Credit Risk Analysis Report

1. Introduction

This report presents a comprehensive Exploratory Data Analysis (EDA) of credit applications. The goal is to uncover patterns, detect anomalies, and derive insights to improve loan approval strategies.

2. Data Preprocessing

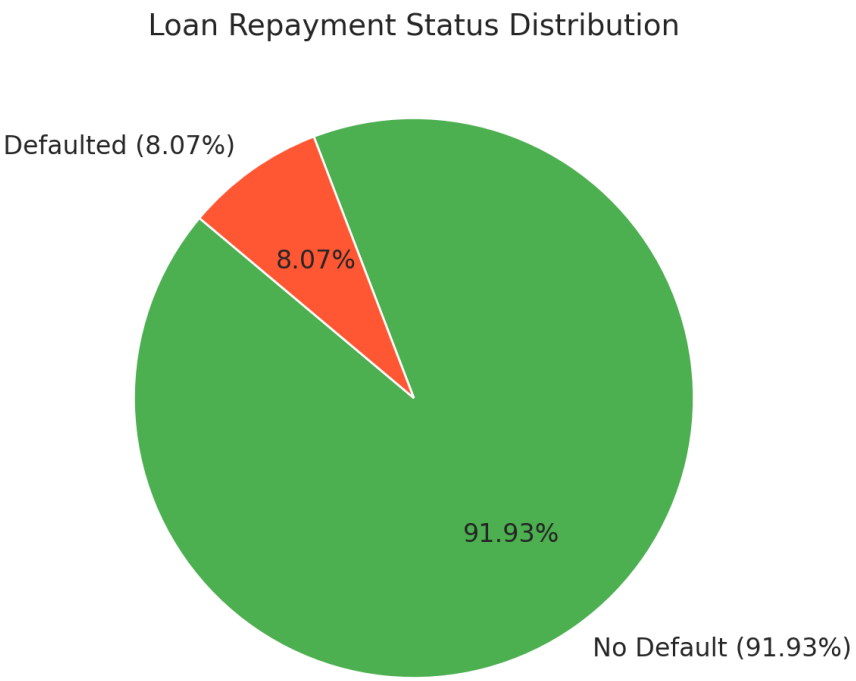
- **Handled missing values** separately for numerical and categorical columns.
- **Removed 41 columns** with more than 50% missing values to improve data quality.
- **Filled missing values** using Mode for categorical and Median for numerical columns.
- **Detected and treated outliers** while preserving valid high-income cases.

3. Target Column Analysis

The target variable represents loan repayment status:

- **0 (No Default - Loan Approved):** 91.93%
- **1 (Defaulted - Loan Not Repaid):** 8.07%

The distribution of loan repayment statuses was analyzed using bar plots and pie charts.



4. Exploratory Data Analysis (EDA)

EDA was conducted to explore trends in the dataset using various statistical and graphical techniques. The analysis is structured into Univariate, Bivariate, and Multivariate sections for clarity.

4.1 Univariate Analysis

Univariate Analysis examines individual features to understand their distribution. Histograms and boxplots were used to analyze credit amount, income levels, and applicant demographics.

4.2 Bivariate Analysis

Bivariate Analysis explores relationships between two variables. Violin plots were used to compare income across loan statuses, while scatter plots helped analyze the correlation between loan amount and annual income.

4.3 Multivariate Analysis

Multivariate Analysis examines interactions among multiple variables. Pair plots and KDE plots were used to analyze credit amount, repayment history, and applicant profiles.

5. Key Insights from EDA

- **No Default Rate (Loan Approved):**
  91.93% of applicants successfully repaid their loans
- **Default Rate (Loan Not Repaid):**
  8.07% of applicants defaulted on their loans
- **Loan Type Distribution:**
  90.5% Cash Loans, 9.5% Revolving Loans
- **Gender-Based Trends:**
  66% Female, 34% Male
- **Property Ownership:**
  69% Own Real Estate, 31% Do Not
- **Income Groups:**
  51% in 'Working' category
- **Education Levels:**
  71% have Secondary Education, 24% have Higher Education

6. Handling Outliers

Outliers were detected in numerical columns using box plots. However, they were not removed, as some extreme values represented valid high-income individuals. This approach ensures that financial patterns remain intact.

7. Merging Previous Application Data

To gain deeper insights, the dataset was merged with previous loan application records. This allowed tracking of repayment history and identifying trends in customer borrowing behavior.

8. Conclusion

The analysis provided valuable insights into customer behavior and credit risk assessment. These insights can be leveraged to refine credit policies and improve risk management strategies.