Lending Club Case Study Analysis

Executive PG Programme in Machine Learning & AI - IIIT Bangalore

Analysis by-
1. Charu Garg
2. Chaitanya

18 Nov’24

Introduction

Background: The lending industry faces significant challenges in managing loan defaults. Understanding the factors that contribute to loan defaults can help lenders mitigate risks and improve their lending strategies.

Business Problem: The project aims to identify the key risk factors associated with loan defaults (charged-off loans) and provide insights to minimize defaults.

Objective: Perform Exploratory Data Analysis (EDA) to analyze factors influencing loan defaults and identify key risk factors.

Dataset Overview

File: loan.csv

Rows: 39,717

Columns: 111

Data_Dictionary: Contains mapping of variable names for column details.

Dataset used has key variables such as loan amount, annual income, interest rate, debt-to-income ratio, employment length, loan grades, home ownership status, loan status, etc.

Data Cleaning

Initial Cleaning:

Dropped rows where the target feature loan_status = "CURRENT".

First, deleted columns with 100% null values and then with more than 50% missing data.

Checked for duplicate rows or completely blank rows; no such cases found.

Deleted columns with unique values as they are irrelevant for our analysis.

Mapping and Exclusion:

Mapped column names with the shared dictionary file Data_Dictionary.xlsx

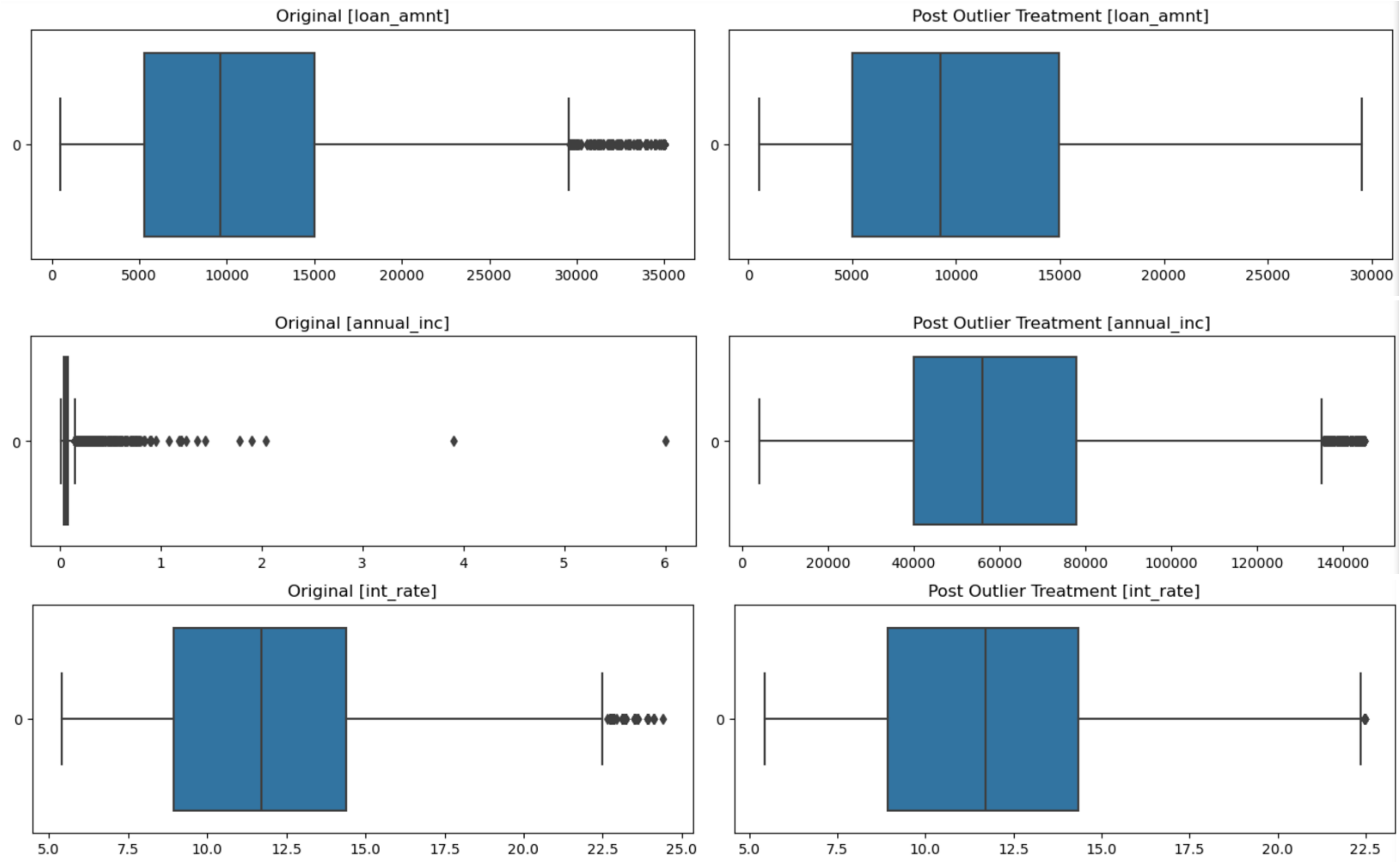
Excluded features that correspond to post-approval of the loan or are irrelevant.

Further Cleaning:

Imputed missing values using mode and median values.

Converted data types as necessary.

Treated outliers using the Interquartile Range (IQR) method and Z-score method, removing outliers beyond ±2 standard deviations for key variables like annual_inc.


Original [loan_amnt]
0
0 5000 10000 15000 20000 25000 30000 35000
Post Outlier Treatment [loan_amnt]
0
0 5000 10000 15000 20000 25000 30000
Original [annual_inc]
0
0 1 2 3 4 5 6
Post Outlier Treatment [annual_inc]
0
0 20000 40000 60000 80000 100000 120000 140000
Original [int_rate]
0
5.0 7.5 10.0 12.5 15.0 17.5 20.0 22.5 25.0
Post Outlier Treatment [int_rate]
0
5.0 7.5 10.0 12.5 15.0 17.5 20.0 22.5

Univariate Analysis Summary

Customer Insights:

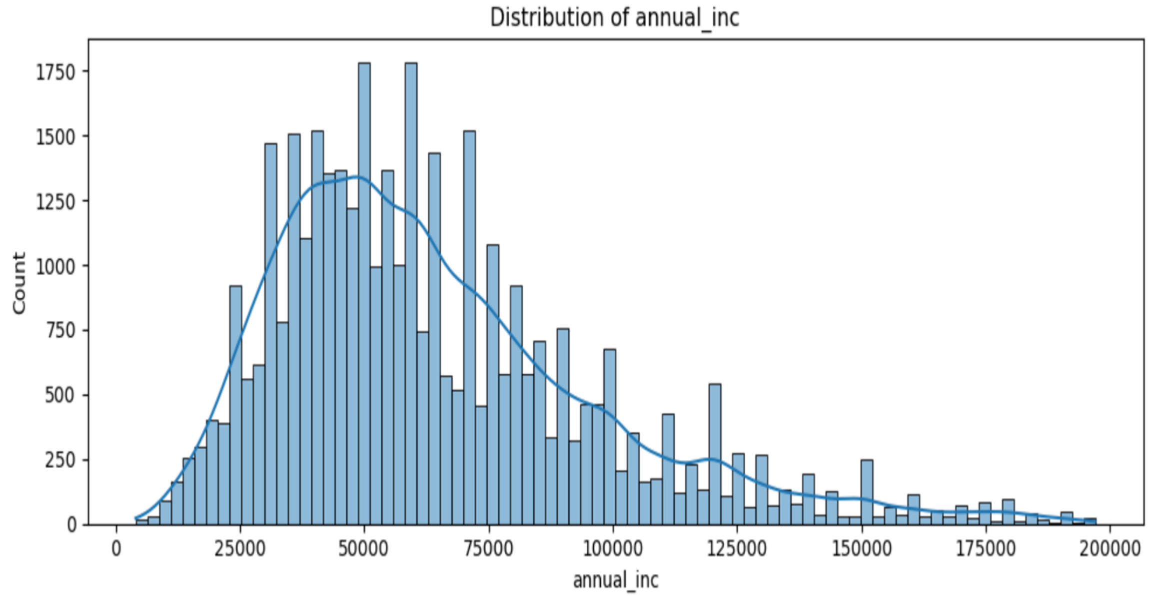
- Most homeowners are either renting or have a mortgage.
- The highest number of loan applications are for debt consolidation.
- The majority of customers have an employment length of 10+ years.
- Most loan applicants have an annual income between 40K-60K.
- The majority of debt-to-income ratios range from 8 to 19.

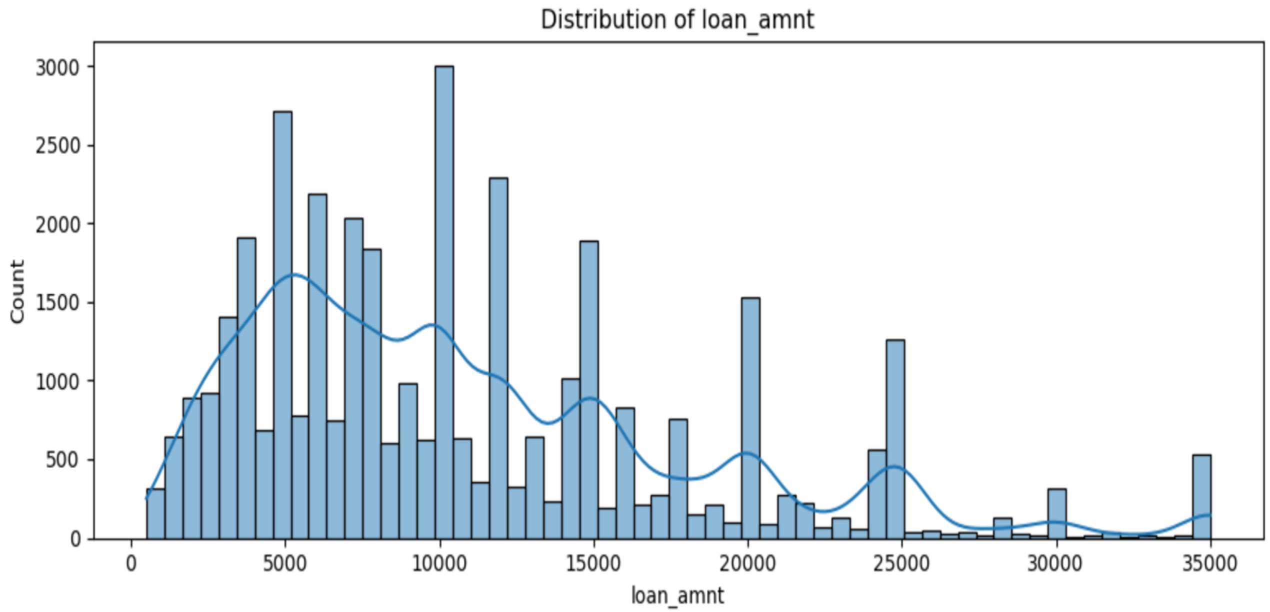
Loan Characteristics:

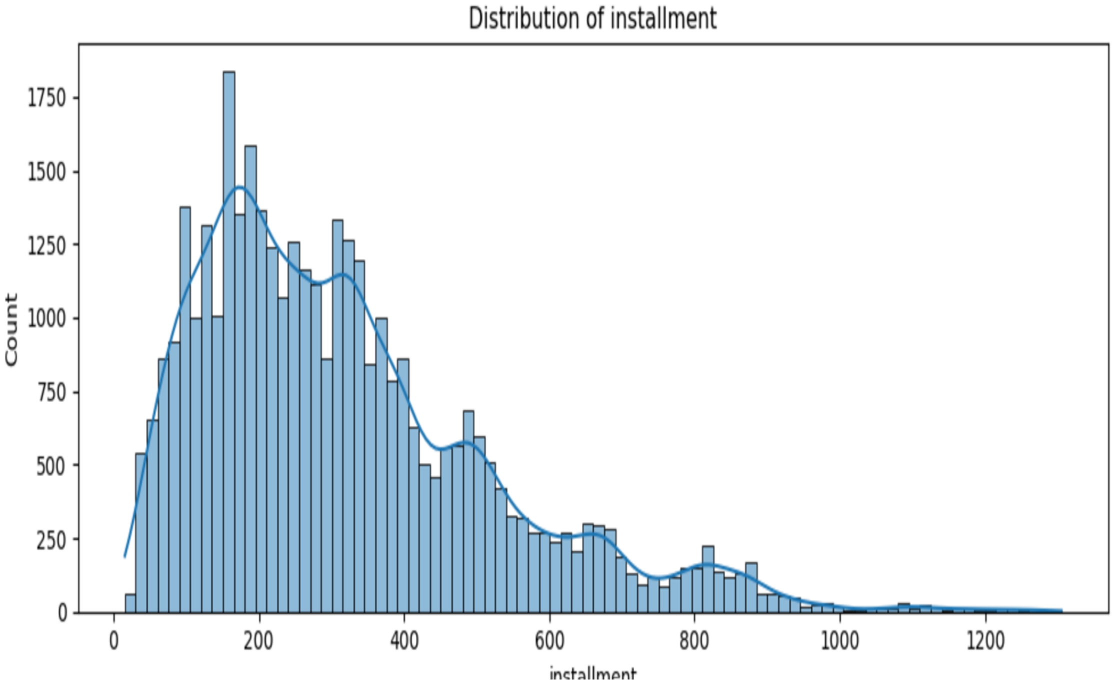
- Most loan amounts fall within the range of 5K to 15K.
- The majority of interest rates fall within the 9% to 11.5% range.
- The majority of installment amounts are between 160 and 400.
- Most loan applications are for a term of 36 months.
- The majority of loan applications fall under Grade B.

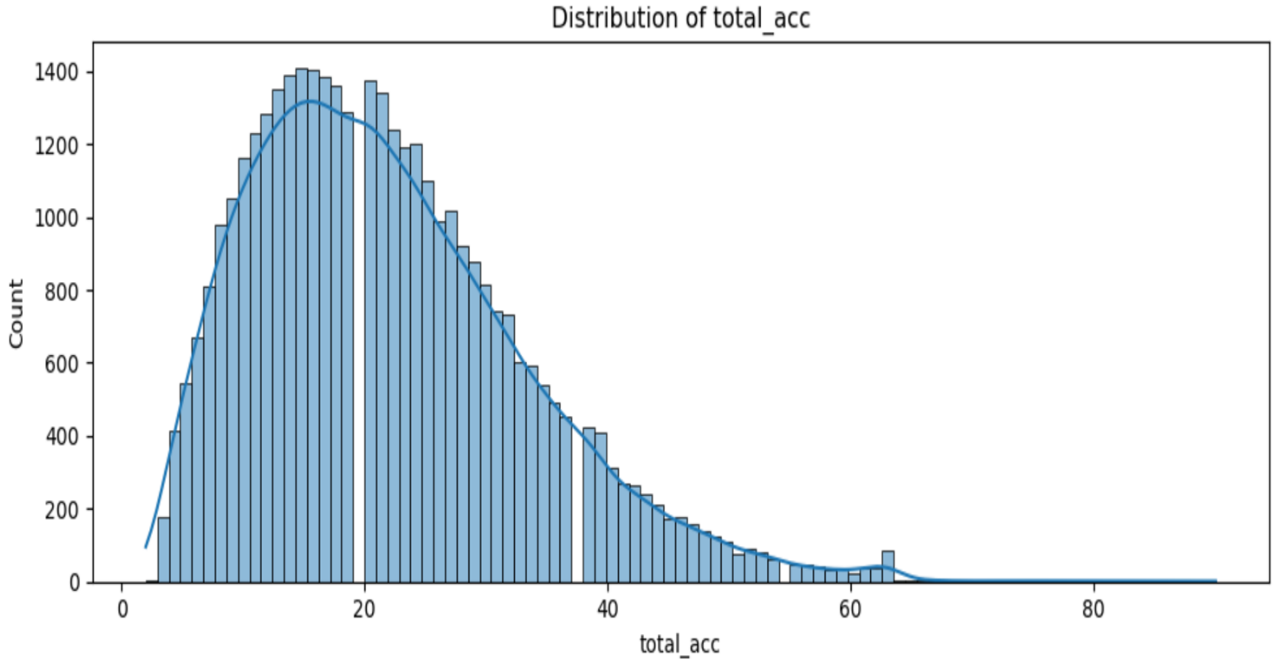
Temporal Trends:

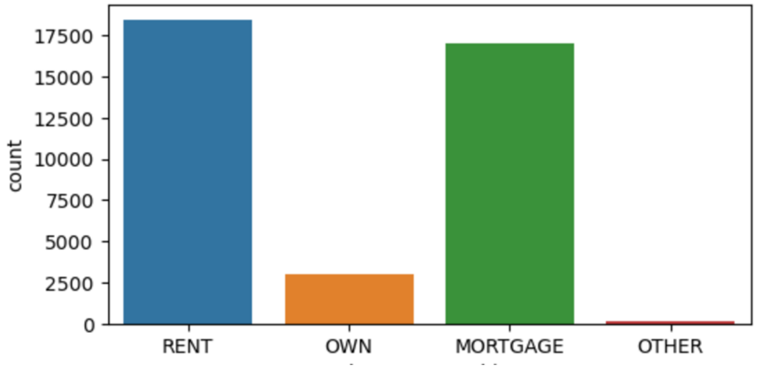
- Loan application counts are increasing year over year.
- The highest loan application volume is in December.
- The lowest loan applications are in the first half of the year (Jan-Jun).

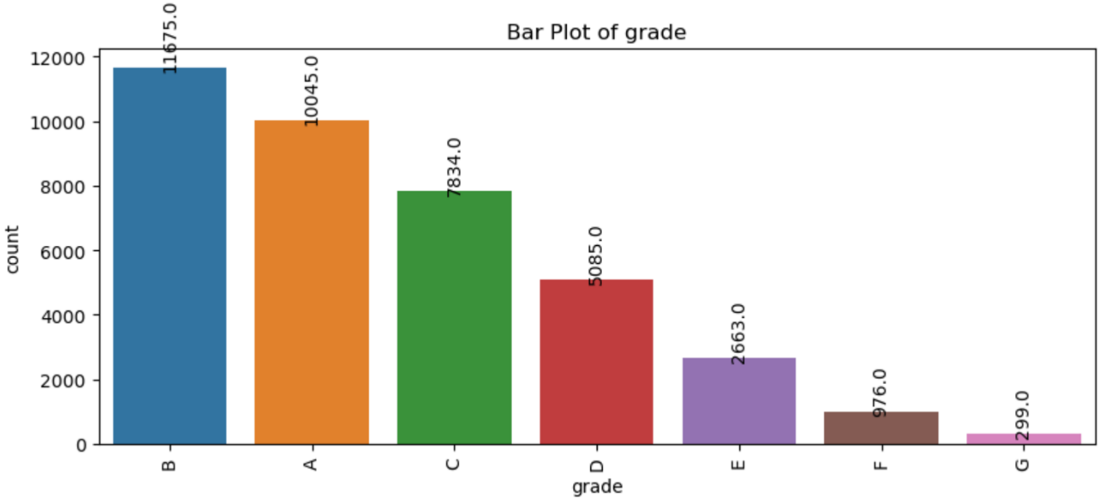

Distribution of annual_inc
Count
1750
1500
1250
1000
750
500
250
0
0
25000
50000
75000
100000
125000
150000
175000
200000
annual_inc

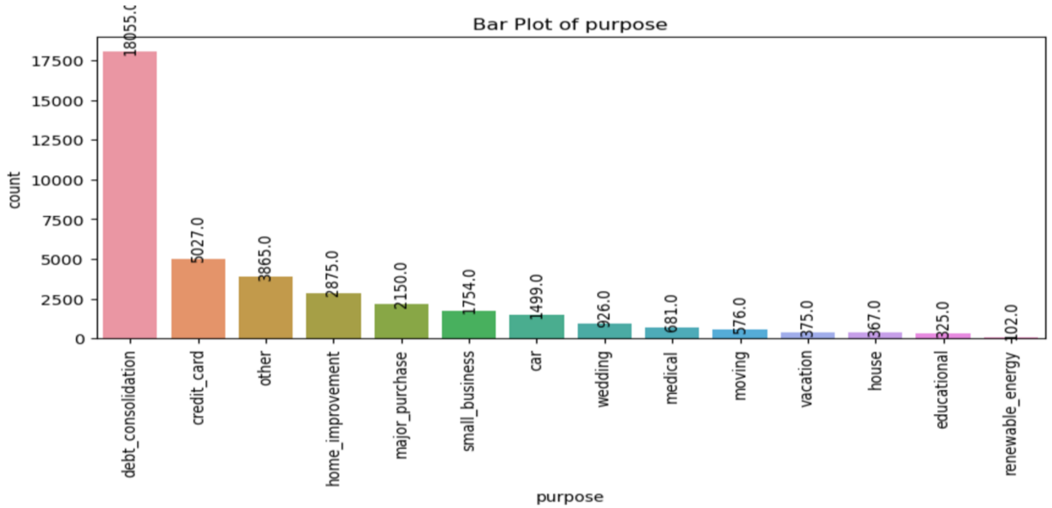

Distribution of loan_amnt
Count
3000
2500
2000
1500
1000
500
0
0
5000
10000
15000
20000
25000
30000
35000
loan_amnt

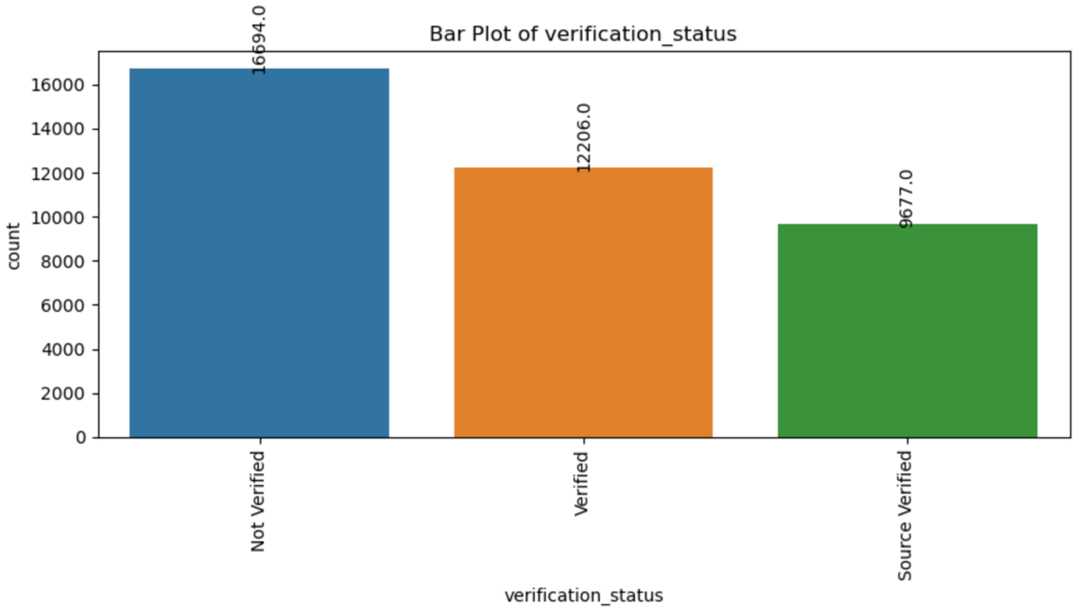

Distribution of installment
Count
1750
1500
1250
1000
750
500
250
0
0
200
400
600
800
1000
1200
installment

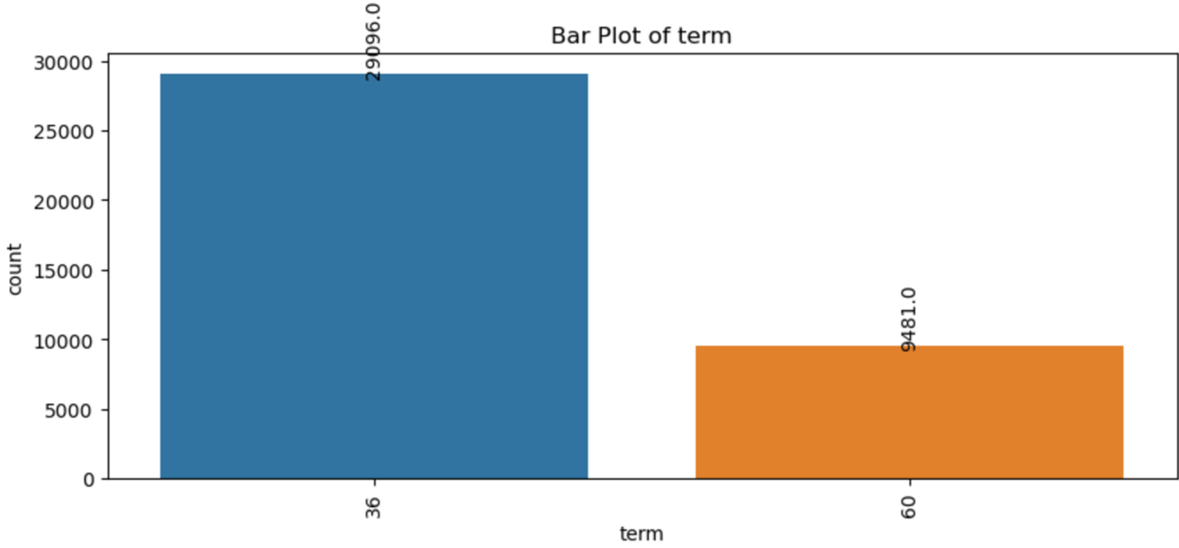

Distribution of total_acc
Count
1400
1200
1000
800
600
400
200
0
0
20
40
60
80
total_acc


count
RENT
OWN
MORTGAGE
OTHER


Bar Plot of grade
count
grade
B 11675.0
A 10045.0
C 7834.0
D 5085.0
E 2663.0
F 976.0
G 299.0

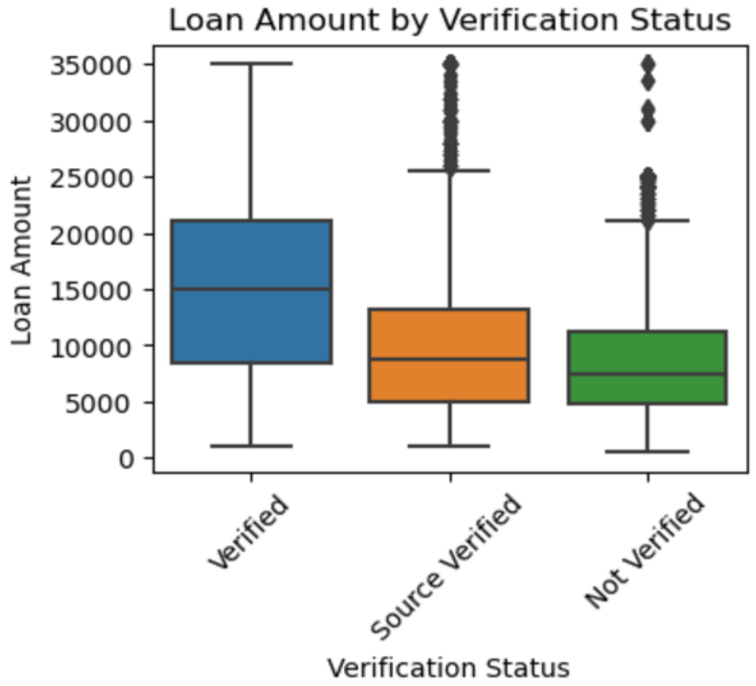

Bar Plot of purpose
count
purpose
debt_consolidation 18055.0
credit_card 5027.0
other 3865.0
home_improvement 2875.0
major_purchase 2150.0
small_business 1754.0
car 1499.0
wedding 926.0
medical 681.0
moving 576.0
vacation 375.0
house 367.0
educational 325.0
renewable_energy 102.0

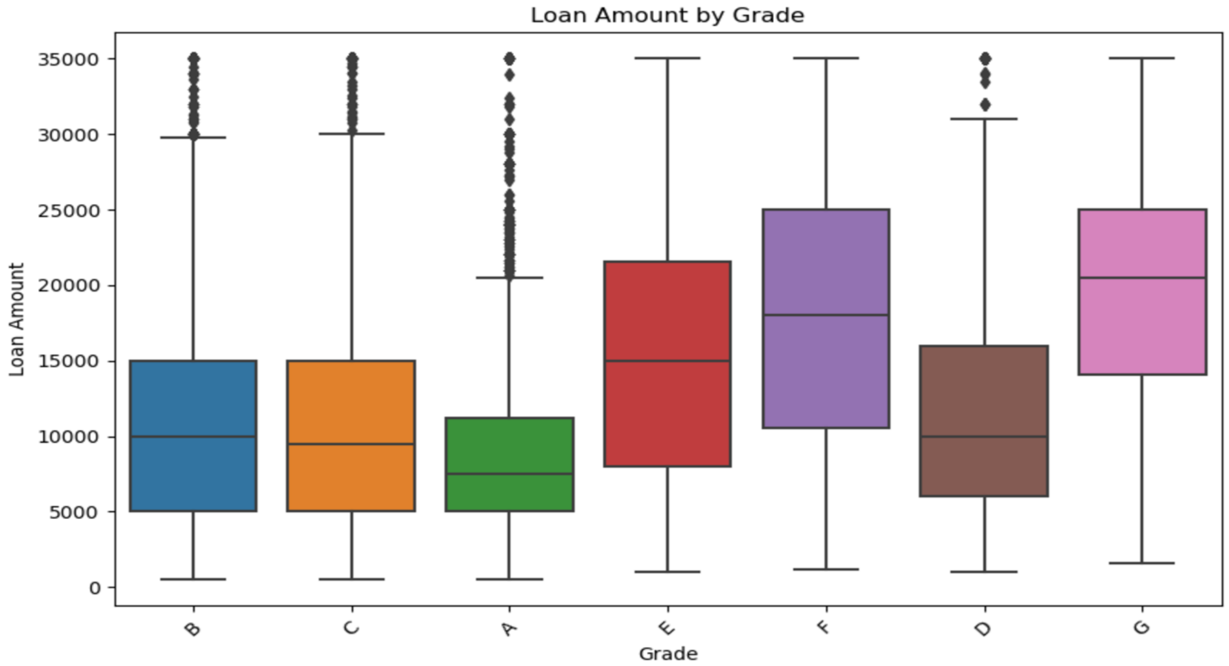

Bar Plot of verification_status
count
verification_status
Not Verified 16694.0
Verified 12206.0
Source Verified 9677.0

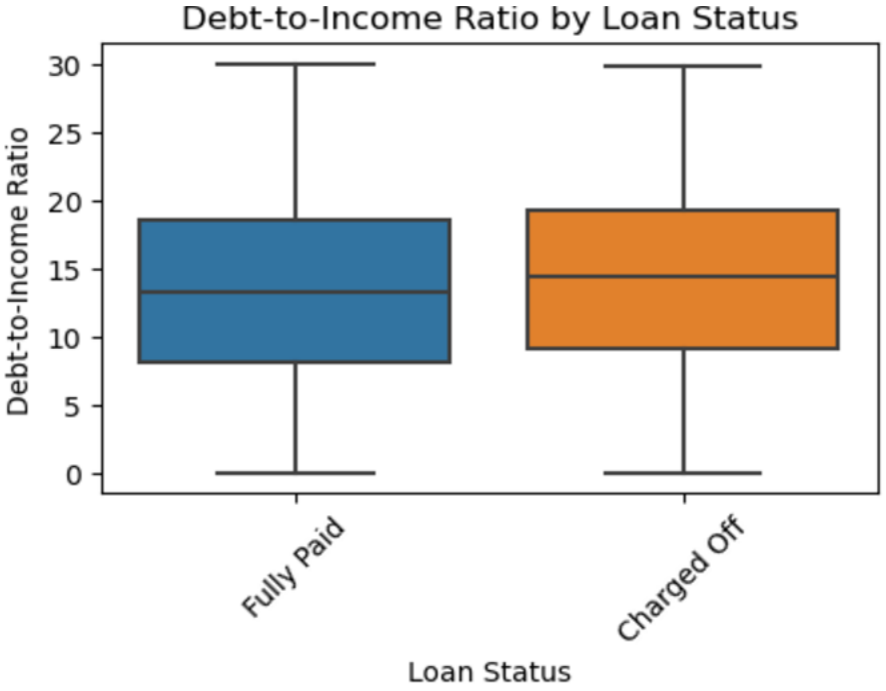

Bar Plot of term
count
term
36 29096.0
60 9481.0

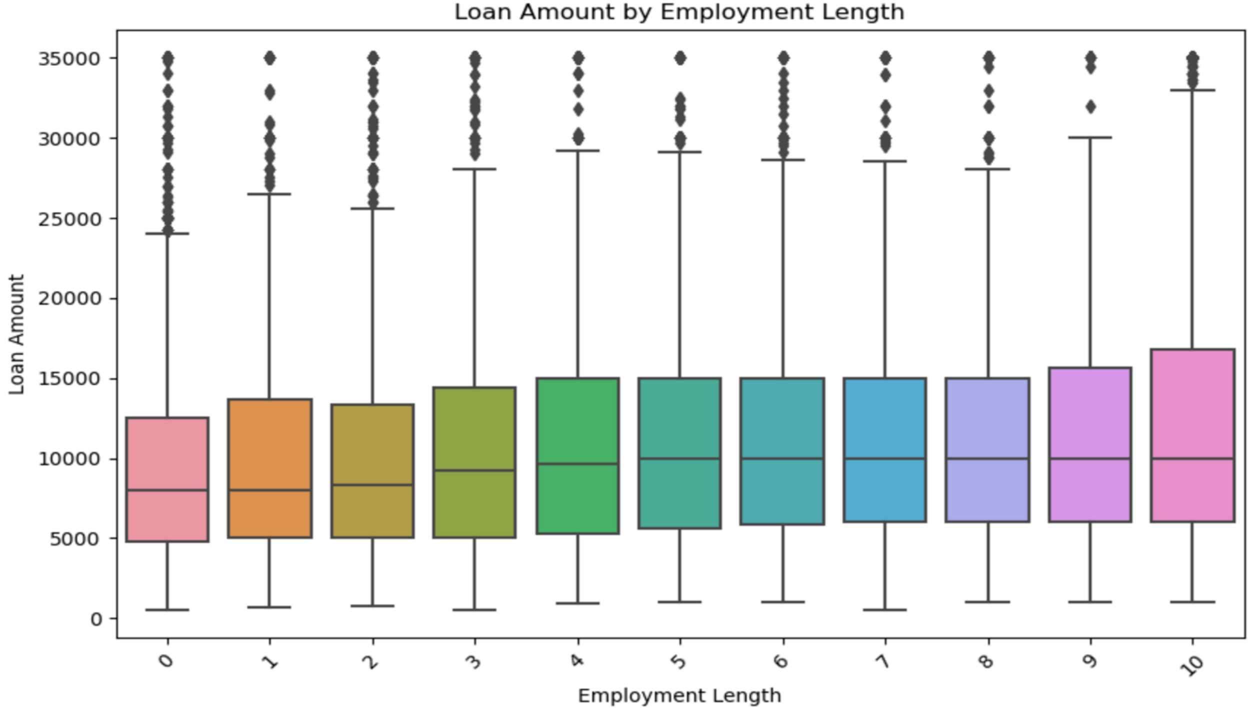
Bivariate Analysis Summary

- **Interest Rates and Loan Status**: Charged-off loans tend to have higher interest rates compared to fully paid loans.
- **Loan Amounts and Grade**: Loans with a grade of 'G' are typically associated with higher loan amounts.
- **Debt-to-Income Ratio and Default Rates**: Higher debt-to-income ratios are linked to increased default rates.
- **Employment Length and Loan Amounts**: Longer employment lengths are correlated with higher loan amounts.
- **Verification Status and Loan Amounts**: Verified loans are generally associated with higher loan amounts.


Loan Amount by Verification Status
Loan Amount
35000
30000
25000
20000
15000
10000
5000
0
Verified
Source Verified
Not Verified
Verification Status

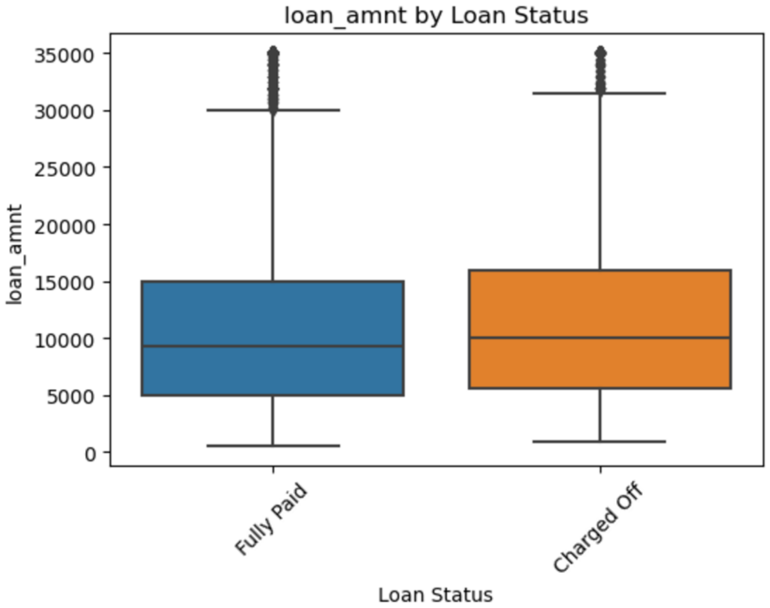

Loan Amount by Grade
Loan Amount
35000
30000
25000
20000
15000
10000
5000
0
B
C
A
E
F
D
G
Grade

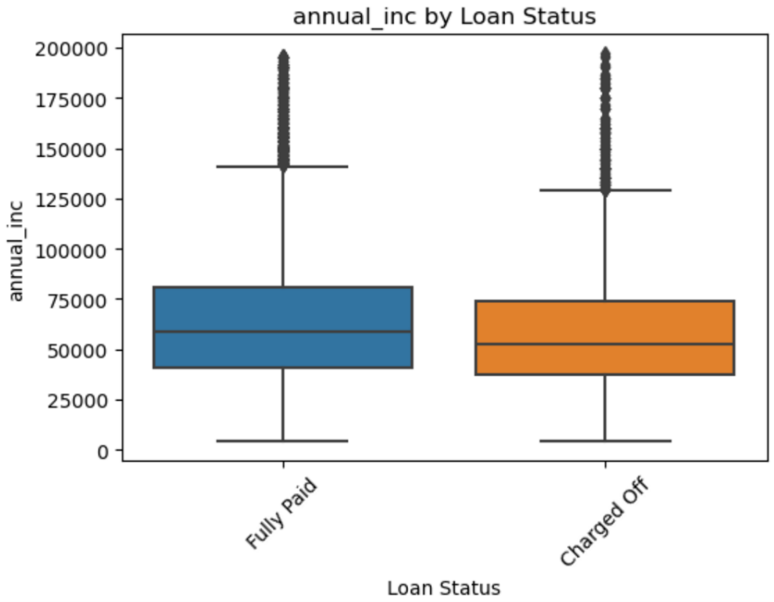

Debt-to-Income Ratio by Loan Status
Debt-to-Income Ratio
30
25
20
15
10
5
0
Fully Paid
Charged Off
Loan Status

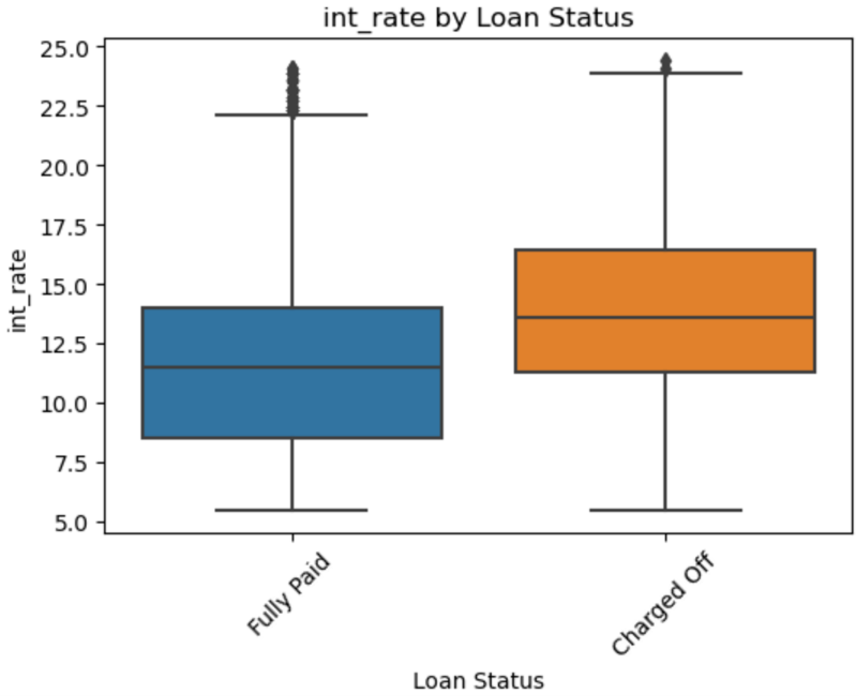

Loan Amount by Employment Length
Loan Amount
35000
30000
25000
20000
15000
10000
5000
0
0
1
2
3
4
5
6
7
8
9
10
Employment Length

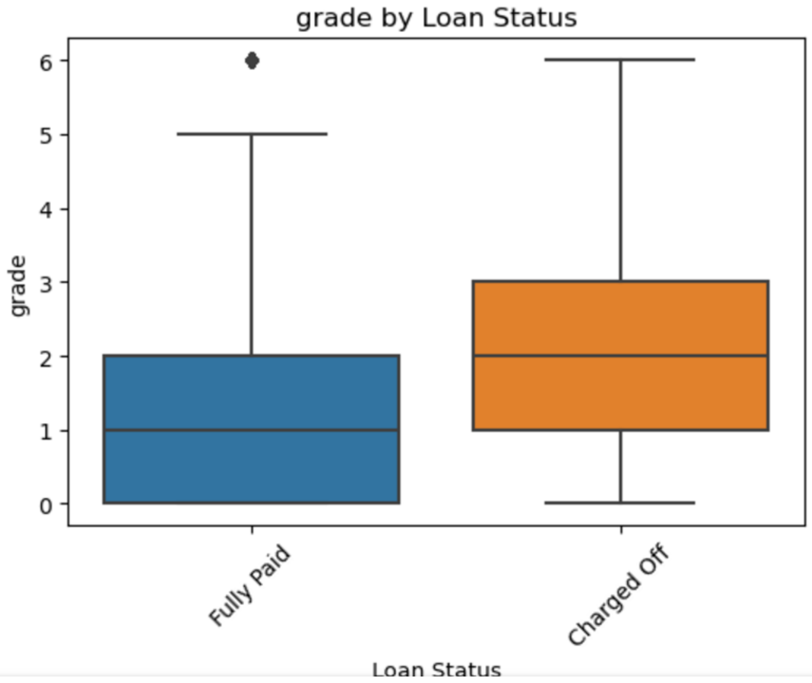
Summary of Charged-Off Cases Analysis

- **Loan Amount**: Borrowers with larger loan amounts are at a higher risk of default.
- **Annual Income**: Borrowers in charged-off cases generally have lower annual incomes.
- **Interest Rate**: Charged-off loans are associated with higher interest rates.
- **Debt-to-Income Ratio (DTI)**: The debt-to-income ratio is slightly higher in charged-off cases.
- **Employment Length**: The median and upper end of employment length are slightly higher in charged-off cases.
- **Grades**: Charged-off loans are associated with lower grades (numerically higher).


loan_amnt by Loan Status
loan_amnt
35000
30000
25000
20000
15000
10000
5000
0
Fully Paid
Charged Off
Loan Status


annual_inc by Loan Status
annual_inc
200000
175000
150000
125000
100000
75000
50000
25000
0
Fully Paid
Charged Off
Loan Status


int_rate by Loan Status
int_rate
25.0
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
Fully Paid
Charged Off
Loan Status


grade by Loan Status
grade
6
5
4
3
2
1
0
Fully Paid
Charged Off
Loan Status

Conclusion

- **Key Risk Factors**: Higher loan amounts, lower annual incomes, higher interest rates, and higher debt-to-income ratios are associated with higher default risks.

- **Additional Risks**: Even borrowers with longer employment lengths and lower loan grades are at risk.

- **Recommendations**: Implement strategies to mitigate default risks, such as offering financial counseling, adjusting loan terms, or implementing stricter lending criteria.

Please note:
"For detailed analysis and plots, please refer to the accompanying Jupyter notebook file (ipynb)."

Thanks