

Lending Club Analysis

Data Understanding

- Here, after looking into the data we found that some of the important information which could help us better in deciding the defaulting behavior of the loan was actually unavailable. Some of them are:
 - **FICO score** – which summarizes about the credit score of the borrower
 - **Avg_cur_bal**- which summarizes about how much balance the member is maintain on average. This could have given us insight on his expenditure pattern
 - **Tax lien** – to understand if the member is failing to pay the tax timely
- Around 50% of the data column has N/A value

Making Improvement on these area could have increased the data quality even better.

Data Cleaning

- Column with only N/A values are removed
- Some column with constant values were identified and removed.
- Sanity Checks were performed:
 - To find duplicate member information
 - To see if the funding amount was more than what was asked in loan
 - To see if there are some value with future payment date whose loan status is already paid-off whereas whose future payment date was missing but loan status was active/current
 - Some column with extreme value than other values were identified for eg: annual income

Data Manipulation

- Looking into the data after basic cleaning and sanity check the types of the data and the null-count were observed to find out misclassified data type due to some junk characters and N/A value.

Some misaligned data-types observed were:

int-rate : should be float not object

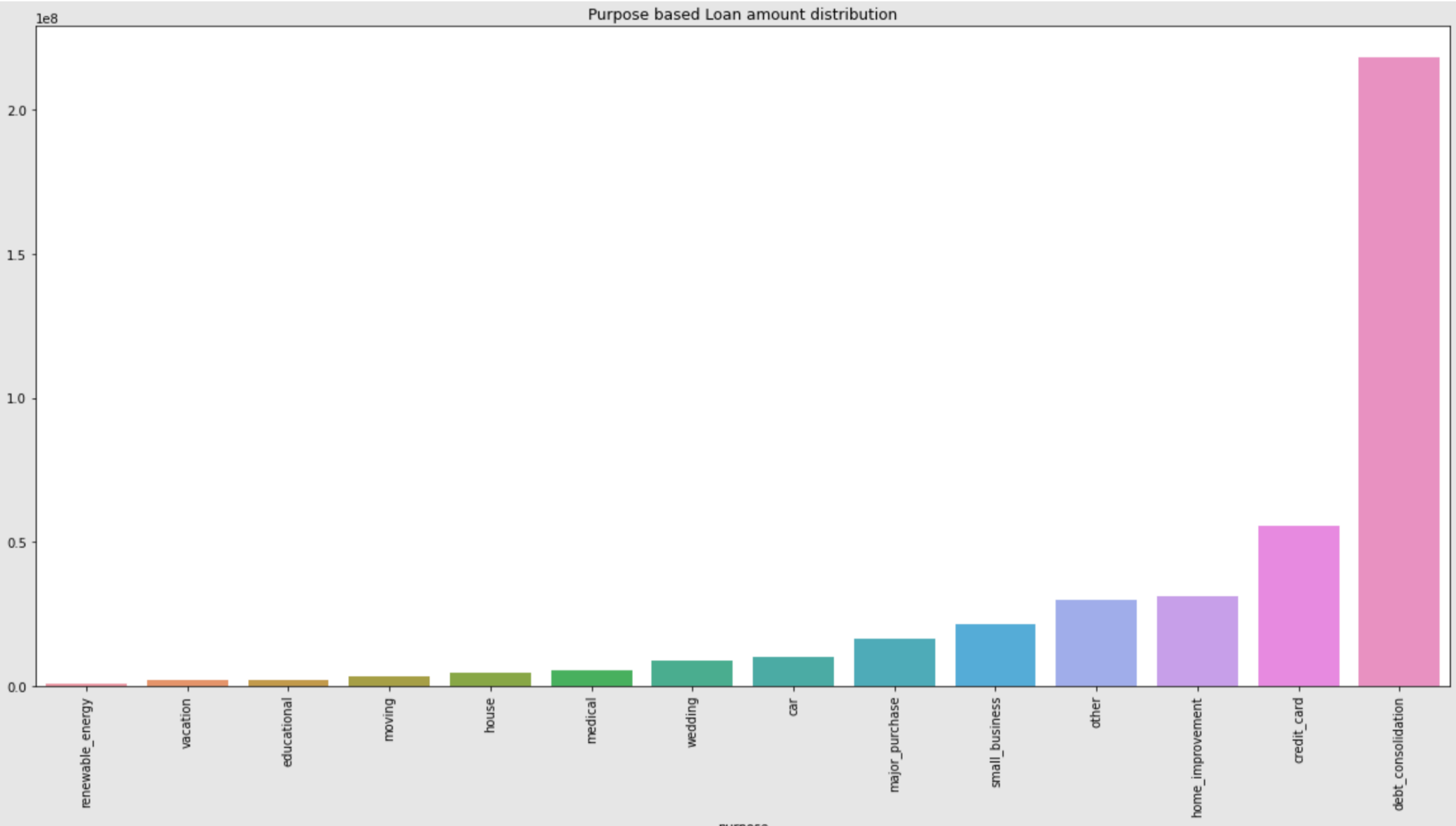
term,emp-lenght,zip_code: should be int not object

issue_d,last_pymnt_d,next_payment_d : should be datetime not object

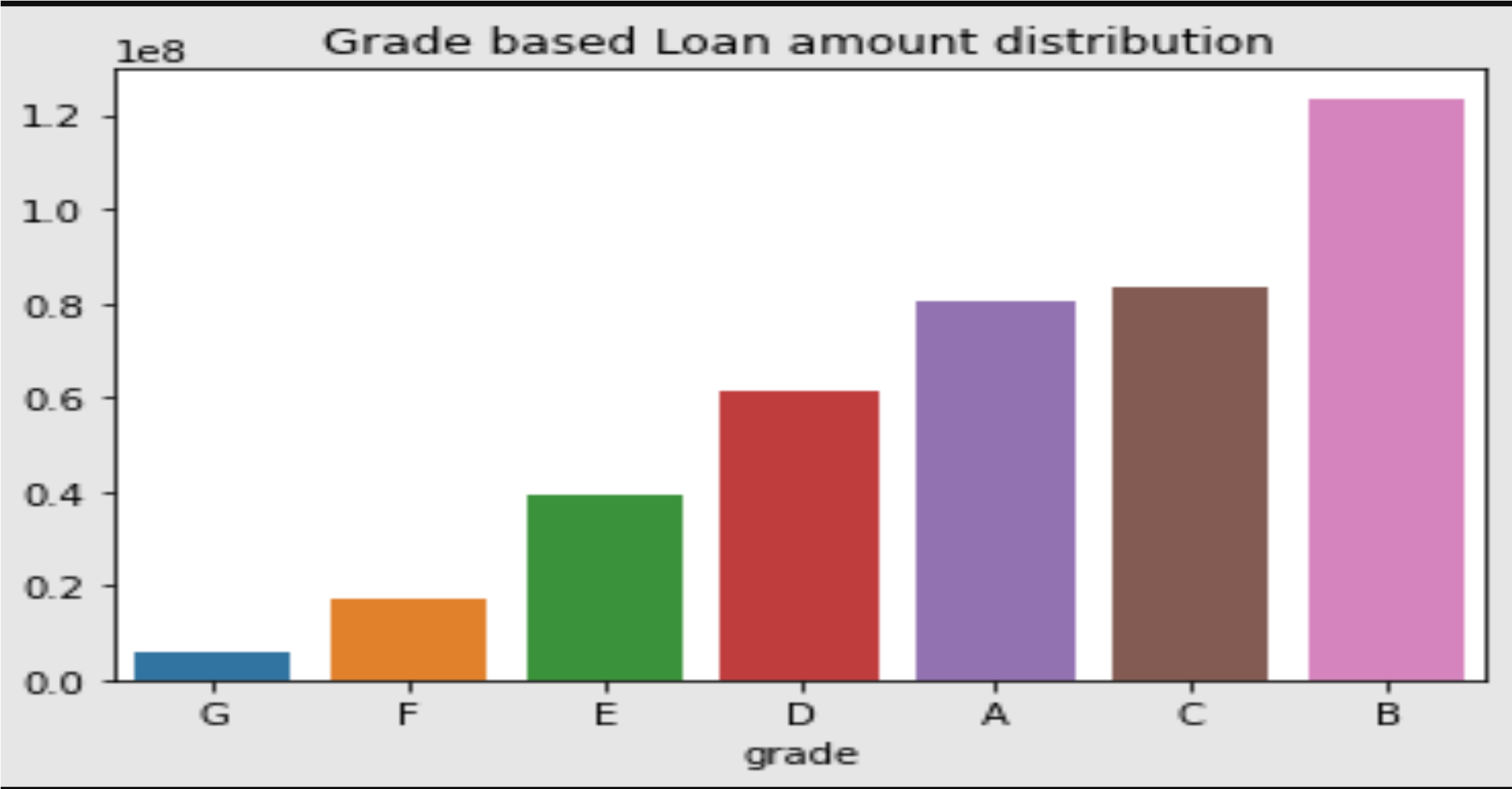
Data Analysis (Univariate and Segmented)

- Distribution of the overall loan amount over different category
- Distribution of the different grades of the loan
- Distribution of application type for the loan disbursed
- What is the relation in loan application category wrt to his income or average balance
- Variation in the loan disbursed amount
- Variation in the loan tenure
- Average collection of installments
- Variation in the interest rate for different category of loan i.e loan grades

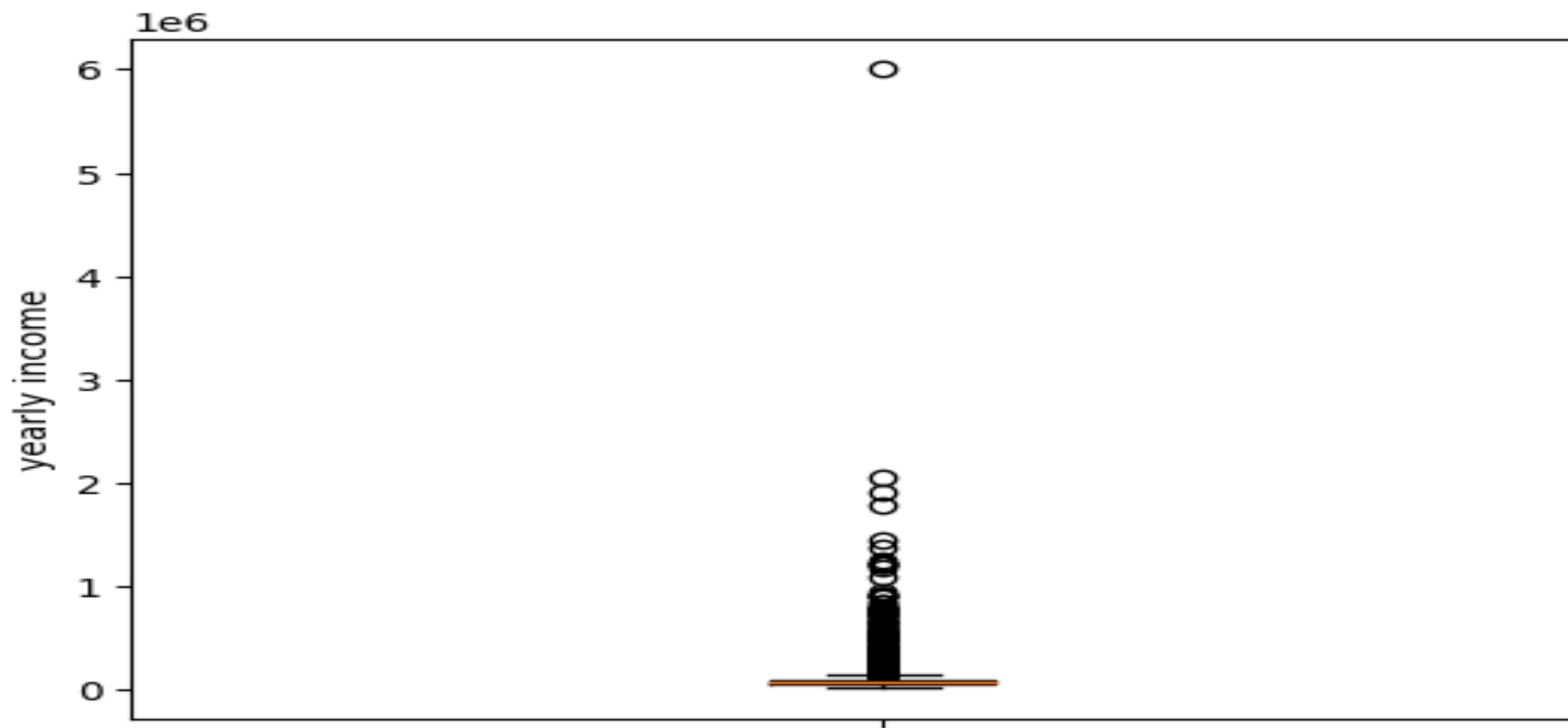
Seems, lending club is more famous choice for debt consolidation purpose.



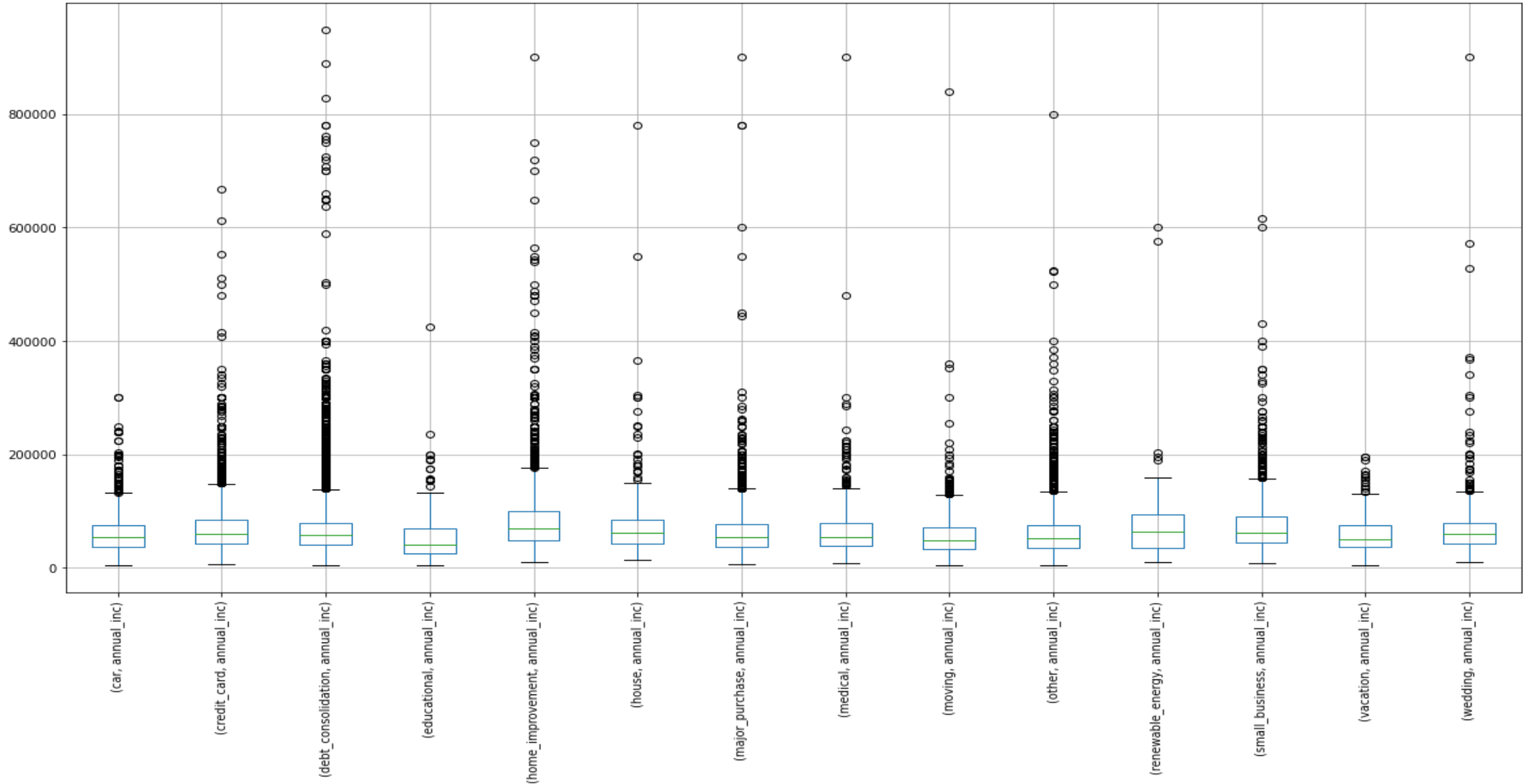
Most of the member loan request are graded as “B”



Some richy-rich are also involved in the lending club they might make the analysis skewed . So its better to keep them out of the analysis as of now.



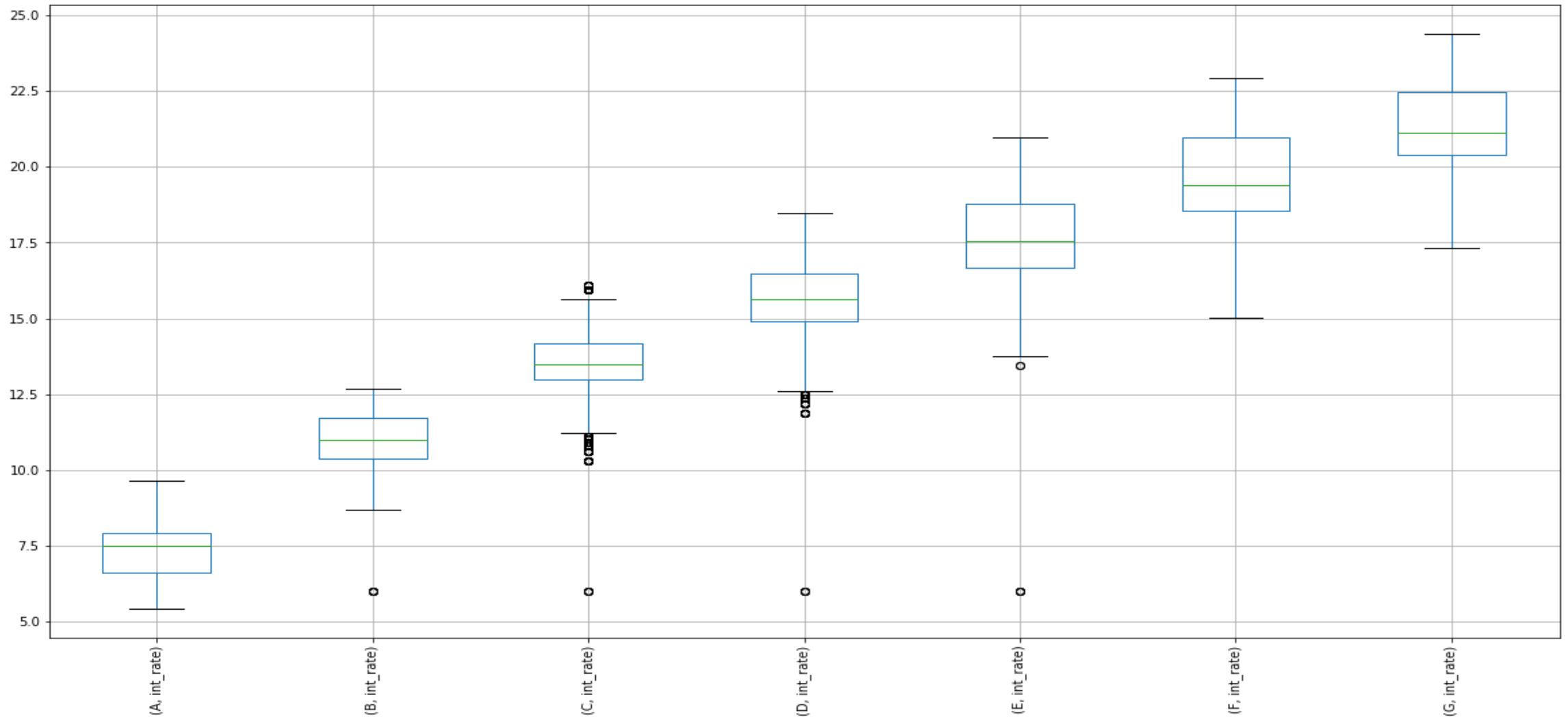
Keeping richy-rich out some lower-middle class groups seems to be using lending club to finance their dream of education, vacation and 4-wheelers.



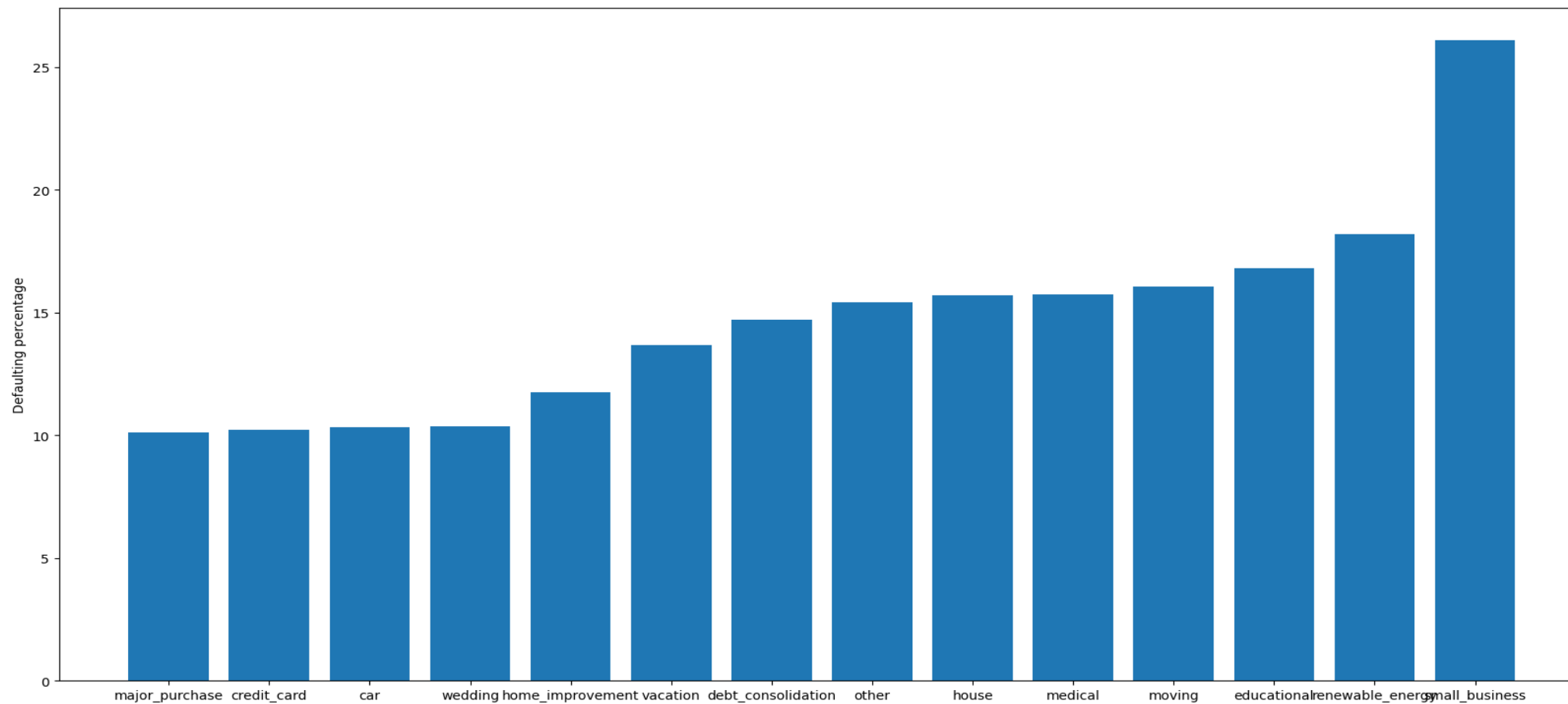
Data Analysis(Bi-variate,Multivariate)

- Relationship with the Grade and interest rate
- If some address has more defaulting members.
- loan category having high defaulting tendency
- interest rate with loan defaulting tendency
- loan grade with defaulting tendency
- emp. length with loan defaulting tendency
- public bankrupt declared application result in defaulting or not (pub_rec & pub_rec_bankruptcies)
- high loan amount and less tenure result in default or not
- annual income with defaulting tendency
- if higher open credit line has any relation with the defaulting rate.
- dti ratio and defaulting tendency relation
- verification status and defaulting tendency
- term or loan with defaulting tendency (term_in_months)

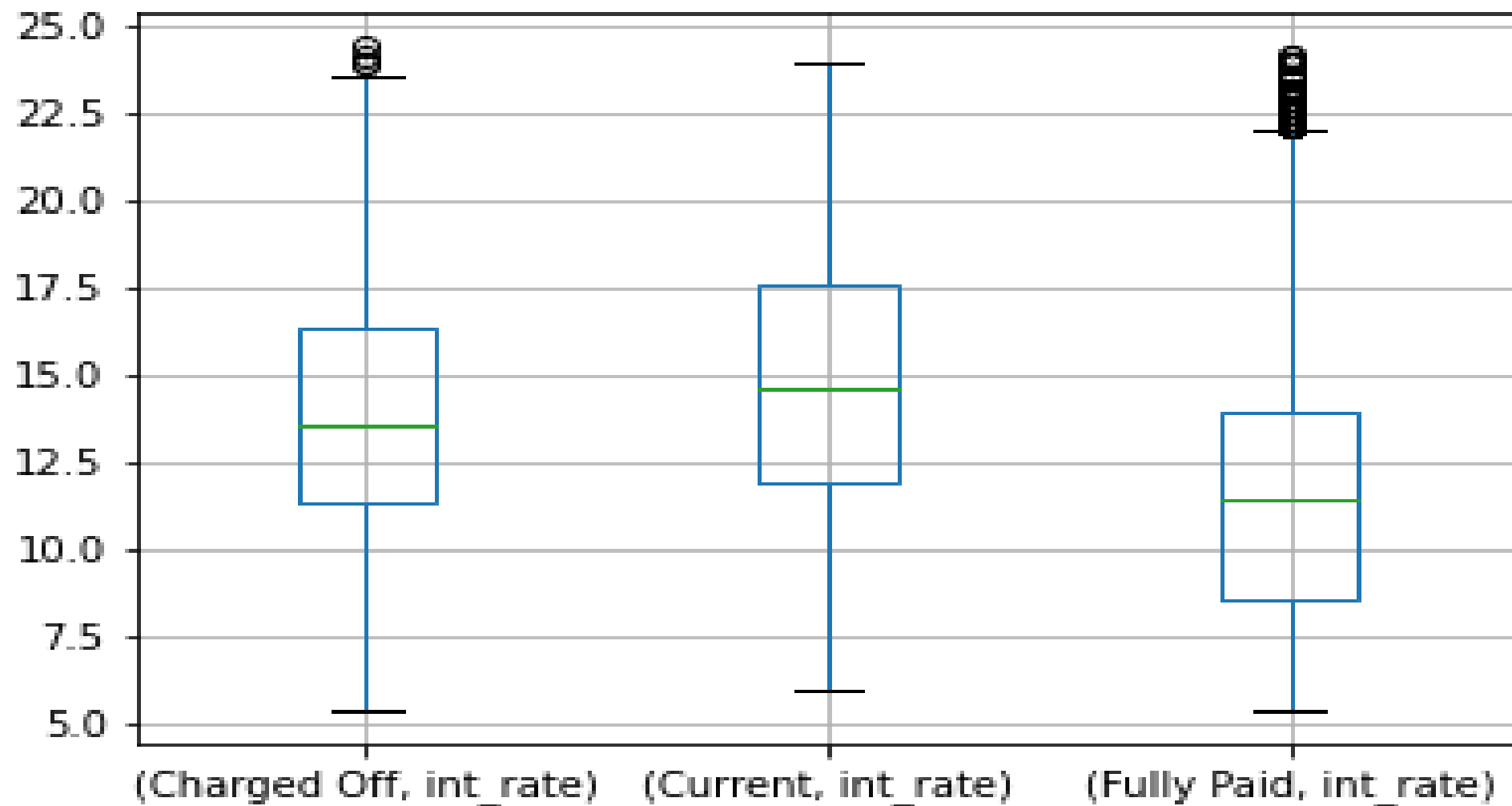
Lower Loan Grade will have to pay more, perhaps they are more risky client of lending business.



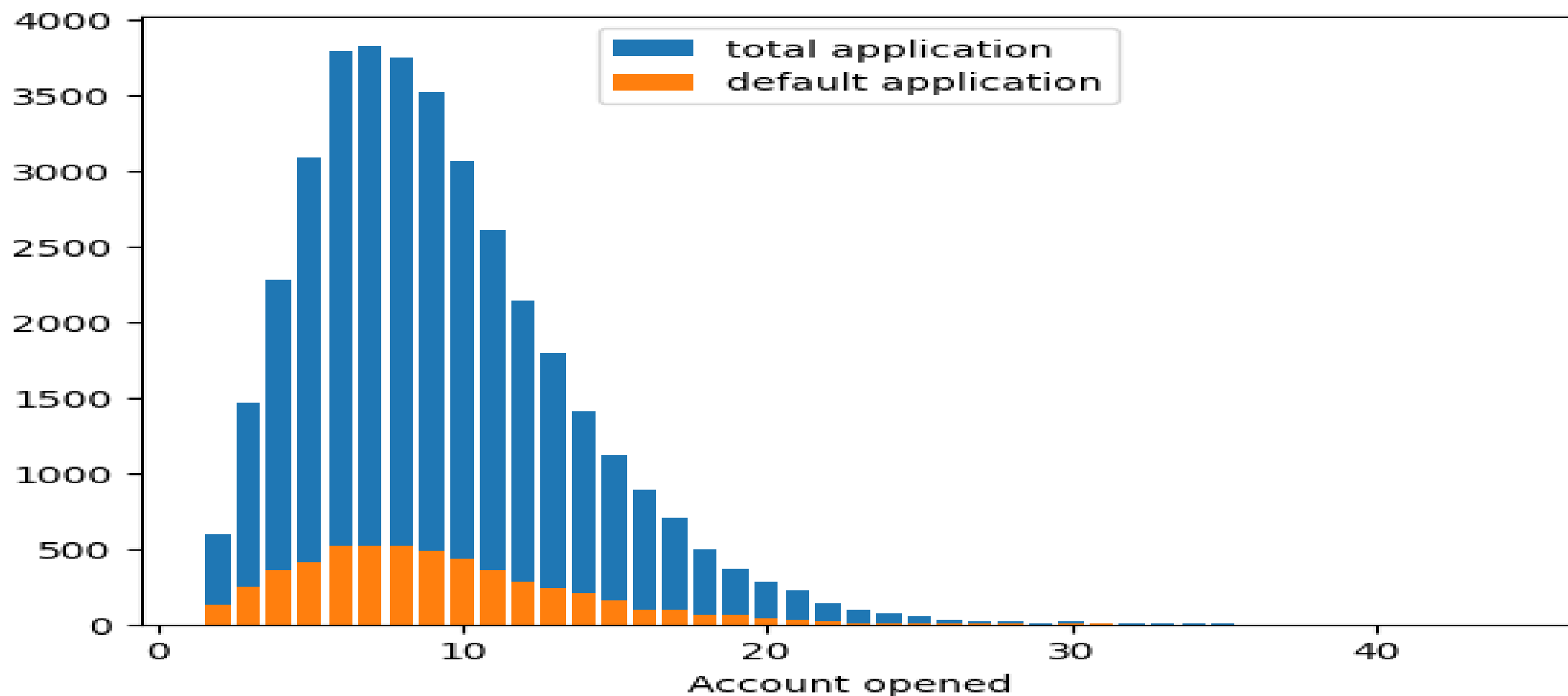
“Taking money for small business”, investor should think first before lending



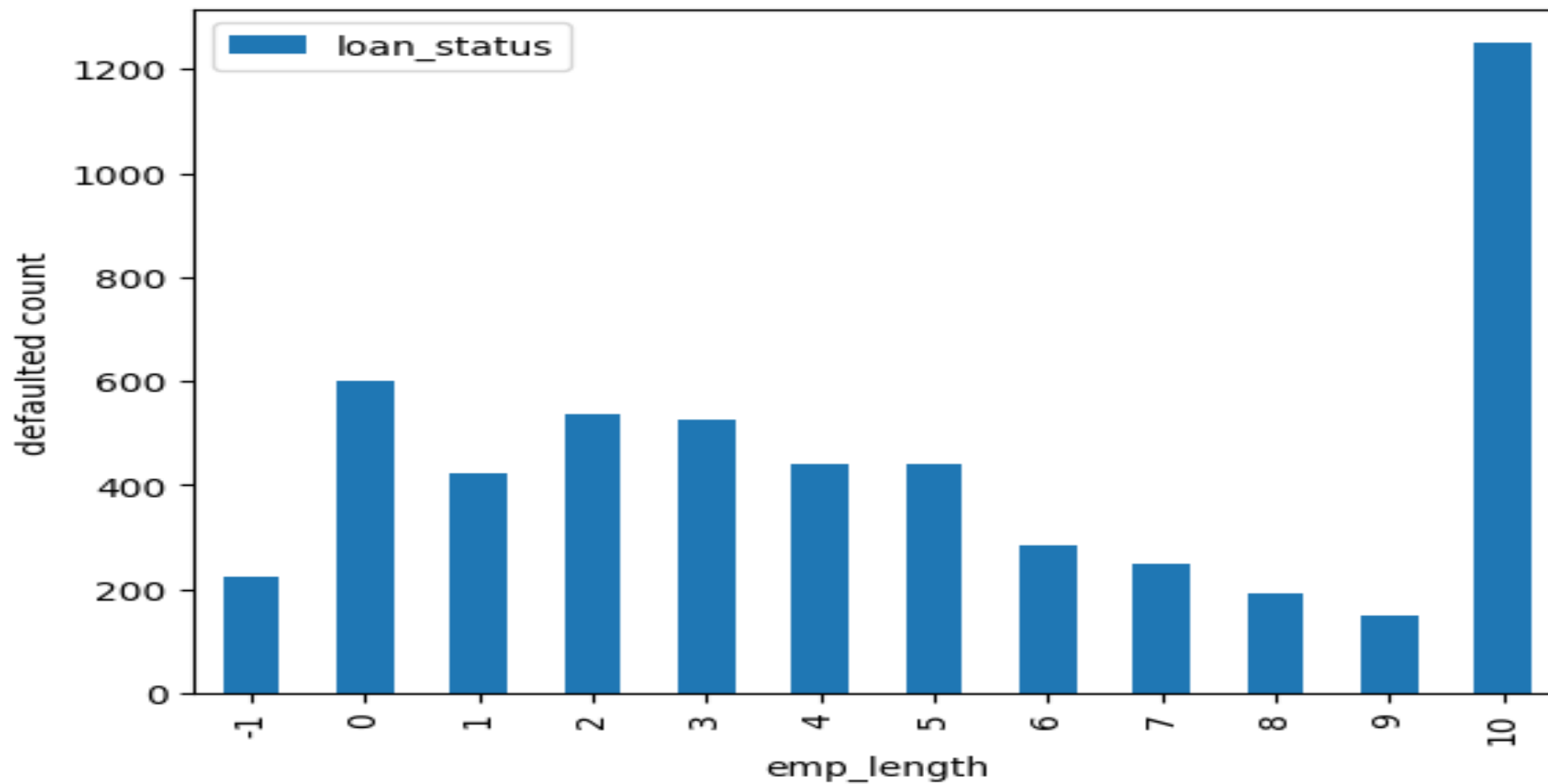
Lending at higher rate or lower rate , doesn't impact on being defaulted loan.



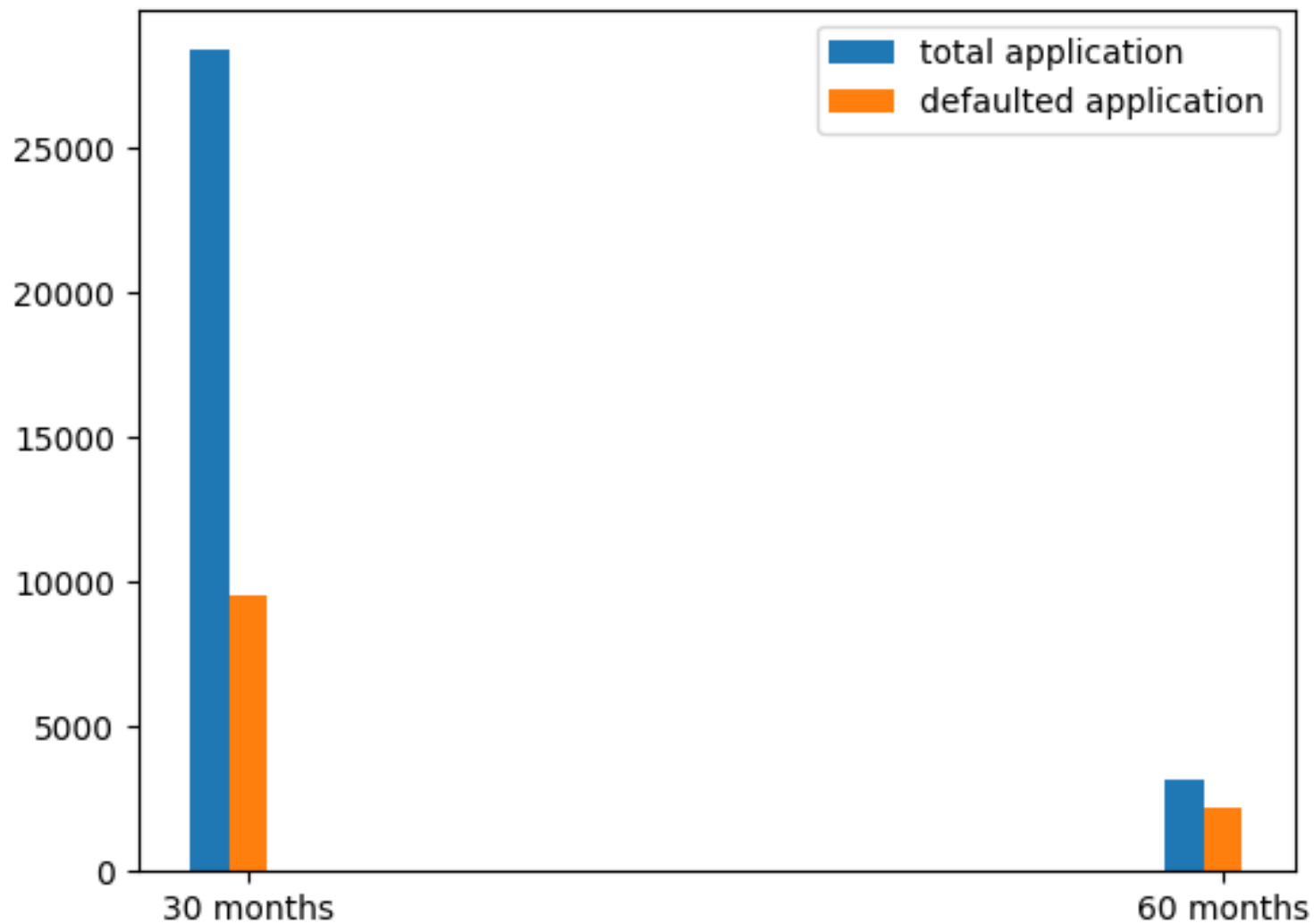
Number of account opened in several banks, not necessarily result in defaulting but we never know.



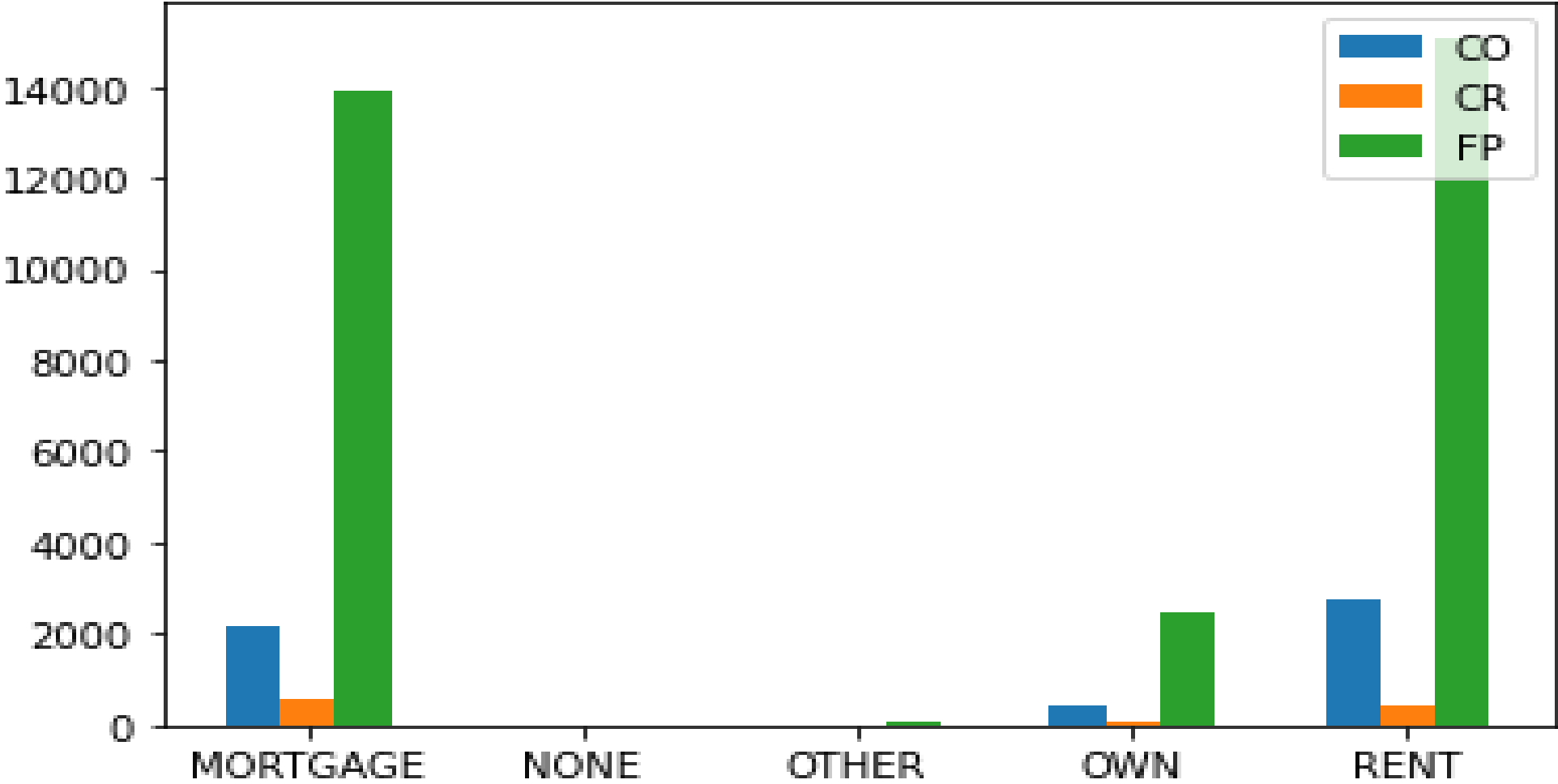
Employed for more that 10 years or between 0-3 year, risk is there to lend money



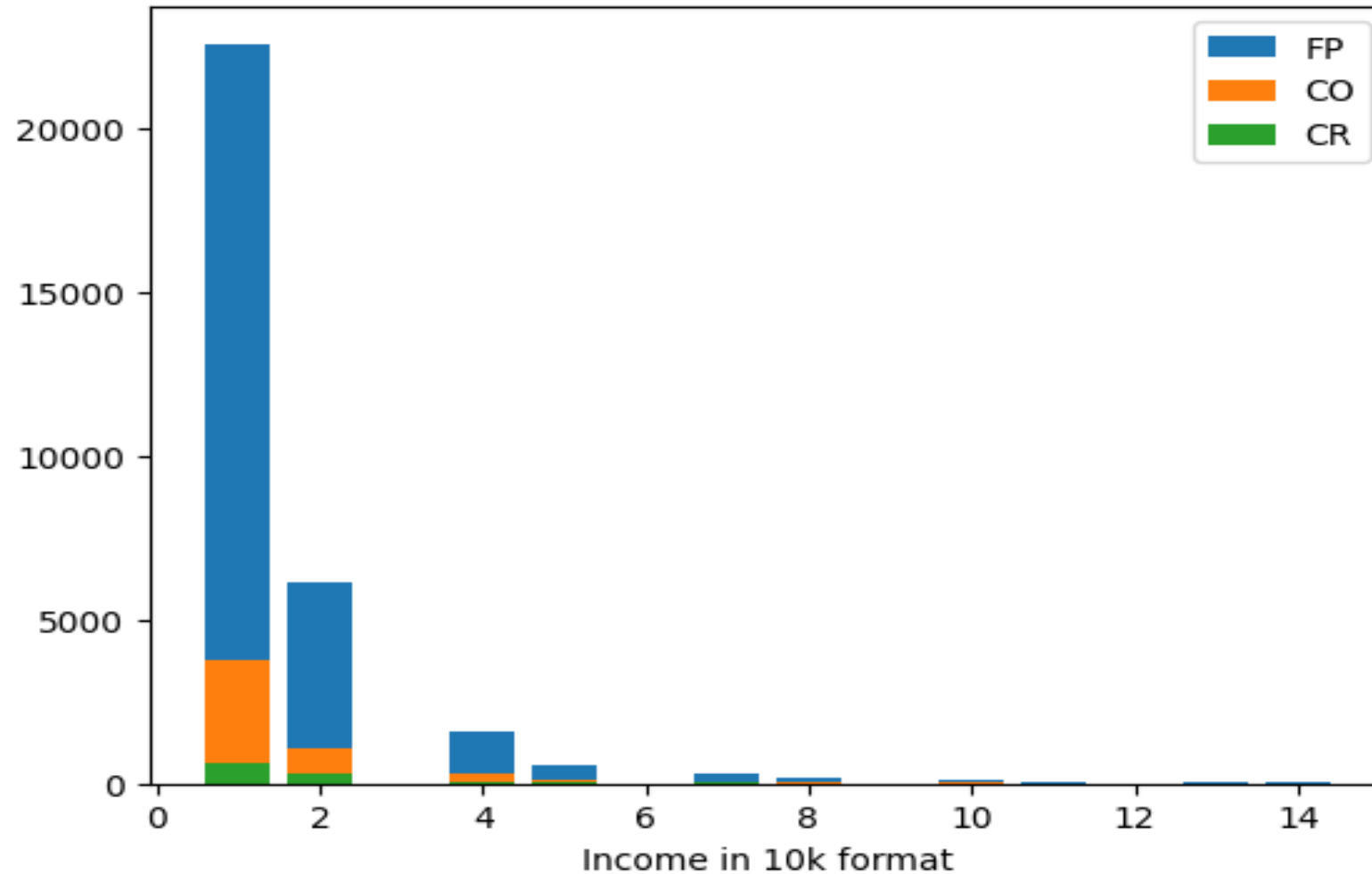
Most of the loan taken for the 60 month tenure is defaulted, so we should be careful while investing in the long term loan.



People with mortgage and rented house has relatively higher changes of being default in loan



With the lower income the noticable chargeoff proportion can be seen which means that if the user is maintaining lower revolving balance then he is likely to be defaulting loan amount



Recommendation

1. When the **purpose** mentioned is as "new business" see detail on the type of business going to be started and its operation and feasibility before lending the sum
2. When the borrowers has their **house not owned** in their name do better analysis before lending "mortgage" or "rented"
3. Look out for **loan grade** "B/C/D", risky to invest.
4. When employed for more years doesnot mean borrower has the cash backup to repay for eg: **emp_length** 10+ years.
5. risk of defaulting increases with the **tenure** of money lended.
6. lower **revolving balance** can increase the chances of defaulting