



# DATA ANALYSIS AND VISUALIZATION WITH PANDAS & MATPLOTLIB: STUDENT DEPRESSION DATASET

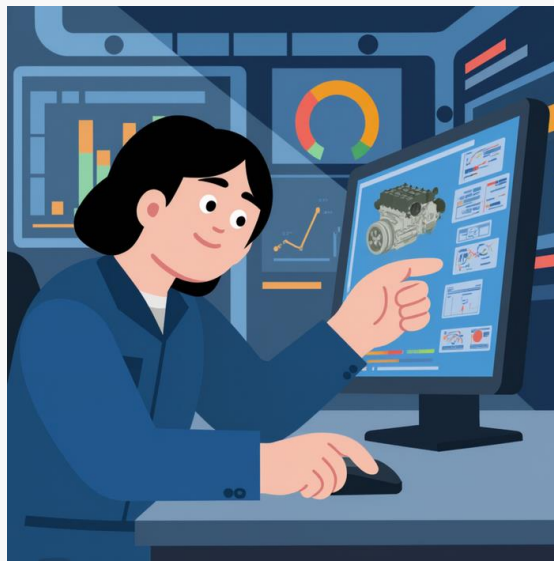


By:  
Sahab Al-Amri  
Maryam AL-Aamri



# WHAT IS THE OBJECTIVE OF THIS PROJECT?

**After this lesson, you will be able to:**



Analyzing real-world  
data using Pandas +  
Matplotlib



Data cleaning + Data  
exploration



Performing statistical  
analysis + Graphing



Extracting useful  
insights

# WHAT HAPPENS IN THE BODY AND BRAIN DURING DEPRESSION?

Depression is a mental disorder that causes persistent sadness and loss of interest in life. It results from stress or a chemical imbalance in the brain, affecting sleep, energy, and concentration.



# DATASET DESCRIPTION:

Data source: [Kaggle](#)  
[Student Depression Dataset](#)

This dataset compiles a wide range of information aimed at understanding, analyzing, and predicting depression levels among students. It is designed for research in psychology, data science, and education, providing insights into factors that contribute to student mental health challenges and aiding in the design of early intervention strategies.




# READING THE DATASET

```
df = pd.read_csv("Student Depression Dataset.csv")
df
```

|   | id | Gender | Age  | City          | Profession | Academic Pressure | Work Pressure | CGPA | Study Satisfaction | Job Satisfaction | Sleep Duration    | Dietary Habits | Degree  | Have you ever had suicidal thoughts ? | Work/Study Hours |
|---|----|--------|------|---------------|------------|-------------------|---------------|------|--------------------|------------------|-------------------|----------------|---------|---------------------------------------|------------------|
| 0 | 2  | Male   | 33.0 | Visakhapatnam | Student    | 5.0               | 0.0           | 8.97 | 2.0                | 0.0              | 5-6 hours         | Healthy        | B.Pharm | Yes                                   | 3.0              |
| 1 | 8  | Female | 24.0 | Bangalore     | Student    | 2.0               | 0.0           | 5.90 | 5.0                | 0.0              | 5-6 hours         | Moderate       | BSc     | No                                    | 3.0              |
| 2 | 26 | Male   | 31.0 | Srinagar      | Student    | 3.0               | 0.0           | 7.03 | 5.0                | 0.0              | Less than 5 hours | Healthy        | BA      | No                                    | 9.0              |
| 3 | 30 | Female | 28.0 | Varanasi      | Student    | 3.0               | 0.0           | 5.59 | 2.0                | 0.0              | 7-8 hours         | Moderate       | BCA     | Yes                                   | 4.0              |

# DATA CLEANING PROCESS

 After reading the dataset we preform data cleaning

```
#Data Cleaning
df.drop(columns=['Work Pressure'], inplace=True)
df.drop(columns=['Study Satisfaction'], inplace=True)
df.drop(columns=['Job Satisfaction'], inplace=True)
df.drop(columns=['Have you ever had suicidal thoughts ?'], inplace=True)
df.drop(columns=['Family History of Mental Illness'], inplace=True)
df.drop(columns=['Financial Stress'], inplace=True)
```

```
df.drop(df.index[200:], inplace=True)
```

|     | id  | Gender | Age  | City          | Profession | Academic Pressure | CGPA | Sleep Duration    | Dietary Habits | Degree   | Work/Study Hours | Depression |
|-----|-----|--------|------|---------------|------------|-------------------|------|-------------------|----------------|----------|------------------|------------|
| 0   | 2   | Male   | 33.0 | Visakhapatnam | Student    | 5.0               | 8.97 | 5-6 hours         | Healthy        | B.Pharm  | 3.0              | 1          |
| 1   | 8   | Female | 24.0 | Bangalore     | Student    | 2.0               | 5.90 | 5-6 hours         | Moderate       | BSc      | 3.0              | 0          |
| 2   | 26  | Male   | 31.0 | Srinagar      | Student    | 3.0               | 7.03 | Less than 5 hours | Healthy        | BA       | 9.0              | 0          |
| 3   | 30  | Female | 28.0 | Varanasi      | Student    | 3.0               | 5.59 | 7-8 hours         | Moderate       | BCA      | 4.0              | 1          |
| 4   | 32  | Female | 25.0 | Jaipur        | Student    | 4.0               | 8.13 | 5-6 hours         | Moderate       | M.Tech   | 1.0              | 0          |
| ... | ... | ...    | ...  | ...           | ...        | ...               | ...  | ...               | ...            | ...      | ...              | ...        |
| 195 | 947 | Male   | 33.0 | Kanpur        | Student    | 2.0               | 9.54 | 7-8 hours         | Healthy        | MA       | 11.0             | 0          |
| 196 | 949 | Female | 20.0 | Pune          | Student    | 5.0               | 6.17 | More than 8 hours | Healthy        | Class 12 | 7.0              | 1          |
| 197 | 953 | Female | 25.0 | Agra          | Student    | 2.0               | 9.95 | More than 8 hours | Unhealthy      | B.Pharm  | 10.0             | 1          |
| 198 | 955 | Male   | 24.0 | Hyderabad     | Student    | 5.0               | 9.05 | 7-8 hours         | Healthy        | B.Tech   | 8.0              | 1          |
| 199 | 956 | Male   | 23.0 | Agra          | Student    | 4.0               | 9.24 | Less than 5 hours | Healthy        | B.Com    | 7.0              | 1          |

# ADDING NULL VALUES TO AGE COLUMNS

```
#change Age to null in 2 rows where id is [2,5]
df.loc[[2,5], 'Age'] = np.nan
df
```

|     | id  | Gender | Age  | City          | Profession | Academic Pressure | CGPA | Sleep Duration    | Dietary Habits | Degree   | Work/Study Hours | Depression |
|-----|-----|--------|------|---------------|------------|-------------------|------|-------------------|----------------|----------|------------------|------------|
| 0   | 2   | Male   | 33.0 | Visakhapatnam | Student    | 5.0               | 8.97 | 5-6 hours         | Healthy        | B.Pharm  | 3.0              | 1          |
| 1   | 8   | Female | 24.0 | Bangalore     | Student    | 2.0               | 5.90 | 5-6 hours         | Moderate       | BSc      | 3.0              | 0          |
| 2   | 26  | Male   | NaN  | Srinagar      | Student    | 3.0               | 7.03 | Less than 5 hours | Healthy        | BA       | 9.0              | 0          |
| 3   | 30  | Female | 28.0 | Varanasi      | Student    | 3.0               | 5.59 | 7-8 hours         | Moderate       | BCA      | 4.0              | 1          |
| 4   | 32  | Female | 25.0 | Jaipur        | Student    | 4.0               | 8.13 | 5-6 hours         | Moderate       | M.Tech   | 1.0              | 0          |
| ... | ... | ...    | ...  | ...           | ...        | ...               | ...  | ...               | ...            | ...      | ...              | ...        |
| 195 | 947 | Male   | 33.0 | Kanpur        | Student    | 2.0               | 9.54 | 7-8 hours         | Healthy        | MA       | 11.0             | 0          |
| 196 | 949 | Female | 20.0 | Pune          | Student    | 5.0               | 6.17 | More than 8 hours | Healthy        | Class 12 | 7.0              | 1          |
| 197 | 953 | Female | 25.0 | Agra          | Student    | 2.0               | 9.95 | More than 8 hours | Unhealthy      | B.Pharm  | 10.0             | 1          |
| 198 | 955 | Male   | 24.0 | Hyderabad     | Student    | 5.0               | 9.05 | 7-8 hours         | Healthy        | B.Tech   | 8.0              | 1          |
| 199 | 956 | Male   | 23.0 | Agra          | Student    | 4.0               | 9.24 | Less than 5 hours | Healthy        | B.Com    | 7.0              | 1          |

200 rows × 12 columns

# CHECK IF THERE ARE NULL VALUES THEN FILL THE MISSING VALUES WITH FORWARD FILL



```
df.isnull().sum()
```

|                   |   |
|-------------------|---|
| id                | 0 |
| Gender            | 0 |
| Age               | 2 |
| City              | 0 |
| Profession        | 0 |
| Academic Pressure | 0 |
| CGPA              | 0 |
| Sleep Duration    | 0 |
| Dietary Habits    | 0 |
| Degree            | 0 |
| Work/Study Hours  | 0 |
| Depression        | 0 |
| dtype: int64      |   |

```
df["Age"] = df["Age"].ffill()  
df
```

|     | id  | Gender | Age  | City          | Profession | Academic Pressure | CGPA | Sleep Duration    | Dietary Habits | Degree   | Work/Study Hours | Depress |
|-----|-----|--------|------|---------------|------------|-------------------|------|-------------------|----------------|----------|------------------|---------|
| 0   | 2   | Male   | 33.0 | Visakhapatnam | Student    | 5.0               | 8.97 | 5-6 hours         | Healthy        | B.Pharm  | 3.0              |         |
| 1   | 8   | Female | 24.0 | Bangalore     | Student    | 2.0               | 5.90 | 5-6 hours         | Moderate       | BSc      | 3.0              |         |
| 2   | 26  | Male   | 24.0 | Srinagar      | Student    | 3.0               | 7.03 | Less than 5 hours | Healthy        | BA       | 9.0              |         |
| 3   | 30  | Female | 28.0 | Varanasi      | Student    | 3.0               | 5.59 | 7-8 hours         | Moderate       | BCA      | 4.0              |         |
| 4   | 32  | Female | 25.0 | Jaipur        | Student    | 4.0               | 8.13 | 5-6 hours         | Moderate       | M.Tech   | 1.0              |         |
| ... | ... | ...    | ...  | ...           | ...        | ...               | ...  | ...               | ...            | ...      | ...              |         |
| 195 | 947 | Male   | 33.0 | Kanpur        | Student    | 2.0               | 9.54 | 7-8 hours         | Healthy        | MA       | 11.0             |         |
| 196 | 949 | Female | 20.0 | Pune          | Student    | 5.0               | 6.17 | More than 8 hours | Healthy        | Class 12 | 7.0              |         |
| 197 | 953 | Female | 25.0 | Agra          | Student    | 2.0               | 9.95 | More than 8 hours | Unhealthy      | B.Pharm  | 10.0             |         |
| 198 | 955 | Male   | 24.0 | Hyderabad     | Student    | 5.0               | 9.05 | 7-8 hours         | Healthy        | B.Tech   | 8.0              |         |
| 199 | 956 | Male   | 23.0 | Agra          | Student    | 4.0               | 9.24 | Less than 5 hours | Healthy        | B.Com    | 7.0              |         |

## CHECK IF THERE IS ANY LEFT NULL VALUES

```
#To check if there is more null values  
df.isna().sum()
```

|                   |   |
|-------------------|---|
| id                | 0 |
| Gender            | 0 |
| Age               | 0 |
| City              | 0 |
| Profession        | 0 |
| Academic Pressure | 0 |
| CGPA              | 0 |
| Sleep Duration    | 0 |
| Dietary Habits    | 0 |
| Degree            | 0 |
| Work/Study Hours  | 0 |
| Depression        | 0 |
| dtype: int64      |   |

# INITIAL DATA EXPLORATION

## TO DISPLAY FIRST 10 ROWS OF THE DATASET USE DF.HEAD(10)

```
# Display the first 10 rows of the dataset  
df.head(10)
```

|   | id | Gender | Age  | City          | Profession | Academic Pressure | CGPA | Sleep Duration    | Dietary Habits | Degree   | Work/Study Hours | Depression |
|---|----|--------|------|---------------|------------|-------------------|------|-------------------|----------------|----------|------------------|------------|
| 0 | 2  | Male   | 33.0 | Visakhapatnam | Student    | 5.0               | 8.97 | 5-6 hours         | Healthy        | B.Pharm  | 3.0              | 1          |
| 1 | 8  | Female | 24.0 | Bangalore     | Student    | 2.0               | 5.90 | 5-6 hours         | Moderate       | BSc      | 3.0              | 0          |
| 2 | 26 | Male   | 24.0 | Srinagar      | Student    | 3.0               | 7.03 | Less than 5 hours | Healthy        | BA       | 9.0              | 0          |
| 3 | 30 | Female | 28.0 | Varanasi      | Student    | 3.0               | 5.59 | 7-8 hours         | Moderate       | BCA      | 4.0              | 1          |
| 4 | 32 | Female | 25.0 | Jaipur        | Student    | 4.0               | 8.13 | 5-6 hours         | Moderate       | M.Tech   | 1.0              | 0          |
| 5 | 33 | Male   | 25.0 | Pune          | Student    | 2.0               | 5.70 | Less than 5 hours | Healthy        | PhD      | 4.0              | 0          |
| 6 | 52 | Male   | 30.0 | Thane         | Student    | 3.0               | 9.54 | 7-8 hours         | Healthy        | BSc      | 1.0              | 0          |
| 7 | 56 | Female | 30.0 | Chennai       | Student    | 2.0               | 8.04 | Less than 5 hours | Unhealthy      | Class 12 | 0.0              | 0          |
| 8 | 59 | Male   | 28.0 | Nagpur        | Student    | 3.0               | 9.79 | 7-8 hours         | Moderate       | B.Ed     | 12.0             | 1          |
| 9 | 62 | Male   | 31.0 | Nashik        | Student    | 2.0               | 8.38 | Less than 5 hours | Moderate       | LLB      | 2.0              | 1          |

# TO SHOW DATASET INFORMATION (.INFO())

## DESCRIPTIVE STATISTICS (.DESCRIBE())

`df.info()`

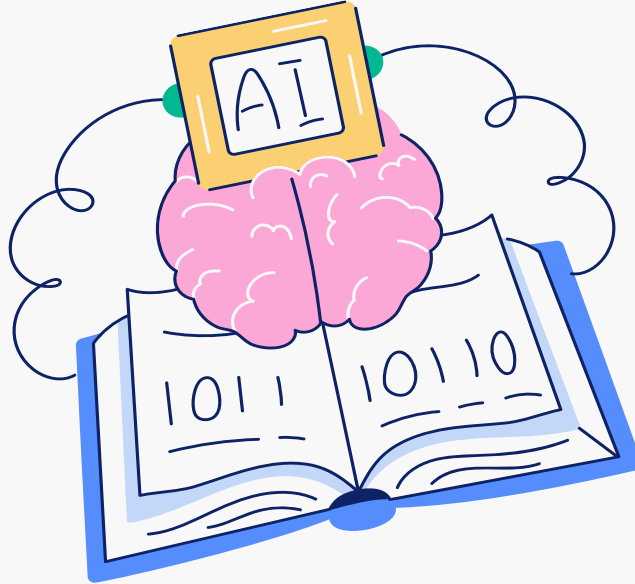
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27901 entries, 0 to 27900
Data columns (total 12 columns):
#   Column                Non-Null Count  Dtype
---  -
0   id                    27901 non-null  int64
1   Gender               27901 non-null  object
2   Age                  27901 non-null  float64
3   City                 27901 non-null  object
4   Profession           27901 non-null  object
5   Academic Pressure    27901 non-null  float64
6   CGPA                 27901 non-null  float64
7   Sleep Duration       27901 non-null  object
8   Dietary Habits       27901 non-null  object
9   Degree               27901 non-null  object
10  Work/Study Hours     27901 non-null  float64
11  Depression            27901 non-null  int64
dtypes: float64(4), int64(2), object(6)
memory usage: 2.6+ MB
```

`df.describe()`

|       | id         | Age        | Academic Pressure | CGPA       | Work/Study Hours | Depression |
|-------|------------|------------|-------------------|------------|------------------|------------|
| count | 200.000000 | 200.000000 | 200.000000        | 200.000000 | 200.000000       | 200.000000 |
| mean  | 526.355000 | 25.505000  | 3.020000          | 7.63090    | 7.240000         | 0.565000   |
| std   | 262.063495 | 4.895899   | 1.406946          | 1.49242    | 3.767409         | 0.497001   |
| min   | 2.000000   | 18.000000  | 1.000000          | 5.11000    | 0.000000         | 0.000000   |
| 25%   | 314.250000 | 21.000000  | 2.000000          | 6.16750    | 4.000000         | 0.000000   |
| 50%   | 536.500000 | 25.000000  | 3.000000          | 7.86000    | 8.000000         | 1.000000   |
| 75%   | 746.750000 | 30.000000  | 4.000000          | 8.94250    | 10.000000        | 1.000000   |
| max   | 956.000000 | 34.000000  | 5.000000          | 9.96000    | 12.000000        | 1.000000   |



# TO COUNT UNIQUE VALUES IN CATEGORICAL COLUMNS



```
df["City"].unique()
```

```
array(['Visakhapatnam', 'Bangalore', 'Srinagar', 'Varanasi', 'Jaipur',  
      'Pune', 'Thane', 'Chennai', 'Nagpur', 'Nashik', 'Vadodara',  
      'Kalyan', 'Rajkot', 'Ahmedabad', 'Kolkata', 'Mumbai', 'Lucknow',  
      'Indore', 'Surat', 'Ludhiana', 'Bhopal', 'Meerut', 'Agra',  
      'Ghaziabad', 'Hyderabad', 'Vasai-Virar', 'Kanpur', 'Patna',  
      'Faridabad', 'Delhi'], dtype=object)
```

# TO CALCULATE THE AVERAGE PERFORMANCE SCORE OF EACH STUDENT

```
df["Average_Score"] = df[["CGPA"]].mean(axis=1)  
df.head(10)
```

|   | id | Gender | Age  | City          | Profession | Academic Pressure | CGPA | Sleep Duration    | Dietary Habits | Degree   | Work/Study Hours | Depression | Average_Score |
|---|----|--------|------|---------------|------------|-------------------|------|-------------------|----------------|----------|------------------|------------|---------------|
| 0 | 2  | Male   | 33.0 | Visakhapatnam | Student    | 5.0               | 8.97 | 5-6 hours         | Healthy        | B.Pharm  | 3.0              | 1          | 8.97          |
| 1 | 8  | Female | 24.0 | Bangalore     | Student    | 2.0               | 5.90 | 5-6 hours         | Moderate       | BSc      | 3.0              | 0          | 5.90          |
| 2 | 26 | Male   | 24.0 | Srinagar      | Student    | 3.0               | 7.03 | Less than 5 hours | Healthy        | BA       | 9.0              | 0          | 7.03          |
| 3 | 30 | Female | 28.0 | Varanasi      | Student    | 3.0               | 5.59 | 7-8 hours         | Moderate       | BCA      | 4.0              | 1          | 5.59          |
| 4 | 32 | Female | 25.0 | Jaipur        | Student    | 4.0               | 8.13 | 5-6 hours         | Moderate       | M.Tech   | 1.0              | 0          | 8.13          |
| 5 | 33 | Male   | 25.0 | Pune          | Student    | 2.0               | 5.70 | Less than 5 hours | Healthy        | PhD      | 4.0              | 0          | 5.70          |
| 6 | 52 | Male   | 30.0 | Thane         | Student    | 3.0               | 9.54 | 7-8 hours         | Healthy        | BSc      | 1.0              | 0          | 9.54          |
| 7 | 56 | Female | 30.0 | Chennai       | Student    | 2.0               | 8.04 | Less than 5 hours | Unhealthy      | Class 12 | 0.0              | 0          | 8.04          |
| 8 | 59 | Male   | 28.0 | Nagpur        | Student    | 3.0               | 9.79 | 7-8 hours         | Moderate       | B.Ed     | 12.0             | 1          | 9.79          |
| 9 | 62 | Male   | 31.0 | Nashik        | Student    | 2.0               | 8.38 | Less than 5 hours | Moderate       | LLB      | 2.0              | 1          | 8.38          |

# CALCULATE THE MAXIMUM CGPA AND THE STUDENT THAT HAS THE HIGHEST CGPA

```
max_CGPA = df["CGPA"].max()
print("The maximum CGPA:",max_CGPA)
```

The maximum CGPA: 9.96

```
higher_CGPA_Student = df[df["CGPA"] == df["CGPA"].max()]
higher_CGPA_Student
```

|            | id  | Gender | Age  | City | Profession | Academic Pressure | CGPA | Sleep Duration    | Dietary Habits | Degree | Work/Study Hours | Depression | Average_Score |
|------------|-----|--------|------|------|------------|-------------------|------|-------------------|----------------|--------|------------------|------------|---------------|
| <b>185</b> | 882 | Female | 29.0 | Pune | Student    | 1.0               | 9.96 | Less than 5 hours | Healthy        | LLM    | 6.0              | 0          | 9.96          |

# CALCULATE THE MINIMUM CGPA AND THE STUDENT THAT HAS THE LOWEST CGPA

```
# the minimum CGPA
min_CGPA = df["CGPA"].min()
min_CGPA
```

np.float64(5.11)

```
lowest_CGPA_Student = df[df["CGPA"] == df["CGPA"].min()]
lowest_CGPA_Student
```

|            | id  | Gender | Age  | City    | Profession | Academic Pressure | CGPA | Sleep Duration    | Dietary Habits | Degree   | Work/Study Hours | Depression | Average_Score |
|------------|-----|--------|------|---------|------------|-------------------|------|-------------------|----------------|----------|------------------|------------|---------------|
| <b>156</b> | 764 | Male   | 20.0 | Lucknow | Student    | 5.0               | 5.11 | More than 8 hours | Unhealthy      | Class 12 | 11.0             | 1          | 5.11          |

# FILTER ROWS USING CONDITIONS (.LOC[], .ISIN(), .BETWEEN())

using isin method will return the student that are from "Delhi", "Hyderabad", "Pune" only

```
# using isin will return the student that are from "Delhi", "Hyderabad", "Pune" only
```

```
df[df["City"].isin(["Delhi", "Hyderabad", "Pune"])]
```

|     | id  | Gender | Age  | City      | Profession | Academic Pressure | CGPA | Sleep Duration    | Dietary Habits | Degree   | Work/Study Hours | Depression | Average_Score |
|-----|-----|--------|------|-----------|------------|-------------------|------|-------------------|----------------|----------|------------------|------------|---------------|
| 5   | 33  | Male   | 25.0 | Pune      | Student    | 2.0               | 5.70 | Less than 5 hours | Healthy        | PhD      | 4.0              | 0          | 5.70          |
| 47  | 305 | Female | 30.0 | Hyderabad | Student    | 1.0               | 5.86 | Less than 5 hours | Healthy        | BSc      | 1.0              | 0          | 5.86          |
| 54  | 335 | Male   | 24.0 | Hyderabad | Student    | 2.0               | 8.04 | 5-6 hours         | Unhealthy      | M.Tech   | 8.0              | 0          | 8.04          |
| 58  | 343 | Male   | 23.0 | Pune      | Student    | 5.0               | 5.74 | 5-6 hours         | Healthy        | BA       | 11.0             | 1          | 5.74          |
| 70  | 414 | Male   | 33.0 | Hyderabad | Student    | 5.0               | 7.94 | Less than 5 hours | Moderate       | B.Ed     | 8.0              | 0          | 7.94          |
| 87  | 491 | Male   | 19.0 | Pune      | Student    | 3.0               | 9.05 | More than 8 hours | Moderate       | Class 12 | 1.0              | 1          | 9.05          |
| 88  | 494 | Female | 24.0 | Hyderabad | Student    | 3.0               | 6.38 | Less than 5 hours | Healthy        | MCA      | 11.0             | 1          | 6.38          |
| 93  | 522 | Male   | 22.0 | Pune      | Student    | 2.0               | 9.88 | More than 8 hours | Healthy        | BSc      | 0.0              | 0          | 9.88          |
| 95  | 525 | Female | 34.0 | Hyderabad | Student    | 3.0               | 5.32 | 7-8 hours         | Unhealthy      | B.Tech   | 7.0              | 1          | 5.32          |
| 99  | 535 | Male   | 21.0 | Delhi     | Student    | 5.0               | 8.10 | More than 8 hours | Moderate       | B.Arch   | 9.0              | 1          | 8.10          |
| 106 | 587 | Female | 33.0 | Delhi     | Student    | 2.0               | 5.51 | Less than 5 hours | Unhealthy      | BSc      | 12.0             | 1          | 5.51          |
| 110 | 595 | Female | 18.0 | Delhi     | Student    | 5.0               | 9.79 | Less than 5 hours | Unhealthy      | Class 12 | 10.0             | 1          | 9.79          |
| 115 | 617 | Male   | 32.0 | Delhi     | Student    | 3.0               | 5.51 | 7-8 hours         | Moderate       | LLB      | 7.0              | 1          | 5.51          |
| 129 | 687 | Female | 22.0 | Pune      | Student    | 1.0               | 6.17 | 7-8 hours         | Healthy        | BCA      | 10.0             | 1          | 6.17          |
| 138 | 710 | Male   | 34.0 | Pune      | Student    | 3.0               | 8.50 | More than 8 hours | Moderate       | B.Tech   | 3.0              | 0          | 8.50          |
| 163 | 793 | Female | 27.0 | Pune      | Student    | 3.0               | 5.59 | Less than 5 hours | Healthy        | PhD      | 6.0              | 1          | 5.59          |
| 167 | 813 | Female | 24.0 | Pune      | Student    | 3.0               | 8.74 | 7-8 hours         | Unhealthy      | B.Arch   | 0.0              | 1          | 8.74          |

USING .LOC[] METHOD FOR LOOK FOR FIRST 3 STUDENT AND THEIR ACADEMIC PRESSURE .AND

```
# Look for first 3 student and their Academic Pressure
df.loc[:2,"Profession":"Academic Pressure"]
```

|   | Profession | Academic Pressure |
|---|------------|-------------------|
| 0 | Student    | 5.0               |
| 1 | Student    | 2.0               |
| 2 | Student    | 3.0               |

```
# age of student that are between 20,25
df[df["Age"].between(20,25)]
```

|     | id  | Gender | Age  | City      | Profession | Academic Pressure | CGPA | Sleep Duration    | Dietary Habits | Degree   | Work/Study Hours | Depression | Average_Score |
|-----|-----|--------|------|-----------|------------|-------------------|------|-------------------|----------------|----------|------------------|------------|---------------|
| 1   | 8   | Female | 24.0 | Bangalore | Student    | 2.0               | 5.90 | 5-6 hours         | Moderate       | BSc      | 3.0              | 0          | 3.633333      |
| 4   | 32  | Female | 25.0 | Jaipur    | Student    | 4.0               | 8.13 | 5-6 hours         | Moderate       | M.Tech   | 1.0              | 0          | 4.376667      |
| 10  | 83  | Male   | 24.0 | Nagpur    | Student    | 3.0               | 6.10 | 5-6 hours         | Moderate       | Class 12 | 11.0             | 1          | 6.700000      |
| 16  | 120 | Male   | 25.0 | Nashik    | Student    | 5.0               | 6.51 | Less than 5 hours | Unhealthy      | M.Ed     | 2.0              | 1          | 4.503333      |
| 17  | 132 | Female | 20.0 | Ahmedabad | Student    | 5.0               | 7.25 | 5-6 hours         | Healthy        | Class 12 | 10.0             | 1          | 7.416667      |
| ... | ... | ...    | ...  | ...       | ...        | ...               | ...  | ...               | ...            | ...      | ...              | ...        | ...           |
| 190 | 918 | Male   | 24.0 | Kolkata   | Student    | 3.0               | 9.44 | 7-8 hours         | Unhealthy      | B.Ed     | 11.0             | 1          | 7.813333      |
| 196 | 949 | Female | 20.0 | Pune      | Student    | 5.0               | 6.17 | More than 8 hours | Healthy        | Class 12 | 7.0              | 1          | 6.056667      |
| 197 | 953 | Female | 25.0 | Agra      | Student    | 2.0               | 9.95 | More than 8 hours | Unhealthy      | B.Pharm  | 10.0             | 1          | 7.316667      |
| 198 | 955 | Male   | 24.0 | Hyderabad | Student    | 5.0               | 9.05 | 7-8 hours         | Healthy        | B.Tech   | 8.0              | 1          | 7.350000      |
| 199 | 956 | Male   | 23.0 | Agra      | Student    | 4.0               | 9.24 | Less than 5 hours | Healthy        | B.Com    | 7.0              | 1          | 6.746667      |

86 rows × 13 columns

USING .BETWEEN() METHOD FOR RETURN THE STUDENT AGE THAT ARE BETWEEN 20, 25

# GROUP DATA BY A CATEGORY AND CALCULATE AGGREGATED STATISTICS

```
#return how many student in each city  
df.groupby("City").size()
```

| City          |    |
|---------------|----|
| Agra          | 8  |
| Ahmedabad     | 7  |
| Bangalore     | 4  |
| Bhopal        | 8  |
| Chennai       | 4  |
| Delhi         | 4  |
| Faridabad     | 3  |
| Ghaziabad     | 6  |
| Hyderabad     | 8  |
| Indore        | 2  |
| Jaipur        | 8  |
| Kalyan        | 14 |
| Kanpur        | 4  |
| Kolkata       | 10 |
| Lucknow       | 11 |
| Ludhiana      | 7  |
| Meerut        | 4  |
| Mumbai        | 5  |
| Nagpur        | 5  |
| Nashik        | 7  |
| Patna         | 6  |
| Pune          | 10 |
| Rajkot        | 5  |
| Srinagar      | 11 |
| Surat         | 9  |
| Thane         | 9  |
| Vadodara      | 4  |
| Varanasi      | 4  |
| Vasai-Virar   | 8  |
| Visakhapatnam | 5  |

dtype: int64

```
# to know which student has highest academic pressure on his city  
df.groupby("City")["Academic Pressure"].mean().sort_values(ascending=False)
```

| City          |          |
|---------------|----------|
| Bangalore     | 4.000000 |
| Ahmedabad     | 3.857143 |
| Delhi         | 3.750000 |
| Bhopal        | 3.750000 |
| Vasai-Virar   | 3.500000 |
| Nagpur        | 3.400000 |
| Rajkot        | 3.400000 |
| Lucknow       | 3.363636 |
| Faridabad     | 3.333333 |
| Mumbai        | 3.200000 |
| Hyderabad     | 3.125000 |
| Kalyan        | 3.071429 |
| Kolkata       | 3.000000 |
| Indore        | 3.000000 |
| Meerut        | 3.000000 |
| Patna         | 3.000000 |
| Ludhiana      | 3.000000 |
| Jaipur        | 2.875000 |
| Srinagar      | 2.818182 |
| Pune          | 2.800000 |
| Visakhapatnam | 2.800000 |
| Varanasi      | 2.750000 |
| Nashik        | 2.714286 |
| Thane         | 2.666667 |
| Agra          | 2.625000 |
| Chennai       | 2.500000 |
| Kanpur        | 2.500000 |
| Ghaziabad     | 2.500000 |
| Vadodara      | 2.500000 |
| Surat         | 2.222222 |

Name: Academic Pressure, dtype: float64

# PERFORM DATA VISUALIZATION USING MATPLOTLIB

## Relationship: Age vs Depression (Line Plot)

### Question:

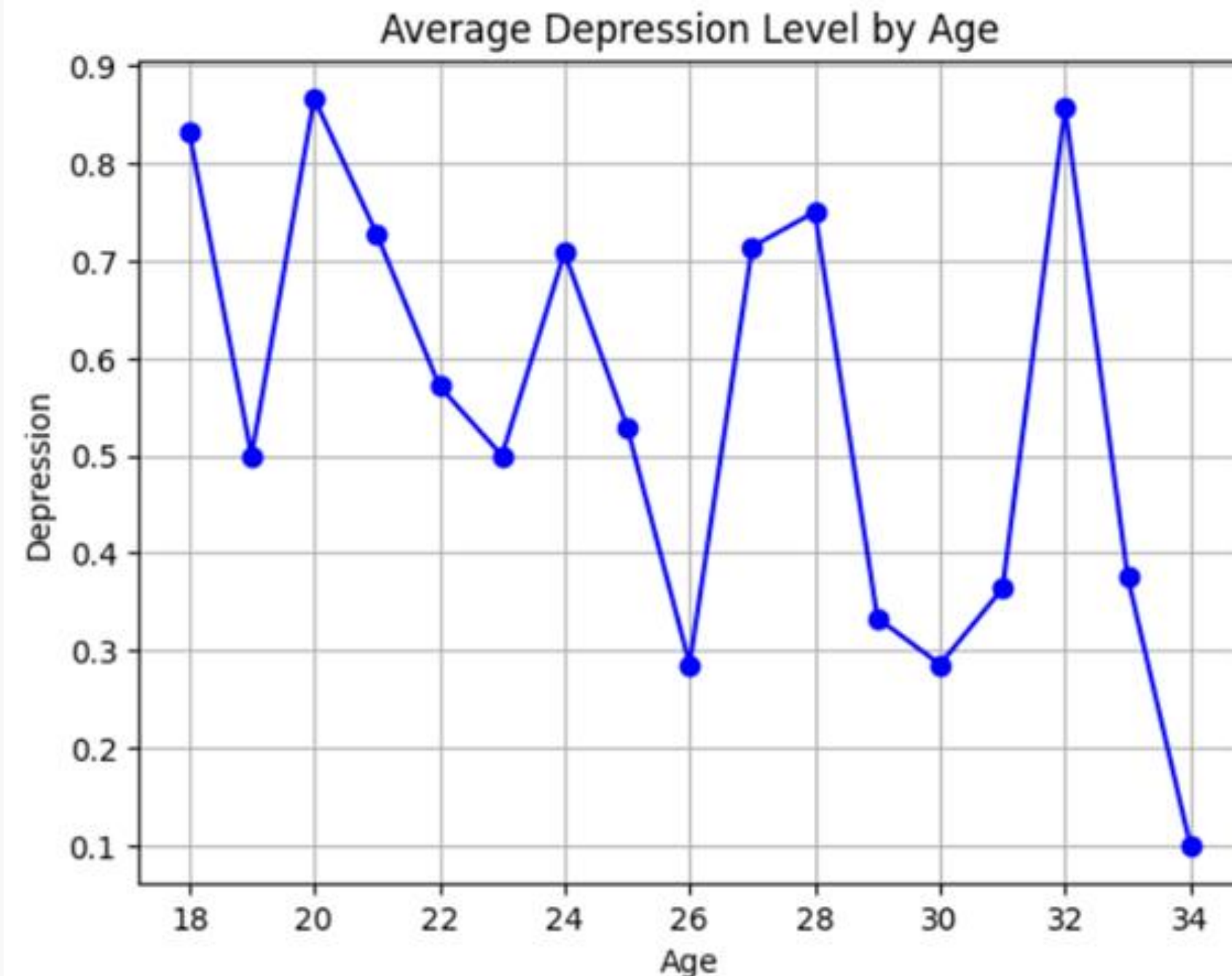
Is depression related to age?

### Insight:

- Depression levels fluctuate across ages
- Slight decrease after early 20s → older students show more stability

```
df_sorted = df.sort_values(by="Age")
age_group = df_sorted.groupby("Age")["Depression"].mean().reset_index()
plt.plot(age_group["Age"], age_group["Depression"], marker = "o", linestyle="-", color='blue')

plt.title("Average Depression Level by Age")
plt.xlabel("Age")
plt.ylabel("Depression")
plt.grid(True)
plt.show()
```



# PERFORM DATA VISUALIZATION USING MATPLOTLIB

## Sleep Duration & Academic Pressure (Bar Plot)

### Question:

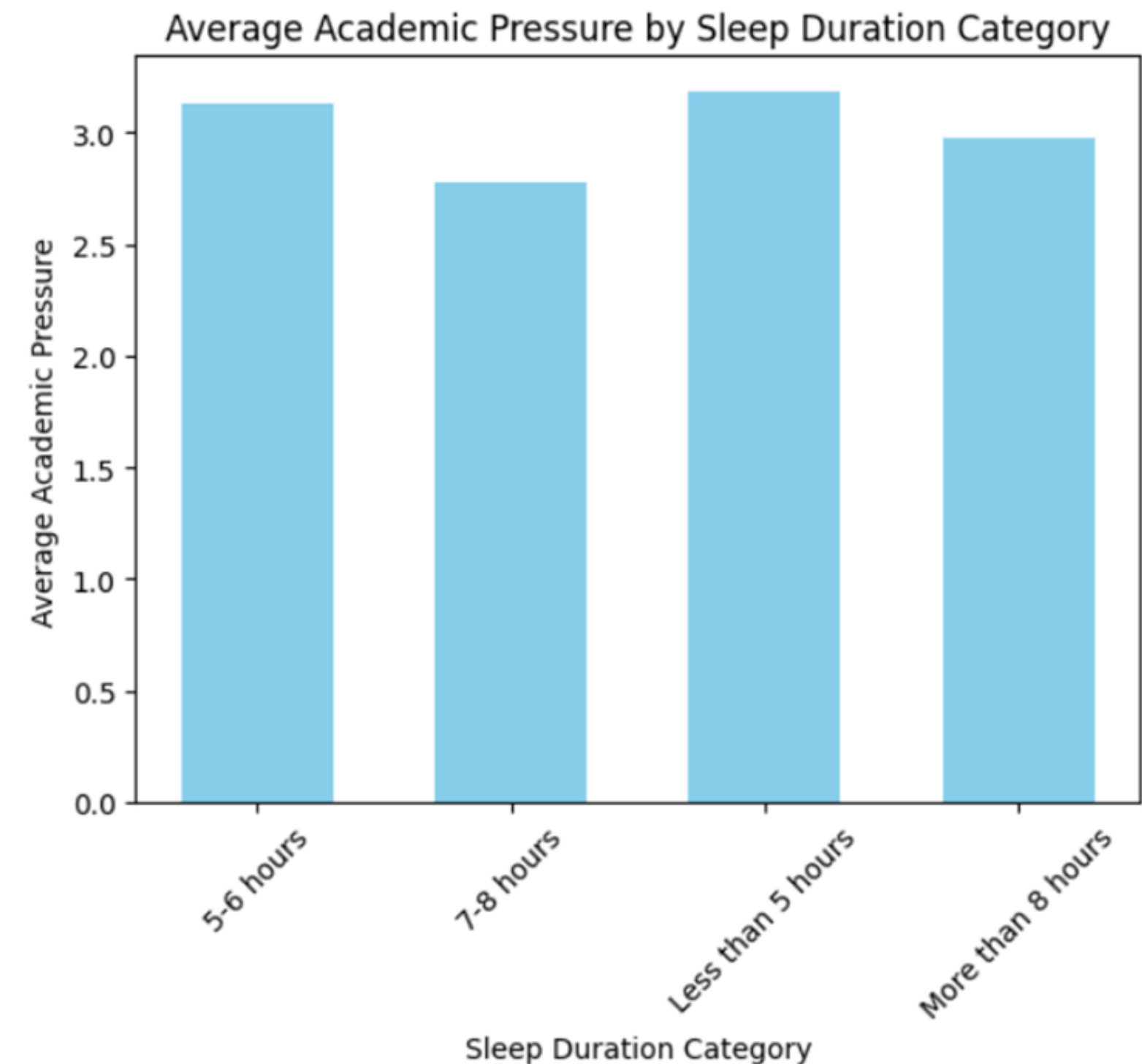
How does sleep affect academic pressure?

### Insight:

- Students sleeping <5 hours or 5–6 hours → highest pressure
- Students sleeping 7–8 hours → lowest academic pressure

```
avg_pressure = df.groupby("Sleep Duration")["Academic Pressure"].mean()

plt.bar(avg_pressure.index, avg_pressure.values, color="skyblue", width=0.6)
plt.title("Average Academic Pressure by Sleep Duration Category")
plt.xlabel("Sleep Duration Category")
plt.ylabel("Average Academic Pressure")
plt.xticks(rotation=45)
plt.show()
```



# PERFORM DATA VISUALIZATION USING MATPLOTLIB

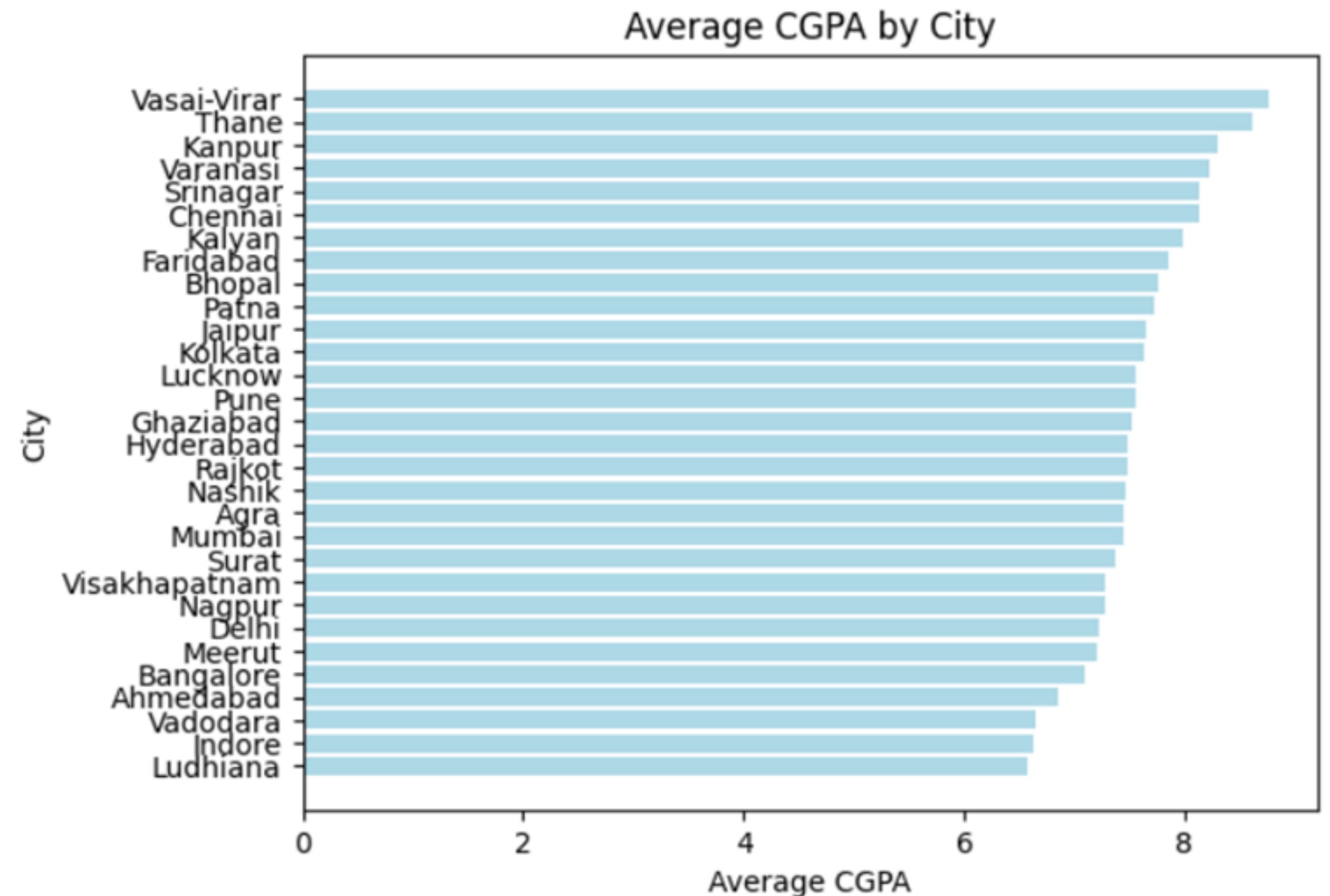
## Academic Performance by City (Horizontal Bar Plot)

### Insight:

- Highest CGPA: Vasai-Virar, Thane, Kanpur
- Lowest CGPA: Indore, Vadodara, Ludhiana

```
avg_cgpa = df.groupby("City")["CGPA"].mean().sort_values()

plt.barh(avg_cgpa.index, avg_cgpa.values, color="lightblue")
plt.title("Average CGPA by City")
plt.xlabel("Average CGPA")
plt.ylabel("City")
plt.show()
```



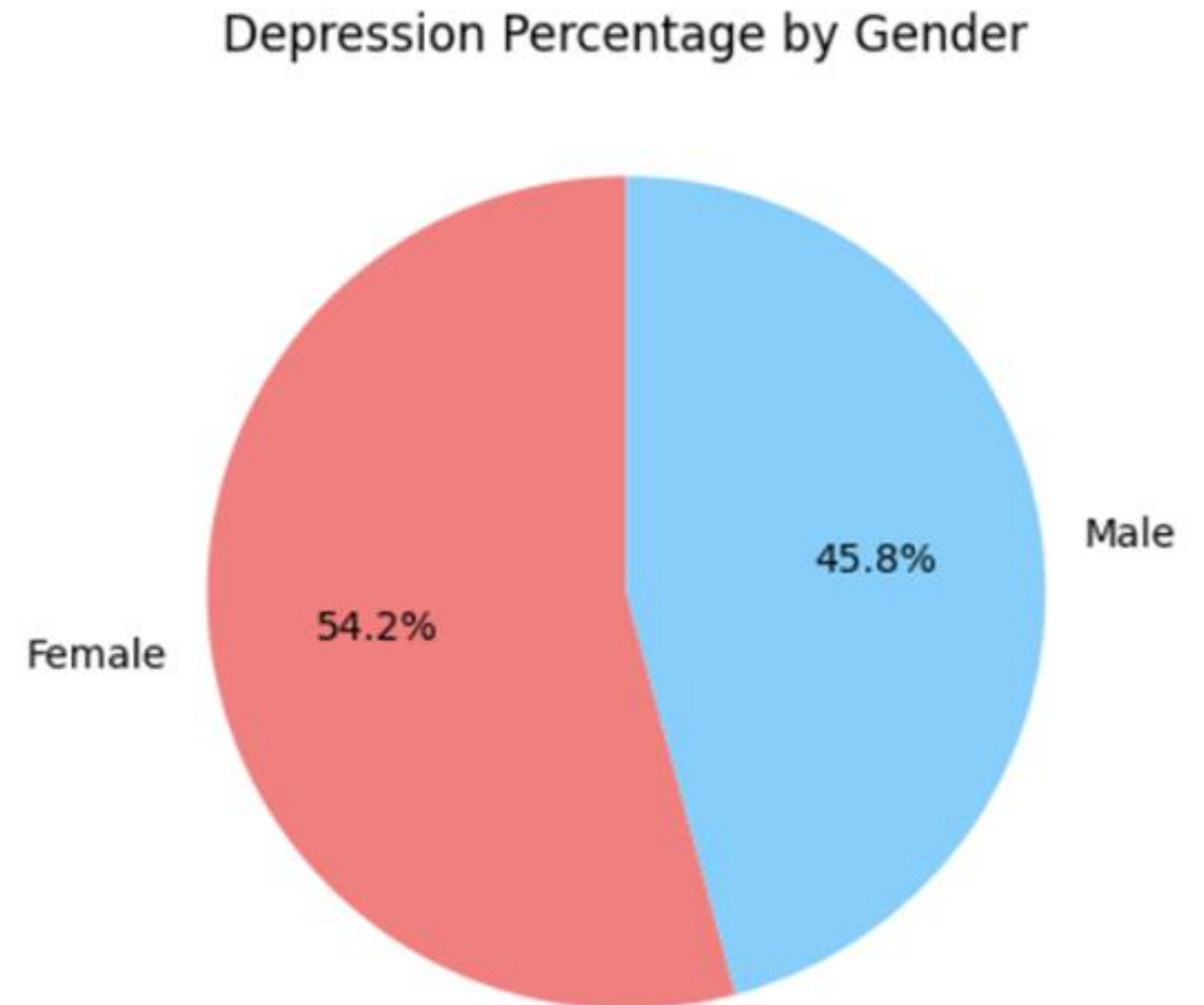
# PERFORM DATA VISUALIZATION USING MATPLOTLIB

## Depression by Gender (Pie Chart)

### Insight:

- Female: 54.2%
- Male: 45.8%
  - Slightly higher depression among female students

```
depression_rate = df.groupby("Gender")["Depression"].mean() * 100
plt.pie(depression_rate,
        labels=depression_rate.index,
        autopct='%1.1f%%',
        startangle=90,
        colors=["lightcoral", "lightskyblue"])
plt.title("Depression Percentage by Gender")
plt.show()
```



# PERFORM DATA VISUALIZATION USING MATPLOTLIB

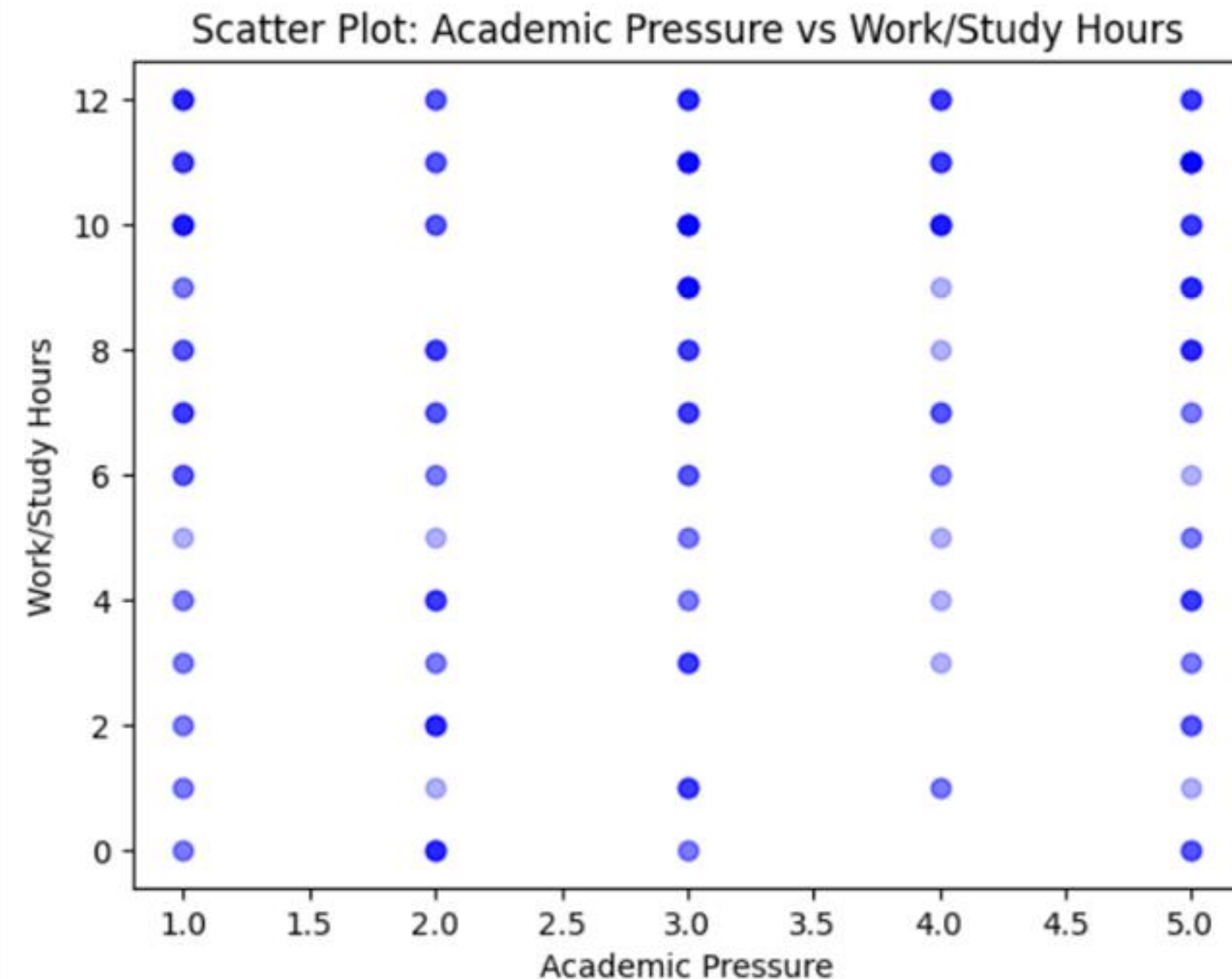
```
plt.scatter(df["Academic Pressure"], df["Work/Study Hours"], color="blue", marker="o", alpha=0.3)
plt.title("Scatter Plot: Academic Pressure vs Work/Study Hours")
plt.xlabel("Academic Pressure")
plt.ylabel("Work/Study Hours")

plt.show()
```

## Academic Pressure & Study Hours (Scatter Plot)

### Insight:

- Longer study/work hours → higher academic pressure
- Suggests that workload difficulty, not just time, affects stress



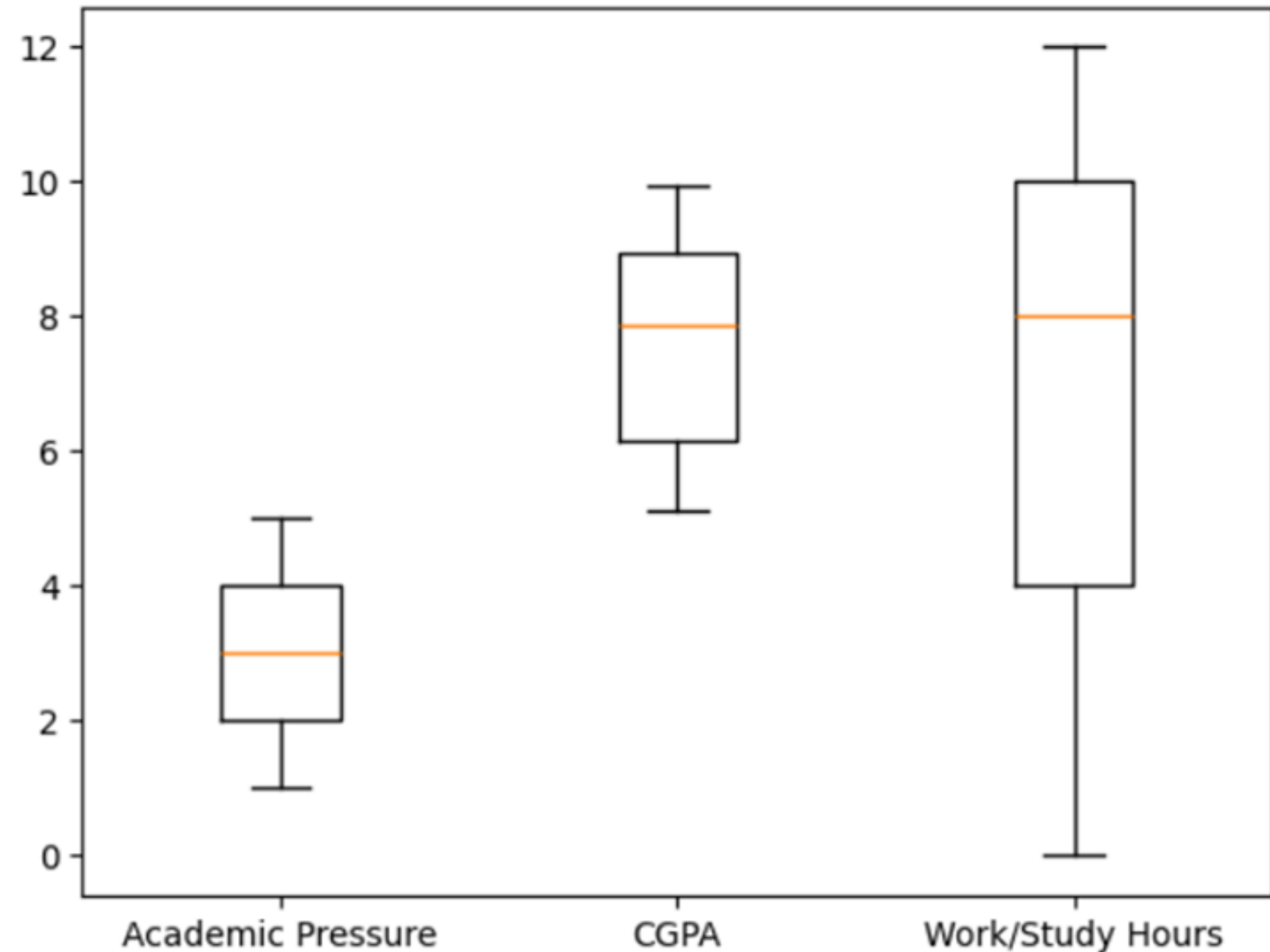
# PERFORM DATA VISUALIZATION USING MATPLOTLIB

## Variability in Factors (Box Plot)

### Insight:

- **Work/Study Hours** → highest variability
- **Academic Pressure** → lowest median
- **CGPA** → relatively stable across students

```
plt.boxplot([df["Academic Pressure"], df["CGPA"], df["Work/Study Hours"]],  
            tick_labels=["Academic Pressure", "CGPA", "Work/Study Hours"]  
            )  
plt.show()
```



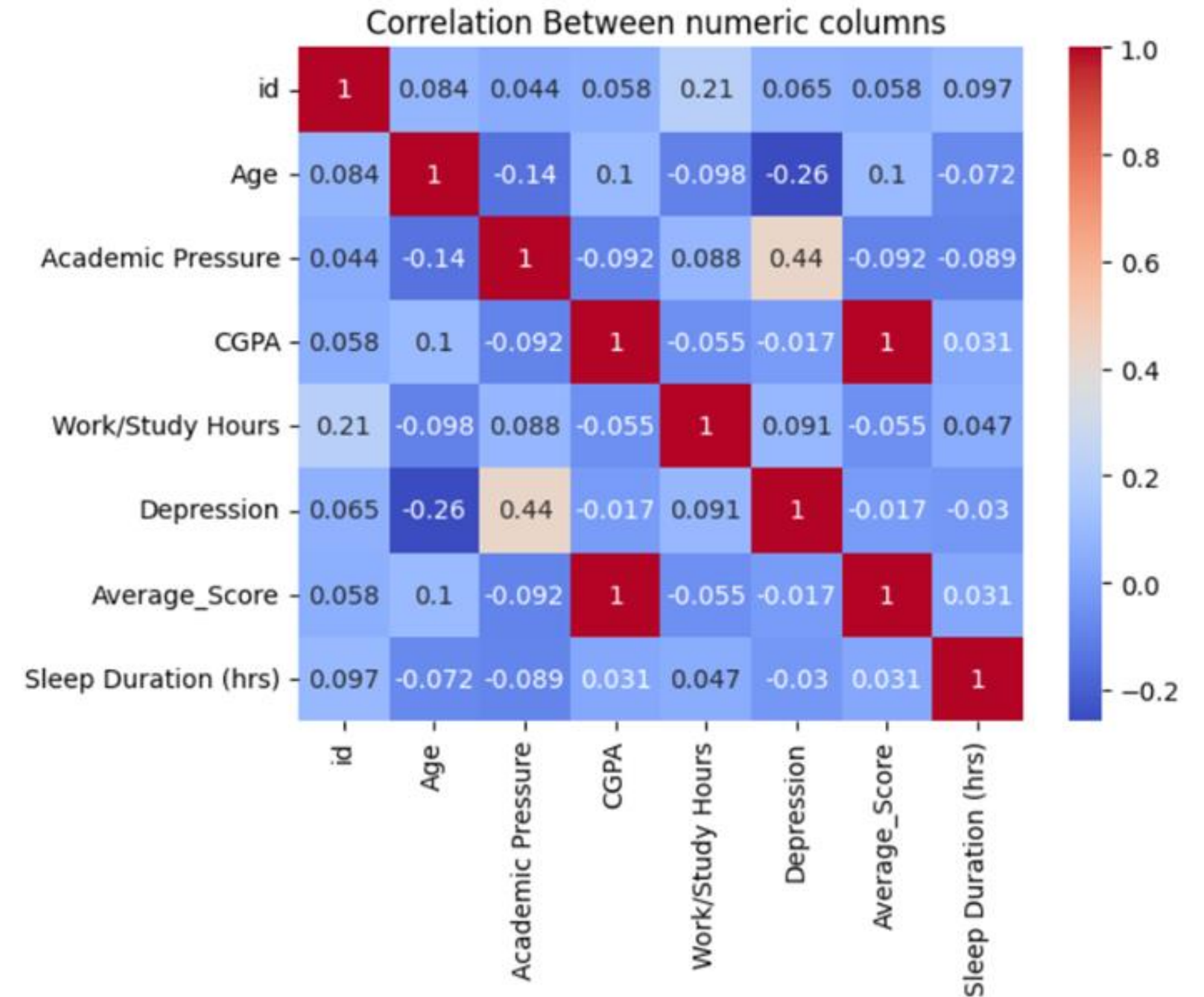
# PERFORM DATA VISUALIZATION USING MATPLOTLIB

## Correlation Heatmap

### Key Correlations:

- Depression ↑ when **Academic Pressure** ↑
- Depression ↓ as **Age** ↑  
→ Younger students with high stress are more vulnerable

```
corr = df.select_dtypes(include='number').corr()  
sns.heatmap(corr, annot=True, cmap="coolwarm")  
plt.title("Correlation Between numeric columns ")  
plt.show()
```



# ADVANCED TASKS (CHALLENGE)

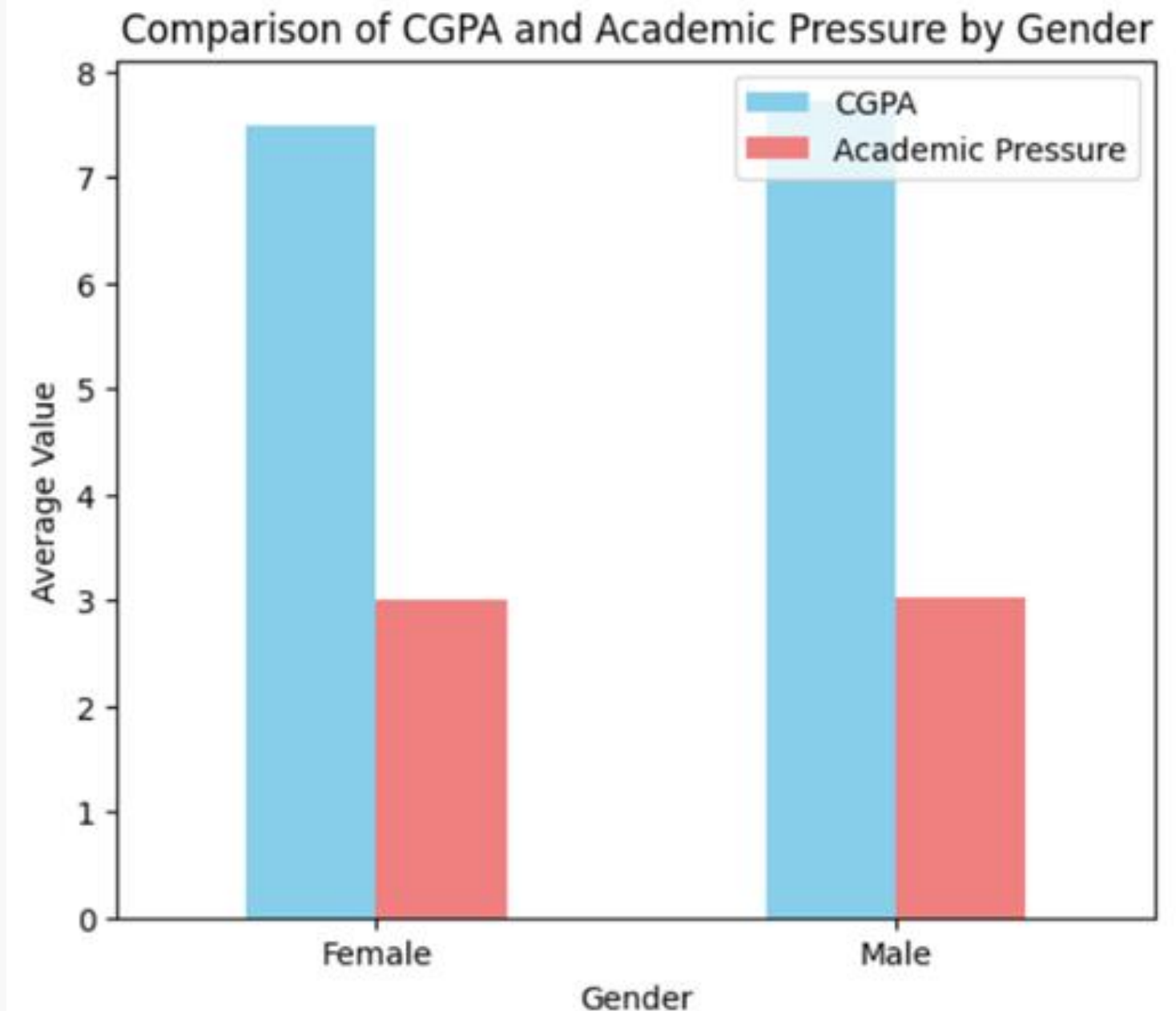
## Gender Comparison: CGPA vs Academic Pressure

### Insight:

- Academic pressure similar for males & females (~3)
  - Females slightly outperform in CGPA
    - Stress affects both genders equally, but females maintain better grades
- form marginally better academically.

```
#Compares two numeric variables (CGPA and Pressure) for each gender group side by side
grouped = df.groupby("Gender")[["CGPA", "Academic Pressure"]].mean()

grouped.plot(kind="bar", figsize=(6,5), color=["skyblue", "lightcoral"])
plt.title("Comparison of CGPA and Academic Pressure by Gender")
plt.ylabel("Average Value")
plt.xticks(rotation=0)
plt.show()
```



## ADVANCED TASKS (CHALLENGE)

## Combine multiple plots in one figure

## Insights:

Depression levels vary by degree (MBA, MPhil higher)

Healthy diet → longer, better sleep

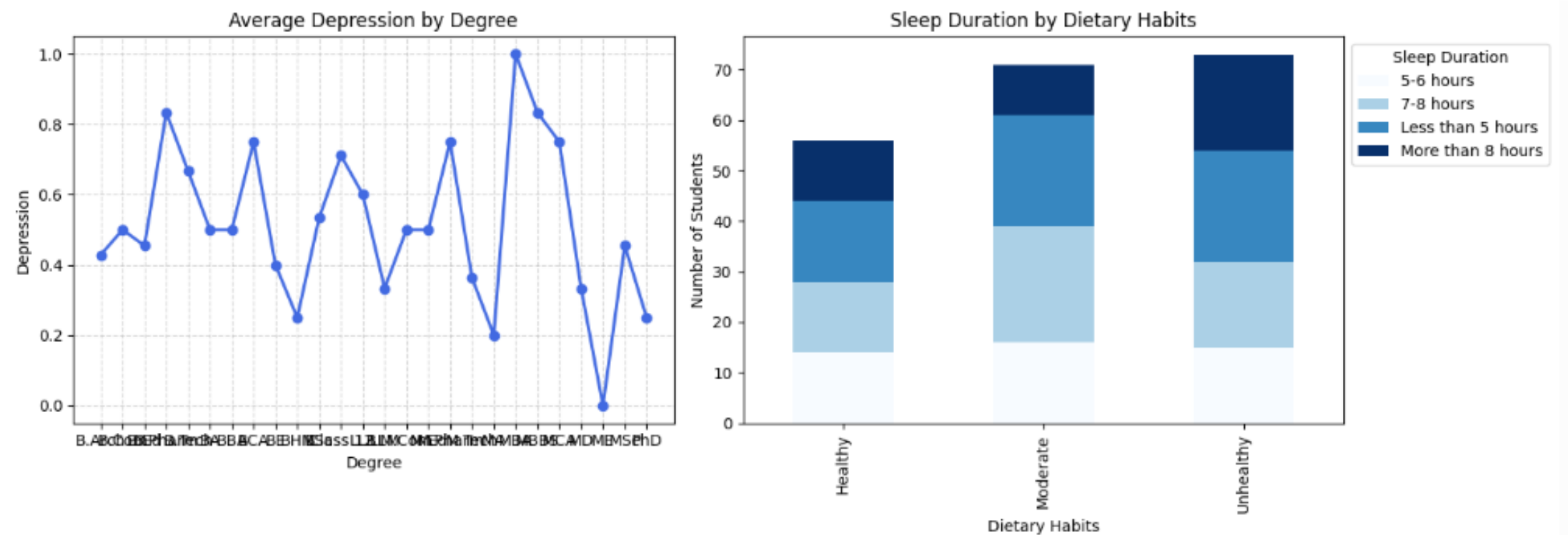
Unhealthy diet → poor or irregular sleep patterns

```
# Create figure with 2 subplots
fig, axs = plt.subplots(1, 2, figsize=(14,5))

# ----- Plot 1: Line plot (Average Depression by Degree) -----
avg_hours_by_age = df.groupby("Degree")["Depression"].mean()
axs[0].plot(avg_hours_by_age.index, avg_hours_by_age.values, color="royalblue", marker='o', linewidth=2)
axs[0].set_title("Average Depression by Degree")
axs[0].set_xlabel("Degree")
axs[0].set_ylabel("Depression")
axs[0].grid(True, linestyle="--", alpha=0.5)

# ----- Plot 2: Bar chart (Sleep Duration by Dietary Habits) -----
sleep_counts = df.groupby("Dietary Habits")["Sleep Duration"].value_counts().unstack().fillna(0)
sleep_counts.plot(kind="bar", ax=axs[1], stacked=True, colormap="Blues")
axs[1].set_title("Sleep Duration by Dietary Habits")
axs[1].set_xlabel("Dietary Habits")
axs[1].set_ylabel("Number of Students")
axs[1].legend(title="Sleep Duration", bbox_to_anchor=(1,1))

plt.tight_layout()
plt.show()
```



# INSIGHTS STORY:

- After analyzing the student depression dataset, I found several interesting patterns. Female students showed slightly higher depression levels (54.2%) than males (45.8%). Younger students were more likely to feel depressed, especially when academic pressure was high. The data also showed that longer study hours did not always mean higher pressure, suggesting that stress is more related to workload difficulty than study time.
- Students with healthy diets and regular sleep (6–8 hours) had better mental well-being, while unhealthy diets were linked to poor or irregular sleep. CGPA remained mostly stable for all students, showing that good performance does not always mean better mental health. Work and study hours peaked around age 26, then gradually decreased with age.
- Overall, the study shows that academic pressure, poor sleep, and unhealthy habits increase depression, while a balanced lifestyle and proper rest can improve students' mental health and academic balance.



**THANKS FOR LISTENING!**  
**THE END!**

