LENDING CLUB CASE STUDY

Analysis by
-Ranjith Kumar J
-Sakshi Singh

AGENDA

- Problem Statement
- Data Understanding
- Data Cleaning and Manipulation
- Data Analysis & Summary
  - Univariate Analysis
  - Bivariate Analysis
  - Correlation Analysis
- Conclusion

PROBLEM STATEMENT

Lending club faces significant financial losses due to credit loss, which occurs when 'risky' borrowers, identified as 'charged-off' defaulters, refuse to repay or abscond with loaned funds. This challenge necessitates a stringent evaluation of loan applications to mitigate the risk of default.

DATA UNDERSTANDING

The dataset loan.csv includes information about past loan applicants and their default status. The goal is to identify patterns that predict the likelihood of default. This analysis can inform decisions such as denying loans, reducing loan amounts, or charging higher interest rates to risky applicants.

LEADING ATTRIBUTE

Fully paid - applicant has fully paid the loan (the principal and the interest rate)

Current - applicant is in the process of paying the instalments, i.E. The tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

Charged-off - applicant has not paid the instalments in due time for a long period of time, i.E. He/she has defaulted on the loan

Loan Rejected - the company had rejected the loan (because the candidate does not meet their requirements etc.). Since the loan was rejected, there is no transactional history of those applicants with the company and so this data is not available with the company (and thus in this dataset)

DATA CLEANING & MANIPULATION

- Read CSV file
- Identify & drop-off duplicate Rows & Columns
- Identify & drop-off Rows & Columns with Null,NAN,NA,0 values
- Identify & drop-off descriptive and identification columns(unique fields)
- Identify if any derived columns are necessary
- Conversion of data to the appropriate data types
- Processing for outlier data
- Imputing values in Columns

UNIVARIATE ANALYSIS

Categorical Variables:

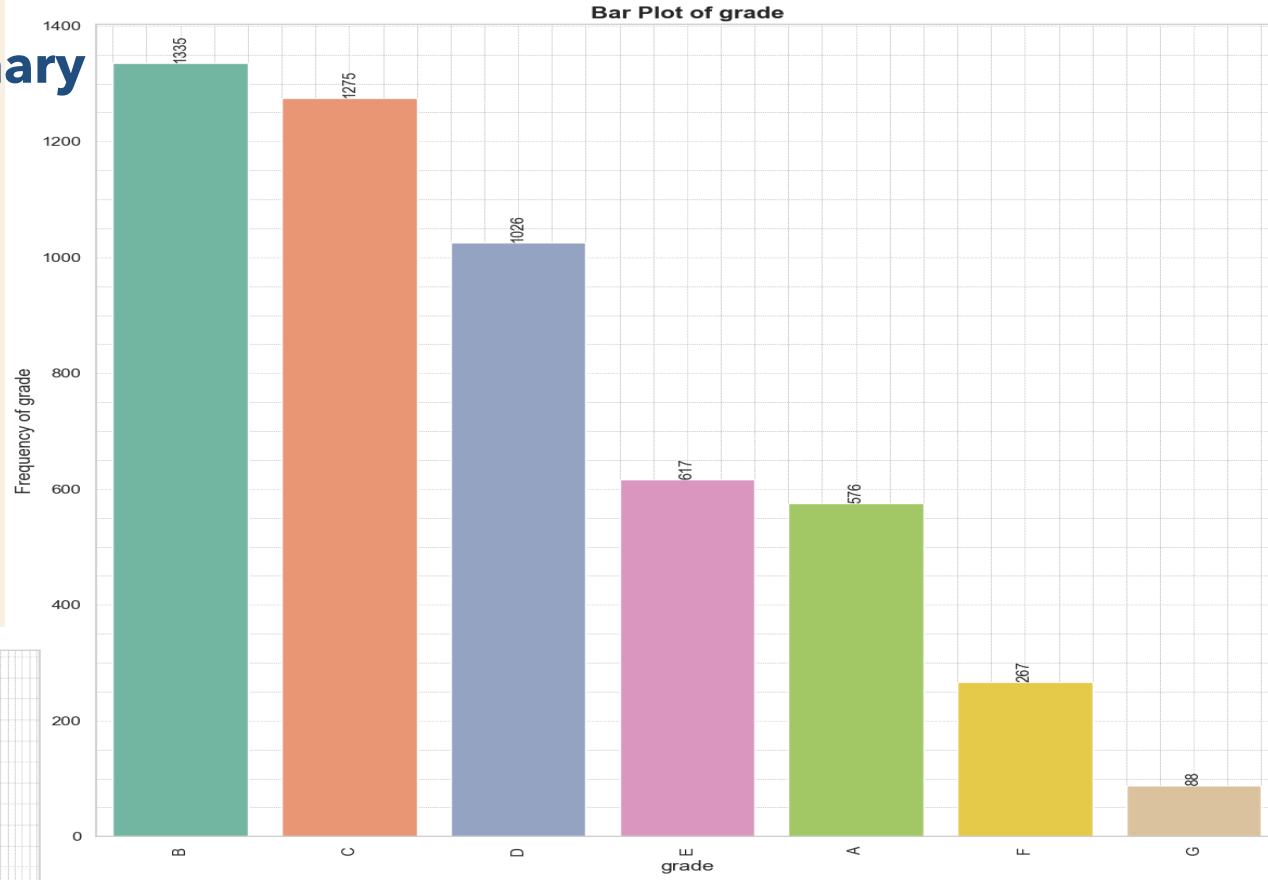
- Grade
- Sub-Grade
- Term (36/60 months)
- Employment Length
- Loan Purpose
- Home Ownership

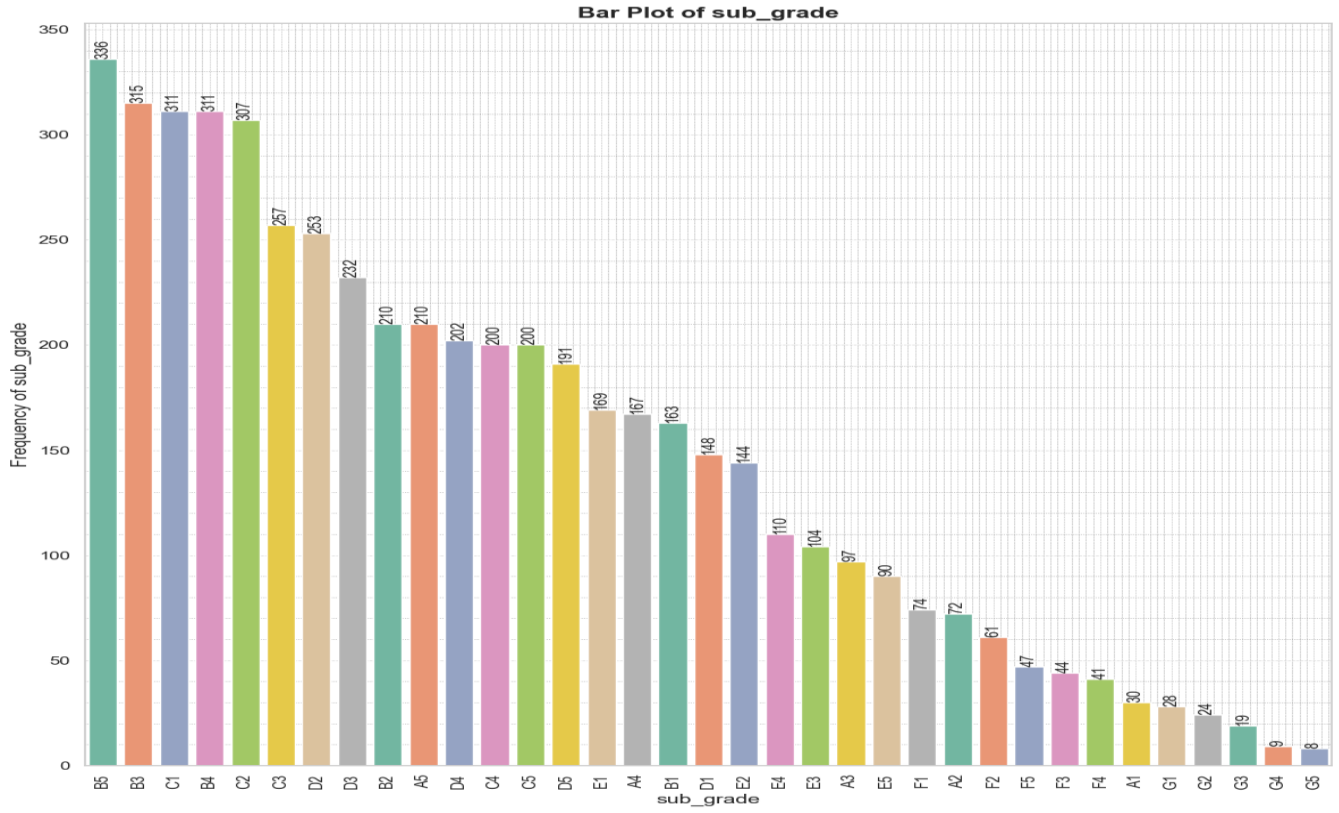
Quantitative Variables:

- Annual income
- Interest rate
- Debt to Income Ratio (DTI)
- Loan amount
- Revolving Line Utilization Rate

Univariate Analysis : Grade , Sub-Grade & Summary

- **Grade: Grade B** has the most "Charged off" loan applicants (1,335), indicating the highest default risk. **Grade G** has the fewest (88), indicating the lowest default risk

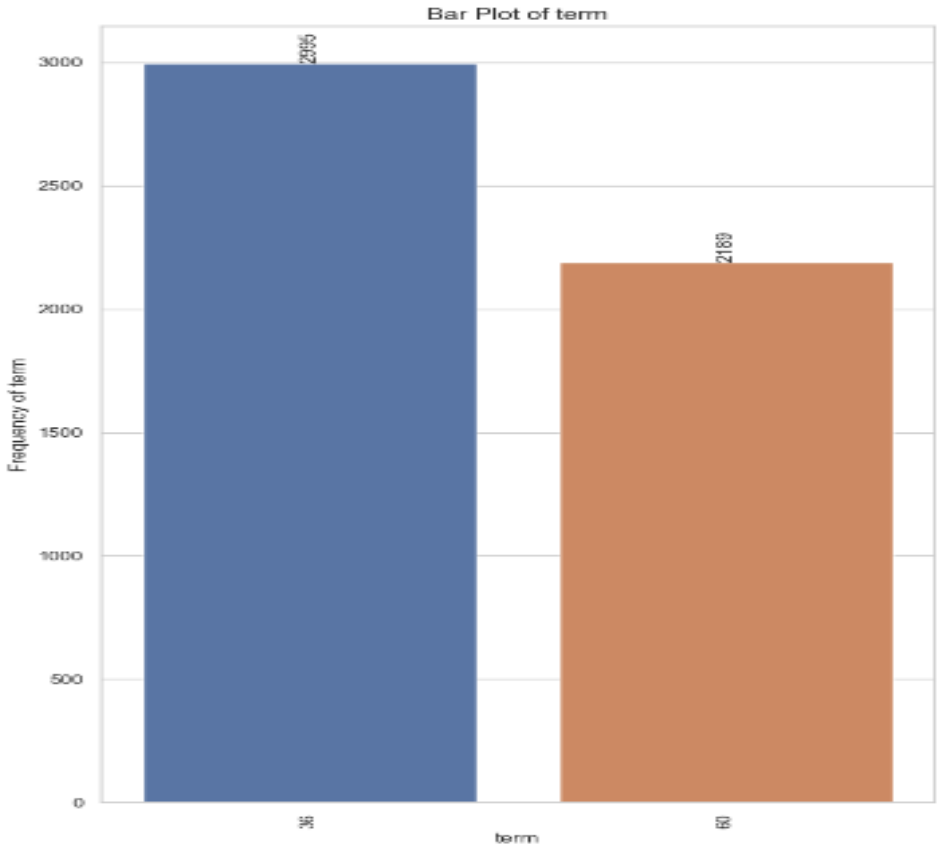


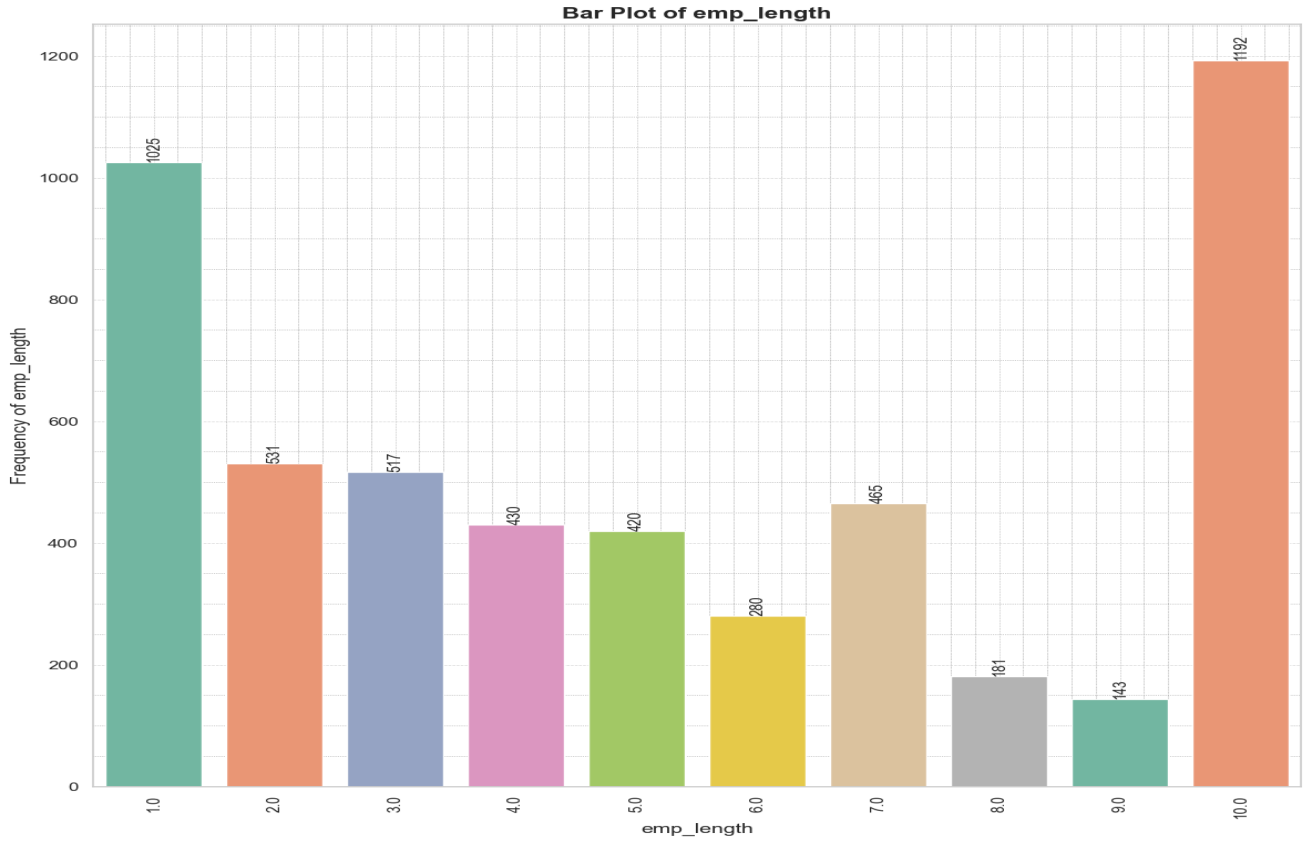


- **Sub-Grade:** Within Grade B, **B5** shows the highest default risk, while **G5** shows the lowest

Univariate Analysis : Term , Employment Length and Summary

- **Term: loans** with a **36-month** term had the highest number of defaults (2,995), suggesting shorter terms are associated with higher default rates.

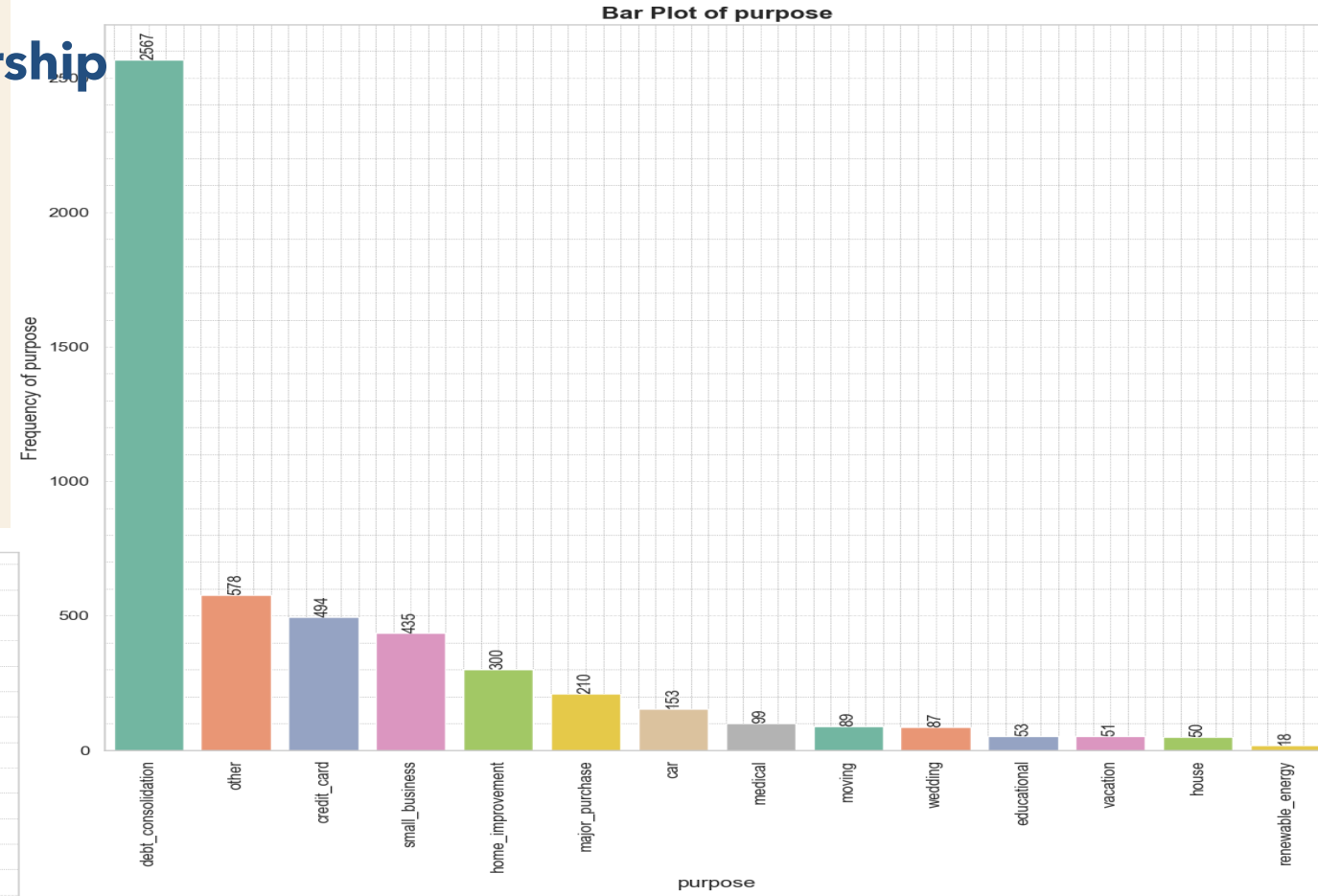


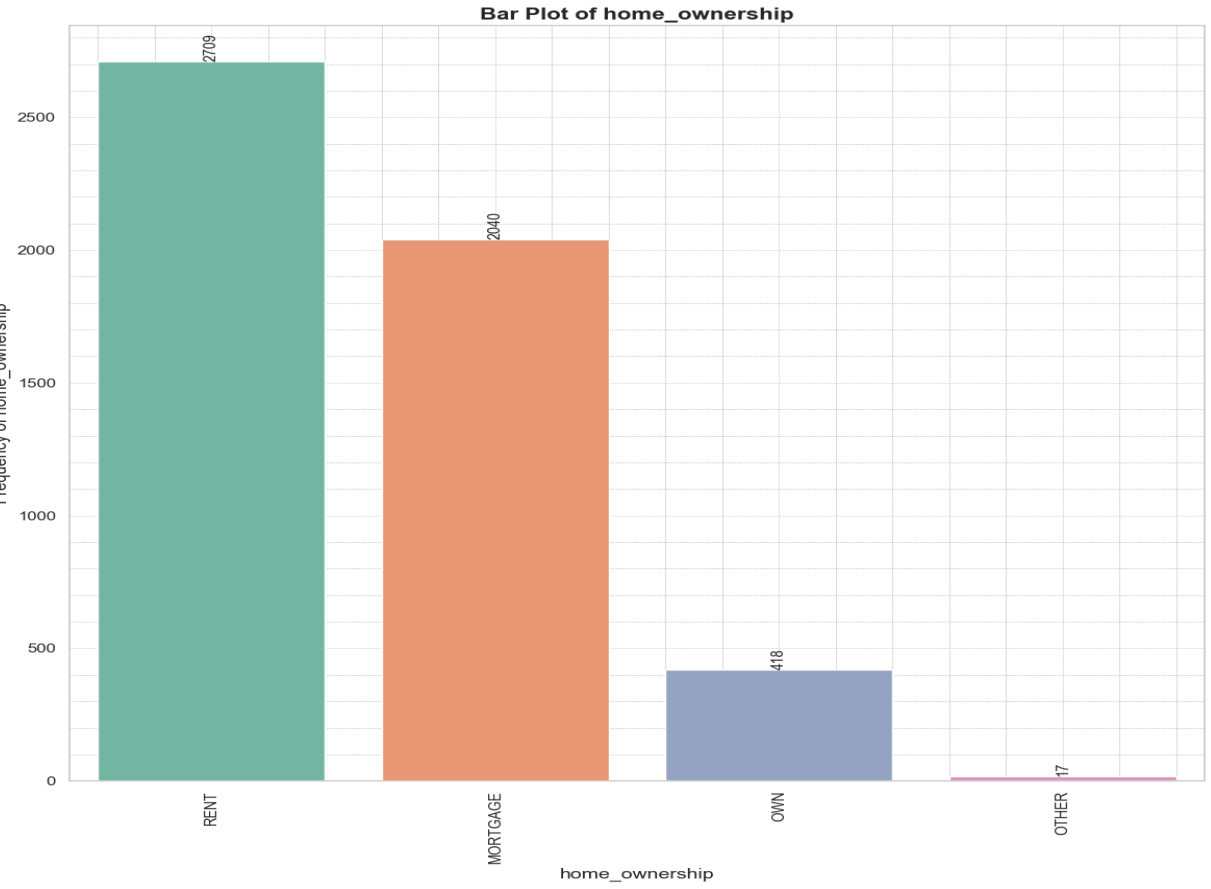


- **Employment Term:** Applicants employed for over **10 years** had the most defaults (1,192), indicating long-term employment does not guarantee repayment success.

Univariate Analysis : Purpose , Home Ownership & Summary

- **Purpose : debt consolidation** was the top reason for default (2,567 applicants). The lending company should be cautious with these loans, as they are often used to pay off other debts.



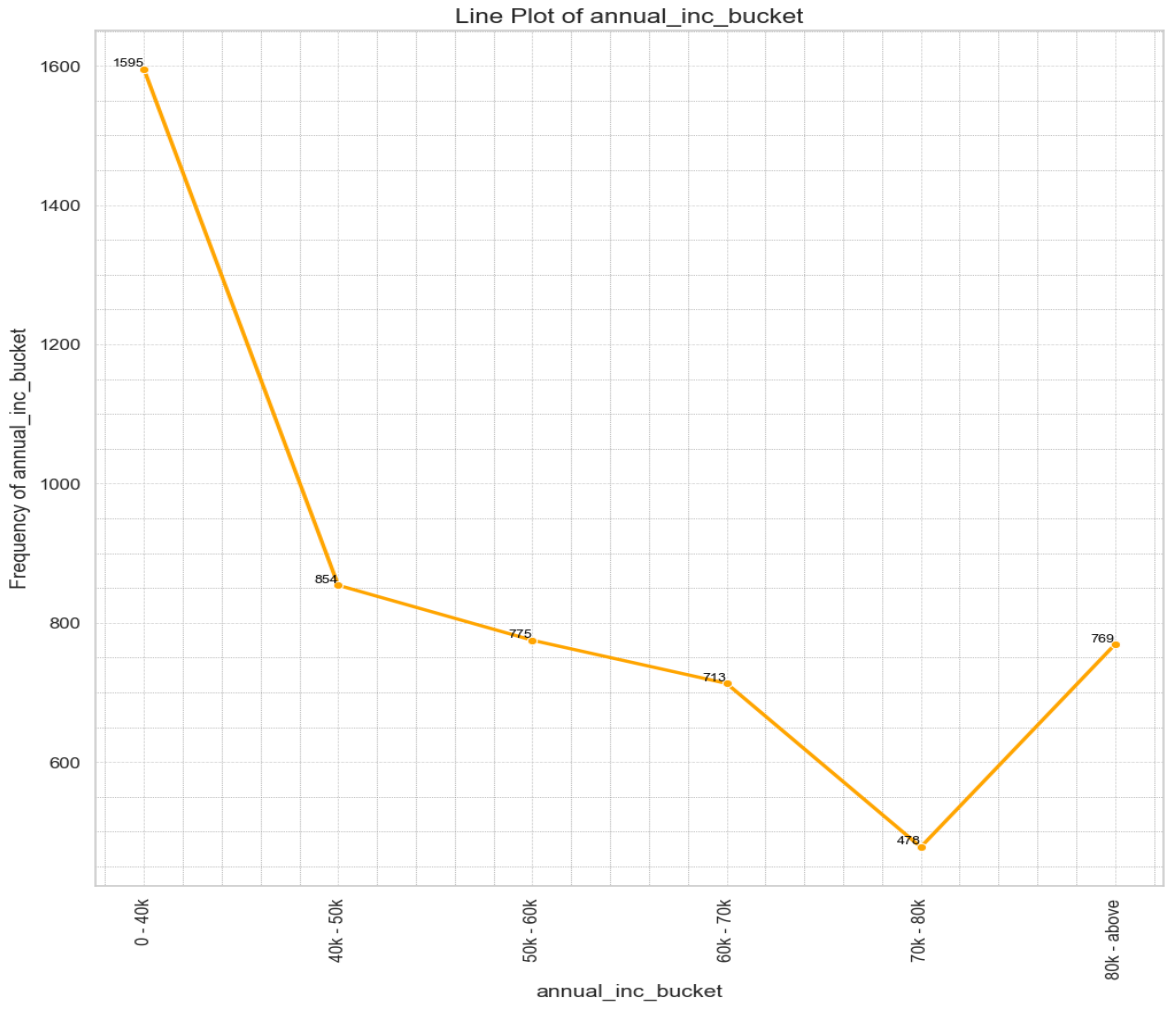


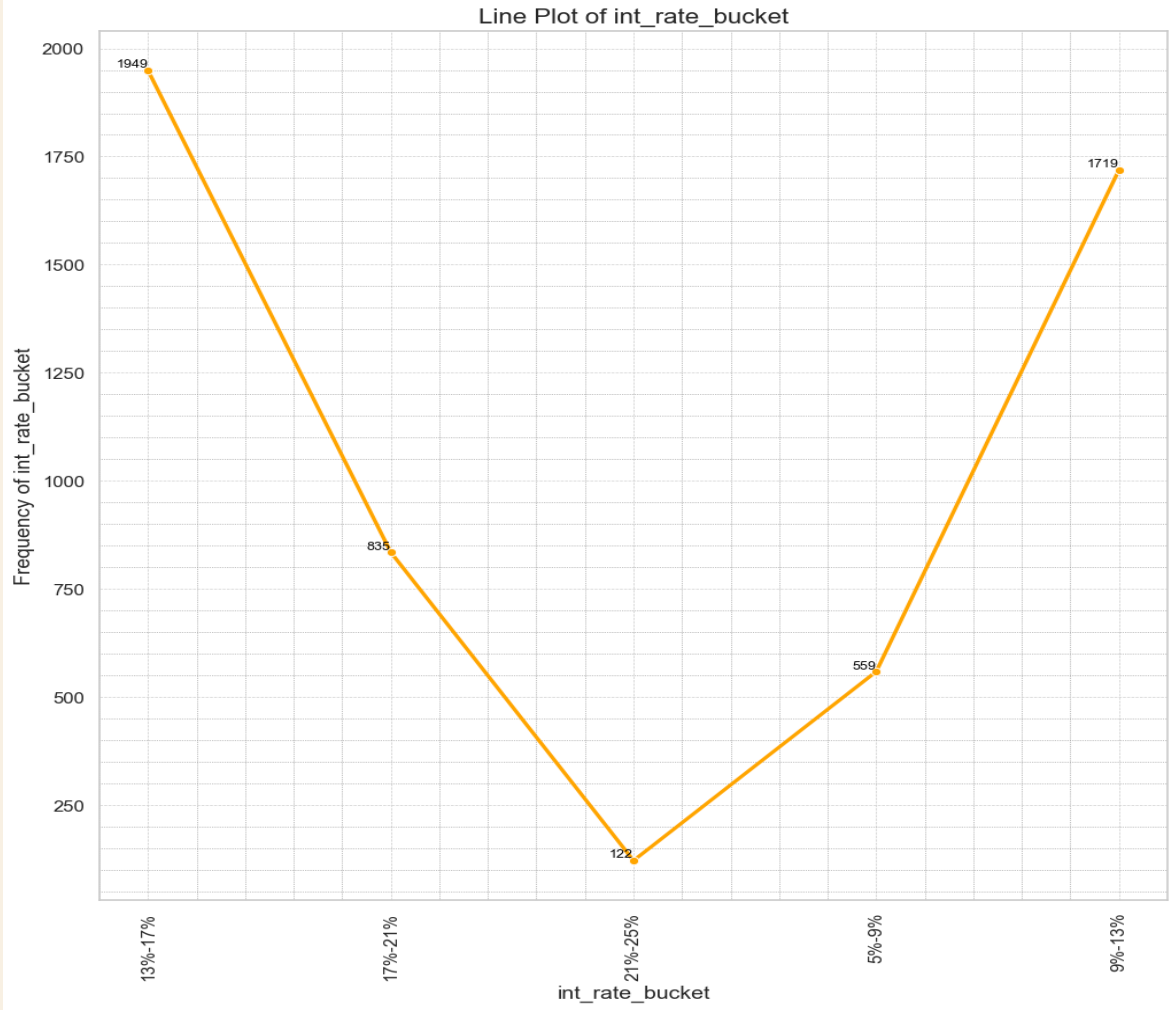
- **Home Ownership:** Most defaults were from applicants in **rented homes** (2,709). The lending company should assess the financial stability of renters, who may be more vulnerable to economic changes.

Univariate analysis : Annual Income ,Interest Rates & Summary

- **Annual Income: 1,595** defaults were from applicants earning less than **$40,000** annually. The company should scrutinize low-income applicants more closely, verifying income and repayment capacity thoroughly.

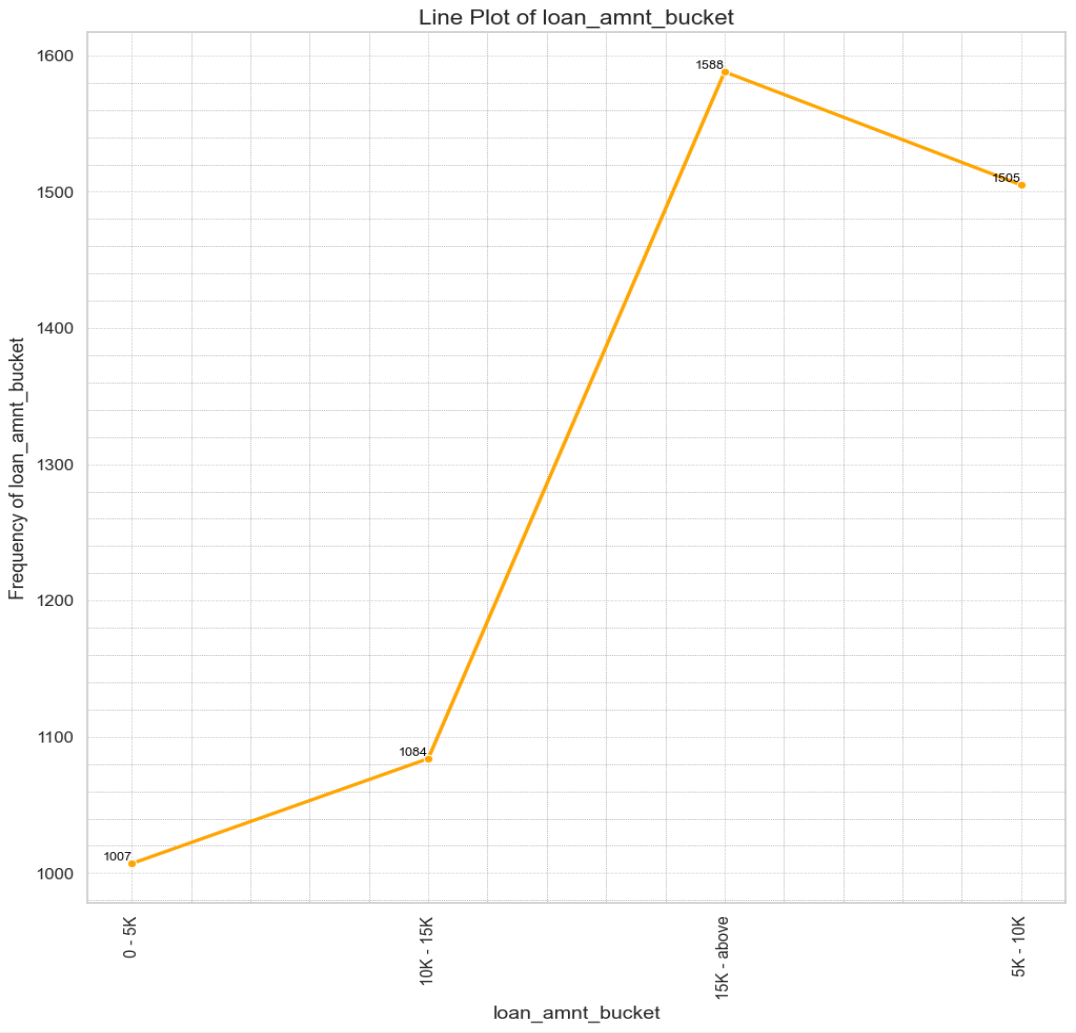
- **Interest Rates:** A significant number of defaults (1,949) occurred with interest rates between **13%-17%**. The company should consider offering lower rates to mitigate default risk.



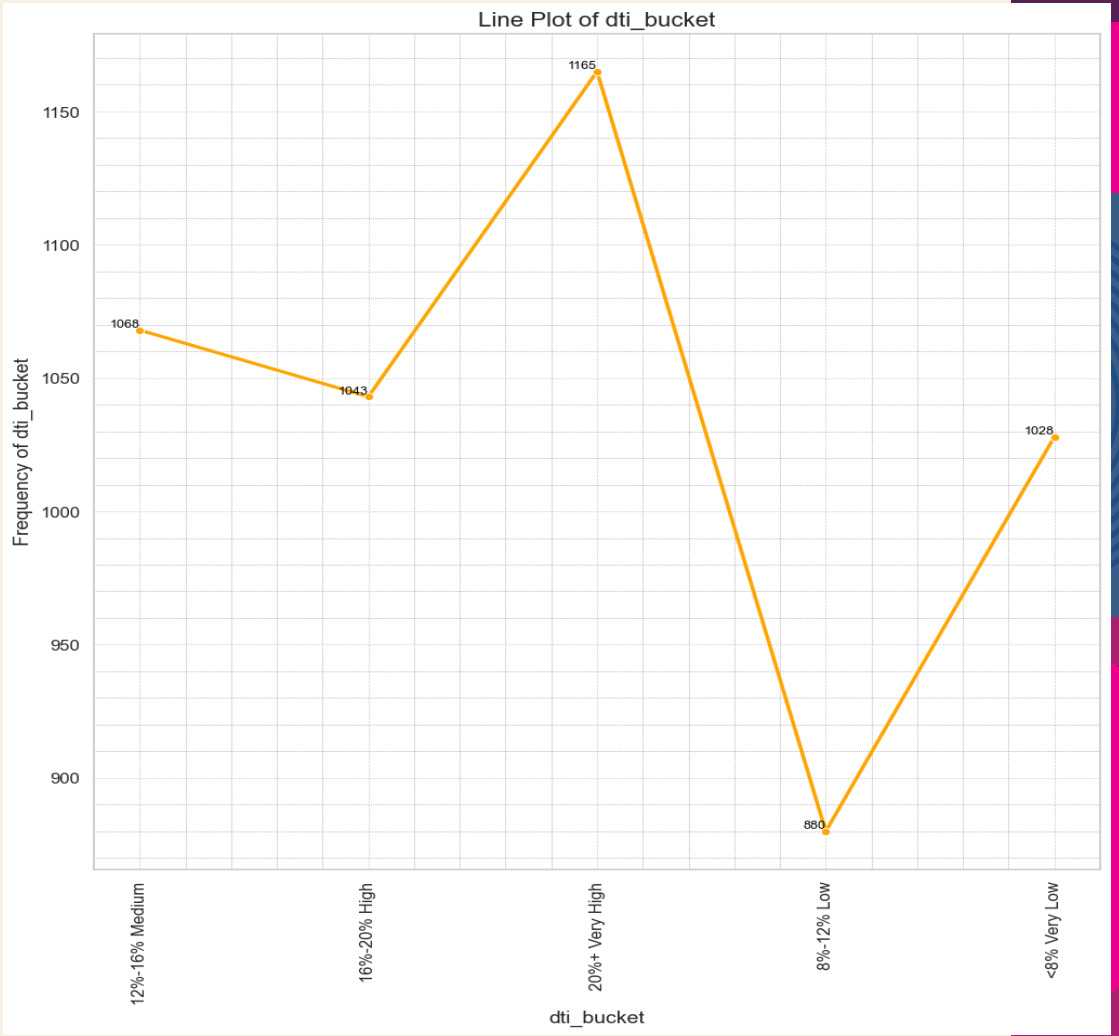


Univariate Analysis : Loan Amount , Debt to Income Ratio & Summary

- **Loan Amount: 1,588** defaults involved loans of **$15,000** or more. High loan amounts should be granted only to applicants with strong credit histories and repayment capabilities.

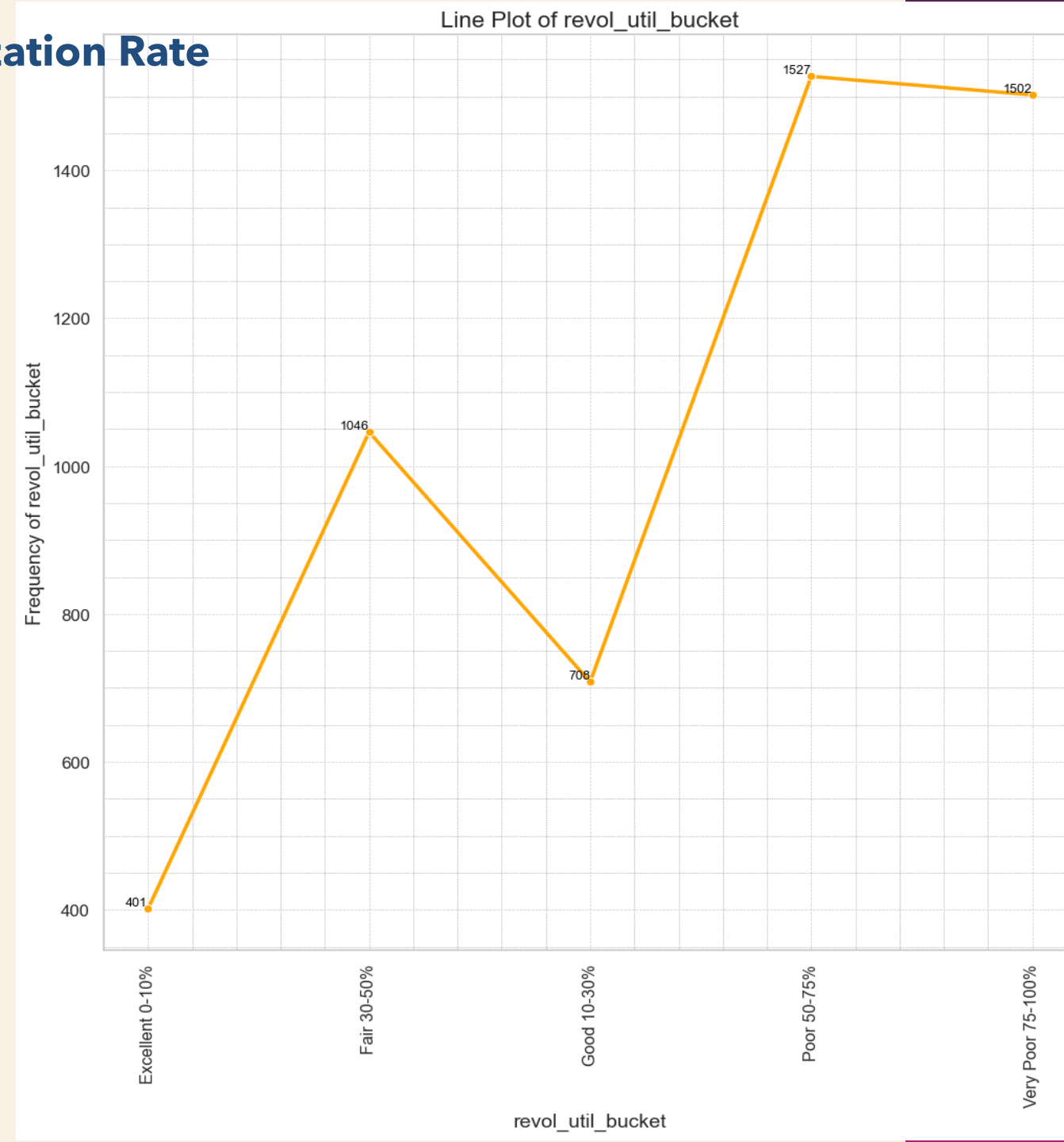


- **Debt-to-Income Ratio: 1,165** defaults were from applicants with high debt-to-income ratios. Strict ratio requirements should be enforced to prevent lending to over-indebted individuals



Univariate Analysis : Revolving Line Utilization Rate

- **Revolving Line Utilization Rate (RLUR):** Defaults were highest among those with **Poor (50%-75%)** and **Very Poor (75%-100%)** RLUR, at 1,527 and 1,502 respectively. RLUR should be a key consideration in lending decisions to gauge credit utilizations to gauge credit utilization.



BIVARIATE ANALYSIS

For Bivariate Analysis, **Loan_Status** is considered as one variable, representing Fully_Paid and Charged_Off states. The **Default Rate** or **Charged_Off percentage** is considered as the second variable against various other parameters such as grade, sub_grade, emp_length, purpose, home_ownership, and verification_status.

Categorical variables

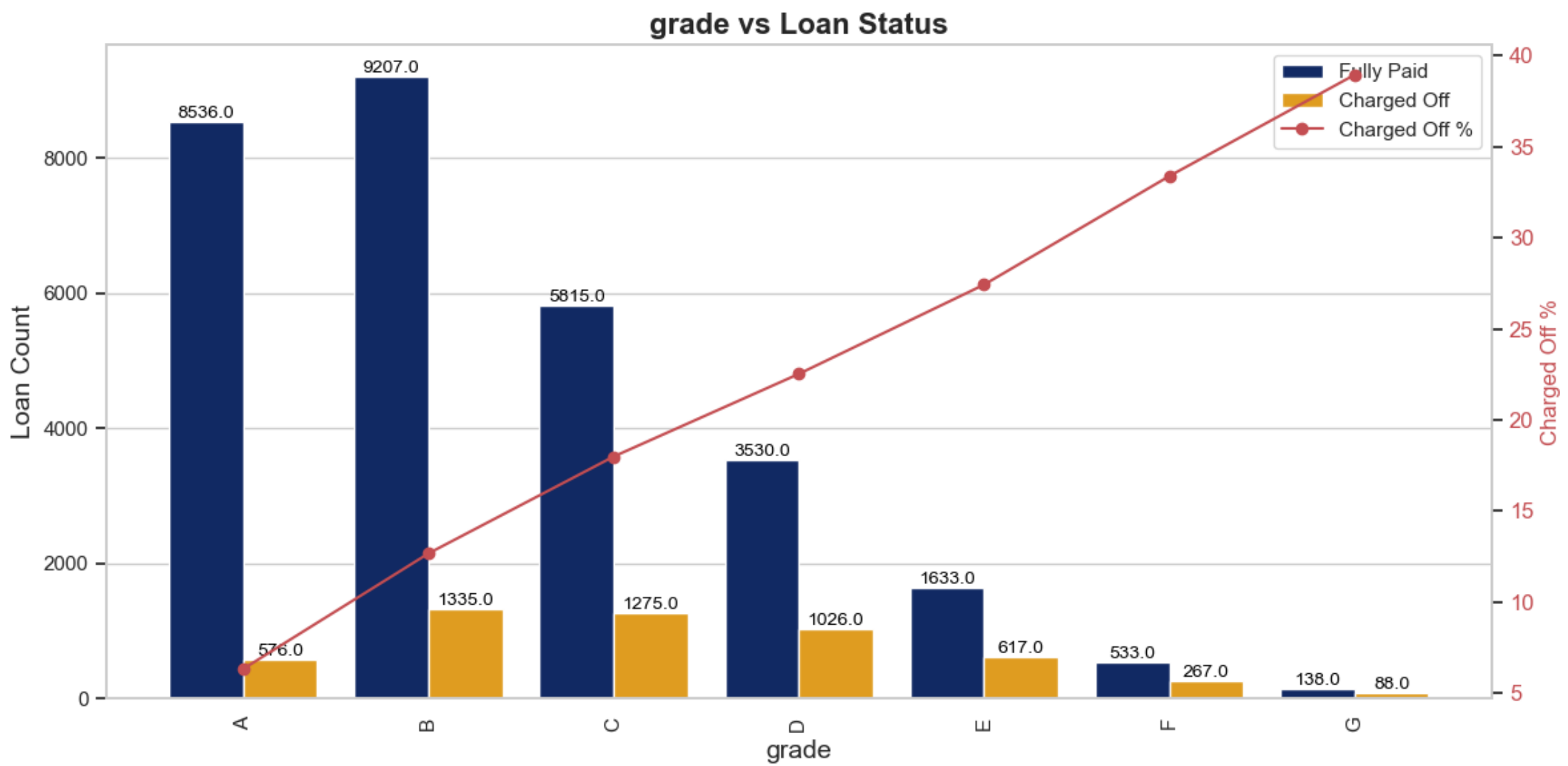
- Grade
- Sub grade
- Employment length
- Loan purpose
- Home Ownership
- Verification Status

Quantitative variables

- Annual Income
- Int Rate
- Debt to Income Ratio
- Loan Amount
- Revolving Line Utilization rate

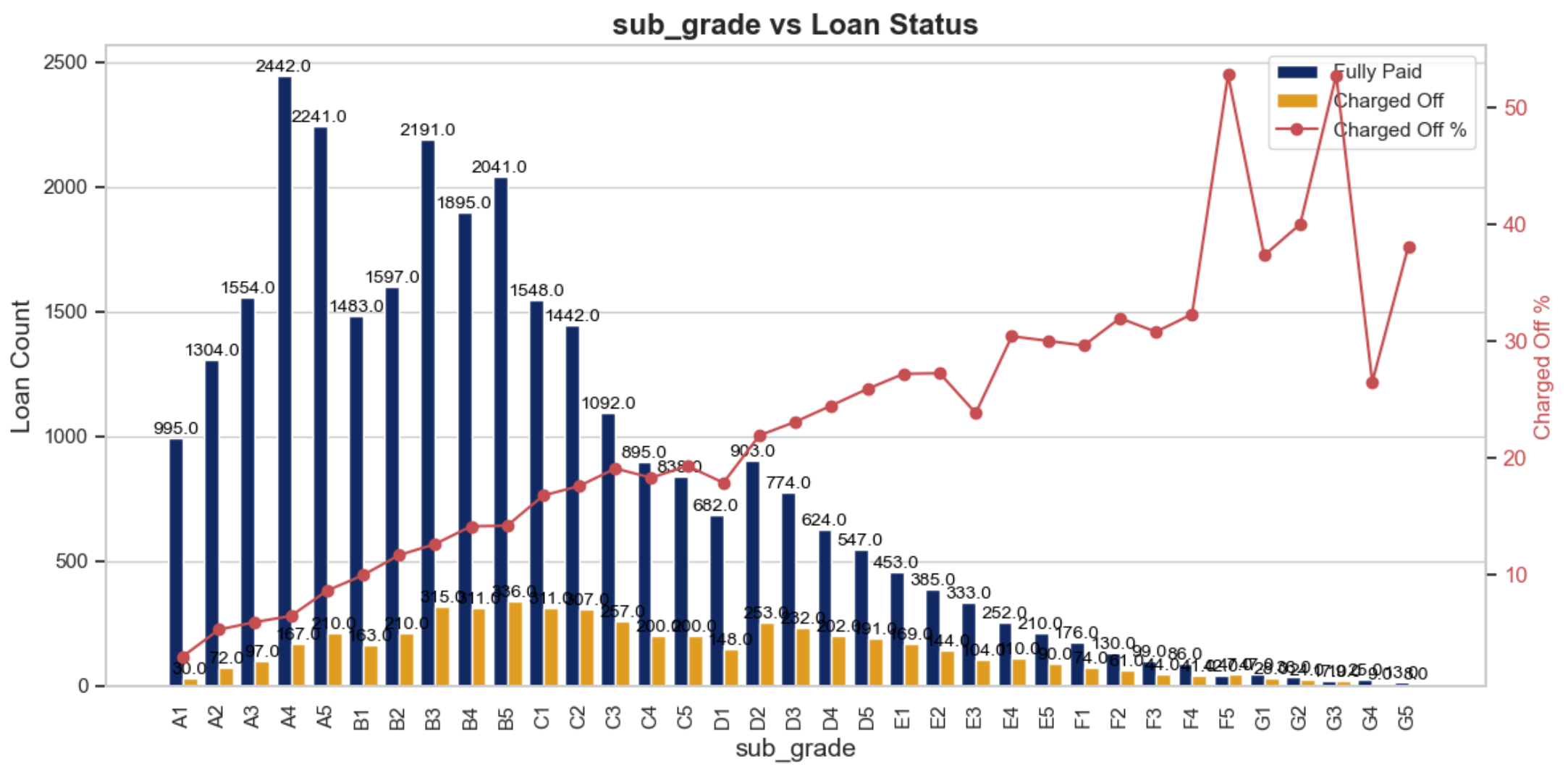
Bivariate analysis : Grade and Loan Status

Grade B has a high number of **default applicants (1335),** whereas in relative numbers, **grade G** has **38%** of applicants defaulting.



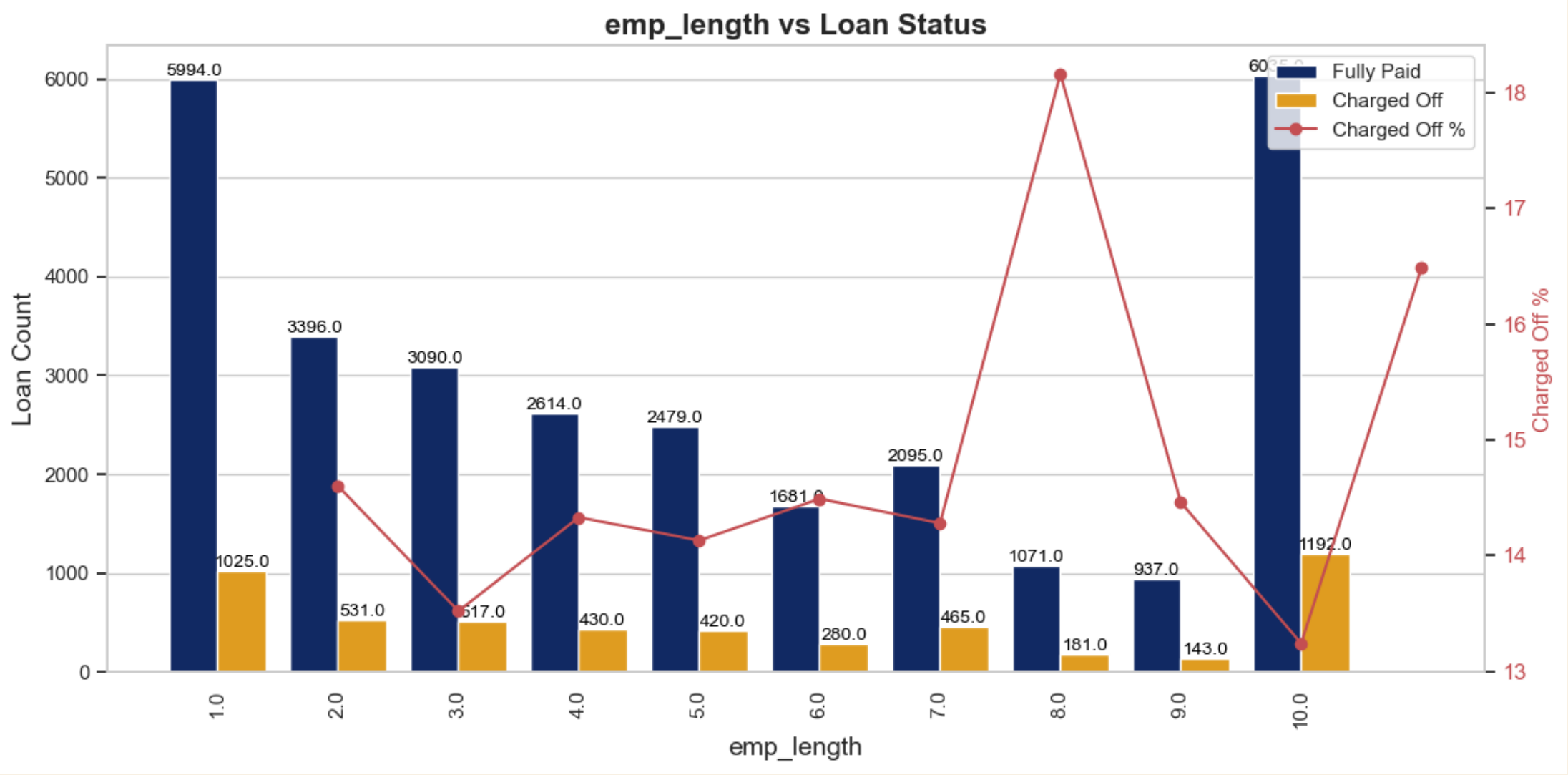
Bivariate Analysis : Sub Grade and Loan Status

- Sub Grades B3, B4, and B5 as they are more likely to charge off in absolute numbers. Also, note subgrades F3 and G3, where the default rate is more than 50%.



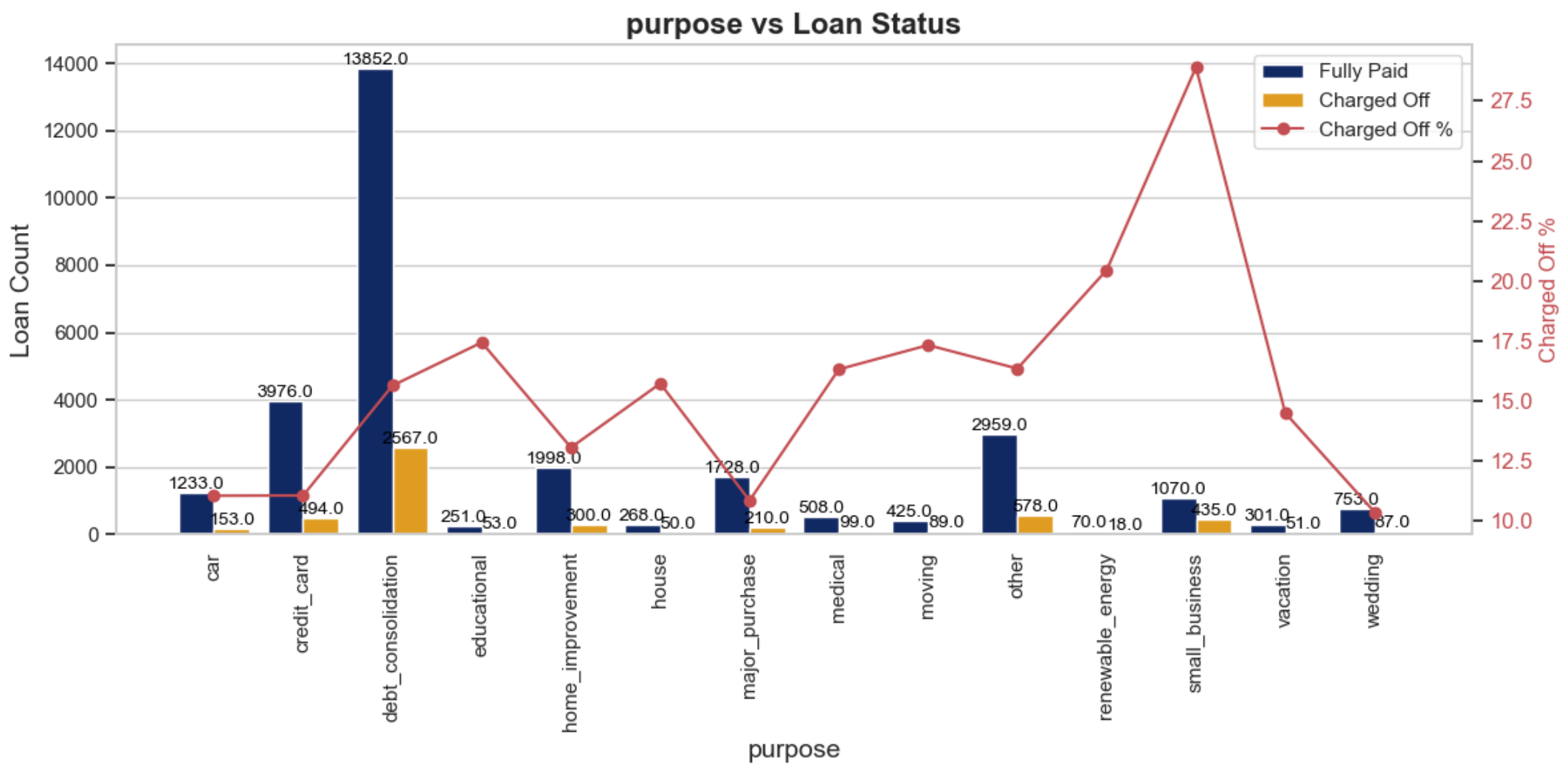
Bivariate Analysis : Emp Length and Loan Status

- Loan applicants with 10+ years of experience are more likely to default in absolute numbers, whereas in relative numbers, applicants with 8 years of experience have an 18% default rate.



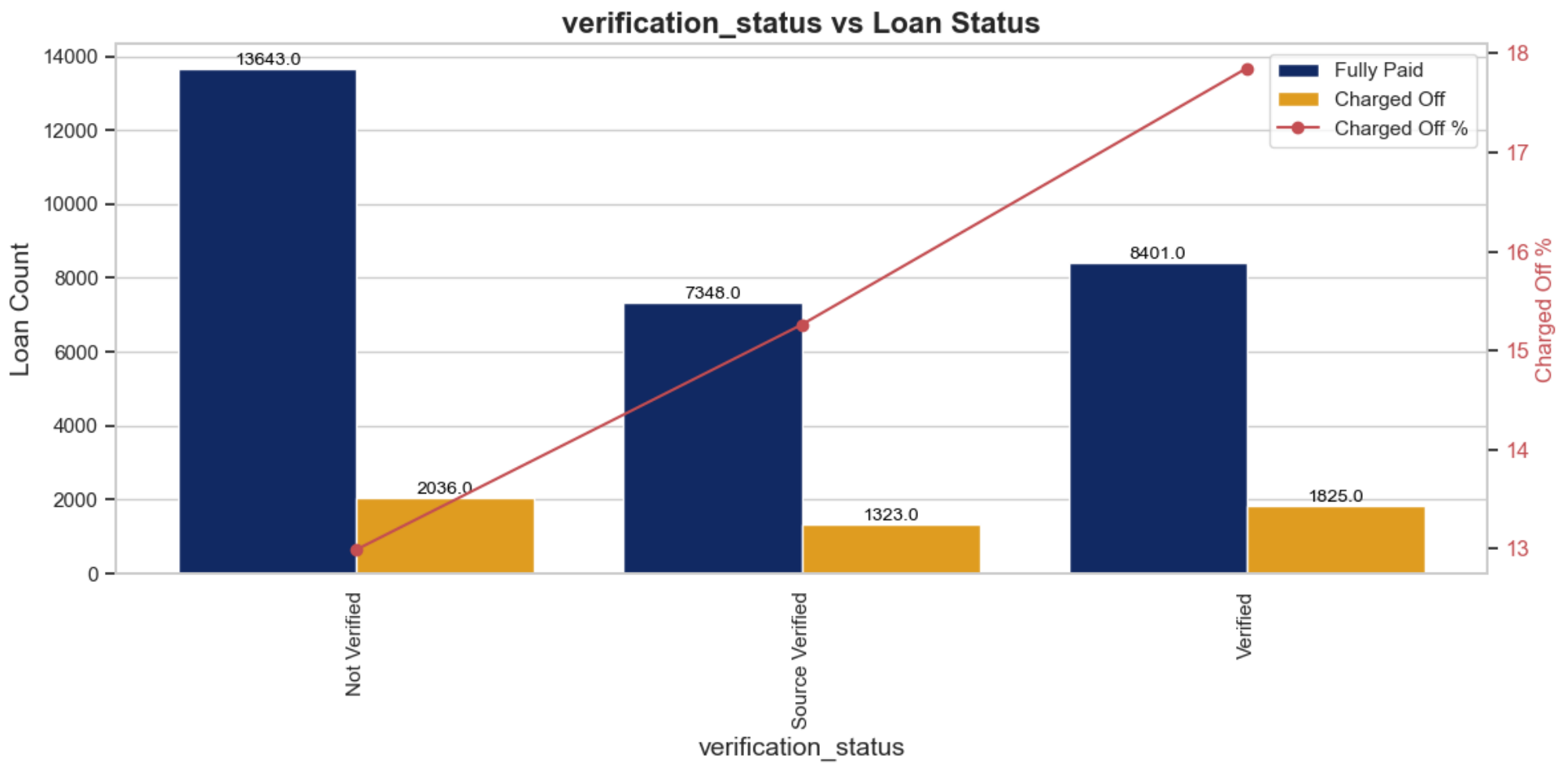
Bivariate Analysis : Purpose and loan status

- Applicants seeking loans for debt consolidation are high in absolute terms, which may result in a vicious debt cycle if financial discipline isn't followed. In relative terms, loans taken for small businesses have a higher percentage of defaults.



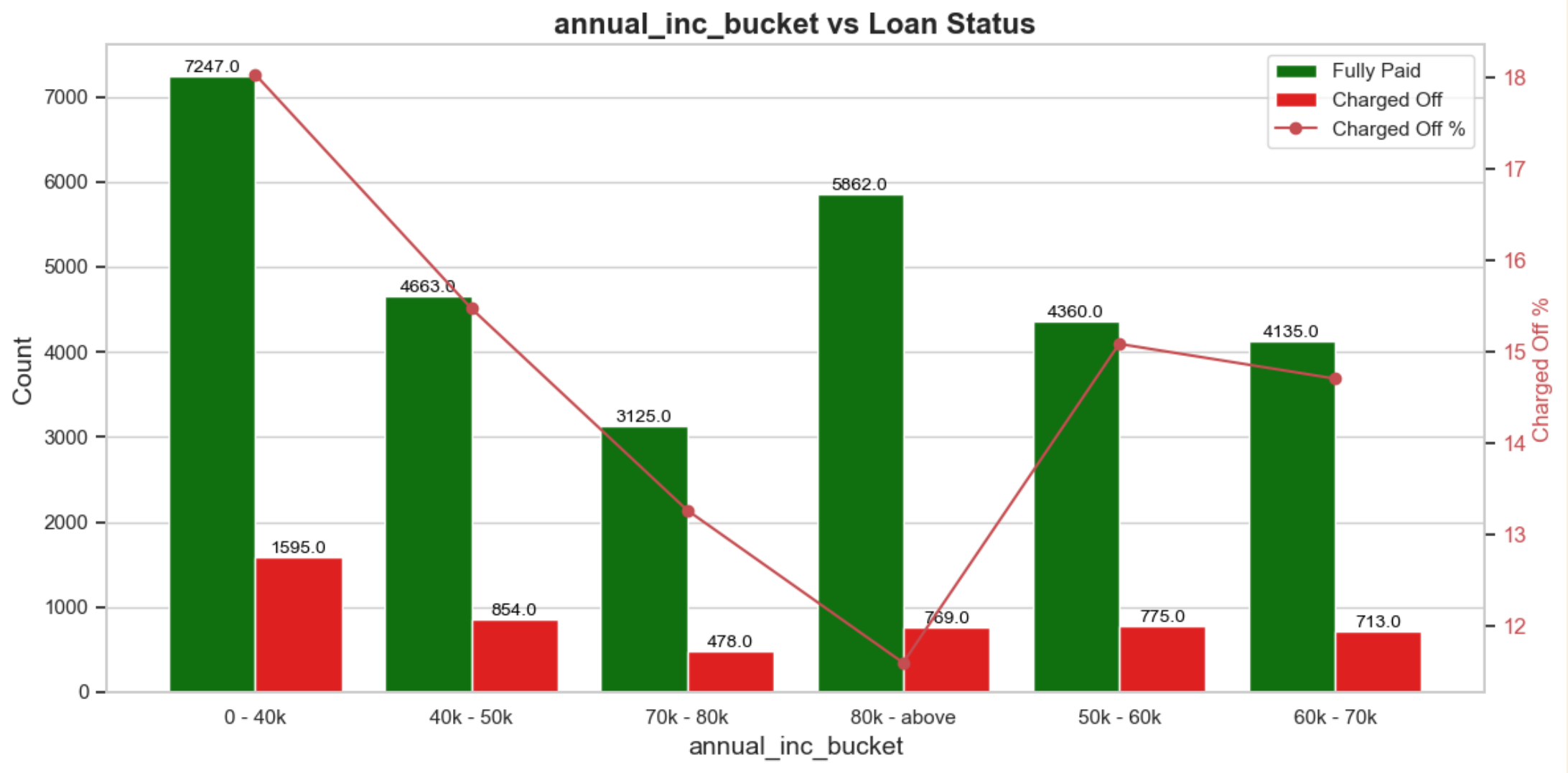
Bivariate analysis : verification status and loan status

- Verified loan applicants are defaulting more in relative terms than those who are not verified. The company should review its verification process to ensure it effectively assesses applicant creditworthiness and consider improvements or adjustments.



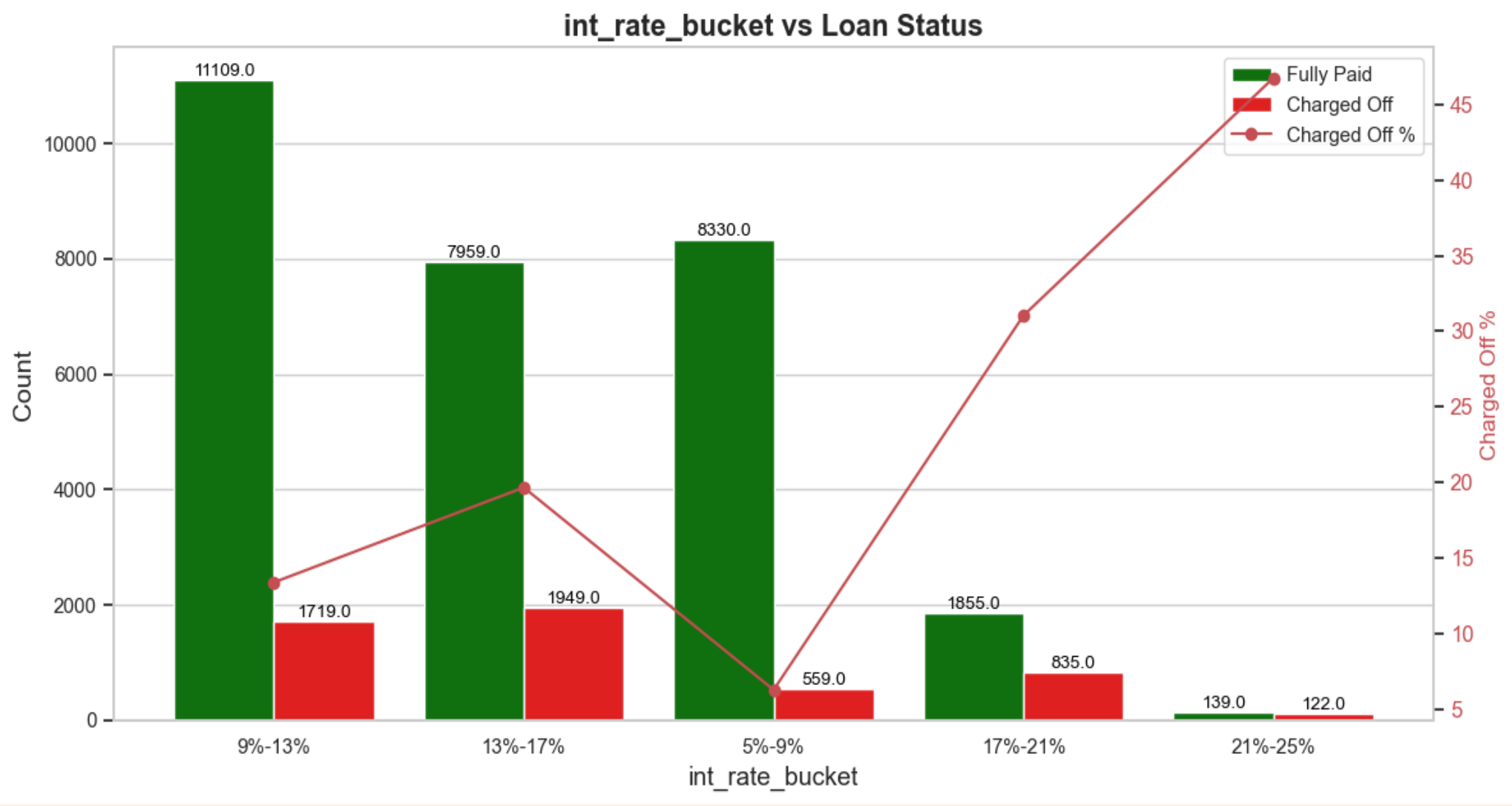
Bivariate Analysis : Annual Income and loan status

- Applicants with annual incomes less than $40,000 have a higher likelihood of defaulting both in absolute and relative terms.



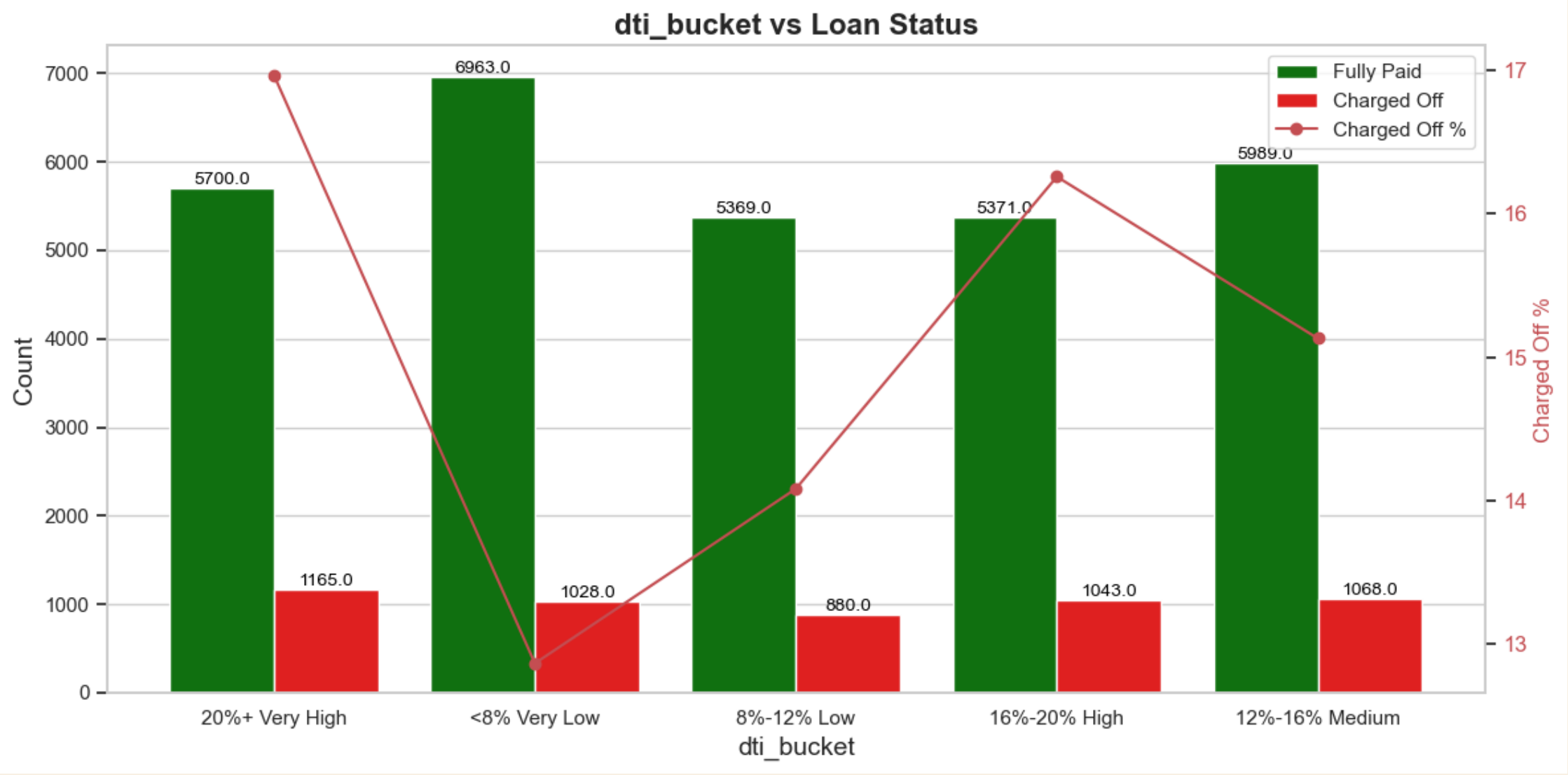
Bivariate Analysis : Interest Rates and Loan Status

- Interest rates in the 13%-17% range are associated with defaults in absolute terms, whereas the default possibility is high in the high-interest rate bucket of 21-25% in relative terms.



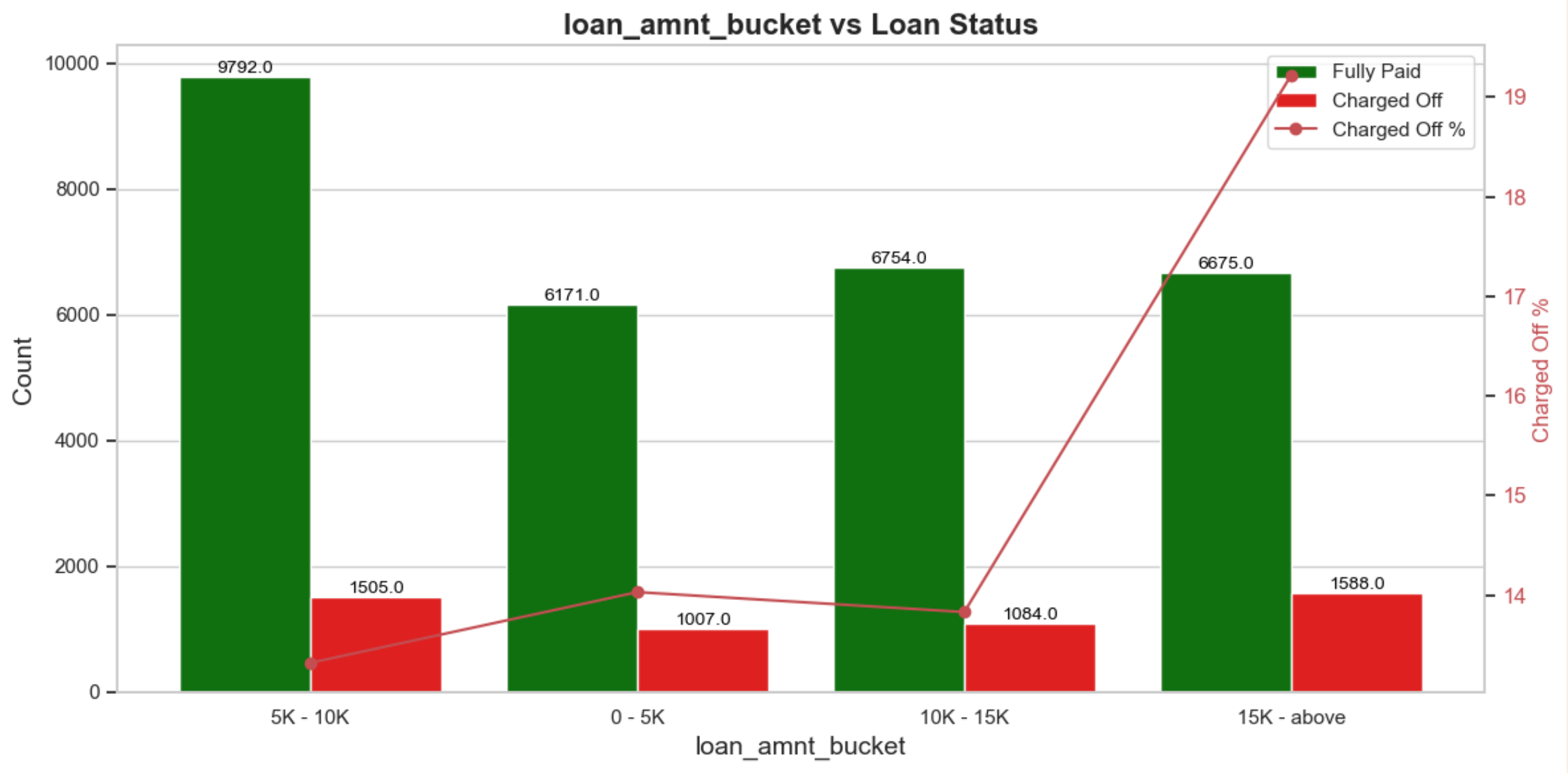
Bivariate Analysis : Debt to Income Ratio and Loan Status

- Applicants with a high DTI ratio (>20%) have a higher likelihood of defaulting both in absolute and relative terms. Hence, this should be a strong criterion for the lending company.



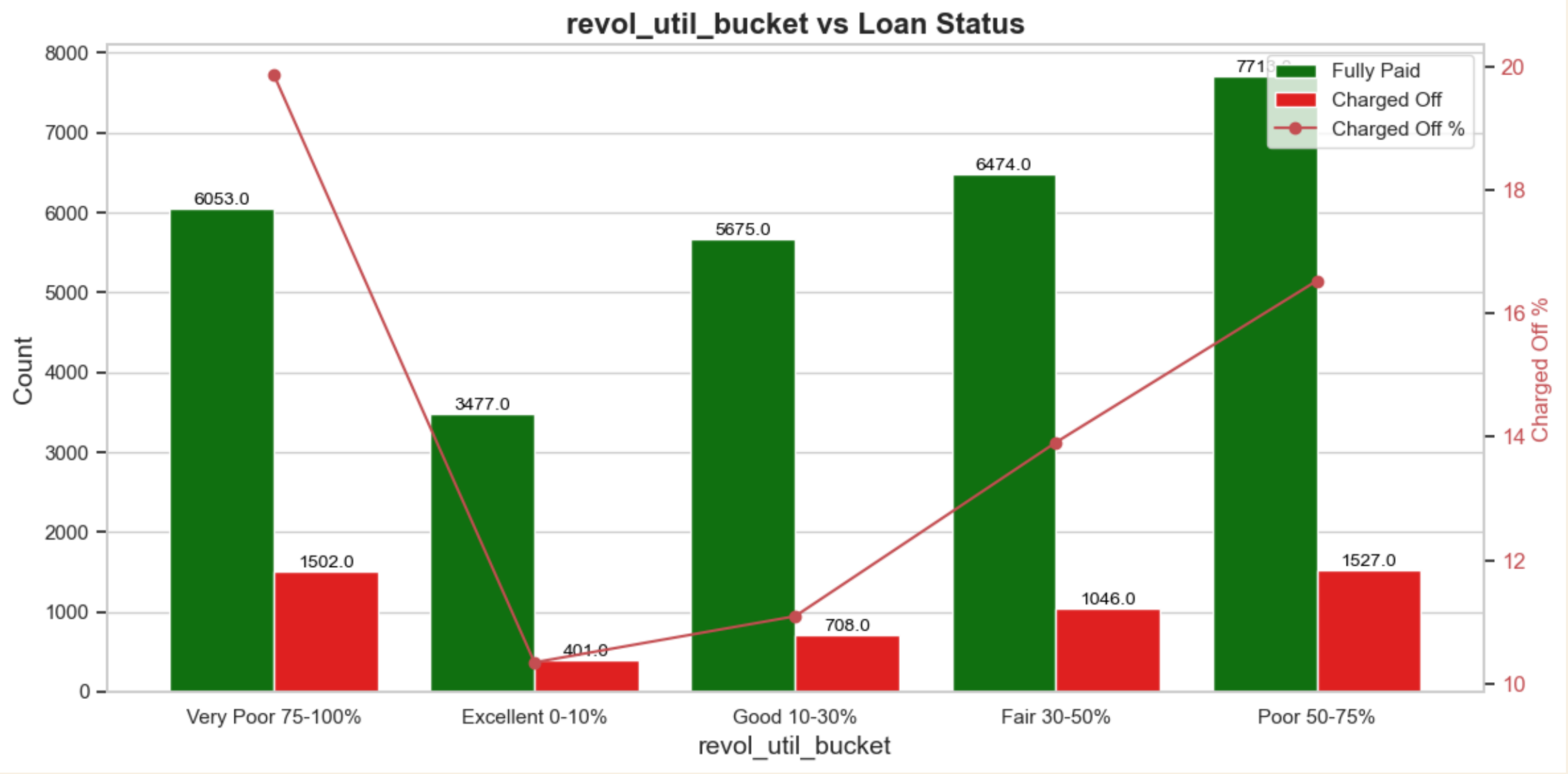
Bivariate Analysis : Loan Amount and Loan Status

- Applicants applying for loan amounts of $15,000 or higher are more likely to default both in absolute(1588) and relative(19%) terms.



Bivariate Analysis : Revolving Line Utilization Rate and Loan Status

- Applicants with poor (50-75%) RLUR are highly likely to default in absolute terms, whereas very poor (75-100%) RLUR applicants are highly likely to default in relative terms.



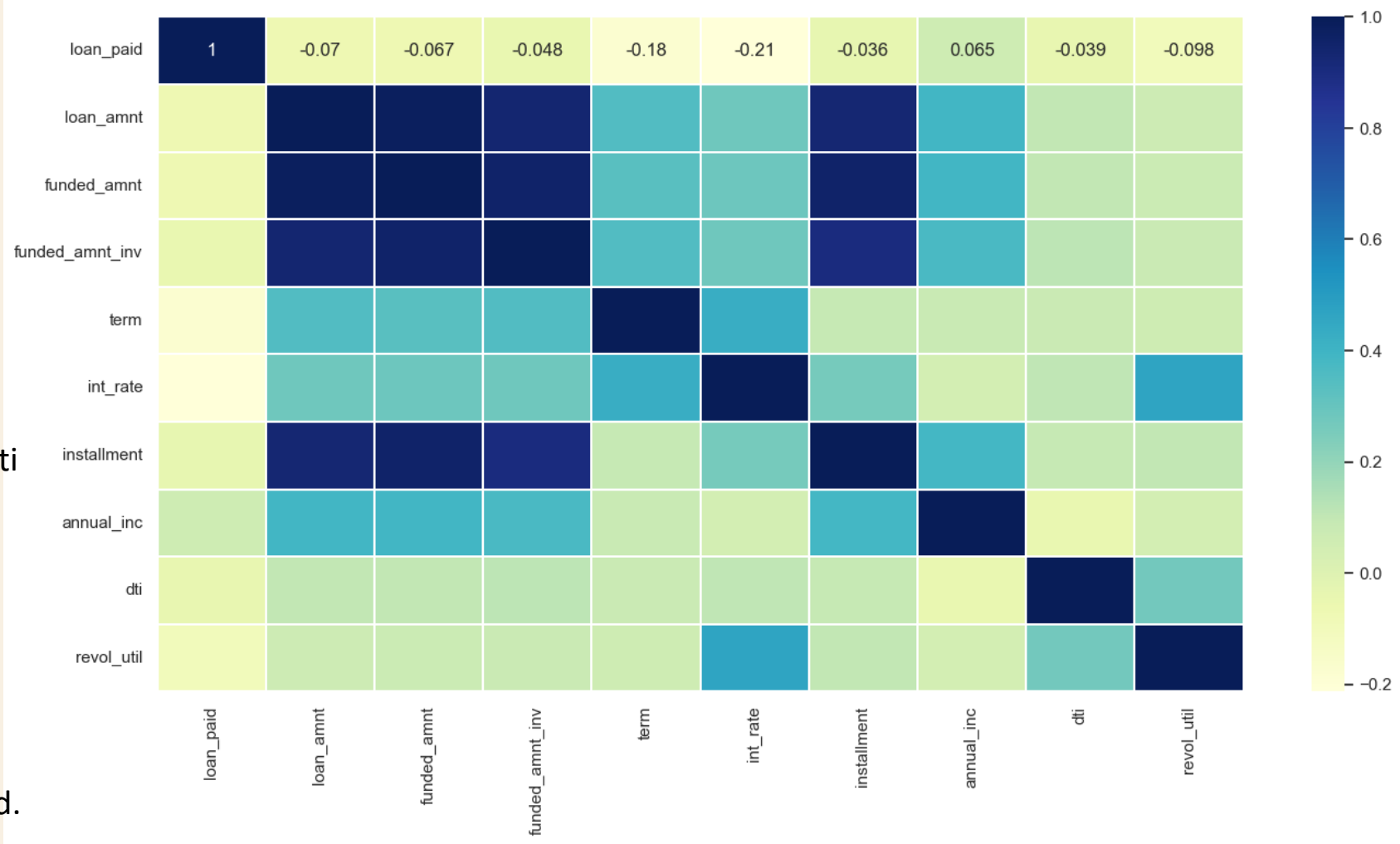
CORRELATION ANALYSIS

Correlation Analysis Summary

Strong Correlation
loan_amnt shows a strong correlation with funded_amnt, int_rate, and funded_amnt_inv.
Term is strongly correlated with the interest rate.
Installment is strongly correlated to loan_amount,funded_amnt, funded_amnt_inv.
Revol_util is strongly correlated to int_rate and dti

Weak Correlation
loan_paid exhibits weak correlations with most variables.
Emp_length has weak correlations with most variables.

Negative Correlation
int_rate has a negative correlation with loan_paid.



CONCLUSION

Univariate Analysis Summary : Grade, Term, Purpose, Annual Income, Interest Rate, Debt-Income ratio and Revolving Line Utilization Rate are major variable which causes default in absolute numbers

Bivariate Analysis Summary : Annual Income, Debt-Income ratio and Revolving Line Utilization Rate are major variables which could result in loan default in absolute and relative terms

Correlation Analysis Summary : Loan Amount shows a strong correlation with funded amount, Interest Rate, and funded amount investors.

Loan term is strongly correlated with the interest rate.

Installment is strongly correlated to Loan Amount, Funded Amnt, Funded Amnt Inv.

Revolving Line Utilization Rate is strongly correlated to Interest Rates and Debt-Income ratio

Loan paid and employment length exhibits weak correlations with most variables.

THANK YOU

-Ranjith Kumar J
-Sakshi Singh