**# 데이터명세**

1. **서울열린데이터광장**

# **서울시 저소득 한부모가족 통계** : 저소득 한부모 가족의 현황을 제공하는 일반통계

* + 1. 저소득 한부모 가정 : 다음에 해당하는 세대주인 모 또는 부와 그에 의하여 양육되는 만 18세 미만의 자녀로 이루어진 가정
       1. 배우자와 사별 또는 이혼하거나 배우자로부터 유기된 자
       2. 정신 또는 신체장애로 인하여 장기간 근로능력을 상실한 배우자를 가진 자
       3. 미혼모(사실혼 관계에 있는 자는 제외)
       4. 배우자의 생사가 분명하지 아니한 자
       5. 배우자 또는 배우자 가족과의 불화 등으로 인하여 가출한 자
       6. 배우자의 장기복역으로 인하여 부양을 받을 수 없는 자
       7. 국내에 체류하고 있는 외국인중 대한민국 국민과 혼인하여 대한민국 국적의 아동을 양육하고 있는자자
  1. **서울시 소득인식수준 통계** : 서울시민의 가구주가 인식하는 주관적 소득수준 정도
  2. **자치구 단위 서울 생활인구(내국인)** : 자치구 단위 서울 생활인구(내국인) 정보
  3. **서울시 다문화가정 현황 통계** : 서울시 각 자치구의 초/중/고등학교의 다문화가정 학생수 정보
     1. 다문화가정 : 학생의 국적과는 상관없이 부모의 국적에 따라 국제결혼가정과 외국인가정으로 구분
     2. 국제결혼가정 : 친부모중 한 명만 외국국적을 가지고 있을 경우
     3. 외국인가정 : 친부모 두 명 모두 외국 국적인 경우
  4. **서울시 보통가구 통계** : 서울시 보통가구의 가구수
  5. **서울시 CCTV 설치현황** : 차치구 CCTV 설치 현황 관련으로 명칭, 주소, 위도/경도(좌표) 정보

1. **스마트치안빅데이터플랫폼**
   1. **112신고 영향요소 융합데이터** : 경찰관서(관할지역)별, 신고성별, 연령대, 발생요일 경찰서비스 종합112신고, 관할지역소득수준, 상주인구 데이터 통계자료 제공
   2. **청소년비행/학교폭력 영향 데이터** : 쇠퇴 및 범죄 유발 지역과 청소년비행학교폭력 연관 영향요인 데이터
   3. **학교폭력데이터** : 접수년월별 경찰관서별 신고성별 접수긴급유형별 사건종별 발생요일별 발생시간대 학교폭력 신고건수를 집계하여 통계자료로 제공
2. **학교알리미(크롤링)**
   1. 학교폭력교육시간
   2. 학교폭력실태조사 참여현황
   3. 학교폭력피해시간
   4. 학교폭력피해신고현황
   5. 학교폭력피해유형
   6. 학교폭력피해장소
3. **공공데이터포털**
   1. **경찰관서 위치 주소 현황** : 지역경찰청별 소속관서 표시, 지방청, 경찰서, 관서명, 지구대, 파출소, 주소 정보 제공
   2. **소상공인시장진흥공단 상권정보** : 영업 중인 전국 상가업소 데이터 제공(상호명, 업종코드, 업종명, 지번주소, 도로명주소, 위도, 경도 등)
4. **나이스**
   1. **학교기본정보** : 학교 기본정보에 대한 학교명, 소재지, 주소, 위경도, 전화번호, 홈페이지주소, 남녀공학여부, 주야구분, 개교기념일 등을 확인할 수 있는 정보
5. **카카오API**
   1. 경찰관서 위치 정보 위경도 매핑
   2. 학교기본정보 위경도 매핑

**# 스케일링( : 무슨 데이터로 스케일링 Scale with what)**

1. 학교폭력피해장소 : 학교폭력피해응답수(학교알리미)
2. 학교폭력피해유형 : 학교폭력피해응답수(학교알리미)
3. 학교폭력피해시간 : 학교폭력피해응답수(학교알리미)
4. 학교폭력교육시간 : 각 학교 1학급당 평균 피해시간이므로 스케일링 진행 안함
5. 학교폭력피해신고현황 : 학교폭력피해응답수(학교알리미)
6. 학교폭력실태조사 - 학교폭력피해응답수 : 고등학교 각각의 학교의 학생 수
7. 한부모가족지원법\_수수급자\_재가보호\_가구원수 : 서울특별시 자치구 주민등록상 인구
8. 국민기초생활보장법수급자\_가구원수 : 서울특별시 자치구 주민등록상 인구
9. 소득인식수준\_중간 / 소득인식수준\_고소득 : 각 자치구의 평균이므로 스케일링 진행 안함
10. 생활인구 : 각 자치구의 고등학교 생활인구 평균이므로 스케일링 진행 암함
11. 다문화가족 : 서울특별시 자치구 가구수
12. 단란주점수,성인게임장수, 술집수, 유흥주점수 : 각 자치구의 개별 특성이므로 스케일링 진행 안함
13. 총 신고수, 청소년비행신고수, 학교폭력신고수 : 고등학교 각각의 학교의 학생 수
14. 사건종별가정폭력 : 서울특별시 자치구 가구수
15. 사건종별데이트폭력 : 서울특별시 자치구 주민등록상 인구
16. 월평균소득금액 : 각 자치구의 평균이고 개별특성이므로 스케일링 진행 안함
17. 남녀공학구분명(남고,여고,남녀공학) : 서울특별시 자치구 학교 수
18. 고등학교구분명(일반고, 특성화고, 자율고, 특목고) : 서울특별시 자치구 학교 수
19. 고등학교일반실업구분명(일반계, 전문계) : 서울특별시 자치구 학교 수

**# 환경지표, 위해지표, 경감지표로 분류 : Reason**

1. 학교폭력발생에는 다양한 원인이 존재한다. 심각한 피해를 일으키는 학교폭력을 예방하기 위해서는 학교폭력의 근본적인 원인을 파악하는 것이 필수이다. 한 개인이 어떤 사람으로 자라며 어떻게 행동하는가는 그 사람의 의식 구조에 따라 결정되는 수가 많으며 그 의식 구조의 형성에는 개인이 자라온 가정, 환경, 학교, 사회의 환경에 의해서 결정된다. 따라서 학교폭력의 사회 환경적 요인에 환경지표, 학교폭력의 학교 환경적 요인으로 위해지표, 경감지표로 분류하였다.
   1. 각 자치구별 환경적 요인이 학교폭력위험도에 영향을 미칠 것이라 파악
   2. 개별 고등학교 특성이 학교폭력위험도에 영향을 미칠 것이라 파악
      1. 학교폭력 발생에 positive한 영향을 주는 요소와 학교폭력 발생에 negative한 영향을 주는 요소

**# 환경지표(학교폭력의 사회 환경적 요인)**

**학교폭력위험도에 부정적 영향을 미치는 서울특별시 자치구 환경을 반영하여 점수를 산정한 상대적 지표이다.**

1. 학교폭력신고수& 학교폭력피해응답수와 자치구별 학교폭력위험도에 영향을 미치는 변수들과의 상관관계 파악
2. 상관분석 결과를 바탕으로 유의미한 변수 선정 : 상관계수
   1. 한부모가족지원법수급자\_재가보호\_가구원수 : 0.569231
   2. 국민기초생활보장법수급자\_가구원수 : 0.641538
   3. 소득인식수준\_중간 : 0.511252
   4. 소득인식수준\_고소득 : -0.241538
   5. 성인게임장수 : 0.324399
   6. 사건종별가정폭력 : 0.693846
   7. 사건종별데이트폭력 : 0.480769
   8. 월평균소득금액(백만) : -0.456154
   9. 다문화가족 : 0.370000
3. kmeans 군집화 진행
   1. Standard Scaler 사용
   2. Elbow Method / Silhouette Score 을 고려하여 최적 k(군집 수) 선정 -> k=5
   3. 군집화 결과
      1. 0번 군집 : 학교폭력위험도 낮음
      2. 1번 군집 : 학교폭력위험도 매우 높음
      3. 2번 군집 : 학교폭력위험도 중간
      4. 3번 군집 : 학교폭력위험도 매우 낮음
      5. 4번 군집 : 학교폭력위험도 높음
   4. 상관계수의 크기를 바탕으로 각 자치구의 학교폭력위험도에 영향을 미치는 변수에 가중치를 부여하여 학교별 환경지표 점수 계산
   5. 환경지표는 서울특별시 자치구의 환경적 특성이 학교폭력위험도에 미치는 영향에 대한 상대적 지표로 같은 자치구에 위치하고 있는 고등학교의 환경지표 score, cluster, 학교폭력위험도는 같다.

**# 상관분석 : Reson**

1. 상관분석(Correalation Analysis)란 두 변수간의 선형의 관계가 있는지에 대한 분석으로 상관분석의 결과로 도출되는 상관계수는 “-1~+1”의 범위를 가지는 숫자로 두 변수의 직선성의 관계를 표현한다.
2. 상관계수가 크면 두 변수간의 직선성이 매우 높고, 상관계수가 낮으면 두 변수간의 직선성이 약하다는 것을 의미한다.
3. 학교폭력발생에 영향을 미치는 사회 환경적 요인의 변수를 파악하기 위해 상관분석 진행
4. 사회 환경적 요인의 변수의 직선성 파악(높은지, 낮은지)
5. 유의미한 사회 환경적 요인의 변수를 파악하여 학교폭력 위험도에 높은 수준의 상관관계가 있는 변수 선택

**# 위험도 분류 Kmeans : Reason**

1. 분류 모델에는 Kmeans, Decision Tree, Random Forest 등이 존재, 주된 목표가 훈련 데이터를 기반으로 새로운 데이터의 레이블을 예측하는 것이 아닌 데이터를 그룹으로 클러스터링 하는 것이기 때문에 비지도학습인 Kmeans를 채택
2. 또한 단순 군집화가 목적이므로 확장성, 해석 용이성, 유연성이 우수한 Kmeans를 채택
3. 다른 군집화 모델에 비해 KMeans가 적절

**# Standard Scaler 사용 : Reason**

1. scale이 큰 feature의 영향이 비대해지는 것을 방지

**# TSNE(Kmeans에서 차원 축소) : Reason**

1. 3개 이상의 변수를 2차원에 나타내고, 데이터들간의 상대적 거리를 보존하기 위해 사용

**# 학교폭력위험도 가중에 영향을 미치는 환경지표에 사용한 변수들에 대한 근거 : 논문**

* **한부모가족지원법수급자\_재가보호\_가구원수**

1. 김영미. (2016). 청소년의 폭력피해경험과 심리적 적응 및 비행문제. 한국콘텐츠학회논문지, 16(1), 470-481.
2. 김희선. "생태학적 관점에서 학교폭력 유발요인이 학교폭력에 끼치는 영향." 국내석사학위논문 이화여자대학교 교육대학원, 2014. 서울
3. 박진희. “학교폭력에 영향을 미치는 지역사회요인의 상호작용 효과 분석.” 박사학위논문, 2016

* **국민기초생활보장법수급자\_가구원수**

1. 김희선. "생태학적 관점에서 학교폭력 유발요인이 학교폭력에 끼치는 영향." 국내석사학위논문 이화여자대학교 교육대학원, 2014. 서울
2. 박진희. “학교폭력에 영향을 미치는 지역사회요인의 상호작용 효과 분석.” 박사학위논문, 2016

* **소득인식수준\_중간\_고소득**

1. 치안정책연구소 범죄수사연구실. (2013). 학교폭력 가해행동에 영향을 미치는 요인에 관한 연구: 가정적 요인을 중심으로
2. 박진희. “학교폭력에 영향을 미치는 지역사회요인의 상호작용 효과 분석.” 박사학위논문, 2016

* **청소년유해업소 : 성인게임장수**

1. 김원영 and 김경식. (2018). 한국 청소년의 학교폭력 가해행동 관련 유발변인에 대한 메타분석. 중등교육연구, 66(3), 783-814.

* **사건종별가정폭력**

1. 김형희, 김용섭. (2022). 청소년의 학대경험이 학교폭력가해경험에 미치는 영향에 대한 종단연구: 학교폭력피해경험과 공격성의 매개효과를 중심으로. 한국컴퓨터정보학회논문지 , 27(8), 203-210.
2. 김재엽, 이지현, 정윤경. (2008). 청소년들의 가정폭력노출경험이 학교폭력가해행동에 미치는 영향. 한국아동복지학, (26), 31-59.

* **사건종별데이트폭력**

1. 이화숙. "대학생의 가정폭력 및 학교폭력 경험과 데이트폭력 가해와의 관계에서 젠더감수성의 조절효과." 국내박사학위논문 신라대학교 일반대학원, 2016. 부산
2. 김동기 and 사공은희. (2007). 학교폭력 피해경험이 데이트폭력 가해행동에 미치는 영향과 사회적 지지의 조절효과에 관한 연구. 청소년학연구, 14(5), 99-124.
3. 조상현. (2020). 데이트폭력의 발생 원인에 관한 연구. 한국범죄심리연구, 16(3), 165-176.

* **월평균소득금액(백만)**

1. 김영미. (2016). 청소년의 폭력피해경험과 심리적 적응 및 비행문제. 한국콘텐츠학회논문지, 16(1), 470-481.
2. 김희선. "생태학적 관점에서 학교폭력 유발요인이 학교폭력에 끼치는 영향." 국내석사학위논문 이화여자대학교 교육대학원, 2014. 서울
3. 박진희. “학교폭력에 영향을 미치는 지역사회요인의 상호작용 효과 분석.” 박사학위논문, 2016

* **다문화가족**

1. 김도원(Do Won Kim);문인수(In Soo Moon). (2022). 다문화 청소년 학교폭력 피해 영향요인. 한국범죄심리연구, 18(3), 37-48. 10.25277/KCPR.2022.18.3.37
2. 부민서, 심혜인 and 조윤오. (2018). 다문화가정 자녀의 취약성과 보호자 특성이 학교폭력 피해경험에 미치는 영향. 한국범죄학, 12(3), 123-140.

**# 위해지표(학교폭력의 학교 환경적 요인)**

**학교폭력위험도 가중에 영향을 미치는 서울특별시 고등학교별 위해 요소를 반영하여 점수를 산정한 상대적 지표이다.**

1. 학교 500m이내 유흥업소 개수, 학교폭력피해장소(교실안), 학교폭력피해유형(신체폭력, 집단따돌림), 학교폭력피해시간(쉬는시간)을 학교폭력위험도 가중에 영향을 미치는 요소로 선정
2. kmeans 군집화 진행
   1. Standard Scaler 사용
   2. Elbow Method / Silhouette Score 을 고려하여 최적 k(군집 수) 선정 -> k=6
   3. 군집화 결과 : 군집화 결과에 따른 위해지표 점수
      1. 0번 군집 : 학교폭력위험도 매우 안전 : 0
      2. 1번 군집 : 학교폭력위험도 매우 위험 : 100
      3. 2번 군집 : 학교폭력위험도 주의 : 60
      4. 3번 군집 : 학교폭력위험도 위험 : 80
      5. 4번 군집 : 학교폭력위험도 보통 : 40
      6. 5번 군집 : 학교폭력위험도 안전 : 20

**# 경감지표(학교폭력의 학교 환경적 요인)**

**학교폭력위험도 경감에 영향을 미치는 서울특별시 고등학교별 경감 요소를 반영하여 점수를 산정한 상대적 지표**

1. 학교 500m이내 cctv개수, 학교 500m이내 경찰관서 개수, 학교폭력예방교육시간을 학교폭력위험도 경감에 영향을 미치는 요소로 선정
2. kmeans 군집화 진행
   1. Standard Scaler 사용
   2. Elbow Method / Silhouette Score 을 고려하여 최적 k(군집 수) 선정 -> k=5
   3. 군집화 결과 : 군집화 결과에 따른 경감지표 점수
      1. 0번 군집 : 학교폭력위험도 매우 위험 : 20
      2. 1번 군집 : 학교폭력위험도 매우 안전 : 100
      3. 2번 군집 : 학교폭력위험도 안전 : 80
      4. 3번 군집 : 학교폭력위험도 위험 : 40
      5. 4번 군집 : 학교폭력위험도 보통 : 60

**# 종합지표(환경지표 + 위해지표 - 경감지표)**

**서울특별시 내 고등학교 별 사회 환경적요인과, 학교 환경적 요인에 대한 학교폭력위험도 가중에 영향을 미치는 환경지표와 위해지표, 학교폭력위험도 경감에 영향을 미치는 경감지표를 반영하여 지표 생성**

1. 환경지표, 위해지표, 경감지표가 학교폭력위험도에 차지하는 비중이 다르므로 상관계수의 크기로 가중치를 부여하여 고등학교 개별 학교폭력위험도 score 생성
   1. 환경지표 : 0.457842 가중치 부여
   2. 위해지표 : 0.263726 가중치 부여
   3. 경감지표 : 0.278432 가중치 부여
2. score 0~100 점수로 스케일링 이유 : 일반인들이 이해함에 있어 편리함을 위해(학교폭력위험도에 대한 판단이나 의사결정에 있어 0~100점으로 스케일링 진행하는 것이 일반인들에게 이해가 편하다고 판단)
3. score 기준 min 20% 40% 60% 80% max 백분위수 파악하여 고등하교 개별 학교폭력위험도 파악
   1. min 0.000000
   2. 20% 40.477257
   3. 40% 51.627009
   4. 60% 61.481197
   5. 80% 71.747387
   6. max 100.000000
   7. 학교폭력위험도 매우 위험 : score > 71.747387
   8. 학교폭력위험도 위험 : score > 61.481197
   9. 학교폭력위험도 보통 : score > 51.627009
   10. 학교폭력위험도 안전 : score > 40.477257
   11. 학교폭력위험도 매우 안전 : else

**# 종합지표 학교 total score**

1. 학교폭력의 사회 환경적 요인에 해당하는 환경지표와 학교폭력의 학교 환경적,개인적 요인에 해당하는 위해지표, 경감지표 각 지표의 점수를 바탕으로 환경지표+위해지표-경감지표로 계산하여 학교폭력 위험점수 계산(ex. 위험점수가 높을수록 학교폭력위험도가 높음)

**# 종합지표 학교폭력위험도**

1. 학교폭력 위험점수를 바탕으로 백분위수를 파악하여 5단계로 분류(매우위험, 위험, 보통, 안전, 매우안전)

**# 학교폭력 종합지표의 학교폭력위험도를 5단계 : Reason**

1. 교육부 학교폭력 조치별 적용 세부 기준 고시 자료에서 기본 판단 요소를 5단계로 분류함을 참고하여 학교폭력위험도를 5단계로 분류

(<https://drive.google.com/file/d/1-LQs5bO5v6rIS4X00MnIMrizRiaytArt/view?usp=sharing>)

**# YOLO**

1. 학교폭력 실시간 CCTV 객체탐지를 위해 Yolo v5를 사용
   1. Yolo v5 : 다른 객체 탐지 모델에 비해 속도가 빠르고 상대적으로 light하여 리소스가 제한된 장치에 적응하기 쉽다는 장점이 있음.
2. Yolo 학습 데이터에는 roboflow Violence Detection dataset 1500개 이미지 사용

(<https://universe.roboflow.com/nuscrimesocietydatasets/violence-9gmjx>)