

서울시 음주운전 사고 분석을 통한 맞춤형 단속 및 예방 정책 제언

1. 분석 배경과 목적

“우리나라 음주운전 적발건수, 일본의 7배 수준이라는데...왜?”

우리나라의 음주운전 적발 건수는 일본의 7배 수준이라는 분석이 나왔다. 우리나라와 일본은 음주운전 단속 기준이 혈중알코올농도 0.03%로 유사함에도 불구하고, 일본은 2001년부터 음주운전 규제를 강화하여 성숙한 교통 문화를 구축했다. 우리나라도 음주운전에 대한 경각심을 높이기 위해 다양한 시도를 하고 있으며, 특히 부산은 음주운전 취약 지역과 사고 빈발 지역에서 집중 단속을 시행해왔다. 그 결과, 올해 상반기 부산의 음주운전 교통사고가 약 25% 감소했다. 이러한 성공 사례를 바탕으로, **서울시의 음주운전과 관련된 요인들을 확인하고자 다양한 분석방법을 사용하고, 결과를 해석하여 맞춤형 단속 및 예방 정책을 제안하고자 한다.**

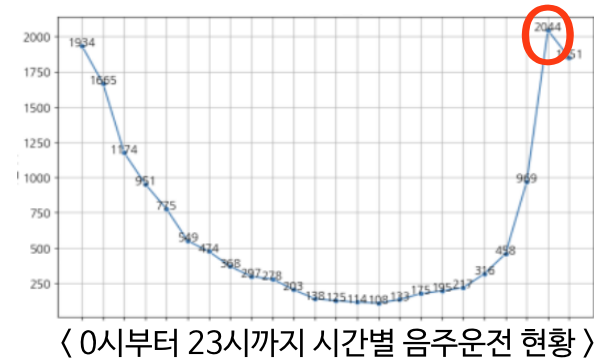


2. 서울시 음주운전 현황

- 기초 통계 분석

2022년 서울 음주운전 현황에 대한 기초 통계를 확인했다. 서울 2022년 음주운전 적발 횟수는 전체 15,511건이다. 그 중 강남구의 음주운전 적발 횟수가 1,908건으로 가장 높으며, 다른 자치구에 비해 7.5배 높다. 음주운전자의 성별은 남성 이 여성보다 10배 더 많다. 또한, 연령대는 30, 20, 40대 순으로, 전체 78%를 차지한다.

오른쪽 그래프와 같이, 시간별 음주운전 현황은 22시에 2,775건으로 가장 높고, -02시까지 높은 비율이며, 두번째 그림에서 요일별 음주운전 현황에서 토요일에 2,775건으로 가장 높다. 우리팀은 먼저, SGIS의 자치구별 음주율과 음주운전 횟수의 상관분석을 진행했는데, 음주율과 음주운전의 상관계수는 -0.16으로, 음주율과 음주운전의 관련성이 전혀 없음을 확인하였다. 그렇다면, 음주운전에 영향을 미치는 요인들이 무엇인지 관련된 요인의 연관성을 파악하고, 예측모델을 생성하여 서울시의 음주운전 집중 단속 지역을 예측하고자 한다.



〈0시부터 23시까지 시간별 음주운전 현황〉



〈월요일 ~ 일요일까지 요일별 음주운전 현황〉

3. 데이터 소개

[1] 데이터 수집 (대상:서울시)

데이터명	데이터 내용	연도	출처
음주운전 적발 신상정보	성별, 적발횟수, 나이, 알콜농도, 측정일시, 관할경찰서	2022	경찰청
행정동별 자동차 등록대수 현황, 가로등 위치 정보, 도시고속도로 교통 CCTV 설치위치 정보, 지하철역 엘리베이터 위치정보, 등	연료별 자동차 등록 현황, 가로등의 위도, 경도 정보, CCTV의 시설물ID, 설치위치, 위치좌표, 엘리베이터의 노드ID, 산업별 종사자 수, 서울시 지하철 승객수, 버스 이용자 수, 시군구/읍면동 코드 .. 등	2023	서울 열린 데이터 광장
고위험 음주를 지리정보	음주한 사람 중 5잔 이상을 주 2회 이상 마시는 사람의 비율	2022	SGIS 통계지리정보서비스
상가(상권)정보	영업 중인 상가업소의 상호명, 업종코드, 도로명주소, 경도, 위도 등	2024	소상공인시장진흥공단
버스정류소 위치정보	버스정류소 (X,Y)위치정보	2022	서울시 교통정보
성연령별_직장인구(SKT)	서울시 각 자치구별 남녀 20대~50대 직장인구 집계	2022	SDC 통계데이터센터
성연령별_주거인구(SKT)	서울시 각 자치구별 남녀 20대~50대 주거인구 집계	2022	SDC 통계데이터센터

4. 분석결과

[1] 자치구별 음주운전 특징 파악

: 음주운전을 데이터를 요일별, 월별, 시간대별, 연령대별로 각각 K-means 군집분석을 통해, 자치구별 음주운전의 성향을 파악함

1) 요일별 군집 분석



→ **[빨강]**: 큰 차이는 없지만 목,금,토요일에 집중됨.
→ **[노랑]**: 토요일만 집중됨. 일요일은 현저히 낮음.
→ **[파랑]**: 토요일, 일요일을 포함한 주말에 집중됨.
→ **[검정]**: 토요일도 높으나, 일요일에 가장 높음.

실루엣 계수: 0.53, 다빈치 지표: 0.93
비슷한 패턴으로 묶인 상태, 클러스터가 적절히 분리됨

2) 월별 군집 분석



→ **[빨강]**: 5월부터 10월까지 비중이 높음.
→ **[노랑]**: 4월과 5월에 비중이 높지만, 음주운전율이 전체적으로 고르게 분포
→ **[파랑]**: 6월부터 12월, 1월까지 비중이 높지만, 전체적으로 고르게 분포.
→ **[검정]**: 6월과 7월, 12월에 집중적으로 높음. 상,하반기 결산 등으로 유추됨.

실루엣 계수: 0.15, 다빈치 지표: 1.57
실루엣 계수가 낮고, 다빈치 지표가 1보다 크다. 뚜렷한 특징은 없다고 판단. 월마다 28일,30일,31일 일수로 인한 차이가 커서, 분리가 어렵다고 판단됨.

3) 시간대별 군집 분석



→ **[빨강]**: 0~4시(40%)에 가장 높음. 20~24시는 30%를 차지
→ **[노랑]**: 20~24시, 0~4시 33%로 비슷하게 집중됨.
→ **[파랑]**: 20~24시(39%)에 가장 높음, 0~4시는 35%를 차지

실루엣 계수: 0.45, 다빈치 지표: 0.89
비슷한 패턴으로 묶인 상태, 클러스터가 적절히 분리됨

4) 연령대별 군집 분석



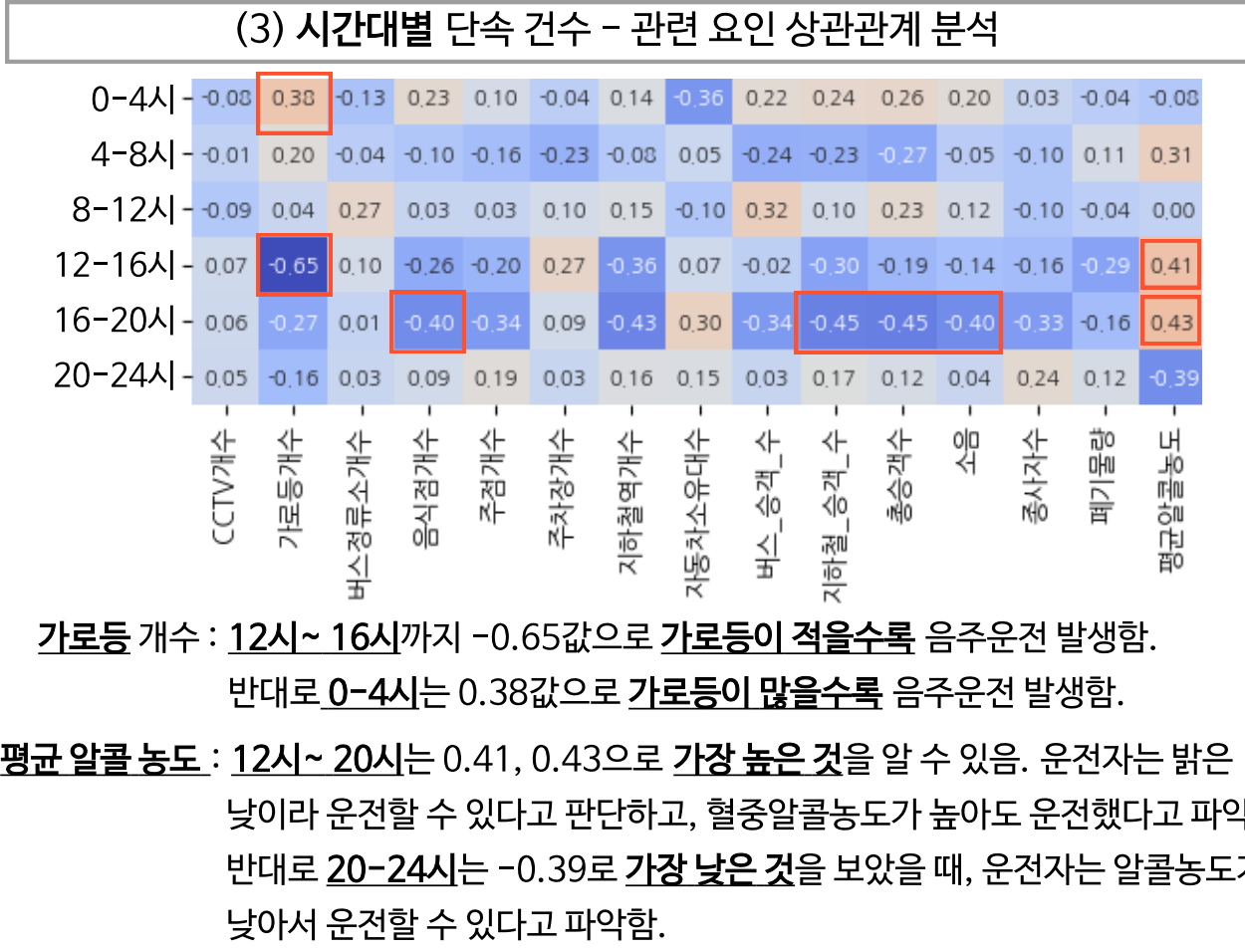
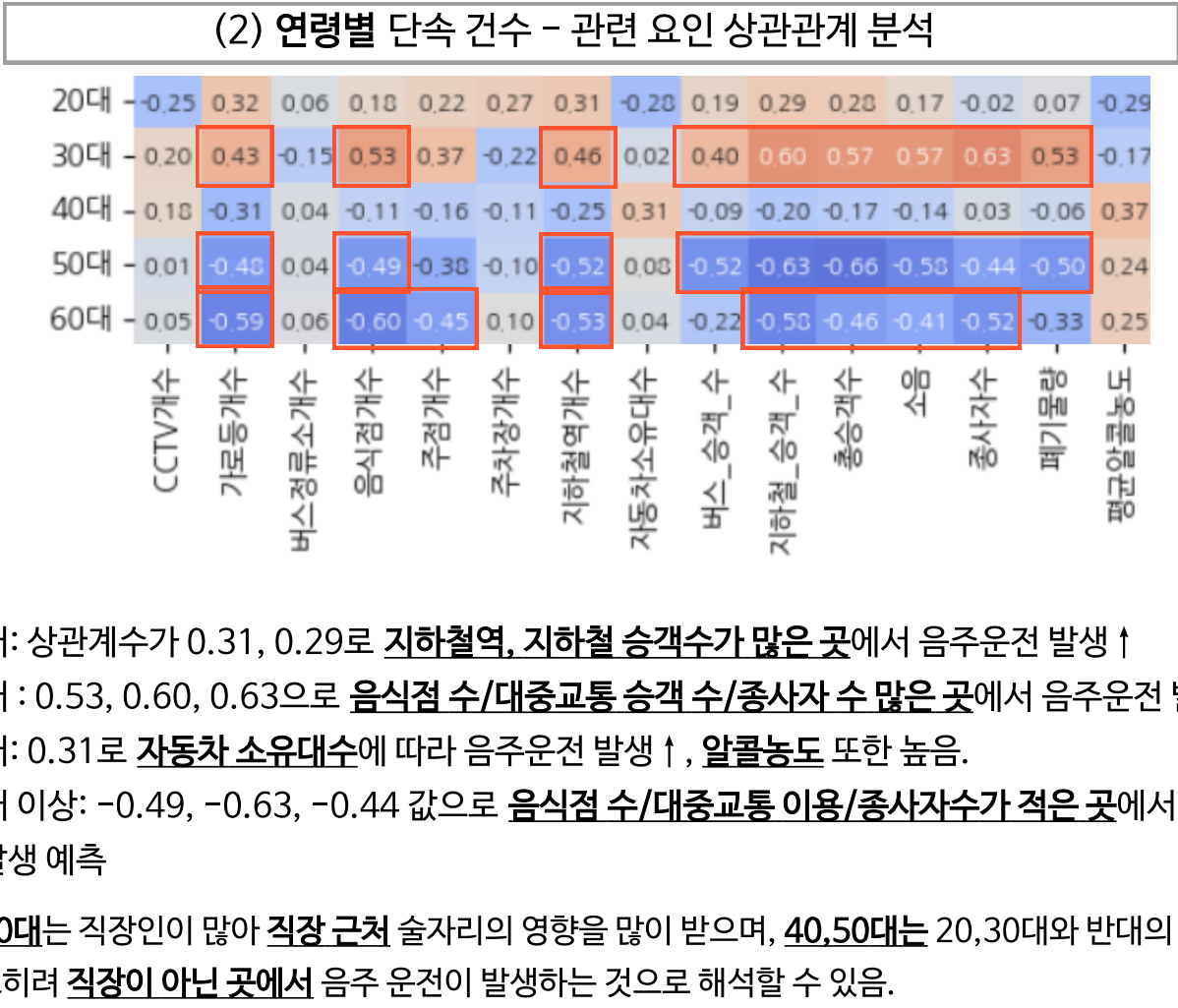
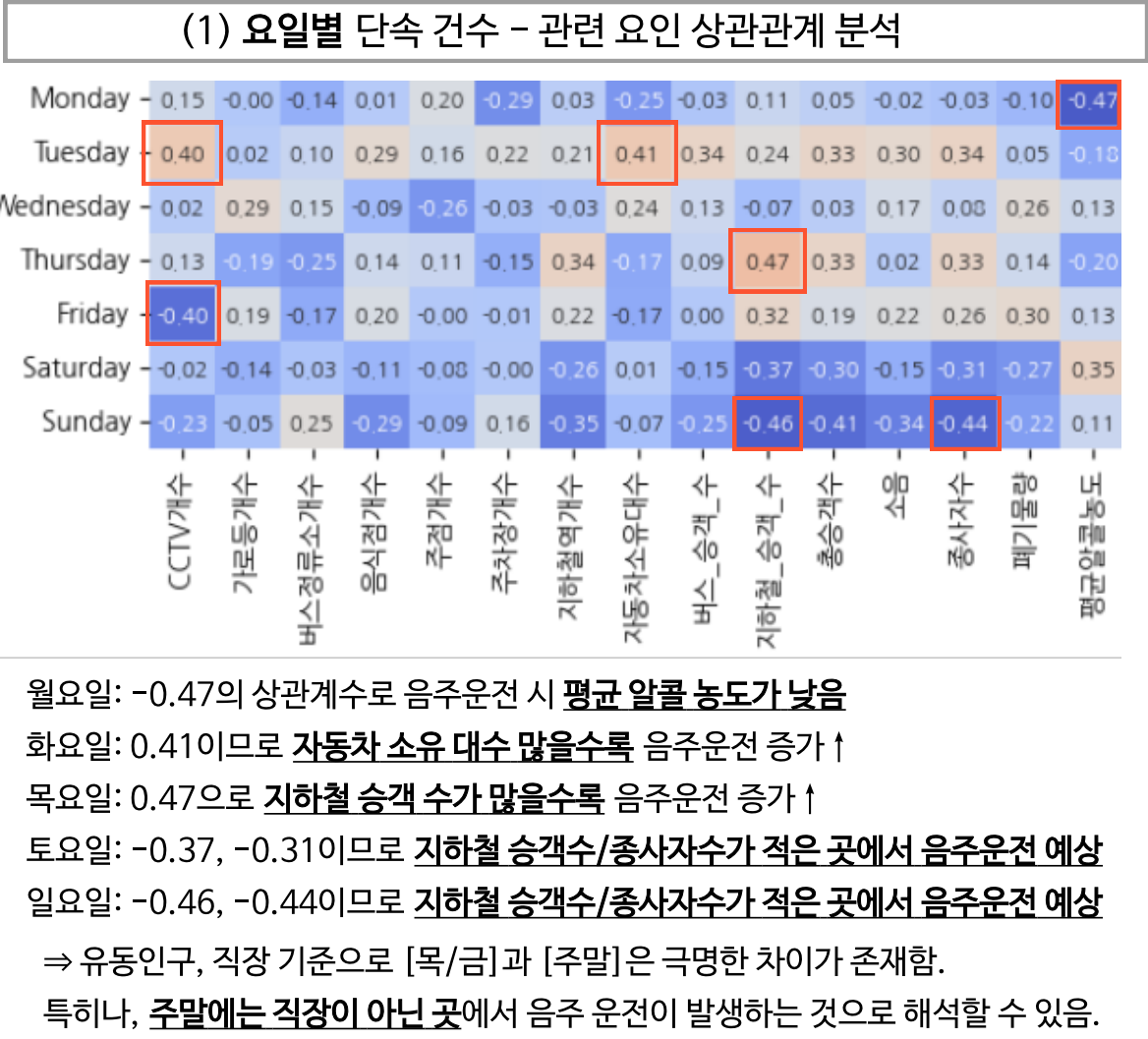
→ **[빨강]**: 30대(31%), 40대(26%)으로 가장 많음.
→ **[노랑]**: 20대(31%), 30대(27%)으로 가장 많음.
→ **[파랑]**: 20대, 30대, 40대, 50대 각각 20%로 고루 분포함.
→ **[검정]**: 30대(33%)가 가장 높고, 20대, 40대는 22%를 차지함.

실루엣 계수: 0.43, 다빈치 지표: 0.91
비슷한 패턴으로 묶인 상태, 클러스터가 적절히 분리됨

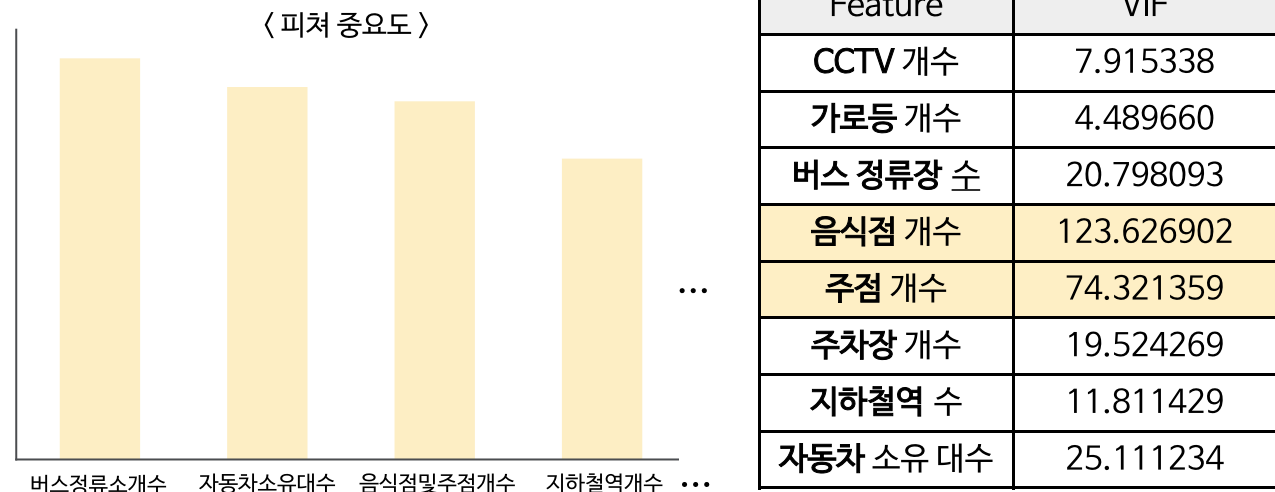
* 군집분석의 평가지표인, 실루엣계수/다빈치 지표 값을 확인함. 클러스터의 차이와 내부 밀집도 확인하여 실루엣계수가 0.5 이상일때, 각 군집의 특성이 분리됨, 다빈치지표가 1 이하일때, 클러스터 내부 요소들이 밀집되어 잘 모여있는 상태.

[2] 음주운전 사고와 관련 요인 분석 : 음주운전의 다양한 요인들의 영향을 해석하기 위해 1번째로 먼저 상관분석을 진행, 2번째로 다중공선성을 확인하고 랜덤포레스트 기법을 이용하여, 모델에 적용할 요인들을 선정함.

1) 상관분석 결과



2) 랜덤포레스트 결과



(1) 음식점 수와 주점 수의 VIF 값이 123, 74로 유독 높게 나타나, 다중공선성 문제를 해결하기 위해 하나로 통합.

(2) 랜덤포레스트의 피쳐 중요도를 확인, 음주운전에 영향을 주는 상위 4개의 요인인, 버스 정류소 개수, 음식점 및 주점 개수, 지하철역 개수, 자동차 소유 대수를 예측모델에 학습할 요인으로 결정.

[3] 예측모델 (다변량회귀모델)

* 버스 정류소 개수, 음식점 및 주점 개수, 지하철역 개수, 자동차 소유 대수를 예측모델에 학습함.

〈데이터를 토대로 학습시킨 예측 모델〉
결정계수 (R^2) : 77%의 설명력을 가짐.
평균절대오차 (MAE) : 9.67
MAE 계산 결과 9.67 값으로, 모델을 사용하기에 큰 문제가 없을 것으로 예상됨.
⇒ 모델을 이용해, 서울시 행정동별 (연간) 음주운전 건수를 예측한 결과 [표1]과 같이 나타남.

행정동명	예측 음주운전 건수
강남구 역삼1동	359.891
강남구 압구정동	170.207
강남구 논현2동	150.275
강남구 논현1동	134.800
강남구 신사동	129.687
:	:

[표1]

5. 결론 (예상효과)

위와 같이 서울시 음주운전 데이터를 통해, 음주운전의 주요 원인, 발생 패턴을 파악했다. 분석 결과를 바탕으로, 아래와 같은 **자치구별 맞춤형 단속 및 예방 정책을 제안한다.**

(1) 각 자치구별 음주운전 집중 단속 시간대

군집 분석으로, 서울 25개의 각 자치구는 표와 같이 집중 단속을 해야하는 요일과, 시간대, 월별로 결과를 확인할 수 있음.

자치구	요일	월	시간	연령
강남구	토	4,5,9월	0~4시	30대
강동구	토	6,12월	20~24시	30~40대
강북구	토,일	7,9,10월	20~24시	20~30대
강서구	목,금,토	4,5,9월	20~24시	30대
:	:	:	:	:

(2) 음주운전 빈발 지역 선정

예측모델을 이용해, 아래와 같이 50건 이상 음주운전 건수가 높은 지역(총 92개)을 선정함.



예시) 강동구



길동에서 예상 130건을 52주로 나누어, **토요일마다 3건의 음주운전이 예상된다.** 위에 따르면 100명중 1명이 음주운전을 하는 비율로 계산하여 (3건 * 100명) **약 300명의 음주단속을 시행**한다. 경찰청에 따르면 2명이 한 시간에 25건을 단속할 때, 따라서 길동에 토요일 20~24시에 **6명의 경찰 인력이 필요하다.** 동일한 방식으로, 성내2동과 성내3동, 천호2동을 계산하여 해당 행정동에 **3~4명의 인력이 필요하다.** 또한 음주측정기와 순찰차량, 통신문방이 필요하며, 계산 결과 **2억원(연간 인건비) + 2억원(장비비용) = 4억원**으로 계산된다. 이런 단속방법을 통해서 기존 음주운전 데이터에서, 강동구의 토요일 20~24시 음주운전이 20% 감소한다고 가정하여, **정규분포 시뮬레이션**을 실행하였다.

〈강동구 22년 기존 음주운전 vs 시뮬레이션 예상 음주운전〉



5천만 원으로 산정할 때, 예상되는 **음주운전 감소 효과로 인해 총 25억 원(51건 x 5천만 원)의 경제적인 비용 절감**이 가능할 것으로 보인다. 이는 강동구의 경찰 인력 및 장비 배치에 대한 투자 비용을 충분히 상쇄하고도 남는 금액으로, 음주운전 감소를 통해 **지역사회의 안전성을 크게 향상**시킬 수 있을 것이다.

기존의 2022년도 강동구의 음주운전 적발 수는 총 524건이었으며, 이 중 **약 51건(약 10%)이 감소할 것으로 예측**했다. 한국 교통연구원 2021년 교통사고비용 조사에 따르면, 소득손실 비용, 의료비용, 행정비용, 물적피해비용, 심리 비용 등 사고 1건당 평균 도로교통사고 비용을

이렇게 시뮬레이션을 통해 강동구의 예상효과를 확인해보았다. 이러한 모델을 서울의 각 자치구에 적용하여 단속 정책을 제안할 수 있다. 이외에도 경기도 및 다른 지역에도 적용하여, **음주운전의 발생 위치를 미리 파악하고 예방할 수 있다.** 하지만, 음주운전을 효과적으로 예방하는 해결책은 음주 단속만으로 해결하기에 한계가 있다. 따라서 아래와 같은 방법도 제안한다.

- * 분석결과로 파악한 **음주운전 예상 지역에 음주 감지 장치 및 도로 카메라를 설치**하여, 음주운전을 사전에 감지하고 즉각 대응할 수 있는 시스템을 구축한다.
- * 분석결과로 30대와 40대의 음주운전이 빈번한 지역과 20대의 음주운전이 빈번한 지역을 알 수 있었다. 이러한 **각 지역의 음주운전자 대상**으로, **타겟화하여 음주운전의 위험성을 알리는 것**이 필요하다.
- * 또한 종사자 수가 많은 지역일수록 음주운전이 빈번하므로, **사내 교육과 회사 자리에서의 안전 관리**가 필요하다.

이와 같은 **맞춤형 단속 및 예방 정책**을 통해 **음주운전 사고를 줄이고, 안전한 교통 환경을 조성**할 수 있을 것이다. 이러한 정책들이 효과적으로 실행되기 위해서는 서울시와 경찰, 교육 기관, 지역 사회의 긴밀한 협력이 필수적이다. 이를 통해 음주운전으로 인한 무고한 피해가 없는 안전한 나라로 성장할 수 있을 것이다.