

LENDING CLUB CASE STUDY

SUBMISSION

Name: Ankur Pandey

Name: Priyanshu

Business Objective

As lending company we want to analyse the past loan data and come up with factors that are precursors for a loan default.

This analysis address two types of risks:

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

So by analysing data we want to come up with driving variables/factors that can indicate loan defaults. The company can utilise this knowledge for its portfolio and risk assessment.

Dataset Analysis

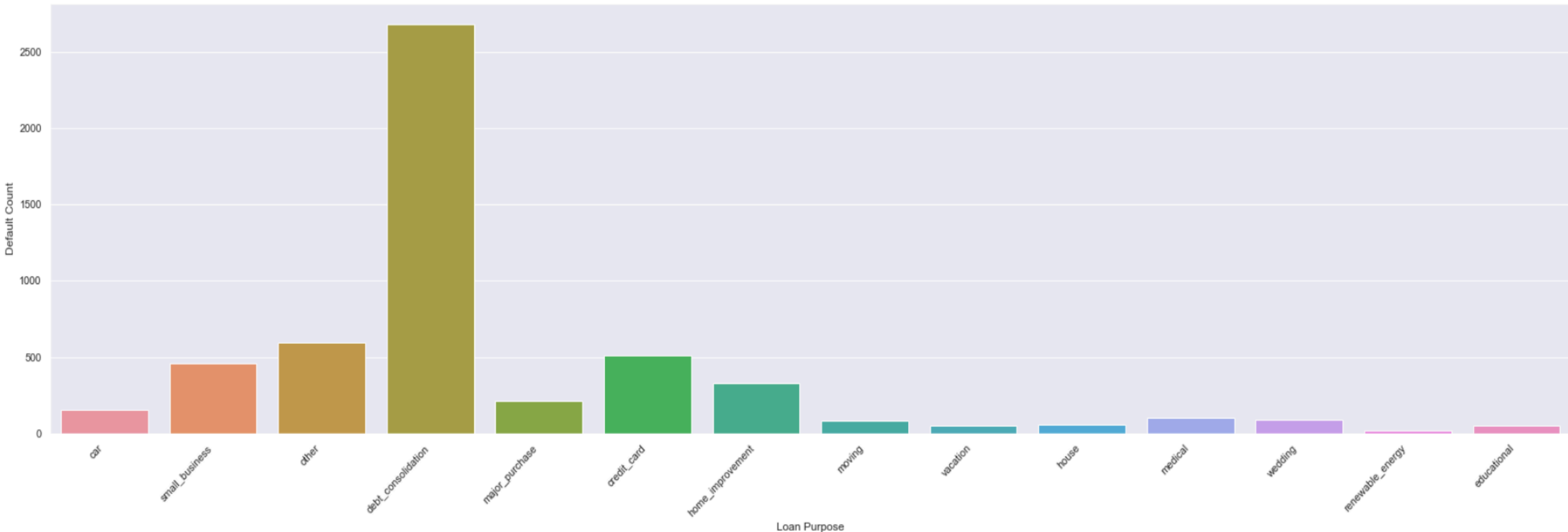
- Check for missing values and Nan values in rows/column's
- Ideally, remove column's having more than 90% missing values
- Ideally, remove the row having more than 5% missing values
- Convert date column's to date time as *Derived Metrics*
- Bin the variables across discrete categories
- Univariate analysis
- Segmented univariate analysis
- Bivariate analysis
- Bivariate analysis across categorical variables

Analysis Approach

Following variables were considered for this case study analysis:

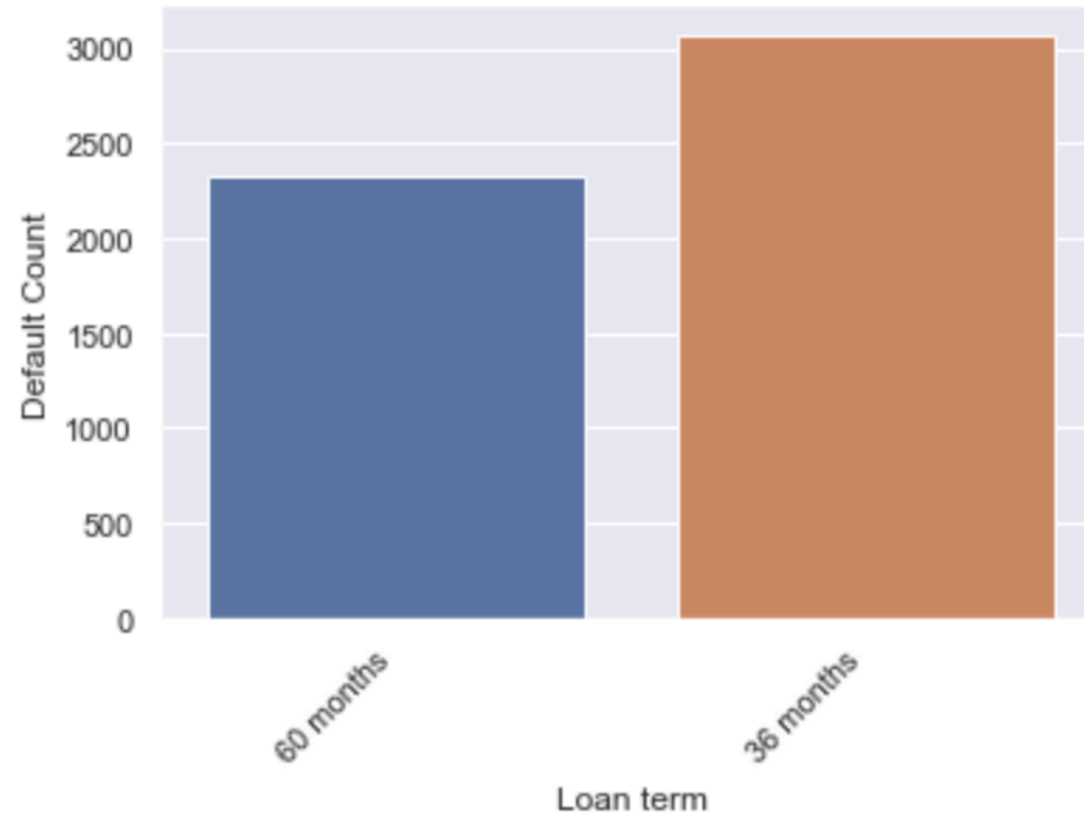
- Loan Purpose
- Loan Grade
- Home Ownership
- Loan Issue Year
- Loan Sub-Grade
- Verification Status
- Interest Rates
- Installment Amount
- Annual Income
- Employment Length
- Loan Amount
- DTI (A ratio calculated using the borrower's total monthly debt payments on the total debt obligations)
- Earliest Credit Line
- Customer Address state
- Number of open credit lines

Univariate Analysis: loan purpose



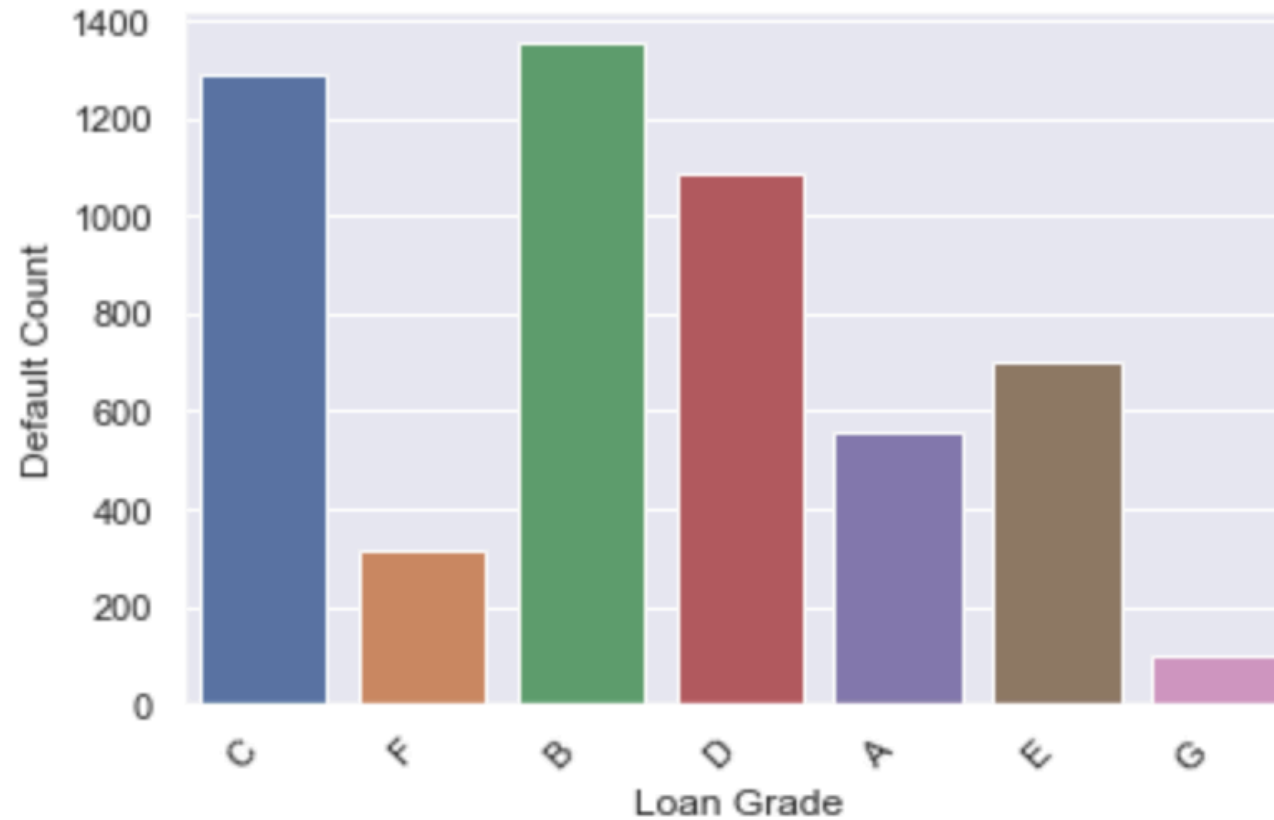
Observations: This suggest debt consolidation have major defaults over 2500 loan applications. Which is significantly higher than the rest of the loan purposes

Loan Term



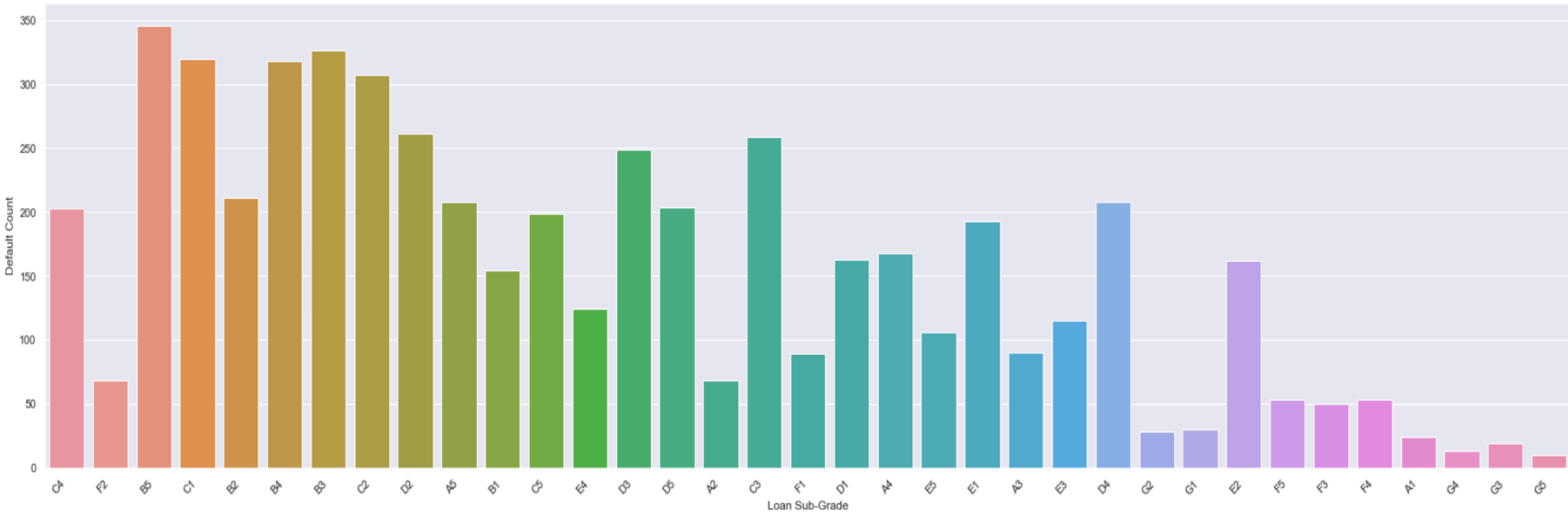
Observations: Its clear from the graph that 36 months loan term have more default rates that 60 month but difference is not much.

Loan Grade



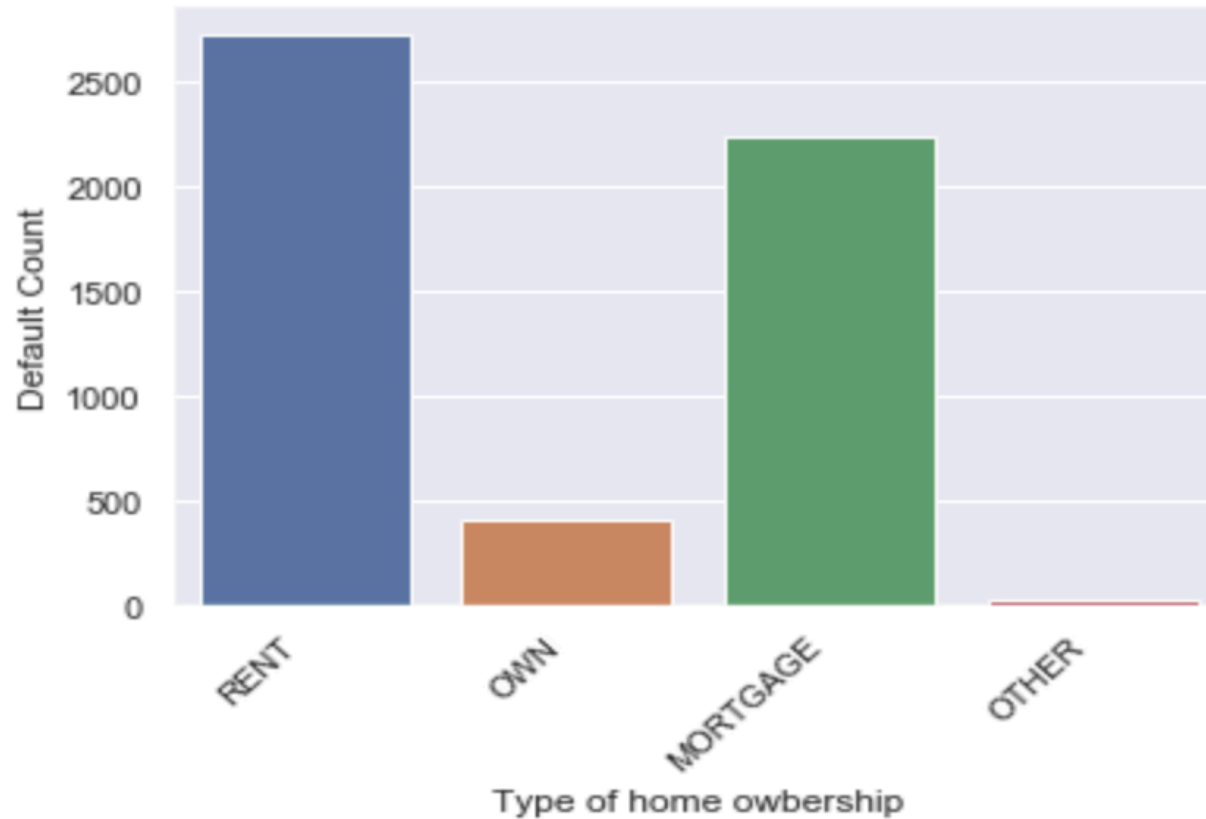
Observations: This suggest Loan Grade B,C and D are most defaulted loan grades

Loan Sub Grade



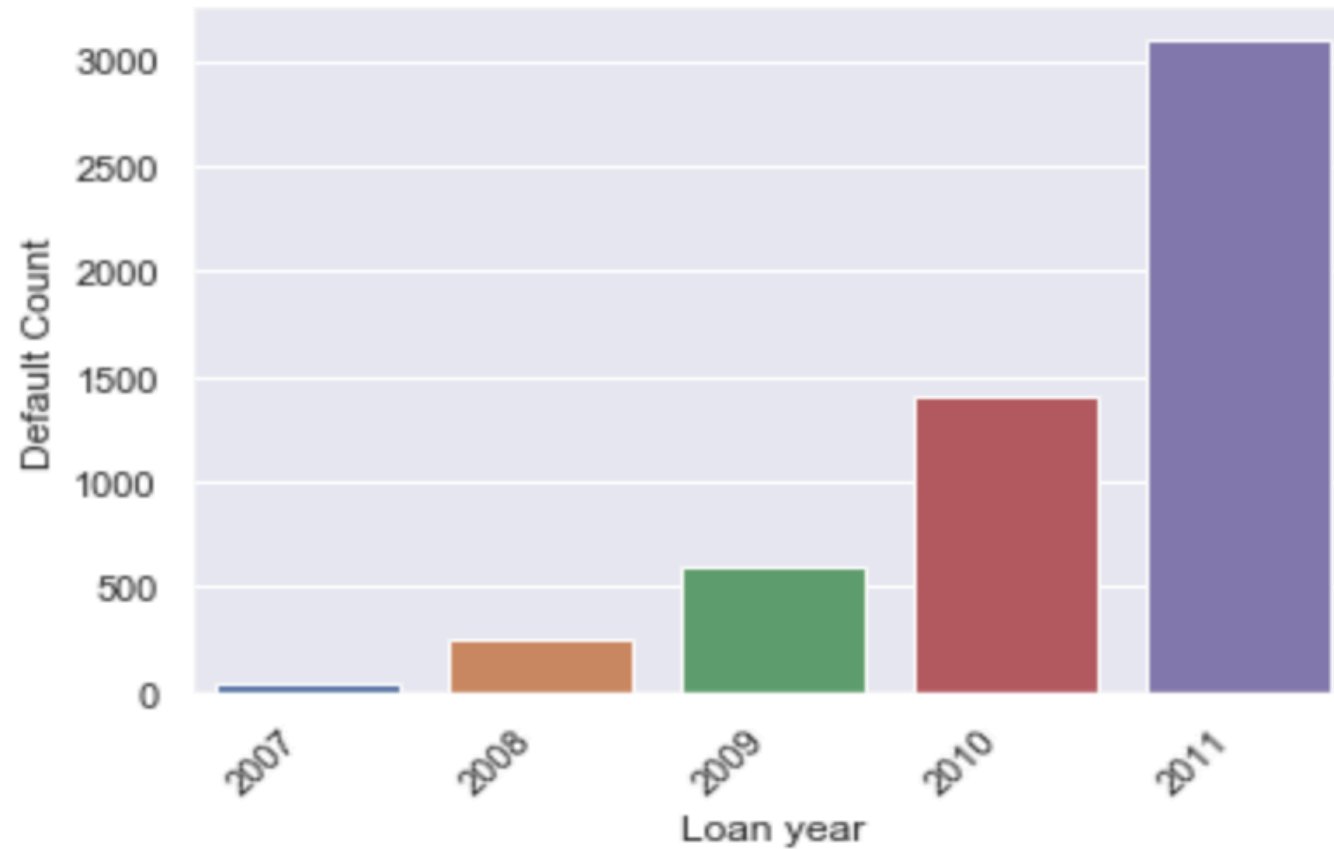
Observations: This observation confirm that observation in previous Grade analysis except few subgrades Loan Grade B, C and D have more loan defaults number.

Home Ownership



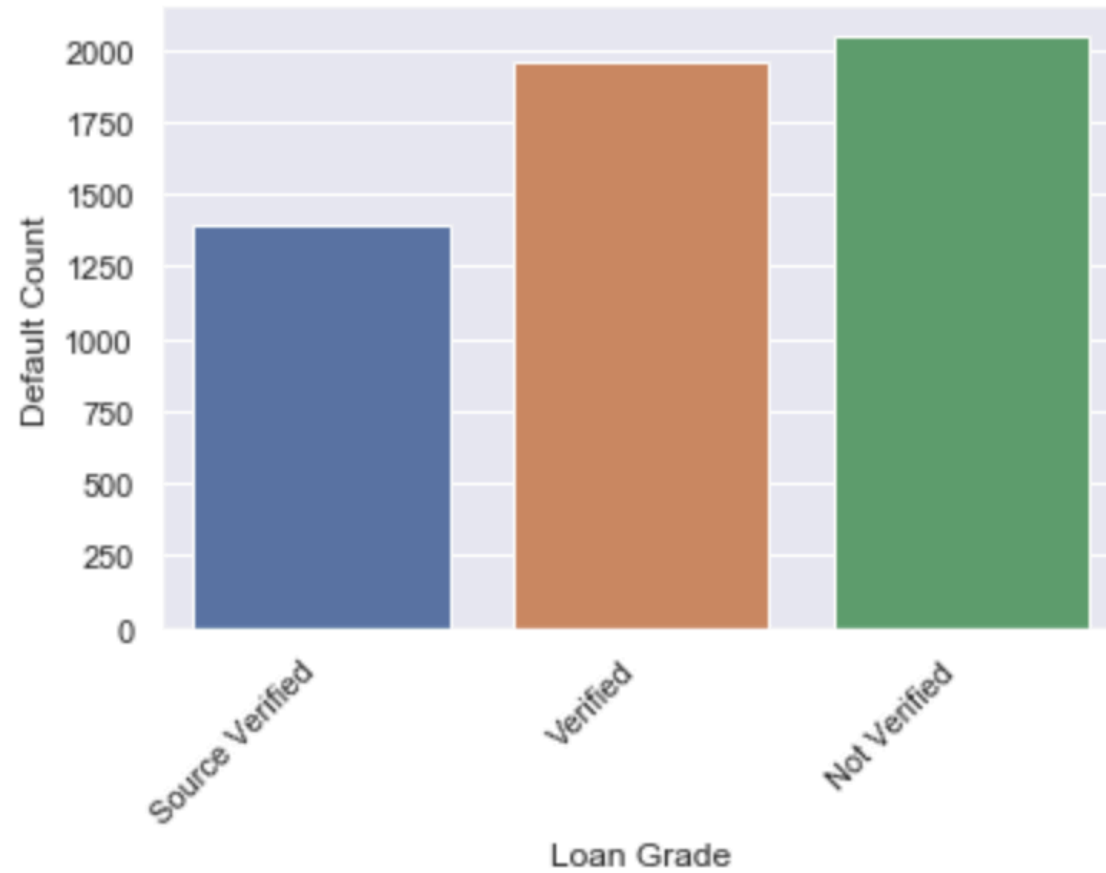
Observations: This suggest people living on rent have over 2500 loan defaults and is followed by Mortgage which is close to 2200. Other types like OWNED property and “others” are NOT that significant.

Loan Issue year



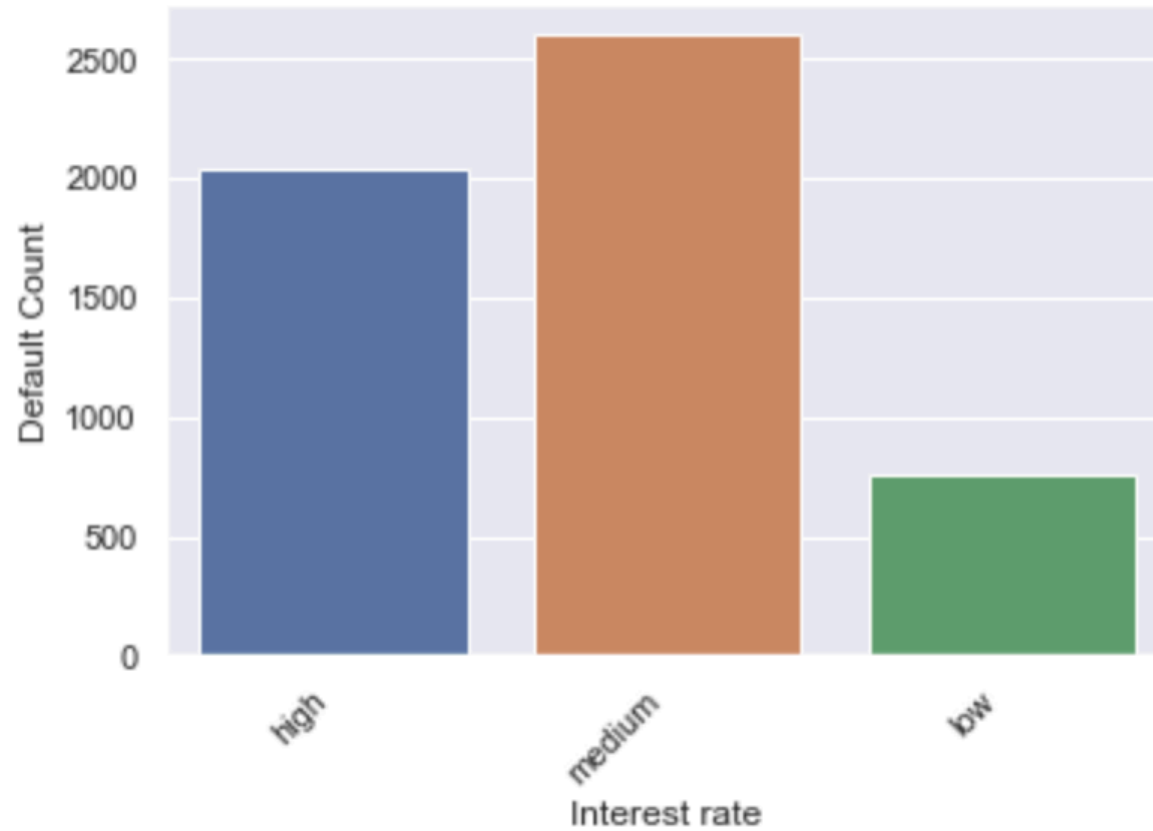
Observations: This suggest loan issued in year 2011 were defaulted more than any other year in past. The trend is increasing year by year.

Verification Status



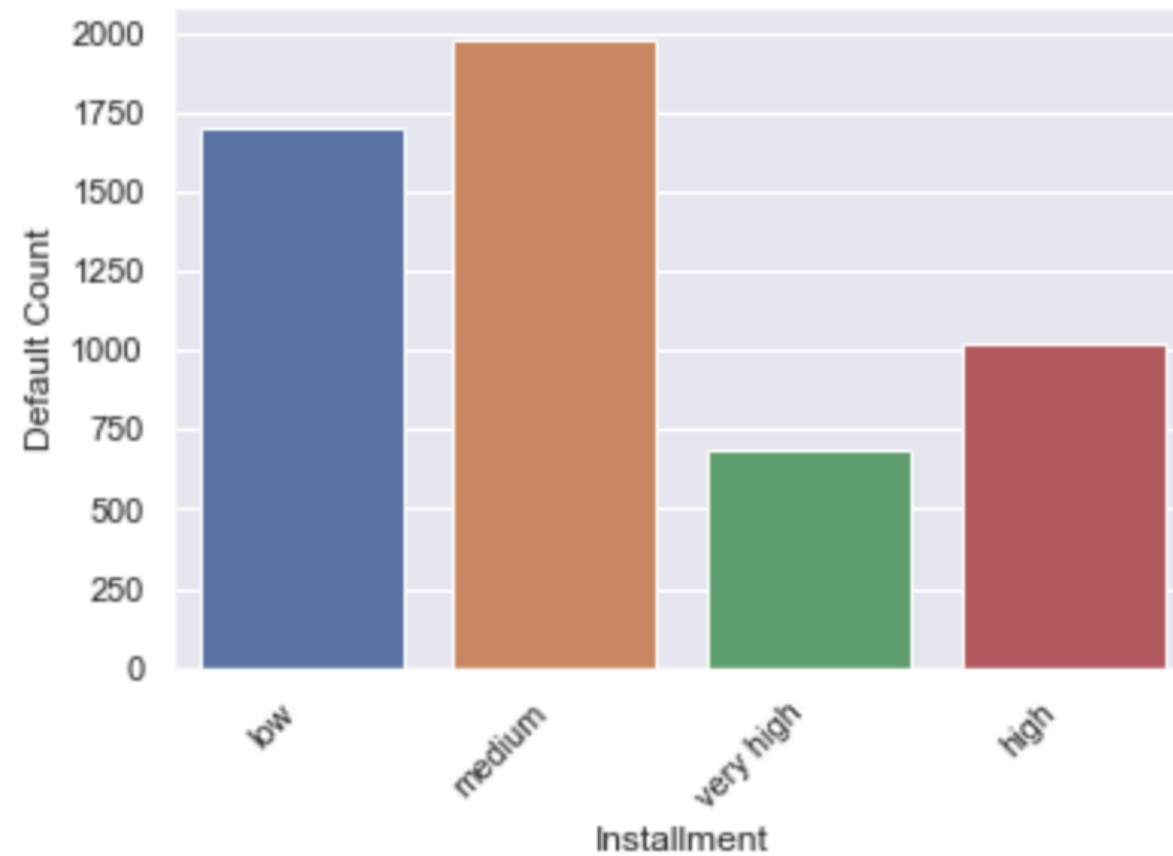
Observations: Graph suggests the loan which had source verified have least number of defaults as compared to verified or Not verified applications. However the difference isn't significant.

Interest Rate



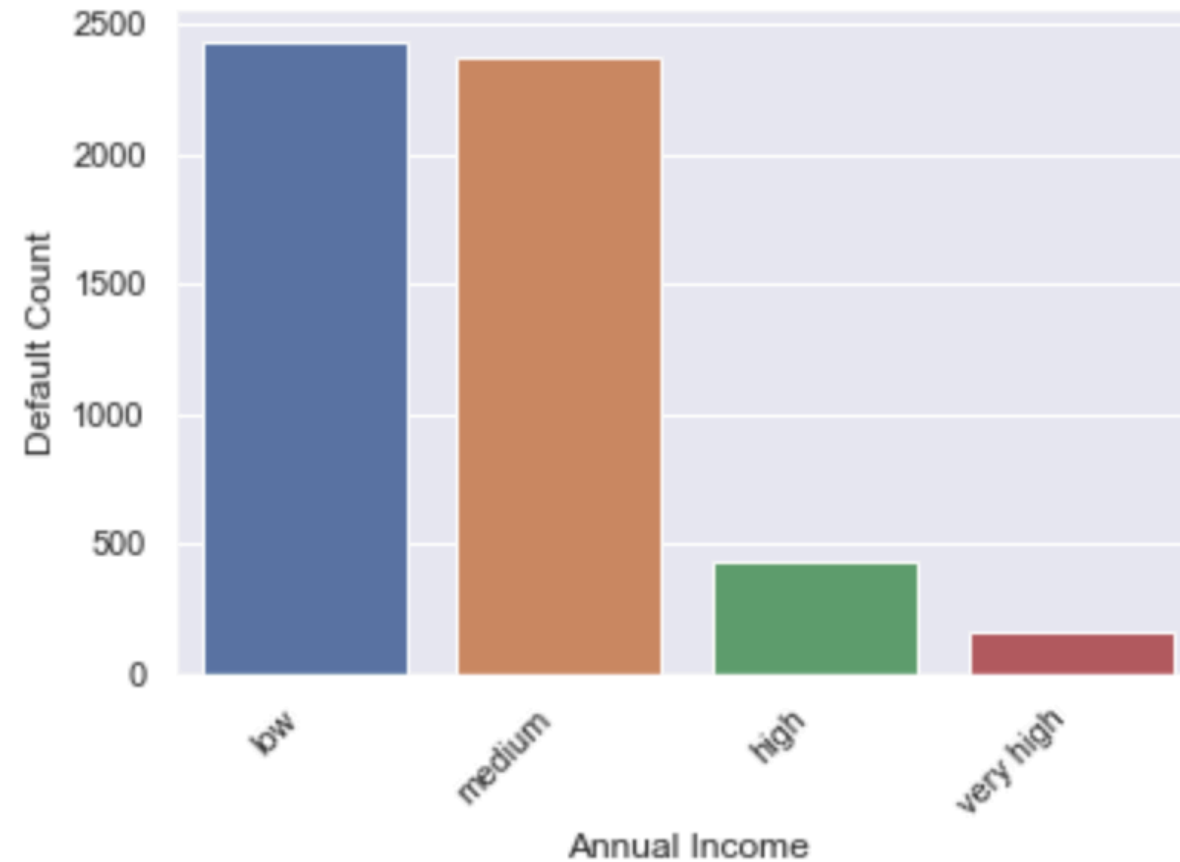
Observations: Graph suggests that loans with medium rate of interest (10.1-15%) have more defaults above 2500.

Installment Amount



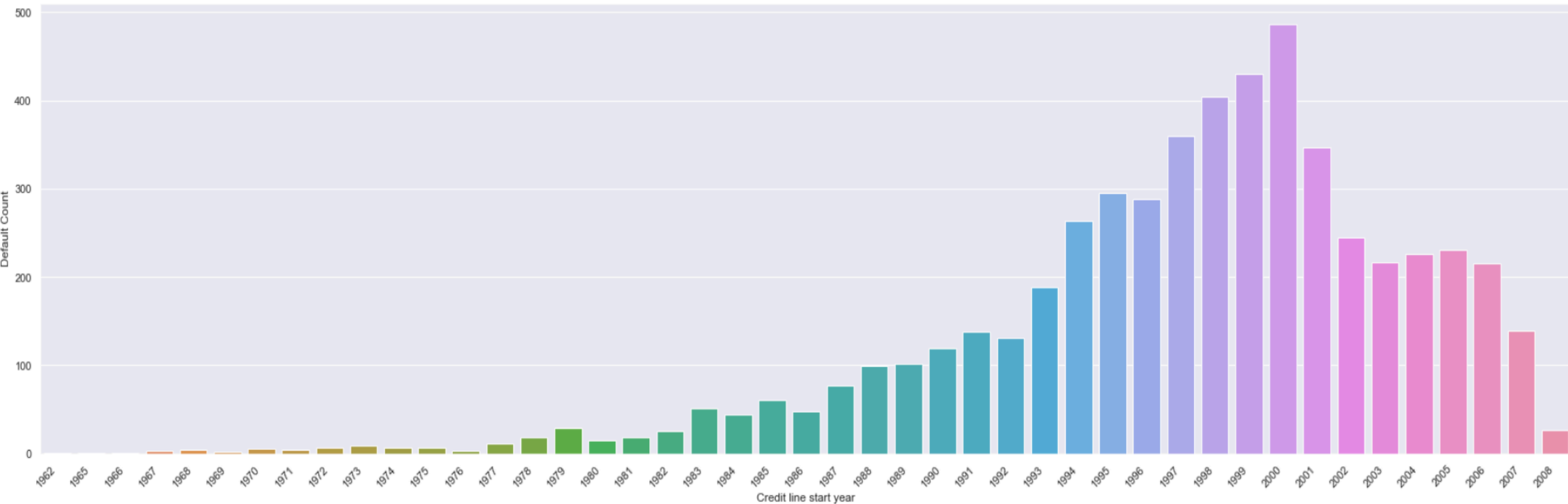
Observations: This graph suggests EMI of medium category (\$200 - \$400) and low category (less than \$200) have more default rate.

Annual Income



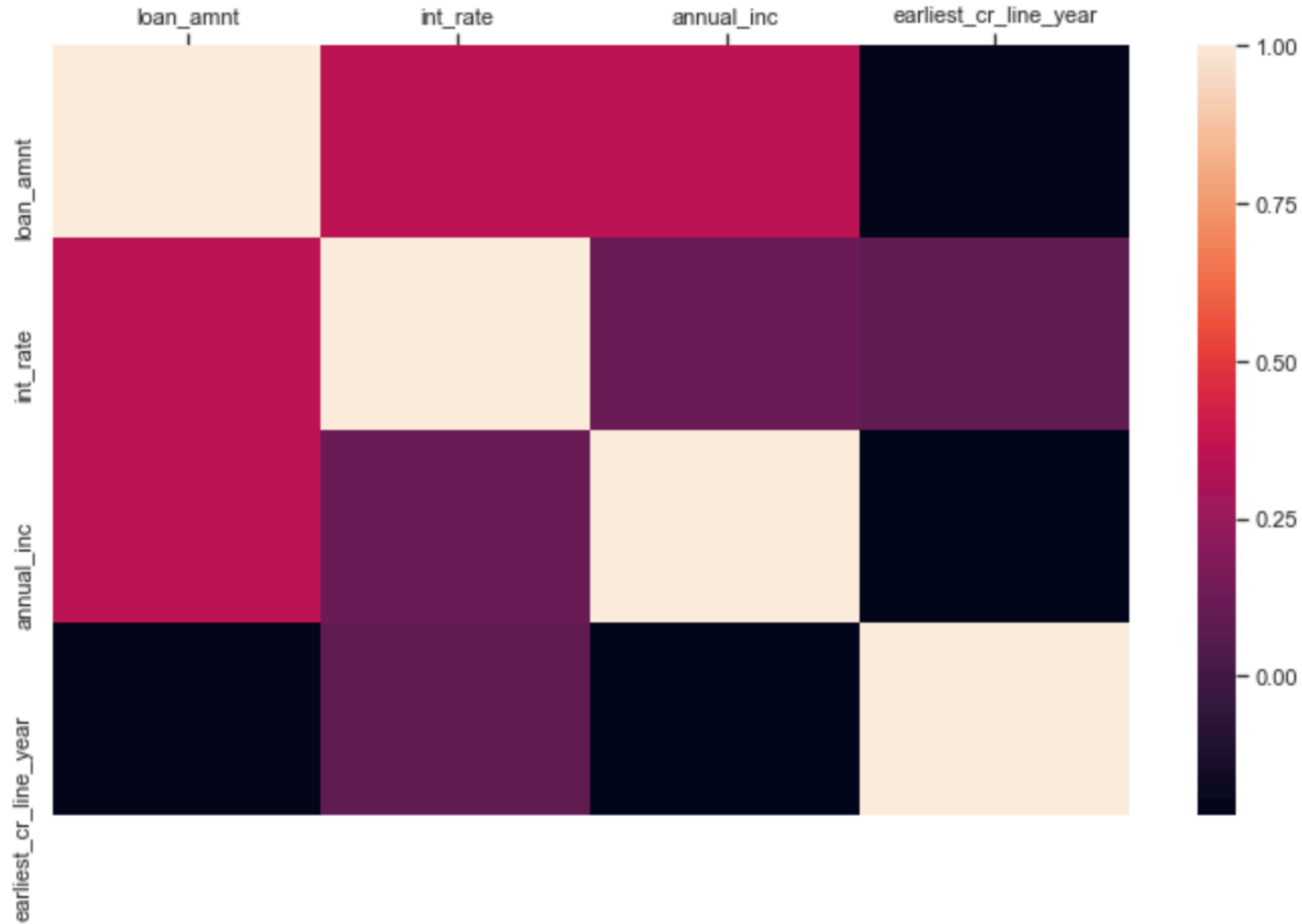
Observations: We can observed that people having low (less than \$50000) and medium (\$50000 to \$100000) income range are high in defaults.

Earliest Credit Line



Observations: We can observed that applicants having credit line started after 1993 have defaulted more and this trend reached its peak in year 2000. While after that data shows a serious decline in number of defaults.

Bivariate Analysis: Correlation

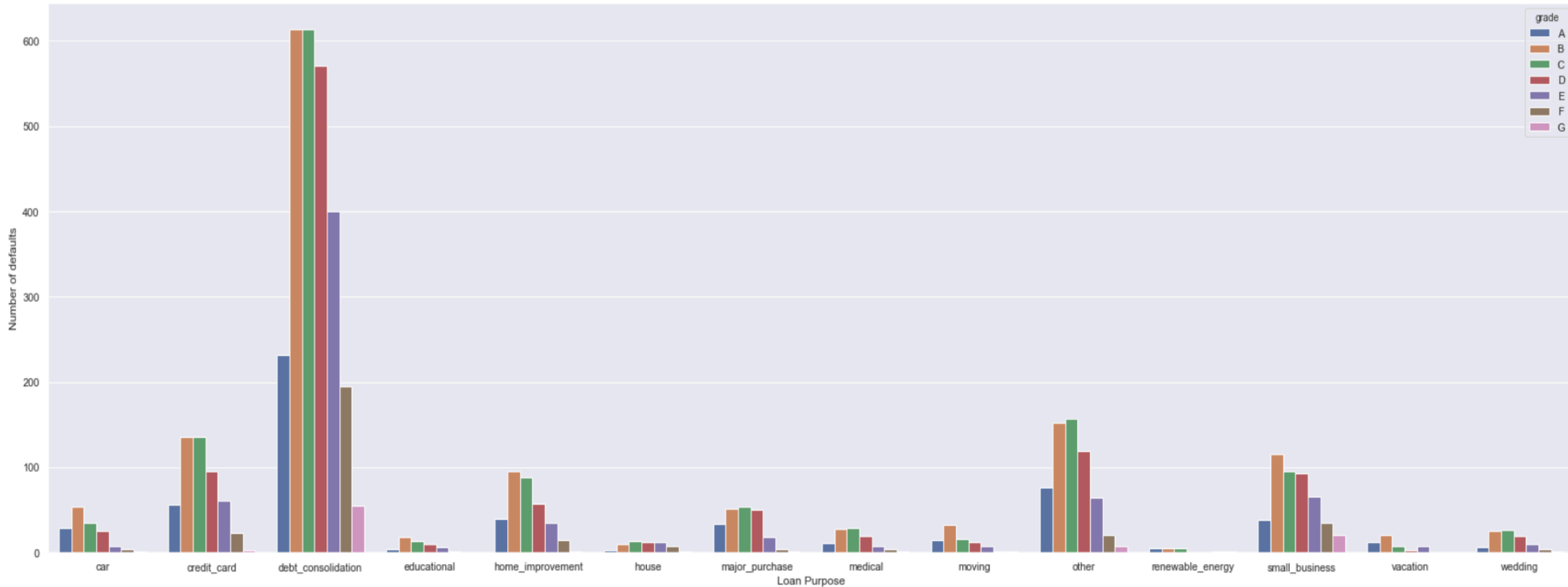


Observations:

Annual income and Loan amount have a negative correlation.

Annual income, Loan amount and Interest rate are Somewhat correlated to each other around 0.34.

Grade Vs Purpose

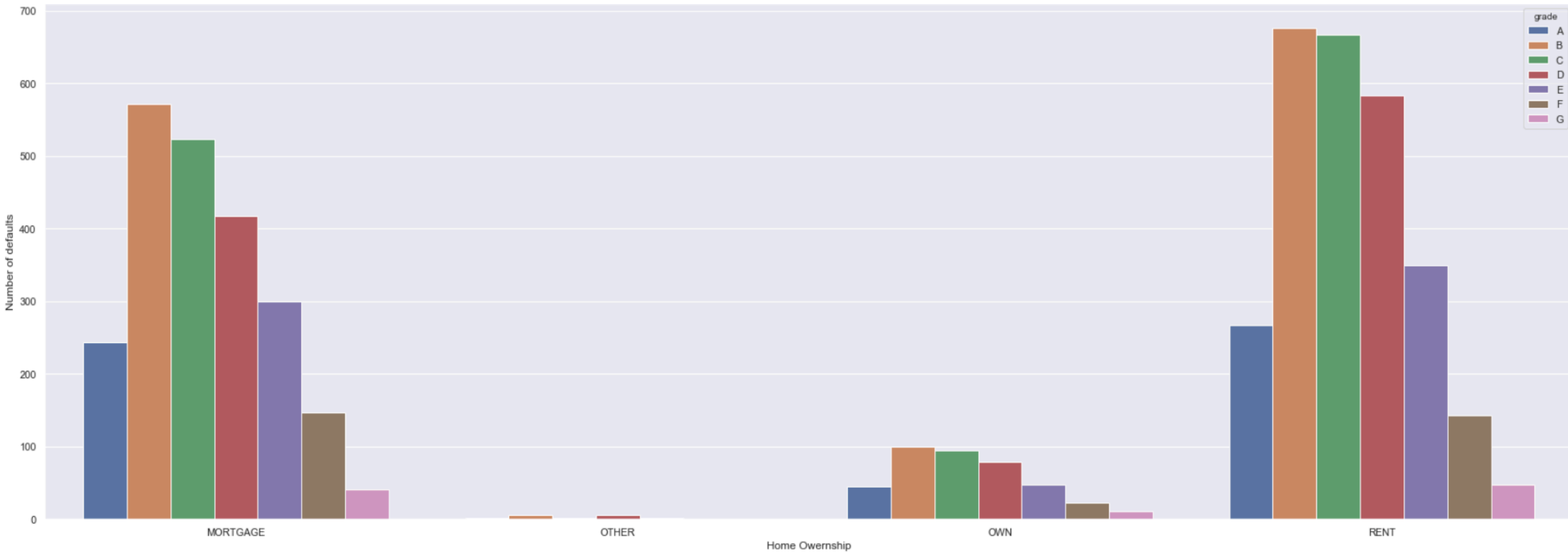


Observations: From univariate analysis of “purpose” we observed that “debt_consolidation” have higher no. of defaults.

Similarly in grade’s univariate analysis, we observed loan grade B, C and D types have more defaults.

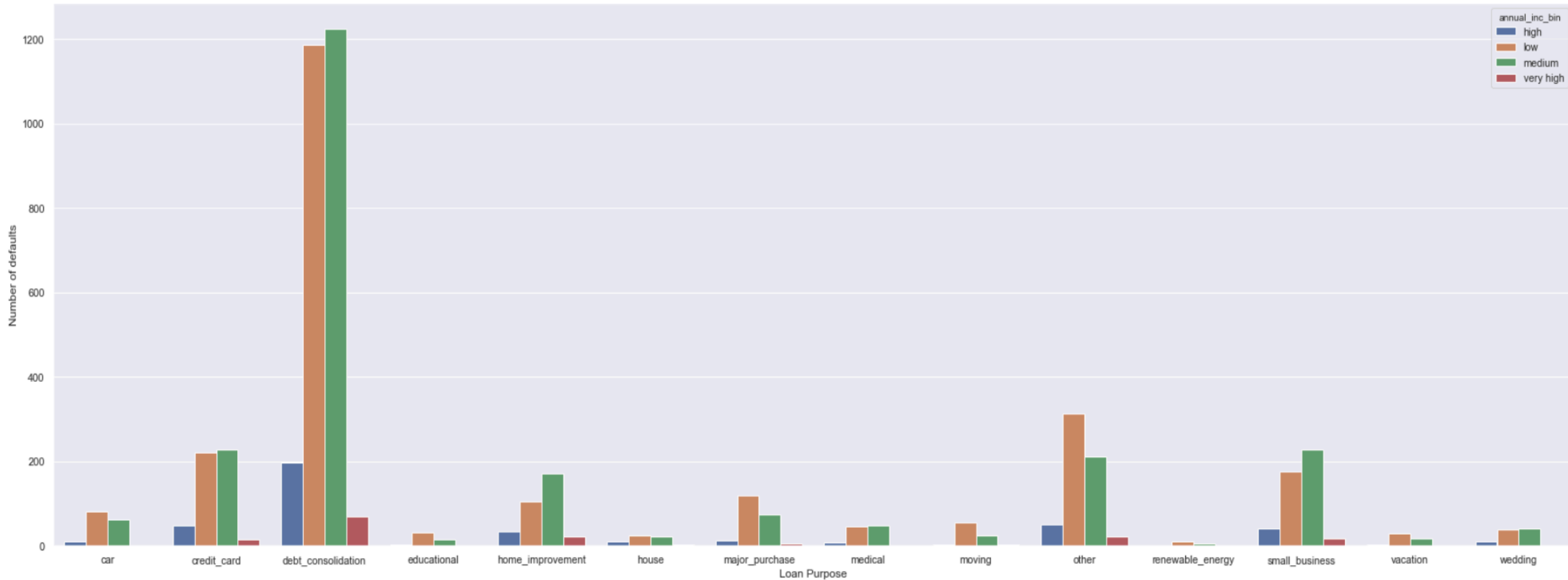
While comparing these two we can confirm the observation that loan grade B, C and D have higher loan defaults comparing with other grade in almost all loan purposes.

Grade Vs Home Ownership



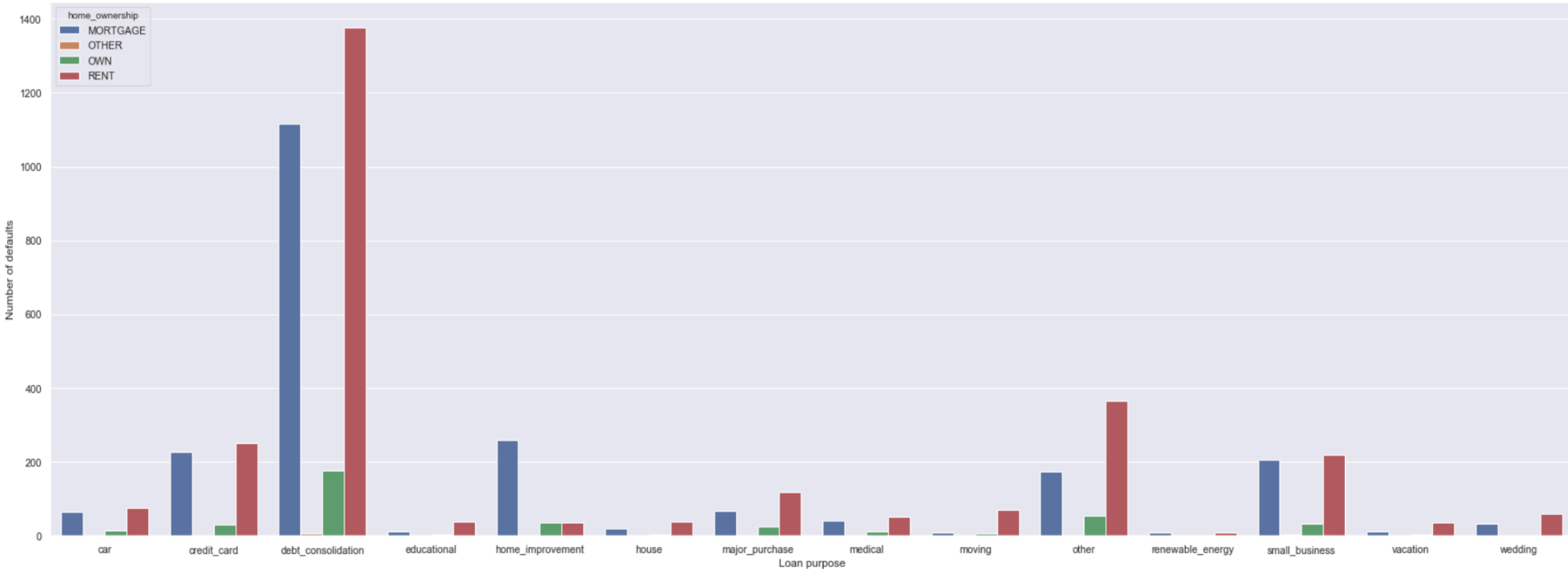
Observations: From univariate analysis of home_ownership we observed that “rent” and “mortgage” have higher no. of defaults. Similarly in grade univariate analysis, we observed loan grade B, C and D types have more default. While comparing these two we can confirm the observation that loan grade B, C, and D have higher loan defaults comparing with other grades in almost all type of “home_ownership”.

Purpose Vs Annual Inc.



Observations: From univariate analysis of purpose we observed that “debt_consolidation” is higher no. of default. Similarly for annual_income we observed that people having low (less than \$50000) and medium (\$50000 to \$100000) income range are high in defaults. While comparing these two we concluded that this observation is valid for all types of loan purpose categories.

Purpose Vs Home Ownership



Observations: From univariate analysis of purpose we observed that “debt_consolidation” is higher no. of defaults.

Similarly for “home_ownership” we observed that “rent” and “mortgage” have higher no. of defaults.

While comparing these two we can confirm the observation that expect “home improvement” type of loan purpose, loans on RENT type of ownership is higher in number of defaults followed by mortgage type.

Conclusions

The major driving factors/indicator for loan defaults:

1. Loan Purpose
2. Income level
3. Loan Grade
4. Home ownership
5. Installment
6. Interest rates

Minor variables that affect default rates are:

1. Loan Term
2. Loan Amount
3. Employment length
4. Earliest credit line
5. Number of credit lines
6. Verification status

Conclusions

A red flag can be raised if :

Loan properties:

- **Loan Purpose is Debt consolidations**
- **Loan Grade is B, C and D**
- **Installment is less than \$400**
- Interest Rate is 10-15%

Applicants properties:

- **Income level is less than \$100000**
- **Applicant is on Rental or mortgage property**
- Earliest credit line is after 1992 but before 2008