

Lending Club Case Study

STUDY GROUP:

RAMAGOVINDARAJA THIRUVADI

SHASHWATH S

Problem Statement

Our task is to analyse a dataset of historical loan information from “*consumer finance company*” to find insights into features that may influence loan default rates. Our goal was to identify feature patterns and correlations that could help in assessing loan risk and making informed lending decisions.

Analysis Approach

Started with data cleaning to ensure the quality and integrity of the dataset for analysis. We removed columns with 100% null values and corrected data types for key columns. We then conducted a univariate analysis to understand the distribution of individual variables, followed by bivariate analyses to explore relationships between two variables, particularly in relation to loan default.

Univariate Analysis

Univariate Analysis has been done on following Variables.

Loan Amounts: Showed a wide range of requested amounts, indicating diverse borrowing needs. Most loans disbursed are at lower end of loan amounts requested (5000 to 10000).

Interest Rates: Shows a broad range of rates, suggesting varying rates based on qualified applicants. Most of the Interest Rates are Ranging from (7.5 to 15.0)

Employment Length: Indicated that a significant number of applicants had 10+ years of employment. i.e., with stable employment

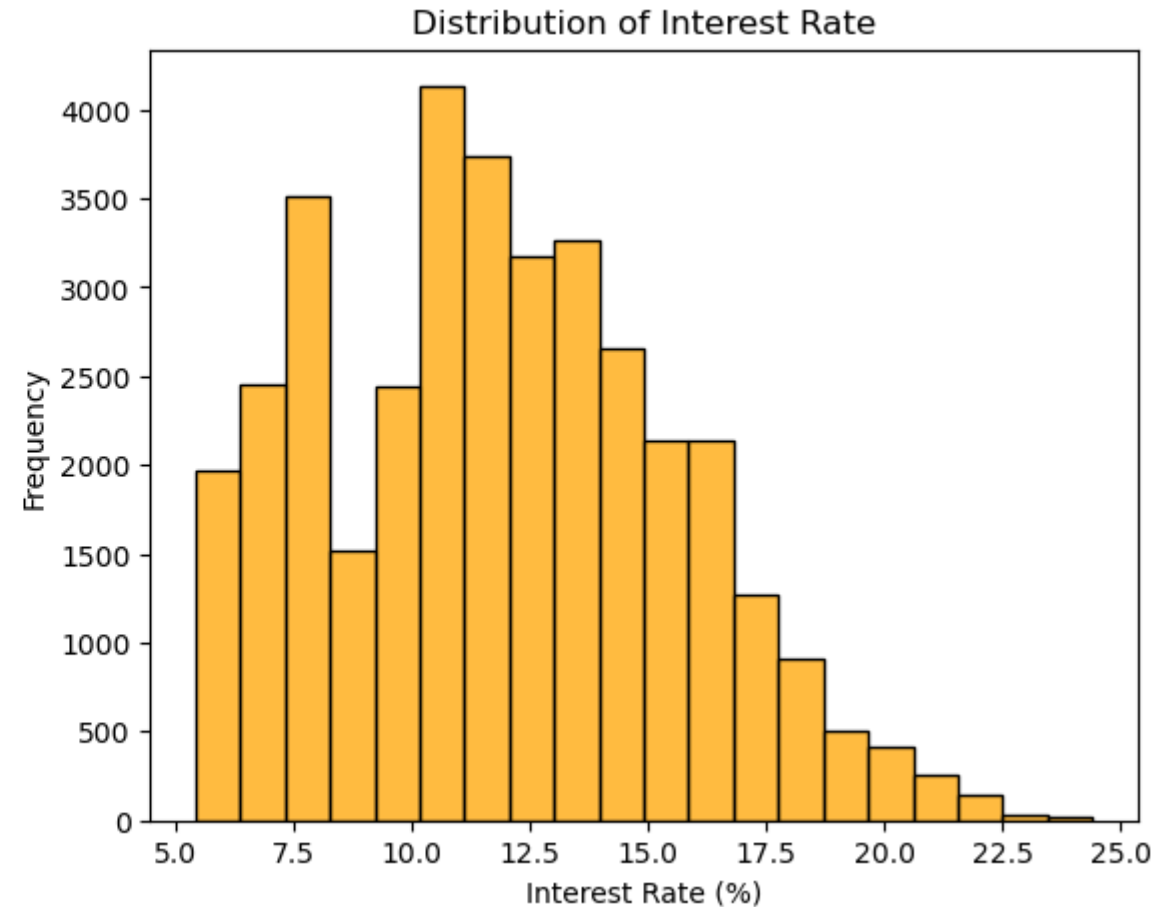
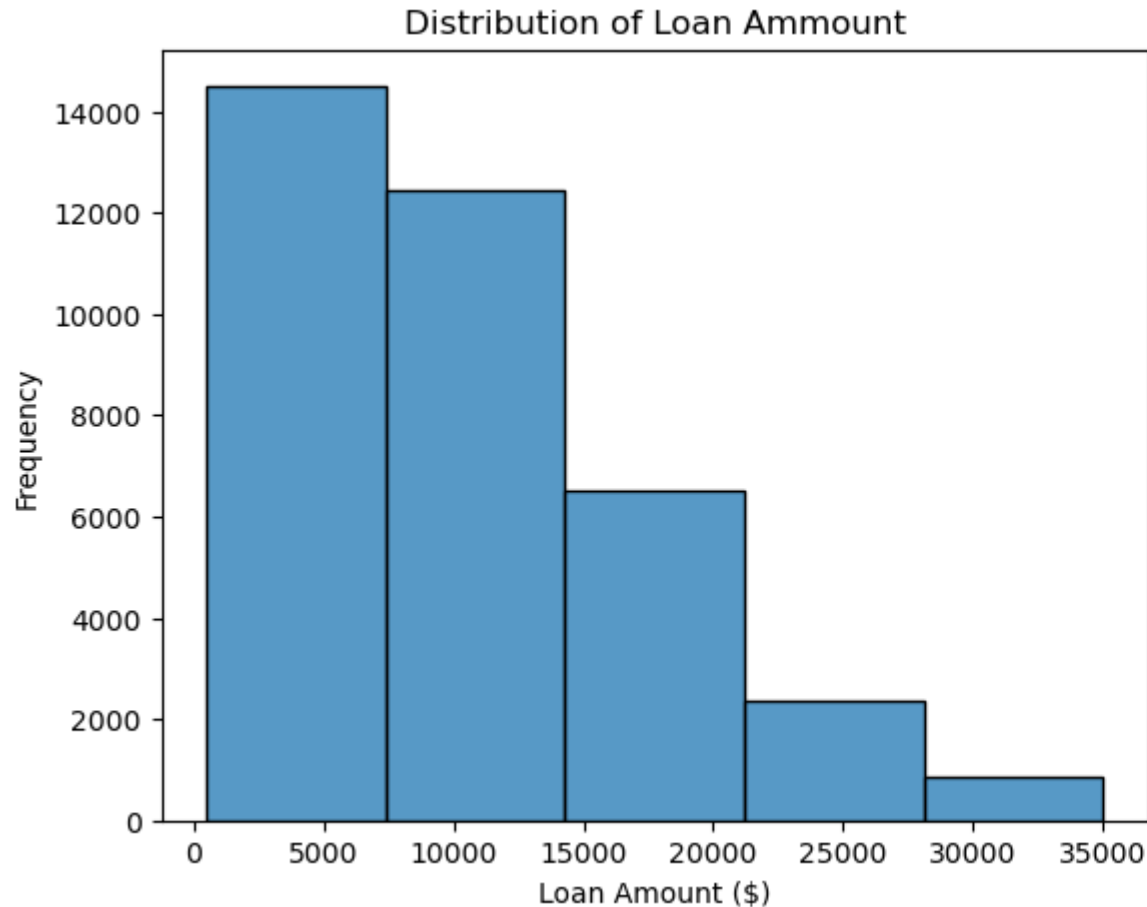
Debt-to-Income Ratio: Shows a right-skewed distribution, with most applicants having a DTI below 30%, but some displayed high debts.

Annual Income: Most of the loan applicant have a stable income ranging from 40000 to 80000.

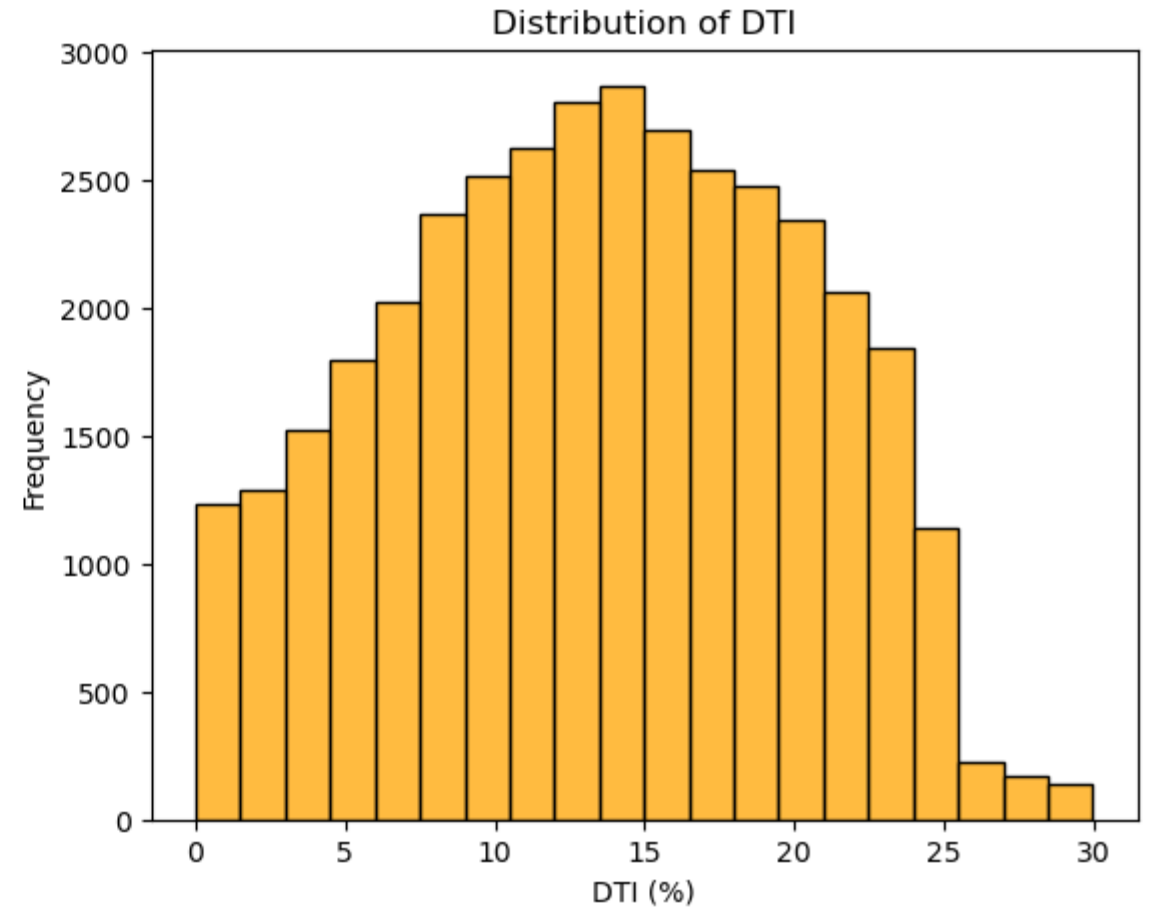
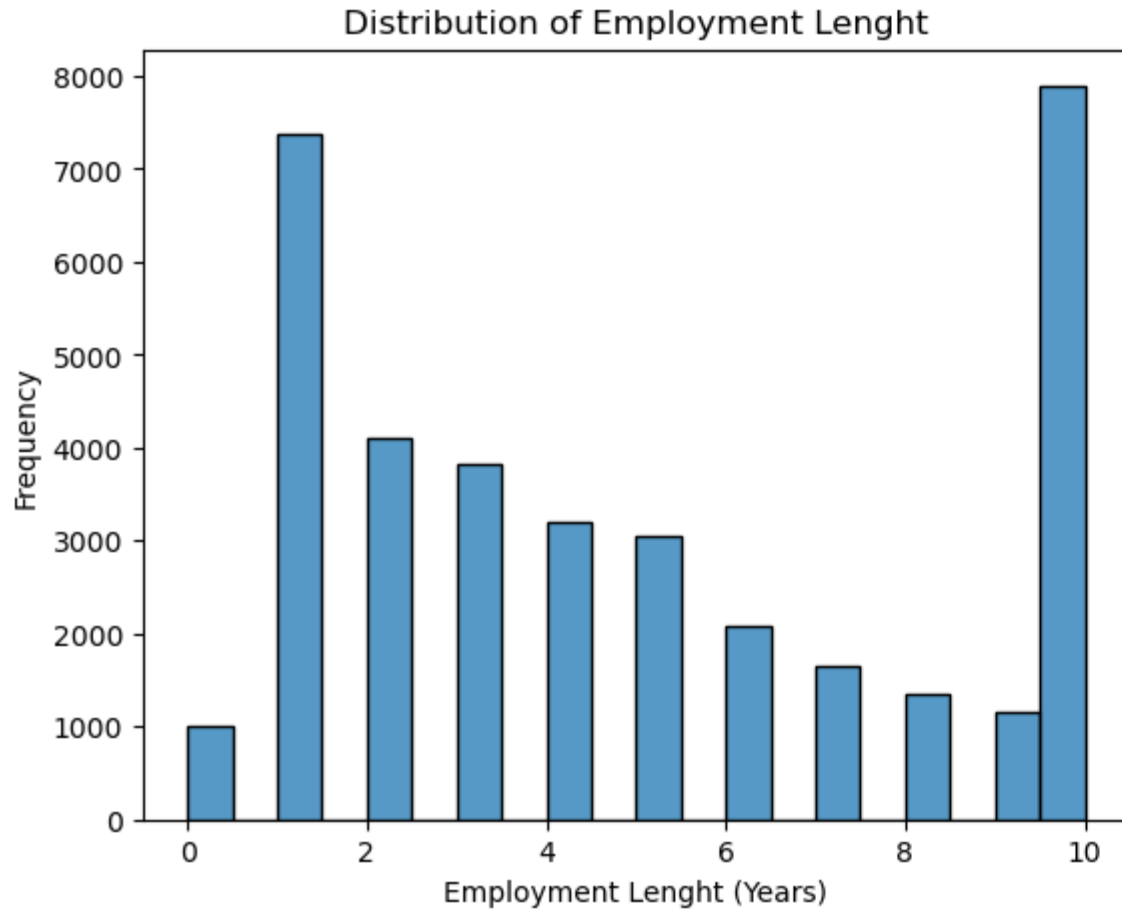
House Owner Ship: Most of the Loan applicants are living under Rented or Mortgaged Property.

Purpose: Most of the Loans are taken for Debt Consolidation Purpose.

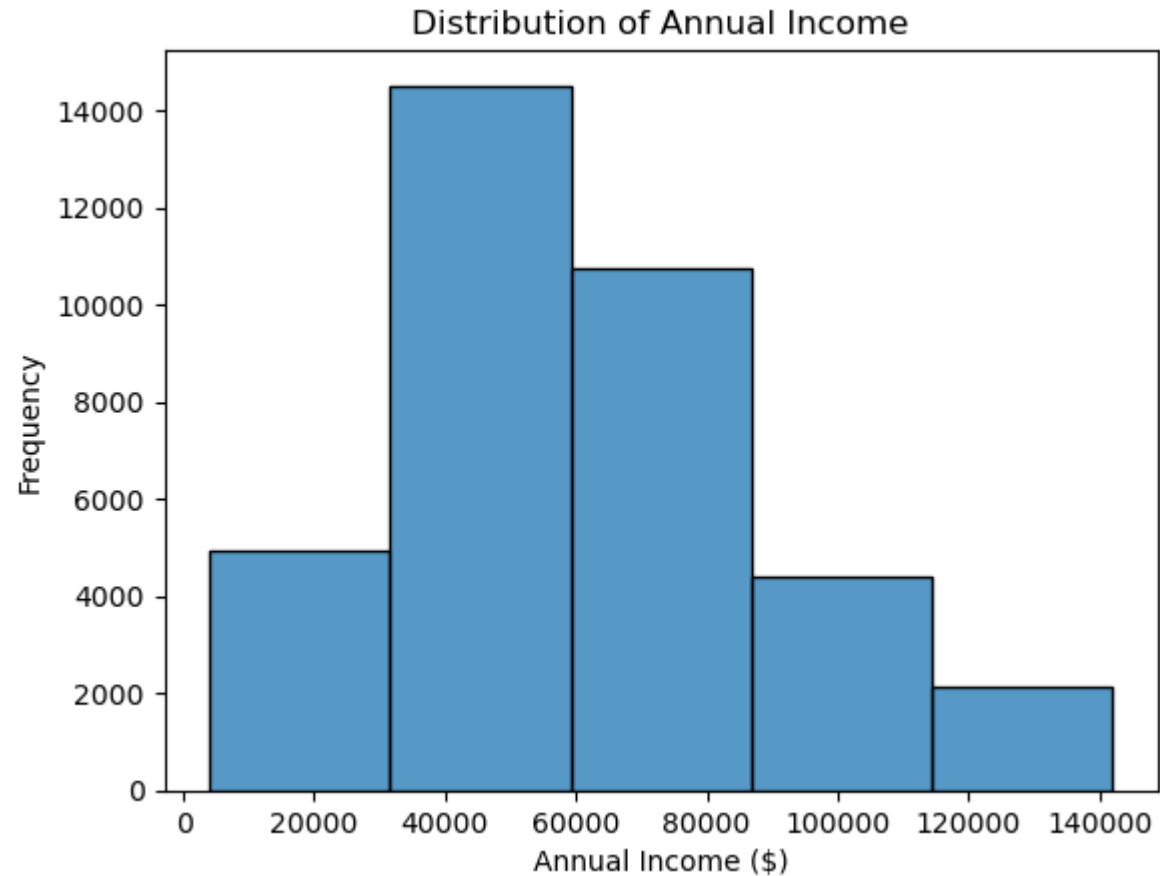
Distribution of Loan Amounts & Interest Rates



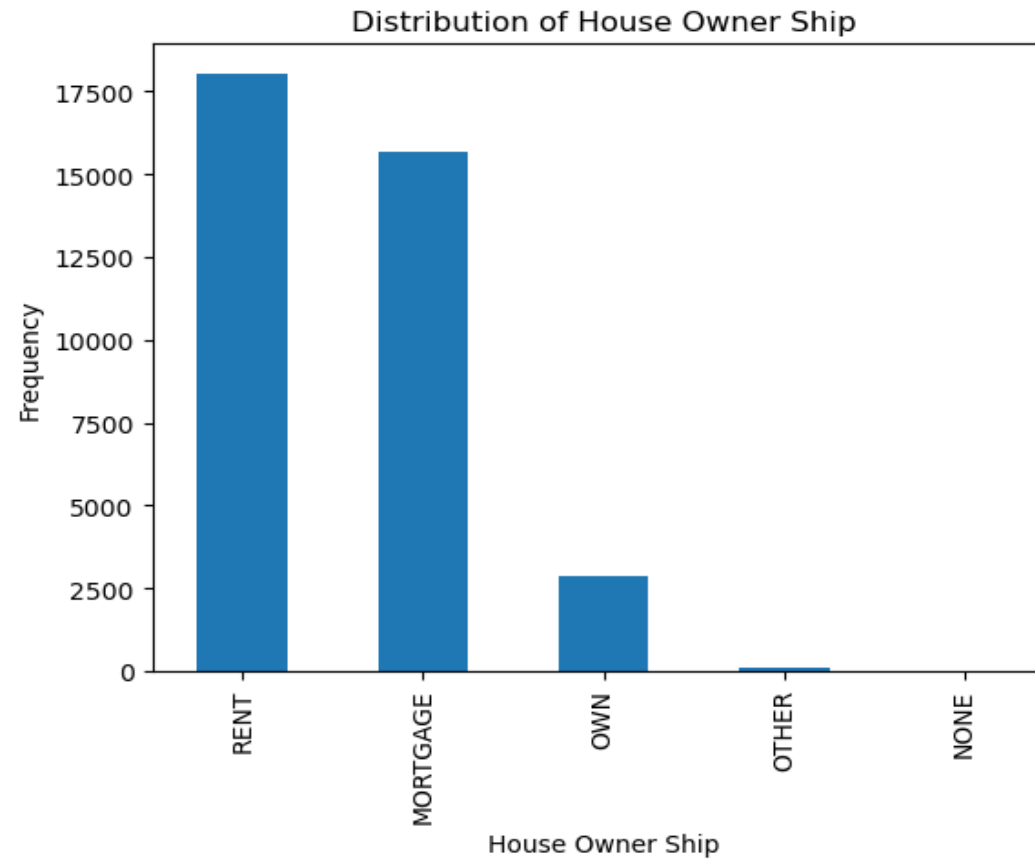
Distribution Employment Length & DTI



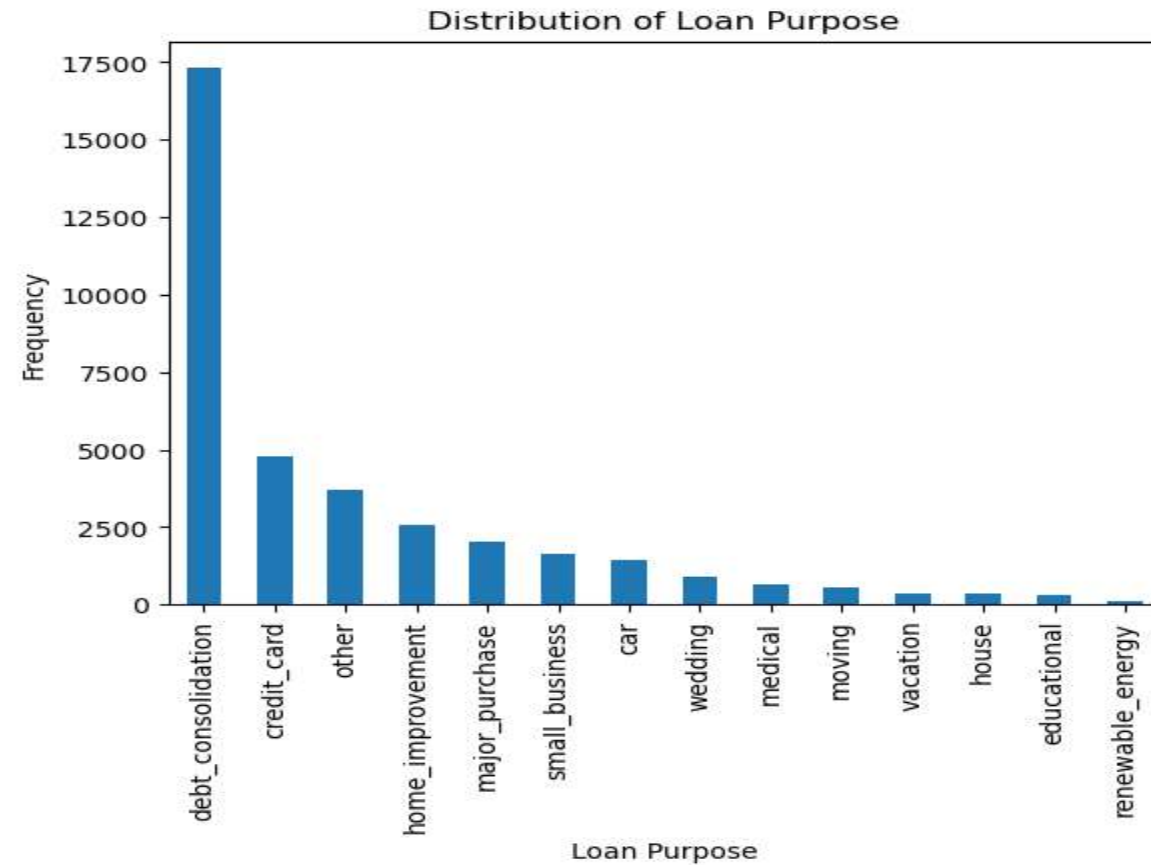
Dist. Of Annual Income



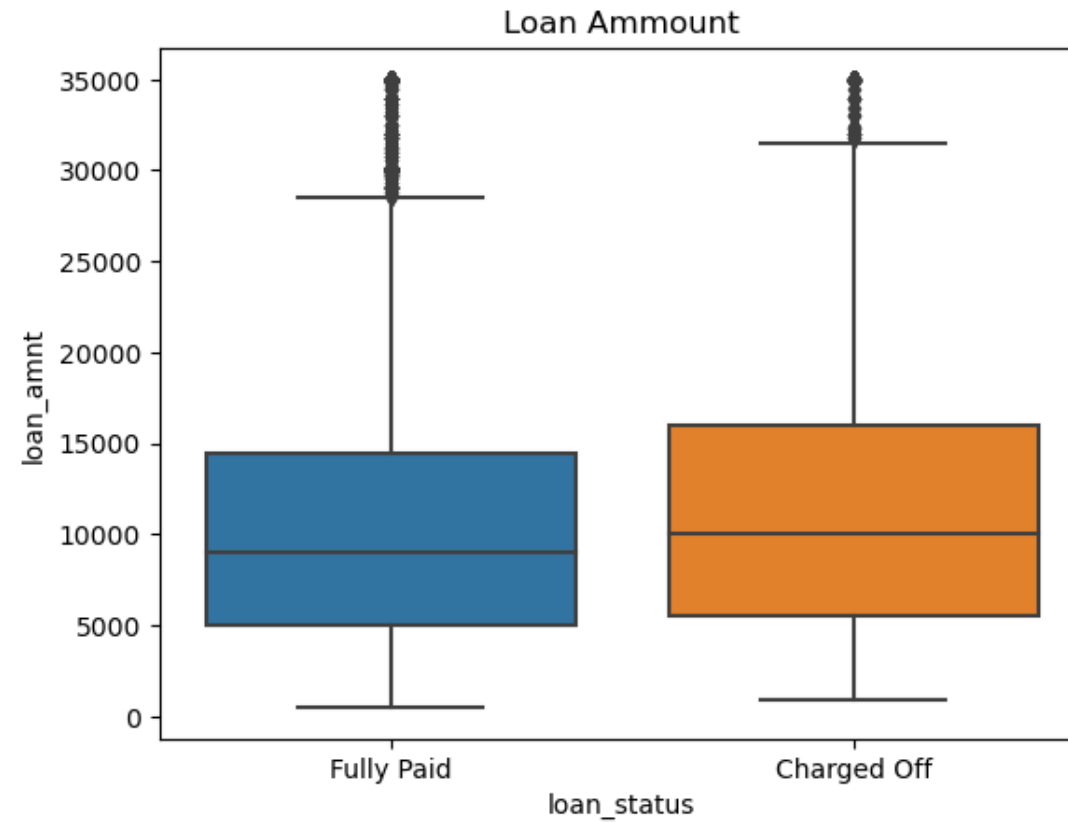
Dist. Of House Owner Ship



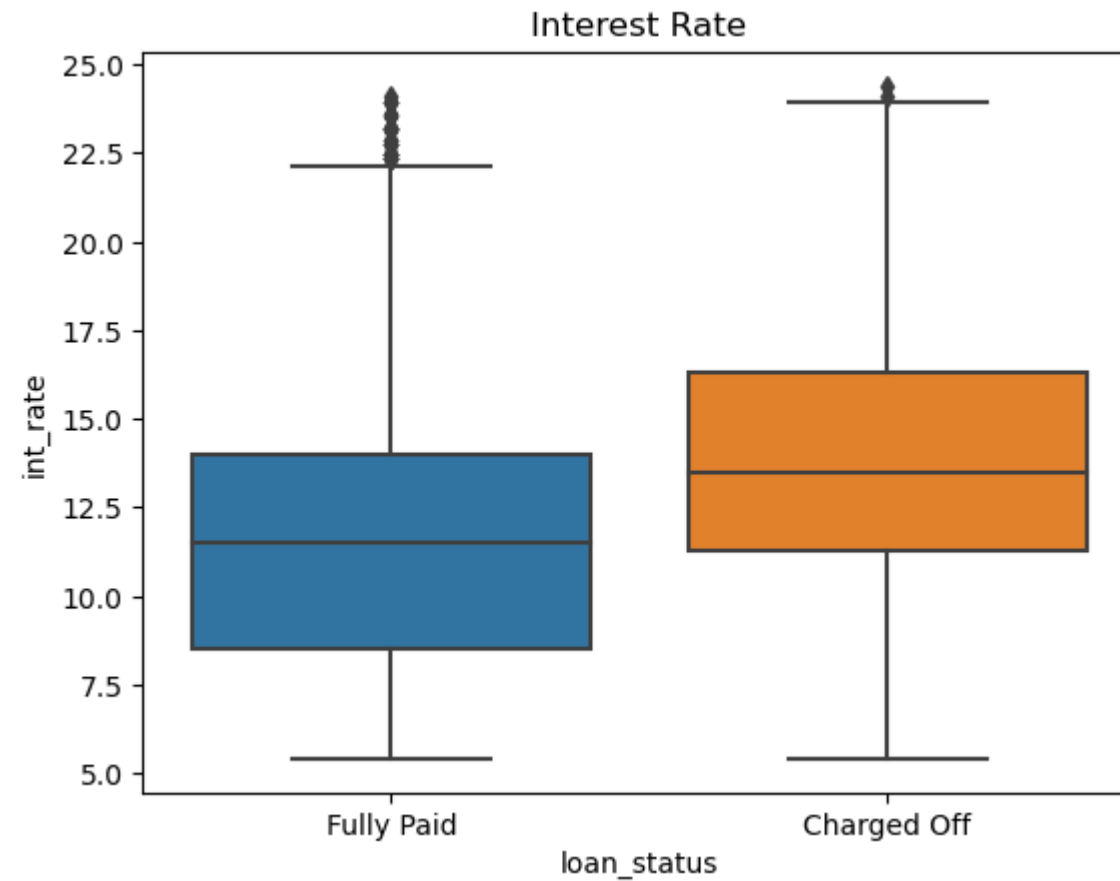
Dist. Of Loan Purpose



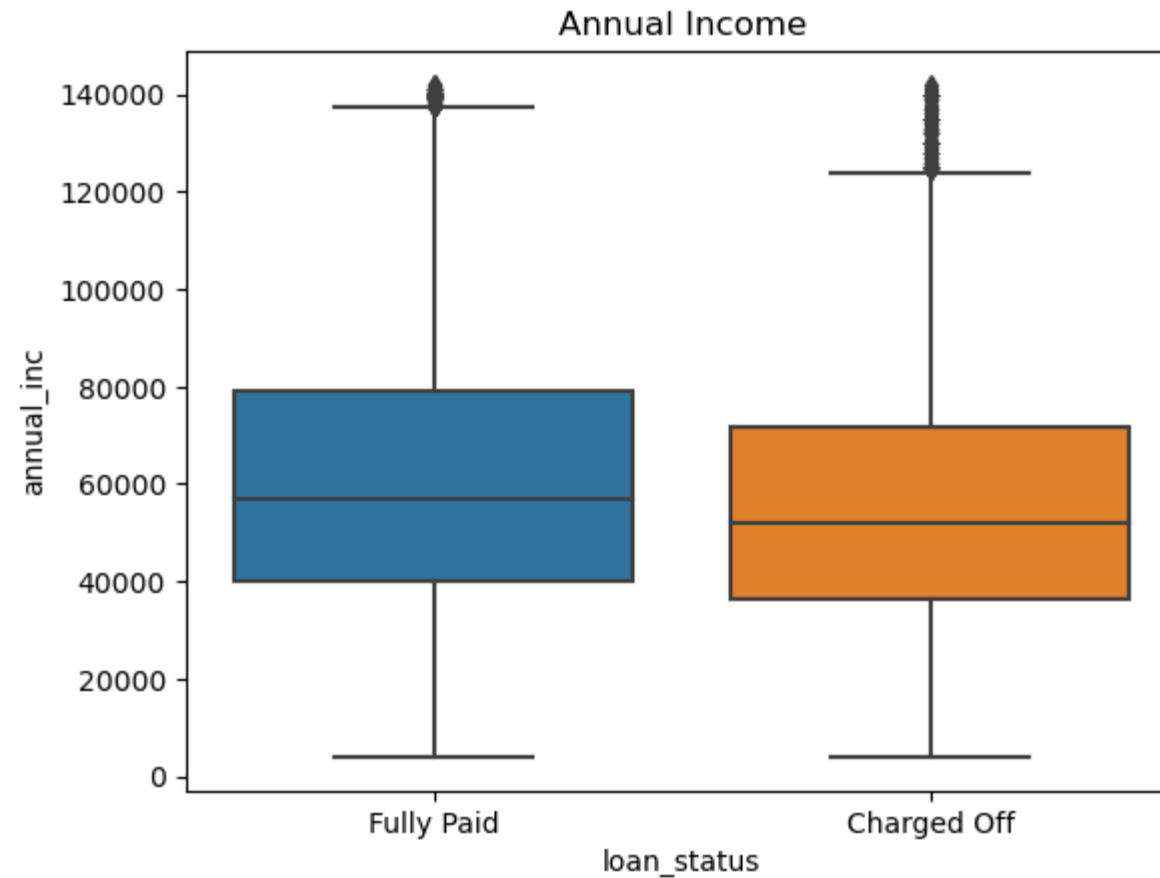
Segmented Univariate Analysis on Loan Amount



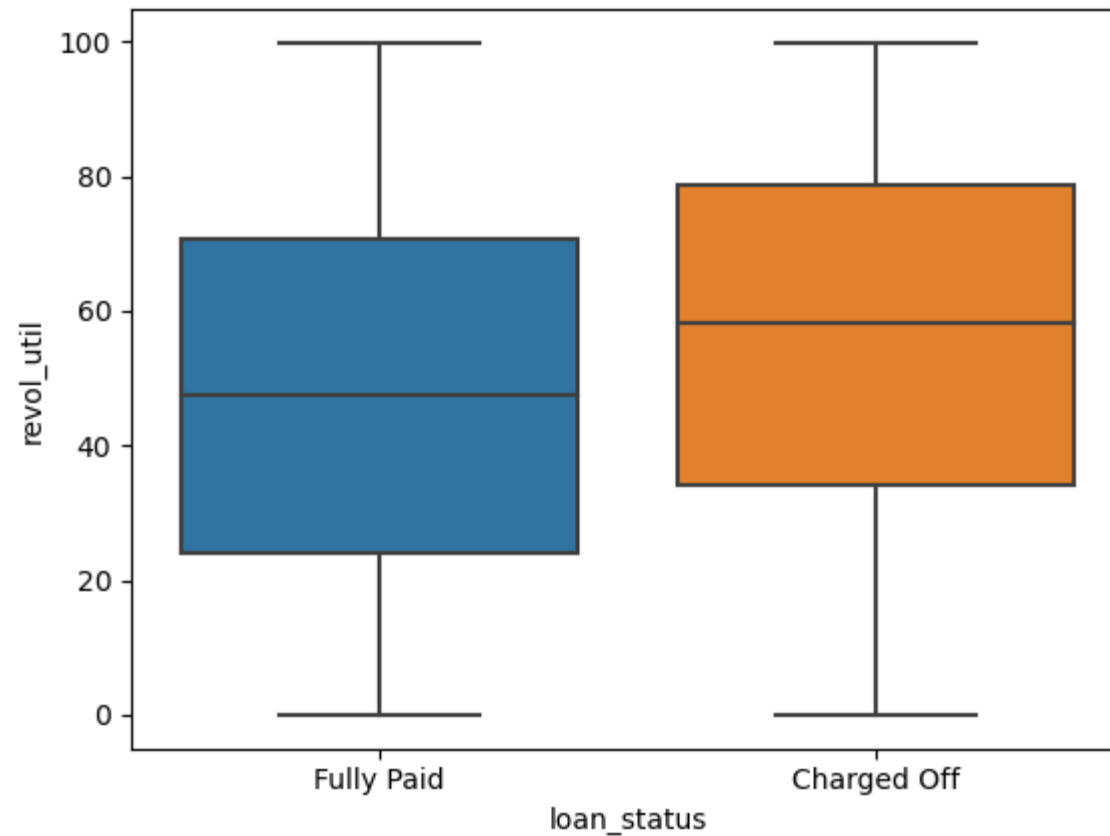
Segmented Univariate Analysis on Interest Rate



Segmented Univariate Analysis on Annual Income



Segmented Univariate Analysis on Revolve Utilization



Observations from Univariate Analysis

Loan Amount: Majority of Fully paid loans falls under 5000 to 15000 & majority of defaulted loans falls under 5000 to 17000. By looking at the median of these values ,risk of loan to be defaulted is higher if loan amount > 10000.

Interest Rates: Majority of interest for fully paid loan is around 8% to 14% & majority of interest rate for defaulted loan is around 11% to 16%. By looking at median of these values, the risk of loan to be charged off will be higher if Interest Rate is > 13%.

Annual Income: Majority of loan applicant who fully paid loan have income of 40000 to 80000 & Majority of the loan applicant who defaulted their loan have income of 35000 to 70000, by looking at median of these values the risk of charged off loan is more if loan applicant income is < 50000.

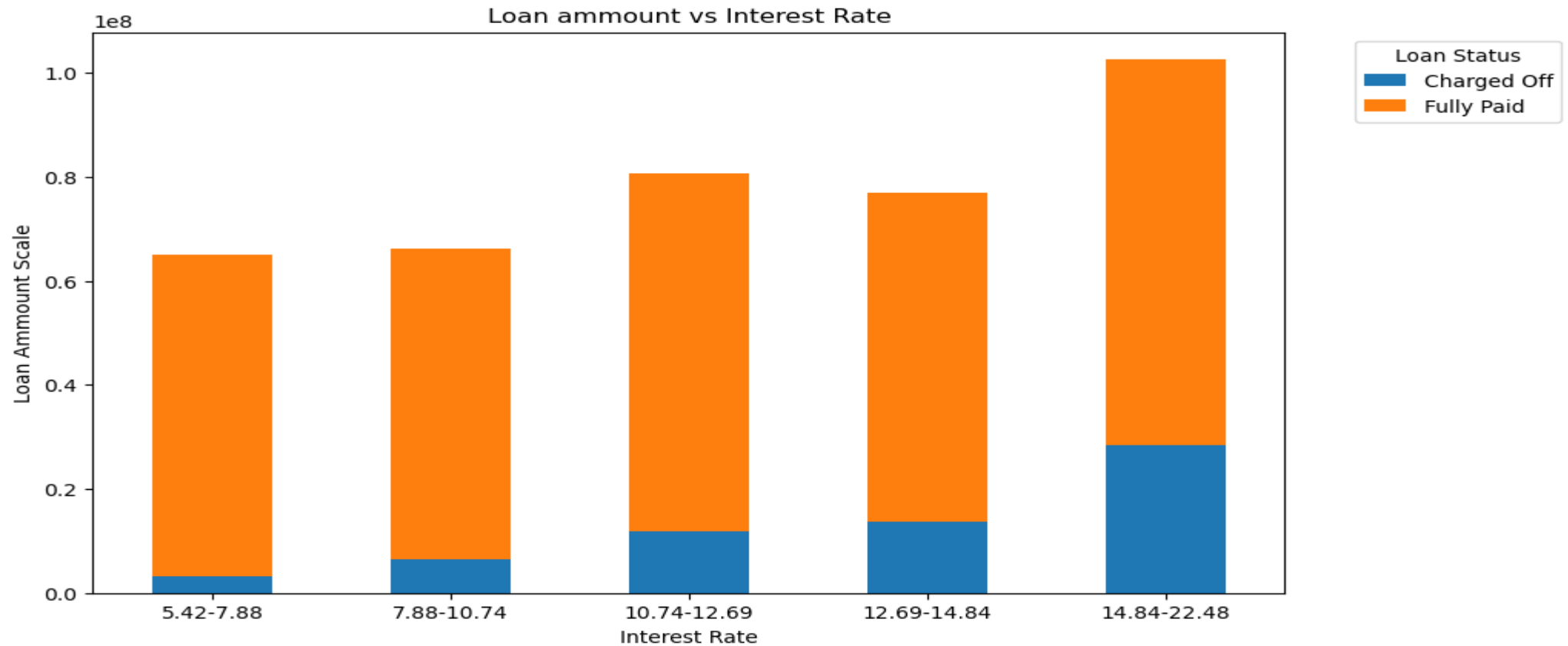
Revolve Utilization: Majority of fully paid loan has a revolve utilization line of 25% to 70% & majority of defaulted loan have revolve utilization of 30% to 75%. looking at the median the risk of loan to be defaulted is higher if Revolve Utilization Line is > 50%.

Bivariate Analysis

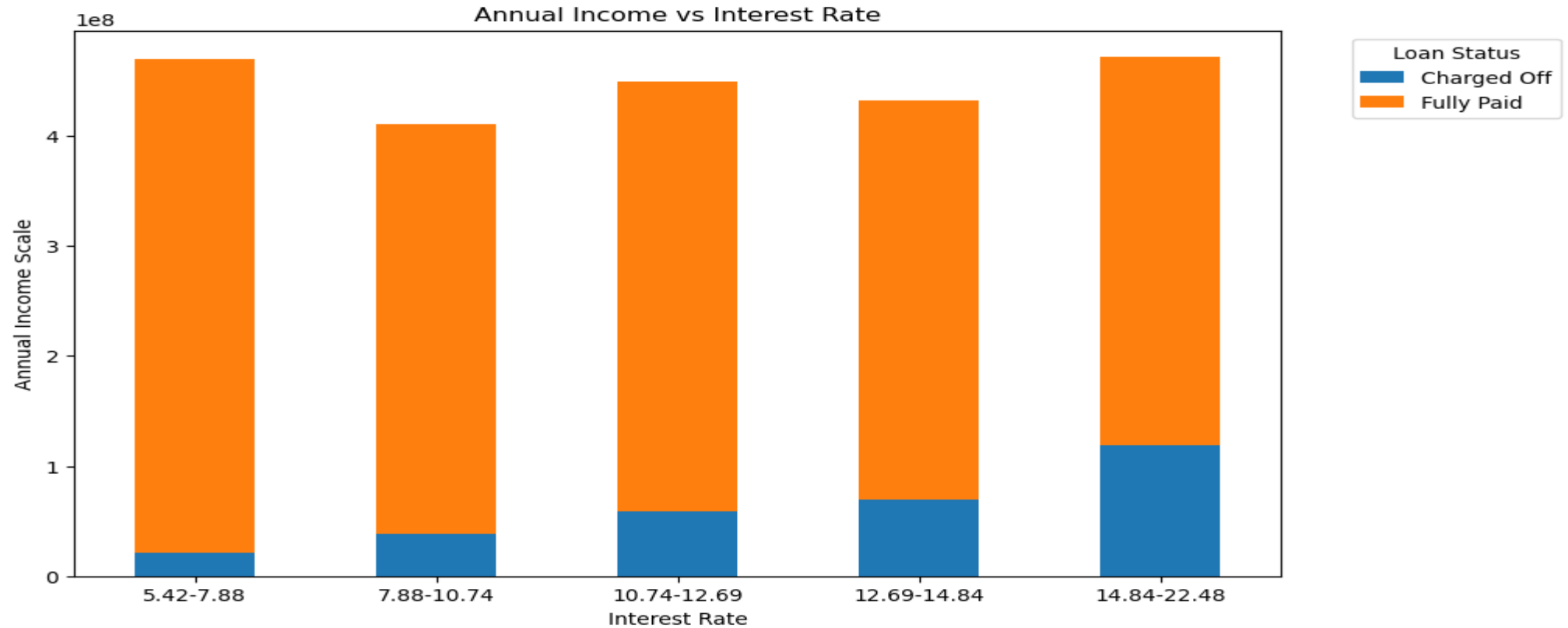
For performing Bivariate Analysis, Buckets has been created for following Quantative Variable.

1. Loan Amount
2. Interest Rate
3. Instalment
4. Annual Income
5. DTI
6. Revolve Utilization.

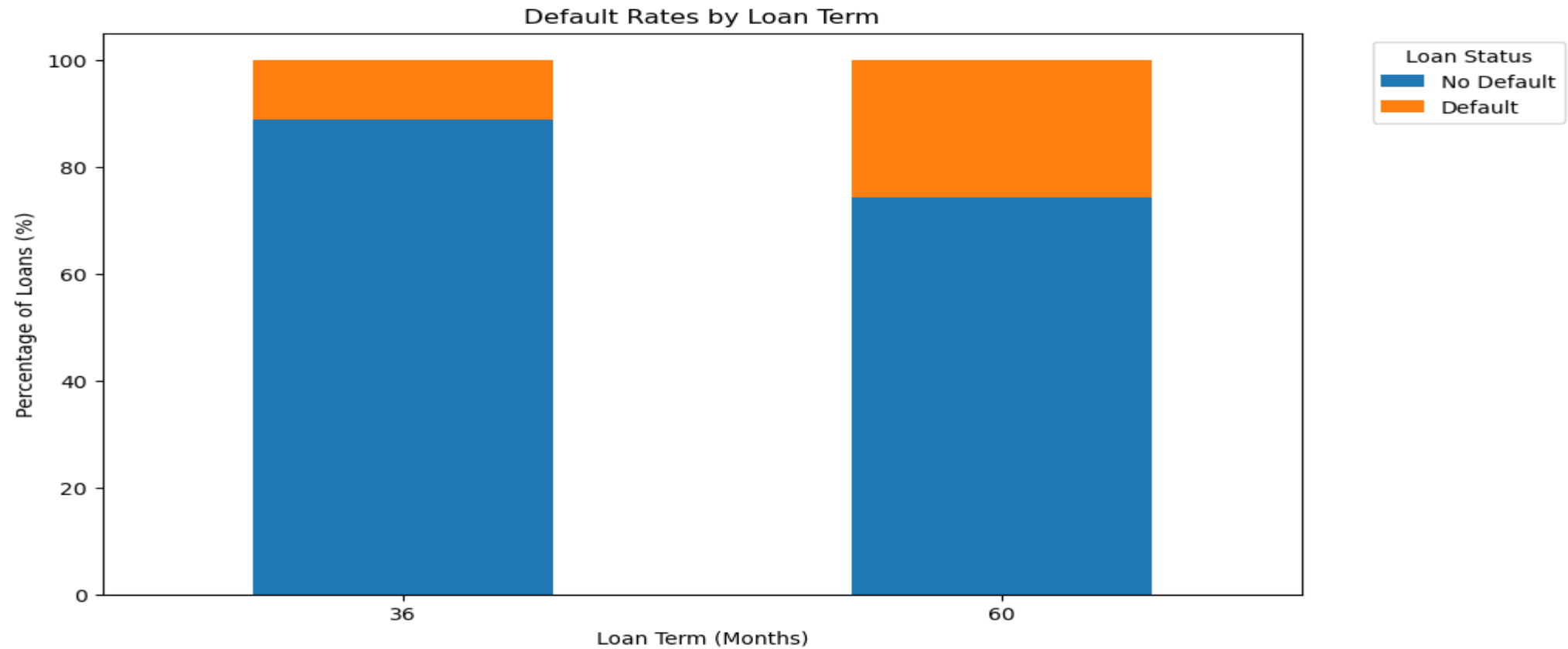
Loan Amount vs Interest Rate



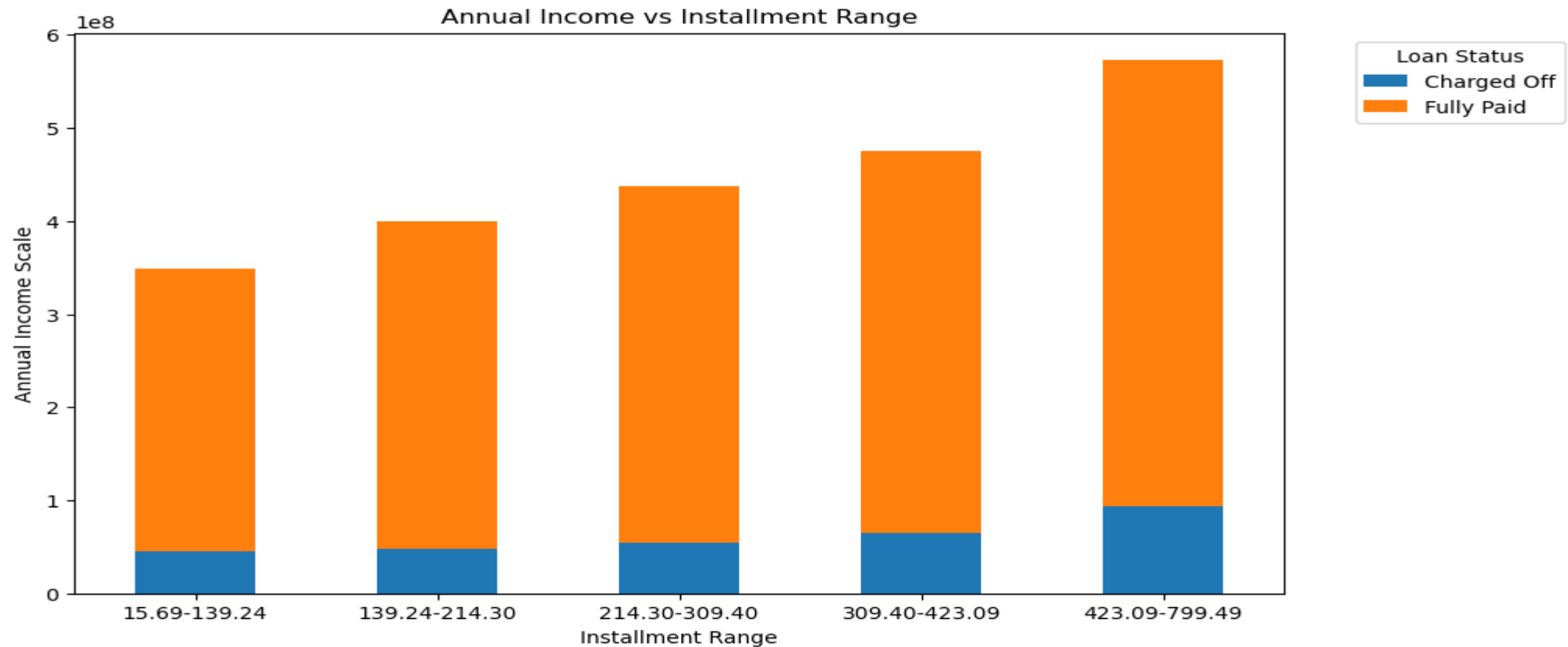
Annual Income vs Interest Rate



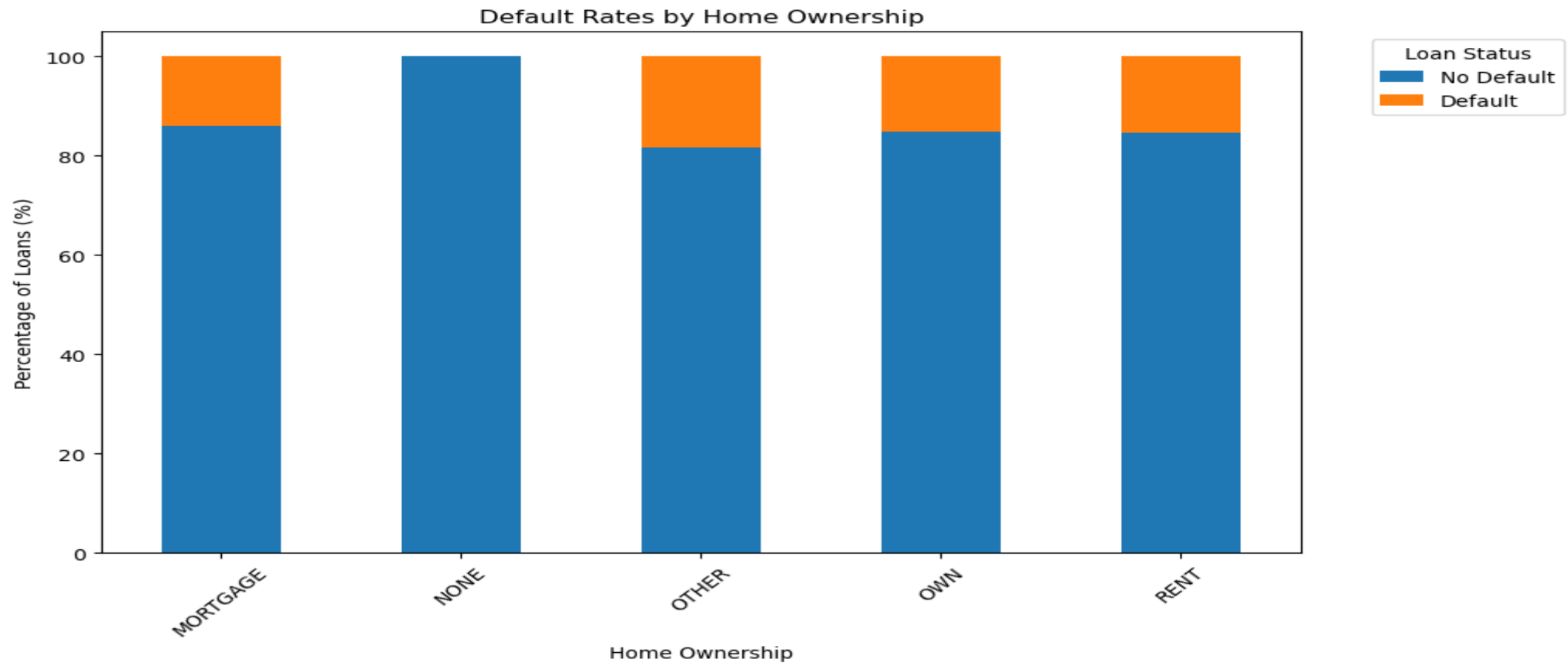
Loan Amount vs Loan Term



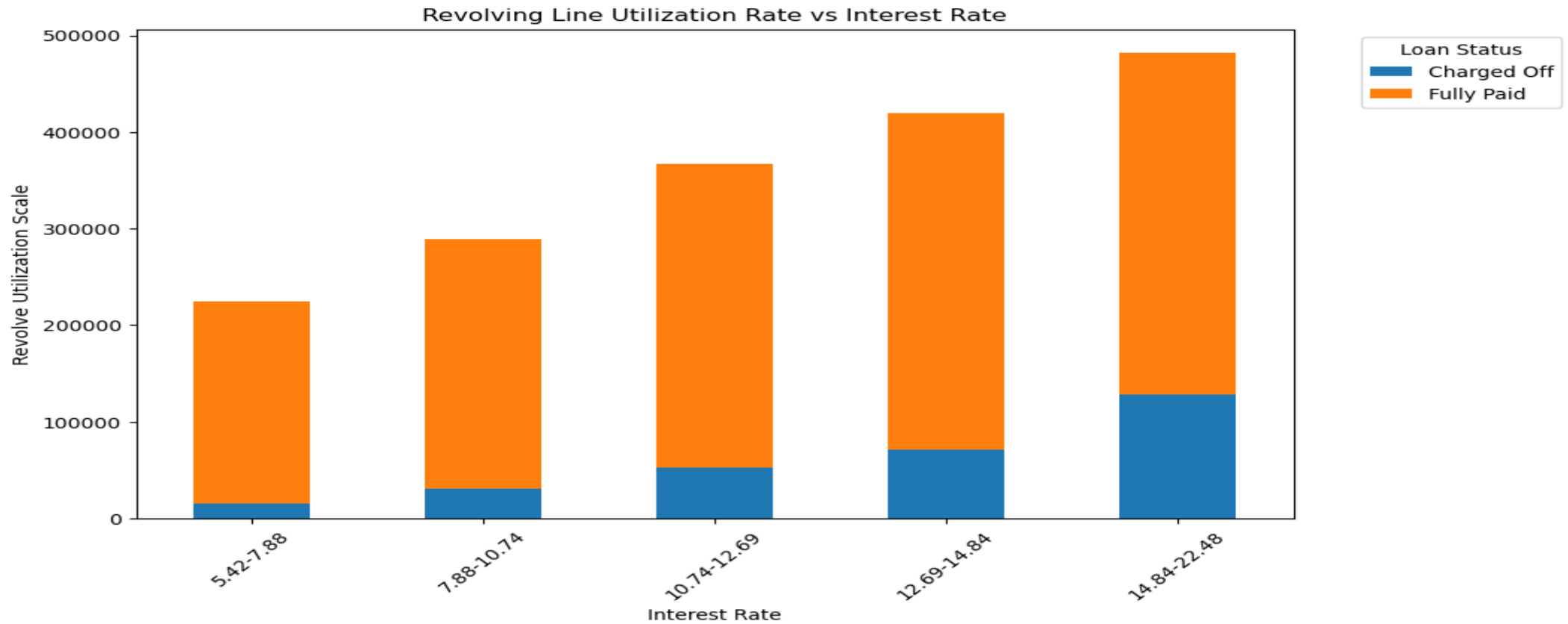
Annual Income vs Installment



Loan Status vs Home Owner Ship



Revolve Utilization vs Interest Rate



Observations from Bivariate Analysis

- The chance of loan to be defaulted is a higher when there is a higher interest rate for lower amount of loan.
- The chance of loan to be defaulted is higher when interest rate is high & applicant annual income is low.
- The chance of loan to be defaulted is higher when both Interest rate & DTI is high.
- The chance of loan to be defaulted is higher when income of an applicant is less & monthly installment is high.
- The chance of loan amount to be defaulted is high when loan amount is high & loan applicant does not own a house.
- The chance of loan amount to be defaulted is high when both interest rate & Revolving Utilization Rate is high.

Conclusion

Based on the analysis, it suggests that loan default risk is not dependent on a single factor. Larger loan amounts, higher interest rates, longer loan terms, higher credit utilization and applicants with history of delinquency appears to be associated with higher default rates. These findings can help the “*consumer finance company*” in performing their risk assessment based on the suggested features when considering loan approvals.