LENDING CLUB CASE STUDY

PREPARED BY:

MADHAVARAPU NAGAMANOHAR
CHAITRA MOKHASI

PROBLEM STATEMENT

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.

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default. The company can utilize this knowledge for its portfolio and risk assessment.

BUSINESS UNDERSTANDING

Lending Club is a consumer finance company which specializes in lending various types of loans to urban customers based on the applicant's profile.

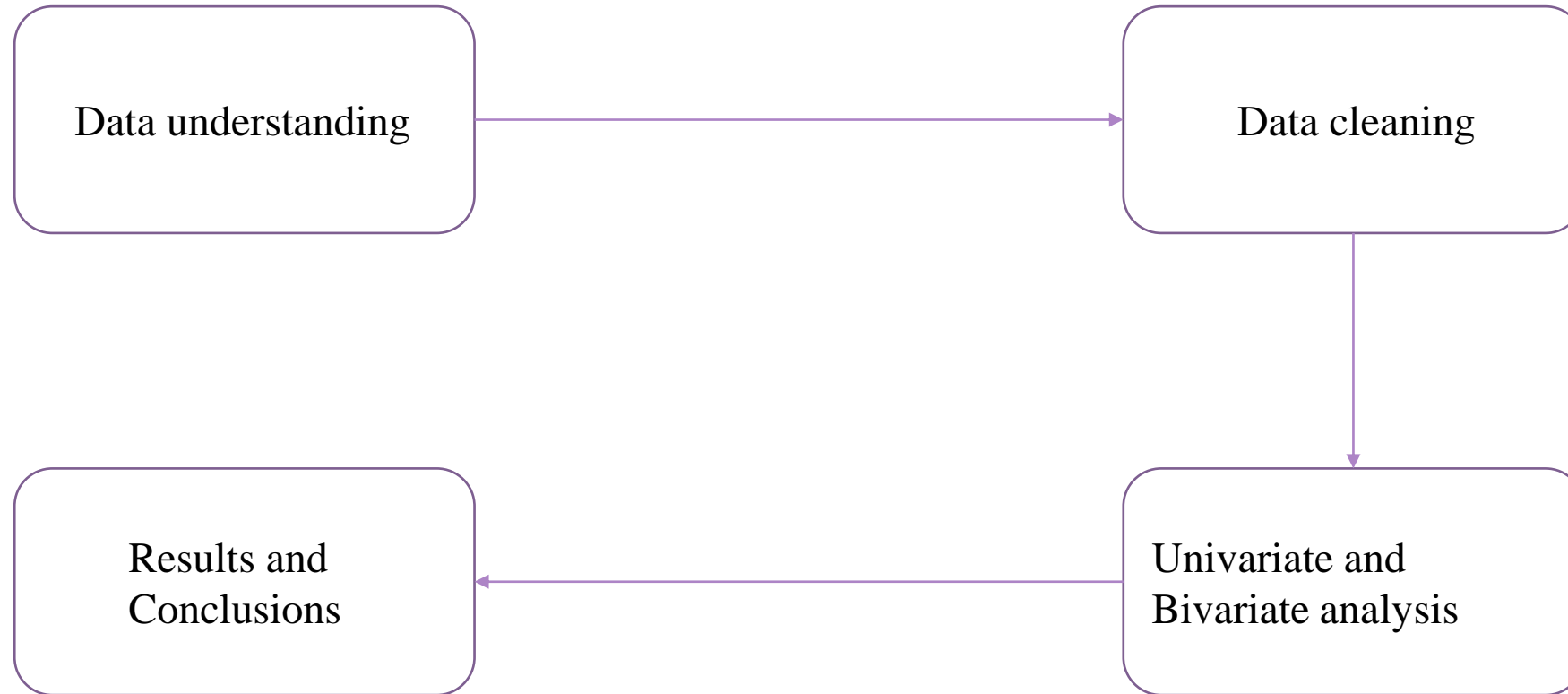**Types of risks are associated with the company's decision** :

If the applicant is **,**

1. Likely to repay the loan

2. Not likely to repay the loan

Types of decisions that could be taken by the company :

1. Loan accepted
   - Fully paid : Applicant has fully paid the loan
   - Current: Applicant is in the process of paying the instalments
   - Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e., he/she has **defaulted** on the loan

2. Loan rejected

Analysis and Approach

```
┌─────────────────────────┐                    ┌─────────────────────────┐
│                         │                    │                         │
│   Data understanding    │ ─────────────────> │     Data cleaning       │
│                         │                    │                         │
└─────────────────────────┘                    └─────────────────────────┘
                                                             │
                                                             │
                                                             ▼
┌─────────────────────────┐                    ┌─────────────────────────┐
│                         │                    │                         │
│   Results and           │ <───────────────── │   Univariate and        │
│   Conclusions           │                    │   Bivariate analysis    │
└─────────────────────────┘                    └─────────────────────────┘
```

DATA CLEANING

- Separate the data as Numerical and Categorical.
- Get the missing percentage of the all the columns and drop the columns where the missing percentage is more than 50.
- Dropping the columns which are not useful for insights and unwanted.
- Treat the missing values with median for the numeric data.(as it is not sensitive for the outliers)
- Treat the missing values with the mode for the categorical data.

UNIVARIATE ANALYSIS

- Box plot will be used to find out the outliers in univariate analysis.
- Column 'annual_inc' has more outliers. [Referring the boxplots from next slide]
- Treating outliers by getting the percentile range.
- Filtering the data below or equal to variation which found in percentile range.

Box plots to check the outliers:
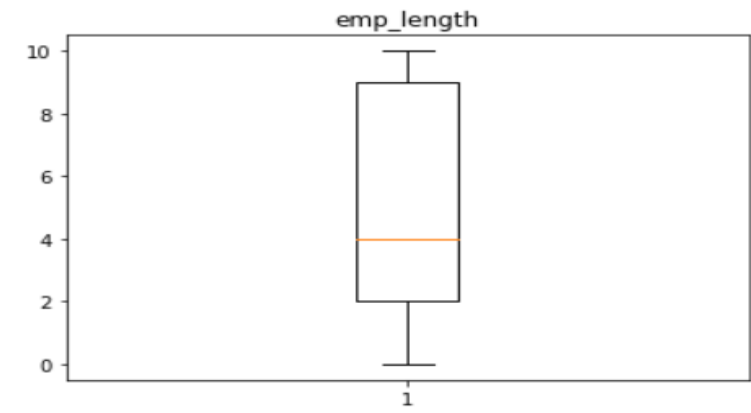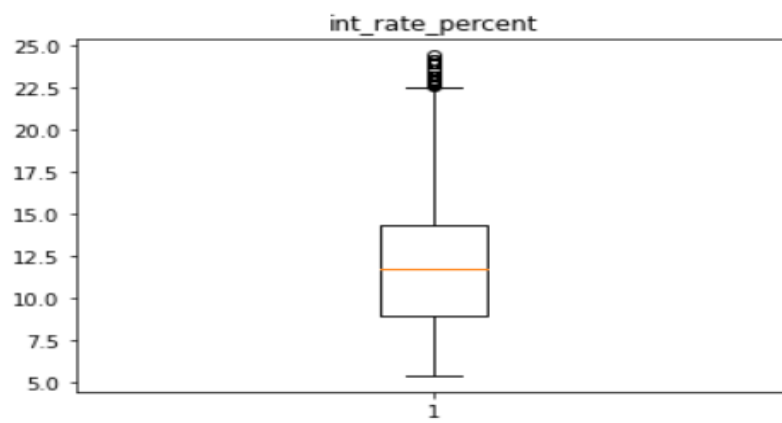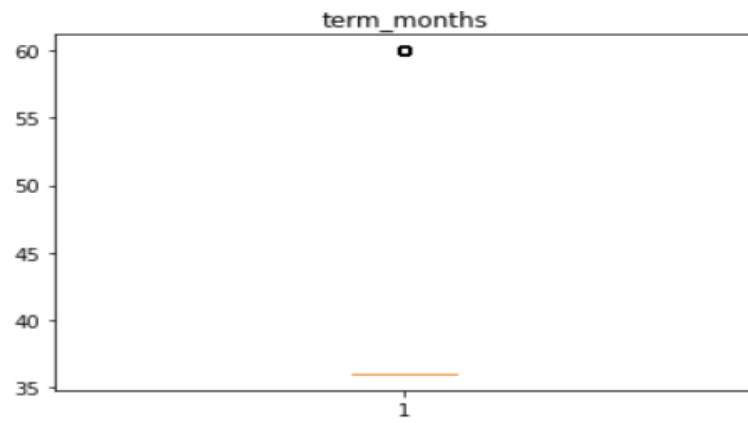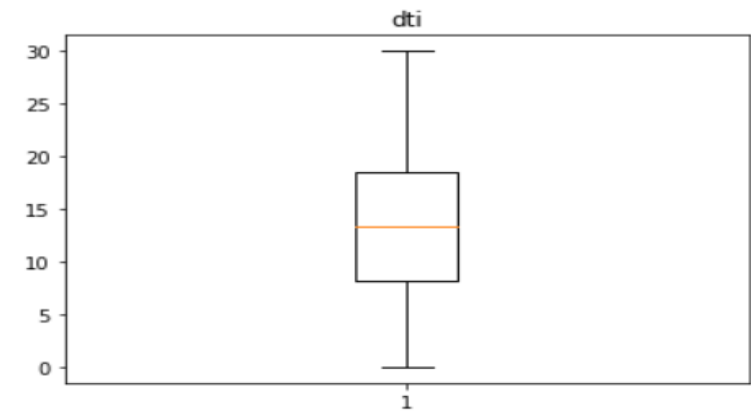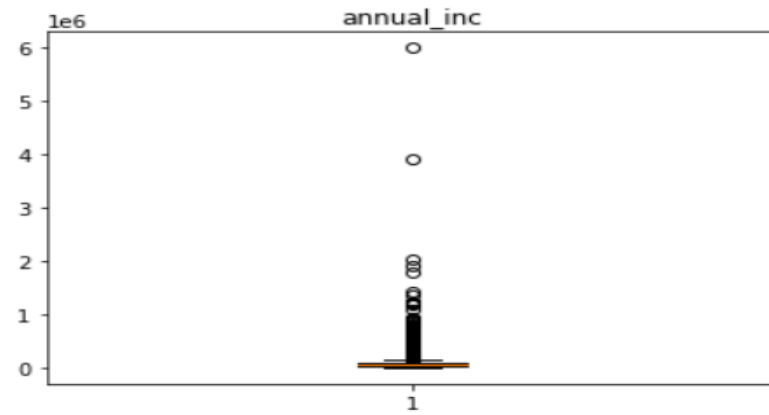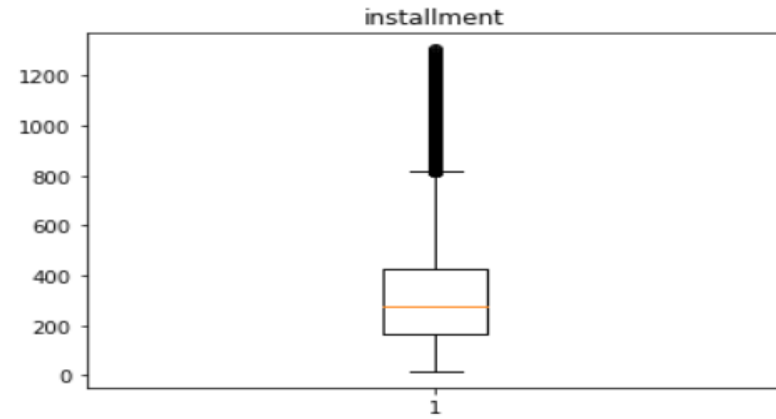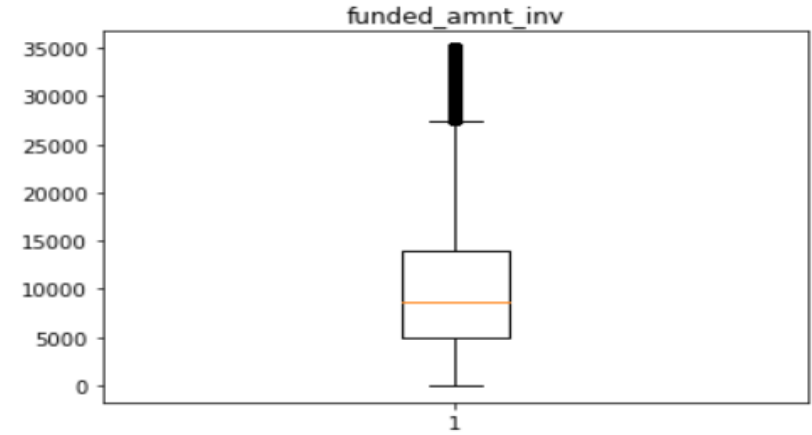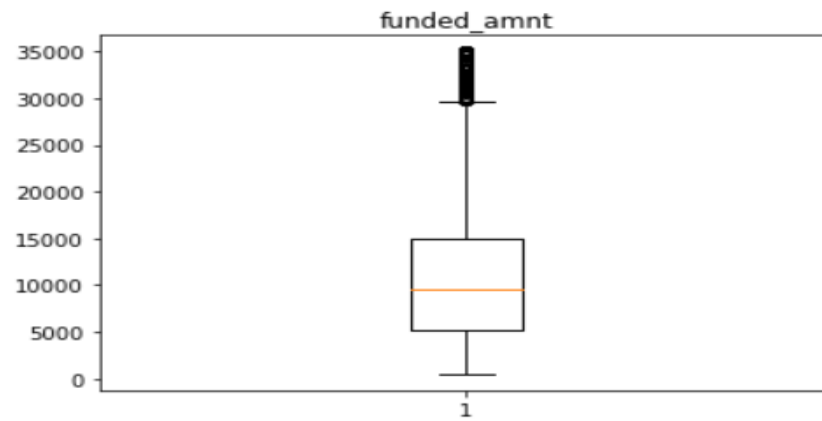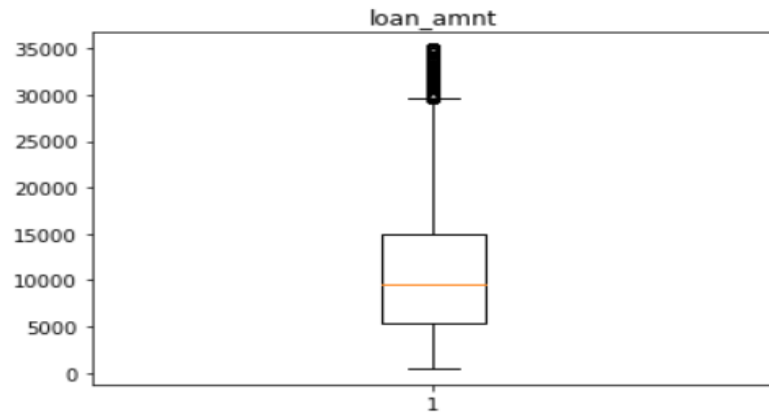
Driving factors derived from the univariate analysis:

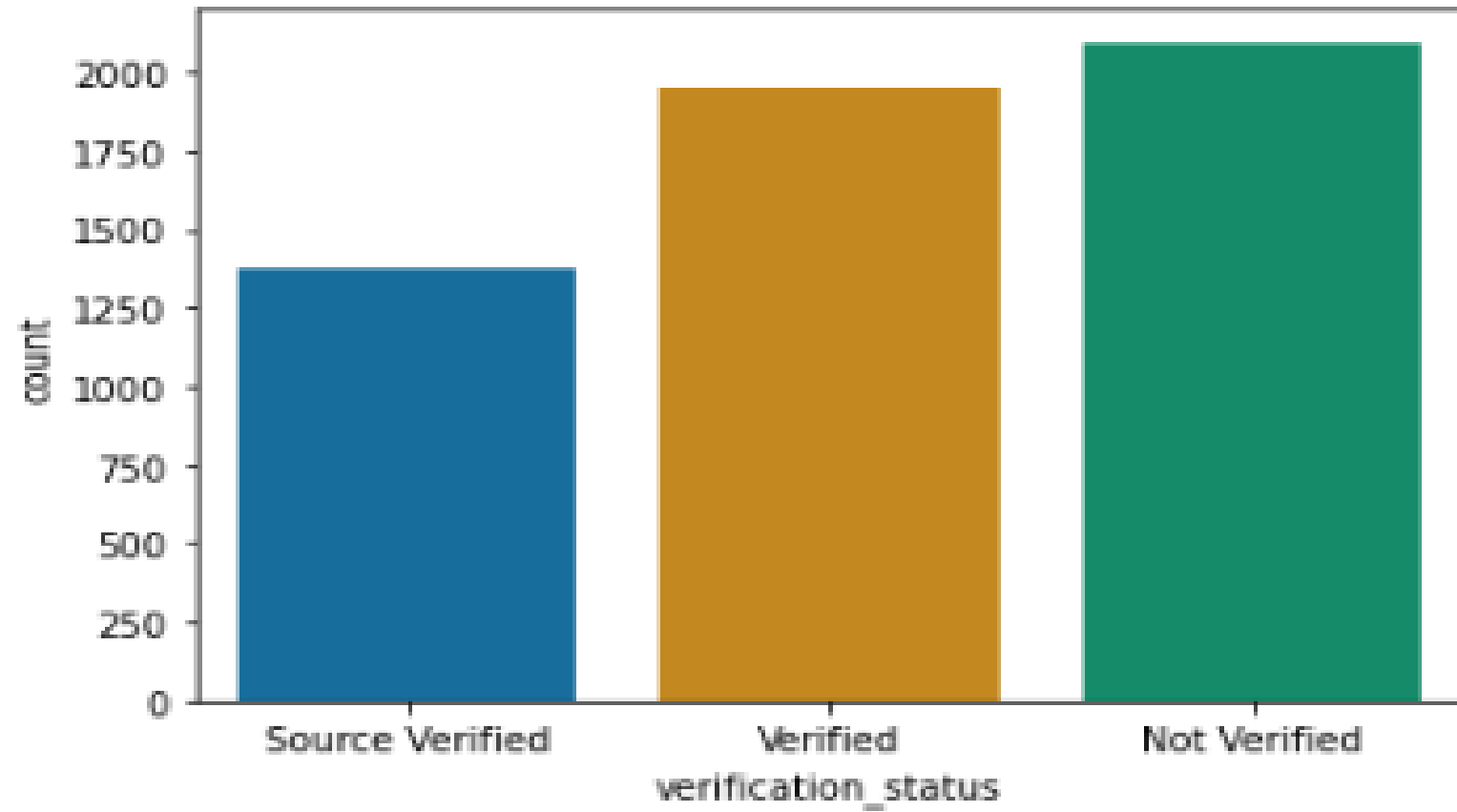**1.** Applicants having Grade as 'B' have more probability of defaulting.
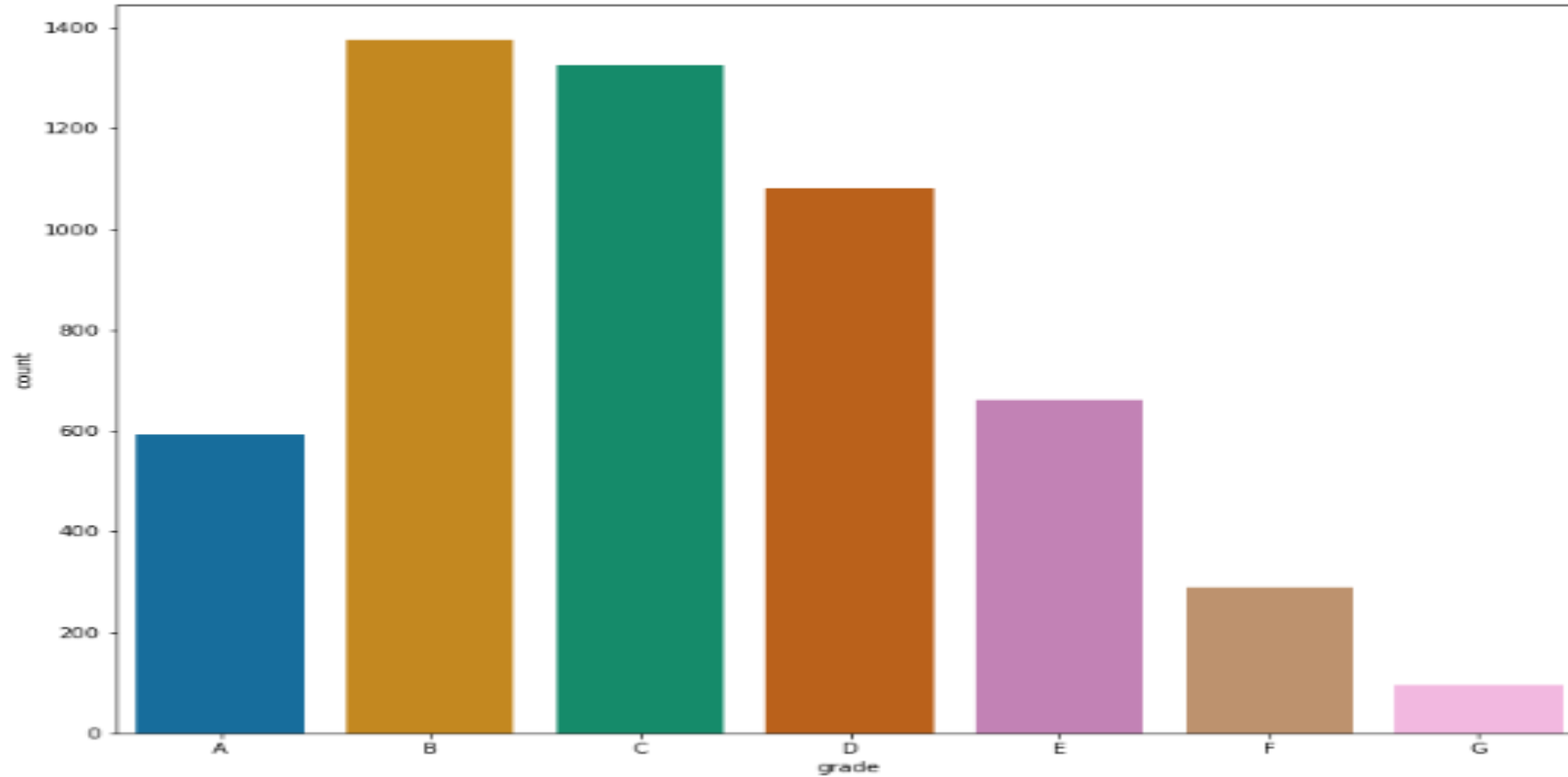
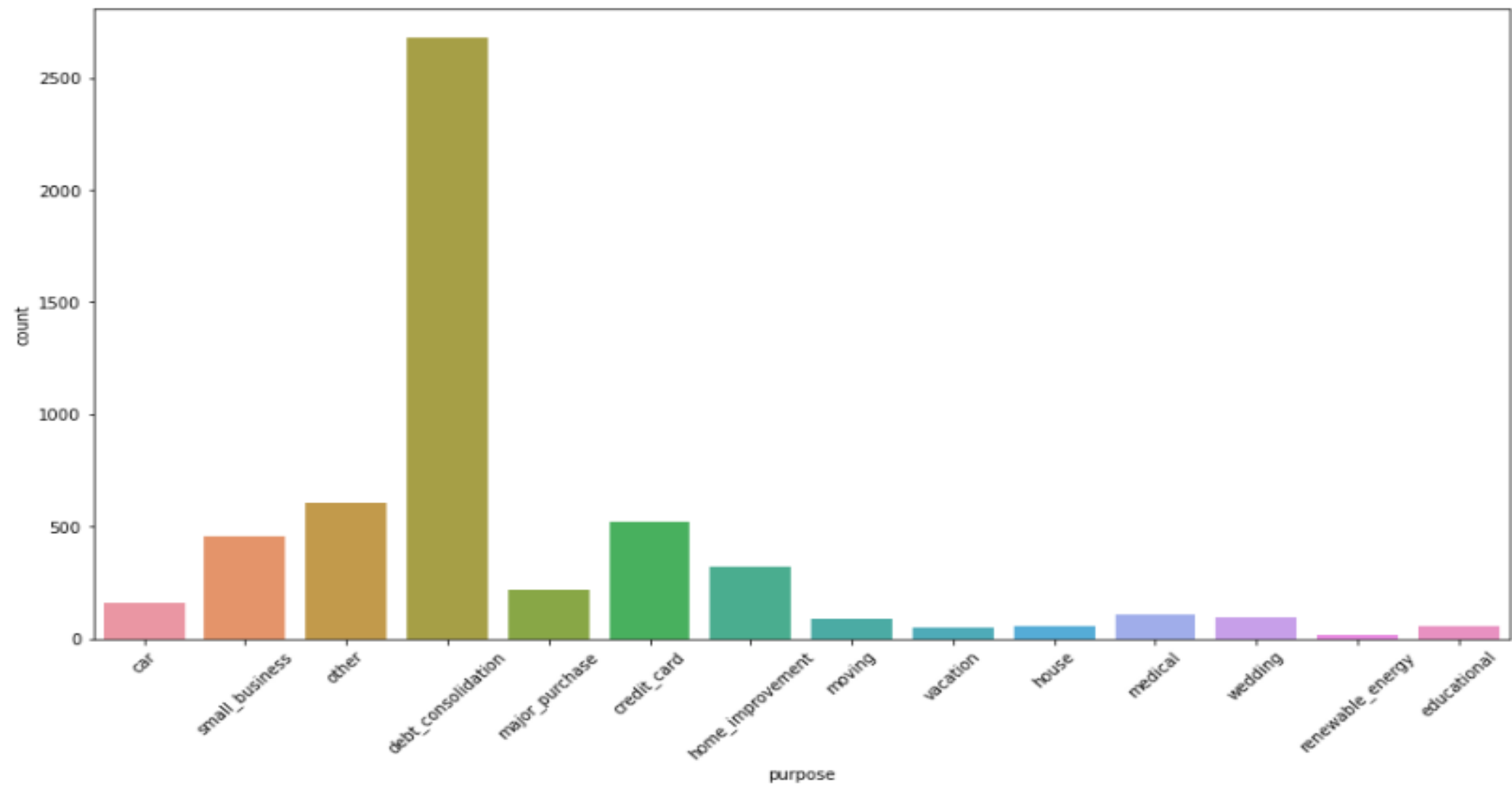**2.** Applicants having subgrade as 'B5' have more probability of defaulting.
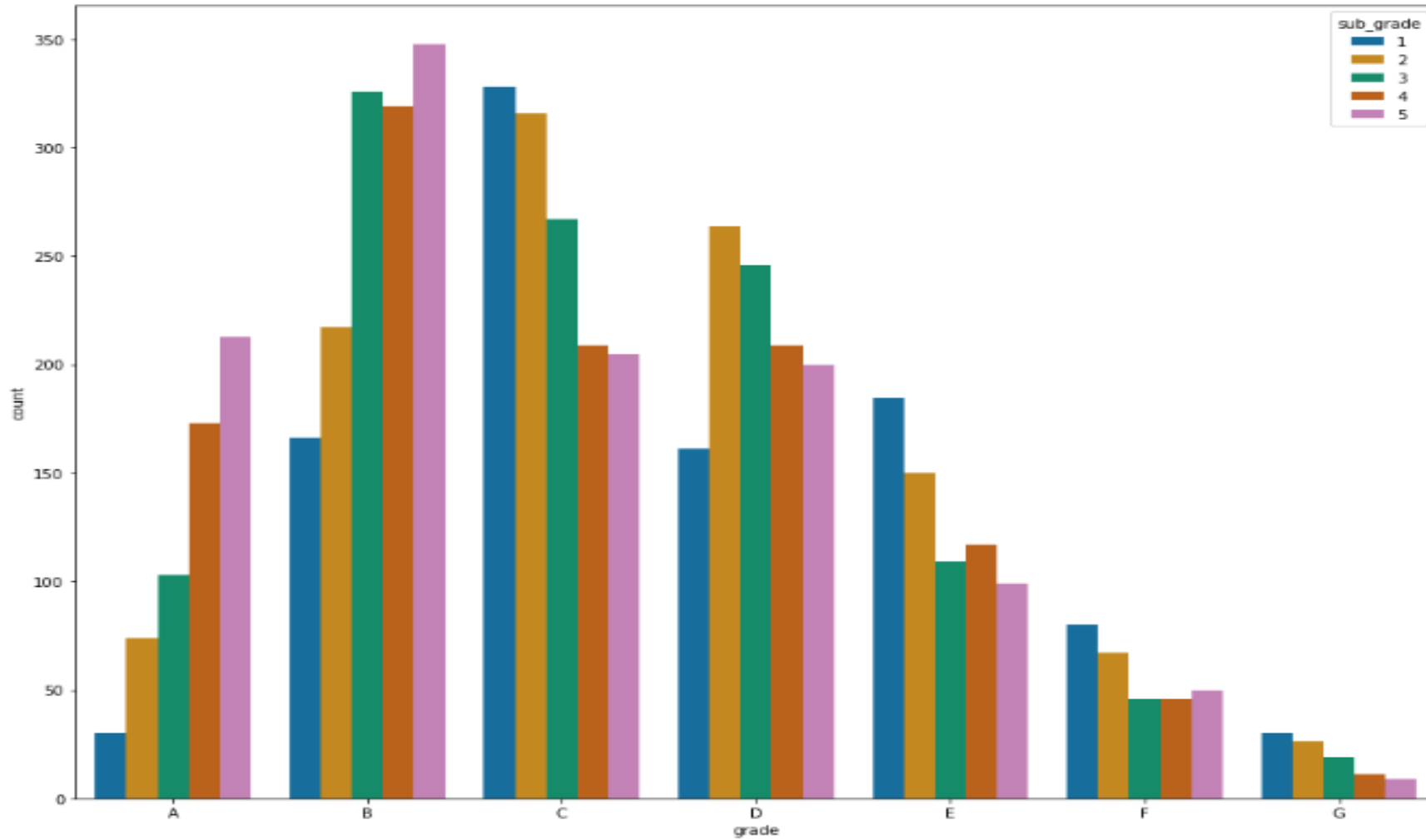
3. Applicants having house ownership as 'RENT' have more probability of defaulting.
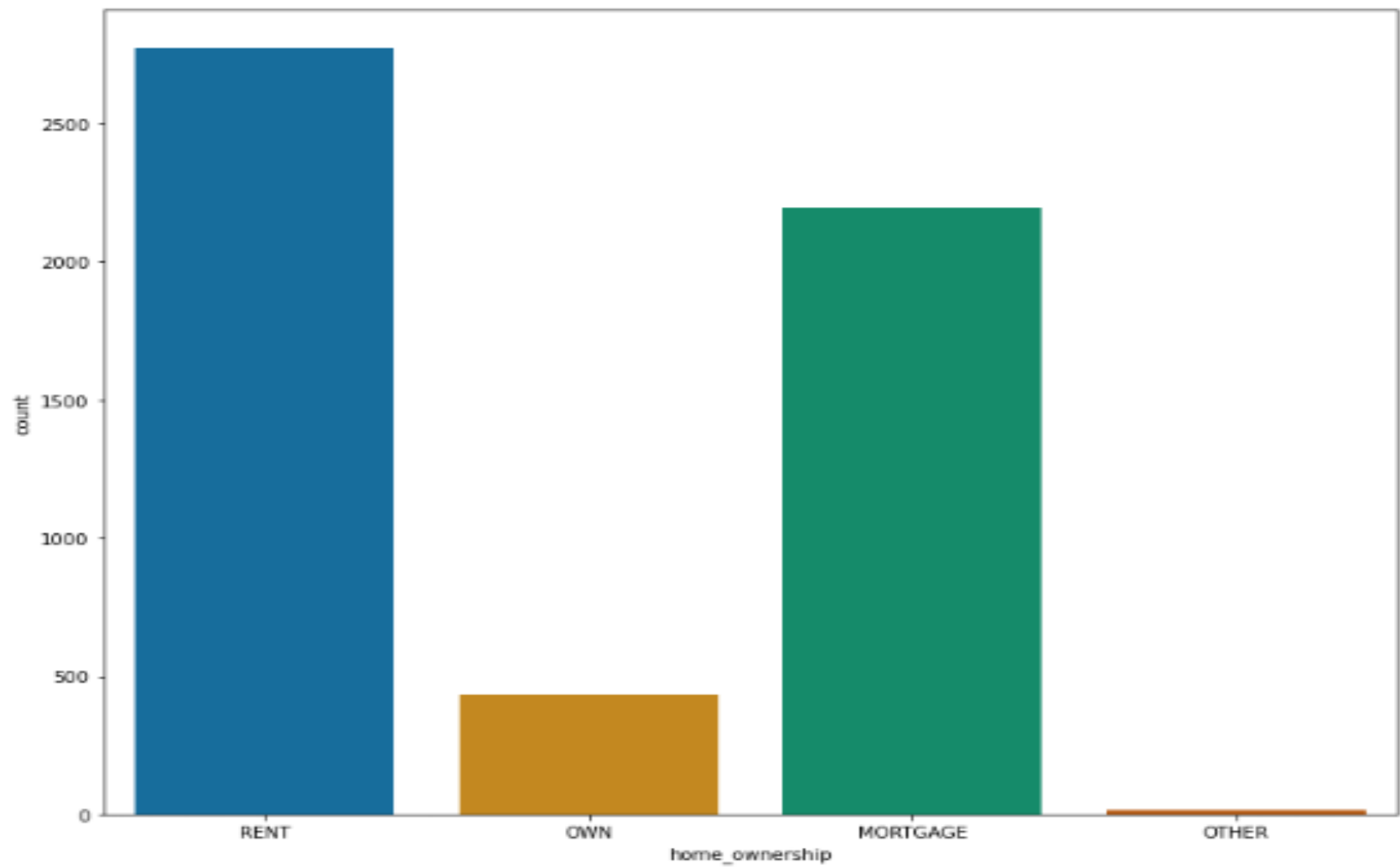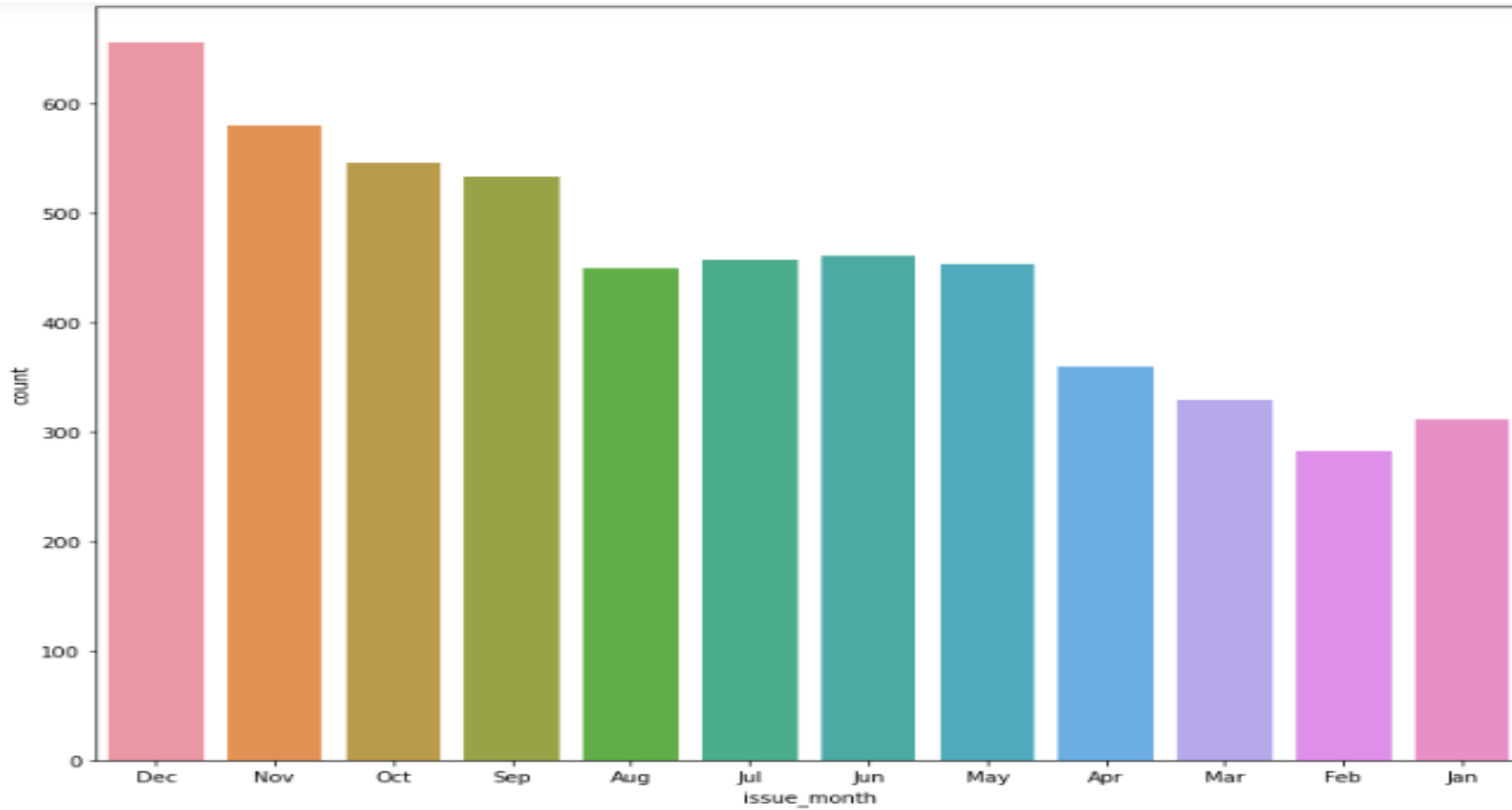
**4.** Applicants having verification status of loan as 'Not Verified' have more probability of defaulting.

5. Applicants having purpose as 'debt_consolidation' have more probability of defaulting.

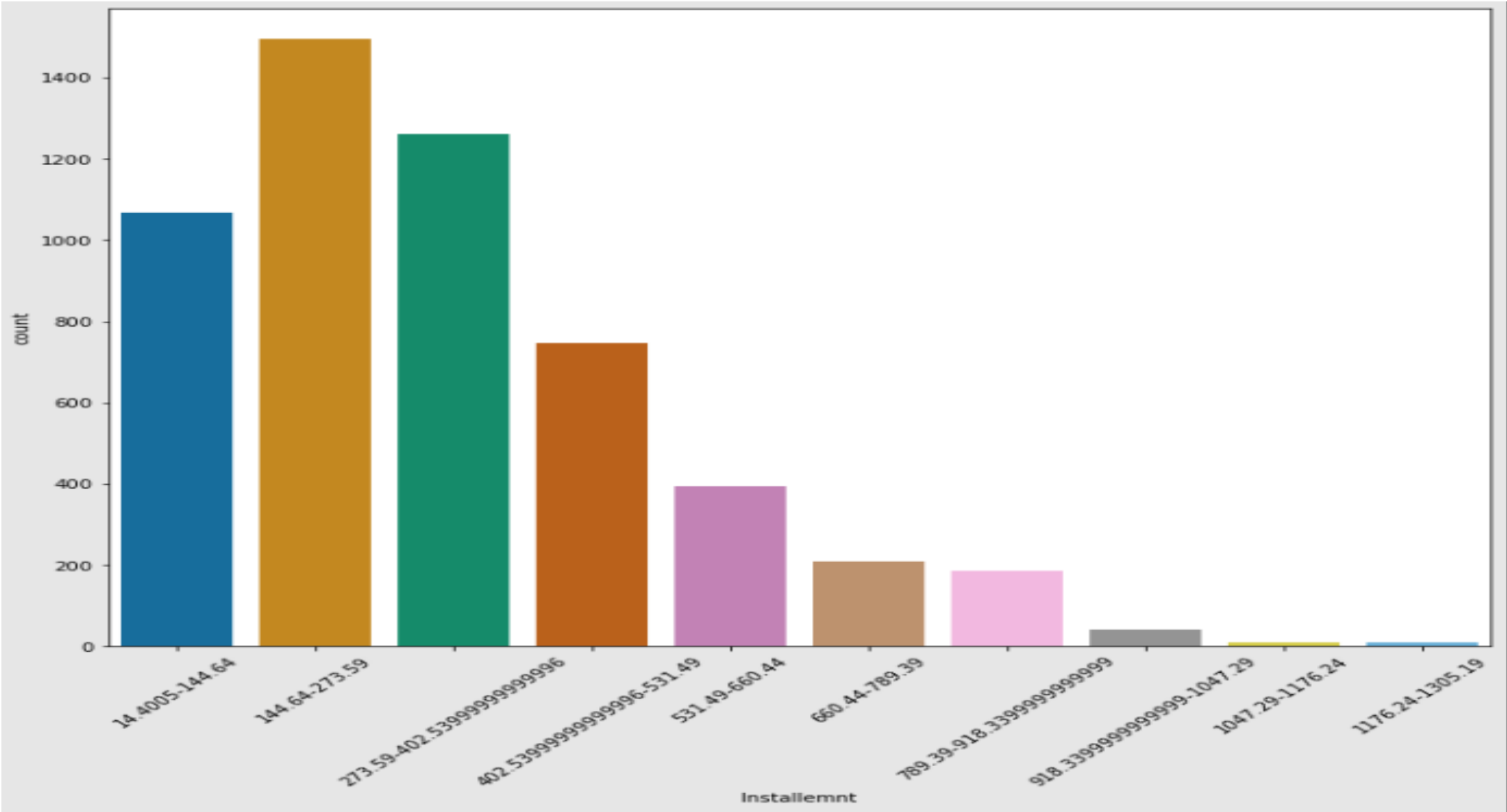**6.** Applicants applied in the month of November and December have more probability of defaulting.

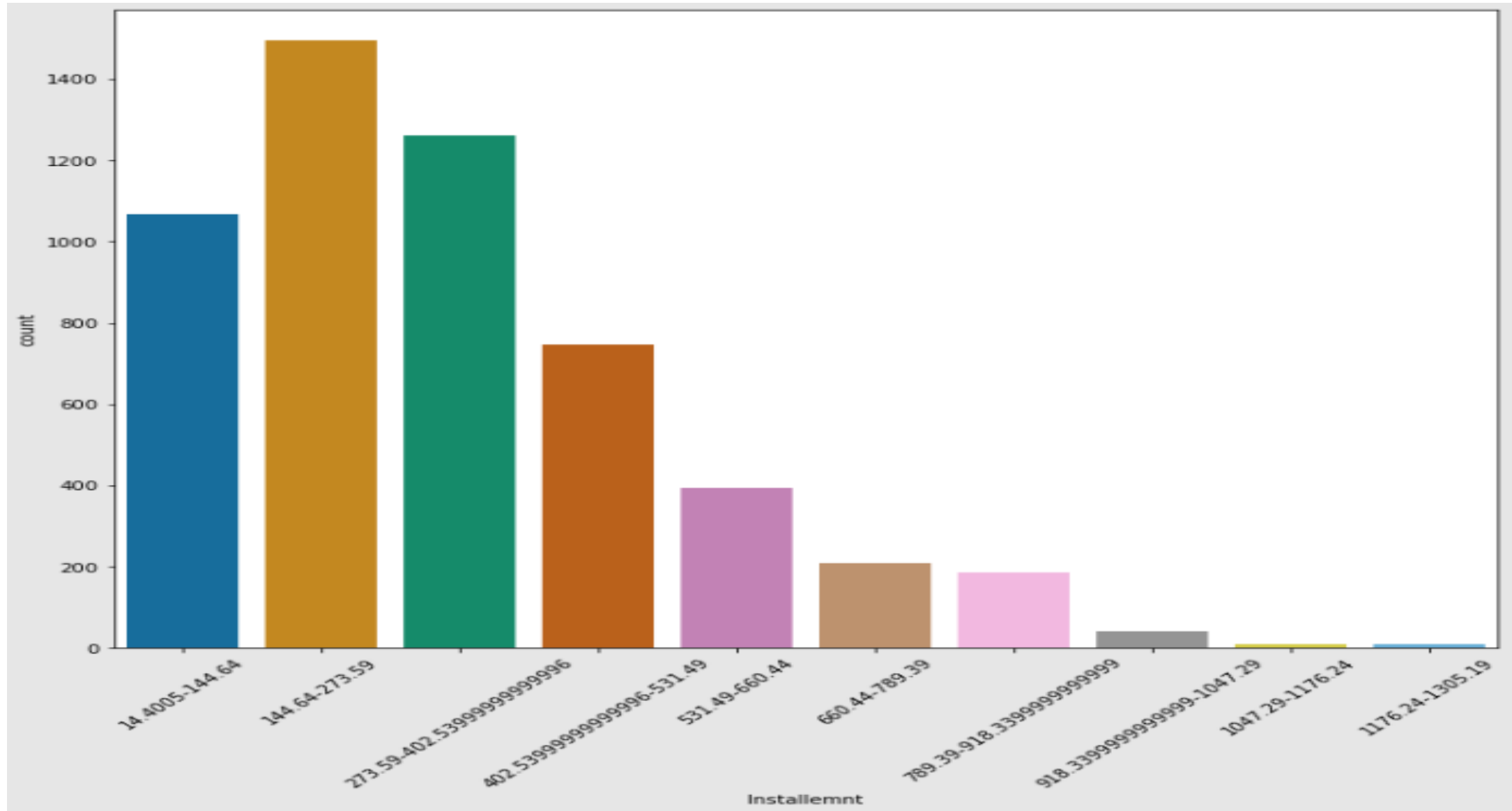**7.** Applicants having loan amount in the range of 5500-10000 have more probability of defaulting.
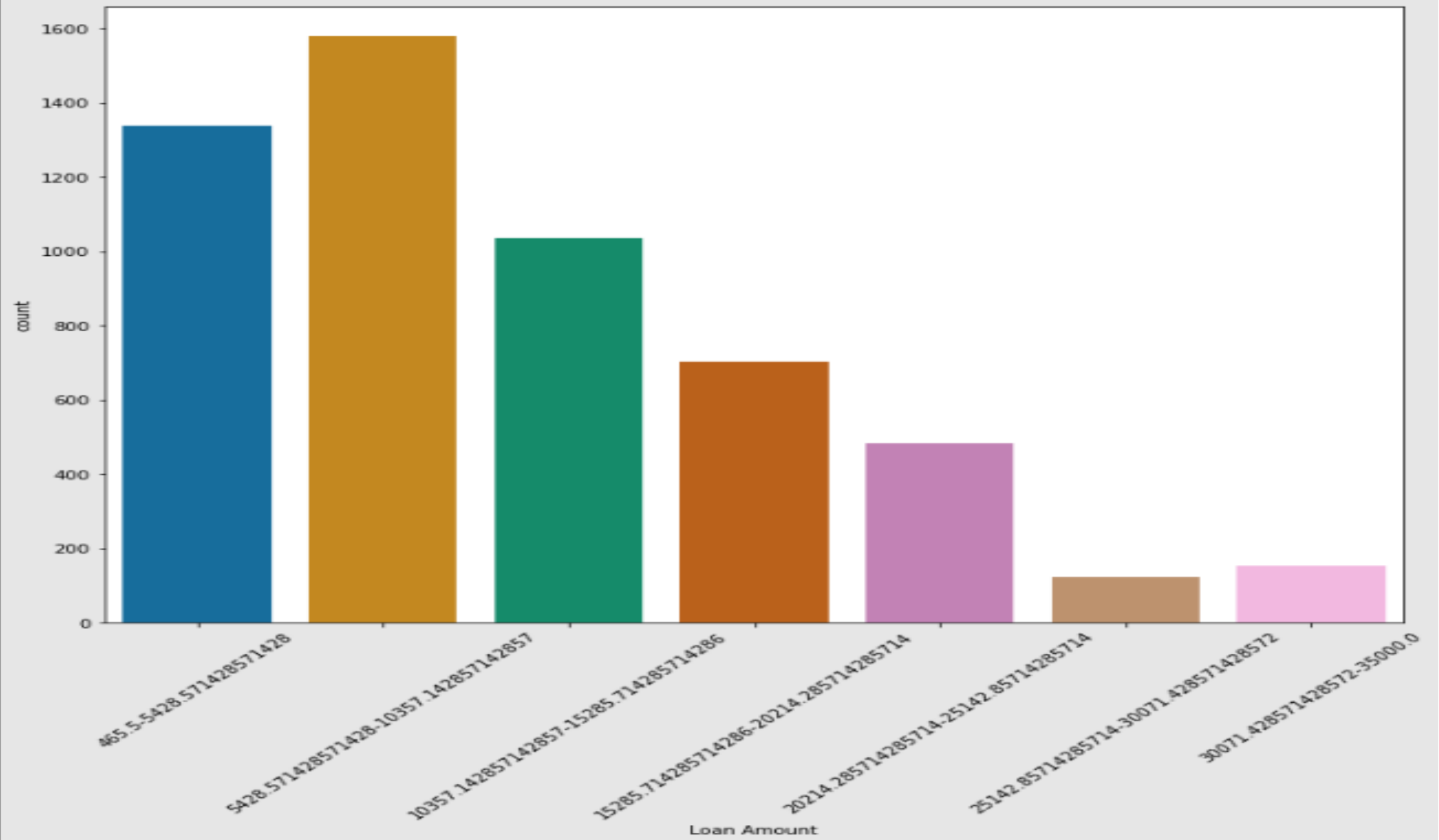
8. Applicants having funded amount by investor in the range of 5500-10000 have more probability of defaulting.
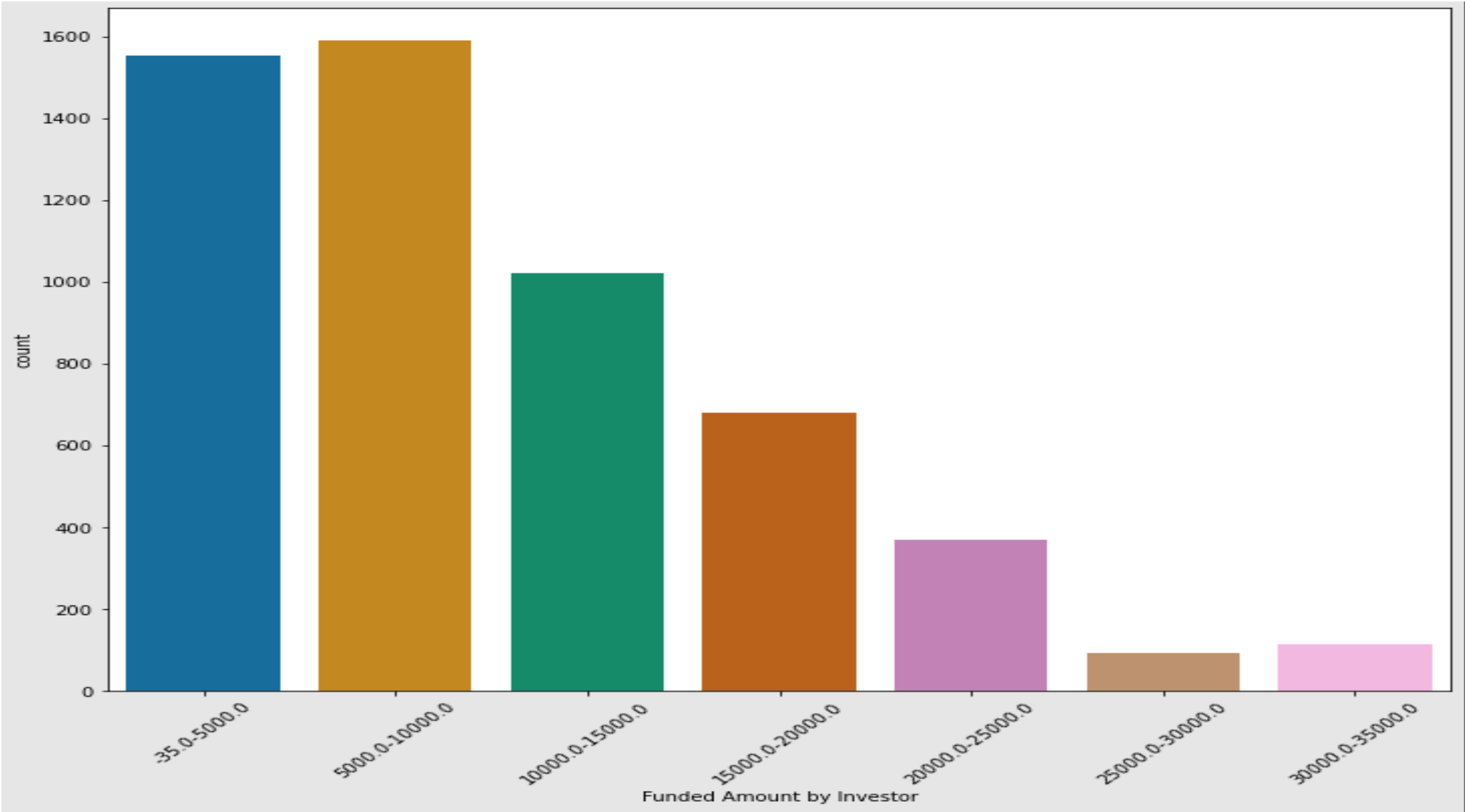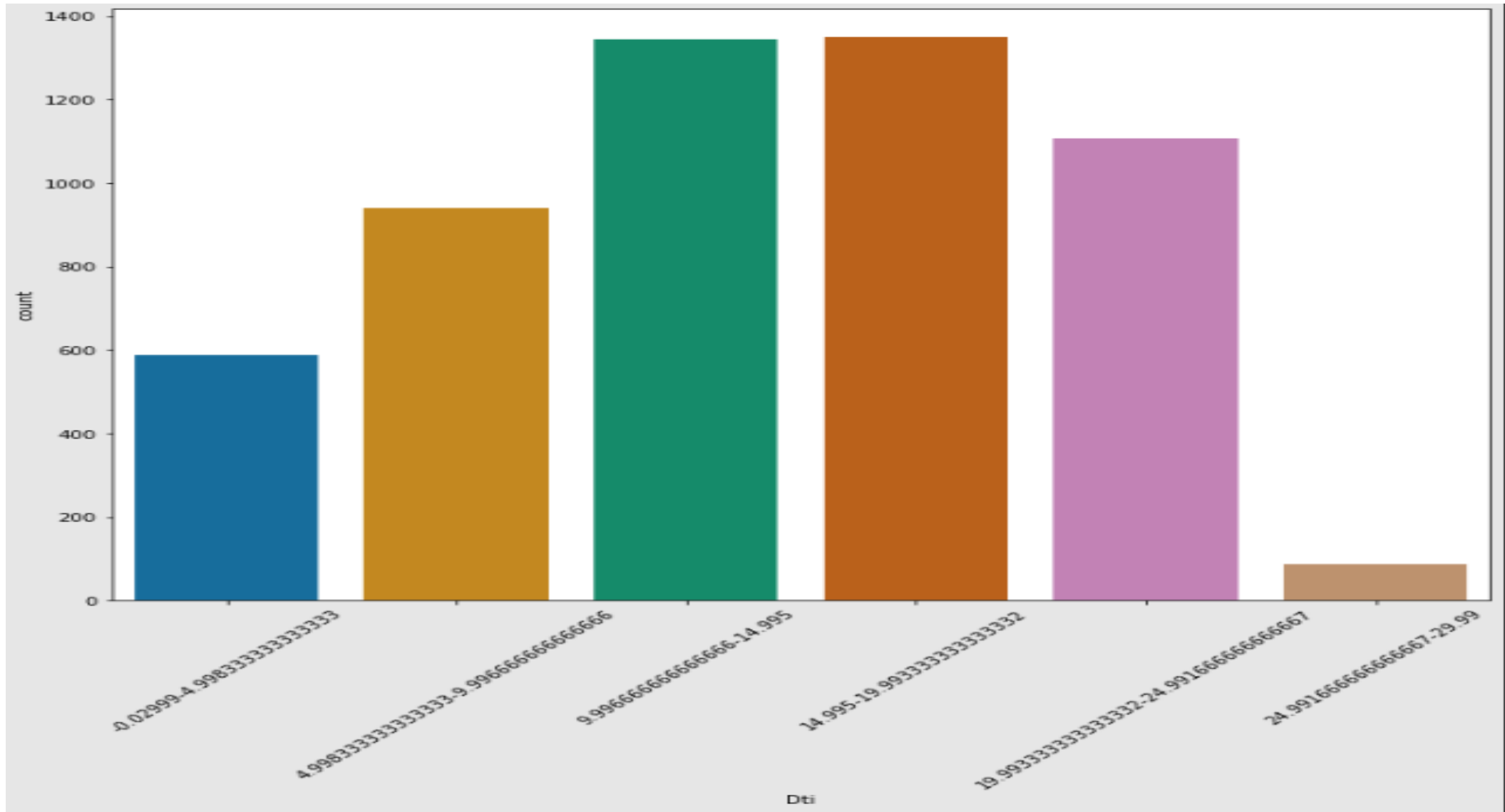
9. Applicants having installments in the range of 144 -273 have more probability of defaulting.
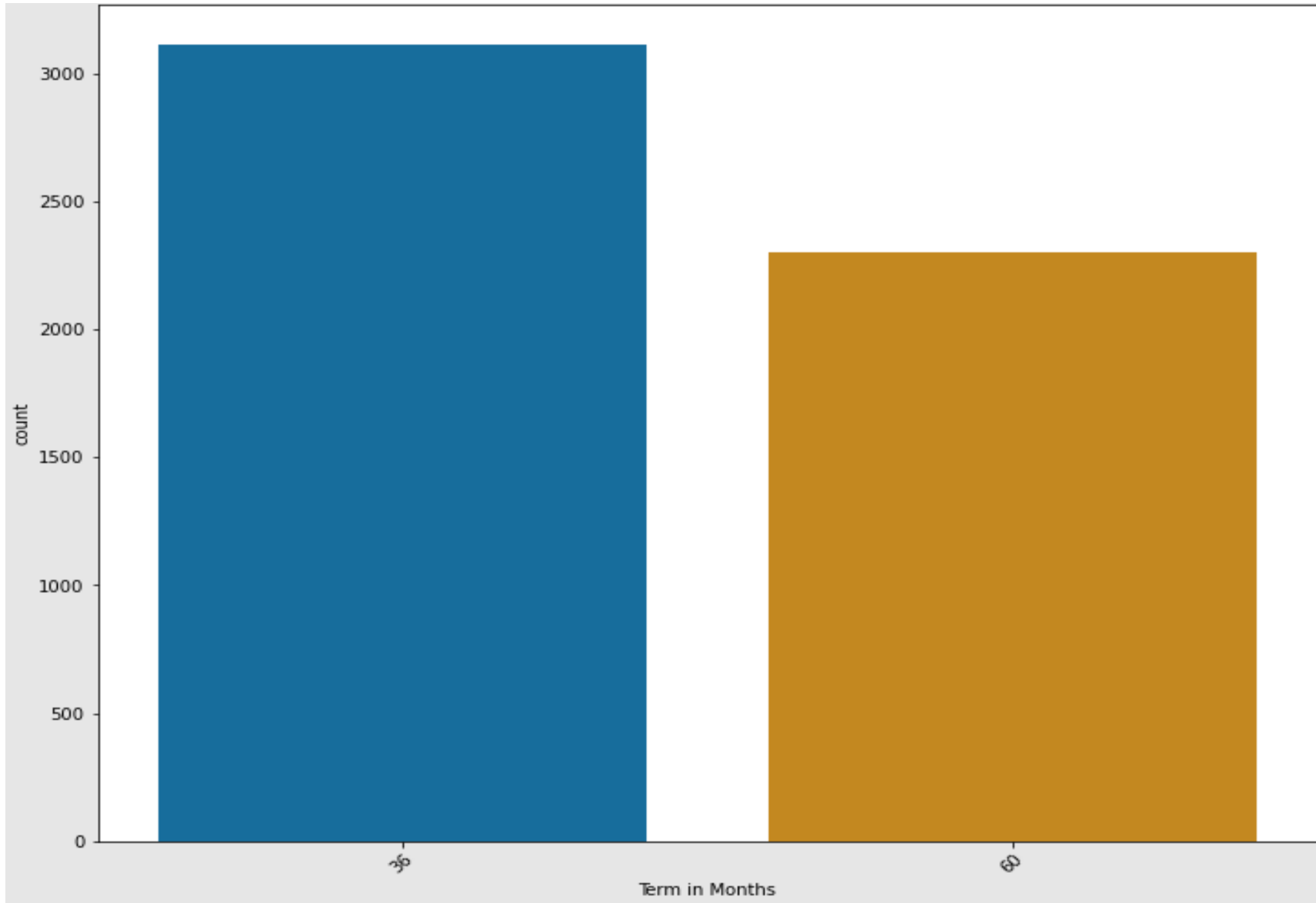
10. Applicants having annual income in the range of 31200 -58402 have more probability of defaulting.
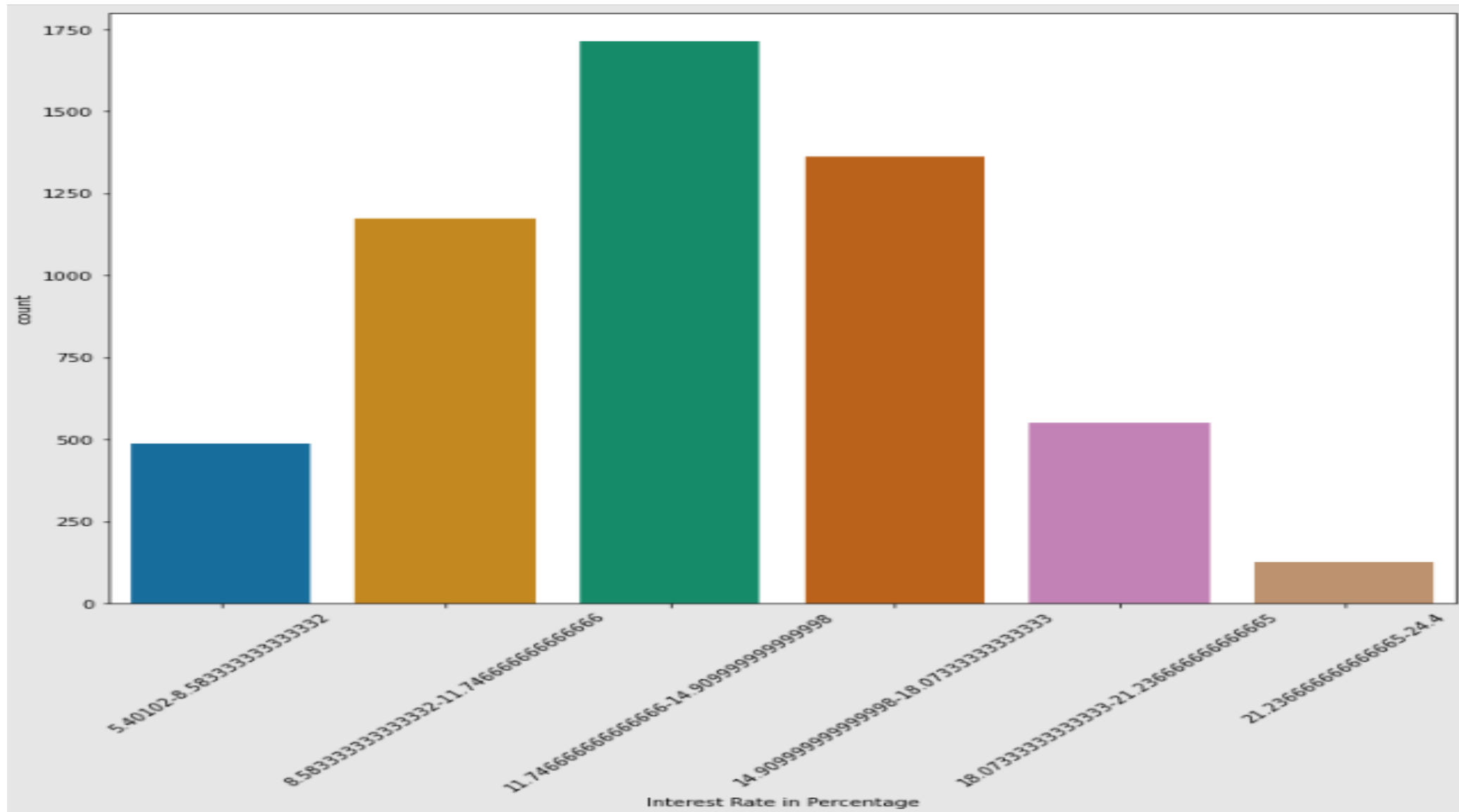
11. Applicants having Dti in the range of 10 -20 have more probability of defaulting.
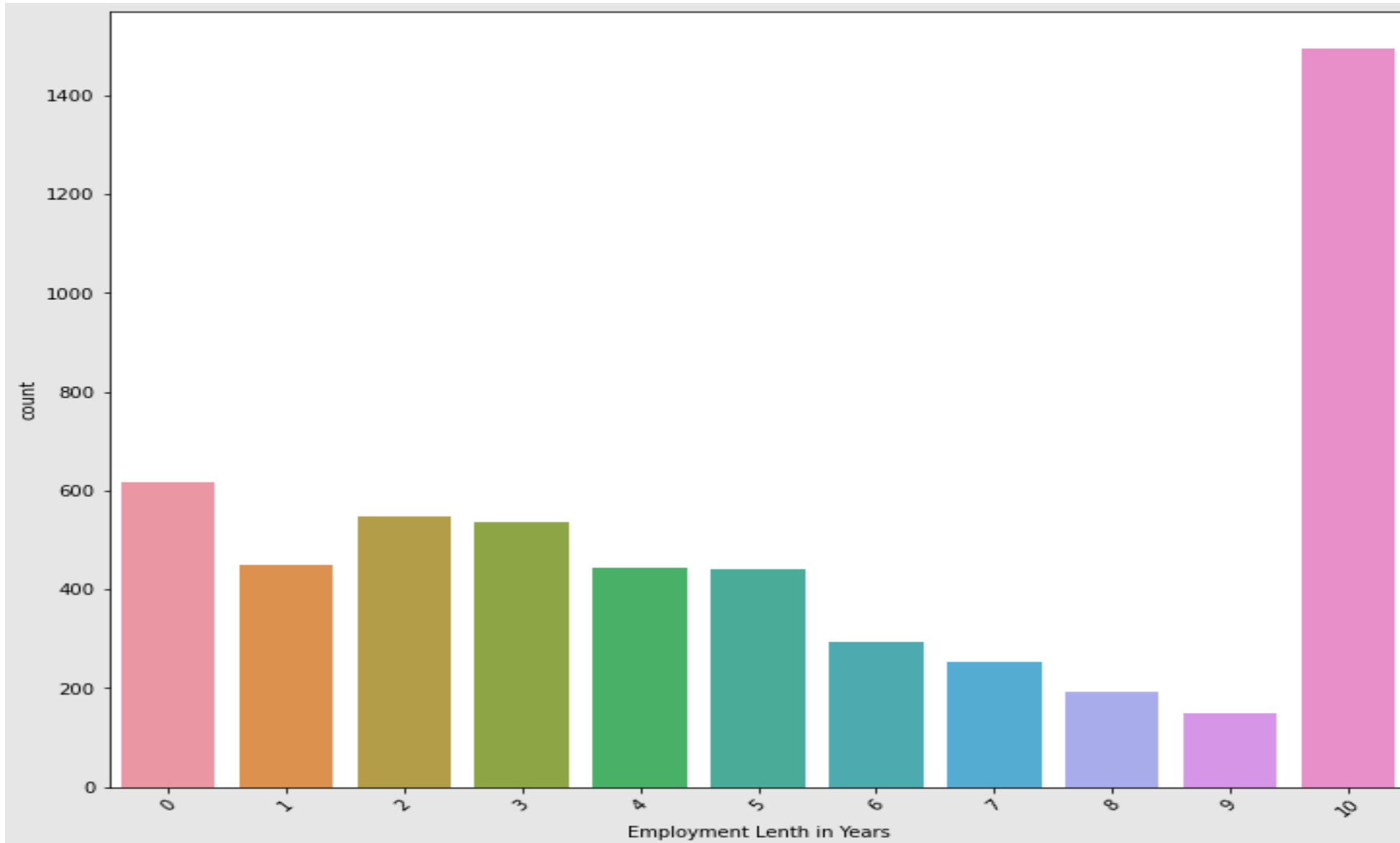
12. Applicants having loan term for 36 months have more probability of defaulting.

13. Applicants with interest rate 11.7% - 15% have more probability of defaulting.

14. Applicants with employment length of '10 or more' have more probability of defaulting.
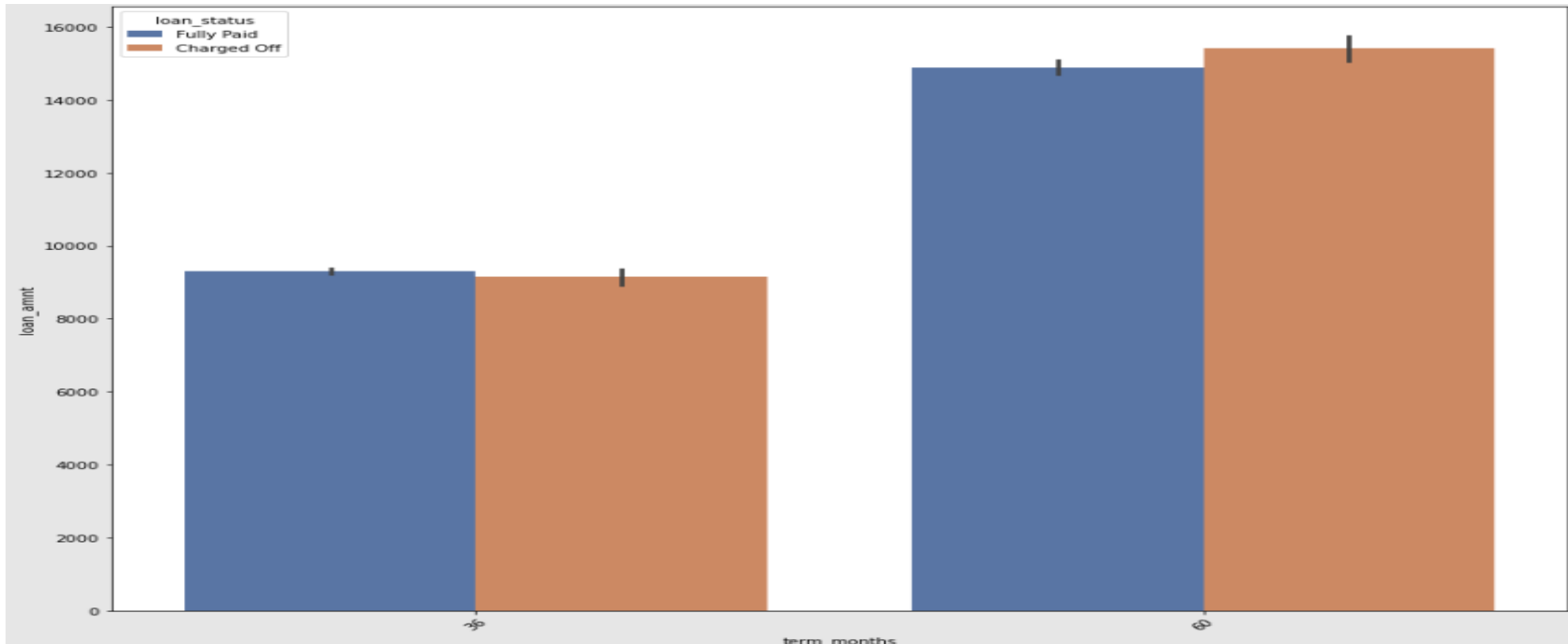
BIVARIATE ANALYSIS

Analysing the relationship between loan amount with other variables to get the driving factors.

**Driving factors derived from the bivariate analysis:**

1. Applicants who have taken loan amount in the range 14k-16k with loan term as 36 months have more probability of defaulting.
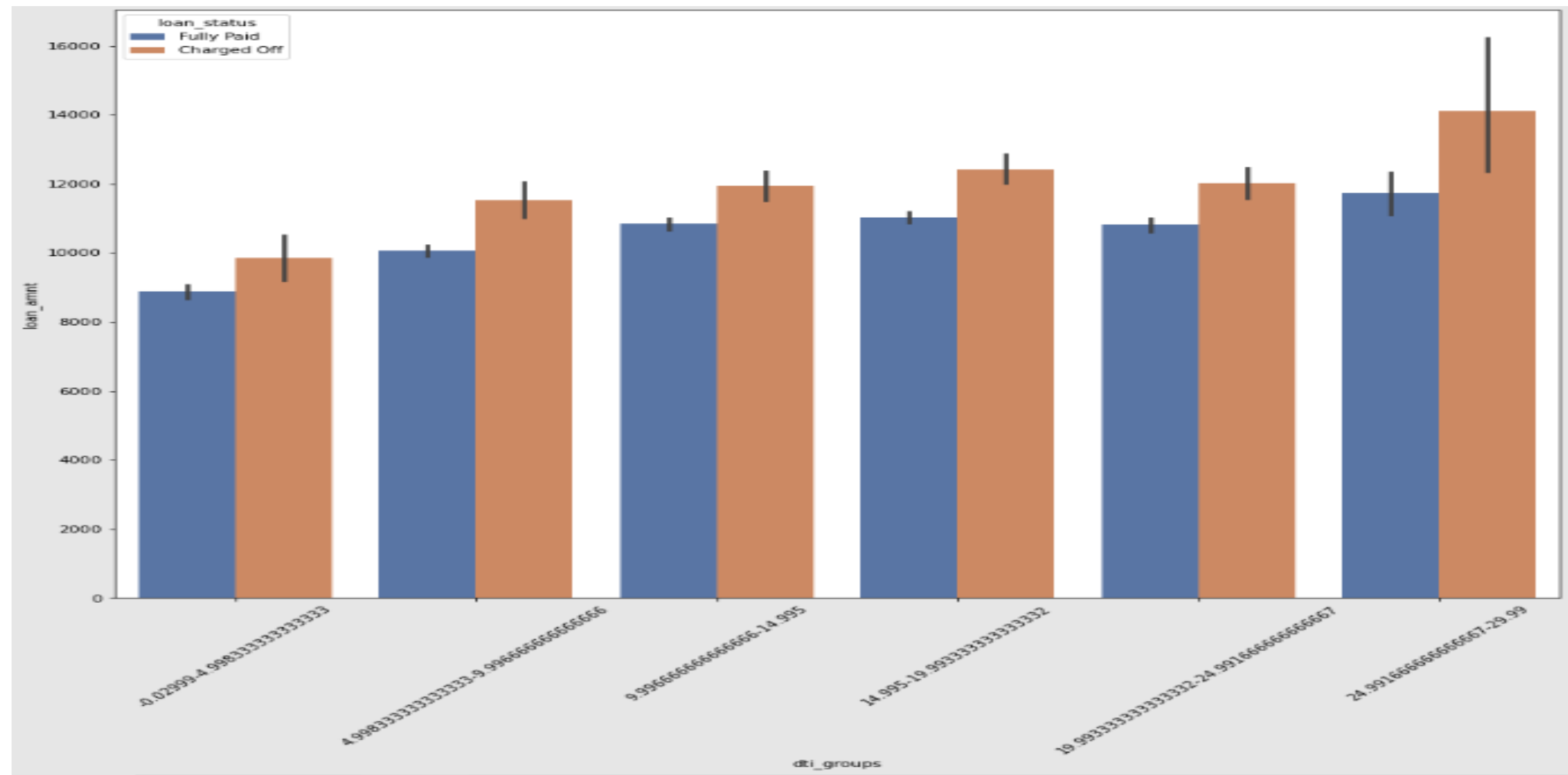
2. Applicants who have taken loan amount in the range of 30k-35k with installment range of 1176-1305 have more probability of defaulting.
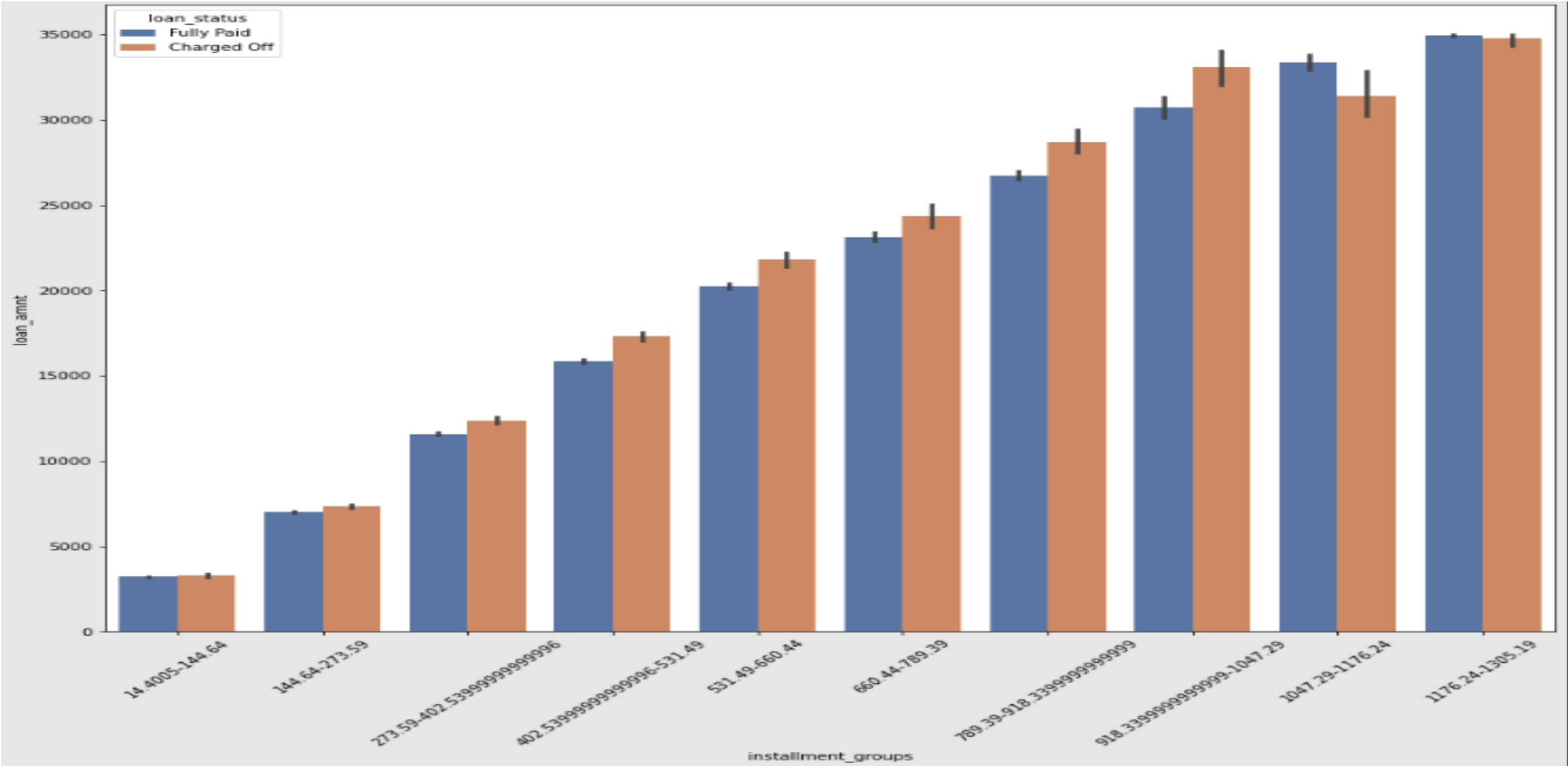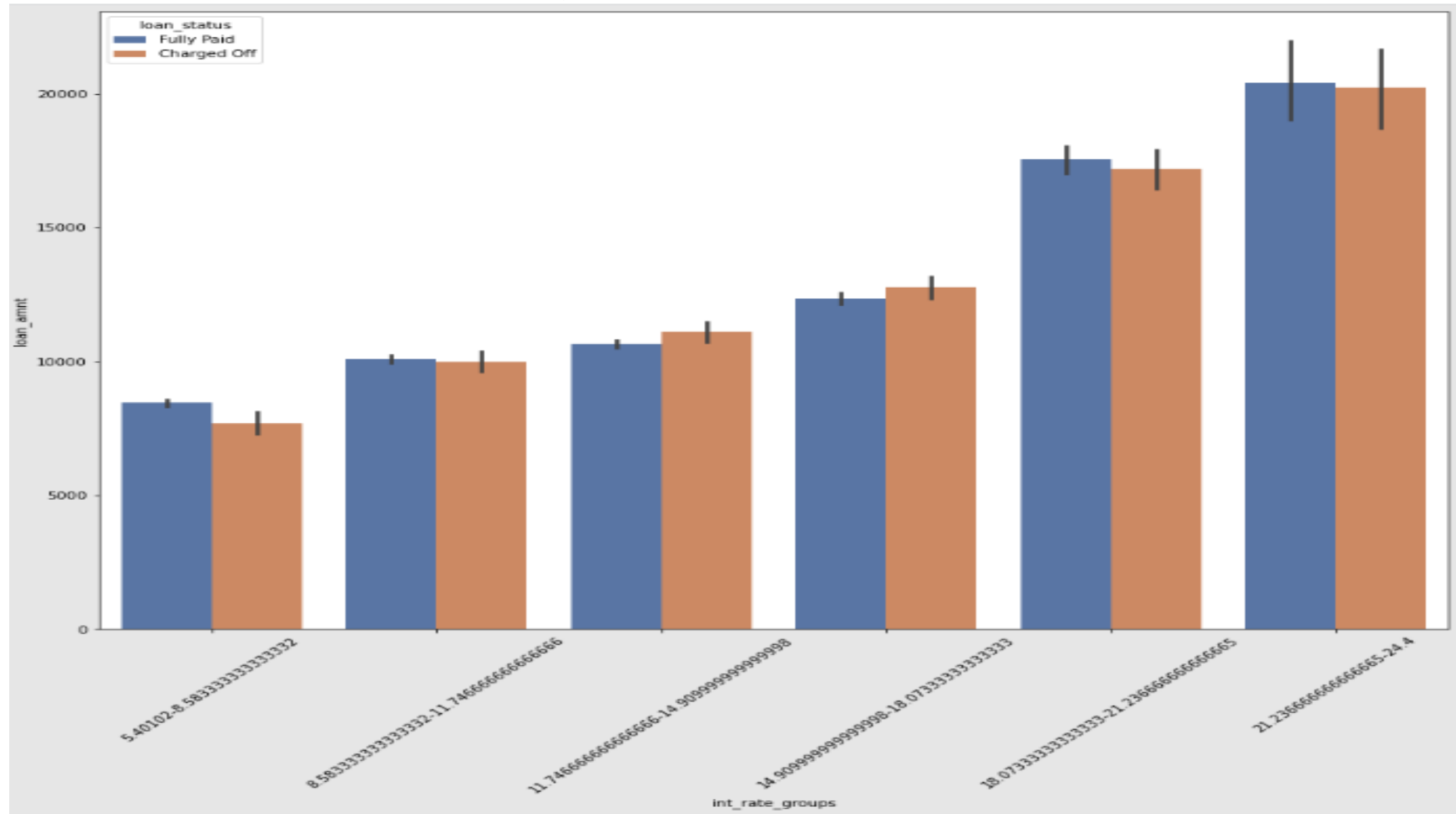
3. Applicants who have taken loan amount in the range 15k-20k with annual income range of 112k-126k have more probability of defaulting.
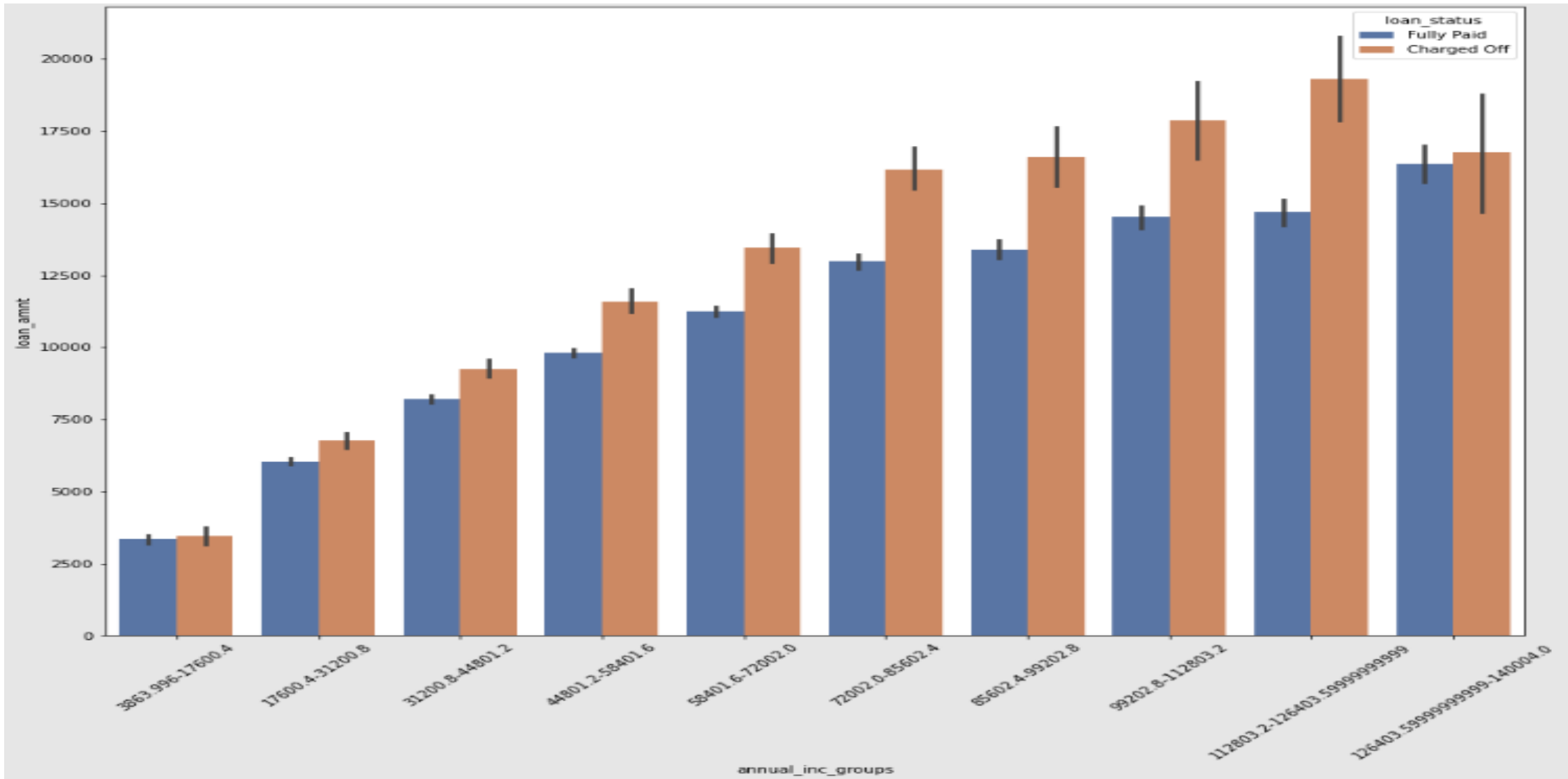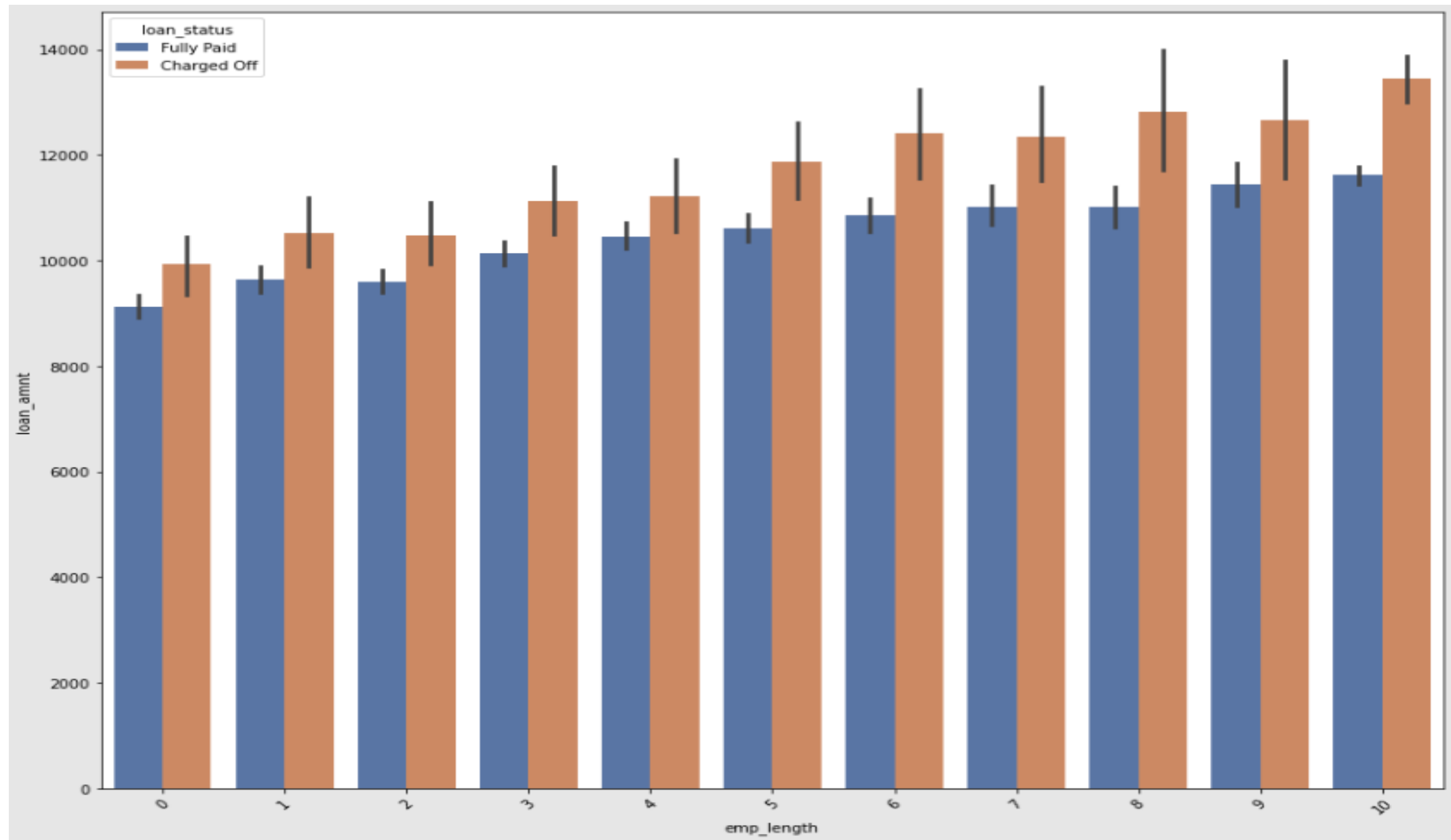
4. Applicants who have taken loan amount in the range 12k-14k with dti range of 25-30 have more probability of defaulting.
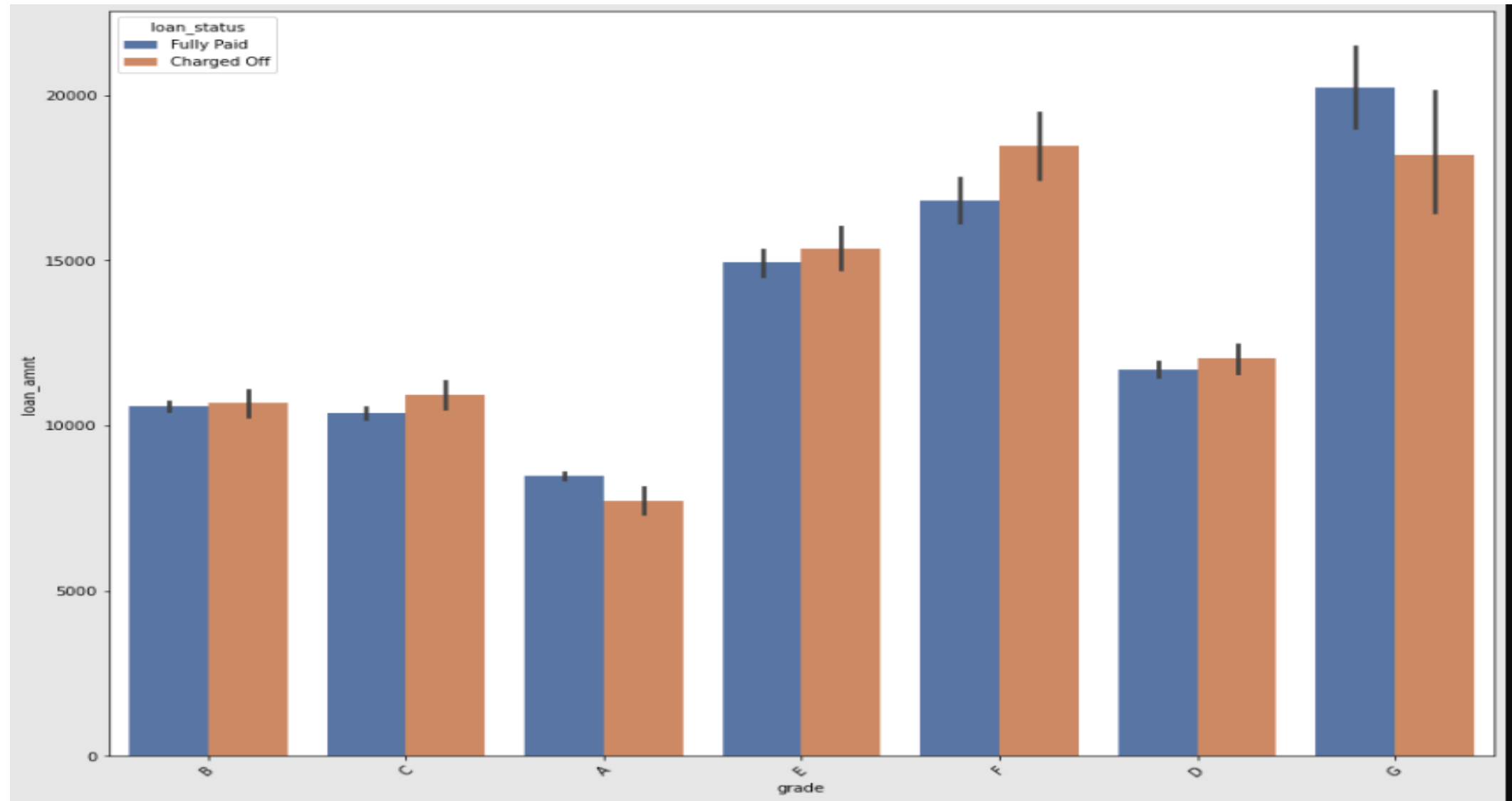
5. Applicants who have taken loan amount in the range 17k-23k with interest rate between 21-25% have more probability of defaulting.
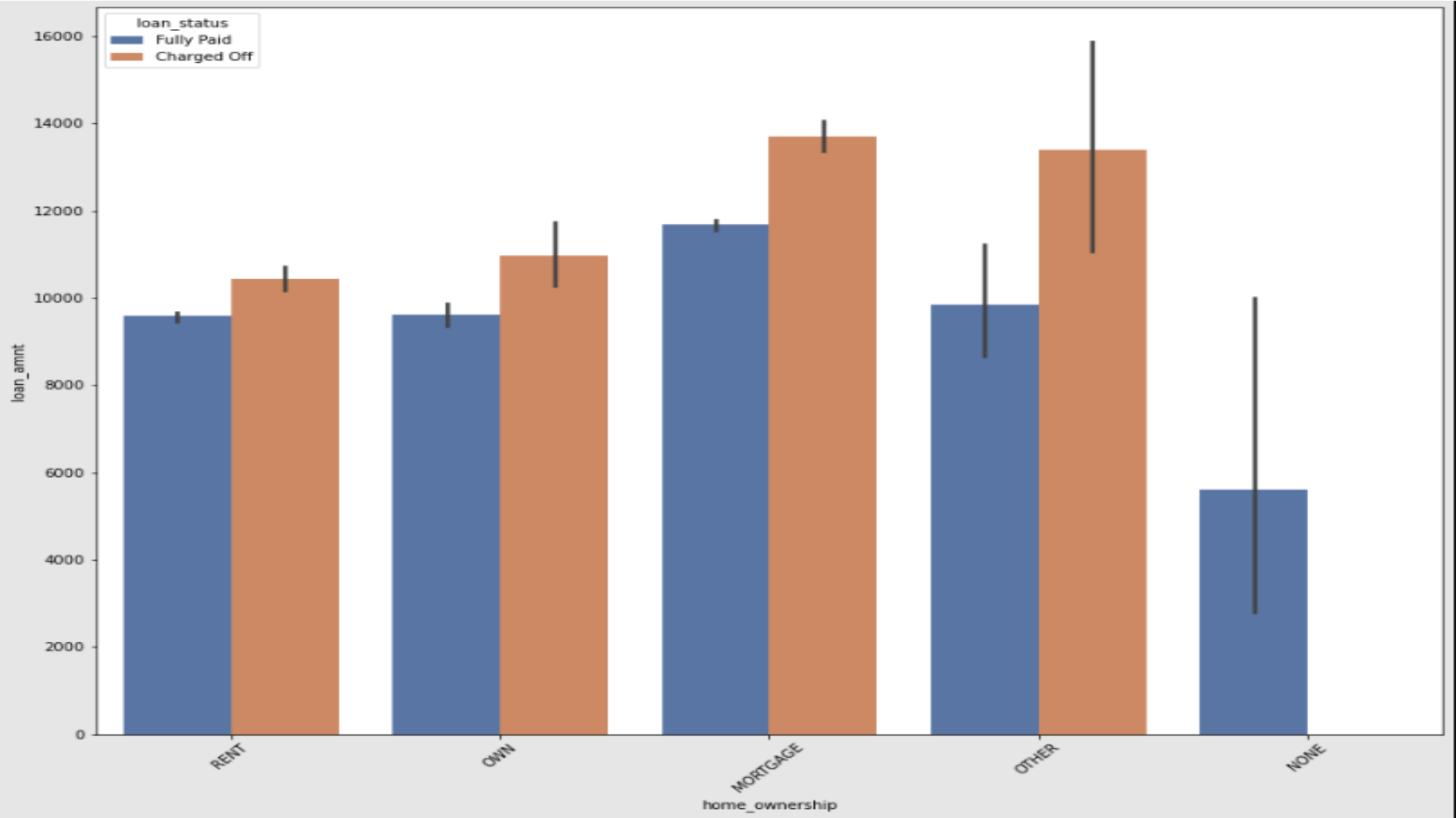
6. Applicants who have taken loan amount in the range 12k-14k with employment length as 10 years or more have more probability of defaulting.
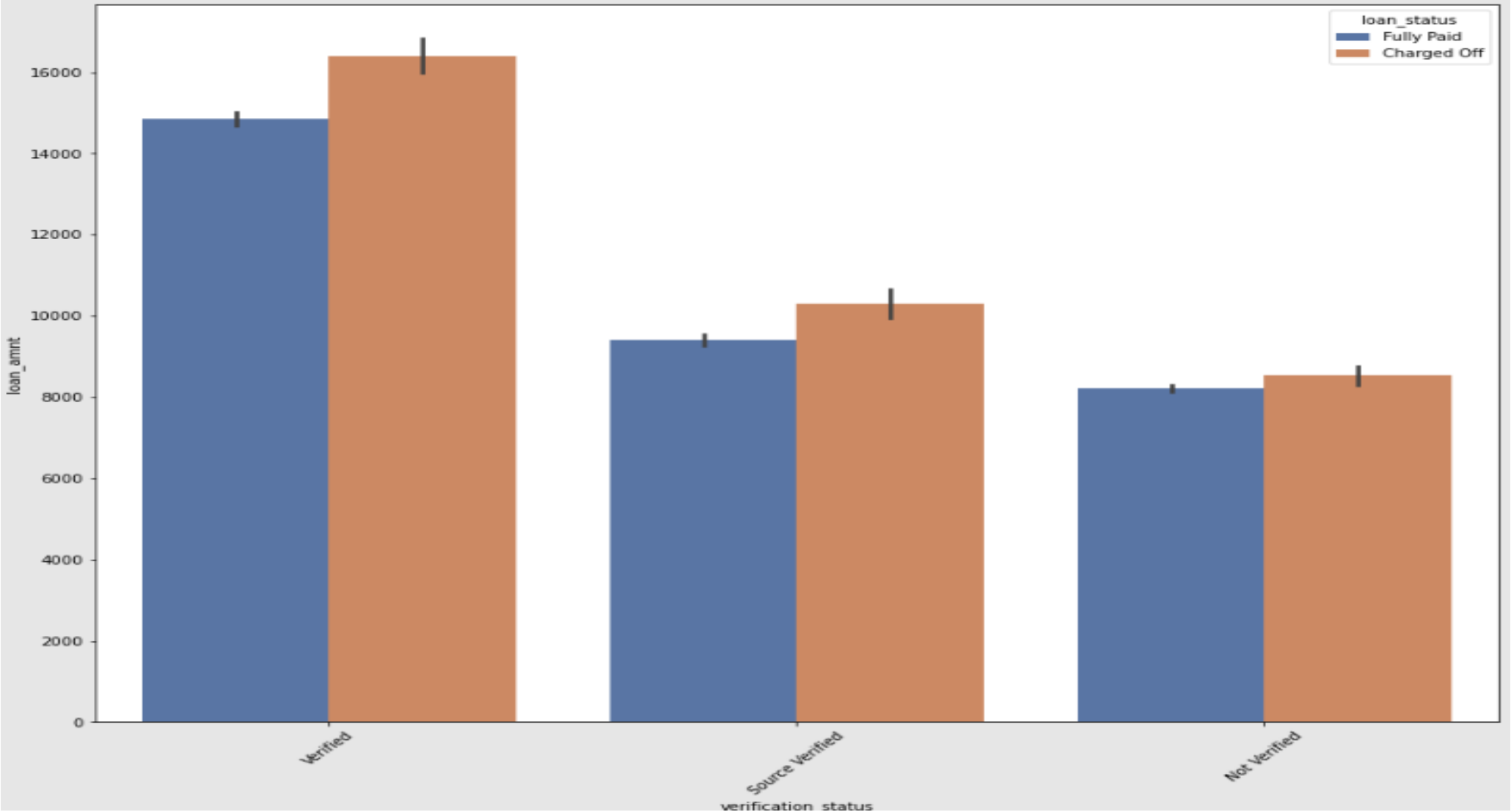
7. Applicants who have taken loan amount in the range 17k-20k with grade as F have more probability of defaulting.
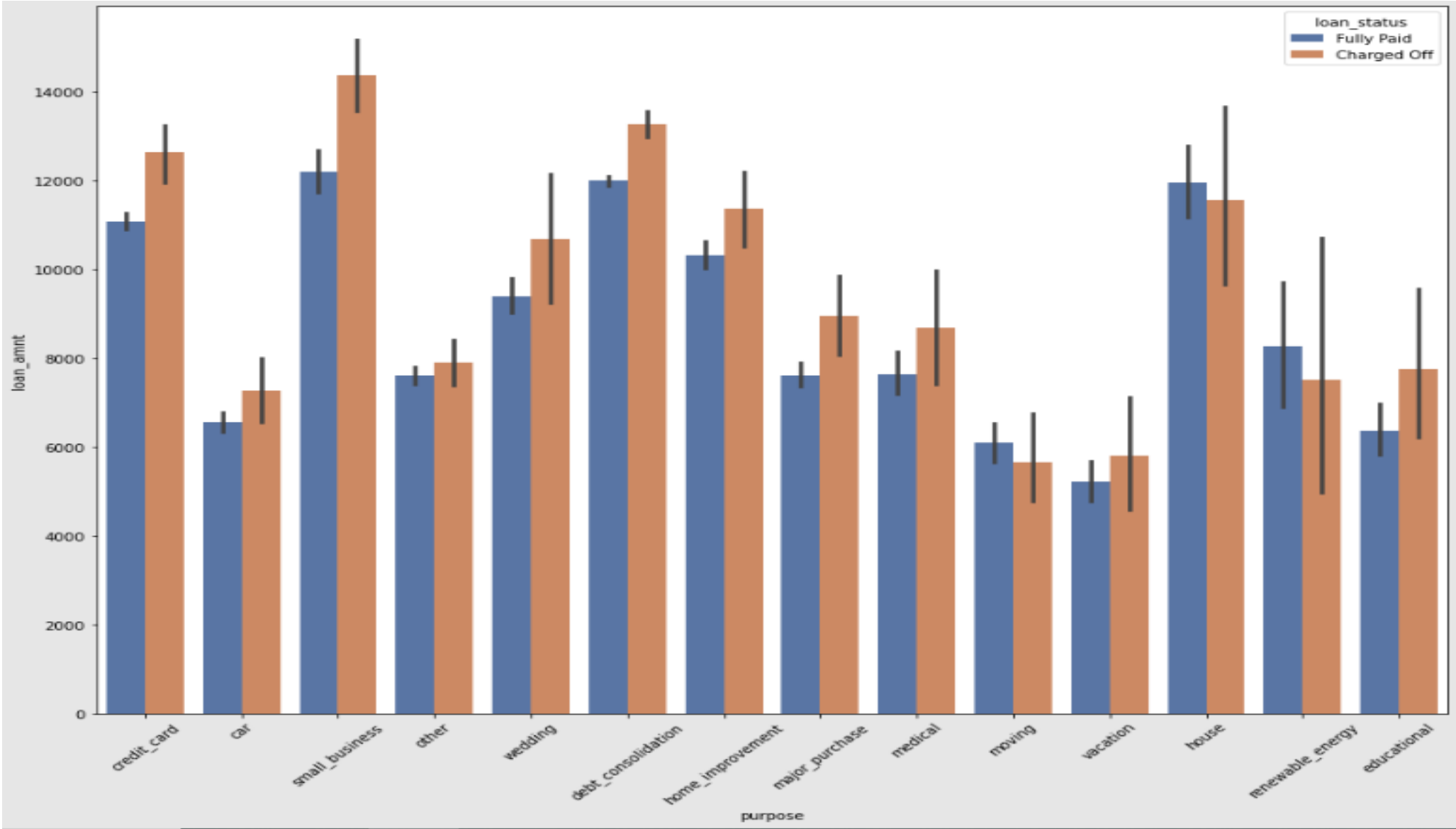
8. Applicants who have taken loan amount in the range 12k-14k with home ownership as mortgage have more probability of defaulting.
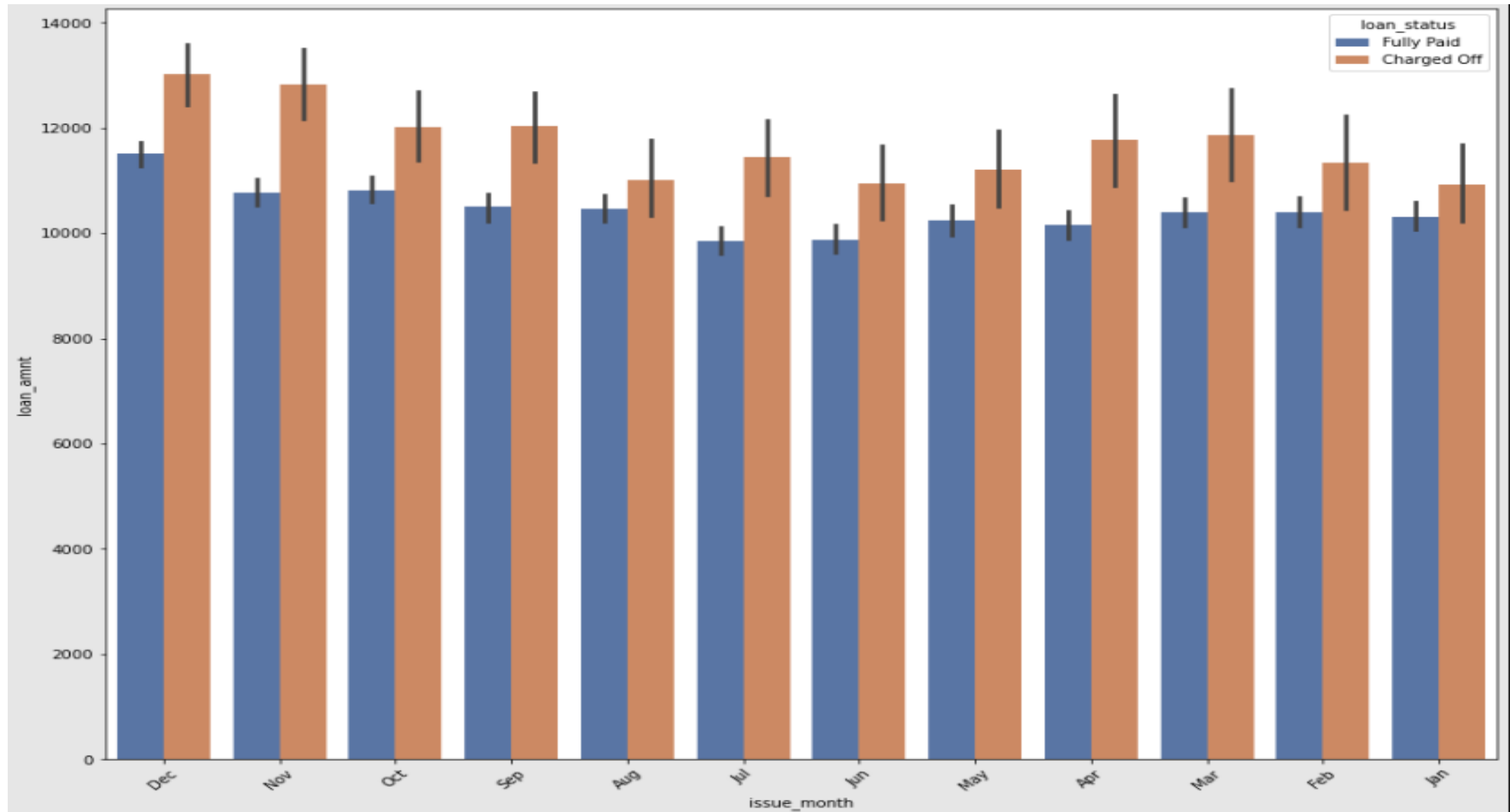
9. Applicants who have taken loan amount in the range 15k-17k with verification status as verified have more probability of defaulting.

10. Applicants who have taken loan amount in the range 12k-17k with purpose as small business have more probability of defaulting.

11. Applicants who have taken loan amount in the range 12k-14k with issue month of November and December have more probability of defaulting.

THANK YOU