

데이터 분석 결과보고서

서울시 음주운전 사고 분석 및 예측



1. 개요

분석/시각화 목적

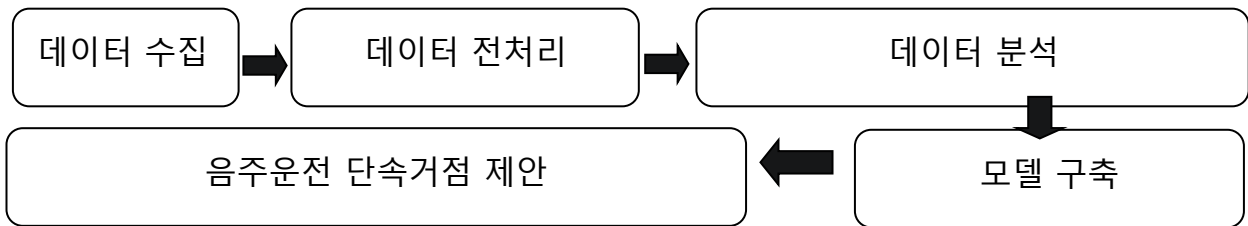
- 서울시 음주운전 사고 가능성 예측
- 음주운전 관련 정책 제안(단속 거점/시간 및 날짜(기간) 제안)

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

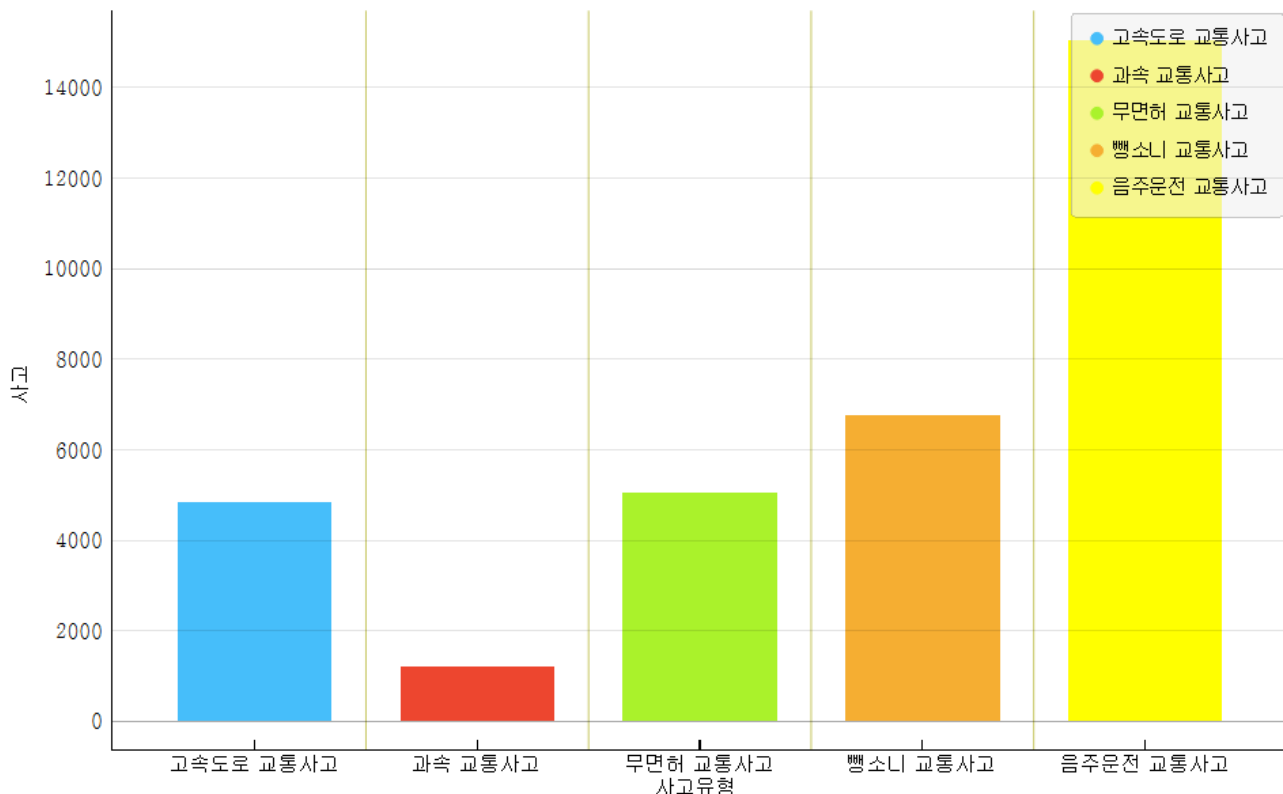
- Python: 회귀분석(Decision Tree, Random Forest, XGboost, SVR), 상관분석(Heatmap)
- RStudio: 데이터 시각화(ggplot2)
- Orange

2. 분석/시각화 결과 상세내용

■ 분석/시각화 상세내용

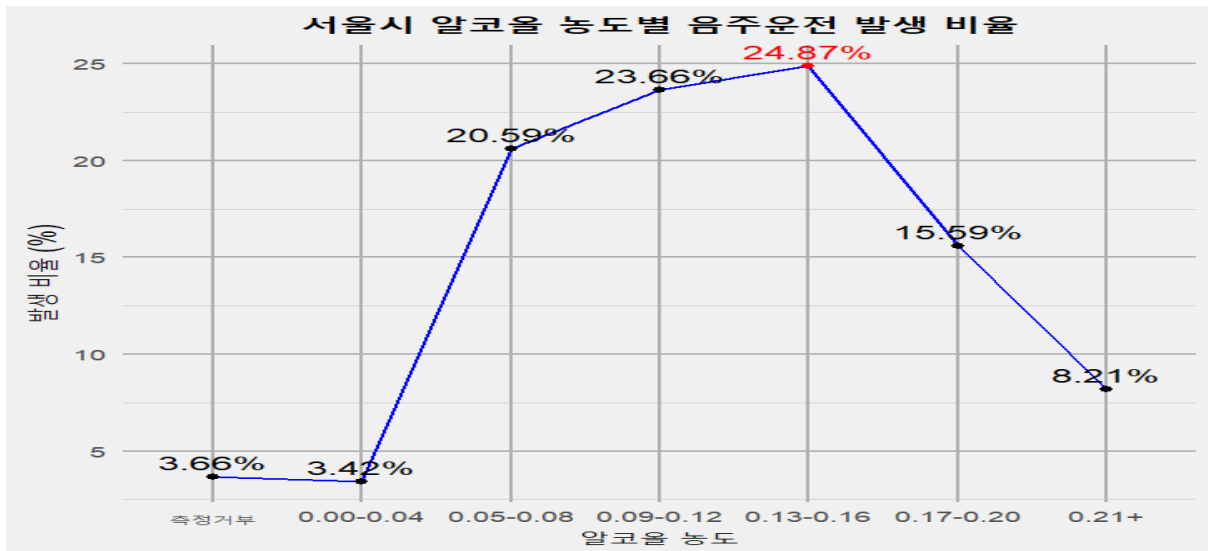
● 데이터 수집

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.

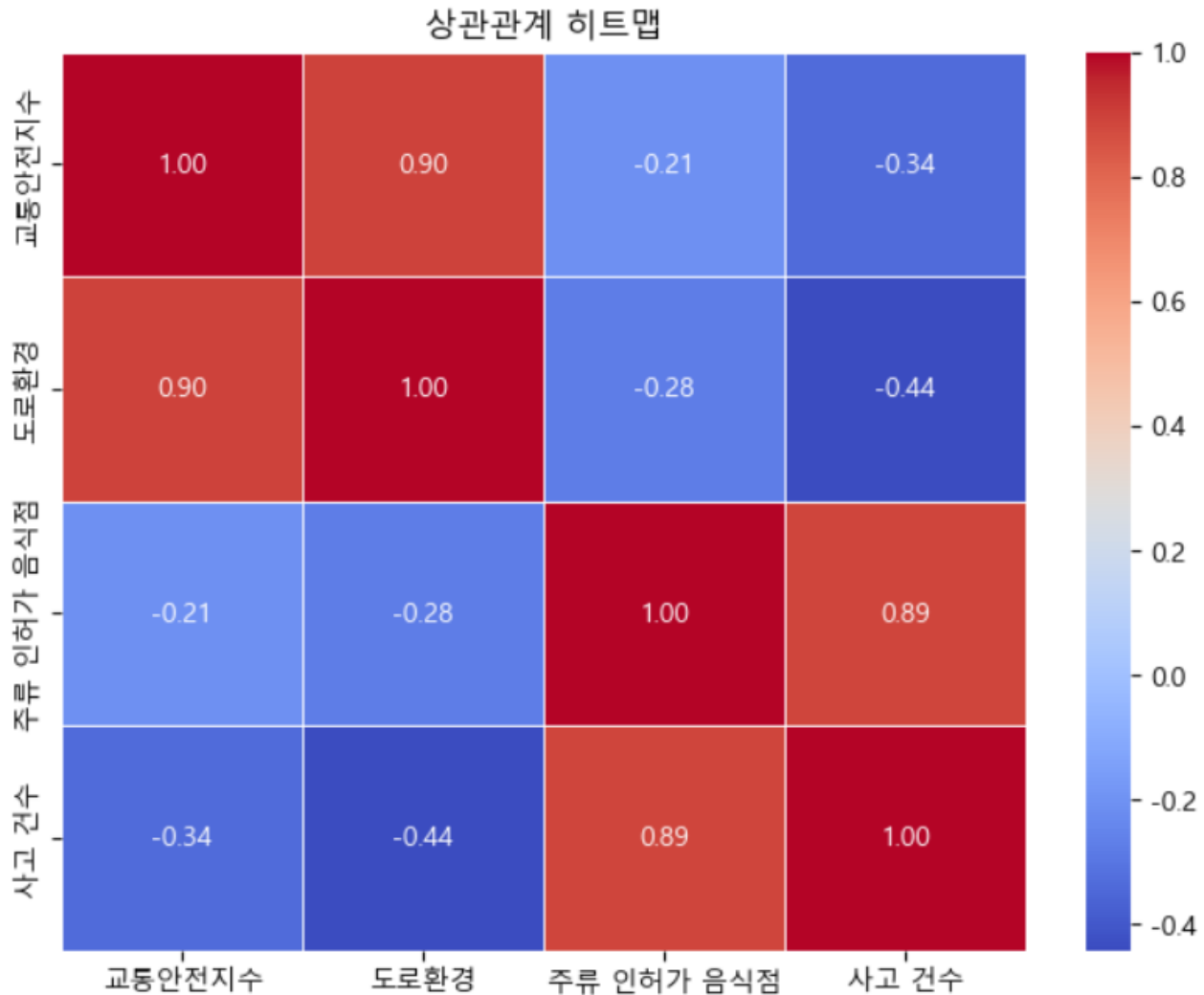
● 탐색적 분석(EDA) 및 변수탐색



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

■ 『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.

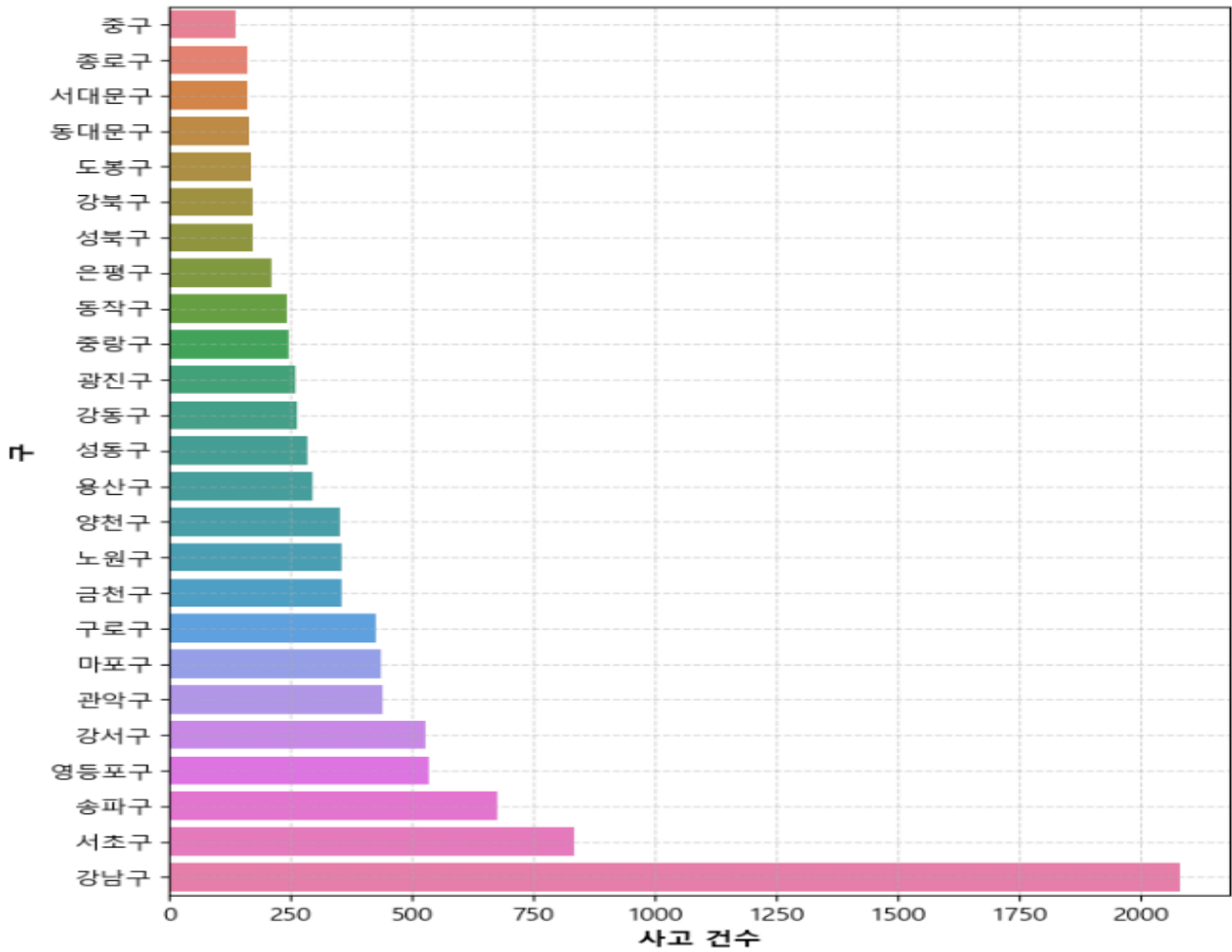
	교통안전지수	도로환경	주류 인허가 음식점	사고 건수
교통안전지수	1.0	0.8958988907793200	-0.209492435198395	-0.34044009509147200
도로환경	0.8958988907793200	1.0	-0.27816051324162700	-0.44268832831030300
주류 인허가 음식점	-0.209492435198395	-0.27816051324162700	1.0	0.8876303341641190
사고 건수	-0.34044009509147200	-0.44268832831030300	0.8876303341641190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



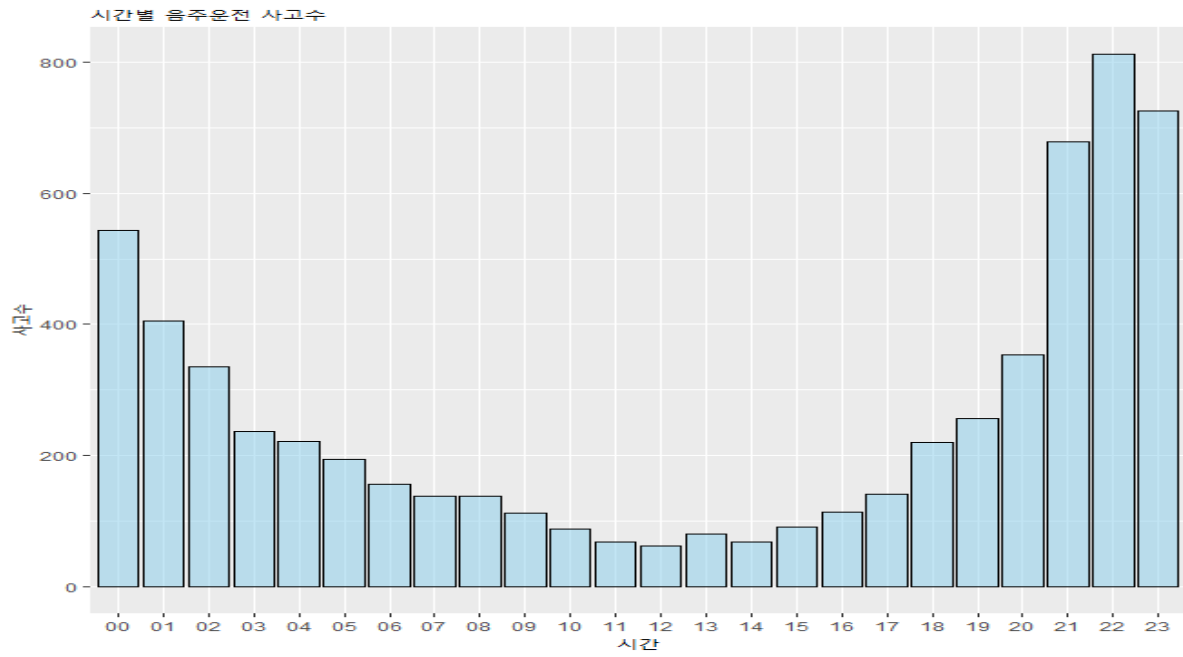
■ 교통안전지수와 도로환경은 사고 건수를 줄이는 데 중요한 역할을 함 교통안전지수와 도로환경이 좋을수록 사고 건수가 줄어드는 경향 있다.

■ 주류 인허가 음식점의 수는 사고 건수와 매우 강한 양의 상관관계를 가지므로, 이러한 지역에서 음주운전 단속을 강화하는 것이 사고를 줄이는 데 효과적일 수 있다.

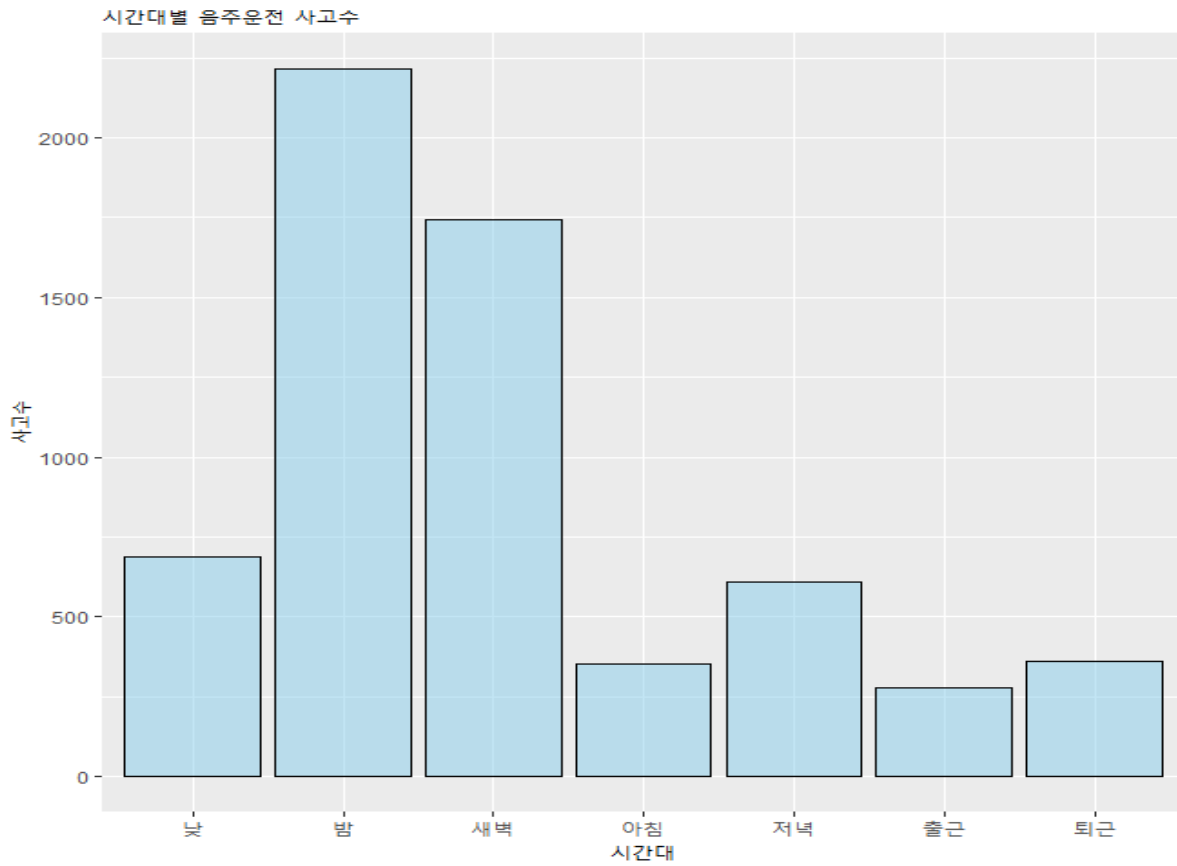
시간별 음주운전 사고건수



- 2021~2023 년 전국 구별 시간대별 음주운전 사고건수에서 추출하여 분석함. 사고건수인 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.



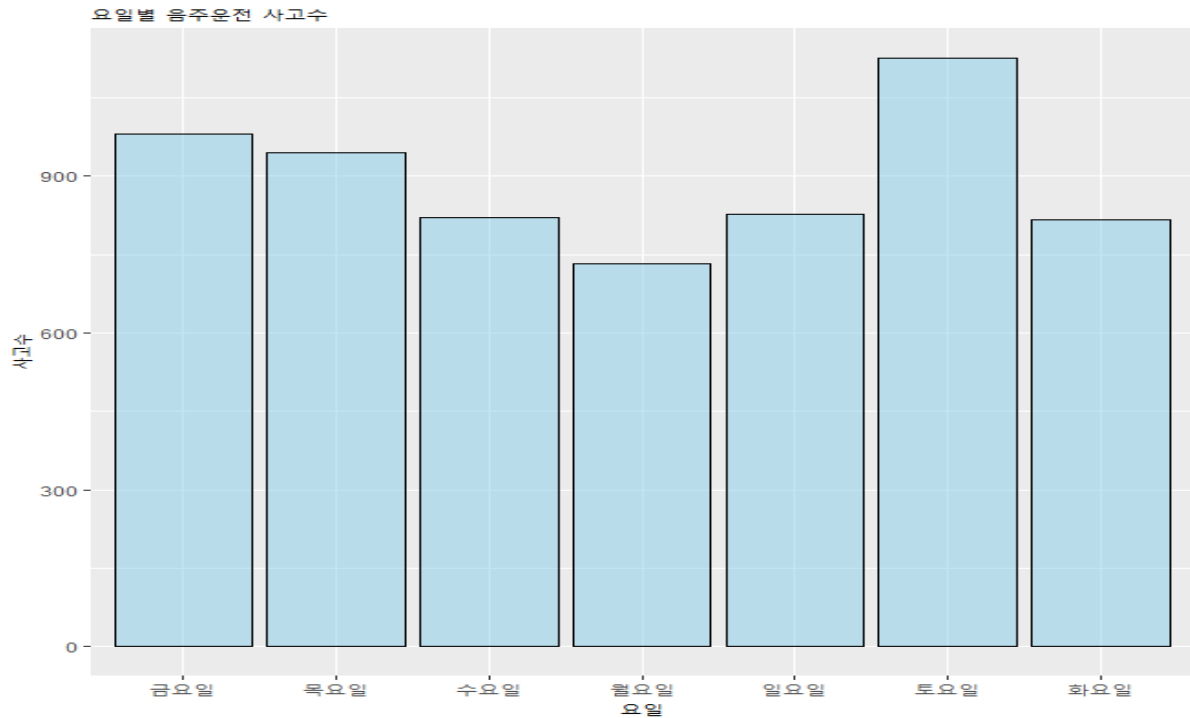
- 시간대에 주기성이 존재함을 확인할 수 있다.
- 낮에 감소했다가 밤과 새벽에 음주운전 사고가 증가하는 것을 알 수 있었다.



■ 2021~2023 년 서울시 시간대별 음주운전 사고건수 데이터

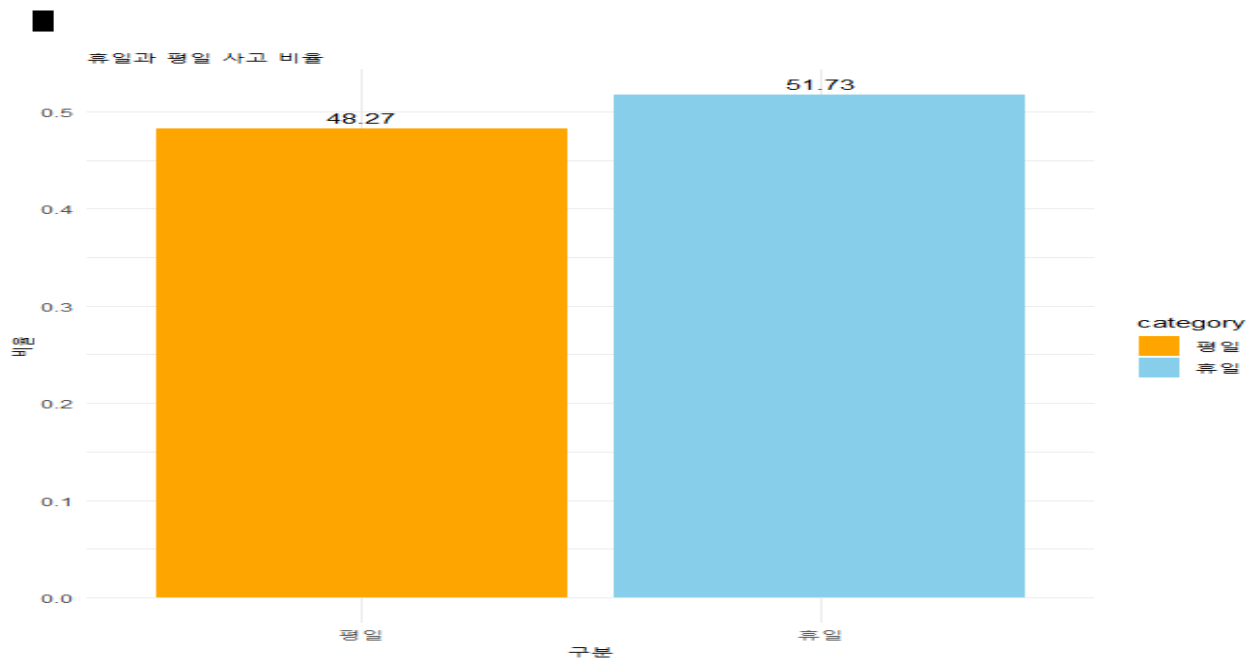
■ 밤과 새벽 시간의 음주운전 발생 빈도가 높았다.

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



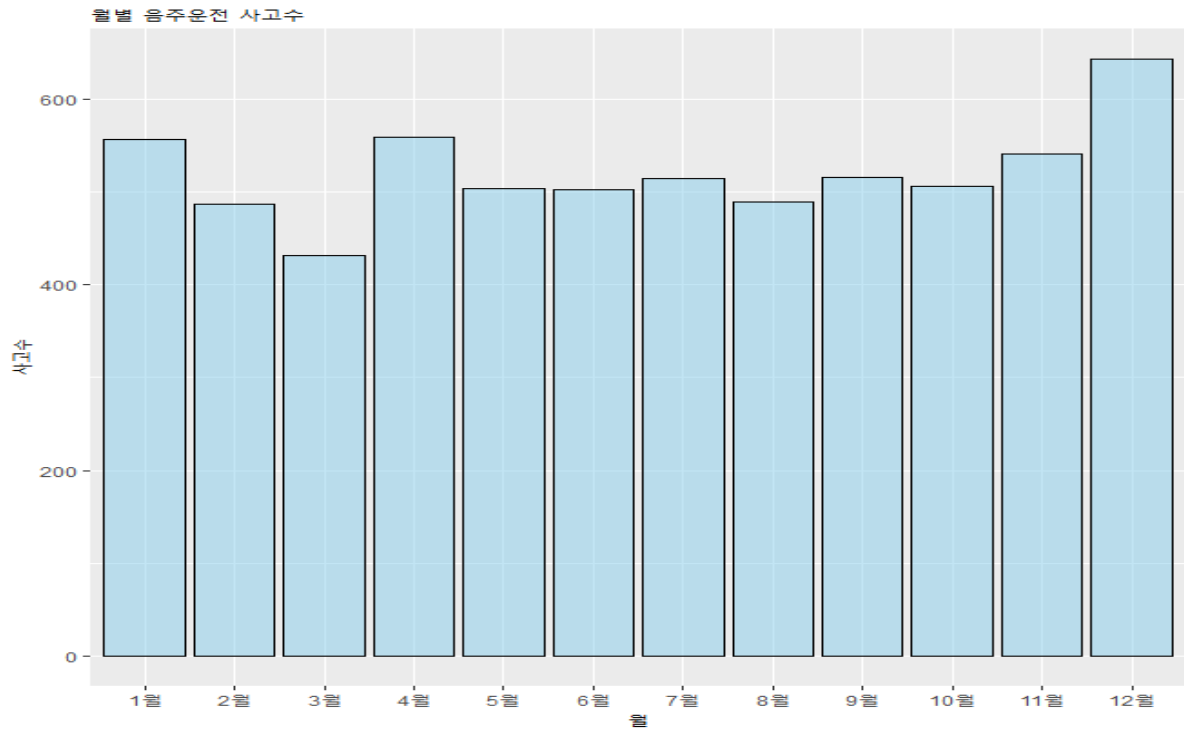
■ 2021~2023 년 서울시 요일별 음주운전 발생건수 데이터

■ 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

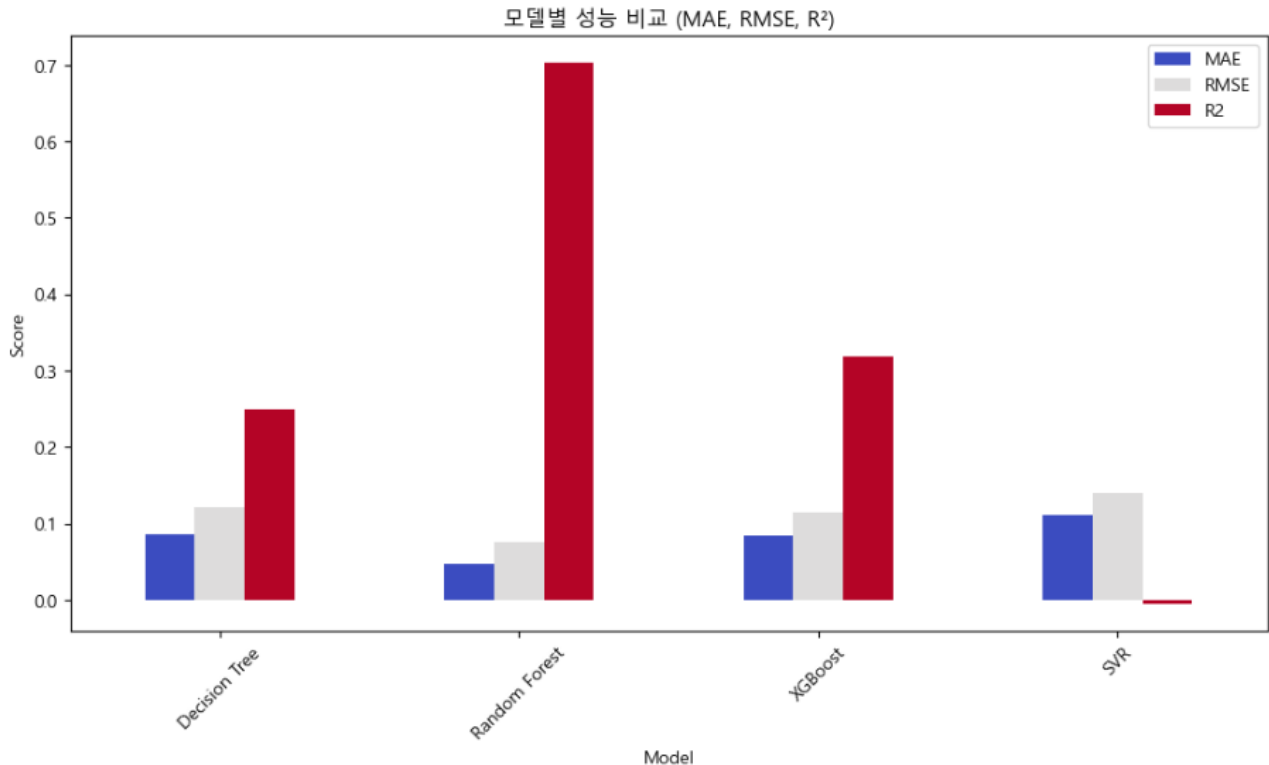
시군구	요일	사고시간
서울특별시 강남구 신사동	토요일	0
서울특별시 서초구 서초동	토요일	1
서울특별시 강남구 논현동	수요일	22

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



음주운전 사고 발생 가능성을 예측하기 위해 Decision Tree, Random Forest, XGBoost, SVR의 회귀모델의 성능을 비교한 결과이다. MAE(평균 절대 오차), RMSE(평균 제곱근 오차), R²(결정 계수)를 기준으로 하였다. Random Forest 회귀모델의 R²값이 가장 높고 MAE와 RMSE가 가장 낮아 우수한 성능을 보였다. 따라서 Random Forest 회귀모델을 최종 선택하였다.

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 확률을 반환한다.

```
# 예측 함수 정의
def predict_accident_probability(day, hour, district):
    day_encoded = label_encoders['사고요일'].transform([day])[0]
    hour_encoded = label_encoders['사고시'].transform([hour])[0]
    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
    return probability

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

probability = predict_accident_probability(day, hour, district)
print(f'🚗 {day} {hour} {district}에서 사고 발생 확률: {probability:.2%}')

🚗 목요일 23시 서울특별시 강남구 청담동에서 사고 발생 확률: 83.76%
```

■ 최종 결과

구	동	요일	사고시간	상세주소	음주 사고 발생 가능성
강남구	신사동	토요일	0	성수대교남단교차로	99.67
강남구	청담동	목요일	23	학동사거리	83.76
강남구	논현동	수요일	22	한국산업경제연구원	77.76
강남구	역삼동	목요일	23	북악하나로교차로	58.6
관악구	신림동	월요일	0	난곡사거리	50.78
성동구	성수 1 가동	금요일	23	성동구 남측	48.78
서초구	신사동	토요일	1	예술의전당앞삼거리	35.5
양천구	목동	목요일	21	신트리사거리	30.4
강남구	대치동	일요일	23	코스콤사거리	30.2
구로구	구로동	금요일	21	시립운동장오거리	29.54
금천구	가산동	금요일	23	시흥 4 동사거리	28.14
서초구	반포동	토요일	22	서래마을입구	27.59
관악구	신림동	화요일	2	난곡사거리	26
서대문구	북가좌동	금요일	22	이대역사거리	25.95
영등포구	대림동	수요일	23	신길로사거리	25.24
중랑구	상봉동	토요일	23	상봉지하차도오거리	24.2
광진구	구의동	금요일	3	영동대교북측	21.68
종로구	가회동	수요일	23	독립문역사거리	21.35
영등포구	대림동	금요일	19	창원초교앞 교차로	19.71
마포구	대흥동	토요일	1	양화동동영중포구입근	19.52
동작구	노량진동	금요일	21	수산시장입구교차로	17.18
강남구	대치동	월요일	2	한국산업경제연구원	16.93
강동구	길동	목요일	4	생태경과업교차로	16.52
강서구	방화동	월요일	4	방화역사거리	15.24
서초구	방배동	금요일	17	신사역사거리	15.02

본 연구에서는 서울특별시 내에 음주운전 사고 발생 가능성이 가장 높은 지역을 분석하여, 각 구 별로 가장 위험성이 높은 지역을 하나씩 선정하였다. 이를 통해 구별 주요 사고 위험 지역을 파악 하고, 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 시 간대와 요일이 존재함을 확인할 수 있었다. 주로 심야(22~24 시) 및 새벽 시간대에 사고 발생 가 능성이 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

- | |
|--|
| <ul style="list-style-type: none">■ 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 API

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>.