Lending Club Case Study

Submitted by:
Keshav Gupta
Deepika Hegde

Contents

Problem Statement

Lending Club faces challenges in loan approval, balancing the risk of financial loss from defaults and lost business from rejections. This case study uses EDA to analyze factors influencing loan defaults, aiding better decision-making.

Data Analysis With Respect To Business Objective

- **Key Analysis**
  - Risk Assessment: Evaluate risks based on borrower traits and loan terms.
  - Trend Analysis: Identify loan performance trends, e.g., grades or home types prone to default.

Data Cleaning and Preprocessing Steps

- **Dropped Columns**:
  - Columns with more than 60% null values, as they provide no value for analysis.
  - Columns that are irrelevant for analysis, such as Id, Member-id, and Emp_Title.
  - Columns containing a single unique value, e.g., Collections_12_mths and application_type.
- **Modified Columns**:
  - Removed the text "months" from the term column.
  - Removed the % symbol from int_rate and revol_util and converted them to float.
- **Imputed Values**:
  - For missing values in emp_length, used the mode of annual_inc to fill the corresponding rows.
- **Merged Values**:
  - Combined Source Verified and Verified into a single category.
- **Dropped Rows**:
  - Removed rows with a Current loan status, as ongoing loans offer no valuable insights.
- **Converted continuous to categorical variables**
  - Annual_income, interest_rate, dti, installment are converted from continuous to categorical variables using binning.

Key Features In Dataset*

- Loan Amount
- Term
- Interest Rate
- Grade
- Sub Grade
- Annual Income
- Debt-to-Income Ratio
- Loan Status
- Home Ownership
- Purpose
- Public recorded bankruptcies
- Funded amount
- Installment
- Employee Length
- Verification status
- DTI
- Delinq_2yrs
- Pub_rec_bankrupties
- Revol_bal
- Installments

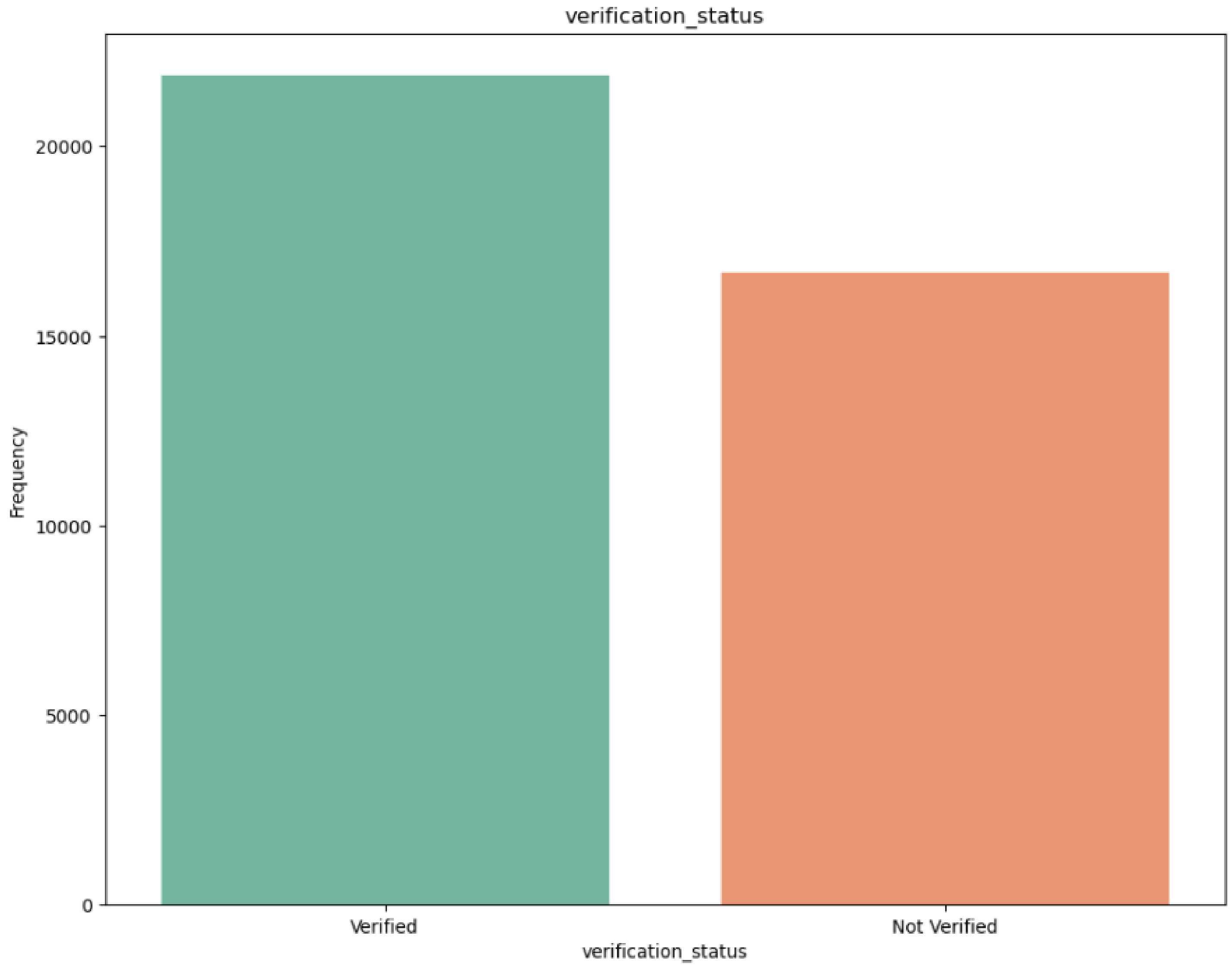
*These are the key features left after cleaning the dataset out of 111 columns.

Unordered categorical

Univariate Analysis

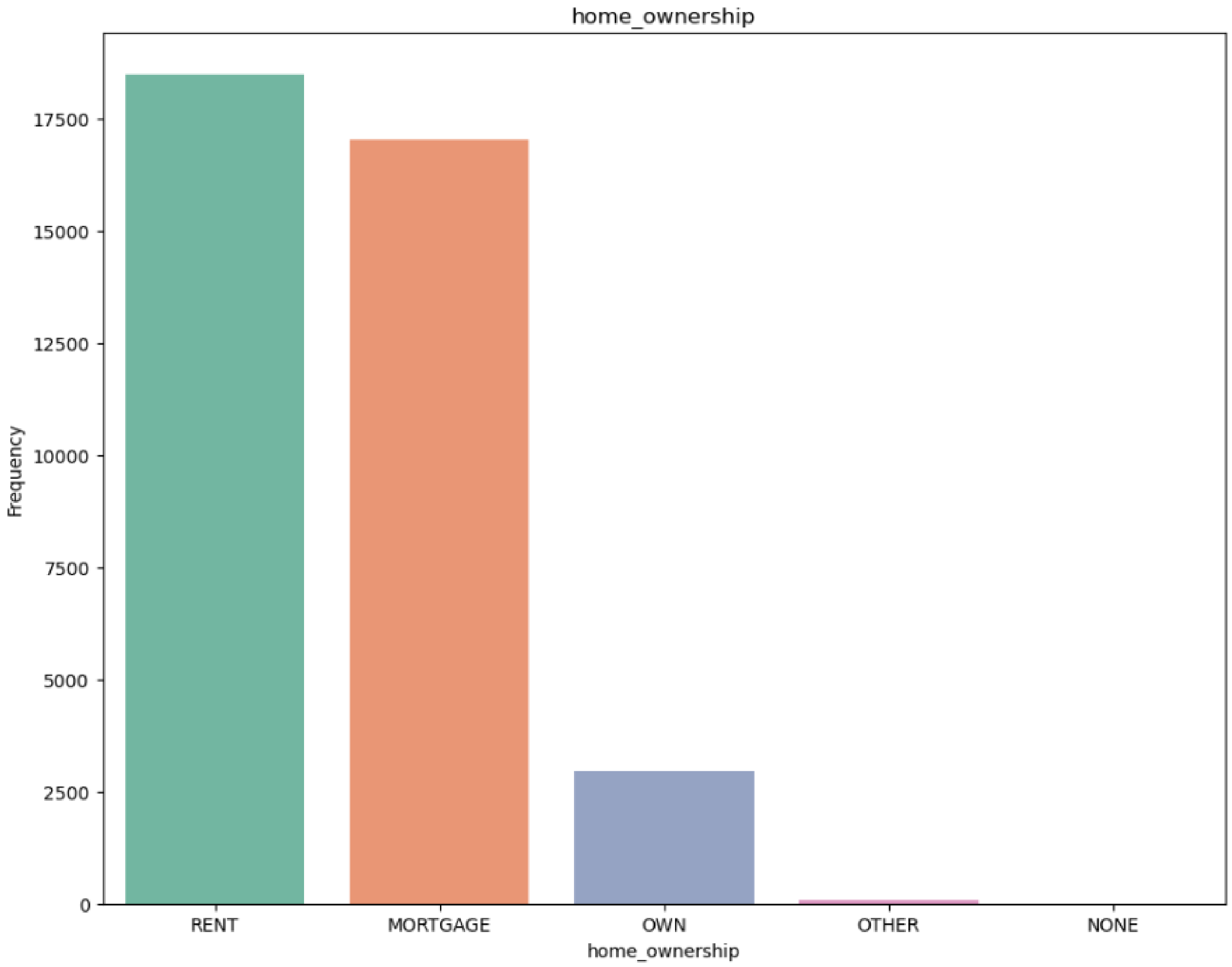
Verification Status

A higher number of borrowers are verified compared to those who are unverified.



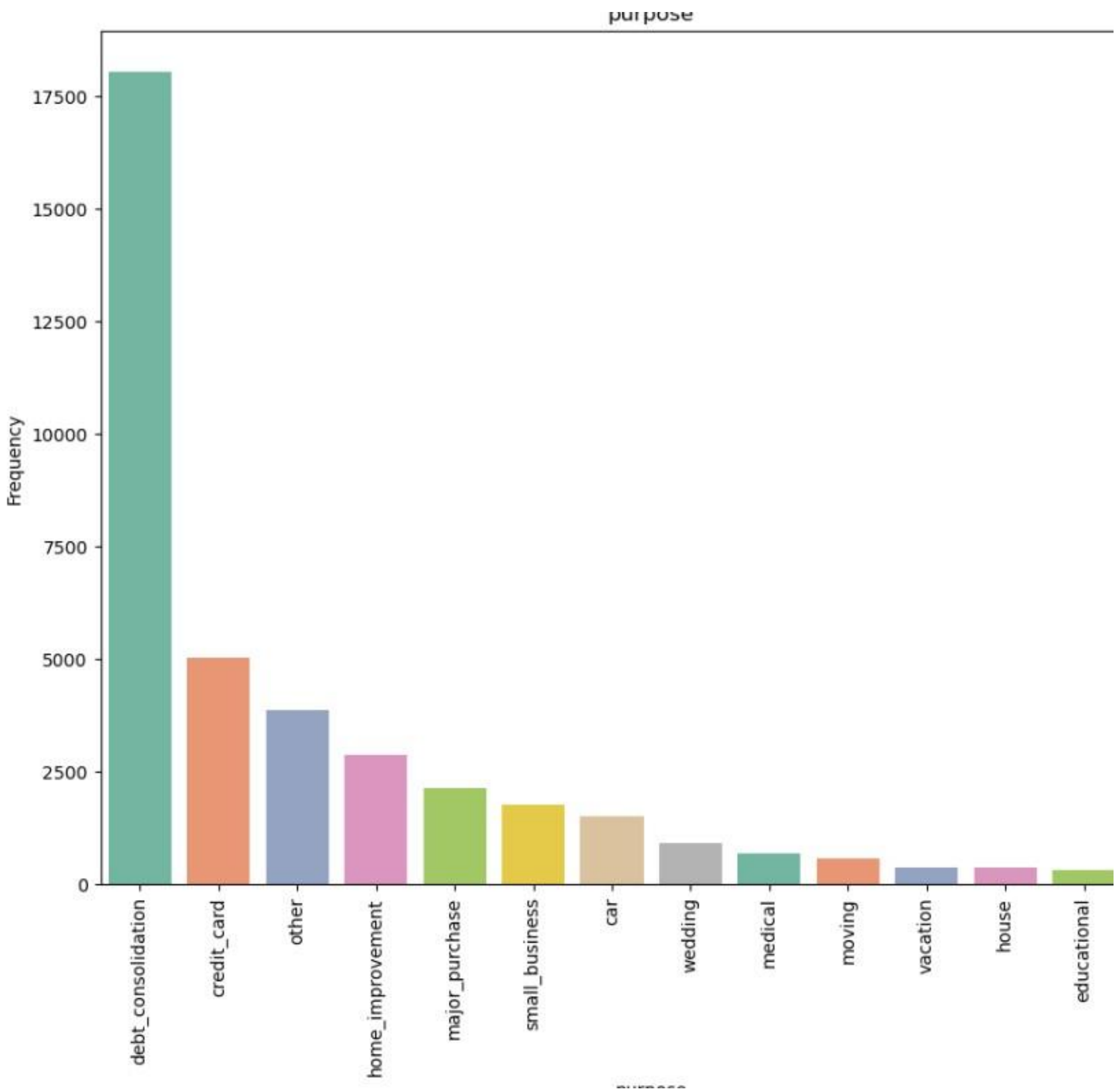
Home Ownership

Individuals with rented or mortgaged homes are more likely to take loans compared to those who own their homes outright.



Purpose

Most number of loans are taken for debt consolidation.

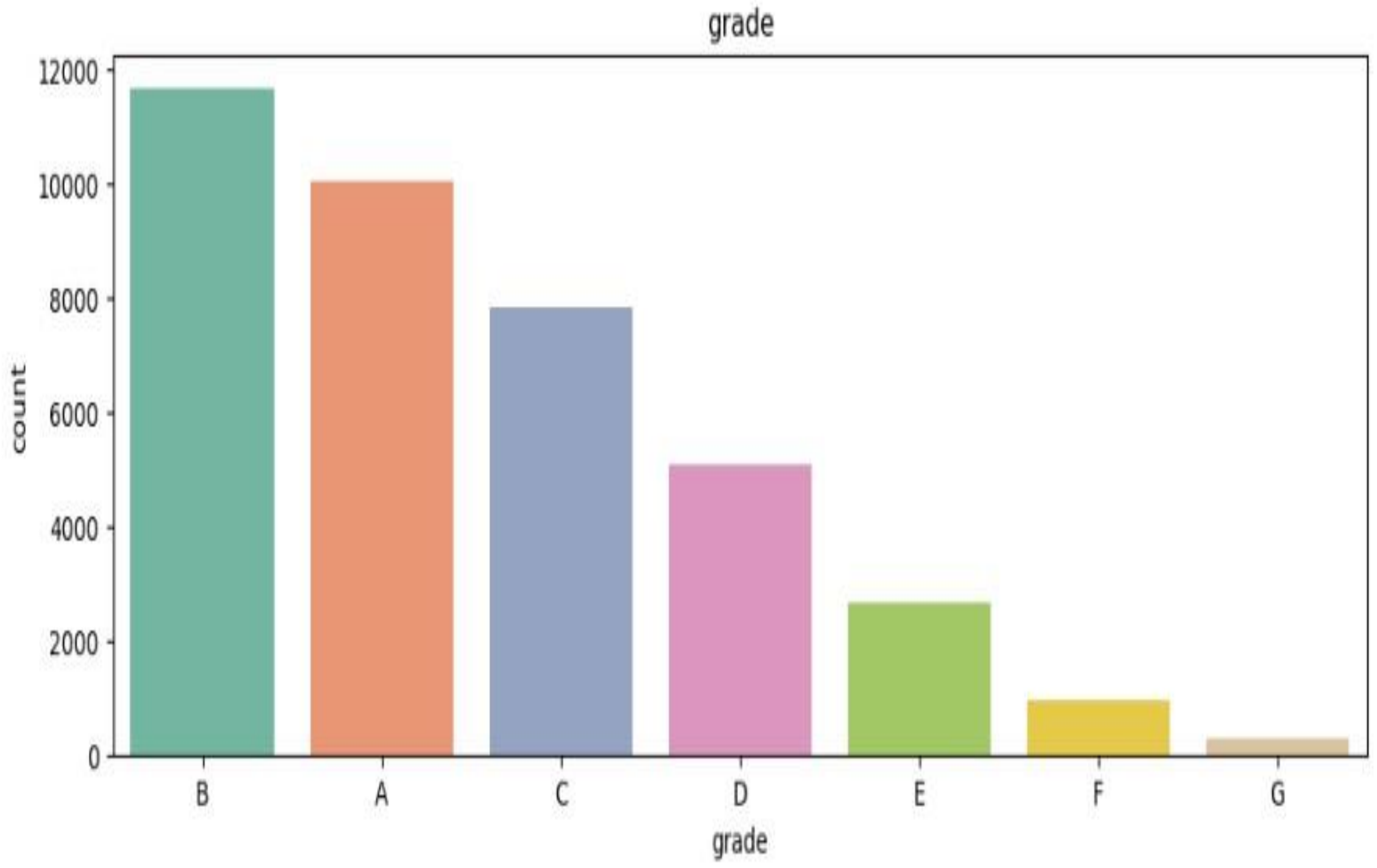


Ordered categorical

Univariate Analysis

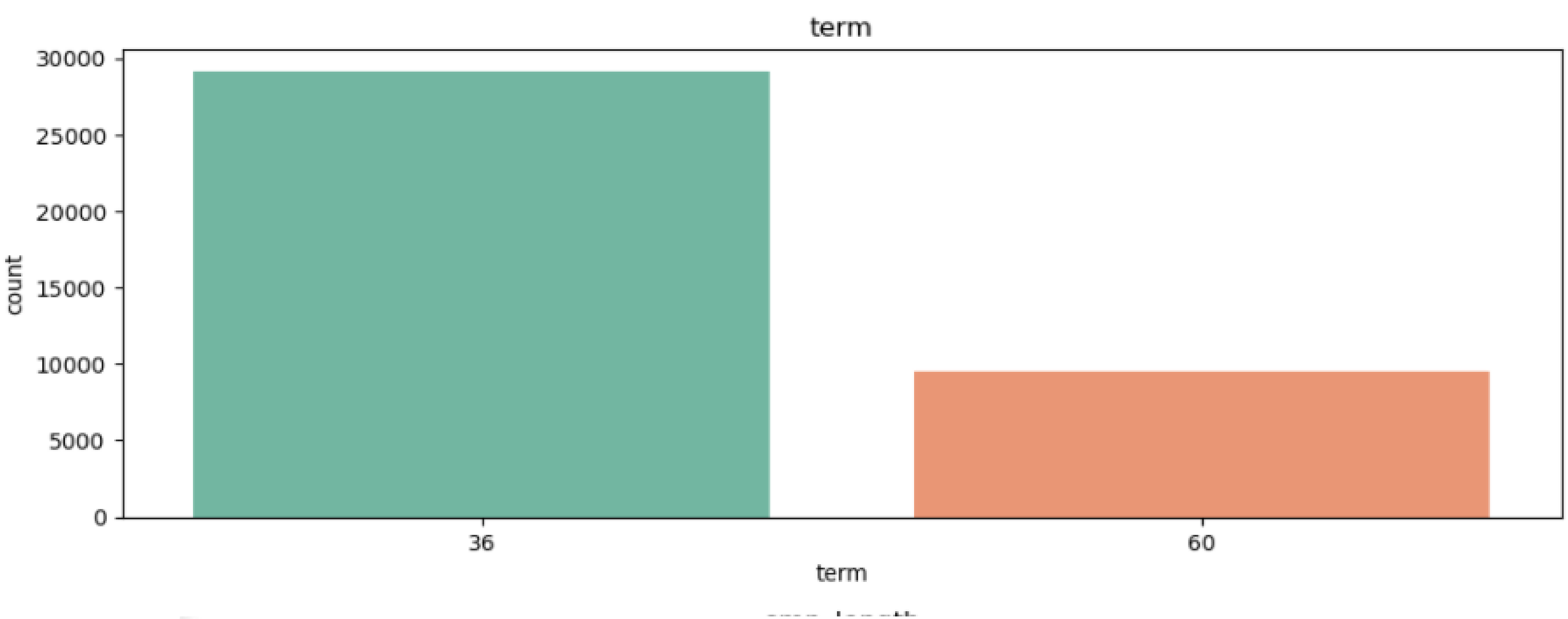
Grade

Grades A and B are the two most commonly assigned loan grades.



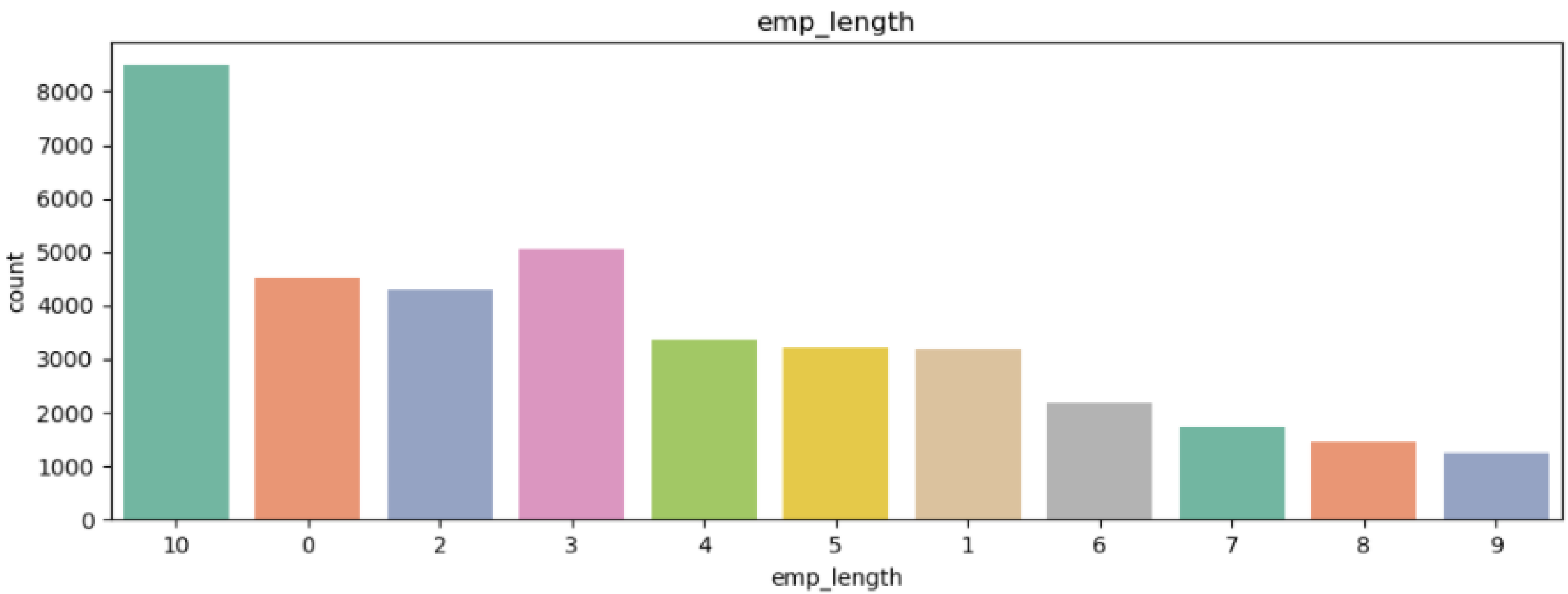
Term

Borrowers tend to take loans for shorter period (36 Months)



Employee Length

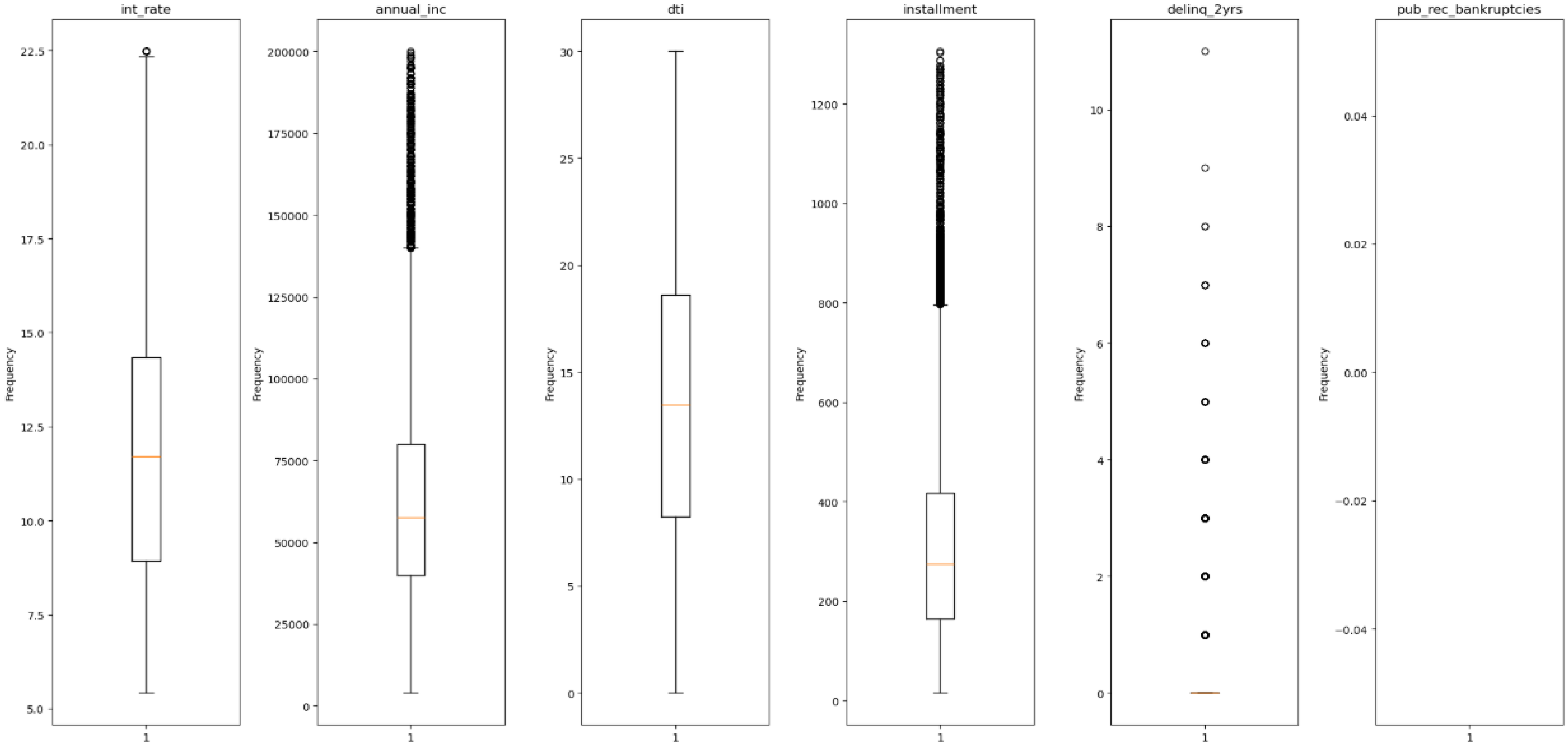
People with higher employment length tend to take more loans.



Quantitative

Univariate Analysis

Univariate Analysis – Continuous variables

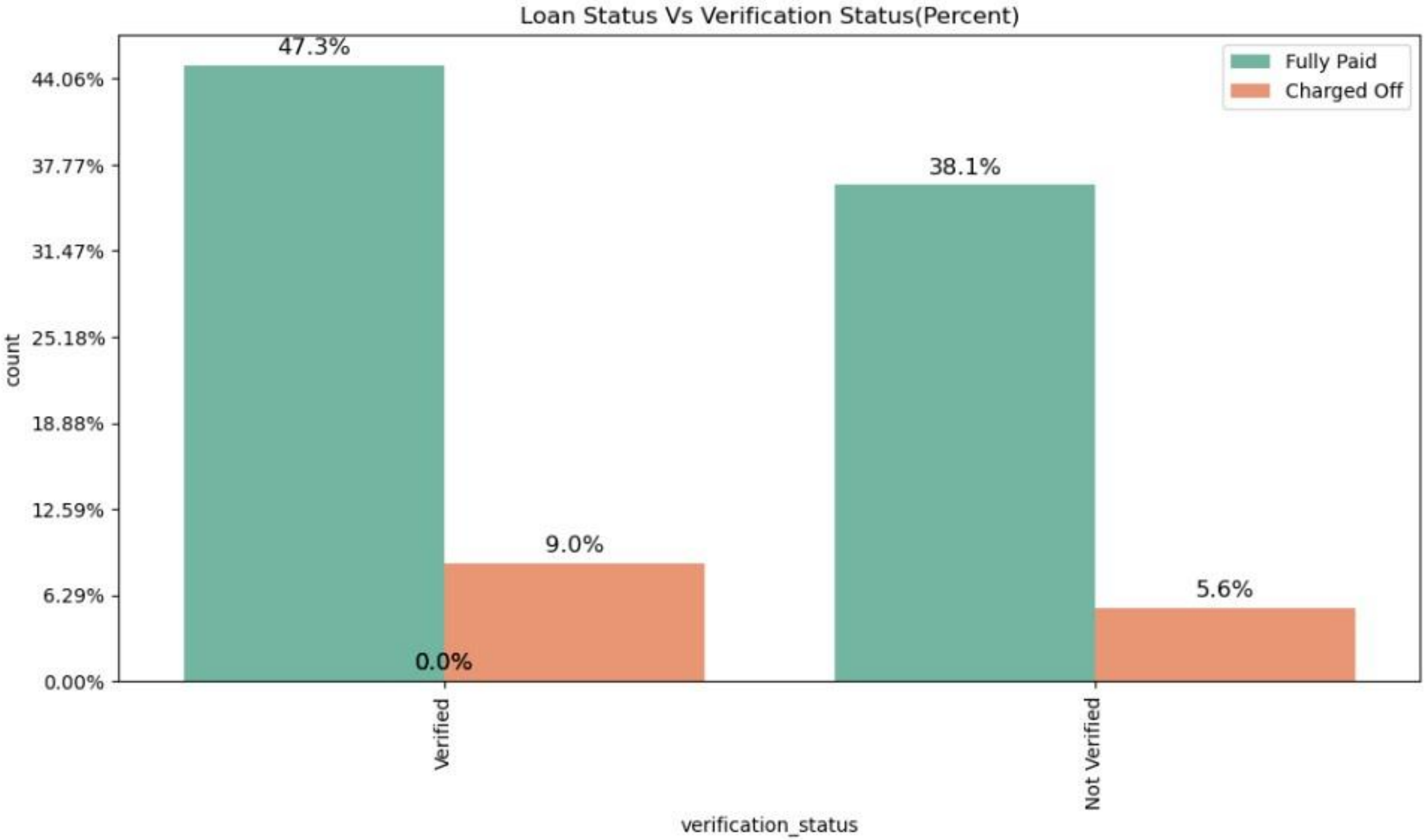


Unordered Categorical

Bivariate Analysis

Loan Status vs Verification Status

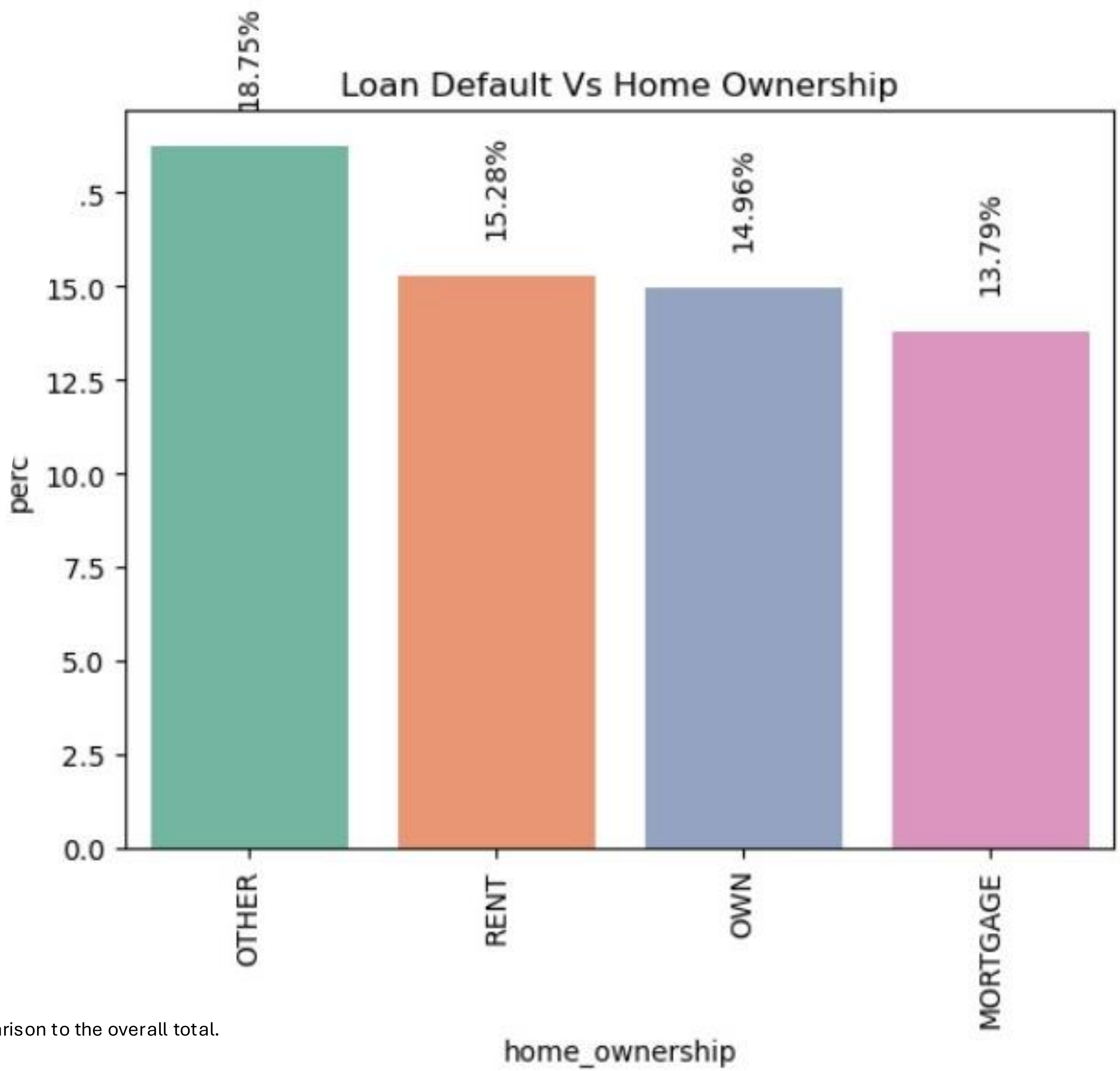
Loan applicants with a verified verification status have a higher percentage of loans that are defaulted or charged off, with 9.0% for verified applicants compared to 5.6% for unverified applicants.



Note: The percentage represents the proportion within the category, not a comparison to the overall total.

Loan Default vs Home Ownership

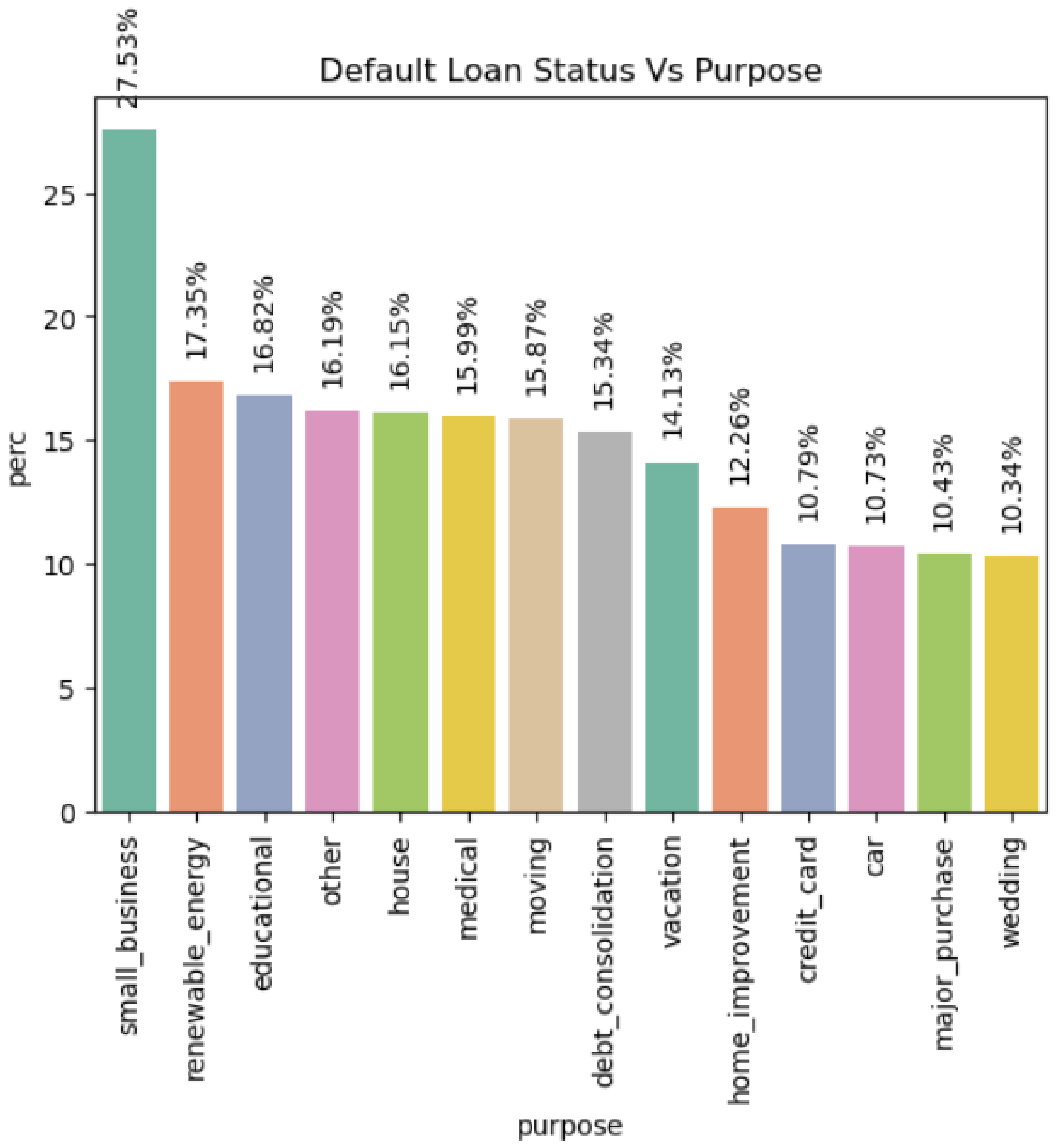
Loan applicants in the 'Other' category have a higher likelihood of being charged off, with a rate of 18.75%.



Note: The percentage represents the proportion within the category, not a comparison to the overall total.

Loan Status vs Purpose

Small businesses are more likely to default, with a default rate of 27.35%.



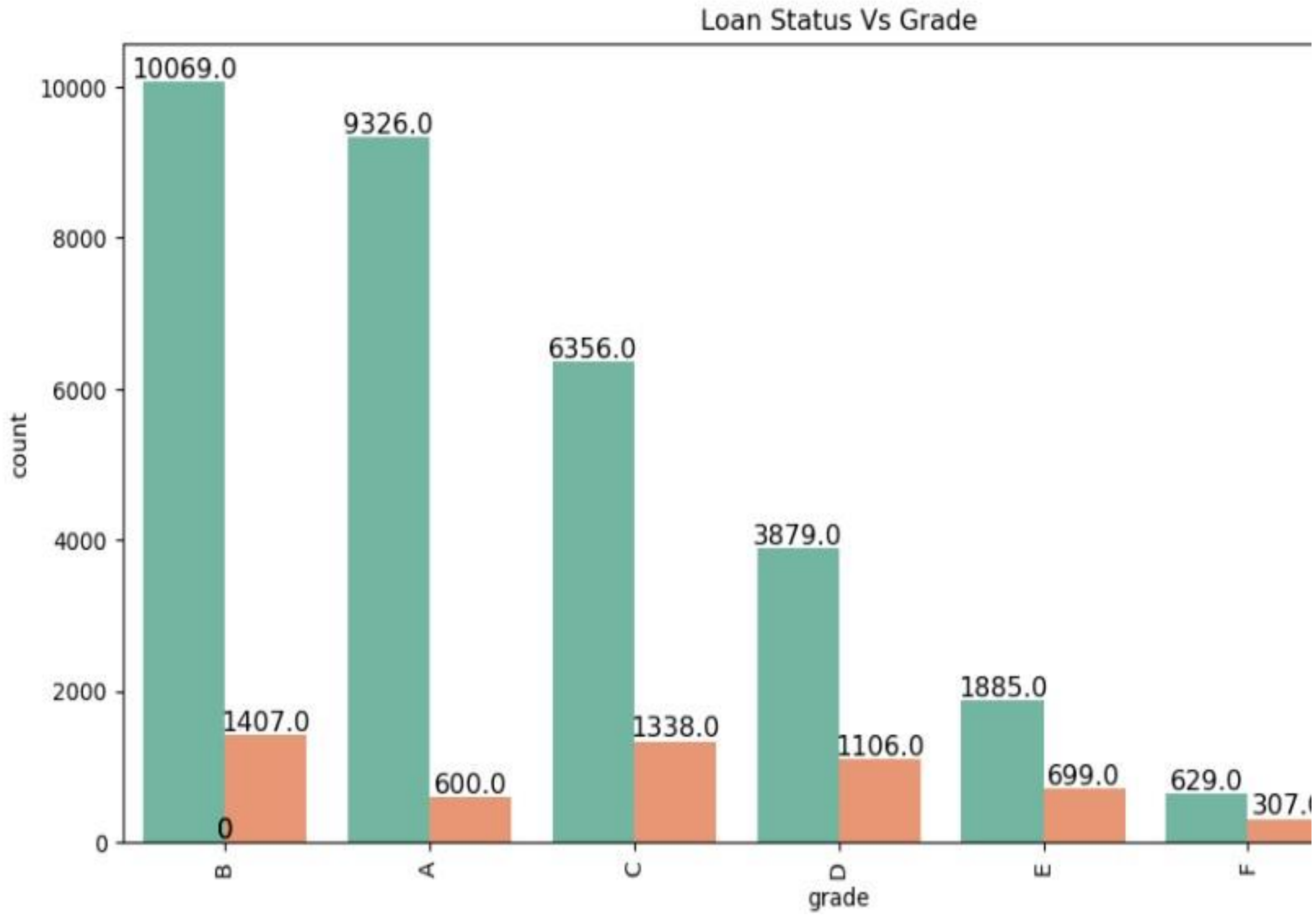
Note: The percentage represents the proportion within the category, not a comparison to the overall total.

Ordered Categorical

Bivariate Analysis

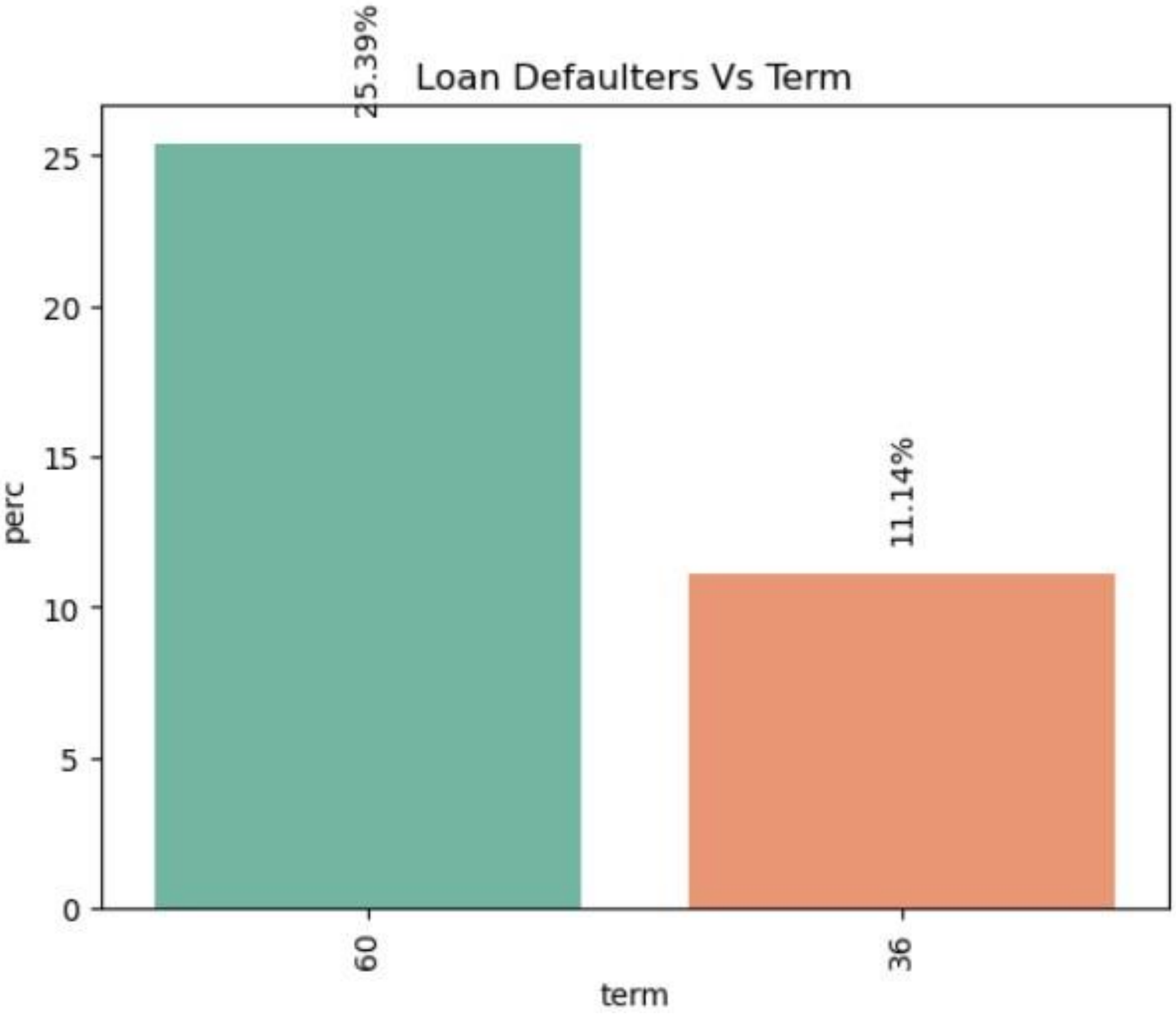
Loan Status vs Grade

Loan applicants in grades B, C, and D have the higher number of charge-offs.



Loan Defaulters vs Term

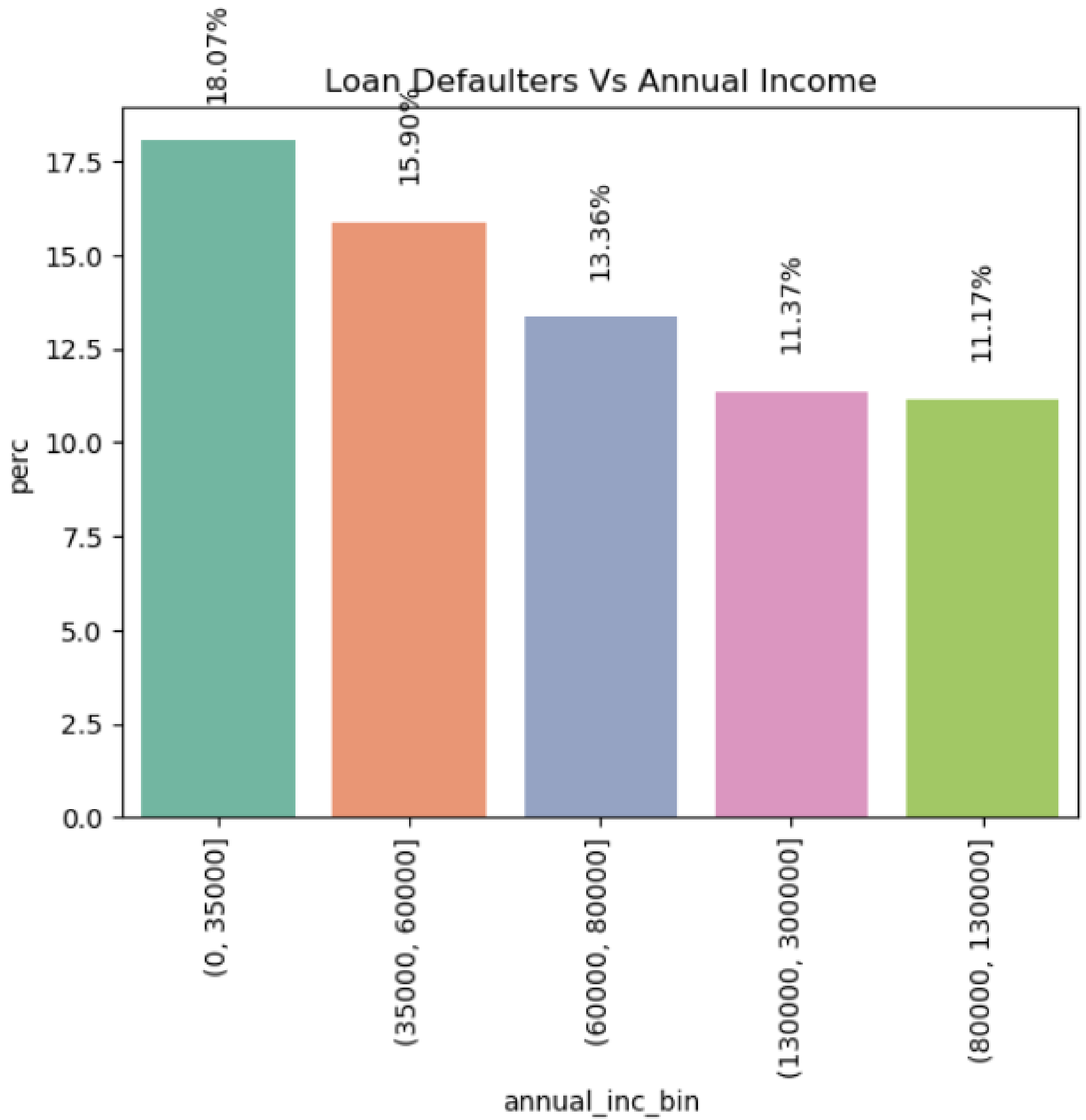
Loan applicants with longer repayment terms are more likely to default, with a default rate of 25.75%.



Note: The percentage represents the proportion within the category, not a comparison to the overall total.

Loan Status vs Annual Income

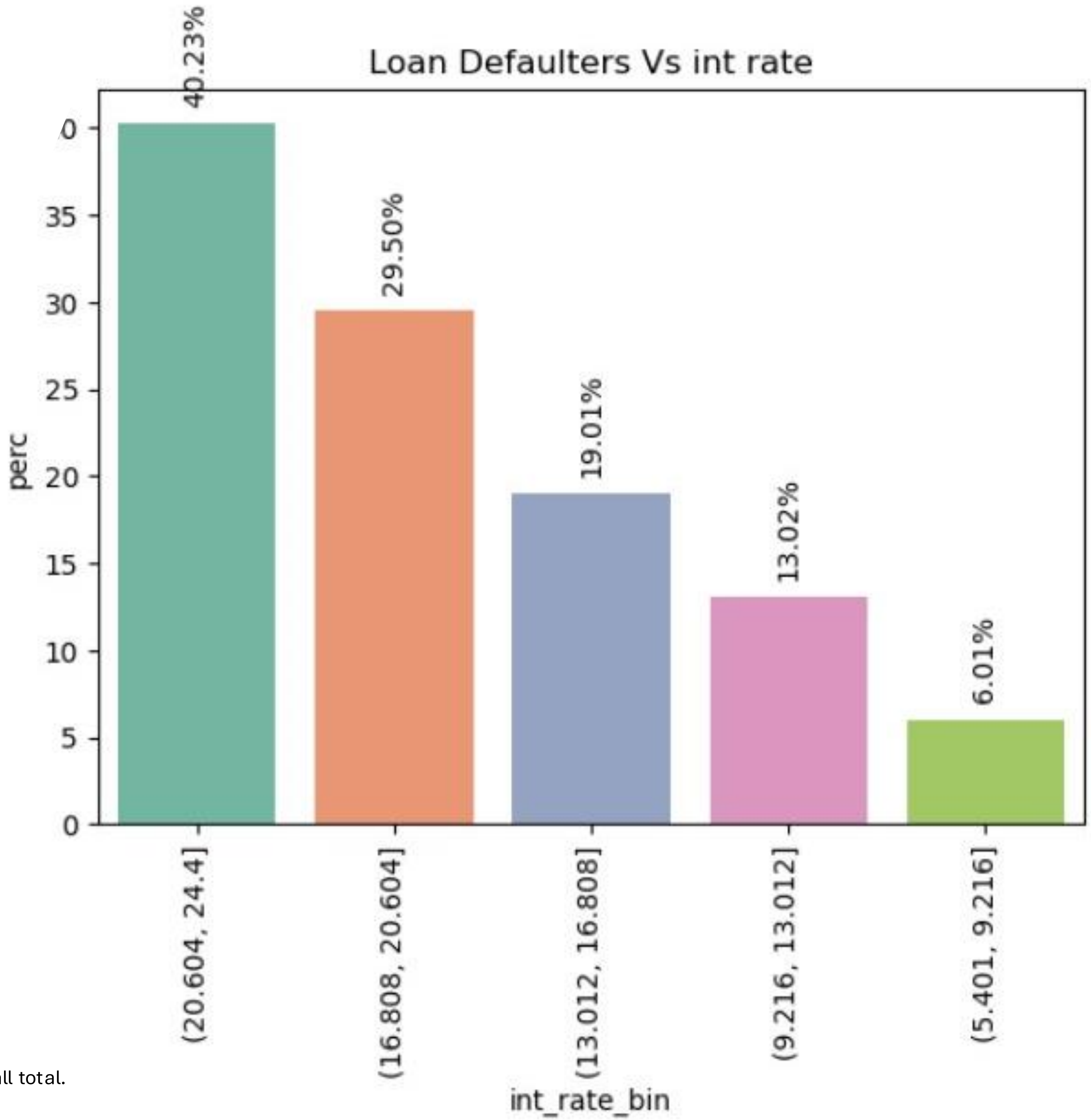
People with an annual income of less than 35,000 are more likely to default on loans, with a default rate of 18.07%.



Note: The percentage represents the proportion within the category, not a comparison to the overall total.

Loan Status vs Interest Rate

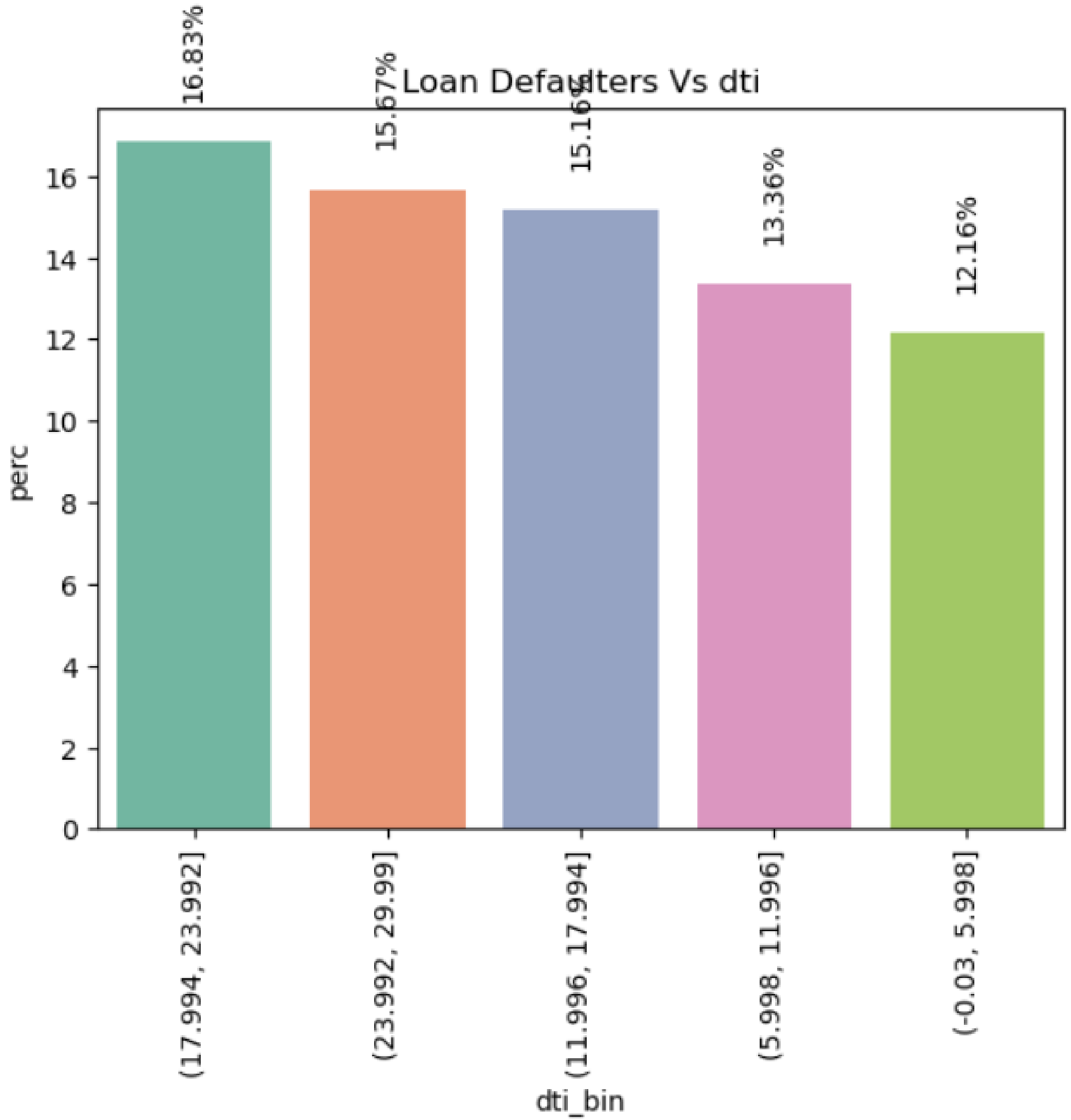
40.8% of loans with an interest rate above 20% default, indicating that loans with higher interest rates have a greater likelihood of default.



Note: The percentage represents the proportion within the category, not a comparison to the overall total.

Loan Defaults vs dti(debt to income)

Borrowers with higher dti tend to default more.



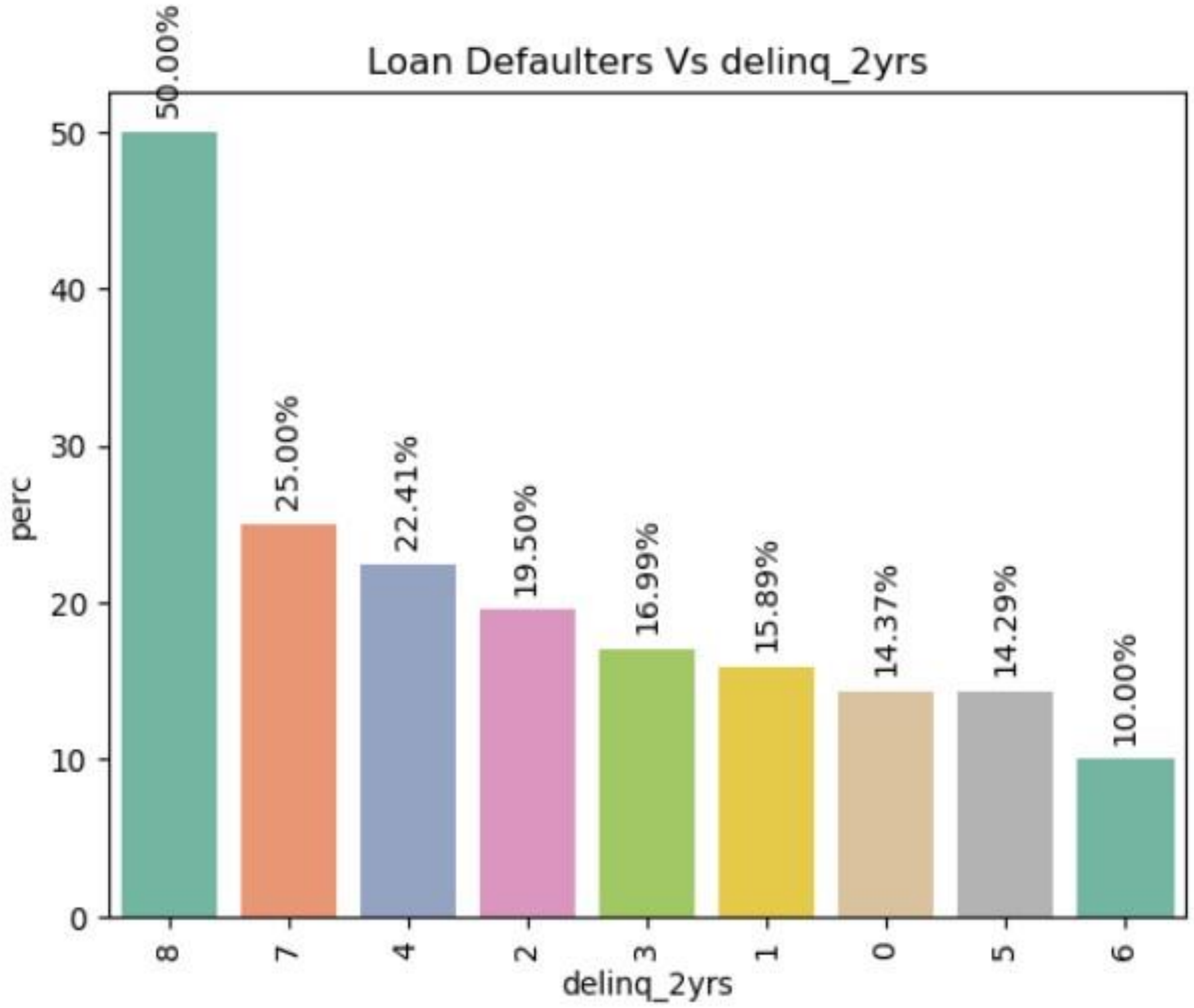
Note: The percentage represents the proportion within the category, not a comparison to the overall total.

Quantitative

Bivariate Analysis

Loan Defaulters Vs Delinq_2yrs

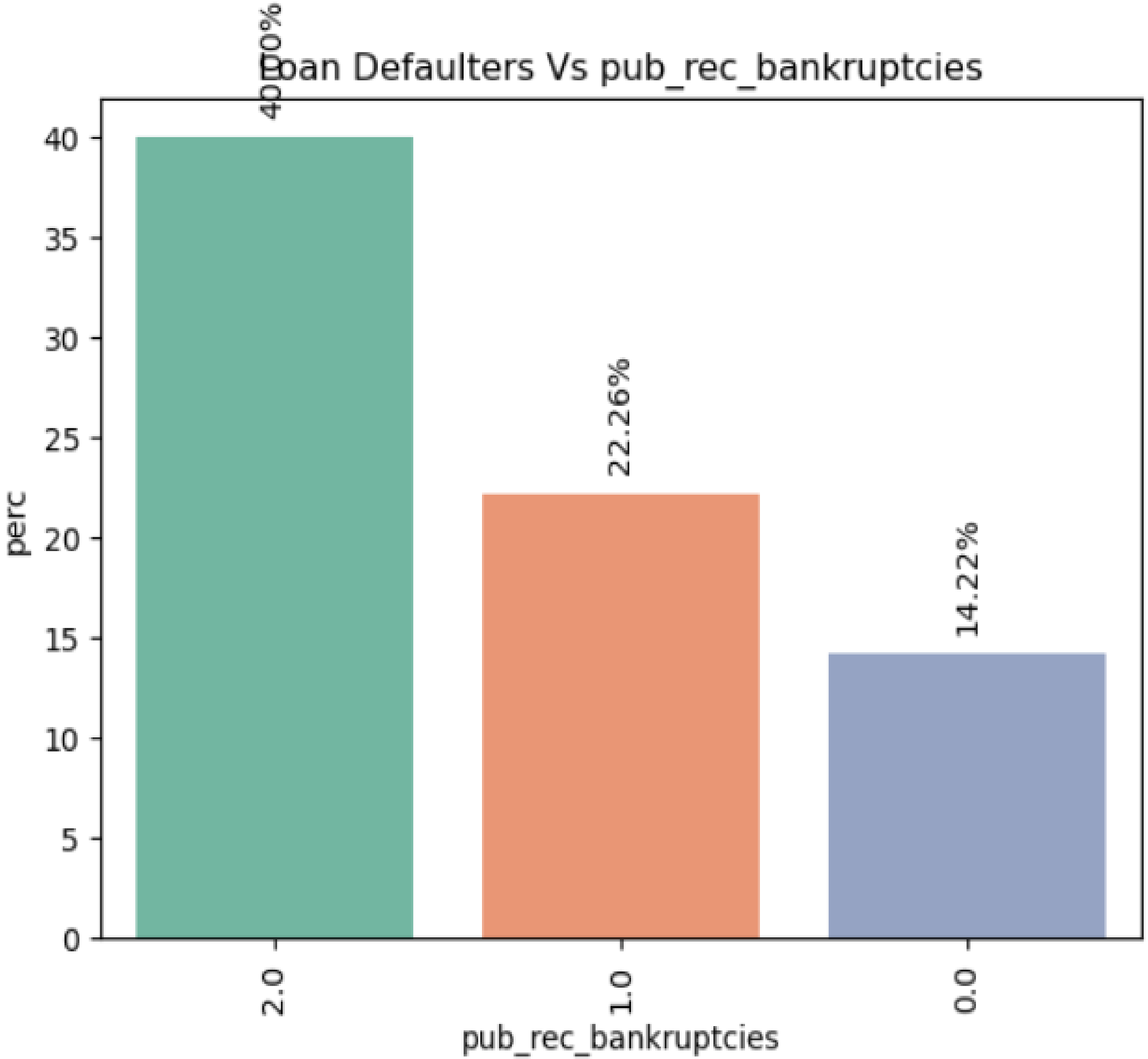
Borrowers who have had 8 or more delinquencies in the past 2 years are more likely to default, with a default rate of 50%.



Note: The percentage represents the proportion within the category, not a comparison to the overall total.

Loan Defaulters vs Pub recorded bankruptcies

Borrowers with 2 or more public bankruptcies are more prone to default, with a default rate of 40%.



Note: The percentage represents the proportion within the category, not a comparison to the overall total.

Recommendations

- Small business borrowers have a higher likelihood of defaulting.
- Borrowers with public bankruptcies are more prone to default.
- Higher debt-to-income (DTI) ratios are associated with increased defaults.
- Long-term loans have higher default rates, suggesting a focus on short-term loans may reduce risk.
- The current verification process is ineffective; institutions should strengthen it.
- Extra caution is advised when lending to borrowers with homeownership status listed as "Others."
- Lenders should exercise caution when approving loans for lower-income borrowers.