# Big Data

Depression in Students

Leonardo Kenji Minemura Suazo

# Introduction

Depression among students is a growing problem. Many students go unnoticed even if they suffer from depression. Students do not know what habits may lead them to depression. On the other hand, institutions do not know which students to focus on to provide them with support.

The objective of this analysis is to identify what type of students are prone to depression and also to identify what kind of lifestyle generates a high level of stress in students. The purpose of this is to make students aware of the lifestyles that may cause them significant stress. It also aims to identify students who may need help to prevent them from falling into depression or to reduce their stress levels. Additionally, it seeks to help students lead a more balanced life.

# Data Selection

To identify students prone to depression, two datasets were selected.

### Depression Student Dataset – Main Datasetl

This dataset examines the connection between mental health and demographic, academic, and lifestyle aspects. It includes attributes such as gender, age, academic pressure, student satisfaction, sleep duration, diet, study hours, financial stress, family history of mental health issues, depression, and suicidal thoughts. This dataset helps identify correlations between these factors in different demographic groups.

**Dataset Content**
- o **Numerical:** 5 columns (Age, Academic Pressure, Study Satisfaction, Study Hours, Financial Stress).
- o **Categóricas**: 6 columnas (Gender, Sleep Duration, Dietary Habits, Suicidal Thoughts, Family History, Depression).
- o **Total entries**: 502

### Student Lifestyle Dataset – Complementary Dataset

This dataset provides a detailed view of students' lifestyles and their correlation with academic performance, represented by GPA. It contains 2,000 records of daily habits related to studies, extracurricular activities, sleep, socialization, and physical activities. Each student's stress level is calculated based on their sleep and study hours, offering insights into how stress levels affect academic performance.

**Dataset Content**
- o **Numerical**: 7 columnas (Student ID, Study Hours, Extracurricular Hours, Sleep Hours, Social Hours, Physical Activity Hours, GPA).
- o **Categóricas**: 1 columnas (Stress Level).

○ **Total entries**: 2000

## How Do They Complement Each Other?

✓ The Depression Student Dataset focuses on mental health and well-being, while the Student Lifestyle Dataset focuses on academic performance. Together, they allow for an examination of the interaction between lifestyle, mental health, and academic performance.

✓ Additionally, they share variables such as study hours, sleep duration, and stress level, which can be analyzed together.

✓ Findings in one dataset can validate discoveries in the other.

# Exploratory Data Analysis

## Depression Student Dataset

**Statistical Description of Numerical Values:** No irregular data observed.



**Dataset Information:** The data type is appropriate and well-balanced.
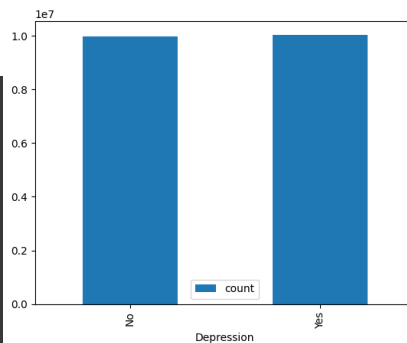


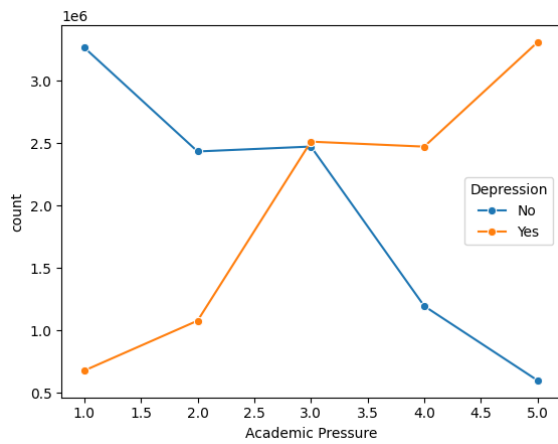

Gender: Males have a higher percentage of depression than females.
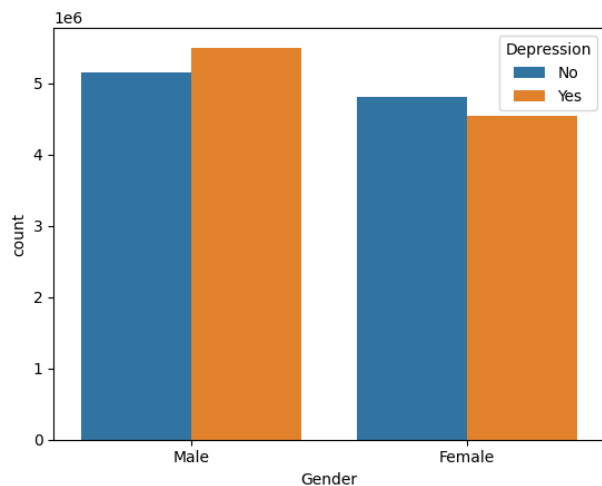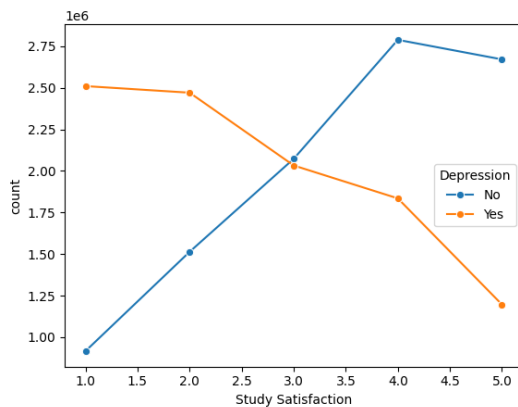
<u>Age:</u> As people age, the number of individuals with depression decreases.
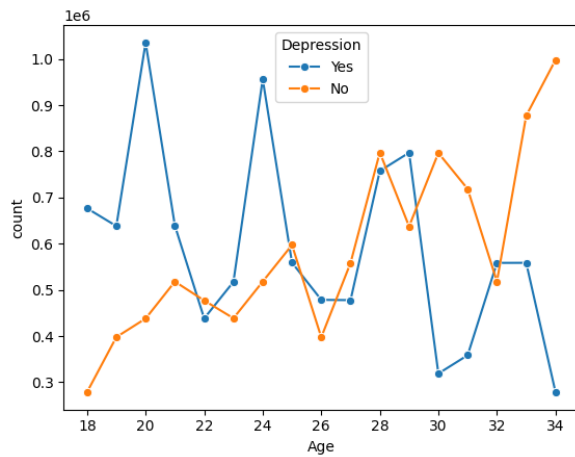


<u>Academic Pressure:</u> The higher the academic pressure, the higher the depression levels.



<u>Academic Satisfaction</u>: Students with depression tend to be dissatisfied with their academic life.

<u>Sleep Duration:</u> Students who sleep more than 8 hours tend to have depression.
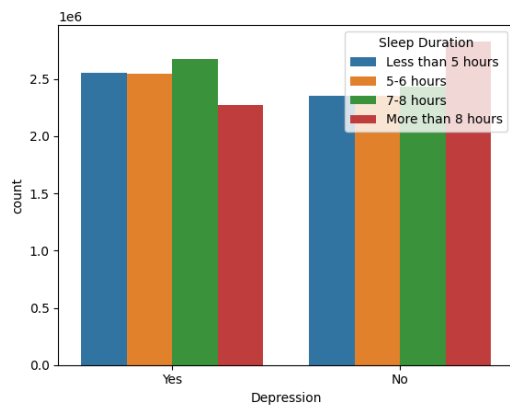


<u>Dietary Habits:</u> Students without depression tend to have better dietary habits than those with depression.



<u>Suicidal Thoughts:</u> Having a suicidal thought is highly related to depression
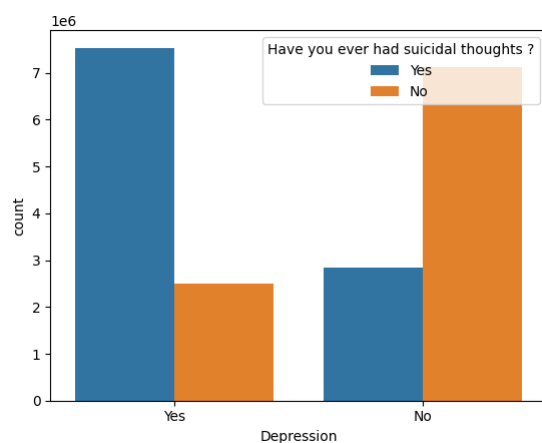
<u>Family history of mental health:</u> Having a family history of mental health problems appears to increase the likelihood of experiencing depression, although it does not seem to be a major factor.
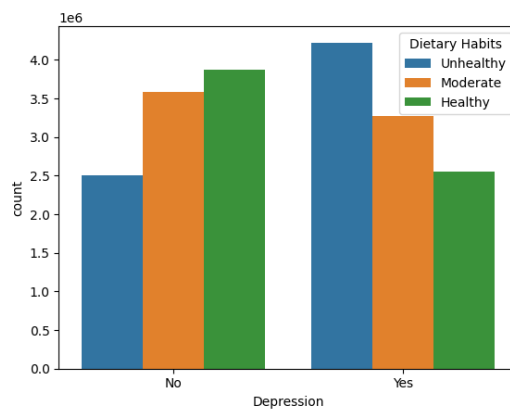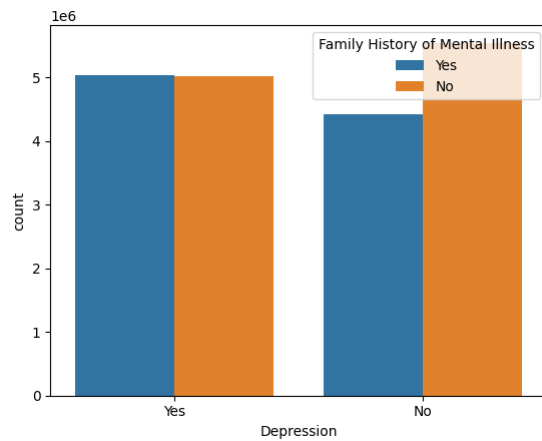


<u>Study Hours:</u> More study hours correlate with a higher likelihood of depression.
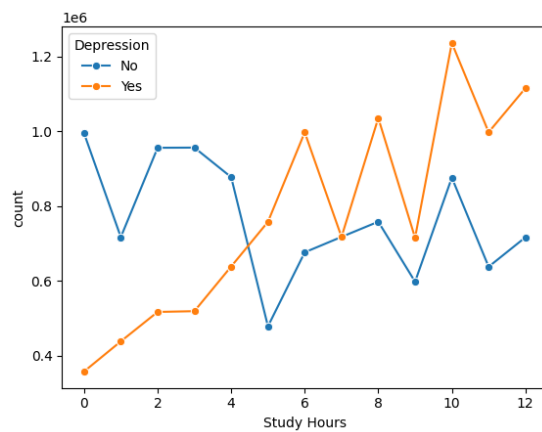


<u>Financial Stress:</u> The number of students with depression increases as financial stress increases.

Correlation Analysis: No significant correlation was observed between numerical variables.

Correlation Matrix

| | Age | Academic Pressure | Study Satisfaction | Study Hours | Financial Stress |
|---|---|---|---|---|---|
| Age | 1.00 | 0.03 | -0.01 | 0.05 | -0.10 |
| Academic Pressure | 0.03 | 1.00 | -0.10 | 0.05 | 0.00 |
| Study Satisfaction | -0.01 | -0.10 | 1.00 | -0.01 | -0.04 |
| Study Hours | 0.05 | 0.05 | -0.01 | 1.00 | 0.02 |
| Financial Stress | -0.10 | 0.00 | -0.04 | 0.02 | 1.00 |

**Student Profiles**

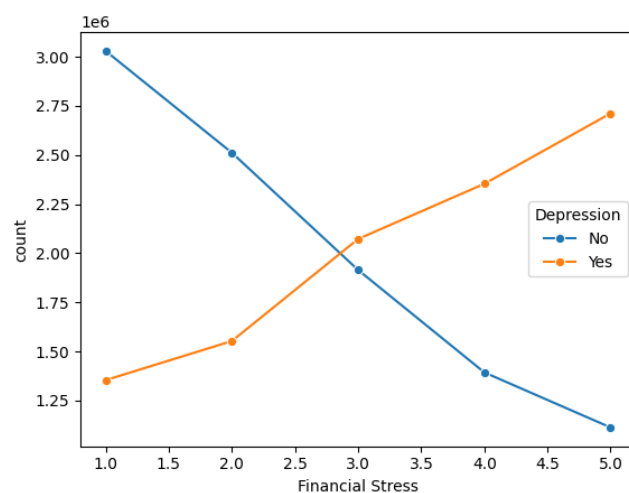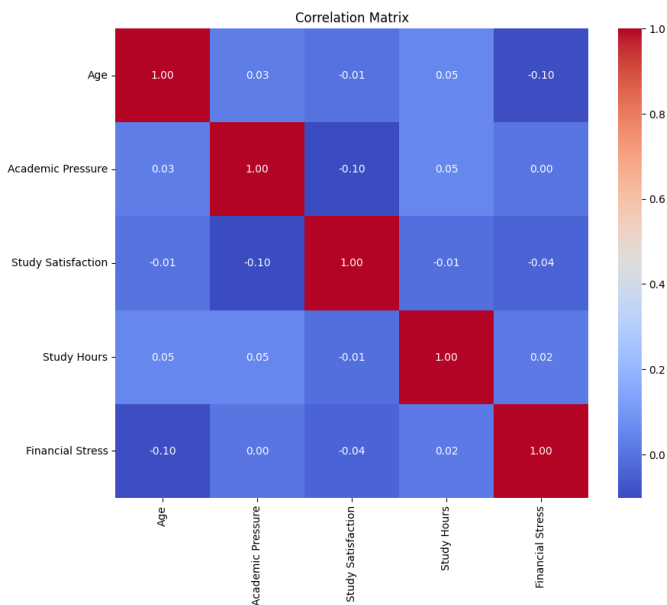| Students with Depression | Students without Depression |
|---|---|
| ⬇ Young | ⬇ Older |
| ⬇ Higher academic stress | ⬇ Lower academic stress |
| ⬇ Unsatisfied with their studies | ⬇ Satisfied with their studies |
| ⬇ Study more hours | ⬇ Study fewer hours |
| ⬇ Higher financial stress | ⬇ Lower financial stress |
| ⬇ Sleep fewer hours | ⬇ Sleep more than 8 hours |
| ⬇ Unhealthy diet | ⬇ Healthy diet |
| ⬇ Have had suicidal thoughts | ⬇ Have not had suicidal thoughts |

## Student Lifestyle Dataset

Statistical Description of Numerical Data: No irregular values observed.

```
+-------+-----------------+------------------+-------------------------------+------------------+-----------------+--------------------------------+------------------+------------+
|summary|       Student_ID|Study_Hours_Per_Day|Extracurricular_Hours_Per_Day|Sleep_Hours_Per_Day|Social_Hours_Per_Day|Physical_Activity_Hours_Per_Day|               GPA|Stress_Level|
+-------+-----------------+------------------+-------------------------------+------------------+-----------------+--------------------------------+------------------+------------+
|  count|          10001415|          10001415|                       10001415|          10001415|         10001415|                        10001415|          10001415|    10001415|
|   mean|1000.5568067118503| 7.475299035221635|              1.990032960335842| 7.50209025420637| 2.7044079862761197|              4.328169763956297|3.1159857760059584|        NULL|
| stddev| 577.4398921663416| 1.423534645819052|              1.1553979279577553|1.4607453354917748|1.6882445006401976|              2.513044320882225|0.2985637796901996|        NULL|
|    min|                 1|               5.0|                            0.0|               0.0|              0.0|                             0.0|              2.24|        High|
|    max|              2000|              10.0|                            4.0|              10.0|              6.0|                            13.0|               4.0|    Moderate|
+-------+-----------------+------------------+-------------------------------+------------------+-----------------+--------------------------------+------------------+------------+
```

Dataset Data Types: No inconsistencies in data types.

```
[('Student_ID', 'int'),
 ('Study_Hours_Per_Day', 'double'),
 ('Extracurricular_Hours_Per_Day', 'double'),
 ('Sleep_Hours_Per_Day', 'double'),
 ('Social_Hours_Per_Day', 'double'),
 ('Physical_Activity_Hours_Per_Day', 'double'),
 ('GPA', 'double'),
 ('Stress_Level', 'string')]
```
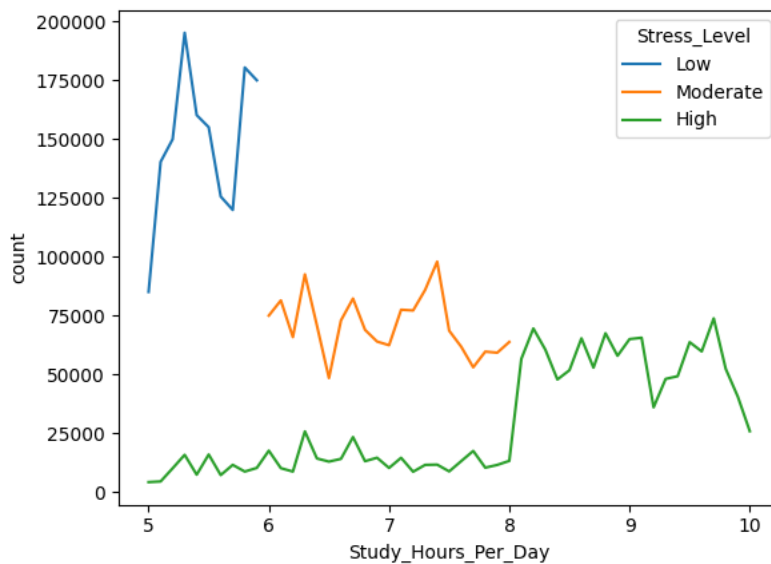
## Category Data Quantity
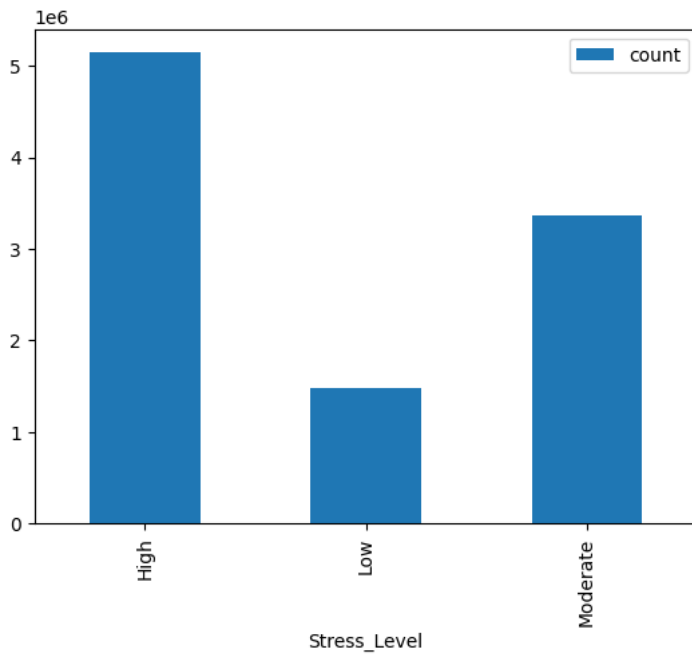


There are many data points for high stress levels, while low stress levels have fewer data points.

This data imbalance may affect visualizations, so data will be balanced for equal category levels.
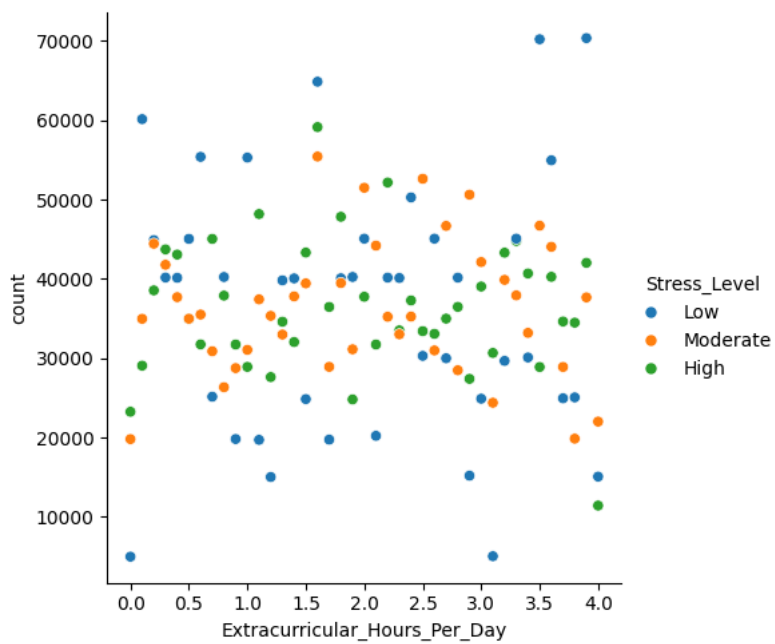
## Study Hours per Day

More study hours correspond to higher stress levels.

Students with low stress study between 5-6 hours per day.
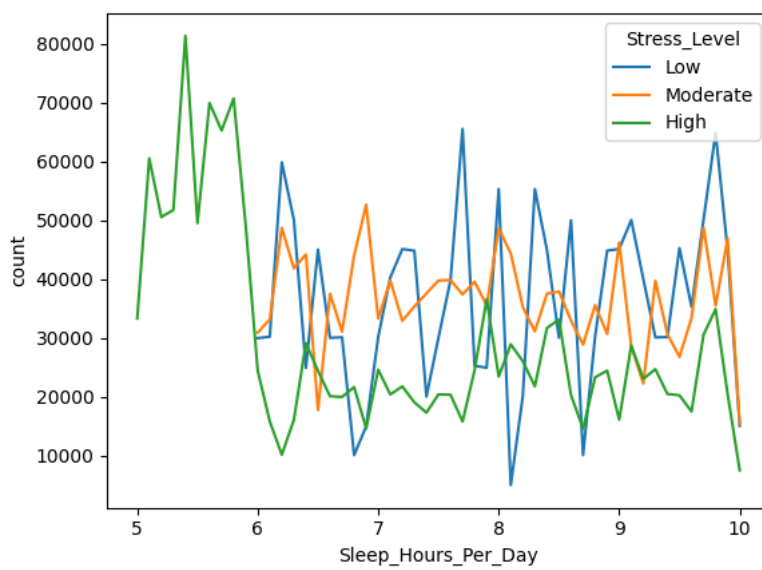
<u>Extracurricular Hours per Day:</u> No significant relationship observed.
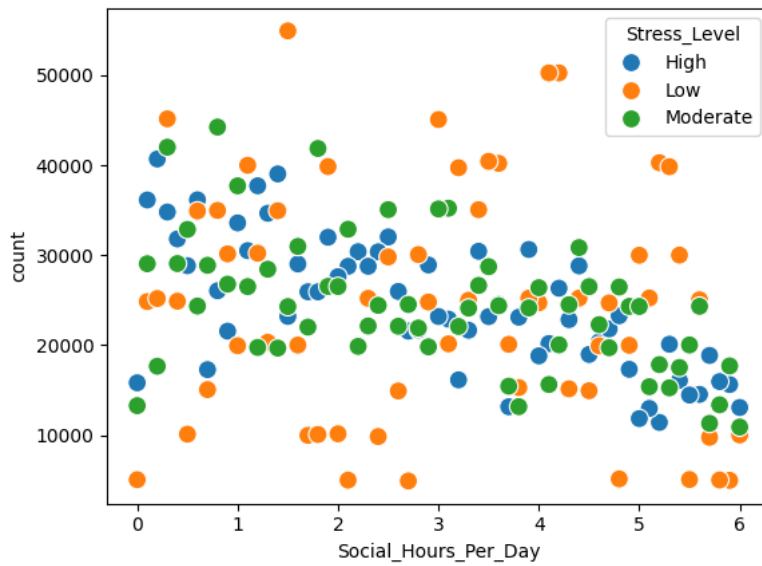


<u>Sleep Hours per Day</u>
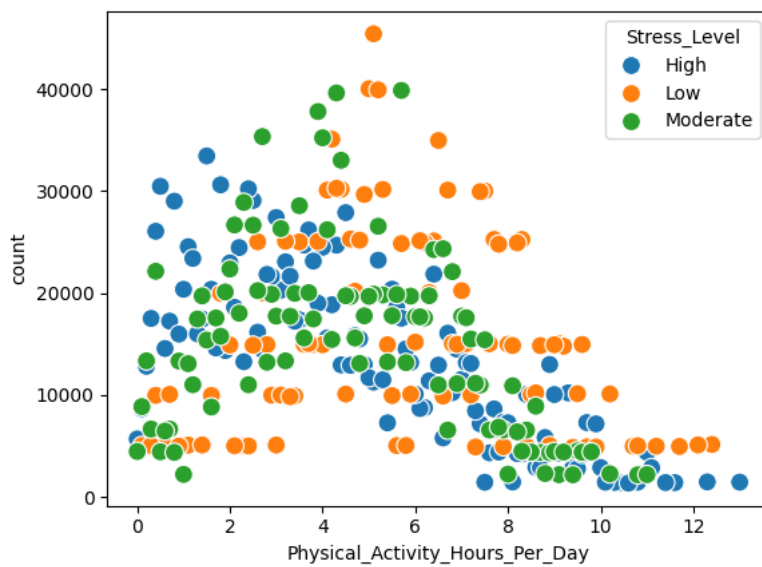
Less sleep corresponds to higher stress levels.

More sleep corresponds to lower stress levels.

<u>Social Hours per Day:</u> As social hours increase, the number of students with low or medium stress decreases.
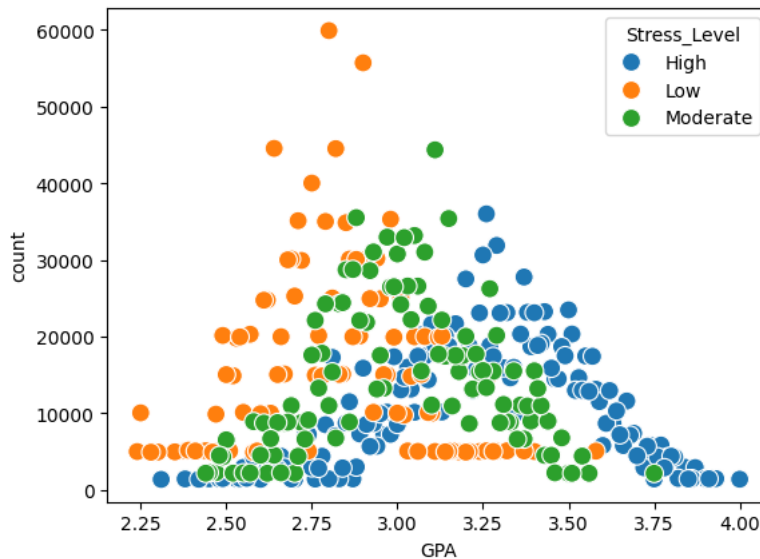


<u>Physical Activity Hours:</u> Less physical activity correlates with higher or moderate stress levels.

A higher GPA is associated with higher stress levels.

Lower stress is linked to a lower GPA.



**Profiles**

| High Stress Level | Low Stress Level |
| --- | --- |
| Many study hours per day | Few study hours per day |
| Less sleep | More sleep |
| Less social time | More social time |
| Less physical activity | More physical activity |
| High GPA | Low GPA |

# Data Cleaning

As shown in the exploratory data analysis, no data cleaning is necessary. The dataset is ready for analysis.

# Detailed Analysis and Value Proposition

### Relationship Between Study Hours and Stress/Depression

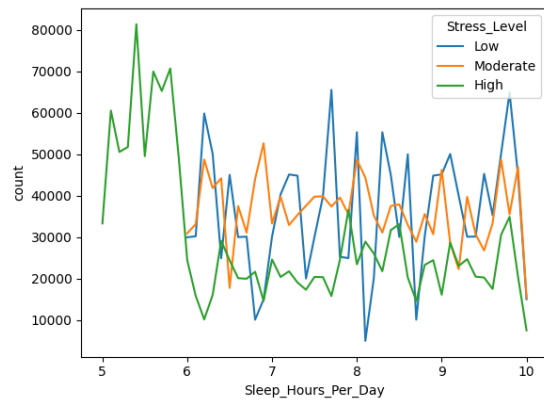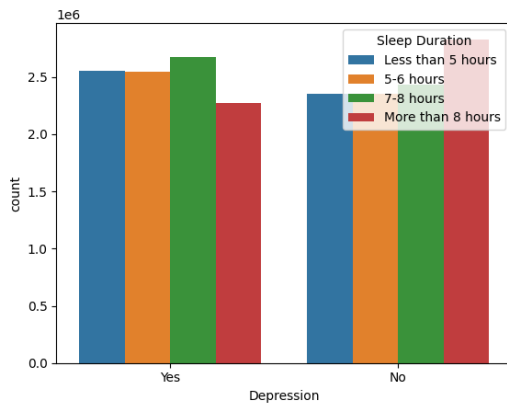As study hours increase, so do depression levels and student stress.

Students who study less than 5 hours tend to not have mental health problems.
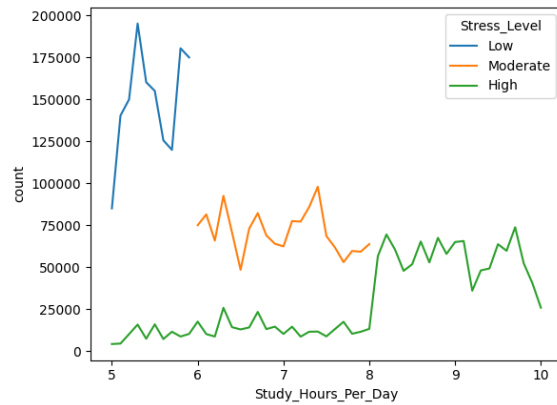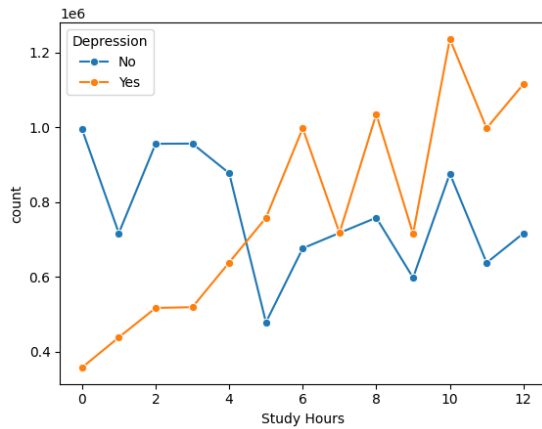
## Impact of Sleep on Mental Health

Lack of sleep negatively affects mental health.

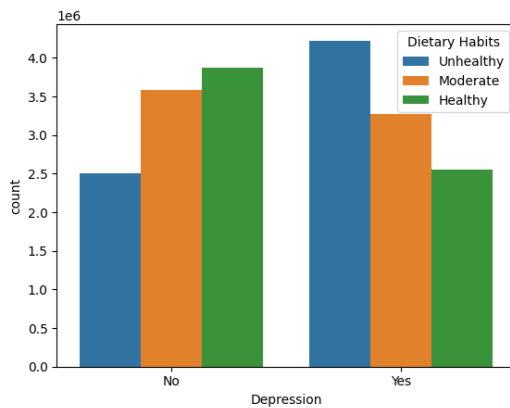Sleeping around 8 hours seems ideal for mental health.


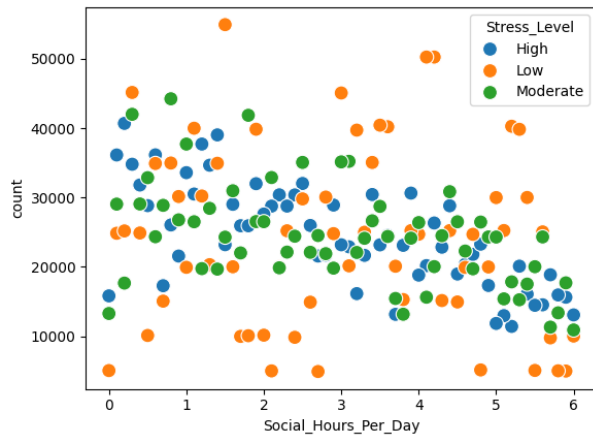
## Role of Diet in Depression

Students with poor dietary habits are more likely to suffer from depression.

Eating healthy appears to be an effective remedy for depression.



## Social Hours as a Way to Reduce Stress

While it is hard to determine if social hours directly improve mental health, cases of moderate and high stress decrease as social hours increase.
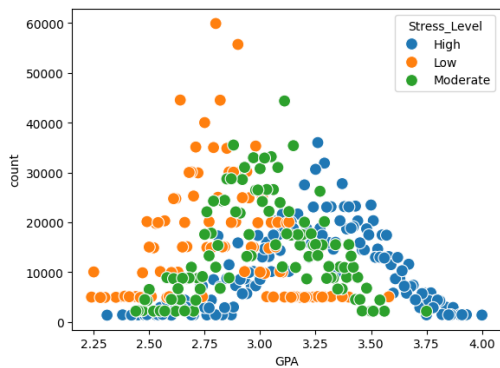
**Relationship Between Academic Performance (GPA) and Stress**

Happy students tend to have better mental health.

Higher academic pressure worsens mental health.

High GPA students experience greater academic pressure and less academic satisfaction.
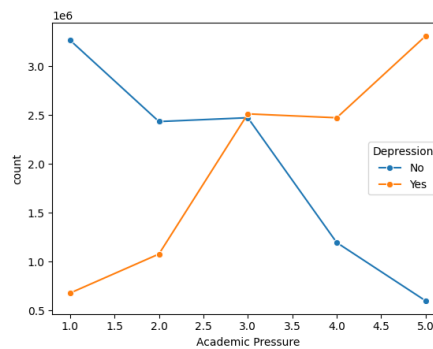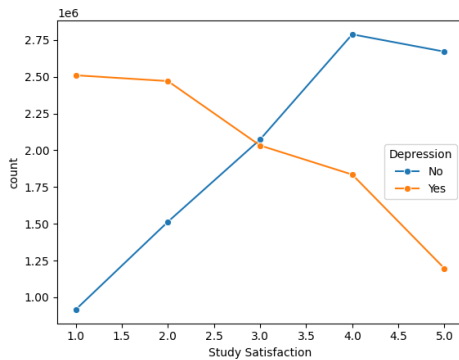


**Recommendations for Students**

Students should strive for a balance between their academic activities and their physical, mental, and social well-being. Based on the conducted analyses, the following points are essential:

1. **Prioritize Sleep:** Sleeping between **7 and 8 hours per day** is fundamental for maintaining good mental health and reducing stress levels. Students should avoid excessive late-night study sessions.

2. **Incorporate Regular Physical Activity:** Physical activity, even if light, helps **reduce stress levels** and **improve mood**. Dedicating at least **30 minutes per day** to exercises such as

walking, running, or playing sports is highly recommended.

3. **Maintain a Balanced Diet:** A nutritious diet improves both physical and mental well-being. It is important to include vitamin- and mineral-rich foods and avoid excessive consumption of processed foods.

4. **Encourage Social Relationships:** Spending quality time with friends or family helps prevent mental health issues. Dedicating time to socialization can be just as important as study hours.

5. **Balance Study Hours:** While achieving good academic performance is important, it is equally crucial to avoid prolonged study sessions that lead to burnout. It is recommended to plan efficient study schedules and combine study time with breaks to rest.Recomendaciones para las instituciones

Based on the findings, institutions should implement the following strategies:

1. **Reducing Academic Workload:** Avoid overloading students with excessive assignments and exams. It is important to offer a **balanced and reasonable curriculum.**

2. **Creating Spaces for Socialization:** Provide areas and events that promote social interaction among students, such as **clubs, workshops, or extracurricular activities** that foster a sense of community.

3. **Raising Awareness About Mental Health:** Implement **educational campaigns** highlighting the importance of sleep, physical activity, and socialization for mental well-being. Additionally, offer **workshops** to help students manage stress.

4. **Providing Support Services:** Have **counselors or psychologists** accessible to students, allowing them to seek help when facing high levels of stress or mental health issues.

5. **Monitoring Student Well-being:** Implement **systems to identify at-risk students** based on indicators such as academic performance, attendance, and direct feedback.

## Actionable Steps

For Students:

- Establish a **fixed sleep schedule** and stick to it.
- Incorporate at least **30 minutes of daily physical activity** into their routine.
- Prepare **healthy meals** and avoid foods high in saturated fats or sugar.
- Schedule **social gatherings or activities** at least once a week.
- Design a **study plan** that includes regular breaks for rest.

For Institutions:

- Review **academic policies** and adjust workload according to student needs.
- Organize events such as **wellness fairs, recreation days, and integration activities.**
- Establish **psychological support centers** where students can receive confidential help.

- Promote **regular surveys** to measure students' stress levels and academic satisfaction.
- Create **mentorship programs** to help students manage their schedules and study strategies more effectively.

## Conclusiones

The analysis reveals that **students' mental well-being and academic performance** are strongly influenced by their lifestyle. Factors such as **sleep duration, physical activity, socialization, and dietary habits** play a crucial role in reducing stress and preventing depression. Additionally, **excessive study hours and academic pressure** not only increase stress levels but can also negatively impact students' mental health.

It is concluded that **both students and educational institutions have an active role** in creating a balanced environment that fosters both learning and overall well-being. Students should strive to **maintain a healthy balance in their lives**, while institutions have the **responsibility to create policies and spaces that promote a sustainable and enriching educational experience.**

Implementing the proposed recommendations—both at the **individual and institutional levels**—can help **identify and support at-risk students, prevent mental health deterioration, and foster a more inclusive and healthier learning environment.** This joint effort has the potential **not only to improve students' quality of life but also to optimize their academic performance and personal development.**