Loan Default Prediction Report:

Dataset Description and Preprocessing Steps:

The dataset is loaded using pandas, and initial exploratory data analysis (EDA) is conducted, including checking for missing values and class distribution. Key preprocessing steps include:

- Handling missing values using mean imputation.
- Encoding categorical variables using one-hot encoding.
- Defining features (X) and target variable (y).

Models Implemented and Rationale:

The notebook implements multiple machine learning models, including:

Logistic Regression: Chosen for its interpretability and baseline performance.

Random Forest Classifier: Selected for its ability to handle non-linearity and feature importance analysis.

SMOTE (Synthetic Minority Over-sampling Technique): Applied to address class imbalance.

Key Insights and Visualizations:

Visualizations using Matplotlib and Seaborn highlight class distribution and relationships between features. Some key findings include:
- The dataset exhibits class imbalance, requiring balancing techniques.
- Feature importance analysis helps in understanding key predictors of loan default.

Challenges Faced and Solutions:

Challenges identified include:
Class Imbalance: Addressed using SMOTE.
Missing Values: Handled through mean imputation.
Feature Encoding: One-hot encoding applied to categorical variables.