EDA Lending Club Case Study

EDA : Bank Loan Default Risk Analysis

PROBLEM

- We have to use EDA to analyze the insights present in the dataset to ensure that the applicants who are capable of repaying the loan are not rejected.
- When the company receives a loan application, the company has to decide for loan approval based on the applicant's profile.
- Two types of risks are associated with the bank's decision:

➢If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
➢If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

- The company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

SOLUTION

- Understand the driving factors (or driver variables) behind loan default
- By leveraging Pandas analytical capabilities, this method facilitates data-driven strategies tailored to specific business objectives, fostering efficiency and accuracy in decision-making processes.
- The study adopted both qualitative (case study) and quantitative methods respectively

APPROACH

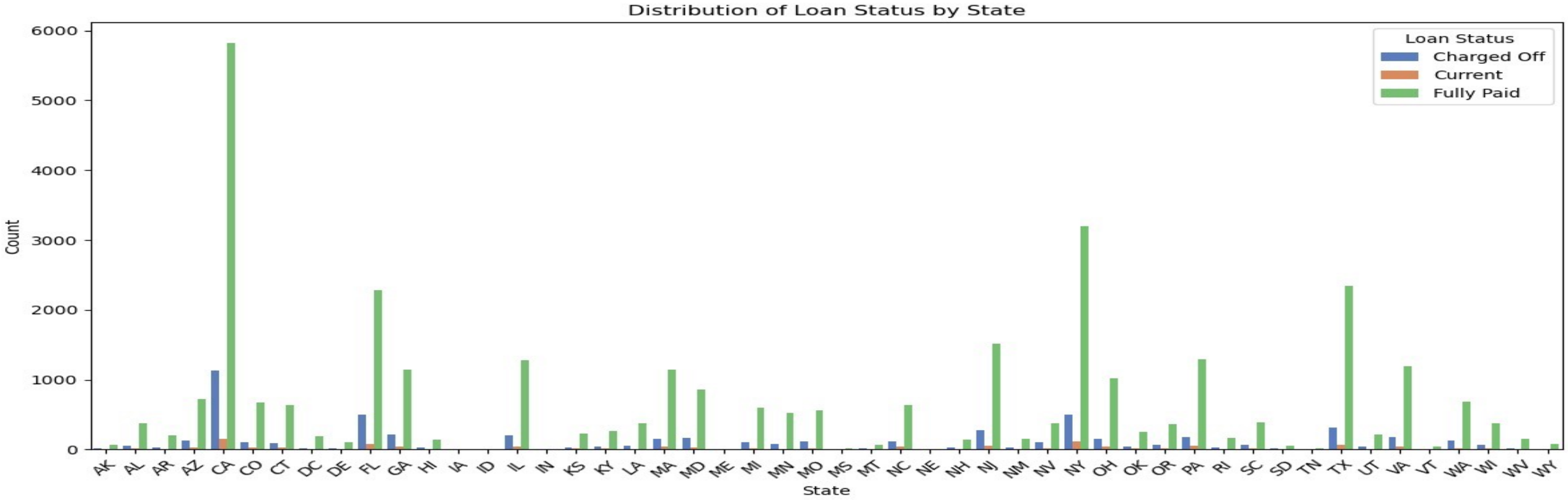
- Find insights into how variables influence each other and the relative prevalence of different categories, empowering informed decision-making.
- Utilizing correlation analysis and percentage breakdowns from each column of the dataset in pandas provides a comprehensive understanding of inter-column relationships and category distributions.

Loan default can be defined as the inability of a borrower to fulfil his or her loan obligation as at when due

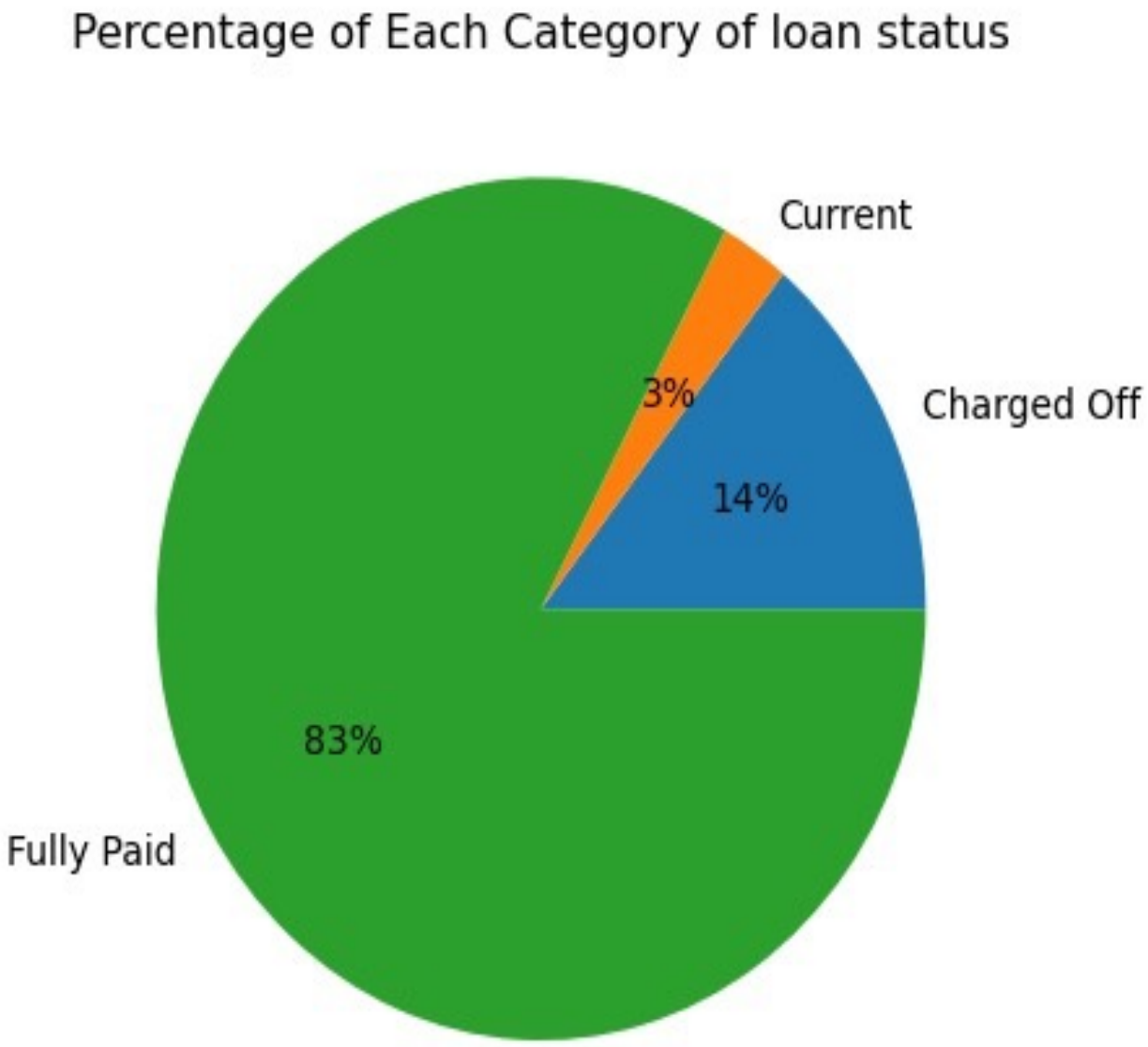
"Borrowing money on what promises repayment is the most common fraud."

- Edward George Earle Lytton Bulwer-Lytton

Analyzing Data

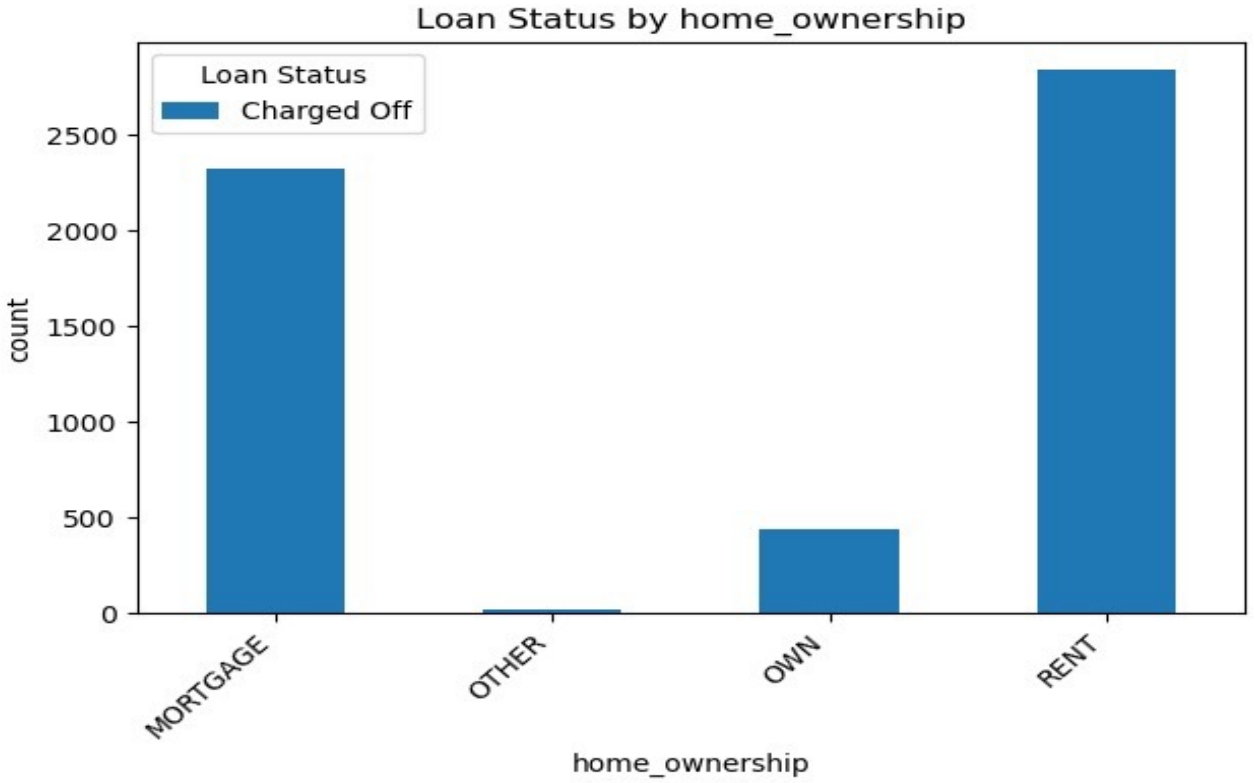


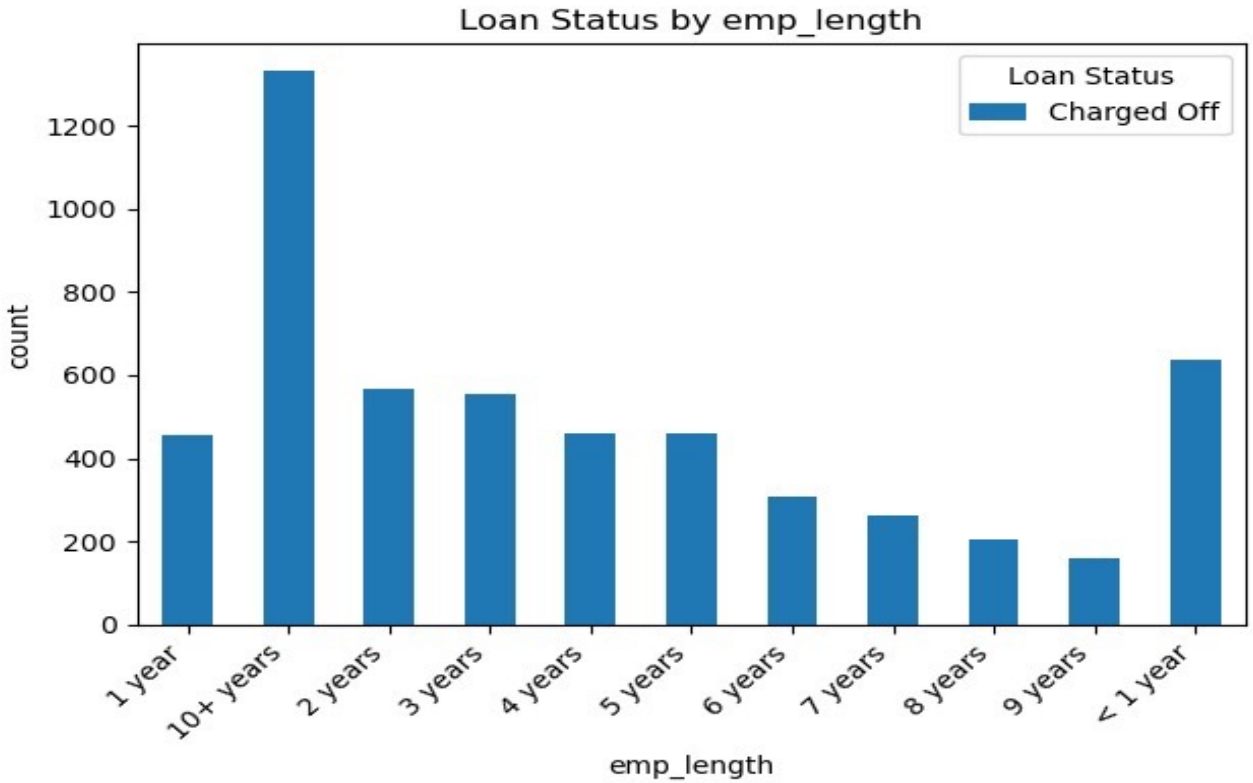
Inference :In our dataset, there appears to be a positive correlation between population size and the number of bank defaulters. This suggests that regions with larger populations tend to have a higher incidence of bank defaulters. Further analysis would be needed to determine the underlying factors contributing to this relationship, such as economic conditions, unemployment rates, or demographic characteristics



- Approximately 15% of loans end in charged offs, indicating a notable proportion compared to those fully and currently paid.

Univariate Analysis

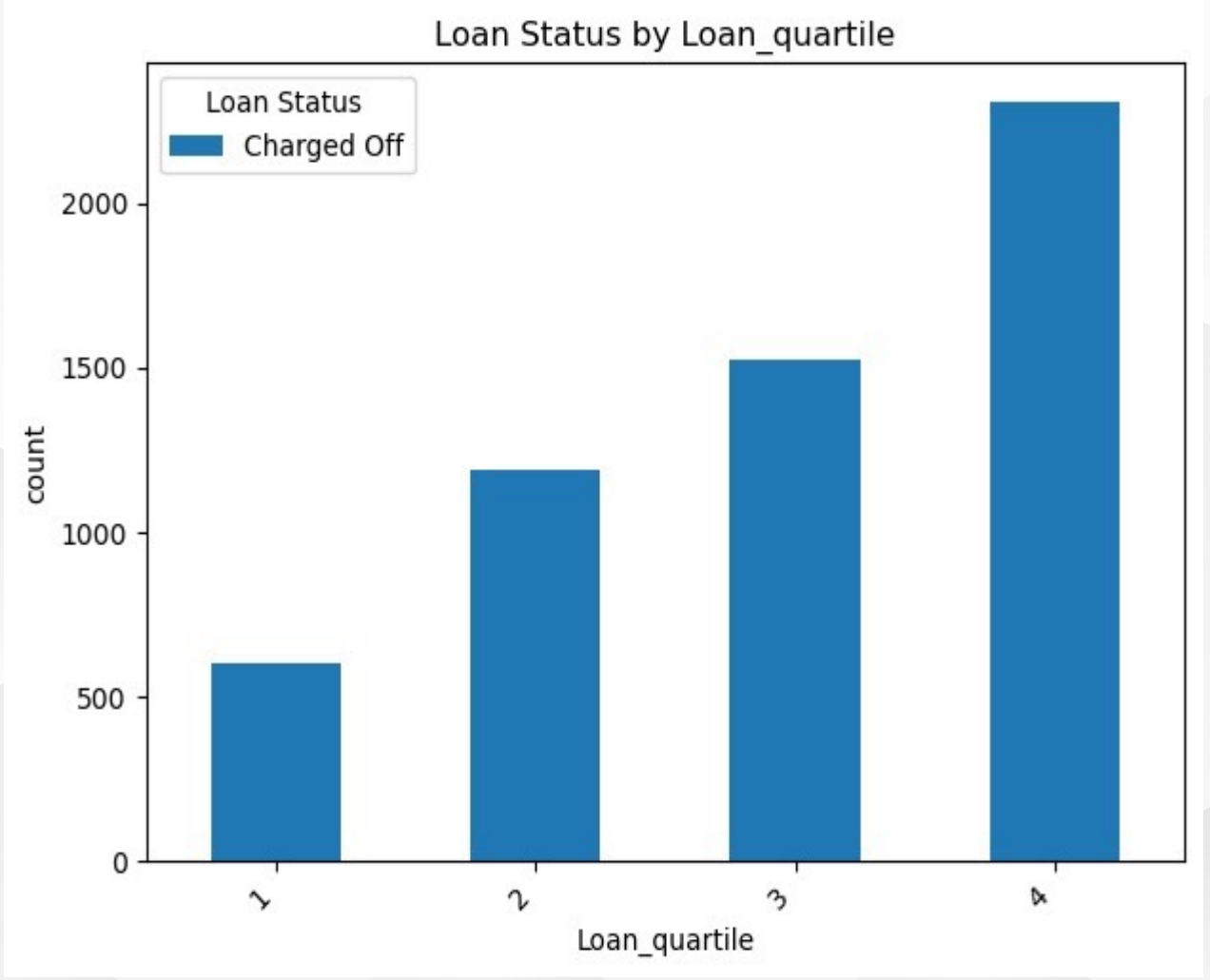


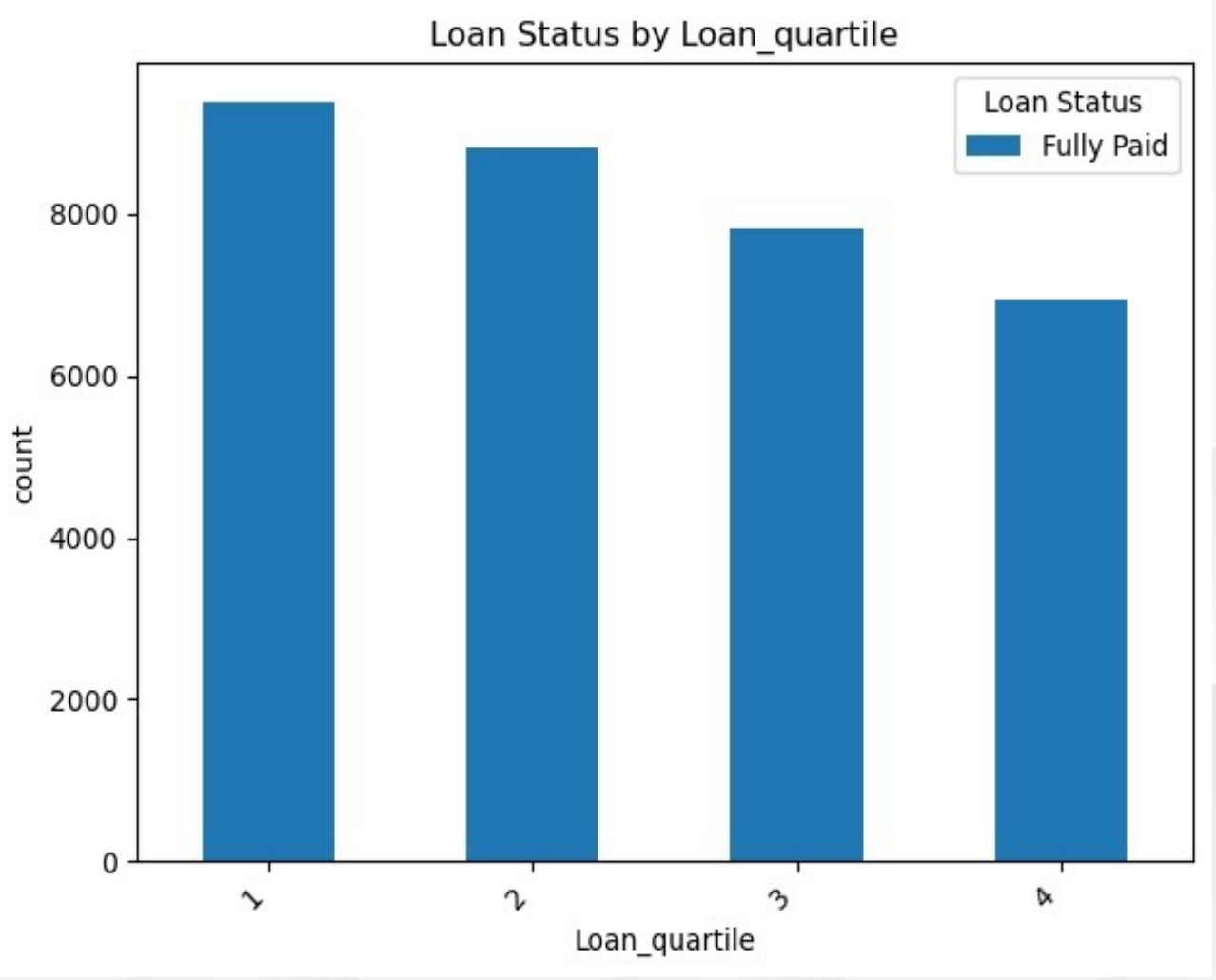


- The borrowers who have their home ownership as Mortgage and Rent are the prime defaulters of the loan

- Borrowers who are employed less than 1 year are more prone to be defaulters.
- As the Emp length increases the charged off rate decreases.

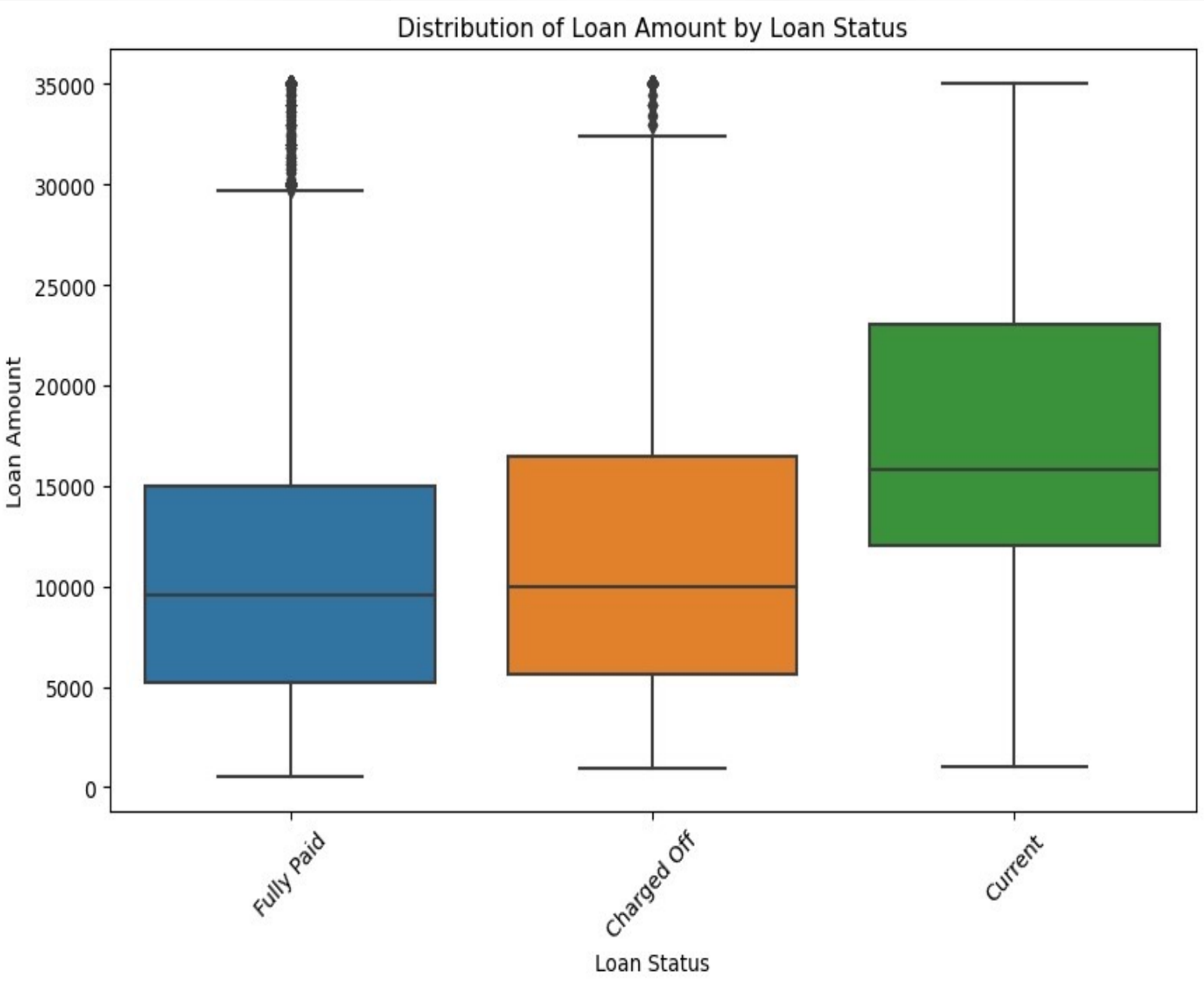
*10+ years is a cumulative count so ignoring that





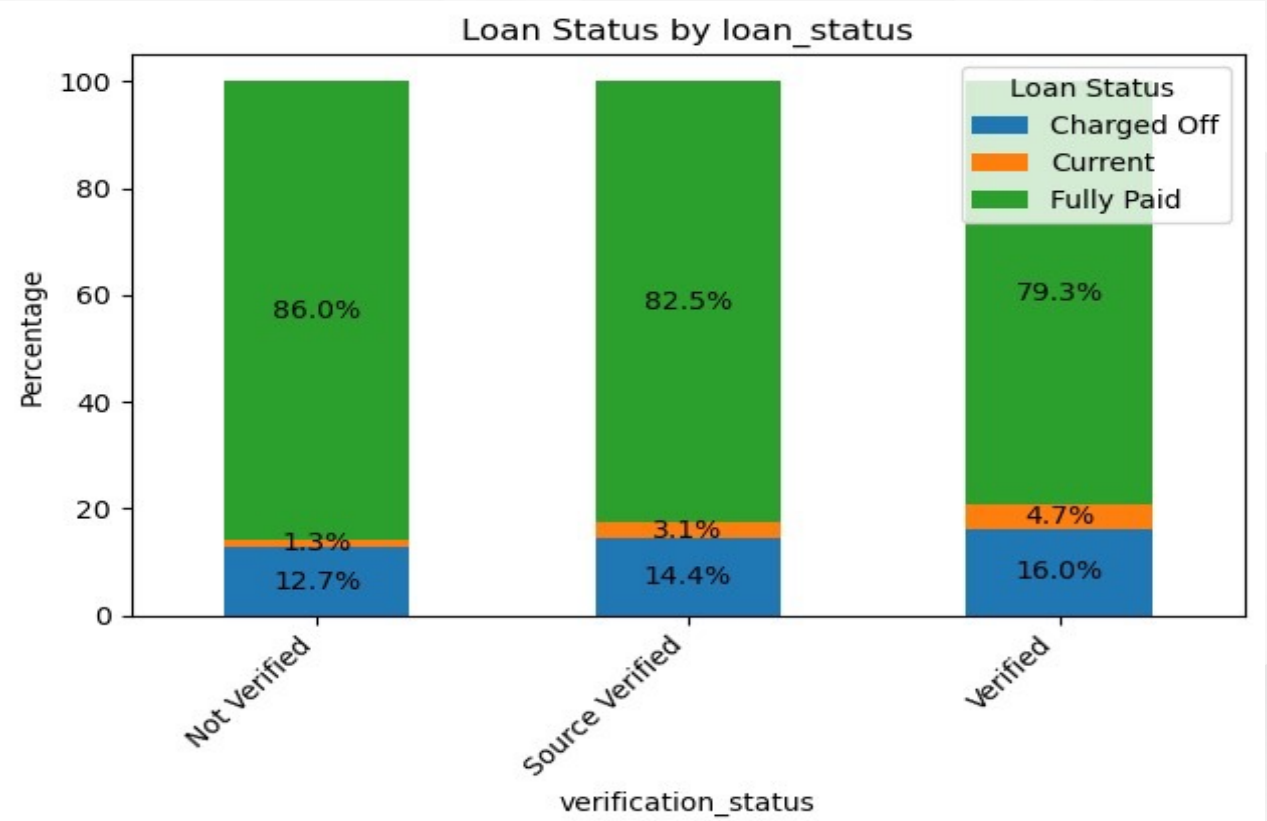
- In our analysis, we observed a trend where borrowers were more likely to default on their loans when faced with higher interest rates. This suggests a significant correlation between interest rates and loan charge-offs, highlighting the impact of borrowing costs on borrower repayment behaviour.

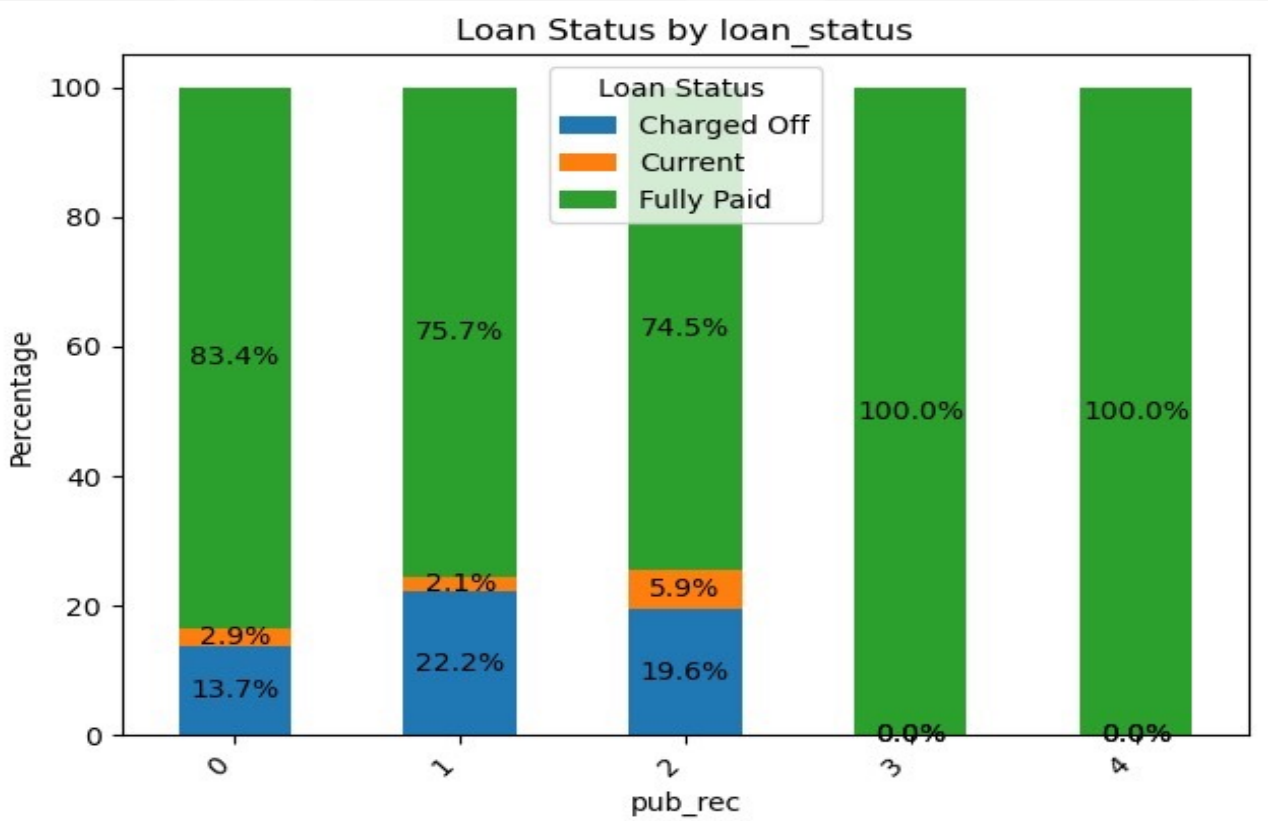
- * Interest rate increases with each quartile
  4>3>2>1



- In our investigation, we found no discernible impact of loan amount on loan status.
- This implies that loan defaulters are present across a range of loan amounts, indicating that the likelihood of default is not significantly influenced by the size of the loan.
- Therefore, other factors beyond loan amount may play a more prominent role in determining loan default rates.

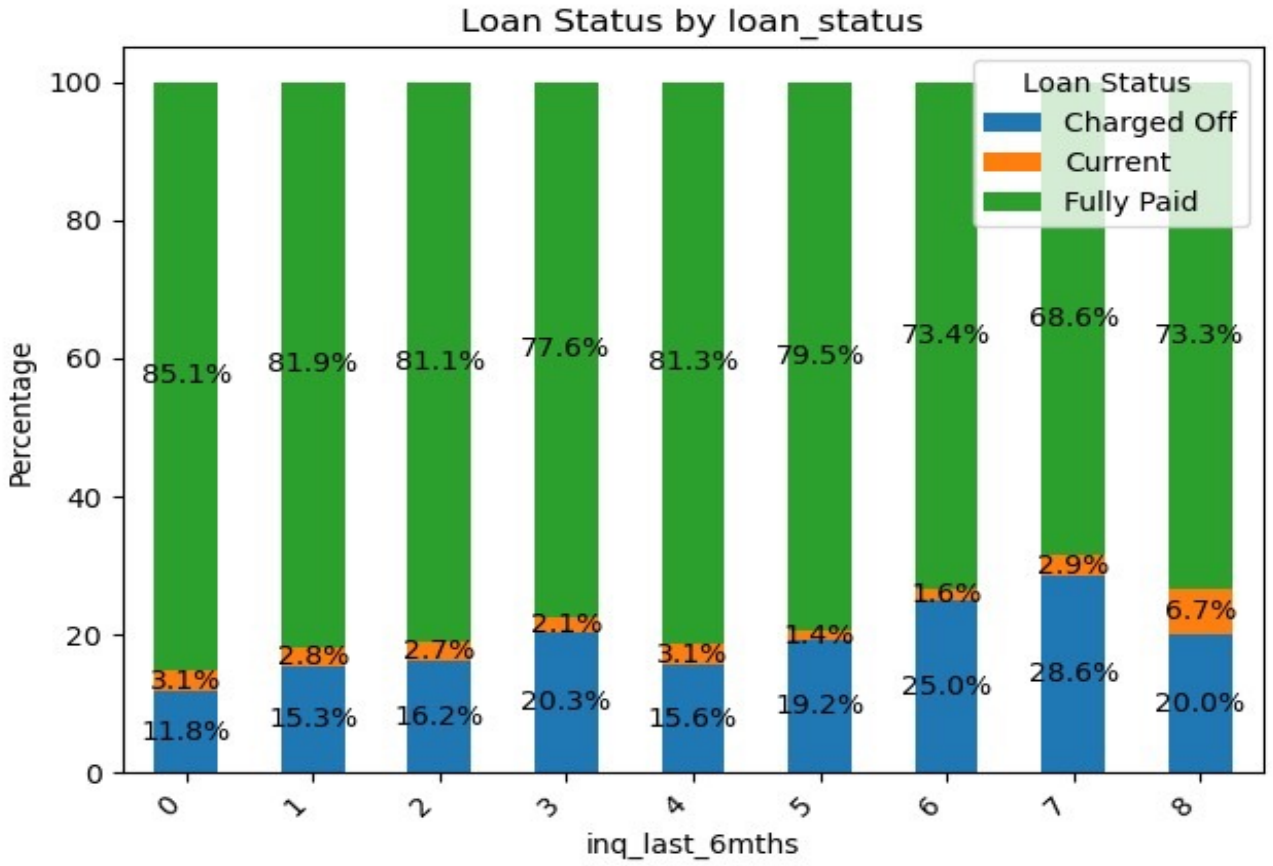
Bivariate Analysis

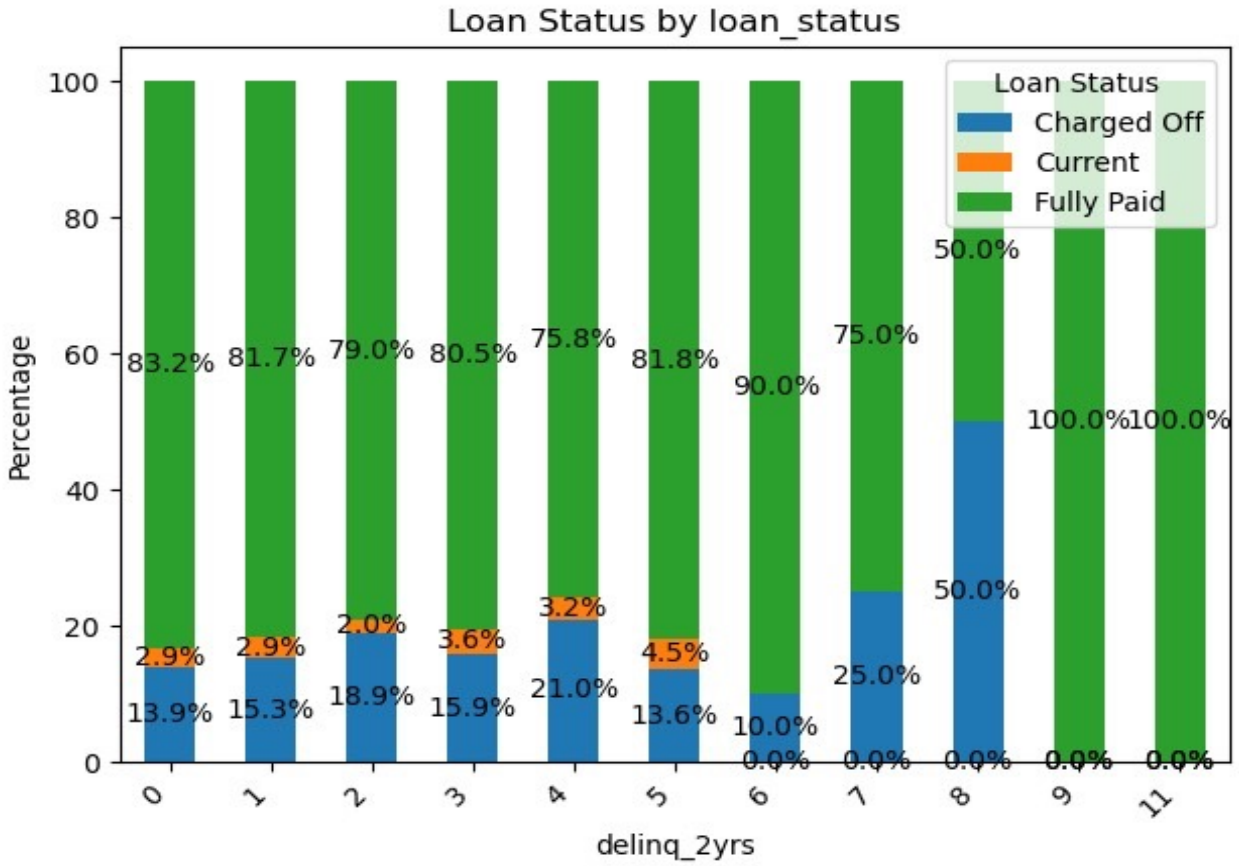




- The verification status of the borrower does not significantly impact the loan status. This implies that whether borrower information was verified or not has little correlation with loan default rates.

- The presence of public records (“pub_rec”) with values of 0, 1, and 2 is associated with a higher frequency of defaulters, while records with values of 3 and 4 are associated with zero defaulters.

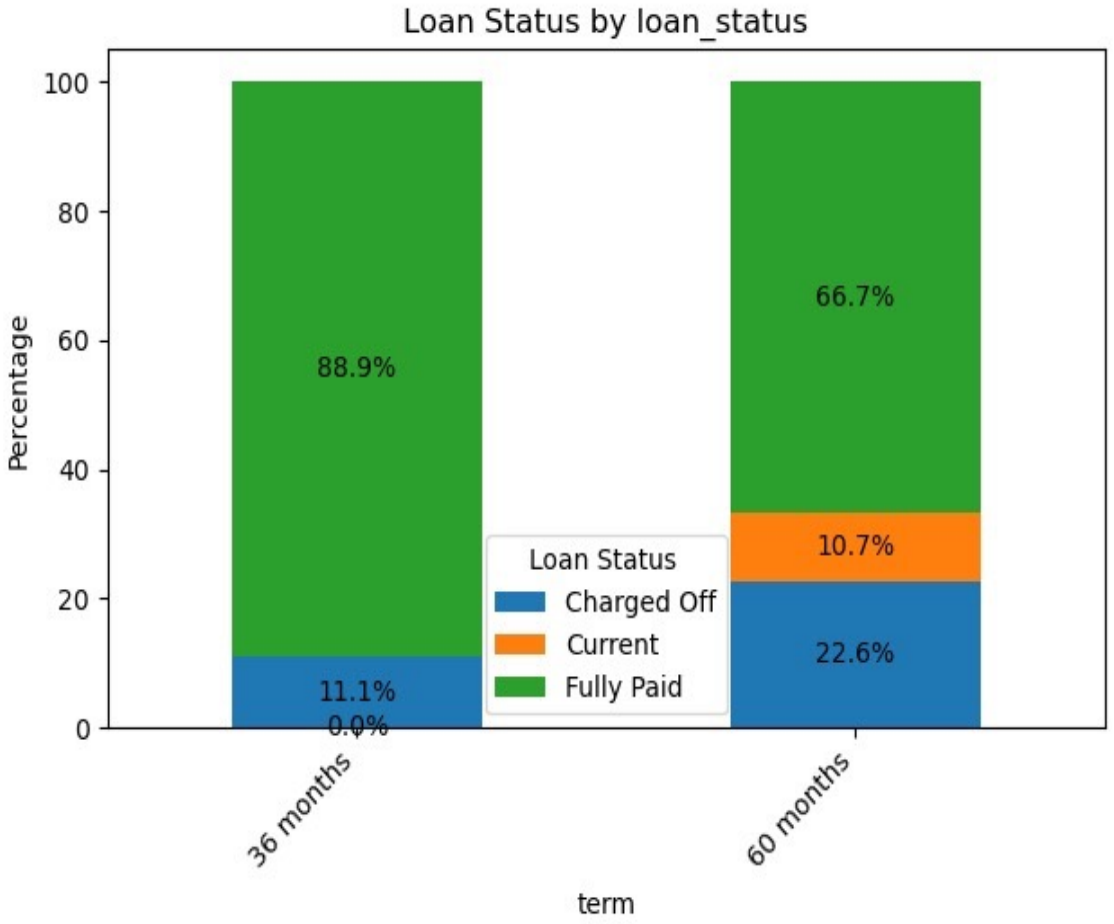


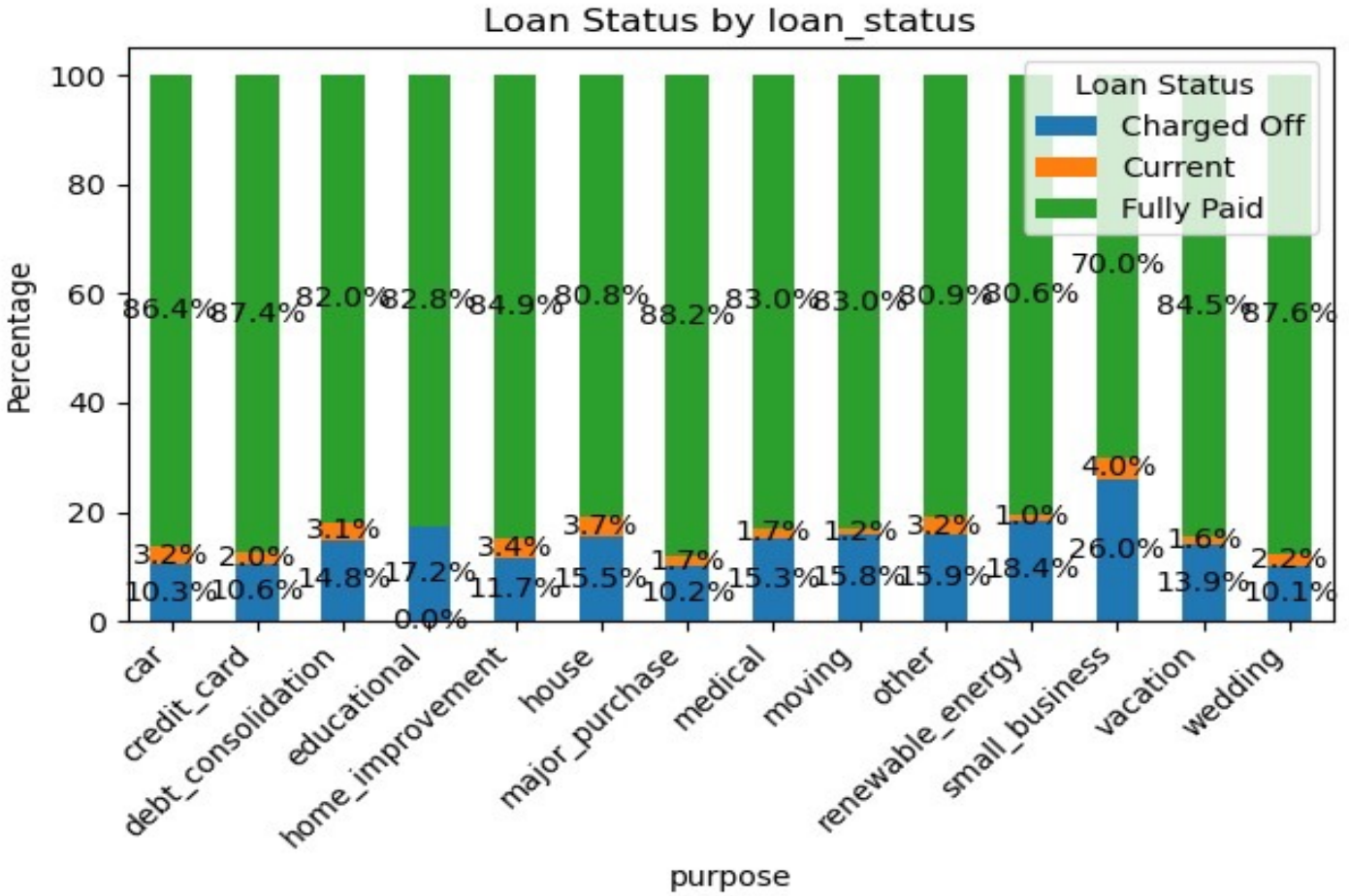


The observed variation in trends may be because of the following reasons:

- Seasonal Variations
- Interest Rate Changes
- Marketing or Promotional Activities
- Financial Planning or Budgeting.

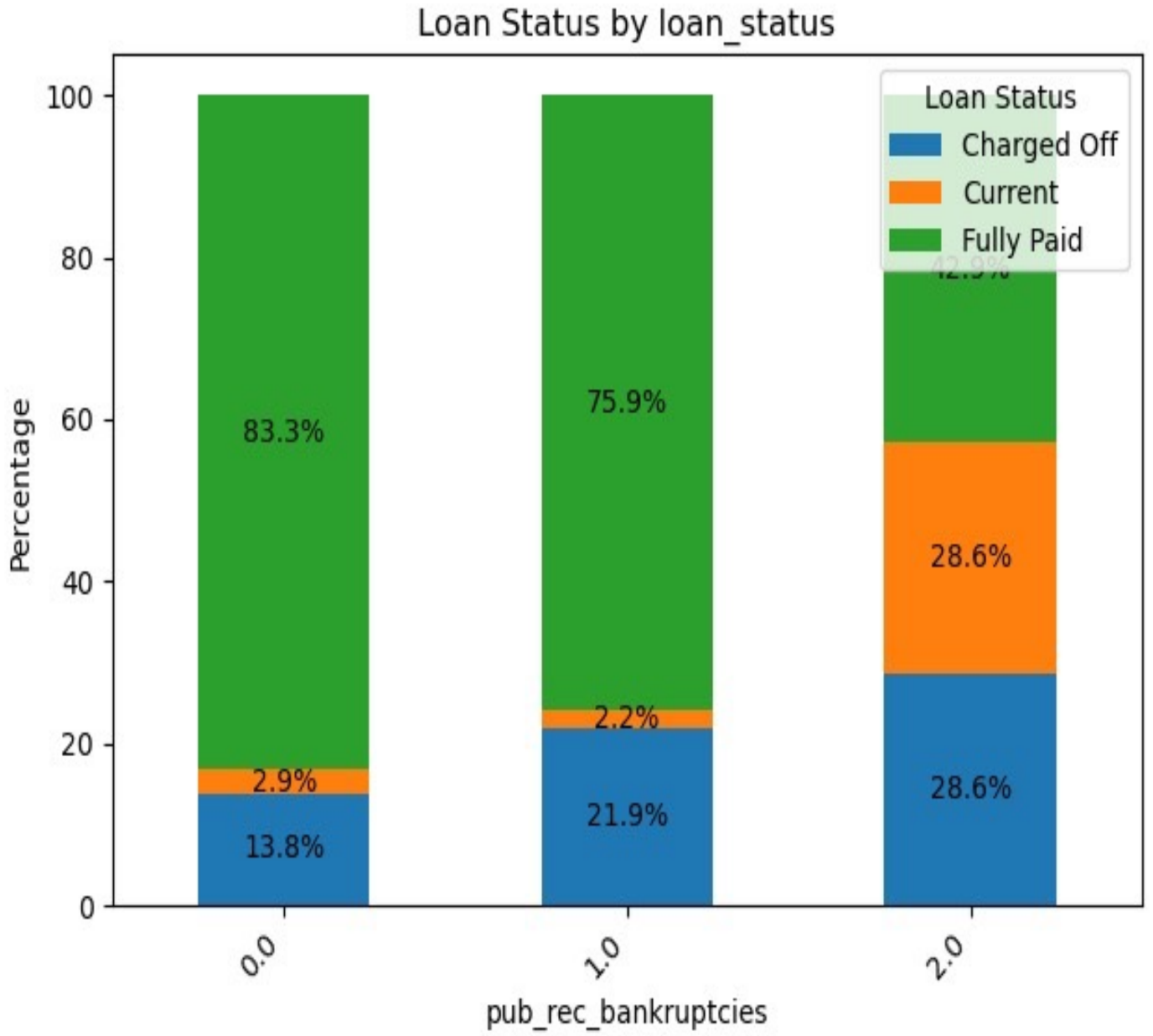
- Loans are more likely to be fully paid when the delinquency count in the borrower's credit file for the past two years exceeds eight instances.





- As observed, an extended loan term is positively correlated with an increased occurrence of defaulters.

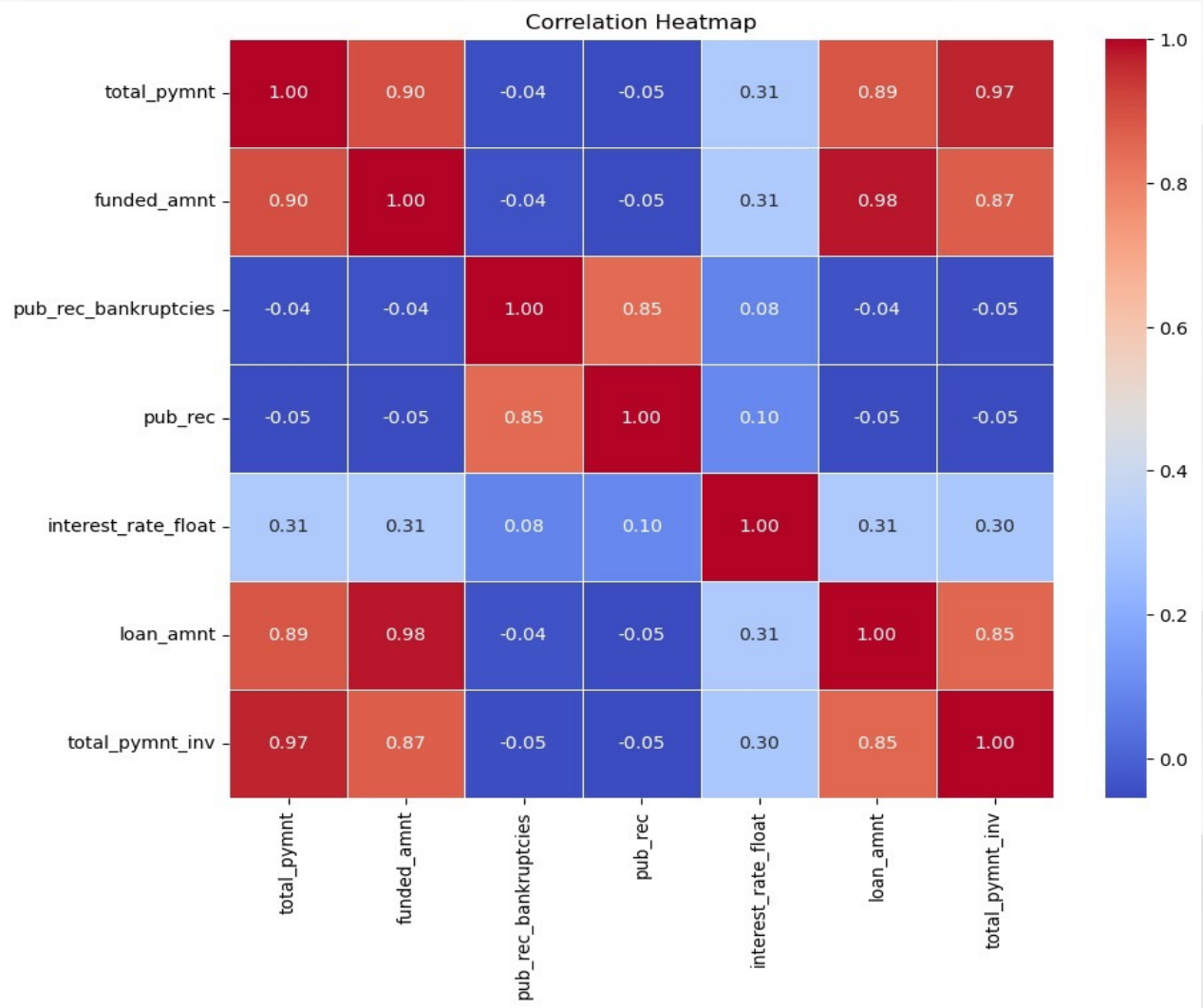
- The percentage of charged-off loans associated with small businesses is notably higher compared to other categories. Hence, it is advisable to exercise caution when considering borrowers seeking loans for small business purposes.



- The relationship between public bankruptcies and loan charge-offs is notably pronounced: as the frequency of public bankruptcies rises, we observe a corresponding increase in the rate of charged-off loans.

- This suggests a significant association between these two factors, indicating that a higher incidence of public bankruptcies tends to coincide with a greater likelihood of loan defaults.

Correlation Analysis



Inferences drawn from correlation Heat Map:

1. 'total_pymnt' exhibits a strong positive correlation with 'funded_amnt' and 'loan_amnt', as well as with 'total_pymnt_inv'.

2. Additionally, 'total_pymnt' shows a positive correlation with 'interest_rate_float'.

3. Conversely, there is a slight negative correlation between 'total_pymnt' and 'pub_rec', as well as 'pub_rec_bankruptcies'.

4. It can be concluded that 'total_pymnt' and 'total_pymnt_inv', followed by 'loan_amnt' and 'funded_amnt', are highly correlated."

Conclusion:

1. Charge-offs represent approximately 15% of loans, a notable proportion compared to fully and currently paid loans.
2. Strong correlations exist between 'total_pymnt' and 'total_pymnt_inv', as well as 'loan_amnt' and 'funded_amnt'.
3. Borrowers with less than one year of employment are more likely to default, while longer employment lengths correlate with lower default rates.
4. Higher interest rates are associated with increased borrower default rates.
5. Loan defaults occur across various loan amounts, indicating independence from loan size.
6. Longer loan terms are linked to higher default rates.
7. Small business loans have a higher charge-off percentage compared to other purposes, suggesting caution in lending.
8. Charge-off rates increase with higher instances of public bankruptcies.
9. Borrowers with pub_rec values of 0, 1, and 2 have more defaulters, while values of 3 and 4 have zero defaulters.
10. Mortgage and rent-based home ownership are associated with higher loan default rates.

Thank you

Overall, our analysis reveals valuable insights into the factors influencing loan defaults. By understanding the correlations and trends identified, we can make more informed lending decisions to mitigate risks and optimize loan portfolio performance. Moving forward, leveraging these insights can aid in developing more effective lending strategies and improving overall loan portfolio management.