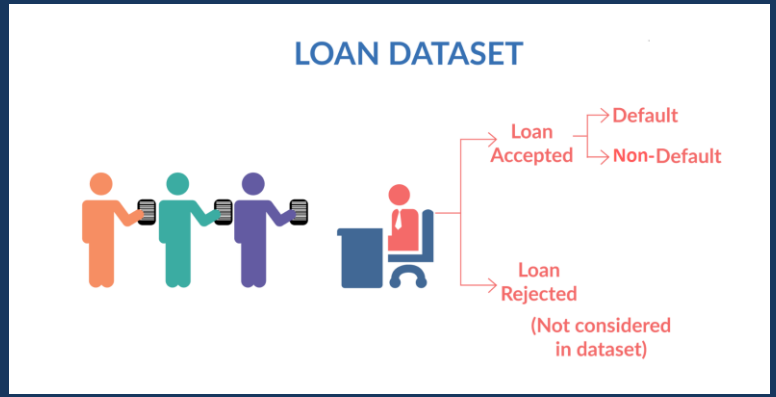
LENDING CLUB CASE STUDY

Contributor : Ishita Das and Arindam Dutta

BUSINESS PURPOSE

Understanding the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.



PROBLEM STATEMENT

With the given data set that contains the complete loan data for all loans need to perform EDA to conclude the factors of default loan and thus control the credit risk.

ANALYSIS APPROACH

Data set analysis

- Understanding the columns (111 cols) of the data set
- Figuring out the driving variable to understand the risks for loan payments

Data Clean up

- Duplicate rows
- Remove and completely null rows and columns
- Remove columns which has higher number of null values

Data Handling

- Convert columns to desired data type
- Create Derived column where required
- Remove columns which hold similar values and are not relevant for analysis

Univariate Analysis

- Distribution and variance of data
- Outliers analysis and removal

Bivariate Analysis

- Analyze the impact of different continuous and categorical variables on driver variable loan_status

STRATEGY FOR BIVARIATE ANALYSIS

New Categorical variable creation -

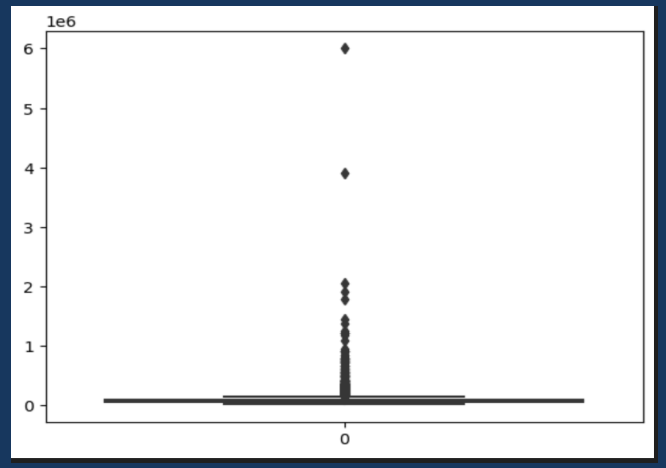
We have created few of the categorical variables like int_rate_groups, open_acc_groups, revol_util_groups. ...

Purpose was to create bins for the continuous variables to perform our analysis easily.

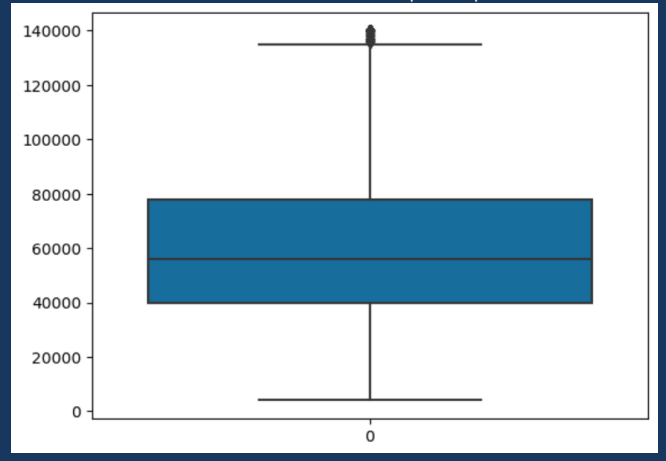
UNIVARIATE ANALYSIS- 1

Annual Income

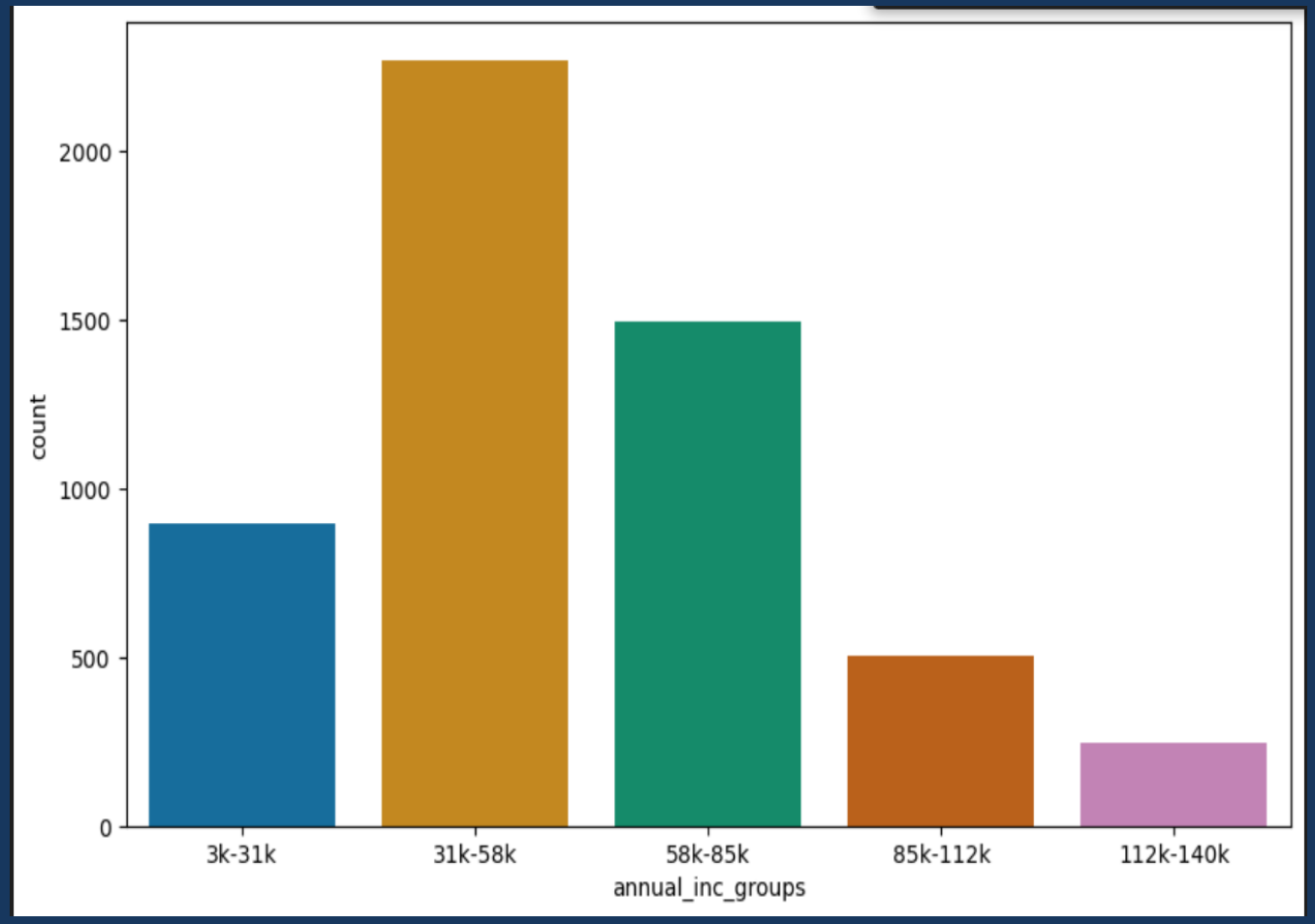
significant outliers for the field `annual_inc`



Outliers are removed based on IQR principle. Post outliers' removal



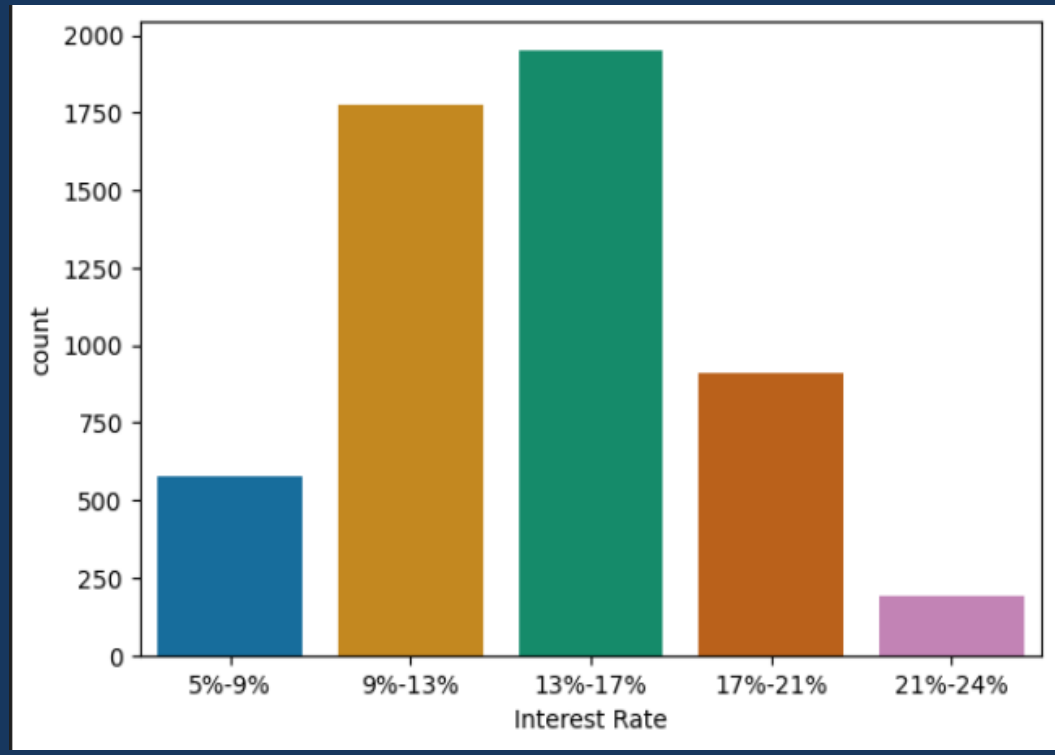
annual_inc_groups - A new variable created to cut the data into bins and visualize the data



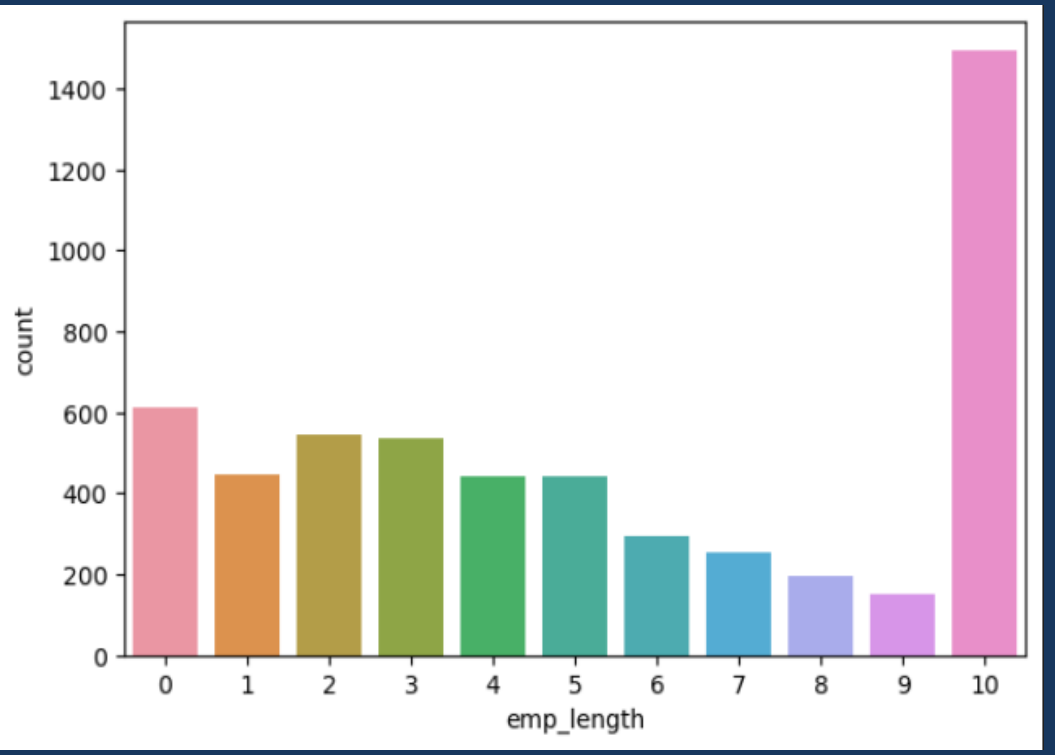
UNIVARIATE ANALYSIS- 2

Analysing interest rate wrt the interest rate bins created and emp_length

Interest Rate

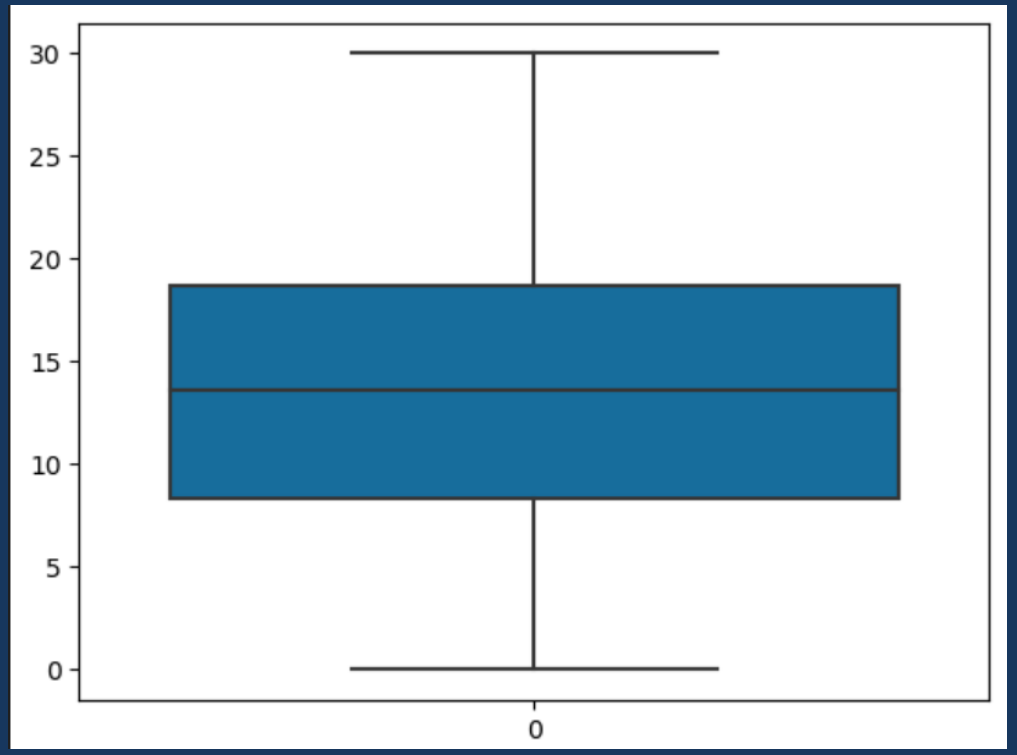


Years of Employment



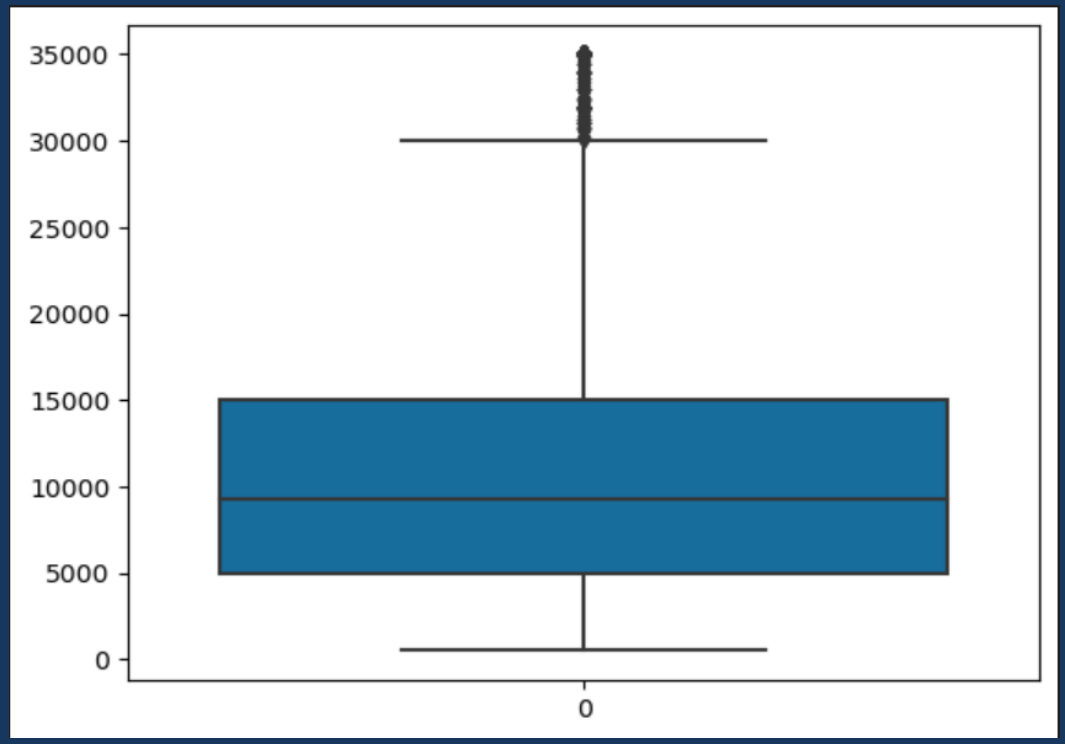
UNIVARIATE ANALYSIS -3

- DTI - Plot on `dti` looks pretty good, no outliers are found in the data set



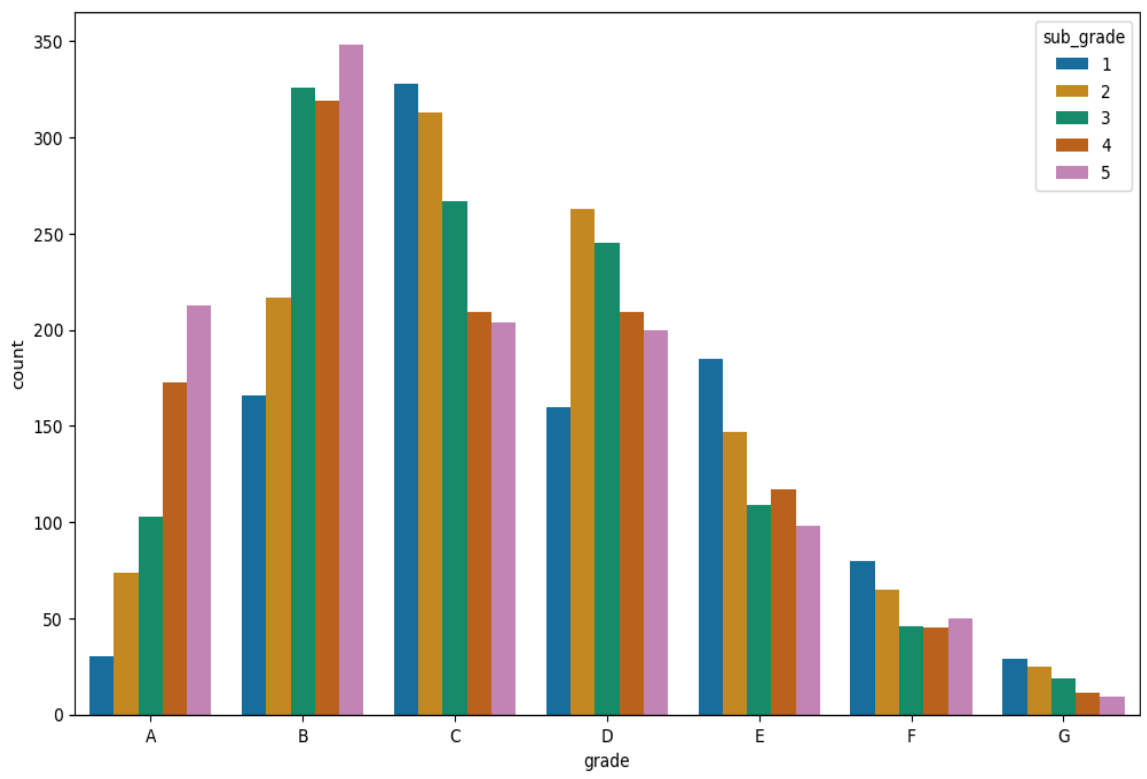
- loan_amnt

Though there are some values far from distribution, the distribution is continuous and there is no need to remove outliers / extreme values for these above columns

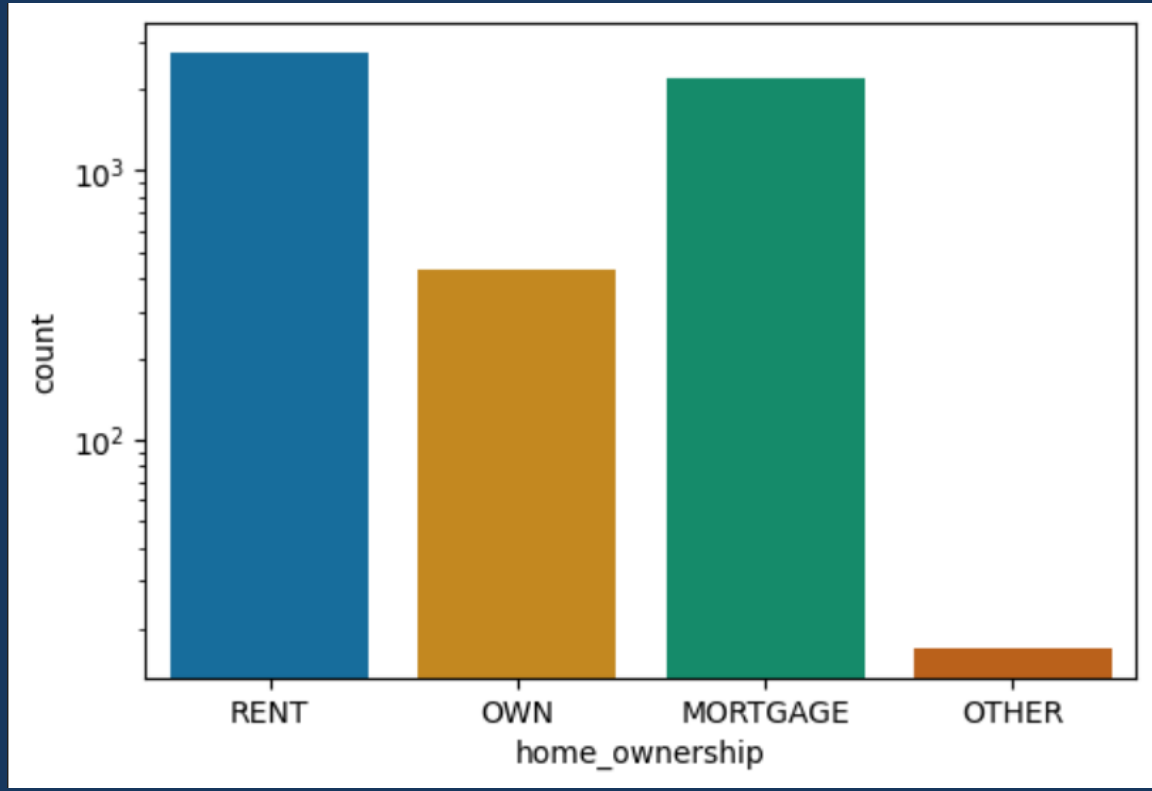


UNIVARIATE ANALYSIS- 4

Grade - As we already have grade column, extracting only subgrade (int level value). We are analysing and visualizing only the defaulter data. So, using only 'Charged Off' loan_status for below plots
Field `grade` A,B,C,D are quite dominant amongst all. Grade B counts is almost 350 , whereas grade G is less than 50.



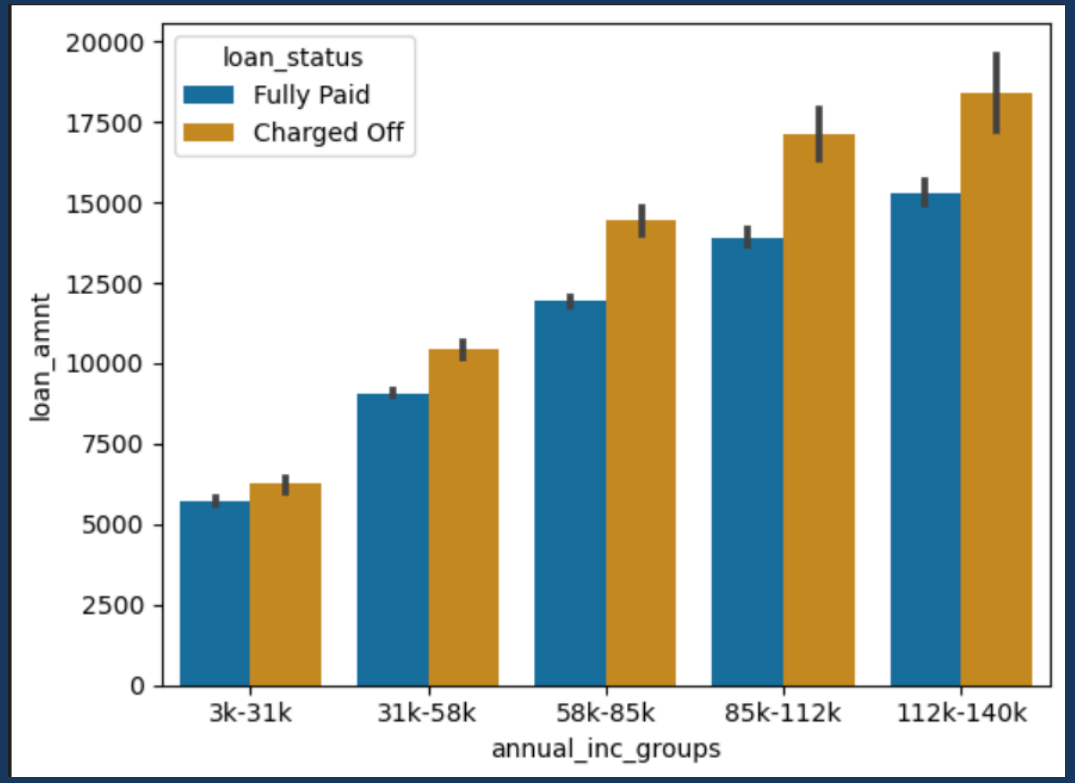
Home Ownership - Most of the loan applicants house type is either Rent or Mortgage. A very small portion of the applicants have their own houses.



BIVARIATE ANALYSIS - 1

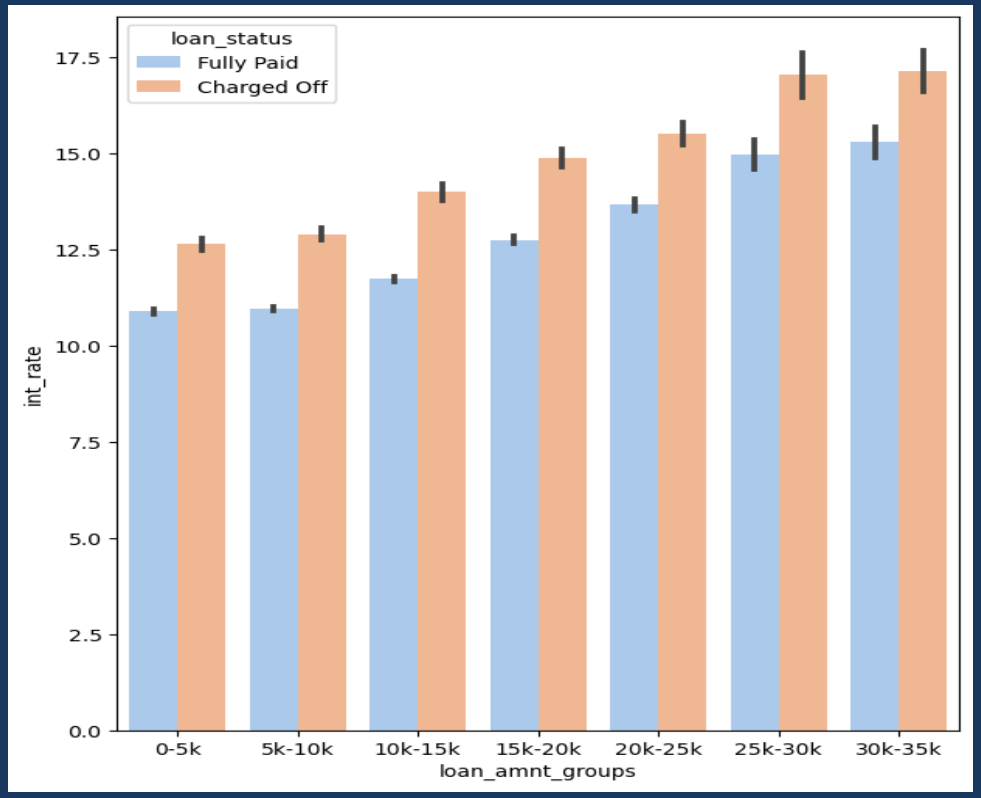
- ## Annual income and Loan amount

  Visualizing the annual income groups wrt loans which are Fully paid or Charged Off



- ## Interest Rate and Loan amount

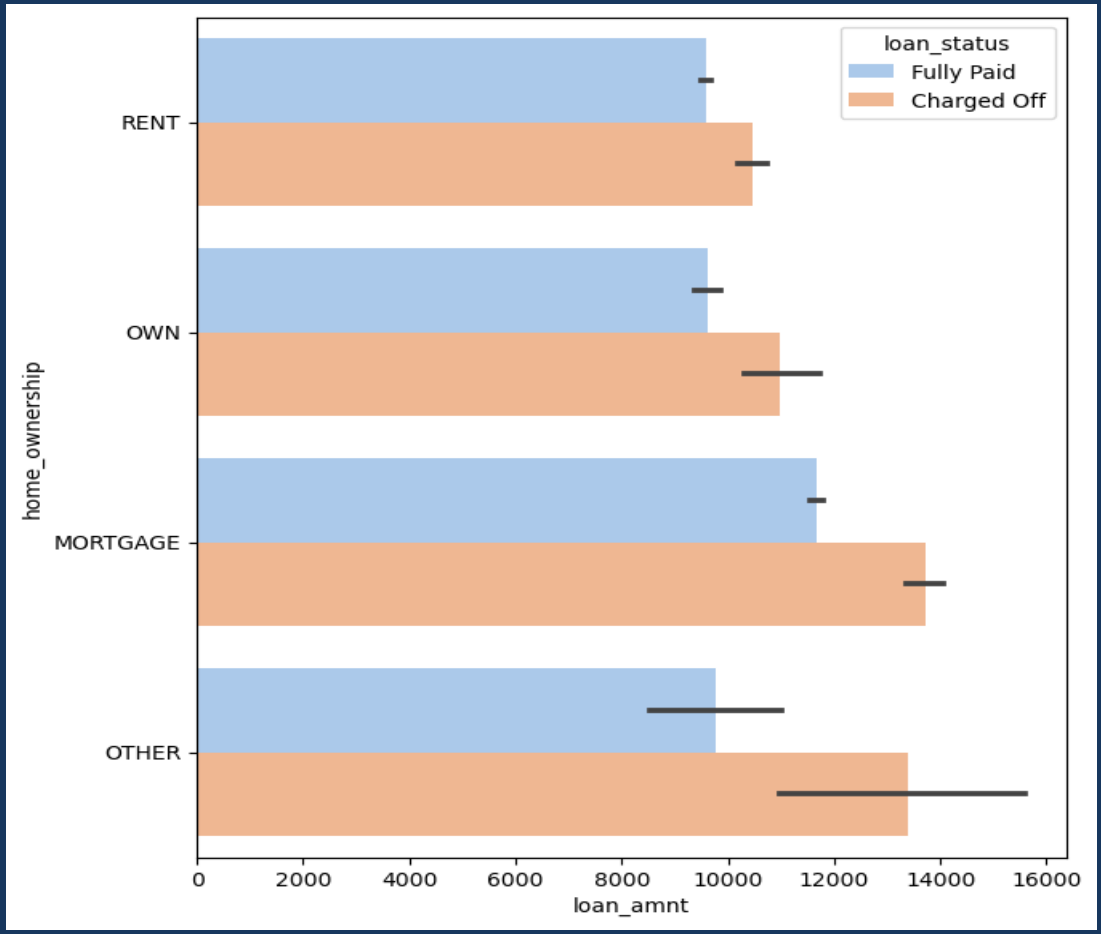
  We get a clear indication that chances of loans getting charged off is very low if the interest rate is low. It gets higher with the increase of interest rate and reaches the pick at 27% and more interest rate.



BIVARIATE ANALYSIS - 2

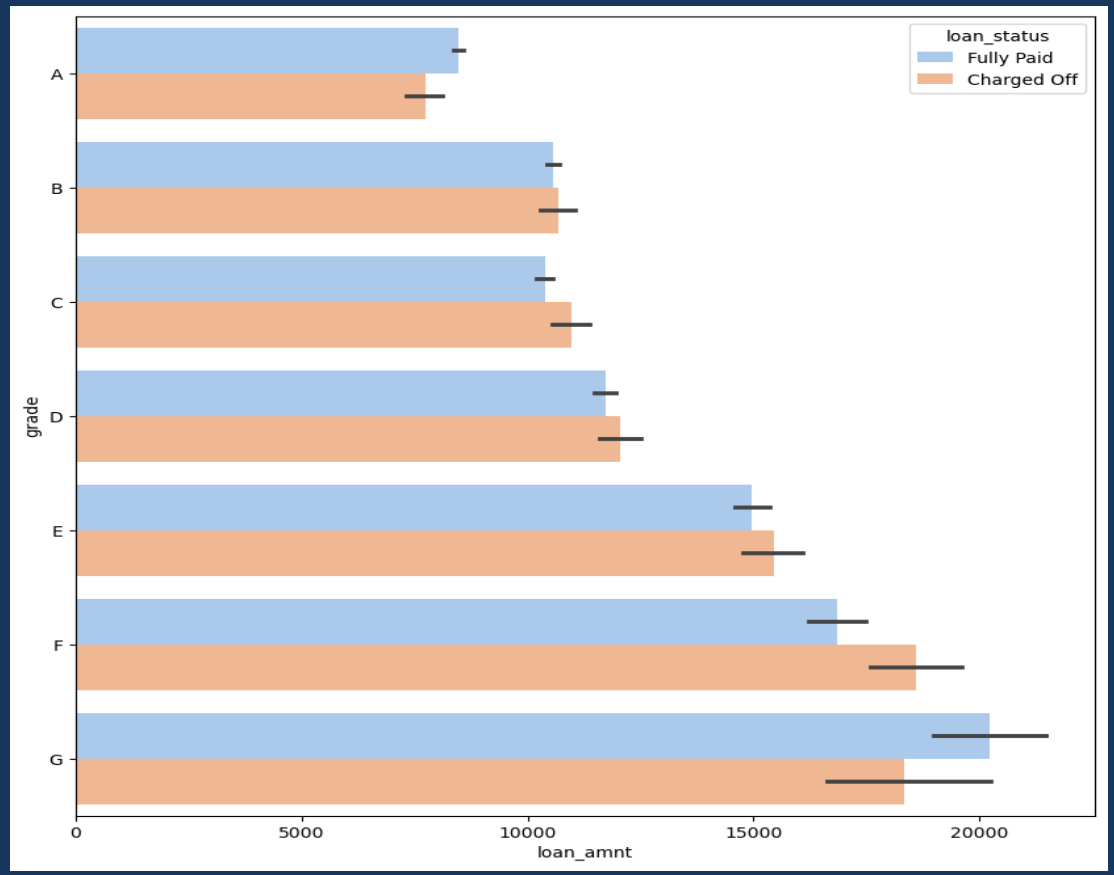
Loan amount and House Ownership

The above plot reveals a very eminent trend that loan term 60 months are much more prone to get charged off than of the loan terms 36 months.



Grade and loan amount

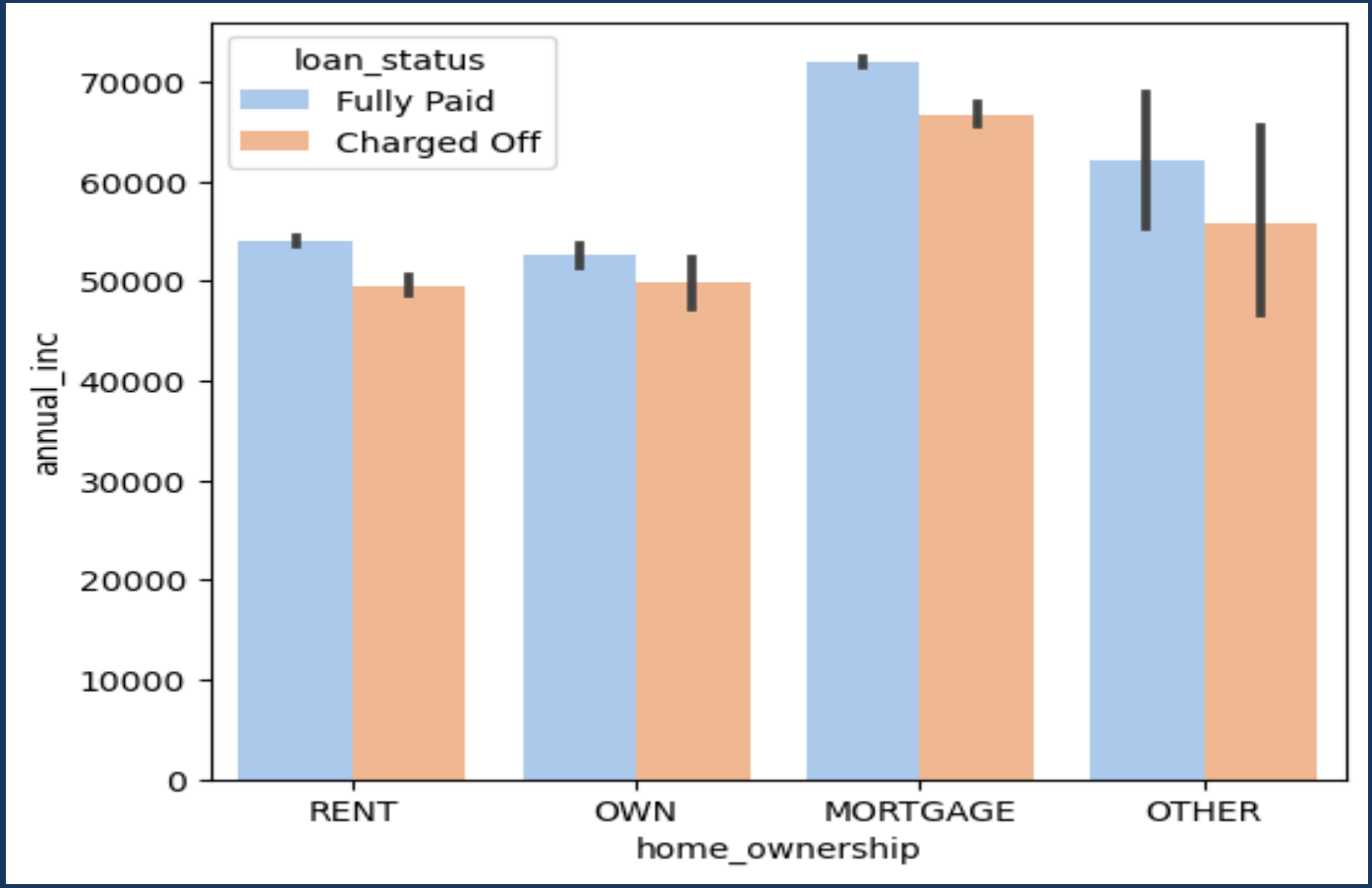
We find an almost linear line between loan grade and the charged off loans. As the loan grade increases the chances of a loan to be charged off increases. It also proves that higher grade loan bears much more credit risk for the lenders.



BIVARIATE ANALYSIS - 3

Annual Income and House Ownership

Loan Default is higher for people who already has mortgage for home

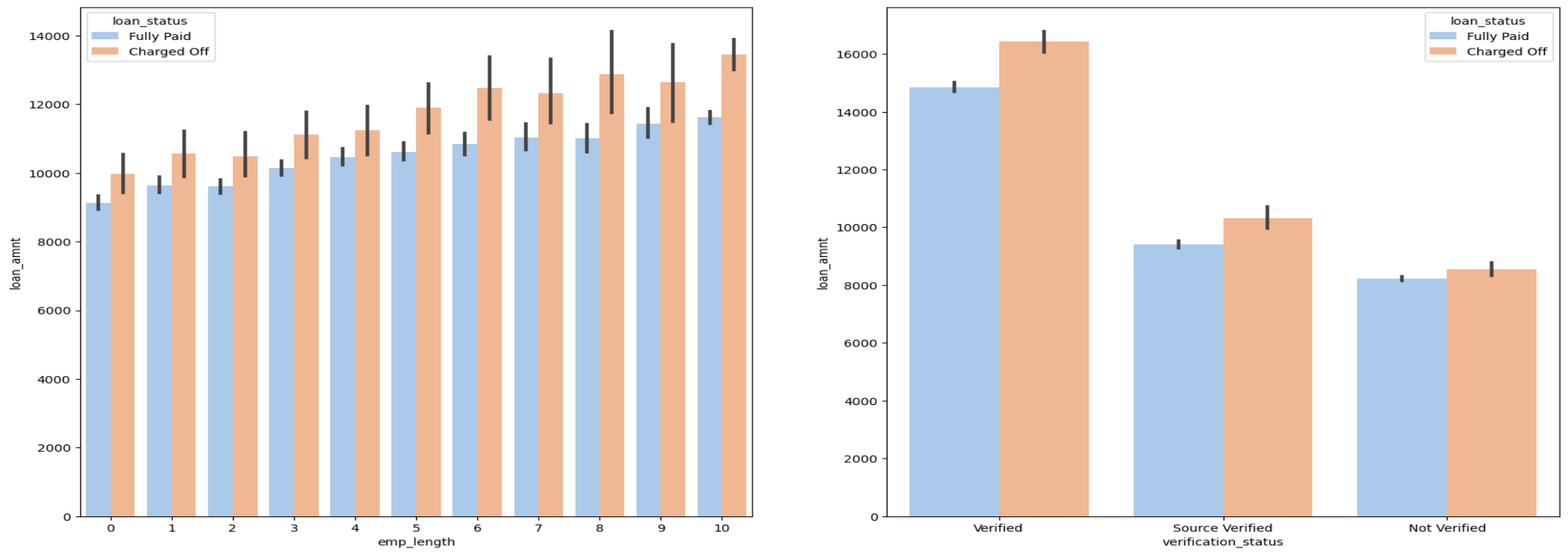


BIVARIATE ANALYSIS - 4

Employment length and Charged Off loan ratio and verification status

We see most loans are charged off when the employment years is 10 years or above,

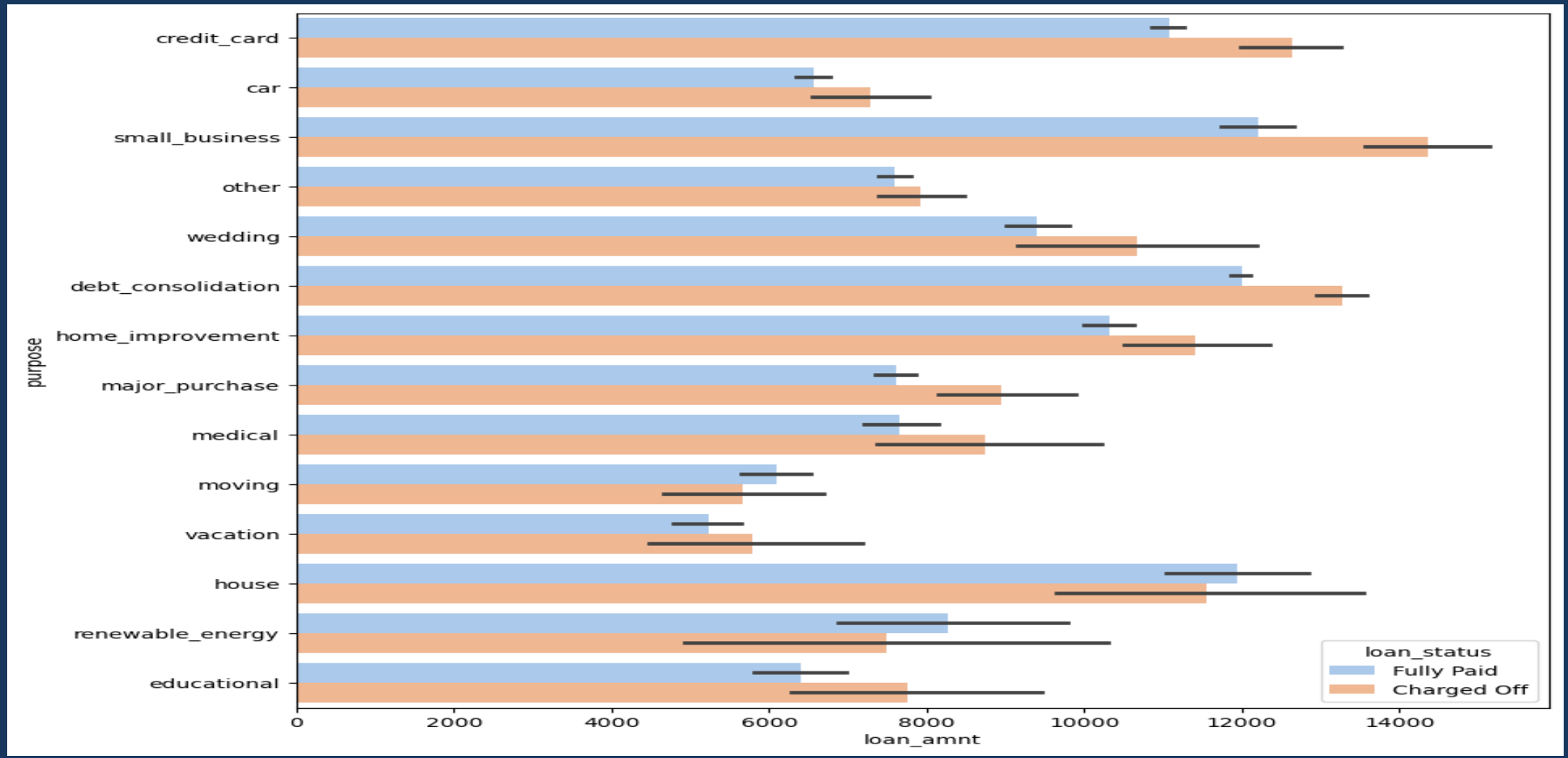
verification status does not have much influence on Charged off loans.



BIVARIATE ANALYSIS - 5

Purpose and Charged Off loan ratio

We see that for small business the chance is highest that the loan may get charged off. And it is the lowest for loan taken for marriage.



INFERENCES

- Applicants taking loan for 'home improvement' and have income of 60k -70k
- Applicants whose home ownership is 'MORTGAGE and have income of 60-70k
- Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k
- Applicants who have taken a loan in the range 30k - 35k and are charged interest rate of 15-17.5 %
- Applicants who have taken a loan for small business and the loan amount is greater than 14k
- Applicants 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 10yrs 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%

THANK YOU