Lending Club Case Study

Sohail & Sumithra

Goal and Objective

Goal:

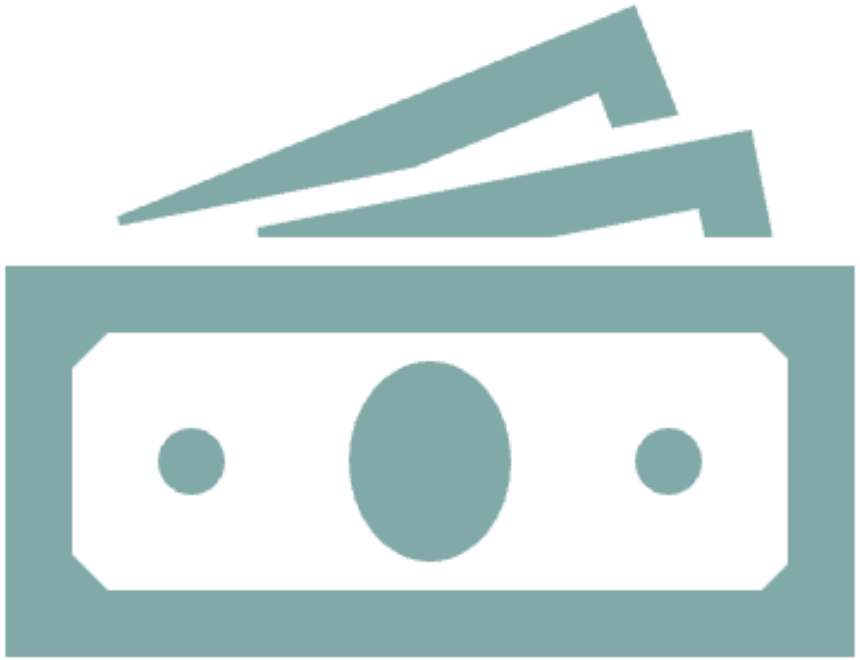
The goal is identify attributes of a loan which influence its repayment status i.e to identify which loans are likely to fully paid and which are in the risk of default.

Objective:

•● Reduce loss of business to the company by identifying applicants that are likely to repay the loan

•● Reduce financial loss by identifying applicants who are not likely to repay the loan

•Methodology:
We will use exploratory data analysis to identify borrower and loan attributes that influence the loan repayment status



Understanding the data

1. The loan dataset contains information for 39717 loans aganist 110 features

2. A significant number of columns have missing values. Some of the columns like acc_now_delinq, tot_coll_amt, tot_cur_bal, open_acc_6m, open_il_6m, open_il_12m etc are entirely missing. It indicates that these columns were not filled by the applicant/bank or were not applicable to loans in the data

3. The dataset contains numerical and categorical data. Some numerical data like 'int_rate' might be stored as string objects due to presence of '%'. These might require cleaning to convert them to numerical values for qunatative analysis.

Understanding the data

1. The loan dataset contains information for 39717 loans against 110 features

2. A significant number of columns have missing values. We dropped all columns with more than 75% missing value. We will be working with 54 remaining columns

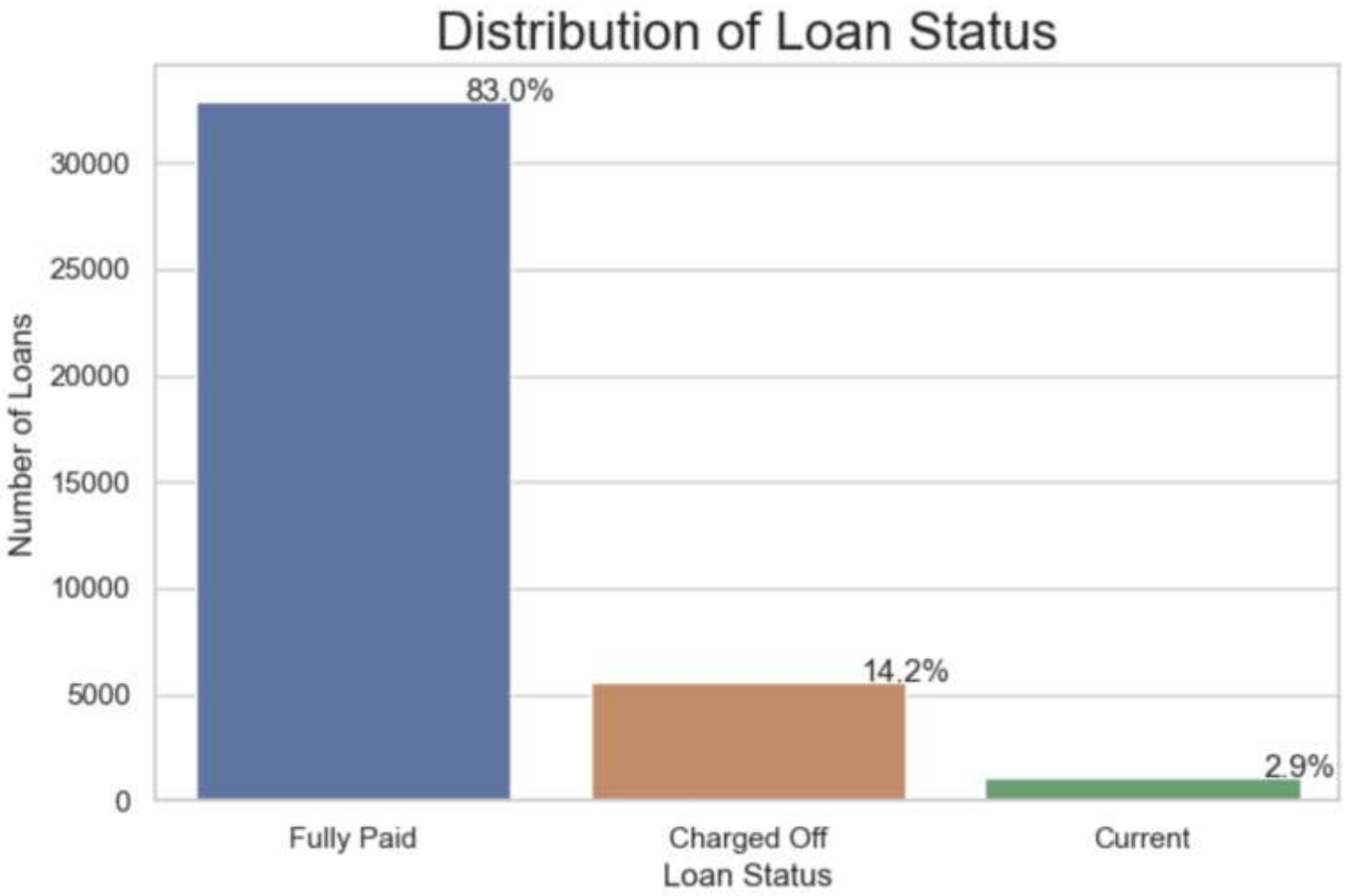
3. The dataset contains numerical and categorical data. Some numerical data like 'int_rate' might be stored as string objects due to presence of '%'. These might require cleaning to convert them to numerical values for quantitative analysis. Similarly, 'issue_d', 'earliest_cr_line', 'last_pymnt_d ', 'last_credit_pull_d' have been converted to the datetime format

3. 'int_rate' which is could be one of the most important driver of loan_status (based on business understanding) has been converted to float64 for quantitative analysis.

4. Some columns like 'revol_util' which are important from business point of view have been Imputed with relevant values values of columns like

Univariate Analysis

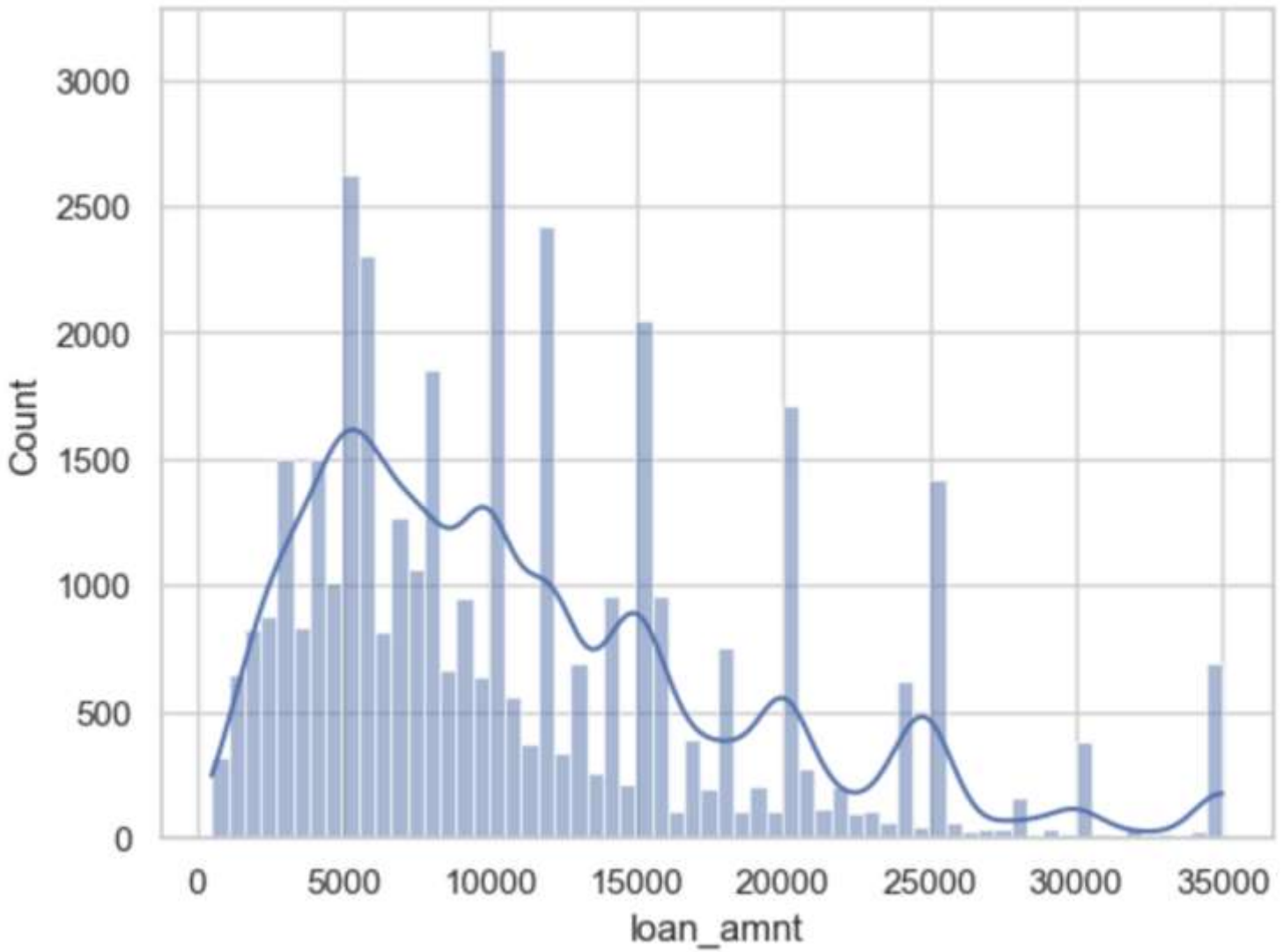
Loan Status Distribution



- Almost 82% of the loan records are paid
- 14% borrowers have defaulted on their loan repayment
- ~3% borrowers are currently servicing their loan repayments
- The distribution of Loan status show that the majority of loans is in the "Fully Paid" category, with a significant number in the "Charged Off" category, indicating default. This is critical for identifying patterns related to defaults.

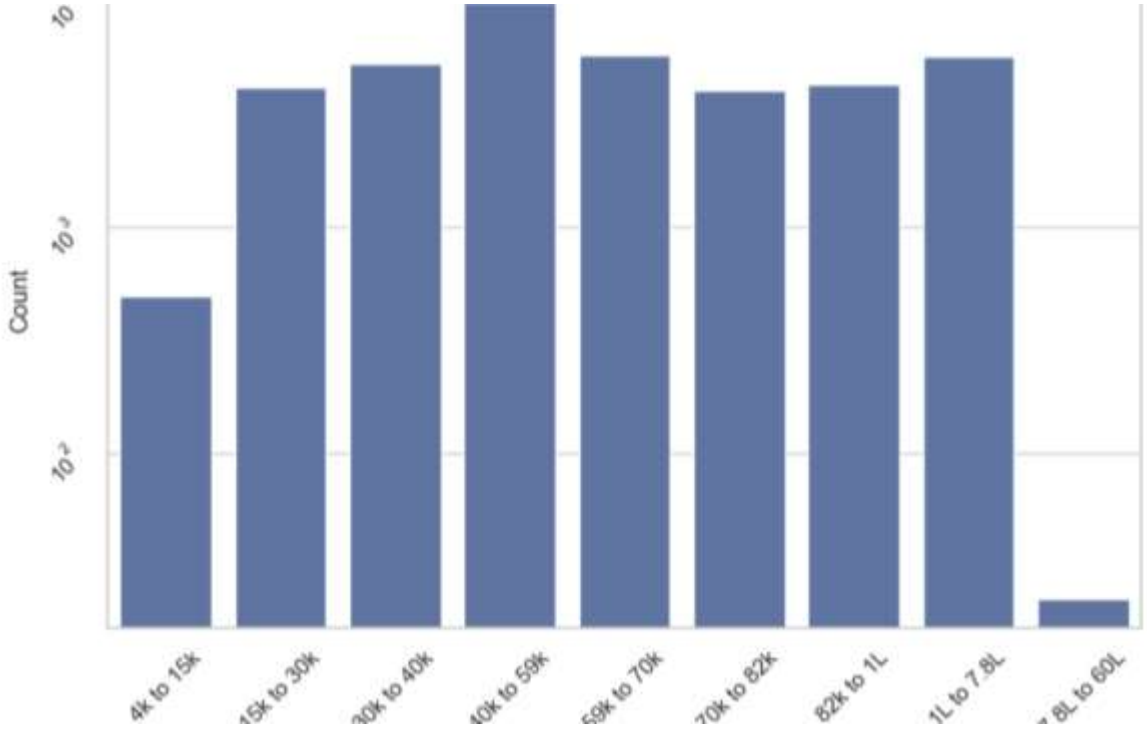
Loan amount distribution

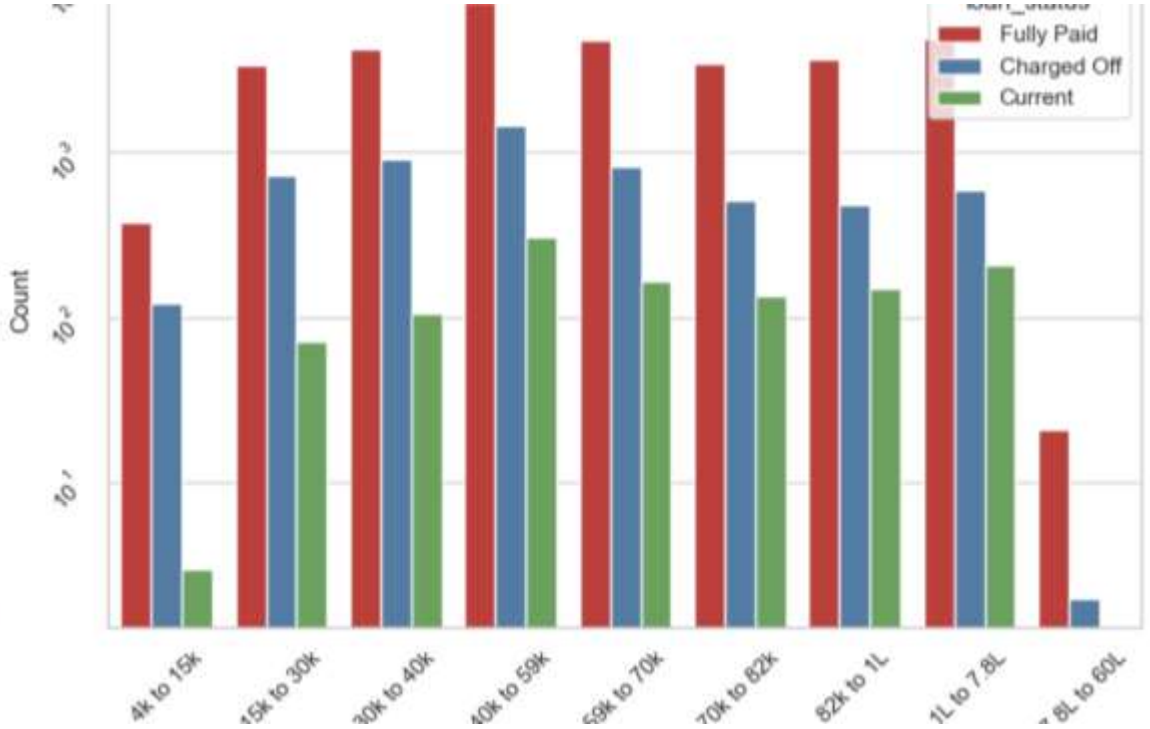
The distribution of loan amounts shows a right-skewed pattern, indicating that smaller loan amounts are more common than larger ones.



Annual income distribution

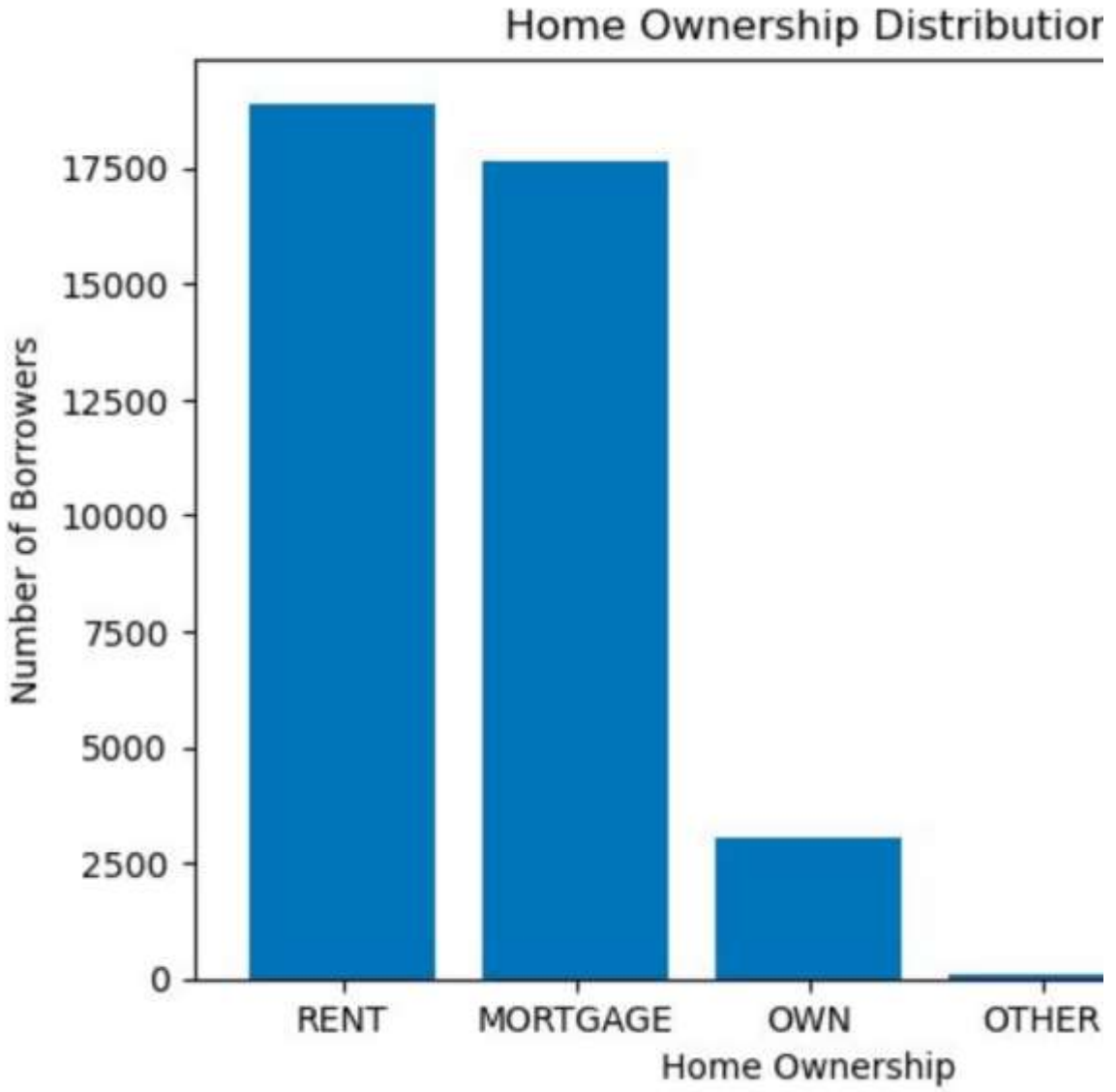
- Higher number of applicants are in the range of 15k to 7.8 L
- Significantly higher default rate in lower income range





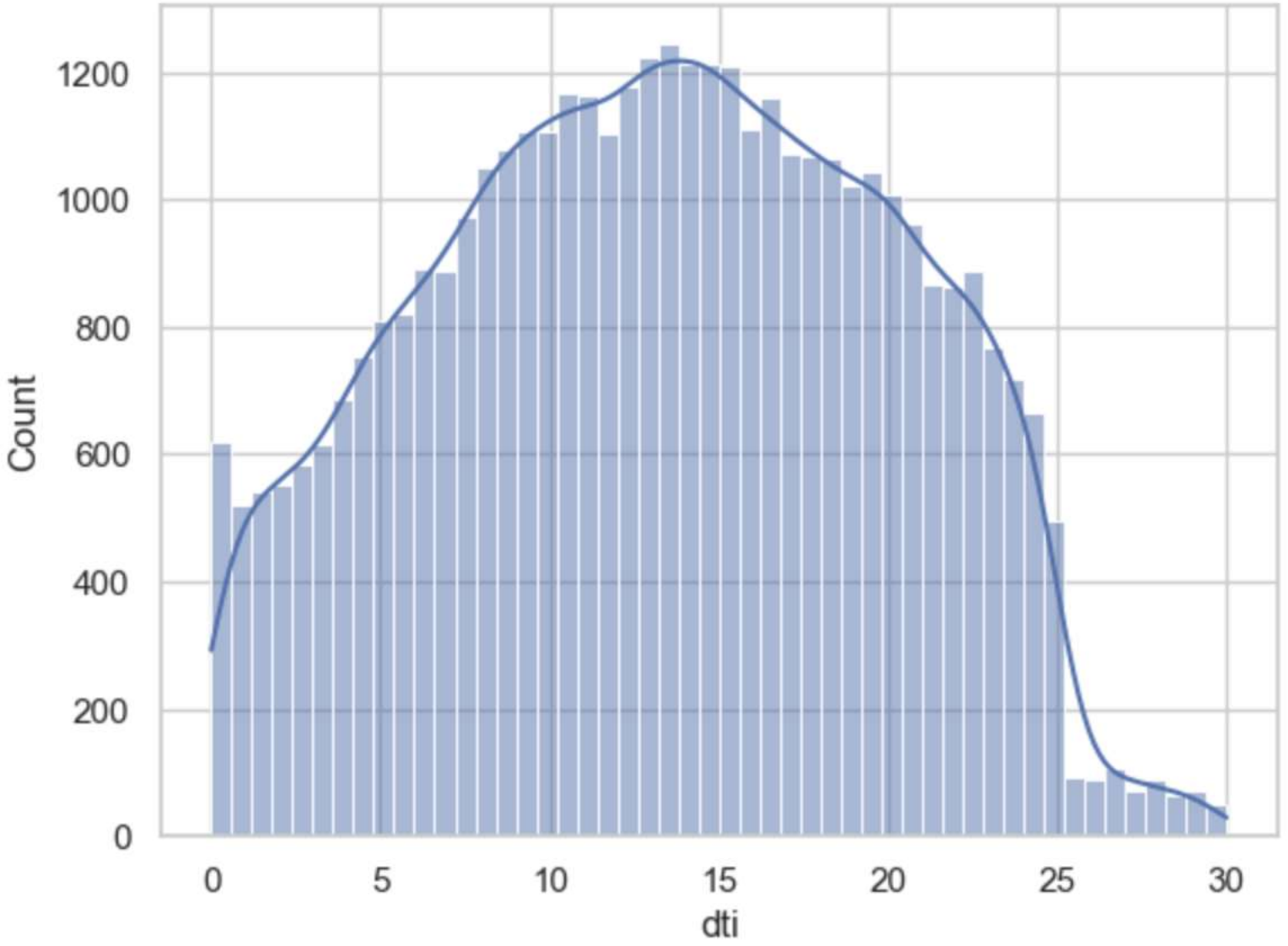
Home Ownership

The distribution of the home ownership show that most of borrowers rent their home or have a mortgage against the home. Only a small number have indicate own home ownership



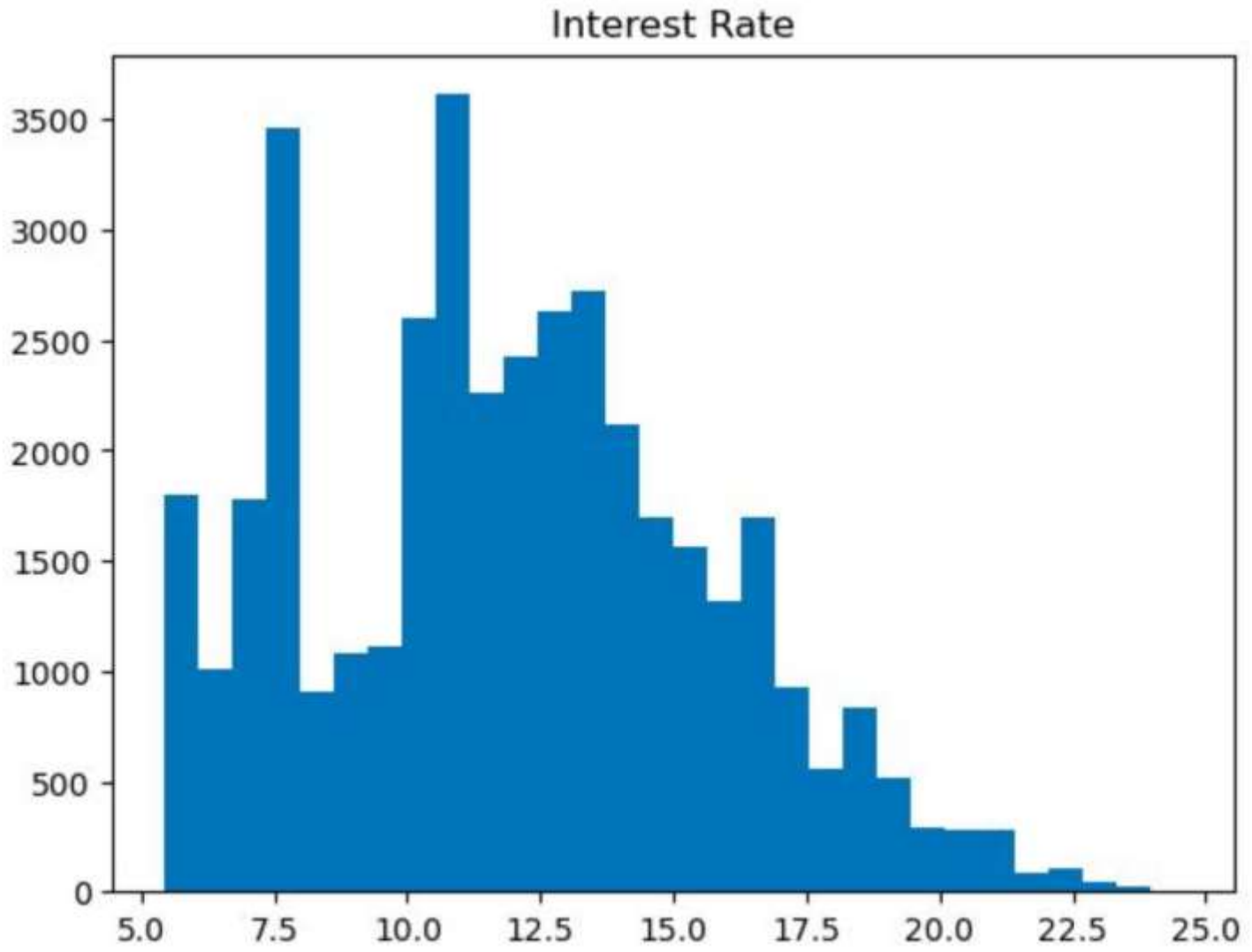
DTI (Debt-to-Income Ratio) Distribution

The DTI ratio has a relatively uniform distribution across a wide range of values, with a slight right skew.



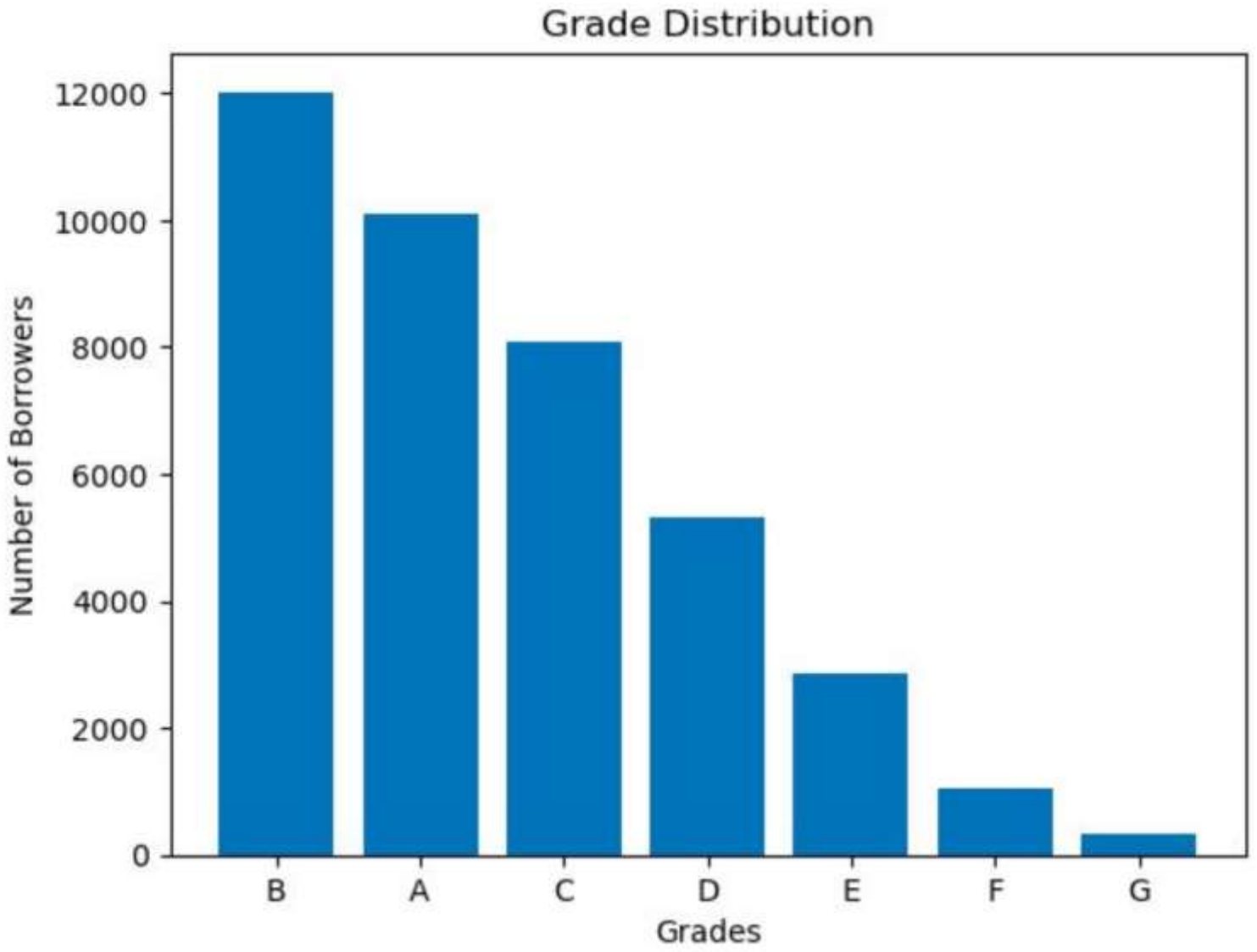
Interest Rate Distribution:

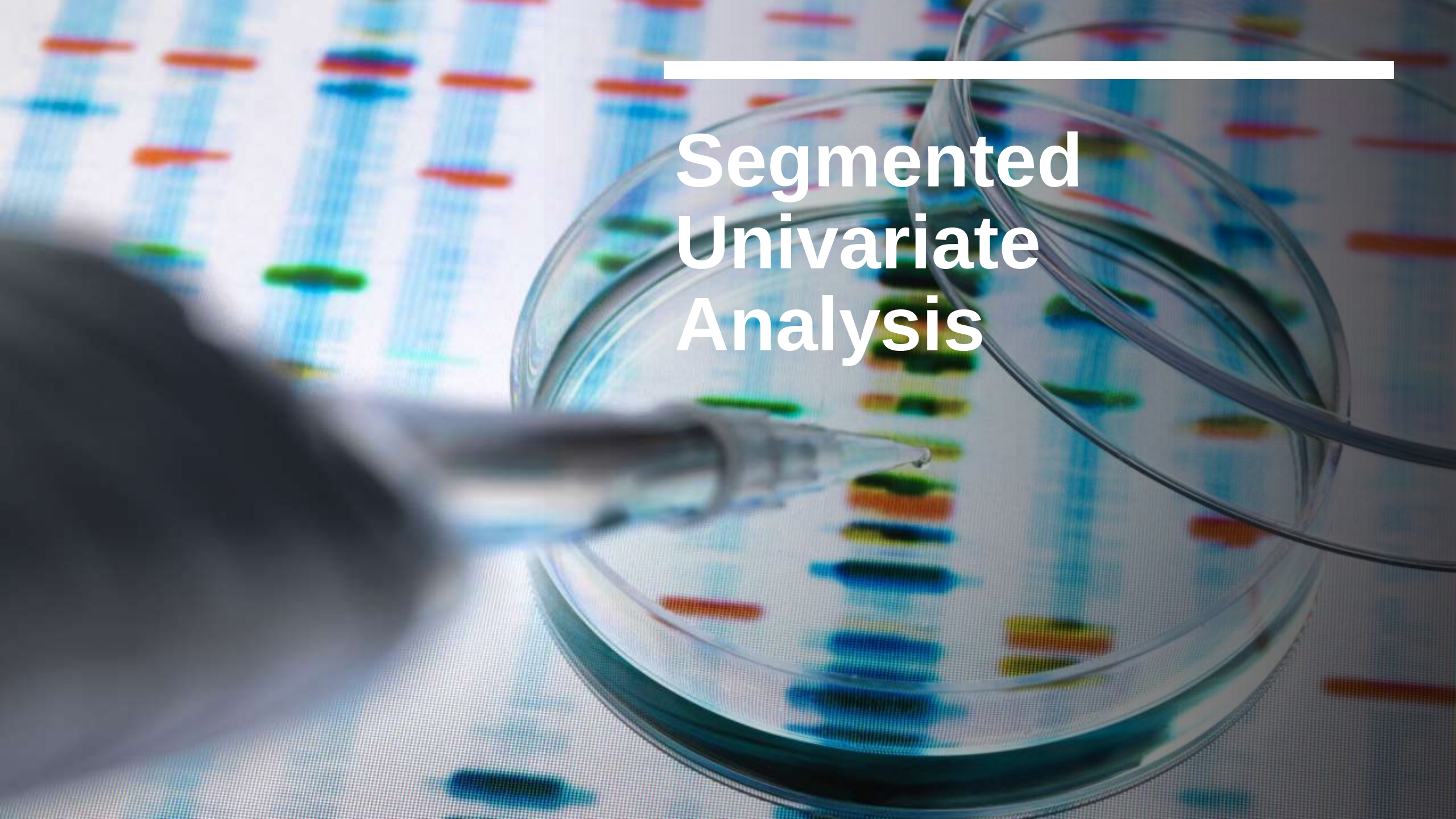
The interest rate distribution does not readily indicate any conventional pattern, reflecting different risk categories assigned to loans



Grade Distributio

The distribution of grades show that most borrowers are assigned in the grade B, followed by A, C, D and so on

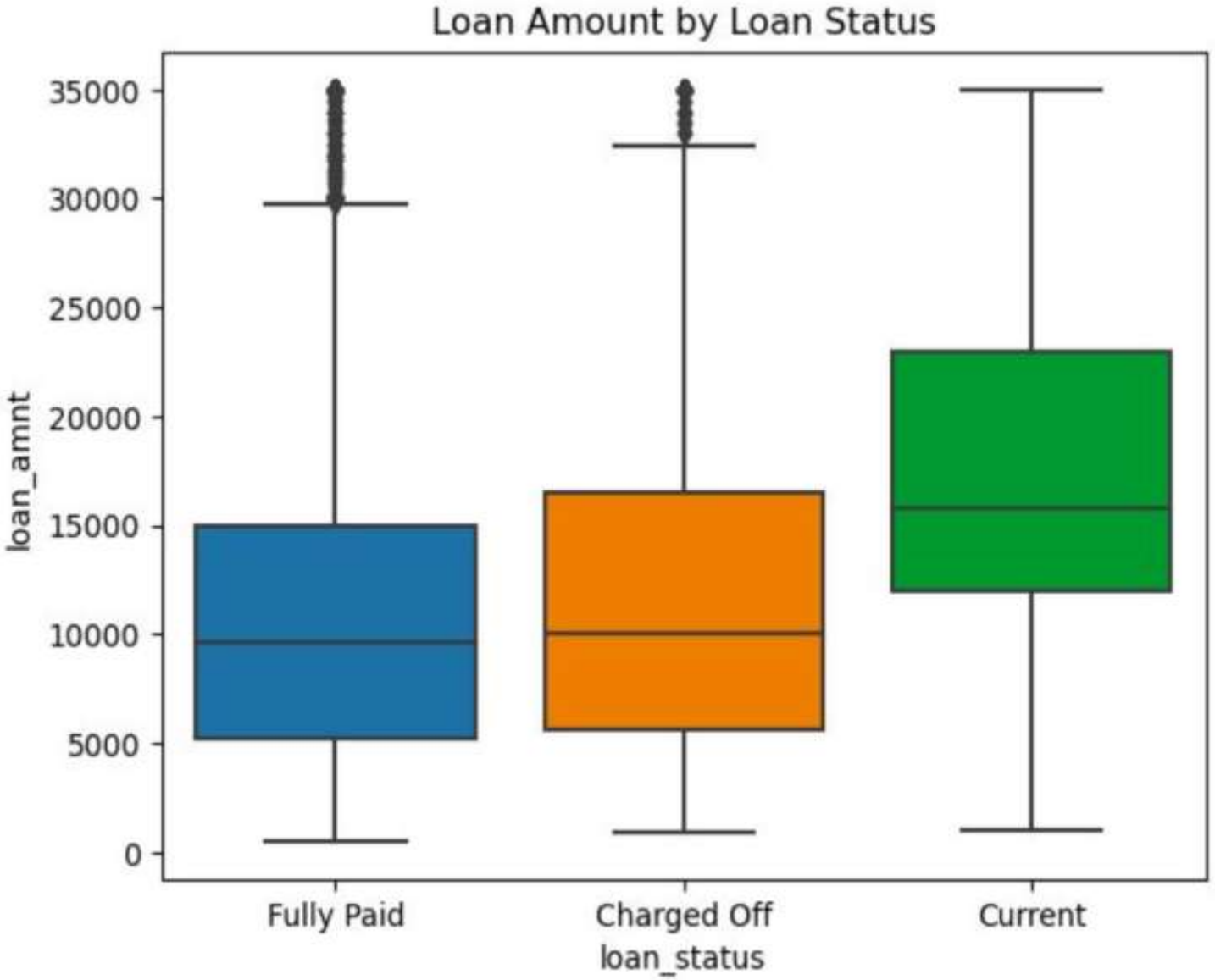




Segmented Univariate Analysis

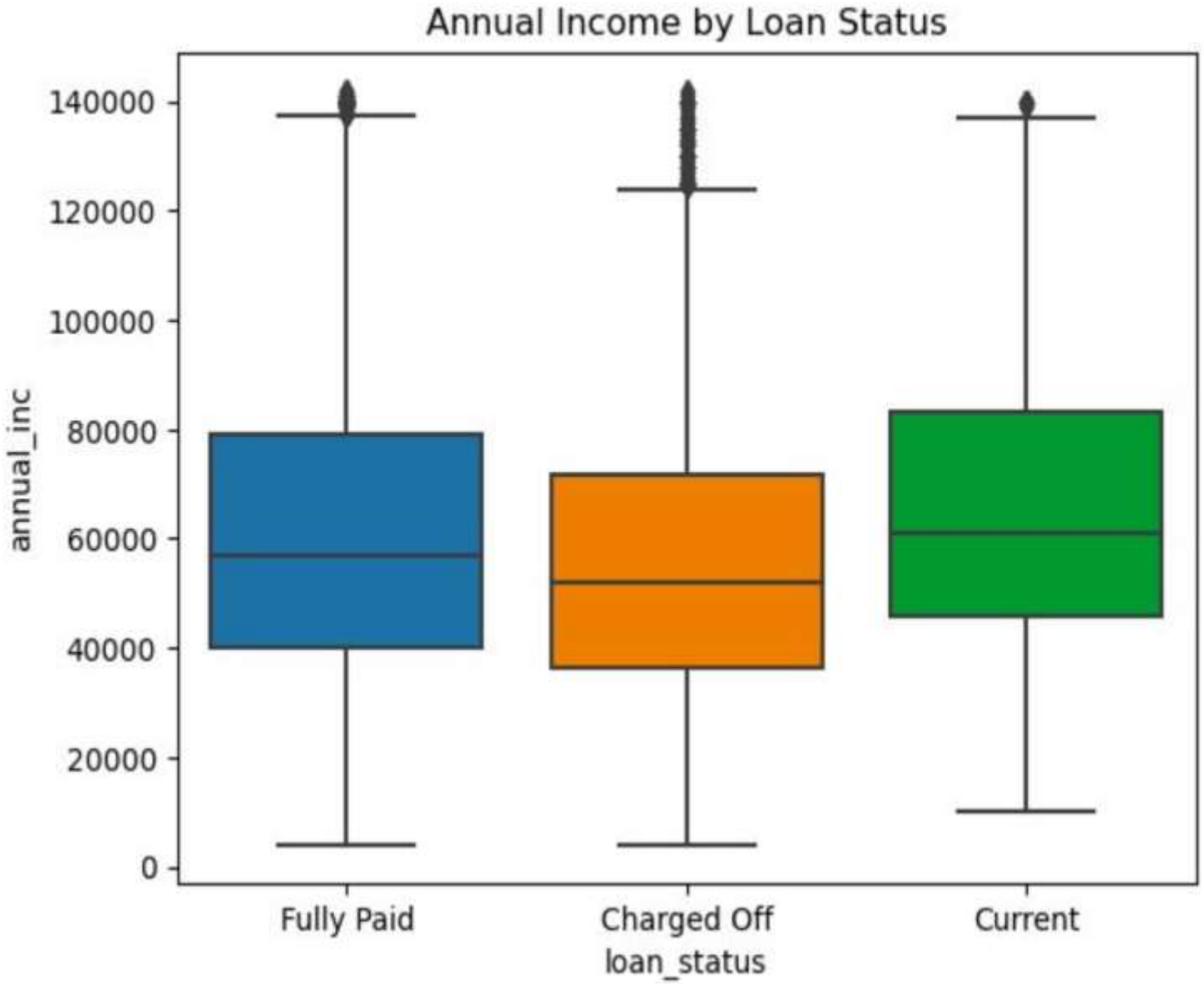
Loan Amount vs. Loan Status

The median loan amount for charged-off loans is slightly higher than for fully paid loans, suggesting that higher loan amounts may have a slightly increased risk of default. However, the difference is not substantial, indicating other factors also play a significant role in the likelihood of default.



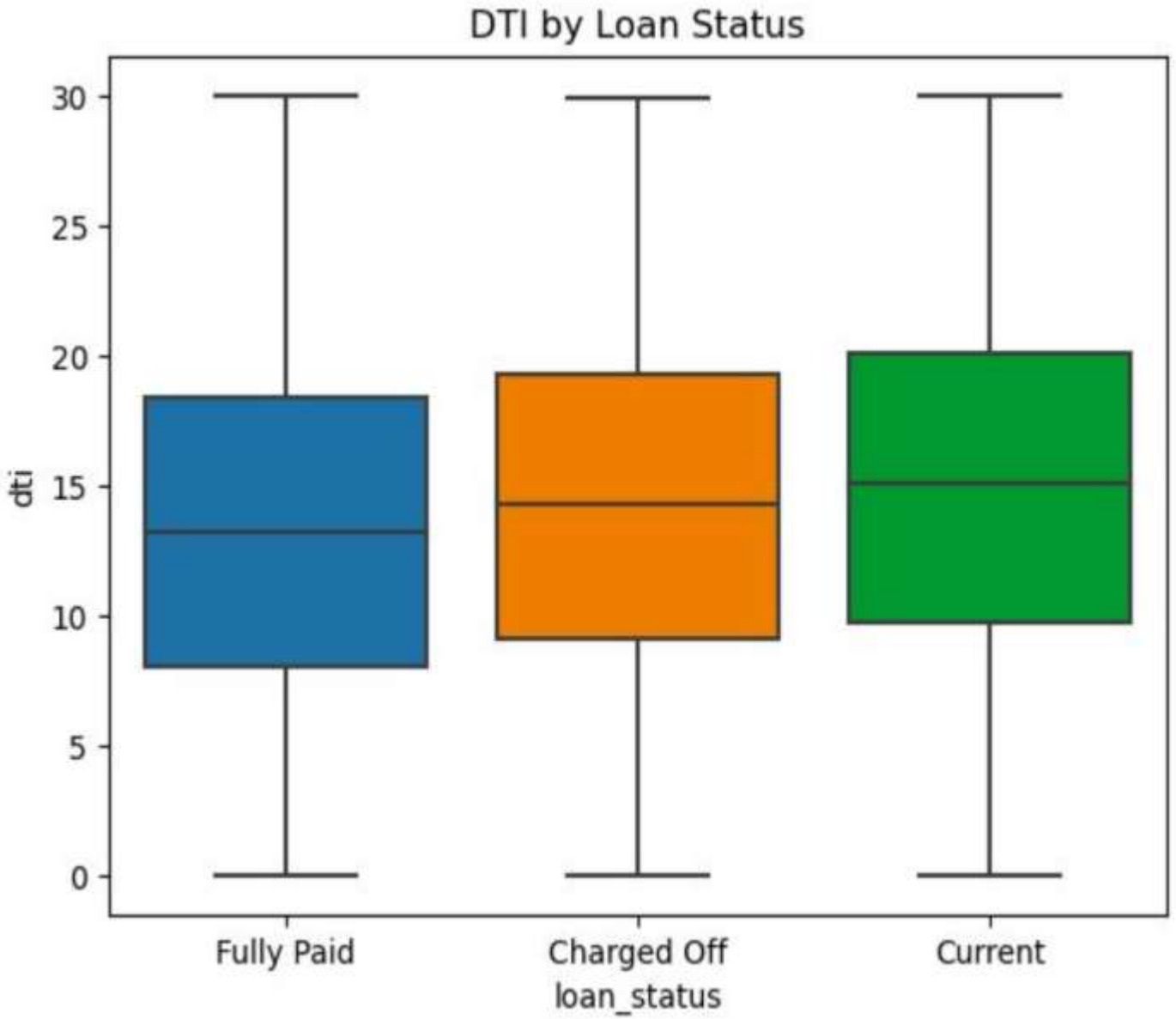
Annual Income vs. Loan Status

The median annual income appears to be lower for charged-off loans compared to fully paid loans, though the difference is not very pronounced. This suggests that borrowers with lower incomes might be at a higher risk of default, but income alone is not a definitive predictor.



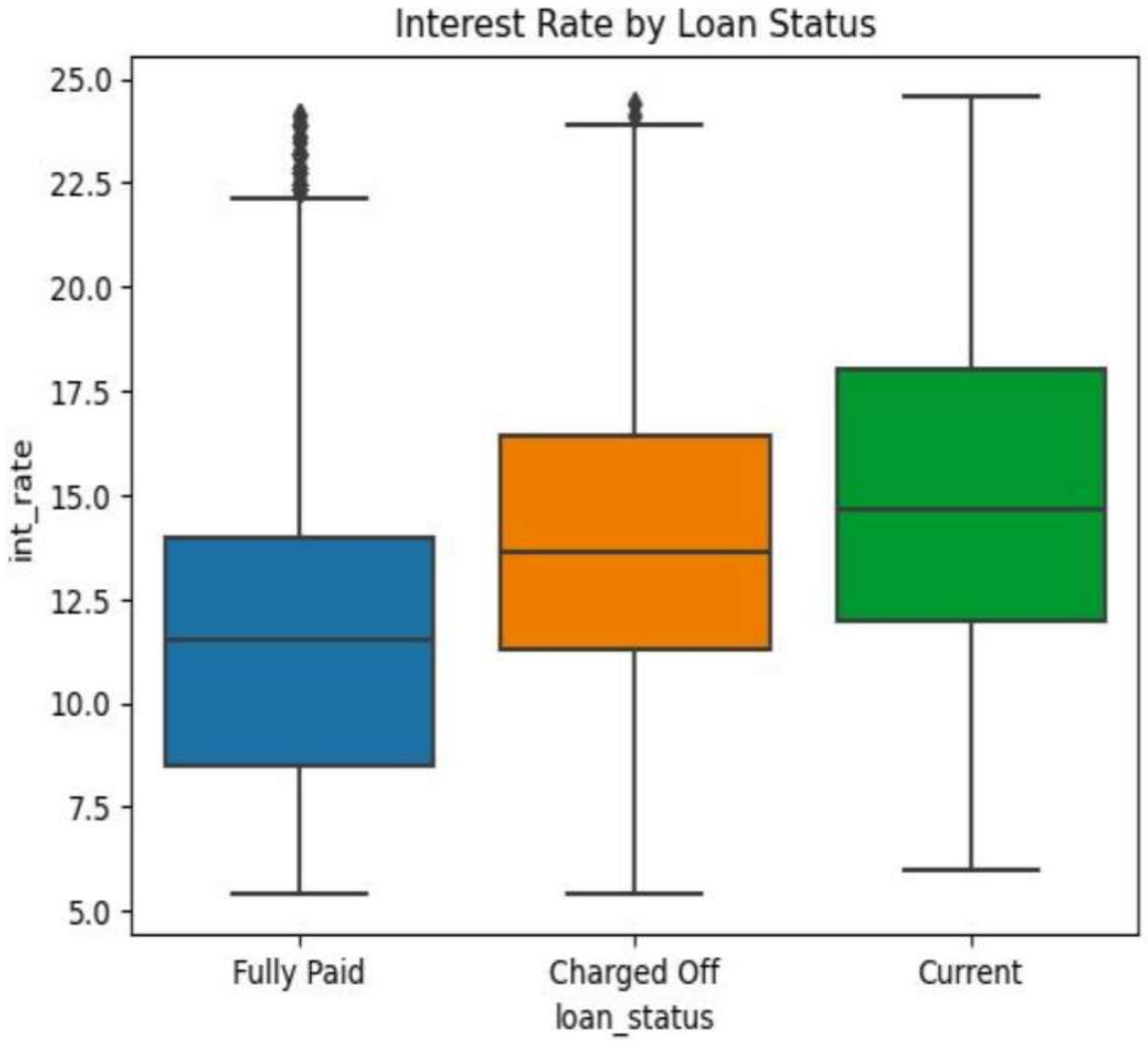
DTI vs. Loan Status

The median DTI ratio is higher for charged-off loans than for fully paid loans. This indicates that a higher debt-to-income ratio is associated with an increased risk of default, which aligns with the expectation that borrowers with higher financial obligations relative to their income are more likely to face difficulties in repaying their loans.

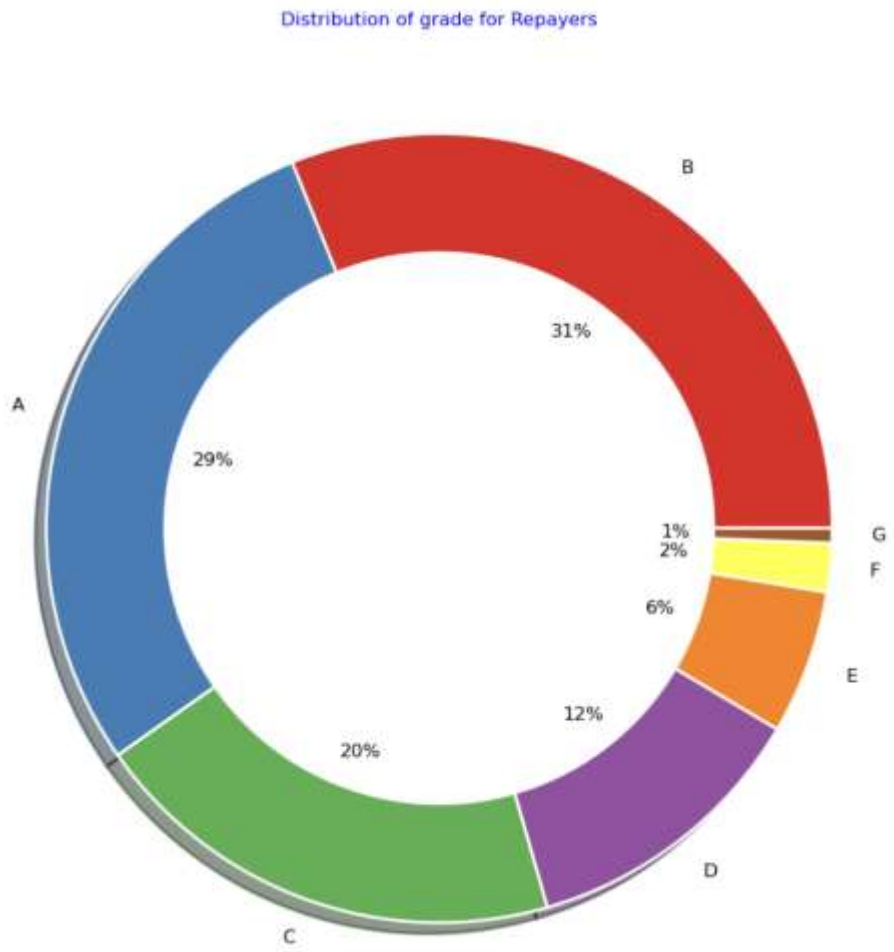


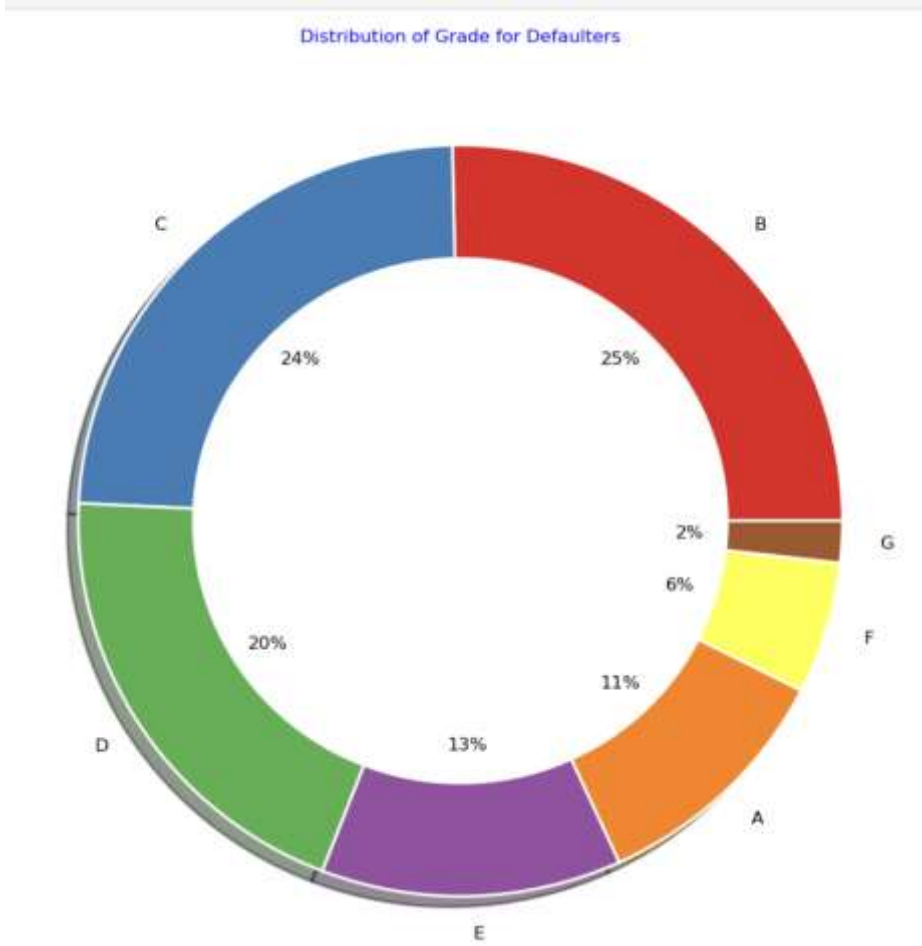
Interest Rate vs. Loan Status

There is a noticeable difference in the interest rates between fully paid and charged-off loans, with charged-off loans having higher median interest rates. This suggests that loans with higher interest rates, indicative of higher risk as assessed by the lender, are more likely to default.



Grade vs Defaulting rate

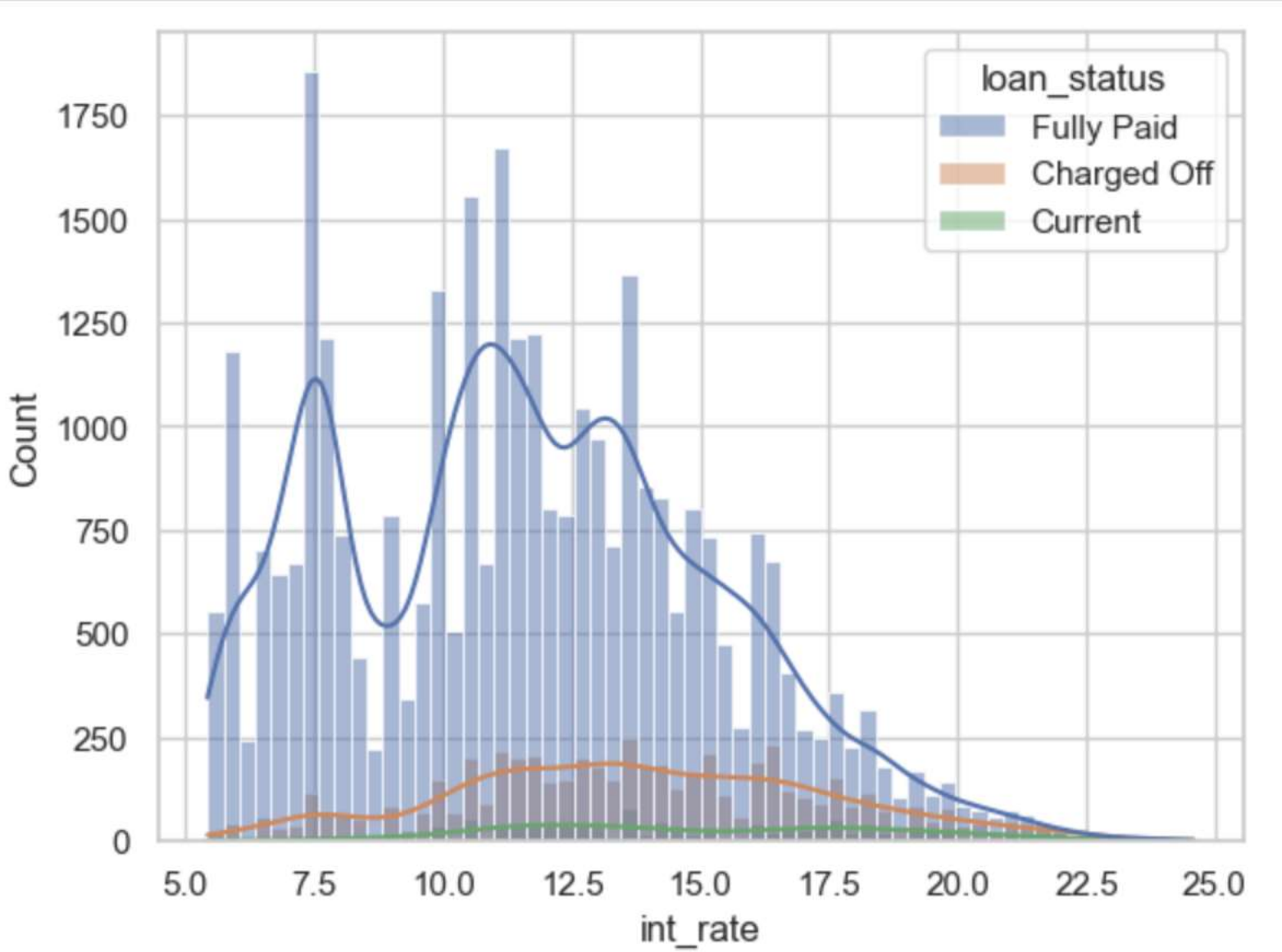




Applicants with grades D, E, F, G are in higher risk of defaulting

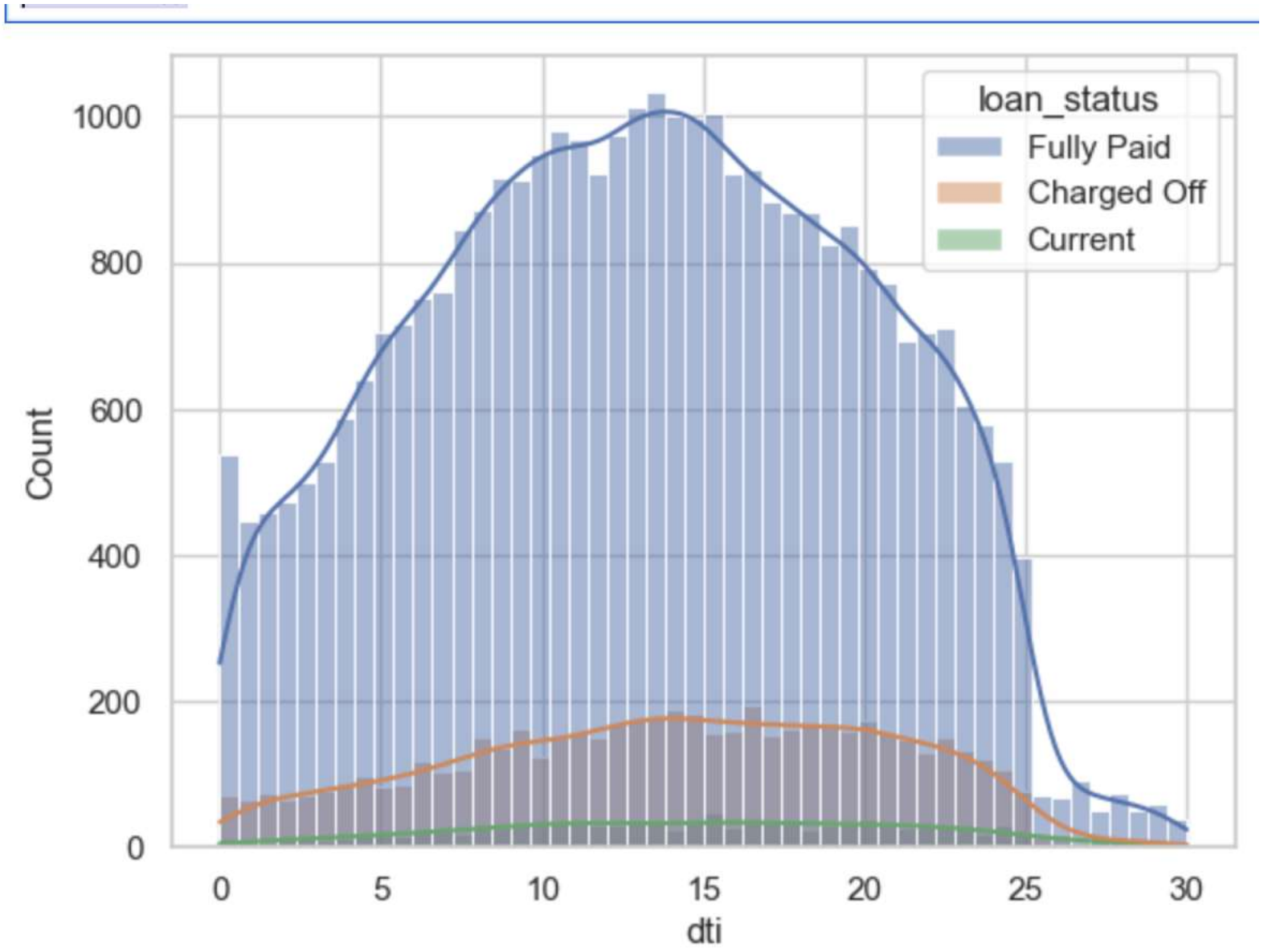
Interest rate vs Default rate

As the interest rate increases rate of defaulting also increases



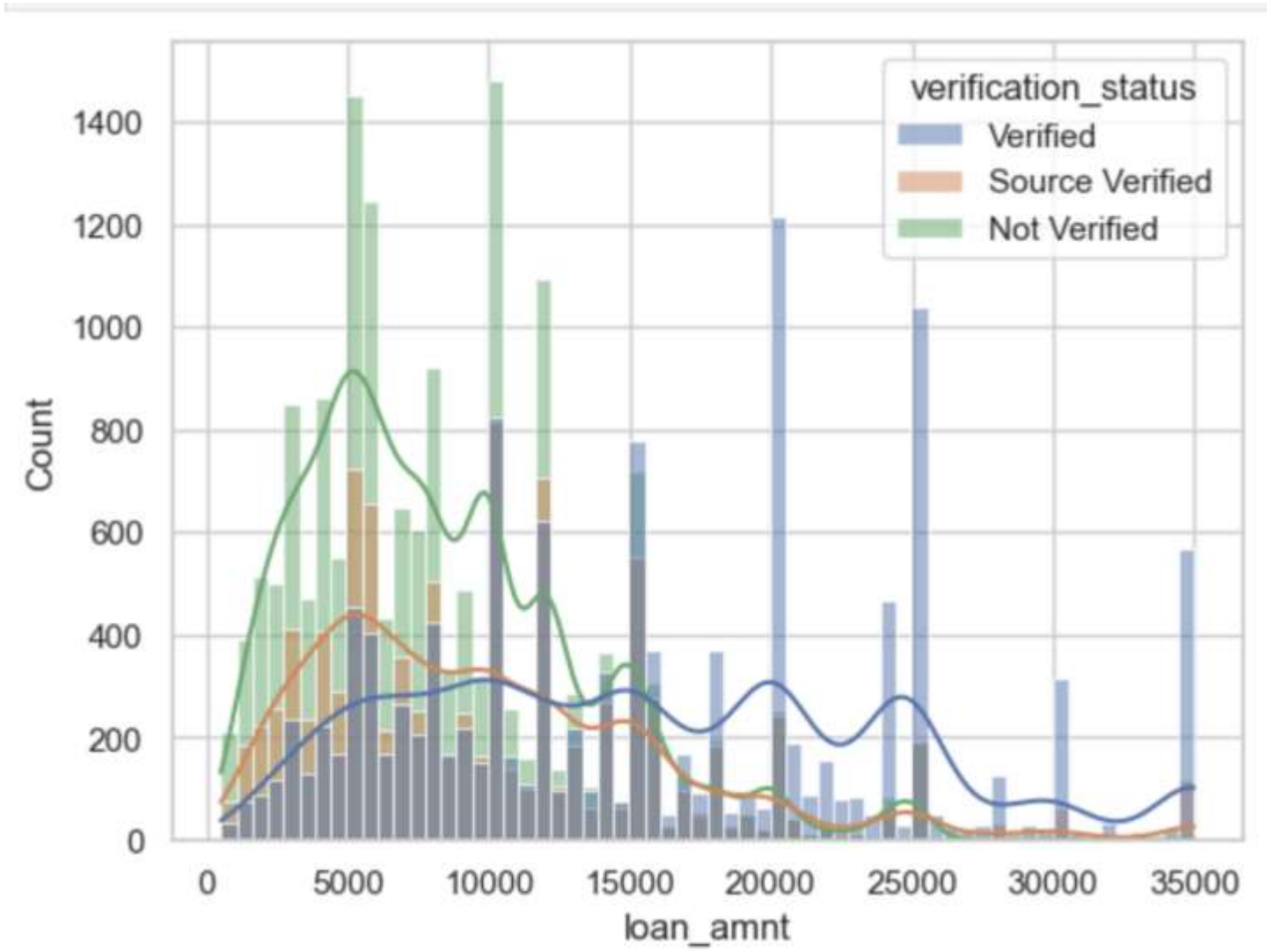
DTI vs Loan Status

Most applicants are in the range of 8.0 to 20.0. Default rate is also high in the same range



Loan amount vs Verification Status

Lower loan amounts have lesser verification done which needs to be increased to reduced default rate



Recommendations

Recommendations

Based on these findings to mitigate the risk of loan defaults, the company could consider the following actions:

1.DTI Ratios: Applicants with high DTI ratios could be subjected to more stringent review processes.

2.Interest Rate Adjustments: Higher interest rates correlate with higher default rates. Adjusting lending criteria or pricing for these loans could help balance risk and return.

3.Annual Income (annual_inc): Lending club should do additional judgements of those borrowers with relatively lower annual income due to higher risk of default. Income range of 40-70K are more likely to default.

4.Loan Amount Limits: Implementing stricter limits or higher scrutiny for larger loan amounts could also help manage default risk.

5.Verification is not done for lower loan amounts and defaulting is also high in the same range, hence verification for lower loan amounts also needs to done.

Ideally the company should target borrowers with higher income but for relatively lower income borrowers loan default risk can be mitigated for them by providing lower loan amounts, have lower installments, lower interests etc