

LENDING CLUB

CASE STUDY

Case Study by

**ARUN KUMAR GOPINATH
CHIRANJIV DIMRI**

AGENDA

003	Executive Summary	009	Chi Square Test
004	Problem Statement, Analysis	010	Multivariate Analysis
005	Univariate Analysis	011	Final Assessment
006	Univariate Analysis		
007	Bivariate Analysis		
008	Bivariate Analysis		

EXECUTIVE SUMMARY

We aim to help the company, a leader in online loans, address a key challenge: reducing losses from borrowers who default on their loans ('charged-off').

By analyzing data, we will identify the factors that indicate a higher risk of default.

We aim to enable smarter lending decisions, minimize financial losses, and improve the company's risk assessment processes.

Our ultimate goal is to uncover the drivers of loan defaults and contribute to building a safer, more reliable lending portfolio.



PROBLEM STATEMENT & ANALYSIS APPROACH

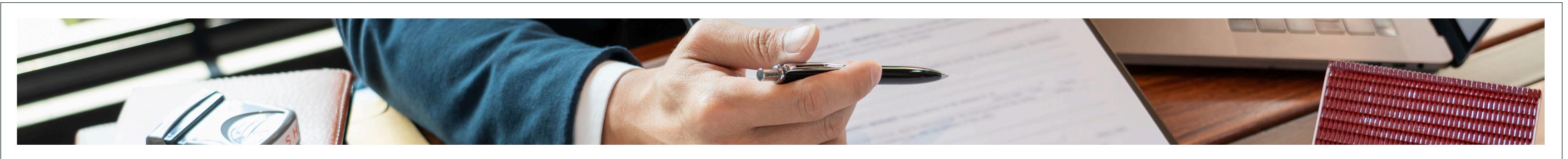
Problem Statement

To identify factors that influence loan defaults (charged-off applicants) using EDA, enabling better risk assessment and minimizing financial losses. The analysis focuses on consumer and loan attributes to uncover patterns associated with default risks.

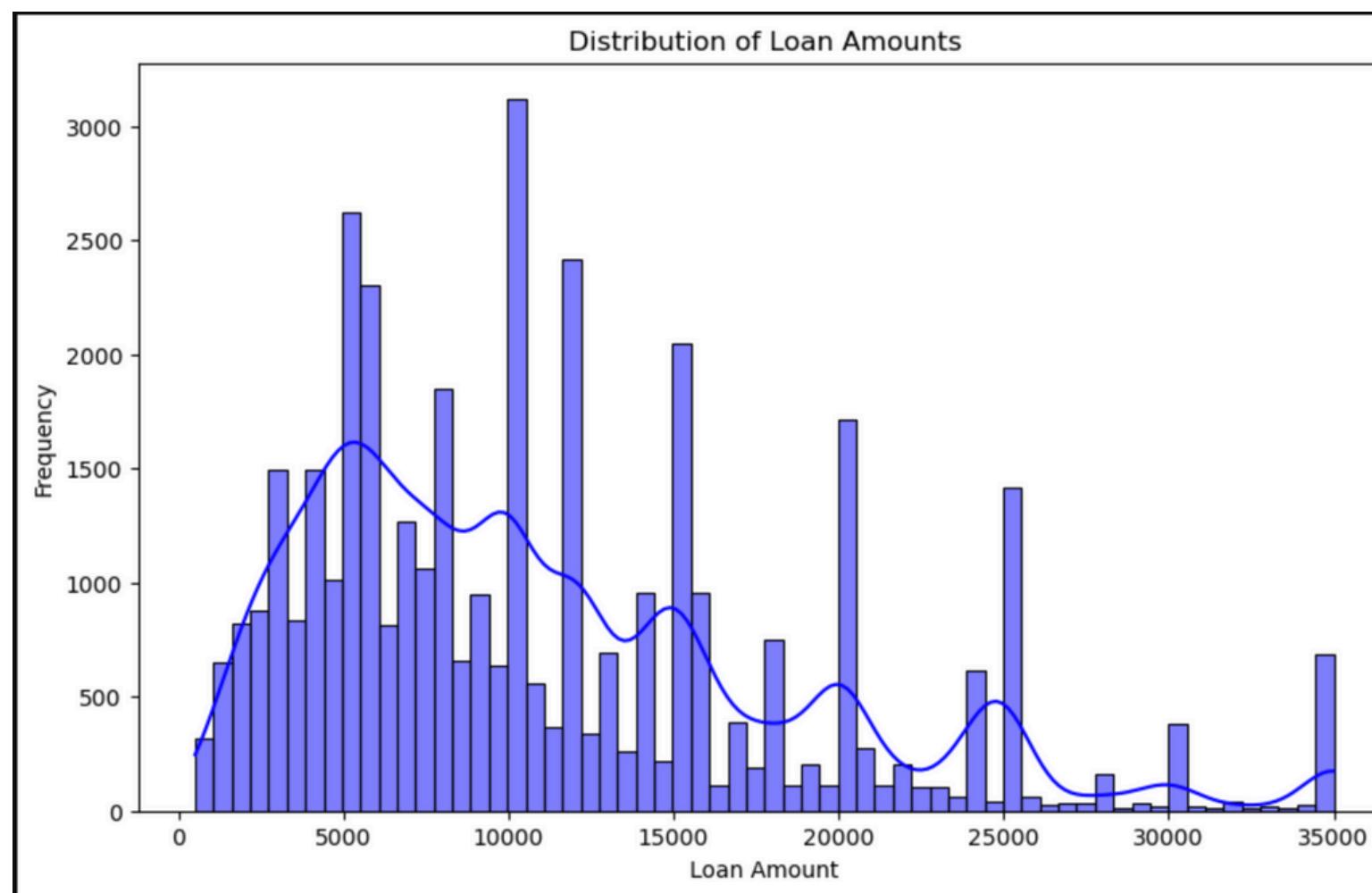
Analysis Approach

It involves cleaning and exploring the data through univariate and bivariate analysis to identify key predictors of defaults. Insights are visualized and summarized to recommend strategies for rejecting or adjusting loans for high-risk applicants.





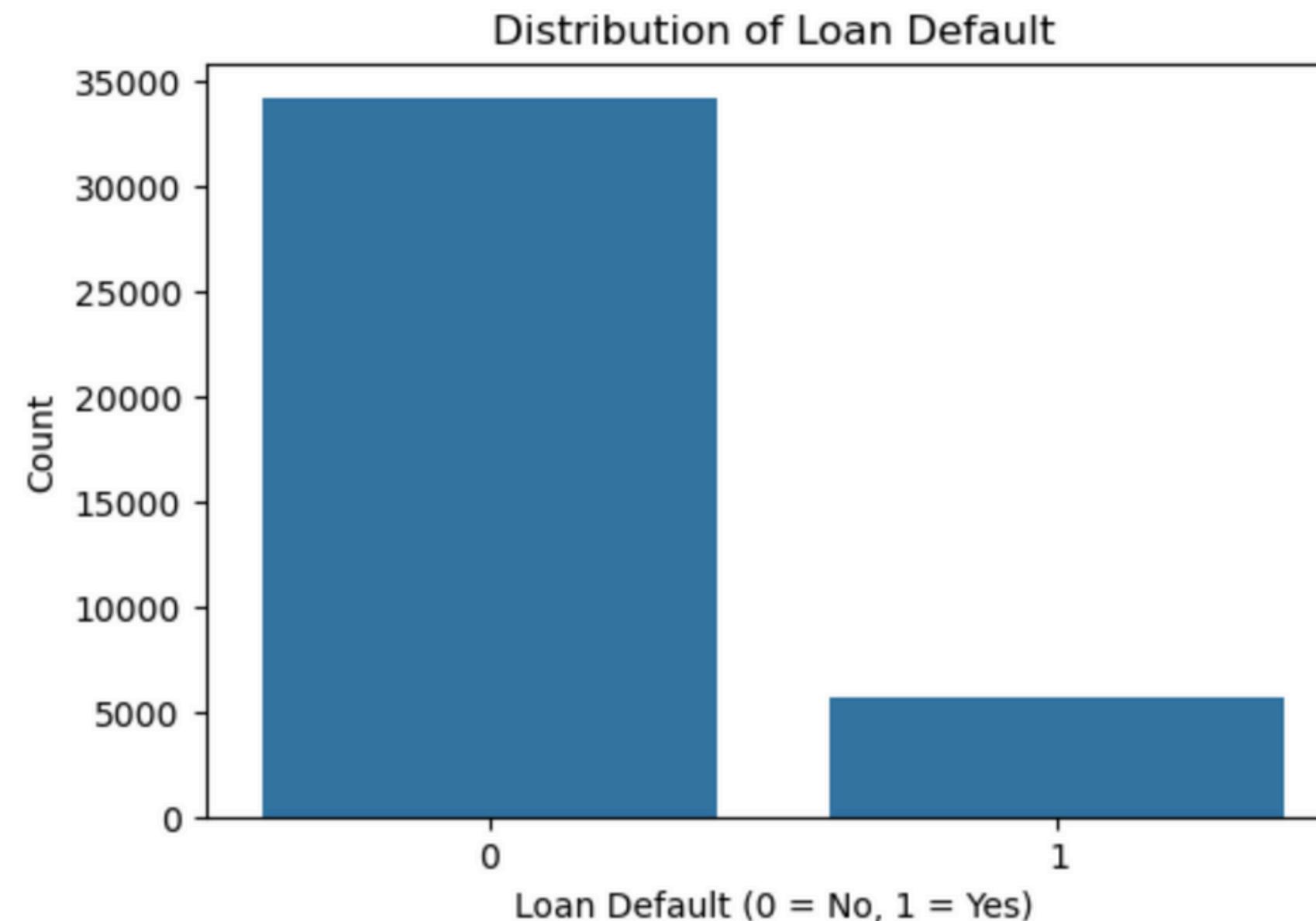
UNIVARIATE ANALYSIS



The histogram coupled with a density plot helps us to visualise the distribution of the loan amounts requested by applicants. This information is crucial to assess the typical loan sizes sought by customers.

Based on the visualisation we can say that approx Rs. 10,000 is the most asked-for loan amount.

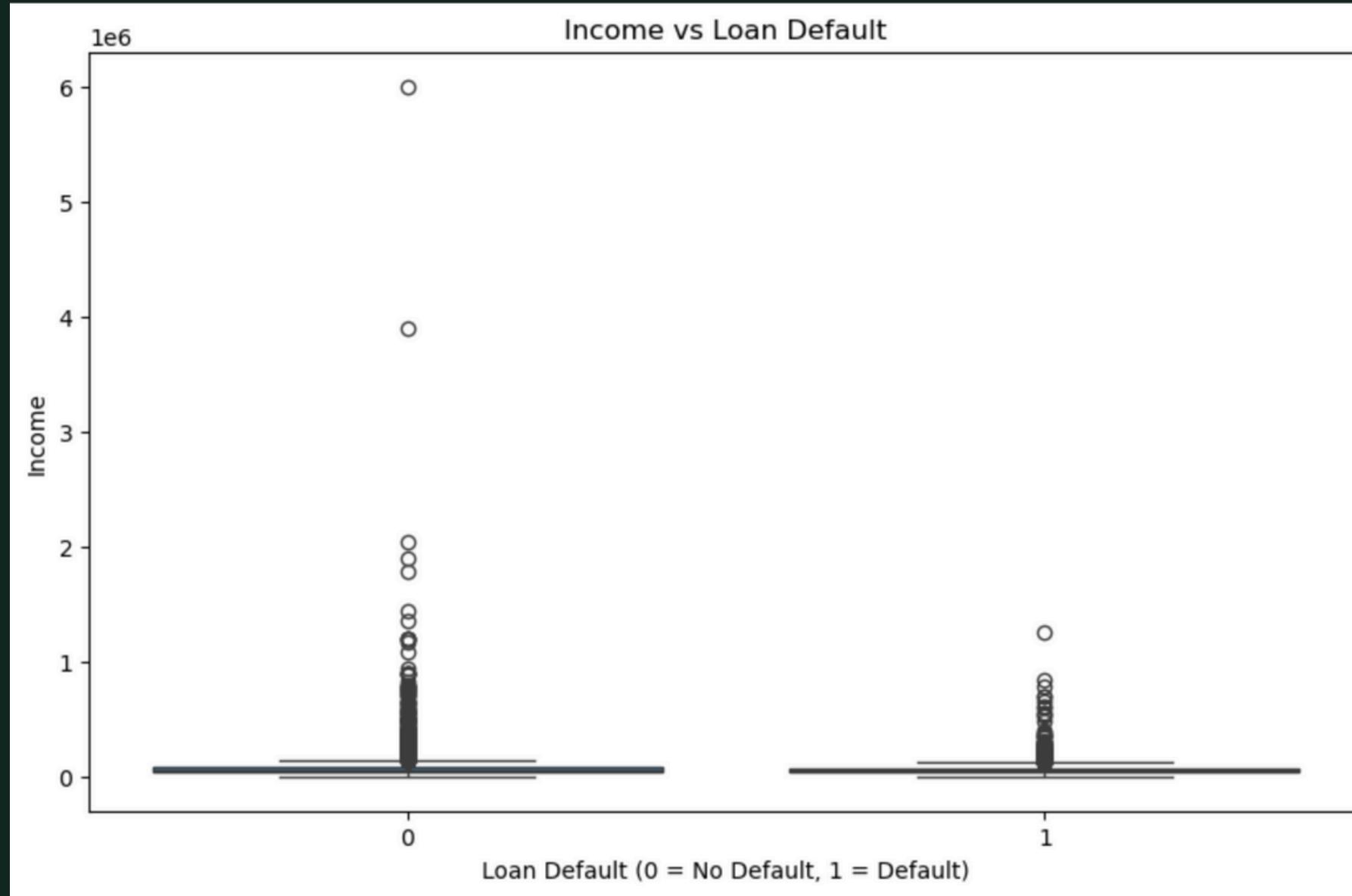
Majority of loan amounts fall between the bracket of Rs. 5-15,000. Followed by Rs. 20-25,000



This visualization aids in assessing the proportion of loans that have defaulted, guiding risk assessment strategies, loss mitigation efforts, decision-making processes regarding loan approvals, and performance evaluation in managing credit risk.

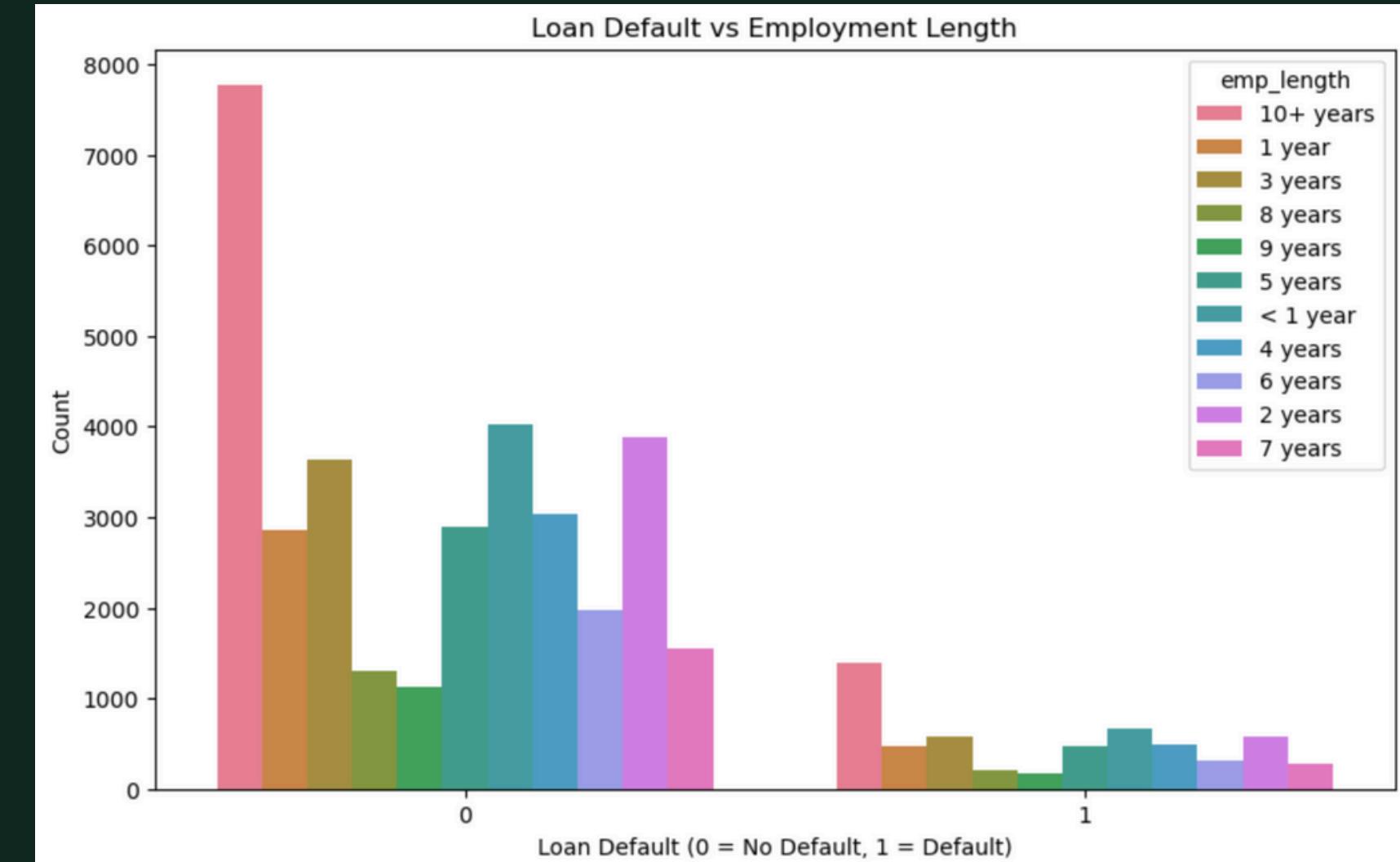
We can observe from the countplot that the number of loan defaults is around 5000.

BIVARIATE ANALYSIS



The boxplot comparing loan defaults against income provides insights into the relationship between income levels and loan defaults. It helps in understanding if there is a correlation between income and the likelihood of default. A higher median income for non-defaulted loans compared to defaulted loans suggests that applicants with higher incomes are less likely to default. This helps us assess risk based on income levels and potentially adjust lending criteria or interest rates accordingly to reduce default risk for lower-income applicants.

The countplot analyzing loan defaults based on employment length offers insights into how the length of employment relates to loan defaults. By visualizing the distribution of loan defaults across different employment lengths, we can assess that there is a pattern indicating that more experienced or stable employees are less likely to default on loans. This analysis can help in refining risk assessment strategies, setting policies for applicants with varying employment lengths, and developing targeted approaches to minimize default rates based on the stability and length of employment.

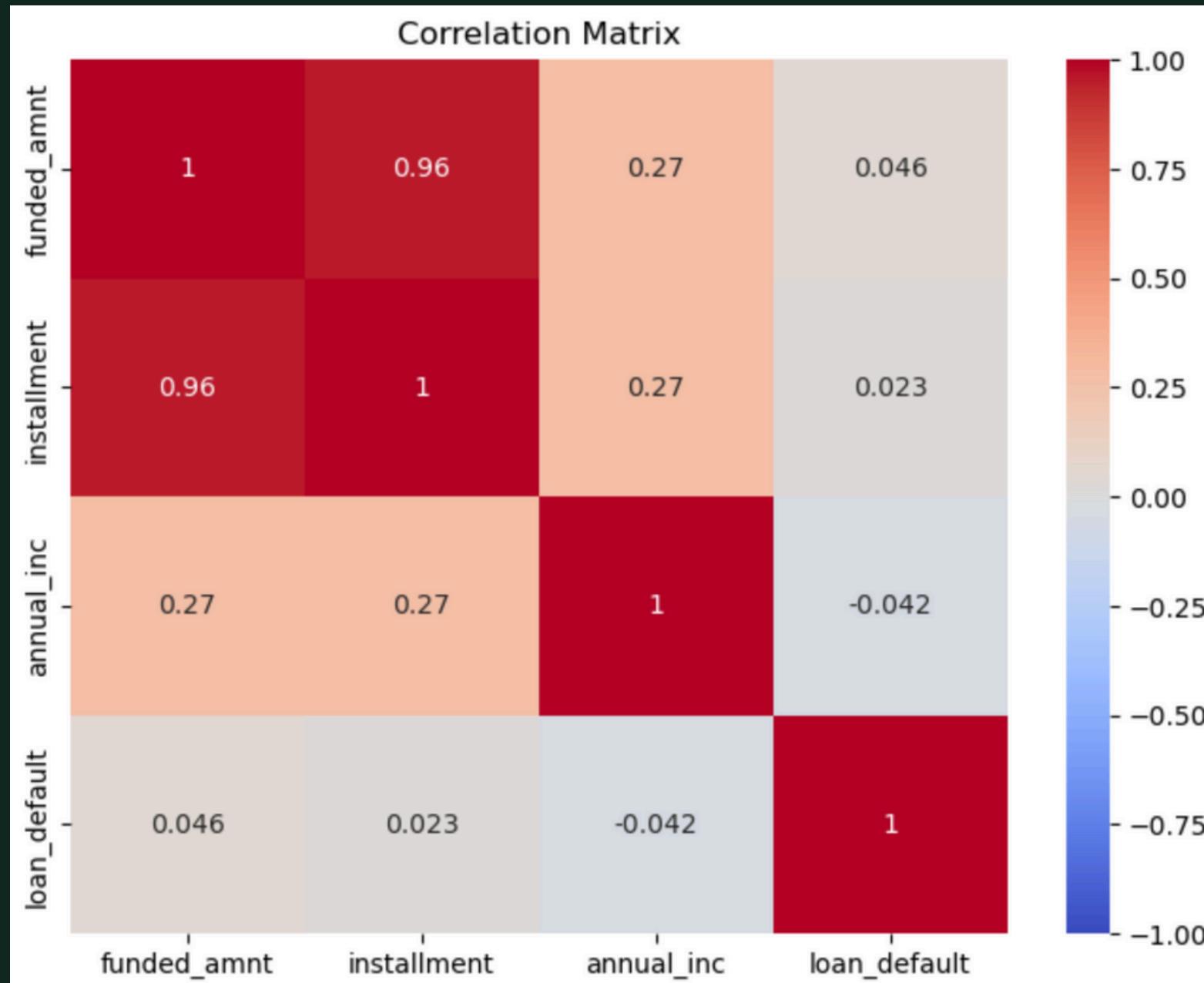


CHI SQUARE TEST

LOAN DEFAULT VS EMPLOYMENT LENGTH	LOAN DEFAULT VS INSTALLMENT CATEGORIES
<ul style="list-style-type: none">Chi-Square Test p-value: 0.13048318160324276The value of 0.13 approx suggests that there is no significant correlation between 'Loan Default' and 'Employment Length'. This is as the p-value is over 0.05.	<ul style="list-style-type: none">Chi-Square Test p-value: 1.5667787359233254e-06 (0.00000157)The value of 0.00000157 approx suggests that there is significant correlation between 'Loan Default' and 'Installment categories'. This is as the p-value is way under 0.05.



MULTIVARIATE ANALYSIS



The correlation matrix highlights the following observations-

1. The funded amount and the installment has a positive relation suggesting that a higher loan amount has a proportionately higher installment amount.
2. A significantly negative correlation between Annual Income & loan default suggests that individuals with higher annual income are less likely to default on loans, implying that income is a predictor of default risk.



FINAL ASSESSMENT

LIST OF TOP-10 MOST RISKY APPLICANTS:

- 1.921363
- 2.910779
- 3.992162
- 4.1232359
- 5.1231984
- 6.1081056
- 7.910970
- 8.976453
- 9.874161
- 10.1047042

THANK YOU

Case Study by

**ARUN KUMAR GOPINATH
CHIRANJIV DIMRI**