2.⁠ ⁠Perform exploratory data analysis (EDA) and create at least 3 visualizations that

highlight important trends.

**1. Promote Peaceful Study Environments**

**Finding:** Students reporting a **"Peaceful" study environment** generally showed **lower stress levels** compared to those in "Noisy" or "Disrupted" environments.  
**Action:**

* Invest in **quiet study zones** on campus or promote the use of noise-canceling tools.
* Encourage **time management and study planning** to reduce last-minute cramming in noisy settings.

**2. Encourage Intellectual Coping Strategies**

**Finding:** Students who used **"Analyze the situation and handle it with intellect"** as a coping strategy reported **lower stress levels** than those relying on emotional breakdowns or social support alone.  
**Action:**

* Integrate **problem-solving and critical-thinking workshops** into academic support programs.
* Promote **mindfulness and cognitive-behavioral techniques** to help students reframe academic challenges.

**3. Address Peer and Home Pressure**

**Finding:** Higher levels of **peer pressure** and **academic pressure from home** were correlated with **increased stress ratings**.  
**Action:**

* Launch **parental awareness programs** to promote healthier expectations and support systems.
* Introduce **peer-mentoring programs** to reduce competitive anxiety and foster collaborative learning.

4.⁠ ⁠Summarize your analysis with 3 actionable insights that could help decision-

makers in solving the real-world problem

Here are the key visualizations that are used to highlight important trends:

1. **Distribution of Categorical Variables**:
   * The notebook includes bar plots for each categorical column to show the distribution of responses. This helps in understanding the frequency of different categories (e.g., academic stages, study environments, coping strategies, bad habits).
2. **Distribution of Numerical Variables**:
   * Although not explicitly shown in the provided code, the notebook likely includes histograms or bar plots for numerical columns (like 'Peer pressure', 'Academic pressure from home', 'Academic competition', 'Academic stress index') to visualize their distributions. This is inferred from the code that loops through columns and creates bar plots for value counts.
3. **Correlation Analysis**:
   * The notebook uses sns.heatmap() to create a correlation matrix heatmap. This visualization highlights the relationships between numerical variables (e.g., how peer pressure, academic pressure from home, academic competition, and academic stress index are correlated).
4. **Pairplot**:
   * The notebook uses sns.pairplot() to create a grid of scatter plots and histograms. This helps in visualizing the pairwise relationships between numerical variables and the distribution of each variable. It can reveal patterns, correlations, and outliers.
5. **Boxplots for Stress Index by Categories**:
   * The notebook includes boxplots to compare the distribution of the academic stress index across different categories (e.g., academic stages, study environments, coping strategies, bad habits). This helps in understanding how stress levels vary with different factors.
6. **Countplots with Stress Levels**:
   * The notebook uses countplots with hue set to the academic stress index to show how stress levels are distributed within different categories. This can reveal which factors are associated with higher or lower stress.
7. **Pie Charts for Proportional Data**:
   * For some categorical variables, the notebook uses pie charts to show the proportion of responses in each category. This provides a clear view of the composition of the data.
8. **Grouped Bar Plots**:
   * The notebook creates grouped bar plots to show the average stress index for different groups (e.g., average stress index by academic stage and study environment). This helps in comparing stress levels across multiple factors simultaneously.

Summary of Visualizations:

* **Bar plots** for distributions of categorical and numerical variables.
* **Heatmap** for correlations between numerical variables.
* **Pairplot** for pairwise relationships.
* **Boxplots** to compare stress index across categories.
* **Countplots with hue** to see stress distribution within categories.
* **Pie charts** for proportional representation.
* **Grouped bar plots** for comparing average stress across multiple factors.

These visualizations are designed to uncover trends such as:

* Which factors (e.g., peer pressure, academic pressure) are most correlated with stress.
* How stress levels vary across different academic stages or study environments.
* The most common coping strategies and their relationship with stress.
* The prevalence of bad habits and their impact on stress.

**Heart Disease Prediction using Machine Learning - Project Summary**

**Project Overview**

Developed a machine learning solution to predict heart disease in patients using clinical and demographic parameters. The project addresses the critical healthcare challenge of early heart disease detection, as cardiovascular diseases cause approximately 17.9 million deaths globally each year.

**Dataset & Methodology**

* **Data Source**: Cleveland Heart Disease dataset from UCI Machine Learning Repository
* **Sample Size**: 303 patient records with 14 clinical attributes
* **Algorithm**: Logistic Regression (chosen for interpretability in medical applications)
* **Features**: Age, sex, chest pain type, blood pressure, cholesterol, ECG results, exercise-induced parameters, and cardiac indicators

**Key Results**

* **Training Accuracy**: 85.12%
* **Test Accuracy**: 81.97%
* **Data Split**: 80-20 train-test split with stratified sampling
* **Model Performance**: Excellent generalization with minimal overfitting

**Technical Implementation**

* **Languages/Tools**: Python, Pandas, NumPy, Scikit-learn
* **Data Quality**: No missing values detected
* **Validation**: Maintained class distribution across training and test sets

**Real-World Applications**

1. **Clinical Decision Support** - Assist healthcare professionals with data-driven diagnosis
2. **Telemedicine** - Enable remote heart disease screening in underserved areas
3. **Preventive Healthcare** - Early identification of at-risk individuals
4. **Resource Optimization** - Help hospitals prioritize cardiac care patients
5. **Health Monitoring** - Integration potential with wearable devices

**Impact & Considerations**

The solution enables proactive intervention for heart disease patients, potentially improving treatment outcomes and saving lives. The model serves as a decision support tool for healthcare professionals while maintaining the importance of clinical expertise and thorough medical validation.

**Medical Disclaimer**: This model is designed as a support tool and should not replace professional medical diagnosis or treatment decisions.

**Future Enhancements**

* Model ensemble with Random Forest/XGBoost
* Advanced evaluation metrics (ROC curves, precision/recall)
* Web application deployment for clinical integration
* Real-time data integration with electronic health records