

# 데이터 분석 결과보고서

---

서울시 음주운전 사고 분석 및 예측



## 1. 개요

### 분석/시각화 목적

- 서울시 음주운전 사고 가능성 예측
- 음주운전 관련 정책 제안(단속 거점/시간 및 날짜(기간) 제안)

### 목적 및 필요성

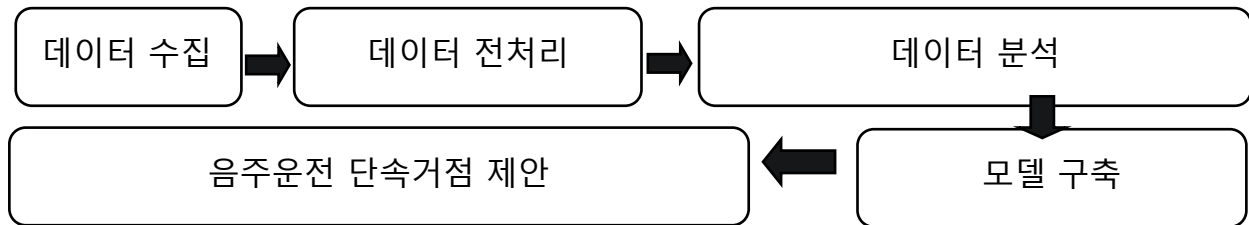
- 서울시 음주운전 사고 분석 목적
  - 서울시는 매년 음주운전 사고를 막기 위해 단속을 강화하고 음주운전 예방 교육 프로그램을 운영하고 있음에도 불구하고 매년 사망자가 발생하는 사고 유형 중 하나다. 본 프로젝트는 2021-2023 년 데이터를 기준으로 음주운전의 실태를 파악하고 사고에 영향을 주는 요인들을 분석함으로써 음주운전 사고 발생 지역을 예측하고자 한다.
- 서울의 음주운전 사고 심각성
  - 2022 년 시도별 음주운전 교통사고 건수를 비교한 결과 서울이 약 2400 건으로 전국 2 위이다.
  - 서울시는 인구수 대비 사고 발생 비율이 전국 하위 3 위임에도 불구하고 서울시 강남구는 전국 2 위를 기록하고 음주운전 사고 다발 지역 수가 20 곳으로 평균 9 곳임을 비교하면 약 2 배이상의 결과가 도출되었다.
  - 대한민국에서 높은 인구 밀집도와 교통량 보유한 도시로 음주운전 사고 발생 시 더 많은 인명 피해와 교통 혼잡을 유발할 수 있다.
  - 타 도시에 비해 야간 경제활동이 활발하므로 강남, 홍대 등 유흥 지역을 중심으로 음주운전 사고가 빈번하게 발생할 수 있다.
  - 시군 구별 자동차등록대수가 전국 1 위로 확인되었다.
- 기존 교통환경의 한계
  - 서울시 열린 데이터 광장에서 2021-2022 년 서울시 교통불편 민원신고 현황을 확인한 결과 택시의 승차거부가 2022 년 2921 건으로 2021 년 1464 건에 비교하여 약 2 배 상승한 것으로 나타났다. 또한 버스 관련 민원신고도 2021 년 4856 건에서 2022 년 5755 건으로 상승한 것으로 확인되었다.

=>음주운전 사고 비율이 서울시 강남구가 가장 높은 것을 확인하여 서울시 구별로 음주운전 사고 건수를 매핑하여 위치를 확인할 것이다.

=>서울시 내 음주운전 사고 건수 상위 5 지역을 선정해 어느 시간, 요일, 시간대에 발생하는지 확인 및 도로환경, 주류 인허가 음식점 개수, 교통안전지수 등 음주운전 사고에 미치는 변수를 확인할 예정이다.

## 분석/시각화 프로세스

### 프로젝트 로드맵



#### ● 수집 데이터

- 도로교통공단 데이터
- 경찰청 데이터
- 서울시 열린데이터

#### ● 전처리 과정

- 도로교통공단 데이터에서 2022 년 교통사고 사고유형별 건수 데이터 추출해 사고유형 빈도를 시각화
- 도로교통공단 자료를 활용하여 2021-2023 년 서울시 음주운전 사고 발생 사고시, 사고요일, 시간대, 월별을 시각화
- 경찰청 2022 년 알코올농도, 성별, 나이, 음주운전 적발건수 데이터에서 알코올 농도 별 음주운전 발생건수를 시각화
- 서울시 열린데이터 내 서울시 일반음식점 인허가 정보 데이터를 활용하여 영업중인 음식점 및 음식점 위치를 구별, 동별로 필터링

#### ● 분석 데이터

- 시군 구별 시간대 별, 요일 별, 월 별 음주운전 교통사고 수
- 서울시 일반음식점 인허가 정보
- 음주운전 다발구역
- 서울시 구별 인구수
- 서울시 음주운전 사고 지역 도로환경, 교통안전지수
- 전국 교통사고 유형

#### ● 분석 방법

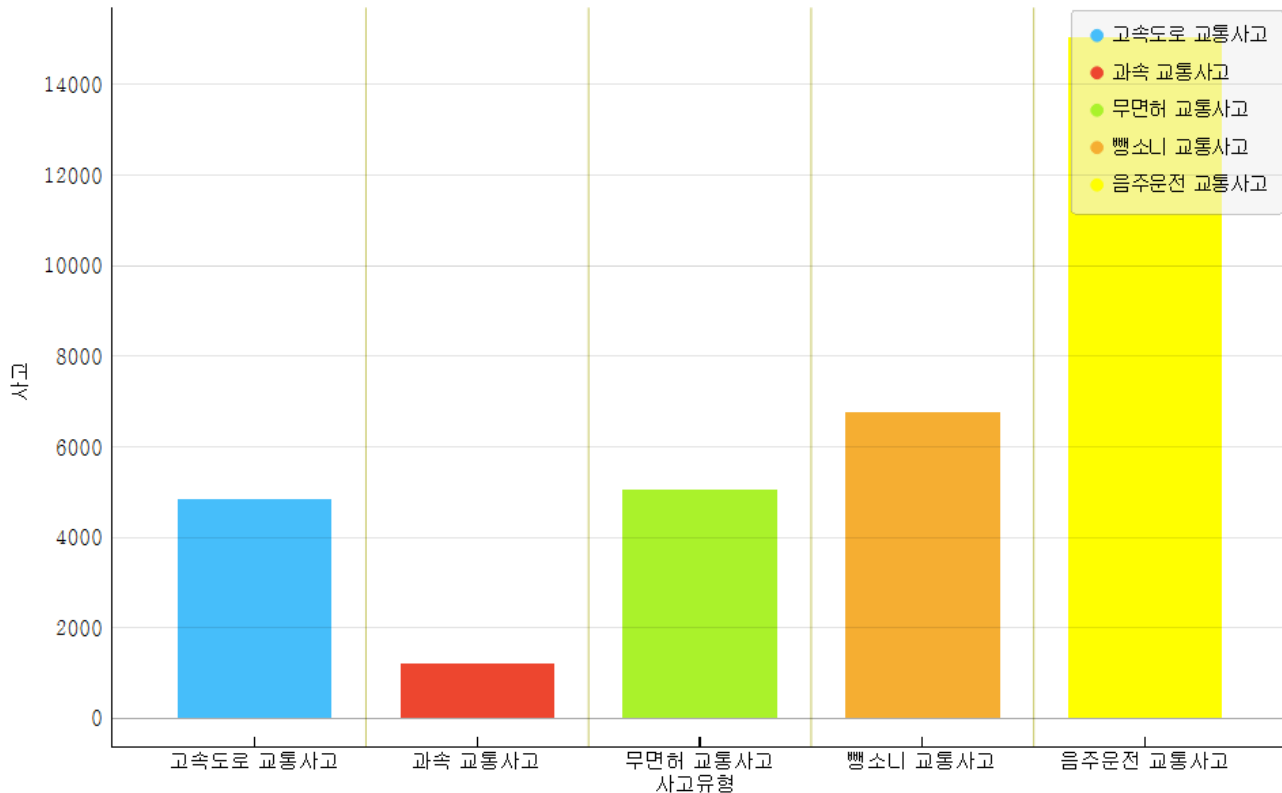
- Python: 회귀분석(Decision Tree, Random Forest, XGboost, SVR), 상관분석(Heatmap)
- RStudio: 데이터 시각화(ggplot2)
- Orange

## 2. 분석/시각화 결과 상세내용

### 분석/시각화 상세내용

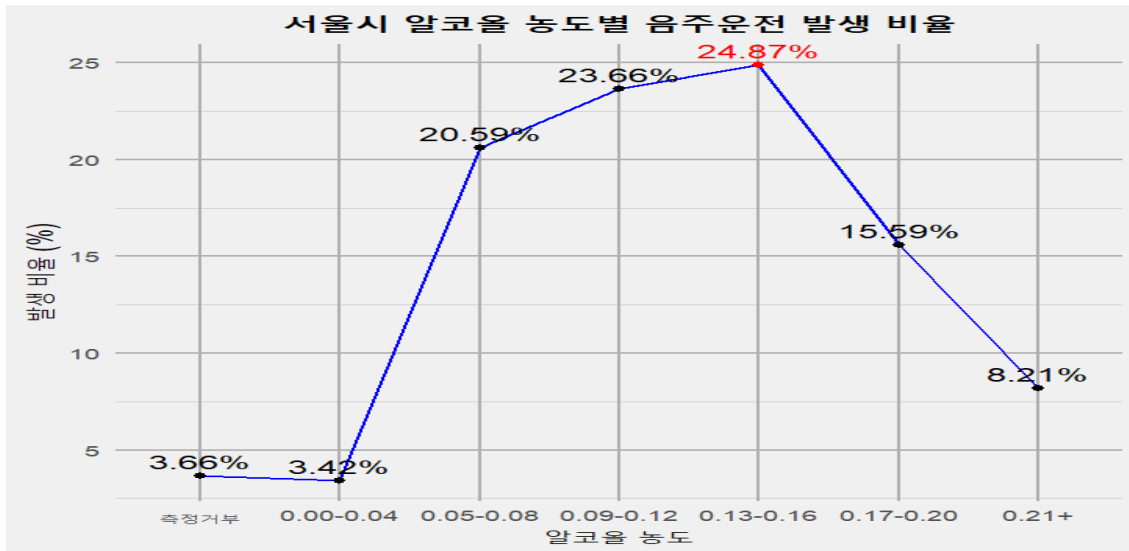
#### ● 데이터 수집

## 교통사고 유형



- 도로교통공단 데이터에서 2022 년 교통사고 사고유형별 건수 데이터 추출하였다.
- 교통사고 유형 중 음주운전이 1 등으로 약 15000 건이 발생함. 하위 4 개의 유형을 합친 값과 음주운전 건수의 값이 비슷하므로 음주운전의 발생 건수가 상당한 것으로 확인되고 있다. 본 연구를 통해 음주운전에 미치는 요인을 분석하고, 음주운전 사고 지역 상위 5 곳을 예상하고자 한다.

●탐색적 분석(EDA) 및 변수탐색

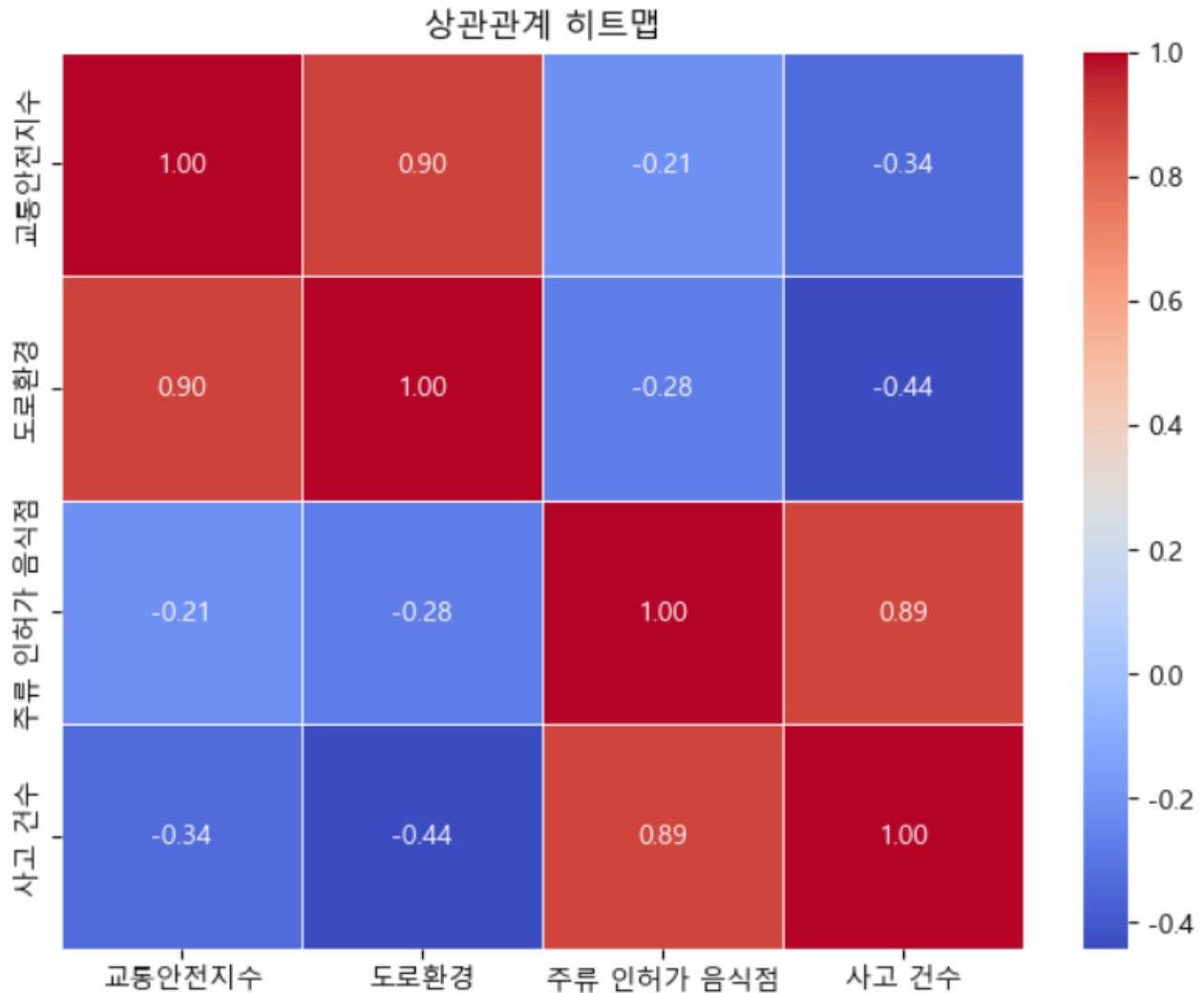


■ 2022 년 경찰청 음주운전 데이터에서 알코올농도를 범주화하고, 음주운전 발생건수를 추출하여 알코올 농도 별 음주운전 발생비율을 시각화 하였다.

■ 『도로교통법』 제 44 조 제 4 항에서 운전이 금지되는 술에 취한 상태의 기준이 운전자의 혈중알코올농도가 0.03% 이상인 경우로 규정하고 있다. 그래프에서 측정거부와 알코올 농도 0~0.04%를 제외한 수치가 약 93%로 도로교통법에 기준이 되는 0.03% 이상이 상당 부분을 차지하고 있다. 본 연구를 통해 음주운전에 미치는 요인을 분석하고 서울시 지도를 음주운전 건수 및 음주운전에 미치는 요인으로 시각화 하고자 한다.

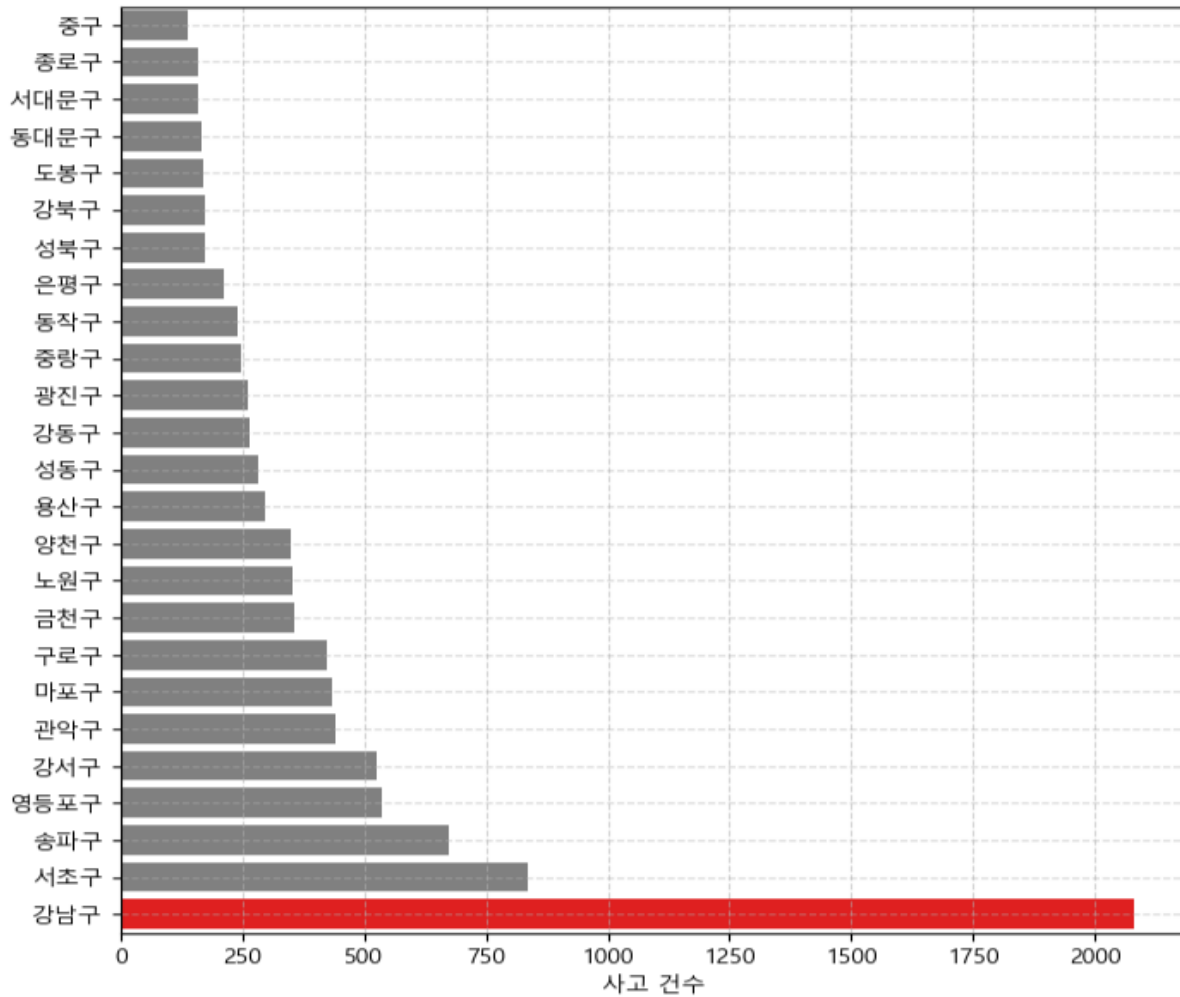
## ● 상관관계

	교통안전지수	도로환경	주류 인허가 음식점	사고 건수
교통안전지수	1.0	0.8958988907793200	-0.209492435198395	-0.34044009509147200
도로환경	0.8958988907793200	1.0	-0.27816051324162700	-0.44268832831030300
주류 인허가 음식점	-0.209492435198395	-0.27816051324162700	1.0	0.8876303341641190
사고 건수	-0.34044009509147200	-0.44268832831030300	0.8876303341641190	1.0

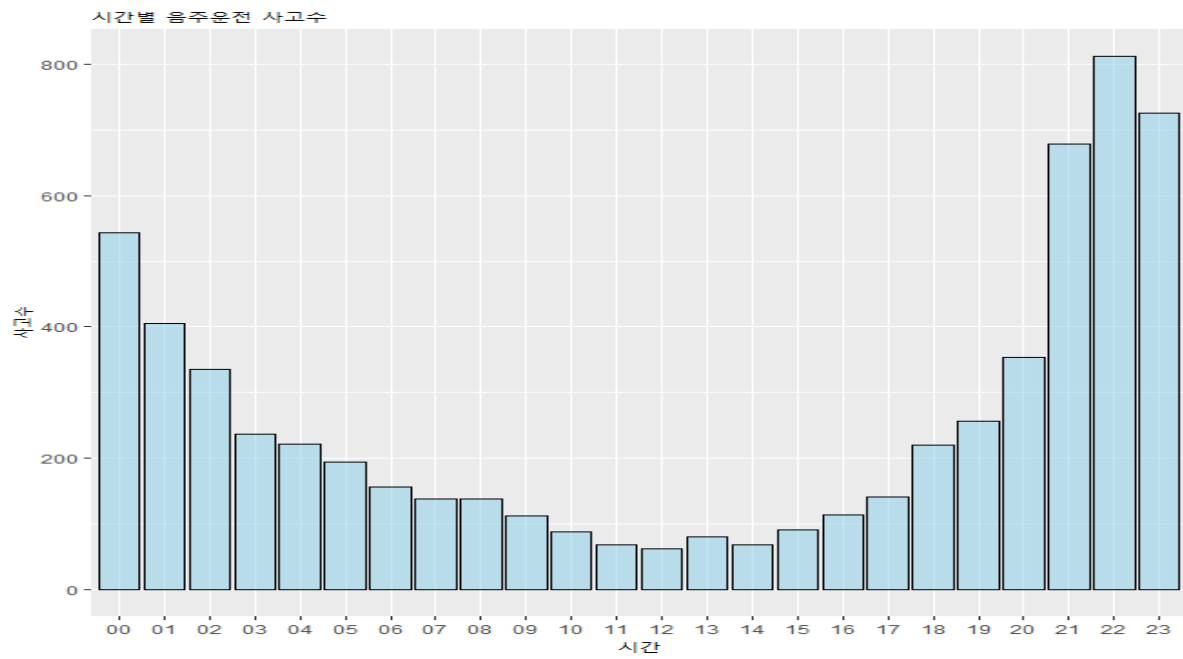


- 교통안전지수와 도로환경은 사고 건수를 줄이는 데 중요한 역할을 하는것으로 나타났다. 교통안전지수와 도로환경이 우수할수록 사고 건수가 줄어드는 (음의상관관계)경향 있다.
- 주류 인허가 음식점의 수는 사고 건수와 매우 강한 양의 상관관계를 가지므로, 이러한 지역에서 음주운전 단속을 강화하는 것이 사고를 줄이는 데 효과적일 수 있다.

구별 음주운전 사고 건수



- 2021~2023 년 전국 구별 시간대별 음주운전 사고건수에서 추출하여 분석하였다. 데이터만 필터링하여 서울시 구별 음주운전 발생건수를 시각화하였다.
- 서울시 강남구에서 약 898 건이 넘는 음주운전 발생. 뒤를 이어 서초구 송파구가 약 438 건의 음주운전이 발생하는 것을 알 수 있다. 이는 유흥가 및 상업지구 밀집 지역이 사고 다발 지역임을 의미한다.

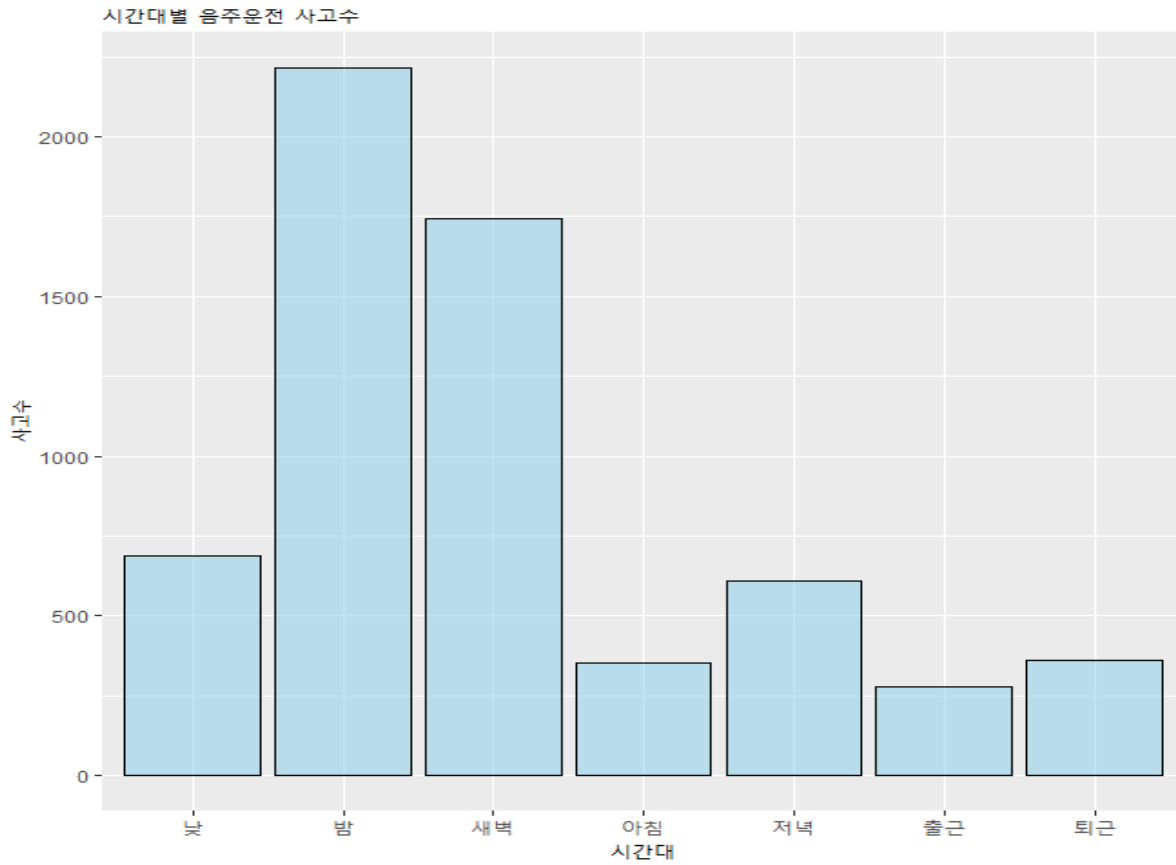


■ 시간대에 주기성이 존재함을 확인할 수 있다.

■ 심야 시간대 (22 시~02 시)에 사고 발생이 집중되었다.

이는 술자리 후 이동하는 시간대에 사고가 많이 발생하는 패턴을 보인다.

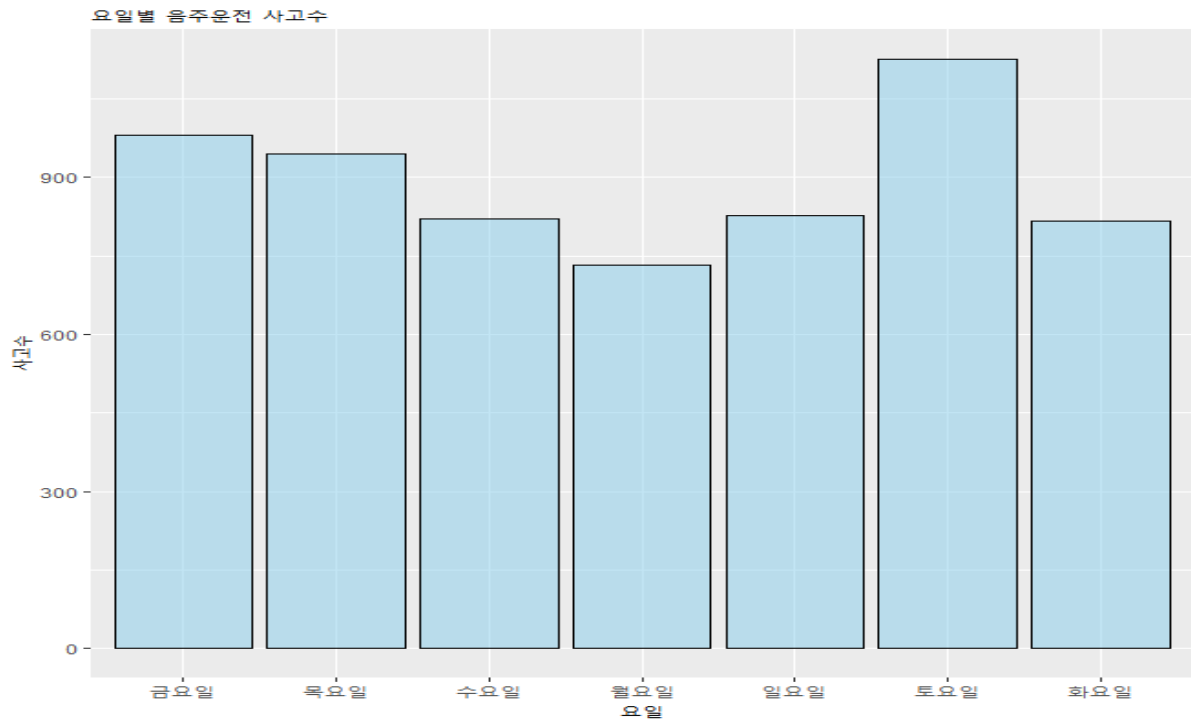




■ 2021~2023 년 서울시 시간대별 음주운전 사고건수 데이터

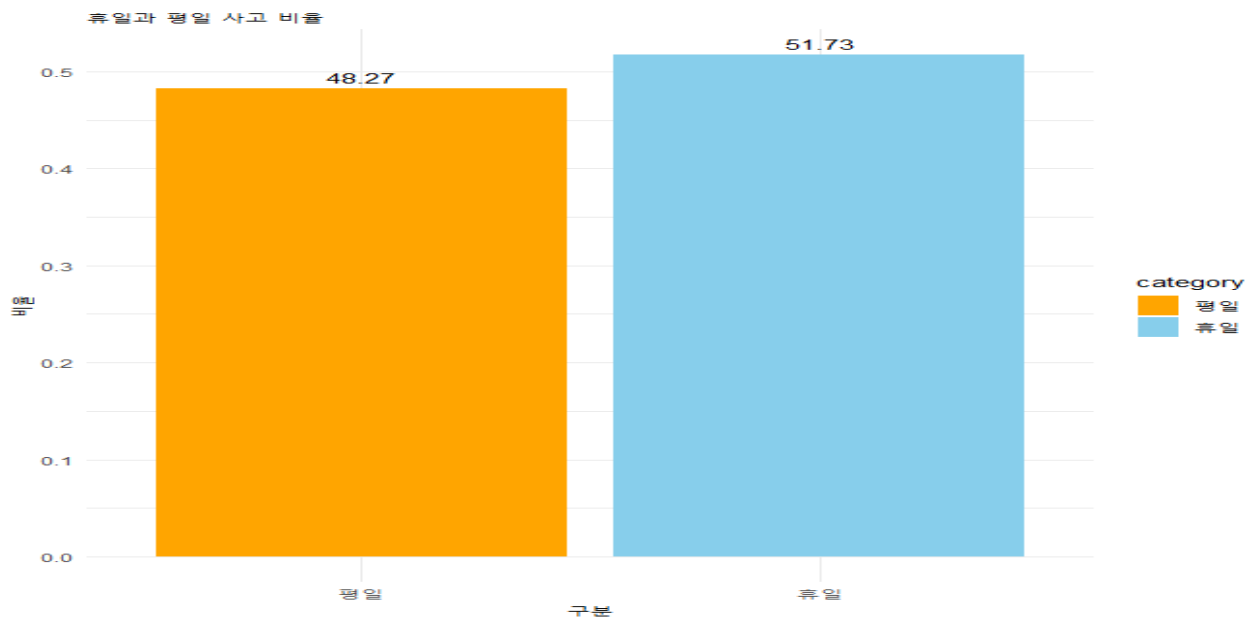
■ 밤과 새벽 시간의 음주운전 발생 빈도가 높았고, 아침과 출근시간대에는 가장 적었음을 확인할 수 있었다.

새벽: 0 시 ~ 5 시  
 아침: 5 시 ~ 7 시  
 출근: 7 시 ~ 9 시  
 낮 : 9 시 ~ 17 시  
 퇴근: 17 시 ~ 19 시  
 저녁: 19 시 ~ 21 시  
 밤 : 21 시 ~ 24 시



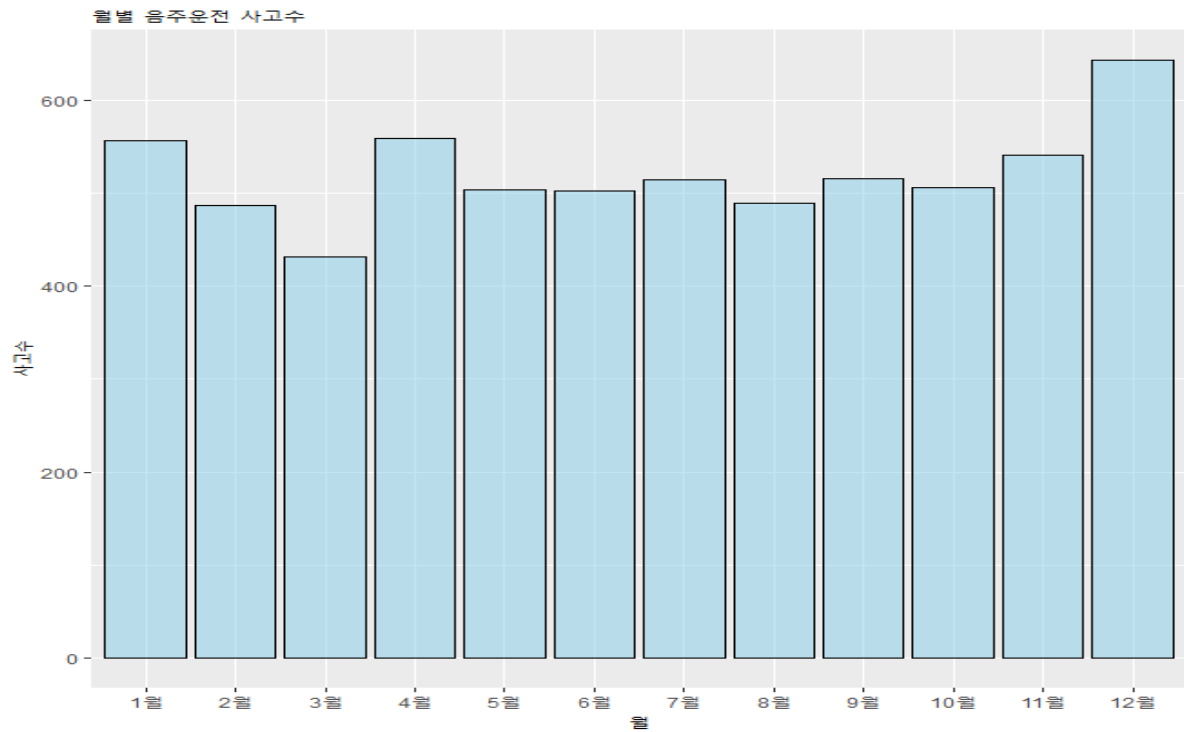
■ 2021~2023 년 서울시 요일별 음주운전 발생건수 데이터

■ 서울시 음주운전 발생 요일은 목요일에 증가해 월요일에 감소하는 추세를 보였다.



■ 2021~2023 년 서울시 요일별 음주운전 발생건수 데이터에서 월요일~금요일을 평일로, 토요일, 일요일, 공휴일을 휴일로 묶어 발생건수를 비율로 시각화하였다.

■ 서울시 평일(월-금)과 휴일(토,일,공휴일)을 비교한 결과 근소하게 휴일 비율이 높았다.



■ 2021~2023 년 서울시 월별 음주운전 발생건수 데이터 연말인 12 월 빈도가 높았고, 3 월이 가장 적었다. 나머지 월들은 비슷하게 나타났다.

## 결과 해석

### ● 요인분석

■ 요일: 토요일, 금요일, 목요일이 상위 3 위 안에 들고 있다.

■ 시간: 22 시, 23 시, 21 시가 상위 3 위 안에 들고 있다.

■ 시군구: 토요일 0 시 강남구 신사동, 토요일 1 시간 서초구 서초동, 수요일 22 시 강남구 논현동 이 상위 3 위 안에 들고 있다.

요일	사고건수
토요일	1126
금요일	980
목요일	945

사고시간	사고건수
22	812
23	726
21	678

시군구	요일	사고시간
서울특별시 강남구 신사동	토요일	0
서울특별시 서초구 서초동	토요일	1
서울특별시 강남구 논현동	수요일	22

Subcategory\_Frequencies

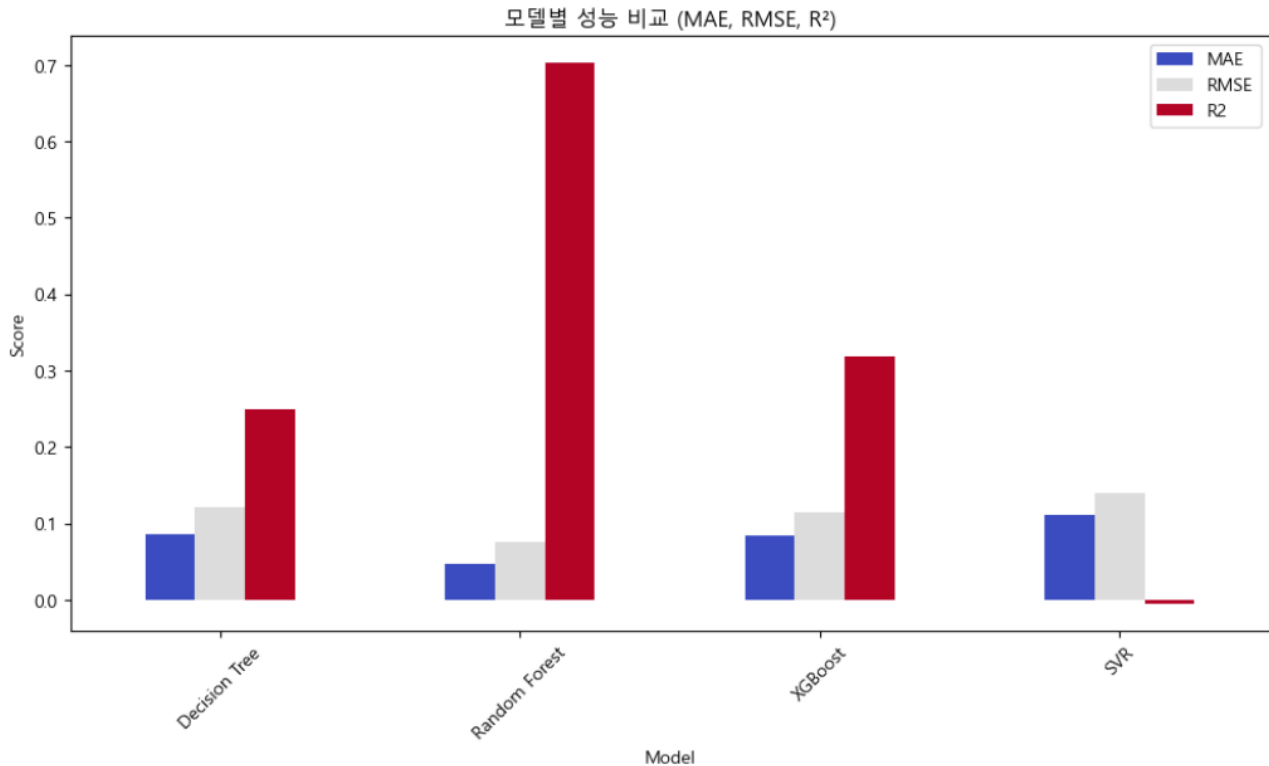
소분류	빈도
역삼동	3183
신림동	2404
신사동	2311
봉천동	2279
화곡동	2247
서초동	2178
구로동	2105
논현동	2060
서교동	1871
여의도동	1721
상계동	1667
마곡동	1478
수유동	1466

=> tableau 지도 시각화한 결과와 더해, 강남구 신사동 도산공원사거리 부근에서 토요일 0 시경, 서초구 서초동 예술의 전당 앞 삼거리부근에서 토요일 1 시경, 강남구 논현동 한국산업경제연구원 부근에서 수요일 22 시경 집중적인 단속 활동을 펼치는 것이 효과적일 것이라 판단된다.

## 음주운전 사고 발생 가능성 예측

	Model	MAE	RMSE	R <sup>2</sup>
0	Decision Tree	0.084977	0.120797	0.248559
1	Random Forest	0.046626	0.075963	0.702846
2	XGBoost	0.084134	0.115037	0.318512
3	SVR	0.111723	0.139736	-0.005535

<Figure size 1200x600 with 0 Axes>

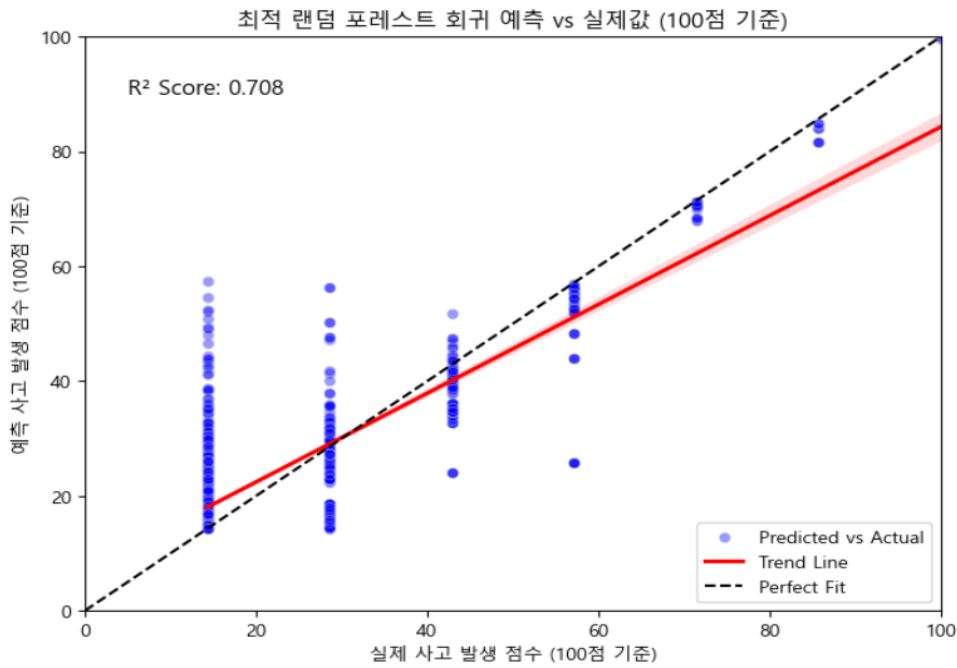


사고요일, 사고시, 시군구 변수를 이용하여 모델 학습을 진행하였다.

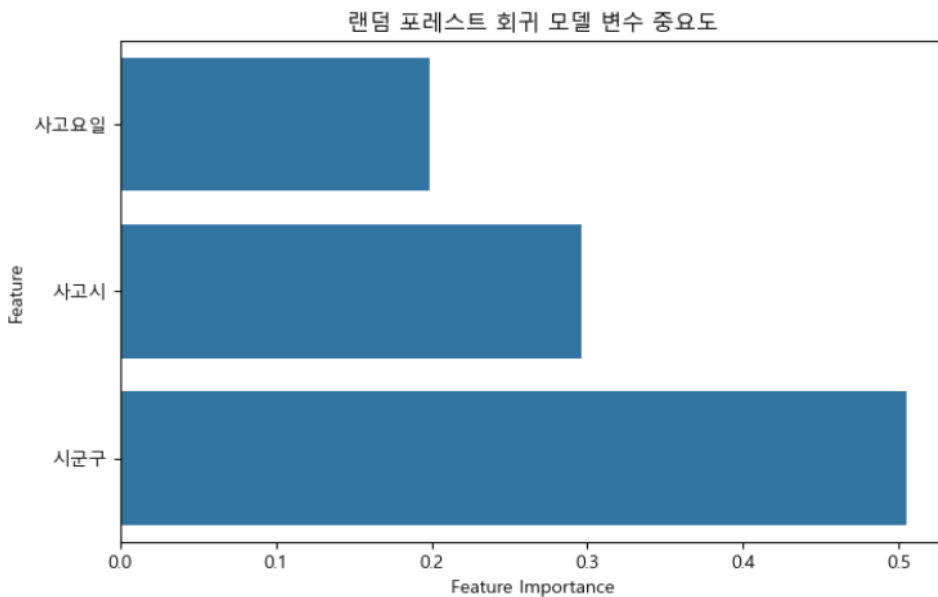
음주운전 사고 발생 가능성을 예측하기 위해 Decision Tree, Random Forest, XGBoost, SVR 의 회귀모델의 성능을 비교한 결과이다. MAE(평균 절대 오차), RMSE(평균 제곱근 오차), R<sup>2</sup>(결정 계수)를 기준으로 성능을 평가했다.

Random Forest 회귀모델의 R<sup>2</sup>값이 가장 높고 MAE 와 RMSE 가 가장 낮아 우수한 성능을 보였다. Decision Tree 는 보통 과적합 위험이 있지만, Random Forest 모델은 여러 트리를 조합해 안정적인 예측이 가능하다. XGBoost 도 강력한 모델이지만, 이 데이터셋에서는 Random Forest 보다 R<sup>2</sup>가 낮아 최적의 선택이 아니었다.

따라서 Random Forest 회귀모델을 최종 선택하였다.



다음은 실제값과 예측값간의 차이를 그래프로 나타낸 것이다. 결정계수는 0.708 로 우수한 성능을 보였지만, 정확한 예측값(대각선)과 일부 차이가 나는 구간도 존재했다.



사고요일, 사고시간, 시군구 총 3 개의 변수를 넣어 RandomForest 모델을 학습시켰다. 위 3 가지 변수중에서 시군구 변수가 회귀모델에 가장 큰 영향을 미치는 요인으로 나타났다. 따라서 시군구 정보에 따라 강남구 신사동, 청담동, 논현동 중심의 단속 전략이 필요하다.

```
def predict_accident_score(day, hour, district):
    day_encoded = label_encoders['사고요일'].transform([day])[0]
    hour_encoded = label_encoders['사고시'].transform([hour])[0] # "00시"도 변환 가능
    district_encoded = label_encoders['시군구'].transform([district])[0]

    input_data = pd.DataFrame([[day_encoded, hour_encoded, district_encoded]],
                               columns=['사고요일', '사고시', '시군구'])

    probability = model.predict(input_data)[0] # 사고 발생 확률 예측 (0~1)
    score = probability * 100 # 🚗 100점 기준 변환
    return score

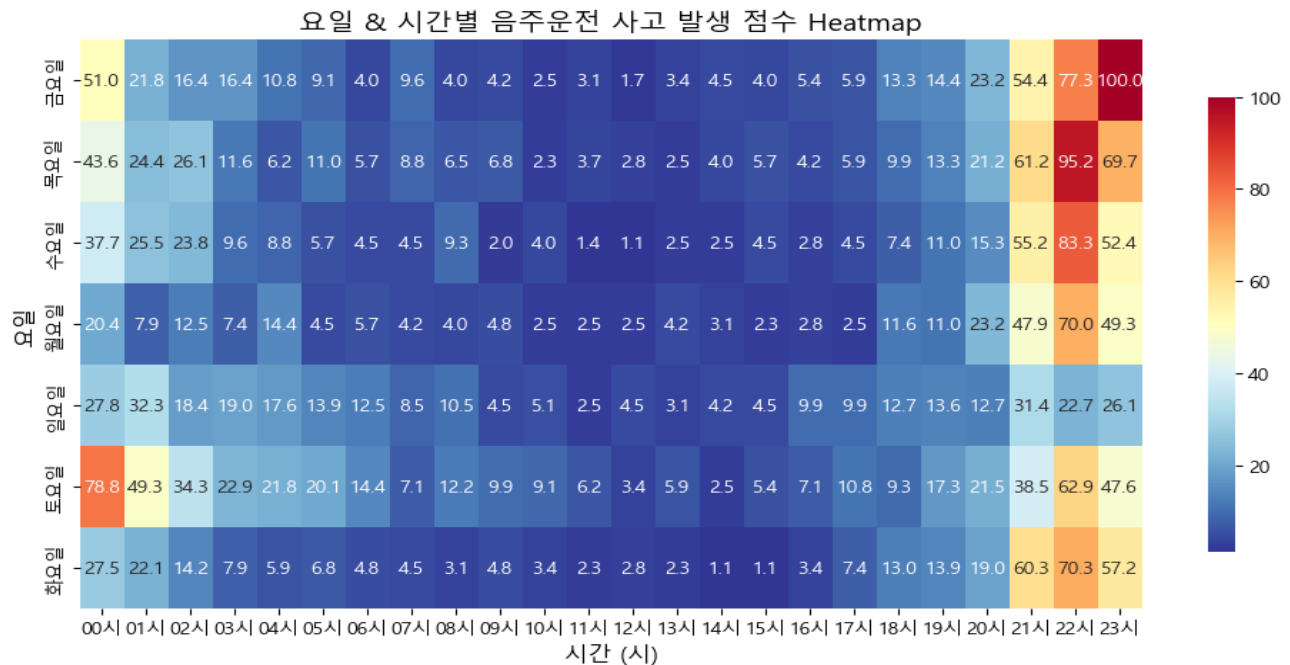
# 특정 조건에서 사고 발생 점수 예측
day = '목요일'
hour = '23시'
district = '서울특별시 강남구 청담동'

score = predict_accident_score(day, hour, district)
print(f'🚗 {day} {hour} {district}에서 사고 발생 점수: {score:.2f}')
```

🚗 목요일 23시 서울특별시 강남구 청담동에서 사고 발생 점수: 83.76

하단의 코드 블록은 특정 조건(요일, 시간, 지역)에 따라 사고 발생 확률을 예측하는 함수이다. 사용자가 요일, 시간, 지역 정보를 입력하면 모델이 해당 조건에서의 사고 발생 점수를 반환한다. 100 점만점 기준으로 점수가 클 수록 사고 발생 가능성이 높다는 의미이다.

아래 그림은 요일과 시간별 음주운전 사고 발생 점수 상관관계이다.



금요일과 토요일 야간(22 시~02 시) 에 음주운전 사고 발생 점수가 가장 높았다.

이는 술자리 후 귀가하는 인구 증가와 관련이 있는 것으로 보인다.

따라서 해당 시간대 집중 단속 강화, 심야 대중교통 확충, 대체 교통수단 제공 등의 정책이 필요하다.

■ 최종 결과

구	동	요일	사고시	정규화 사고 접수	구체적 위치 추천
강남구	신사동	토요일	00 시	100	신사역사거리 부근
서초구	서초동	토요일	01 시	100	예술의전당앞삼거리 부근
강남구	청담동	목요일	23 시	73.5	청담스퀘어 부근
강남구	논현동	수요일	22 시	73.5	교보타워사거리 부근
강남구	역삼동	목요일	02 시	73.5	경복아파트교차로 부근
강남구	역삼동	수요일	00 시	73.5	센터필드교차로 부근
강남구	역삼동	수요일	23 시	73.5	르네상스호텔사거리 부근
강남구	신사동	금요일	23 시	61.2	신사오거리 부근
서초구	서초동	금요일	22 시	51	1370 부근
금천구	시흥동	목요일	21 시	51	시흥 4 동사거리 부근
강남구	청담동	금요일	22 시	51	청담사거리 부근
서초구	서초동	금요일	00 시	51	신사역사거리 부근
관악구	신림동	토요일	23 시	51	난곡사거리 부근
강남구	삼성동	화요일	22 시	51	강남구청역 교차로 부근
강남구	역삼동	월요일	22 시	51	개나리아파트교차로 부근
강남구	청담동	토요일	00 시	51	유니팜약국 부근
관악구	신림동	월요일	00 시	51	난곡사거리 부근
강남구	역삼동	금요일	23 시	51	르네상스호텔사거리 부근
송파구	잠실동	월요일	22 시	51	잠실역사거리 부근
강남구	삼성동	화요일	21 시	51	KTF 타워 부근
강남구	논현동	토요일	00 시	51	한국산업경제연구원 부근
강남구	논현동	수요일	00 시	51	한국건설기술회관 부근
강남구	신사동	수요일	01 시	51	성수대교남단교차로 부근
성동구	성수동 1 가	금요일	23 시	51	성수대교 부근
강남구	역삼동	목요일	22 시	51	차병원사거리 부근
영등포구	여의도동	금요일	23 시	51	여의도공원앞 부근
강남구	청담동	목요일	22 시	51	영동대교남단교차로 부근
서초구	서초동	목요일	22 시	51	예술의전당앞삼거리 부근
송파구	가락동	화요일	22 시	32.7	가락시장역교차로 부근
강남구	논현동	토요일	01 시	32.7	거평타운오피스텔 부근



본 연구에서는 서울특별시 내에 음주운전 사고 발생 가능성이 가장 높은 지역을 분석하여, 각 구별로 가장 위험성이 높은 지역을 하나씩 선정하였다. 이를 통해 구별 주요 사고 위험 지역을 파악하고, 해당 지역에서의 교통안전 대책 수립에 참고할 수 있도록 하였다. 특정 구(강남구, 서초구, 관악구 등)에서 사고 발생 가능성이 높은 지역이 확인되었으며, 구마다 사고가 집중되는 주요 시간대와 요일이 존재함을 확인할 수 있었다. 주로 심야(22~24 시) 및 새벽 시간대에 사고 발생 가능성이 높았다.

#### 기대효과

- 음주운전 발생 빈도가 높은 지역의 구체적 장소와 시기를 분석하여 음주운전 단속 추천지 목록을 만들었다.
- 사고 발생 가능성이 높은 장소와 시간대에 집중적으로 단속하여, 음주운전을 효과적으로 단속할 수 있으며, 음주운전 사고 발생 가능성을 줄일 수 있다.
- 본 연구에서 도출된 음주운전 단속 추천지 분석 결과는 향후 음주운전 예방 정책 수립을 위한 기초 자료로 활용될 수 있으며, 이는 음주운전 사고 발생 예방에 기여할 수 있으리라 기대된다.

### 3. 기타

#### 건의 사항

- 2021-2023 년 서울시 음주운전 사고 지역 위도, 경도 데이터 필요
- 2021~2023 년 서울시 음주운전 사고 지역 지번주소

## 활용 데이터 및 참고 문헌 출처

### ■ 도로교통공단 데이터

2021~2022 년판 시군구별 등록인구

2021~2022 년판 시군구별 자동차등록대수

2022 년판 부문별 교통사고(과속, 음주운전, 무면허, 뺑소니, 고속도로)

음주운전 사고다발지역정보

2022 년 전국 구별 음주운전 발생건수(요청데이터)

2022 년 전국 구별 시간대별 음주운전 사고건수(요청데이터)

2022 년 전국 구별 요일별 음주운전 발생건수(요청데이터)

2022 년 전국 구별 월별 음주운전 발생 데이터(요청데이터)

### ■ 경찰청 데이터

2022 년 음주운전 데이터(알코올농도, 음주운전 발생건수)

### ■ 서울시 열린데이터

서울시 주민등록인구(연령별/동별) 통계

서울시 일반음식점 인허가 정보

서울시 교통불편 민원신고 현황 통계

### ■ 이외 데이터

기획팀 홍젤리, "3 가지 기준으로 음주 운전실태 데이터 분석하기", NEWSJELLY, 2023.08.02,

<https://newsjel.ly/archives/newsjelly-report/data-storytelling/15276>.