Lending Club Case Study

There are 2 main points to this analysis

First we have to understand this problem. Here we find that we are to assess the probability of default of credit before lending so that we may sanction the correct projects and reject those applications which have a high probability of default. Both cases(not sanctioning the right applications and sanctioning the wrong applications) may lead to revenue loss.

- 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

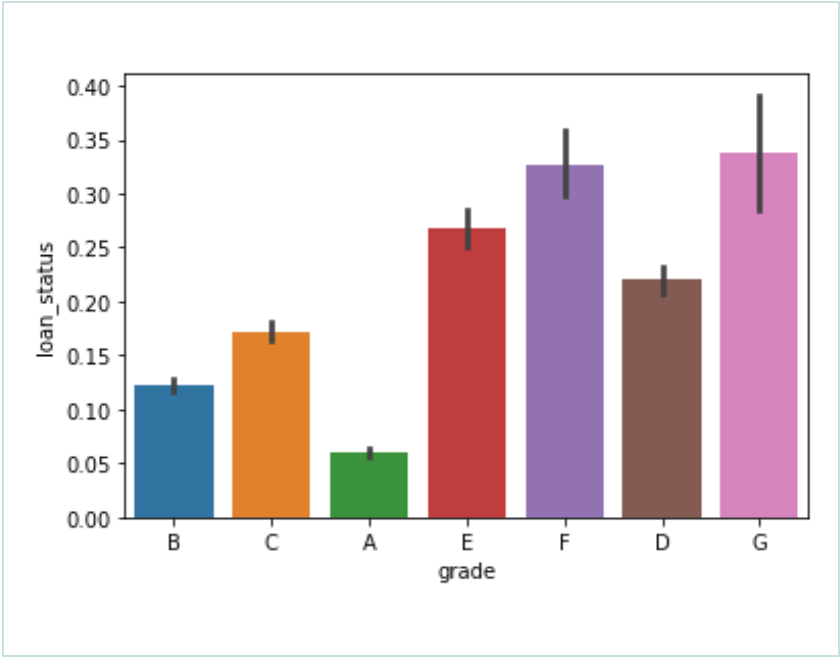
Before undertaking any sort of analysis, first we take a cursory look at this data and find that there were many missing data, redundant data and data which is not relevant to our project.

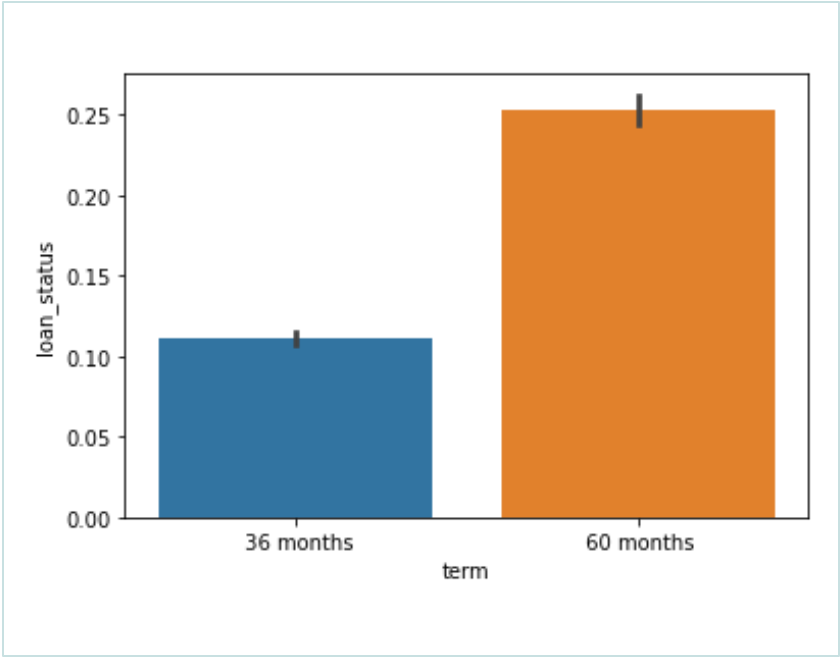
Below is the overview of the data provided to us after some operations

- Rows : 39717
- Columns : 111
- Missing values : 2263364
- Unique values :
- id 39717
- member_id 39717
- loan_amnt 885
- funded_amnt 1041
- funded_amnt_inv 8205
- tax_liens 1
- tot_hi_cred_lim 0
- total_bal_ex_mort 0
- total_bc_limit 0
- total_il_high_credit_limit 0

etc

- Dropping the data which have null values above 60% and data which is not relevant to our project, we end up with 49 columns
- Further we some columns which has all zero values and one column employee title has no relevance to this project, we drop these too.
- Further we replace all the null value with a placeholder ‘0’ value and our final data for analysis is complete
- Our final data has 39717 rows and 43 columns
- Before analysis, we convert some data which is ‘char’ data type to float and int to complete the analysis
- The first analysis to be done is finding out the overall data rate which we find out through mean and find out it is 0.15 overall



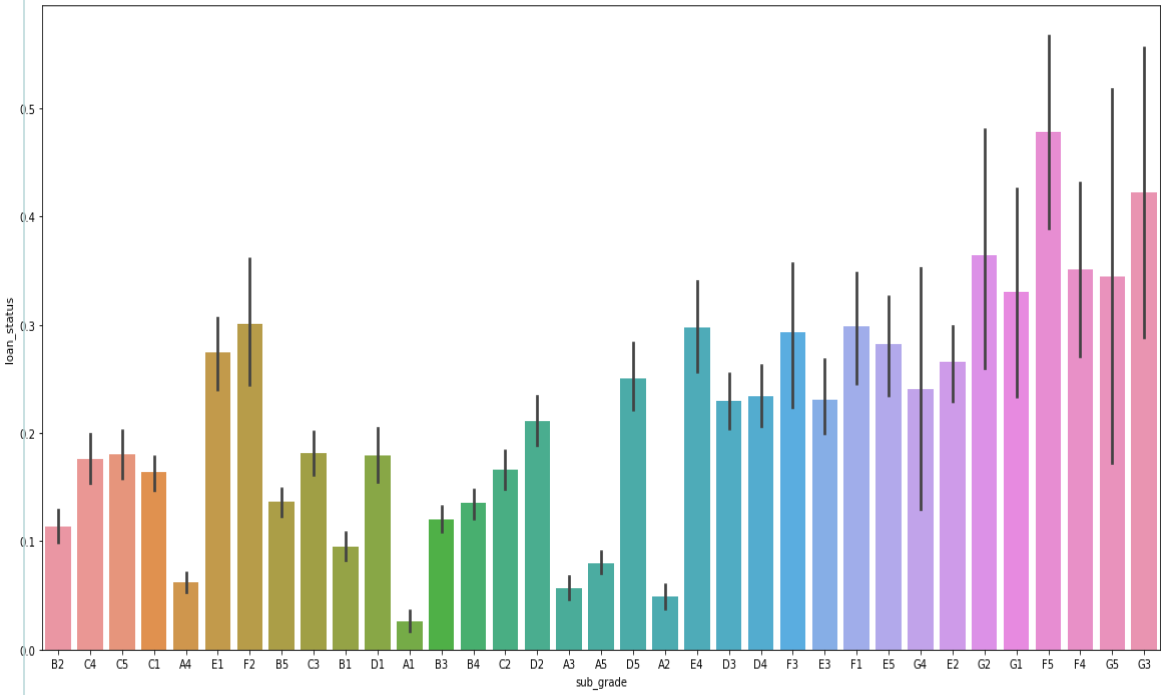


Univariate analysis of grade assigned to a loan against the default rate

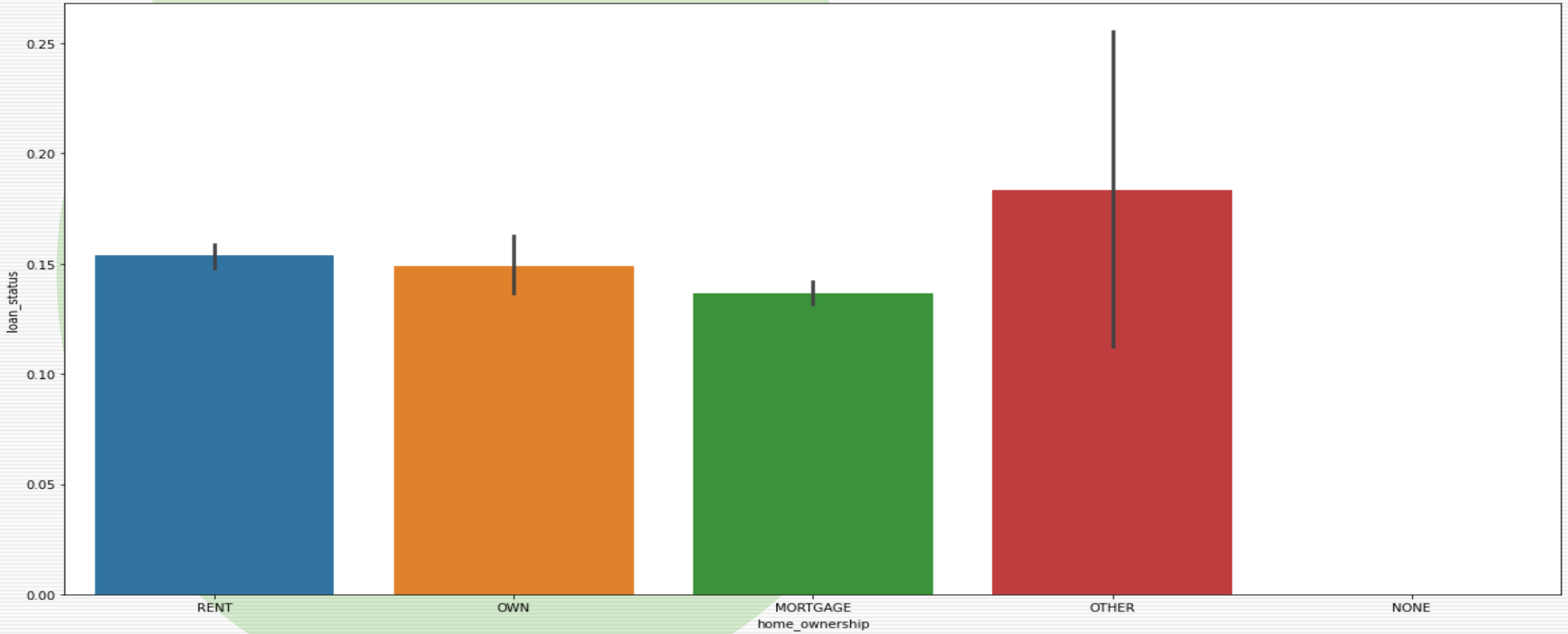
- We see that the higher the grade(A being the highest ,F being the lowest) the lower the rate of default (with the exception of G&F score).
- We also see that the higher the term of loan the higher the rate of default

Univariate analysis of sub-grade assigned to a loan against the default rate

- We see that as expected low graded loans have higher rate of default(once again with the exception of F&G)



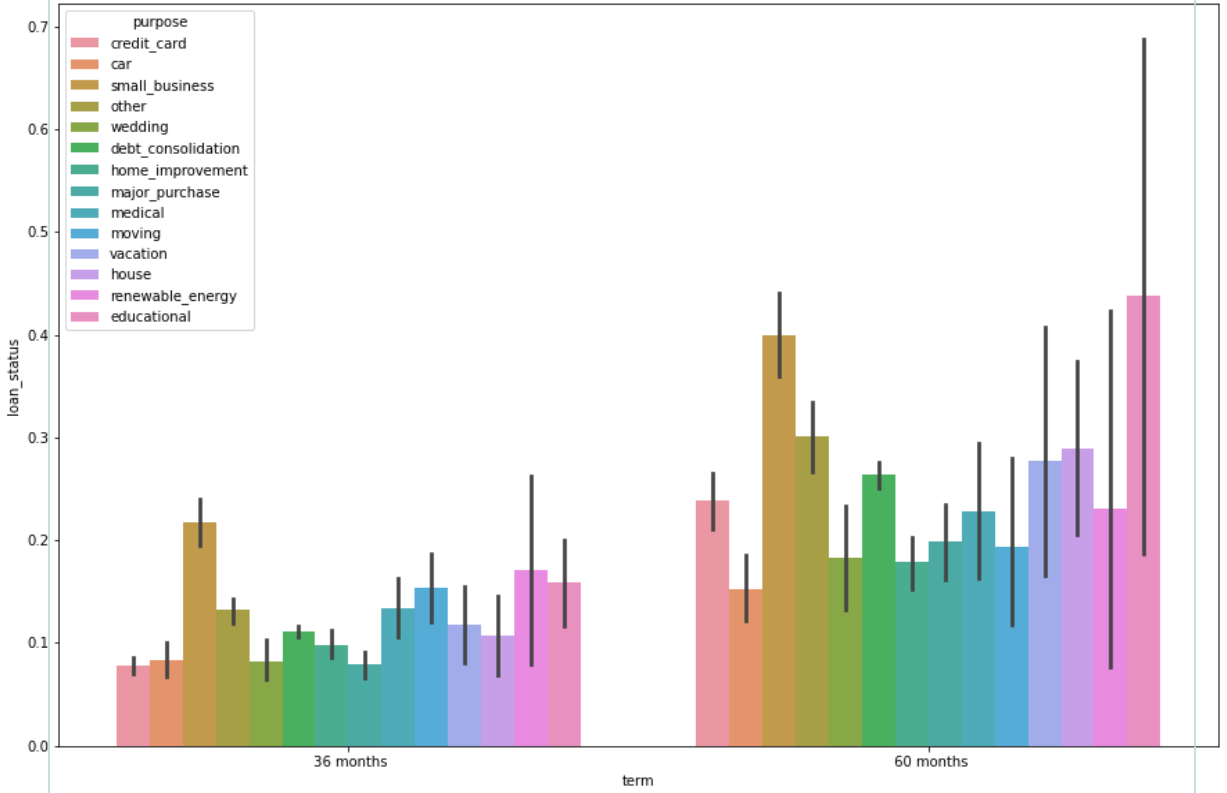
Univariate analysis of home ownership against the default rate



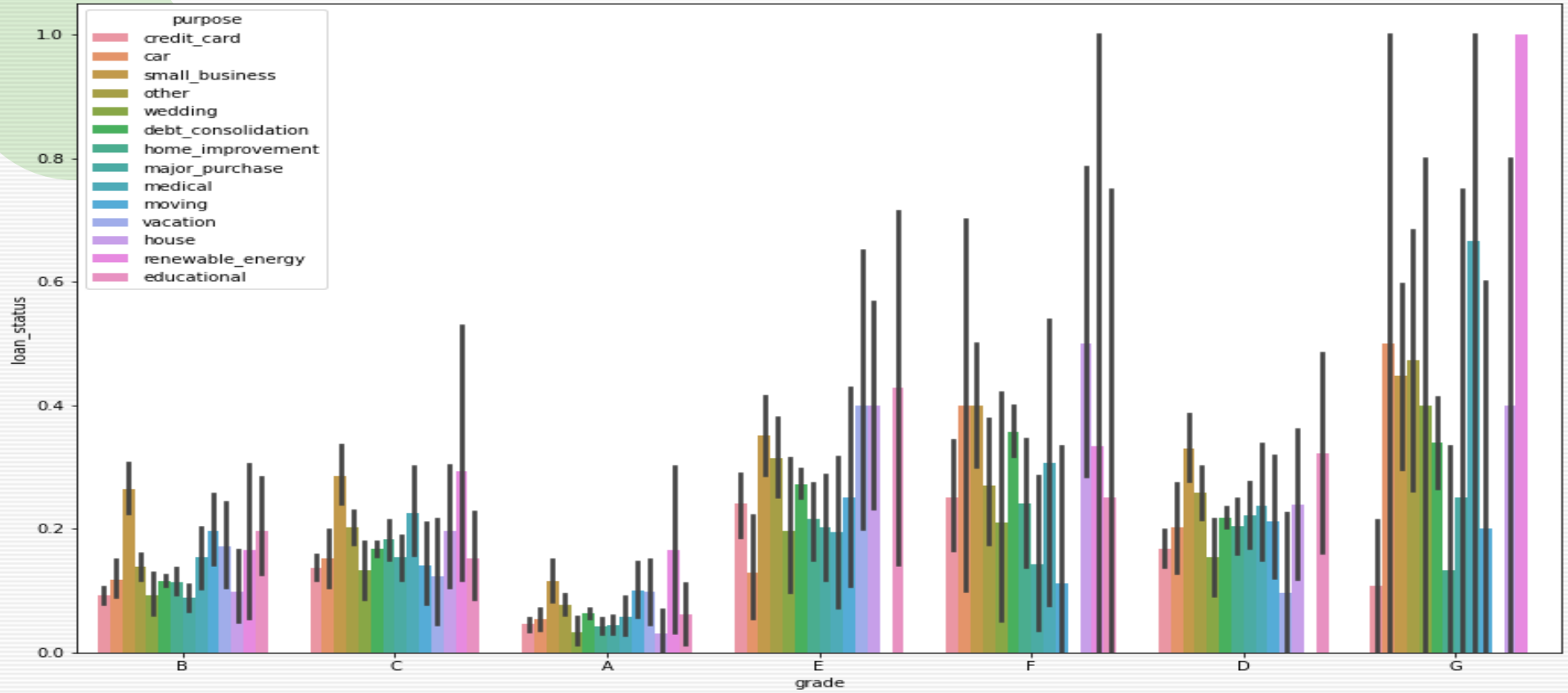
- We see that home ownership is not a great qualifier in sanctioning or rejecting a proposal

Bivariate analysis of purpose of loan, term of loan with the loan status

- We see that for low term loan, credit card has the lowest rate of default while for long term loan, car loan has the lowest rate of default.
- Conversely, small business has the highest rate of default for small term loan and education loan for the long term loan.

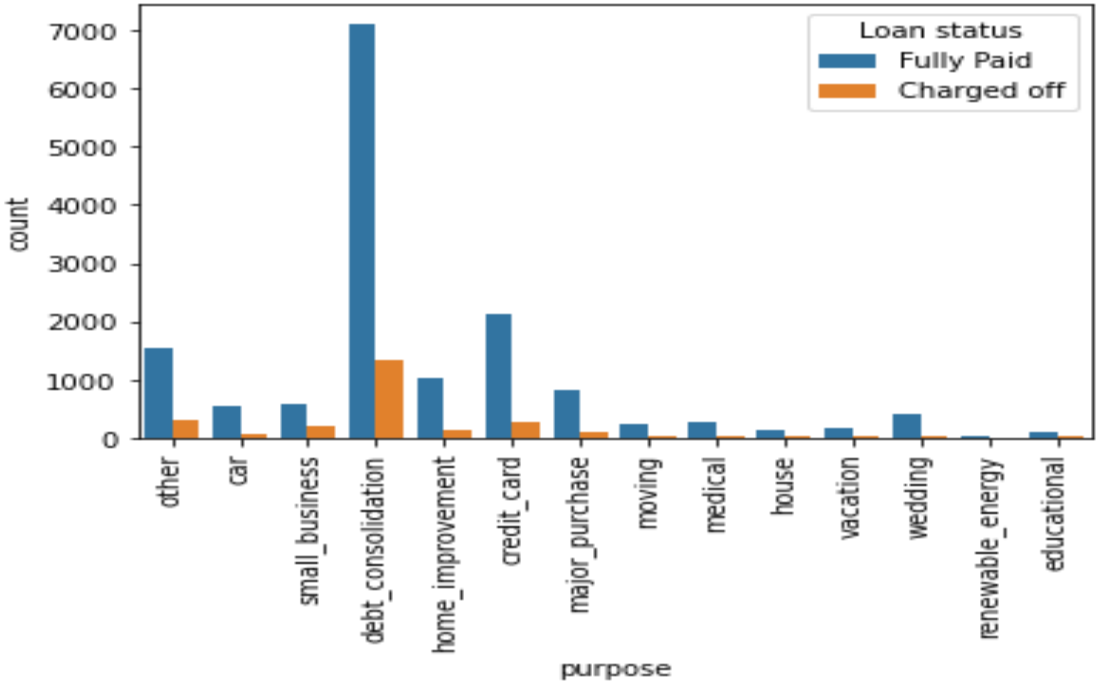


Bivariate analysis comparing grade and purpose of loan to default rate



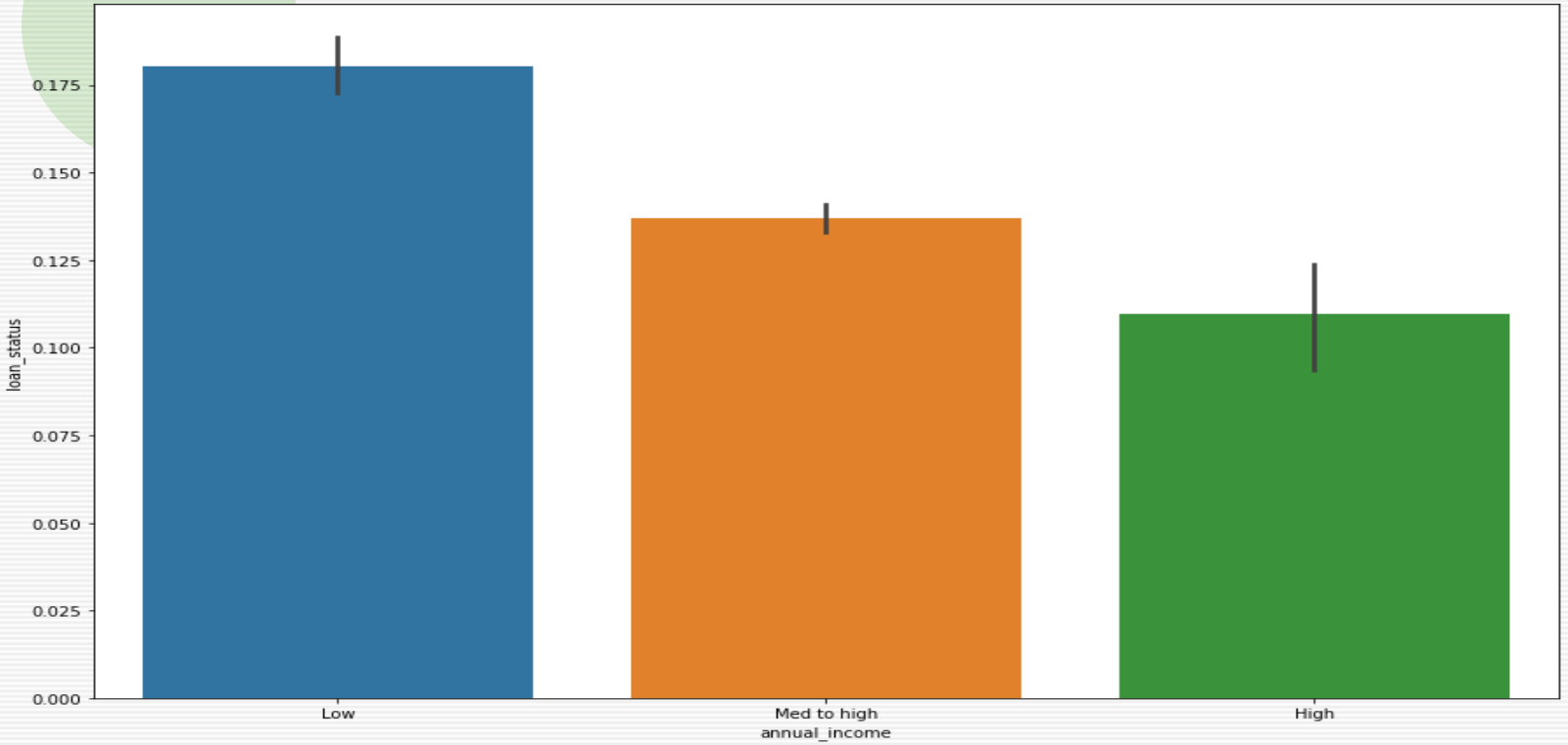
- We see that F&G graded loan specially education loan and renewable energy has the highest chance of default while A rated credit card has the lowest chance of default

Next, we are interested in counting how many loans defaulted with respect to purpose of the loan



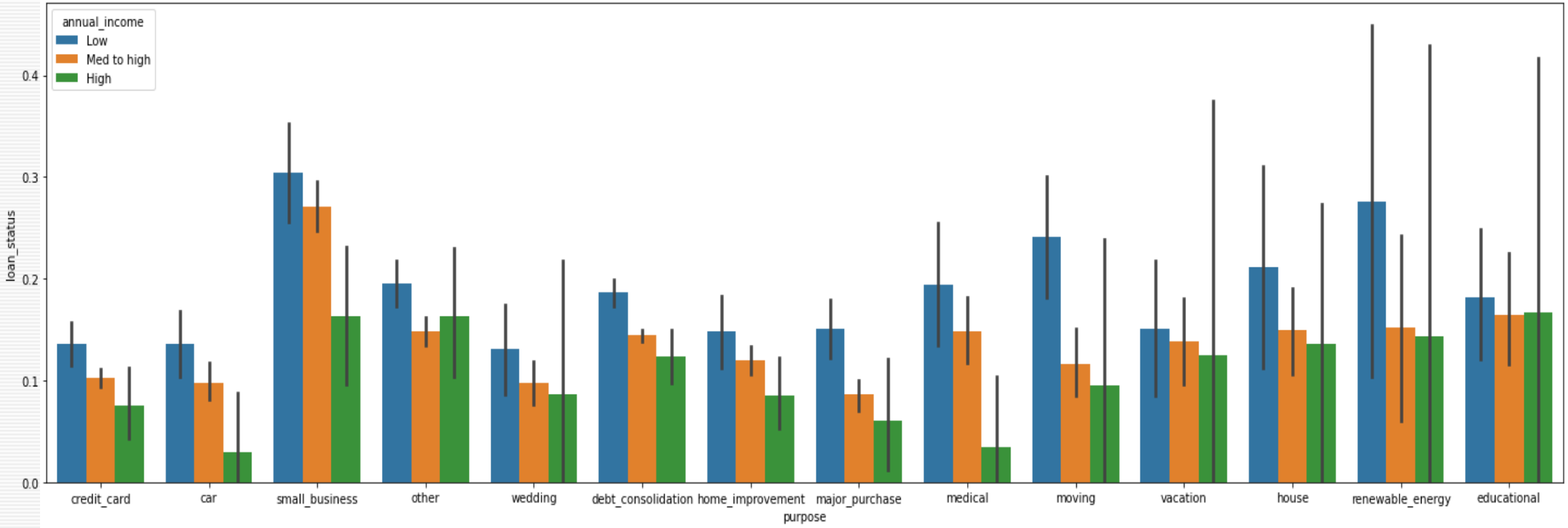
- We see that around 4200 credit card applications were fully repaid and near about 200 defaulted while most of the debt consolidation loan defaulted with only a few fully repaid

Default rate vary around income status

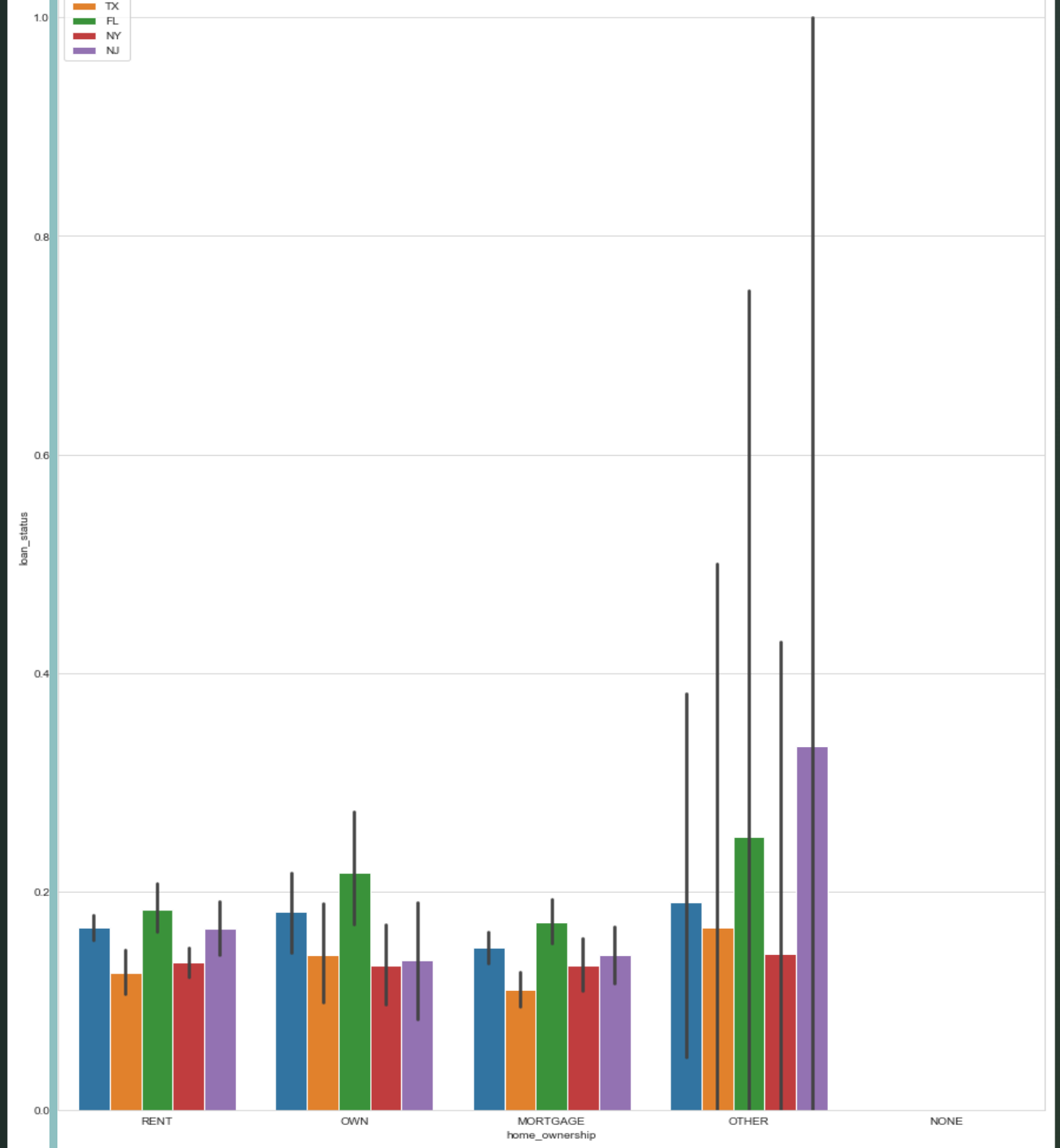


- For this we divide annual income>150000 as high,40000 to 150000 as medium and below that as low.
- We see that low income customers have higher rate of default while high income customers have low rate of default.

Bivariate analysis comparing the purpose of loan and income with default rate



- Here we see that customers with high income taking car loan and medical loan have the lowest rate of default while the customers with low income taking small business loan and renewable energy loan have the highest rate of default. An interesting thing to note here is that chances of education loan and other loan default are almost the same across the three income criteria



Comparing the house ownership in 5 major states and default rate

- We see that new Jersians without homes have the highest rate of default while Texans who have taken a mortgage have the lowest rate of default. In general, Floridians(either rented, own home, or mortgaged) who own a home have a high rate of default