Lending Club Case Study

SUBMISSION

Group Members:

Mukul Patel

Nagaraj Gandge

Lending Club Case Study

Background:

Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

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:

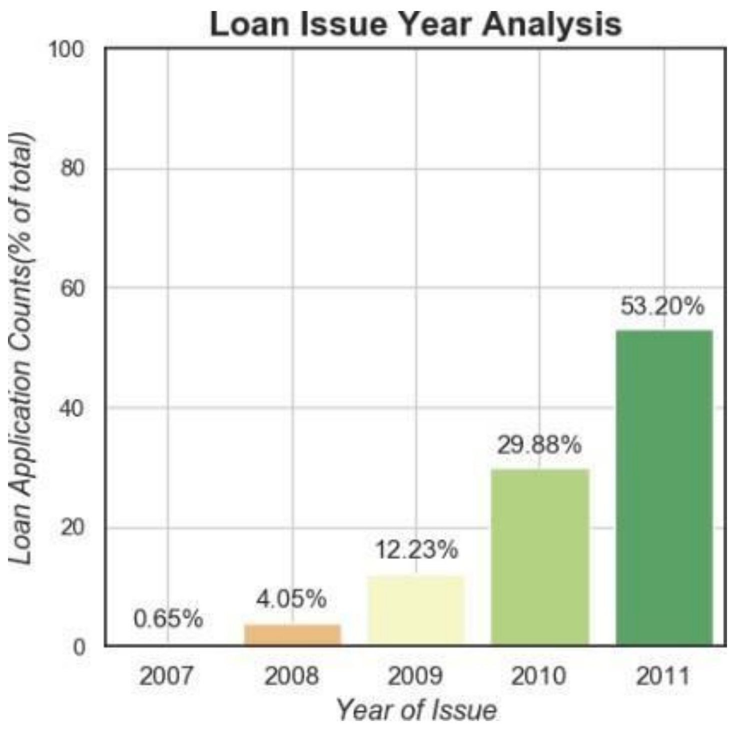
- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company

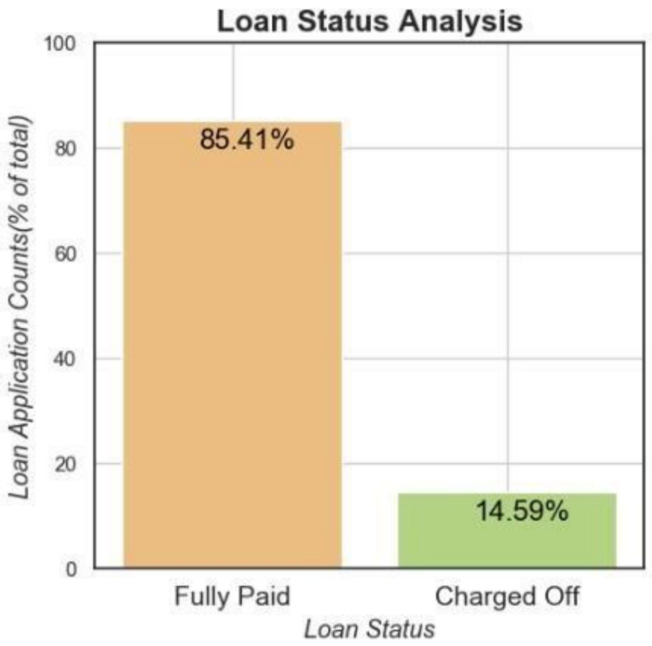
Objectives :

- Identify variables that provide strong indicators of potential loan default thus helping Lending club to decide approval/rejection of loan.
- Identification of such applicants using EDA is the aim of this case study.

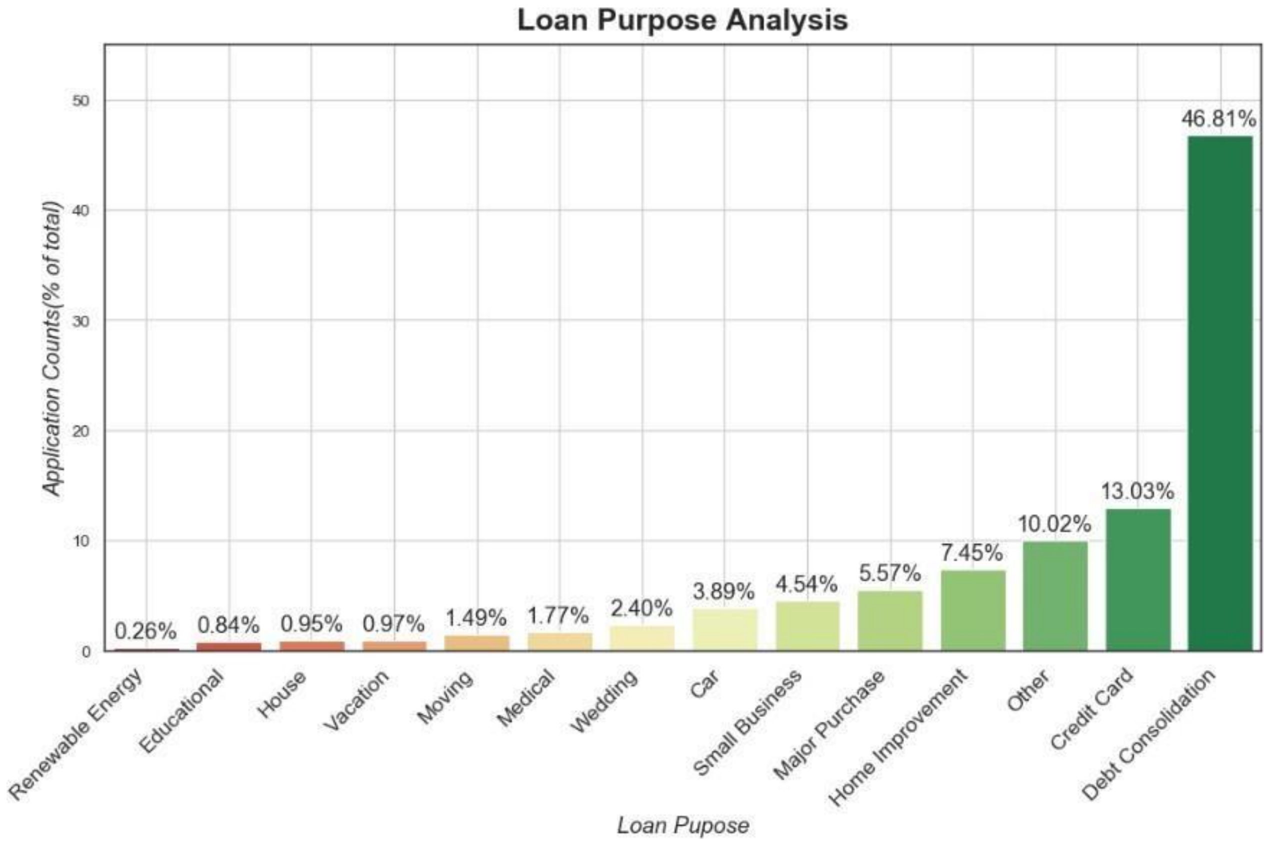
Univariate Analysis: Insights

- Overall Default Rate stands at 14.59%.
- Approval rate of loans jumped by around 78% from the year 2010 to 2011.
- Approved Loan amount almost doubled between years’ 2010 and 2011.





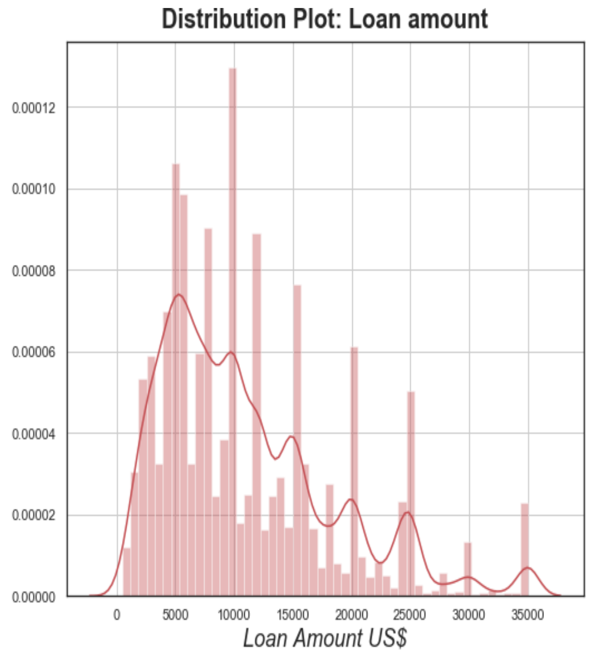
Univariate Analysis: Insights

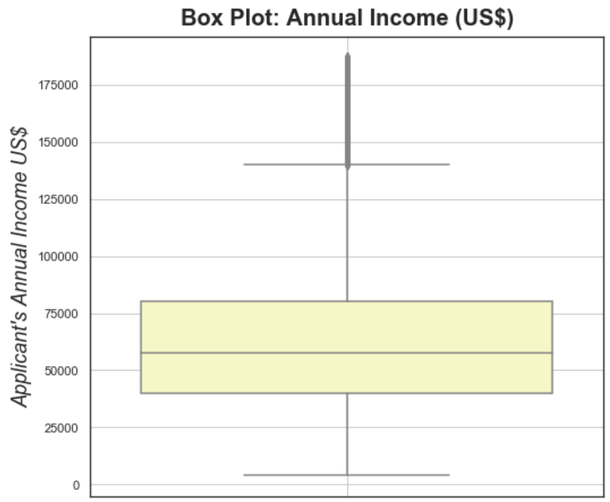


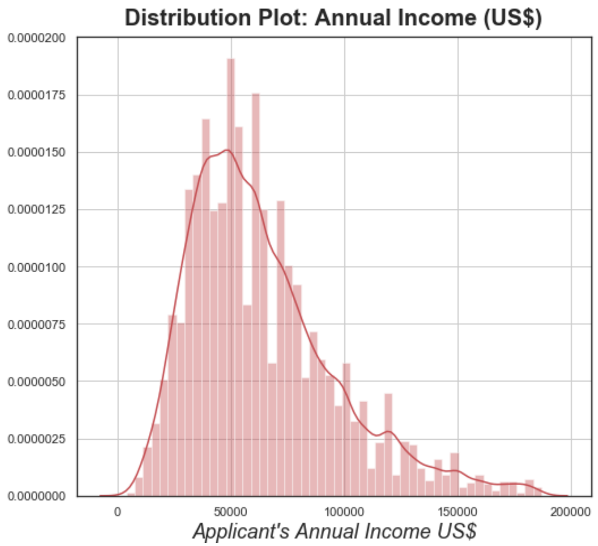
- ‘Debt Consolidation’ followed by ‘Credit card’, are the top two reasons for loan applications amongst approved loans.

Univariate Analysis: Insights

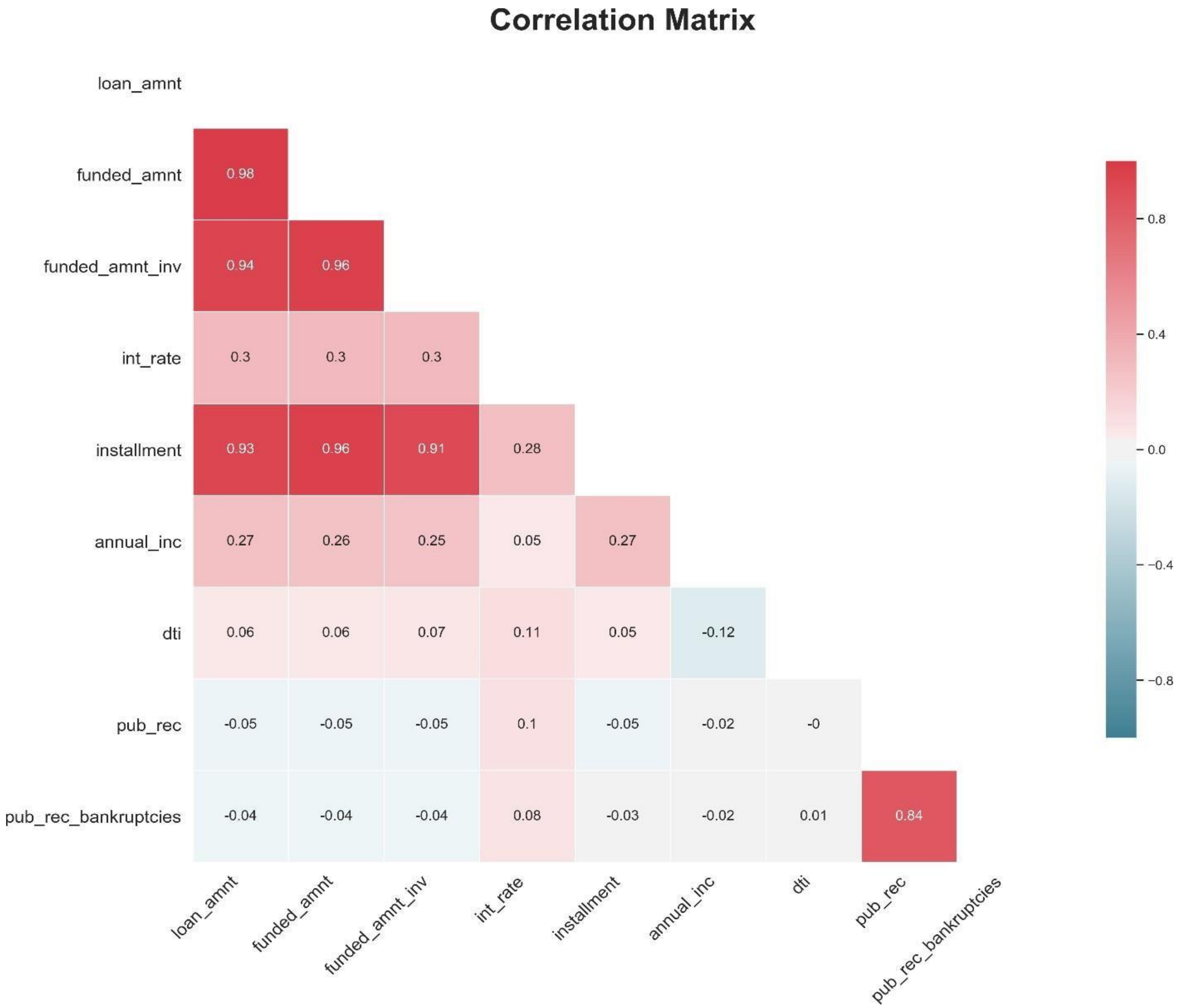
- 50% of loan applicants request a loan amount between 5.3K USD and 15K USD.
- Annual Income was highly skewed as expected; ignoring the outliers, the average salary of applicants is 68.78K USD.
- 50% of the applicants earn between 40K USD and 82K USD annually.





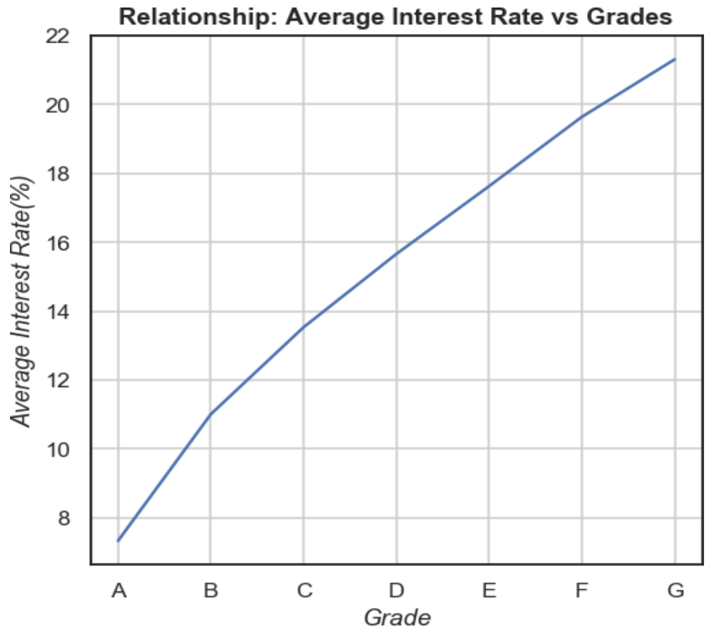


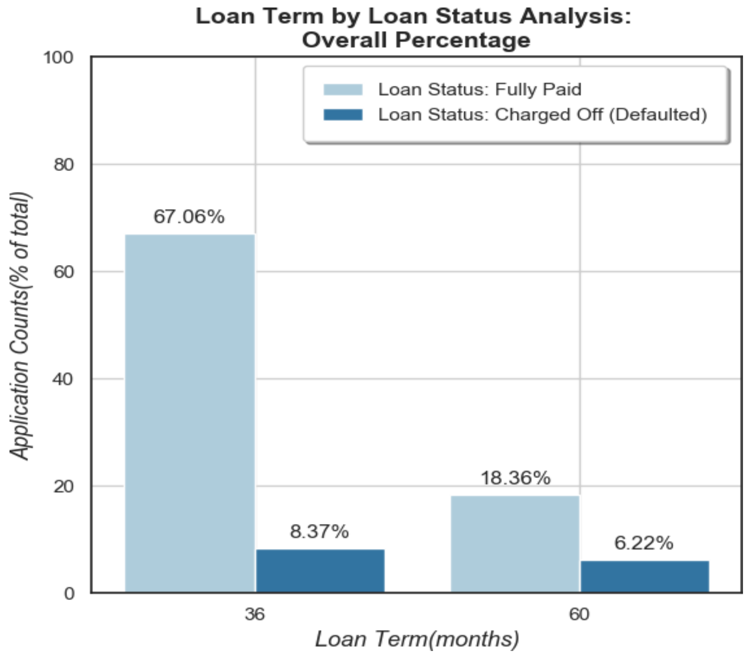
Bivariate Analysis



- Pearson Correlation plot between numeric/continuous variables retained
- Numbers are indicative of the strength of correlation between variables

Bivariate Analysis

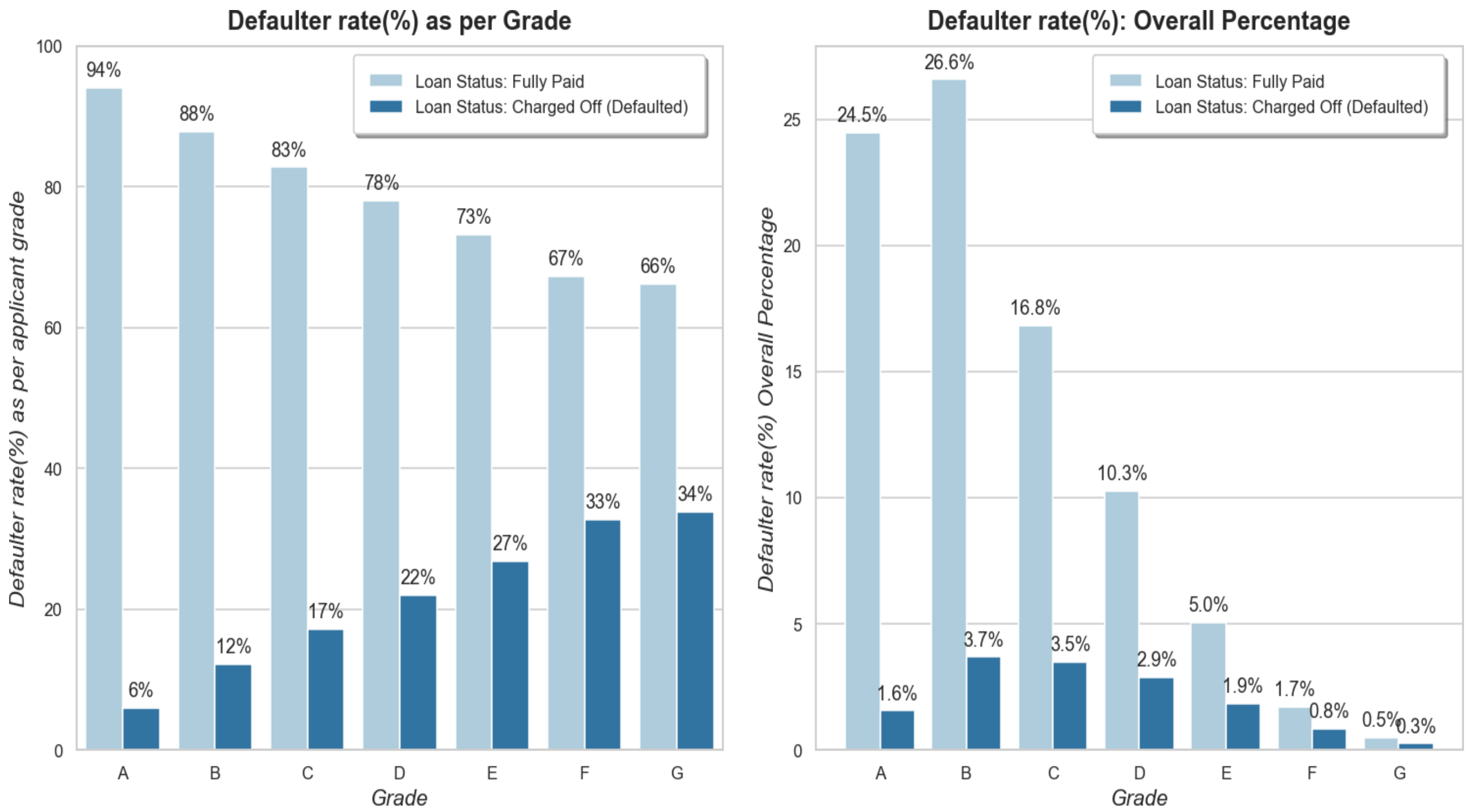




- Average Interest rates increase as the risk associated with the applicants increase, which can be denoted by the grades assigned to applications. It is an effective process to follow.

- Loan approved for 36 months seems to have a greater % of defaulters as compared to loans approved for 60 months term.

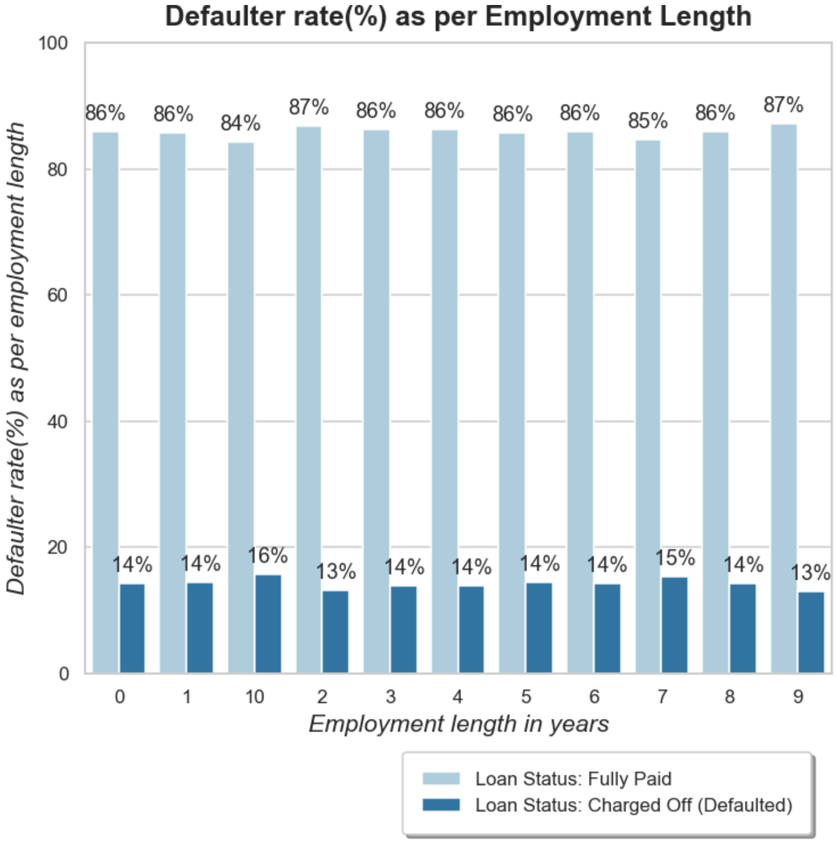
Bivariate Analysis: Grade Vs Loan Status

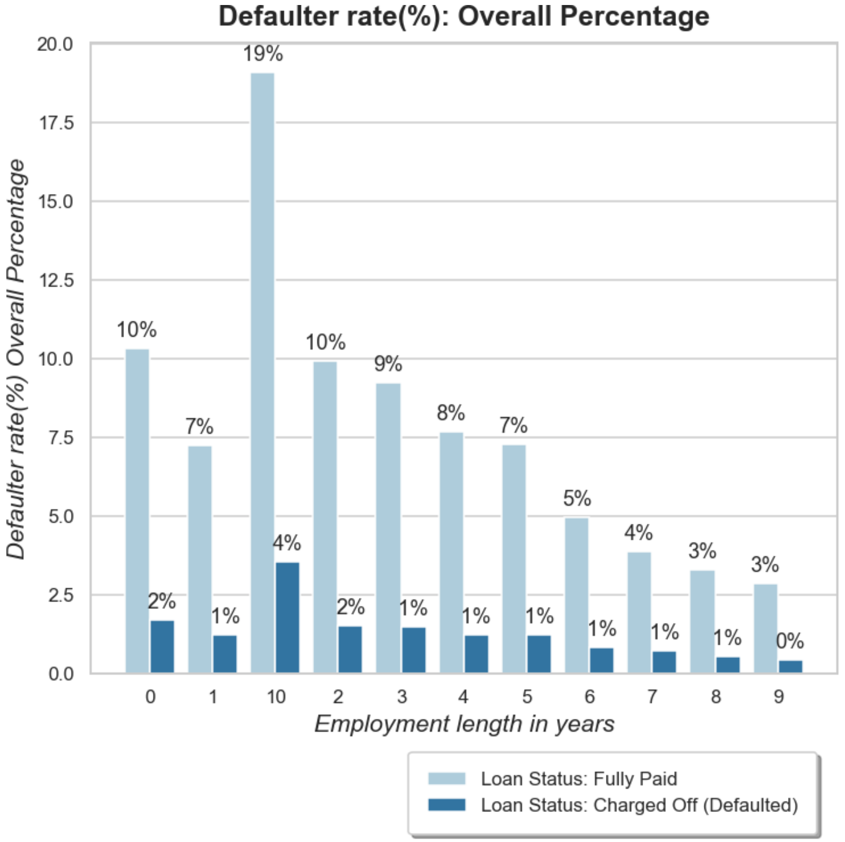


- Percentage of defaulters is significantly high within the grade categories, 'F' and 'G'. LC should be more vigilant while approving loan applications for high risk applicants.

- Largest Proportion of approved loan defaulters belong to Grade 'B'(3.7%) and 'C'(3.5%) category.

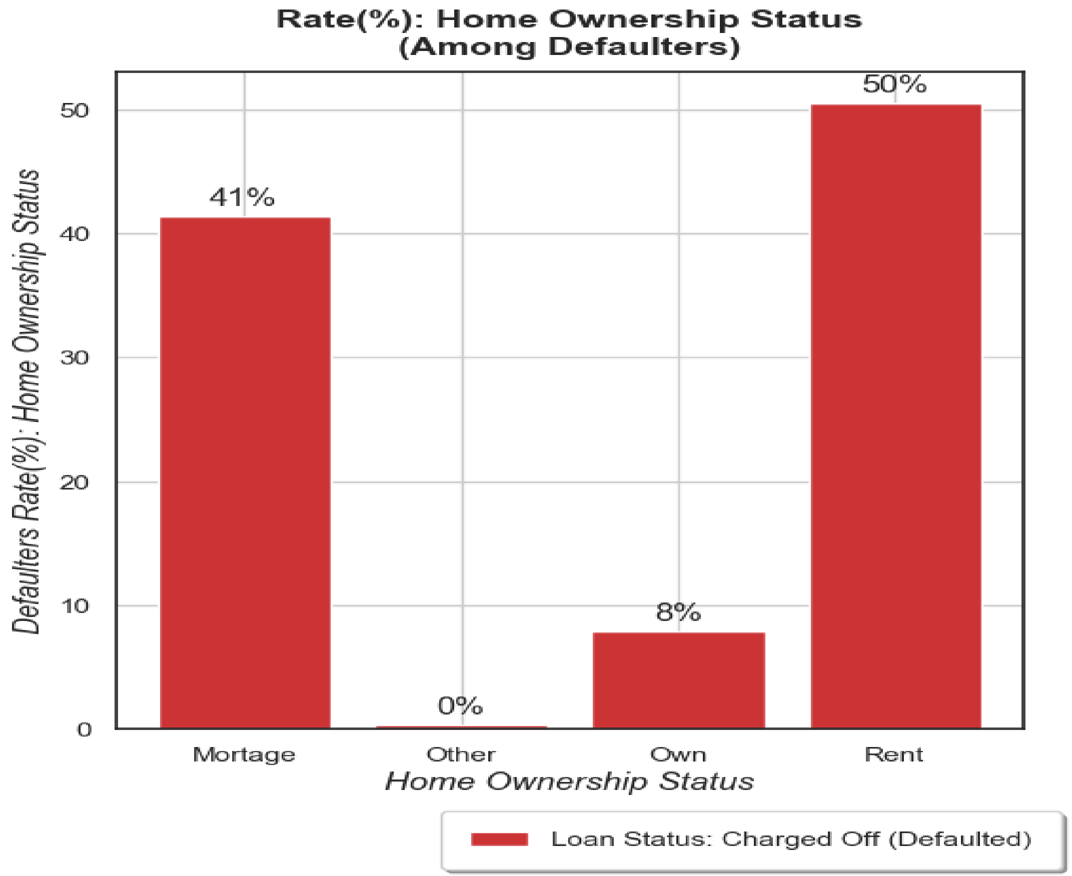
Bivariate Analysis: Loan Status Vs Grade & State





- Maximum percentage of defaulters are found in the 10 years (4%) employment year category. After it, we see applicants with less years of experience i.e. around 0 to 2 years, following the trend.

Bivariate Analysis



- 91 % of the defaulters already have a mortgage taken out or live on rent.

Conclusion

Driving Factors (or driver variables):

- Grade :- Default Rate is high in high risk loan applicants. It would be important for LC to thoroughly vet high risk loan applications.

- Annual Income :- Applicants from 'Low'(<=45K USD) and 'Medium'(45K-90K USD) income group have a greater share of defaulted loans.

- Employment Length : Maximum number of defaulters have 10/10+ years of experience and 0 to 2 years of experience. Hence, LC should be take this aspect into consideration while lending loans.

- Loan Purpose :- The top two reasons for loans are debt consolidation and credit card. Such applications should be carefully assessed.

Thank You