Lending Club Case Study

Presented By:

Shreyas M S

Shikhar Vashisht

Problem Statement

Lending Club faces a critical challenge in loan decision-making: balancing missed business opportunities with financial losses. Declining loans to creditworthy applicants, who are likely to fully repay their loans, results in potential revenue loss, while approving high-risk applicants, who are more likely to be charged off, leads to defaults and credit loss. This case study aims to identify key risk factors associated with loan defaults, enabling data-driven, balanced, and informed lending decisions to minimize financial setbacks and optimize lending strategies.

Overview of Loan Status and Dataset

Fully Paid	Charged Off	Current
The borrower has fully repaid the loan, including both principal and interest.	The borrower has failed to make required payments, resulting in a loan default.	The borrower is making installment payments, and the loan is active with no defaults.

Our analysis targets loans with clear outcomes ("Fully Paid" or "Charged Off") to uncover trends and insights that can guide business decisions.

The dataset contains the complete loan data for all loans issued through the time period 2007 to 2011.

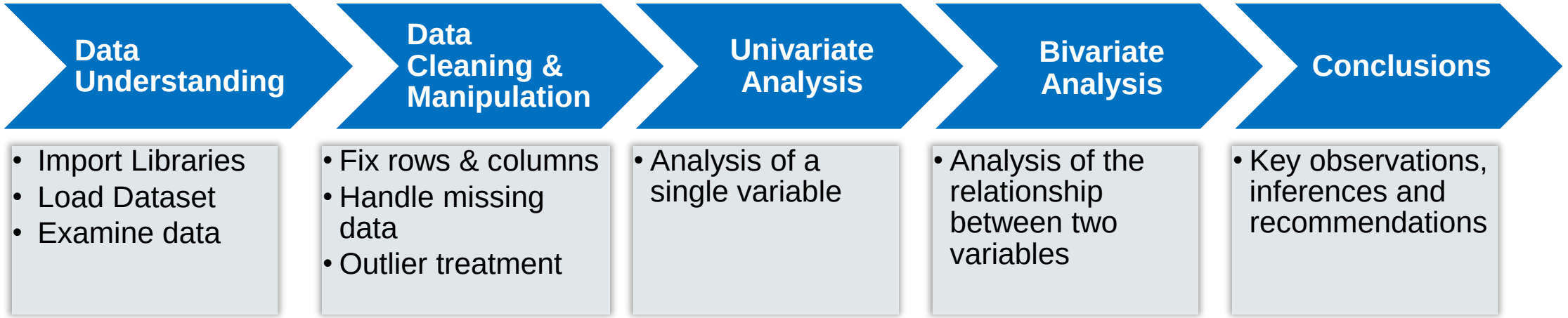
Objective

This analysis aims to

- ✓Perform Exploratory Data Analysis to uncover key patterns and characteristics that distinguish high-risk (likely-to-default) applicants from low-risk ones.
- ✓Utilize insights to enable the company to reduce credit losses by identifying and mitigating risk factors effectively, ultimately improving decision-making around loan approvals.

Through this EDA, we aim to provide actionable insights that balance risk management with business growth, allowing the company to optimize its lending strategies based on historical applicant and loan data.

Approach and Methodology



Data Understanding

Numerical Columns

Continuous

Loan Amount (loan_amnt)
Funded Amount (funded_amnt)
Interest Rate (int_rate)
Annual Income (annual_inc)
Debt-To-Income Ratio (dti)
Installments (installment)

Discrete

Public Record Bankruptcies (pub_rec_bankruptcies)

Categorical Columns

Unordered

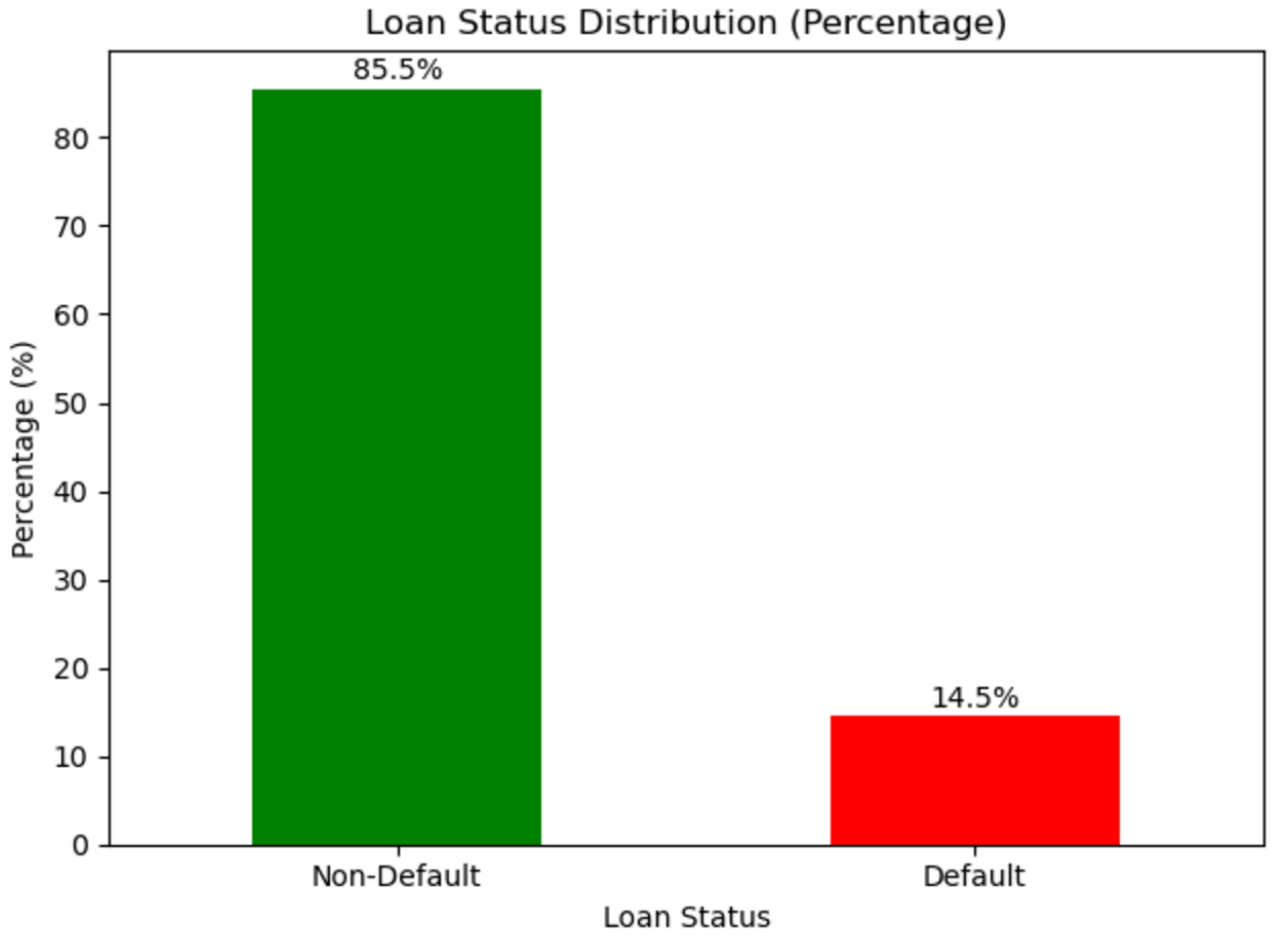
Address State (addr_state)
Home Ownership (home_ownership)
Loan Purpose (purpose)
Loan Status (loan_status)
Loan Paid (loan_paid)
Verification Status (verification_status)

Ordered

Grade (grade)
Sub Grade (sub_grade)
Employee Length (emp_length)
Term (term)
Loan Amount Bucket (loan_amnt_bucket)
Funded Amount Bucket (funded_amnt_bucket)
Interest Rate Bucket (int_rate_bucket)
Annual Income Bucket (annual_inc_bucket)
Debt-To-Income Bucket (dti_bucket)
Issue Year/Month/Quarter (issue_y, issue_m, issue_q)

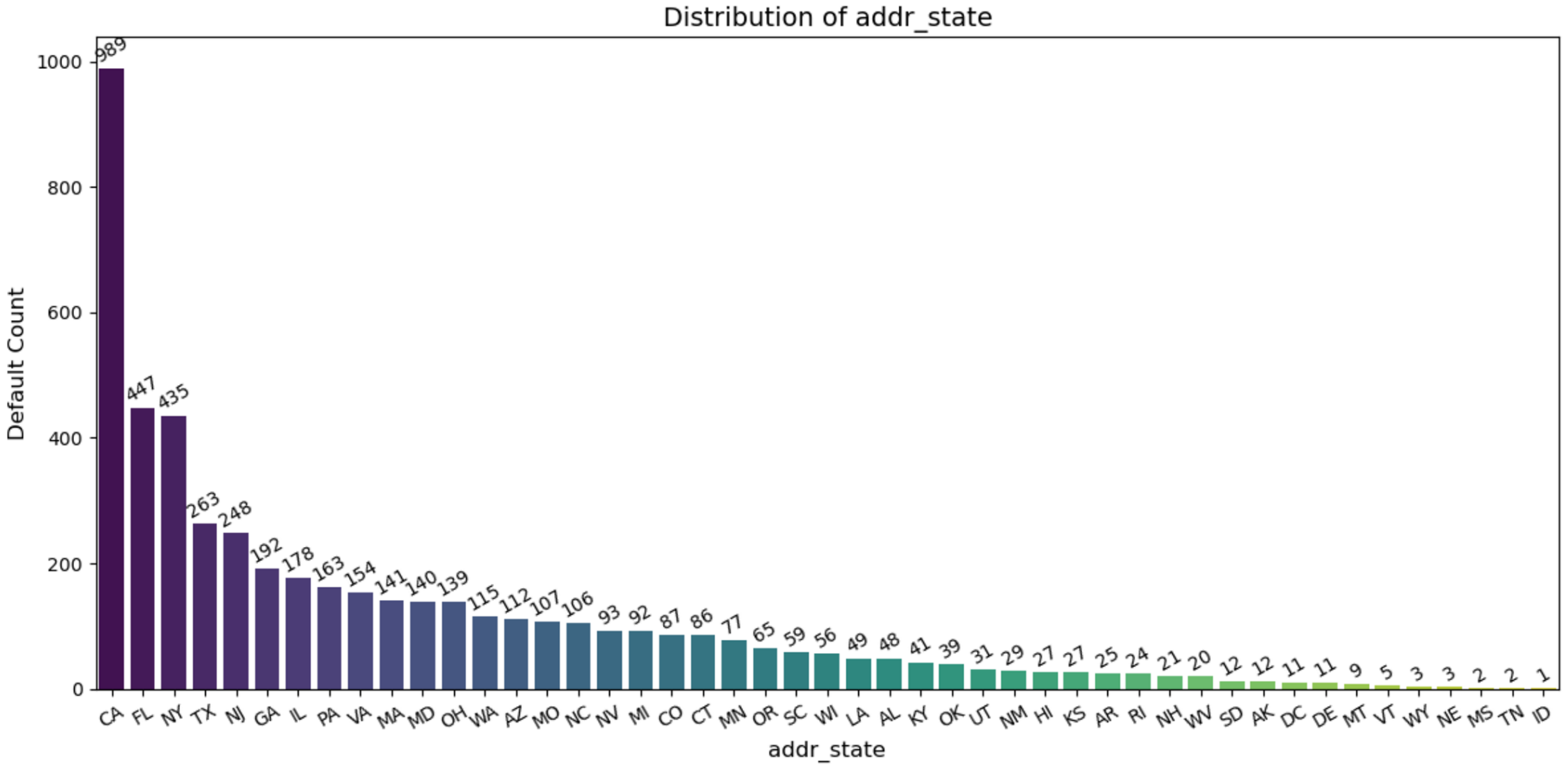
Data Understanding

Overall Default Rate is 14%

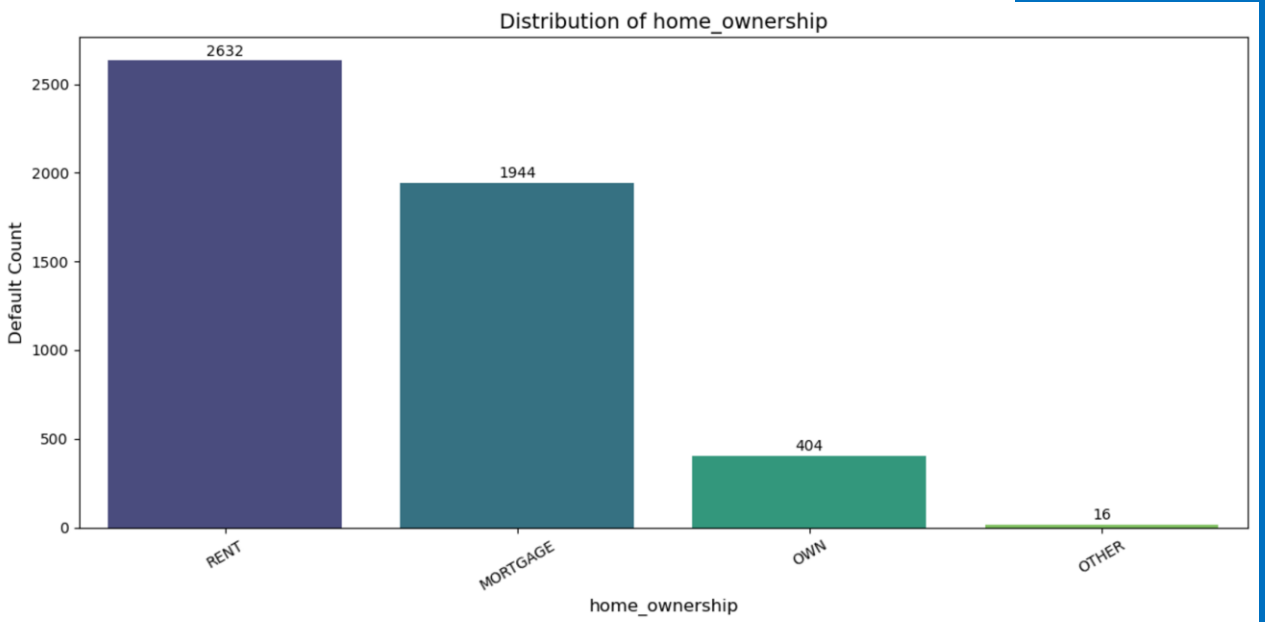


Univariate Analysis

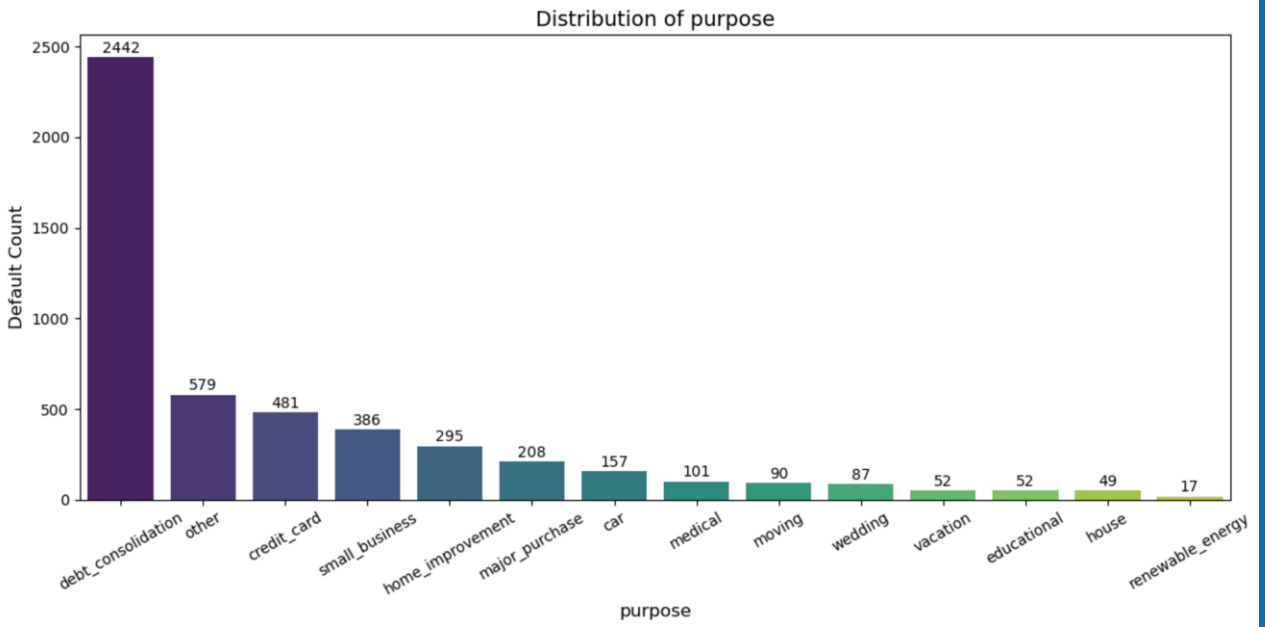
Categorical Variables



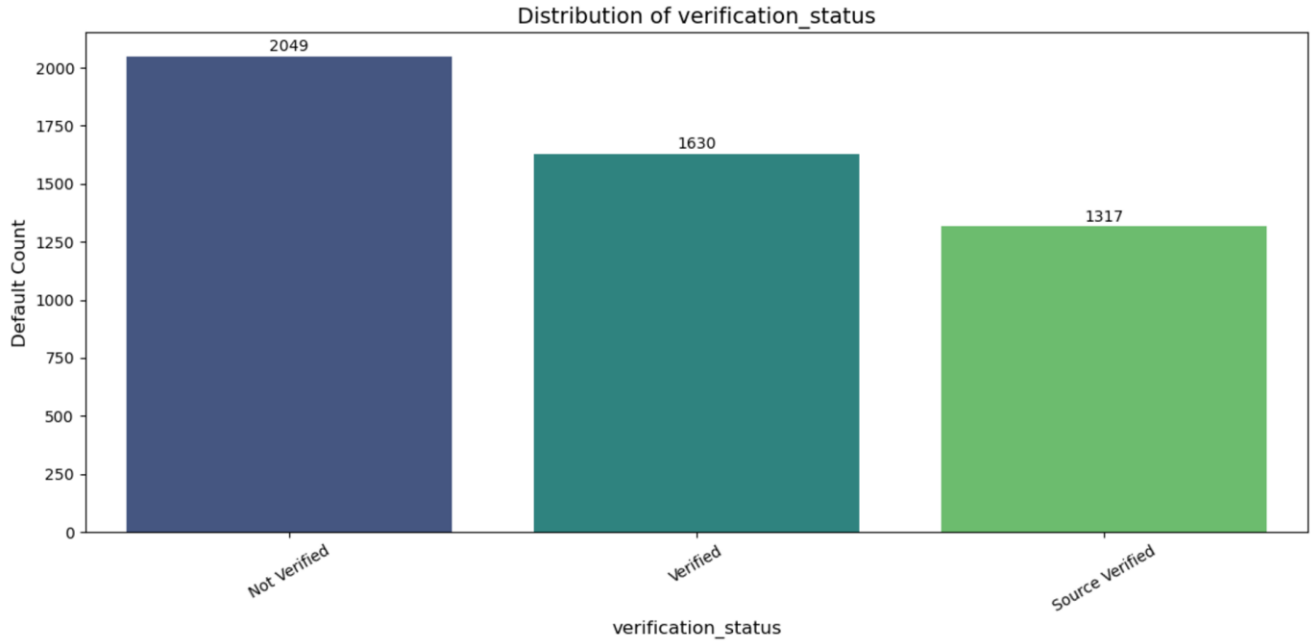
Address State - California (CA) recorded the highest number of "Default" loan applicants, totaling 989, highlighting a significant risk area for the lending company. To address this, the company could introduce tailored loan products, enhance financial literacy programs, or implement dynamic interest rates to proactively mitigate risks.



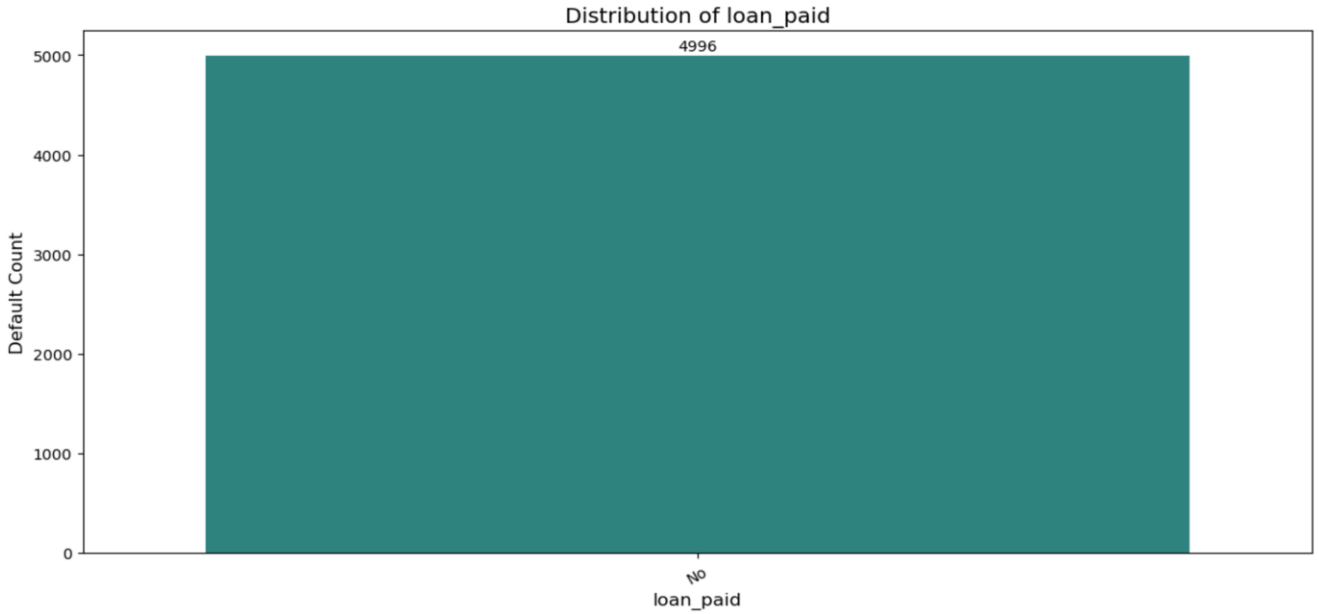
Home Ownership - The majority of "Charged Off" loan applicants, amounting to 2,632 individuals, resided in rented accommodations. The company could assess the financial stability of renters more thoroughly by incorporating additional parameters, such as rental expenses and savings, into the credit evaluation process.



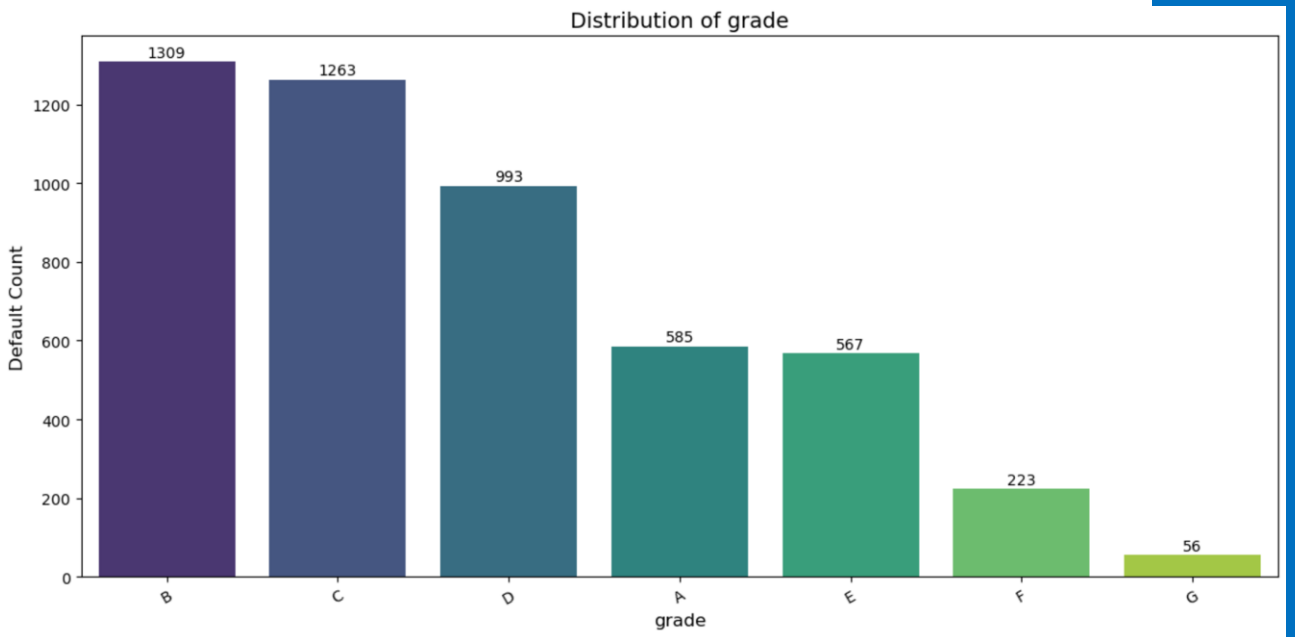
Purpose - A significant number of "Charged Off" loan applicants, totaling 2,442 individuals, cited debt consolidation as their primary loan purpose. The company should implement stricter approval criteria for debt consolidation loans, including evaluating debt-to-income ratios, assessing debt causes, and providing financial counseling.



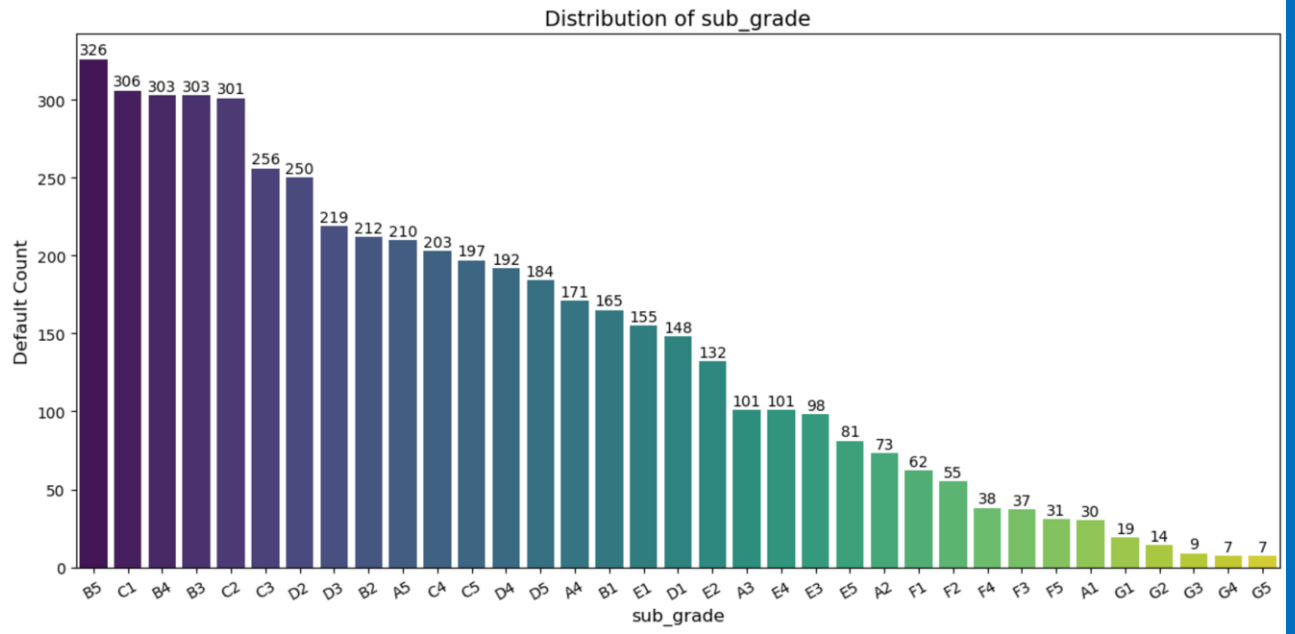
Verification Status - Loans marked as 'Not Verified' represent the largest portion of charged-off loans, totaling 2,049 applicants. This indicates a potential gap in the verification process that could be contributing to higher default rates. To reduce the risk of charged-offs, the company should strengthen the income verification process, and implement more rigorous criteria for loan approval.



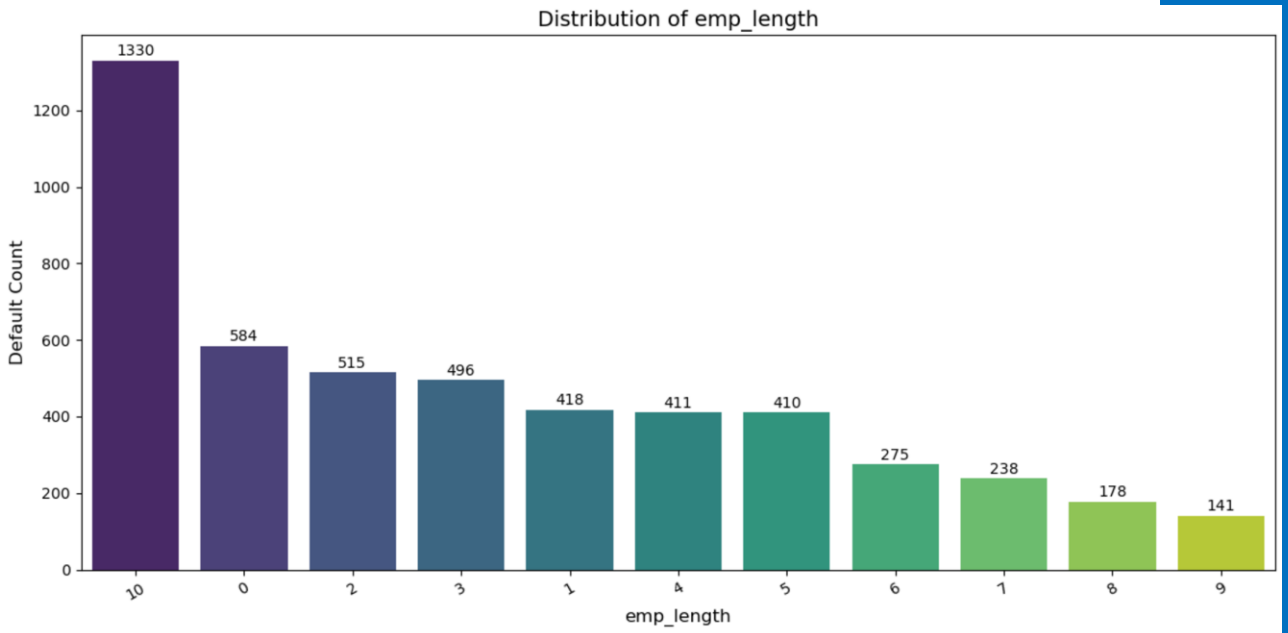
Loan Paid - A considerable portion of loan participants, totaling 4,996 individuals, defaulted on their loans, indicating a critical risk for the lending company. To mitigate the risk associated with loan defaults, the company should adopt a more comprehensive credit risk assessment process, utilizing advanced predictive models and incorporating alternative data sources.

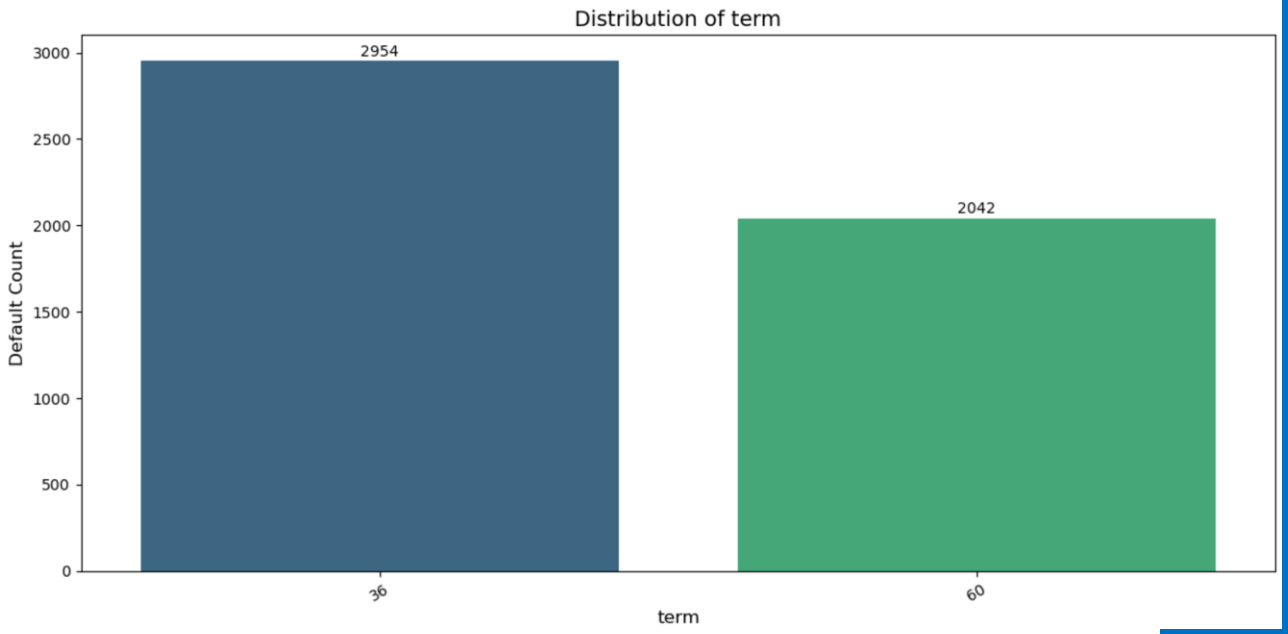


Grade - Credit Grade B had the highest number of loan defaults, with 1,309 applicants, closely followed by Grade C with 1,263 applicants. This suggests that borrowers in these credit grades are at a higher risk of default, indicating the need for enhanced credit evaluation processes for these groups.



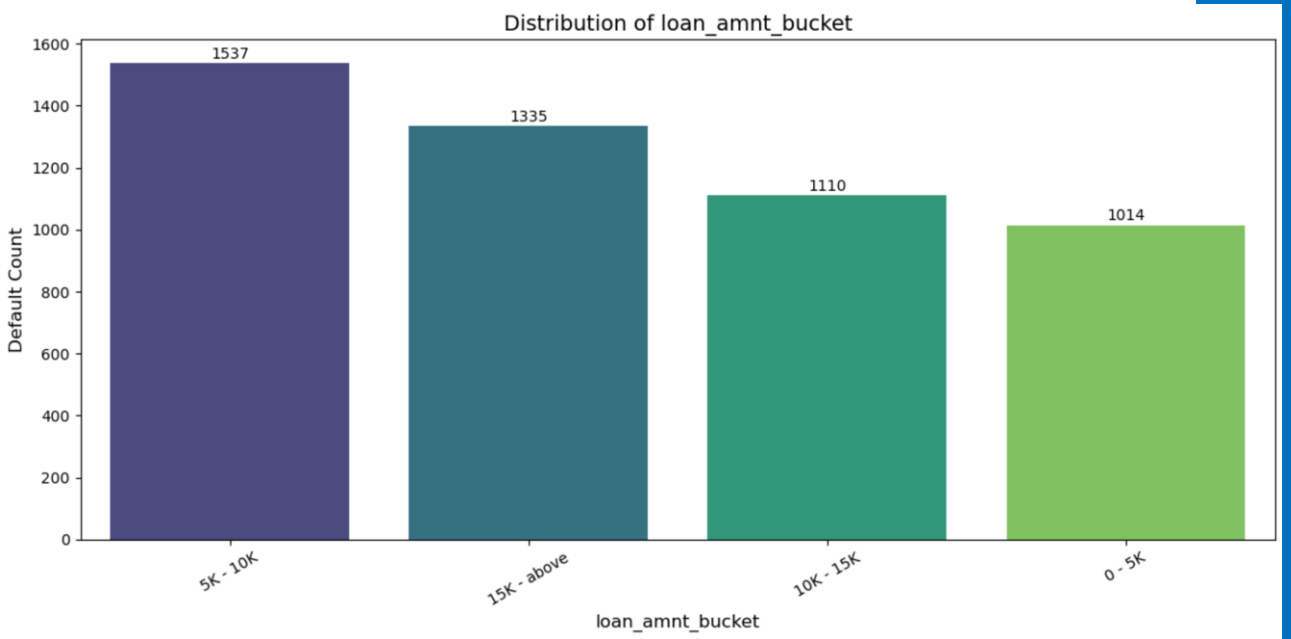
Sub Grade - Sub-grades B3, B4, B5, C1, and C2 are notably linked to high default rates, with B5 recording the highest default rate, accounting for 326 loan defaults. This granular level of credit segmentation highlights specific risk areas within broader credit grades, requiring targeted interventions.

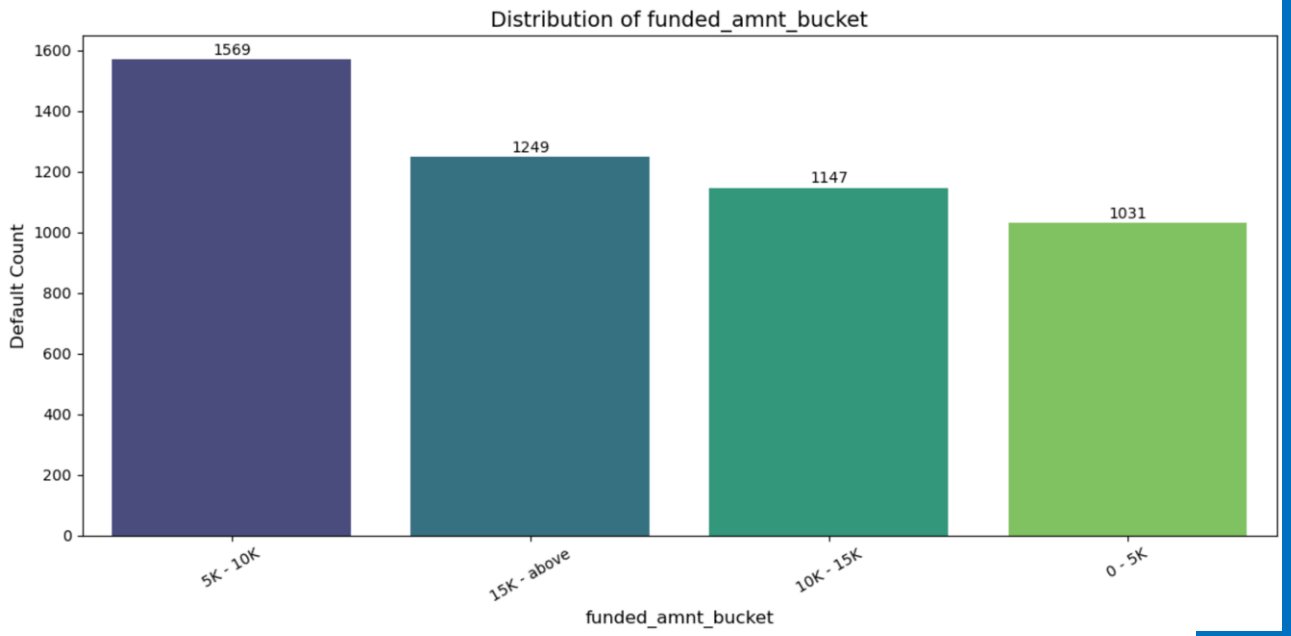




Employee Length - Applicants employed for over 10 years had the highest number of loan defaults, totaling 1,330. This suggests that a long employment history does not always correlate with loan repayment ability, and additional financial assessments may be necessary.

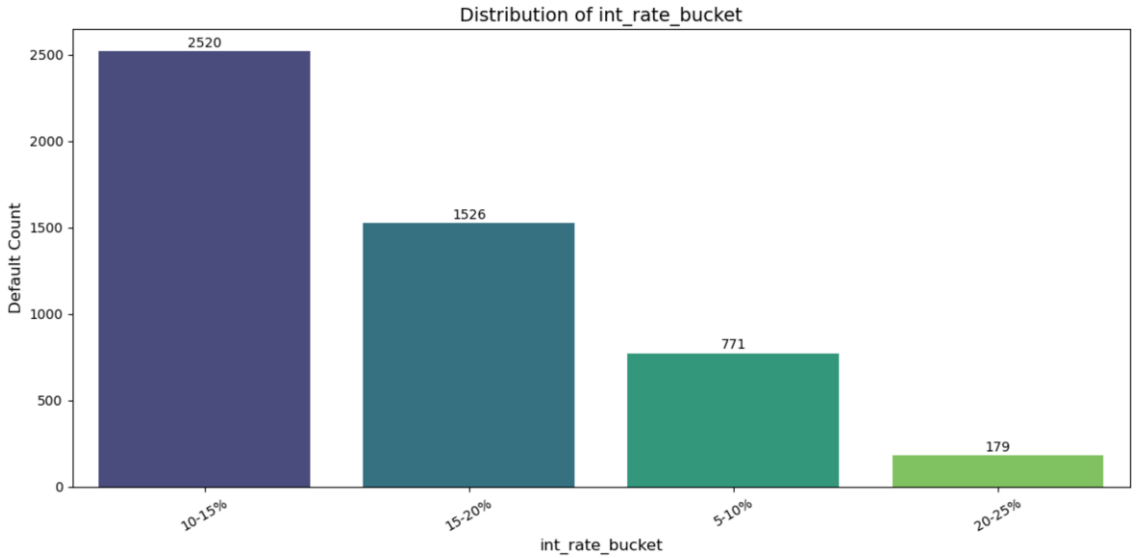
Term - Short-term loans of 36 months were the most popular among defaulters, with 2,954 cases. This indicates that shorter repayment durations may pose challenges for borrowers, potentially due to higher monthly payment obligations.



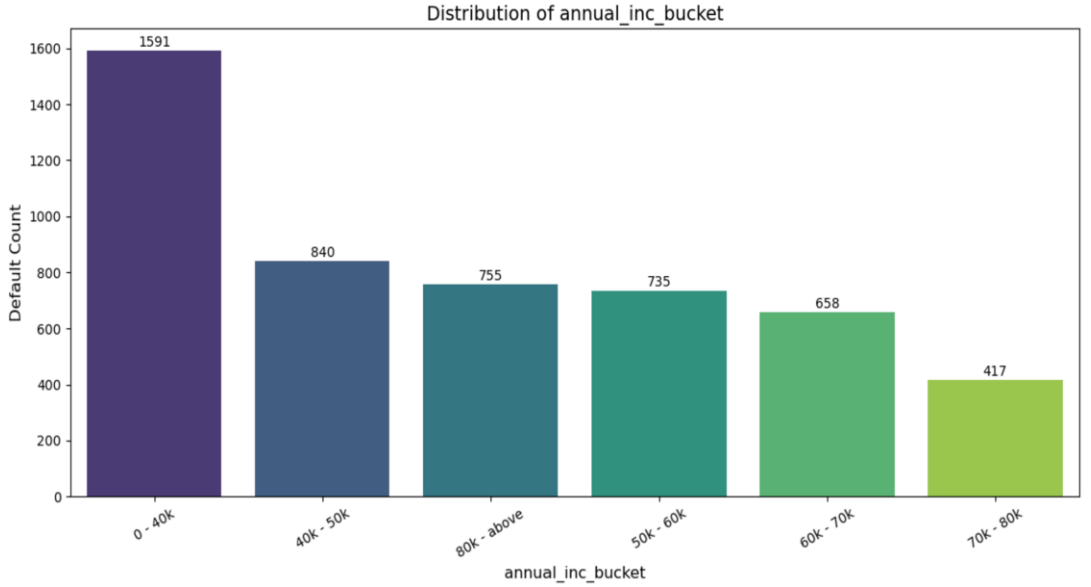


Loan Amount - Loan amounts between 5,000 and 10,000 USD had the highest defaults, with 1,537 applicants. This bucket may represent a sweet spot for moderate-risk loans that require careful scrutiny during approval.

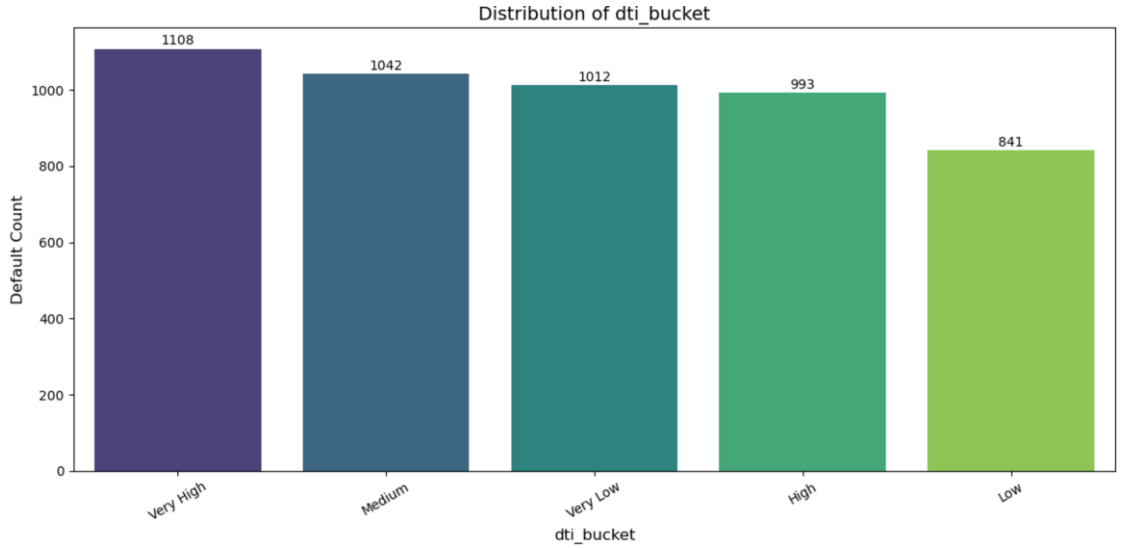
Funded Amount - Funded loans in the 5,000 – 10,000 USD range also saw the most defaults, with 1,569 cases. The alignment with the loan amount bucket underscores the need for thorough funding evaluations within this range.



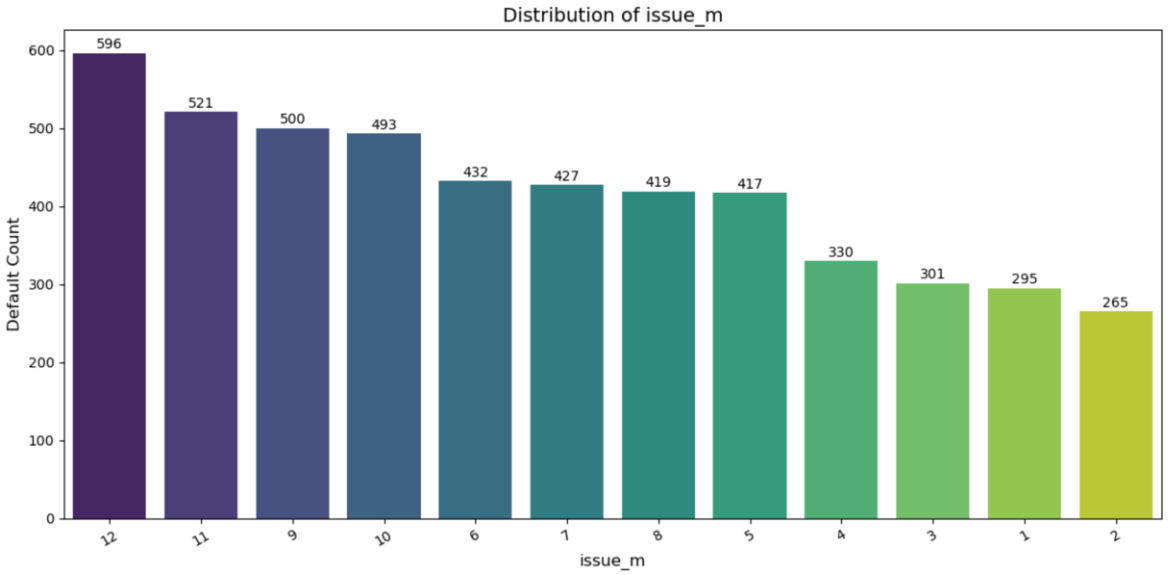
Interest rate - The 10–15% interest rate bucket accounted for the most defaults, with 2,520 cases. This suggests that mid-range interest rates may not significantly deter risky borrowers, necessitating more dynamic interest structuring.

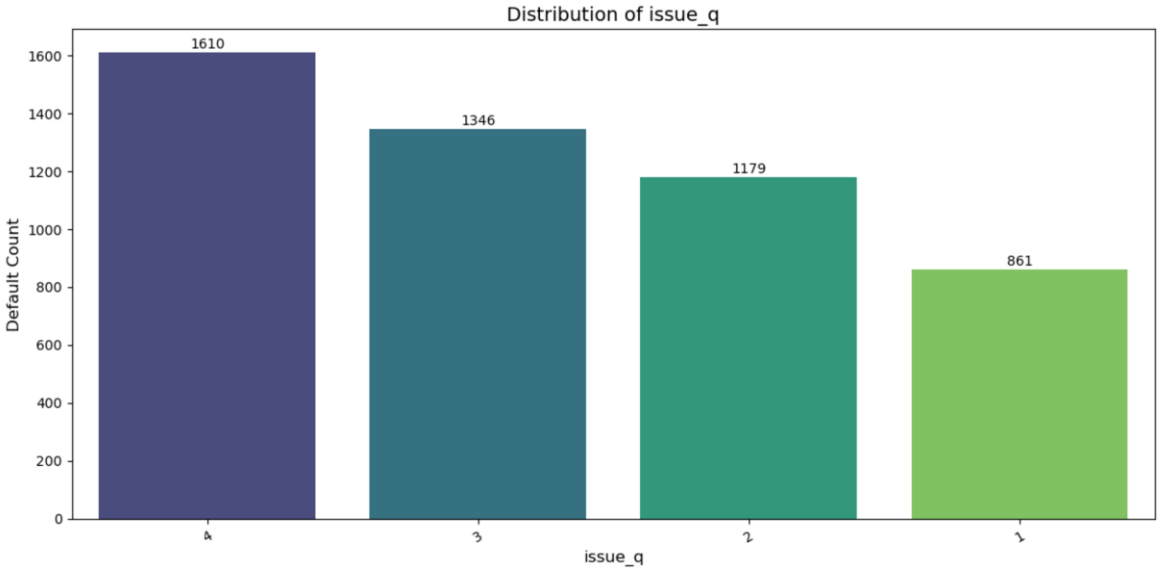


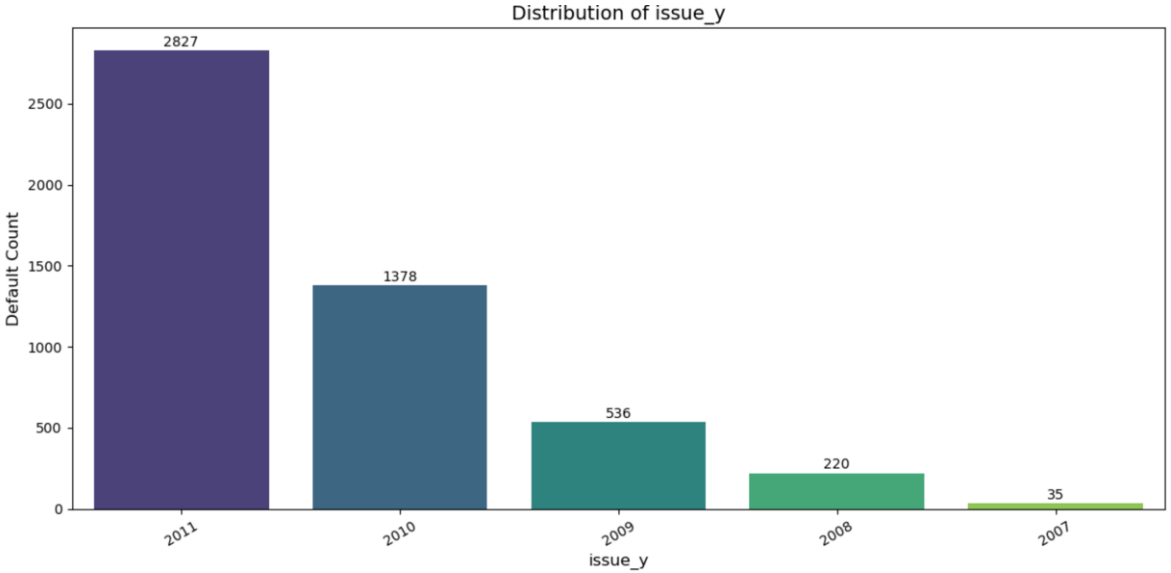
Annual Income - Borrowers with annual incomes between 0 and 40,000 USD showed the highest default rate, with 1,591 cases. Lenders should prioritize verifying applicants' income to ensure accuracy and mitigate the risk of defaults.



DTI - Borrowers with very high DTI ratios experienced 1,108 defaults, signaling the importance of stricter DTI thresholds in the loan approval process to minimize risk.

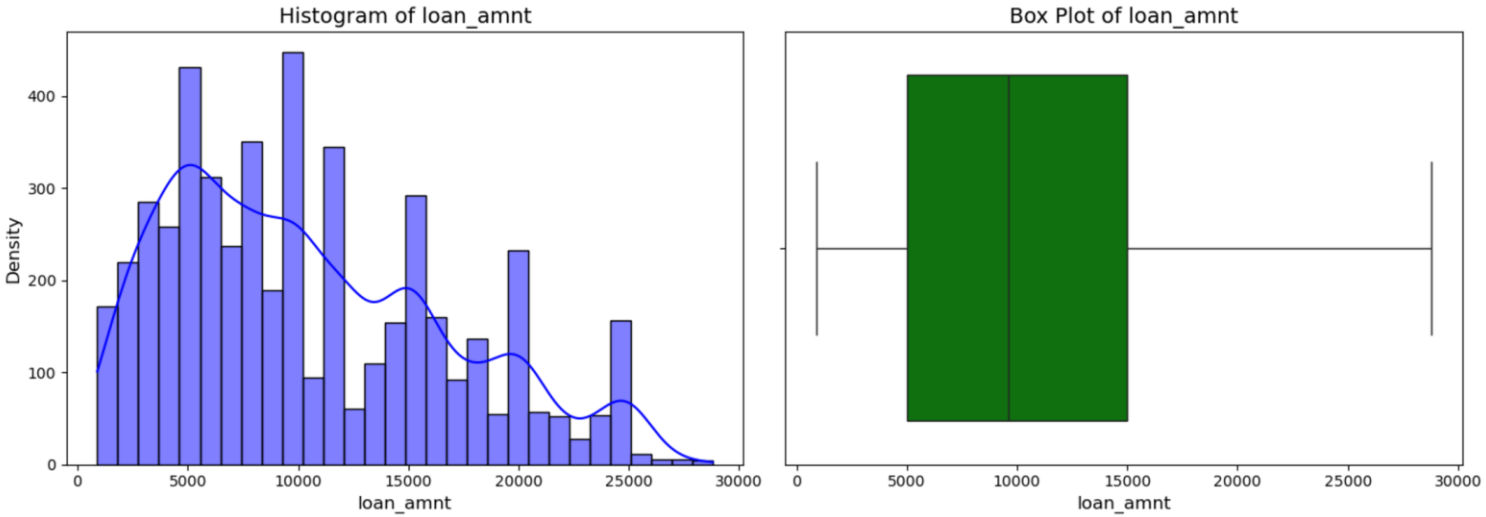




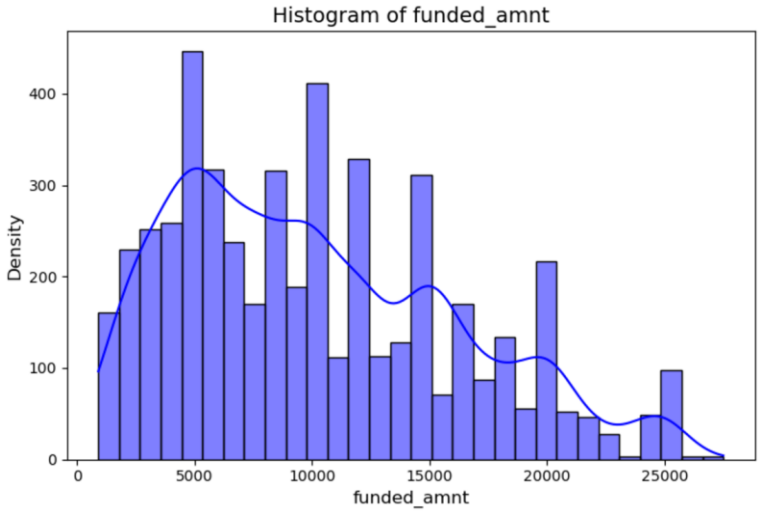


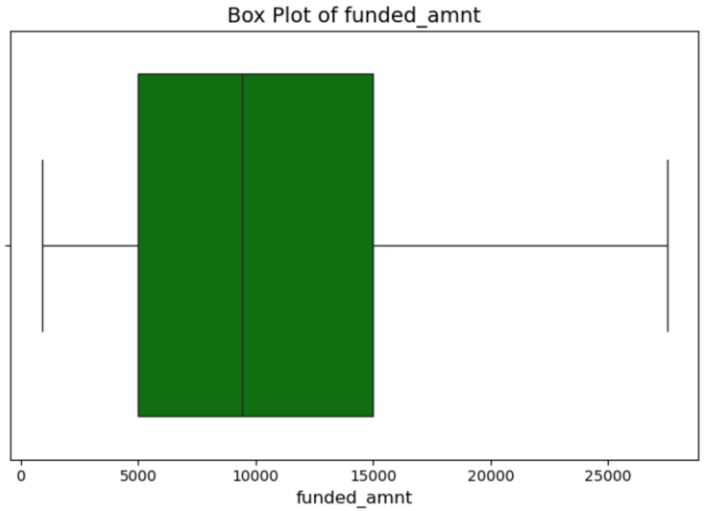
Issue Date - Seasonal Trends and Economic Factors: The fourth quarter had the highest defaults, with 1,610 cases, particularly in December, which saw 596 defaults. This trend could be linked to financial stress during the holiday season. The year 2011 recorded 2,827 defaults, a significant increase (105%) compared to 1,378 in 2010, possibly reflecting broader economic challenges.

Numerical Variables

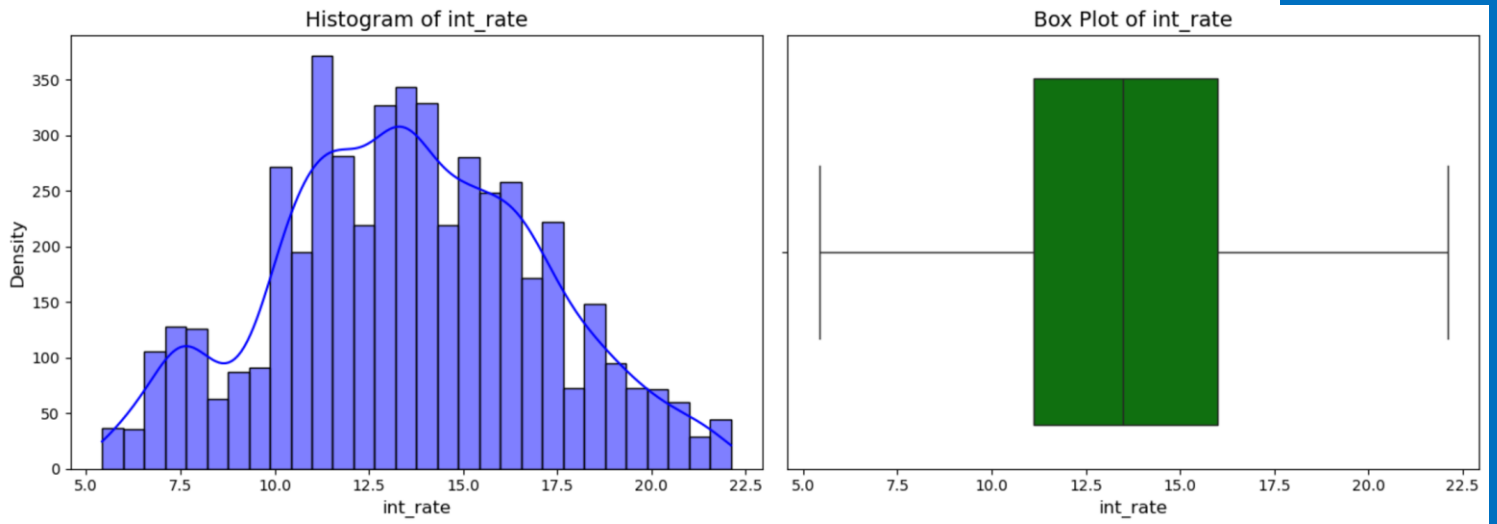


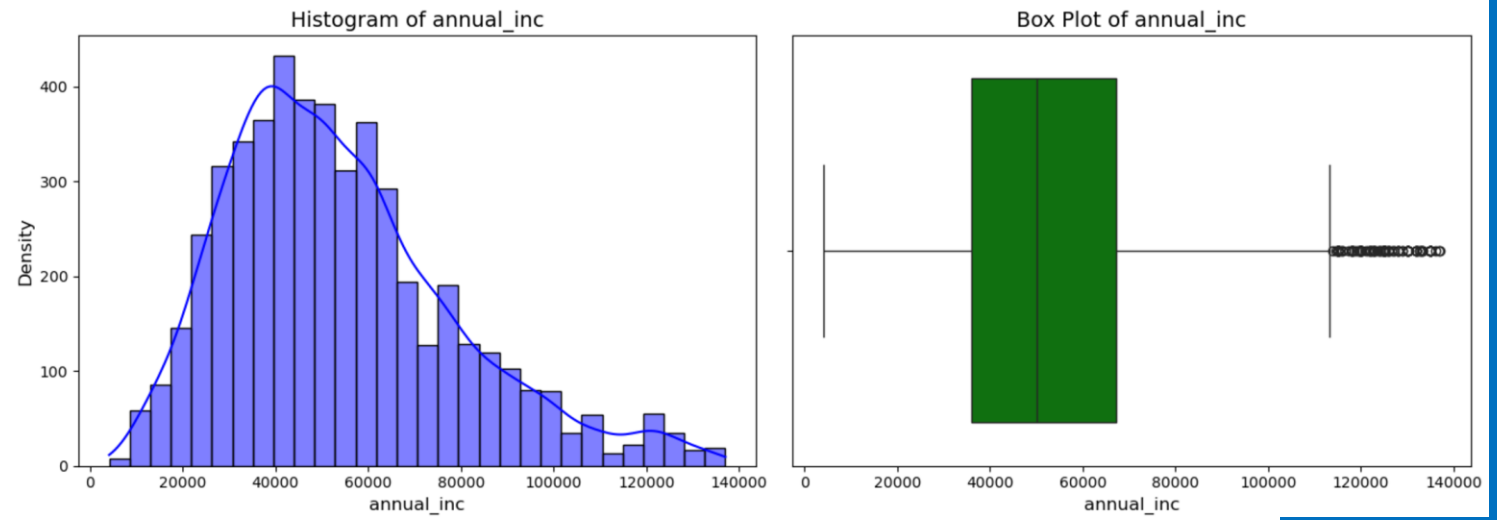
Loan Amount - Among charged-off loans, most loan amounts fall between 5,000 and 15,000 USD, suggesting that moderate-sized loans are more prone to default. The typical charged-off loan amount is around 10,000 USD, but there is significant variation in loan sizes among defaulters.





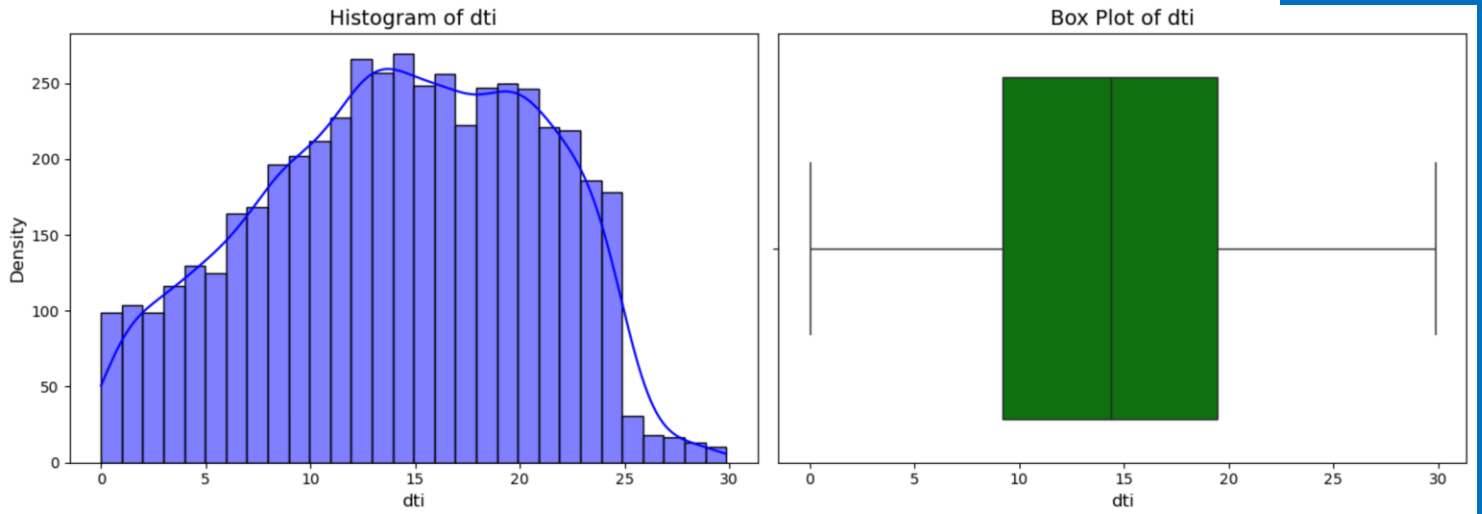
Funded Amount - Most charged-off loans have funded amounts between 5,000 and 15,000 USD, indicating increased default risk in this range. Funded amounts exceeding 25,000 USD are less commonly associated with defaults, possibly due to stricter approval processes or lower borrower demand.

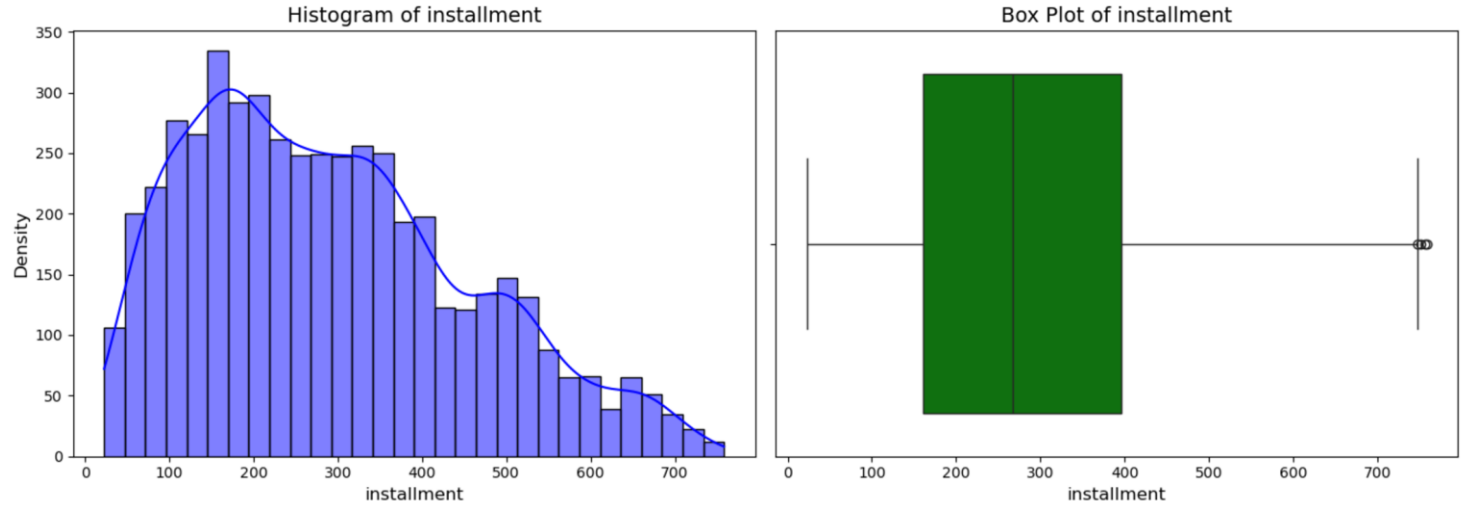




Interest rate - The majority of charged-off loans have lower interest rates, with most defaults occurring in the 10-15% range, while loans with rates above 20% are less common among defaults.

Annual Income - Borrowers with annual incomes between 0 and 40,000 USD experienced the highest default rates, indicating that lower-income borrowers are more likely to default; the majority of loans fall within this income range, with a few borrowers earning more than 120,000 USD.



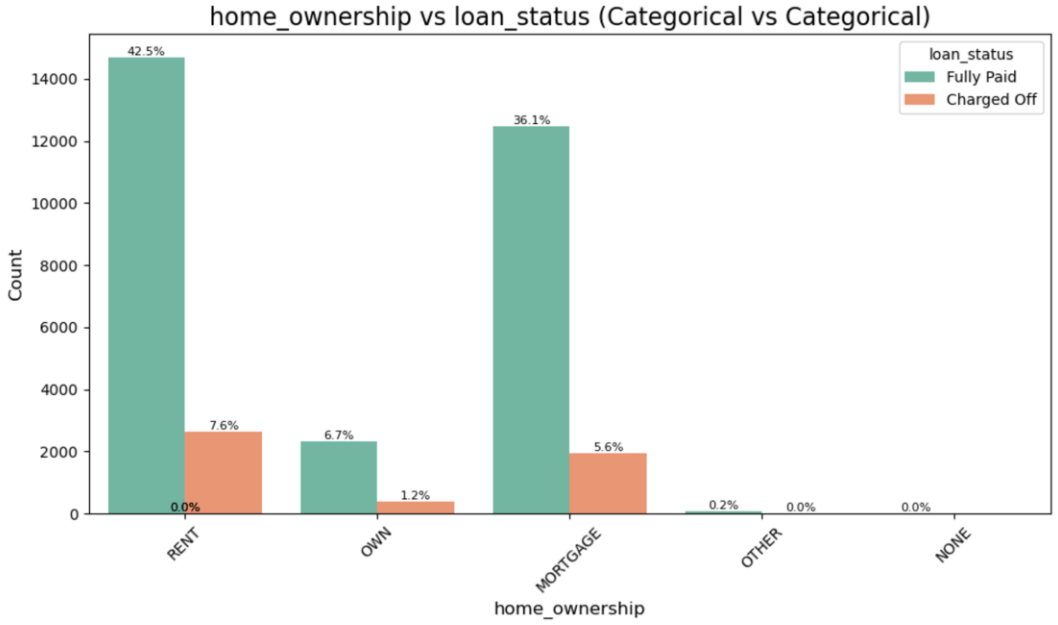


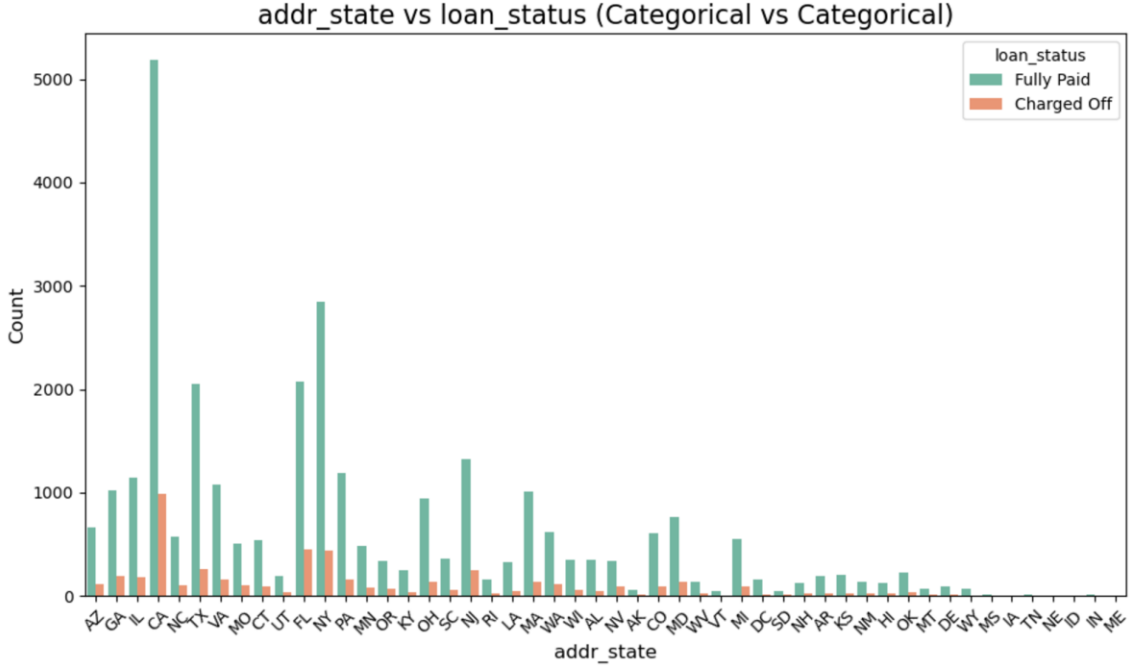
DTI - The majority of borrowers have lower DTI ratios, falling within the Very Low to Medium categories. Despite this, defaults are more concentrated among borrowers with Very High DTI ratios (above 20), indicating that borrowers with high debt relative to income are at greater risk of default.

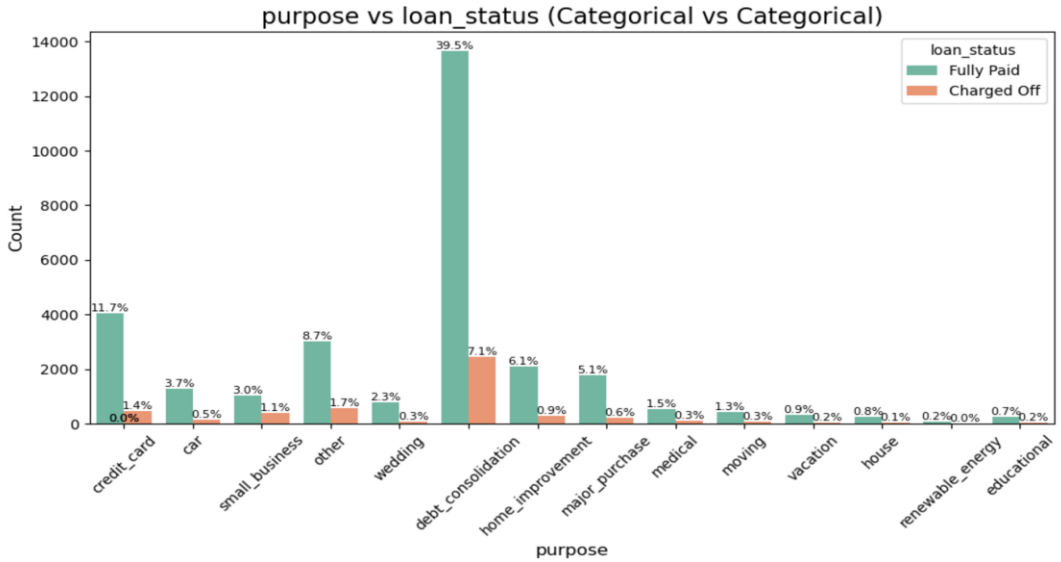
Installments - The majority of charged-off loans have smaller installments, typically ranging from 160 to 400 USD, with a few loans exceeding 600 USD. This suggests that while smaller installments are more common, the higher monthly payments associated with loans over 600 USD may place more financial strain on borrowers, potentially contributing to defaults

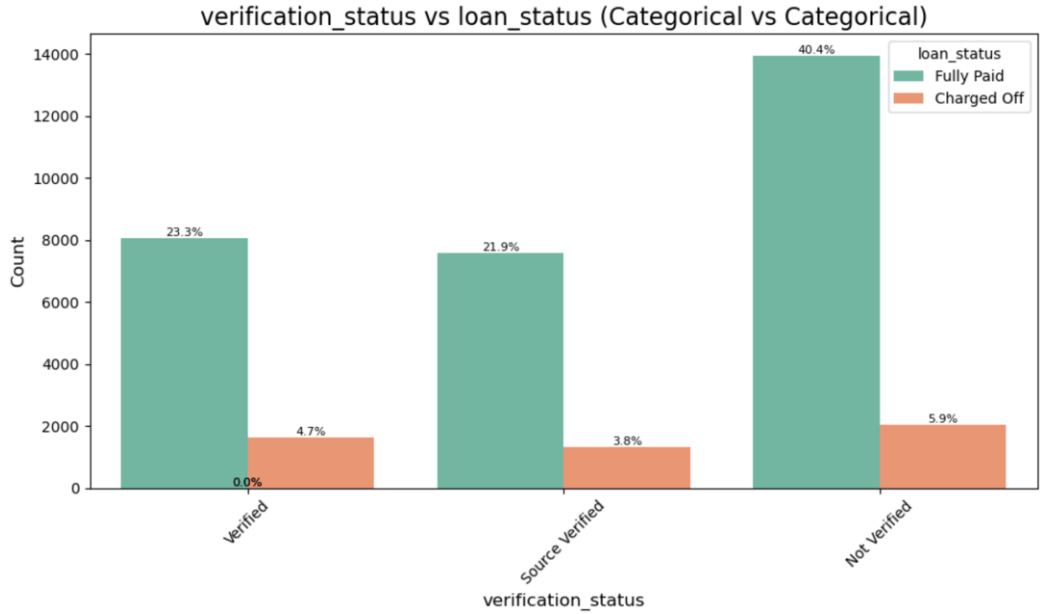
Bivariate Analysis

Categorical Variables vs Loan Status

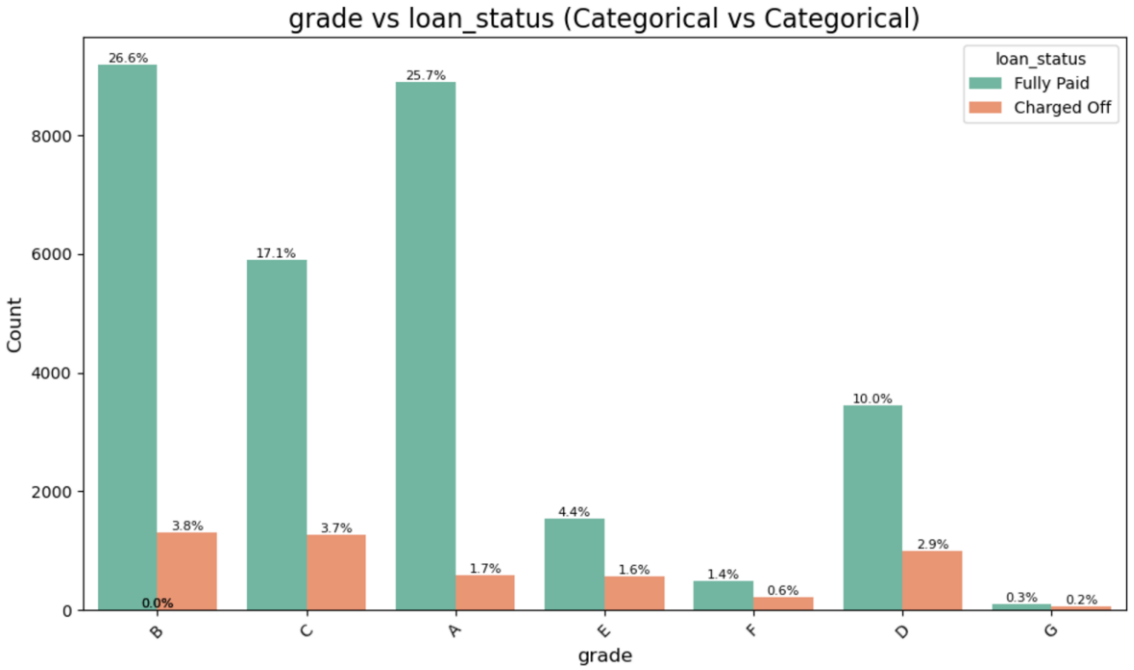


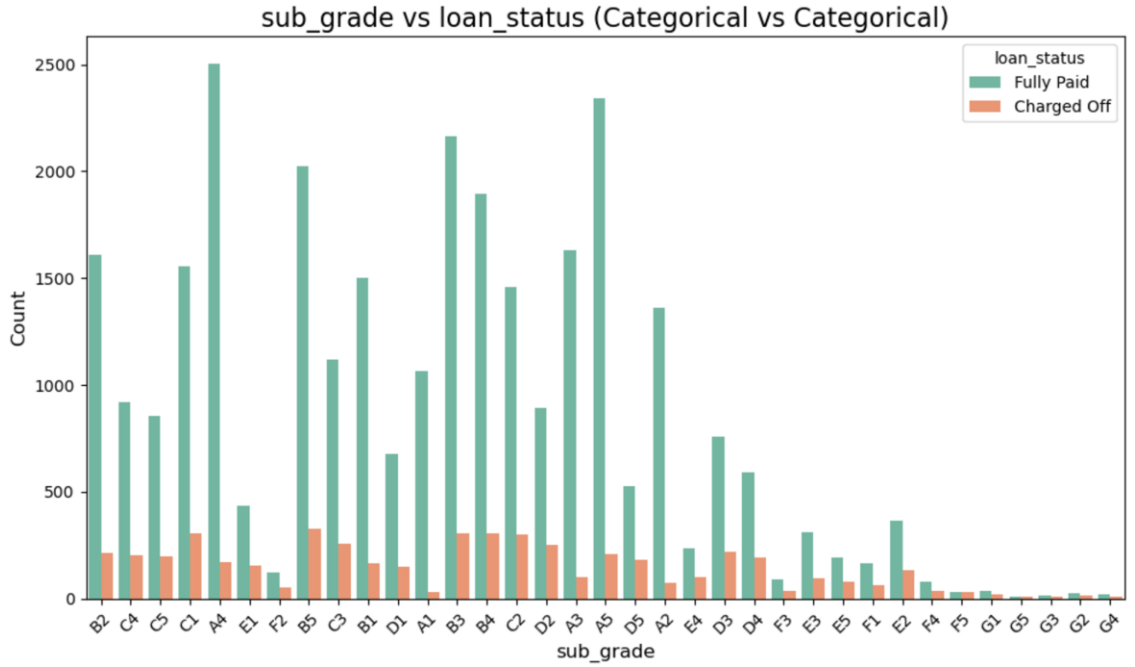


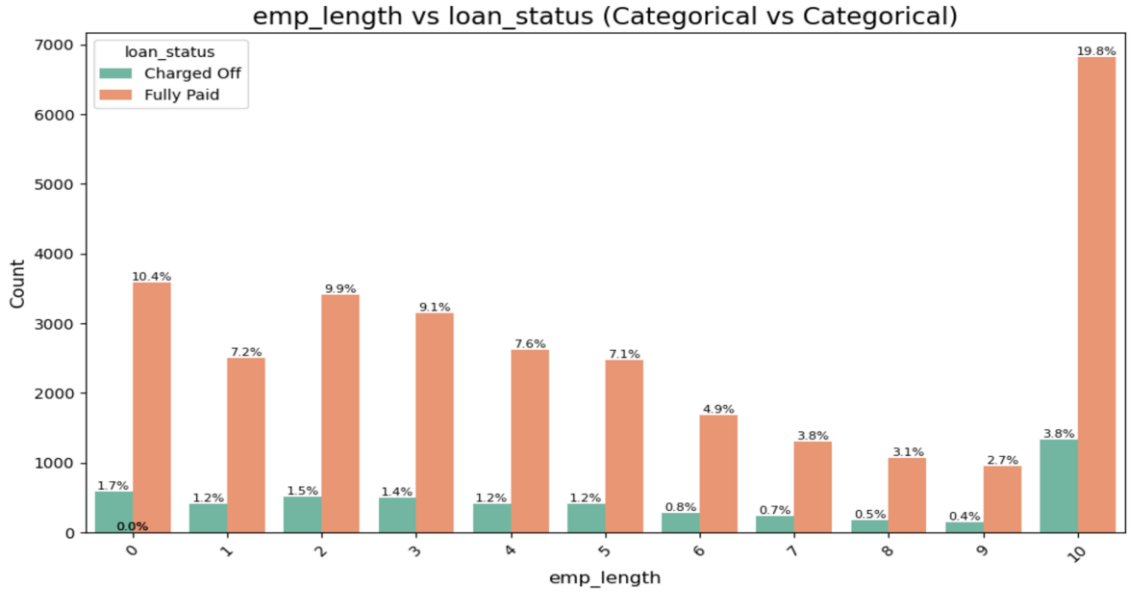


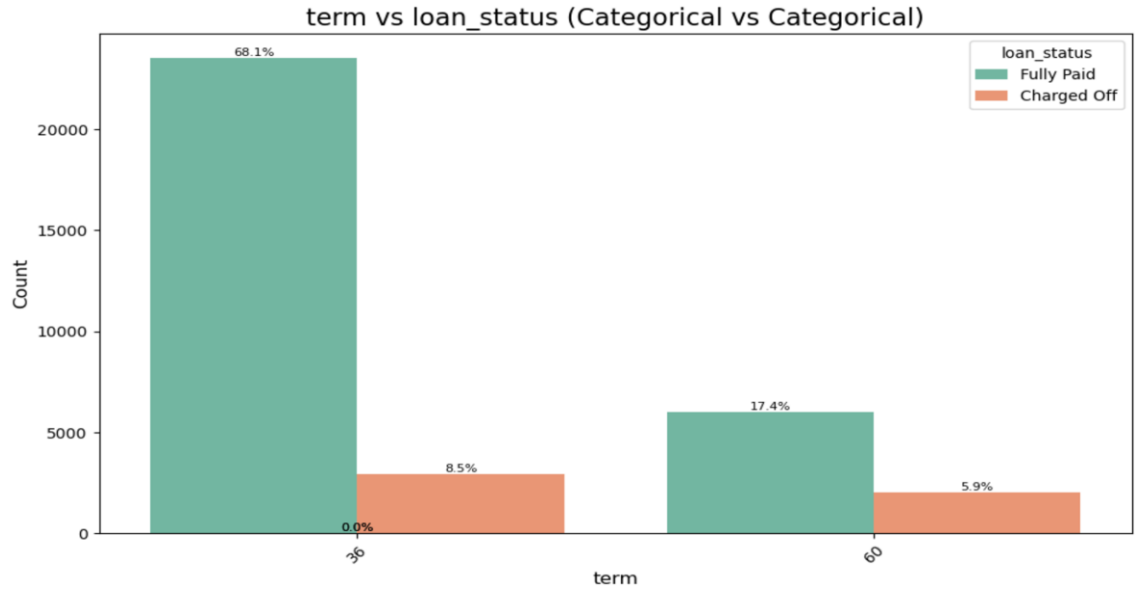


Categorical Variables vs Loan Status

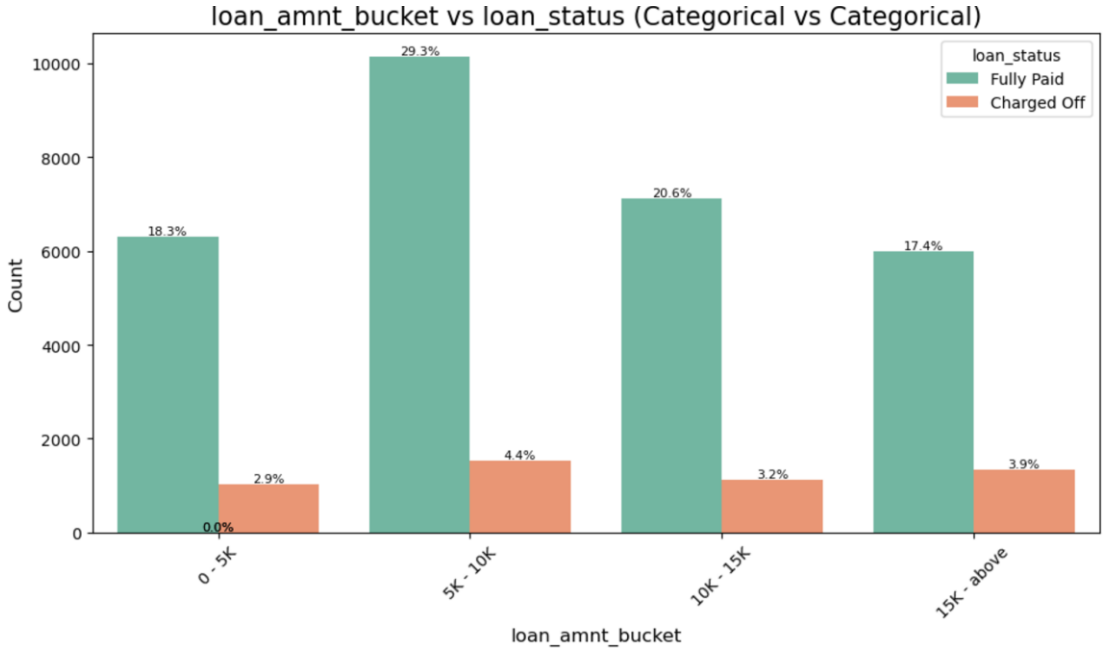


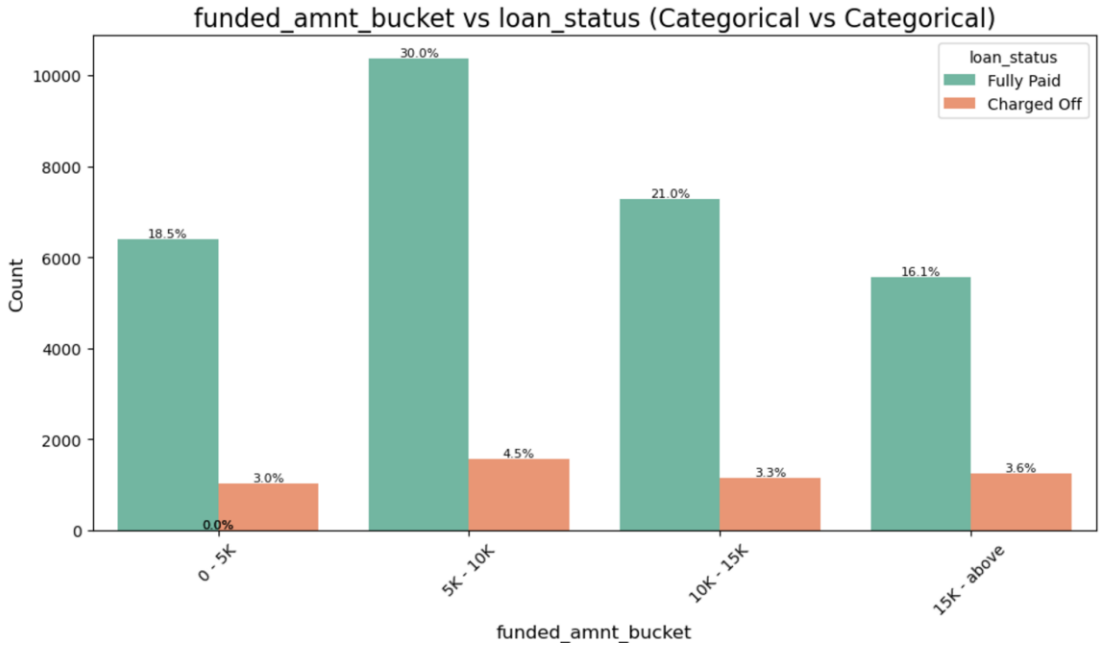


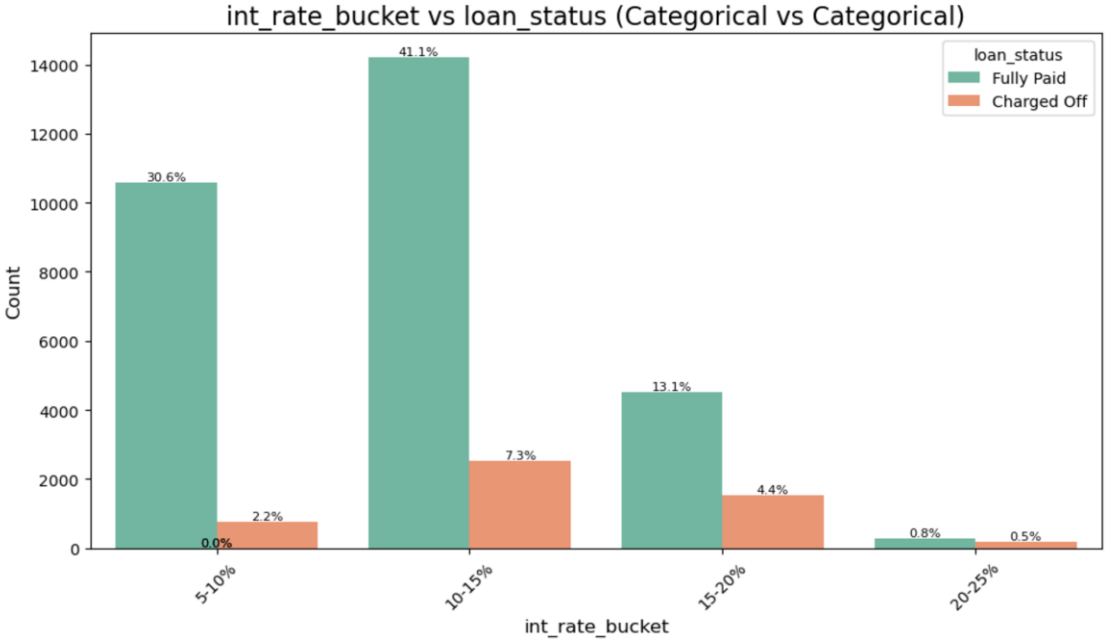


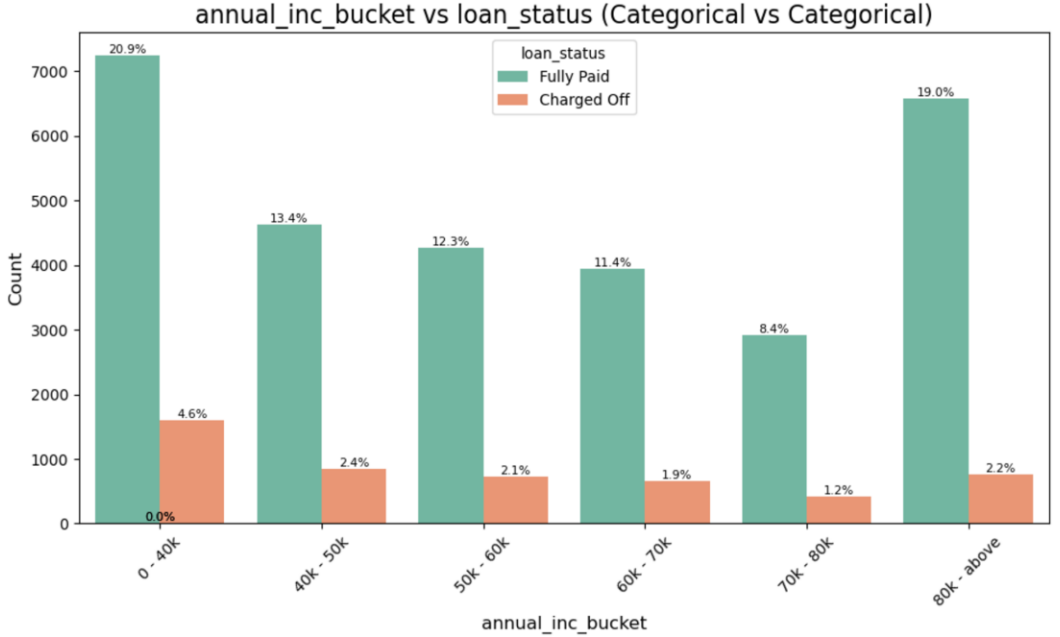


Categorical Variables vs Loan Status

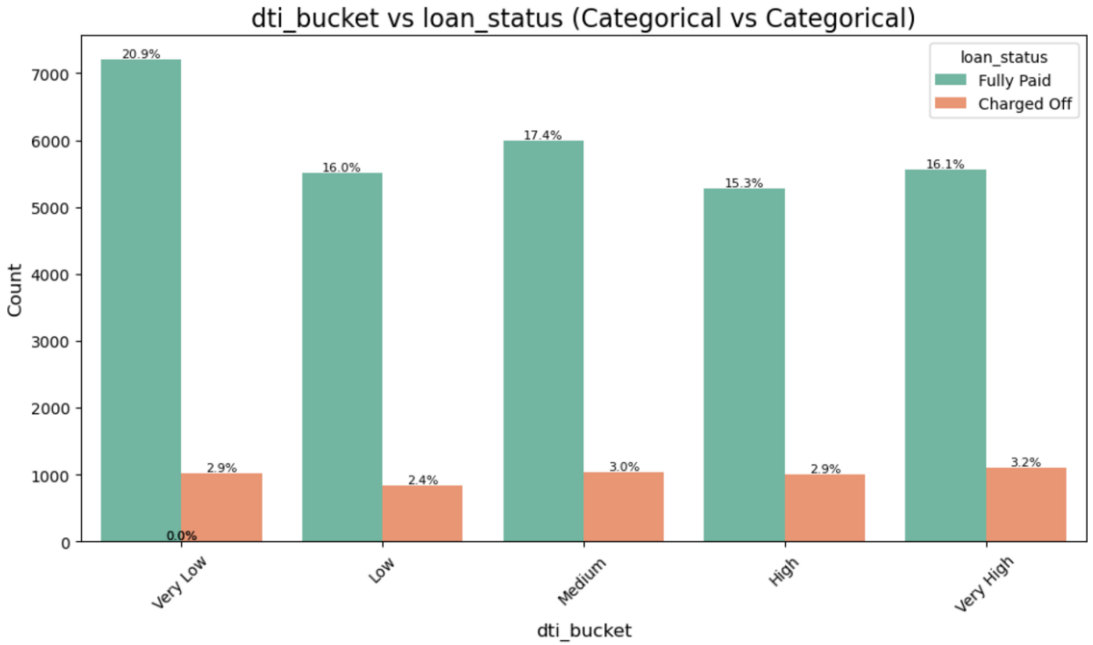


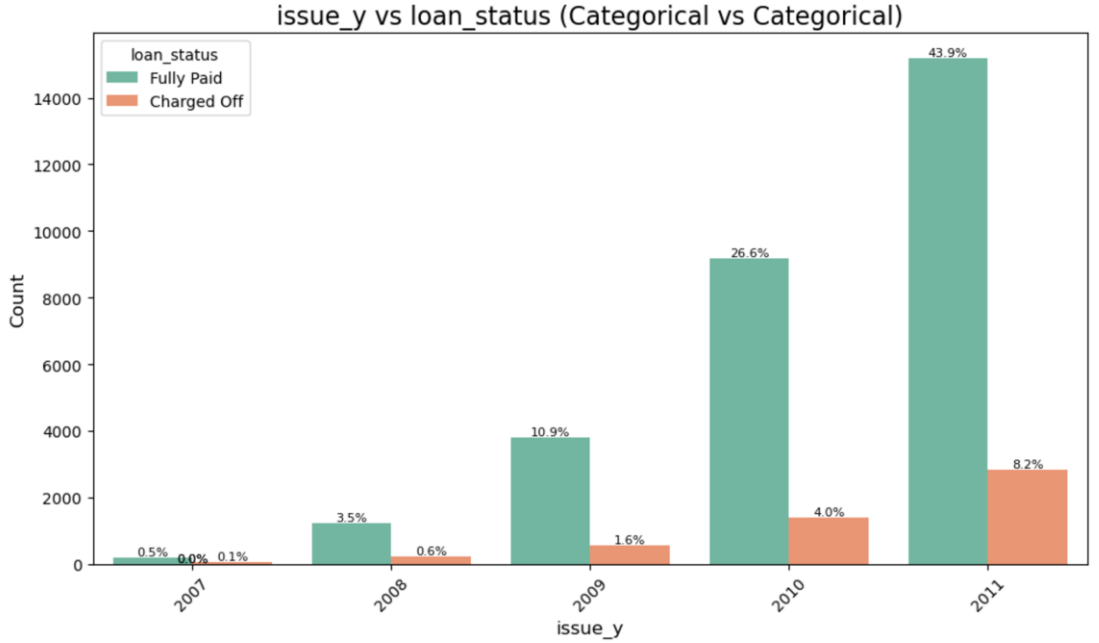


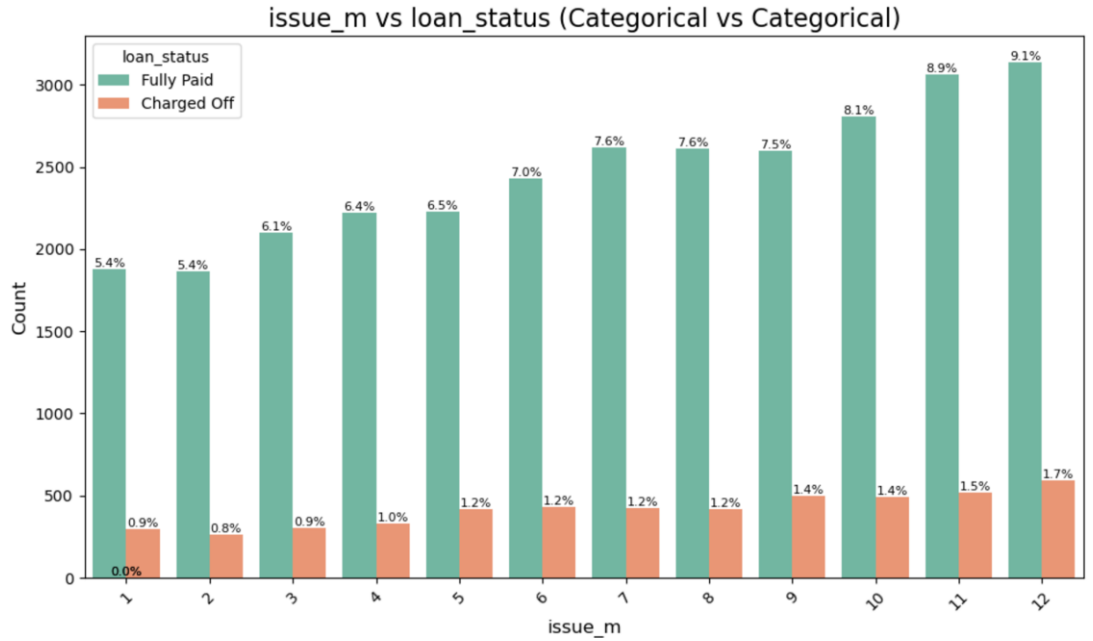


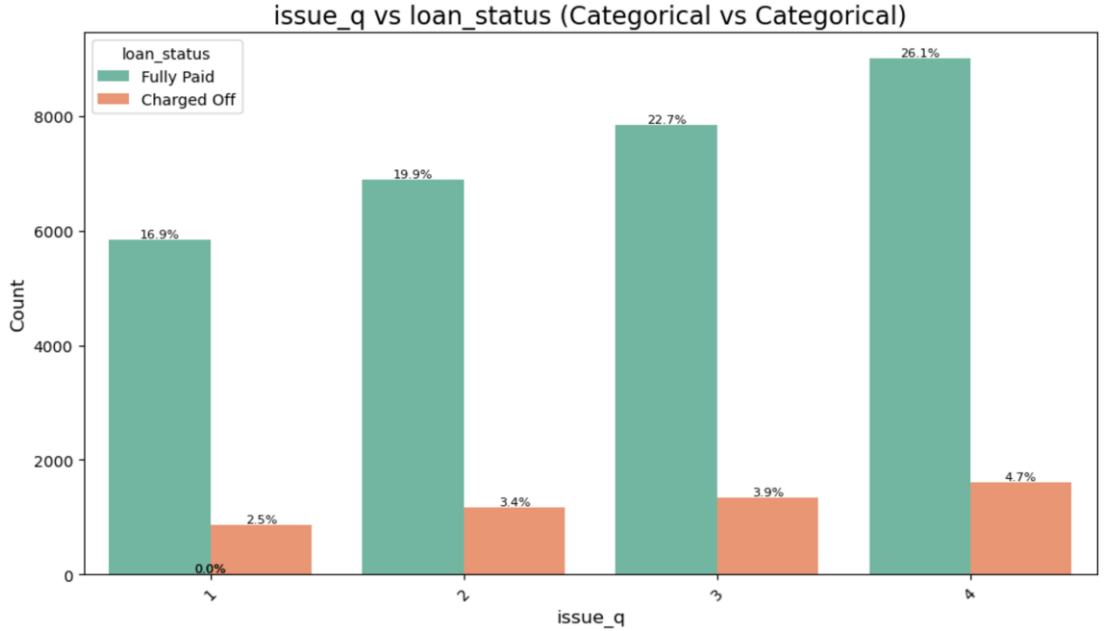


Categorical Variables vs Loan Status

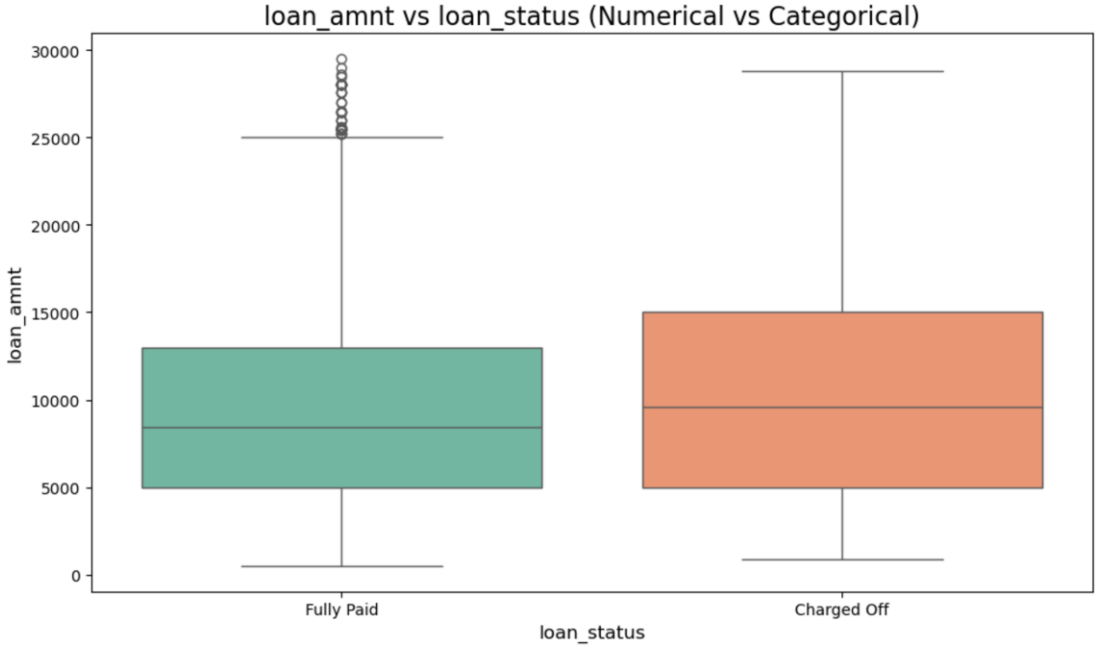


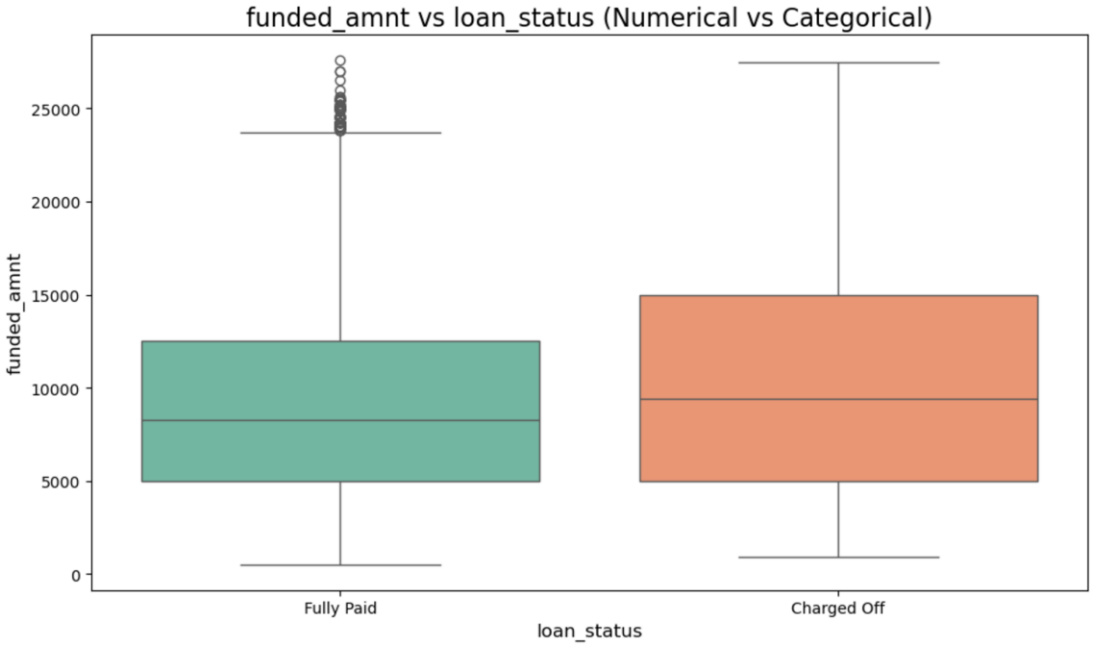


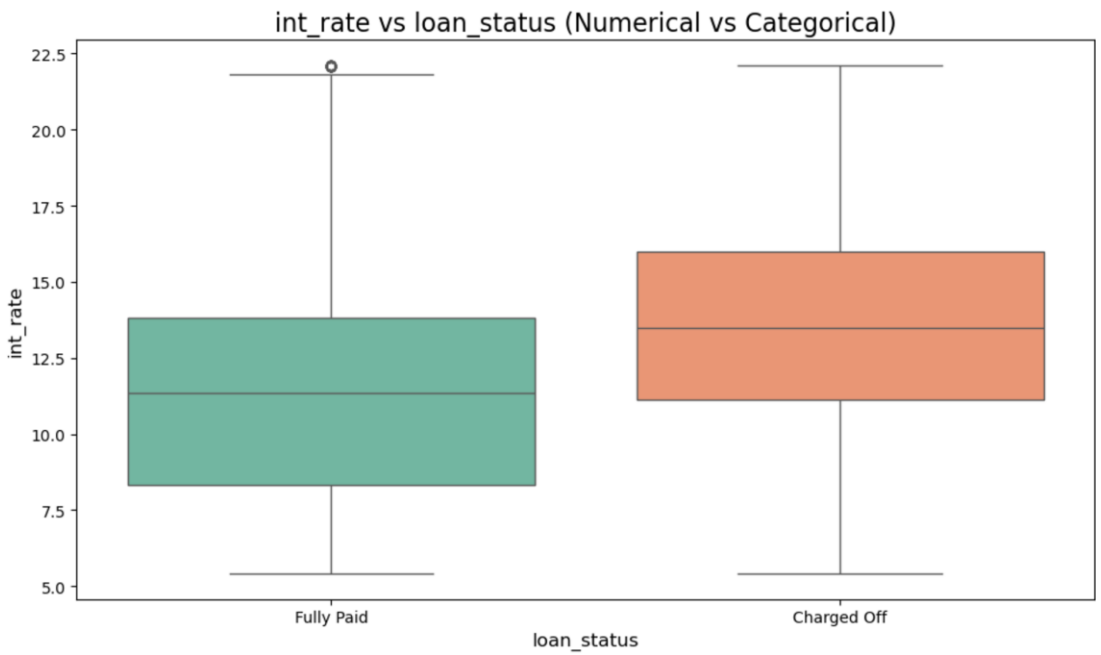


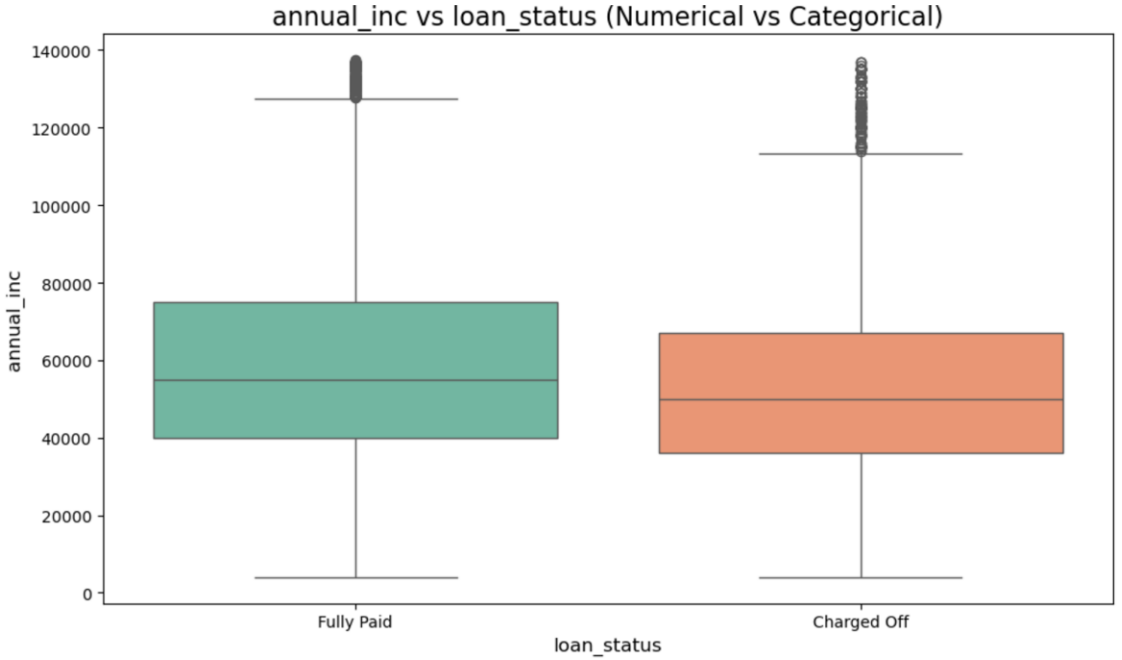


Numerical Variables vs Loan Status

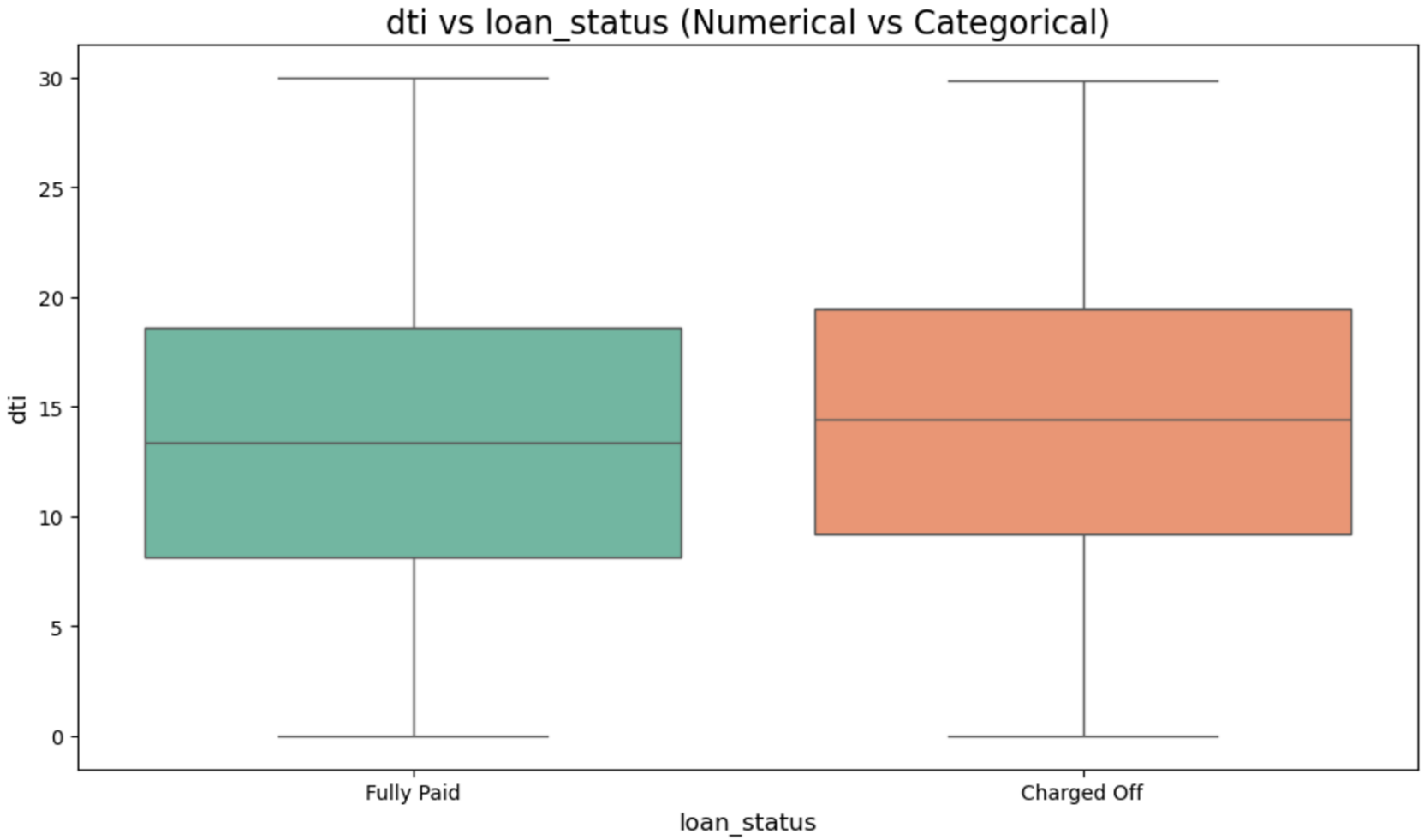




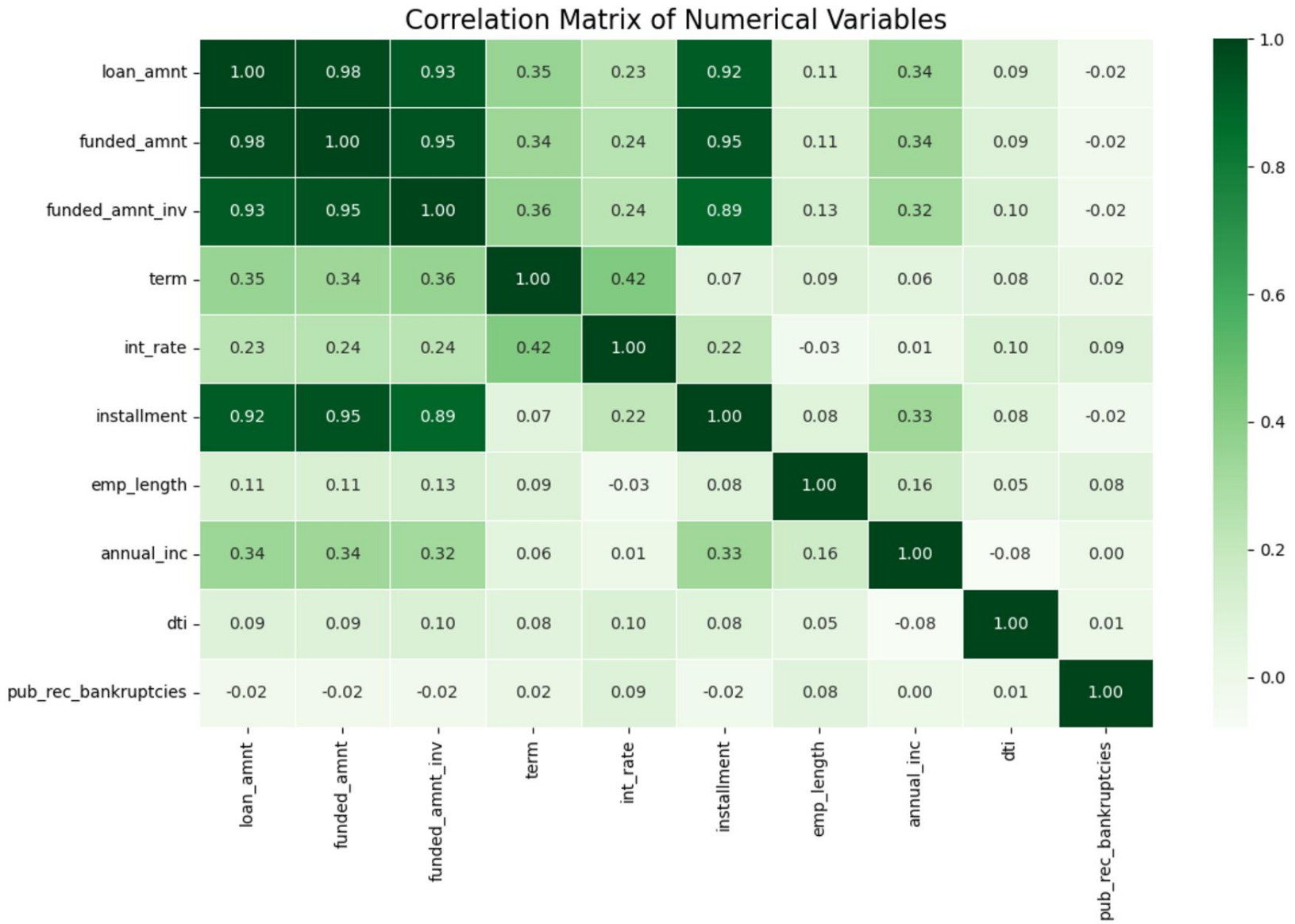




Numerical Variables vs Loan Status



Numeric vs Numeric Variables - Correlation Matrix



Observations:

Strong Correlation
installment has a strong correlation with funded_amnt, loan_amnt, and funded_amnt_inv

Moderately Strong Correlation
term has a moderately strong correlation with interest rate

Weak Correlation
1. *annual_inc* has a weak correlation with *loan_amount* and *funded_amnt*
2. *dti* has weak correlation with most of the fields
3. *emp_length* has weak correlation with most of the fields
4. *pub_rec_bankrupticies* has weak correlation with almost all of the fields

Bivariate Analysis: Inferences and Recommendations

1. Renters vs. Homeowners:
Inference: Renters, as opposed to homeowners, are more likely to default on loans. This may reflect the financial instability often associated with renting, such as a lack of long-term assets or potentially higher living expenses.
Recommendation: Lenders should implement stricter risk assessments for applicants who rent, including evaluating their overall financial stability more rigorously. Consider offering tailored loan products or repayment terms that are more suitable for renters, or increase interest rates for this group to offset higher risk.

2. Regional Risk Factors (States):
Inference: States like California, Florida, and New York have a higher likelihood of loan defaults. This could be influenced by regional economic conditions, such as higher cost of living or regional economic downturns.
Recommendation: Lenders should apply a region-specific risk model, with higher risk premiums for borrowers from states with higher default rates. Additionally, further analysis should be done to understand the underlying economic factors in these states, which could help tailor loan offerings or repayment terms.

3. Debt Consolidation:
Inference: Debt consolidation loans have the highest number of defaults. This suggests that borrowers seeking to consolidate their debt may already be in financial distress, potentially with high existing debt and limited repayment capacity.
Recommendation: Lenders should assess the total debt load and repayment capacity more thoroughly for debt consolidation applicants. Offering financial counselling services or evaluating the long-term sustainability of the borrower's financial situation can help reduce defaults in this category.

4. Income Verification:
Inference: Loans with "Not Verified" income have the highest default rate. This highlights the risk of inaccurate or misrepresented income information, leading to higher default likelihood.
Recommendation: Lenders should prioritize verifying applicants' income to ensure accuracy and mitigate the risk of defaults. Additionally, offering lower loan amounts or higher interest rates for borrowers with unverified income can help manage risk more effectively.

5. Loan Grades and Defaults:
Inference: Loan Grades B, C, and D contribute significantly to defaults. These borrowers may be at higher risk due to weaker credit histories or financial instability.
Recommendation: Lenders should introduce stricter underwriting criteria for applicants in these grades. Furthermore, it may be beneficial to offer more customized loan terms, such as higher interest rates or shorter loan durations, for applicants in these grades.
6. Loan Sub-Grades:
Inference: Sub-grades B3, B4, B5, C1, and C2 are especially associated with high default rates. These sub-grades may indicate borrowers with riskier profiles.
Recommendation: Loan applications within these sub-grades should be subjected to enhanced scrutiny, including a deeper analysis of financial health and debt-to-income ratios. Consider using alternative credit scoring models to assess the risk more accurately for these sub-grades.

7. Employment Duration:
Inference: Applicants with over 10 years of employment history have the highest defaults, followed by those employed for less than 1 year. A long employment history does not necessarily correlate with loan repayment capacity, suggesting that other financial factors should be prioritized.
Recommendation: Lenders should prioritize income stability and debt-to-income ratios over employment duration for better risk assessment.

Bivariate Analysis: Inferences and Recommendations

8. Short-Term Loans:
Inference: Short-term loans (36 months) have a higher default rate, possibly due to the higher monthly payment burden.
Recommendation: Lenders should consider offering more flexible repayment options for short-term loans, such as extending the loan term or adjusting the payment schedule, to reduce borrower strain and defaults. Additionally, implementing more rigorous financial assessments for short-term loan applicants can help minimize risk.

9. Loan Amounts:
Inference: Loan amounts between 5,000 and 10,000 USD have the highest default rates, followed by larger loans. This suggests that the loan amount range may correspond to a threshold where borrowers struggle with repayment.
Recommendation: Loans in this amount range should be more carefully scrutinized, particularly for applicants with weaker financial profiles. Offering smaller loan amounts or longer repayment terms could help mitigate defaults in this category.

10. Funded Loans and Defaults:
Inference: Funded loans in the 5,000 – 10,000 USD range also experience the highest default rates, aligning with the loan amount bucket.
Recommendation: Lenders should focus on improving funding evaluations for loans in this range, ensuring that borrowers' financial situations are thoroughly assessed. Offering more flexible repayment terms or adjusting the loan approval process for this range may help reduce defaults.

11. Interest Rates:
Inference: Interest rates in the 10-15% and 15-20% ranges are associated with the highest defaults, suggesting that these rates may not effectively deter high-risk borrowers.
Recommendation: Lenders should consider more dynamic interest rate structures, possibly offering lower rates to borrowers with stronger financial profiles while adjusting the rates based on riskier factors like credit score and debt-to-income ratio.

12. Income and Defaults:
Inference: Borrowers with annual incomes between 0 and 40,000 USD exhibit the highest default rates, indicating that lower income levels correlate with higher default risk.
Recommendation: Lenders should place more emphasis on assessing the financial capacity of borrowers with lower incomes. Offering smaller loan amounts or providing tailored repayment options, such as income-based repayment plans, can help mitigate defaults in this group.

13. Debt-to-Income Ratios (DTI):
Inference: Borrowers with very high DTI ratios are more likely to default, indicating that higher debt relative to income is a key predictor of loan default.
Recommendation: Lenders should establish stricter DTI ratio limits during the underwriting process to reduce the likelihood of defaults. Implementing policies that prevent high DTI borrowers from accessing larger loans or offering counseling for debt management can also help mitigate risk.

14. Seasonal Trends and Economic Factors:
Inference: The fourth quarter, particularly December, sees the highest defaults, which may be linked to increased financial strain during the holiday season. Additionally, the 105% increase in defaults in 2011 suggests broader economic challenges during that period.
Recommendation: Lenders should anticipate higher default rates during the fourth quarter and adjust their lending strategies accordingly, possibly by offering more flexible repayment terms or temporary relief options. Monitoring broader economic trends and adjusting risk models to account for seasonal fluctuations could help reduce defaults during these peak times.

THANK YOU