LENDING CLUB CASE STUDY

By : Varun Mohite and Nakul Gupta

Contents

Business Objective

- The objective is to identify the driving factors (or driver variables) behind loan default at the largest online loan marketplace offering various types of loans like personal loans, business loans and financing of medical procedures.
- Identification of the variables can be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.
- Analyze data and generate insights and recommendations which will help Lending Club for its portfolio and risk assessment.

Data Information

- Historical data containing the loan details for all loans issued.
  - The data contains 111 columns and 39717 rows.
  - The primary attribute in data is ‘loan_status’ which has 3 distinct data values:
    - ✓Fully-Paid: Signifies customers who have successfully repaid their loans.
    - ✓Charged-Off: Indicates customers who have been labeled as "Charged-Off" or have defaulted on their loans.
    - ✓Current: Represents customers whose loans are presently in progress and, thus, cannot provide conclusive evidence regarding future defaults.

    Note - For the purposes of this case study, rows with a "Current" status will be excluded from the analysis.

Analysis Approach

- Understanding business KPIs - Studying data dictionary and data to understand variables and identify target variable.

- Data Preprocessing –
  - Data Cleaning (remove Null values, single value and duplicates, outlier treatment etc.)
  - Data Transformation (converting the data into a suitable format)
  - Data Reduction (selecting required KPIs for analysis)
  - Data Discretization (binning ,clustering)

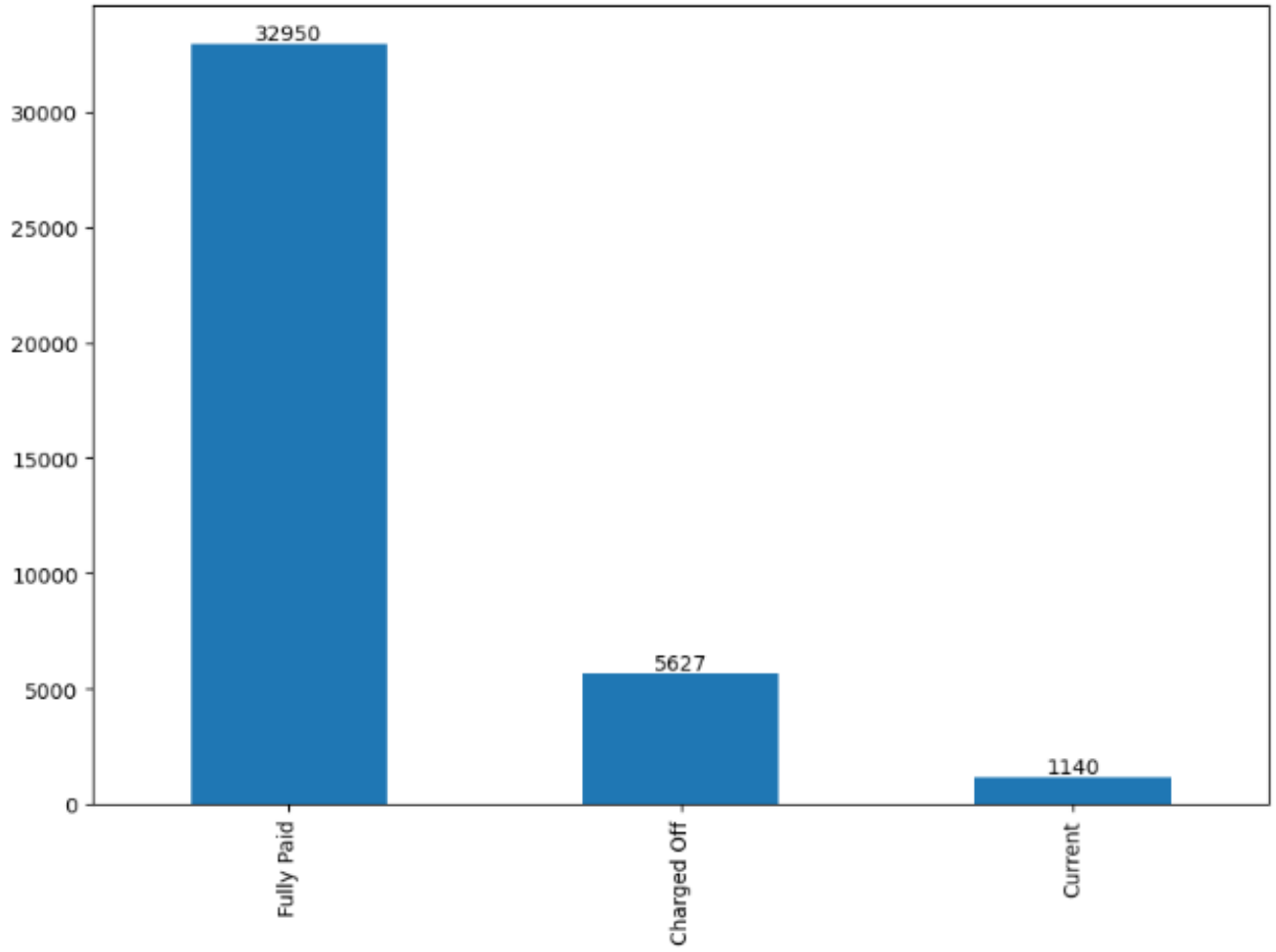
- Univariate Analysis - Check distributions and frequencies of various numerical and categorical variables. Create derived variables if required.

- Segmented Univariate Analysis - Analyze variables against segments of other variables.

- Bivariate Analysis - Analysis of two variables for determining the empirical relationship between them.

- Summarization - Publish insights and observations

Analysis - Overall Loan Status

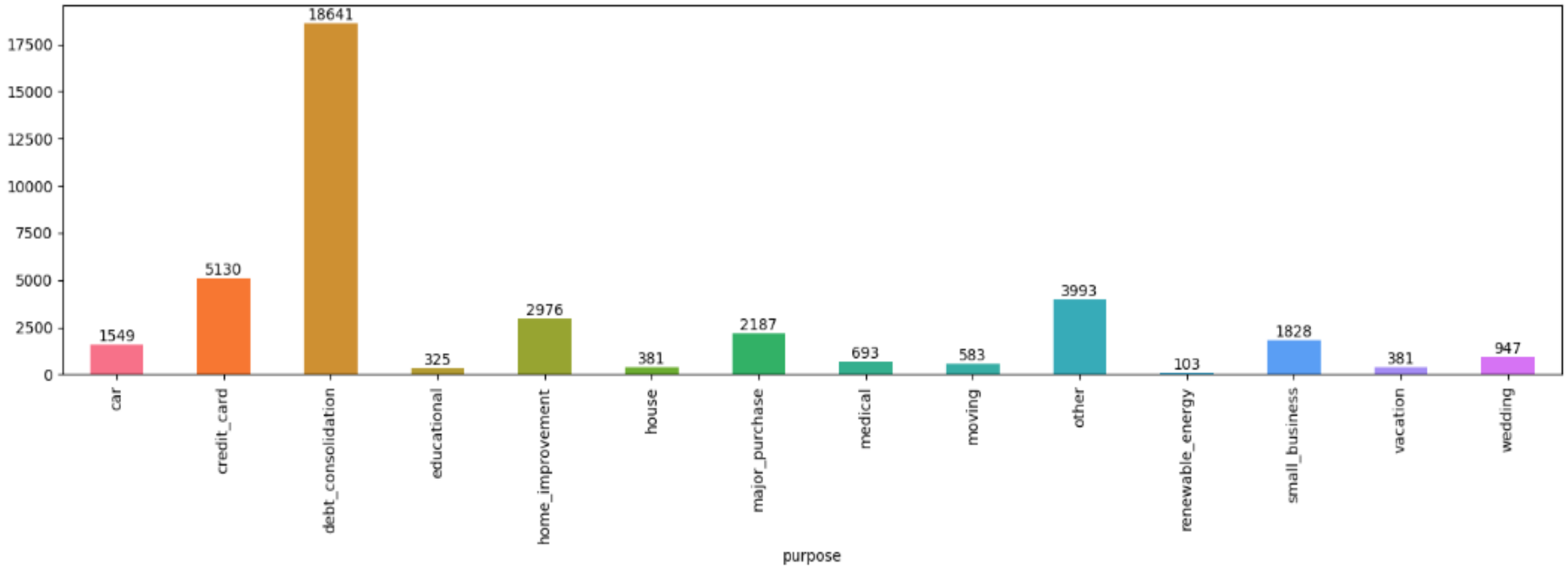


Current loan status –

Fully paid – 32950
Charged off – 5627
Current – 1140
Total – 34617

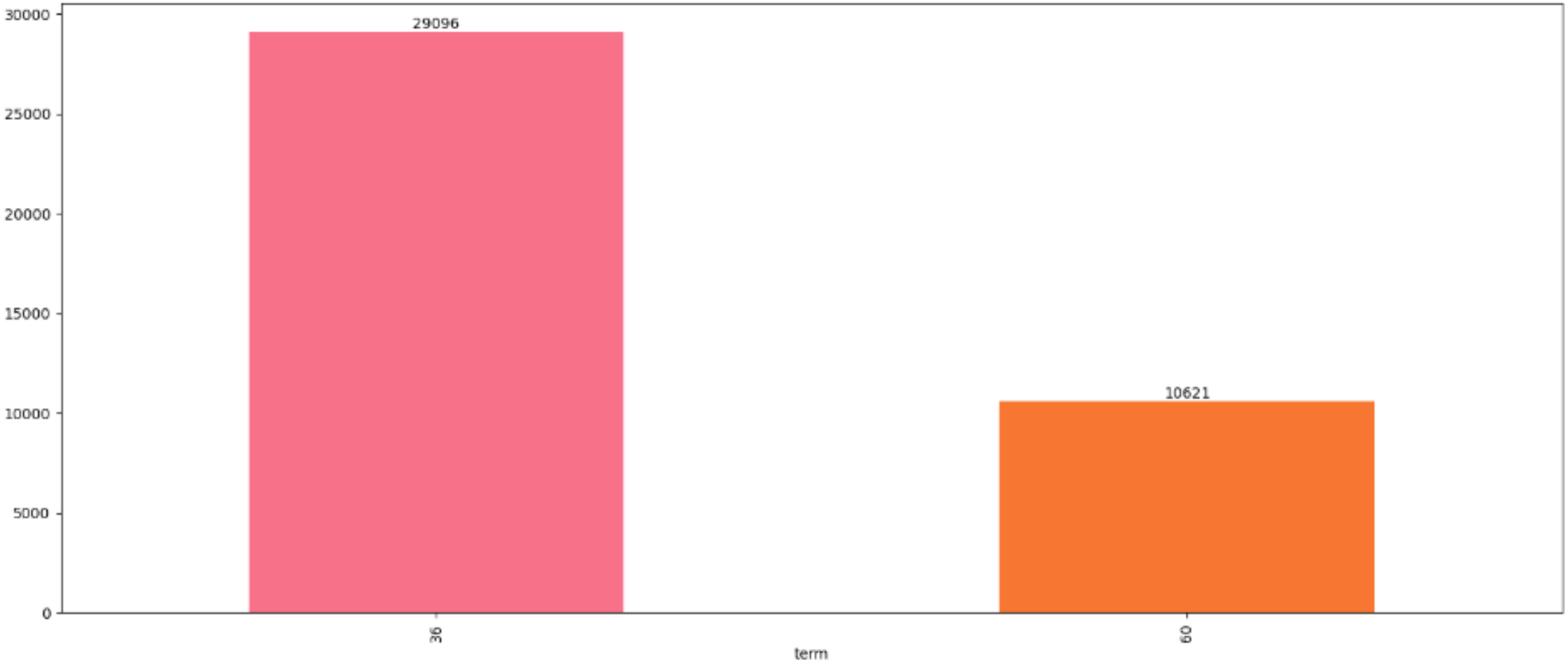
We can observe that 16.26% of total loans are charged off or defaulted.

Analysis : Purpose & Default Loan Status



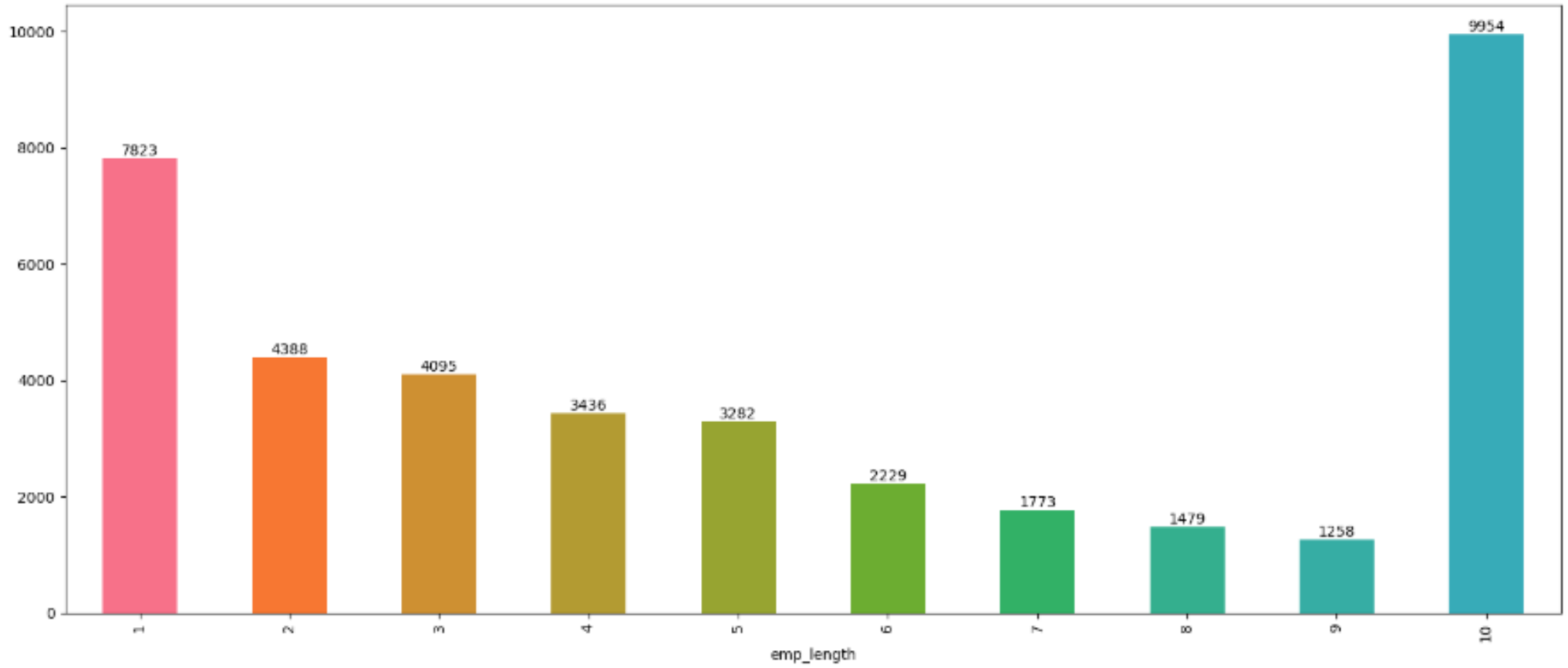
~47% of the “Charged off” loans were taken as debt consolidation loans. Care must be taken by the lending company when approving loans for debt consolidation purposes, as most of the debt consolidation loans have defaulted.

Analysis : Term & Default Loan Status



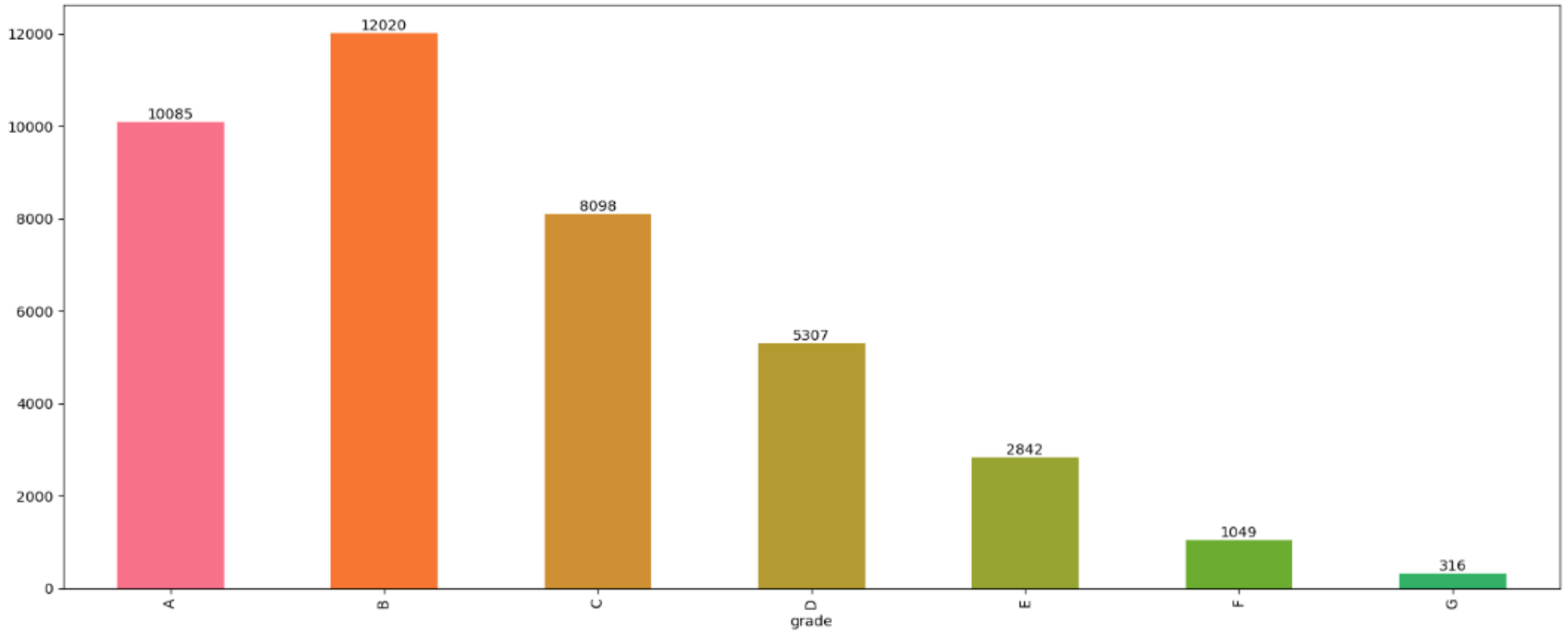
57% “Charged off” loans were short-term loans with a duration of 36 months. This suggests that a significant portion of applicants who experienced loan default chose shorter repayment terms.

Analysis : Employment Length & Default Loan Status



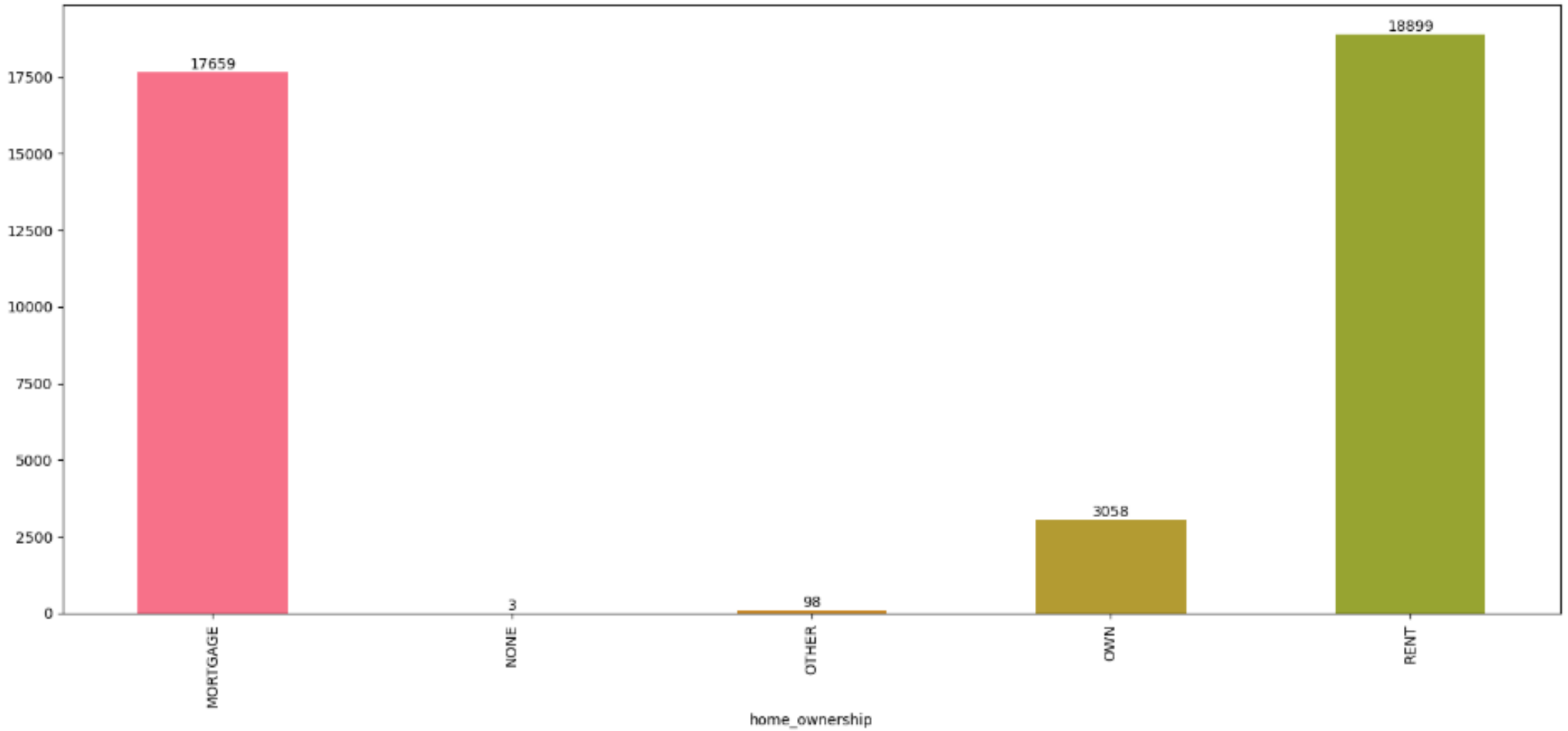
Applicants who had been employed > 10 years accounted for the highest number of "Charged off“ loans. This indicates that long-term employment history doesn’t guarantee successful loan repayment.

Analysis : Grade & Default Loan Status



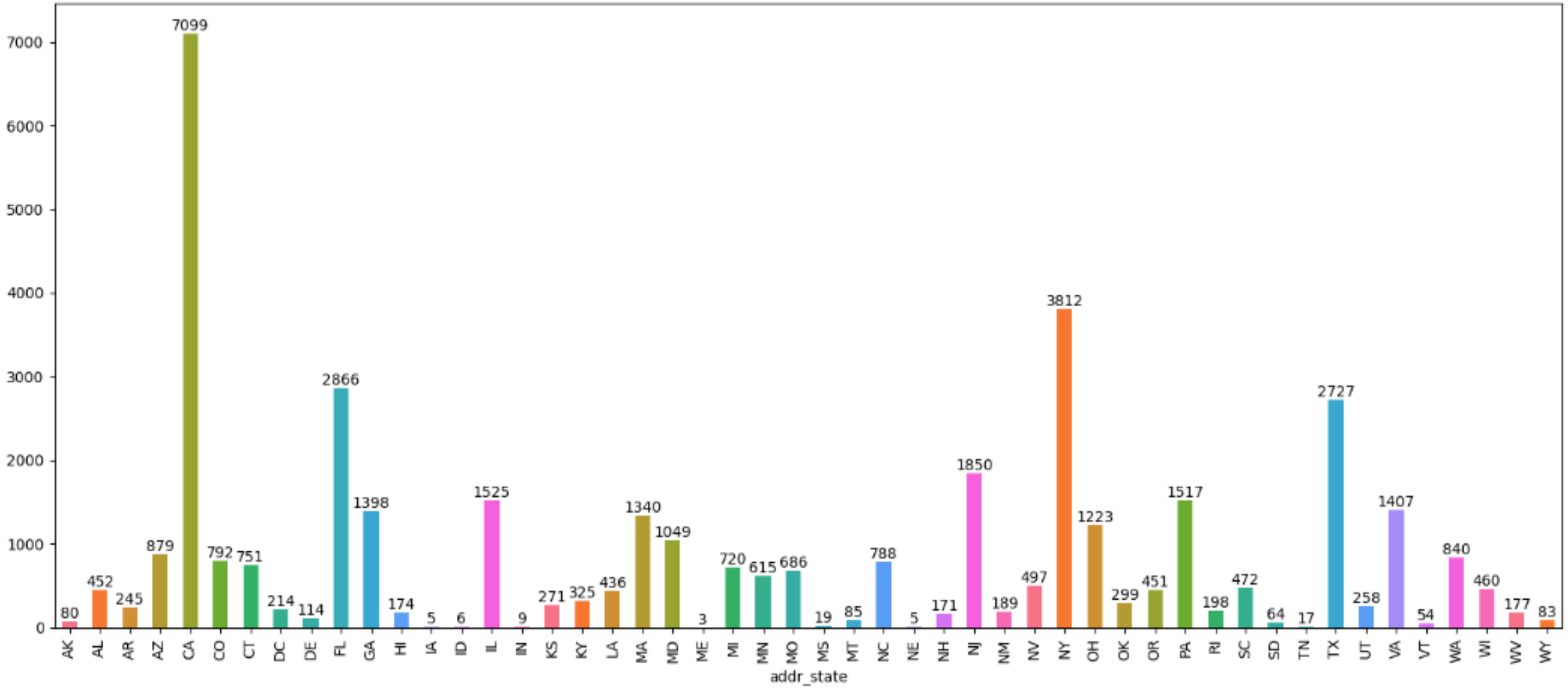
Grade B & Grade C applicants constitutes ~50% of the defaulters. Significant increase (58%) in default rate is observed from Grade A to Grade B while a declining trend is observed from Grade C to Grade G.

Analysis : Home ownership & Default Loan Status



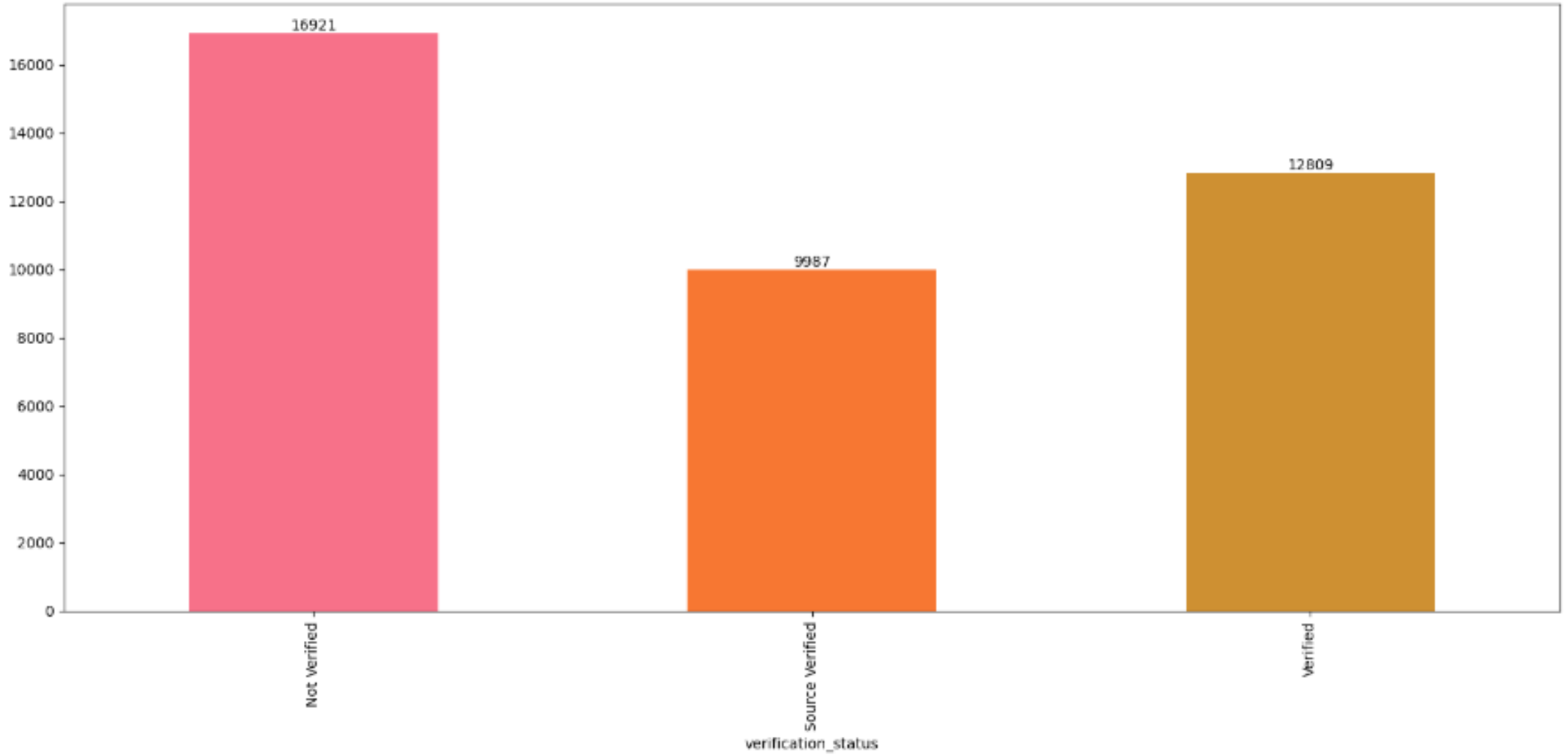
~50% of "Charged off" loan consumer lived in rented houses followed closed by mortgage houses.

Analysis : State & Default Loan Status



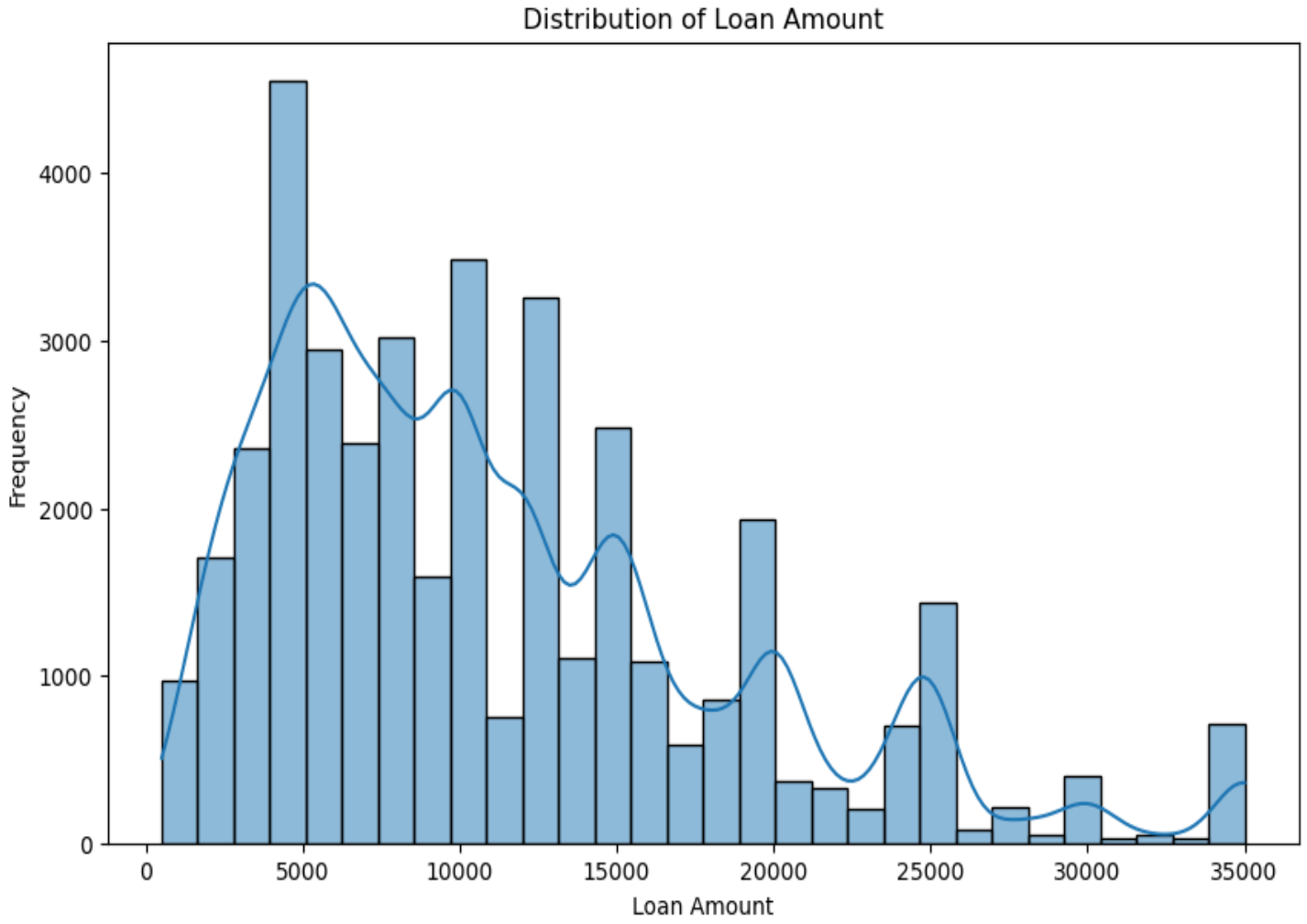
California observes high default rate followed by New York, Florida and Texas

Analysis : Verification status & Default Loan Status



The number of loan applicants who have been verified are defaulting more than those not verified.

Analysis : Distribution of loan amount

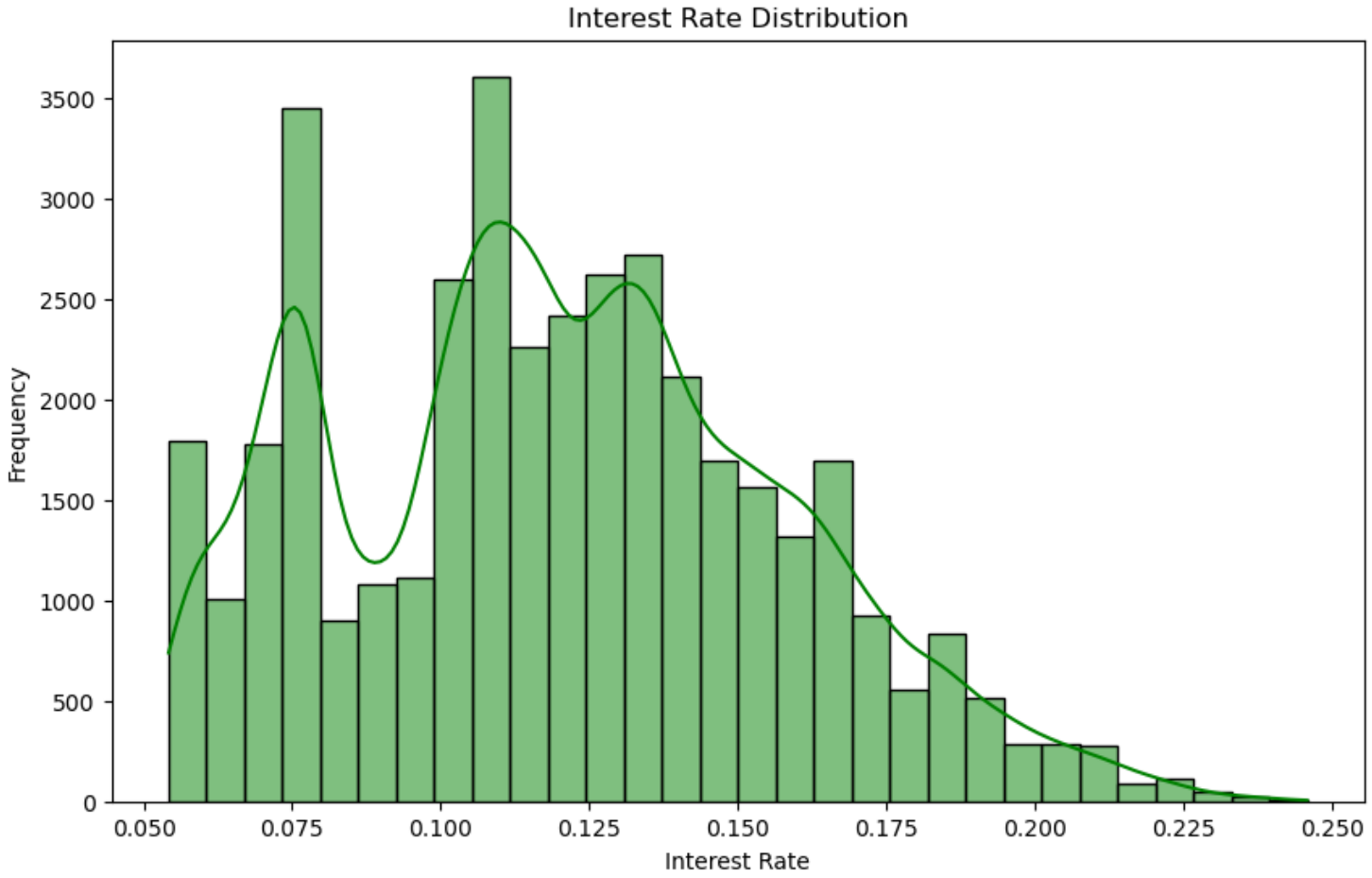


Visualizing the distribution of loan amounts to gain insight into the typical loan sizes that borrowers take.

Overall, the applied loan amount distribution is slightly right-skewed with mean greater than the median. Most of the loans granted are below $15000.

Customers applying for a loan up to $5000 is exponentially greater than any other amount followed by $10000. Less customers apply for loans above $15000 amount.

Analysis : Interest rate distribution

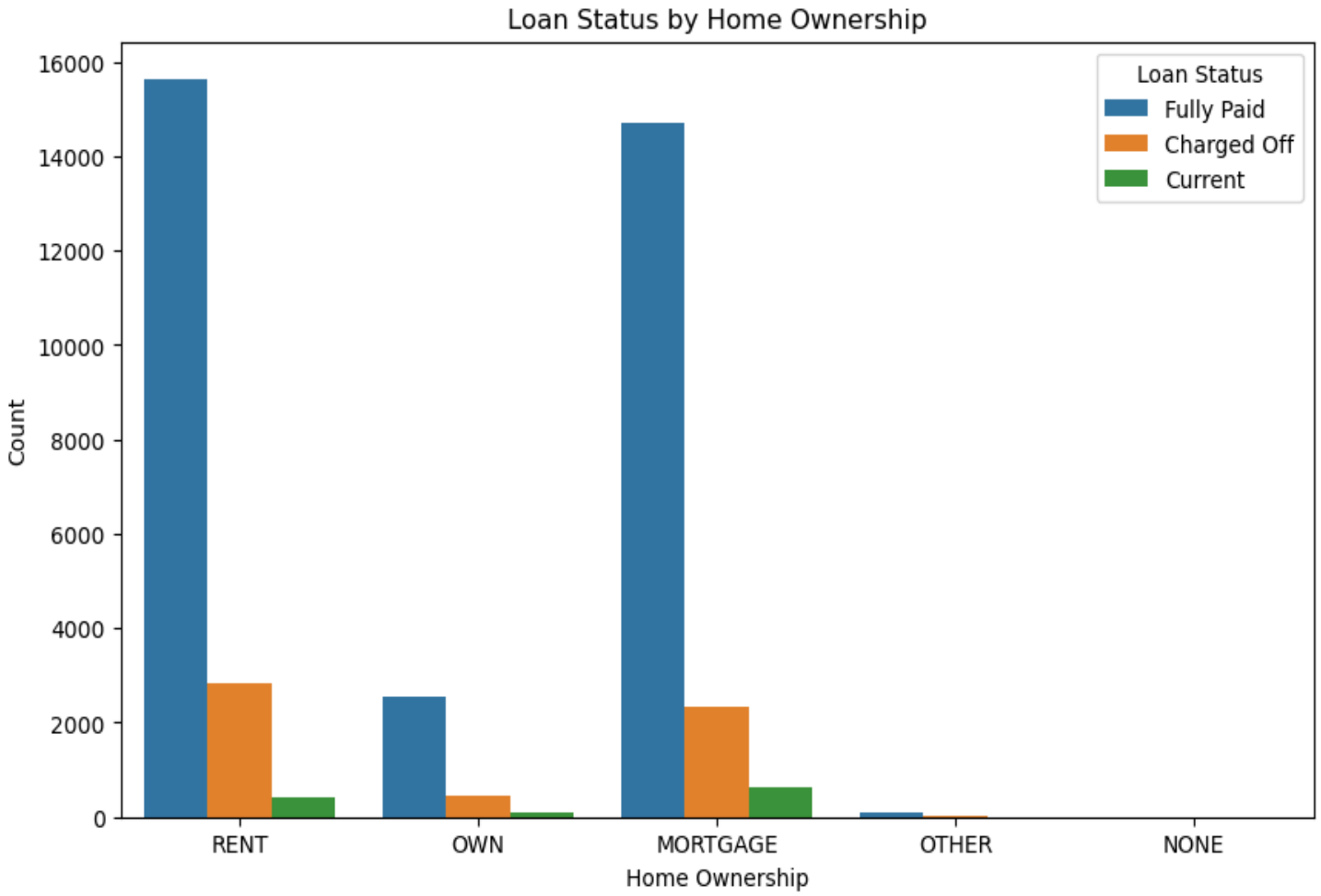


We use this histogram plot to visualize the distribution of interest rates and understand their spread across loans.

The interest rate has been mostly varying between 7.5% to 15% in most cases.

This graph does not take into consideration the type of loan so the range is higher than expected.

Analysis : Loan status by Home ownership

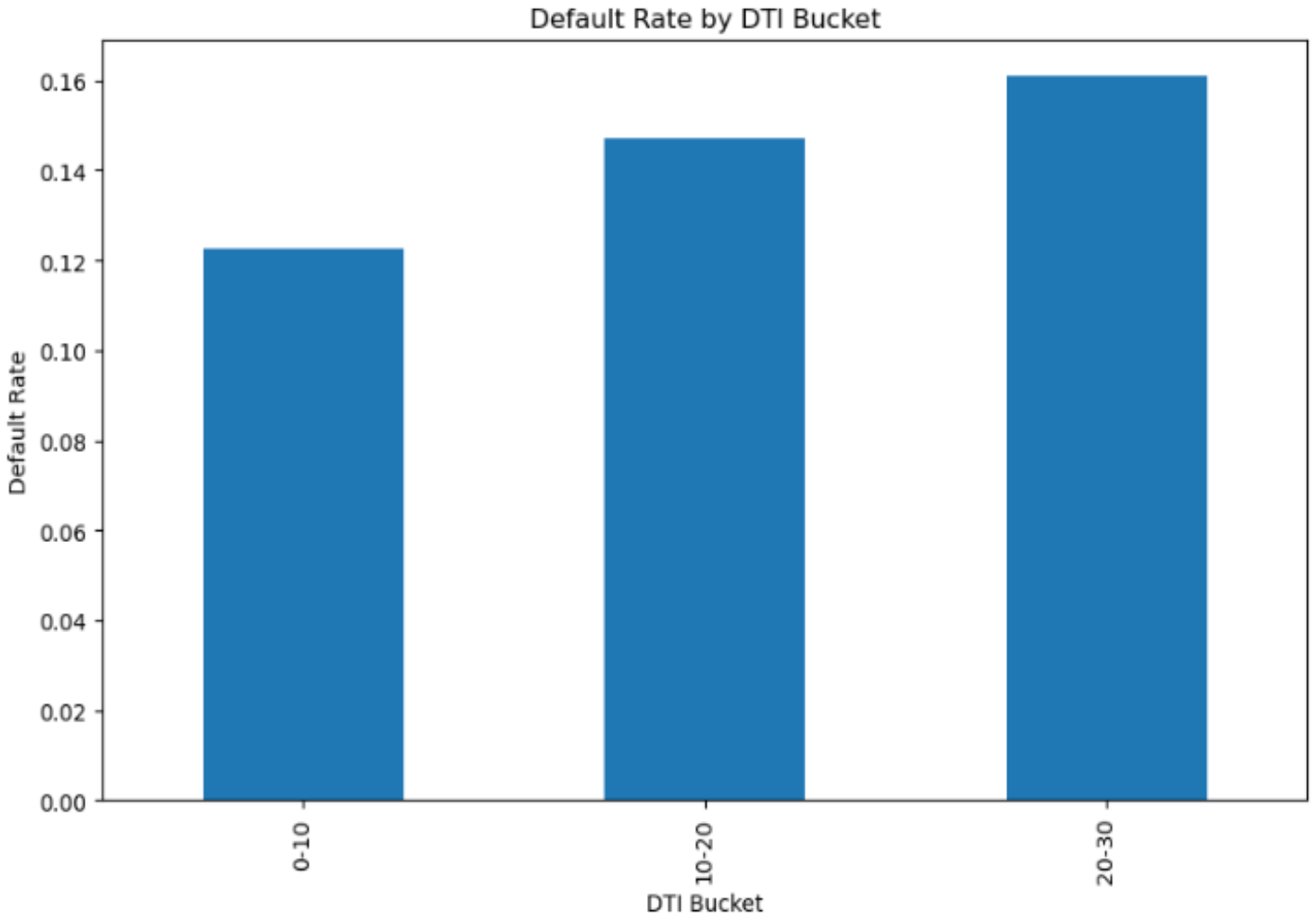


This histogram displays the correlation between loan status with home ownership.

Loan status distribution varies among different home ownership categories considerably.

We used seaborn countplot to categorize data based on ‘home_ownership’ and ‘color_code’ by ‘loan_status’

Analysis : Default rate by DTI bucket

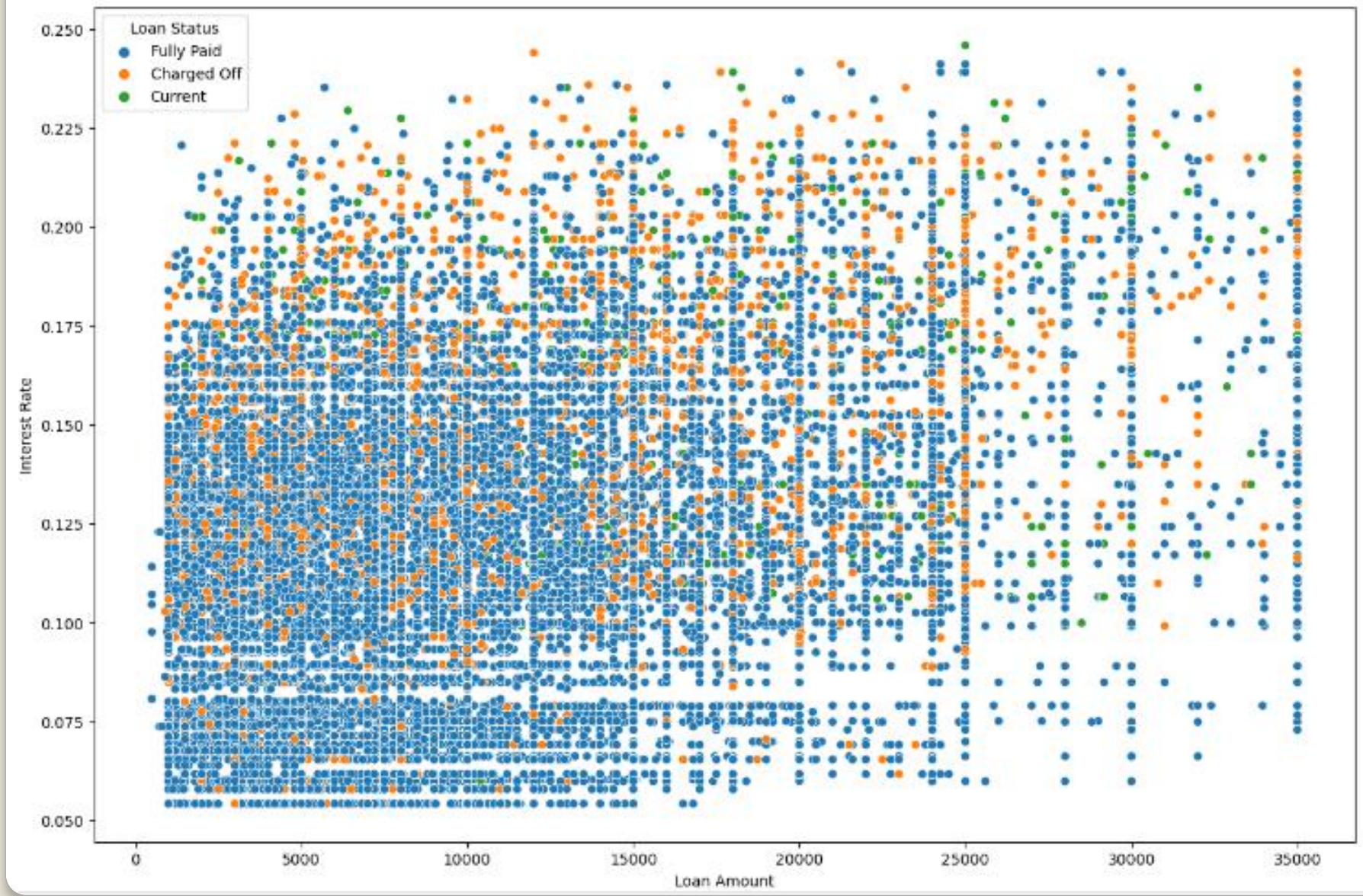


A great insight can be observed with the use of Derived metrics.

In this bar plot, I have decided to use bins that segment the DTI ratio into four categories for a more nuanced analysis.

Here we calculate the default rate within each DTI bucket. We're specifically looking at charged off loans. We are visualizing the default rate by DTI bucket more efficiently.

Analysis: Loan Amount vs Interest Rate by Loan status

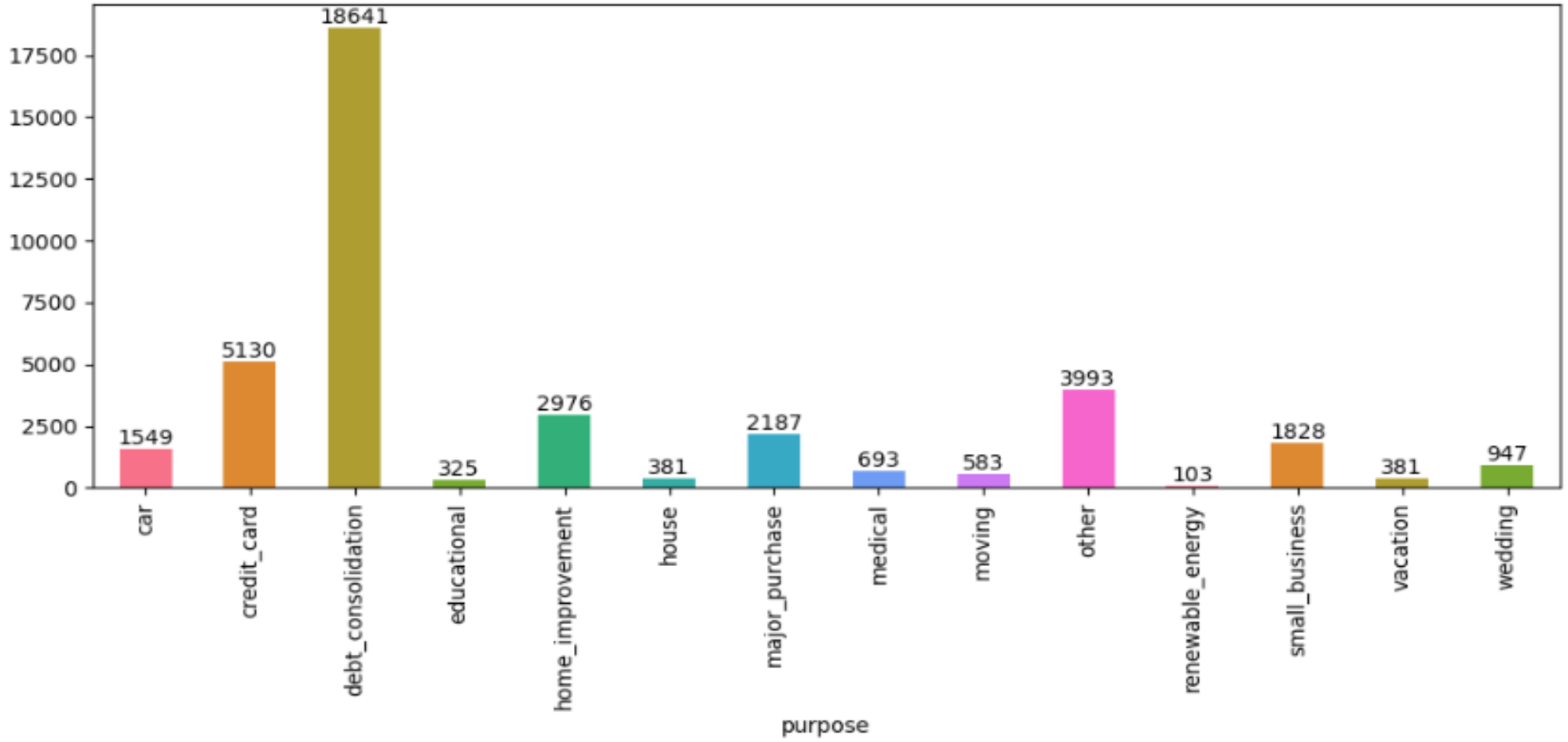


This scatter plot is being used to demonstrate the use of bivariate analysis.

Plot visualizes the relationship between the loan amount and interest rate.

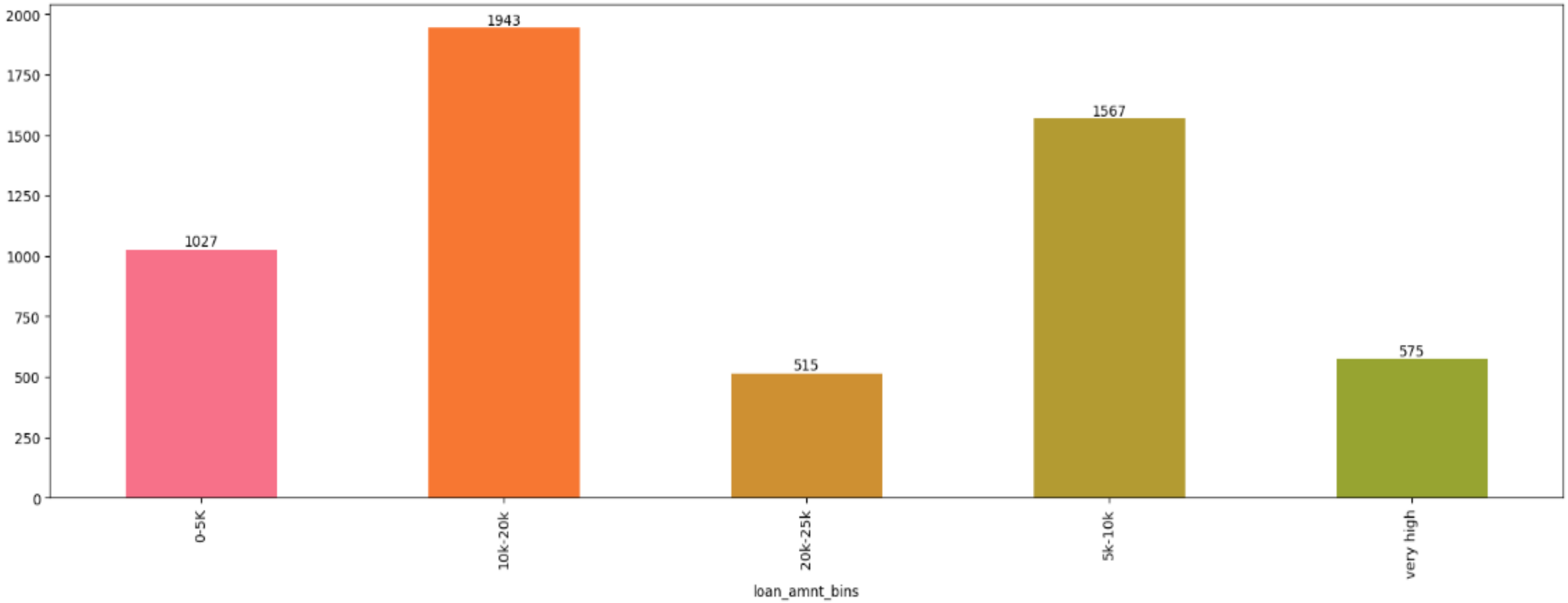
Interesting to see how different loan statuses are distributed acros the two variables.

Analysis : Loan purpose



Consumers who have interest rate in the range of 10-15% default the most

Analysis : Default rates across funded amounts on bins created



Visualizing using bar graph plot the default rates across funded_amnt based on bins created.

Insights:

1. **Debt-to-Income Ratio (DTI):** I've observed that higher DTI ratios are tied to an increased likelihood of loan default. Specifically, borrowers with DTIs above 30% demonstrate a significantly higher default rate than those with lower DTIs. This finding underscores the critical role that debt management plays in financial stability.

2. **Loan Amount:** I've also found a noticeable correlation between the amount of the loan and default rates. Particularly, larger loans, such as those exceeding ₹20,000, are more prone to default. This insight suggests that the size of the loan is a crucial factor to consider in risk assessment.

3. **Charged off loans:** 50% of the “Charged off” loan was due to Debt consolidation. The lending company needs to exercise caution when approving loans for debt consolidation purposes, as it has maximum number of defaults.

4. **Interest Rates:** Higher interest rates are linked to higher default rates. This indicates to me that borrowers subjected to higher interest rates face a greater risk of financial distress. It's a clear sign that interest rates are not just numbers but significant indicators of potential financial strain.

Insights (contd..) :

5. **Home Ownership:** Another interesting insight I've uncovered is that renters exhibit slightly higher default rates compared to homeowners. This leads me to believe that housing stability might play a role in a borrower's capacity to fulfill loan obligations.

6. **Employment Length:** Contrary to what one might expect, my analysis shows that the length of employment has minimal impact on default rates. This was surprising to me and suggests that job stability may not be as pivotal in determining loan repayment ability as previously assumed.

7. **Employment history:** Applicants who had been employed for 10 years or more than 10 years accounted for the highest number of "Charged off".

8. **Lower income borrowers:** The borrowers who are in lower income groups have maximum tendency to default the loan and it generally decreases with the increase in the annual income.

Recommendations:

1. Exercise caution when approving loans for debt consolidation purposes

2. Exercise caution before granting loans who stay in rented & mortgaged homes.

3. Stop approving loans to people with derogatory public records and bankruptcies record.

4. Exercise caution when approving higher loan amount to lower grade (particularly G,E,F) for longer terms.

5. Exercise caution while granting loans in states like California, Florida & New York.

6. Exercise caution while granting loan to borrowers who are in lower income groups (particularly in range of 20 -60k)

7. Exercise caution while granting loan for short tenure.

GitHub Repository Link:

https://github.com/thenakulgupta/LendingClubCaseStudy

Thank You