

Lending Club Case Study

By

Shaheena & Lokesh

Problem Description

Consumer finance company which specialises in lending various types of loans to urban customers. When the company receives a loan application, the company must decide for loan approval based on the applicant's profile.

Two types of risks are associated with the bank's decision

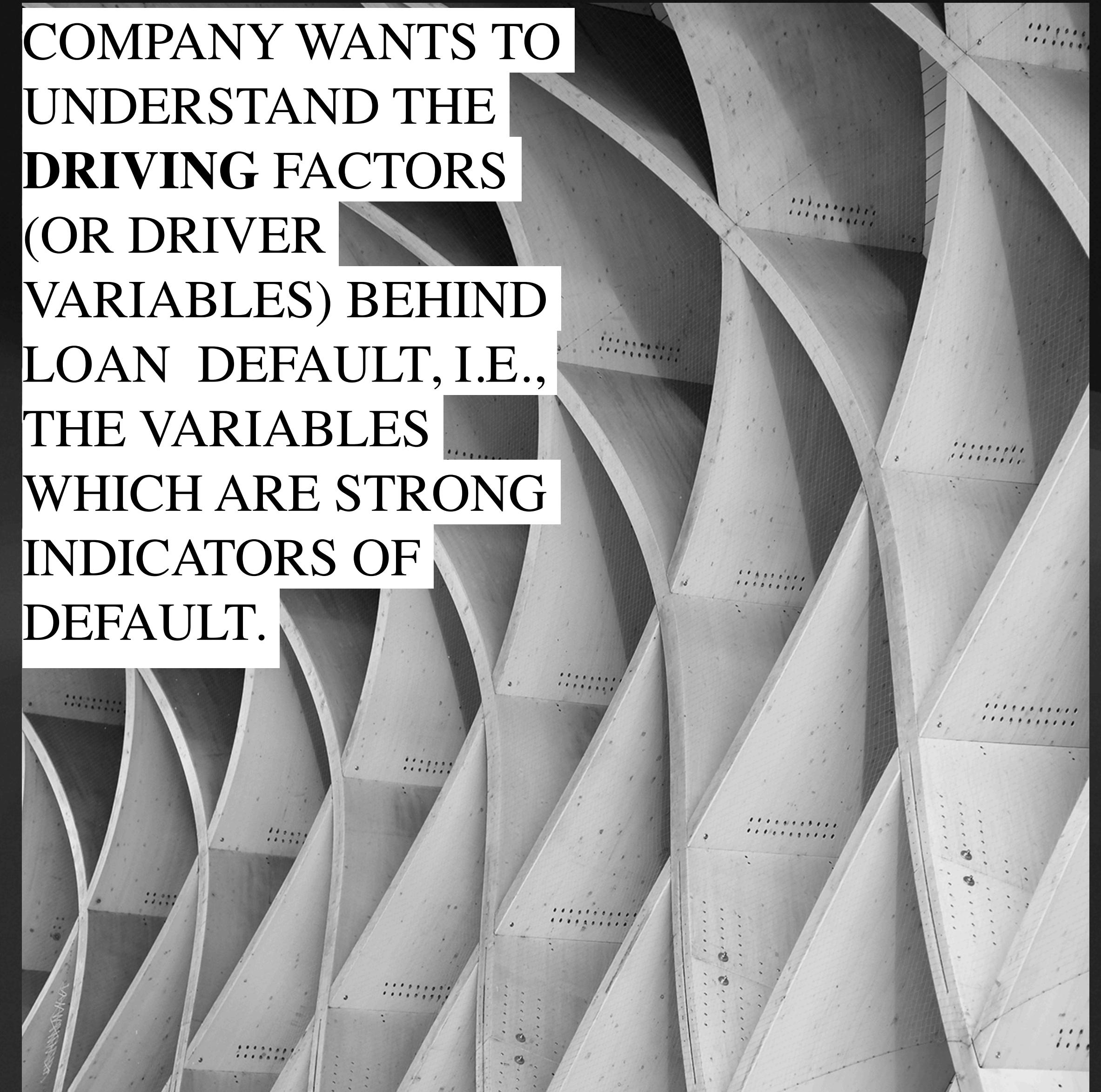
If the applicant is likely to repay the loan,
then not approving the loan results in a loss of business to the company

If the applicant is not likely to repay the loan,
i.e. he/she is likely to default, then approving the loan may lead to a financial loss fo the company

Objective Case Study



COMPANY WANTS TO UNDERSTAND THE DRIVING FACTORS (OR DRIVER VARIABLES) BEHIND LOAN DEFAULT, I.E., THE VARIABLES WHICH ARE STRONG INDICATORS OF DEFAULT.



Exploratory Data Analysis

Technologies Used :

- Python/Pandas/NumPy/Matplotlib
- Load.csv DataSet Provided .

Steps of Analysis

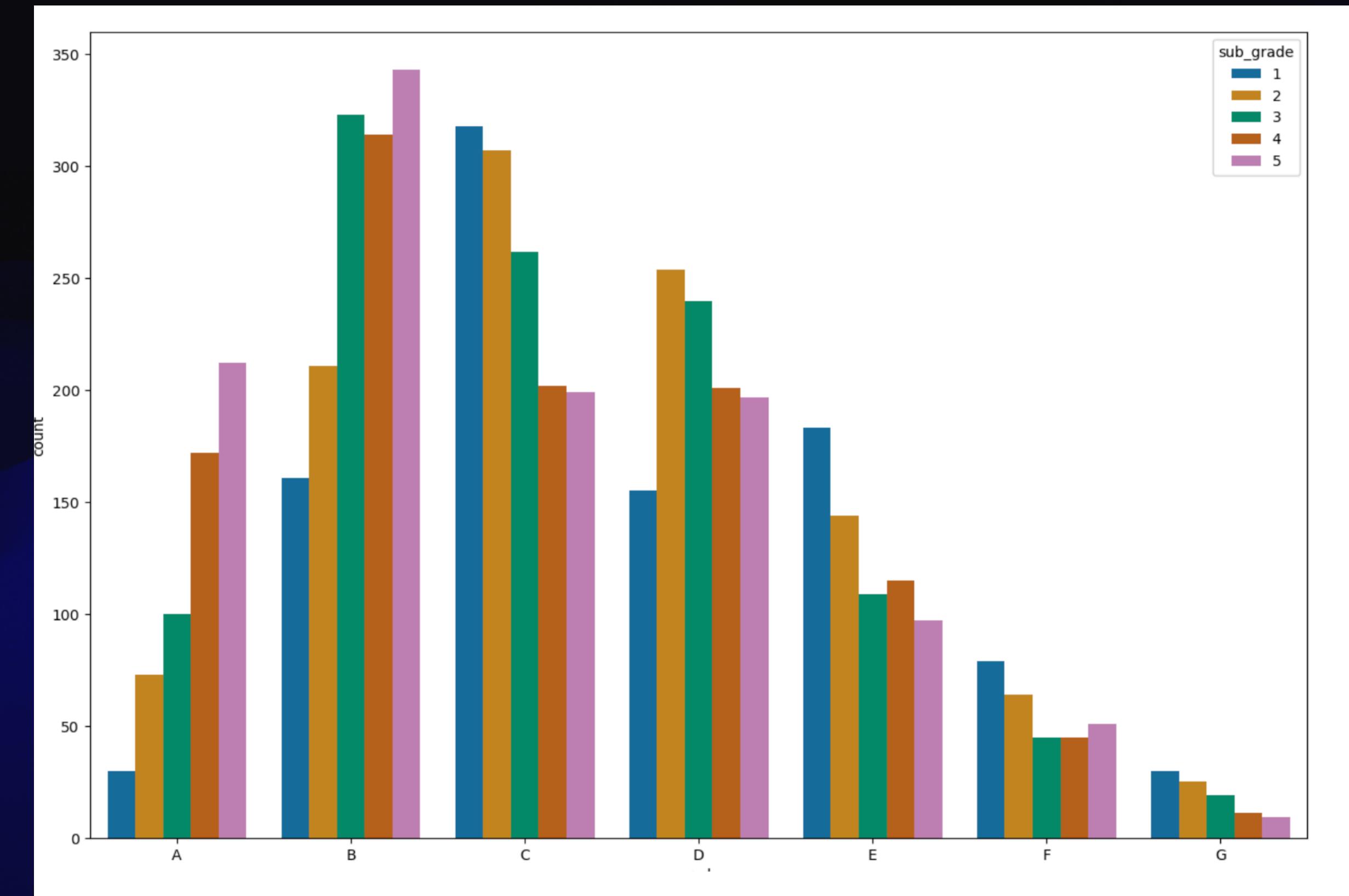
- Data Sourcing
- Data Cleaning
- Univariate Analysis
- Bivariate Analysis
- Derived Metrics

Observation

Term : Loan with term 36 months have higher Default percentage compared to 60 months.

Grade: Loan with 'B' grade have higher percentage of default

Sub Grade: Loan with Grade and subgrade 'B5' has higher percentage of default



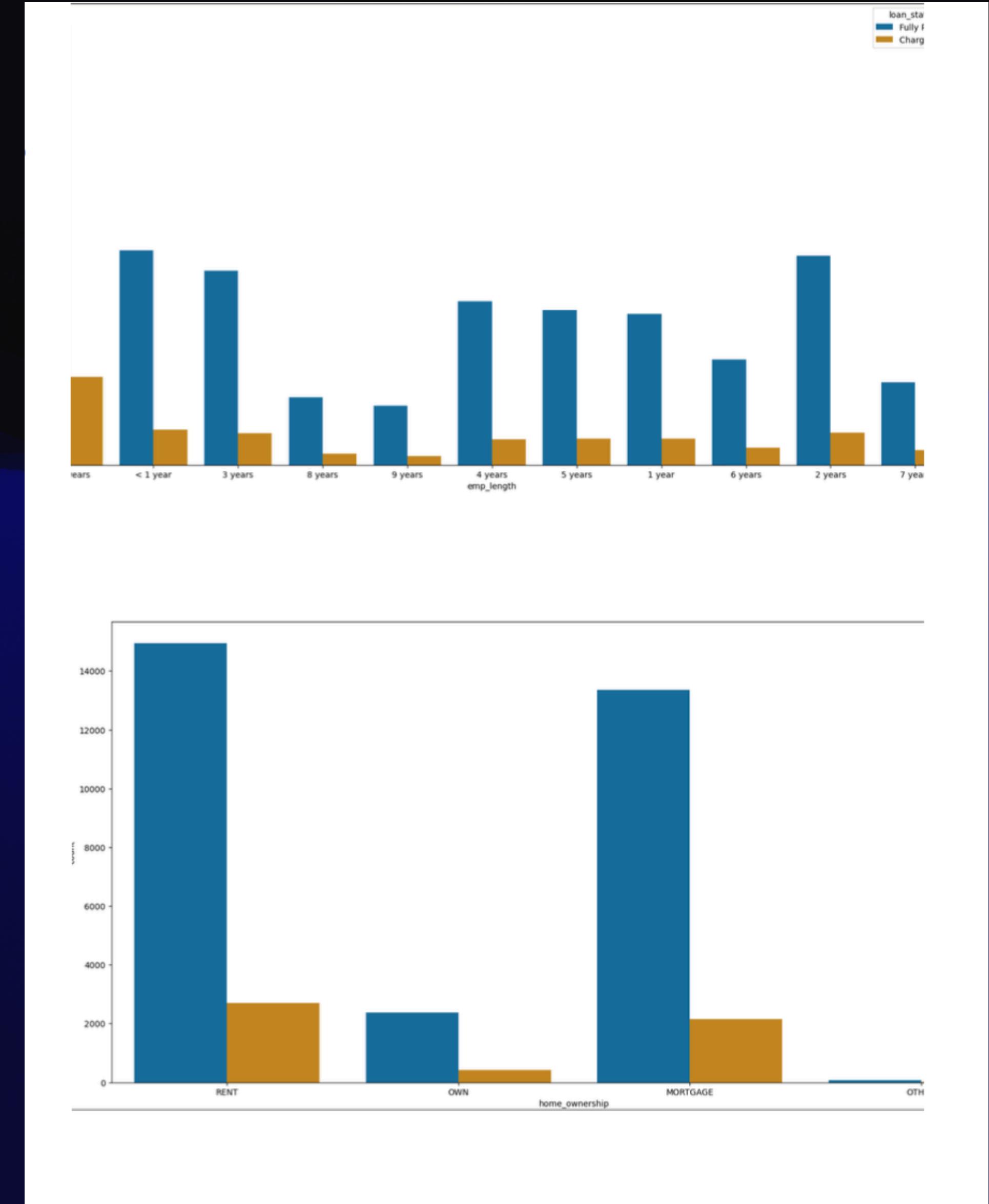
Observation

Employee Length: Employees with experience more than 10 years have higher percentage of default. This might be because As experience increases his/her age increases and other responsibilities might increase might result in default.

Home Ownership: People whole have a rented house defaulted more when compared to people who have Mortgaged their property or having own house.

This might be because people living in rent do not have much to lose. Others have their Home lost as part of recovery process

Verification Status: People whose incomes have been verified are very less prone to default. As their incomes are verified that means the values, they quoted are correct and decision we have taken based on their income levels will be accurate.

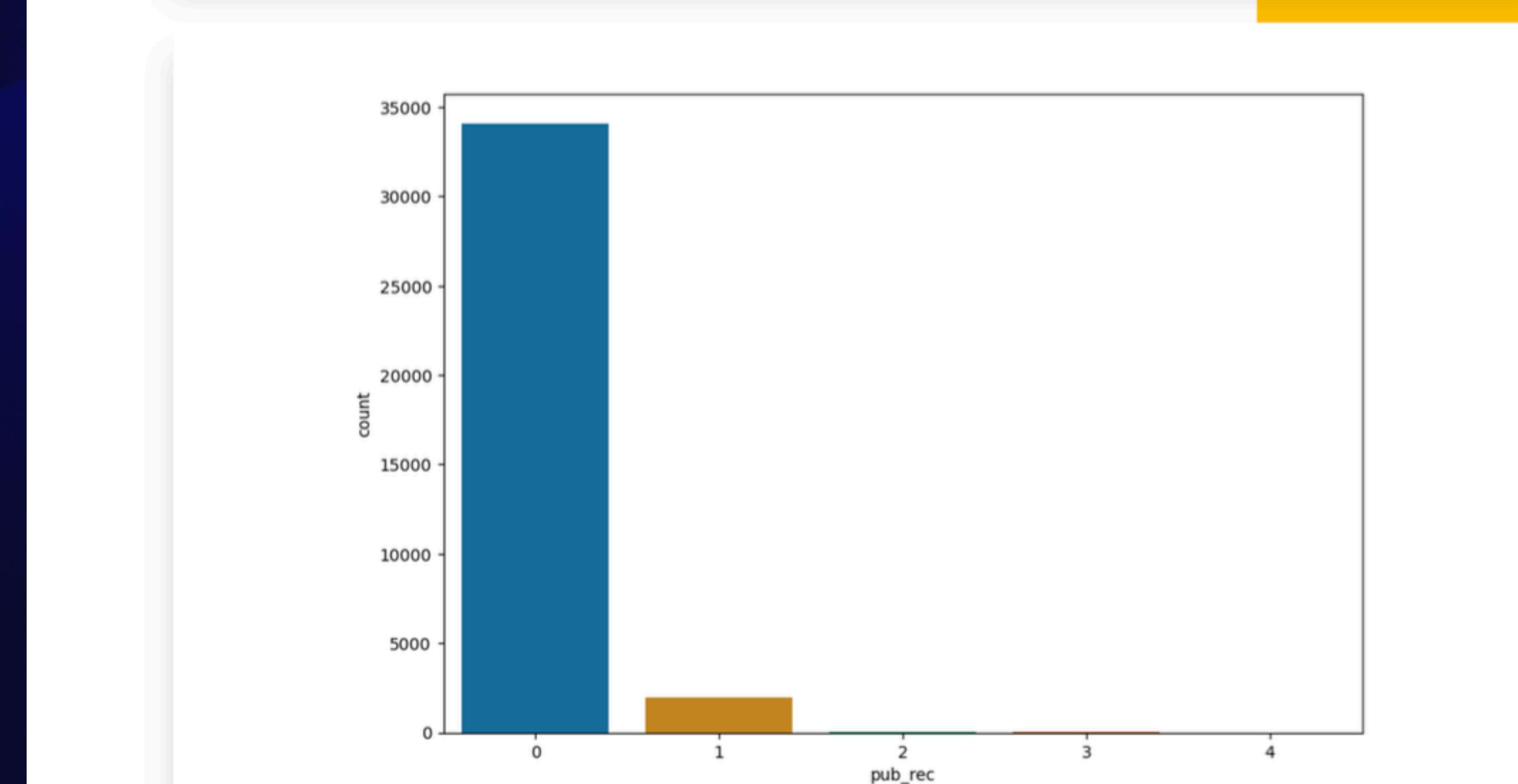
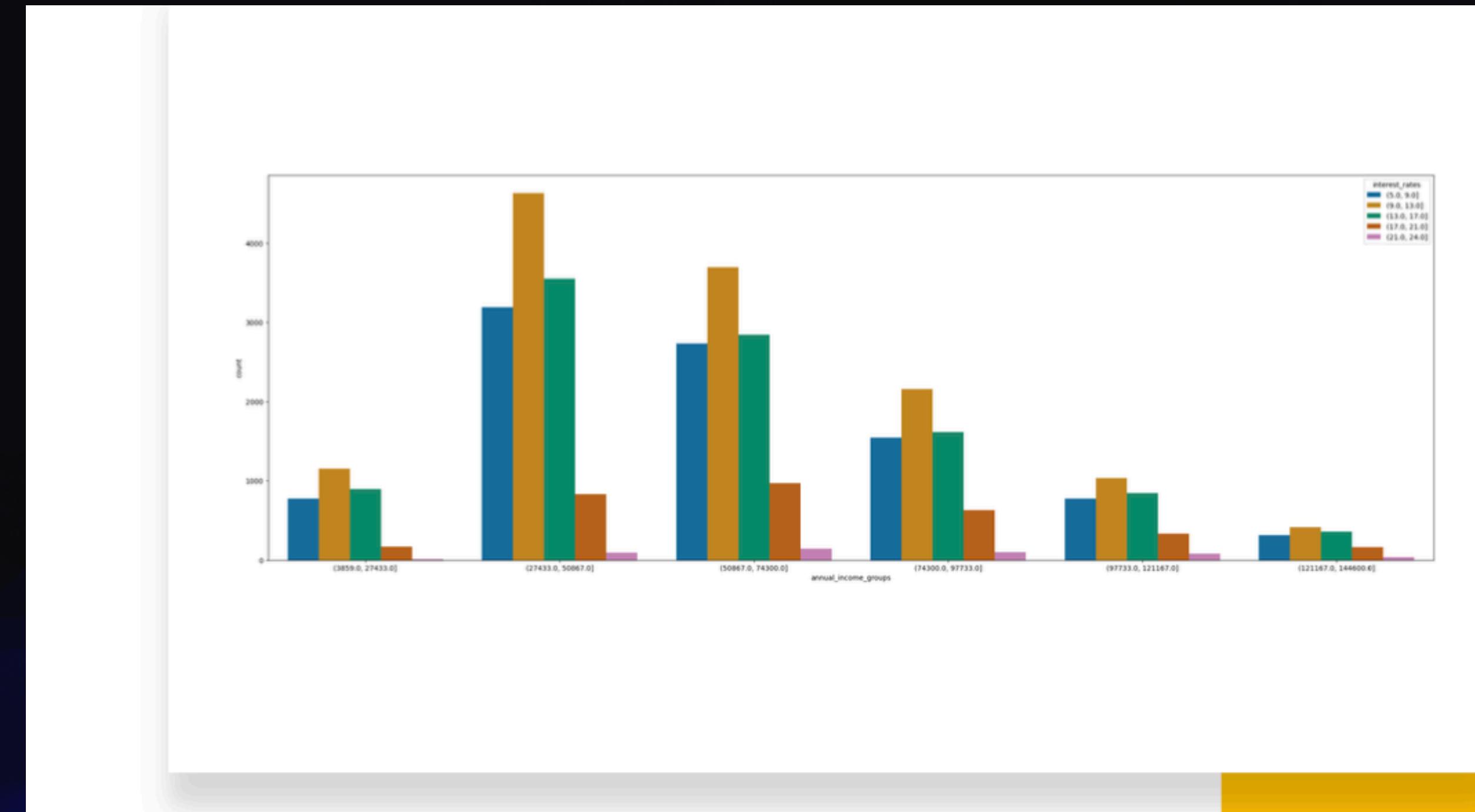


Observation

annual_income: People with annual income in range 27k-51k are prone to high default. From the derived metrics analysis of Monthly income to the installments we can observe - in this income group high percentage of the monthly income is going in installments which might be the reason for the high defaulters.

pub_rec: People with 0 public records are prone to higher number of defaulters.

revol_util: People with 40-60 % of revol_util are higher prone to Default



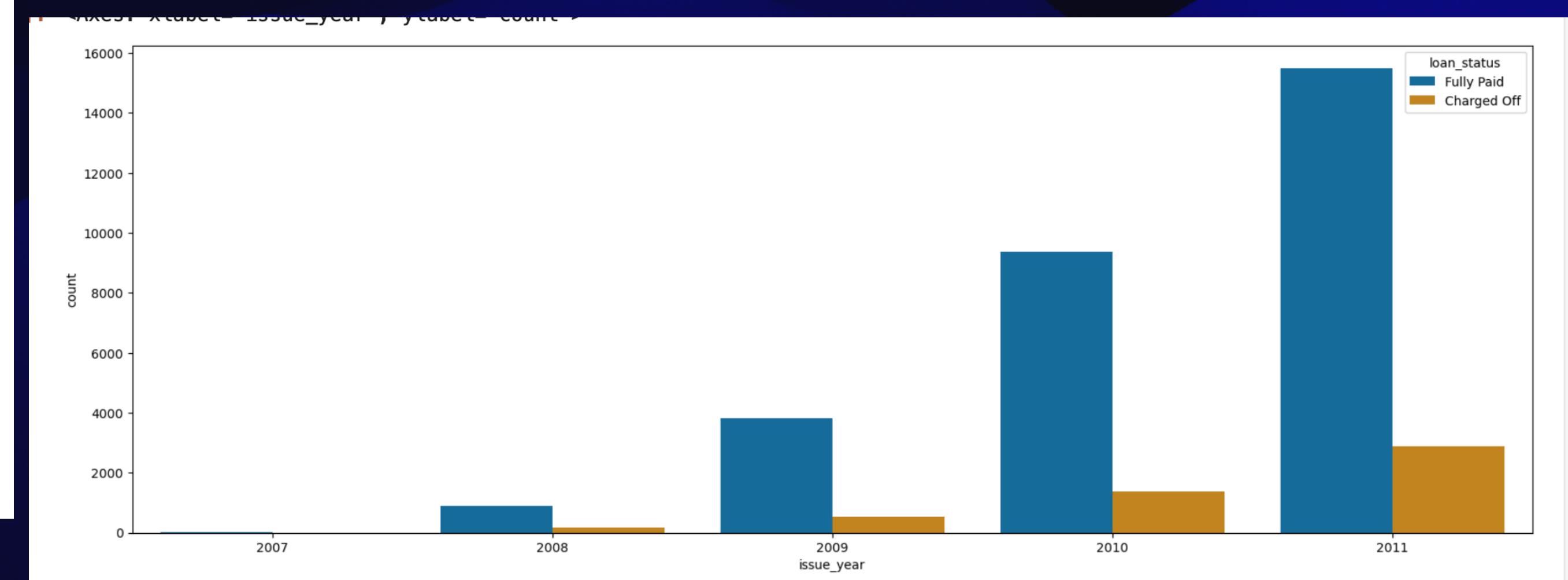
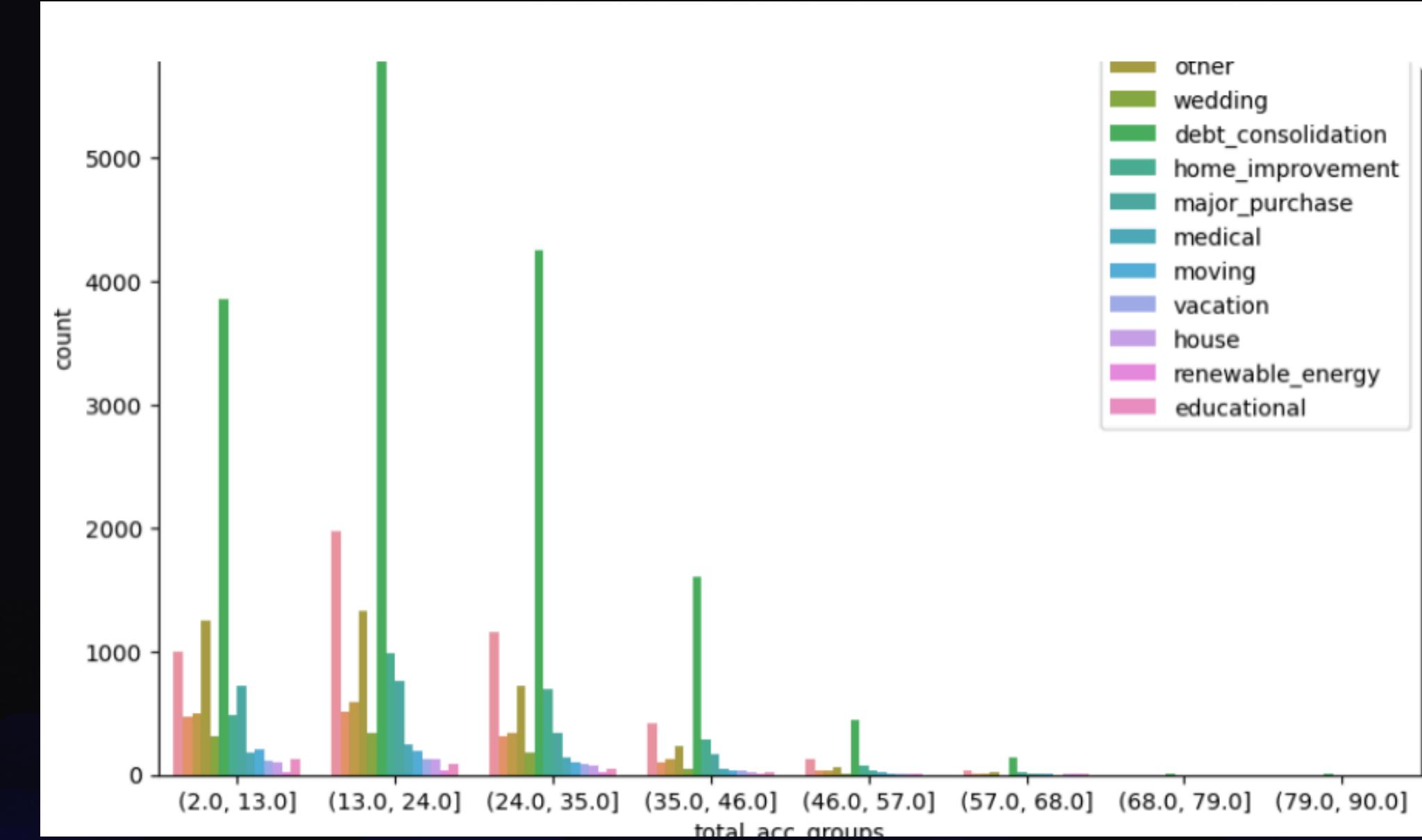
Observation

Purpose: People who takes loan for the purpose of debt consolidation are more prone to default.

People who have more accounts opened are taking more loans/accounts to close other loans. In below figure we can observe the same.

Loans taken in the year 2011 are having higher defaulters. We can analyze more on this phenomenon with supporting data to see the inflation rates and other metrics which could be the reason for the same.

Loans taken in the month of December are having higher defaults.



Thank You