**Question 1 : Bean Dataset**

Load Dataset1 and do any 6 of the following:

1. Calculate the mean, median, and standard deviation for the "Area" and "Perimeter" columns. What do these values indicate about the distribution of these features?
2. Find the range and interquartile range (IQR) for "MajorAxisLength" and "MinorAxisLength." What do these values suggest about the shape of the objects in the dataset?
3. Analyze the distribution of the "AspectRation." What does the aspect ratio reveal about the shapes of the objects?
4. How many unique "Class" categories are present in the dataset?
5. Calculate the frequency distribution of each "Class." Which class appears the most, and which appears the least?
6. Is there a relationship between "Area" and "Perimeter"? Use a scatter plot to visualize this relationship and calculate the correlation coefficient.
7. Create a boxplot for "Area" across different classes. What does the boxplot reveal about the distribution of "Area" for each class?
8. Use a heatmap to show the correlation between different shape features (e.g., "Compactness," "ShapeFactor1," "AspectRation"). What patterns or relationships do you observe?
9. Plot a scatter plot of "MajorAxisLength" vs. "MinorAxisLength" and color the points based on the "Class." What does the plot suggest about the relationship between the two features for different classes?
10. Create a scatter plot for "Roundness" vs. "Compactness" for each class. What patterns can you identify from the scatter plot?

**Question 2 : Adult Dataset**

**Load Dataset2 and do the following**

Use the given dataset to build a machine learning model for predicting whether an individual's "salary" is >50K or <=50K.

1. **Data Preprocessing:** Outline the steps you would take to preprocess the data, including handling categorical variables, normalizing numerical features, and dealing with any missing values.
2. **Model Selection:** Choose at least three machine learning algorithms (e.g., Logistic Regression, Decision Tree, Random Forest, Gradient Boosting) to train and evaluate on this classification task. Justify why you selected these models.
3. **Model Evaluation:** Evaluate the models using appropriate metrics such as accuracy, precision, recall, F1-score, and ROC-AUC score. Which model performs the best, and what might explain this result?
4. **Feature Importance Analysis:** Use feature importance techniques (e.g., coefficients from Logistic Regression, feature importance from Decision Trees) to identify the most influential features for predicting "salary." How do features such as "age," "education-num," and "occupation" impact the predictions?
5. **Hyperparameter Tuning:** Perform hyperparameter tuning on the best-performing model using techniques such as grid search or random search. What parameters did you tune, and what impact did the tuning have on the model's performance?

**Question 3 : Adult Dataset (Optional Question)**

**Build a Flask application for the best performance model found above.**