

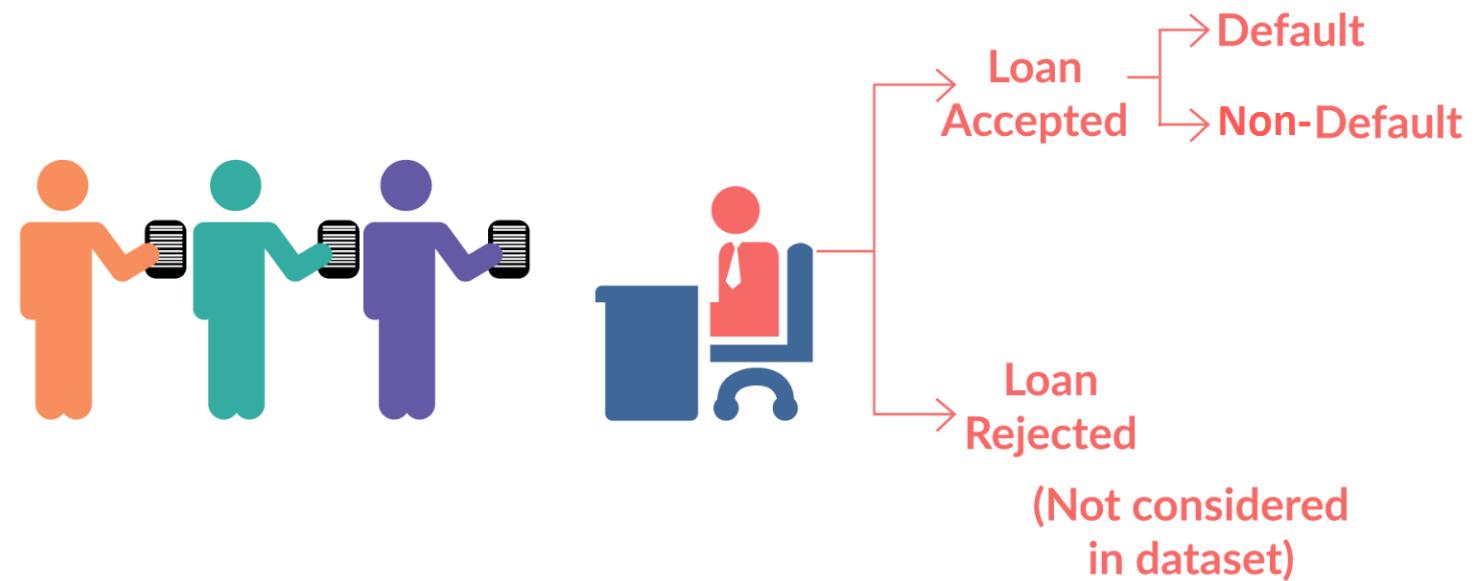
LENDING CLUB CASE STUDY





Goal of this Case Study is to gain an industry experience on how real business problems are solved using EDA (Exploratory Data Analysis). In this case study, apart from applying the techniques in EDA, also developed a basic understanding of Risk Analytics in Banking and Financial Services and understand how data is used to minimise the risk of losing money while lending to customers.

LOAN DATASET



This case study comes from a consumer finance company which specialises in lending various types of loans to urban customers. 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:

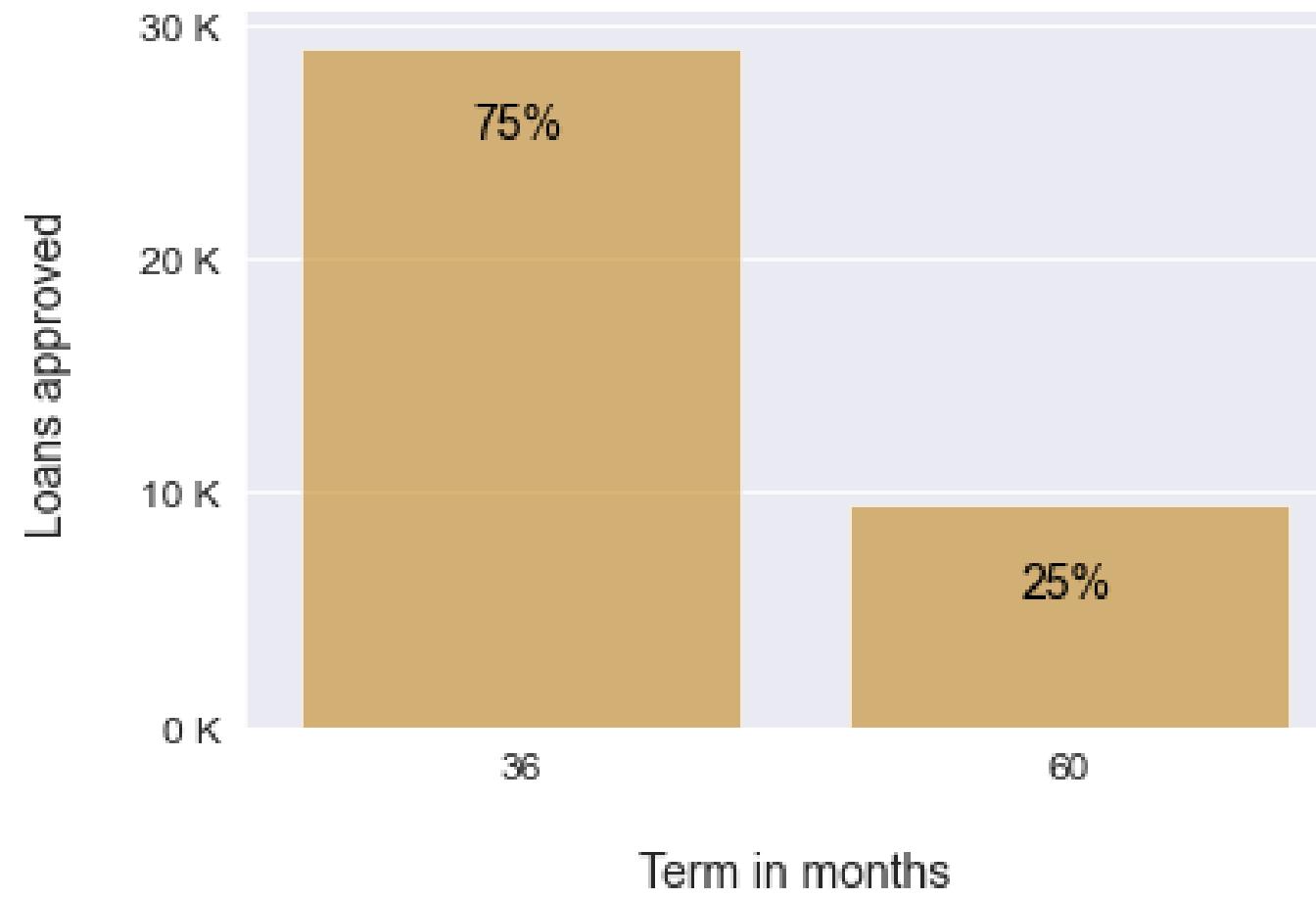
1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
2. If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

Data Cleaning

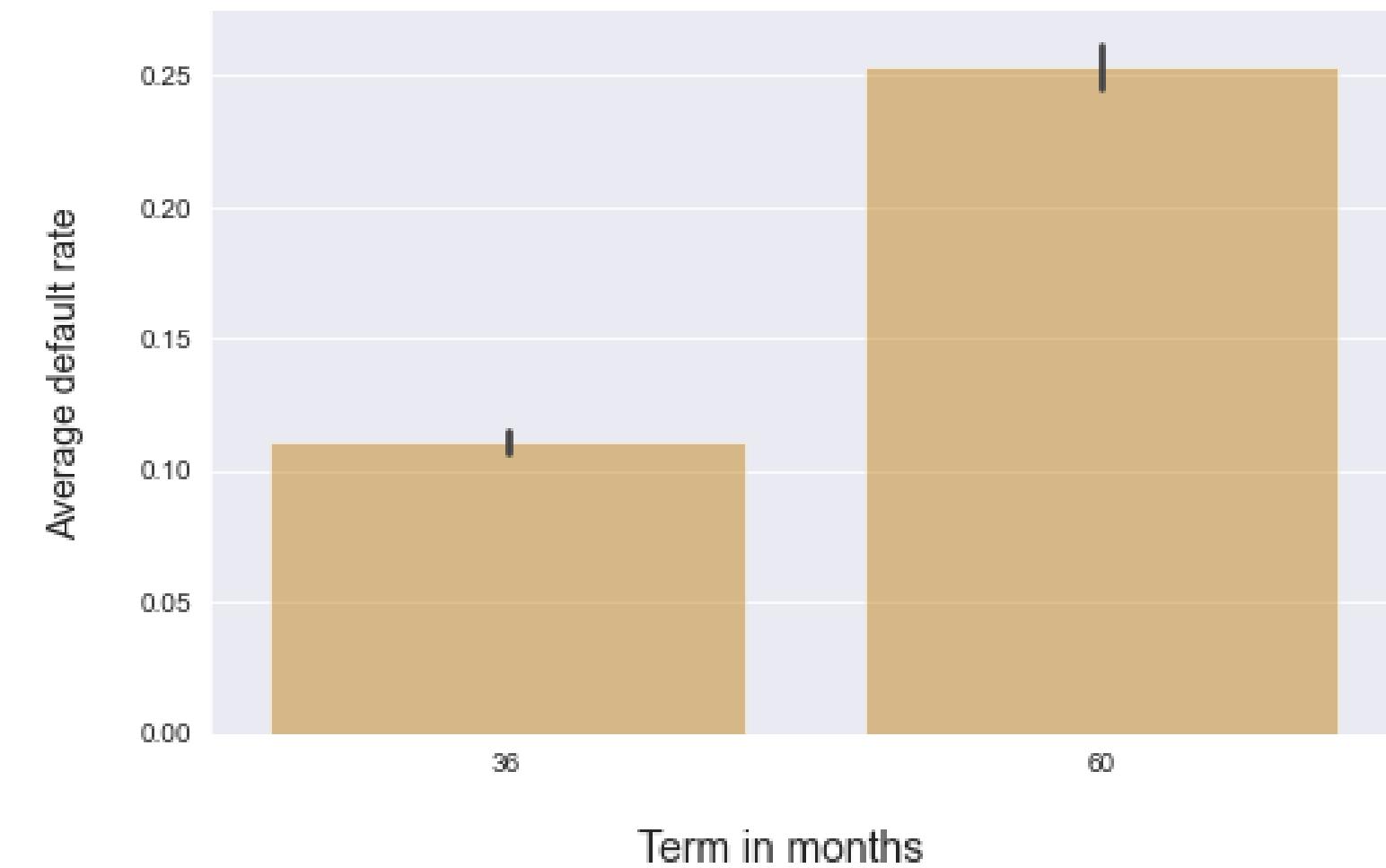
1. The customer behavior variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval. Hence it is safe to remove customer behavior variables from the analysis.
2. Removed columns with no statistical significance such as with only one value.
3. Removed columns with high null values.
4. Dropped redundant columns.
5. Standardized the values across columns.
6. Removed columns highly correlated to other columns.

Univariate Ananlysis

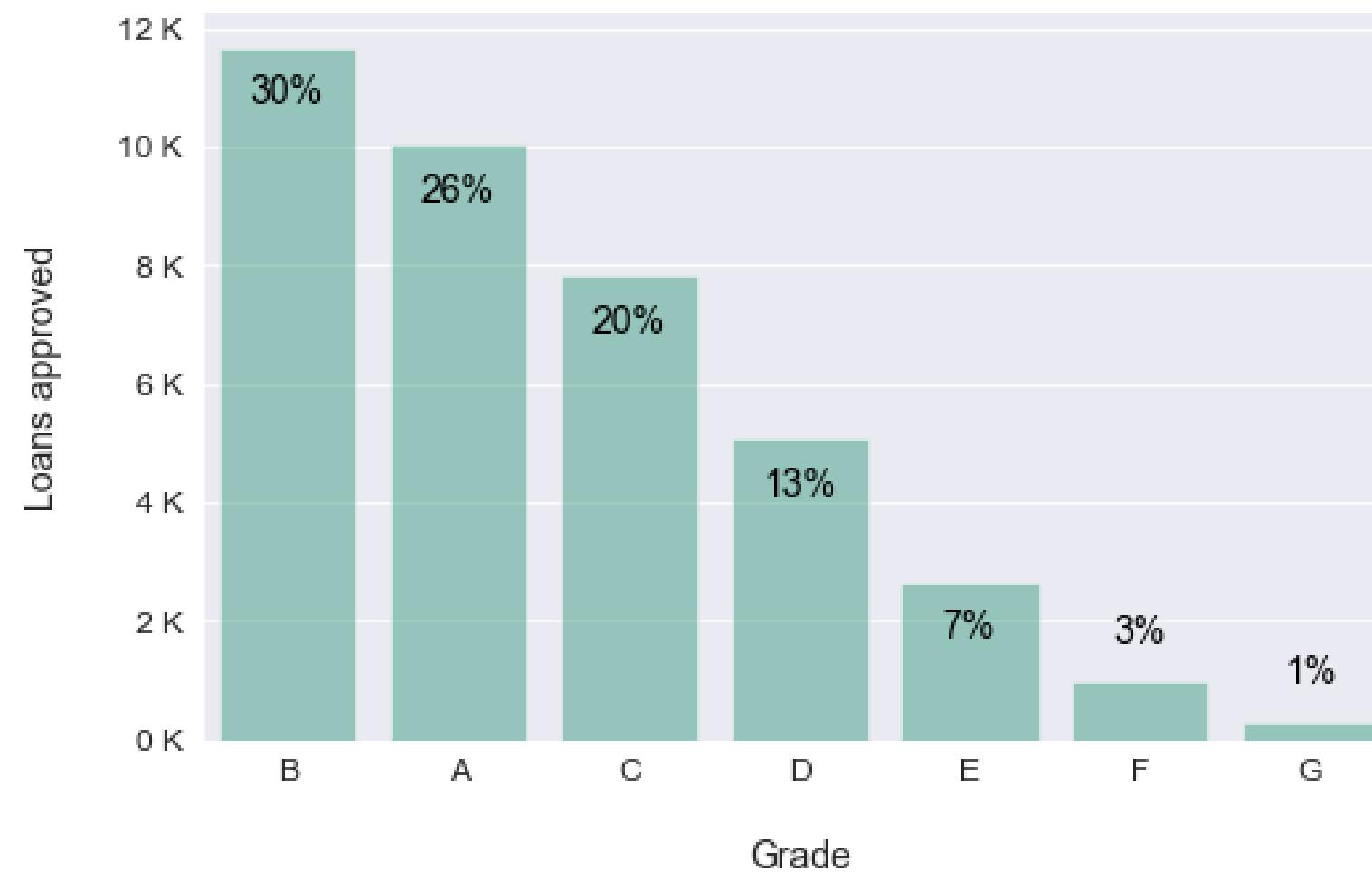
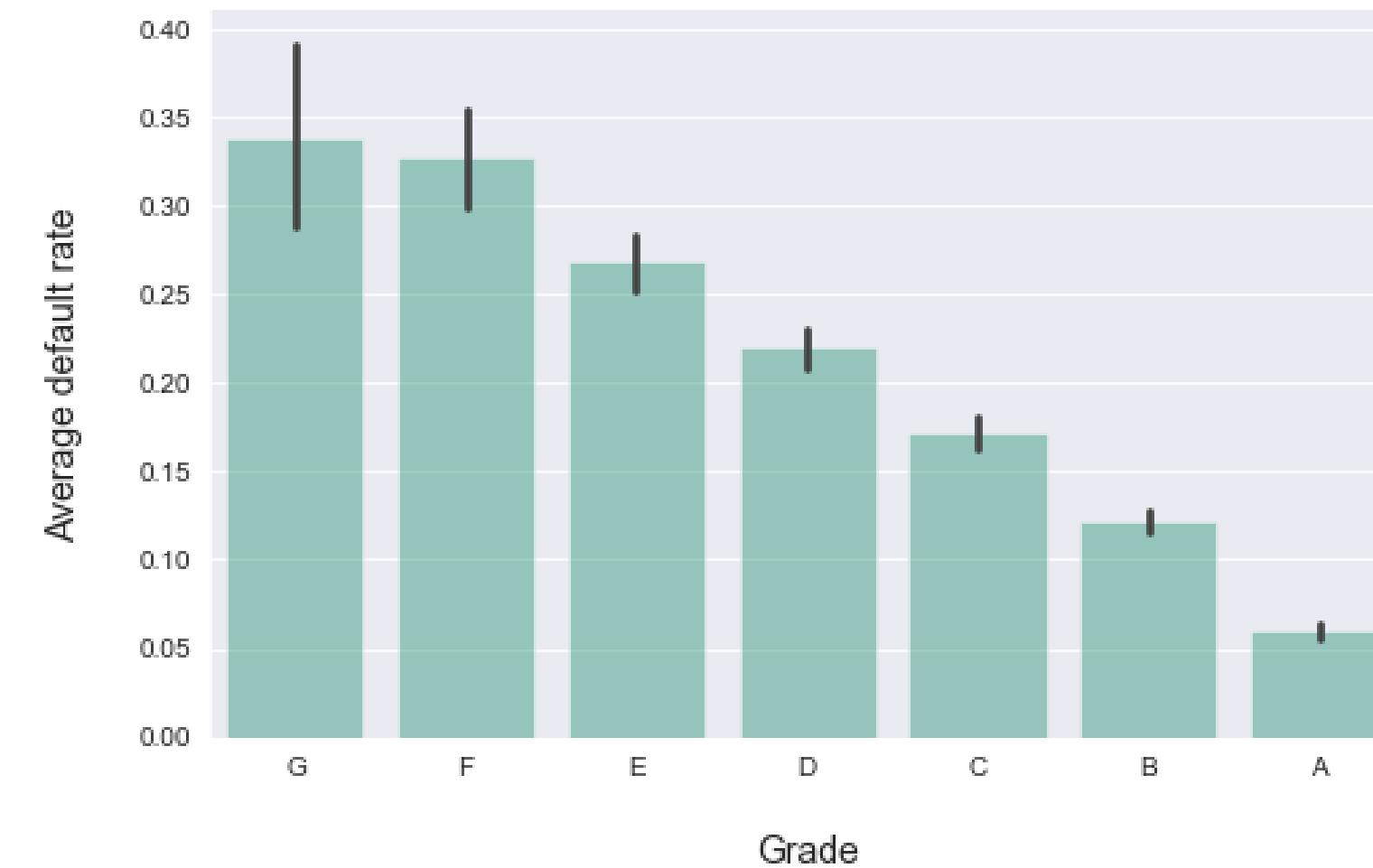
Aggregated count of loans approved based on Term



Average default rate based on term

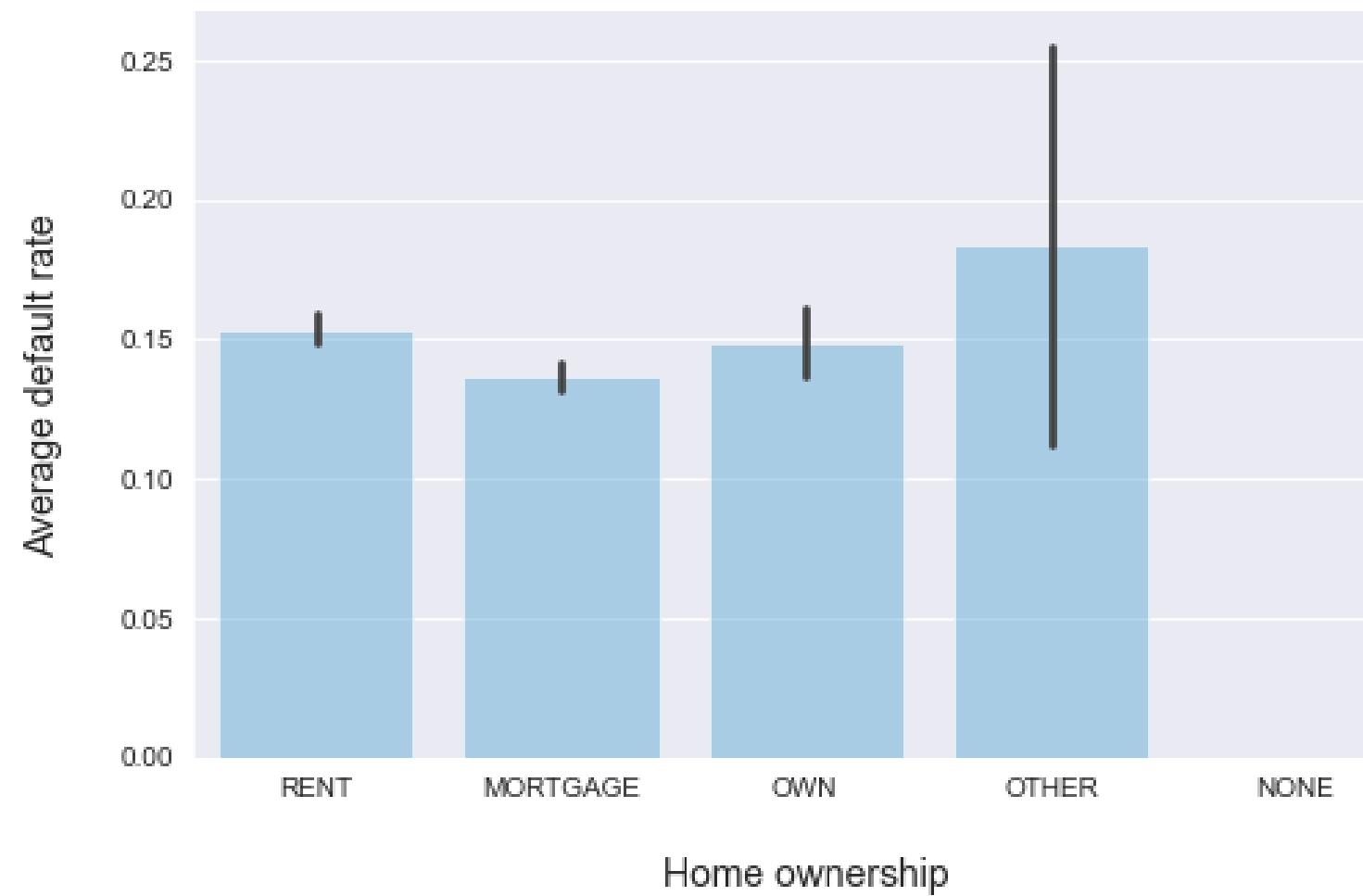


More loans were given to applicants with short term of tenure.
Risk of default is more for applications with more tenure/term duration.

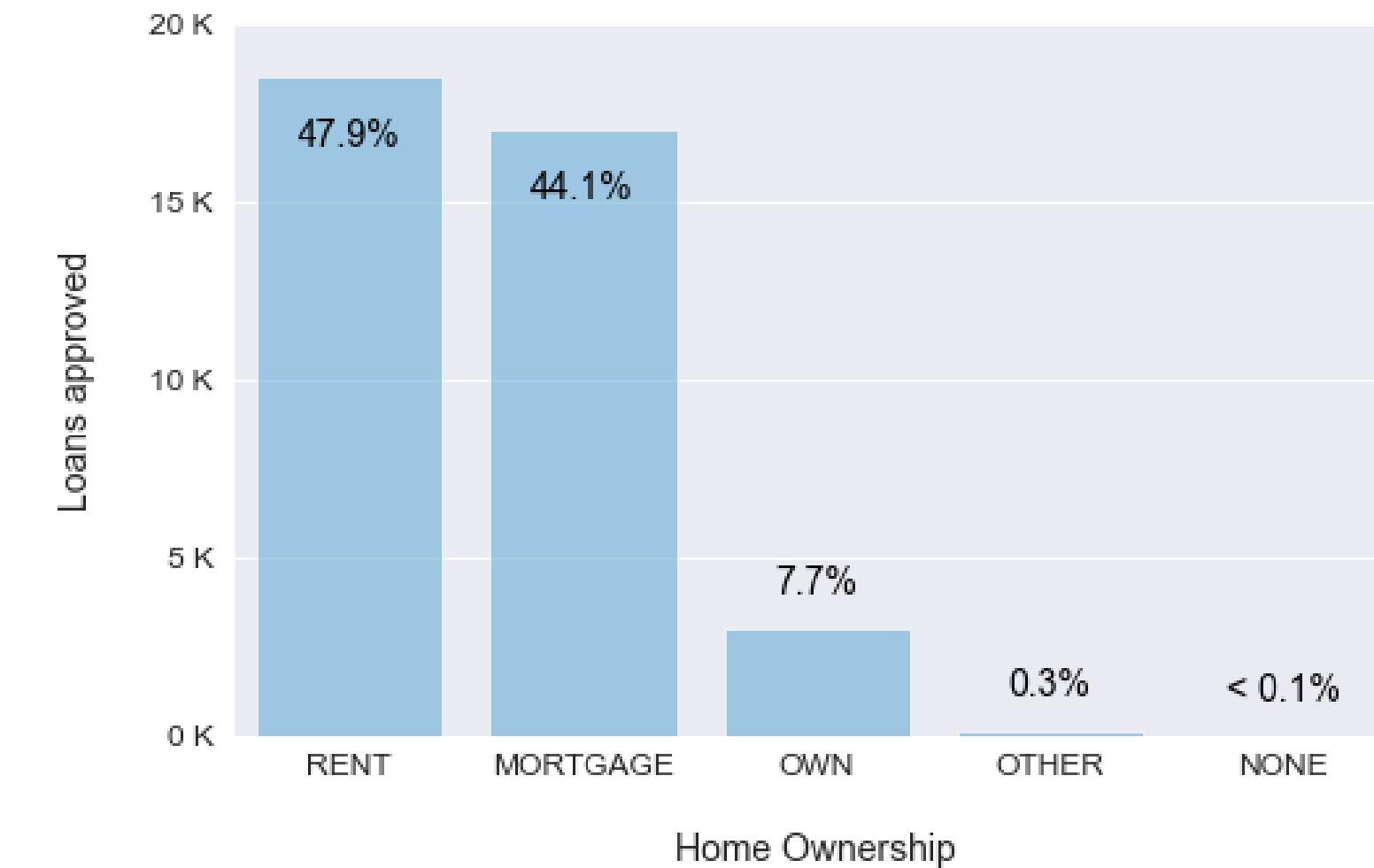
Aggregated count of loans approved based on Grade**Average default rate based on Grade**

More loans are given to High grade applications.
Evidently low grade applications have high risk of getting default.

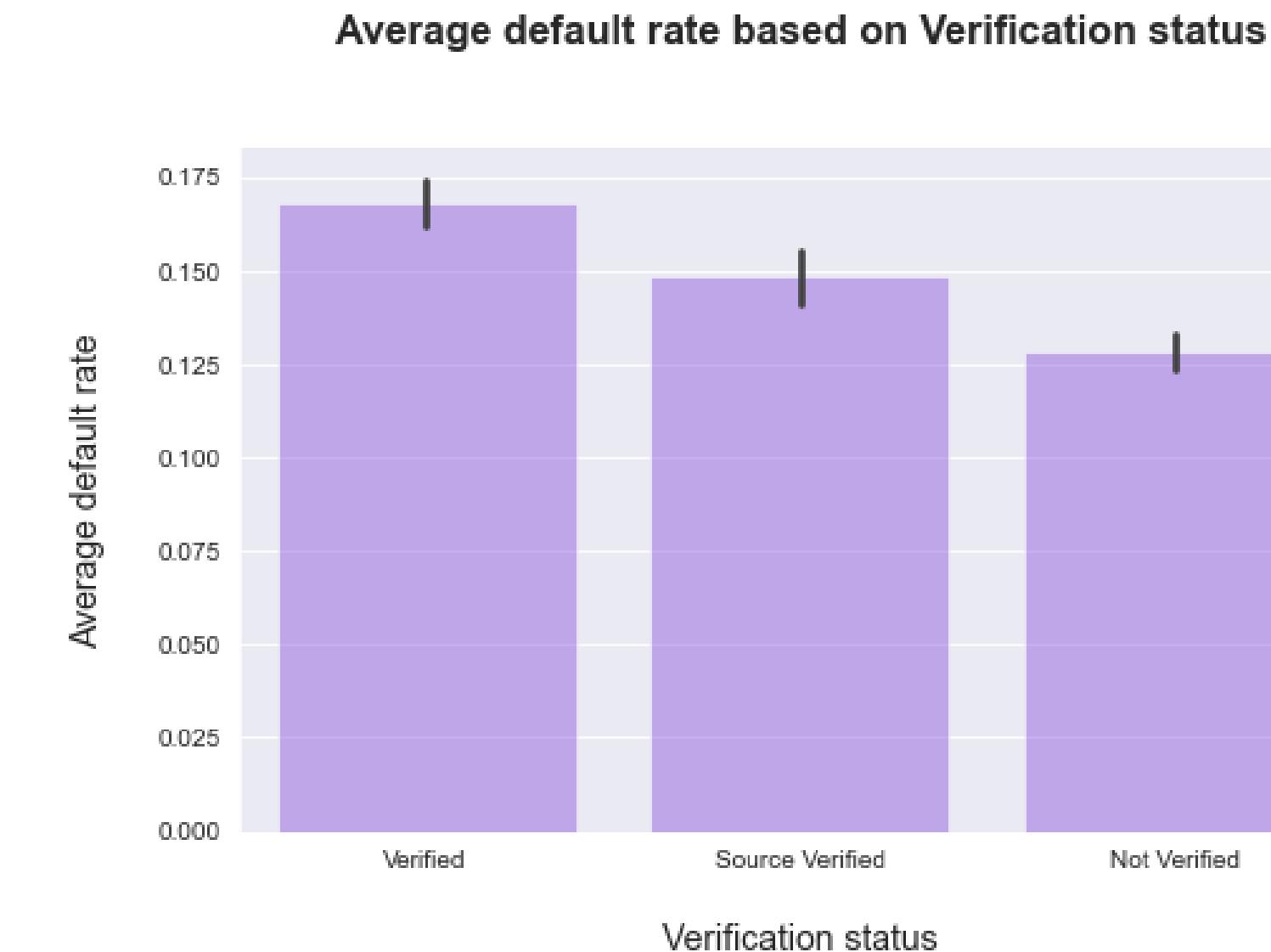
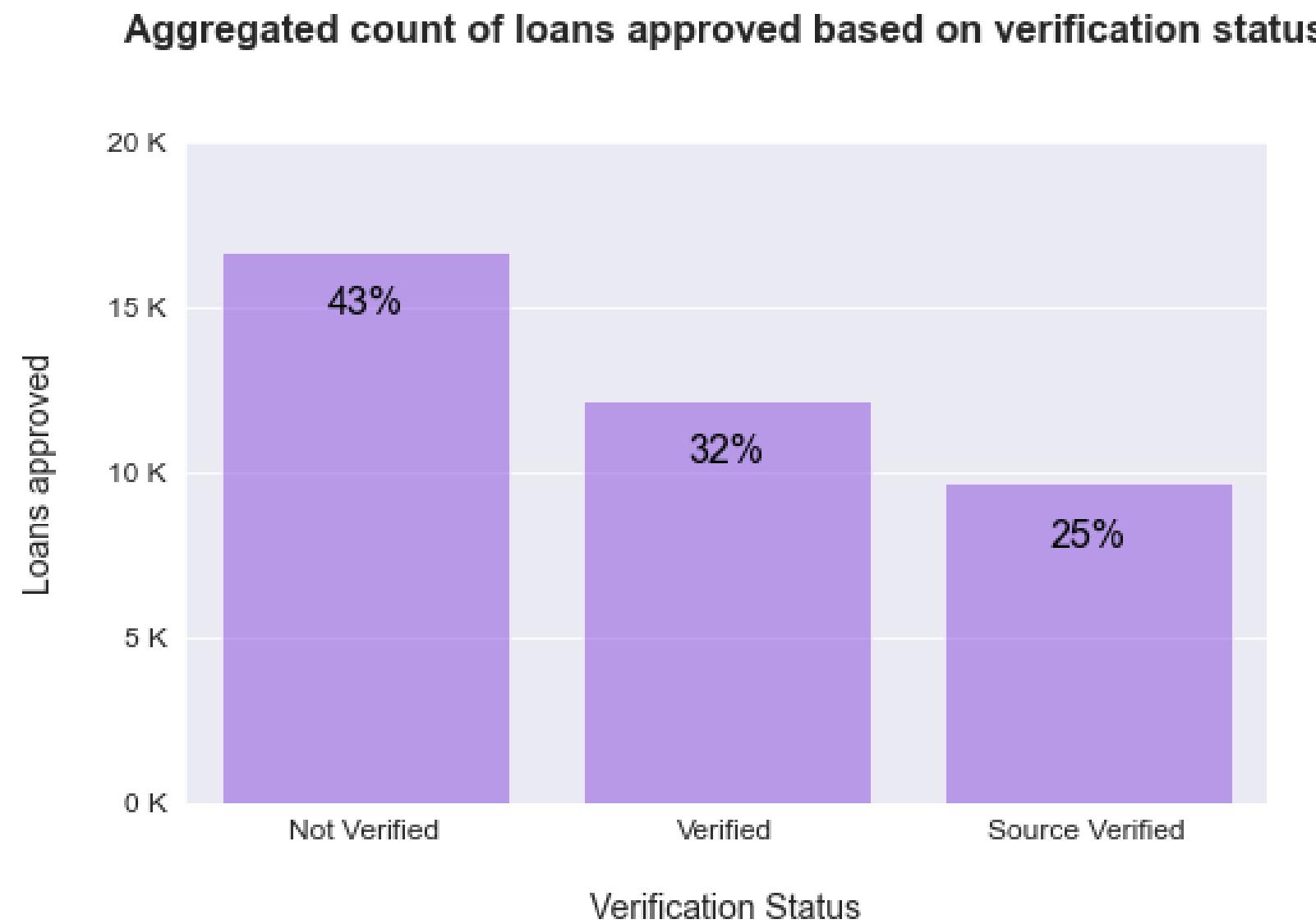
Average default rate based on Home ownership



Aggregated count of loans approved based on Home Ownership

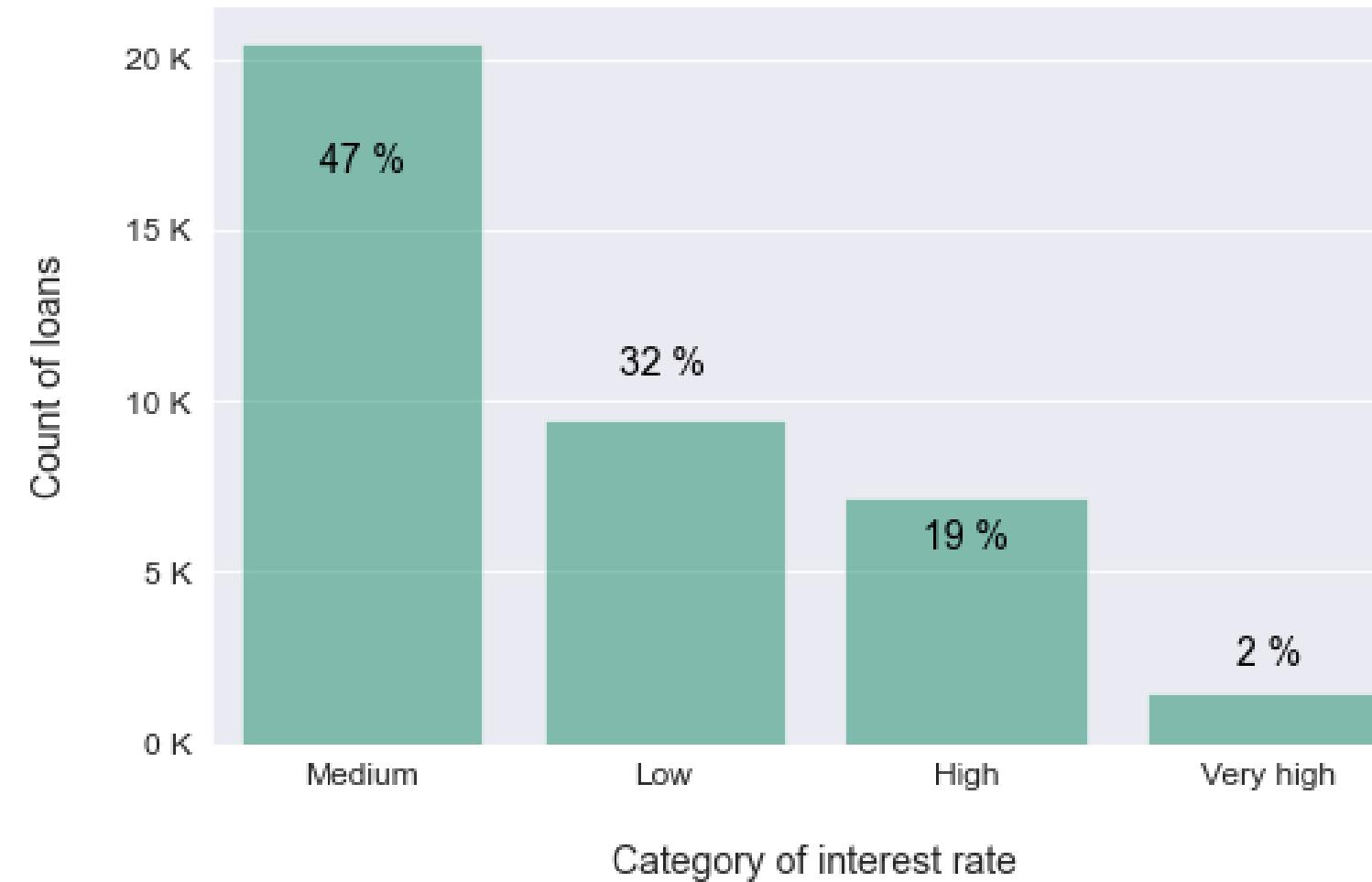


surprisingly Rent house and mortgage house owners were given more loans however, risk is same for almost every category of home ownership.

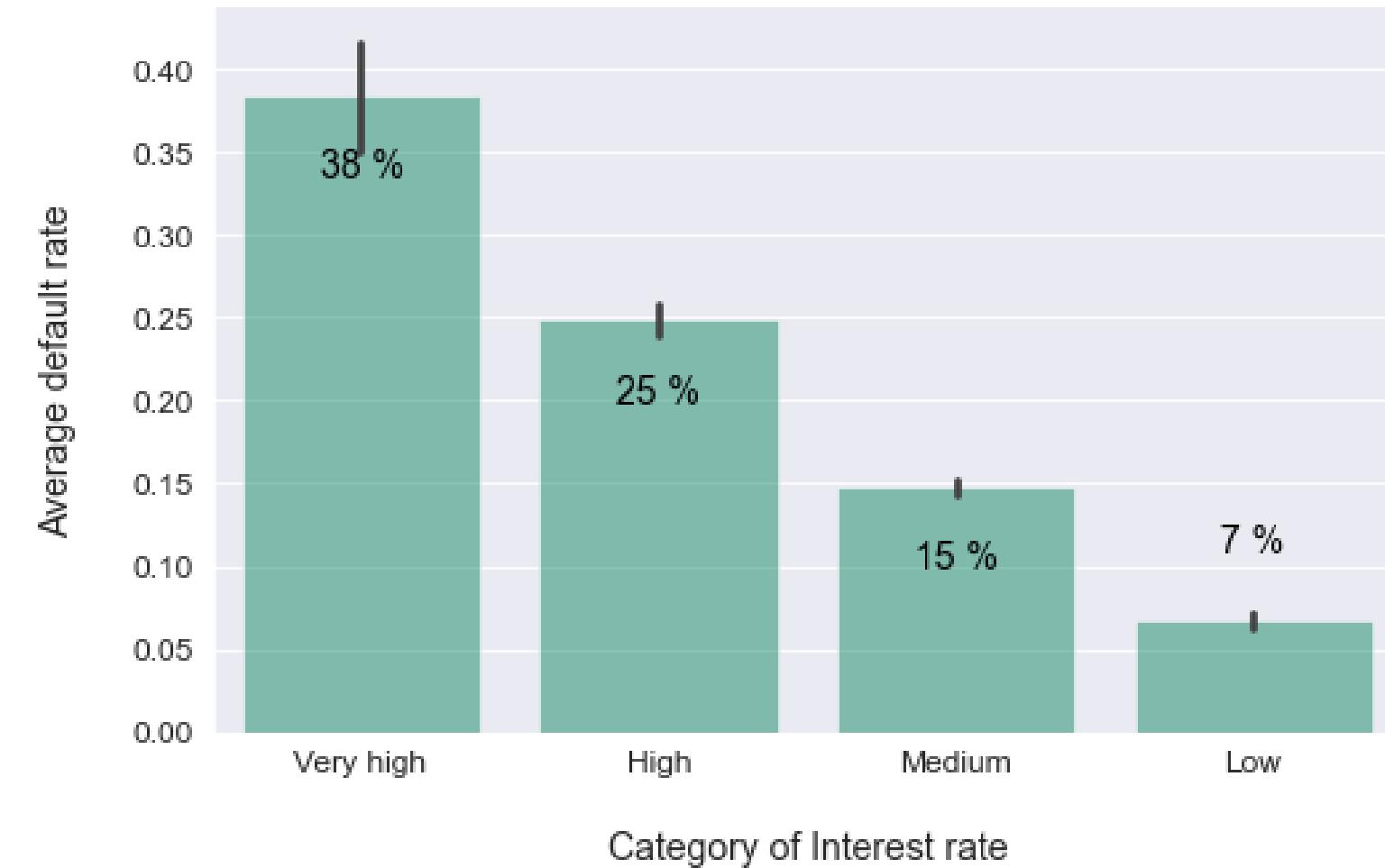


Non verified sources were approved more loans.
and surprisingly risk of default is less for nor verified sources.

Categorized count of loans approved based on interest rate

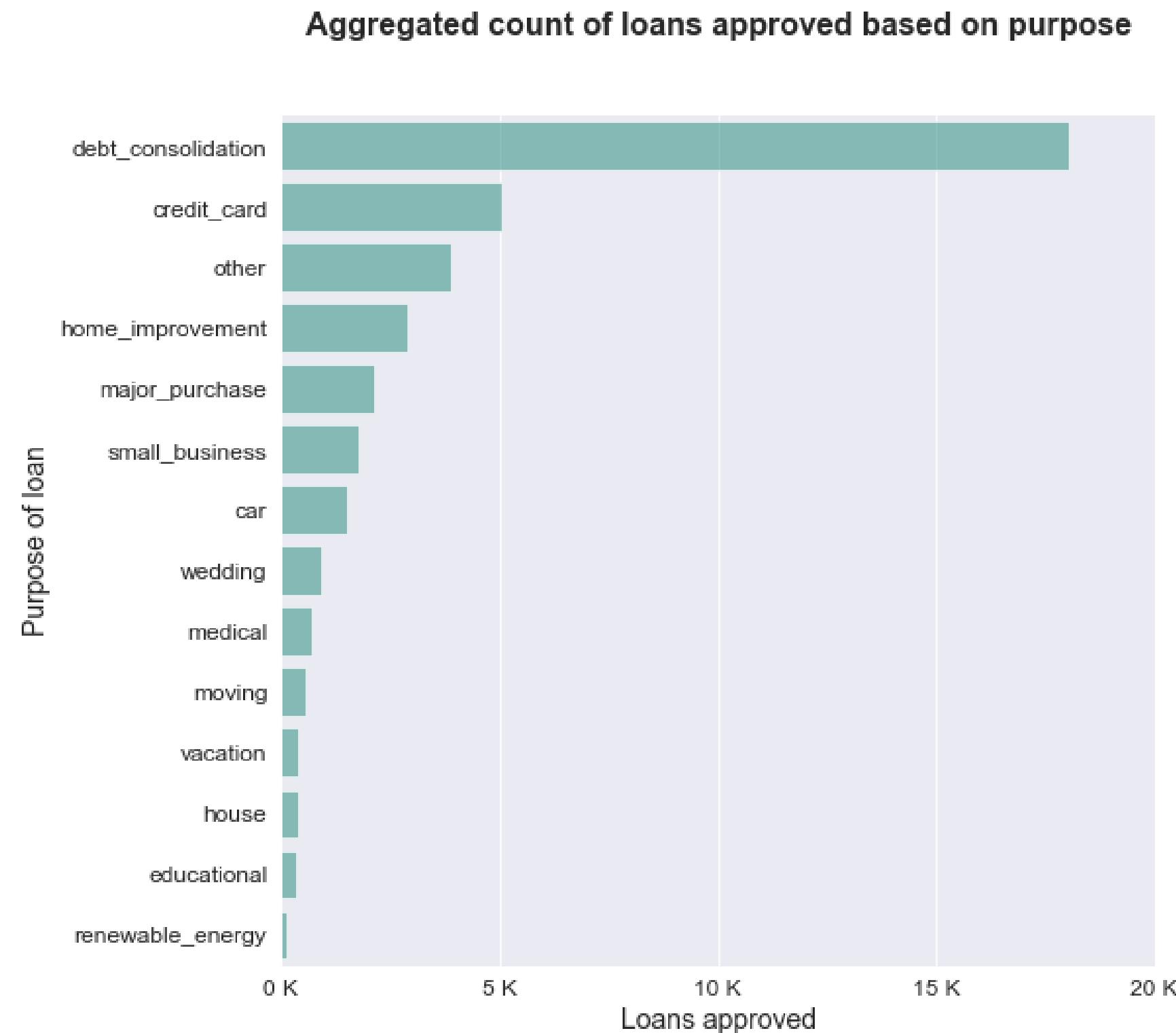


Average default rate based on Interest rate type



The above plot depicts that the approvals with very high interest rates are most likely to default compared to other categories. While approvals with low and medium interest are less likely to default. Hence it is a wise choice to approve more loans with low and medium interest rates and reduce any approvals which require high and very high interest rates to reduce the risk of credit loss.

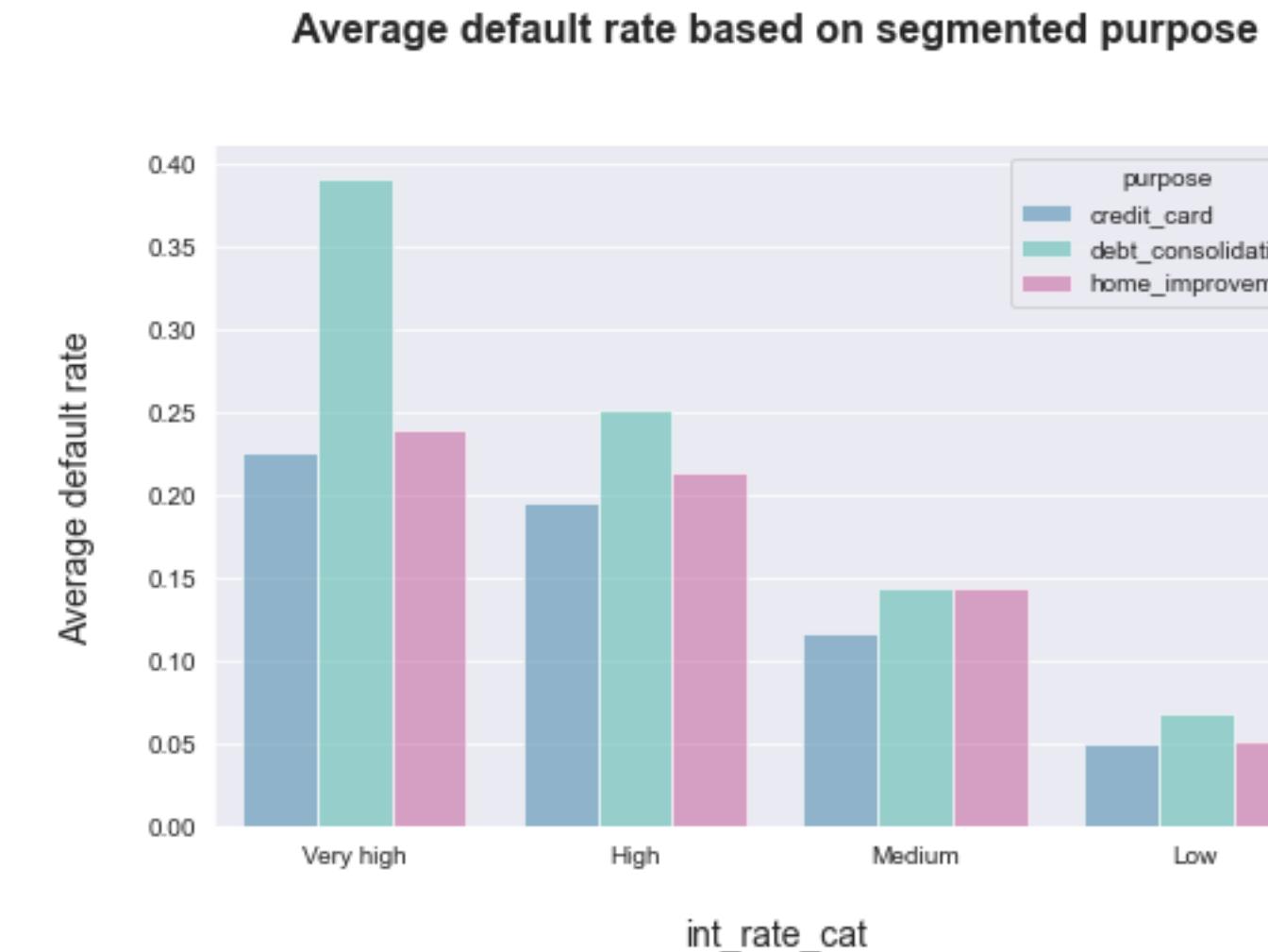
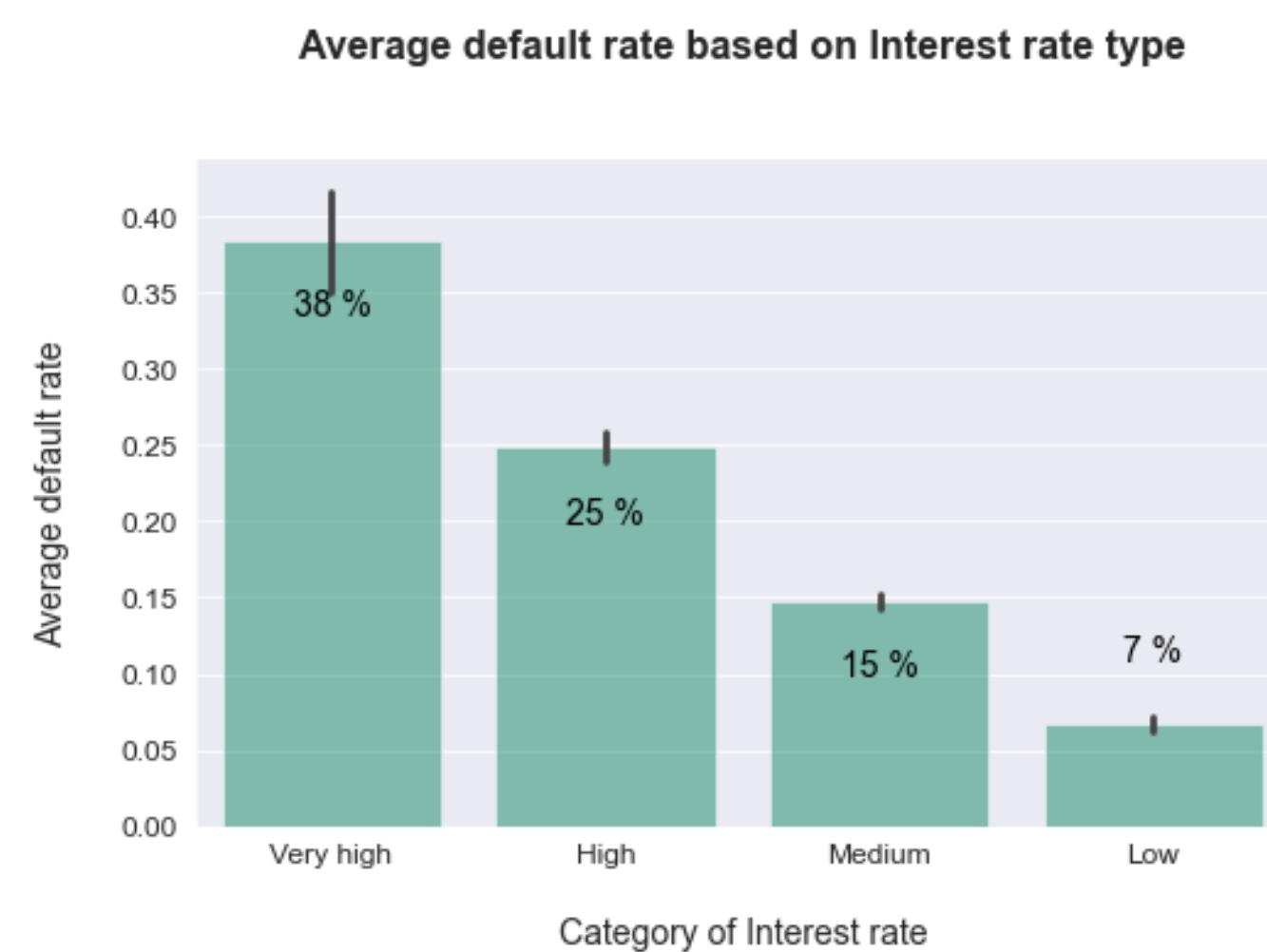
Important factors



Excluding the category of other, the three major purposes of taking loans are:

1. debt_consolidation,
2. credit_card
3. home_improvement

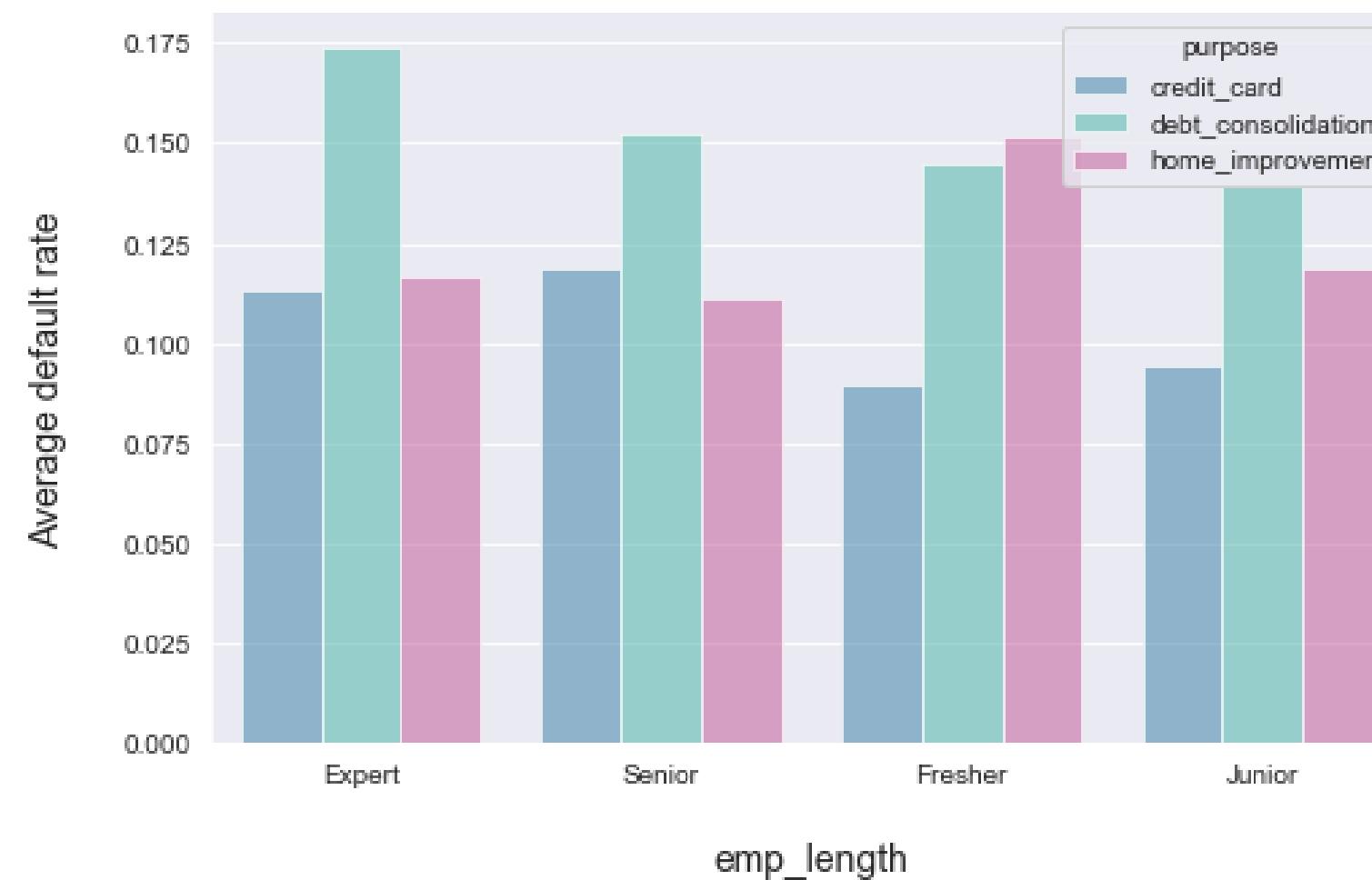
Segmented Univariate Ananlysis



The above plot depicts that the approvals with very high interest rates are most likely to default compared to other categories. While approvals with low and medium interest are less likely to default. Hence it is a wise choice to approve more loans with low and medium interest rates and reduce any approvals which require high and very high interest rates to reduce the risk of credit loss.

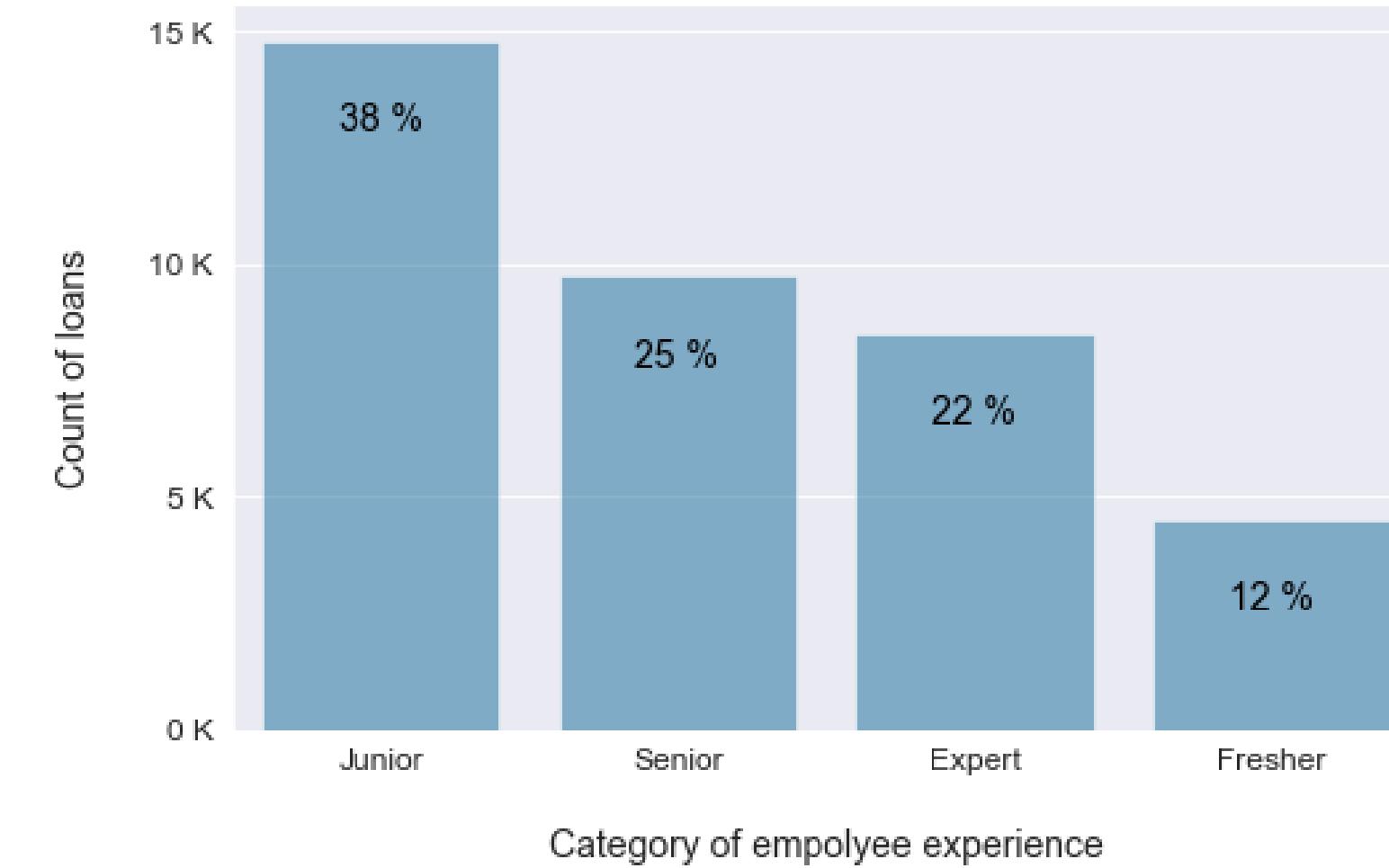
Default rate of Loans taken on debt consoliation, increases with increase in interest rate.

Average default rate based on segmented purpose



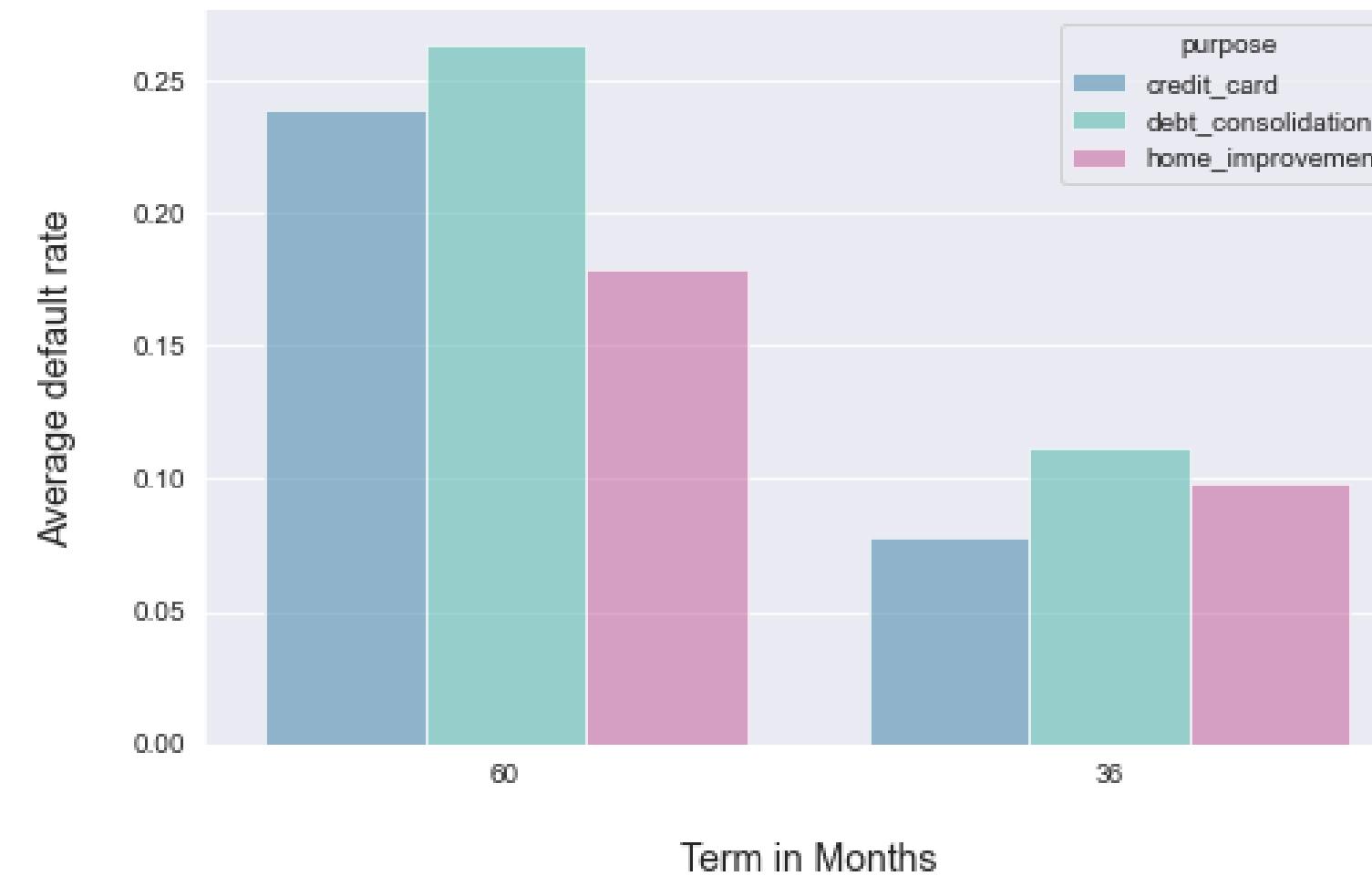
Home improvement loans are more tend to default in all categories. For freshers it is risky to give loans for home improvement.

Categorized count of loans approved based on employee experience

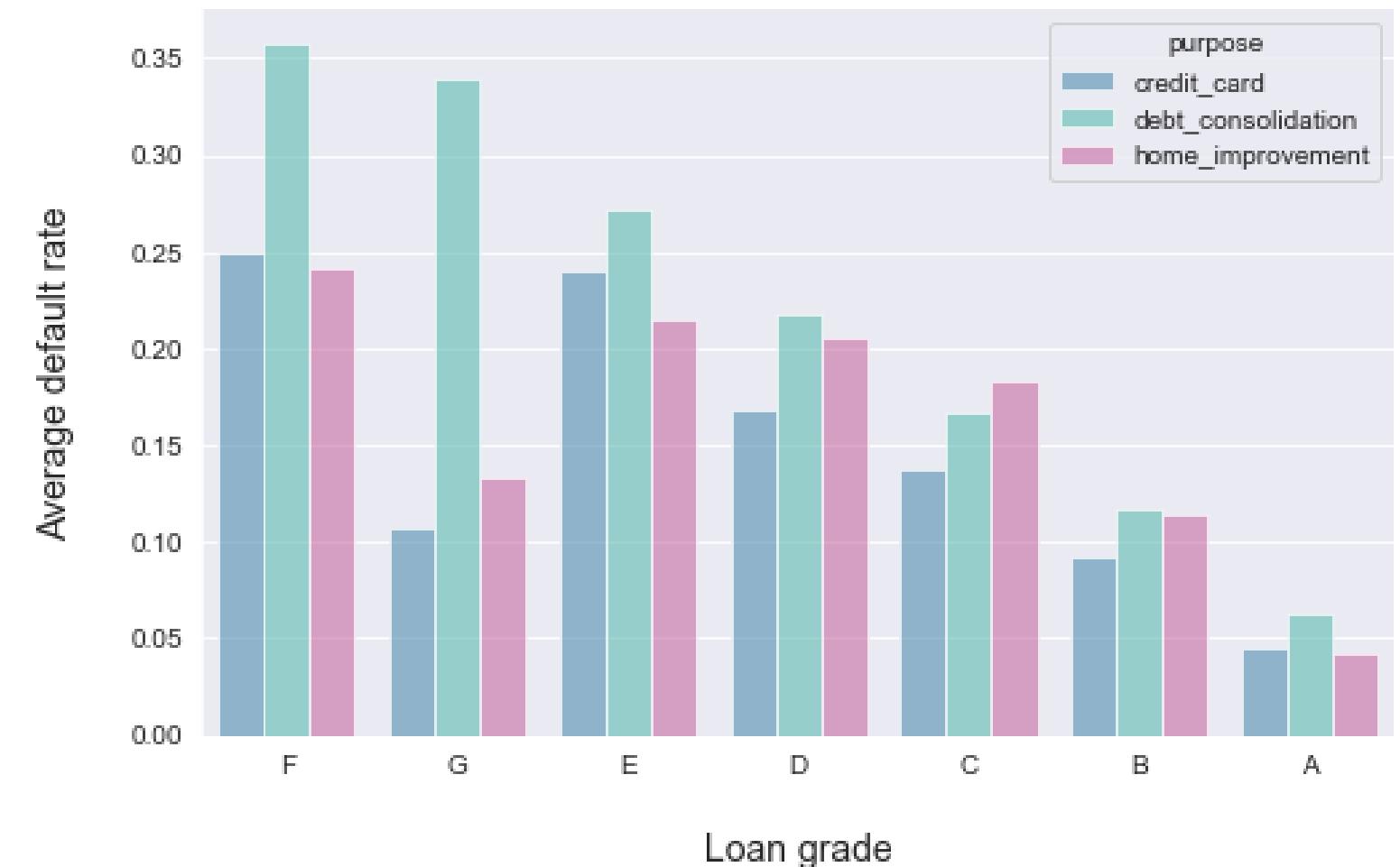


The above plot depicts that the employee with junior level experience has more chances of getting his loan approved compared to fresher level. And the chances decrease over increase in experience. This may be because of decrease in service period.

Average default rate based on segmented purpose



Average default rate based on segmented purpose

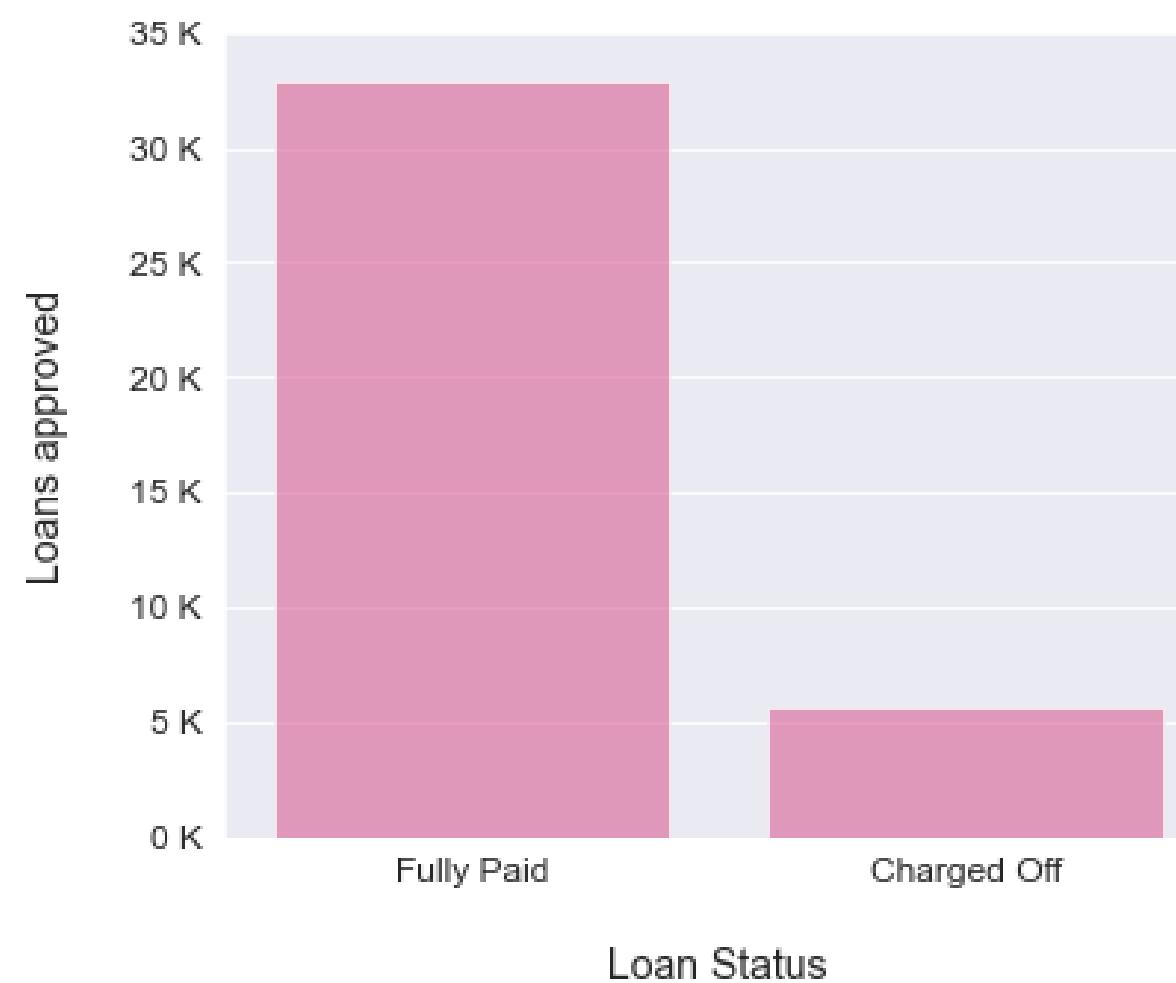


Long term loans are more likely to get defaulted, and especially loans taken for credit card and debt consolidation has high risk of default.

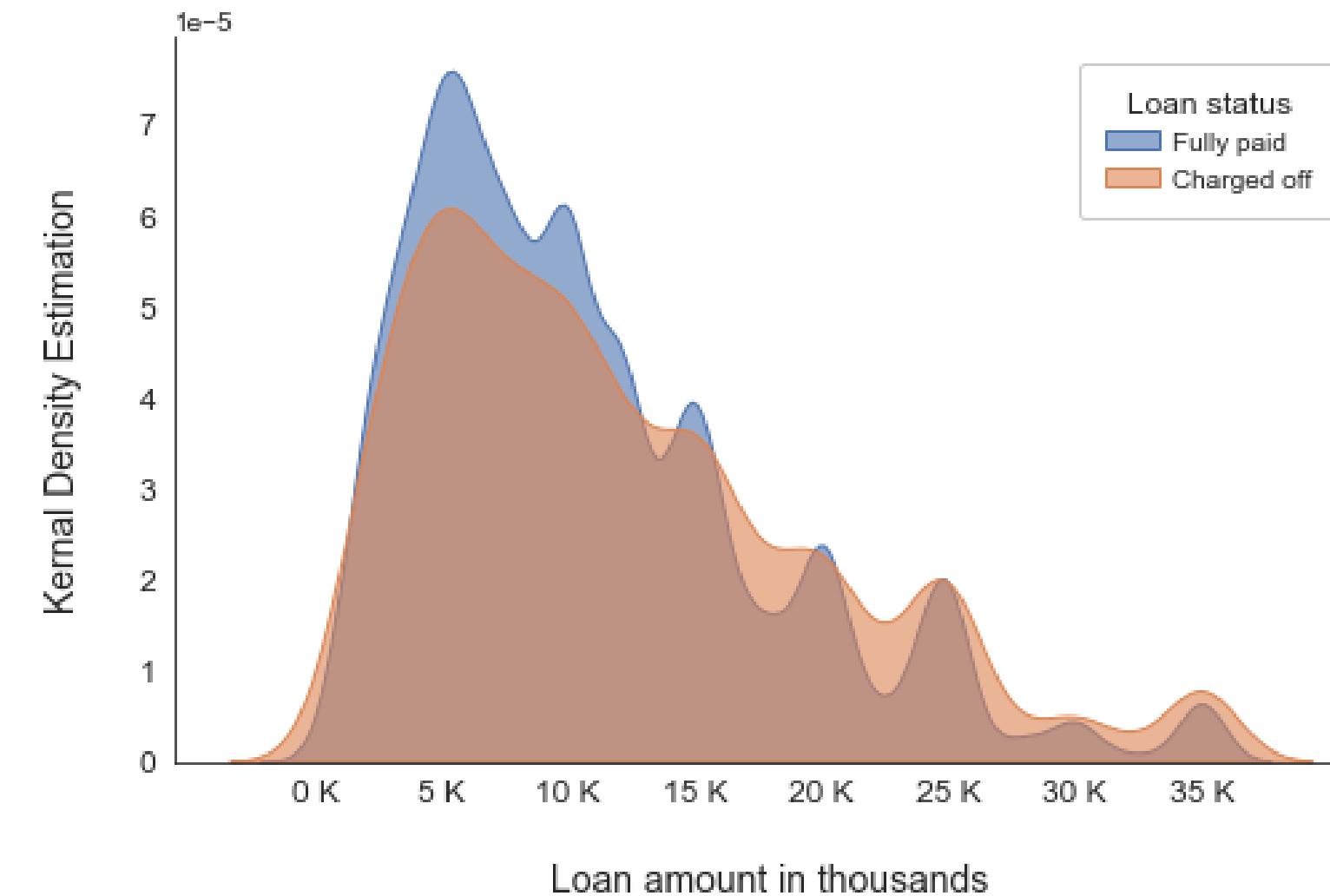
Debt consolidation has the higher chances of default as the grade of loans decreases. The same trend is observed in the credit card purpose.

Bivariate Ananlysis

Aggregated count of loans approved based on loan status



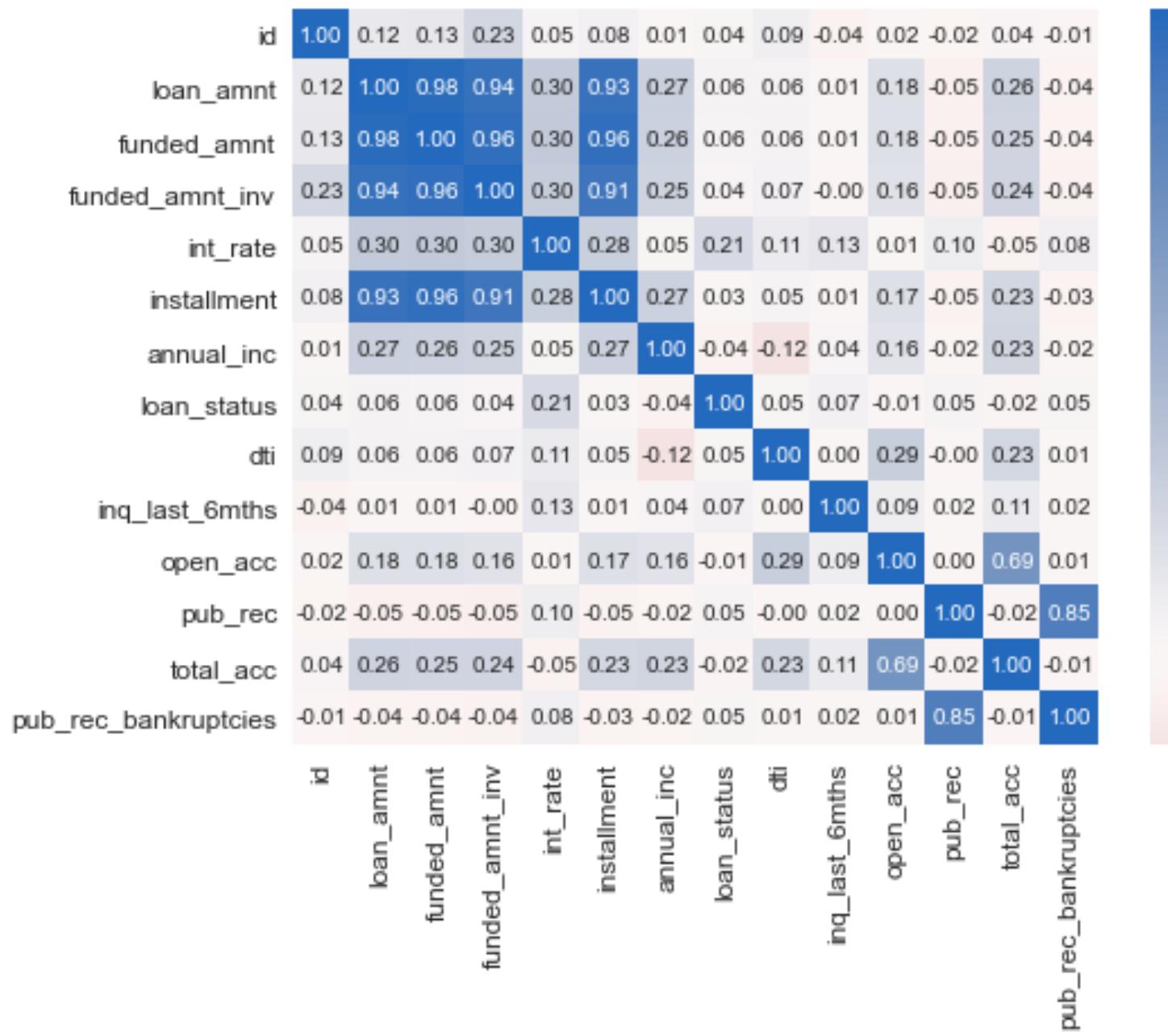
Distribution of loan amount



It is evident that loans with higher amout are more likely to default compared to smaller loans. Let us see whether company is leveraging this insight.

Multivariate Ananlysis

Loan dataset - correlation map



Heat map is used to analyze the correlation between variable. This provides a base understanding of the interrelated data.

Heatmap of numerical variables

Conclusion

The company can decrease the credit risk using the insights generated from the Exploratory data analysis. The analysis not only provides the trends but the key insights that effect the default rate for the loan approvals. This better equips the company to make informed decisions based on historical data to drive the business to more profit path.



Credits

Please find the team details below:

Team Member Detail:

- Name: Vamsi Krishna P
- Email ID: vamshi.krishna.prime@gmail.com
- Phone no: 9491392912

Group Facilitator:

- Name: Richa
- Email ID: richabhardwaj897@gmail.com
- Phone No: 919560394072