



Lending Loan Case study

Comprehensive Analysis and Key Insights

By Namratha Grandhi and Gaurav Sahoo

20-11-2024

Loan Data Analysis



Problem Statement

- **Objective:** Analyse loan data from 2007 to 2011 to identify trends, key factors influencing loan defaults, and actionable insights. Analyse the given loan dataset to gain insights into key factors that influence loan defaults, loan approvals, or other relevant metrics.
- **Business Context:** Understanding loan data can help financial institutions improve decision-making processes, mitigate risks, and optimize lending practices.
- **Description:** The dataset includes detailed loan data from 2007 to 2011.
- **Key Features:** Mention critical columns (e.g., loan_amnt, int_rate, grade, loan_status).

A	1 Description
1	LoanStatNew
2	acc_now_delinq The number of accounts on which the borrower is now delinquent.
3	acc_open_past_24mths Number of trades opened in past 24 months.
4	addr_state The state provided by the borrower in the loan application
5	all_util Balance to credit limit on all trades
6	annual_inc The self-reported annual income provided by the borrower during registration.
7	annual_inc_joint The combined self-reported annual income provided by the co-borrowers during registration.
8	application_type Indicates whether the loan is an individual application or a joint application with two co-borrowers.
9	avg_cur_bal Average current balance of all accounts
10	bc_open_to_buy Total open to buy on revolving bankcards.
11	bc_util Ratio of total current balance to high credit/credit limit for all bankcard accounts.
12	chargeoff_within_12_mths Number of charge-offs within 12 months
13	collection_recovery_fee post charge off collection fee
14	collections_12_mths_ex_med Number of collections in 12 months excluding medical collections



Project Overview

The primary goals of this project are:

1. Data Cleaning and Preprocessing:

1. Identify and address missing values, duplicate records, and irrelevant columns.
2. Standardize data formats by removing unnecessary suffixes and ensuring uniformity across numerical fields.

2. Exploratory Data Analysis (EDA):

1. Understand the relationships between key financial variables, such as loan amounts, interest rates, and debt-to-income ratios.
2. Visualize distributions and correlations using tools like correlation matrices and plots.

3. Insights and Recommendations:

1. Derive actionable insights from the analysis to improve loan processing, risk assessment, and portfolio management.

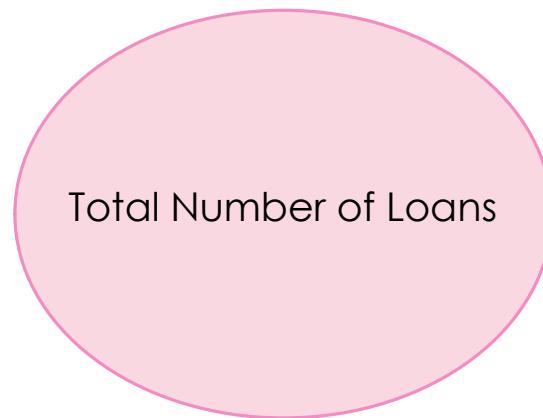
This project emphasizes clear and structured data handling, enabling more accurate analysis and decision-making in the financial domain.

• Dataset Overview and Initial Data Exploration •



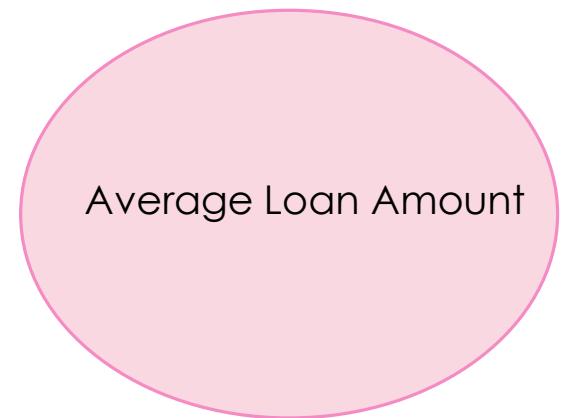
Total Loan Amount
\$445,602,650

The overall sum of loans issued



Total Loans
39717

The total count of loans provided



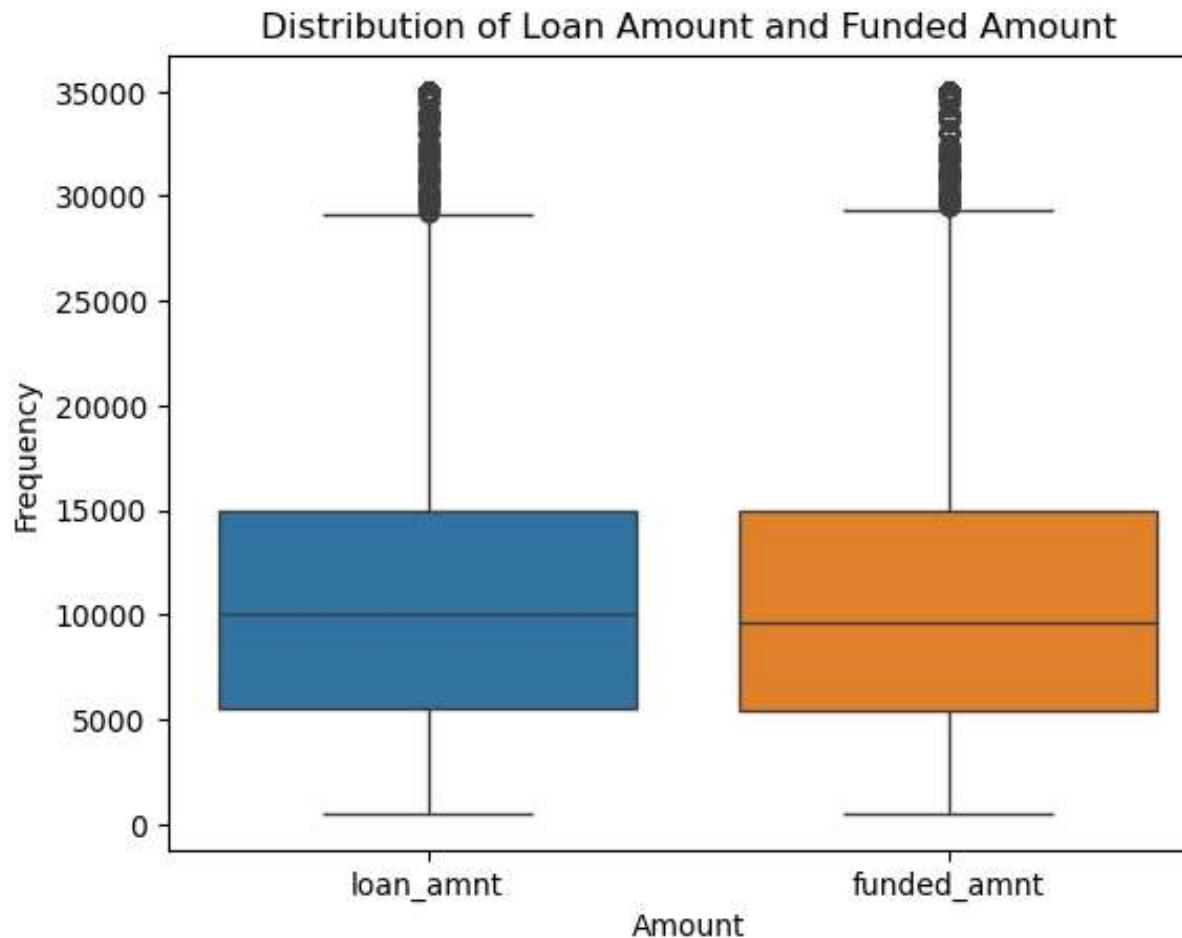
Average Loan Amount
\$ 11,219.44

Mean Value of loans

• Data Cleaning and Preprocessing •

- Several columns consist entirely of either NaN values or zeros, and these can be identified for removal or handling.
- The loan term column includes a suffix indicating months, which can be stripped for cleaner data processing.
- Similarly, the interest rate column includes a suffix indicating %, which can be stripped as well.
- There are a few irrelevant columns, such as id, member_id, emp_title, title, url, and zip_code, which can be dropped as they are not required for the analysis.

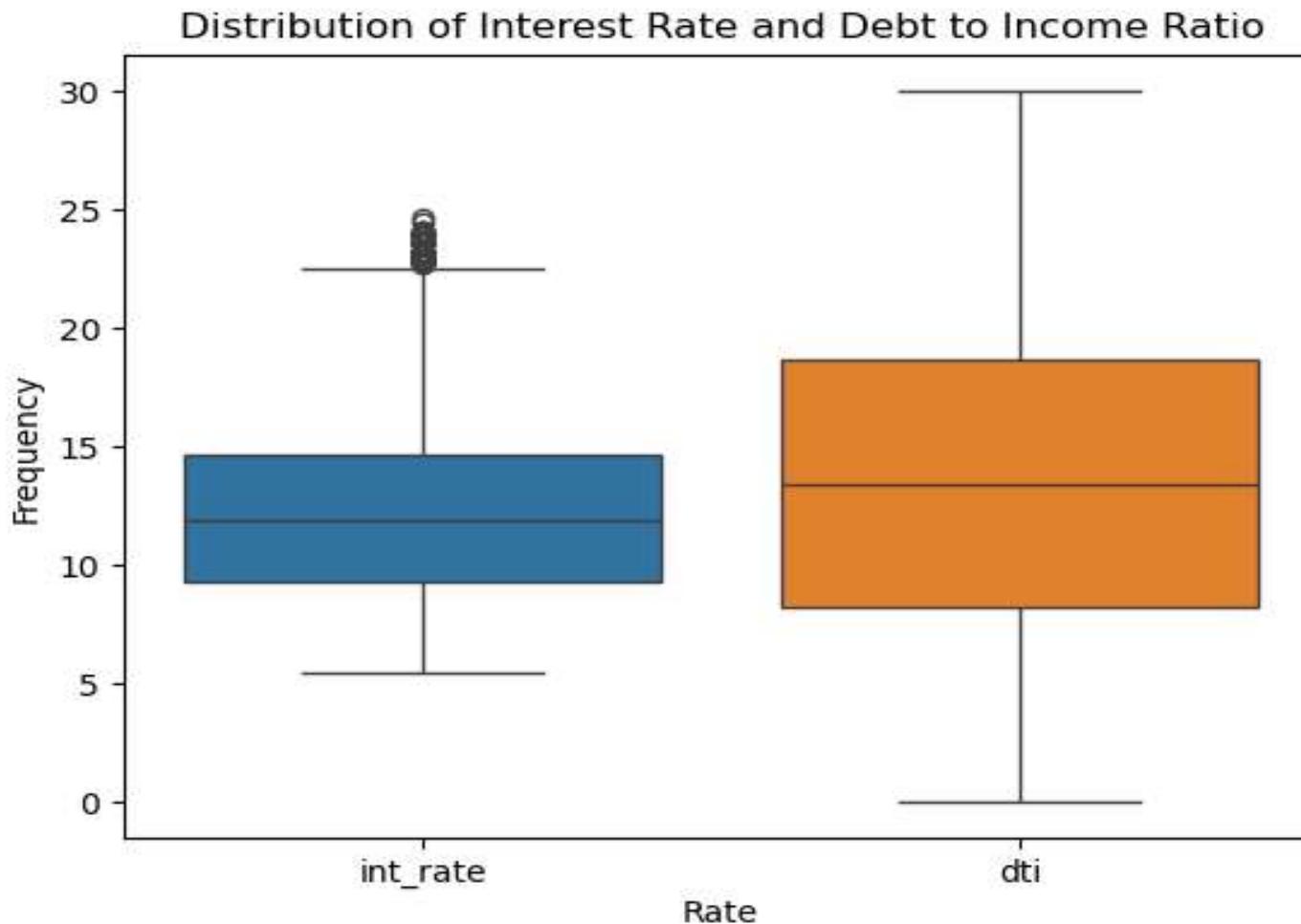
Univariate Analysis



ANALYSIS:

- The loan amount is mostly between 5500 to 15000 where outliers are mostly above 30000.
- The funded amount is mostly between 5400 to 15000 where outliers are mostly above 30000.

Univariate Analysis

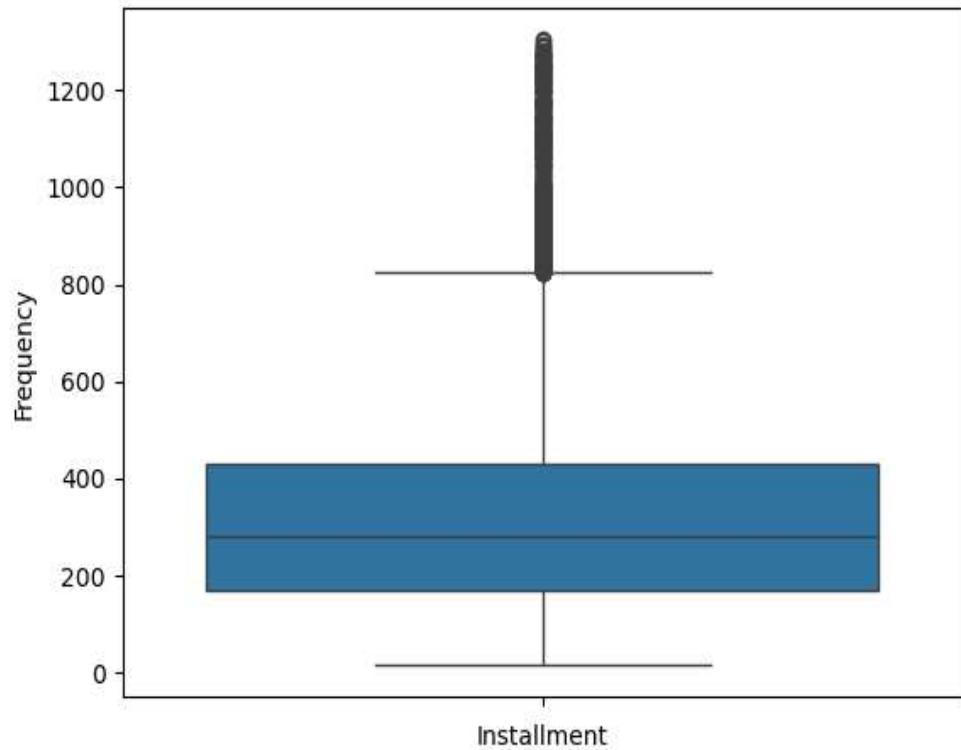


ANALYSIS:

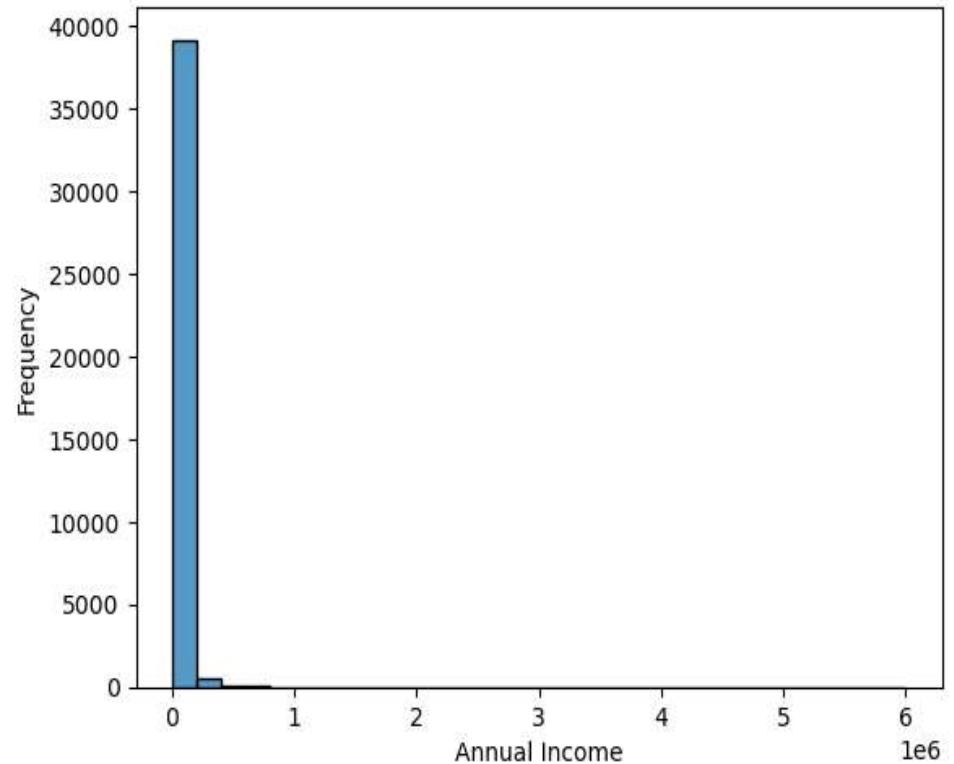
- The interest rate is mostly between 9.25% to 14.59% where outliers are above 22.5%.
- The Dept to income ratio is mostly between 8.17% to 18.6%

Univariate Analysis

Distribution of Installment



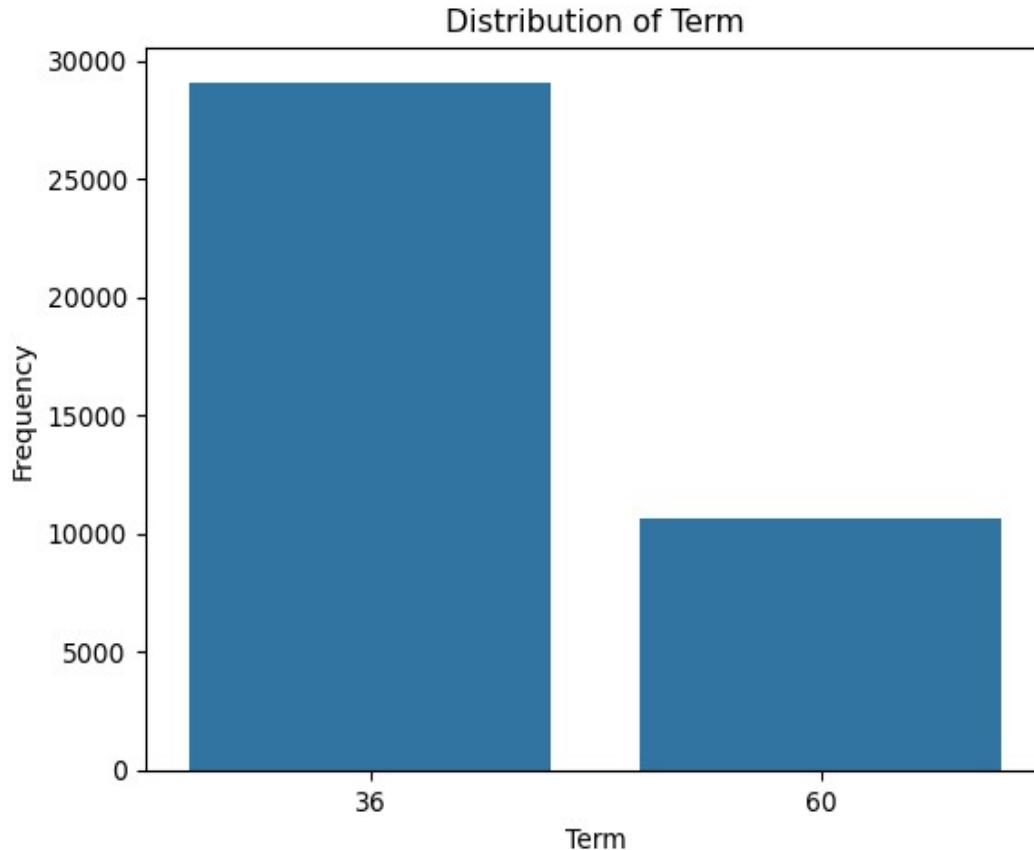
Distribution of Annual Income



ANALYSIS:

- The installment is mostly between 167.02 to 430.78 with outliers are above 800.
- Most loan applicants has annual income below 100,000

Univariate Analysis

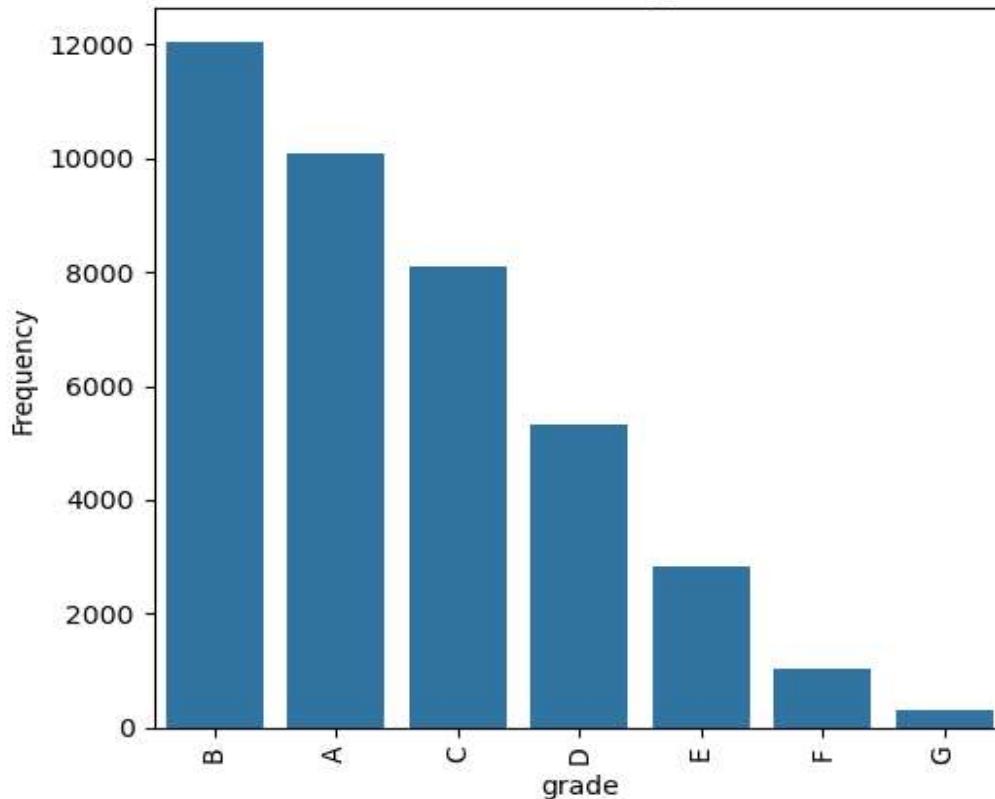


ANALYSIS:

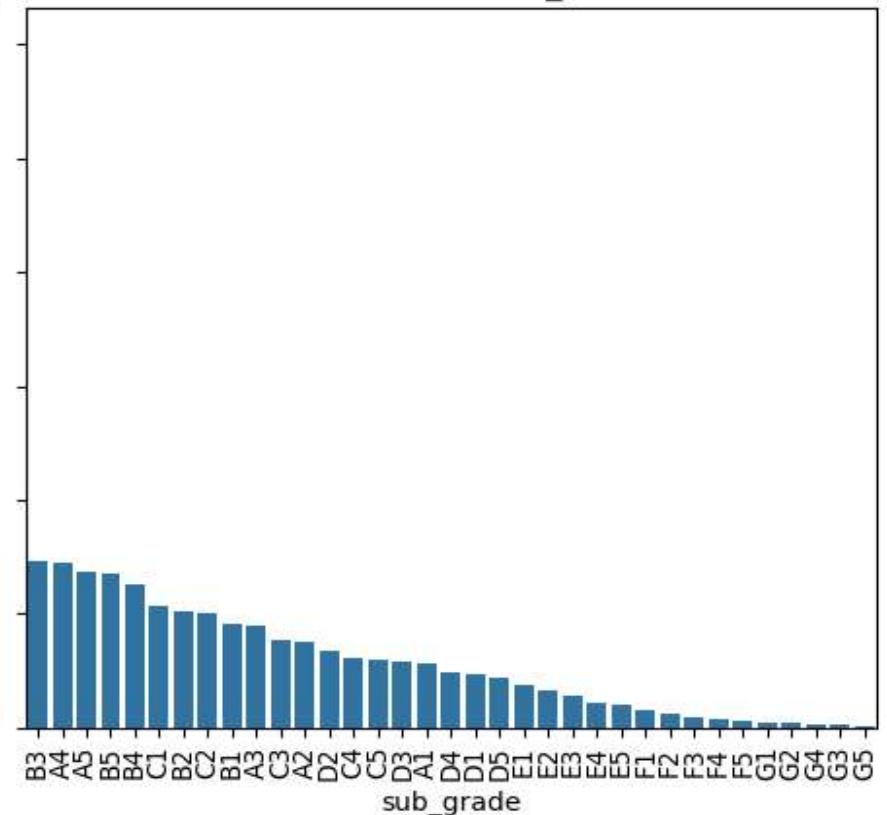
- ❑ Loan taken for 3 years is almost thrice the loan taken for 5 years. So mostly 3 years loan is preferred.

Univariate Analysis

Distribution of grade



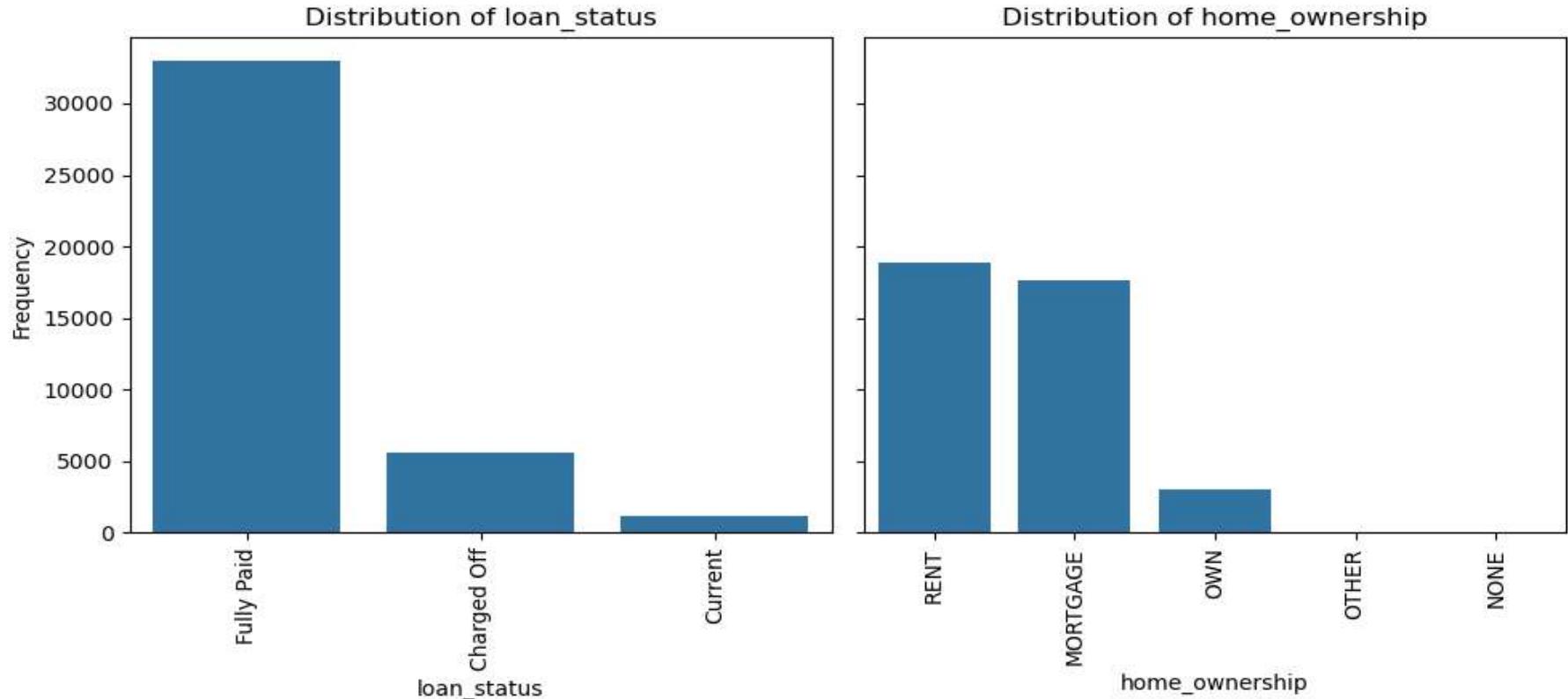
Distribution of sub_grade



ANALYSIS:

- The majority of loans are taken by borrowers with Grade B, A, and C, with the highest concentrations observed in sub-grades B3, A4, A5, B5, B4, and C1.

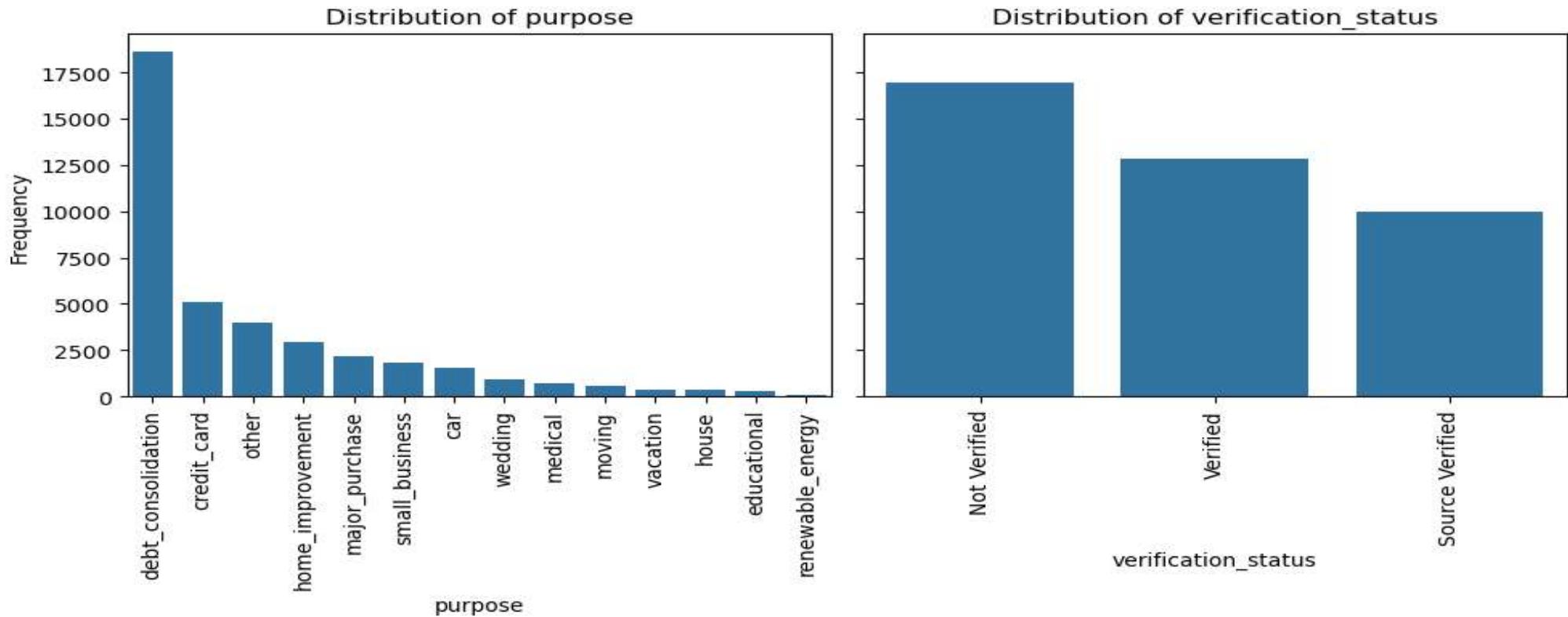
Univariate Analysis



ANALYSIS:

- The number of charged-off loans is approximately one-sixth of the total number of fully paid loans, indicating a significant proportion of defaults
- A substantial number of loans are taken by individuals who rent or mortgage their homes, suggesting that homeownership status is a key factor for loan applicants.

Univariate Analysis

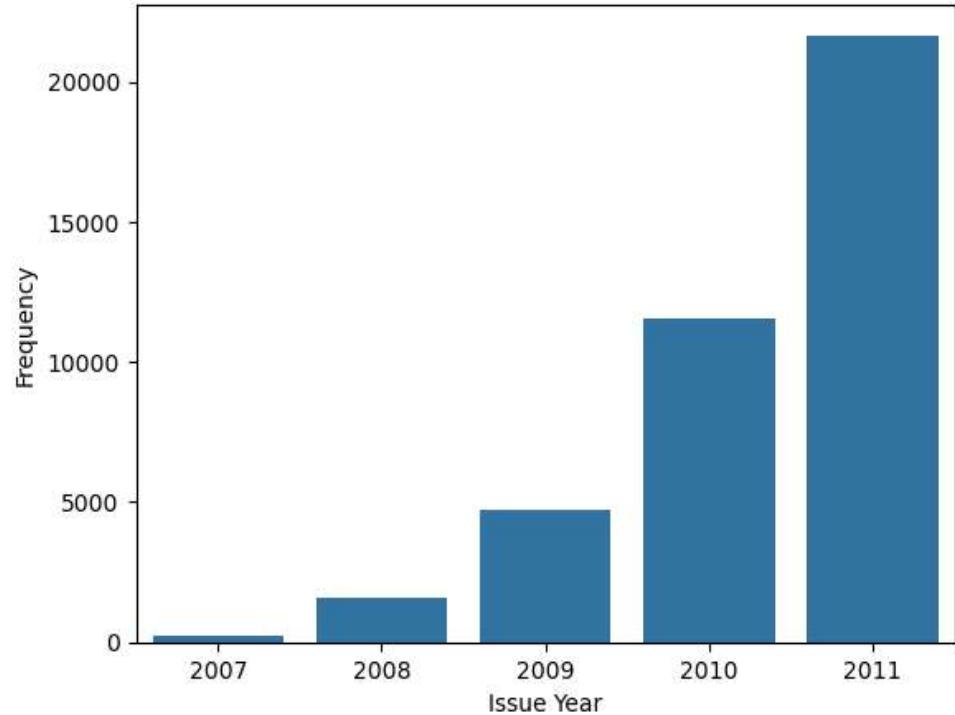


ANALYSIS:

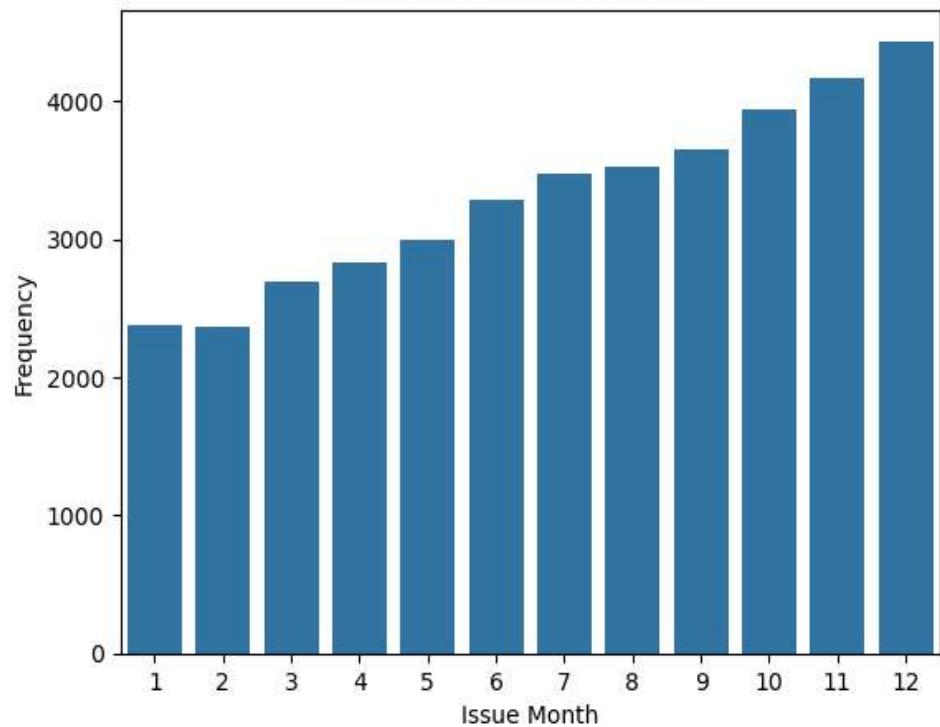
- ❑ The most common reason for taking out loans is debt consolidation, reflecting a trend of borrowers seeking to streamline or reduce existing debt obligations.
- ❑ Majority of loans are granted to individuals whose income source is not verified, potentially indicating more lenient lending practices or reliance on other factors for loan approval

Univariate Analysis

Distribution of Issue Year



Distribution of Issue Month

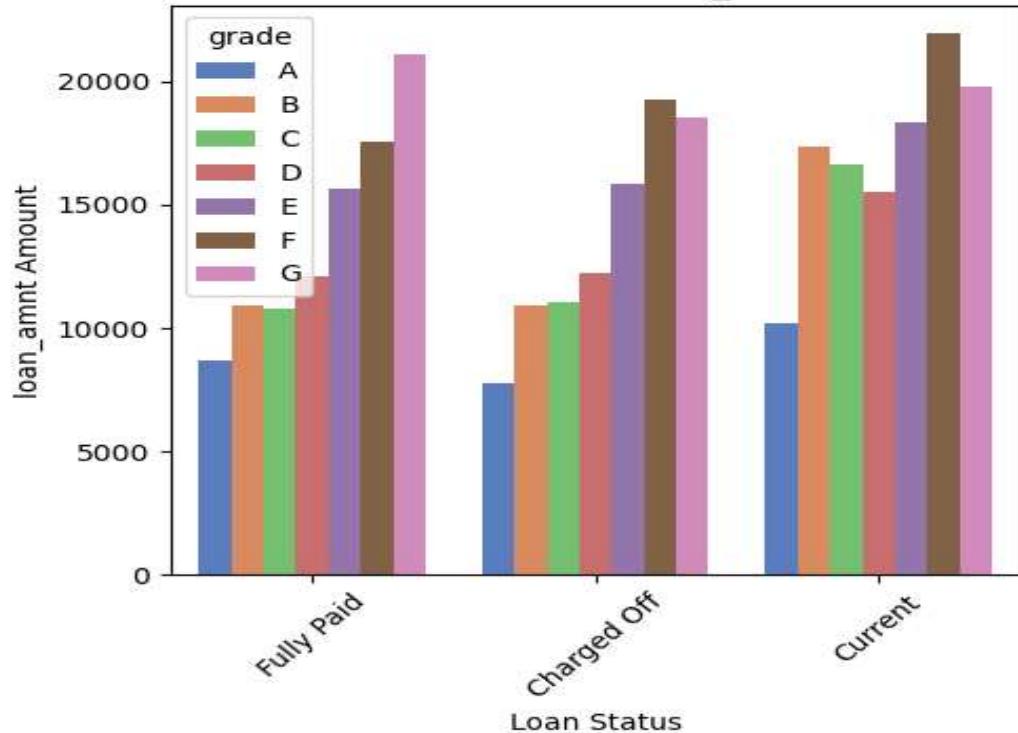


ANALYSIS:

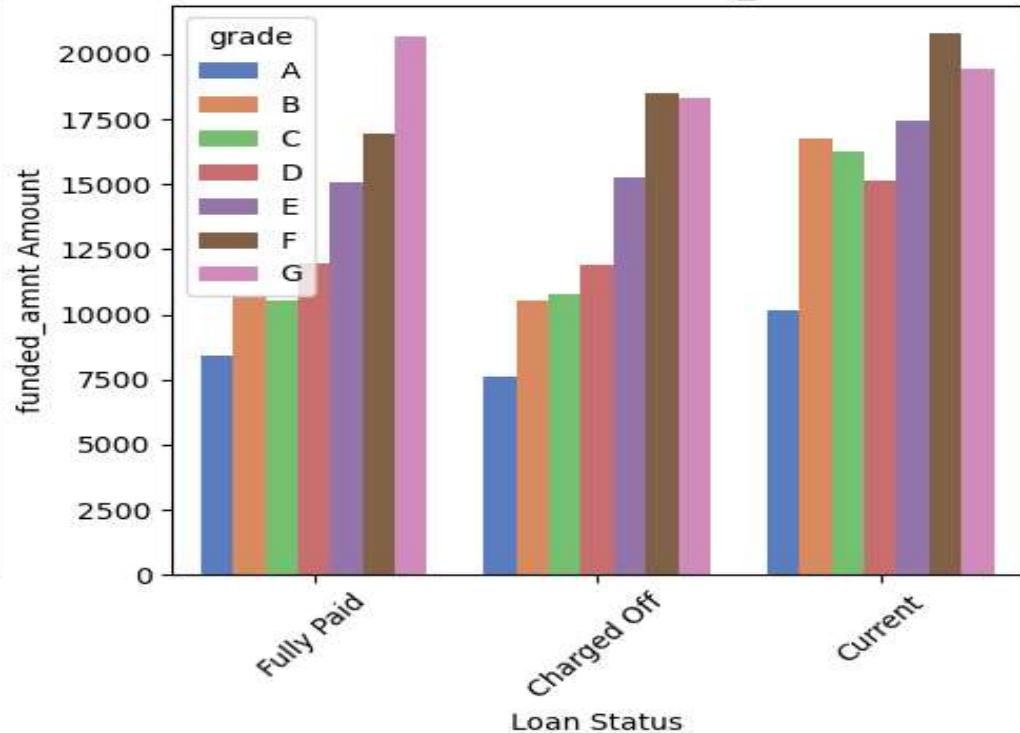
- The number of loans issued has consistently increased year over year, reflecting a growing demand for loans over time.
- Loan issuance shows an upward trend within each year, with the loan count increasing steadily from January to December.

Bivariate/Multivariate Analysis

Loan Status vs loan_amnt



Loan Status vs funded_amnt

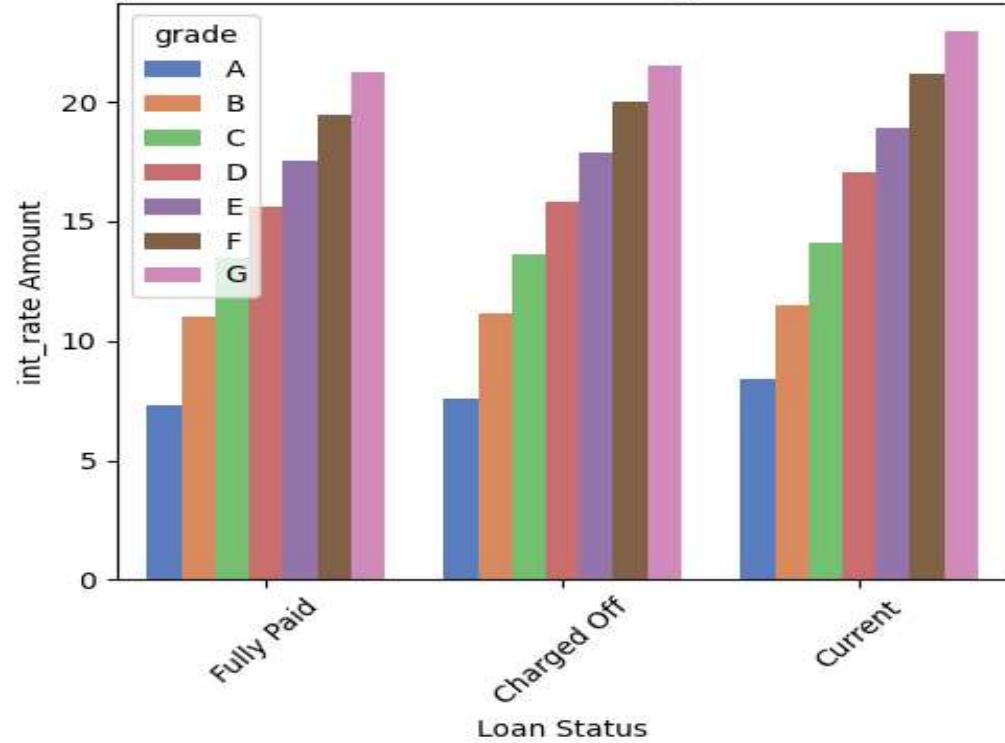


ANALYSIS:

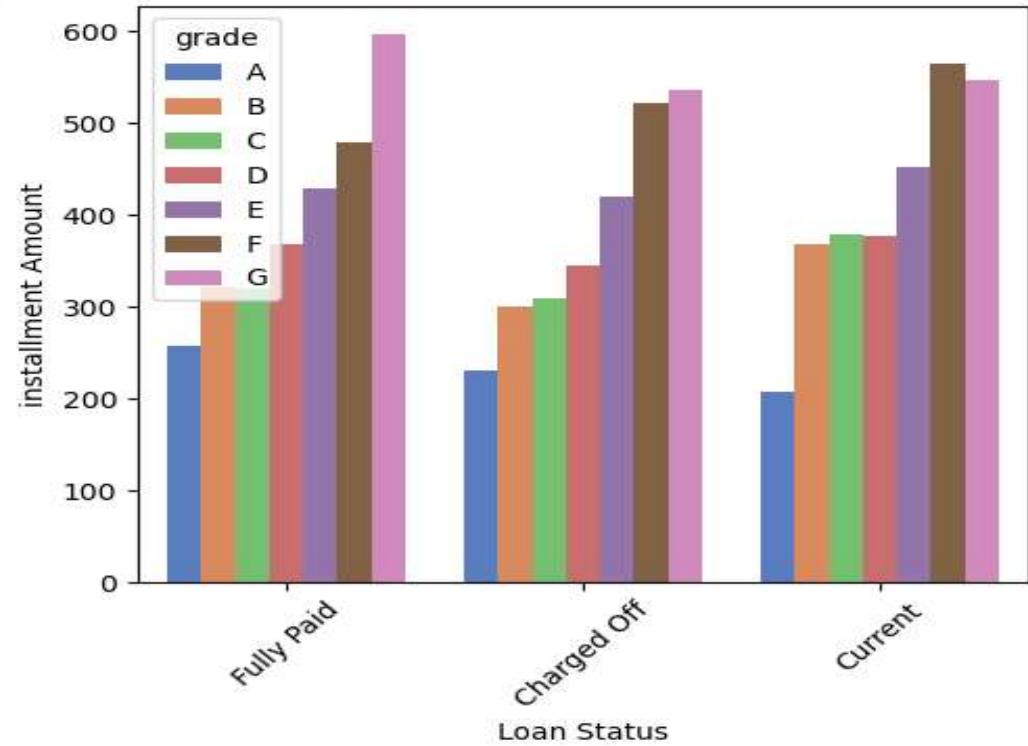
- **Loan Amount vs Loan Status:** Higher loan grades (e.g., A and B) are associated with smaller loan amounts, likely due to better creditworthiness.
- **Funded Amount vs Loan Status:** Lower-grade borrowers tend to receive larger funded amounts, potentially increasing the risk of default.

Bivariate/Multivariate Analysis

Loan Status vs int_rate



Loan Status vs installment

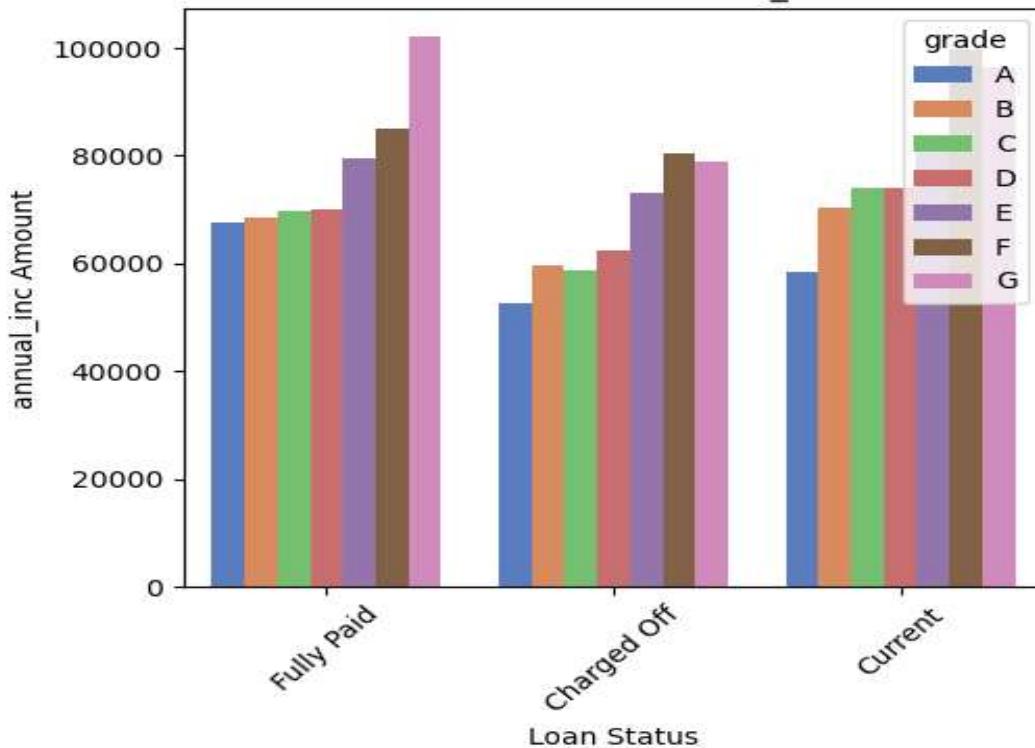


ANALYSIS:

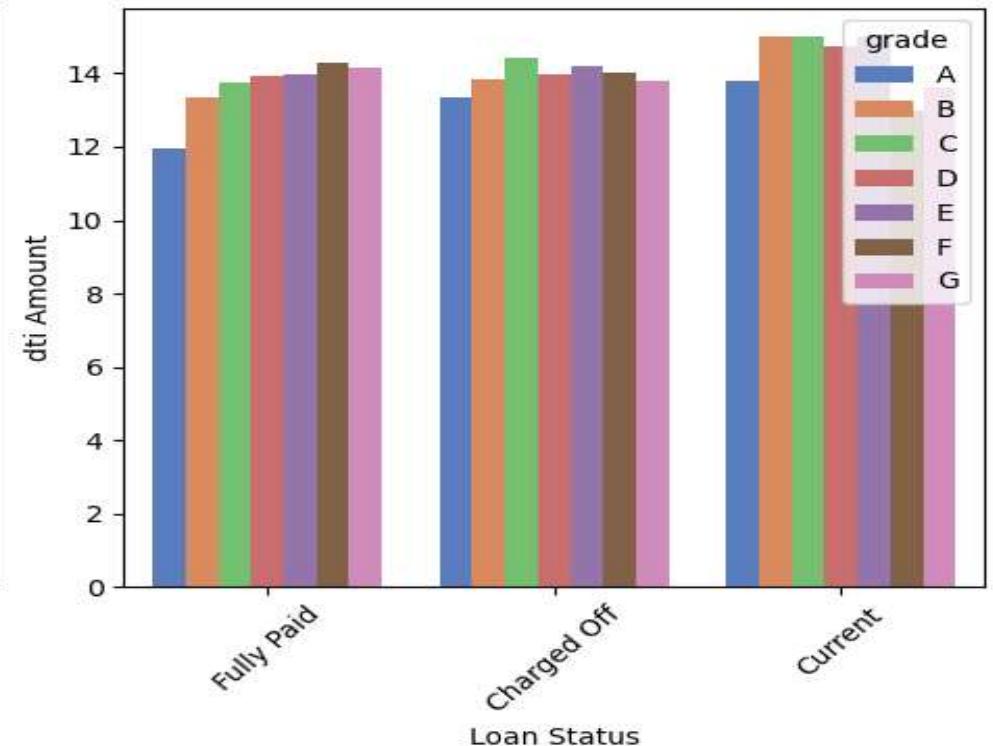
- ❑ **Interest Rate vs Loan Status:** Higher interest rates in lower grades may contribute to higher Charged Off.
- ❑ **Installment vs Loan Status:** Since installments increase with grade, higher installments, especially in lower-grade loans, may increase likelihood of Charged Off.

Bivariate/Multivariate Analysis

Loan Status vs annual_inc



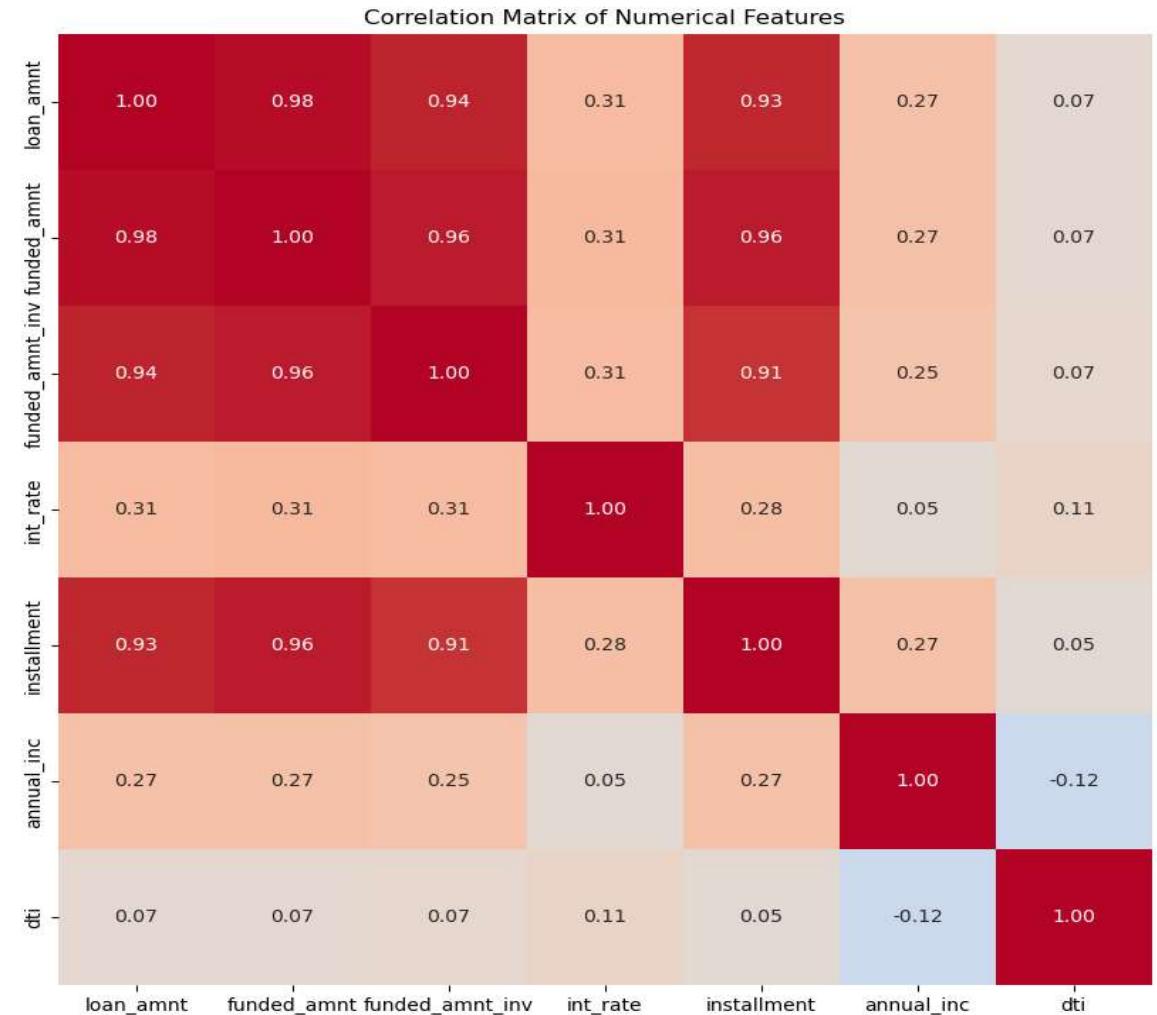
Loan Status vs dti



ANALYSIS:

- **Annual Income vs Loan Status:** Borrowers with higher incomes but lower grades are more likely to default or keep loans in the Current status.
- **Debt to Income Ratio vs Loan Status:** Lower-grade loans are riskier due to high DTI ratios, as borrowers are likely over-leveraged.

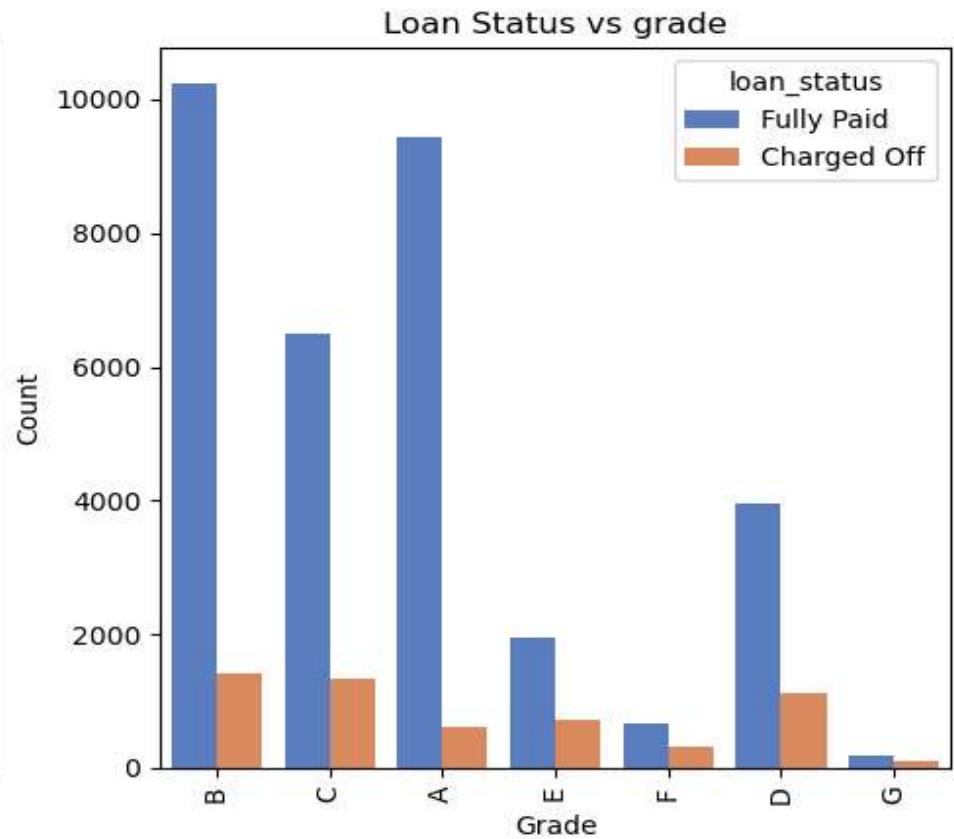
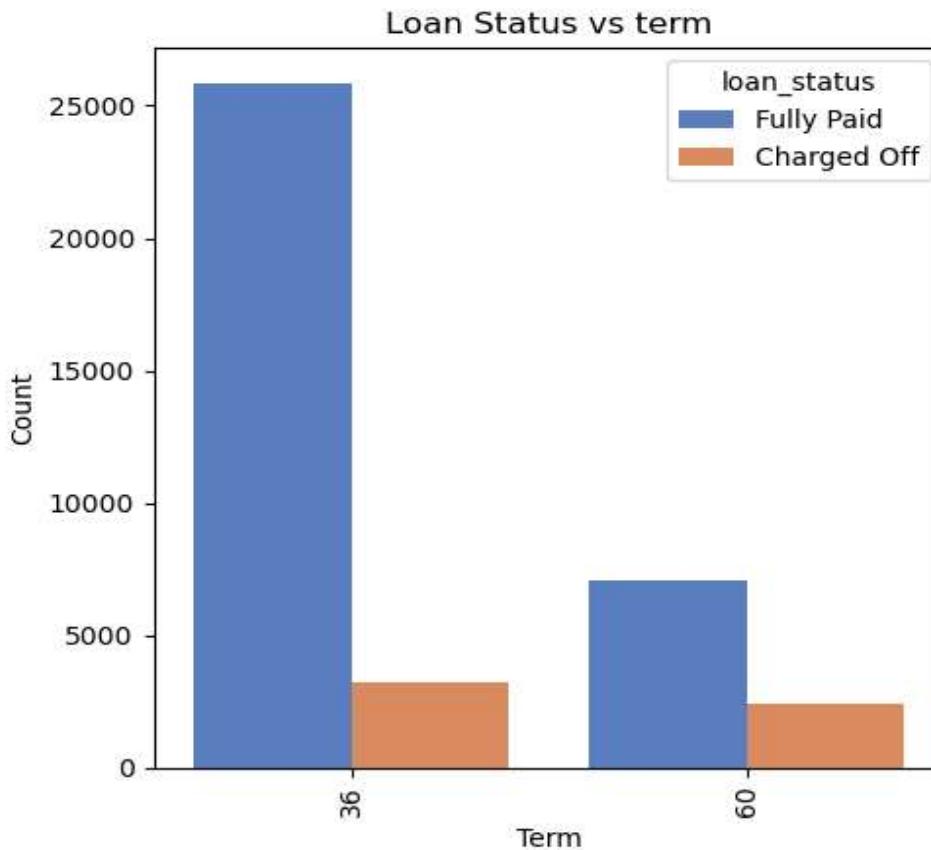
Bivariate/Multivariate Analysis



ANALYSIS:

- ❑ **Loan Amount and Funded Amount (0.98):** These two variables are almost perfectly correlated, since the funded amount is closely tied to the loan amount approved.
- ❑ **Loan Amount and Installment (0.93):** Larger loan amounts naturally result in higher installment payments.
- ❑ **Funded Amount and Installment (0.96):** Funded amounts also heavily influence installment size.
- ❑ **Funded Amount and Funded Amount Inv (0.96):** Funded amount and Funded amount by investors are also highly correlated and directly proportional to each other.

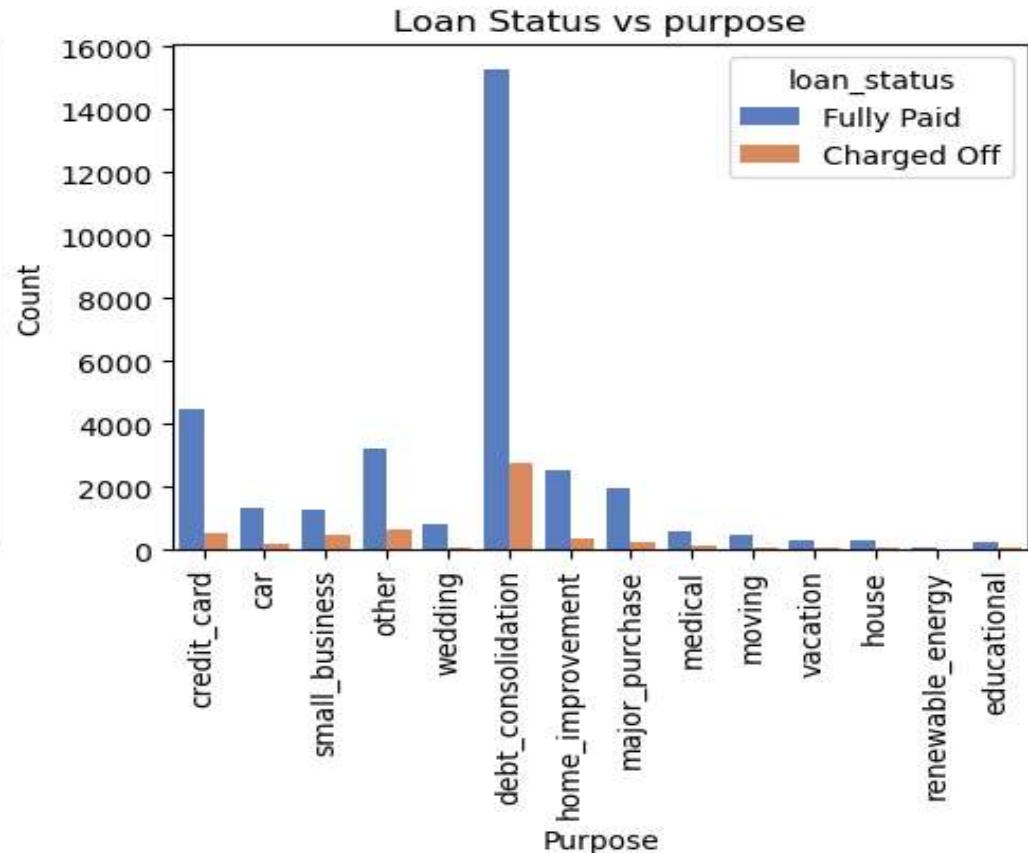
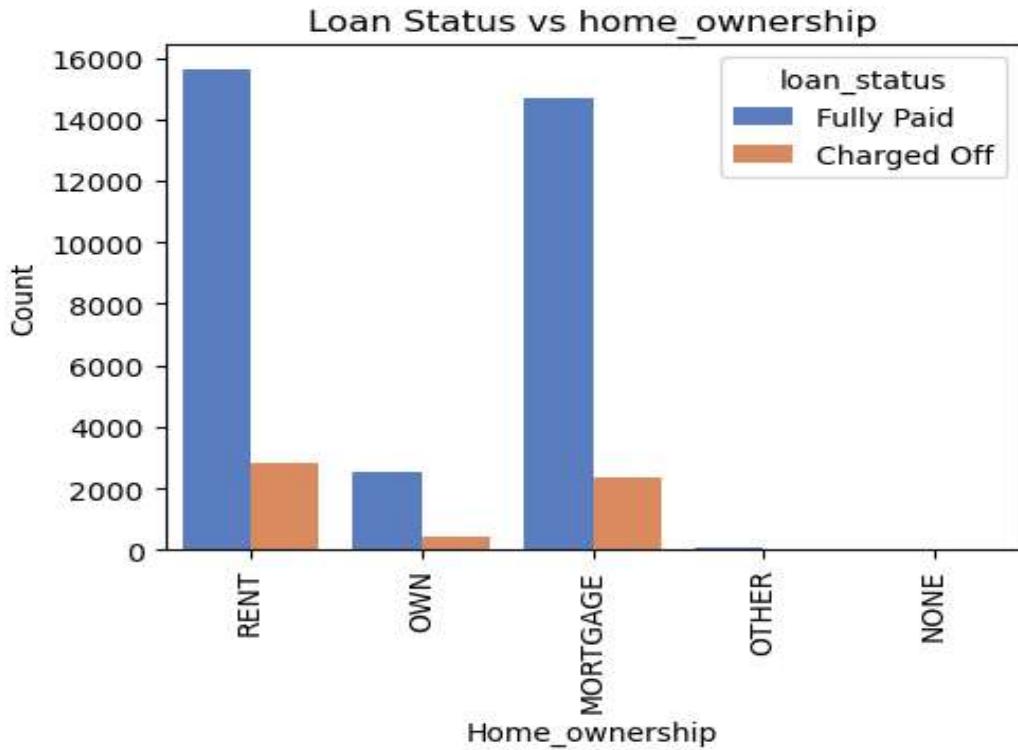
Bivariate/Multivariate Analysis



ANALYSIS:

- Term vs Loan Status:** Loans with a 36-month term are less likely to default than 60-month term loans.
- Grade vs Loan Status:** A-grade loans have a very low risk compared to the other grades.

Bivariate/Multivariate Analysis

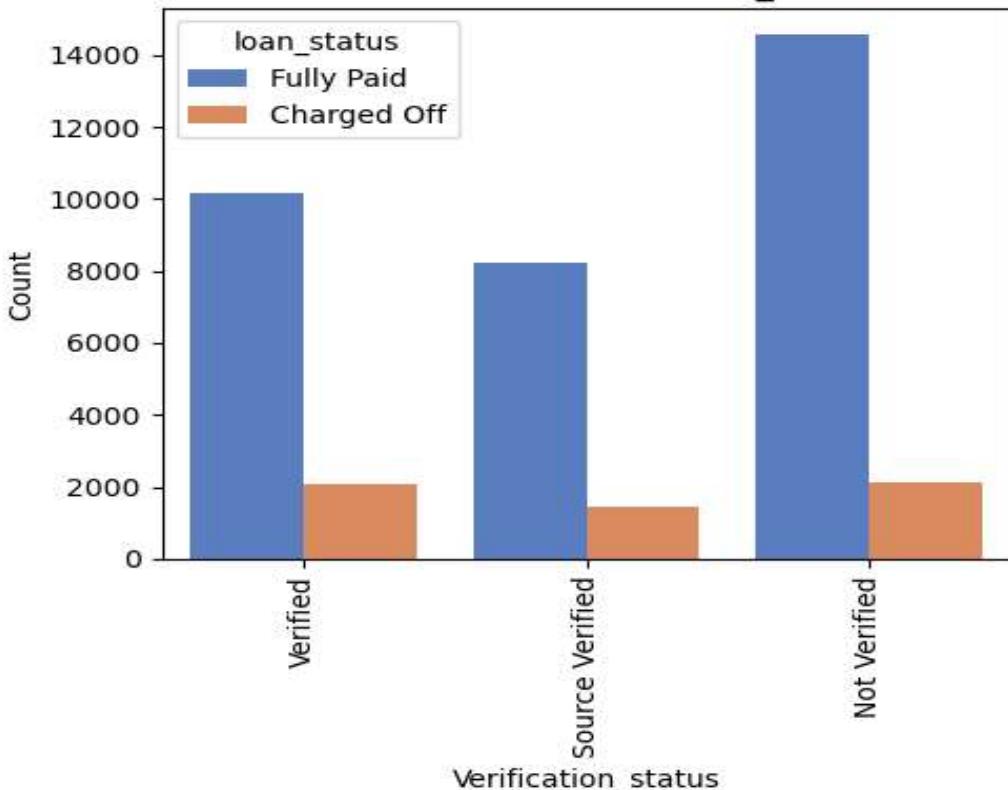


ANALYSIS:

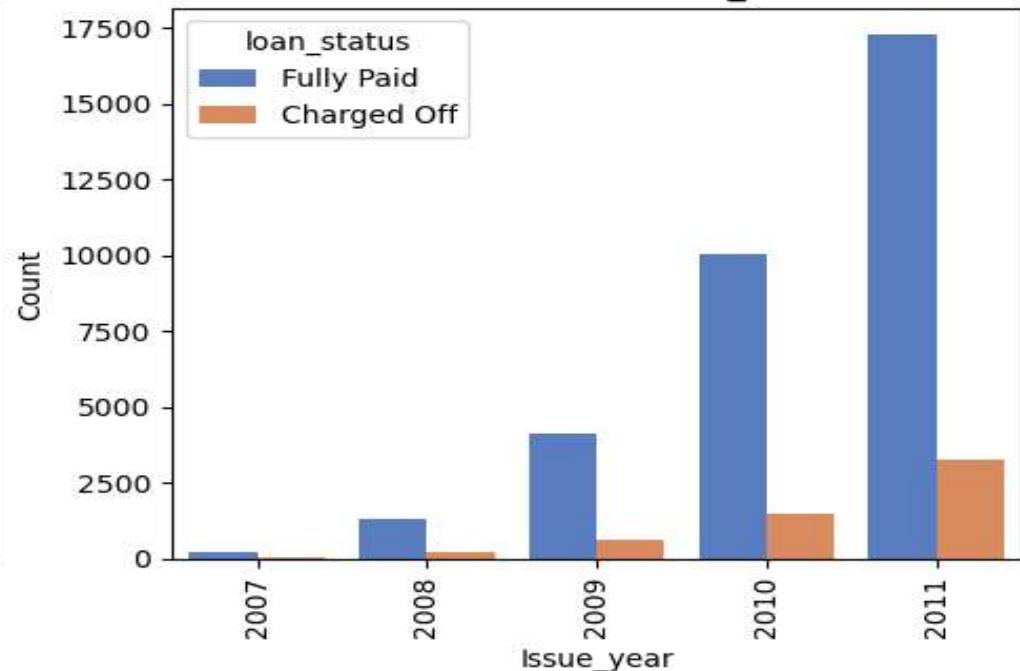
- **Home Ownership vs Loan Status :** Homeowners are more likely to fully repay loans than renters or mortgagers .
- **Purpose vs Loan Status:** Loans for debt consolidation are most commonly repaid, but most defaulters are seen in the same category as well.

Bivariate/Multivariate Analysis

Loan Status vs verification_status



Loan Status vs issue_year



ANALYSIS:

- ❑ **Verification Status vs Loan Status:** Surprisingly loans without verified income are less likely to default.
- ❑ **Issue Year vs Loan Status:** Fully paid loans increased every year, but the defaulters increased as well.

Overall Analysis

Risk Factors:

Loans with lower grades (E, F, G) tend to have higher loan amounts, interest rates, and DTI ratios, increasing the likelihood of default.

Income Influence:

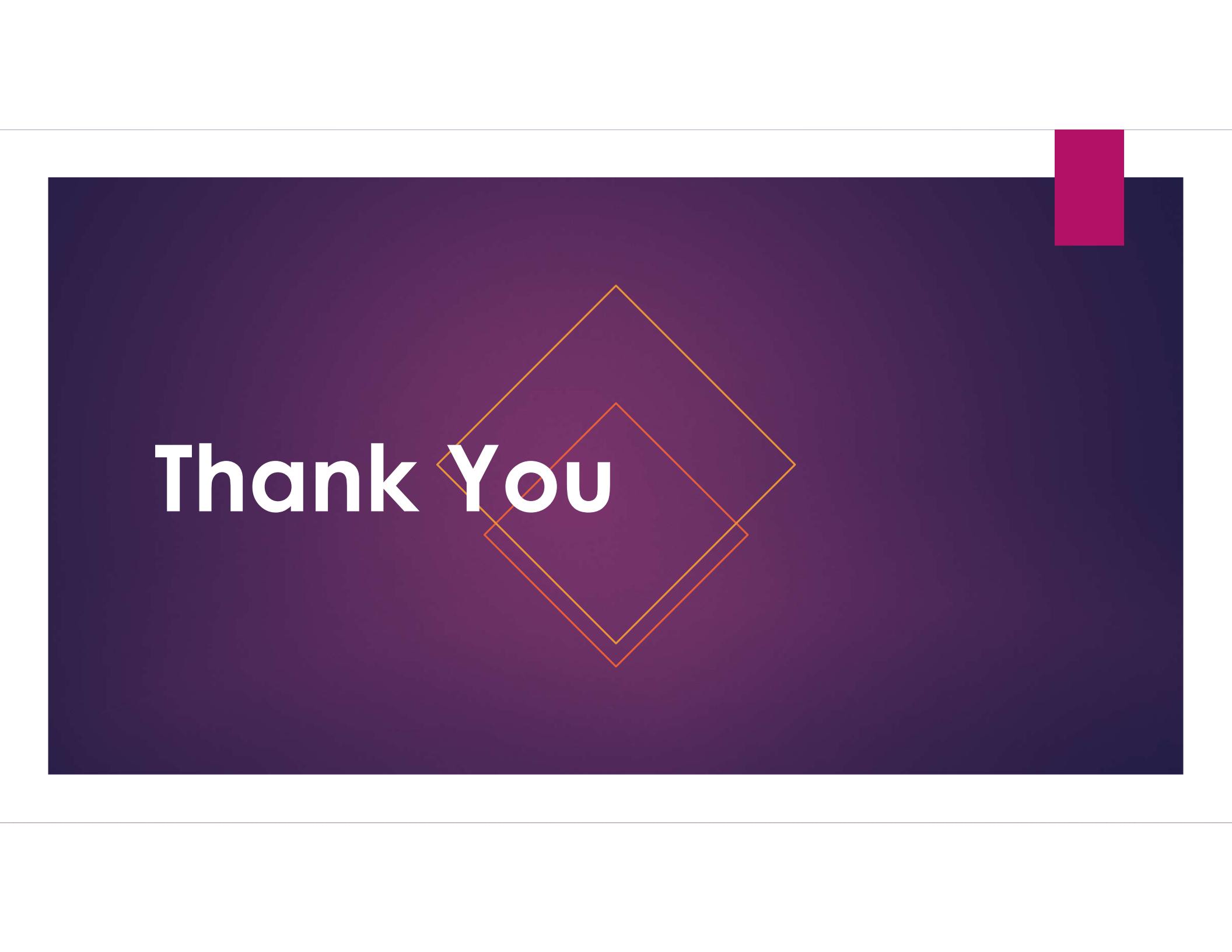
Borrowers with higher incomes are more likely to fully repay loans.

Current Loans:

These loans often have higher loan amounts, funded amounts, and installments, particularly in lower grades, indicating they are still at risk of default.

Key Findings:

Shorter Terms, Higher Grades, Home Ownership, and Debt Consolidation: Loans with these factors are more likely to be fully paid.



Thank You

