Loan Default Risk Analysis

...

Overview

Harmony Consumer Finance Company specializes in lending various types of loans to urban customers.

When a loan application is received, the company must make a decision to approve or reject the loan based on the applicant's profile.

The company wants to understand the driving factors behind loan default.

Understanding the risks

Risk 1

If the applicant is likely to repay the loan, rejecting the loan results in a loss of business

Risk 2

If the applicant is not likely to repay the loan, i.e., default, approving the loan may lead to a financial loss

Objective:

Identify patterns and factors that indicate if a loan applicant is likely to default.

Use Exploratory Data Analysis (EDA) techniques to understand how consumer attributes and loan attributes influence the tendency of default.

Provide insights and recommendations to help the company make informed decisions and minimize financial losses.

Data

- The dataset contains information about past loan applicants
- The data includes various attributes such as demographics, credit history, loan characteristics, and repayment status
- The target variable is the loan status, which indicates whether the loan was fully paid, current, or charged-off (defaulted)

Analysis

Analysis Approach

- Data understanding
- Univariate analysis
- Bivariate Analysis
- Multivariate Analysis
- Visualization

Univariate analysis

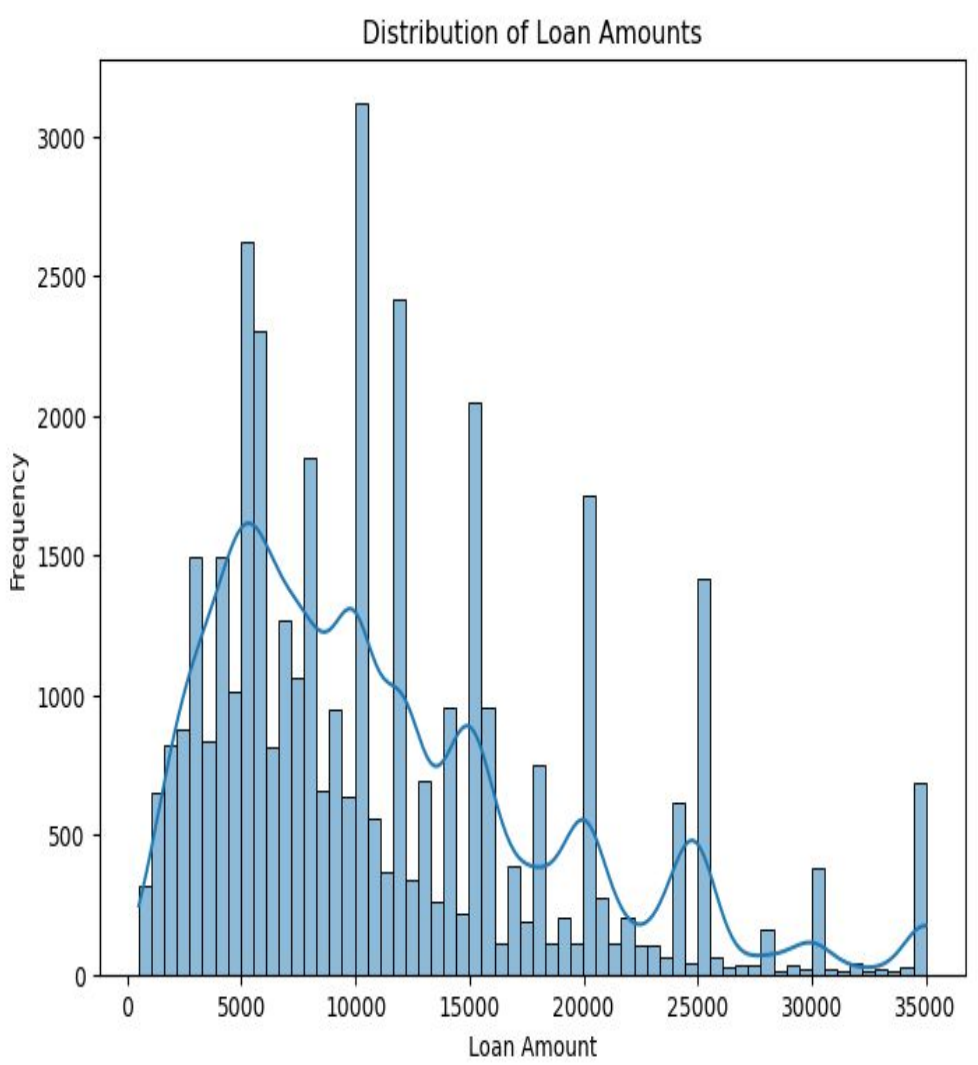
Loan Status Distribution:

- Majority of loans are fully paid (82.9%)
- Smaller proportion of loans are charged off (14.1%)

Loan Amount Distribution:

- Most loans fall in the range of $5,000 to $15,000
- Distribution is right-skewed

Visualization: Histogram of Loan Amounts



Bivariate analysis

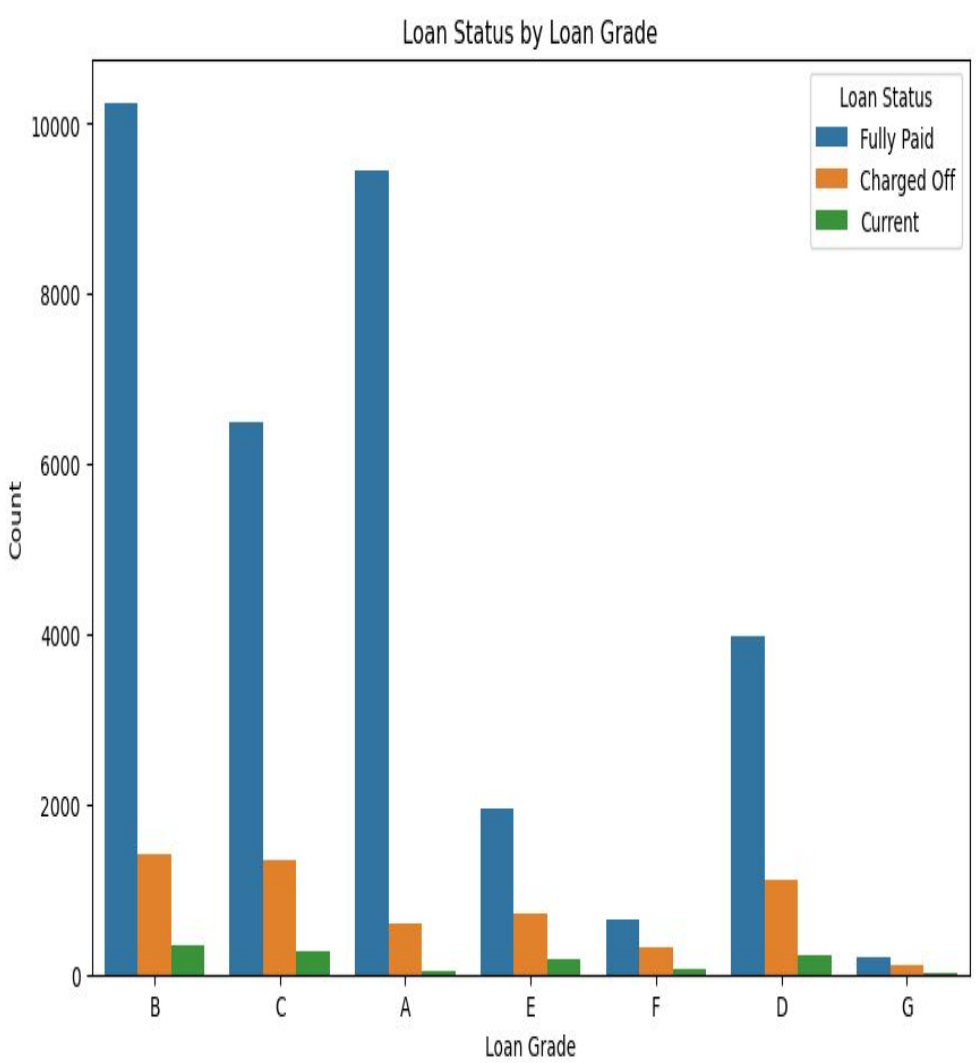
Loan Grade vs. Loan Status:

- Higher loan grades (A and B) have a lower proportion of defaults
- Lower loan grades (C, D, and E) have a higher proportion of defaults

Employment Length vs. Loan Status:

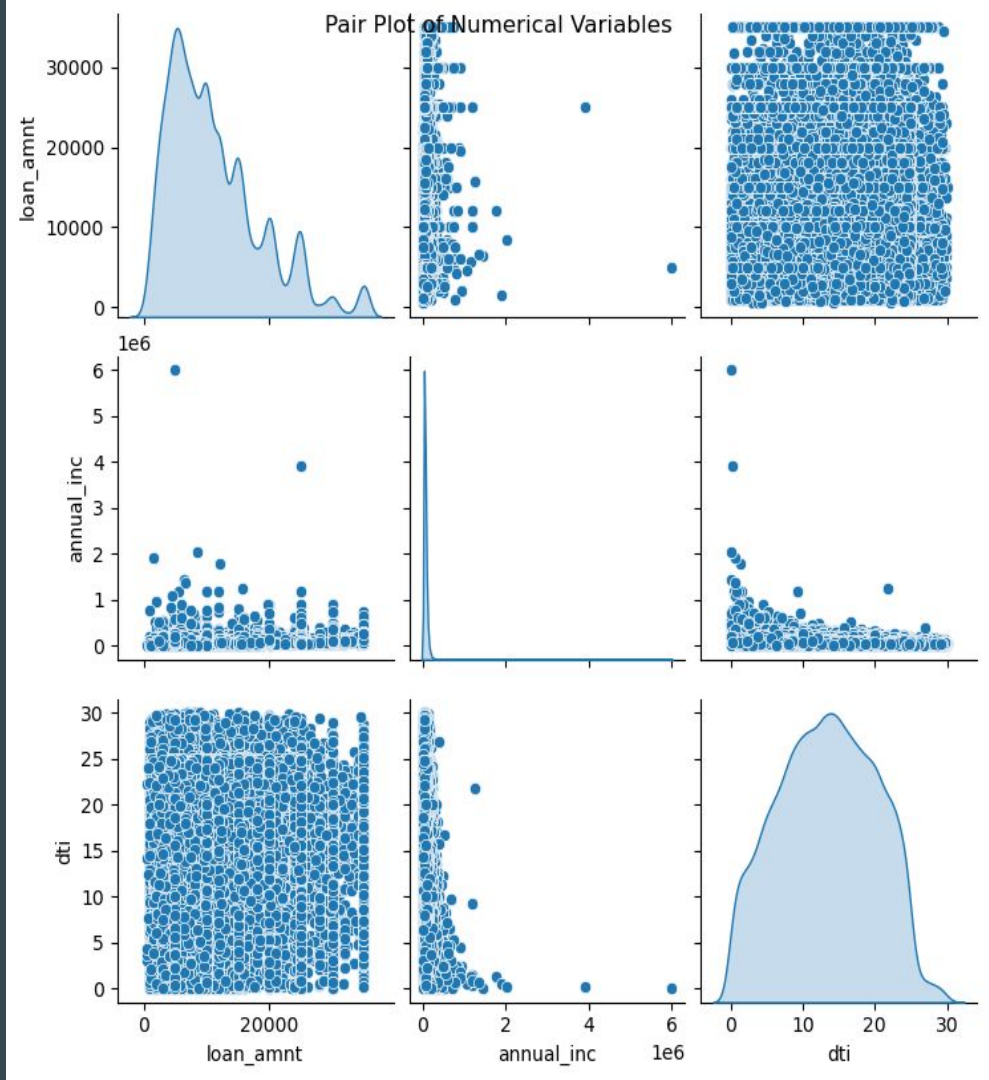
- Borrowers with longer employment lengths (10+ years) have a lower default rate
- Shorter employment lengths have a higher default rate

Visualization: Count Plots of Loan Grade and Employment Length vs. Loan Status



Multivariate analysis

- loan_amnt: The distribution of loan amounts, which appears to be right-skewed.
- annual_inc: The distribution of annual income, which is heavily right-skewed.
- dti: The distribution of debt-to-income ratios, which appears to be right-skewed as well.



Key insights & recommendation

Interest Rates:

- Loans with interest rates above 15% have a significantly higher default rate (33.3%) compared to the overall default rate (14.2%)
- Recommendation: Implement stricter lending criteria for high-interest loans

Employment Stability:

- Borrowers with longer employment lengths (10+ years) have a lower default rate
- Recommendation: Consider employment stability as a key factor in loan approval decisions

Visualization: Pair Plot of Numerical Variables (Loan Amount, Interest Rate, Annual Income, Debt-to-Income Ratio)

The Team

Jagruti Shah

Data Analysist

Kishor

Data Analysist