Lending Club Case Study

upGrad

Lending club is one of the largest online loan marketplace, facilitating personal, business, and medical loans. Our mission is to help borrowers access lower interest rates quickly. However, we encounter significant credit losses from risky applicants who default on their loans.

This case study aims to identify the key factors that indicate loan default, allowing us to reduce credit loss and improve our portfolio and risk assessment strategies.

To enhance our understanding, we will explore risk analytics and the importance of various indicators.

Group Members:

Rahul Srivastava

Raj Srivastava

Approach to Addressing Loan Default Risk

Data understanding-

Working with the Data Dictionary and getting knowledge of all the columns and their domain specific uses

Data Cleaning and Manipulation –

Data quality issues are addressed in the right way (missing value imputation, outlier treatment and other kinds of data redundancies, etc.).

Data analysis –

Analyze consumer and loan attributes.

Identify patterns and correlations that indicate potential defaults.

Provide insights that can guide the company's lending strategies

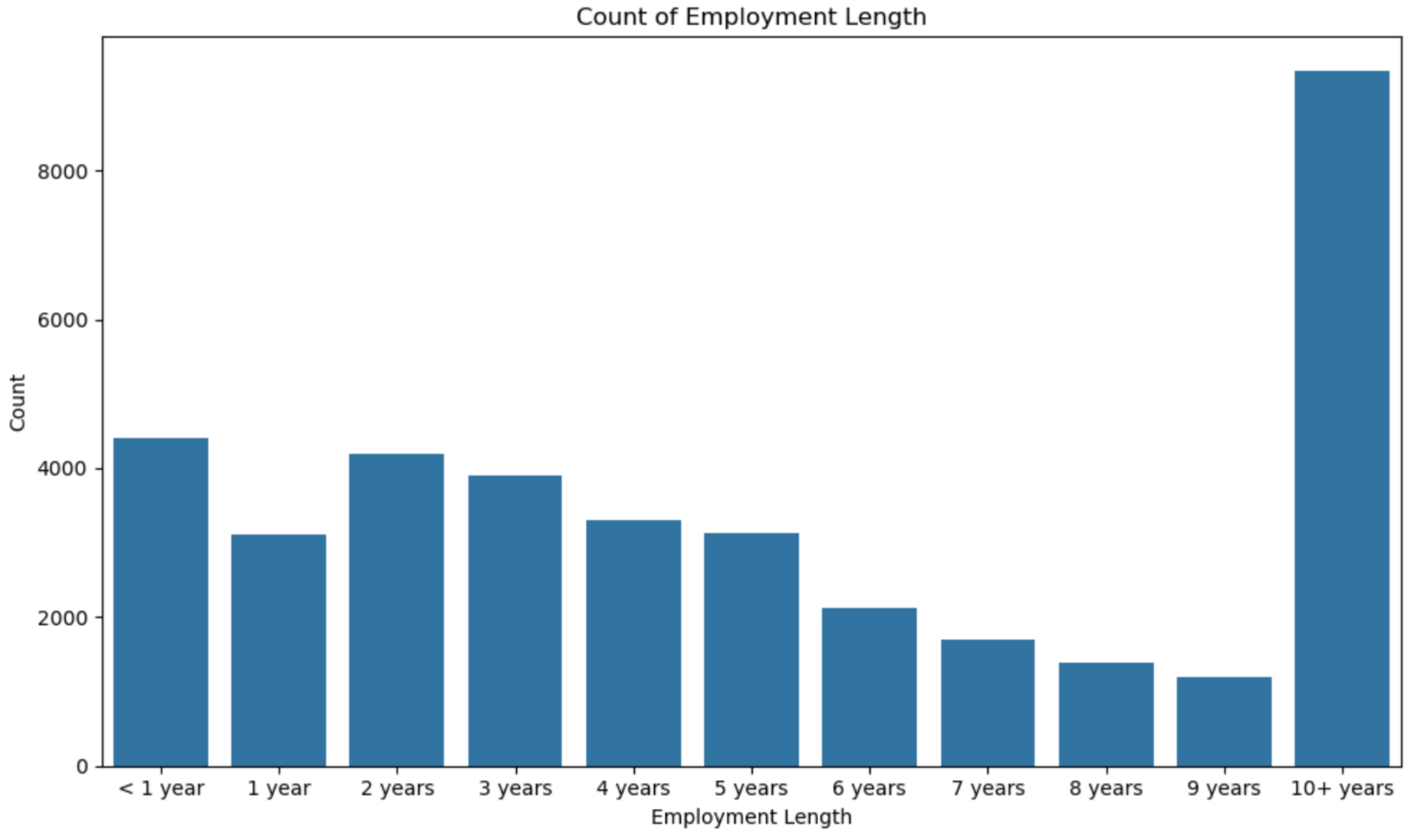
Recommendation –

Analysing all plots and recommendations for reducing the loss of business by detecting driving factors which contribute to loan defaulters.

Univariate Analysis

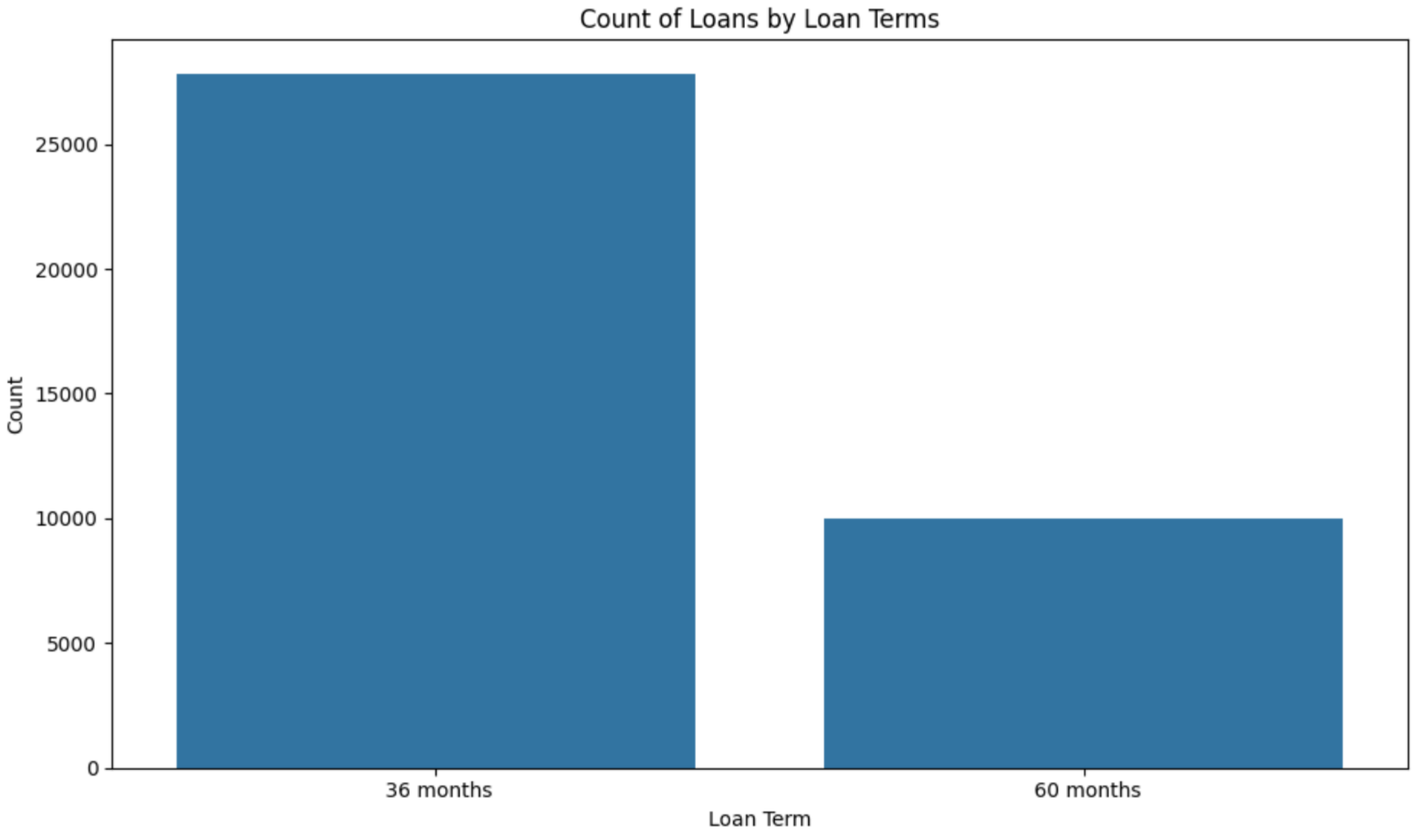
Loans with Employment Length

- We can clearly see that users with over 10 years of experience have taken significantly more loans compared to employees with less experience.



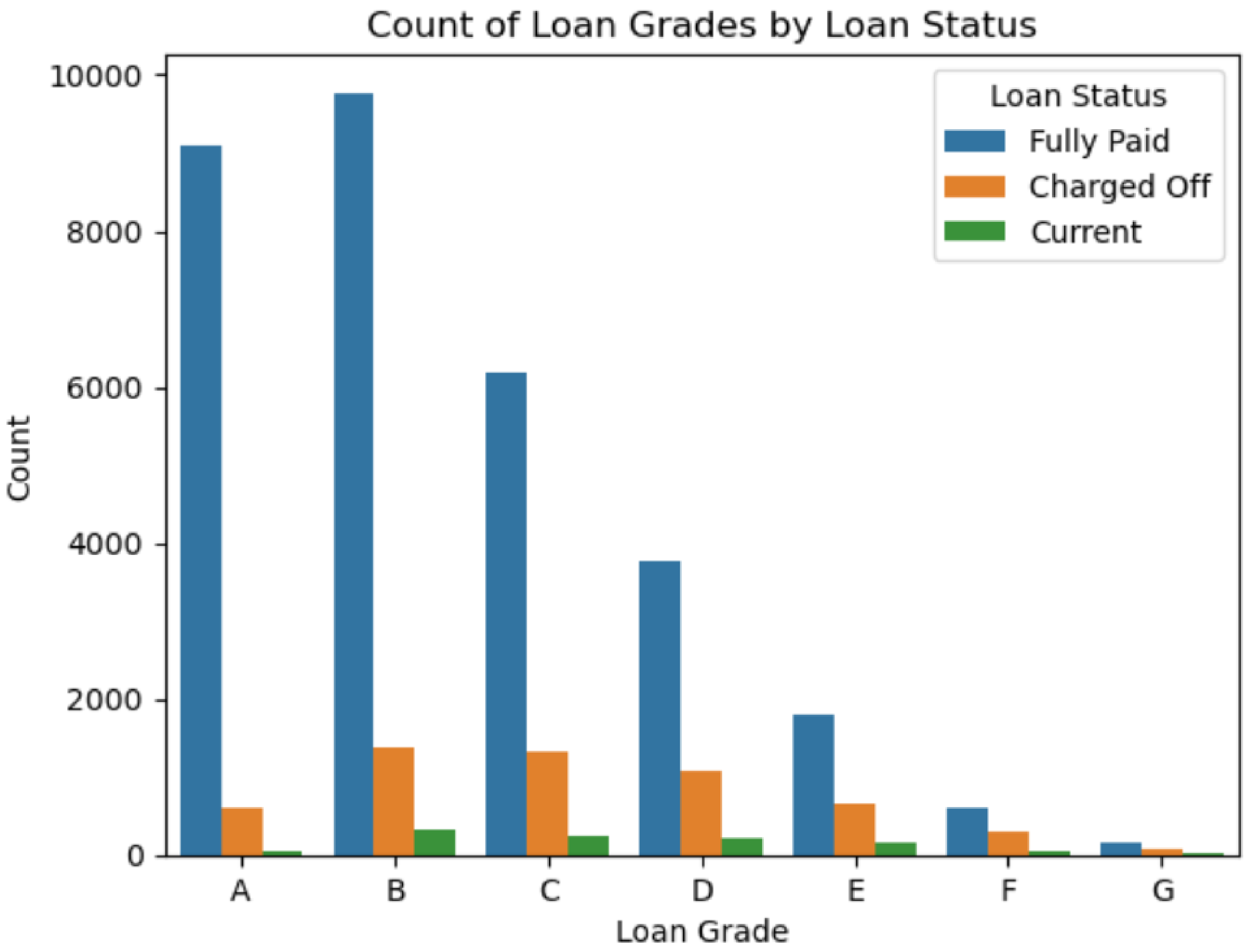
Loans by Tenure

- The graph clearly shows that more people prefer the 36-month loan over the 60-month loan.



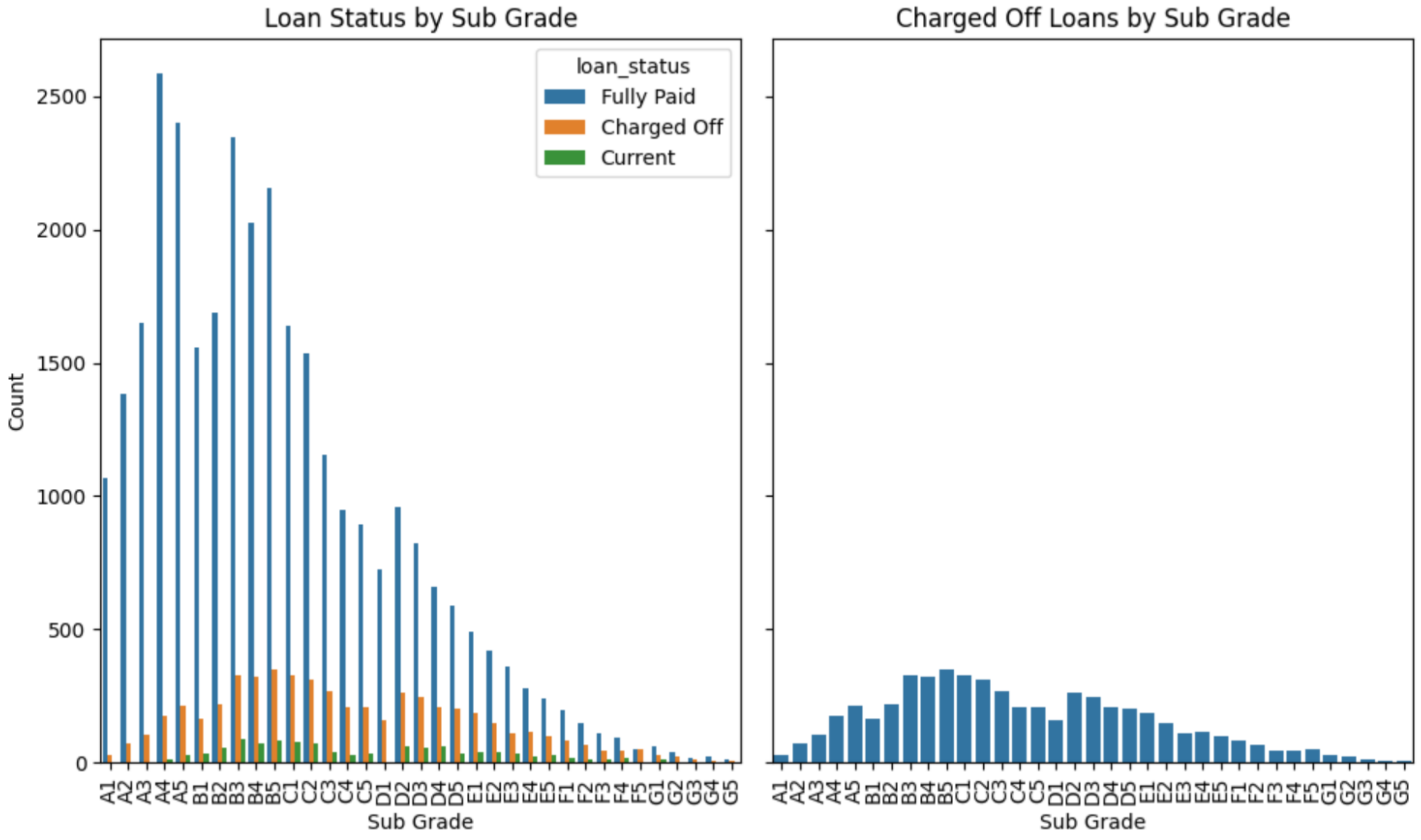
Loans by Grade and Status

- The graph shows the distribution of loans based on their grade, along with the distribution of loan status for each grade.
- A-grade and B-grade loans are more preferable, as they are considered more secure.



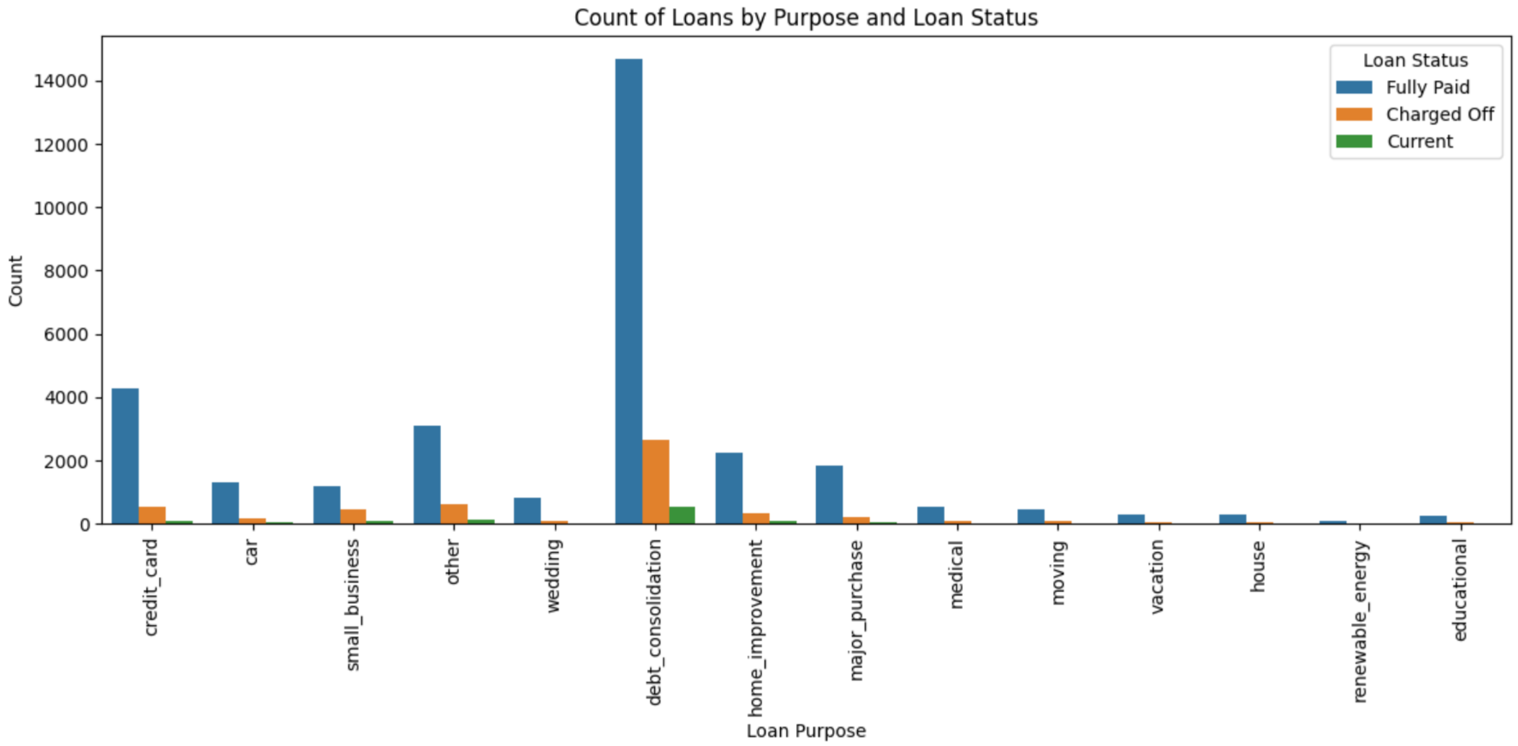
Loans by Sub Grade and Status

- The left graph shows the absolute distribution of loans in each subgrade.
- The right graph shows the absolute distribution of default loans in each subgrade.



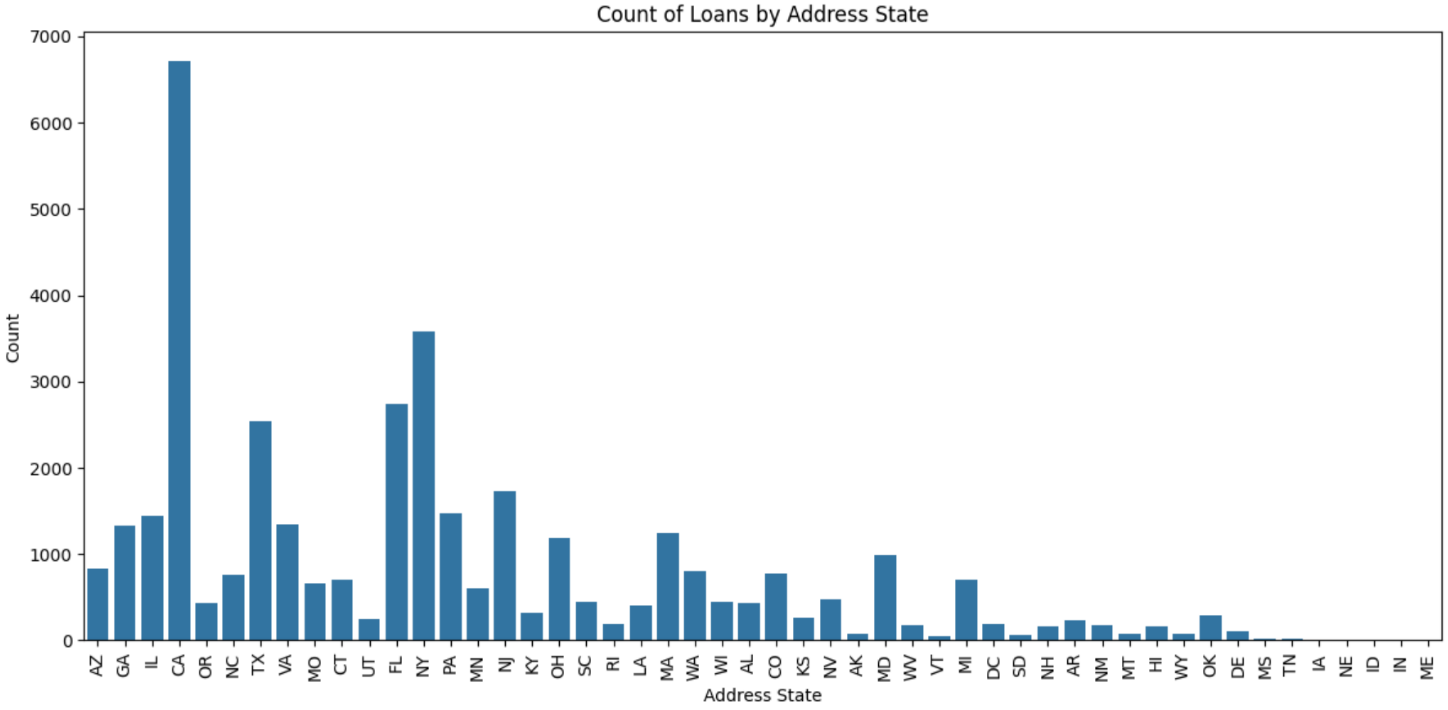
Loans by Purpose and Status

- Most loans are taken out for either debt consolidation or credit card repayments.



Loans by Address State

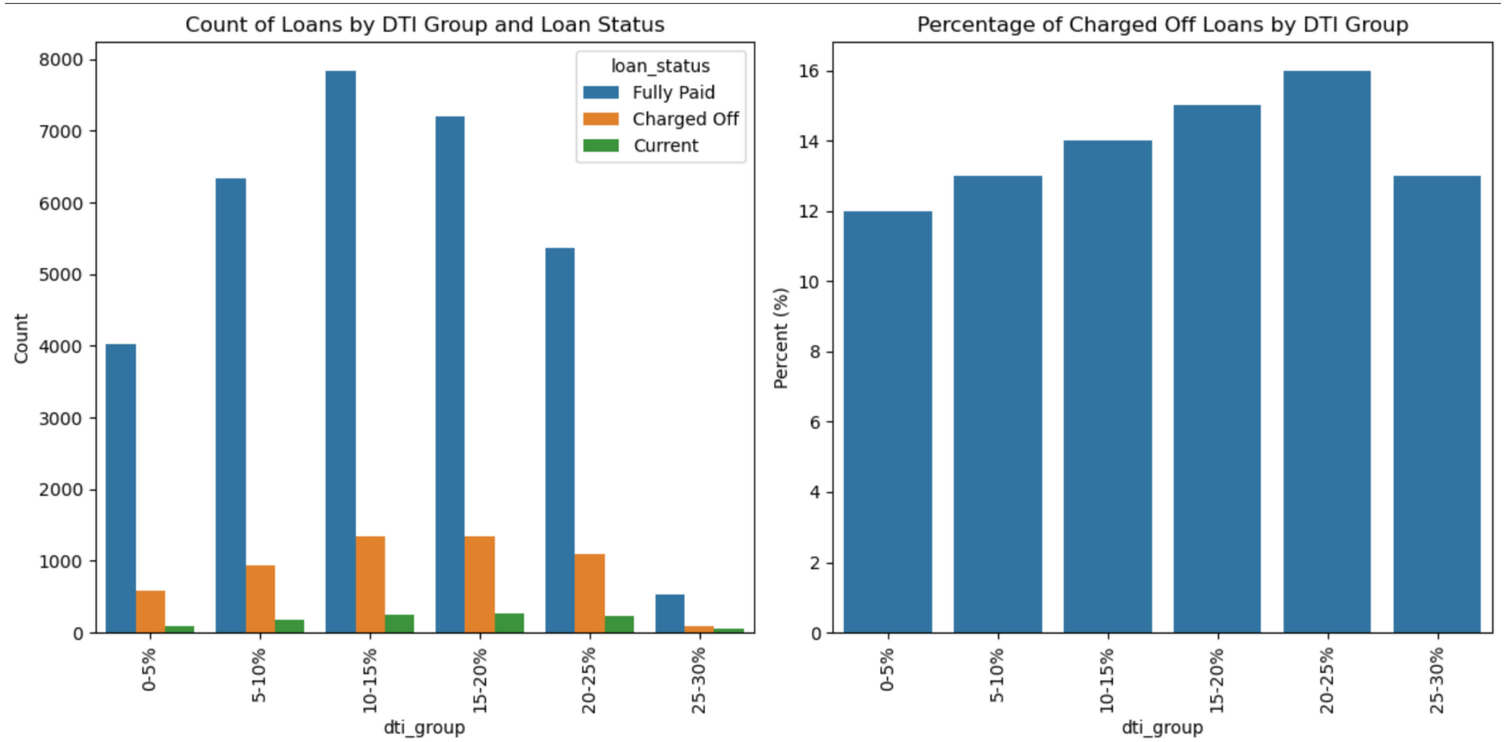
- It is clearly visible in the graph that most borrowers are from California (CA).



Bivariate Analysis

Loans by DTI group and Status

- We have analyzed the count of loans in relation to the Debt-to-Income (DTI) ratio, along with loan status.
- It is clear that the probability of default increases as the DTI ratio rises, as shown in the subplot representing the percentage of charged-off loans.



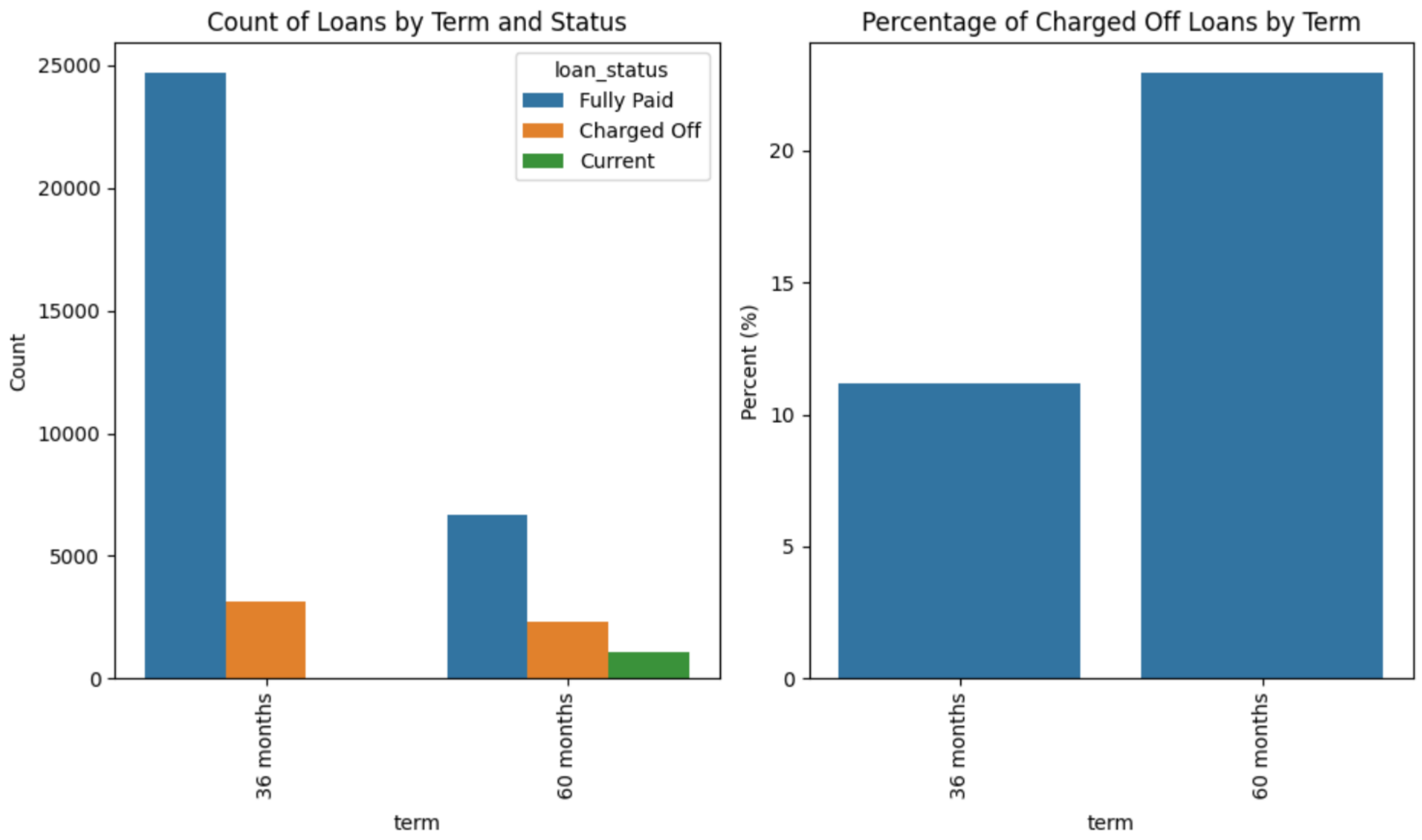
Loans by Tenure and Status

- We have plotted the count of loans with respect to different loan tenures.

- The graph shows the distribution of loans based on loan status and loan term.

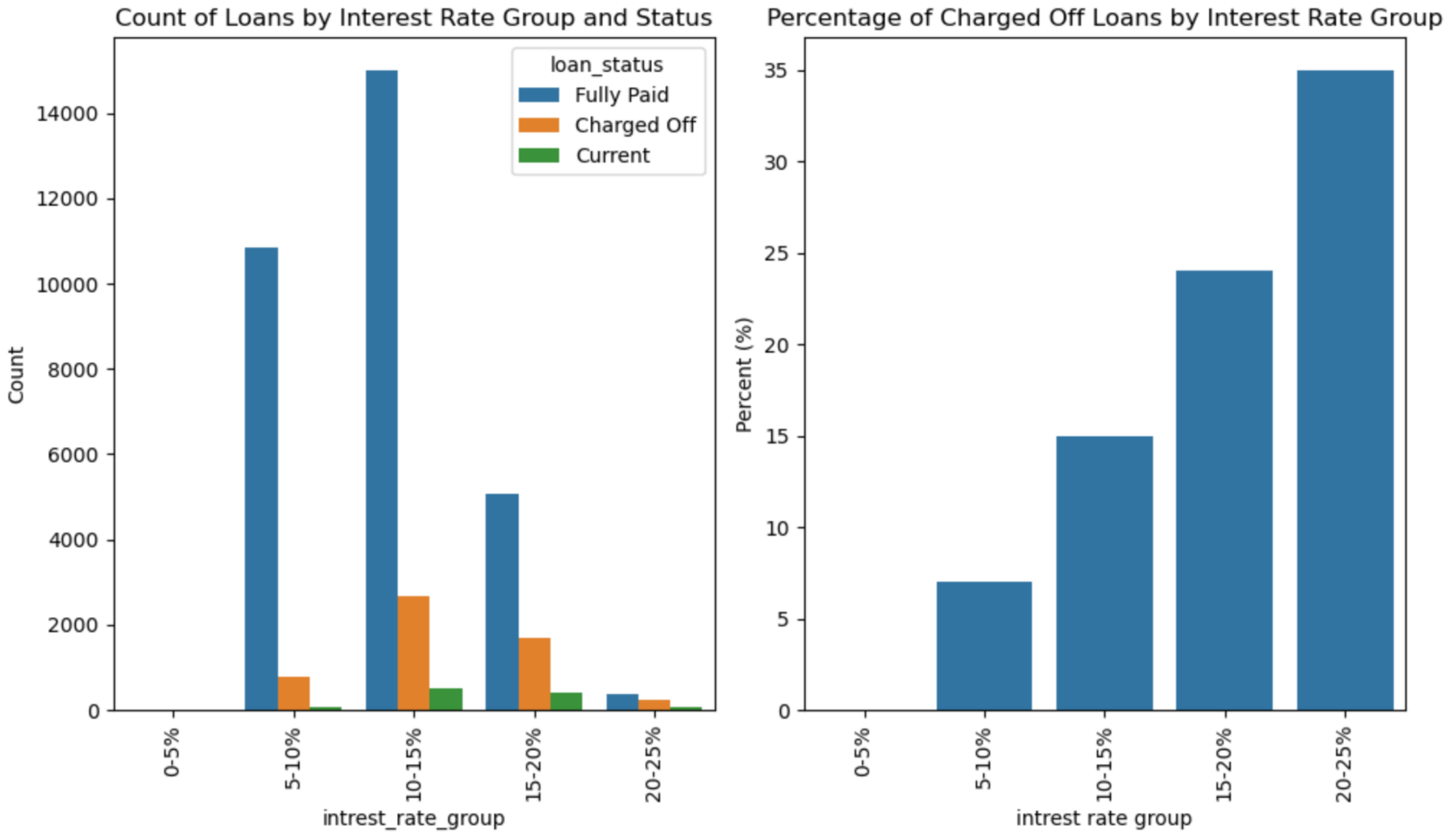
- It is clear that more people have taken loans with a tenure of 36 months compared to 60 months.

- Additionally, we observe that more borrowers have defaulted on loans with a 60 month tenure.



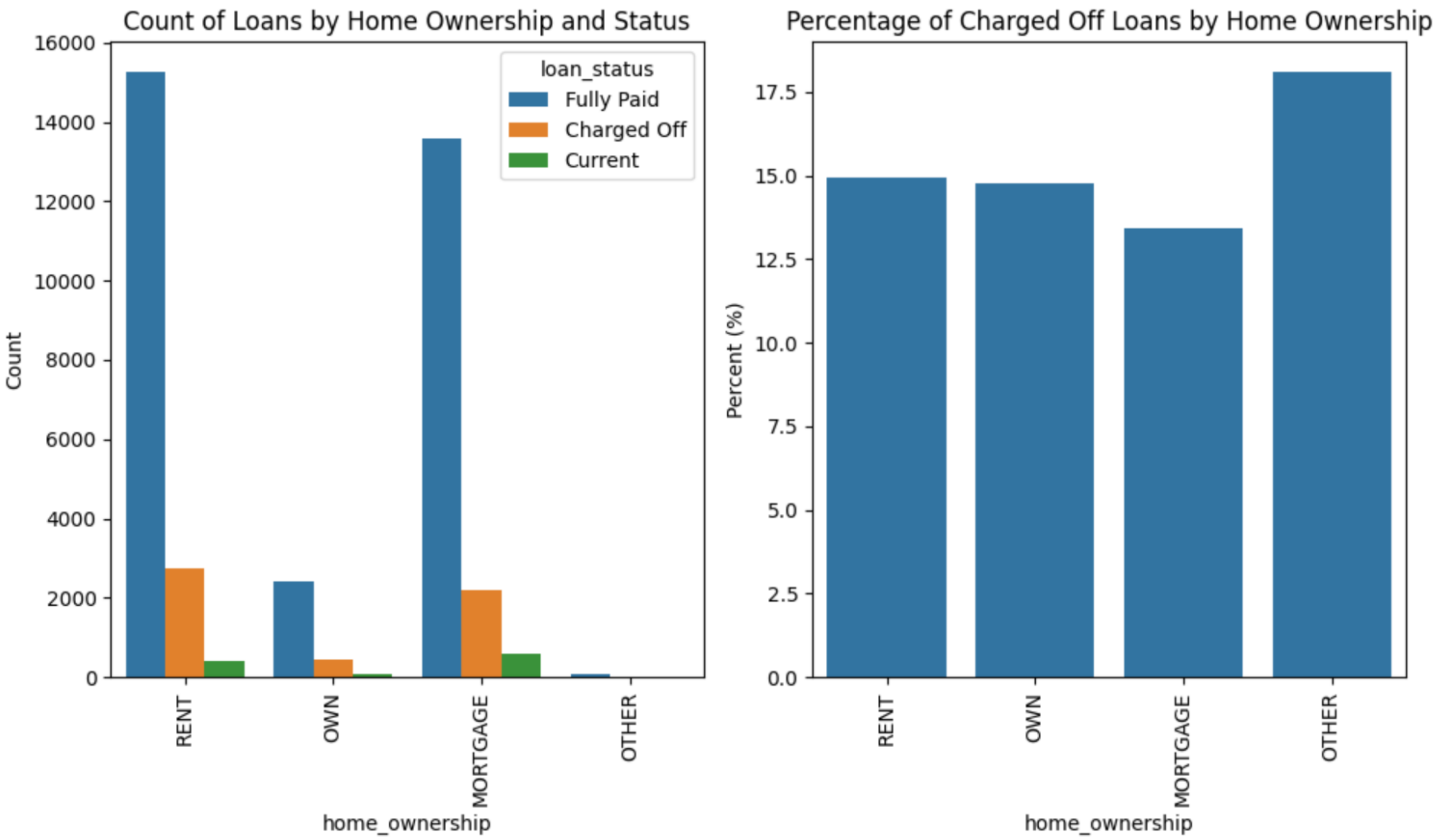
Loans by Interest Rate Group and Status

- The left graph indicates that the majority of borrowers have interest rates between 10-15%.
- The right graph demonstrates that as interest rates increase, the likelihood of loan default also rises.
- We can also conclude that defaulters tend to increase with higher interest rates.



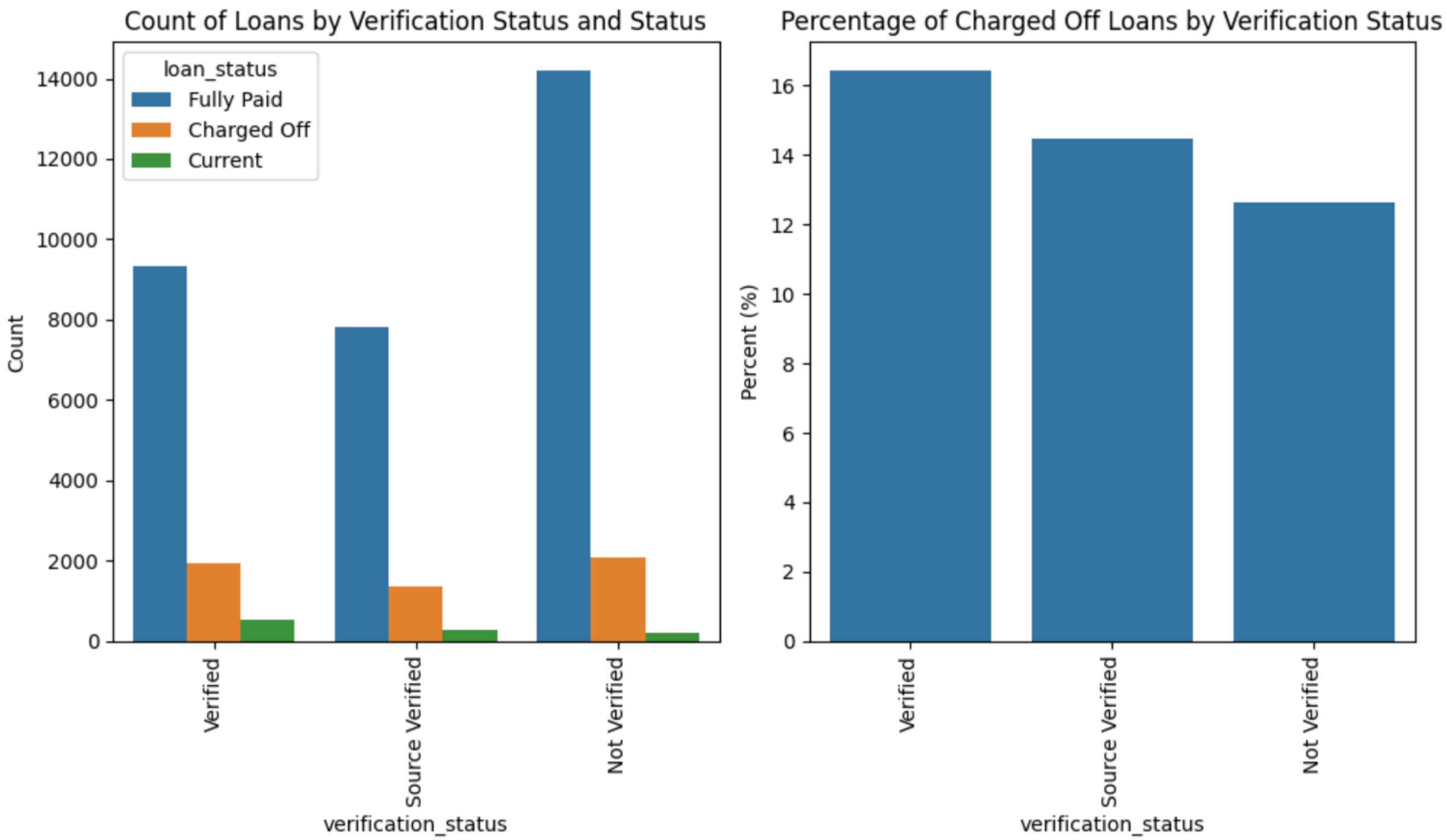
Loans by Home Ownership and Status

- We can see from the right graph that the "Other" category has a high probability of default.
- However, since the data for this category is very limited, we can ignore it.
- Keeping that in mind, we observe that for a larger set of borrowers, home ownership **does not** play a vital role in identifying loan defaulters.



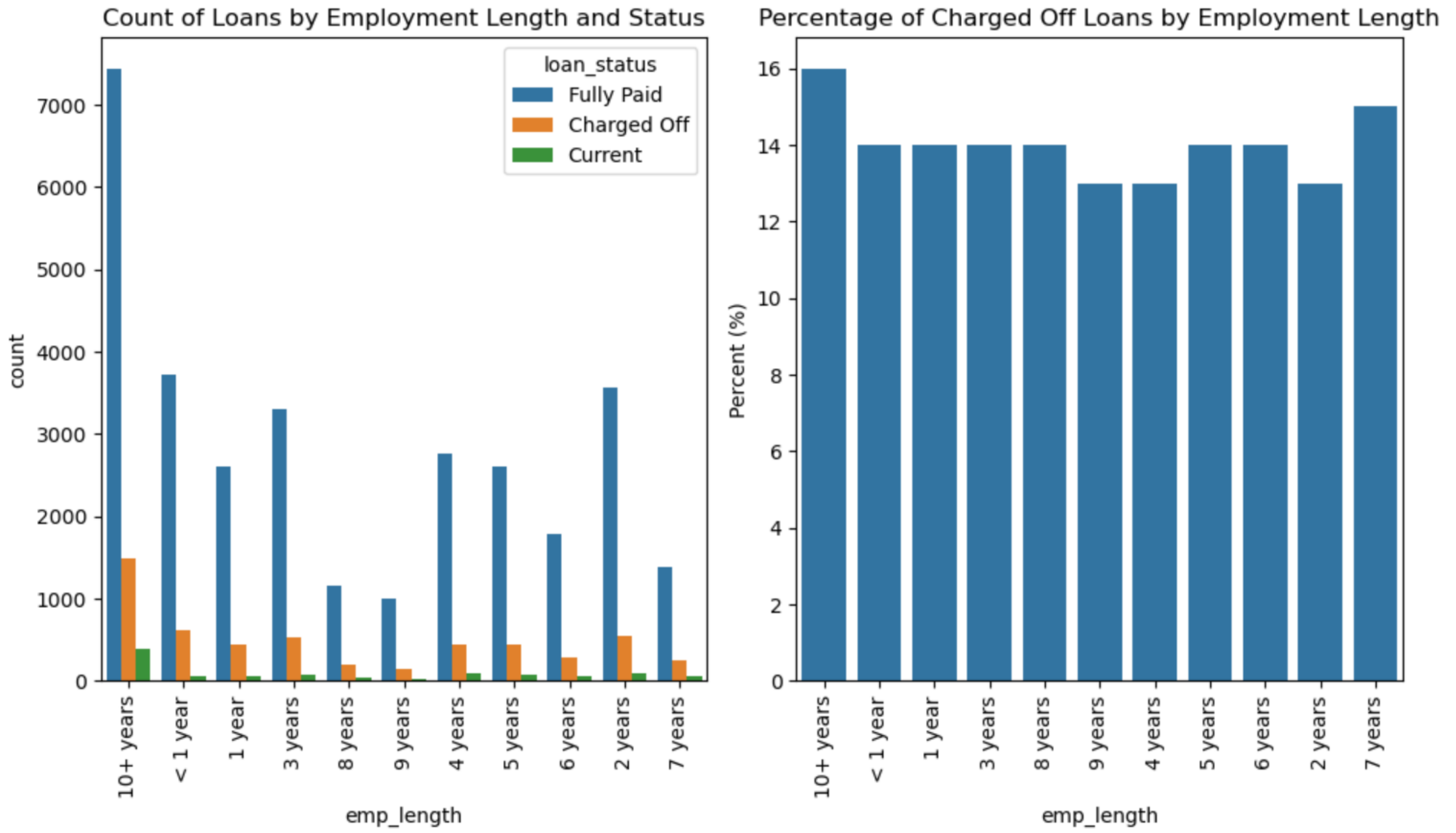
Loans by Income Verification and Status

- We can clearly see that income source verification is **not** a driving factor for defaulters, as people with verified income have a similar probability of default.



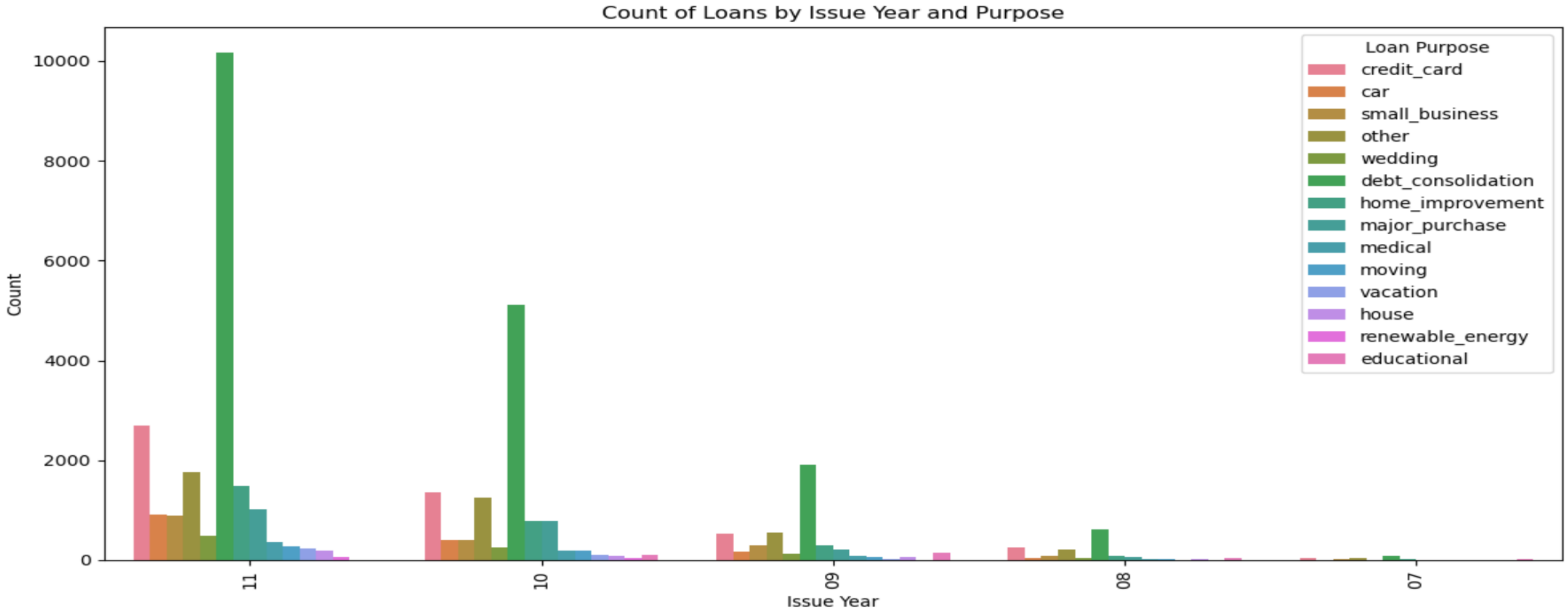
Loans by Employment Tenure

- Borrowers with over 10 years of experience take out more loans, leading to higher numbers of defaults and fully paid loans.
- However, when we compare relative rates across the different experience levels, the 10+ year group doesn't have a significantly higher rate.
- Overall, years of experience aren't a strong predictor of defaults, so we can consider the 10+ year group as being in a cautious category.



Loans by Issue year and Purpose

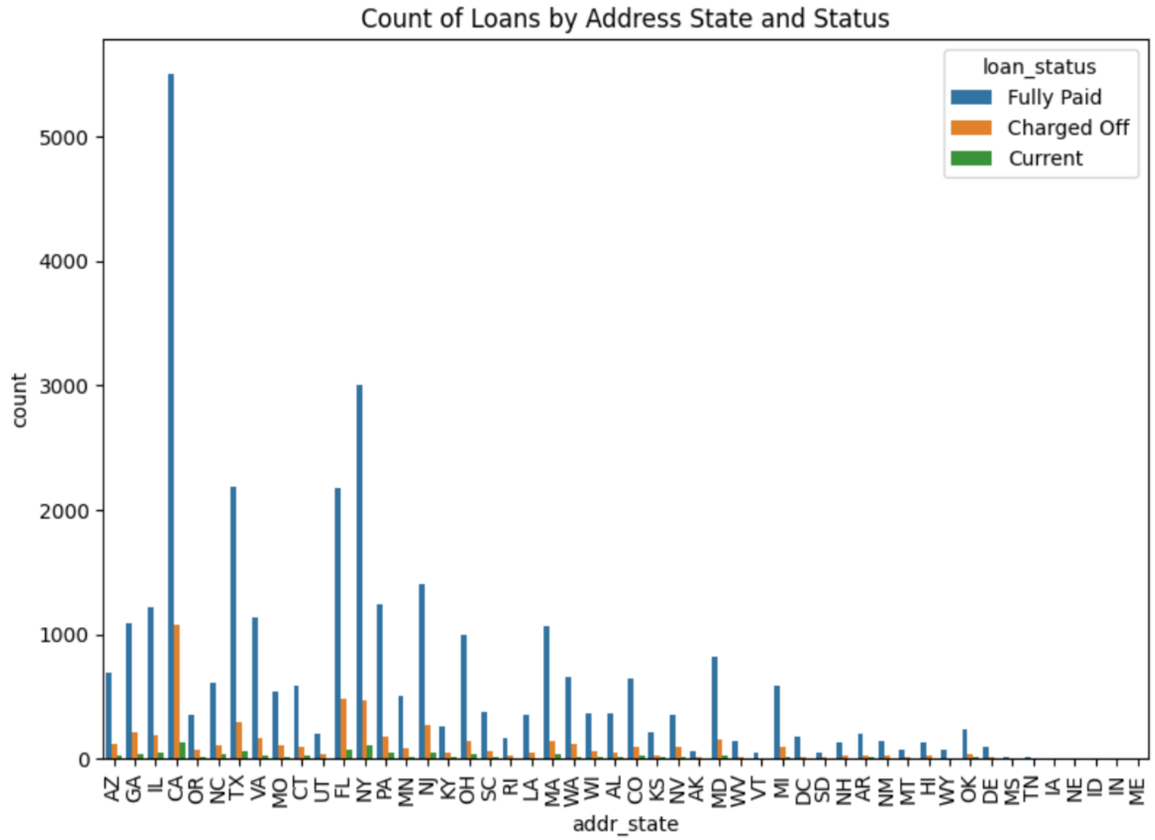
- Between 2011 and 2008, a significant portion of loans was primarily taken out for the purpose of debt consolidation.
- This trend highlights borrowers' focus on consolidating their existing debts into a single loan, likely to simplify payments and potentially secure lower interest rates.

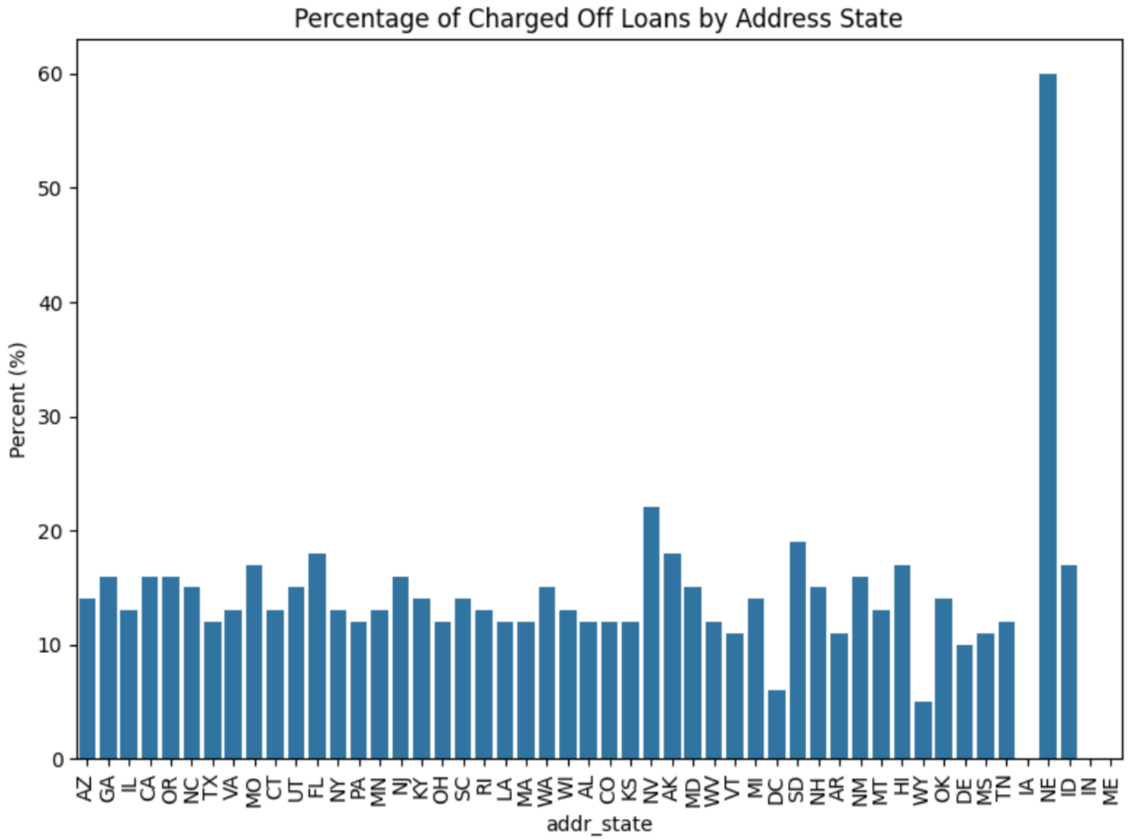


upGrad

Loans by Address State and Status

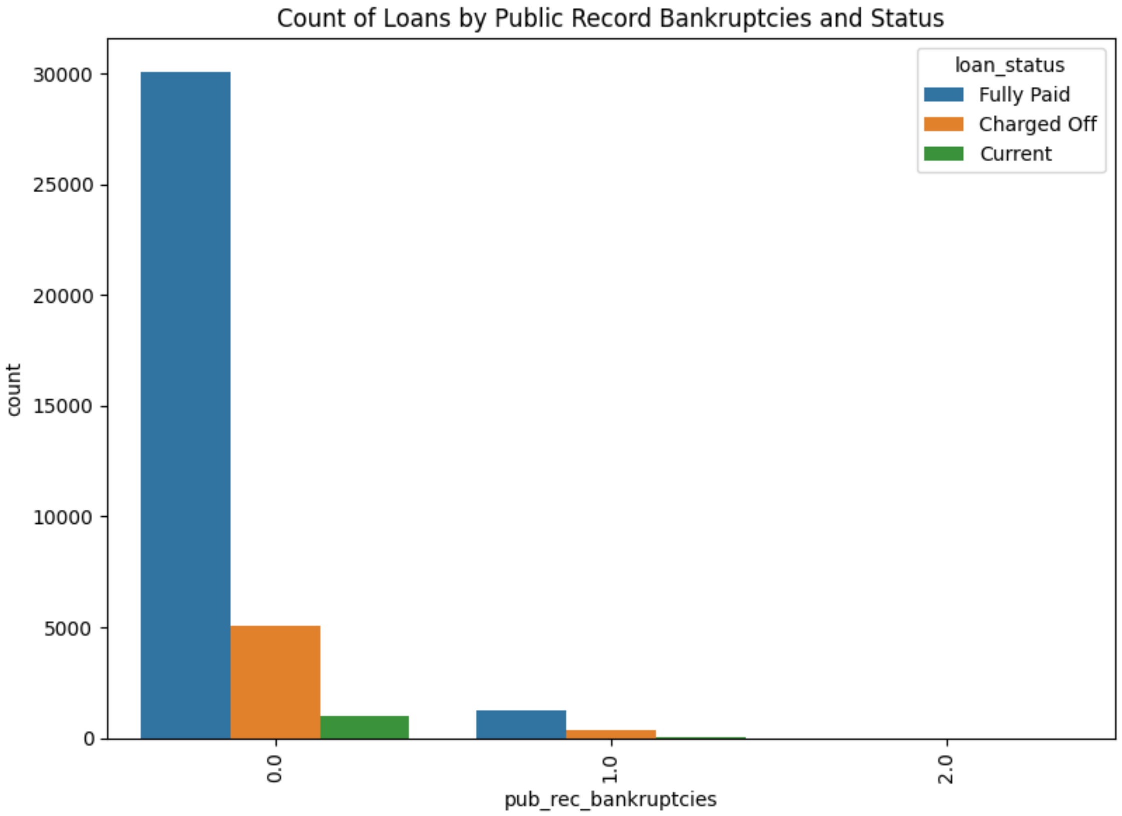
- From the left graph, we see that most borrowers are from California (CA).
- However, the right graph indicates that borrowers from Nebraska (NE) have relatively higher chances of default.
- Since the data for Nebraska is very limited, we can ignore that spike and conclude that the address state does not play a vital role in identifying defaulters.

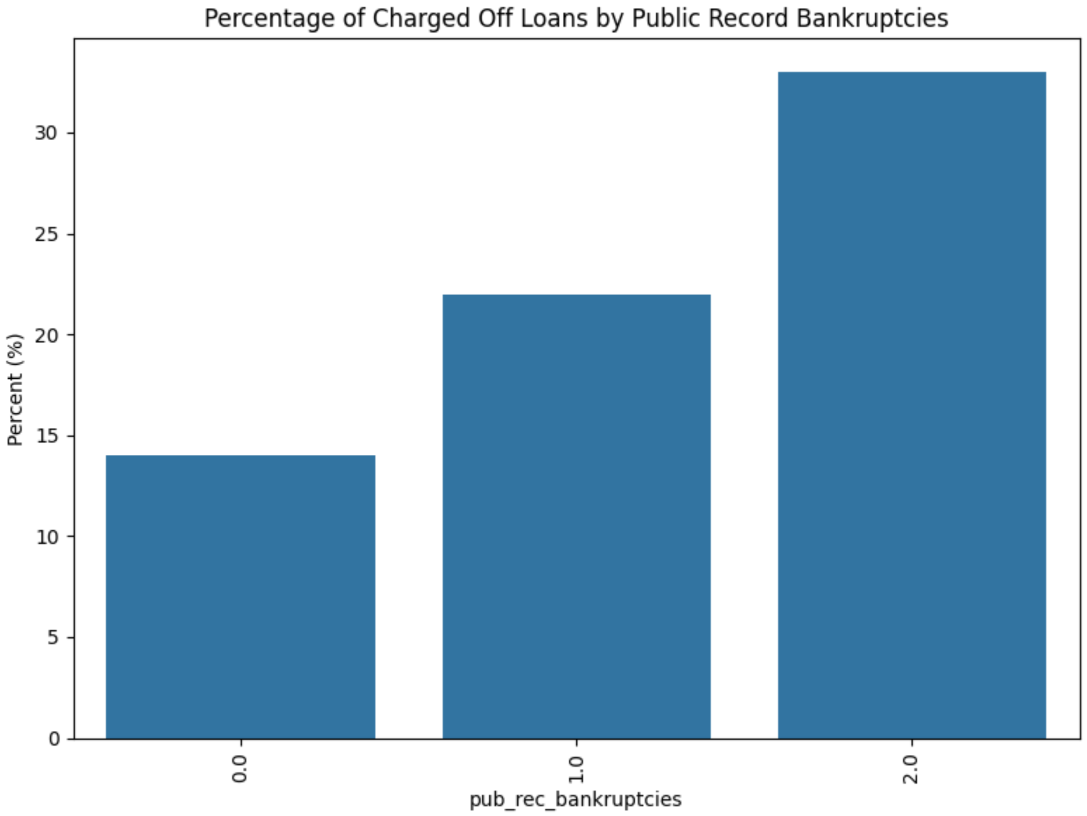




Loans by Public Bankruptcy and Status

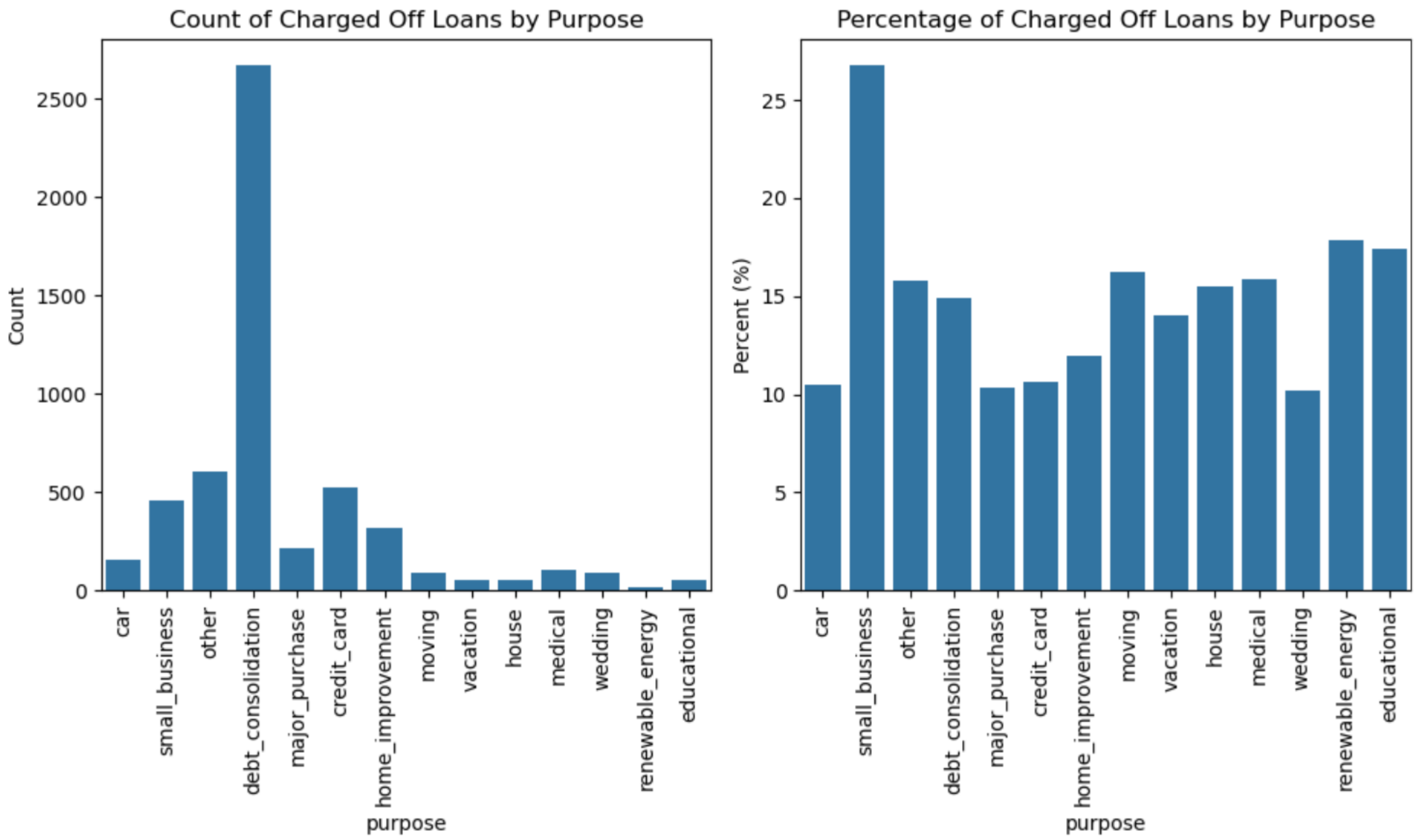
- We see that individuals with bankruptcy records have a higher chance of default, but since the data is limited, we can ignore this factor





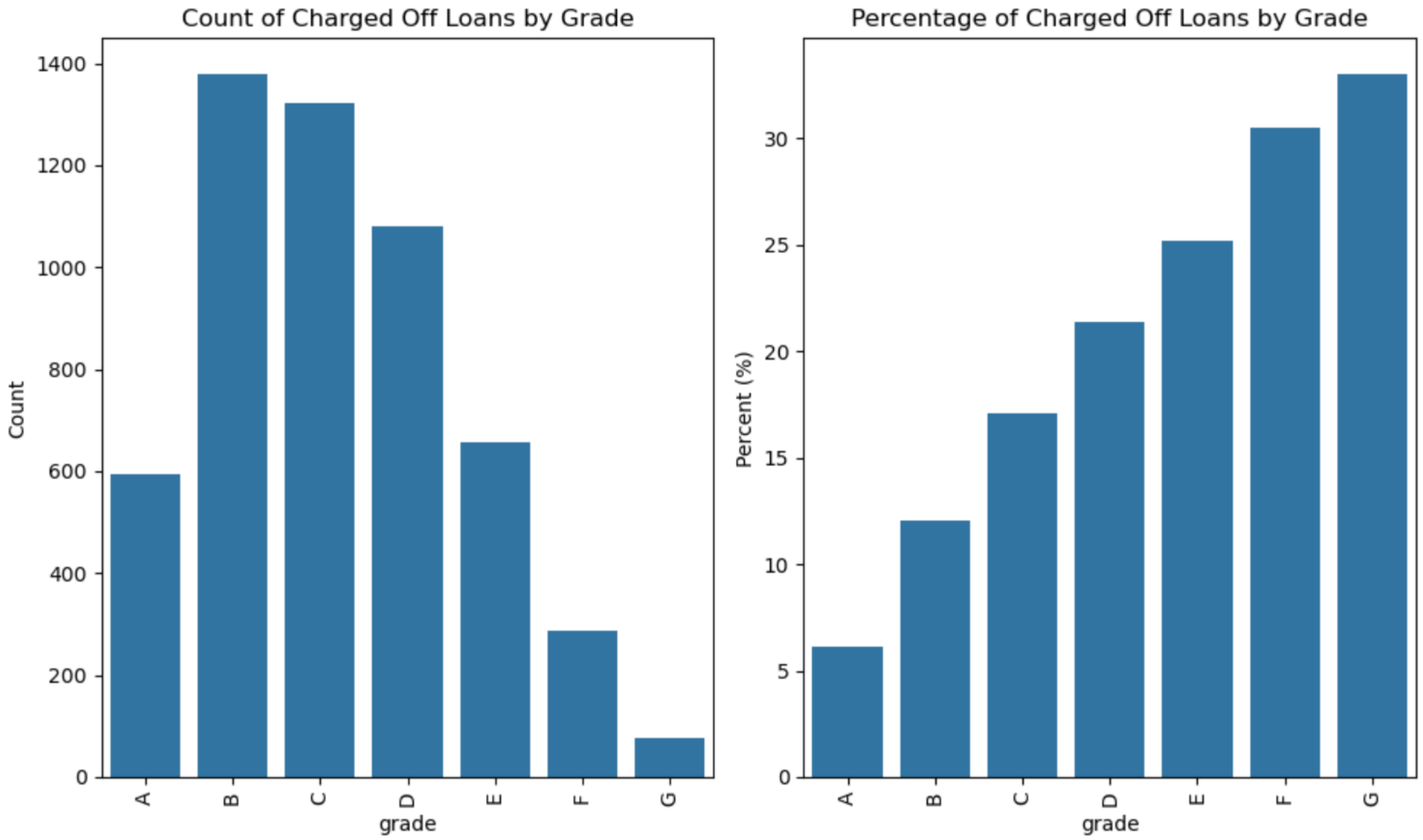
Defaulted Loans with respect to Purpose

- If we look at the absolute numbers, loans taken for debt consolidation have a high likelihood of defaulting, although the total number of loans for this purpose is also substantial.
- However, when we examine the categories relatively, loans taken for **small businesses** show a higher probability of default.



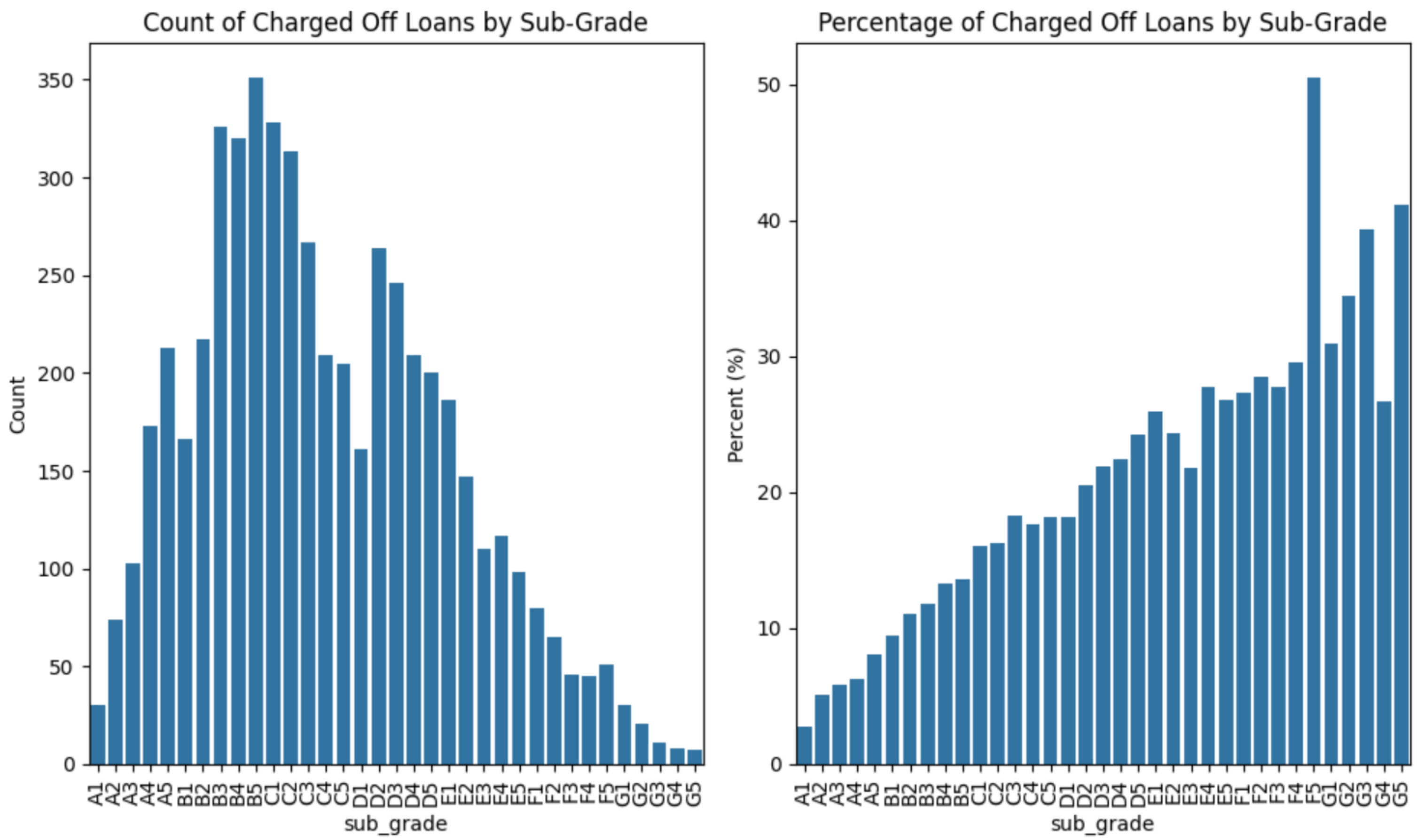
Defaulted Loans with respect to Grade

- In absolute terms, B, C, and D grade loans have the highest default rates.
- However, when we look at the grades relatively, the probability of default increases as we move from A to G.



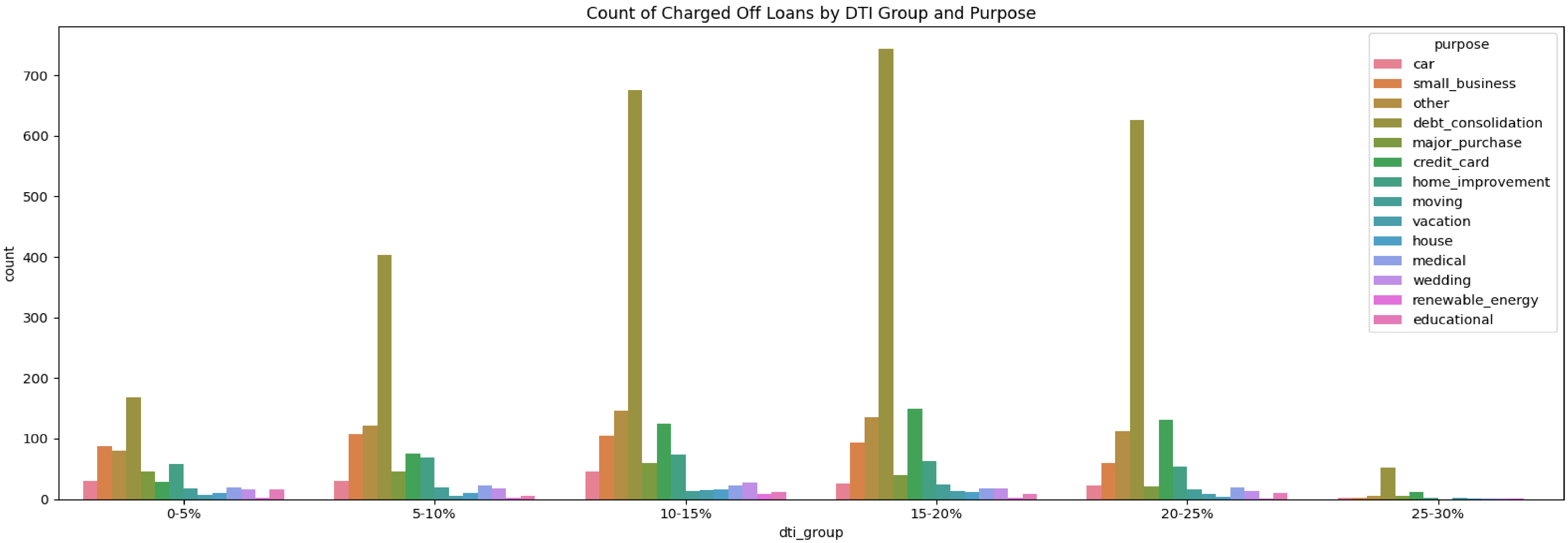
Defaulted Loans by Sub-Grade

- In the initial analysis, one might be confused into thinking that subgrades B to C are the driving factors for loan defaults.
- However, in the detailed visual analysis of the percentage of defaults by subgrade, we can clearly see that subgrade is **directly proportional** to loan default.
- Additionally, F5 has being the most risky sub grade and around 50% of the loans have been defaulted from this sub grade.

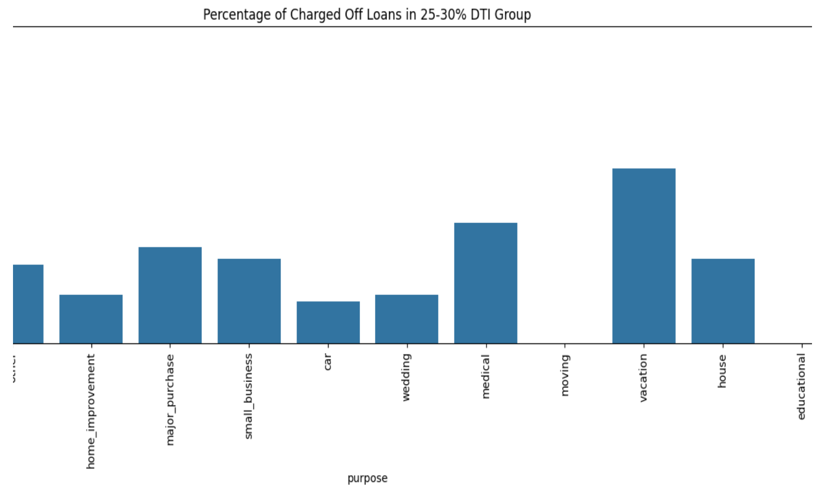


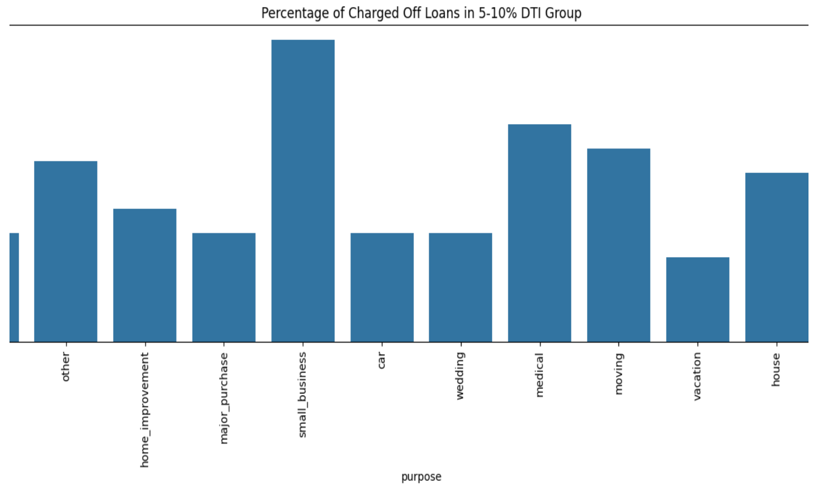
Charged Off Loans by DTI Group and Purpose

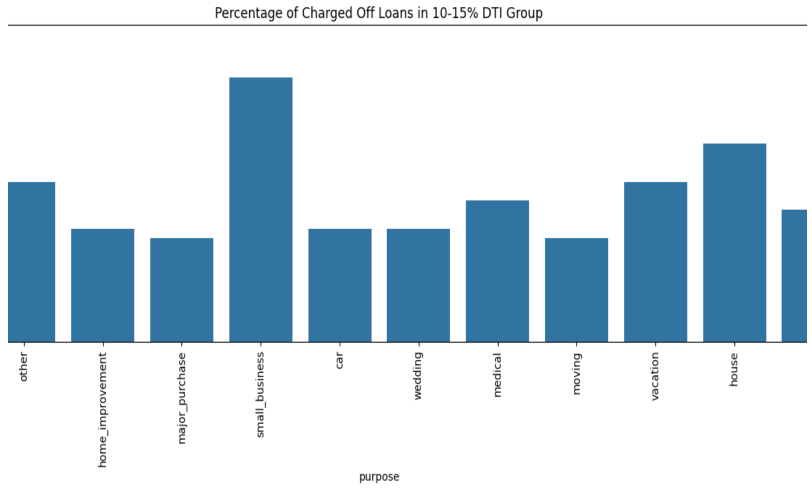
- Small business loans are more likely to default when borrowers have a DTI ratio between 0-25%.
- Educational loans tend to default more often if the borrower's DTI is either between 0-5% or 20-25%.
- Renewable energy loans have a higher default risk for borrowers with a DTI ratio of 10-15% or 25-30%.

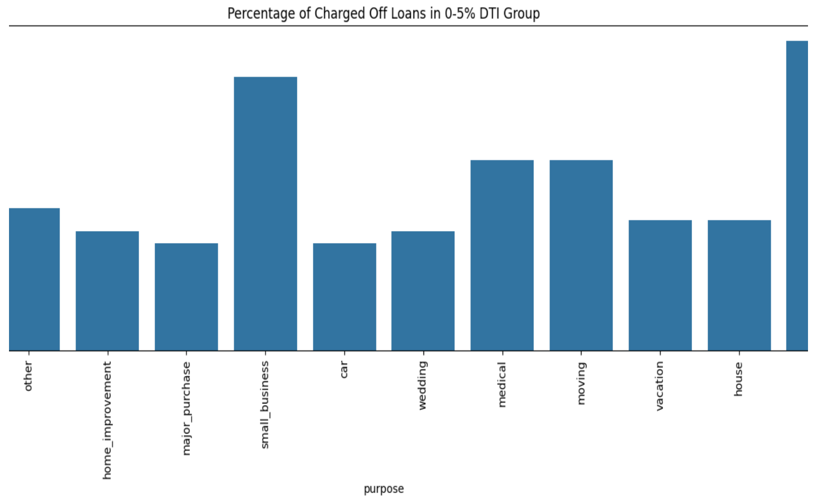


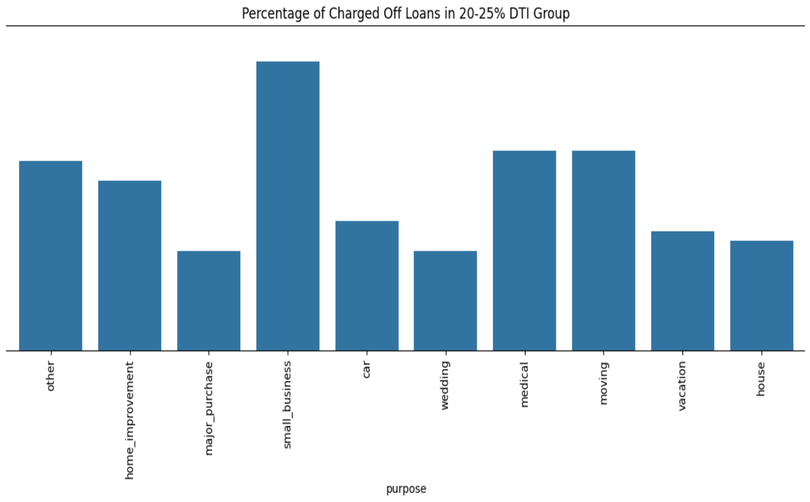
Charged Off Loans by DTI Group and Purpose (contd.)

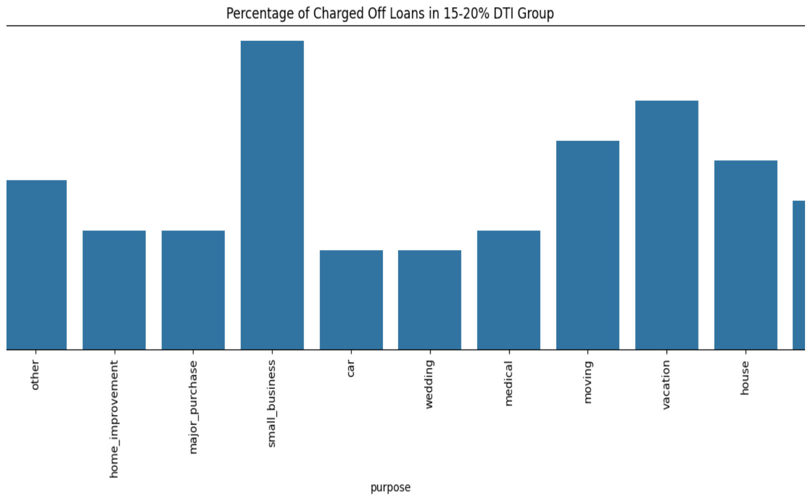






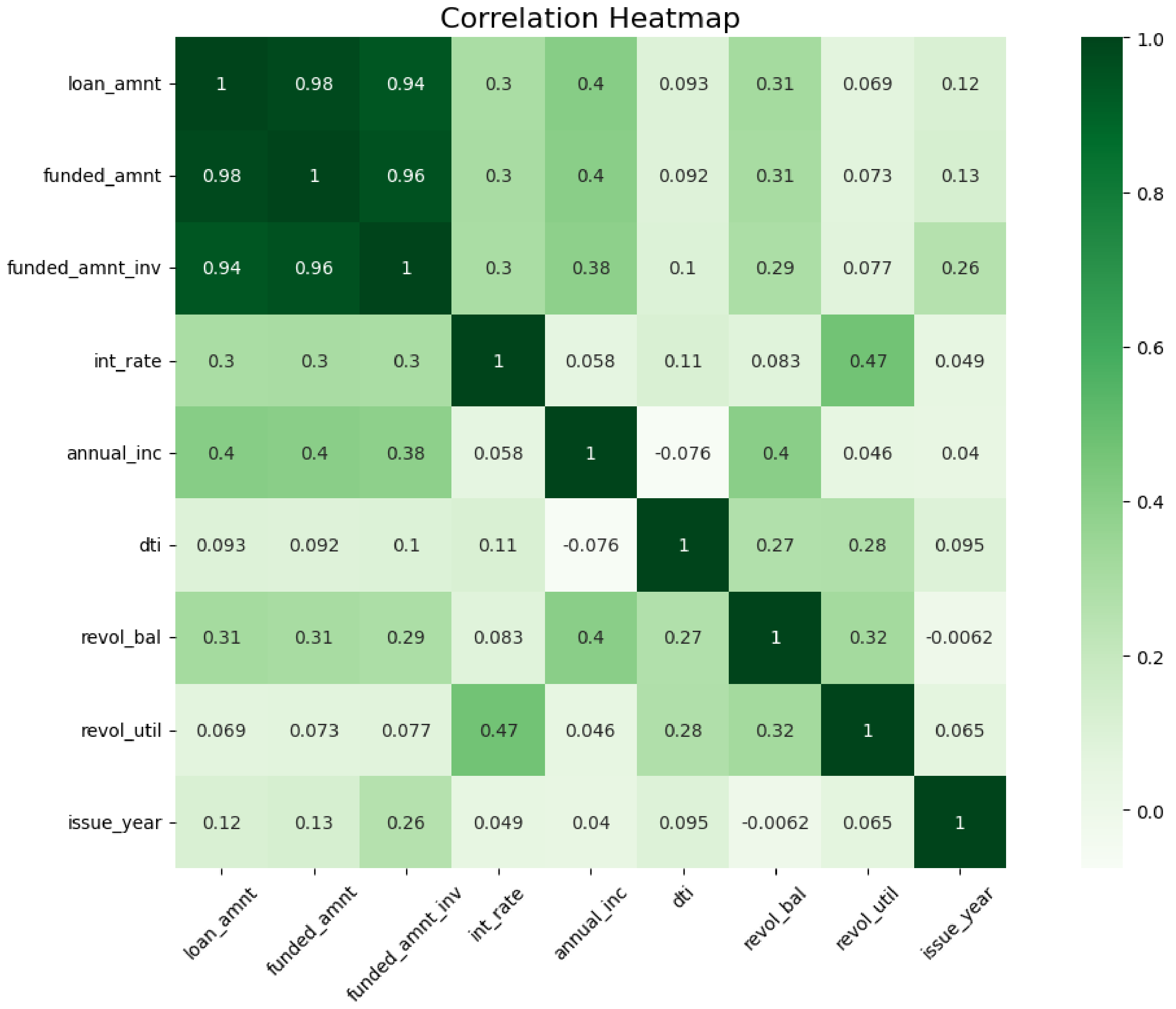






Correlation Heatmap

- The loan amount has a high positive correlation with both the funded amount and the investor-funded amount.
- However, the driving factor for charged off loan is not conclusive in the given heatmap.



Conclusions

upGrad

Loan Default Driving Factors

- Debt-To-Income (DTI) Ratio is directly proportional
- Loan Grades and sub-grades is directly proportional
- Loan Terms (Tenure) of 60 months is riskier than 30 months.
- Loan Purpose, specifically small business seems to be risky.
- Interest Rate

Loan Default Non-Driving Factors

- Home Ownership
- Address State
- Income Verification Status
- Employment Length