구 북아현동	차대사람 보행자
3	2019010100100236	201901011700	화요일	서울특별시 중랑구 상봉동	차대차
4	2019010100100249	201901011800	화요일	서울특별시 구로구 천왕동	차대차

날씨 데이터의 도시와 일시(년/월/일/시간)를 활용해

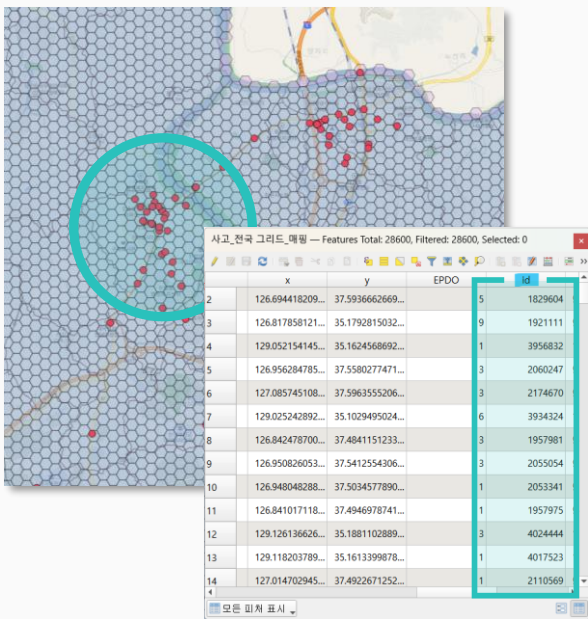
도시와 일시가 같은 사고번호에 기온, 강수량, 풍속 데이터 추가

	사고번호	기온	강수량	풍속
0	2019010100100002	-5.5	0.0	1.0
1	2019010100100141	-3.7	0.0	2.5
2	2019010100100170	-1.5	0.0	2.7
3	2019010100100236	-1.8	0.0	2.2
4	2019010100100249	-2.7	0.0	2.6

# 데이터 전처리 - (5) 교차로, 단속카메라, 버스정류장, 어린이 보호구역, 음식점 데이터

## 사고 위치와 그리드 연결

사고 위치 표시 및  
사고가 위치한 **그리드 id**와  
**사고번호** 연결

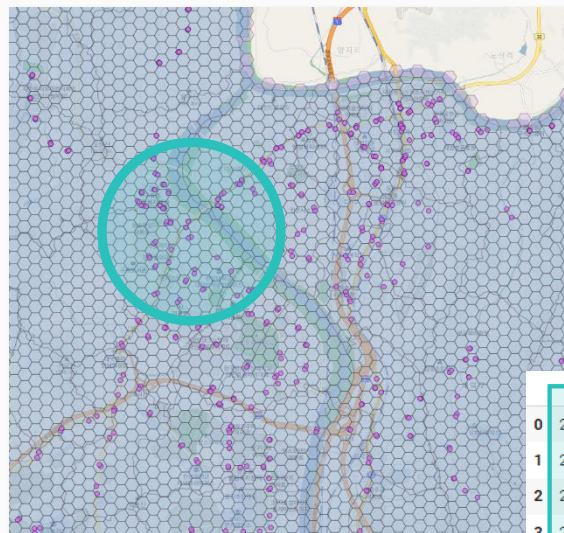


모든 데이터 2019-2021로 date 필터링하여 사용

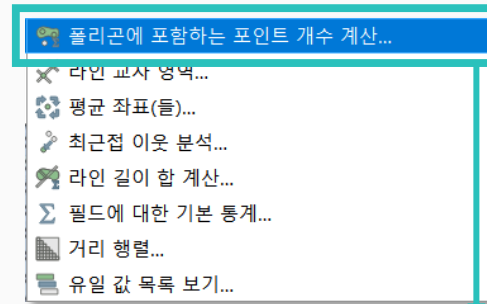
## 그리드 내 데이터의 개수 집계

각 데이터의 위치 표시 및  
사고가 위치한 **그리드 내 각 데이터의 개수** 집계, 각 **사고번호의 그리드 id**와 연결

ex) ① **버스정류장 위치** 표시



② 사고가 위치한 **그리드의 버스정류장 개수** 집계,  
그리드 id를 기준으로 **사고번호와 버스정류장 개수** 연결



	사고번호	grid_id	교차로수	단속카메라수	버스정류장수	스쿨존수	음식점수	유유주점수
0	2019010100100002	1980515.0	2	0.0	4.0	0.0	23.0	0.0
1	2019010100100141	2114043.0	1	0.0	6.0	0.0	0.0	0.0
2	2019010100100170	2060247.0	2	1.0	6.0	0.0	3.0	0.0
3	2019010100100236	2174670.0	1	0.0	0.0	0.0	59.0	9.0
4	2019010100100249	1957981.0	1	1.0	4.0	1.0	0.0	0.0

# 데이터 전처리 - (6) 최종 데이터셋

행 기준: 사건(사고번호)

## 제외한 컬럼

사고번호, 시군구, x, y, 연월, 위치, grid\_id: 위험도 예측 모델을 돌리는 데 불필요

TAAS 사고 데이터

위치 및 일시로  
연결한 데이터

그리드 id로  
연결한 데이터

사고일시	요일	사고유형	법규위반	노면상태	도로형태	가해운전자	차종	가해운전자	연령	EPDO	이륜차	인구	기온	강수량	풍속	단속카메라수	스쿨존수	음식점수	유흥주점수	단란주점수	교차로수	버스정류장수
0	새벽	화요일	차대차	안전운전불이행	건조	단일로 - 기타	이륜	청년	3	15623	404049.0	-5.5	0.0	1.0	0.0	0.0	23.0	0.0	0.0	0.0	2.0	4.0
1	점심	화요일	차대차	안전거리미확보	건조	단일로 - 기타	승용	중년	3	15623	404049.0	-3.7	0.0	2.5	0.0	0.0	0.0	0.0	0.0	0.0	1.0	6.0
2	점심	화요일	차대사람	보행자보호의무위반	건조	단일로 - 기타	이륜	청년	3	15623	404049.0	-1.5	0.0	2.7	1.0	0.0	3.0	0.0	0.0	0.0	2.0	6.0
3	저녁	화요일	차대차	신호위반	건조	교차로 - 교차로안	이륜	청년	3	15623	404049.0	-1.8	0.0	2.2	0.0	0.0	59.0	9.0	3.0	1.0	0.0	0.0
4	저녁	화요일	차대차	안전운전불이행	젖음/습기	교차로 - 교차로안	이륜	청년	3	15623	404049.0	-2.7	0.0	2.6	1.0	1.0	0.0	0.0	0.0	0.0	1.0	4.0

위험도(예측할 값)

## Part 3

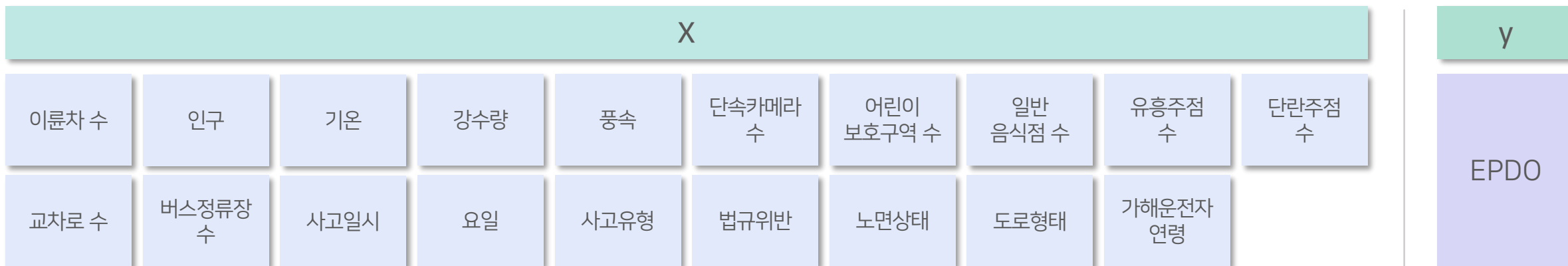
# 위험도 분석



# 모델링 & 튜닝 - (1) X, y 변수

EPDO를 예측하는 회귀모델 학습을 위한 X, y변수 설정

- X : 19개 변수
- Y : EPDO



$$\text{EPDO (대물피해환산계수)} = \text{사망자수} \times 12 + \text{중상자수} \times 5 + \text{경상자수} \times 3 + \text{부상자수}$$

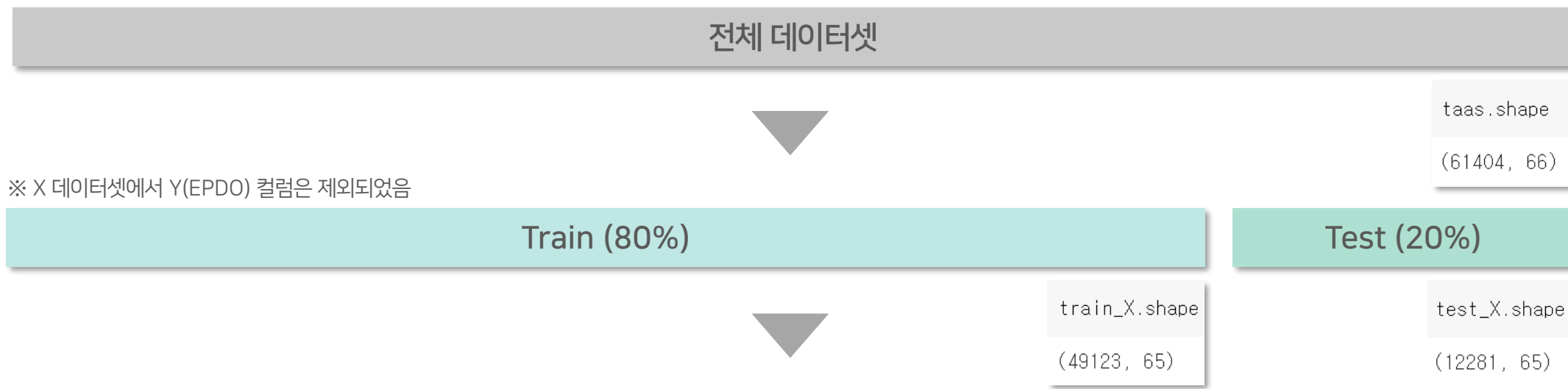
\* EPDO : 교통사고 피해 정도에 가중치를 적용하여 교통사고 잦은 지점을 선정하는 방법

8개 범주형 컬럼 → 50개 컬럼

사 고 일 시	요 일	사 고 유 형	법 규 위 반	노 면 상 태	도 로 형 태	가 해 운 전 자	차 종	가 해 운 전 자 연 령
새벽	화요일	차대차	안전운전불이행	건조	단일로 - 기타		이륜	청년
사 고 일 시 「새벽」	요 일 「화요일」	사 고 유 형 「차대차」	법 규 위 반 「안전운전불이행」	노 면 상 태 「건조」	도 로 형 태 「단일로 - 기타」	가 해 운 전 자 「이륜」	차 종 「이륜」	가 해 운 전 자 연 령 「청년」

# 모델링 & 튜닝 - (3) train/test 데이터 분리

하나의 데이터셋을 train/test 데이터셋으로 분리



이후 Grid Search 진행 시 Cross Validation (교차검증) 진행, CV=5





# 모델링 & 튜닝 - (4) feature selection

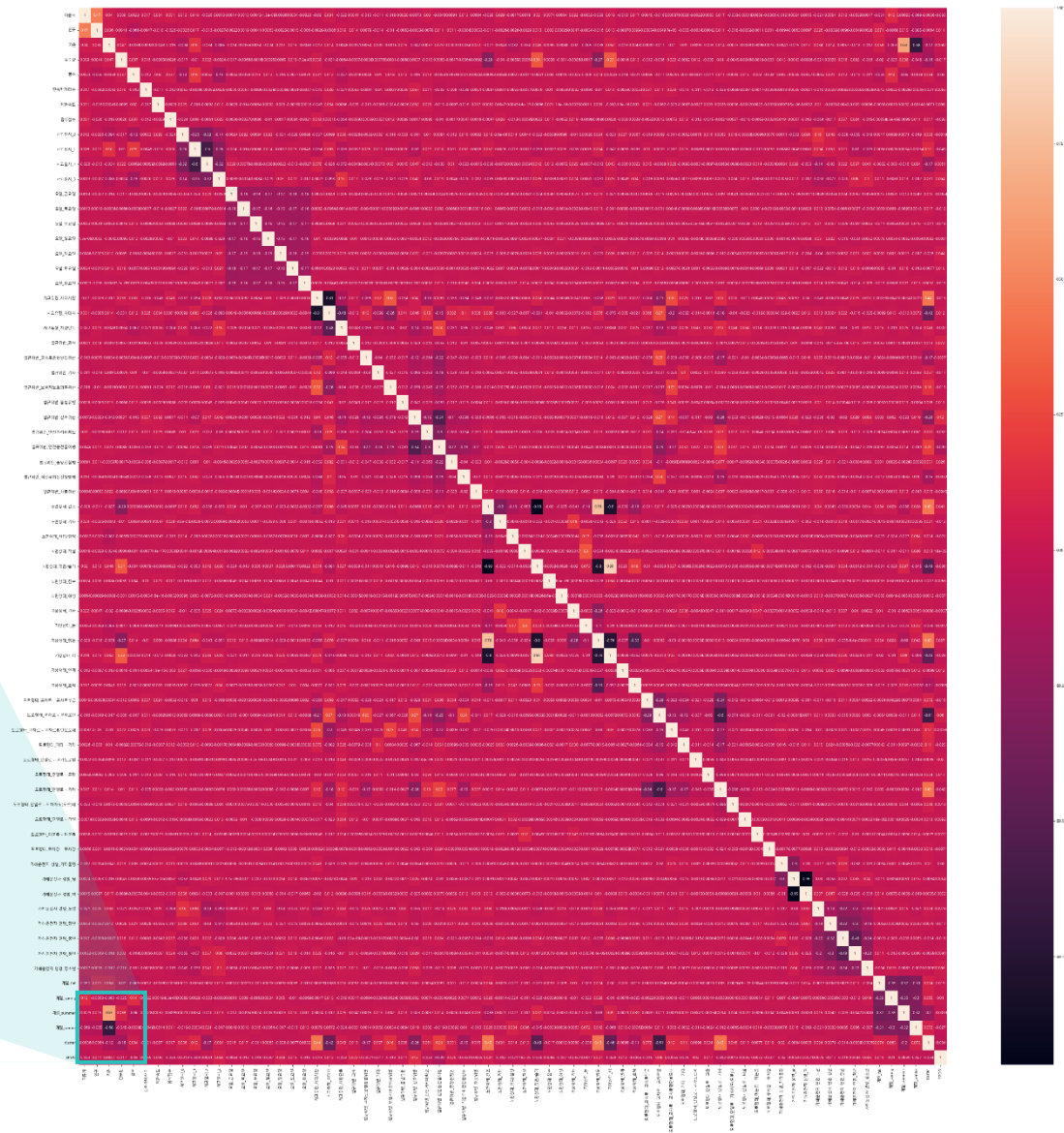
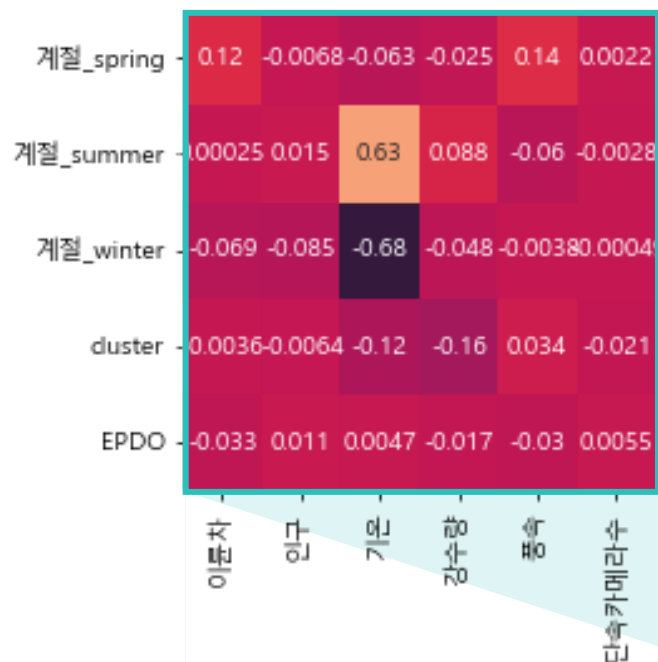
변수들 간의 상관성 분석 후 유의미한 feature 제외하고 drop

ex )

계절\_summer와 기온 비례( $r=0.63$ ) / 계절\_winter와 기온 반비례 ( $r=-0.68$ )

→ 계절과 기온, 두 변수가 모두 존재할 필요X

※ r은 상관계수



## 연속형 데이터 Robust Scaling

Train 데이터셋 scaling한 만큼 Test 데이터셋 scaling 진행

	이륜차	인구	기온	강수량	풍속	단속카메라수	스쿨존수	음식점수	유흥주점수	단란주점수	교차로수	버스정류장수
<b>57087</b>	14526.0	427684.0	4.3	0.0	0.7	2.0	0.0	11.0	0.0	0.0	6.0	0.0
<b>59182</b>	25328.0	155511.0	23.2	0.0	0.7	0.0	0.0	35.0	5.0	7.0	4.0	0.0
<b>29830</b>	6487.0	175799.0	27.0	0.0	1.7	0.0	0.0	32.0	0.0	1.0	4.0	0.0
<b>44166</b>	11559.0	424947.0	31.8	0.0	1.1	0.0	0.0	15.0	8.0	0.0	1.0	2.0
<b>27591</b>	8660.0	272532.0	22.4	0.0	2.8	0.0	0.0	12.0	5.0	2.0	0.0	0.0



## • Train 데이터셋

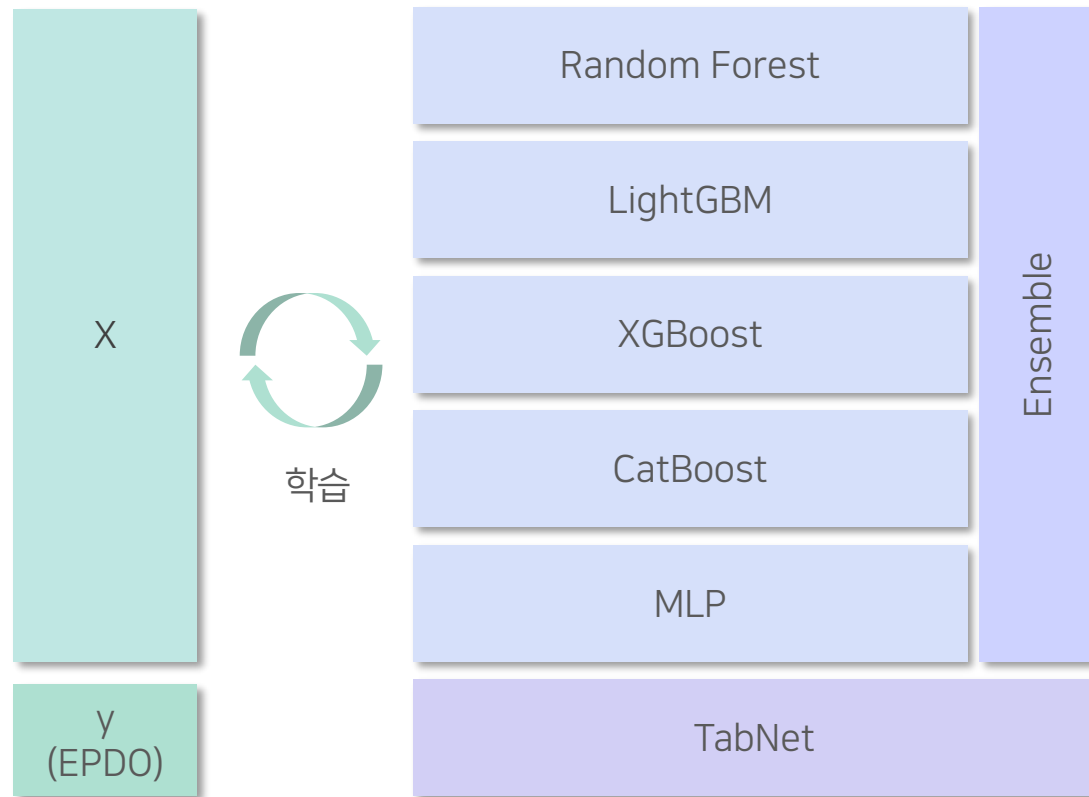
	이륜차	인구	기온	강수량	풍속	단속카메라수	스쿨존수	음식점수	유흥주점수	단란주점수	교차로수	버스정류장수
<b>0</b>	-0.034003	0.260688	-0.794872	0.0	-0.888889	2.0	0.0	0.190476	0.0	0.0	2.5	-0.333333
<b>1</b>	1.024498	-0.968689	0.416667	0.0	-0.888889	0.0	0.0	1.333333	5.0	7.0	1.5	-0.333333
<b>2</b>	-0.821754	-0.877050	0.660256	0.0	-0.333333	0.0	0.0	1.190476	0.0	1.0	1.5	-0.333333
<b>3</b>	-0.324743	0.248325	0.967949	0.0	-0.666667	0.0	0.0	0.380952	8.0	0.0	0.0	0.333333
<b>4</b>	-0.608819	-0.440117	0.365385	0.0	0.277778	0.0	0.0	0.238095	5.0	2.0	-0.5	-0.333333

## • Test 데이터셋

	이륜차	인구	기온	강수량	풍속	단속카메라수	스쿨존수	음식점수	유흥주점수	단란주점수	교차로수	버스정류장수
<b>0</b>	0.881823	4.279171	0.839744	0.0	0.166667	0.0	0.0	-0.333333	0.0	0.0	-0.5	-0.333333
<b>1</b>	-0.827927	-0.090799	-0.852564	0.0	-0.777778	0.0	0.0	5.333333	0.0	0.0	1.5	-0.333333
<b>2</b>	0.106124	0.166759	-0.967949	0.0	0.333333	0.0	0.0	-0.333333	0.0	0.0	-0.5	0.333333
<b>3</b>	-0.286428	-0.485652	-1.551282	0.0	1.111111	0.0	0.0	-0.238095	0.0	0.0	0.0	-0.333333
<b>4</b>	-0.192161	0.727496	-0.089744	0.0	0.833333	0.0	0.0	-0.333333	0.0	0.0	-0.5	-0.333333

EPDO 예측을 위한 회귀모델 선택 및 학습 진행

모델	설명
Random Forest	여러 개의 의사결정나무(Decision Tree)를 조합하여 안정적이고 정확한 예측을 수행하는 앙상블(Ensemble) 기법
LightGBM	일반적인 그래디언트 부스팅(Gradient Boosting) 방법을 사용하면서도 빠른 학습과 예측 속도, 과적합 방지 등의 성능 향상을 위한 최적화 기법을 적용한 부스팅 알고리즘
XGBoost	트리 분류기와 부스팅 알고리즘을 활용하여 높은 예측 성능을 가지며, 과적합 방지와 빠른 학습 및 예측 속도 등이 특징인 알고리즘
CatBoost	범주형 특성을 다루는데 있어서 탁월한 예측 성능을 가지며, 자체적으로 과적합을 방지하는 기능을 가진 부스팅 알고리즘
MLP	인공 신경망 기반의 다층 퍼셉트론으로, 복잡한 비선형 문제를 다룰 수 있는 대표적인 머신러닝 알고리즘
Ensemble	여러 개의 머신러닝 모델을 결합하여 예측을 수행하는 앙상블 기법 본 실험에서는 위 5개 머신러닝 모델들을 결합함
TabNet	특징 중요도를 고려하여 특징 선택과 유용한 정보 추출을 동시에 수행하는 DNN 기반의 정형 데이터 딥러닝 모델



Grid Search CV를 진행하여 최적의 하이퍼파라미터 선정

Random Forest	LightGBM	XGBoost
<pre># Define the parameter grid to search param_grid = {     'n_estimators': [50, 100],     'max_depth': [5, 10],     'min_samples_split': [5, 10],     'min_samples_leaf': [1, 2, 4] }</pre>	<pre># Define the hyperparameters to tune param_grid = {     'num_leaves': [5, 10],     'learning_rate': [0.001, 0.01, 0.1],     'n_estimators': [100, 500, 1000] }</pre>	<pre># Define the hyperparameters to be tuned param_grid = {     'n_estimators': [50, 100, 200],     'max_depth': [3, 5, 7],     'learning_rate': [0.001, 0.01, 0.1] }</pre>
CatBoost	MLP	TabNet
<pre># Define the hyperparameters to tune param_grid = {     'depth': [3, 5, 7, 10, 20],     'learning_rate': [0.001, 0.01, 0.1, 0.3, 0.5] }</pre>	<pre># Define the hyperparameters to be tuned param_grid = {     'hidden_layer_sizes': [(10,), (50,)],     'activation': ['logistic', 'tanh', 'relu'],     'solver': ['lbfgs', 'sgd', 'adam'],     'alpha': [0.001, 0.01, 0.1] }</pre>	<pre># Define the hyperparameters to be tuned param_grid = {     'n_steps': [4],     'n_independent': [1, 2, 3],     'n_shared': [1, 2, 3], }</pre>

## 모델 비교

모델 평가 지표	Random Forest	LightGBM	XGBoost	CatBoost	MLP Regressor	Ensemble (좌측 5개 모델 앙상블)	TabNet
RMSE	2.264	2.264	2.259	2.257	2.261	2.257	2.266
RMSLE	0.398	0.397	0.397	0.397	-	-	0.398

## 최적 모델 선정

CatBoost의 RMSE가 가장 낮으므로,  
이론차 사고 위험도 분석 모델로 CatBoost 선정

# 변수 중요도 해석 - SHAP Feature Importance

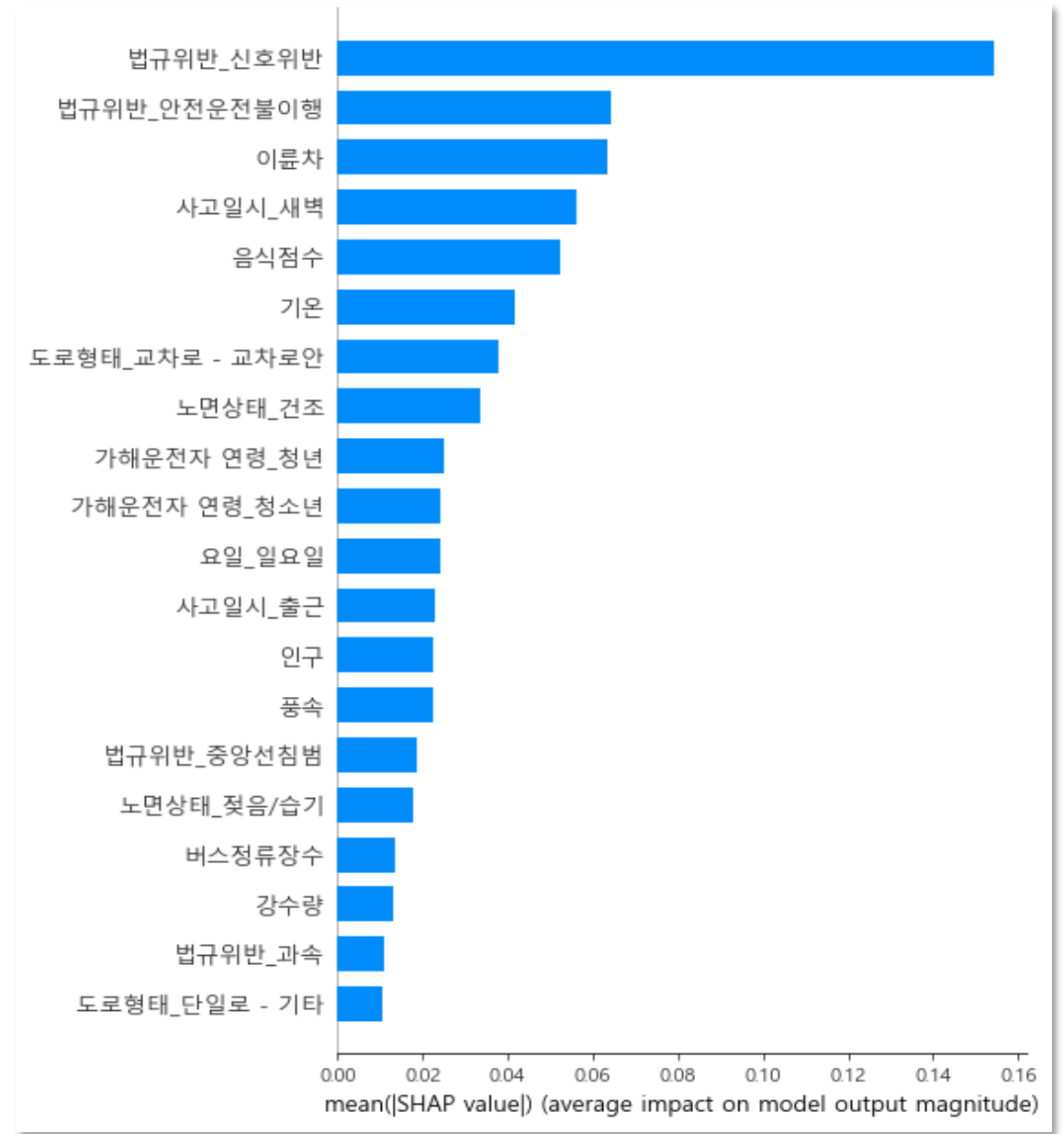
## SHAP 변수 중요도

Shap (SHapley Additive exPlanations)

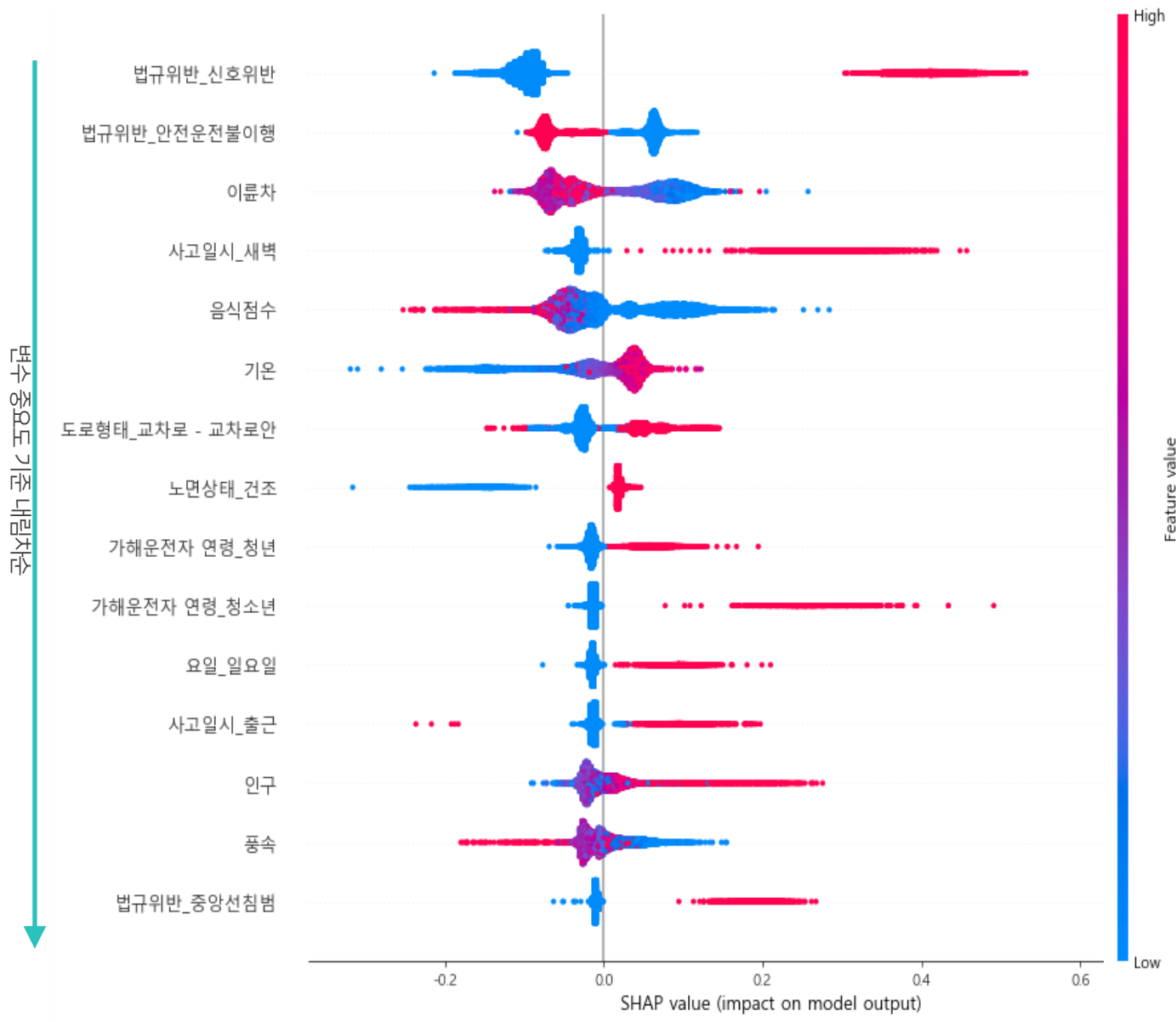
: ML 모델에서 각 변수를 하나씩 빼고 더하면서  
해당 변수가 y 예측의 정확도에 얼마나 기여하는지 계산하여  
각 변수의 중요도를 수치화하는 방법

## 중요도 Top 20 변수

	분류	변수
범주형	법규위반	신호위반, 안전운전불이행, 중앙선침범, 과속
	사고일시	새벽, 출근
	도로형태	교차로 안, 단일로
	노면상태	건조, 젖음/습기
	가해운전자 연령	청년, 청소년
	요일	일요일
	연속형	이륜차 등록 대수, 음식점 수, 기온, 인구, 풍속, 버스정류장 수, 강수량



# 변수 중요도 해석 - SHAP Summary Plot



## SHAP value 해석 방법

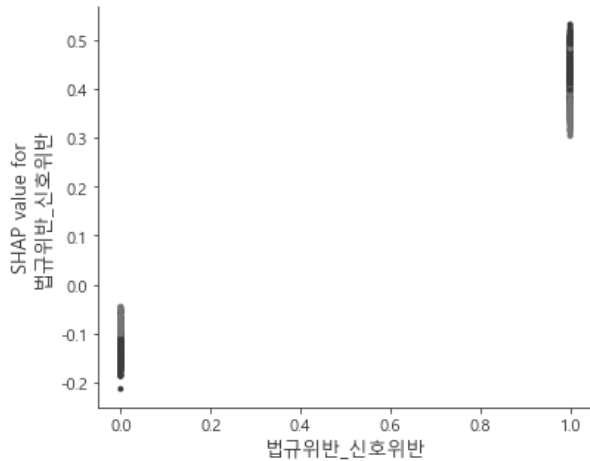
- 색상 : 각 변수의 기여 방향
  - 빨간색 : 변수 값이 높으면
  - 파란색 : 변수 값이 낮으면
- 영역 : y 변수에 대한 작용 방향
  - 양의 영역 : 예측 결과에 긍정적 요인
  - 음의 영역 : 예측 결과에 부정적 요인

※ 단, 인과관계가 아닌 기여 방향의 경향성으로 해석해야 함

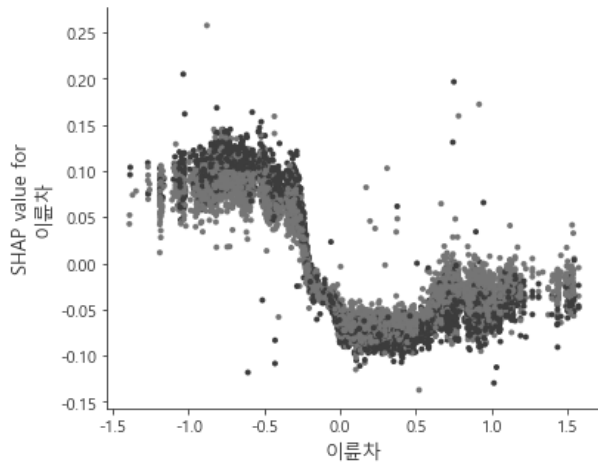
## 변수별 SHAP value 해석 예시

- 법규위반\_신호위반이면 사고 위험도가 **높은** 경향 있음
- 법규위반\_안전운전불이행이면 사고 위험도가 **낮은** 경향 있음
- 신고된 이륜차가 많으면 사고 위험도가 **낮은** 경향 있음
- 사고일시\_새벽이면 사고 위험도가 **높은** 경향 있음
- 음식점 수가 많으면 사고 위험도가 **낮은** 경향 있음
- 도로형태\_교차로 안이면 사고 위험도가 **높은** 경향 있음

# 변수 중요도 해석 - SHAP Dependence Plot



▶ 법규위반\_신호위반이면  
사고 위험도가 **높은** 경향 있음  
(사고 위험도에 긍정적인 영향)



▶ 신고 이륜차가 많으면  
사고 위험도가 **낮은** 경향 있음  
(사고 위험도에 부정적인 영향)

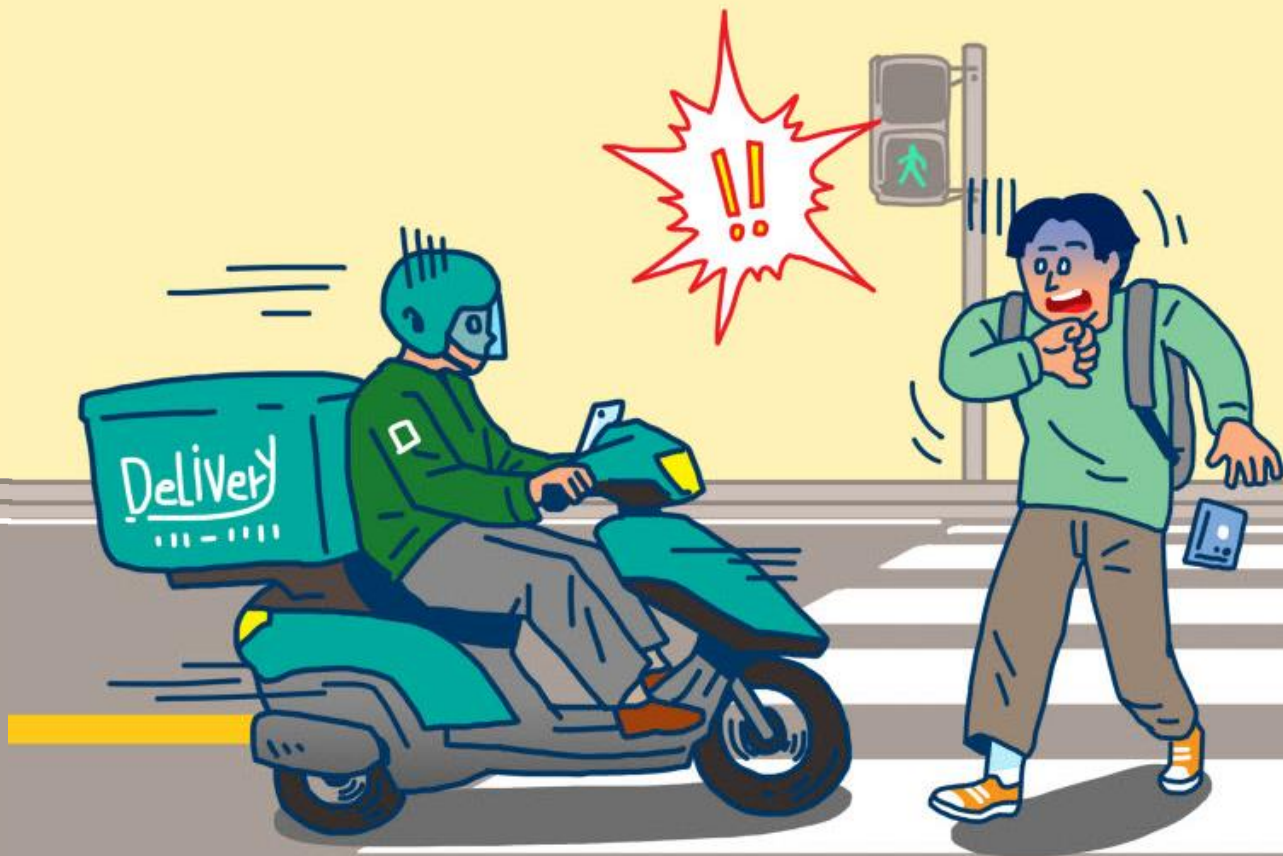
## Top15 변수별 SHAP value 해석

예측 기여 경향성	중요도 Top15 변수
값이 커지면 사고 위험도가 <b>높은</b> 경향이 있는 변수들	<ul style="list-style-type: none"> <li>• 법규위반_신호위반*</li> <li>• 사고일시_새벽*</li> <li>• 기온*</li> <li>• 도로형태_교차로 - 교차로 안*</li> <li>• 노면상태_건조*</li> <li>• 가해운전자 연령_청년*</li> <li>• 가해운전자 연령_청소년*</li> <li>• 요일_일요일</li> <li>• 사고일시_출근</li> <li>• 인구</li> <li>• 법규위반_중앙선침범</li> </ul>
값이 커지면 사고 위험도가 <b>낮은</b> 경향이 있는 변수들	<ul style="list-style-type: none"> <li>• 법규위반_안전운전불이행*</li> <li>• 이륜차*</li> <li>• 음식점 수*</li> <li>• 풍속</li> </ul>



Part 4

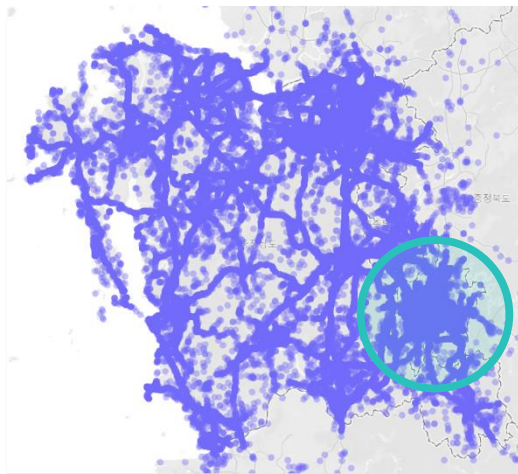
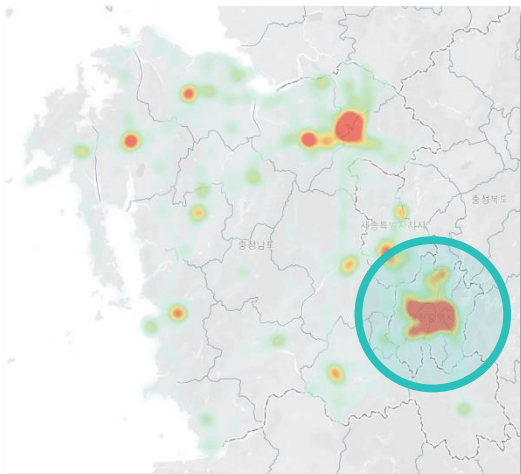
# 이륜차 사고 위험지역 선정



## 대전 위험지역 선정

### 대전 선정 배경 및 위험지역 선정 방법

- 충남: 최근 교통사고 예방 위한 노력 증가, 실제 사고건수 감소
- 세종: 교통환경수치 전국 상위권
- 대전: 제공 데이터 상으로 사고 밀집 지역이었음



대전 지역의 이륜차 사고 위험지역 그리드를 선정하여  
맞춤형 개선을 통해 사고를 줄이고자 함

#### [ 위험지역 선정 방법론 ]


대전, 가해차종 이륜차 사고, 사고 건수  $\geq 2$ 인 그리드 중

- 1) 그리드 별로 평균 X features 생성
- 2) 모델을 활용하여 EPDO 예측

# 이론차 사고 위험지역 산출 공식

이론차 사고 위험도 산출을 위해 **PSI(Potential for Safety Improvement, 잠재적 교통안전개선지수)** 개념 활용

- $EPDO_{\text{예측}}$  : 해당 그리드 내에서 발생할 사고의 EPDO 예측값
- $E(EPDO_{\text{실제}})$  : 해당 그리드 내에서 발생했던 사고들의 EPDO 평균값
- $N$  : 해당 그리드 내에서 발생했던 사고 건수
- $\max(EPDO_{\text{실제}})$  : 해당 그리드 내에서 발생했던 사고들의 EPDO 최댓값

- 
- ① (예측 EPDO - 실제 EPDO) 값이 클수록 위험할 가능성이 높았던 지역
  - ② 실제 EPDO 값이 클수록 위험도가 높은 지역
  - ③ 사고가 많이 발생한 지역일수록 위험도가 높은 지역
  - ④ EPDO 최댓값으로 나눔 : 비슷한 위험도의 사고가 지속 발생한 지역을 추구

$$\text{이론차 사고 위험도} = \frac{\left( EPDO_{\text{예측}} - E(EPDO_{\text{실제}}) \right) \times E(EPDO_{\text{실제}}) \times N}{\max(EPDO_{\text{실제}})}$$

\* PSI : 미국 주도로 교통 행정관 협회(AASHTO)가 발간한 도로 안전 편람(Highway Safety Manual, 2010)에 제시된 지수로 미국에서는 교통안전사업의 우선순위를 결정하는데 활용

## CatBoost 훈련

최적 하이퍼파라미터의 CatBoost 모델 train\_X 데이터에 대해 훈련

```
from catboost import CatBoostRegressor

cb = CatBoostRegressor(depth=7, learning_rate=0.01, random_state=42)
cb.fit(train_X, train_y)

print(f"RMSE for CatBoost : {round(mean_squared_error(test_y, cb.predict(test_X))*0.5,3)}")
print(f"RMSLE for CatBoost : {round(mean_squared_log_error(test_y, cb.predict(test_X))*0.5,3)}")
```

944:	learn: 2.2437217	total: 7.7s	remaining: 448ms
945:	learn: 2.2436950	total: 7.71s	remaining: 440ms
946:	learn: 2.2436479	total: 7.72s	remaining: 432ms
947:	learn: 2.2436065	total: 7.73s	remaining: 424ms
948:	learn: 2.2435599	total: 7.74s	remaining: 416ms
⋮			
990:	learn: 2.2419585	total: 8.15s	remaining: 74.1ms
991:	learn: 2.2419377	total: 8.16s	remaining: 65.8ms
992:	learn: 2.2419175	total: 8.18s	remaining: 57.6ms
993:	learn: 2.2418718	total: 8.19s	remaining: 49.4ms
994:	learn: 2.2418163	total: 8.2s	remaining: 41.2ms
995:	learn: 2.2417876	total: 8.2s	remaining: 33ms
996:	learn: 2.2417134	total: 8.21s	remaining: 24.7ms
997:	learn: 2.2416665	total: 8.22s	remaining: 16.5ms
998:	learn: 2.2416324	total: 8.23s	remaining: 8.24ms
999:	learn: 2.2415496	total: 8.24s	remaining: 0us

RMSE for CatBoost : 2.257  
RMSLE for CatBoost : 0.397

train\_X, test\_X, train\_y, test\_y

지역 = 서울특별시, 6 광역시, 세종특별자치시, 충청남도

가해차종 = 이륜차 or 피해차종 = 이륜차인 데이터



지역 = 대전광역시

가해차종 = 이륜차

데이터로 한정 !

## 그리드 index 데이터프레임

- ❶ grid\_id 별 사고 건수 column 생성(count)
- ❷ grid\_id 별 최대 EPDO 값 column 생성(max\_EPDO)
- ❸ grid\_id 별 사고 건수가 여러 개일 경우 한 grid\_id로 합칠 때 해당 grid\_id의 column별 평균값으로 대체

4 평균값으로 대체된 더미 변수들 전처리(0.5 이하일 경우 : 0 / 0.5 초과일 경우 : 1)

→ 더미 변수들의 경우 0 또는 1로 구분되어야 하기에 0.5, 0.66과 같은 평균값 그대로 입력되는 것은 옳지 않음

```
df.iloc[:, 12:] = np.where(df.iloc[:, 12:] <= 0.5, 0, 1)
```

[illegible]

# 상위 위험지역 선정 - (3) 그리드별 EPDO 예측 및 최종 위험도 산출

## 최종 예측 & 위험도 결과

- 1 CatBoost를 통해 그리드 index 데이터프레임에 대해 예측

```
result = cb.predict(df)
```

- 2 위험도 공식에 필요한 실제 EPDO, 최대 EPDO, 예측 EPDO, 사고 건수를 담은 데이터프레임 생성(df\_final\_sub)

- 3 위험도 공식을 통해 그리드별 이륜차 사고위험도 계산 후 내림차순으로 정렬

```
df_final_sub['danger'] = ((df_final_sub['EPDO_reg'] - df_final_sub['EPDO']) * df_final_sub['EPDO'] * df_final_sub['count']) / df_final_sub['max_EPDO']
```

- 4 describe() 통해 각 수치형 column 값 요약

```
df_final_sub.head()
```

	EPDO	max_EPDO	EPDO_reg	count
grid_id				
2360893.0	15.0	15	4.259322	1
2360897.0	1.0	1	4.370910	1
2360900.0	6.0	6	4.078577	1
2362631.0	8.0	8	4.187592	1
2362636.0	17.0	17	4.748920	1

```
df_final_sub.head()
```

	EPDO	max_EPDO	EPDO_reg	count	danger
grid_id					
2442397.0	2.000000	3	4.222009	4	5.925356
2404243.0	3.000000	3	3.911659	6	5.469957
2397313.0	3.000000	3	4.318848	4	5.275394
2482275.0	3.000000	3	4.654945	3	4.964835
2452802.0	2.333333	3	4.364635	3	4.739703

위험도 상위 3개 그리드

```
df_final_sub.describe()
```

	EPDO	max_EPDO	EPDO_reg	count	danger
count	319.000000	319.000000	319.000000	319.000000	319.000000
mean	4.614741	6.451411	4.153211	2.793103	-0.800605
std	1.693472	3.109630	0.232635	1.320312	3.053326
min	2.000000	3.000000	3.606623	2.000000	-15.954534
25%	3.366667	5.000000	4.014312	2.000000	-2.289134
50%	4.500000	6.000000	4.114330	2.000000	-0.331422
75%	5.500000	8.000000	4.251385	3.000000	1.669682
max	12.000000	18.000000	4.997837	10.000000	5.925356



# 대전 위험지역 분석 - (1) 1위 그리드 : 서구 둔산동 크로바네거리 부근

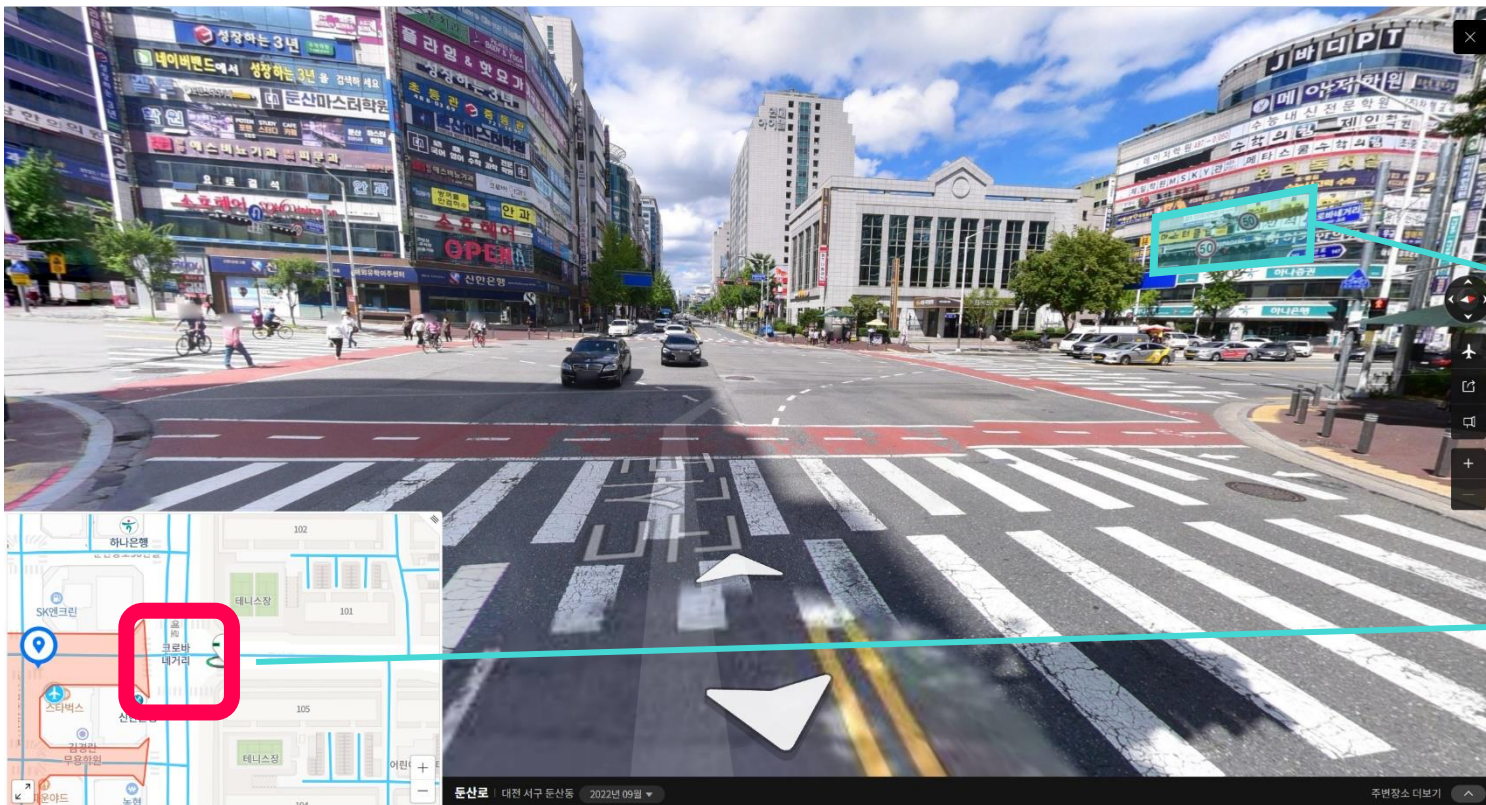


## 위험도에 기여한 중요 변수

위험도를 높이는 데에 기여한 변수		위험도를 낮추는 데에 기여한 변수	
서구의 이륜차 신고 대수	약 9천 대	법규위반_신호위반	0건
법규위반_안전운전불이행 사고 건수	0건	교차로 수	6개
음식점 수	0개	어린이 보호구역 수	1개

단속카메라 수	어린이 보호구역 수	음식점 수	유형주점 수	단란주점 수	교차로 수	버스 정류장 수
0 개	1 개	0 개	0 개	0 개	6 개	2 개

※ 버스 정류장이 평균(0.65)보다 많은 도로, 교차로 수 또한 평균(0.91) 보다 큰 수치이지만 단속카메라 개수가 0개



## 도로 특성

- 신호과속단속장비 존재(제한속도: 50)
  - 불법주정차단속 cctv 존재
  - 주변에 어린이보호구역 존재 (제한속도: 30)

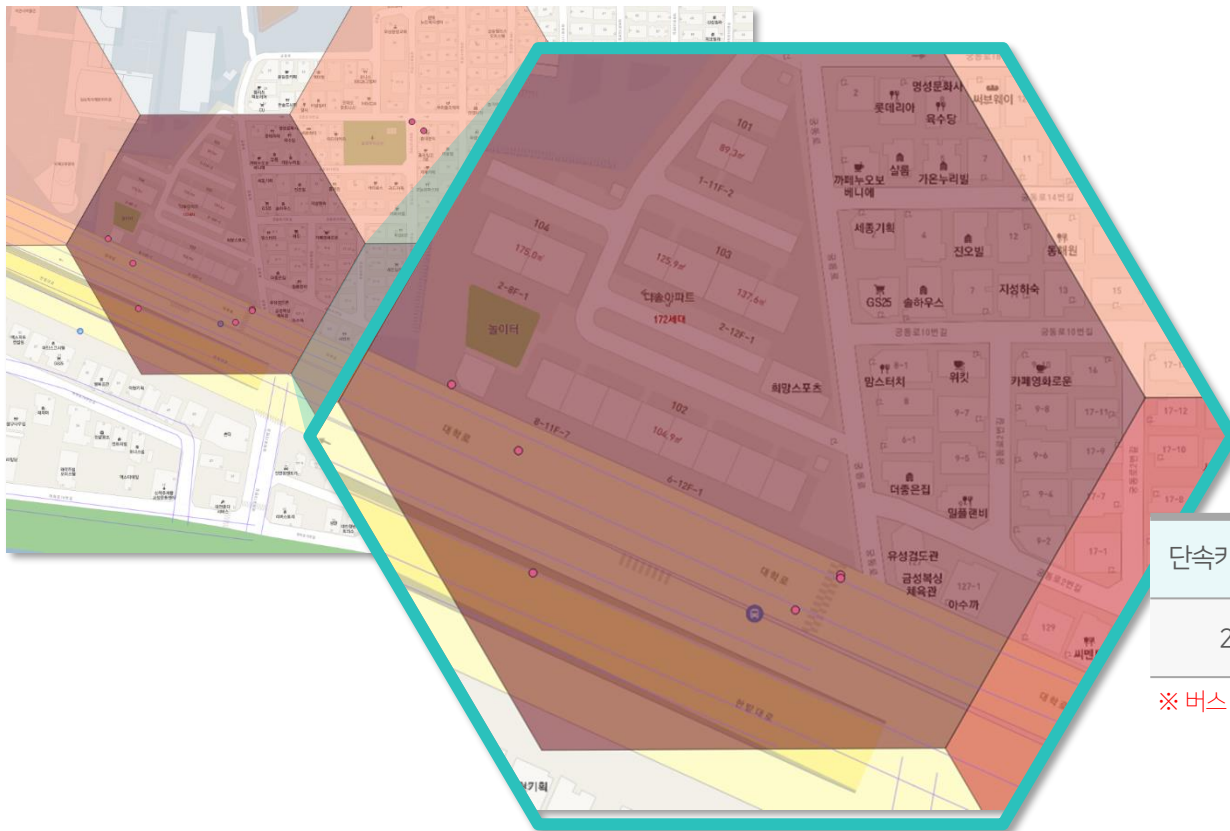
→ 교차로 내에서 사고 발생 위험성 존재

## 제언

- 어린이보호구역이 교차하는 문예로 교차로의 신호과속단속장비 위치에 이륜차단속카메라 설치
- 어린이 보호구역에 불법주정차 단속 카메라 등하교 시간에 인력 배치



# 대전 위험지역 분석 - (2) 2위 그리드 : 유성구 공동 다솔아파트 버스정류장 부근



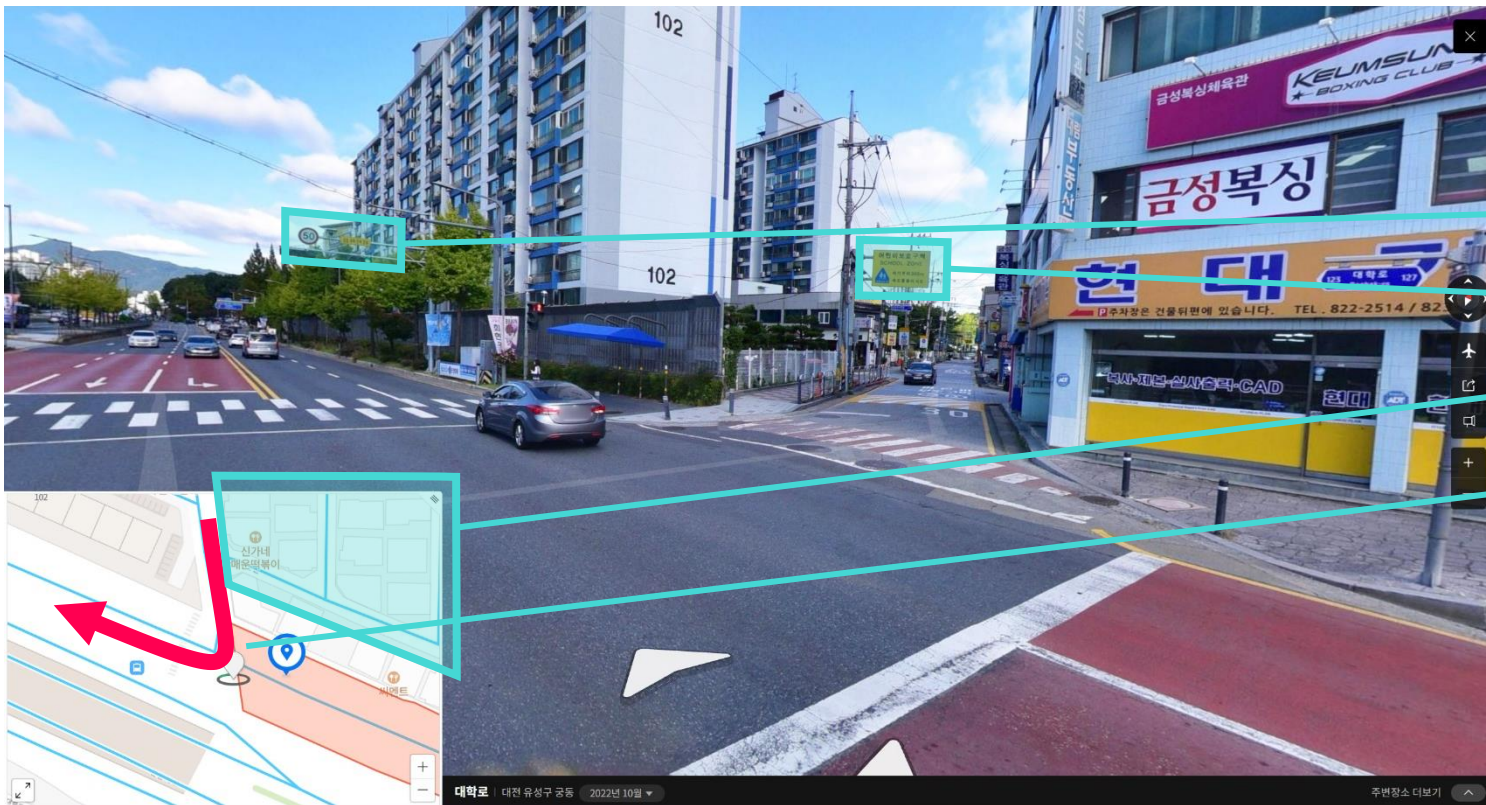
## 위험도에 기여한 중요 변수

위험도를 높이는 데에 기여한 변수		위험도를 낮추는 데에 기여한 변수	
유성구의 이륜차 신고 대수	약 8천 대	음식점 수	139건
법규위반_안전운전불이행 사고 건수	2건	법규위반_신호위반	1건
가해운전자 연령_청년	5건	기온	7°

단속카메라 수	어린이 보호구역 수	음식점 수	유흥주점 수	단란주점 수	교차로 수	버스 정류장 수
2 개	0 개	139 개	0 개	2 개	0 개	<b>1 개</b>

※ 버스 정류장이 평균(0.65)보다 많은 도로, 버스정류장 부근 사고 다발 지역

## 대전 위험지역 분석 - (2) 2위 그리드 : 유성구 궁동 다솔아파트 버스정류장 부근



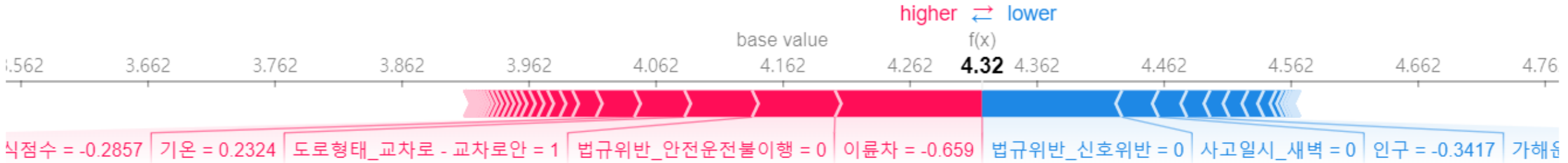
### 도로 특성

- 신호과속단속장비 존재(제한속도: 50)
- 꺾는 순간 어린이보호구역(제한속도: 30)
- 어린이보호구역 골목에 음식점 다수 존재
- 궁동로 골목에서 나오는 차량이 대학로 차량을 주시하다가 횡단보도에서 사고 발생할 가능성 높음 (실제 횡단보도 사고가 있었음)

### 제언

- 궁동로 골목에서 대학로로 진입하는 차량과 대학로의 차량을 인식할 수 있는 대학로에 이륜차단속카메라와 신호위반 단속 카메라 설치
- 궁동로 골목 - 대학로 방향 궁동로 진입로에 반사경 설치
- 청년 이륜차 운전자와 배달 라이더 교육 및 면허 취득 자격 요건 강화

# 대전 위험지역 분석 - (3) 3위 그리드 : 유성구 봉명동 도안마을 1단지 정문 부근



## 위험도에 기여한 중요 변수

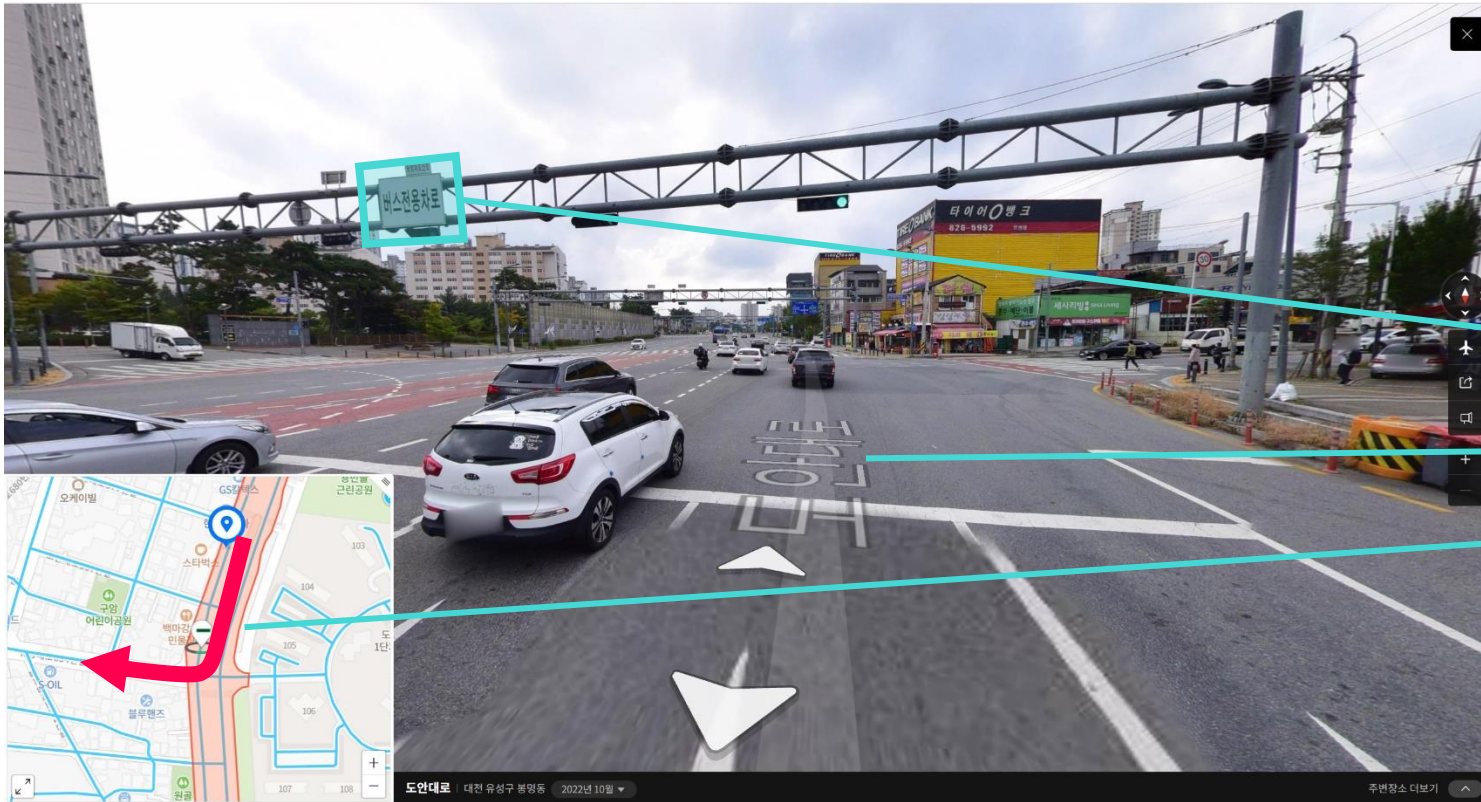
위험도를 높이는 데에 기여한 변수		위험도를 낮추는 데에 기여한 변수	
유성구의 이륜차 신고 대수	약 8천 대	법규위반_신호위반	2건
법규위반_안전운전불이행 사고 건수	2건	사고일시_새벽	0건
도로형태_교차로 - 교차로 안	3건	인구	29만 명

단속카메라 수	어린이 보호구역 수	음식점 수	유흥주점 수	단란주점 수	교차로 수	버스 정류장 수
0 개	0 개	1 개	0 개	0 개	0 개	1 개

※ 버스 정류장이 평균(0.65)보다 많은 도로이지만, 단속카메라 개수가 0개



## 대전 위험지역 분석 - (3) 3위 그리드 : 유성구 봉명동 도안마을 1단지 정문 부근



### 도로 특성

- 단속카메라 없음
- 버스전용차로 있음
- 버스전용차로 제외 왕복 10차선 도로
- 도안대로에서 유성대로654번길로 꺾을 때  
속도제한: 50 → 30으로 바뀜

### 제언

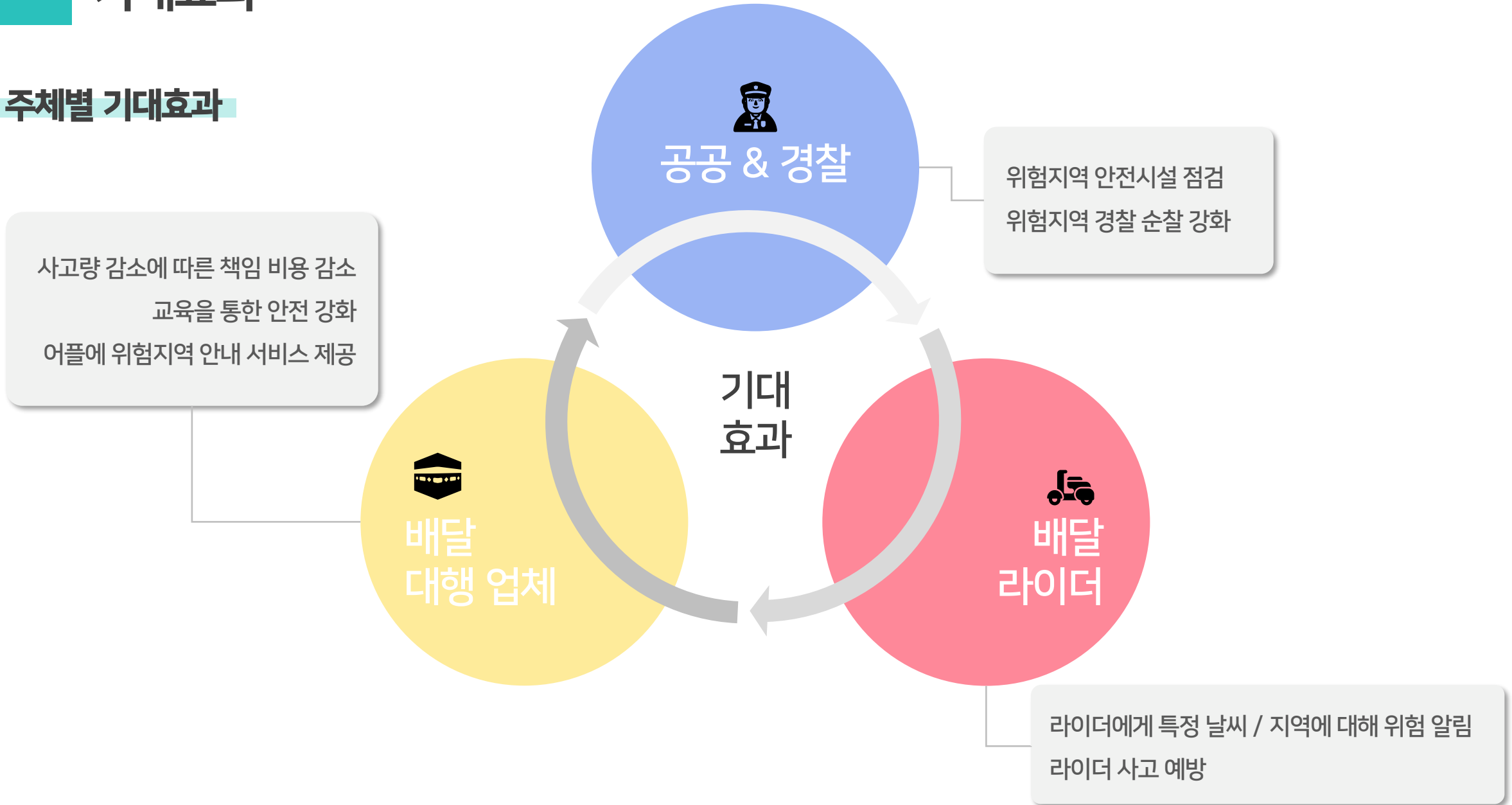
- 많은 차량을 인식할 수 있는 교차로 내 신호과속단속장비 위치에 이륜차단속카메라 설치
- 교차로 내 사고 다발 지역에 주의 표지판 및 단속카메라 설치

## Part 5

# 결론



## 주체별 기대효과



## ☑ 의의

그리드 내 다양한 변수들을 고려하여 이륜차 교통사고 위험도 정량화

행정구역 기준이 아닌 그리드 데이터 사용으로 더 세부적인 공간 데이터 분석 가능

다양한 머신러닝, 딥러닝 모델 적용 및 비교를 통한 최적의 모델 선정, 파라미터 튜닝으로 예측 정확도 향상

이륜차 교통사고 심각도에 영향을 미치는 요인들을 활용해 정책적 시사점 도출

## ☑ 한계

실시간 위험도 추출 어려움

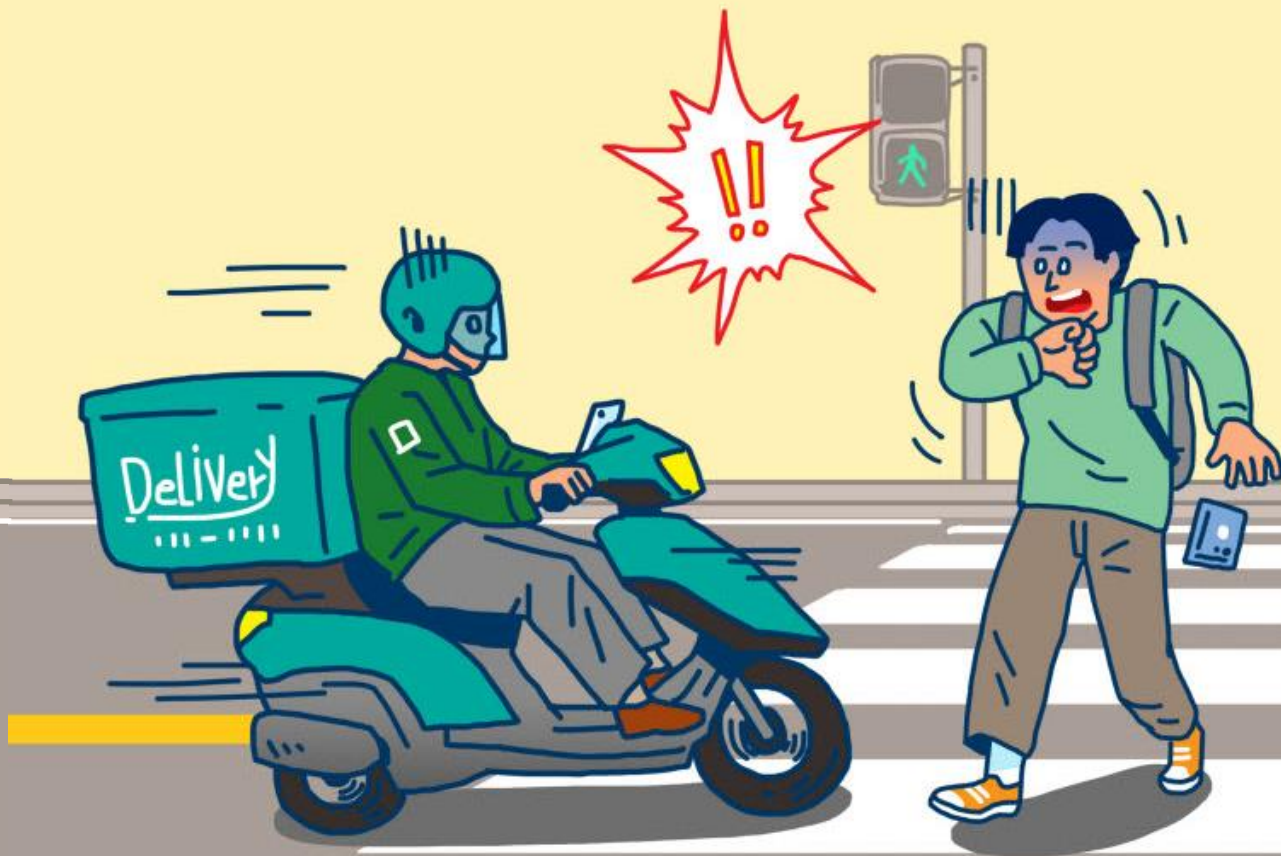
시군구 단위의 데이터로 인해서 정확한 그리드별 데이터 반영 불가

2022년 사고 데이터의 부재로 인해 최근 데이터를 활용하지 못함



Part 6

# 활용 데이터 및 참고문헌 출처





# 활용 툴·데이터 및 참고문헌 출처



활용 툴	활용 분야
Tableau	서론 데이터 시각화
QGIS	그리드 데이터 생성 및 시각화
Python (pandas, numpy, scikit-learn, matplotlib, seaborn, skimage, shap, pytorch 등)	데이터 시각화, 분석, 모델링

논문
임선호, 박은미, & 장현봉. (2009). 교통사고율에 영향을 미치는 요인 분석. 대한교통학회지, 27(4), 41-53.
신정수 외 3명. (2021). 대전시 교통사고 위험지역 분석 및 교통안전시설물 추가 설치 구역 제안. 2021년 한국컴퓨터종합학술대회 논문집
최준희 외 3명. (2021). DNN 기반 지역별 이륜차 사고 심각도 예측. 2021년 한국ITS학회 추계학술대회, pp.349-354
이한빈 외 6명. (2021). 이륜차사고 영향요인 종합분석을 통한 위험지역 도출. 디지털문화아카이브지, Vol.4, No.2
권철우, & 장현호. (2021). XGBoost 를 활용한 이륜자동차 교통사고 심각도 비교분석. 한국 ITS 학회논문지, 20(4), 1-12.

기사	기사 상세
[칼럼] 이륜차 사고 교통법규 위반이 가장 큰 문제	<a href="https://www.motorcycle-story.com/post/4307">https://www.motorcycle-story.com/post/4307</a>
오토바이도 무인단속 추진...후면번호판, '카메라+AI'로 단속	<a href="https://www.etnews.com/20210108000102">https://www.etnews.com/20210108000102</a>

데이터	제공 플랫폼	데이터 상세
부문별 교통사고현황	TAAS	16-21년도 교통사고 유형별 현황
이륜차 법규위반 데이터	공공데이터포털	경찰청_이륜차 위반법규별 단속 현황.csv
대전, 충남, 세종 치안 데이터	스마트치안 빅데이터 플랫폼	NPA2020 KP2020, KP2021
이륜차 사고 데이터	TAAS	19-21년도 가해차종-이륜차 또는 피해차종-이륜차 사고 정보
단속 카메라 데이터	공공데이터포털	경찰청_무인교통단속카메라_20220819.csv 대전광역시_무인교통단속카메라_20221215.csv 세종특별자치시경찰청_무인교통단속카메라_20220520.csv
어린이 보호구역 데이터	공공데이터포털	전국어린이보호구역표준데이터.csv
버스 정류장 데이터	공공데이터포털	국토교통부_전국 버스정류장 위치정보 부산광역시_버스 정류소 정보(SHP)
교차로 데이터	국가교통정보센터	전국표준노드링크
음식점 데이터	LOCALDATA	19-21년도 영업 일반음식점 정보 19-21년도 영업 단란주점영업 정보 19-21년도 영업 유흥주점영업 정보
기후 날씨 데이터	기상자료개방포털	종관기상관측(ASOS) 방재기상관측(AWS)
이륜차 데이터	KOSIS	이륜차신고현황_ 시도별
인구 데이터	KOSIS	행정구역(시군구)별 인구수

# 2023년 지역 치안 안전 데이터 분석 공모전 이륜차 교통사고 위험도 분석 및 대전광역시 위험지역 선정

## 감사합니다.

폴보아즈(Pol-Boaz)

신재욱 (jwshin0908@naver.com)

김혜연 (sally879@naver.com)

박규연 (20203065@kookmin.ac.kr)

최유진 (youjin0100450@sookmyung.ac.kr)

