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

#### **프로젝트 개요**

**목적**: 서울시 음주운전 사고 다발지역의 특성을 분석하고, 이를 바탕으로 효과적인 음주운전 단속 지역과 시간을 선정하여 사고를 줄이는 데 기여한다.

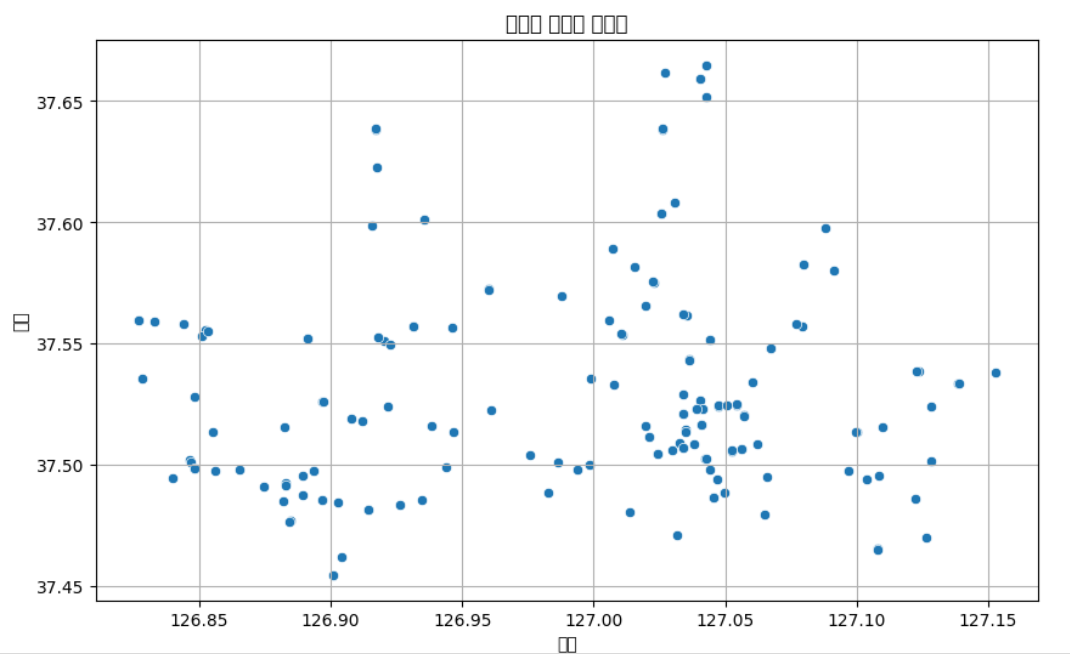
음주운전 사고의 특성을 파악하고, 각 지역별 특징과의 연관성을 검증한다

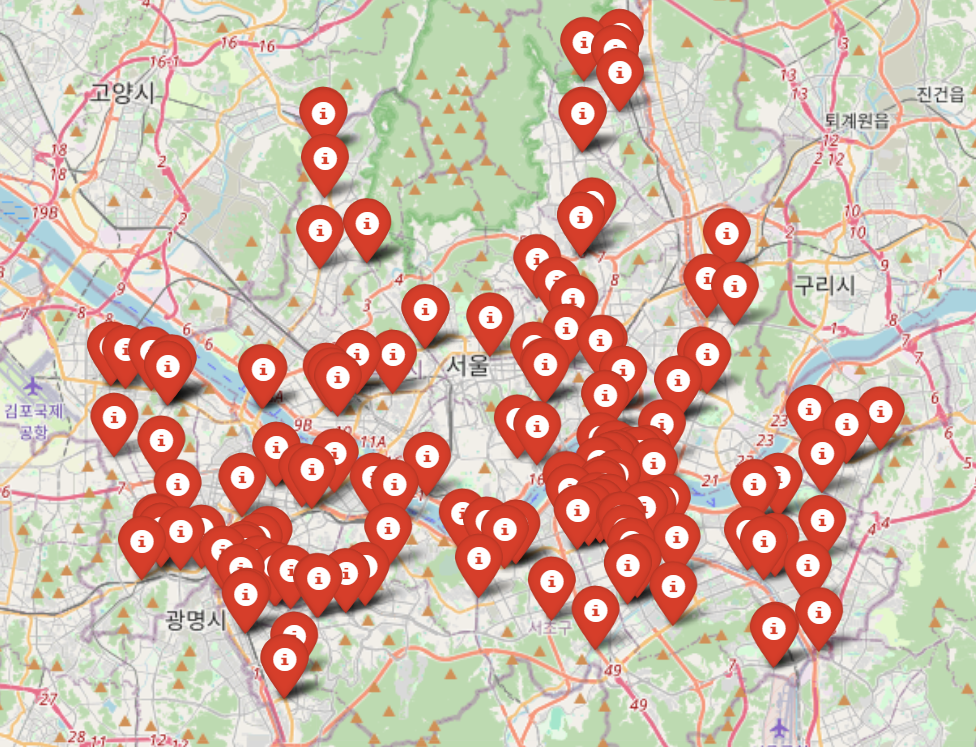
특정 요인을 파악하고, 음주운전 예상 지역을 파악하여 음주단속을 지원한다.( 이 요인이 음주사고와 관련되므로, 이런 요인이 있는 이런 지역은 -시 -요일에 집중 단속한다.)

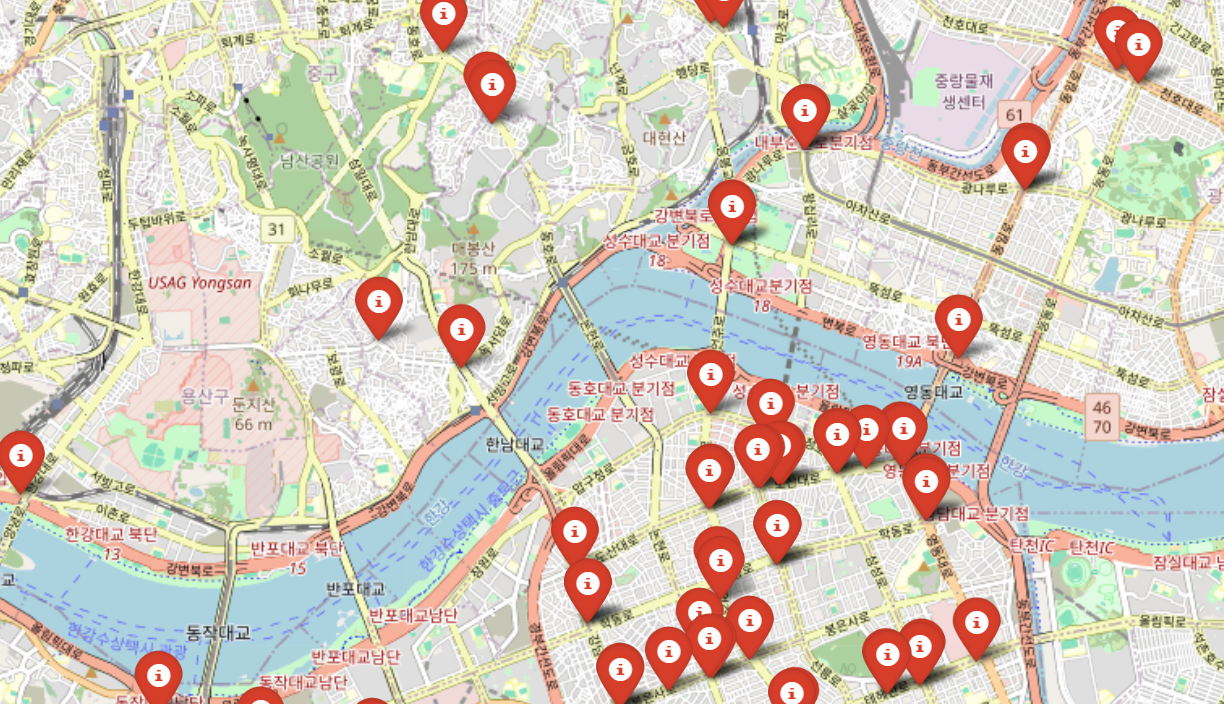
적은 예산과 시간, 인력을 통해서 효과적인 음주운전 사고를 예방한다.

#### **절차**:

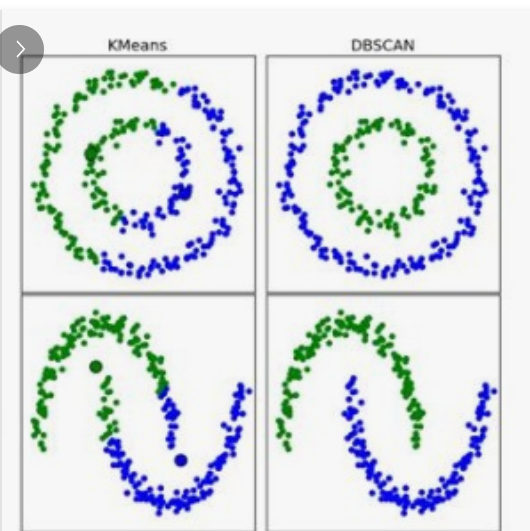
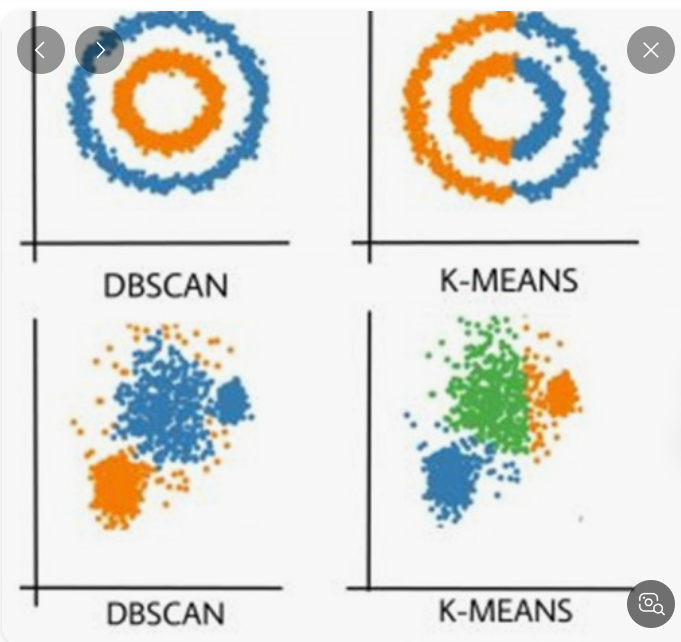
* + 음주운전 사고 다발지역을 파악한다. (**위도와 경도** 데이터로 표시 가능)

 사고지 전체 데이터(위도, 경도를 축으로 표기)

 (사고지역 **folium 시각화 샘플**) 서울 전체

 서울 강남구, 성동구 근처

* + 사고 다발지역 근처의 술집, 음식점, 주차장, 대중교통, 교통량, CCTV, 회사 건물 등의 위치와 사고 발생의 연관성을 분석한다.
  + k-means 와 DBSCAN 차이

k-means : 사고 지역을 위치 기반으로 군집

DBSCAN : 사고 지역의 특성(교통량, 음식점 수 등) 기반으로 군집

* + 분석 결과를 바탕으로 효과적인 음주운전 단속 지역과 날짜와 시간을 선정한다.
    - 예시 1: A 지역은 술집과 회사 건물이 많기 때문에 회식을 통한 음주운전의 가능성이 높고, 평일 저녁에 단속을 실시해야 한다.
    - 예시 2: B 지역은 대중교통이 많이 이용되는 곳으로, 주말 밤에 음주운전 사고가 많이 발생하므로, 주말 밤에 집중 단속을 해야 한다.
    - 예시 3: C 지역은 주차장이 많고, 사람들이 차를 이용하는 빈도가 높아 주말 낮 시간대에 단속이 필요하다.

**수정본**

1. 분석할 데이터 구조 (예상)

음주운전 사고DB

| key value | 사고지 | 위도 | 경도 | 날짜 | 요일 | 시간 | 술집  (반경 ~m) | 대중교통 | 그 외 데이터 … |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| a1 | 서울특별시 강남구 … | 소수값 | 소수값 | 5월 20일 18시 36분 | 월 | 18:36 | 19곳 | 5곳 | … |

| key value | 사고지 | 위도 | 경도 | 날짜 | 요일 | 시간 | 술집  (반경 ~m) | 대중교통 | 그 외 데이터 … |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ~~동 | 서울특별시 강남구 … | 소수값 | 소수값 | 5월 20일 18시 36분 | 월 | 18:36 | 19곳 | 5곳 | … |

단속 걸린 DB

| key value | 걸린 지역 | 위도 | 경도 | 날짜 | 요일 | 시간 | 술집  (반경 ~m) | 대중교통 | 그 외 데이터 … |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| b1 | 서울특별시 강남구 … | 소수값 | 소수값 | 5월 20일 18시 36분 | 월 | 18:36 | 19곳 | 5곳 | … |

노란색 셀은 반드시 필요한 데이터

1. 각 사고지와 음주 걸린 지역에 교집합이 있는지, 그리고 특징을 파악함.

예상 결론:

| 클러스터 | 속한 지역 | 평균 위도 | 평균 경도 | 최빈 요일 | 평균 시간대 | 평균 술집 수 | 대중교통수 (이용량) | 등 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A | a1,a3,a4 … |  |  | 월 | 21:35 | 32 | 5 | … |
| B | a2,a6,a7 … |  |  | 토 | 16:49 | 26 | 10 | … |

A 지역은 술집과 회사 건물이 많기 때문에 회식을 통한 음주운전의 가능성이 높다.

B 지역은 대중교통이 많이 이용되는 곳이다.

C 지역은 주차장이 많고, 사람들이 차를 이용하는 빈도가 높다.

1. 검증 또는 대조(해석)

: 실제 A지역에서 언제, 어느 시간대에 많이 걸렸는지, 사고가 발생했는지

| 클러스터 | 속한 지역 | 평균 위도 | 평균 경도 | 최빈 요일 | 평균 시간대 | 평균 술집 수 | 대중교통수 (이용량) | 등 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A | a1,a3,a4 … |  |  | 월 | 21:35 | 32 | 5 | … |
| B | a2,a6,a7 … |  |  | 토 | 20:49 | 26 | 10 | … |

A 지역은 평일 저녁에 사고가 많아, 회식을 통한 음주운전의 가능성이 높다.

B 지역은 주말 밤에 사고 많다. 대중교통이 많이 이용되는 시간에 사고 가능성이 높다.

C 지역은 주말 낮 시간대에 사고건수가 많다. 주말에 사람들이 차를 이용하는 빈도가 높은 것이 원인으로 보인다. 등

1. 분류 모델 생성

| 클러스터 | 평균 술집 수 | 대중교통수 (이용량) | 등 이 외 지역 데이터들 | 단속 추천 요일 | 단속 추천 시간대 |
| --- | --- | --- | --- | --- | --- |
| A | 32 | 5 | … | 평일 | 20시 이후 |
| B | 26 | 10 | … | 주말 | 18시 이후 |

1. 사고지(단속지) 특성에 따른 군집마다 단속 요일과 시간대 데이터를 배정한다. (train + test)
2. 위 (1)데이터로 분류모델을 학습한다. (train data)
3. 학습한 모델이 잘 분류 되었는지 확인한다. (test data)
4. 결론 : 이 분류 모델을 통해서 경기도, 충청도 등 다른지역에도 적용하면, 단속 구역과 시간을 파악 가능하다.

#### **2. 팀 구성 및 역할 분담**

* **공윤하 (포스터 제작)**
  + 포스터 제작
  + 관련 데이터 수집 (대중교통, 술집, 음식점, 주차장, CCTV, 회사 건물 위치 등)
  + 최종 결과 정리 및 문서화
* **권소현 (데이터 전처리)**
  + 관련 데이터 수집 (대중교통, 술집, 음식점, 주차장, CCTV, 회사 건물 위치 등)
  + 데이터 정리 및 전처리 (불필요한 데이터 제거, 결측치 처리 등)
  + 수집 데이터의 정합성 검토
* **김도연 (데이터 분석 및 시각화)**
  + 데이터 정리 및 전처리
  + 전처리된 데이터를 바탕으로 군집화 및 상관분석 수행
  + 분석 결과 시각화 (지도 위에 클러스터 표시, 산점도 등)
  + 시각화 결과 해석 및 인사이트 도출

|  | 공윤하 | 권소현 | 김도연 |
| --- | --- | --- | --- |
| 주제 선정 | 0 | 0 | 0 |
| 데이터 수집 | 0 | 0 |  |
| 데이터 전처리 |  | 0 | 0 |
| 데이터 분석 |  |  | 0 |
| 데이터 시각화 |  |  | 0 |
| 포스터 제작 | 0 |  |  |

#### **3. 프로젝트 진행 절차 (구체화)**

1. **데이터 수집 및 전처리**
   * 필요한 데이터 목록 작성
     + 대중교통 위치 데이터
     + 술집 위치 데이터
     + 음식점 위치 데이터
     + 주차장 위치 데이터
     + CCTV 위치 데이터
     + 회사 건물 위치 데이터
     + 교통량 데이터
   * 각 데이터의 수집 (**서울 지역만**, **위도 경도 데이터 필수 or 각 구별/동별 데이터**)
   * 분석할 데이터 구조 (예상)

음주운전 사고DB

| key value | 사고지 | 위도 | 경도 | 날짜 | 요일 | 시간 | 주변입지  (반경 ~m) | 술집 | 대중교통 | 그 외 데이터 … |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A1 | 서울특별시 강남구 … | 소수값 | 소수값 | 5월 20일 18시 36분 | 월 | 18:36 | 미정 | 19곳 | 5곳 | … |

단속 걸린 DB

| key value | 걸린 지역 | 위도 | 경도 | 날짜 | 요일 | 시간 | 주변입지  (반경 ~m) | 술집 | 대중교통 | 그 외 데이터 … |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| B1 | 서울특별시 강남구 … | 소수값 | 소수값 | 5월 20일 18시 36분 | 월 | 18:36 | 미정 | 19곳 | 5곳 | … |

각 사고지와 음주 걸린 지역에 교집합이 있는지, 그리고 특징을 파악함.

예상 결론:

A 지역은 술집과 회사 건물이 많기 때문에 회식을 통한 음주운전의 가능성이 높다.

B 지역은 대중교통이 많이 이용되는 곳이다.

C 지역은 주차장이 많고, 사람들이 차를 이용하는 빈도가 높다.

검증 또는 대조(해석)

: 실제 A지역에서 언제, 어느 시간대에 많이 걸렸는지, 사고가 발생했는지

A 지역은 평일 저녁에 사고가 많아, 회식을 통한 음주운전의 가능성이 높다.

B 지역은 주말 밤에 사고 많다. 대중교통이 많이 이용되는 시간에 사고 가능성이 높다.

C 지역은 주말 낮 시간대에 사고건수가 많다. 주말에 사람들이 차를 이용하는 빈도가 높은 것이 원인으로 보인다.

분류 모델 생성

사고지(단속지) 특성에 따른 군집마다 단속 요일과 시간대 데이터를 배정한다. (train + test)

위 (1)데이터로 분류모델을 학습한다. (train data)

학습한 모델이 잘 분류 되었는지 확인한다. (test data)

이 분류 모델을 통해서 경기도, 충청도 등 다른지역에도 적용하면, 단속 구역과 시간을 파악 가능하다는 결론

* + 수집된 데이터를 정리하고 전처리 (불필요한 데이터 제거, 결측치 처리 등)

1. **데이터 분석 및 시각화**
   * 전처리된 데이터를 바탕으로 군집화 및 상관분석 수행
     + 사고 다발지역 주변의 요인들(대중교통, 술집, 주차장, CCTV 등)의 수를 기준으로 클러스터링
   * 분석 결과를 시각화
     + 지도 시각화: 사고 다발지역과 각 요인의 위치 표시
     + 산점도: 각 요인과 사고 발생 건수 간의 상관관계 시각화
2. **보고서 작성 및 발표 자료 준비**
   * 데이터 수집, 전처리, 분석 결과를 받아 정리 및 요약
   * 분석 과정과 결과로 포스터 제작
3. **최종 점검 및 피드백**
   * 팀 전체: 최종 결과를 점검
   * 포스터 자료에 대한 피드백 및 수정
   * 최종 보고서와 발표 자료 제출

#### **4. 기대 효과**

* 서울시 음주운전 **사고 다발지역의 특성을 체계적으로 분석**함으로써, 사고 발생 **원인**에 대한 이해를 높인다.
* 분석 결과를 바탕으로 **효과적인 음주운전 단속 지역을 선정**하여, 음주운전 사고를 줄이는 데 기여할 수 있다.
* 프로젝트를 통해 데이터 분석, 시각화, 보고서 작성 등 다양한 역량을 개발하고 협업 능력을 향상시킨다.

#### **5. 일정 계획**

| **No** | **주요 활동** |
| --- | --- |
| 1 | 프로젝트 기획 및 역할 분담, 데이터 수집, **지원서 작성** |
| 2 | 데이터 전처리 |
| 3 | 데이터 분석 (군집화) |
| 4 | 분석 결과 시각화 |
| 5 | 포스터 작성 |
| 6 | 최종 점검 및 피드백, 자료 제출 |

#### **6. 필요 자원**

* **소프트웨어**: Python (pandas, matplotlib, seaborn, scikit-learn, folium 등)
* **데이터 소스**:
  + 서울시 공공 데이터 포털
  + 네이버 지도 API (위치 데이터 수집)

#### **7. 참고 문헌 및 자료**

* 서울시 공공 데이터 포털: https://data.seoul.go.kr
* Python 데이터 분석 도구 및 라이브러리 공식 문서
* 관련 학술 논문 및 보고서

=> 신청서: [통계데이터센터 (kostat.go.kr)](https://data.kostat.go.kr/sbchome/index.do)

분석주제(타이틀): **서울시 음주운전 사고 다발지역의 특성 분석 및 단속 방법을 제안**

분석과제 요약: 서울 지역 음주운전 사고 다발 지역을 파악하고, 각 사고지 특성에 따라 단속 방법을 제안

이용장소: 서울 중구

| 서울 중구 | 서울특별시 중구 세종대로 9길 42, 부영빌딩 7층 데이터 안심구역 (02-310-9405, 9050) |
| --- | --- |
| 서울 강남 | 서울특별시 강남구 언주로 721 서울세관 별관 4층 통계데이터 서울강남센터(02-3438-8551~2) |
| 서울 상암 | 서울특별시 마포구 매봉산로 31, 에스플렉스센터 스마티움 15층 통계데이터 서울상암센터 (02-2124-2992) |
| 서울 국회 | 서울특별시 영등포구 의사당대로1, 국회도서관 통계데이터 서울국회센터(02-6788-4416) |
| 서울대 | 서울특별시 관악구 관악로1, 서울대학교 사회과학대학 신양학술정보관(16-1동) 4층 401호(02-880-8785) |

응모분야: 포스터

참가형태/ 유형: 팀/ 대학생

분야: 정책 수립 개선 방향 제시

활용자료:

통계데이터센터 보유데이터

승인용 마이크로데이터

공공용 마이크로데이터

빅데이터 활용지표

지도로 본 빅데이터

일자리행정통계 재현자료