LENDING CLUB LOAN DATA

An exploratory data analysis on the data

Created by: Saptarshi Ghosh

CLEANING THE DATA

- The excel file came with 111 columns out of which many were null columns.
- A closer look and a few steps of data cleaning led to identifying a set variables which could explain the data in a manner worthy of analysis.
- The variables identified had been categorized into:

  - Numerical Variables

  - Categorical Variables.

NUMERICAL VARIABLES

- revol_util-Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
- loan_amnt-The listed amount of the loan applied for by the borrower
- int_rate-Interest Rate on the loan
- annual_inc-The self-reported annual income provided by the borrower during registration.
- total_acc-The total number of credit lines currently in the borrower's credit file

CATEGORICAL VARIABLES

- emp_length-Employment length in years
- pub_rec_bankruptcies-number of public record of bankruptcy
- Term-The number of payments on the loan. Values are in months and can be either 36 or 60.
- Grade-LC assigned loan grade
- home_ownership-The home ownership status provided by the borrower during registration. Our values are: RENT, OWN, MORTGAGE, OTHER.
- verification_status-Indicates if income was verified by LC, not verified, or if the income source was verified
- loan_status-Current status of the loan
- Purpose- Purpose mentioned by borrower while applying for loan
- delinq_2yrs-The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
- inq_last_6mths-number of loan enquiries in last 6 months
- pub_rec-Number of derogatory public records

BUSINESS OBJECTIVE

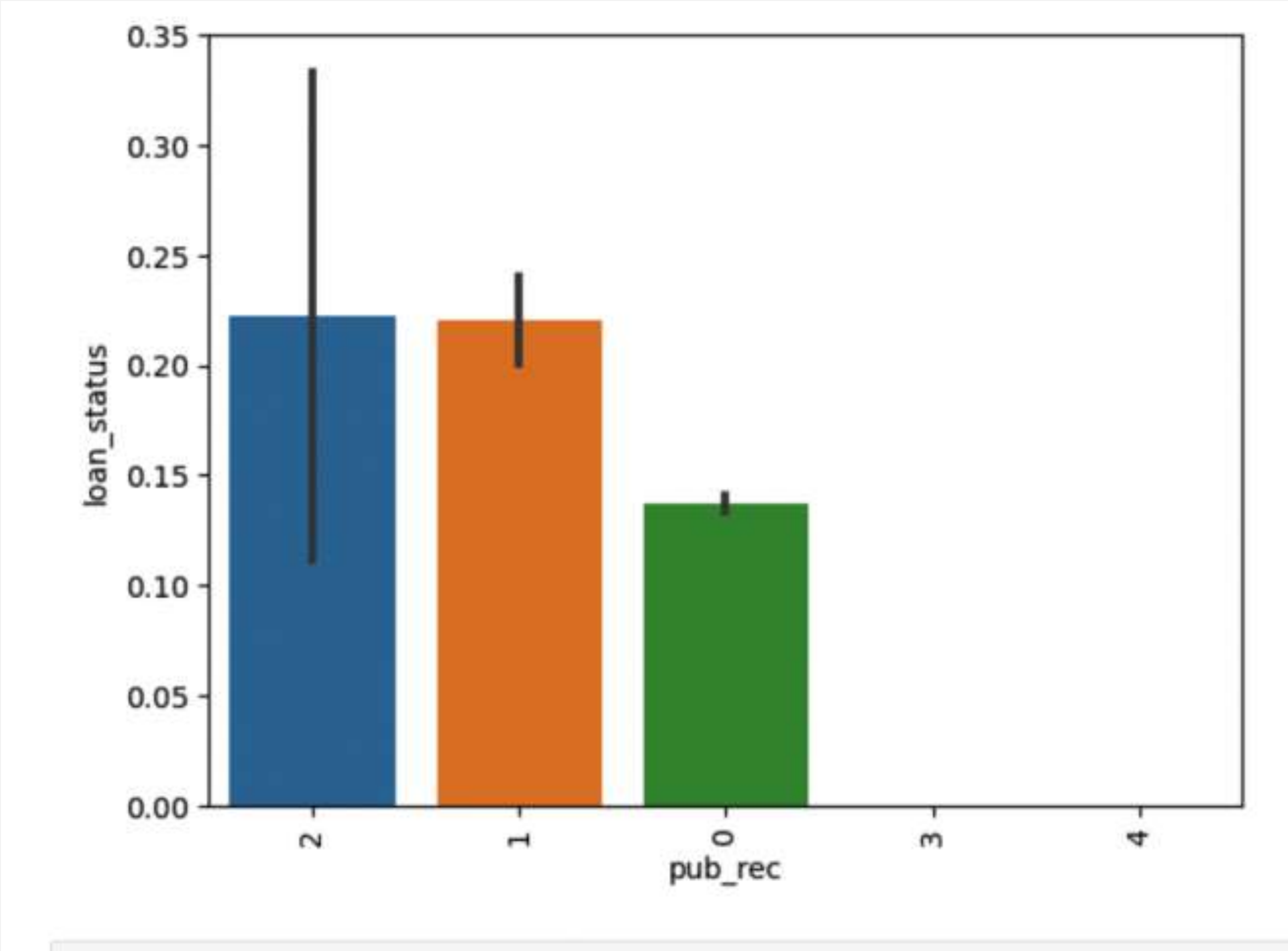
- Understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default
- Borrowers who **default** cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters’
- Reduce financial loss of the organization

PLAN OF ANALYSIS

- Univariate and segmented analysis based on ‘proportion’ to find insights and strong factors for defaulting on a loan
- Bivariate analysis to identify combination of factors affecting loan defaulters and other valuable insights
- Derived metrics based on business requirement to come up with solutions and recommendations

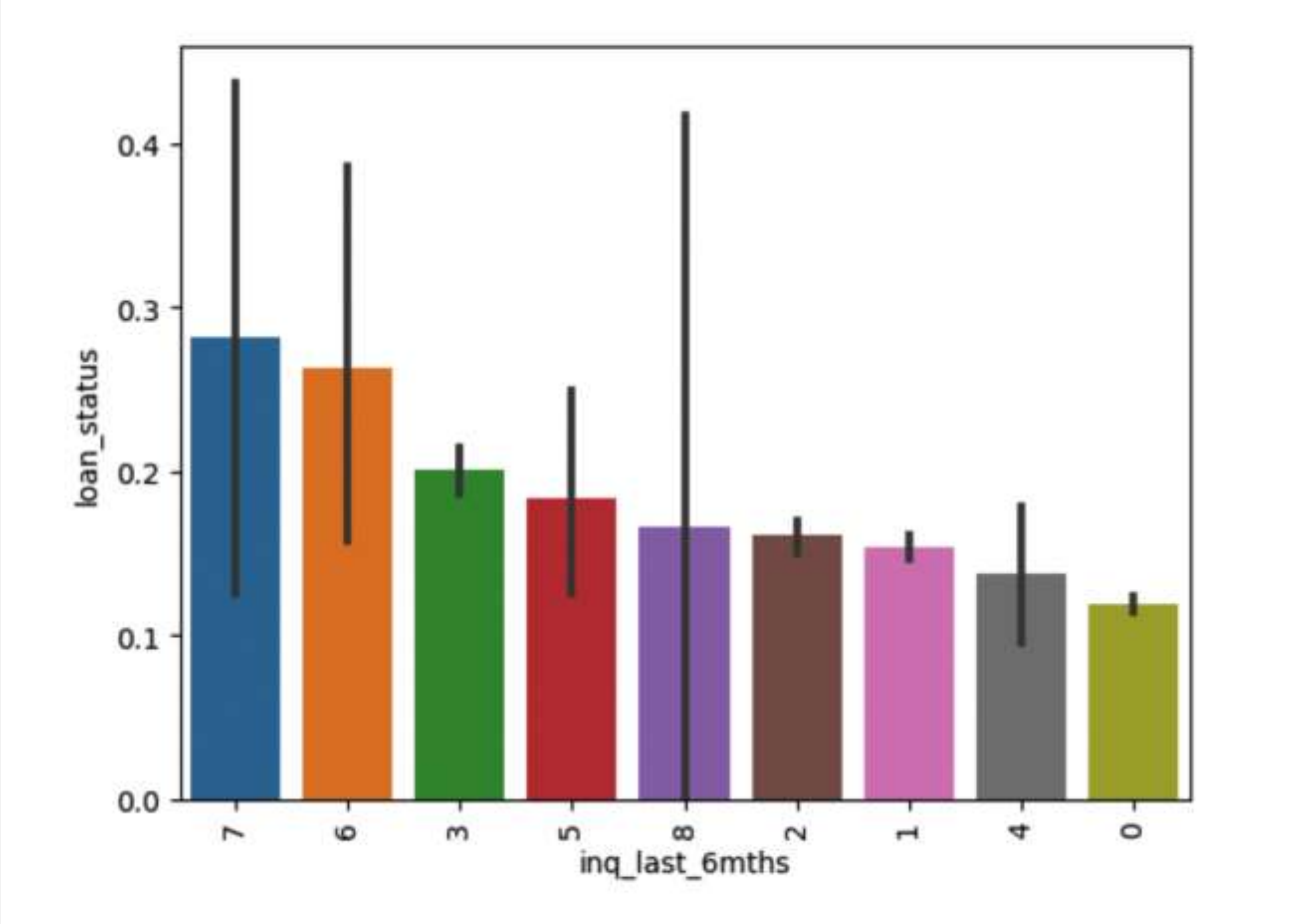
INSIGHTS FROM UNIVARIATE ANALYSIS

- Number of 'public records of bankruptcy' directly affects the chances of defaulting



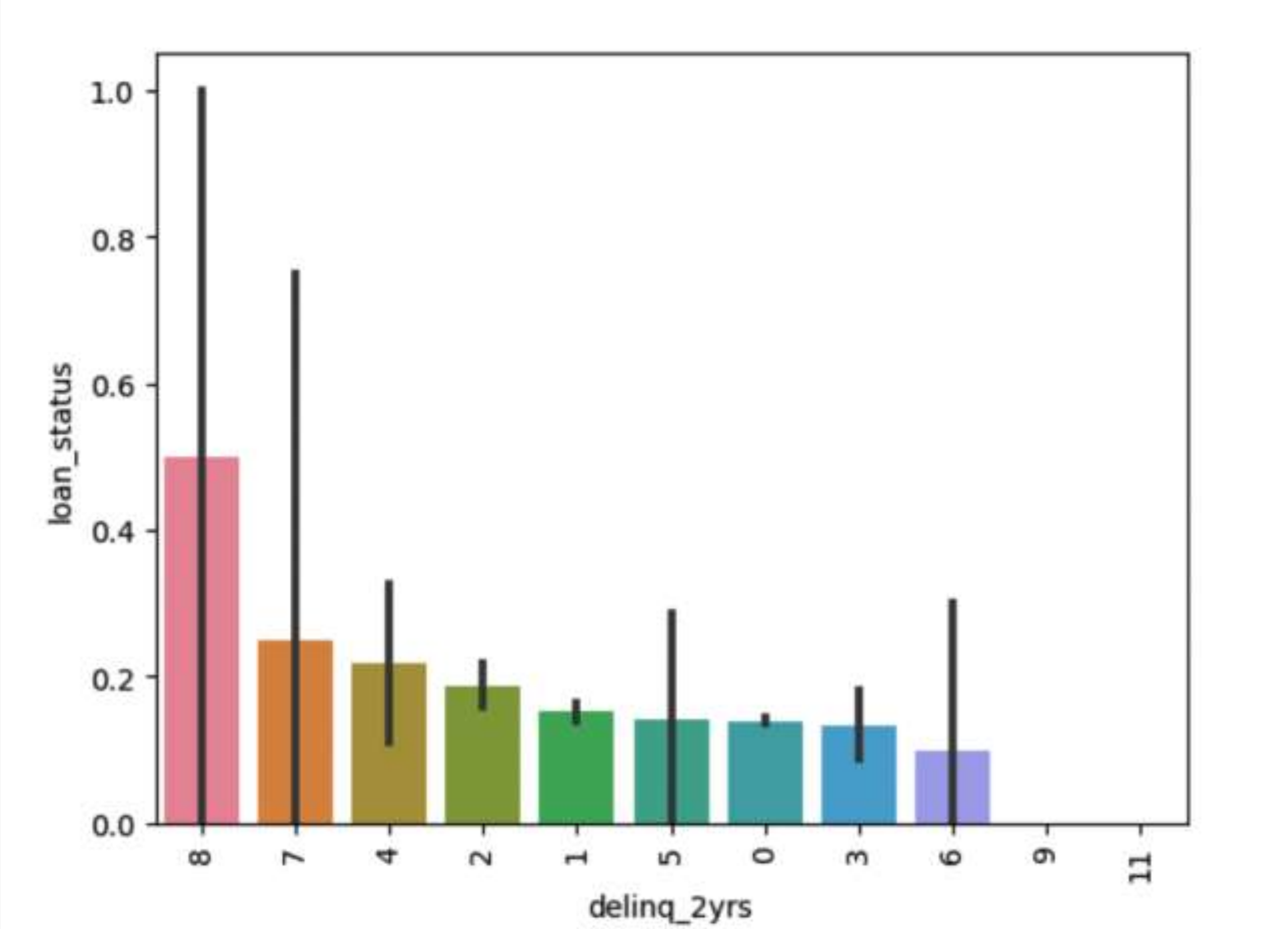
INSIGHTS FROM UNIVARIATE ANALYSIS

- Customers making more loan inquiries in the last 6 months are more prone to being defaulters as compared proportionately with fully paid customers.
- The trend of the graph clearly shows that more number of enquiries can be a red flag.



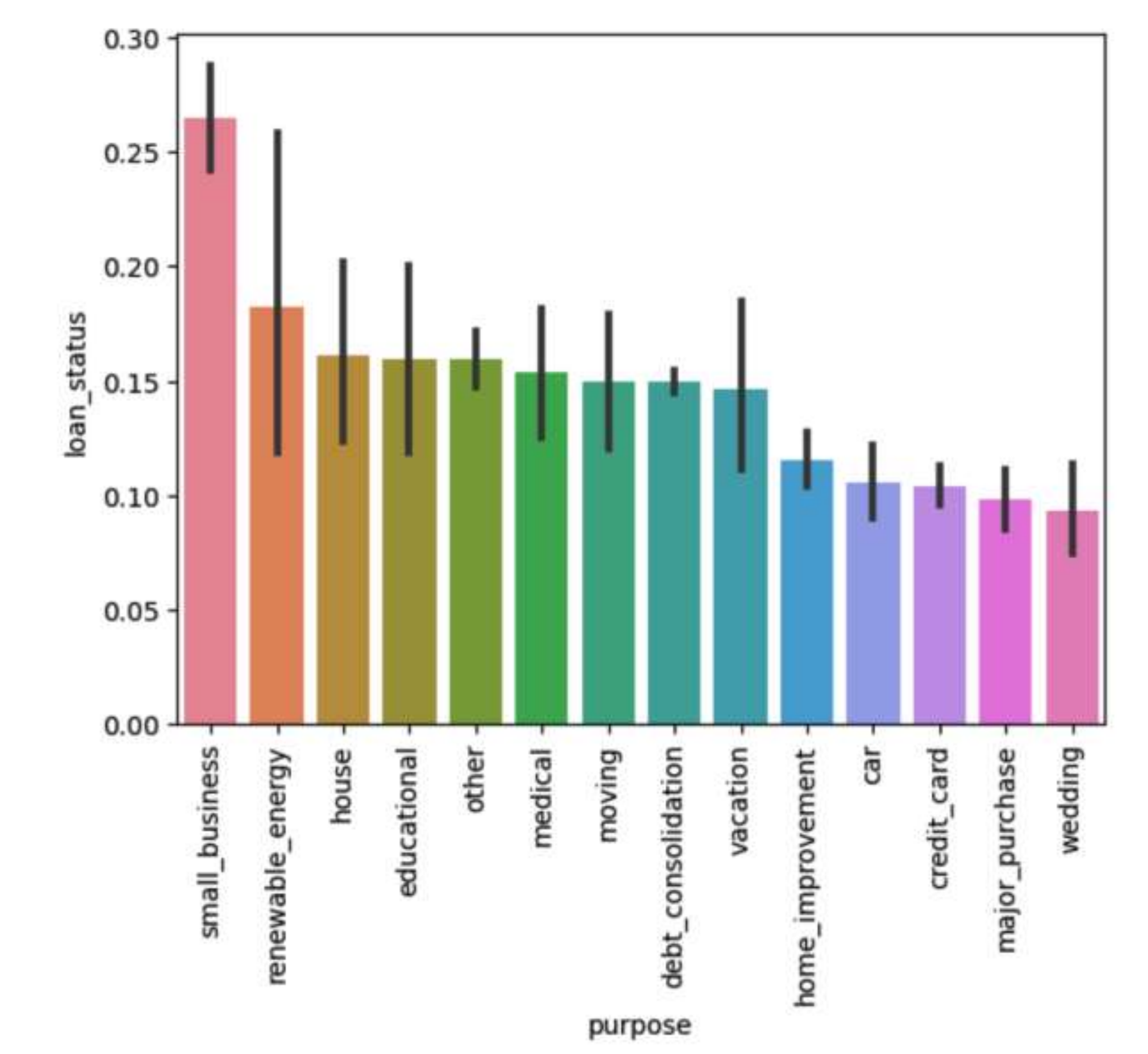
INSIGHTS FROM UNIVARIATE ANALYSIS

- Customers who have record of being delinquent more than 30 days past due have more risk of defaulting
- The trend of the graph clearly shows that more number of delinquency record an be a red flag.



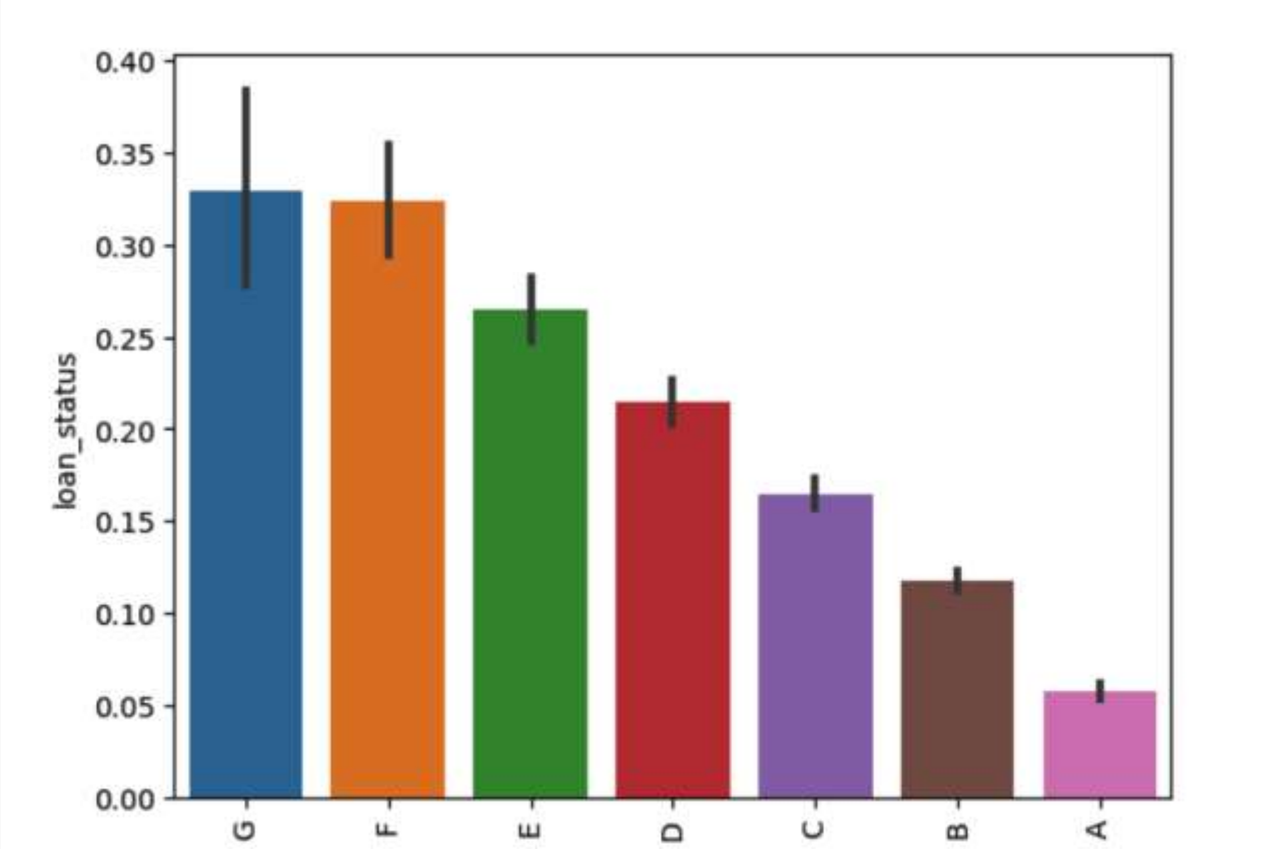
INSIGHTS FROM UNIVARIATE ANALYSIS

- Loans availed for the purpose of 'Small Business' have significantly high risk of defaulting.
- However loans taken for personal use like-credit card, major purchase , wedding etc are less likely to be defaulted.



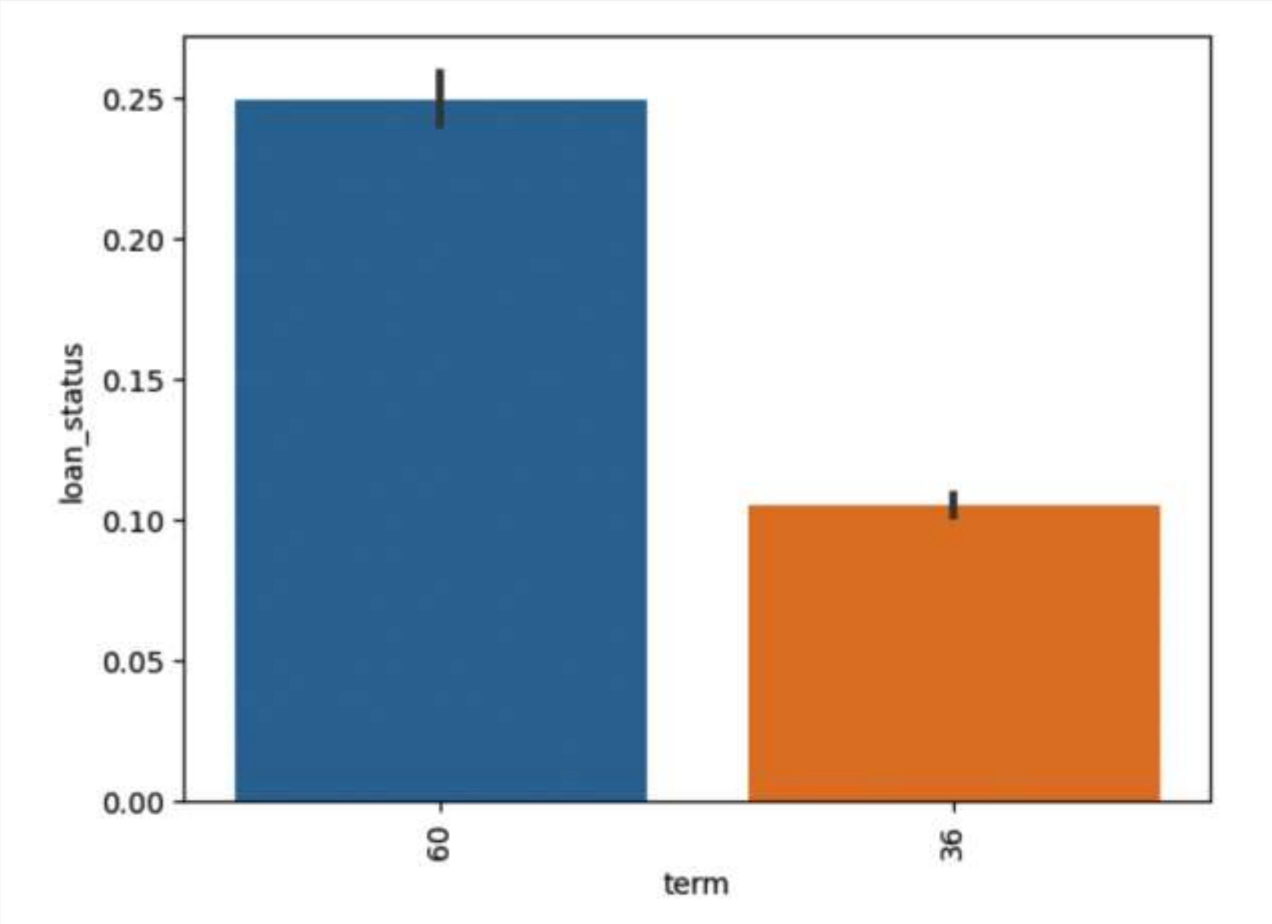
INSIGHTS FROM UNIVARIATE ANALYSIS

- Higher the 'loan grade' higher is the chances of defaulting
- The graph clearly shows that loan grade G is highly likely to be defaulted.



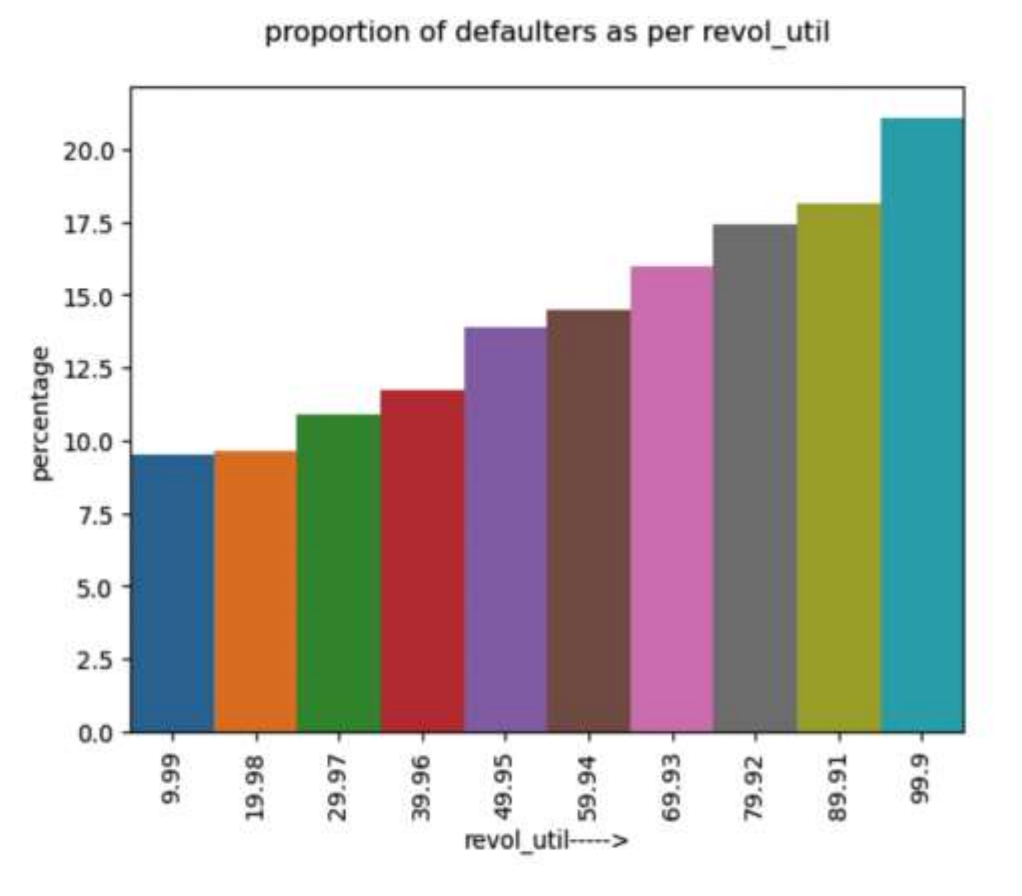
INSIGHTS FROM UNIVARIATE ANALYSIS

- Long term loans tends to get defaulted more than the short term loans



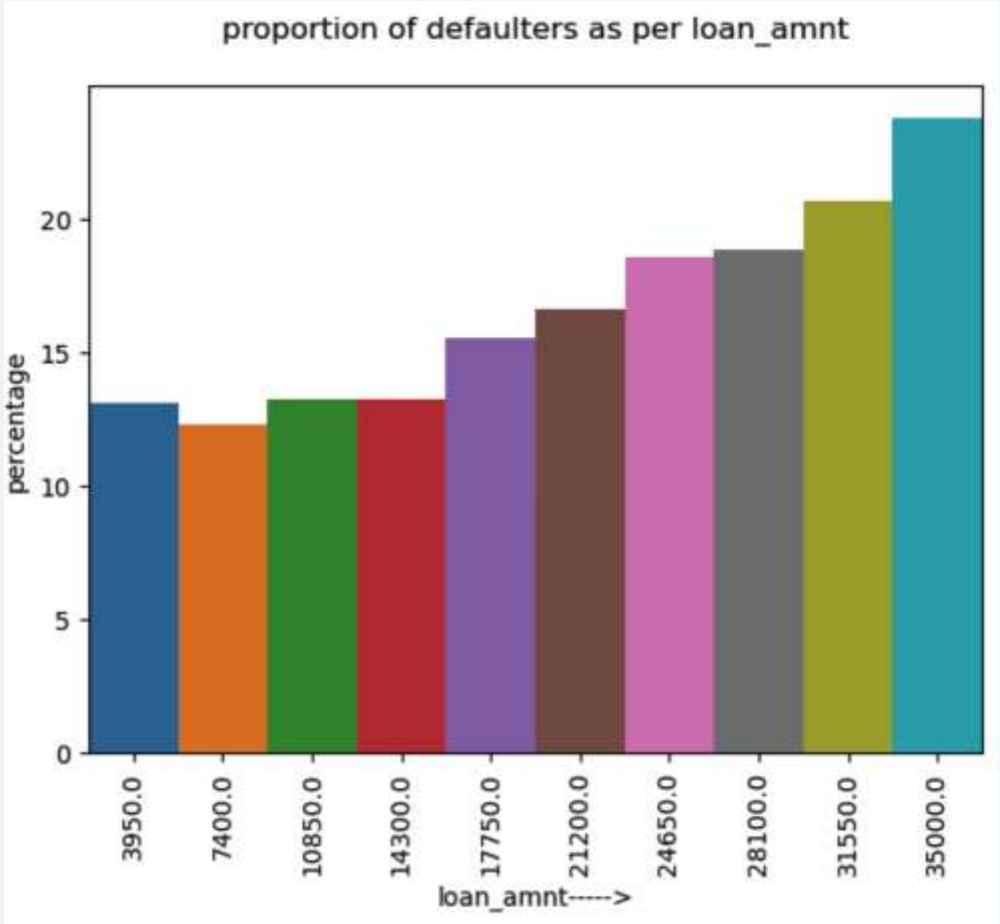
INSIGHTS FROM UNIVARIATE ANALYSIS

- Customers having higher 'credit utilization rate are more prone to defaulting



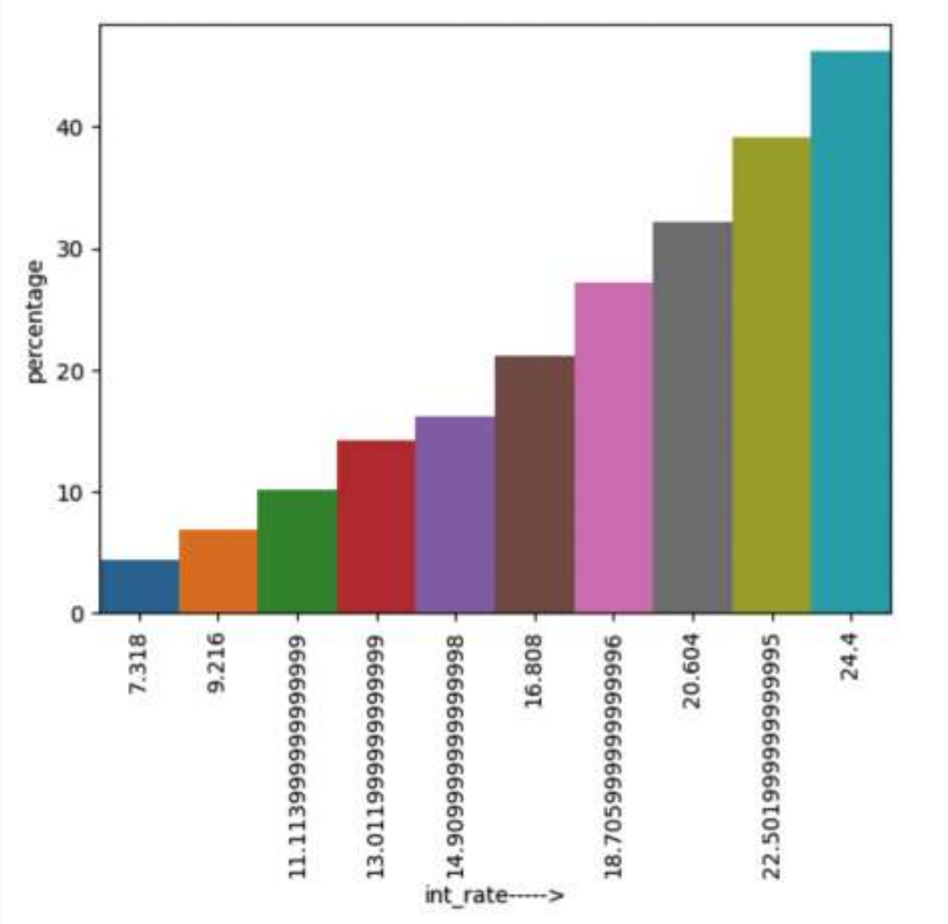
INSIGHTS FROM UNIVARIATE ANALYSIS

- Customers having higher loan amount are more prone to be defaulters.



INSIGHTS FROM UNIVARIATE ANALYSIS

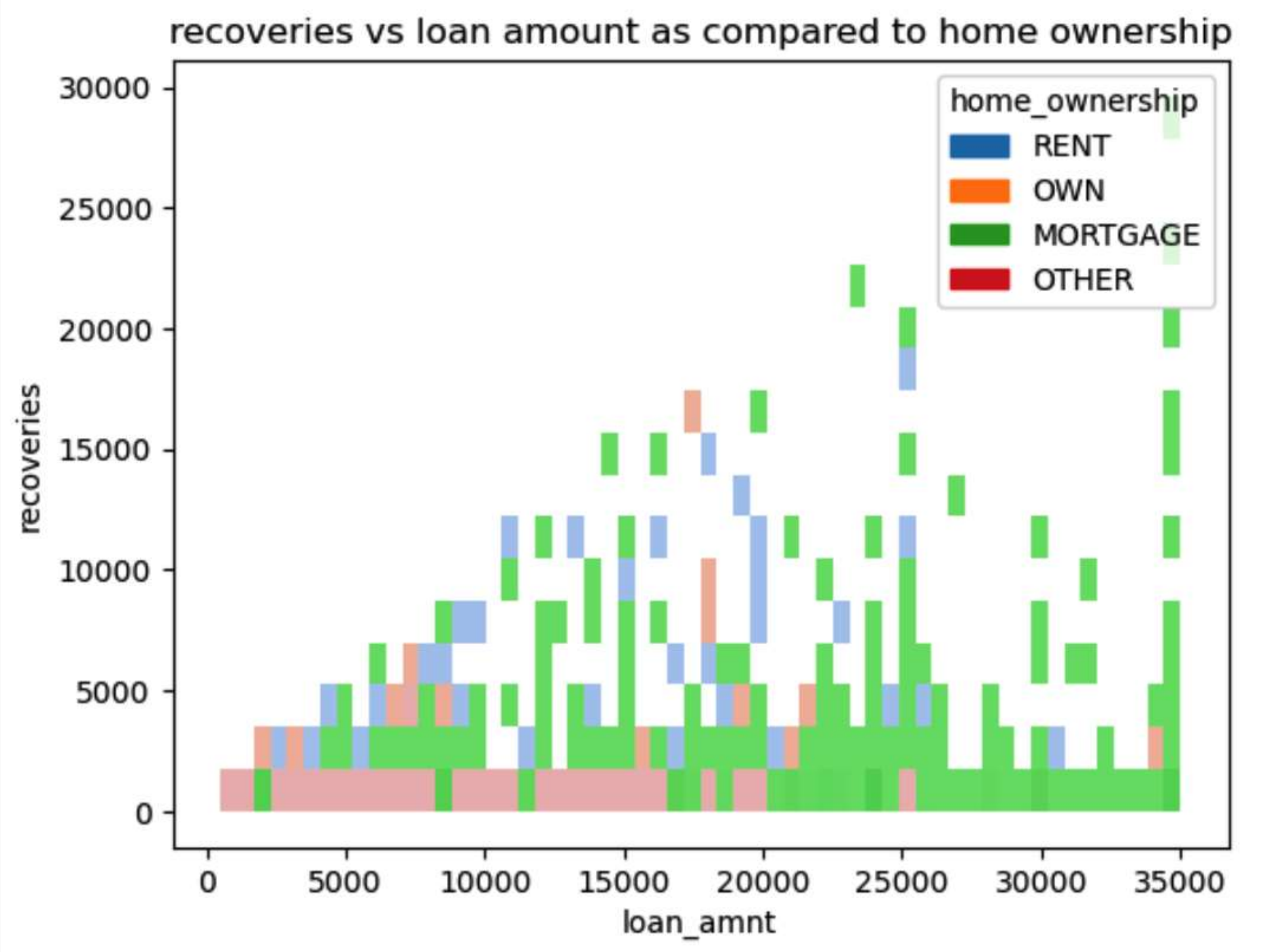
- Clearly there is a strong indication of defaulters when the interest rate is high.



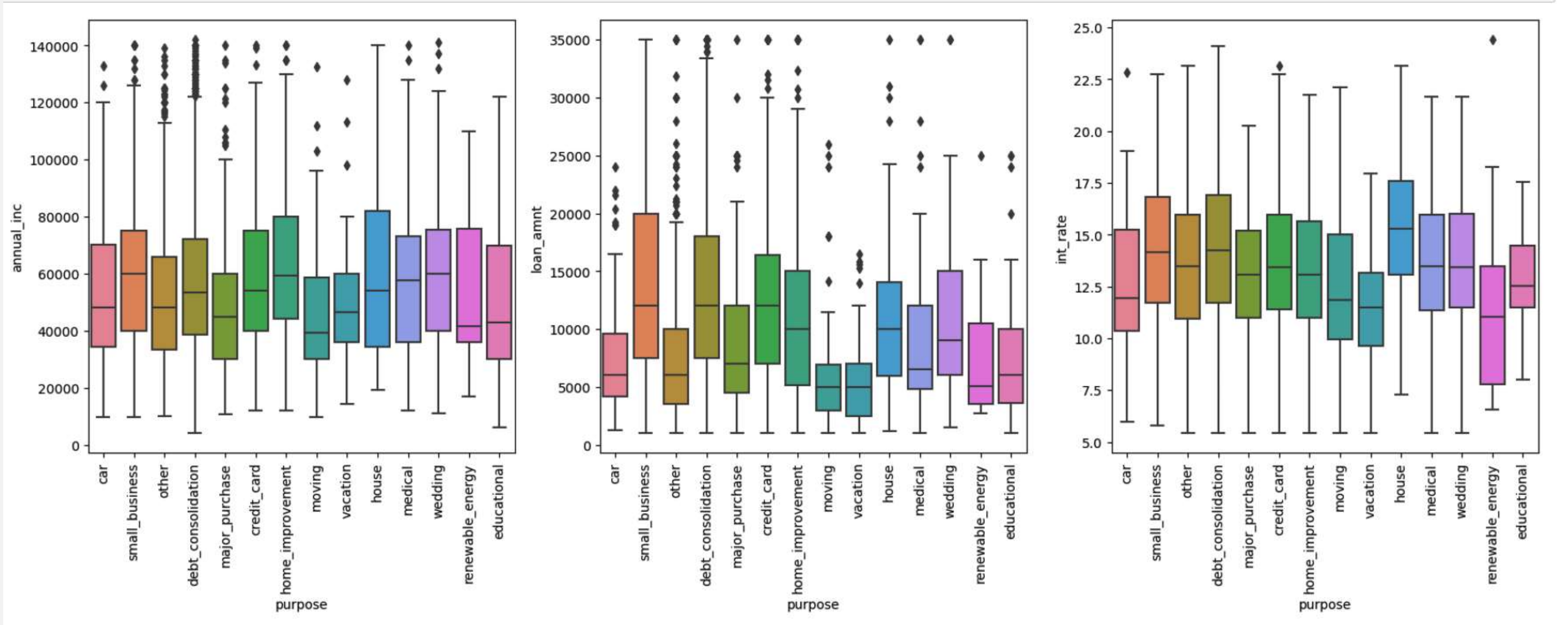
BIVARIATE ANALYSIS

INSIGHTS FROM BIVARIATE ANALYSIS

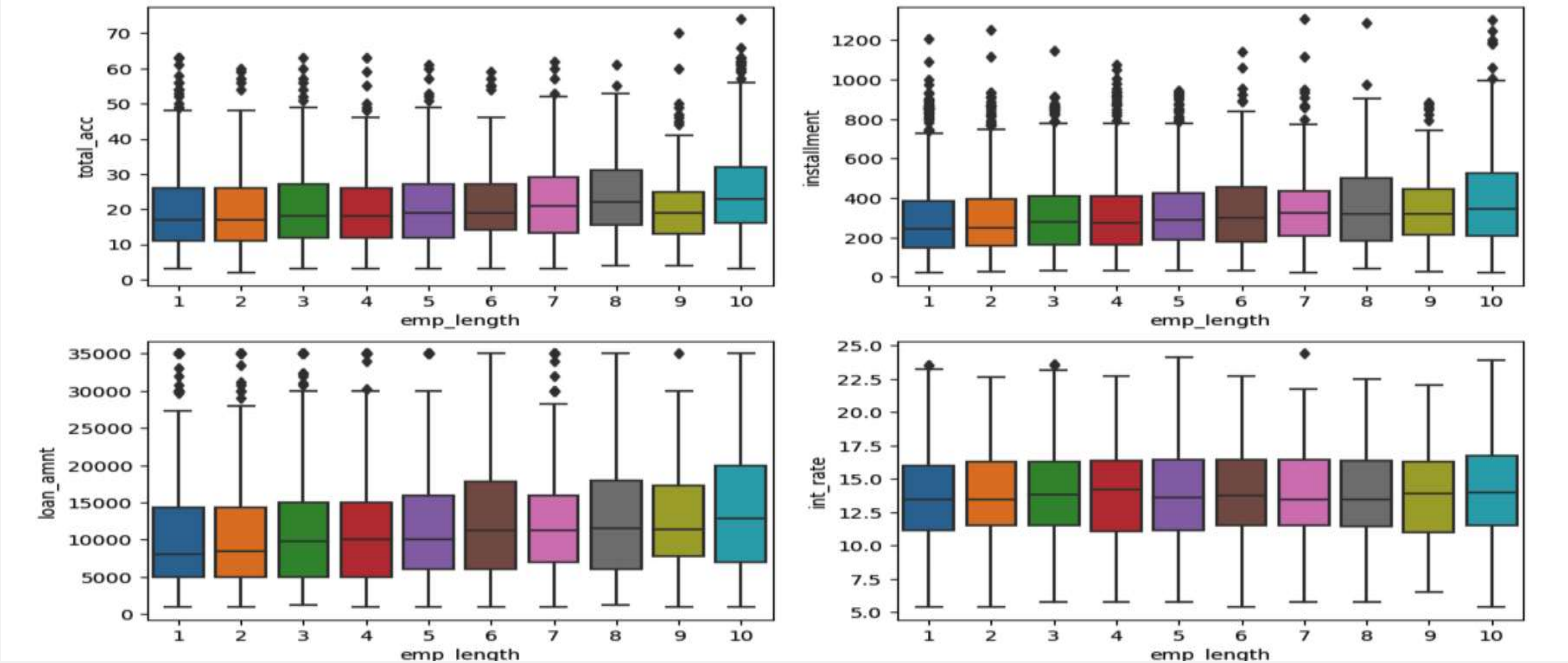
- For customers who have defaulted on the loan it is clearly seen that recovery is more when they have given their property on mortgage.
- It is worthy to note that while giving loans to risky customers it is advisable to take their property as mortgage.



IT IS CLEAR THAT LOANS TAKEN FOR SMALL BUSINESSES DEFAULT THE MOST. HOWEVER THEIR HIGH LOAN AMOUNT , INTEREST RATE AND HIGH INCOME ENSURE MORE BUSINESS PROFIT FOR BANKS. IT IS ADVISED TO TAKE COLLATERAL FROM THEM LIKE MORTGAGE TO ENHANCE RECOVERY AND PROFIT.

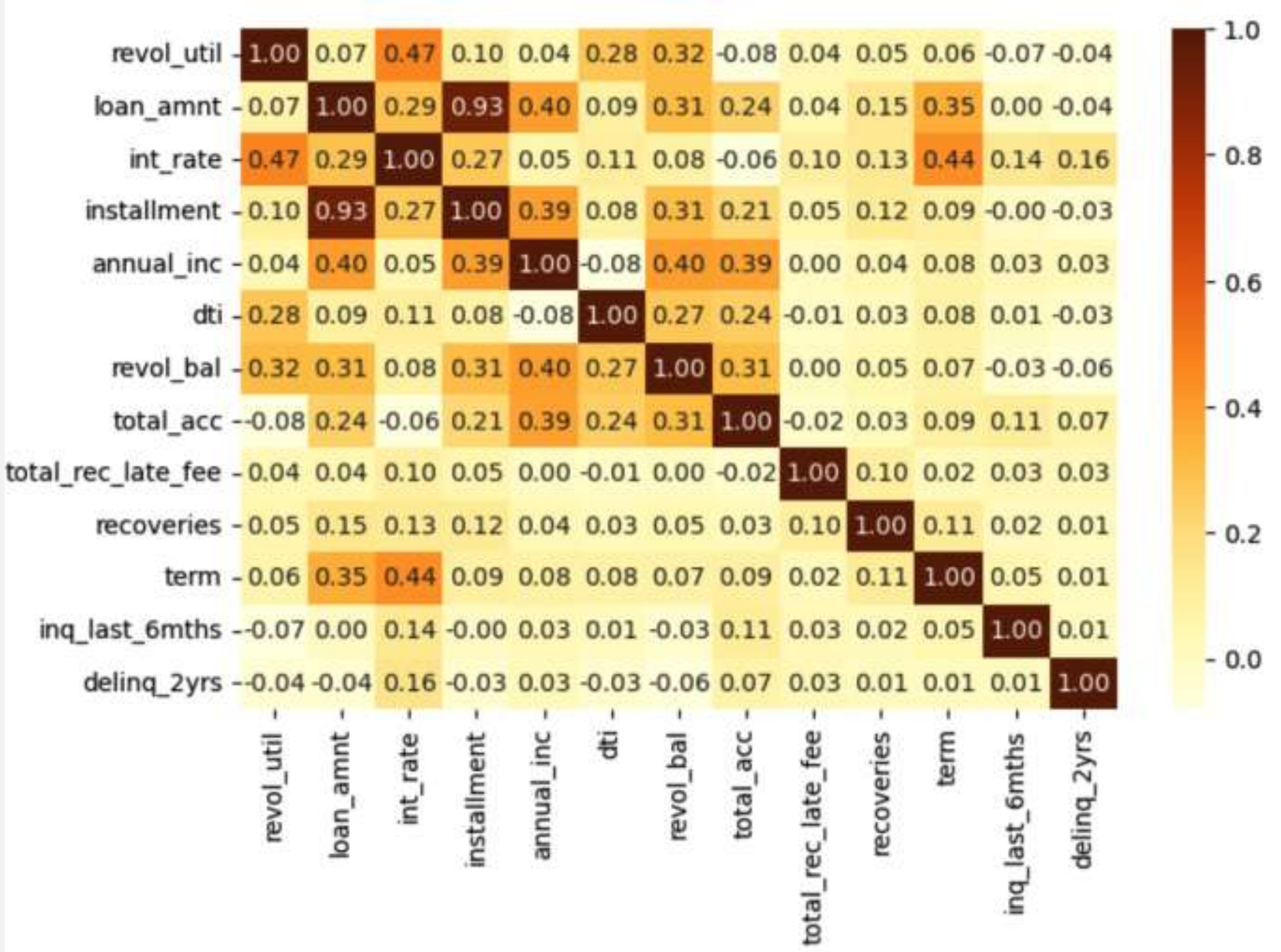


CUSTOMERS WHO HAVE HIGHER EMPLOYMENT LENGTH TEND TO TAKE HIGHER LOAN AMOUNT AT HIGHER INTEREST RATE. ALSO THEY HAVE MORE NUMBER OF OPEN CREDIT ACCOUNTS AND INSTALMENT AS A RESULT THEY ARE OFTEN BURDENED AND TEND TO DEFAULT MORE. TO HAVE A GOOD PROFIT BY GIVING LOAN TO THIS CUSTOMERS , ONE OPTION IS TO KEEP THEIR PROPERTIES AS MORTGAGE TO EVEN OUT ON RECOVERIES.



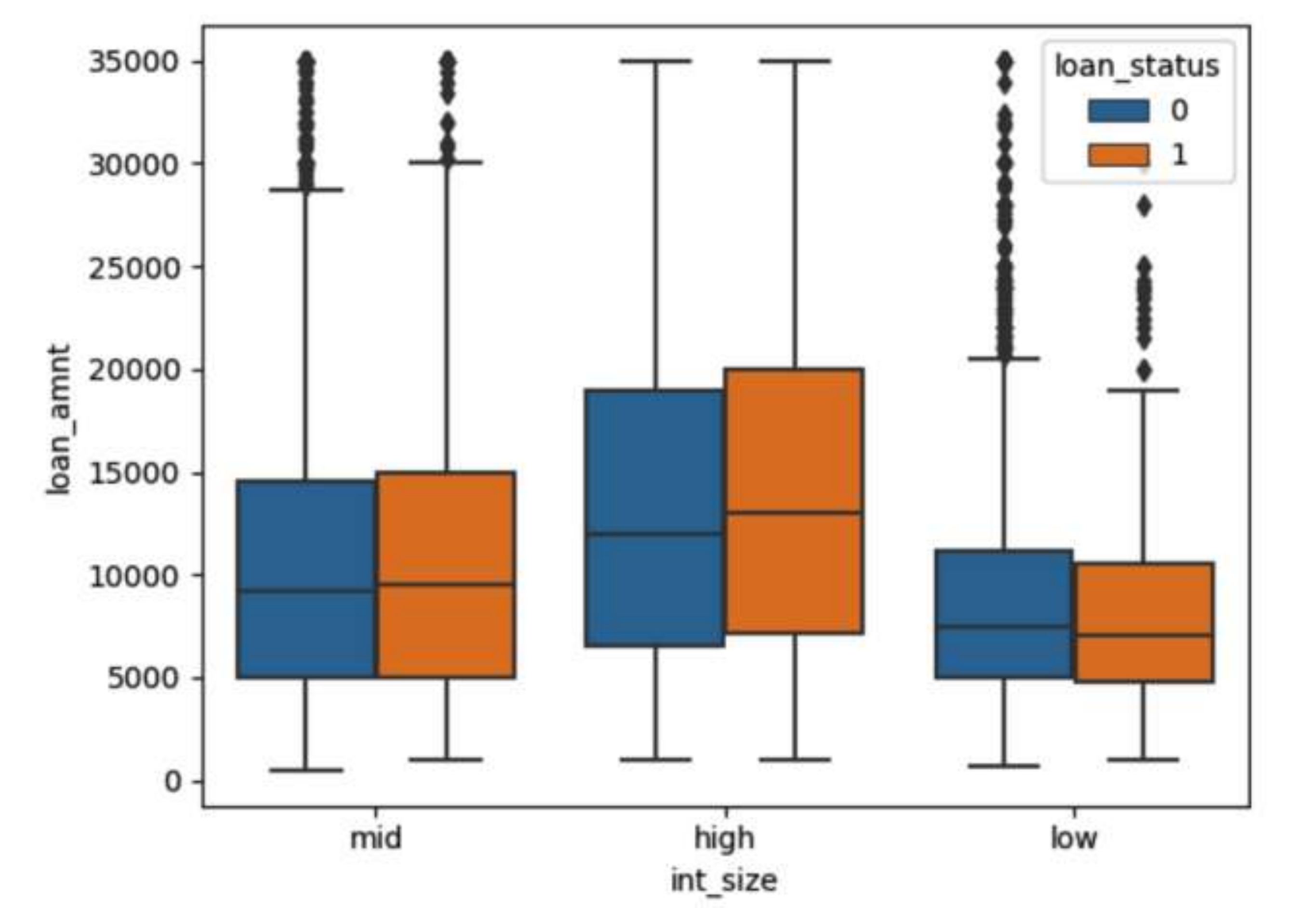
INSIGHTS FROM BIVARIATE ANALYSIS

- The heat map does show that we have a clear positive correlation between a few variables like:
- .interest rate and loan amount
  .interest rate and loan term
  .interest rate and credit utilization(revol_util)



INSIGHTS FROM BIVARIATE ANALYSIS

- Clearly from the graph we can see that higher loan amount with higher interest rate is a very strong indicator of defaulters. We can calculate the risk factor of a customer by taking the product of loan amount and interest rate

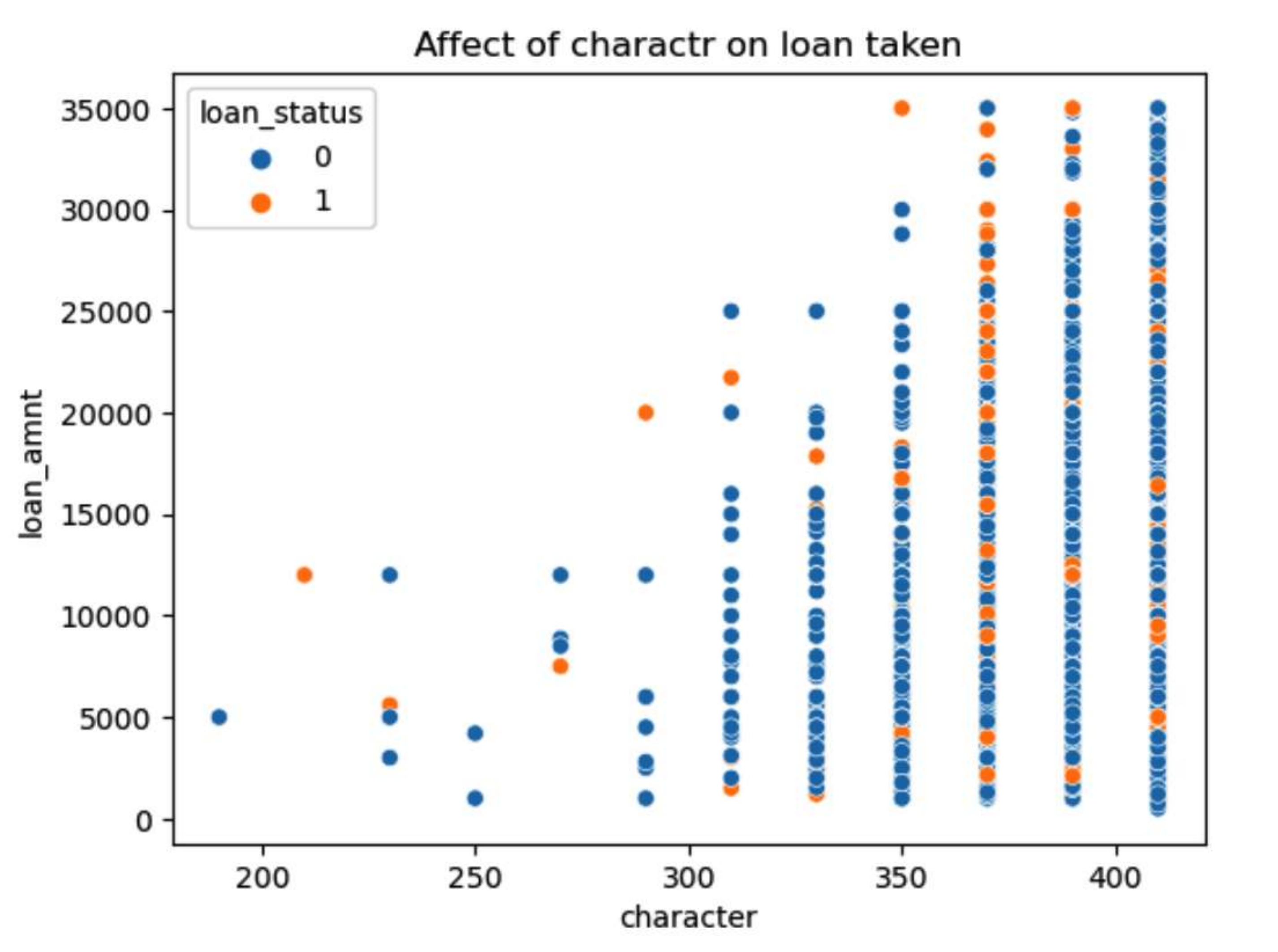


BUSINESS DRIVEN METRICS

- I have created 3 new business driven metrics by the name:
- - Character: it takes into account the character aspect of customer like bankruptcy record , delinquent record , enquiries made and assigns a score to them.
- - Risk Factor- it takes into account the loan amount asked by the borrower and interest rate given to customer. And assigns a risk to them
- - Capacity – it takes into account the

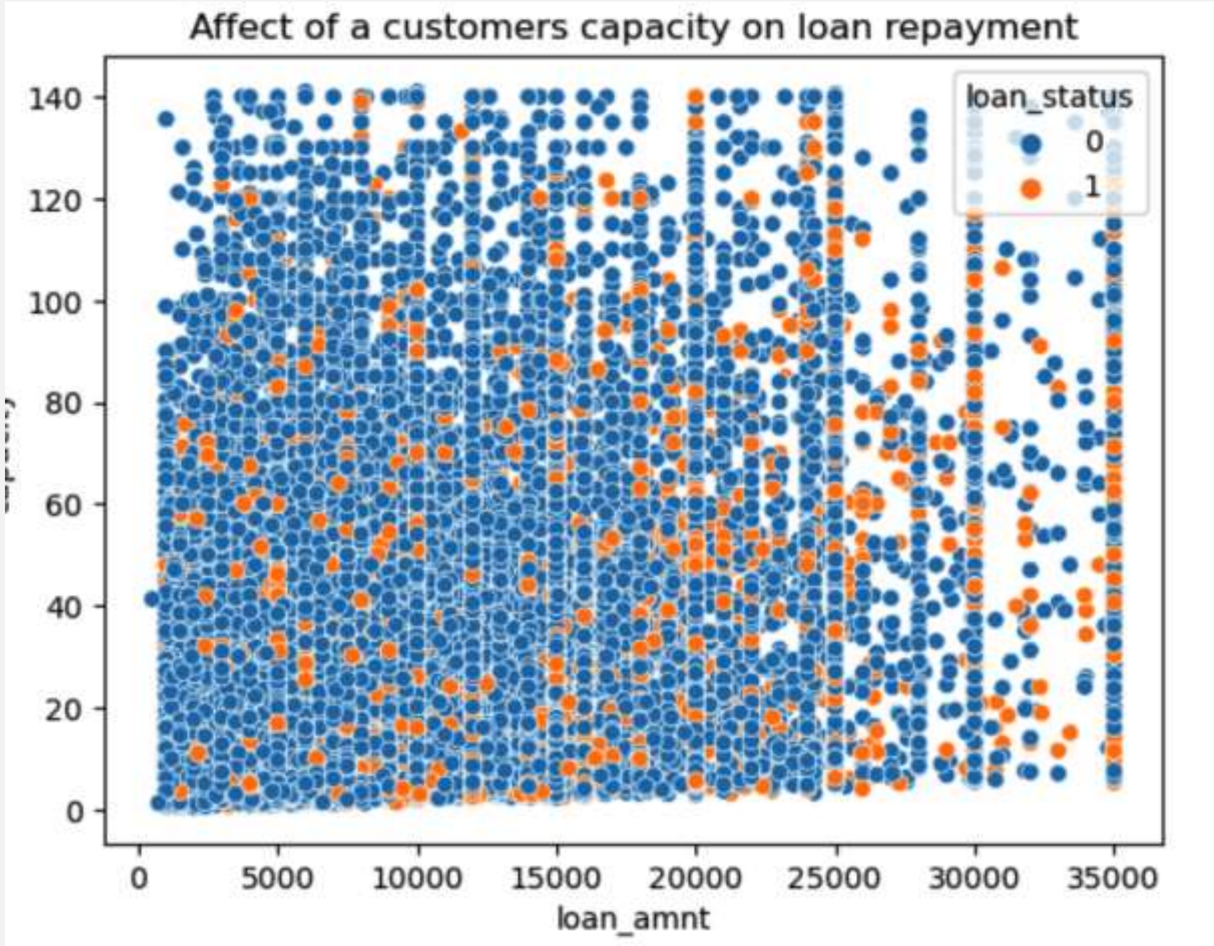
INSIGHTS FROM DERIVED METRICS

- The graph above clearly shows that most defaulters lie below the 375 score and the density of such defaulters increase with increase of loan amount



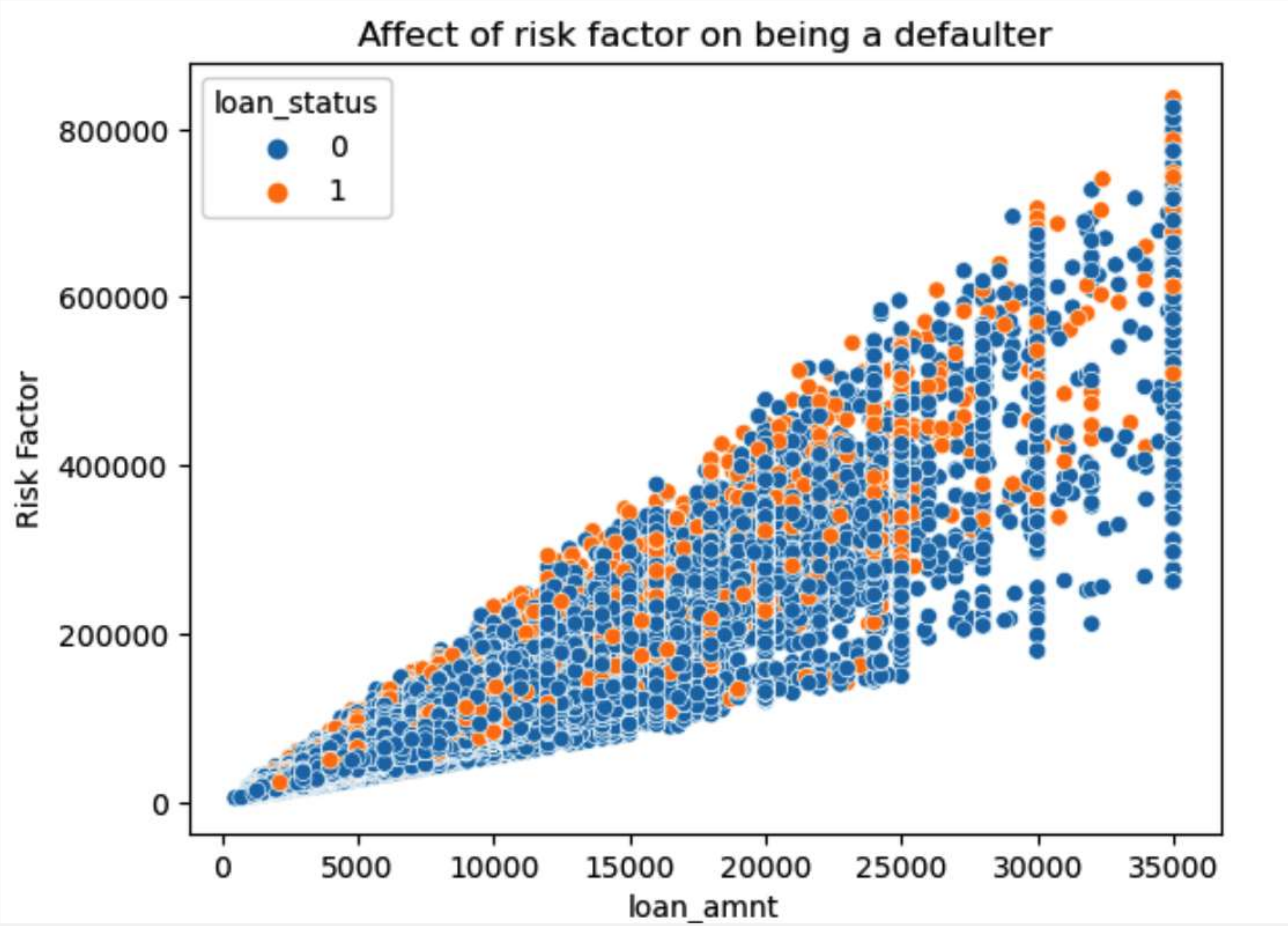
INSIGHTS FROM DERIVED METRICS

- It is evident from the graph that customers who have taken high loan amounts and have capacity below 100 are major defaulters.



INSIGHTS FROM DERIVED METRICS

- The graph clearly shows that customers who have taken higher loan amount at higher risk are more prone to being defaulters



RECOMMENDATIONS

- As per the business driven metrics framed , it is evident that customers who have a character score above 375,capacity score below 100 are low risk customers and low risk factor are low risk customers.
- Customers who have high risk factor and higher loan amount can be given loan only at some collateral like keeping their property at mortgage, so that the recovery compensates for their default
- A shorter loan term leads to higher chances of being fully paid. So by shortening the loan tenure and offering lesser interest rate can help reduce the loan defaults.
- .Small businesses have customers of high risk factor so they must provide collateral for availing loan.