Lending Club Case Study

Sneha Sarkar

9th November 2022

Contents

Problem Statement

- Lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.

- The main objective is to be able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

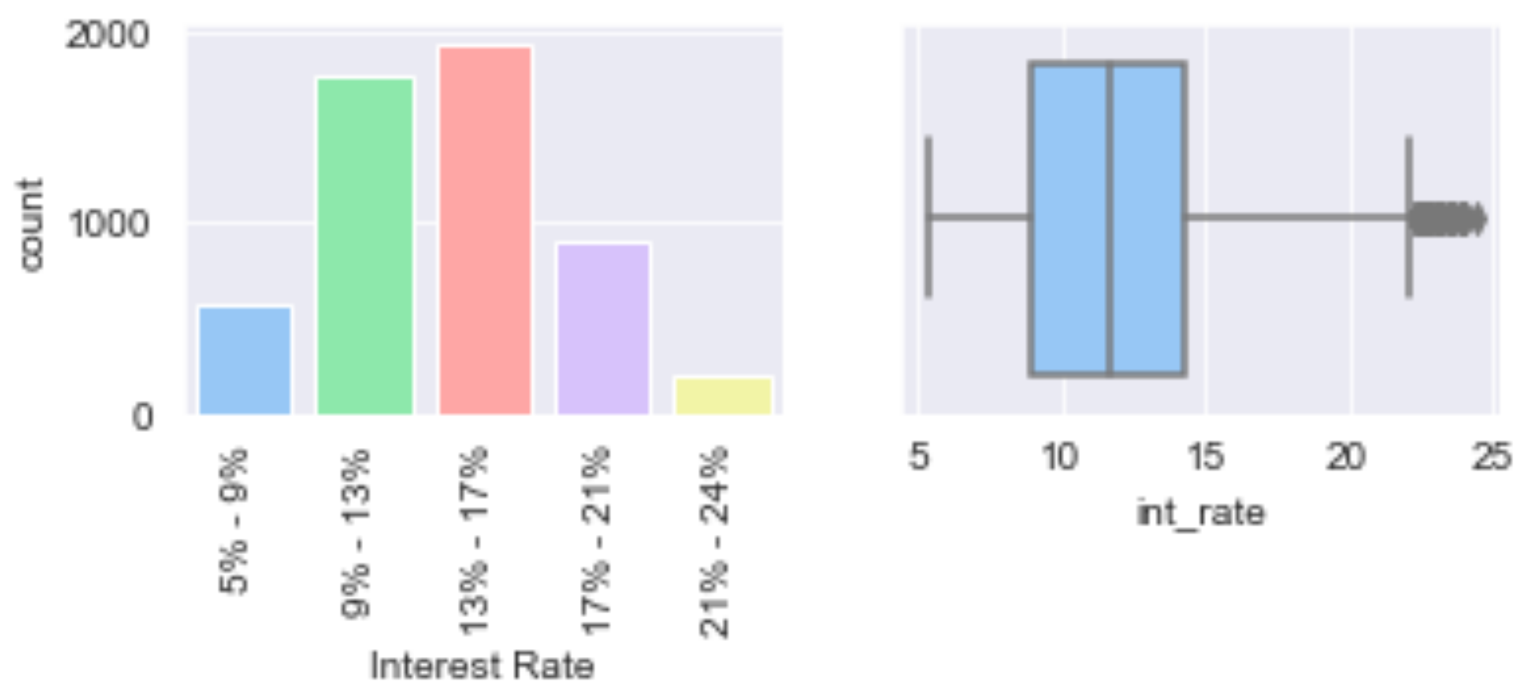
- Perform an analysis to understand the driving factors (or driver variables) behind loan default, i.e.the variables which are strong indicators of default. The company can utilise this knowledge for its portfolio and risk assessment.

Analysis Approach

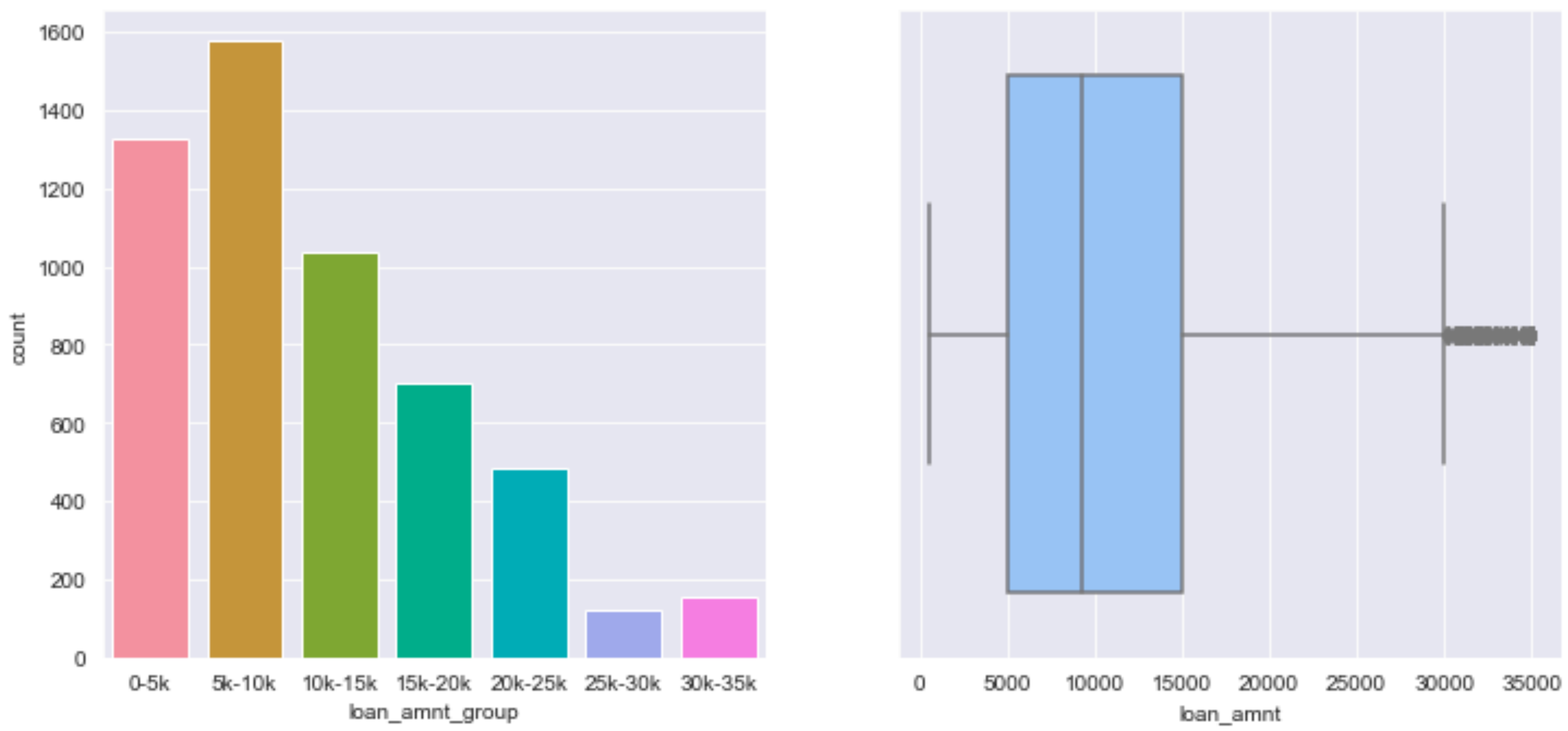
- Data Cleaning
- Univariate Analysis
- Bivariate Analysis

Univariate Analysis on defaulter data

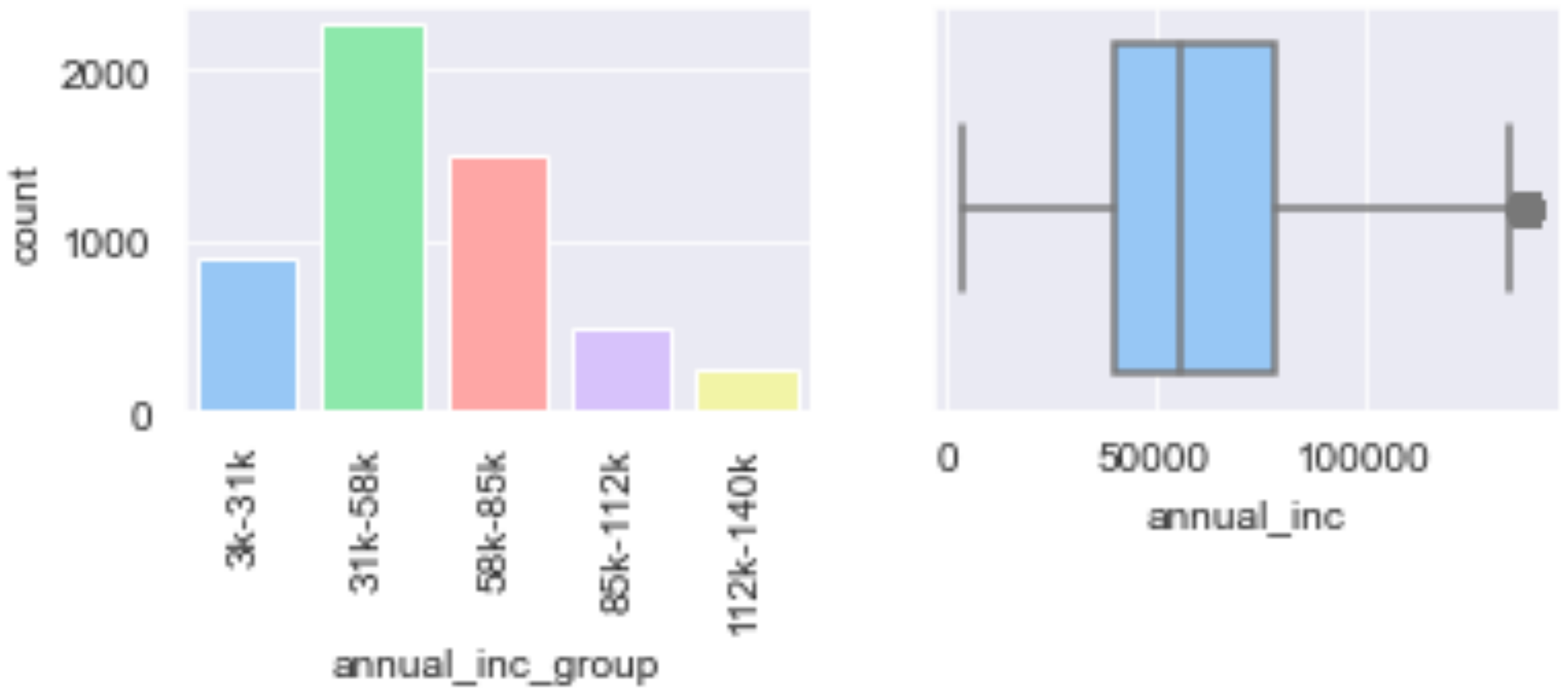
Interest Rate - Average interest rate is 12%. Loans with interest rates ranging from 13% - 17% have the most chances of being charged off, whereas 21% - 24% have the least.



Most of the Loan amounts are in range of 5000 - 15000. Loan amounts of 5k-10k have the most chances of being charged off, whereas 30k-35k have the least.

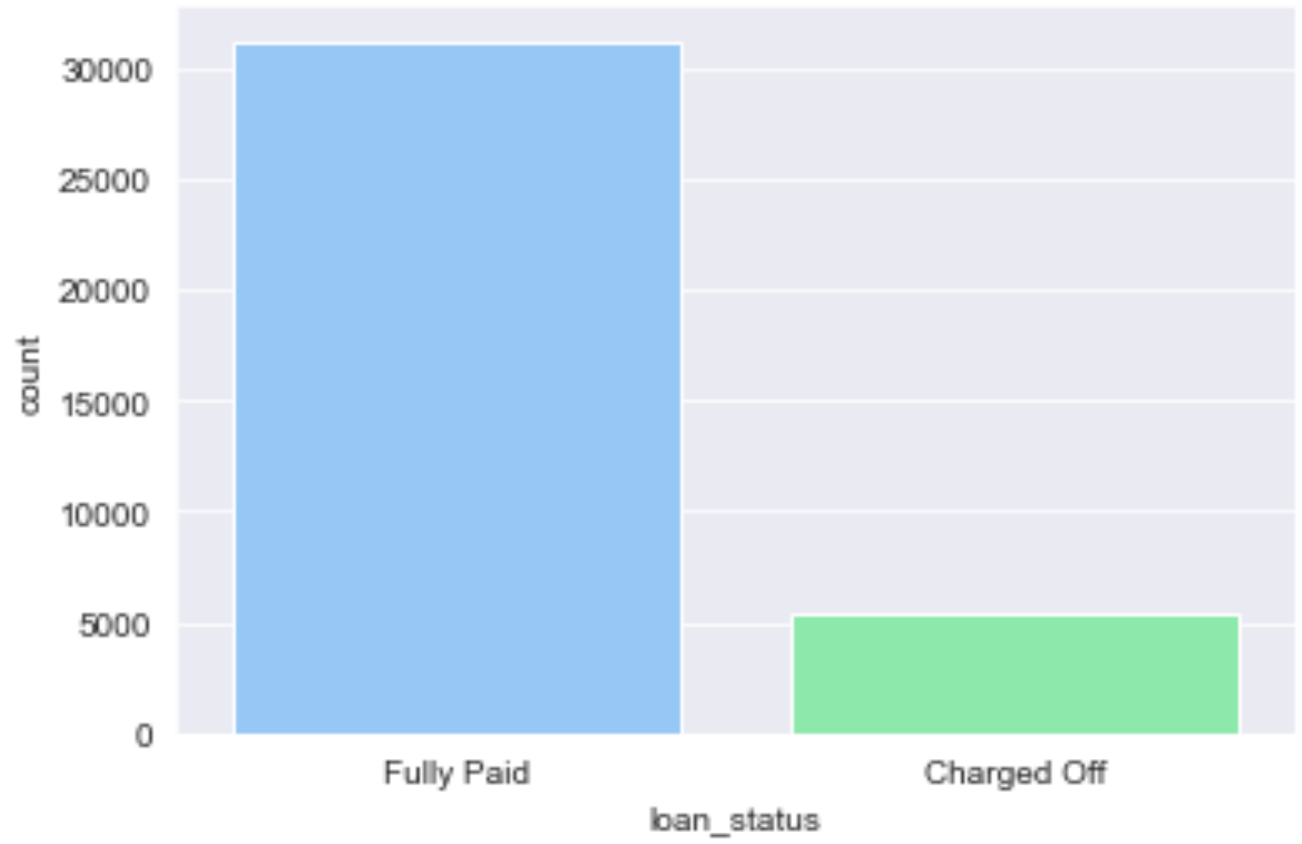


Borrower's Annual incomes are in range of 40000- 80000. Most defaulters are in the income range of 31k - 58k. Incomes ranging from 112k - 140k have the least chances of being charged off.

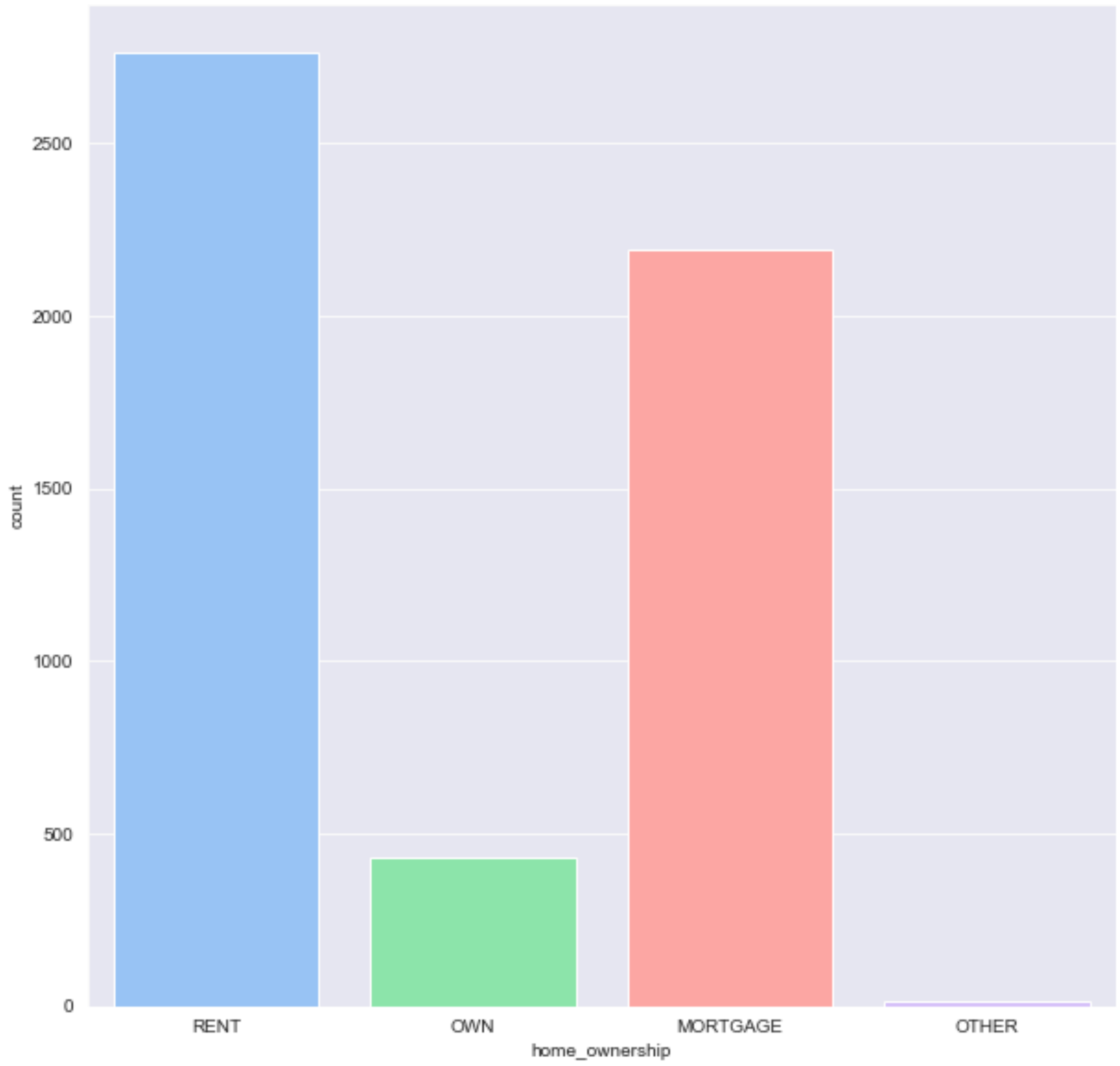


Univariate Analysis

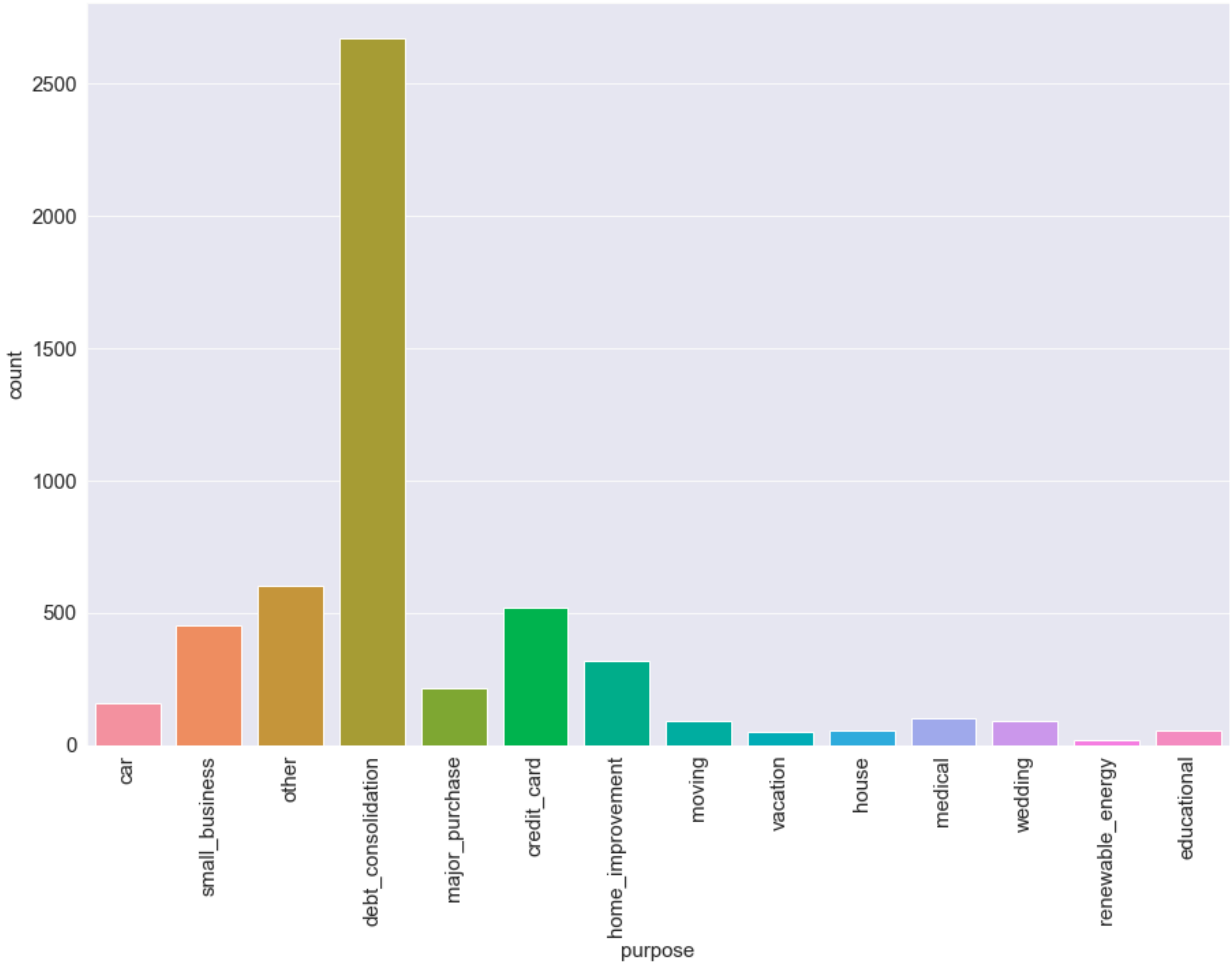
The number of loans charged off is much smaller than the fully paid ones.



Most of them live in rented homes or mortgaged their homes. So, statistically people in rented or mortgaged homes have high chances of being defaulters.

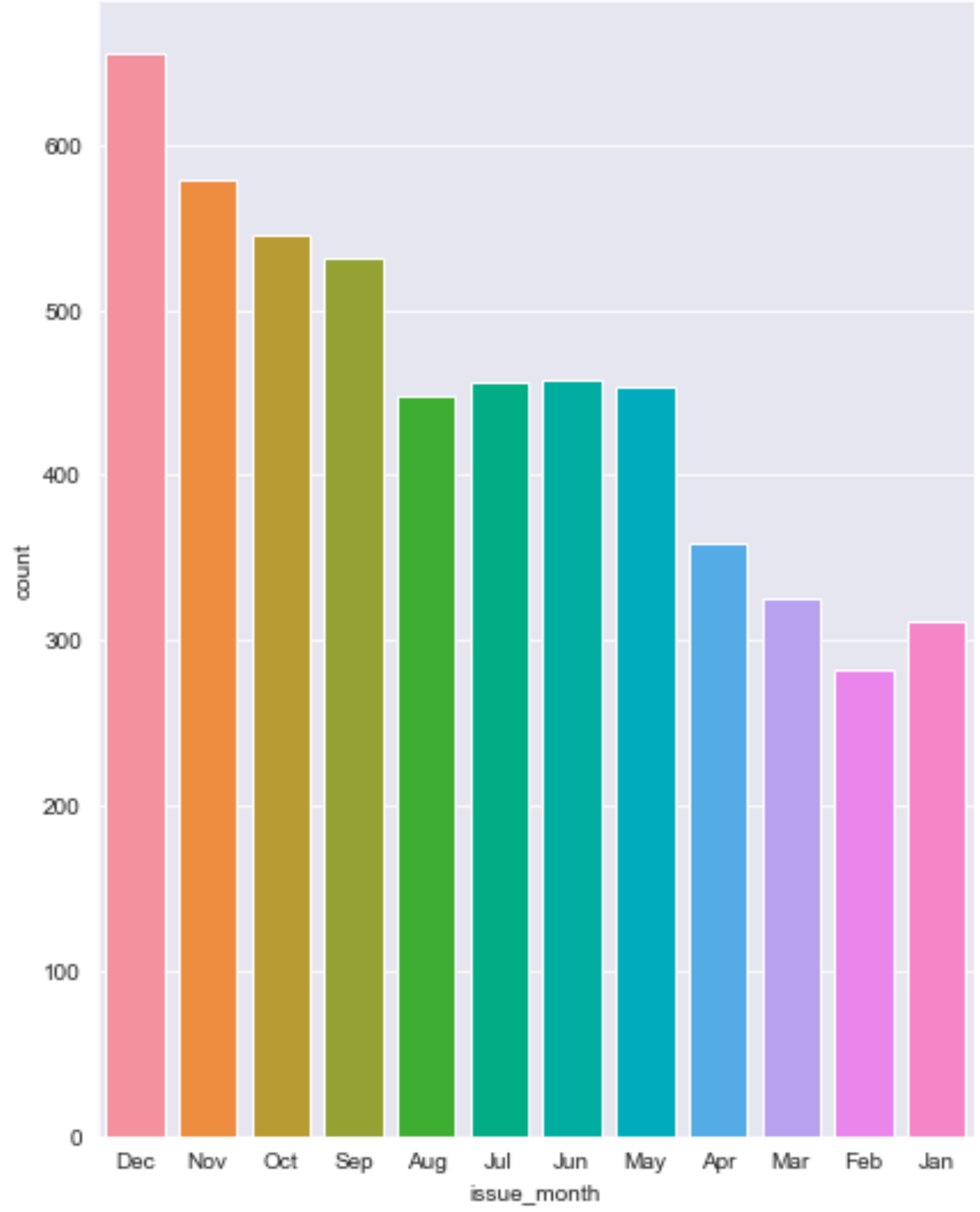


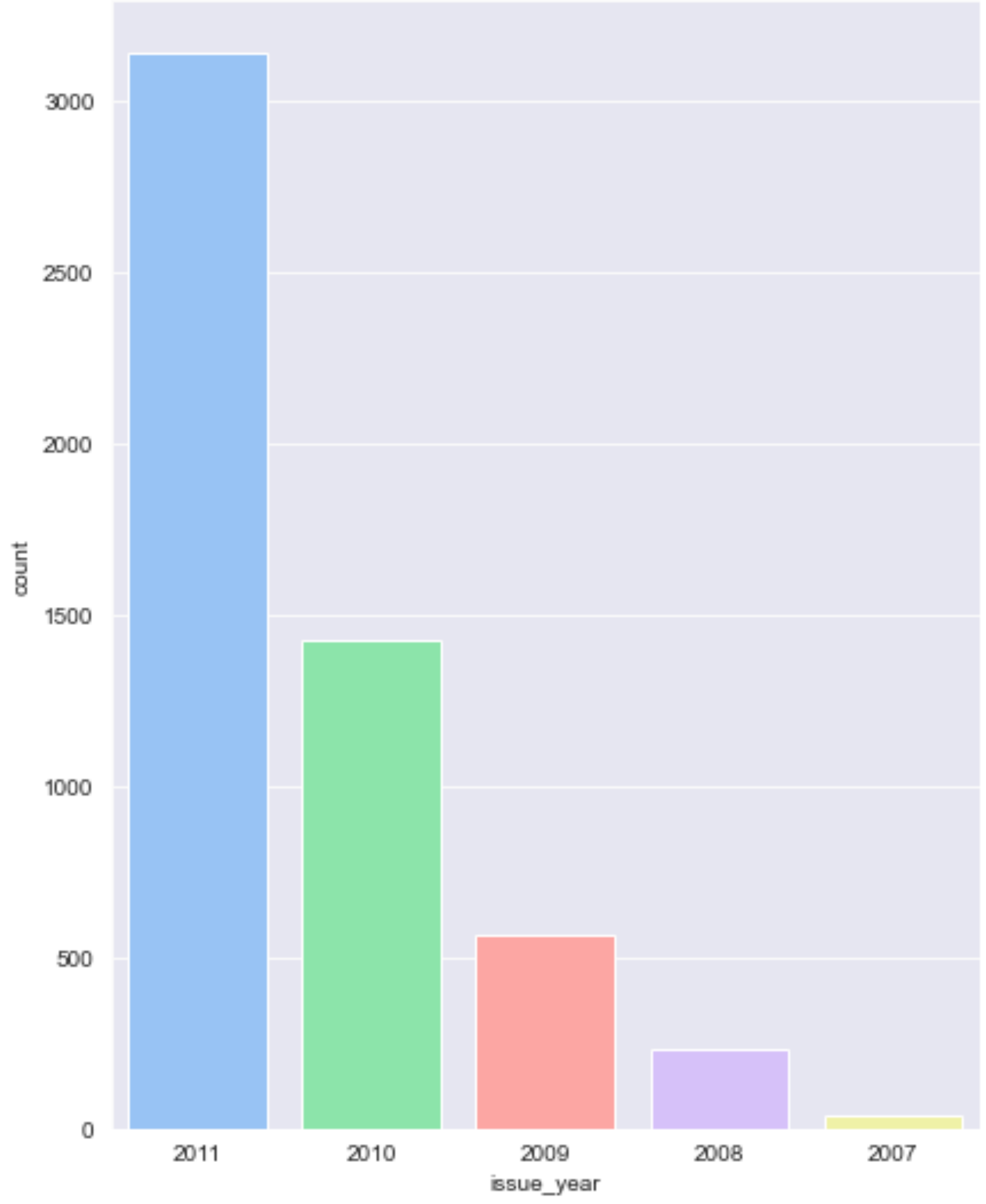
Most of the defaulter loans were taken for the purpose of debt consolidation & paying credit card bills. Loans for renewable energy have the least chances of being charged off.



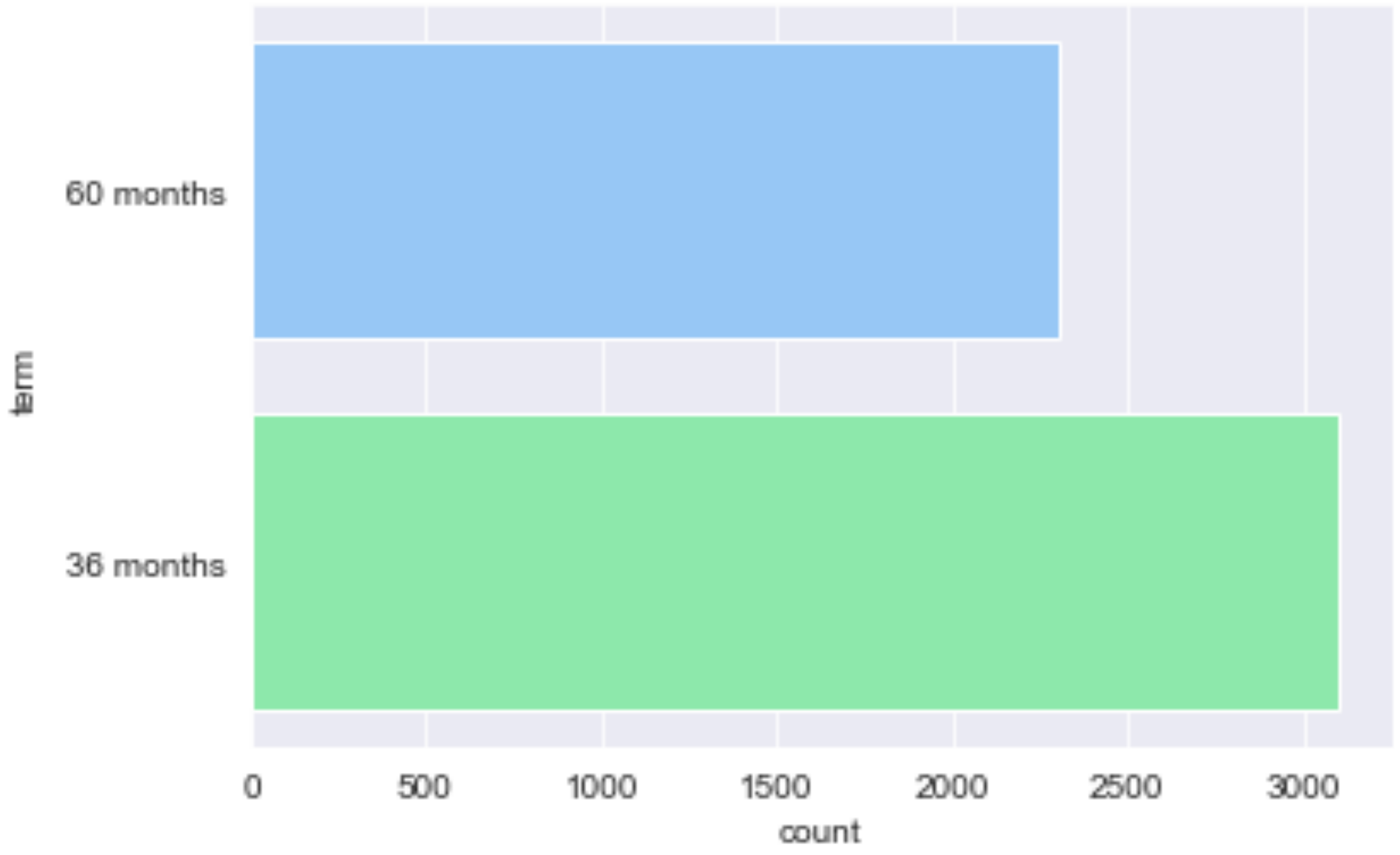
Univariate Analysis on defaulter data

Counts of defaulter loan application is increasing every passing year as well as every month from jan-dec. Maximum defaults occurred for loans issued in the month of December of the year 2011.



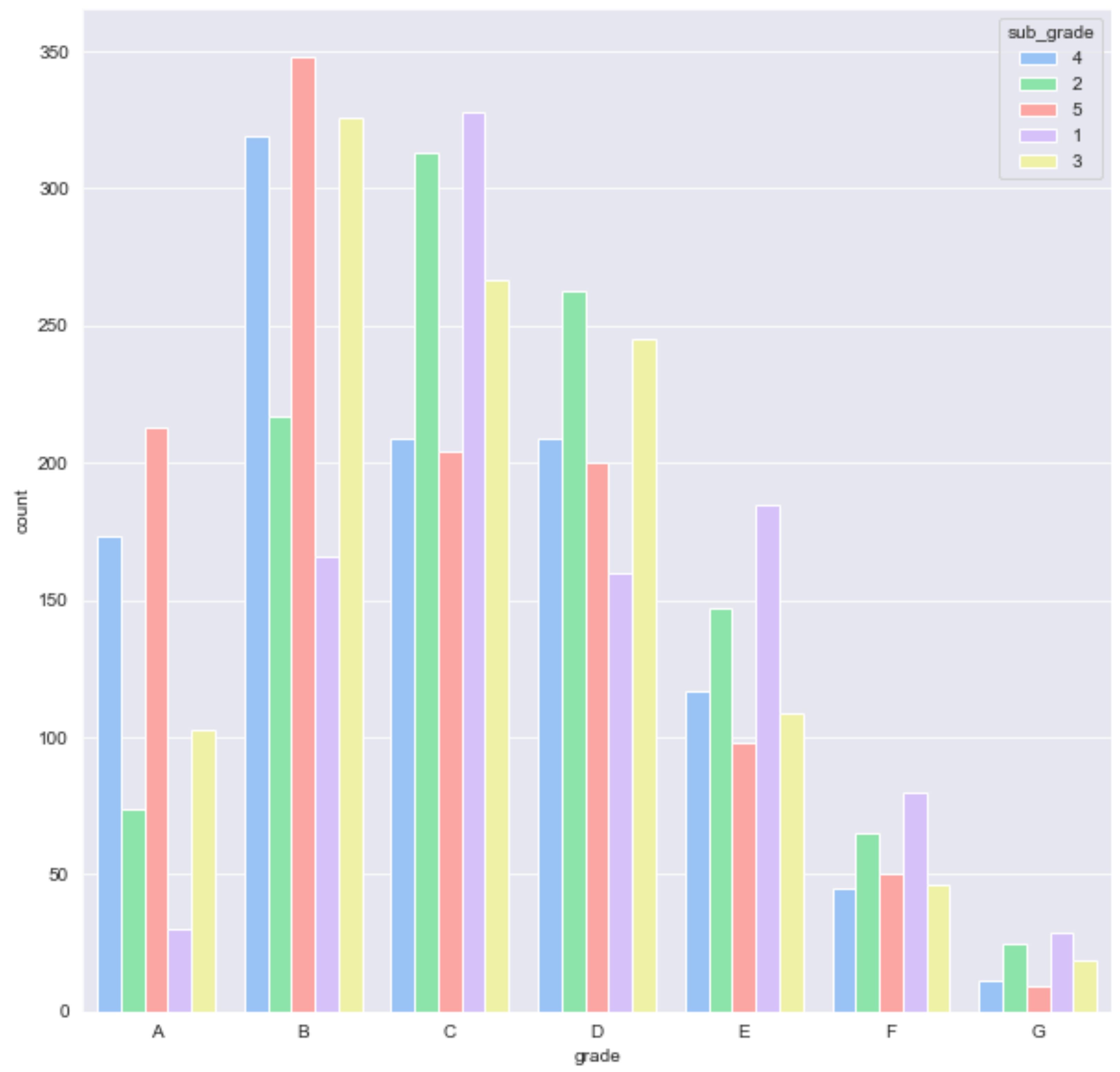


Those who had taken loan to repay in 60 months had more % of number of defaulters as compared to applicants who had taken loan for 36 months.

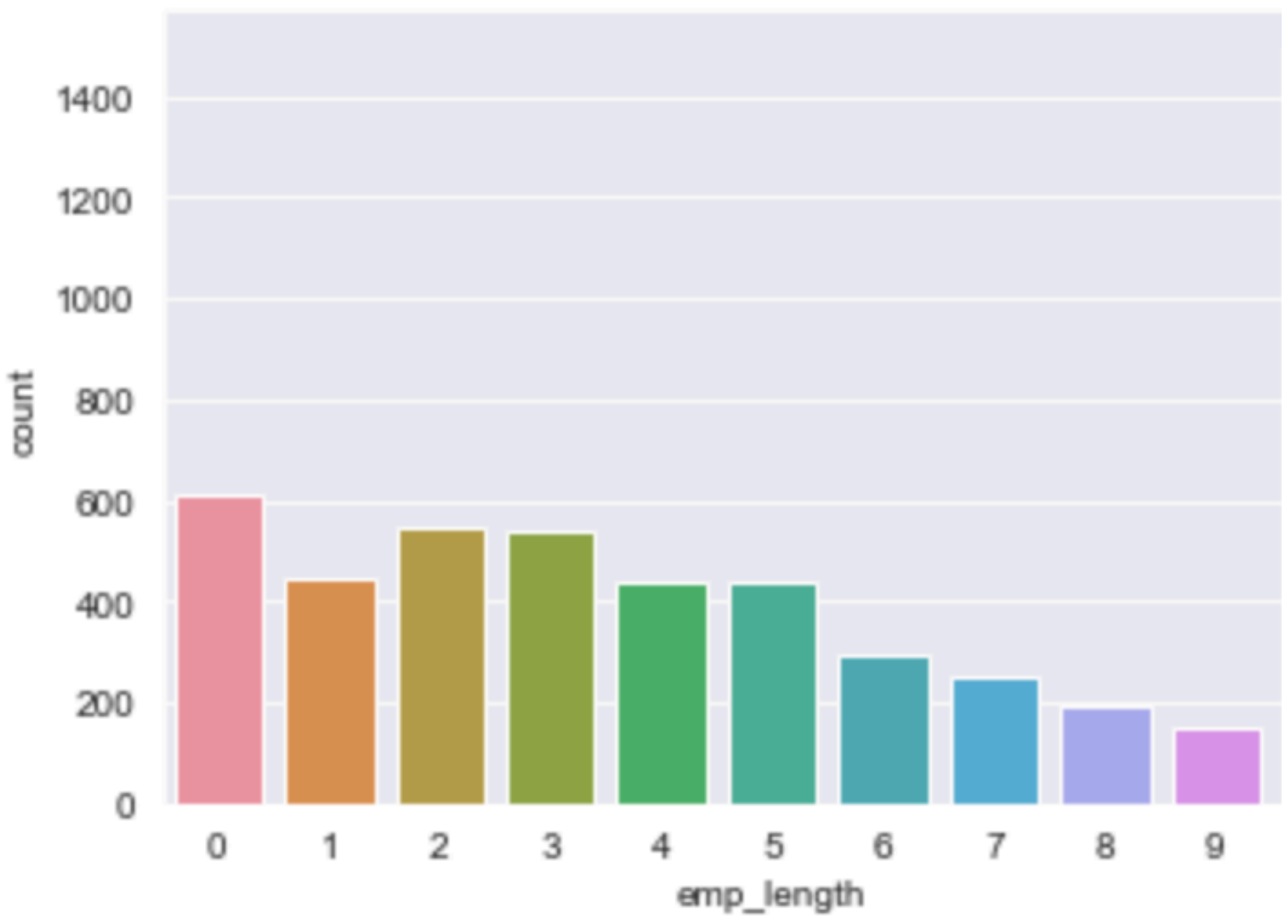


Univariate Analysis on defaulter data

We have maximum number of defaulters in the grade B with sub grade 5.

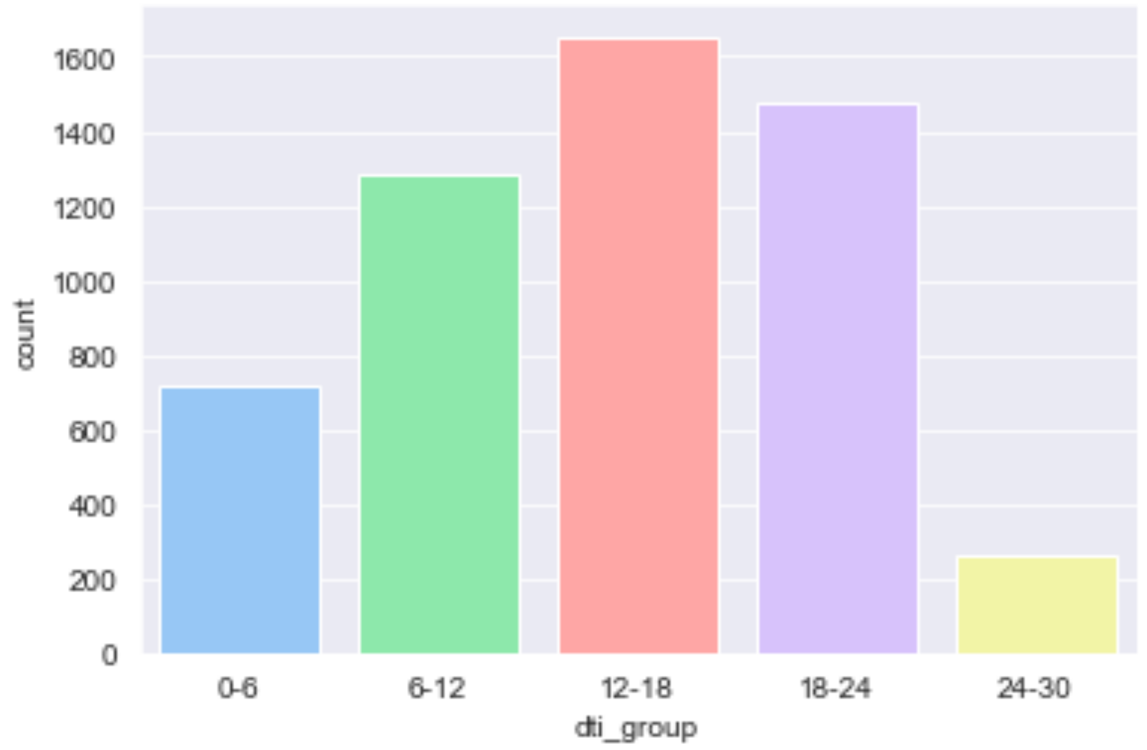


Those who are not working or have less than 1 year of work experience have high chances of getting charged off. It makes sense as with less or no experience they don't have source of income to repay loan. Rest of the applicants have more or less same chances of getting charged off.

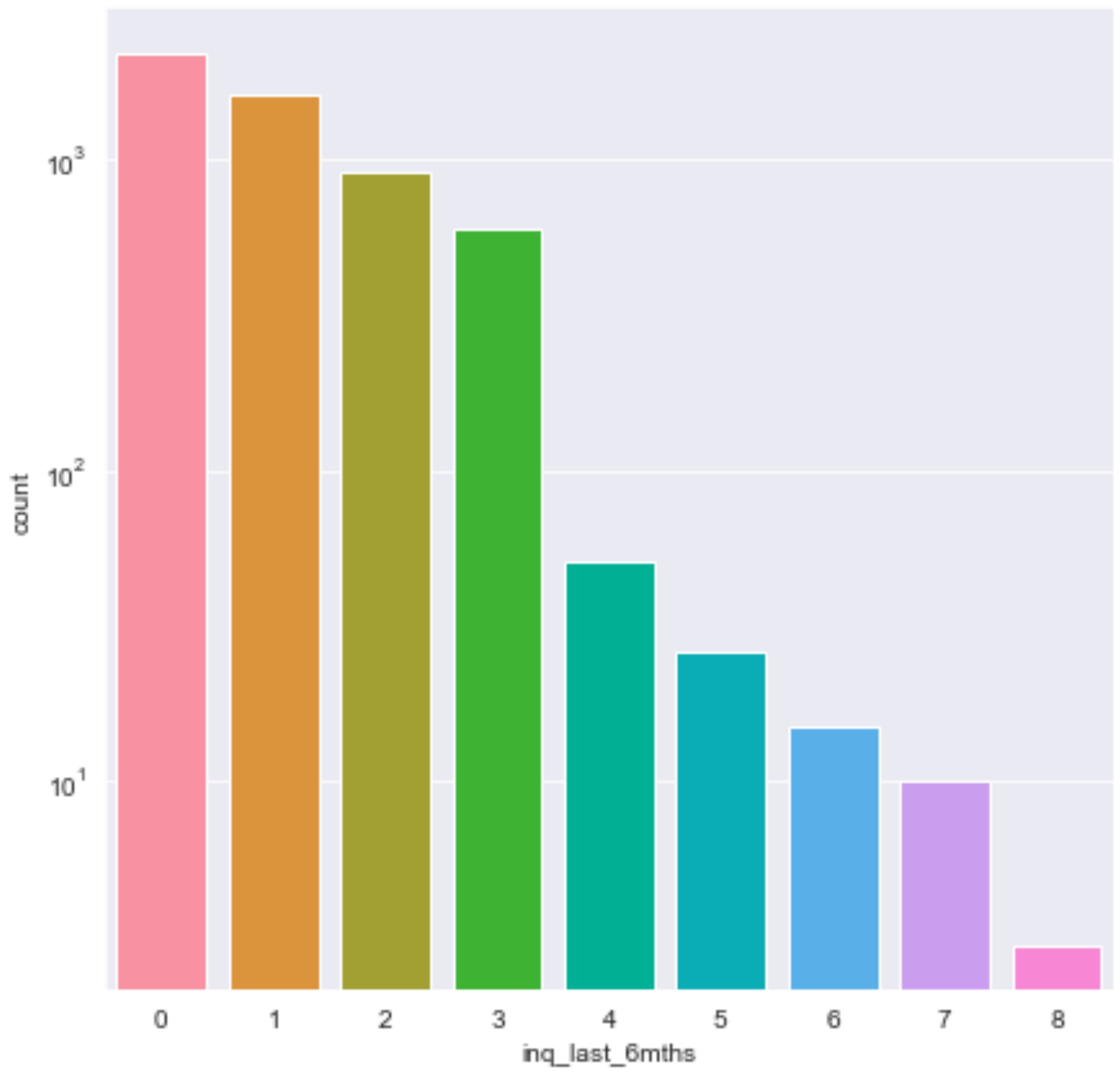


Univariate Analysis on defaulter data

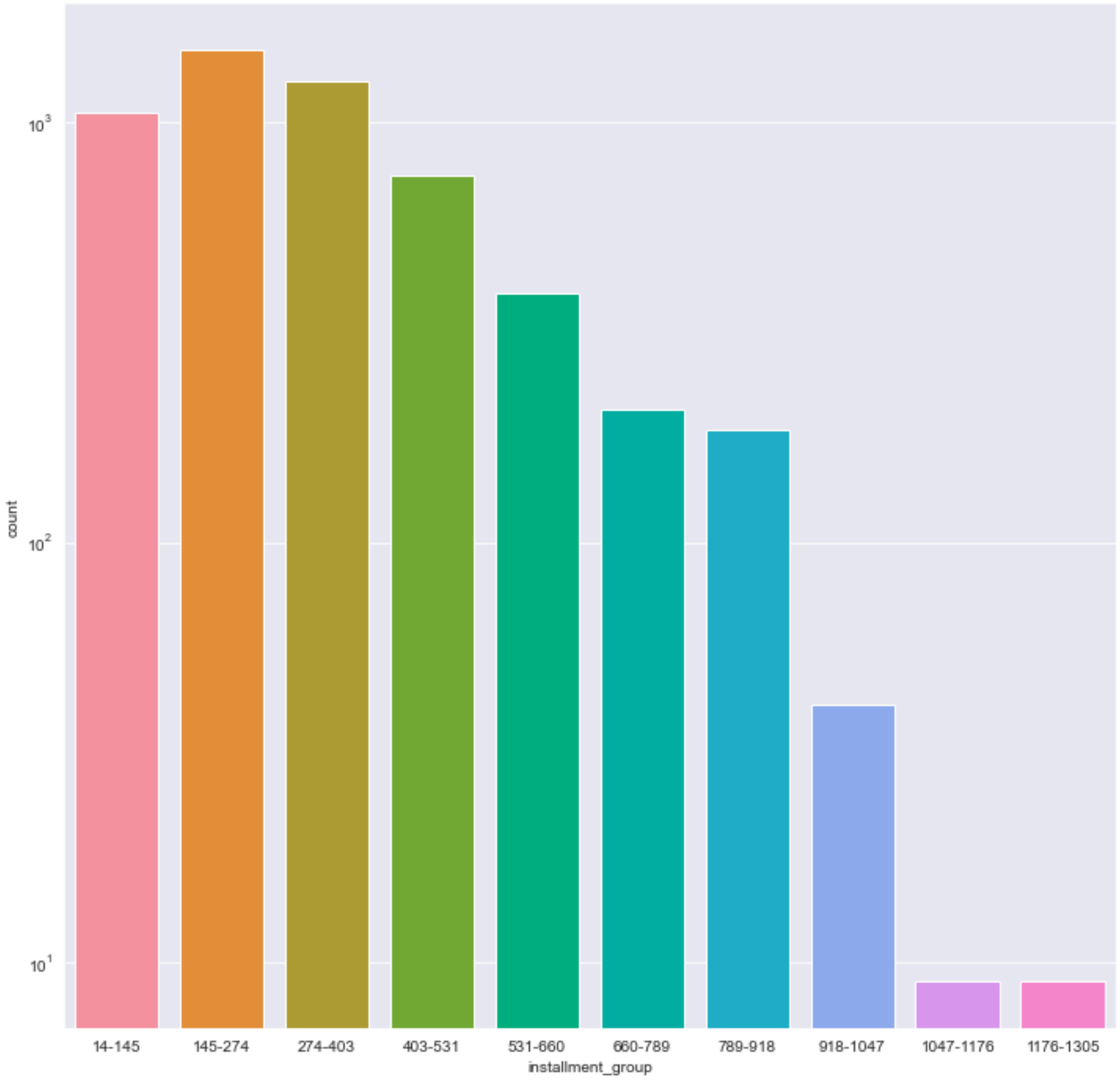
Debt income ratio i.e. dti in the range of 12-18 has the most chances of defaulting.



Number of inquiries in last 6 months is 0 have higher cases of defaulters.

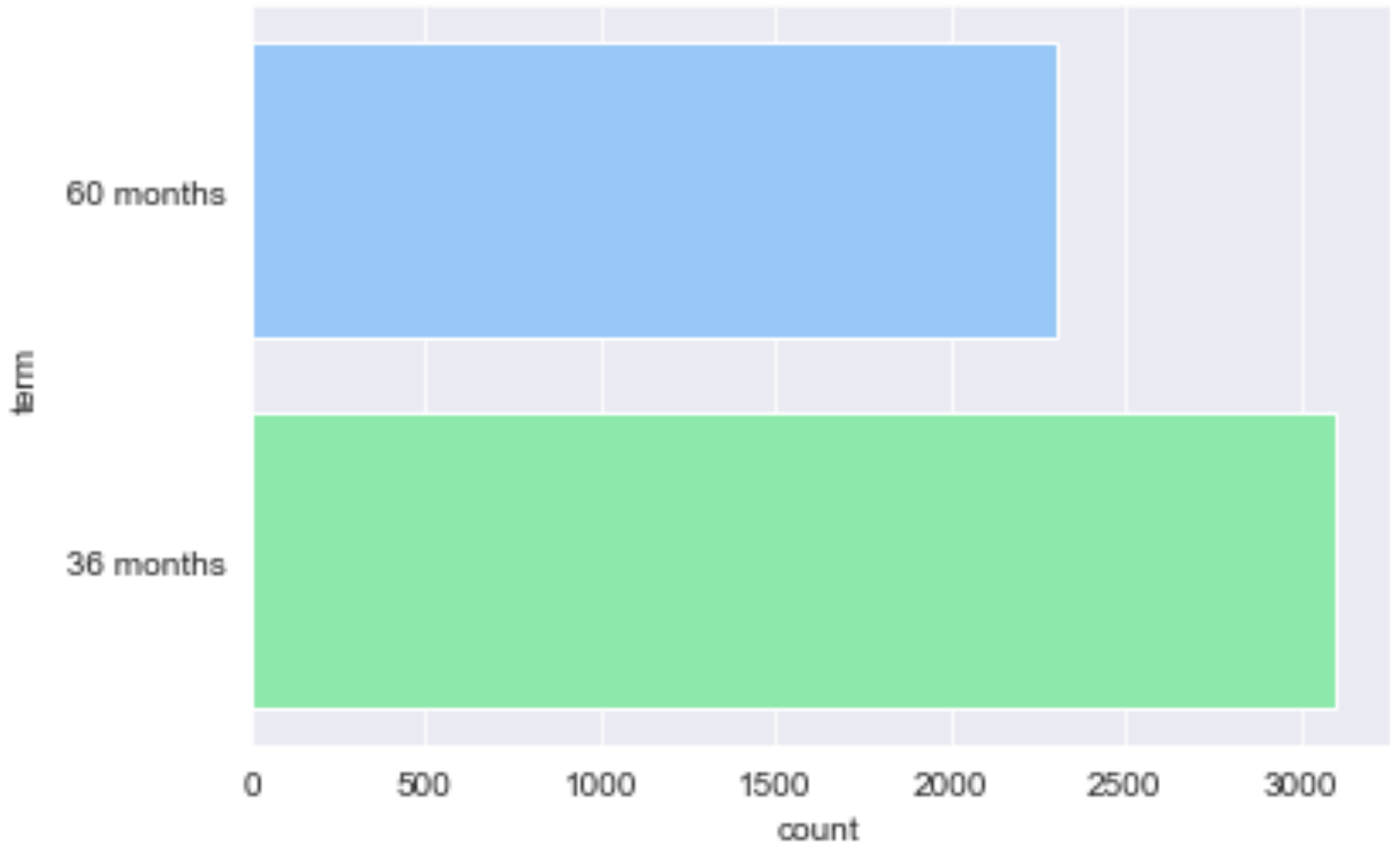


Installment i.e. The monthly payment owed by the borrower was 145 - 274, have higher cases of defaulting.

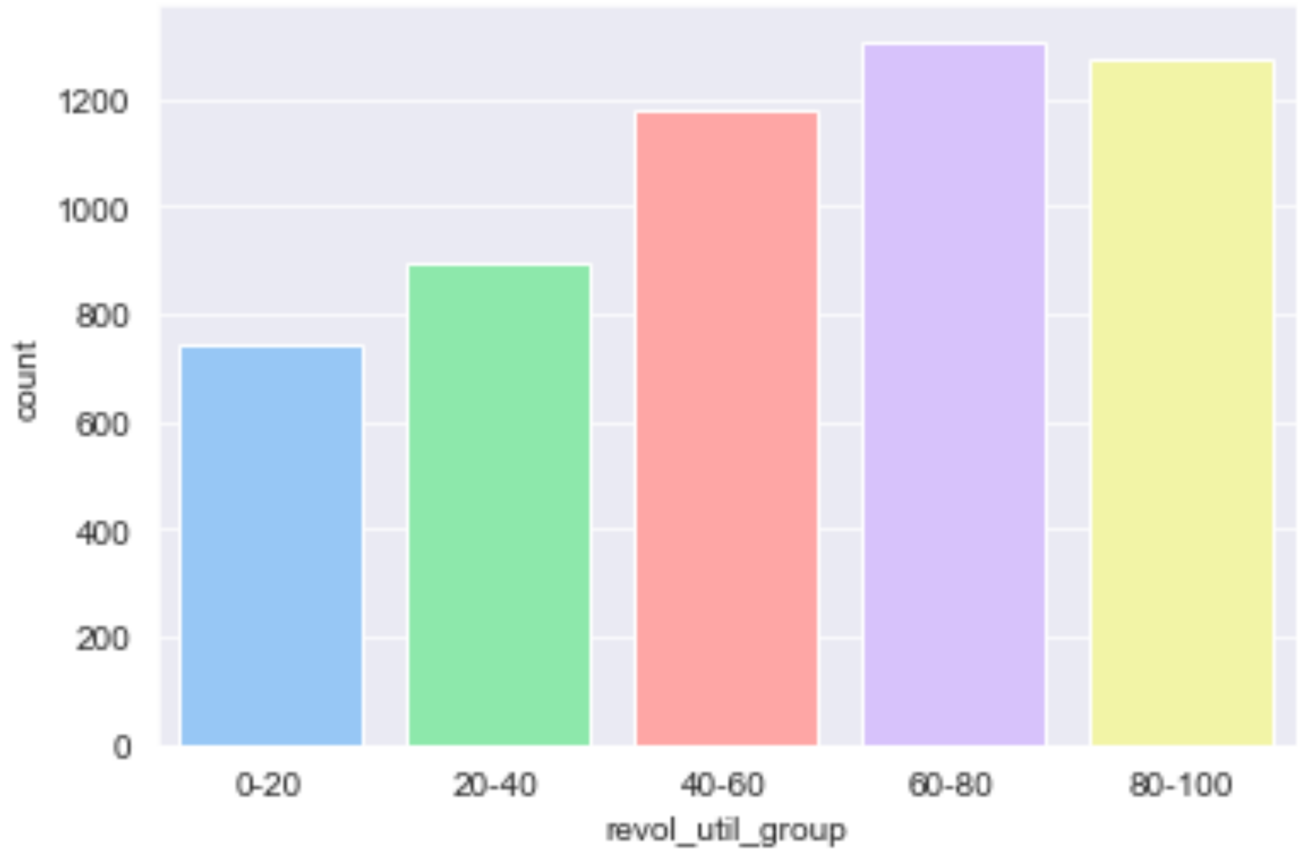


Univariate Analysis on defaulter data

Term i.e. the number of payments on the loan of loan is 36 months have higher cases of defaulters.



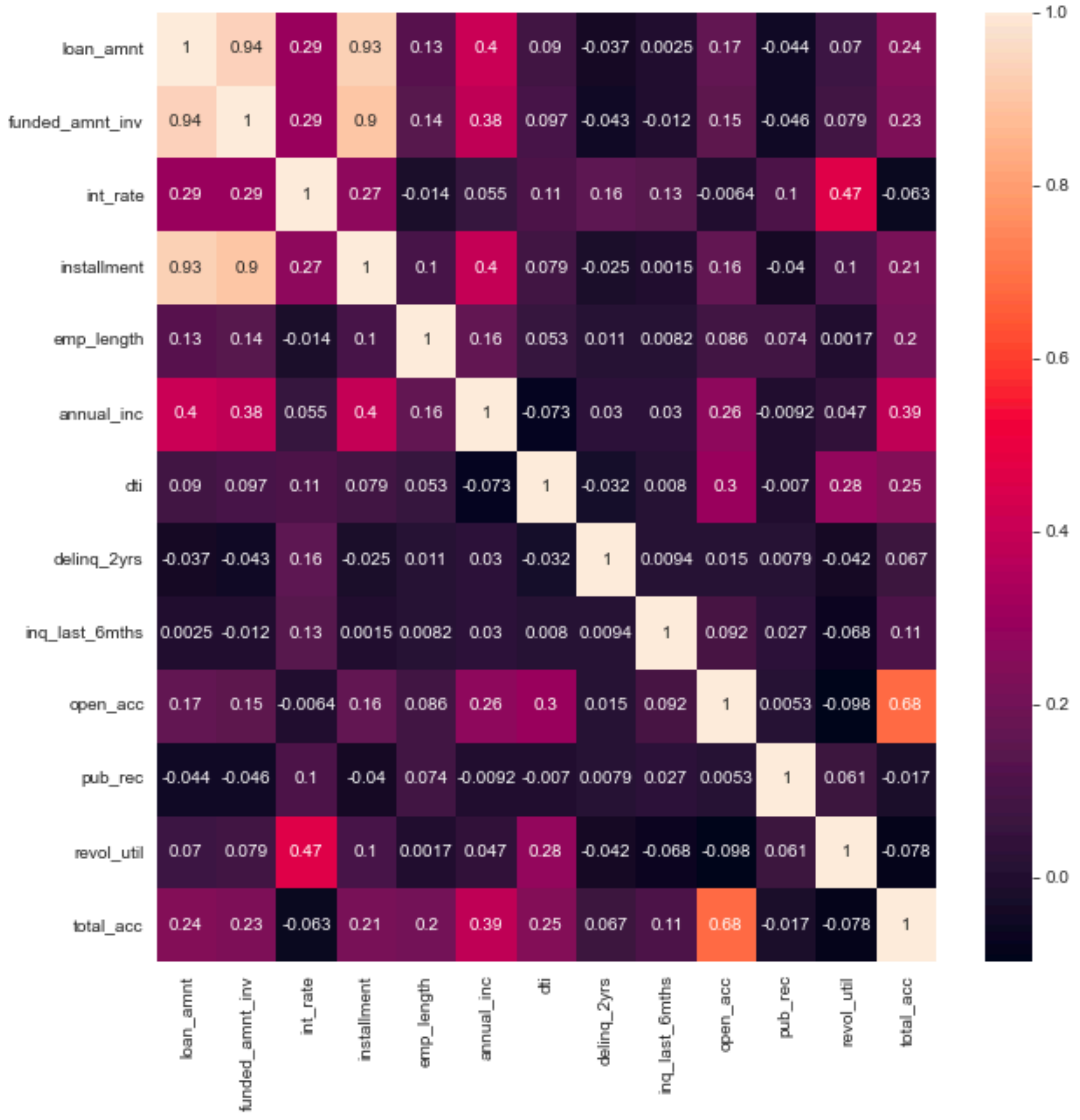
Revolving line utilization rate is 60-80 has higher cases of defaulters, but it is not very conclusive as the other values are also close.



Verification status is not conclusive as the values are quite close here as well.

Bivariate Analysis

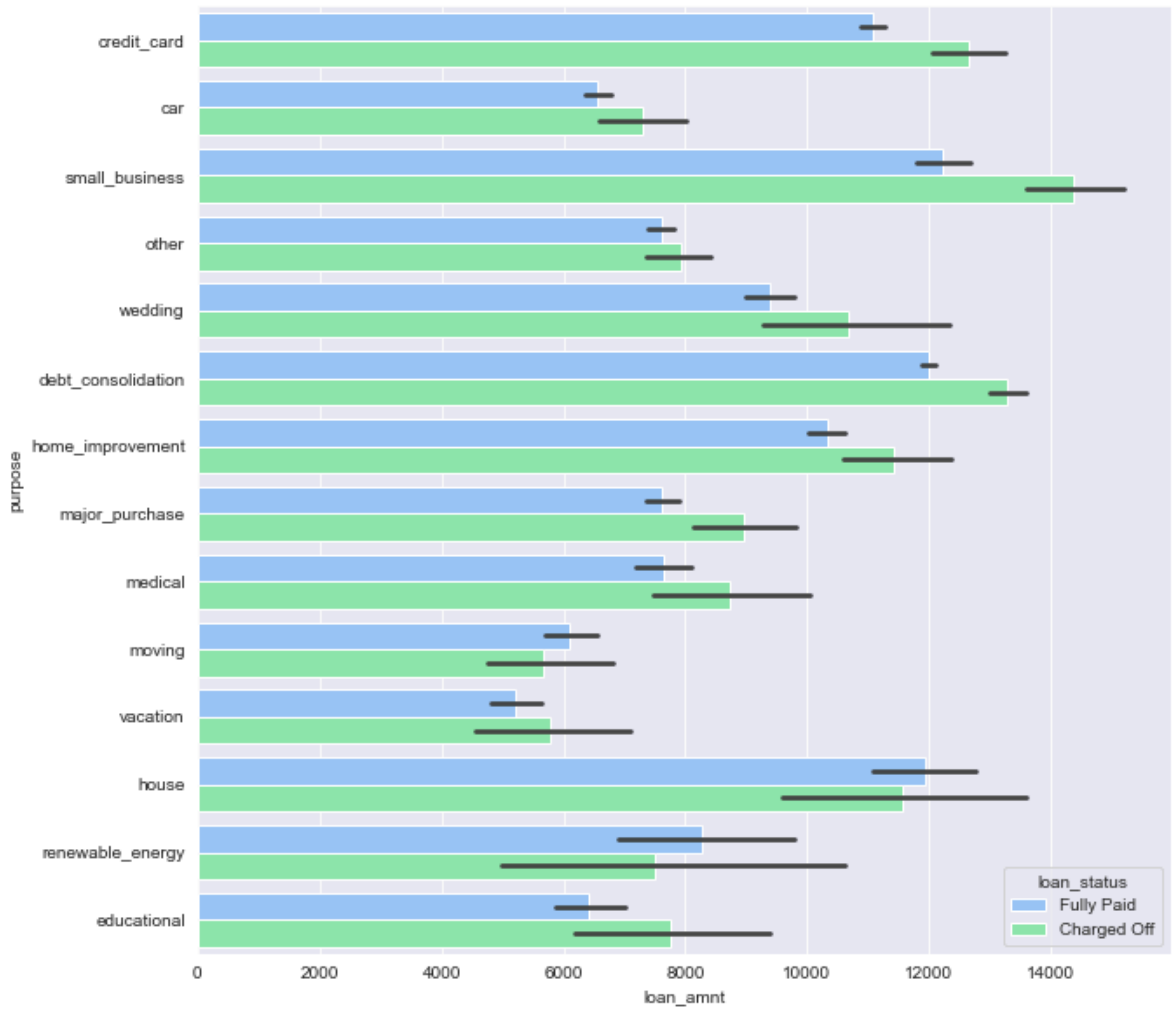
- Observation is that Loan amount, funding amount by investor are strongly correlated.
- Annual income with DTI(Debt-to-income ratio) is negatively correlated.
- Debt income ratio is the percentage of a consumer's monthly gross income that goes toward paying debts i.e. that means when annual income is low DTI is high & vice versa.
- Positive correlation between annual income and employment years i.e. that means income increases with work experience.



Correlation Heat Map

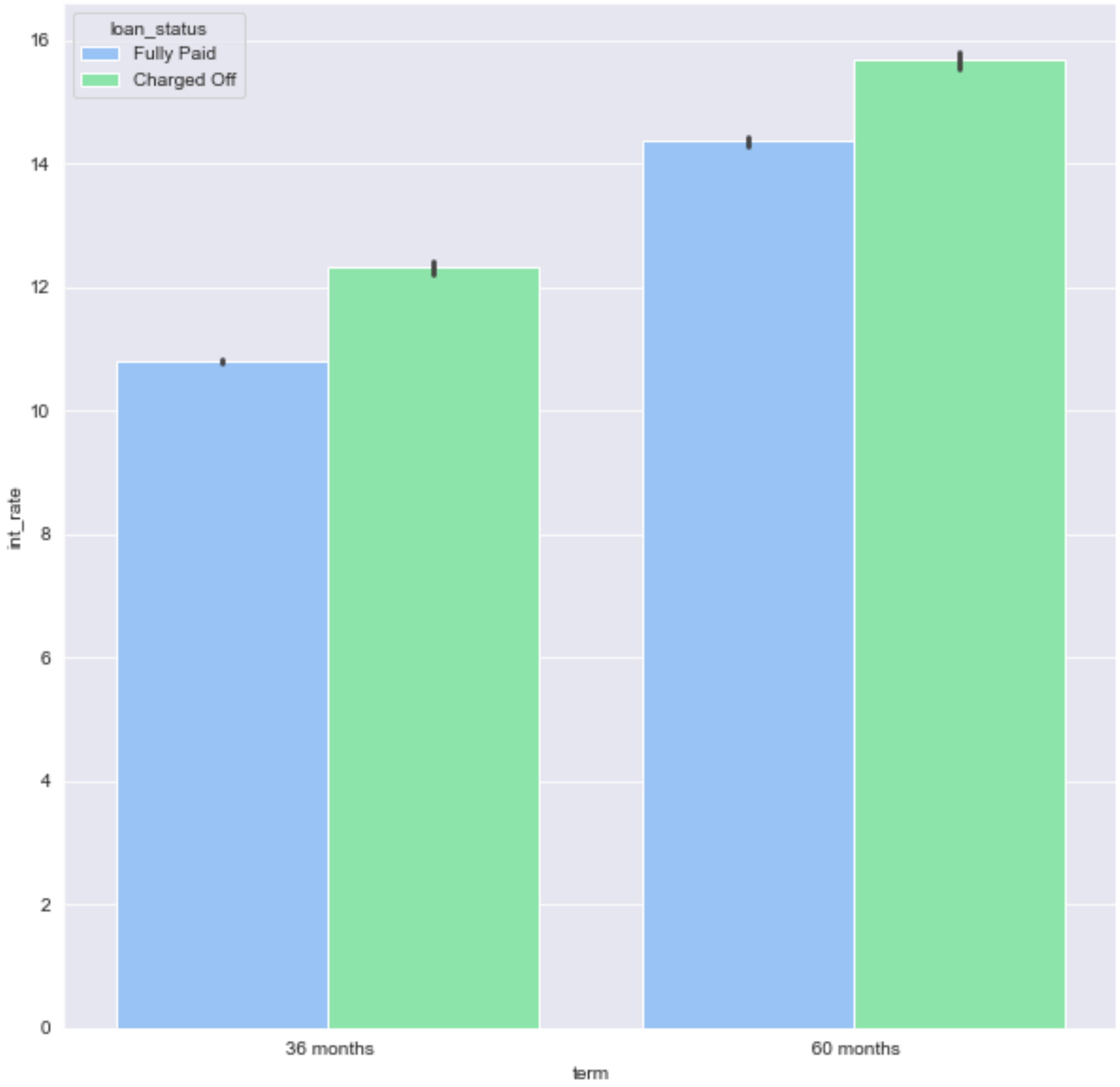
Bivariate Analysis

Loan Amount vs Loan Purpose



Highest loan amounts are for small business, debt consolidation & houses. The defaulters all have higher loan amounts compared to the people who paid off their loans across all categories except moving, house & renewable energy.

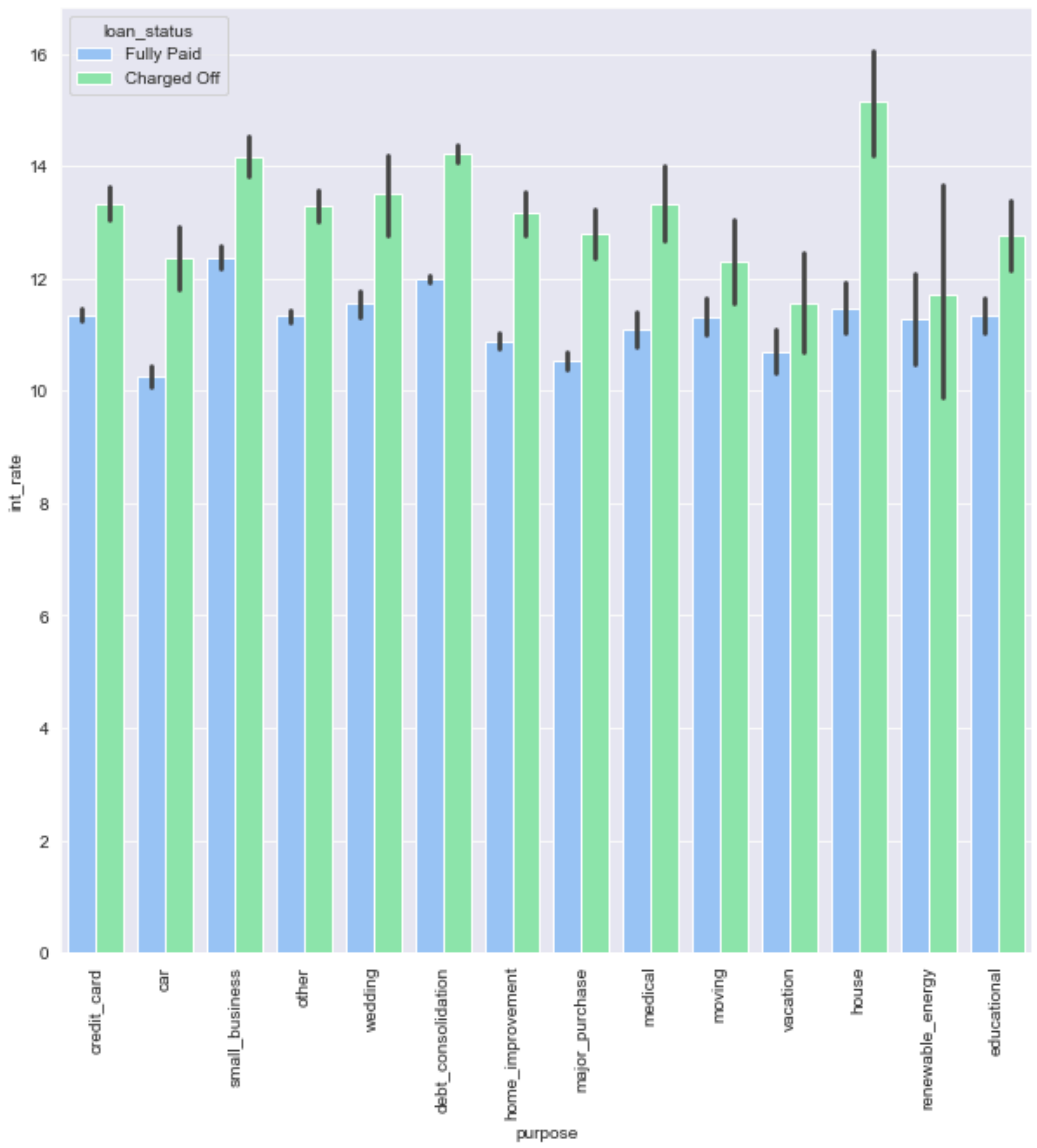
Term vs Interest Rate



It is clear that average interest rate is higher for 60 months loan term. Most of the loans issued for longer term had higher interest rates for repayment.

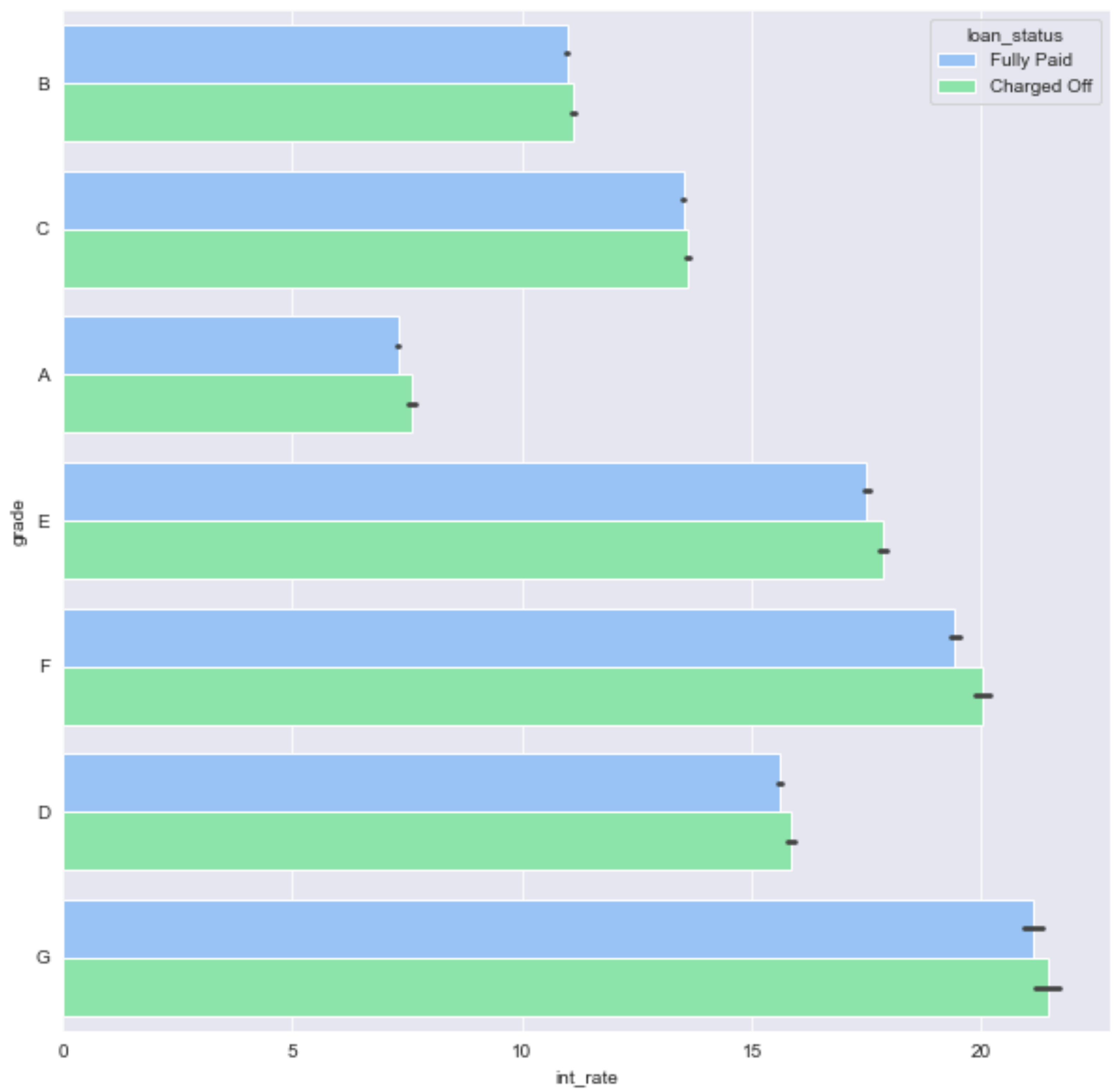
Bivariate Analysis

Purpose vs Interest Rate



It is clear that average interest rate is highest for house, small business purpose & debt consolidation. These are cases where most defaulters are as well.

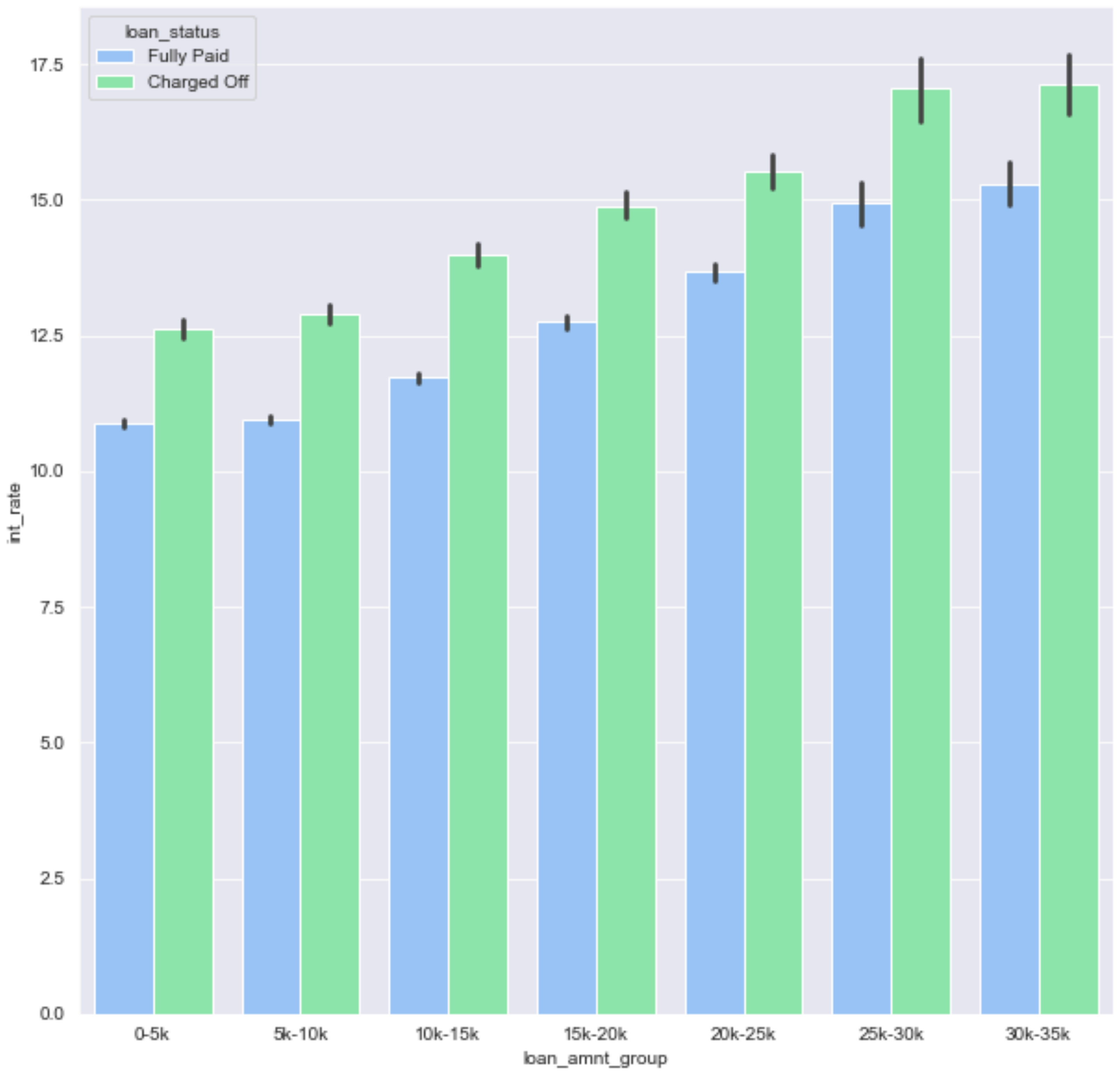
Grade vs Interest Rate



A-grade is a top letter grade for a lender to assign to a borrower. The higher the borrower's credit grade, the lower the interest rate offered to that borrower on a loan. It is clear that interest rate is increasing with grades moving from A to G. Hence defaulters also increasing.

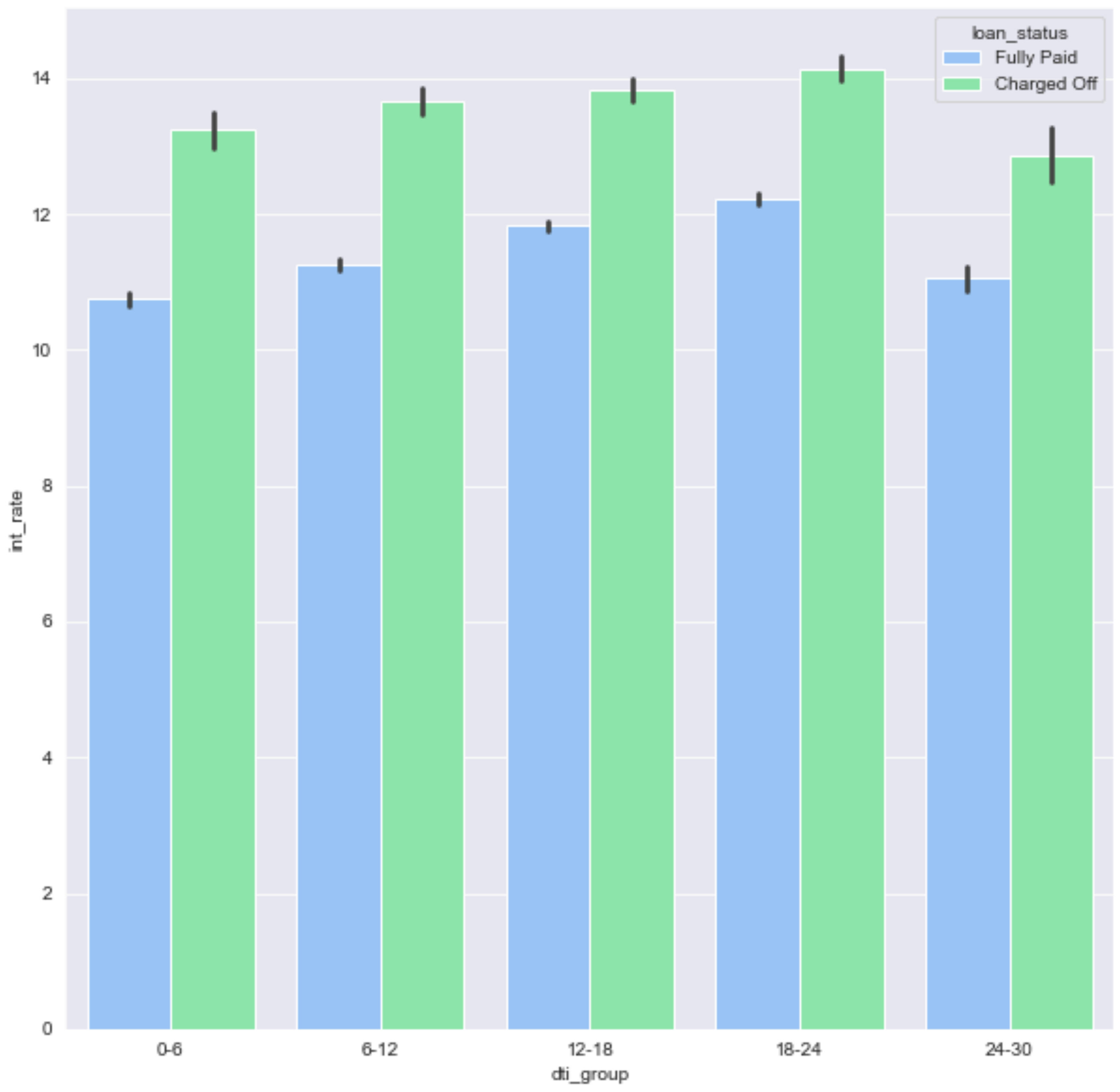
Bivariate Analysis

Loan Amount vs Interest Rate



Higher loan amounts mean higher interest rates and within an interest rate group the defaulters are ones with the higher interest rate.

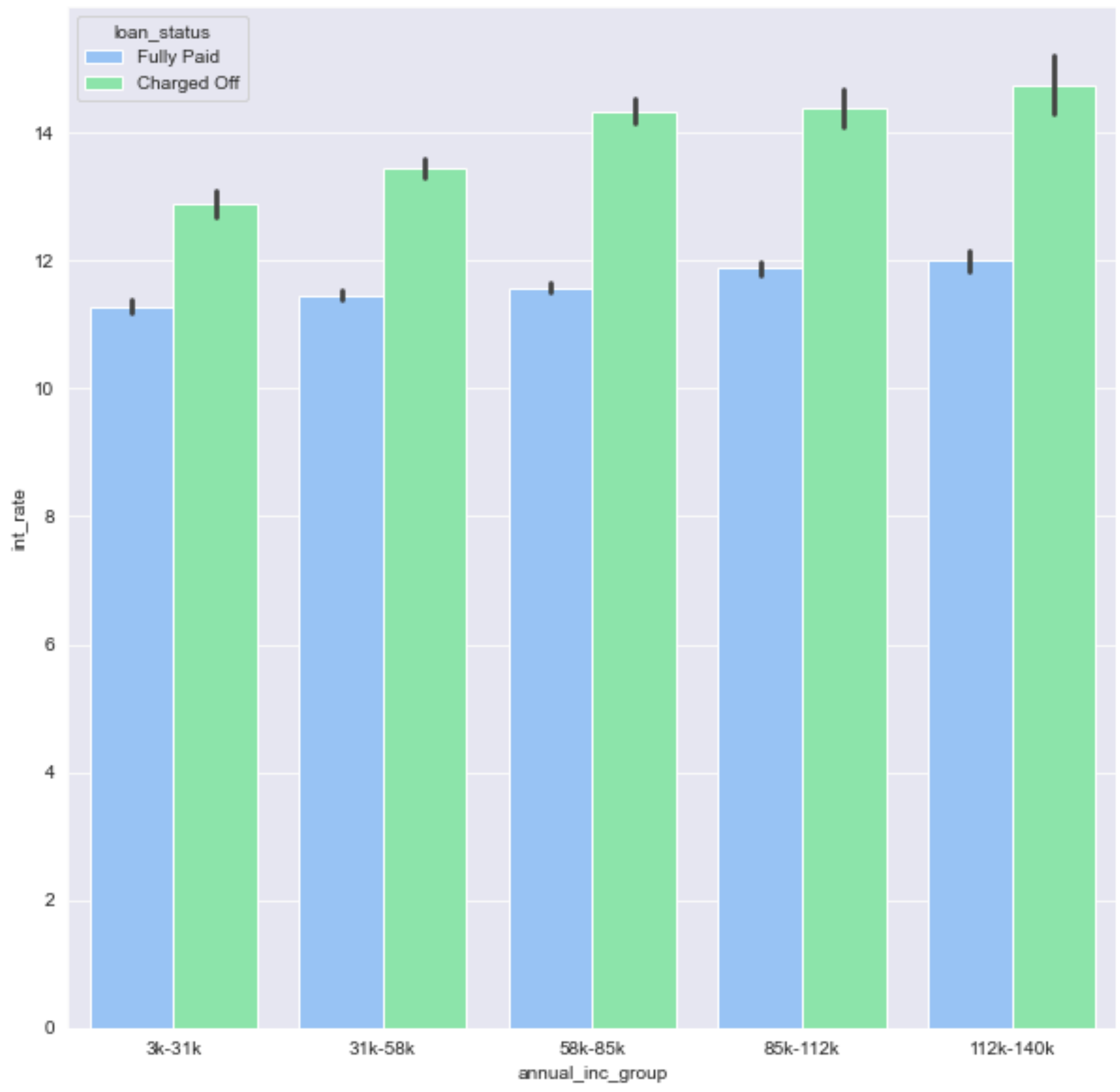
DTI vs Interest Rate



If your DTI is low enough you may get a lower interest rate. Plot shows no significant variation but there is slight increase in interest rate with increase in DTI.

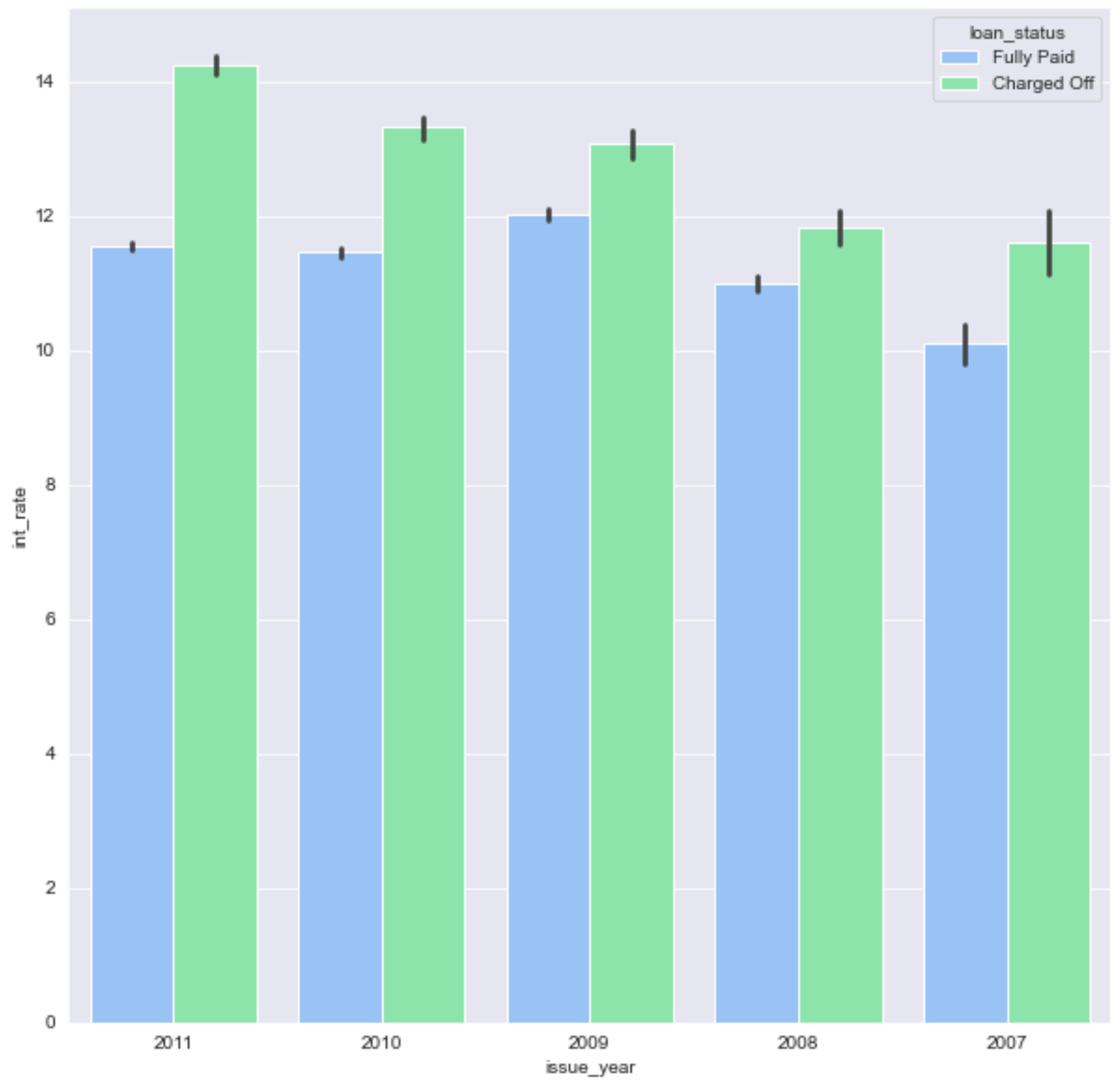
Bivariate Analysis

Annual Income vs. Grade



The ones getting 'charged off' have lower annual incomes than the ones who 'paid fully' for each and every grade (i.e. at same interest range).

Year vs Interest Rate



Interest rate is increasing slowly with increase in year.

Summary

- Number of loans defaulted along with their interest rates have been going up through the years.
- Maximum defaults occurred for loans issued in the month of December of the year 2011.
- Lower grade loans have much higher chances of defaulting compared to others. Loan amounts of 5k-10k have the most chances of being charged off, whereas 30k-35k have the least. We have maximum number of defaulters in the grade B with sub grade 5.
- Loans with interest rates ranging from 13% - 17% have the most chances of being charged off, whereas 21% - 24% have the least.
- Lower income means higher chances of defaulting. Most defaulters are in the income range of 31k - 58k. Incomes ranging from 112k - 140k have the least chances of being charged off.
- Statistically people in rented or mortgaged homes have high chances of being defaulters.
- Most of the defaulter loans were taken for the purpose of debt consolidation & paying credit card bills. Loans for renewable energy have the least chances of being charged off.
- Those who had taken loan to repay in 60 months had more % of number of defaulters as compared to applicants who had taken loan for 36 months.
- Those who are not working or have less than 1 year of work experience have high chances of getting charged off. It makes sense as with less or no experience they don't have source of income to repay loan. Rest of the applicants have more or less same chances of getting charged off.
- A-grade is a top letter grade for a lender to assign to a borrower. The higher the borrower's credit grade, the lower the interest rate offered to that borrower on a loan. It is clear that interest rate is increasing with grades moving from A to G. Hence defaulters also increasing.
- It is clear that average interest rate is highest for house, small business purpose & debt consolidation. These are cases where most defaulters are as well.
- The ones getting 'charged off' have lower annual incomes than the ones who 'paid fully' for each and every grade (i.e. at same interest range).
- But luckily, the number of loans charged off is much smaller than the fully paid ones.