Lending Club Case Study

STUDY GROUP:

RAMAGOVINDARAJA THIRUVADI

SHASHWATH S

Problem Statement

Our task is to analyse a dataset of historical loan information from “*consumer finance company*” to find insights into features that may influence loan default rates. Our goal was to identify feature patterns and correlations that could help in assessing loan risk and making informed lending decisions.

Analysis Approach

Started with data cleaning to ensure the quality and integrity of the dataset for analysis. We removed columns with 100% null values and corrected data types for key columns. We then conducted a univariate analysis to understand the distribution of individual variables, followed by bivariate analyses to explore relationships between two variables, particularly in relation to loan default.

Univariate Analysis

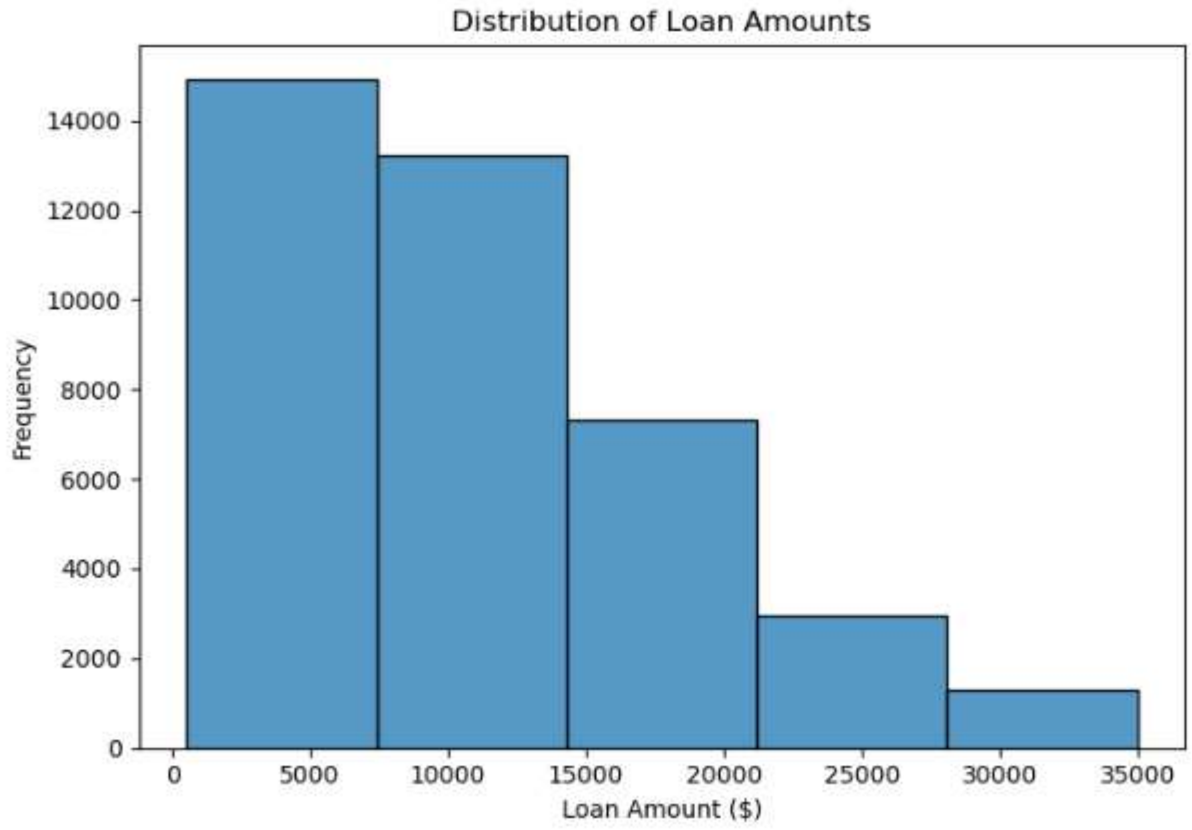
Loan Amounts: Showed a wide range of requested amounts, indicating diverse borrowing needs. Most loans disbursed are at lower end of loan amounts requested.

Interest Rates: Shows a broad range of rates, suggesting varying rates based on qualified applicants.

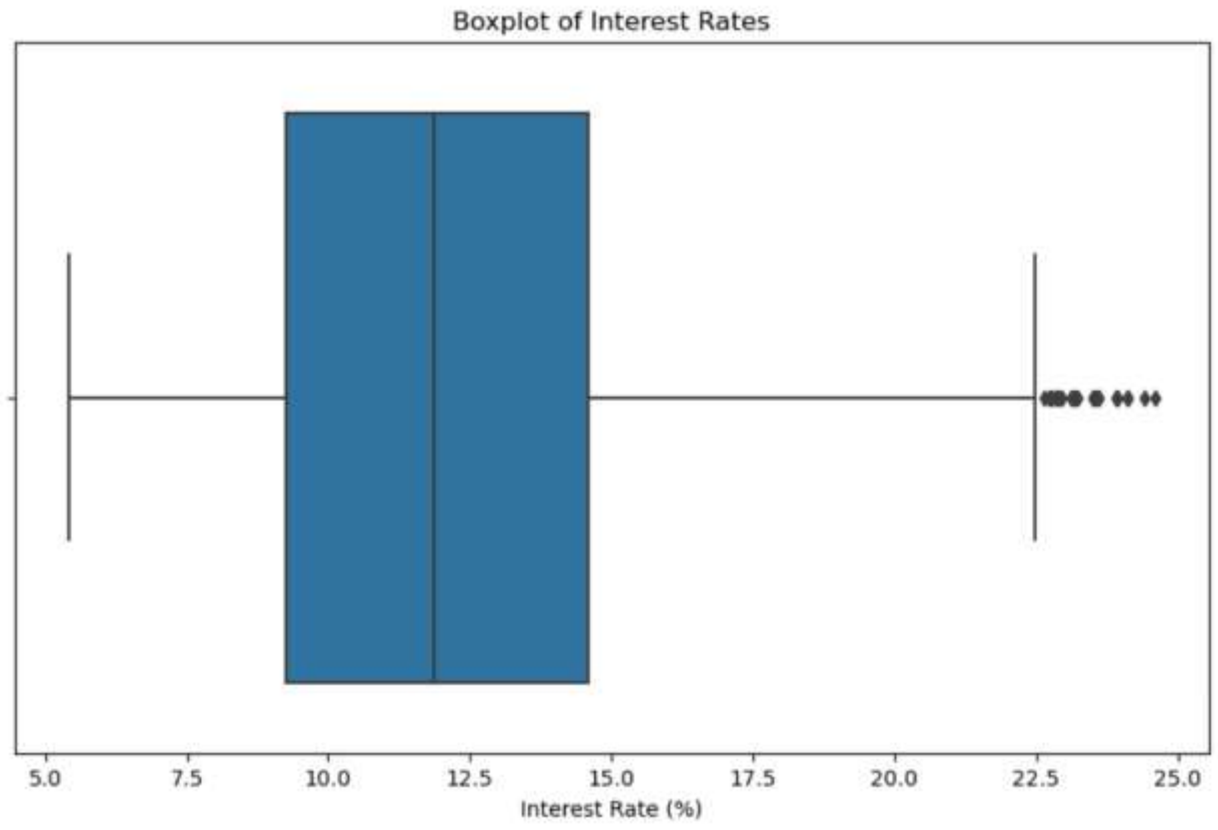
Employment Length: Indicated that a significant number of applicants had 10+ years of employment. i.e., with stable employment

Debt-to-Income Ratio: Shows a right-skewed distribution, with most applicants having a DTI below 30%, but some displayed high debts.

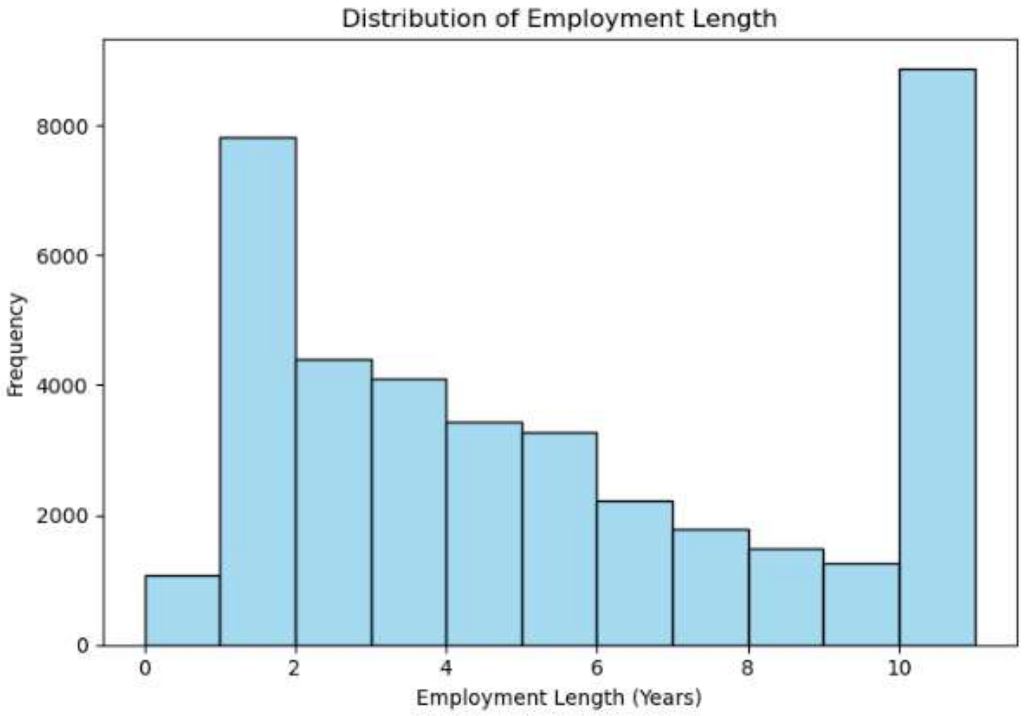
Distribution of Loan Amounts

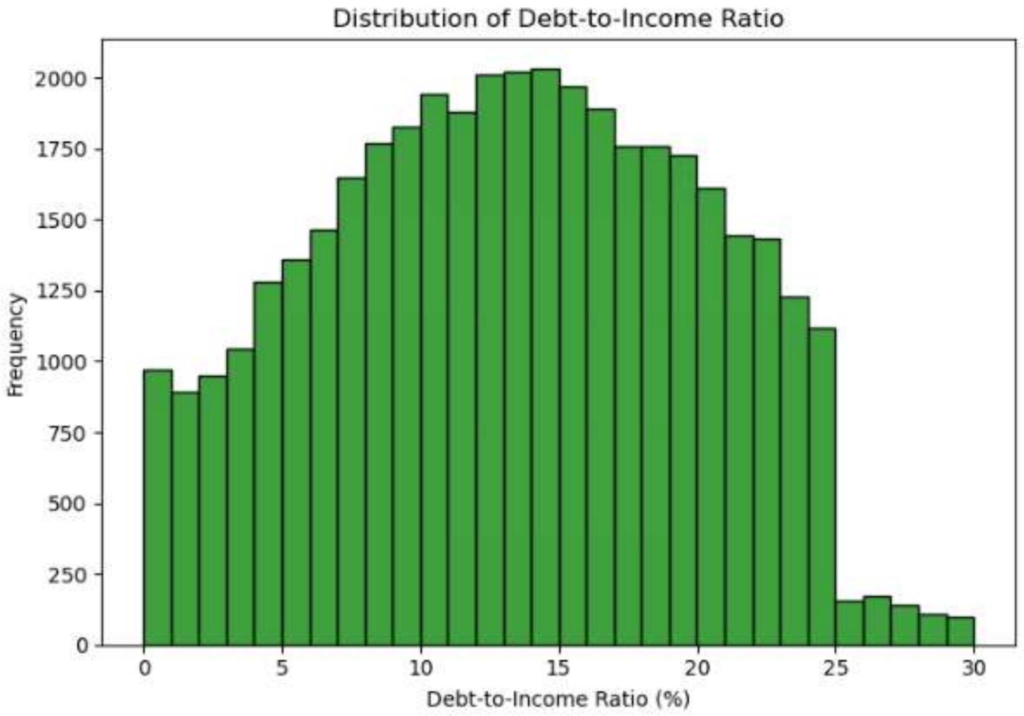


Distribution of Interest Rates



Dist. Of Employment Length & DTI



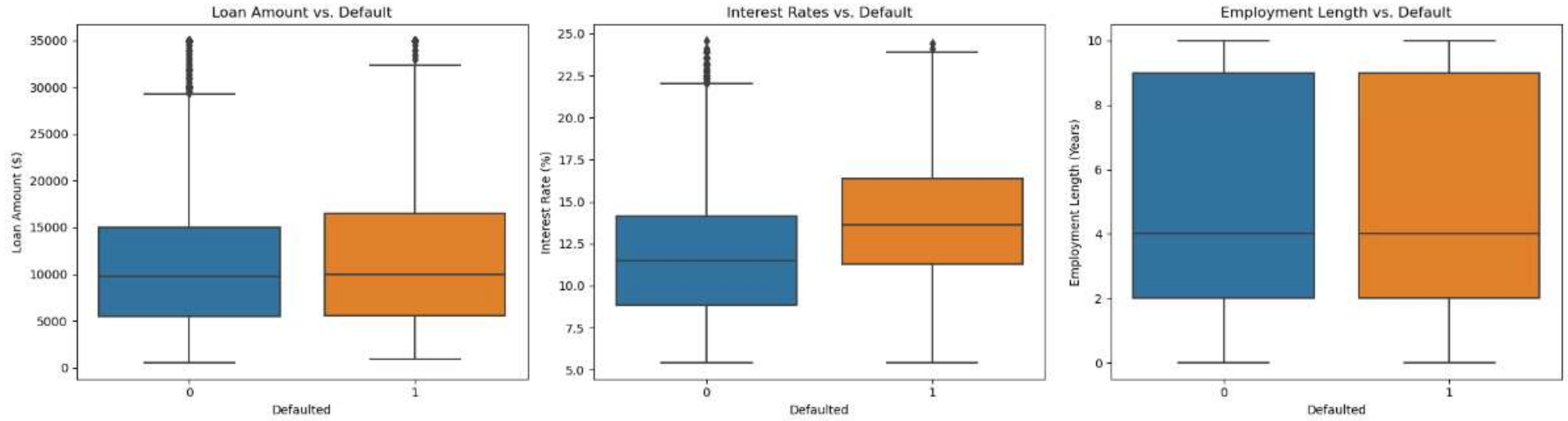


Bivariate Analysis

Loan Amount vs. Default: Slightly higher loan amounts were associated with defaults, implying higher amounts may be riskier. i.e., the median amount of defaults is slightly higher when loam amounts are higher.

Interest Rate vs. Default: Higher interest rates corresponded with defaults, potentially reflecting the higher risk priced into these rates.

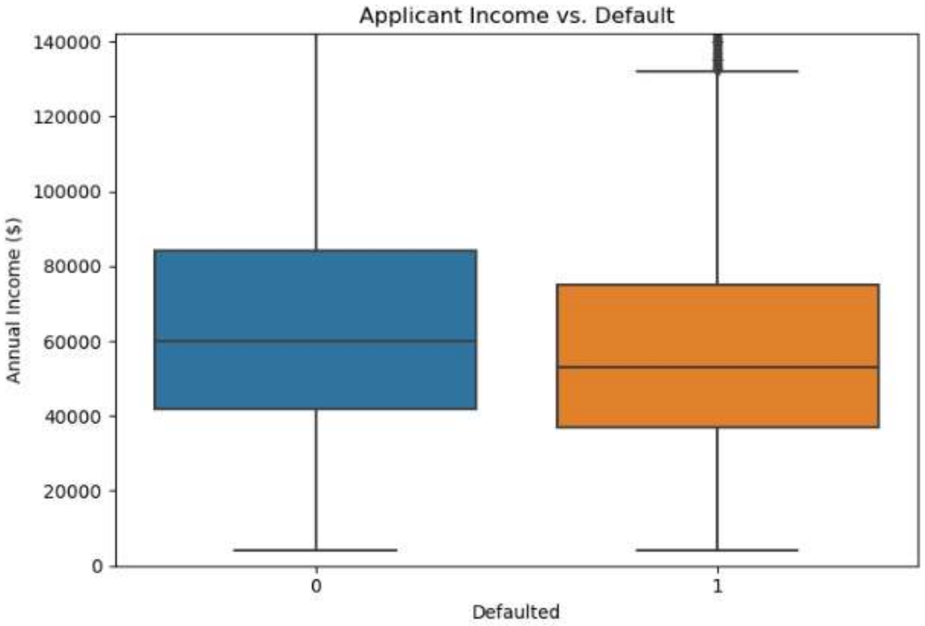
Employment Length vs. Default: No significant difference was found, suggesting that employment length alone may not be a strong predictor of default.

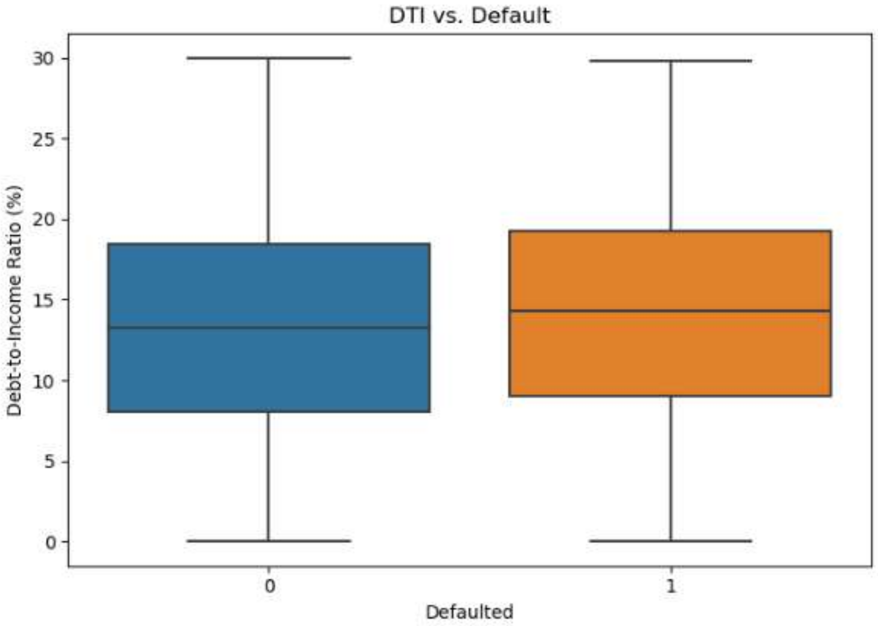


Bivariate Analysis Contd.

Income vs. Default: Lower incomes were slightly more associated with defaults, hinting at financial strain as a possible factor in defaults.

Loan Term vs. Default: Longer-term loans (60 months) showed a notably higher default rate than shorter-term loans (36 months), suggesting a correlation between term length and default risk.

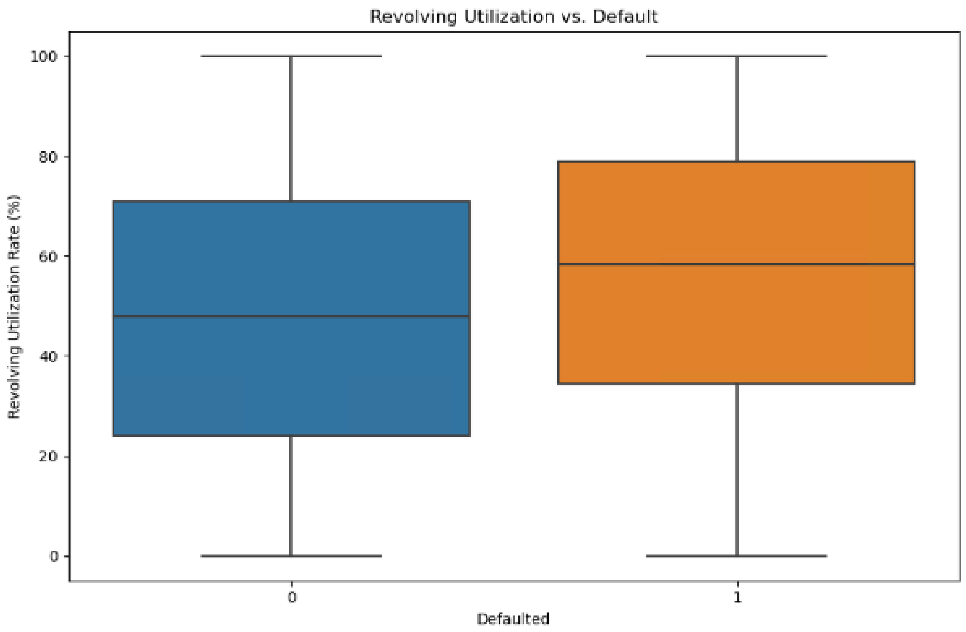


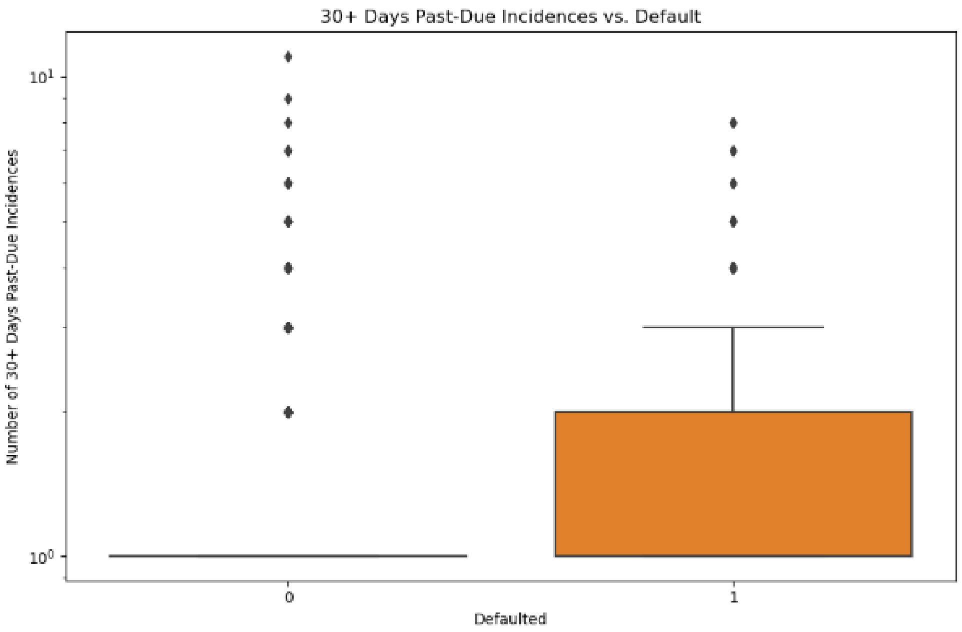


Bivariate Analysis Contd.

Credit Line Utilization vs. Default: Higher utilization rates were observed in defaulted loans, suggesting that borrowers who are using more of their available credit may be at a higher risk of default.

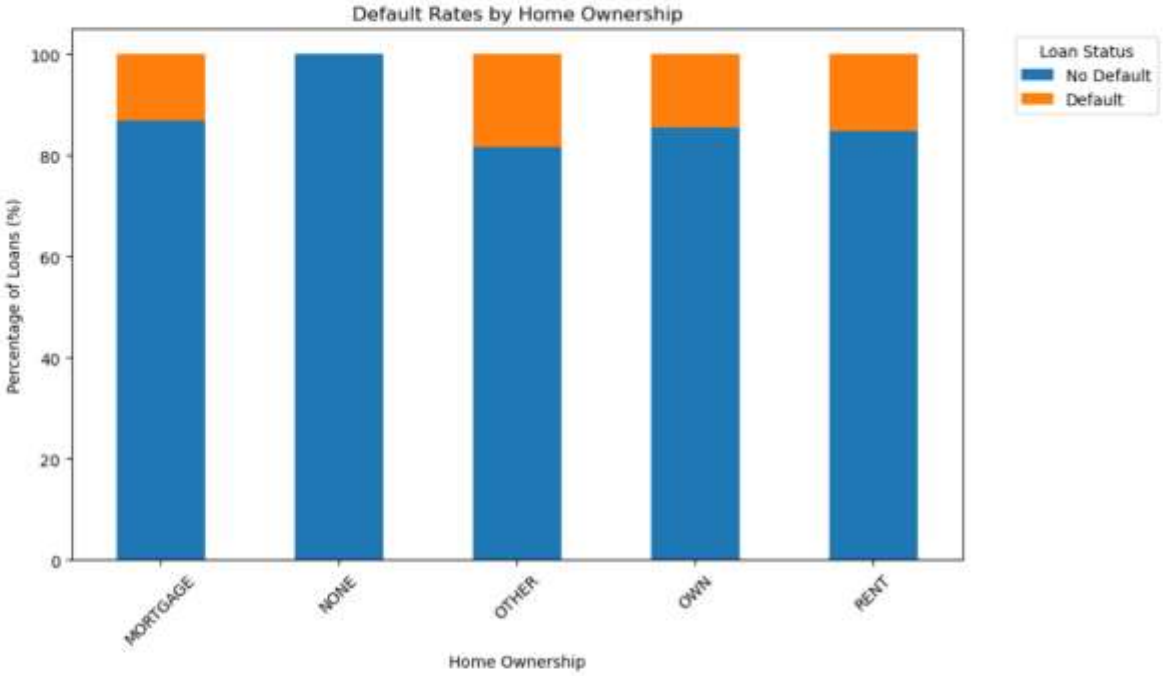
Past-Due Incidences vs. Default: An increased number of past-due events were seen in defaulted loans, enforcing the idea that a history of delinquency is a risk factor for default.





Bivariate Analysis Contd.

Home Ownership vs. Default: Renters and those with 'other' home ownership statuses had slightly higher default rates, indicating a possible link between property ownership and loan repayment reliability.



Conclusion

Based on the analysis, it suggests that loan default risk is not dependent on a single factor. Larger loan amounts, higher interest rates, longer loan terms, higher credit utilization and applicants with history of delinquency appears to be associated with higher default rates. These findings can help the "*consumer finance company*" in performing their risk assessment based on the suggested features when considering loan approvals.