LENDING CLUB CASE STUDY

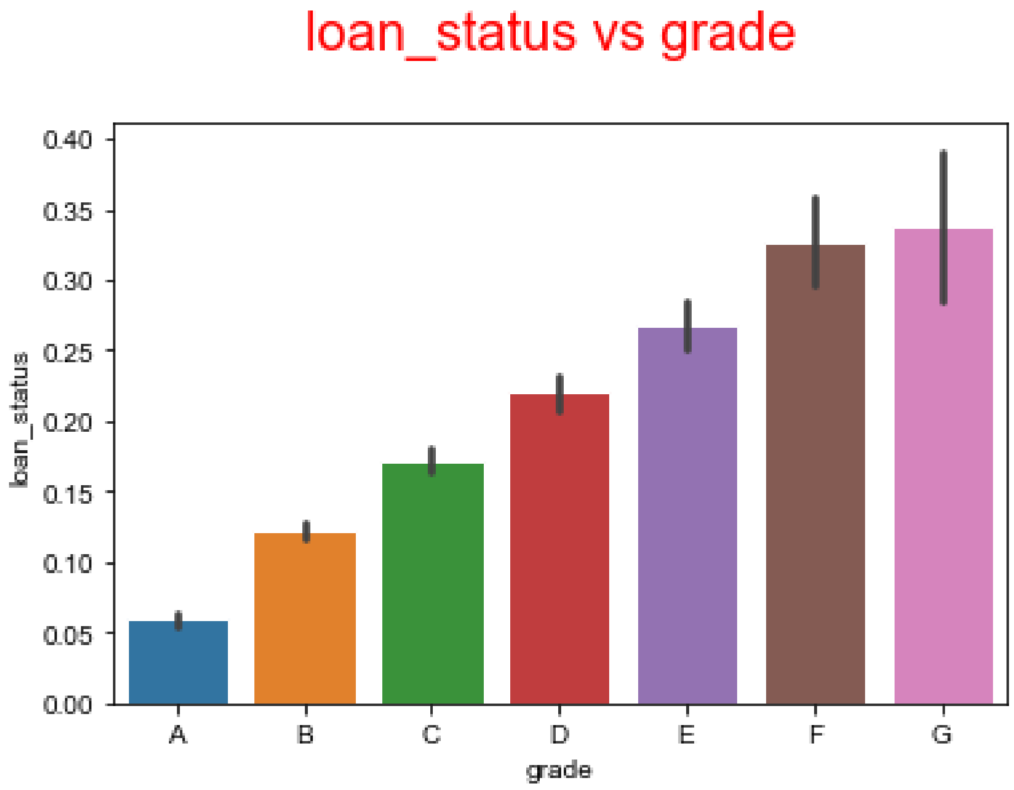
Shuchi Jain

Objective

- The data in the loan.csv file contains information about past loan applicants and whether they ‘defaulted’ or not. The aim is to identify patterns that indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

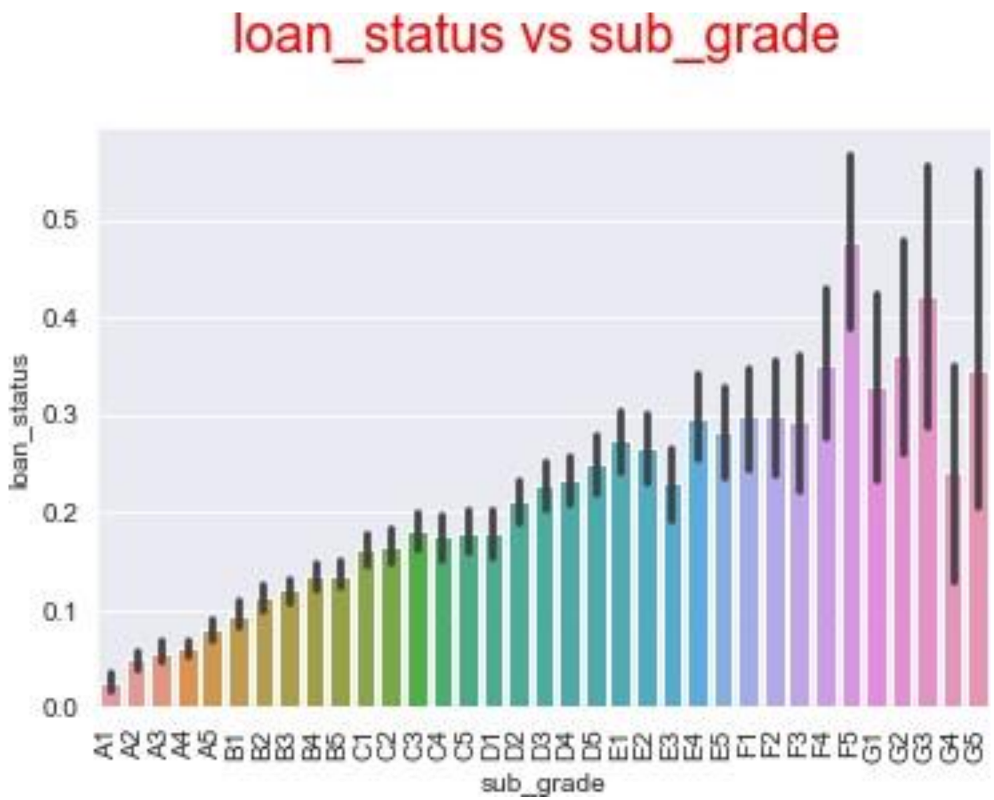
UNIVARIATE AND BIVARIATE ANALYSIS

Analysis (Loan_status vs Grade)



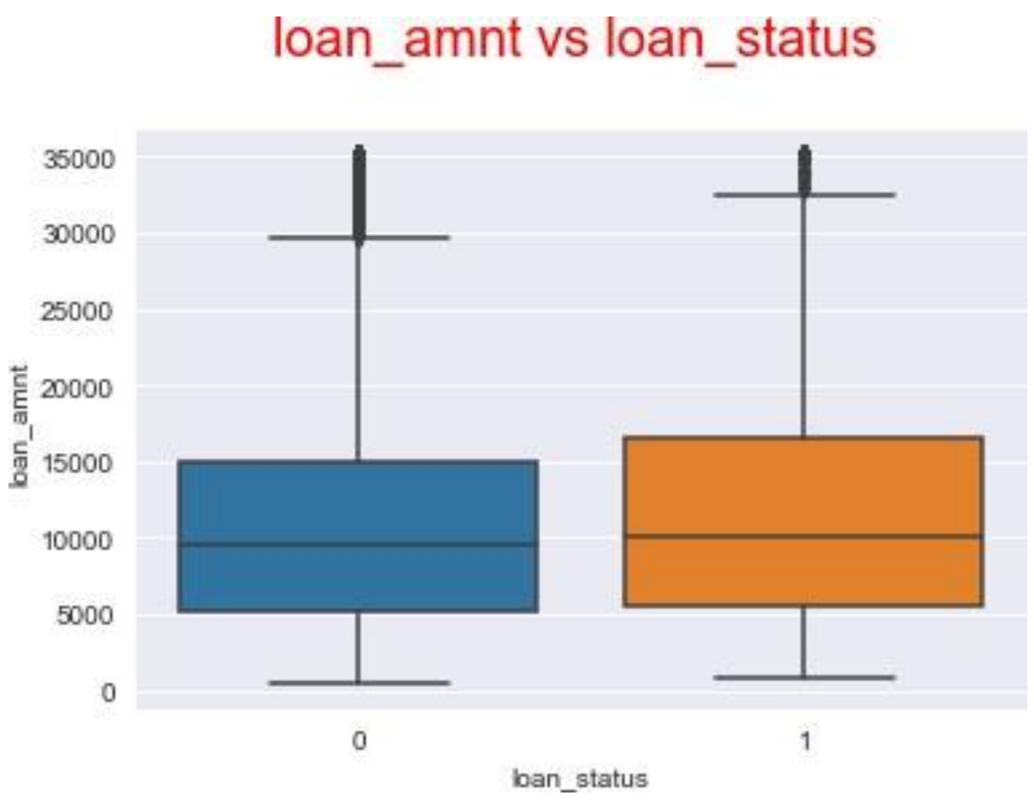
After analysing the dataset with respect to **grades** and **loan_status** we came to the conclusion that the loans belonging to the grades D, E, F and G are considered to be a high-risk cases in increasing order.

Analysis (Loan_status vs Sub_Grade)



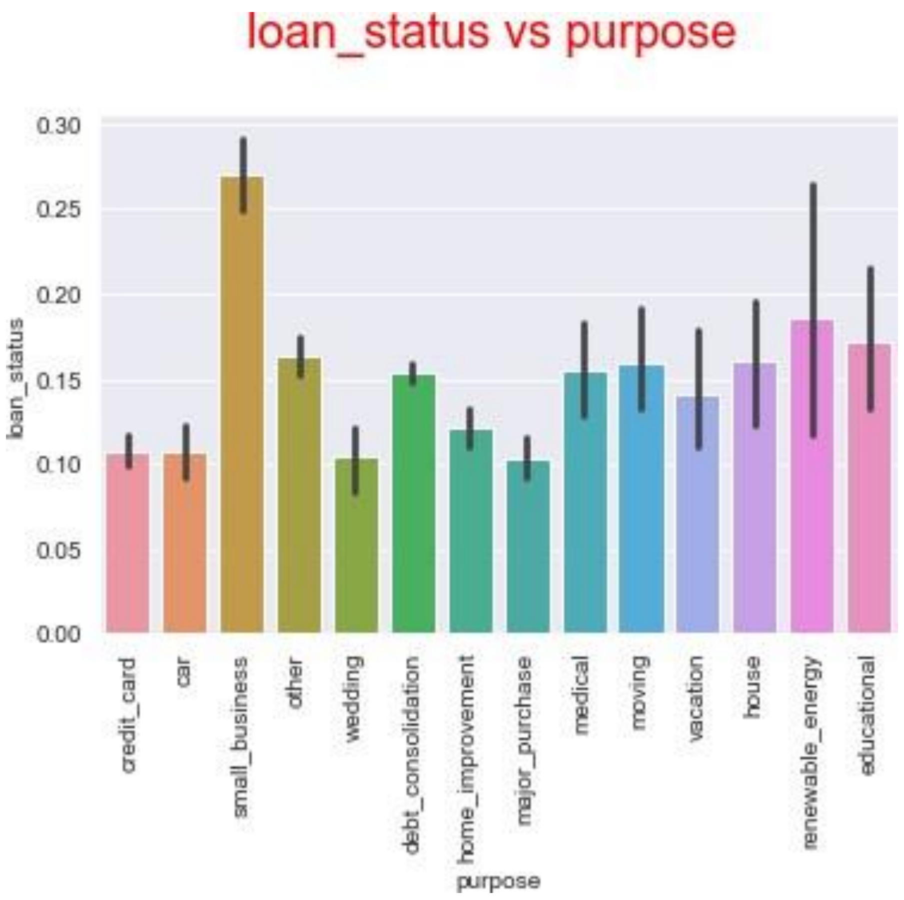
After analysing the dataset with respect to **sub_grades** and **loan_status** we came to the conclusion that the loans belonging to the sub_grades after D are considered to be a high-risk cases in increasing order.

Analysis (Loan_status vs Loan_amnt)



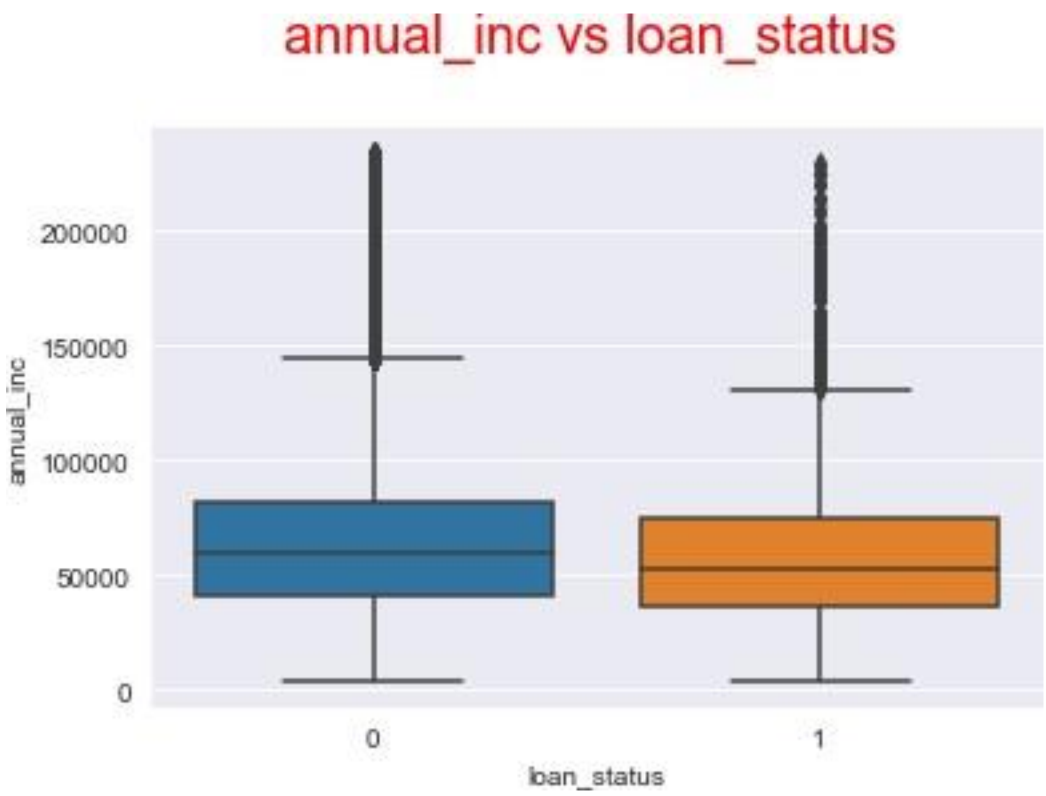
After analysing the dataset with respect to **loan_amnt** and **loan_status** we came to the conclusion that these two columns do not have significant correlation and they do not affect each other.

Analysis (Loan_status vs Purpose)



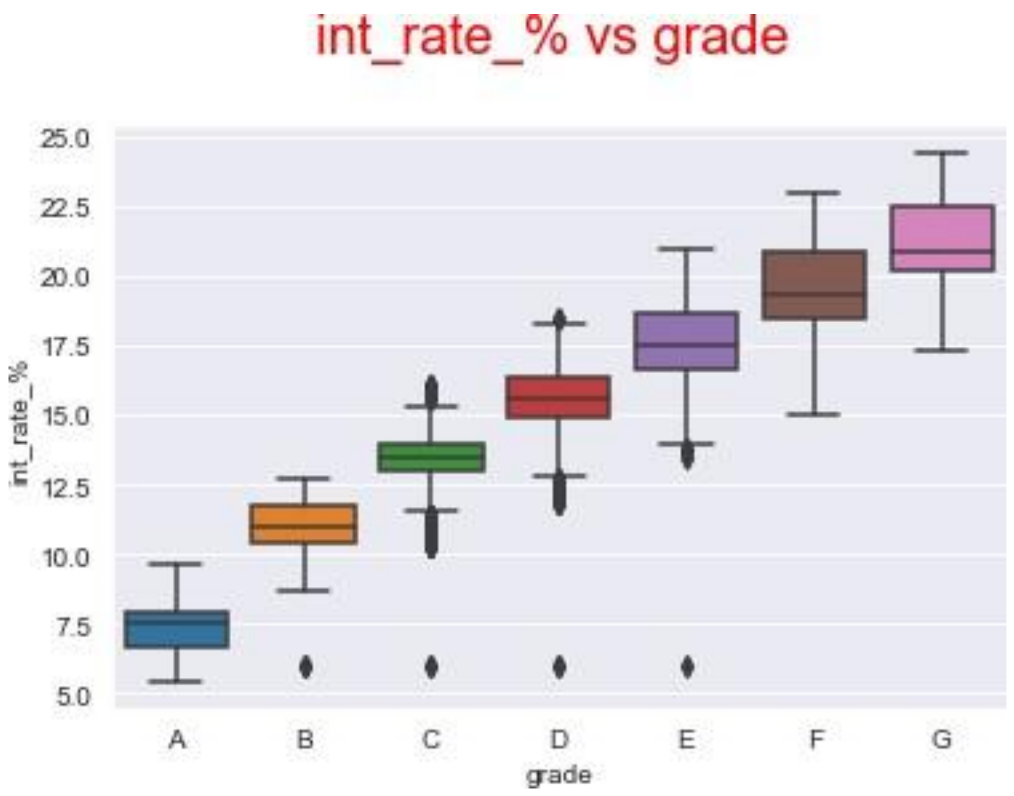
After analysing the dataset with respect to **loan_status** and **purpose** we came to the conclusion that small_business, renewable_energy and educational are having more chances to become a defaulter.

Analysis (Annual_inc vs Loan_status)



After analysing the dataset with respect to **annual_inc** and **loan_status** we came to the conclusion that these two columns do not have any significant correlation. The annual income of the borrower doesn't have any impact on their repaying capacity.

Analysis (int_rate_% vs Grade)

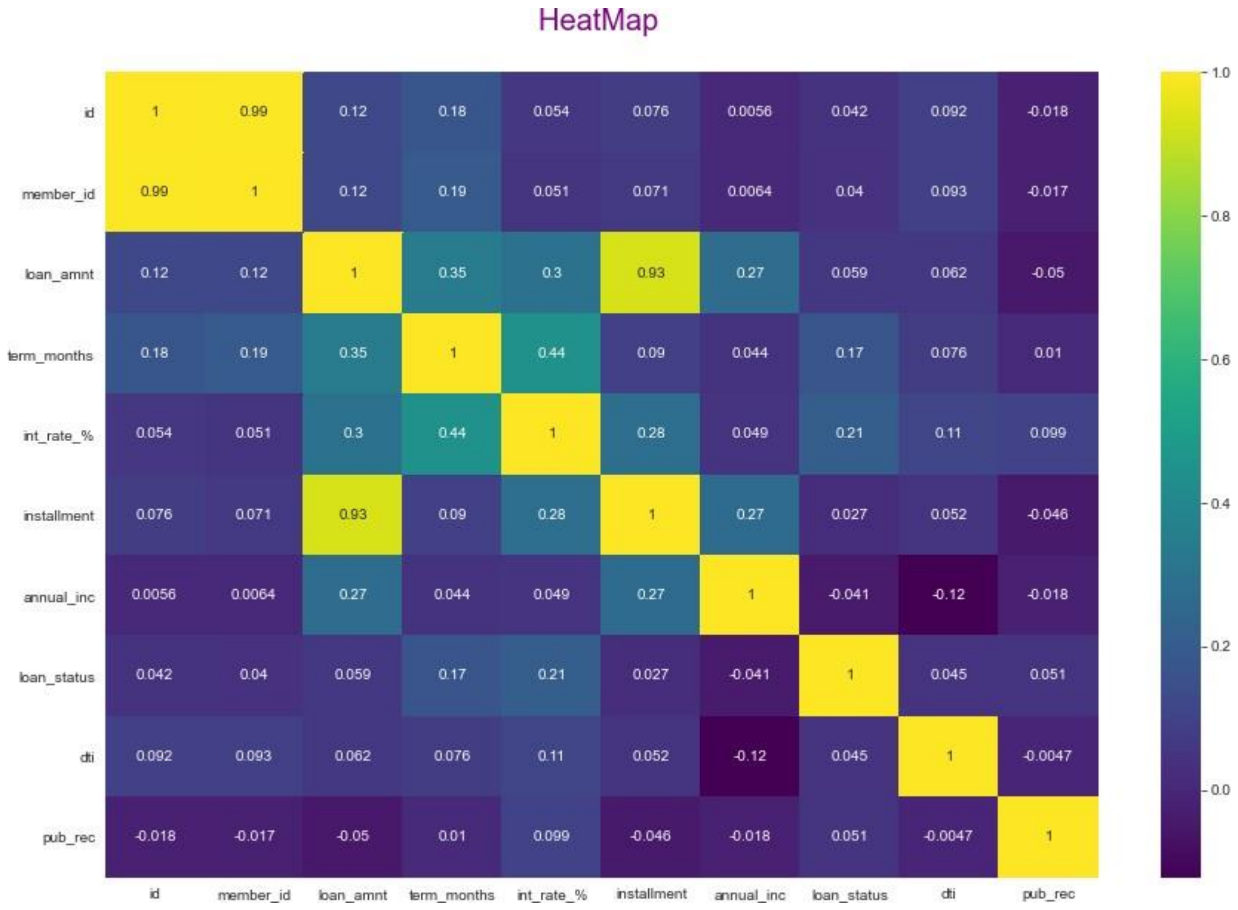


After analysing the dataset with respect to **int_rate_%** and **Grade** we came to the conclusion that both the columns are linearly correlated and as the grades increase the interest rate percentage is increasing.

Analysis (Term_months vs Loan_status)

After analysing the dataset with respect to **term_months** and **loan_status** we came to the conclusion that people opting for 60 months of tenure are more likely to default as compared to 36 months.

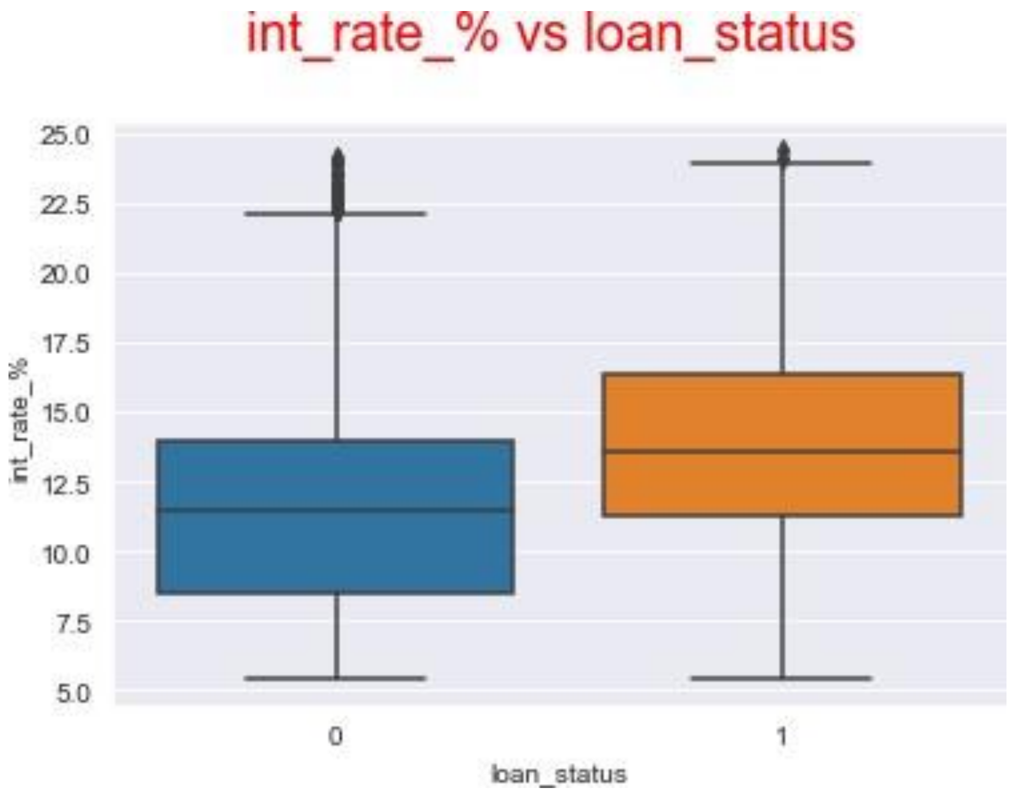
Correlation Matrix



This Heat Map represents all the correlation value for the different columns. It shows which two numeric variables are highly linearly correlated.

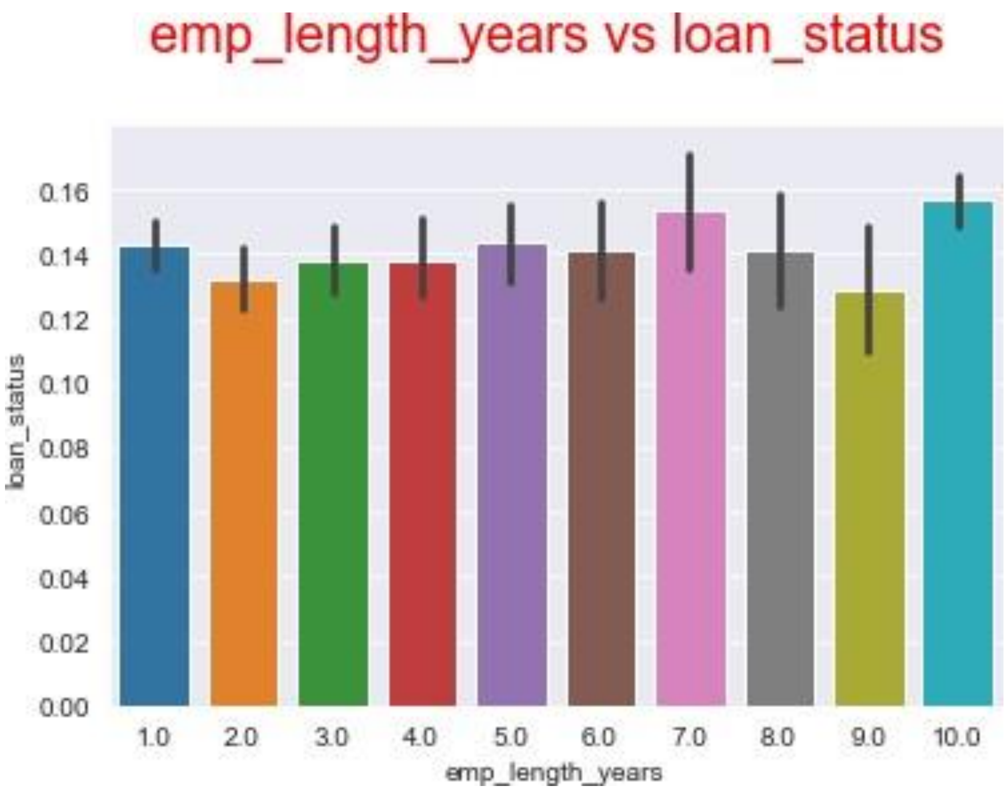
In the correlation matrix, the only non-trivial linear correlation that we can find is between installment and loan_amnt, but they both don't have any correlation with the target variable loan_status. If we check the row(or column) corresponding to loan_status, we don't observe any 'strong' positive or negative linear correlation between loan_status and most other numerical variable.

Analysis (int_rate_% vs Loan_status)



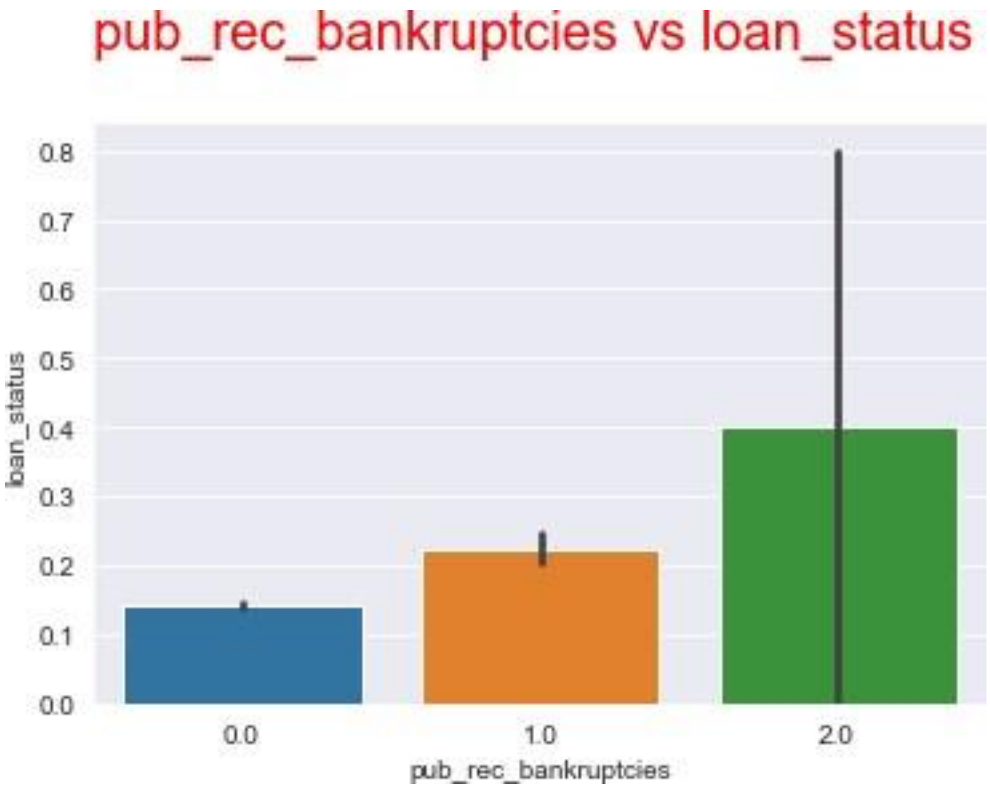
After analysing the dataset with respect to **int_rate_%** and **loan_status** we came to the conclusion that as the interest rate is increasing the number of defaulters are increasing.

Analysis (emp_length_years vs Loan_status)



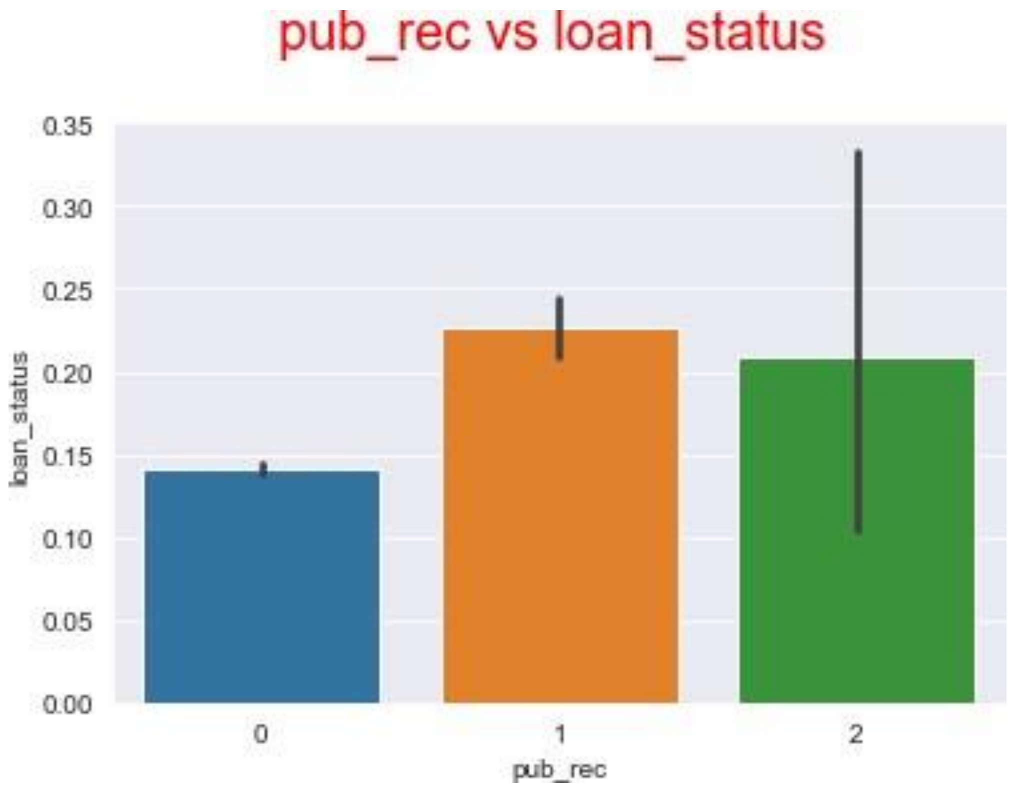
After analysing the dataset with respect to **emp_length_years** and **loan_status** we came to the conclusion that the employment length is not correlated with loan status which means that there is no impact of employment length on the loan status.

Analysis (pub_rec_bankruptcies vs Loan_status)



After analysing the dataset with respect to **pub_rec_bankruptcies** and **loan_status** we came to the conclusion that the people having two public record bankruptcies are highly likely to default.

Analysis (pub_rec vs Loan_status)



After analysing the dataset with respect to **pub_rec** and **loan_status** we came to the conclusion that people having any derogatory public record are more likely to default.

Conclusion

The following are the variables that has considerable impact on the repayment of the loan:

- grade → particularly starting from D has higher default rate
- sub_grade → particularly starting from D2 has higher default rate
- int_rate → higher the interest rate, higher default rate
- term → people opting for 60 months has higher default rate
- purpose → particularly small_business has higher default rate
- pub_rec → having derogatory public record significantly increases the chances of defaulting
- pub_rec_bankruptcies → having public record bankruptcies significantly increases the chances of defaulting