

Introduction

- **Objective:** The primary objective of this analysis is to minimize credit risk by identifying loan applicants who are likely to default. By doing so, the company can reduce financial losses and make informed lending decisions.
- **Business Context:** The company is facing significant credit losses due to defaults on loans. Understanding the factors that contribute to loan defaults can help the company implement effective risk management strategies.

Data Understanding

- **Dataset Description:** The dataset contains detailed information about loan applicants from 2007 to 2011, including their demographic information, loan details, and whether they defaulted.
- **Key Variables:**
 - Loan amount
 - Interest rate
 - Annual income
 - Loan term
 - Loan grade

- **Data Quality Issues:** Initial inspection revealed missing values, outliers, and inconsistencies in the data. Addressing these issues is crucial for accurate analysis.

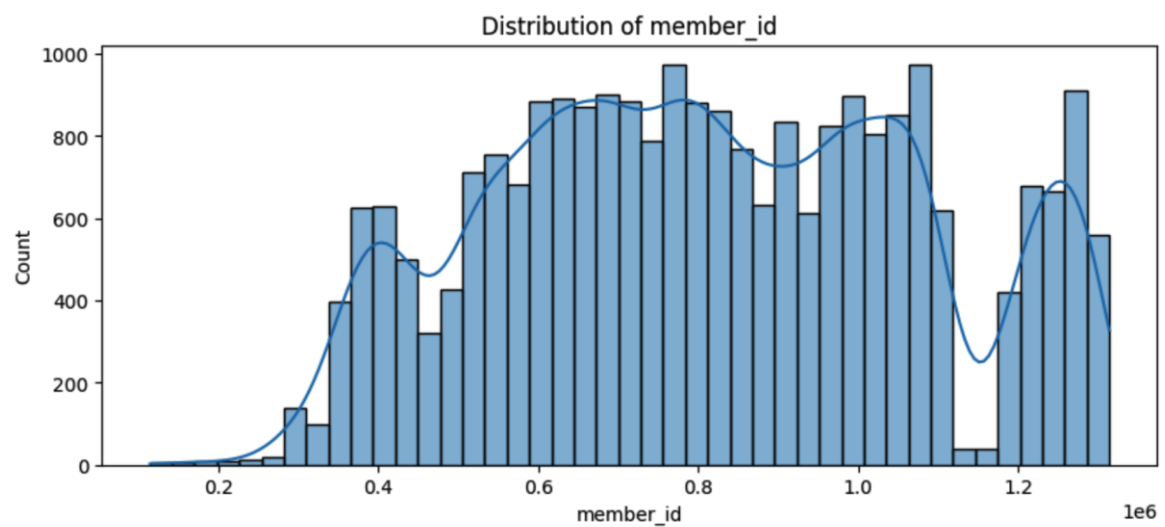
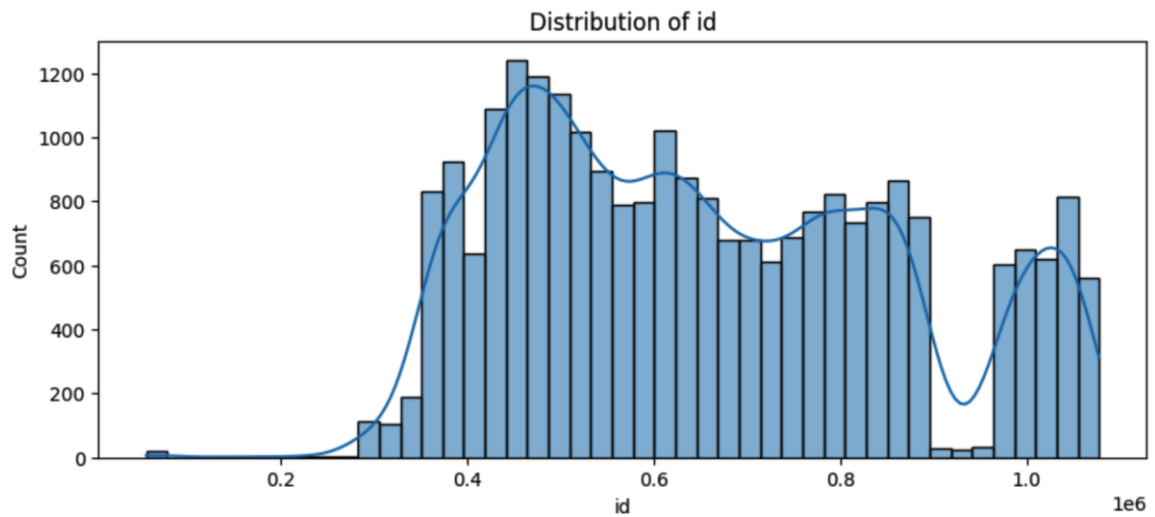
Data Cleaning and Manipulation

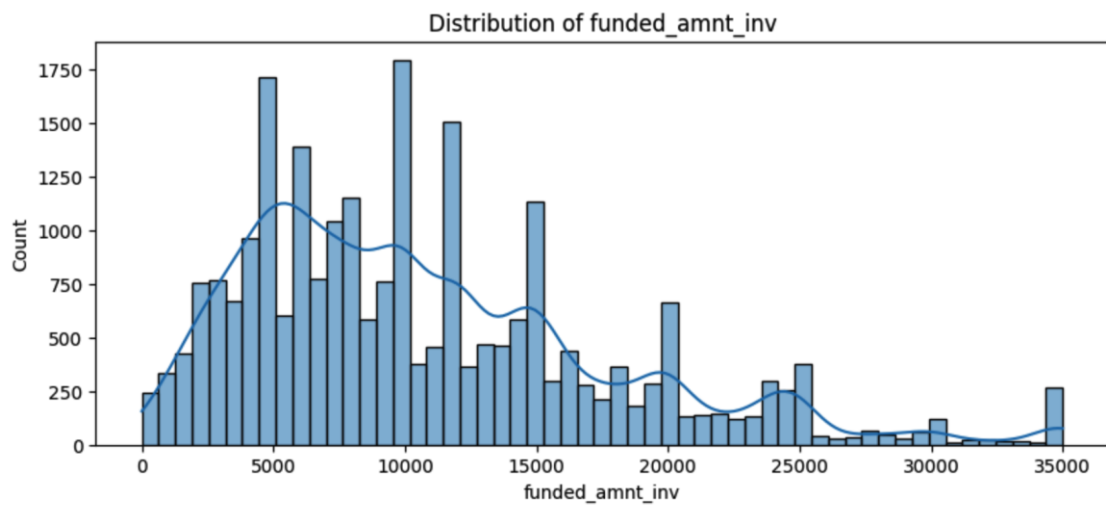
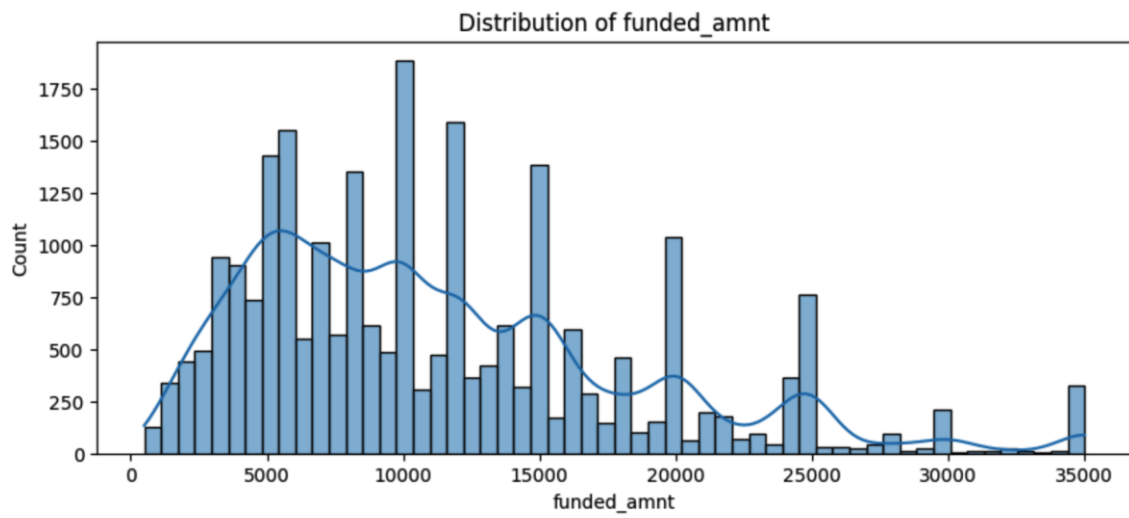
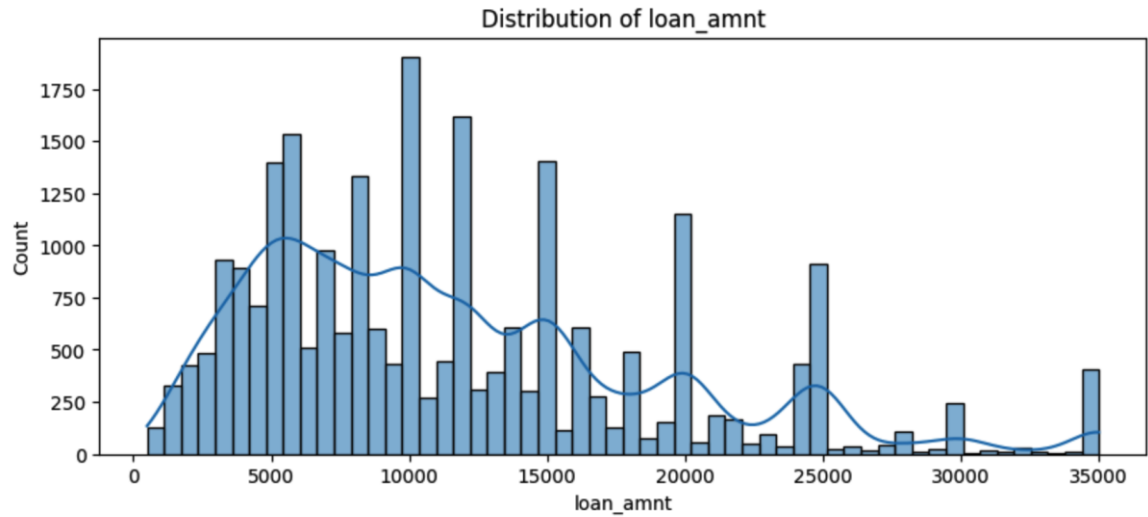
- **Steps Taken:**
 - **Handling Missing Values:** Used imputation techniques for missing data.
 - **Outlier Treatment:** Identified and treated outliers to avoid skewing the analysis.
 - **Data Transformations:** Converted data into suitable formats for analysis (e.g., converting dates, categorizing variables).
- **Example:** Missing values in the 'annual income' field were imputed using the median income of similar applicants.

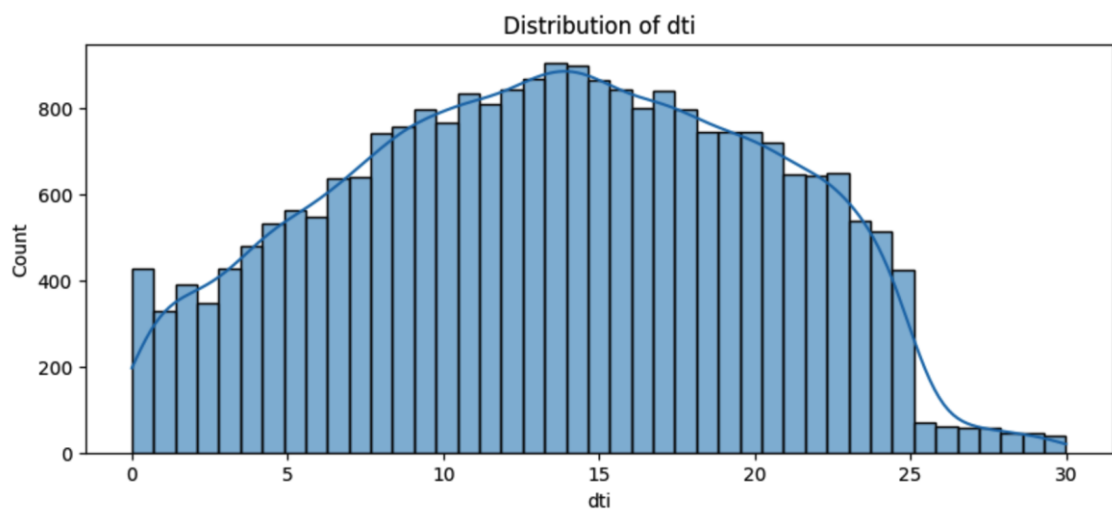
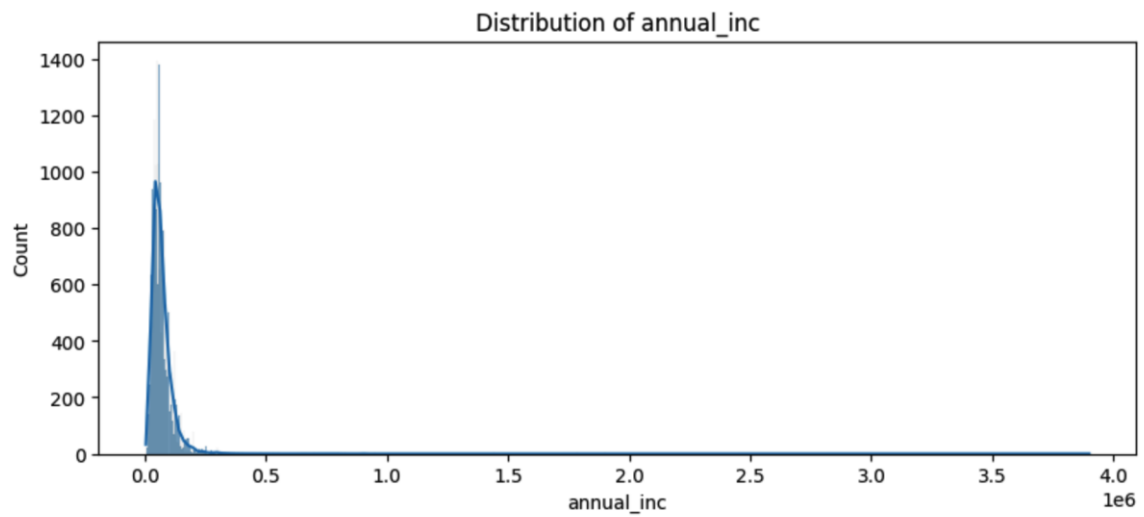
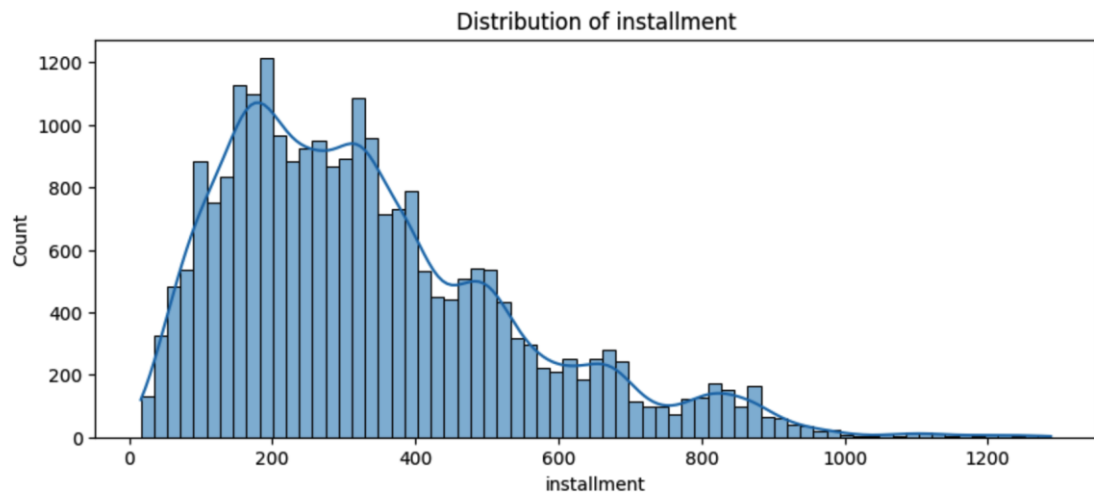
Univariate Analysis

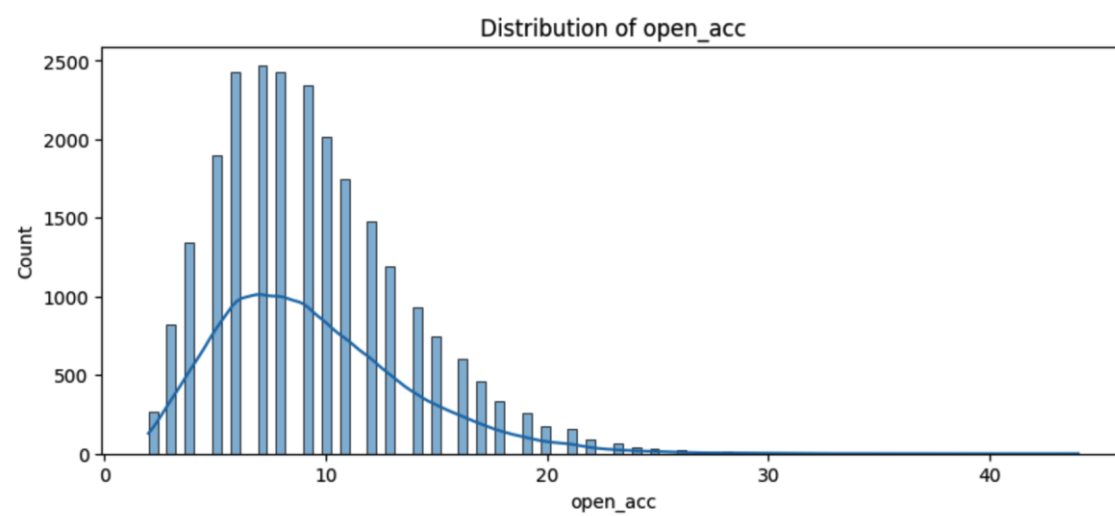
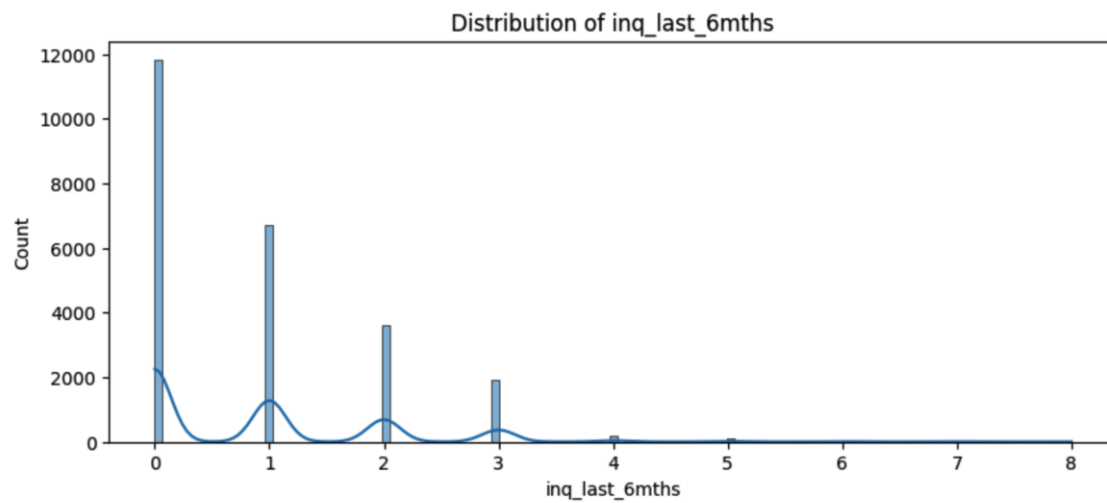
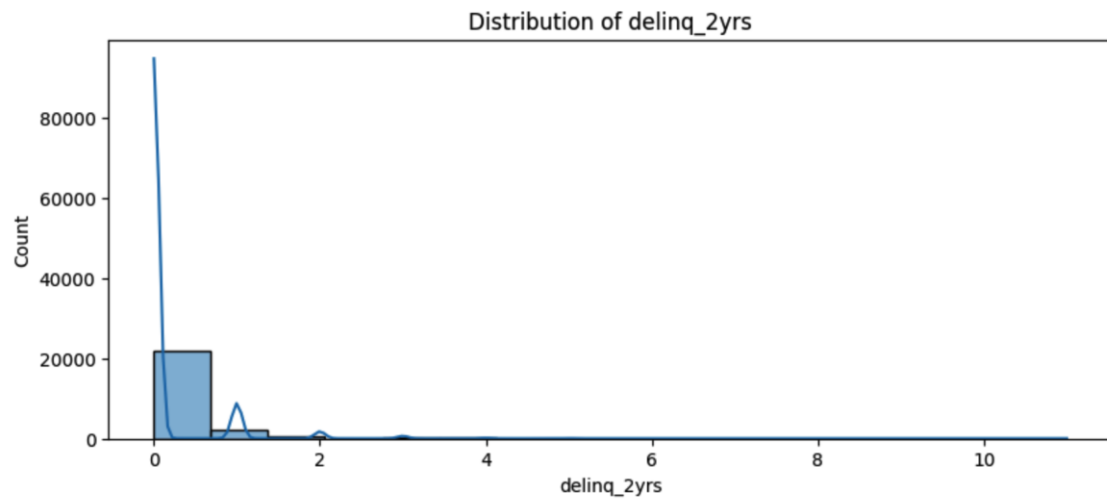
- **Key Variables:**
 - **Loan Amount:** Most loans fall within a specific range, with a few very high amounts.
 - **Interest Rate:** Interest rates vary, with higher rates potentially indicating higher risk.

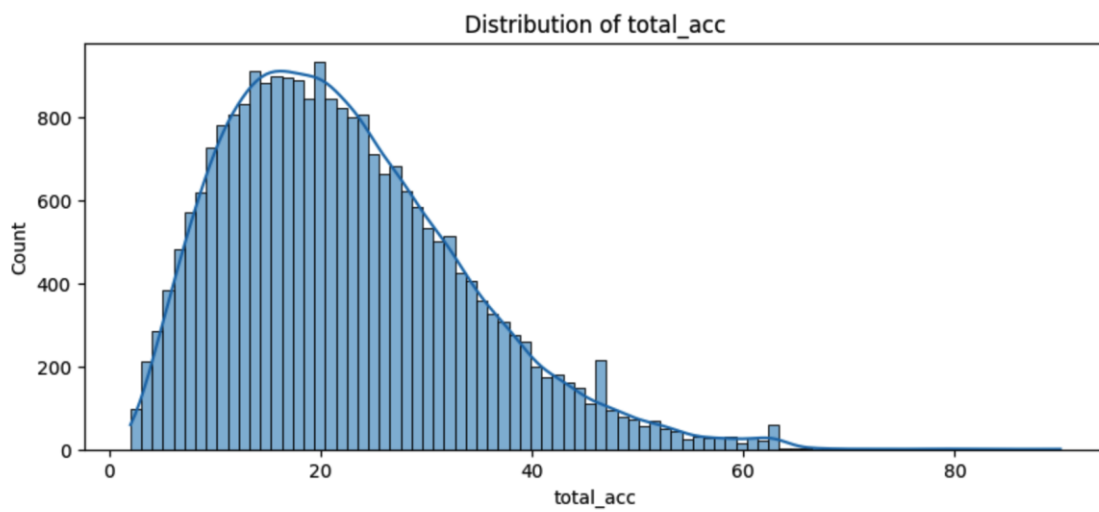
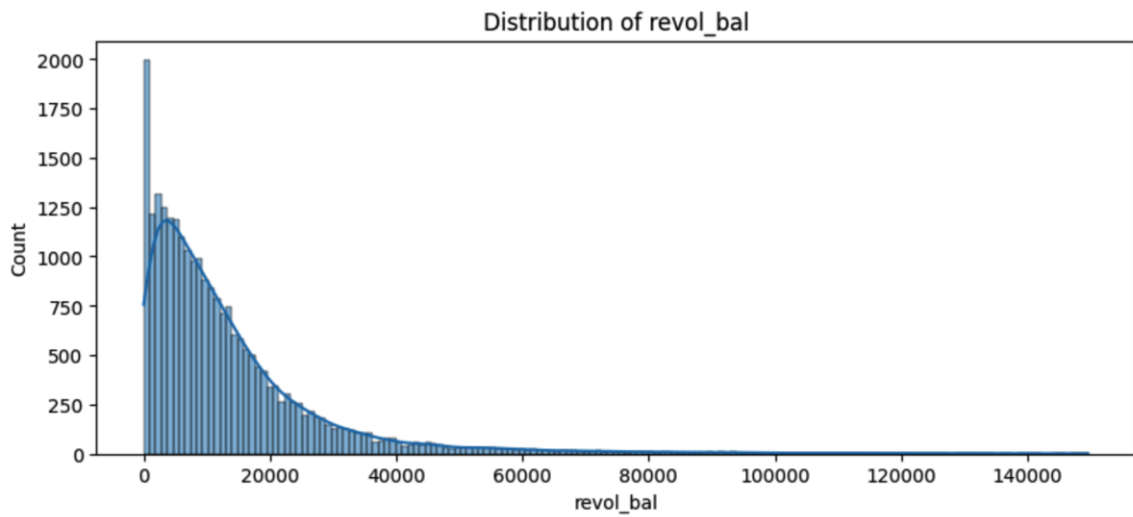
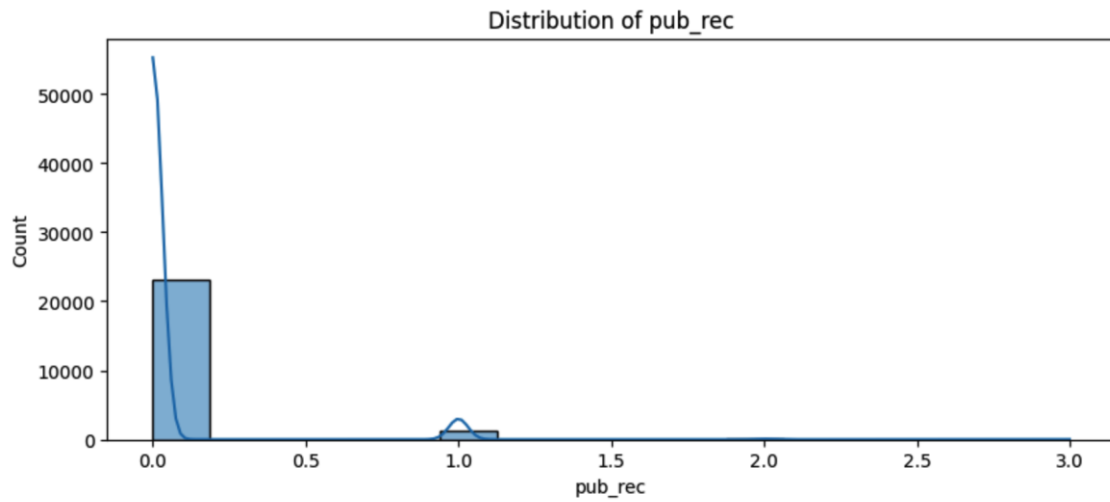
- **Annual Income:** Income levels vary widely among applicants.
- **Insights:** Initial analysis shows that loan amount and interest rate are critical factors. Higher loan amounts and interest rates may be associated with a higher risk of default.

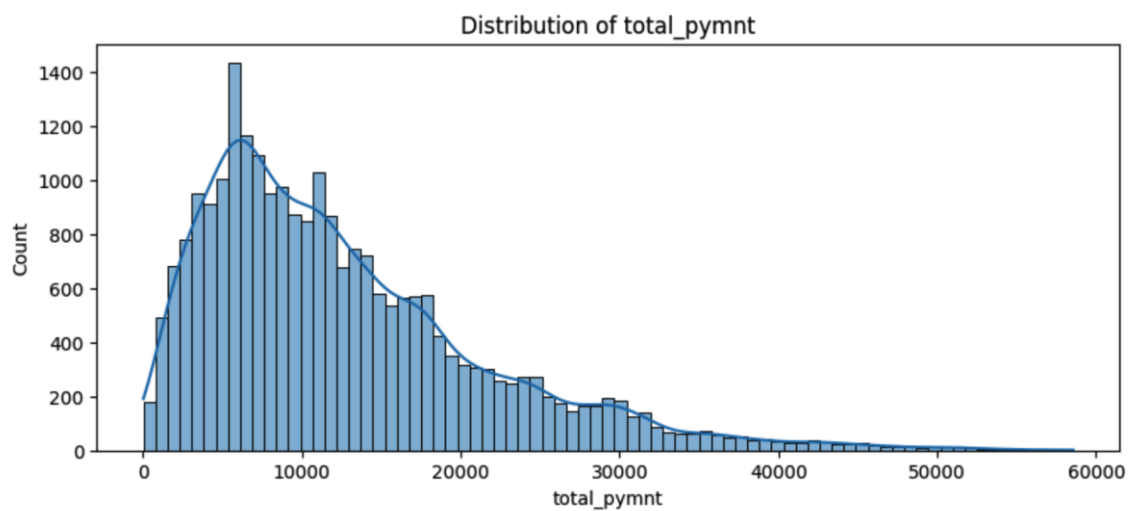
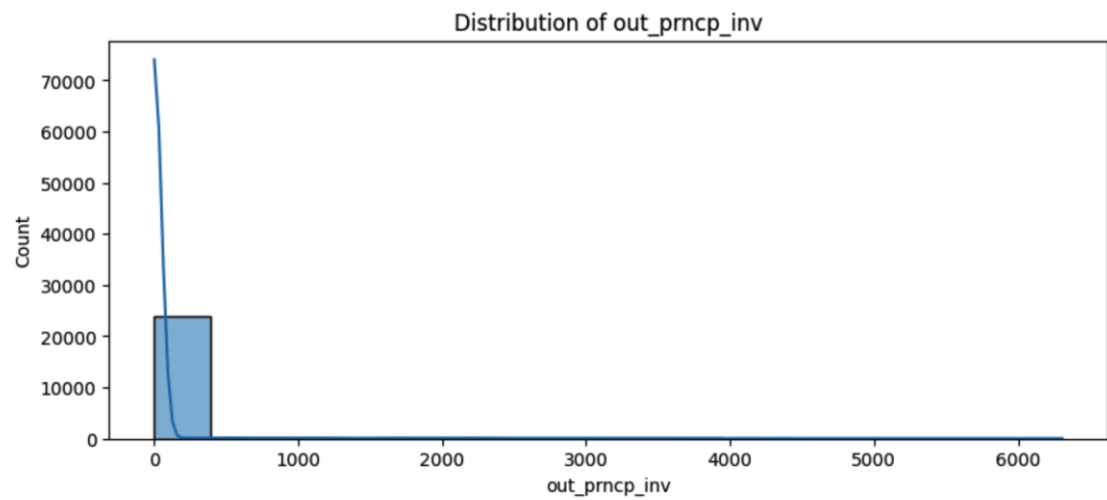
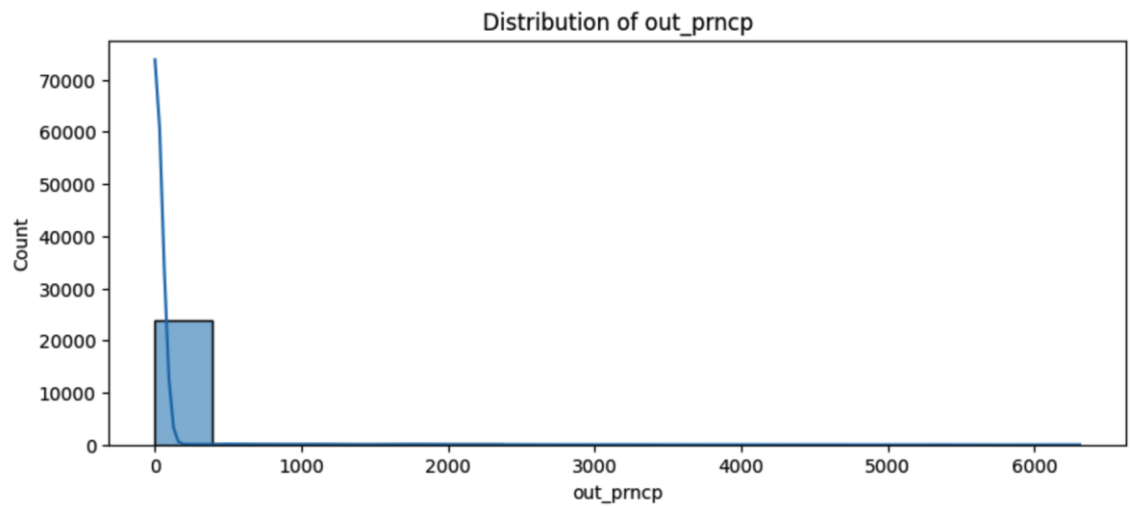


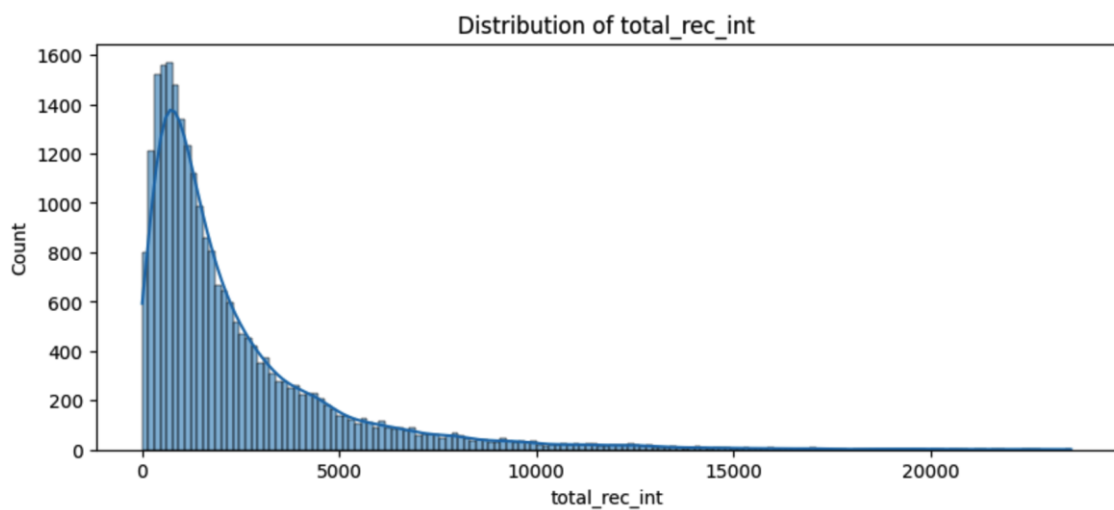
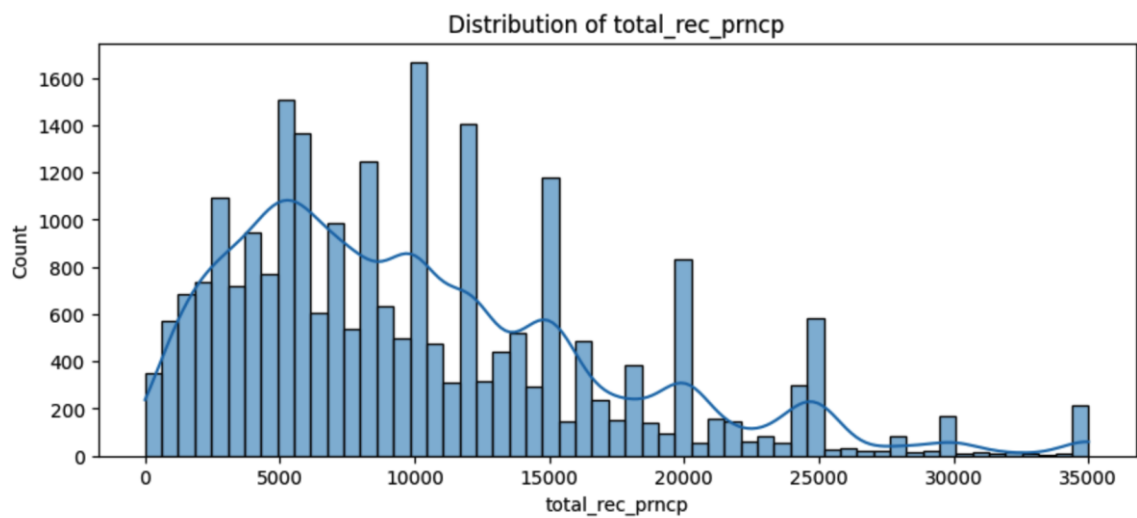
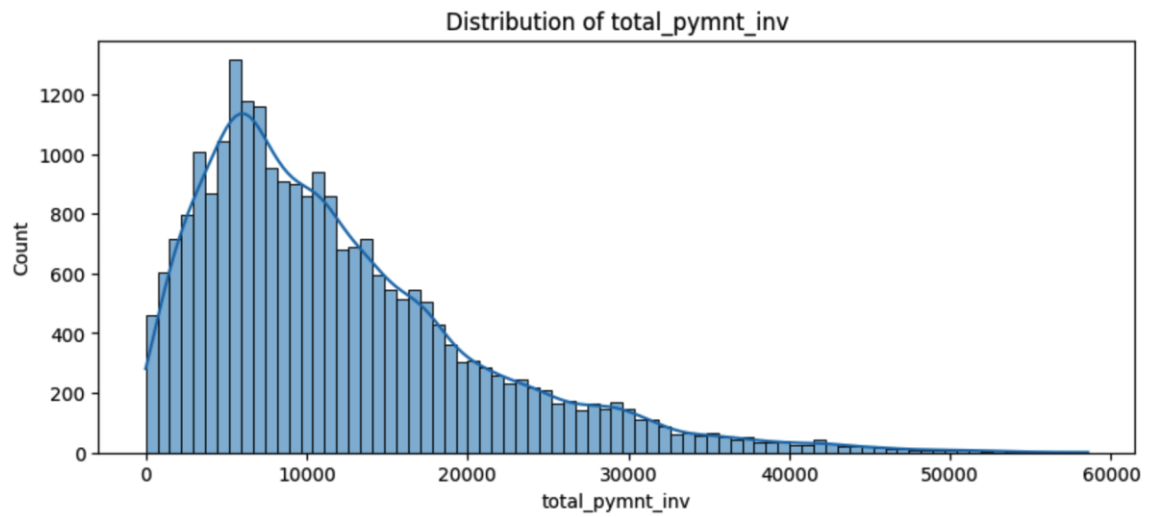


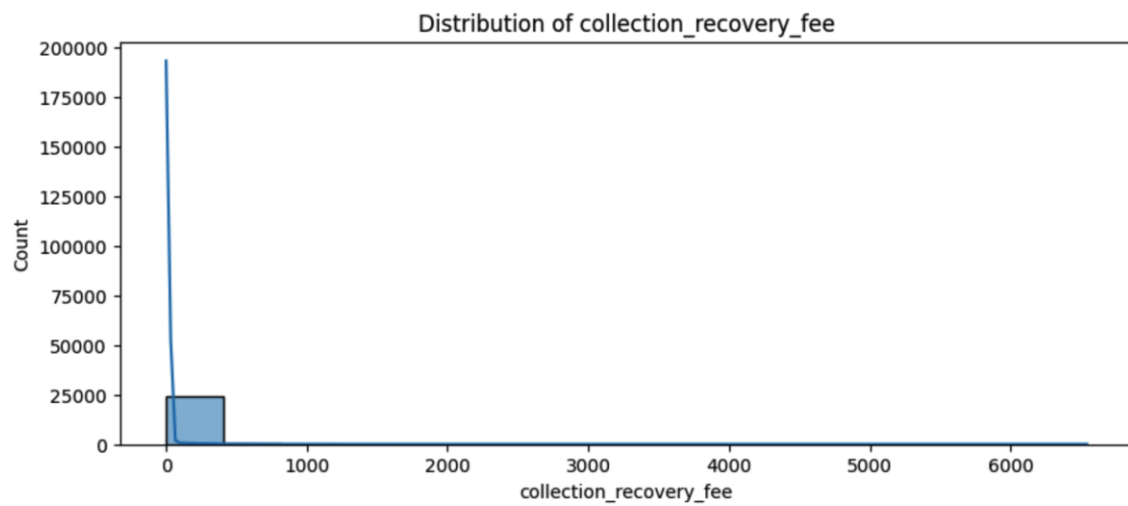
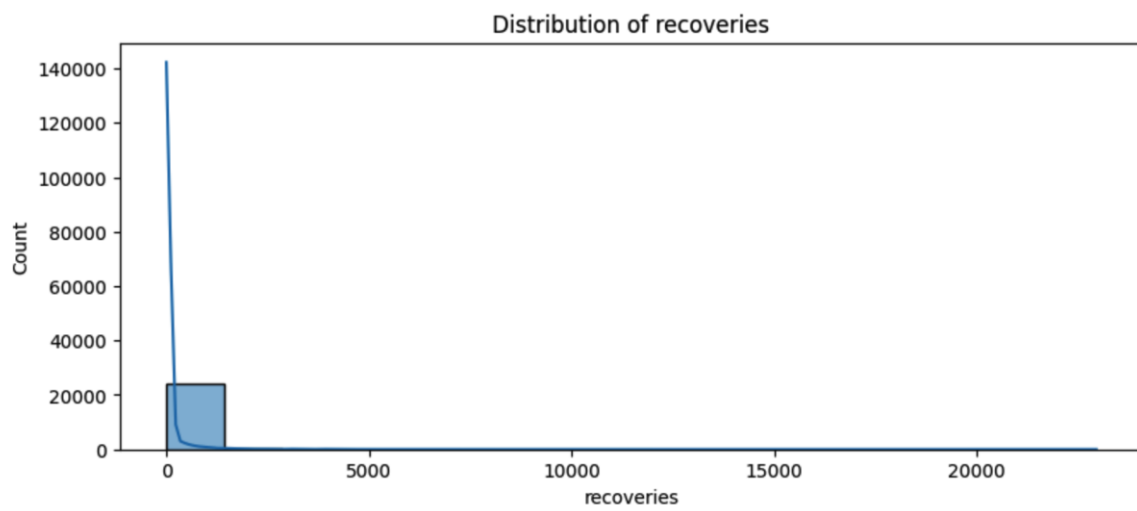
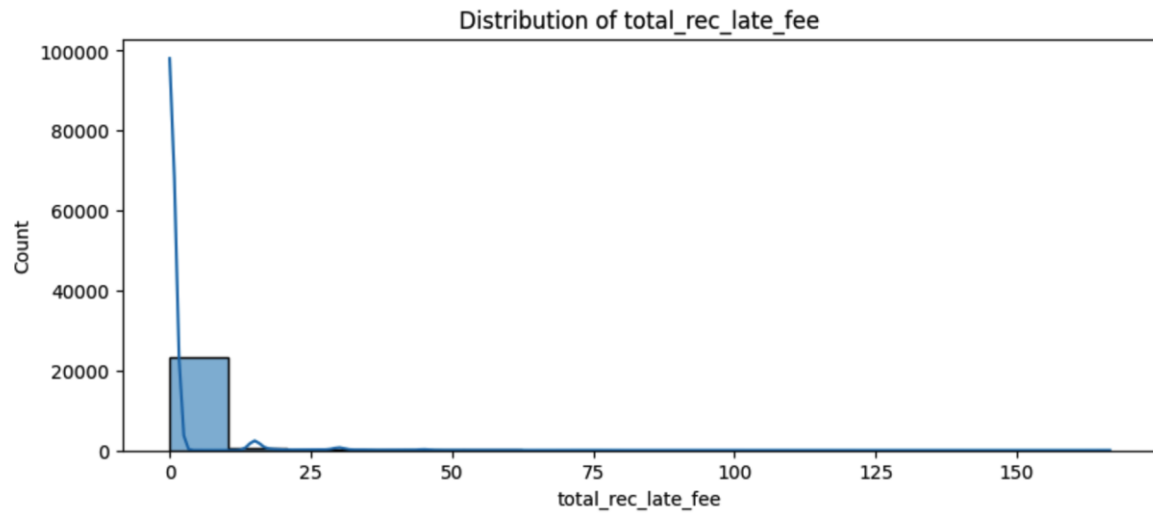


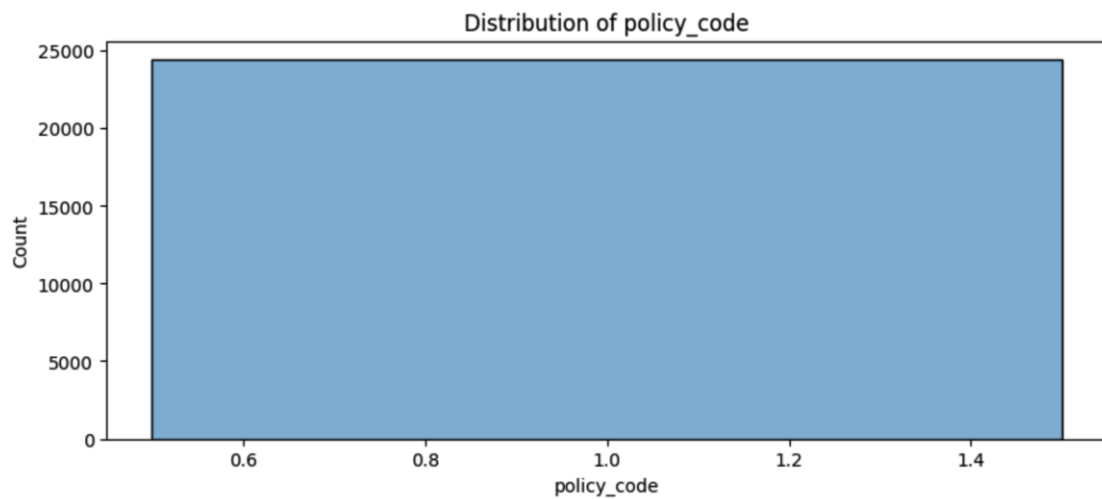
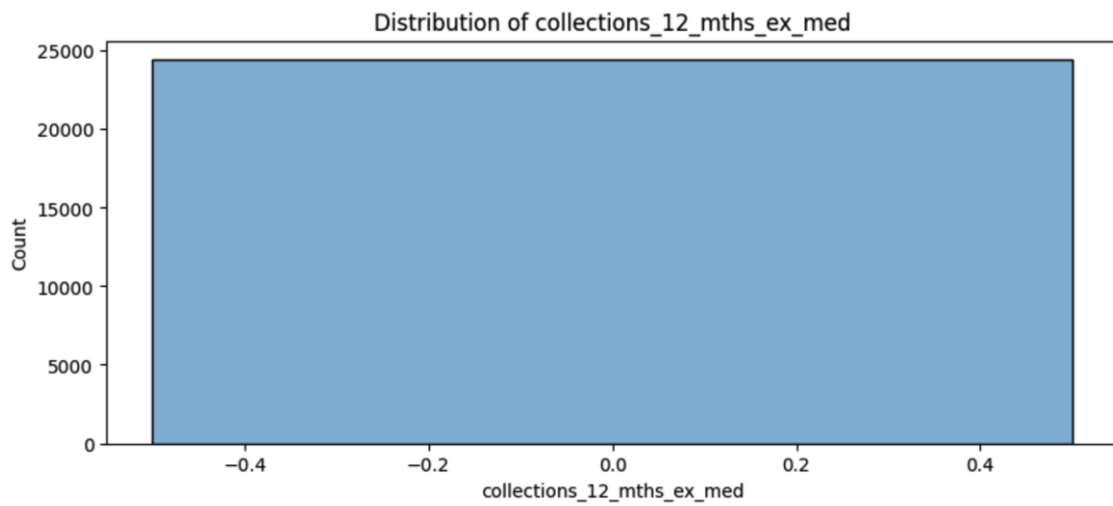
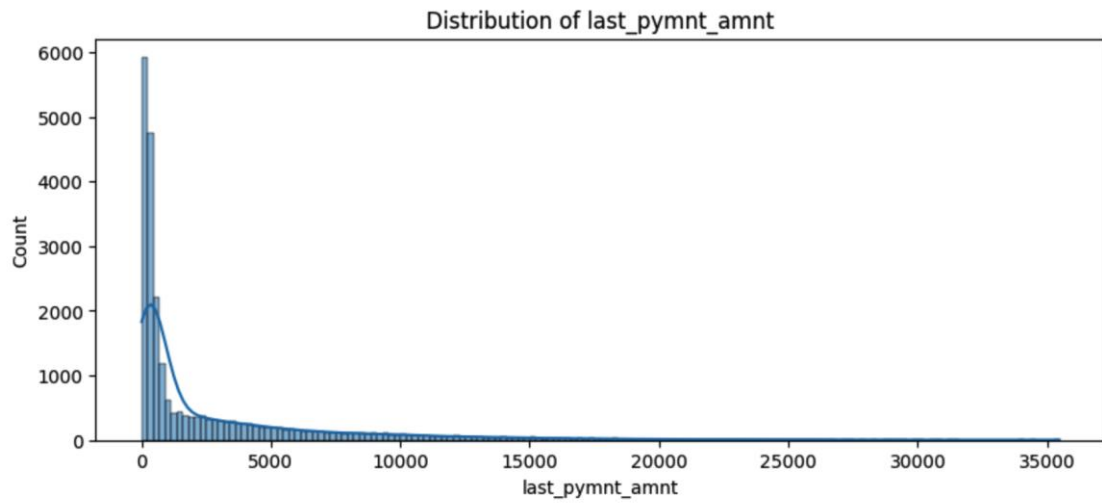


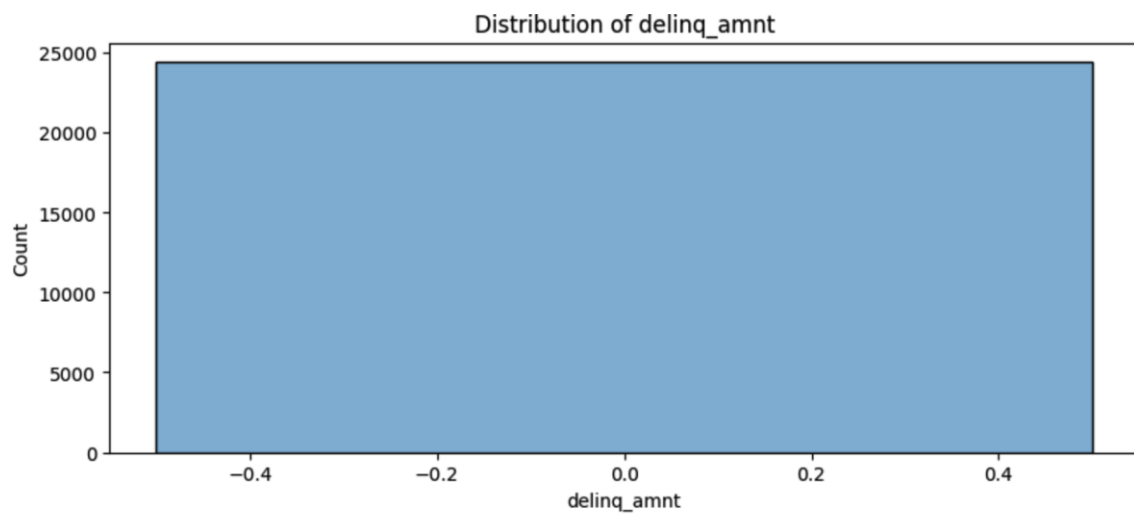
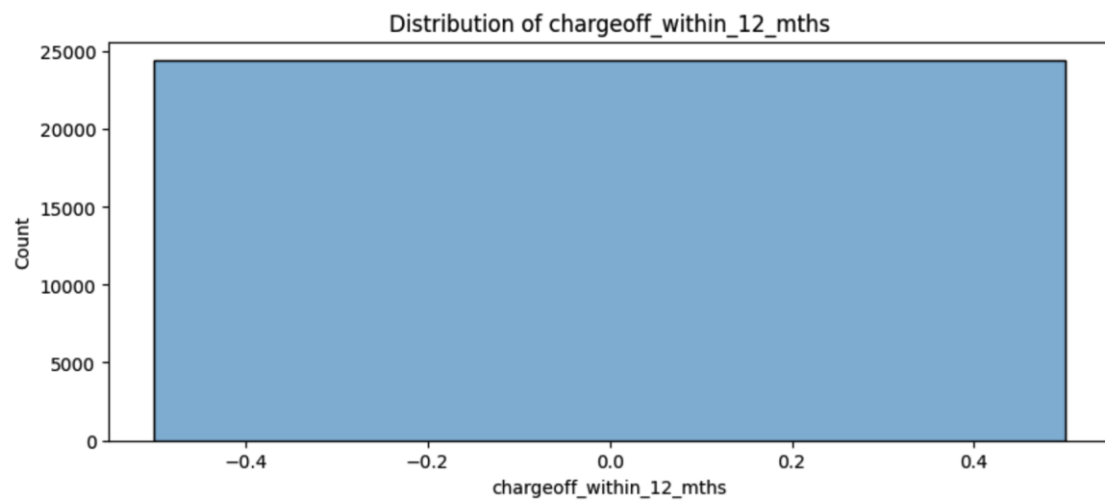
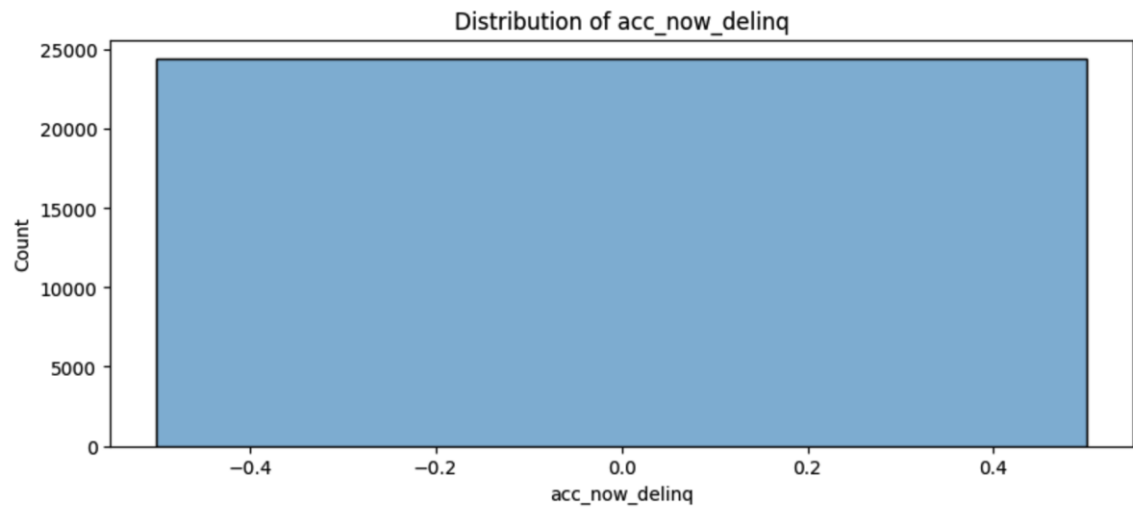


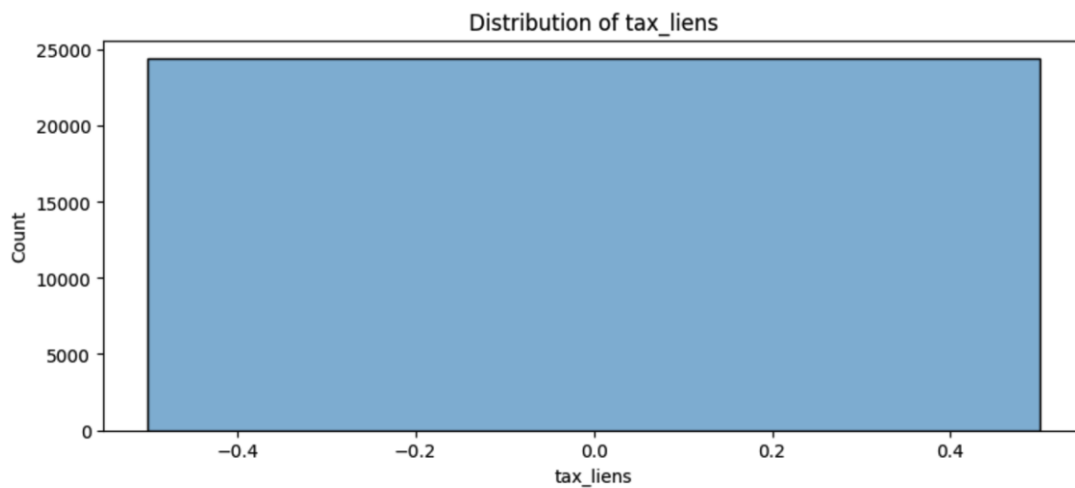
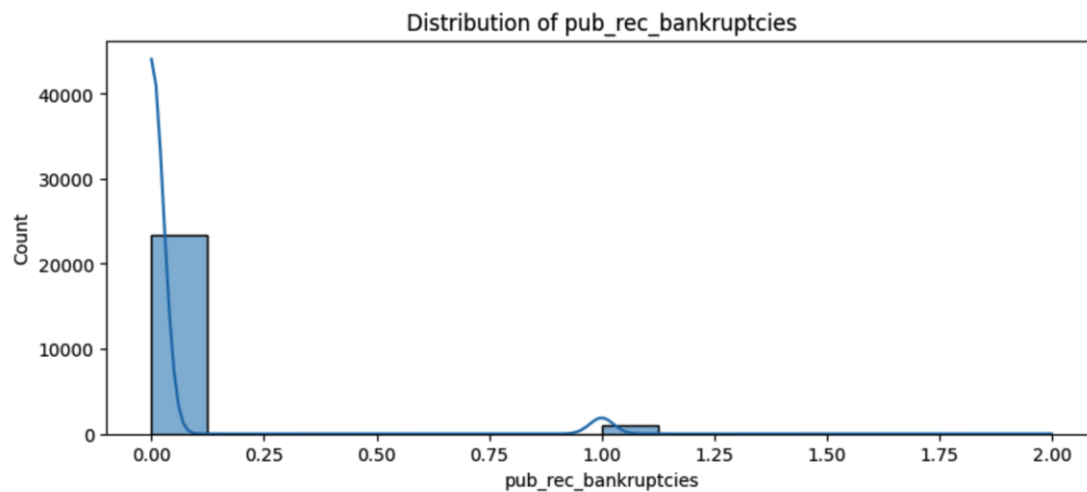






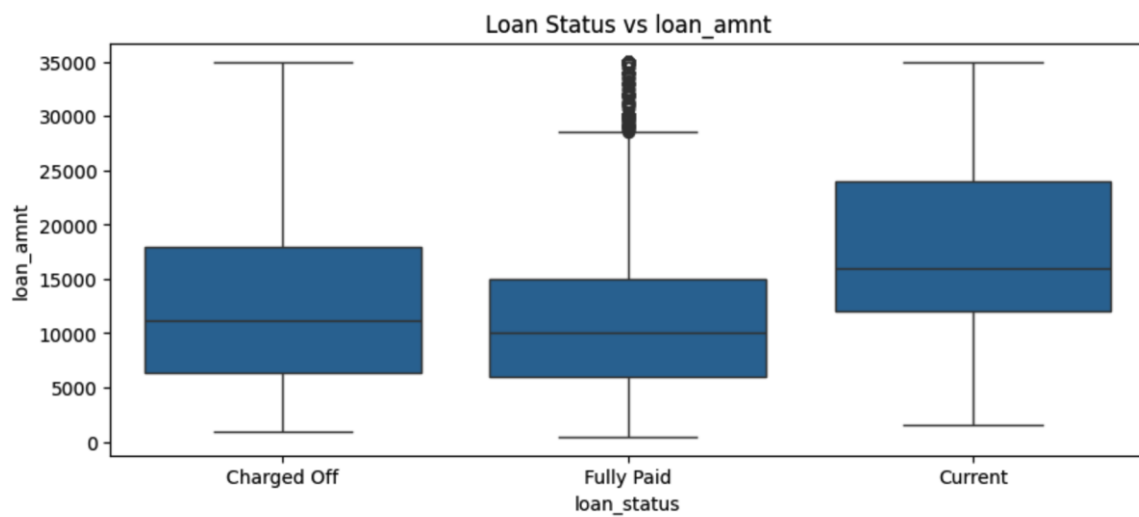
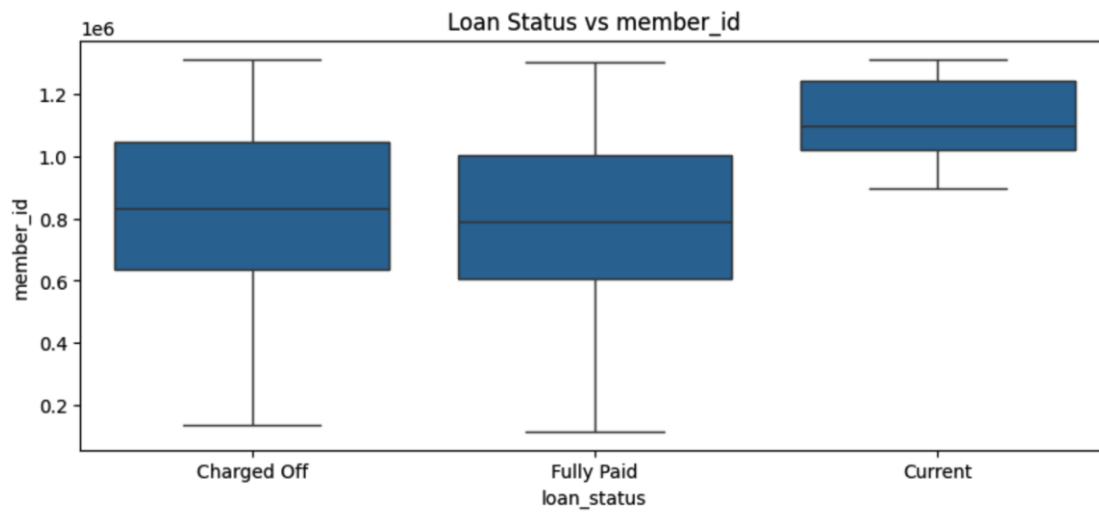
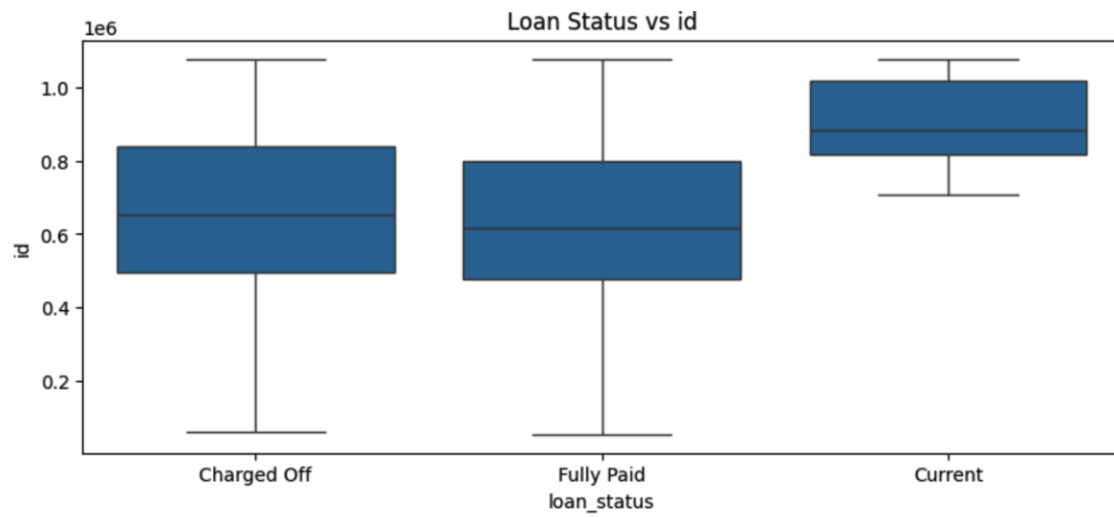


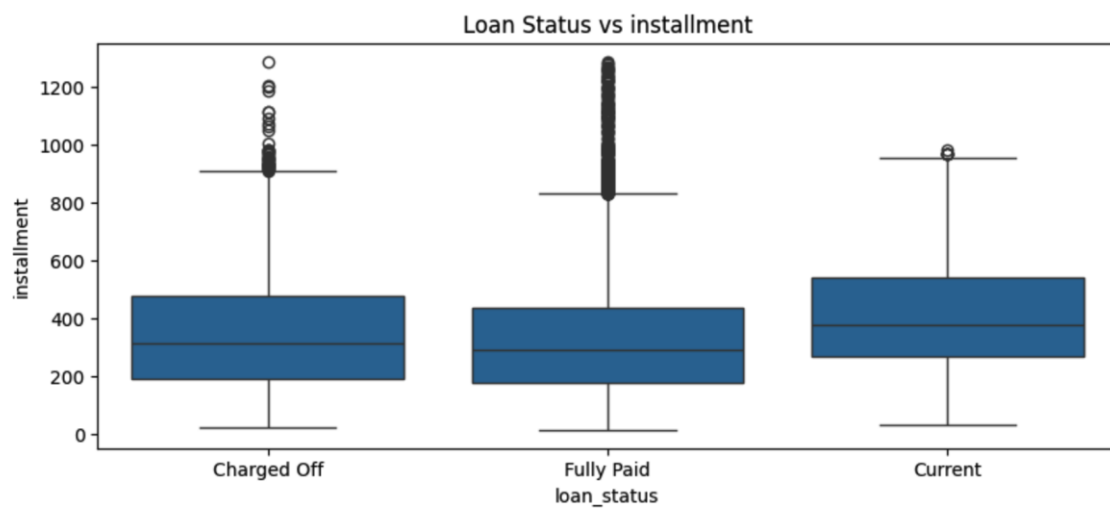
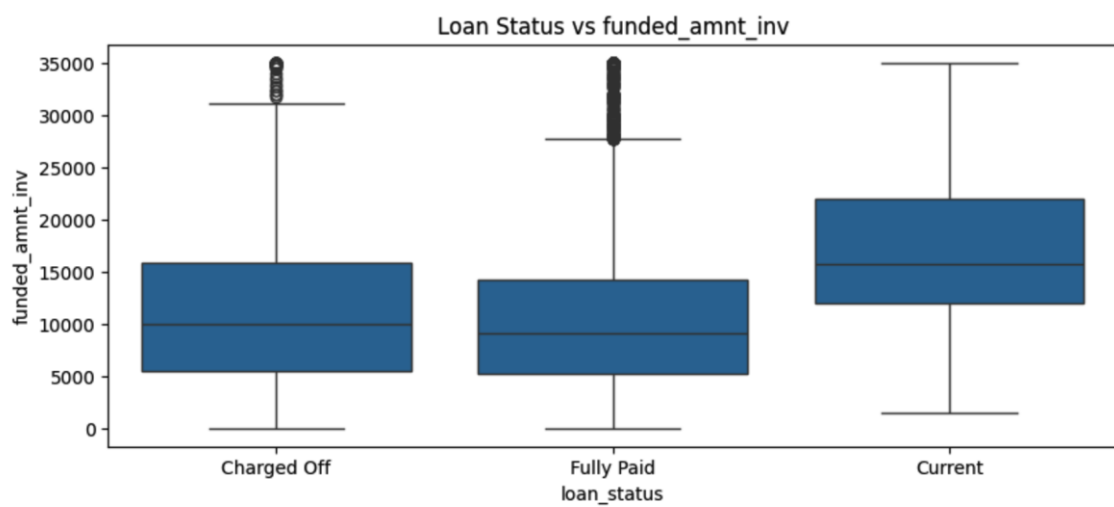
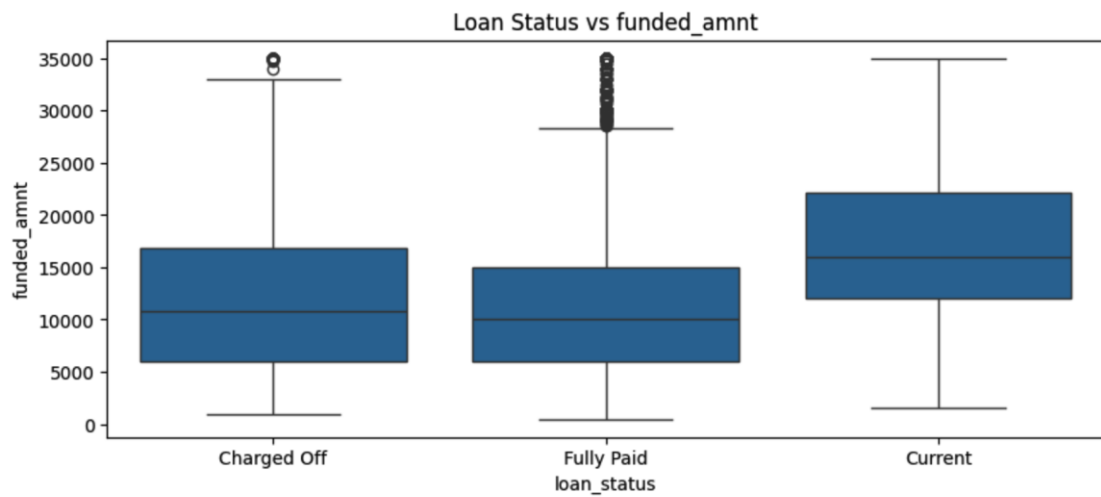


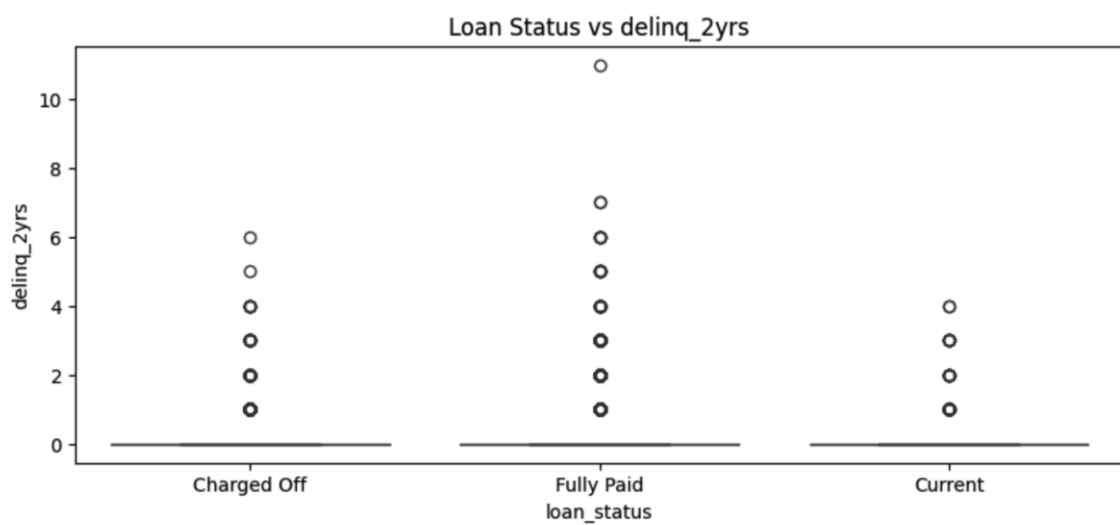
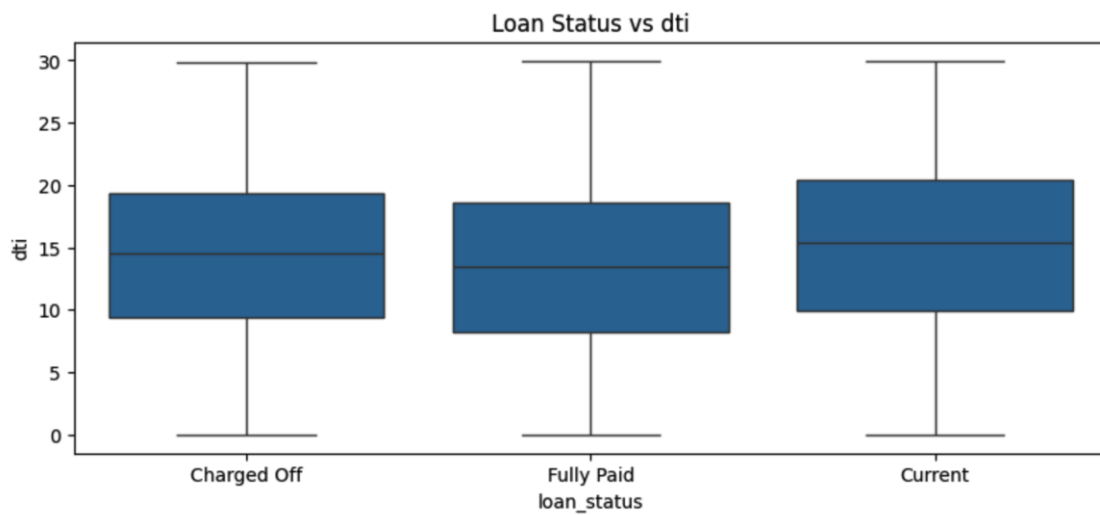
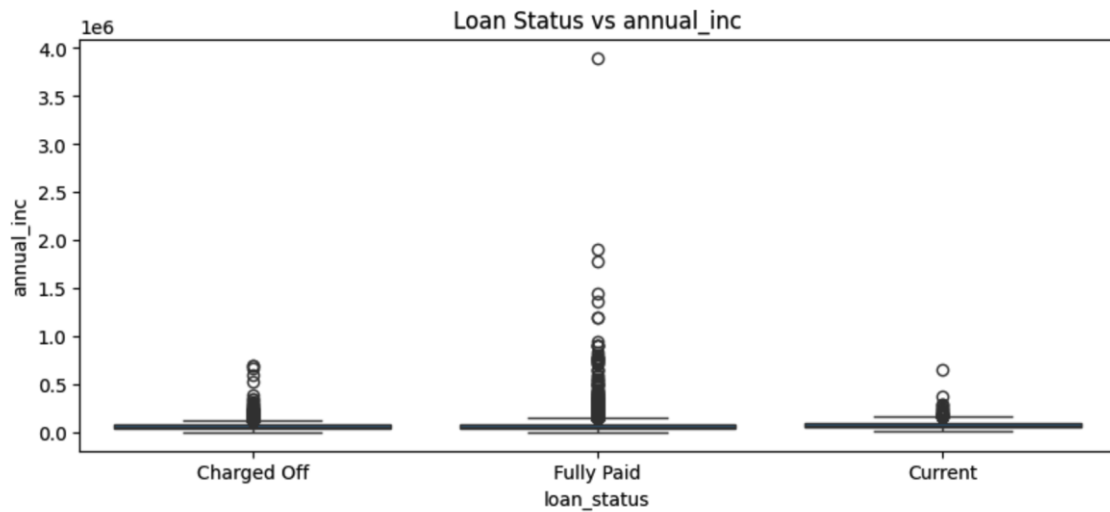


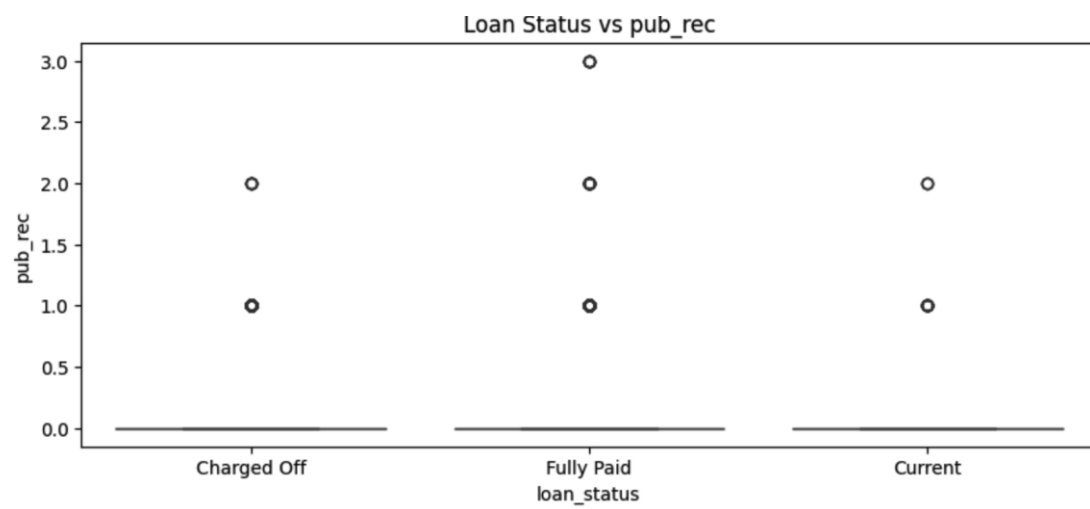
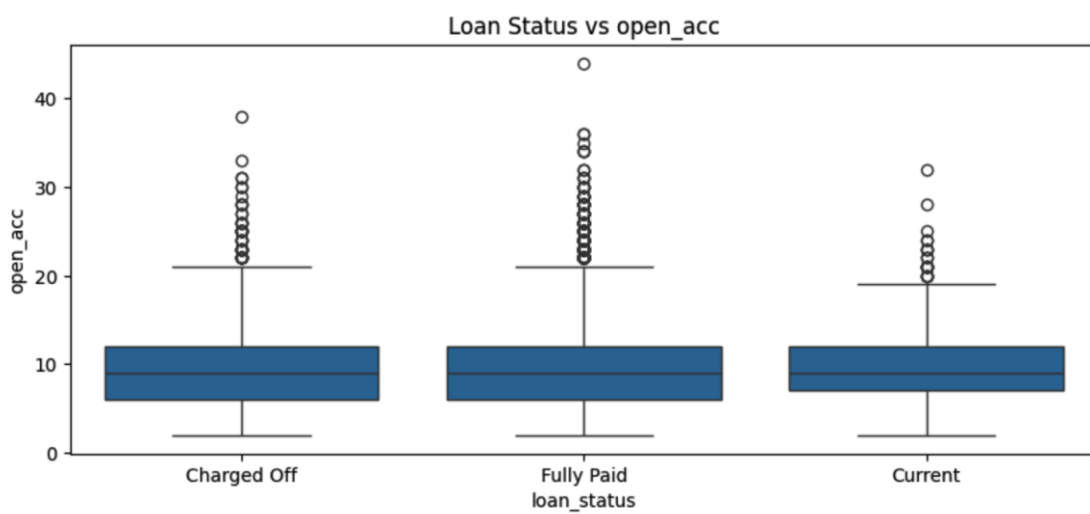
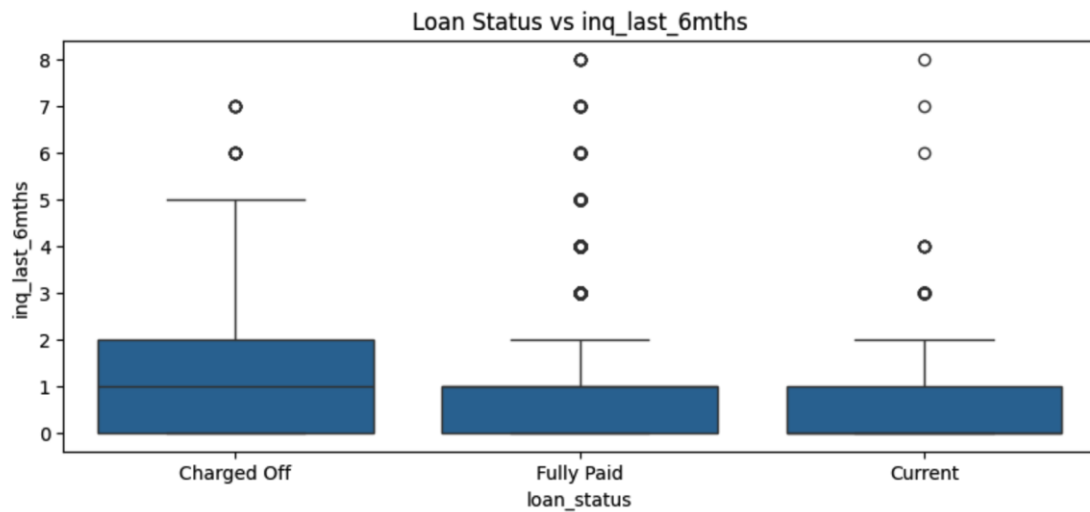
Bivariate Analysis

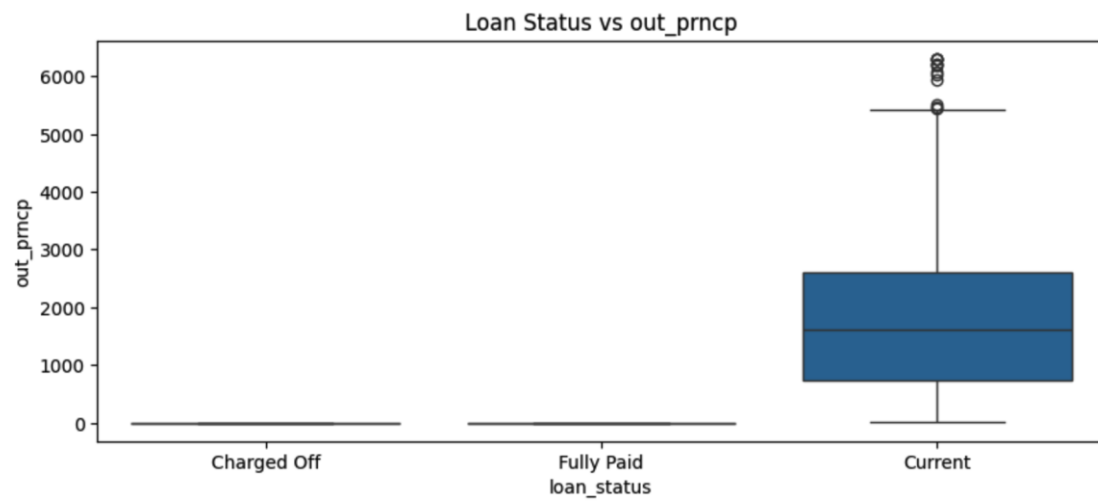
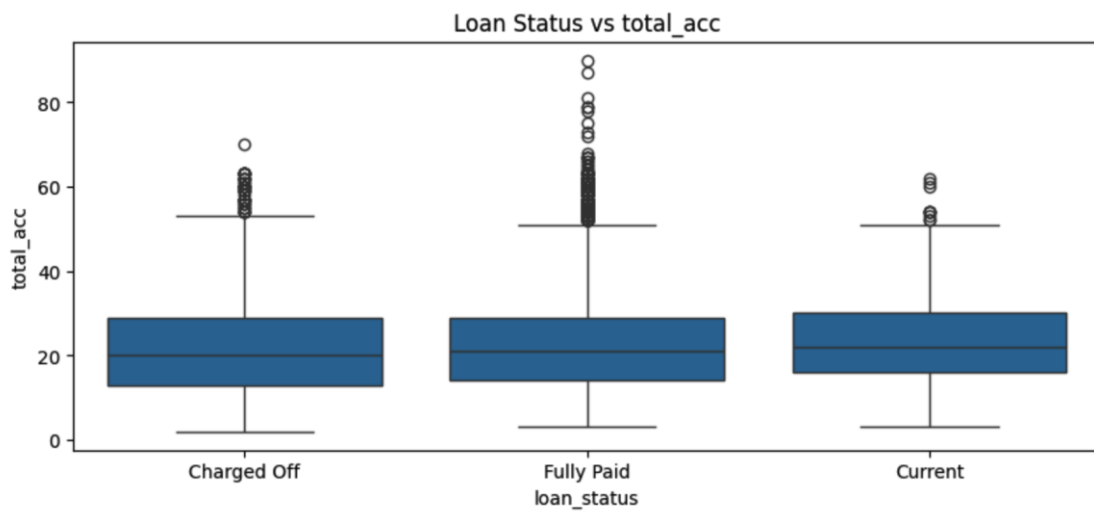
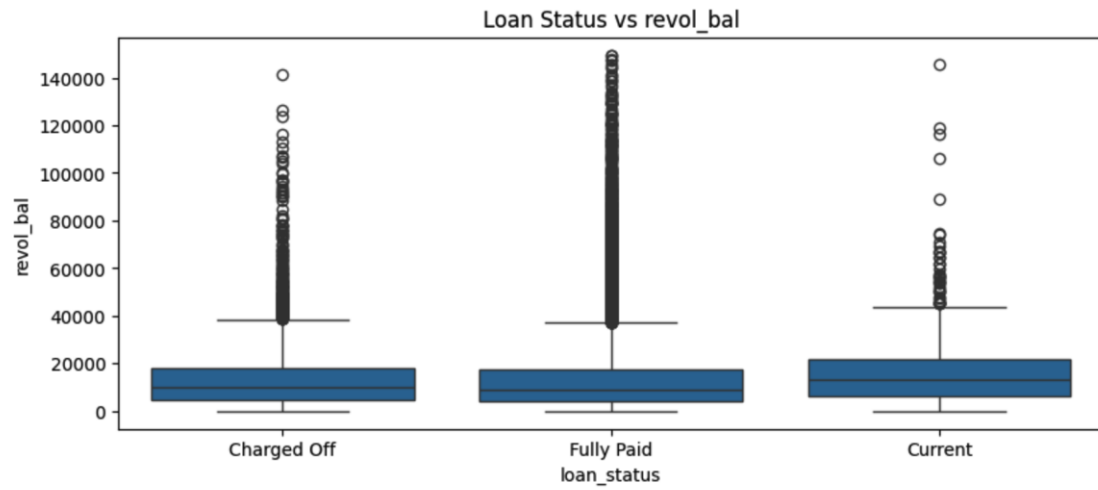
- **Combinations Analyzed:**
 - **Loan Amount vs. Interest Rate:** Higher loan amounts often come with higher interest rates.
 - **Annual Income vs. Default Status:** Applicants with lower annual incomes are more likely to default.
- **Insights:** The relationship between loan amount and interest rate is significant. Higher interest rates are often associated with higher loan amounts and increased risk of default.

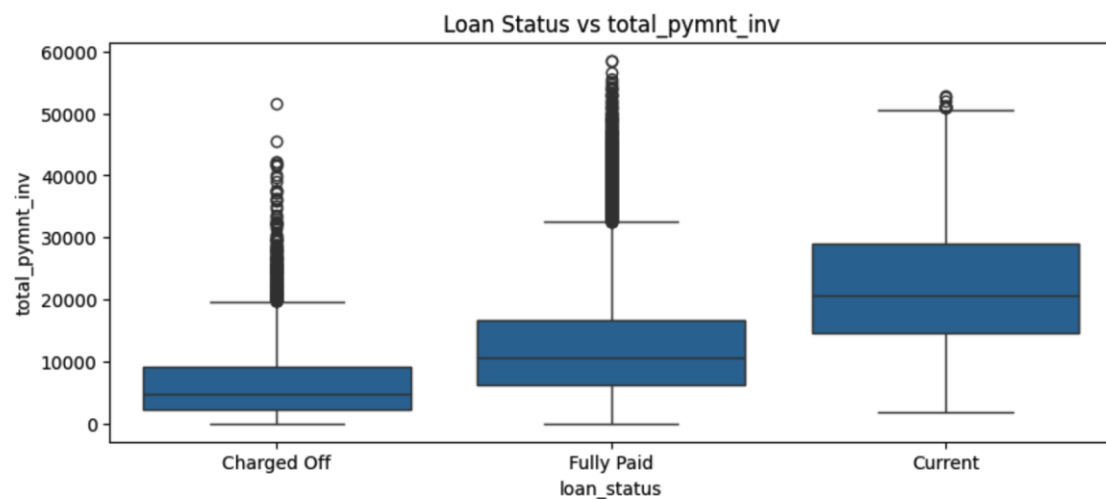
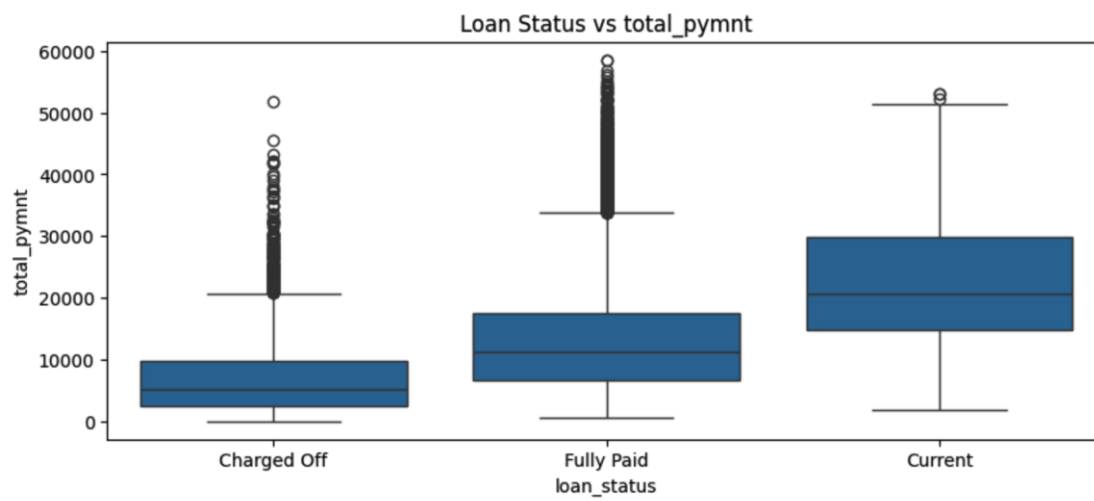
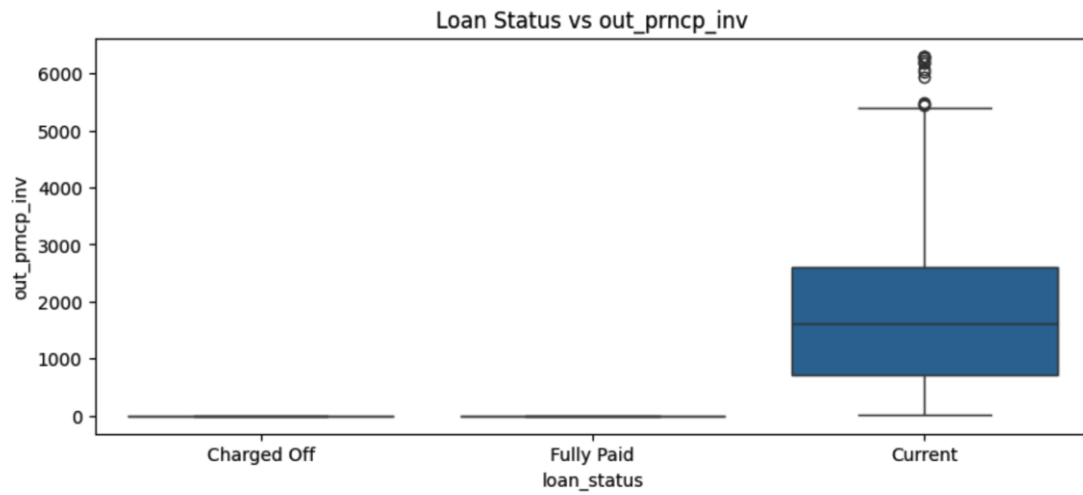


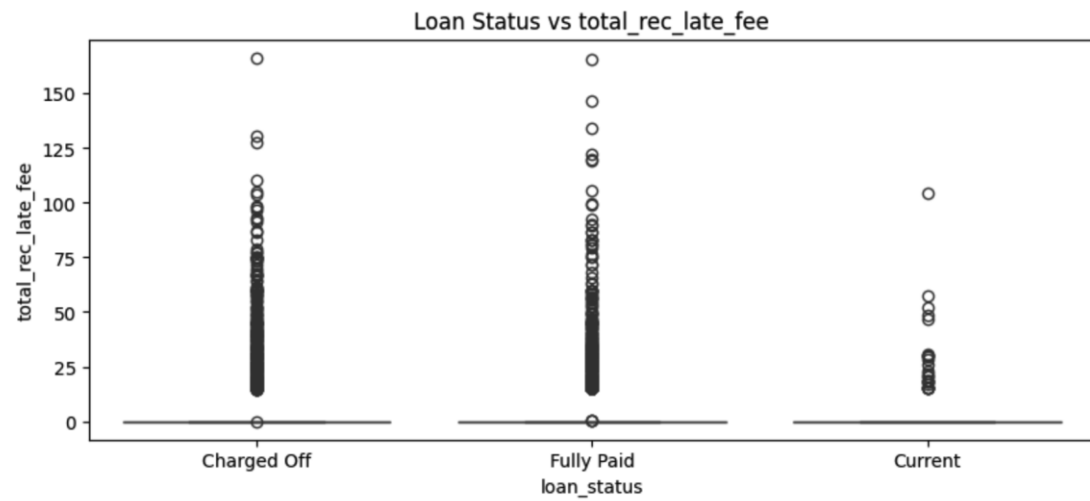
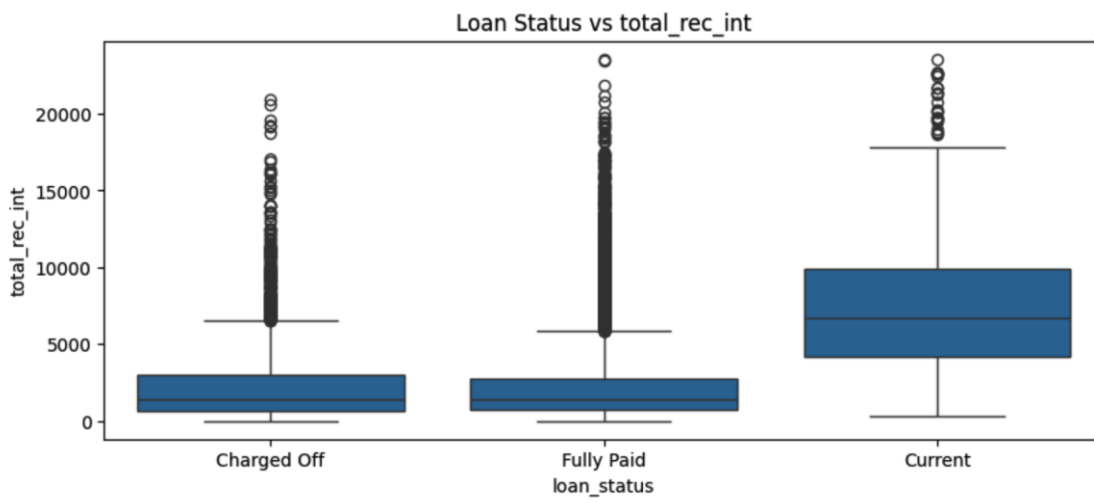
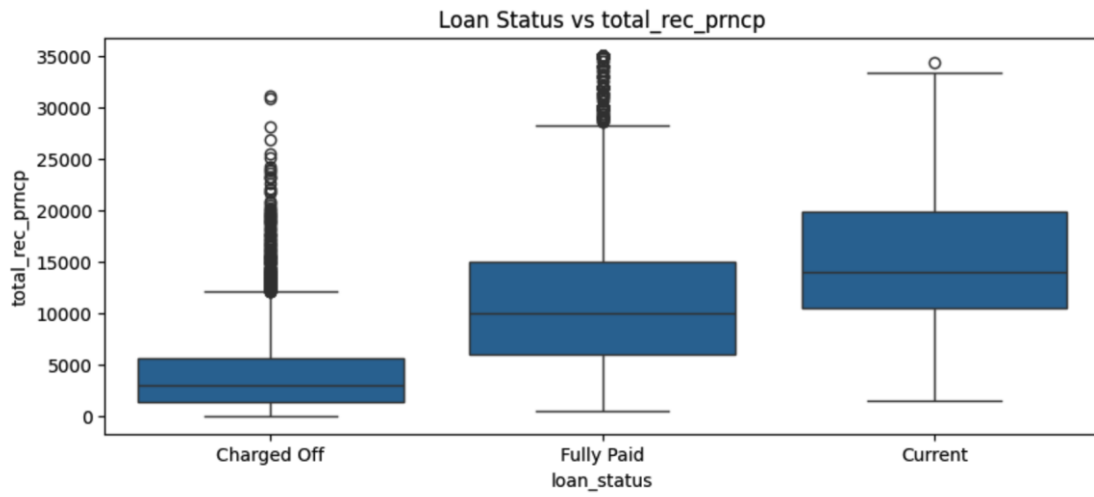


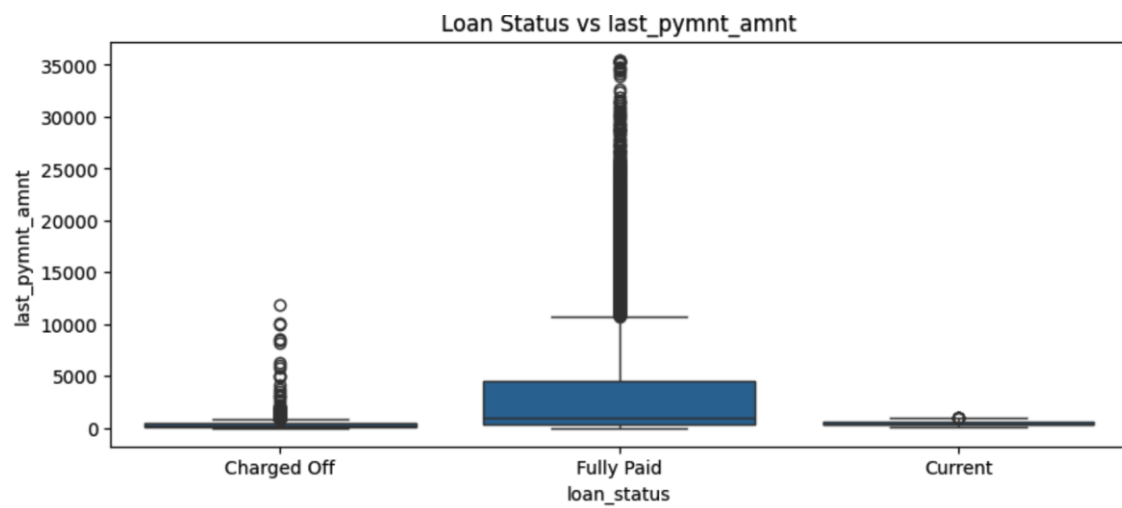
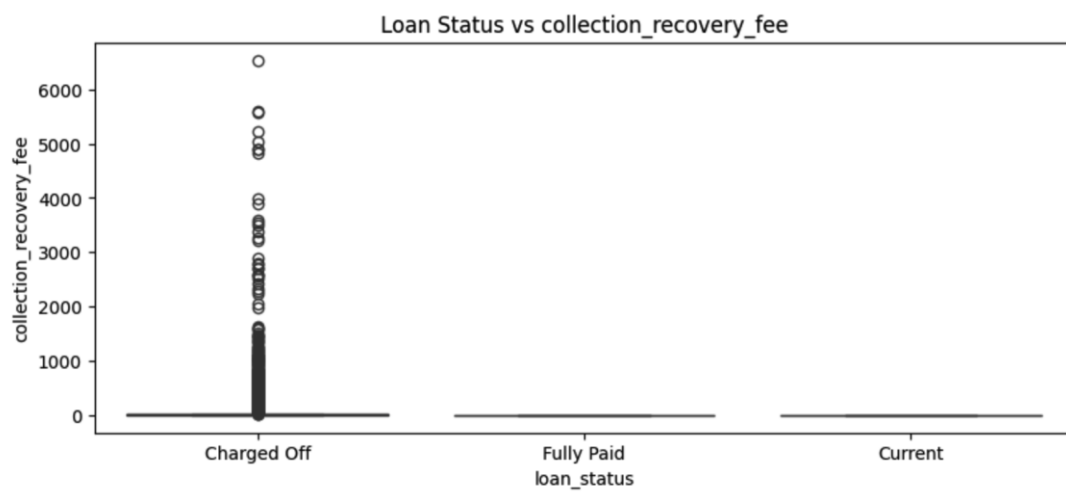
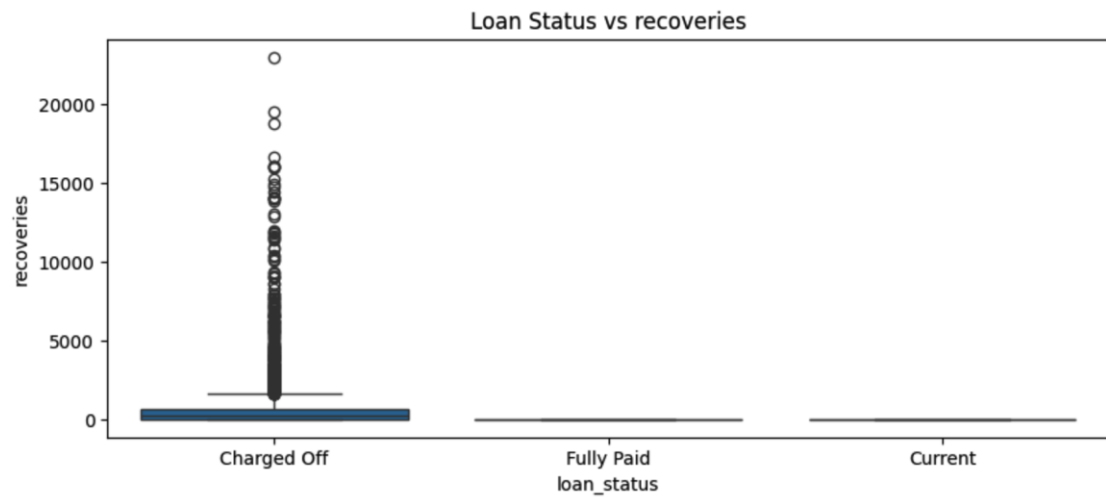


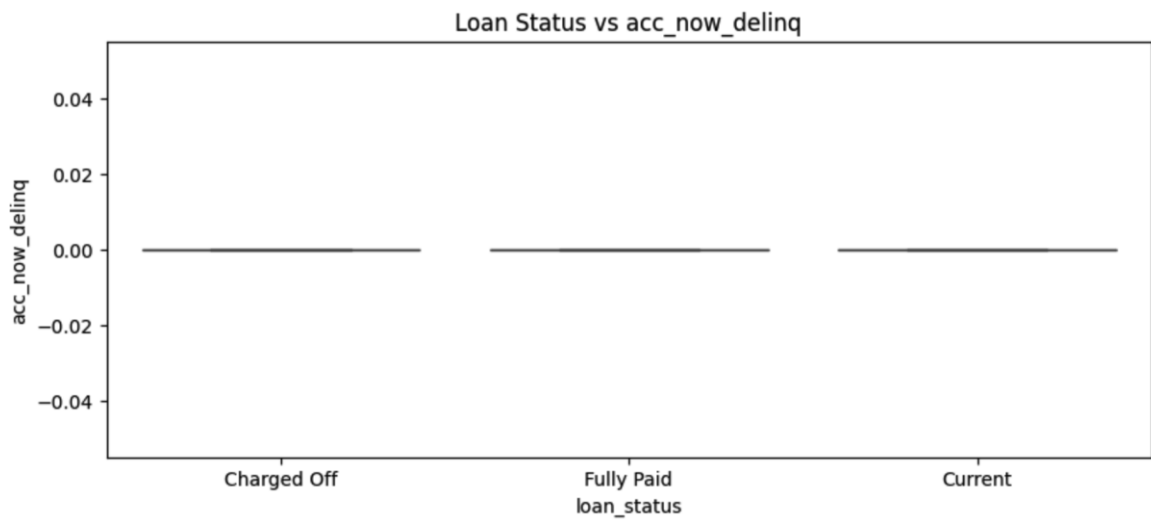
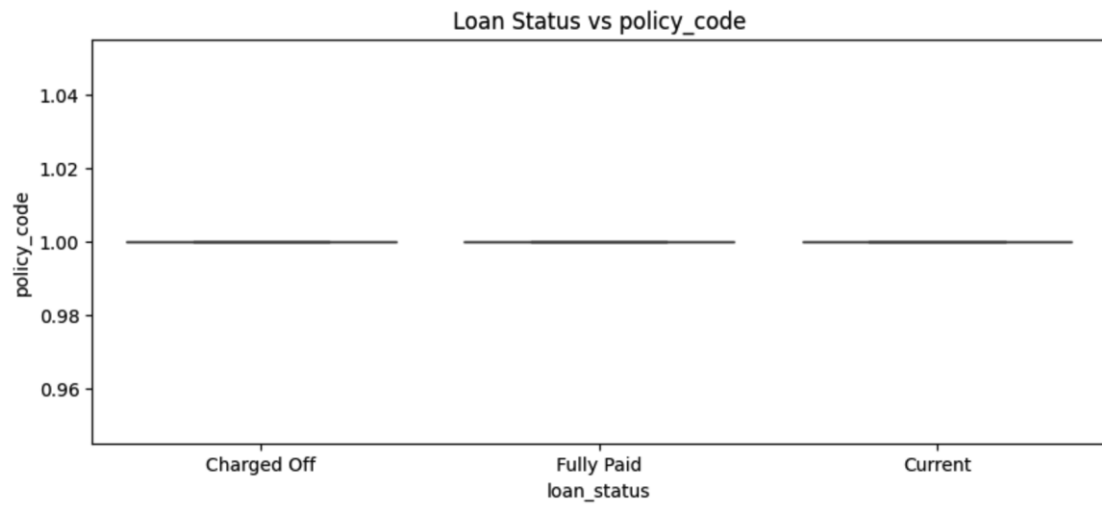
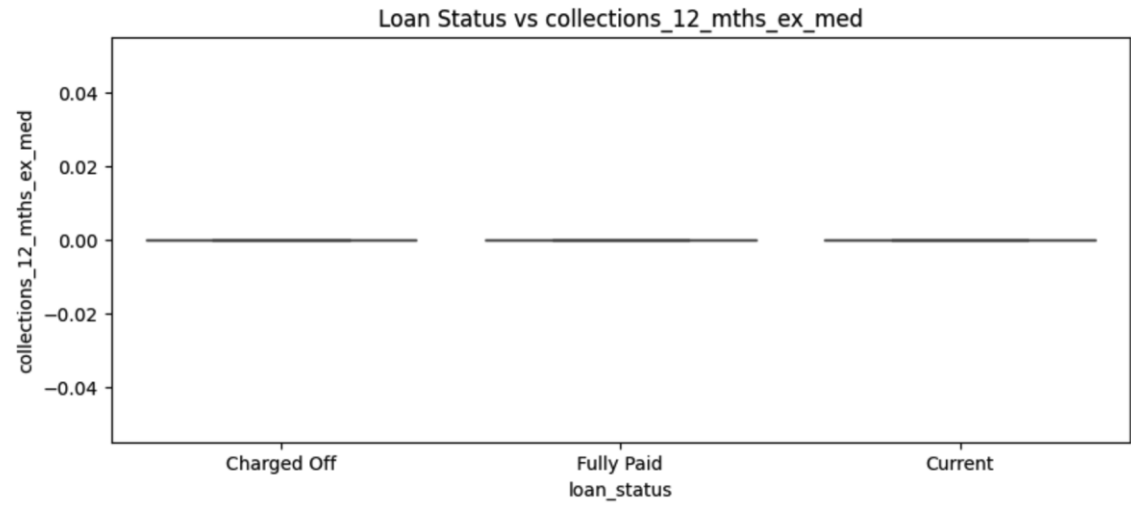


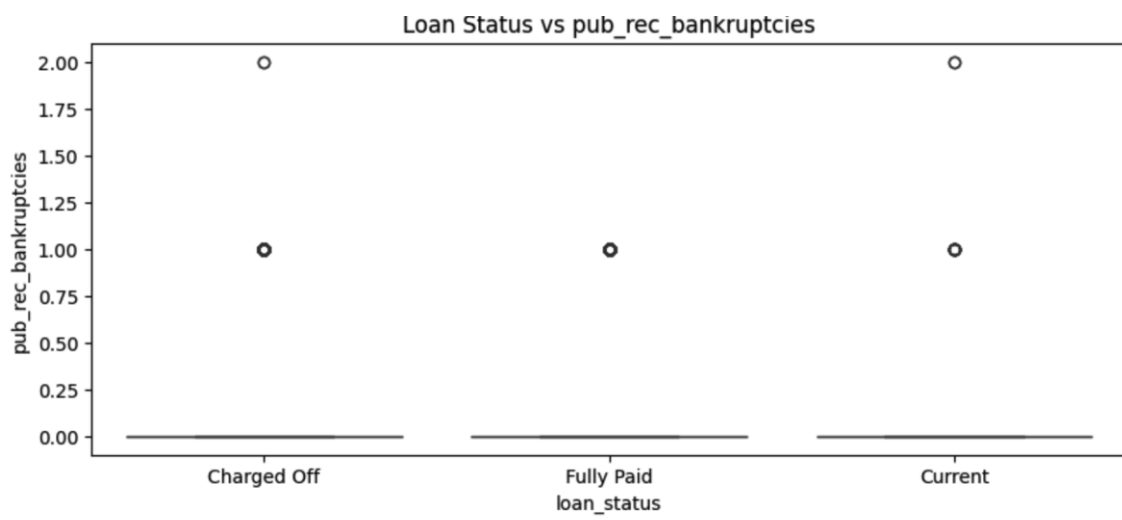
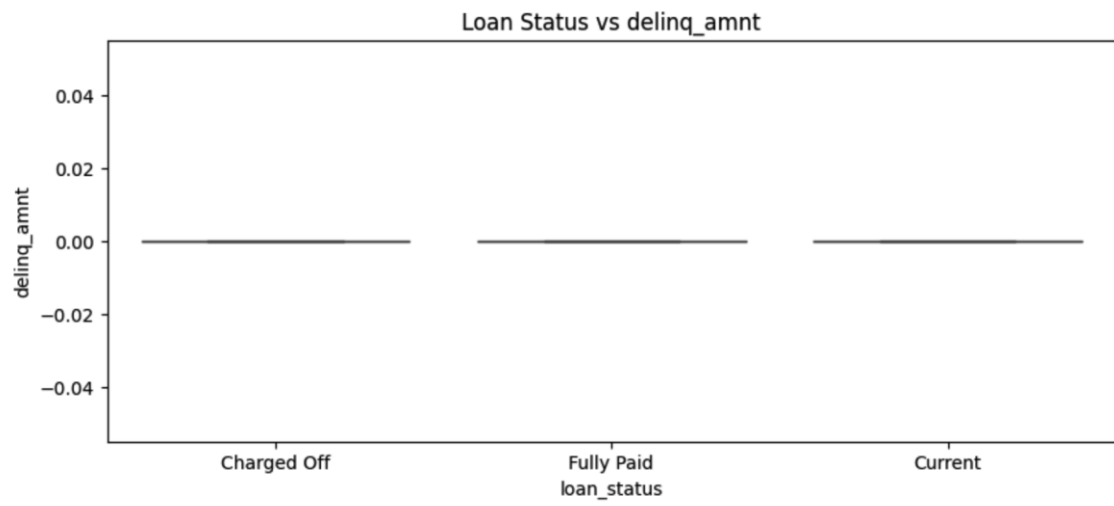
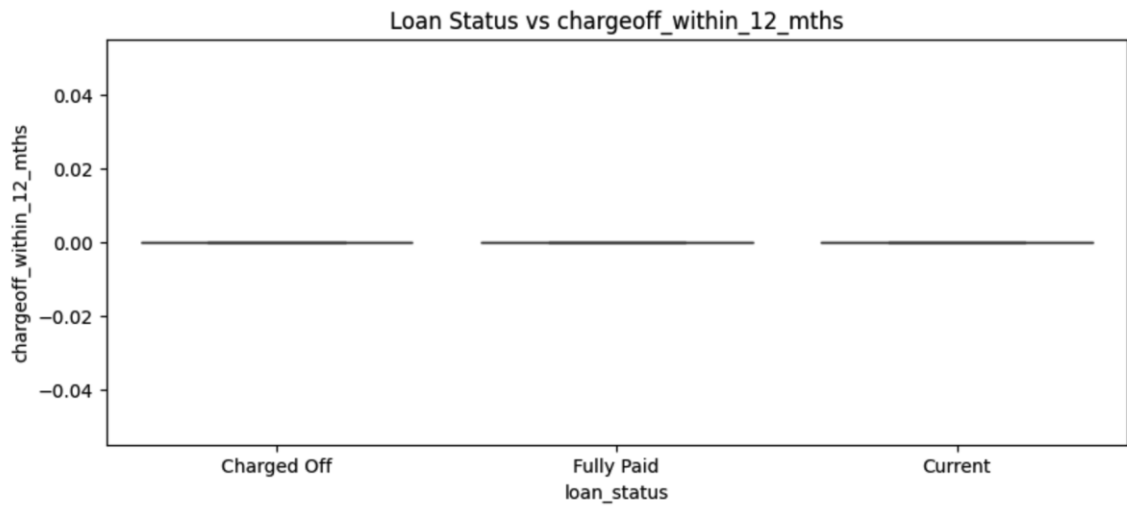


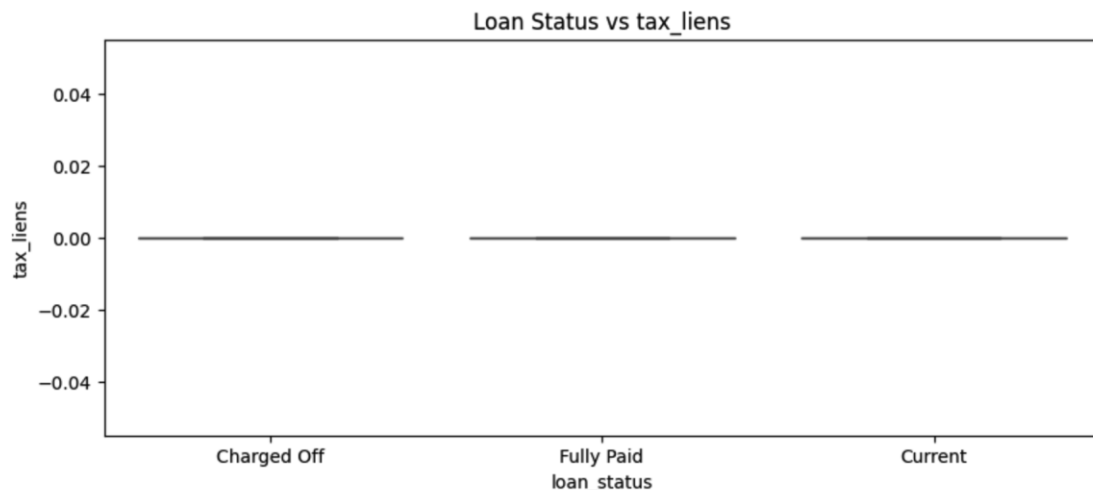












- **Metrics Created:**
 - **Debt-to-Income Ratio:** A crucial metric to assess an applicant's financial health.
 - **Credit Utilization Rate:** Indicates how much credit an applicant is using relative to their available credit.
- **Justification:** These metrics provide additional insights into an applicant's ability to repay the loan. High debt-to-income ratios and credit utilization rates are strong indicators of default risk.

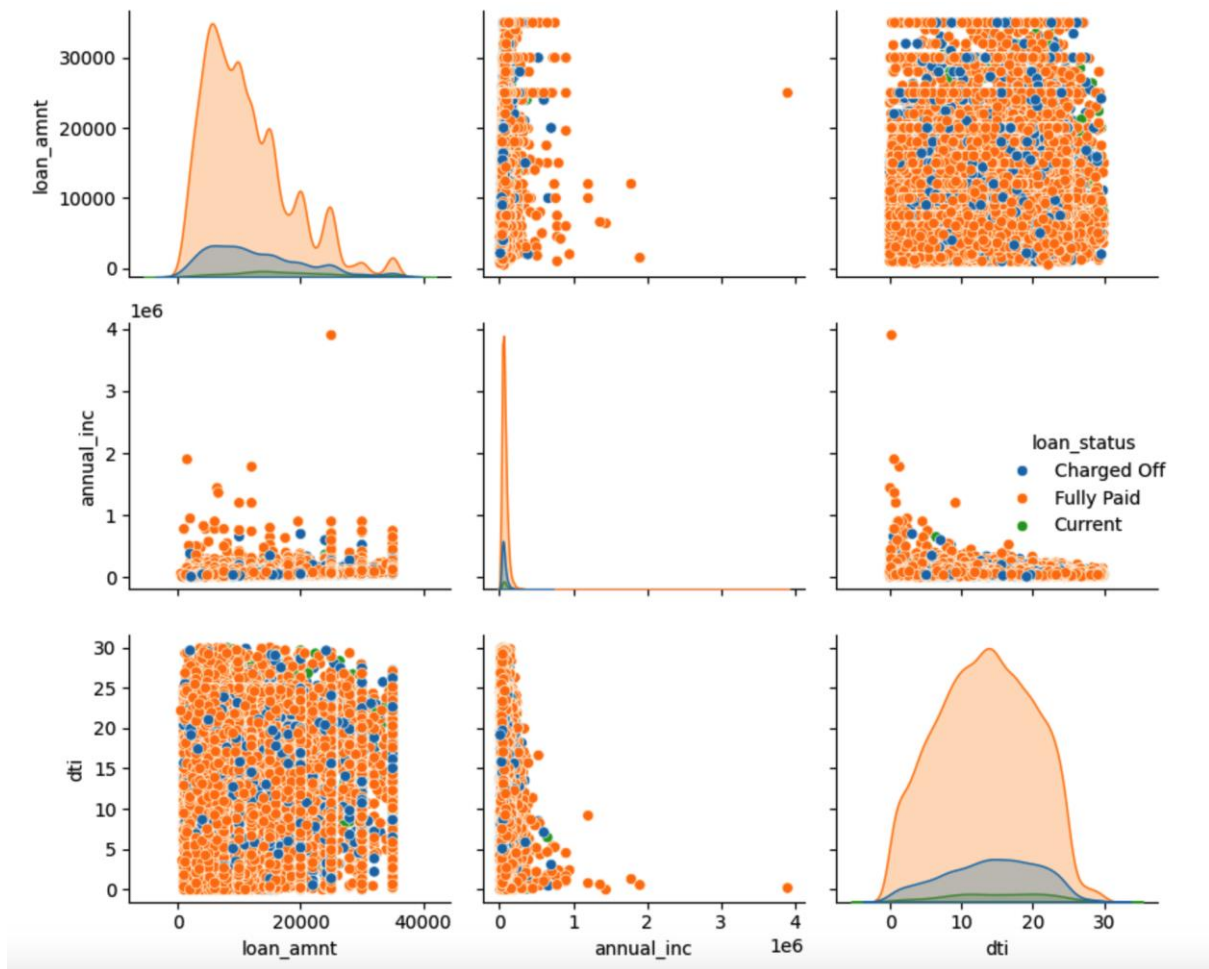
Important Driver Variables

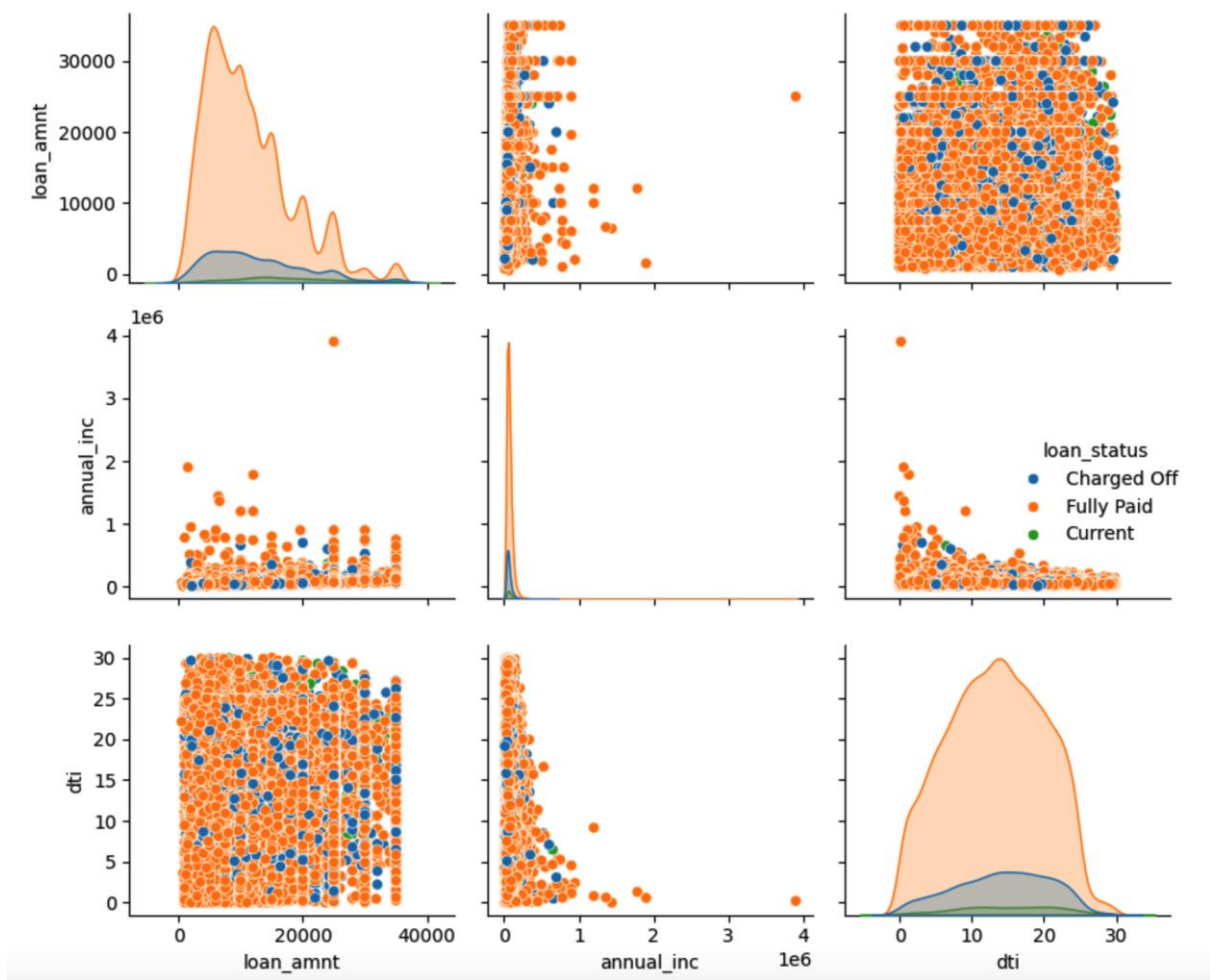
- **Top Predictors:**
 - **Loan Amount:** Larger loan amounts are associated with higher default rates.
 - **Interest Rate:** Higher interest rates correlate with increased default risk.
 - **Loan Term:** Longer loan terms may indicate higher risk.
 - **Grade:** Loan grade assigned by the company is a significant predictor.
 - **Annual Income:** Higher incomes are generally associated with lower default rates.
- **Impact on Default:** These variables significantly influence the likelihood of default. Understanding these factors helps in making informed lending decisions.

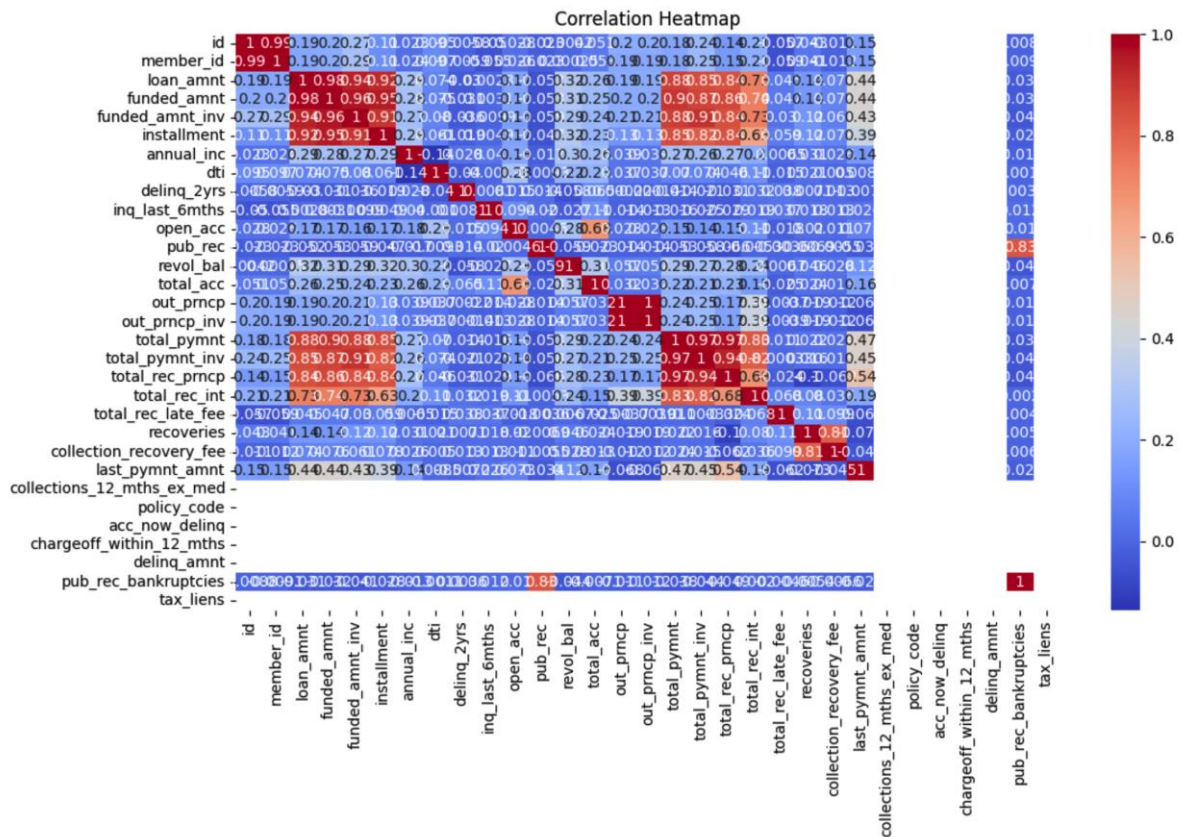
Visualizations

- **Graphs and Charts:**
 - **Histogram of Loan Amounts:** Shows the distribution of loan amounts.

- **Scatter Plot of Interest Rate vs. Loan Amount:** Highlights the relationship between these two variables.
 - **Box Plot of Annual Income by Default Status:** Compares income levels of defaulters and non-defaulters.
- **Interpretation:** Visualizations reveal key patterns and relationships. For example, the scatter plot shows that higher loan amounts often come with higher interest rates, and the box plot indicates that lower income levels are more common among defaulters.







Recommendations

• Actions to Take:

- **Deny Loans to High-Risk Applicants:** Use the identified driver variables to assess risk and deny loans to applicants with high default risk.
- **Adjust Interest Rates:** Offer higher interest rates to riskier applicants to compensate for potential losses.
- **Implement Stricter Loan Terms:** Consider shorter loan terms for high-risk applicants.

• **Implementation:** Integrate these insights into the company's loan approval process. Develop risk assessment models that incorporate the key driver variables identified in this analysis.

Conclusion

- **Summary:** This analysis identified key factors that influence loan defaults, including loan amount, interest rate, loan term, grade, and annual income. These insights can help the company make informed lending decisions and reduce credit risk.

Contact Information

- **Name – Pushkar Mishra**
- **Email – p2k3m_2002@yahoo.com**