Exploratory Data Analysis On Banking Data

A deep dive into data trends, distributions, and insights

Introduction

- Problem Statement

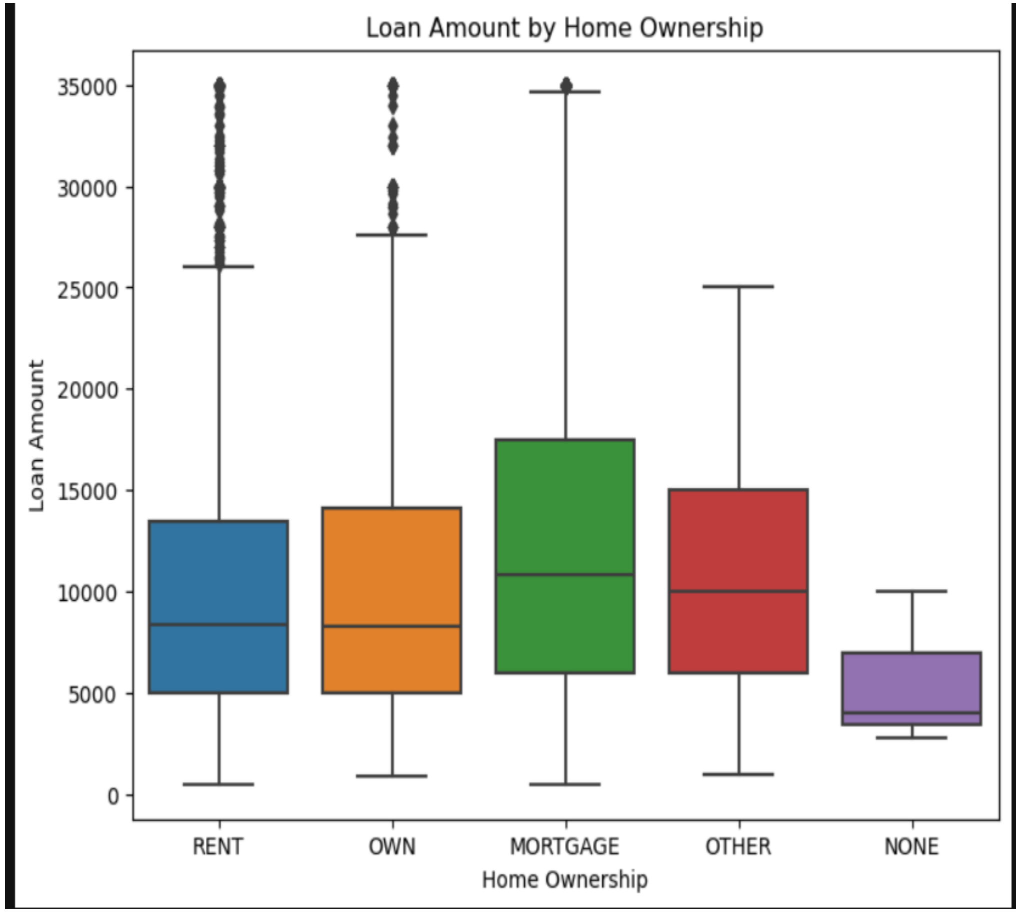
  To Understand the driving factors (or driver variables) behind loan default, i.e., the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

- Description of the dataset

  This dataset contains loan performance data, including information about borrowers and loan statues. The main goal is to explore how borrower characteristics influence loan outcomes(Fully Paid, Charged Off, Current)

Analysis

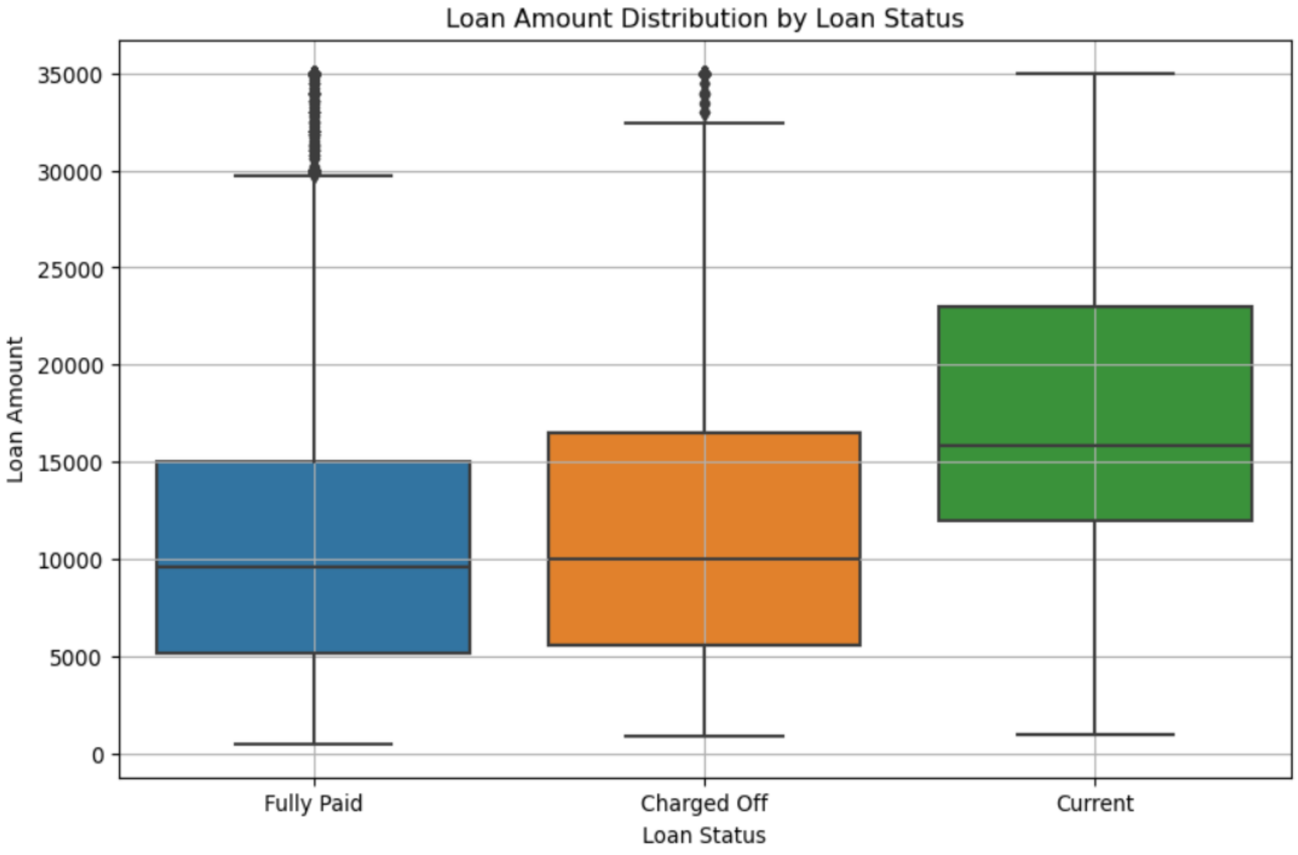
- Graph for Loan Amount and Homeownership



- Mortgage holders tend to take out the largest loans compared to other categories, with a larger spread of loan amounts and many high outliers.
- Renters and those who own their homes tend to take similar-sized loans, with the median around 10,000.
- People who have no home ownership (NONE) take out the smallest loans, with very little variation.

Analysis

- Loan Amount Distribution by Loan Status

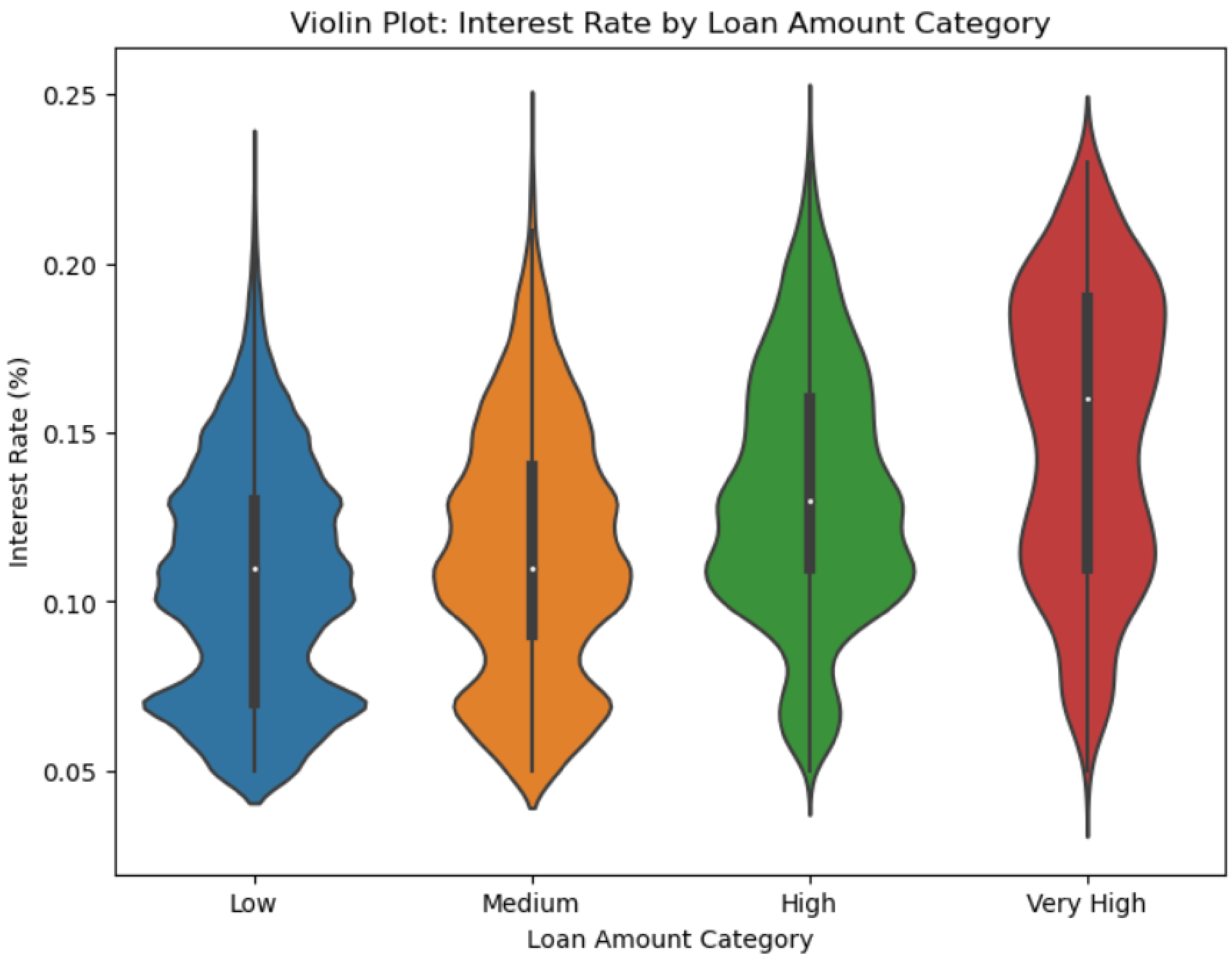


Question: Does a higher loan amount increase the probability of default?

From this, we can refer that higher loan amounts are not strictly correlated with a higher probability of default(charged off). Instead, the defaulted loans(charged off) seems of fall between fully paid and current loans in term of loan amounts.

Analysis

- Visualize the Correlation Between loan_amnt and interest rate



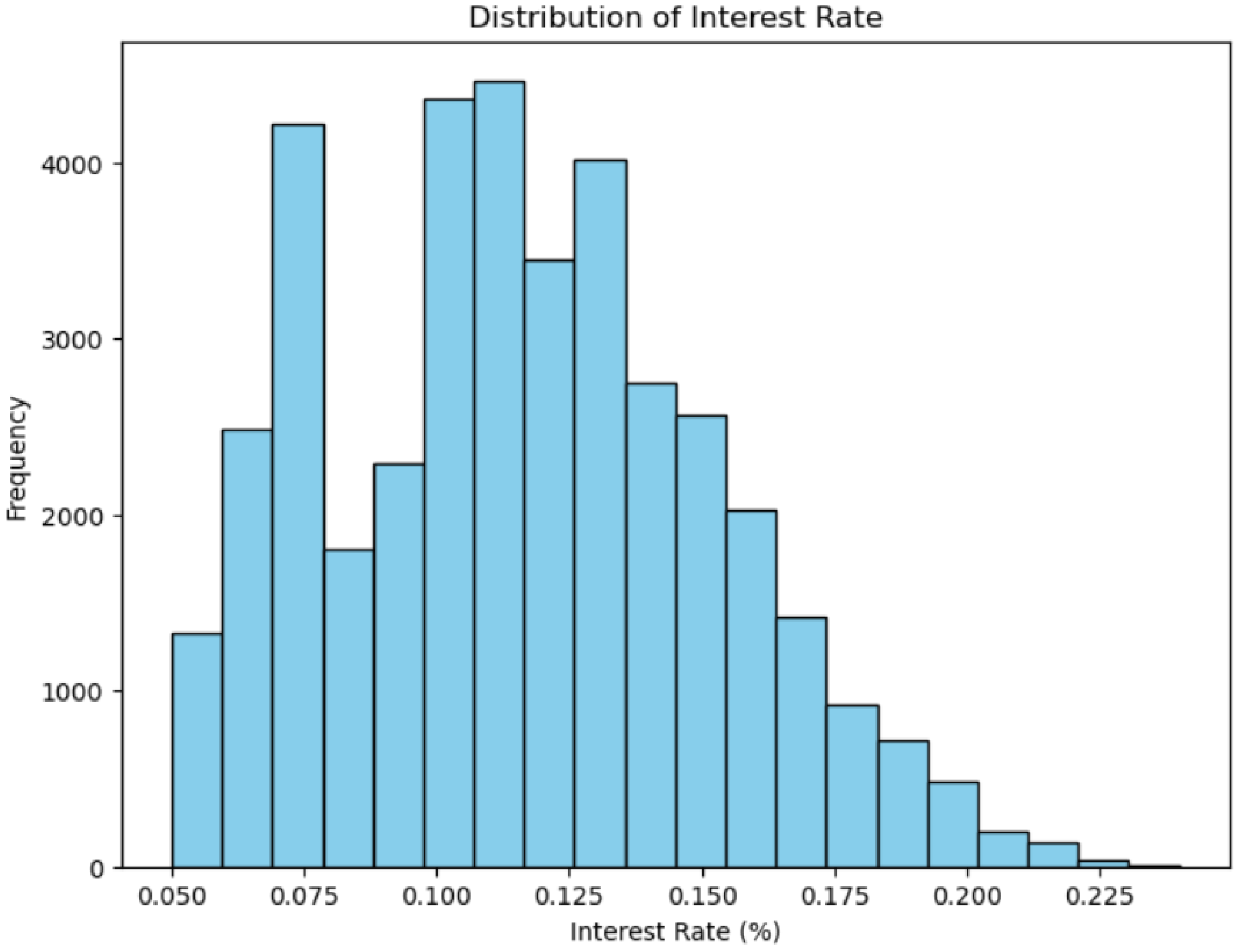
Question**:** Are larger loan typically gives higher interest rate?

Interest rate seems to be fairly consistent across all loan amount categories, with no strong upward or downward trend. The most common interest rate hover around 10-12%, regardless of loan amount.

There is slightly more variability in interest rates for loans in the low and very high loan categories, but the difference are not dramatic.

Analysis

- Interest rate and Term Analysis

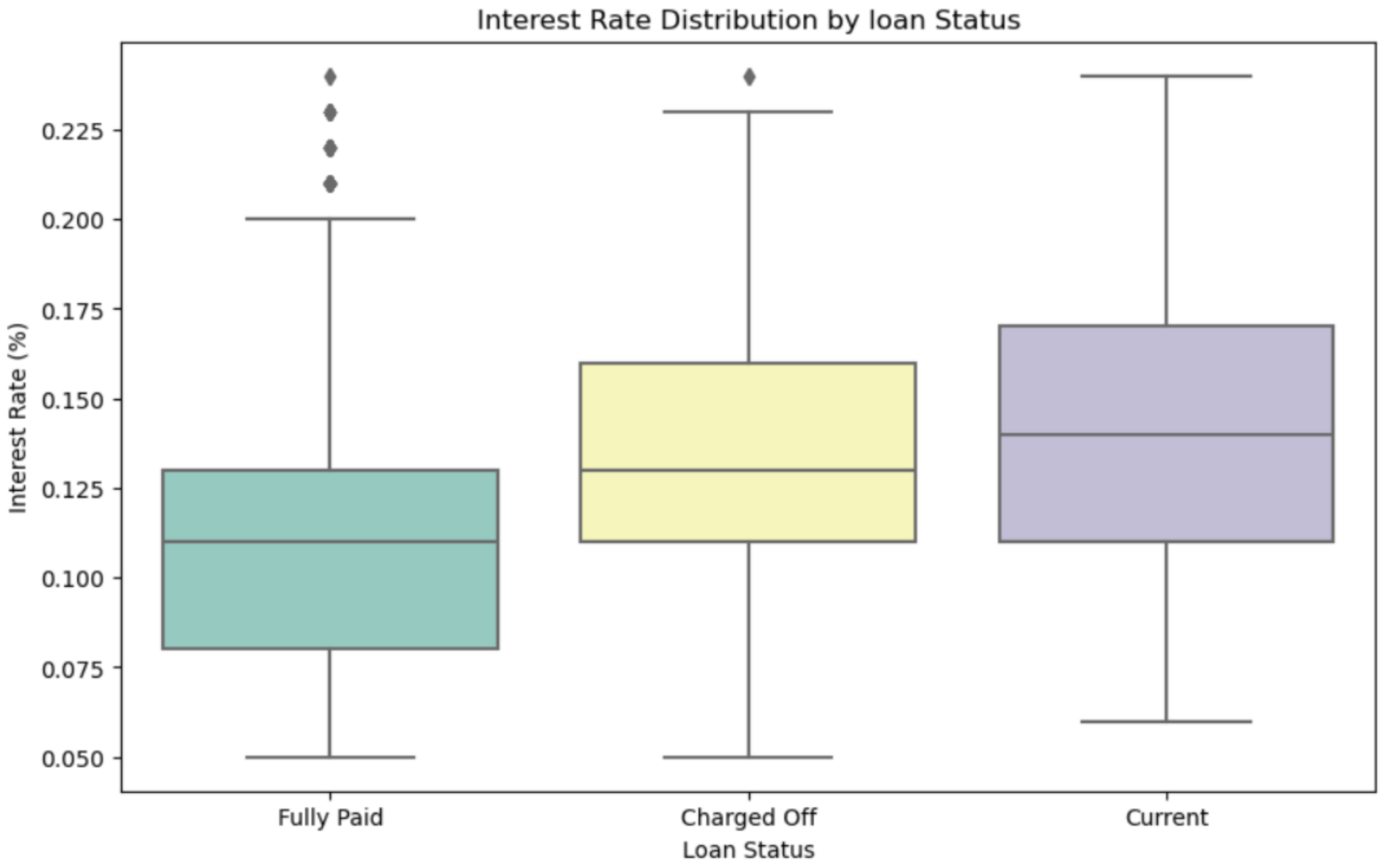


Question**:** Are there significant variation in Interest rate

Yes, There are significant variations in interest rates, with a large portion of loans being distributed around 7.5% to 15%, but with a long tail showing rates upto 22.5%. This is variability might be driven by factor as risk assessment, loan duration, or borrower creaditworthiness.

Analysis

- Interest rate and Loan status



Question**:** Are loans with higher interest rate more likely to default?

Yes, we can see that the average interest rate for charged off is higher then the interest rate of fully paid or current loans. This suggest that loans with higher interest rates are more likely to default.

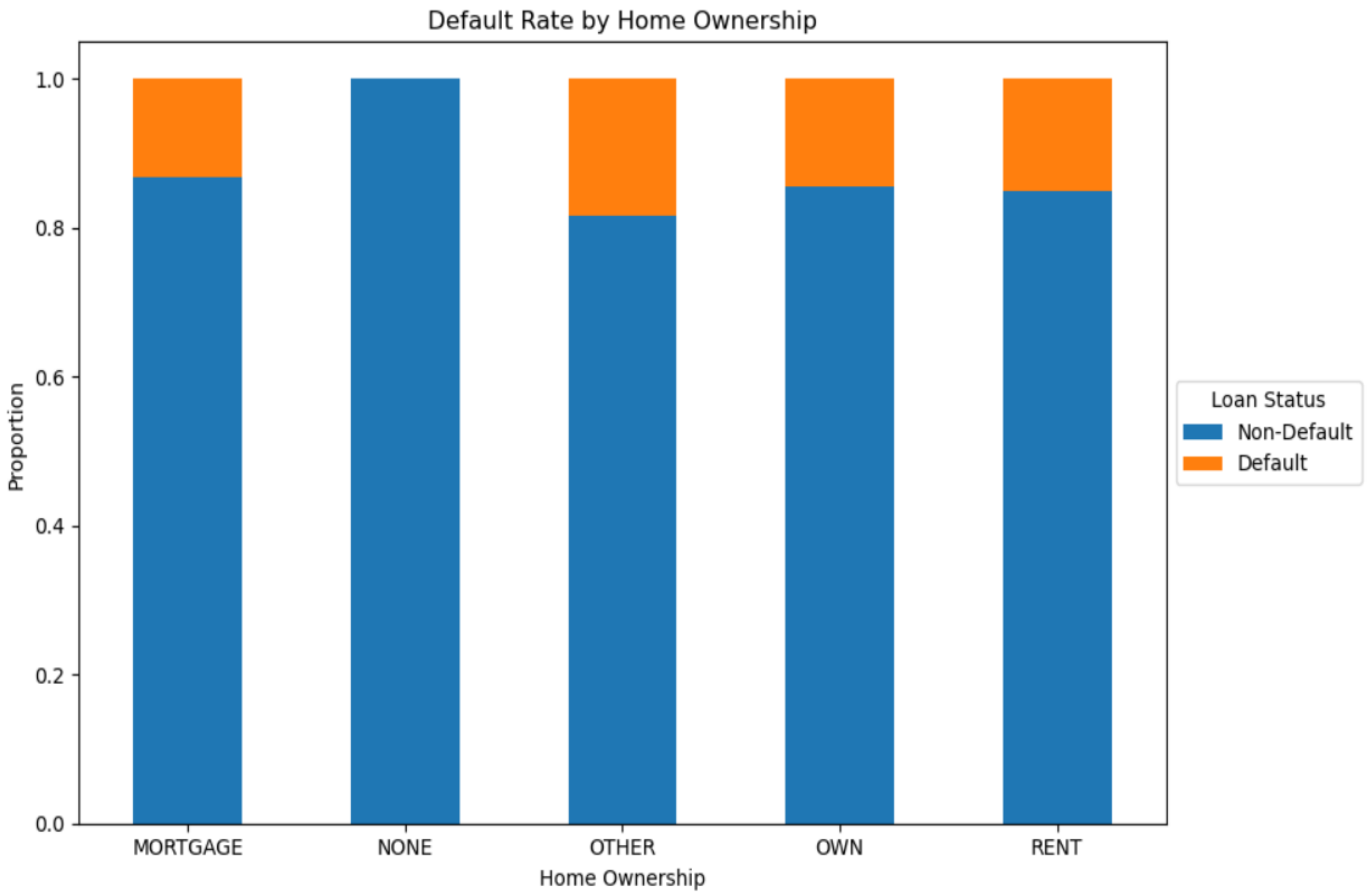
Analysis

- Default rate by employment length

We can see the most defaulters are from shorter employment length.

Analysis

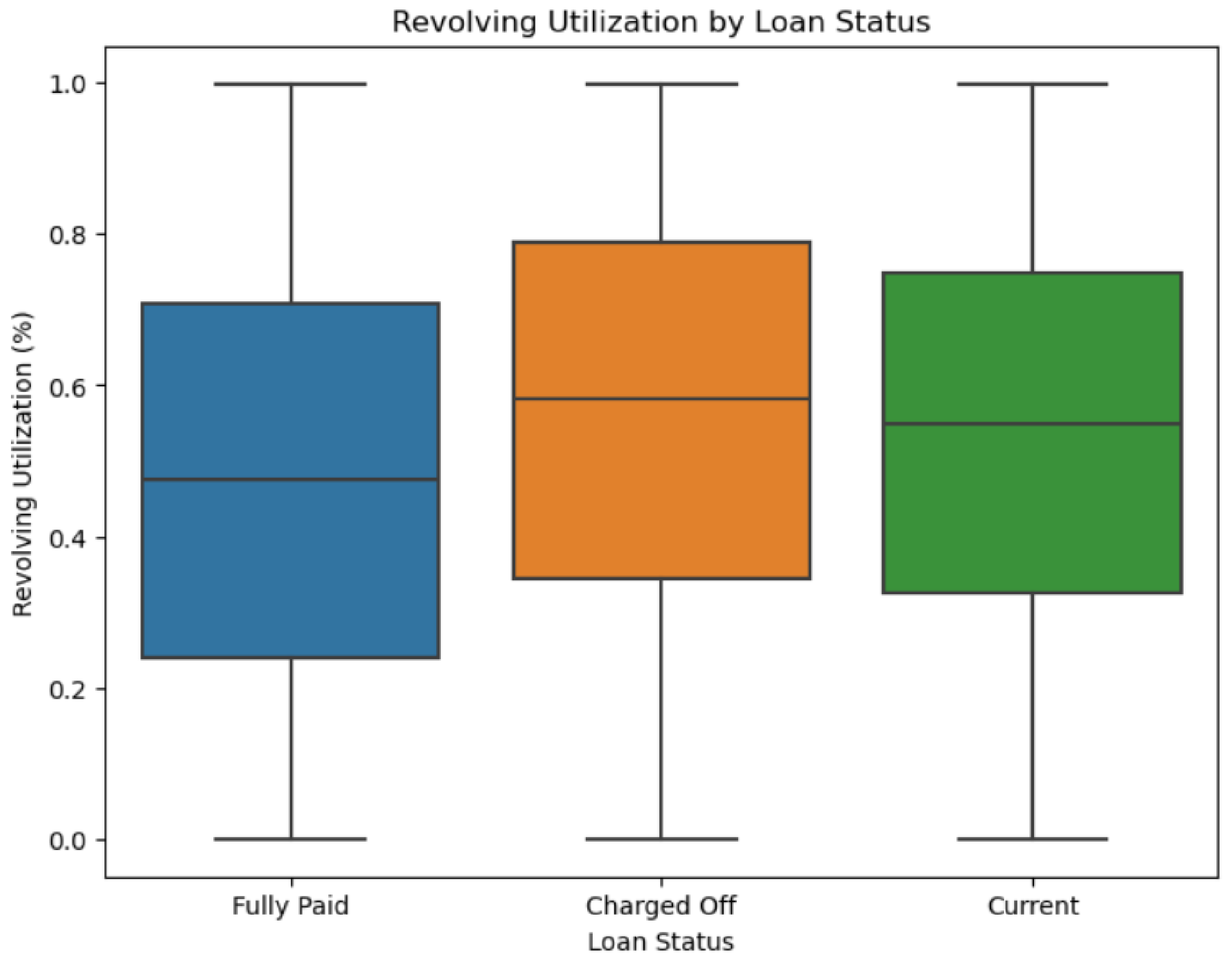
- Default rate by Home Ownership



The orange section representing defaults shows that a higher proportion of people who rent homes and other default on their loans compared to the other categories. Renters and other seem to have a slightly higher risk of default than those who own or have mortgages.

Analysis

- Default Rate by Revolving Line Utilization Rate

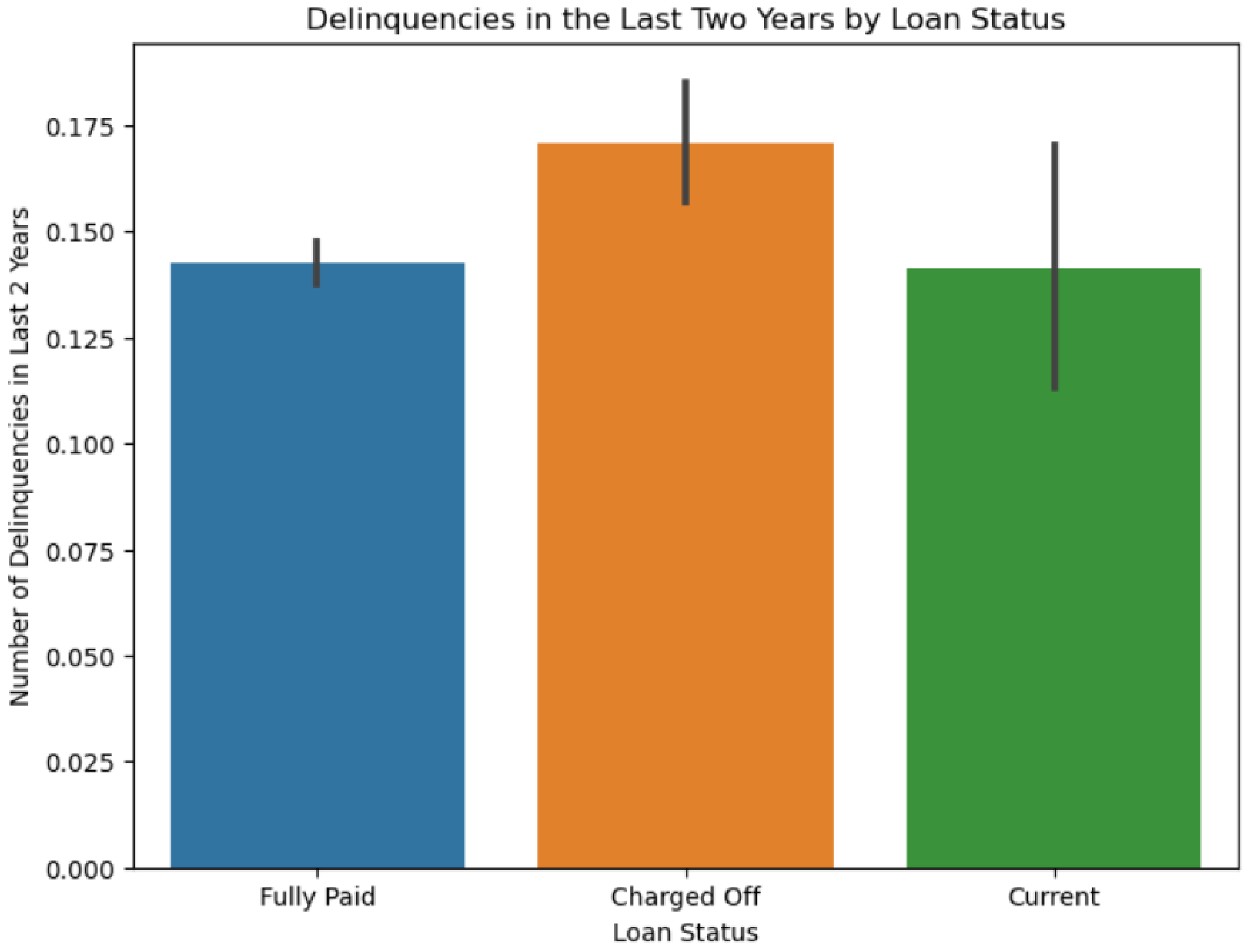


The median revolving utilization for borrowers with a "Charged Off" status (defaulted loans) is higher than that for those who are "Fully Paid" or "Current."

Borrowers who defaulted tend to have a higher credit utilization on average, as indicated by the elevated median and the interquartile range for the "Charged Off" group.

Analysis

- Default Rate by Delinquency in Past 2 Years



The plot shows the number of delinquencies in the last 2 years, grouped by loan status. From the plot it appears that borrowers with recent delinquencies are indeed more likely to be Default

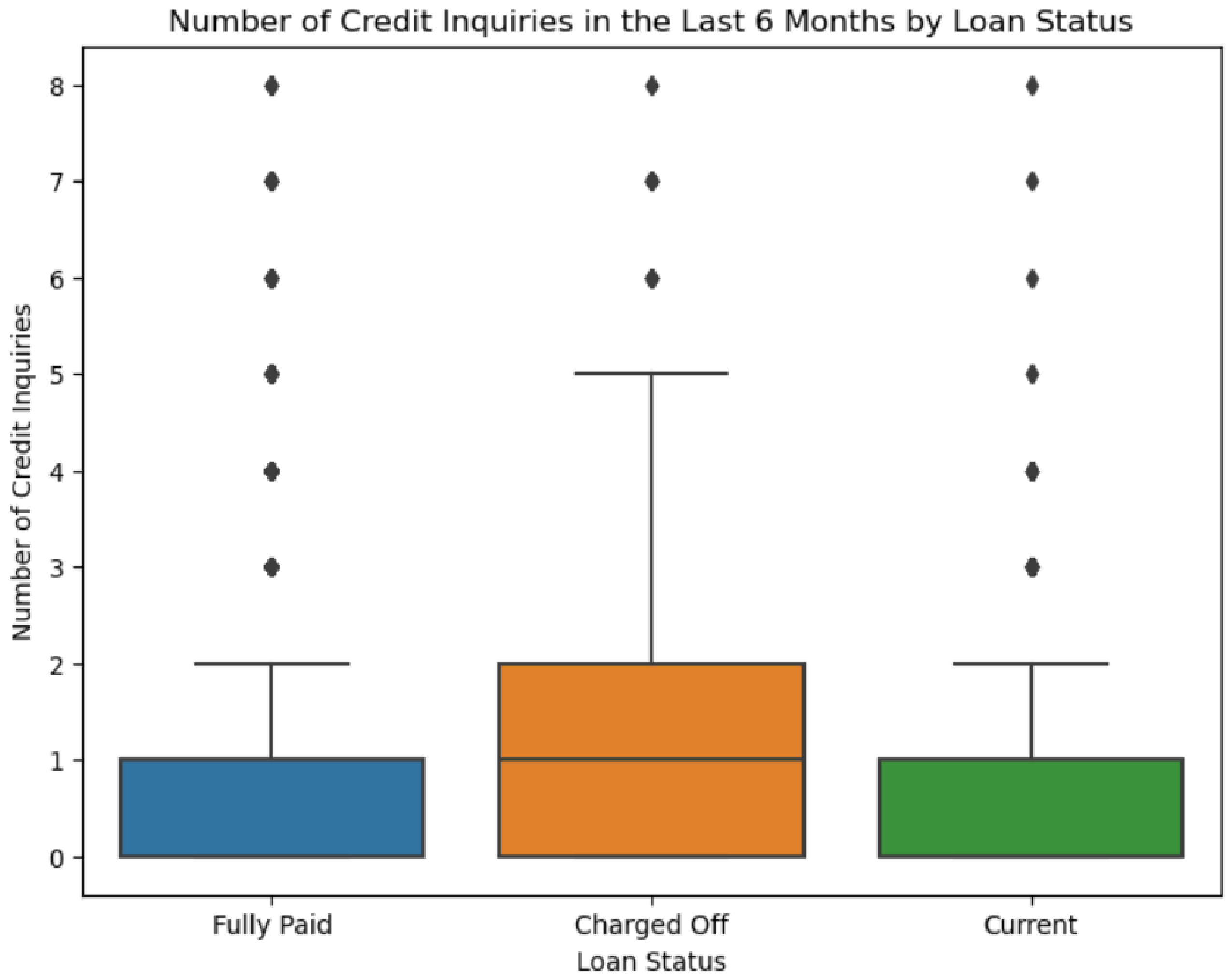
In the Charged Off category, there are multiple borrowers with higher number of delinquencies.

Borrowers in the Fully Paid category tend to have zero or very few delinquencies.

This suggest that borrowers with higher delinquencies are more likely to default compared to those who fully pay their loans.

Analysis

- Default Rate by Credit Inquiries in Past 6 Months

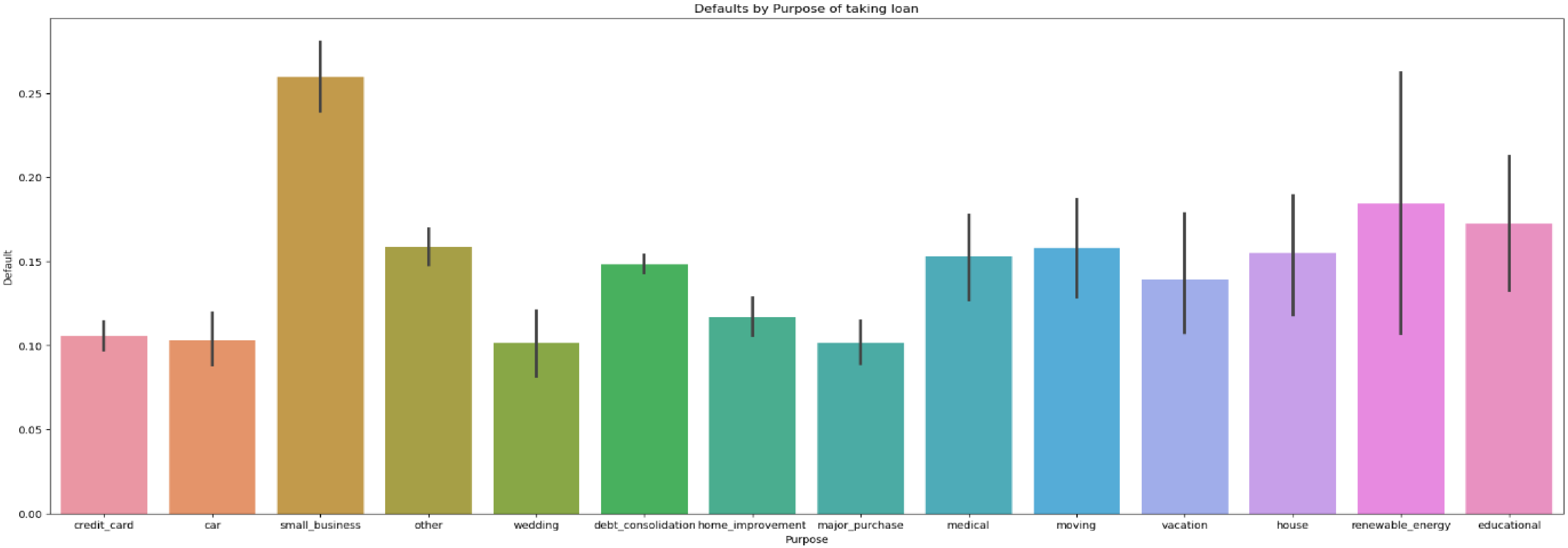


Charged Off loans (loans that went into default) may correlate with borrowers having a higher utilization rate. These borrowers are more likely to have more inquiries, indicating they are seeking additional credit, which can be a warning sign of financial distress

Fully Paid and Current loans: These borrowers tend to have fewer inquiries, suggesting a lower likelihood of financial distress, and often lower credit

Analysis

- Default Rate by Purpose of Taking Loan

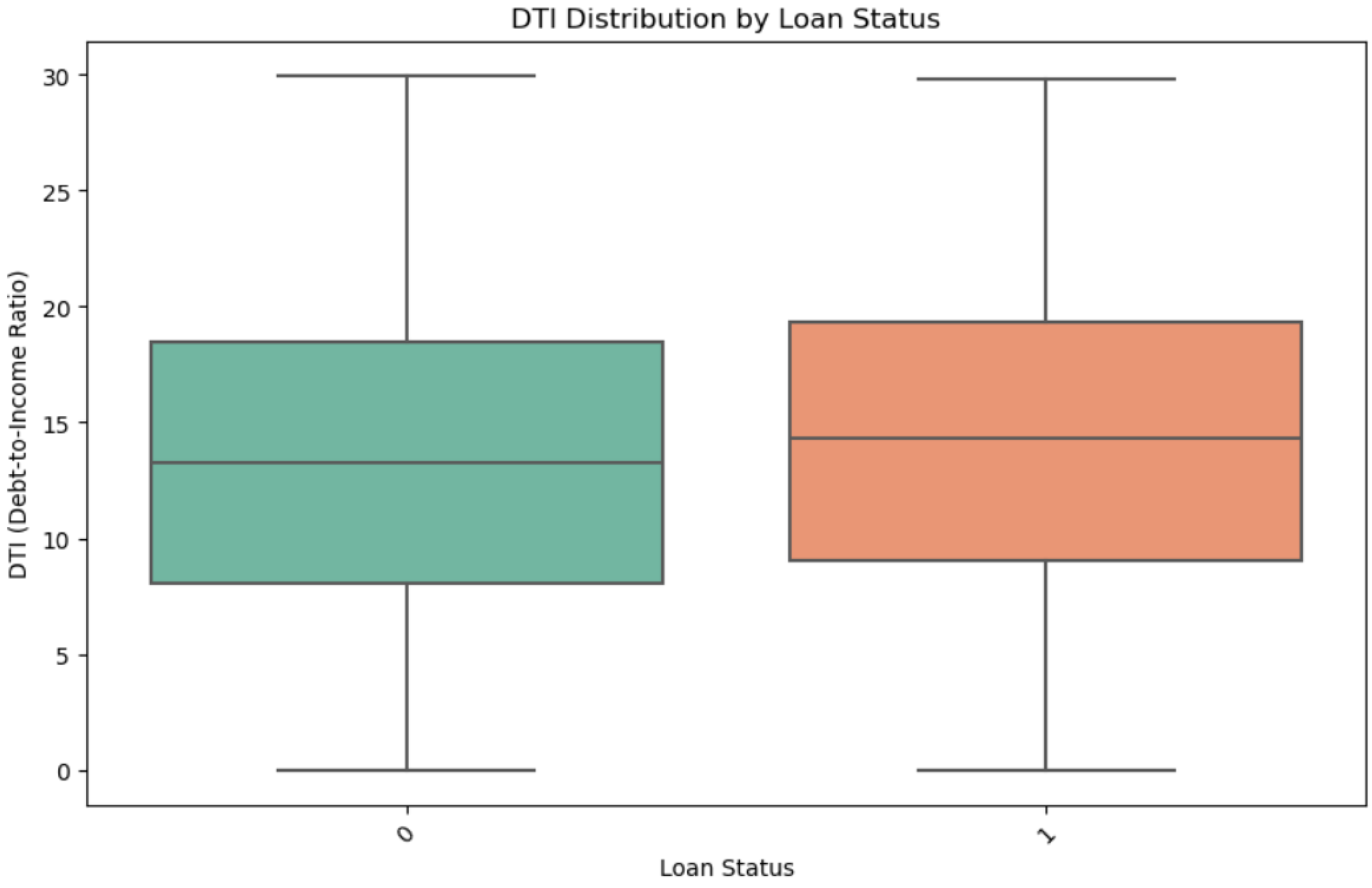


Loans for **small business** have the highest default rate, followed by **educational** loans.

The graph suggests that lenders should exercise caution when approving loans for small business and educational purposes, as these categories have a higher likelihood of default.

Analysis

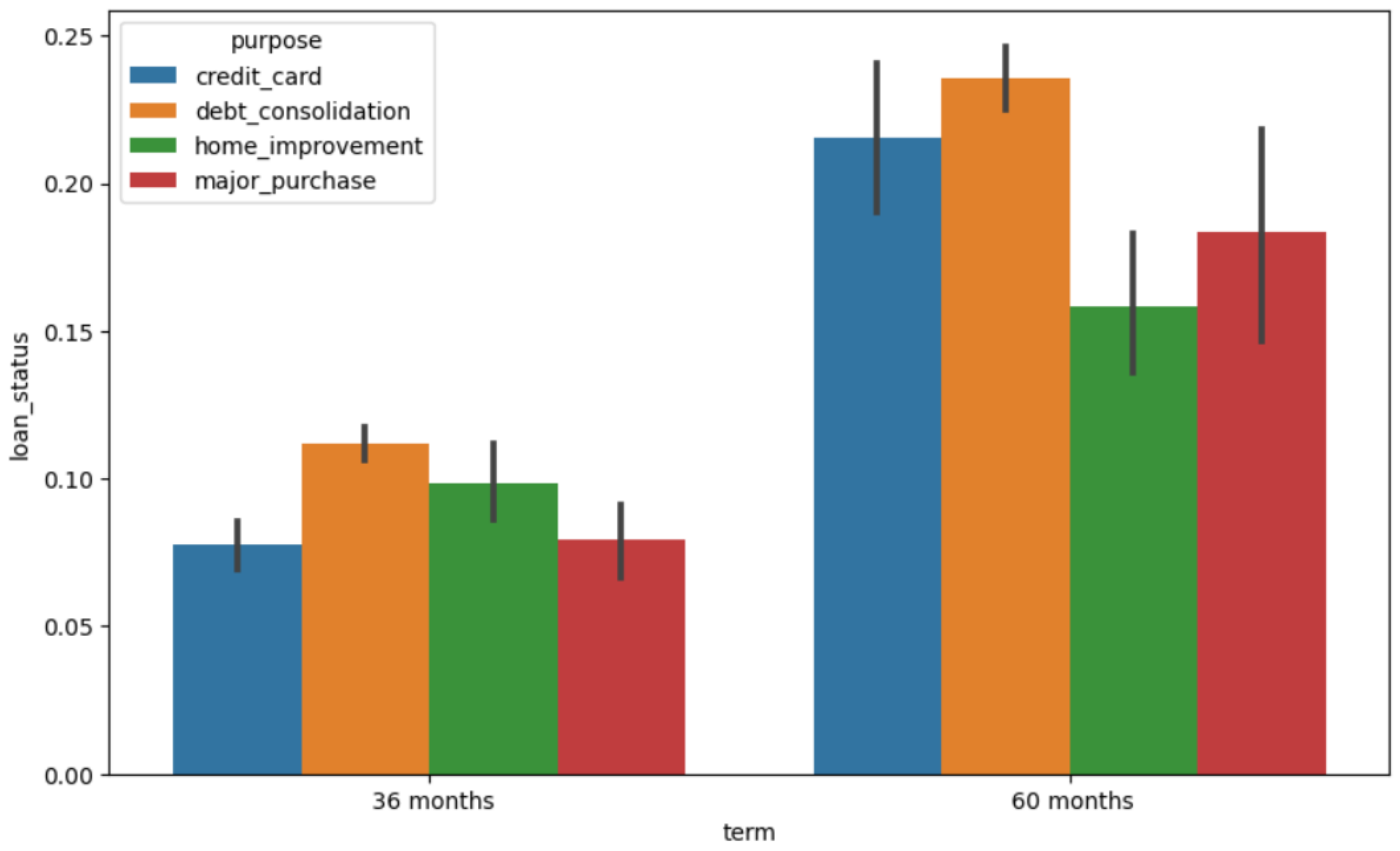
- DTI Distribution



The median DTI for the 0 (non-default) group is slightly lower than the median for the 1 (default) group. This suggests that borrowers with lower DTIs are generally less likely to default.

Analysis

- Loan Status by Top loan amount purpose



For the long tenure we can see the debt consolidation is high.

Conclusion

- Mortgage holders tend to take larger loans compared to renters or those who own their homes. Renters and homeowners have similar loan size, while people without home ownership (category “NONE”) take the smallest loans.
- Higher loan amounts do not necessarily correlate with a higher default probability. Defaulted loan fall in between fully paid and current loans in terms of loan amount, indicating that other factors play a stronger role in defaults.
- The interest rate remains consistent across loan amount, typically between 10-12%. There is more variability at the extremes of loan amounts, but the trend does not show a significant correlation between loan size and interest rate.
- There are significant variations in interest rate with most loan distributed around 7.5-15%. The long tail up to 22.5% suggests interest rates distributed by factors such as loan term, borrower creditworthiness, and risk assessment.
- Loan with higher interest rate are more likely to default, as evidenced by the higher average interest rate of charged off loans compared to fully paid or current loans.
- A higher proportion of defaults across among borrowers with shorter employment lengths, indicating that job stability is a significant factors in loan repayment ability.
- Renters and those classified as “other” show a higher risk of default compared to homeowners or mortage holders. This could be due to less financial stability among renters.

Conclusion

- Borrowers with higher credit utilization rates are more likely to default. The median revolving utilization is higher for defaulted loans, which suggests that borrowers with heavy credit usage may struggle with loan repayments.
- Borrowers with more delinquencies in the past two years are more likely to default. The analysis shows a clear relationship between recent delinquencies and the likelihood of default.
- Borrowers with more credit inquiries in the past six months are more likely to default, as this may indicates financial distress or aggressive seeking of a additional credit.
- Loans taken for small business and educational purpose have the highest default rates, suggesting that these loan purpose are riskier to lenders.
- A higher DTI ratio correlates with a higher likelihood of default, implying that borrowers with high debt relative to their income are at greater risk of defaulting on loans.

Recommendations

- **Risk Assessment for High-Interest Loans:**

  Since loans with higher interest rates are more likely to default, it is essential to strengthen the risk assessment process for borrowers who are offered high-interest rates. Lenders should review borrowers' creditworthiness, employment stability, and financial history more thoroughly before approving high-interest loans.

- **Targeted Loan Policies for Renters and Low-Employment Borrowers:**

  Renters and borrowers with shorter employment lengths are at higher risk of default. Loan policies could be adapted to include stricter eligibility criteria for these groups or provide financial literacy programs to mitigate the risks.

- **Credit Utilization Monitoring:**

  Given that higher credit utilization is linked to defaults, lenders should monitor borrowers' revolving credit usage closely. Implementing automated alerts or early warning systems when utilization crosses certain thresholds could help prevent defaults.

Recommendations

- **Delinquency History as a Key Indicator:**

  Recent delinquencies are a strong predictor of default. Lenders should give significant weight to delinquency history during loan approval, possibly rejecting or setting higher thresholds for those with multiple delinquencies.

- **Additional Scrutiny for High-Risk Loan Purposes:**

  Loans for small business and educational purposes have higher default rates. Lenders should introduce tighter risk evaluation processes for these types of loans or offer them under stricter terms, such as requiring higher down payments or collateral..

- **Debt-to-Income Ratio Guidelines:**

  Loans should be offered more cautiously to individuals with high DTIs. Implementing DTI caps based on loan sizes or offering lower loan amounts to borrowers with high DTIs could reduce the risk of defaults.

Recommendations

- **Credit Inquiry Patterns:**

  Lenders should be wary of borrowers with frequent credit inquiries, as this may signal financial instability. Limiting loan amounts or offering more secure loans (such as secured loans) to these individuals might mitigate default risks.

- **Segmented Loan Pricing Strategies:**

  Interest Rate Adjustments: Since higher interest rates correlate with a higher likelihood of default, lenders should consider **tiered pricing strategies** based on borrower risk profiles. Borrowers with lower default risk (better credit scores, longer employment history, lower DTI ratios) could be offered reduced interest rates, while those with higher risk profiles should receive rates that compensate for the increased default likelihood.

- **Employment Verification and Job Stability Metrics:**

  - Given that borrowers with shorter employment lengths are more prone to default, employment verification should not only confirm the job title but also assess the **stability of the industry** and the borrower's job type (contract vs permanent). Introducing **employment length minimums** or developing a **job stability scoring metric** can enhance loan approval decisions.

Recommendations

- **Homeownership-Based Incentives:**

  Homeownership is associated with lower default risk, particularly for mortgage holders. Lenders should consider creating **incentives or lower interest rates for homeowners**, as they are more financially stable. This could involve designing loan products tailored for homeowners, such as **home equity-based loans** or **preferential terms for those with mortgage histories**.

- **Personalized Loan Repayment Plans:**

  For borrowers at risk of default (e.g., those with high DTI ratios or high revolving utilization), lenders could offer personalized loan repayment plans that include income-based repayments or flexible installment options. This strategy could reduce the pressure on borrowers and help them stay on track with payments.

- **High Utilization Rate Intervention:**

  Borrowers with high revolving credit utilization are more likely to default. To counter this, lenders could offer **credit line reductions** or **limit increases** on revolving credit based on repayment performance. Alternatively, they could offer **consolidation loans** to help borrowers manage and pay down their debt at a lower interest rate..