LENDING CLUB CASE

Ranjit Kumar Yadav

AGENDA

General Information

Data Cleaning

Checking for Missing Values

Standardizing the Data

Finding Outliers

Visualizing the Data Using:

- Univariate Analysis
- Bi-Variate Analysis

GENERAL INFORMATION

The company Lending Loans is a major online loan marketplace that facilitates personal loans, business loans, and medical procedure financing through a user-friendly online platform, providing borrowers with access to lower-interest-rate loans.

Similar to many lending firms, the primary financial risk is associated with lending to applicants considered 'risky,' resulting in credit loss when borrowers fail to repay. In this context, 'charged-off' customers are labeled as 'defaulters,' causing the most significant financial loss.

The focus of this case study is to employ Exploratory Data Analysis (EDA) to identify factors influencing loan default, aiming to reduce risky loans and mitigate credit loss.

The company seeks to understand the key variables or indicators contributing to loan default, with the intention of utilizing this knowledge for portfolio management and risk assessment.

DATA CLEANING

- In the loan data sheet there multiple fields with null values, we need to process these null value as a part of the data cleaning. These fields don't provide any information related to the loan and in any kind they are not use full to get any information, so removing these data from the data set
- There is no member have taken the more than 1 loan, so we can remove this column as this column don't provide any information regarding the loan defaulters
- Also removing the following columns: 'policy_code','delinq_amnt','tax_liens','pub_rec_bankruptcies','collections_12_mths_ex_med','acc_now_delinq','pub_rec_bankruptcies', "url", "title", "emp_title", "zip_code", "last_credit_pull_d", "addr_state","out_prncp_inv","total_pymnt_inv","funded_amnt", "delinq_2yrs", "revol_bal", "out_prncp", "total_pymnt", "total_rec_prncp", "total_rec_int", "total_rec_late_fee", "recoveries", "collection_recovery_fee", "last_pymnt_d", "last_pymnt_amnt", "next_pymnt_d" , "chargeoff_within_12_mths", "mths_since_last_delinq", "mths_since_last_record"
- According to the request made, we have to analyze the significant information about who is likely to be a defualters. So, to find this we don't require the records for the customers who is not a defaulter and not yet fully paid, i.e. the status 'current'.

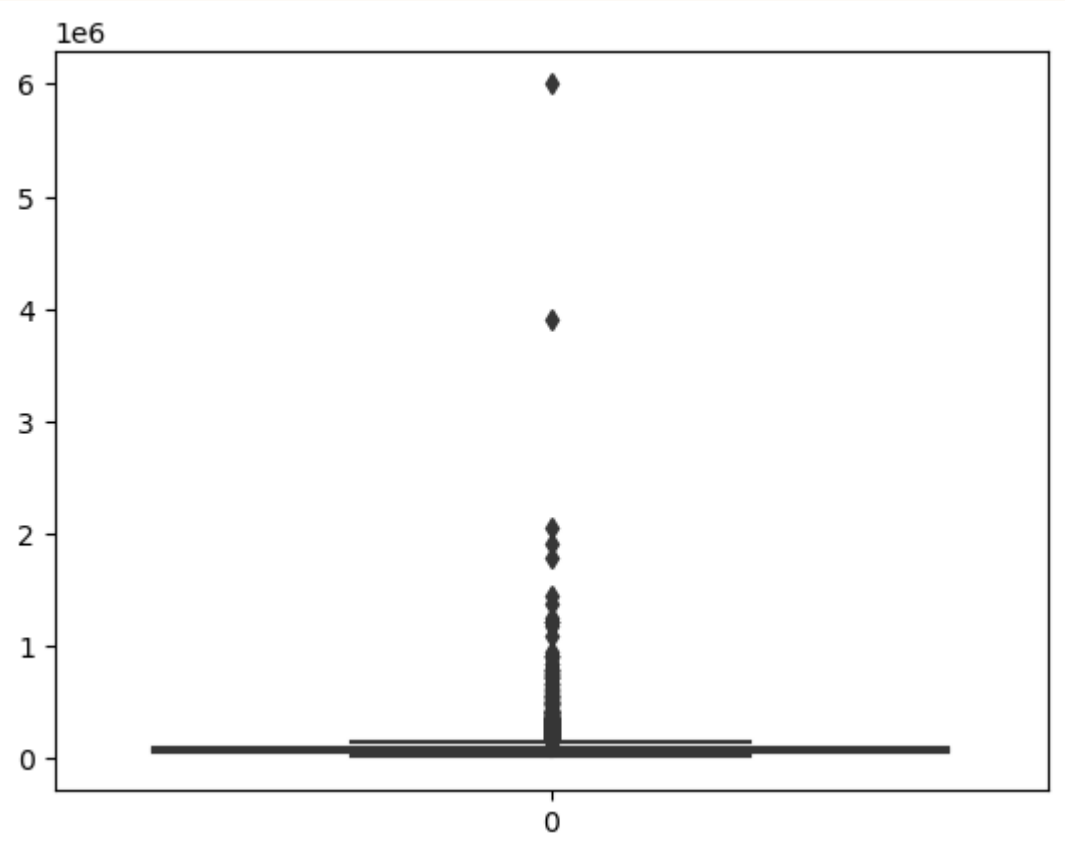
CHECKING FOR MISSING VALUES

- emp_length has 2.68% of missing data
- revol_util have 0.13 % missing data
- Filling the missing values with value which is used in maximum for emp_length
- And dropping the records for revol_util as the % of missing value is less than 1% which will not impact much

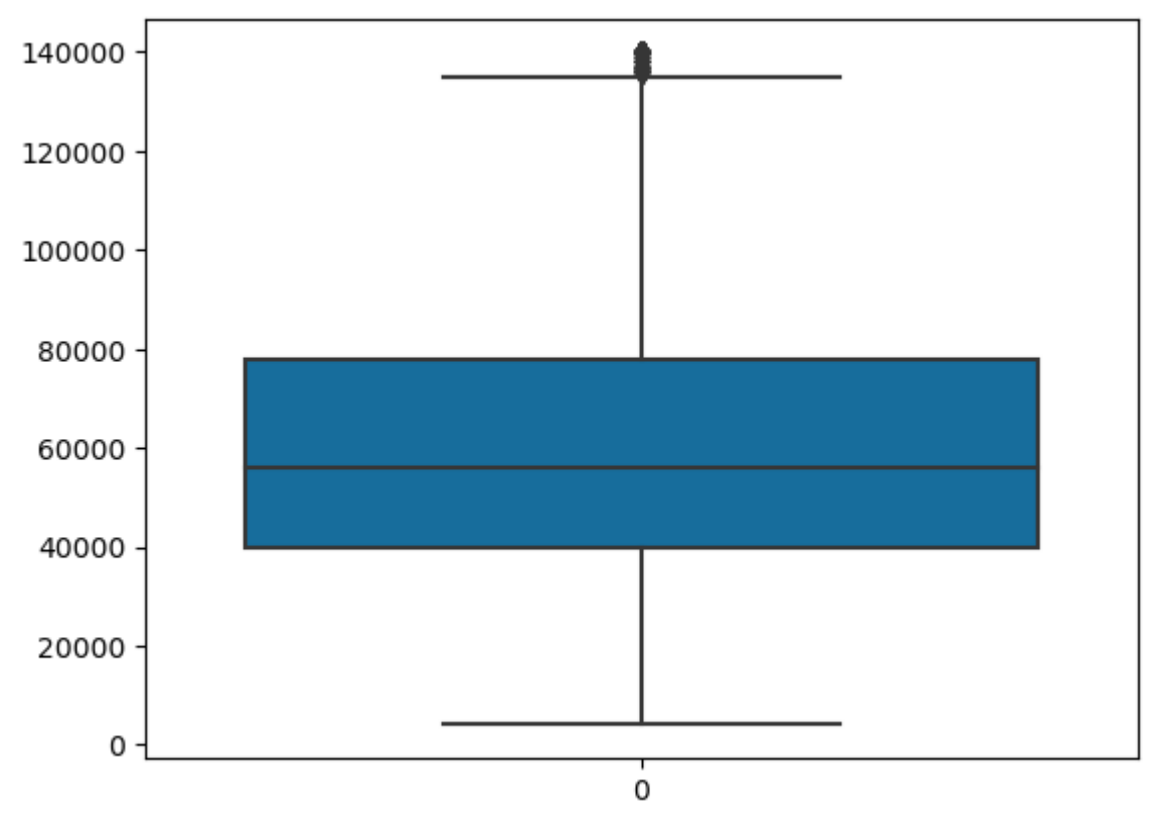
STANDARDIZING THE DATA

- "revol_util" column although described as an object column, it has continuous values. So we need to standardize the data in this column"
- "int_rate" and "term" are also same as above
- "emp_length" < 1 year is assumed as 0 and 10+ years is assumed as 10

FINDING OUTLIERS



It shows that an outlier is present in the data. So, removing them.
To understand and remove the outlier, we can check the quantile variations of the "annual_inc"

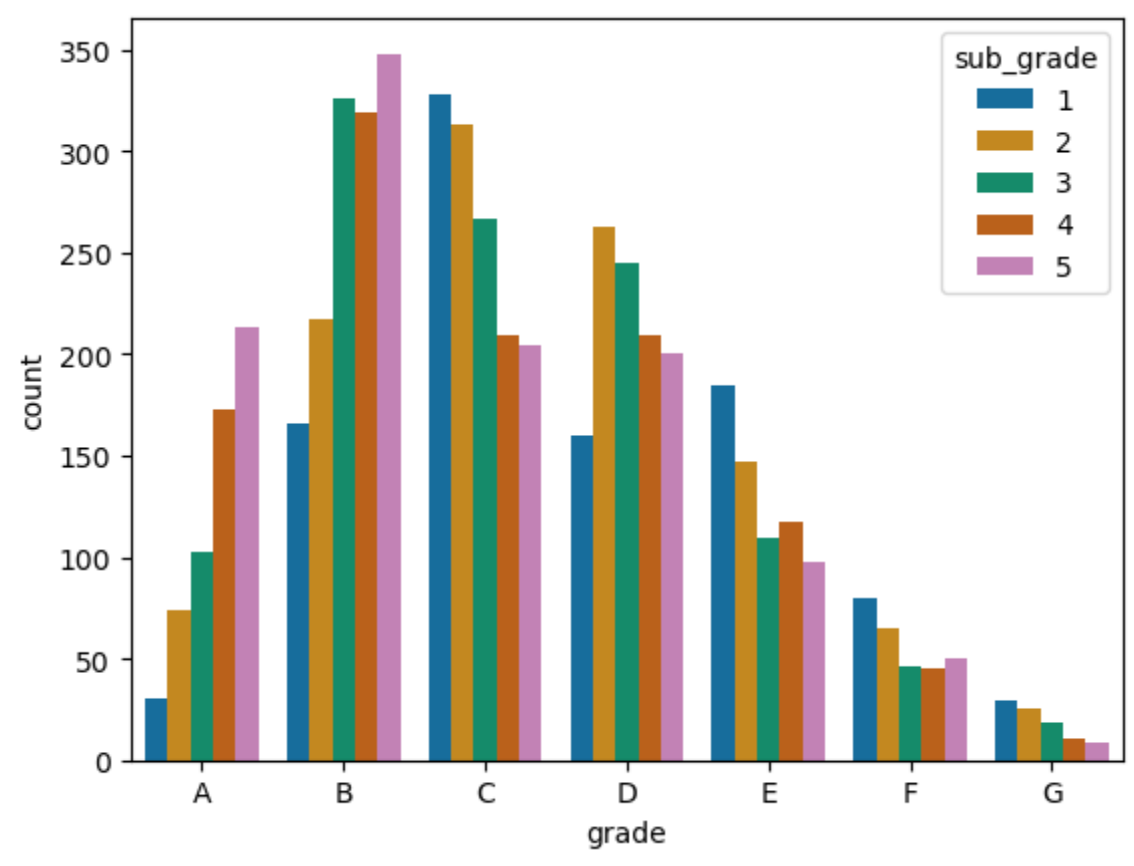


After removing the outlier using the threshold 95%

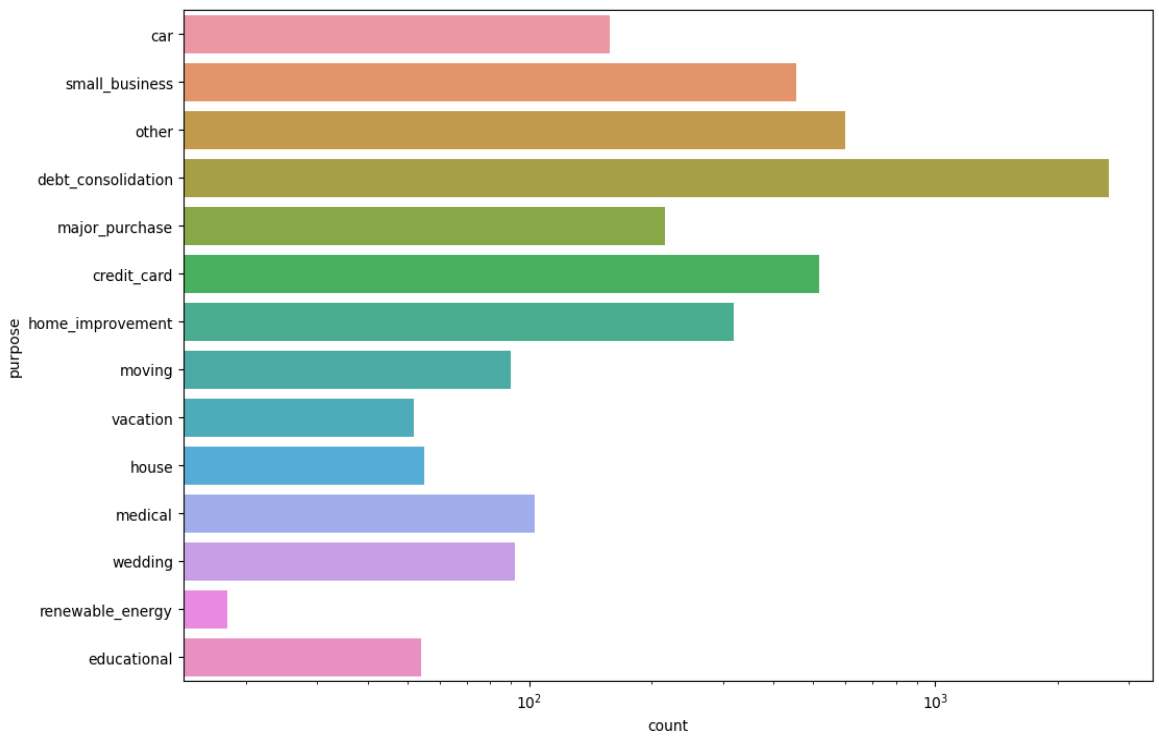
VISUALIZING THE DATA

Univariate Analysis:

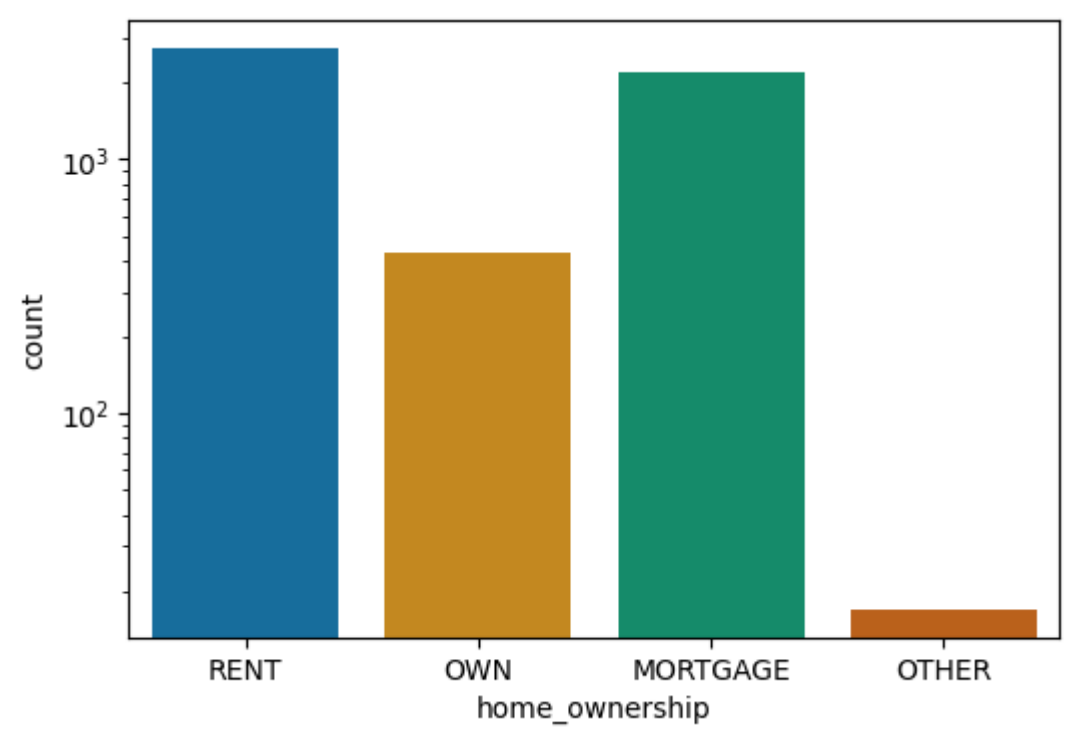
It shows that Grade B5 is the highest in defaulter list



Univariate Analysis Cont..:

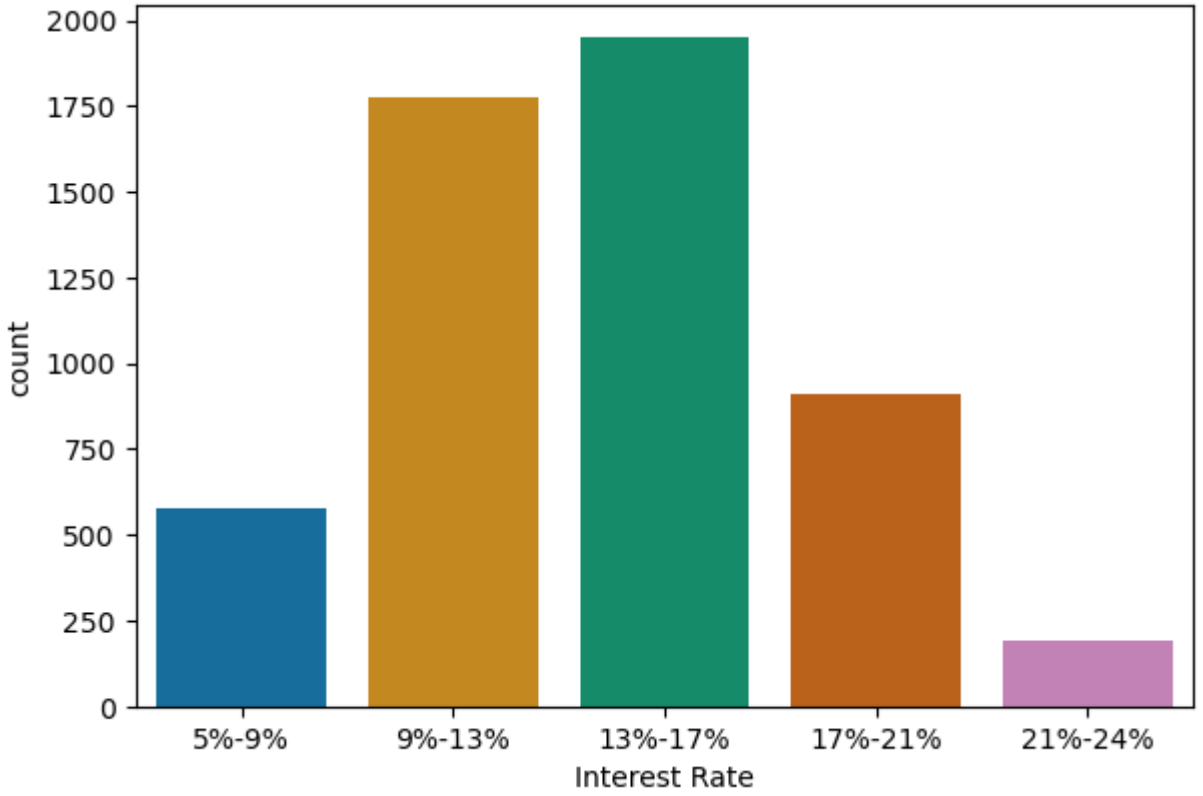


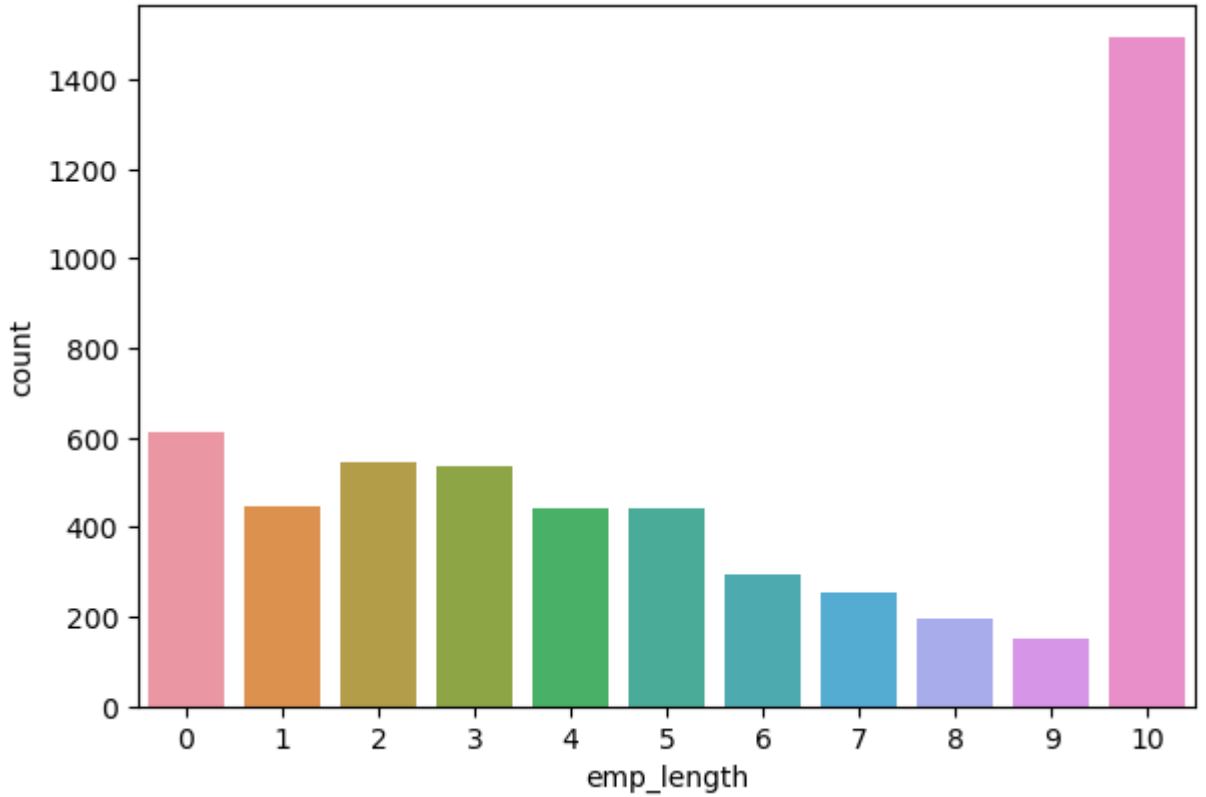
Purpose as 'debt_consolidation' is highest in defaulters



'Rented' house is the highest in defaulters

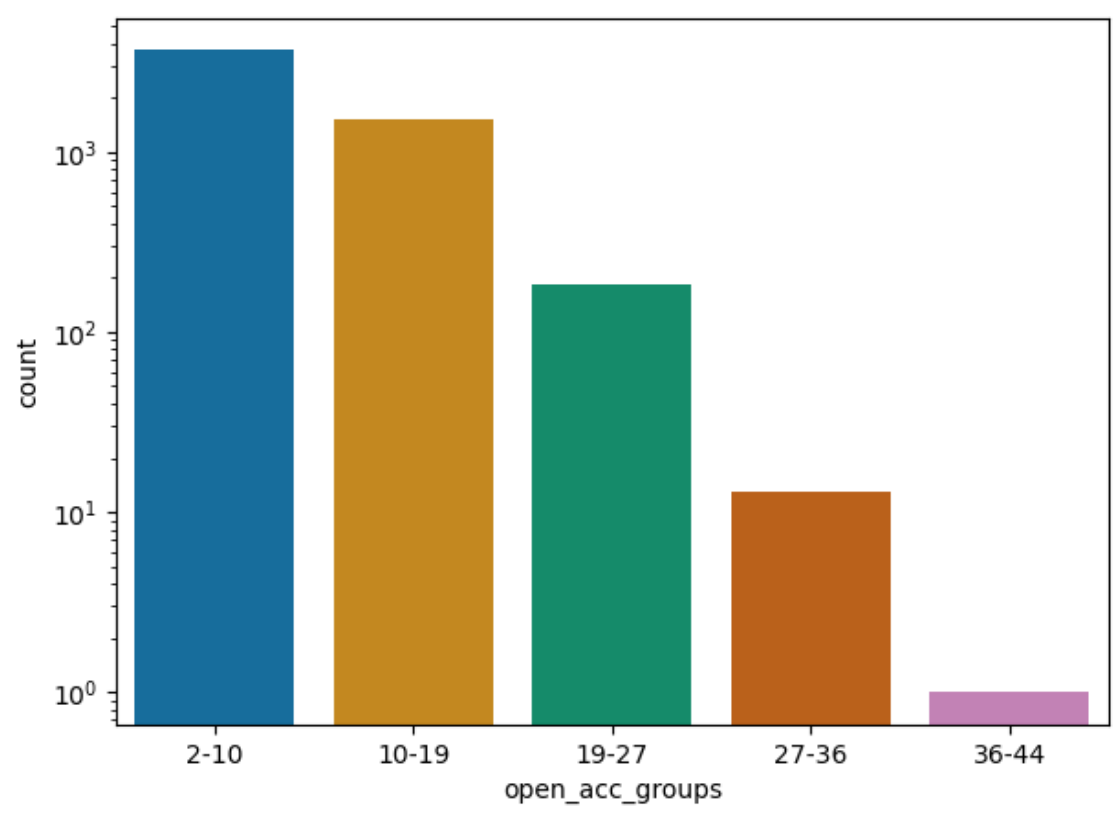
Univariate Analysis Cont..:



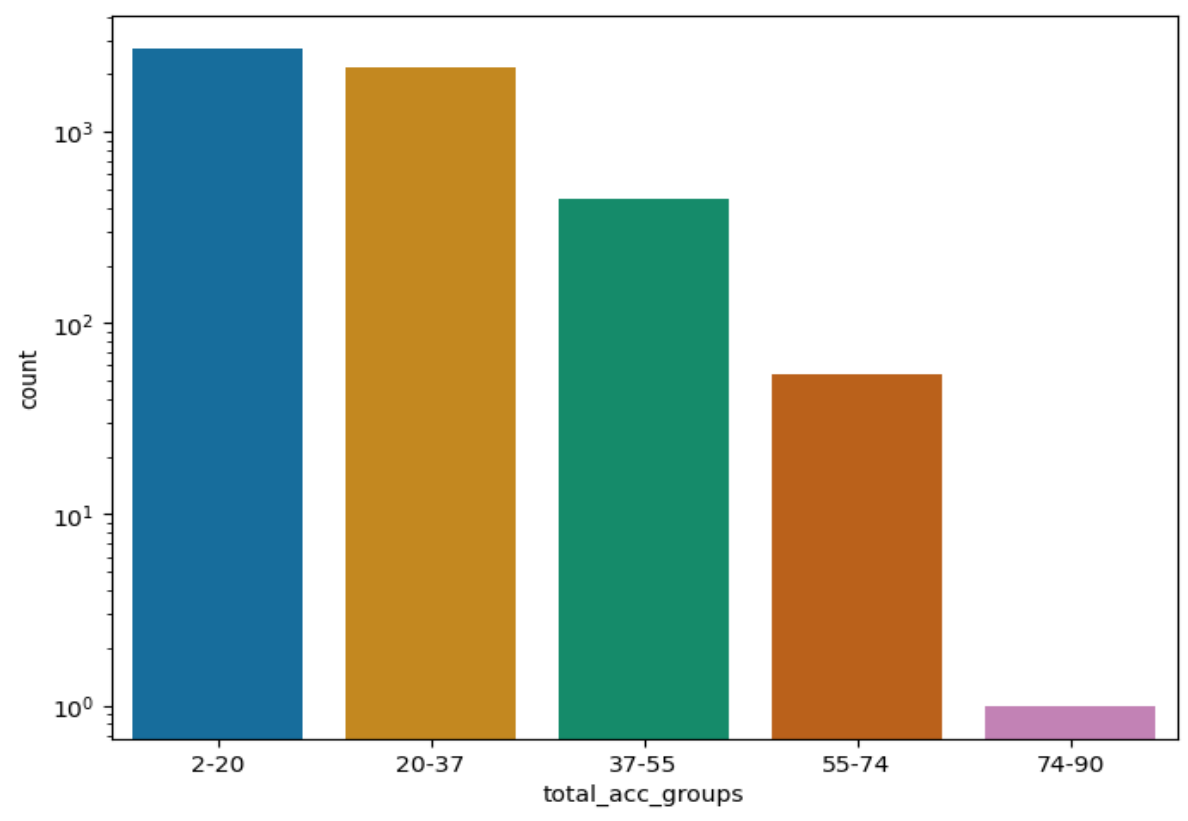


where rate of interest is 13-17% the defaulters are higher and also where the emp_lentgth is 10 the defaulters are higher.
There for here we can assume by the data that the customer having em_length=>10 and interest rate in between 13 - 17% are seen to be defaulter, but will continue to further data analysis to get to more clear factors for defaulter.

Univariate Analysis Cont..:

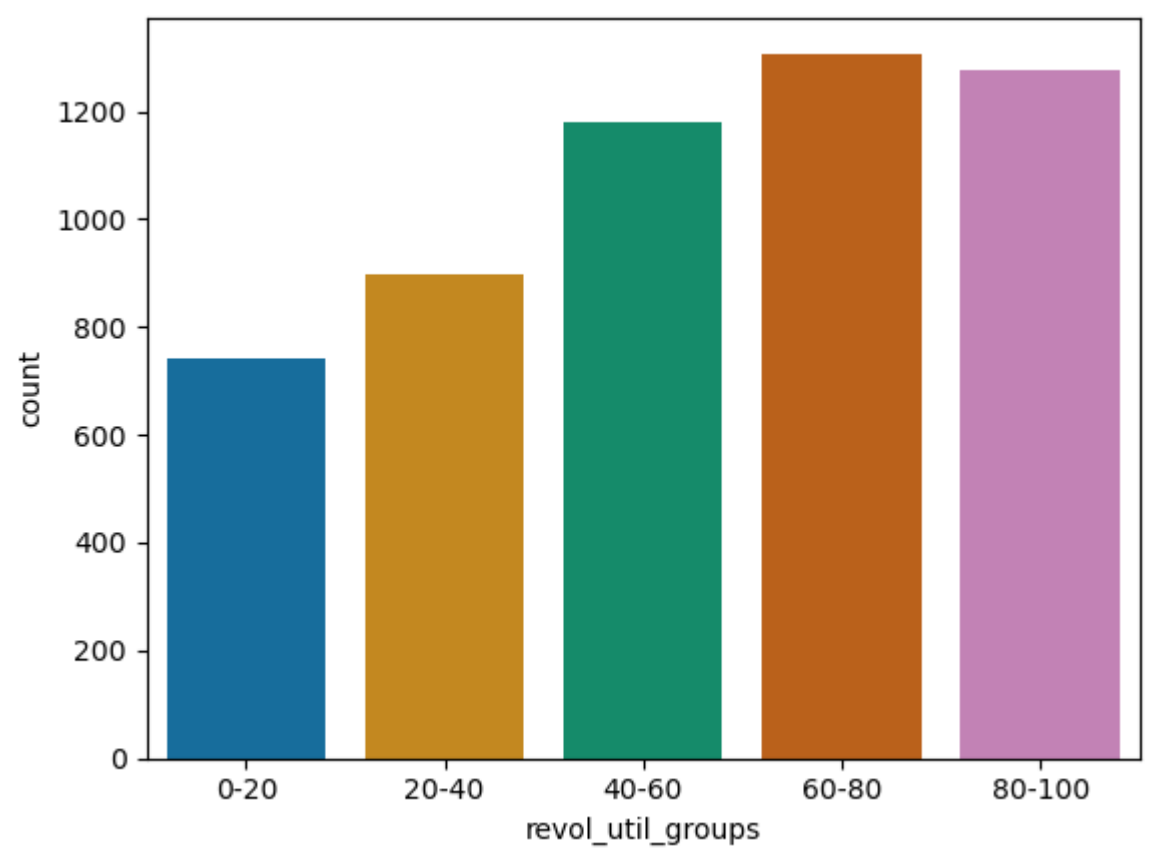


The defaulters are more where the open_acc is in between 2-10

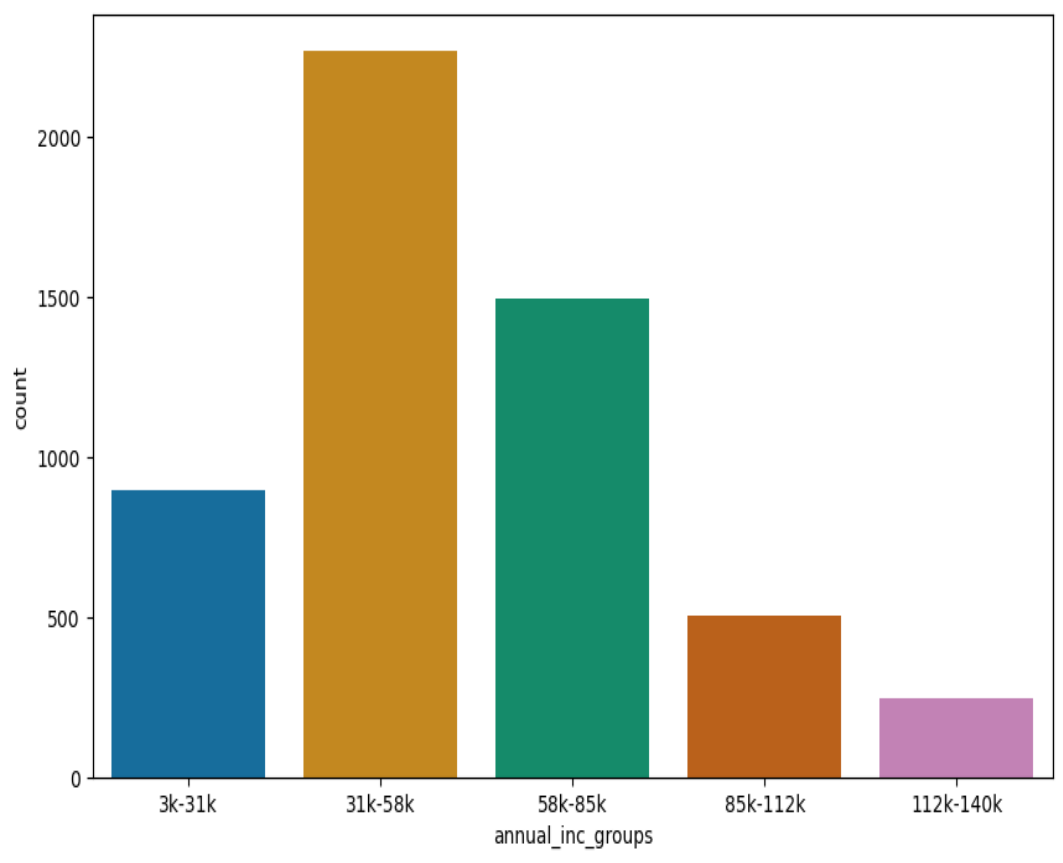


defaulters are higher if the total_acc is in between 2-20

Univariate Analysis Cont..:

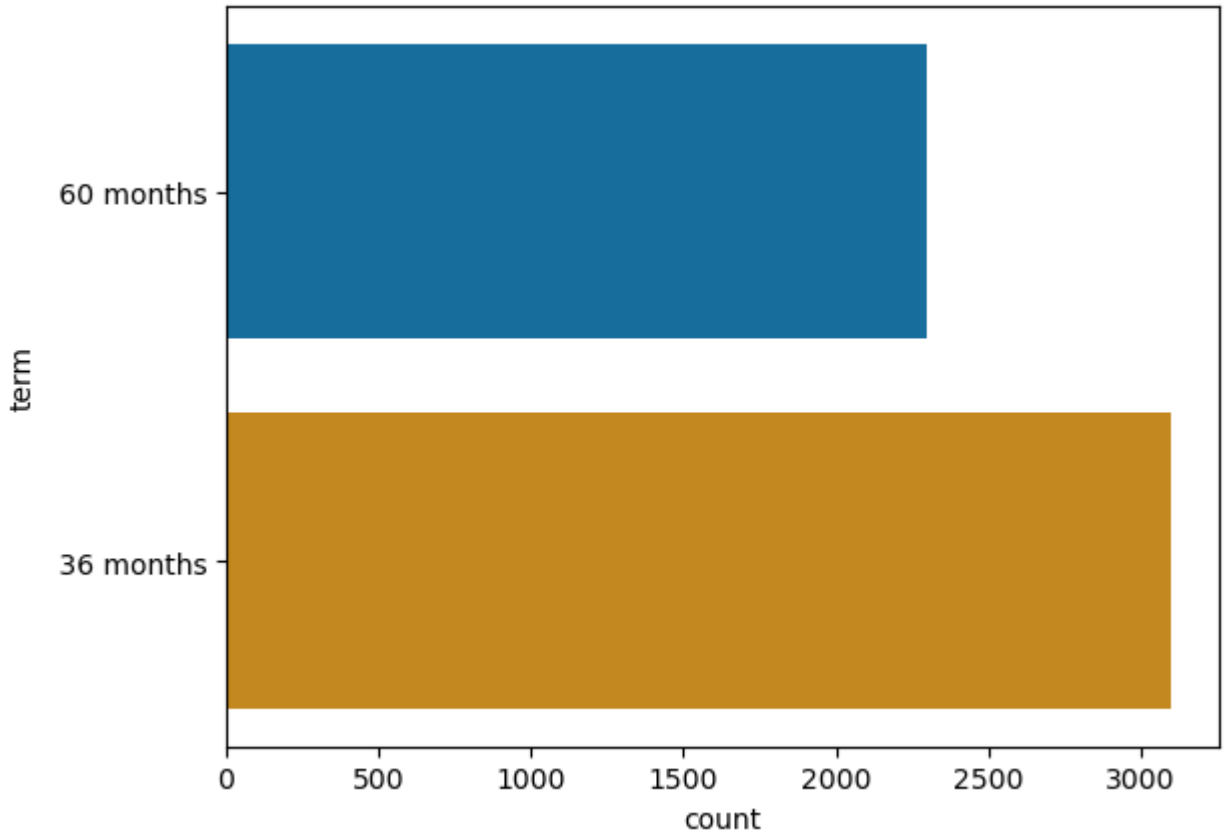


defaulters are higher if the revol_util is greater than 80

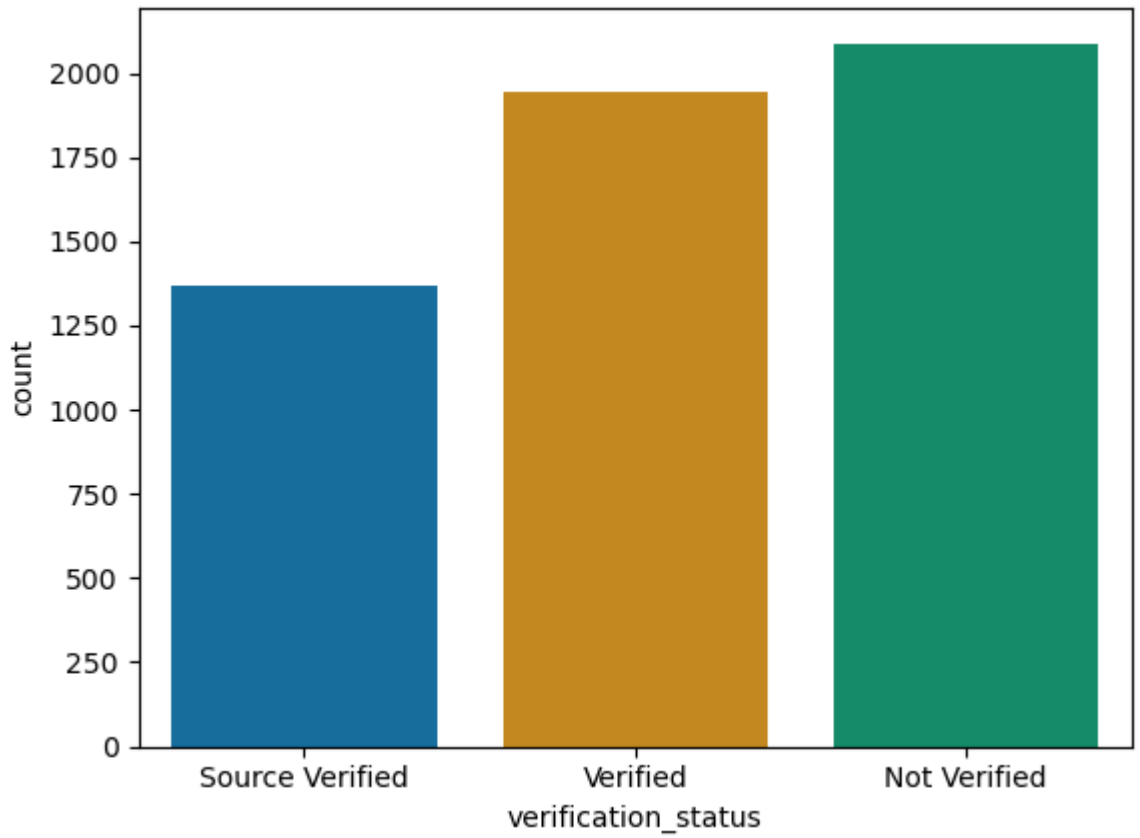


defaulters are higher if the annual_income in between 31-58 thousand

Univariate Analysis:



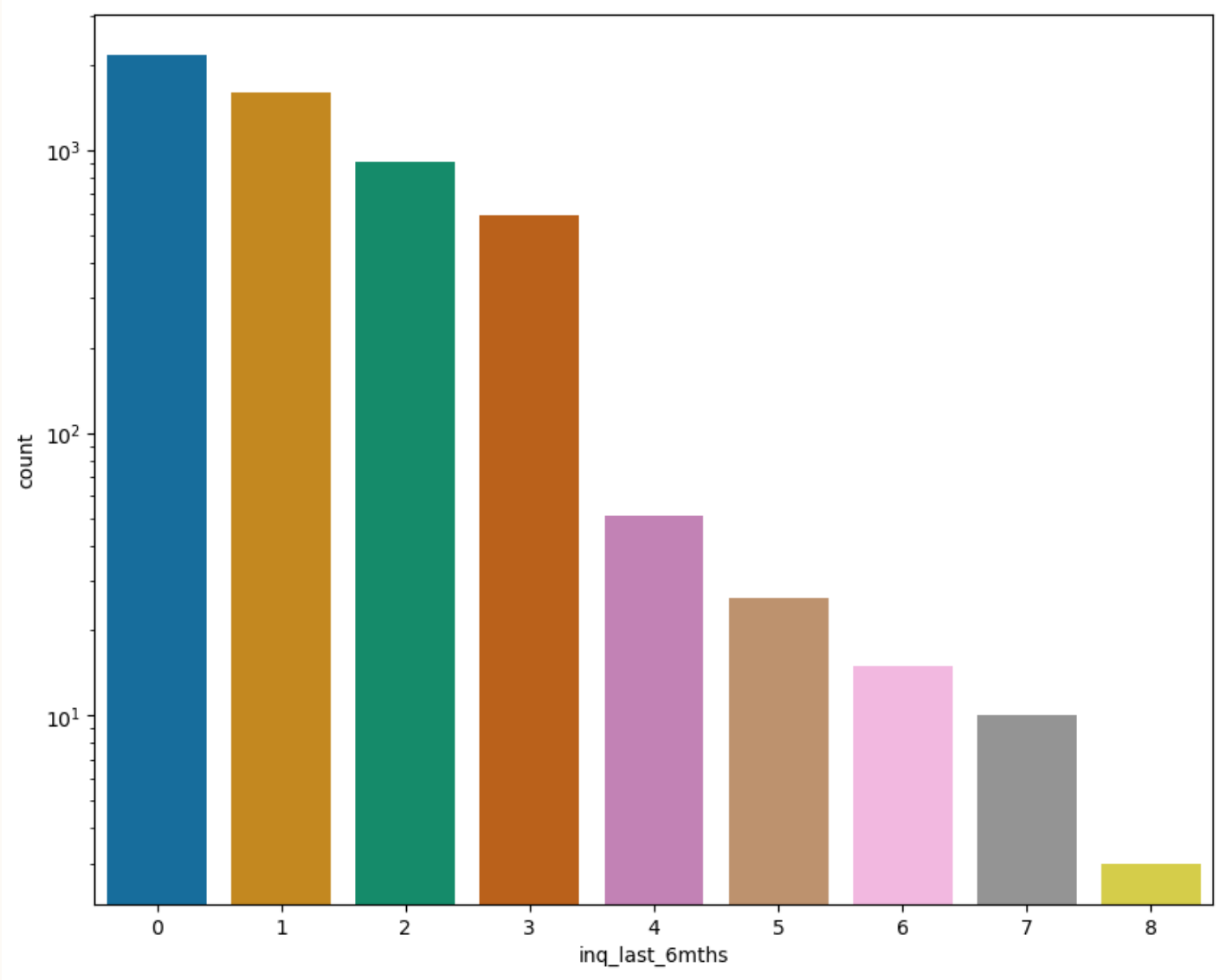
defaulters are higher if the term is 36 months



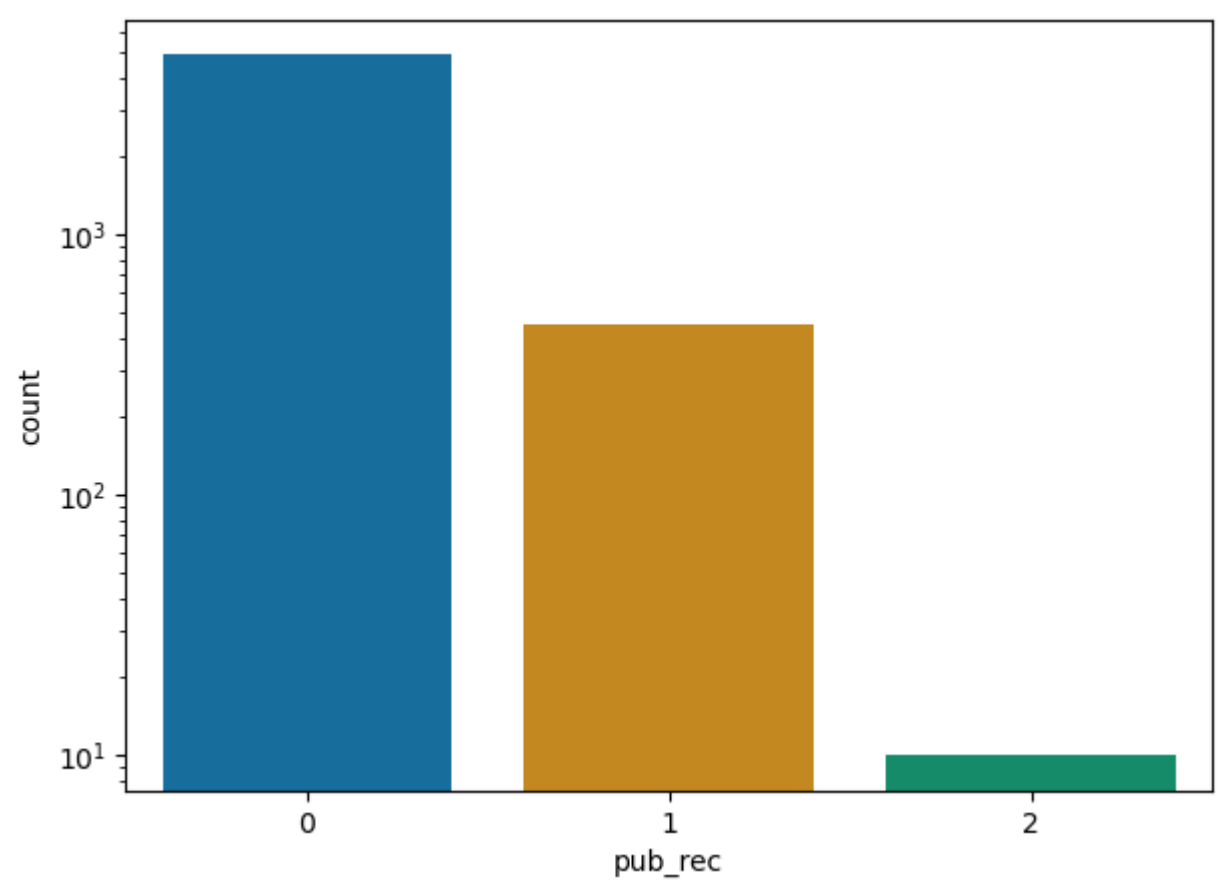
defaulters are higher if the source is not verified

Univariate Analysis:

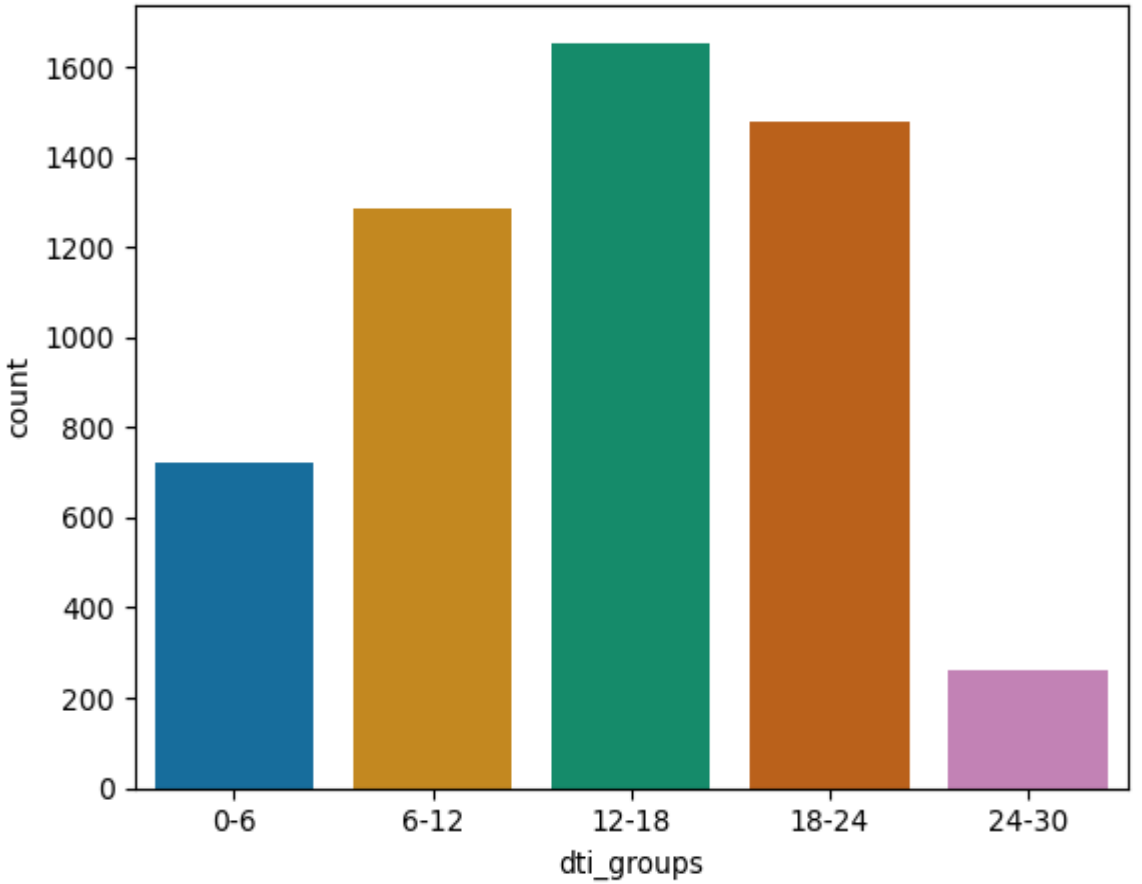
defaulters re higher if the inquiry in last 6 months are 0



Univariate Analysis:

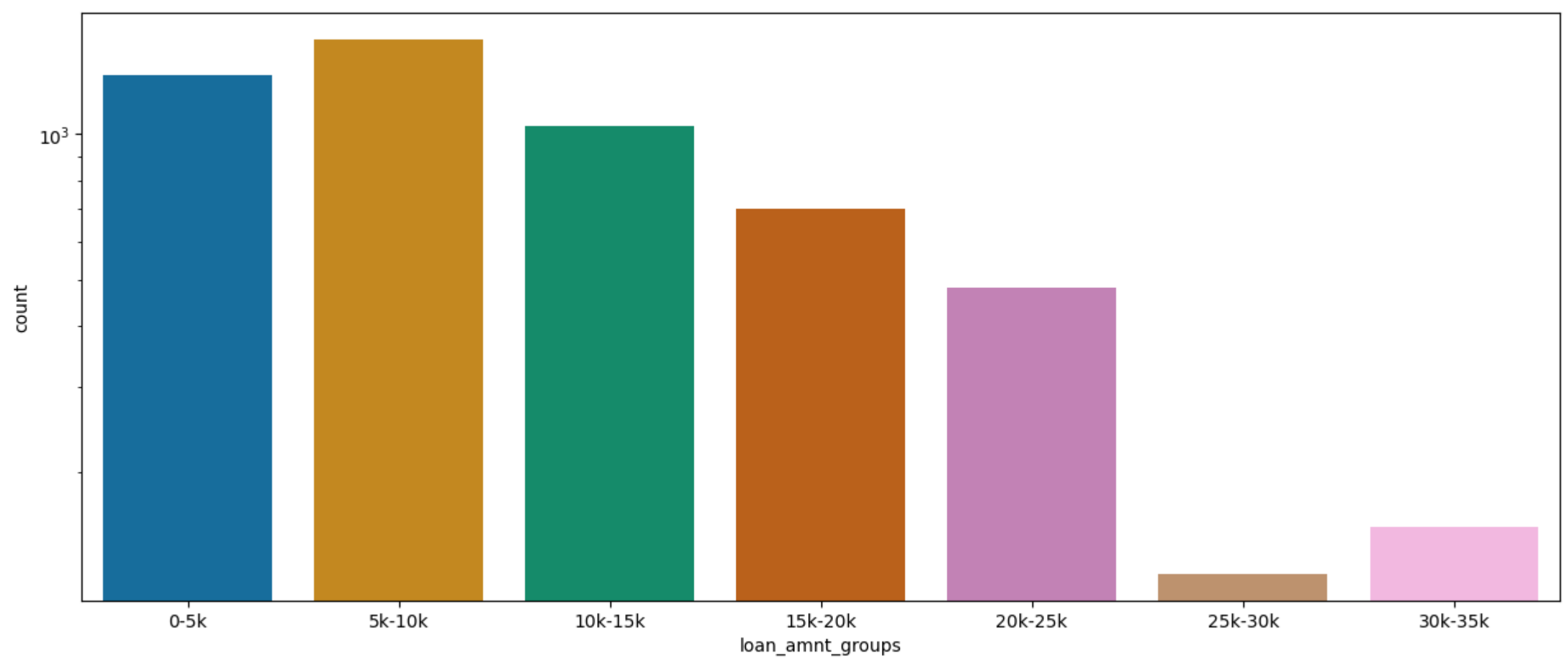


defaulters are higher if the pub_rec is 0



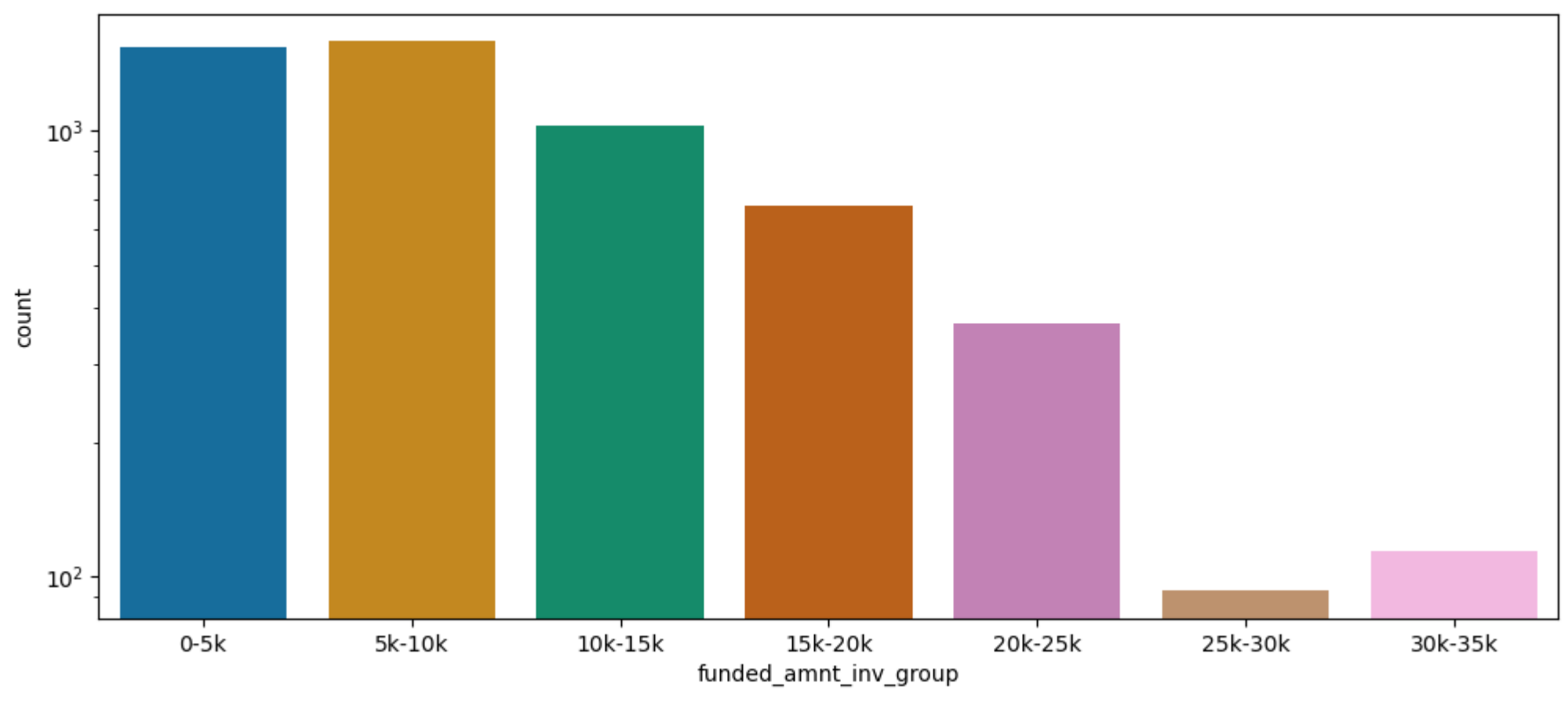
defaulters are higher if the dti in between 12 - 18

Univariate Analysis:



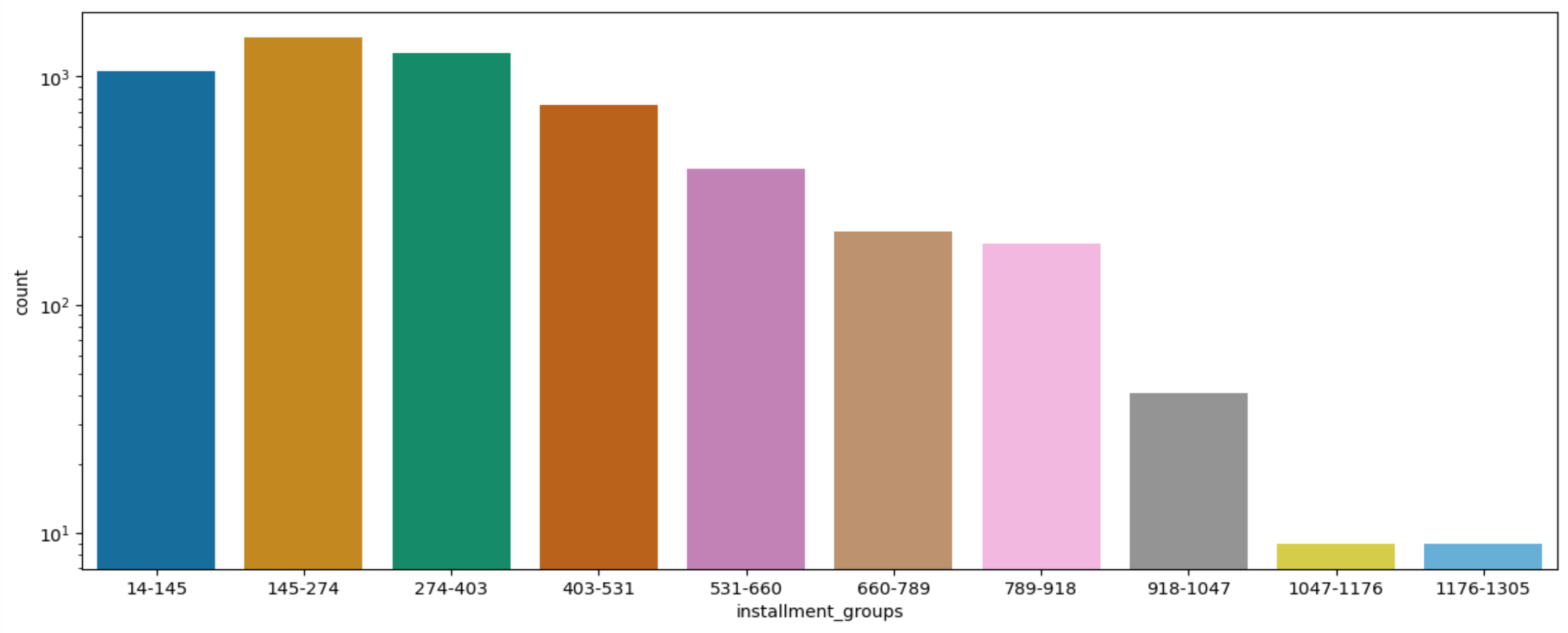
defaulters are higher if the loan amount is in between 5000 - 10000

Univariate Analysis:



defaulters are higher if the funded loan amount is in between 5000 - 10000

Univariate Analysis:



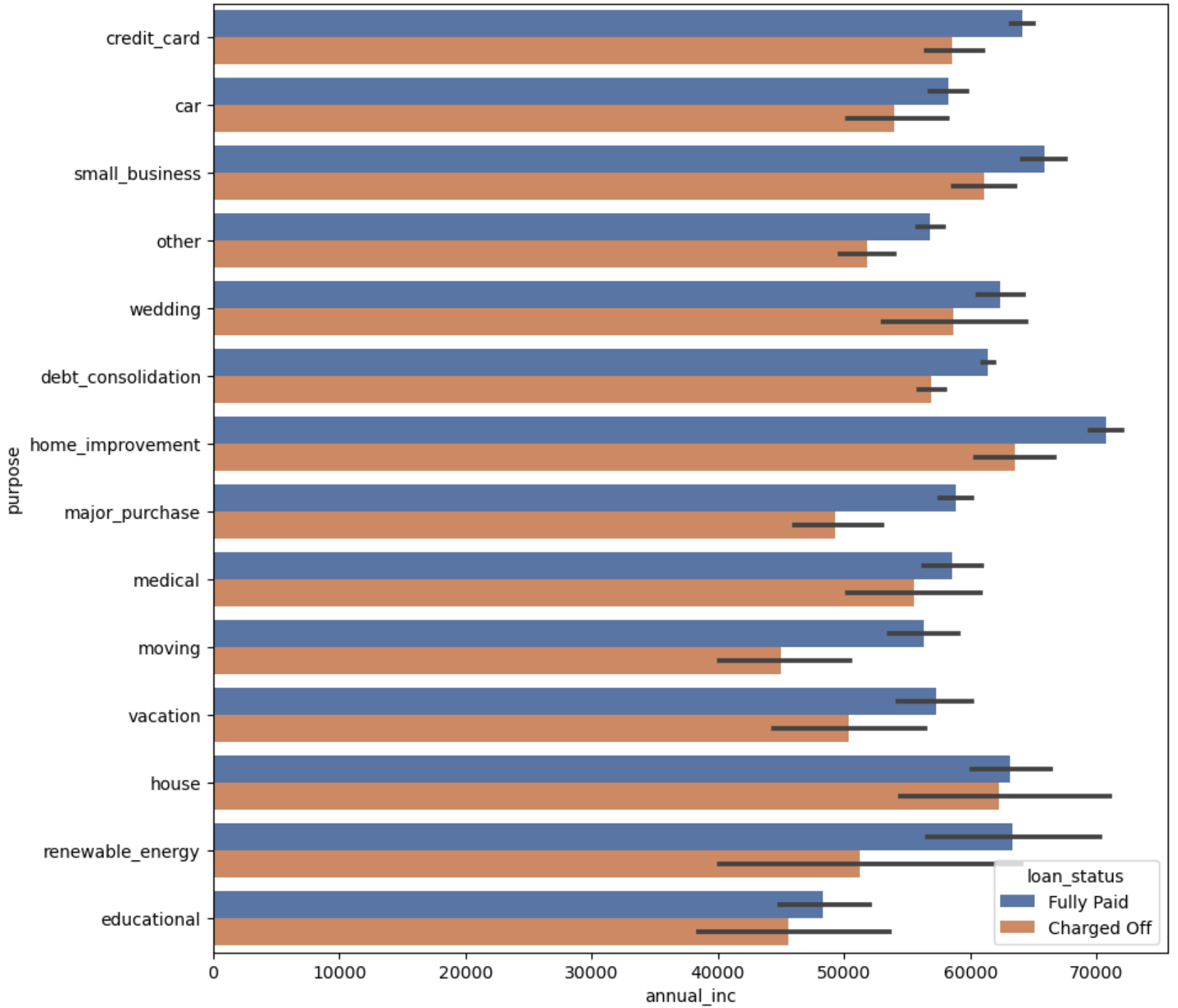
defaulters are higher if the instalment amount in between 145 - 274

With all the above analysis we can draw the conclusions as follows:

- Grade is B
- Sub Grade is B5 level
- Rented House
- taken loan for debt clearance
- Loan interest rate of 13 to 17%
- Loan amount in between 5000 to 10000
- Monthly installments in between 145 to 274
- Term of 36 months
- When loan status is not verified
- When loan inquiry in last 6 month is 0
- Open accounts in between 2-10
- Total account in between 2-20
- Employment tenure is 10
- Income range in between 30000 to 50000
- dti in between 12 to 18
- derogatory public records is 0

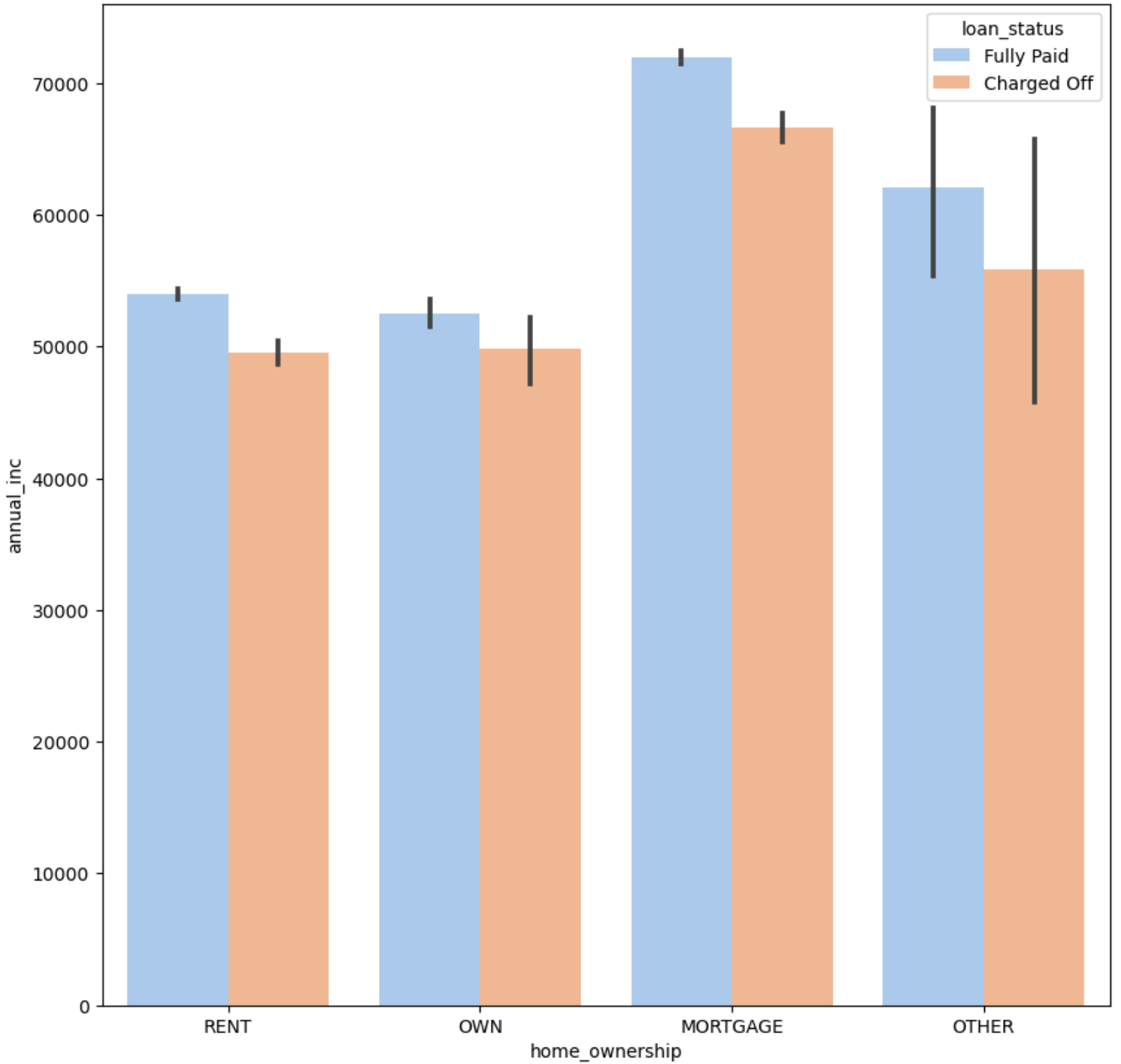
Bivariate Analysis:

Customer having salary in between 60 to 70 thousand and purpose is 'home_improvement' has higher chances of being defaulter

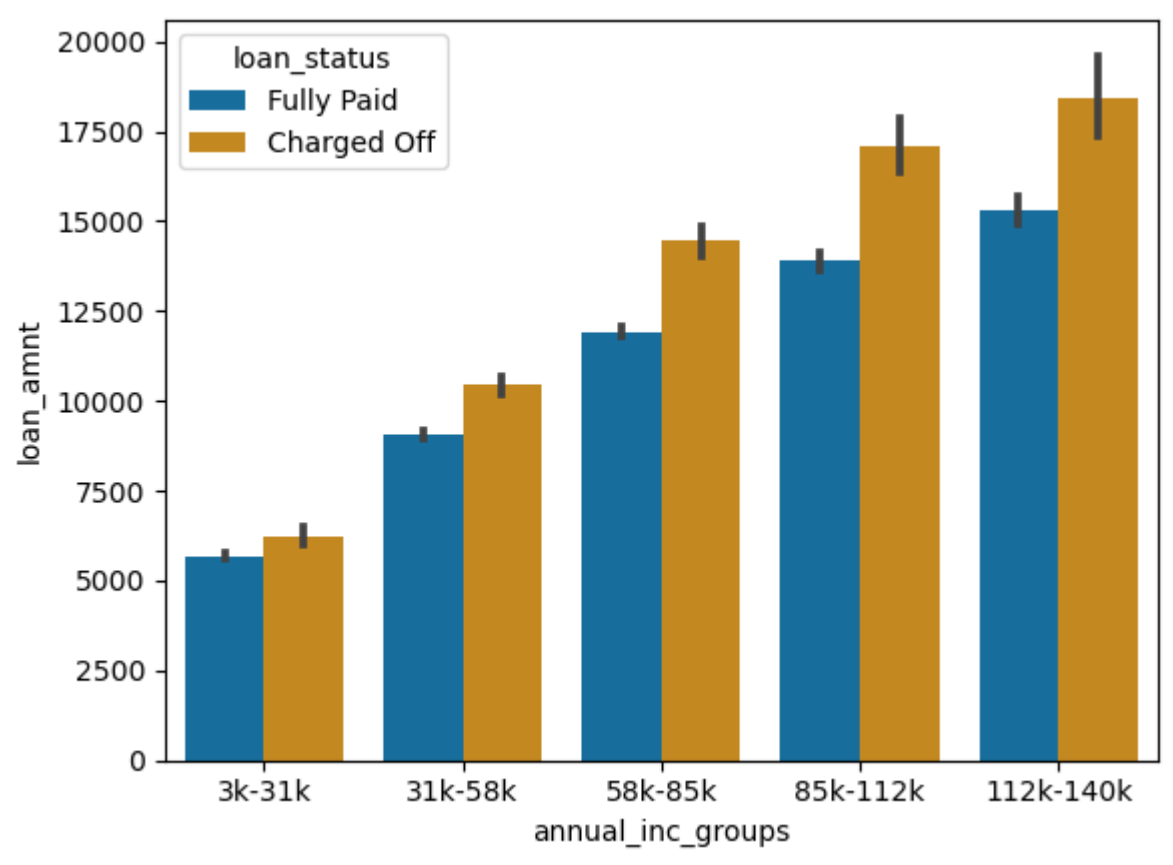


Bivariate Analysis:

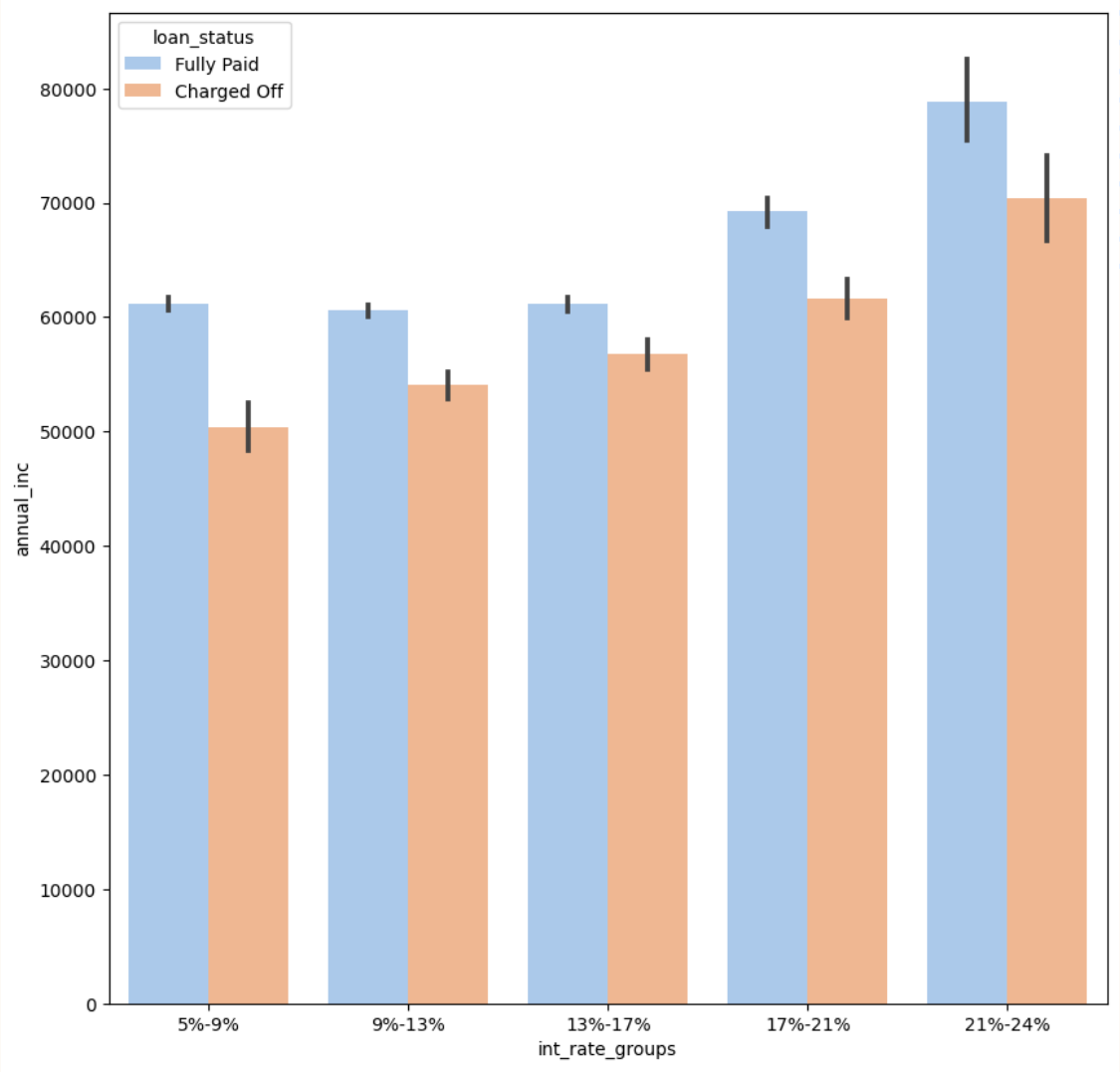
Customer having salary in between 60 to 70 thousand and having mortgage on home has more chances of being defaulter



Bivariate Analysis:



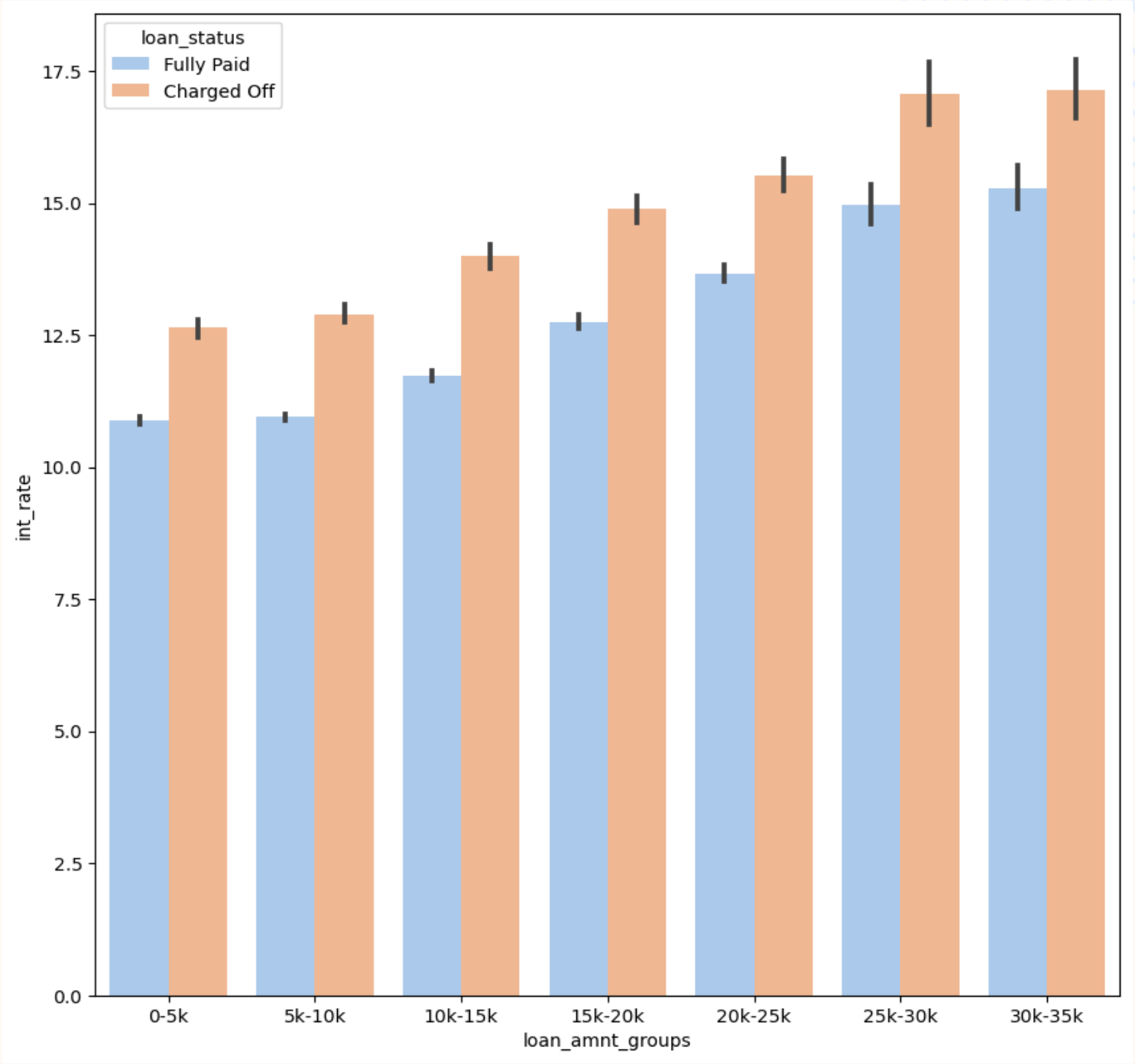
Customer having highest income has got large amount of loans. And also seen that the defaulters are more in number across all income groups



Customer having annual_inc in between 60 to 80 k has more defaulters with 21 - 24% rate of interest

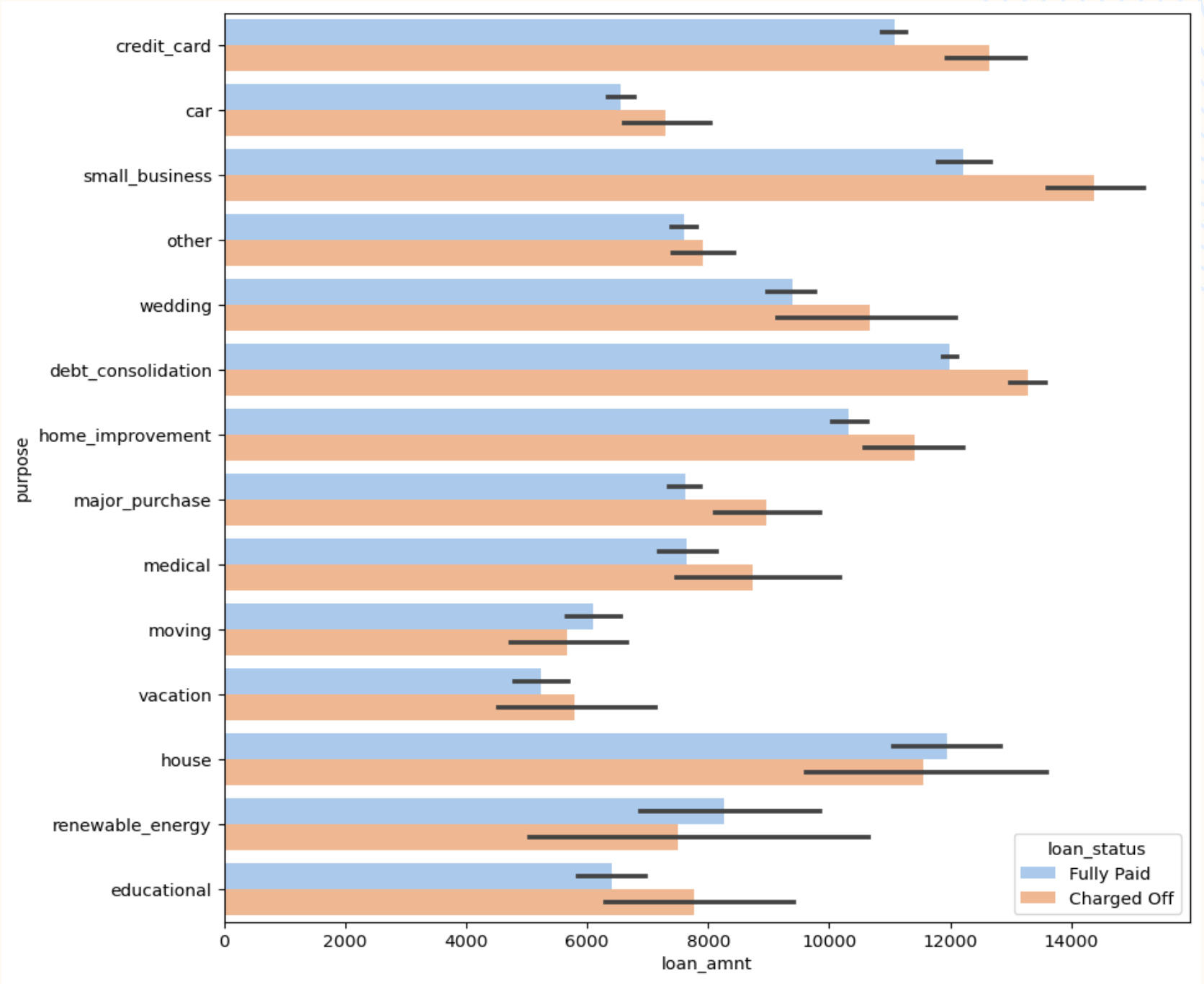
Bivariate Analysis:

Loan amount between 30 to 35 k with interest rate as 15 to 17.5% has highest number of charged off



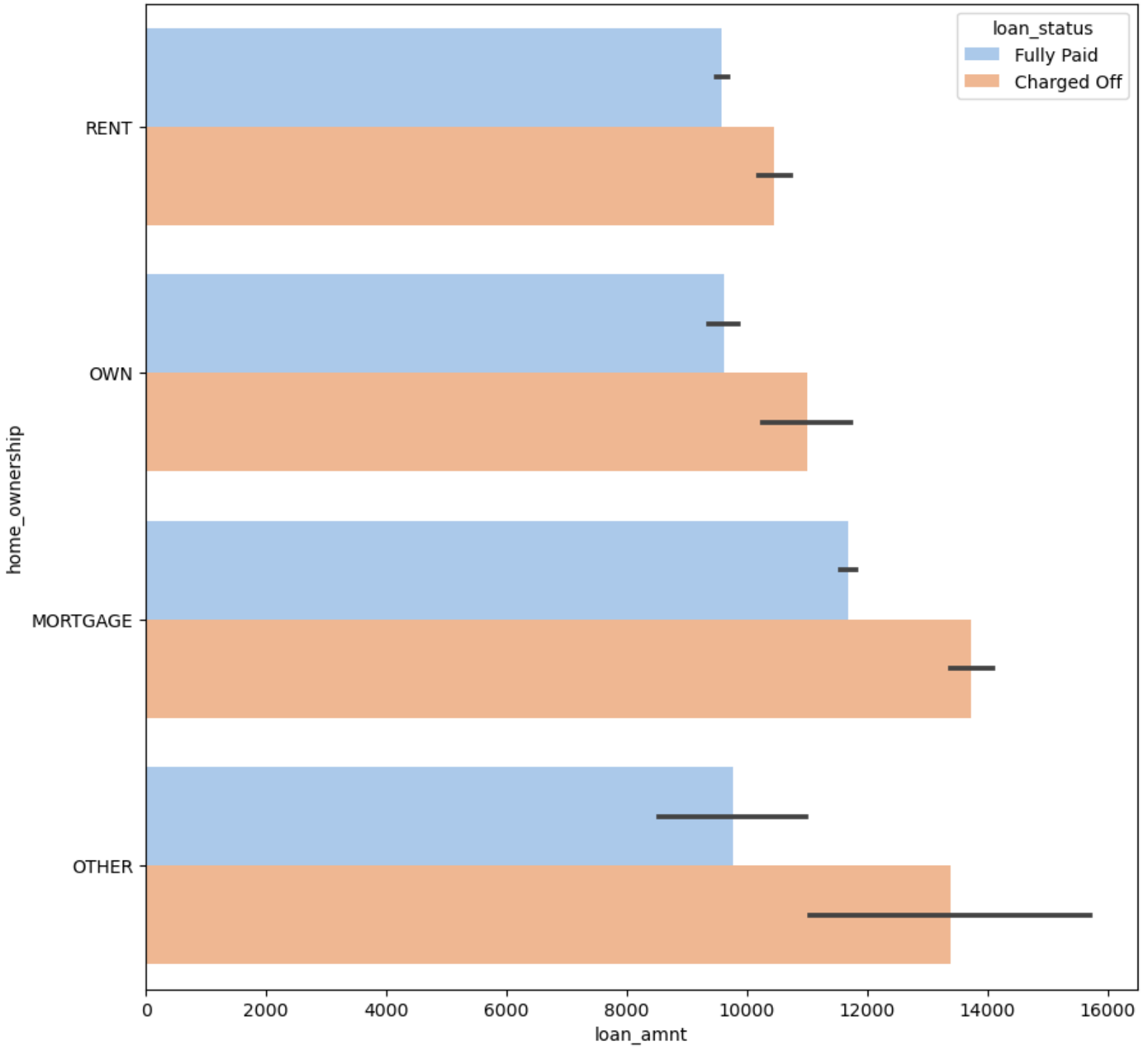
Bivariate Analysis:

Customer purpose is small business and loan amount in between 12 - 14 k has the highest number of defaulters



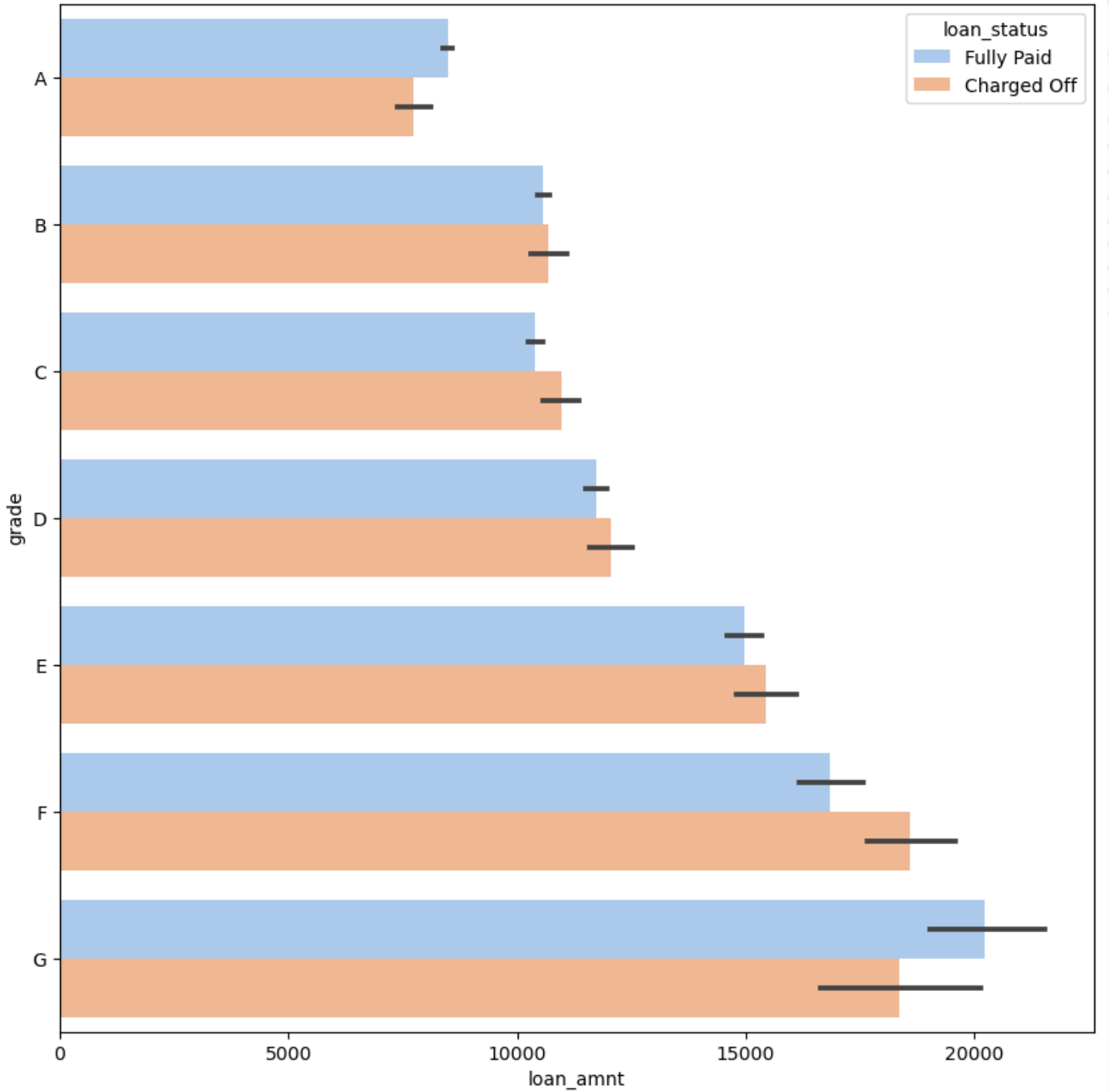
Bivariate Analysis:

Customer having loan of amount 12 - 16 k and home is mortgage and other has been highest in defaulters

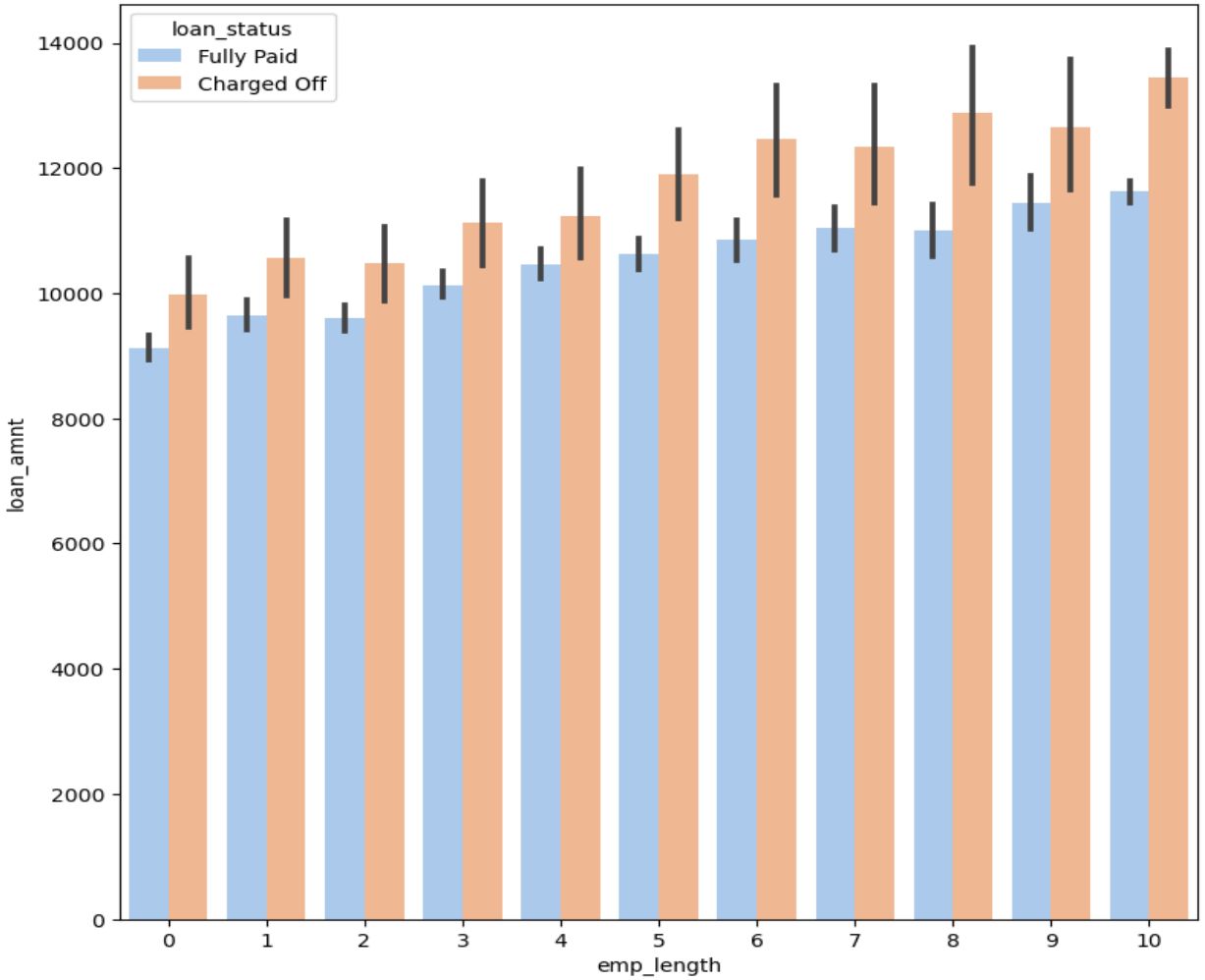


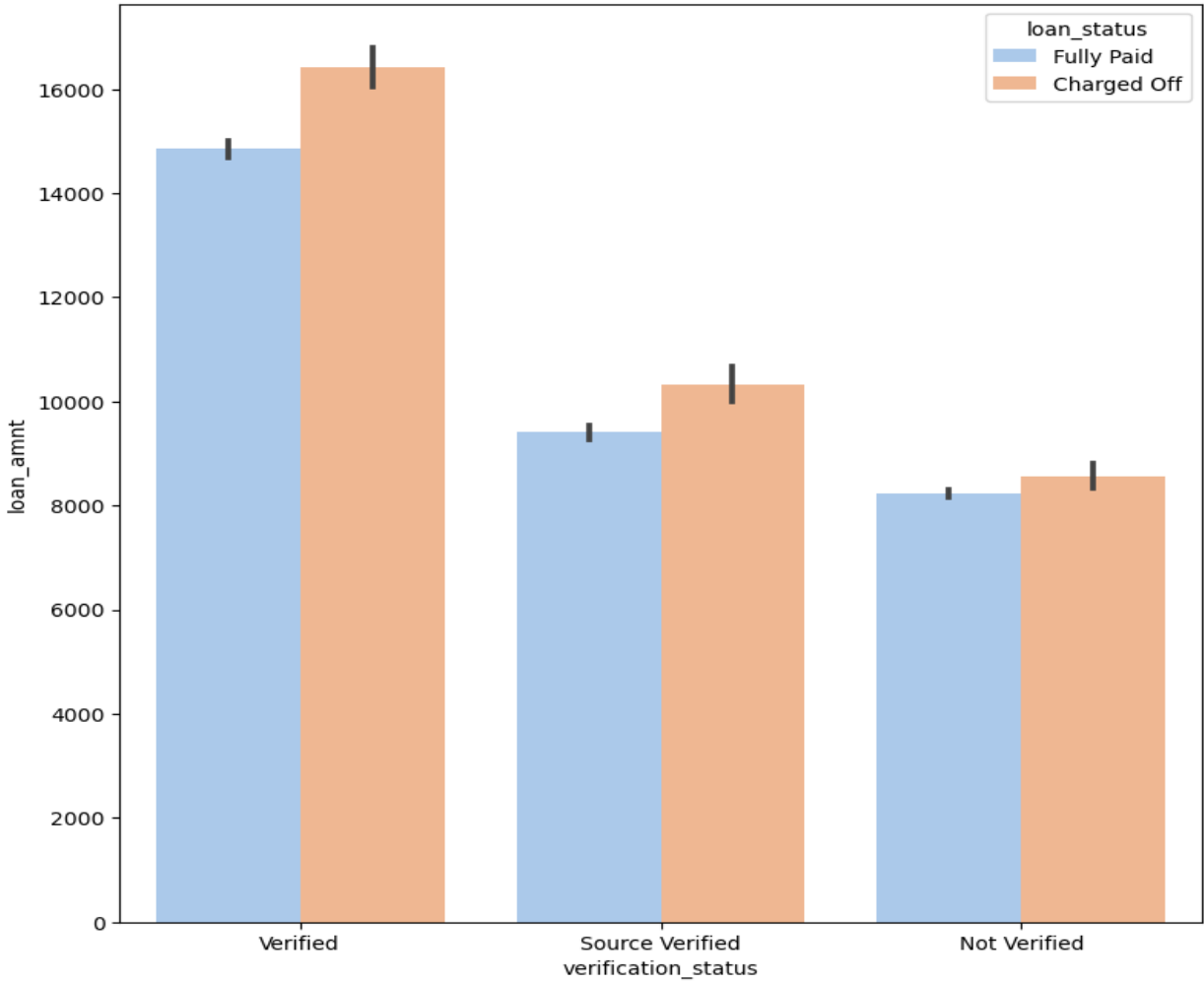
Bivariate Analysis:

Customer with grade F and loan amount in between 16 to 20 k has high chances of getting defaulter



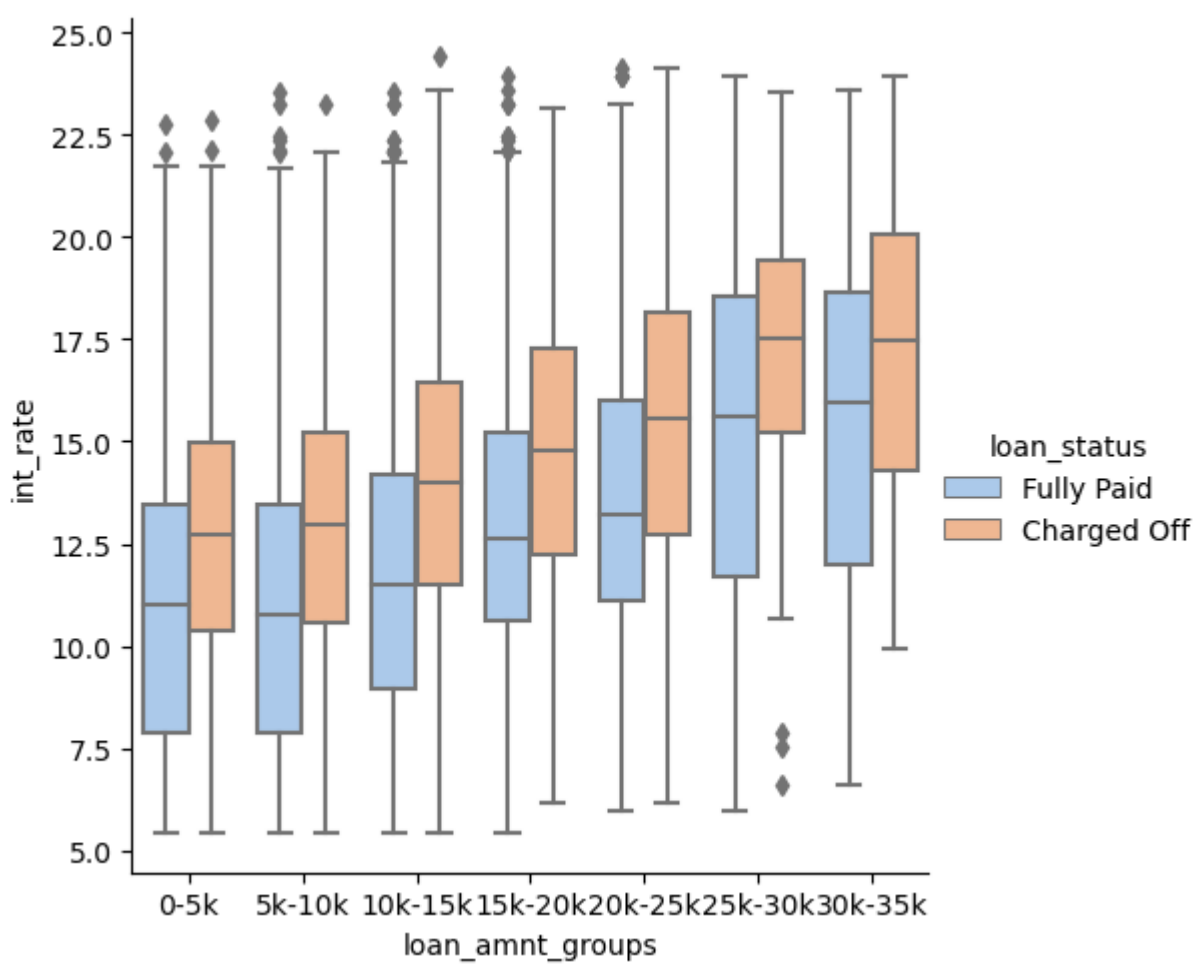
Bivariate Analysis:



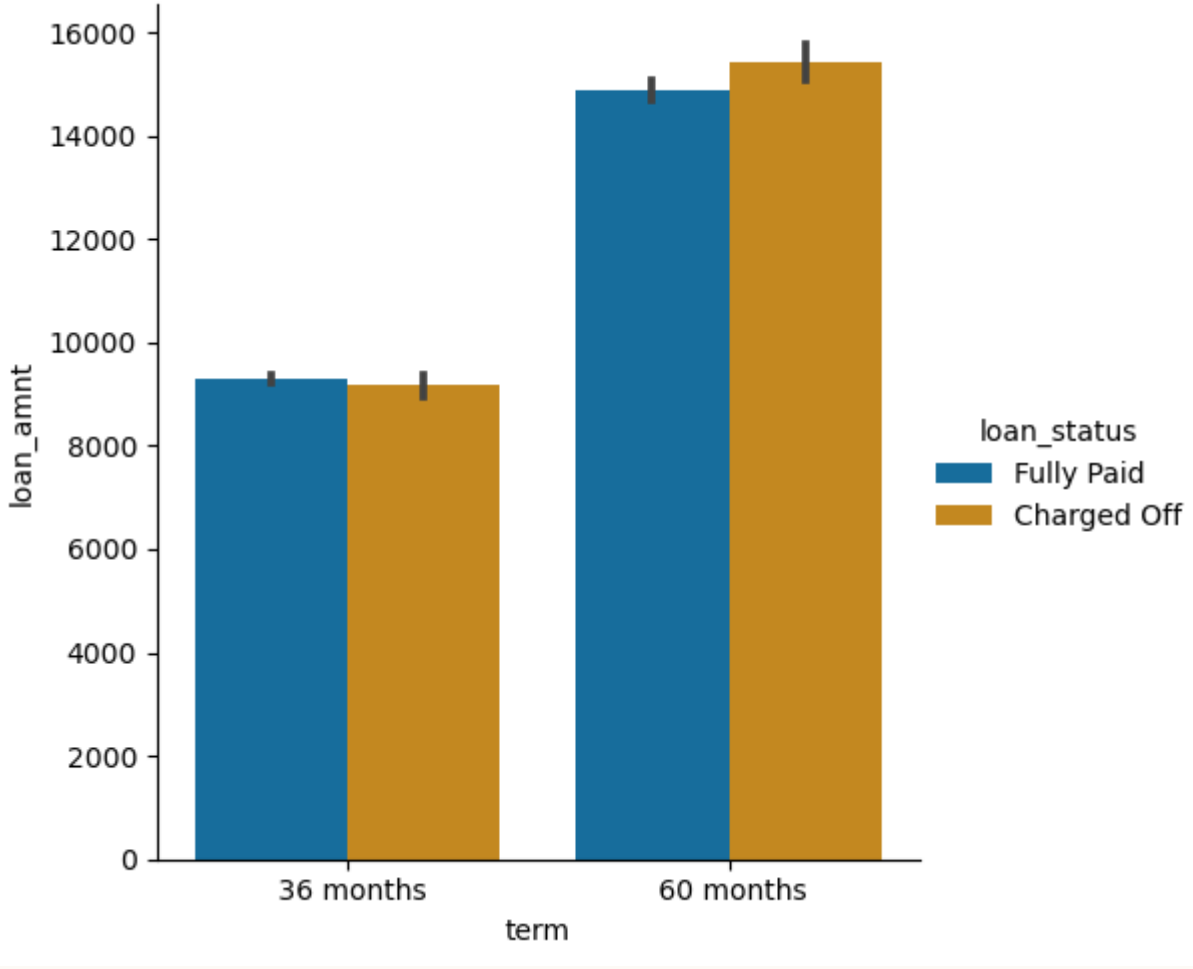


customer with longer working tenure got large amount of loan. it also shows that verified loan have higher amount of loan. Customer is a defaulter when the loan amount is 12 - 14 k and the working tenure is 10, and also if the source is verified but the loan amount is more than 16k is a defaulter

Bivariate Analysis:



Higher int. rate has higher defaulters.



Long term loans have large loan amount, larger the loan amount larger the defaulter

Bivariate Analysis:

With all the above analysis we can draw the conclusions as follows:

- Customer taking loan for 'home improvement' and having income in between 60k -70k
- Customer with home ownership as 'MORTGAGE and have income of 60-70k
- Customer whose rate of interest is in between 21-24% and having an income of 70k-80k
- Customer whose loan amount in range 30k - 35k and interest rate of 15-17.5 %
- Customer whose purpose for loan is 'small business' and the loan amount is greater than 14k
- Customer whose home ownership is 'MORTGAGE and have loan of 14-16k
- When grade is F and loan amount is between 15k-20k
- When employment length is 10 yrs. and loan amount is 12k-14k
- When the loan is verified and loan amount is above 16k
- For grade G and interest rate above 20%

THANKYOU