

주제 : 초중고 학생들의 건강관리

소주제 1. 학생들의 올바른 식습관과 운동량을 분석 후, 청소년 관리 예측

컬럼사용 : ['학년','음료수' , '년도','우유_유제품','하루30분이상운동','가족흡연' , '아침식사' , '하루수면량','채소(김치제외)','패스트푸드','하루2시간이상게임','라면','몸무게','하루TV시청2시간이상','육류','가족음주' , '주3회이상운동' , '키','자아신체상(체형)','과일','다이어트경험','건강검진_종합소견']

소주제 2. 수면량과 음식 섭취 종류에 따른 자아신체상(체형) 예측

컬럼사용 : [소주제1 컬럼 + '괴롭힘따돌림_초', '괴롭힘따돌림_중고', '가출생각_초', '가출생각_중고', '상담요청_초', '상담요청_중고' 컬럼을 '괴롭힘따돌림', '가출생각', '상담요청']

소주제 3. 집,학교에서 주는 외/내적 영향력이 학생들의 건강에 미치는 효과

컬럼사용 :

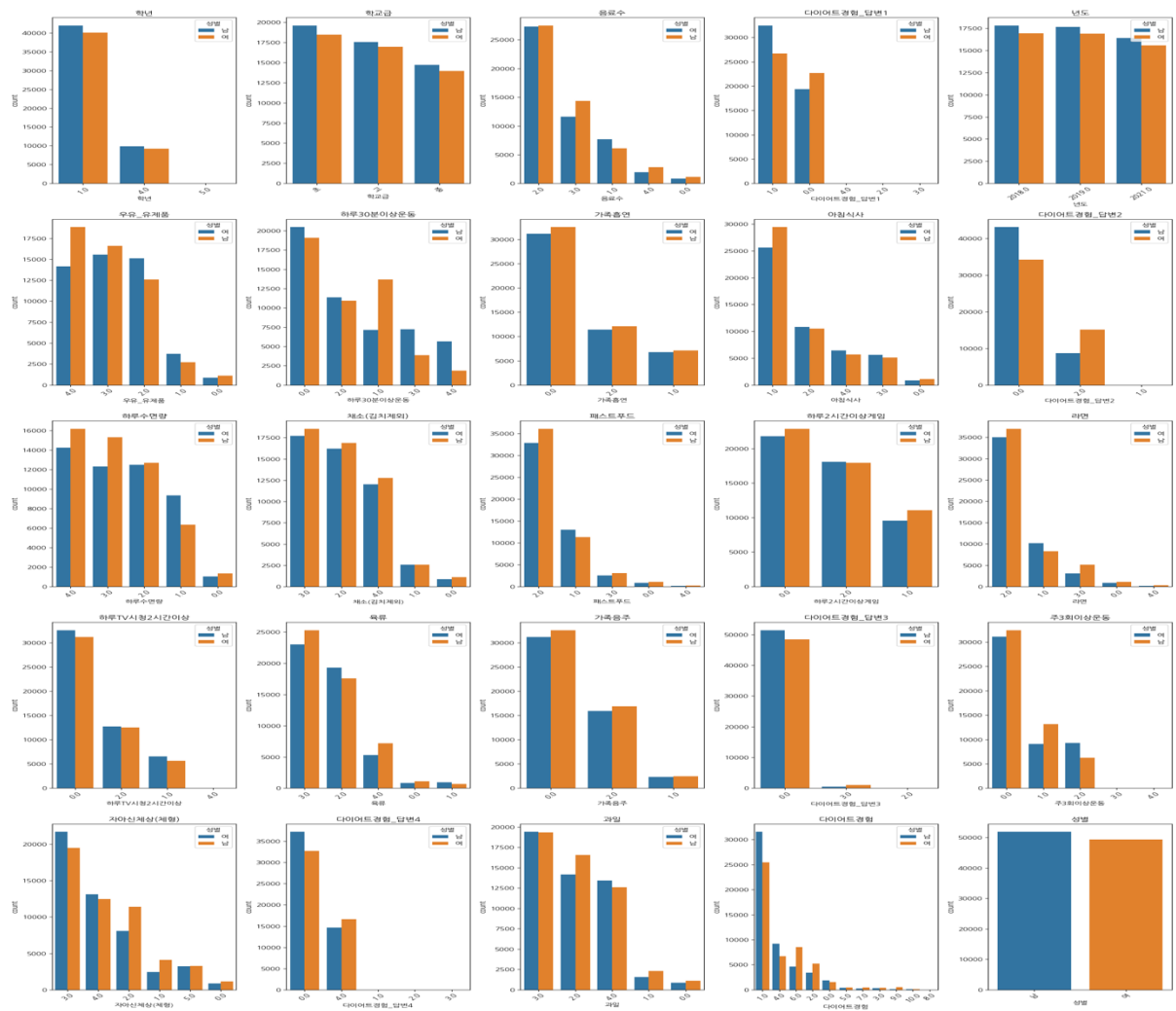
elementary : [학교급, 건강검진_종합소견, 괴롭힘따돌림_초, 현금갈취, 체벌경험, 신체접촉, 무기력감, 수업태도교정, 가출생각_초, 가족흡연, 가족음주, 가족지지, 상담요청_초]

mid_high : [학교급, 건강검진_종합소견, 괴롭힘따돌림_중고, 폭력위협, 가출생각_중고, 가정문제걱정, 상담요청_중고, 고민상담대상, 상담희망, 흡연_음주전문가상담희망, 성문제전문가상담희망, 진로고민]

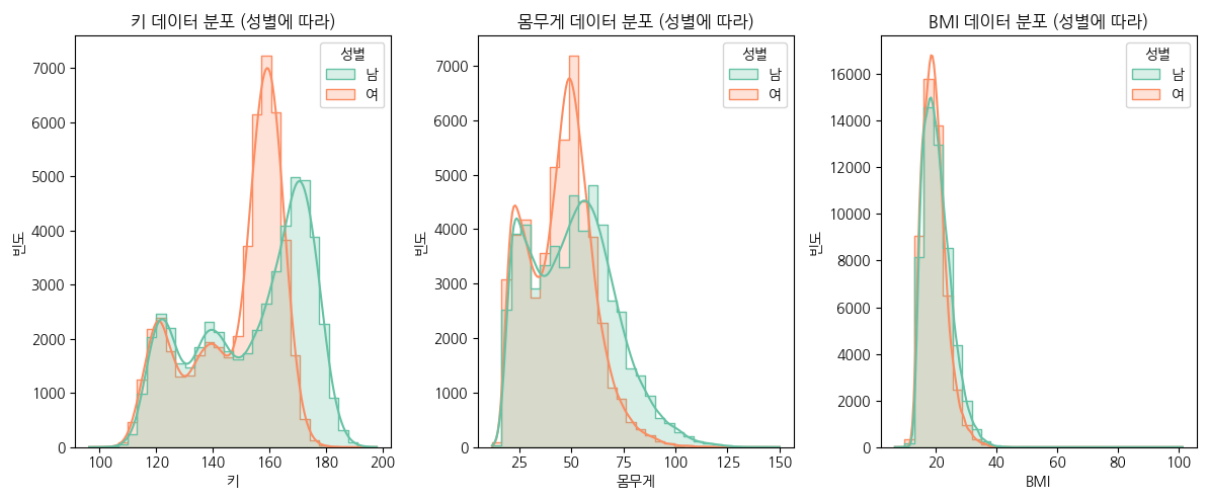
[데이터 결합 + 전처리] : (101245 X 84) 84개의 컬럼, 101245개의 행

1. 2018,2019,2021 데이터프레임에서 공백(' ')값 → NaN 결측치 값으로 변경 → '건강검진_종합소견' 컬럼에서 null 값 삭제
2. 2018,2019,2021 각각의 데이터프레임에 년도 추가
3. 2018,2019,2021 데이터프레임에서 공통으로 가지고 있는 컬럼 값만 가지고 세개의 프레임 하나로 합침
4. 키, 몸무게 null 값인 index 동일 확인 후 → 키 null 값인 데이터 삭제 (키랑 몸무게 결측치 유추해서 채울 수 없으므로 제거)
5. 데이터타입 변경 (모두 object) -> 결측치 0으로 바꾼 후 → 성별, 건강검진_종합소견, 시도, 도시규모, 학교급,충치치아_유무 + col2 리스트 값 : object → 혈당(식전), 키, 혈색소, ALT, AST 등등 나머지 : float
6. 다이어트경험_답변1, 다이어트경험_답변2, 다이어트경험_답변3, 다이어트경험_답변4 합쳐서 다이어트경험 컬럼 생성
7. 현금갈취.1 컬럼과 같이 컬럼명 뒤에 붙은 .1삭제 (현금갈취.1 → 현금갈취)

[남녀 구분 지어 각 칼럼별 비율 확인



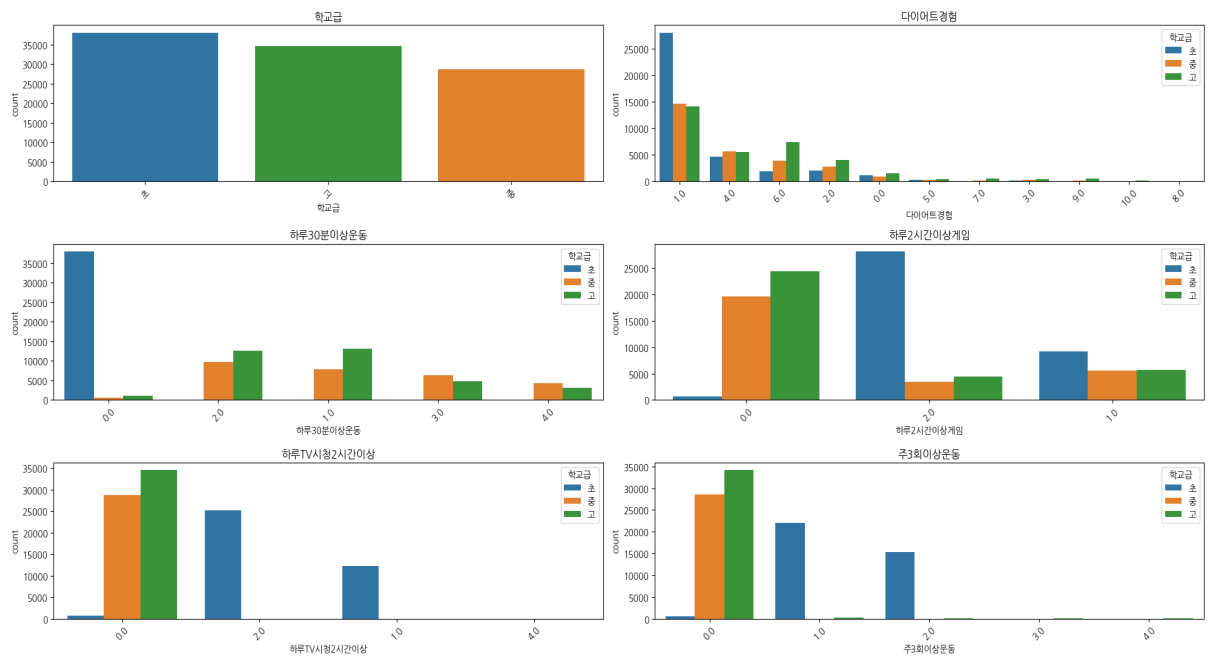
남녀 별 키, 몸무게, BMI 분포 확인



[학교급 별 식습관 확인]



[학교급 별 생활습관 확인]



[건강검진_소견 별 청소년 심리관련 칼럼 확인]

