Lending Club Case Study: Predicting Loan Default

This presentation explores a Lending Club case study, analyzing loan data to predict default risk. We'll delve into data cleaning, analysis, and key insights that can help lenders make informed decisions.

by Raghavendra Tharun

Understanding the Data

Loan Types

Lending Club offers various loan types to urban customers, including personal, auto, and business loans.

Decision Process

The company evaluates loan applications, considering factors like credit history, income, and debt-to-income ratio.

Data Preparation

1

Data Cleaning

Removing irrelevant or incomplete data points to ensure data accuracy and reliability.

2

Outlier Treatment

Identifying and addressing extreme values that could skew analysis results.

3

Data Imputation

Filling in missing values using appropriate methods to maintain data integrity.

Univariate and Bivariate Analysis

Univariate Analysis

Examining individual variables to understand their distribution and characteristics.

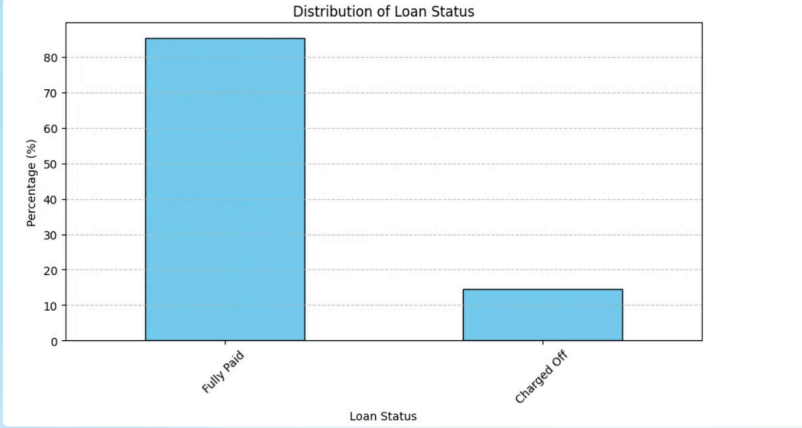
Bivariate Analysis

Exploring relationships between pairs of variables to identify potential correlations and patterns.

Percentage distribution of loan_status between charged off and Fully paid

Loan_Status

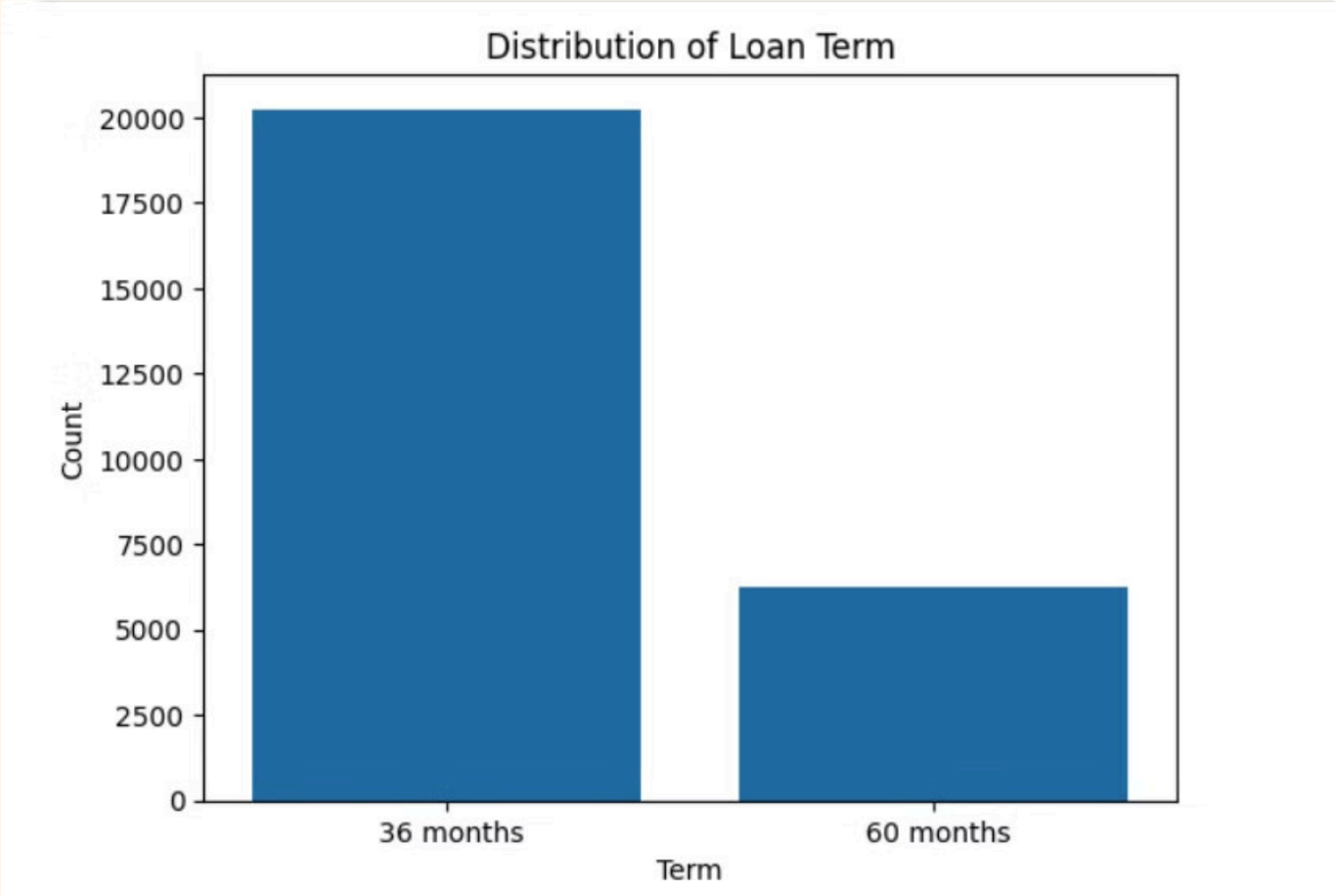
Fully Paid	85.41
Charged Off	14.59



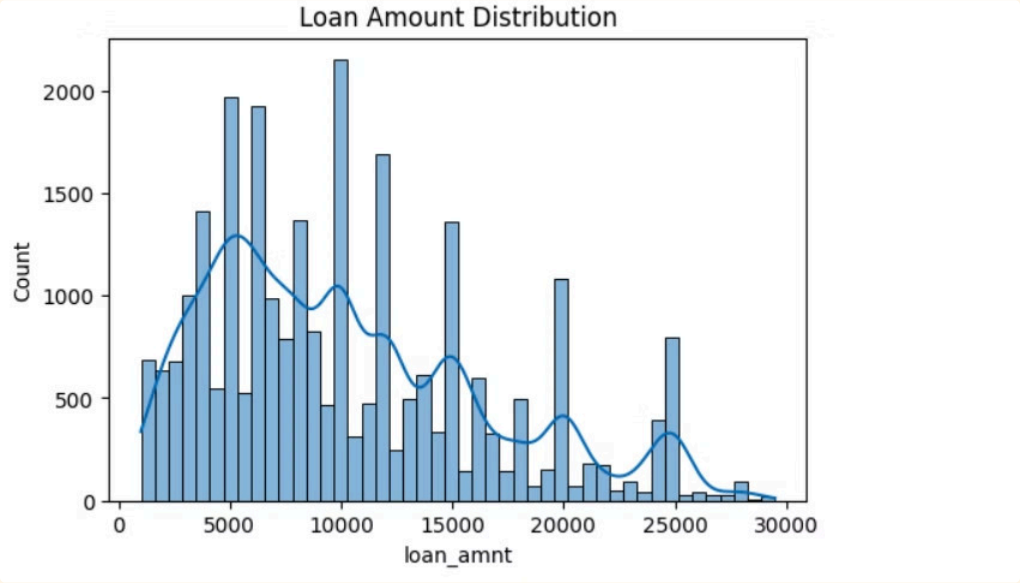
The term most borrowers prefer (or have opted for).

Insights:

Most term loan have taken in 36 months than 60 months term.Additionally having more risky on 60 term plan.



Distribution of Loan Amounts



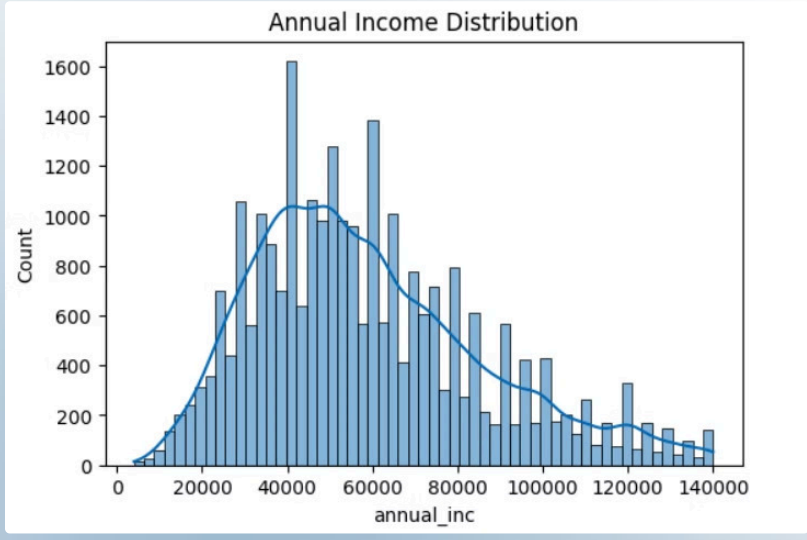
Insights:

- The majority of loans are in the $5,000 to $10,000 range, with over 2,000 loans at $10,000.

Borrower Income Distribution

Insights:

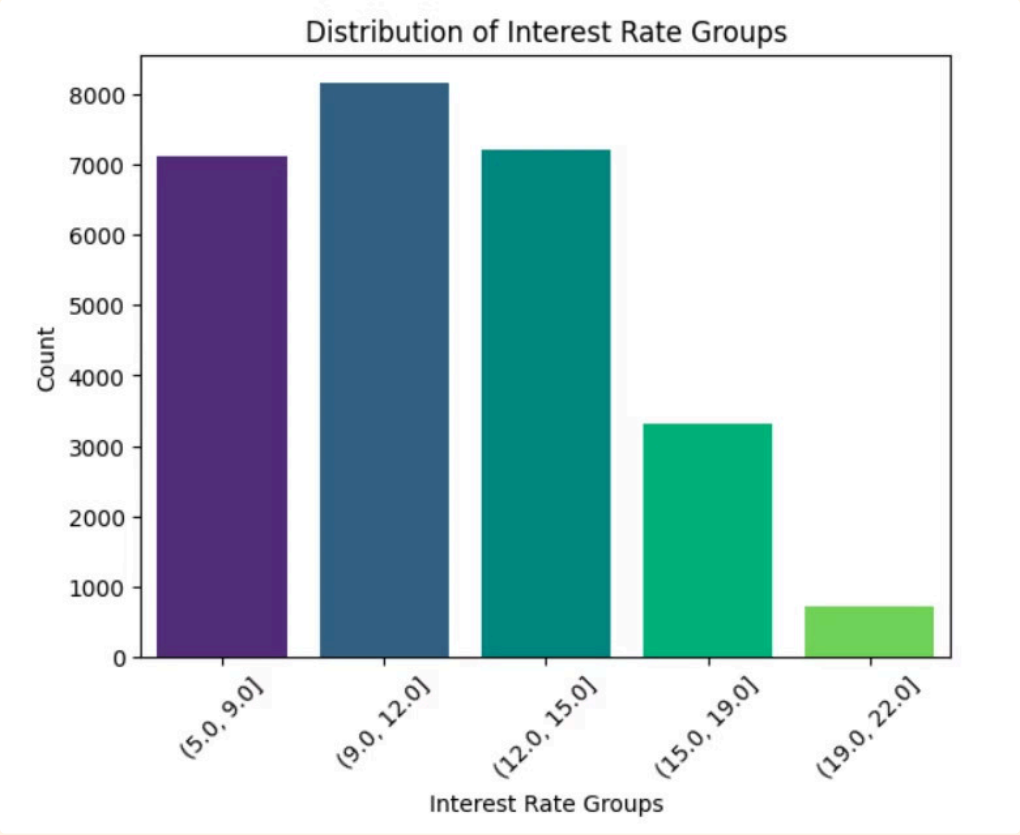
Most annual income relying on the 40000 and 60000 with the count of 1600 from the visualization



Explore Interest Rate Groups in Interactive Bins

Insights:

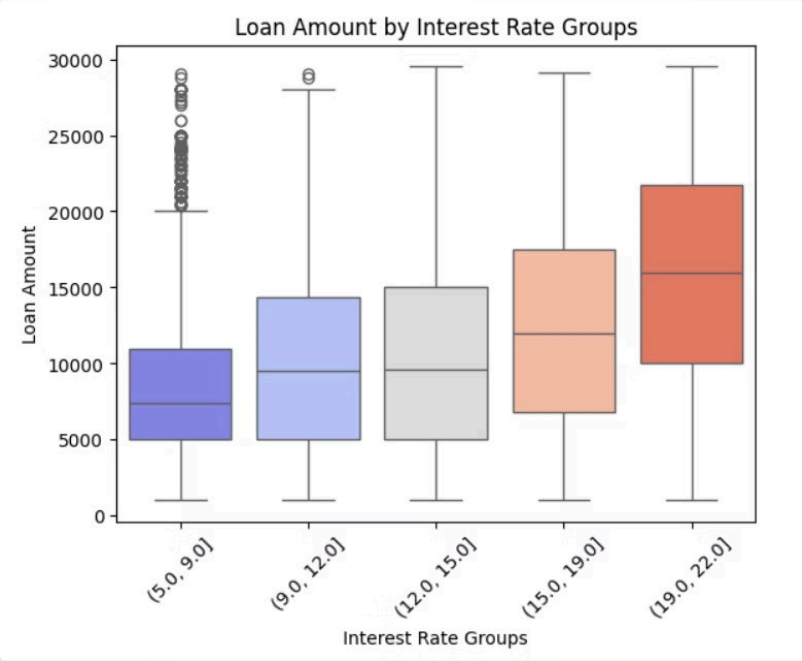
1. Gain Valuable Insights into Interest Rate Distribution Across the Dataset with this Plot
2. The majority of interest rates fall within the range of (9.0, 12.0].

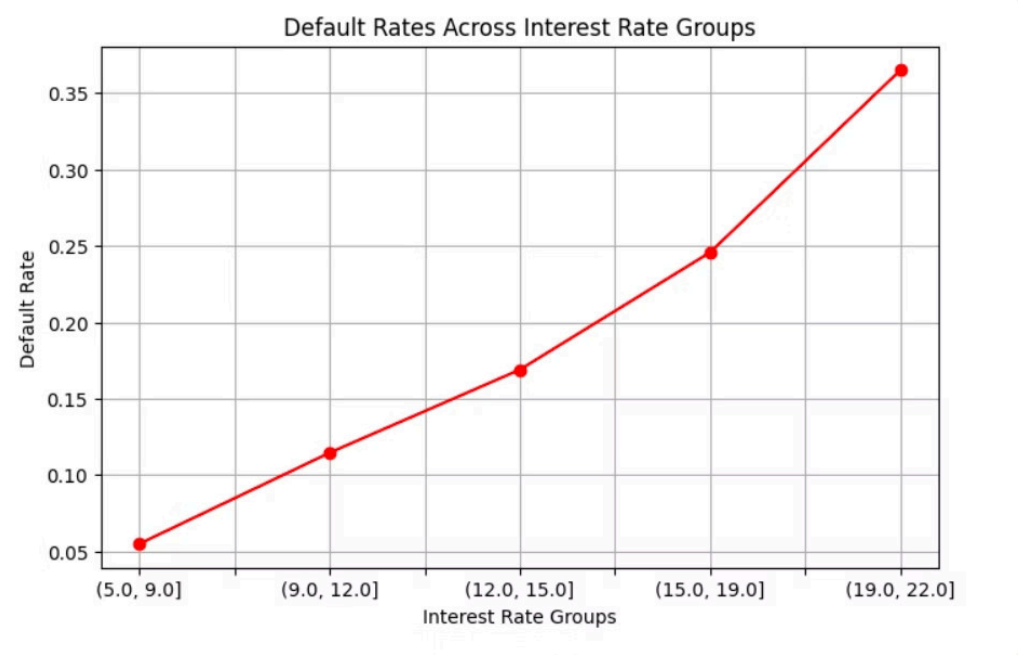


Identify loan amount outliers by interest rate

Insights:

- Discover additional outliers beyond $20,000 loan amounts within the (5.0, 9.0] interest rate range.





Analyze default rates by interest rate group

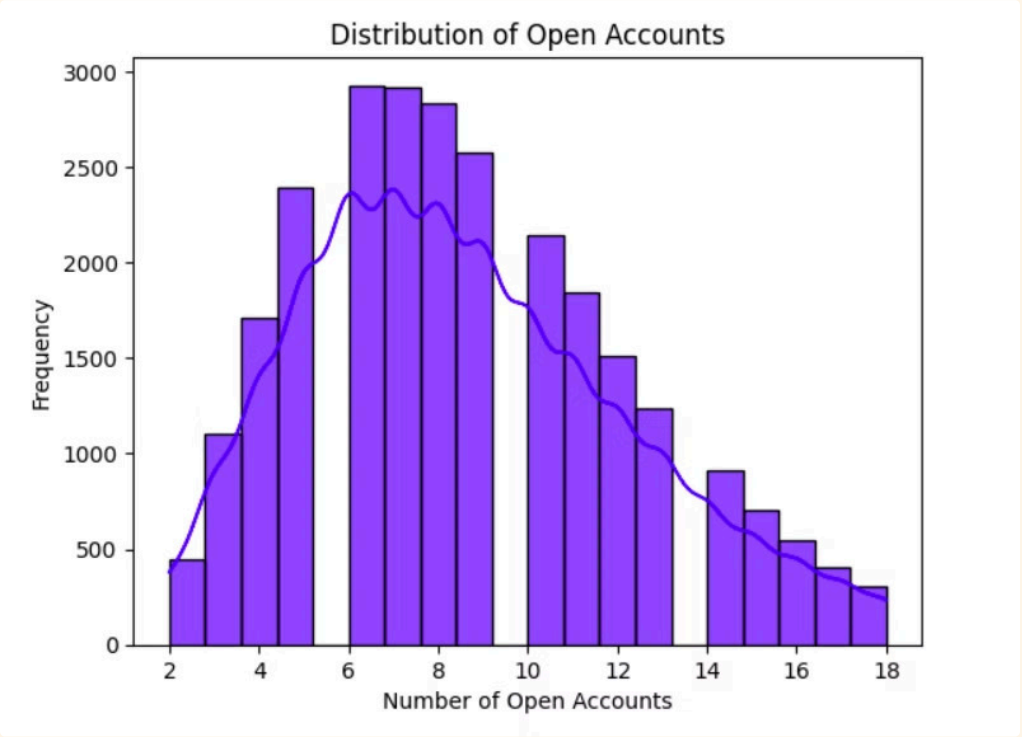
Insights:

- Dive into financial risk assessment with this essential visualization, spotlighting the interest rate groups (19.0,22.0] at the highest risk of default.

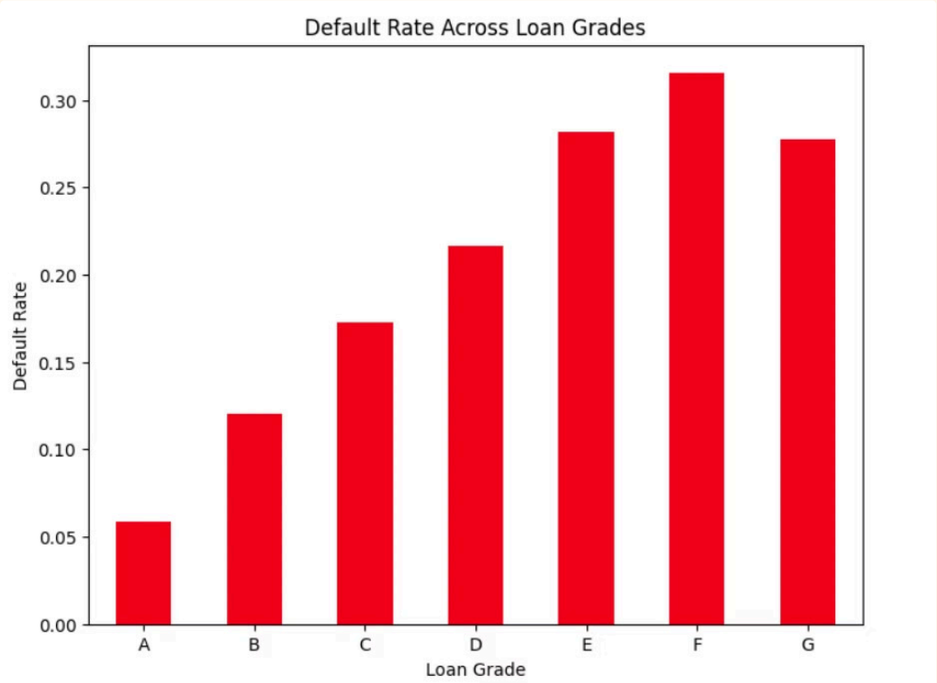
Open Accounts Distribution

Insights:

- This plot reveals if most borrowers have a small or large number of open accounts. It can help identify trends like borrowers with many accounts being more or less risky.
- Here more open accounts are relying on 6 to 10 that's more risky



Calculate default rate for each grade



Insights:

- **High-Risk Grades**:

  If the bars for 'E', 'F', or 'G' are significantly higher, these grades represent high-risk borrowers.

- **Low-Risk Grades**:

  If 'A' and 'B' grades have minimal default rates, these are safer borrower categories, suitable for low-risk lending strategies.

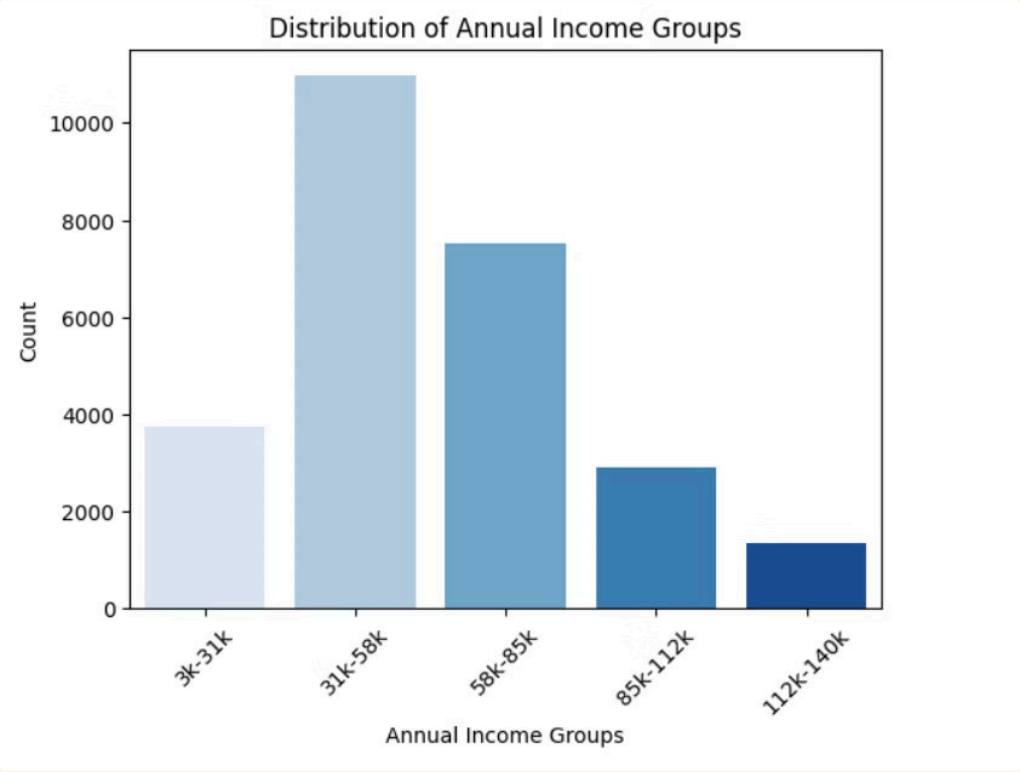
- The plot shows that grades E, F, and G have higher default rates, indicating greater risk in lending credits to these groups.

- **Actionable Insights**:
  - **For Risk Reduction**: Limit loans issued to 'F' and 'G' grades or impose higher interest rates.
  - **For Balanced Portfolio**: Increase lending to 'A', 'B', and 'C' grades while cautiously managing 'D' grades.

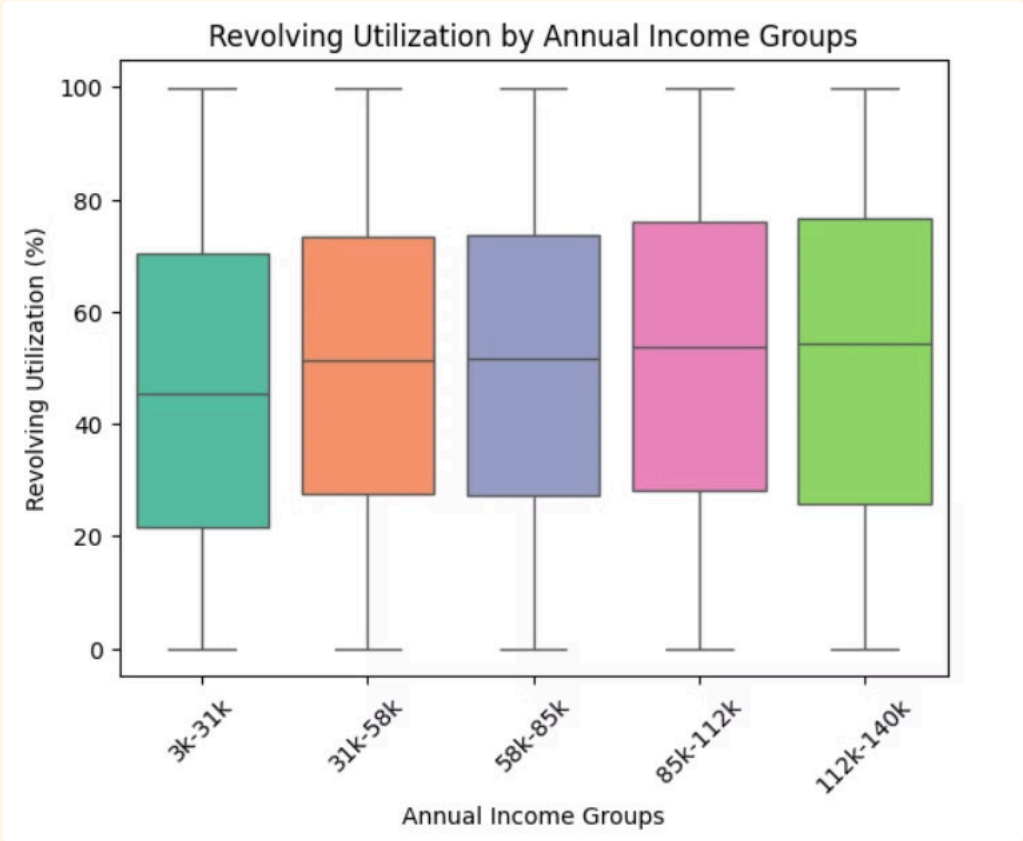
To visualize how borrowers are distributed across income groups.

Insights:

1. This helps assess the income levels of the majority of borrowers and whether there's a concentration of high or low-income earners.
2. The plot indicates higher risk for individuals with an income level between 31k and 58k.

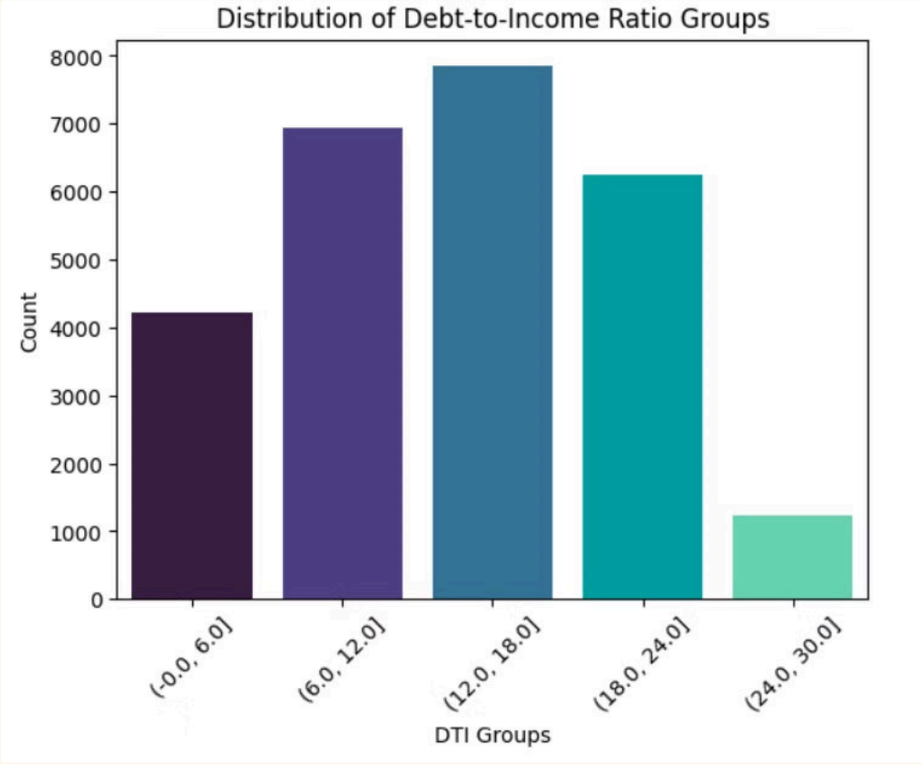


To compare the levels of revolving credit utilisation among different income groups.



Insights:

- Group with highest median revolving credit utilization: 112k-140k
- Group with highest mean revolving credit utilization: 85k-112k



To examine how borrowers are distributed based on their debt-to-income ratio (DTI)

Insights:

- Percentage distribution of borrowers across DTI groups:
  dti_groups

  (-0.0, 6.0] - 15.95

  (6.0, 12.0] - 26.22

  (12.0, 18.0] - 29.61

  (18.0, 24.0] - 23.55

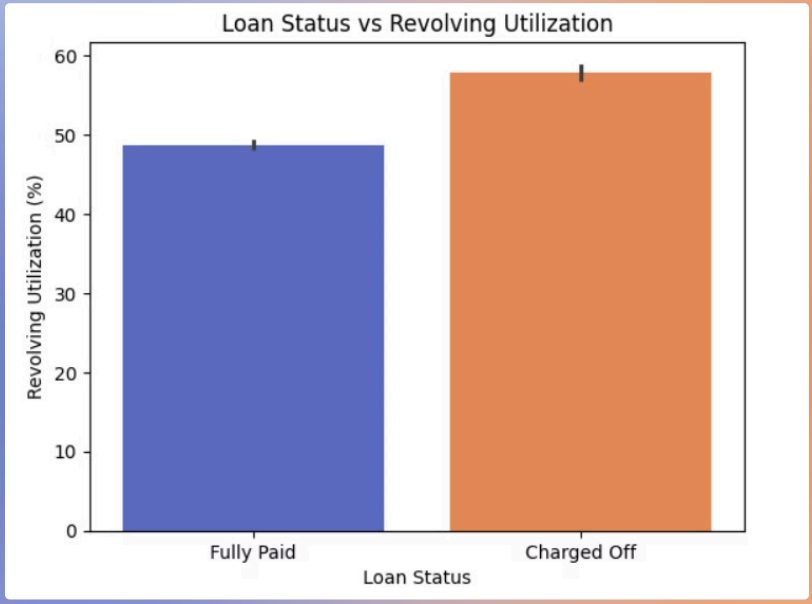
  (24.0, 30.0] - 4.68

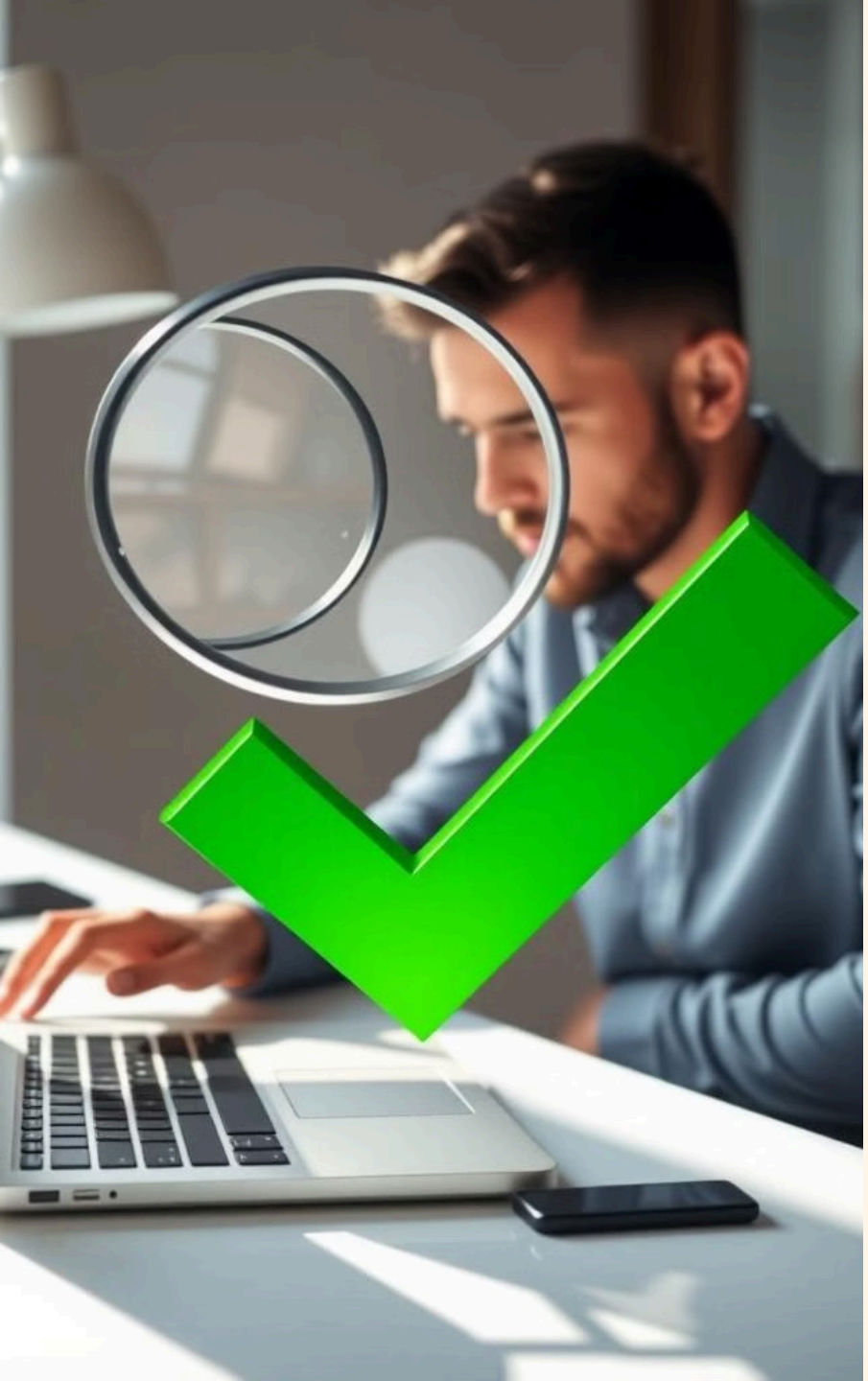
- The most common DTI group is: (12.0, 18.0], with 29.61% of borrowers.

Comparing average revolving utilization for each loan status.

Insights:

- Average Revolving Utilization by Loan Status:
  Fully Paid 48.72
  Charged Off 57.87
- Difference in revolving utilization between Charged Off and Fully Paid loans: 9.14%





Recommendations for Loan Approval

High-Risk Factors

- Loan amount over $15,000
- Loan term of 60 months
- Lower loan grades (D-G)
- Annual income below $58,000
- Debt consolidation, other, or small business loan purposes
- Debt-to-income ratio above 12%
- Multiple credit inquiries in the past 6 months
- Previous bankruptcy
- Revolving line utilization rate over 60%
- 10+ years of employment