Bharghav R & Ishan Bhowmik

Lending Club Case Study

Problem Statement

- Lending Club is a financial services company specializing in urban consumer lending. They offer a diverse portfolio of loan products, including personal loans, business loans, and medical expense financing. Through their online platform, borrowers can access quick loan approvals at competitive interest rates.
- While many clients successfully repay their loans, Lending Club faces challenges with loan defaults. A significant portion of borrowers struggle to meet their repayment obligations, leading to financial difficulties for both the company and the individuals involved.
- To address this issue, Lending Club is conducting a comprehensive analysis to identify the root causes of loan defaults and develop effective strategies to mitigate this risk. Their goal is to enhance their lending practices, improve borrower outcomes, and maintain a sustainable business model.

Data quality

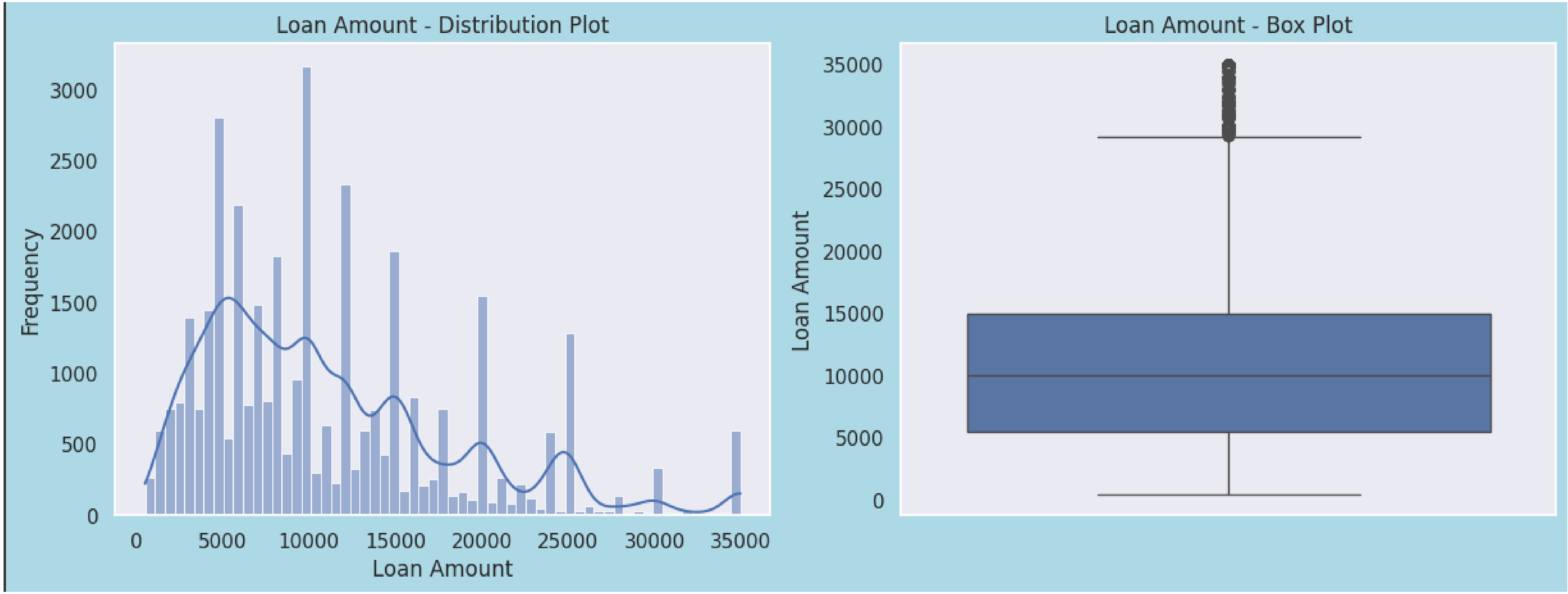
- Conducted a thorough examination of missing values across all columns and calculated their percentages to prioritize cleaning efforts.
- Removed columns with more than 30% missing values to maintain the dataset's integrity and analytical reliability.
- Identified and eliminated columns with single unique values and removed redundant columns, including borrower metadata and non-contributing variables.
- Standardized text-based columns by trimming whitespace and ensuring consistent formatting to prevent data entry errors.
- Converted columns containing date values from string to datetime format and extracted year, month, and day components for more granular temporal analysis.

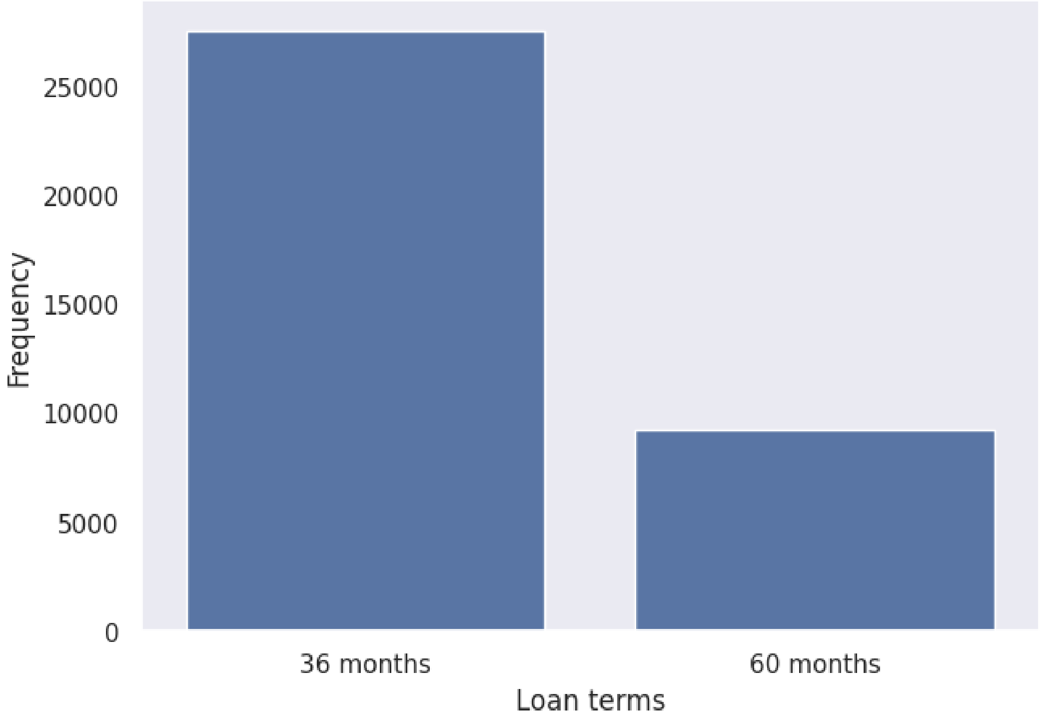
Data Cleaning

- Eliminated irrelevant columns such as member_id, url, title, emp_title, zip_code, etc to focus on the most impactful features for predicting loan defaults.
- Converted columns containing percentage values (int_rate and revol_util) from string to float by removing the '%' character, and ensured all date columns are appropriately converted to datetime format.
- Removed rows containing null values in critical columns where imputation was not feasible, ensuring the remaining dataset is complete and reliable.
- Created a new feature, approved_loan_amnt_percentage, representing the ratio of approved loan amount to requested loan amount, providing additional insight into loan approval patterns.
- Rechecked for any remaining null values and verified the data types of all columns to ensure they are appropriate for analysis.

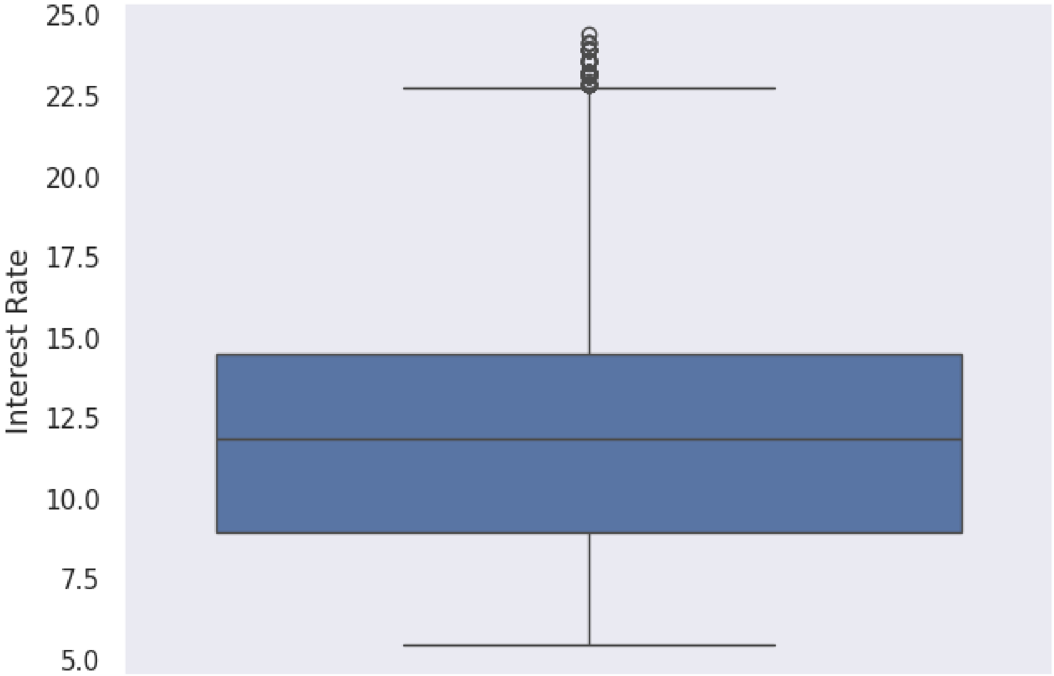
Univariate Analysis

- The data reveals a clear preference for loan amounts around 10,000, with this being the most common choice. Significantly fewer borrowers opt for either the minimum (500) or maximum (35,000) loan amounts available.

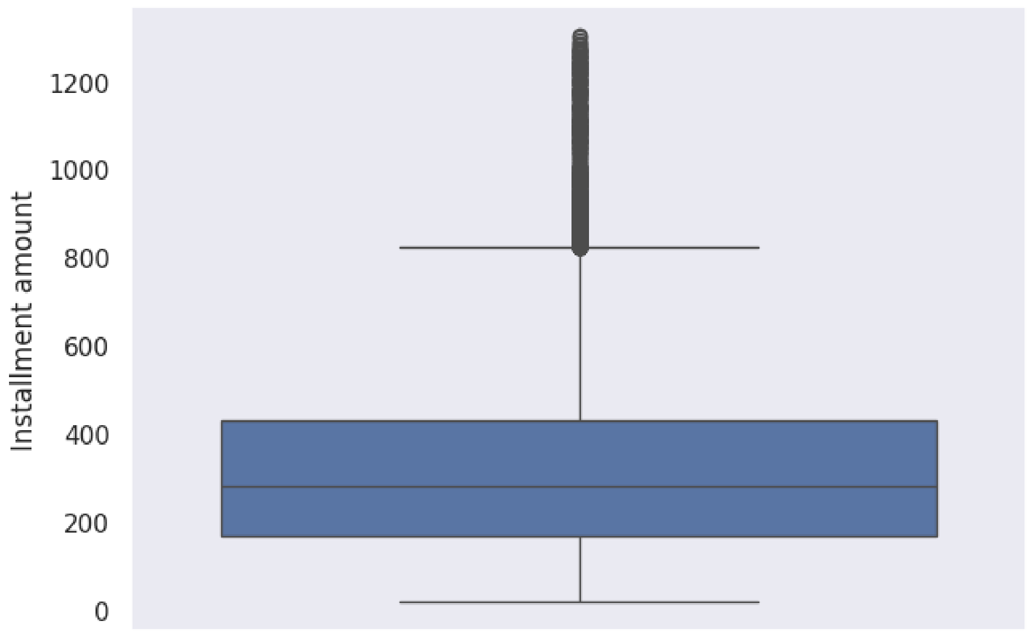


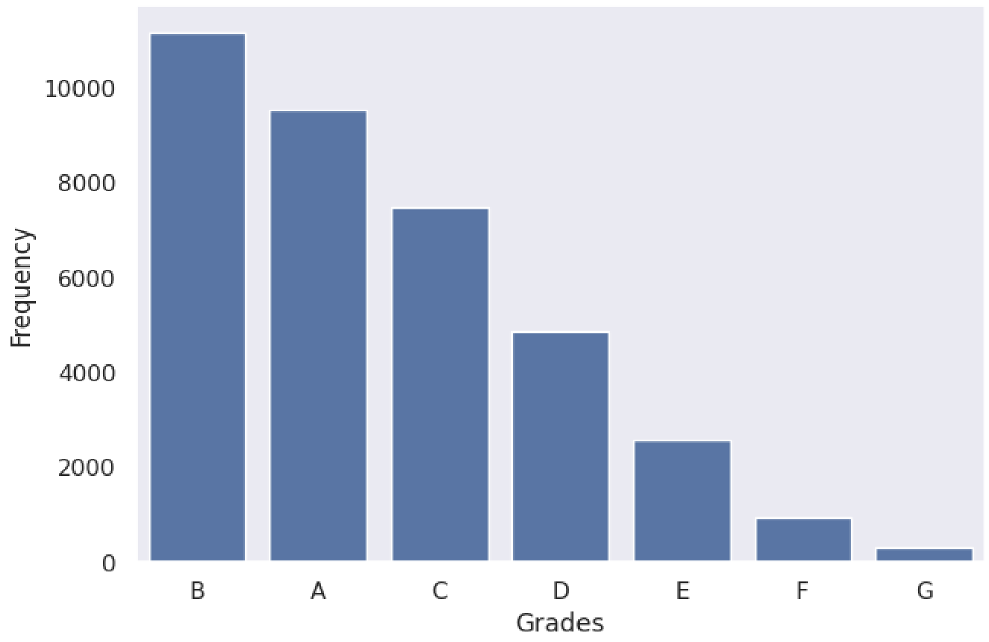


Borrowers overwhelmingly favor shorter loan terms (36 months), likely due to a desire to pay off their debts more quickly and potentially pay less interest overall.



The boxplot and descriptive statistics show outliers with interest rates exceeding 22%, indicating that a small portion of borrowers face significantly higher borrowing costs. This could be due to various factors such as lower credit scores, riskier loan types, or specific loan terms.

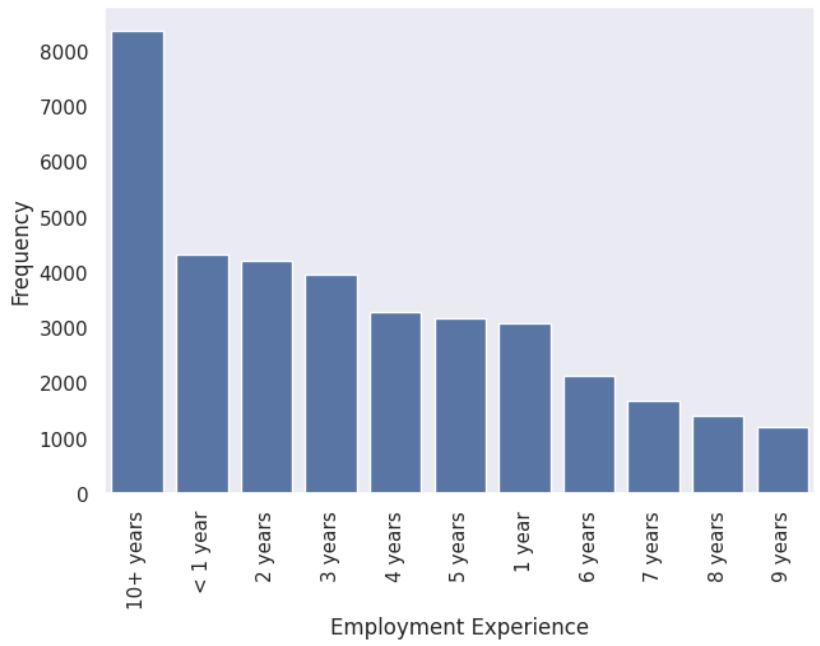




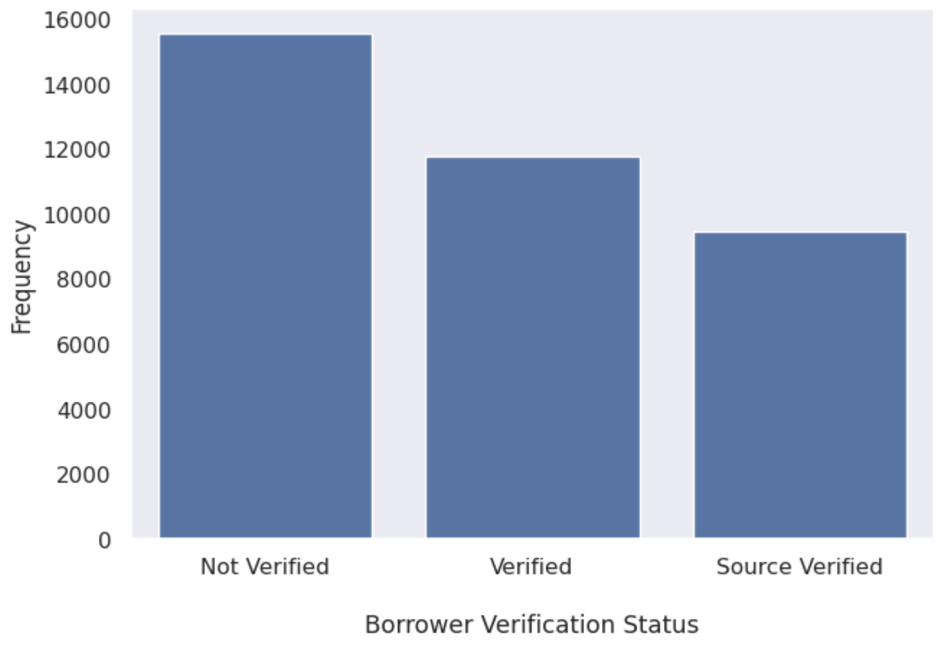
The majority of borrowers have monthly installment payments between 167.73 and 429.36, suggesting that these amounts are generally manageable for most individuals.

A small portion of borrowers face significantly higher monthly payments (up to 1305), likely due to larger loan amounts, shorter terms, or higher interest rates. These individuals could be at higher risk of default and require closer scrutiny.

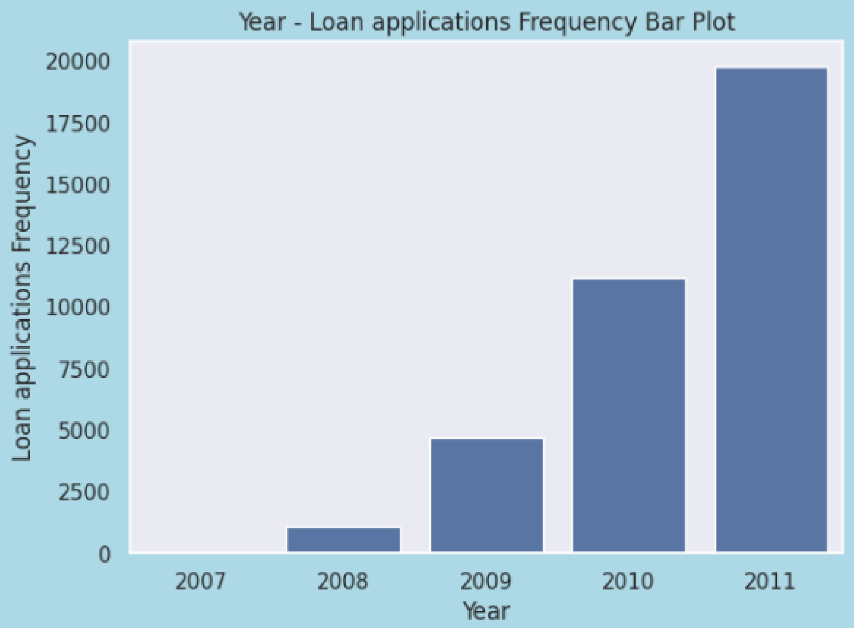
The data indicates that 56% of borrowers fall within the top two credit grades (A and B). This is a positive sign for the lending platform, as it suggests a significant portion of borrowers are considered low-risk and more likely to repay their loans on time. This aligns with the finding that a substantial share of borrowers are creditworthy.

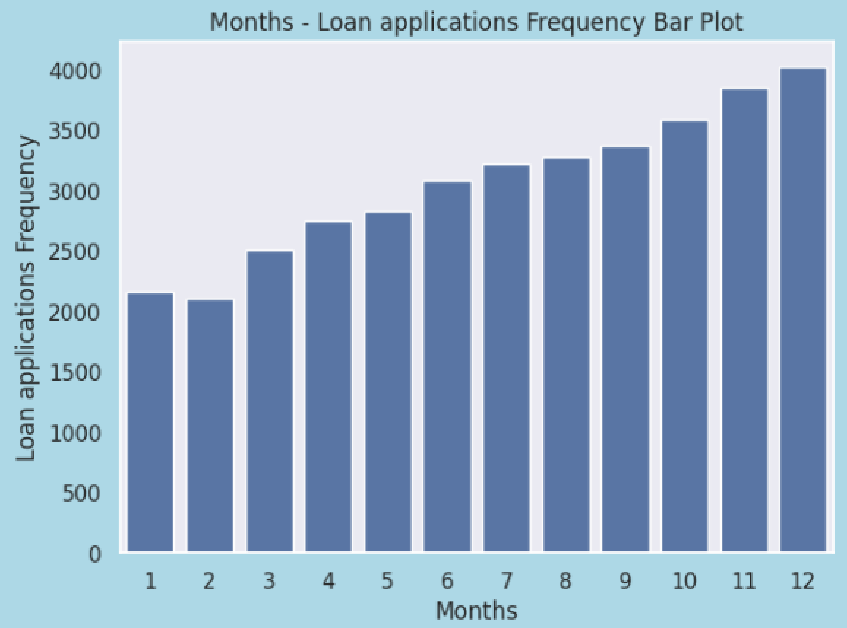


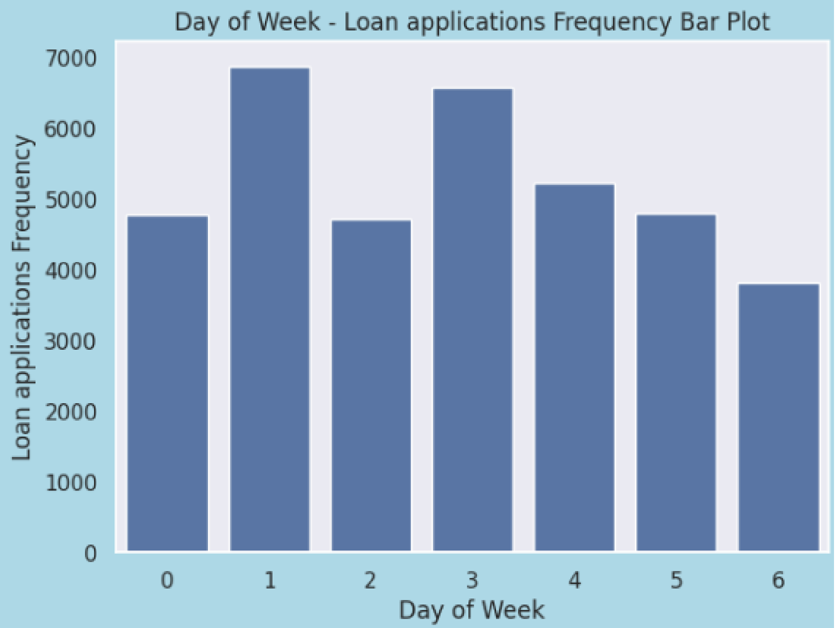
The graph shows most borrowers have a lot of work experience (10+ years), which is good because they are probably stable in their jobs and can pay back the loan.



The largest segment of borrowers (42%) falls under the "Not Verified" category, meaning their stated income hasn't been confirmed. This raises significant concerns about the accuracy of their financial information and their ability to repay loans.







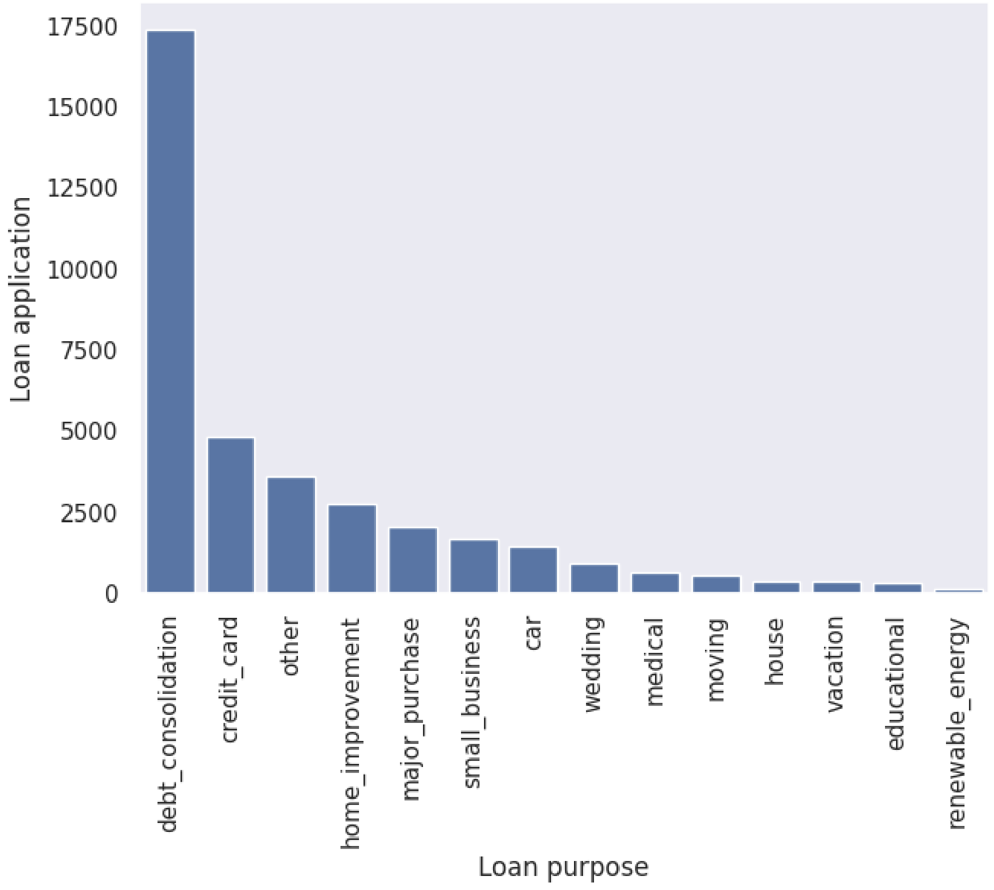
Loan applications experienced a significant surge between 2007 and 2011, nearly doubling each year. This indicates a growing demand for loans during this period and suggests a potential expansion of the lender's customer base or increased marketing efforts.

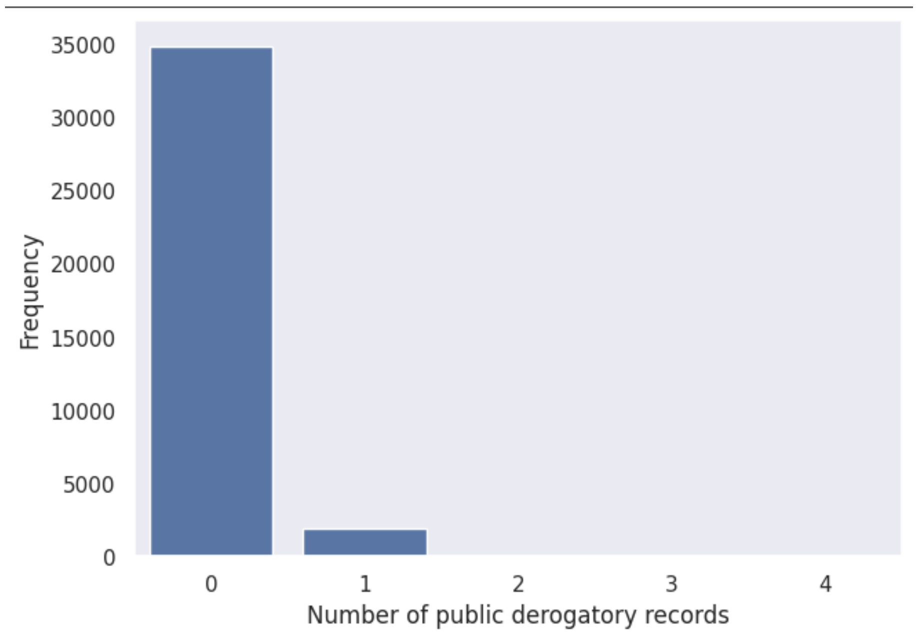
There is a clear seasonal pattern in loan applications, with a steady increase from the beginning of the year and peaking in the last quarter (October to December). This could be attributed to various factors like holiday spending, year-end financial planning, or specific promotional campaigns.

Loan applications are most frequent on Mondays and Wednesdays, followed by a decline over the rest of the week. This suggests borrowers may prefer to initiate loan applications at the beginning and middle of the work week, possibly due to convenience or aligning with financial planning cycles.

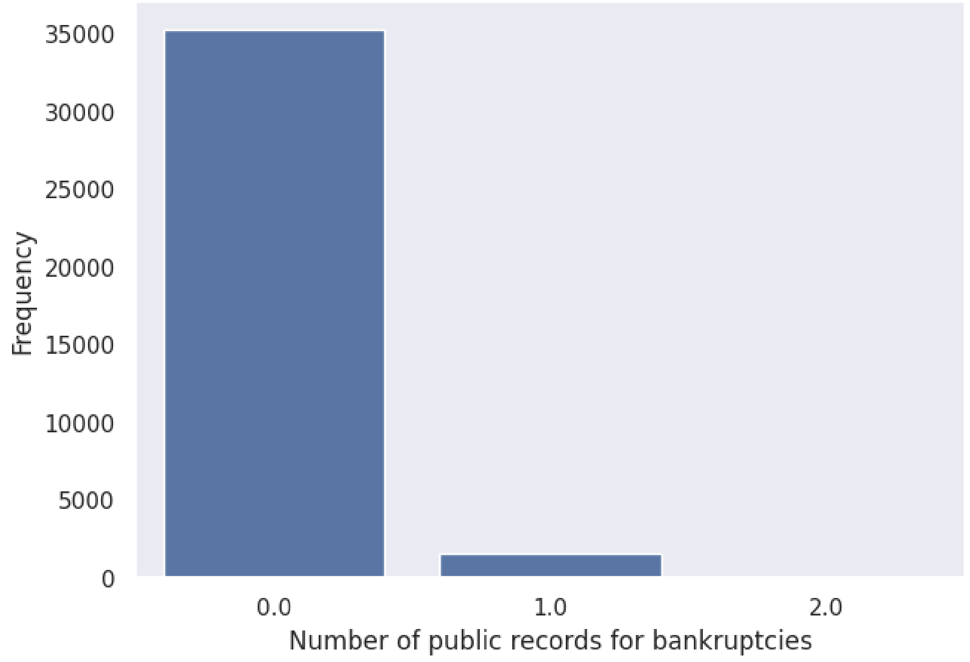
- The chart overwhelmingly demonstrates that borrowers primarily seek loans for debt consolidation (47%), highlighting the prevalence of managing existing debts as a key financial concern. This suggests that many individuals may be struggling with multiple debts and are seeking to simplify their repayment by consolidating them into a single loan.

- While debt consolidation is the leading purpose, credit card debt emerges as a substantial secondary reason (13%). This emphasizes the widespread reliance on credit cards and the potential challenges individuals face in managing their credit card balances. This could indicate a need for financial education or alternative credit solutions to address the issue of credit card debt.





The vast majority of borrowers (95%) have no public derogatory records, signifying a low-risk borrower pool and a positive outlook for loan repayment.



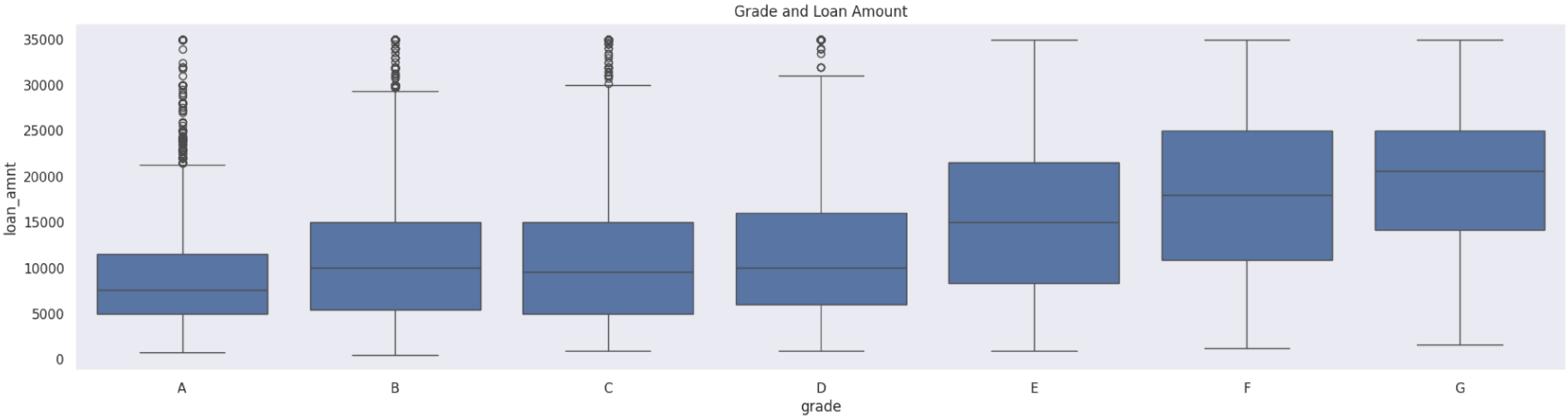
The vast majority of borrowers (approximately 96%) have no record of bankruptcy. This demonstrates a high degree of financial responsibility and stability among the borrower pool, suggesting a lower risk of default.

Conclusion from univariate analysis

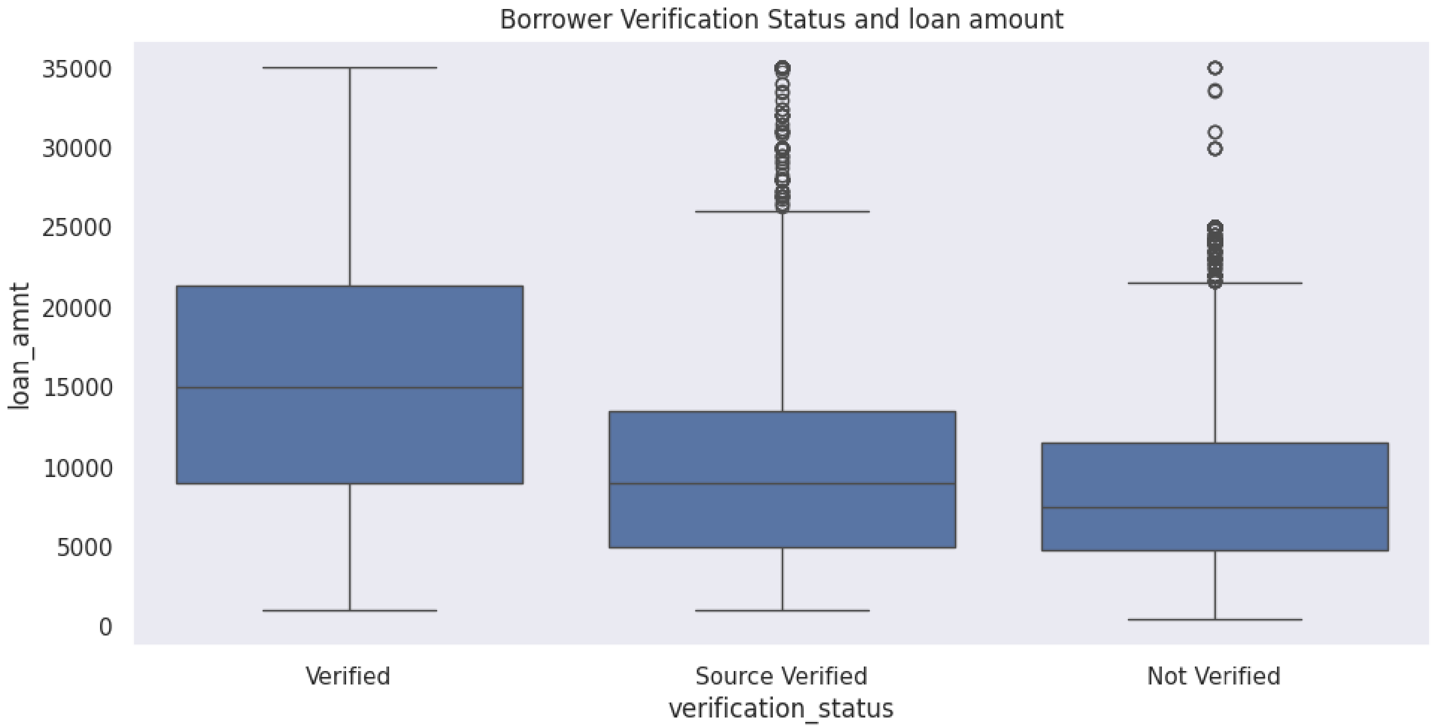
- The loan portfolio predominantly consists of conservative, low-risk borrowers who prefer moderate loan amounts and shorter terms, resulting in a healthy 85% repayment rate. However, a significant portion of unverified borrowers (42%) presents a potential risk that requires attention.
- The increasing demand for loans, primarily for debt consolidation and credit card refinancing, suggests growing financial pressures. This trend may indicate incomes not keeping pace with inflation or changing spending habits, highlighting the need for targeted financial education initiatives.
- The low representation of high-risk borrowers (those with derogatory records or bankruptcies) in the portfolio suggests these individuals may be seeking alternative financing options. This, combined with geographic concentrations in lending, presents both challenges and opportunities for market expansion and risk management.

Segmented Univariate Analysis

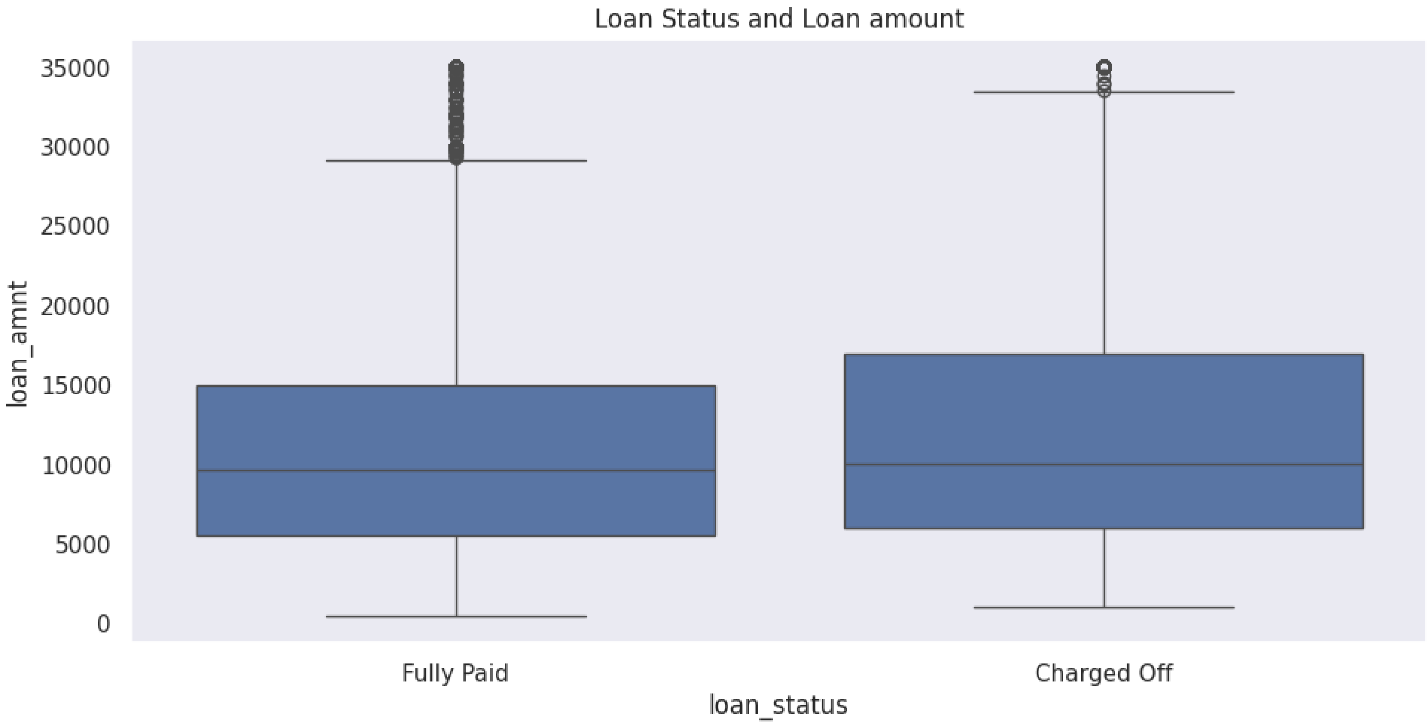
- The boxplot visually confirms a distinct trend where borrowers with lower credit grades (especially C and below) tend to take out larger loan amounts compared to those with higher grades (A and B). This pattern suggests that lenders may be willing to extend larger loans to individuals with lower credit scores, potentially at higher interest rates to compensate for the increased risk.



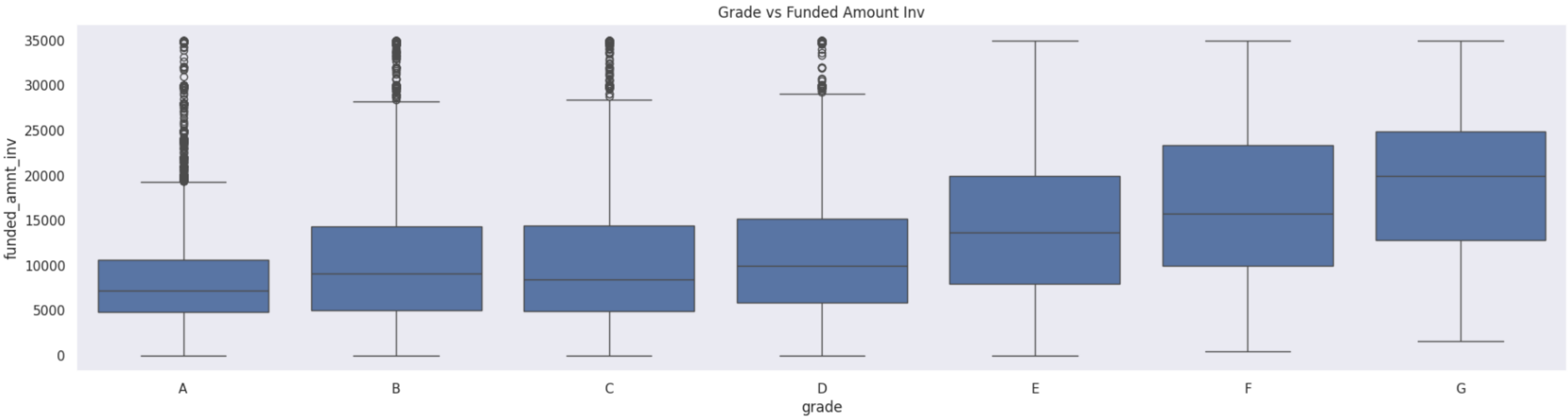
- The income verification process for a significant portion (42%) of borrowers remains incomplete, raising concerns about the accuracy of their financial information and potential risks for the lender.



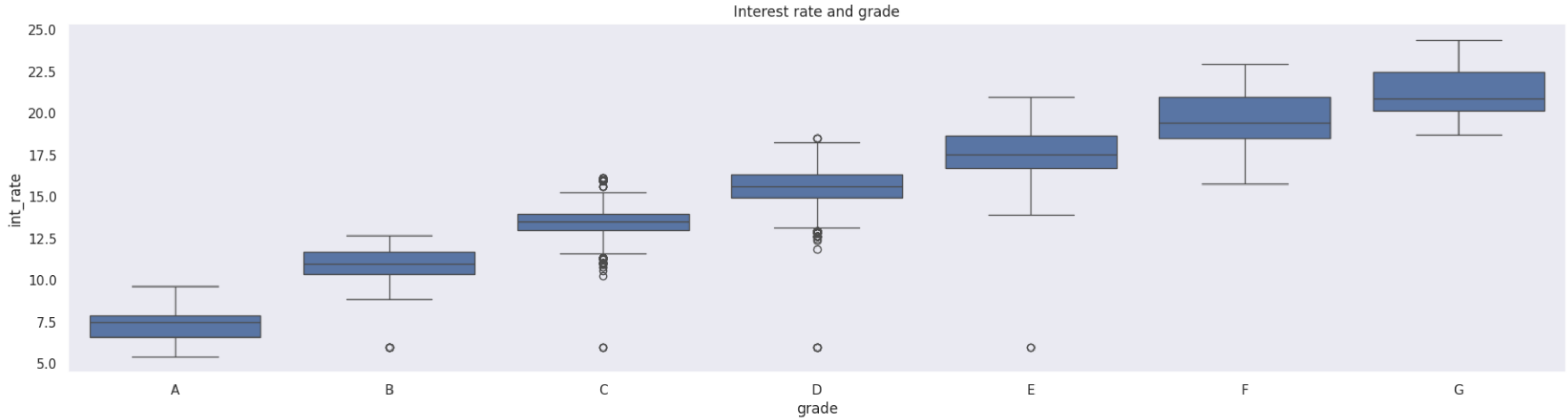
- The analysis reveals a notable positive correlation between loan amounts and the risk of default, suggesting that larger loan sizes may pose a greater challenge for borrowers in terms of repayment.



- This illustrates a clear trend where borrowers with lower credit grades (D to G) tend to receive larger loan amounts compared to those with higher grades (A and B) from investors. This suggests a higher risk tolerance from lenders for lower-grade borrowers, potentially offset by higher interest rates charged to these individuals, leading to increased potential profits. However, this strategy also carries a higher risk of loan defaults.



- A clear positive correlation between loan grade and interest rate: as the loan grade decreases, the interest rate charged to the borrower increases. This strategy is likely employed to mitigate the increased risk associated with lower credit grades, potentially leading to higher profits for the lender. However, this approach also carries a heightened risk of default, as observed in the higher default rates among borrowers with higher interest rates.

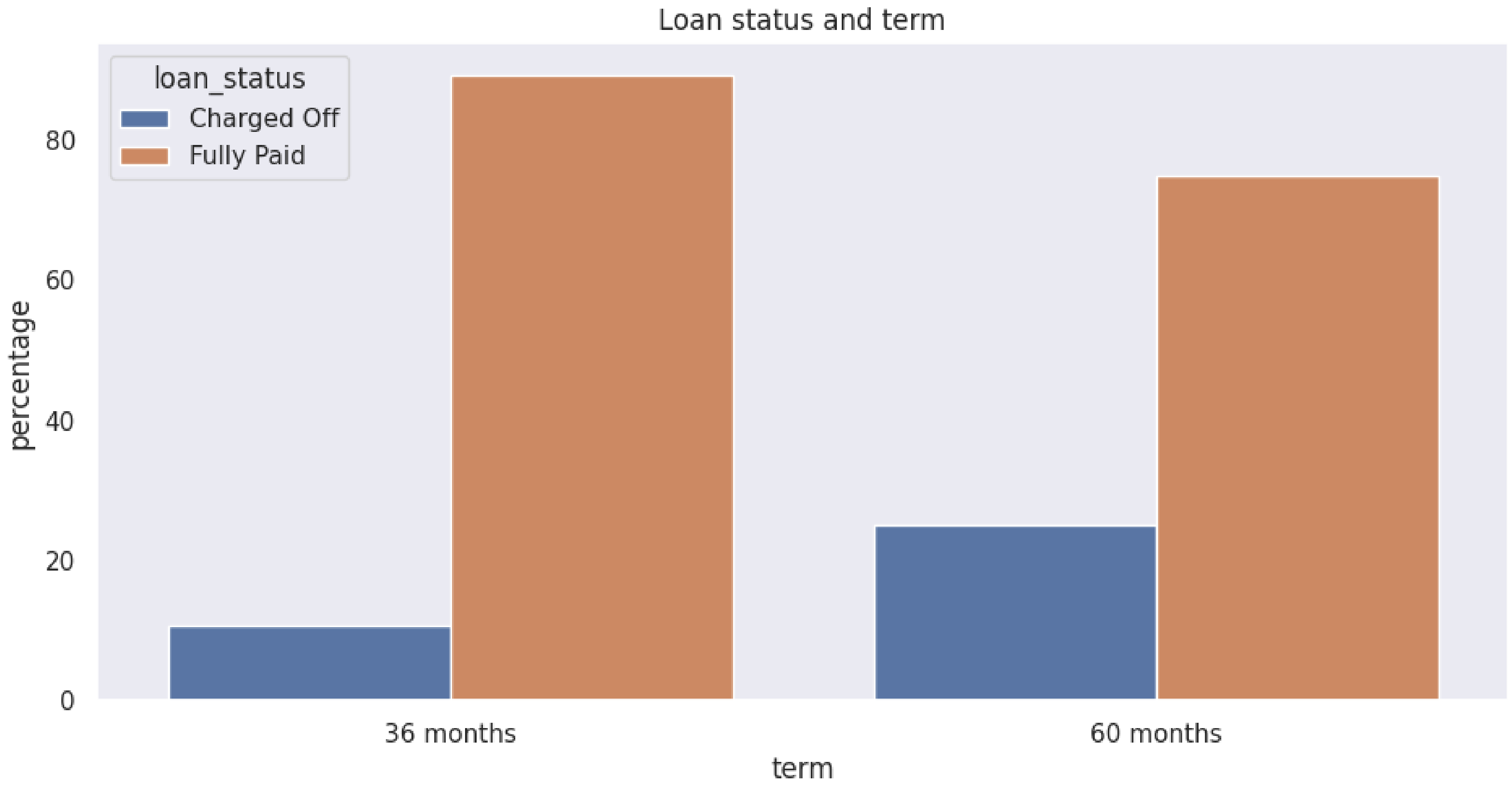


Conclusion from Segmented Univariate Analysis

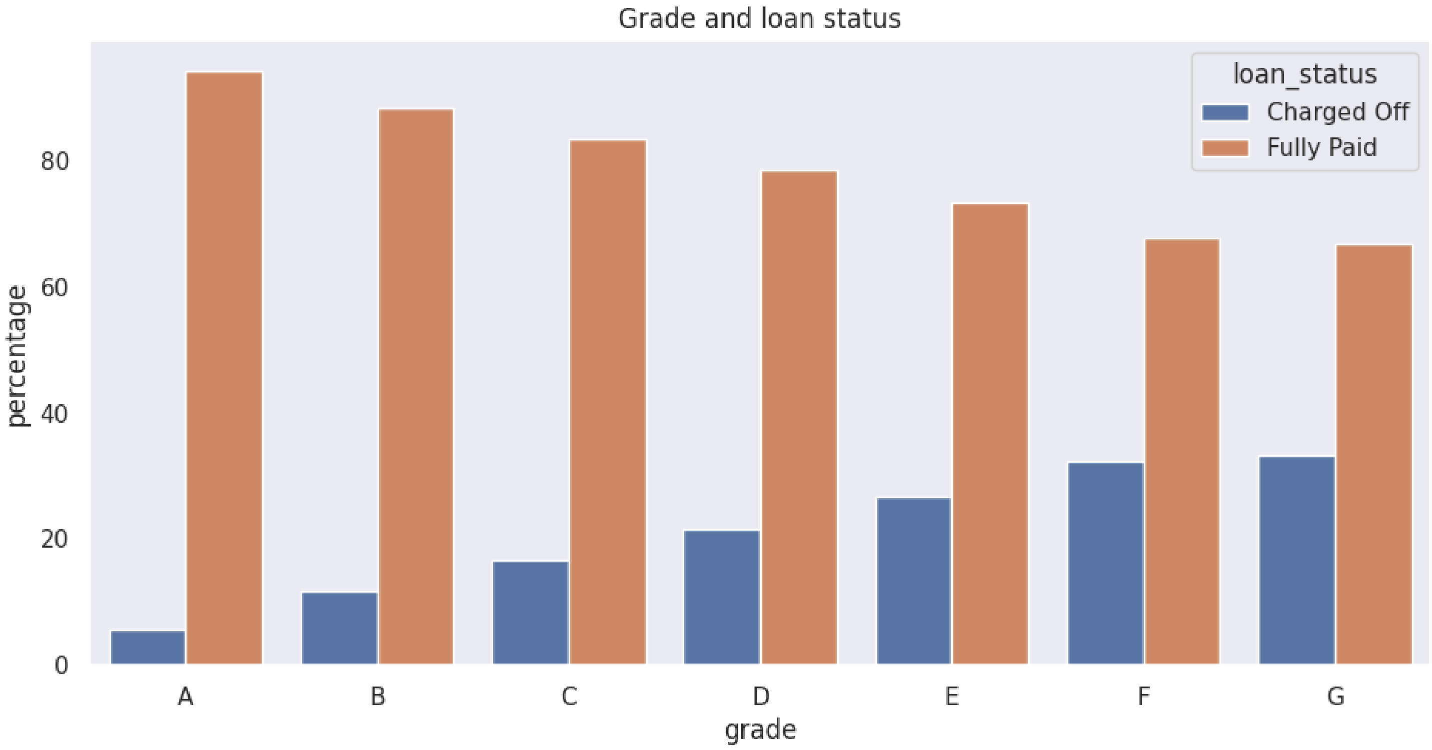
- Loan terms and credit grades significantly influence borrowing patterns. Longer terms (60 months) are associated with larger loan amounts, while lower credit grades tend to borrow more at higher interest rates. This risk-based pricing strategy may lead to increased default exposure for the lender.
- Income verification processes show concerning gaps, with 42% of borrowers lacking full verification. Counterintuitively, unverified incomes receive lower interest rates, raising questions about the effectiveness of current risk assessment practices and potentially increasing the lender's vulnerability.
- Multiple factors contribute to loan default risk, including larger loan amounts, higher interest rates, and elevated debt-to-income ratios. Credit grades show a gradual increase in DTI ratios as they decrease, and the presence of public derogatory records correlates with higher default rates, emphasizing the need for comprehensive borrower risk evaluation.

Bivariate Analysis

- Borrowers with 60-month loans default more frequently than those with 36-month terms, suggesting that longer loan durations correlate with higher default rates. This trend highlights the increased risk associated with extended repayment periods.

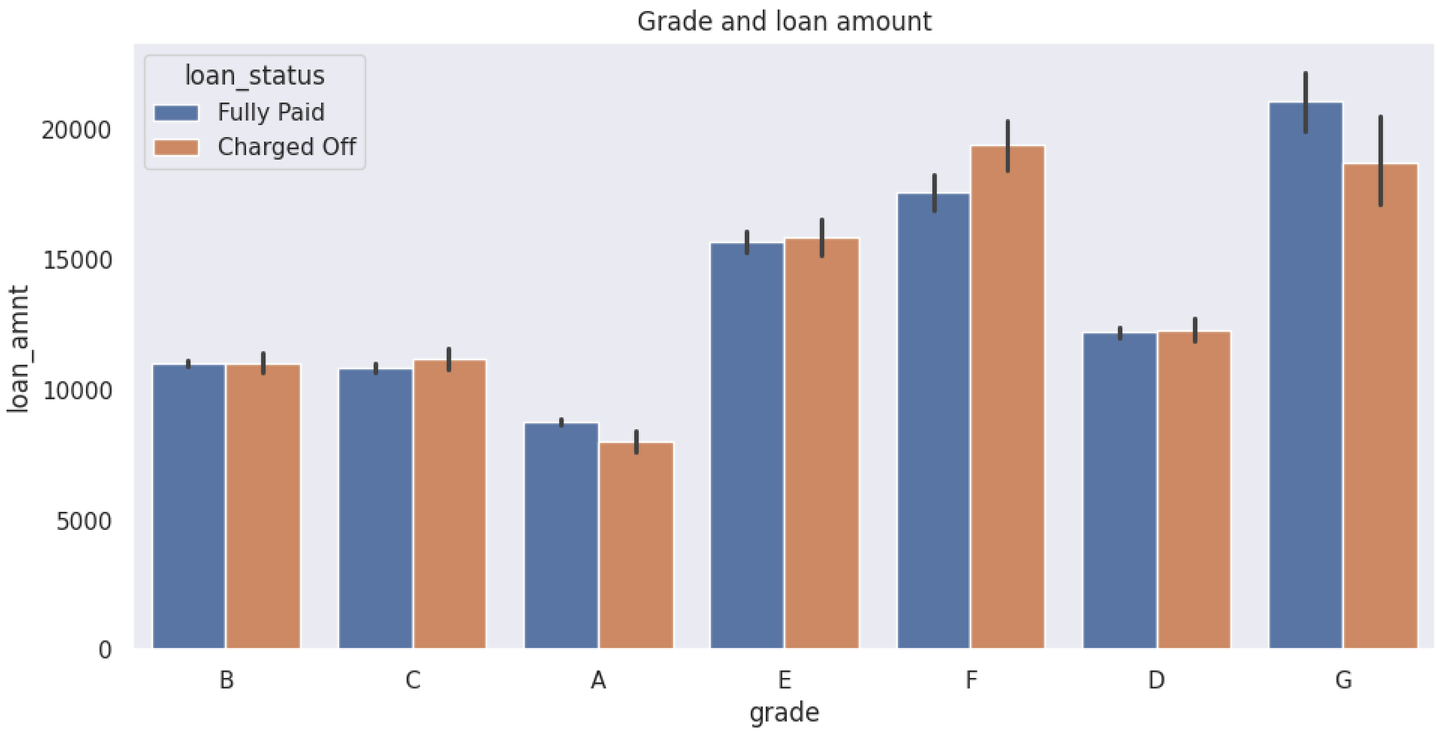


- Analysis reveals a distinct pattern in the relationship between loan grades and default rates. As loan grades decrease from higher to lower quality, there is a corresponding increase in the number of defaulted loans. This trend aligns with our earlier observations from the univariate analysis, further reinforcing the significant impact of loan grades on default likelihood.

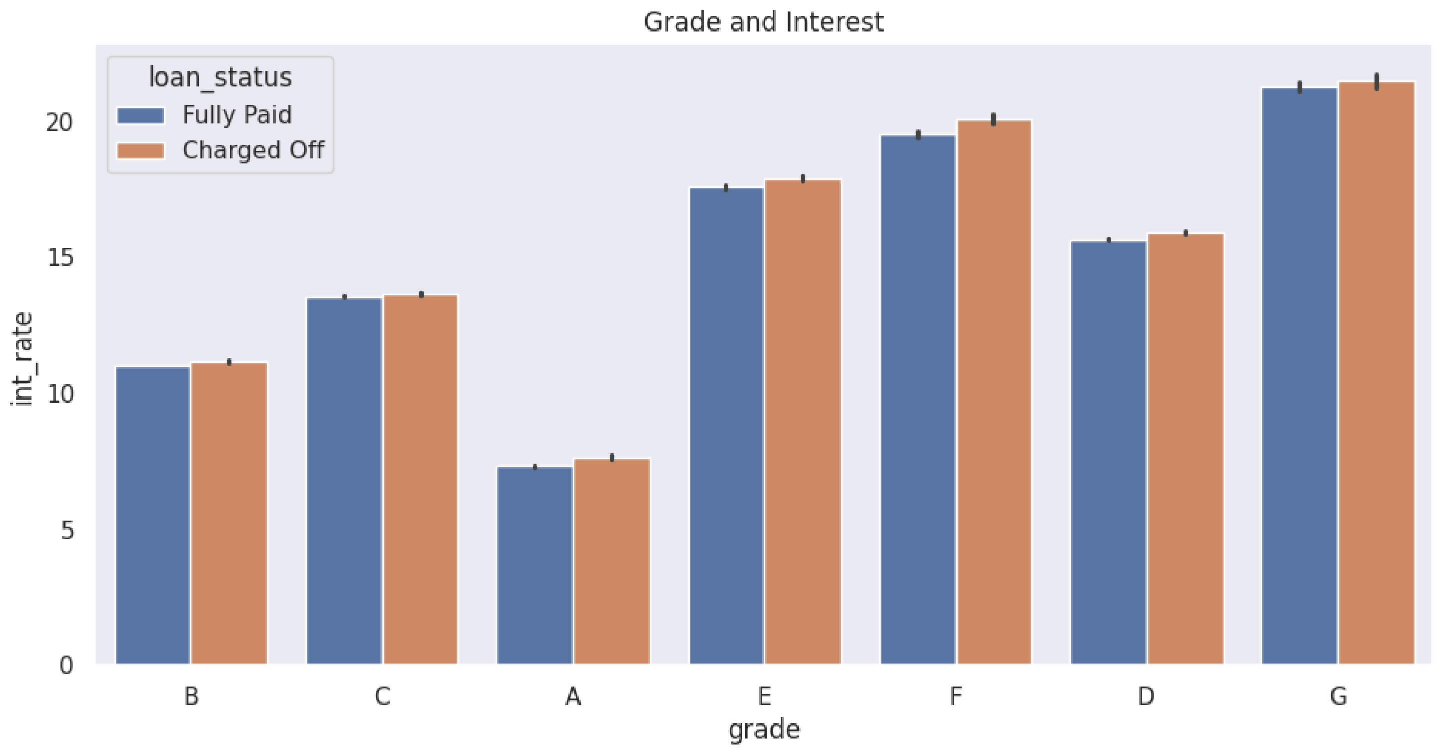


- Lower-grade borrowers show higher default rates and, surprisingly, receive larger loan amounts compared to higher-grade borrowers.

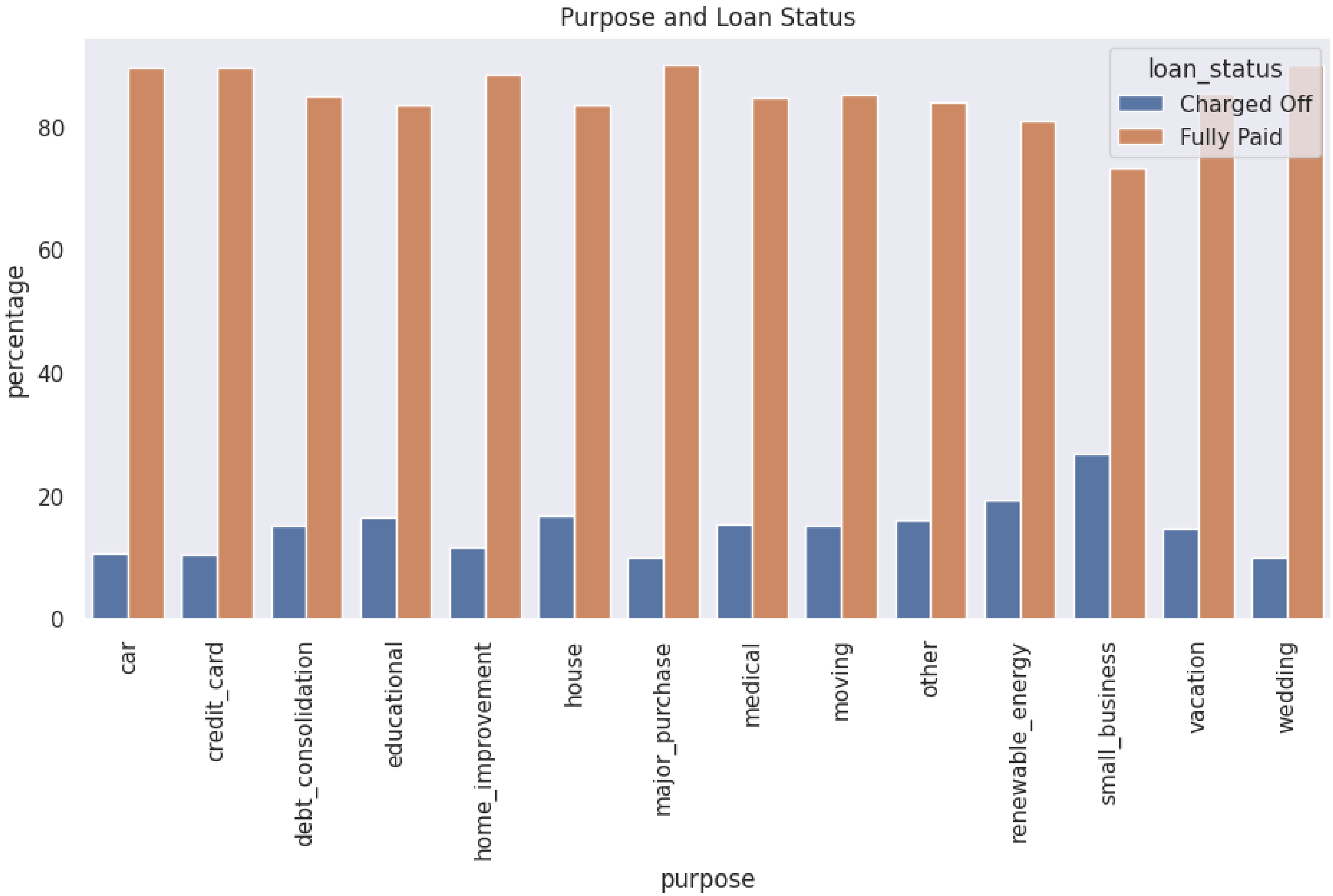
- This suggests that lenders club may be offering larger loans at higher interest rates to lower-grade borrowers, potentially contributing to increased default rates.

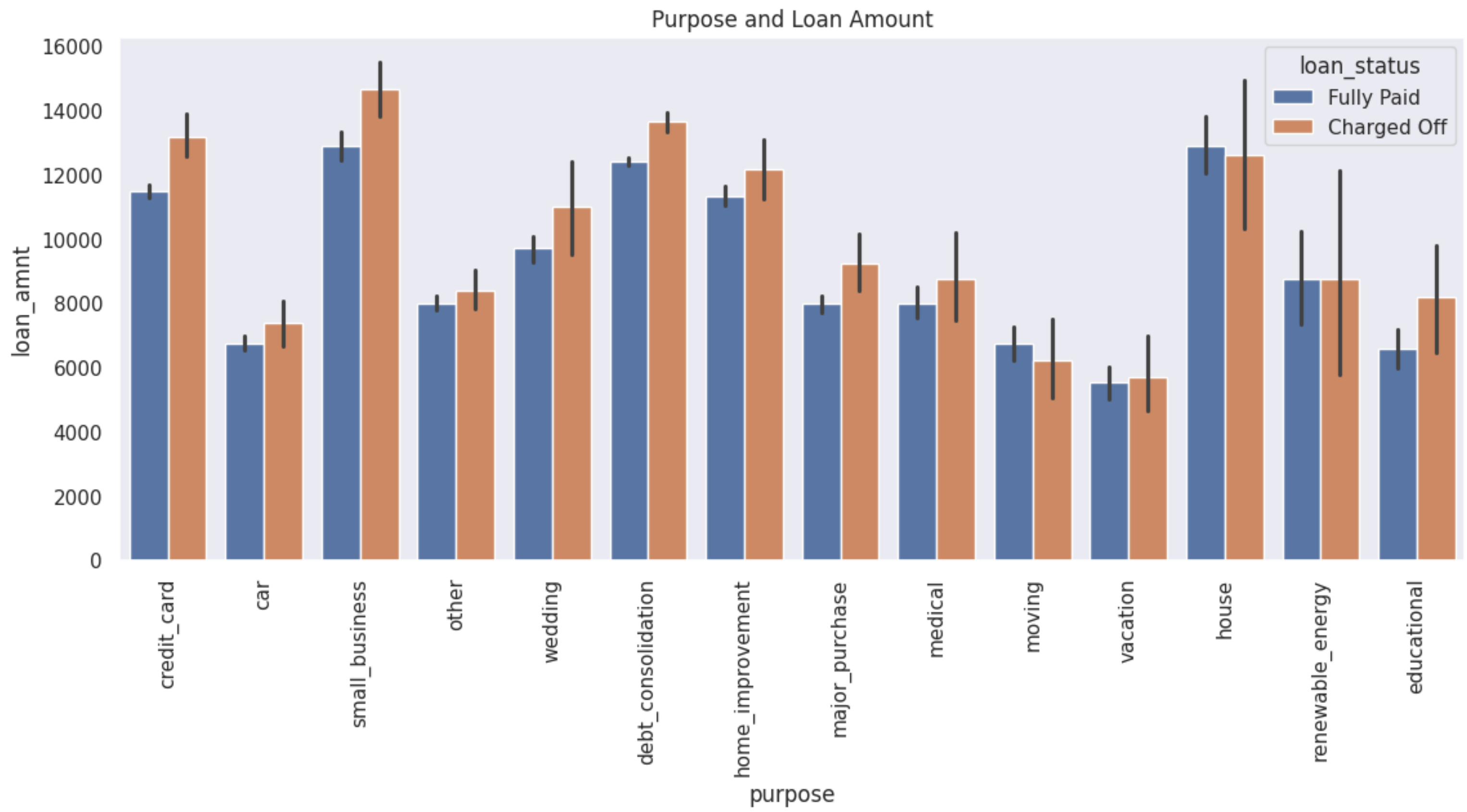


- This analysis confirms our earlier hypothesis: the combination of lower credit grades, larger loan amounts, and higher interest rates significantly increases the likelihood of loan default.



- Small business loans show a higher tendency for default compared to loans taken for other purposes.





- Borrowers with small businesses who defaulted on their loans have, on average, taken larger loan amounts compared to other borrower categories. This trend suggests a potentially higher risk profile for small business loans, highlighting the need for careful consideration in lending practices for this segment.

Conclusion from Bivariate Analysis

- Loan terms, amounts, and grades significantly influence default rates. Longer terms (60 months) and lower credit grades correlate with higher default rates. Surprisingly, lower-grade borrowers receive larger loan amounts at higher interest rates, potentially exacerbating default risk.
- Interest rates and debt-to-income (DTI) ratios show clear relationships with loan performance. Higher interest rates and DTI ratios are associated with increased defaults across all loan terms. However, the differences are often subtle, suggesting that even minor adjustments to lending criteria could significantly impact repayment rates.
- Loan purpose and verification status offer additional insights into default risk. Small business loans show a higher tendency for default and are often associated with larger loan amounts. Interestingly, a borrower's verification status does not appear to be a strong predictor of default rates, challenging assumptions about the importance of income verification in risk assessment.

Conclusion For Lending Club Case Study

- Loan characteristics strongly influence default risk. Longer terms, lower credit grades, and higher interest rates correlate with increased defaults. Lower-grade borrowers often receive larger loans at higher rates, exacerbating risk.
- Borrower financial health is crucial in predicting defaults. Lower credit grades, higher debt-to-income ratios, and public bankruptcy records contribute to increased default risk. However, income verification status has less impact than expected.
- Loan purpose and amount affect default probability. Small business loans show higher default rates and larger amounts. Minor adjustments to lending criteria could significantly impact repayment rates across borrower segments.