Loan Default Analysis and Recommendations

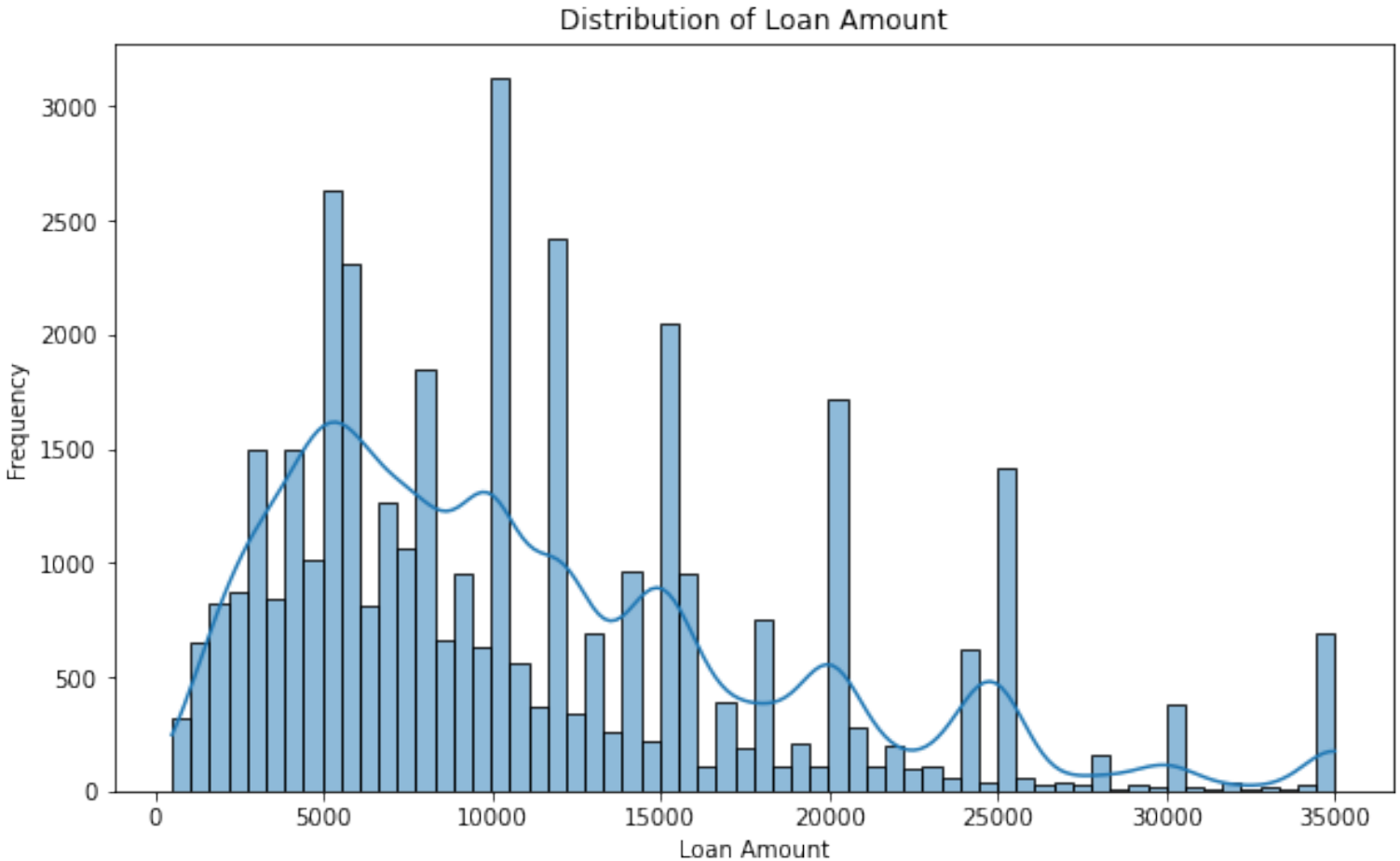
Executive Summary

This report presents an exploratory data analysis (EDA) of loan data to identify driving factors behind loan default. The analysis aims to help the largest online loan marketplace reduce credit loss by identifying risky loan applicants. The report provides insights into the distribution of loan amounts, interest rates, loan status, and other key variables, as well as correlations and relationships between them.

Distribution of Loan Amount

The distribution of loan amounts is skewed to the right, indicating that most loans are for smaller amounts. This suggests that the company may want to focus on smaller loan amounts to reduce credit loss.

Histogram of Loan Amount

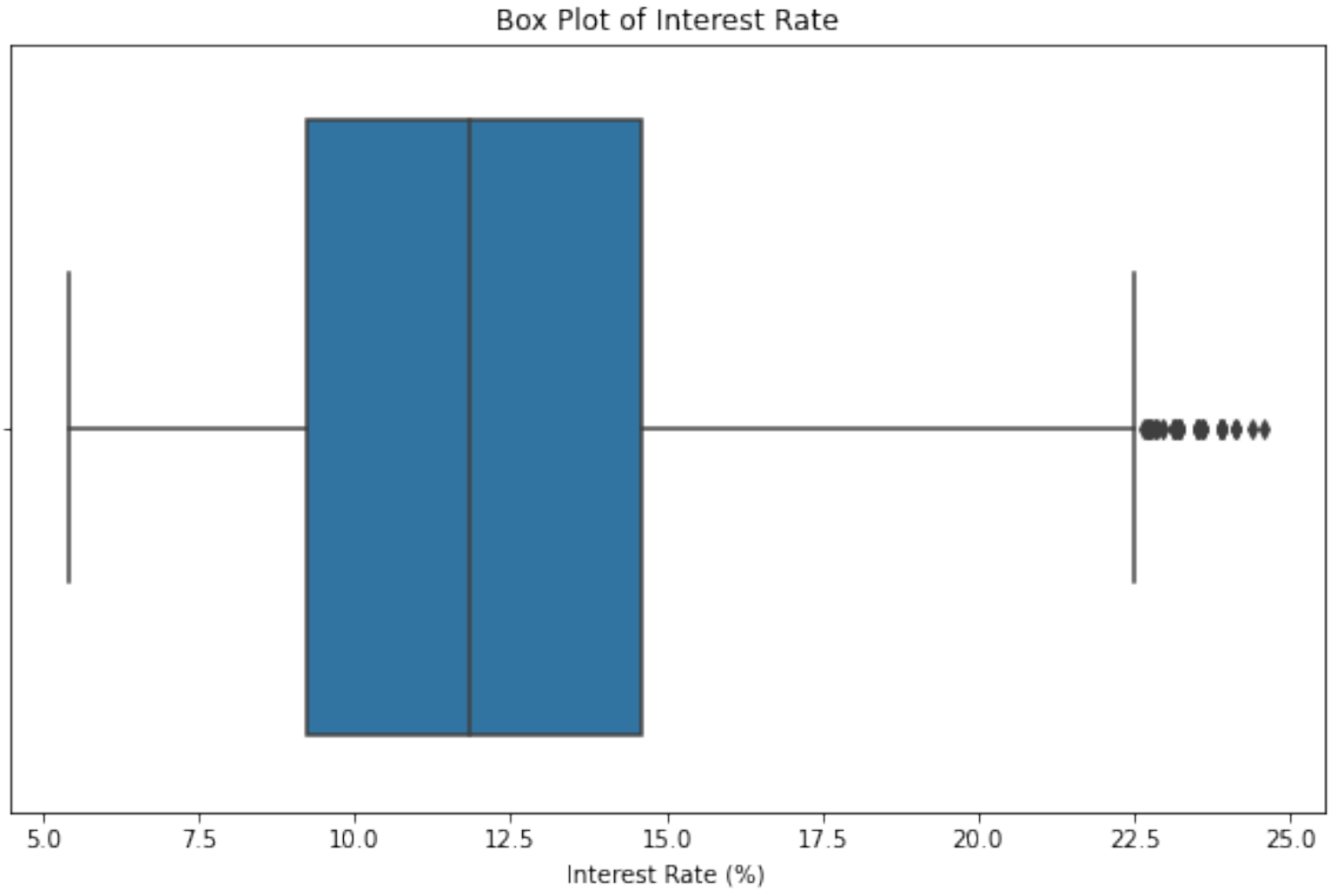


Histogram of Loan Amount

Distribution of Interest Rate

The interest rates are relatively consistent, with a median value around 12%. This suggests that interest rates may not be a significant driver of loan default.

Box Plot of Interest Rate

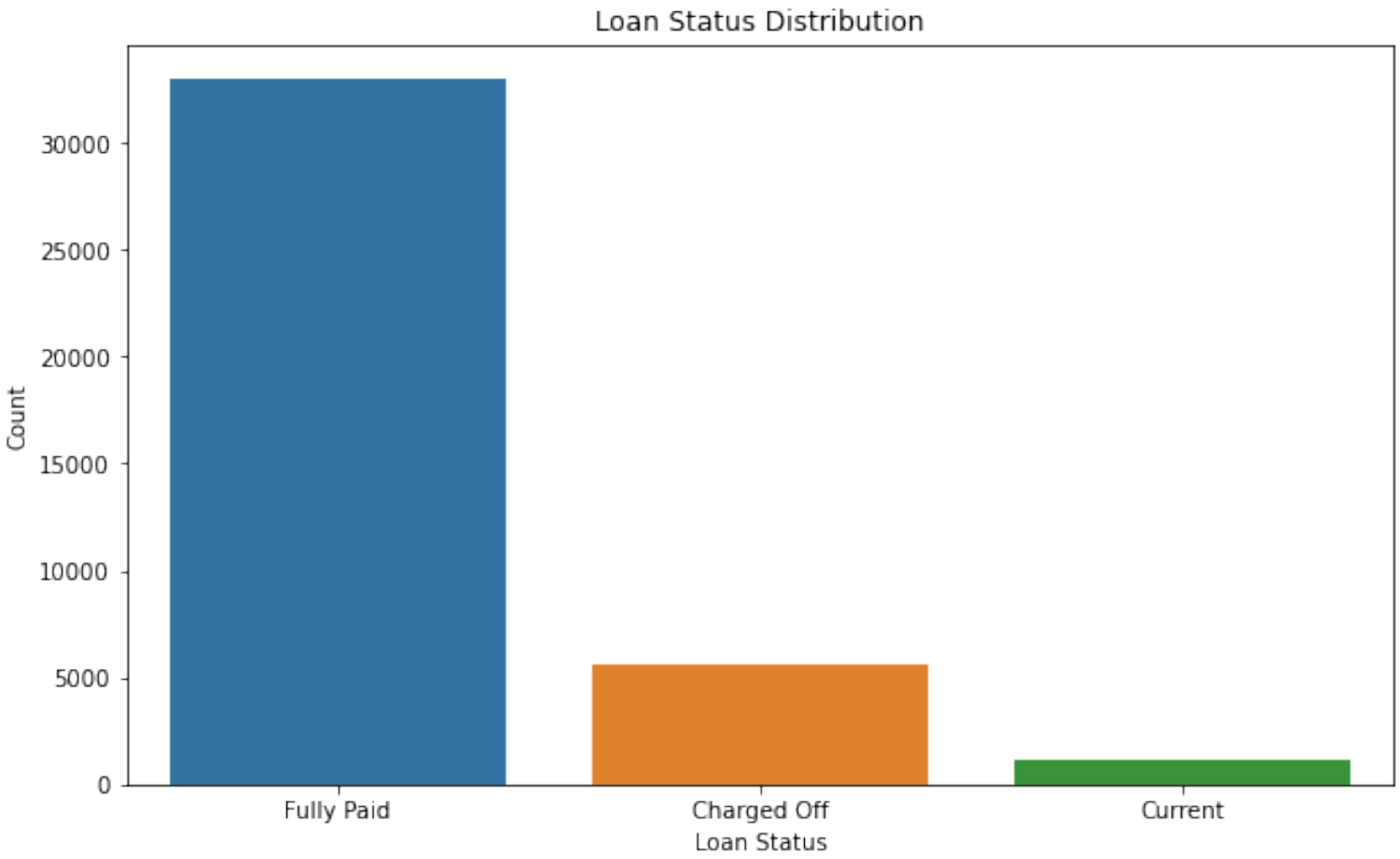


Box Plot of Interest Rate

Distribution of Loan Status

The majority of loans are fully paid, while a significant proportion are charged off. This highlights the importance of identifying risky loan applicants to reduce credit loss.

Bar Plot of Loan Status

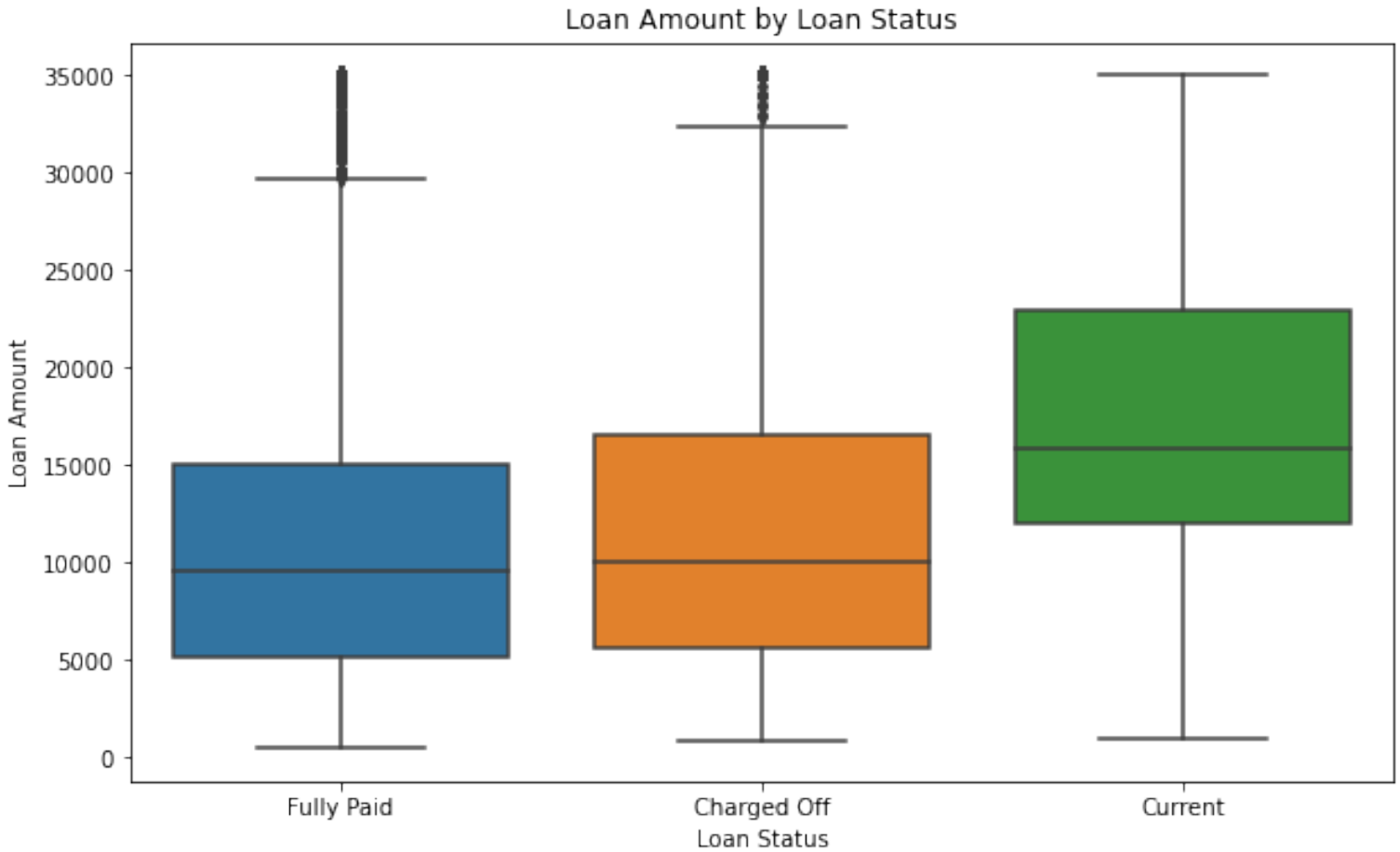


Bar Plot of Loan Status

Loan Amount by Loan Status

Charged off loans tend to have higher loan amounts compared to fully paid loans. This suggests that loan amount may be a driver of loan default.

Box Plot of Loan Amount by Loan Status

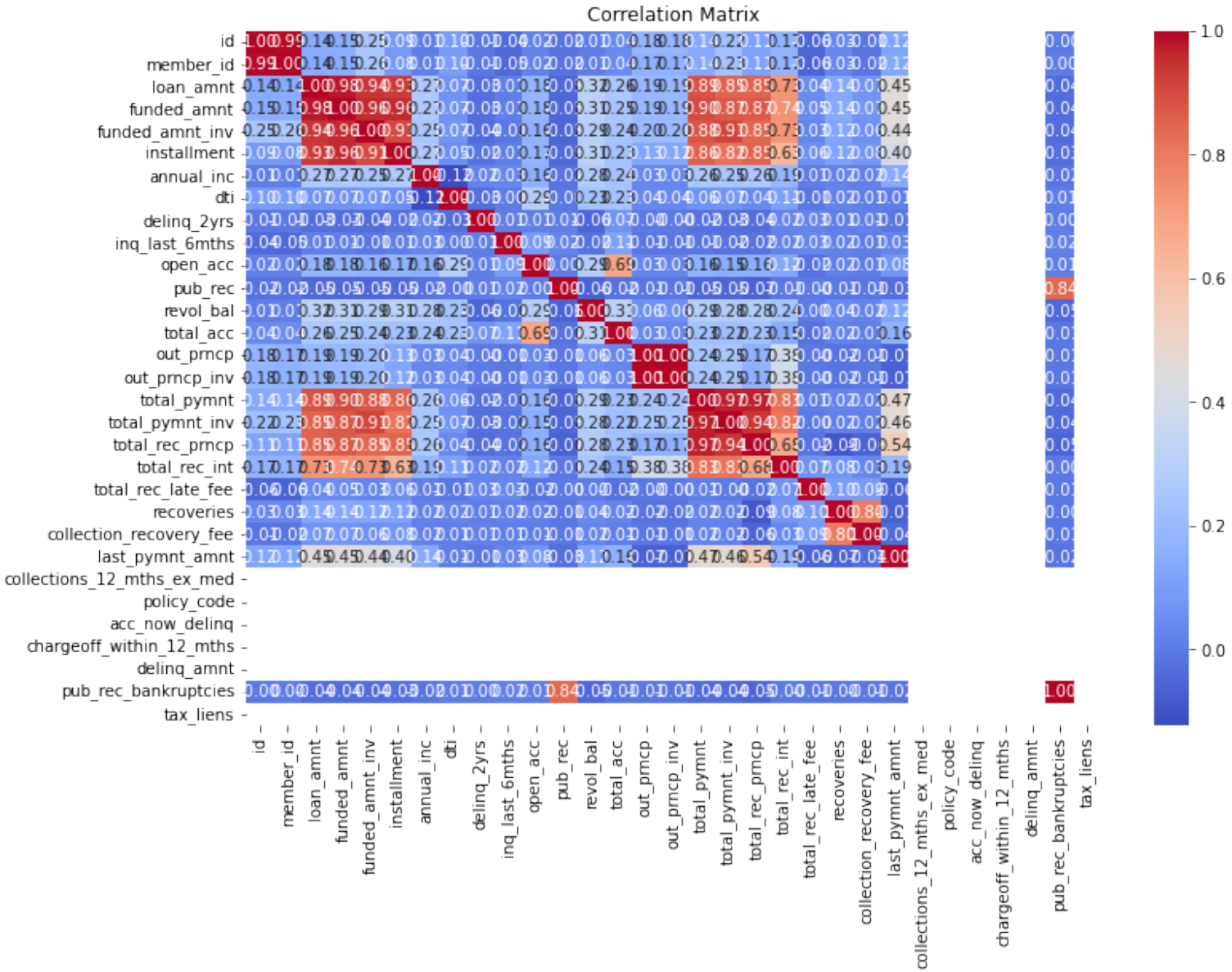


Box Plot of Loan Amount by Loan Status

Correlation Matrix

There are strong correlations between loan amount, interest rate, and installment amount. This suggests that these variables may be interrelated and influence loan default.

Correlation Matrix

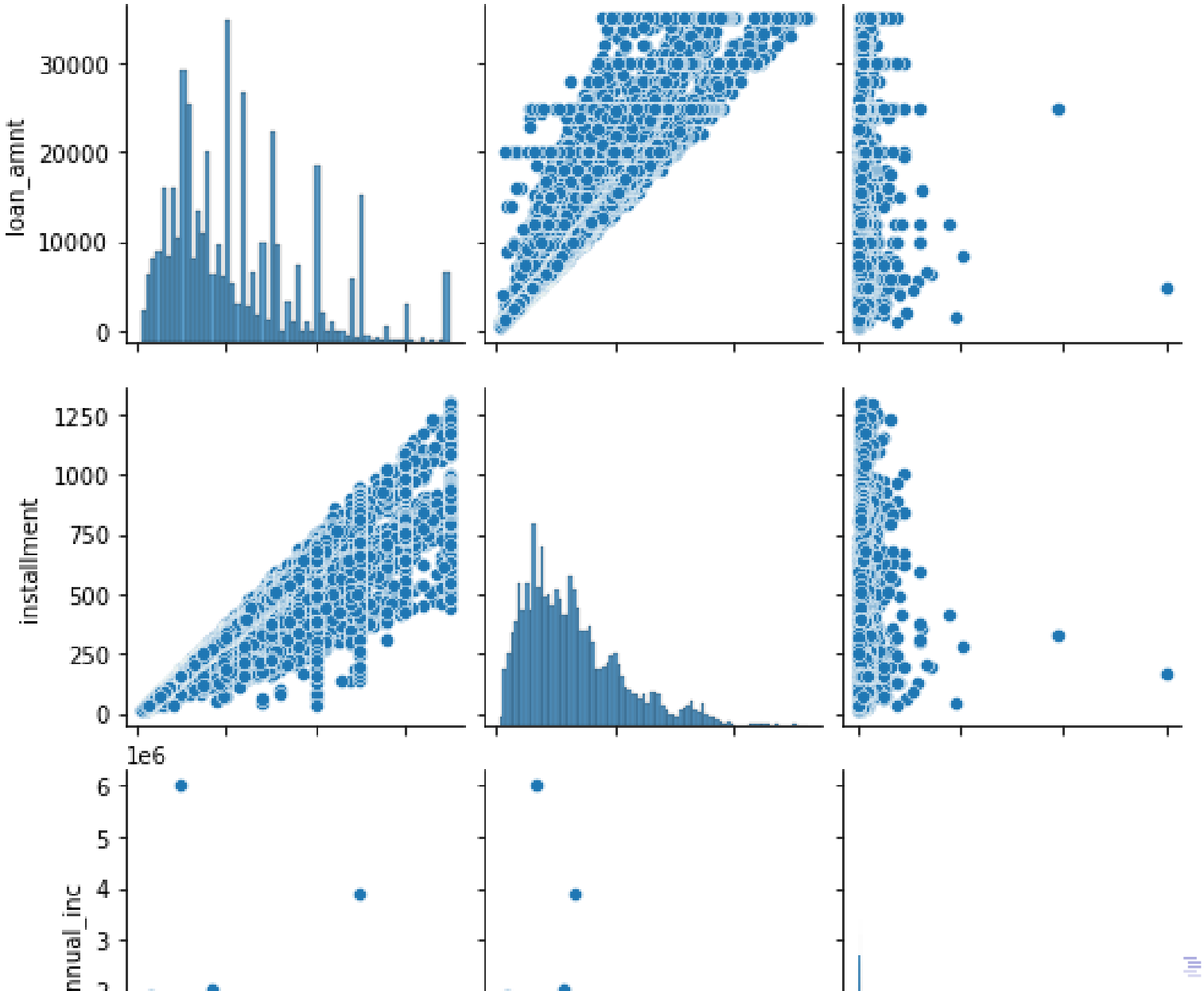


Correlation Matrix

Pair Plot

The pair plot reveals relationships between loan amount, interest rate, installment amount, and annual income. This suggests that annual income may be an important factor in loan default.

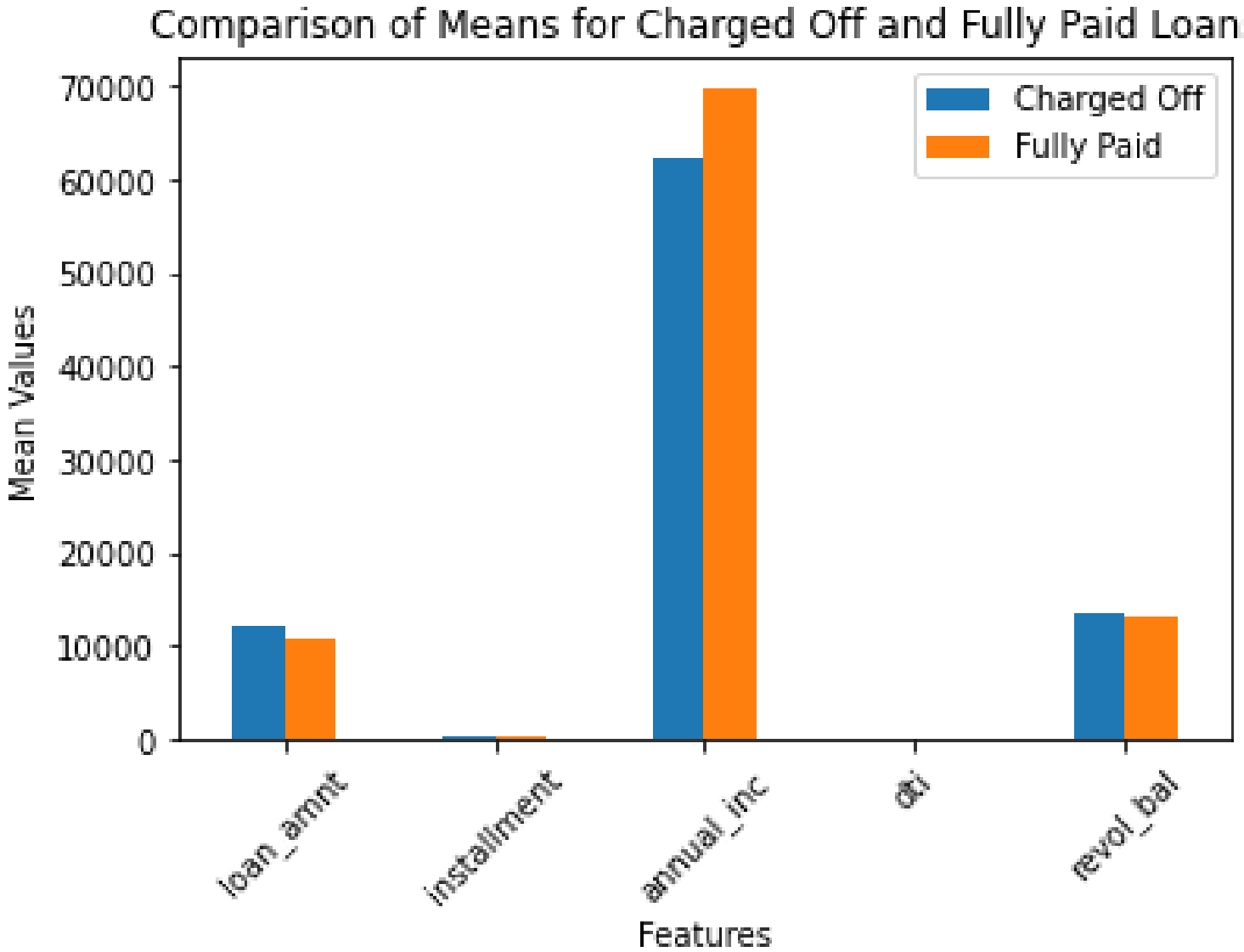
Pair Plot


loan_amnt
30000
20000
10000
0
installment
1250
1000
750
500
250
0
1e6
6
5
4
3
nnual_inc

Comparison of Means

Charged off loans have higher means for loan amount, interest rate, and installment amount compared to fully paid loans. This suggests that these variables may be drivers of loan default.

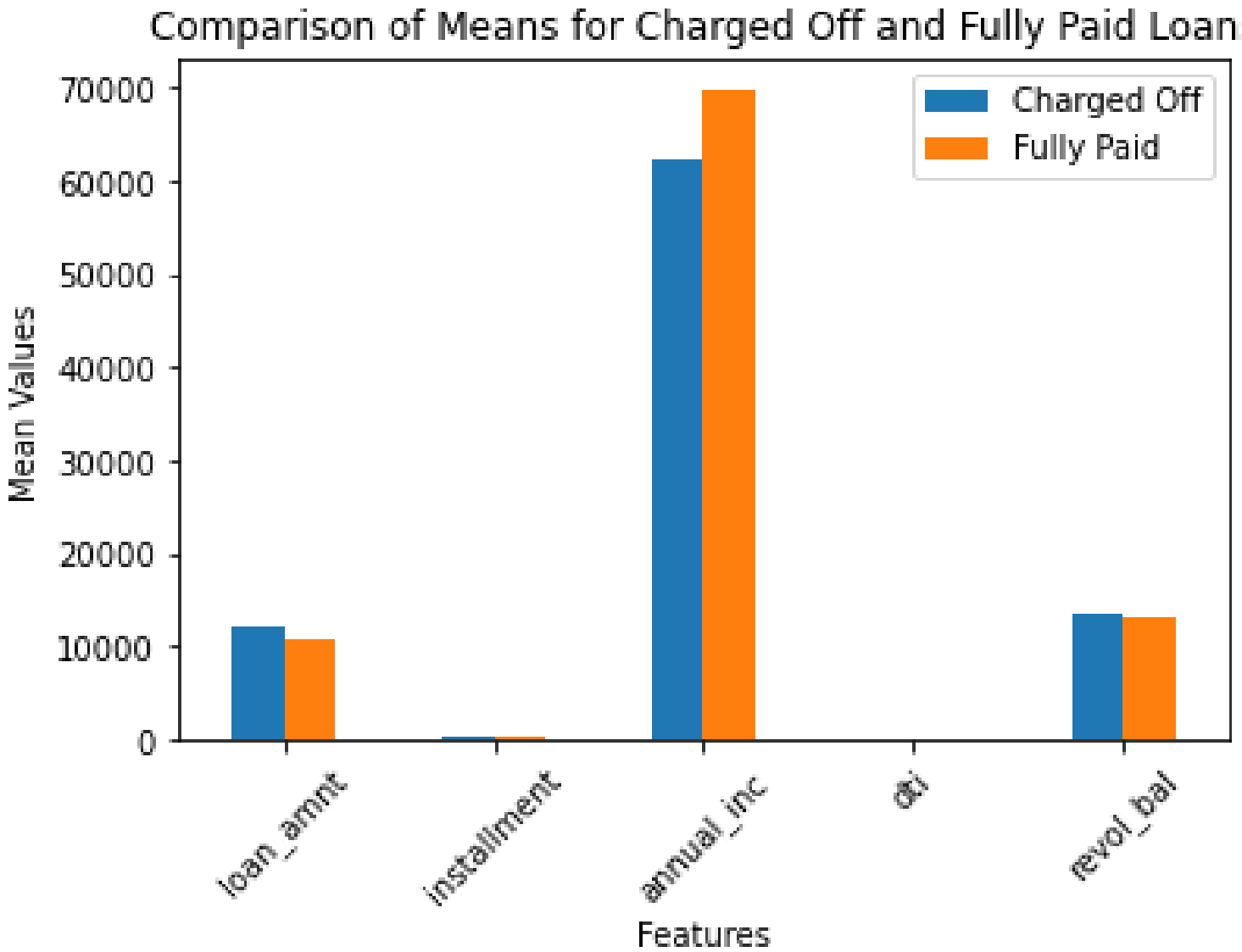
Comparison of Means



Distribution of Annual Income

The distribution of annual income is similar for charged off and fully paid loans, with a peak around $50,000. This suggests that annual income may not be a significant driver of loan default.

Histogram of Annual Income

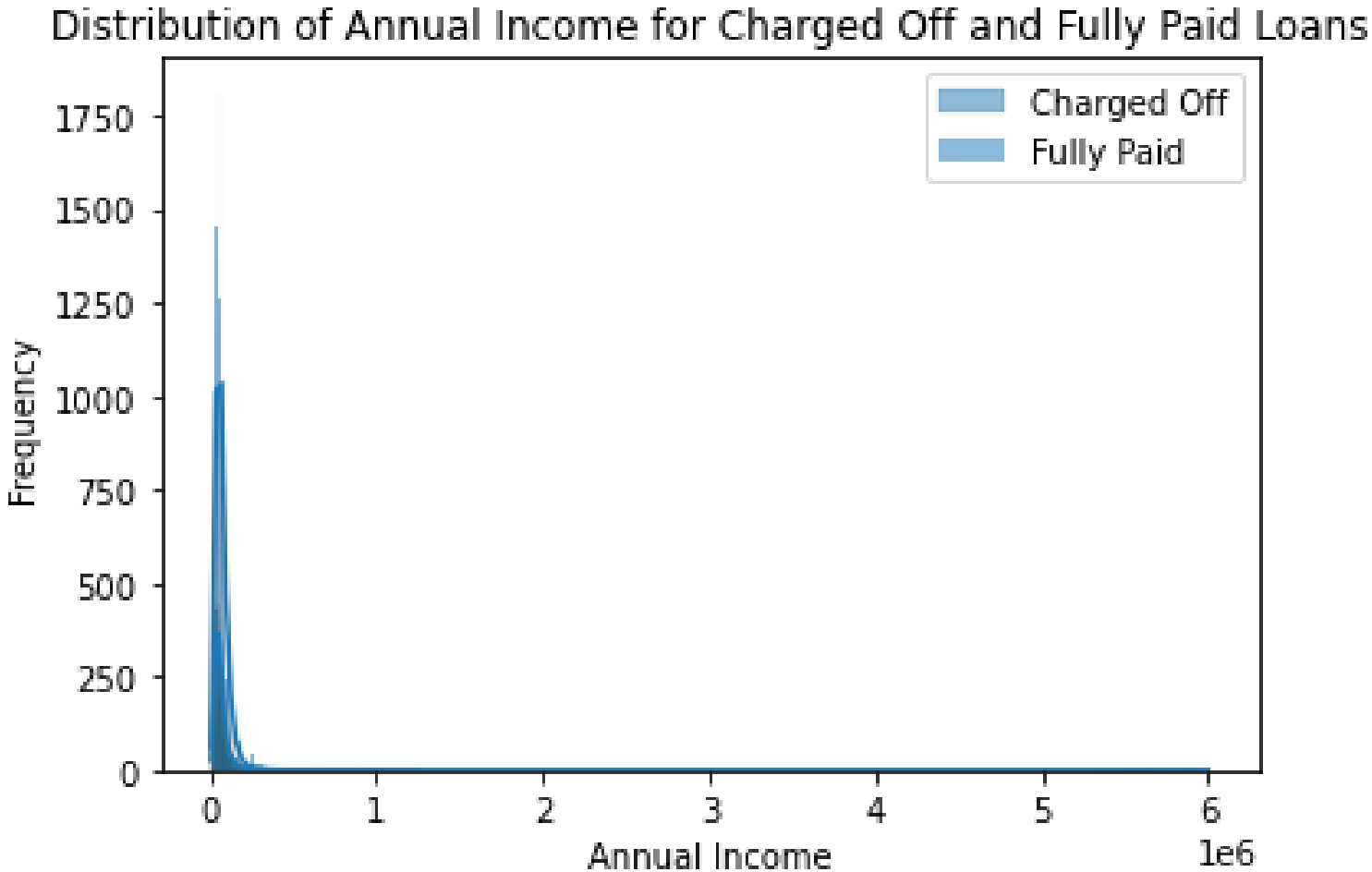


Histogram of Annual Income

Distribution of DTI Ratio

The distribution of DTI ratio is similar for charged off and fully paid loans, with a peak around 20. This suggests that DTI ratio may not be a significant driver of loan default.

Histogram of DTI Ratio



Histogram of DTI Ratio

Recommendations

Based on the analysis, we recommend the following:

1. **Focus on smaller loan amounts**: The company may want to focus on smaller loan amounts to reduce credit loss, as most loans are for smaller amounts.
2. **Monitor loan amount and interest rate**: The company should monitor loan amount and interest rate, as they appear to be related to loan default risk.
3. **Evaluate risk factors more deeply**: Further analysis should be conducted to assess the impact of additional factors on loan default.