Loan Default Risk Analysis

Problem Statement

- Minimizing credit loss by identifying risky borrowers.
- Use data to identify the factors of loan defaults.

Dataset Overview

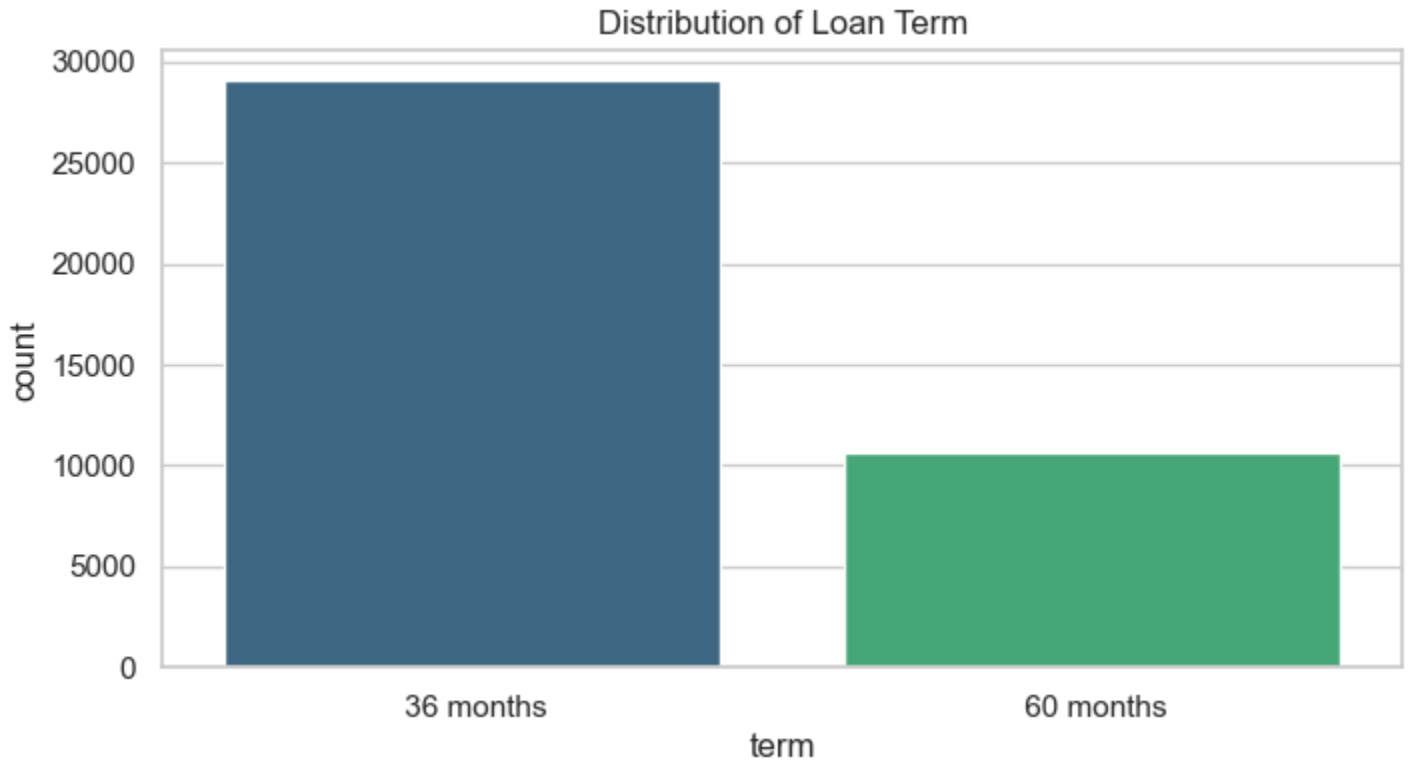
- Description of the dataset
- Data involve Loan amount, interest rate, term, grade, loan status, etc.
- Size of Data: 39717 Row x 111 column

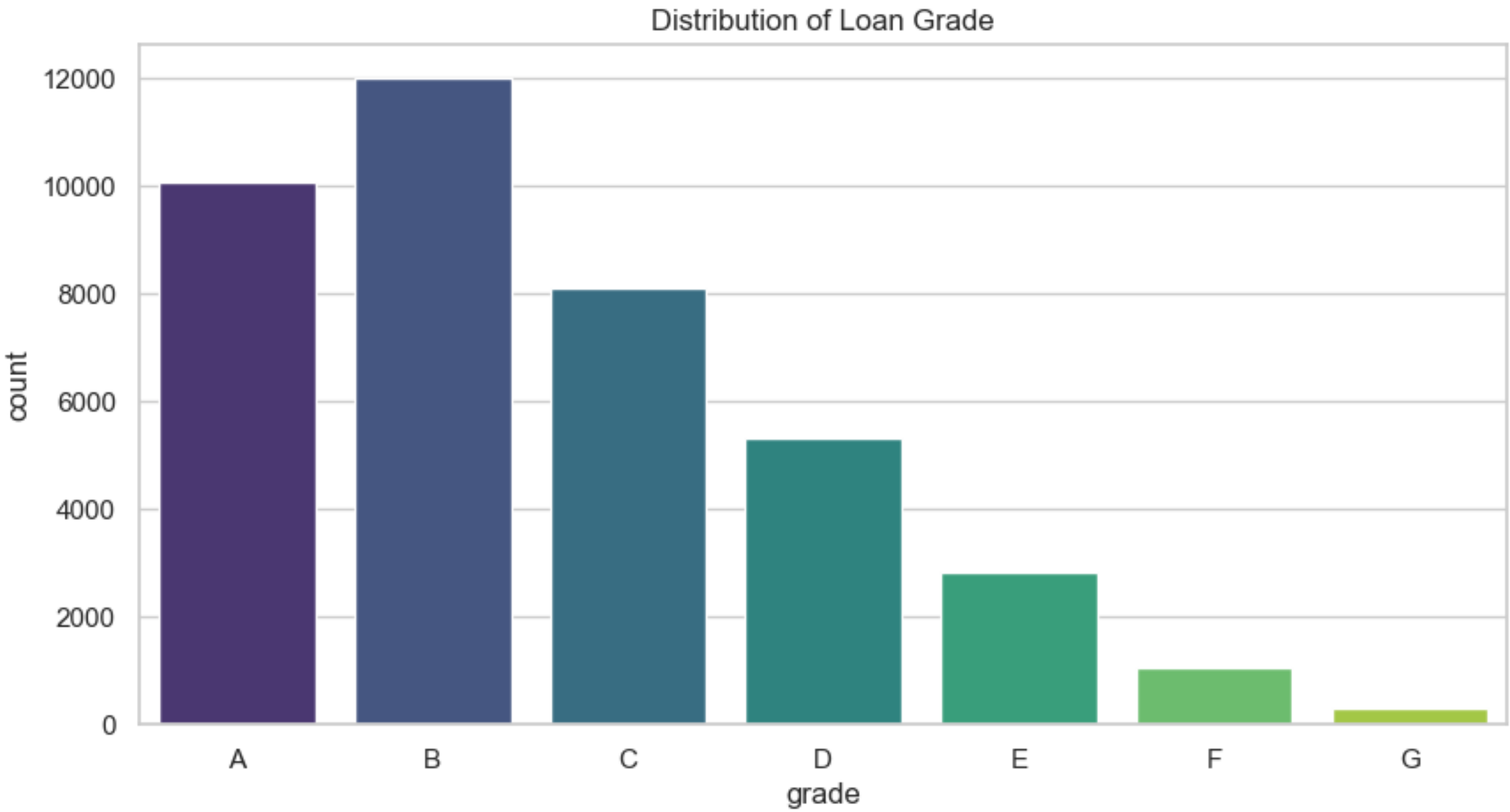
Just cleaned the data

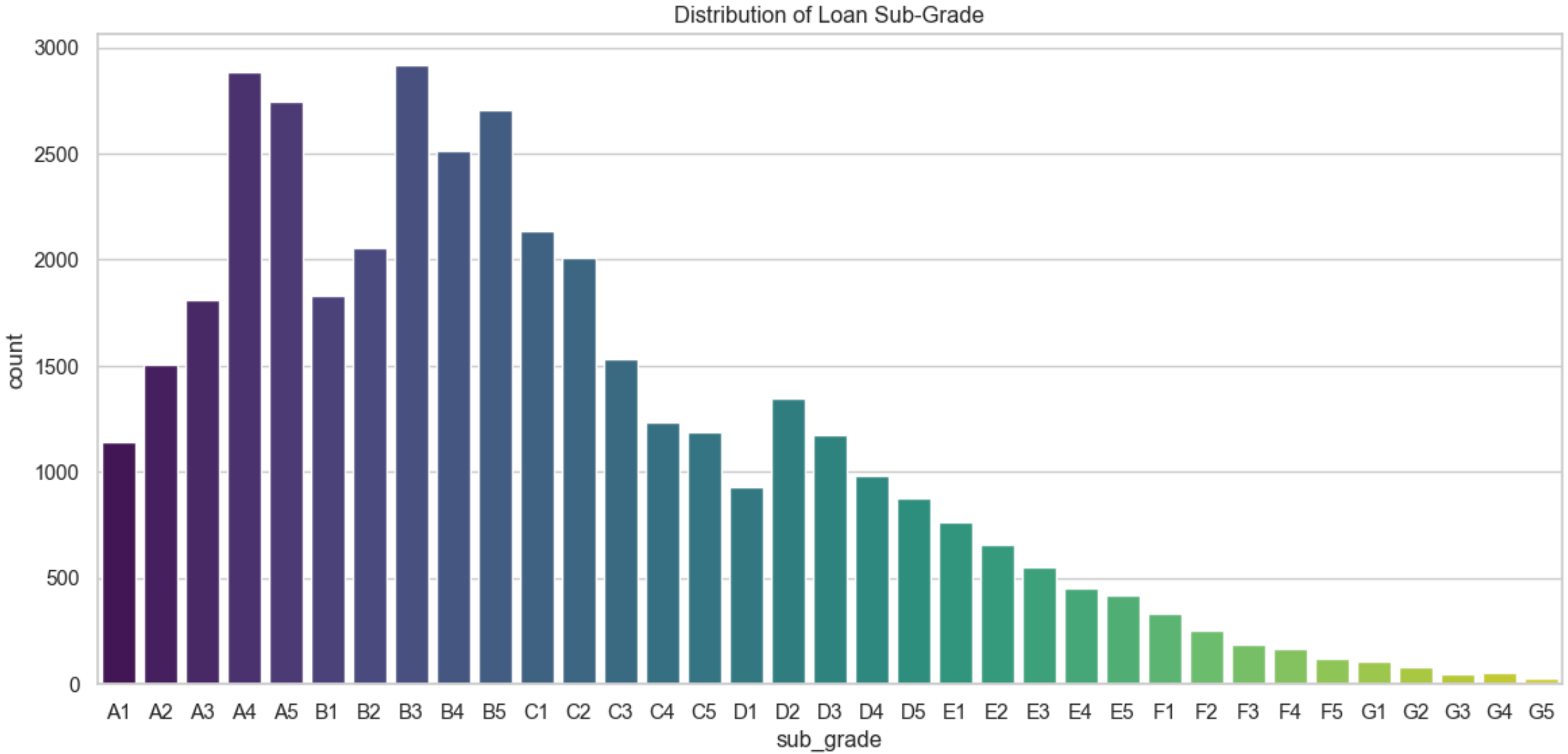
- First we filter the data and removed the missing value data from the data.
- Then after filtring the data it become 39717 row and 54 Column

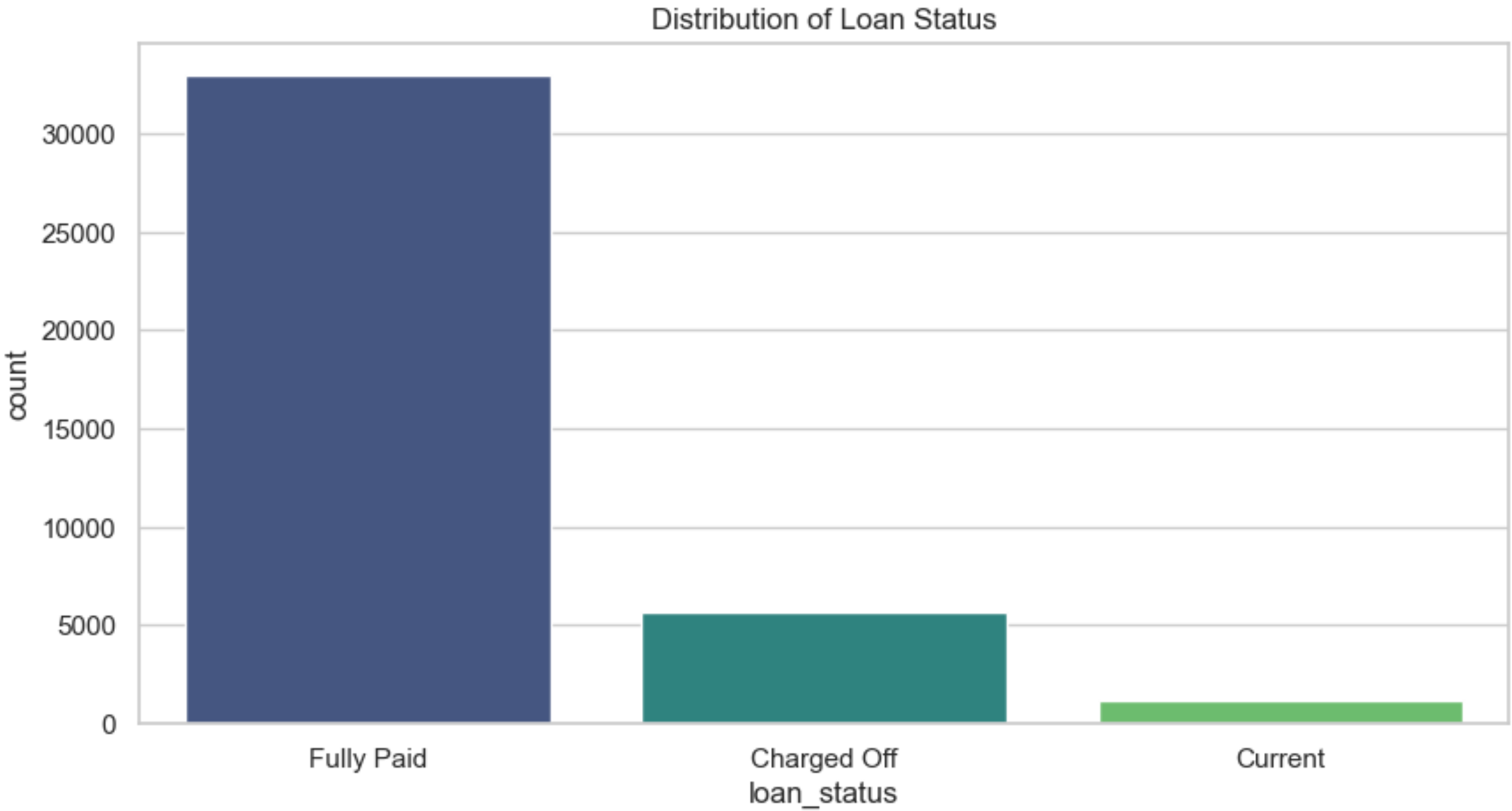
Univariate Analysis

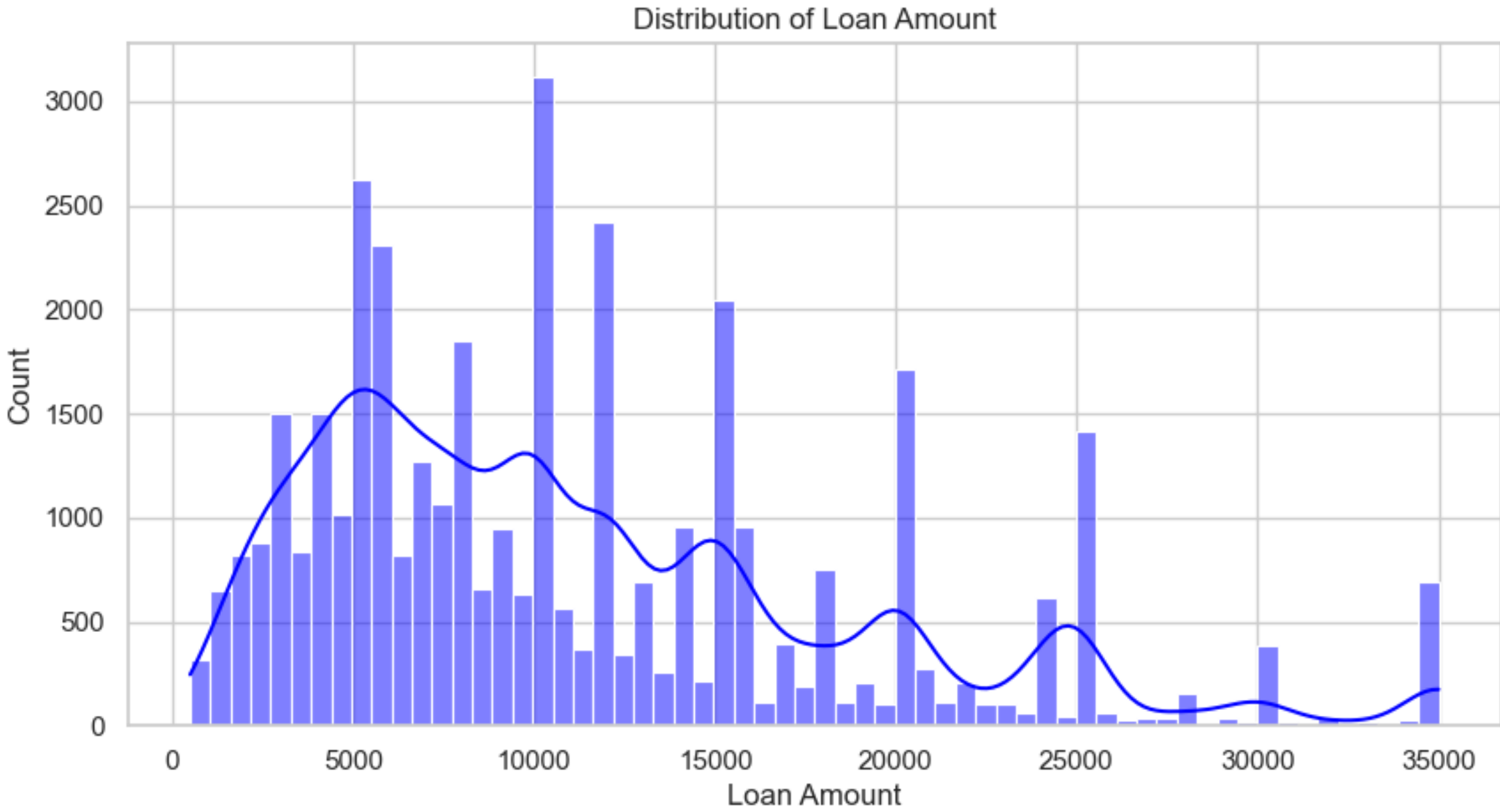
- Lots of loans are between$10,000 and 15,000.
- Lots of borrowers have annual incomes between 40,000 and 80,000.

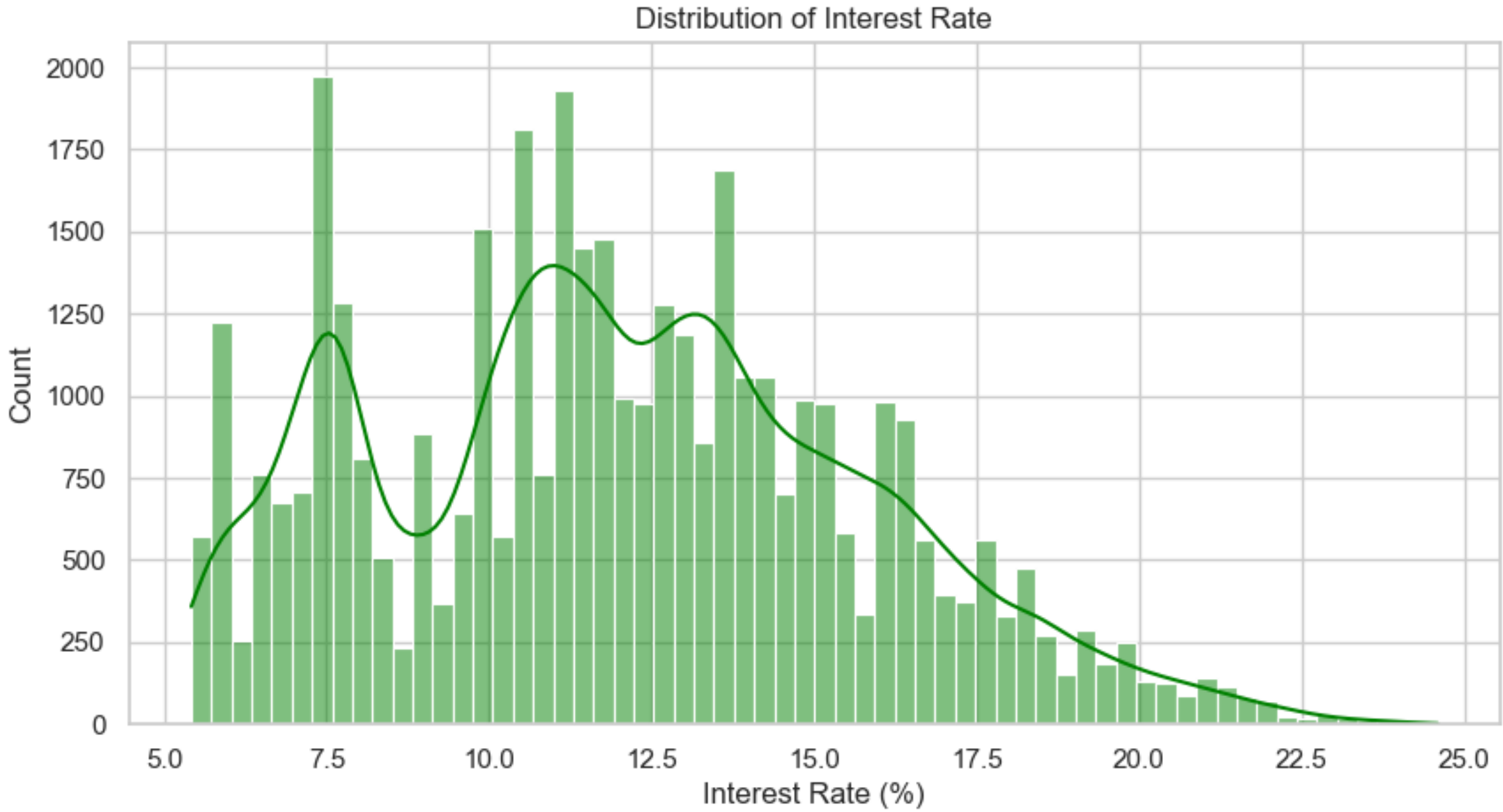


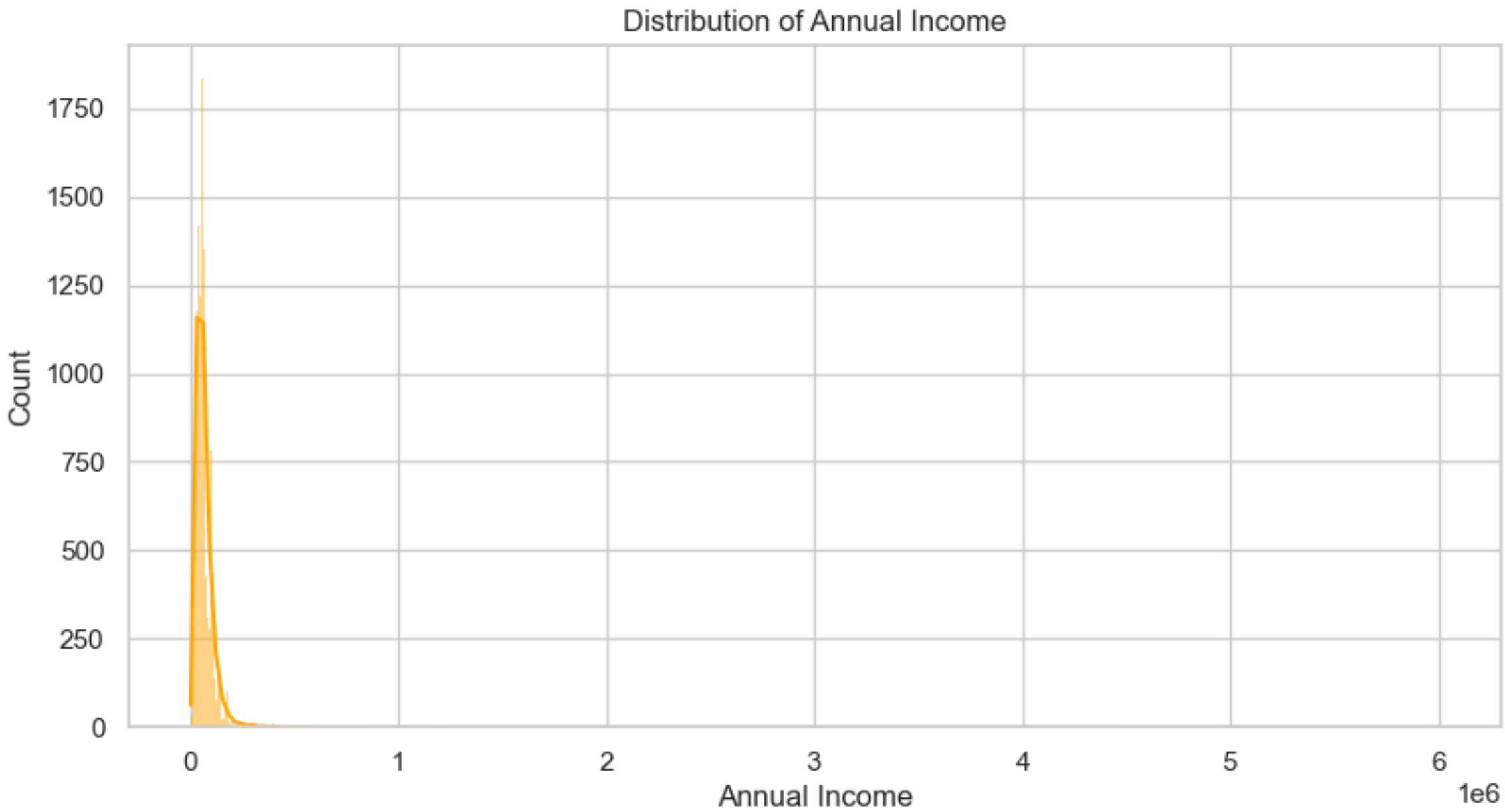

Distribution of Loan Grade
12000
10000
8000
6000
4000
2000
0
count
A
B
C
D
E
F
G
grade


Distribution of Loan Sub-Grade
count
3000
2500
2000
1500
1000
500
0
A1 A2 A3 A4 A5 B1 B2 B3 B4 B5 C1 C2 C3 C4 C5 D1 D2 D3 D4 D5 E1 E2 E3 E4 E5 F1 F2 F3 F4 F5 G1 G2 G3 G4 G5
sub_grade

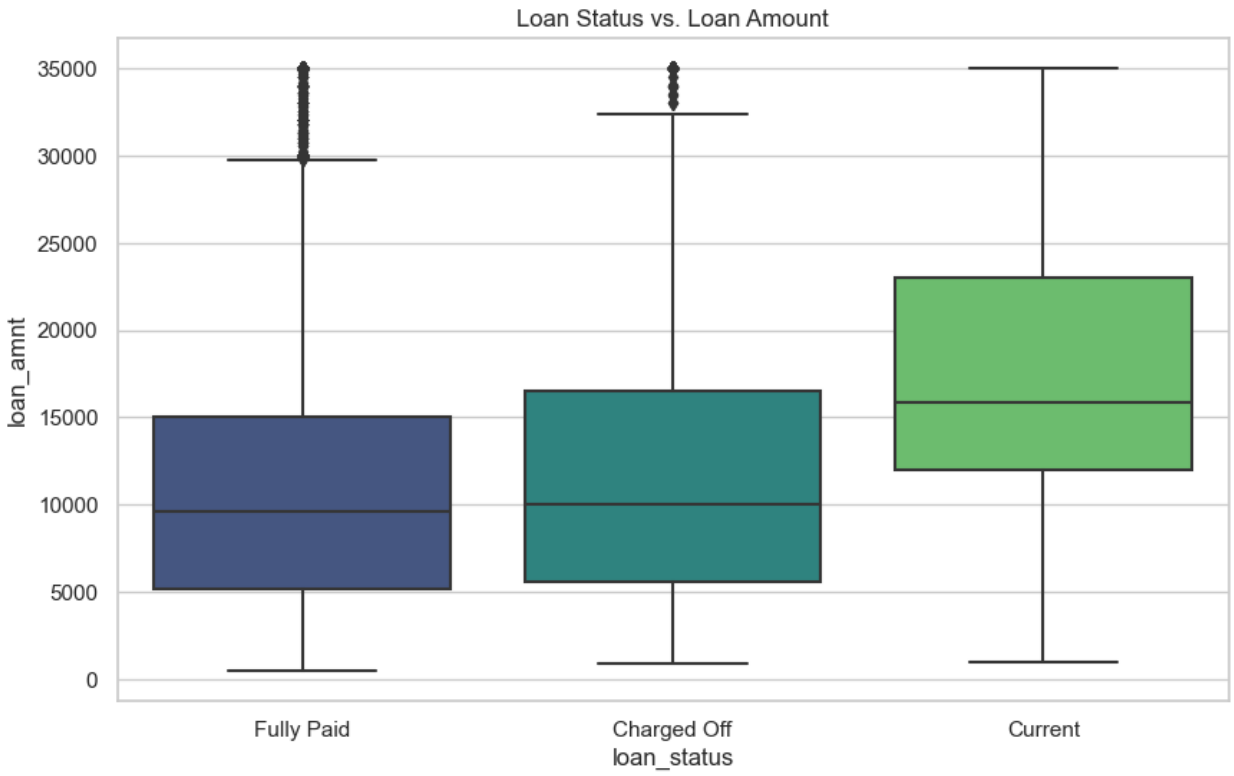

Distribution of Loan Status
30000
25000
20000
15000
10000
5000
0
count
Fully Paid
Charged Off
Current
loan_status

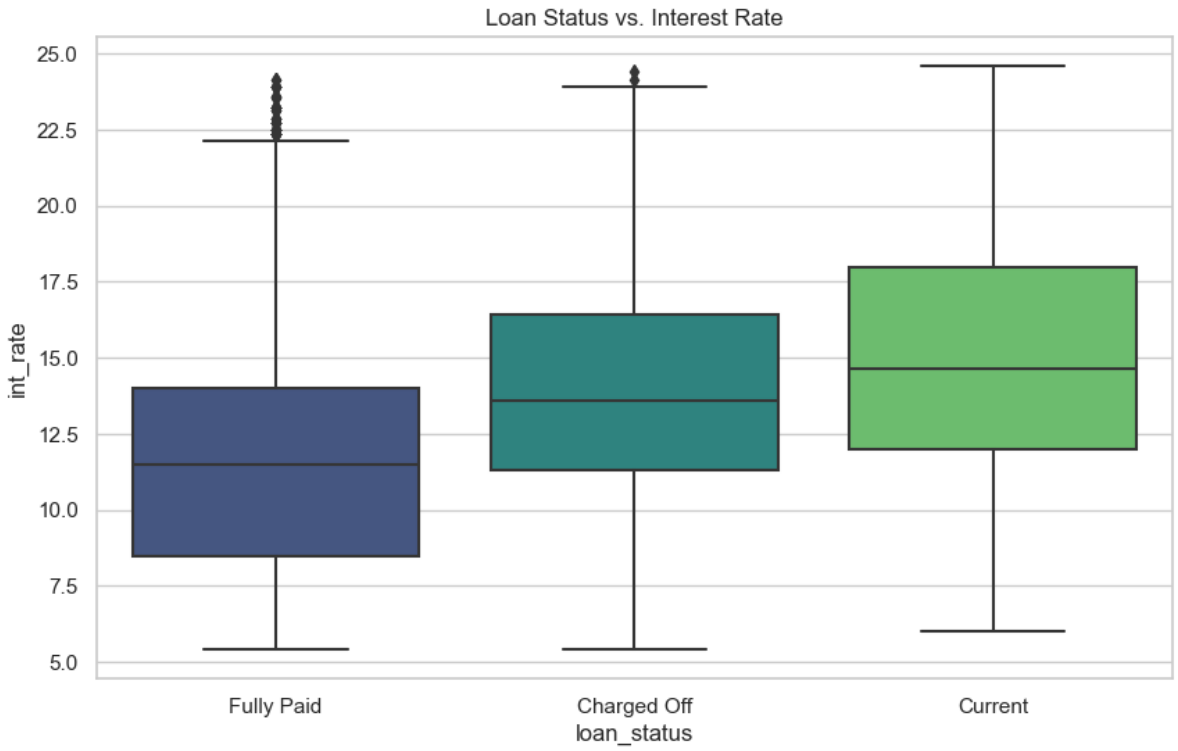

Distribution of Loan Amount
Count
3000
2500
2000
1500
1000
500
0
0
5000
10000
15000
20000
25000
30000
35000
Loan Amount


Distribution of Interest Rate
Count
2000
1750
1500
1250
1000
750
500
250
0
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
Interest Rate (%)

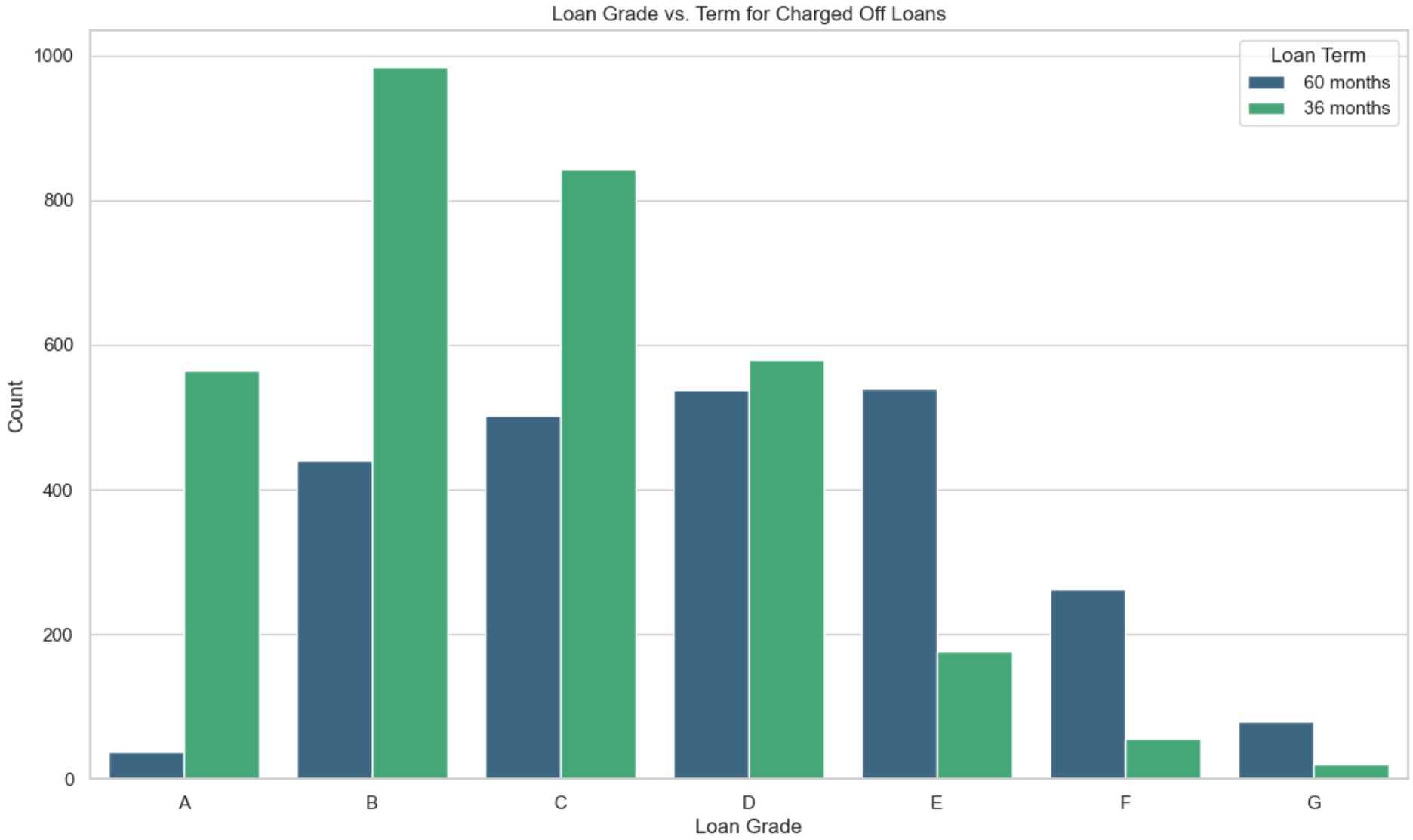

Distribution of Annual Income
Count
1750
1500
1250
1000
750
500
250
0
0
1
2
3
4
5
6
Annual Income
1e6

Bivariate Analysis





Multivariate analysis



Conclusions

1. Be careful when approving loans for high-risk borrowers.
2. Take higher interest rates for borrowers with lower credit scores manage risk.
3. Give small loans to borrowers who are at higher risk.