Lending Club Case Study

Harshavardhan Katarki

Deepak Katarki

Initial Steps

- Libraries are loaded

- Data is cleaned
  - Both for numerical values and categorical values

Univariate Analysis: Loan Status Distribution

- Dominant Status - Fully Paid: There is a large majority (approximately 83%) of loans are in the 'Fully Paid' status, indicating successful repayment. So it is good to have.

- Significant Charged Off Rate: About 14% of loans are 'Charged Off', signifying a notable portion of loans where borrowers have defaulted. Need to have mesures to mitigate these.

- Minor Proportion of Current Loans: Only around 2.9% of loans are currently ongoing ('Current'). This is a small fraction compared to fully paid and charged-off loans. So this is good.

- Indicator of Good Credit Decisions: The high percentage of 'Fully Paid' loans suggests effective credit decision-making and loan management. Good to have.

- Need for Risk Mitigation: The presence of a substantial 'Charged Off' rate points to the need for improved risk assessment and mitigation strategies.

- Portfolio Health: The good proportion of loans being fully paid is a positive sign of the portfolio's overall health.
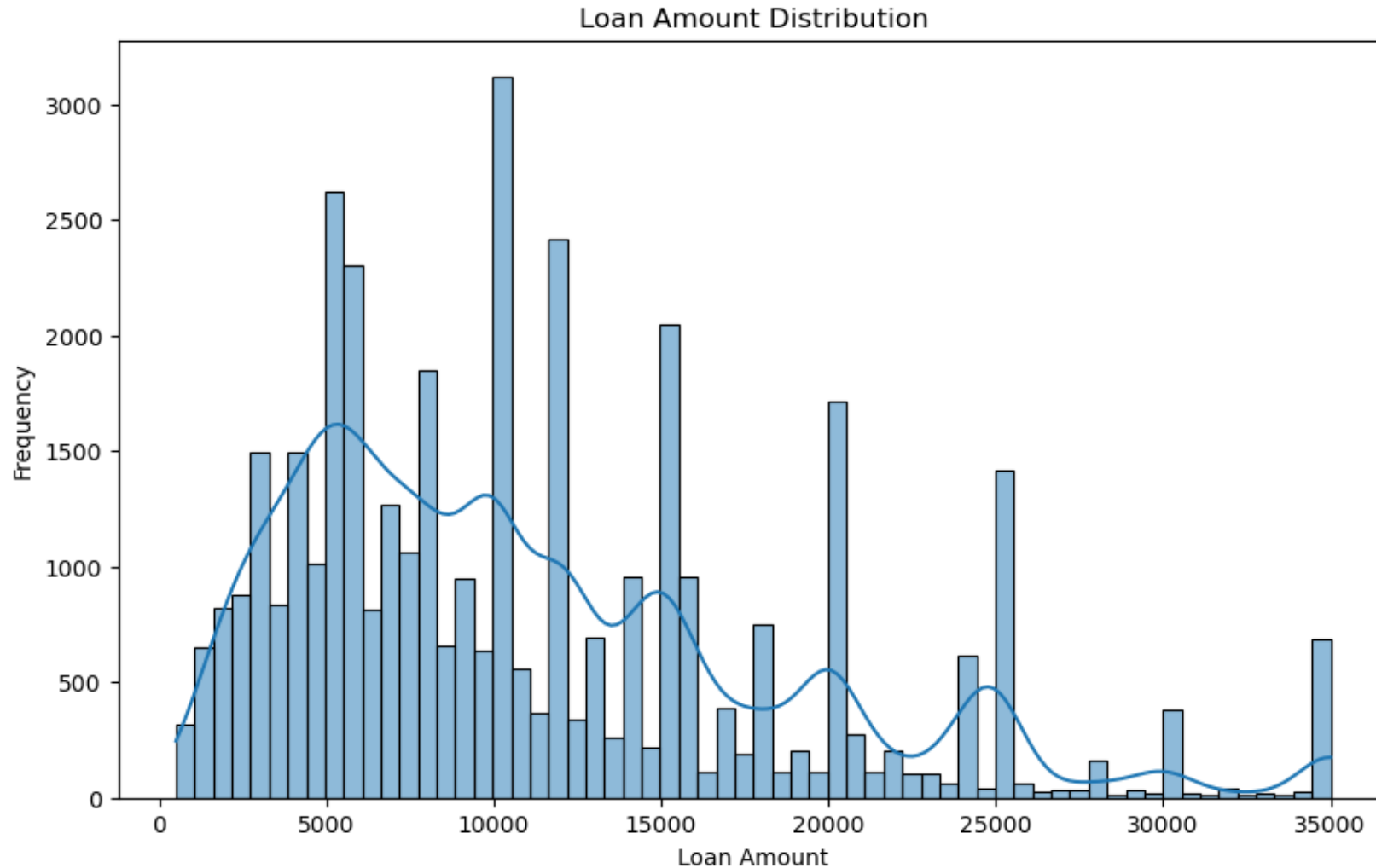
Univariate Analysis: Loan Status Distribution

Loan Status Distribution
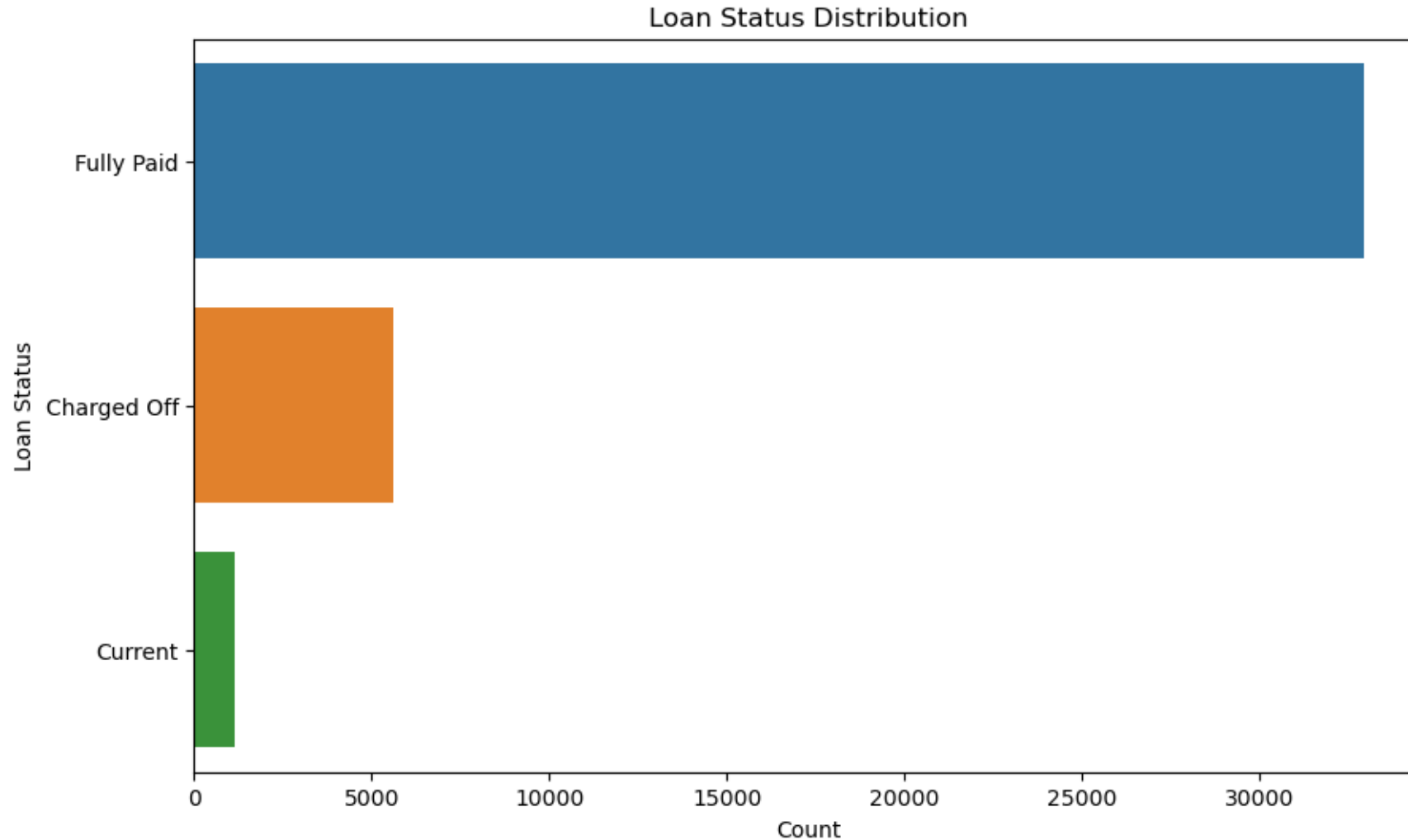
Univariate Analysis: Loan Amount Distribution

- Average Loan Amount: The average (i.e., mean) loan amount is approximately $11,219.

- Range of Loan Amounts: Loan amounts range from a minimum of $500 to a maximum of $35,000.

- Standard Deviation: The standard deviation is about $7,457. This is indicating variability in the loan amounts.

- Most Common Loan Amounts: The median loan amount is $10,000, suggesting that half of the loans are below this amount and half are above.

- Distribution Spread: The 25th percentile is at $5,500 and the 75th percentile is at $15,000, indicating a spread of loan amounts in the middle 50 percent of the data.

- Loan Amount Preferences: The data suggests a preference or tendency towards mid-range loan amounts among borrowers.
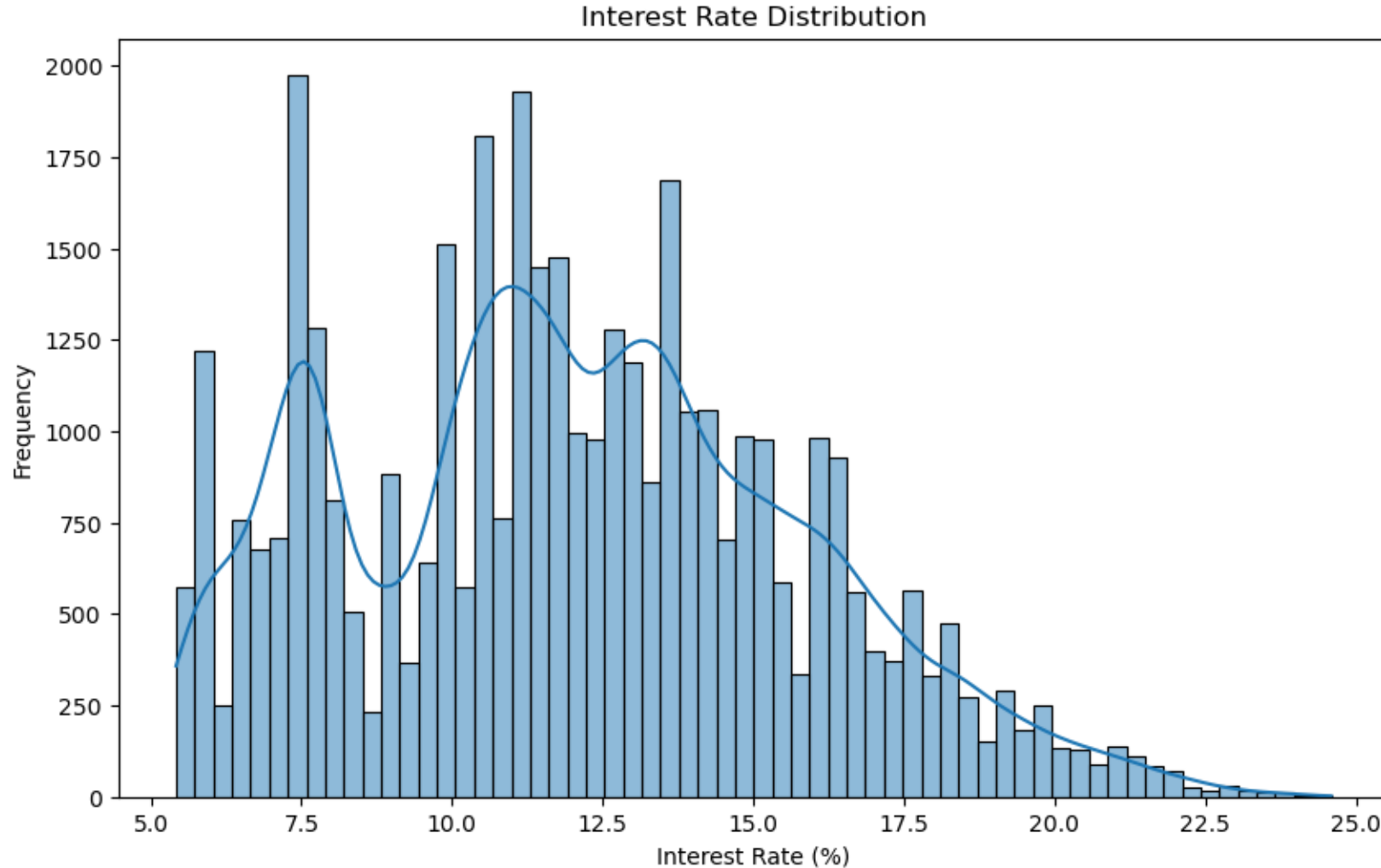
Univariate Analysis: Loan Amount Distribution

Loan Amount Distribution

Univariate Analysis: Interest Rate Distribution

- Average Interest Rate: The mean interest rate across all loans is approximately 12.02%. This is indicating the average cost of borrowing.

- Higher Rates for Riskier Loans: The presence of higher rates (up to 24.59%) may indicate loans issued to individuals with higher risk factors such as lower credit scores or unstable income.

- Variability in Rates: There is a notable variation in interest rates (standard deviation of about 3.72%). This is suggesting diverse risk profiles among borrowers.

- Most Common Rates: The median interest rate is 11.86%, with half of the loans having rates below this value and half above. This is indicating the central tendency of the rates.

- Range of Interest Rates: Interest rates range from a low of 5.42% to a high of 24.59%. This is reflecting a wide spectrum of creditworthiness and loan terms.

- Competitive Rates for Lower Risks: The lower end of the spectrum (5.42%) suggests competitive rates offered to borrowers with better credit profiles.
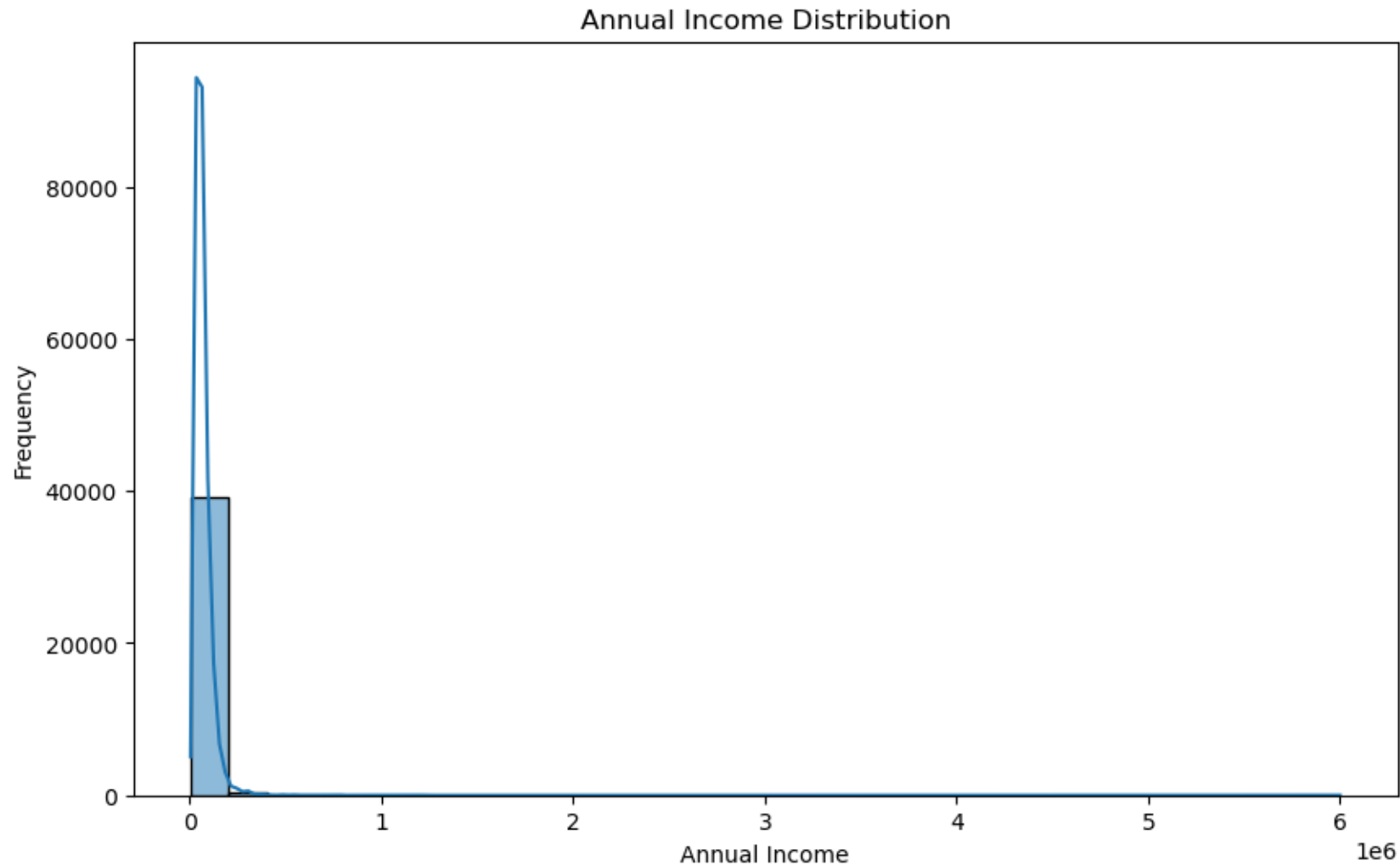
Univariate Analysis: Interest Rate Distribution

Interest Rate Distribution

Univariate Analysis: Annual Income Distribution

- Average Annual Income: The average annual income of borrowers is approximately $68,969. This is indicating the general income level of the loan applicants.

- Median Income: The median annual income is $59,000, which is lower than the average. This is hinting at a right-skewed distribution where a smaller number of high-income individuals push the average up.

- Variability in Income Levels: The standard deviation is quite high (around $63,794). This is indicating substantial variability in the income levels of the borrowers.

- Wide Range of Incomes: There's a significant range in annual incomes, from as low as $4,000 to as high as $6,000,000, suggesting a diverse borrower base.

- Income Distribution: 25% of borrowers earn $40,404 or less, and 75% earn $82,300 or less. This is implying that the majority of borrowers are in the middle to lower income brackets.

- Implications for Lending: The wide range in incomes, along with the higher average compared to the median, suggests that the loan data includes both high-earning borrowers and those with relatively modest incomes, which could influence loan terms and risk assessments.
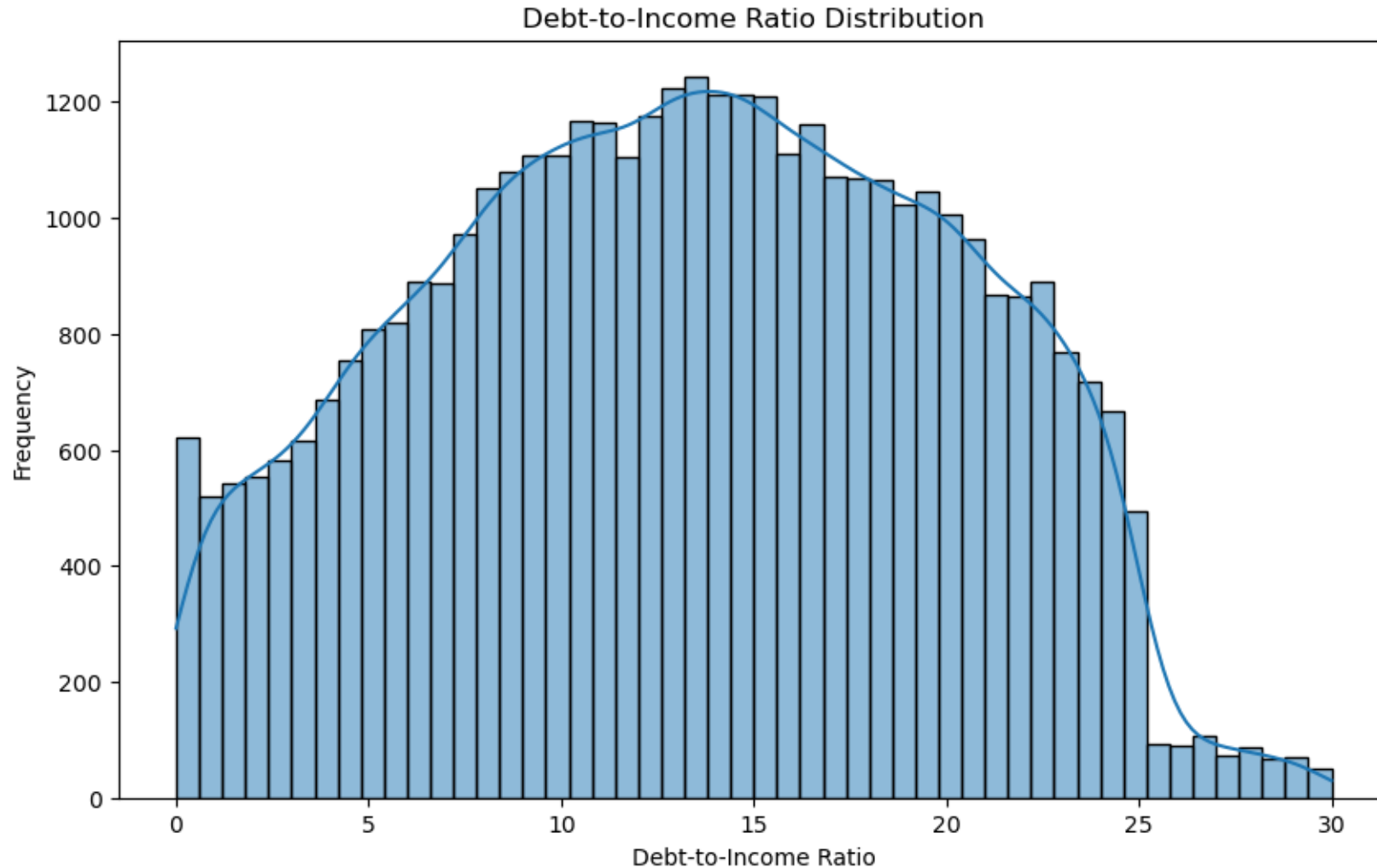
Univariate Analysis: Annual Income Distribution

Univariate Analysis: Debt-to-Income Ratio Distribution

- Average DTI Ratio: The average DTI ratio is approximately 13.32%. This is indicating the average proportion of borrowers' income going towards debt payments.

- Variability in DTI Ratios: There is notable variability in DTI ratios among borrowers (standard deviation of about 6.68%), reflecting different levels of debt burden relative to income.

- Median DTI Ratio: The median DTI ratio is 13.4%, suggesting that half of the borrowers have a lower debt burden relative to their income and half have a higher burden.

- Income Utilization for Debts: A higher DTI ratio means a larger portion of income is used for debt payments. This can be an important factor in assessing loan affordability and risk.

- Range of DTI Ratios: DTI ratios range from 0% to nearly 30%. This is showing a wide spectrum of debt burden among borrowers.

- Creditworthiness Indication: Lower DTI ratios might indicate better creditworthiness and financial stability, while higher ratios could signal potential risk in lending.

Univariate Analysis: Debt-to-Income Ratio Distribution

Debt-to-Income Ratio Distribution

Univariate Analysis: Employment Length Distribution

- Dominant Employment Length: The most common employment length is '10+ years'. This is accounting for approximately 23% of borrowers. This suggests a significant portion of borrowers have a long employment history.

- Stable Employment Indicator: A large number of borrowers with over 10 years of employment may indicate financial stability, which is a positive sign for lenders.

- Middle-Range Employment: Employment lengths of 4 to 9 years represent a smaller but notable portion of the borrower base. This is reflecting a mix of mid-career individuals.

- Distribution Across Categories: Employment lengths are well-distributed across different categories, with '2 years', '3 years', and '< 1 year' also being common durations, each constituting around 11% to 12% of borrowers.

- Variety in Employment Lengths: The presence of borrowers across all employment lengths suggests a diverse borrower base in terms of job stability and career stage.

- Short-Term Employment Representation: The data also includes a significant proportion of borrowers with less than one year of employment, highlighting the inclusion of relatively new entrants into the workforce or those who may have recently changed jobs.

Univariate Analysis: Employment Length Distribution
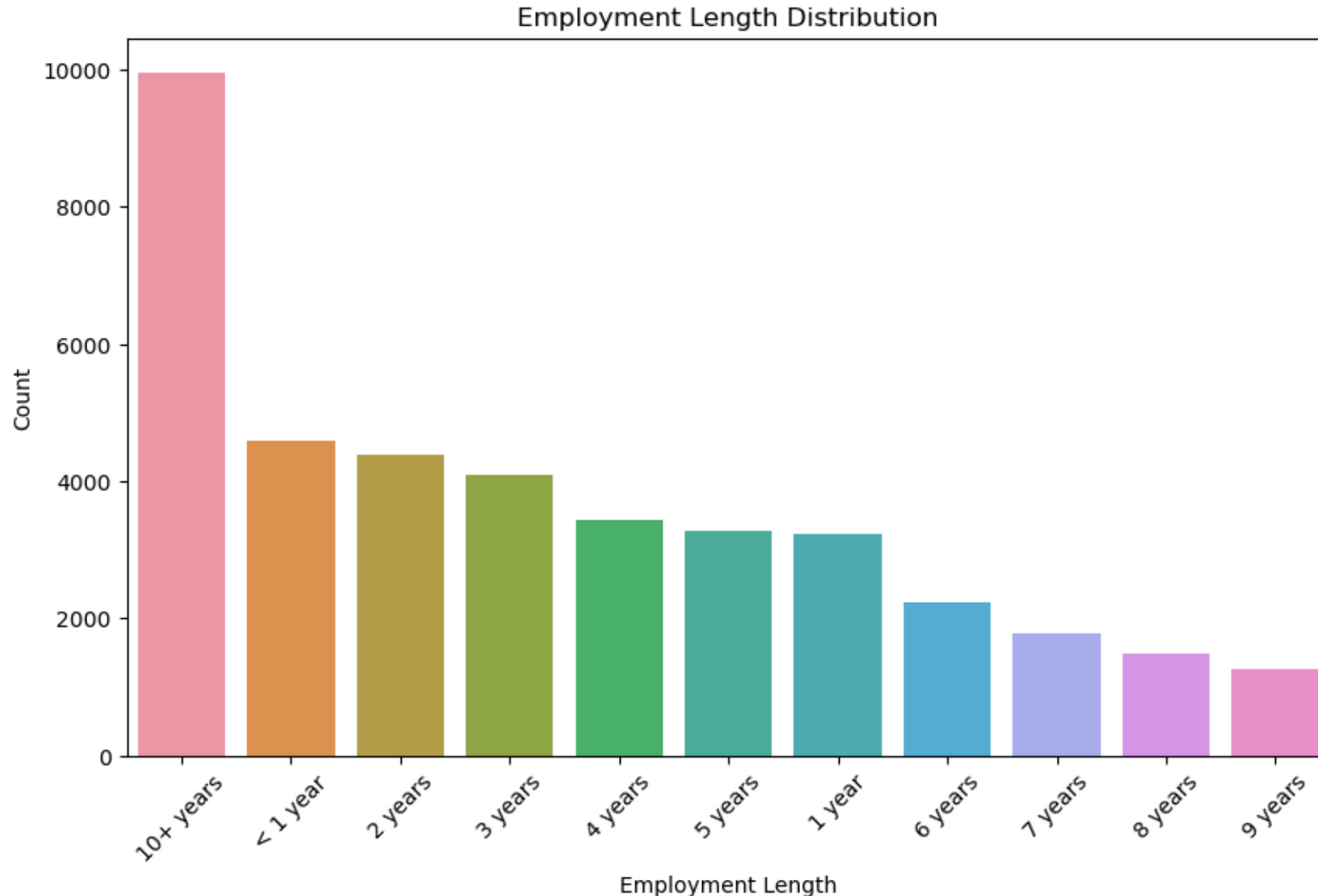
Bivariate Analysis: Loan Status vs Loan Amount

- Variation Across Loan Statuses: The distribution of loan amounts varies significantly across different loan statuses.

- Similar Ranges for Fully Paid and Charged Off: Both 'Fully Paid' and 'Charged Off' categories have a wide range of loan amounts, but the upper quartile of 'Charged Off' loans is notably higher.

- Implications for Risk Management: The analysis suggests the need for closer scrutiny of higher loan amounts, as they may carry a higher risk of not being fully repaid.

- Higher Amounts in Charged Off Loans: 'Charged Off' loans tend to have higher median loan amounts compared to 'Fully Paid' loans, suggesting that higher loan amounts might be associated with a greater risk of default.

- Current Loans Distribution: Loans that are currently 'Current' have a distribution similar to 'Fully Paid' loans. This is indicating similar loan amount preferences among borrowers who are actively repaying.

- Outliers and Variability: The presence of outliers, especially in the 'Charged Off' category, indicates that a few loans have significantly higher amounts. This could be risk factors for default.
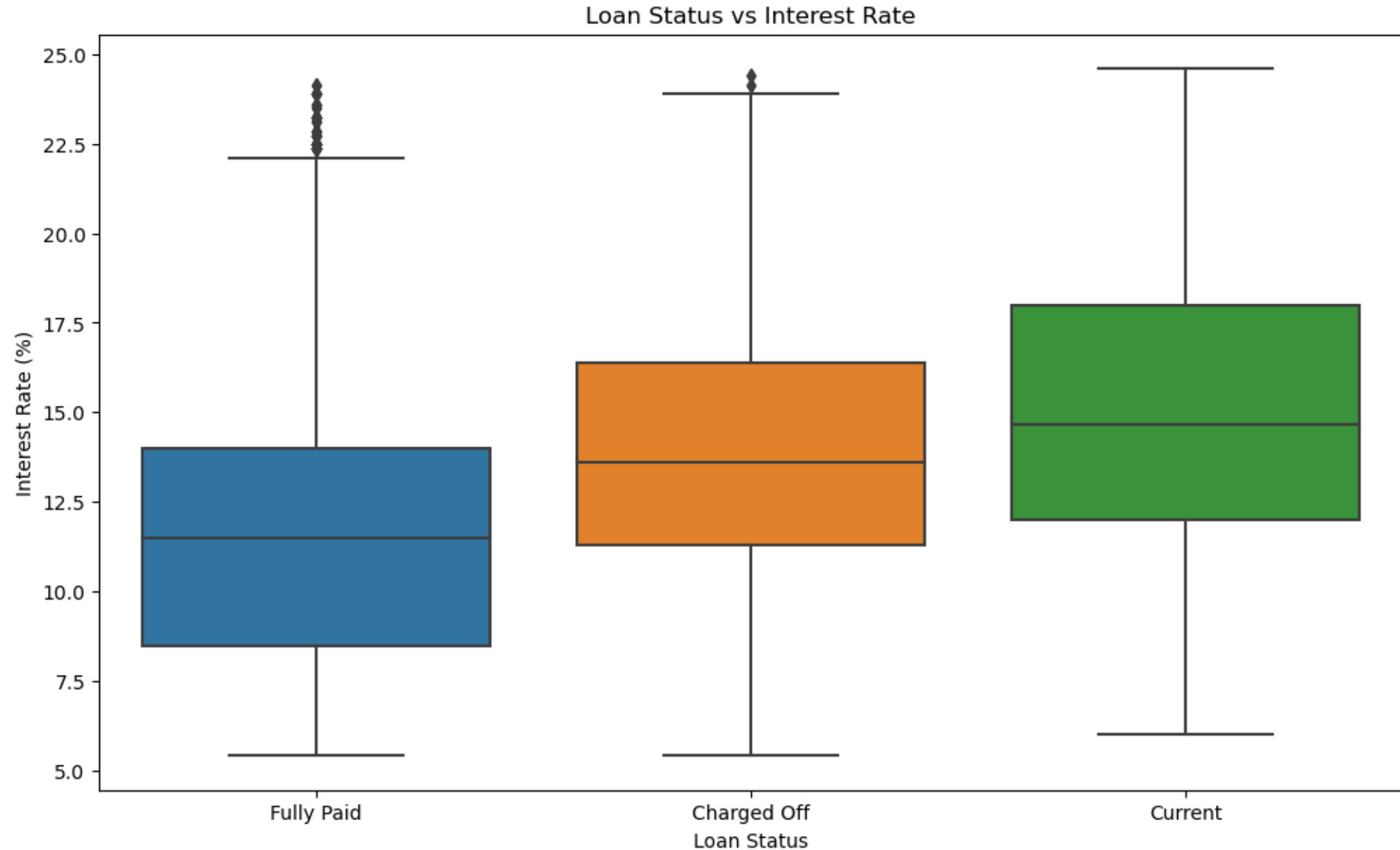
Bivariate Analysis: Loan Status vs Loan Amount

Loan Status vs Loan Amount
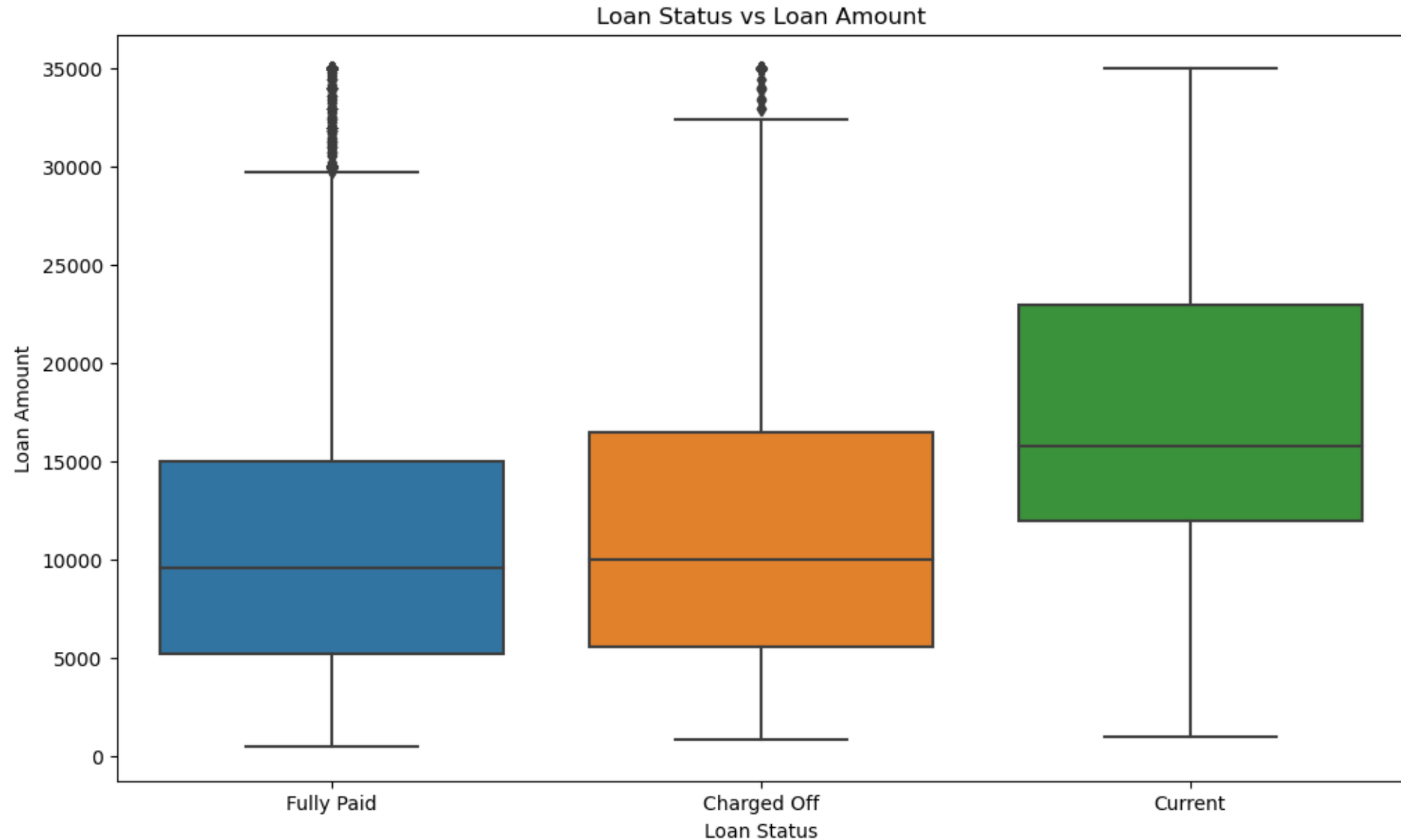
Bivariate Analysis: Loan Status vs Interest Rate

- Higher Rates for Charged Off Loans: There's a noticeable trend of higher interest rates associated with 'Charged Off' loans, which suggests that loans with higher interest rates are more likely to default.

- Importance of Interest Rate in Loan Performance: The clear distinction in interest rates across different loan statuses underscores the importance of interest rate as a factor in loan performance and borrower's ability to repay.

- Variability in Rates: The range of interest rates is quite broad for all loan statuses, but particularly so for 'Charged Off' loans, suggesting that higher risk loans (as indicated by higher rates) don't always result in default, but the risk is increased.

- Lower Rates for Fully Paid Loans: Loans that are 'Fully Paid' generally have lower interest rates, indicating that borrowers with lower rates are more likely to successfully repay their loans.

- Current Loans Interest Rates: Loans currently in the 'Current' status have interest rates that are somewhat in between those of 'Fully Paid' and 'Charged Off' loans. This might reflect ongoing loans with varying risk profiles.

- Risk Assessment Implication: The higher interest rates for 'Charged Off' loans could reflect the lender's initial risk assessment, where higher rates were likely offered to riskier borrowers.

Bivariate Analysis: Loan Status vs Interest Rate

Loan Status vs Interest Rate

Bivariate Analysis: Loan Amount vs Annual Income

- Outliers and Exceptions: Some outliers are present where individuals with relatively lower annual incomes have high loan amounts. This could be cases requiring special attention for risk assessment.

- Implications for Lending Policies: Understanding this relationship can help in tailoring lending policies and loan offers, ensuring that borrowers are not over-leveraged relative to their income.

- Positive Correlation: There appears to be a positive correlation between loan amount and annual income. This is indicating that borrowers with higher incomes tend to take larger loans.

- Variability at Lower Income Levels: There is significant variability in loan amounts for borrowers with lower annual incomes, suggesting that factors other than income might play a significant role in determining the loan amount for these individuals.

- Cap on Loan Amounts: Despite increases in income, the loan amounts seem to plateau. This suggests that there might be a cap on the loan amounts that can be borrowed, regardless of higher income levels.

- Diverse Borrower Base: The scatterplot shows a wide range of incomes and loan amounts, indicating a diverse borrower base with varying financial needs and capabilities.
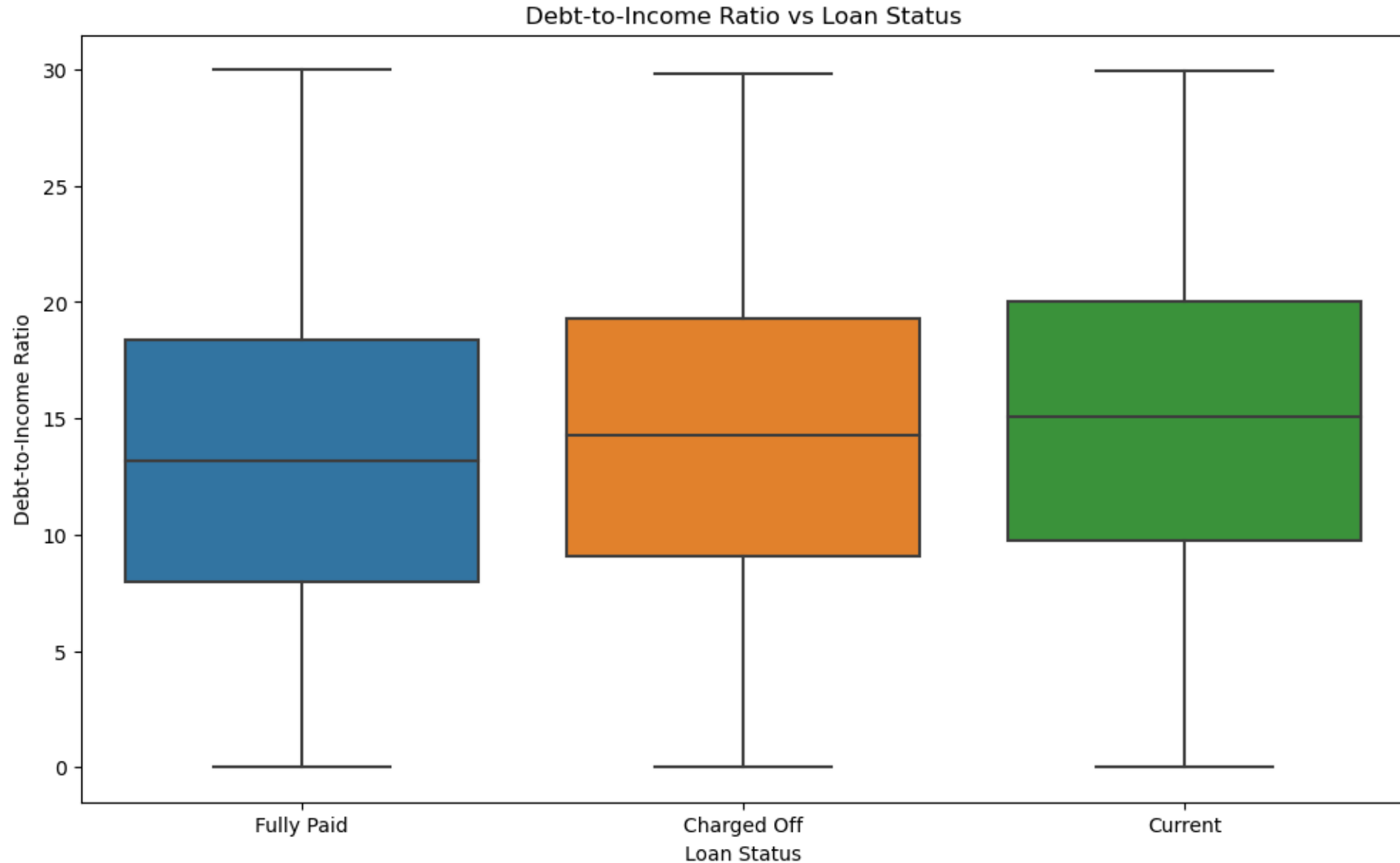
Bivariate Analysis: Loan Amount vs Annual Income

Loan Amount vs Annual Income
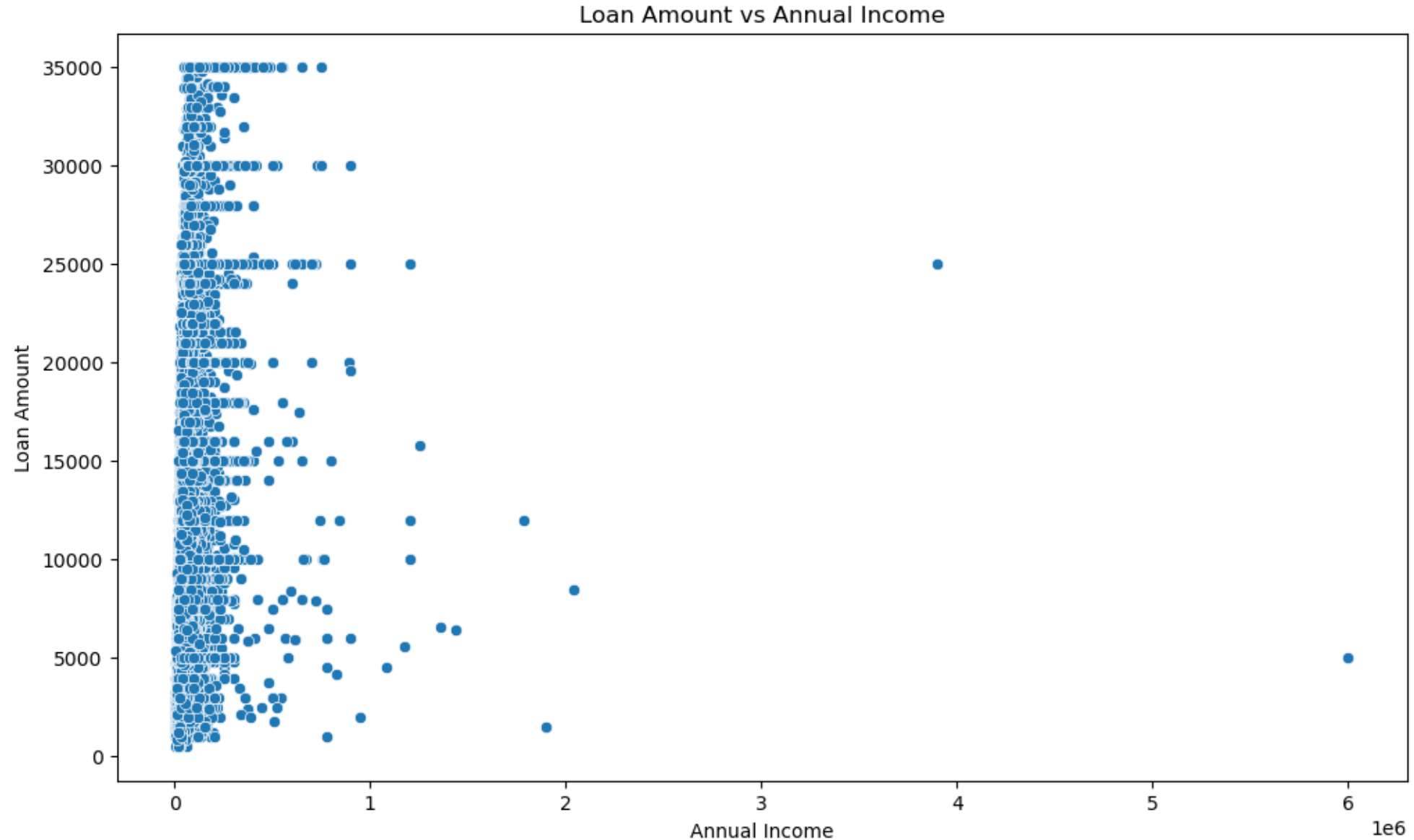
Bivariate Analysis: Debt-to-Income Ratio vs Loan Status

- DTI Range for Current Loans: Loans that are currently 'Current' show a DTI distribution similar to 'Fully Paid' loans, which might indicate that ongoing loans have manageable debt levels relative to income.

- Debt Burden and Loan Performance: While DTI is an important measure of debt burden, its similar distribution across loan statuses suggests that other factors, in addition to DTI, are influential in determining loan performance.

- Risk Assessment Implication: Lenders might consider DTI as one of several factors in assessing loan risk, given its modest differentiation power in loan performance.

- Similar DTI Across Loan Statuses: The median DTI appears to be relatively similar across different loan statuses. This is suggesting that DTI alone may not be a strong differentiator in predicting loan default.

- Slight Increase in DTI for Charged Off Loans: The 'Charged Off' loans exhibit a slightly higher median DTI compared to 'Fully Paid' loans, indicating a potential but modest association between higher DTI and loan default.

- Variability and Outliers: There is noticeable variability and presence of outliers in the DTI for all loan statuses, particularly for 'Charged Off' loans. This may imply that extreme DTI values could be linked to higher default risks.

Bivariate Analysis: Debt-to-Income Ratio vs Loan Status

Debt-to-Income Ratio vs Loan Status

Multivariate Analysis: Loan Status, Loan Amount, and Interest Rate

- Higher Loan Amounts in Charged Off Loans: 'Charged Off' loans tend to have higher loan amounts compared to 'Fully Paid' loans. This suggests a correlation between higher loan amounts and increased likelihood of default.

- Loan Amount Distribution: The distribution of loan amounts is fairly consistent across different loan statuses, but with a noticeable increase in the median and upper quartiles for 'Charged Off' loans.

- Risk Management Implications: The analysis underscores the importance of considering both loan amount and interest rate in assessing loan risk. It suggests a need for a balanced approach in loan underwriting, taking into account both the size of the loan and the cost to the borrower.

- Higher Rates Indicative of Higher Risk: The elevated interest rates for 'Charged Off' loans likely reflect the higher risk assessed at the time of loan origination.

- Current Loans Characteristics: Loans in the 'Current' status exhibit loan amounts and interest rates that fall between those of 'Fully Paid' and 'Charged Off' loans, suggesting ongoing loans have varied risk profiles.

- Interest Rate Variation Across Loan Statuses: The interest rates are generally higher for 'Charged Off' loans than for 'Fully Paid' loans, indicating that loans with higher interest rates are more prone to default.
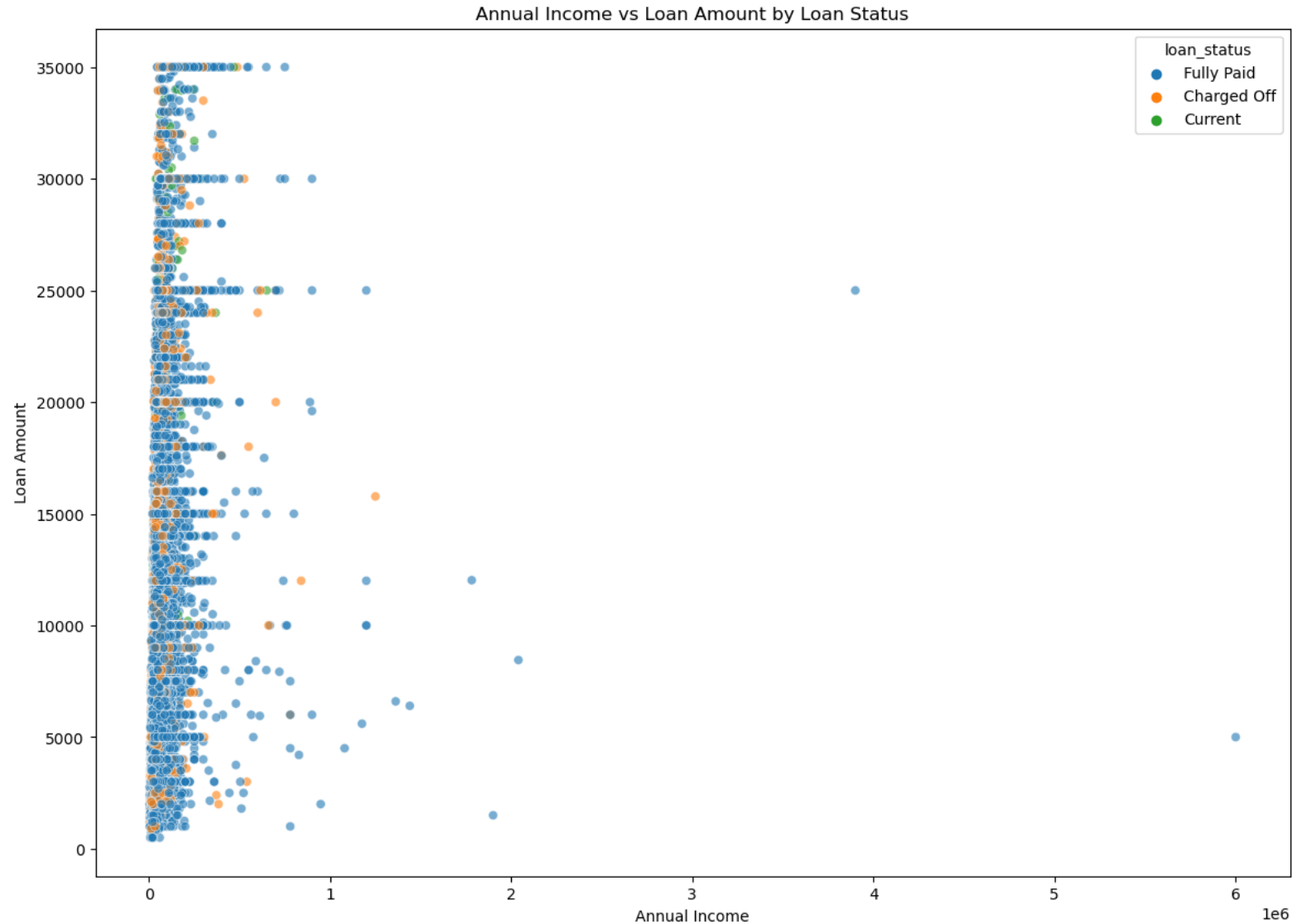
Multivariate Analysis: Loan Status, Loan Amount, and Interest Rate
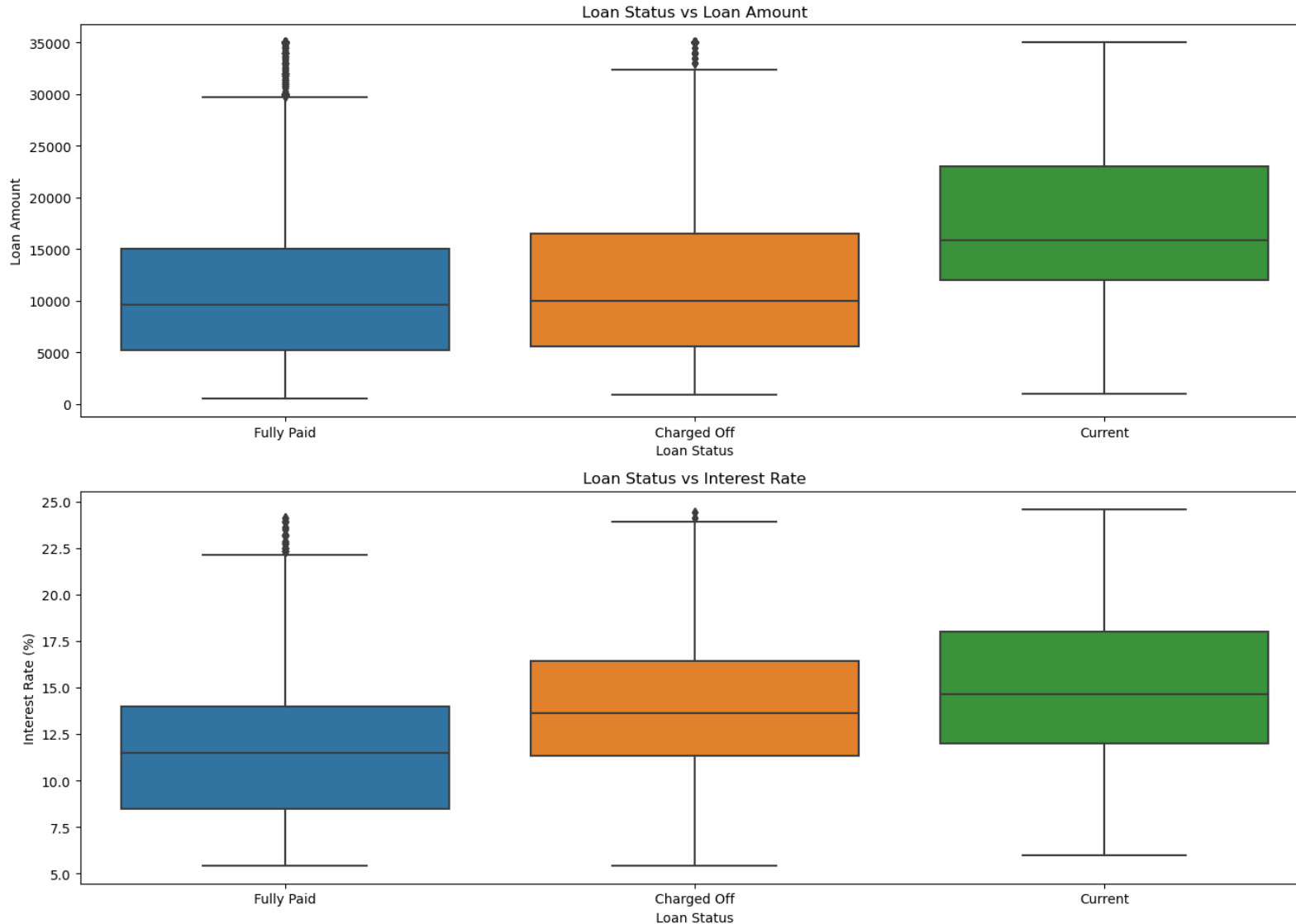
Multivariate Analysis: Annual Income, Loan Amount, and Loan Status

- Variability Among Loan Statuses: While 'Fully Paid' and 'Charged Off' loans are distributed across the full range of incomes, 'Charged Off' loans are more prevalent at lower income levels.

- Wide Range of Incomes and Loan Amounts: The scatterplot shows a broad range of both annual incomes and loan amounts across different loan statuses. This is indicating a diverse borrower base.

- Correlation Between Income and Loan Amount: There appears to be a positive correlation between annual income and loan amount, with higher incomes generally associated with larger loan amounts.

- Higher Loan Amounts in Certain Default Cases: For some 'Charged Off' loans, especially at lower income levels, the loan amounts are relatively high. This is suggesting potential over-leveraging of borrowers.

- Income as a Factor in Loan Performance: While income is an important factor, the presence of 'Charged Off' loans across the income spectrum suggests that high income alone does not guarantee loan repayment.

- Current Loans Distribution: Loans in the 'Current' status are spread across various income levels and loan amounts, with no clear distinction from 'Fully Paid' or 'Charged Off' loans.

Multivariate Analysis: Annual Income, Loan Amount, and Loan Status

Annual Income vs Loan Amount by Loan Status

Multivariate Analysis: Debt-to-Income Ratio, Loan Amount, and Loan Status

- Fully Paid Loans Distribution: 'Fully Paid' loans are also distributed across the entire range of DTI ratios, which is indicating that a high DTI ratio does not singularly determine loan repayment success.

- Higher Loan Amounts Across DTI Ranges: There is a wide range of loan amounts across different DTI ratios, indicating that loans are extended to borrowers with varied debt burdens.

- Charged Off Loans' Characteristics: 'Charged Off' loans are present across all DTI levels, but they seem more concentrated in higher DTI ratios, suggesting that higher debt burdens might be linked to an increased likelihood of default.

- Current Loans Distribution: Loans categorized as 'Current' are evenly spread across different DTI ratios and loan amounts. This is reflecting the ongoing status of these loans.

- Risk Management Implications: While DTI is a critical factor in assessing borrowers' ability to manage new debt, this analysis suggests it should be considered alongside other factors, like loan amount, for a comprehensive risk assessment.

- Interplay of Loan Amount and DTI: The relationship between loan amount and DTI for different loan statuses indicates that both factors, along with others, play a role in determining loan performance.

- Distribution Across Loan Statuses: The scatterplot shows that loans of various amounts and DTI ratios are spread across all loan statuses, with no immediate, distinct pattern separating 'Fully Paid', 'Charged Off', and 'Current' loans.

Multivariate Analysis: Debt-to-Income Ratio, Loan Amount, and Loan Status

Debt-to-Income Ratio vs Loan Amount by Loan Status