Loan Default Analysis

UPGRAD EDA ASSIGNMENT

BHUMIKA SHARMA

Problem Statement

Identify the factors that influence loan defaults in a consumer finance company. By analyzing historical loan data, the goal is to recognize patterns that can predict whether a loan applicant is likely to default, helping the company to minimize financial risks while optimizing business growth.

Analysis Approach

1. **Data Cleaning and Preprocessing**: Begin by handling missing values, encoding categorical variables, and scaling numerical features to prepare the data for analysis.

2. **Exploratory Data Analysis (EDA)**:

   1. **Univariate Analysis**: Examine the distribution of individual variables to understand the general characteristics of loan applicants.

   2. **Bivariate Analysis**: Explore relationships between different variables, particularly focusing on how they relate to the target variable (loan default).

   3. **Outlier Detection and Removal**: Identify and address any anomalies in the data that could distort the analysis.

3. **Feature Selection**: Determine the most significant variables that are strong indicators of default through EDA and statistical methods.

4. **Insights and Recommendations**: Use the findings from EDA and modeling to provide actionable insights, such as identifying high-risk applicants or adjusting loan offerings to mitigate potential losses.
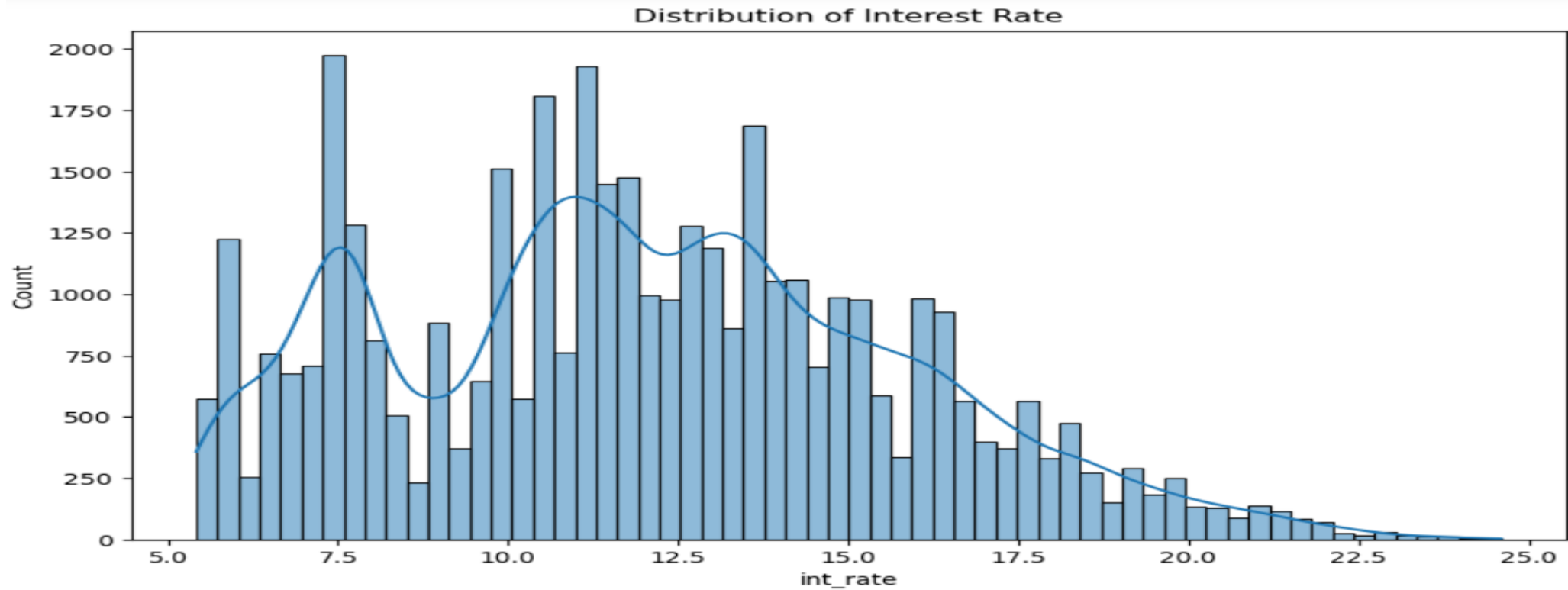
# Insight 1: Most loans are either fully paid or charged off, with fewer in the "current" status.
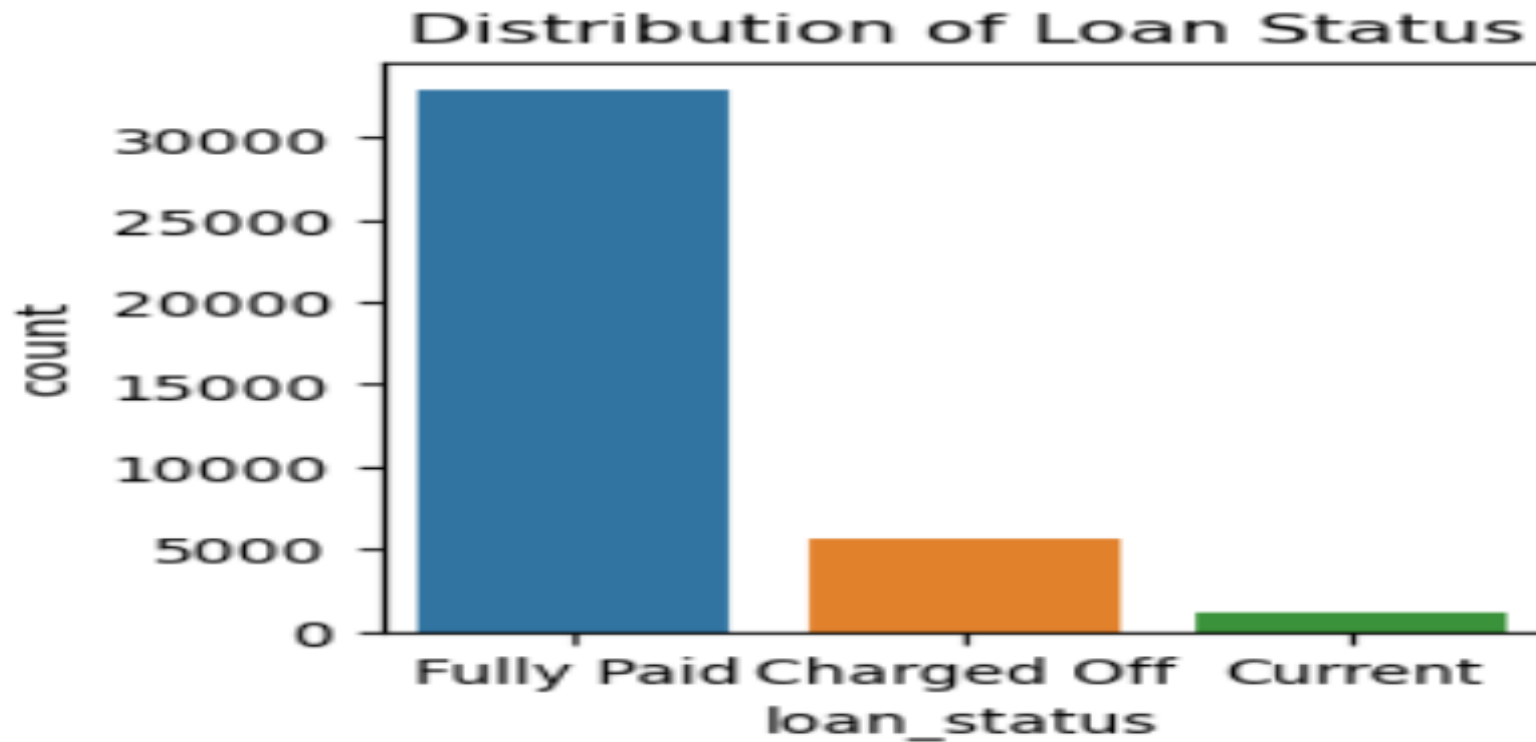
# # Insight 2: Most loans are in the range of $5,000 to $15,000.

Distribution of Loan Amount
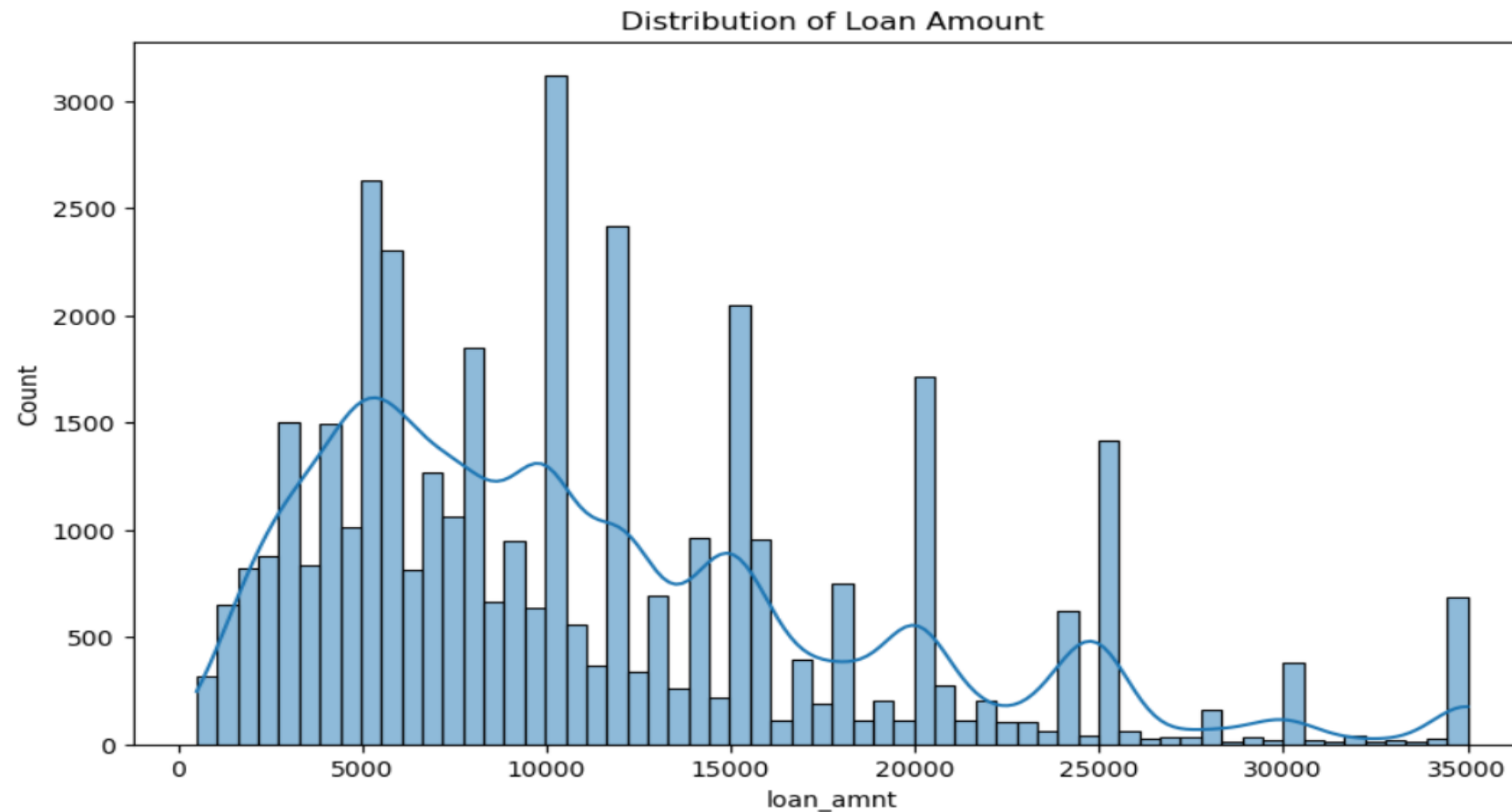
Insight 3: Interest rates are mostly between 10% and 20%.

Distribution of Interest Rate

Insight 4: Charged-off loans tend to have slightly higher loan amounts compared to fully paid loans.
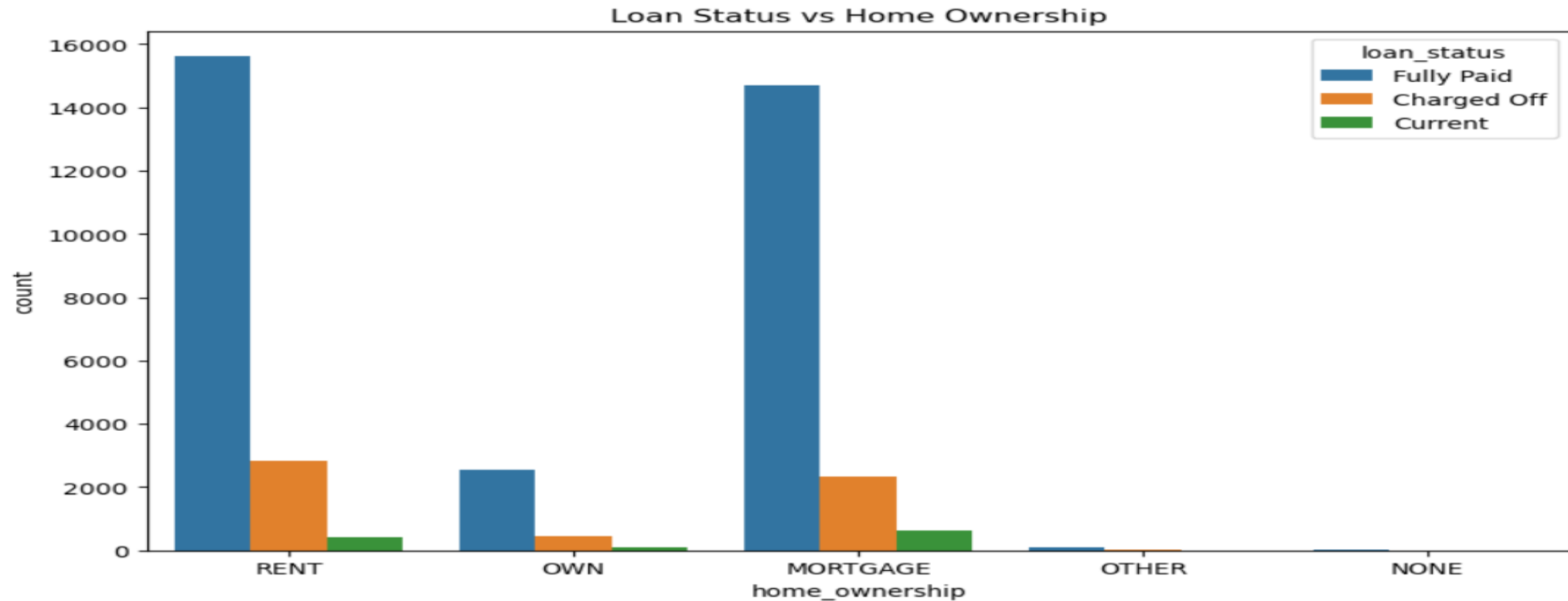

Loan Status vs Loan Amount
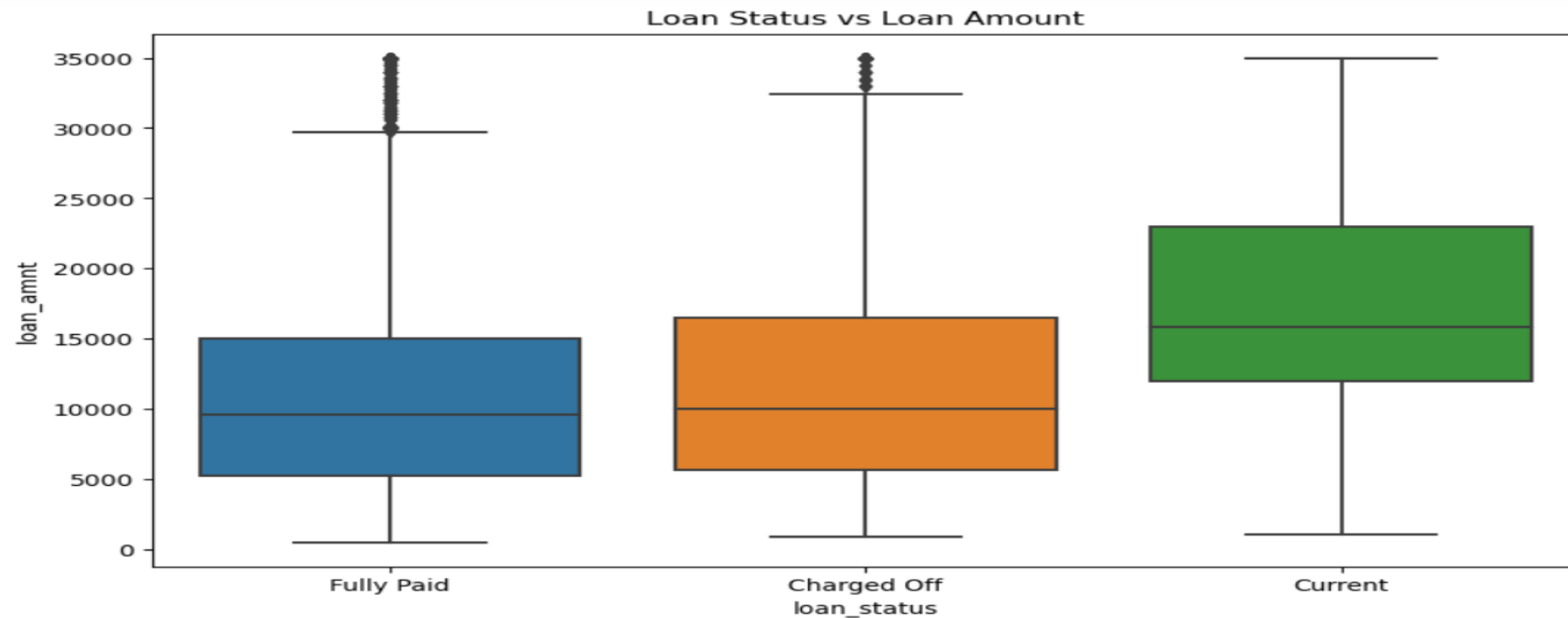
Insight 5: Charged-off loans have significantly higher interest rates compared to fully paid loans.
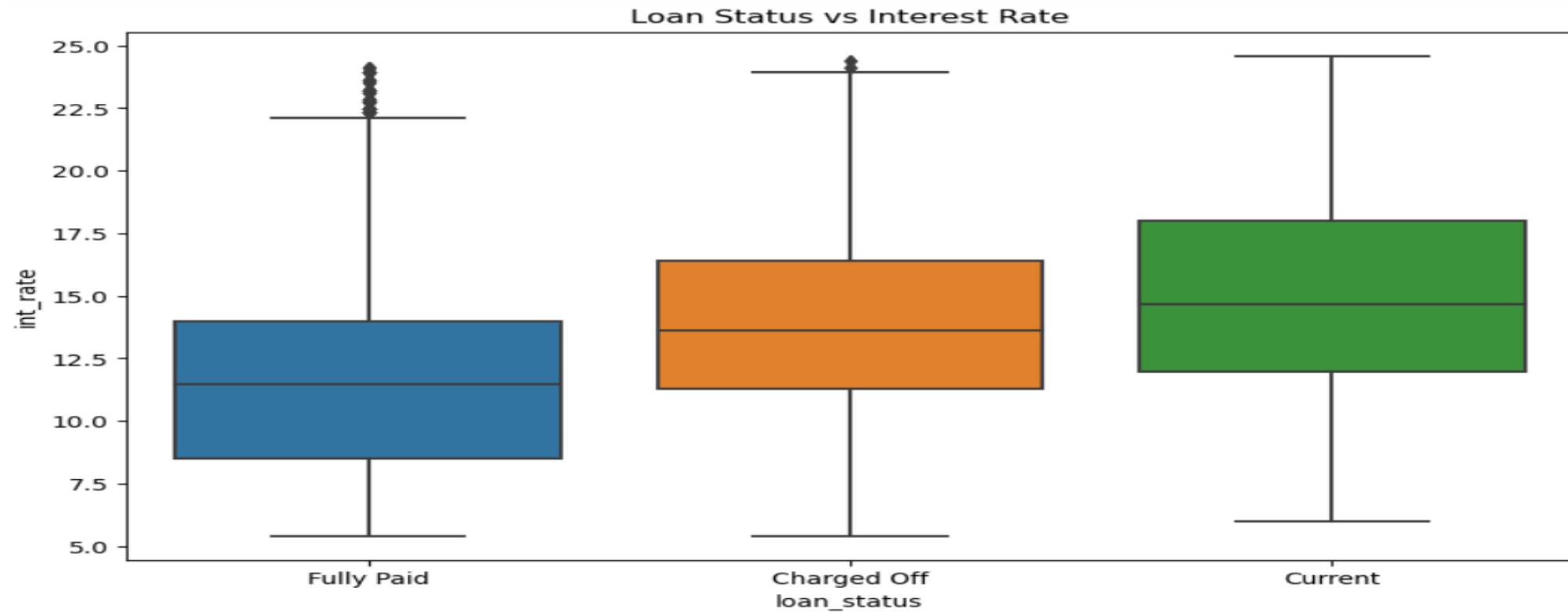
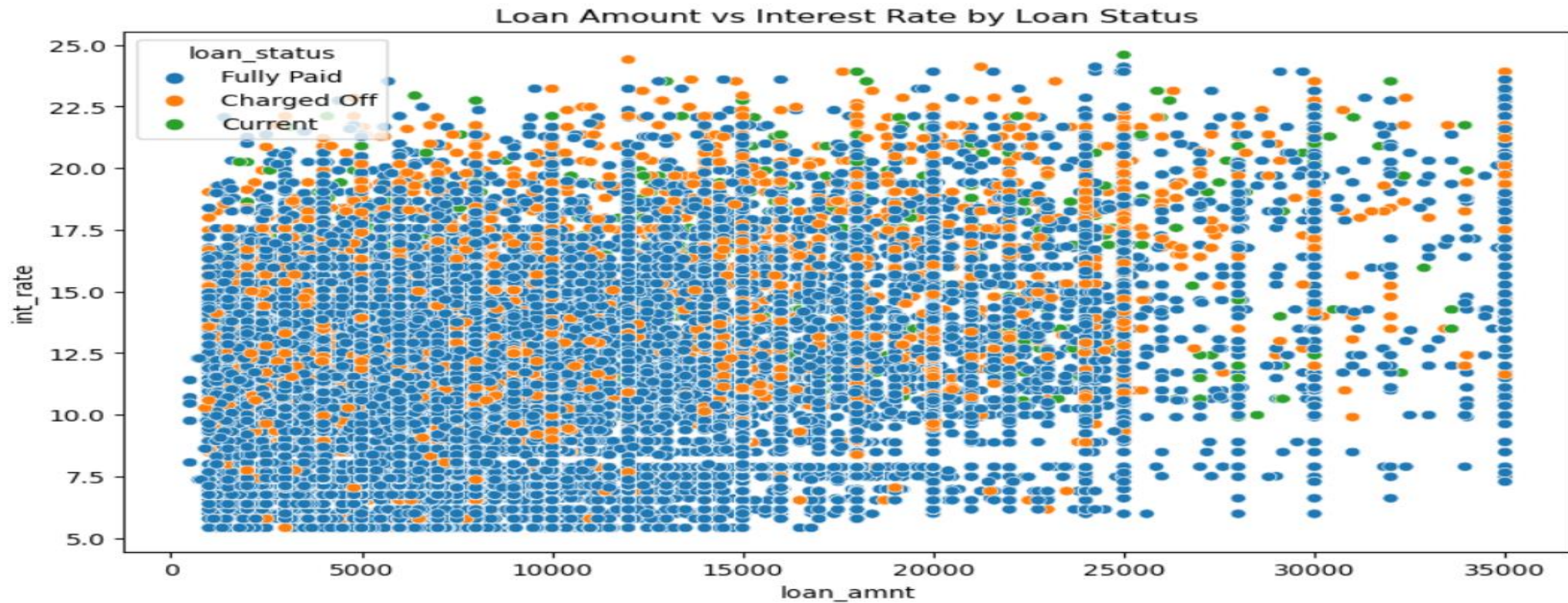# Insight 6: Borrowers with shorter employment lengths are more likely to default.

Loan Status vs Employment Length
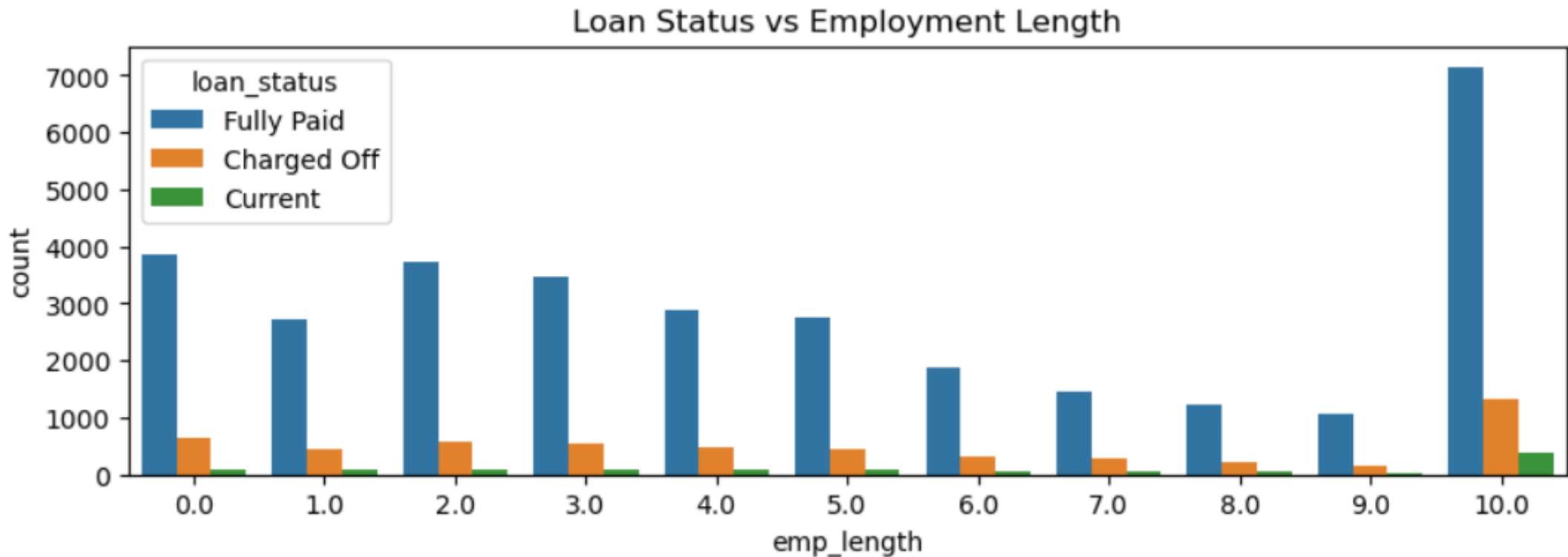
Insight 7: Borrowers who rent are more likely to default compared to those who own a home.

Insight 8: Charged-off loans tend to have both higher loan amounts and interest rates.
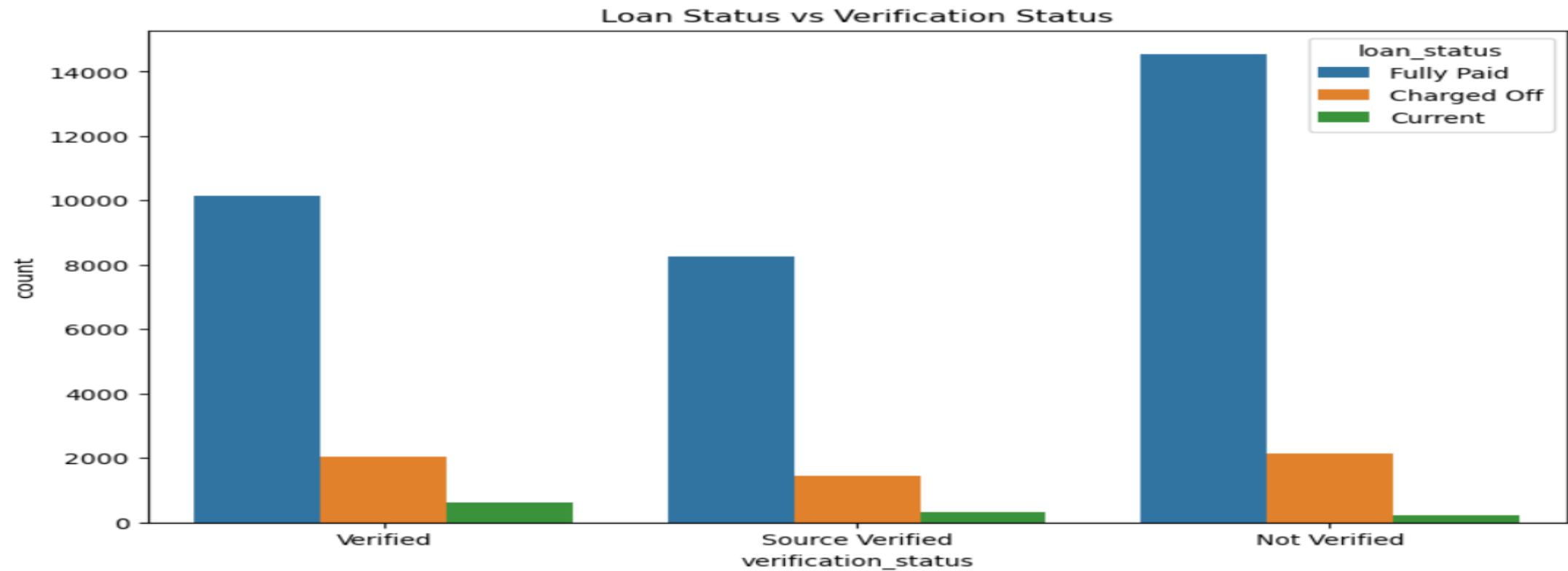

Loan Amount vs Interest Rate by Loan Status

Insight 9: Loans taken for small businesses and debt consolidation have higher default rates.
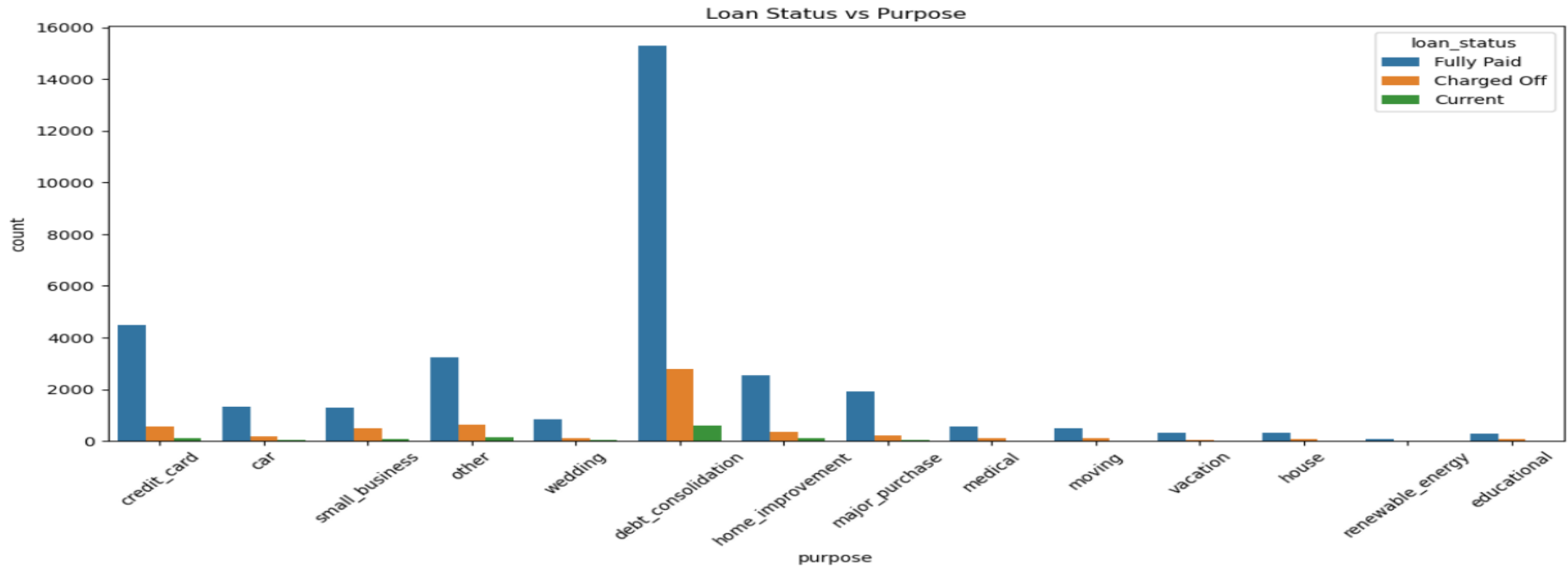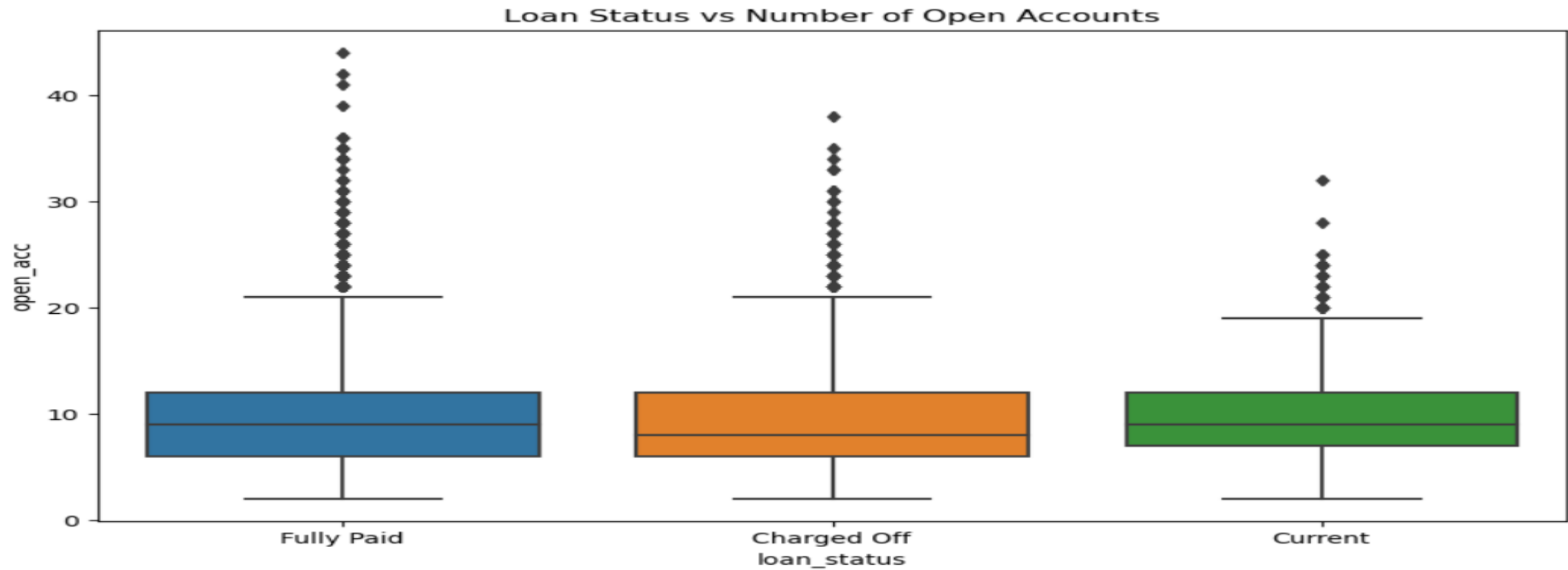
# Insight 10: Lower grades (e.g., D, E, F) are associated with higher default rates.
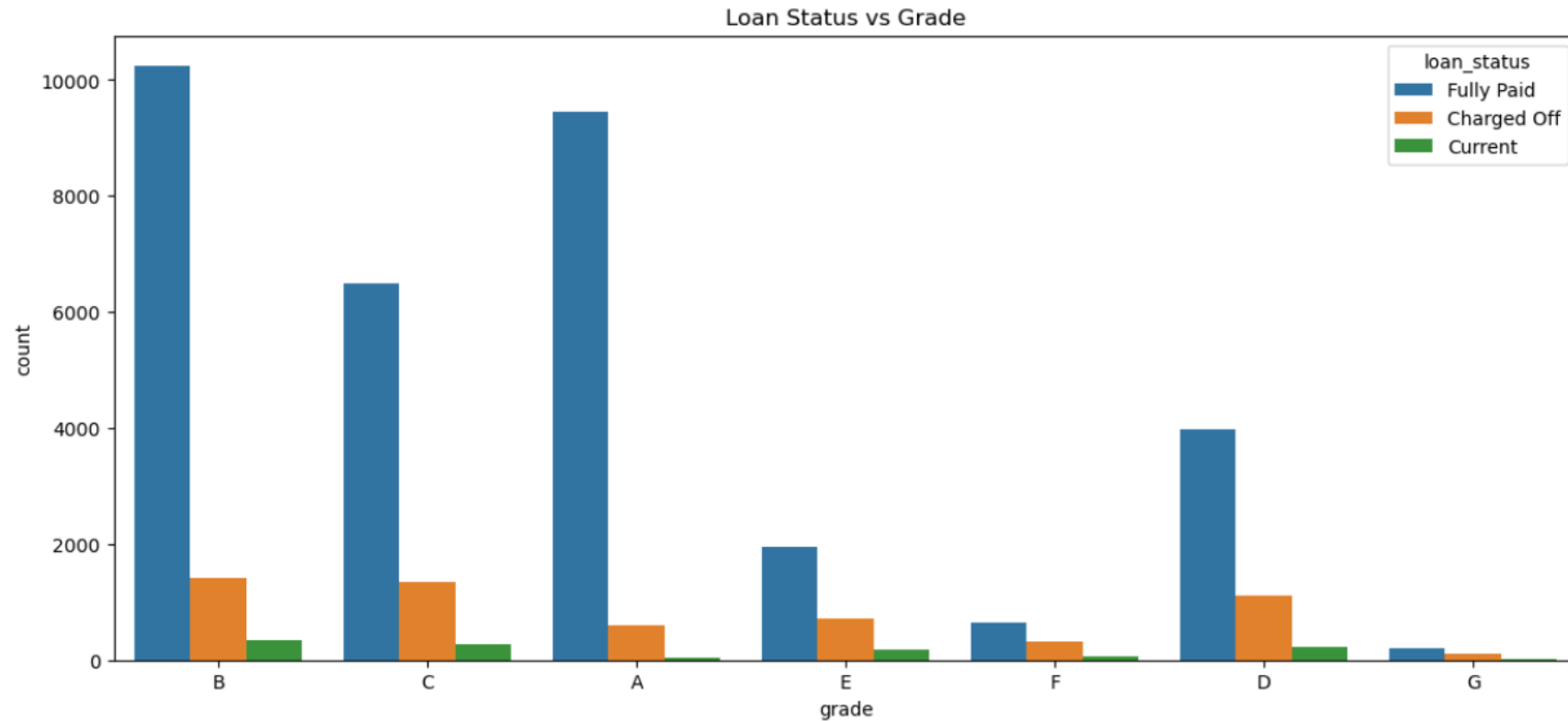
Insight 11: Borrowers whose income is not verified have a higher likelihood of default.

Insight 12: Borrowers with fewer open accounts tend to default more.

Loan Status vs Number of Open Accounts

Insight 13: Shorter employment lengths coupled with higher loan amounts are strong indicators of default.
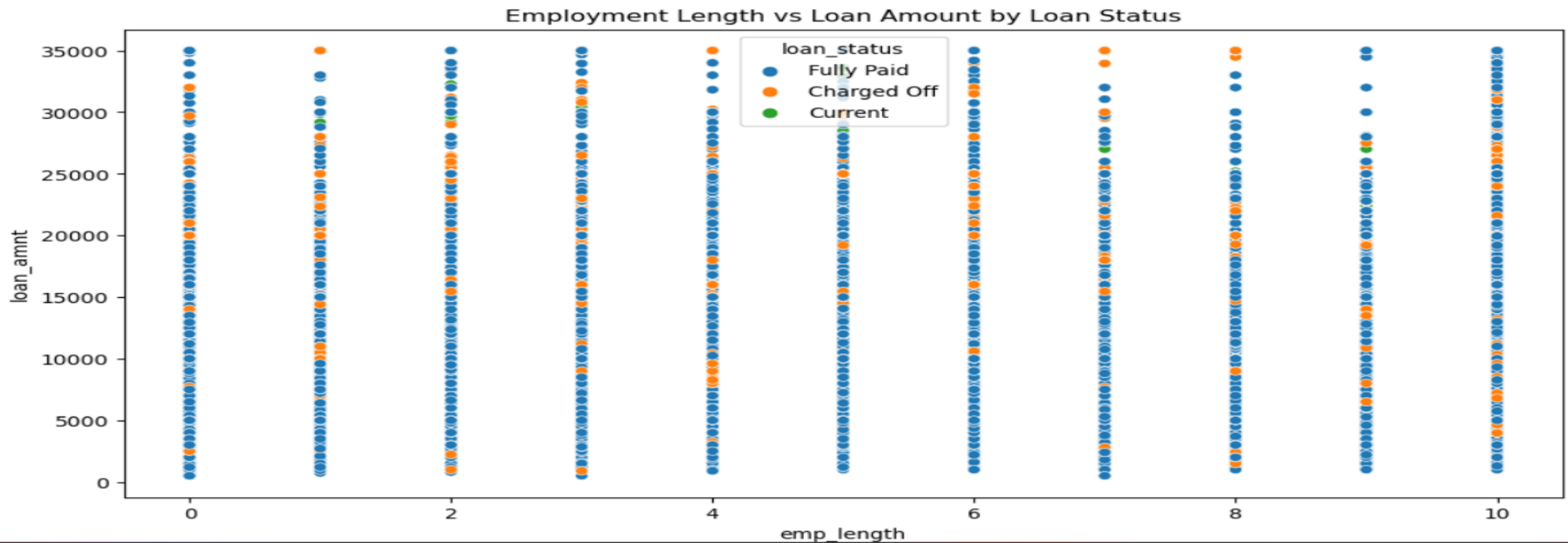
Recommendations to Avoid Loan Defaults

Risk-Based Loan Pricing: Adjust loan amounts and interest rates based on the borrower's risk profile, including income, employment length, and debt-to-income ratio.

Strengthen Credit Assessments: Prioritize borrowers with longer credit histories, higher incomes, and lower debt-to-income ratios. Implement stricter lending criteria for those with poor credit grades, unverified incomes, or public records.

Employment Stability Checks: Give preference to borrowers with longer, stable employment histories, and consider additional verification for those with shorter employment lengths.

Targeted Loan Products: Be cautious with loans for high-risk purposes like small businesses and debt consolidation, and consider offering these with stricter terms or higher interest rates.

Housing Status Consideration: Consider the borrower's housing status in risk assessments, as renters tend to have higher default rates compared to homeowners.

Enhanced Income Verification: Implement robust income verification processes to reduce the likelihood of defaults from borrowers with unverified or lower incomes.