Lending Club Case Study

Group Members:

- Moumita
- Jashwanth

Introduction

- The consumer finance industry faces risks when deciding to approve or reject loan applications based on applicant profiles.
- Two major risks are associated with loan decisions:

  1. **Loss of business** if a reliable borrower is denied a loan.
  2. **Credit loss** if a borrower defaults after approval.
- The dataset contains past loan applicant data, focusing on those who defaulted, which is critical for identifying high-risk borrowers.
- The analysis aims to identify factors that indicate whether a borrower is likely to default, helping reduce credit losses.
- By identifying risky applicants, the company can:

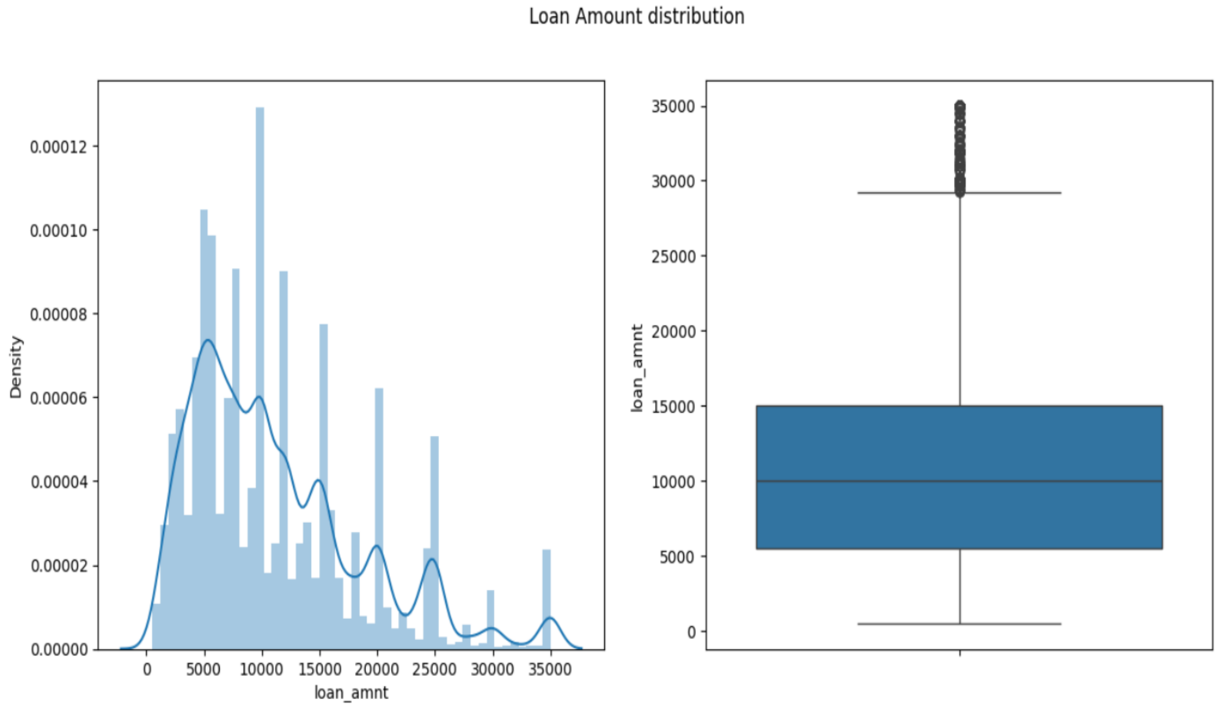
  1. Reject or minimise high-risk loans.
  2. Offer higher interest rates for riskier borrowers

Problem Solving Methodology:

- **Understanding the features and data**:
- Explored the dataset to comprehend the features and their significance.
- **Cleaning the data**:
- Handled missing values, and outliers, and ensured data consistency for accurate analysis.
- **Univariate analysis**:
- Analyzed the frequency distribution of each feature to understand individual characteristics.
- **Segmented univariate analysis**:
- Performed targeted univariate analysis to explore feature distributions within specific categories.
- **Bivariate analysis**:
- Examined the relationships and correlations between two features to identify patterns.
- **Drawing inferences**:
- Derived insights based on the analysis to identify key factors influencing loan default.

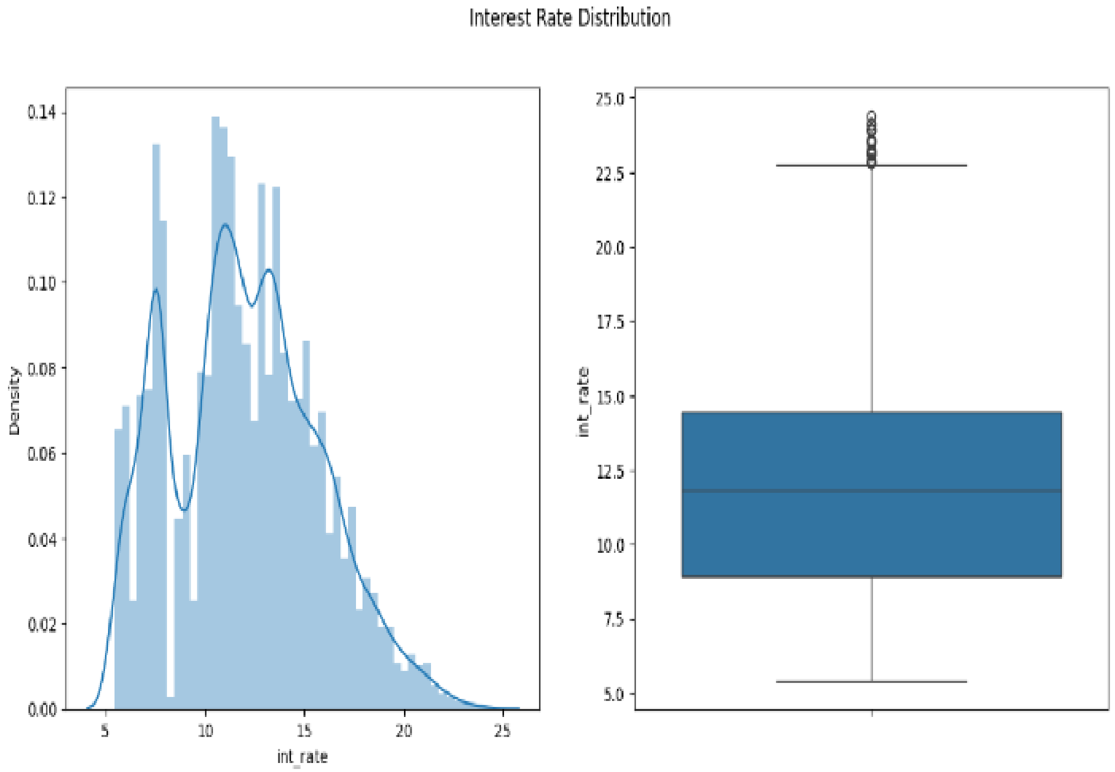
Inference:

- As we see, the distribution of loan amount follows the Gaussian curve which is right-skewed.
- Majority of the customers are having a loan amount of around 10000 and the rest of the people are also having more than 10000



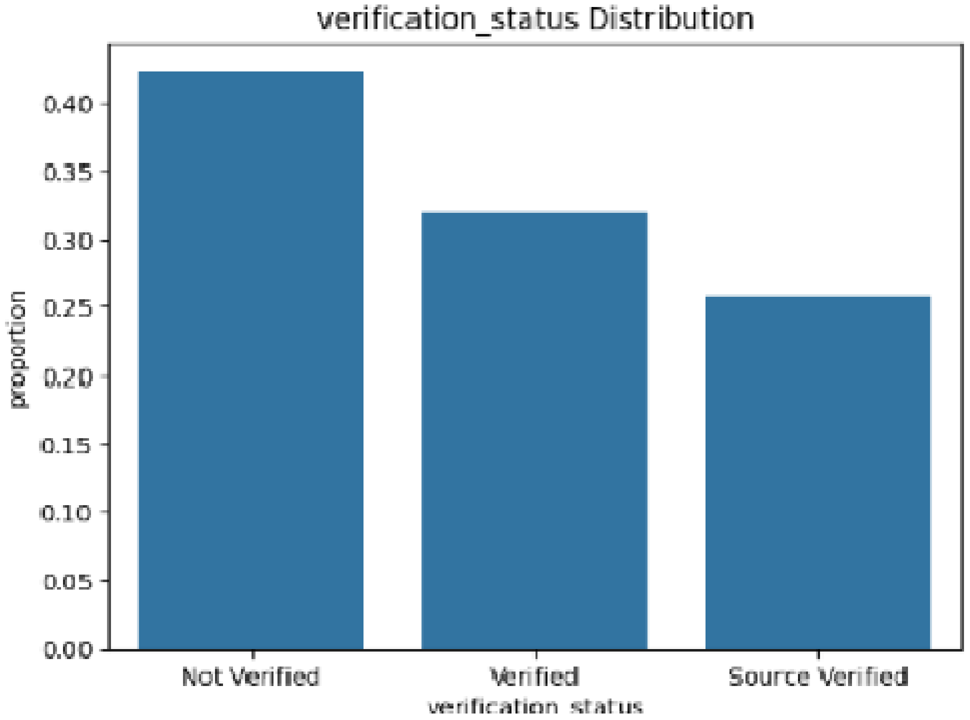
Inference:

- Based on the analysis of the two plots, we can conclude that the majority of interest rates fall within the range of 9% to 14.5%, with a median rate of approximately 12.5%.
- A small number of borrowers, however, took loans with significantly higher interest rates, reaching up to 22.5%.



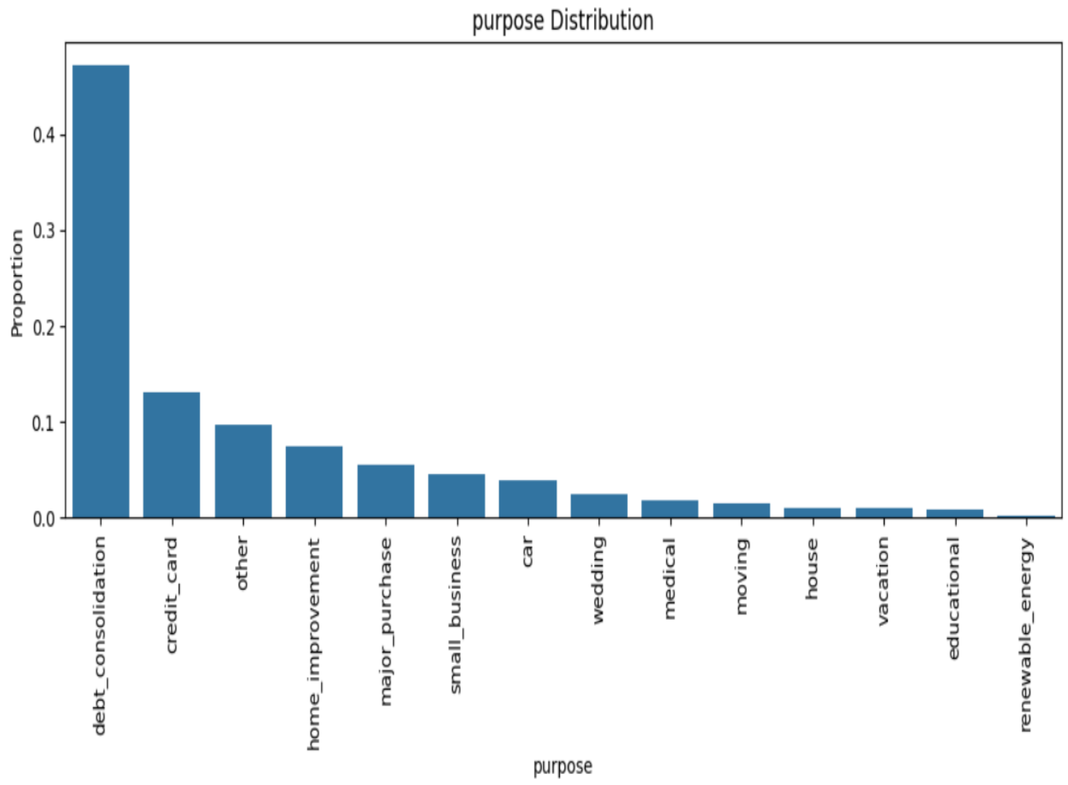
Inference:

- As we can observe, the majority of the customer got their loan sanctioned from Lending Club with proper background verification.
- More than 40% of the customers were grated the loans without verification, which can cause the more defaulters and credit loss to the company



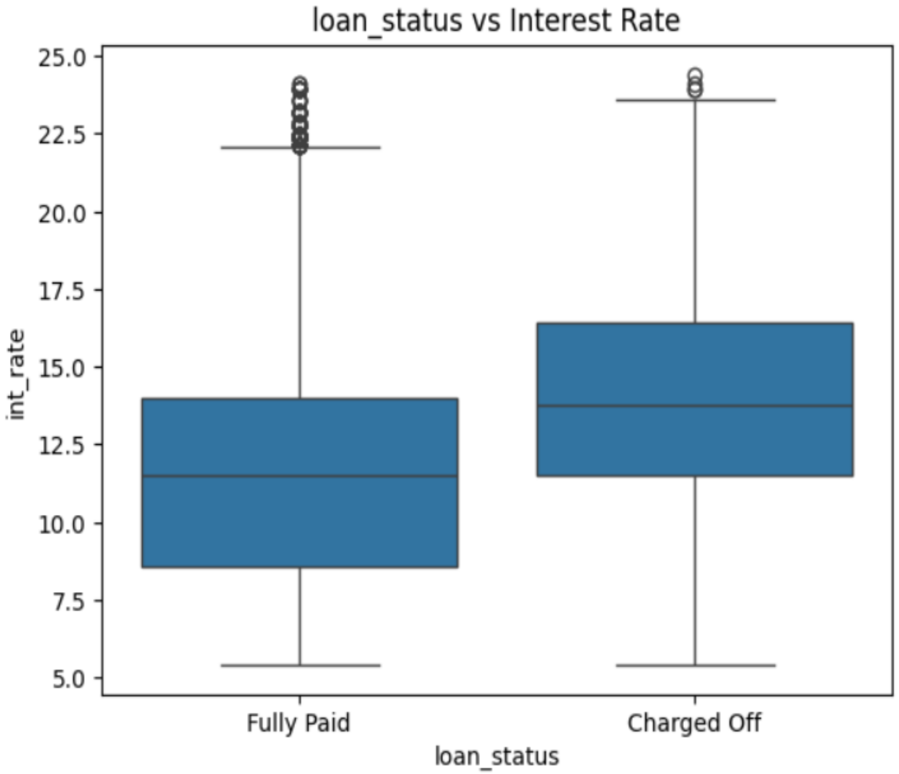
Inference

- Customers who are having the loan purpose as debt consolidation have the higher take followed by the credit card holders and home improvements



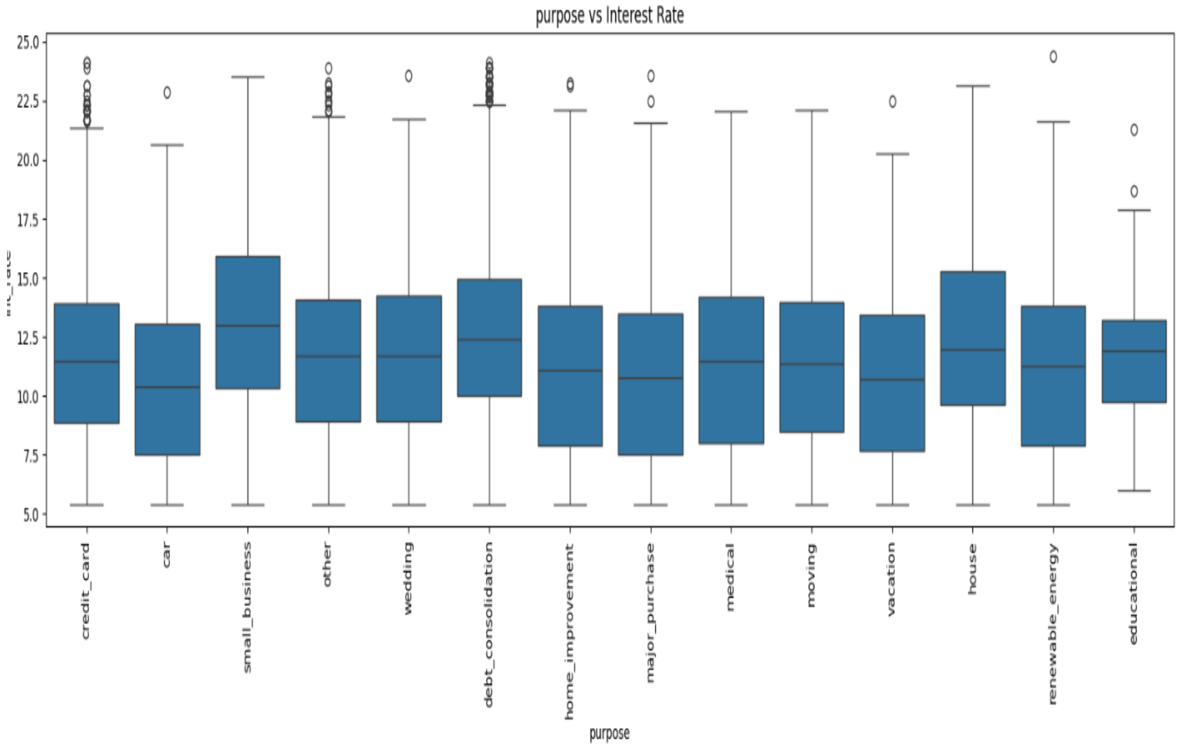
Inference:

- We can infer that the customers who have high interest rates have been prone to default on the loan with most of the interest rates being around 14%
- While the customers who are having lesser interest rates are able to clear the loans.



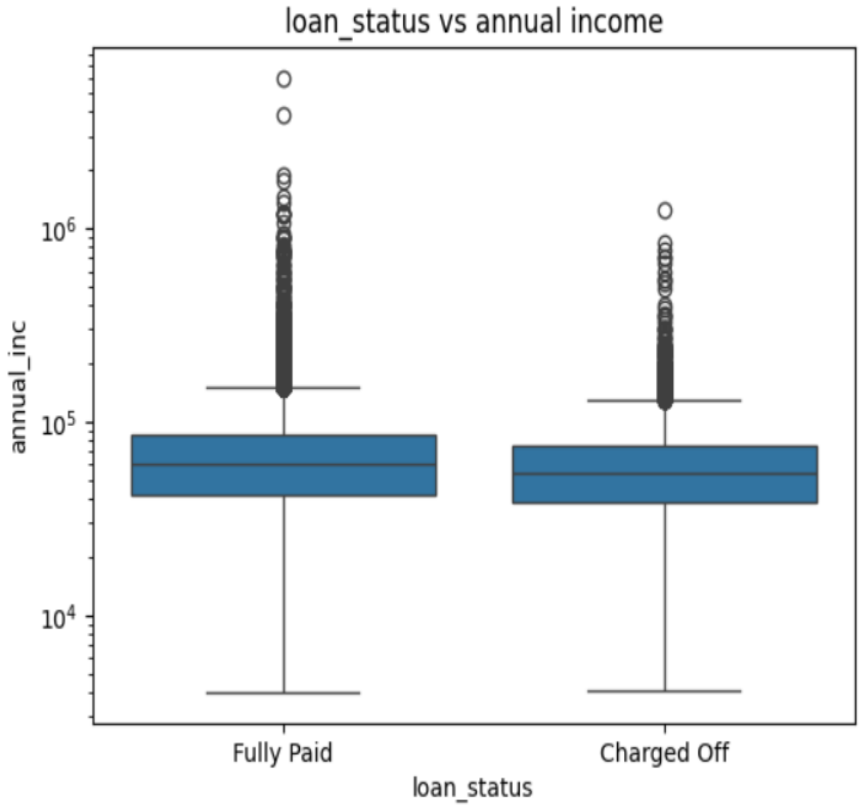
Inference:

- Customers for the loan purposes like small business, debt consolidation and house loans have higher interest rates than others.
- Also, a good volume of customers have taken the higher interest loans who are having debt consolidation and credit card payments etc



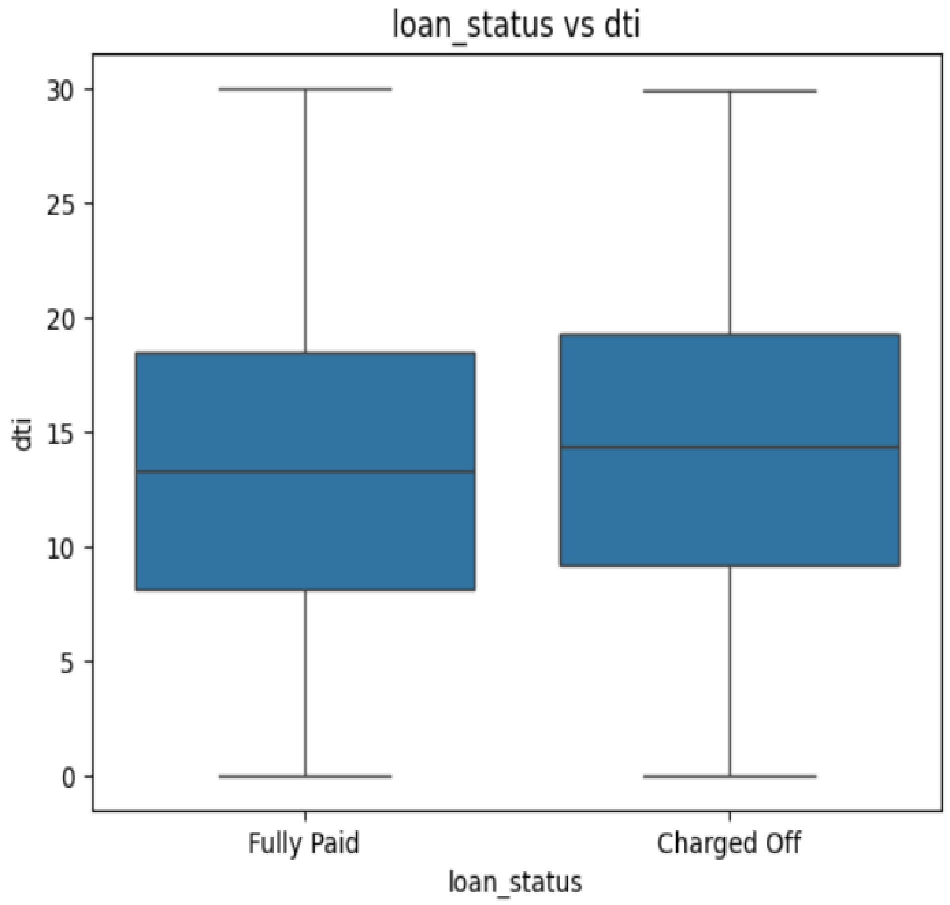
Inference

- Customers having a higher annual income can successfully repay the loan.
- While customers who have less annual income are prone to default on the loan.



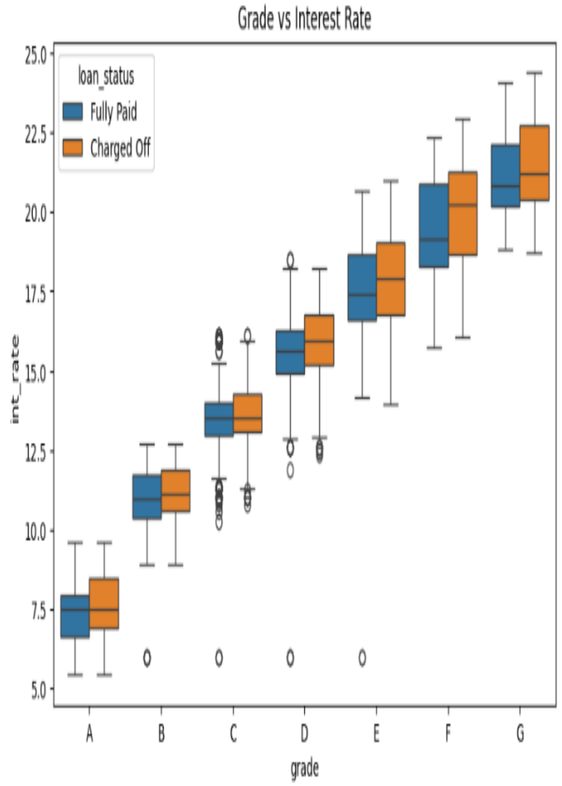
Inference:

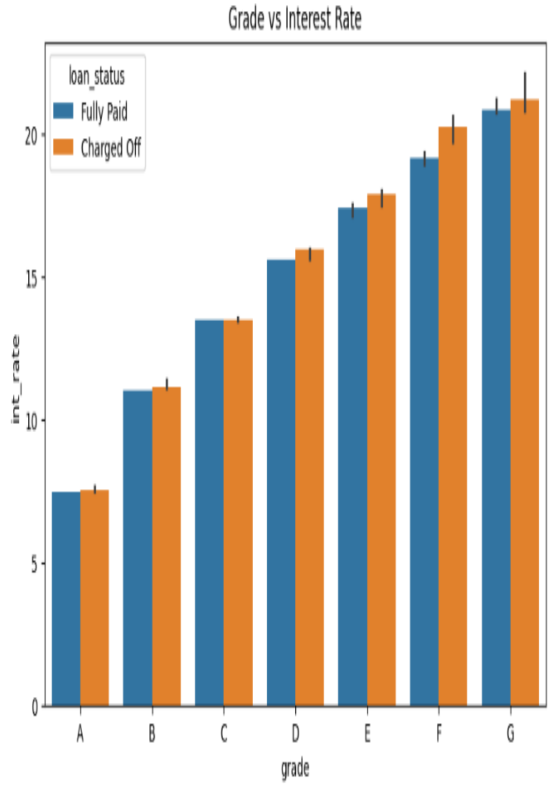
- Although it’s not too significant and evident from dti feature, but there is a slight indication that customers having more dti have more chances to default on the loan.
- Customers having relative lesser dti are able to close the loan payment successfully.



Inference:

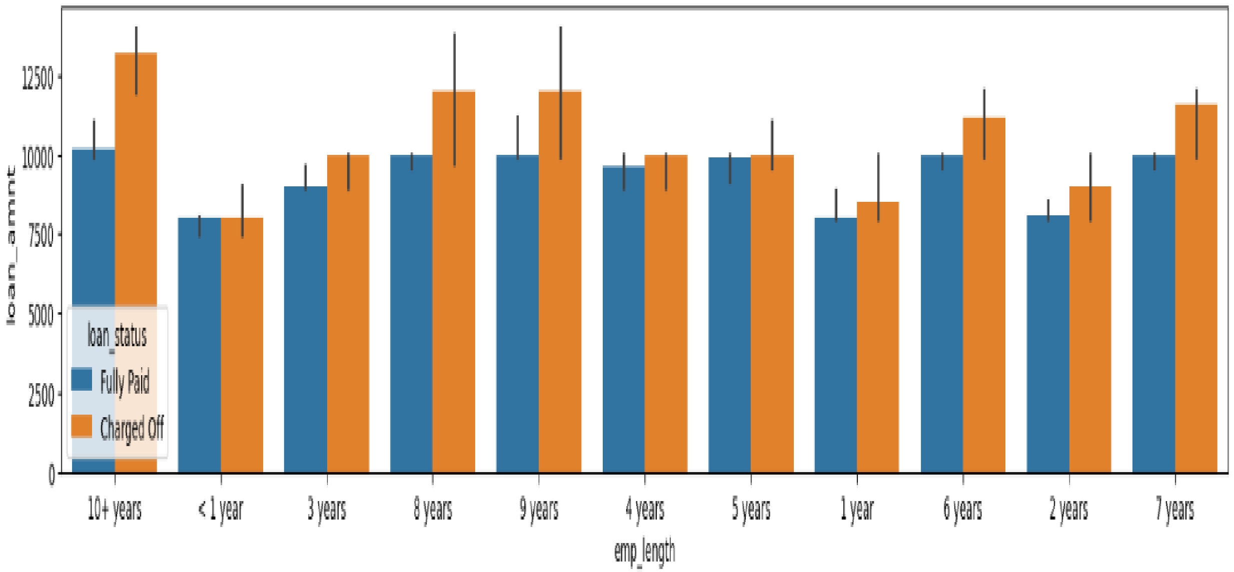
- As the grade of the customer increases, the interest also linearly increases with in turn leads to a higher rise in the defaulter’s chances.
- Customer having A, B and C grades can clear the loans with fewer loan default rate





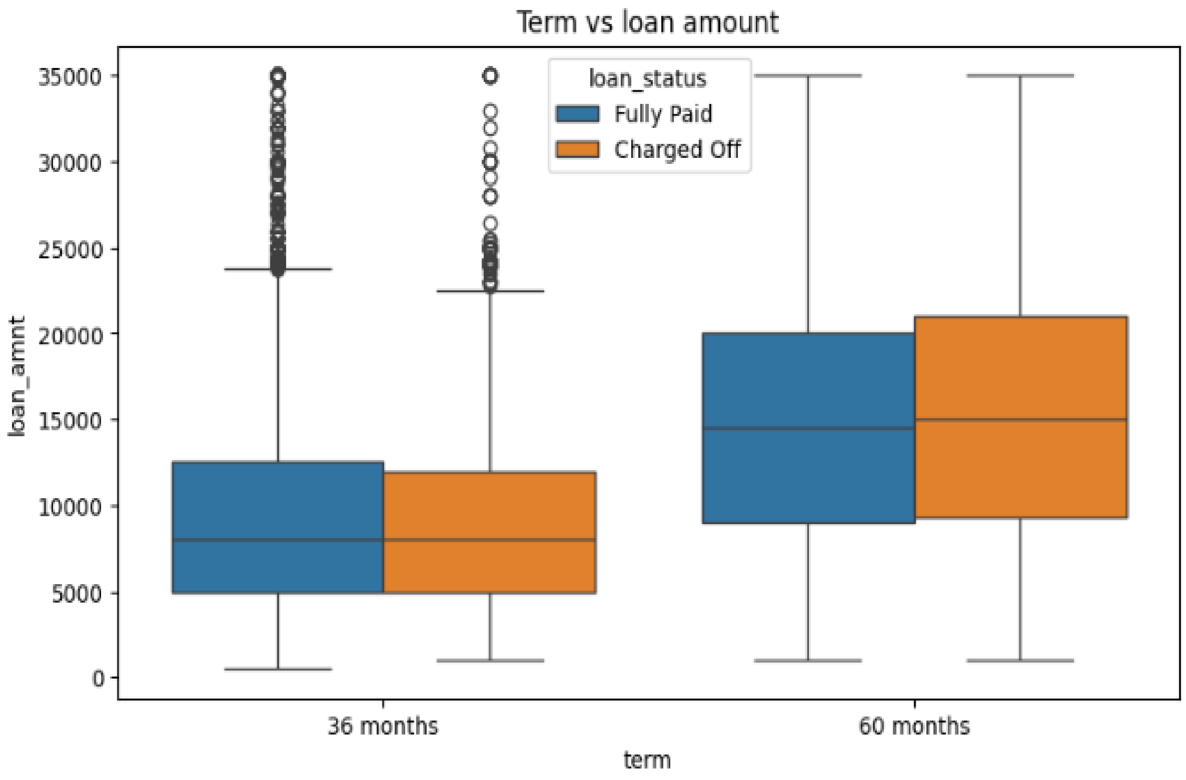
Inference:

- Customers with high employment length have defaulted the loan more than less employment lengths.
- Shorter loan tenures make is yielding the good loan recovery and good credit maintenance for the company.



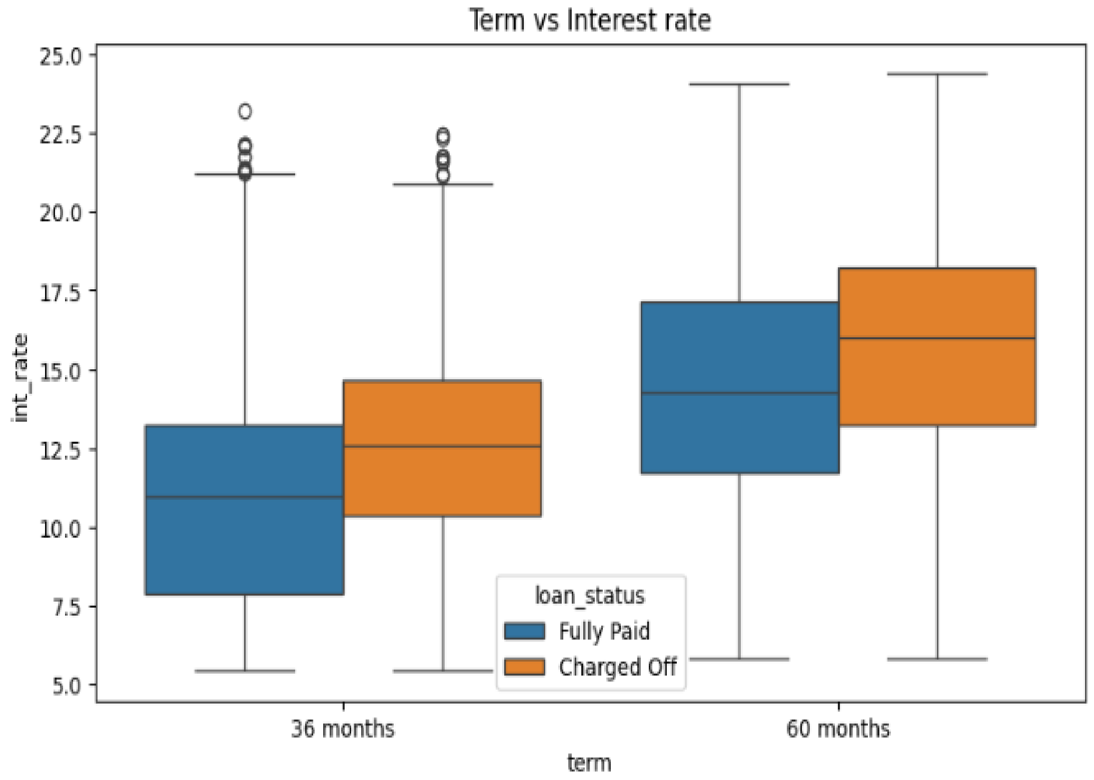
Inference:

- Customers with short term were able to clear the loan amount effectively.
- The customers who are having higher term loan payments are subjected to default on the loan.



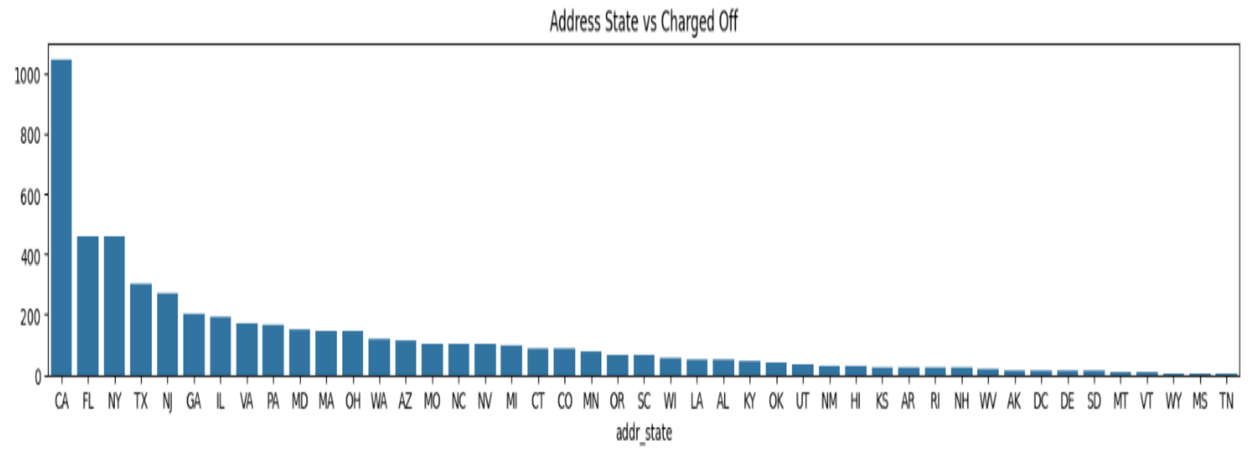
Inference

- Customers who opted to clear the loan for higher terms are subjected by the company to pay the higher interest rates.
- This has counter-fired as many of the customers have defaulted on the loan because of the higher interest rates and higher term.



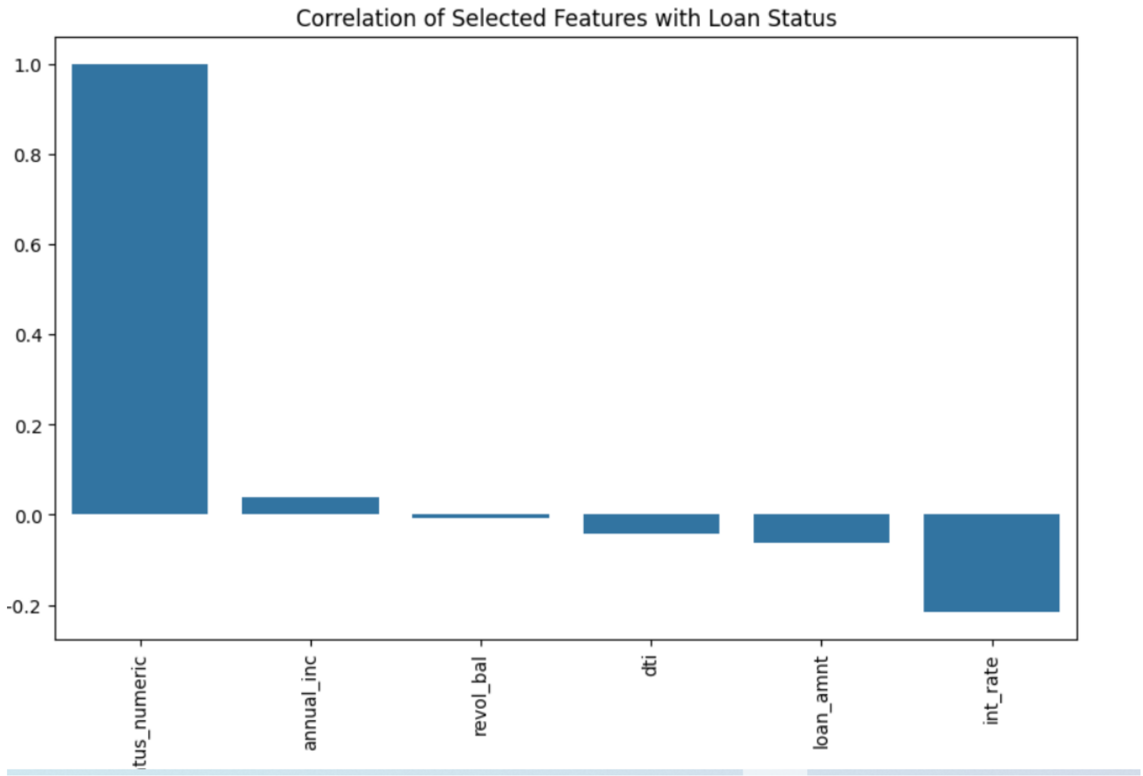
Inference

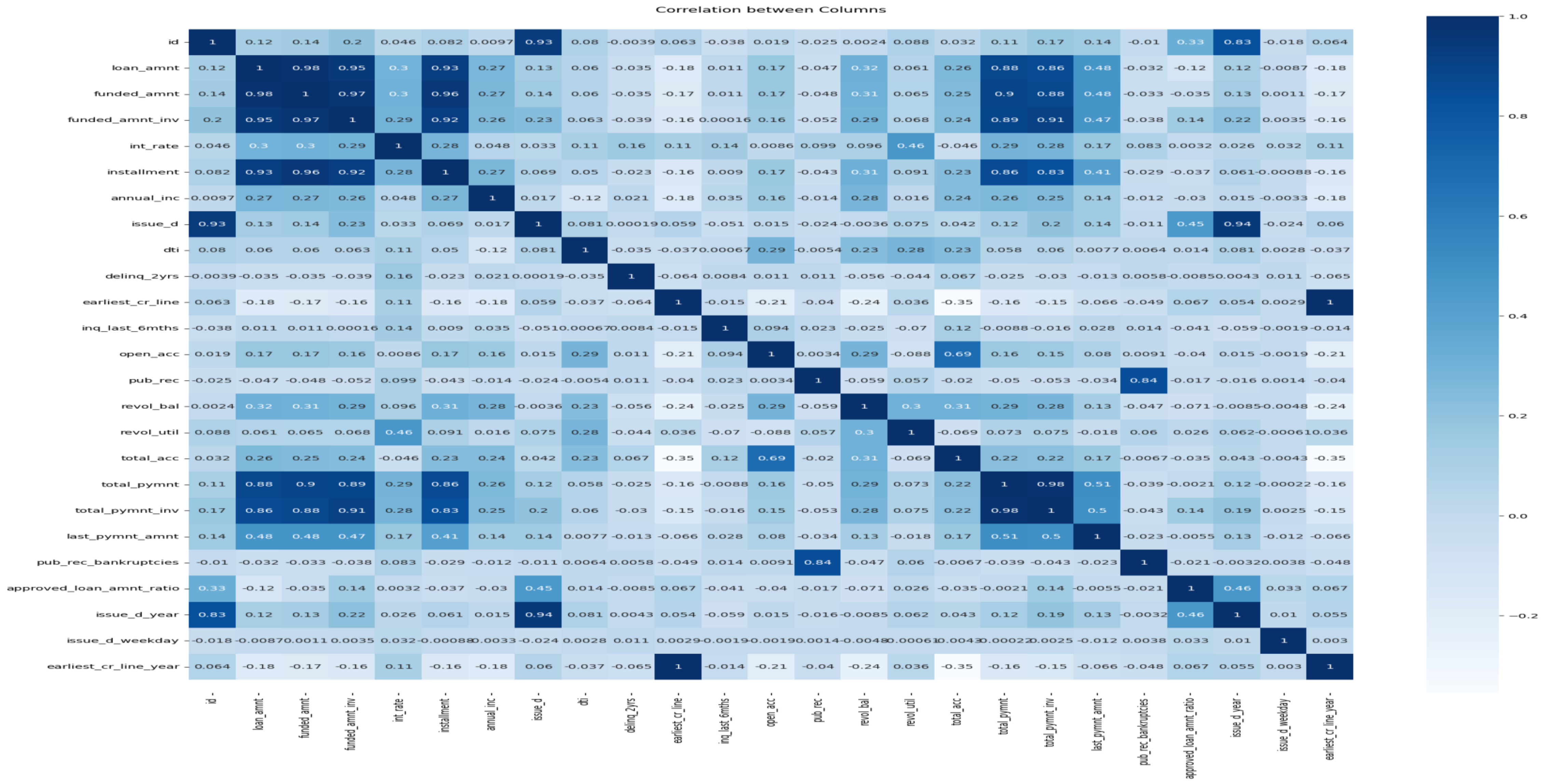
- It’s evident that customers from the CA, FL, and NY have a very high charged Off compared to customers from the other states.



Inference

- We can see from the Pearson correlation, that high interest rates, high loan amounts, and high DTI have a negative correlation with the loan status




Correlation between Columns
id
loan_amnt
funded_amnt
funded_amnt_inv
int_rate
installment
annual_inc
issue_d
dti
delinq_2yrs
earliest_cr_line
inq_last_6mths
open_acc
pub_rec
revol_bal
revol_util
total_acc
total_pymnt
total_pymnt_inv
last_pymnt_amnt
pub_rec_bankruptcies
approved_loan_amnt_ratio
issue_d_year
issue_d_weekday
earliest_cr_line_year
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

Inference:

1. **Strong Positive Correlations**:

Loan Amount, Funded Amount, Funded Amount Inv : These variables are highly correlated, as expected since they represent different aspects of the loan amount.

Total Payment, Total Payment Inv: Similarly, these variables are strongly correlated, indicating a direct relationship between the total amount paid and the total amount invested.

2. **Moderate Positive Correlations:**

Loan Amount, Interest Rate: There's a moderate positive correlation, suggesting that larger loans might have higher interest rates.

Installment, Loan Amount: A larger loan amount typically corresponds to higher installments.

3. **Weak or No Correlations:**

Public Records Bankruptcies, Interest Rate: There's a weak correlation or no correlation between these variables, indicating that bankruptcy history doesn't significantly influence interest rates.

4. **Specific Observations:**

Interest Rate: While it has moderate positive correlations with loan amount and installment, it doesn't seem to be strongly correlated with other variables.

Delinquency: Delinquency (delinq_2yrs) doesn't have strong correlations with most other variables, suggesting it might be a relatively independent factor.

Public Records Bankruptcies: As mentioned earlier, it doesn't seem to be strongly correlated with interest rates.

Conclusions:

- The Lending Club should exercise heightened caution when offering loans to small businesses, as they have a higher likelihood of default.
- It is advisable not to issue loans with a repayment tenure exceeding 8 years, as shorter terms are more effective in ensuring loan recovery.
- Special attention should be given to applicants from California (CA), Florida (FL), and New York (NY), as these states exhibit higher default rates.
- The company must maintain tight control over interest rates, as higher rates tend to increase the risk of default, leading to significant credit losses.
- A robust customer verification process is critical. Unverified applicants have shown a significantly higher tendency to default on their loans.
- Borrowers in lower credit grades, particularly F and G, should be closely monitored, as they are more likely to default.