

# LENDING CLUB CASE STUDY

## Identifying Factors Influencing Loan Default

*Date:* **22nd May, 2024**

*Submitted by:*

**Saranya Thampi**

**Chandan Sahoo**



# INTRODUCTION

## Objective

- Understand key factors influencing loan defaults and borrower behavior.
- Identify high-risk borrowers based on the data.
- **'Dataset'**: Overview of the loan dataset used for analysis.

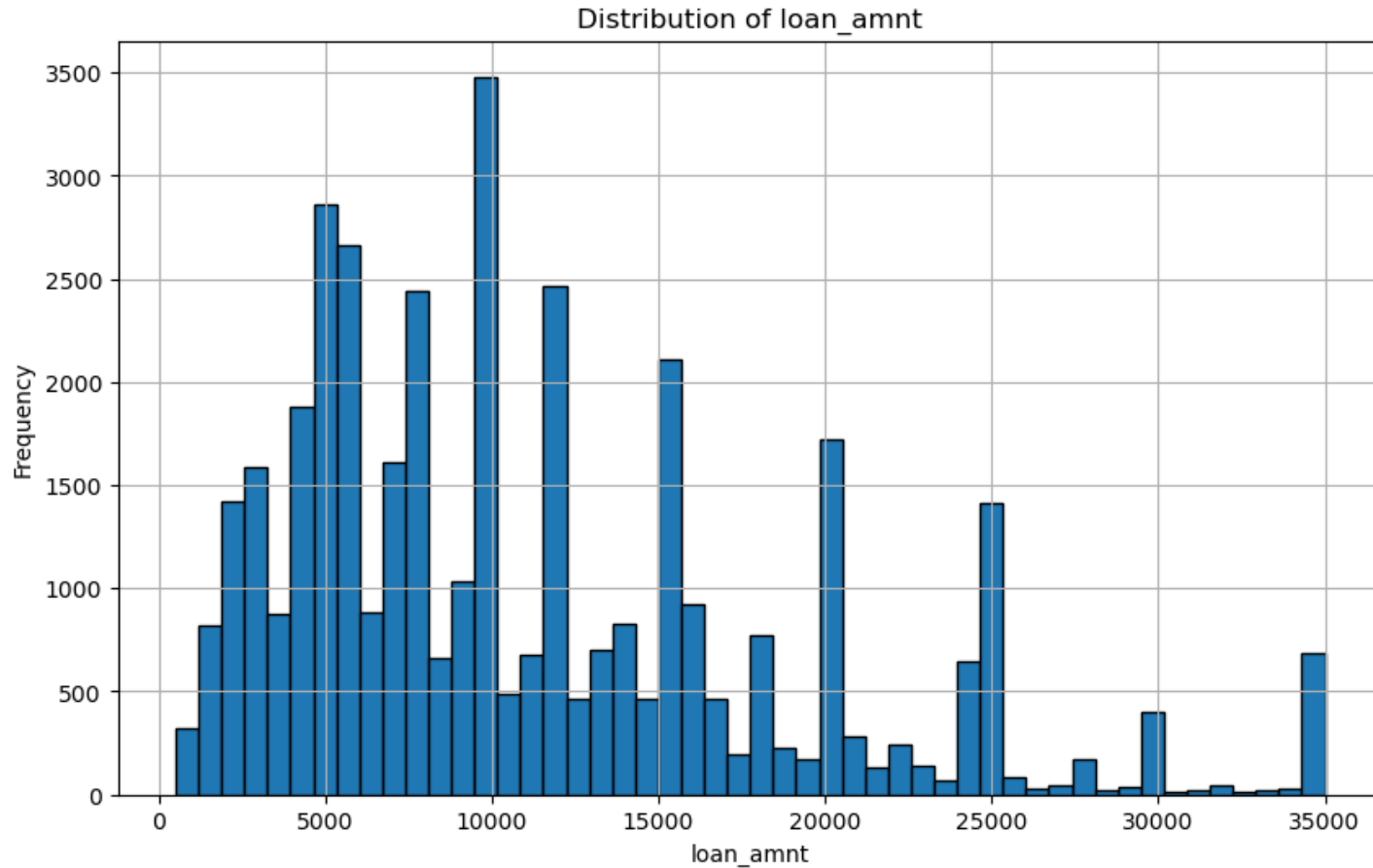
# DATA OVERVIEW

- The dataset contains various columns related to loan details, borrower information, and loan performance.
- **Initial Exploration** :Summary statistics, data types, and missing values.

# DATA CLEANING STEPS

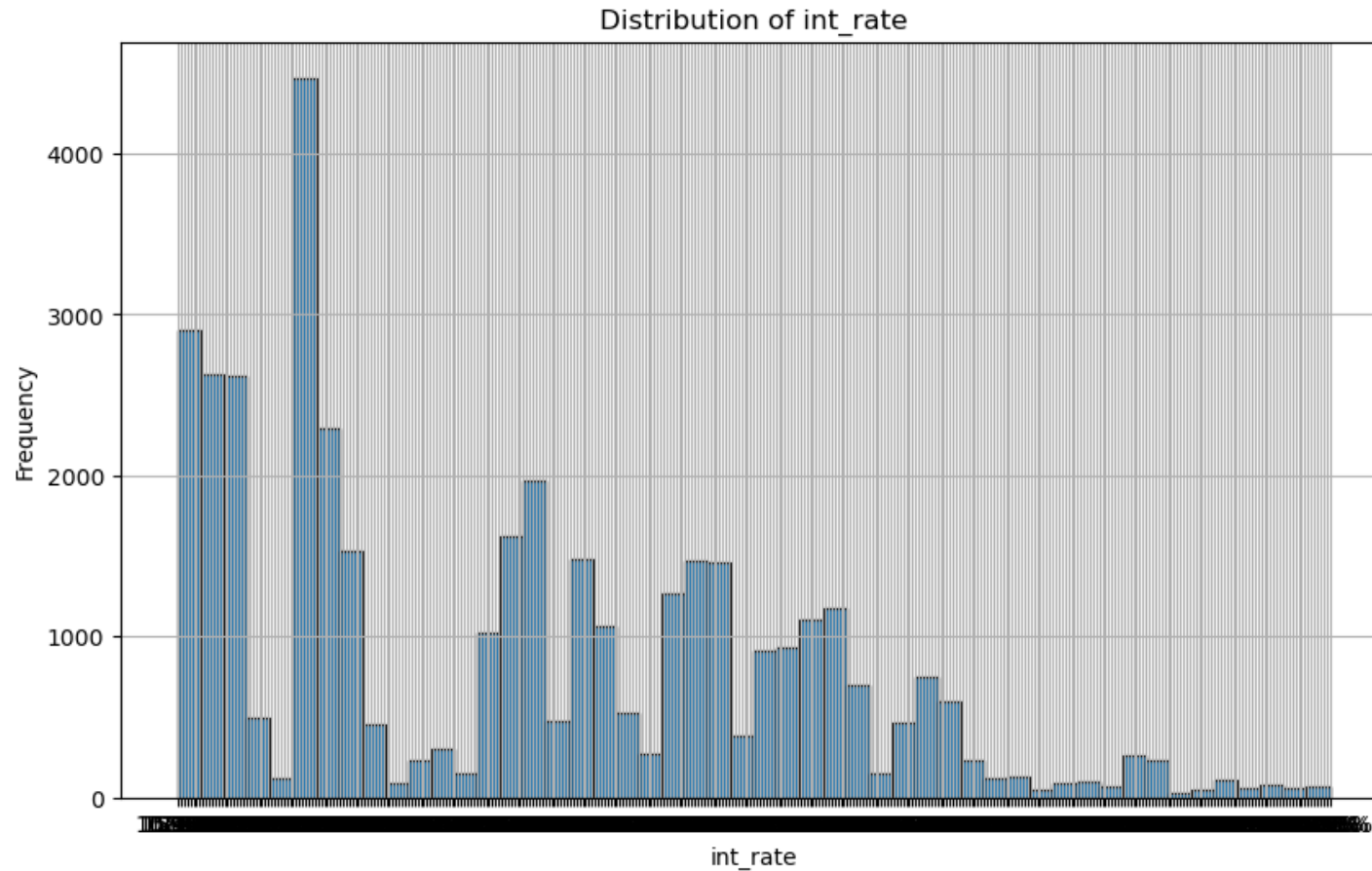
- **Irrelevant Columns:** Dropped columns like 'member\_id', 'application\_type', etc.
- **Missing Values:** Handled missing values by removing columns with all NaNs.
- **Data Types:** Converted relevant columns to categorical, datetime, etc.
- **New Metrics:** Created 'Credit Loss' and other derived metrics.

# DISTRIBUTION OF LOAN AMOUNT



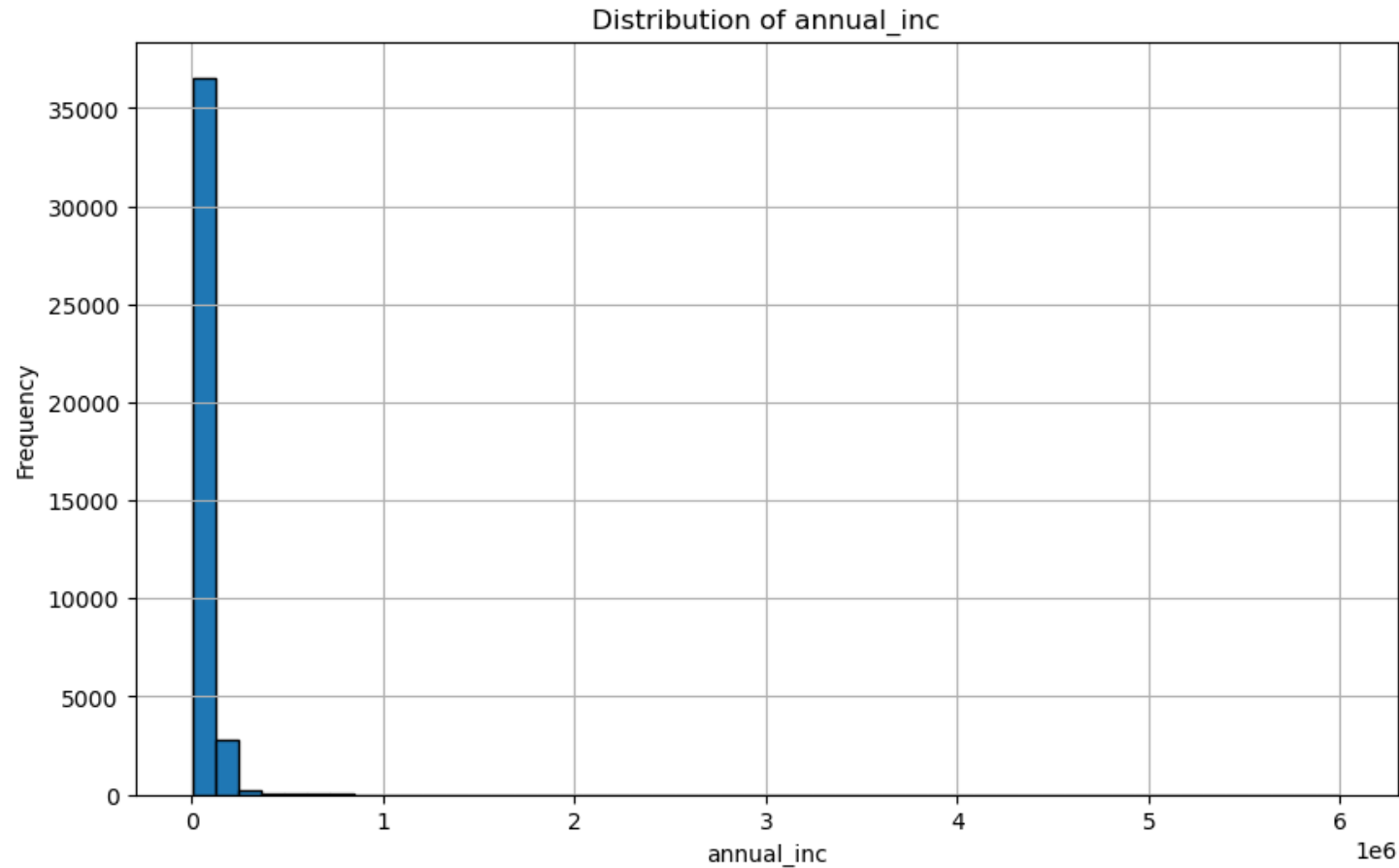
- Most loans are within a specific range, with some outliers on the higher end.
- Most loan amounts are concentrated at the lower range.

# DISTRIBUTION OF INTEREST RATE



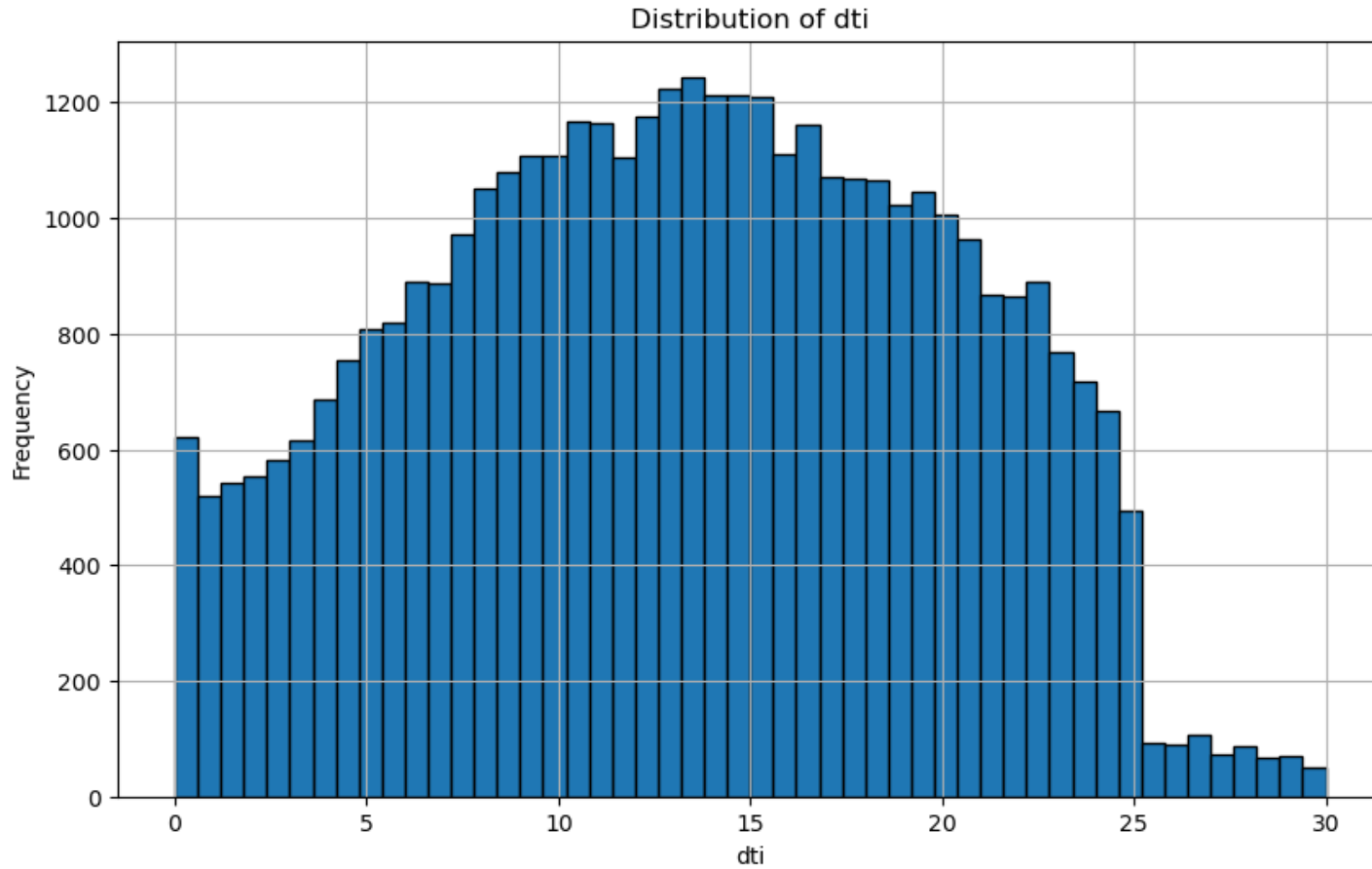
- Interest Rates are right skewed – meaning there are more loans at lower interest rates with a long tail extending to higher rates.
- The second peak suggests another cluster of loans at higher interest rates.

# DISTRIBUTION OF ANNUAL INCOME



- Annual Incomes are right skewed, with most borrowers are from the lower income groups.

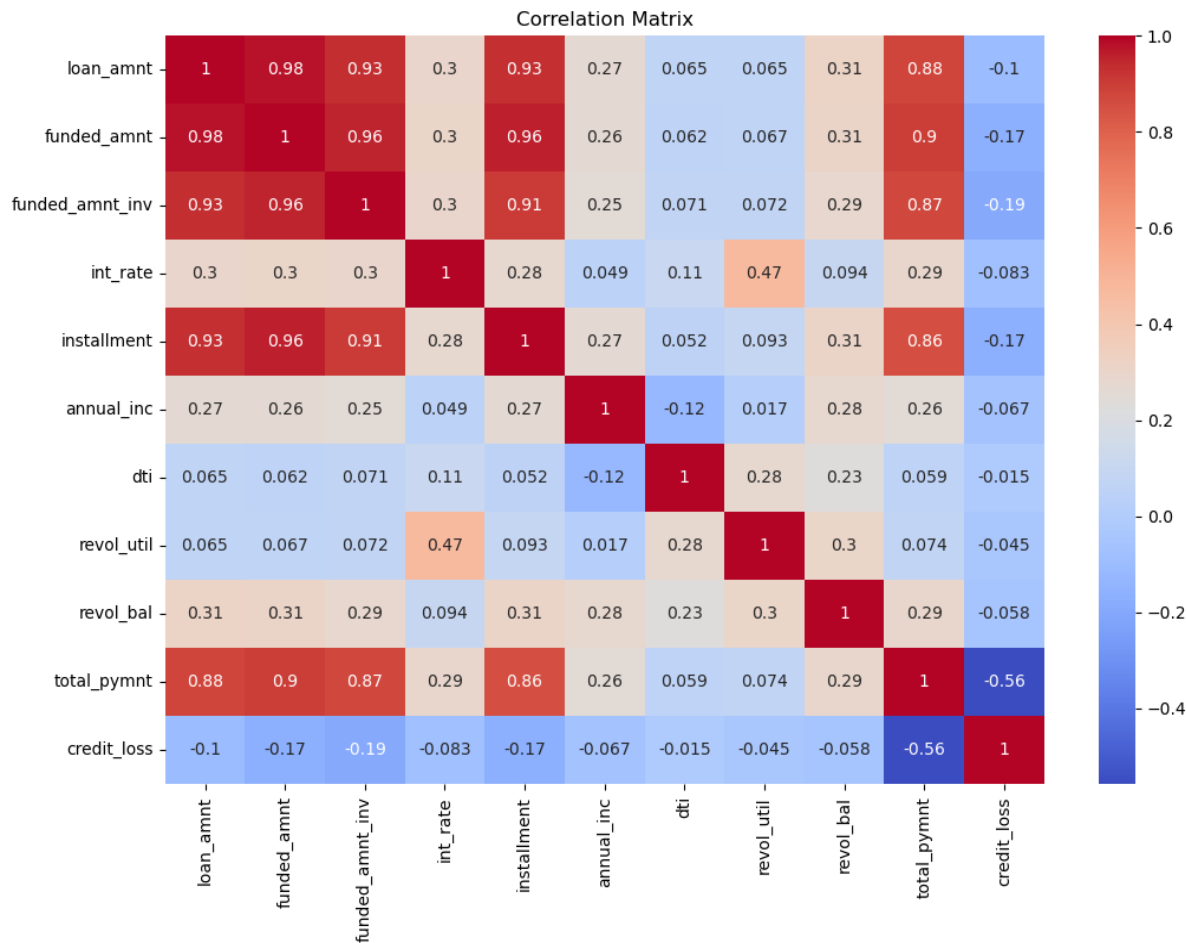
# DISTRIBUTION OF DEBT-TO-INCOME



- Most borrowers have a moderate DTI, with some outliers in the higher end.
- The DTI distribution shows. Most of the borrowers are within moderate range.



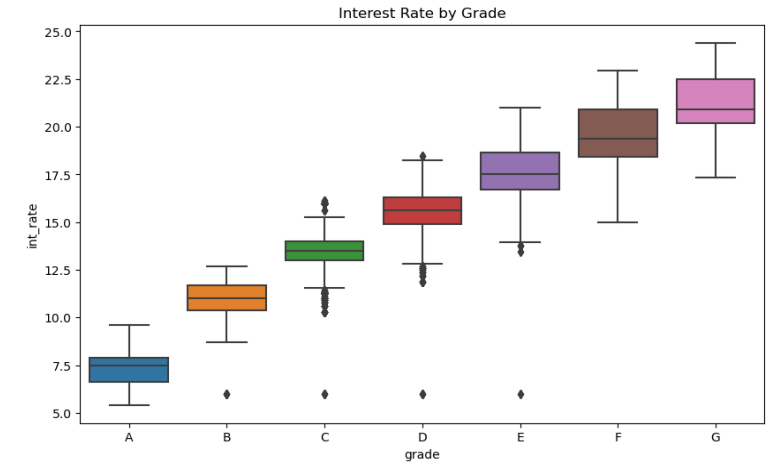
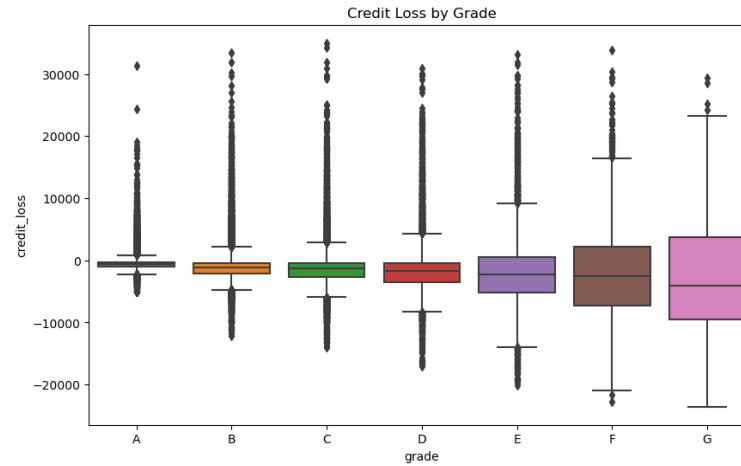
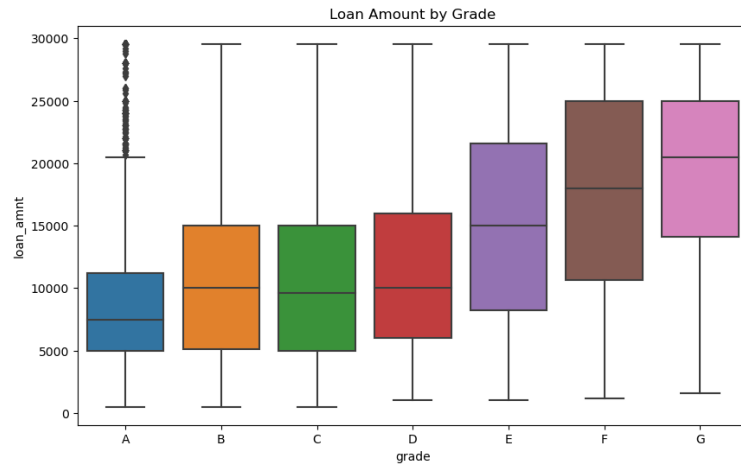
# CORRELATION ANALYSIS



The Correlation Matrix is a table showing correlation coefficients between variables. Each cell in the table shows the correlation between two variables. The values range from  $-1$  to  $1$ .

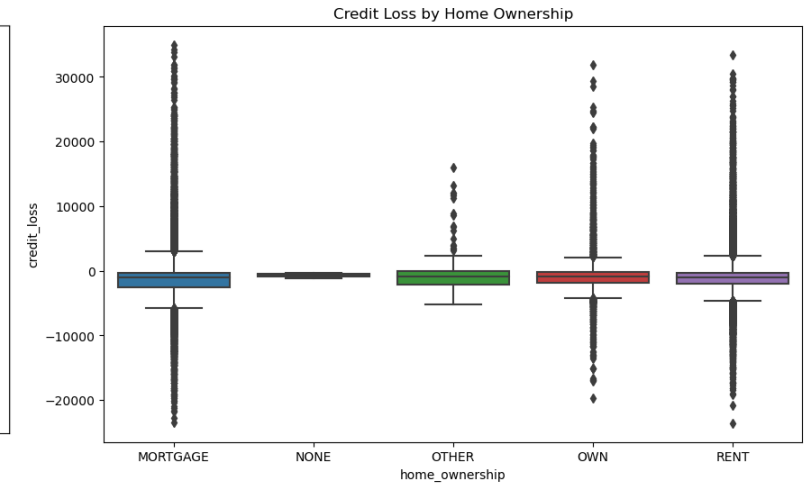
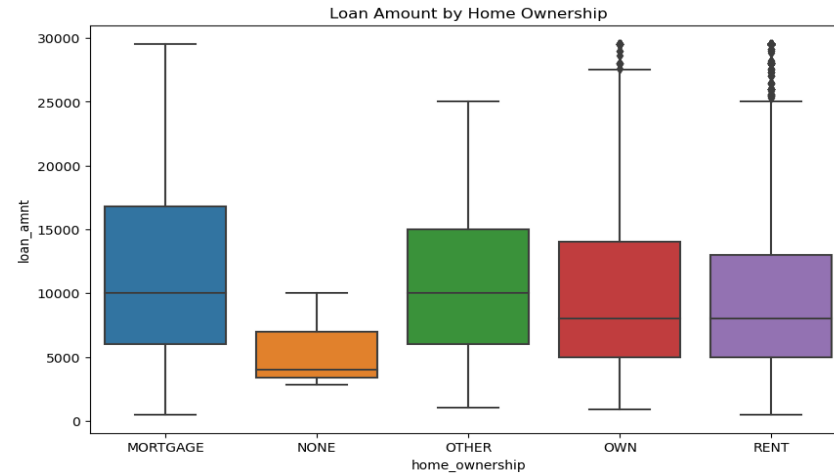
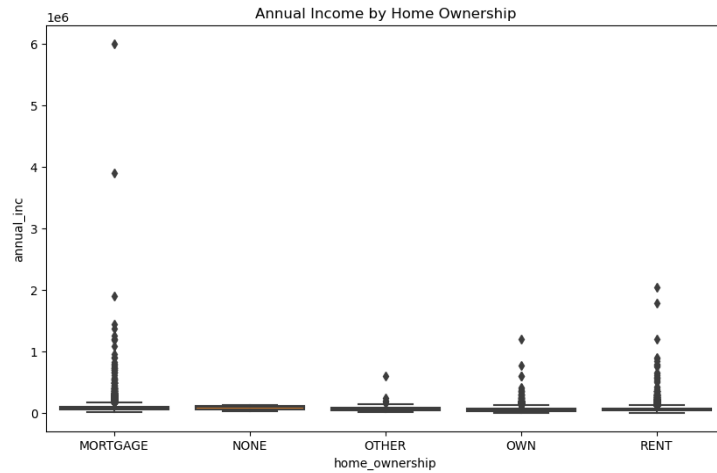
- **High Positive Correlations:** Variables that are highly positively correlated (e.g., 'loan\_amnt' and 'funded\_amnt') tend to move in the same direction. Understanding these relationships can help identify key factors that grow together.
- **High Negative Correlations:** Variables that are highly negatively correlated (e.g., 'dti' and 'annual\_inc') tend to move in opposite directions. These relationships can indicate trade-offs or balancing factors in financial metrics.

# LOAN GRADE ANALYSIS



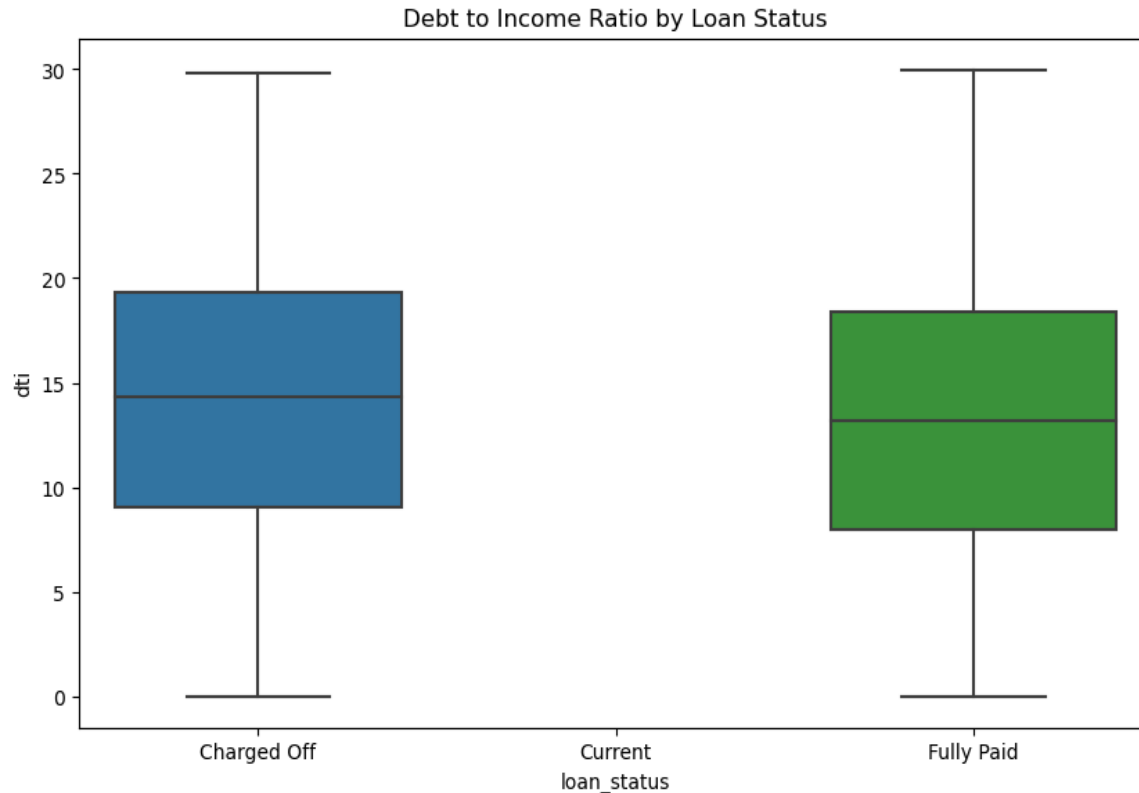
- Higher loan grades (A, B, C) are associated with lower interest rates and lower credit losses.
- Lower grades (D, E, F, G) have higher interest rates and higher credit losses.

# HOME OWNERSHIP ANALYSIS



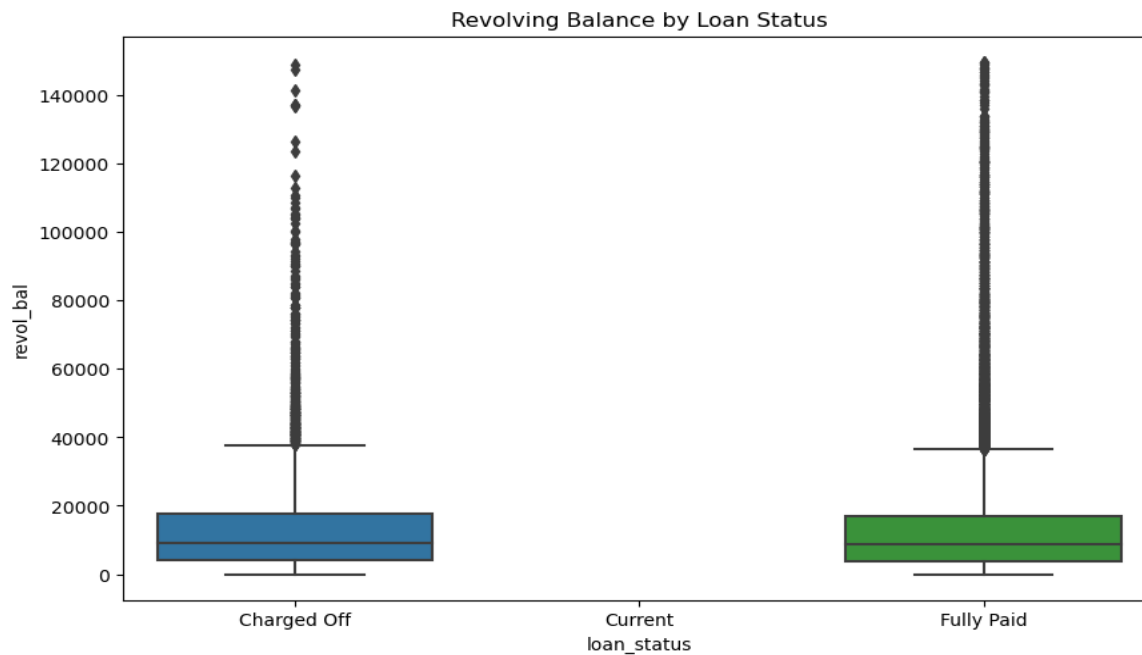
- Homeowners tend to request larger loan amounts.
- Homeowners have higher annual incomes compared to renters and others.

# LOAN STATUS ANALYSIS BY DEBIT-TO-INCOME RATIO



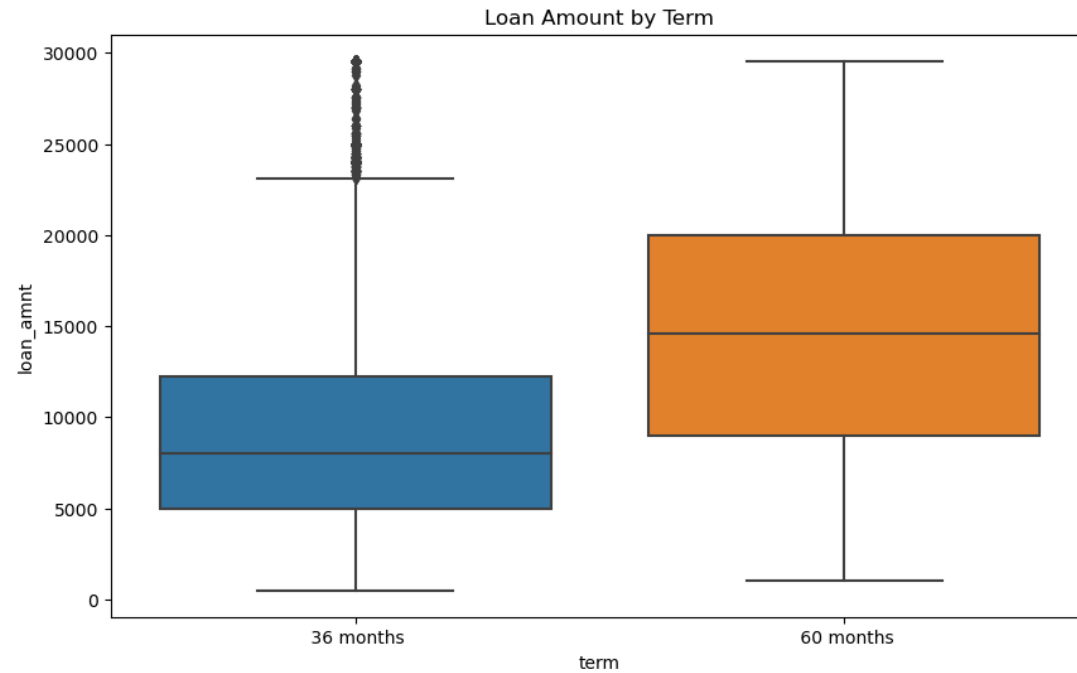
- Loan statuses with higher median DTI ratios indicate that loans in those categories are associated with borrowers who have higher levels of debt relative to their income.
- For example, if 'Charged Off' or 'Default' statuses show higher median DTI ratios compared to 'Fully Paid' status, it suggests that borrowers with higher DTI ratios are more likely to default or be charged off.

# LOAN STATUS ANALYSIS BY REVOLVING BALANCE



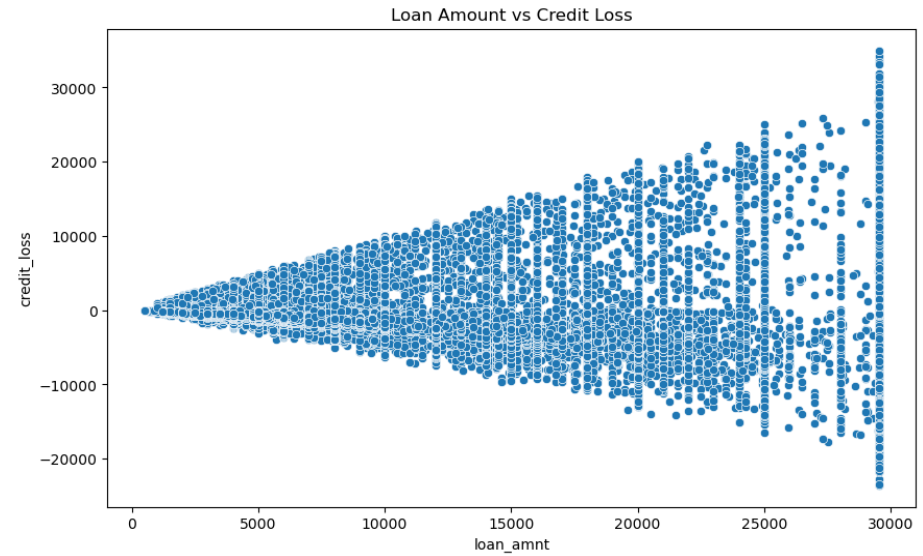
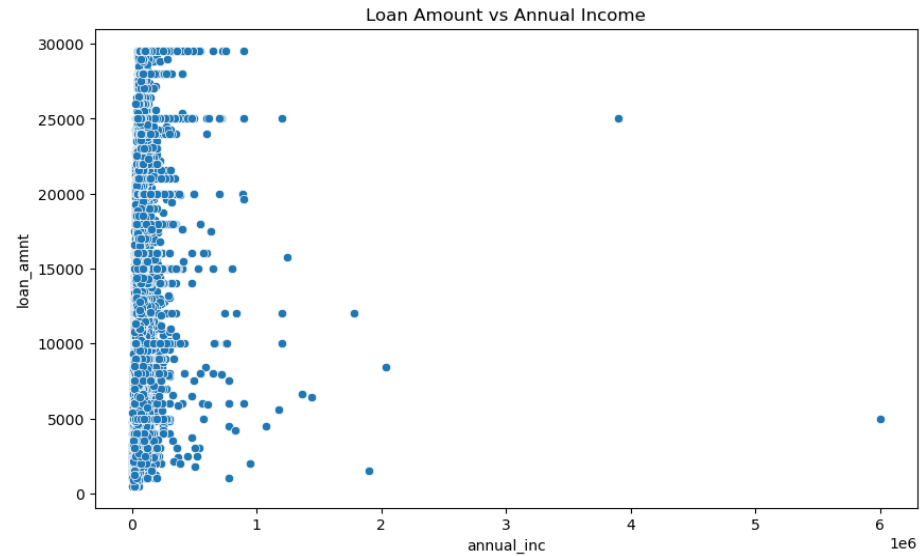
- Higher revolving balances may indicate higher financial stress, leading to worse loan outcomes.

# LOAN AMOUNT BY TERM



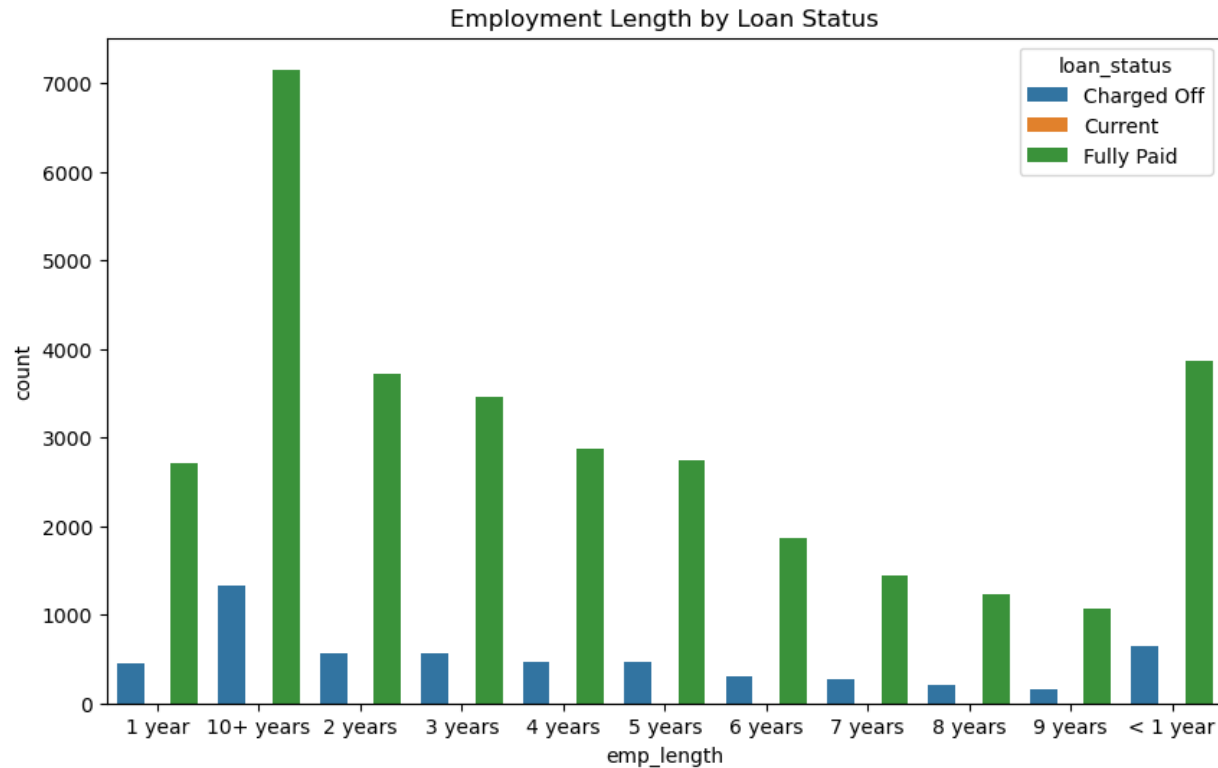
- Longer-term loans (60 months) tend to have higher loan amounts.

# SCATTER PLOT ANALYSIS



- Higher loan amounts correlate with higher annual incomes and higher credit losses.

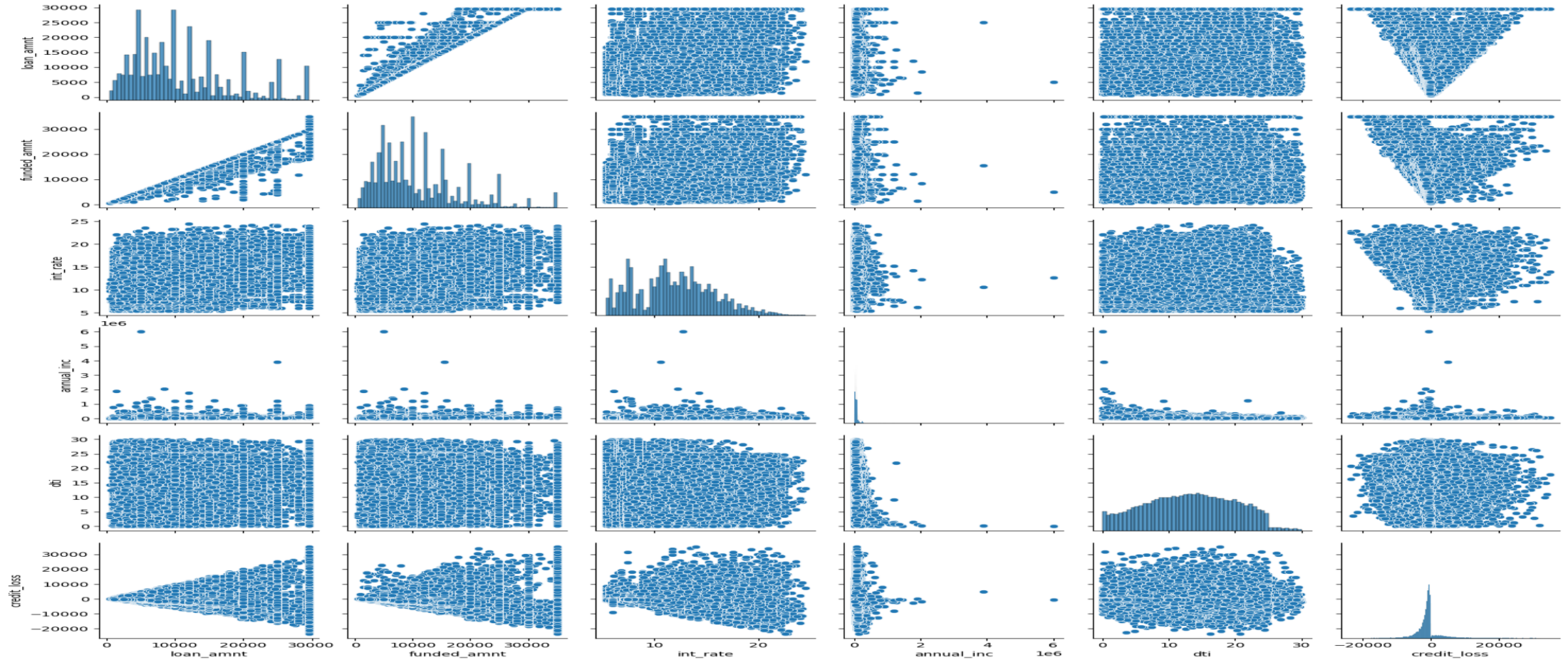
# EMPLOYMENT LENGTH ANALYSIS



- Shorter employment lengths are associated with higher probabilities of default.



# PAIR PLOT ANALYSIS



# Insights from the Pair Plot:

## Positive Correlations

- **loan\_amnt** and **funded\_amnt** show a strong positive correlation, it indicates that the loans requested are generally funded in similar amounts.
- **annual\_inc** and **loan\_amnt** show a positive correlation, it suggests that borrowers with higher incomes tend to request larger loans.

## Negative Correlations

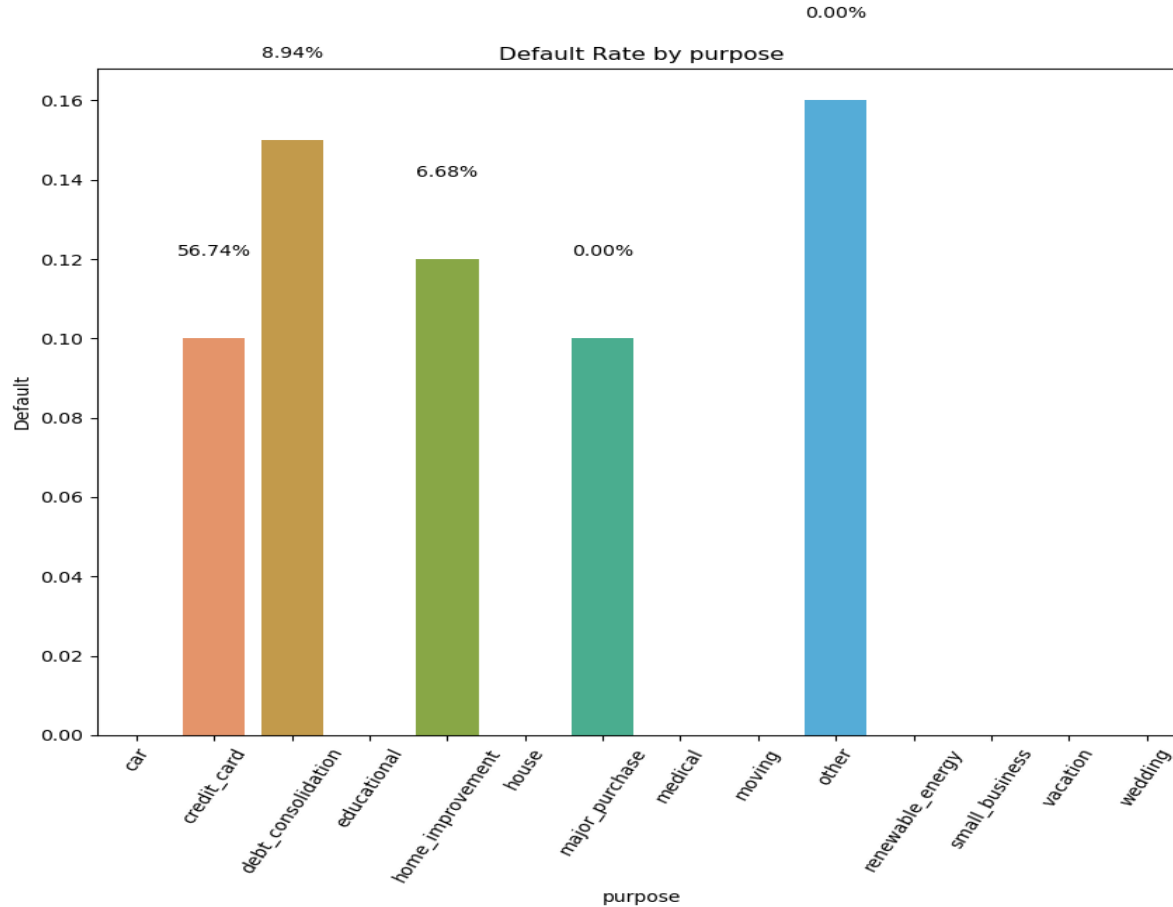
- **dti** and **annual\_inc** show a negative correlation, it indicates that higher-income borrowers have lower debt-to-income ratios.

## No Correlations

- Some pairs of variables show a scattered pattern with no clear trend, it indicates no strong relationship between those variables.

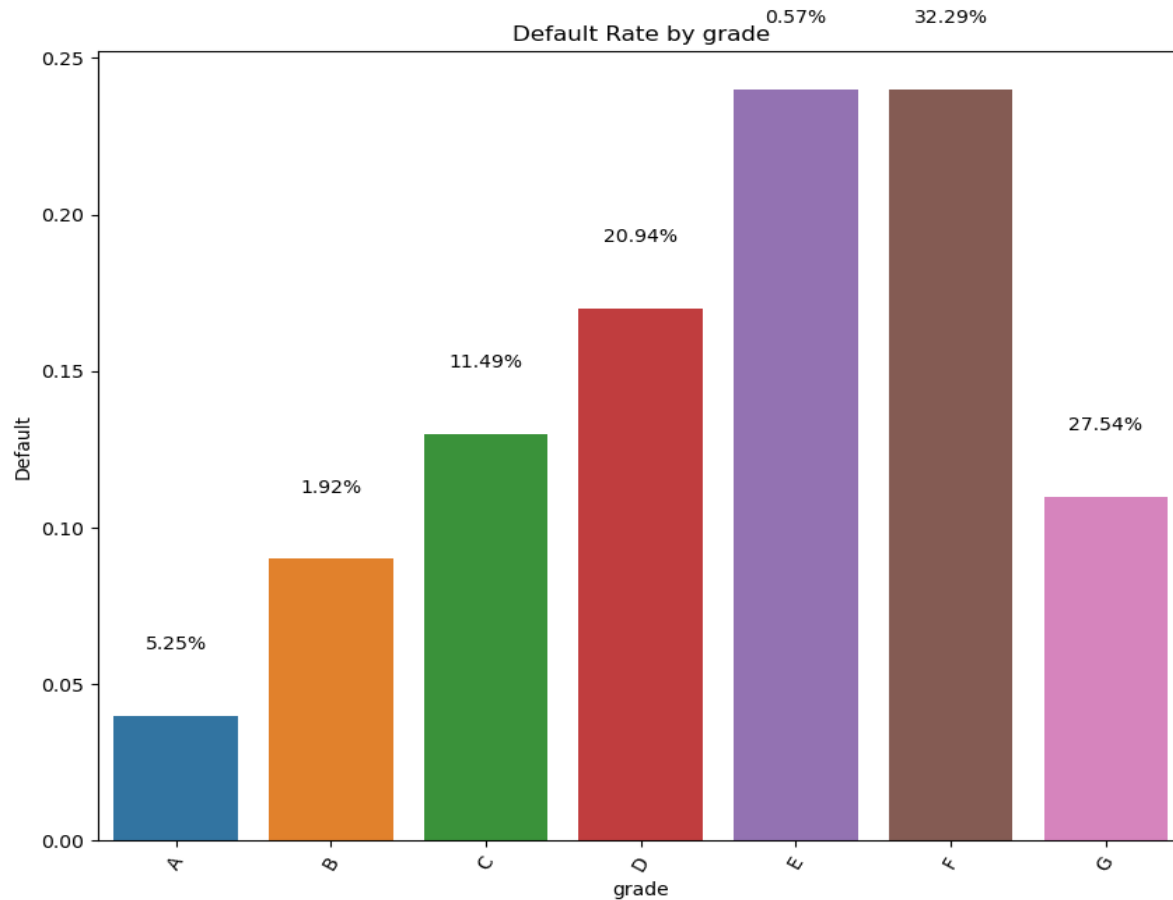


# DEFAULT RATE BY LOAN PURPOSE



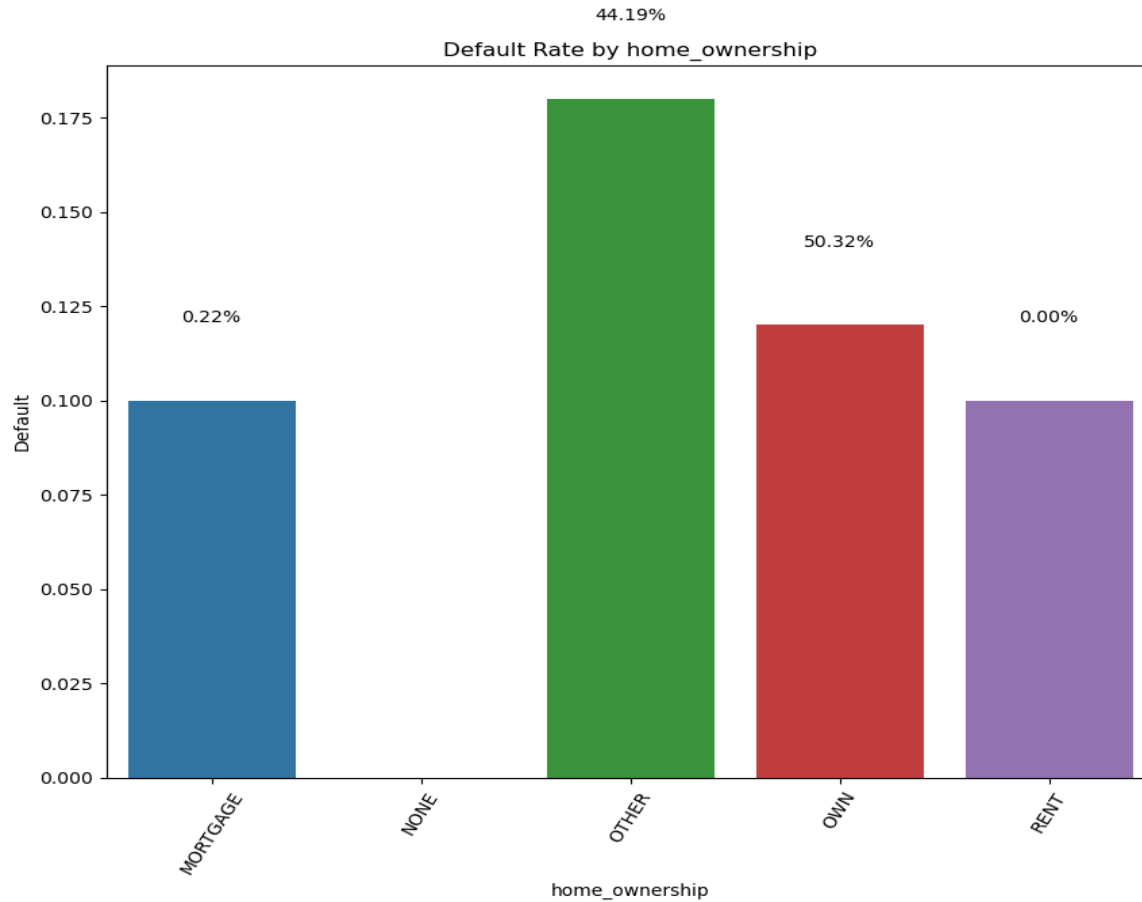
- Default rates vary significantly across loan purposes.
- Highest Default rates are across Debt collection

# DEFAULT RATE BY LOAN GRADE



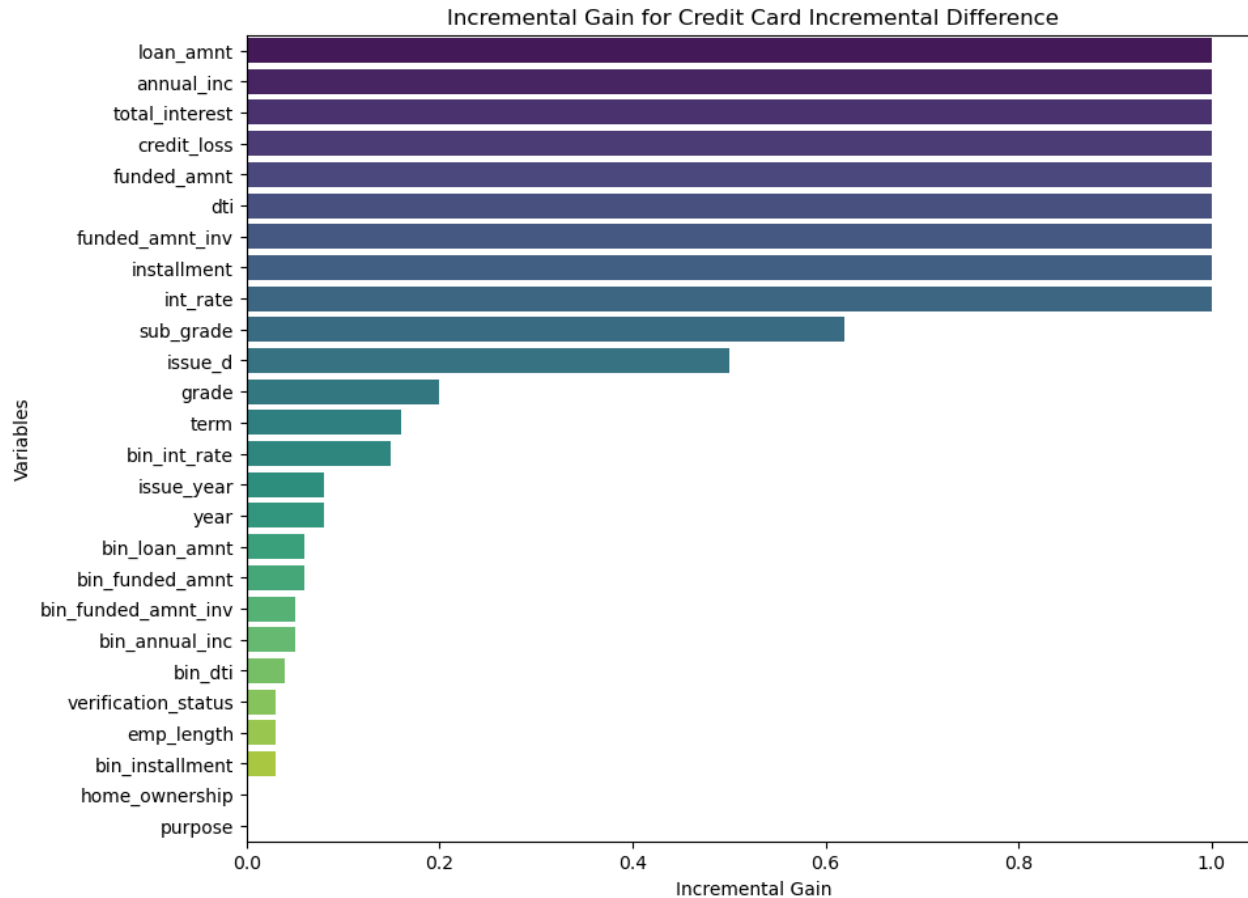
- Higher default rates observed in lower-grade loans (Grades E, F).
- Lower default rates in higher-grade loans (Grades A, B, C).

# DEFAULT RATE BY HOME OWNERSHIP



- Default rates vary based on home ownership status.
- Higher default rates among renters and those with mortgages.
- Lower default rates among homeowners.

# INCREMENTAL GAIN FOR CREDIT CARD LOANS



- Features like **int\_rate** and **annual\_inc** show high incremental gains.
- Identification of the most impactful features for credit card loan defaults.
- Helps in targeted risk management and policy adjustments.

# SUMMARY OF KEY FINDINGS

## Factors influencing default:

- Higher interest rates.
- Higher debt-to-income ratios.
- Lower annual incomes.
- Non-homeowners.
- Shorter employment lengths.
- Larger loan amounts.

# CONCLUSION

Borrowers with these characteristics have higher probabilities of default.





THANK YOU

