CREDIT RISK ANALYSIS

Data Preparation

- **Data Loading and Exploration**: The notebook starts by loading the dataset and performing exploratory data analysis (EDA). It involves checking for missing values, understanding data distributions, and visualizing relationships between variables using plots.

Feature Engineering

- **Feature Selection and Encoding**: The dataset features are processed, and categorical variables are encoded. Features are selected based on their importance or correlation with the target variable (credit risk).

Model Building

- **XGBoost Model**: Performed 4 different models to evaluate which model gives high accuracy and model is trained on the processed data to predict credit risk.
- **Model Evaluation**: The model's performance is evaluated using metrics like accuracy, precision, recall, F1-score, and ROC AUC.

Visualization

- **Actual vs. Predicted Credit Risk**: The notebook includes a plot showing the relationship between actual and predicted credit risk values, helping visualize model performance.

Results

- **Model Metrics**:
    - Accuracy: 80%
    - Precision: 70%
    - Recall: 65%
    - F1-Score: 67%
    - ROC AUC: 85%

Conclusion

The logistic regression model demonstrates good predictive performance with an ROC AUC of 85%, indicating a strong balance between sensitivity and specificity in predicting credit risk.