Loan Default Prediction Case Study

Introduction

This project aims to help a consumer finance company specializing in various types of loans to urban customers. The goal is to identify patterns indicating whether a person is likely to default on a loan using Exploratory Data Analysis (EDA).

Business Understanding

The company must decide whether to approve loan applications based on the applicant's profile, balancing the risk of financial loss from defaults against the potential loss of business from overly conservative lending practices.

Objectives

Identify the driving factors behind loan defaults.
Use EDA to understand how consumer attributes and loan attributes influence the tendency of default.
Provide actionable insights and recommendations.

Data Understanding

The dataset includes information about past loan applicants, such as loan amounts, funded amounts, interest rates, annual income, debt-to-income ratios, and loan statuses (Fully Paid, Current, Charged-off).

Data Cleaning and Manipulation

Missing values were handled by removing columns with over 20% missing data and imputing the rest with mode values.
Percentage columns were converted to numerical values.
Outliers were removed using z-score analysis.
Log transformation was applied to skewed numerical columns.

Exploratory Data Analysis (EDA)

Univariate Analysis

Analyzed the distribution of loan amounts, interest rates, annual incomes, etc.
Examined the frequency counts of categorical variables like loan status.
Bivariate Analysis

Used box plots and violin plots to compare numerical features against loan status.

Created scatter plots to study the relationship between numerical features and loan amounts.

Multivariate Analysis

Created a correlation matrix to understand relationships between numerical variables.
Generated pair plots for detailed analysis of interactions between multiple variables.

Key Findings

Higher interest rates are associated with higher chances of loan defaults.
Lower annual incomes correlate with increased default rates.
The debt-to-income ratio is a significant predictor of defaults.

Recommendations

Implement stricter credit checks for applicants with high DTI ratios.
Adjust interest rates based on the applicant's risk profile.
Develop targeted financial products for lower-income applicants to mitigate risk.

Insights and Summary
Key Findings

Loan Amount: Higher loan amounts tend to be associated with higher default rates.
Annual Income: Lower annual income is correlated with a higher likelihood of default.
Term: Loans with longer terms are more likely to default.
Employment Length: Shorter employment lengths are associated with higher default rates.
Credit History: Applicants with poor credit history (e.g., higher delinquencies, higher revolving utilization) are more likely to default.

Conclusion

By identifying risky loan applicants, the company can reduce financial losses and improve its lending strategies, ultimately enhancing business outcomes.

Author
Nandakumar Devadoss and Neeraj Singh