# 서울시 음주운전 사고 분석을 통한 맞춤형 단속 및 예방 정책 제안

### 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

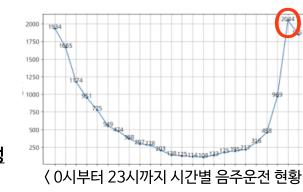
혈중알코올농도 0.03%로 유사함에도 불구하고, 일본은 2001년부터 음주운전 규제를 강화하여 성숙한 교통 문화를 구축했다. 우리나라도 음주운전에 대한 경각심을 높이기 위해 다양한 시도를 하고 있으며, 특히 부산은 음주운전 취약 지역과 사고 빈발 지역에서 집중 단속을 시행해왔다. 그 결과, 올해 상반기 부산의 음주운전 교통사고가 약 25% 감소했다. 이러한 성공 사례를 바탕으로, 서울시의 음주운전과 관련된 요인들을 확인하고자 다양한 분석방법을 사용하고, 결과를 해석하여 맞춤형 단속 및 예방 정책을 제안하고자 한다.



2. 서울시 음주운전 현황 - 기초통계 분석

2022년 서울 음주운전 현황에 대한 **기초 통계**를 확인했다.

서울 2022년 음주운전 적발 횟수는 전체 <u>15,511건</u>이다. 그 중 <u>강남구</u>의 음주운전 적발 횟수가 1,908건으로 가장 높으며, 다른 자치구에 비해 7.5배 높다. 음주운전자의 성별은 <u>남성</u>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, <u>- 02시까지</u> 높은 비율이며, 두번째 그림에서 요일별 음주운전 현황에서 <u>토요일</u>에 2,775건으로 가장 높다. 우리팀은 먼저, SGIS의 자치구별 음주율과 음주운전 횟수의 상관분석을 진행했는데, 음주율과 음주운전의 상관계수는 - 0.16 으로, 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.

### 3. 데이터 소개

### [1] 데이터 수집 (대상:서울시)

데이터명	데이터 내용	연도	출처
<u>음<b>주운전 적발</b></u> 신상정보	성별, 적발횟수, 나이, 알 <del>콜농</del> 도, 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 시설물I D, 설치위치, 위치좌표, 엘리베이터의 노드ID, 산업별 종사자 수, 서울 시 지하철 승객수, 버스 이용자 수, 시군구/읍면동 코드 등	2023	서울 열린 데이터 광장
고위험 음주율 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

#### [2] 데이터 전처리

- 1. 서울시 자치구별 CCTV, 가로등, 음식점 및 주점, 주차장, 소유 자동차 수, 버스 정류소 및 지하철 역 등 속한 행정동의 평균값으로 산출
- CCTV, 가로등, 버스 정류소 및 지하철역의 '위경도 데이터'를 기준으로 경곗값 구분
- 음식점 데이터 분류명 필터링을 통해 '비주점' 제외 2. 서울시 자치구별 음주운전 단속 및 사고건수 집계
- '신사동' 등 중복된 행정동명으로 발생한 결측치 평균값으로 처리
- 3. 조건별(요일별, 월별, 연령별, 시간대별) 음주운전 적발 비중 집계 4. 자치구 기준 데이터 병합
- 시군구명CCTV수가로등수버스정류소수음식점수...강남구306.77282.00025.409350.318...강동구72.40047.35022.050156.700...:::::

\_\_\_\_ 〈 최종 전처리 데이터 예시 〉

### [3] 데이터 부석 절차 : 다양한 요인들을 파악하기위해, 다양한 분석 방법으로 활용하여, 분석결과의 신뢰도를 높인다

- 1. 클러스터링(k-means): 음주운전 적발 건수를 요일별, 연령대별 등 음주운전의 특성을 확인하기 위함.
- 2. 상관 분석: 음주운전과 각 요인 변수들(CCTV, 음식점 등)의 관계를 확인하고, 어떤 요인으로 음주운전이 발생하는지 파악.
- 3. 랜덤 포레스트 : 앙상블 기법 중 하나로, 회귀모델에 사용될 요인을 선정하기 위해서 실행 후, 피쳐 중요도를 확인함.
- <mark>4. 다변량 회귀분석</mark> : <u>여러 결과(<del>종속</del>변수)</u>들이 <u>여러 원인(독립변수)</u>들에 의해 어떻게 영향을 받는지, <u>음주운전 발생 예측 모델을 생성함.</u>

### 4. 분석결과 [1] 자치구별 음주운전 특징 파악 : 음주운전율 데이터를 요일별, 월별, 시간대별, 연령대별로 각각 K-means 군집분석을 통해, 자치구별 음주운전의 성향을 파악함

### 1) 요일별 군집 분석



- → <mark>[빨강]</mark> : 큰 차이는 없지만 <u>목,금,토요일</u>에 집중됨.
- → [노랑]: <u>토요일만</u> 집중됨. 일요일은 현저히 낮음.
- → [파랑]: 토요일, 일요일을 포함한 <u>주말에 집중</u>됨. → [검정]: 토요일도 높으나, <u>일요일에 가장 높음</u>.

실루엣 계수: 0.53, 다빈치 지표: 0.93 비슷한 패턴으로 묶인 상태, <u>클러스터가 적절히 분리됨</u>

### 2) 월별 군집 분석



- → <mark>[빨강]</mark>: <u>5월부터 10월까지</u> 비중이 높음.
- → <mark>[노랑]</mark>: <u>4월</u>과 <u>5월</u>에 비중이 높지만, 음주운전율이 전체적으로 고르게 분포 → <mark>[파랑]</mark>: <u>6월부터 12월,1월까지</u> 비중이 높지만, 전체적으로 고르게 분포.
- → [검정]: <u>6월</u>과 <u>7월</u>, <u>12월</u>에 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



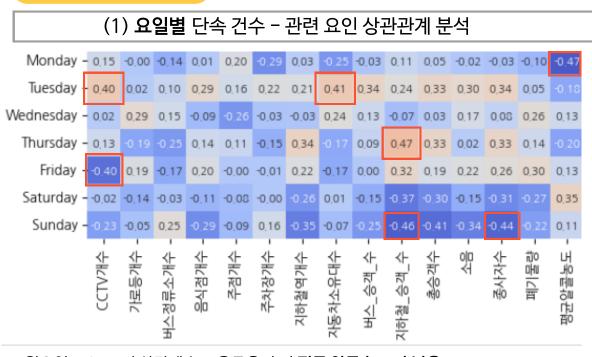
- → <mark>[빨강]</mark>: <u>0-4시</u>(40%)에 가장 높음. 20-24시는 30%를 차지
- → <mark>[노랑]</mark>: **20-24시, 0-4시** 33%로 비슷하게 집중됨.
- → [파랑]: <u>20-24시</u>(39%)에 가장 높음, 0-4시는 35%를 차지

실루엣 계수: 0.45, 다빈치 지표: 0.89 비슷한 패턴으로 묶인 상태, <u>클러스터가 적절히 분리됨</u>



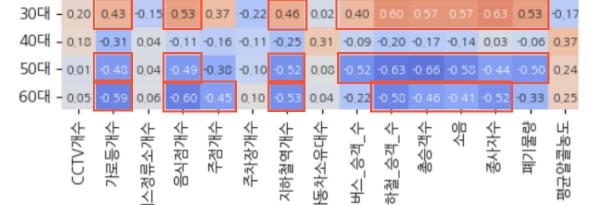
- → [빨강]: <u>30대(31%), 40대(26%)</u>으로 가장 많음.
- → <mark>[노랑]: 20대(31%), 30대(27%)</mark>으로 가장 많음.
- → [파랑]: 20대, 30대, 40대, 50대 각각 20%로 고루 분포함. → [검정]: 30대(33%)가 가장 높고, 20대, 40대는 22%를 차지함.
  - **실루엣 계수: 0.43, 다빈치 지표: 0.91** 비슷한 패턴으로 묶인 상태, <u>클러스터가 적절히 분리됨</u>
- \* 군집분석의 평가지표인, 실루엣계수/다빈치 지표 값을 확인함. 클러스터의 차이와 내부 밀집도 확인하여 실루엣계수가 0.5 이상일때, 각 군집의 특성이 분리됨, 다빈치지표가 1 이하일때, 클러스터 내부 요소들이 밀집되어 잘 모여있는 상이
- [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



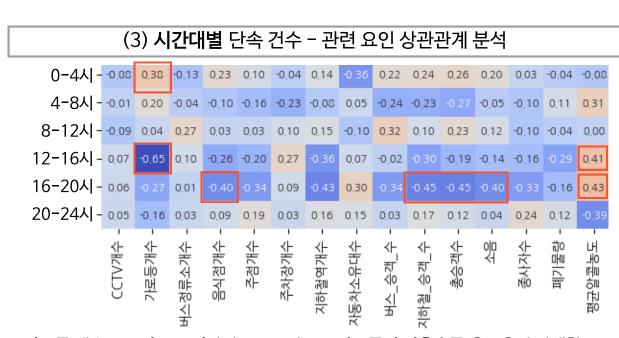
- 월요일: -0.47의 상관계수로 음주운전 시 평균 알콜 농도가 낮음 화요일: 0.41이므로 **자동차 소유 대수 많을수록** 음주운전 증가 ↑
- 목요일: 0.47으로 **지하철 승객 수가 많을수록** 음주운전 증가 ↑
- 토요일: -0.37, -0.31이므로 **지하철 승객수/종사자수가 적은 곳에서 음<del>주운</del>전 예상**
- 일요일: -0.46, -0.44이므로 **지하철 승객수/종사자수가 적은 곳에서 음주운전 예상**
- ⇒ 유동인구, 직장 기준으로 [목/금]과 [주말]은 극명한 차이가 존재함. 특히나. **주말에는 직장이 아닌 곳**에서 음주 운전이 발생하는 것으로 해석할 수 있

### (2) **연령별** 단속 건수 - 관련 요인 상관관계 분석 20대 - 0.25 0.32 0.06 0.18 0.22 0.27 0.31 0.28 0.19 0.29 0.28 0.17 0.02 0.07 0.29



20대: 상관계수가 0.31, 0.29로 **지하철역, 지하철 승객수가 많은 곳**에서 음주운전 발생 ↑ 30대: 0.53, 0.60, 0.63으로 **음식점 수/대중교통 승객 수/종사자 수 많은 곳**에서 음주운전 발생 ↑

- 40대: 0.31로 **<u>자동차 소유대수</u>에 따라 음주운전** 발생↑, **알콜농도** 또한 높음.
- 50대 이상: -0.49, -0.63, -0.44 값으로 **음식점 수/대중교통 이용/종사자수가 적은 곳**에서 음주운 전 발생 예측
- ⇒ 30대는 직장인이 많아 직장 근처 술자리의 영향을 많이 받으며, 40,50대는 20,30대와 반대의 결과로, 오히려 직장이 아닌 곳에서 음주 운전이 발생하는 것으로 해석할 수 있음.



<u>가로등</u> 개수 : <u>12시~ 16시</u>까지 -0.65값으로 <u>가로등이 적을수록</u> 음주운전 발생함. 반대로 <u>0-4시</u>는 0.38값으로 <u>가로등이 많을수록</u> 음주운전 발생함.

평균 알콜 농도: 12시~ 20시는 0.41, 0.43으로 가장 높은 것을 알 수 있음. 운전자는 밝은 낮이라 운전할 수 있다고 판단하고, 혈중알콜농도가 높아도 운전했다고 파악됨. 반대로 20-24시는 -0.39로 가장 낮은 것을 보았을 때, 운전자는 알콜농도가 낮아서 운전할 수 있다고 파악함.

# 2) 랜덤포레스트 결과



	Feature	VIF
	CCTV 개수	7.915338
	<b>가로등</b> 개수	4.489660
	<b>버스 정류장</b> 소	20.798093
	<b>음식점</b> 개수	123.626902
	<b>주점</b> 개수	74.321359
	<b>주차장</b> 개수	19.524269
	지하철역 수	11.811429
	<b>가동차</b> 소유 대수	25.111234

- (1) 음식점 수와 주점 수의 VIF 값이 123, 74로 유독 높게 나타나, <u>다중공선성 문제</u>를 해결하기 위해 <u>하나로 통합.</u>
  - (2) <u>랜덤포레스트의 <mark>피쳐 중요도</mark>를 확인,</u> 3 8 강에 연합용 공도 **사이 4개이** 유입이 <mark>및 4 경로 4</mark>
- 음주운전에 영향을 주는 **상위 4개의** 요인인, **버스 정류소 개**수, 음식점 및 주점 개수, 지하철역 개수, 자동차 소유 대수를

### [3] 예측모델(다변량회귀모델)

- \* <u>버스 정류소 개수, 음식점 및 주점 개수, 지하철역 개수, 자동차 소유 대수</u>를 예측모델에 학습함
- 〈데이터를 토대로 학습시킨 예측 모델 〉

결정계수 (R²) : 77%의 설명력을 가짐

평균절대오차 (MAE) : 9.67

MAE 계산 결과 9.67 값으로, 모델을 사용하기에 큰 문제가 없을 것으로 예상됨.

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

⇒ 모델을 이용해, **서울시 행정동별 (연간) 음주운전 건수를 예측한 결과** [표1]과 같이 나타남.

## 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 정책**을 제안한다.

#### (1) 각 자치구별 음주운전 집중 단속 시간대 |욱 25개이 가 가기그는 피어 가이 지즉 다소은

군집 분석으로, 서울 <u>25개의 각 자치구</u>는 표와 같이 집중 단속을 해야하는 <u>요일과, 시간</u> 대, 월별로 결과를 확인할 수 있음.

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

### (2) 음주운전 빈발 지역 선정

예측모델을 이용해, 아래와 같이 <u>50건 이상 음주운전 건수가 높은 지역(총 92개)을 선정</u>함.



### 예시 ) 강<del>동구</del>

예상 단속 효과를 확인하기 위해, **강동구를 예시**로 계산한다.



- 앞서 (1)집중 단속 시간대 결과에서 강동구의 음주운전 주요 대상은 30~40대로 파악했다. 예측한 음주운전 위험 지역은 <u>길동, 성내2동, 성내3동, 천호2동</u>이 각각 연간 130건, 96건, 66건, 60건으로 예측했다. 실제 이 지역은 다른지역보다 <u>주점 수가 150개 더 많고</u>, 음주운전 가능성이 높다고 예상된다. 강동구의 음주운전 특징은 <u>직장인구가 대부분</u>이며, <u>토요일 20~24시에 집중</u>하여 단속한다. 30,40대의 주거인구는 6.7만명, **직장인구는 9만명**으로 집계되었다. 9만 명 중 **강동구 총 예상음주운전은**
- 30,40대의 구기한구는 0.7한당, <u>작용한구는 9한당으로 합계되었다. 9한 당 중 중 중 중 에당남구운전은</u>
  530건이다. 약 0.5%, 강동구의 <u>남성 음주운전자가 95%이기 때문에</u> 남성으로만 제한한다면 1%, 100명 중 1명이 음주운전을 한다.

길동에서 예상 130건을 52주로 나누어, <u>토요일마다 3건의 음주운전이 예상된다</u>. 위에 따르면 100명중 1명이 음주운전을 하는 비율로 계산하여 (3건 \* 100명) 약 300번의 음주단속을 시행</u>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, 통신장비가 필요하며, 계산 결과 2억원(연간 인건비) + 2억원 (장비비용) = 4억원으로 계산된다. 이런 단속방법을 통해서 기존 음주운전 데이터에서, 강동구의 토요일 20-24시 음주운전이 20% 감소한다고 가정하여, 정규분포 시뮬레이션을 실행하였다.

(강동구 22년 기존 음주운전 vs 시뮬레이션 예상 음주운전)
기존 2022년 8주운전
시뮬레이션 예상 음주운전
기존 음주운전 vs 시뮬레이션 예상 음주운전
기존 2022년 8주운전
지뮬레이션 예상 음주운전

- 기존의 2022년도 강동구의 음주운전 적발 수는 총 524건이었으며, 이 중 약 51건(약 10%)이 감소할 것으로 예측했다. 한국 교통연구원 2021년 교통사고비용 조사에 따르면, 소득손실 비용, 의료비용, 행정비용, 물적피해비용, 심리 비용 등 사고 1건당 평균 도로교통사고 비용을
- 5천만 원으로 산정할 때, 예상되는 음주운전 감소 효과로 인해 총 25억 원(51건 x 5천만 원)의 경제적 비용 절감이 가능할 것으로 보인다. 이는 강동구의 경찰 인력 및 장비 배치에 대한 <u>투자 비용을 충분히 상쇄하고도 남는 금액</u>으로, 음주운전 감소를 통해 **지역사회의 안전성을 크게** 향상시킬 수 있을 것이다.

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 <u>음주 단속만으로 해결하기에 한계가 있다.</u> 따라서 아래와 같은 방법도 제안한다.

- \* 분석결과로 파악한 음주운전 예상 지역에 음주 감지 장치 및 도로 카메라를 설치하여, 음주운전을 사전에 감지하고 즉 각 대응할 수 있는 시스템을 구축한다.
- \* 분석결과로 30대와 40대의 음주운전이 빈번한 지역과 20대의 음주운전이 빈번한 지역을 알 수 있었다. 이러한 <mark>각 지역의 음주운전자 대상으로, 타겟화하여 음주운전의 위험성</mark>
- 을 알리는 것이 필요하다.
   \* 또한 종사자 수가 많은 지역일수록 음주운전이 빈번하므로,
   로, 사내 교육과 회식 자리에서의 안전 관리가 필요하다.

이와 같은 **맞춤형 단속 및 예방 정책을 통해 음주운전 사고를 줄이** 고, 안전한 교통 환경을 조성할 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.