**Part 1.Dataset Preparation and Exploration**

1. **Load the Dataset**
   * Drag the **File** widget onto the canvas.
   * Double-click the **File** widget and load the **adult\_census income** Connect the **File** widget to a **Data Table** widget to inspect the dataset’s contents.
2. **Examine Class Distribution**
   * Add a **Data Info** widget and connect it to the **File** widget to observe class distribution and other dataset properties.
   * Connect **Data Info** to a **Box Plot** widget and check for outliers in each numerical attribute.
3. **Analyze Data Statistics**
   * Add the **Feature Statistics** widget to the canvas and connect it to the **File** widget.
   * Record and interpret summary statistics (mean, median, standard deviation) to understand data variation.

**Part 2. Data Preprocessing and Feature Engineering**

1. **Handle Missing Values**
   * Add the **Select Columns** widget to the canvas and connect it to the **File** widget.
   * In **Select Columns**, remove any columns with excessive missing values (more than 30%).
2. **Impute Missing Data**
   * Add an **Impute** widget and connect it to the **Select Columns** widget.
   * Set **Impute** to replace missing values with the most frequent value for categorical features and the mean for numerical features.
3. **Feature Selection**
   * Add a **Rank** widget and connect it to the **Impute** widget.
   * In the **Rank** widget, choose a ranking method such as **Information Gain** or **ReliefF** to identify the most informative features.
   * Select the top 5 features for modeling.
4. **Principal Component Analysis (PCA)**
   * Add a **PCA** widget and connect it to the **Rank** widget.
   * Set **PCA** to reduce the data to 2 components and observe the explained variance.
   * Connect **PCA** to a **Scatter Plot** widget to visualize the data distribution after dimensionality reduction.

**Part 3. Model Training and Hyperparameter Tuning**

1. **Split Data**
   * Add a **Data Sampler** widget to the canvas and connect it to the **PCA** widget.
   * Set **Data Sampler** to randomly split the dataset into 70% training and 30% testing data.
2. **Train Multiple Models**
   * Add a **Logistic Regression**, **Random Forest**, and **Support Vector Machine (SVM)** widget to the canvas.
   * Connect each model widget to the **Data Sampler** training output (labeled “Data”).
3. **Hyperparameter Tuning**
   * For **Random Forest**
     1. Double-click the widget and set **Number of Trees** to 50, **Max Depth** to 10.
   * For **SVM**
     1. Set **Kernel** to **RBF** and **C** to 1.0. Test additional values of **C** ( 0.5 and 1.5) and note the results.

**Part 4. Model Evaluation and Comparison**

1. **Evaluate Models**
   * Add a **Test & Score** widget to the canvas.
   * Connect **Logistic Regression**, **Random Forest**, and **SVM** to **Test & Score**.
   * Connect the testing data output from **Data Sampler** to **Test & Score**.
2. **Analyze Metrics**
   * In **Test & Score**, observe metrics such as **Accuracy**, **Precision**, **Recall**, **F1 Score**, and **AUC (Area Under the Curve)** for each model.
   * Record the performance of each model, focusing on models that perform best in terms of both accuracy and F1 score.
3. **Plot ROC Curves**
   * Add the **ROC Analysis** widget and connect it to **Test & Score**.
   * Examine the **ROC curve** for each model to evaluate how well the models distinguish between classes.
   * Note which model has the highest AUC, indicating better predictive performance.

**Part 5. Reporting and Insights**

1. **Summarize Findings**
   * Discuss the effectiveness of feature engineering steps like PCA and Rank.
   * Compare model performance and identify the best-performing model.
2. **Interpret Model Metrics**
   * Describe why certain models performed better based on metrics like accuracy, precision, recall, and F1 score.
3. **Recommendations**
   * Provide recommendations for further improvement, such as using a larger dataset or trying additional models.
4. **Take Screenshots**

* Capture screenshots of critical steps and results (feature ranking, PCA scatter plot, ROC curves, and Test & Score results) to include in your report.