Lending Club Case Study

Executive PG Programme in Machine Learning & AI - March 2023

Program by UpGrad

Created and submitted by

- Sagar Shinde

Index

- About the Case Study
- Objectives
- Dataset
- Conclusions

About the Case Study

- 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.

- In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labeled as 'charged-off' are the 'defaulters'.

Objectives

- The core objective of the exercise is to help the company minimize the credit loss. There are two potential sources of credit loss are:

  - Applicant likely to repay the loan, such an applicant will bring in profit to the company with interest rates.*
    ** Rejecting such applicants will result in loss of business**

  - Applicant not likely to repay the loan, i.e. and will potentially default, then approving the loan may lead to a financial loss* for the company

Dataset

The given data contains the information about past loan applicants and whether they ‘defaulted’ or not.

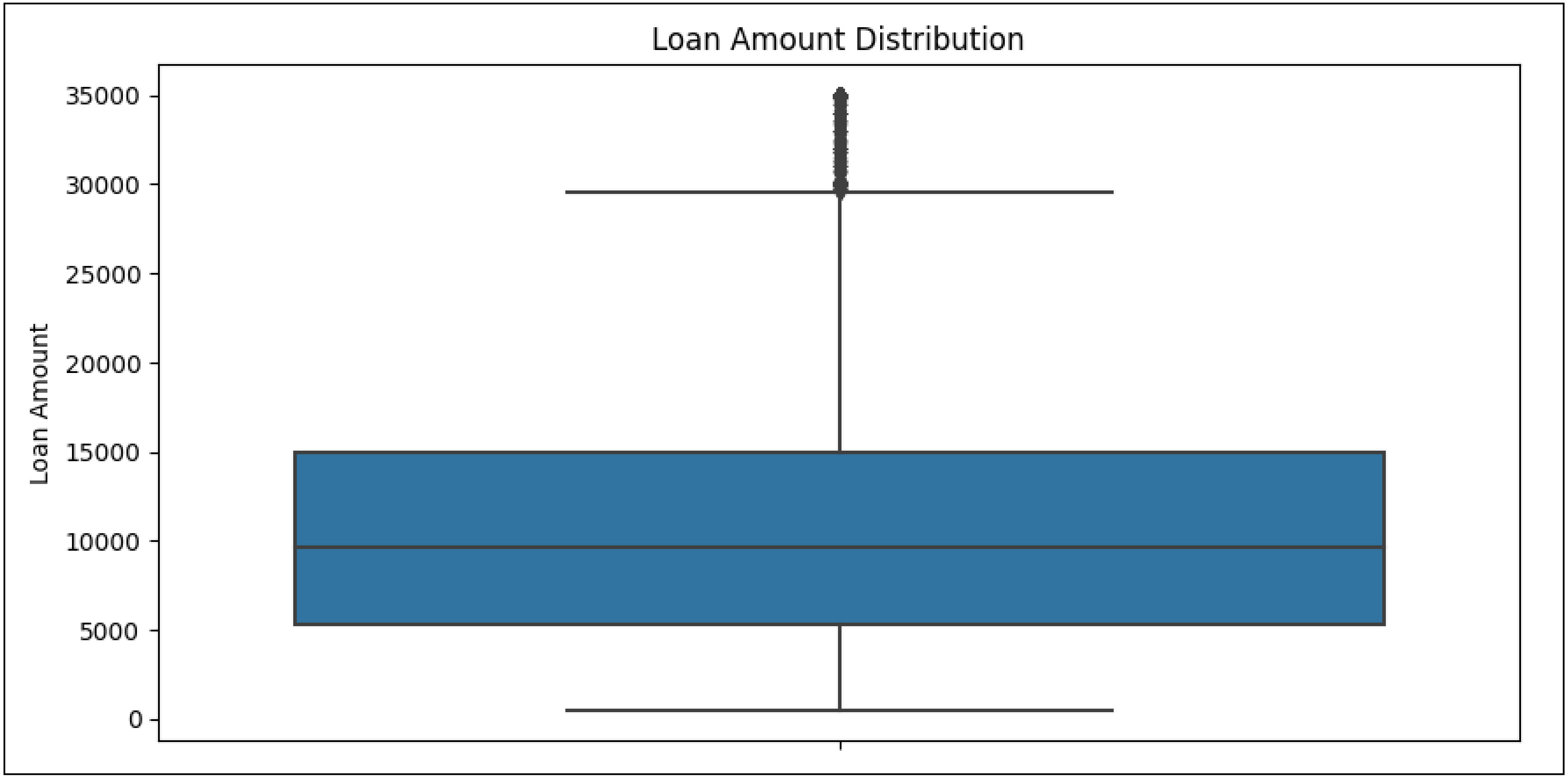
The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

The data contains 39717 rows and 111 columns.

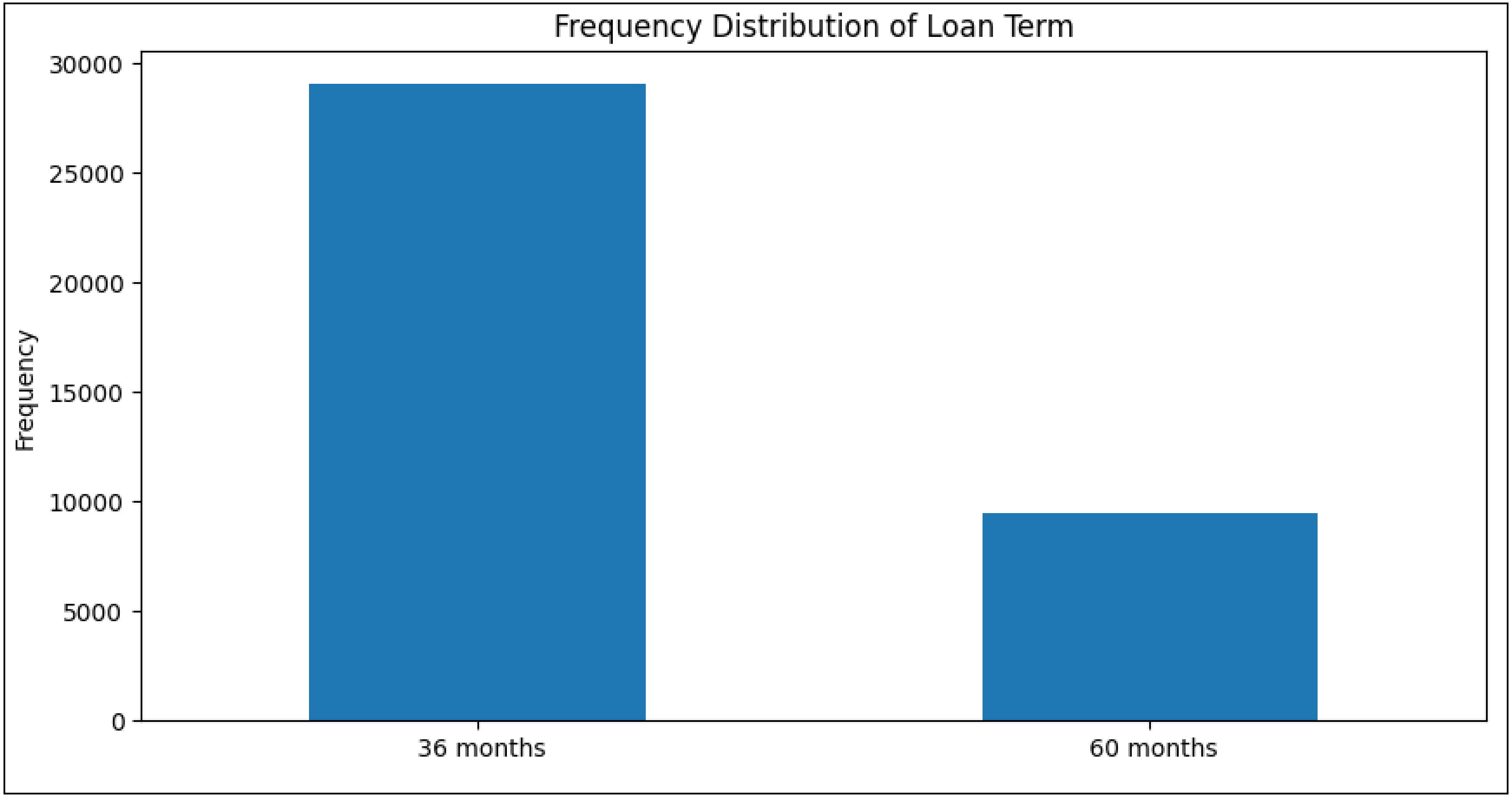
The goal of the analysis is to see who is likely to default and this can only be said in case of either fully paid or charged off loans.

Observations (from visualization of graphs)

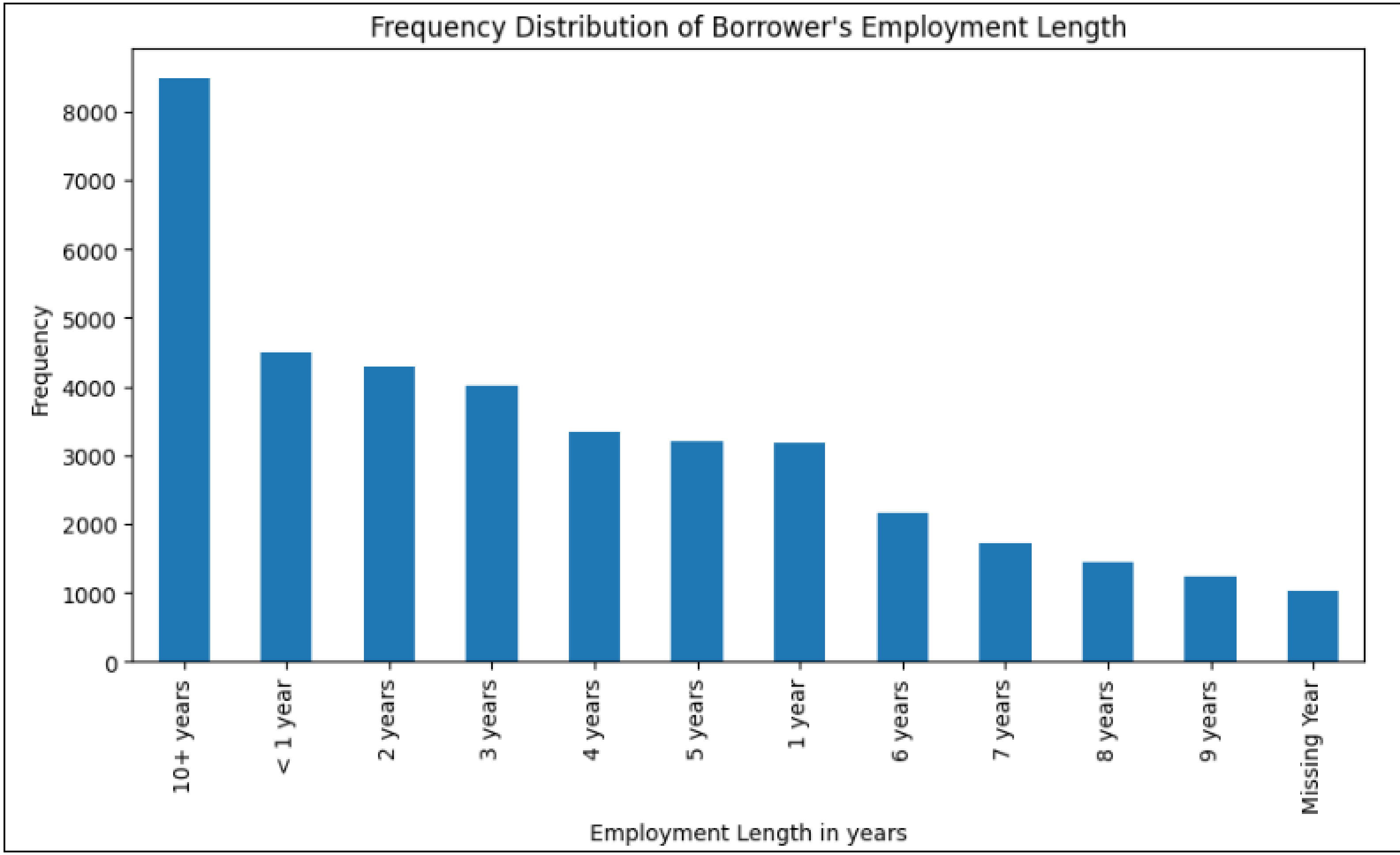
Univariate Analysis



- The loan amount distribution is from 5500 to 15000

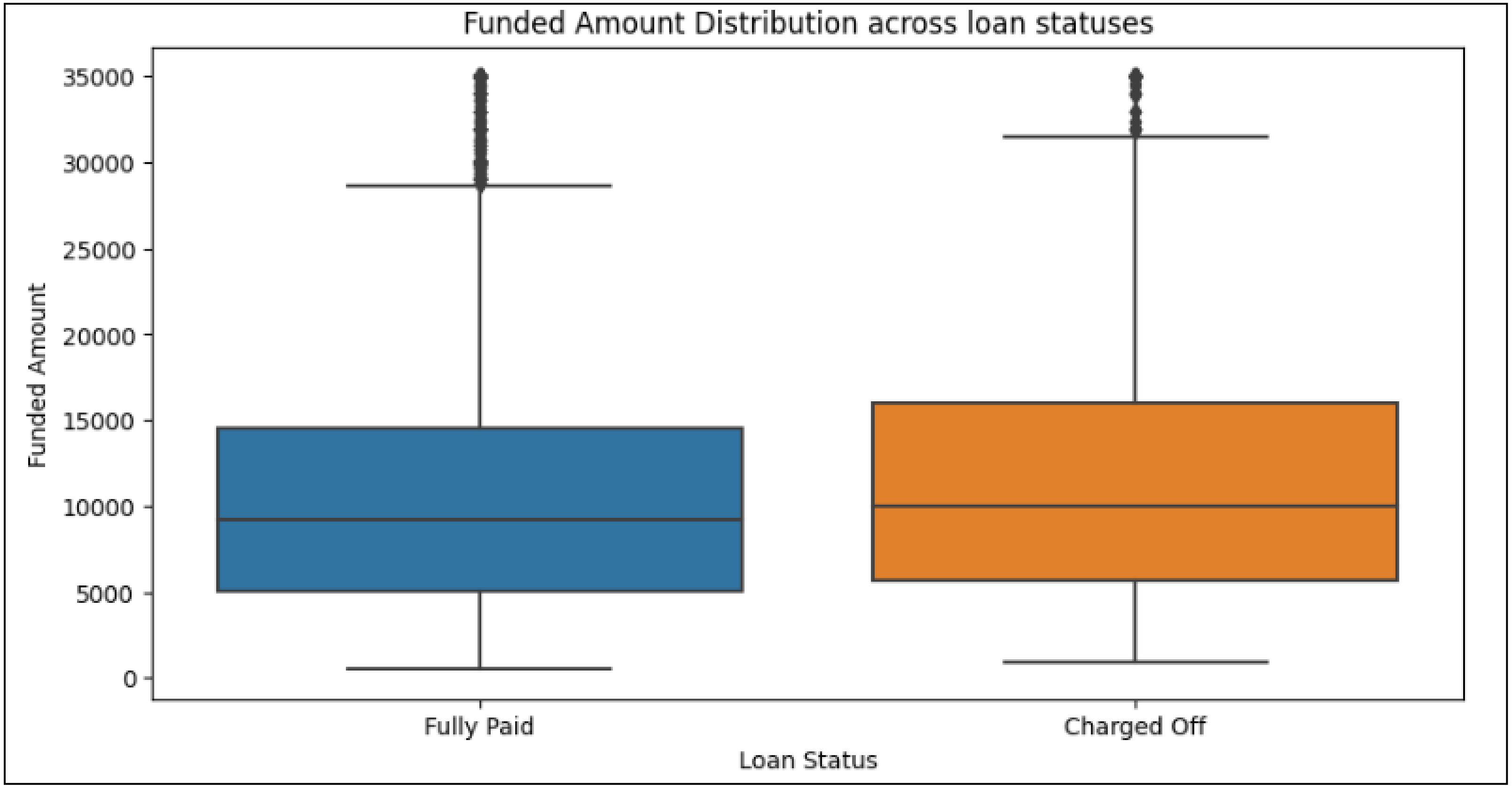


- The loan is taken mostly for lesser tenure i.e. 36 months.

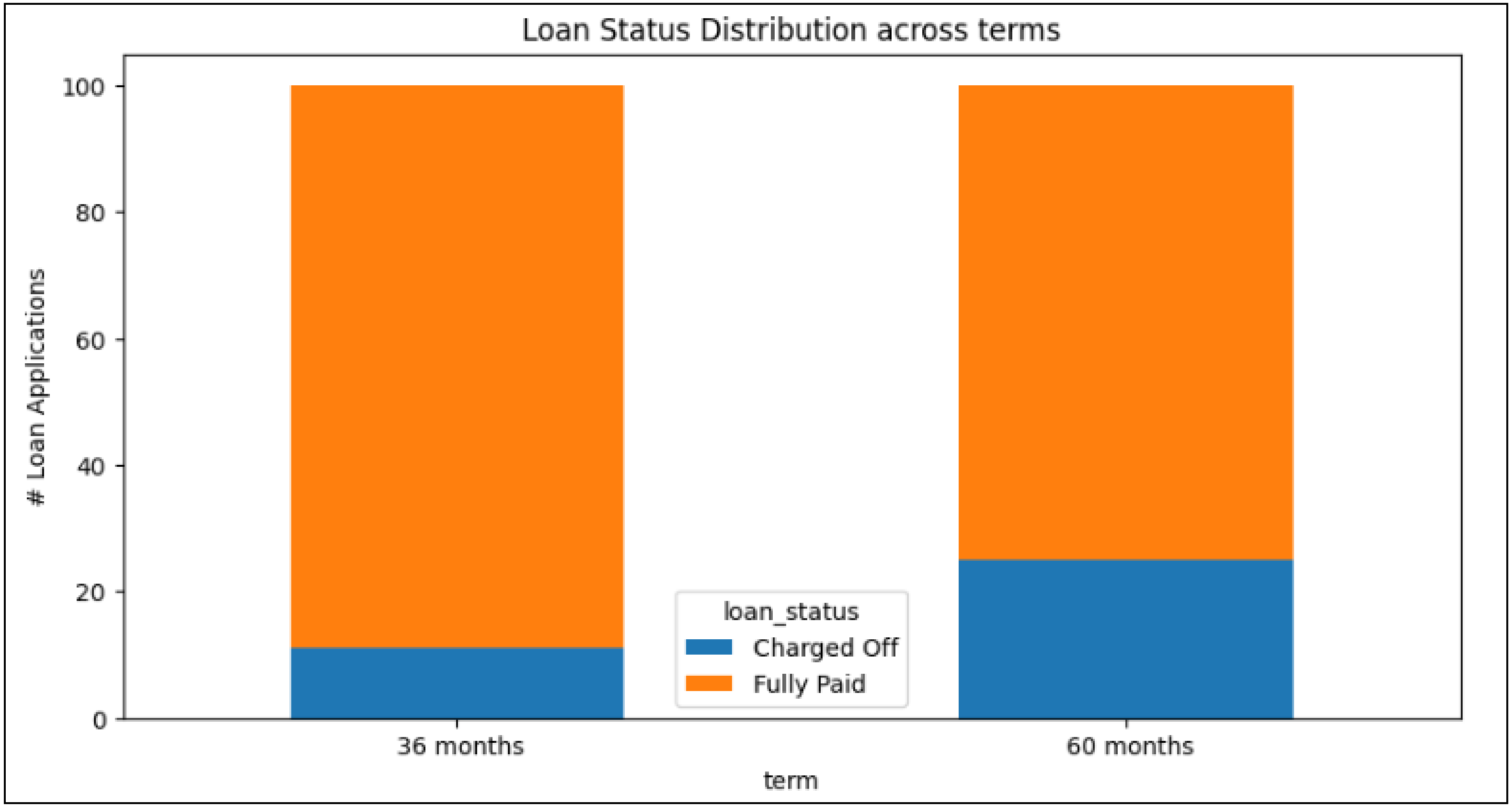


More loans are taken from borrowers whose employment tenure is more than 10 years and also those whose tenure is less than a year.

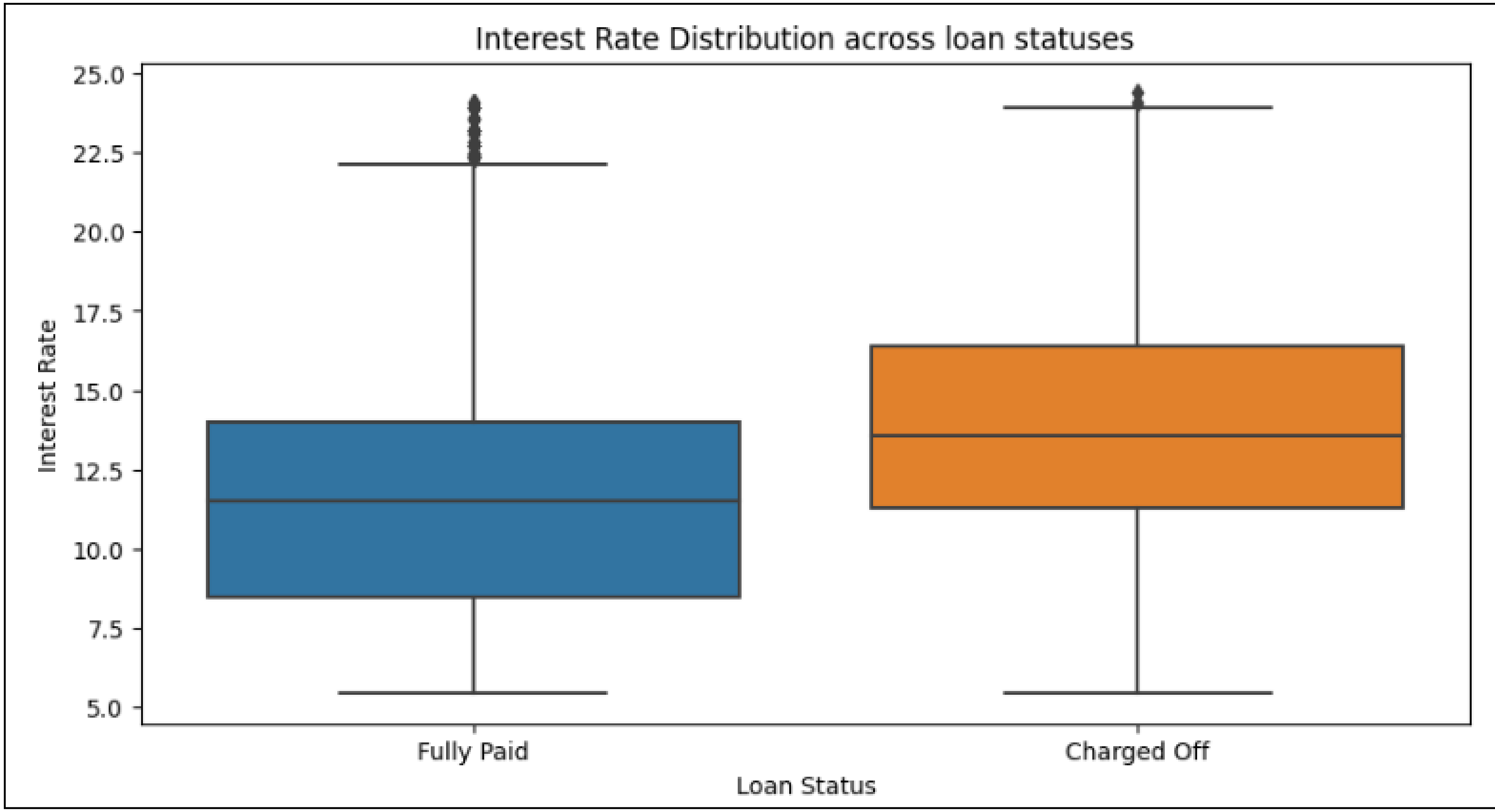
Segmented Univariate Analysis



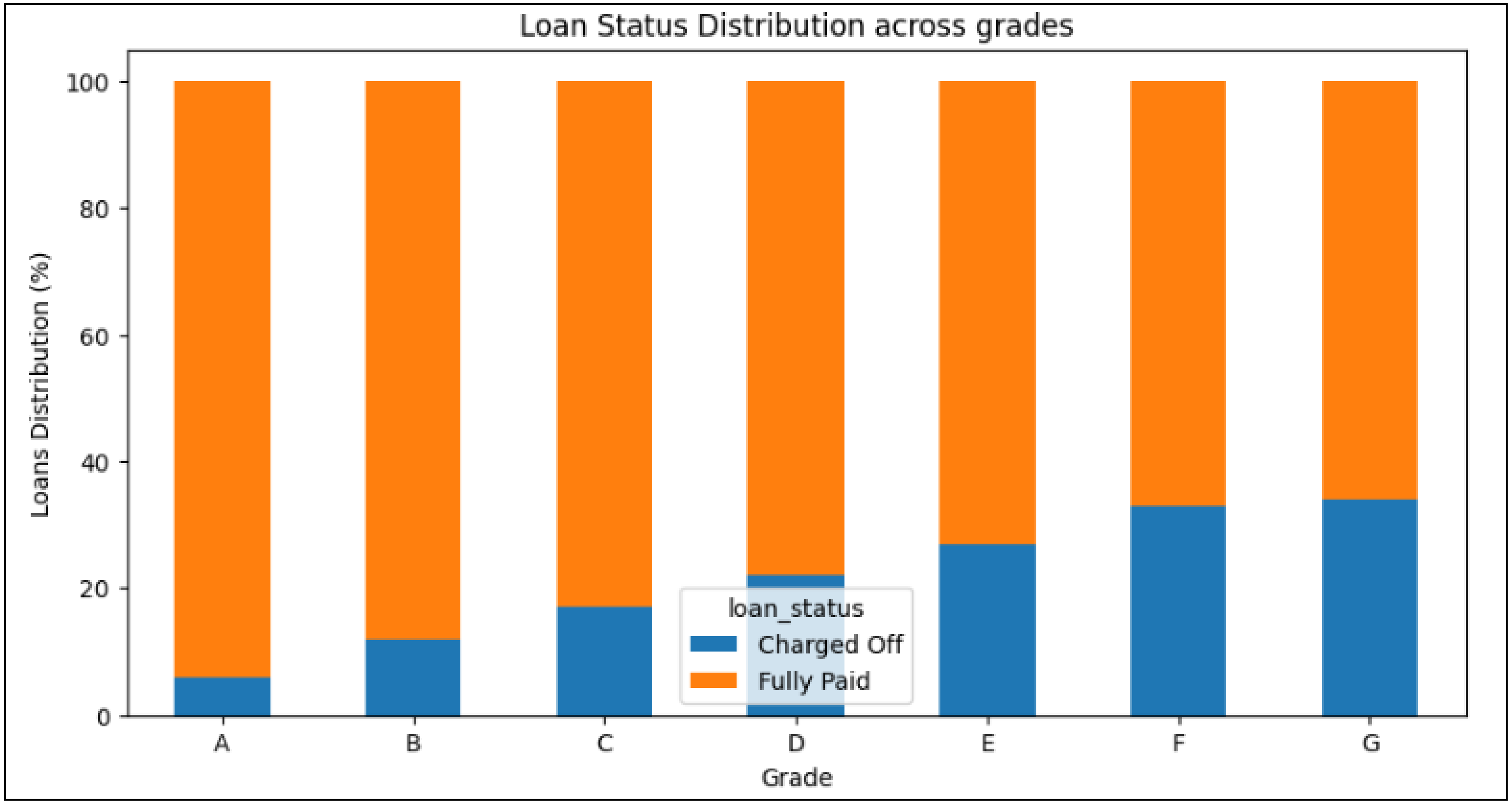
- Loan Applicants with higher Funded Amount tend to default more than that of lower amounts



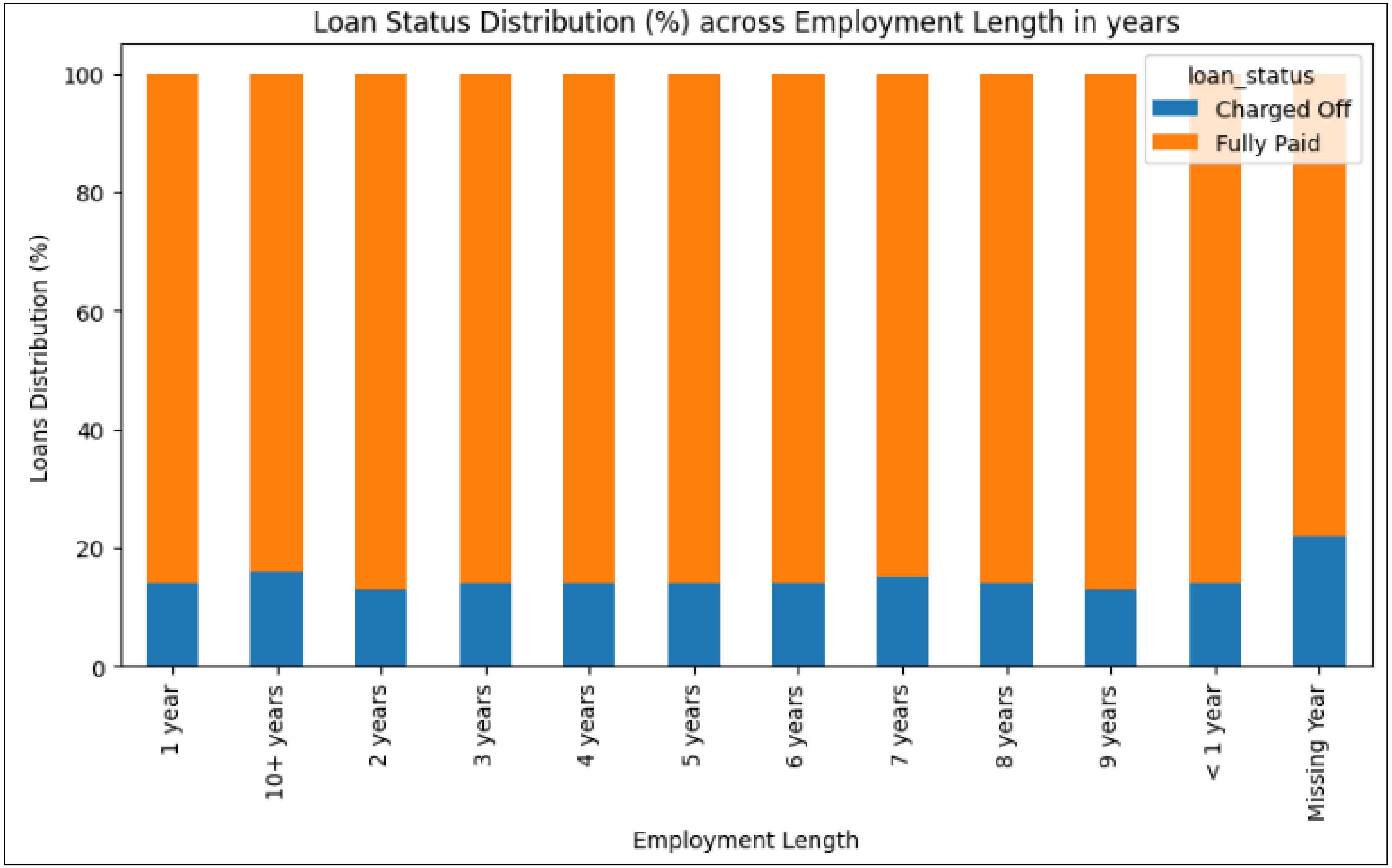
➢ Loan Applicants with the term 60 months are more likely to default than that of 36 months



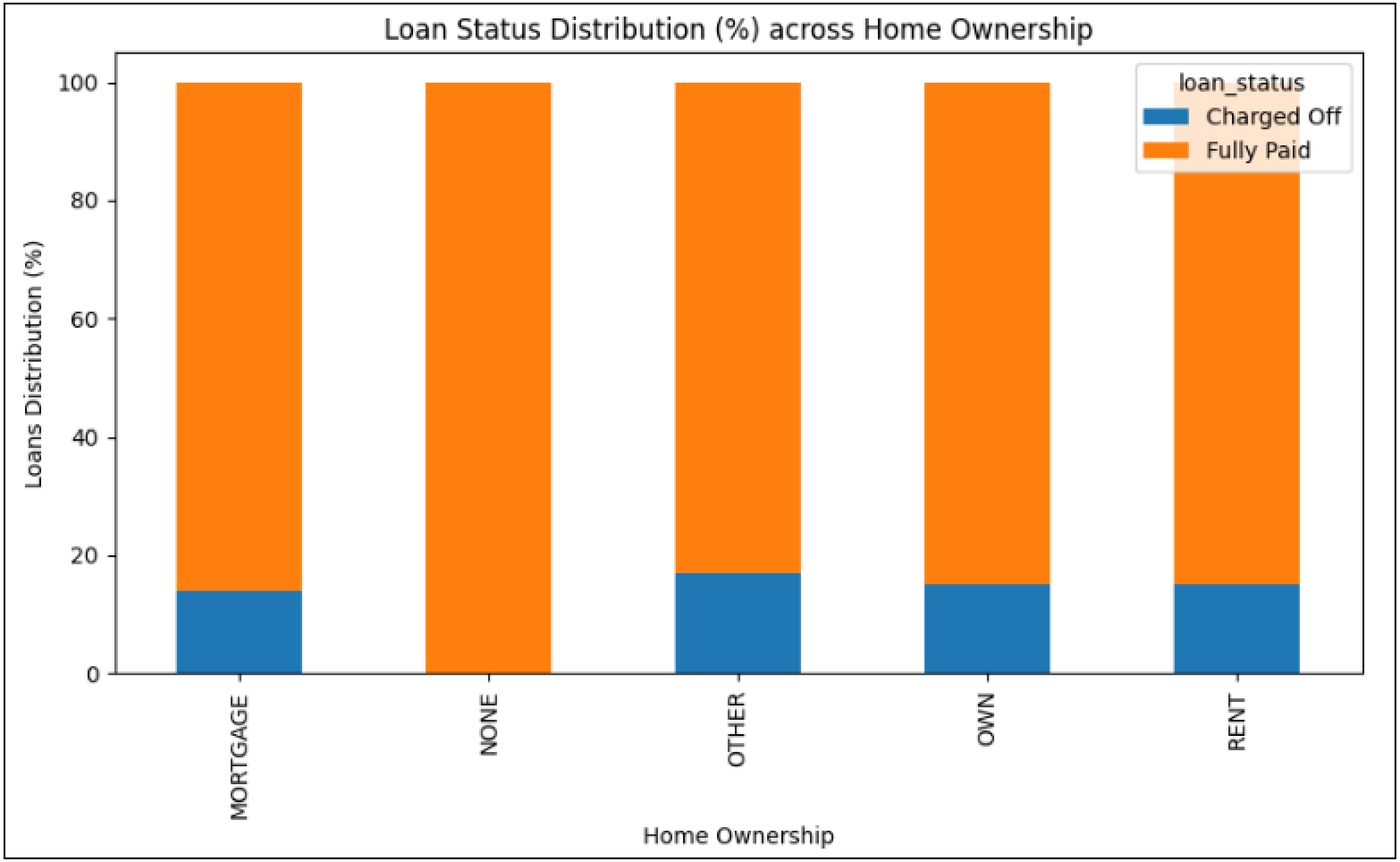
- Loan Applicants who got loan for higher interest rates are more likely to default than that of lower interest rates



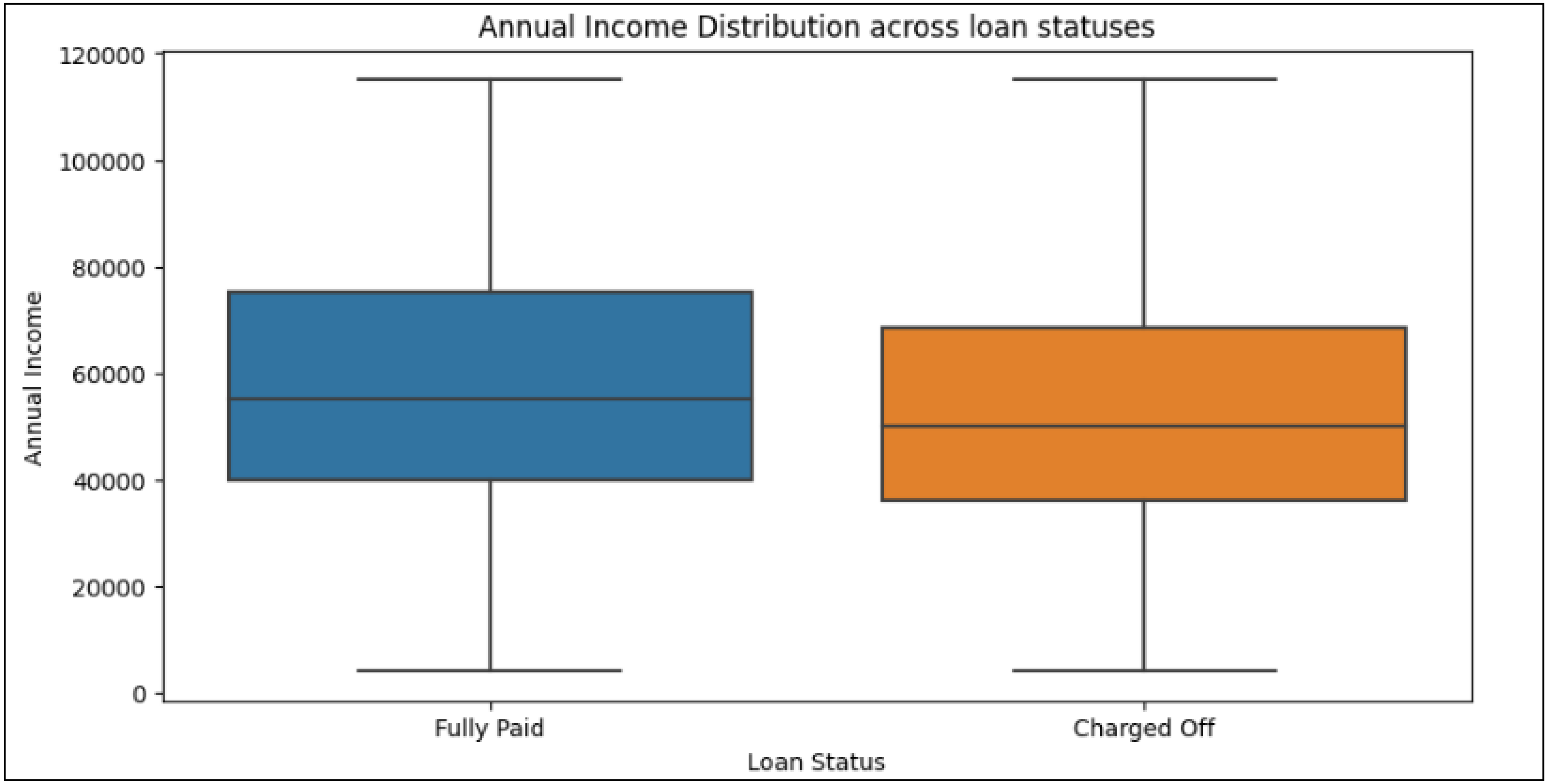
➢ Loan Applicants with grades E, F, G are more likely to default than that of others. This correlates with interest rates analysis since E, F, G grades have higher interest rates



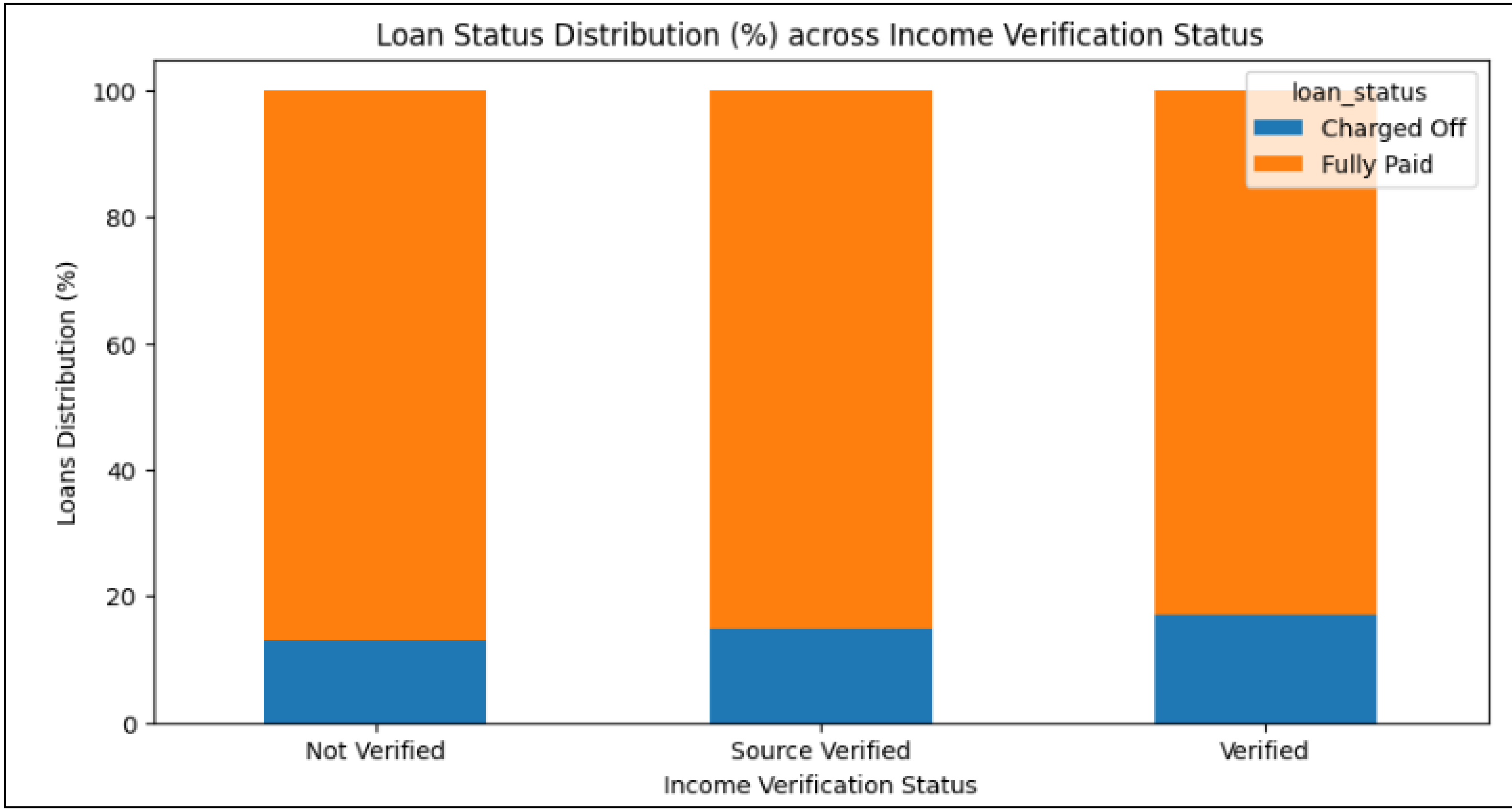
- Loan Applicants with employment length of 10+ years and NA tend to default more than others



➢ Applicants with Home Ownership OTHER tend to default more than others. Interestingly, Home Ownership = None has no defaulters.

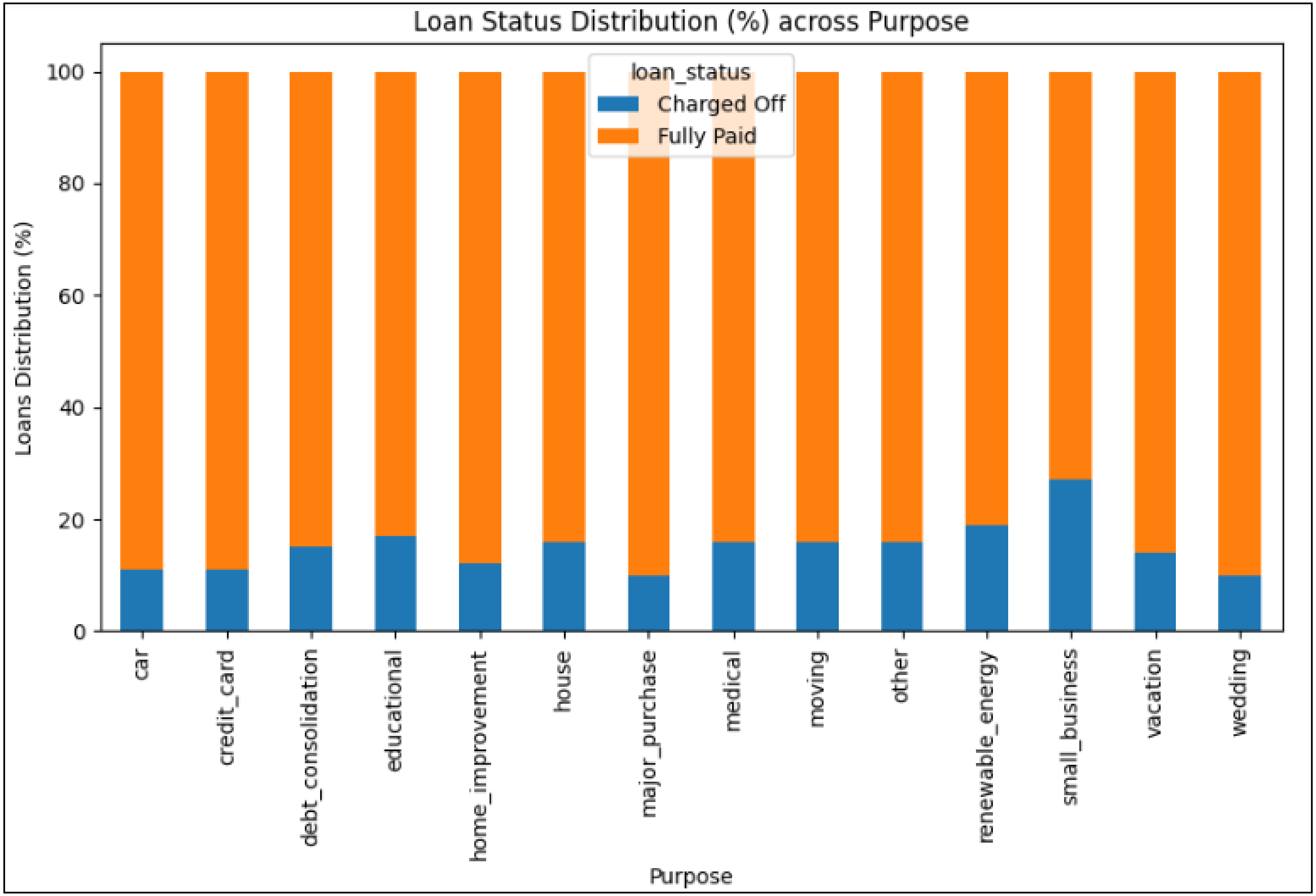


➢ Loan Applicants who have lesser Annual Income are likely to default than that of higher annual income

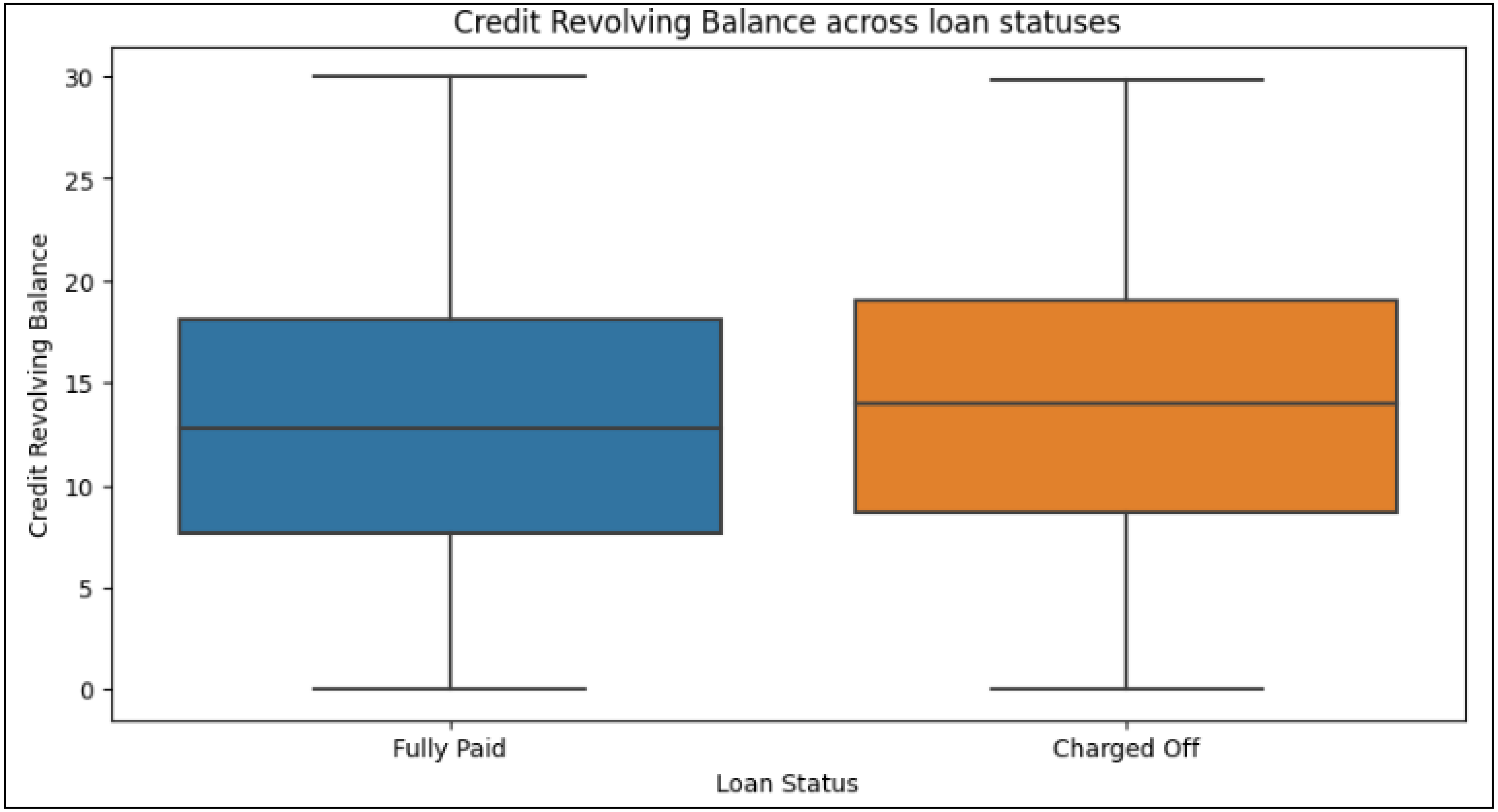


➢ Surprisingly, Loan Applications whose Income is 'Verified', defaults more than others by at least 2%.

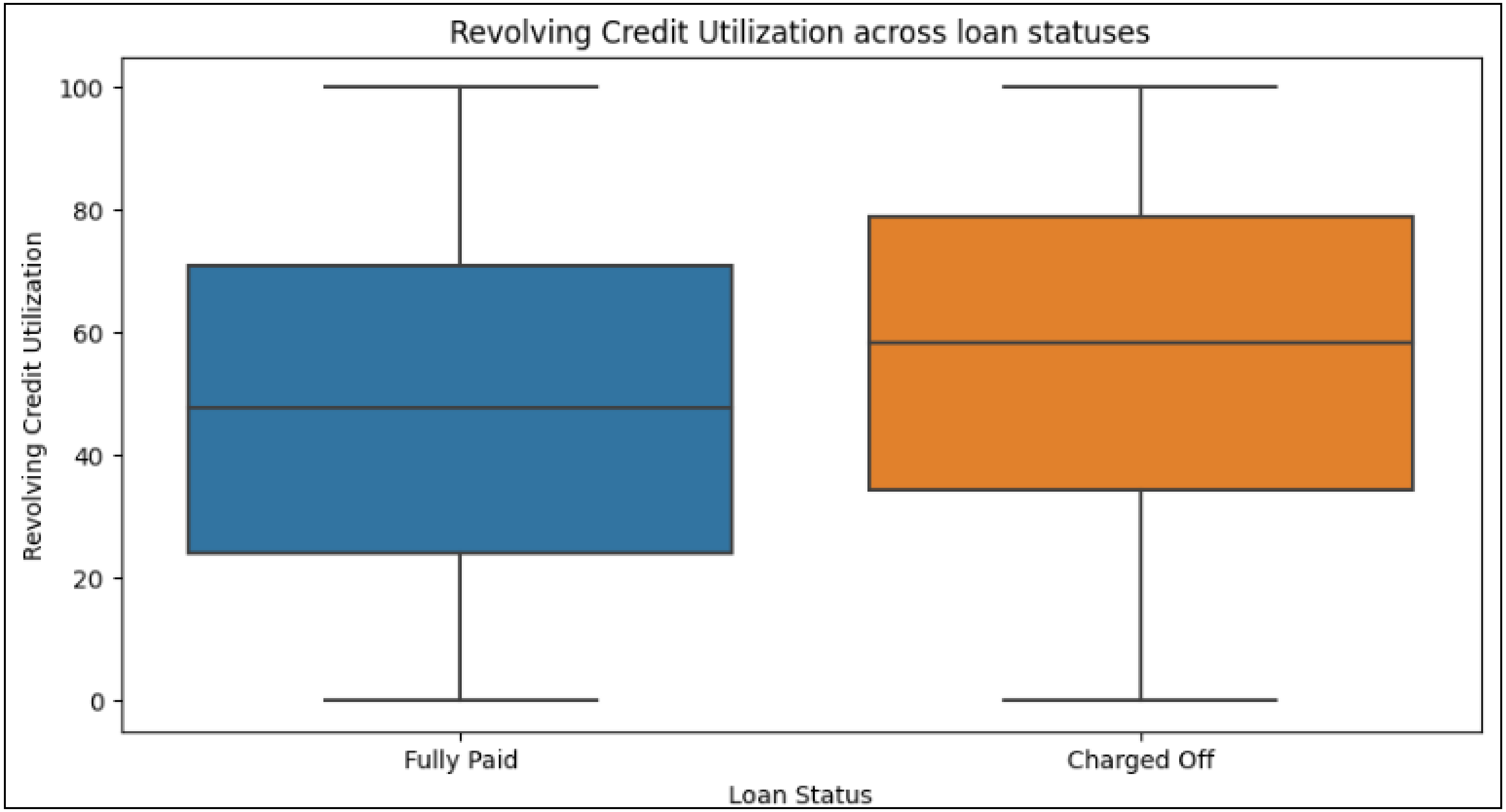
It seems that verification process needs to be revisited



➢ Loan Applicants for the purpose of Small Business tend to default more than other purposes by a huge difference.

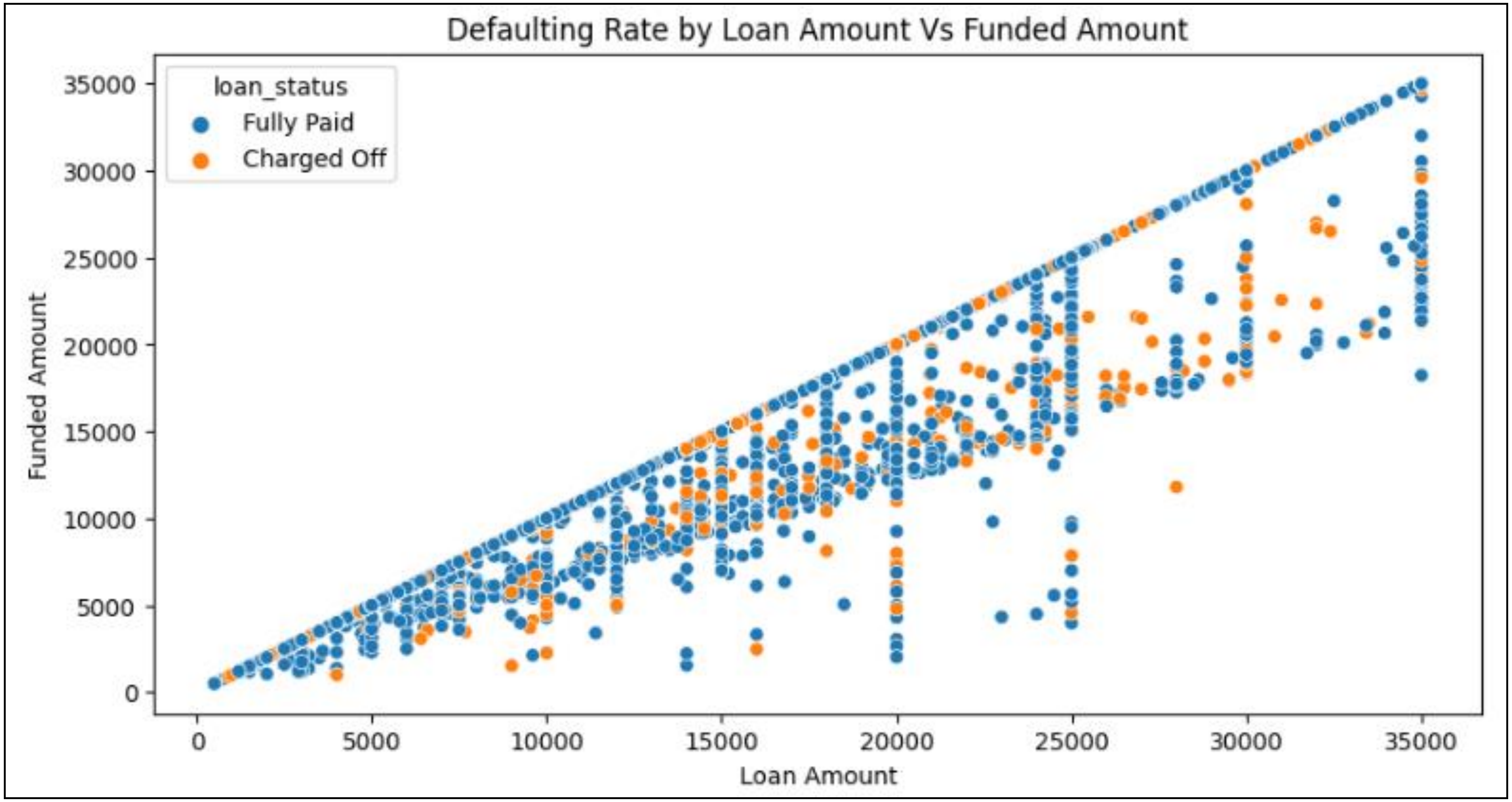


- Loan Applications with higher credit revolving balance are more likely to default than the others. The difference is not striking though.

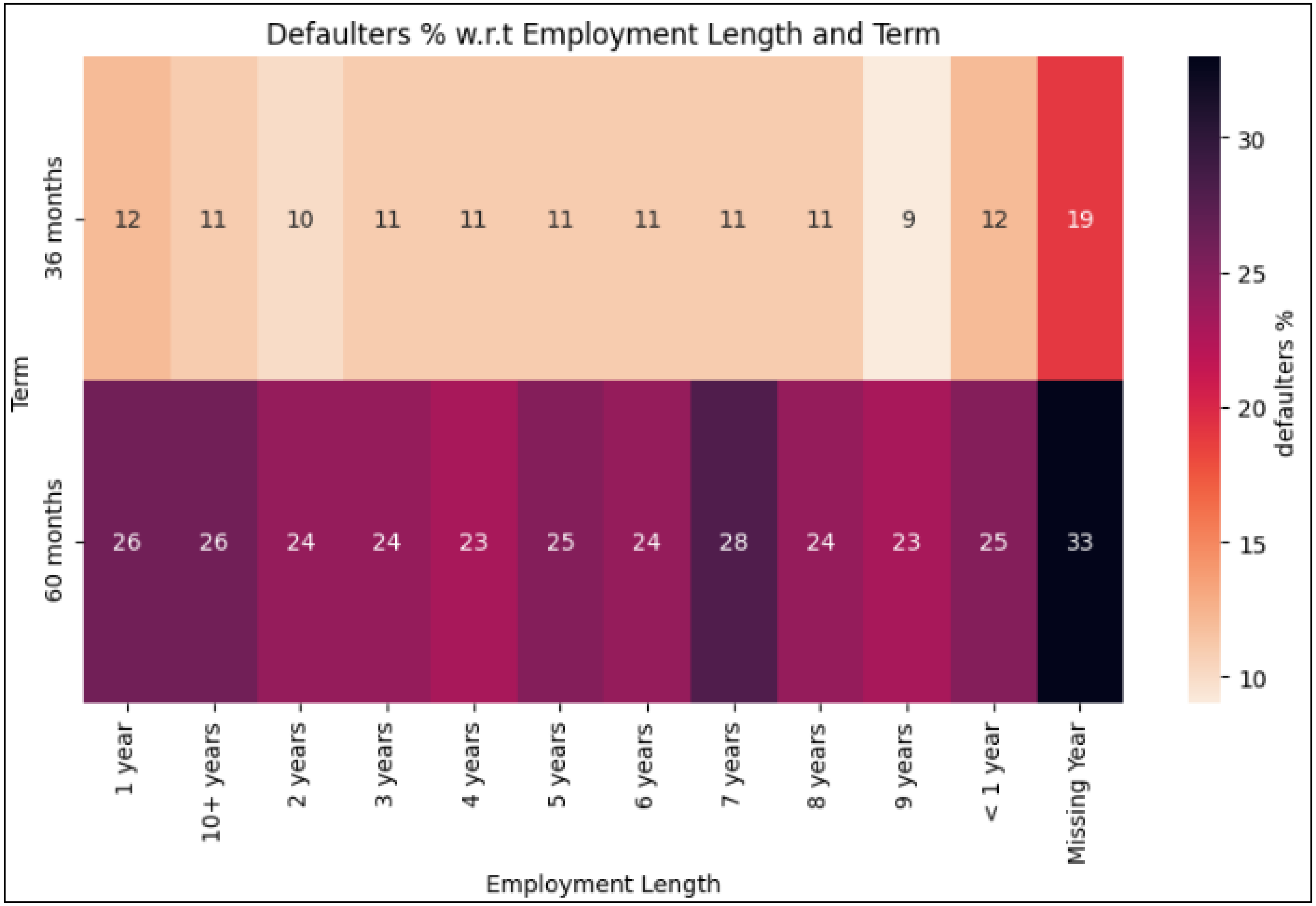


➢ Loan Applications with higher revolving credit utilization tend to default the loan more than others

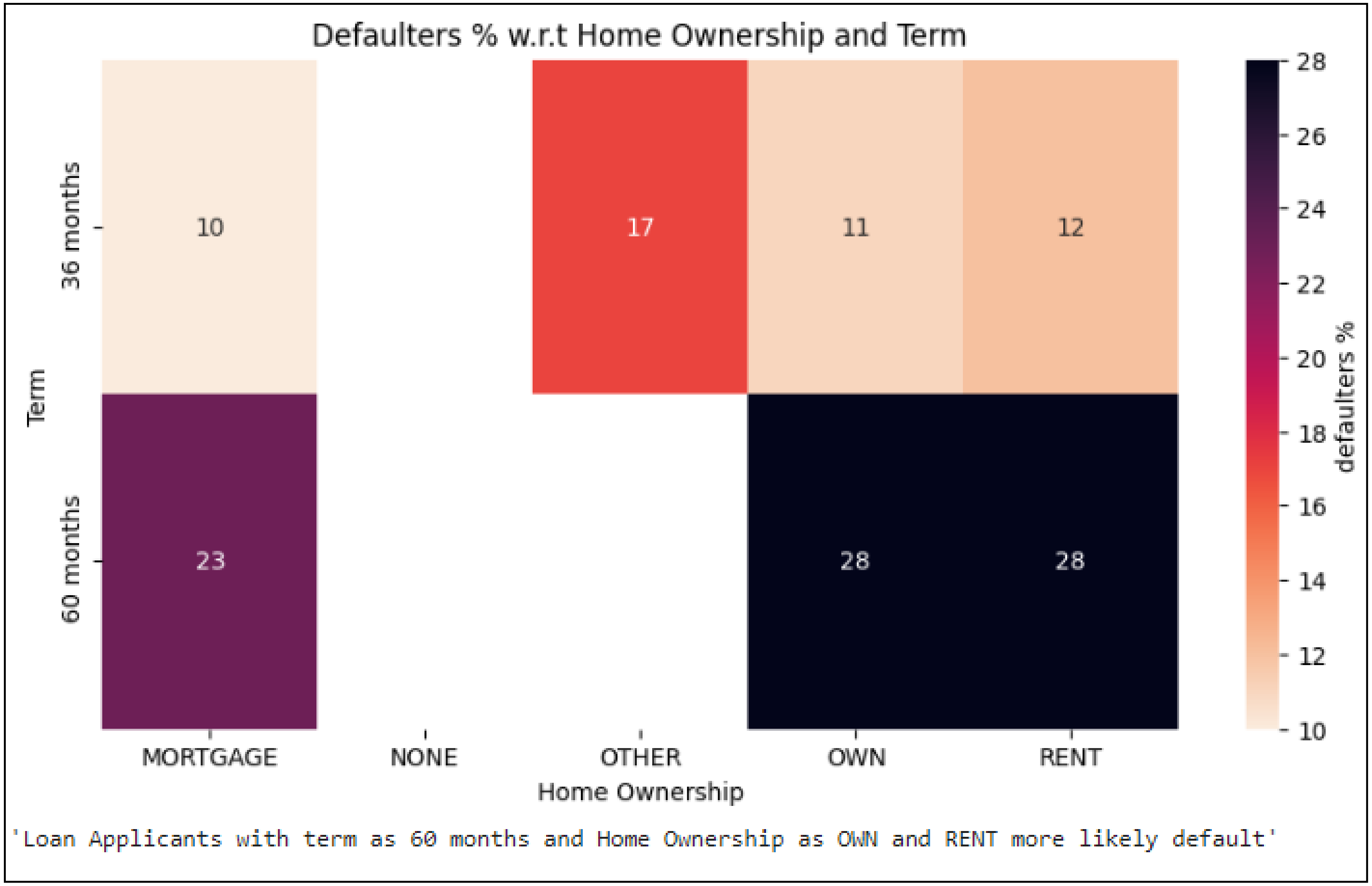
Bivariate Analysis



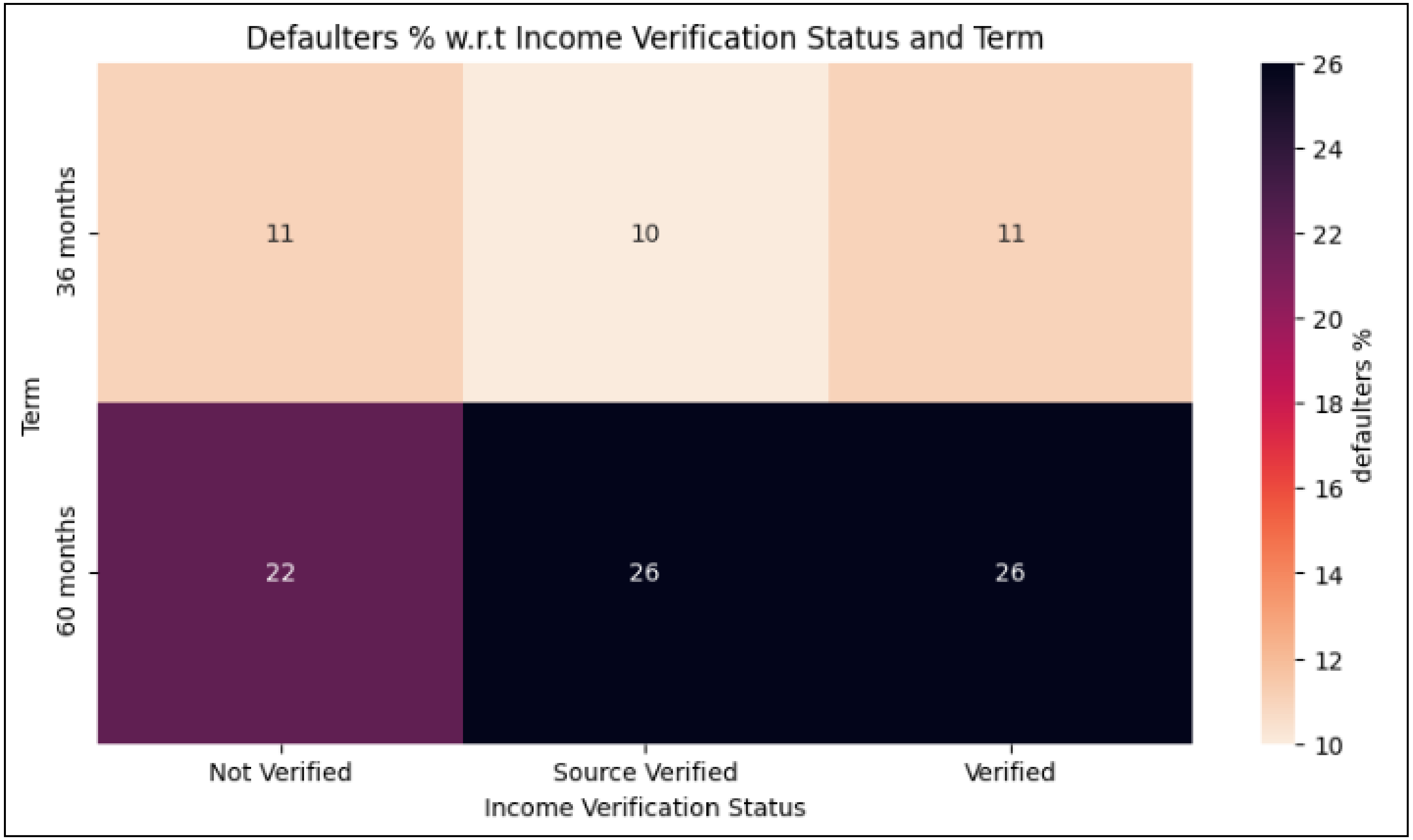
➢ There are less defaulters when Funded Amount is same as the Loan Amount (linear relation)



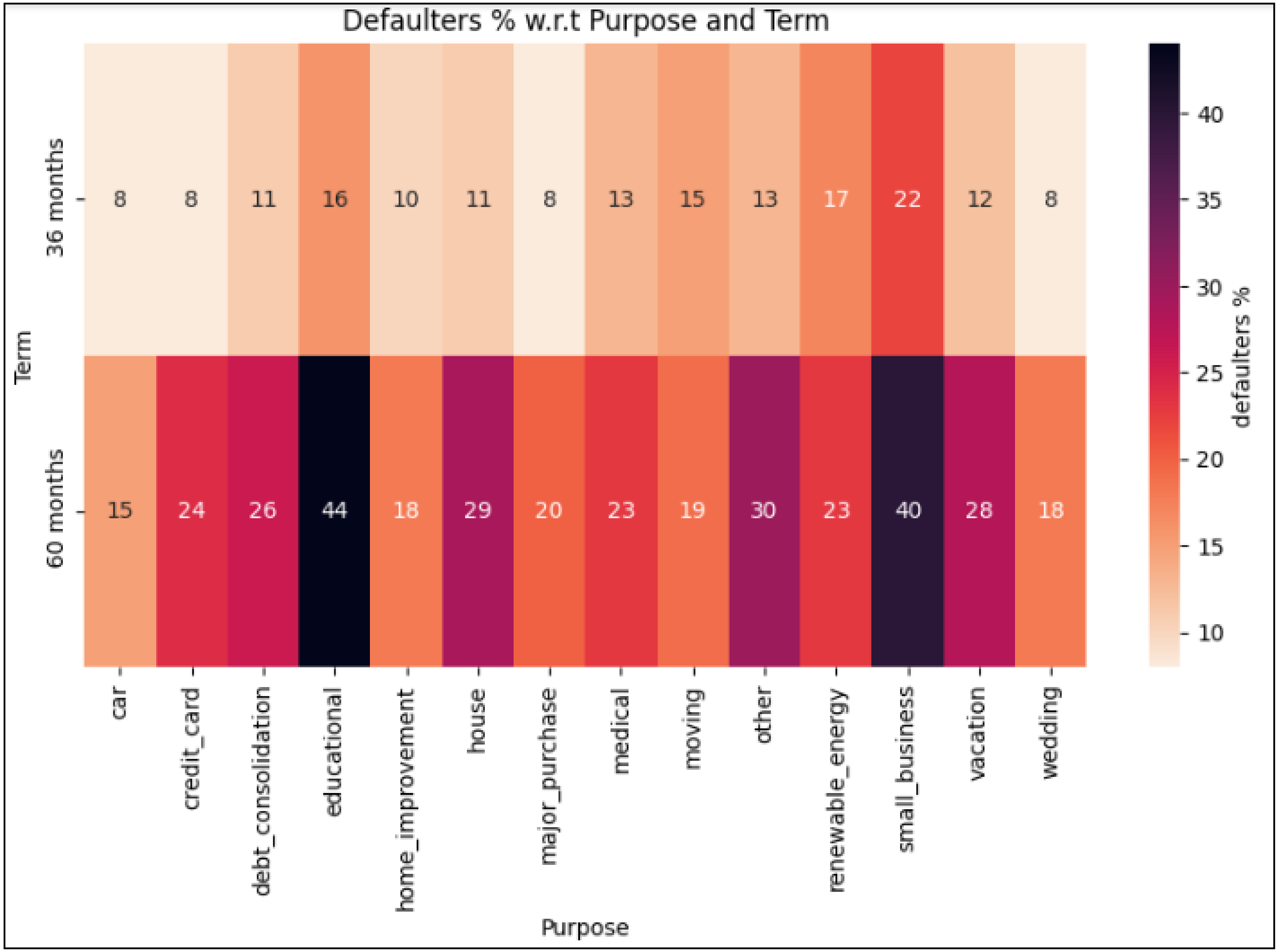
➢ Loan Applicants with employment length of 7 years and NA values tend to default more when the term of the loan is 60 months



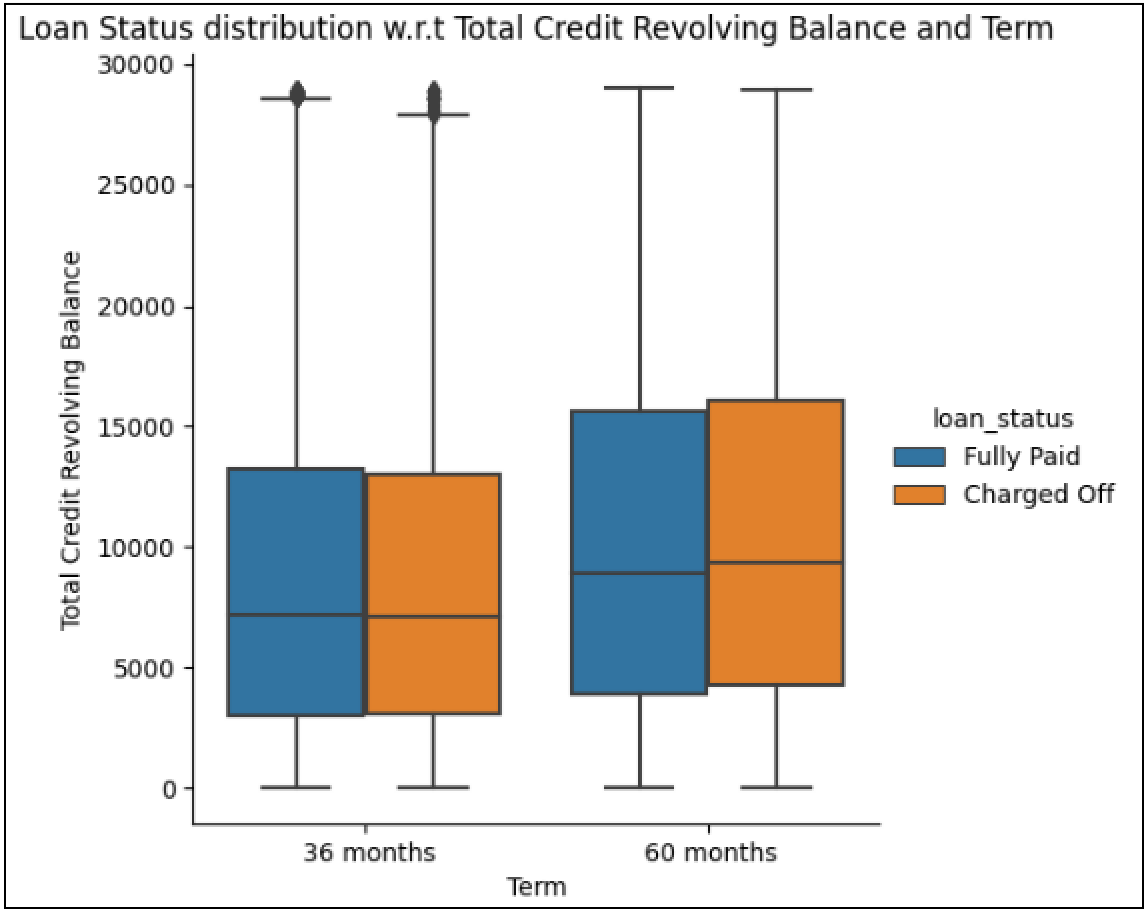
- Loan Applicants with term as 60 months and Home Ownership as OWN and RENT more likely default



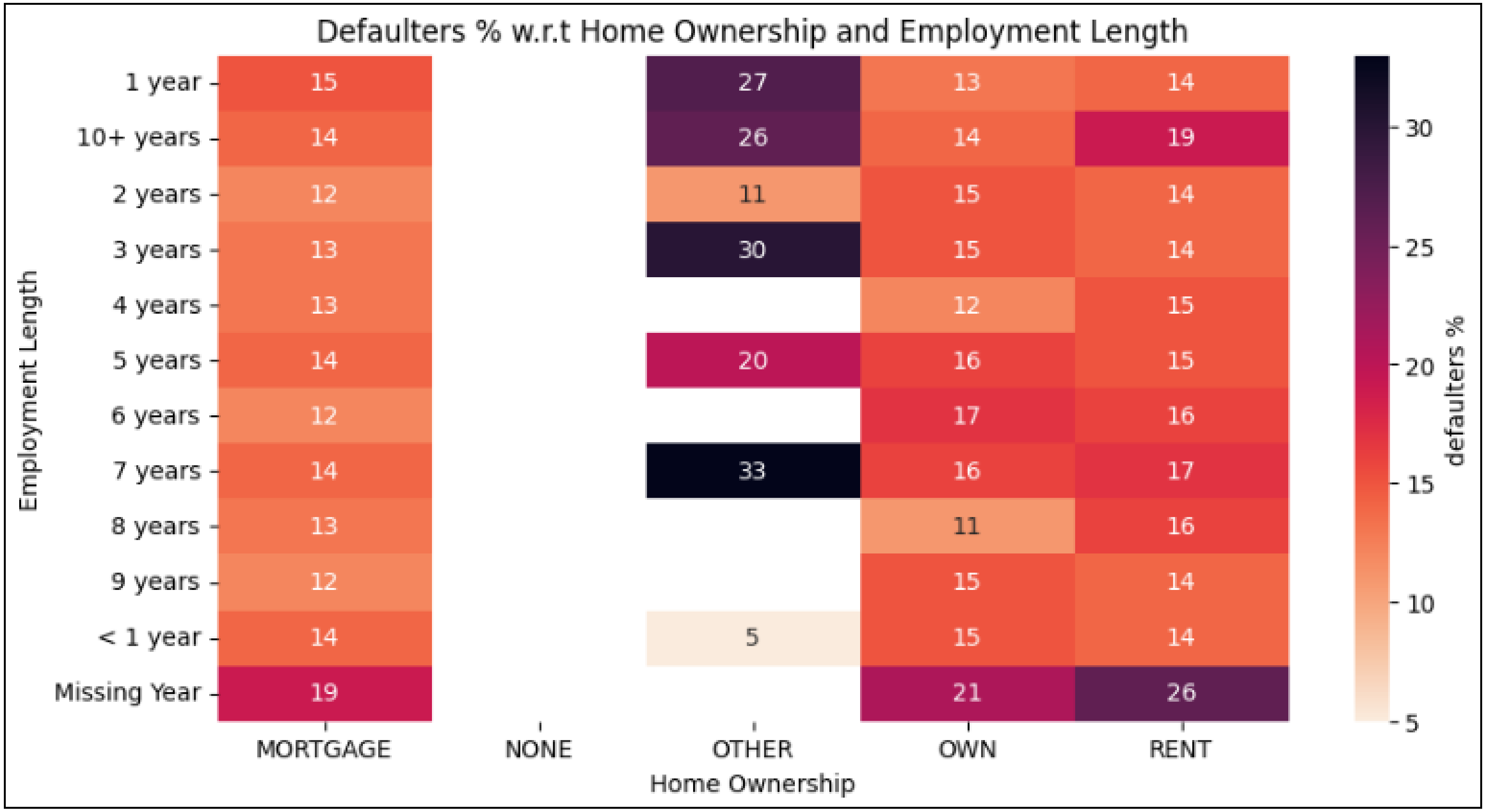
➢ Loan Applications with term as 60 months, their income 'Verified' and 'Source Verified' tend to default more likely than others.



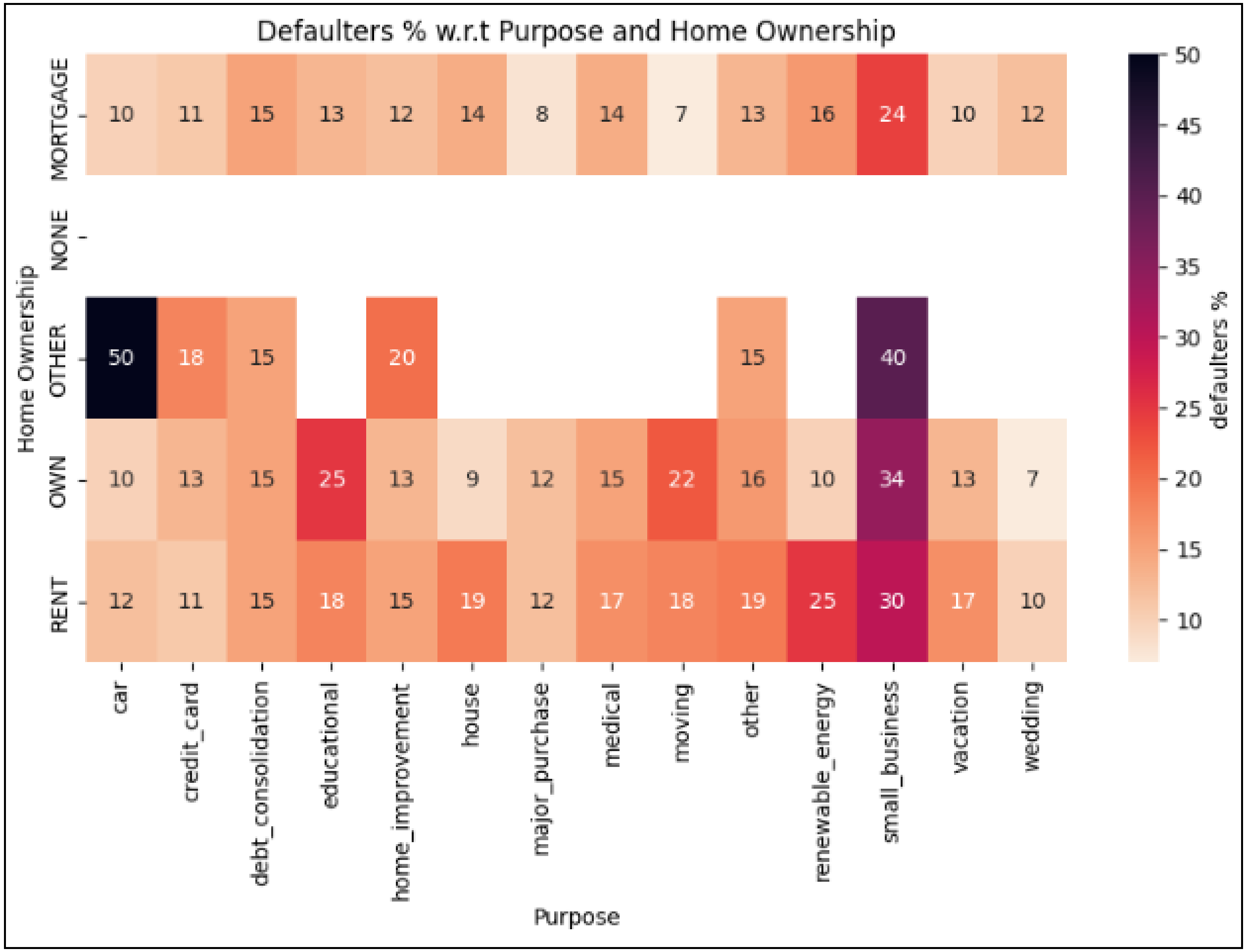
➢ Loan Applicants with term as 60 months are more likely to default when the purpose is educational and small business



➢ Loan Applicants with higher Total Credit Revolving Balance tend to default likely when the term is 60 months



➢ Loan Applicants with OTHER home ownership tend to default more when their employment length is 10+ years, 7 years or 3 years or 1 yea



➢Loan Applicants for the purpose of small business tend to default more across RENT, OWN and OTHER home ownership types. Also, Car - OTHER combination seem to default likely

Conclusions

- Applicants who opted for loans with higher term (60 months) tend to default more than others.

- Applicants who got loan with higher interest rate tend to default more than others.

- Applicants with higher salary mostly have applied loans for "home_improvment", "house", "renewable_energy" and "small_businesses“.

- Applicants applying for long term loans have applied for more loan amount.