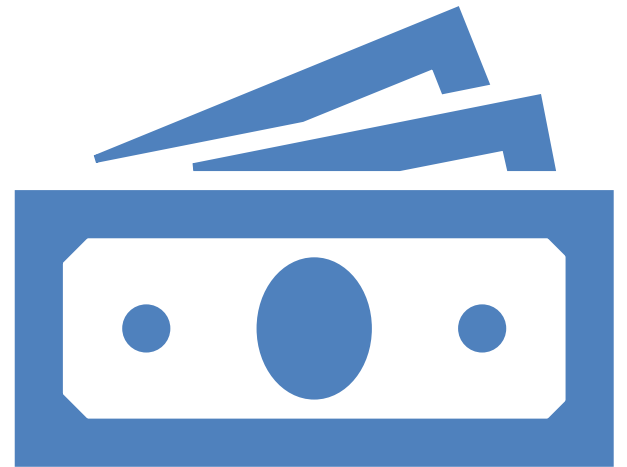


Exploratory Data Analysis on Loan
Dataset

Lending Club Loan Default Analysis

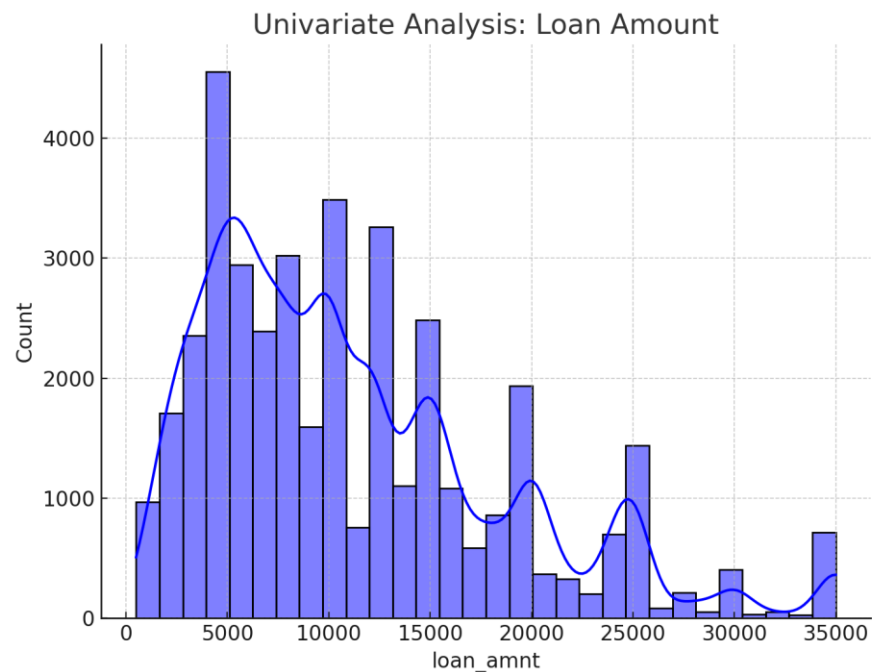


Introduction

- This analysis aims to understand factors influencing loan defaults. Insights are derived using various visualization techniques to assist in better decision-making.

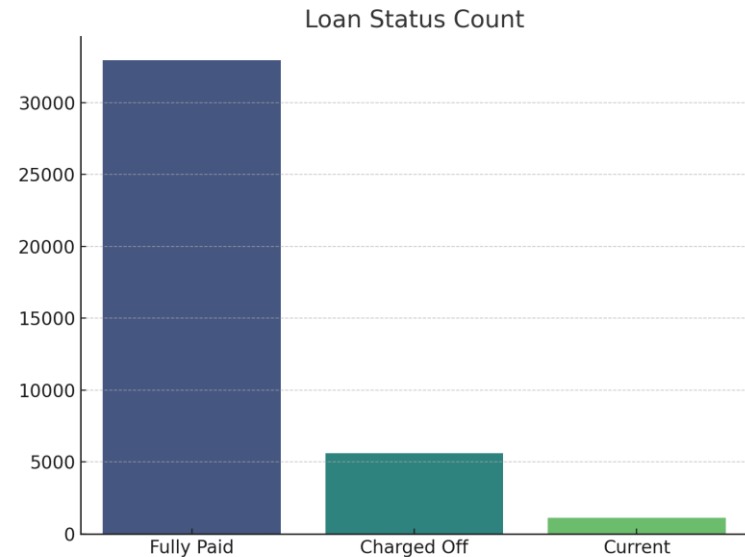
Univariate Analysis: Loan Amount

- Most loans are in the range of \$5,000 to \$15,000, with a peak around \$10,000. This analysis highlights the common loan amounts, useful for understanding customer preferences.
- Conclusion: Loan amounts show a concentrated distribution. Most borrowers prefer moderate loans, indicating typical credit demand patterns.



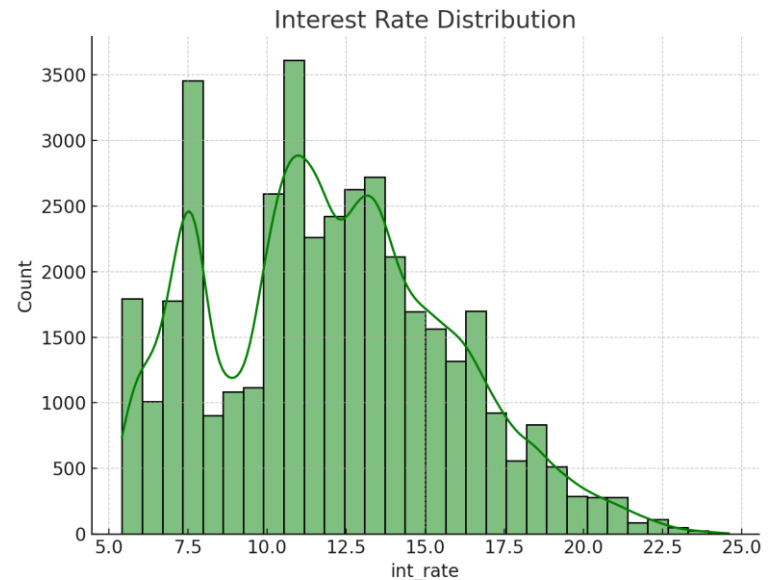
Loan Status Count

- Most loans are 'Fully Paid,' but a significant portion is 'Charged Off.' This shows the need for deeper analysis into factors influencing defaults.
- Conclusion: The company should focus on reducing 'Charged Off' loans through enhanced risk assessment.



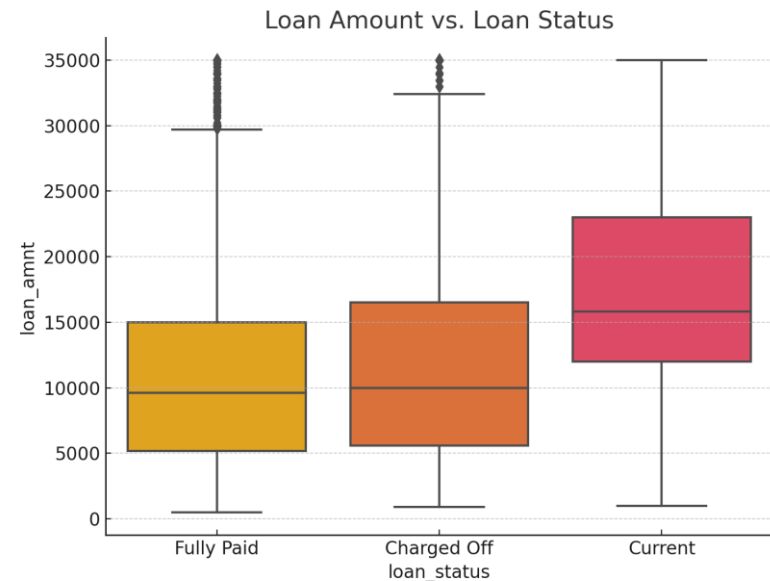
Interest Rate Distribution

- Interest rates are concentrated between 10% and 20%, with higher rates often indicating riskier loans. Understanding interest rates helps assess risk pricing strategies.
- Conclusion: Higher interest rates correlate with riskier borrowers, highlighting the importance of rate adjustments.



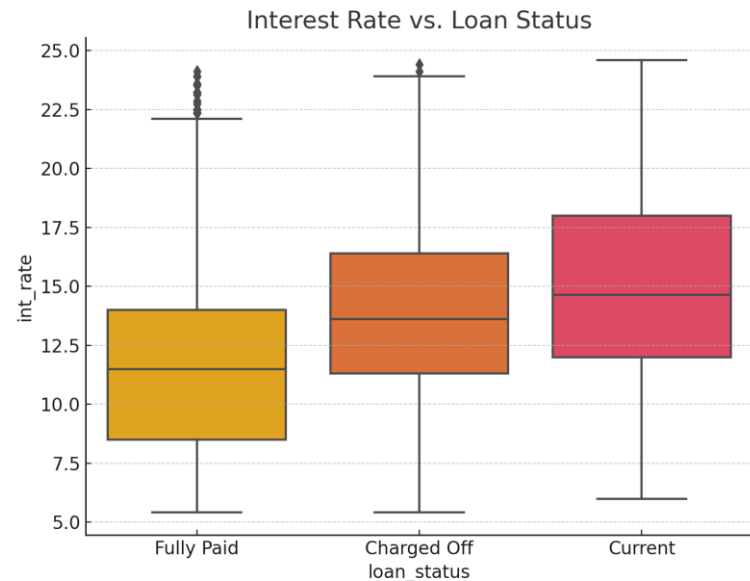
Loan Amount vs. Loan Status

- Defaults occur across all loan amounts, but higher amounts show more variance. Loan amount plays a partial role in understanding default risk.
- Conclusion: High variance in loan amounts for defaulters suggests combining this with other metrics for better predictions.



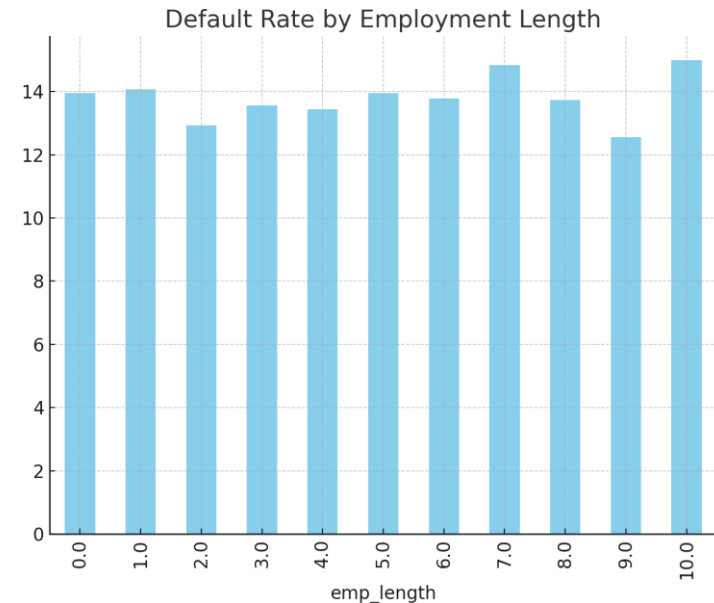
Interest Rate vs. Loan Status

- Defaults are more common in loans with higher interest rates (>15%). This emphasizes the role of interest rates in predicting loan performance.
- Conclusion: Target borrowers with high-interest loans for better evaluation to reduce defaults.



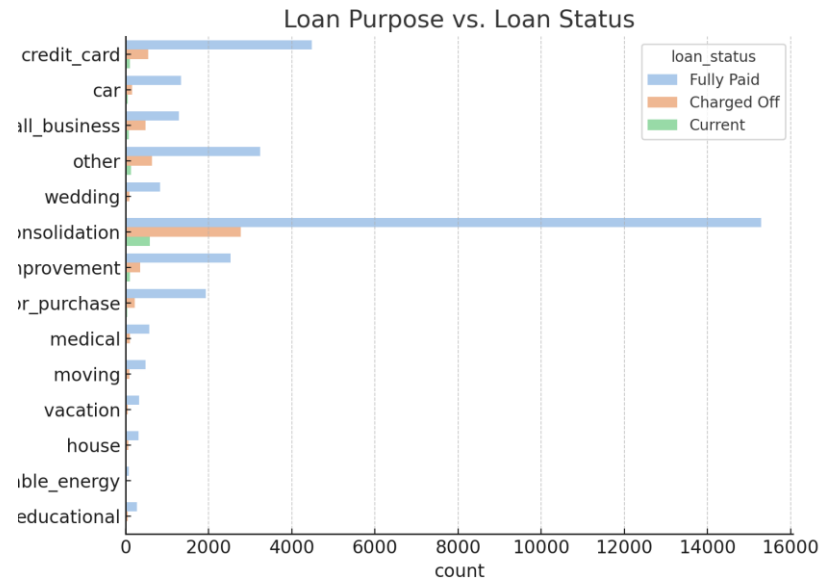
Default Rate by Employment Length

- Default rates vary slightly with employment length, being slightly higher for longer tenures. Employment length provides secondary insights into customer stability.
- Conclusion: While not a primary predictor, employment length provides insights into financial reliability.



Loan Purpose vs. Loan Status

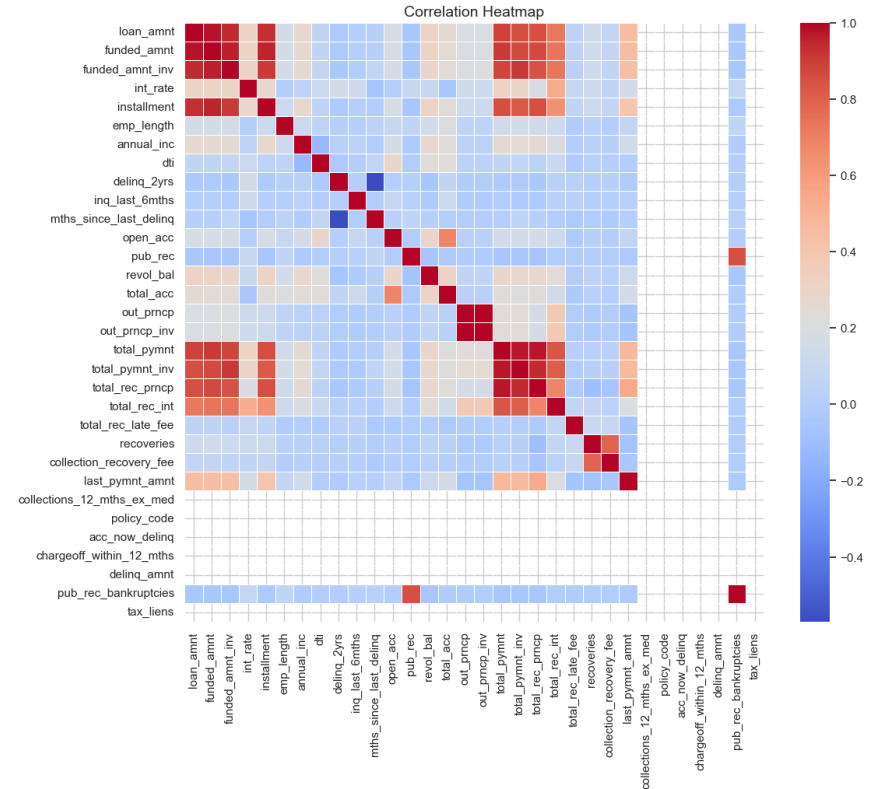
- Loans for 'Debt Consolidation' dominate, but 'Small Business' loans show higher default rates. Understanding loan purpose is critical for portfolio diversification.
- Conclusion: Loan purposes with high default rates require stricter evaluation criteria.



Correlation Heatmap

- Strong correlations exist between 'loan_amnt,' 'funded_amnt,' and 'installment.'
- Moderate correlations with 'int_rate' indicate its role in default prediction.

- Conclusion: High correlations highlight potential redundancy in features; feature selection can improve modeling.



Conclusion

- The analysis highlights critical factors influencing loan defaults, such as interest rates, loan amounts, employment length, and loan purposes. These insights can guide risk-based pricing and improve credit evaluation processes.