Presentation on EDA Case Study

By: Pragati Priya

Purpose

- Credit risk analysis will help the company to make a decision for loan approval based on the applicant's profile. Which controls loss of business to the company and avoid financial loss for the company.

Steps

- 1.Dataunderstanding and sourcing
- 2.check for Data quality issues and Binning
- 3.check for Data imbalance and univariate,segmented univariate & Bivariate analysis,correlation
- 4.Merging of application data with previous application data
- 5.Data analysis by univariate, segmented univariate , Bivariate analysis and correlation
- 6.Recommendations and Risks

Data Cleaning

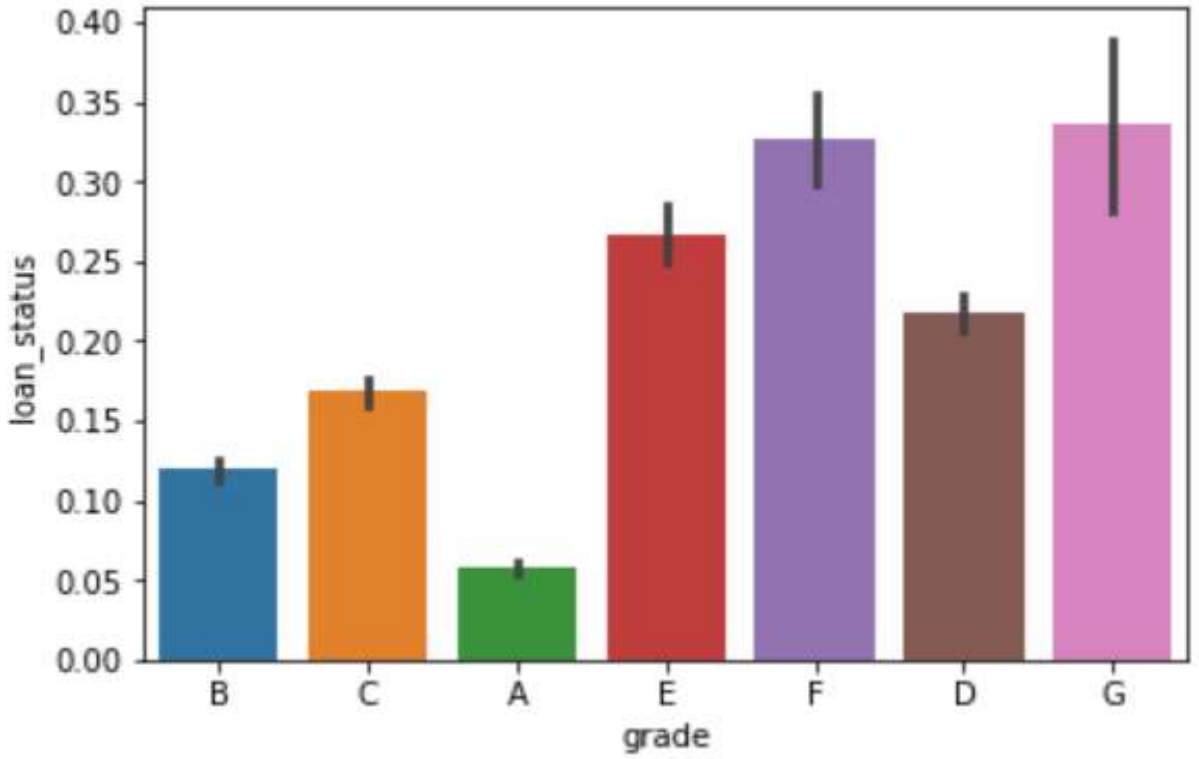
The data is analyzed for the null values. The columns having null values are dropped and after that percentage of null values are cross checked using **df.isnull().sum()** command.

Data Analysis

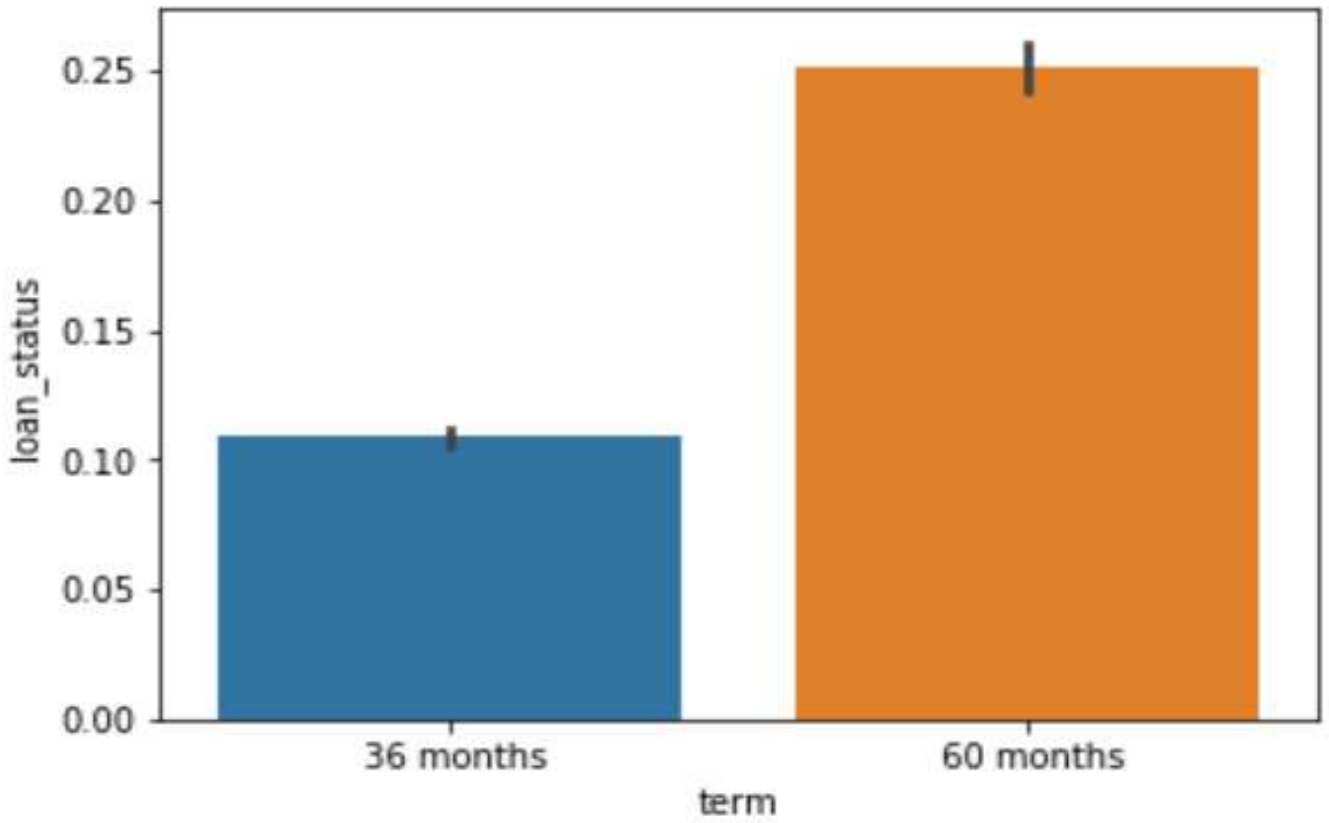
- The objective is to identify predictors of default so that at the time of loan application, we can use those variables for approval/rejection of the loan. There are broadly three types of variables - 1.Those which are related to the applicant (such as age, occupation, employment details etc.), 2.Loan characteristics (amount of loan, interest rate, purpose of loan, etc.) and 3.Customer behaviour variables (which are generated after the loan is approved e.g. delinquent 2 years, revolving balance, next payment date, etc.). Now, the customer behaviour variables are not available at the time of loan application, and thus they cannot be used as predictors for credit approval.

Univariate Analysis

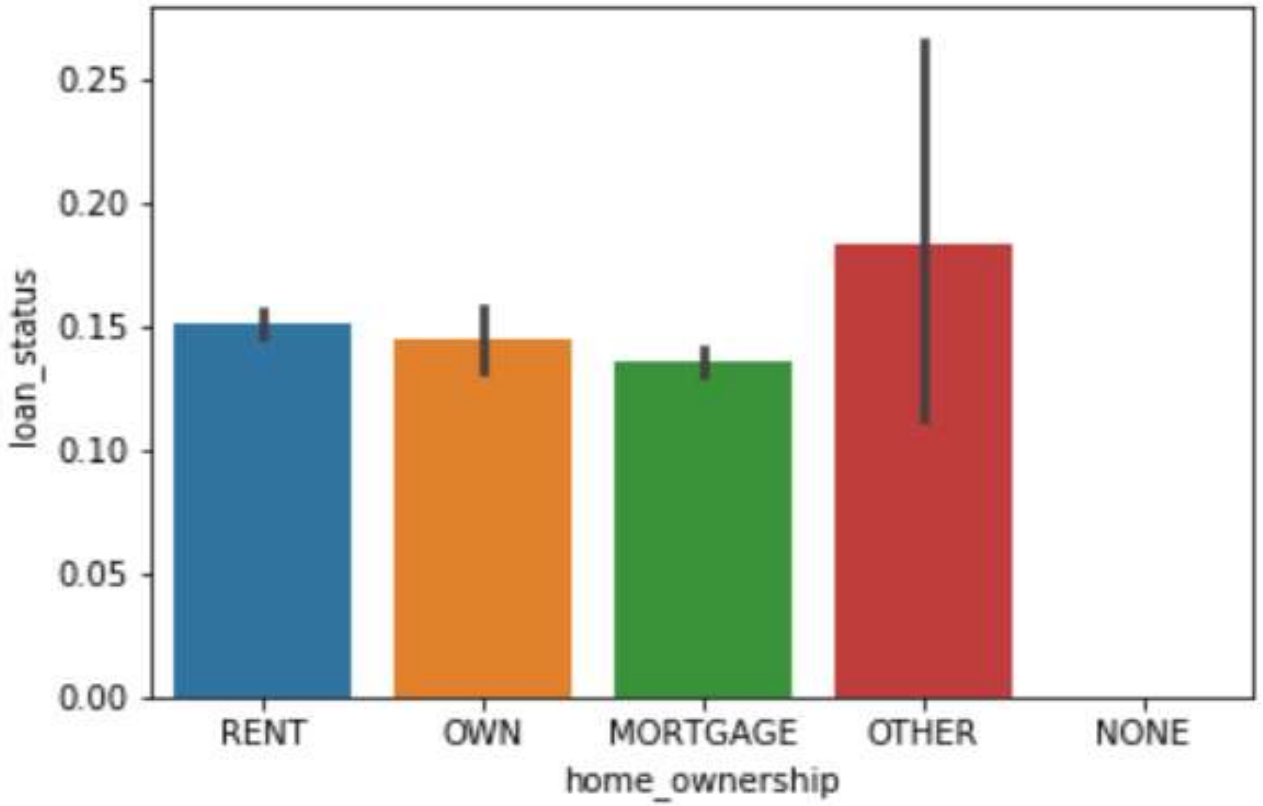
Plotting default rates across grades of the loan:



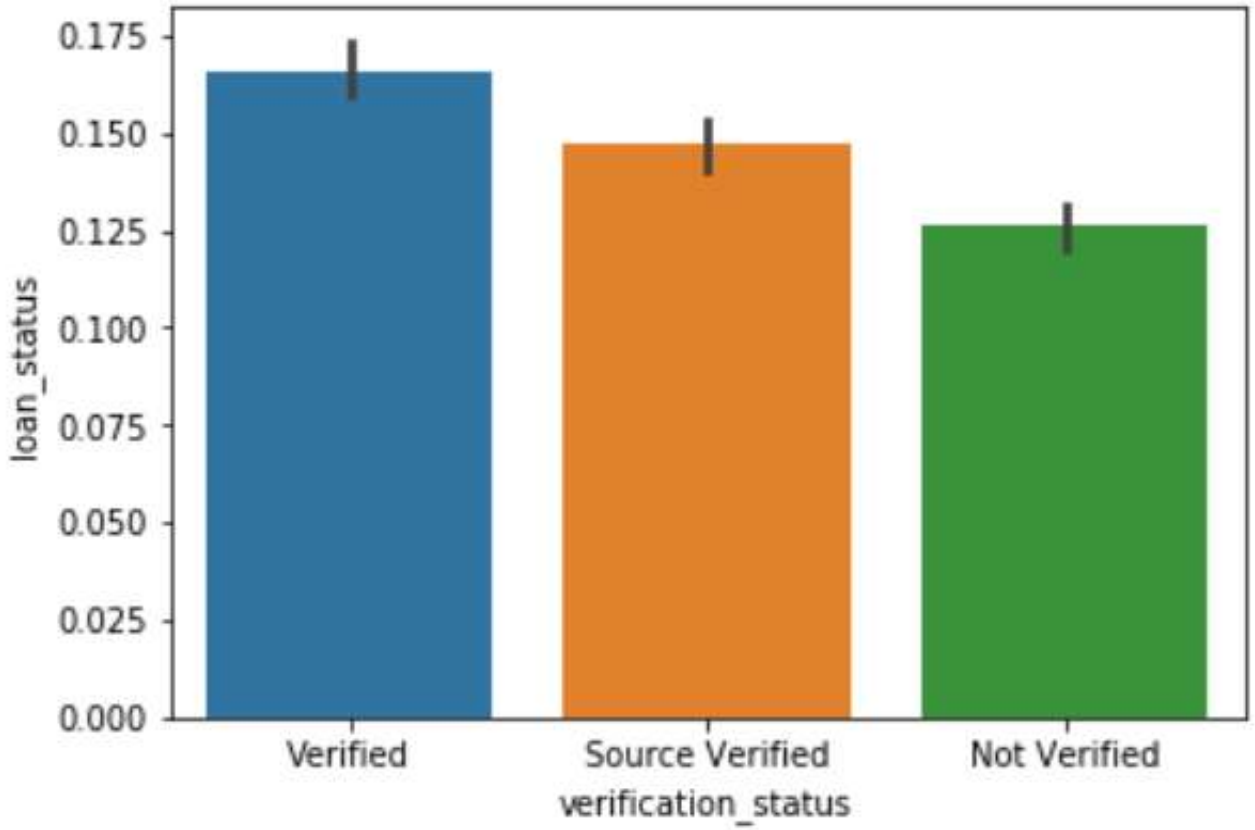
Plot term across loan status:



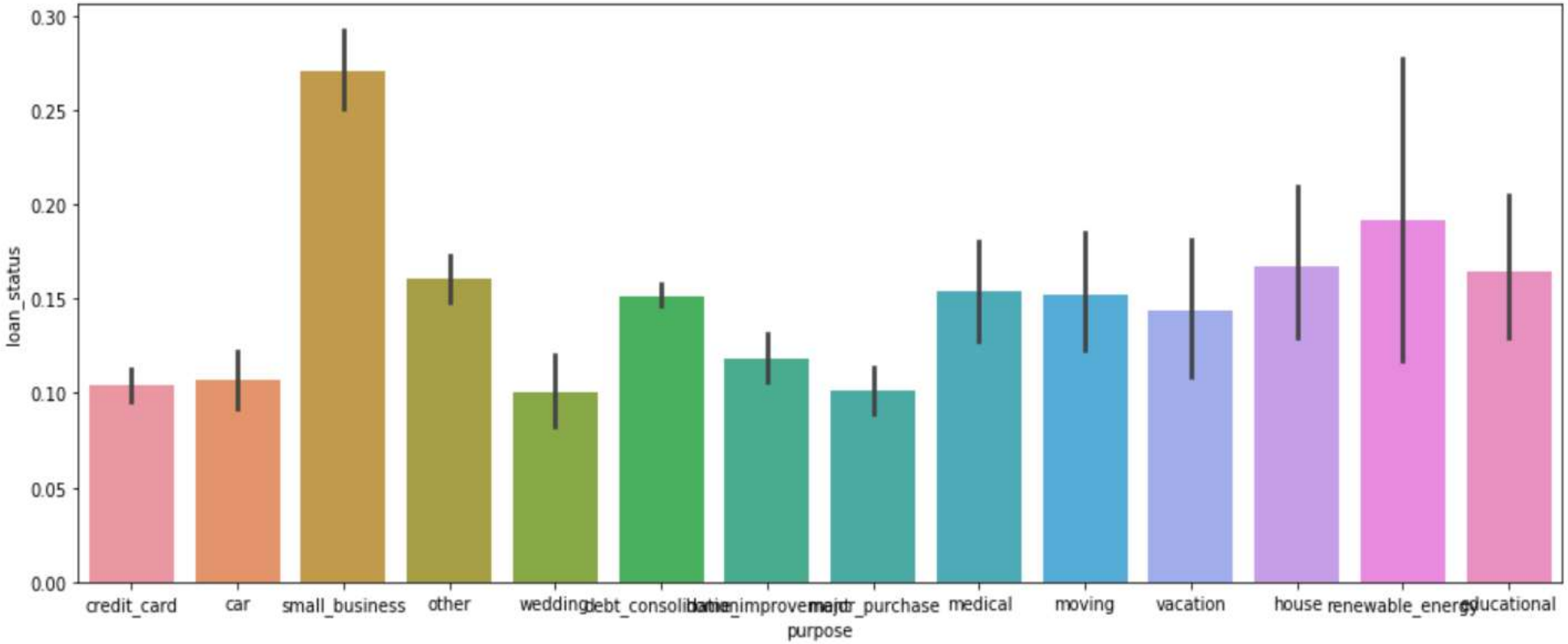
Home ownership across loan status:



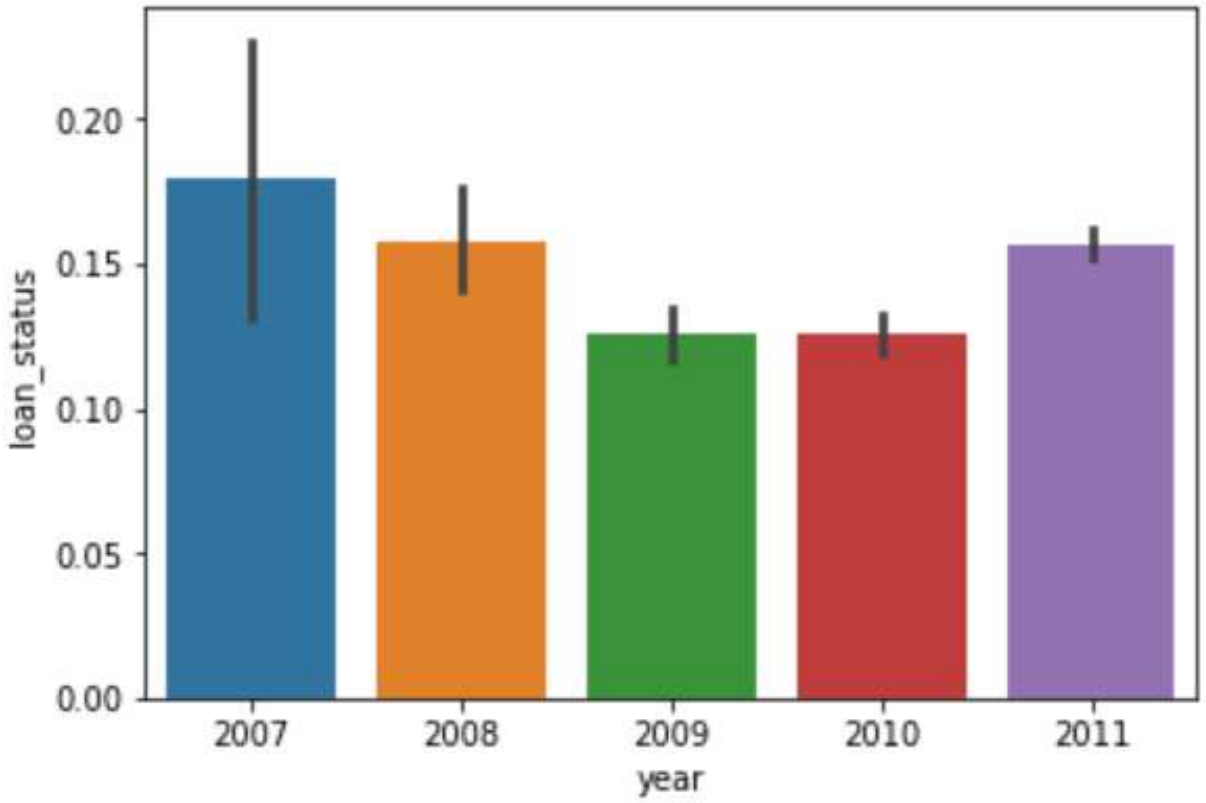
Verification status across loan status:



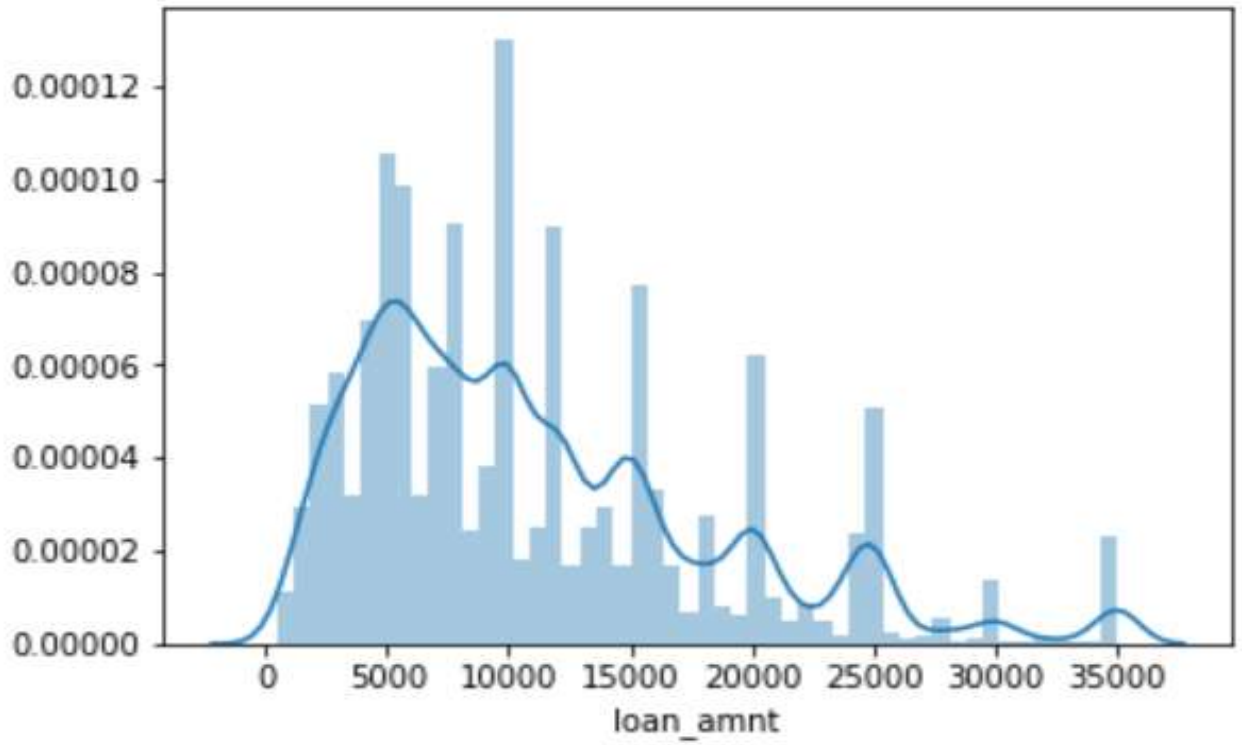
Purpose across loan status:



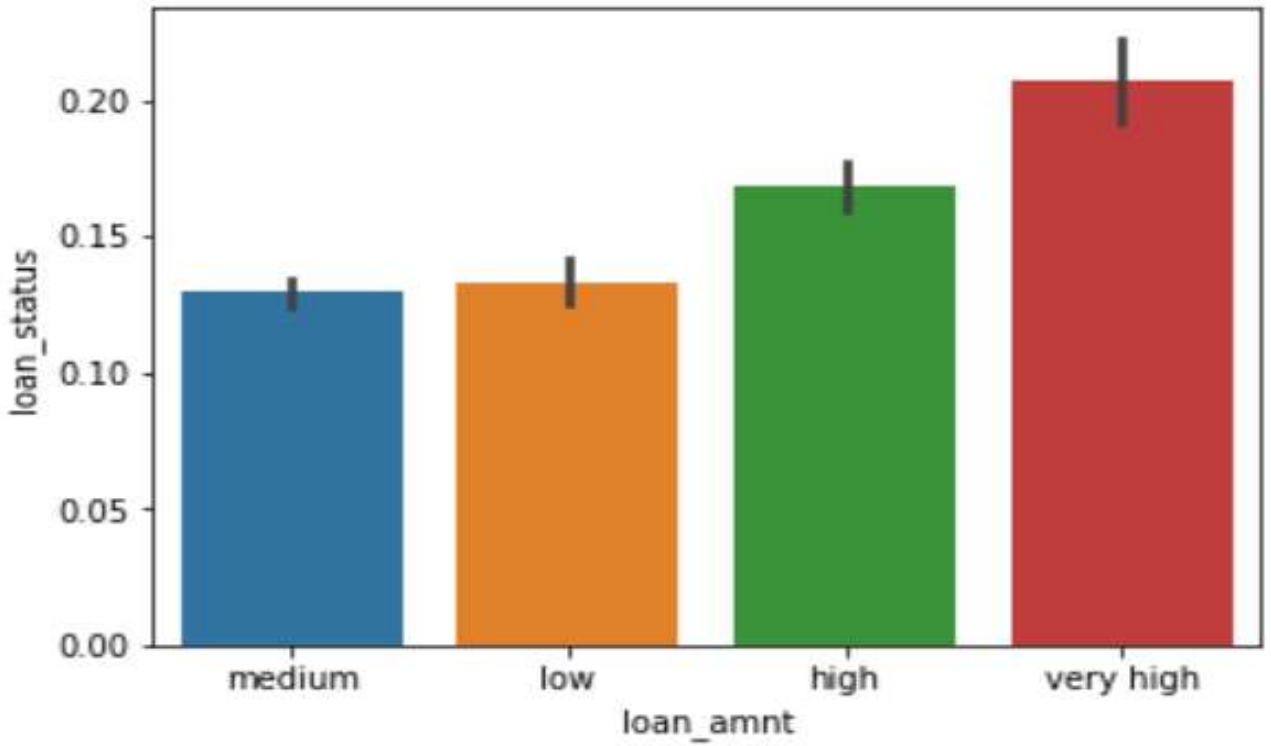
Observe the distribution of loans across years:



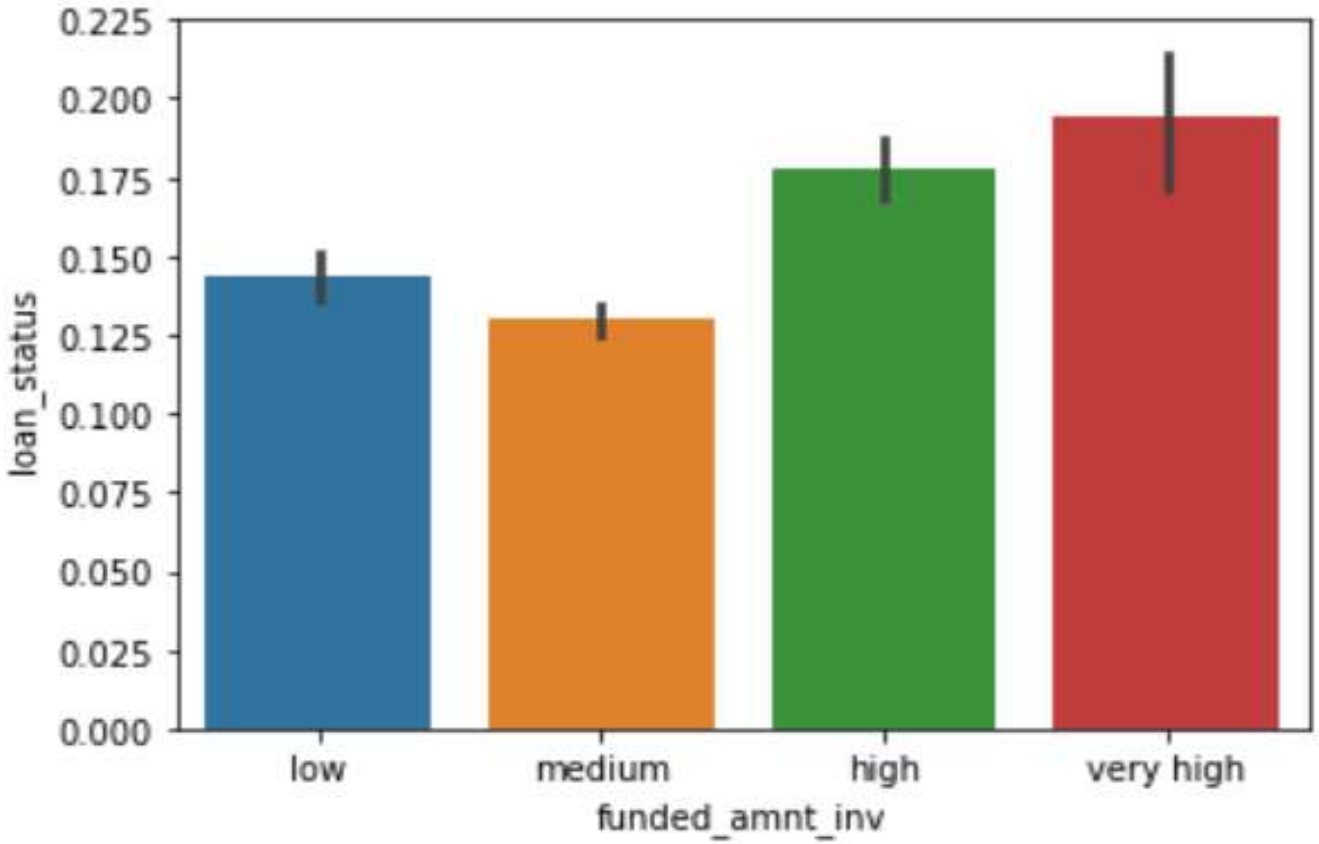
Distribution of loan amount:



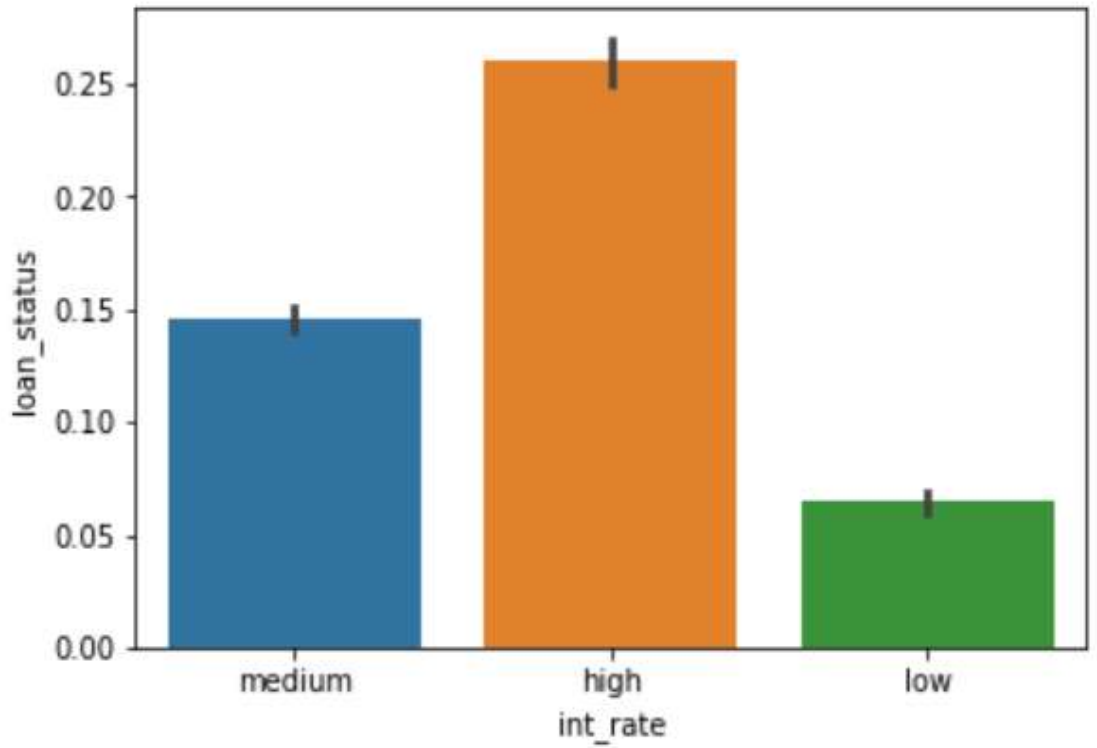
Binning the loan amount:



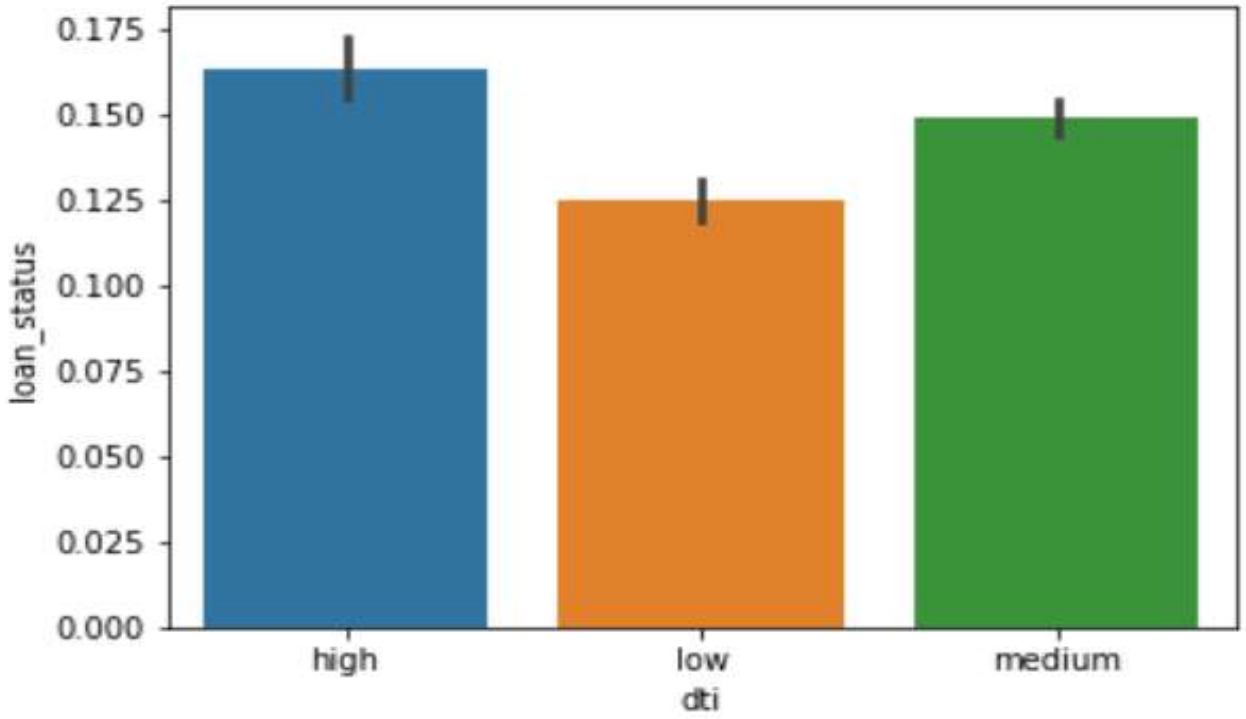
Binning the funded amount:



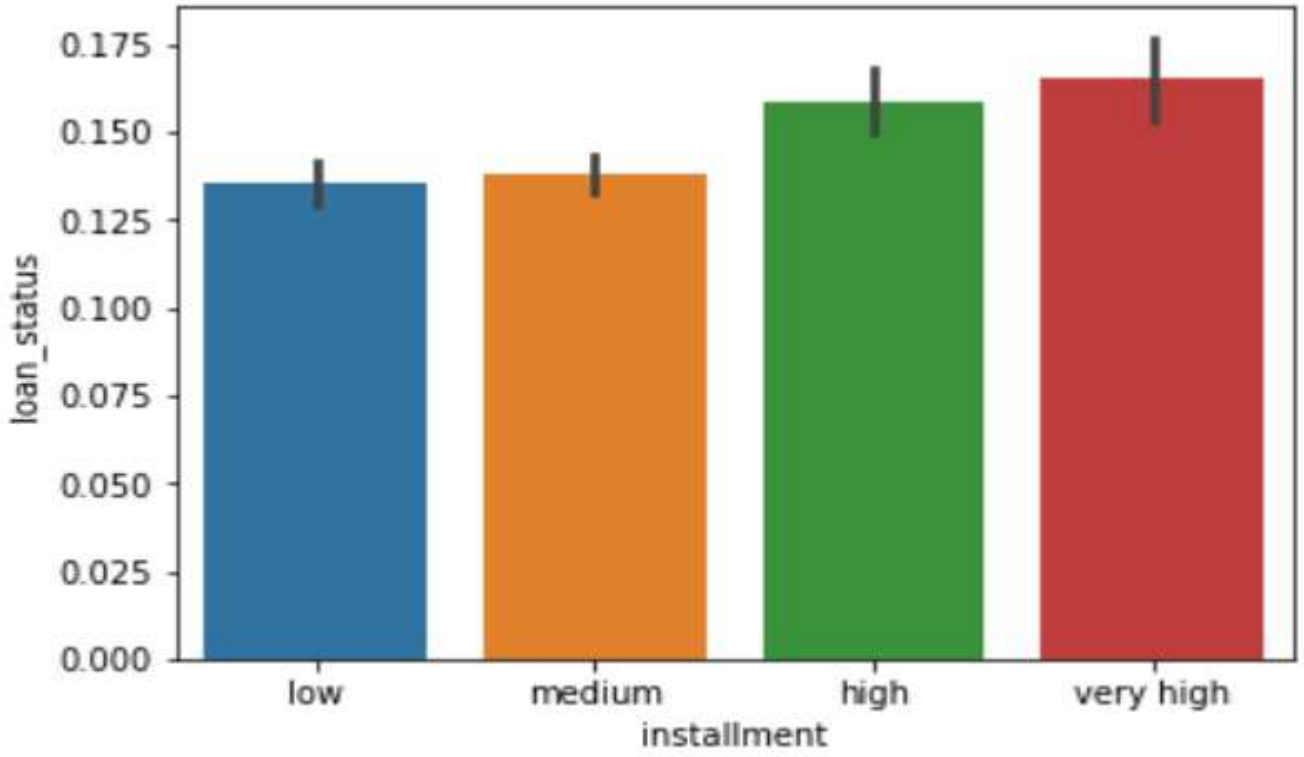
Binning the int rate:



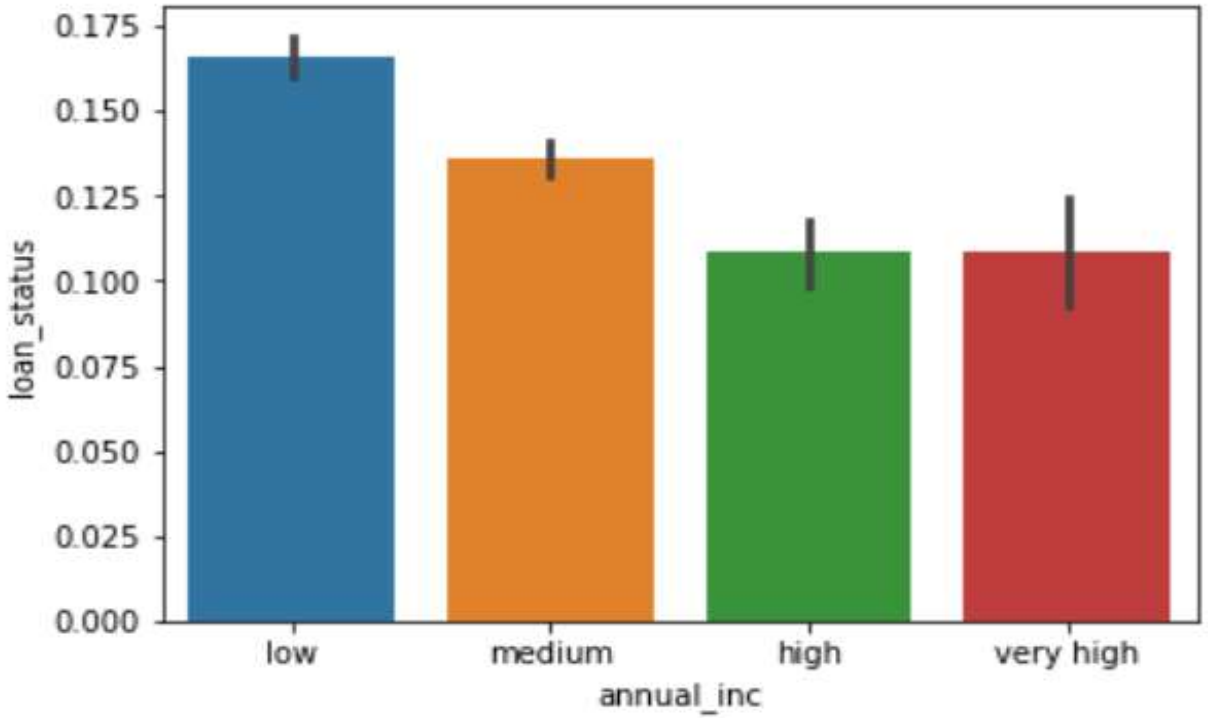
Binning the dti (debt to income ratio):



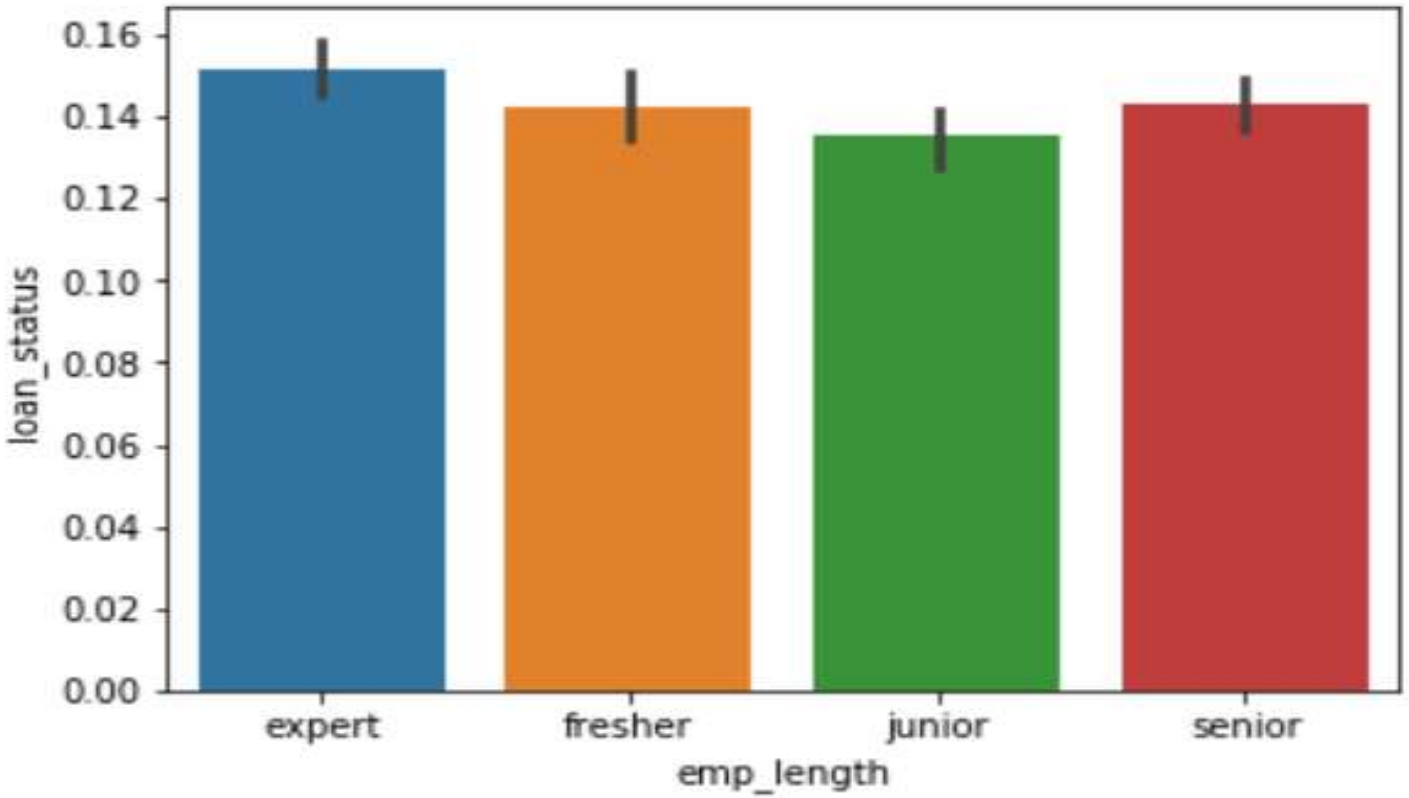
Binning the installment:



Binning the annual income:

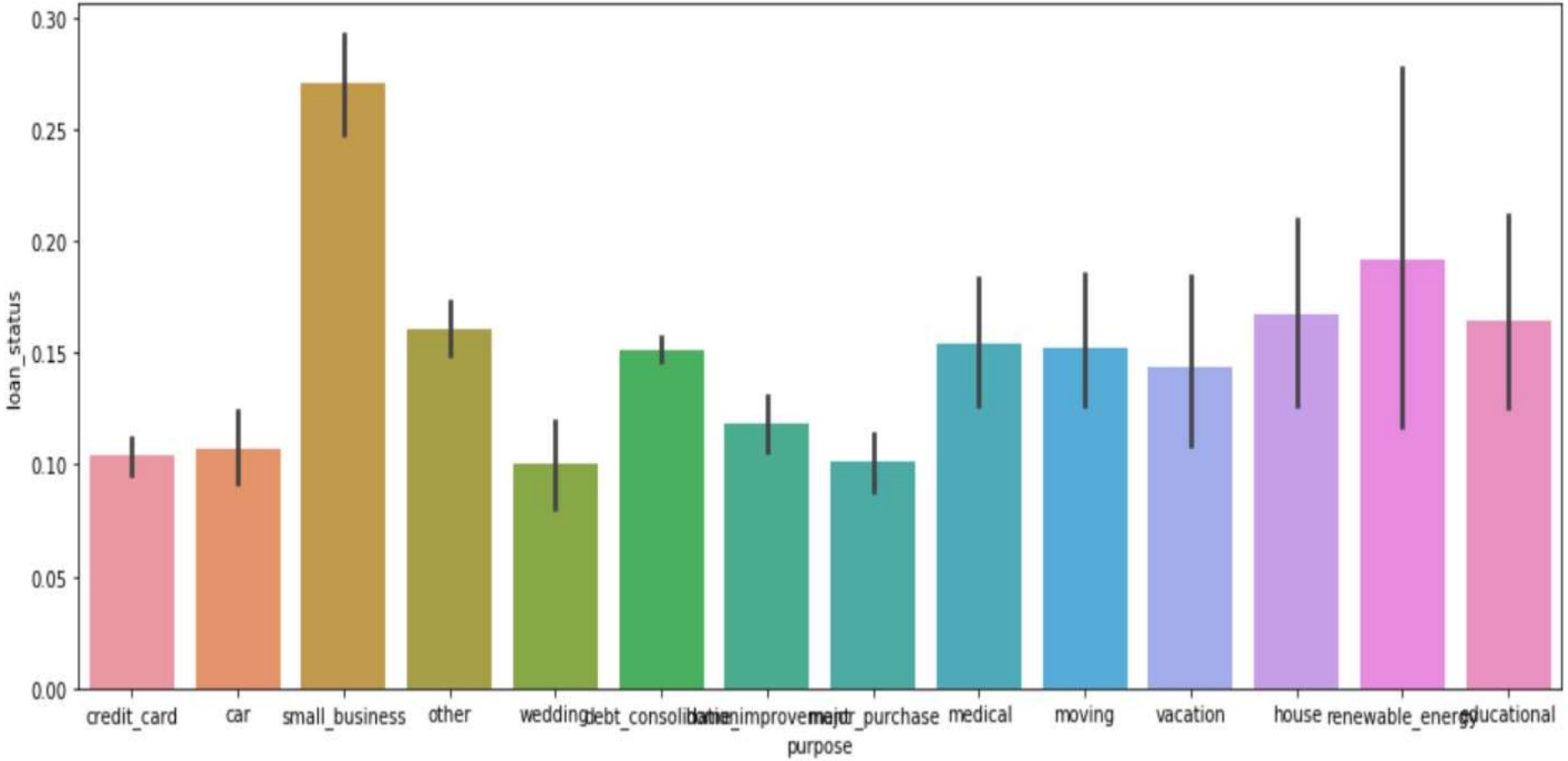


Binning the employment length:

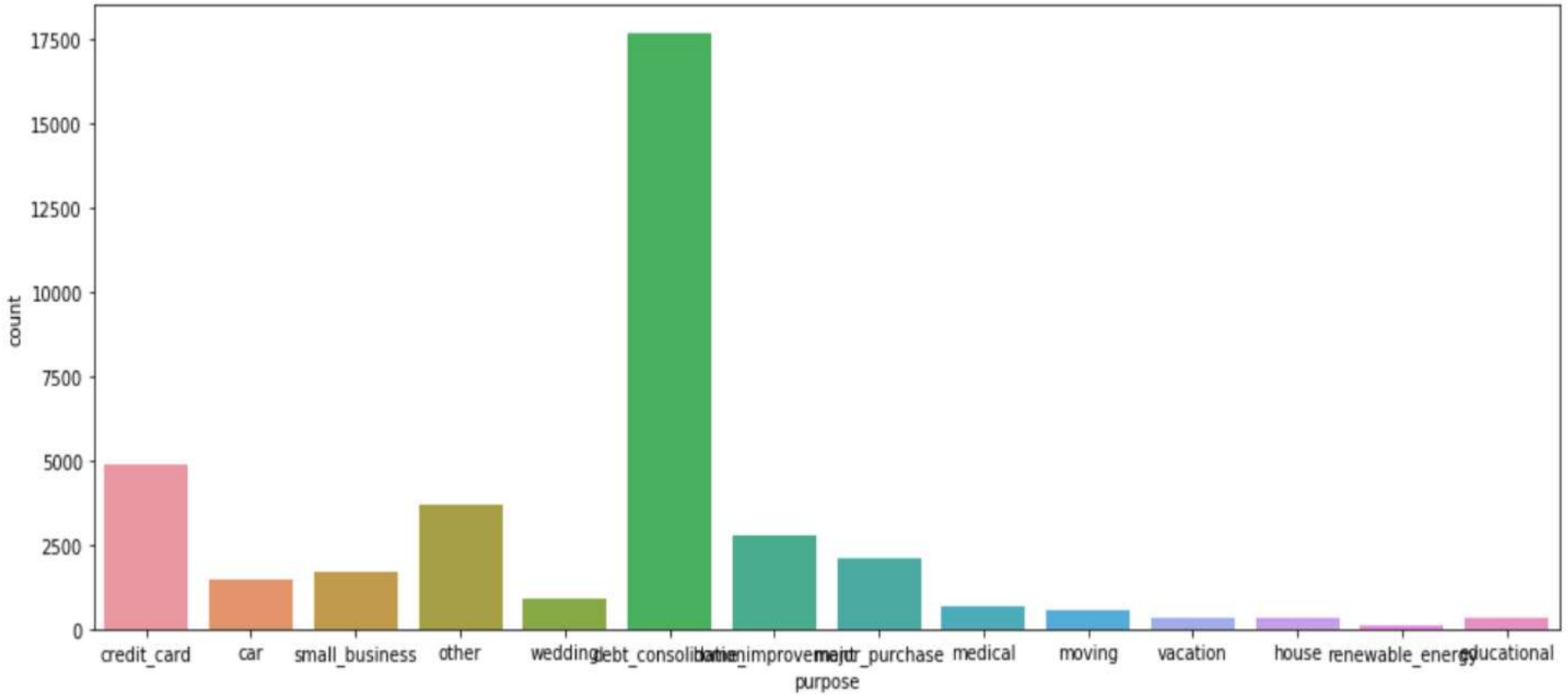


Segmented Univariate Analysis:

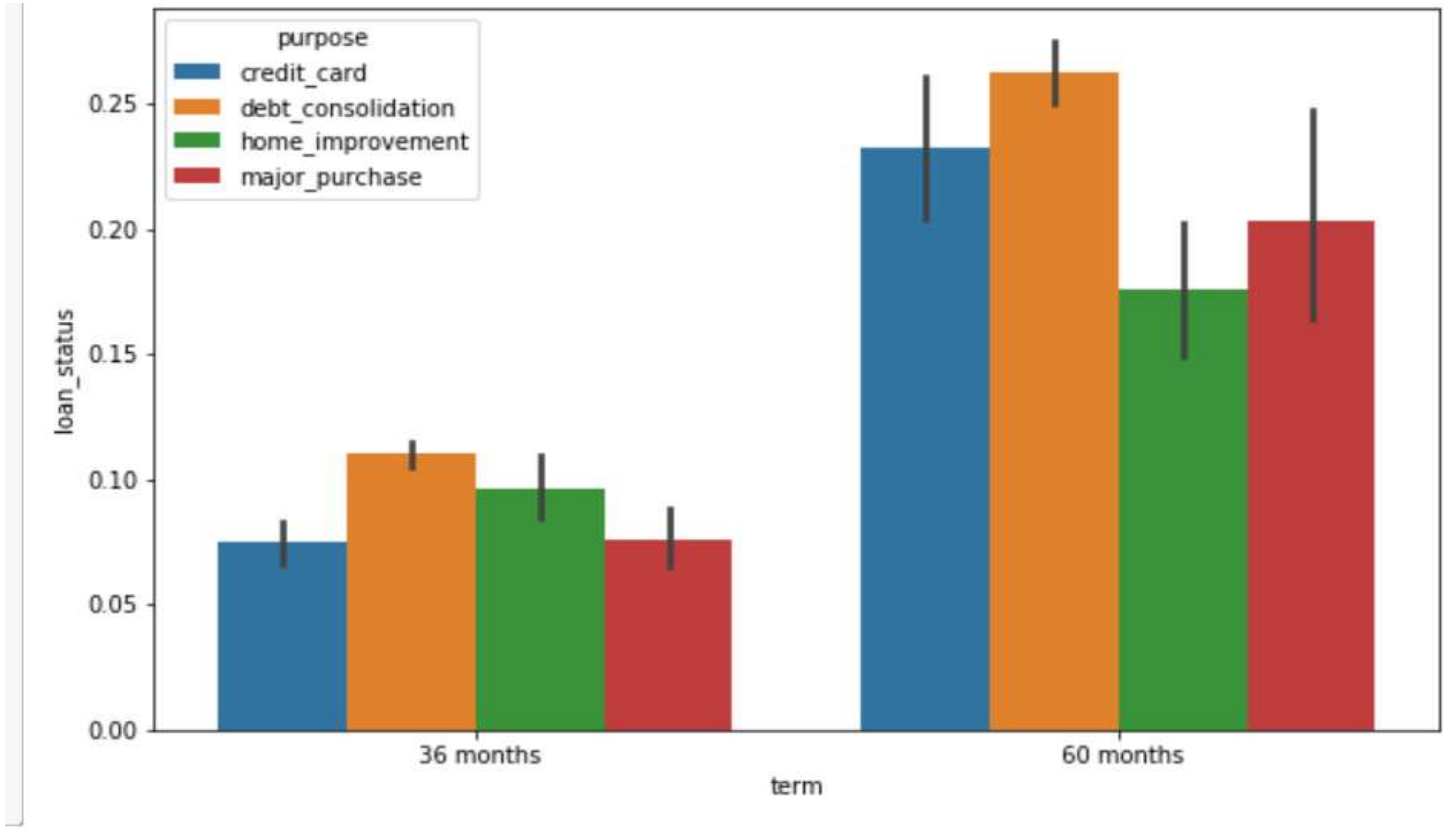
- We have now compared the default rates across various variables, and some of the important predictors are purpose of the loan, interest rate, annual income, grade etc.
- In the credit industry, one of the most important factors affecting default is the purpose of the loan - home loans perform differently than credit cards, credit cards are very different from debt condolidation loans etc.
- This comes from business understanding, though let's again have a look at the default rates across the purpose of the loan.


0.30
0.25
0.20
0.15
0.10
0.05
0.00
loan_status
credit_card
car
small_business
other
wedding
debt_consolidation
home_improvement
major_purchase
medical
moving
vacation
house
renewable_energy
educational
purpose

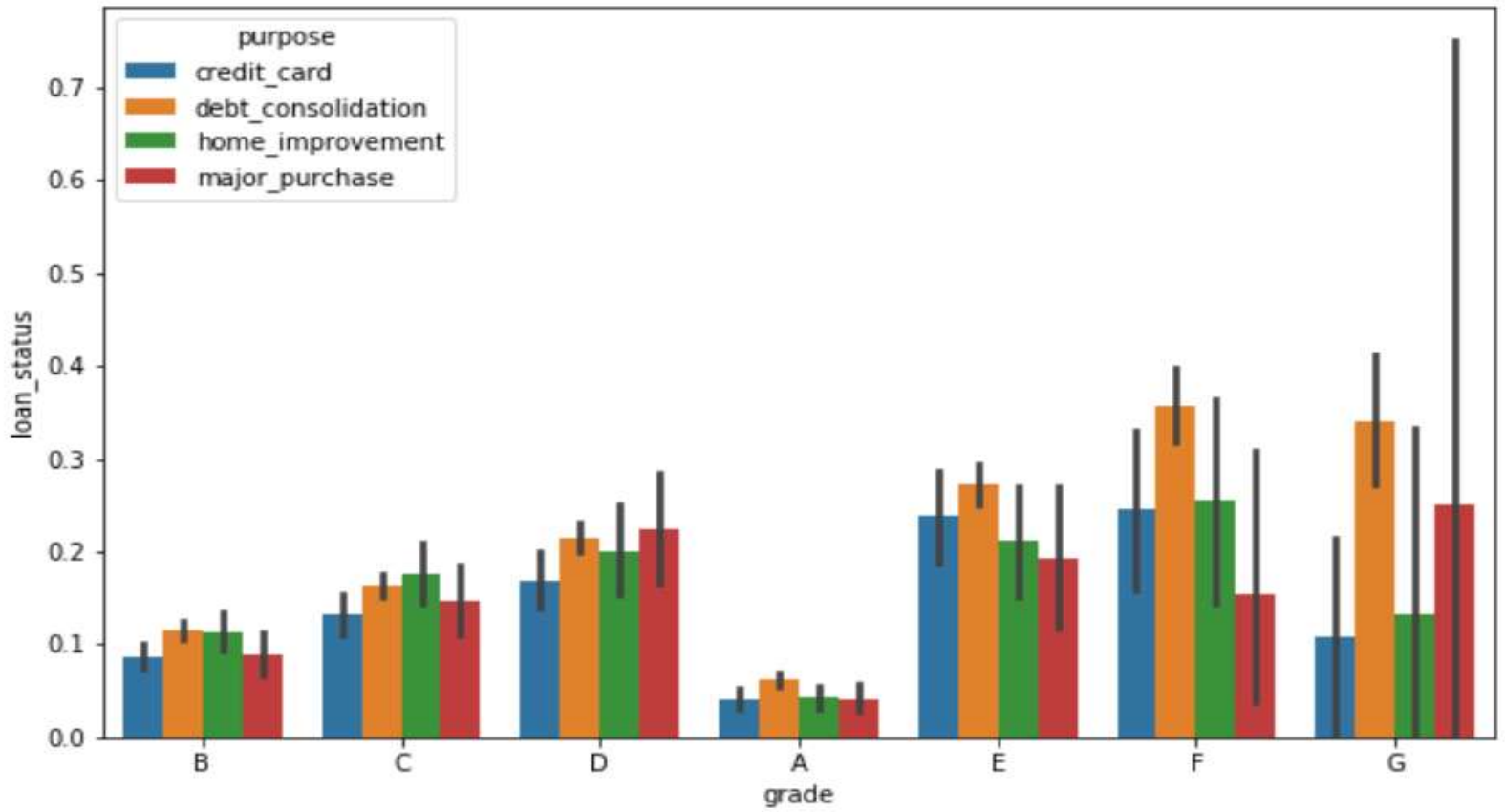
Find the number of loans for each type (purpose) of the loan:



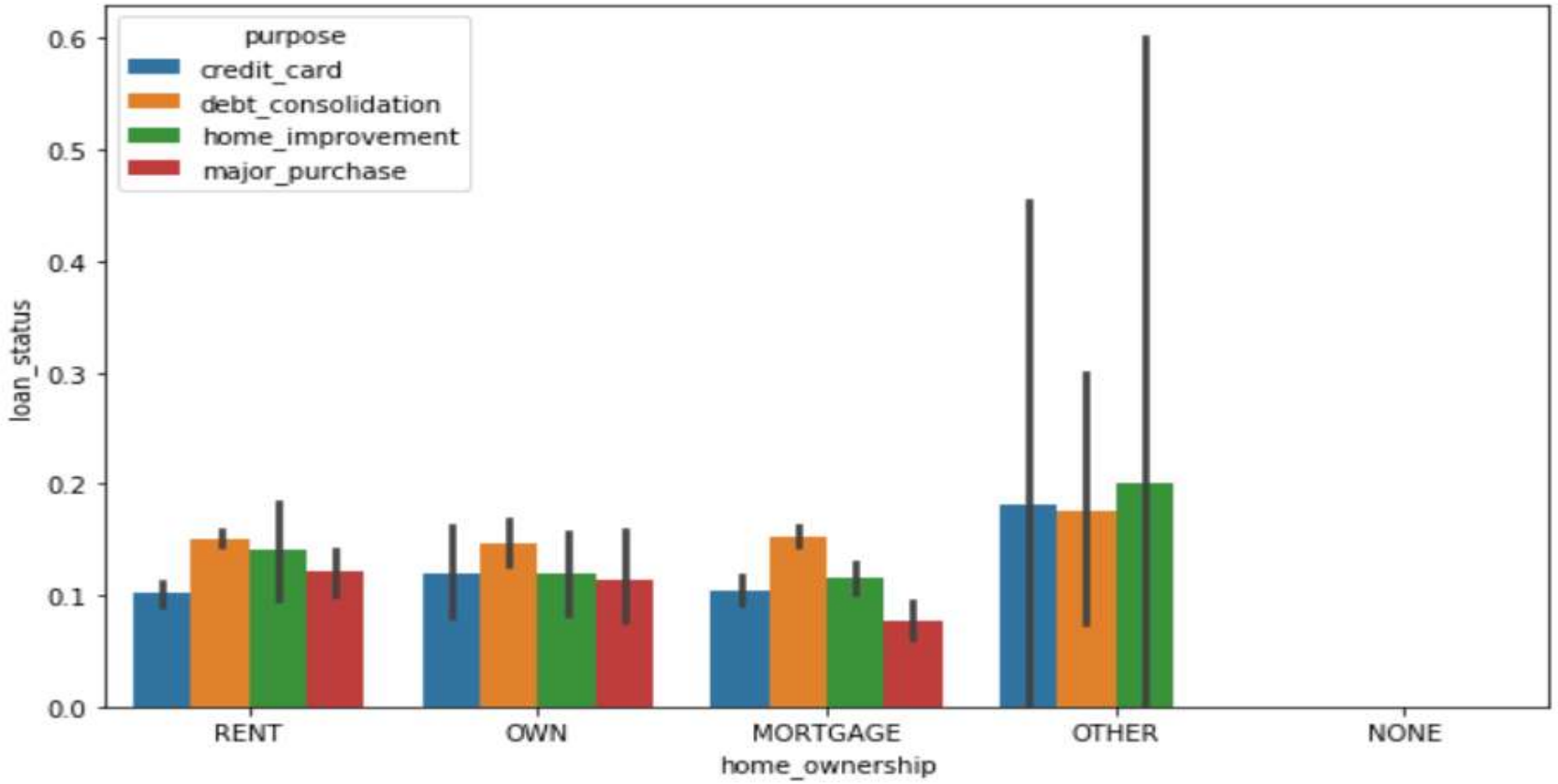
Compare the default rates across two types of categorical variables, purpose of loan (constant) and another categorical variable (which changes):



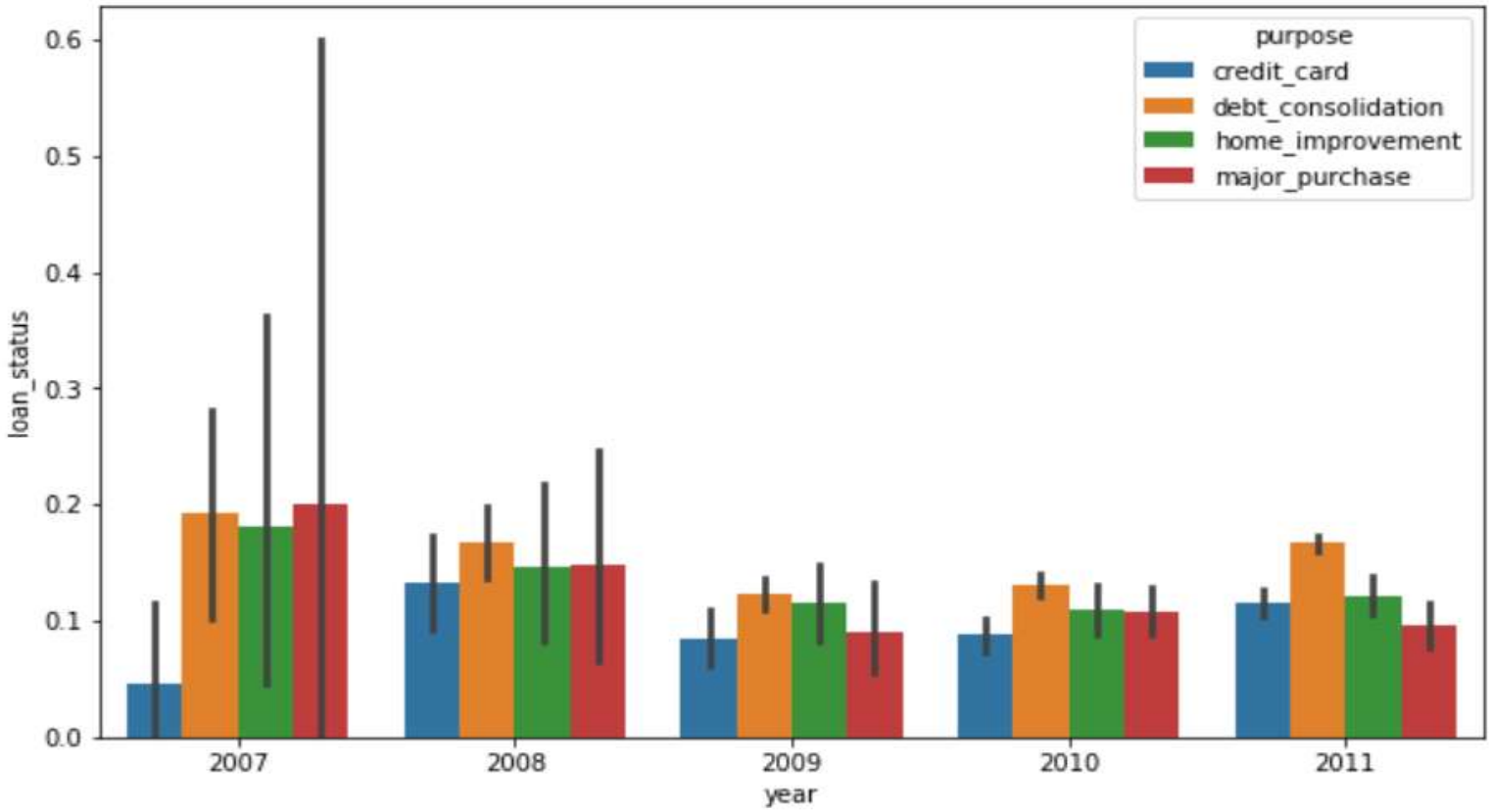
Plot segmented grade:



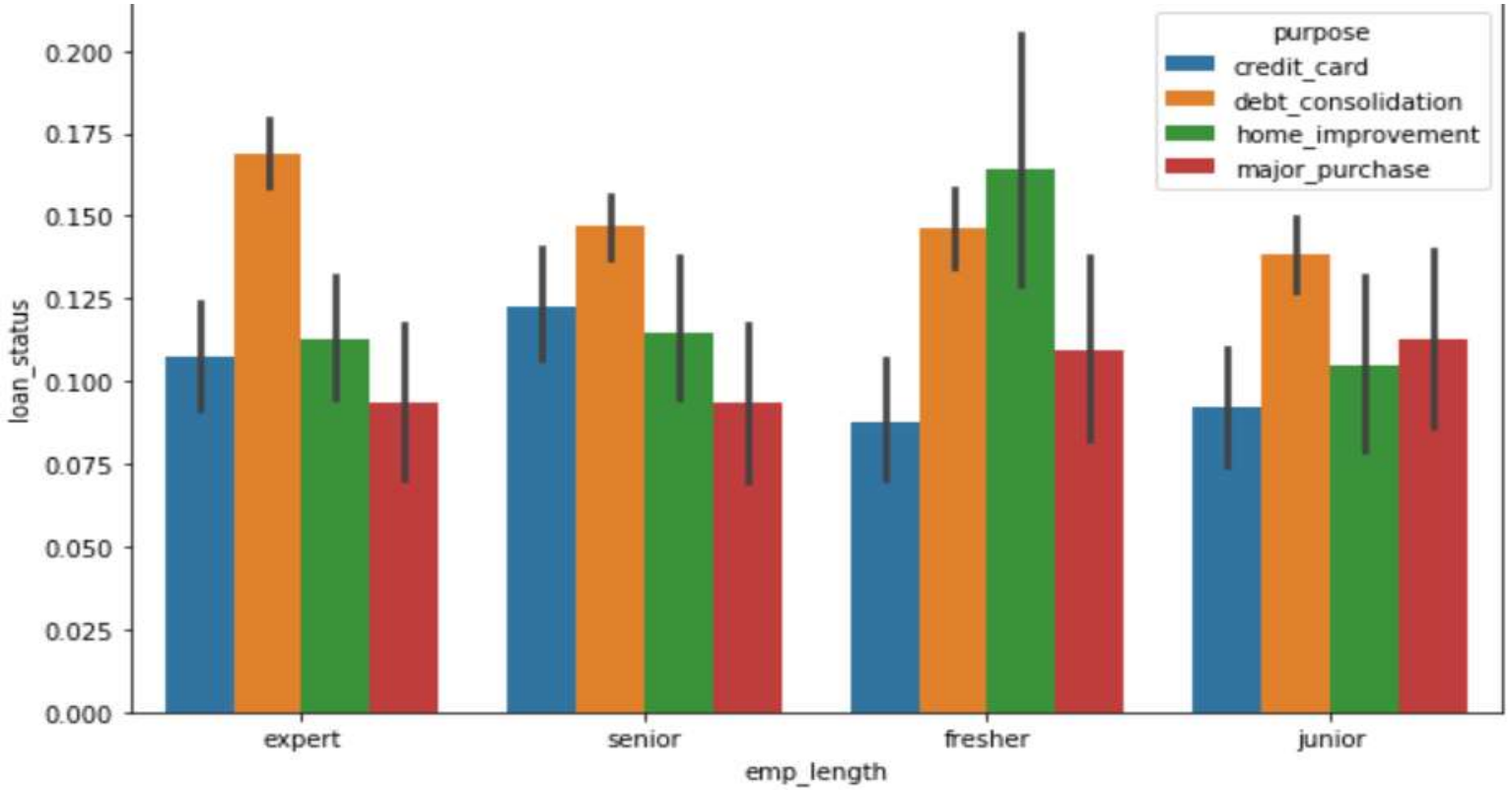
Plot segmented home ownership:



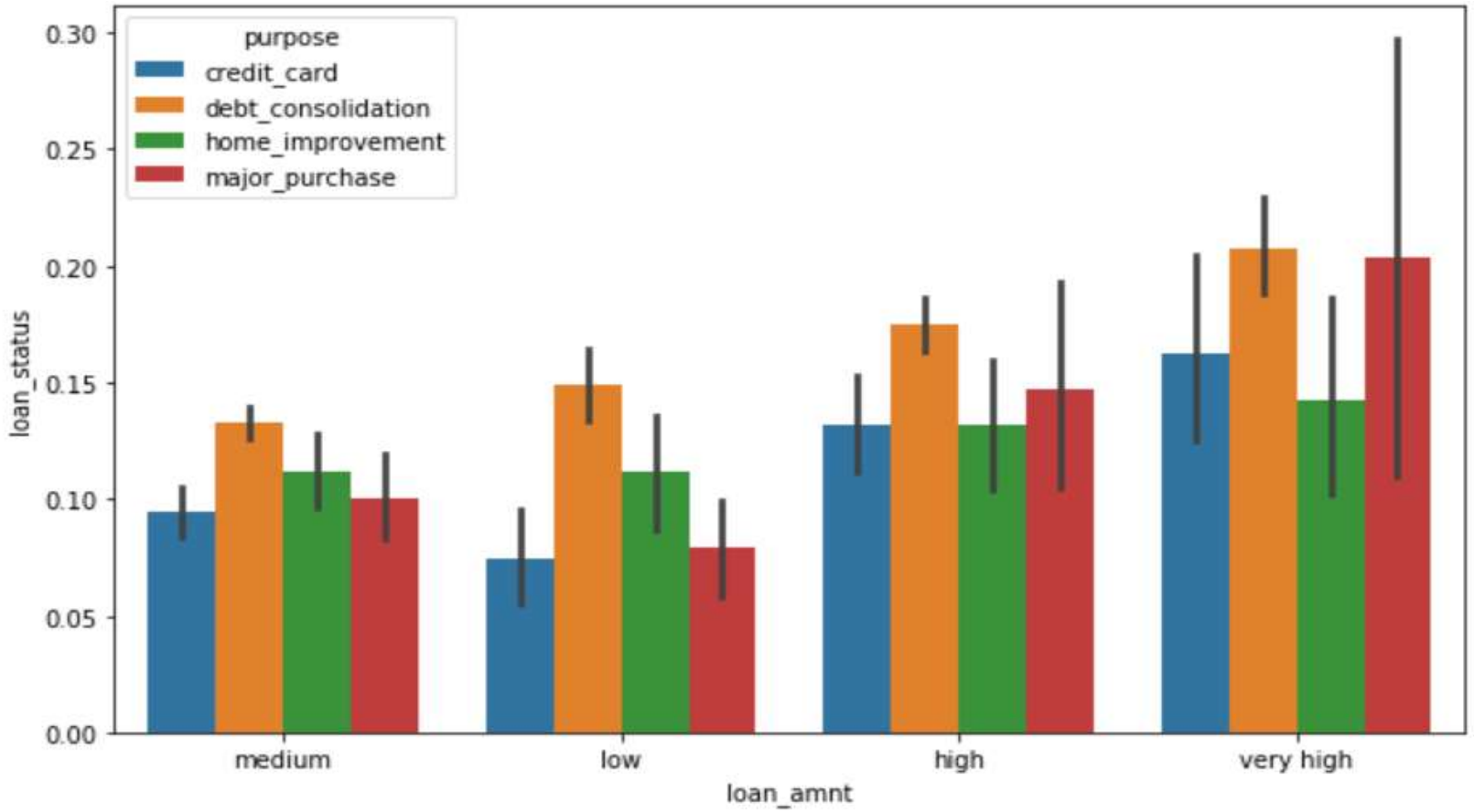
Plot segmented year:



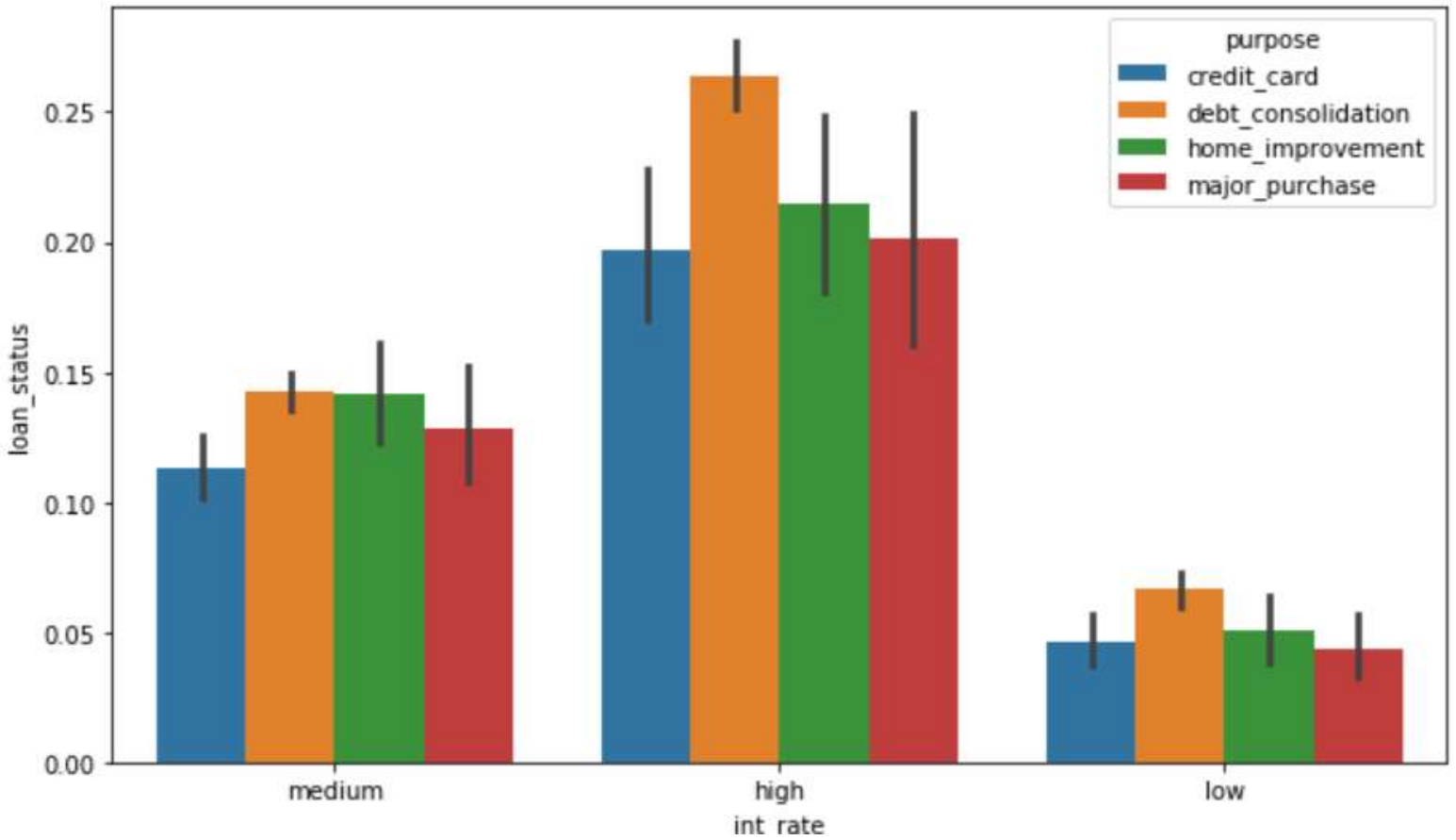
Plot segmented employment length:



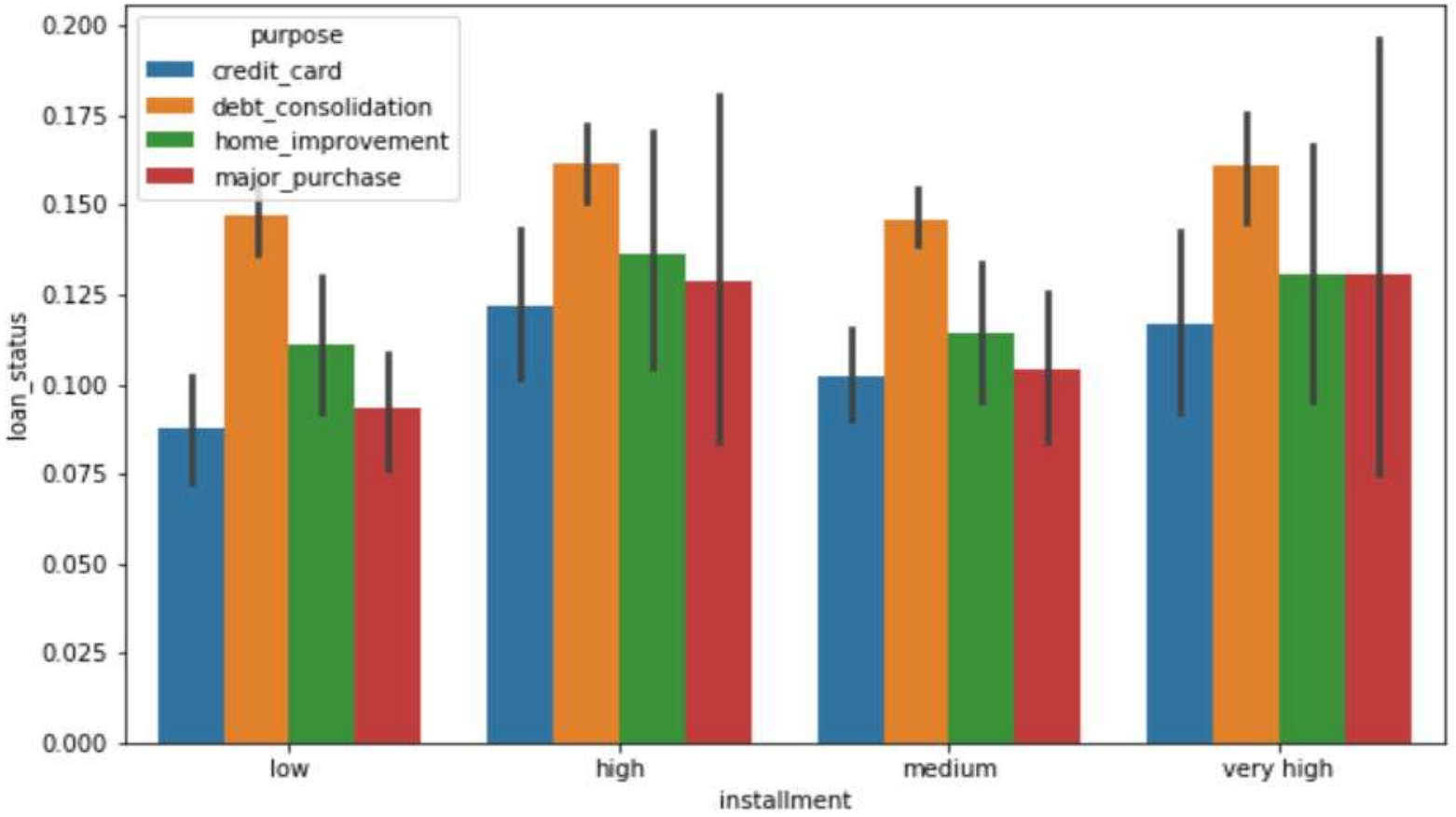
Plot segmented loan amount:



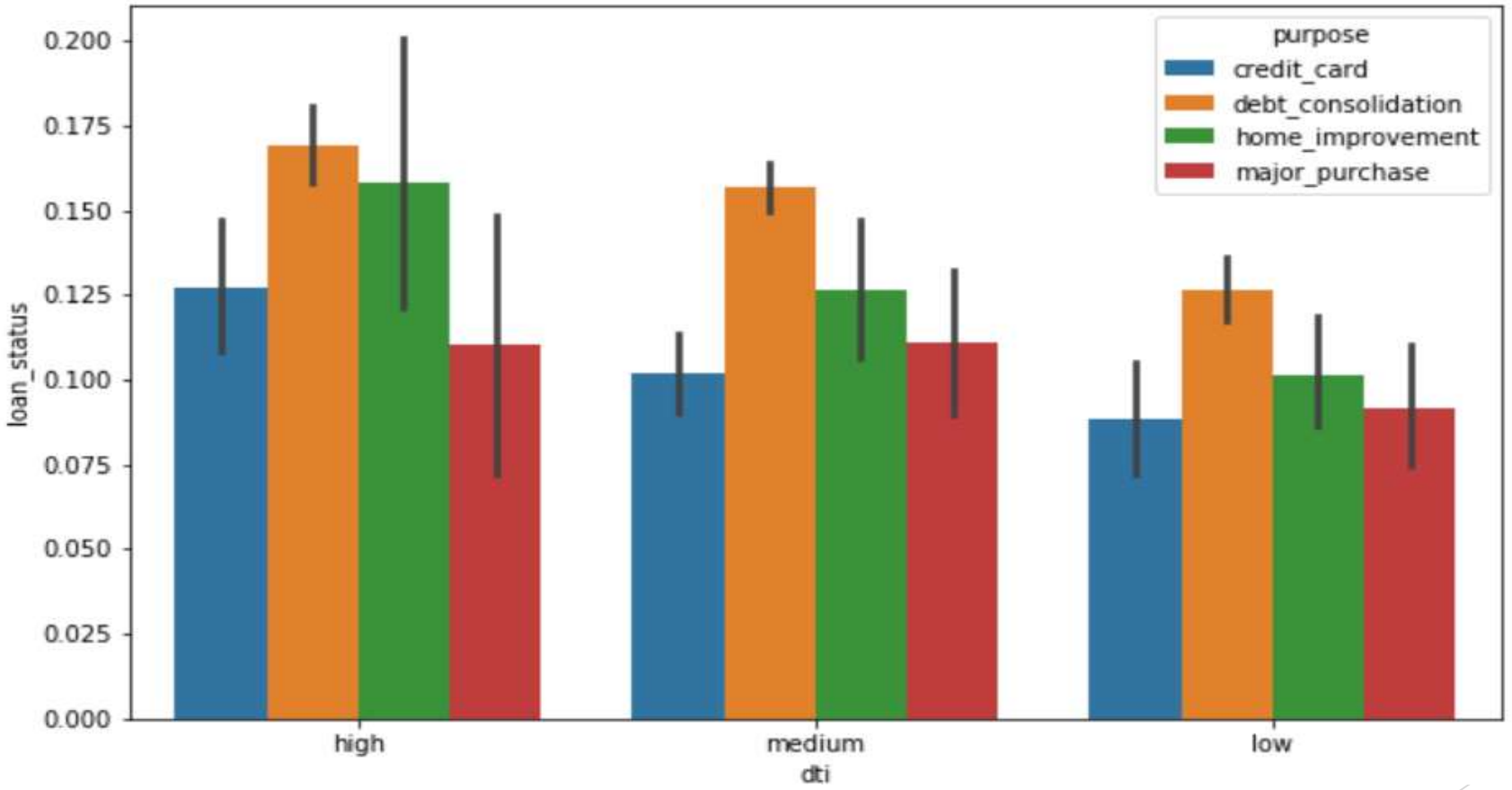
Plot segmented int rate:



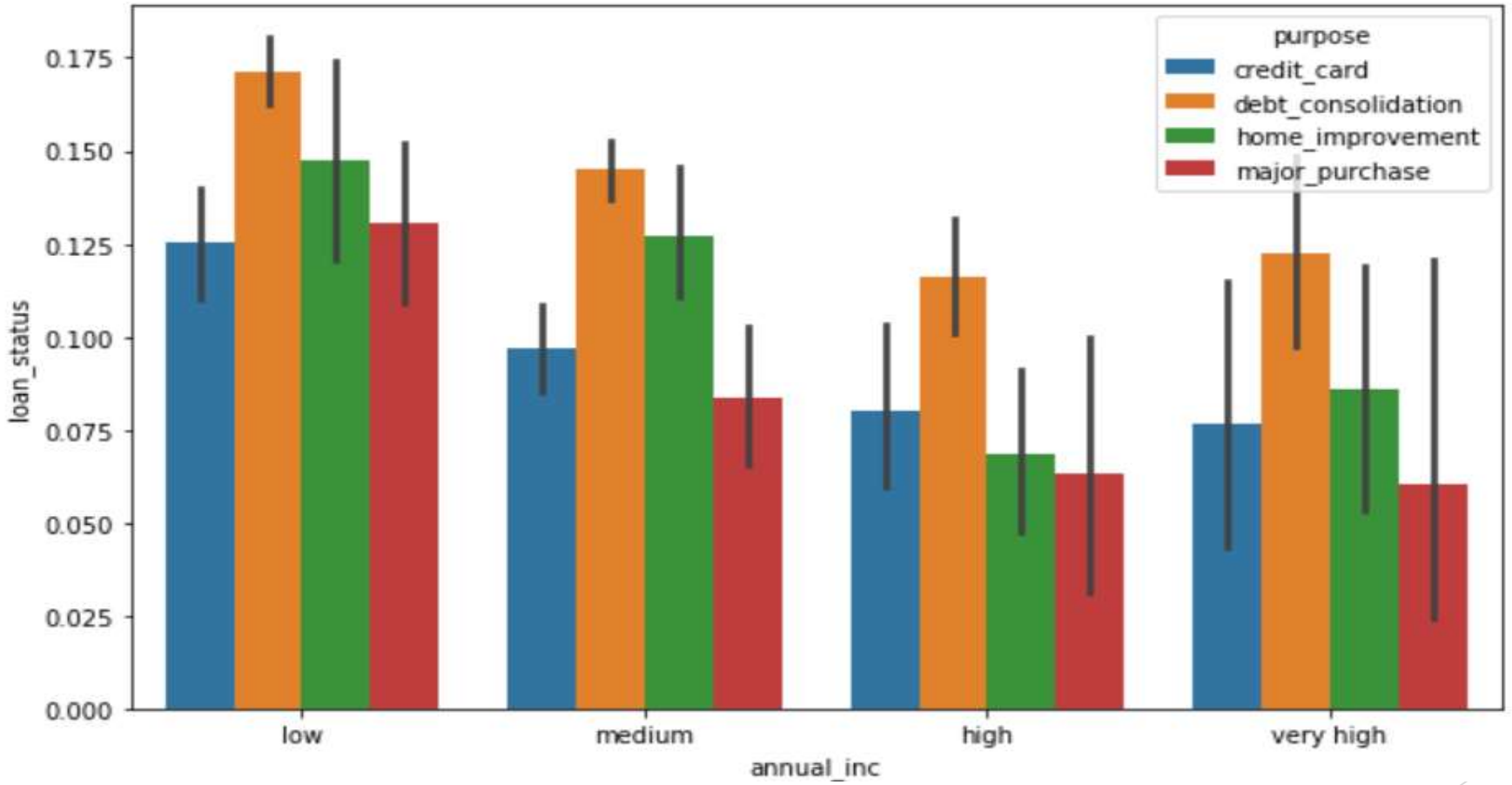
Plot segmented installment:



Plot segmented debt to income ratio(dti):



Plot segmented annual income:



Recommended group where loan can be credited. (Less chance to be defaulter) •

- Clients who are working as a state servant.
- Old people of any income group.
- Client with high income category.
- Old female client.
- Client with higher education (female).
- Any client who's previous loan was approved.
- Widow who has unused previous loan status.
- Refreshed client who has unused loan status previously.

Thankyou!