**Data Science using Python**

# Mini Project Report On

# Student-Mental & Stress and Coping Mechanisms


PREPARED BY

| | |
|---|---|
| Kamal Agrahari | VU4F2223028 |
| Rutvik Gondekar | VU4F2223031 |
| Akash Nahak | VU4F2223034 |
| Sahil Deshmukh | VU4F2223020 |

Class: TE IT-B (Batch -B)

Guided By: Darshana Gajbhiye

# Student-Mental &Stress and Coping Mechanisms

**Problem Statement:**

In today's fast-paced academic environment, students face immense pressure from exams, assignments, social expectations, and future career concerns. This often leads to increased stress, anxiety, and mental health challenges, which can negatively impact their academic performance, personal relationships, and overall well-being.

However, many students lack awareness of effective coping mechanisms and mental health resources. There is a need for a structured approach to help students identify stressors, manage their mental health, and adopt healthy coping strategies. Addressing this issue can lead to improved student productivity, emotional resilience, and overall mental well-being.

**About the dataset:**

S**tudent_stress.csv (Main dataset for training and analysis):**

This dataset contains information about student mental health, stress levels, and coping mechanisms. It can be used for training machine learning models to predict stress levels, identify high-risk students, and recommend coping strategies.

**Dataset Use Cases**

**Predicting Student Stress Levels** – Train a model to classify students into low, moderate, or high stress categories.

**Clustering Student Mental Health Patterns –** Identify groups of students with similar stress levels and behaviors using K-Means clustering.

**Anomaly Detection for High-Risk Students –** Use Isolation Forest to detect students experiencing extreme stress or unhealthy coping mechanisms.

**Coping Mechanism Analysis –** Understand which stress management techniques work best for different student groups.

**Algorithms used:**

**Clustering (K-Means) and Anomaly Detection (Isolation Forest) for Student Mental Stress Analysis:**

In a **Student Mental Stress and Coping Mechanisms** project, unsupervised learning techniques such as **clustering (K-Means)** and **anomaly detection (Isolation Forest)** provide valuable insights into student stress patterns and unusual mental health conditions.

**Clustering (K-Means) for Stress Pattern Identification**
Clustering algorithms like K-Means can group students based on their mental health attributes such as academic pressure, social anxiety, sleep hours, exercise frequency, and coping mechanisms. By identifying student clusters, we can:

- Classify students into different stress levels (Low, Moderate, High).
- Identify common stress patterns among students based on their lifestyle and habits.
- Discover distinct coping mechanism groups, such as students who manage stress through exercise vs. those who seek social support.

**Example Clusters:**
**1** Low Stress Group – Students with balanced sleep, regular exercise, and effective coping strategies.

**2** Moderate Stress Group – Students with occasional stress but some healthy coping mechanisms.

**3** High Stress Group – Students experiencing high academic pressure, poor sleep, and ineffective coping.

**By analyzing these clusters, institutions can design targeted stress management programs and recommend personalized coping strategies.**

# Data Loading and Inspection:

## Data Loading and Inspection

```python
from google.colab import files
uploaded = files.upload()
import pandas as pd

# Load the dataset
data = pd.read_csv("Student_Mental_Stress_and_Coping_Mechanisms.csv", encoding='latin-1')

# Display basic info
data.info()

# Show the first few rows
data.head()
```

Choose Files Student_M...hanisms.csv

- **Student_Mental_Stress_and_Coping_Mechanisms.csv**(text/csv) - 57141 bytes, last modified: 3/28/2025 - 100% done

```
Saving Student_Mental_Stress_and_Coping_Mechanisms.csv to Student_Mental_Stress_and_Coping_Mechanisms (1).csv
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 760 entries, 0 to 759
Data columns (total 20 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   Student ID                          760 non-null    object
 1   Age                                 760 non-null    int64
 2   Gender                              760 non-null    object
 3   Academic Performance (GPA)          760 non-null    int64
 4   Study Hours Per Week                760 non-null    int64
 5   Social Media Usage (Hours per day)  760 non-null    int64
 6   Sleep Duration (Hours per night)    760 non-null    int64
 7   Physical Exercise (Hours per week)  760 non-null    int64
```

13s completed at 12:29 AM

| | Student ID | Age | Gender | Academic Performance (GPA) | Study Hours Per Week | Social Media Usage (Hours per day) | Sleep Duration (Hours per night) | Physical Exercise (Hours per week) | Family Support | Financial Stress | Peer Pressure | Relationship Stress | Mental Stress Level | Counseling Attendance | Diet Quality | Stress Coping Mechanisms | Cognitive Distortions | Family Mental Health History | Medical Condition | Substance Use |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 802-17-3671 | 22 | Female | 2 | 9 | 2 | 12 | 2 | 1 | 1 | 3 | 5 | 9 | No | 1 | Walking or Nature Walks | 4 | No | Yes | 1 |
| 1 | 871-12-8572 | 25 | Female | 0 | 28 | 0 | 6 | 0 | 1 | 1 | 1 | 2 | 9 | Yes | 3 | Meditation | 2 | Yes | No | 1 |
| 2 | 495-13-2672 | 24 | Female | 0 | 45 | 3 | 12 | 10 | 3 | 3 | 1 | 4 | 9 | Yes | 5 | Reading | 1 | Yes | Yes | 3 |
| 3 | 365-77-2496 | 20 | Male | 2 | 8 | 7 | 7 | 4 | 1 | 3 | 2 | 5 | 1 | No | 1 | Social Media Engagement | 2 | Yes | No | 4 |
| 4 | 664-76-5622 | 28 | Male | 0 | 14 | 6 | 8 | 1 | 2 | 4 | 4 | 2 | 7 | Yes | 1 | Exercise | 1 | Yes | No | 3 |

## Data Cleaning and Preprocessing

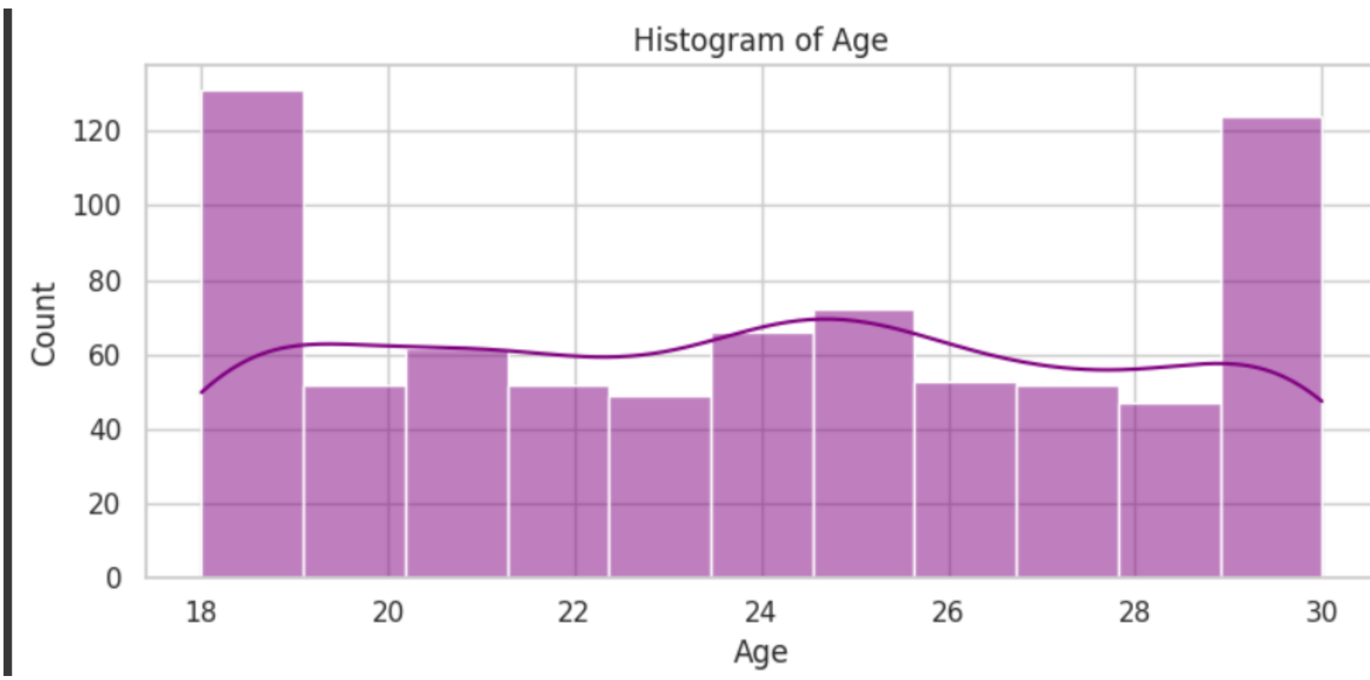
```
Data columns (total 20 columns):
 #   Column                              Non-Null Count  Dtype
---  ------                              --------------  -----
 0   Student ID                          760 non-null    object
 1   Age                                 760 non-null    int64
 2   Gender                              760 non-null    object
 3   Academic Performance (GPA)          760 non-null    int64
 4   Study Hours Per Week                760 non-null    int64
 5   Social Media Usage (Hours per day)  760 non-null    int64
 6   Sleep Duration (Hours per night)    760 non-null    int64
 7   Physical Exercise (Hours per week)  760 non-null    int64
 8   Family Support                      760 non-null    int64
 9   Financial Stress                    760 non-null    int64
 10  Peer Pressure                       760 non-null    int64
 11  Relationship Stress                 760 non-null    int64
 12  Mental Stress Level                 760 non-null    int64
 13  Counseling Attendance               760 non-null    object
 14  Diet Quality                        760 non-null    int64
 15  Stress Coping Mechanisms            760 non-null    object
 16  Cognitive Distortions               760 non-null    int64
 17  Family Mental Health History        760 non-null    object
 18  Medical Condition                   760 non-null    object
 19  Substance Use                       760 non-null    int64
dtypes: int64(14), object(6)
memory usage: 118.9+ KB
```

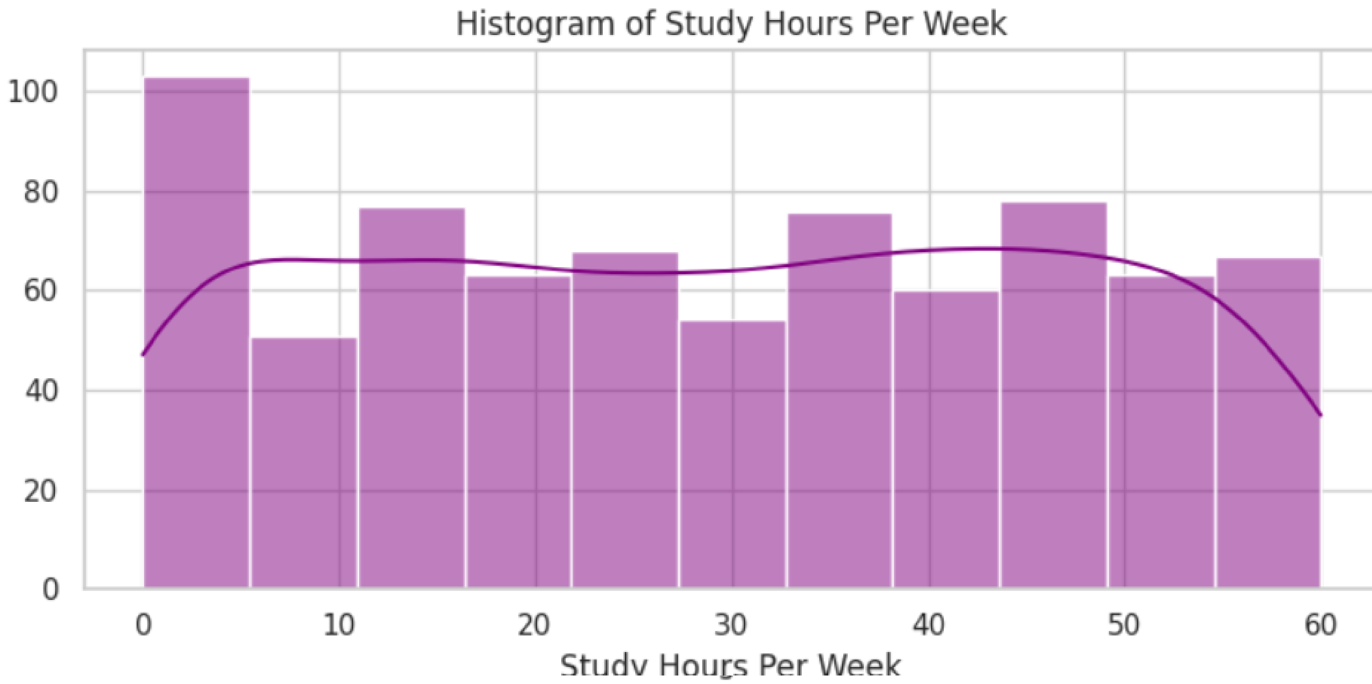
```
Missing values in each column:
Student ID                            0
Age                                   0
Gender                                0
Academic Performance (GPA)            0
Study Hours Per Week                  0
Social Media Usage (Hours per day)    0
Sleep Duration (Hours per night)      0
Physical Exercise (Hours per week)    0
Family Support                        0
Financial Stress                      0
Peer Pressure                         0
Relationship Stress                   0
Mental Stress Level                   0
Counseling Attendance                 0
Diet Quality                          0
Stress Coping Mechanisms              0
Cognitive Distortions                 0
Family Mental Health History          0
Medical Condition                     0
Substance Use                         0
dtype: int64
Initial shape: (760, 20), After removing duplicates: (760, 20)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 760 entries, 0 to 759
Data columns (total 20 columns):
```

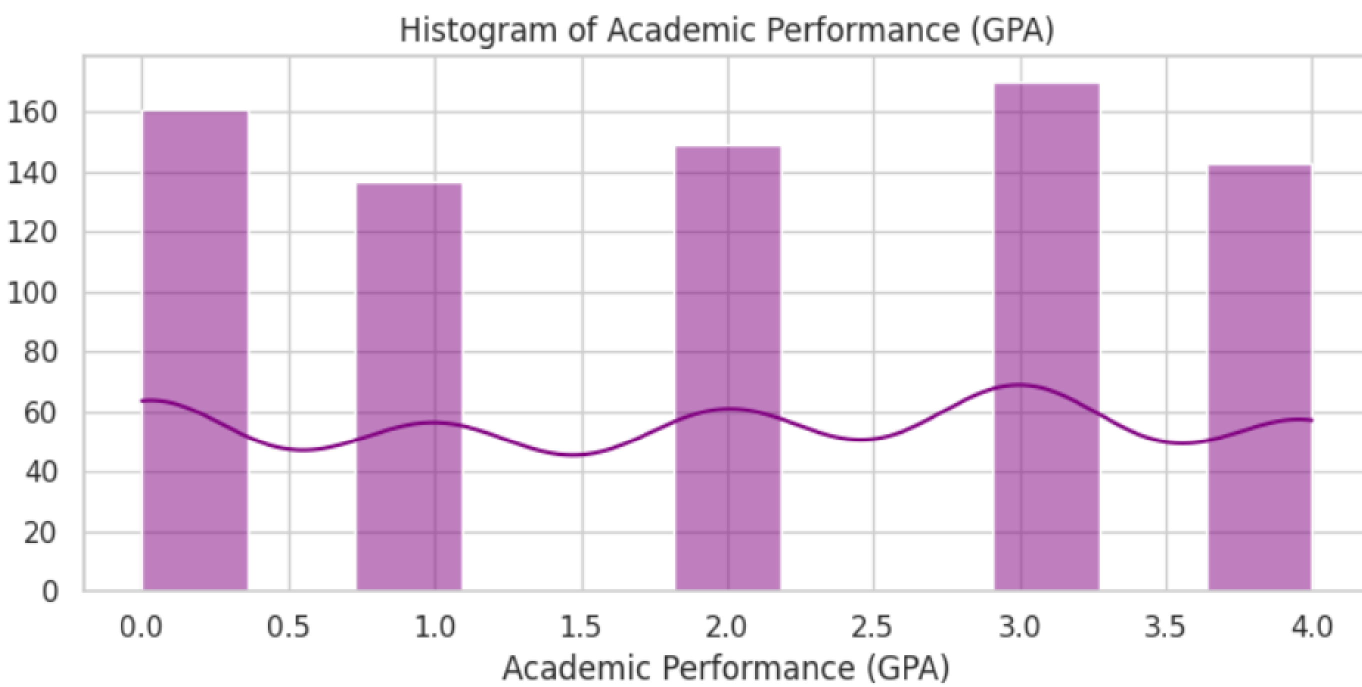
# Exploratory Data Analysis:

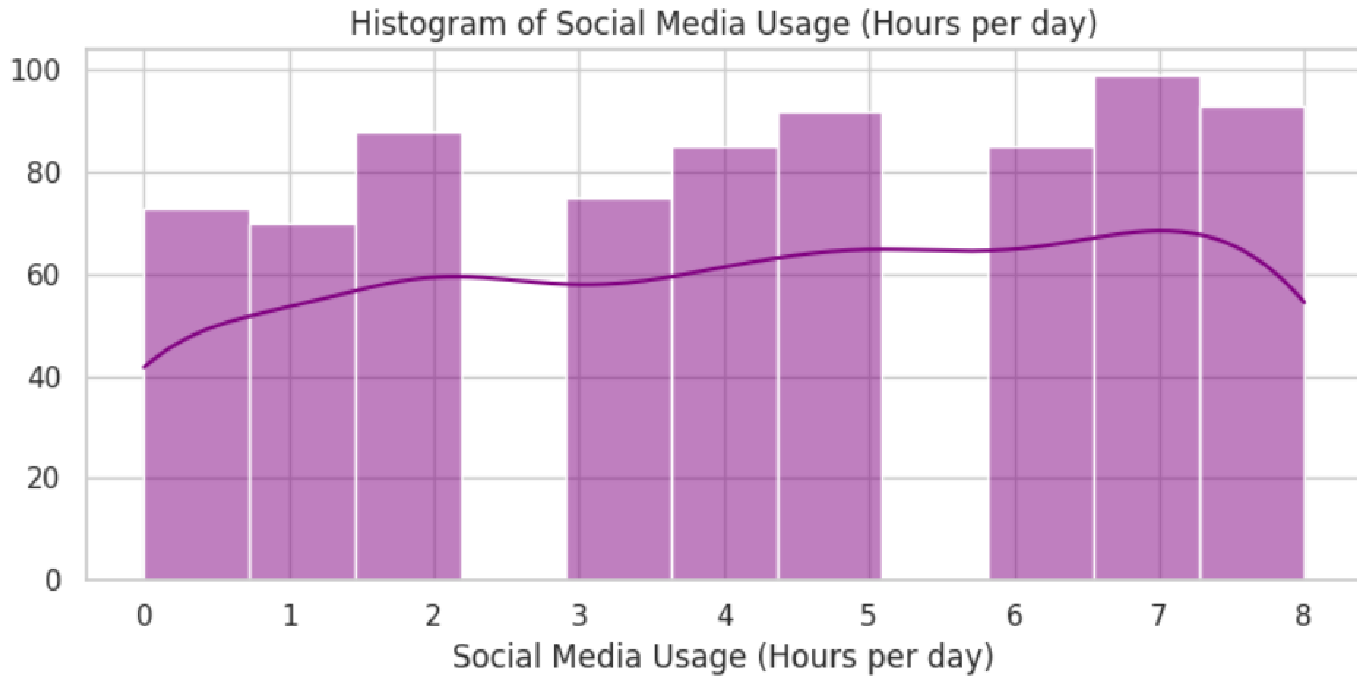
Statistical Summary:

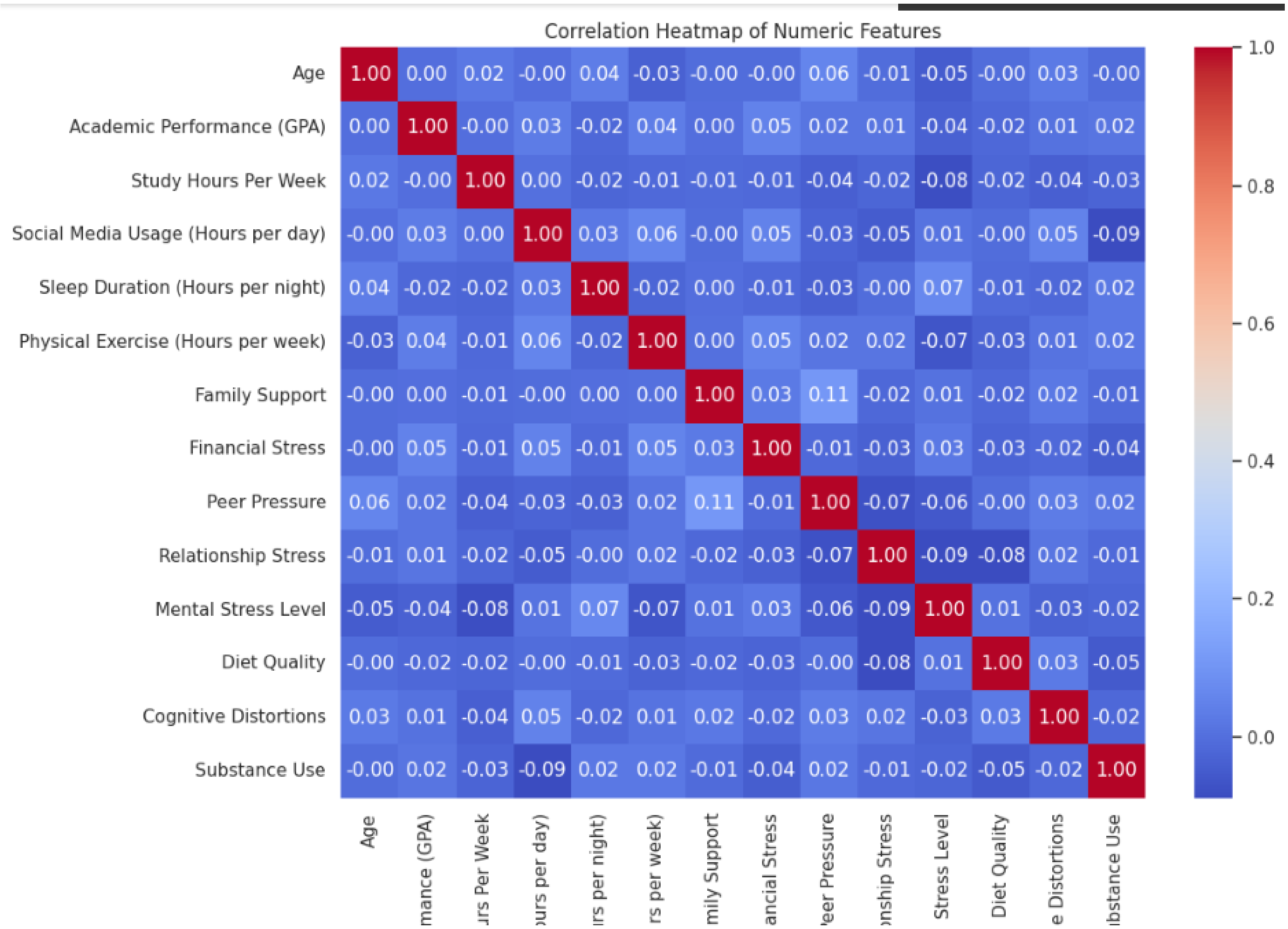
| | Age | Academic Performance (GPA) | Study Hours Per Week | Social Media Usage (Hours per day) | Sleep Duration (Hours per night) | Physical Exercise (Hours per week) | Family Support | Financial Stress | Peer Pressure | Relationship Stress | Mental Stress Level | Diet Quality | Cognitive Distortions | Substance Use |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 760.000000 | 760.000000 | 760.000000 | 760.000000 | 760.000000 | 760.000000 | 760.000000 | 760.000000 | 760.000000 | 760.000000 | 760.000000 | 760.000000 | 760.000000 | 760.000000 |
| mean | 23.919737 | 1.996053 | 29.130263 | 4.234211 | 8.093421 | 4.934211 | 2.972368 | 3.021053 | 2.982895 | 2.963158 | 5.363158 | 3.042105 | 3.007895 | 2.996053 |
| std | 3.819392 | 1.416535 | 18.104662 | 2.565174 | 2.596838 | 3.082679 | 1.413477 | 1.452662 | 1.391096 | 1.369231 | 2.902856 | 1.453111 | 1.432703 | 1.385502 |
| min | 18.000000 | 0.000000 | 0.000000 | 0.000000 | 4.000000 | 0.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 | 1.000000 |
| 25% | 21.000000 | 1.000000 | 13.000000 | 2.000000 | 6.000000 | 2.000000 | 2.000000 | 2.000000 | 2.000000 | 2.000000 | 3.000000 | 2.000000 | 2.000000 | 2.000000 |
| 50% | 24.000000 | 2.000000 | 29.000000 | 4.000000 | 8.000000 | 5.000000 | 3.000000 | 3.000000 | 3.000000 | 3.000000 | 5.000000 | 3.000000 | 3.000000 | 3.000000 |
| 75% | 27.000000 | 3.000000 | 45.000000 | 7.000000 | 10.000000 | 8.000000 | 4.000000 | 4.000000 | 4.000000 | 4.000000 | 8.000000 | 4.000000 | 4.000000 | 4.000000 |
| max | 30.000000 | 4.000000 | 60.000000 | 8.000000 | 12.000000 | 10.000000 | 5.000000 | 5.000000 | 5.000000 | 5.000000 | 10.000000 | 5.000000 | 5.000000 | 5.000000 |

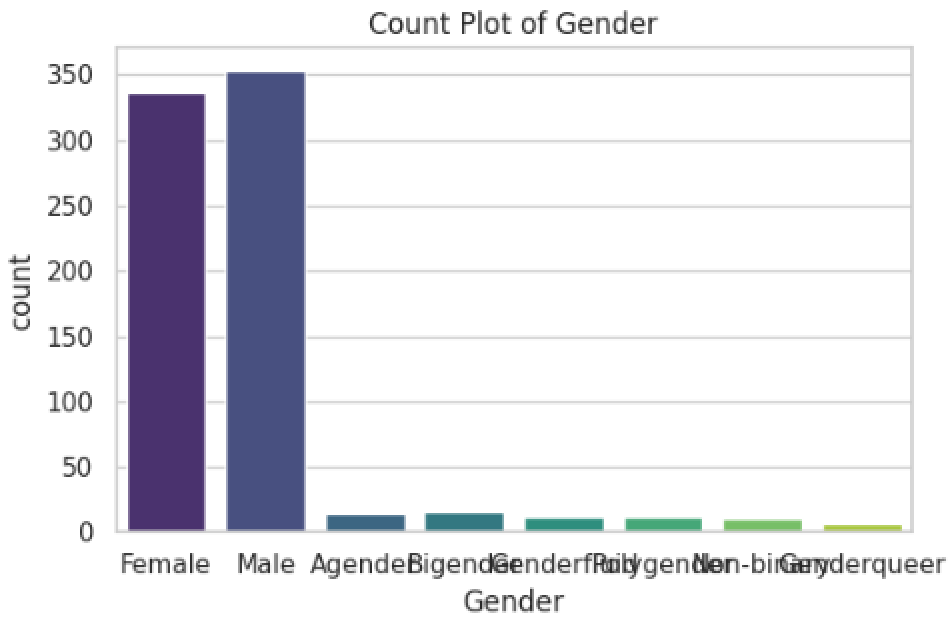


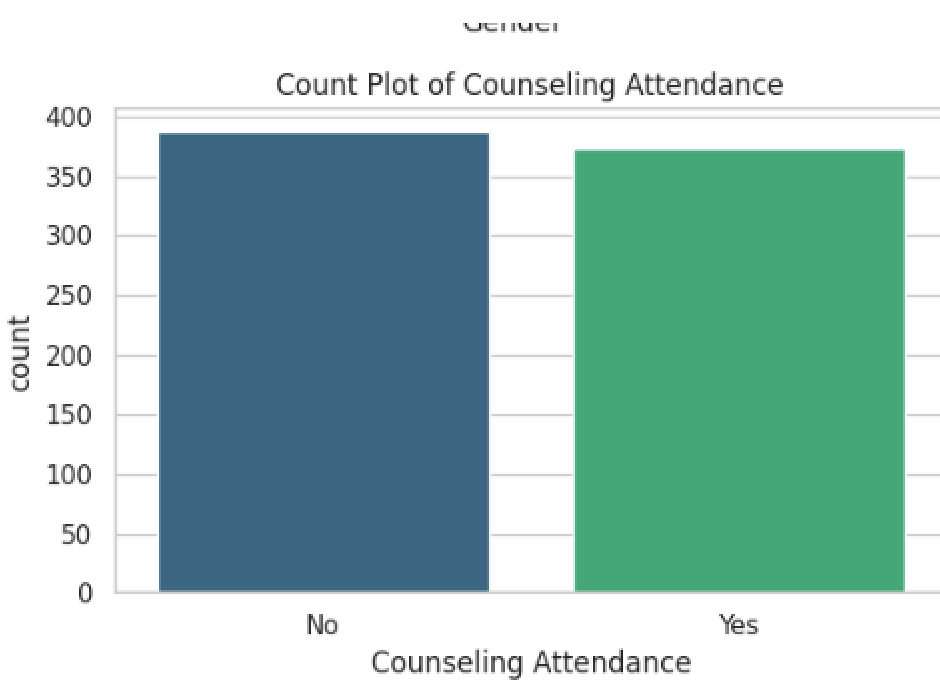

Histogram of Study Hours Per Week
0
20
40
60
80
100
0
10
20
30
40
50
60
Study Hours Per Week


Histogram of Academic Performance (GPA)
0
20
40
60
80
100
120
140
160
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
Academic Performance (GPA)


Histogram of Social Media Usage (Hours per day)
100
80
60
40
20
0
0
1
2
3
4
5
6
7
8
Social Media Usage (Hours per day)


Correlation Heatmap of Numeric Features
Age 1.00 0.00 0.02 -0.00 0.04 -0.03 -0.00 -0.00 0.06 -0.01 -0.05 -0.00 0.03 -0.00
Academic Performance (GPA) 0.00 1.00 -0.00 0.03 -0.02 0.04 0.00 0.05 0.02 0.01 -0.04 -0.02 0.01 0.02
Study Hours Per Week 0.02 -0.00 1.00 0.00 -0.02 -0.01 -0.01 -0.01 -0.04 -0.02 -0.08 -0.02 -0.04 -0.03
Social Media Usage (Hours per day) -0.00 0.03 0.00 1.00 0.03 0.06 -0.00 0.05 -0.03 -0.05 0.01 -0.00 0.05 -0.09
Sleep Duration (Hours per night) 0.04 -0.02 -0.02 0.03 1.00 -0.02 0.00 -0.01 -0.03 -0.00 0.07 -0.01 -0.02 0.02
Physical Exercise (Hours per week) -0.03 0.04 -0.01 0.06 -0.02 1.00 0.00 0.05 0.02 0.02 -0.07 -0.03 0.01 0.02
Family Support -0.00 0.00 -0.01 -0.00 0.00 0.00 1.00 0.03 0.11 -0.02 0.01 -0.02 0.02 -0.01
Financial Stress -0.00 0.05 -0.01 0.05 -0.01 0.05 0.03 1.00 -0.01 -0.03 0.03 -0.03 -0.02 -0.04
Peer Pressure 0.06 0.02 -0.04 -0.03 -0.03 0.02 0.11 -0.01 1.00 -0.07 -0.06 -0.00 0.03 0.02
Relationship Stress -0.01 0.01 -0.02 -0.05 -0.00 0.02 -0.02 -0.03 -0.07 1.00 -0.09 -0.08 0.02 -0.01
Mental Stress Level -0.05 -0.04 -0.08 0.01 0.07 -0.07 0.01 0.03 -0.06 -0.09 1.00 0.01 -0.03 -0.02
Diet Quality -0.00 -0.02 -0.02 -0.00 -0.01 -0.03 -0.02 -0.03 -0.00 -0.08 0.01 1.00 0.03 -0.05
Cognitive Distortions 0.03 0.01 -0.04 0.05 -0.02 0.01 0.02 -0.02 0.03 0.02 -0.03 0.03 1.00 -0.02
Substance Use -0.00 0.02 -0.03 -0.09 0.02 0.02 -0.01 -0.04 0.02 -0.01 -0.02 -0.05 -0.02 1.00
1.0
0.8
0.6
0.4
0.2
0.0




# Predictive Modeling:

```
mse = mean_squared_error(y_test, y_pred)  # Remove squared=False
rmse = np.sqrt(mse)
r2 = r2_score(y_test, y_pred)
print(f'Random Forest Regressor RMSE: {rmse}')
print(f'Random Forest Regressor R2 Score: {r2}')
```

```
Random Forest Regressor RMSE: 3.054300031792968
Random Forest Regressor R2 Score: -0.08722462469733627
```

## Conclusions

The analysis provided a comprehensive overview of the student mental stress dataset. We covered the journey from loading and cleaning the data to exploratory visualization and the development of a predictive model. The Random Forest Regressor, predicting mental stress level using student behavior and support metrics, displayed a particular performance based on the RMSE and R2 scores.

Merits of the approach include:

- Diverse visualization techniques that allow us to gain a multifaceted understanding of the data.
- Step-by-step data cleaning methods to avoid common pitfalls such as missing values and duplicate entries.
- An interpretable modeling approach highlighting feature importance via permutation importance.