Lending Club Case Study

Problem Statement:

- The company wants to identify factors influencing loan default to minimize credit loss.
- The goal is to develop an understanding of how consumer attributes and loan attributes influence the tendency of default.

Analysis Approach:

- We'll start by loading the loan dataset and performing exploratory data analysis (EDA) to understand the data's structure and characteristics.
- Univariate analysis will involve examining individual variables to understand their distributions and identify any anomalies or patterns.
- Bivariate analysis will explore relationships between pairs of variables, particularly focusing on their association with loan default.
- Business driven, Type Driven and Data driven analysis
- We'll use visualizations such as histograms, box plots, and scatter plots to illustrate distributions and relationships.
- Insights from the analysis will be interpreted in business terms, focusing on the implications for loan approval decisions and risk assessment.
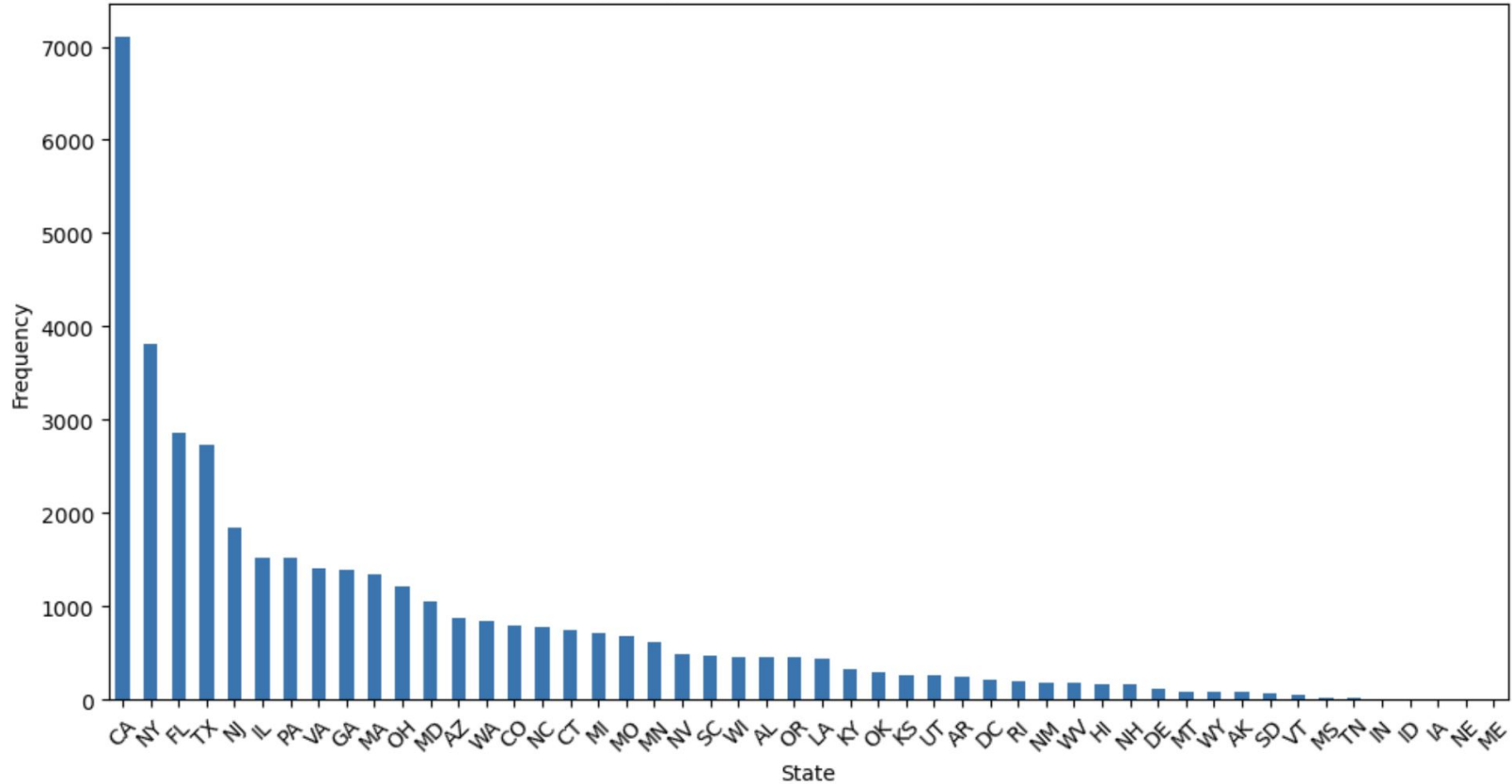
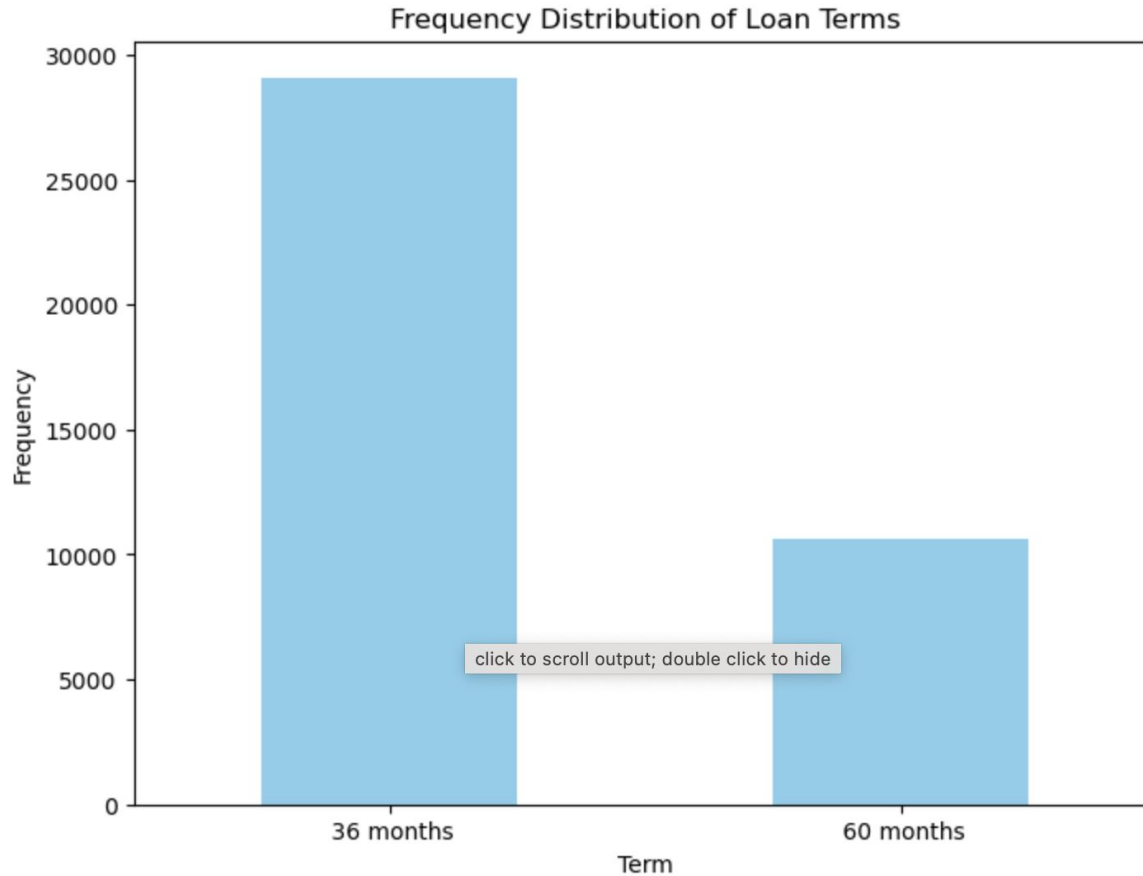# Loan Status Distribution

loan_status Distribution
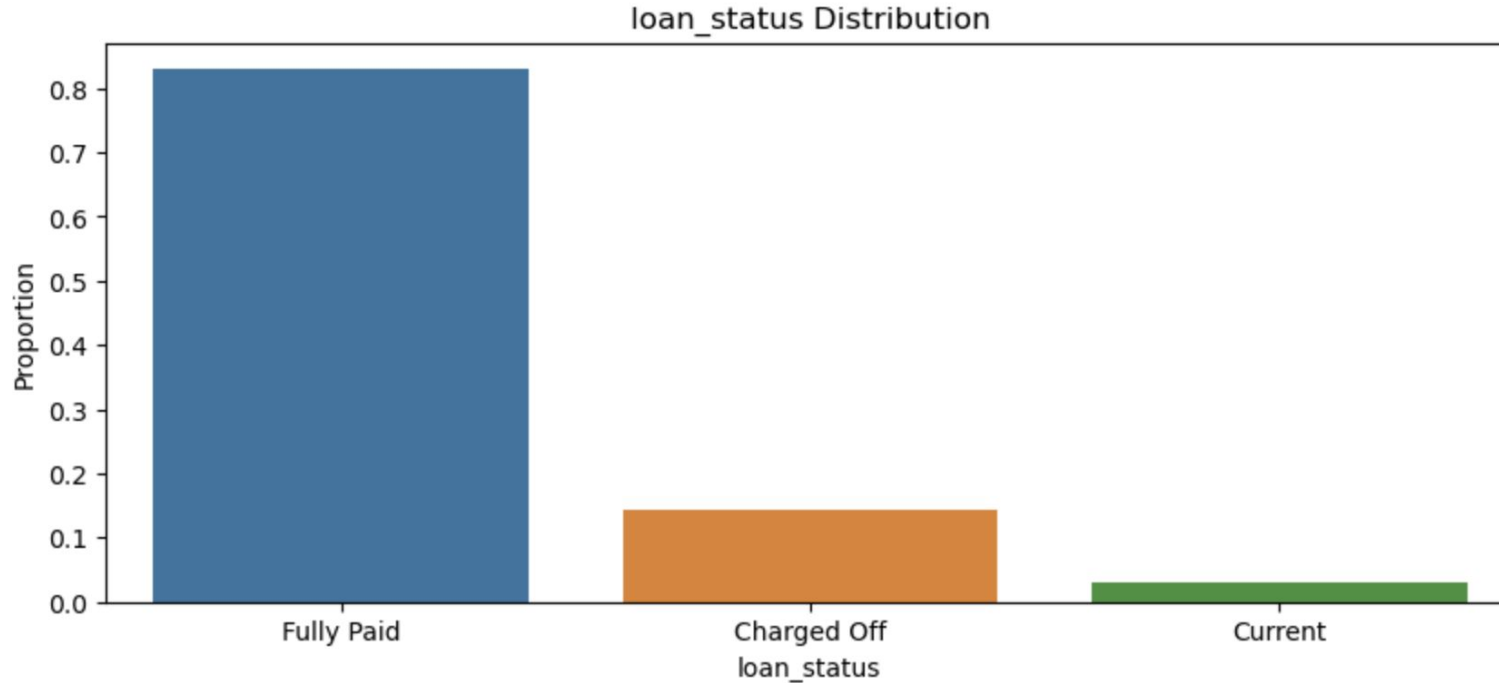
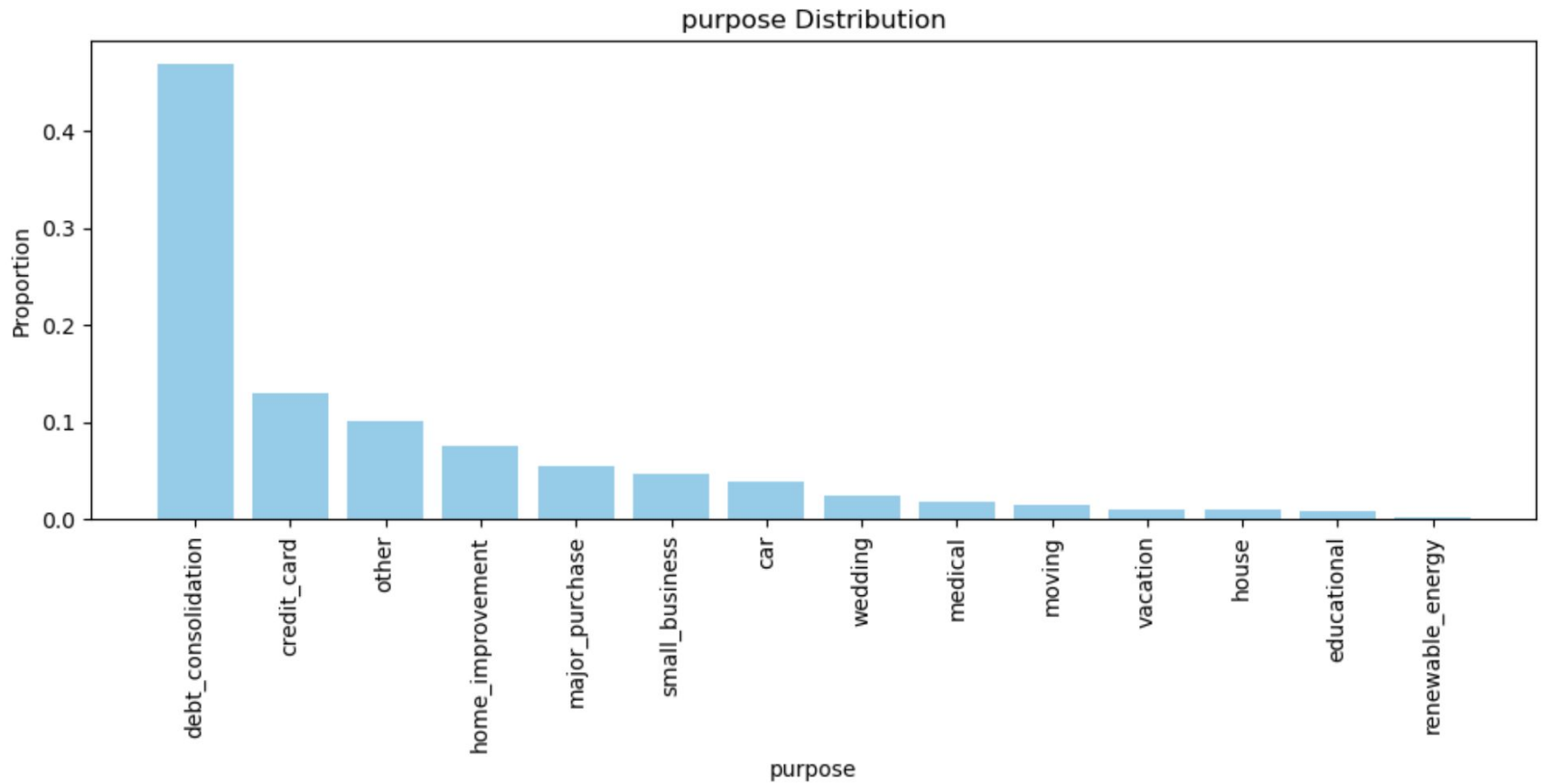**Observation: 85% of borrowers has paid the loan fully ,14% are defaulted the loan and 5% were paying loan currently**

Frequency Distribution of Loans by State

Observation: CA state has the maximum amount of loan applications

Frequency Distribution of Loan Terms

Observation : Majority of the loan applications counts are in the term of 36 months.

purpose Distribution

Observation: More number of people took loan for debt consolidation and a very few people took for renewable energy

Univariate Analysis:

- Loan Status Distribution: About 85% of borrowers paid their loans in full, 14% defaulted, and 5% were in the process of repaying.
- State-wise Distribution: California had the highest number of loan applications.
- Term Distribution: The majority of loan applications had a term of 36 months.
- Loan Purpose Distribution: Debt consolidation was the most common purpose for taking a loan.
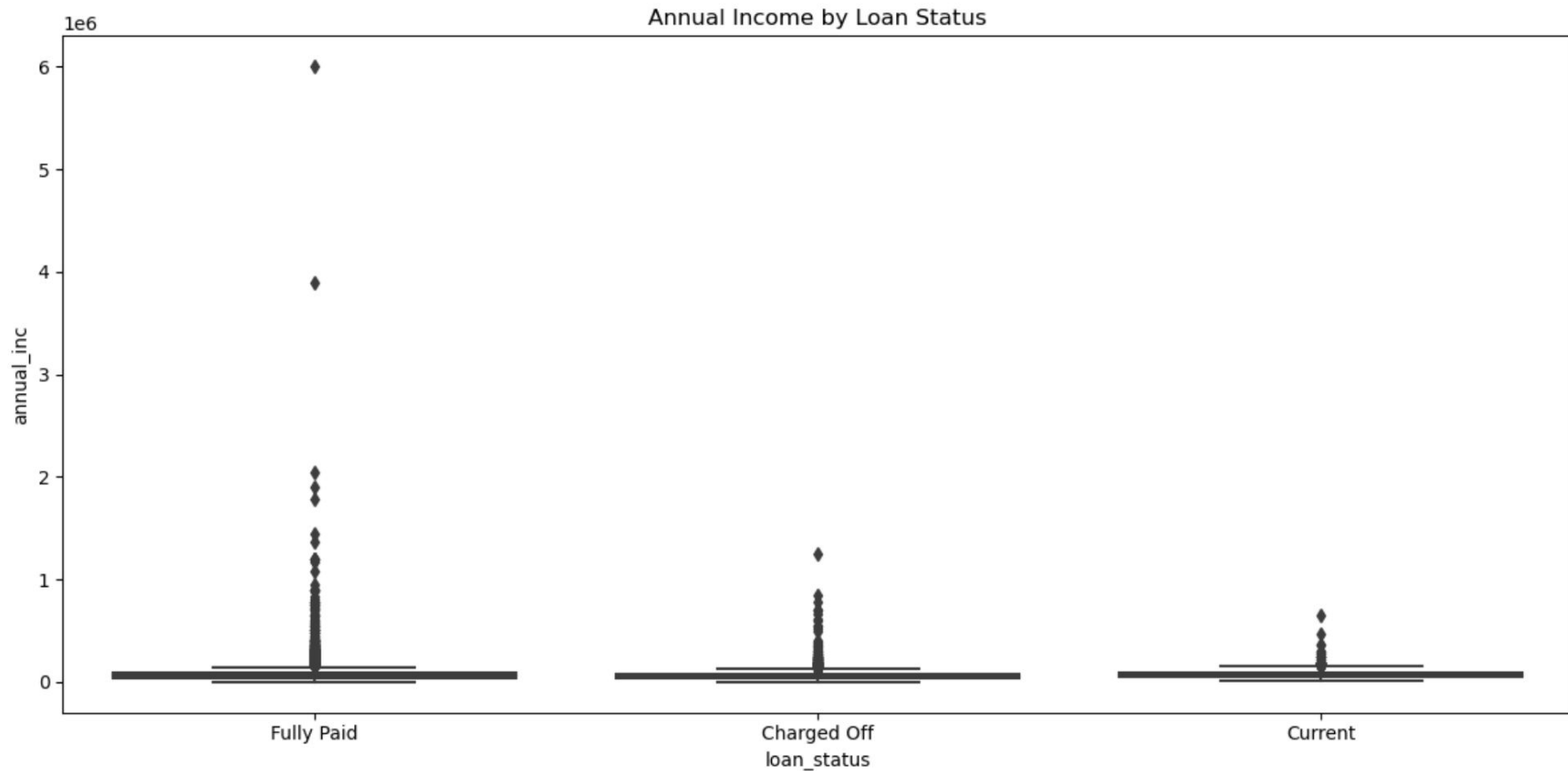
# Bivariate Analysis: Relationship between loan status and other variables
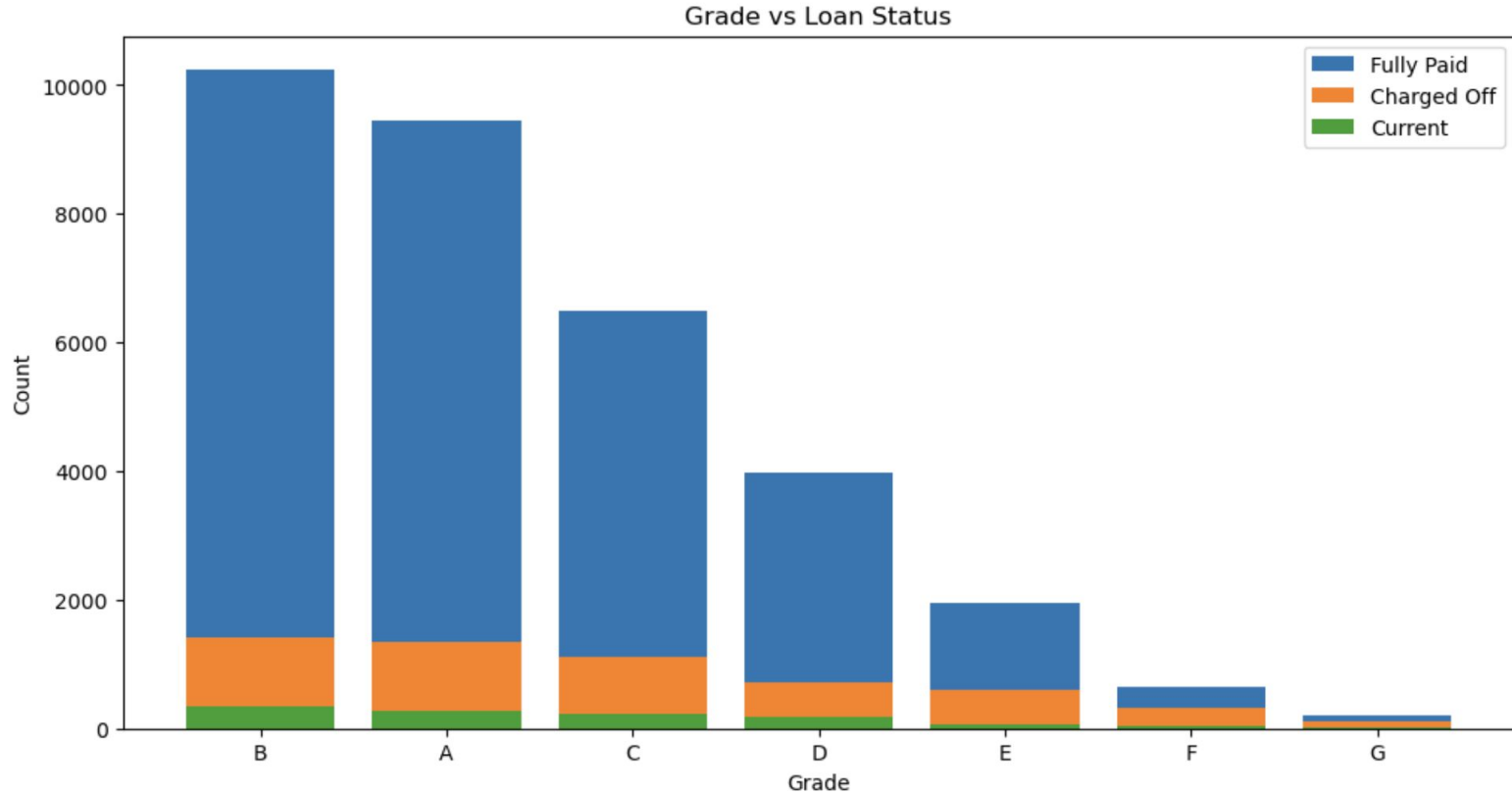
Observation: The loan applicants belonging to Grades B, A, C and D contribute to most number of "Charged Off" loans

Loan Amount by Loan Status
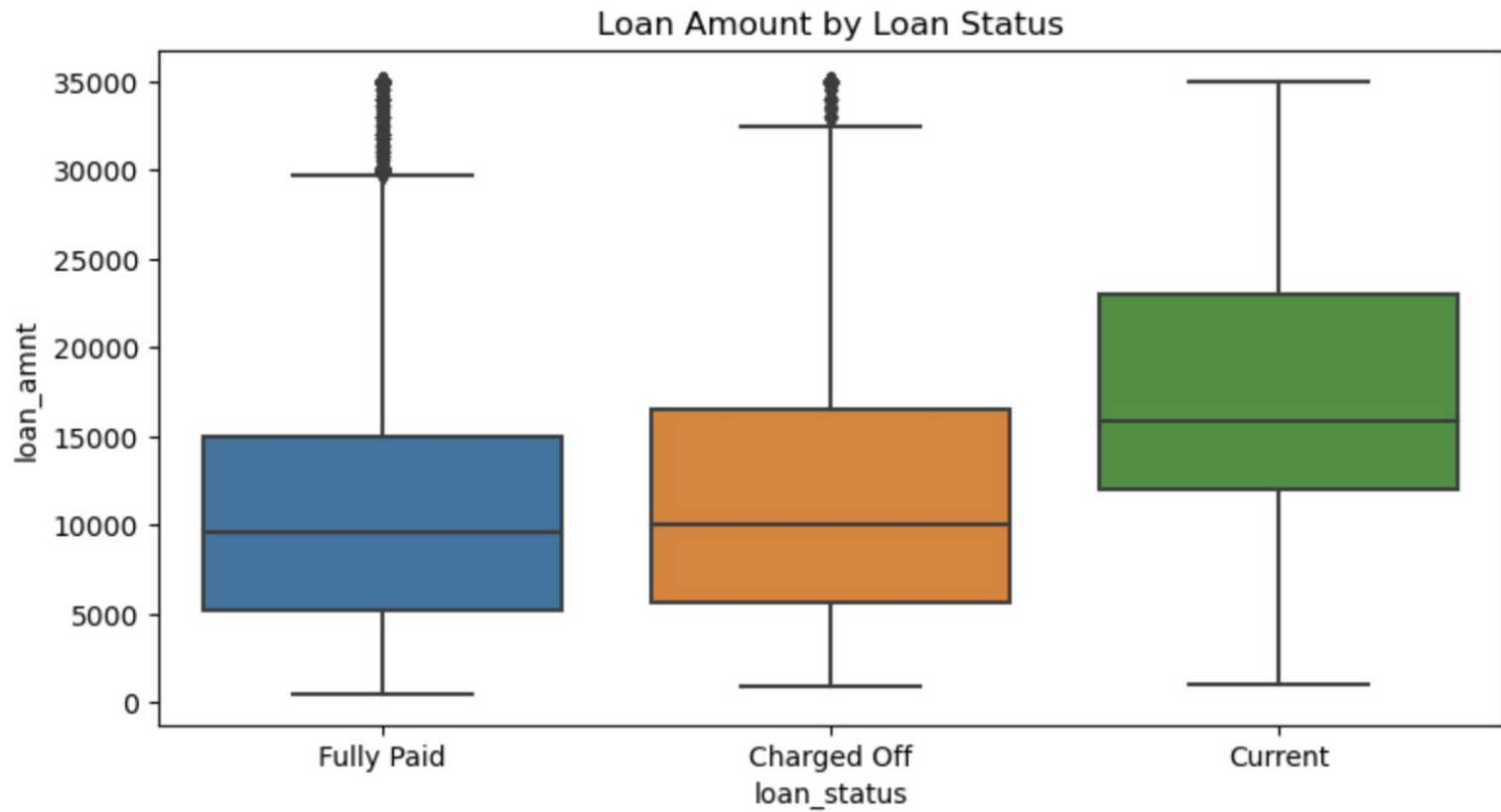
Observation: Charged Off loans have higher loan amounts than Fully Paid ones.

Annual Income by Loan Status

Observation: Charged Off loans have lower annual income than Fully Paid ones.
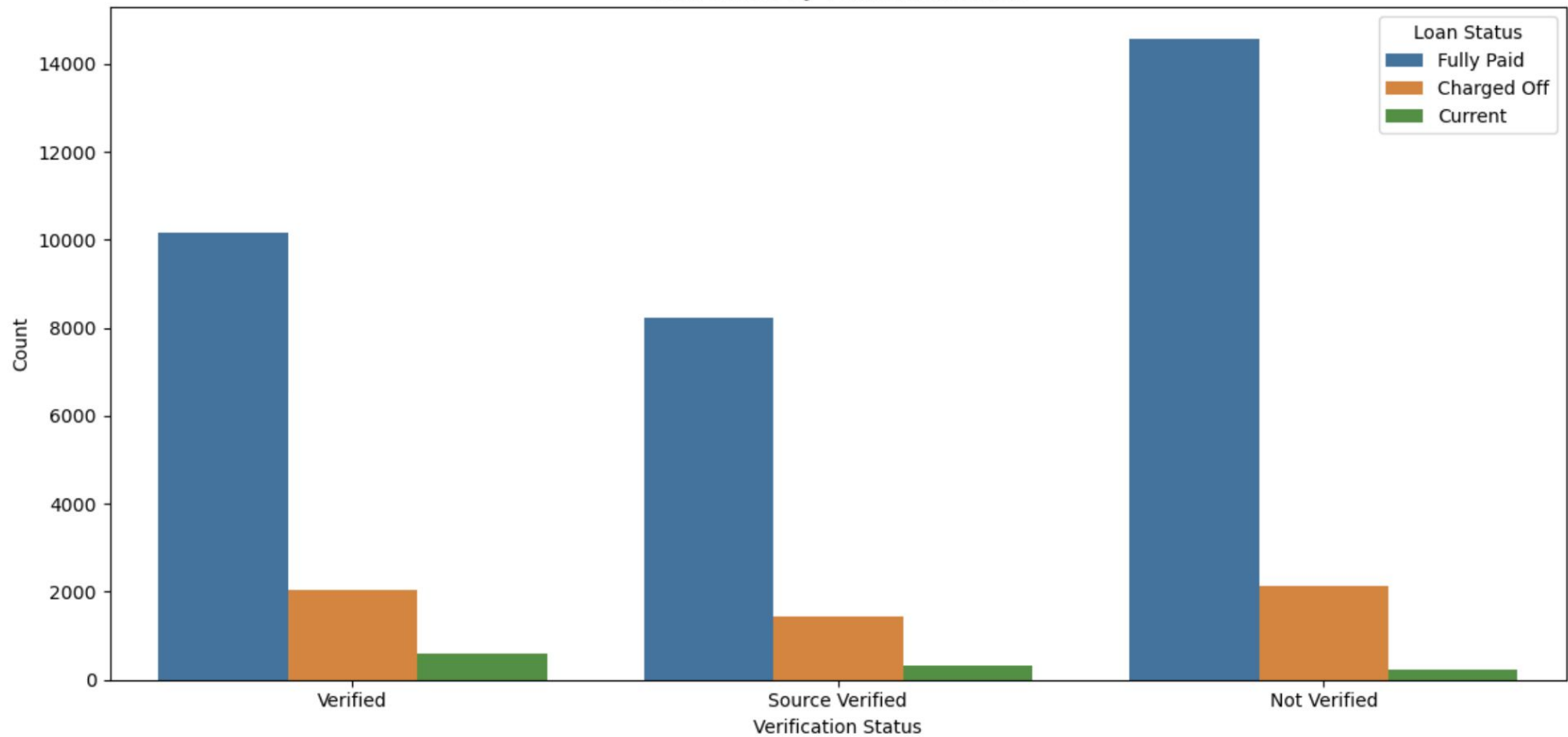
Term vs Loan Status
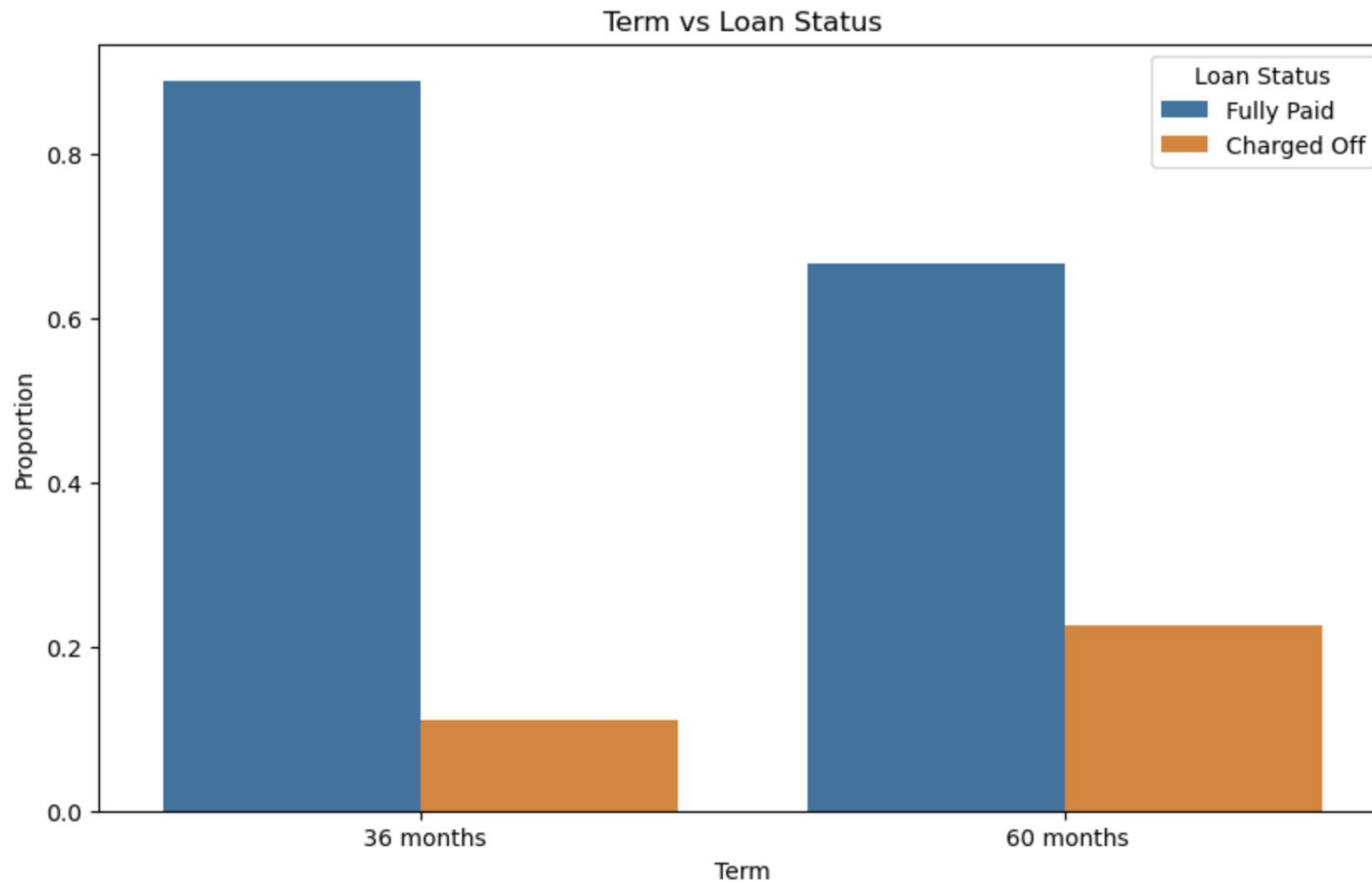
Observation: There are more proportion of borrowers defaulted loan in 60 months term then 36 months. Also the Fully Paid rate is higher in 36 months tenure.
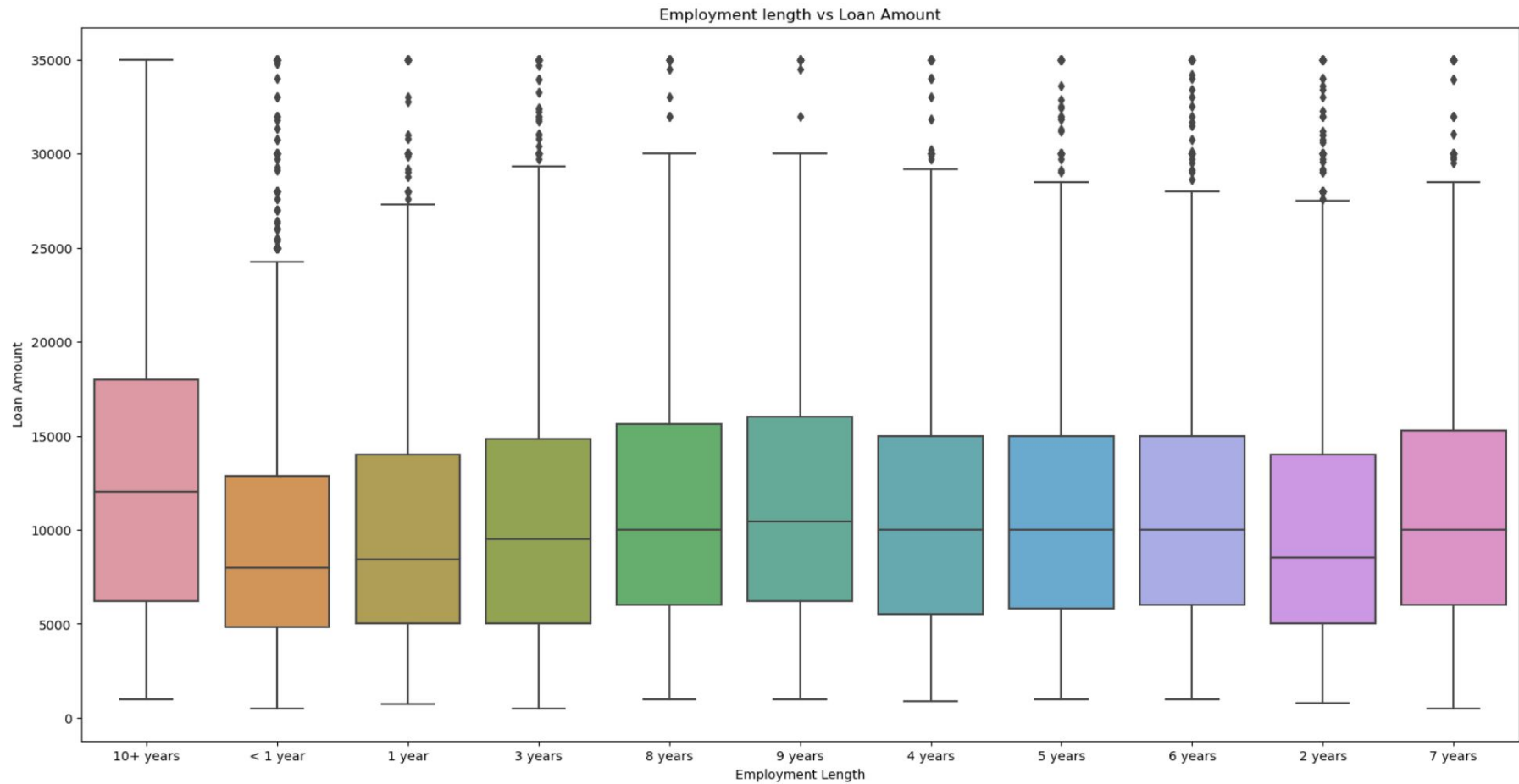
Loan Status by Home Ownership
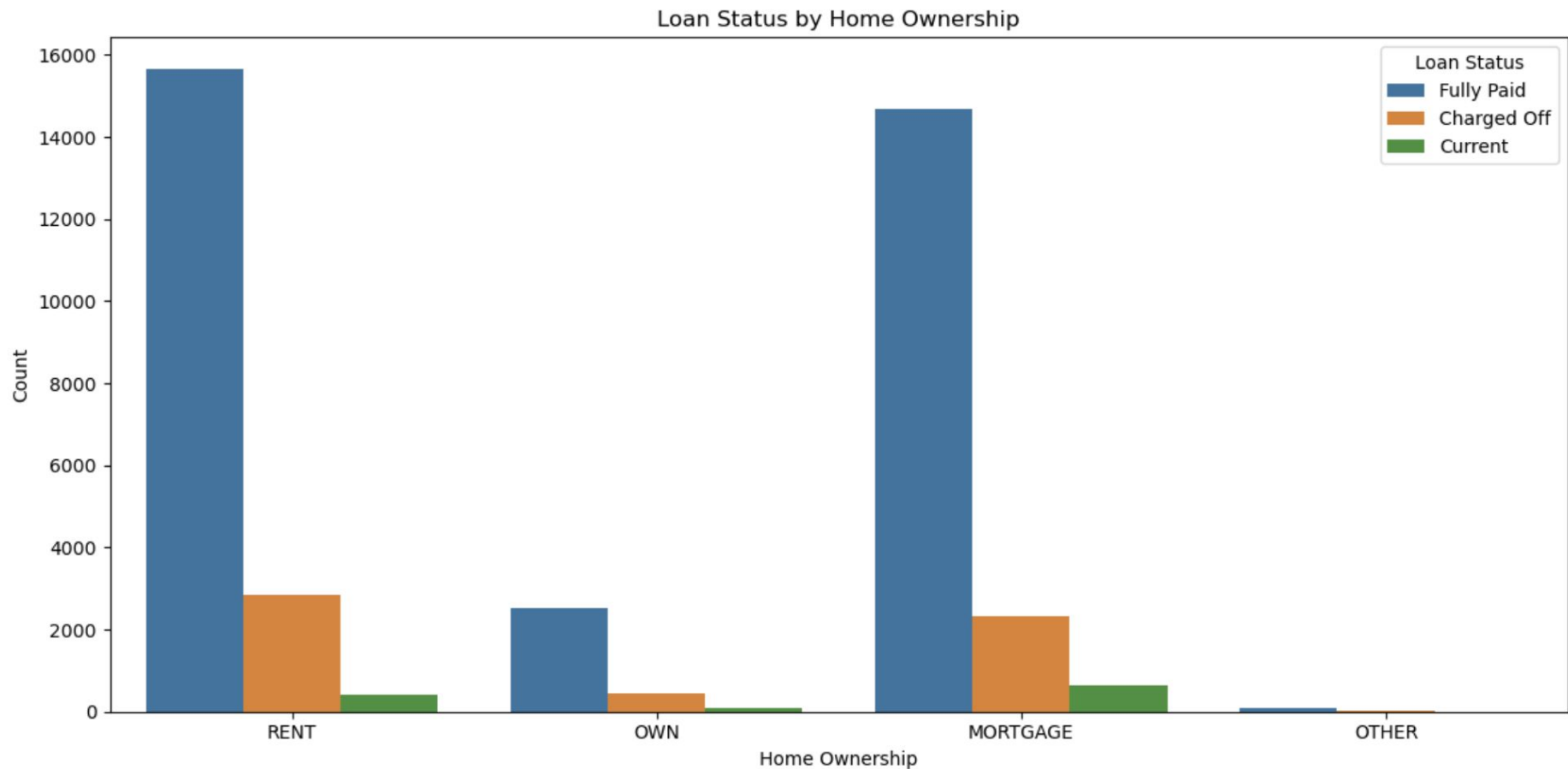
Observation: Overall highest Charge Off numbers are in the category of RENT and MORTGAGE

Loan Status by Verification Status

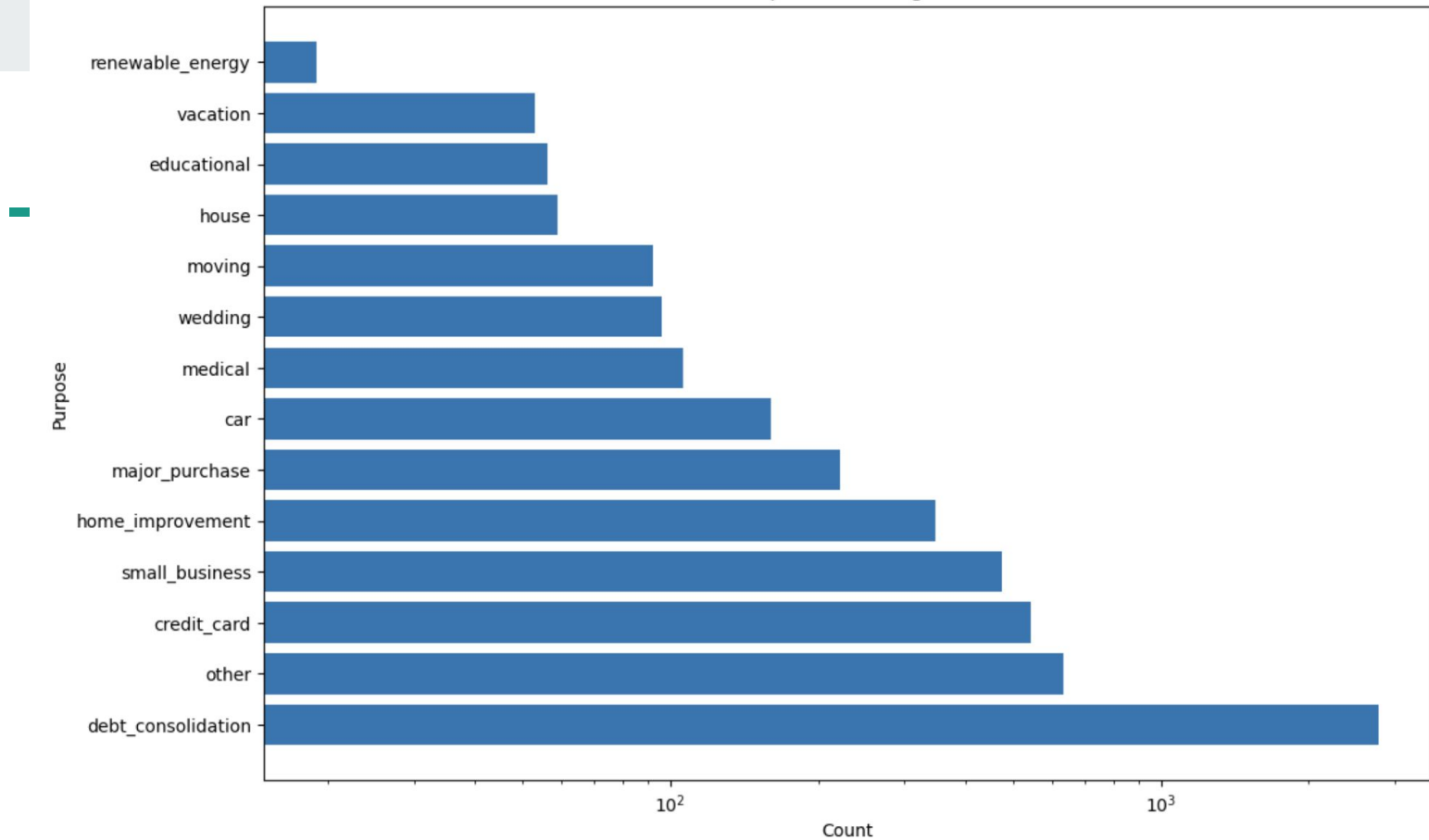**Observation: Majority of loans were given without verification of applicants income.**

Employment length vs Loan Amount

Observation: More borrowers are from 10+ years and least is <1 year

Count of Purpose for Charged Off Loans

Observation: The likelihood of loans being charged off (defaulted - not repaid) is higher when the purpose of the loan is 'debt consolidation'.

Bivariate Analysis:

- Grade vs. Loan Status: Grades B, A, C, and D contributed the most to "Charged Off" loans.
- Loan Amount vs. Loan Status: Charged Off loans had higher loan amounts than Fully Paid ones.
- Annual Income vs. Loan Status: Charged Off loans had lower annual incomes than Fully Paid ones.
- Term vs. Loan Status: The proportion of defaulted loans was higher for 60-month terms compared to 36-month terms.
- Home Ownership vs. Loan Status: Rent and Mortgage categories had the highest number of Charge Offs.
- Verification Status vs. Loan Status: Most loans were given without verifying applicants' income.
- Employment Length vs. Loan Amount: Borrowers with 10+ years of employment had the highest loan amounts.
- Purpose vs. Loan Status: Loans for debt consolidation were more likely to be charged off (defaulted).

Business Driven Metrics

Loan Performance Metrics
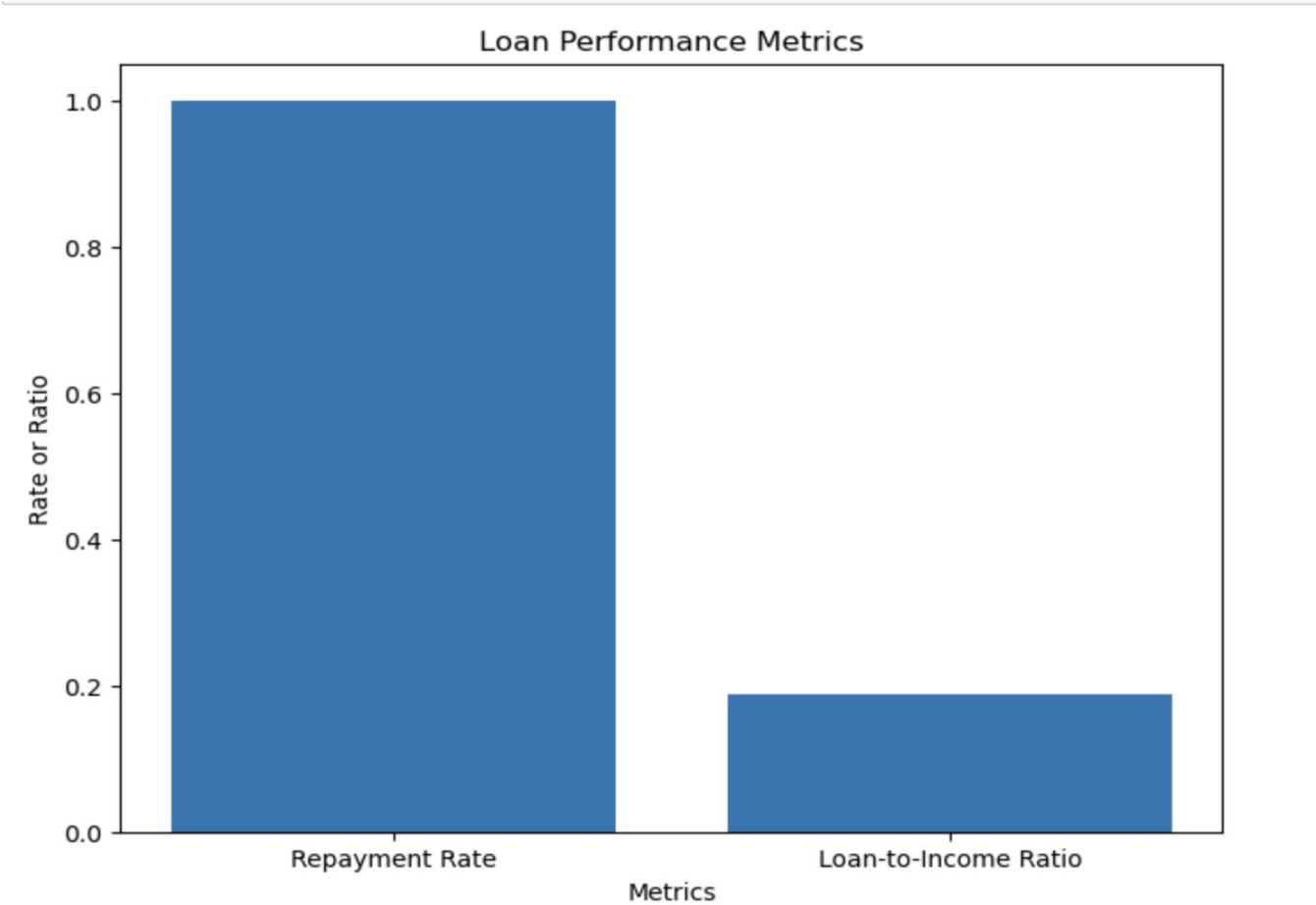
- Most borrowers repay their debts on time, and their debt levels are manageable in relation to their income.

Type Driven Metrics
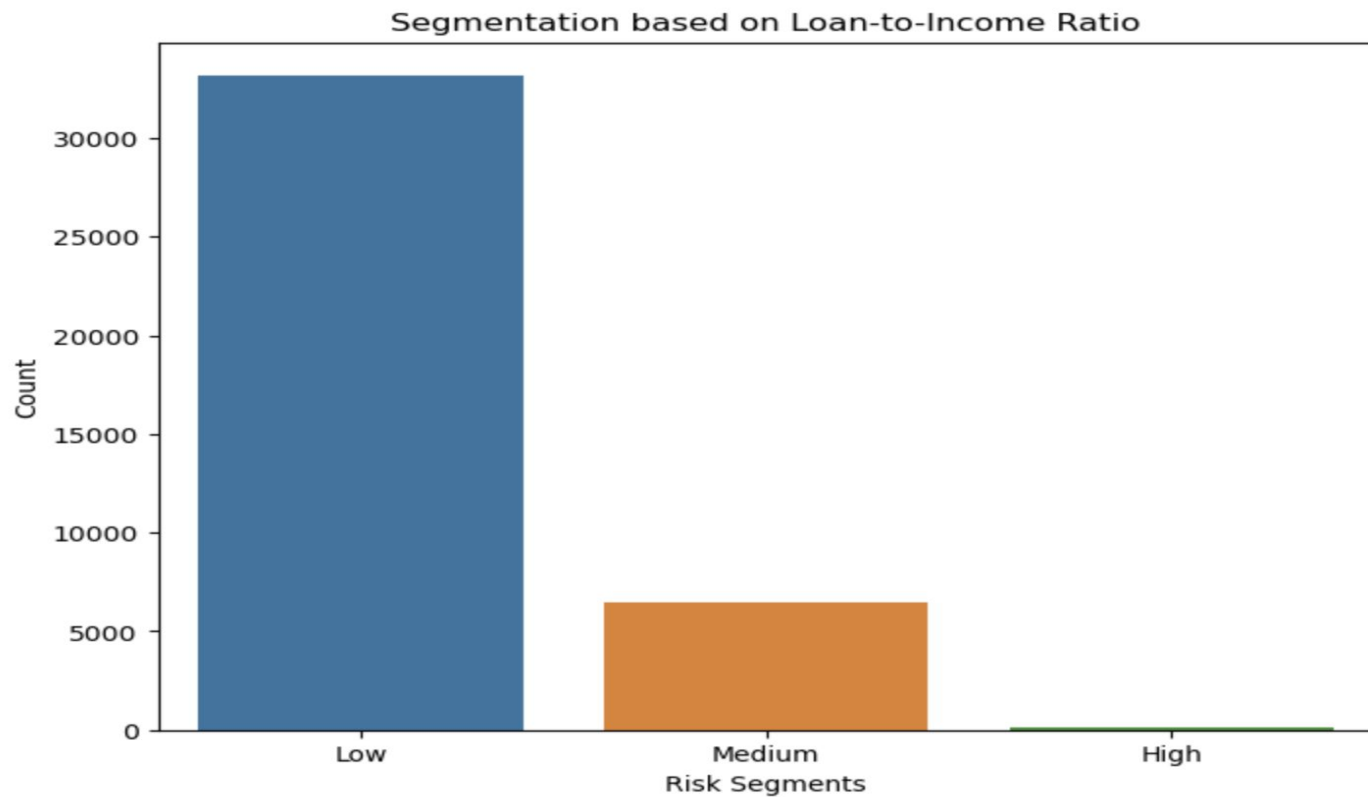
```
Name: count, dtype: int64
home_ownership
RENT            18899
MORTGAGE        17659
OWN              3058
OTHER              98
NONE                3
Name: count, dtype: int64
verification_status
Not Verified        16921
Verified            12809
Source Verified      9987
Name: count, dtype: int64
```

Data driven Metrics

Debt-to-Income Ratio Distribution

Segmentation
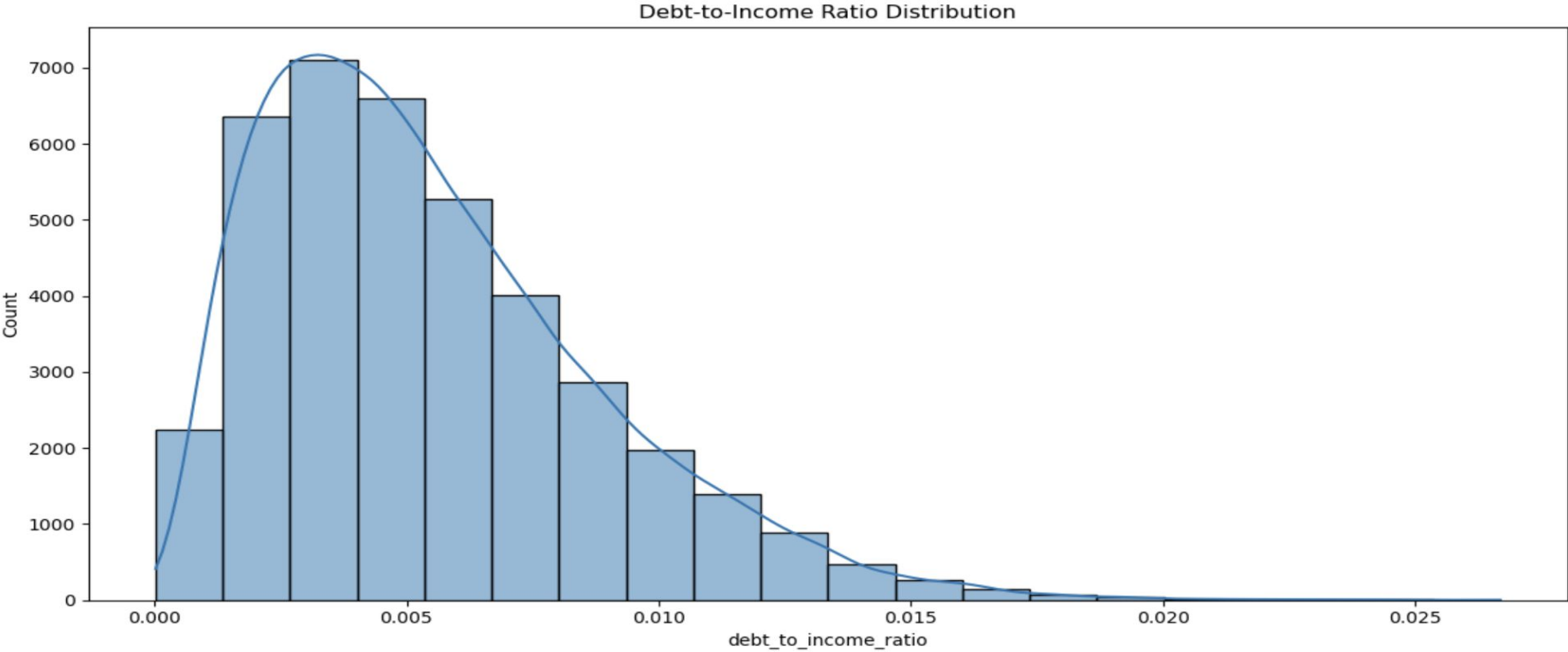
Segmentation based on Loan-to-Income Ratio

Loan-to-Income Ratio was segmented into Low, Medium, and High risk categories based on the loan amount relative to annual income.