Gramener Case Study

Group Name: Datatitans

Parul Jain

Ritisha Swain

Polamuri Mohana Krishna

Abhishek Kumar

Objective:

ABC is a consumer finance company that deals in lending various types of loans to urban customers.

Business Objectives:

1. Identification of Loan Applicant traits that tend to ‘default’ paying back
2. Understand the ‘Driving Factors’ or ‘Driver Variables’ behind Loan Default.
3. Possible Solution to the Issue.

Problem analysis

Understanding Data:

- This data has information about the loans that have been issued by the company. They are mainly of three types Fully Paid off, Current and Defaulted. The data set contains all sorts of information about customer like annual income, employee length, purpose of loan, no. of bankruptcies etc.

Meta Data : A glance at the data can tell us about different types of variables such as

- Variables related to customers demographic and its characteristics
- Variables related to the loan characteristics
- Variables related to the customers behavior characteristics.

The following steps are followed for the analysis:

- Clean and format the provided data for readability.
- Remove the extra columns that might not be useful for analysis.
- Extrapolate the Derived Metrics.
- Look for the outliers and remove them.
- Filter Data Set on Loan Status.
- Perform Univariate and Bivariate analysis on the Data.
- Plot the results of Analysis and Hypnotize the Root of the Issue.

Assumptions:

- As current loans have not yet been completed, we decided to ignore them during analysis.
- Outliers (Based on Annual Income) have not been considered during analysis.

Data Cleaning/Preparation

1. Converting the date to a more understandable format.
2. Remove all columns that :

- don’t change as there is no variance, it cannot help us determine the reason for default. We can save memory and analysis, plotting and data frame transformations are faster
- don’t provide any value. E.g. url, title etc.
- Are redundant E.g. The purpose of loan is a drop down which is already a categorical variable. We don’t need the title column as it becomes redundant.
- have all NA or a single value.

3. Converting the columns to numeric that are read as characters due to the presence of the special characters such as %.
4. Since we are dealing with the aggregate, we may not need the primary keys in this analysis such as id and member_id.
5. Converting the appropriate columns as categorical value.

Univariate Analysis

- For our Analysis we have:

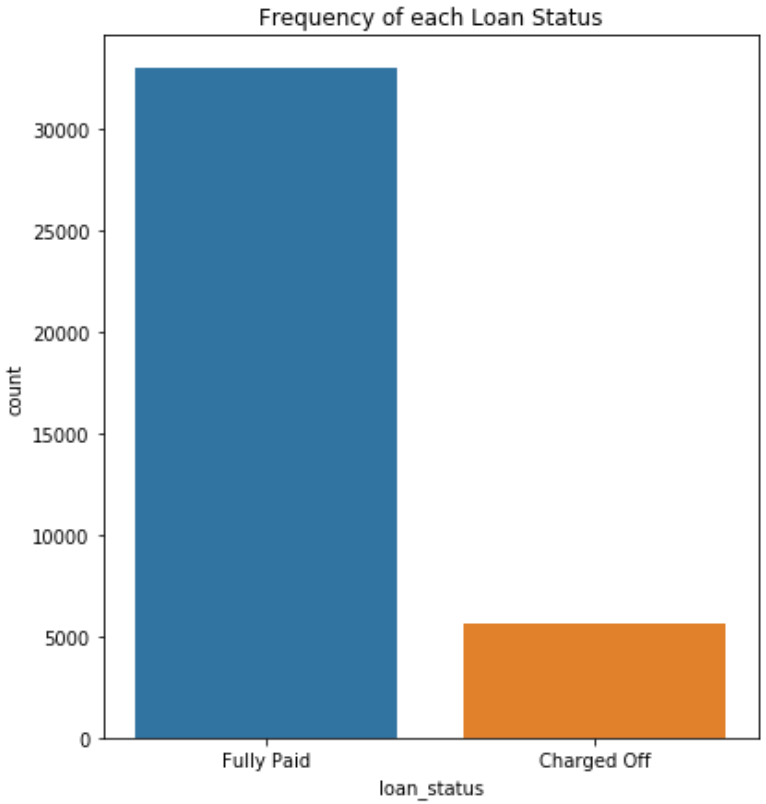
Categorical variables: purpose, home_ownership, emp_length, grade, sub_grade, verification_status and term.

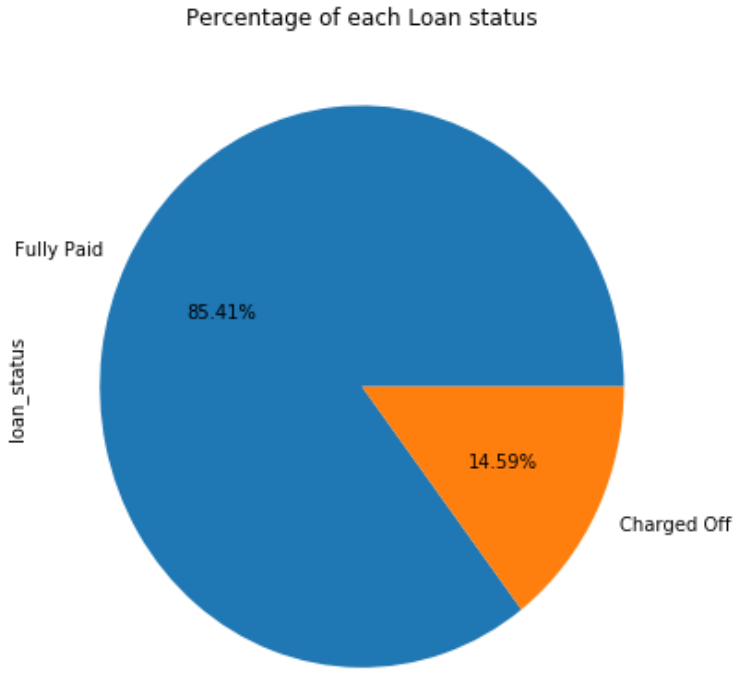
Quantitative Variables: loan_amt, funded_amnt, funded_amnt_inv, int_rate, installment, annual_inc, dti.

Target variable: loan_status.

- We will be considering only Fully paid and Charged Off.
- As Current is still on going loan and we cannot draw any analysis from it
- Fully Paid can be consisdered ad Non Defaulted loans and Charged Off can be termed as Defaulted loans

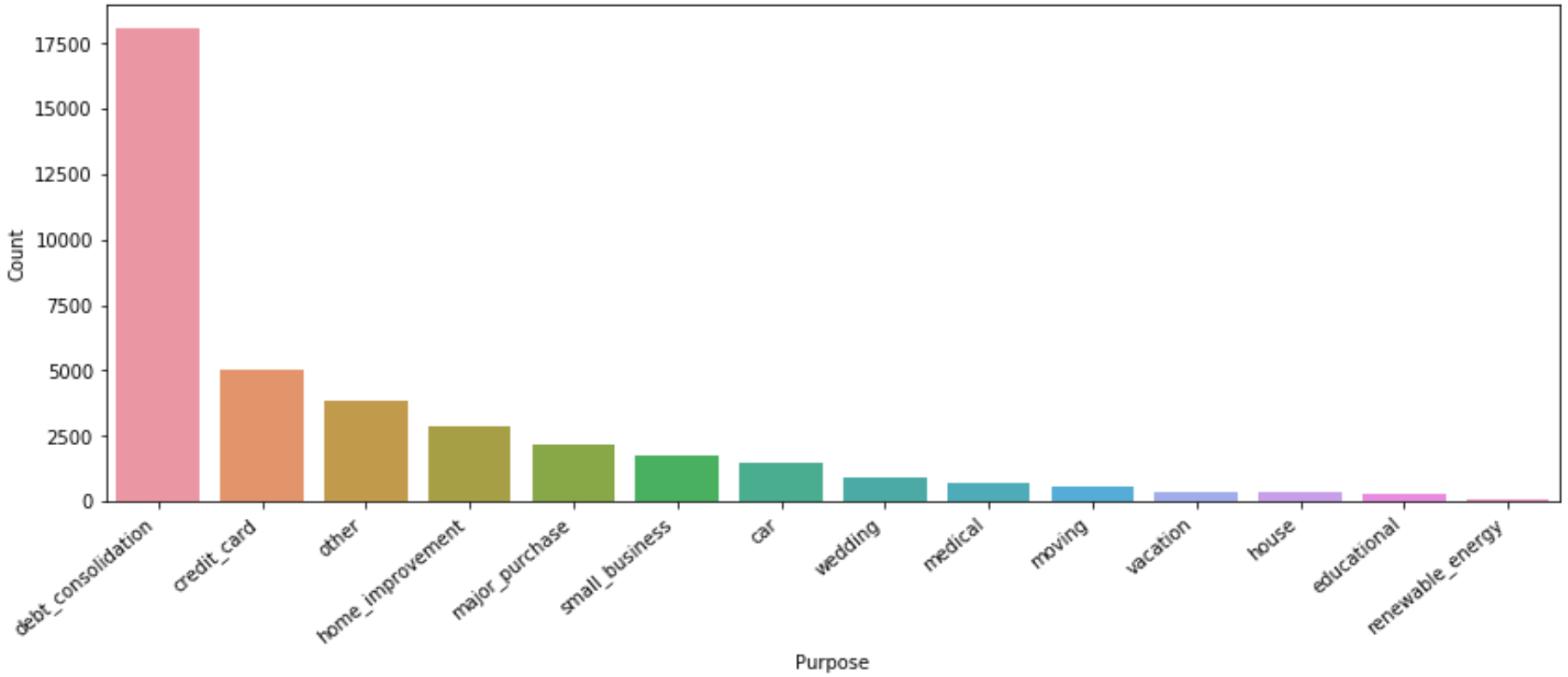
Univariate Analysis





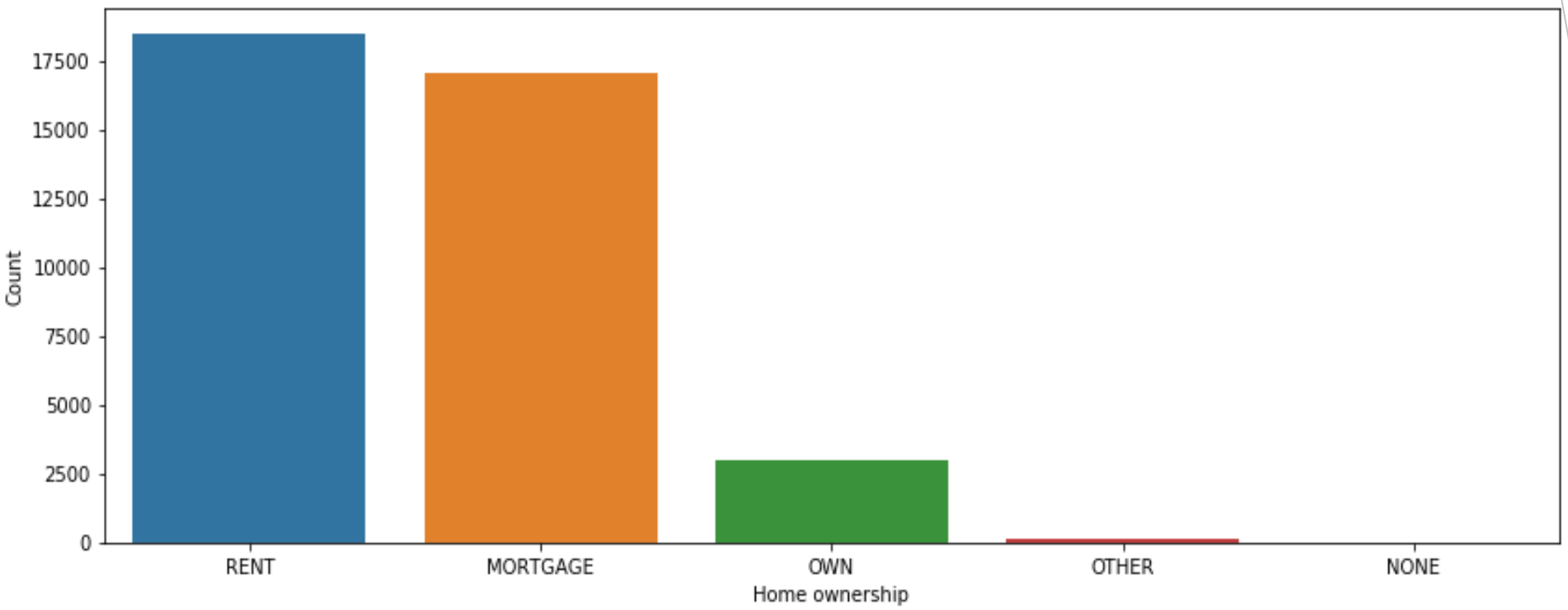
- We can clearly see that 14.6% of the loans are Defaulted

Univariate Analysis



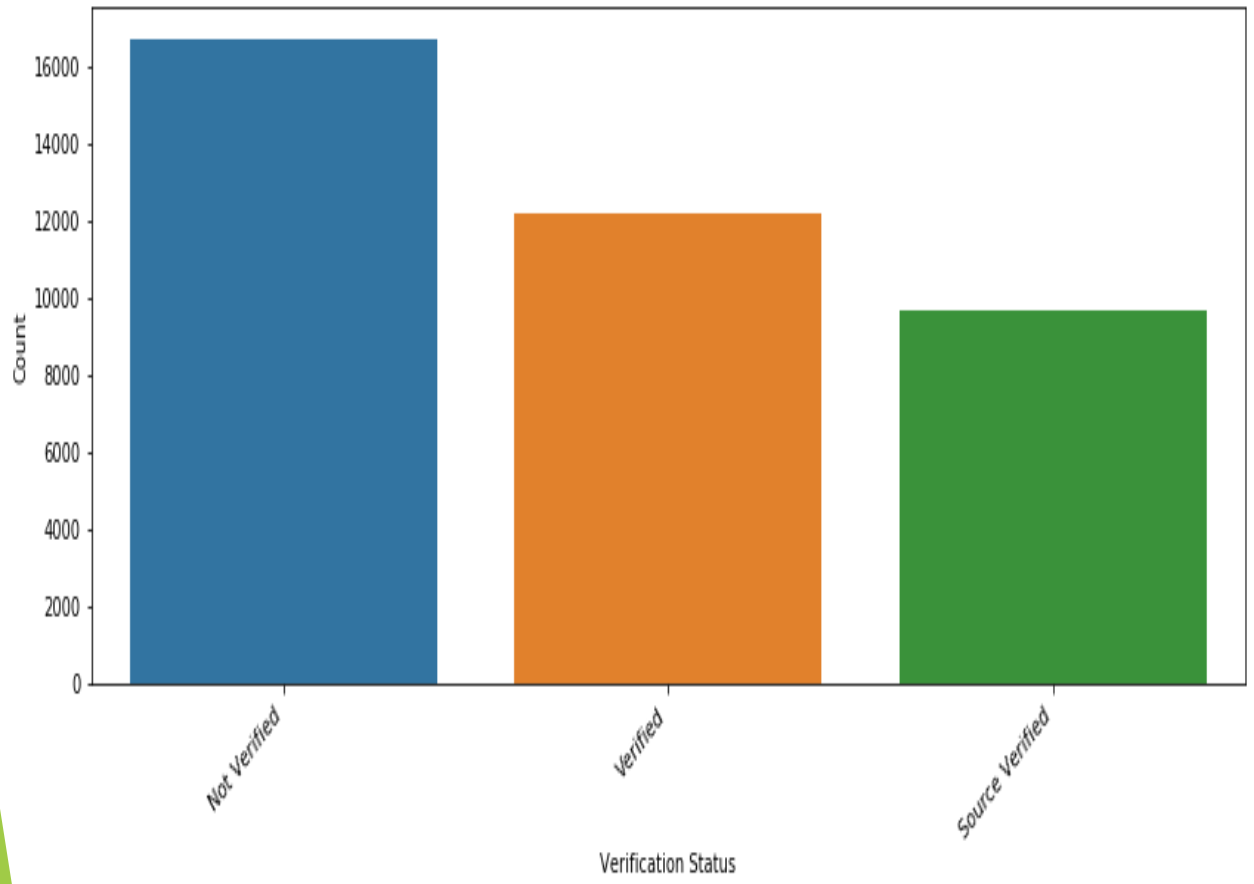
- Most loans are taken for Debt consolidation, credit card bills

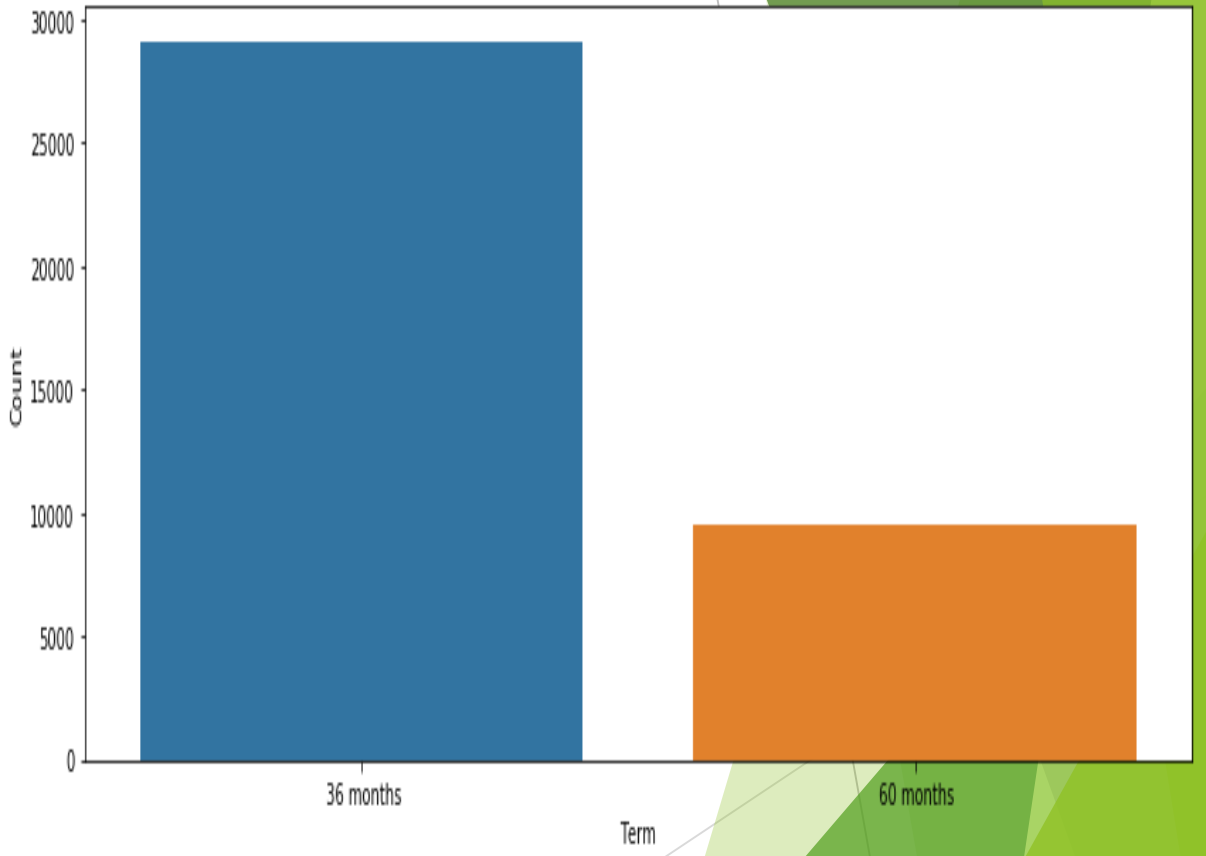
Univariate Analysis



- Large number of loan takers have rented and mortgaged homes

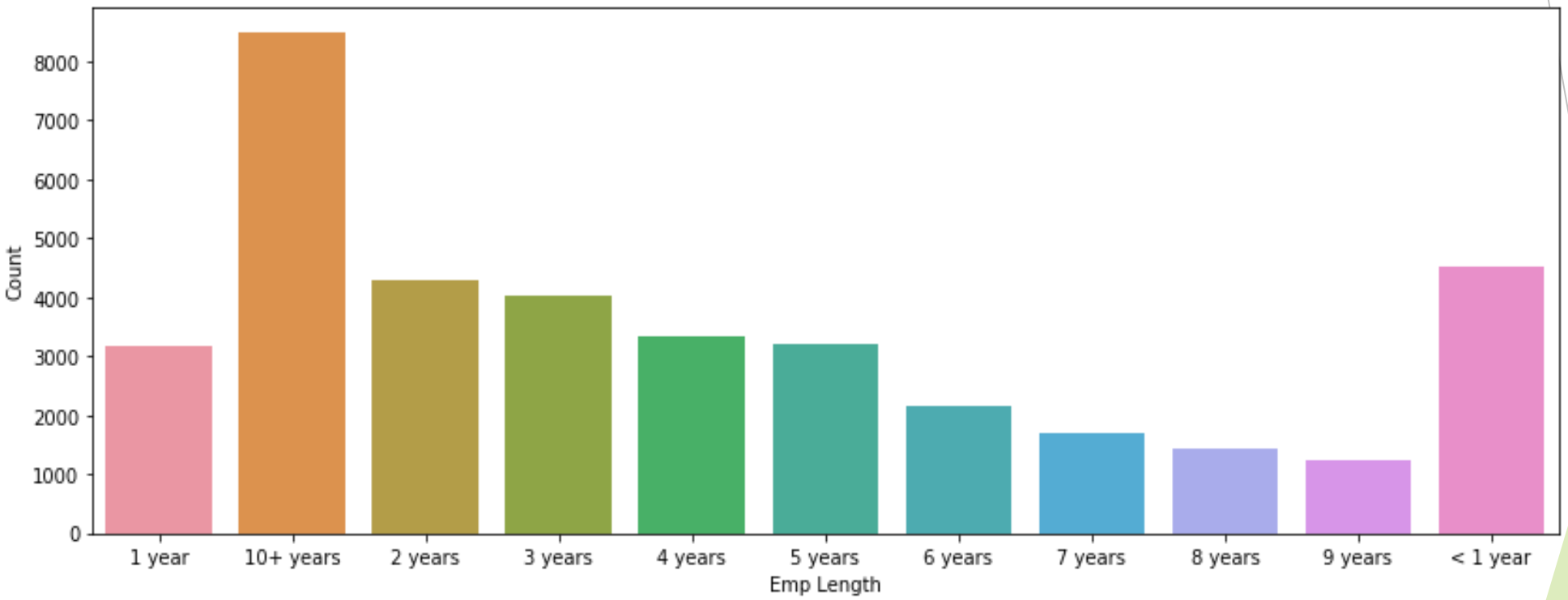
Univariate Analysis



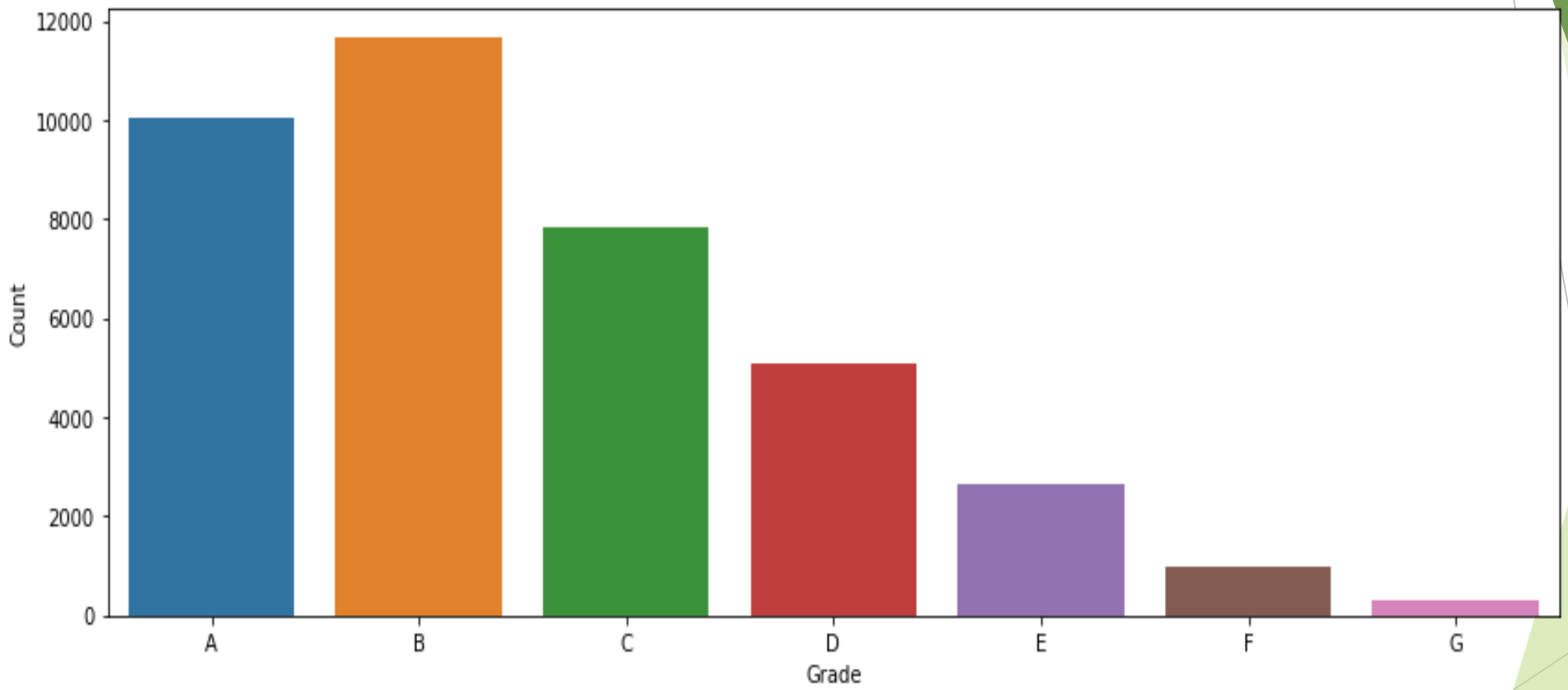


- 36 Months term loans are more.
- For a little surprise most of the loans are not verified.

Univariate Analysis

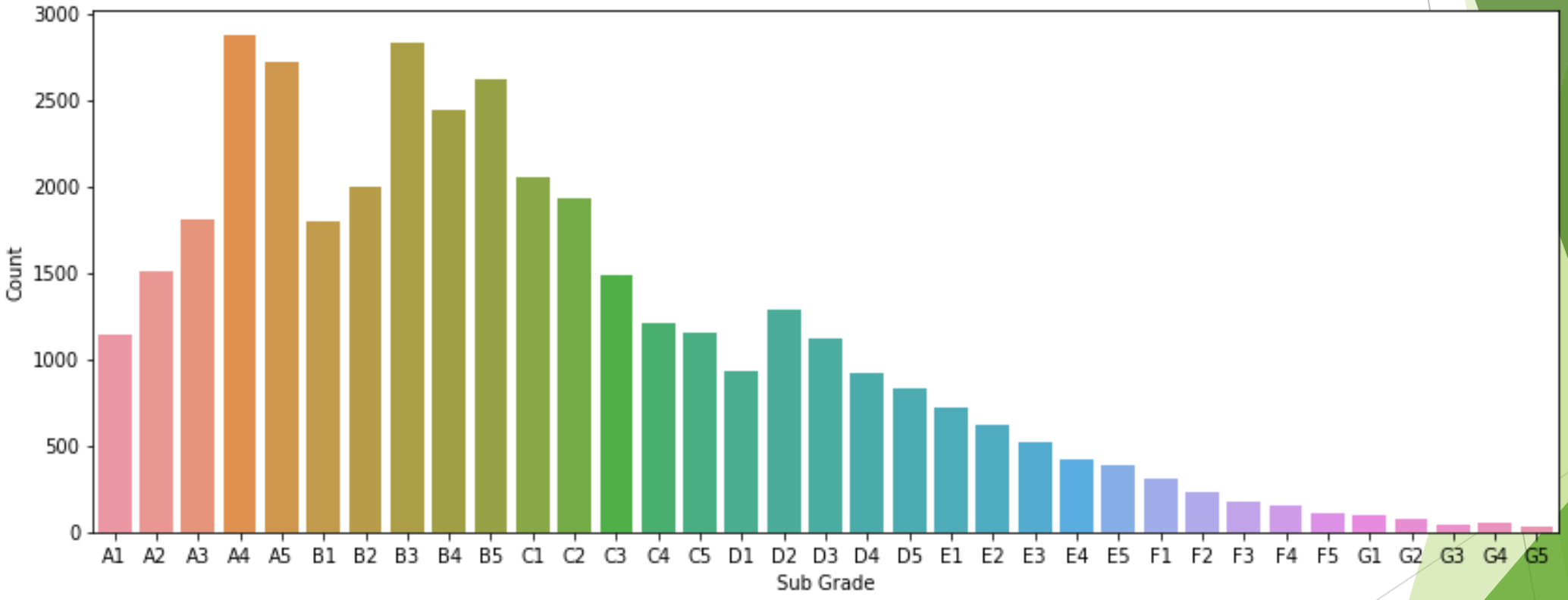


Univariate Analysis

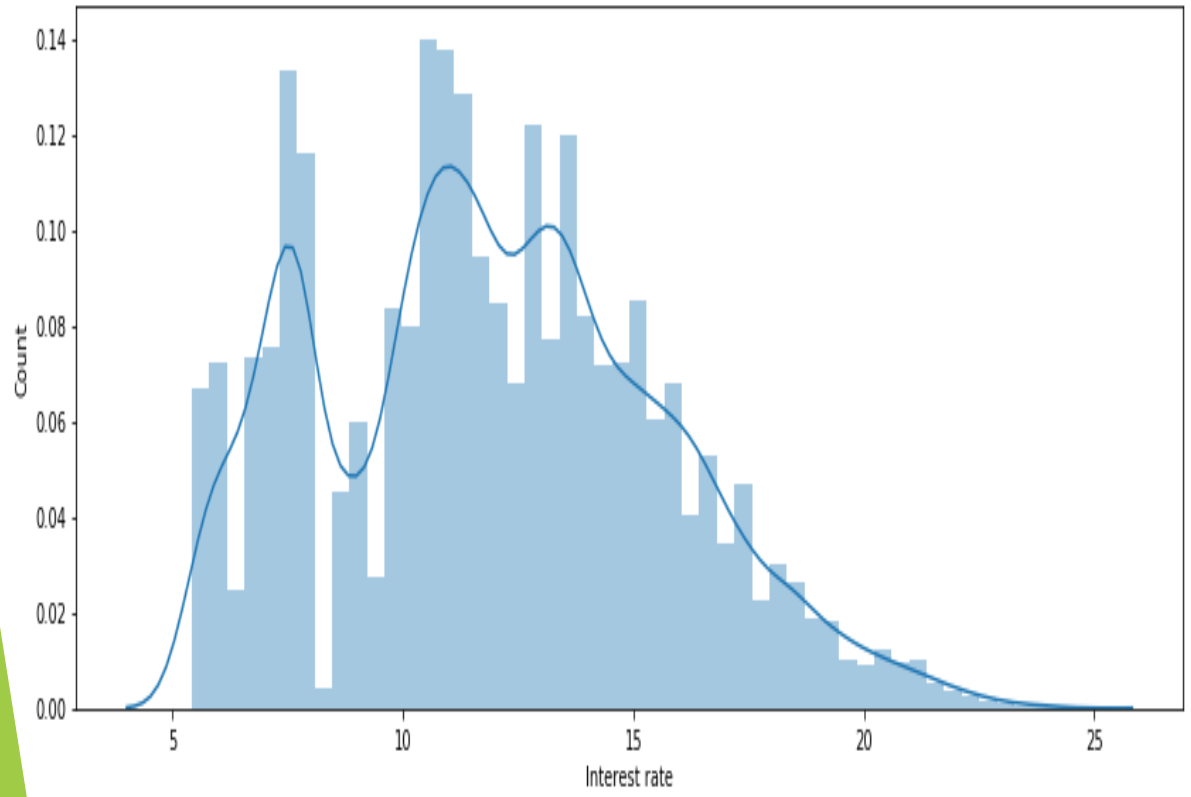


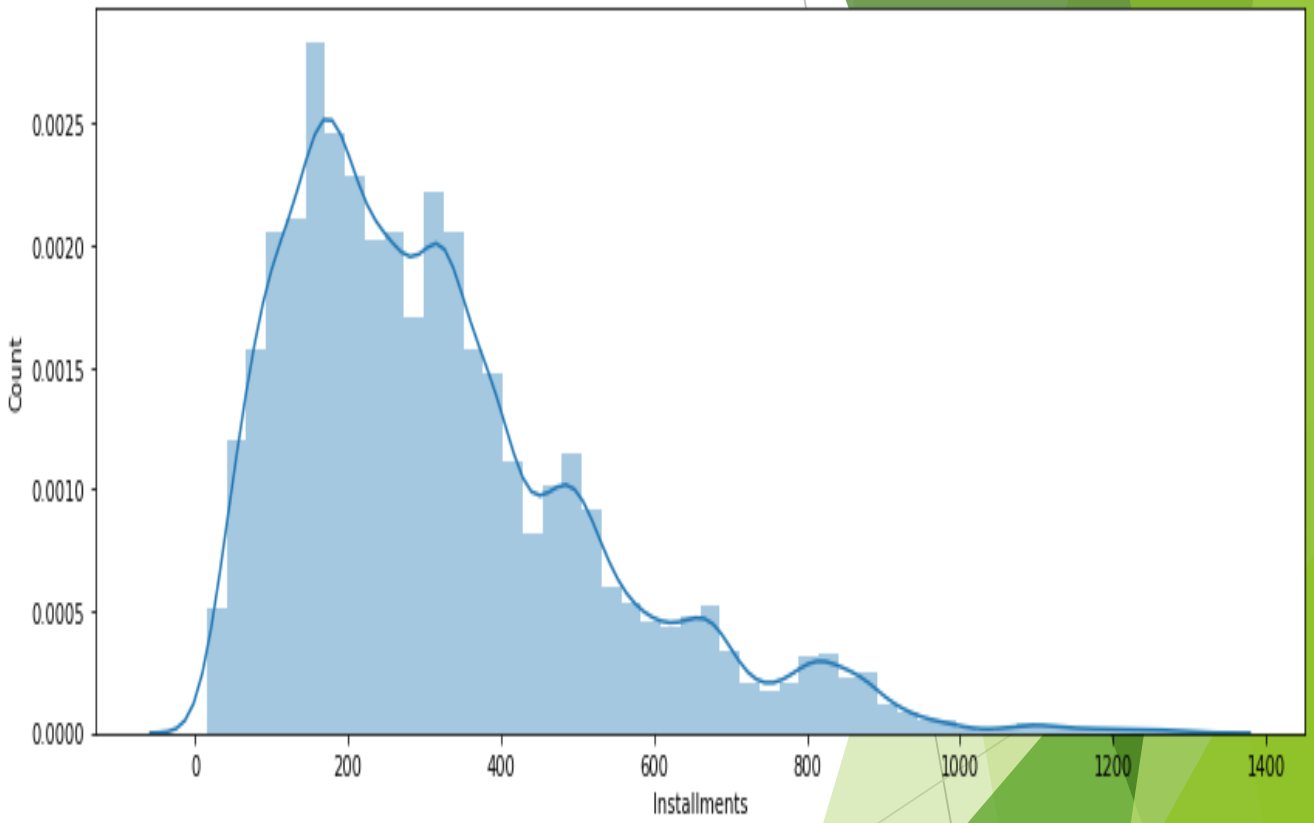
- Grade B loans are more

Univariate Analysis

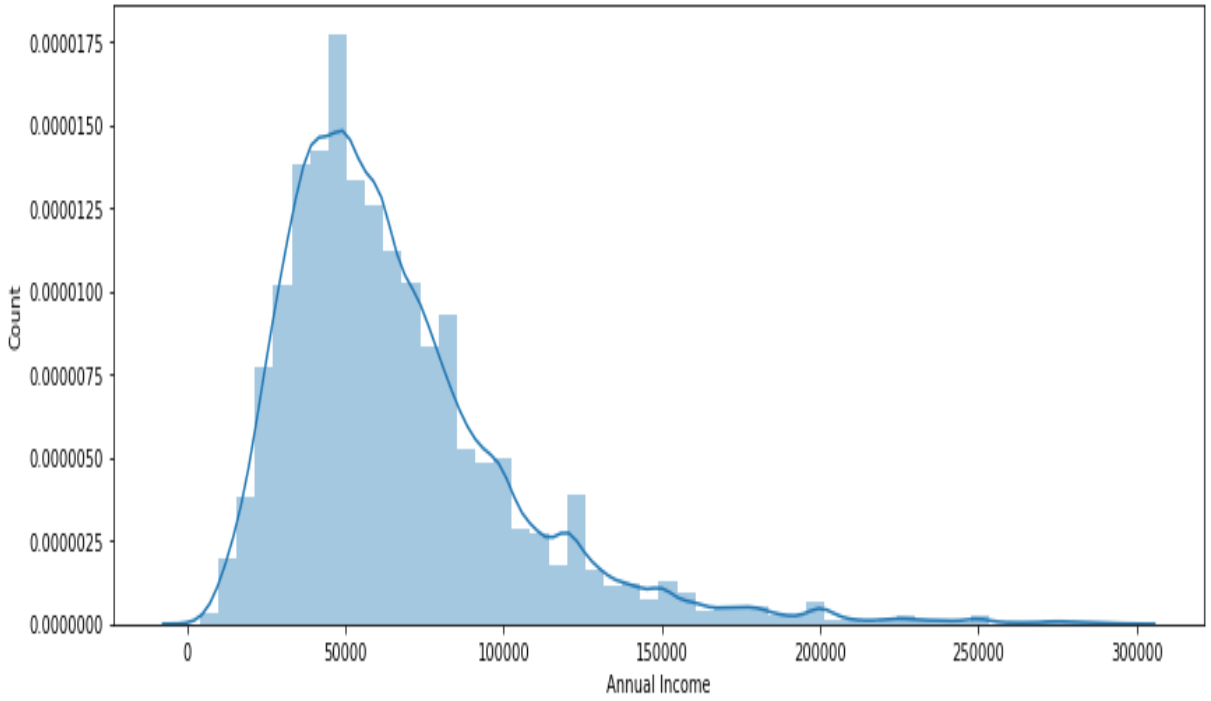


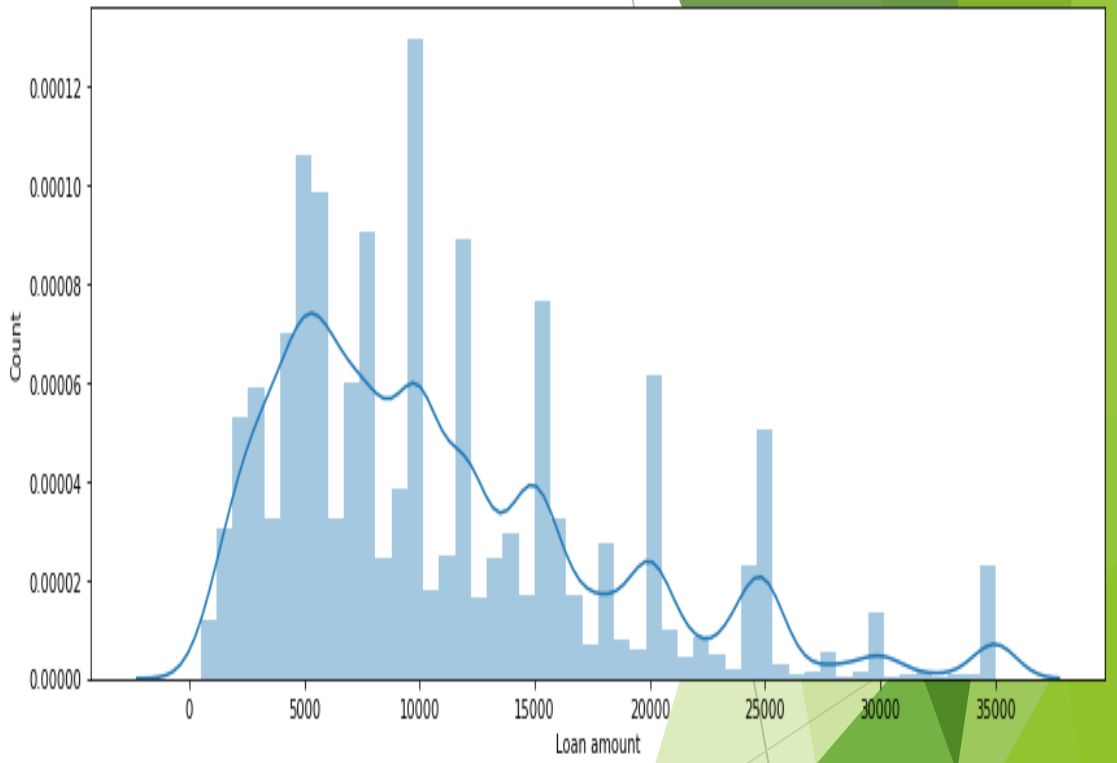
Quantitative Univariate Analysis



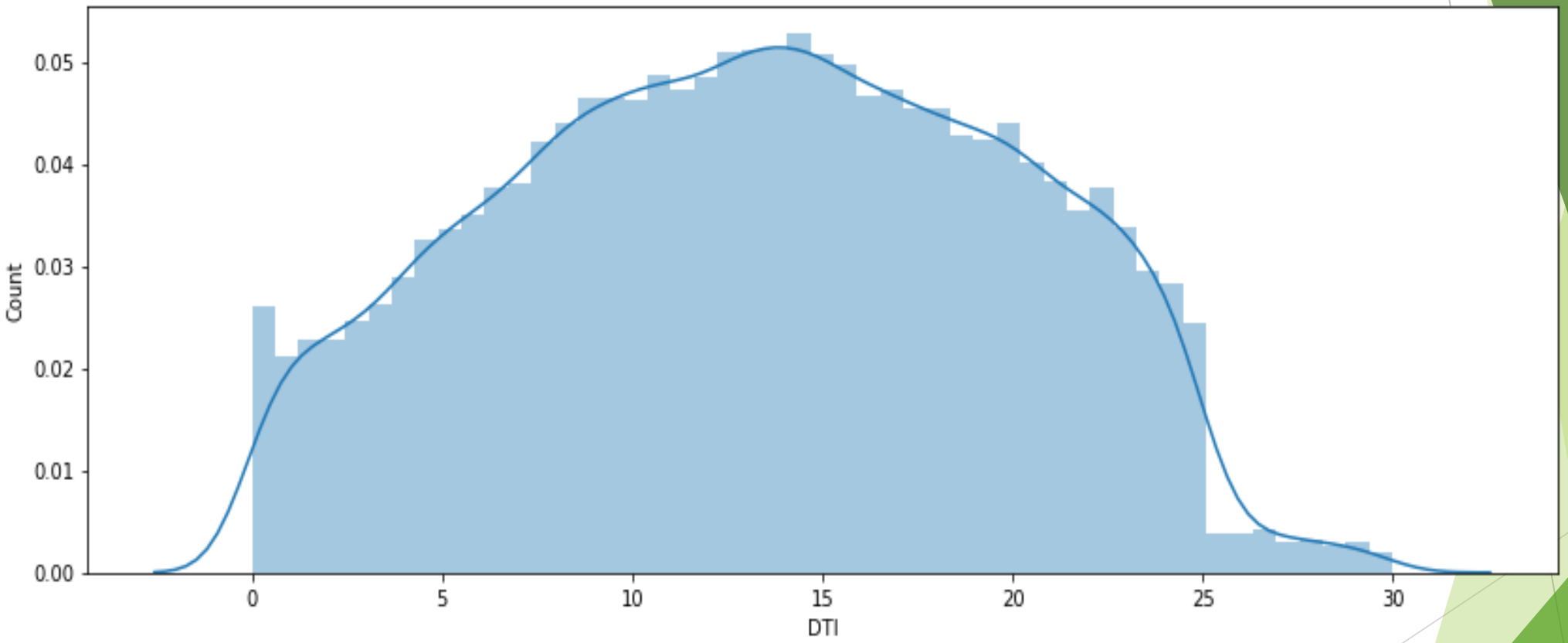


Quantitative Univariate Analysis

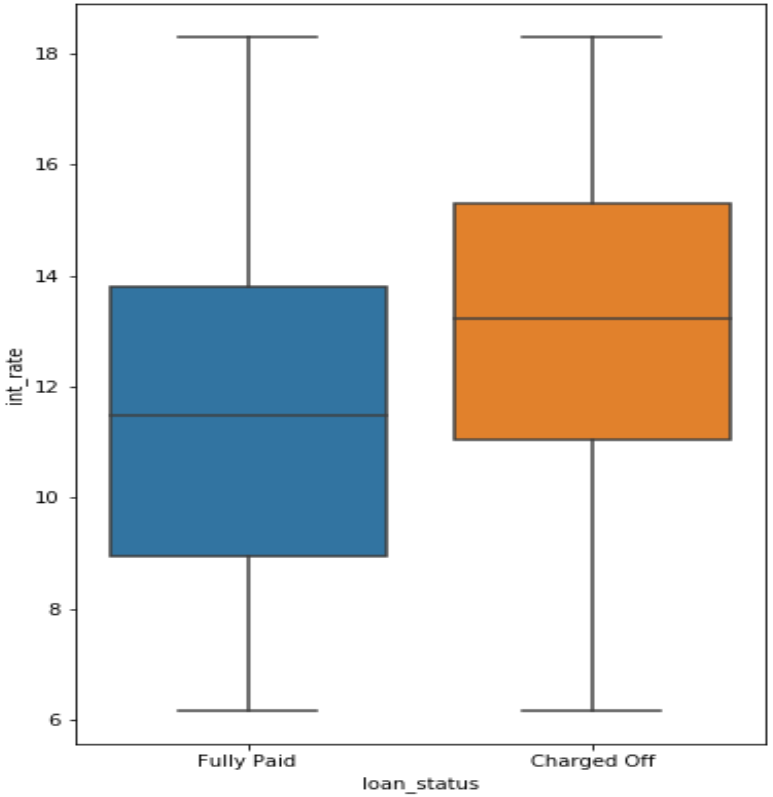


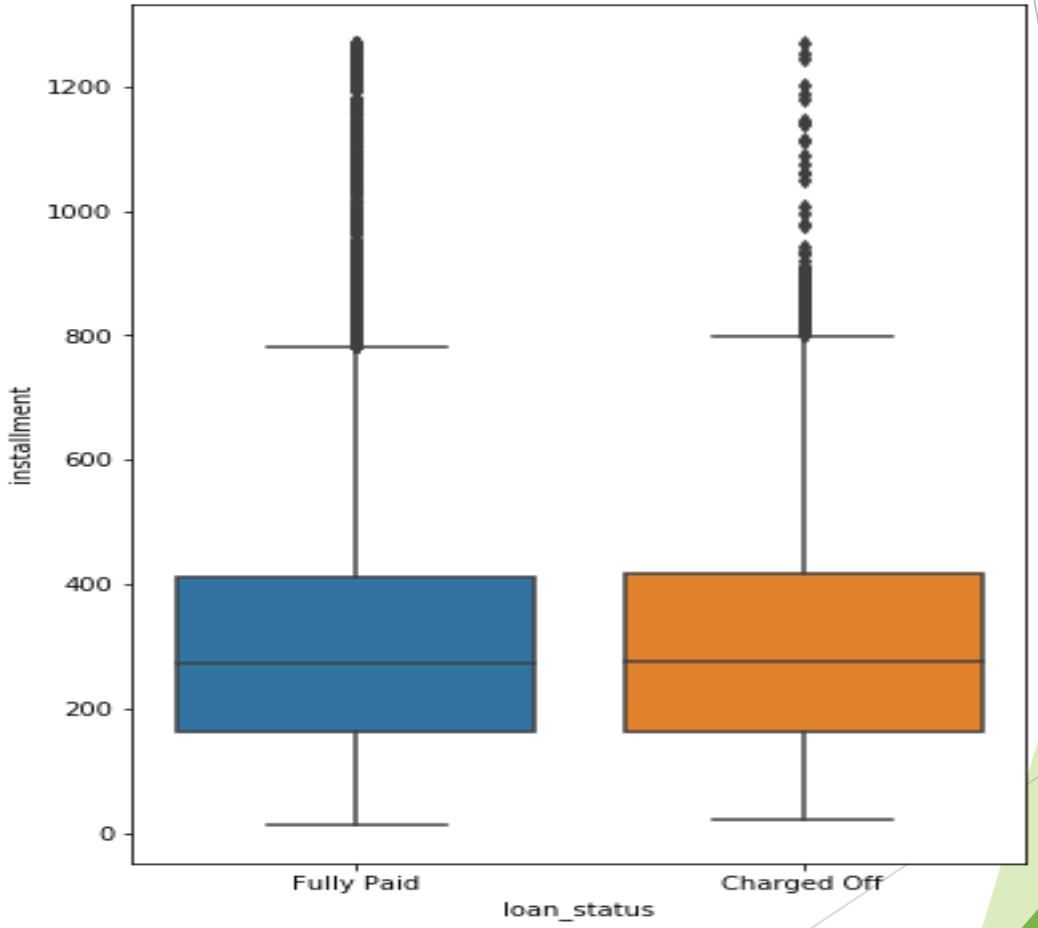


Quantitative Univariate Analysis

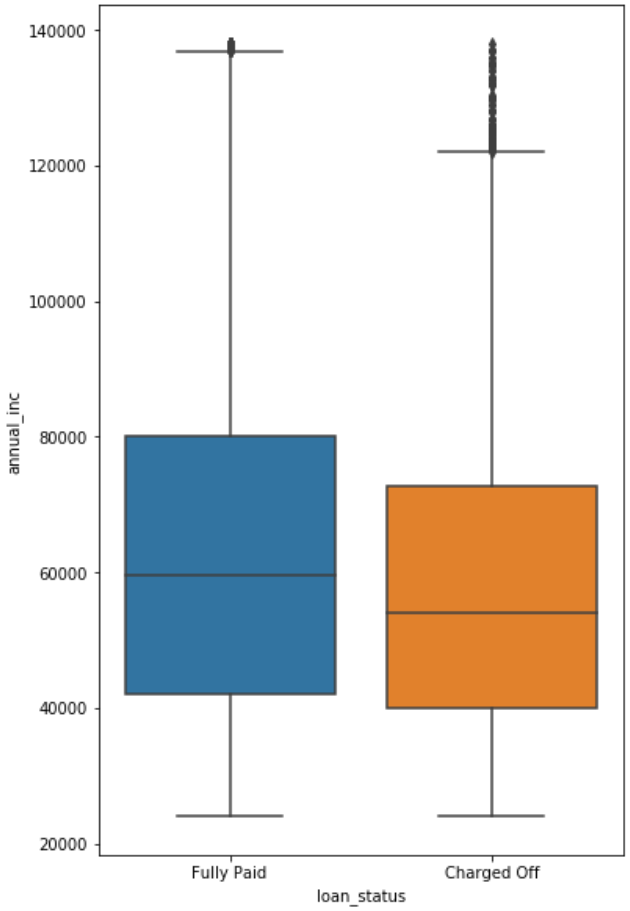


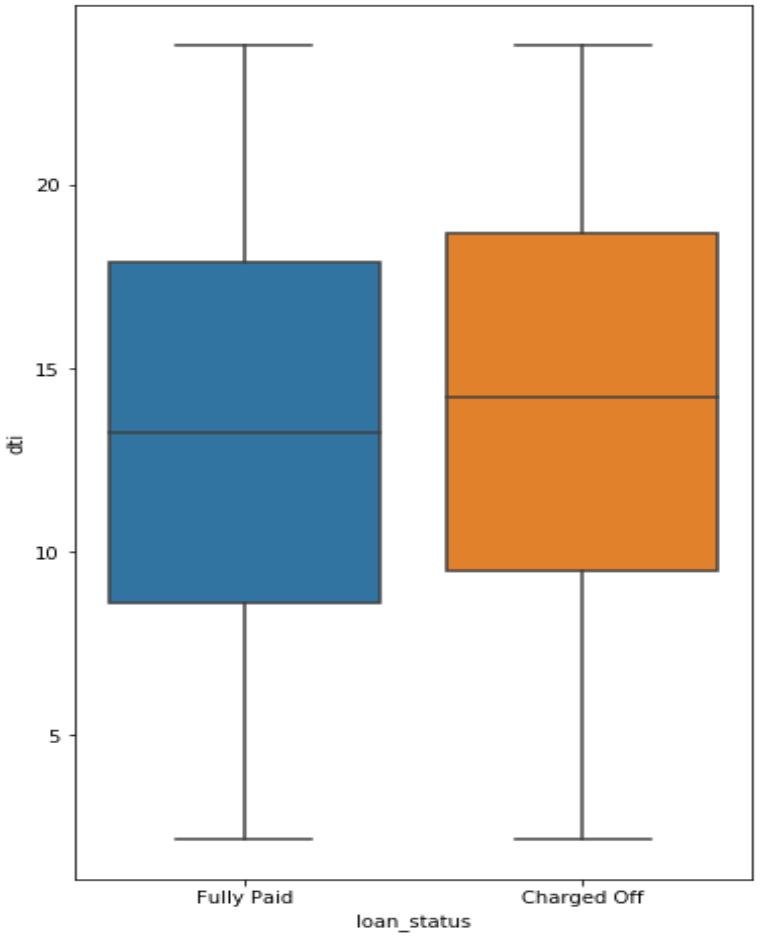
Segmented Univariate Analysis

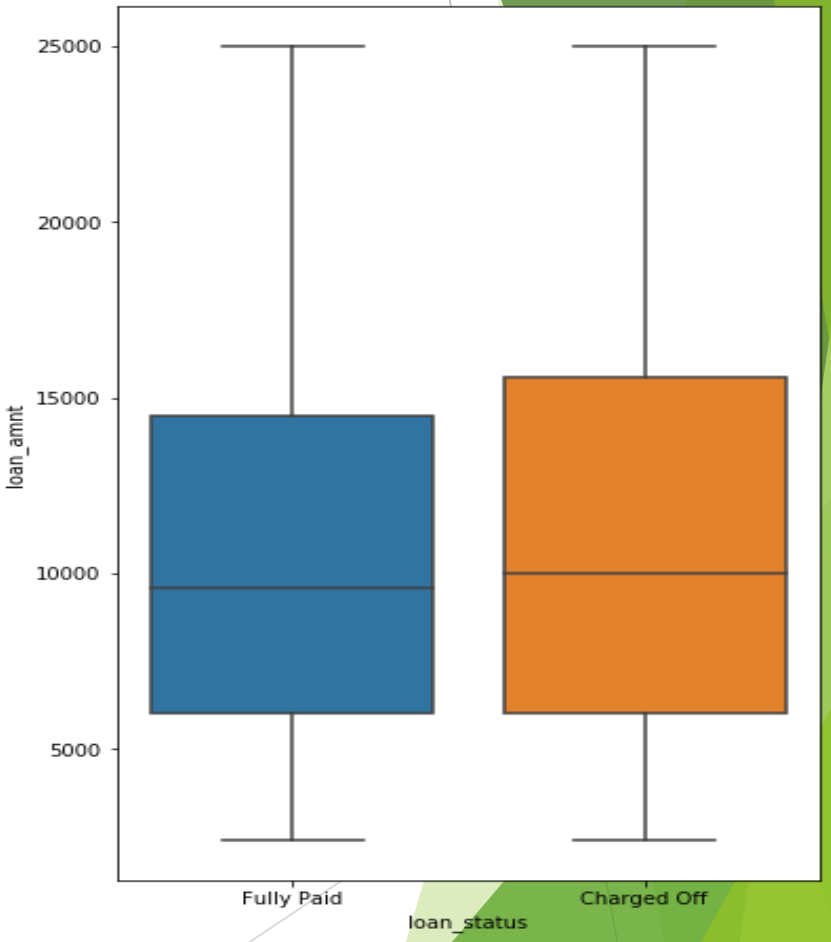




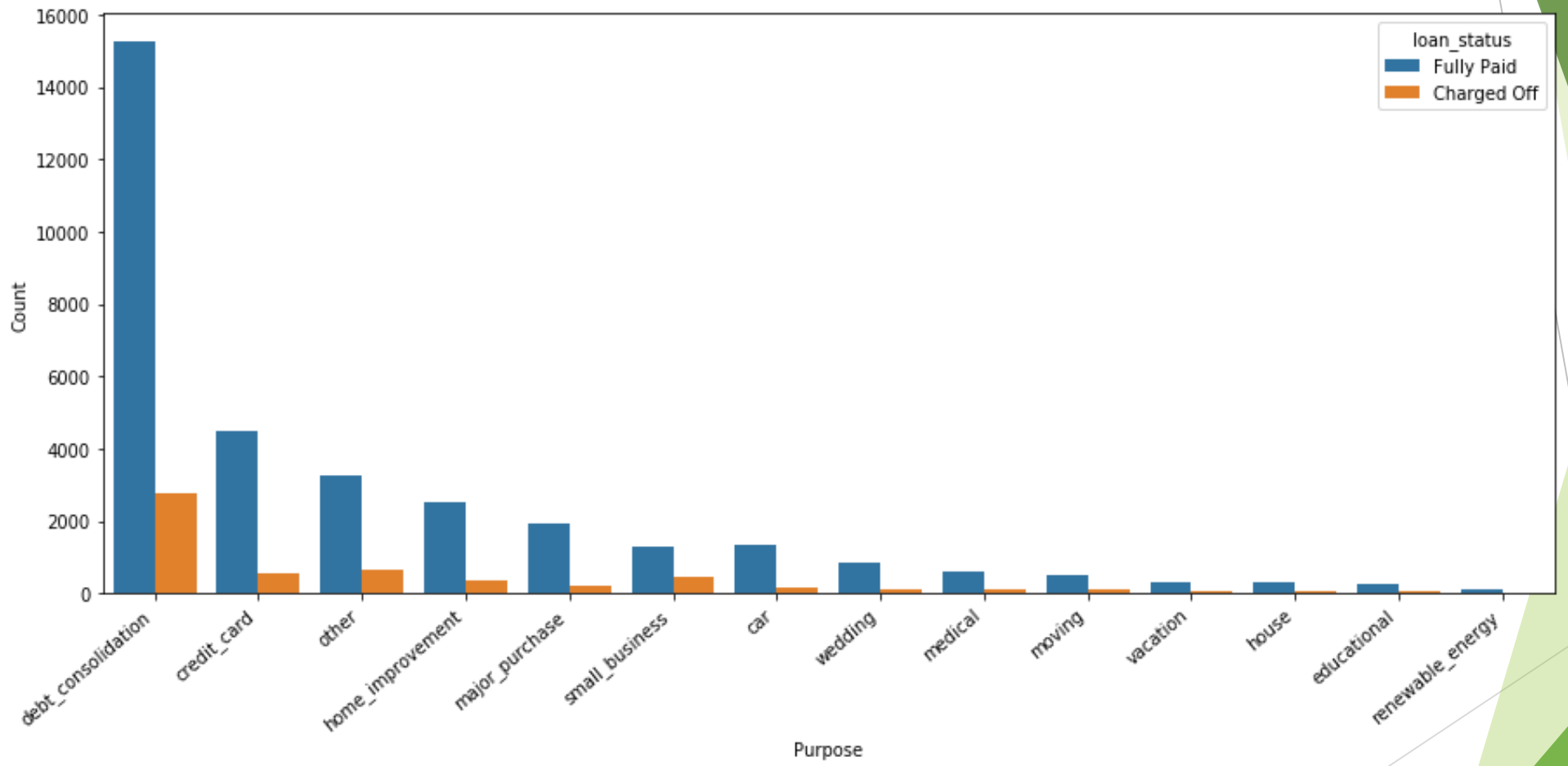
Segmented Univariate Analysis



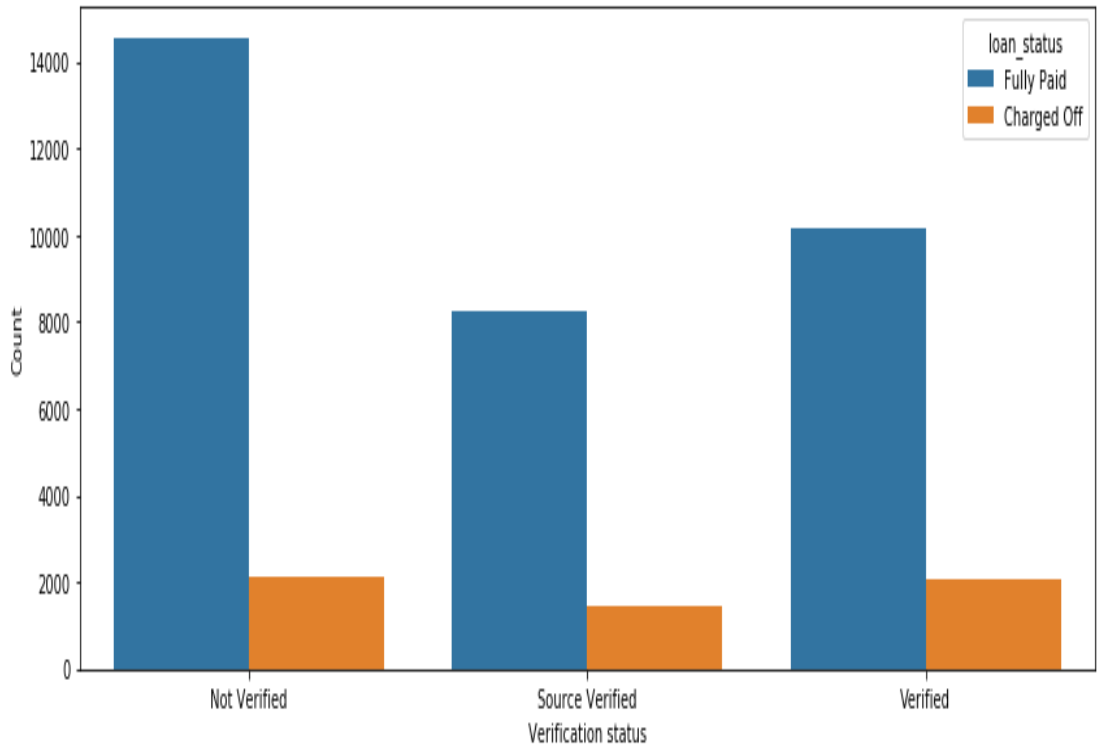


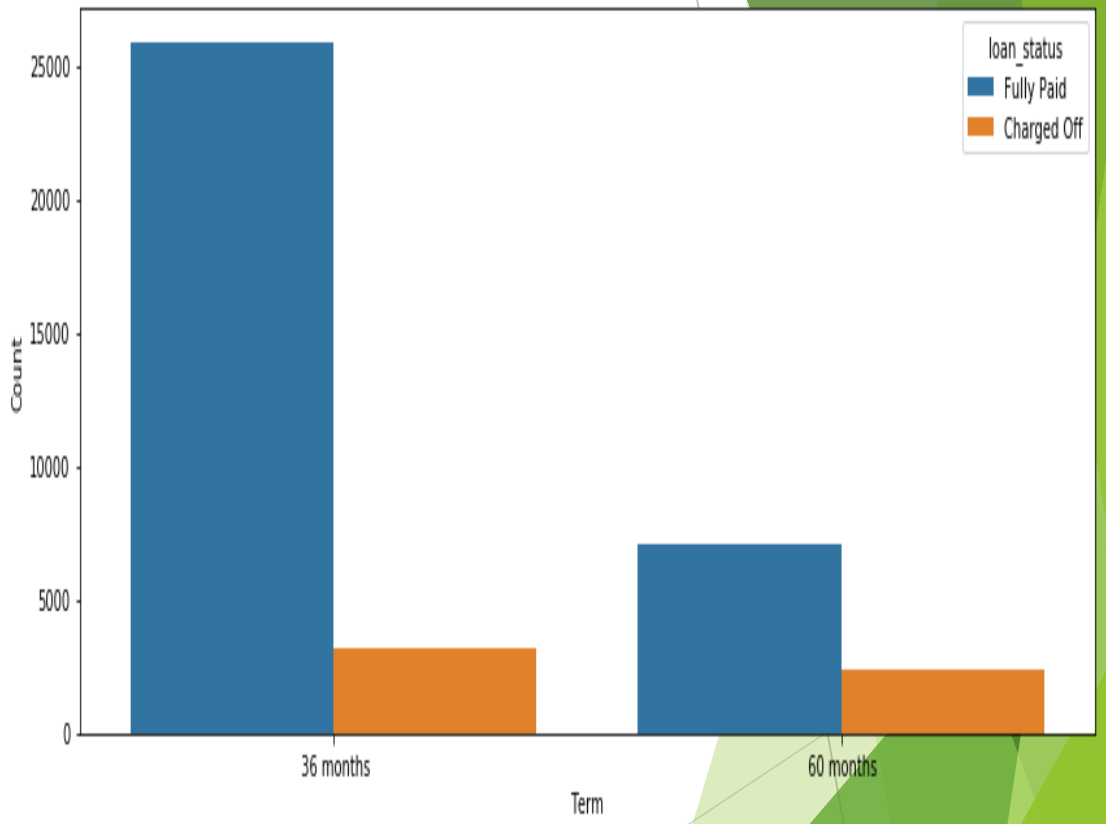


Bivariate Analysis

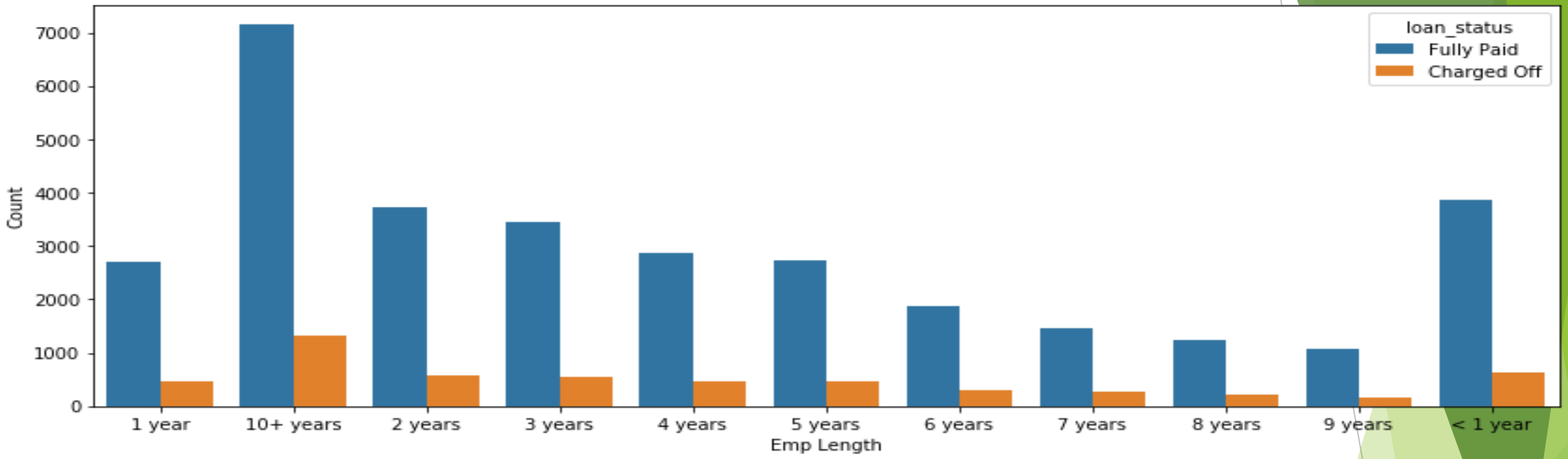


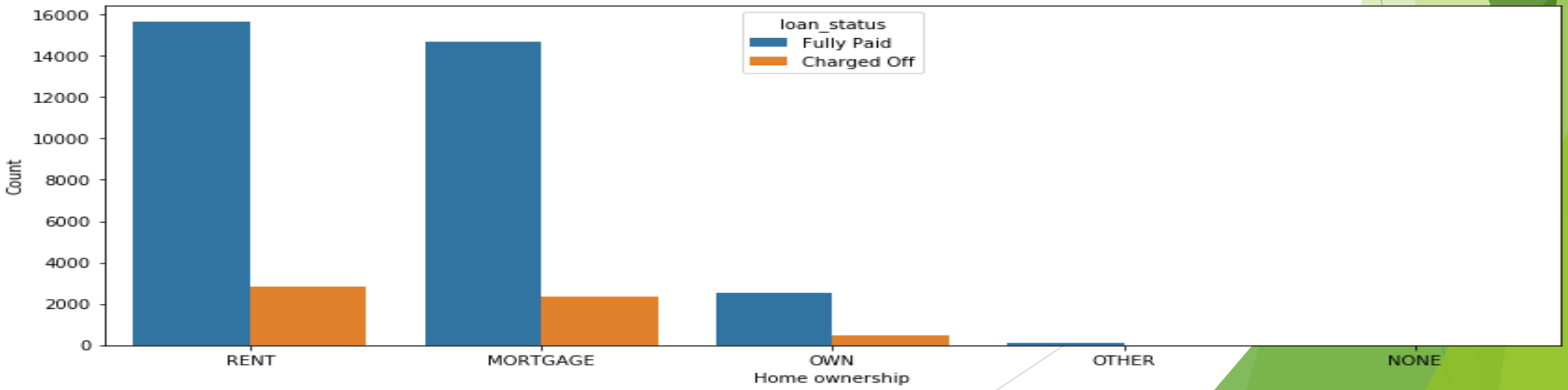
Bivariate Analysis



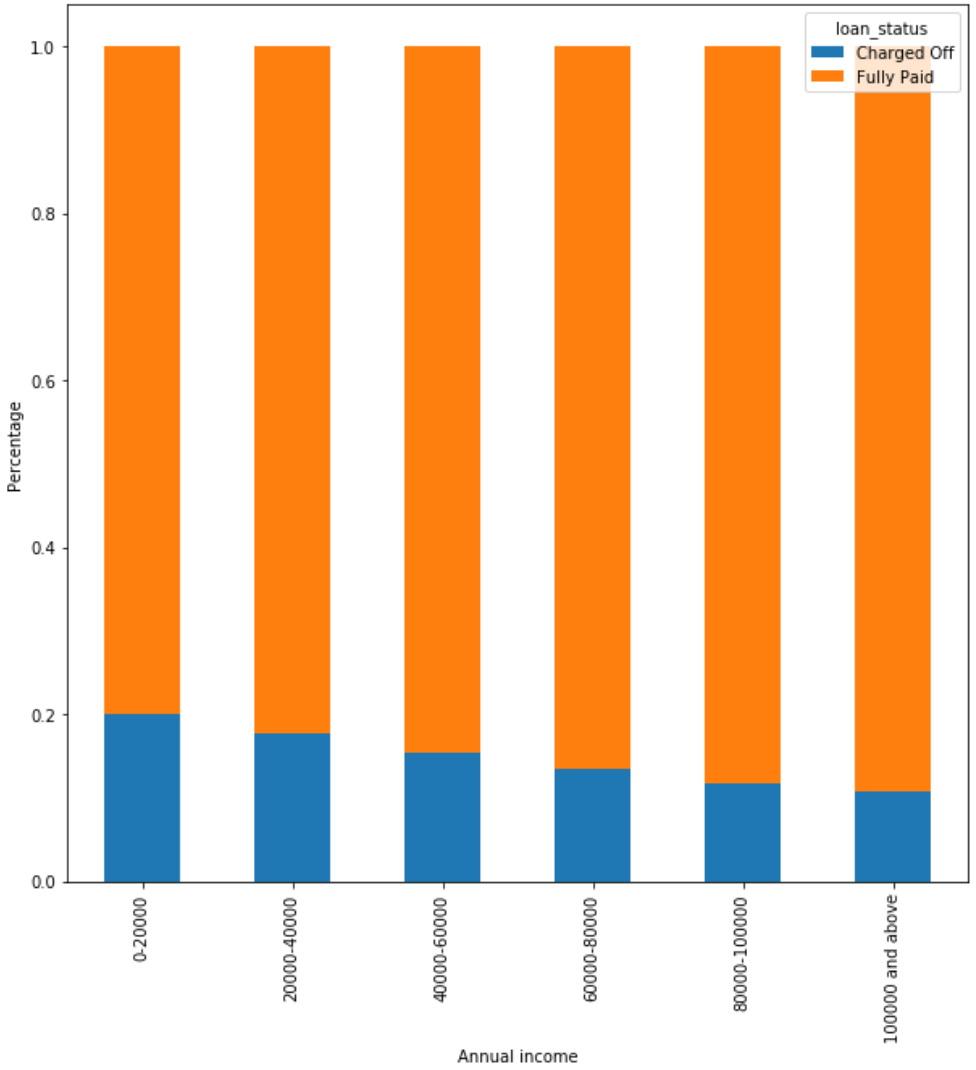


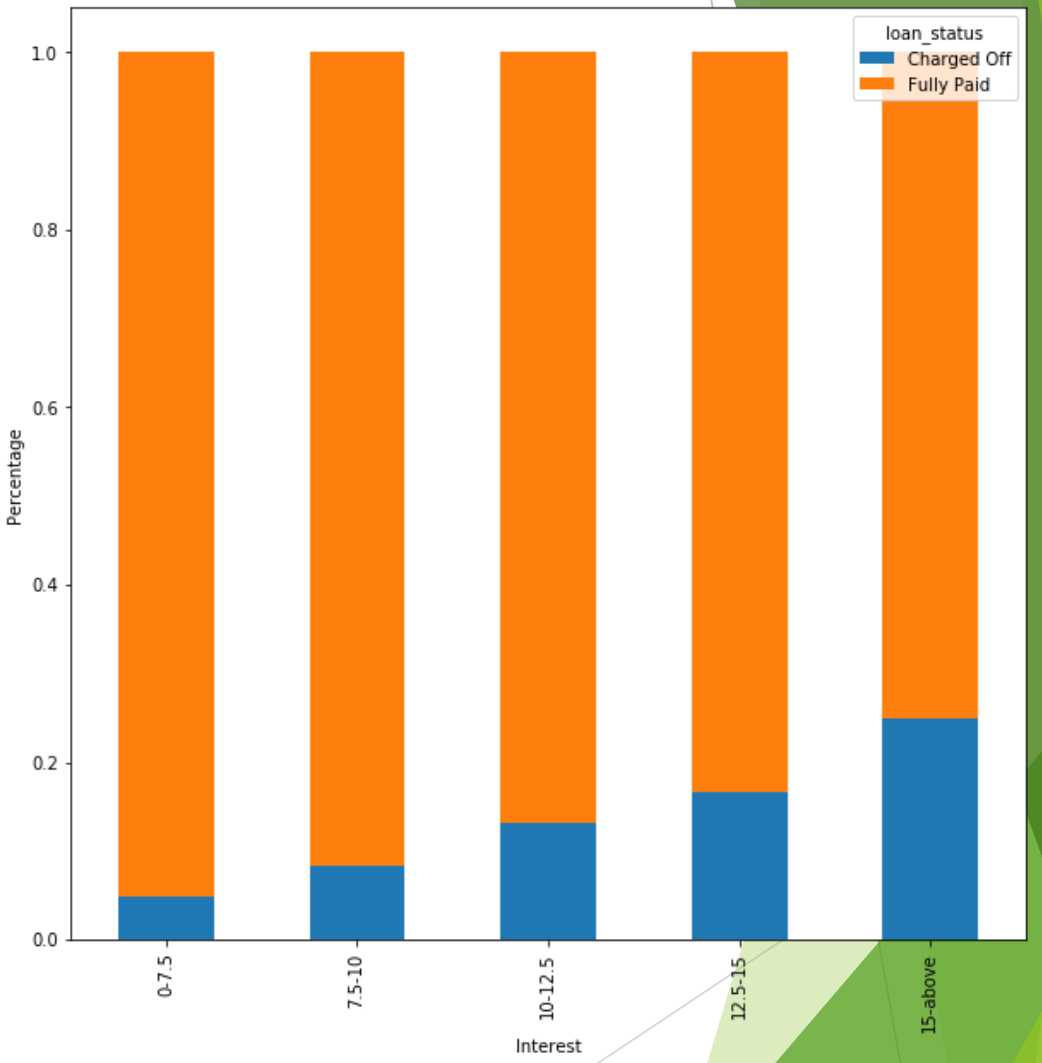
Bivariate Analysis



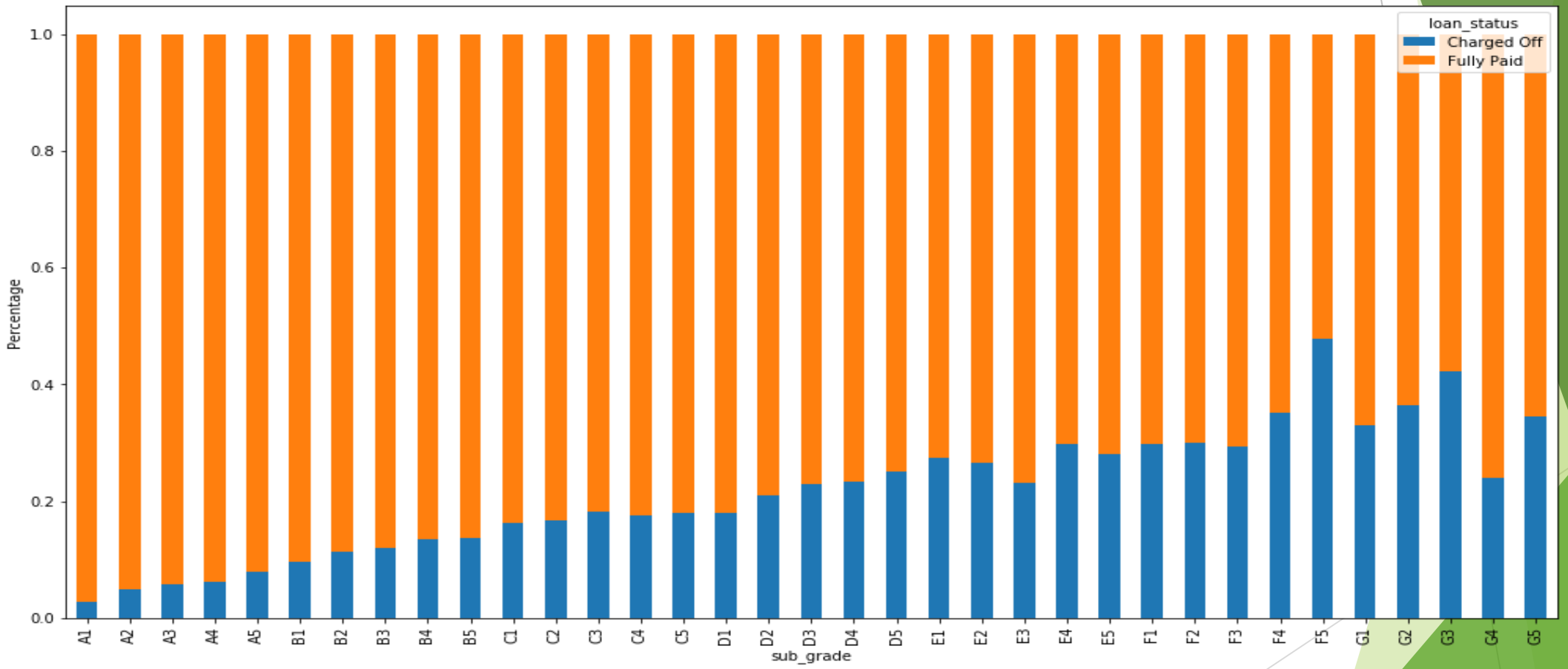


Bivariate Analysis

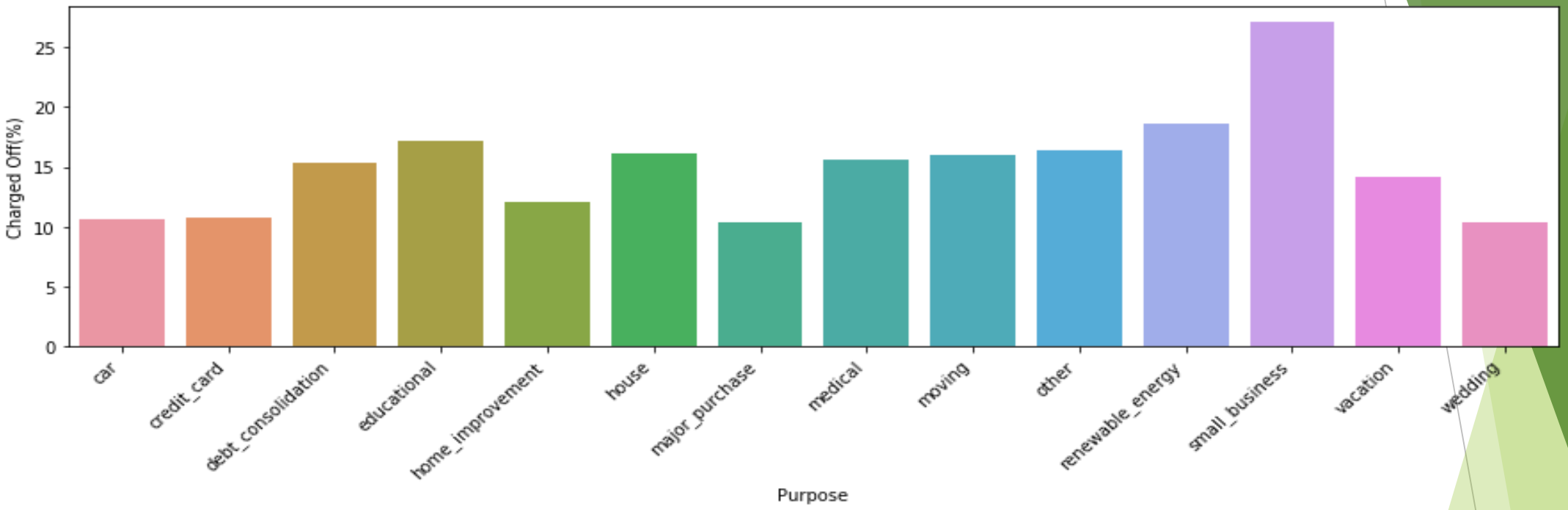


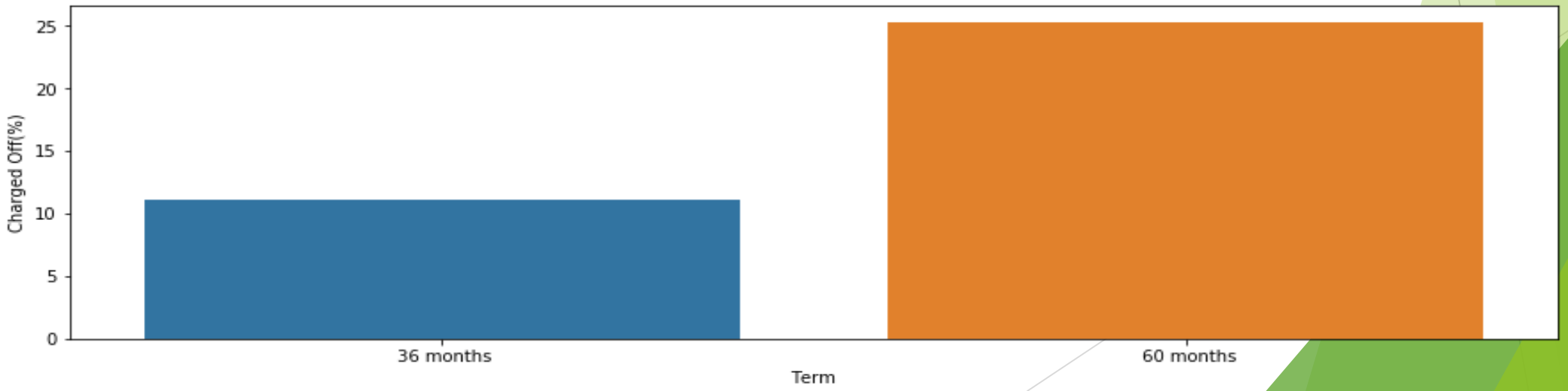


Bivariate Analysis



Conclusion Graphs





Conclusions

- Most of the loans are taken to consolidate debts from other loans.
- Majority of applicants stay in rented or mortgaged house.
- Verification status does not seem to be of significance since most of the applicants are not verified.
- A considerable number of applicants have charged off.
- About 70% of the loans are for a tenure of 36 months.
- Grade does not seem to have any clear pattern - needs more analysis.
- Most of the loan amount is between 5000-20000 USD.
- A large portion of the applicants have an annual income of 50000 to 100000 USD.
- The most prevalent interest rates are 12 to 16 percent.

Conclusions

- DTI needs to be further analyzed.
- Low Annual Income can cause High Default.
- Instalment to income ratio is driving factor. Having high value means high chances of default.
- For 60 months term, if the instalment to income ratio is greater than 9%, there is maximum probability of defaulting
- Loan taken for 60 month interval has higher chances of defaulting.
- Nearly 35% of loans in the interest bracket 20-25% have been defaulted followed by 15-20% interest bracket which has defaulted loan of 24%. So if the loan has higher interest rate (about 15%), there is higher chance of defaulting.
- Purpose of loan: Loan taken for “small business” has high chances of defaulting.

THANK YOU