# LENDING CLUB CASE STUDY

# INTRODUCTION

**Objective**: In this case study, we aspire to analyze the factors contributing to loan defaults. These factors can help lending institutions better assess the risks associated with lending to individual borrowers.

**Dataset**: The dataset contains loan data with various borrower attributes. These include the loan amount, interest rate, annual income, and loan status.

**Approach**: We will perform data transformation, formatting, and exploration to identify key factors that have a significant impact on the loan status. Further, we will conduct a univariate analysis of numerical and categorical variables to understand their distributions and relationships with the target variable, loan status.

**Key Findings**: Our analysis will help to identify the factors associated with higher loan default rates. And it can guide lending institutions in making informed decisions when assessing borrower risk.

By understanding the factors contributing to loan defaults, lending institutions can develop more effective risk assessment measures. They can also improve overall portfolio performance and create decision-making tools to evaluate borrower risk more accurately.

# DATA CLEANING & MANIPULATION

**Handling Missing Values**

- Dropping missing values

- Imputing Mode value for Emp_length and pub_rec_bankruptcies.

**Data Transformation and Formatting**

- Convert 'int_rate' and 'revol_util' columns to numerical data types by removing the '%' sign and converting them to float.

- Convert 'emp_length' column to integer values by extracting numerical data and mapping special cases.

- Change 'issue_d' column to datetime format for better time-based analysis.

```python
#Imputing the null values in the column 'emp_length' with the mode value
loan_df['emp_length'].fillna(loan_df['emp_length'].mode()[0], inplace=True)


#Imputing the null values in the column 'pub_rec_bankruptcies' with the mode value
loan_df['pub_rec_bankruptcies'].fillna(loan_df['pub_rec_bankruptcies'].mode()[0], inplace=True)
```

# DATA CLEANING & MANIPULATION

**Derived Metrics**

- Add new columns 'issue_year' and 'issue_month' to facilitate further exploration of the dataset.
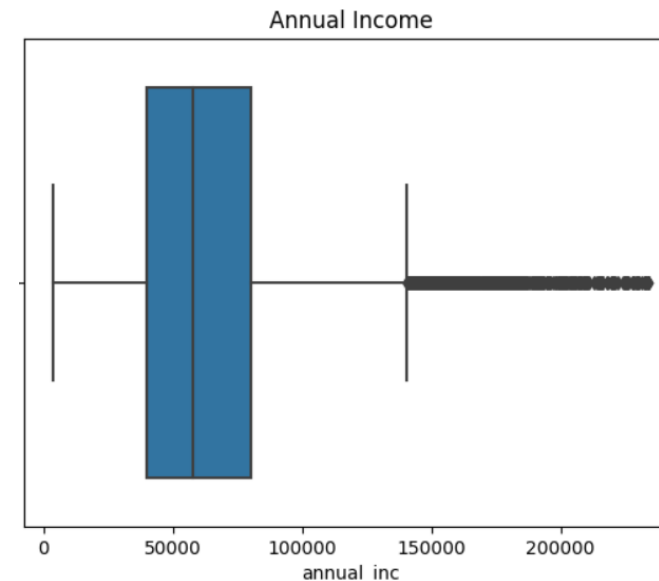
**Handling Outliers**

- Remove outliers from 'annual_inc' column by keeping only the data below the 99th quantile.

**Dropping Highly Correlated Columns**

- Remove columns with high correlation to avoid multicollinearity issues in the analysis.
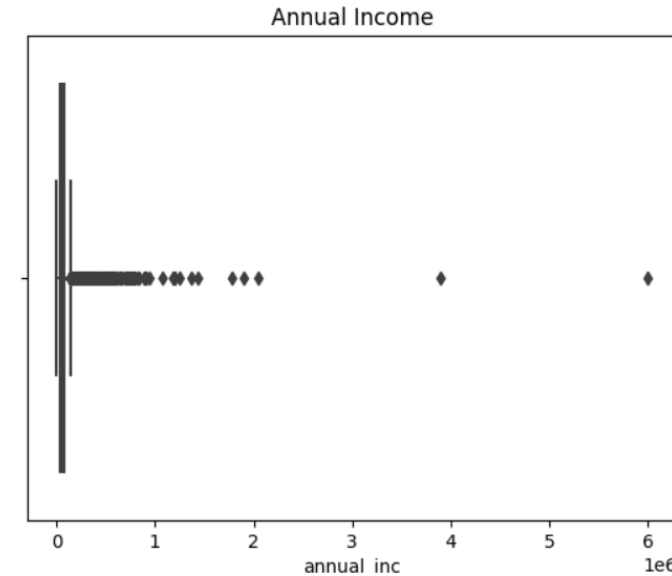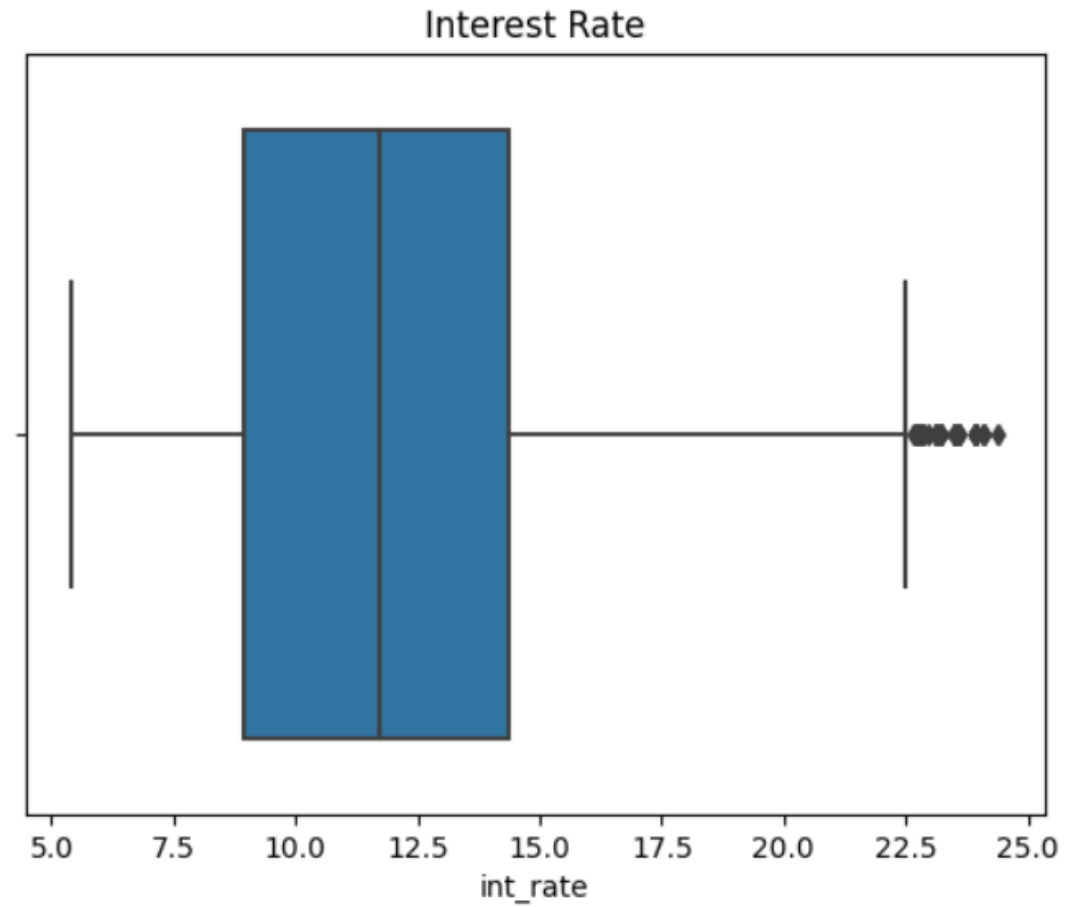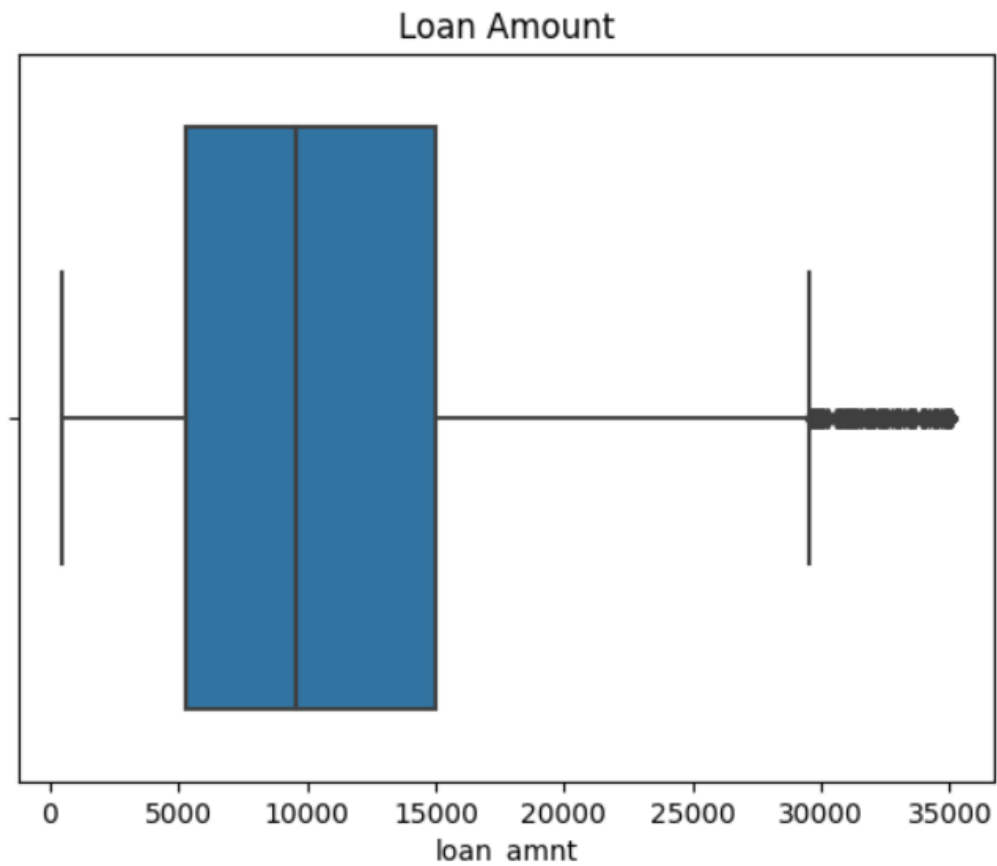


After Removing Outlier

# UNIVARIATE ANALYSIS

**Target Variable:**

- Loan_Status is the target variable.

- "Fully Paid" and "Charged Off" values are considered for analysis.

- Out of 38577 records, 85.4% records have loan_status as 'Fully Paid' and 14.6% records have loan_status as 'Charged Off'.
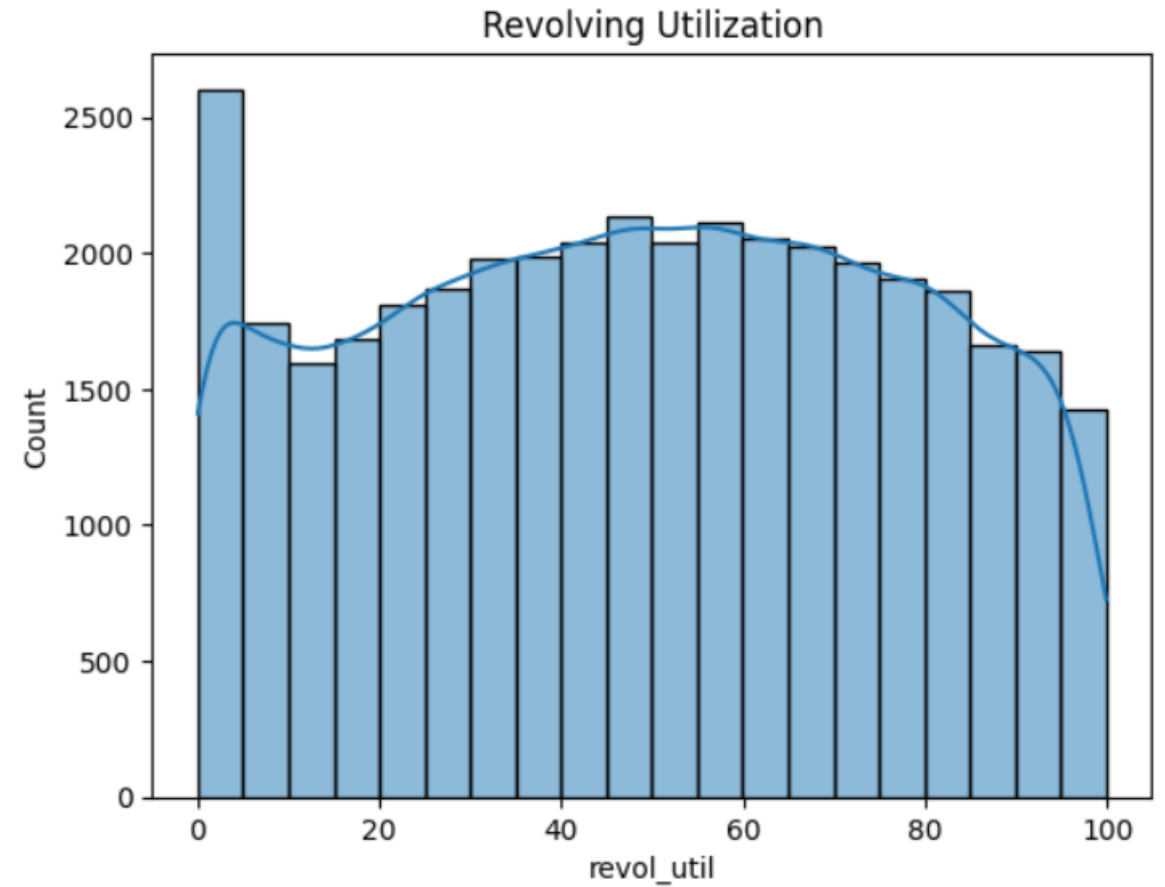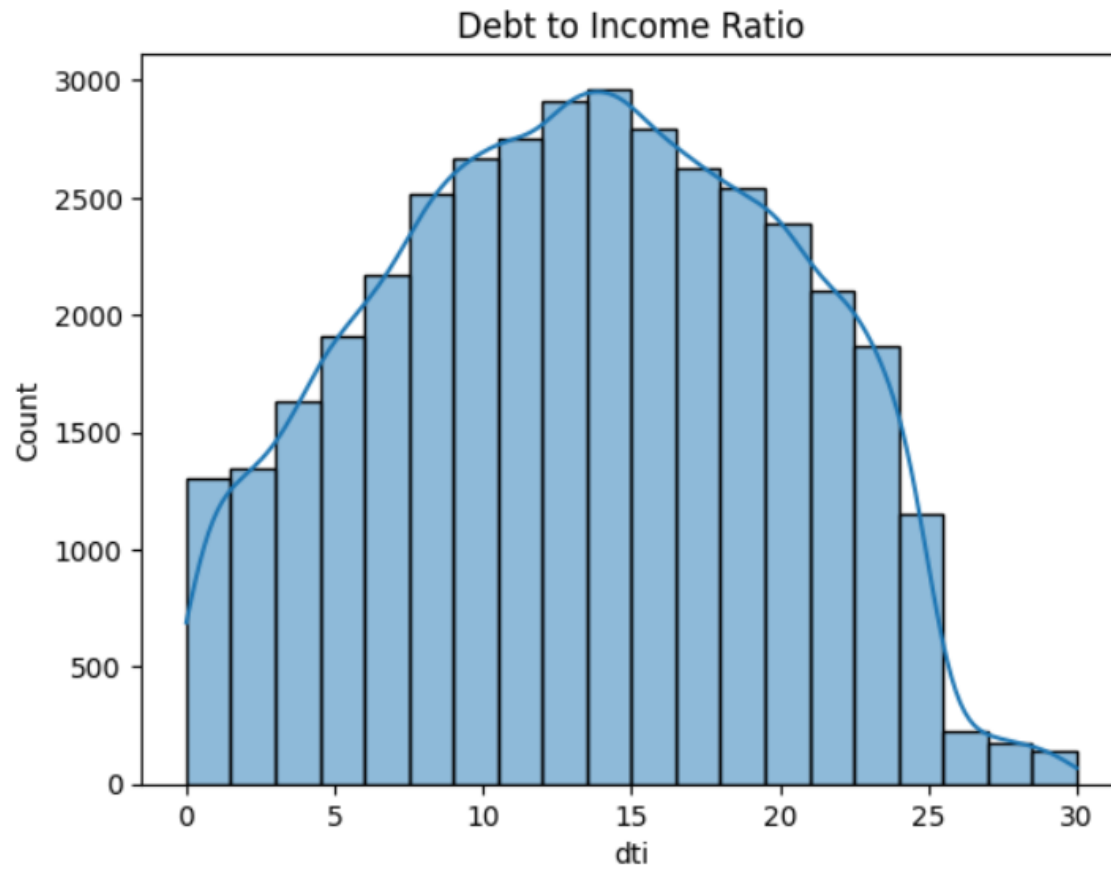
- Most of the loans are fully paid back.
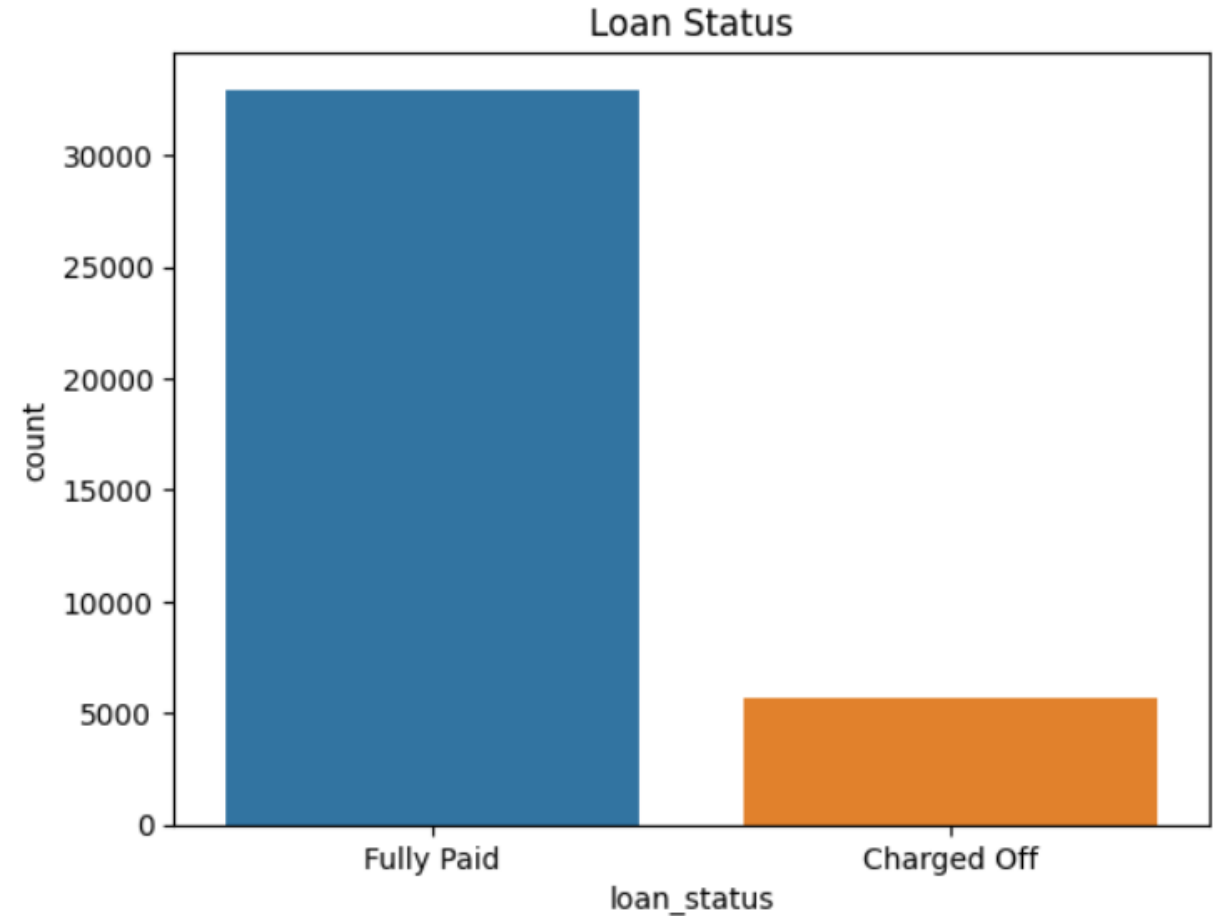


Loan Status

# UNIVARIATE ANALYSIS

**Univariate Analysis of Numerical variables**

# UNIVARIATE ANALYSIS

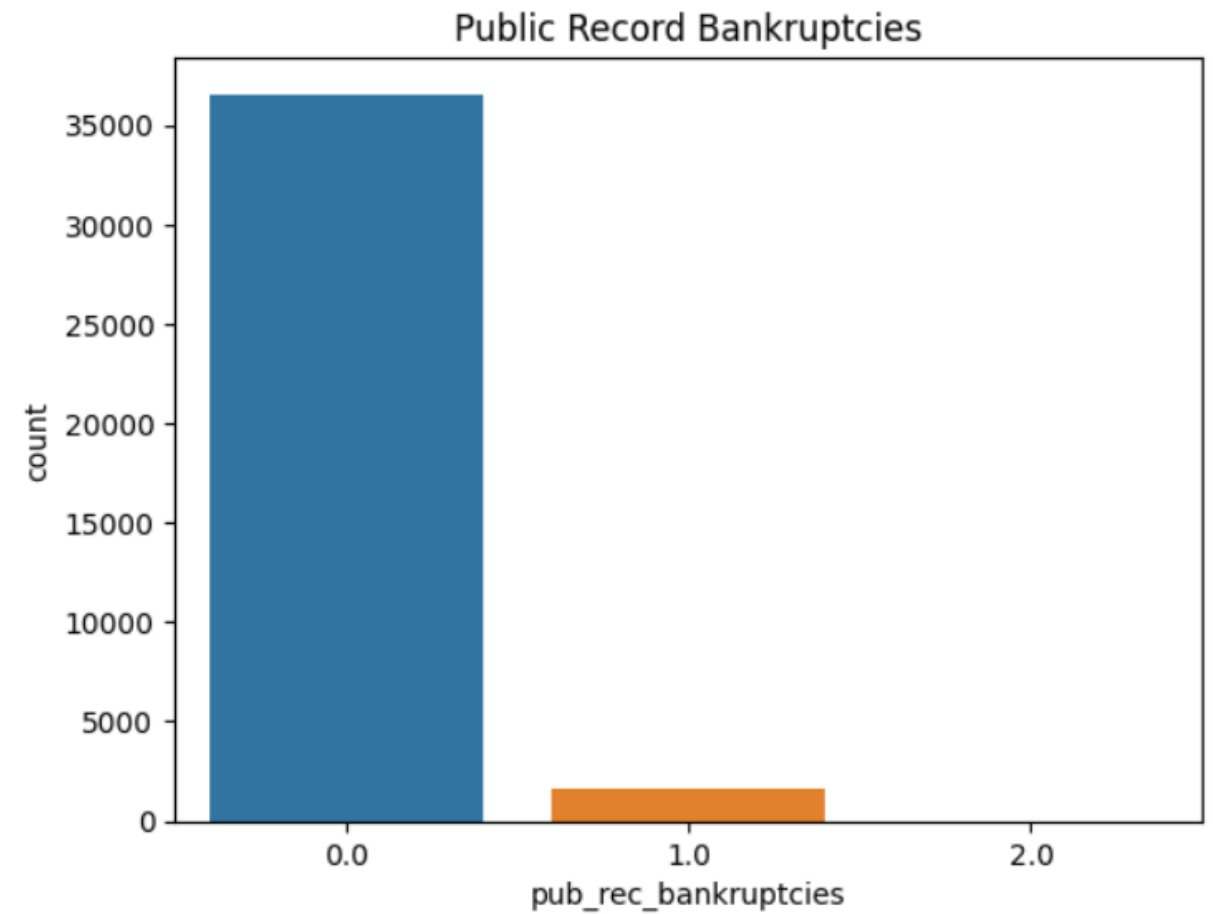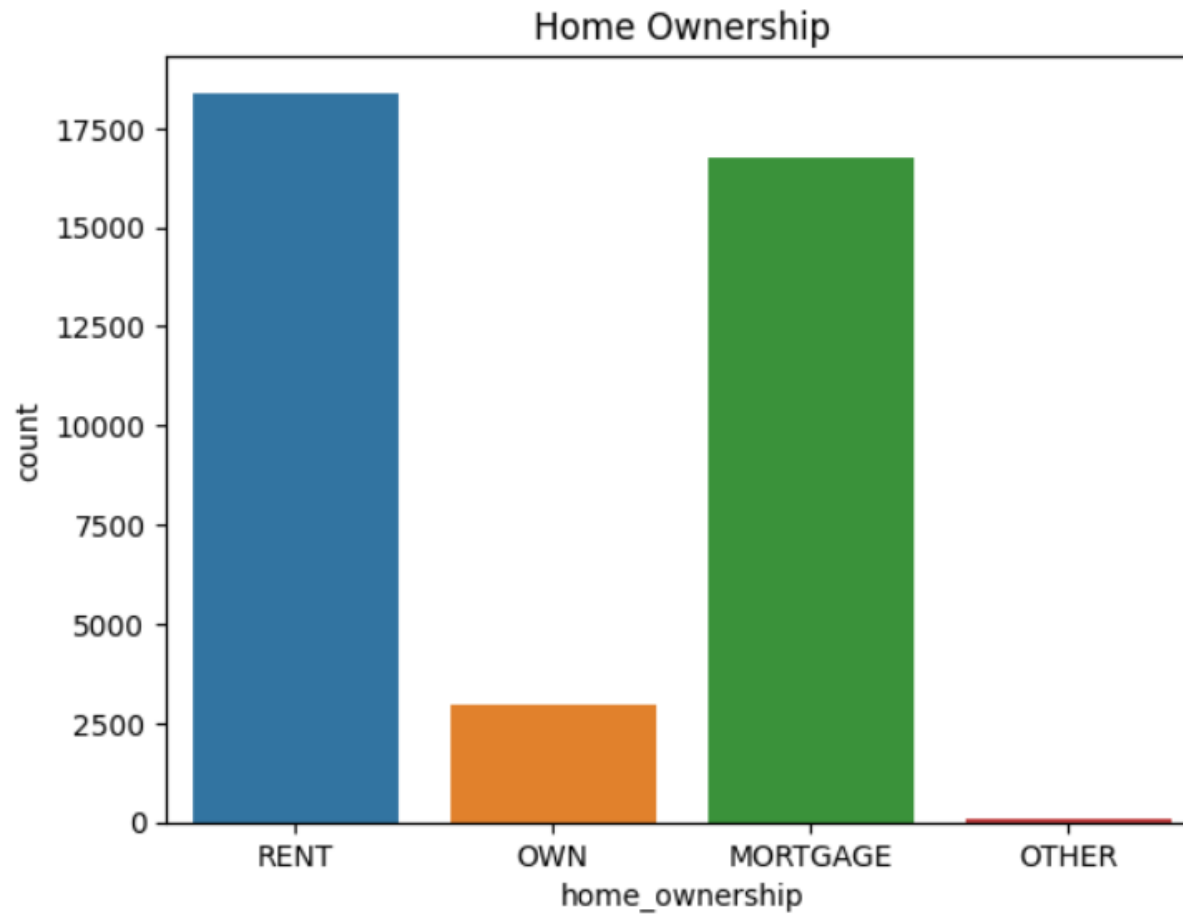**Univariate Analysis of Numerical variables**

# UNIVARIATE ANALYSIS

**Univariate Analysis of Numerical variables - Observations**

- The loan amount has mean value of 11047.02 and median value of 9600.00.The maximum loan amount is 35000.00 and minimum loan amount is 500.00.

- The interest rate is between 5.42% and 24.40% with mean value of 11.93% and median value of 11.71%.

- About 75% of the data has annual income of upto $82000.00.

- The debt to income ratio is between 0.00% and 29.99% with mean value of 13.33% and median value of 13.43%. Lower debt to income ratio indicates higher creditworthiness of the borrower.

- The mean value of revol_util is 48.69, indicating that on average, borrowers are utilizing roughly half of their available revolving credit. it varies widely across the borrower population and may have implications for their ability to repay their debts.
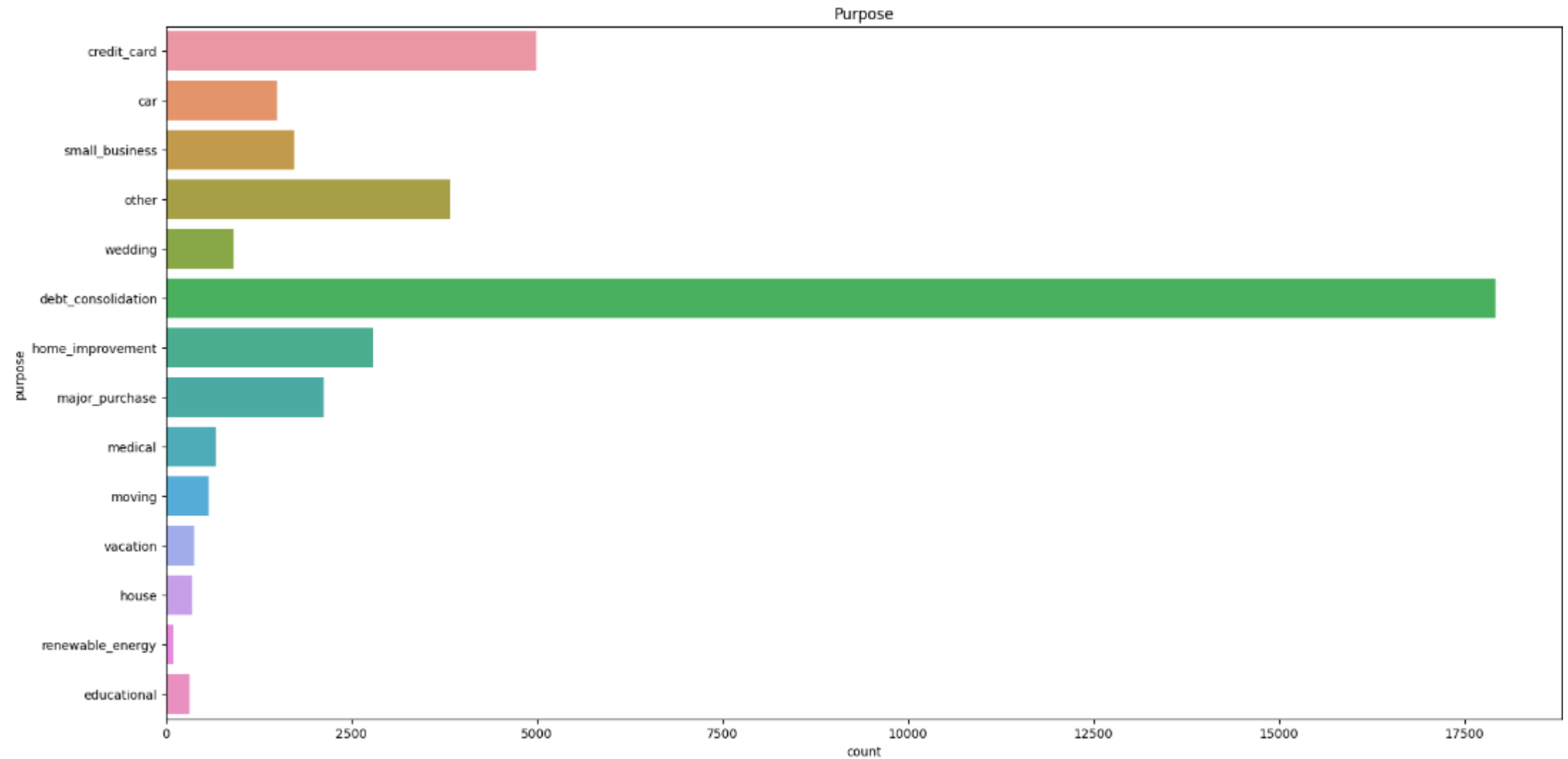
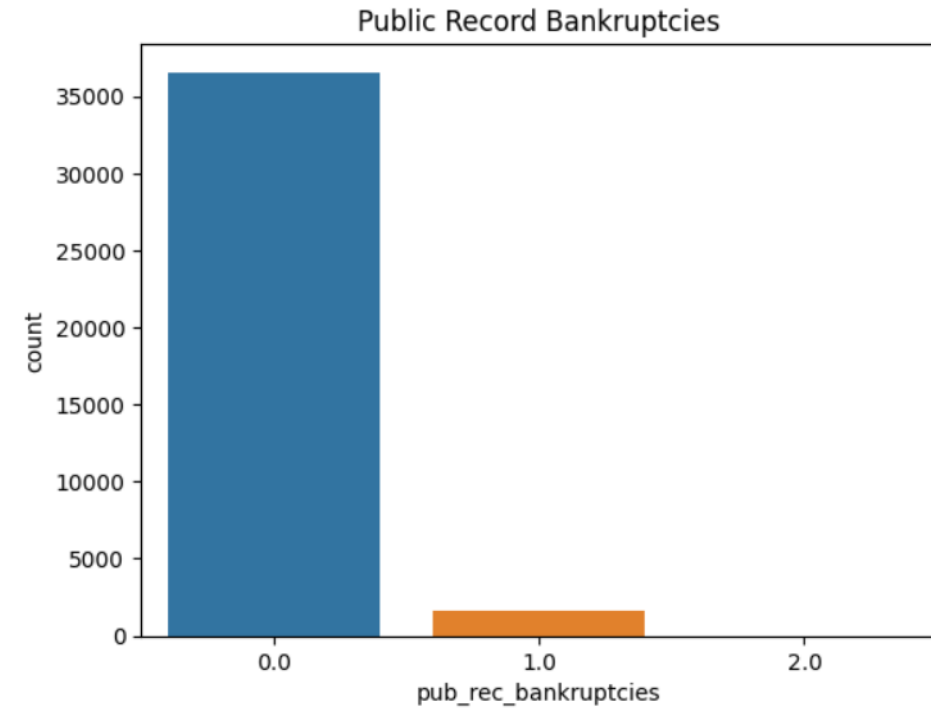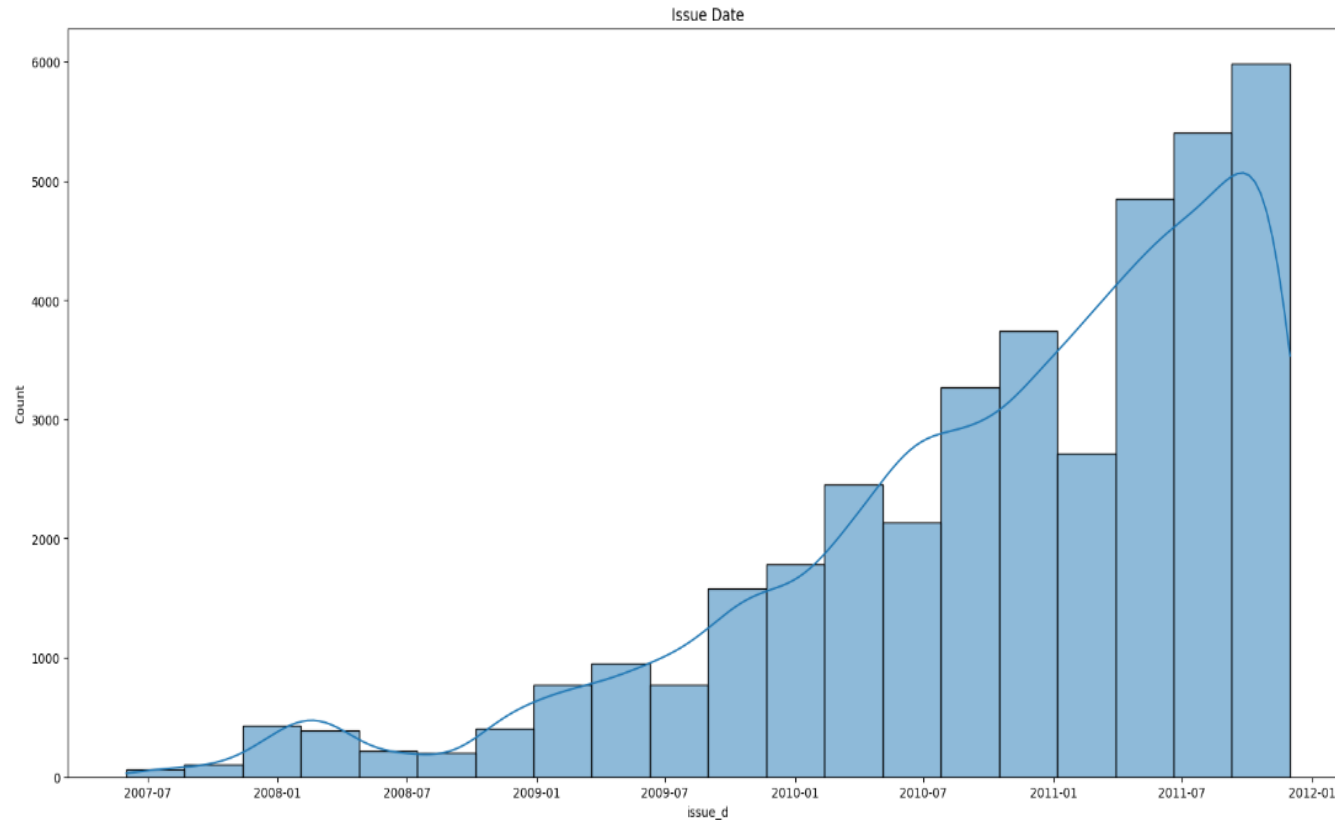# UNIVARIATE ANALYSIS

**Univariate Analysis of Categorical variable**

# UNIVARIATE ANALYSIS

**Univariate Analysis of Categorical variable**

# UNIVARIATE ANALYSIS

**Univariate Analysis of Categorical variable**

# UNIVARIATE ANALYSIS

**Univariate Analysis of Categorical variables - Observations**

- Borrowers have taken loan for 36 months in 75.46% of the cases and for 60 months in 24.54% of the cases. Also charge off rate is higher for 60 months term loans.

- Most of the borrowers have rent or Moragage as their home ownership type

- Most of the borrowers have taken loan for debt consolidation purpose followed by credit card.

- The count of loans issued keeps increasing over the years indicating that the company is growing.

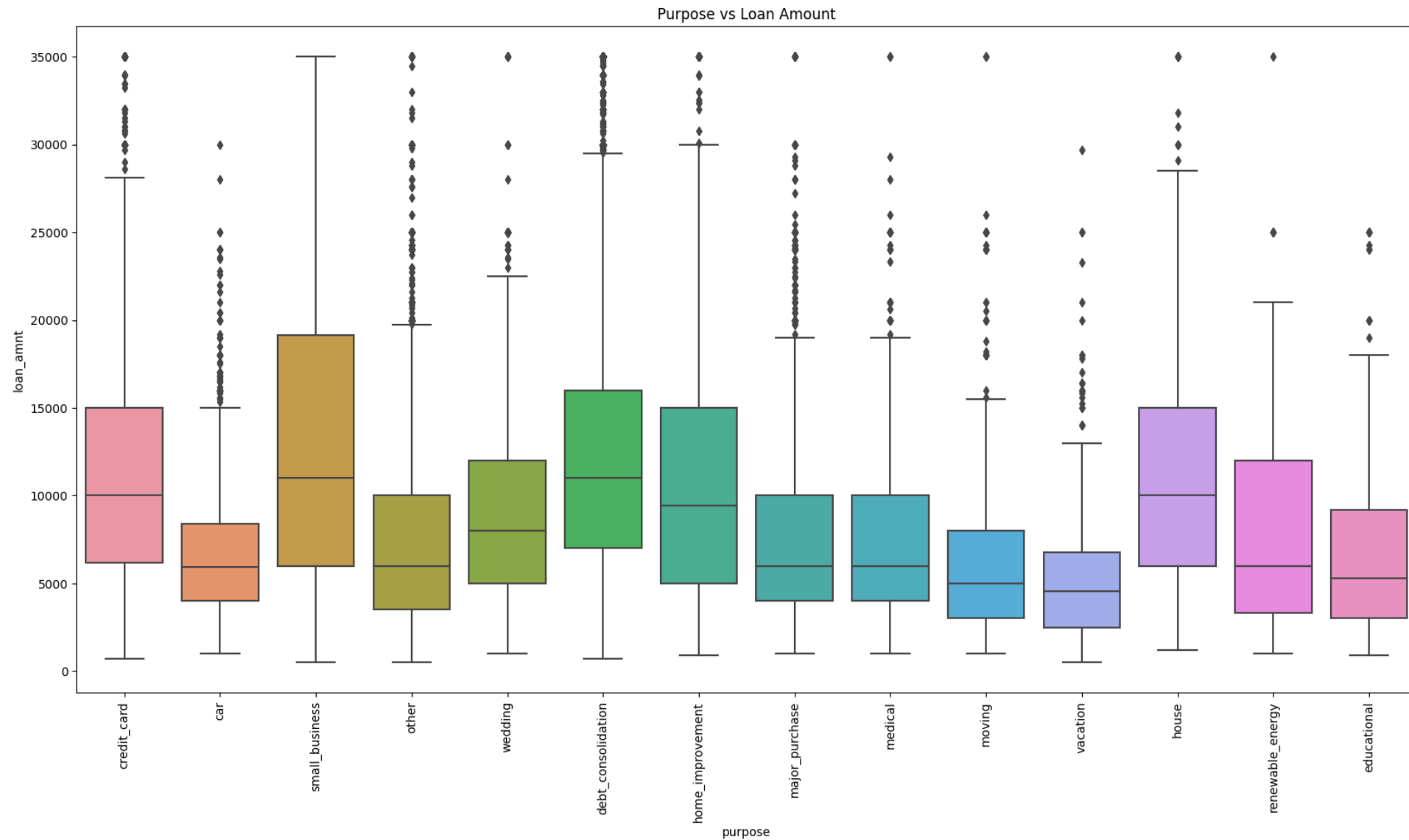- Most of the borrowers have 0 public record bankruptcies.
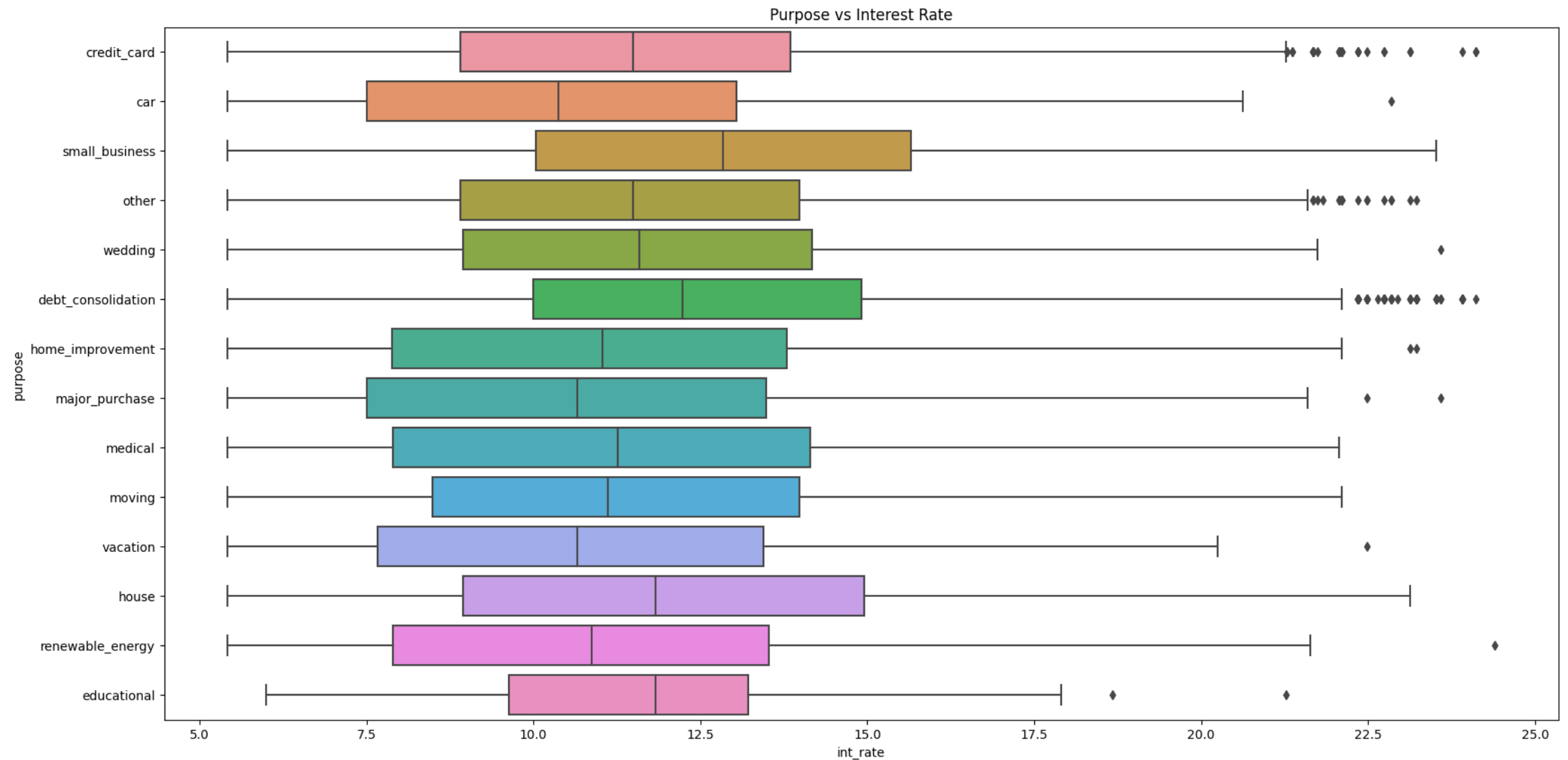
# BIVARIATE ANALYSIS
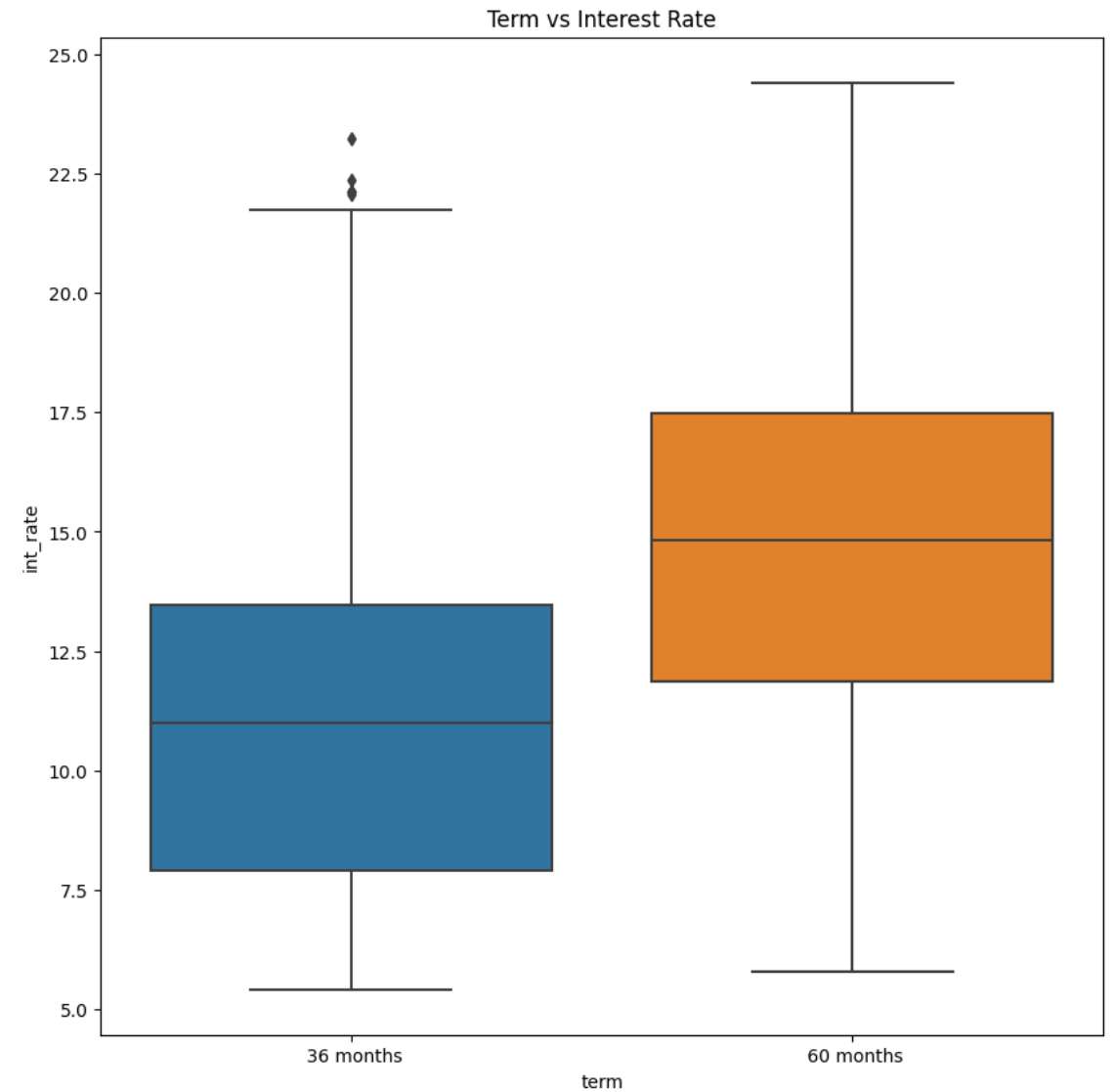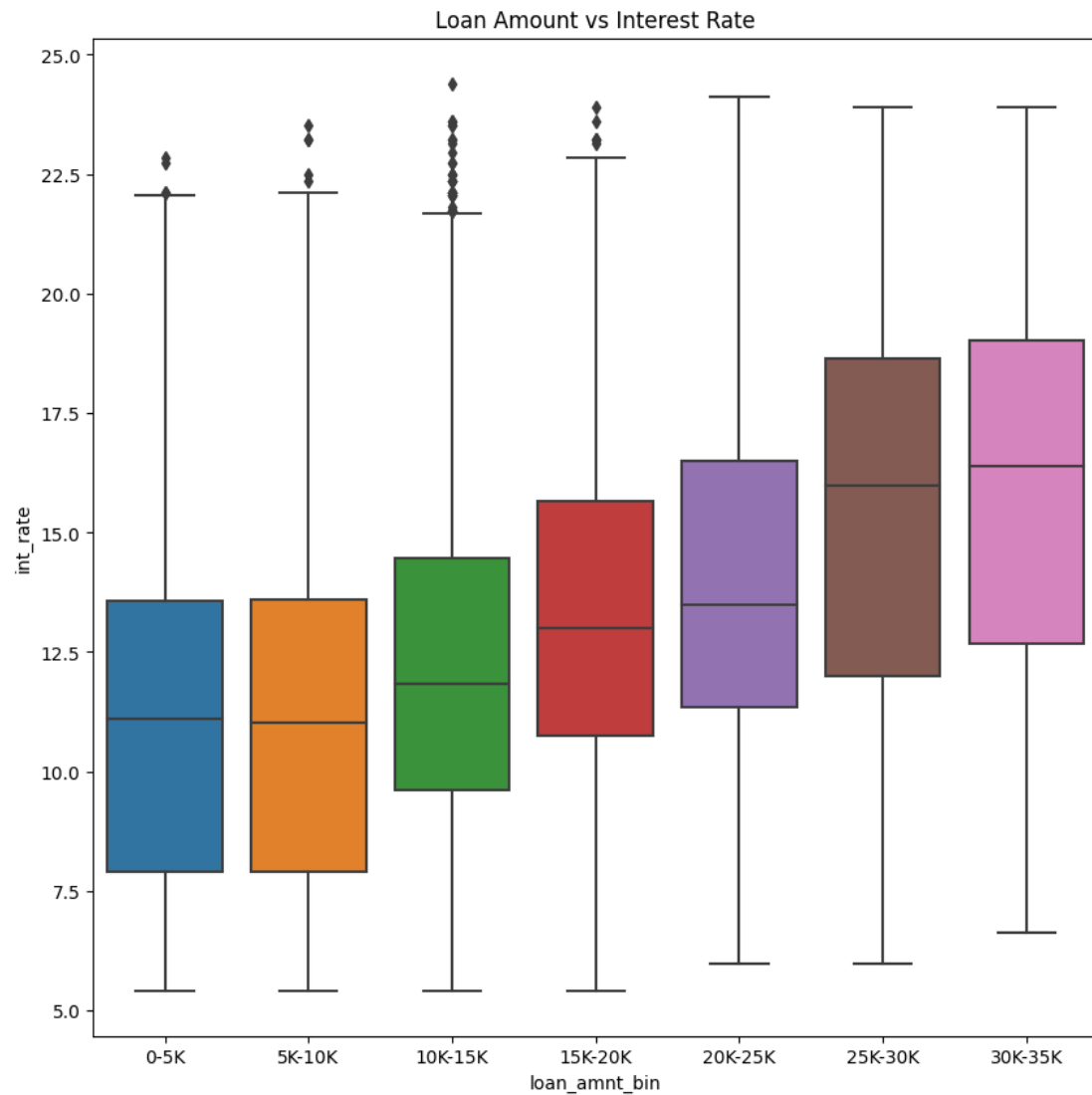
# BIVARIATE ANALYSIS

# BIVARIATE ANALYSIS



Purpose vs Loan Amount

# BIVARIATE ANALYSIS



Purpose vs Interest Rate
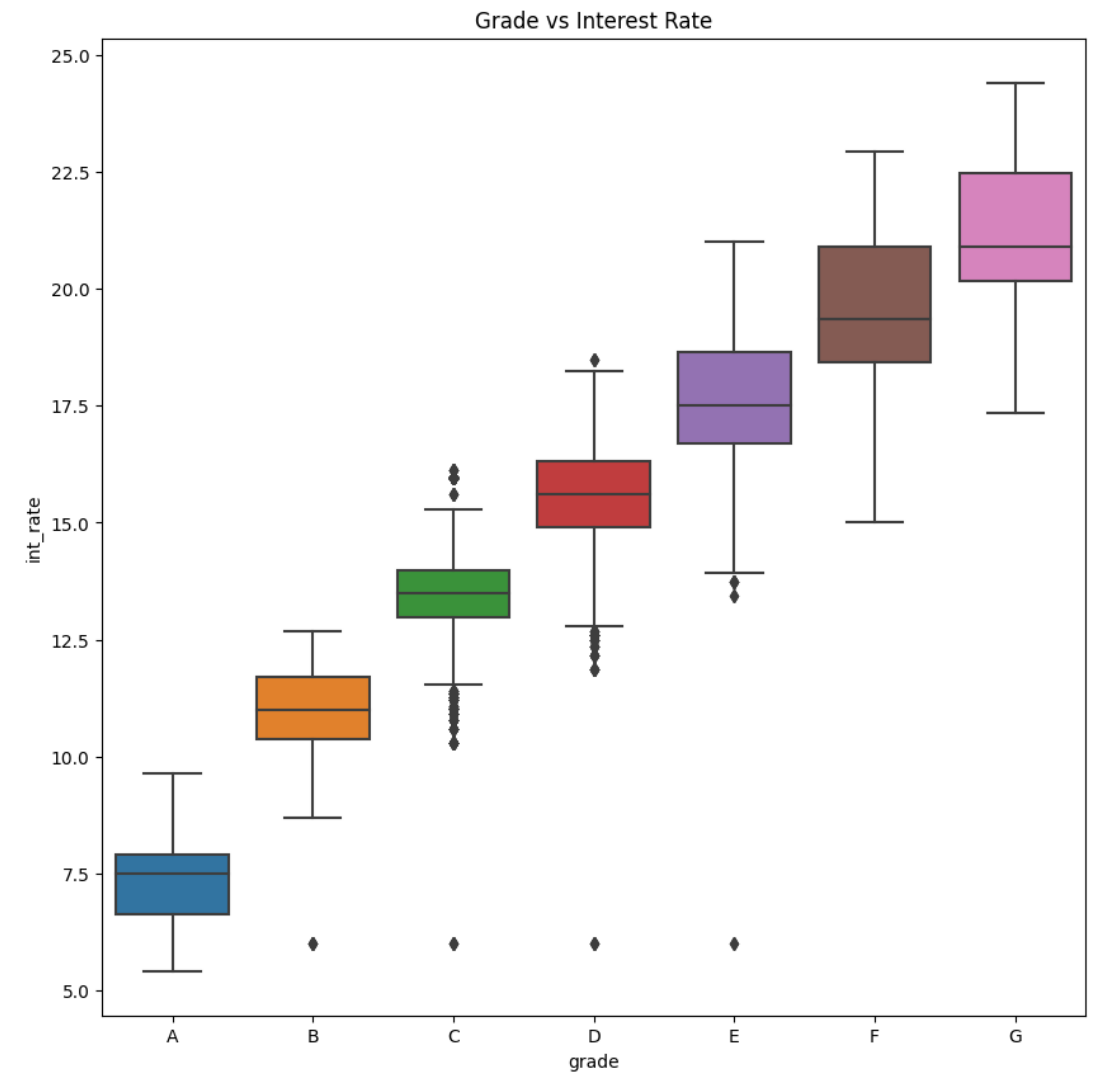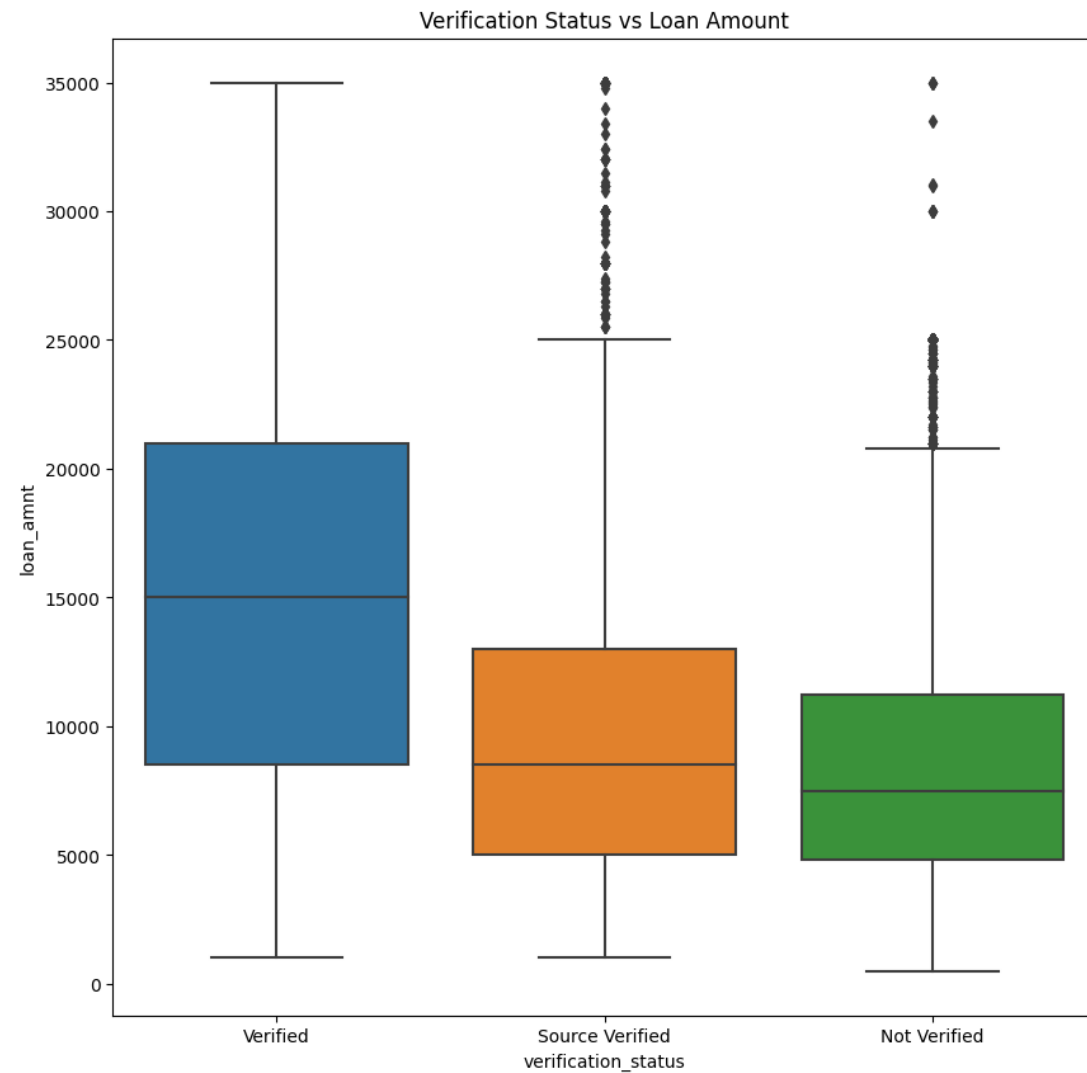
# BIVARIATE ANALYSIS

**Bivariate Analysis - Observations**

- As loan amount increases, interest rate also increases.

- The interest rate is higher for the 60 months term than the 36 months term.

- The loan amount borrowed is higher for verified borrowers than for non verified borrowers.

- As the grade of the loan increases from A to G, the range of interest rate also increases. This implies grade is dependent on interest rate.

- The lower loan amount likely to be taken for term of 36 months and higher loan amount likely to be taken for term of 60 months.

- The loan amount is higher for the purpose of small business followed by debt consolidation and credit card.

- The average interest rate is higher for the purpose of small business followed by debt consolidation. Also the small business tend to have higher interest rate than the other purposes due to its risky nature.
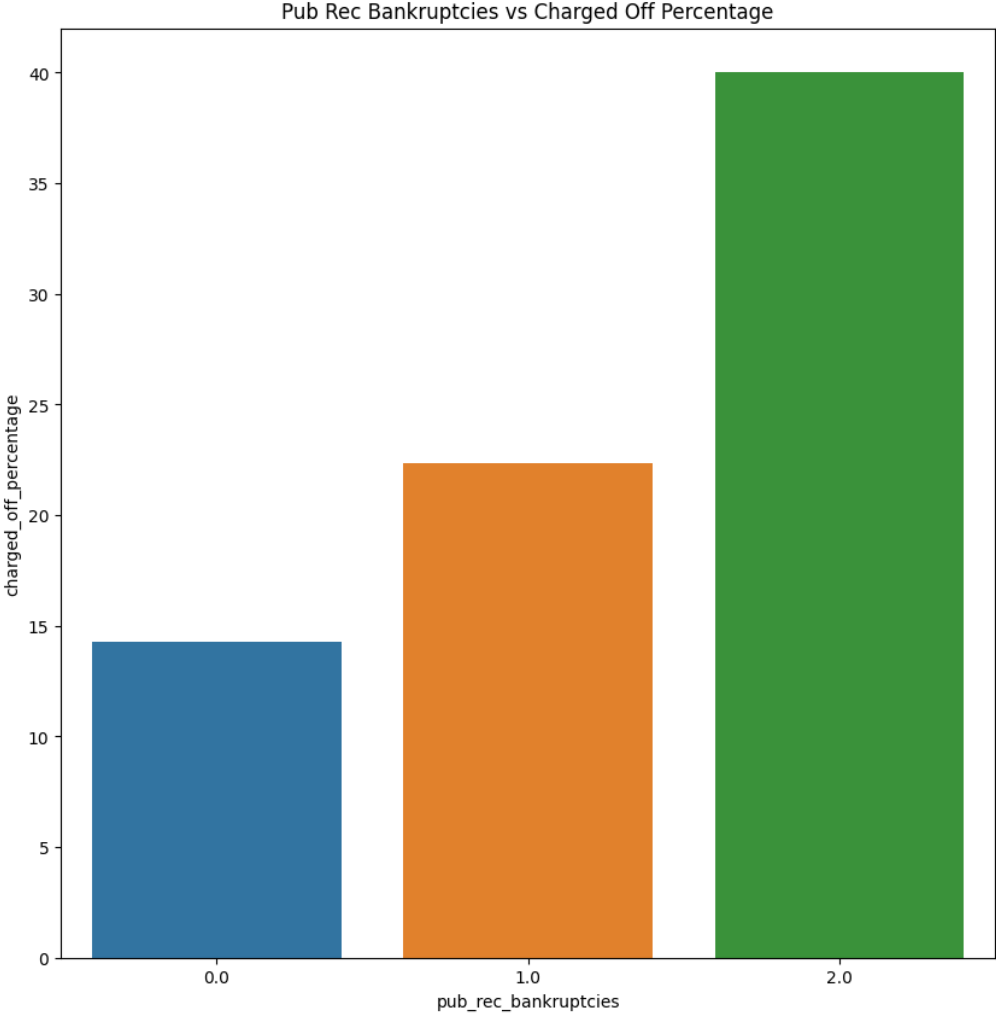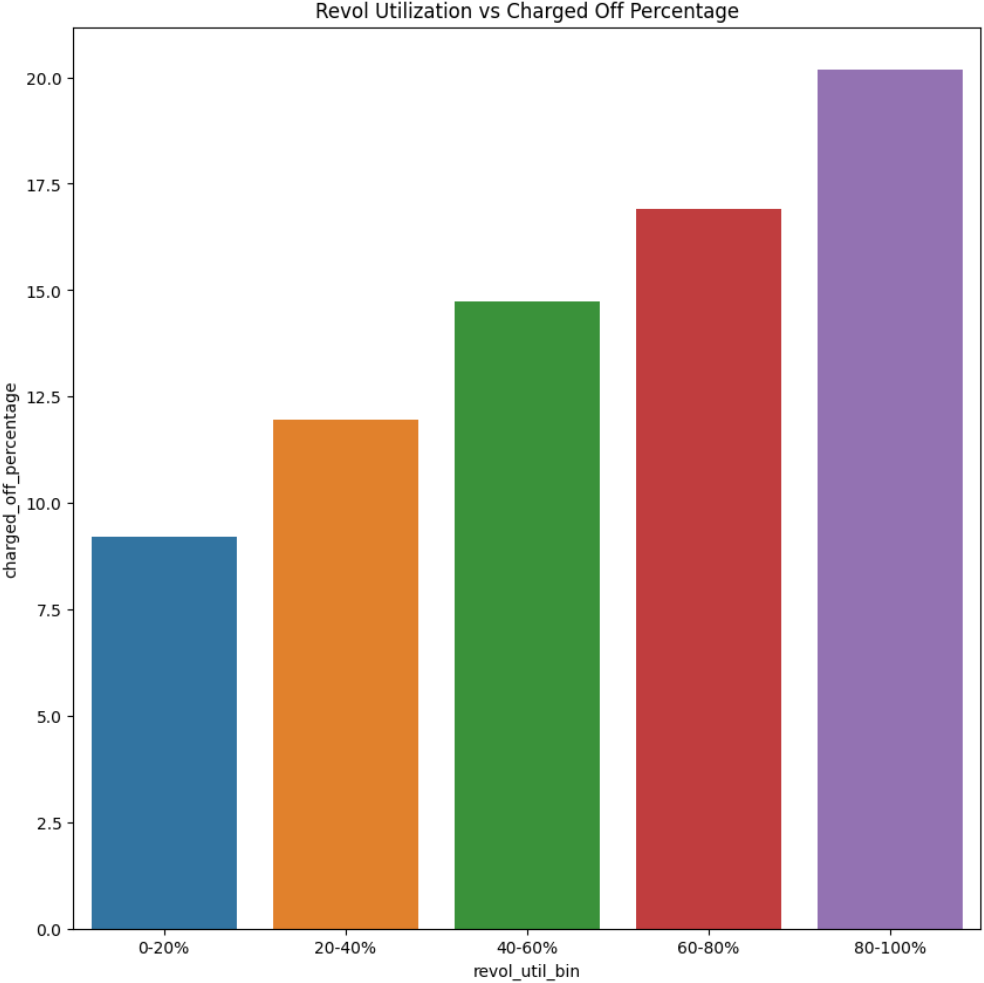
# BIVARIATE ANALYSIS
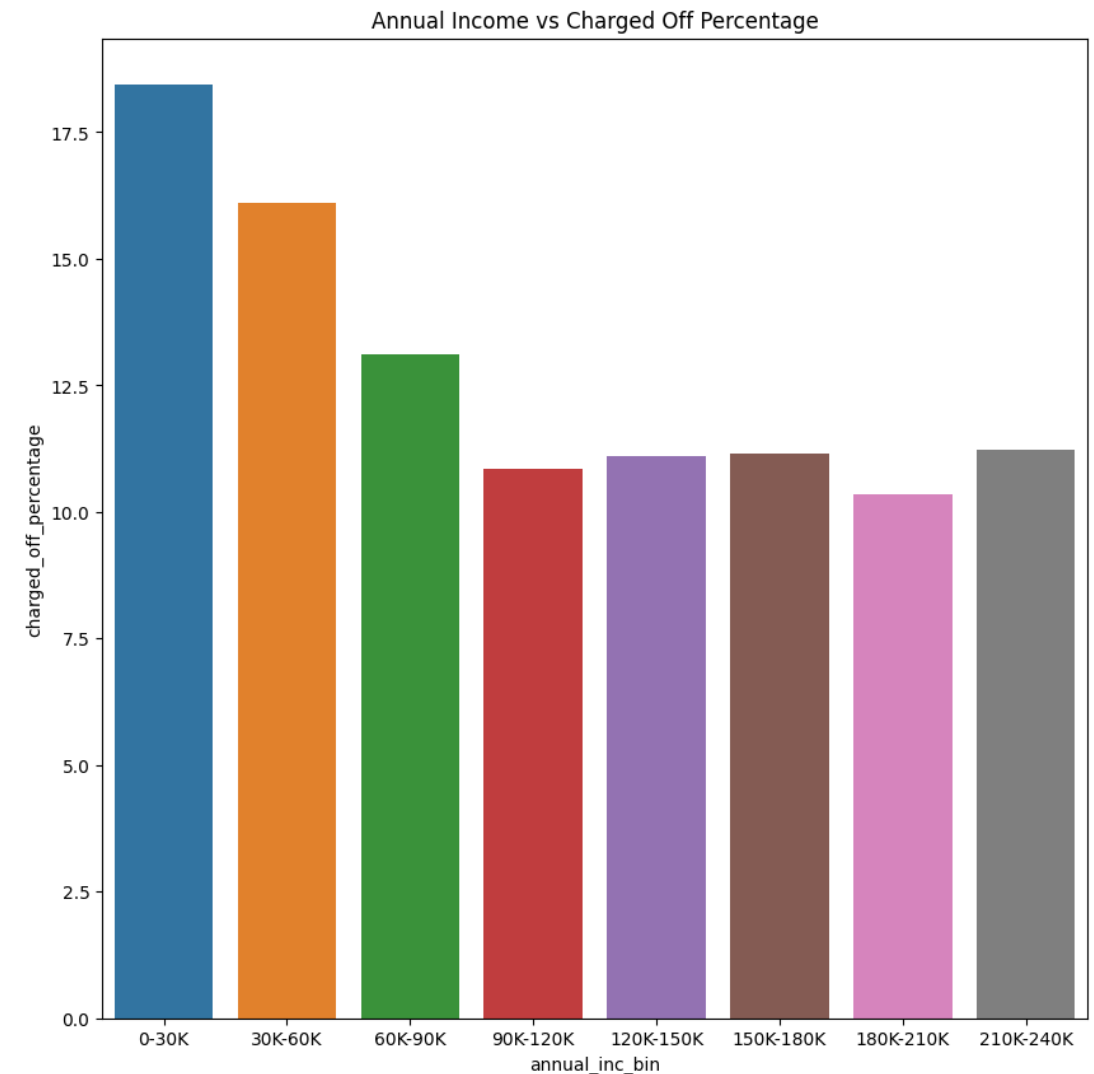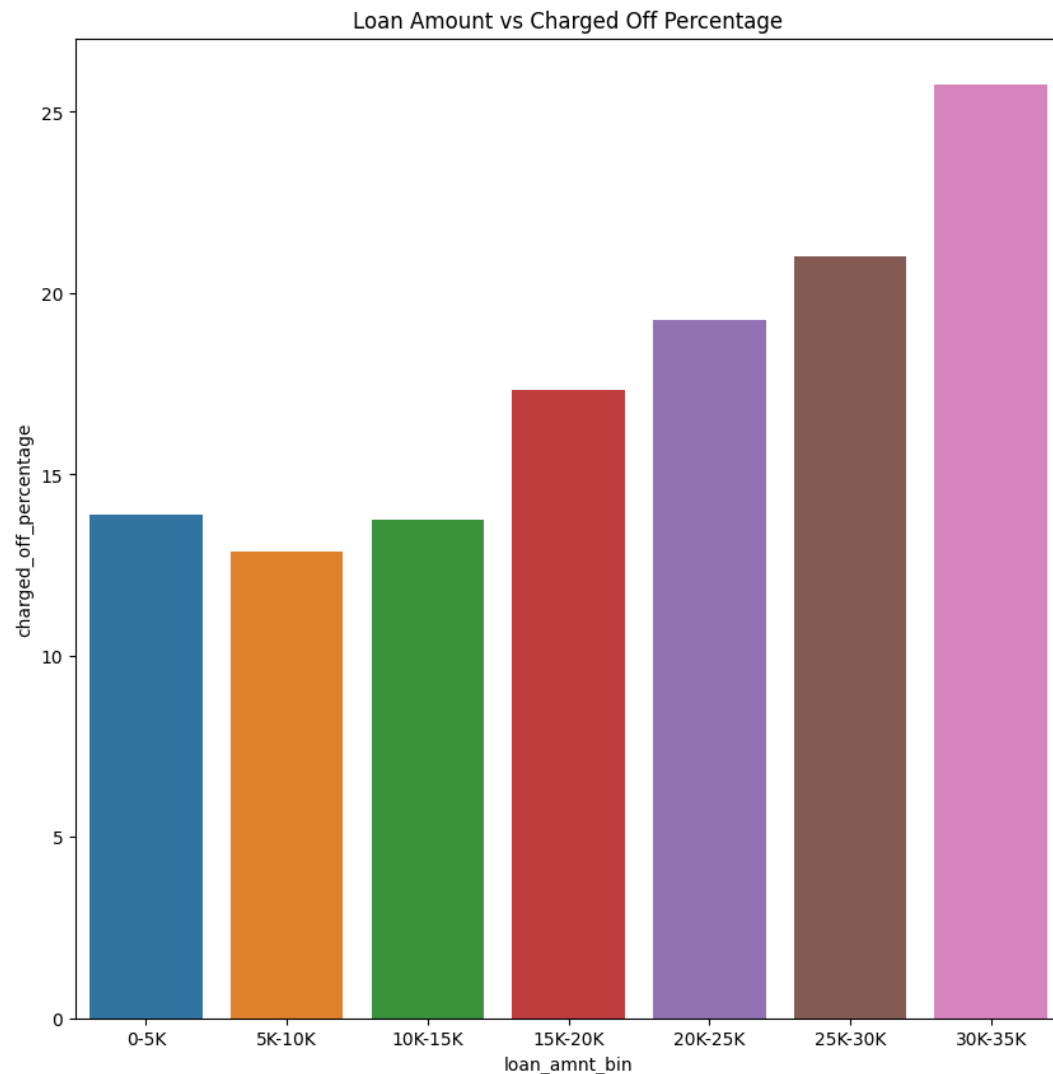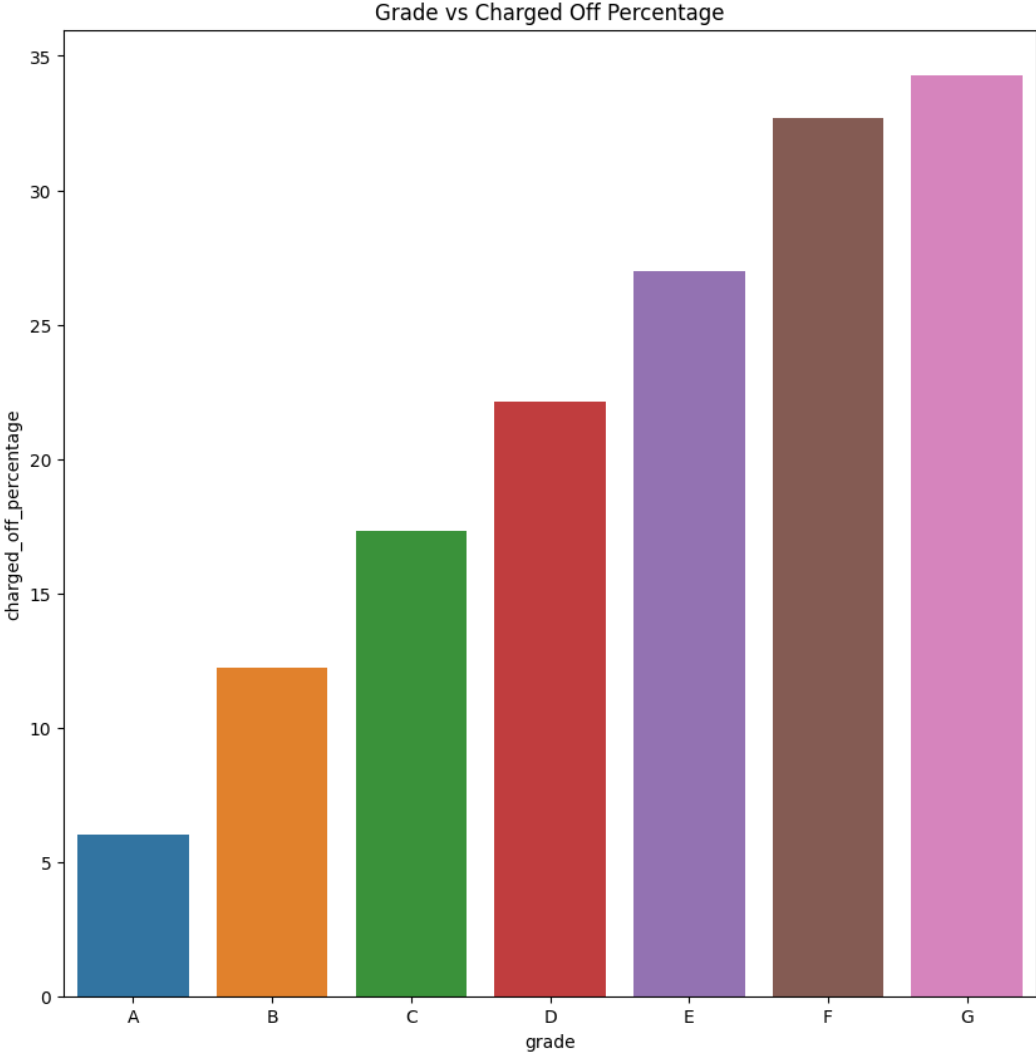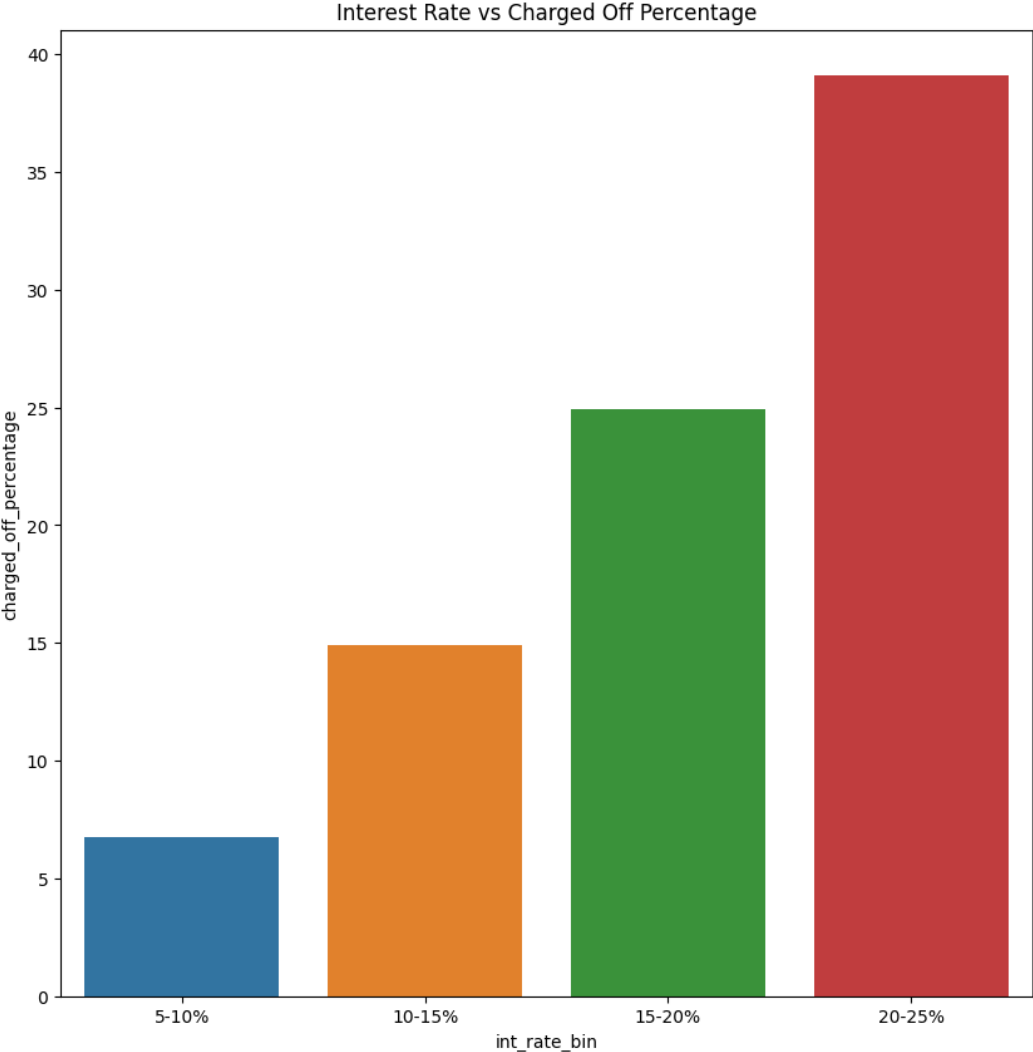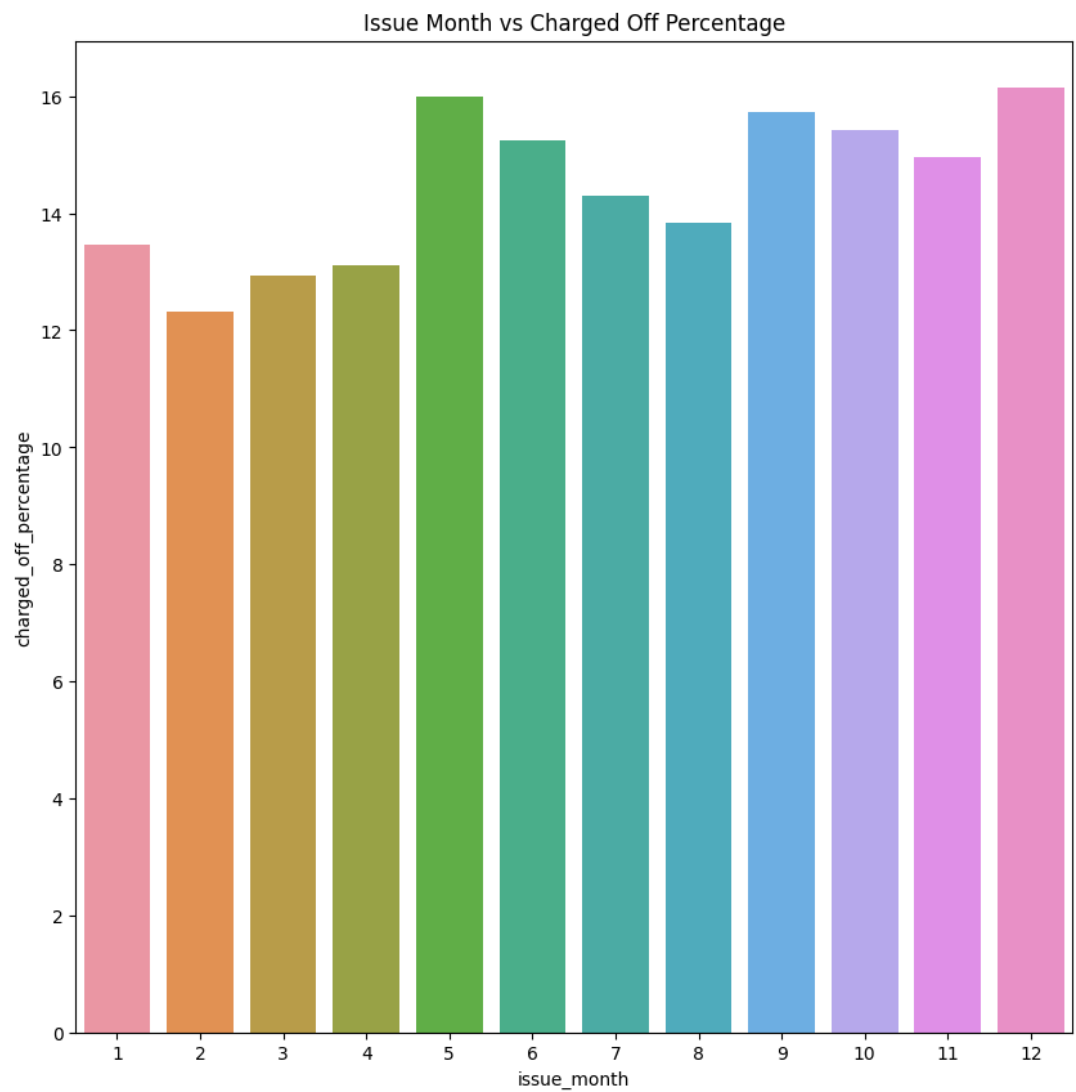
**Bivariate Analysis with Target variable**

# BIVARIATE ANALYSIS

**Bivariate Analysis with Target variable**

# BIVARIATE ANALYSIS

**Bivariate Analysis with Target variable**

# BIVARIATE ANALYSIS

**Bivariate Analysis with Target variable**

# BIVARIATE ANALYSIS

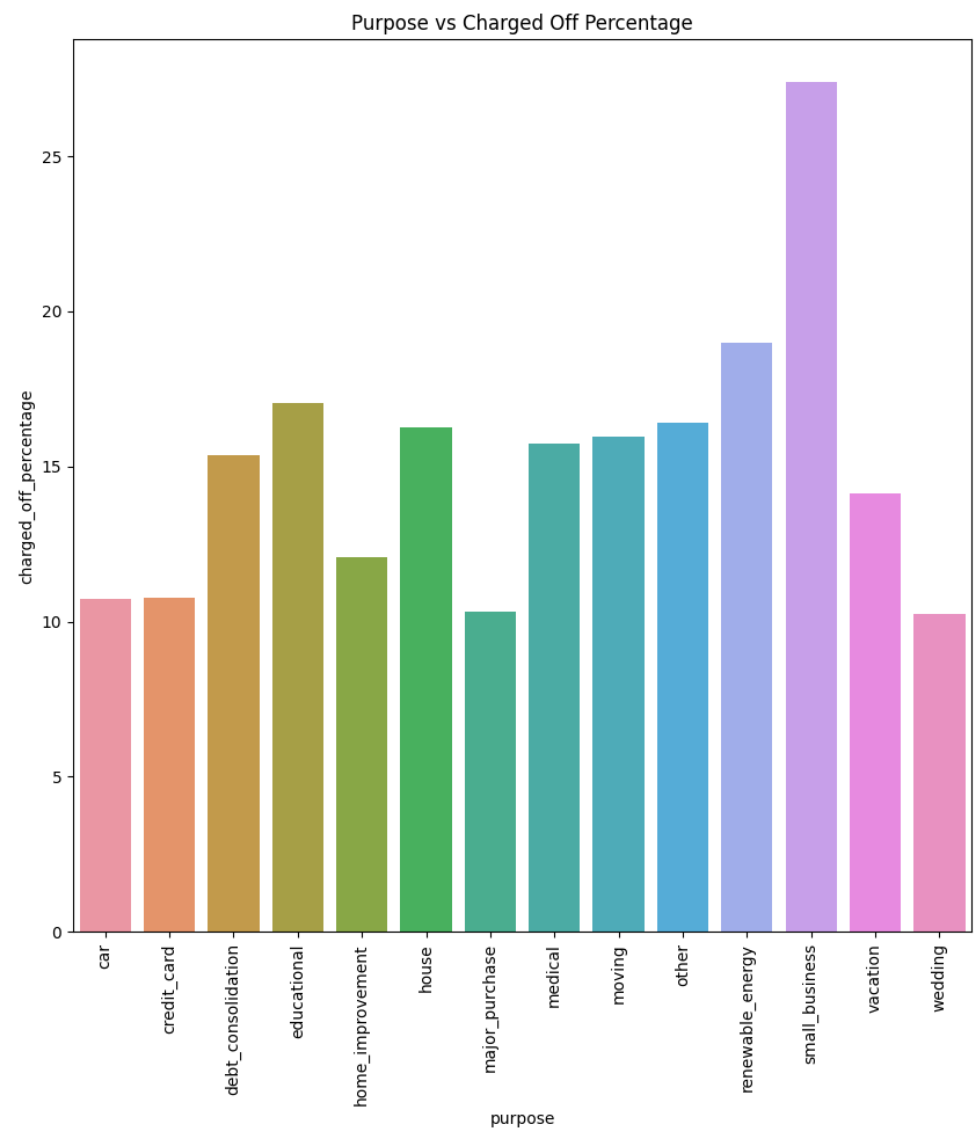**Bivariate Analysis with Target Variable - Observations**

**Numerical Variables**

- Loan Amount: Higher loan amounts (above 15K-20K) are associated with a higher charged off percentage.

- Annual Income: Lower annual income (below 90K) is associated with a higher charged off percentage.

- Interest Rate: Higher interest rates are associated with a higher charged off percentage.

- Debt to Income Ratio: Higher debt to income ratio is associated with a higher charged off percentage.

- Revolving Utilization: Higher revolving utilization (above 80%) is associated with a higher charged off percentage.

# BIVARIATE ANALYSIS

**Bivariate Analysis with Target Variable - Observations**

**Categorical Variables**

- Grade and Sub Grade: Higher grade and sub grade (e.g., Grade G) are associated with a higher charged off percentage.

- Home Ownership: 'Other' home ownership type has the highest charged off percentage.

- Purpose: Small business, renewable energy, and educational loans have the highest charged off percentages.

- Issue Month: December and May have high charged off percentages, while January to April have low charged off percentages.

- Verification Status: 'Verified' borrowers have a higher charged off percentage.

- Term: 60-month term loans have a higher charged off percentage compared to shorter-term loans.
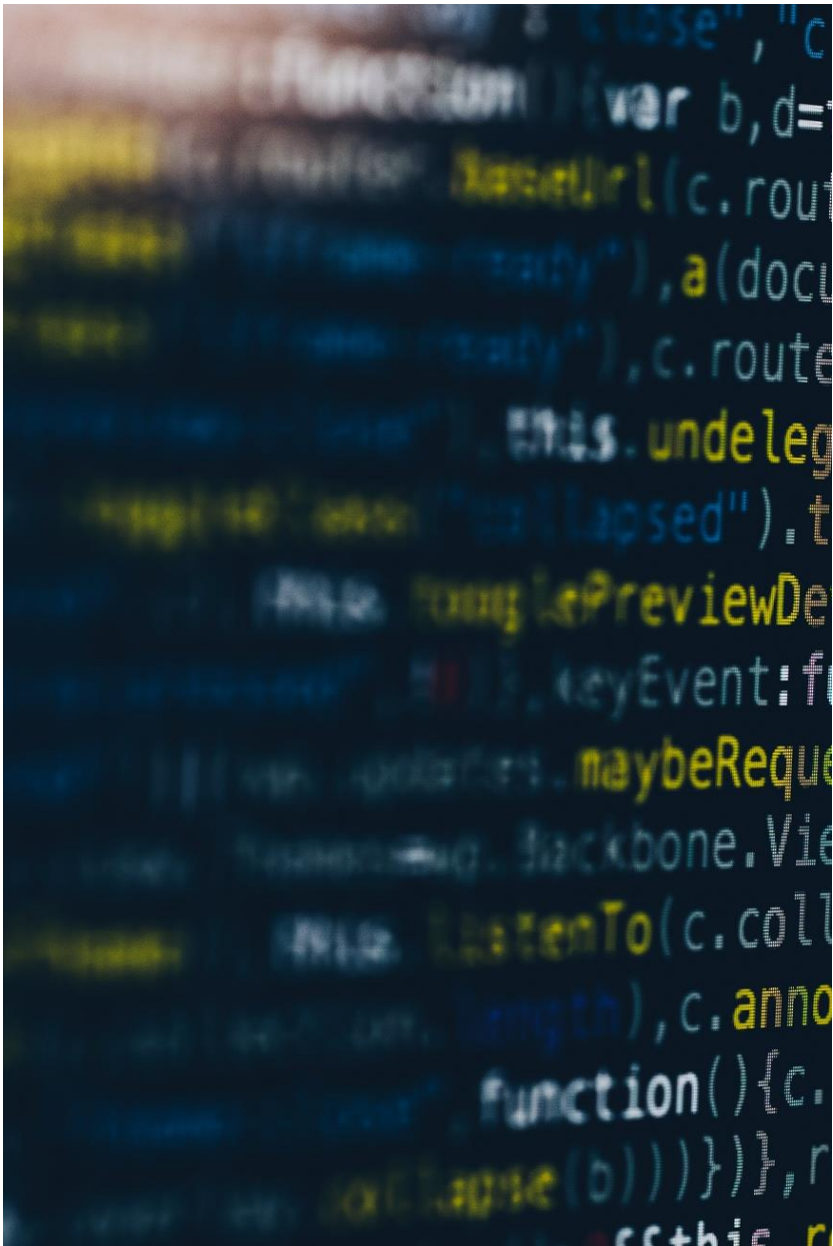
# CONCLUSION

The case study identified various contributing factors. Increased default rates due to high loan amounts, lower annual income, higher interest rates, and heightened borrower risk in terms of grade, subgrade, debt-to-income ratio, revolving utilization, etc. If lending institutions understand these factors, they can take better decisions by evaluating borrower risk, minimizing charged-off loans and improving loan portfolio performance.

The findings can further help lending institutions tailor their loan products and strategies to serve borrowers better while mitigating risks. Financial organizations can create an environment that fosters long-term success for borrowers and lenders when these key factors are incorporated into the risk assessment process.