

LENDINGCLUB CASE STUDY

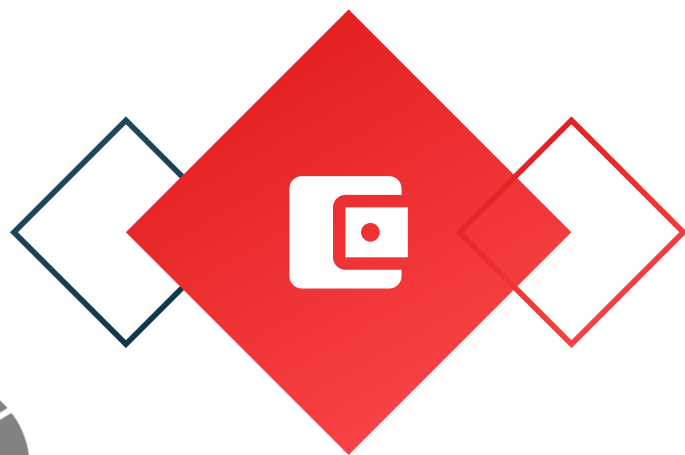
- Vishnu Karnam
- Venu Prakash Reddy



Problem Statement

When the company receives a loan application, the company has to make a decision 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 for the company



Analysis Done



Data Cleaning

Checked and dropped the columns in which more than 90% of the row are null and also dropped other columns which are not useful for our analysis



Univariate analysis

Users can not only present the presentation on the projector or computer, but they can also print out the presentation.



Multivariate analysis

Users can not only present the presentation on the projector or computer, but they can also print out the presentation.



Data Understanding

Types of variables

- Customer (applicant) demographic
- Loan-related information & Characteristics
- Customer behaviour (if the loan is granted)

applicant demographic	Loan characteristics	Customer behaviour
Emp Length	Loan Amount	Laon Grade
Emp Grade	Funded Amount	earliest credit line
Annual income	Funded Amount Investment=	Revolving balance
Zip code	Interest Rate	Recoveries
Emp title	Loan Status	Loan purpose

Data Cleaning

Dropping unwanted columns

- We observed 56 columns have more than 90 % of null values, which will not help us analyze the data, so we have dropped them.
- Dropped the columns with only one unique value as they will not affect the analysis in any manner.
- We have also dropped the columns which contain only 0 and nan.
- We have divided the columns into 3 sections namely Customer information, customer behavior, and loan characteristics.
- As customer behavior values are unknown during the application time, doing analysis with them will not help the lender to predict the outcome of the customer as a defaulter or fully paid based on historical data of other customers so we are dropping those columns.

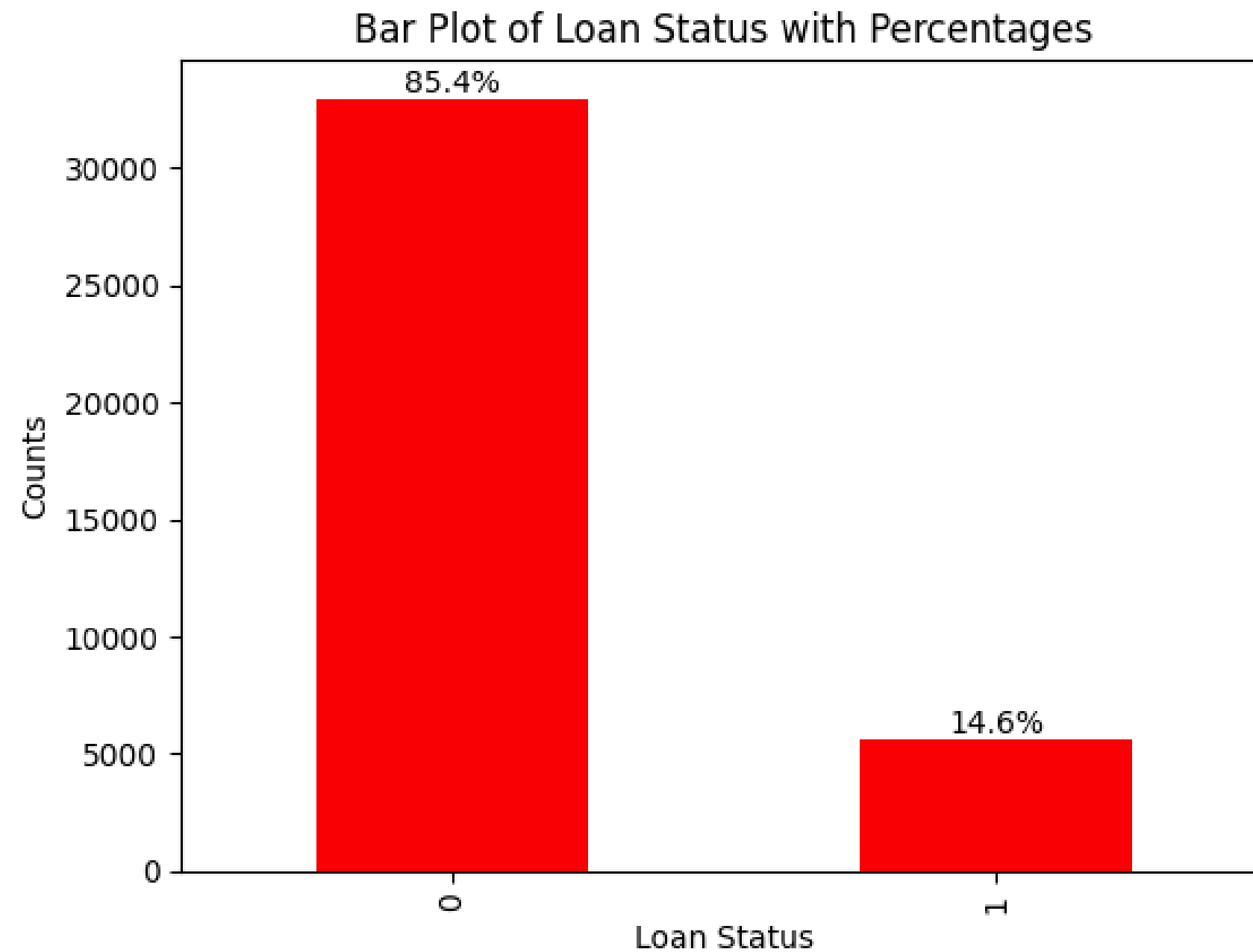
Data Cleaning

Derived values and conversions

- Converted ***emp_length*** from a string to a consumable integer format and stored in the ***emp_length_years*** column
- Converted ***issue_d*** to the date time format.
- Extracted month and year from ***issued_d*** and stored in ***month*** and ***year*** column.
- Removed all the current rows from the ***loan_status*** column and assigned all the fully-paid as 0, and charged-off as 1 for simplified analysis.

Univariate analysis

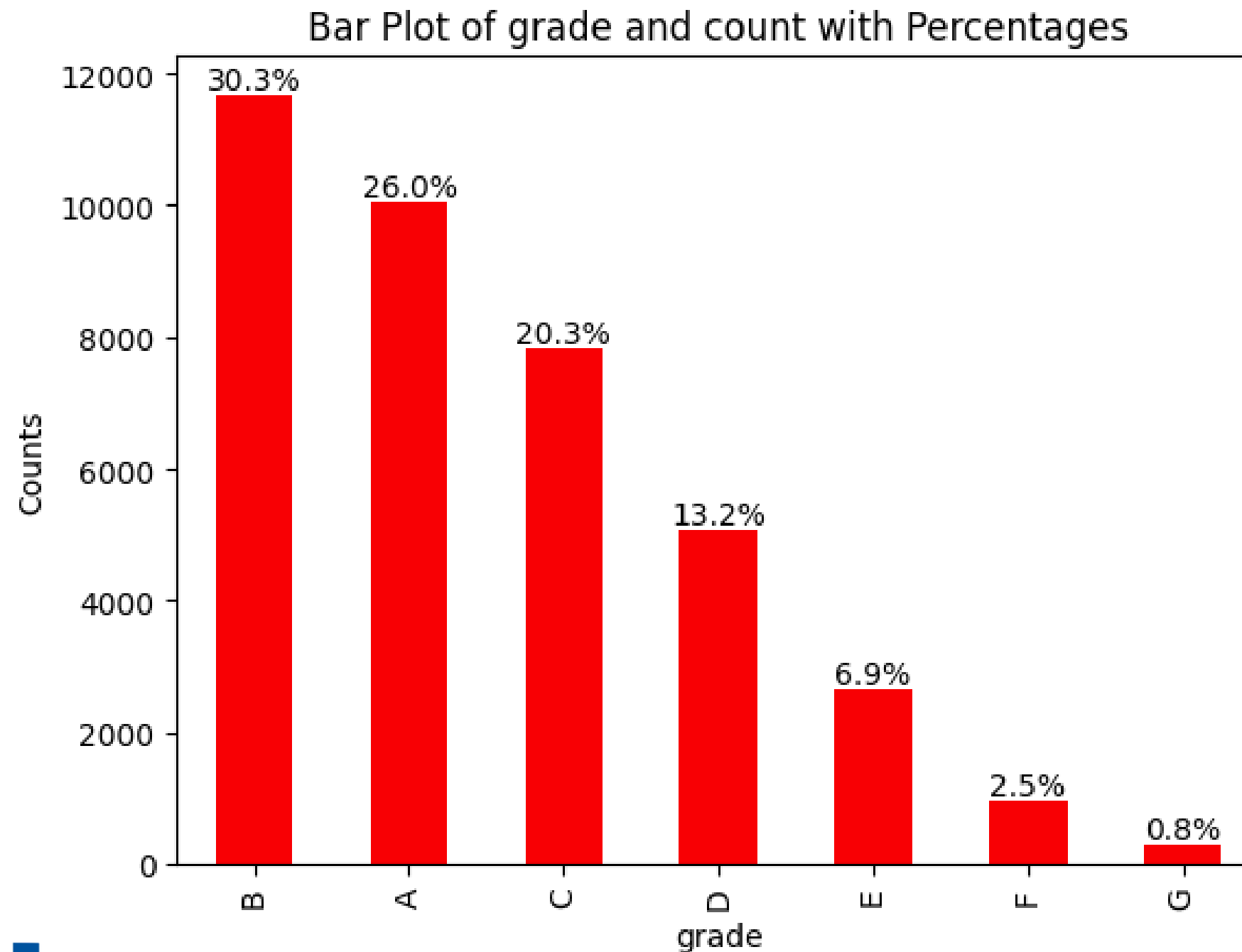
Loan Status



Based on the analysis of the predictor variable **status** we have 85.4 % of the applicants fully paid the loan where as the 14.6% of the application as charged-off.

Univariate analysis

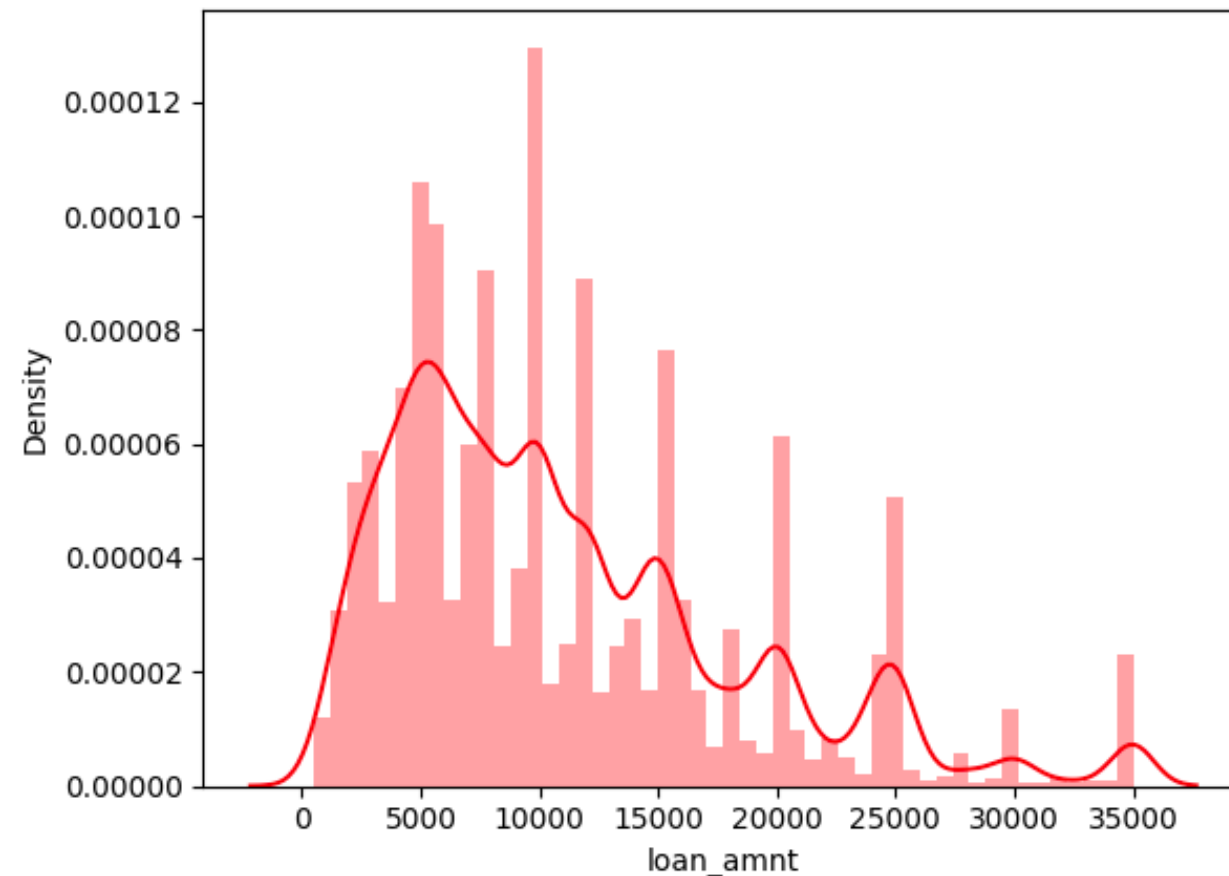
Grade



From the bar chart, we can observe that the applicants with grade **B** have applied for the loan frequently. The applicants with grade **G** applied for the least number.

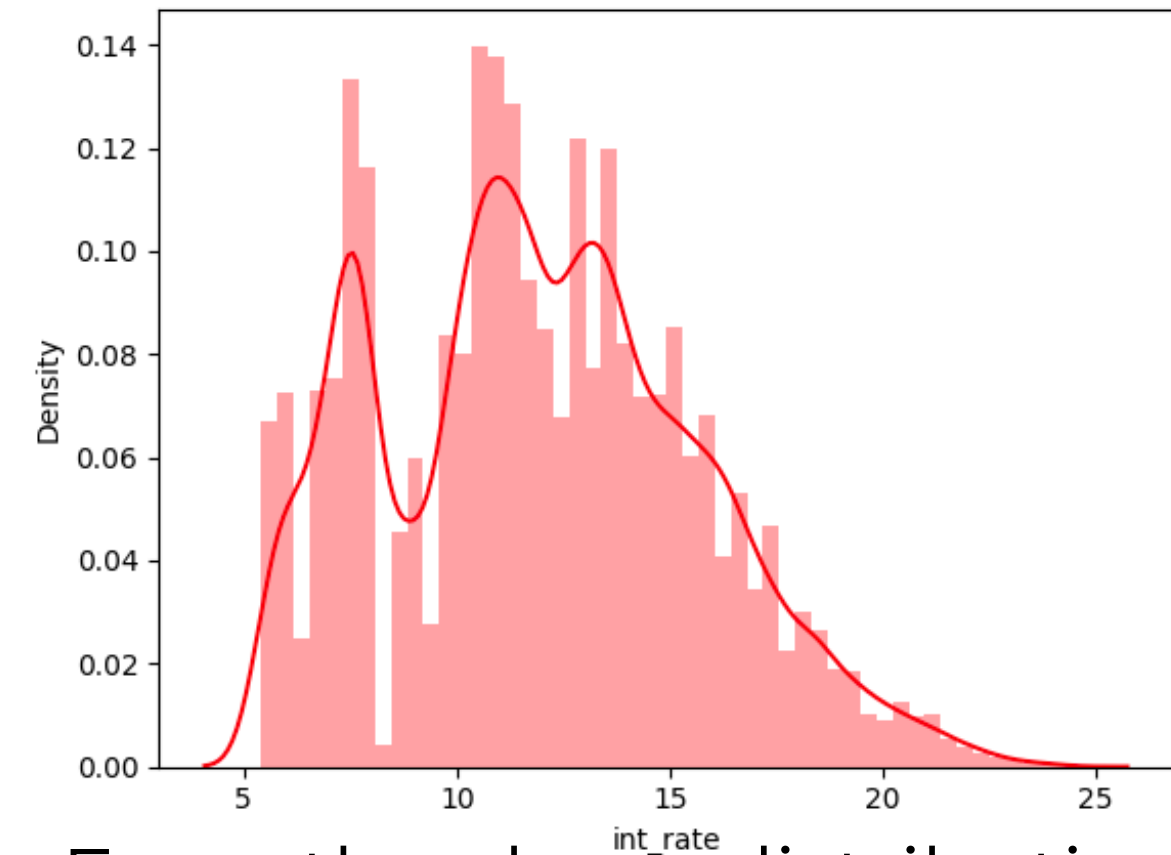
Univariate analysis

Loan Amount



From the above distribution chart, we can observe most of the loan amounts distributed around 5000.

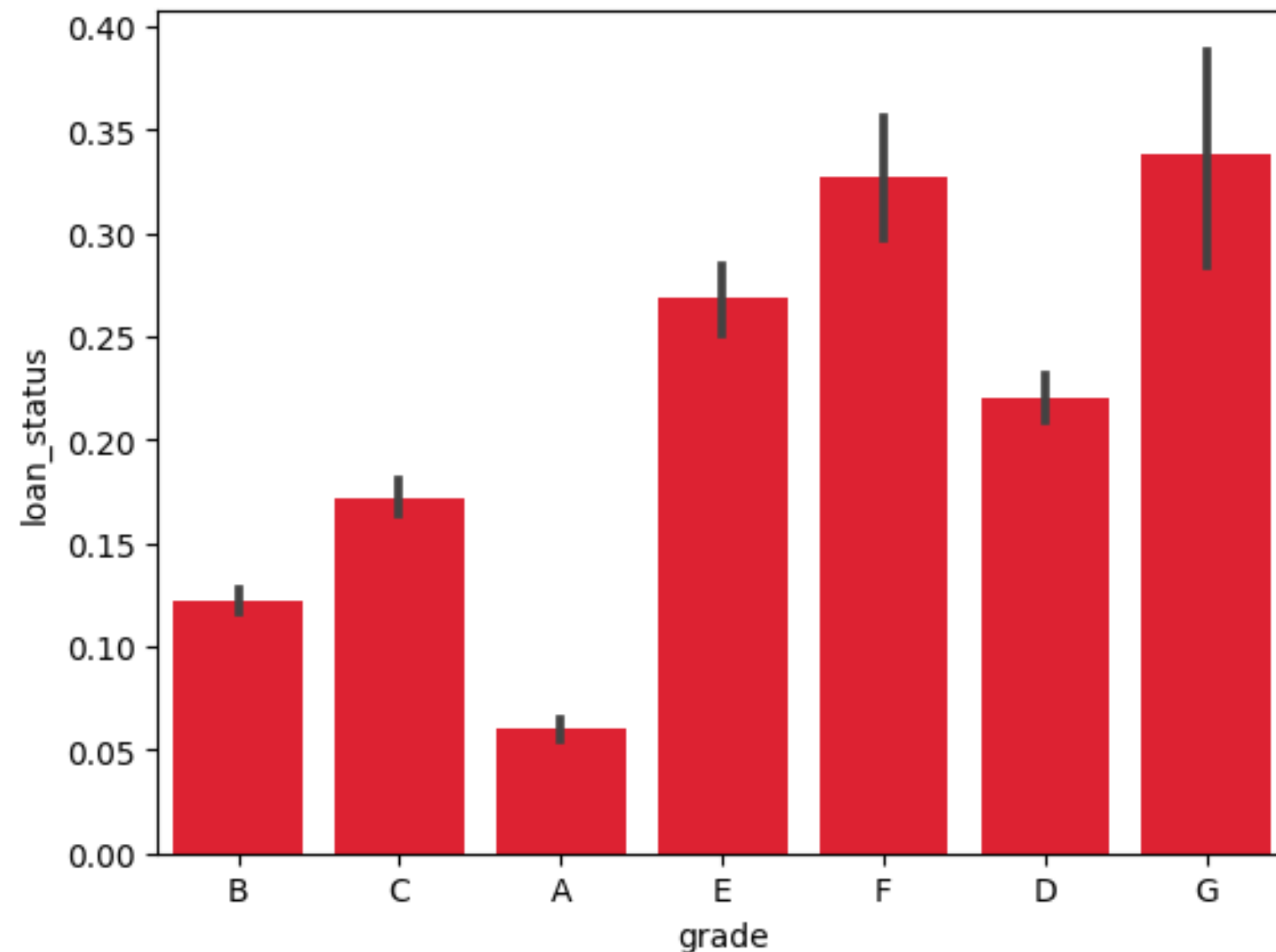
Intrest Rate



From the above distribution chart, we can observe most of the loan interests are distributed around 7.5 or 12.5

Bivariate analysis

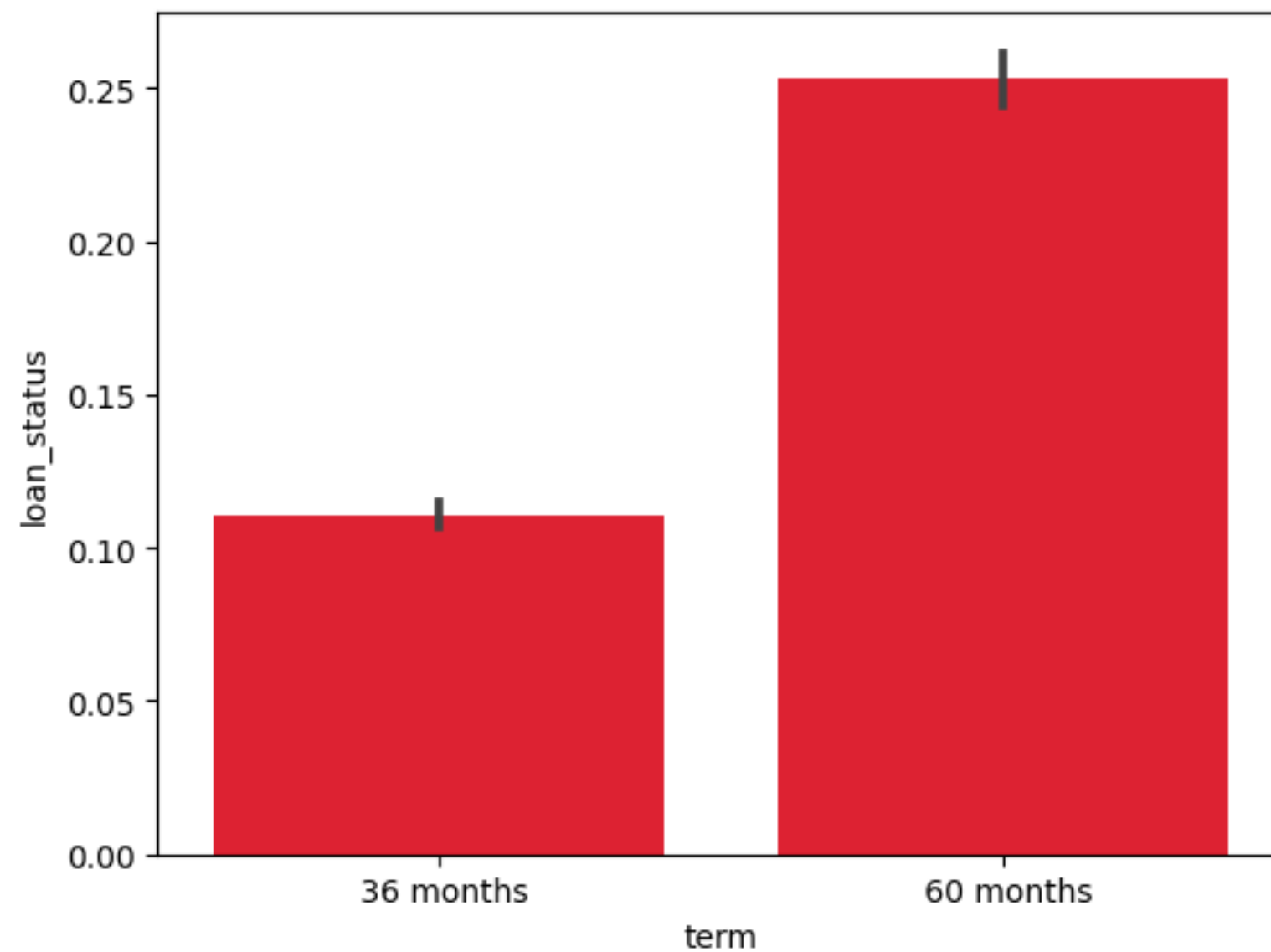
Loan status vs Grade



From the figure its evident that the applicants with the lowest grades are more likely be the defaulters. And the applicants with high grade are most likely to payoff the loan.

Bivariate analysis

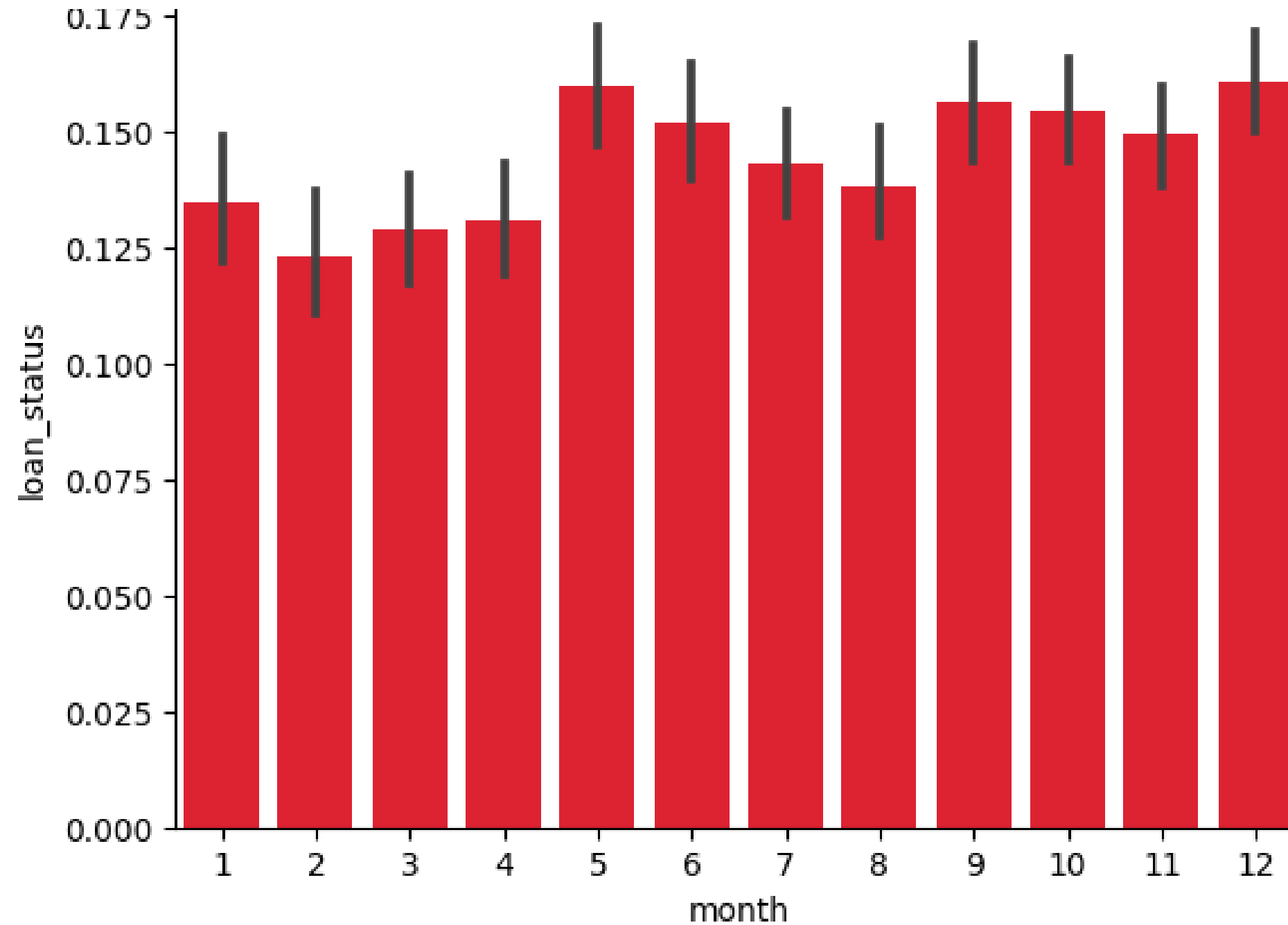
Loan status vs Duration



Loans with short duration are more likely to be repaid but the loans with the longer duration are being the defaulters. The reason could be the mindset of the applicant might change or even his economic status can also change which could be the reason for the more defaulters

Bivariate analysis

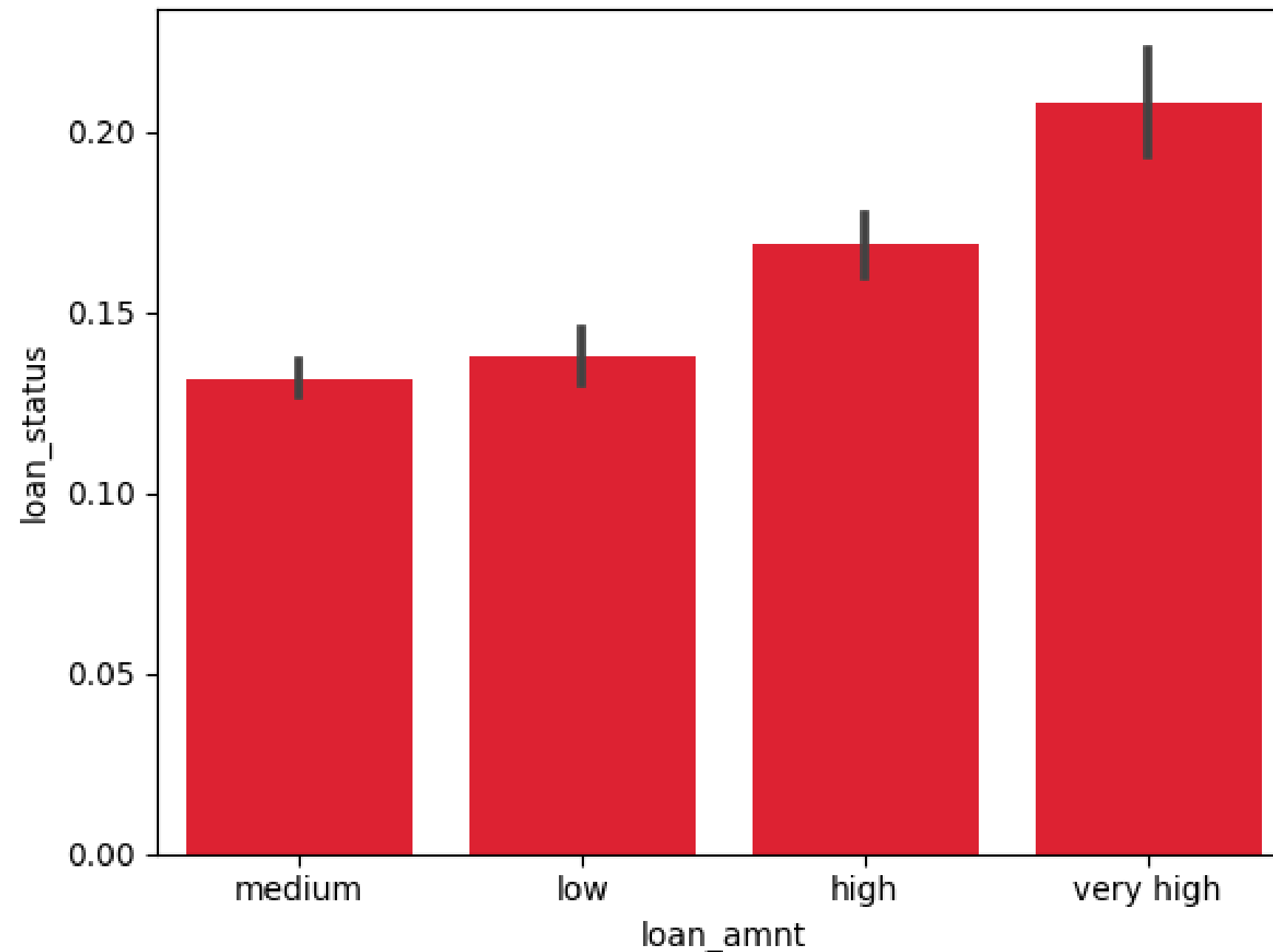
Loan status vs Month



We observe a more number of defaulters in the months of 5,6, 9, 10, 11, and 12. The reasons is because the high school and college admissions.

Bivariate analysis

Loan status vs Amount

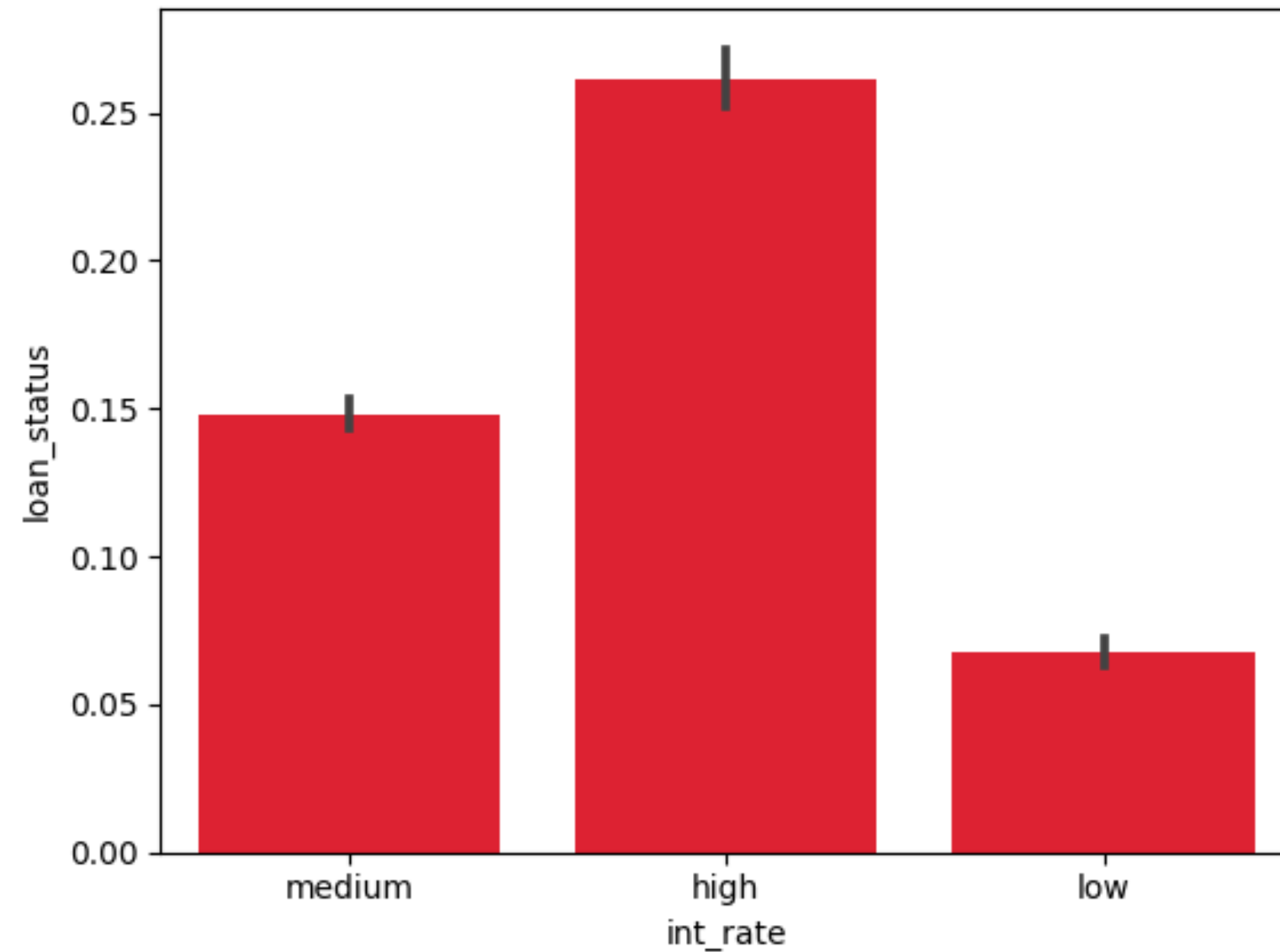


The applicants with high loan amounts i.e. more than 25k are more likely be defaulters.

loan amount low(<5k), medium(5k to 15k) high(15k to 25k) very high(>25k)

Bivariate analysis

Loan status vs Int_rate

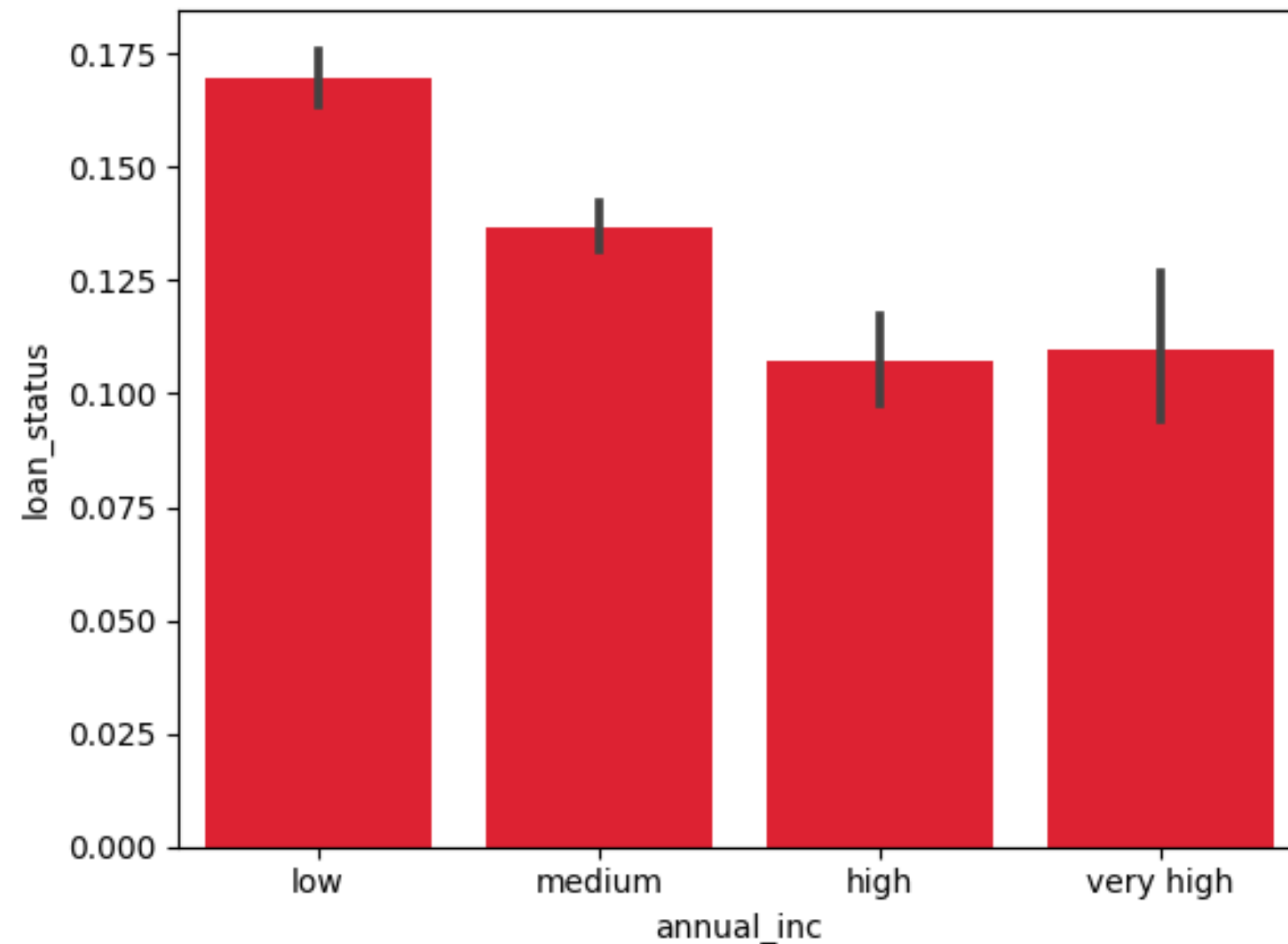


The applicants with high interest rates i.e. more than 15% are more likely to be defaulters.

interest rate low(<10%), medium(10% to 15%) high(>15%)

Bivariate analysis

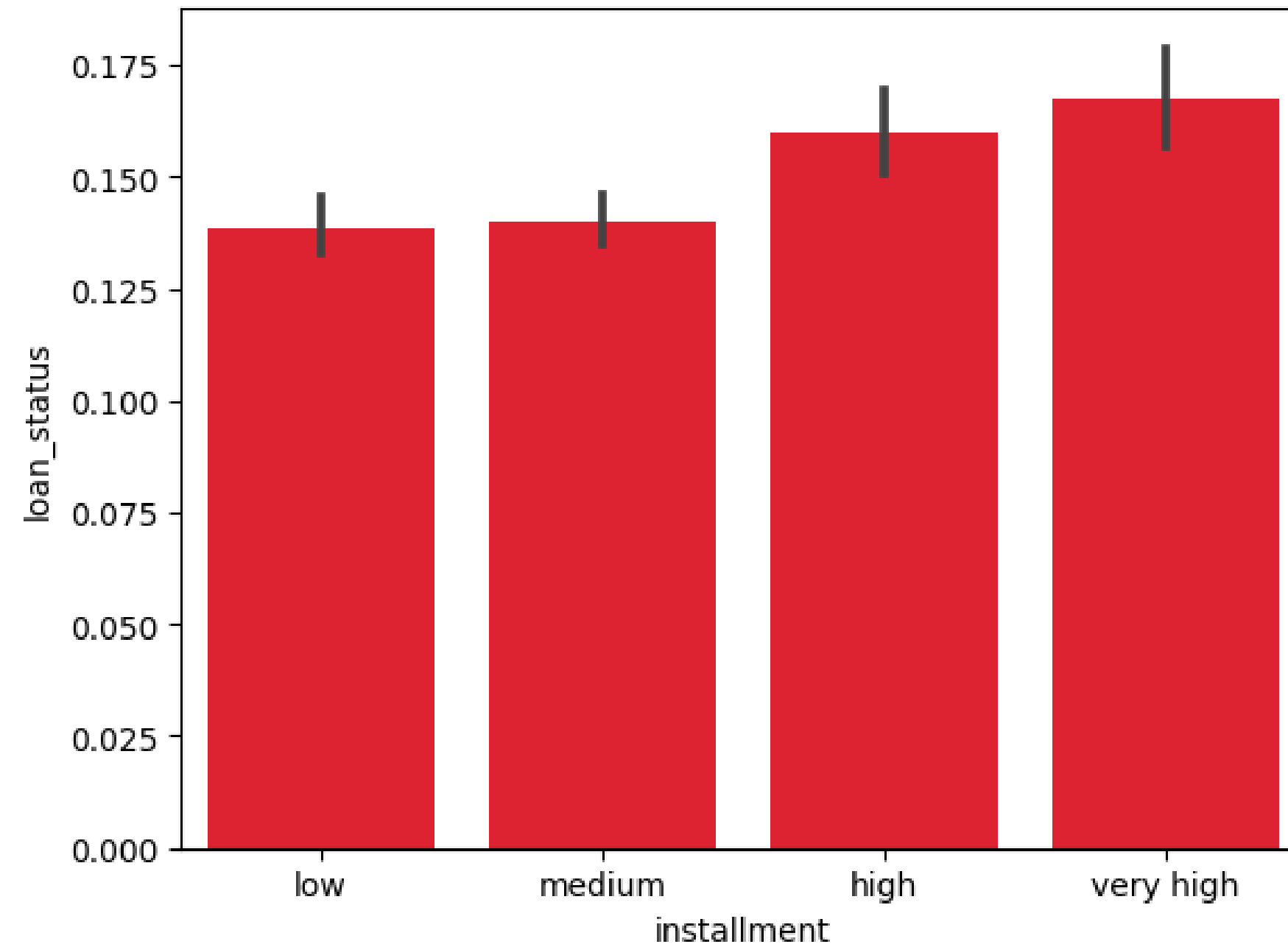
Loan status vs Annual income



The applicants with low annual income i.e. more than 50k are more likely to be defaulters.

Bivariate analysis

Loan status vs Installment

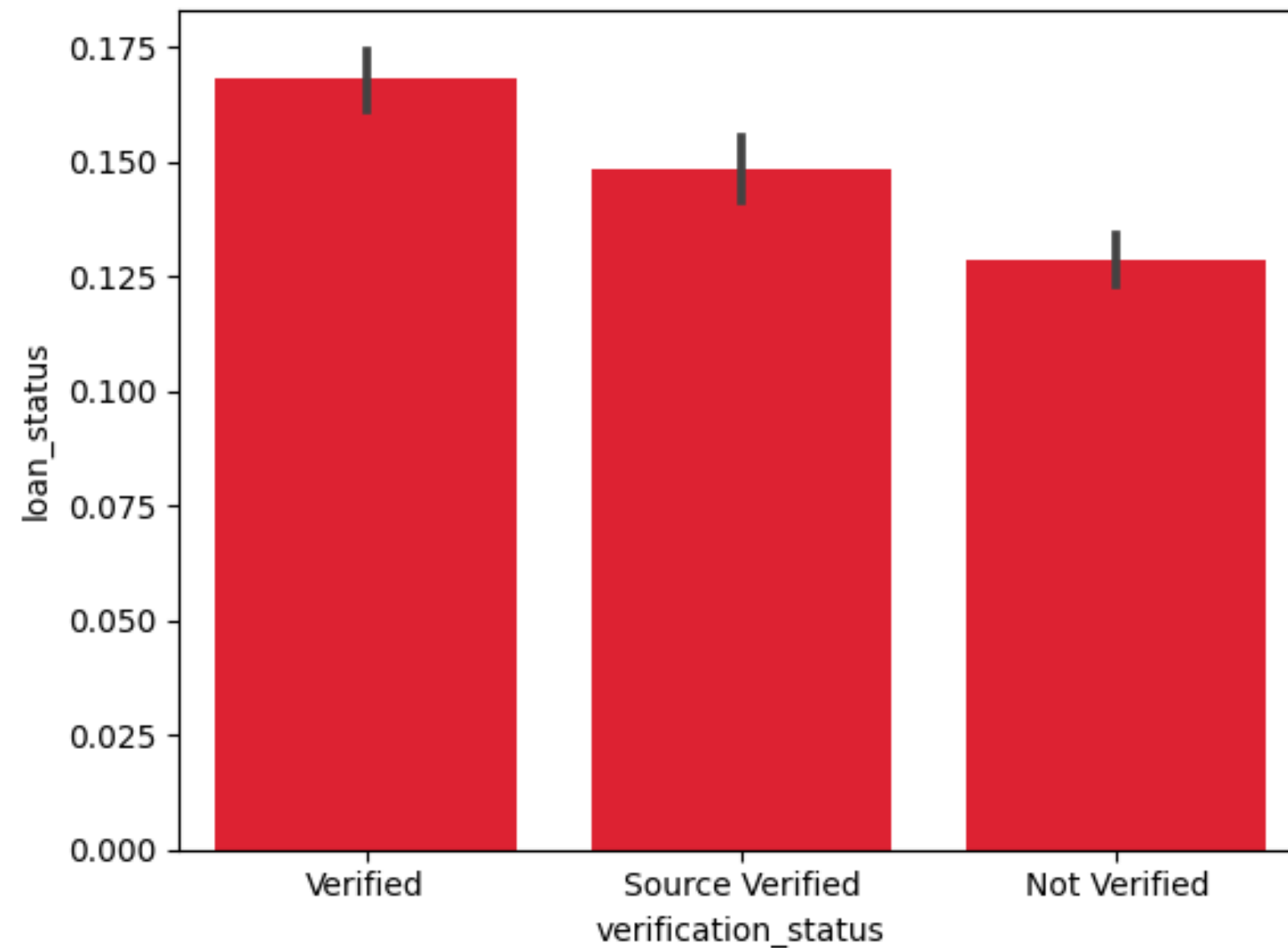


The applicants with High-interest rates i.e. more than 600 are more likely to be defaulters.

Instalment low(<200), medium(200 to 400) high(400 to 600), very hiigh(>600)

Bivariate analysis

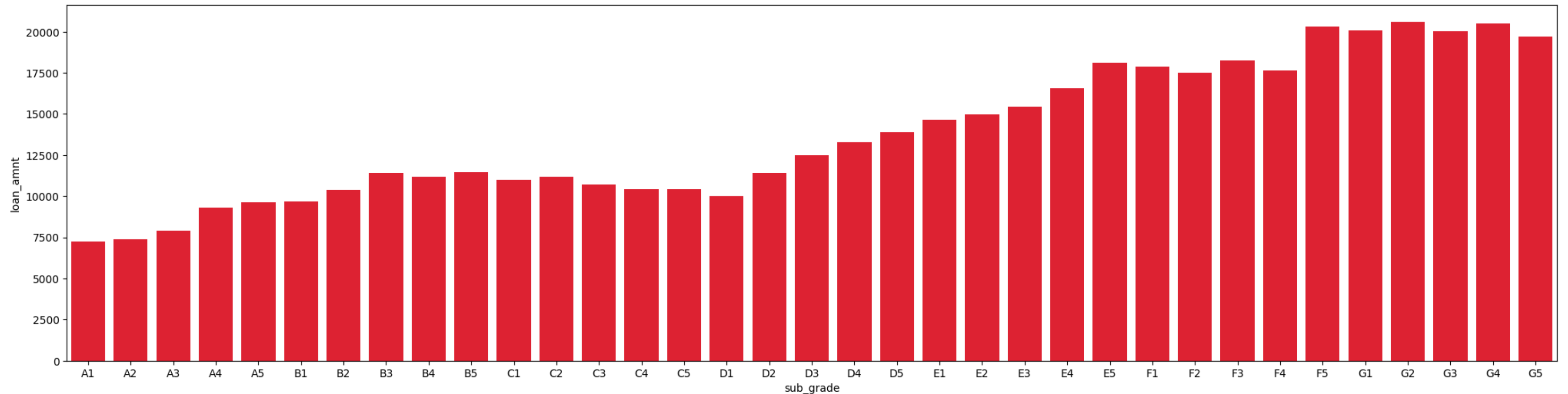
Loan status vs Verification



surprisingly, verified loans default more than not verified

Bivariate analysis

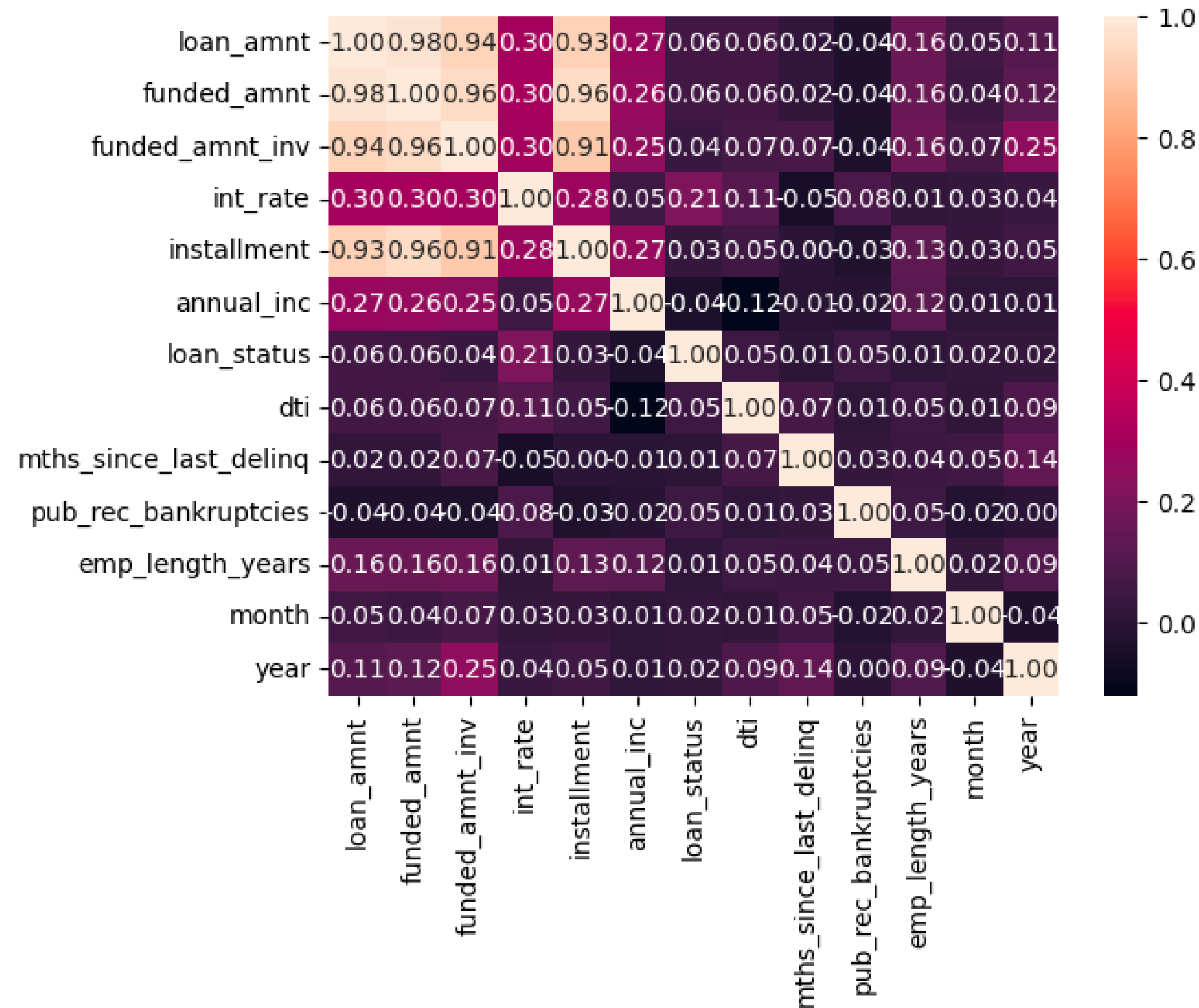
Loan Amount vs Sub Grade



Applicants with lower grades (G) are taking the larger loan amounts when compared to the higher grades (A)

Bivariate analysis

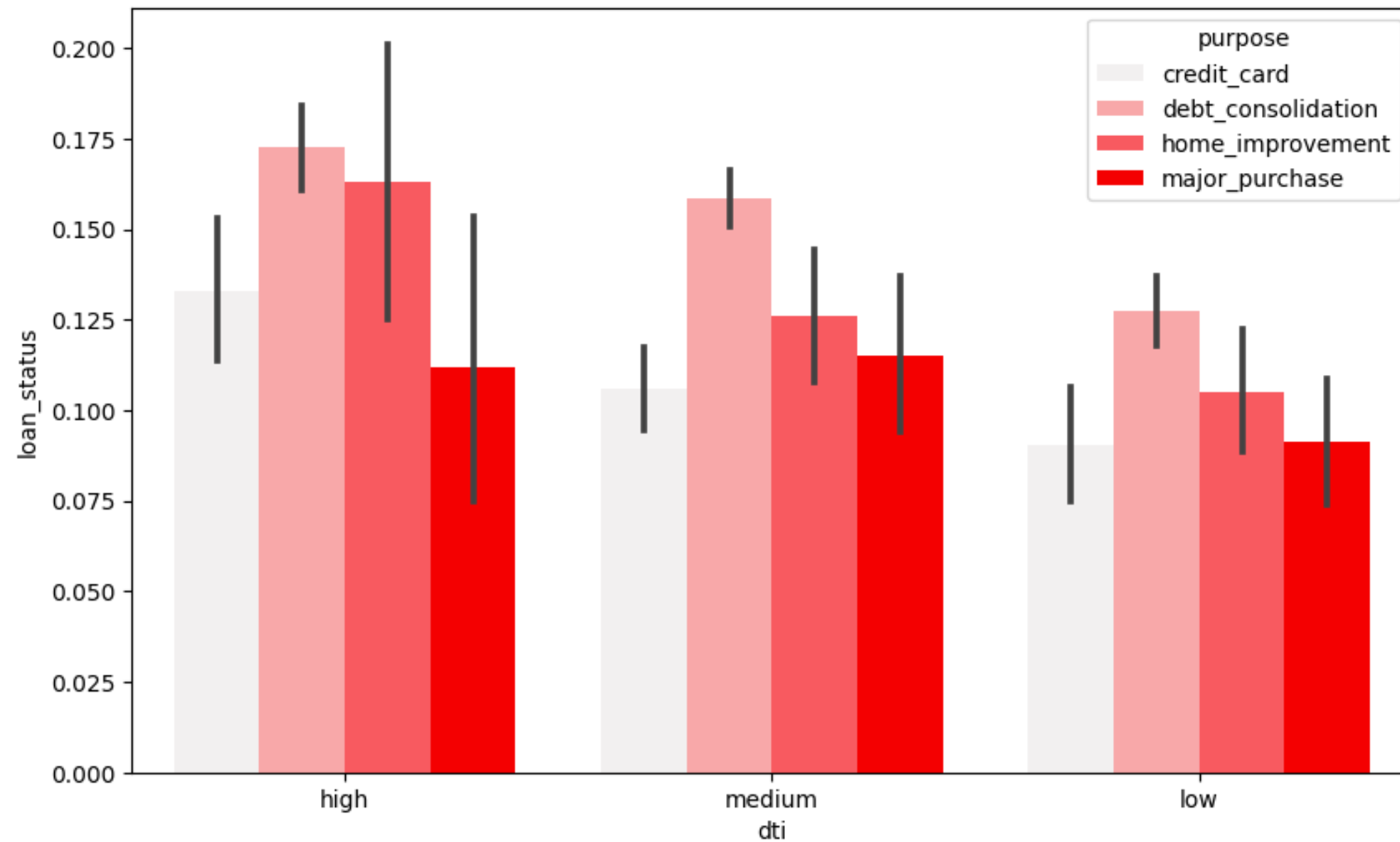
Correlations



Applicants with lower grades (G) are taking larger loan amounts when compared to the higher grades (A)

Segmented analysis

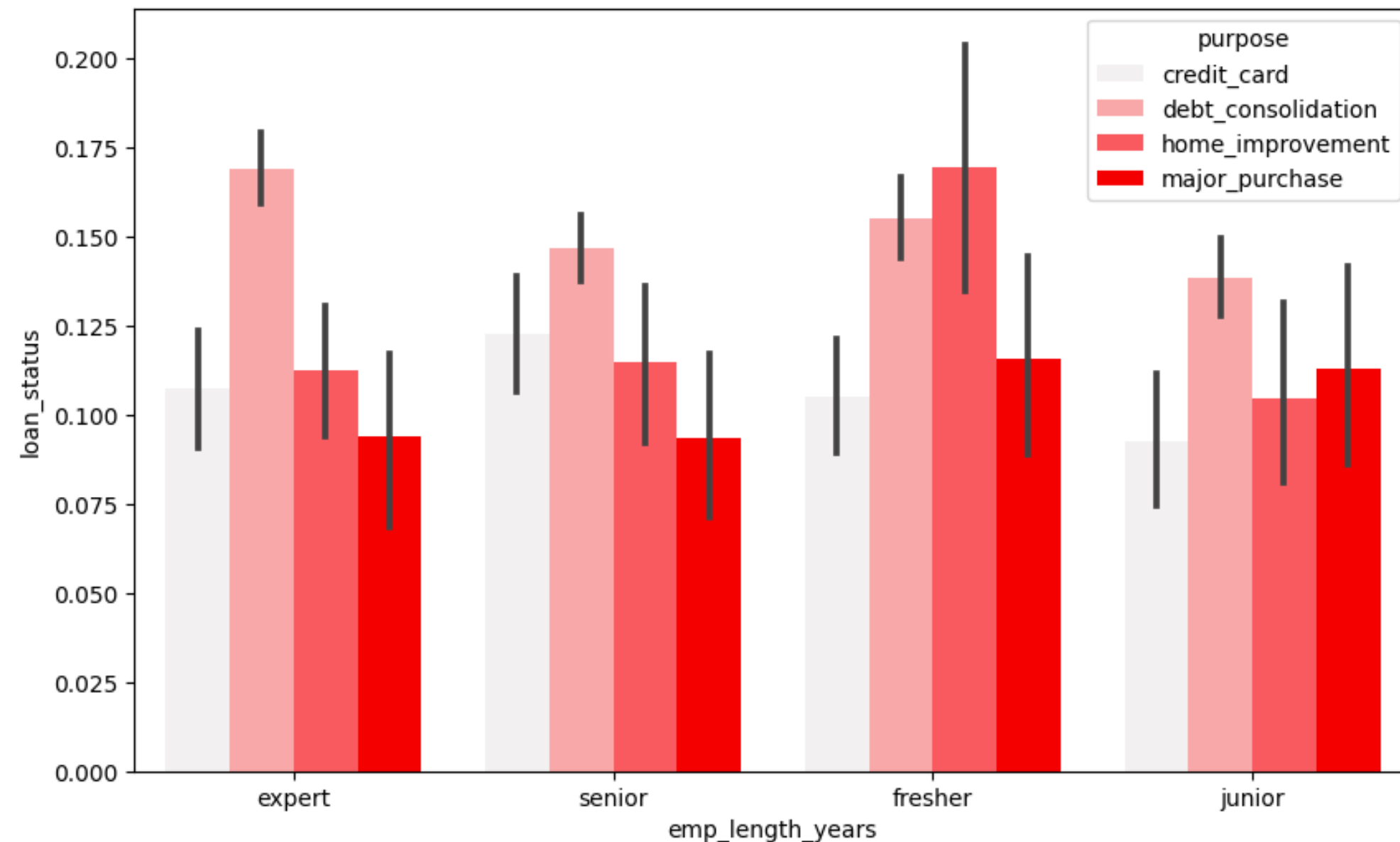
- Loan status vs dti (*hue = purpose*)



From the graph, we can observe the high **dti** of the applicant and the purpose of debt_consolidation being the reason for high defaulters.

Segmented analysis

- Loan status vs emp_exp (*hue = purpose*)



In all the 4 bins we can see the debt_consolidation and home improvement being the top reasons for defaulters.

Conclusions

The major variables which are the determining factors for an applicant to be a defaulter.

- Interest rate
- Grade, Subgrade
- Loan Amount
- Annual Income
- Installment
- Duration
- DTI (debt to income ratio)



Summary

The users with High interest, Low grades or subgrades, high loan amounts, Low annual income, high dti, and users with the purpose of debt consolidation or home improvement are more likely to be the defaulters.





**THANK
YOU**

