You are tasked with developing a model to predict whether patients admitted to the emergency department (ED) should be admitted to the hospital or sent home. This prediction model will help optimize bed management and ensure that patients needing immediate care are identified quickly.

You will use a dataset that contains patient records, including demographics (age, gender), vital signs (blood pressure, heart rate), lab test results, and previous medical history. The target variable in the dataset is binary: `1` indicates that a patient was admitted, while `0` indicates that a patient was not admitted.

1. Data Preparation:

- Load the dataset and perform an initial exploration to understand the features and target variable.
- Split the dataset into 80% training data and 20% testing data. Ensure that both sets are representative of the overall population.

2. Model Development:

- Use logistic regression to develop a binary classification model. Before training the model, ensure that the features are standardized.
- Address any multicollinearity issues that might arise among the features to ensure stable coefficient estimates.

3. Model Evaluation:

- After training, evaluate the model using the test data.
- Display the results using the following:
  - **Confusion Matrix:**To show the counts of true positives, true negatives, false positives, and false negatives.
  - **Precision and Recall:**Calculate these metrics to understand the model's accuracy and its ability to correctly identify patients who need admission.
  - **ROC Curve:**Plot the Receiver Operating Characteristic (ROC) curve to visualize the trade-off between the true positive rate and false positive rate. Also, calculate the Area Under the Curve (AUC) to assess the model’s performance.

4. Visualizations:

- **Sigmoid Function Plot:**Display the sigmoid function curve, which shows how logistic regression maps input features to a probability between 0 and 1.

- Cost Function Graph: Plot the cost function (binary cross-entropy loss) over iterations to show how the model's prediction error decreases during training.

-ROC Curve: Visualize the ROC curve with the AUC score highlighted to indicate the model's effectiveness in distinguishing between the two classes.

-**Residuals Plot:** Display a plot of the residuals (differences between observed and predicted values) to assess how well the model fits the data, particularly to check for any patterns that might suggest a poor model fit.

5. Model Interpretation:

- Interpret the coefficients of the logistic regression model. Identify which features have the most significant impact on the admission decision.
- Visualize the coefficients in a bar chart to show the magnitude and direction (positive or negative) of each feature's influence on the model's predictions.

Expected Output:

- A trained logistic regression model that predicts hospital admissions.
- Confusion matrix, precision, recall, and ROC curve with AUC score.
- Visualizations including the sigmoid function plot, cost function graph, residuals plot,and feature coefficients bar chart.

<u>Evaluation Rubrics</u>

Total:10 Marks
Preprocessing:2 Marks
Visualization:2 Marks
Performance metrics:3 Marks
Sigmoid and Cost function :3 Marks