Leading club Case study

By

Rajneesh and Rishi

Problem Statement

- The problem statement involves analyzing data from a consumer finance company specializing in lending various types of loans to urban customers. The company needs to make informed decisions regarding loan approvals based on applicants' profiles, considering two types of risks:

1. **Risk of Business Loss**: If a likely-to-repay applicant is not approved, it results in lost business opportunities for the company.
2. **Risk of Financial Loss**: If an applicant is likely to default (not repay the loan), approving the loan could lead to financial losses for the company.

Objective

- The objective is to uncover patterns and insights using Exploratory Data Analysis (EDA) to understand how various consumer and loan attributes influence the likelihood of default.

Key Points to Consider

- **Loan Decision Outcomes**: When a person applies for a loan, the company can either accept or reject the application.**Loan Accepted**: If approved, there are three possible outcomes:
  - **Fully Paid**: Applicant has fully repaid the loan amount along with interest.
  - **Current**: Applicant is currently paying off the loan, with no defaults recorded.
  - **Charged-off**: Applicant has failed to make payments for an extended period, resulting in default.
- **Loan Rejected**: Data for rejected applicants are not available since there is no transactional history with the company.

The goals of the analysis include:

- Identifying factors that correlate with loan default rates.
- Developing strategies such as denying loans, reducing loan amounts, or adjusting interest rates for risky applicants based on identified patterns.

Data Quality check and cleaning

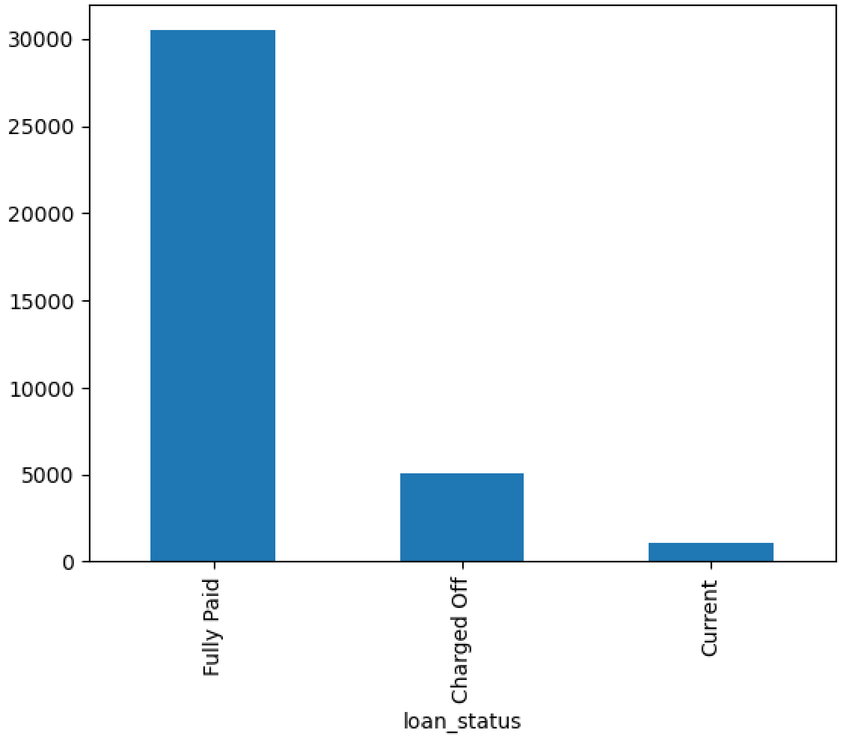
- Available data set named loan.csv is used to perform analysis of the problem statement and reach to objective
- Before starting we need to clean the dataset via checking columns with null values or Nan records
- There were total 111 columns which was reduced to 57 columns after performing this cleanup
- Removed columns which do not play any role in the analysis
- There are columns like emp_title and emp_length which seems to have empty records so clean those as well to remove around 6% and 2% empty data set
- Convert employment length to integer values which can be used as extracted parameter as number of years of service

Parameters for analysis

- term - The number of payments on the loan. Values are in months and can be either 36 or 60.
- int_rate - Interest Rate on the loan
- installment - monthly EMI
- grade and sub_grade - LC assigned loan grade
- emp_title - he job title supplied by the Borrower when applying for the loan.*
- emp_length -Employment length in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
- home_ownership - The homeownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER
- annual_inc - The self-reported annual income provided by the borrower during registration.
- loan_status Current status of the loan
- loan_amnt - The listed amount of the loan applied for by the borrower. If at some point in time, the credit department -educes the loan amount, then it will be reflected in this value.
- verification_status- Indicates if income was verified by LC, not verified, or if the income source was verifie-
- purpose - A category provided by the borrower for the loan request.-
- DTI - A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations, excluding -mortgage and the requested LC loan, divided by the borrower’s self-re-orted monthly income.
- delinq_2yrs - The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years

Extract parameters

- Employment length can be denoted as number of years of service
- Add loan term column convert to number 60 months 5 years and 36months as 3 years
- Add loan amount type as categorical value as low, medium, high and very high
- Add interest rate type as low, medium and high
- Add column for defaulters
- Debt to income ration DTI
  - - Excellent: DTI below 20%.
  - - Good: DTI between 20% and 35%.
  - - Fair: DTI between 35% and 50%.
  - - Poor: DTI above 50%.
- Loan Grades based on
  - A: Represents the most favorable credit quality.
  - - B: Indicates good credit with moderate risk.
  - - C: Suggests average credit with some risk.
  - - D: Implies below-average credit with higher risk.
  - - E: Signifies high risk.
  - - F/G: Reserved for the riskiest borrowers

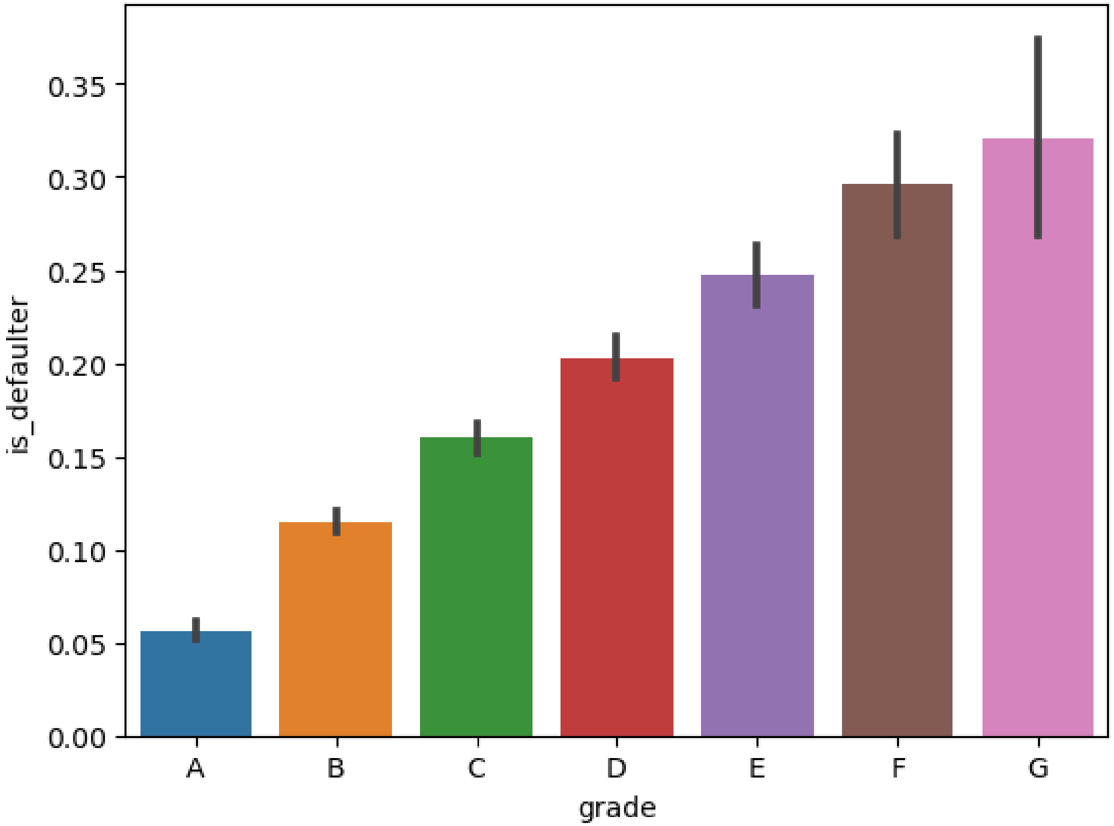


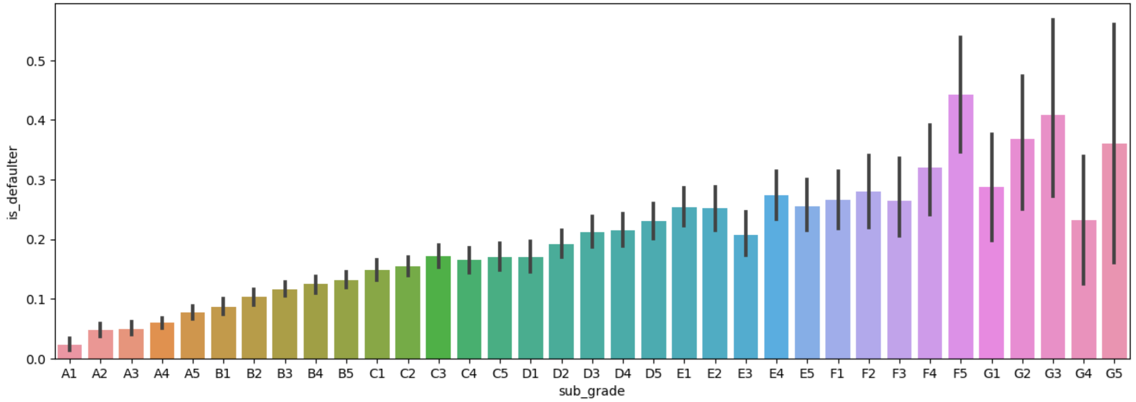
Observation matrix Loan Status – univariate

- Analysis of loan status as univariant analysis
- To be check defaulters we know that charged off so lets create a column where we can mark which all customers are defaulter for analysis
- Around 14 % of defaulter are present
- loan_status Fully Paid 0.833507 Charged Off 0.137326 Current 0.029167

Observation Grade and defaulters

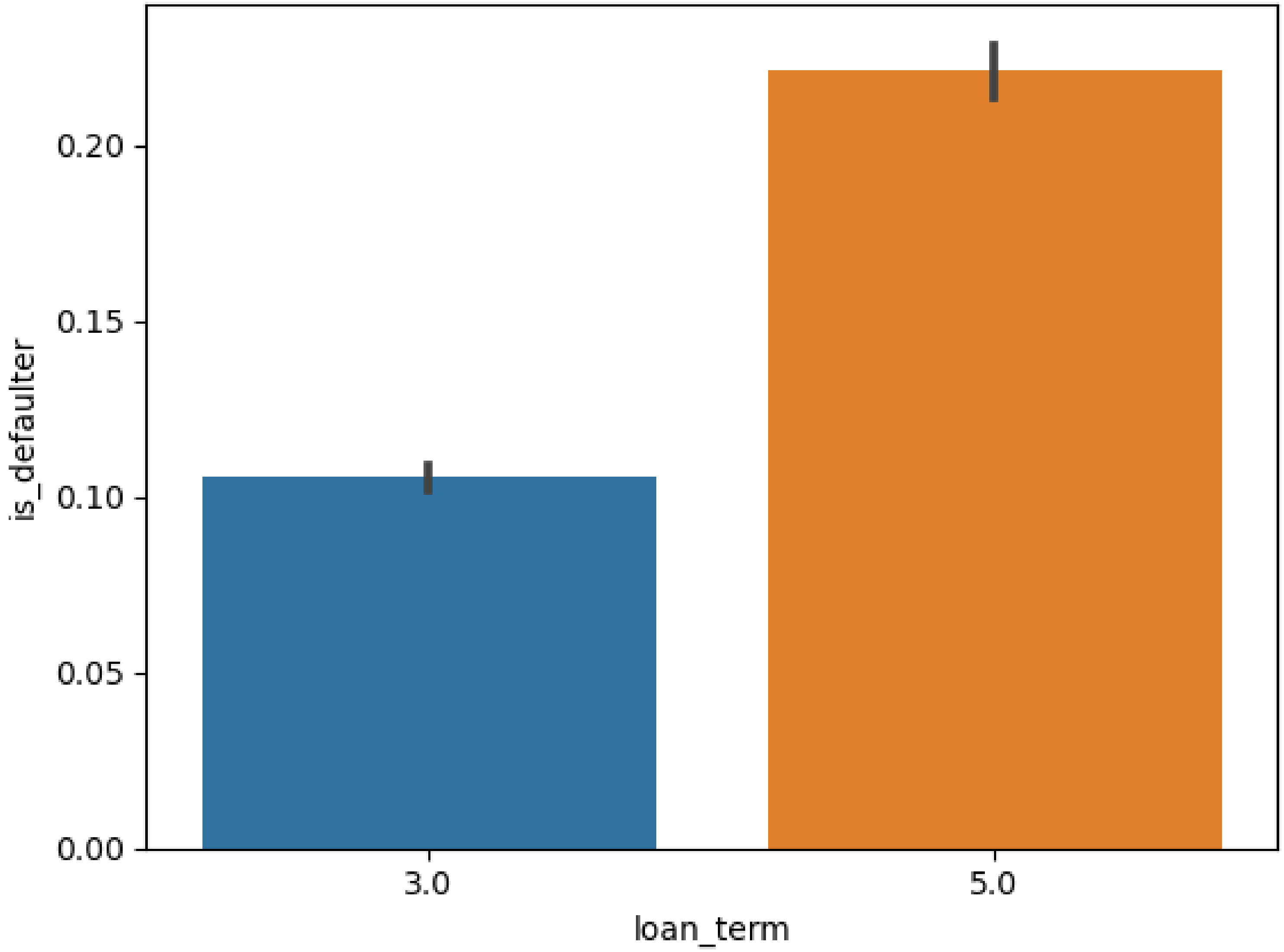
- As we can see that grades with higher risk have more defaulters.
- Grades F and G has more of the defaulters





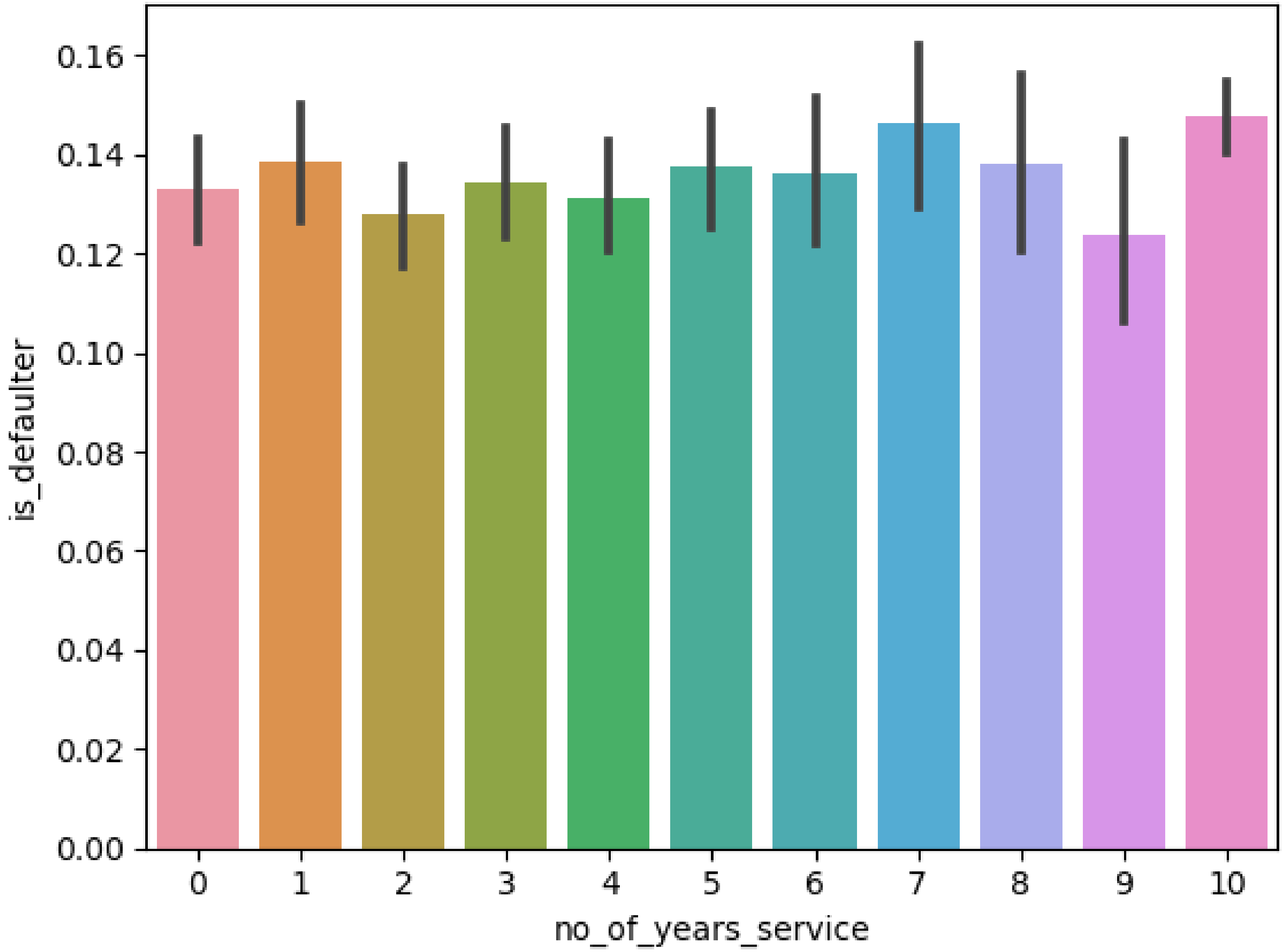
Observation on loan term and defaulters

- As we can see the longer the term loan the more defaulters we have



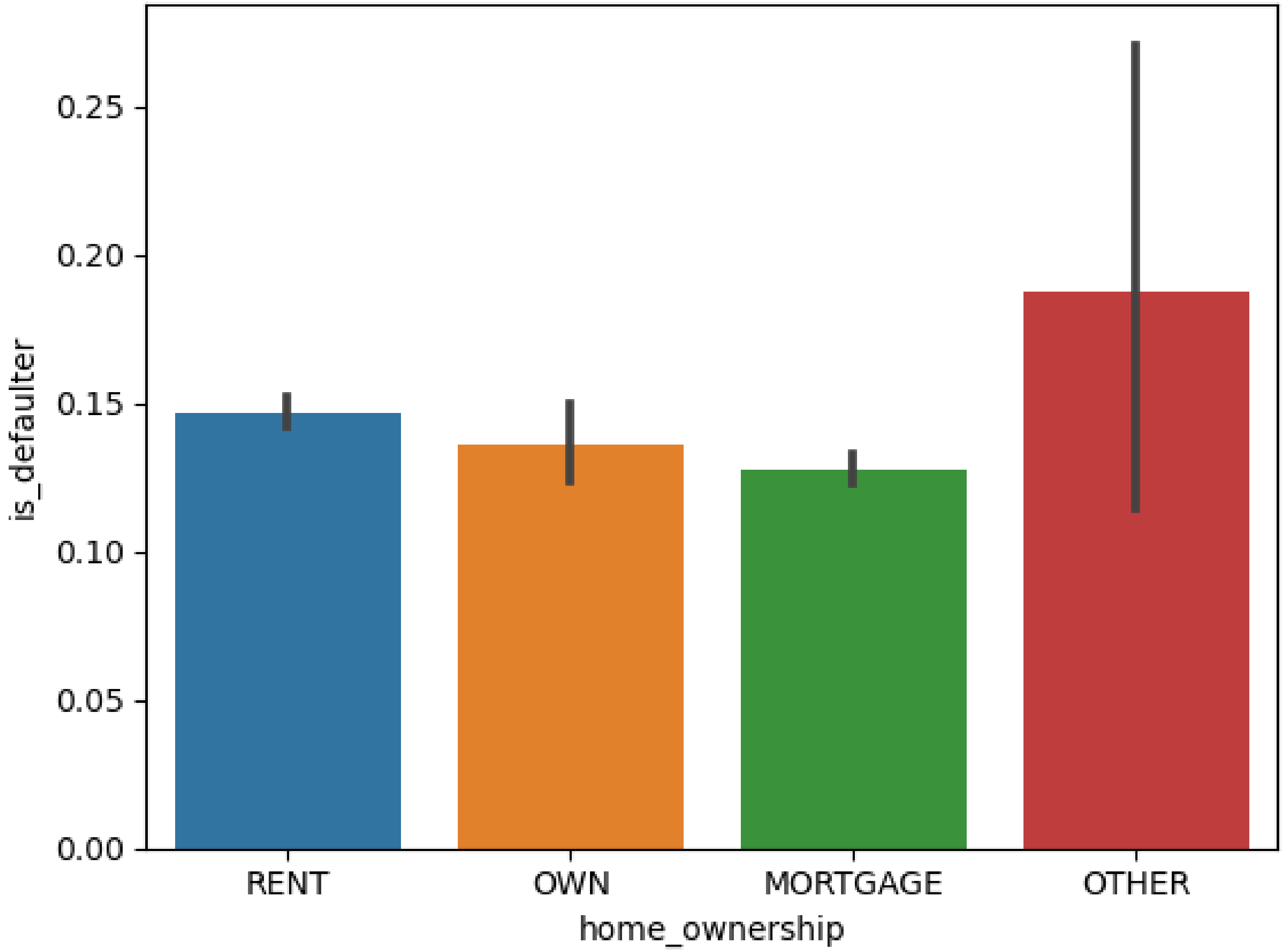
Observation based on number of years of service

- Looks like no of service does not influence on loan defaulters



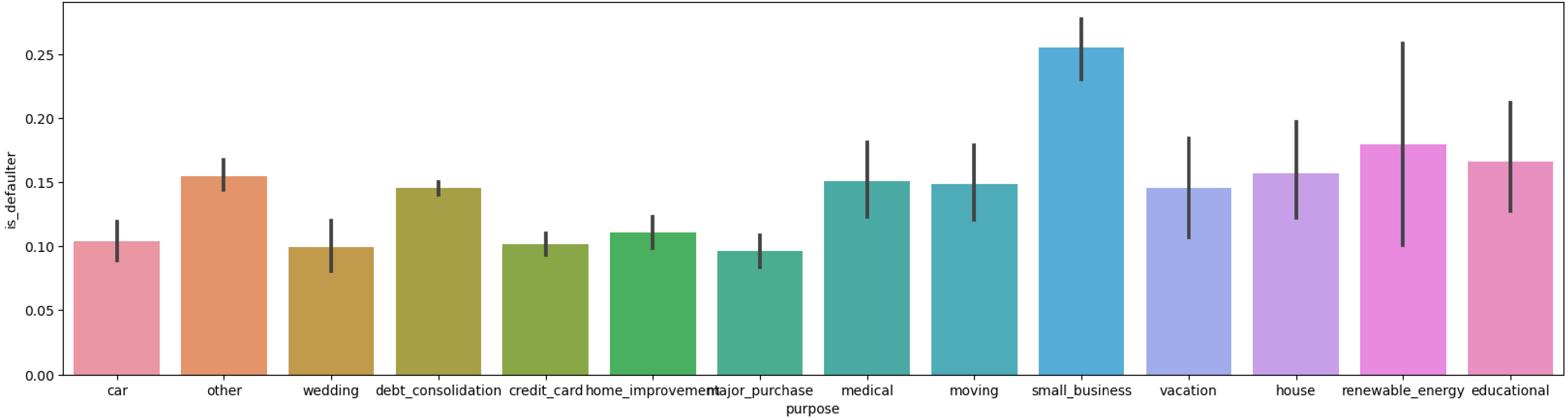
Observation on the basis of home ownership

- We can see that home ownership have impacted the dfaulter list where home is not own , rented or on mortgage



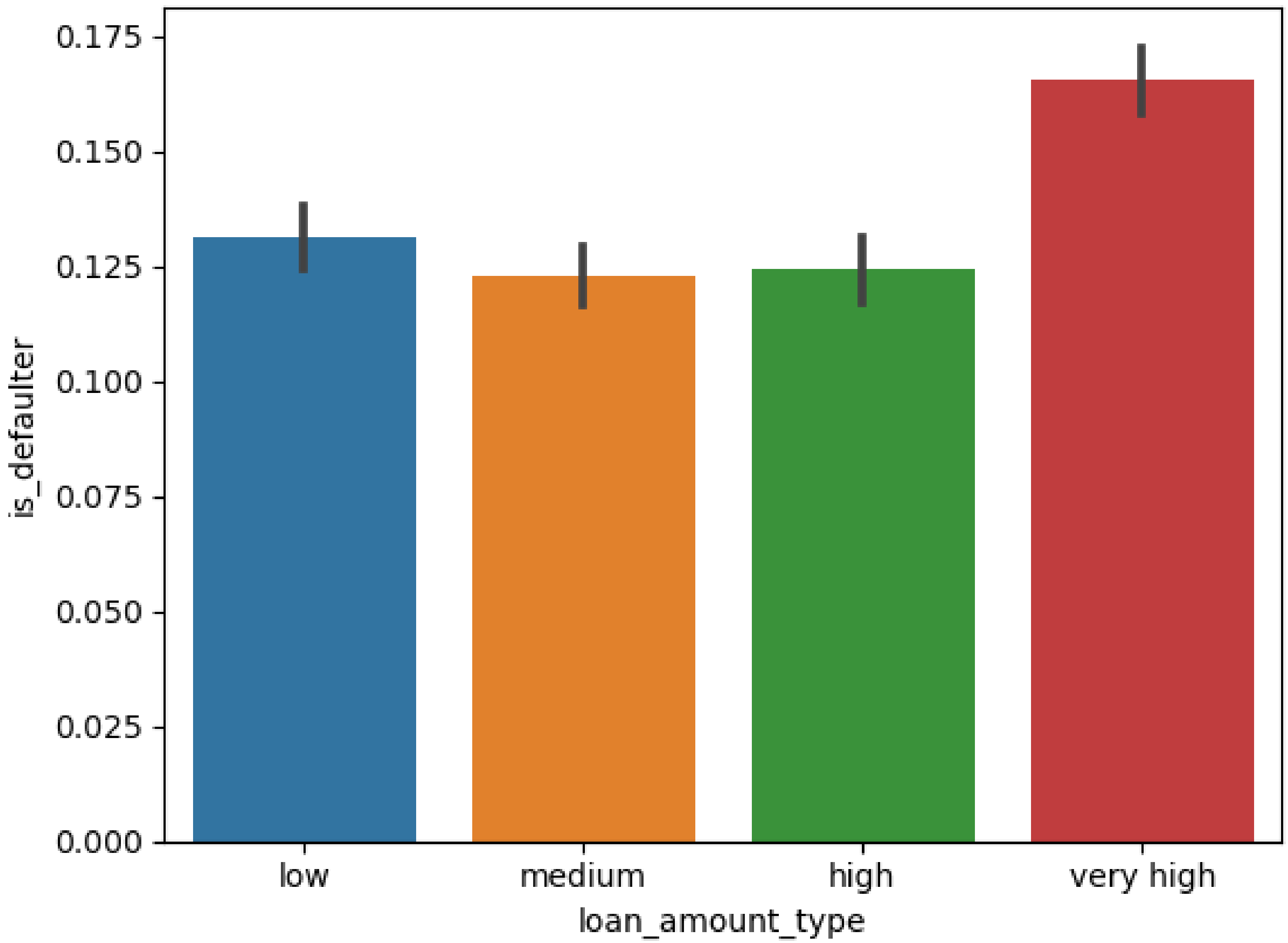
Observation on basis of Purpose of loan

We can see small business tends to fall into defaulter list more then any other purpose for which loan was taken



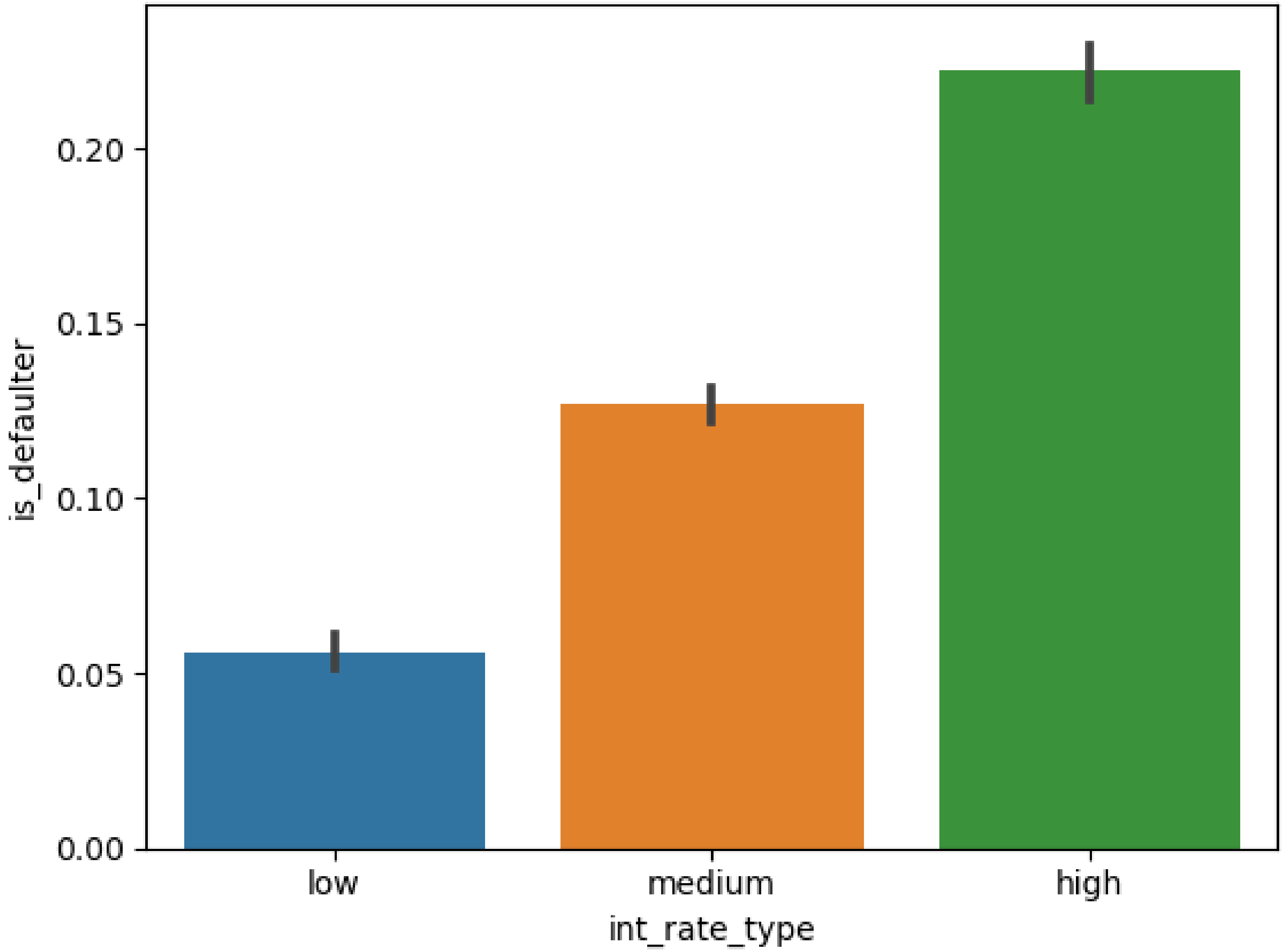
Observation on basis of loan amount type

- Looks like a very high loan amount will result in more defaulter However, we need to look at why low loan amounts have more defaulters than medium

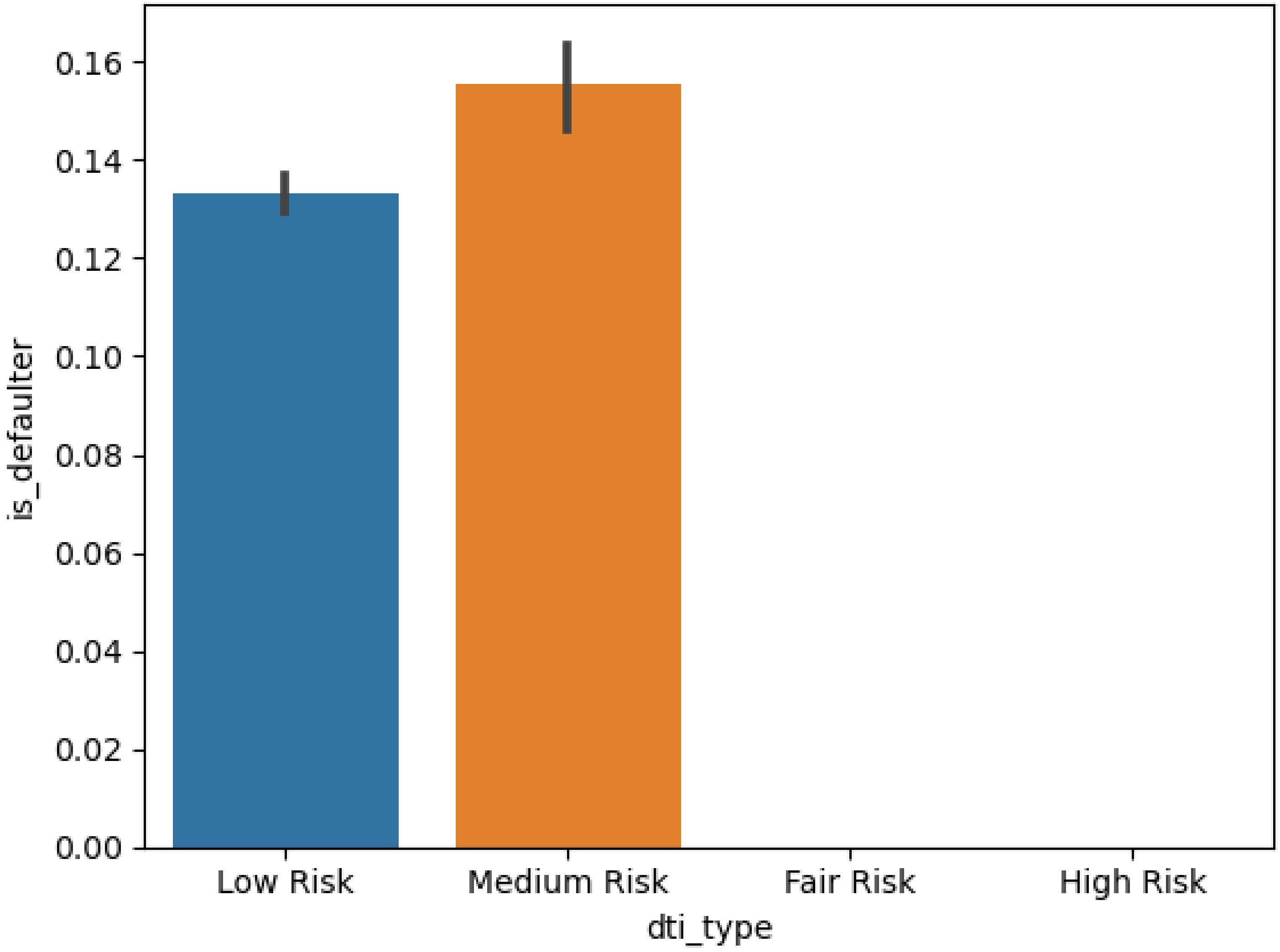


Observation on the basis of High Interest Rates

- As we can see that the higher the interest rate the higher the chances of defaulters



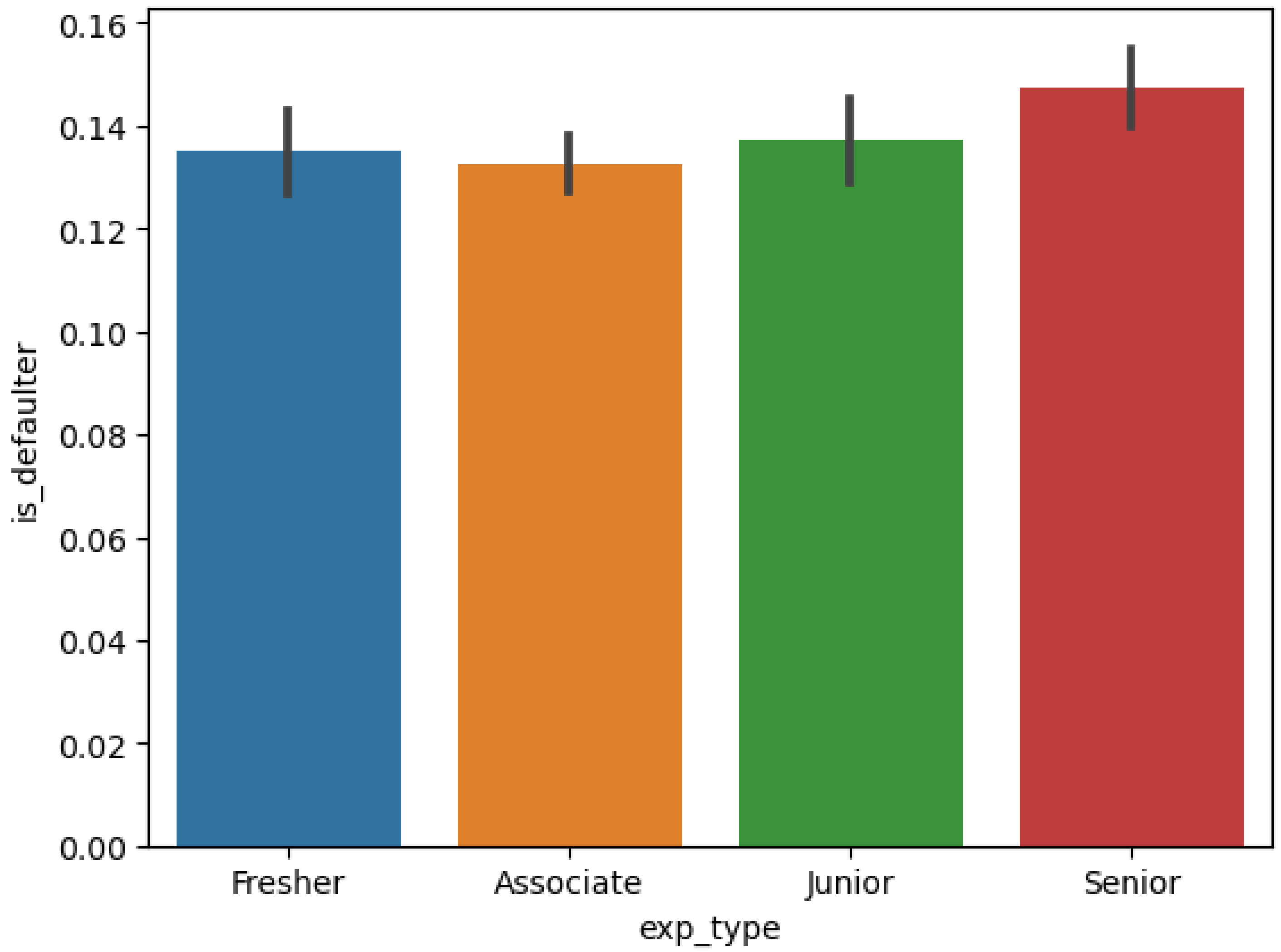
Observation on basis of DTI



- Surprisingly with Low and Medium risk dti, there still are many defaulters.

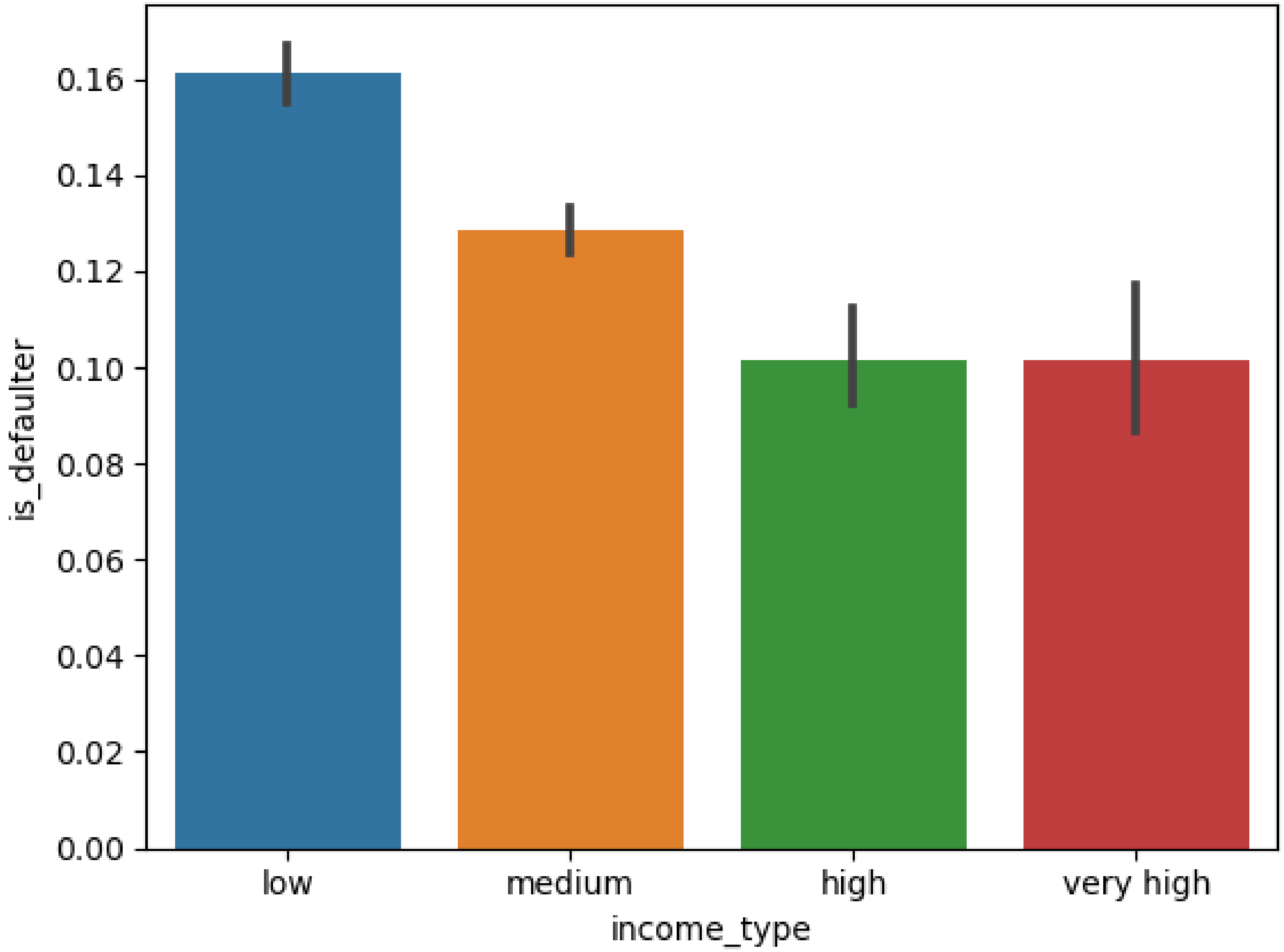
Observation on basis of years of service

- Years of experience does not impact the defaulter list



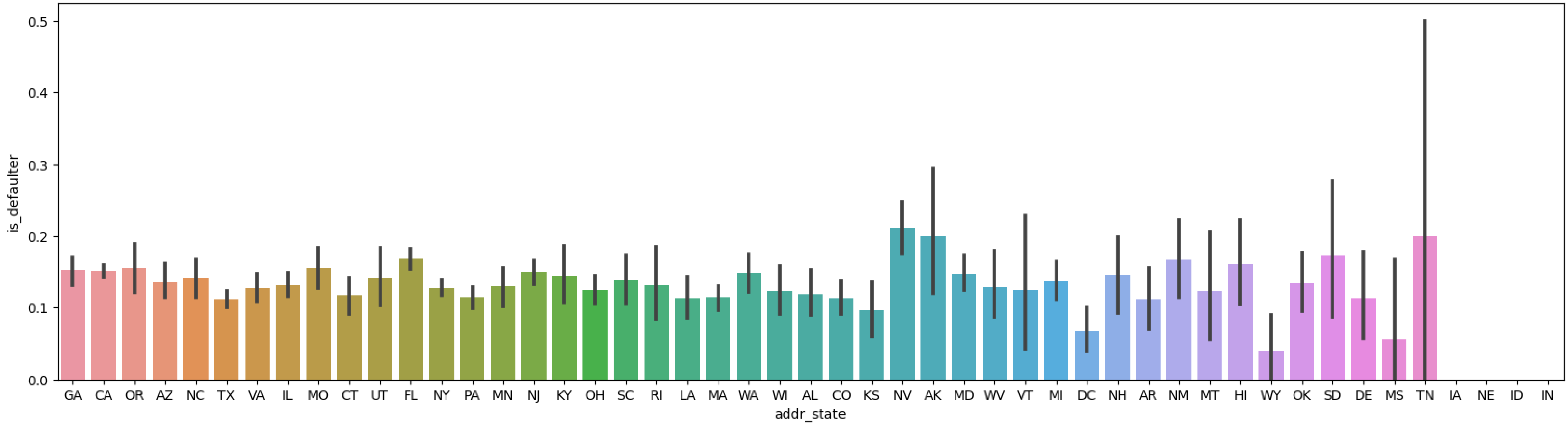
Observation on basis income type

- Lower the income higher will be the chances of loan repayment and being marked as defaulters



Observation based on State

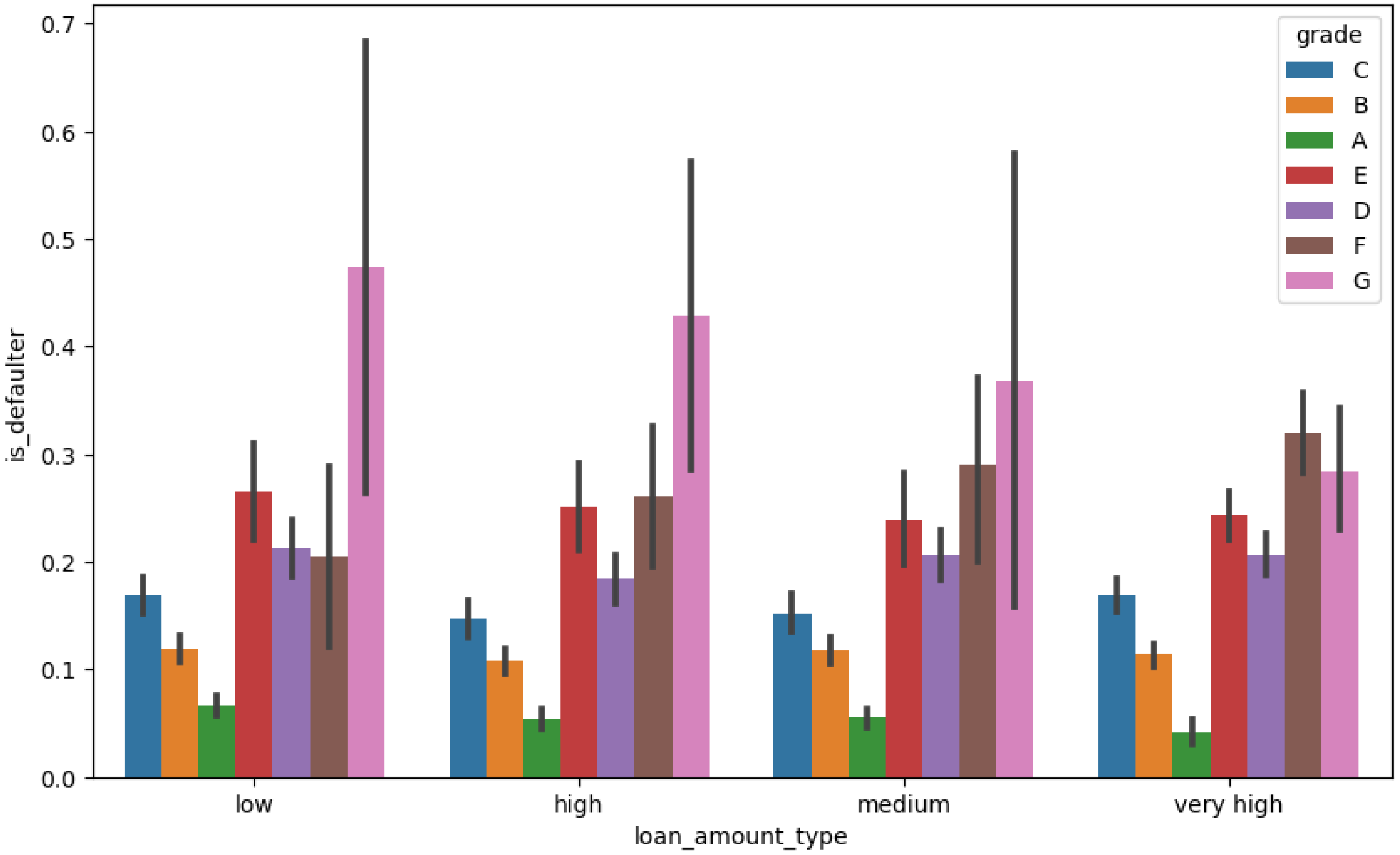
States like IA, NE,ID & IN have no defaulters where as NV and AK have highest defaulters



Bivariate Observations

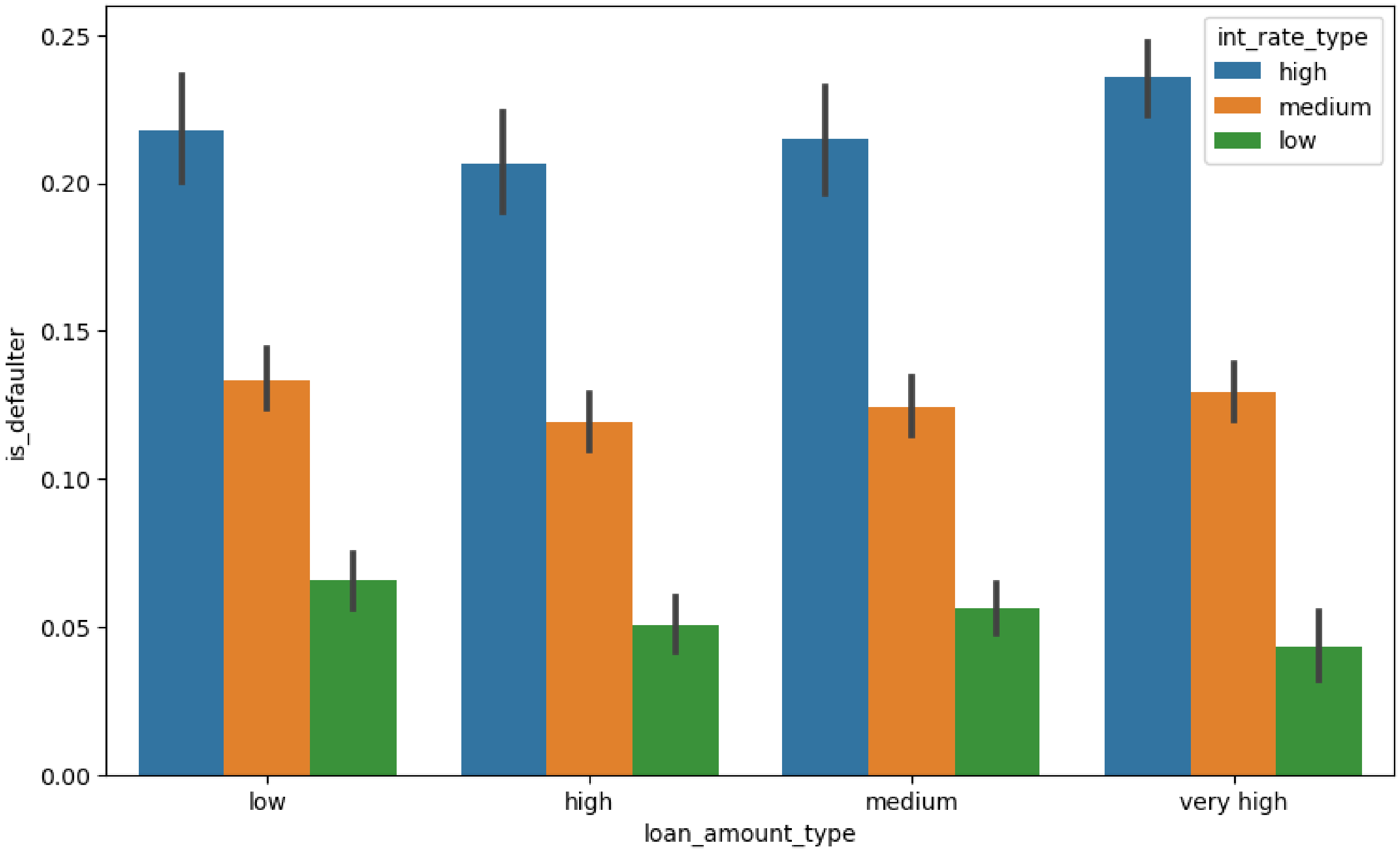
- Loan amount type
- Defaulters
- Grade

- Grade G is having the highest defauler list in every loan amount type segment



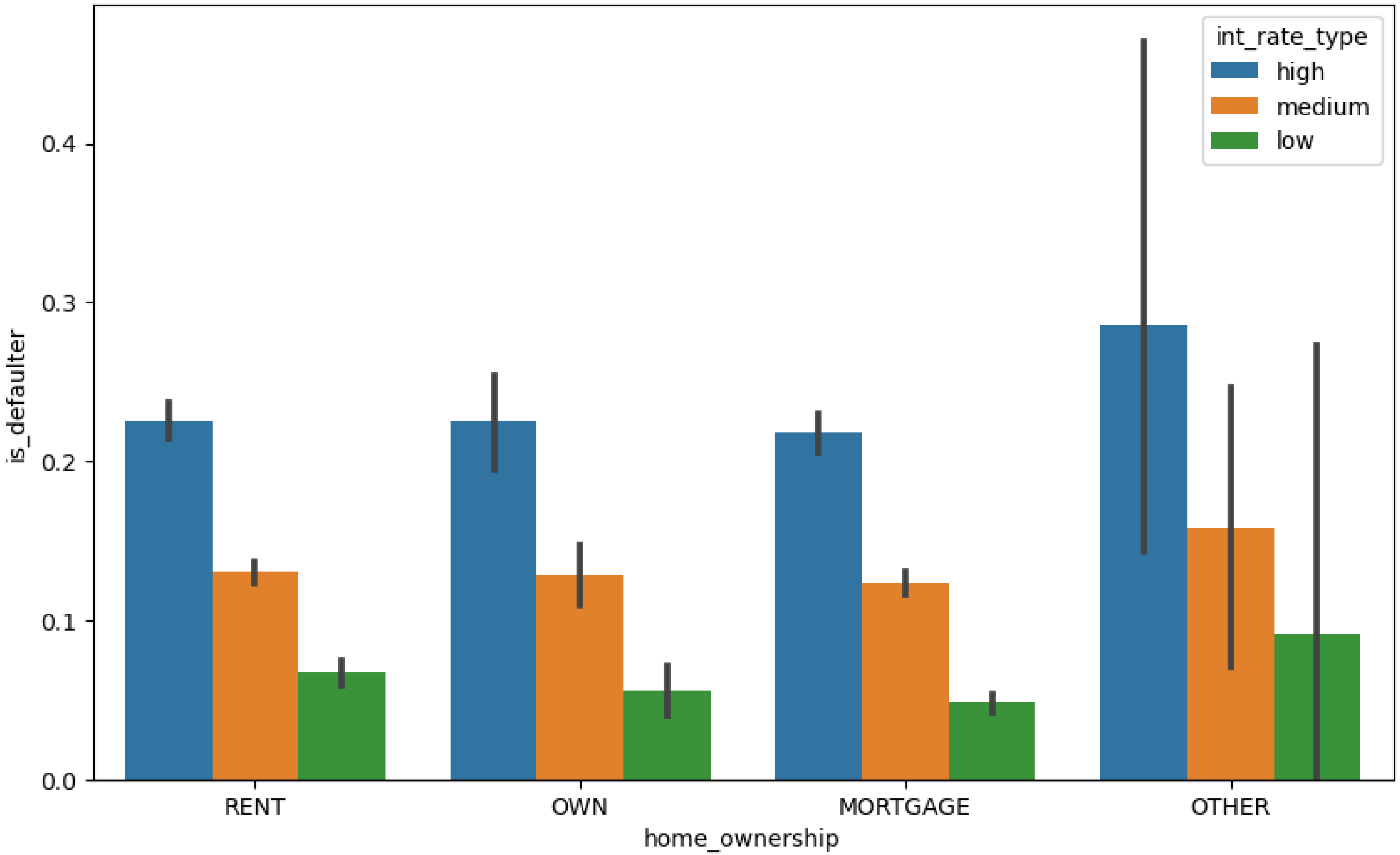
Loan amount, defaulters and interest rate type

- Observation - higher the interset rate higher is the chance of defaulter list for any loan amount type



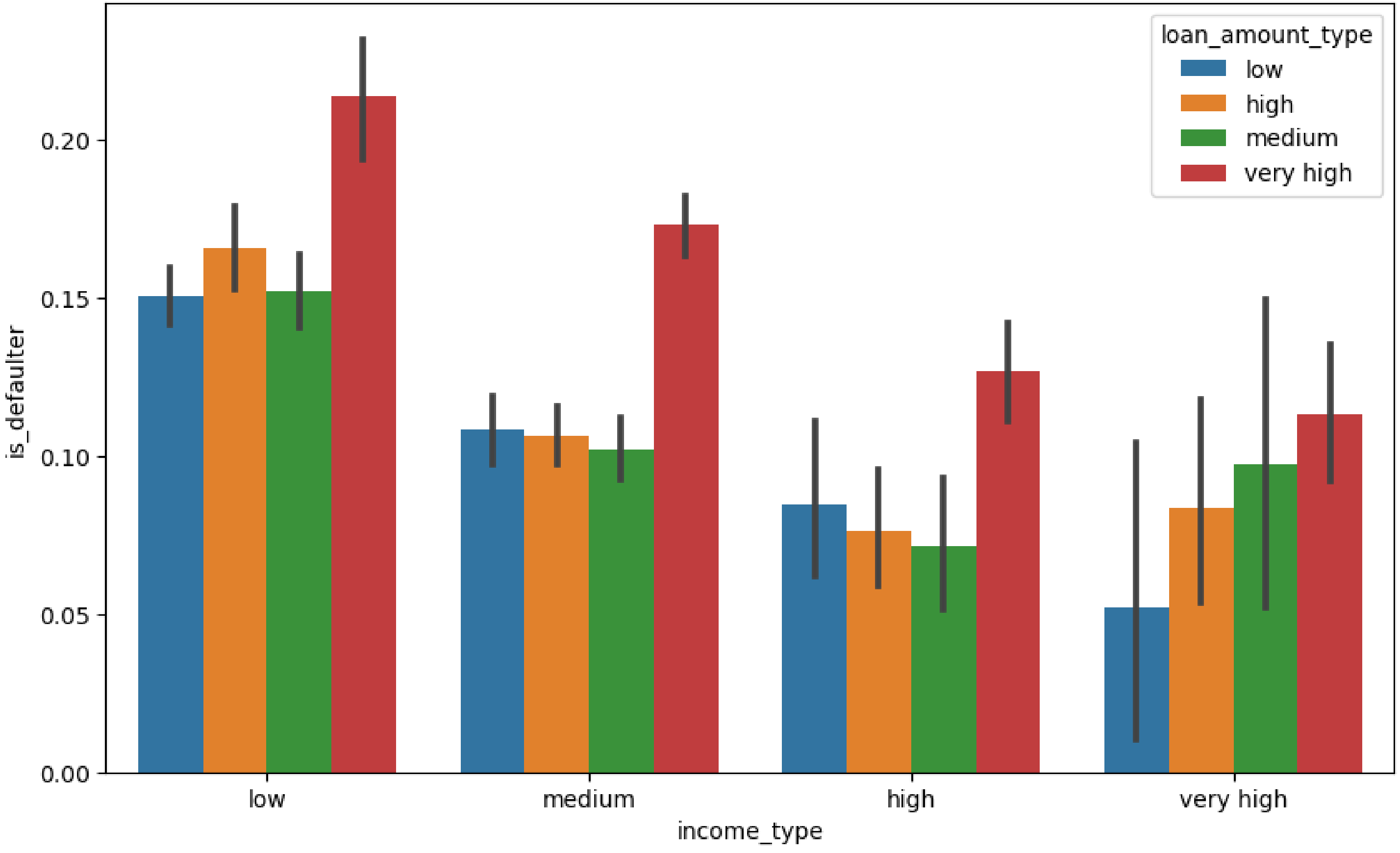
Home ownership, defaulters and Grades

Observation - every home ownership type where interset rate type is high have higher defaulter list

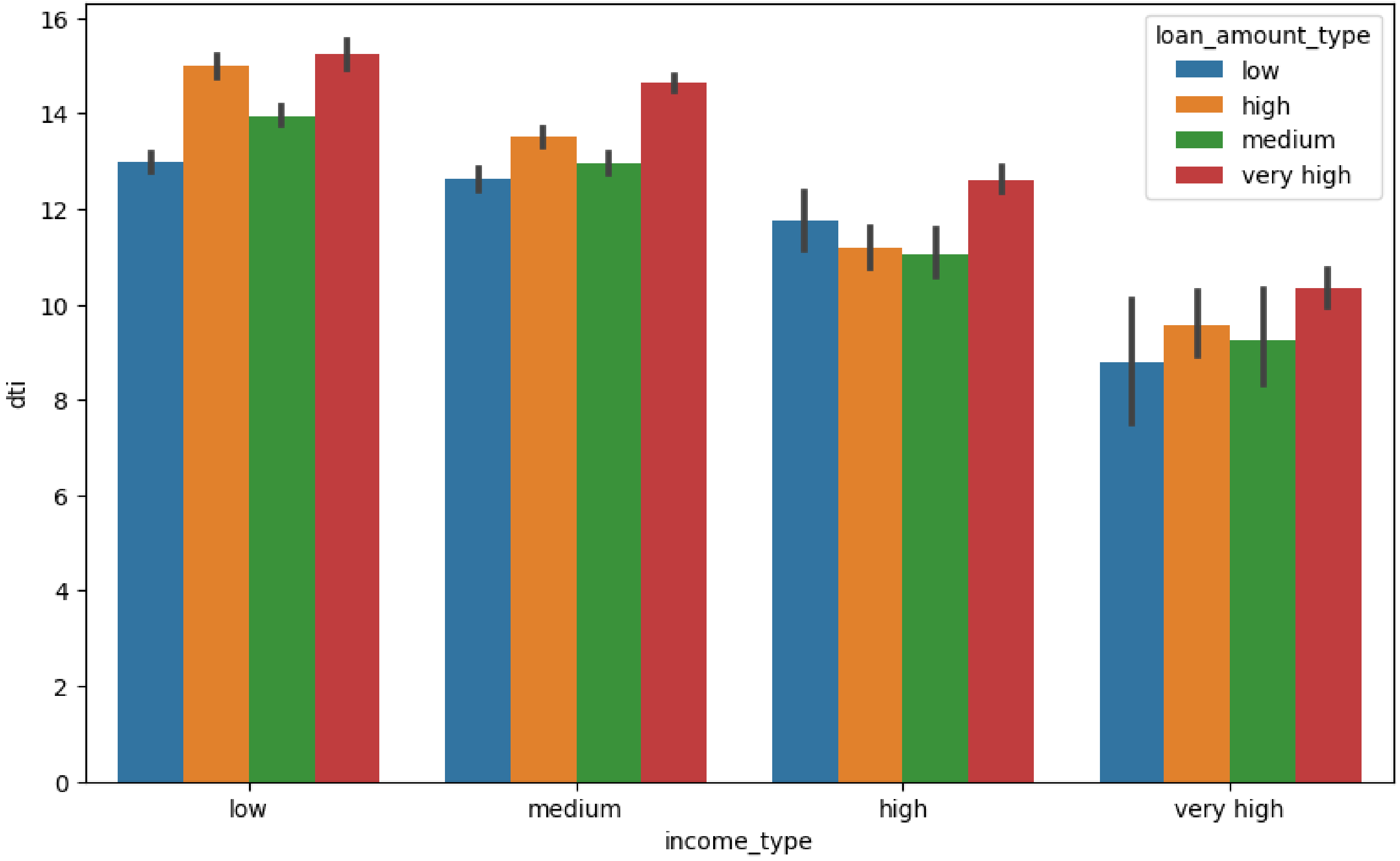


Defaulters, income type and loan amount type

- Observations - lower the income and higer the amount tends to increase the chances of defaulter list



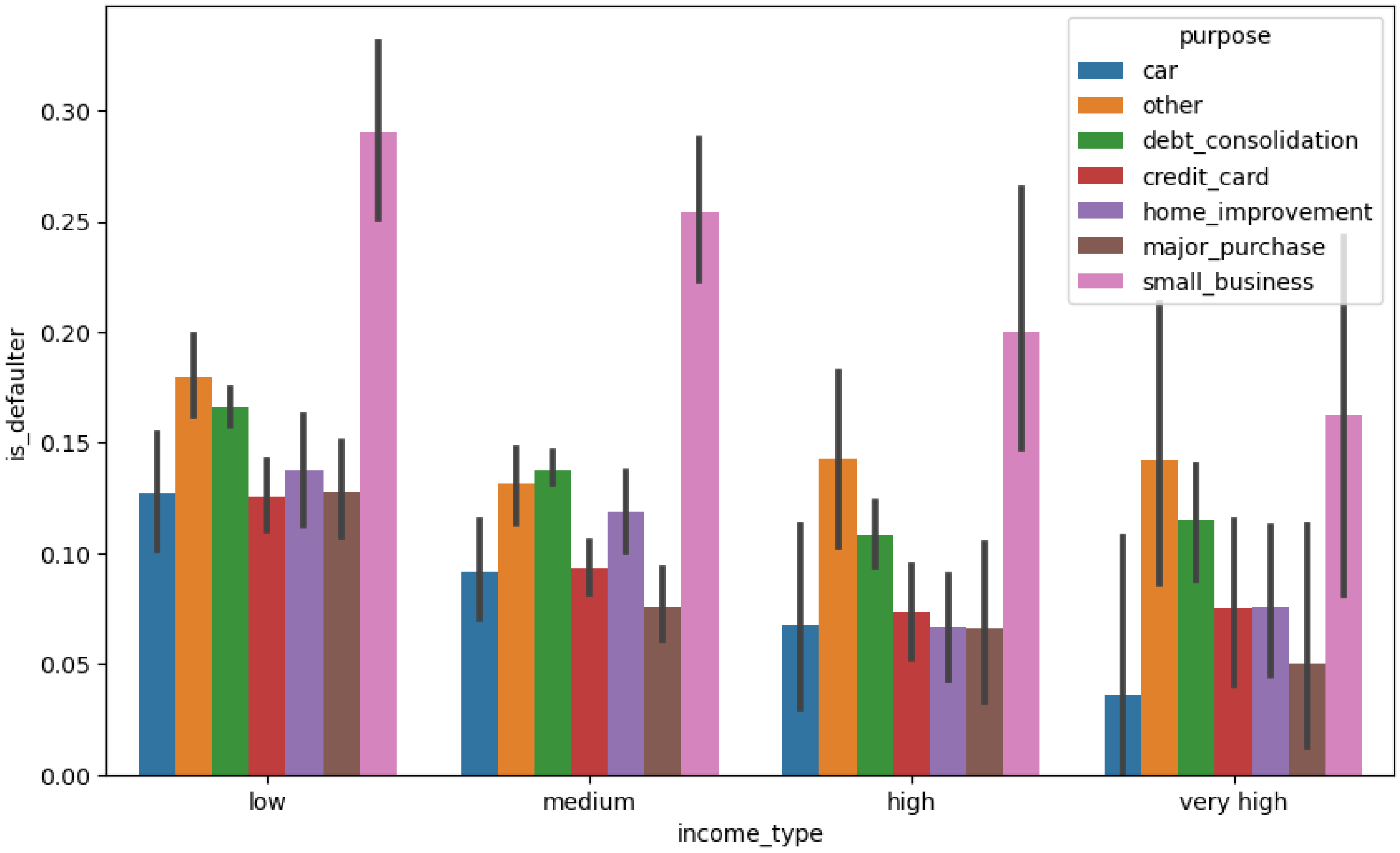
Income type , DTI and Loan amount type



Observation - lower the income and higer the amount tends to increase the chances of defaulter list

Income type , Defaulter and purpose

Observation - Small business with low income tends have higher chance to fall in defaulter list



Final Observation

- Grades indicating higher risk tend to have more defaulters.
- Longer loan terms correlate with increased default rates.
- Number of years of service does not appear to influence loan default rates significantly.
- Home ownership status impacts default rates, particularly when homes are rented or on mortgage.
- Loans for small businesses are more likely to default compared to loans for other purposes.
- Higher loan amounts tend to result in more defaulters, though lower loan amounts have unexpectedly high default rates compared to medium amounts.
- Higher interest rates are associated with higher default rates across different loan amount types.
- Even with low and medium debt-to-income ratios (dti), there are significant numbers of defaulters.
- Lower income levels are associated with higher chances of defaulting on loans.
- States like IA, NE, ID, and IN have no defaulters, whereas NV and AK have the highest default rates.
- Grade G consistently shows the highest number of defaulters across various loan amount segments.
- Higher interest rates are consistently linked with higher default rates across all types of home ownership.
- Lower income coupled with higher loan amounts increases the likelihood of defaulting.
- Small businesses with low incomes are particularly prone to defaulting on loans.

Thanks

Leading Club Case study