

[천재교육XSeTEC]

에듀테크 PM과정 2024학년도 1차 데이터 연구 보고서

스트레스와 학업 성취도의 상관관계 : 에듀테크의 방향성 제시

1팀

김도영, 류예나, 변인섭, 신하연, 이유빈, 윤나현, 허청

보고서 제출일

2024.12.30

목 차

I. 서론

1)연구 주제 및 배경 설명

2)데이터 출처 및 소개

(1)사용한 데이터셋

(2)탐색적 데이터 분석(EDA)

II. 본론

1)가설 설정

2)가설 설정 배경

3)분석 방법

(1)이상치와 결측치

(2)데이터 형 변환 및 범주형 데이터 처리

4)가설 검증

(1)변수들 간의 상관관계 분석

(2)가설 검증

III. 결론

1)결과 요약

2)시사점 및 기대효과

3)한계 및 향후 연구 방향

I.서론

1)연구 주제 및 배경 설명

코로나 19 팬데믹 이후, 디지털 기술과 인공지능(AI)이 일상생활 전반에 빠르게 확산되었다. 사람들은 영화관 대신 집에서 OTT 서비스를 이용하며 여가를 즐기고, AI 기술로 업무와 학업을 효율적으로 수행하는데 익숙해졌다. 이러한 변화는 교육 분야에서 두드러지게 나타났다. 그 중심에 에듀테크(EduTech)가 있다. 에듀테크는 교육(Education)과 기술(Technology)이 결합한 단어로 전통적인 학습 방식에서 벗어나, AI 기반 디지털 교과서와 게임화된 학습 콘텐츠로 학습자에게 흥미를 유발한다. 그러나 에듀테크 산업은 아직 초기 단계에 머물러 있어, 관련 연구가 충분히 이루어지지 않았다. 추후 에듀테크의 효과를 검증하고 개선점을 도출하기 위해서는 학업 성취도에 영향을 미치는 다양한 요소를 면밀히 분석할 필요가 있다.

특히, 학업과 스트레스 간의 상관관계는 교육학 및 심리학 분야에서 오랜 논의의 주제였으며, 에듀테크 산업에서도 이러한 관계를 파악하여 적용하는 것은 중요한 연구 과제로 꼽힌다. 스트레스는 학생들의 학업 성취도에 긍정적 혹은 부정적 영향을 미칠 수 있는 중요한 변수다. 스트레스에 따라 학업 성적이 어떤 변화를 가 통찰을 제시하고자 한다. 궁극적으로, 본 연구는 에듀테크가 학습자의 학업 성취를 극대화하면서도 스트레스여오는지 알기 위해 학생들의 일상 데이터를 분석하여 해답을 찾고자 한다. 따라서 이 보고서는 학업과 스트레스의 상관관계를 분석하고, 에듀테크가 이러한 관계에 미칠 수 있는 영향을 탐구하는 것을 목적으로 한다. 분석하고자 하는 데이터는 2023년 8월 인도 학생의 일상 생활을 전반적으로 기록해 데이터로 만든 자료를 바탕으로 한다. 이를 통해 학업 스트레스의 긍정적, 부정적 효과를 균형 있게 이해하고, 에듀테크가 학습 환경에서 어떻게 활용될 수 있는지에 대한을 완화할 수 있는 방향으로 발전할 수 있는 근거를 마련하는 데 기여할 것이다.

2)데이터 출처 및 소개

(1)사용한 데이터셋

본 보고서에서 분석한 데이터셋은 “Student lifestyle dataset_Daily Lifestyle and Academic Performance of Students¹”이다. 이 데이터셋은 공부시간, 과외활동, 수면, 사교, 신체활동, 스트레스 수준 및 CGPA*에 대한 정보를 포함하고 있다. 학업에 영향을 미치는 요소들에 대한 데이터를 분석해보고자 선택하였다.

(2)탐색적 데이터 분석(EDA)

본 데이터셋에서 제공된 데이터프레임을 활용하여 학업 성취도(GPA)에 영향을 미칠 수 있는 요인들과의 상관관계를 탐색하고자 한다. 해당 데이터셋에서 제공한 변수(컬럼)는 Study hours per day(학업시간), Extracurricular Hours Per Day(교외활동 시간), Sleep Hours Per Day(수면시간), Social Hours Per Day(사회활동 시간), Physical Activity Hours Per Day(신체활동 시간), Stress Level(스트레스 수준), GPA(학업 성적)이다.

- 데이터 컬럼

¹ Student lifestyle dataset_Daily Lifestyle and Academic Performance of Students
학생 생활방식 데이터셋_학생들의 일상생활과 학업 성취도 (SUMIT KUMAR /2024.10 /kaggle)

- Student_ID: 학생 고유 ID
- Study_Hours_Per_Day: 하루 학습 시간
- Extracurricular_Hours_Per_Day: 하루 교외 활동 시간
- Sleep_Hours_Per_Day: 하루 수면 시간
- Social_Hours_Per_Day: 하루 사회 활동 시간
- Physical_Activity_Hours_Per_Day: 하루 신체 활동 시간
- Stress_Level: 스트레스 수준
- GPA: 학업 성적 (Grade Point Average).

해당 데이터셋에서 제공한 주요 수치형 데이터 통계량은 다음과 같다. 평균 공부 시간은 7.48시간, 평균 수면 시간은 7.50시간 등 시간을 파악할 수 있었고, 학업 점수인 GPA 평균 점수는 3.12으로 확인된다. 추가로 평균, 분포, IQR을 통해 수치를 알아보하고자 한다.

- 주요 수치형 데이터 통계량(평균, 중앙값, 최소, 최대 등)
 - Study_Hours_Per_Day (공부 시간)
 - 평균 공부 시간: 7.48시간
 - 분포: 최소 5시간에서 최대 10시간
 - IQR(25%~75%): 하루 6.3~8.7시간 공부
 - Sleep_Hours_Per_Day (수면 시간)
 - 평균 수면 시간: 7.50시간
 - 분포: 최소 5시간에서 최대 10시간
 - IQR(25%~75%): 하루 6.2~8.8시간 수면
 - GPA (학업 성적)
 - 평균 GPA: 3.12
 - 분포: 최소 2.24에서 최대 4.00
 - IQR(25%~75%): 대부분 2.9~3.33 성적 범위

이상치와 결측치를 확인하여 데이터 처리에 오류가 없도록 하였다. 결측치는 없는 것으로 확인된다. 또한 데이터를 변환하여 범주형으로 처리하였다.

- 이상치와 결측치 확인 및 처리 전략 결정
 - 주요 컬럼인 Stress_Level의 결측치 및 이상치가 발견되지 않음.
- 결측치와 이상치 처리 및 결과
 - 결측치 확인 : Study hours per day(학업시간), Sleep Hours Per Day(수면시간), Stress Level(스트레스 수준), GPA(학업 성적)의 결측치가 발견되지 않음.
- 데이터 형 변환 및 범주형 데이터 처리
 - Stress_Level(스트레스 수준)을 범주형 데이터로 변환함.
 - 데이터 타입을 확인했을시, 결과값이 category로 나와, 범주형 데이터로 변환되었다는 것을 알 수 있음.

- Stress_Level(스트레스 수준)의 Low<Moderate<High의 순서를 명시하기 위해서 범주형 데이터를 변환하였음. Low는 1, Moderate는 2, High는 3으로 변환함. 해당 값들은 고정된 범주를 가지기 때문에 변환해줘야 함.

II. 본론

1)가설 설정

스트레스와 학업의 상관관계를 분석하기 위해 다음의 가설을 설정하였다.

가설: 적절한 스트레스는 학업성취도에 효과적이다.

위 가설에 대한 영향을 구체적으로 확인하기 위해 다음과 같은 세부 가설을 설정하였다.

- (i) 스트레스가 있으면 학업 성적이 높을 것이다.
- (ii) 스트레스가 있으면 학업 시간이 높을 것이다.
- (ii-i) 학업 시간이 높으면, 학업 성적이 높을 것이다.
- (번외) 수면 시간이 적으면, 학업 성적이 높을 것이다.

위 두 가지 세부 가설을 통해 스트레스와 학업 성취 간의 관계를 입체적으로 분석하고자 한다. 가설들을 바탕으로, 스트레스가 학업 성적과 학업 시간에 미치는 영향을 검증한다. 스트레스가 학습 동기와 성취에 유의미한 영향을 미칠 수 있는지 살펴보고, 가설이 맞는지 파악하고자 한다. 또한 학업 시간 증가가 실제로 성적 향상으로 이어지는 지에 대한 분석을 통해 실질적인 가설들의 연관성을 찾고자 하였다.

추가로 “수면 시간이 적으면, 학업 성적이 높을 것이다” 가설을 연구하고자 한다. 비록 본 연구의 주제와 직접적인 연관성은 없지만, 수면 시간과 학업 성적 간의 상관관계는 일반적으로 많은 관심을 받는 주제다. 또한, 이번 데이터 탐색 과정에서 신뢰도 높은 결과를 도출할 수 있었기에 부가적으로 해당 요인에 대한 분석을 진행하였다.

2)가설 설정 배경

모든 사람은 크고 작은 스트레스와 싸우며 살아간다. 스트레스는 누구나 겪는 자연스러운 현상이지만, 이를 효과적으로 해소하고 관리하는 일은 결코 쉽지 않다. 스트레스의 원인은 개인에 따라 다양하지만, 학업 성취를 위해 장기간 공부에 몰두하는 것도 중요한 스트레스 요인 중 하나다. 학업 스트레스는 단순히 피로감을 느끼는 것에서 그치지 않고, 집중력 저하, 학습 효율 감소, 심리적 불안과 같은 부정적인 영향을 미친다. 이러한 영향은 궁극적으로 학업 성취도 저하로 이어질 수 있다. 그럼에도 불구하고 많은 사람들은 더 나은 성적을 위해 학업 시간을 늘리고, 여가를 포기하는 것이 당연하다고 여긴다. 하지만 스트레스를 무작정 감수하는 것이 정말 학업 성취도를 높이는 유일한 길일까.

이러한 의문에서 출발하여 이 보고서는 스트레스와 학업 성취 간의 상관관계를 심층적으로 분석하고자 한다. 스트레스의 높고 낮음이 학습 과정과 결과에 어떤 영향을 미치는지 탐구한다.

3)데이터 전처리

(1)이상치와 결측치

Stress_Level 변수는 “Low”, “Moderate”, “High”의 세 가지 범주로 구분되어 있으며, 이 범주에 속하지 않는 데이터는 이상치로 간주될 수 있다. 데이터를 점검해본 결과 세 가지 범주에서 벗어나는 이상치는 존재하지 않았다.

그 외 변수들 간의 이상치는 유의미한 결과로 데이터 분석 시 사용될 예정이다. 결측치 확인 결과 다음과 같이 결측치가 발견되지 않았다.

Study_Hours_Per_Day	0
Sleep_Hours_Per_Day	0
GPA	0
Stress_Level	0

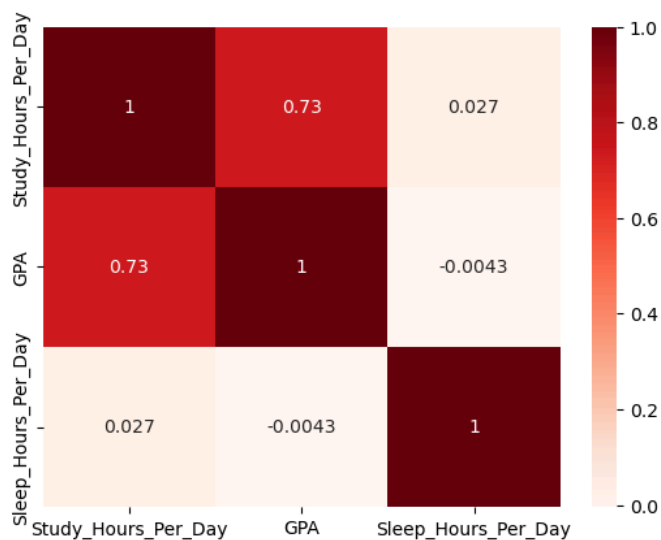
(2)데이터 형 변환 및 범주형 데이터 처리

본 분석에서 핵심 변수인 Stress_Level의 “Low”, “Moderate”, “High”을 범주형 데이터로 명시하기 위해, 범주형 데이터 처리를 진행했다.

4)가설 검증

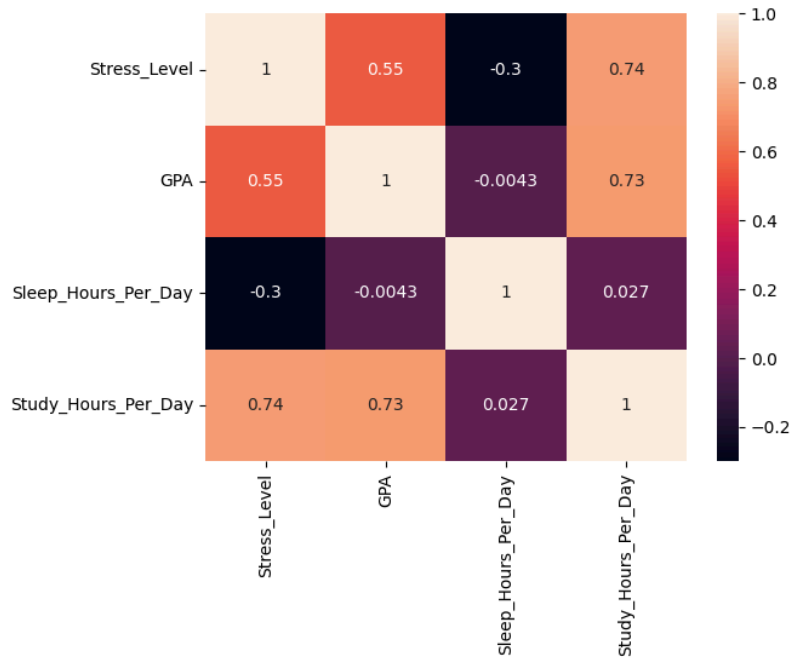
(1)변수들 간의 상관관계 분석

i)수치형 데이터² 간의 상관관계 분석



ii)Stress_Level을 수치형 데이터로 변환한 후, 기존의 수치형 데이터와의 상관관계를 분석

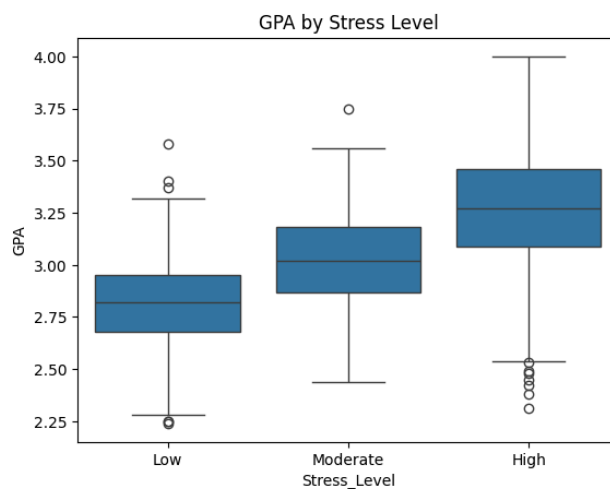
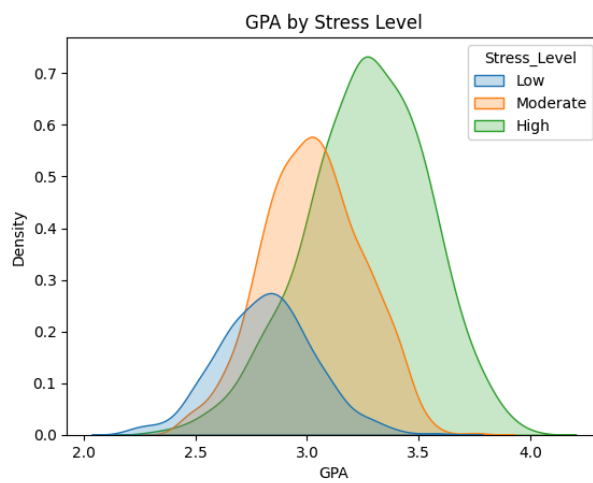
² 수치형 데이터: Study_Hours_Per_Day(학업시간), GPA, Sleep_Hours_Per_Day(수면시간)



(2)가설 검증

i)스트레스가 있으면 학업성적이 높을 것이다.

①Stress_Level과 GPA의 관계 (KDE plot, Box plot)



Stress_Level과 Study_Hours_Per_Day 간의 상관계수는 0.55으로 나타나, 두 변수 간 중등도의 양의 상관관계가 있음을 확인할 수 있다. 스트레스 수준에 따른 GPA 분포를 살펴보면, 스트레스 수준이 “Low”인 경우 GPA는 약 2.5~3.2 사이에 분포하였으며, “Moderate” 수준에서는 GPA는 약 2.8~3.5 사이에 분포하여 평균값과 중앙값이 “Low”수준보다 상승하였다. 스트레스 수준이 “High”인 경우 GPA는 약 3.0~3.8 사이에 분포하며 가장 높은 평균값을 가지는 것을 확인할 수 있다.

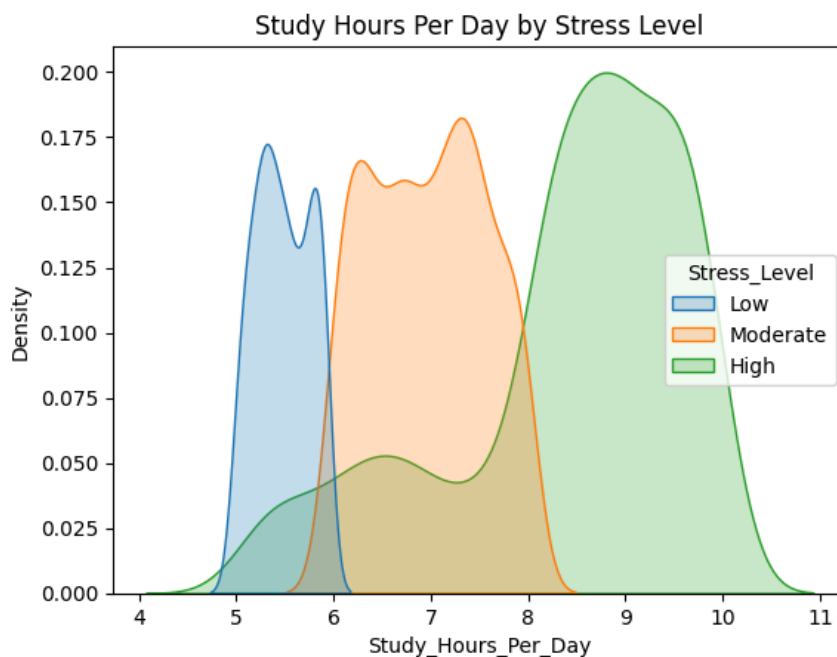
Boxplot 분석결과 스트레스 수준이 “Low”일 때는 이상치가 주로 상단에 나타나는 반면, “High”일 때는 하단에서 이상치가 나타났다.

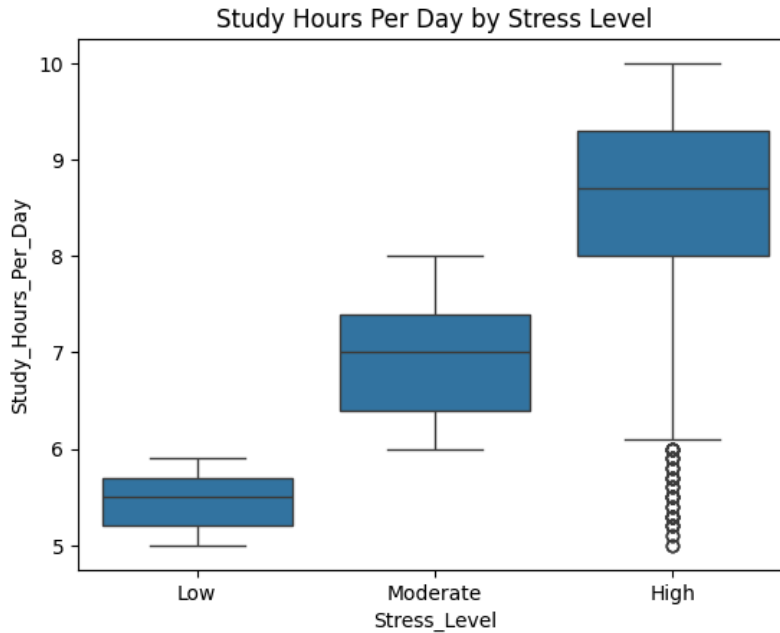
②결론

스트레스 수준이 증가함에 따라 Boxplot의 중앙값 및 평균값이 증가하는 경향과 상관계수를 분석결과를 보아, 스트레스와 GPA 간에 양의 상관관계가 존재함을 알 수 있다. 그러나 이상치의 양상과 분포를 고려할 때, 두 변수 간의 관계는 단순 선형 상관관계가 아닌 로그함수적 혹은 포물선 관계를 가질 가능성이 있음을 시사한다.

ii) 스트레스가 있으면 학습시간이 높을 것이다.

①Stress_Level과 Study_Hours_Per_Day의 관계 (KDE plot, Box plot)





Stress_Level과 Study_Hours_Per_Day 간의 상관계수는 0.74로, 두 변수 간 높은 상관관계를 확인할 수 있었다. 분석 결과, 스트레스 수준이 “Low”인 경우 학업시간은 주로 5 ~ 6시간에 분포하였으며, 스트레스 수준이 “High”인 경우 학업시간의 평균값은 증가하여 8 ~ 10시간 사이에 분포하는 경향을 보였다.

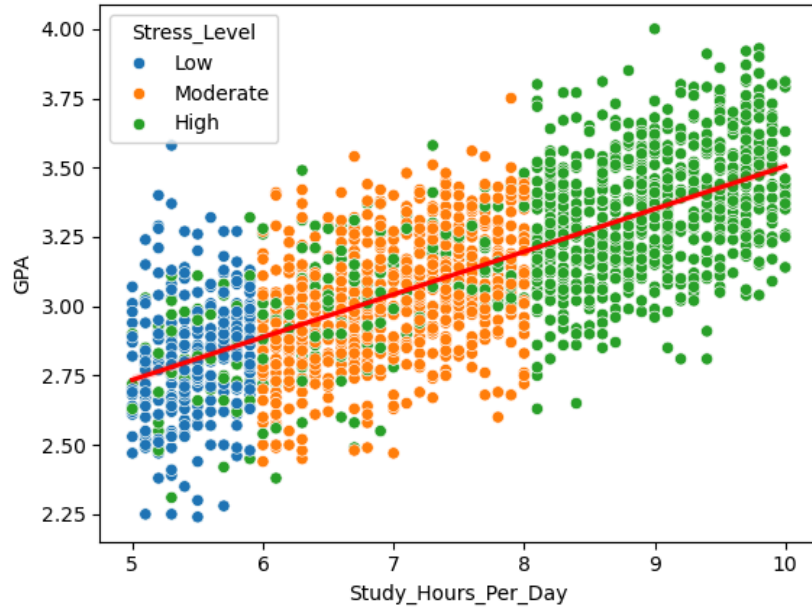
하지만 KDEplot의 High 수준 좌측과 Boxplot의 High 수준 하단영역에서 이상치가 다수 관측되었다. 이를 통해 스트레스 수준이 높아질수록 학업시간의 평균값은 증가하지만, 학업시간의 분산이 커져 학생간의 차이가 두드러지게 됨을 확인할 수 있었다. 이는 스트레스가 적정 수준일 때까지는 학업 시간 증가와 선형적 양의 관계를 보이지만, 과도한 스트레스는 학업시간 증가에 비례적인 긍정적 영향을 미치지 않거나 되려 부정적인 영향을 미칠 가능성이 있음을 시사한다.

②결론

적정 수준의 스트레스는 학업 시간의 증가에 긍정적 영향을 미칠 수 있으나 일정 수준 이상의 과도한 스트레스는 긍정적 영향을 미친다고 보기 어렵다. 즉, 선형적인 상관관계보다는 비선형적인 상관관계(예: 로그함수 혹은 포물선) 관계로 추론할 수 있다.

(ii-i)학업시간이 높으면, 학업성적이 높을 것이다.

①Study_Hours_Per_Day와 GPA의 관계 (Scatter plot) : Stress_Level로 구분



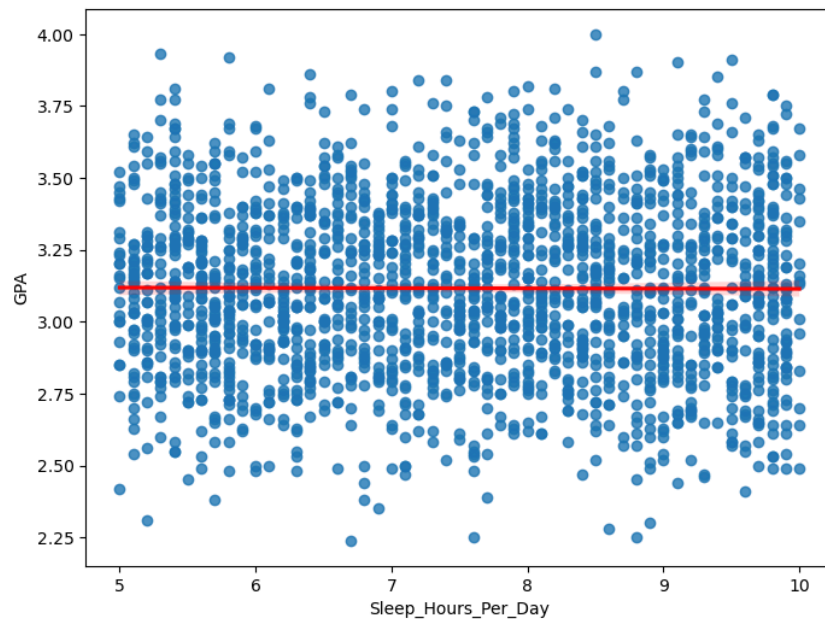
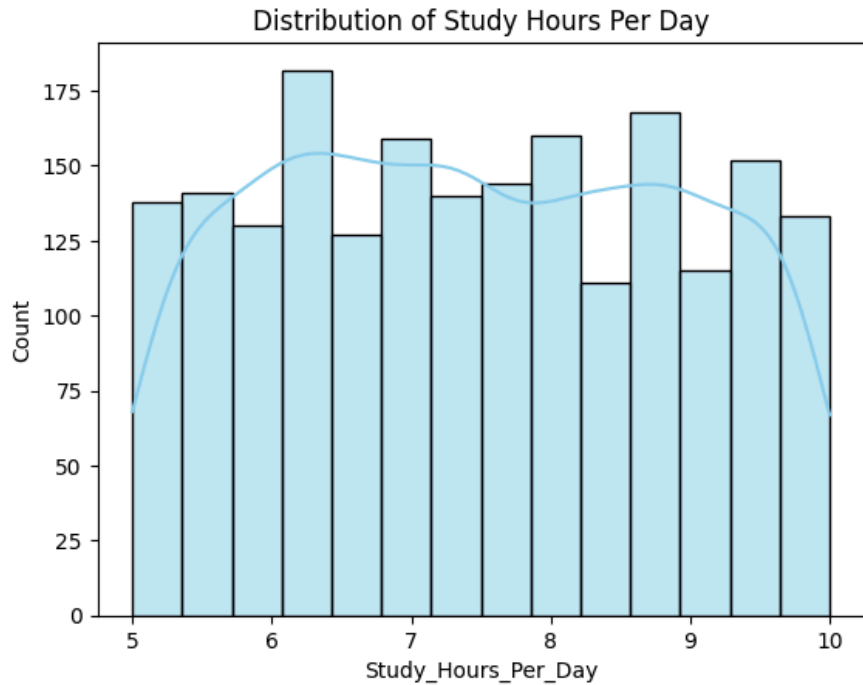
Study_Hours_Per_Day와 GPA간의 상관계수는 0.73로, 높은 선형 상관 관계를 가진다고 볼 수 있다. 그래프에서 스트레스가 낮은 학생들은 파란색으로, 스트레스가 적당한 학생들은 주황색으로, 스트레스가 높은 학생들은 초록색으로 표현했다. 하루 평균 5~6시간 정도 학습하는 학생들의 GPA는 2.7에서 3.0 사이에 분포하여 성적이 낮은 편이다. 하루 평균 6~8시간 정도 학습하는 학생들의 GPA는 2.75~3.50 사이로 범위가 넓게 분포한다. 하루 평균 8~10시간 정도로 더 많은 학습 시간을 투자하는 학생들의 GPA는 3.25 이상으로 성적이 높은 편이다. 이처럼 학업 시간 평균이 길어질수록 학업 성적이 높아지는 경향이 있다.

②결론

그래프가 상승곡선을 그리며, 산점도가 중앙에 잘 밀집된 것을 보아 공부 시간이 증가할수록 학업 성적도 높아지는 걸 확인할 수 있다. i), ii)의 가설과 연관지어 파악하면, 공부 시간이 증가할수록 스트레스 수준도 높아지며, 스트레스 수준이 적정일때 성적 향상 가능성도 크다는 점이 보여진다. 하지만 현재 데이터셋으로는 어떤 스트레스 요인을 받은지 확인할 수 없다. 학업 스트레스가 아닌 다른 외부 요인의 가능성을 배제할 수 없어서 학업 스트레스를 받으면 성적이 향상된다고 확정짓는 것의 위험성을 염두에 두어야 한다.

(번외) 수면 시간이 적으면, 학업 성적이 높을 것이다.

①Sleep_Hours_Per_Day와 Study_Hours_Per_Day의 관계 (KDE plot, Box plot)



또한, 수면시간 0 ~ 6시간을 “Low”, 6 ~ 9시간을 “Balanced”, 9시간 이상을 “High”라고 범주화하였다. 적정 수면시간을 6~9시간으로 설정한 이유는 [Distribution of Study Hours Per Day] 그래프에서 확인할 수 있듯이, 수면시간이 6~8시간 사이에 집중되어 있기 때문이다.

수면시간과 GPA간의 상관계수는 -0.004로 나타나, 두 변수 간 유의미한 선형적 상관관계가 없다고 보았다. “Low”범주의 GPA는 3.111891, “Balanced” 범주의 GPA는 3.123757, “High”범주의 GPA는 3.098407로 나타났다.

②결론

수면 시간에 따른 GPA의 차이가 크지 않으며, 수면 시간이 학업 성취도에 미치는 영향이 미미한 수준임을 시사한다. 더불어, 산점도 및 회귀선으로 시각화해보았을 때, 회귀선이 수평이라 선형의 상관관계가 거의 없다고 볼 수 있다.

Ⅲ. 결론

1)결과 요약

스트레스 수준이 “Low”에 해당하는 학생들은 하루 평균 5~6시간 정도의 학습 시간을 유지하는 경향이 나타났으며, GPA는 주로 2.5에서 3.2 사이에 분포하였다. 스트레스 수준이 “High”에 해당하는 학생들은 하루 평균 8~10시간 정도로 더 많은 학습 시간을 투자하였다. GPA는 3.0에서 3.8 사이로 분포하여 더 높은 경향을 보였다.

위와 같은 상관관계를 바탕으로, 스트레스와 학업 성취 간의 인과관계를 이끌어낼 수 있다. 일정 수준의 스트레스는 학습 동기를 부여하고, 스트레스가 높은 집단이 더 많은 시간을 학업에 할애한다. 그러나 KDE 플롯 High레벨 그래프의 좌측 부분과 박스플롯 High레벨 하단 부분에서 이상치가 다수 관측된다. 이는 선형적 상관관계 보다는 비선형적 관계에 더 적합하다고 해석할 수 있다.

이러한 결과를 종합적으로 분석해 보면, 적정 수준의 스트레스는 학업 동기를 자극하고 학습 성과를 높이는 긍정적인 역할을 할 수 있지만, 스트레스가 지나치게 높아질 경우 오히려 학업 효율성과 성취도를 저하시킬 위험이 있다.

이 보고서는 학업 스트레스를 효과적으로 관리하기 위한 중요한 시사점을 제공한다. 이는 에듀테크 산업이 학업 효율성을 높이면서도 심리적 안정감을 유지할 수 있는 전략을 마련하는 데 기여할 수 있을 것으로 보인다.

2)시사점 및 기대효과

이번 조사를 통해 학습 성취와 스트레스는 비선형 관계임을 확인했다. 그렇다면 학습에 있어 동기부여가 되는 적당한 스트레스를 제공하는 것이 중요할 것이다. 이러한 방법을 제시하는 예로 ‘밀크T’를 들 수 있다. ‘밀크T’는 천재교육의 온라인 학습 프로그램이다. 밀크T는 학습에 게임을 활용하여 학생의 스트레스를 줄여주는 동시에, 학습에 대한 부담감을 덜어주는 방식으로 설계되었다. 유쾌하고 활발한 분위기로 학생과 공부와의 거리감을 줄이고, 자기주도적 학습을 진행하도록 도와준다. 이와 같은 접근은 학습으로 인한 과도한 스트레스를 낮추는 데 효과적이다.

더 나아가, 밀크T는 각 학생에게 적합한 학습 경로를 제시한다. AI 기술을 도입하여 학생의 학습패턴과 방법을 분석하고, 이를 기반으로 맞춤형 학습을 제공한다. 앞으로 AI는 무궁무진한 가능성을 보여준다. 2024년도 10월 뉴스에 의하면 챗GPT를 활용한 기술로 정서적 교감을 통해 어르신의 인지 건강을 돕는다는 사실이 확인되었다. 정신 건강과 인공지능의 연관성을 담은 논문에서도 실제로 “AI가 자살 생각 판별 및 우울증과 같은 정신 질환을 분석”할 수 있다고 판단했다. 앞으로 AI는 학생들이 학습 중 느끼는 감정의 변화를 분석하고, 이를 바탕으로 학생의 스트레스를 관리하면서 더 효과적인 수업을 제공할 수 있을 것이다.

결론적으로, 학습 성취와 스트레스 관리의 균형을 잘 맞춘 밀크T와 같은 에듀테크 제품은 학습 효율을 극대화하며, 이러한 AI 기술은 산업을 더욱 발전시키는 원동력이 될 것이다.

3)한계 및 향후 연구 방향

³ 김서현, 2024-10-21, “[어떻게 생각하십니까] AI와 사랑을... 스트레스 해소 도움 vs 과몰입 문제”, <https://metroseoul.co.kr/article/20241021500613>, 메트로신문

⁴ Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V. (2019). Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Current psychiatry reports*, 21(11), 116. <https://doi.org/10.1007/s11920-019-1094-0>

본 보고서에서는 학업 성취도와 스트레스 간의 연관성을 확인할 수 있었지만, 분석 과정에서 몇 가지 한계가 발견되었다. 먼저, 스트레스 수치가 “Low”, “Moderate”, “High” 이라는 범주로만 제공이 되어 세부적인 분석이 제한되었다. 스트레스의 미세한 변화를 반영하지 못해 구체적으로 어느 정도의 스트레스부터 변곡점을 보이는지, 스트레스와 타 항목 간의 인과관계를 명확하게 파악하기 어려웠다. 또한, 데이터셋에 스트레스의 원인이 포함되지 않아 이 스트레스가 어떤 요인에 의한 것인지 구분할 수 없었다. 후속 연구에서는 스트레스 데이터를 보다 세부적으로 수집하고, 학업과 관련된 스트레스를 구분하여 분석할 수 있다면 더욱 개선될 수 있을 것이다.