LENDING CLUB CASE STUDY

ML 62 batch

Anandh M

Varsha Kannan

OBJECTIVE

To perform Exploratory data analysis, on the provided bank customer data, to come up with the insights on the scenarios when a customer would default the loan.

STEPS FOLLOWED FOR EDA

DATA UNDERSTANDING	DATA CLEANING	UNIVARIATE ANALYSIS	BIVARIATE ANALYSIS	MULTIVARIATE ANALYSIS
Going through the given info from data dictionary	*Performing steps to create a better dataset*	*Analysis based on a single variable*	*Analysis based on comparison with double variable*	*Exploring relationships between the variables*

INSIGHTS

STEP 1: DATA SOURCING

- We are provided with the dataset by UPGRAD, loan.csv dataset and a data dictionary to go through and understand the column details from it.

- Totally there are 39717 rows, and 111 columns

- We can briefly see and understand that there are many columns with null values only, and few columns that require to be filled with missing values.

STEP 2: DATA CLEANING

1. We have removed all the columns containing the null values.

2. We have then found that there are columns containing only single values, so we removed it as it would not help in the analysis.

3. Then we calculated the percentage of null values present in each column of the dataset and removed the columns having null value percentage above 60%.

4. We dropped the columns the irrelevant columns that wouldn't provide us much help with our objective.

5. Dropped the rows containing loan status as "current", as our goal is to predict if in future the customer would default or not.

6. We have handled the missing values using median for numerical columns and mode for categorical columns

7. We have standardized the columns like int_rate, revol_util by removing the % symbol and converting it to a numerical value.

8. Finally we have removed the outliers by visualizing it using box plots and removing it using the percentile method.

STEP 3: UNIVARIATE ANALYSIS OBSERVATIONS

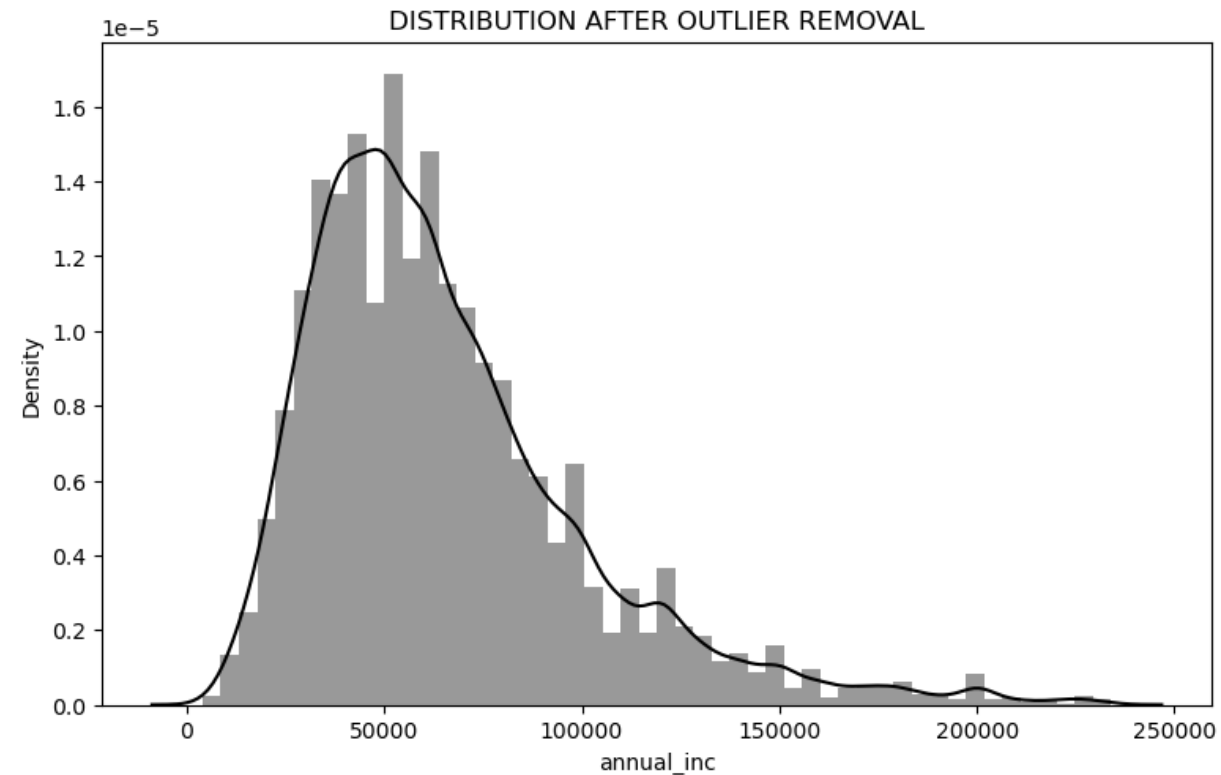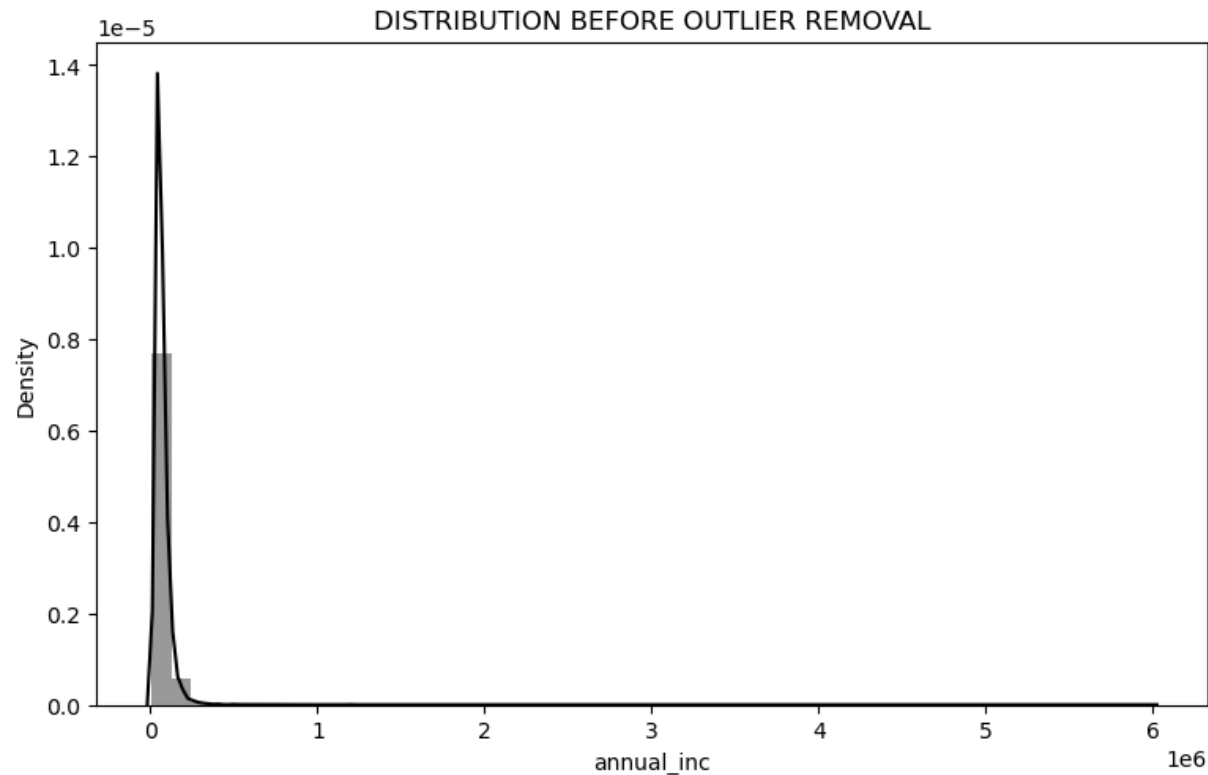We can see a high range of customers in the following scenarios:

- the feature "annual_inc" has outliers in the upper fence

- There is a huge jump between 99 and 100 % of data in "annual_inc" and that can be trimmed.
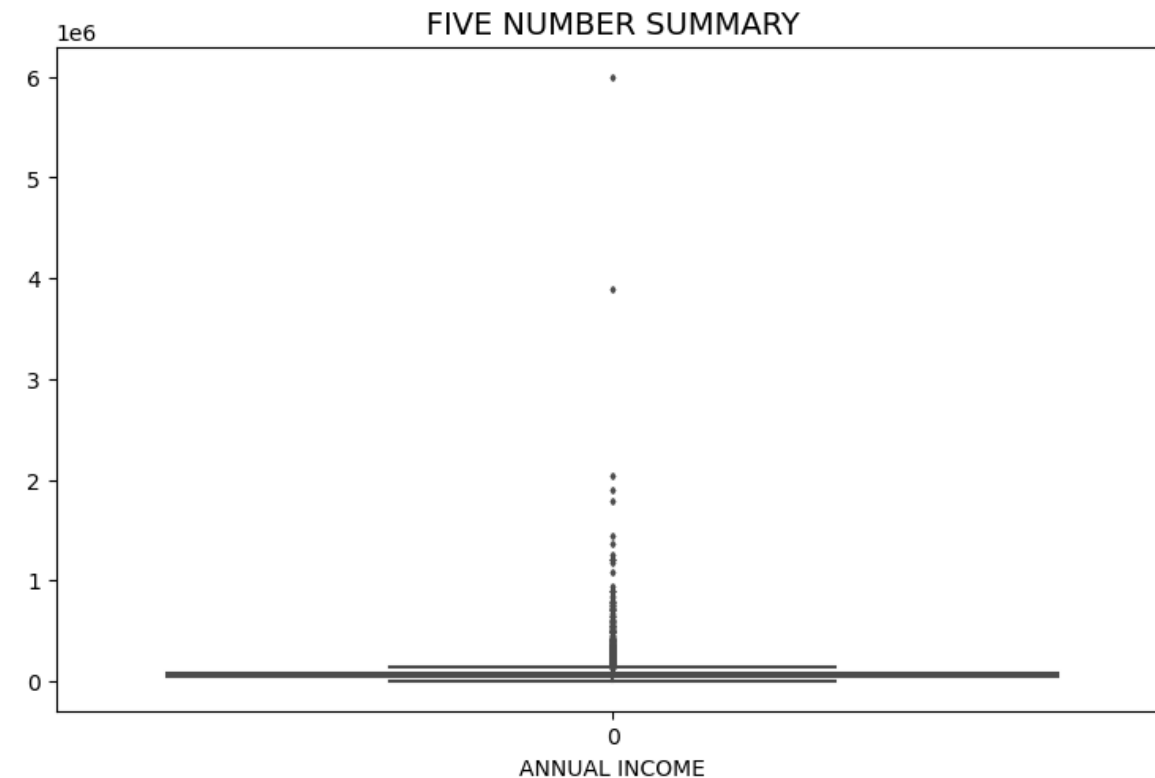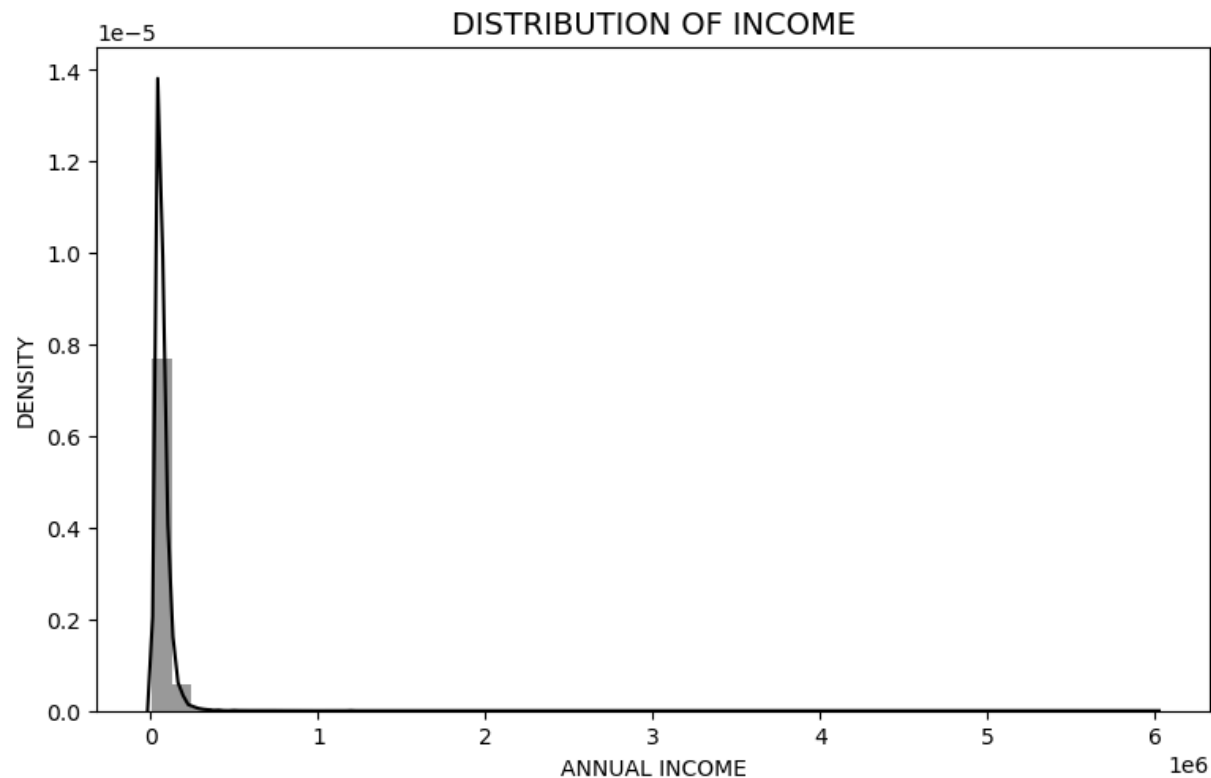
STEP 3: UNIVARIATE ANALYSIS OBSERVATIONS

We can see a high range of customers in the following scenarios:

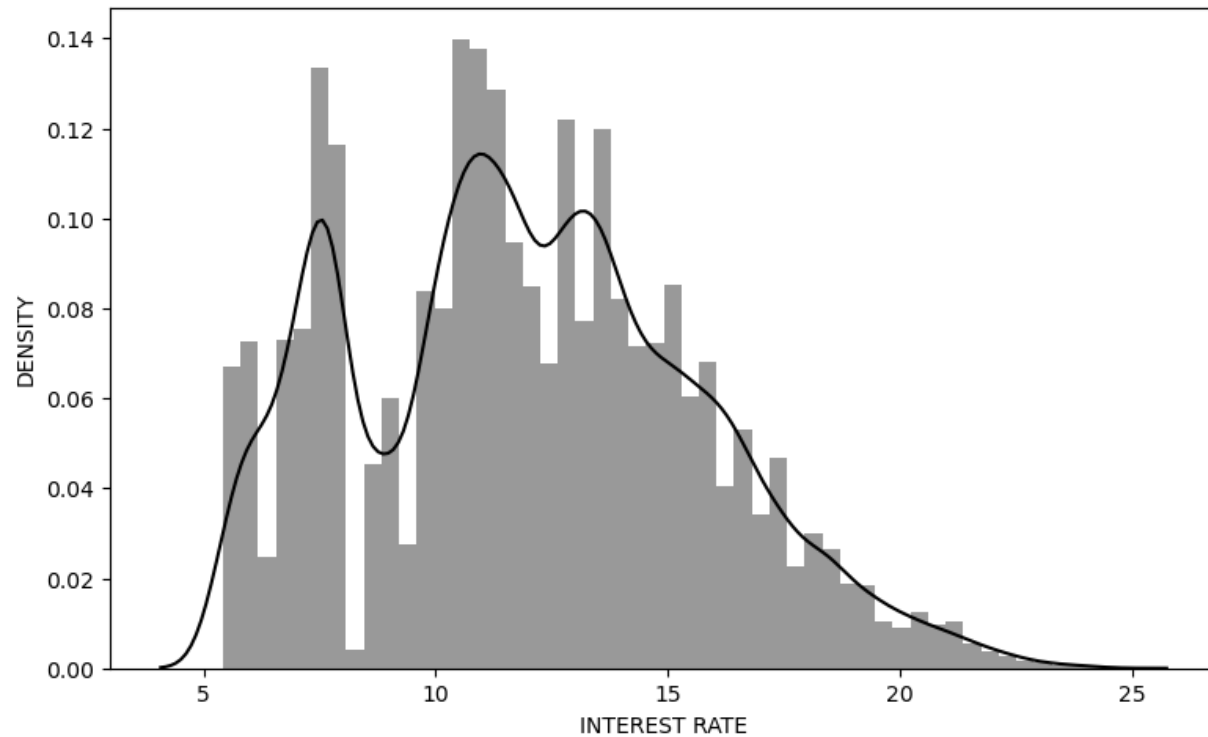- After outlier removal, we can see that the majority of borrower's income lies between 40k-80k range
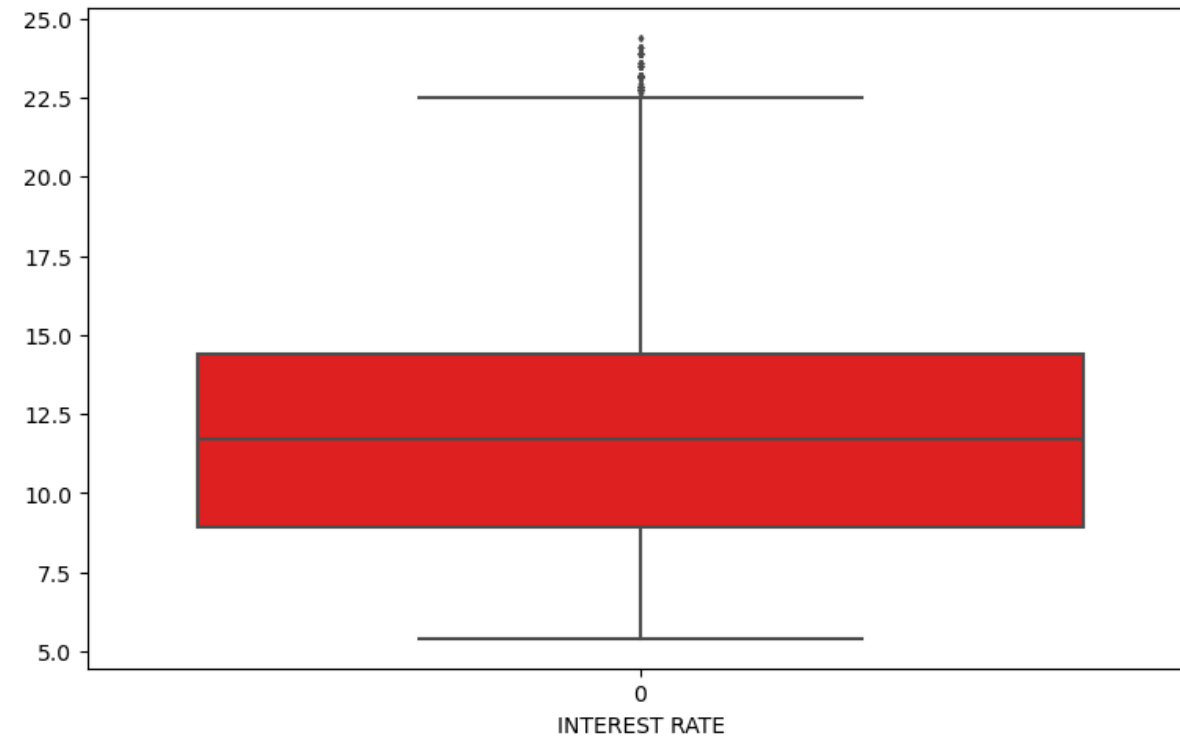
STEP 3: UNIVARIATE ANALYSIS OBSERVATIONS

We can see a high range of customers in the following scenarios:

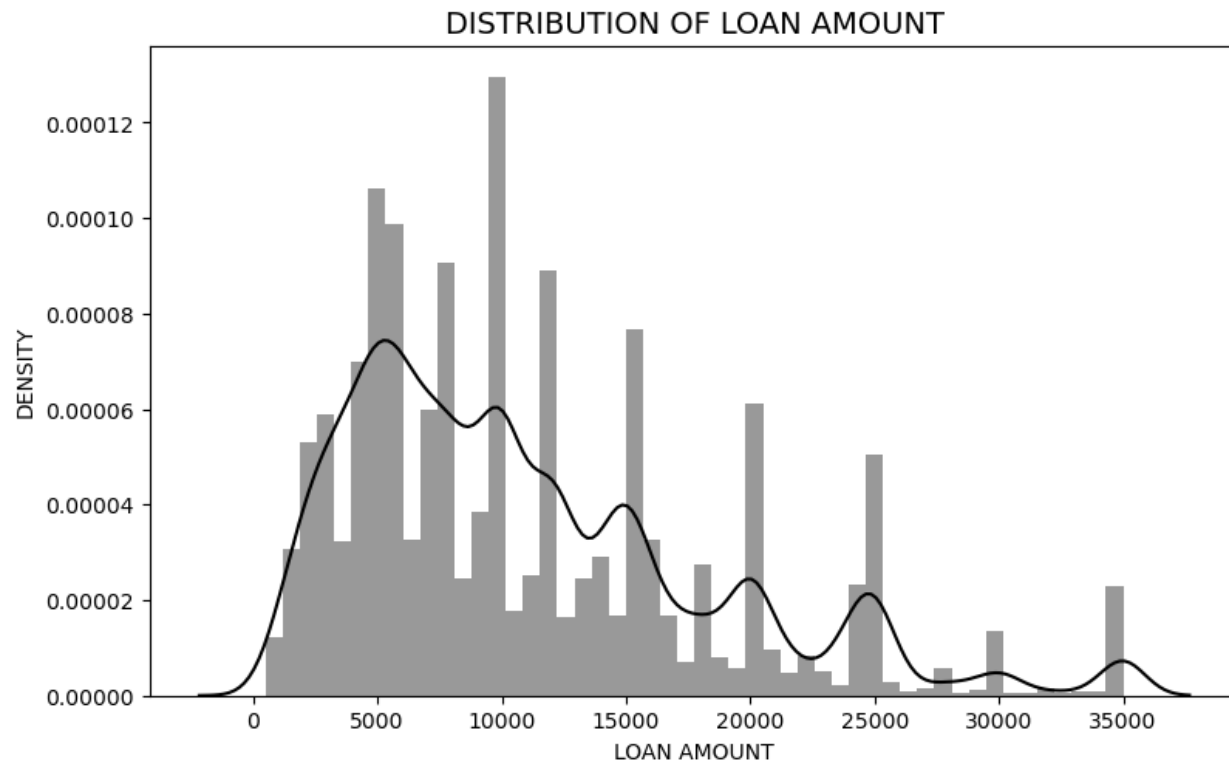- The average interest rate is 12% but there is huge jump after 75 percentile.

STEP 3: UNIVARIATE ANALYSIS OBSERVATIONS

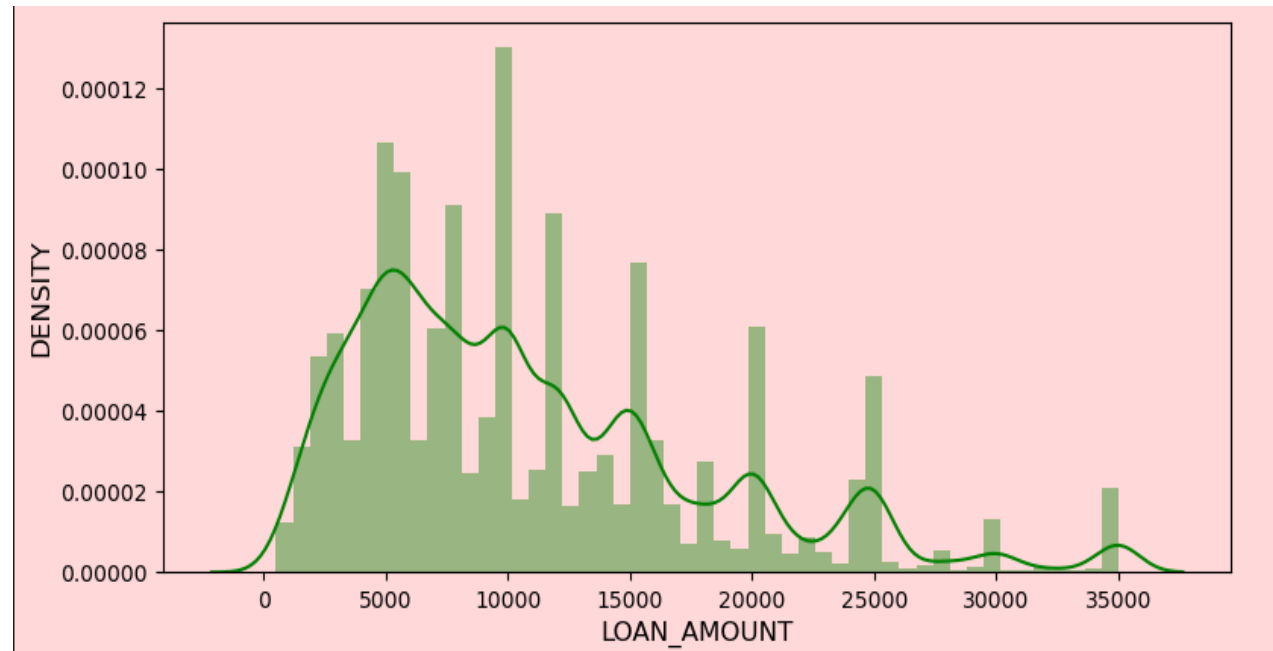We can see a high range of customers in the following scenarios:

- Most of the borrowers prefer to get loan amount between 5k to 15k.

STEP 3: UNIVARIATE ANALYSIS OBSERVATIONS

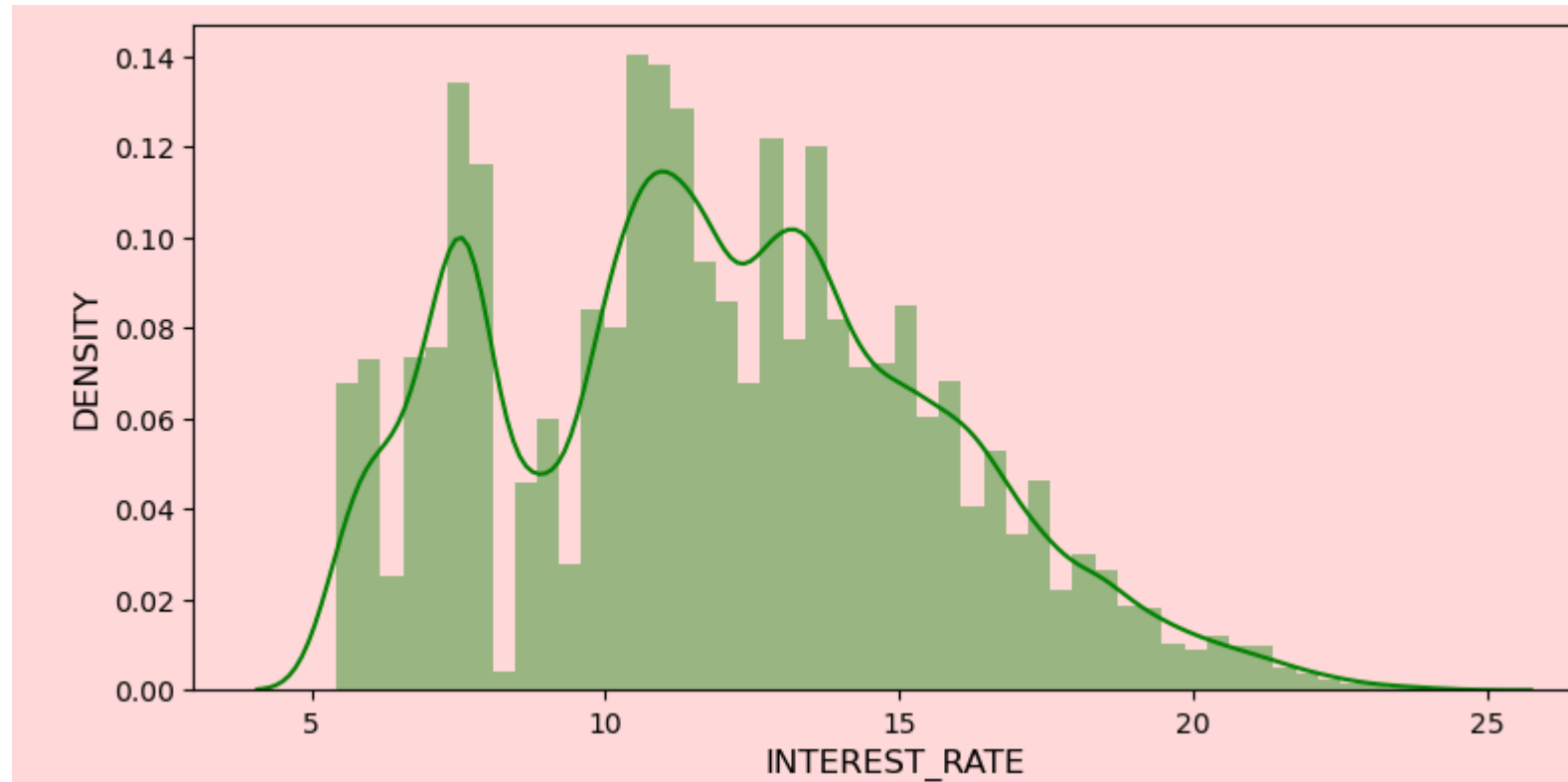We can see a high range of customers in the following scenarios:

- Most of the borrowers prefer to get loan amount between 5k to 15k.

STEP 3: UNIVARIATE ANALYSIS OBSERVATIONS

We can see a high range of customers in the following scenarios:
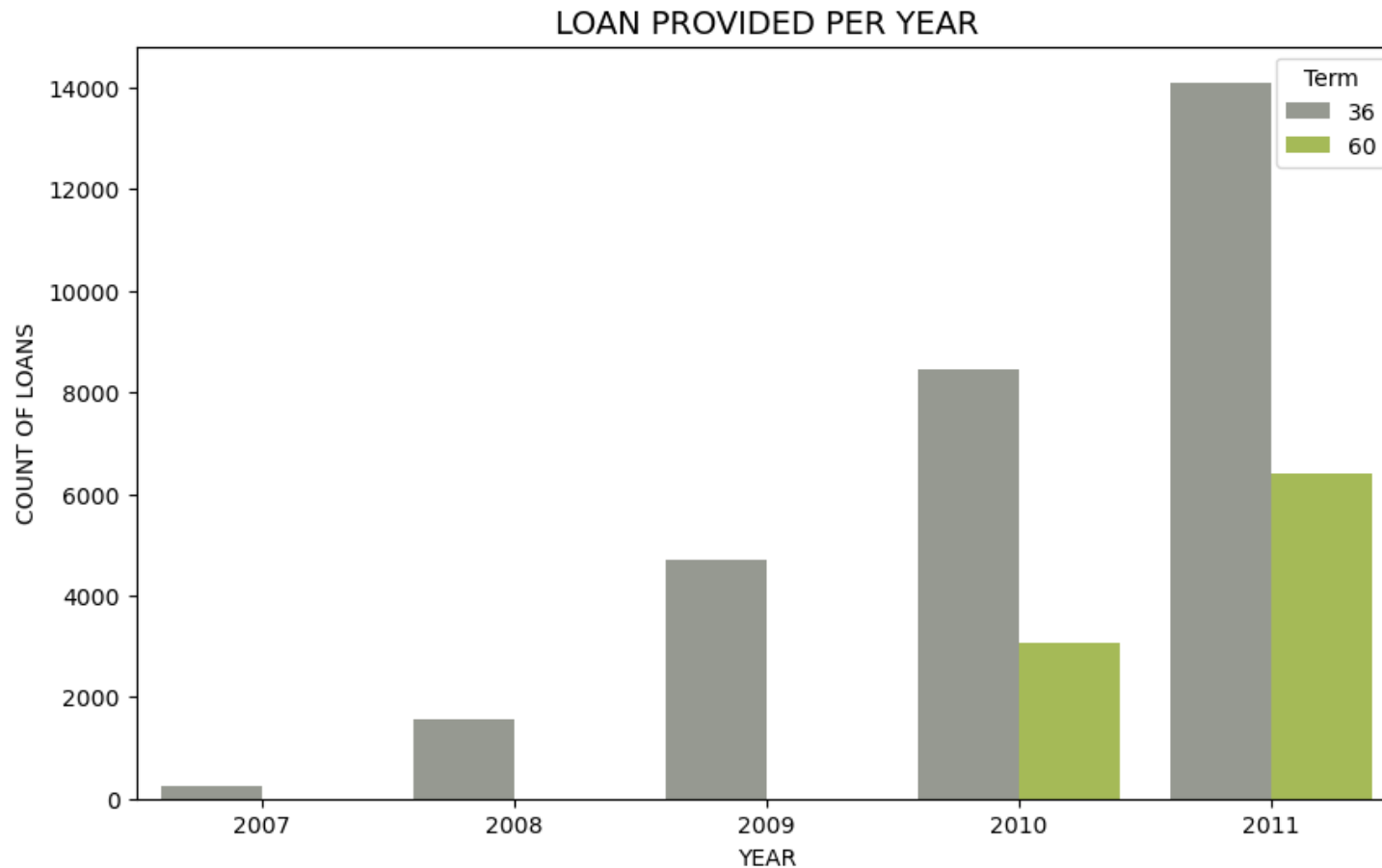
- Almost of the borrowers gets interest rate between 9% to 14%.

STEP 3: UNIVARIATE ANALYSIS OBSERVATIONS

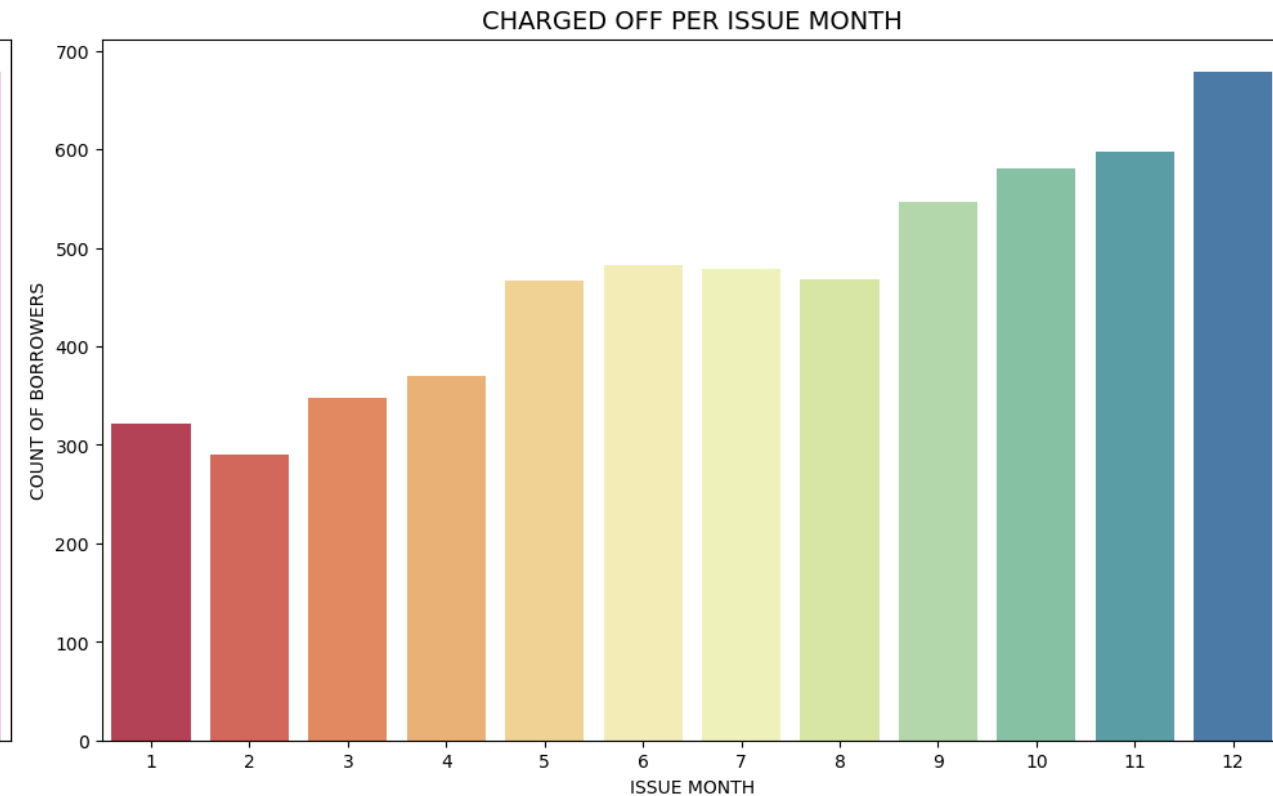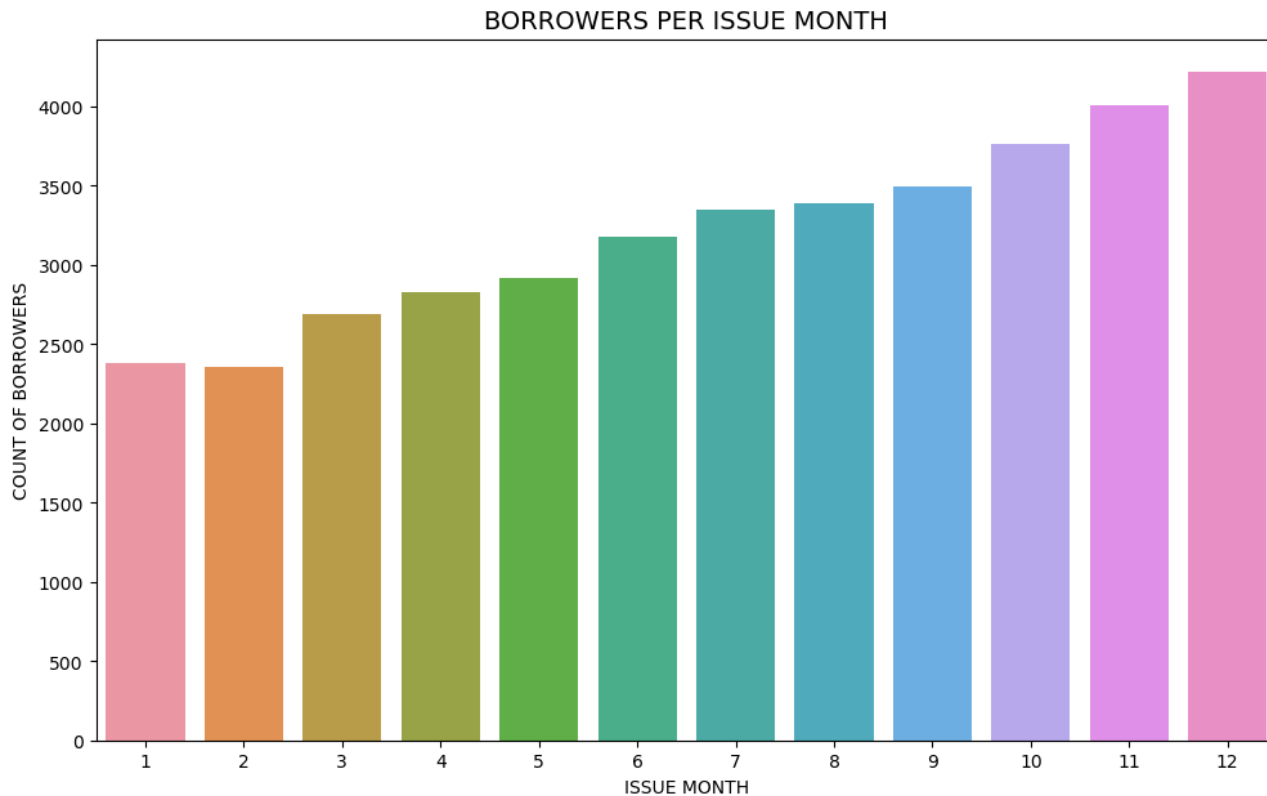We can see a high range of customers in the following scenarios:

- For each year number of borrowers keeps on increasing and they opt for 36 months tenure.

LOAN PROVIDED PER YEAR

STEP 3: UNIVARIATE ANALYSIS OBSERVATIONS

We can see a high range of customers in the following scenarios:
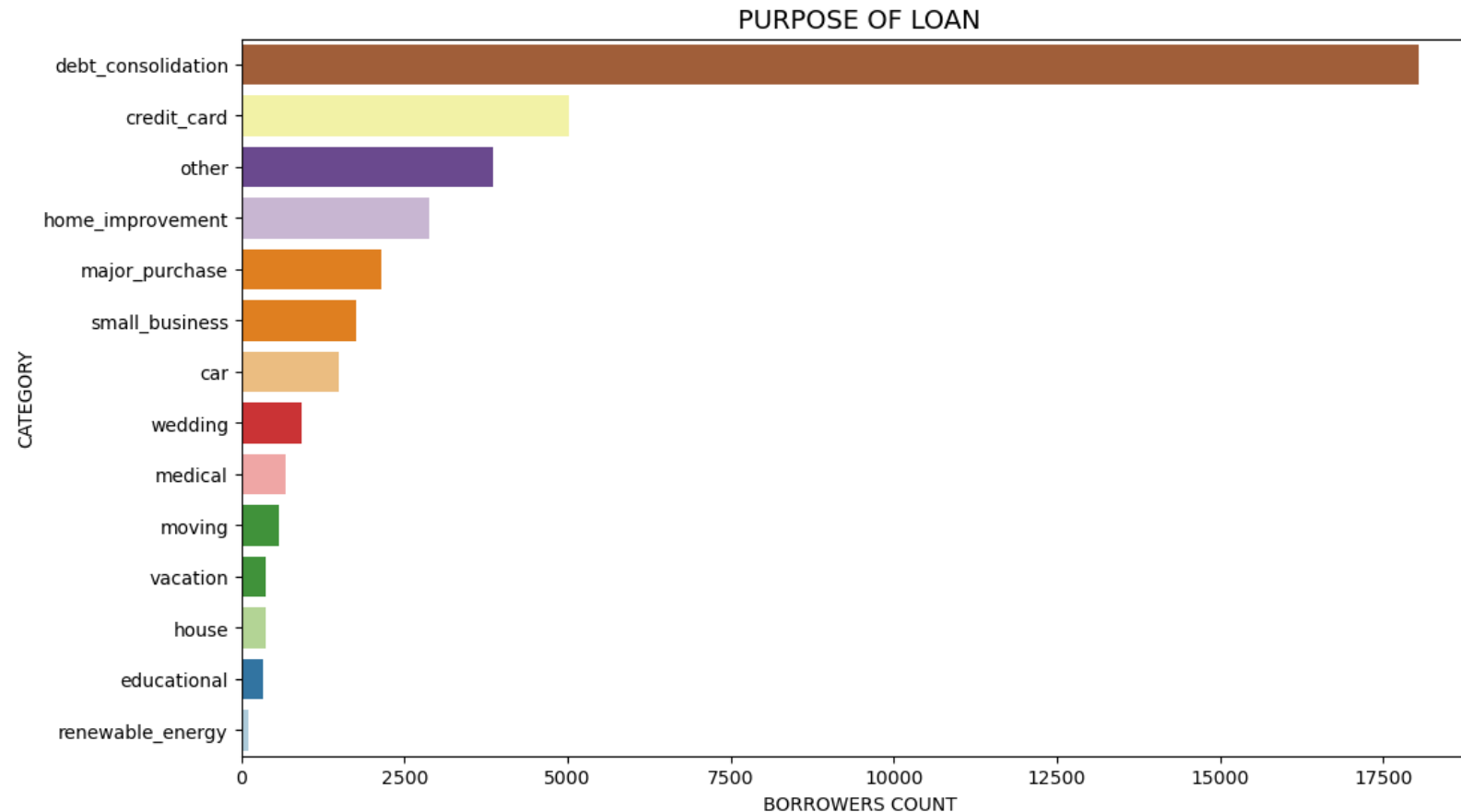
- Loans approved on Q4 are more likely to default.

STEP 3: UNIVARIATE ANALYSIS OBSERVATIONS

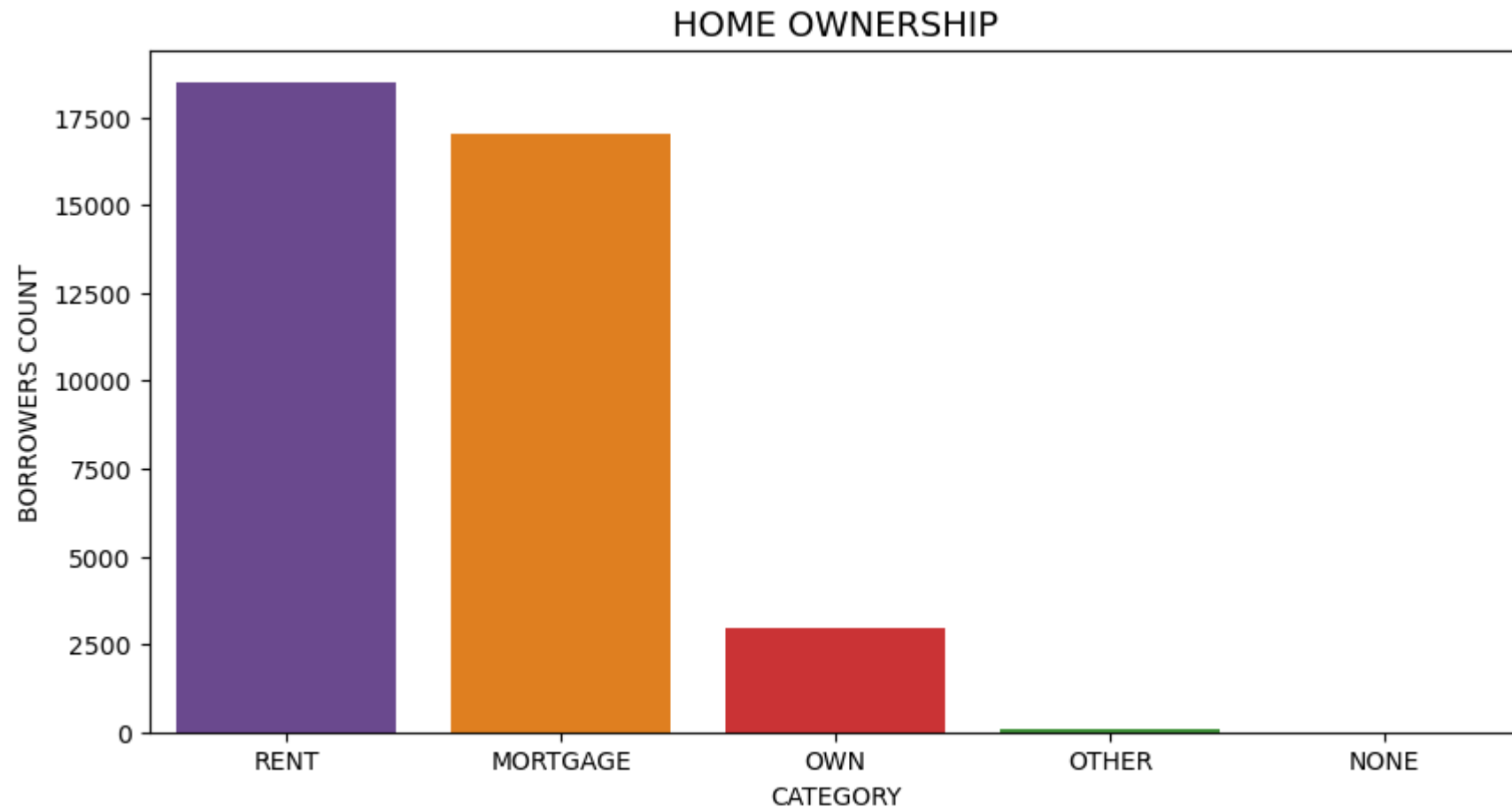We can see a high range of customers in the following scenarios:

- Most of the borrowings happens for debt consolidation and credit card bill payment.

PURPOSE OF LOAN

STEP 3: UNIVARIATE ANALYSIS OBSERVATIONS

We can see a high range of customers in the following scenarios:
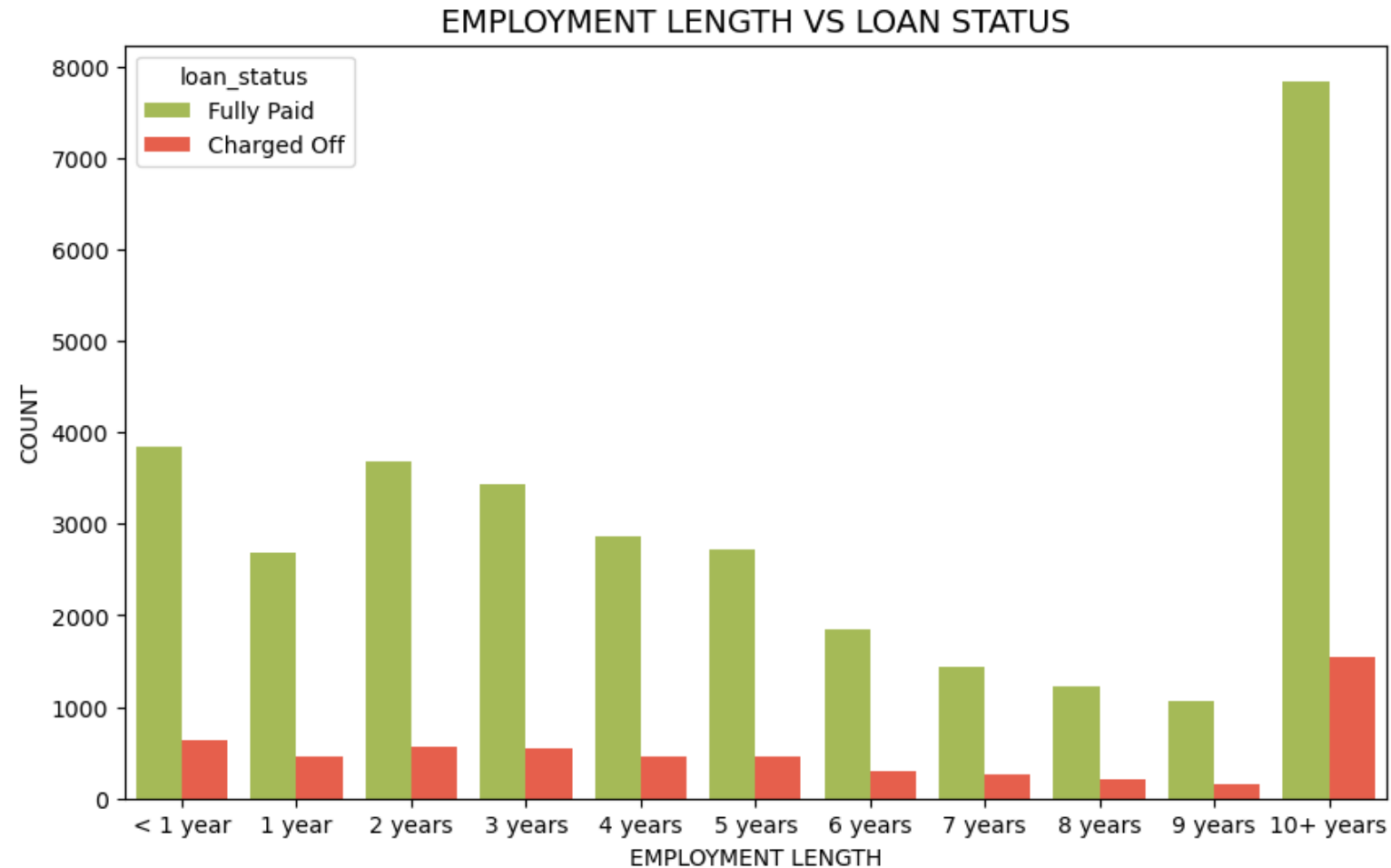
- Very least of the borrowers has a own house.

HOME OWNERSHIP

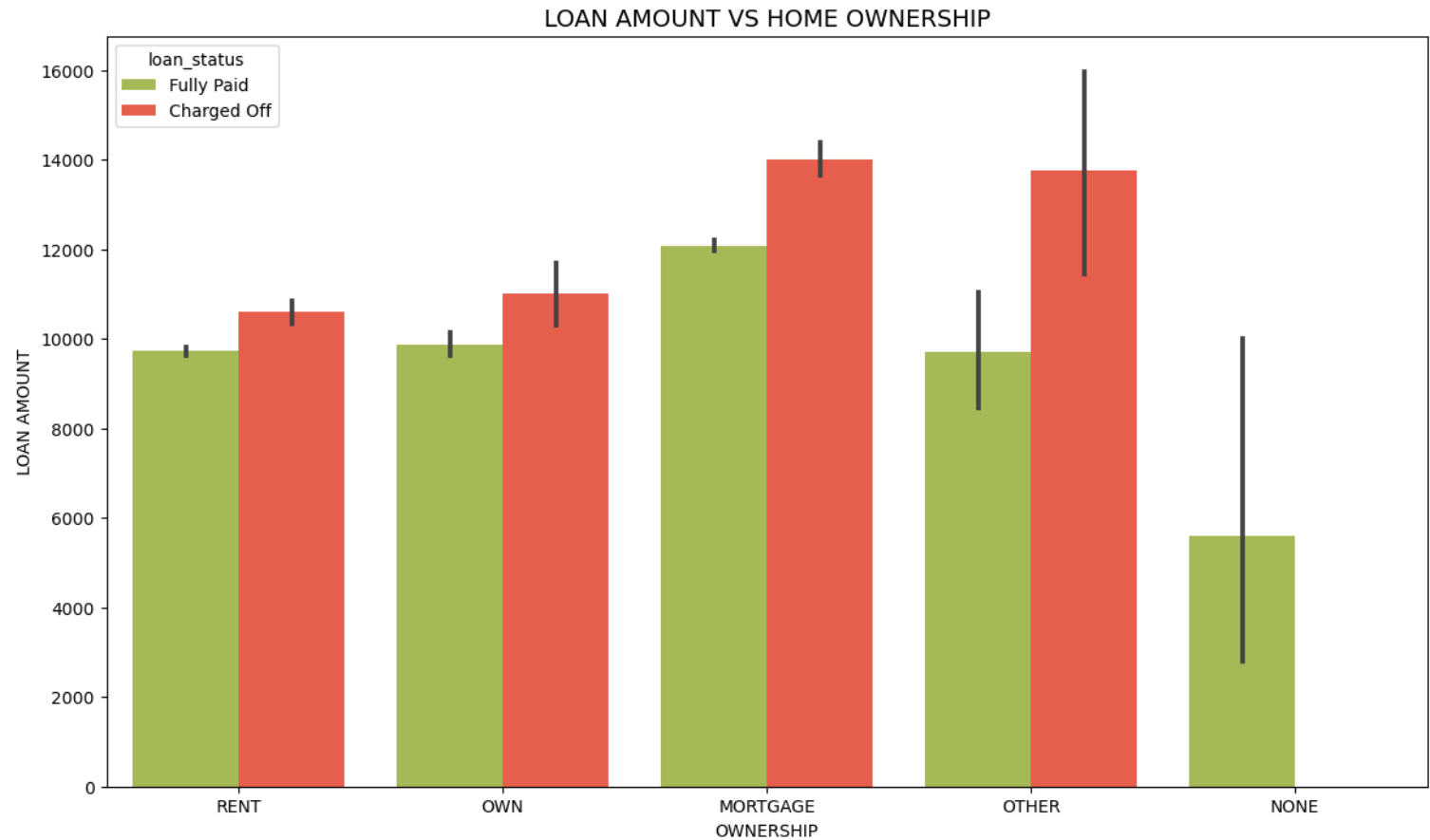# STEP 4: BIVARIATE ANALYSIS OBSERVATIONS

EMPLOYMENT LENGTH VS LOAN STATUS

- Borrowers with 10+ years of employment length are more likely to charge off.

EMPLOYMENT LENGTH VS LOAN STATUS

STEP 4: BIVARIATE ANALYSIS OBSERVATIONS

HOME OWNERSHIP VS LOAN AMOUNT

- Borrowers living on mortgage with loan amount greater than 12k are likely to charge off.

LOAN AMOUNT VS HOME OWNERSHIP

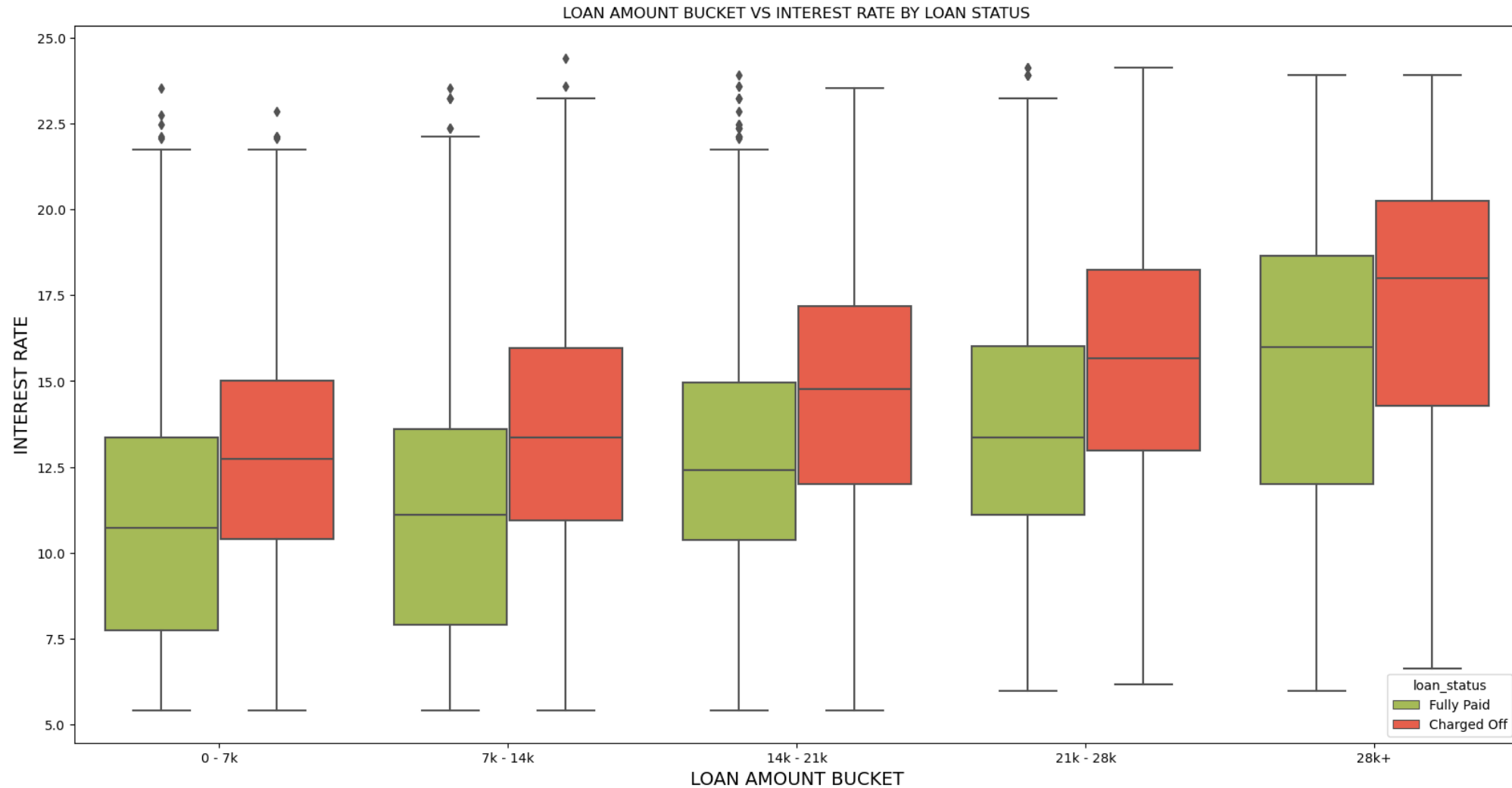# STEP 4: BIVARIATE ANALYSIS OBSERVATIONS

LOAN AMOUNT VS PURPOSE

- Loan for debt consolidation, credit card, small business with average loan amount greater than 12000 are more likely to default.

LOAN AMOUNT (DEFAULT) VS PURPOSE
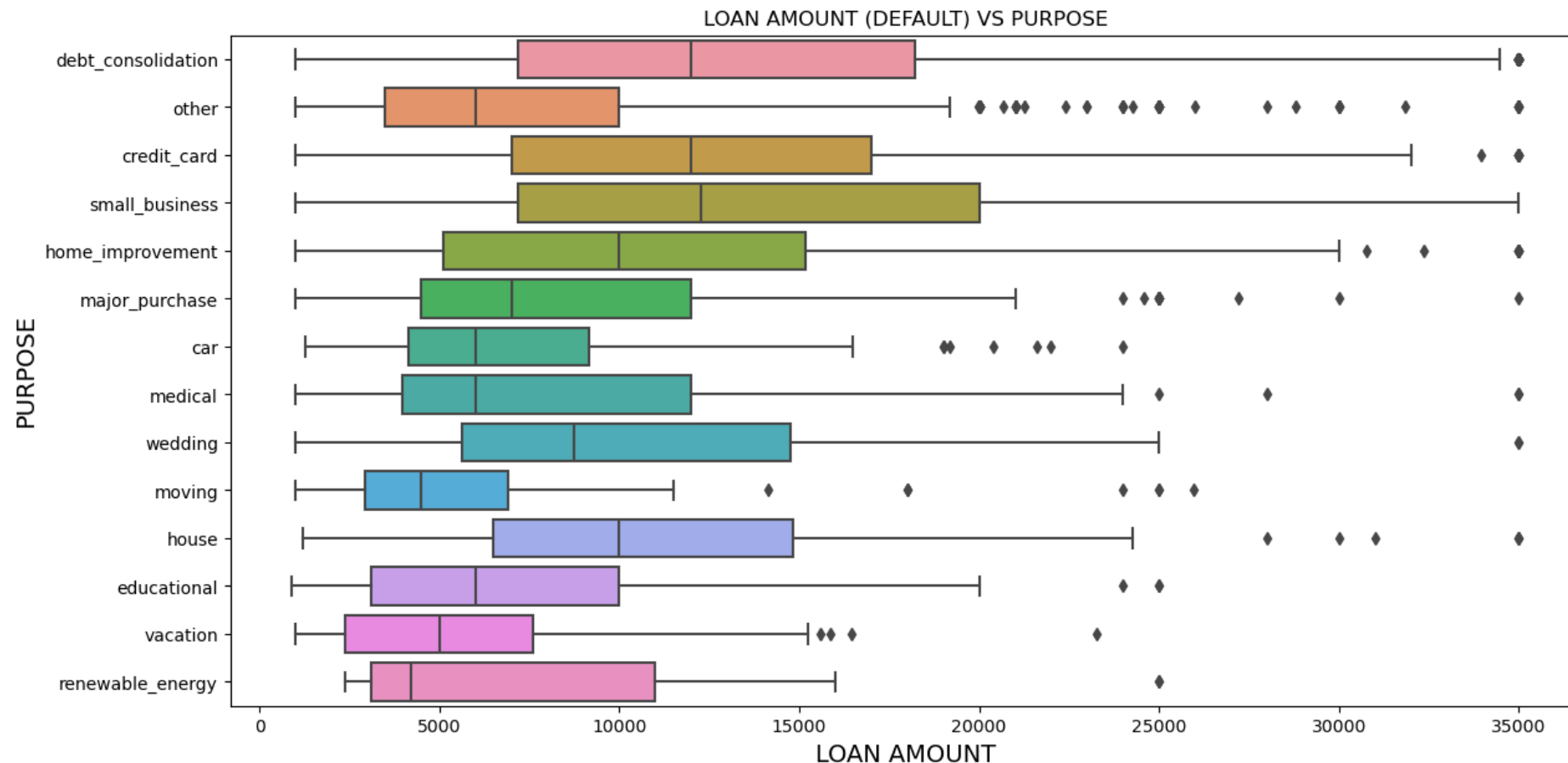
STEP 4: BIVARIATE ANALYSIS OBSERVATIONS

loan amount bucket vs interest rate

- Loan with highest interest rate in all buckets are more likely to charge off.

LOAN AMOUNT BUCKET VS INTEREST RATE BY LOAN STATUS

STEP 4: BIVARIATE ANALYSIS OBSERVATIONS

interest rate vs purpose:
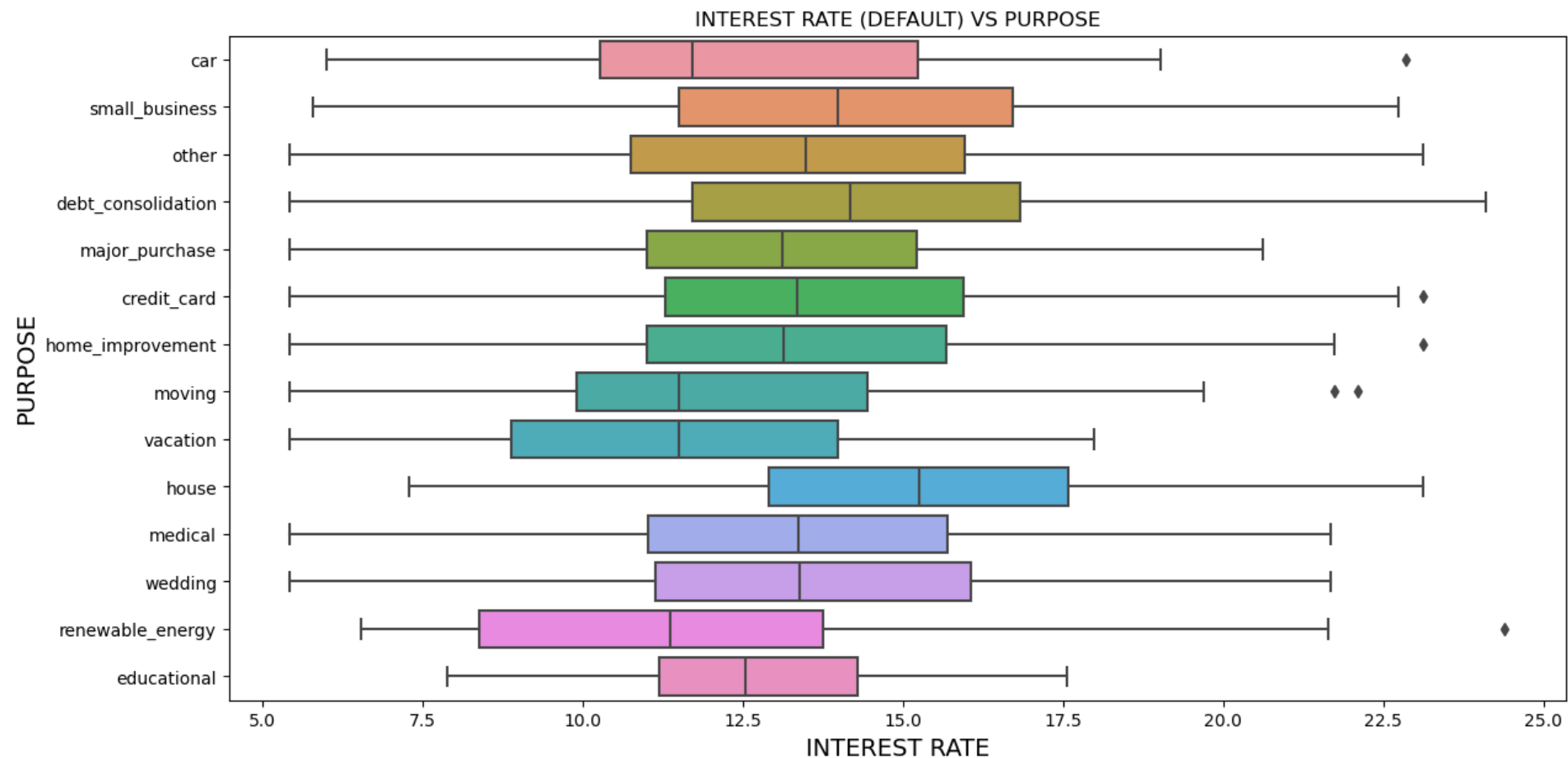
- Loan provided for house, debt consolidation, small business with highest interest rate are more likely to charge off.

STEP 4: BIVARIATE ANALYSIS OBSERVATIONS

INTEREST RATE VS TERM:

- Longer tenure has the higher interest rate and people in this term with interest rate between 15-20% is more likely to charge off. Shorter tenure has the lower interest rate and people in this term with interest rate between 10-15% is more likely to charge off.
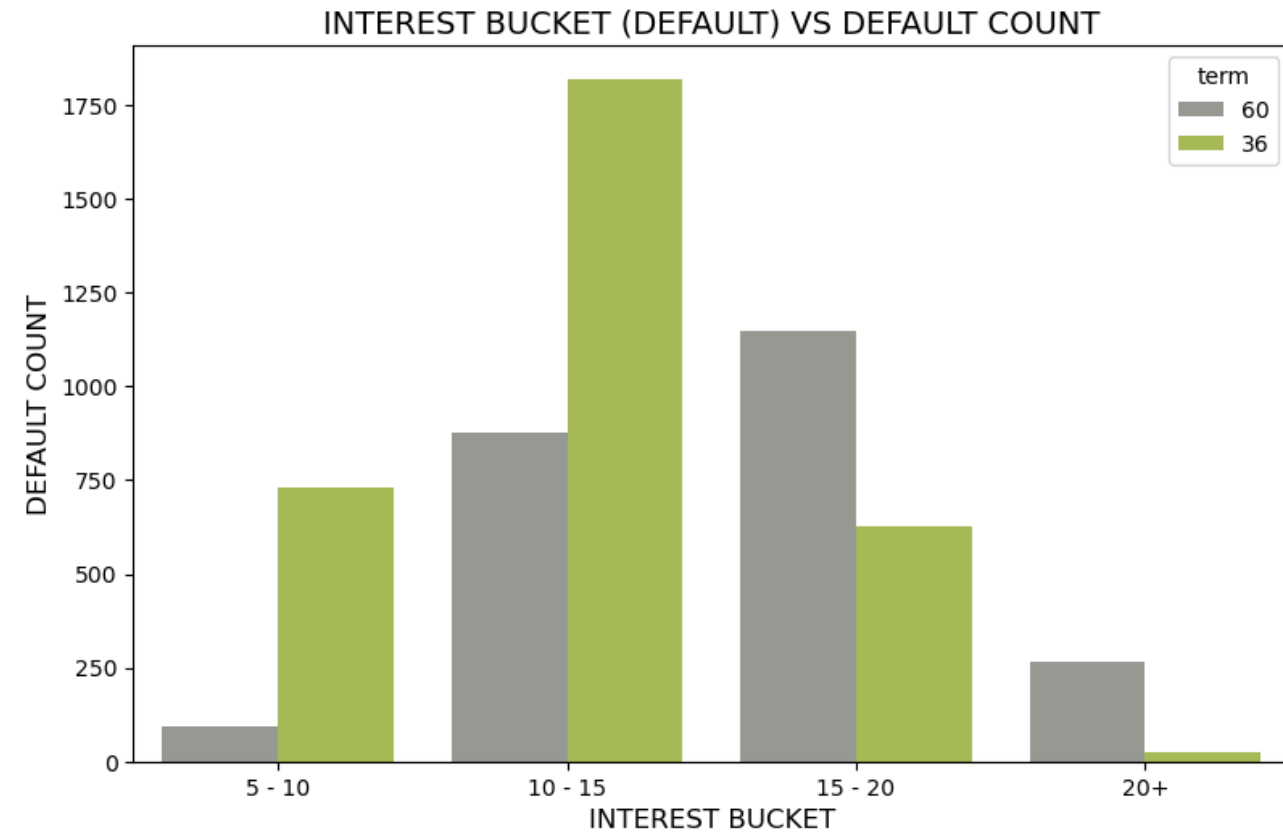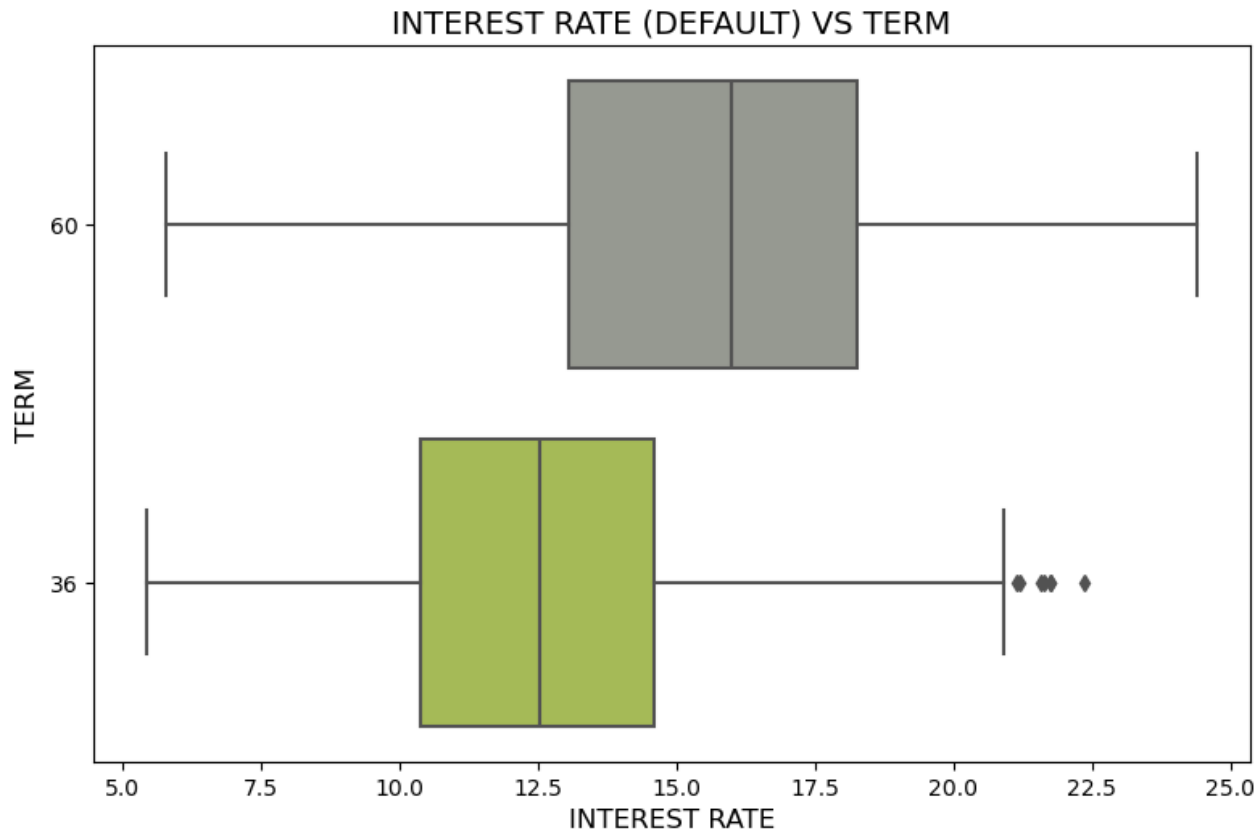
STEP 4: BIVARIATE ANALYSIS OBSERVATIONS
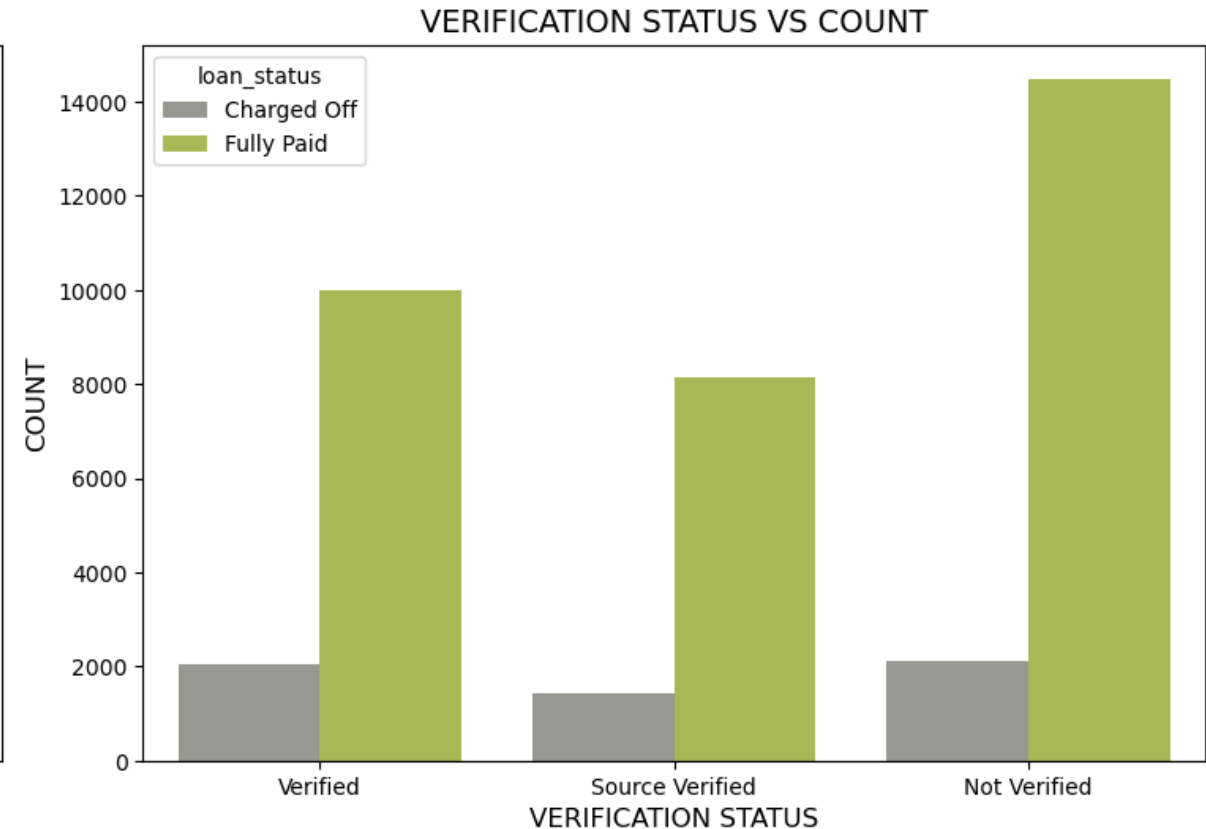
LOAN AMOUNT VS GRADE

- Borrower within grade E,F,G and with average loan amount greater than 15000 are more likely to charge off.

LOAN AMOUNT (DEFAULT) VS GRADE
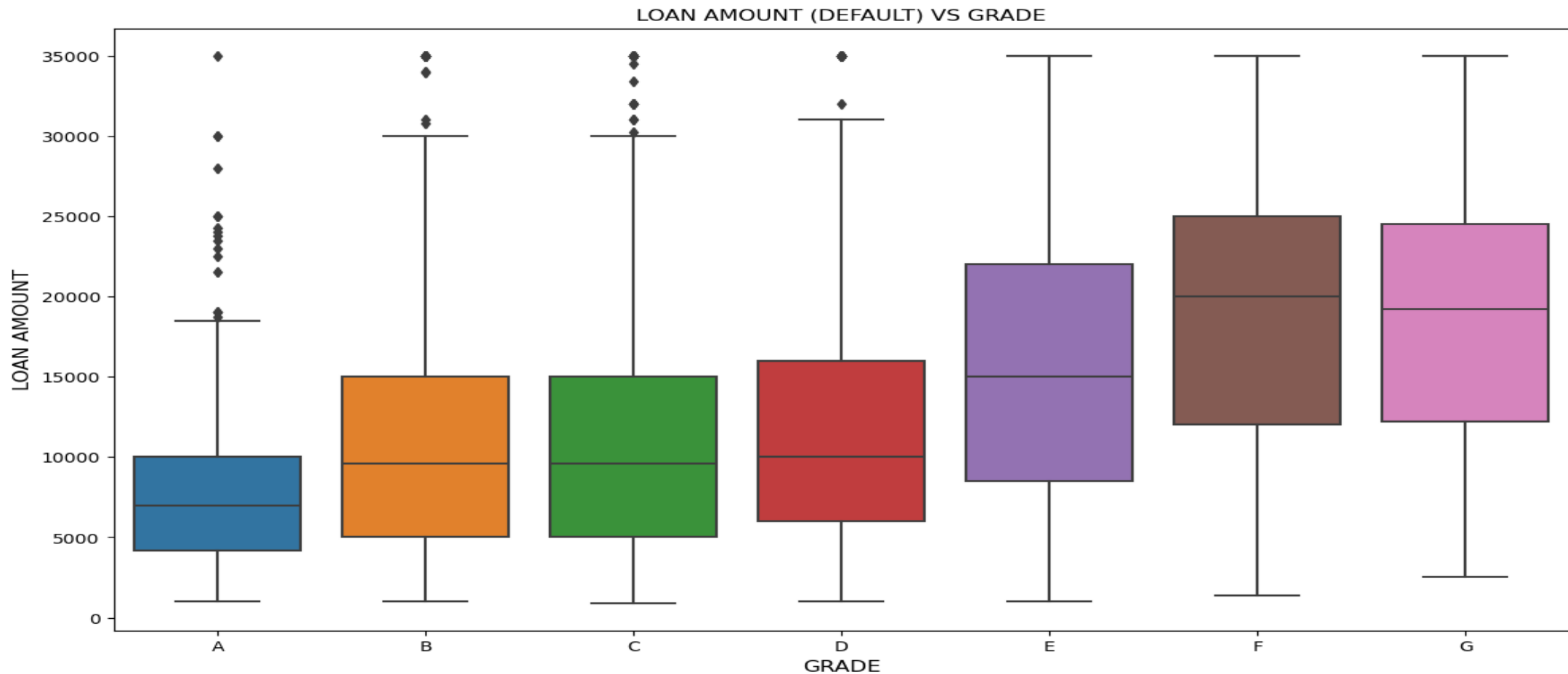
STEP 4: BIVARIATE ANALYSIS OBSERVATIONS
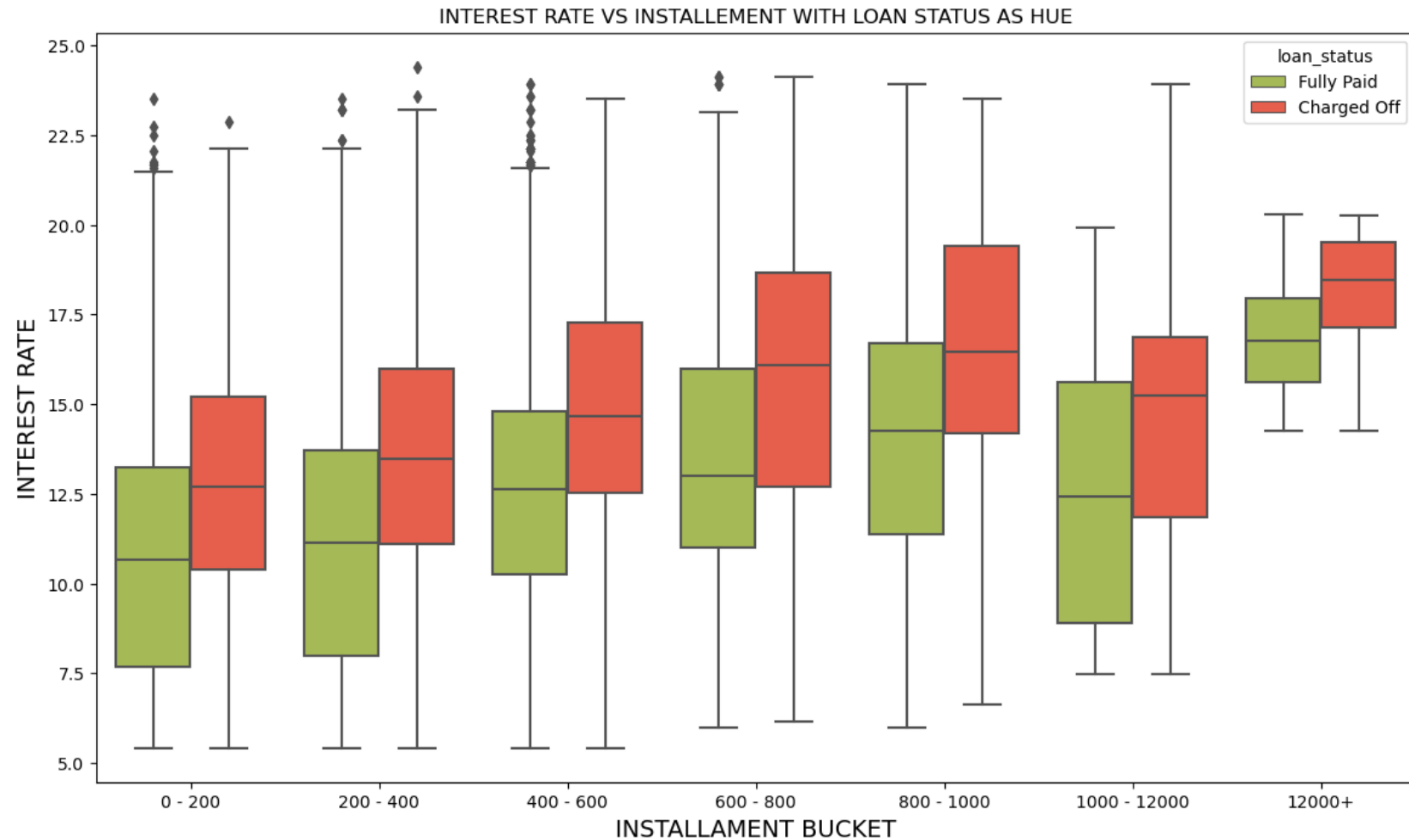
LOAN AMOUNT VS VERIFICATION STATUS

- Borrowers with verification status "not-verified" has loan amount with 5k to 12.5k are more likely to charge off.

- Borrowers with verification status "not-verified" and "verified" has a equal proportion in charge off
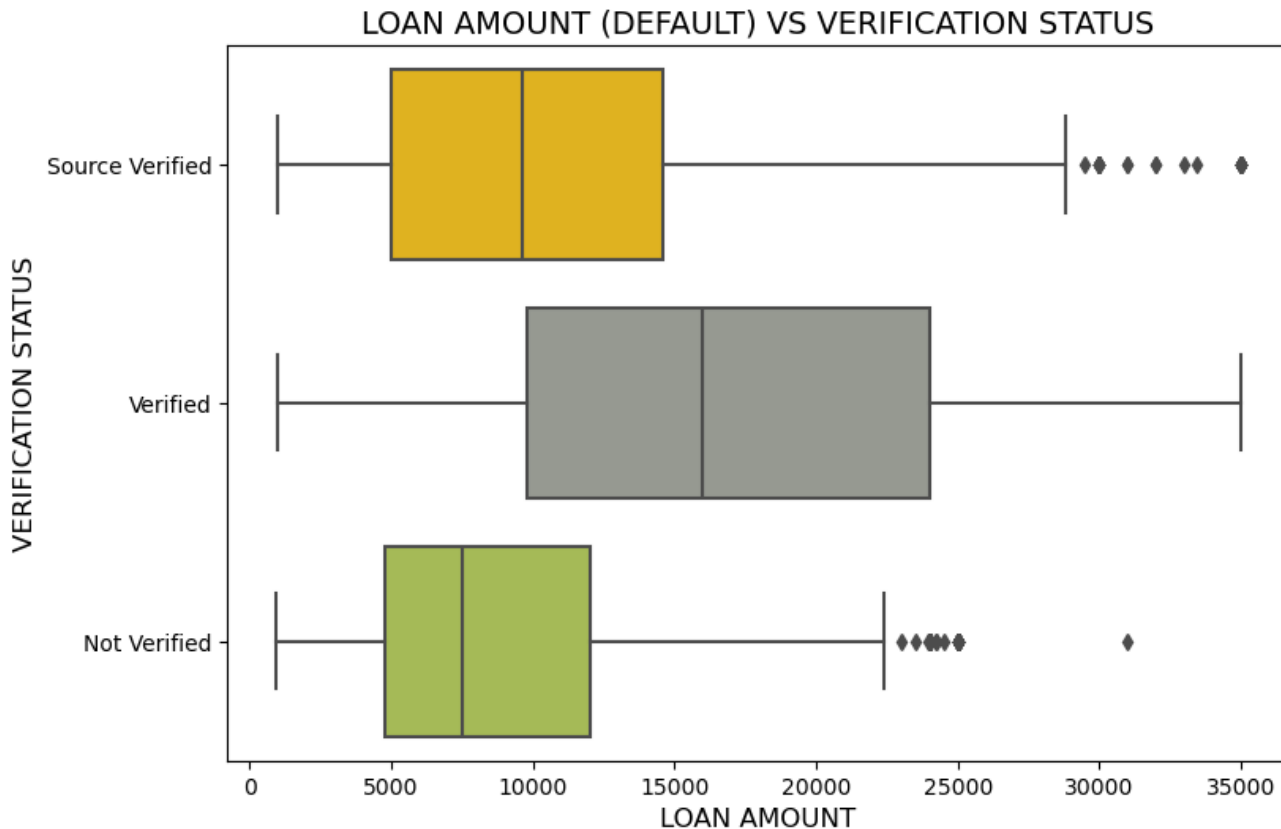
STEP 4: BIVARIATE ANALYSIS OBSERVATIONS

Loan amount vs employment period:

- As the number of installment increase the borrowers with high interest rate are more likely to charge off.

INTEREST RATE VS INSTALLEMENT WITH LOAN STATUS AS HUE

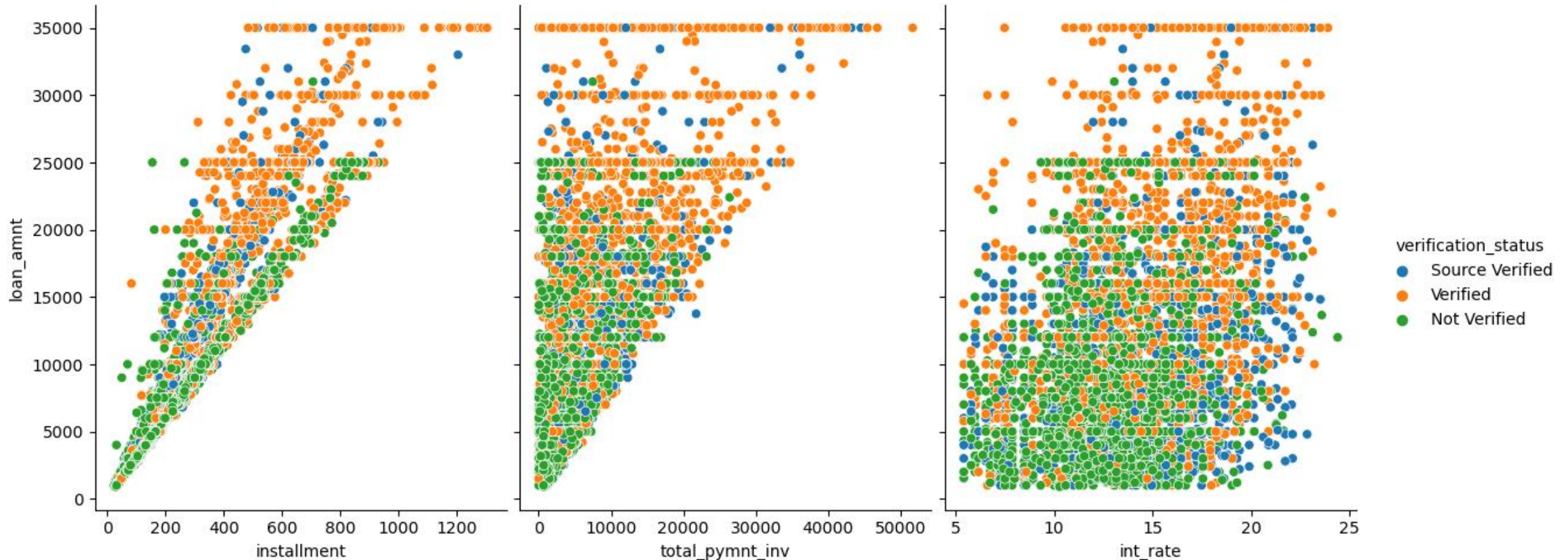# STEP 5: MULTIVARIATE ANALYSIS OBSERVATIONS

Correlation heatmap

- loan_amnt has strong positve correlation with installment, total_payment_inv.
- loan_amnt has intermediate positive correlation with annual_inc, int_rate, total_acc.
- open_acc has strong positive correlation with total_acc.
- pub_rec has strong positve correlation with pub_rec_bankruptcies.
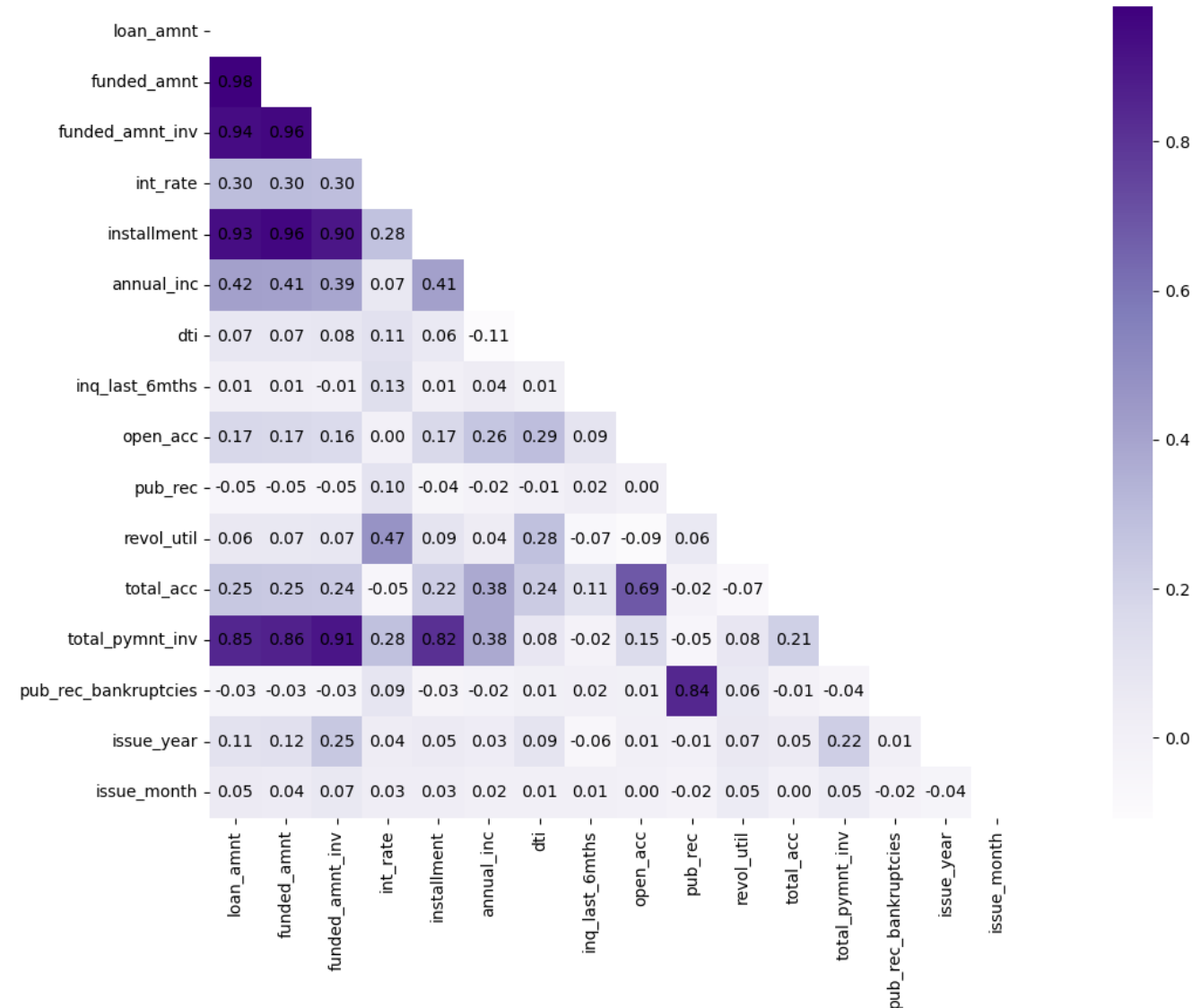
STEP 5: MULTIVARIATE ANALYSIS OBSERVATIONS

Plotting a pair plot for 'installment', 'total_pymnt_inv', 'int_rate' vs loan amount by verification status as hue

- When the loan amount increases, number of installment, total_pymnt_inv, int_rate also increases.
- Borrowers with verification status "Not-Verified" are most likely to get loan less than 25k (Charge off).
- Borrowers with verification status "Verified" are most likely to get loan greater than 25k (Charge off).

STEP 5: MULTIVARIATE ANALYSIS OBSERVATIONS

Plotting a pair plot for 'installment', 'total_pymnt_inv', 'int_rate' vs loan amount by term as hue

- Borrowers with lower term has higher installments comparatively.
- Borrowers with 60 months term has highest share to total_pymnt_inv.
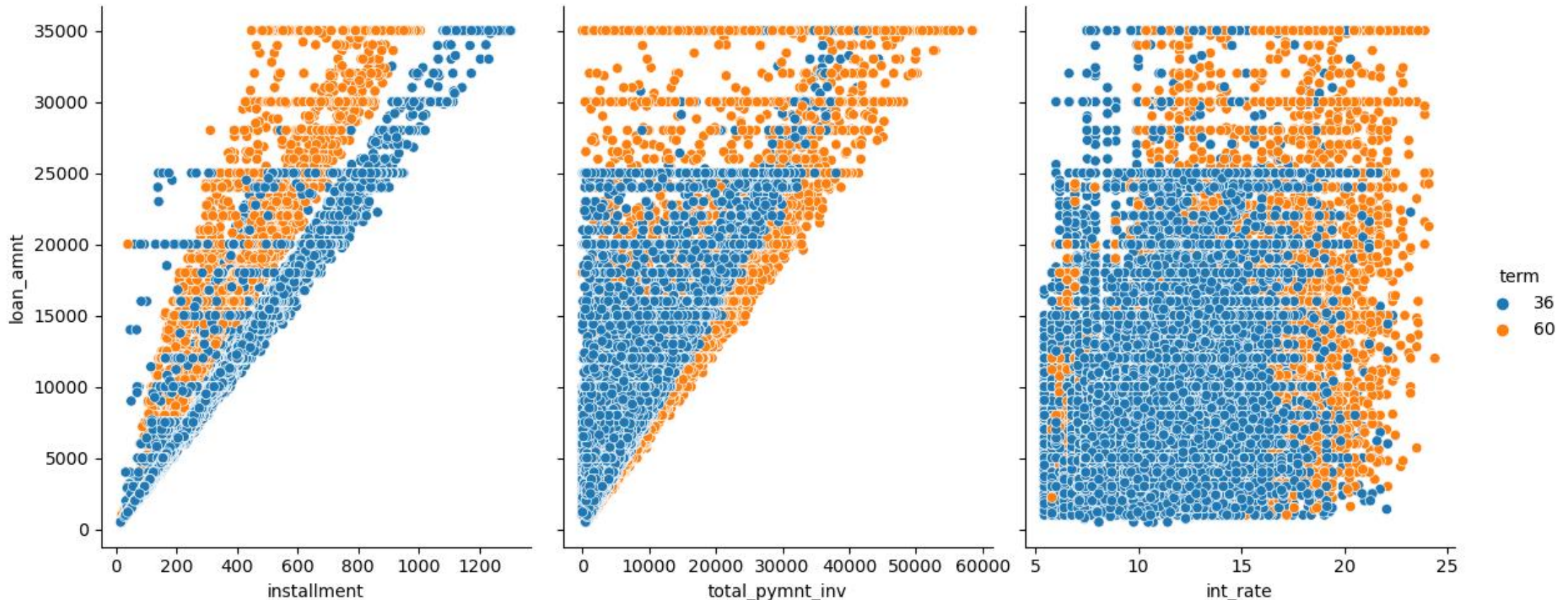
STEP 5: MULTIVARIATE ANALYSIS OBSERVATIONS

Plotting a swarm plot for home ownership vs loan amount by term as hue

- Across all categories of home ownership, borrowers with loan amount less than 25k are most like to be source verified.
- Across all categories of home ownership, borrowers with loan amount greater than 25k gets verified.
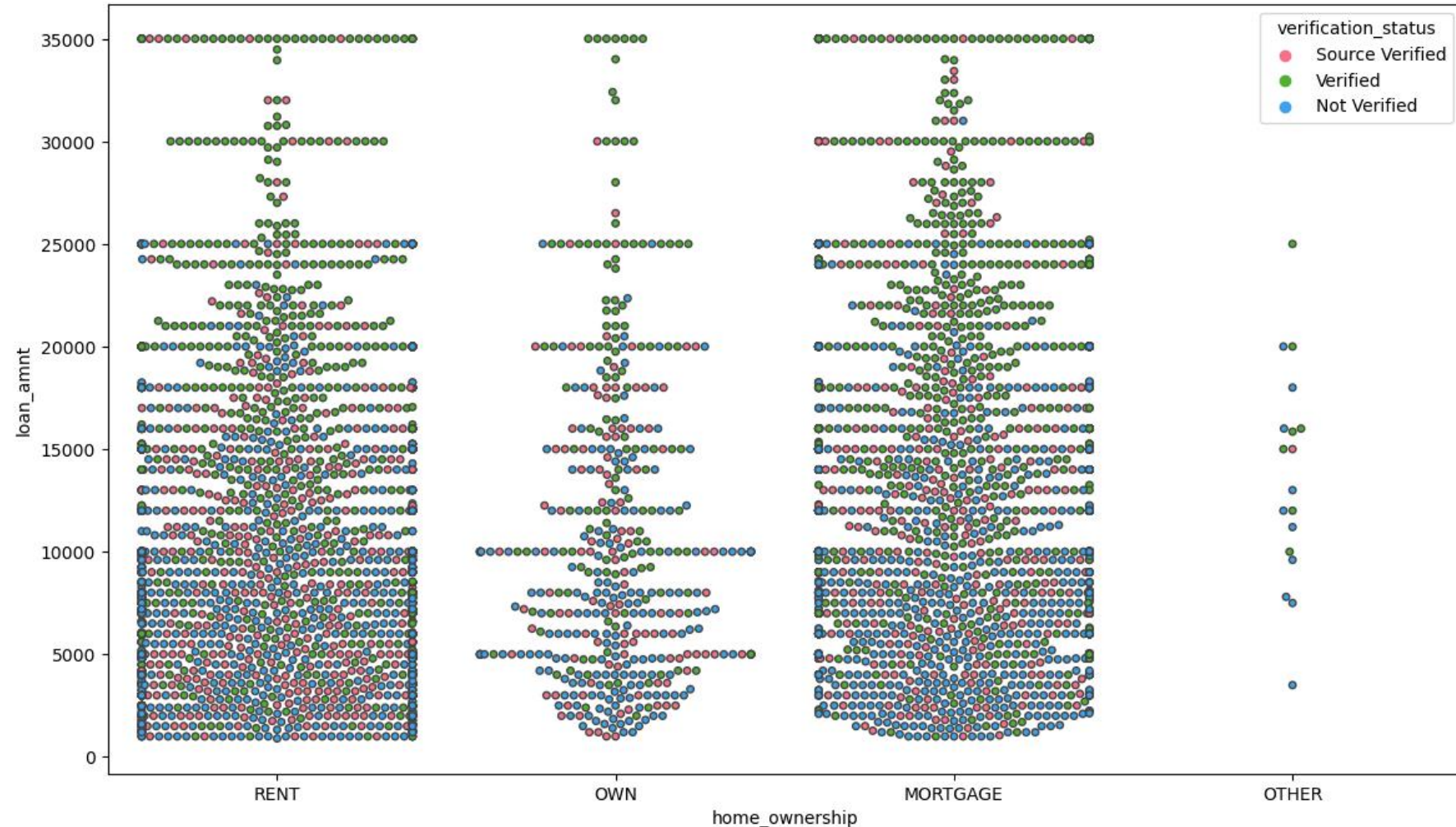
CONCLUSION

It is risky for the bank to give loan when

- Borrowers with high interest and 10+ years of employment length.

- Borrowers living on mortgage with loan amount greater than 12000.

- Loan for debt consolidation, credit card, small business with loan amount greater than 12000.

- For any loan amount provided with an average interest rate higher than 13%.

- Loan provided for house purpose with average interest grater than 15%.

- Loan provided with an average of 12.5% interest rate for 36months of tenure.

- Loan provided with an average of 16.5% interest rate for 60months of tenure.

- Average loan amount greater than 15000 with grade E, F, G.

- Installments between 800 - 12000 with average interest greater than 15%.

- Average interest rate of 17% with installment greater than 12000.

THANK YOU!