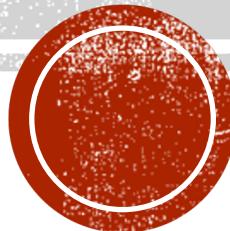


LENDING CLUB CASE STUDY

Group Members:

1. Raina Sharma
2. Rahul Madan





understand the **driving factors** (or **driver variables**) behind loan default, i.e. the variables which are strong indicators of default.



identify risky loan applicants using EDA. The company can utilise this knowledge for its portfolio and risk assessment.

CASE STUDY OBJECTIVES



DATA CLEANING STEPS

1

Fix rows
and
columns

2

Fix missing
values

3

Standardise
values

4

Fix invalid
values

5

Filter data



PRE PREPARATORY STEPS

Removed columns having null values >40%

Dropped columns not required in analysis

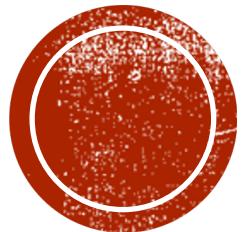
Handled Null values

Corrected data types of columns

Created Derived Columns required for the analysis

Found outliers and removed





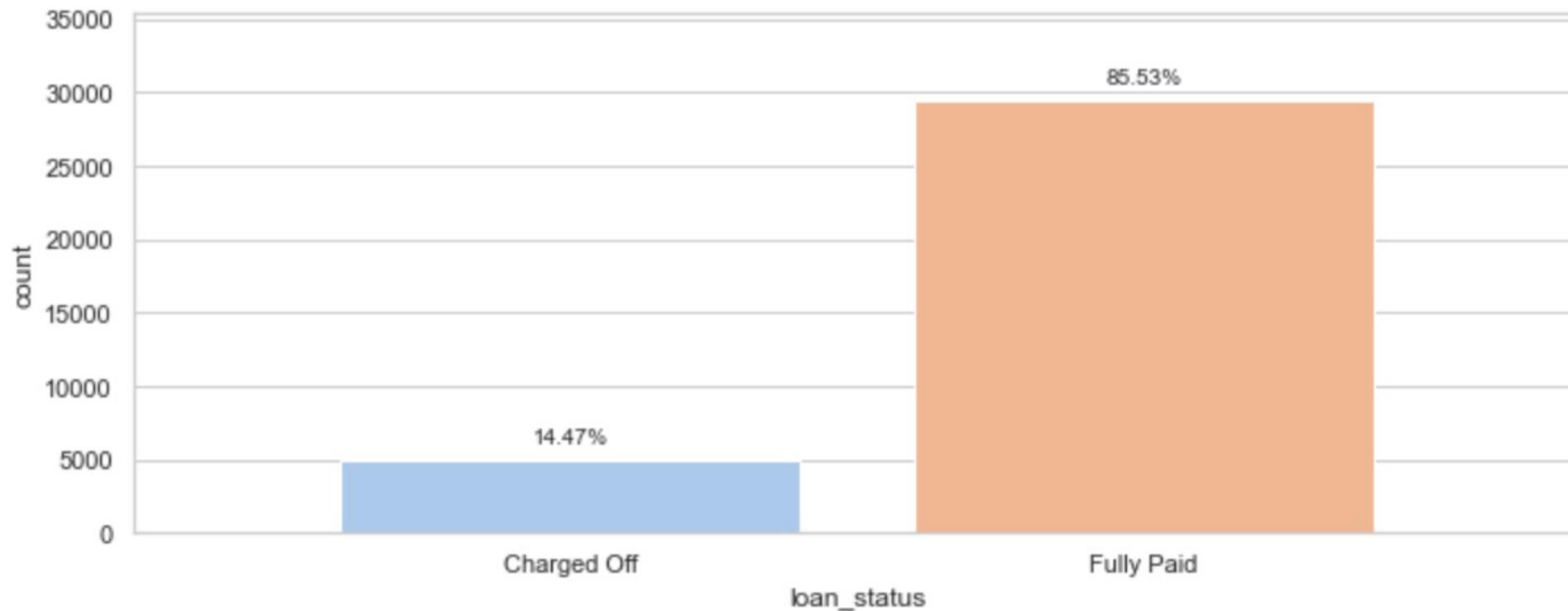
UNIVARIATE ANALYSIS



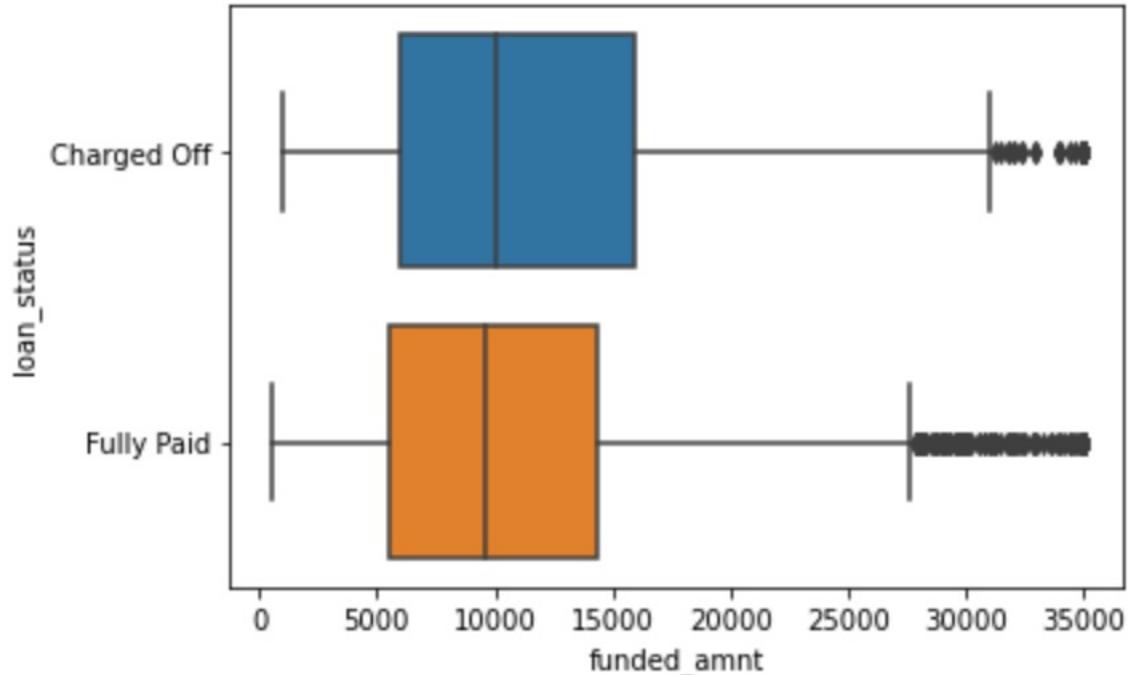
LOAN STATUS

Observation:

- 85% of the loans are successfully paid and only 15% of the loans are defaulted



FUNDED AMOUNT

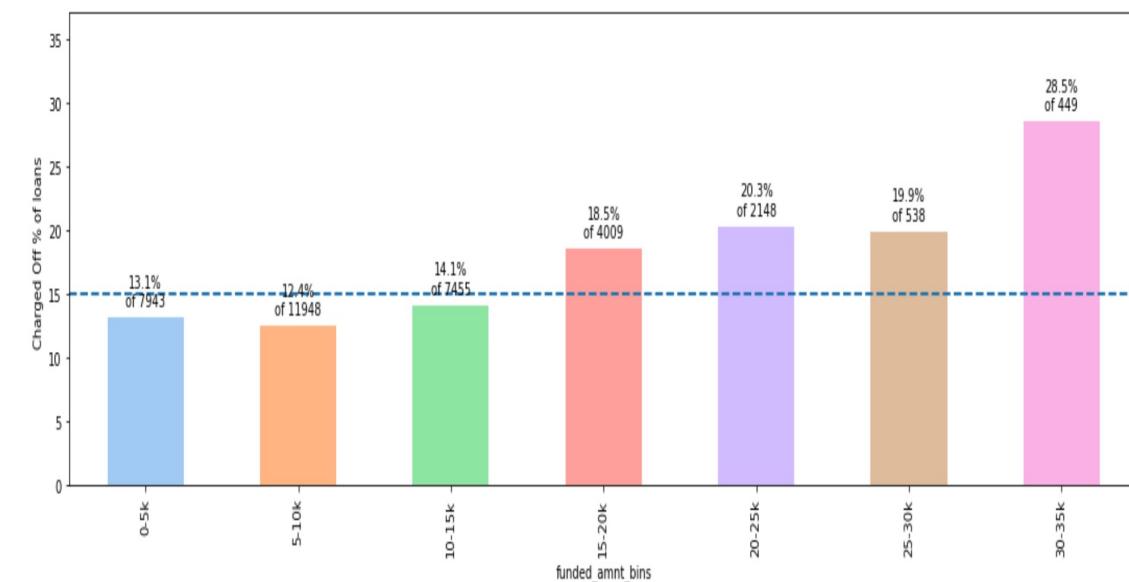
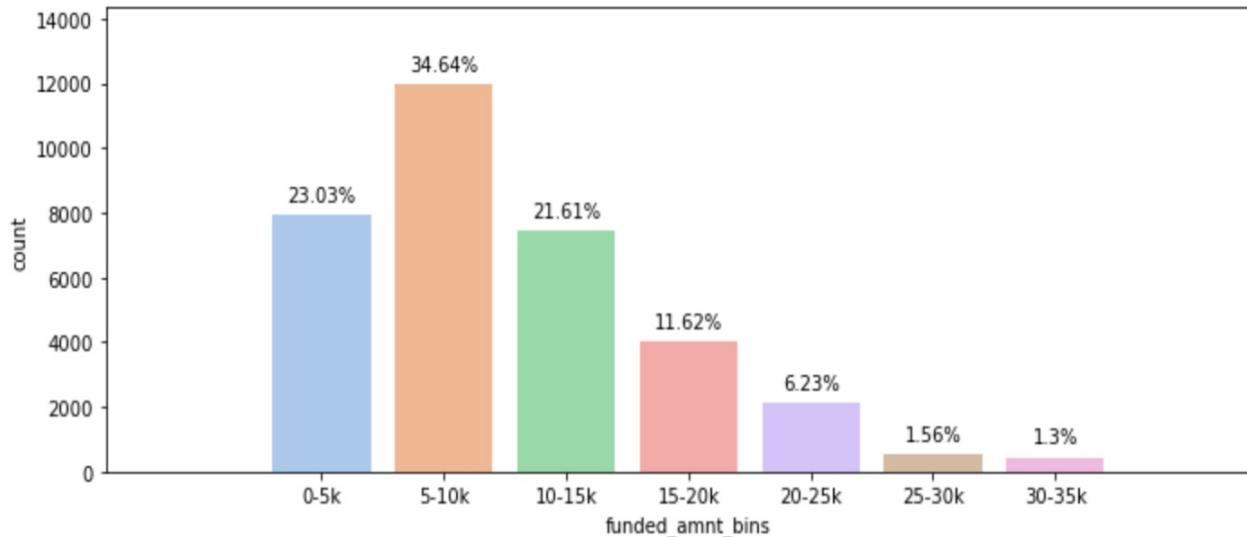


Observation:

- ❖ Overall, the funded amount distribution is slightly right-skewed with mean greater than the median. Most of the loans granted are below 16000 (75 percentile value)
- ❖ Charged off loans have higher average of funded amount than the paid off loans.



FUNDED AMOUNT

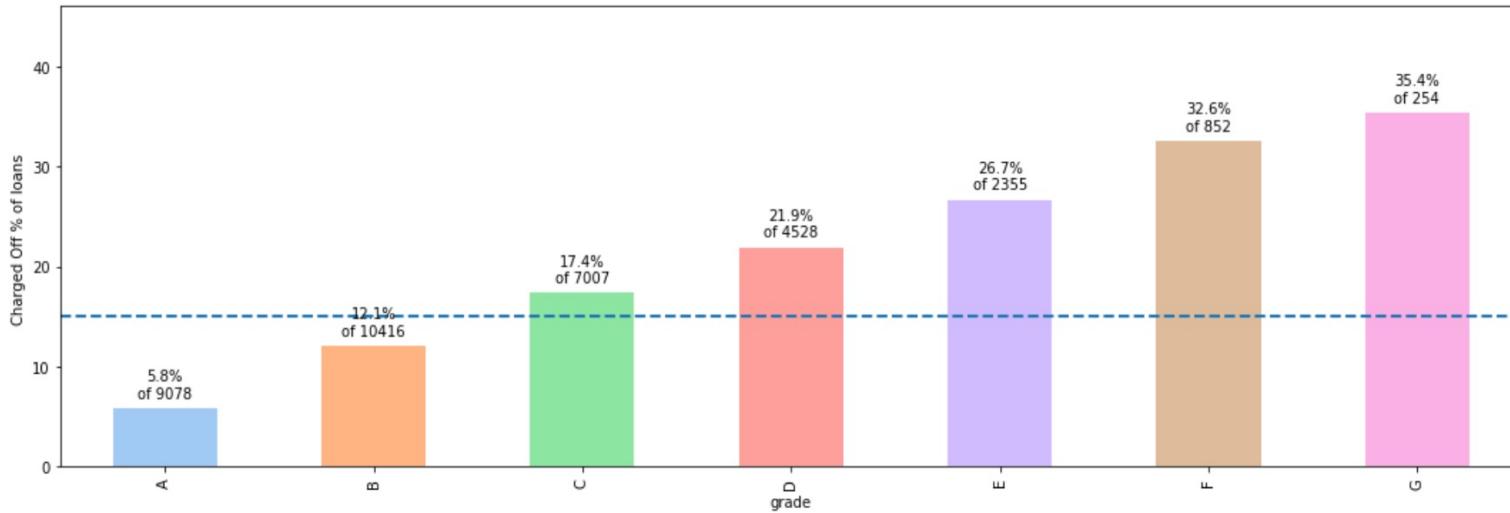


Observation:

- ❖ The % of charged off loans increases substantially with increase in funded amount

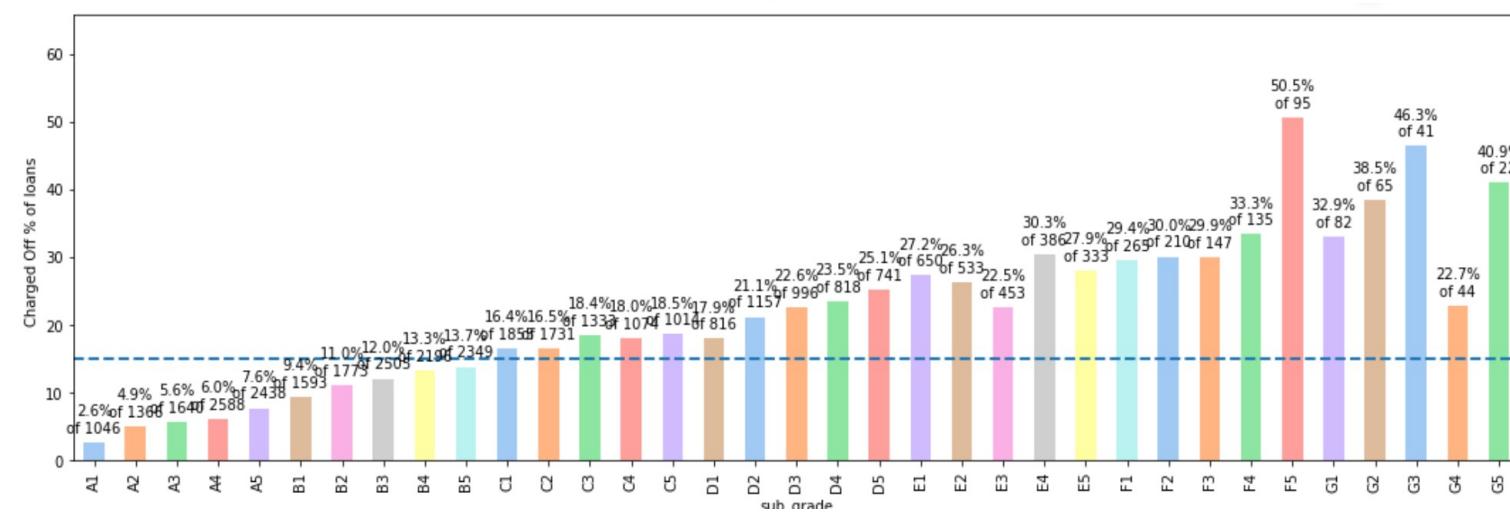


GRADE & SUB GRADE

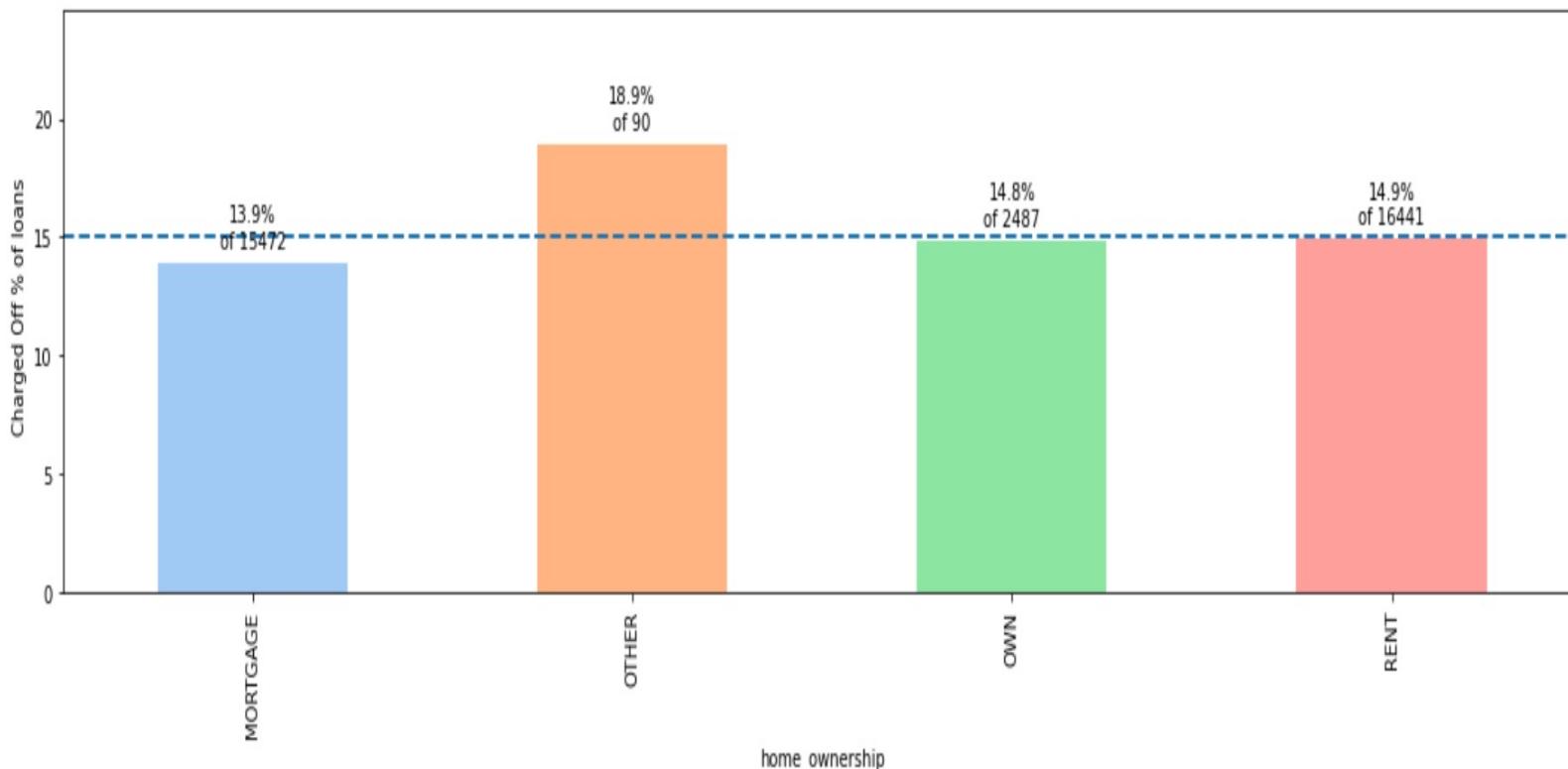


Observation:

- ❖ Chances of defaulting increases with grade from A to G and subgrade from A1-A5, B1-B5, & so forth



HOME OWNERSHIP

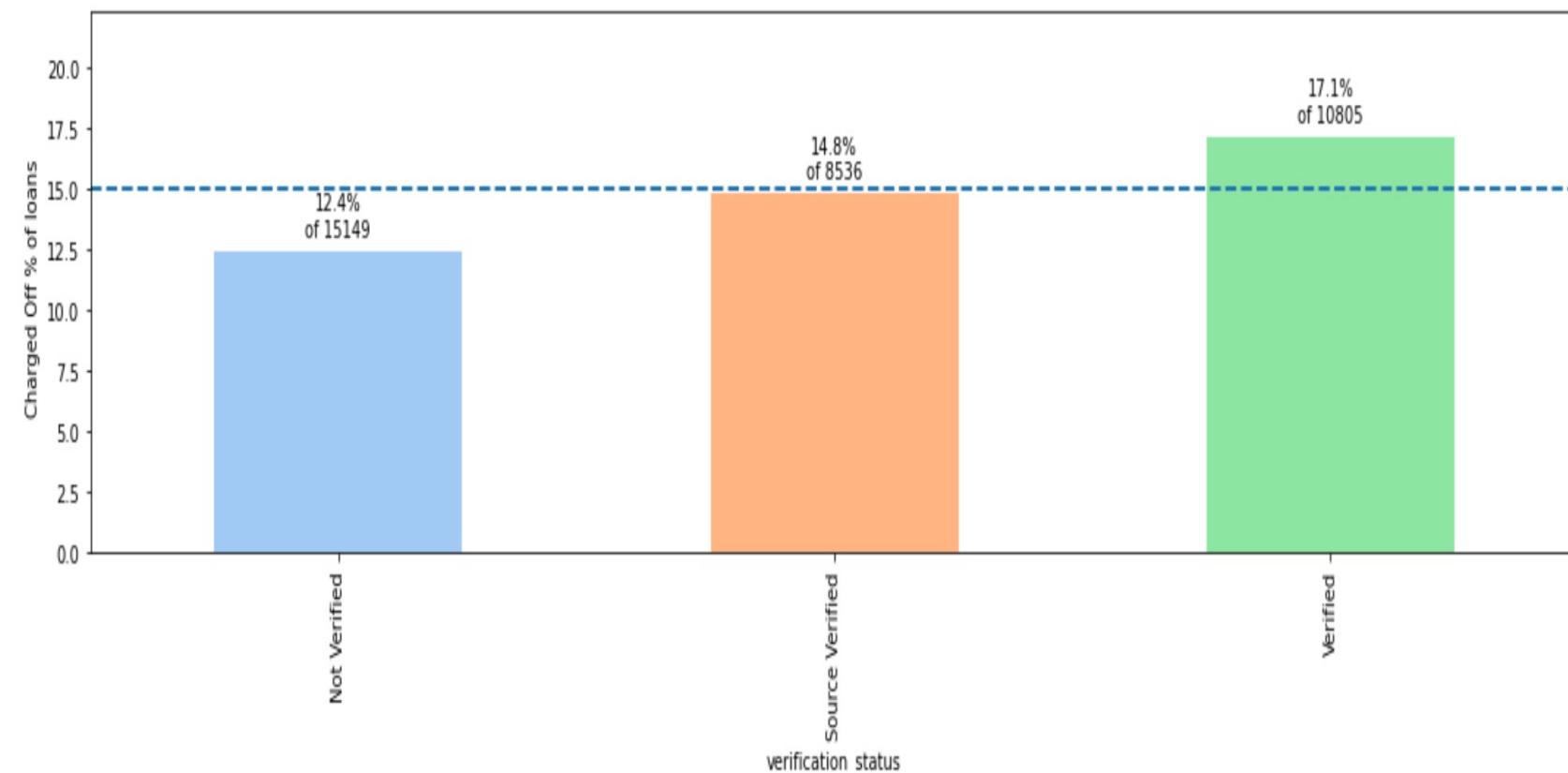


Observation:

- ❖ People with ownership type as Mortgage and Rent are the most who take loans.
- ❖ The default rate is high for people having home ownership other.



VERIFICATION STATUS

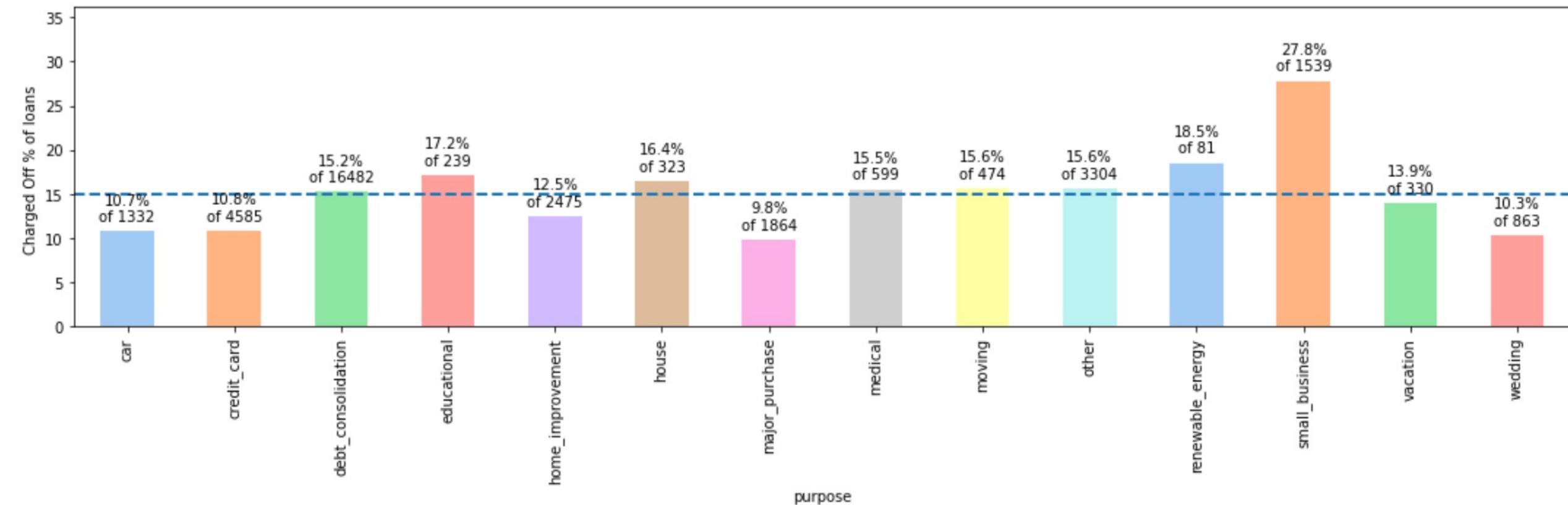


Observation:

- ❖ Majority of people taking loans are not verified.
- ❖ Chances of Defaulting is higher even with the verified verification status.



PURPOSE

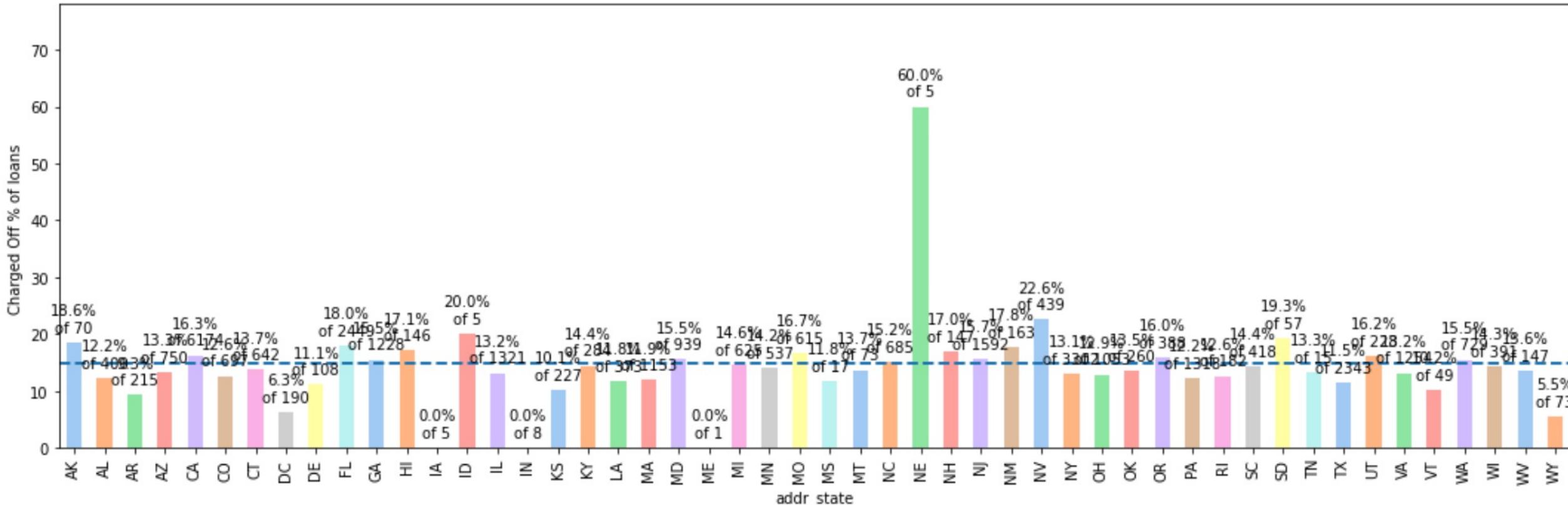


Observation:

- ❖ Majority of people taking loans have credit card or debt consolidation as purpose.
- ❖ Default rate is higher for small business, renewable energy, educational categories.



BORROWER STATE

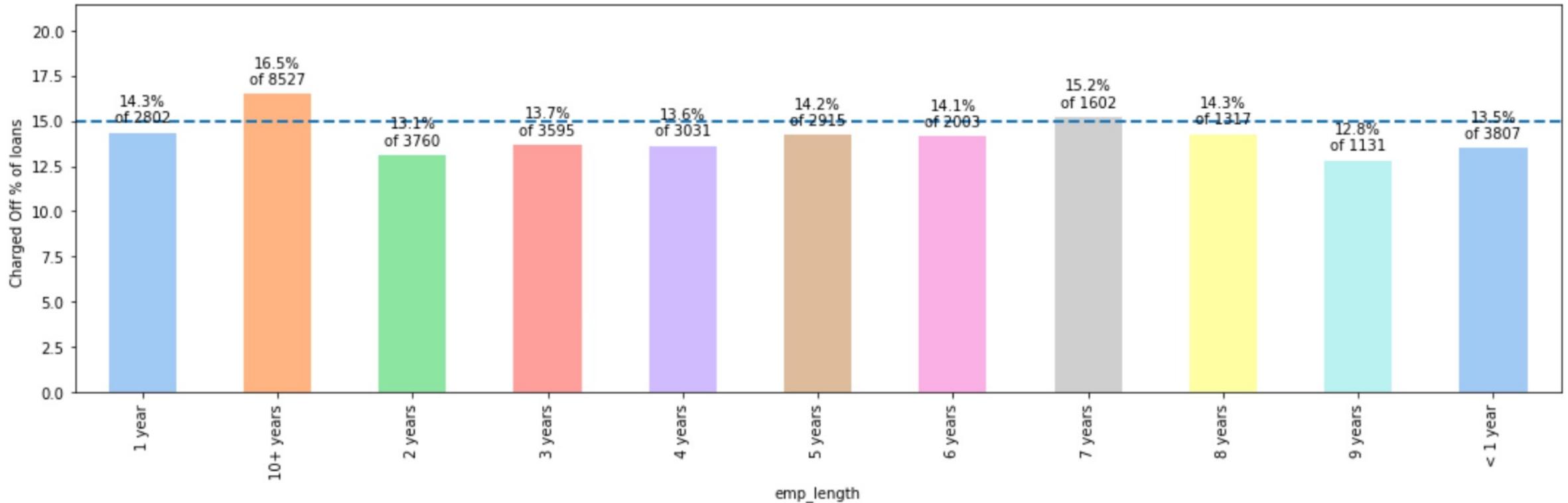


Observation:

- ❖ Most of the borrowers are from California, New York, Florida.
- ❖ Although Nebraska(NE) has high default rate but can be ignored since there are very insignificant number of borrowers. New York has high default rate and with significant % of borrowers.



EMPLOYMENT LENGTH

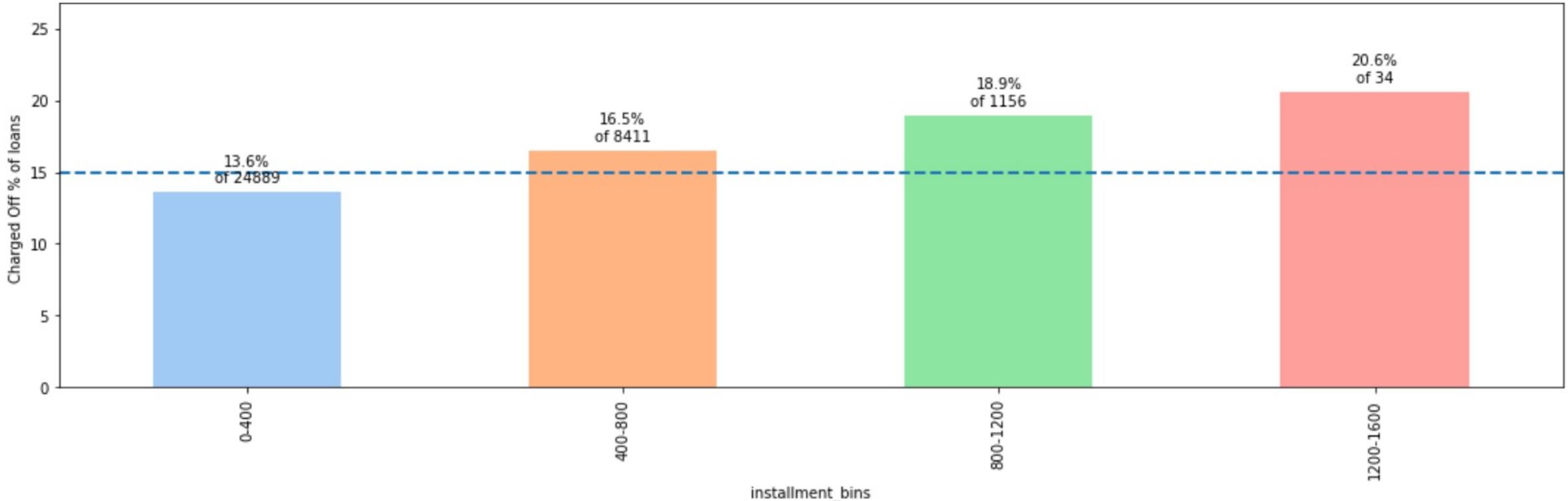


Observation:

- ❖ Most of the borrowers either have experience >10 or are freshers.
- ❖ Chances of defaulting is higher for people with experience > 10 years.



LOAN INSTALLMENT

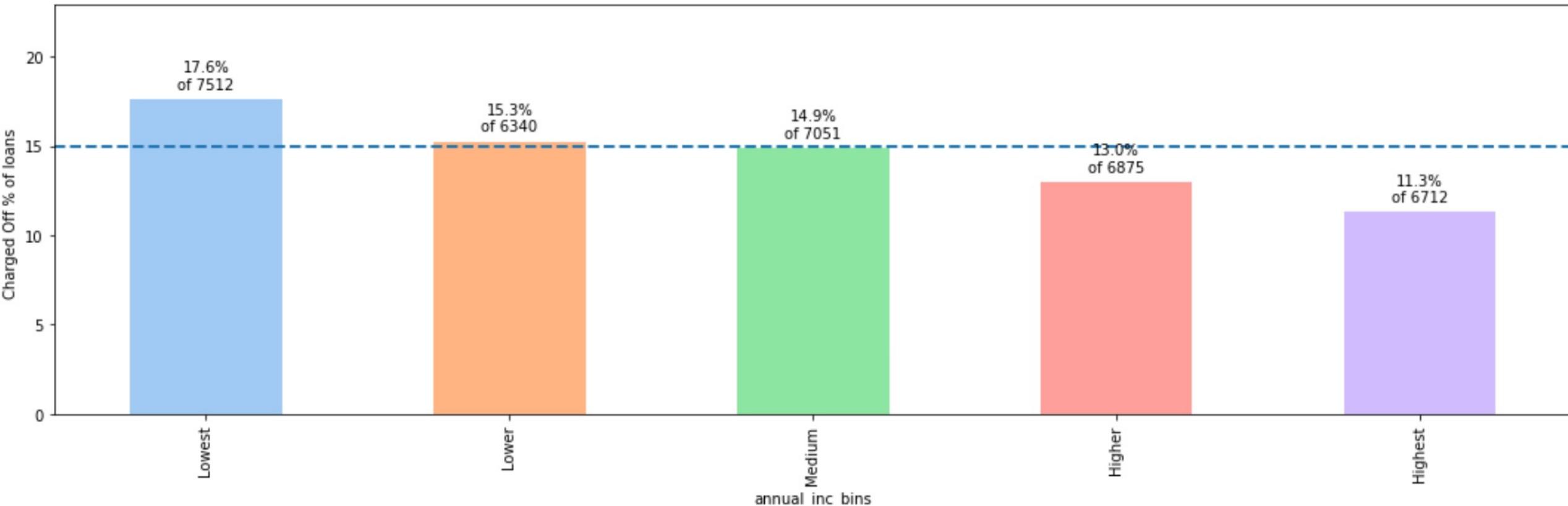


Observation:

- ❖ Chances of defaulting are higher for loans with higher amount of instalment



ANNUAL INCOME

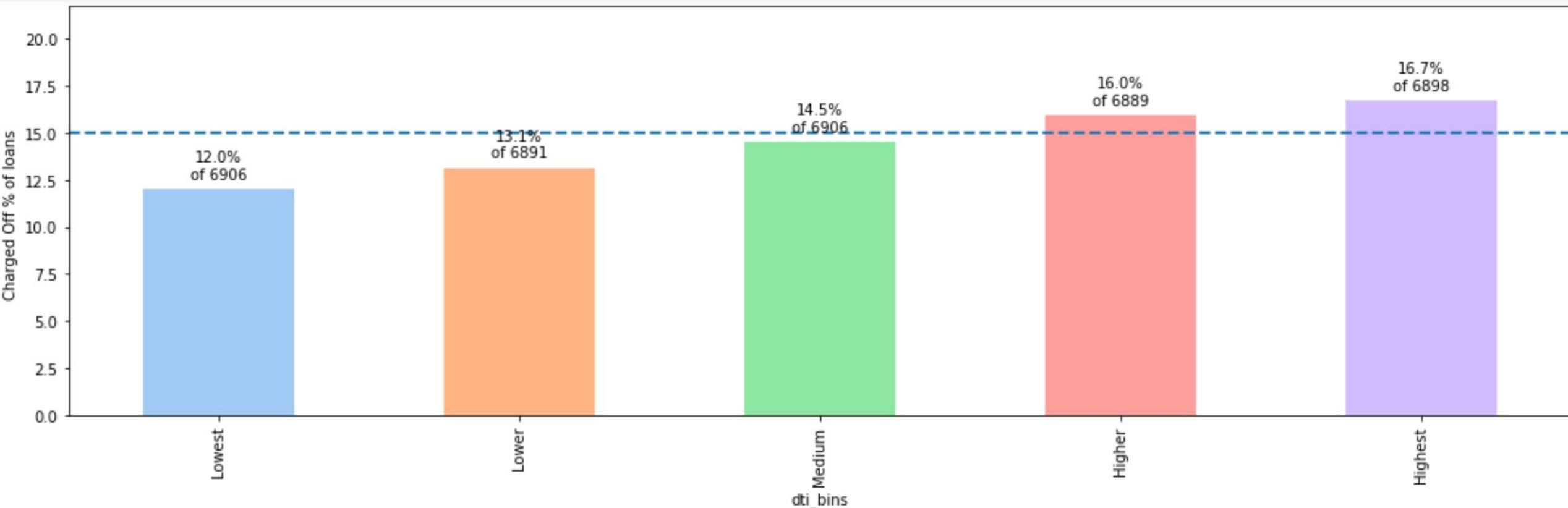


Observation:

- ❖ People with low annual income have high chances of defaulting.



DEBT TO INCOME RATIO

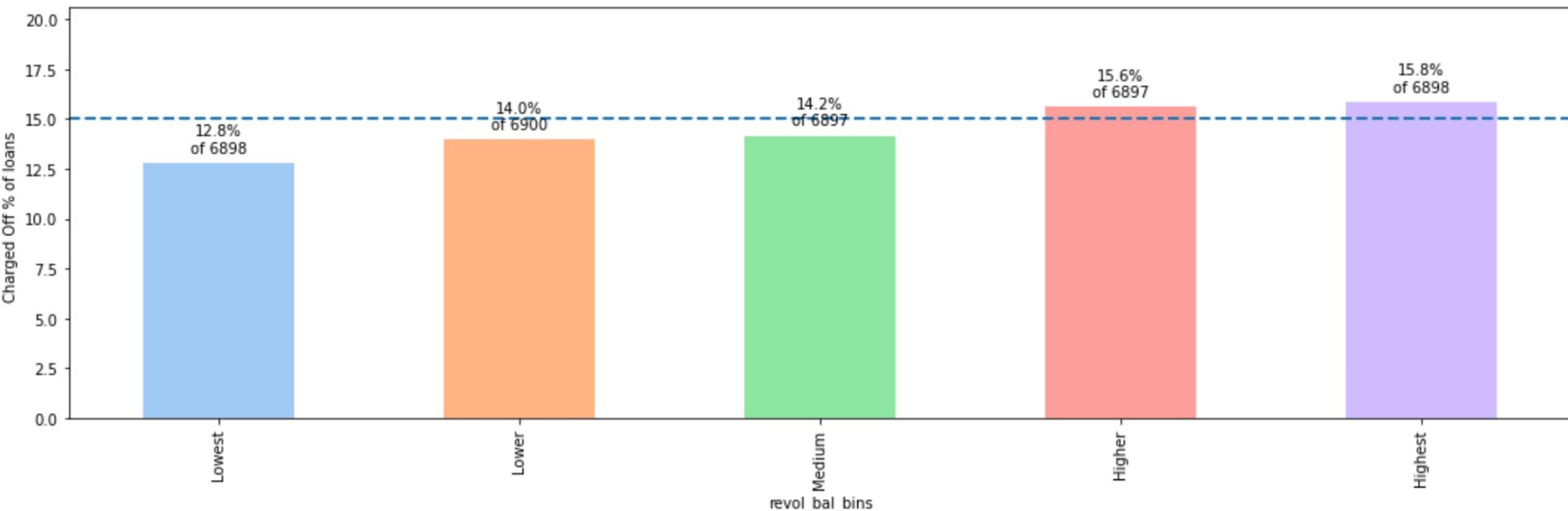


Observation:

- ❖ High DTI (debt to income ratio) results in higher charged off % for loans.



REVOLVING BALANCE

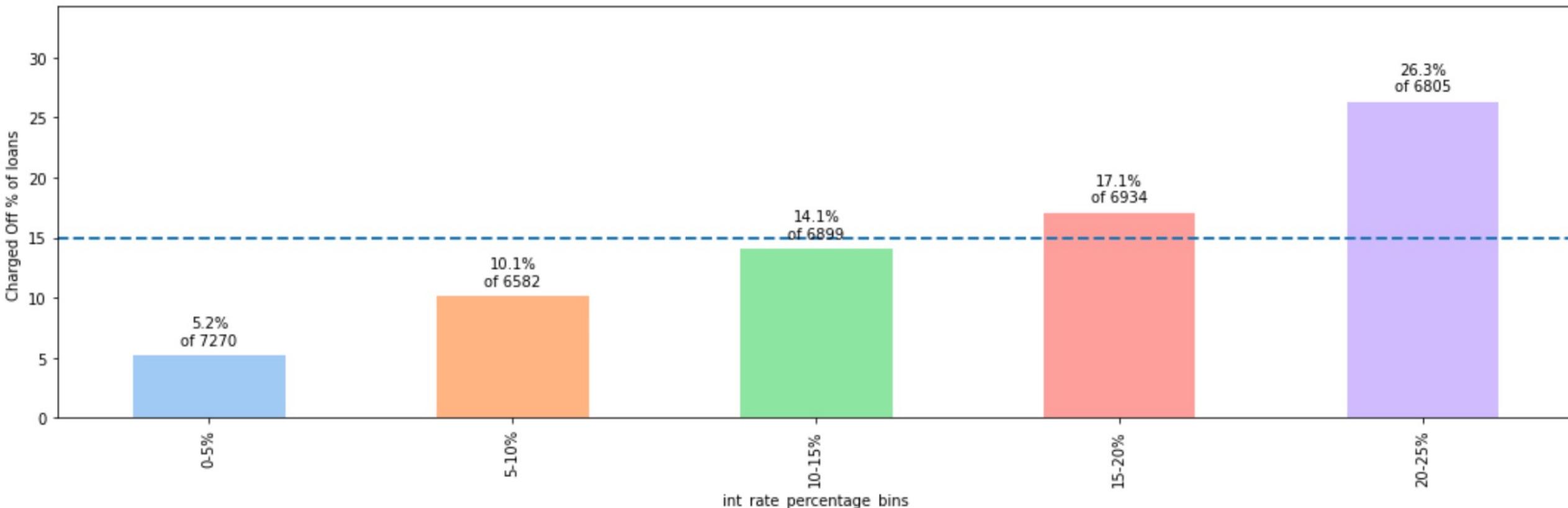


Observation:

- ❖ Higher the revolving balance, bigger the chance of the loan getting defaulted.



INTEREST RATE

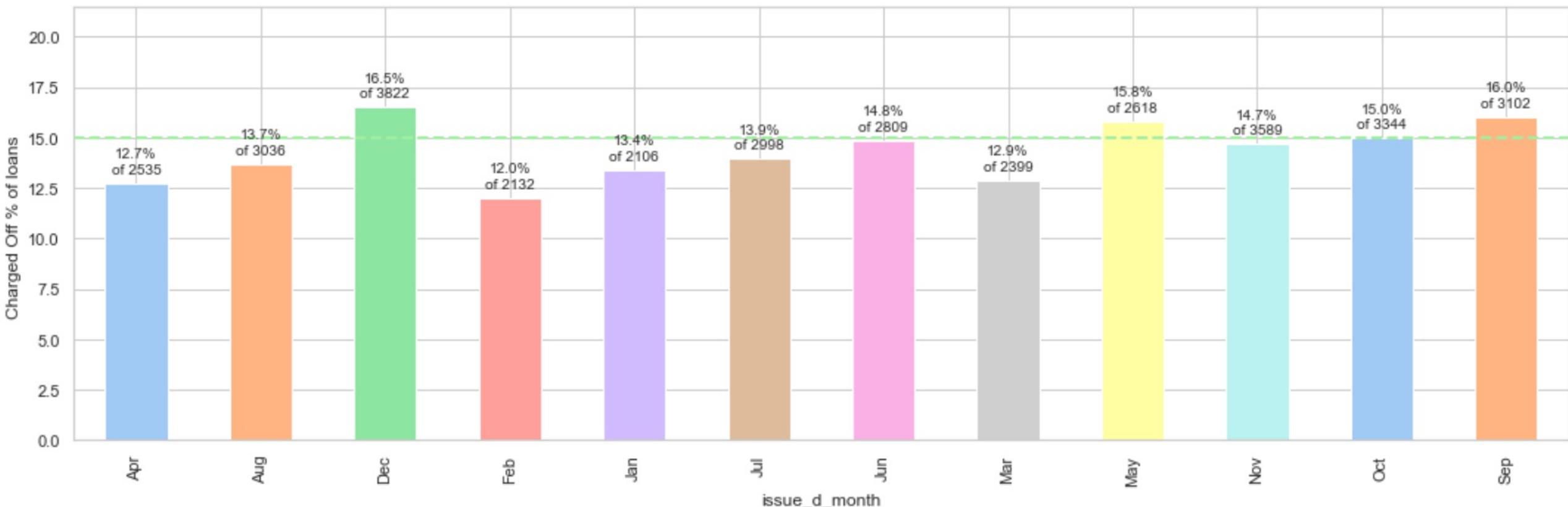


Observation:

- ❖ Loans with high interest rates results into high default rates.



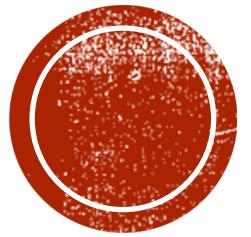
ISSUE DATE



Observation:

- ❖ December has the highest charged off ratio.





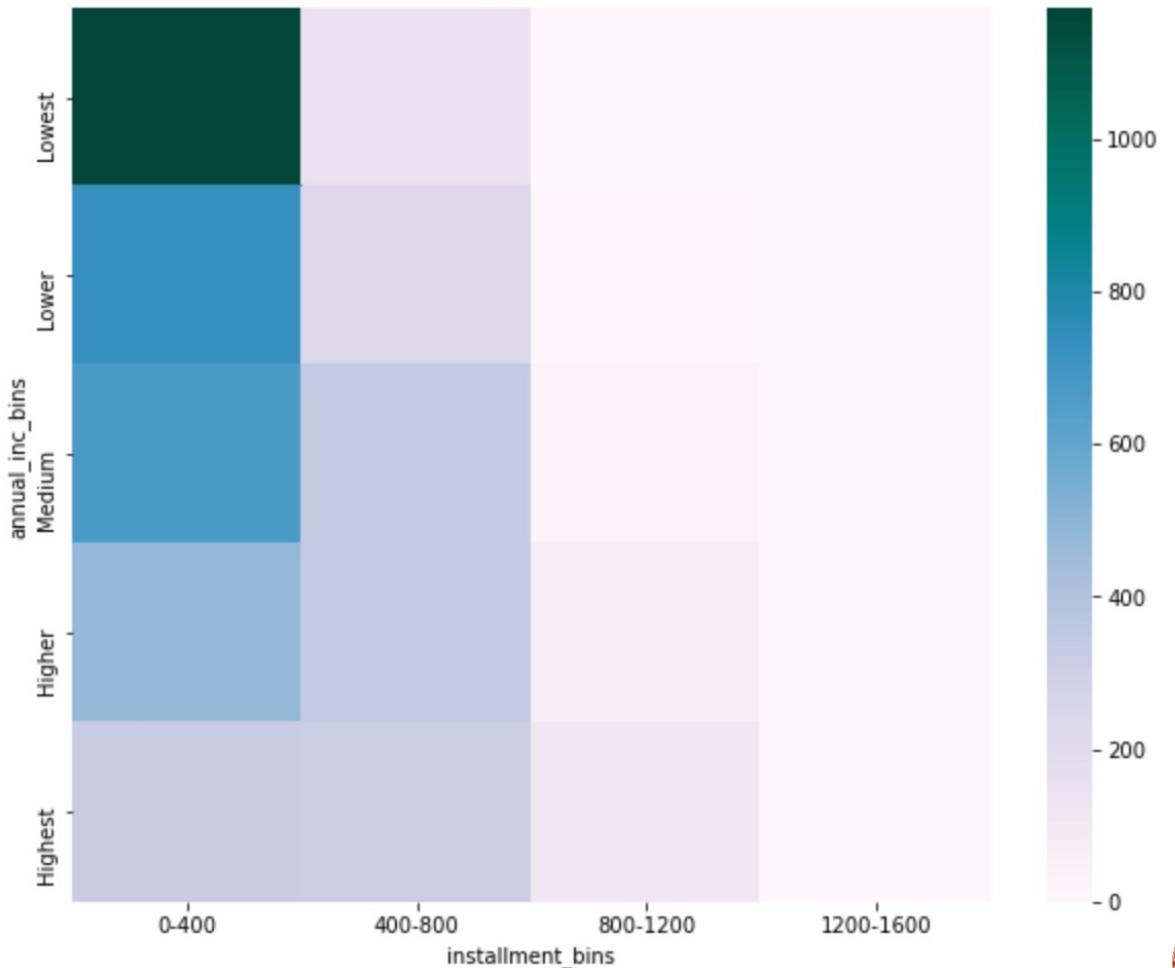
BIVARIATE ANALYSIS



ANNUAL INCOME AND INSTALMENT

Observation:

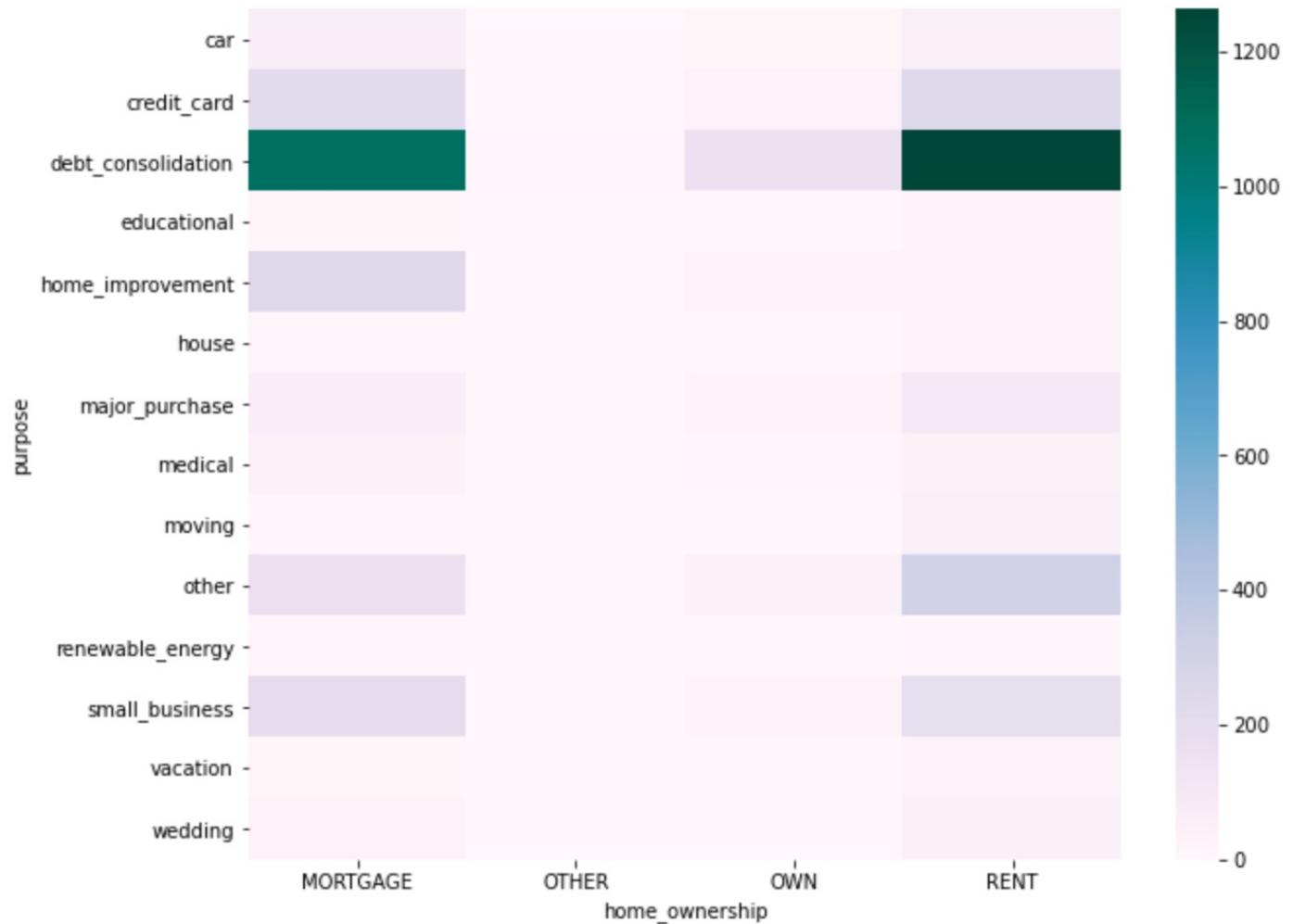
- ❖ Chances of defaulting is high when the borrower has annual income in the lowest segment, even though the instalment is lowest too.



HOME OWNERSHIP AND PURPOSE

Observation:

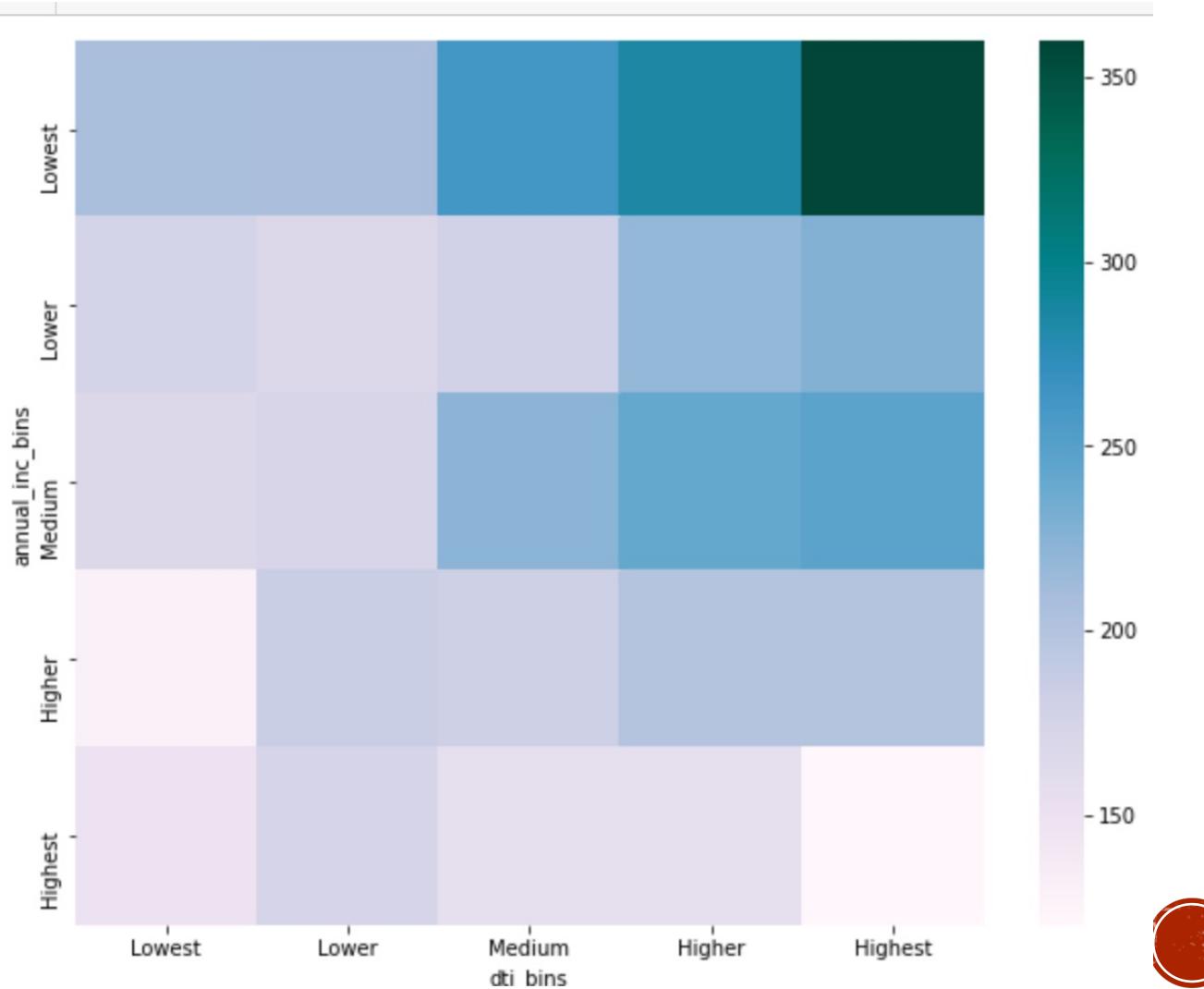
- ❖ Chances of defaulting is high when the borrower living in mortgage or rented home takes loan for the purpose of debt consolidation.



DTI AND ANNUAL INCOME

Observation:

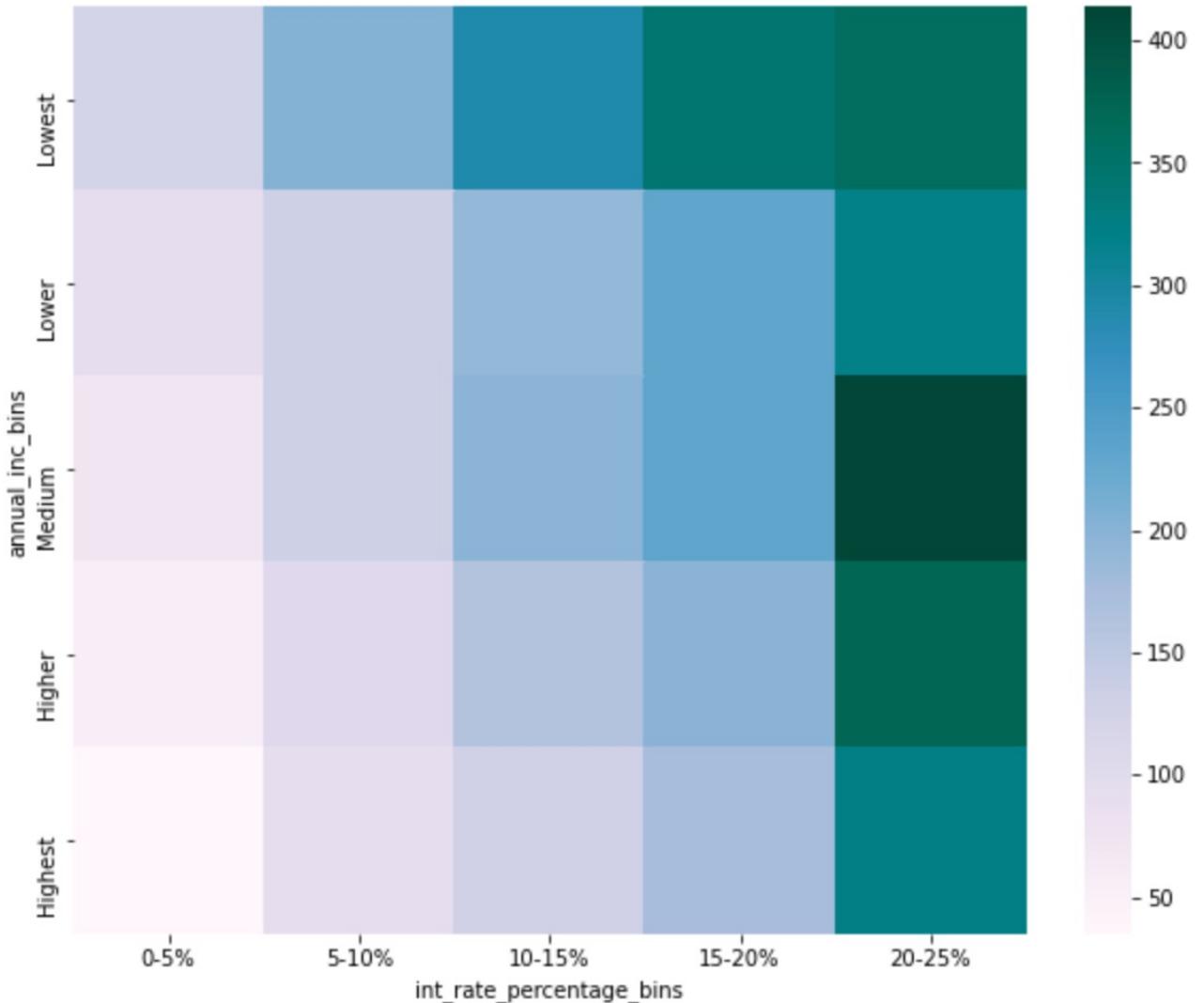
- ❖ Chances of defaulting is high for the borrower with lowest income and high dti(debt to income ratio)



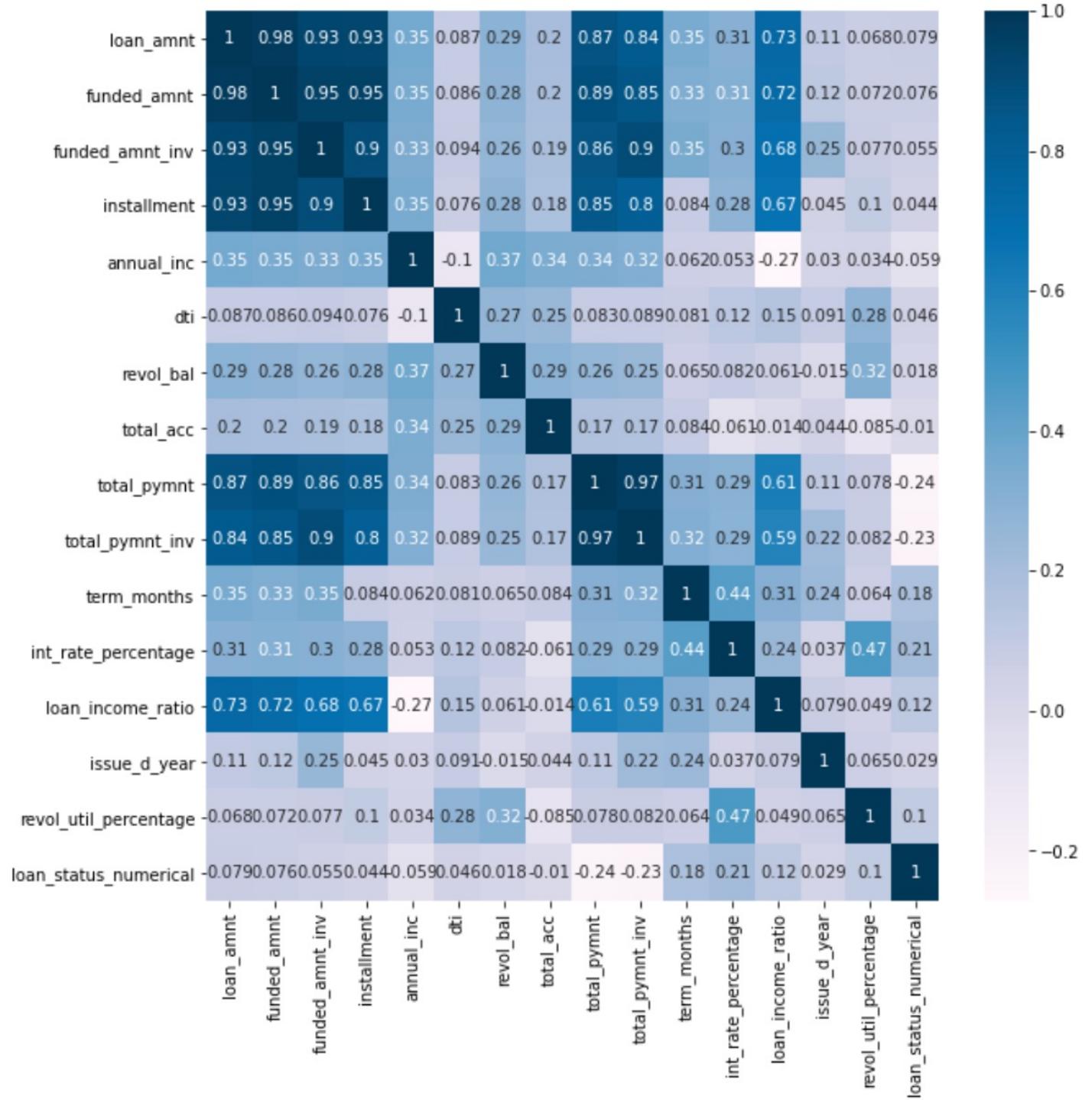
INTEREST RATE AND ANNUAL INCOME

Observation:

- ❖ High interest rate for lower and medium income groups lead to high default rate



CORRELATION MATRIX



Observation:

- ❖ Positive correlation of default loans with loan_amount, funded_amnt, dti, int_rate, issue_d_year and revol_util_percentage. It means the charged off loans % increases with an increase in these values.
- ❖ Negative correlation of default loans with annual_inc, total_payment. It means the charged off loans % increases with an decrease in these values



CONCLUSION

The most impactful drivers are :

Univariate Factors

- Higher interest rate
- Term of the loan (60 months)
- Loan purpose (small business, renewable energy, educational)
- Higher revolving line utilization rate
- Loan grade & sub-grade
- Higher loan amount
- Higher instalment amount
- Lower annual income
- Higher debt to income ratio
- Applicant's address state
- Loan issue month (Dec, May, Sep)

Bivariate Factors

- High loan amount & interest rate for lower income group
- Home ownership (mortgaged and rented) and loan purpose (debt consolidation)
- Residential state and loan purpose
- lowest income and high dti(debt to income ratio)