

4조\_보람상조

# 교통사고 유발요인 분석 및 사망사고 예측

조장 : 우성균

조원 : 김민철, 이성민, 이유경, 안소연

# CONTENTS

01



기획 배경

02



EDA 및 머신러닝 분석

03



한계 및 보완점

04



프로젝트 의의

05



Epilogue

## 역할 분담

우성균\_16 : ML - IMBALANCE CLASSIFICATION: 전체사고중 사망사고 분류예측(데이터 전처리 모듈 제작,  
머신러닝 및 딥러닝 모델 학습, 결과 해석)  
- LSTM: 서울시 시계열 사고건수 예측 (데이터 전처리, 오차율 및 평균 오차율 그래프 제작, 모델 성능 개선)

김민철\_15 : EDA - 서울시 연령별 운전자 비율과 사고율 분석 (2020년)  
- 전국 시간대별 교통사고 사망률과 치사율 분석 (2020년 ~ 2021년)  
ML - LSTM을 이용해 서울시 시계열 사고건수 예측

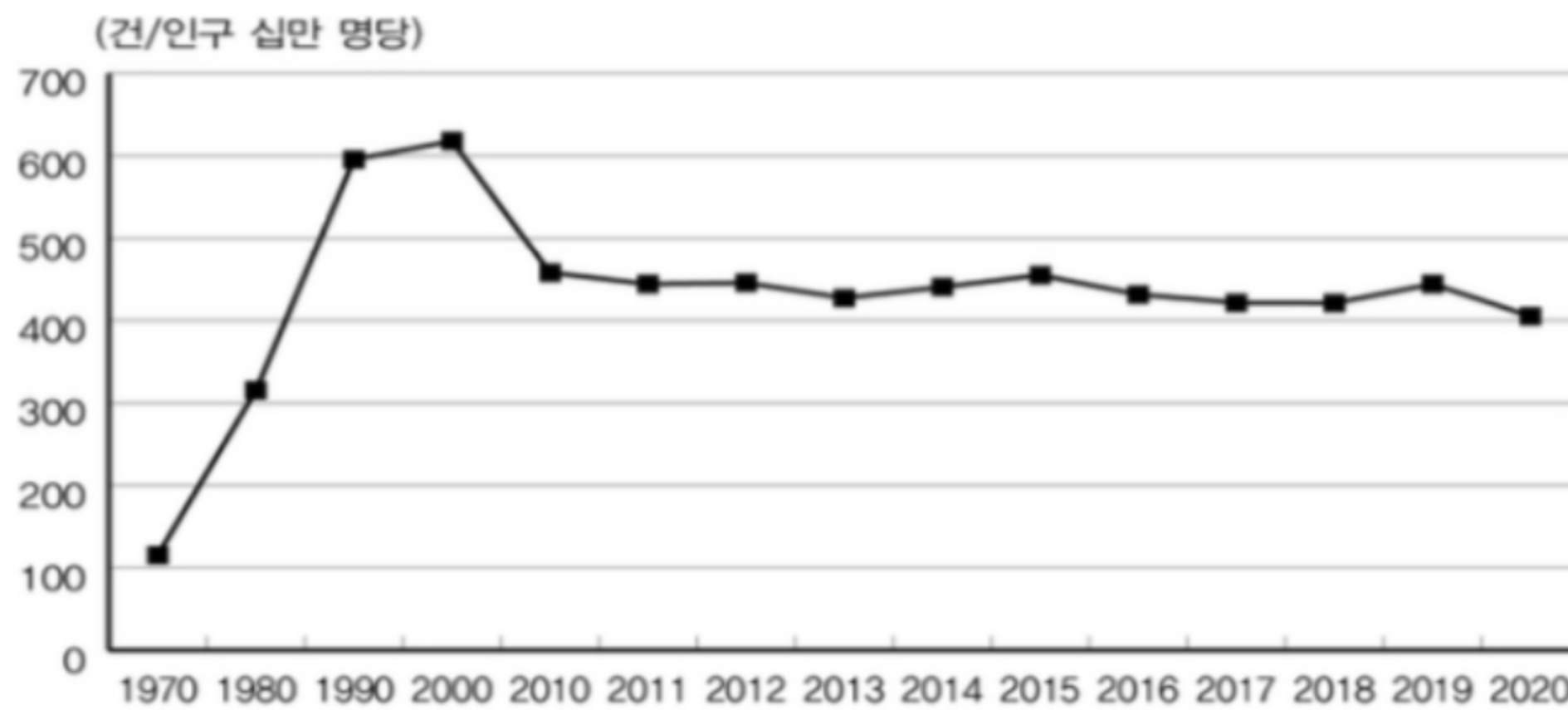
이성민\_15 : EDA - 국내 법규 위반과 사고율에 대한 분석, 해외(OECD) 국가들의 사고율 및 사망률 분석  
ML - RANDOMFORESTREGRESSOR를 이용해 서울 내 사고 건수 예측

이유경\_16 : EDA - 국내 운전면허 시험 난이도와 사고율 및 사망률 분석  
ML - LSTM 서울시 사고건수 예측

안소연\_16 : EDA - 기상 상태별 교통사고 치사율 및 사망률 분석  
ML - LSTM 서울시 사고건수 예측

## 기획 배경

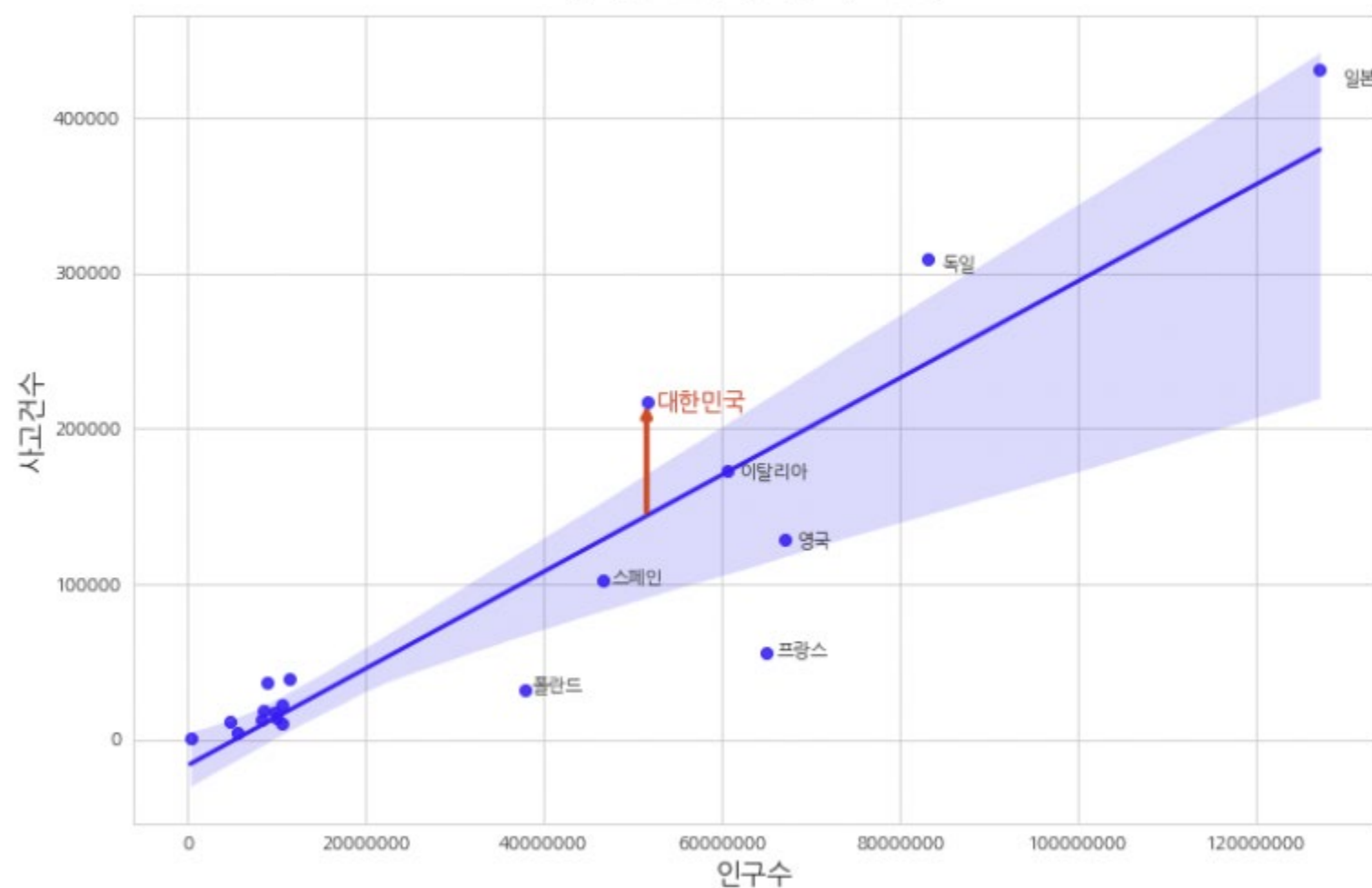
보조 도로교통사고발생률, 1970-2020



교통사고 발생률은 감소하는 추세를 보임

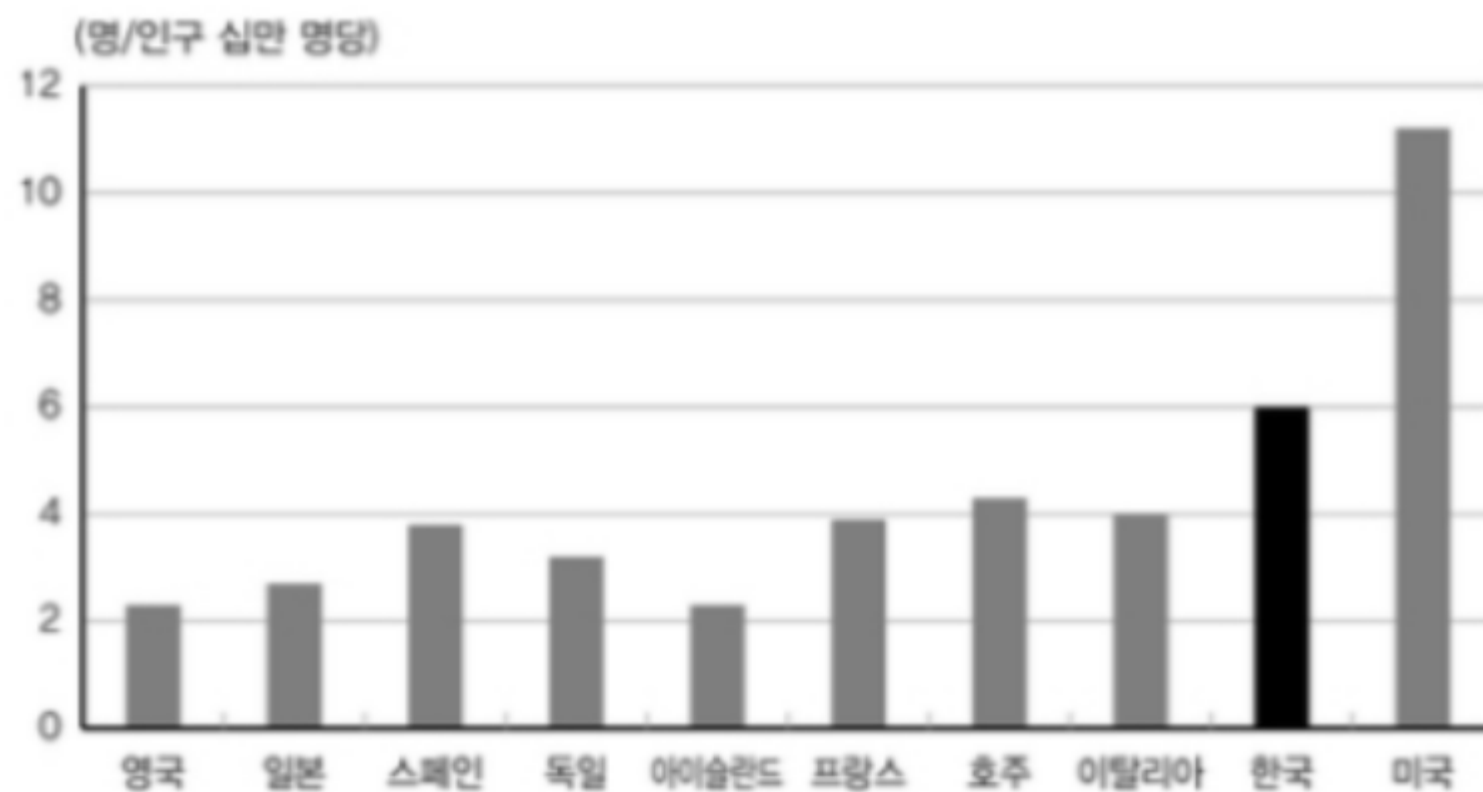
## 기획 배경

국가별 인구수와 사고건수



인구수와 사고건수는 상관 관계를 가지고 선형회귀식을 가지는 데,  
그 회귀식을 웃도는 사고 건수가 발생하고 있음

국제 OECD 주요국의 도로교통사고사망률, 2020



교통사고 사망률이 OECD 주요국 대비 2위  
인구 1천 명당 사고 건수 OECD 13개국 대비 1위

**[ 사고 원인 및 사고 발생시 사망까지 이르는 요인을 분석, 이를 예방할 수 있도록 하고자 함 ]**

## EDA 분석

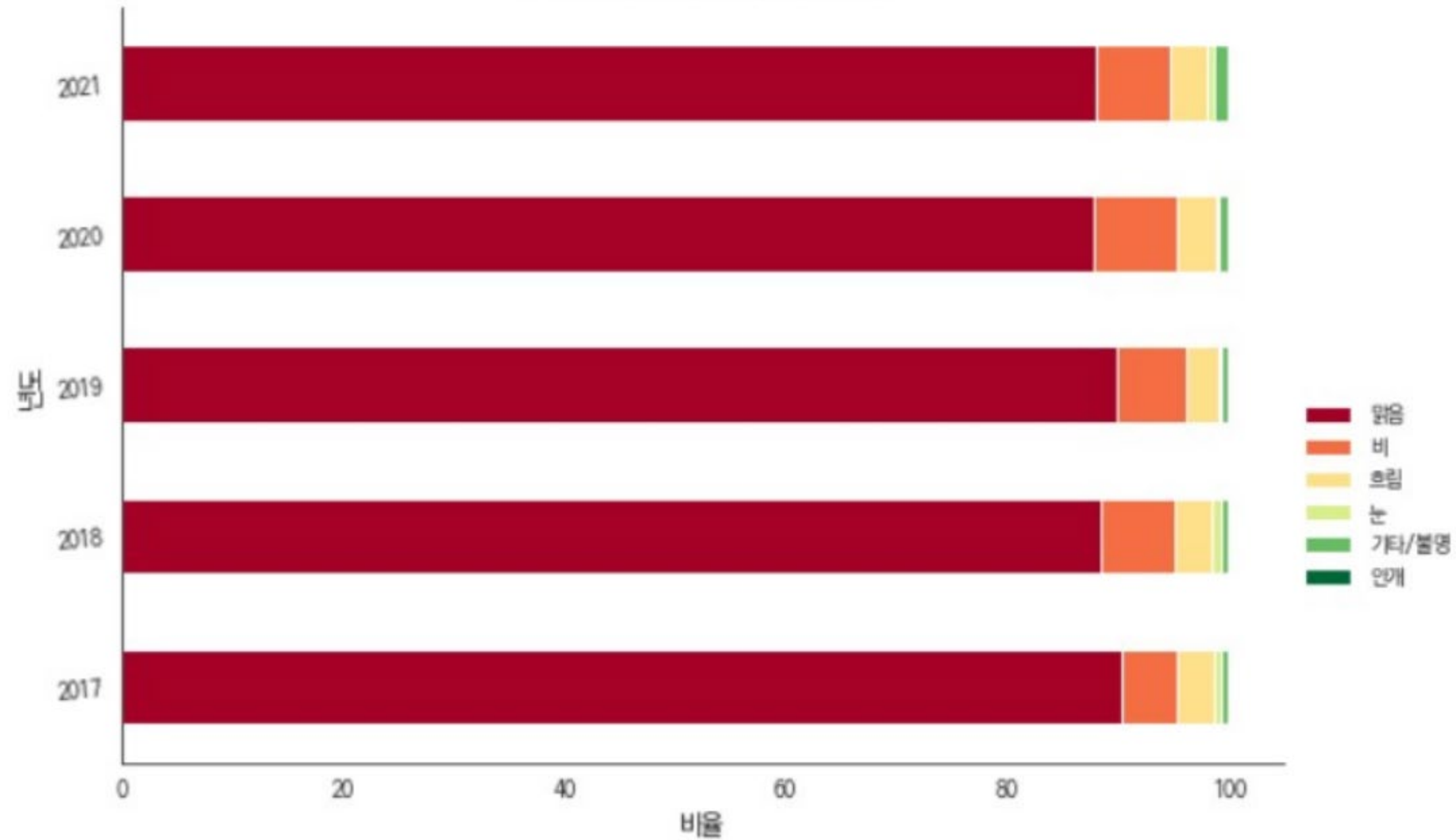
- 기후 : 안개가 낀 상황에서 교통사고 사망률이 가장 높음
- 발생 시간 : 다른 시간대에 대비 02시부터 06시까지의 사망률이 가장 높았음
- 연령대 : 20대 사고율이 64.4%로 압도적으로 높았음
- 법규 : 국내 전체 사망률을 꾸준히 감소하나, 과속으로 인한 사망률이 11%로 높음
- 해외 국가 비교 : OECD 평균 인구 10만 명당 사망자 수는 5.1명, 우리나라 인구 10만 명 당 7.3명으로 평균을 웃돌음

## 머신러닝 분석

LSTM(시계열 데이터) : 평균 오차율 15.2%의 모델을 생성함

## EDA 및 머신러닝 분석\_기후

최근 5년간 교통사고 비율

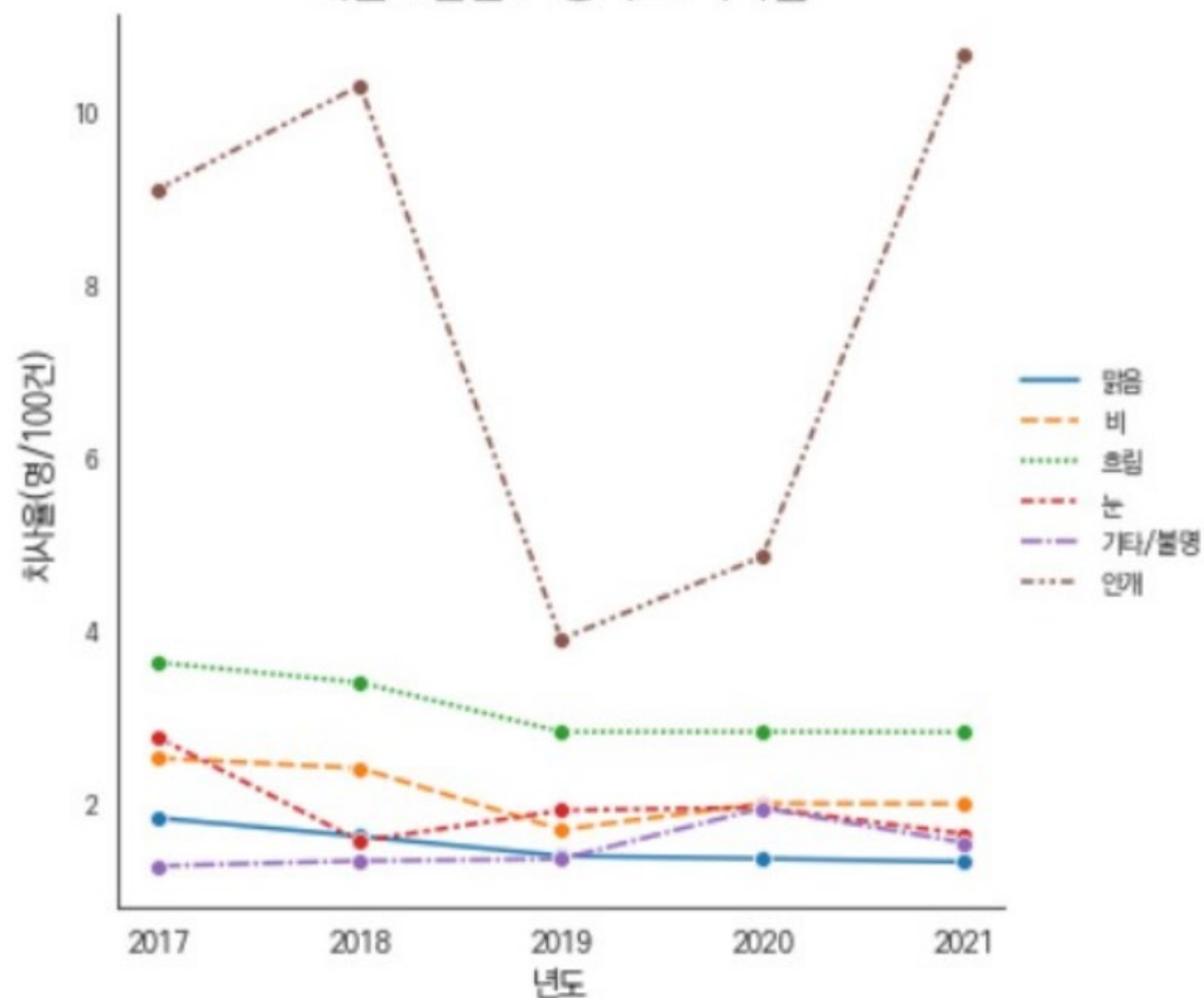


전체 사고별 기상상태의 비율은 맑음 > 비 > 흐림 > 눈 순으로 나타남

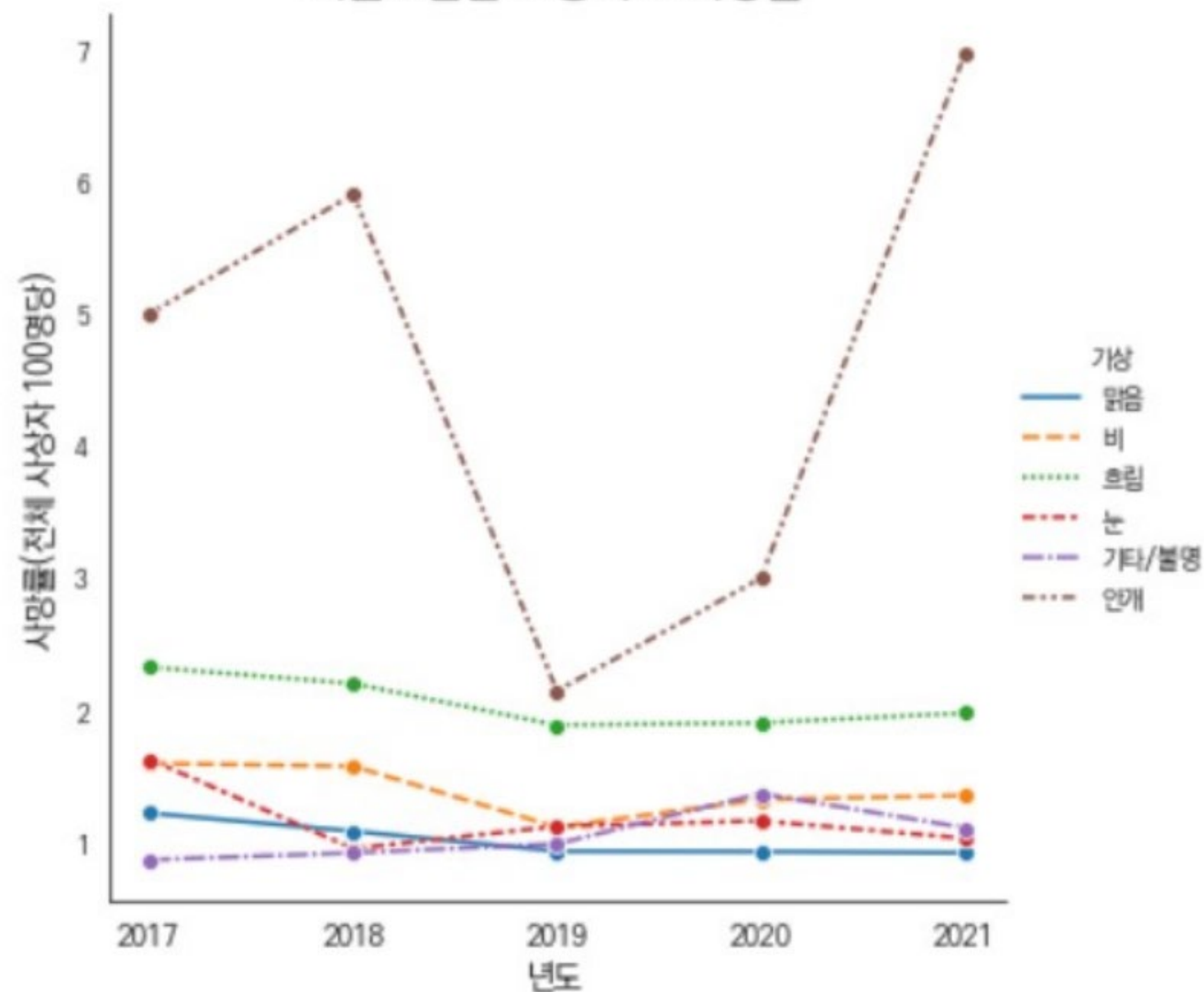


## EDA 및 머신러닝 분석\_기후

최근 5년간 교통사고 치사율



최근 5년간 교통사고 사망률

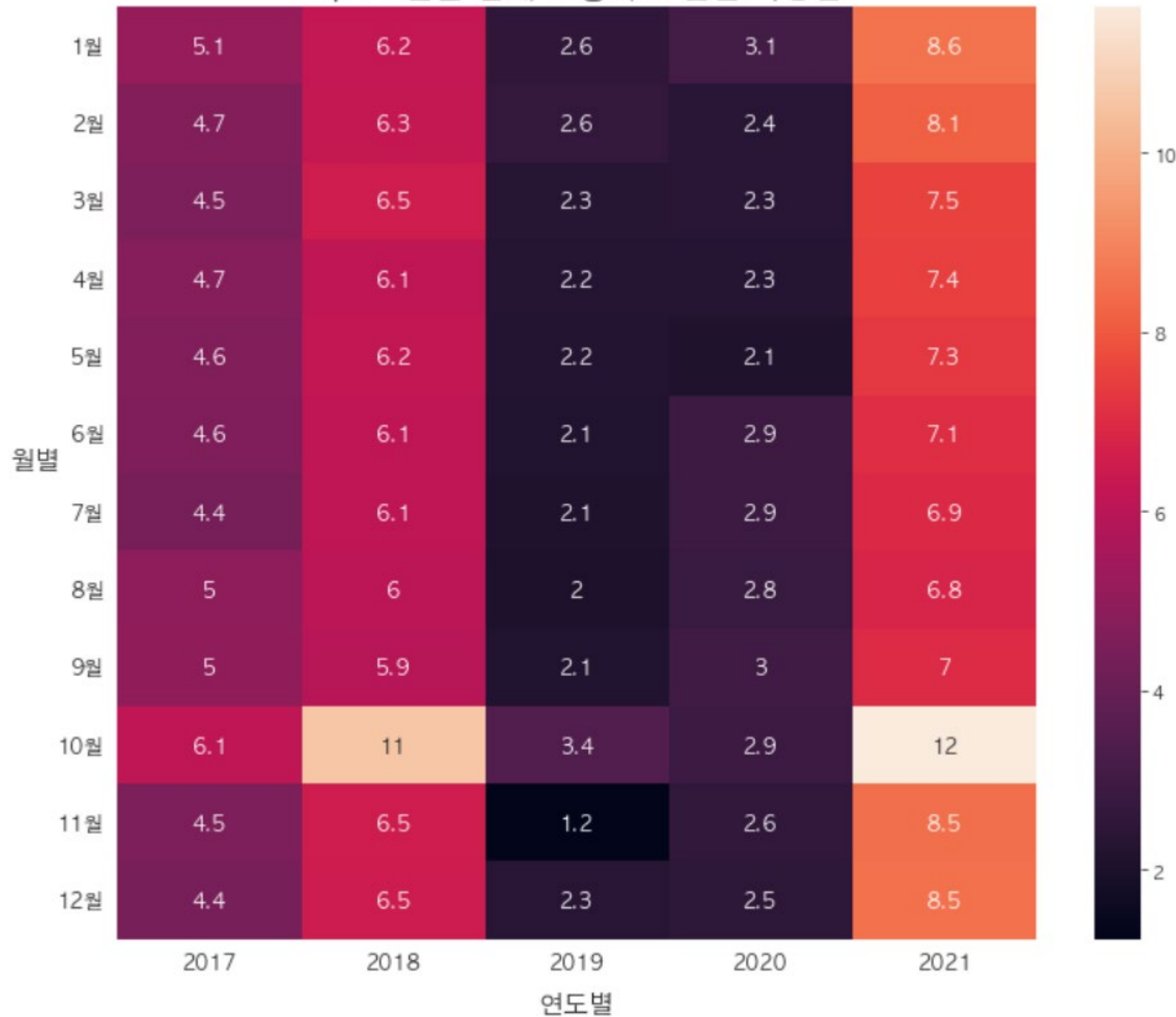


안개 낀 날 치사율 및 사망률이 높게 나타남



## EDA 및 머신러닝 분석\_기후

최근 5년간 안개 교통사고 월별 사망률

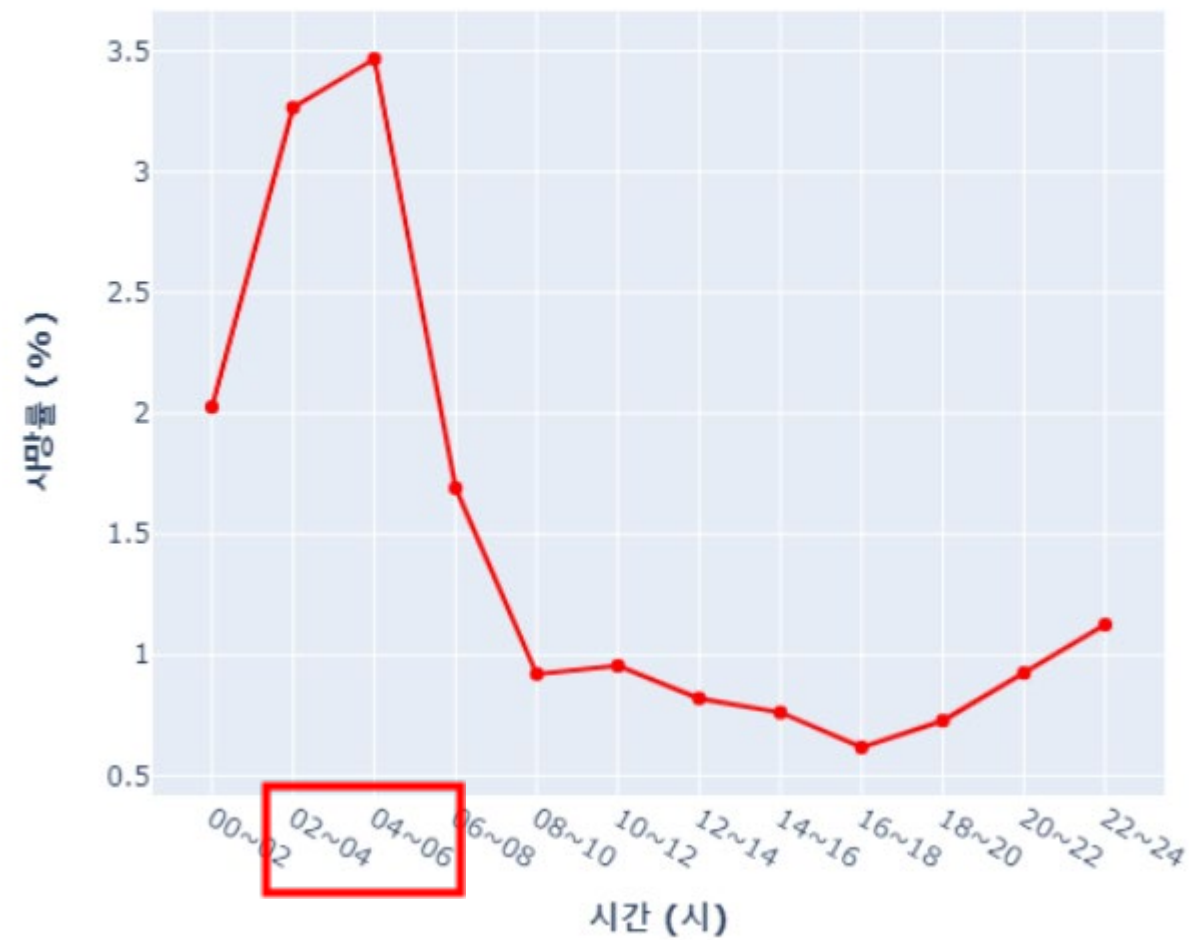


안개로 인해 시야확보가 어려움

평소보다 안전거리 및 서행이 중요함

## EDA 및 머신러닝 분석\_시간

2021년 전국 시간대별 교통사고 사망률



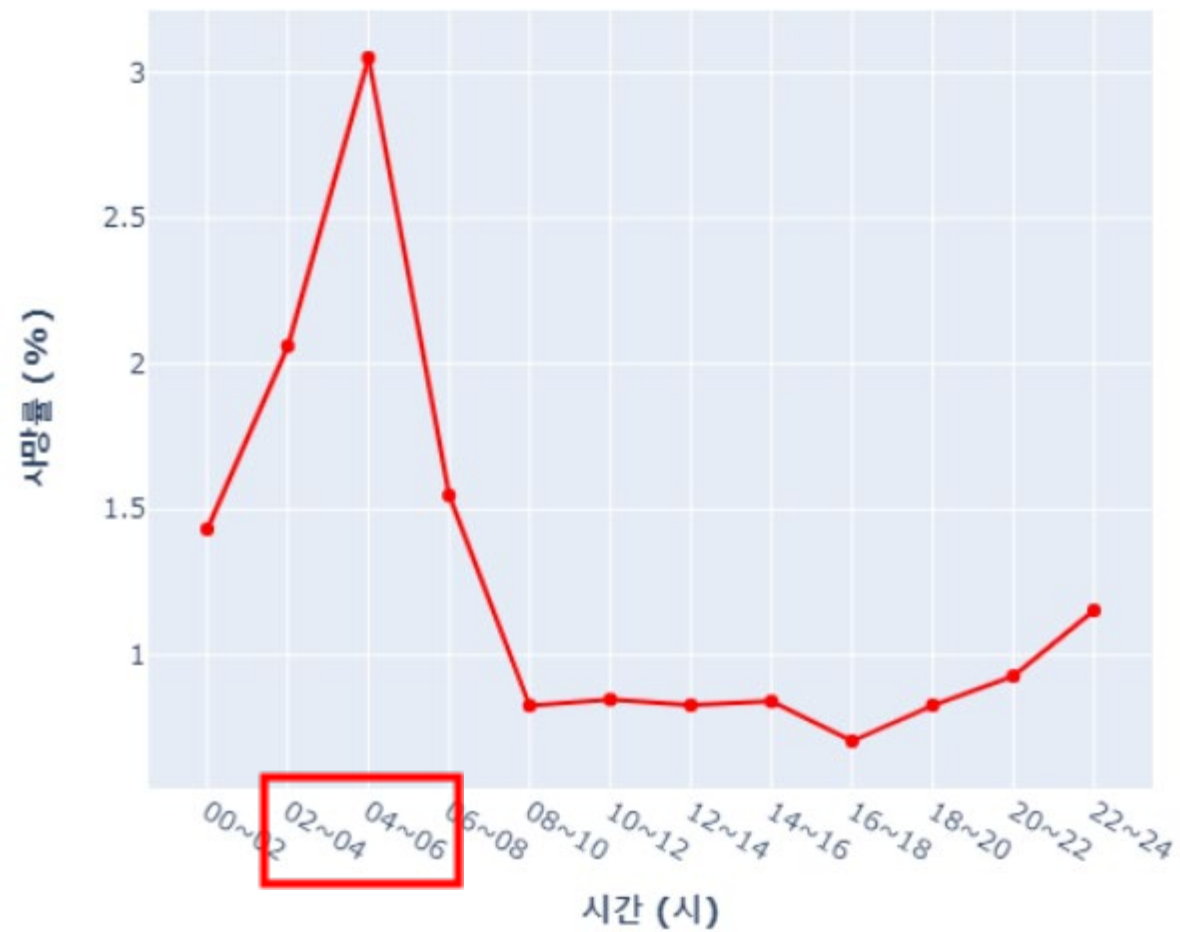
2021년 전국 시간대별 교통사고 치사율



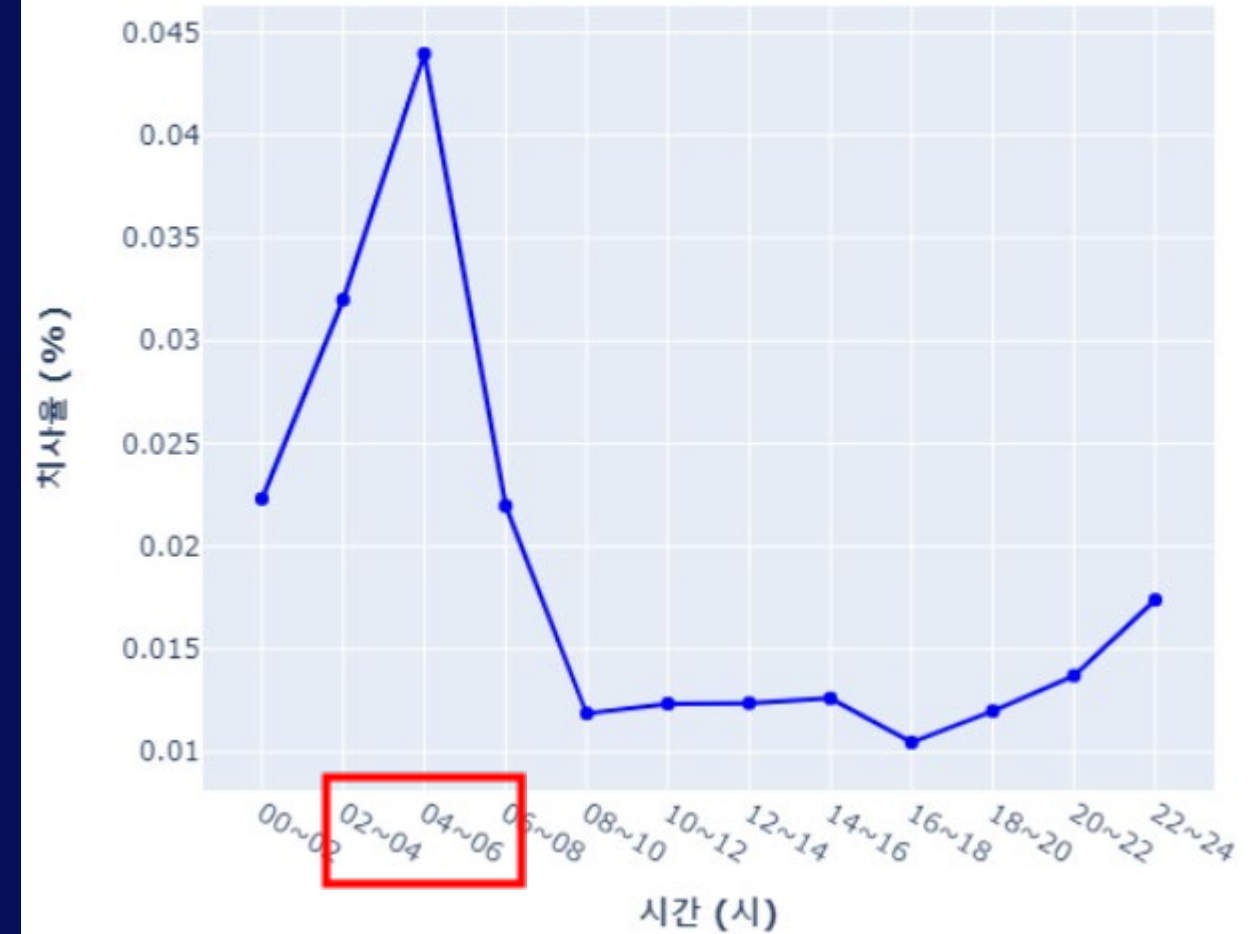
사람과 단속이 적은 새벽시간대  
교통법규를 준수하지 않았을 것이라 추측

## EDA 및 머신러닝 분석\_시간

2020년 전국 시간대별 교통사고 사망률



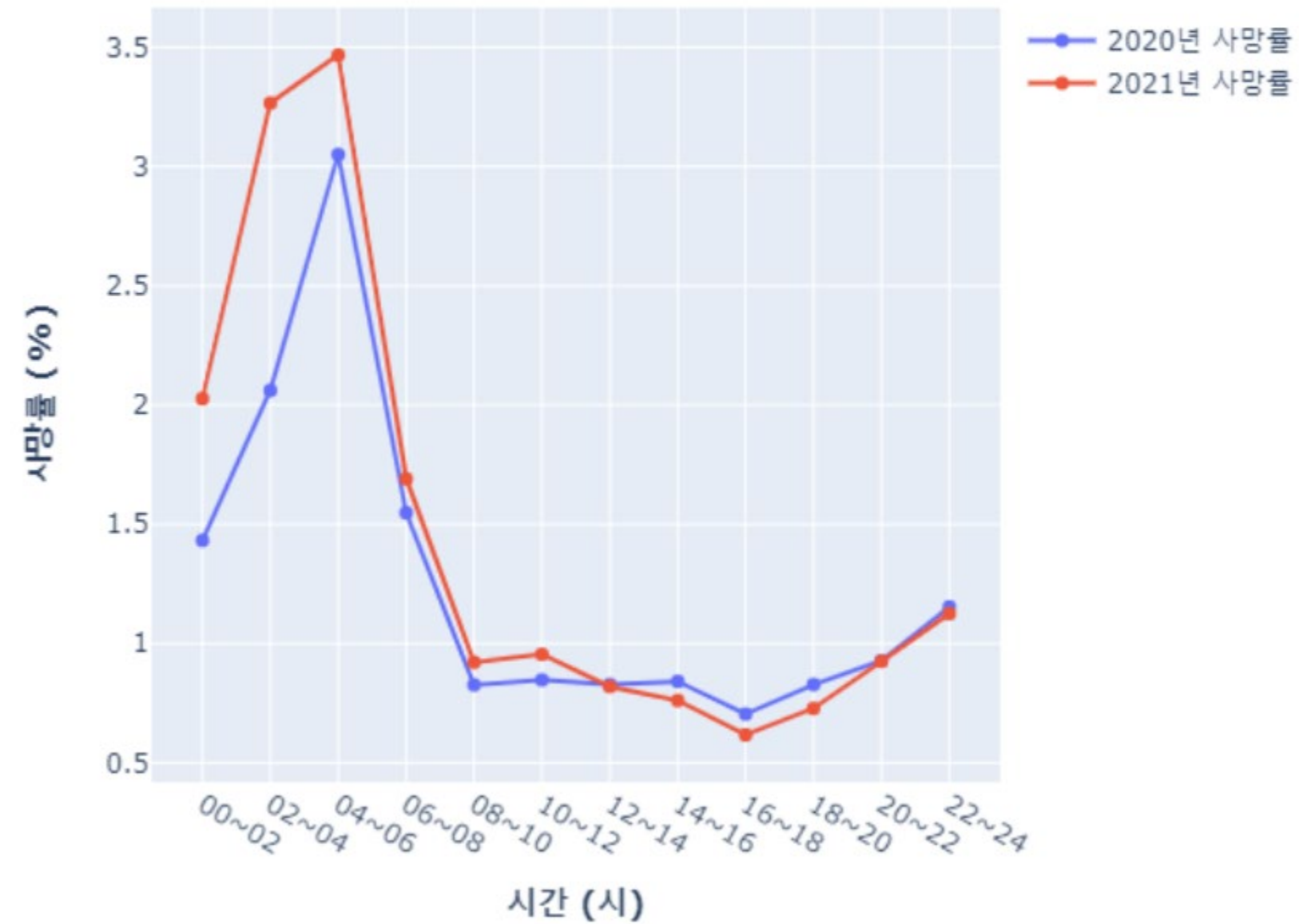
2020년 전국 시간대별 교통사고 치사율



오전과 오후 대비, 새벽 시간대 특히 새벽 02시에서 06시 사이에 사망률이 아주 높음

## EDA 및 머신러닝 분석\_시간

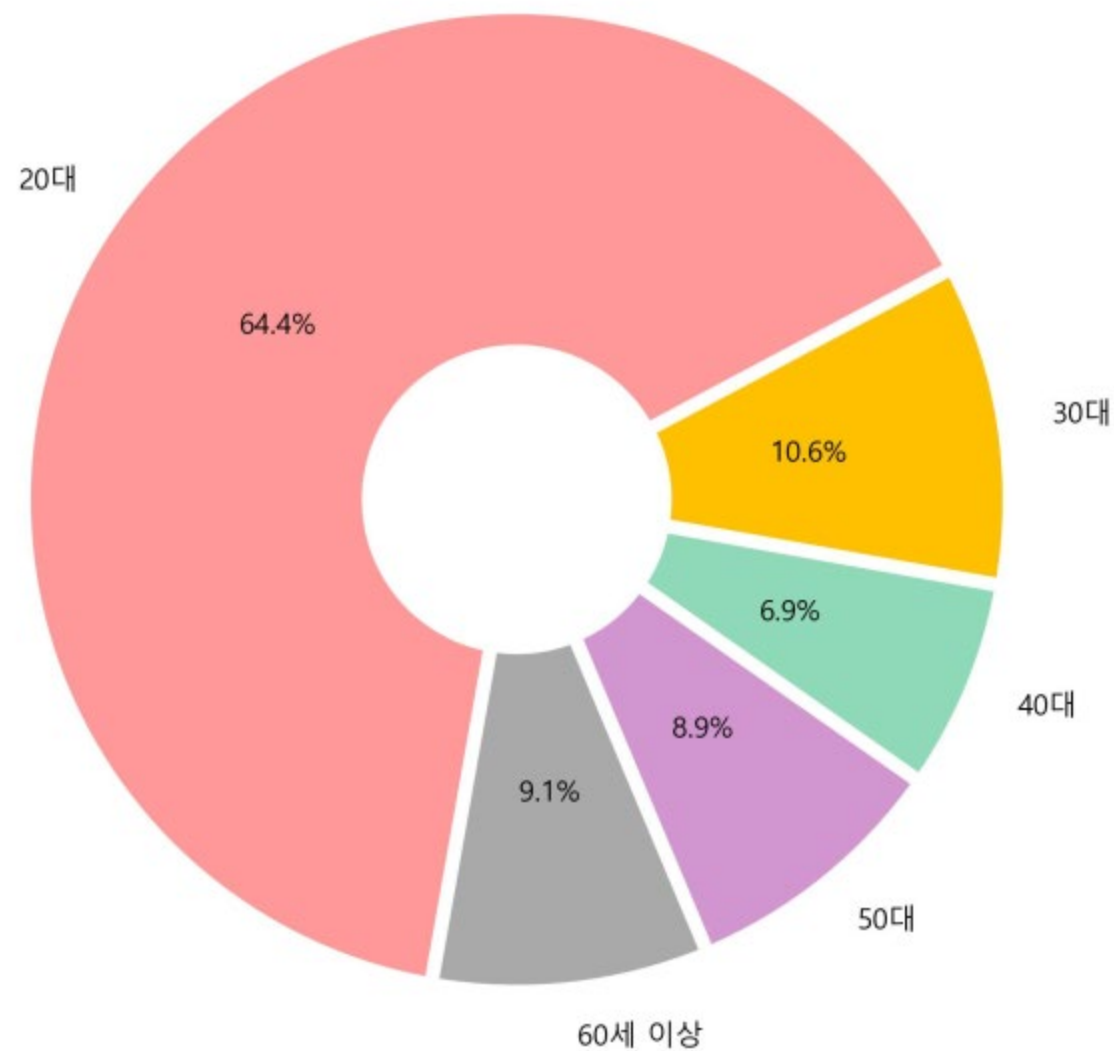
전국 시간대별 교통사고 사망률 (2020년 ~ 2021년)



매년 시간대별 교통사고 사망률은 전체적으로 유사한 형태를 띄고 있음

## EDA 및 머신러닝 분석\_연령대

2020년 서울시 연령대별 사고율



10대가 가장 높고,  
10대를 제외하고 20대 사고율이  
다른 연령보다 6배 이상 높음

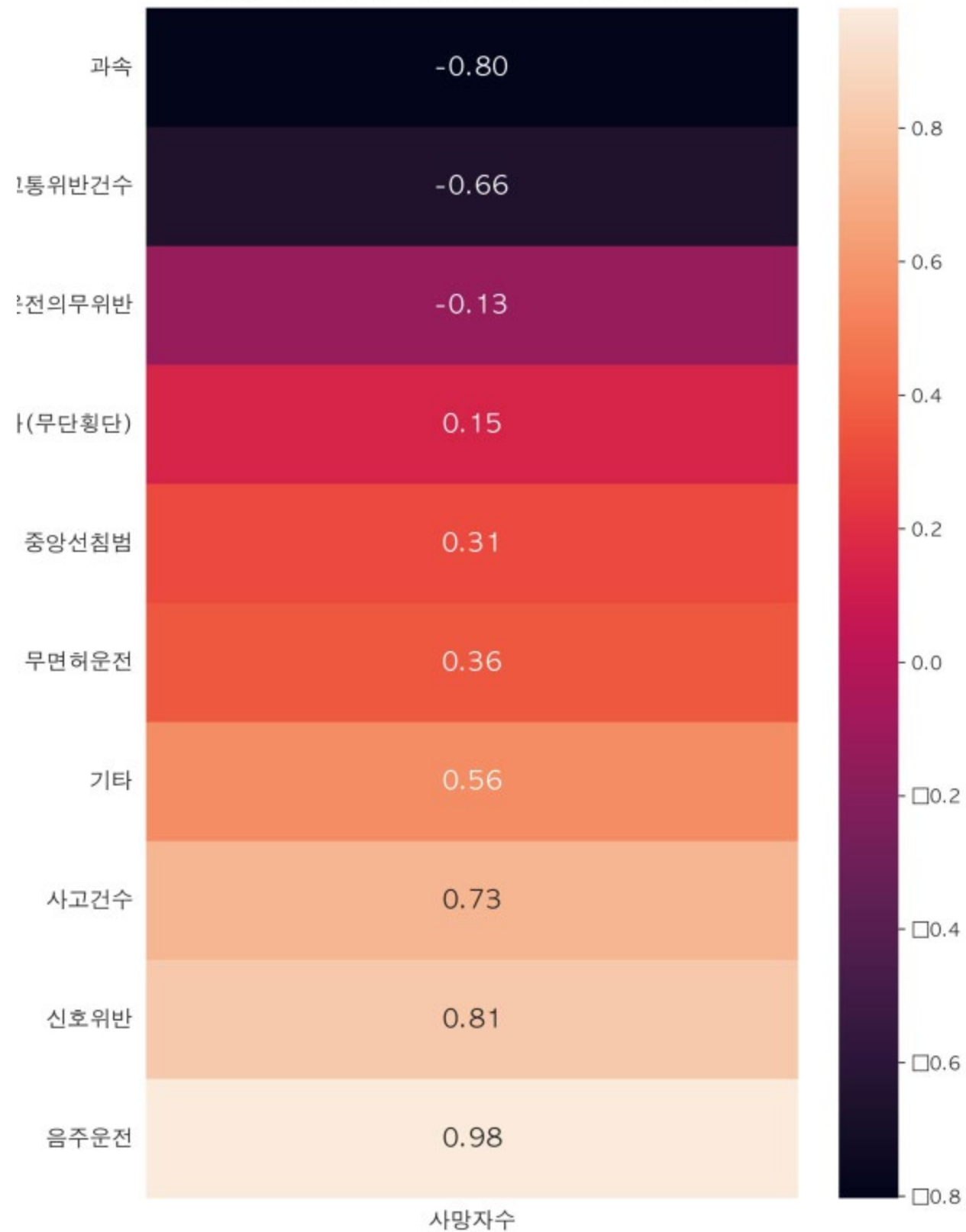


경험이 부족해 발생하는 운전미숙,  
부족한 운전 및 안전 교육을 원인으로 추측



## EDA 및 머신러닝 분석\_법규 위반

법규 위반과 사망자수의 상관관계



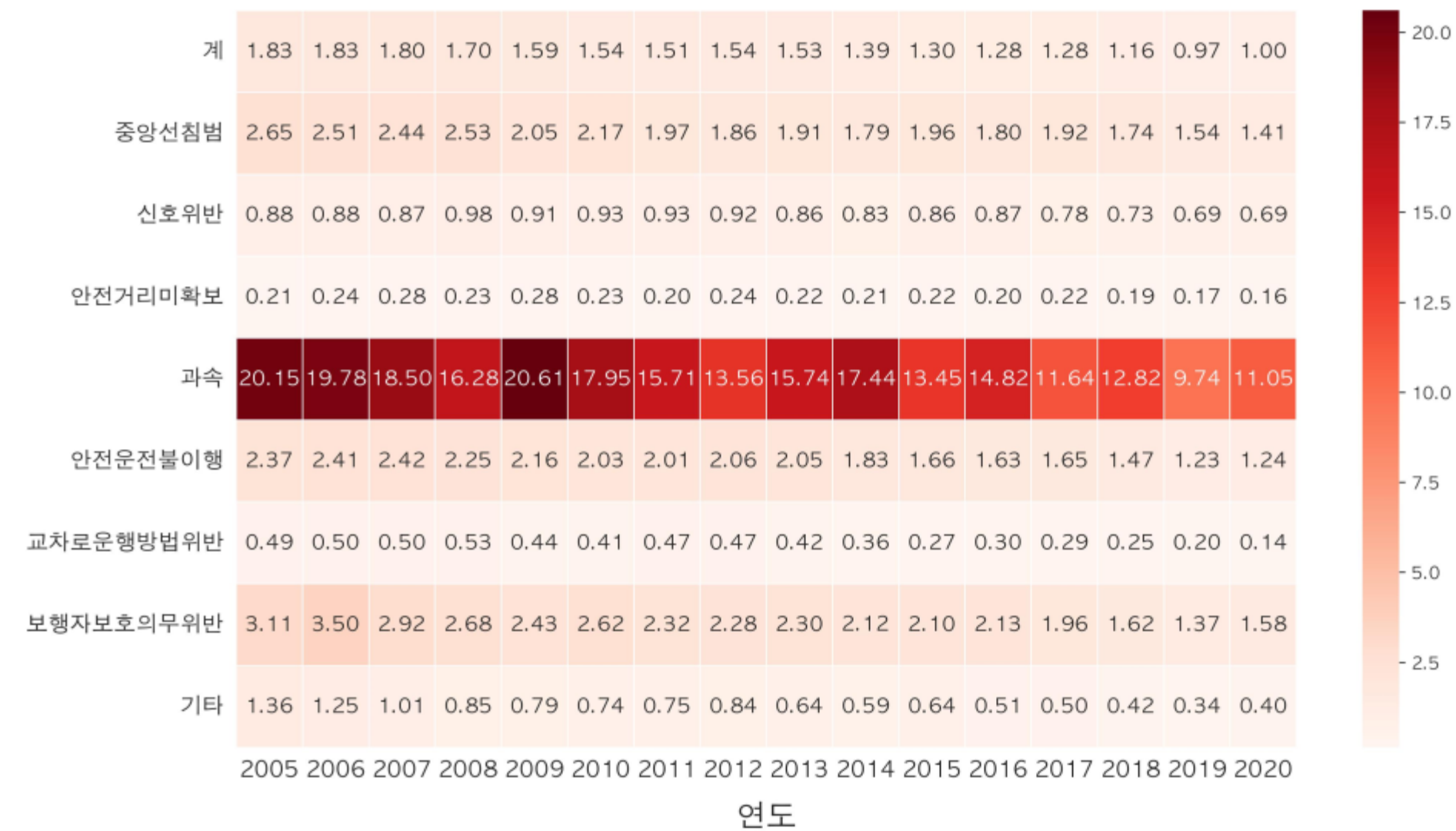
사망자수와 전체 교통법 위반 건수는 음의 상관관계



음주운전과 신호위반은 강한 양의 상관관계

## EDA 및 머신러닝 분석\_법규 위반

법규 위반별 사망자율 변화



법규 위반별 사고의 사망율 추이

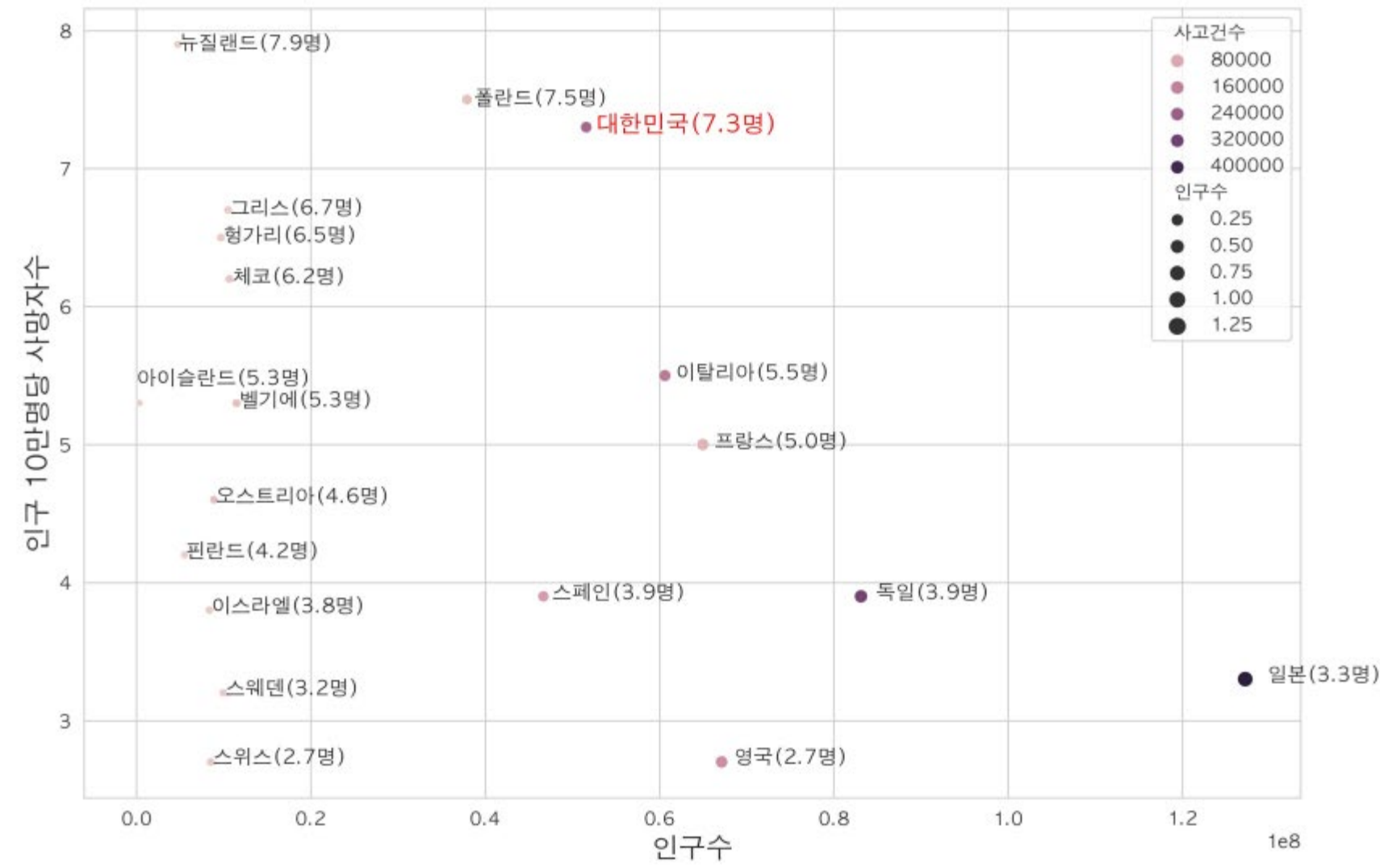


과속으로 인한 사망자는 감소하고 있지만, 다른 법규 위반에 비해 높음



EDA 및 머신러닝 분석\_법규 위반

교통사고 사망율(OECD국가 중 19개국)

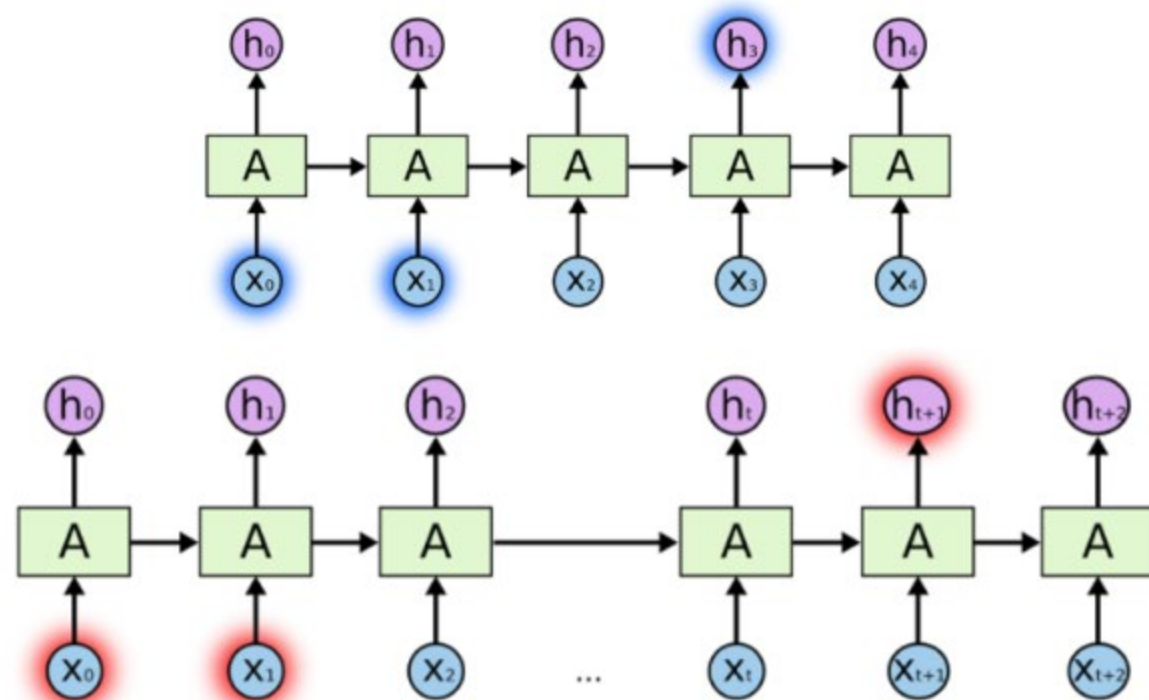
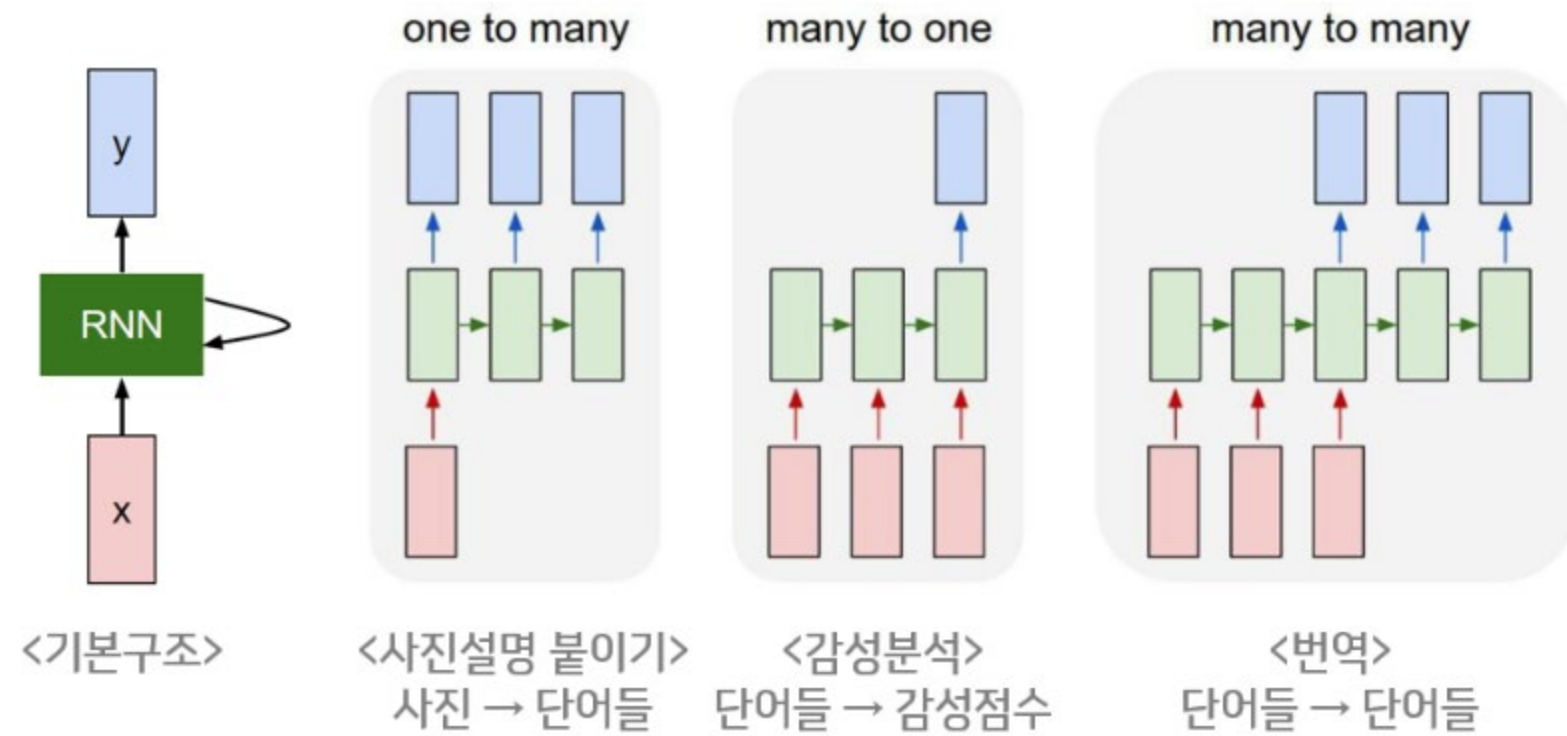


사고건수 대비 교통사고 사망율(OECD국가 중 19개국)

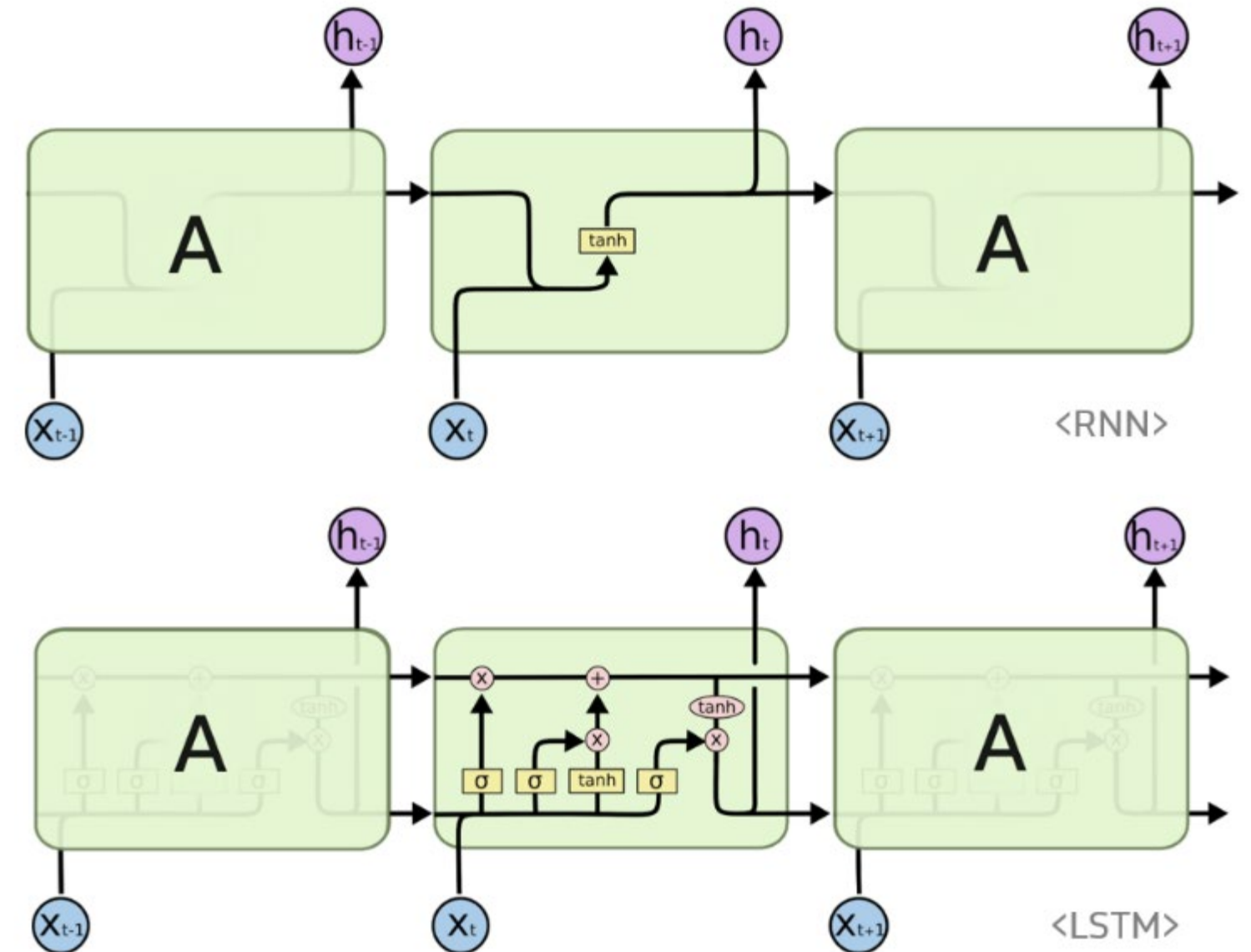


인구 10만명당 교통사고 사망률이 7.3명으로 높은편

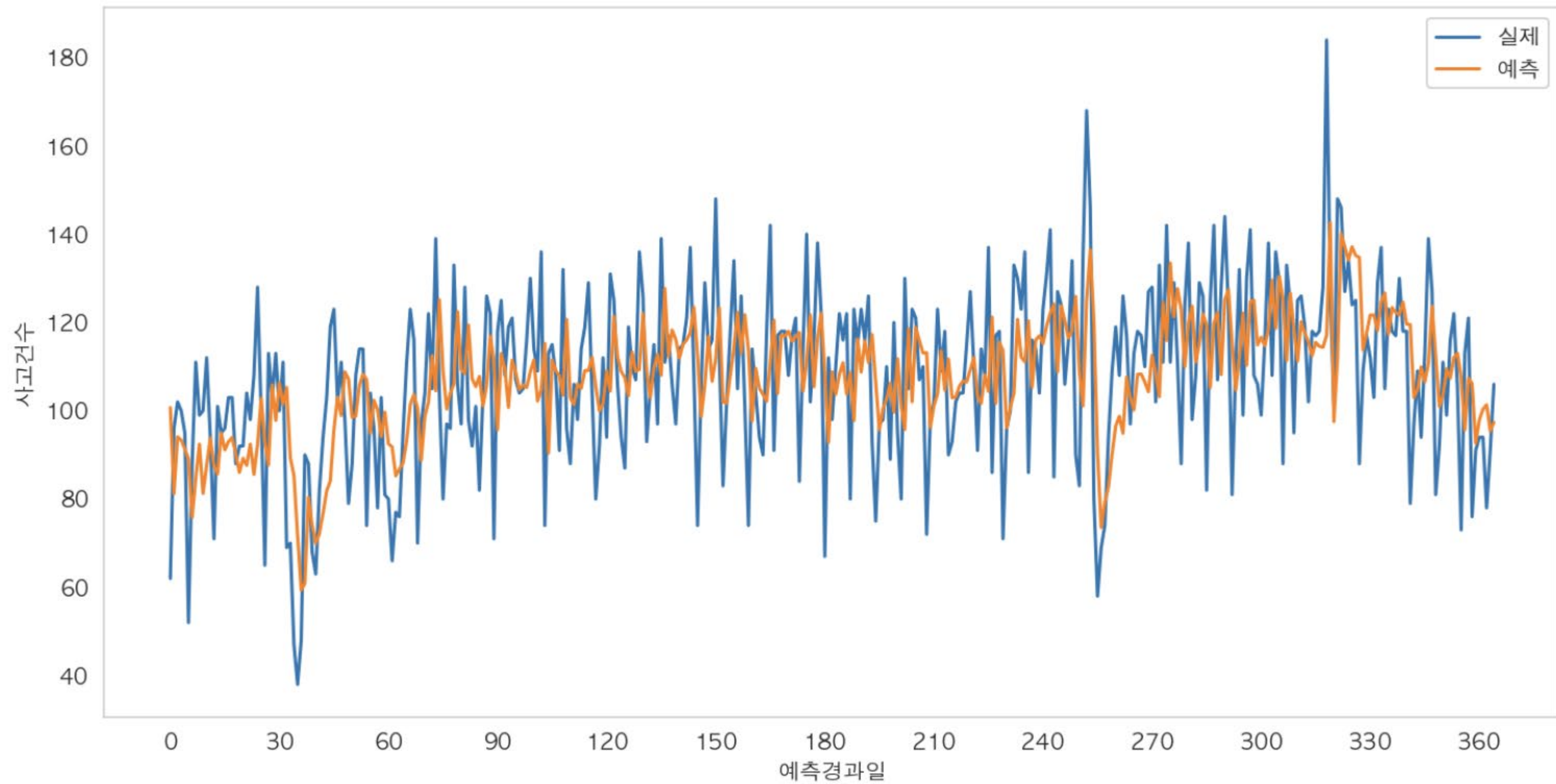
## EDA 및 머신러닝 분석



<관련 정보와 그 정보를 사용하는 지점 사이 거리가 멀 경우 RNN 학습능력 저하>

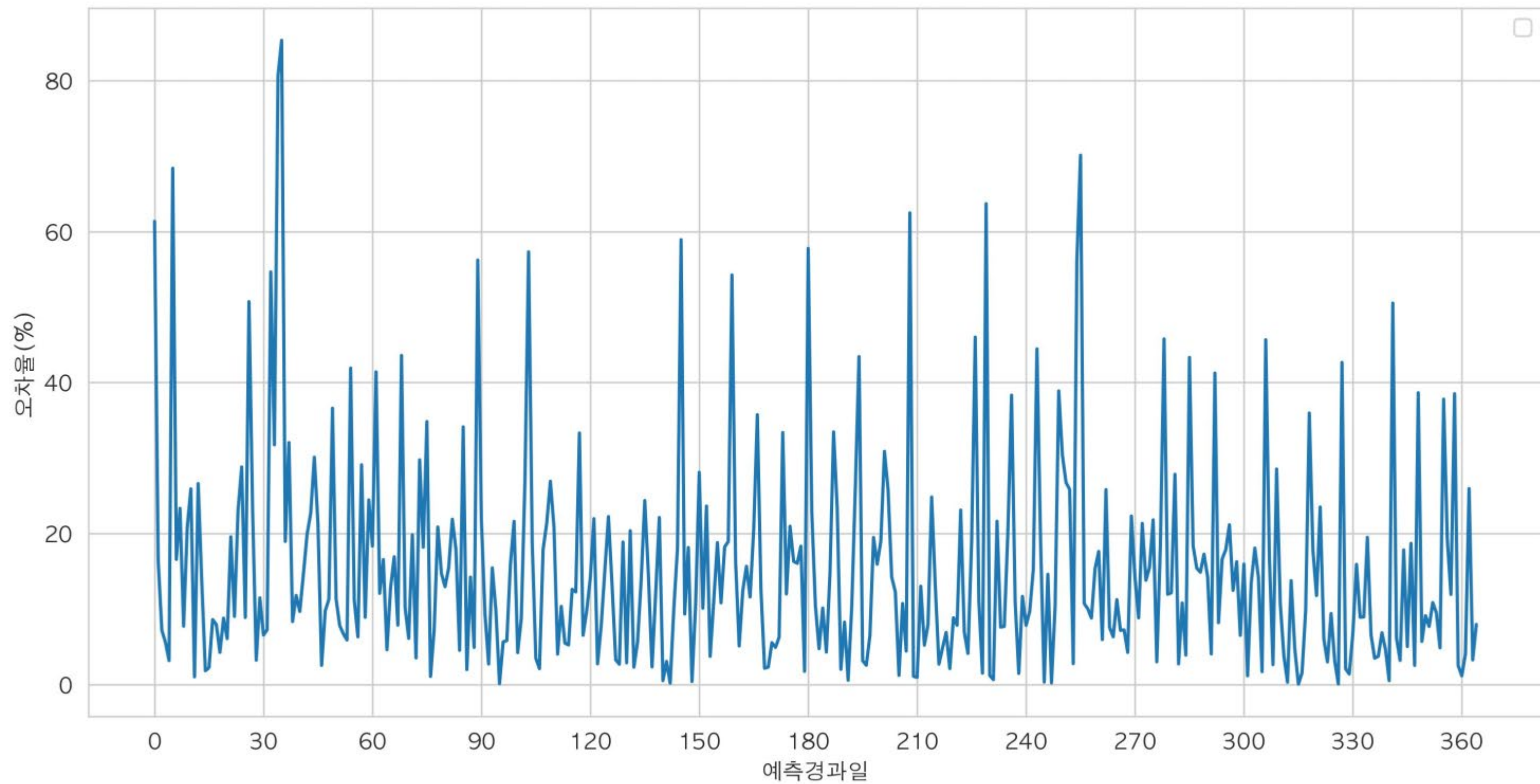


## EDA 및 머신러닝 분석





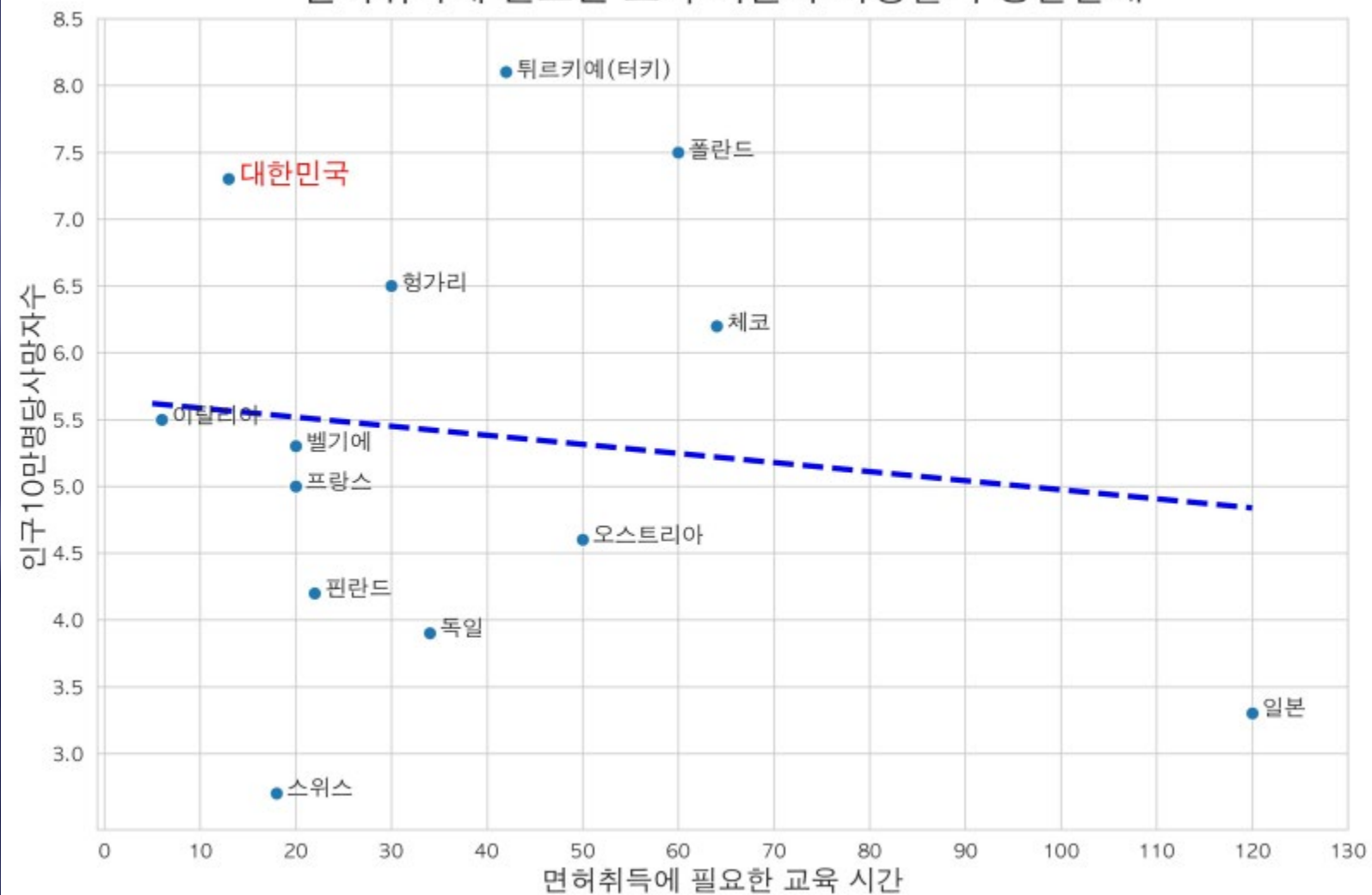
## EDA 및 머신러닝 분석



## 면허취득교육시간과 사망율의 상관관계 비교

## 면허 취득교육시간, 사망율

면허취득에 필요한 교육 시간과 사망율의 상관관계



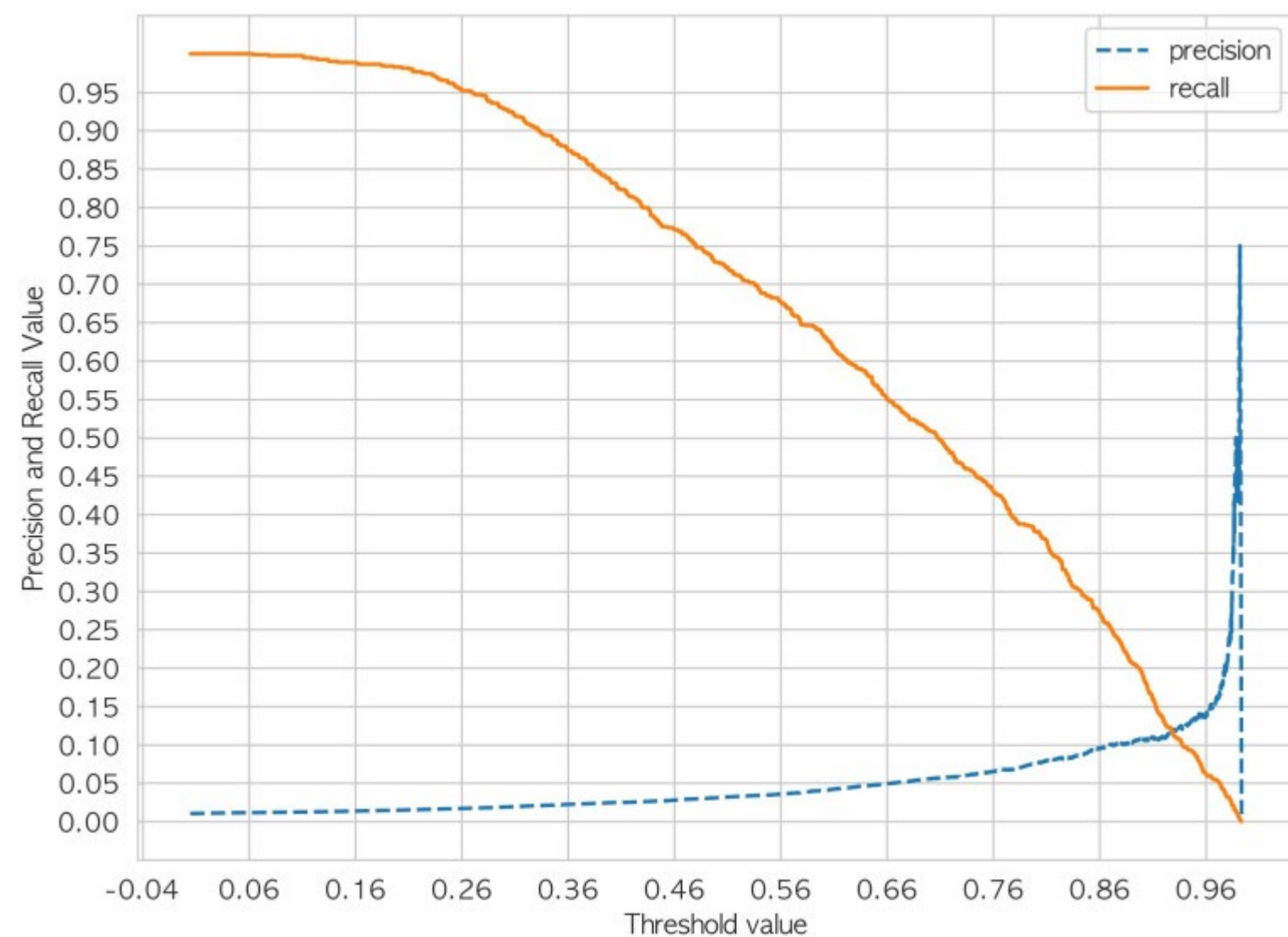
면허취득교육시간과 사망율의 상관 계수 : -0.09

13개국 면허취득 교육시간 평균 : 38.38시간

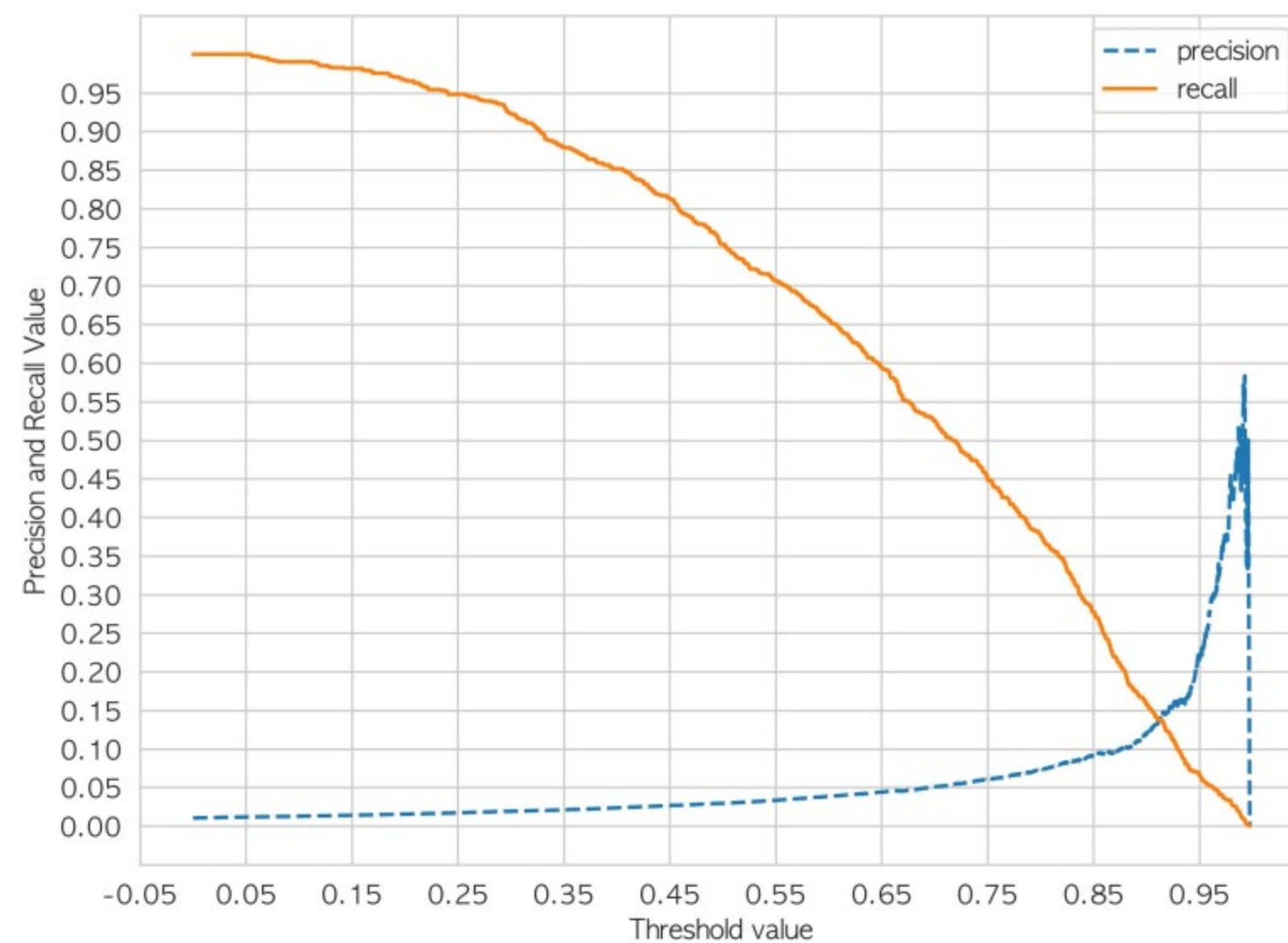
대한민국 면허취득 교육시간 : 13시간

## IMBALANCE CLASSIFICATION: 전체사고중 사망사고 예측

PR-THRESHOLD XGBOOST그래프



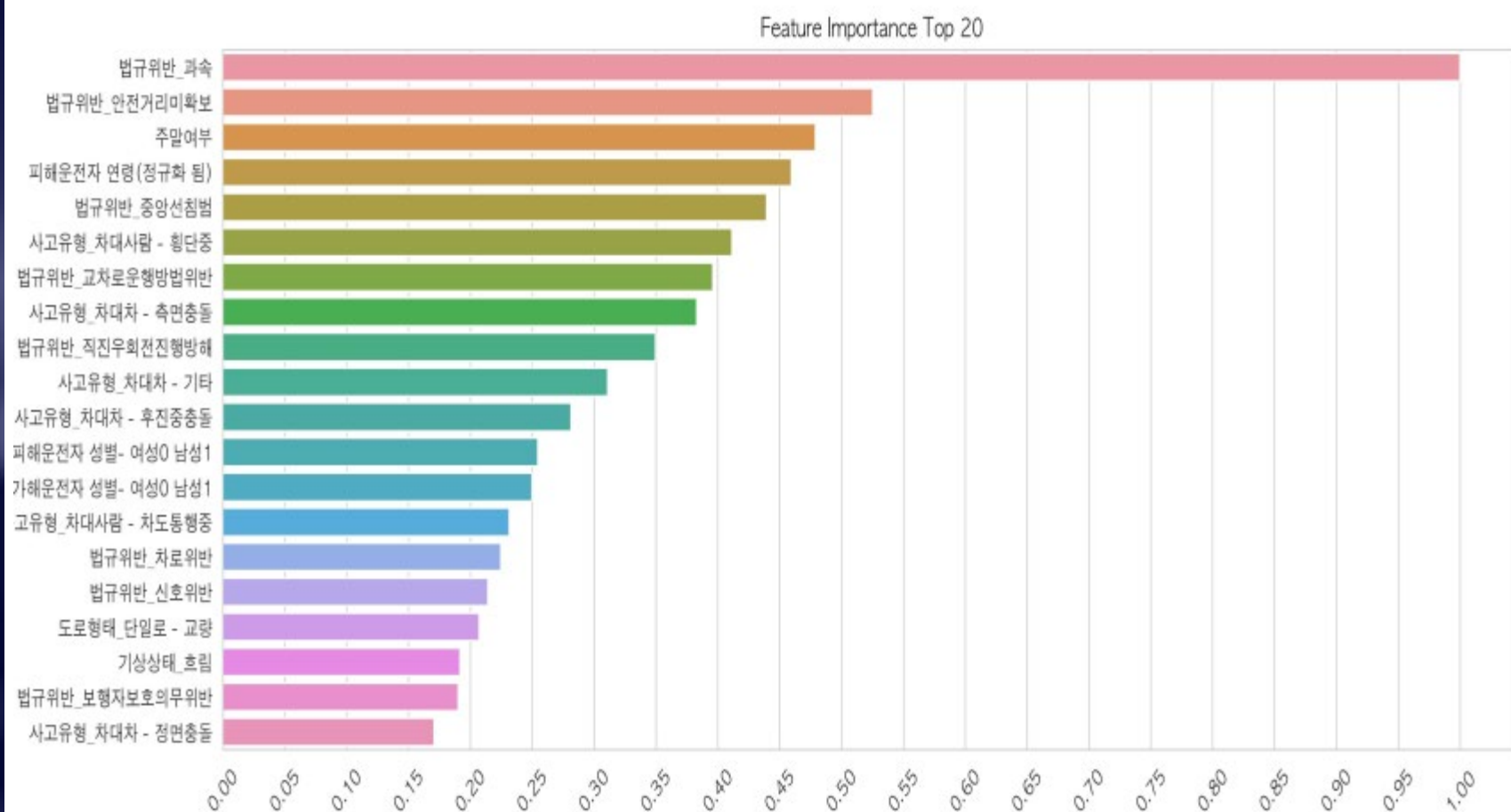
PR-THRESHOLD 로지스틱 회귀 그래프





## 한계 및 보완점

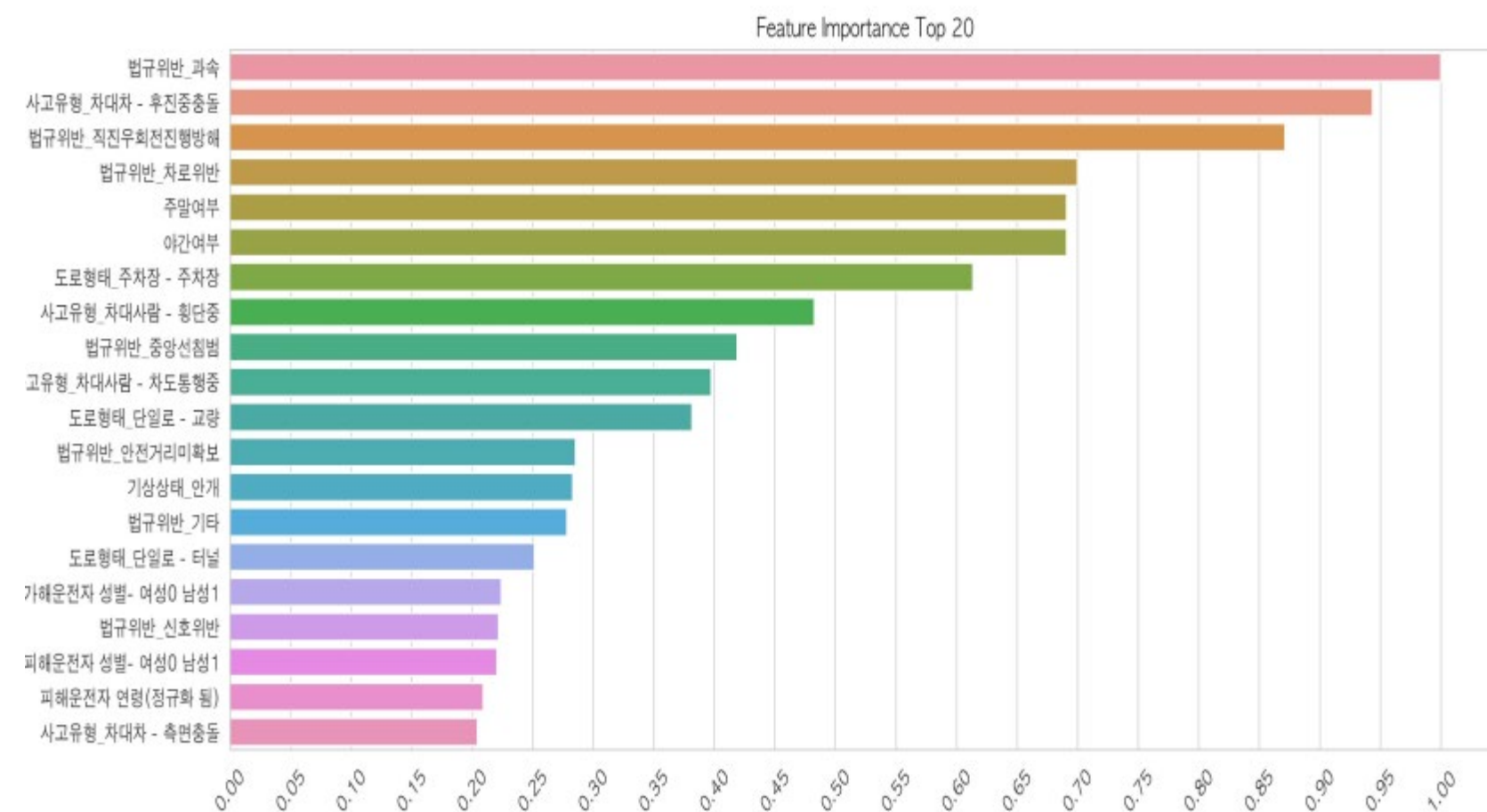
## IMBALANCE CLASSIFICATION: 전체사고중 사망사고 예측

FEATURE IMPORTANCE TOP20 -  
XGBOOST그래프

기준값

recall\_score = 0.961

precision\_score = 0.016

FEATURE IMPORTANCE TOP20 -  
로지스틱 회귀 그래프

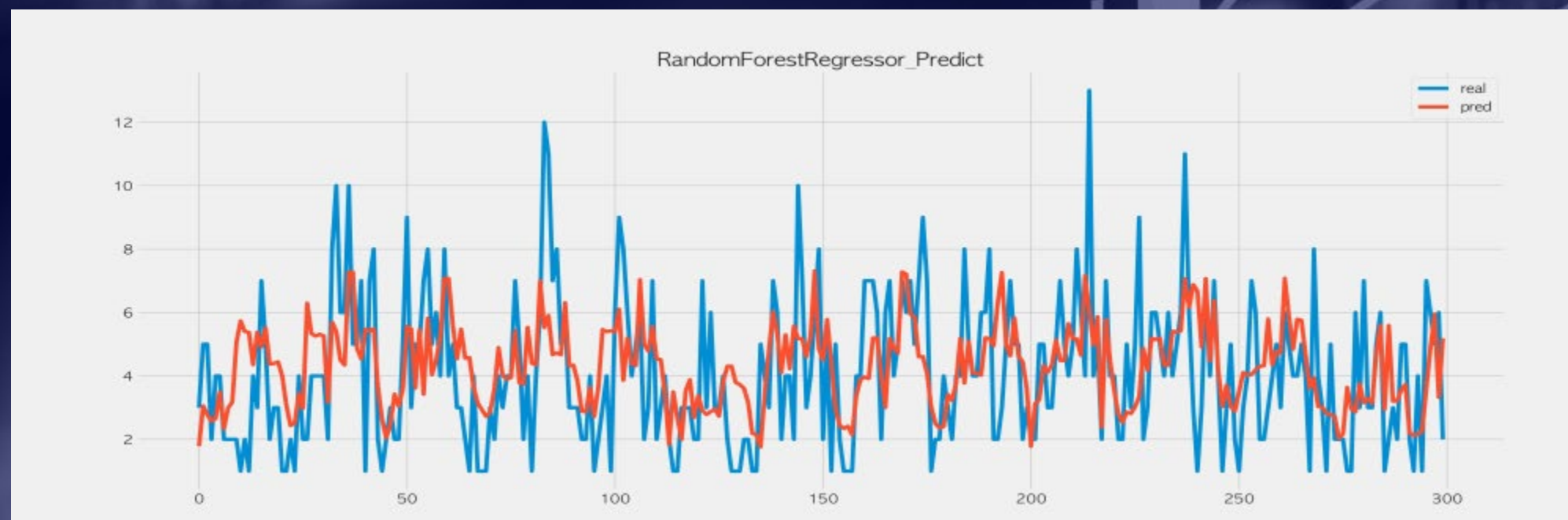
기준값

recall\_score = 0.948

precision\_score = 0.017



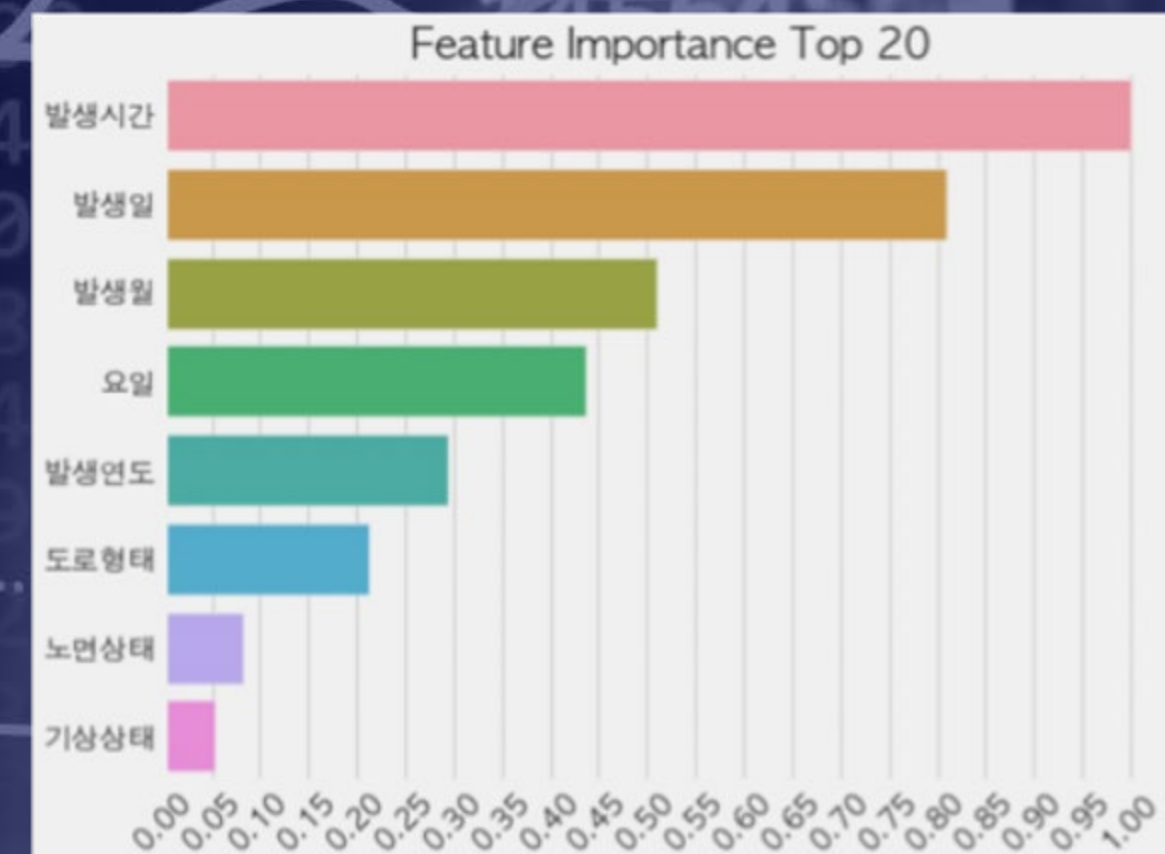
## RANDOMFOREST REGRESSOR

RANDOM FOREST  
REGRESSOR

테스트 데이터 MSE : 4.983  
 테스트 데이터 R<sup>2</sup> score : 0.351  
 오차율 평균 65.58%

FEATURE IMPORTANCE TOP20 -  
RANDOM FOREST

발생시간, 발생일, 발생월, 요일



## 프로젝트 의의



# Carrot





## EPILOGUE

### LESSONS LEARNED

우성균\_16 : 이 프로젝트에서 지금까지 배운 머신러닝 지식들을 응용하고, 모르는 딥러닝 지식들을 구글링해가면서 시행착오해나가는 과정이 즐거웠었고 뜻깊었습니다. IMBALANCED CLASSIFICATION의 성능을 만족할만큼 향상시키기엔 시간과 지식이 부족했던 점이 아쉬움으로 남았습니다. 조장을 처음 해봤는데 다른 조원분들이 열심히 해주셔서 프로젝트 기간 동안 즐겁게 임했던 것 같습니다.

김민철\_15 : 이 프로젝트를 통해 다양한 교통사고 통계 자료들을 EDA화 함으로써 교통사고가 발생하는 다양한 요인들을 알 수 있었습니다. 또한, ML을 시도해 보면서 다양한 경험을 할 수 있었던 프로젝트였습니다.

이성민\_15 : 교통사고를 발생시키는 요인들은 단순하지 않고 복잡하고 다양하다는 사실을 알았으며, 팀원들과 소통하며 다양한 각도에서 교통사고를 바라보고, 해석해 볼 수 있어 뜻깊은 시간이었습니다.

이유경\_16 : 수업 시간에 배운 걸 직접 해보는 과정에서 얻는 것이 많았고 개념이 정리됐습니다. 예측한 것과 다른 결과가 나오면 신기했고 다른 원인을 찾는 과정이 재미있었습니다.

안소연\_16 : 단순히 운전자의 미숙함만이 아니라 여러 가지 요소가 교통사고의 원인이 된다는 것을 알 수 있었고, 공부한 내용을 적용해 볼 수 있어서 많은 도움이 된 프로젝트였습니다.

## 자료 출처

### ML-전체사고중 사망사고 예측

- 도로교통공단 사고분석시스템 taas - 교통사고분석시스템

### ML-서울 교통사고 시계열 예측

- 공공데이터포털 도로교통공단\_서울시 일별 시간별 교통사고 현황

### ML - 서울 교통사고 건수 예측 (\*\*랜덤포레스트)

- 도로교통공단 사고분석시스템 taas - 교통사고 GIS

### EDA - 해외비교

- OECD 국가별 사고(2018년), 도로교통공단 사고분석시스템 taas

### EDA - 국내 법규위반과 사고율 분석

- 국내 도로형태별 교통사고(2000~2020년), 도로교통공단 사고분석시스템 taas

### EDA - 연령별 운전자 사고율 분석

- 서울시 연령별 교통사고 발생 현황(2020년), 서울 열린 데이터 광장

### EDA - 시간대별 사고율 분석

- 전국 시간대별 교통사고 사망률(2020년~2021년), 도로교통공단 사고분석시스템 taas

### EDA - 기후

- 도로종류별 기상상태별 교통사고(2017~2021), 통계청





Q&A



The background of the slide is a dark blue-grey color, featuring a complex pattern of overlapping geometric shapes like squares and rectangles, some of which are filled with a halftone dot pattern. A large, light blue-grey arrow points from the top right towards the center of the slide.

**감사합니다**