**Data Analysis Report**

**Name: Abhijit Naikwadi PRN No:** **240340325001**

**Problem Statement**: The goal of this analysis is to predict bankruptcy based on financial indicators using logistic regression.

**Dataset Overview**: The dataset consists of financial metrics of companies, with Bankrupt as the target variable (1 for bankruptcy, 0 for non-bankruptcy).

**Data Distribution**:

* Histograms and box plots were used to visualize the distribution of key variables.
* For example, the Debt ratio % showed a skewed distribution towards lower values.

**Correlation Analysis**:

* A correlation matrix was computed to identify relationships between variables.
* Strong correlations (>0.5) were found between Bankrupt status and certain financial ratios

**Data Cleaning (EDA)**

**Handling Missing Data:**

* No missing values were found in the dataset, ensuring completeness.

**Test Performed**

**Hypothesis Tests**:

* T-tests were conducted to determine significant predictors of bankruptcy.
* Features like Net profit before tax and Retained Earnings to Total Assets showed significant differences between bankrupt and non-bankrupt companies.

**Building Model**

**Logistic Regression Model**:

* Logistic Regression was chosen for its interpretability and ability to handle binary classification tasks.
* The model was trained on the dataset with class weights balanced due to imbalanced classes.

**Model Performance**:

* **Accuracy:** 84%
* **Precision:** 17%
* **Recall:** 84%
* **F1 Score:** 29%

**Confusion Matrix**:

* The confusion matrix illustrates the model's predictions versus actual outcomes

**Key Findings**:

* Companies with higher Debt ratios and current liability to current assets are more likely to be classified as bankrupt.
* The model achieves high recall but low precision, indicating it correctly identifies bankrupt companies but with many false positives