LENDING CLUB CASE STUDY

- BY MAHESH CHANDRA AND RAVI KRISHNA

Problem Statement

- A finance company specialises in lending various types of loans. When the company receives a loan application, company has to decide on the basis of applicant's profile if loan should be given or not. The risk involved in lending the loan being -

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

Dataset

loan.csv - Loan related data for various users from 2007 to 2011

data_dictionary.csv - Dictionary to used to understand variables defined in the loan dataset

Problem Solving Approach

Data Understanding

Develop understanding of various columns present in loan.csv file

Data Cleaning and Manipulation

Cleaning up irrelevant fields and perform data manipulation for better data readability

Data Analysis

Perform Univariate, Segmented Univariate, Bivariate, Multivariate analysis on the dataset

Conclusion

Develop conclusion from the analysis

Libraries Imported

- Numpy
- Pandas
- Matplotlib
- Seaborn

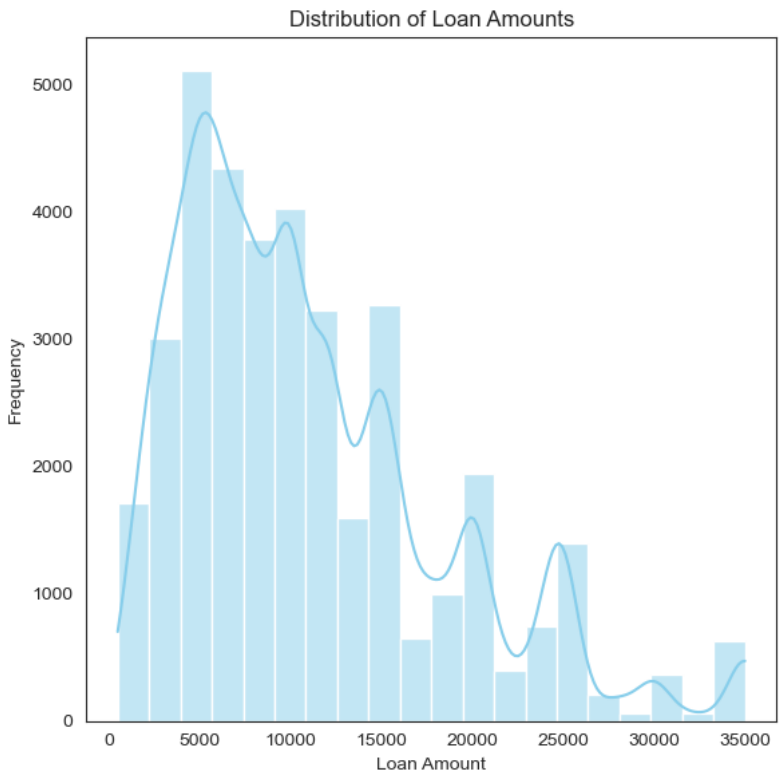
Data Cleaning

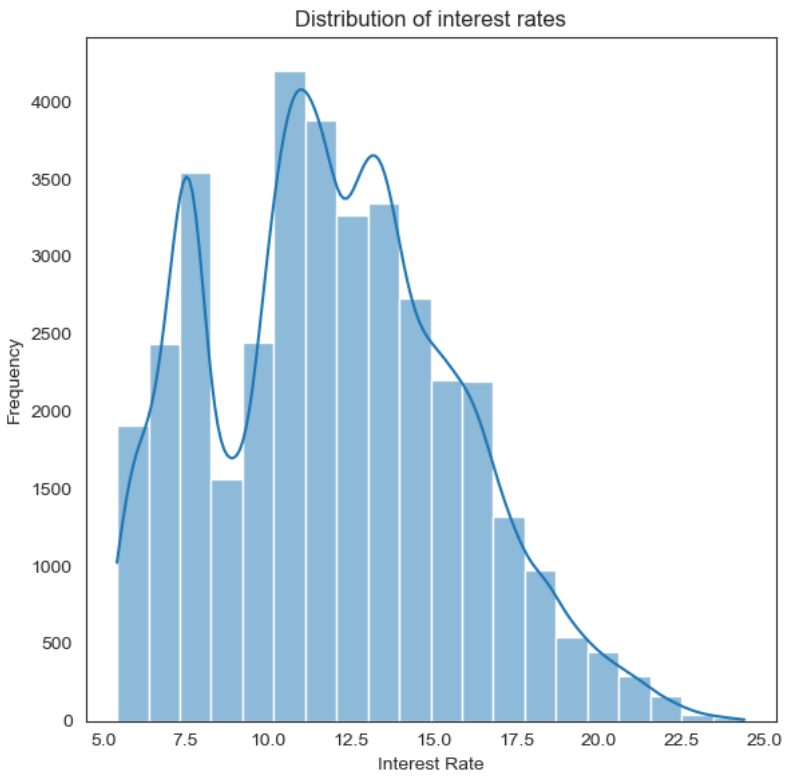
- Removed columns which has majority of the values as null since they are not of any use for our analysis
- Identified variables which were not present at the time of loan distribution so can be ignored

Data Manipulation

- Fields like interest rate, emp_length which are converted to numeric for better analysis
- Ignored values with ongoing loan as they cannot be referenced for analysis
- Extracted issue_month, issue_year from issue_date
- Filtering of Outliers for annual_inc

UNIVARIATE ANALYSIS



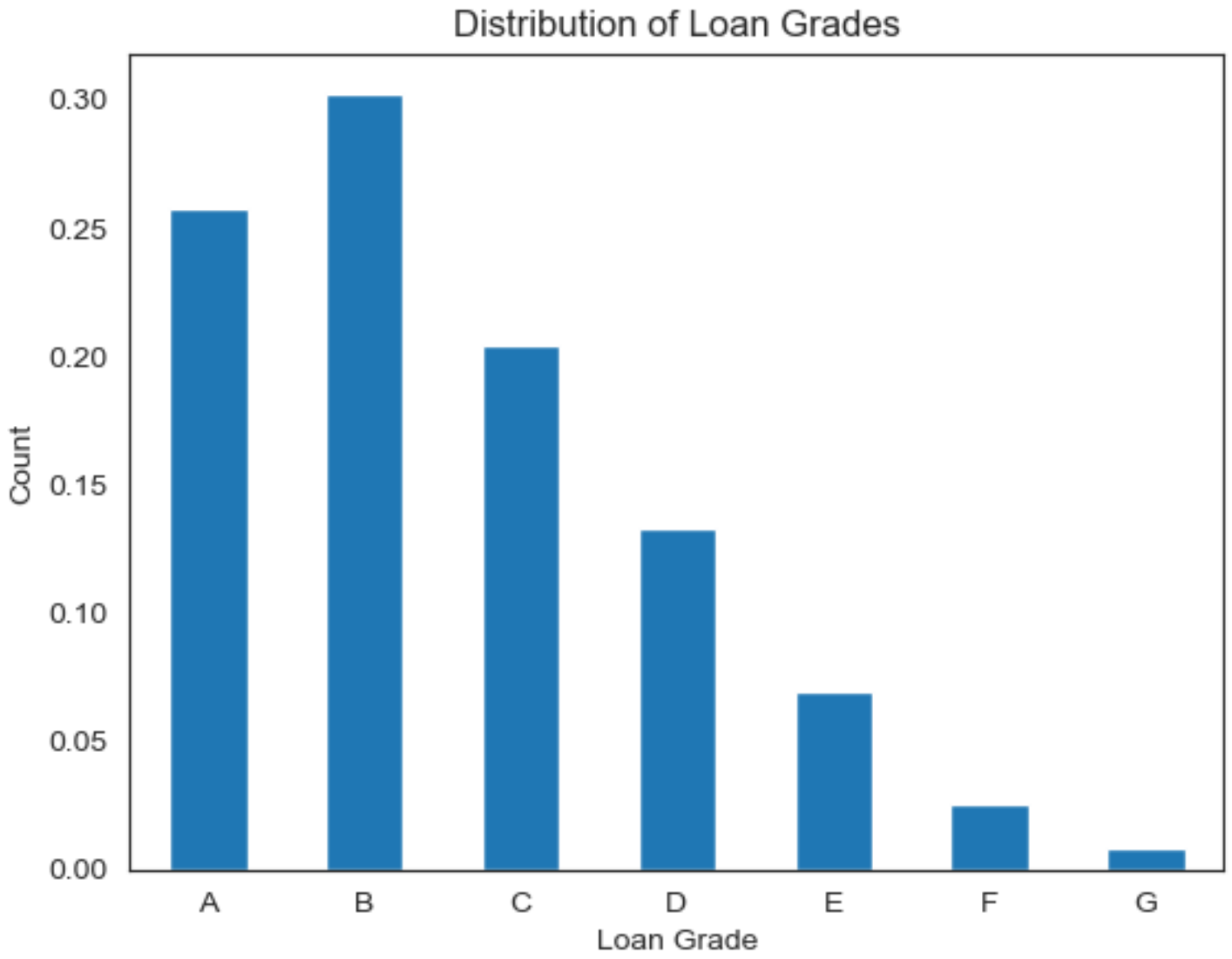


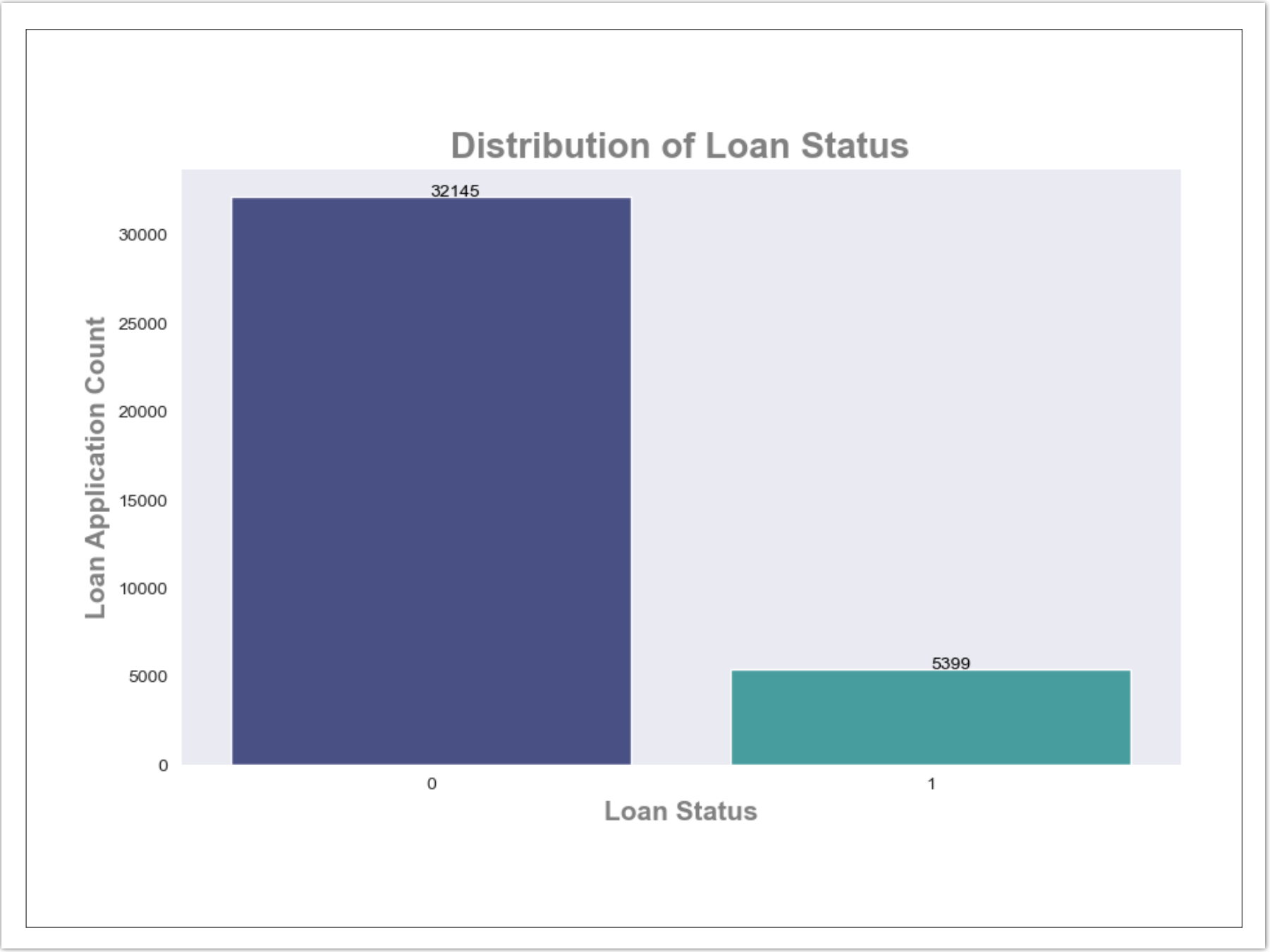
Loan Amount and Interest Rate

- Range of common loan sizes is between 5k to 15k
- The interest rate is normally distributed with range between 10%-15%

Loan Grade

- It's observed that most common loan grade is B and less common loan grade is G





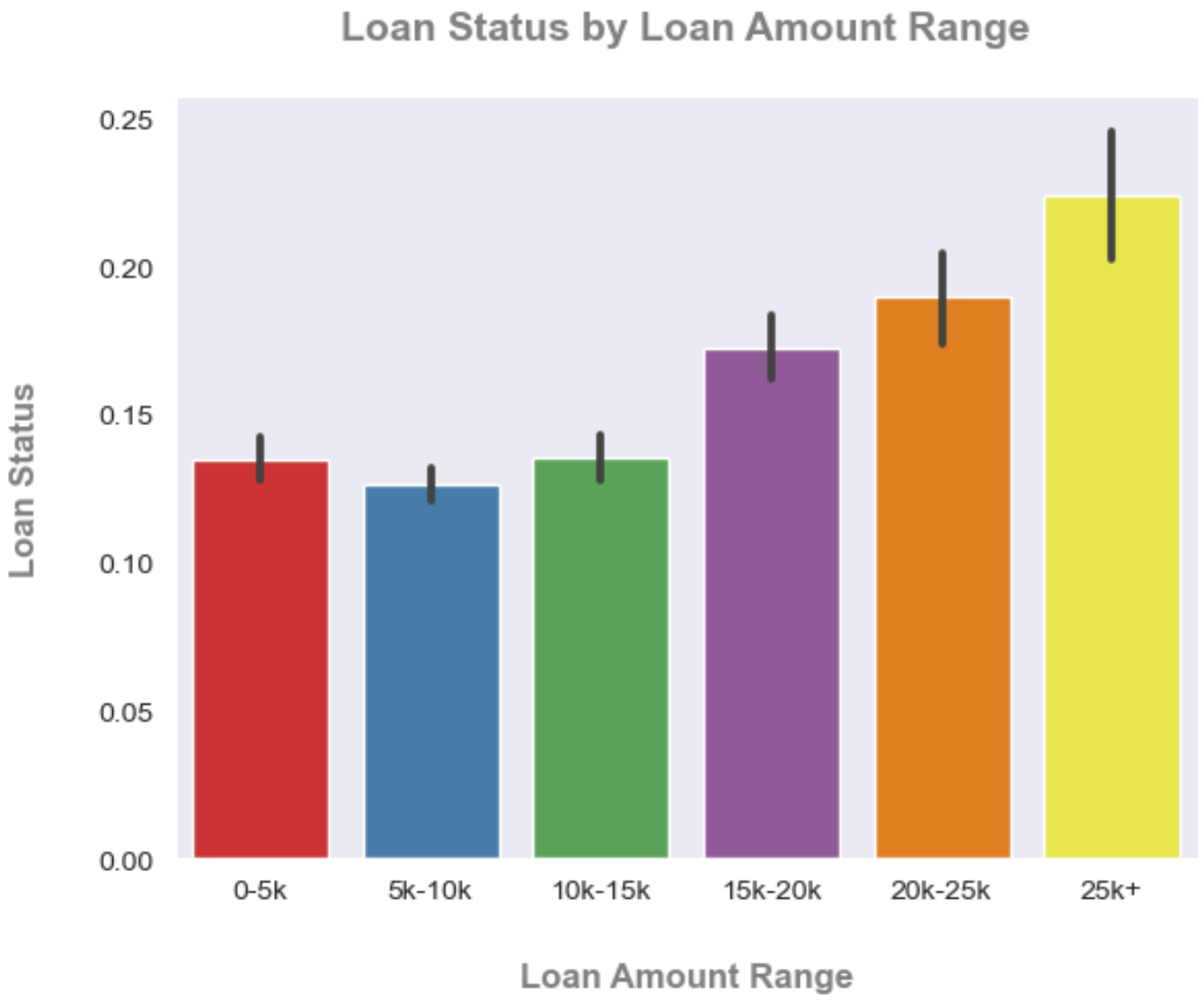
Distribution of Loan Status

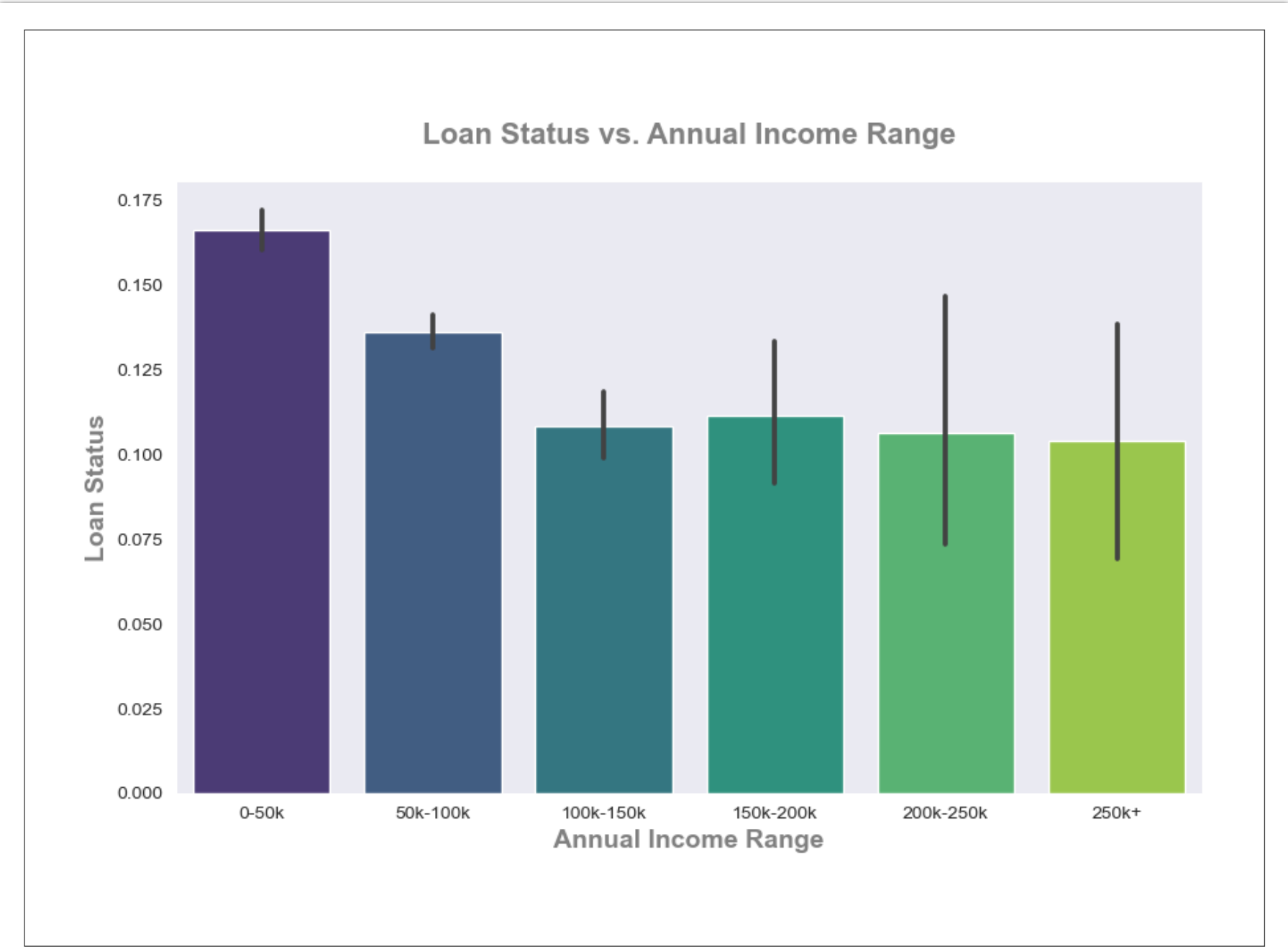
- 0 – Fully Paid
- 1 – Charged Off

SEGMENTED UNIVARIATE ANALYSIS

Loan Status by Loan Amount

- Default rate is higher for loans greater than 20k
- Higher the loan amount, higher the loan charged off rate
- Default rate is lower for loans less than 15k



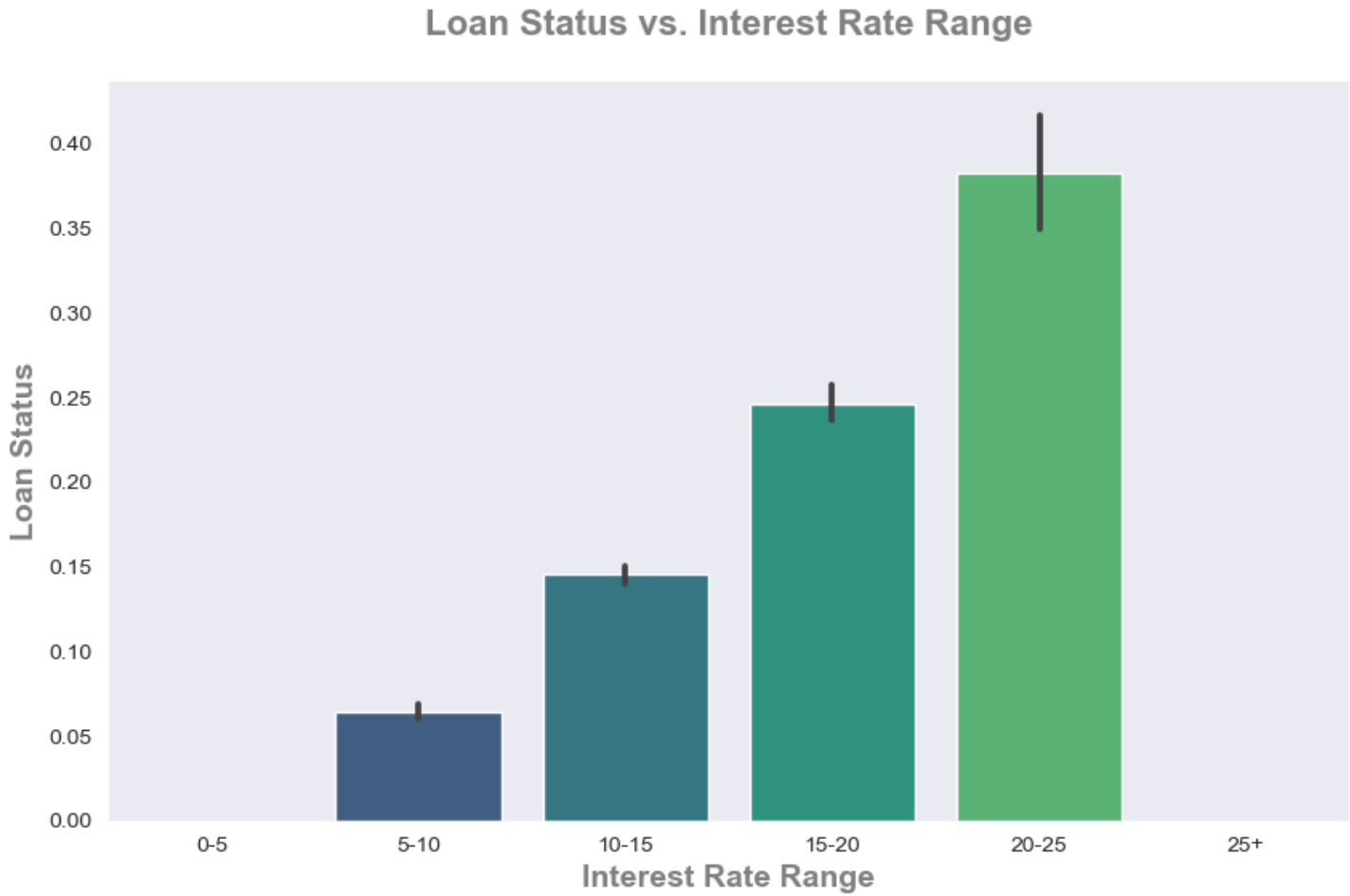


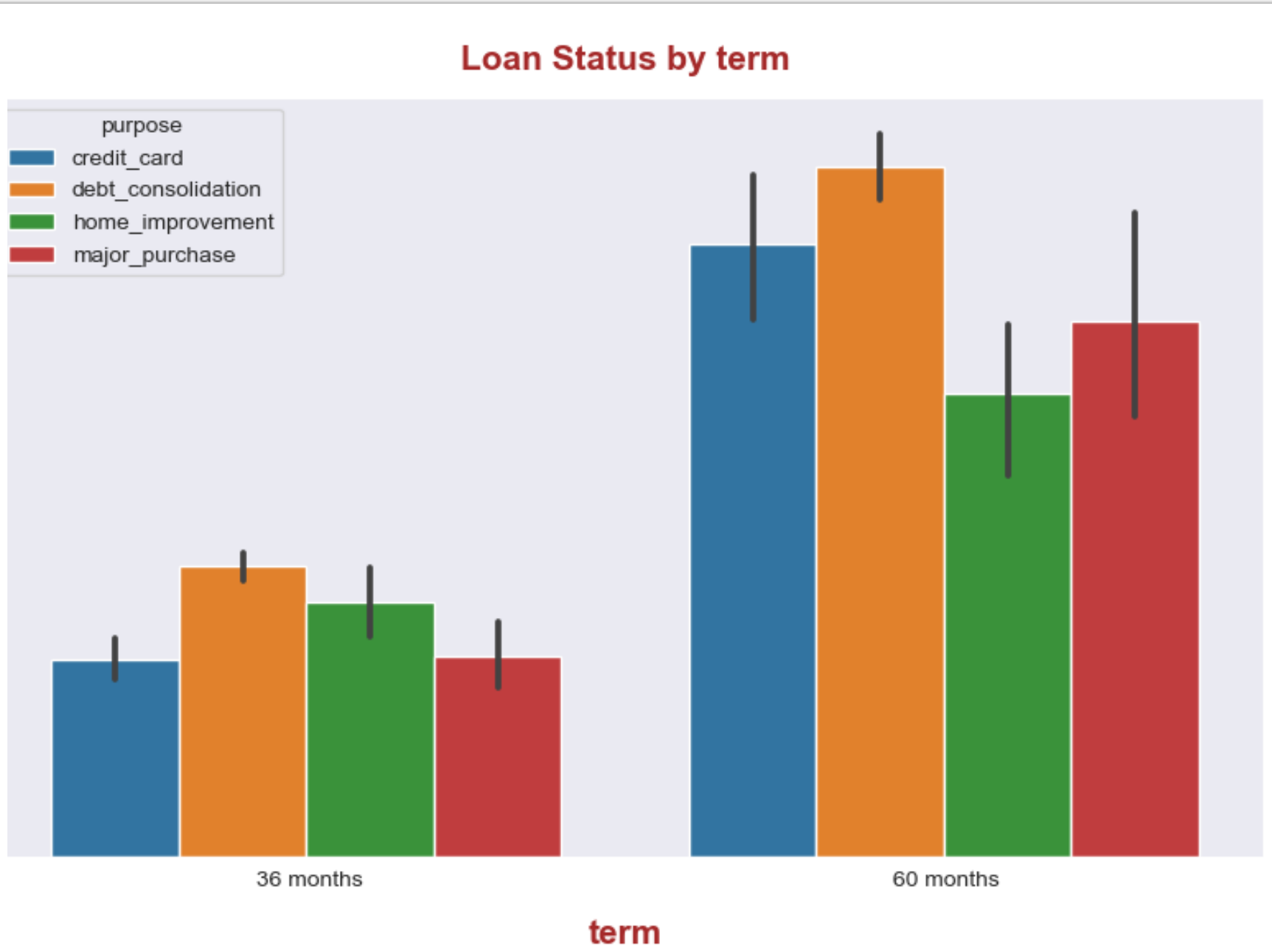
Loan Status vs Annual Income Range

- Lower income range has more probability of defaulting on loan
- Income range of 0-50k are the most probable loan defaulters
- 250k+ income range has the highest probability of paying off loan

Loan Status vs Interest Range

- The higher the interest rate, the higher the chance of defaulting on the loan
- Interest rate range falls between 20% and 25% have the highest default rate





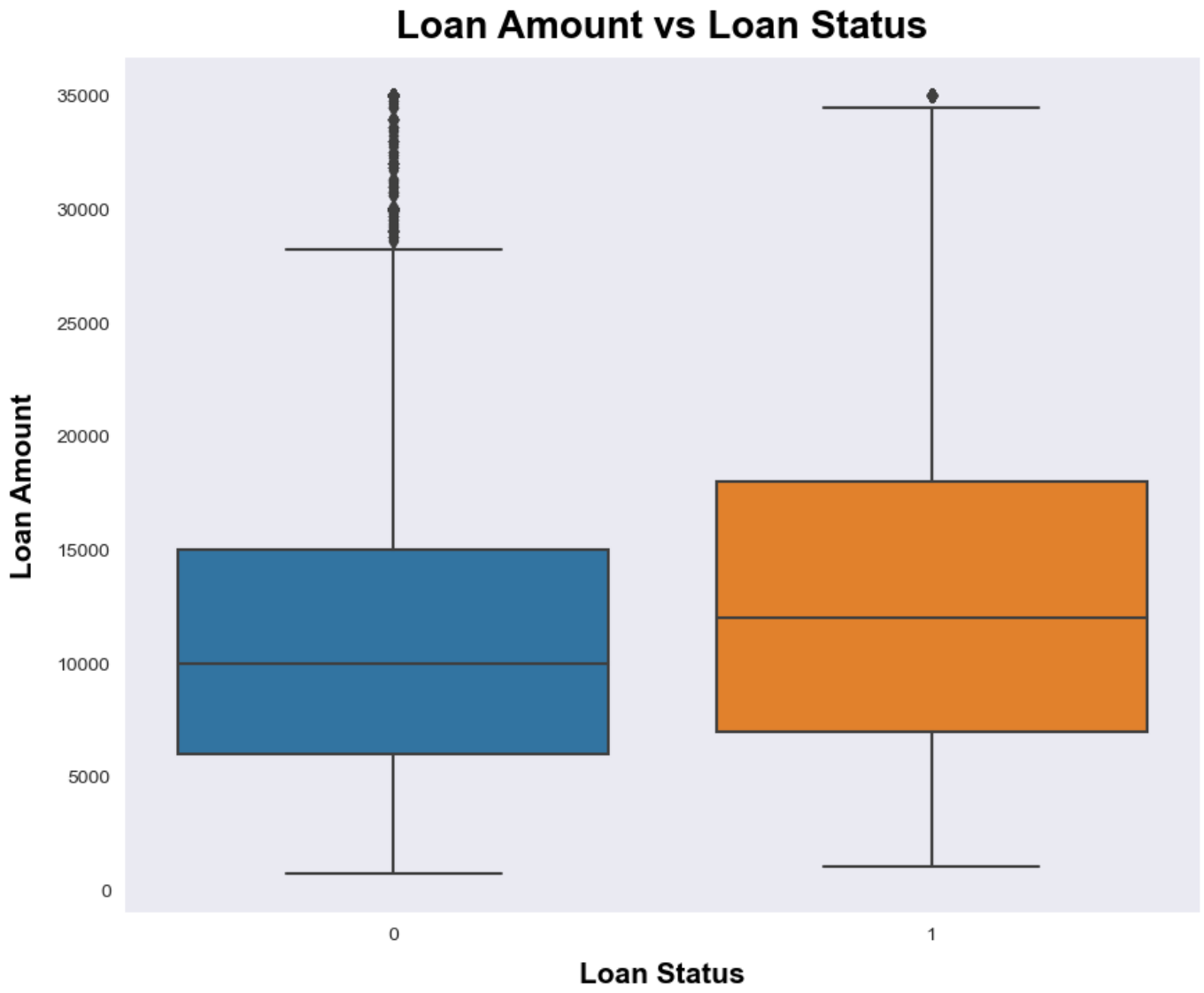
Loan Status vs Loan Term

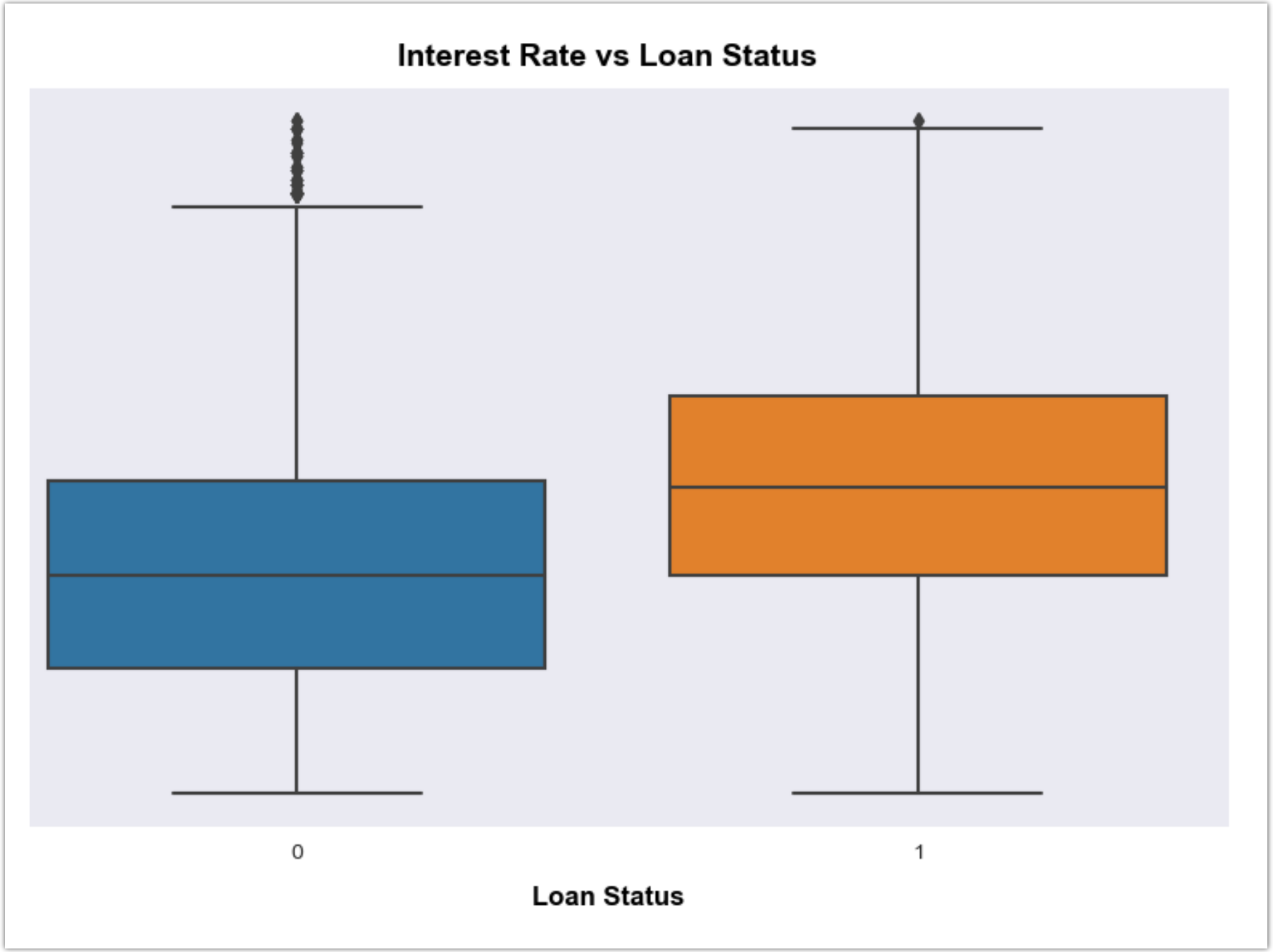
- Loan charged off rate is higher for 60 months term loans than 36 months term
- The loan_status is more likely to be charged off for debt_consolidation than other purposes

BI VARIATE ANALYSIS

Loan Amount vs Loan Status

- The range of loan amount for charged off loans is higher than the range of loan amount for fully paid loans





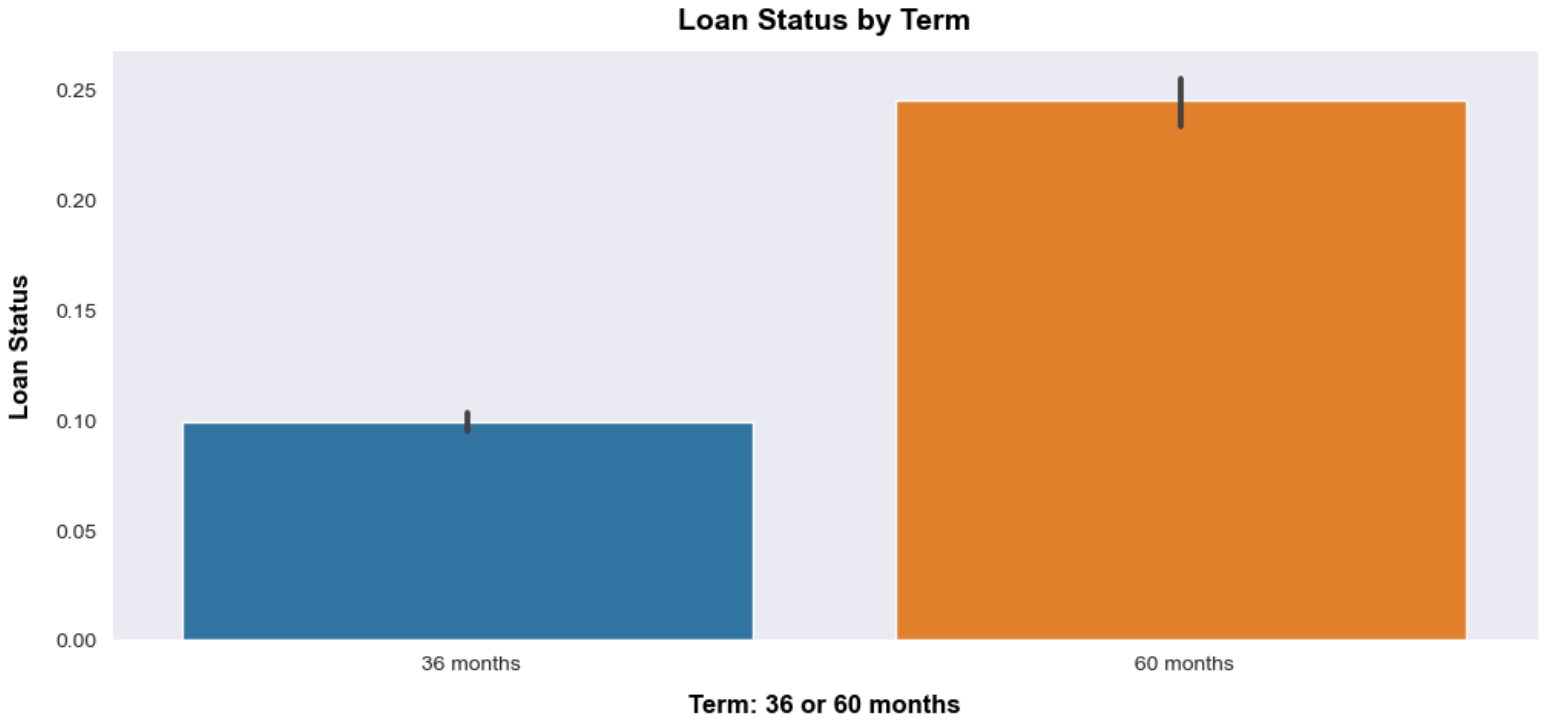
Interest Rate vs Loan Status

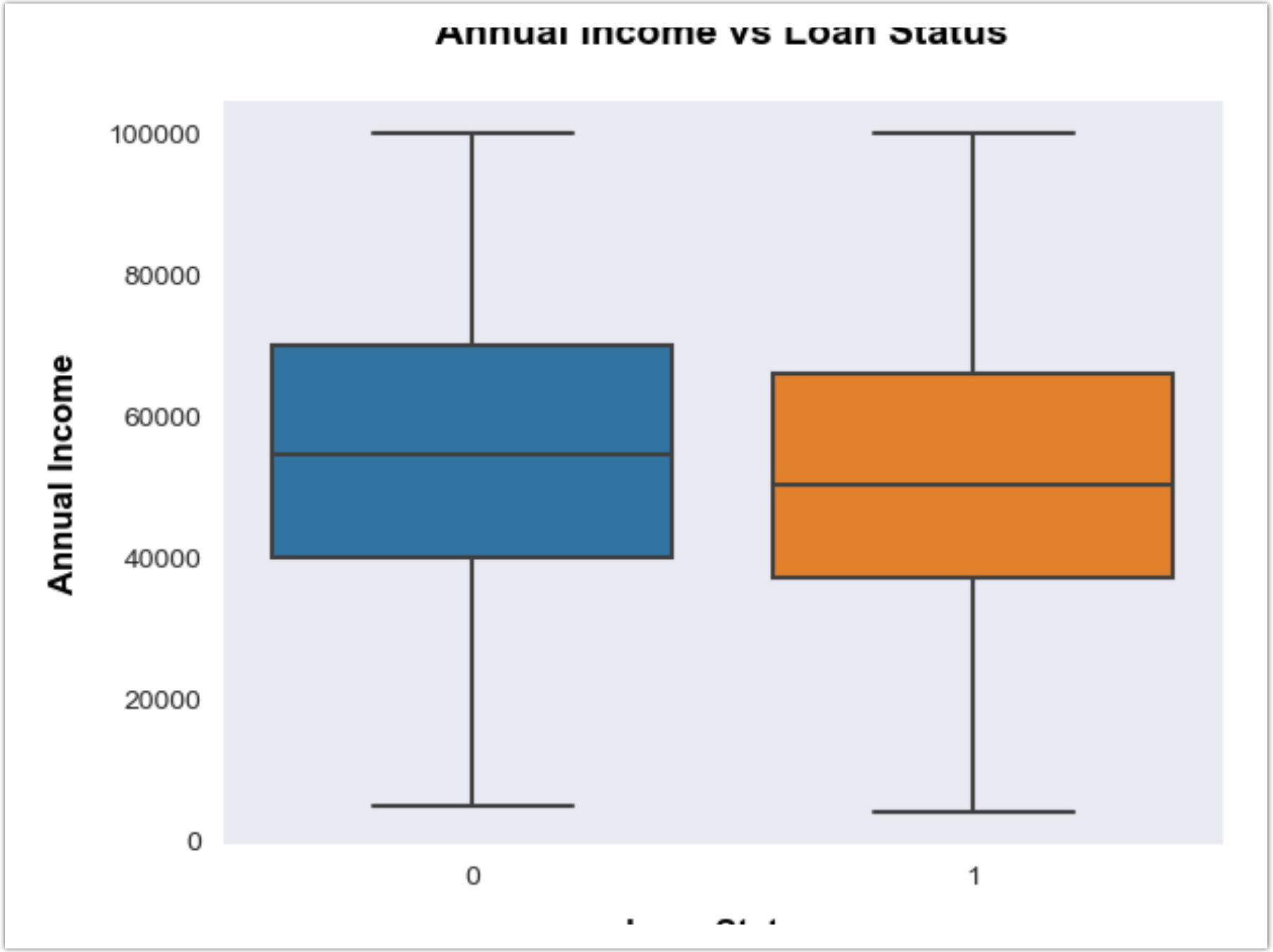
- 0 – Fully Paid
- 1 – Charged Off

- Most of the loans that are charged off have higher interest rates
- The median interest rate for charged off loans is higher than the median interest rate for fully paid loans

Loan Status by Term

- The longer the loan term, more likely the loan will be defaulted





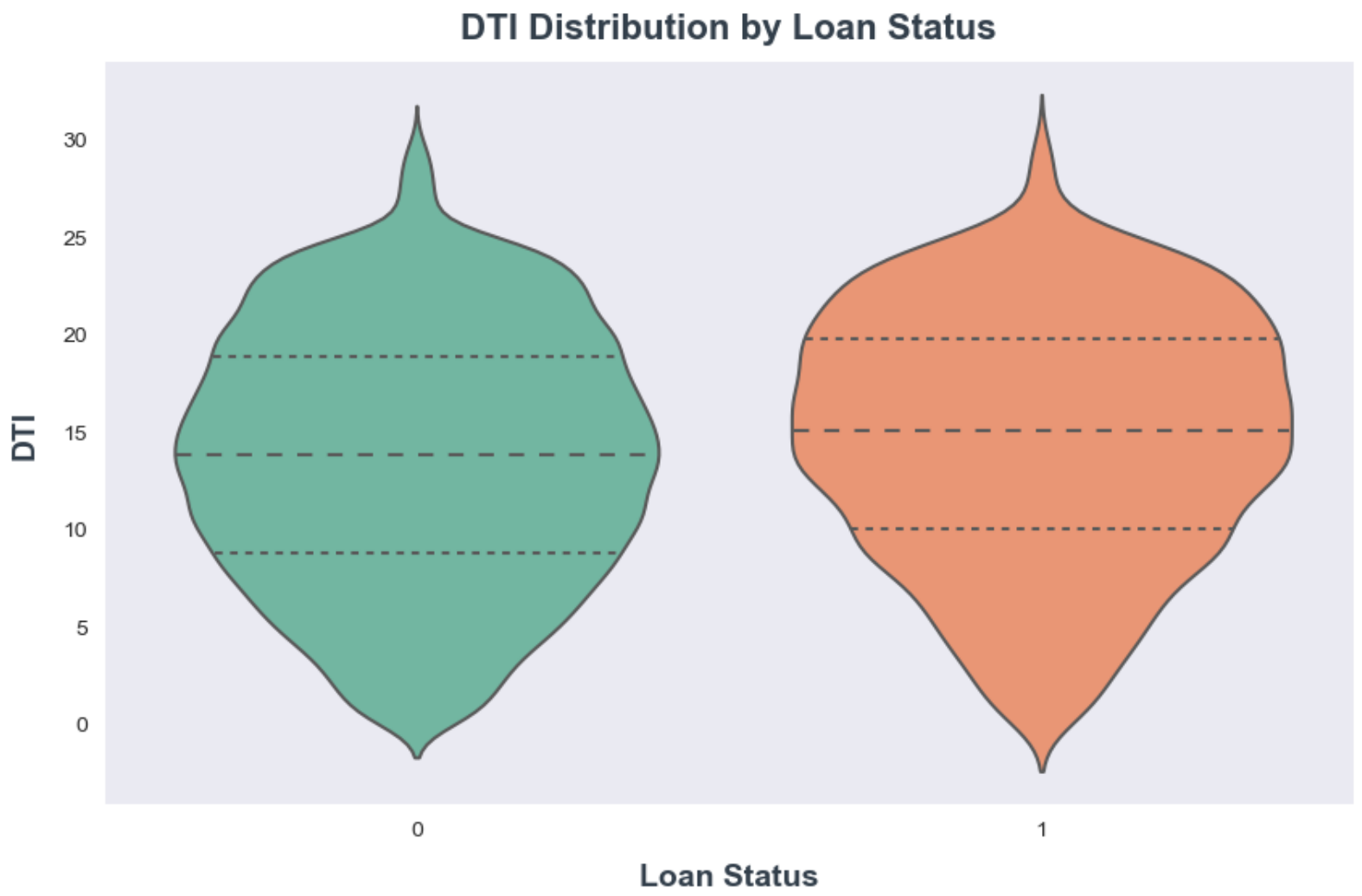
Annual Income vs Loan Status

- 0 – Fully Paid
- 1 – Charged Off

- The boxplot shows that the annual income of the charged off loans is lower than the fully paid loans

Verification Status with Loan Status

- Surprisingly, verified loans have a higher rate of default than unverified loans



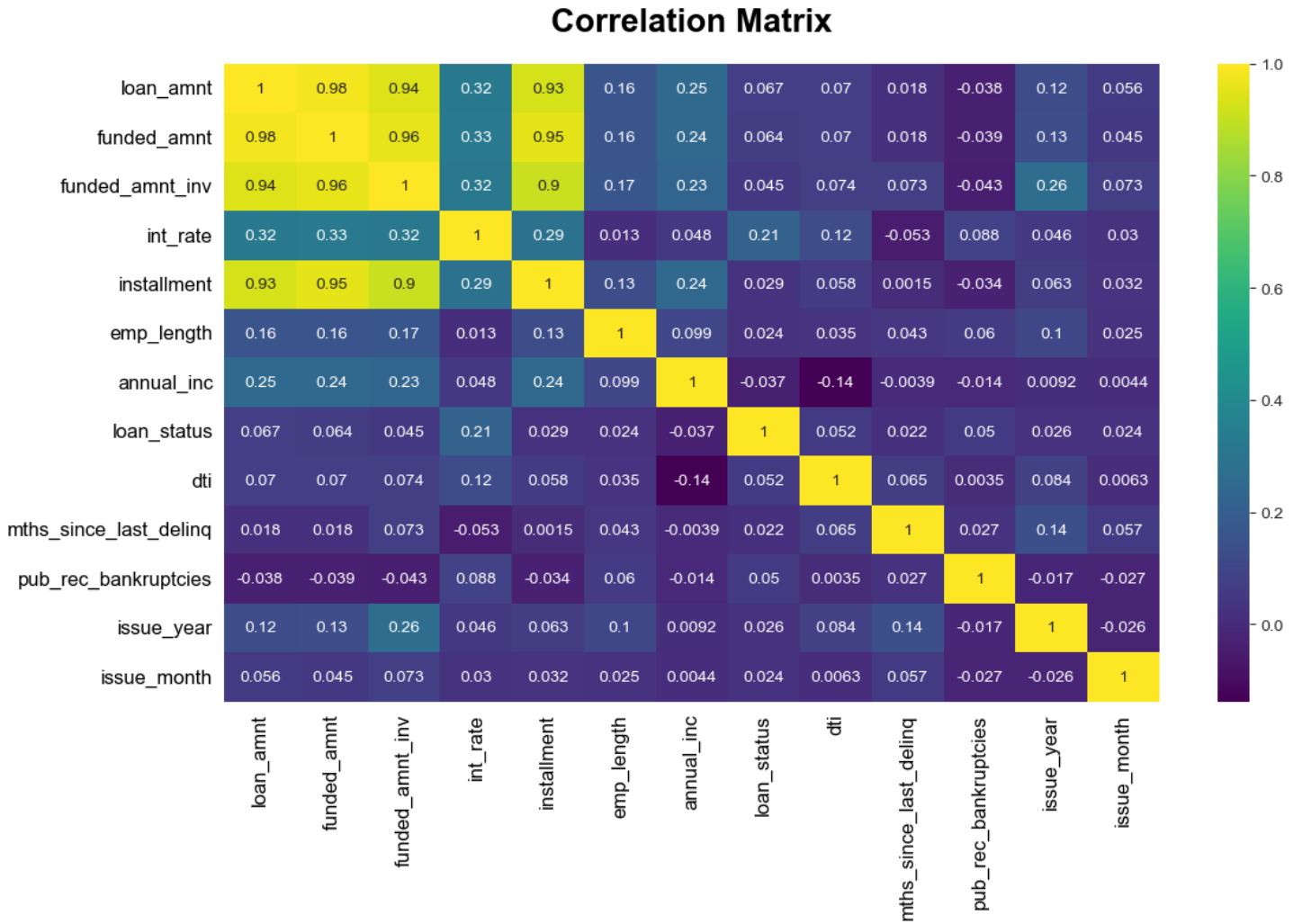
DTI Distribution by Loan Status

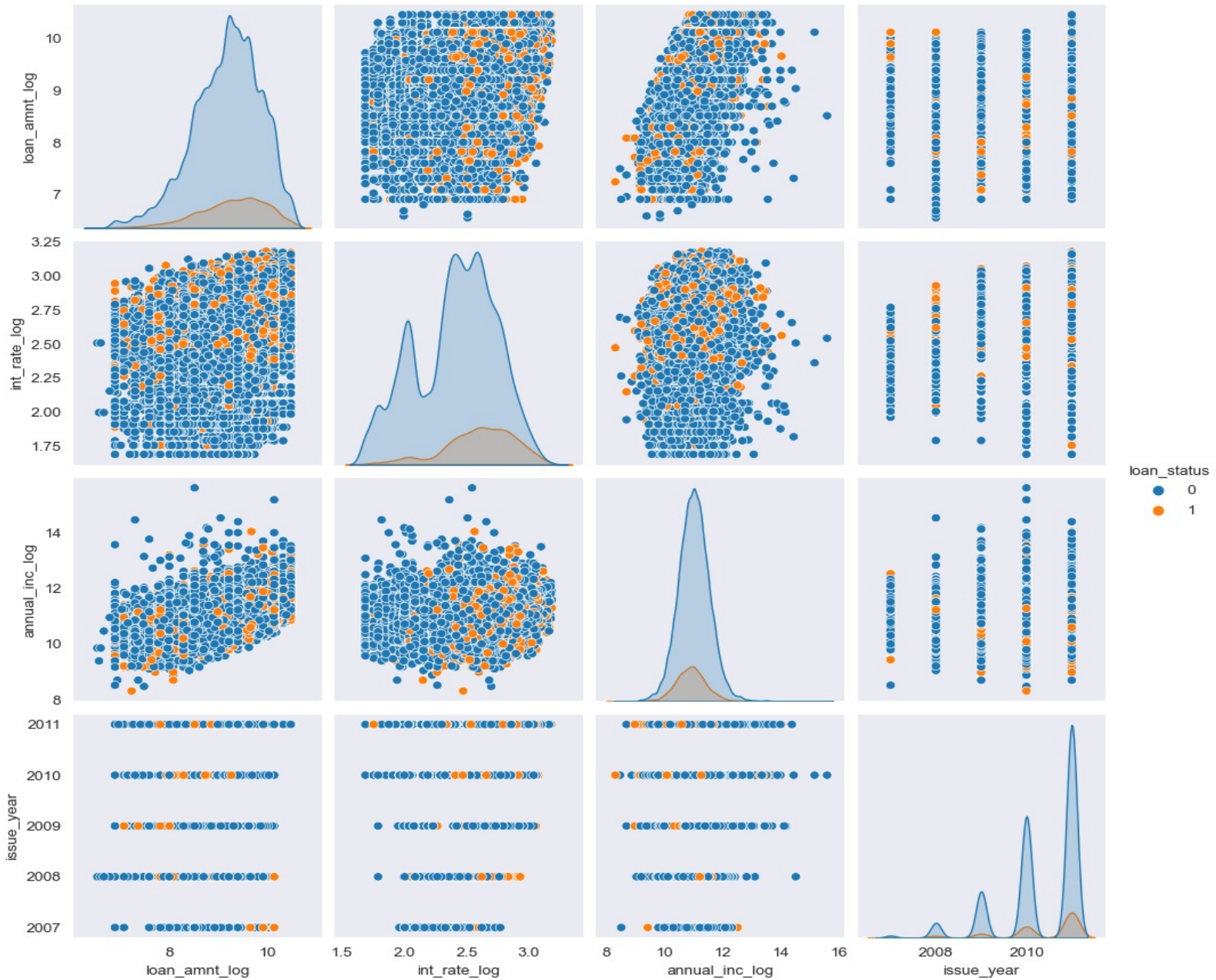
- Mean DTI is higher for charged off loans
- The distribution of DTI for charged off loans is wider than that of fully paid loans

MULTIVARIATE ANALYSIS

Correlation Matrix

- Loan status is highly correlated with int_rate
- Loan status is least correlated with annual income
- Loan amount and funded amount are highly correlated
- Loan amount and funded amount are highly correlated with installment





Pair Plot

- Loan amount is higher for charged off loans
- Interest rate is higher for charged off loans
- Annual income is higher for higher loan amount

CONCLUSION

Continuous Variables

- LOAN_AMOUNT : Default rate is higher for the higher loan amounts. Loan amount greater than 15000 dollars have higher default rate
- FUNDED_AMOUNT : Funded amount greater than 15000 dollars have higher default rate. The median funded amount for charged off loans is higher than that of fully paid loans
- FUNDED_AMOUNT_INVESTED : Funded amount invested greater than 15000 dollars have higher default rate
- INTEREST_RATE : Higher the interest rate, higher the chance of defaulting on the loan. Loans with int rate falling between 20% and 25% have the highest default rate
- ANNUAL_INCOME : Lower income range has more probability of defaulting on loan. Income range of 0-50k are the most probable loan defaulters
- DTI : As DTI increase the default rate increases. Mean value of DTI is greater for charged off loans

Categorical Variables

- TERM : Loan charged off rate is higher for 60 months term loans than 36 months term
- GRADE : As the grade of the loan decreases [A,B,C,D,E,F,G], chances of loan getting defaulted increases
- SUB_GRADE : As the Sub Grade decreases (A1 A2 B1 B2.....) default rate increases. Loans with sub grade F5 is having highest default rate around 48%
- VERIFICATION STATUS : Percent of loan defaulted is higher for verified borrowers
- PURPOSE : Small business loans have the highest default rate

Other observations

- Loan status is highly correlated with int_rate
- Loan status is moderately correlated with loan amount, funded amount, installment
- Loan status is least correlated with annual income
- Loan amount and funded amount are highly correlated
- Loan amount and funded amount are highly correlated with installment

THANK YOU

maheshpaneru22@gmail.com

krissravi36@gmail.com