CREDIT EDA CASE STUDY

BY: JAY HADIYAL

Agenda

- INTRODUCTION
- DATA OVERVIEW
- UNIVARIATE ANALYSIS
- BIVARIATE ANALYSIS
- CORRELATION ANALYSIS
- KEY INSIGHTS AND RECOMMENDATION
- CONCLUSION

INTRODUCTION

CHAPTER-1

PROBLEM STATEMENT

- The consumer finance company specializes in providing various types of loans to urban customers. When the company receives a loan application, it has to decide whether to approve or reject it based on the applicant's profile. Two **types of risks** are associated with the bank's decision:
- If the applicant is **likely to repay the loan**, then not approving the loan results in a **loss of business** to the company.
- If the applicant is **not likely to repay the loan**, i.e., they are likely to default, then approving the loan may lead to **financial loss** for the company.
- The aim is to **identify patterns** indicating that a person is likely to default, which may be used to deny the loan, reduce the loan amount, lend (to risky applicants) at a higher interest rate, etc.

ANALYSIS APPROACH

We conducted a comprehensive analysis using historical loan data to identify key patterns and risk factors associated with loan defaults. Our goal is to provide actionable insights that can enhance decision-making in loan approvals. Below is the flowchart of approach.

DATA OVERVIEW

CHAPTER-2

DATA LOADING

IMPORTING LIBRARIES:-

- **Pandas** – Used for Data Manipulation and Analysis
- **Numpy** – Used for Numerical Computational
- **Seaborn**- Used for Data Visualization

LOADING THE DATASET:-

- Using the ‘loan.csv’ file, a key dataset containing vital information about loan applications.
- Efficiently loading this dataset is the first crucial step in the data analysis.

Here’s how we approach it using the Pandas library:

```
loan_df=pd.read_csv('loan.CSV)
```

DATA CLEANING

HANDLING MISSING VALUES:-

- Detect and understand the extent of missing data in the dataset.
- Use ‘**isnull()**’ and ‘**sum()**’ to find and count the missing values

HANDLING MISSING VALUES:-

- Remove rows/columns with missing data using ‘**dropna()**’.
- Replace missing values using ‘**fillna()**’.

UNIVARIATE ANALYSIS

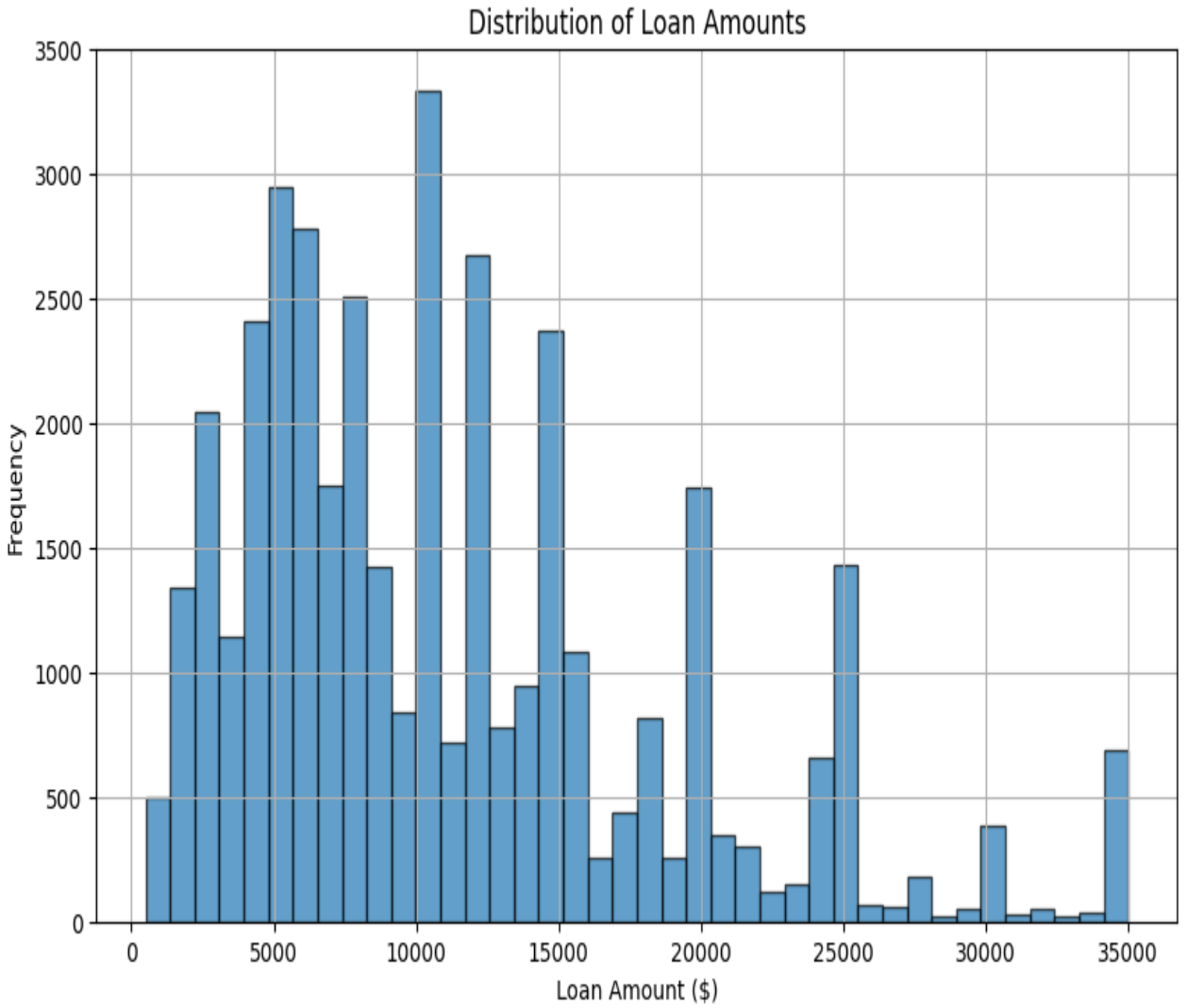
CHAPTER-3

LOAN AMOUNT

Checking the distribution of loan amounts to understand the range and frequency of loan requests.

Insights:-

- The average loan amount is around $11,219, with a significant spread indicated by the standard deviation of $7,456.
- Loans range from $500 to $35,000.

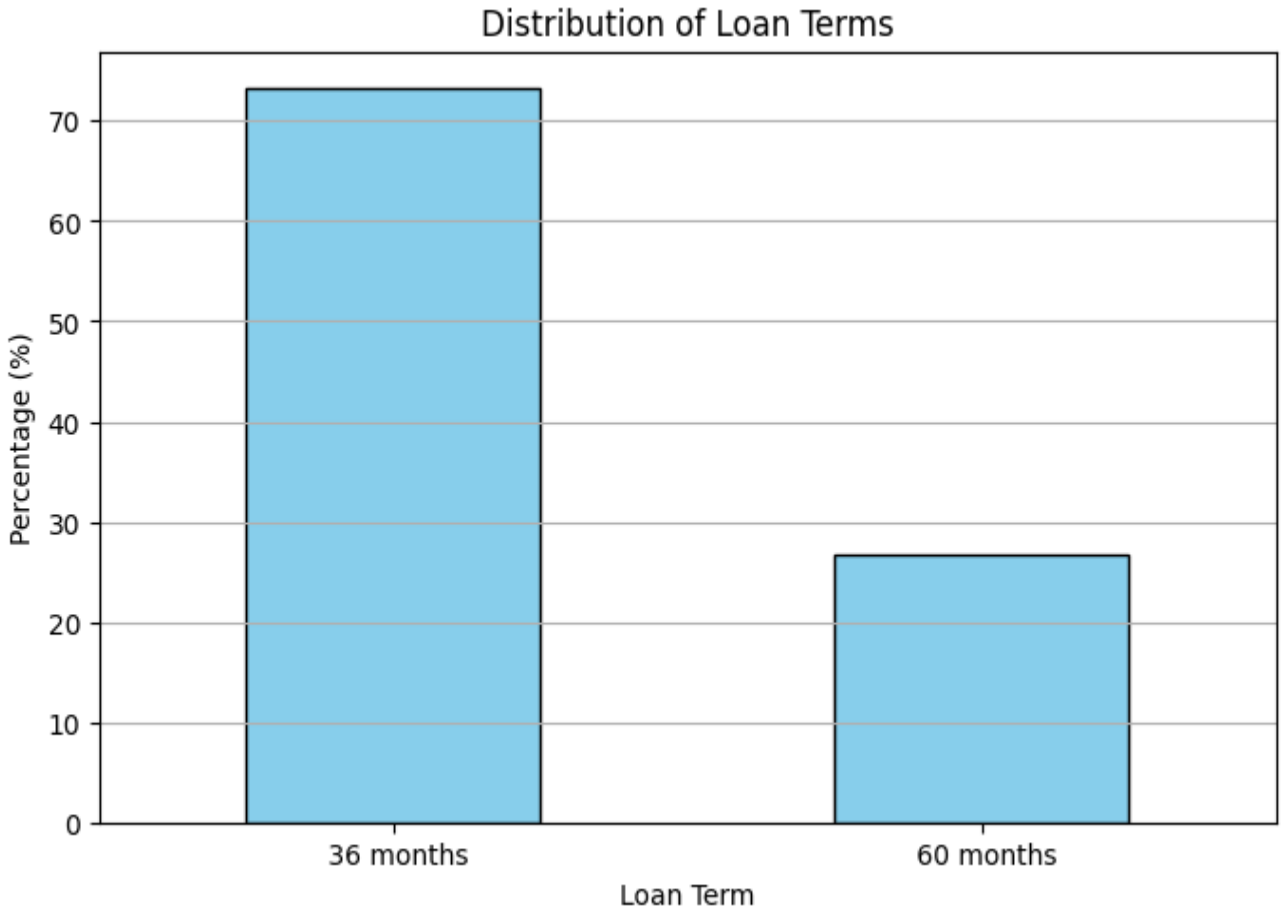


TERM

Explore the distribution of loan terms (e.g., 36 months vs. 60 months) to understand the most common loan duration.

Insights:-

- Most loans have a term of 36 months (approximately 73% of loans), while the remaining have a term of 60 months.

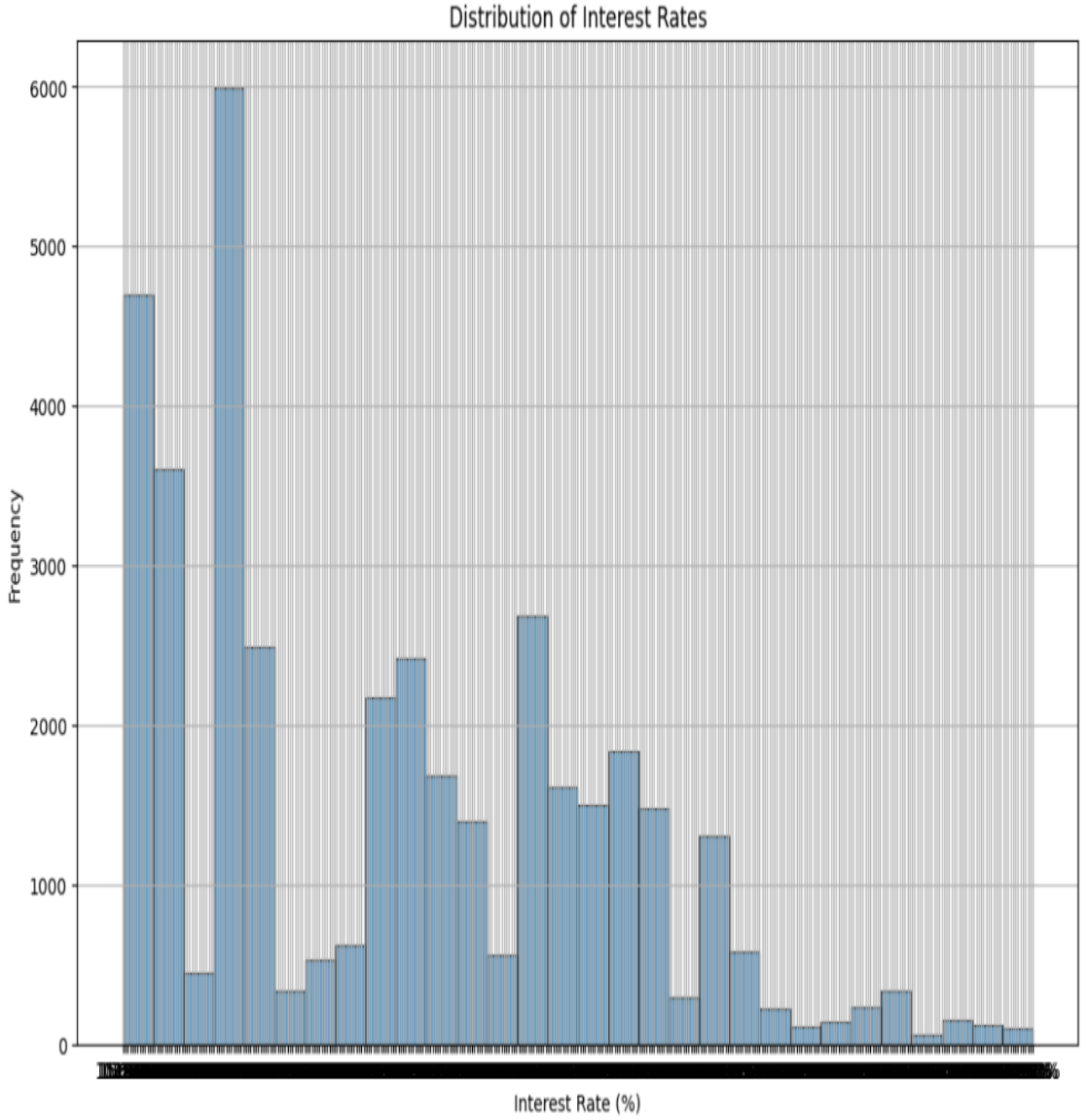


INTEREST RATE

Analyze the distribution of interest rates to identify common interest rate ranges.

Insights:-

- The average interest rate is approximately 12.02%, ranging from 5.42% to 24.59%.
- This indicates variability in interest rates offered to borrowers.

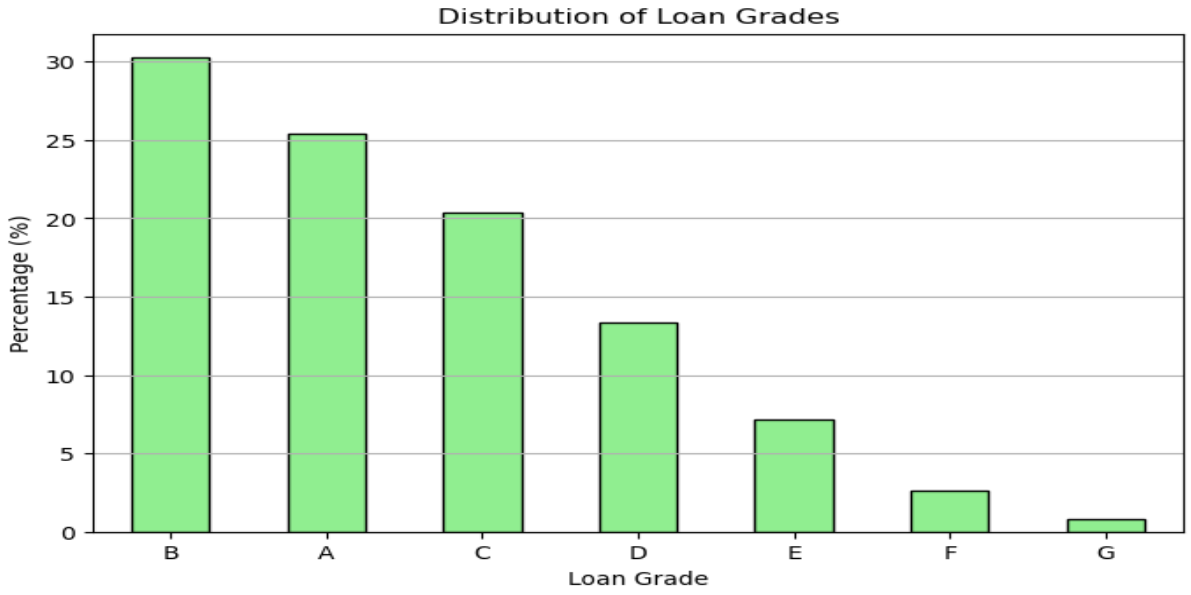


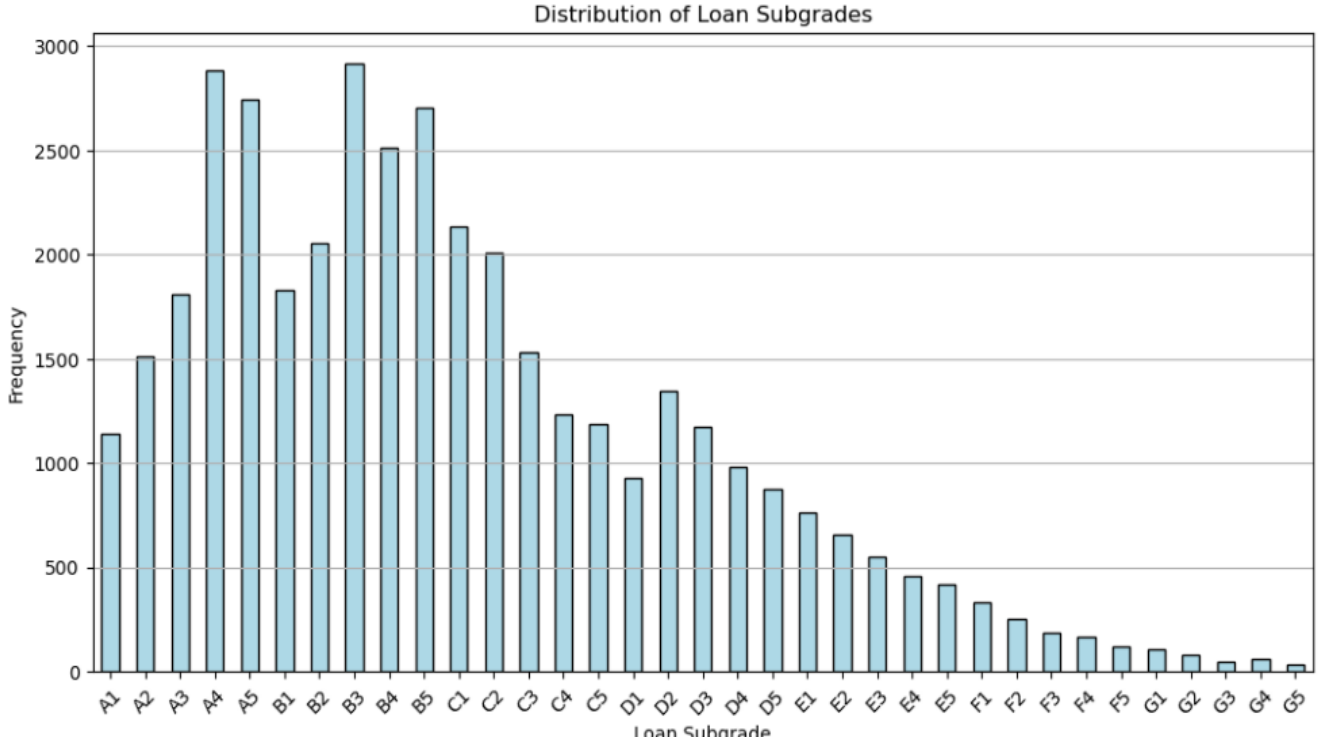
GRADE and SUB_GRADE

Plot the distribution of loan grades and subgrades assigned by LC.

Insights:-

- Grades range from A to G, with subgrades further categorizing the risk within each grade.
- Grade distributions show that higher grades (A and B) have more loans compared to lower grades (F and G).





EMP_LENGTH

Analyze the distribution of employment lengths to understand the tenure of borrowers.

Insights:-

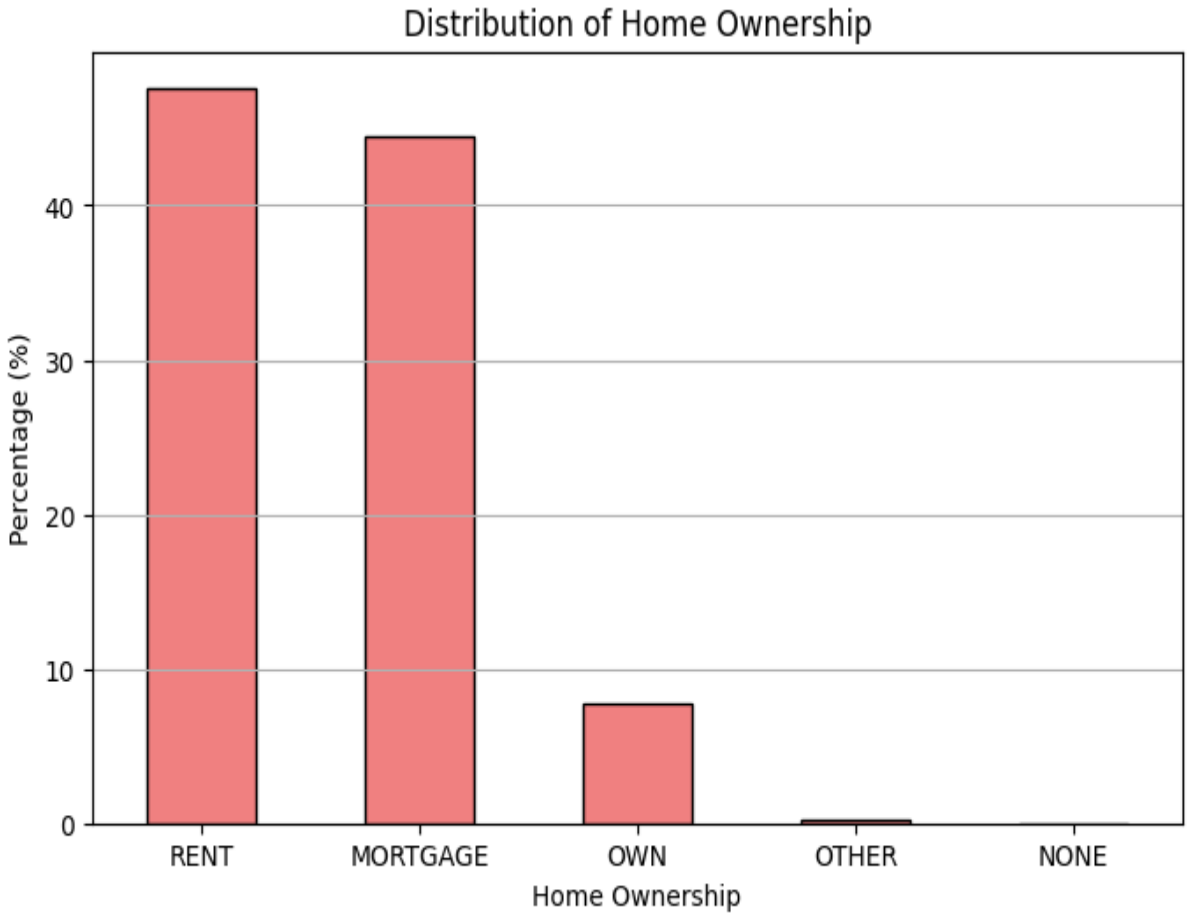
- Employment lengths vary, with a significant portion of borrowers having more than 2 years of employment.
- Borrowers with 10+ years of employment constitute the largest category.

HOME_OWNERSHIP

Explore the distribution of home ownership status among borrowers.

Insights:-

- Most borrowers rent or have a mortgage.
- A small percentage own homes outright or have other arrangements.

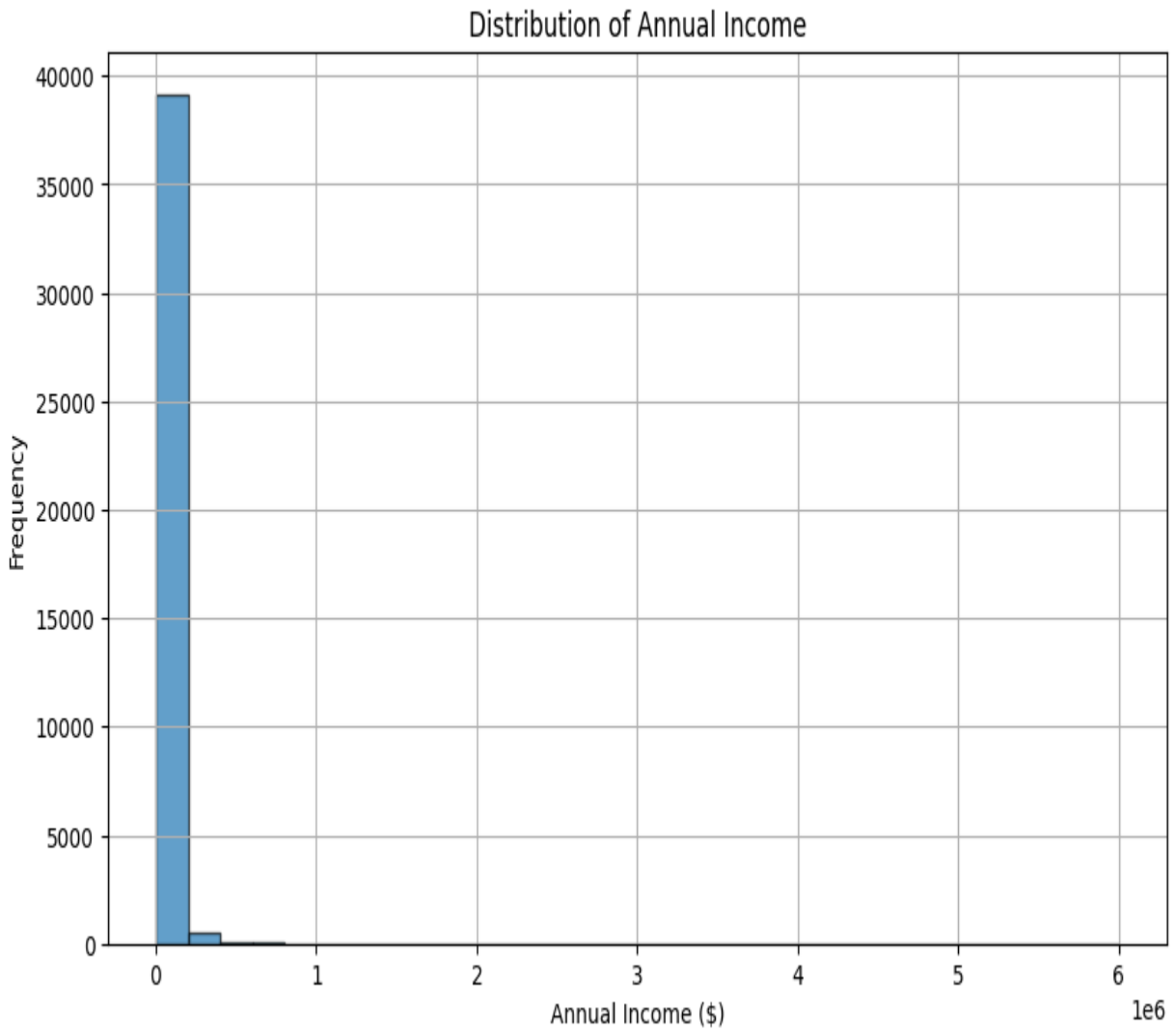


ANNUAL INCOME

Examine the distribution of borrowers' annual incomes to understand income levels.

Insights:-

- The average annual income is approximately $68,969, with considerable variability as indicated by the standard deviation of $63,793.
- Incomes range from $4,000 to $6,000,000.

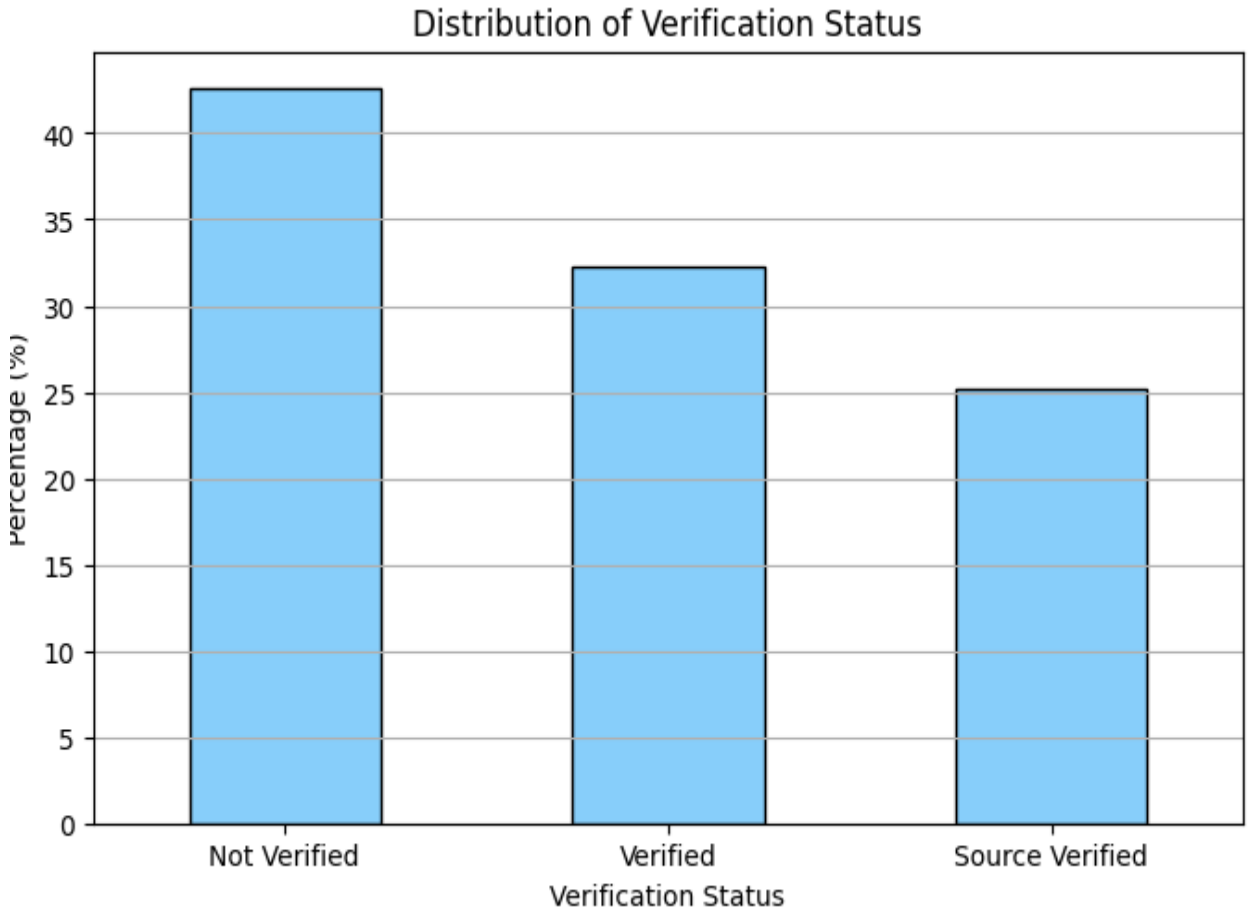


VERIFICATION_STATUS

Explore the distribution of income verification status (verified vs. not verified).

Insights:-

- Verification status varies among loans, with a significant portion being either not verified or verified.

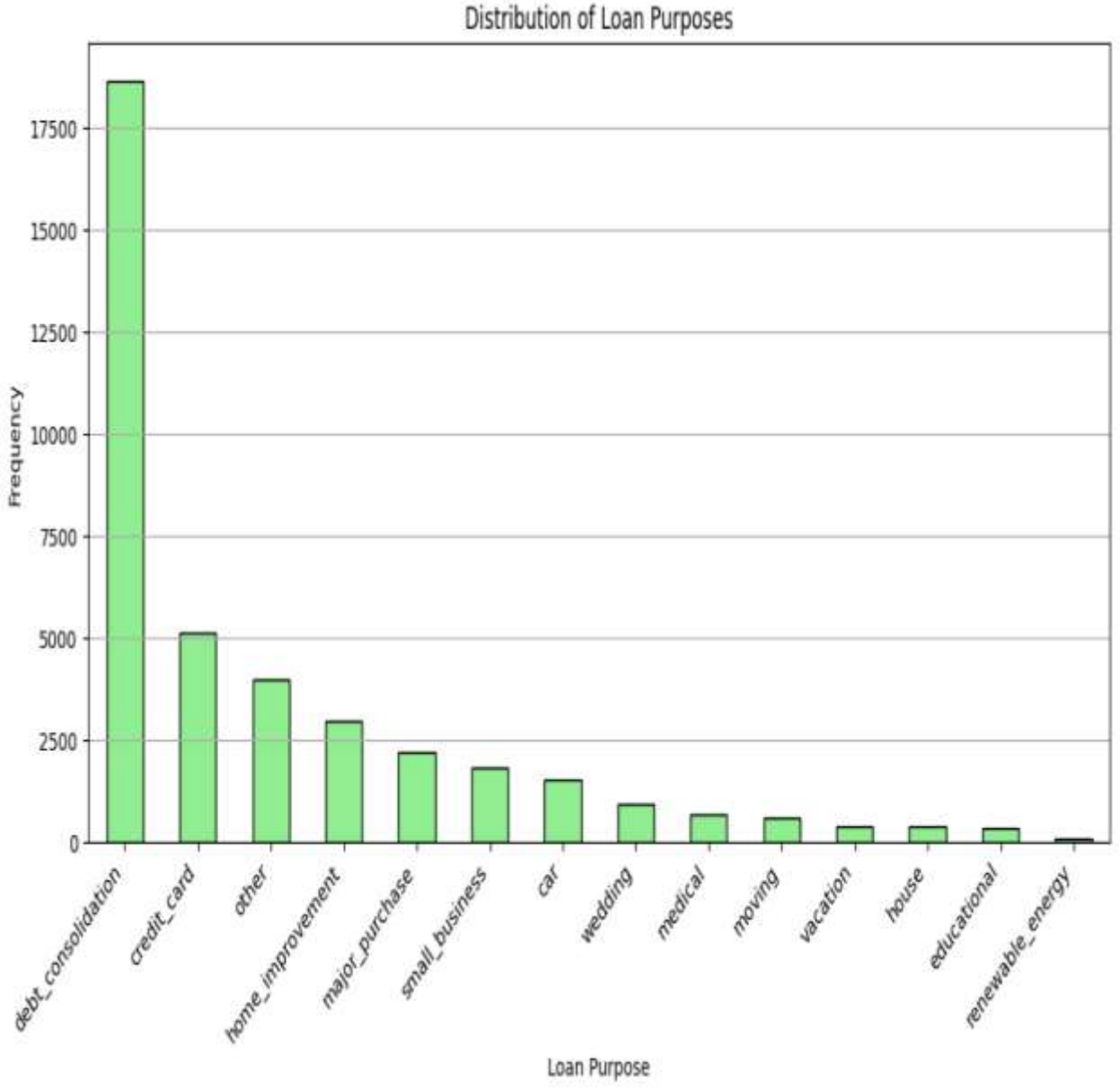


PURPOSE

Analyze the distribution of loan purposes (e.g., debt consolidation, credit card refinancing).

Insights:-

- Debt consolidation and credit card loans are the most common purposes.
- Other purposes such as home improvement, major purchases, and small business loans also appear frequently.

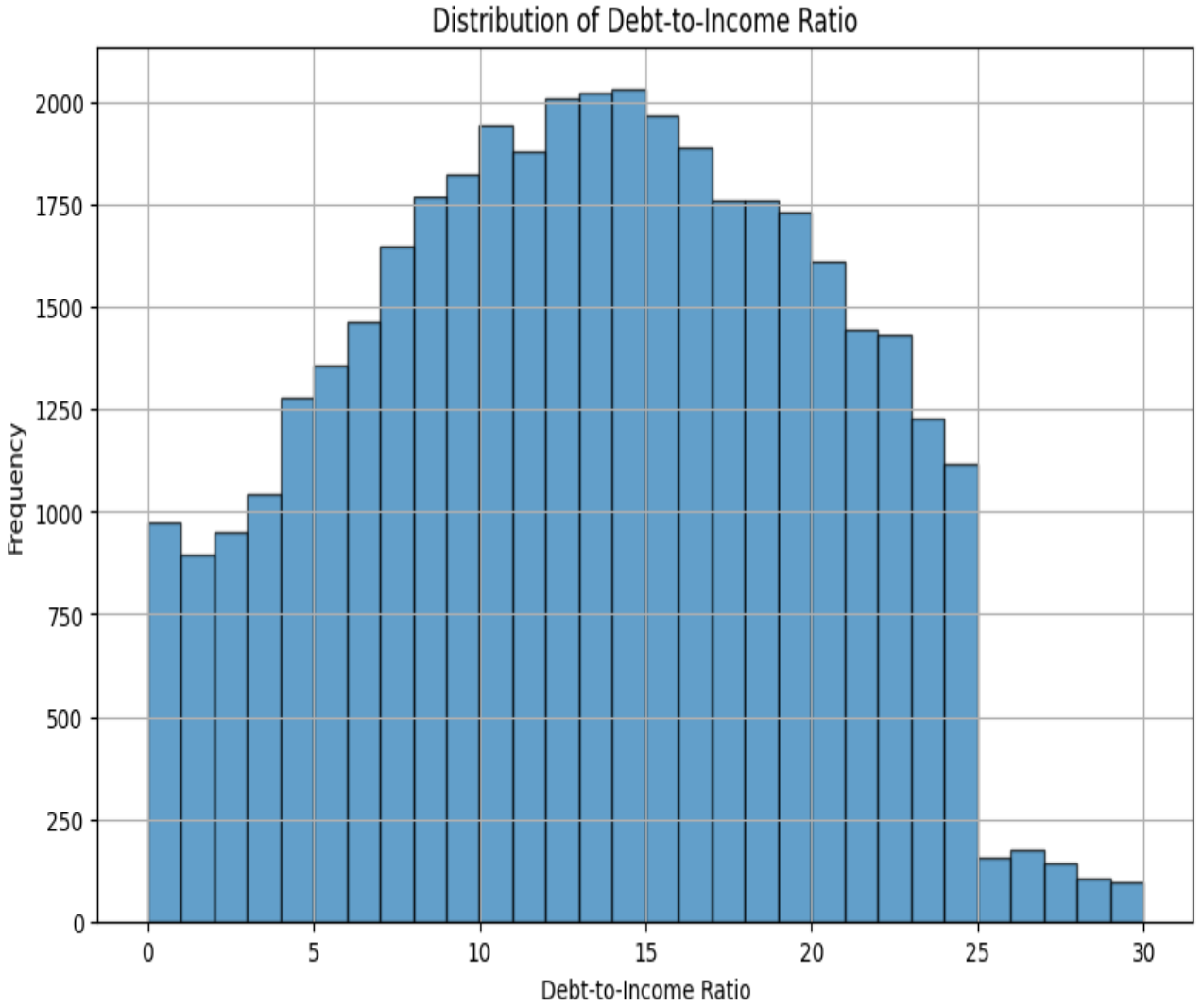


DTI RATIO

Examine the distribution of debt-to-income ratios.

Insights:-

- The average Debt-to-Income (DTI) ratio is approximately 13.32%, ranging from 0% to 29.99%.
- This metric helps assess borrowers' ability to manage additional debt based on existing income and obligations.



BIVARIATE ANALYSIS

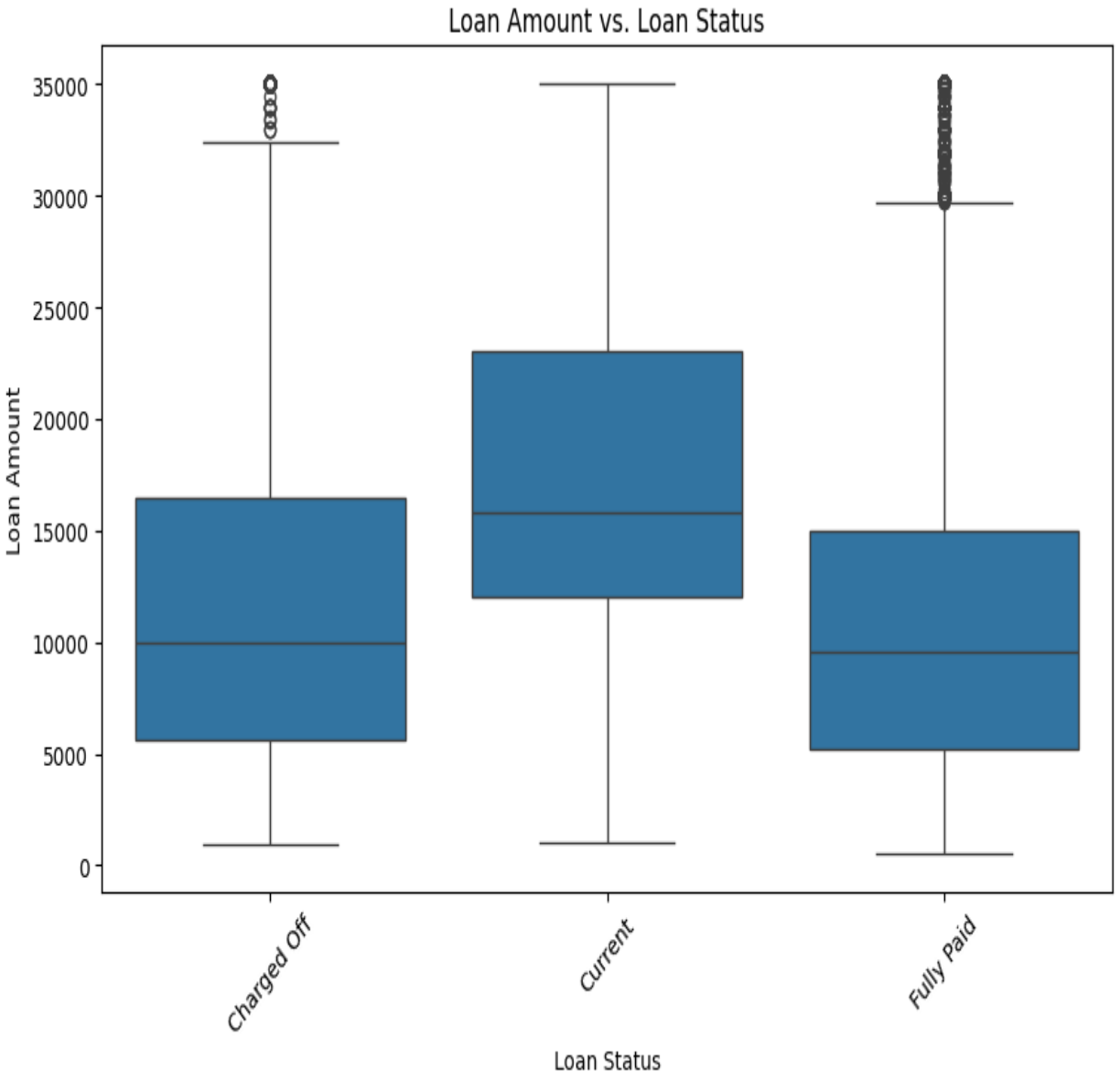
CHAPTER-3

LOAN AMOUNT vs. LOAN STATUS

Box plots to compare loan amounts for different loan statuses.

Insights:-

- Loans that are currently being paid (Current status) have higher average and median loan amounts compared to both charged-off and fully-paid loans. This suggests that larger loans might be more actively managed or have stricter repayment oversight.
- Charged-off loans tend to have higher loan amounts than fully paid loans, indicating that higher loan amounts might pose a higher risk of default.

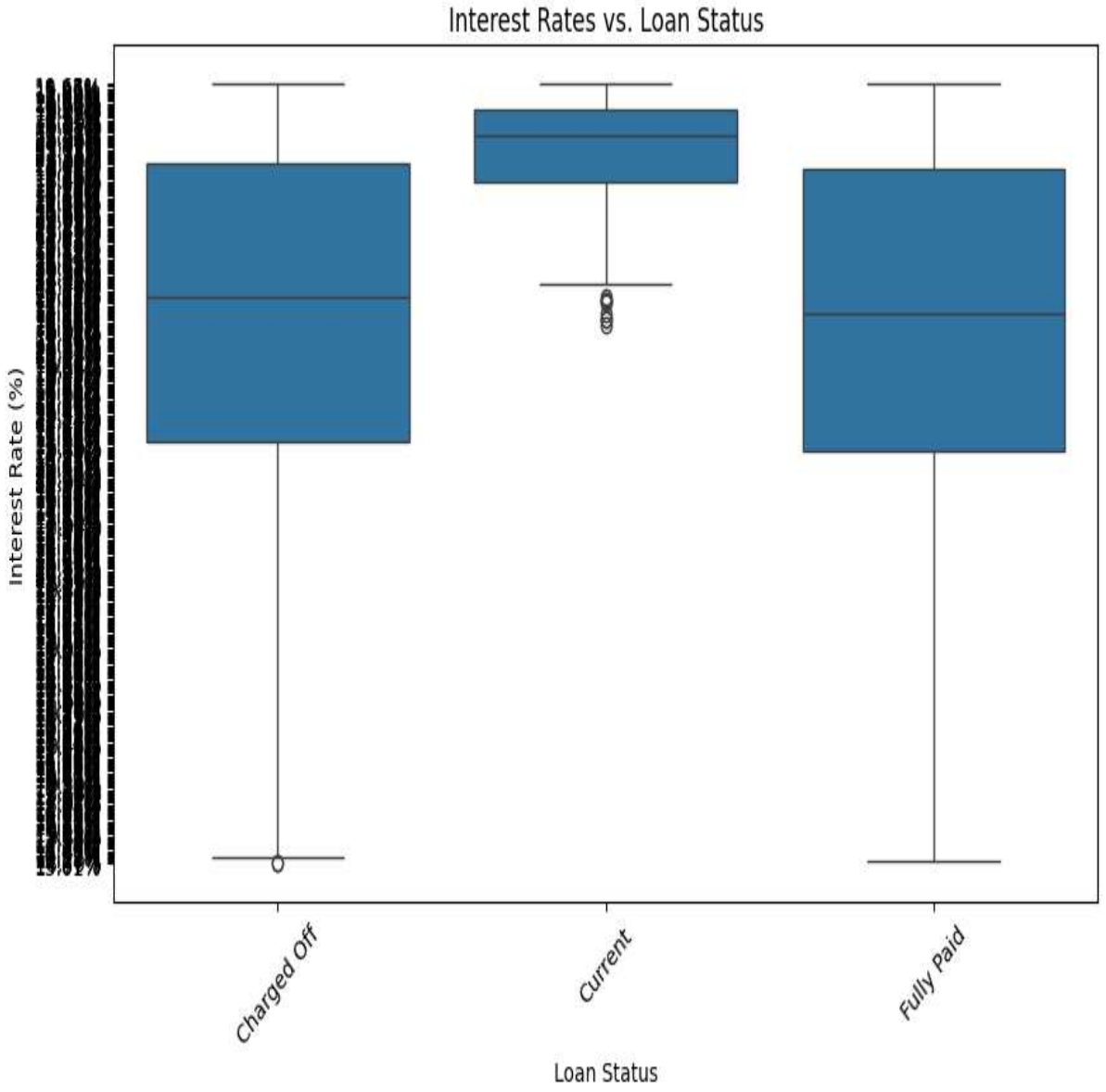


INTEREST RATE vs. LOAN STATUS

Box plots to compare interest rates for different loan statuses.

Insights:-

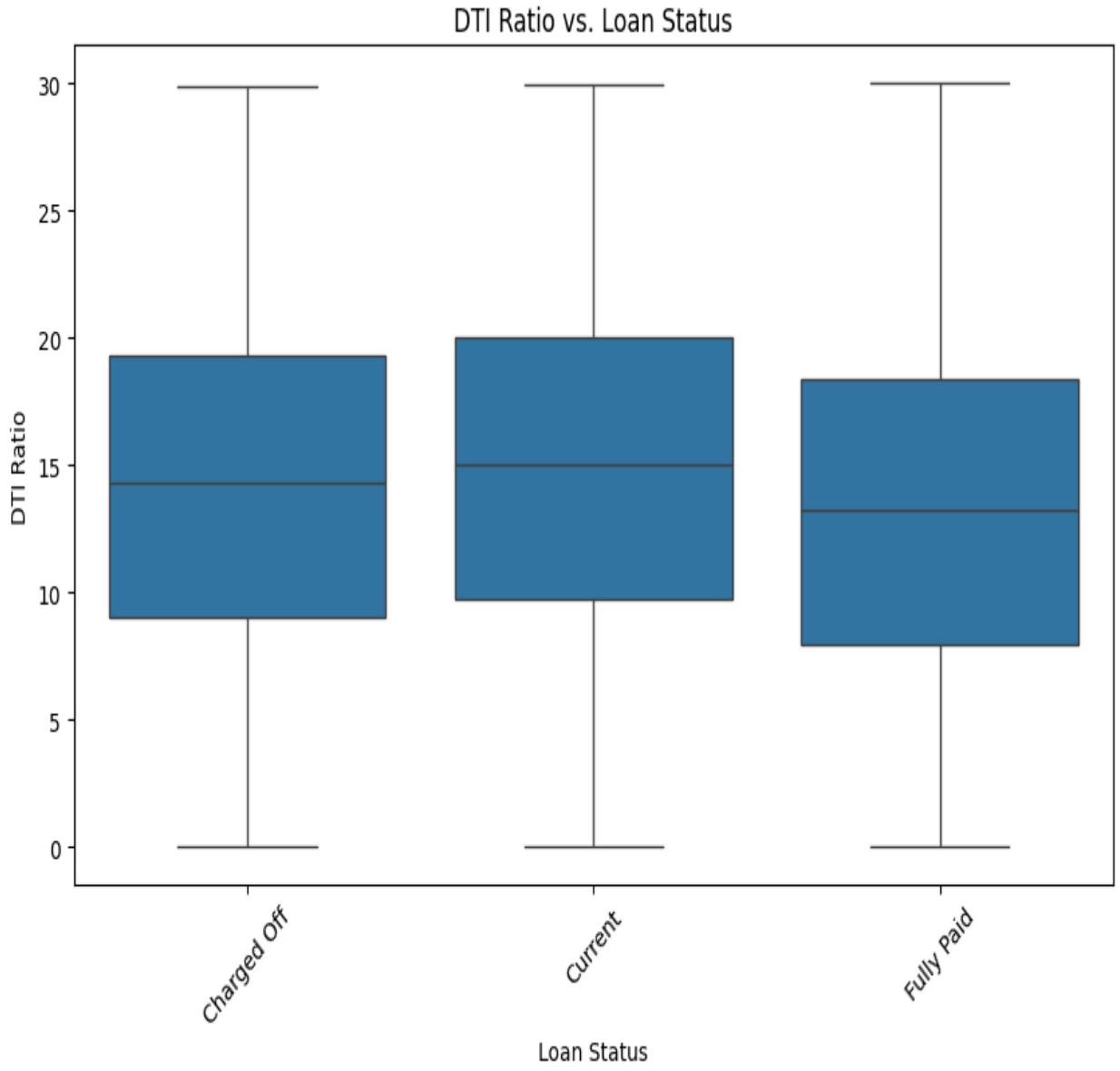
- Loans that are currently being paid (Current status) exhibit the highest average and median interest rates among all loan statuses.
- Charged-off loans have higher interest rates on average compared to fully-paid loans. This implies that higher interest rates could potentially correlate with increased default risk.



DTI RATIO vs. LOAN STATUS

Insights:-

- Loans that are currently being paid (Current status) show the highest average and median Debt-to-Income (DTI) ratios.
- Charged-off loans have higher DTI ratios than fully paid loans, suggesting that borrowers with higher DTI ratios may face greater challenges in loan repayment.

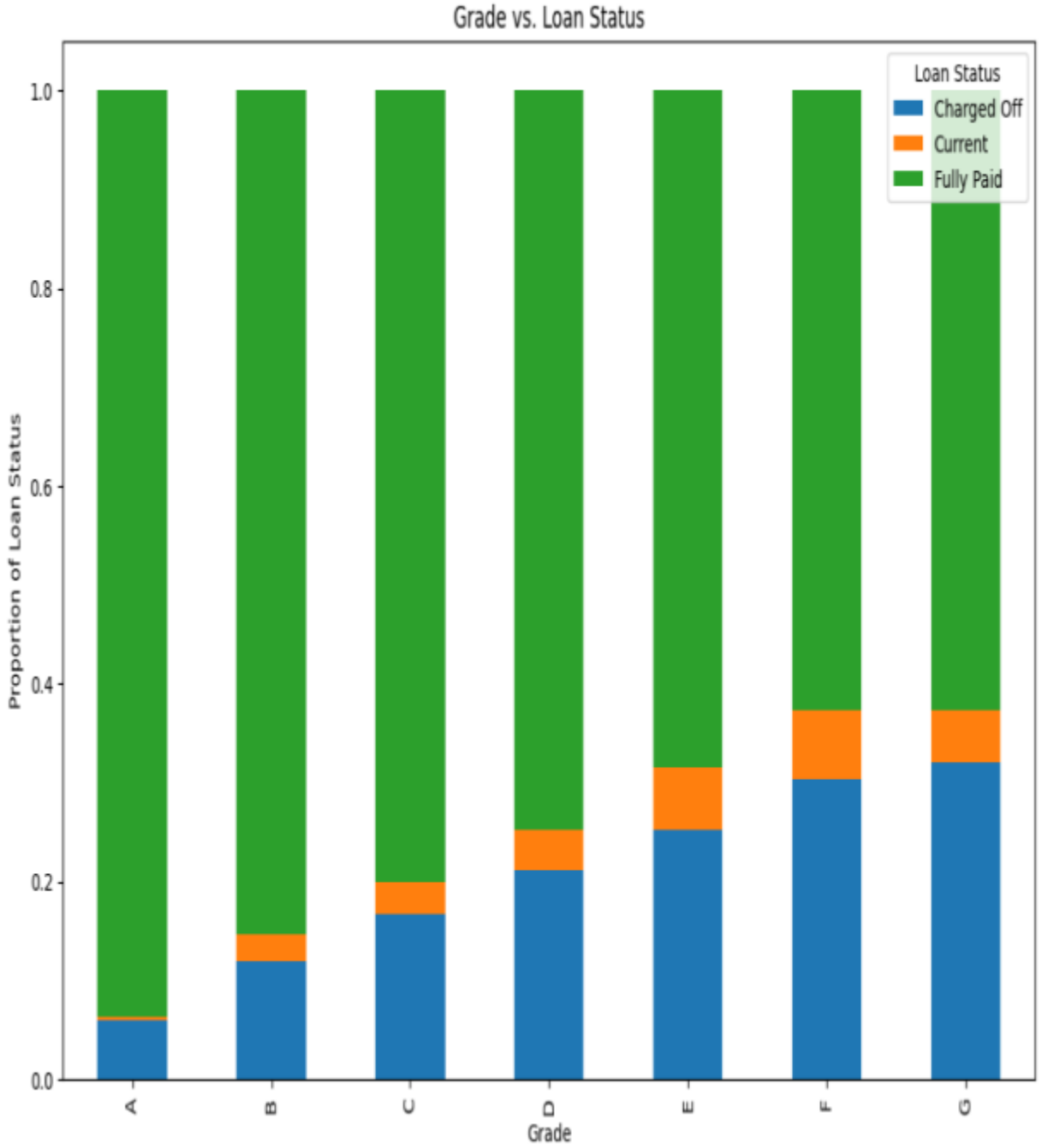


GRADE vs. LOAN STATUS

Used a stacked bar chart to compare the distribution of loan grades across different loan statuses. This will help to see if certain grades are more likely to result in default or full payment.

Insights:-

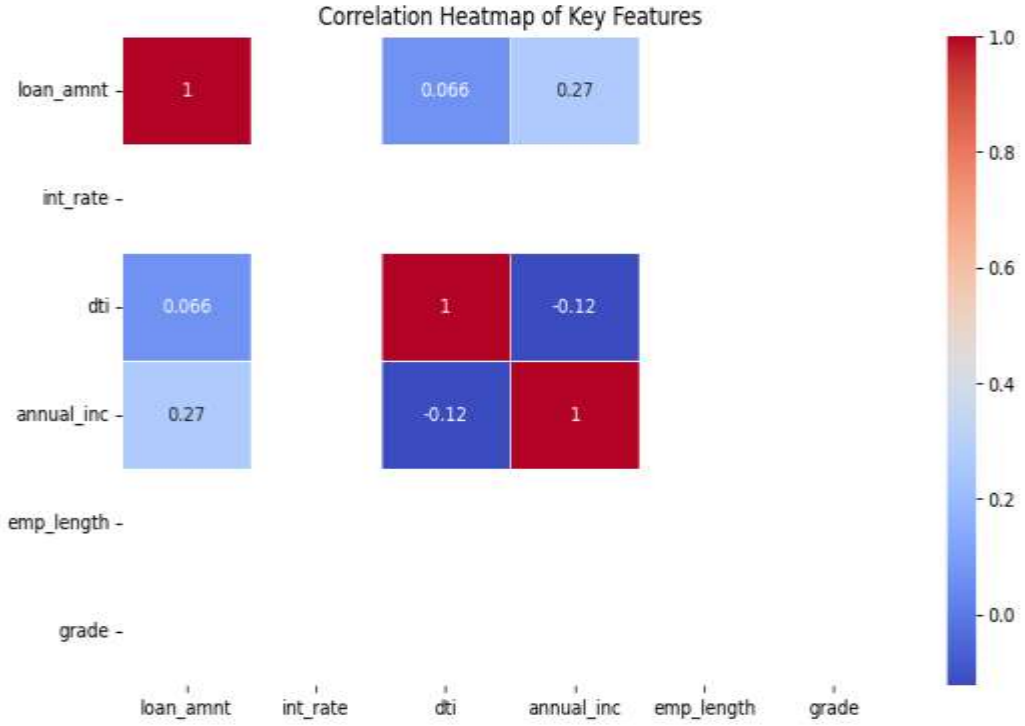
- Loans with higher grades (A and B) have a significantly higher proportion of fully paid loans compared to lower grades (F and G). This indicates that loans with better grades are associated with lower default rates.
- Lower loan grades (F and G) have a higher proportion of charged-off loans, highlighting an increased risk of default with lower-quality loans.



CORRELATIONAL ANALYSIS

CHAPTER-3

Correlation analysis helps us in understanding the relationship between different variables. A strong correlation indicates a significant relationship, which can be used to make informed decisions.



Insights:-

Loan Amount (loan_amnt)

- Annual Income (annual_inc): Shows a moderate positive correlation (0.27) with loan amount. This implies that higher annual income tends to be associated with higher loan amounts.
- Debt-to-Income Ratio (dti): Shows a weak positive correlation (0.066) with loan amount. This suggests a slight tendency for higher loan amounts to be associated with higher DTI ratios, but the relationship is not strong.

Interest Rate (int_rate)

- The correlation with other key features is not shown in this heatmap, suggesting it might not have strong direct correlations with the selected features.

Debt-to-Income Ratio (dti)

- Annual Income (annual_inc): Shows a weak negative correlation (-0.12) with DTI. This indicates that higher annual income slightly tends to be associated with lower DTI ratios.

Annual Income (annual_inc)

- Loan Amount (loan_amnt): As mentioned, shows a moderate positive correlation (0.27) with loan amount.
- Debt-to-Income Ratio (dti): Shows a weak negative correlation (-0.12) with DTI.

CONCLUSION

Based on the analysis, the driving factors (or strong indicators) behind loan default can be summarized as follows:-

- Higher Loan Amounts: There is a correlation between higher loan amounts and increased risk of default. Charged-off loans tend to have higher average loan amounts compared to fully-paid loans.
- Higher Interest Rates: Loans with higher interest rates are more likely to end up as charged-off. This indicates that interest rates play a significant role in determining the risk of default.
- Lower Grades (Credit Grades): Lower credit grades (e.g., grades D, E, F, G) are associated with a higher proportion of charged-off loans. Applicants with lower credit grades pose a higher risk of default.
- Higher Debt-to-Income (DTI) Ratio: Loans with higher DTI ratios are more likely to default. Charged-off loans tend to have higher DTI ratios compared to fully-paid loans.
- Employment Length: While not explicitly mentioned in the summary, employment length can also be a driving factor. Applicants with shorter employment histories or unstable employment records might be at a higher risk of default.
- Annual Income: Although not strongly highlighted in the provided insights, annual income often correlates with the ability to repay loans. Higher-income levels generally indicate better repayment capability.

Thank you