

Univariate analysis

Univariate analysis is deals with only one variable, it does not deal with casual relationships and it mainly used for summarize data, visualize and find interesting pattern in the data.

Continuous Variables

Continuous variables help to understand the central tendency of the data and how the variable is spread. By using various statistical metrics and visualization methods such as:

- Box plot ,
- Distribution Plot
- Violin Plot

A boxplot is used to graphically demonstrate the spread and skewness groups of numerical data using quartiles. Distribution plot is used for comparing the range and distribution for groups of numerical data and Violin plots are used to show the probability density of the data at different values.

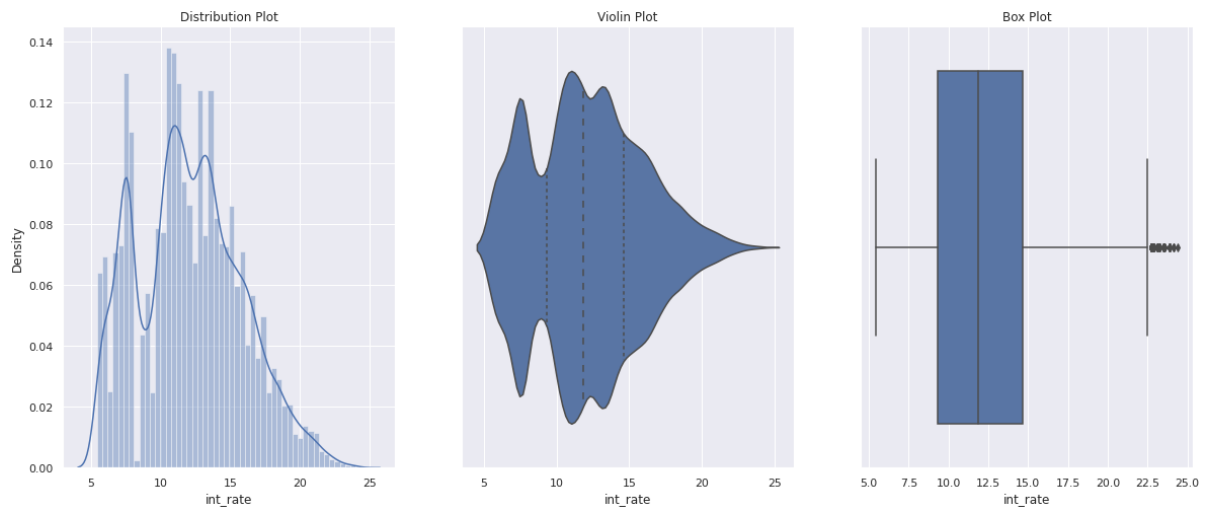
Categorical Variables

For categorical variables, the frequency distribution is used for each category. The visualization methods used for categorical variables are:

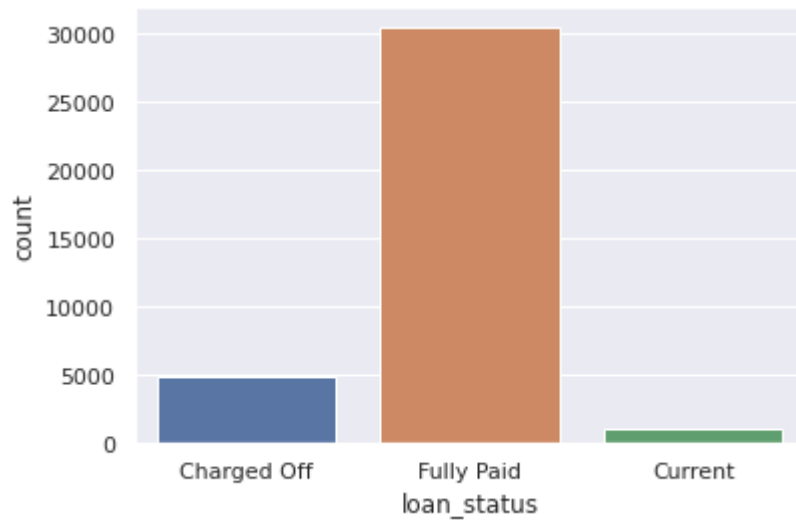
- Bar chart
- Count plot

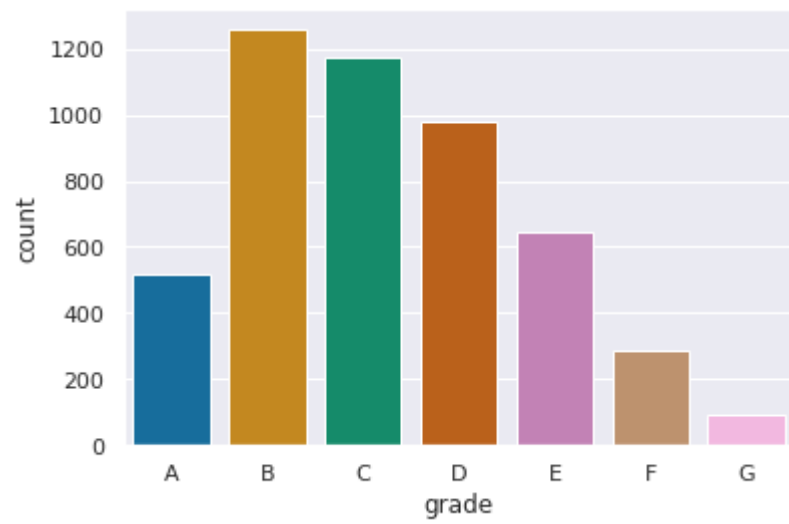
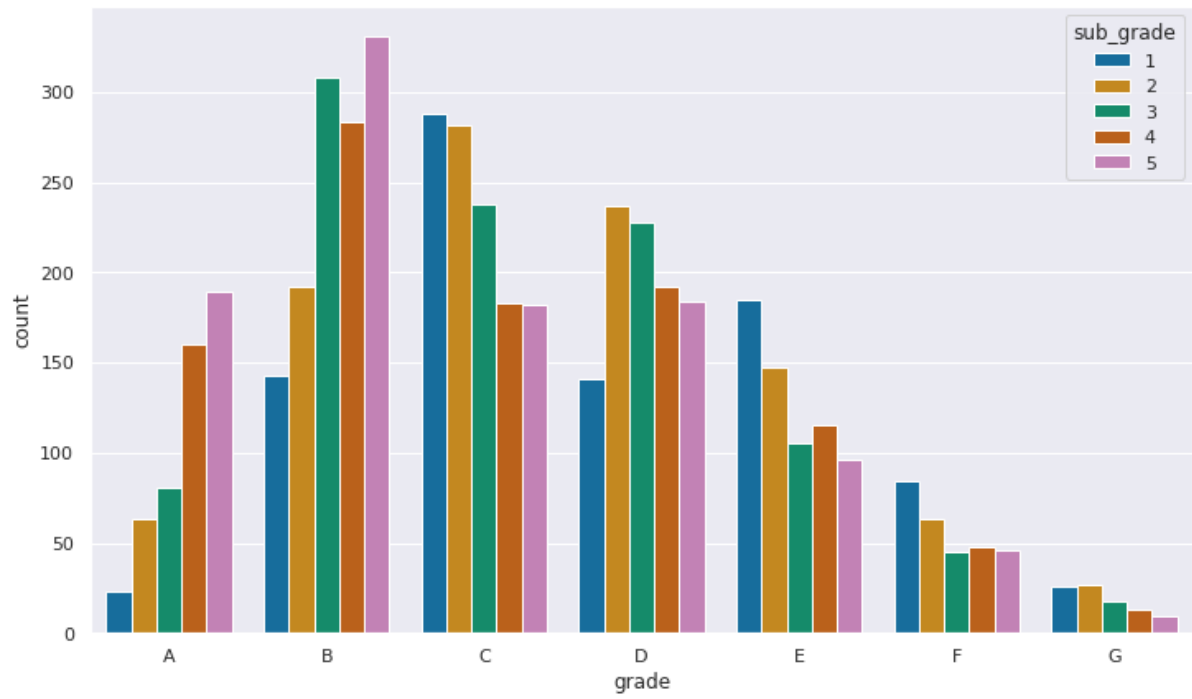
Bar chart used to present categorical data with rectangular bars with heights proportional to the values that they represent

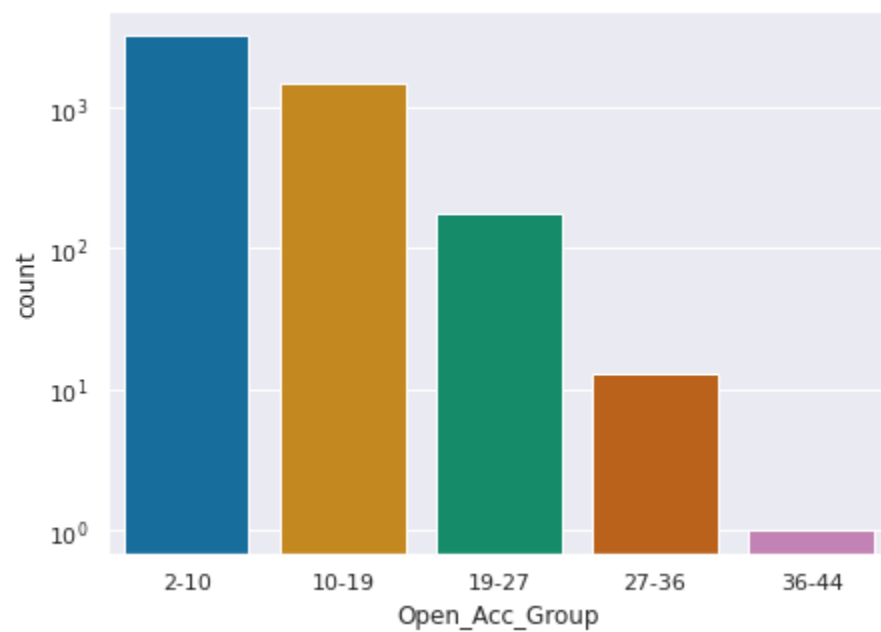
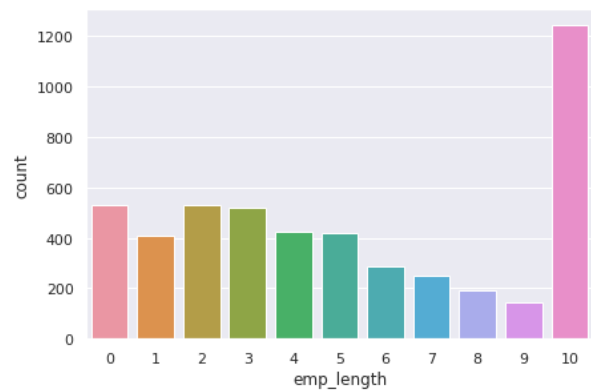
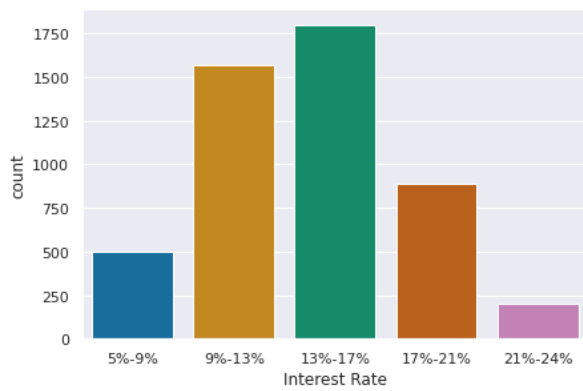
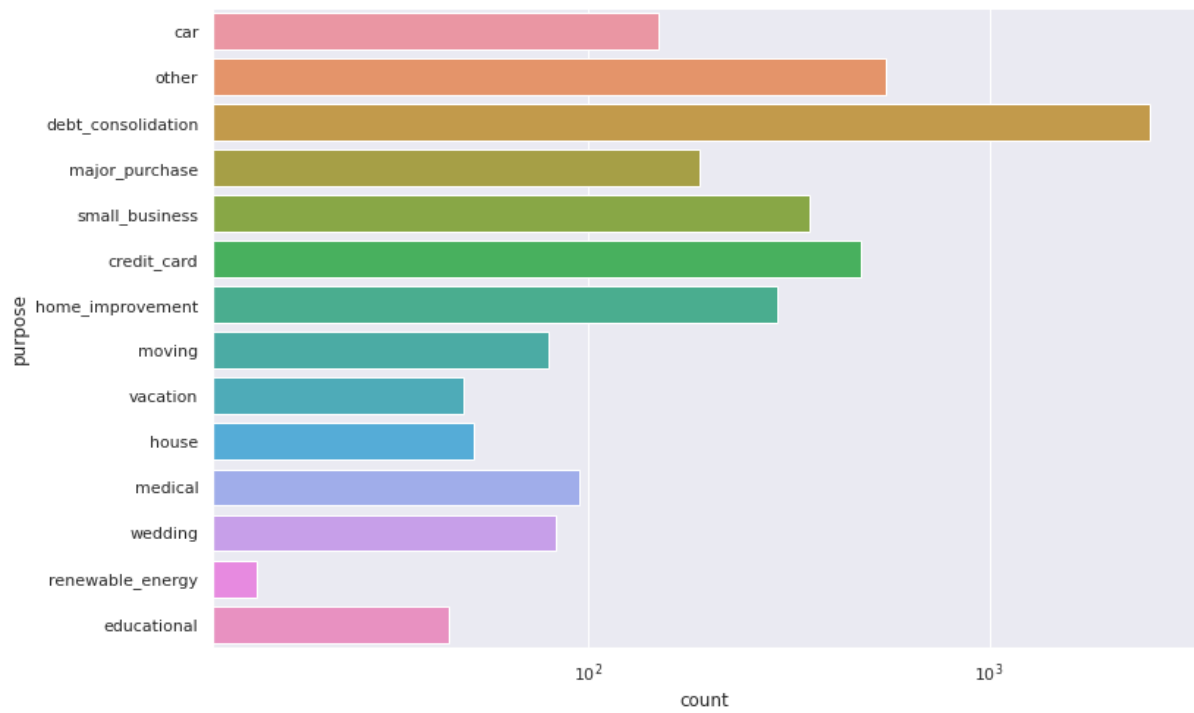


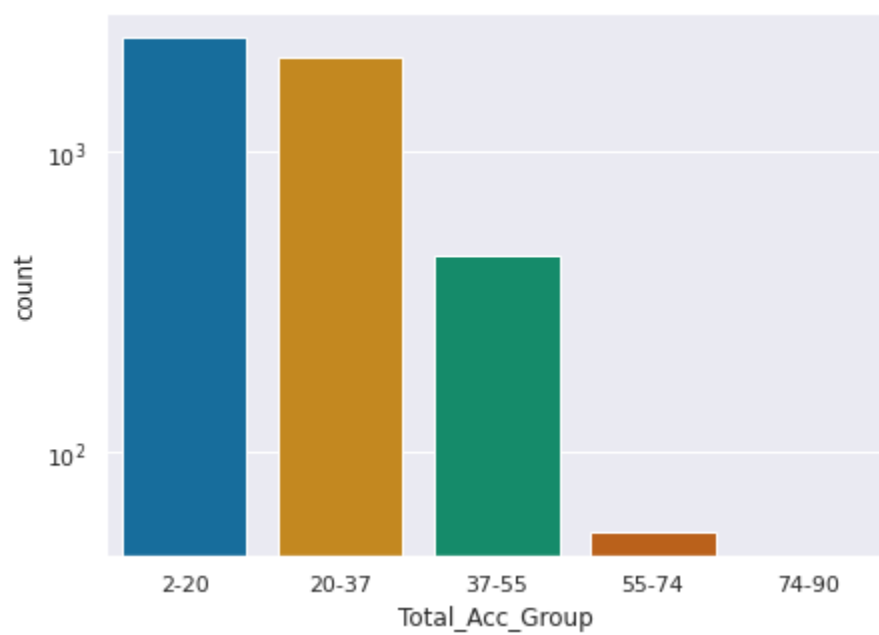
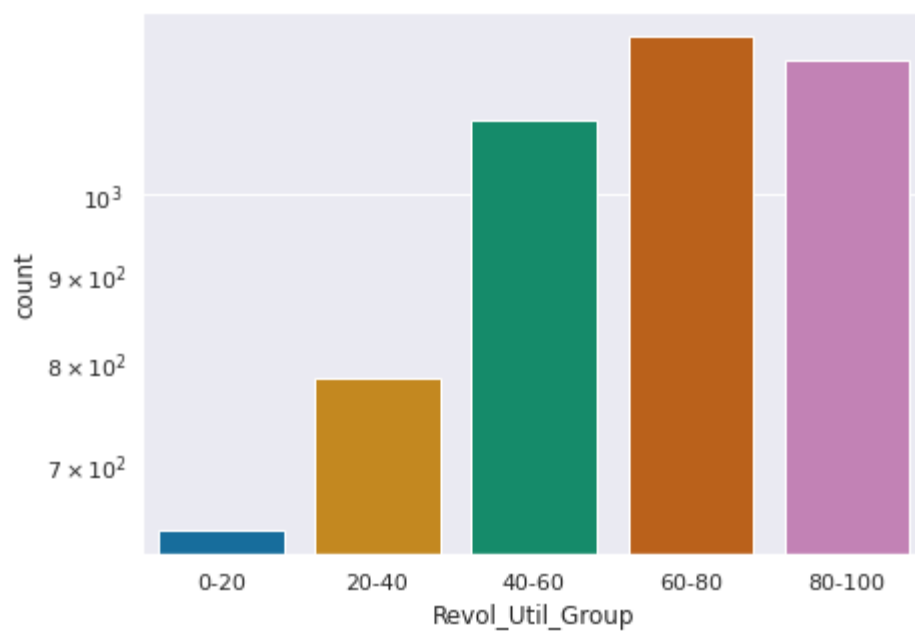


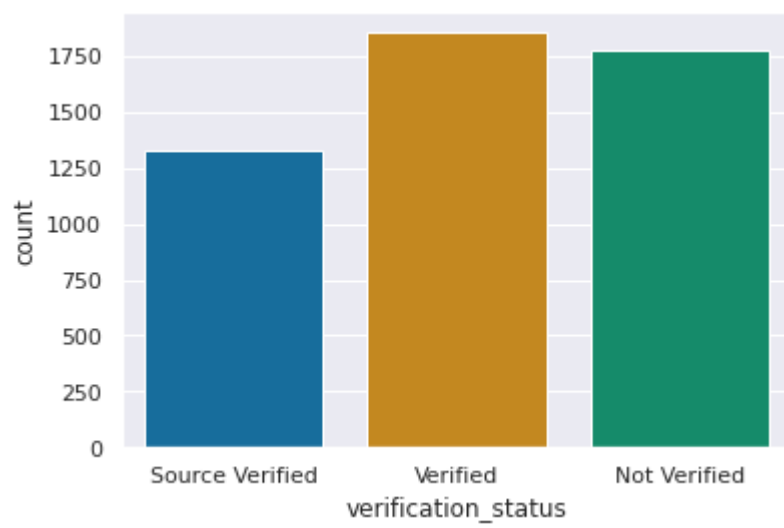
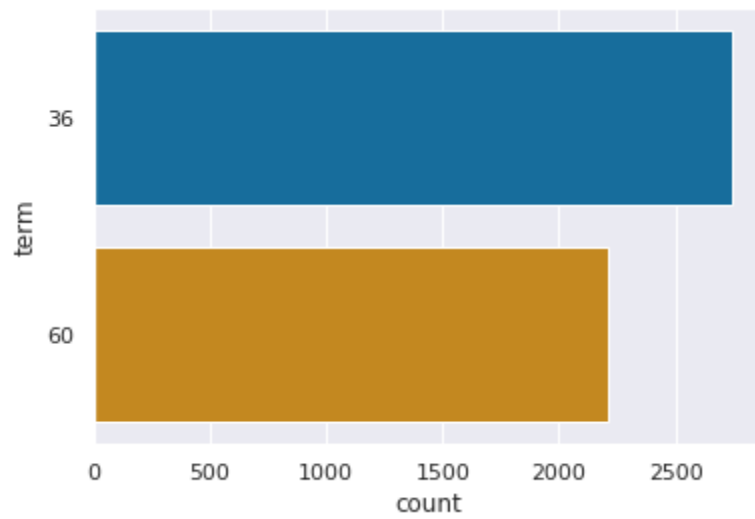
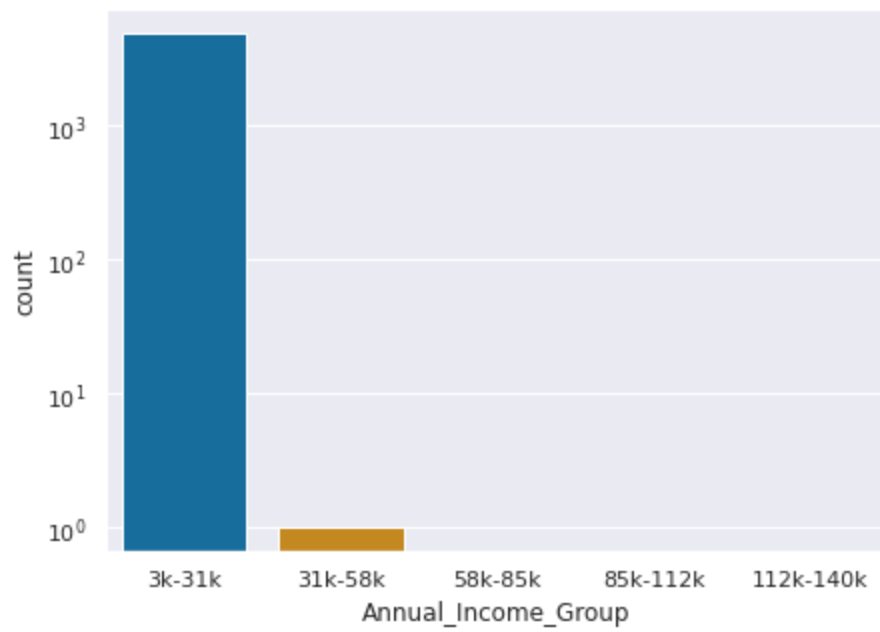
Loan with higher interest are between 10 and 16%

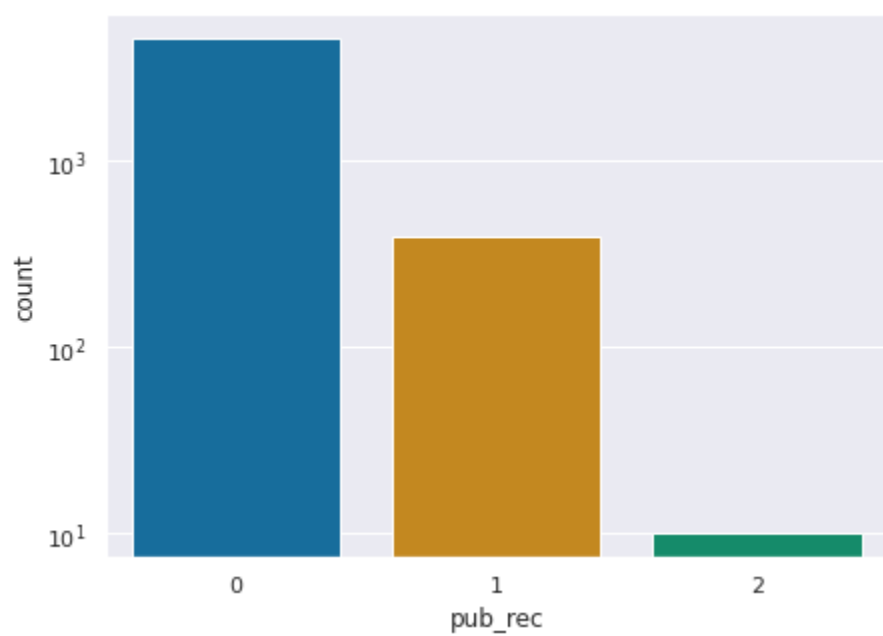
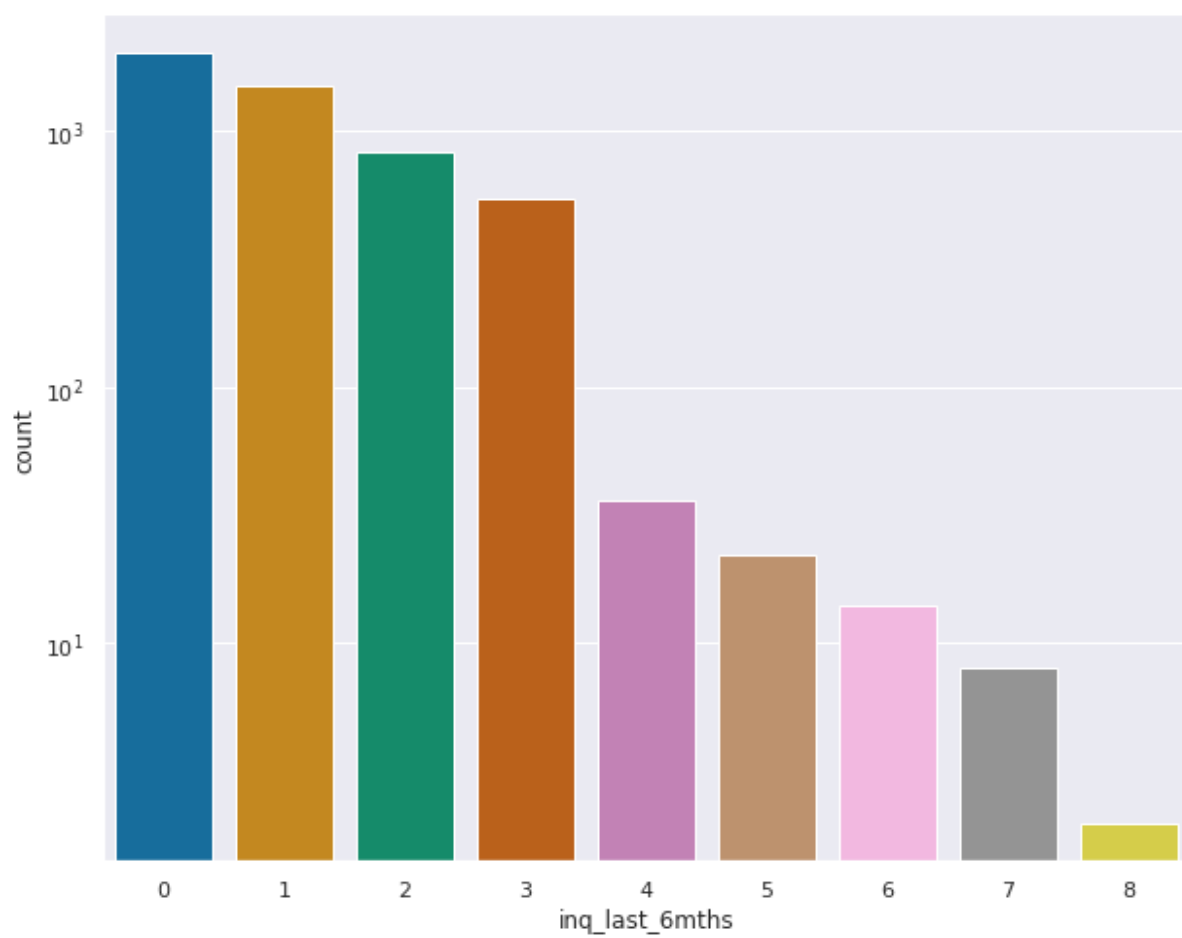


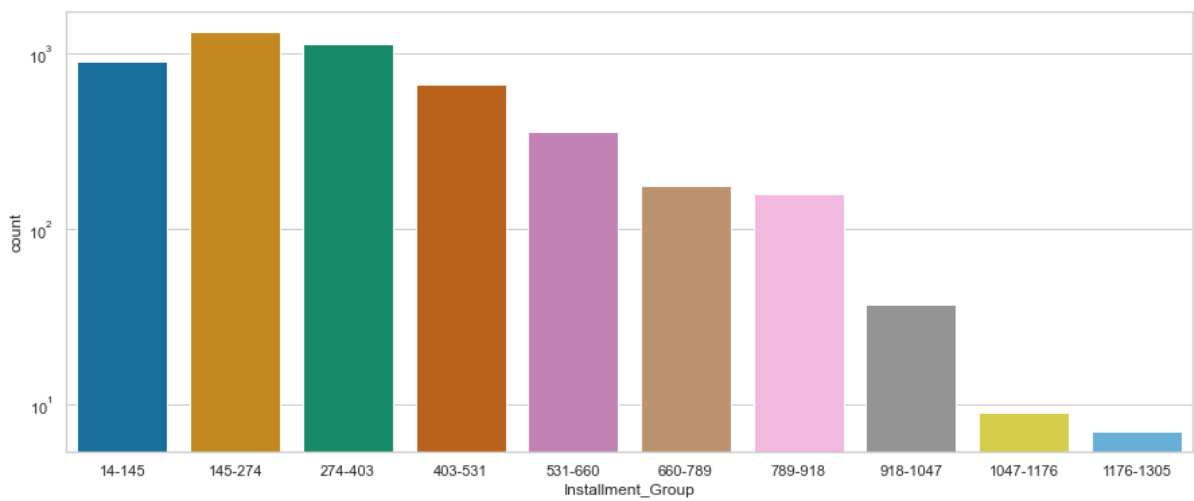
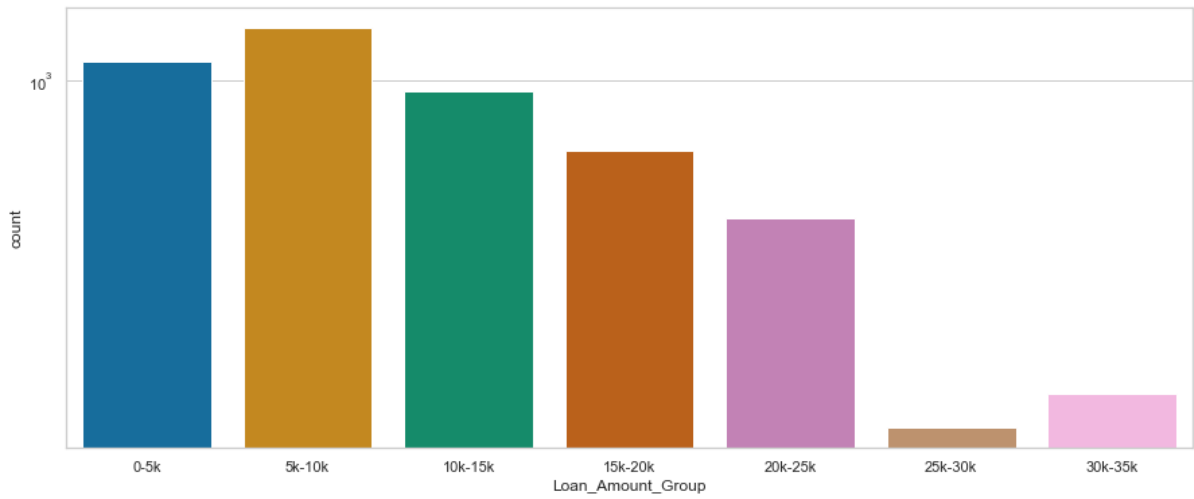
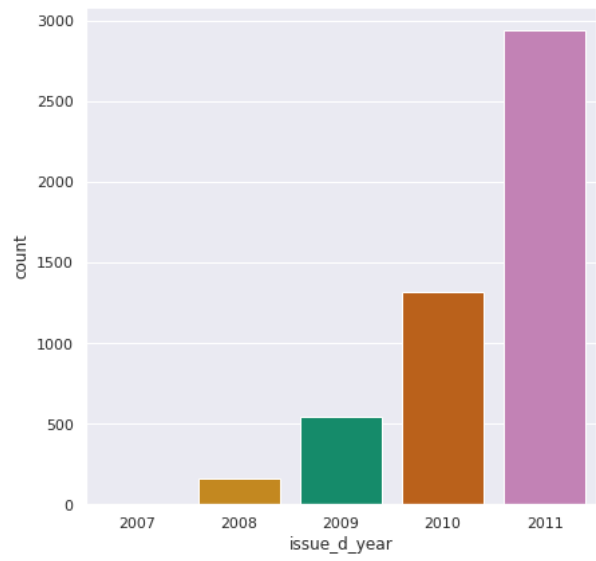
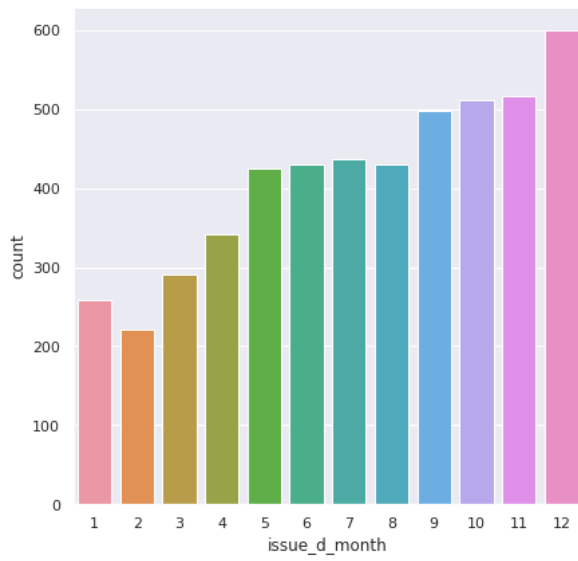


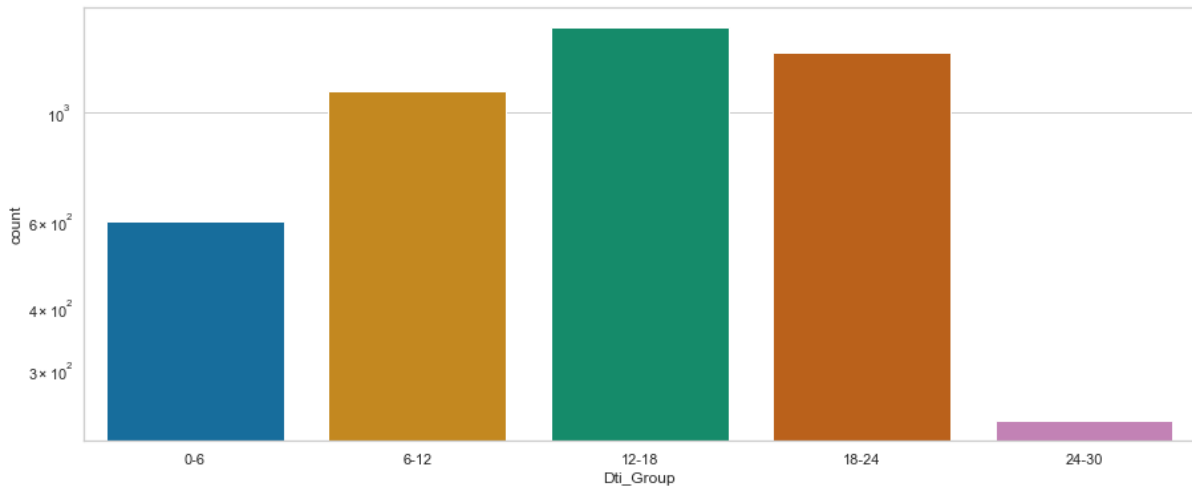








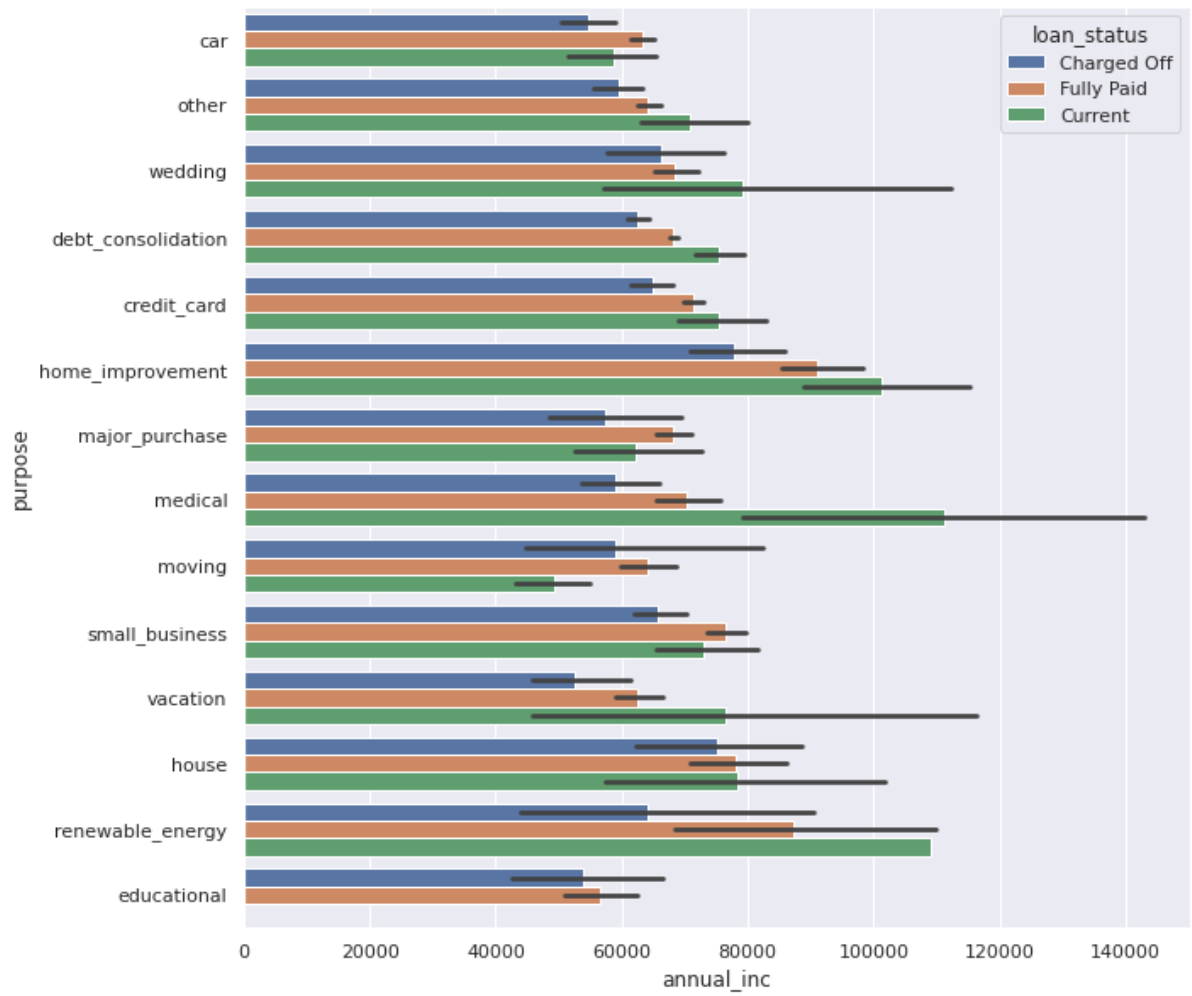


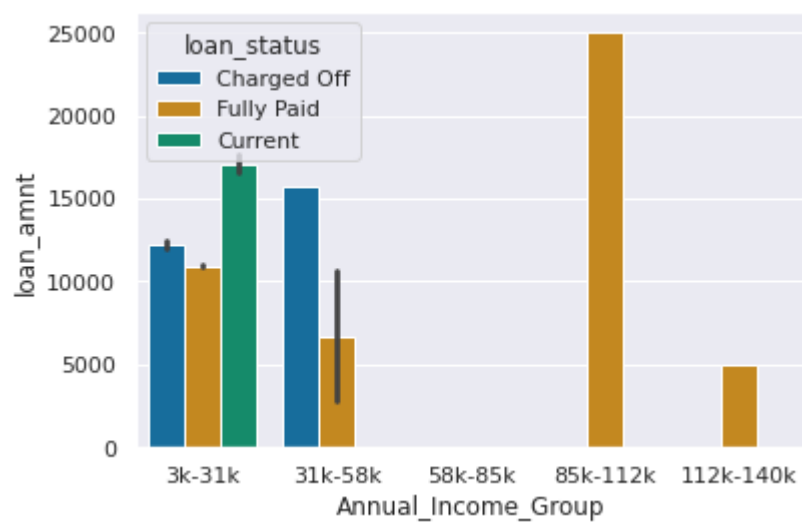
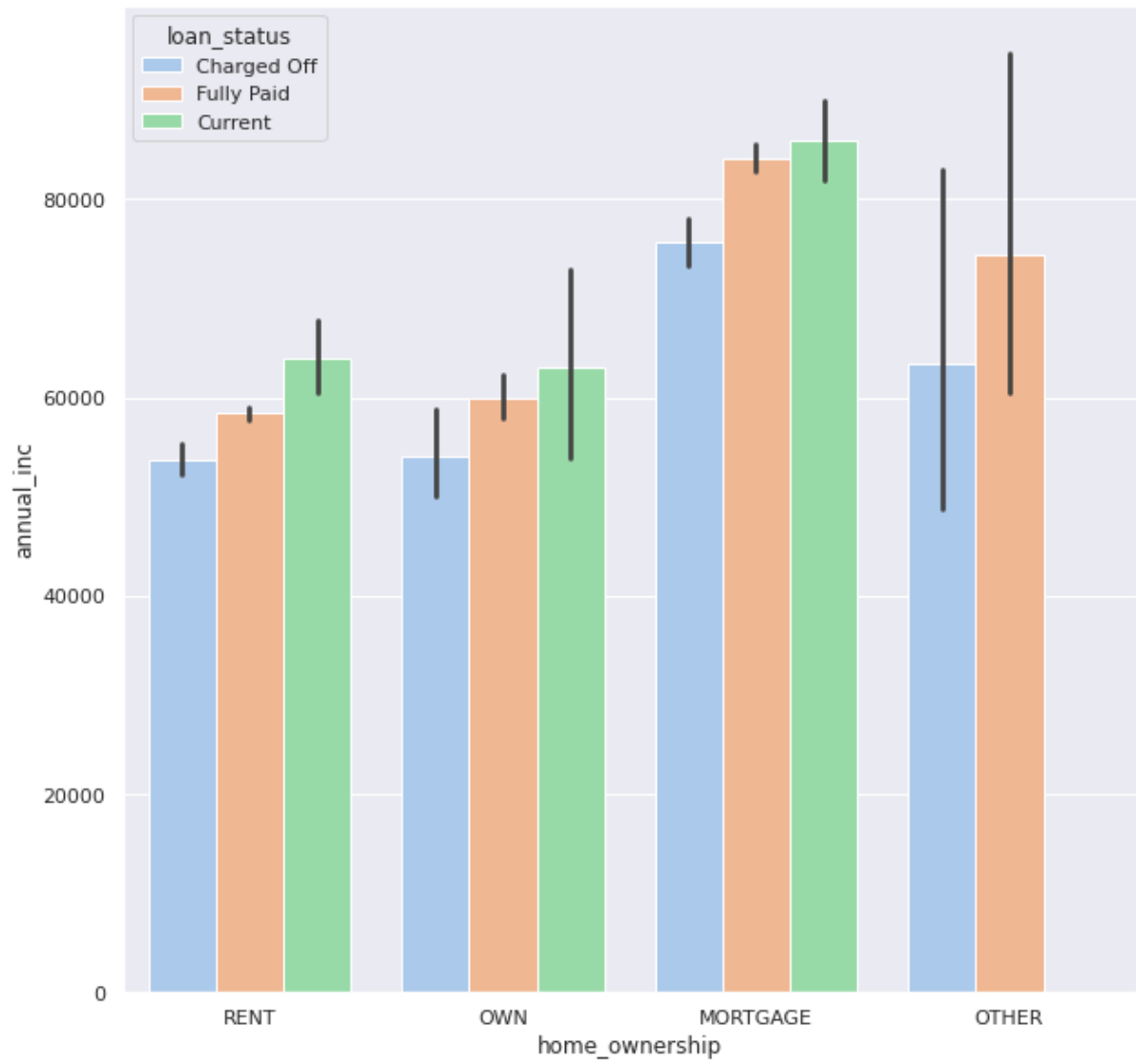


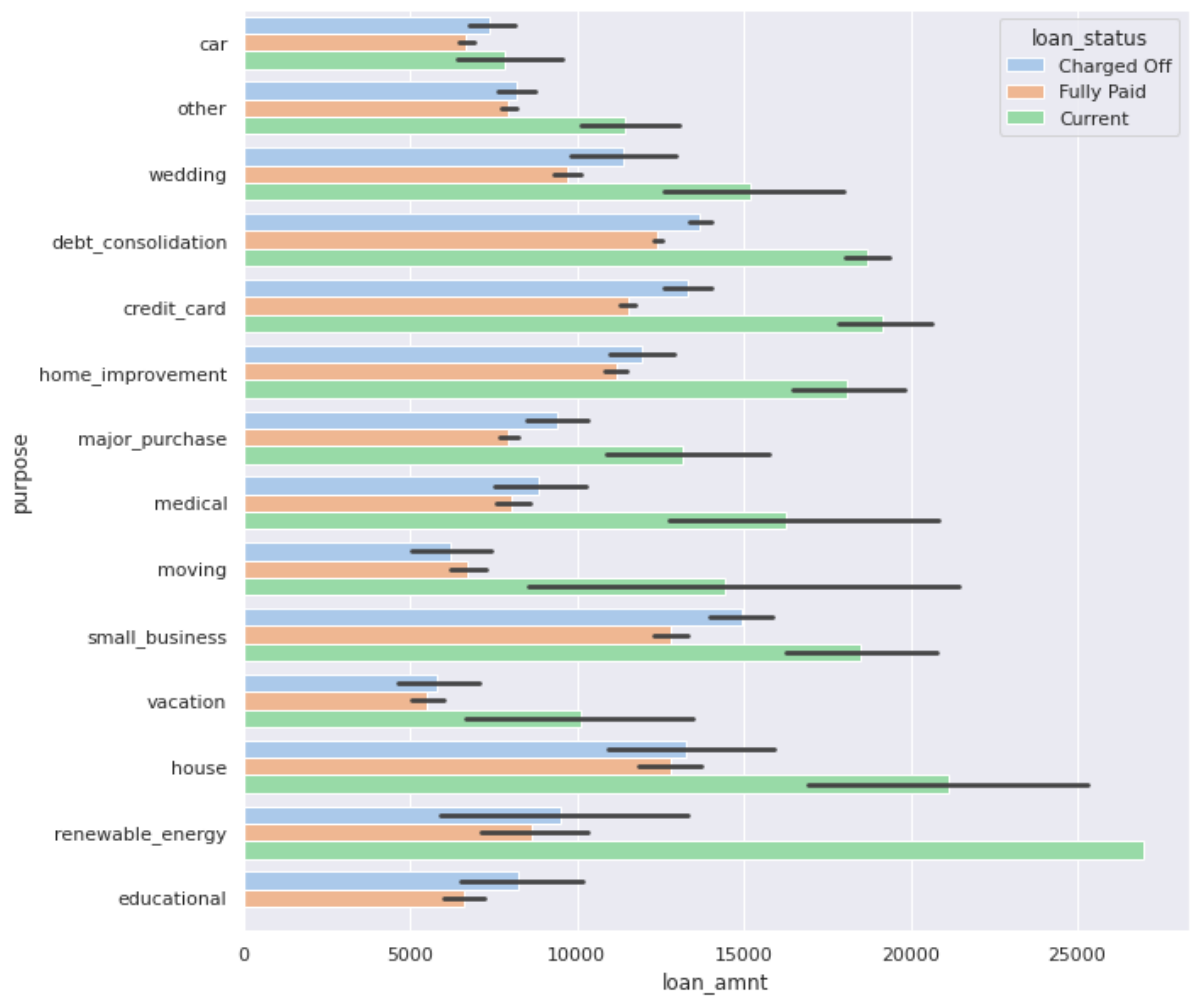
Observations

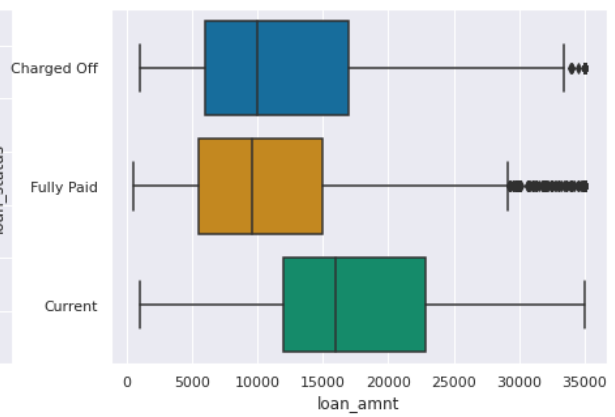
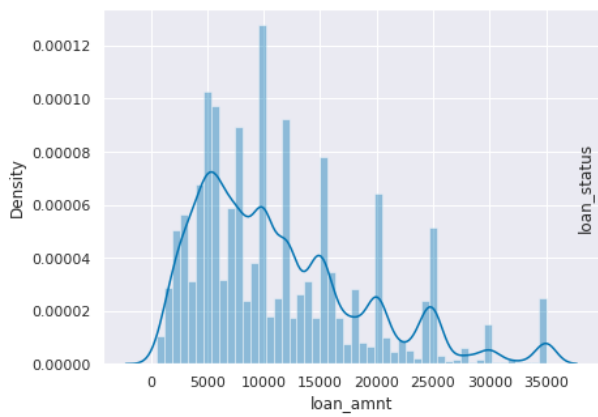
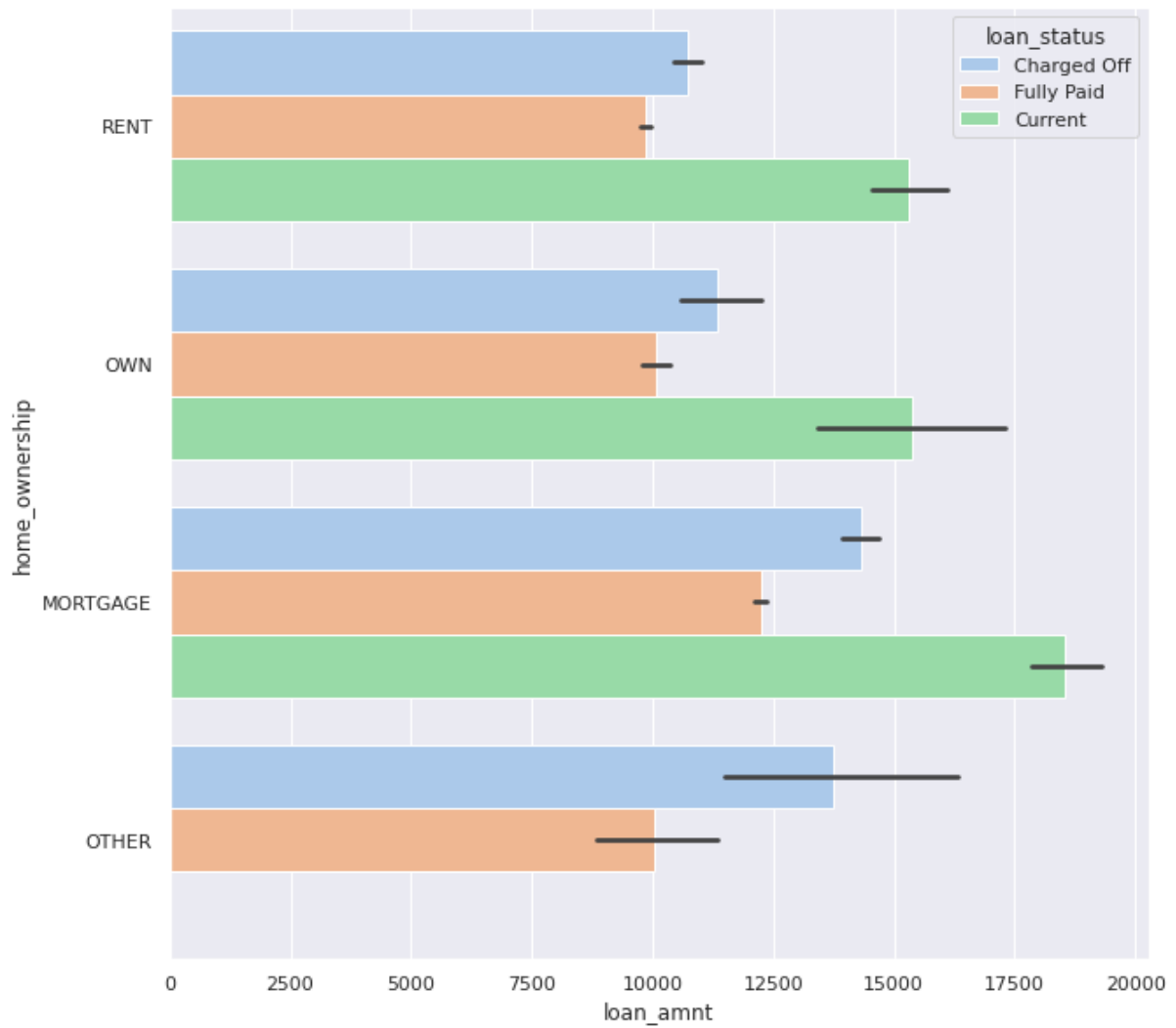
In the above analysis with respect to the charged off loans provide the following insights which has more probability of defaulting loans when

1. Applicants income range of 31000-59000
2. Loan amount is between 5000- 10000 approx
3. Dti is between 12-18
4. When there is no enquiry in last 6 months
5. If loan are in Grade B
6. If the Installment are in range of 145-270
7. Applicant who are levied with interest rate of 13-17%
8. If Applicant who is on Rental are subjected more default
9. If the purpose of the loan is to debt consolidation
10. If the employment length is 10 years
11. If the loan status is not verified
12. However majority of the loan issued in the month dec-2011 had more default perhaps may be due the recession in US

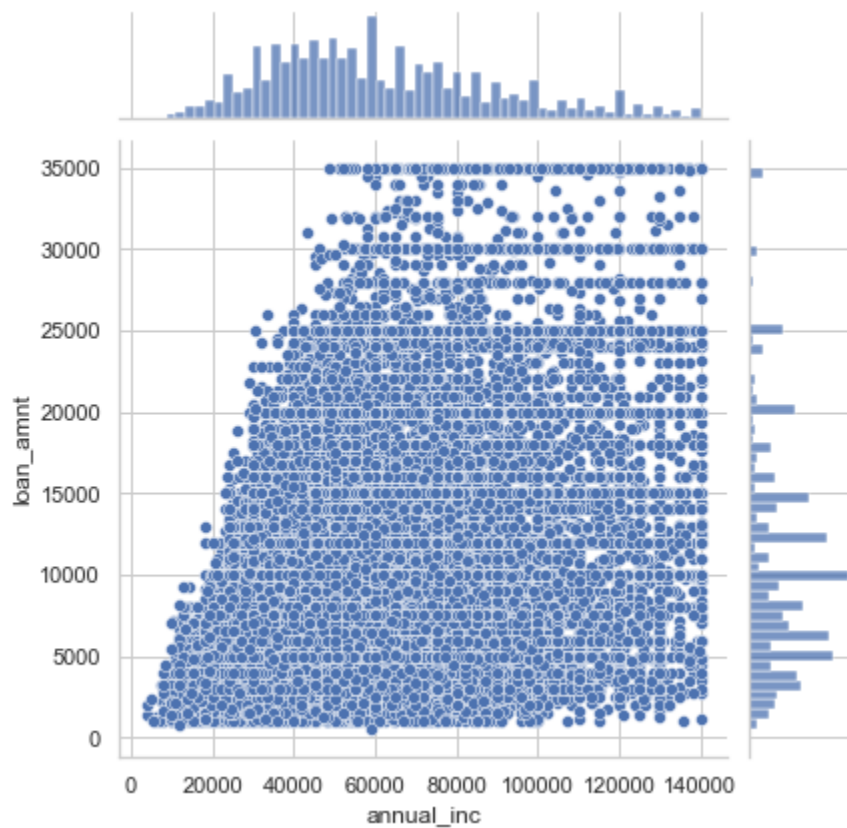




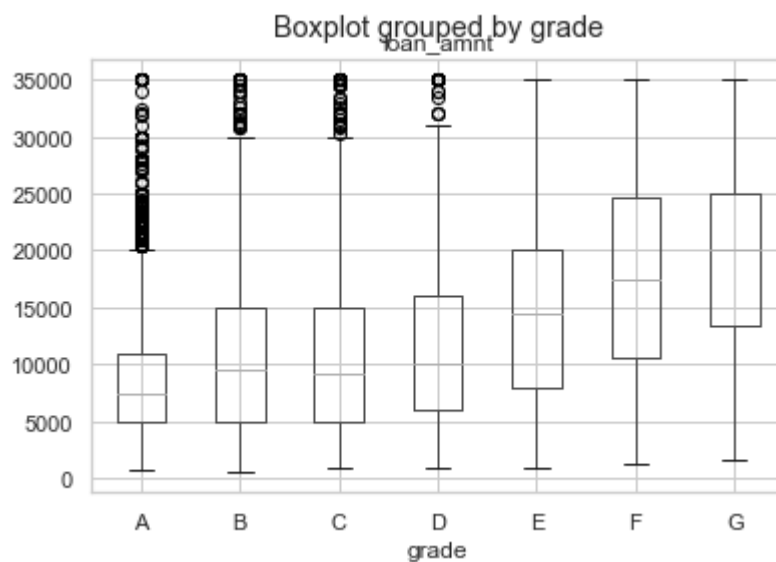




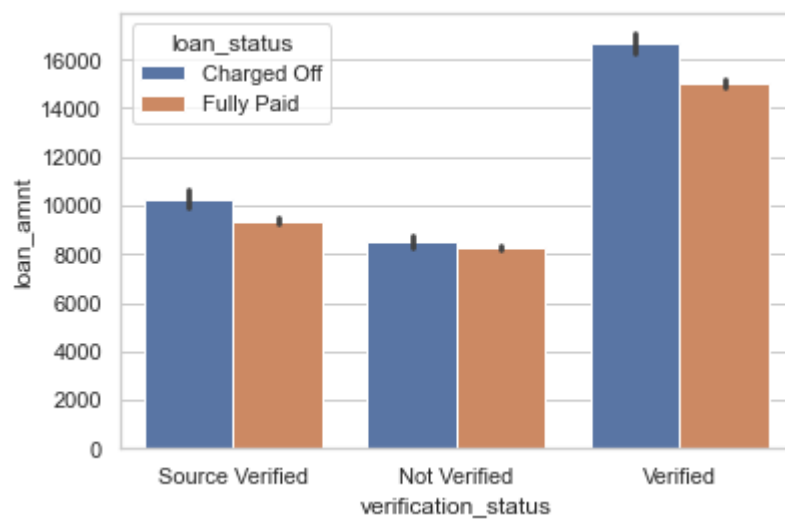
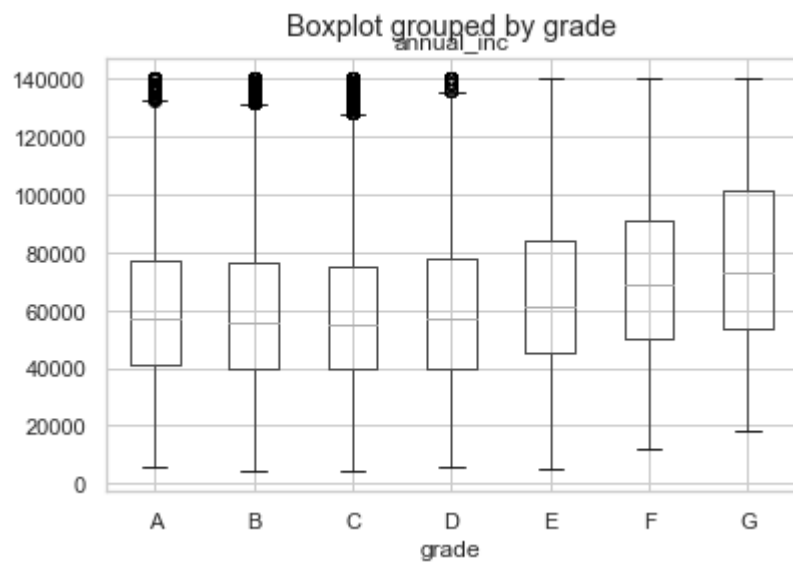
Bivariate Analysis

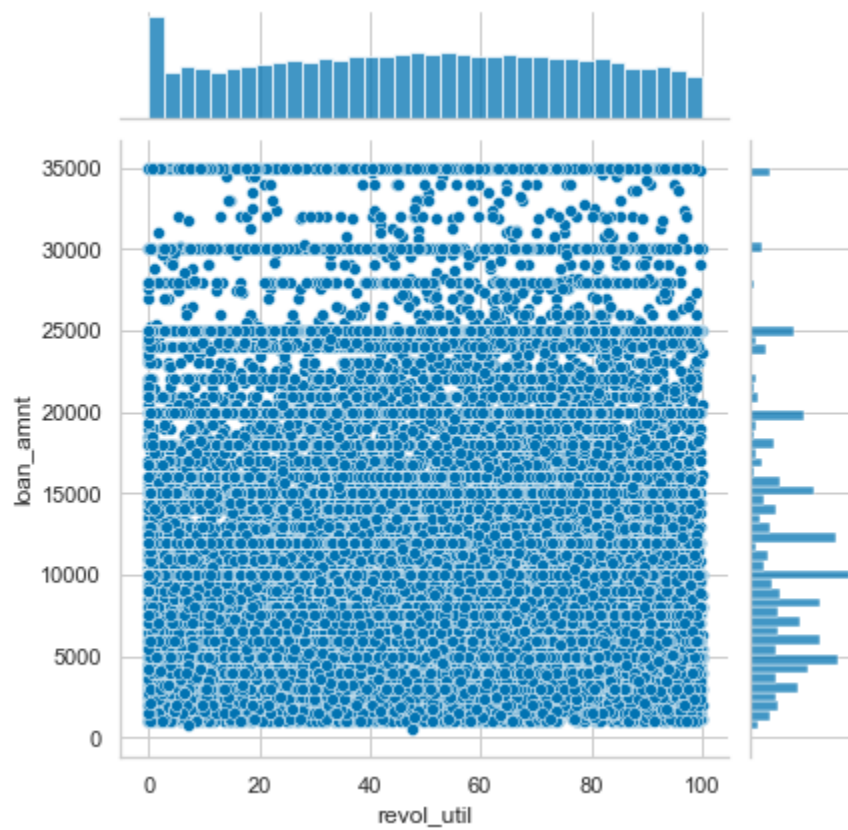


- People who are taking loan 25000 higher has average income lower 50000 , it will be risky loans.

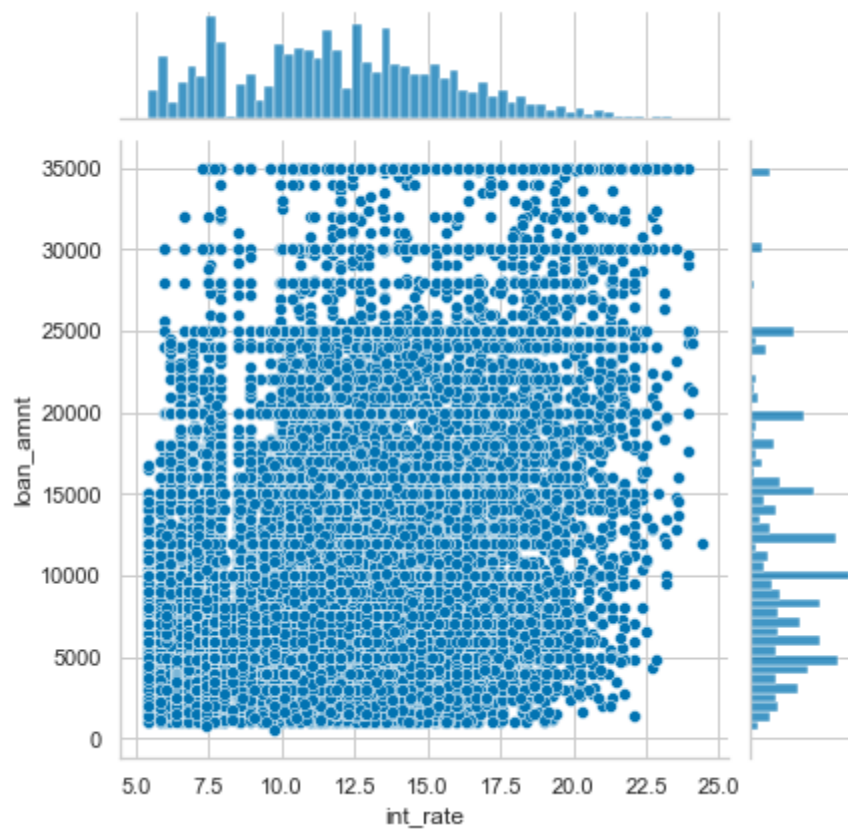


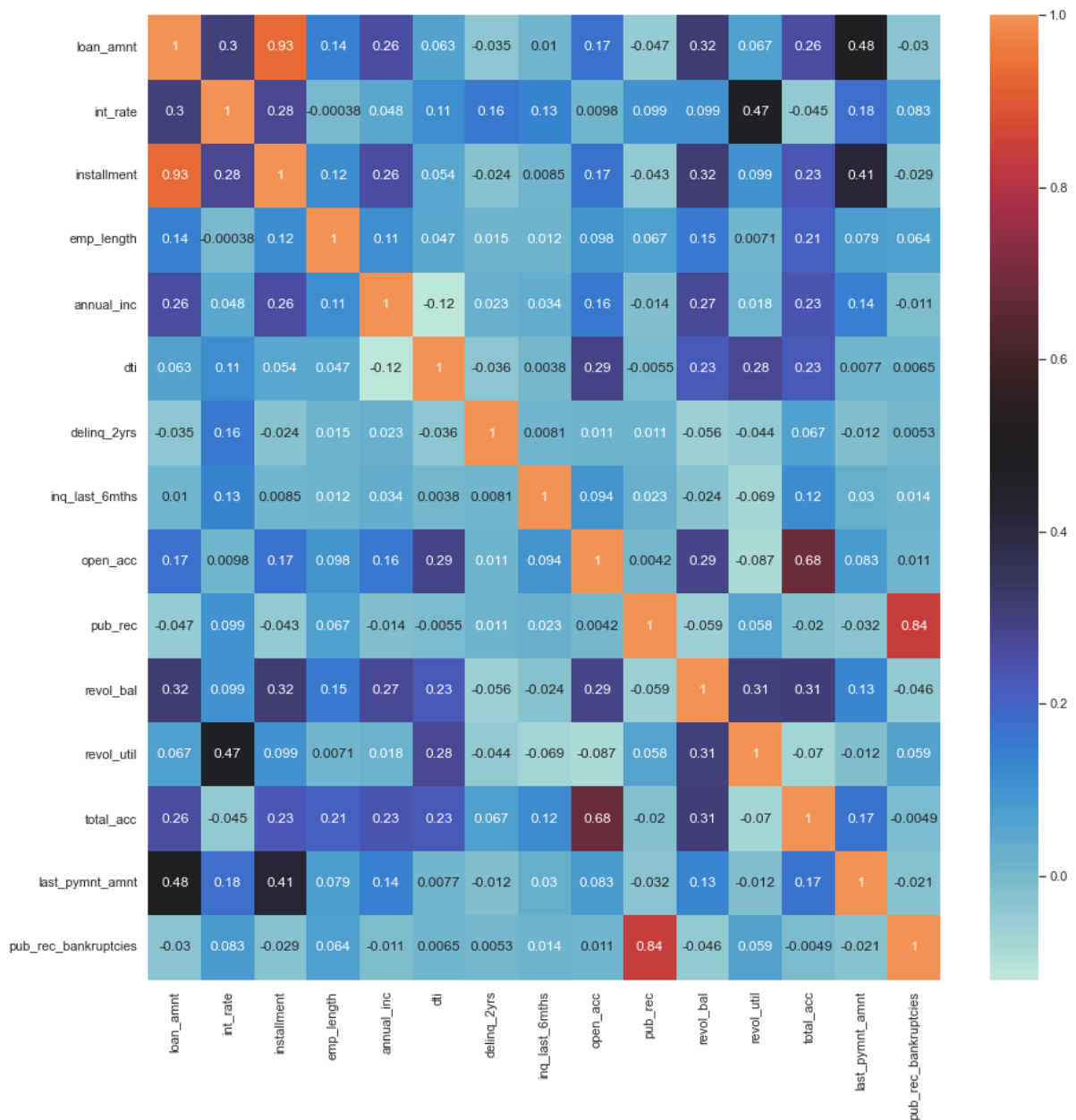
- Median loan for grade G loan being almost 10000 higher than grade of A , B or C ,since the larger loans generally appear to be given lower grade



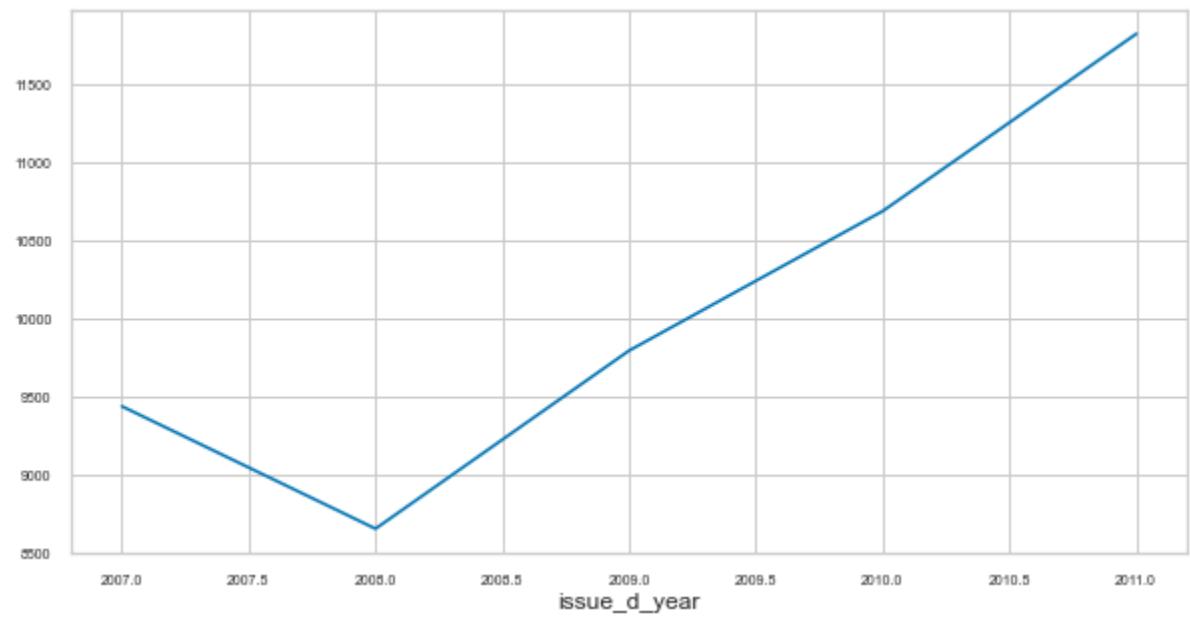
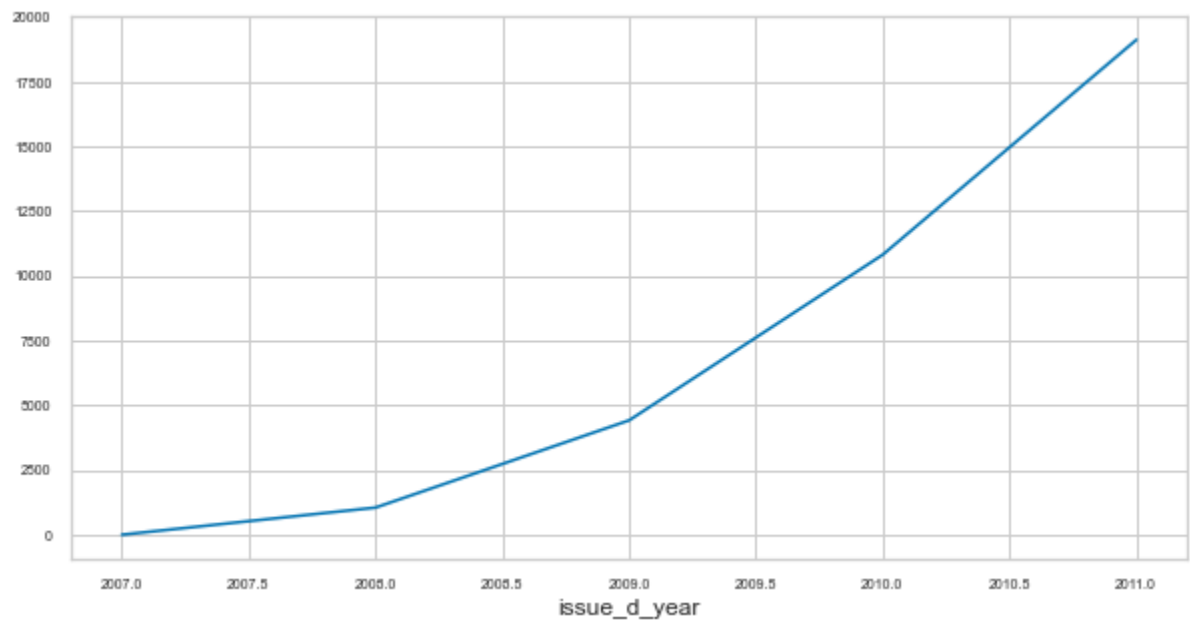


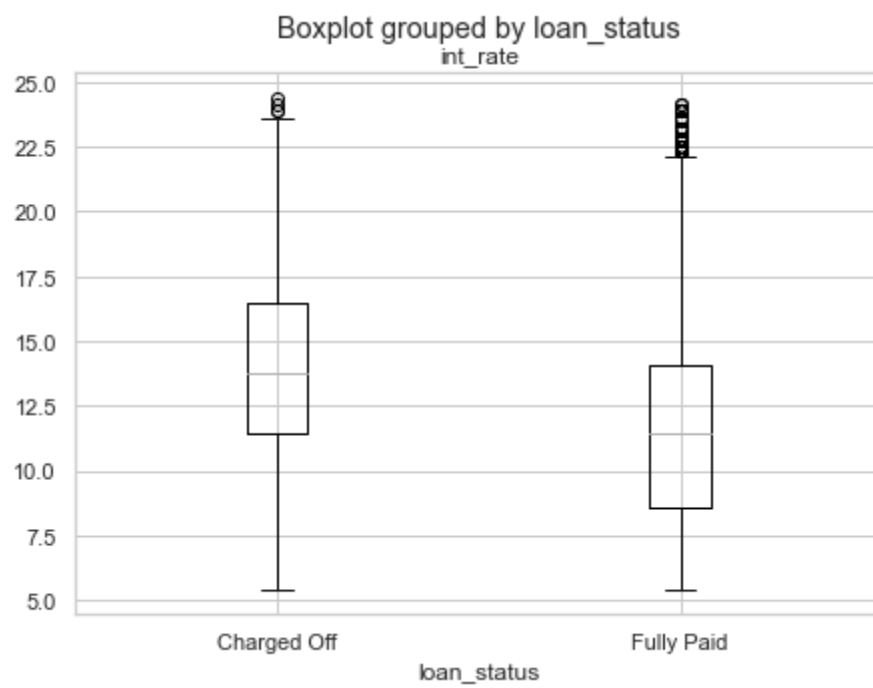
- People who are taking loan 25000 higher has average income lower 50000 , it will be risky loans



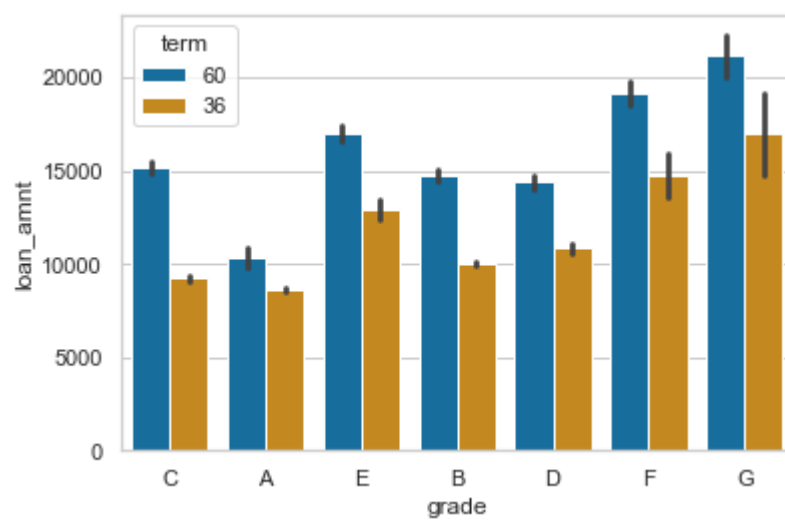


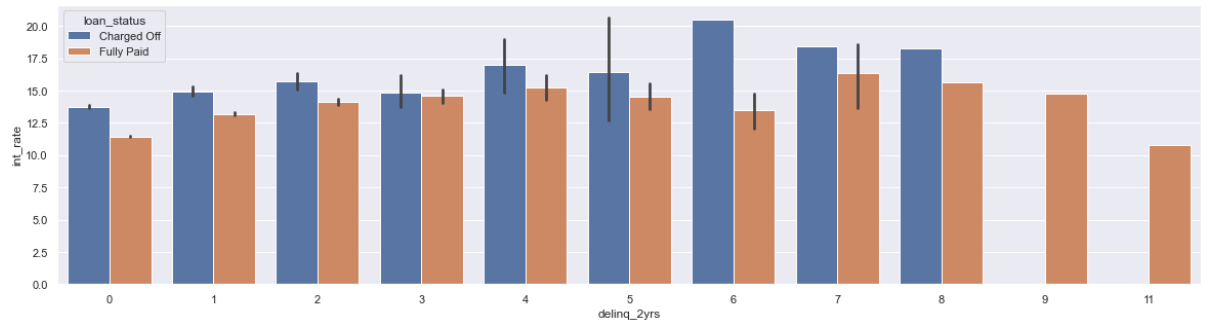
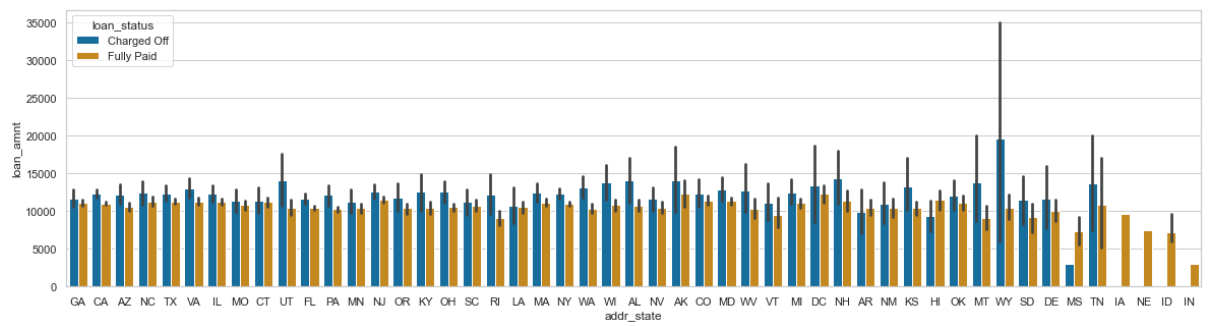
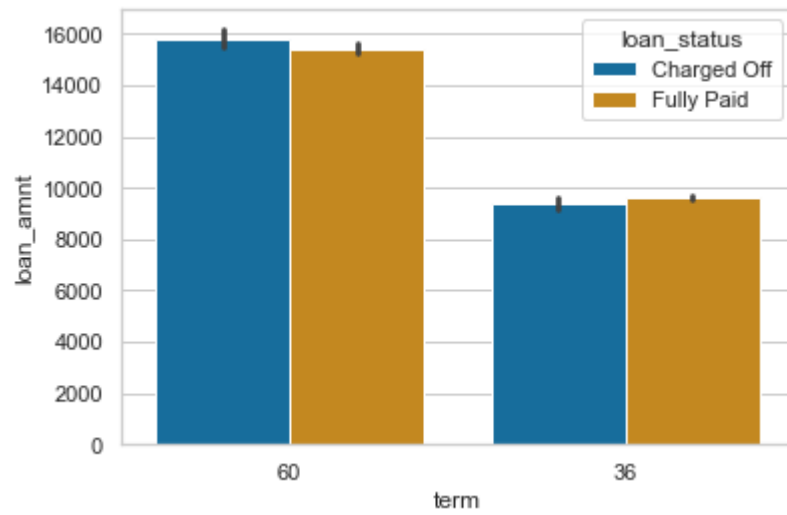
1. loan_amnt is correlated to last_payment with r factor .48
2. int_rate is correlated to revol_util with r factor .47 , this is good sign that loans are charged with higher interest rate from risky loan
3. loan_amnt and revol_bal are correlated with r factor .32 this is not a good sign due higher amount of loan is been approved to high risk borrowers.

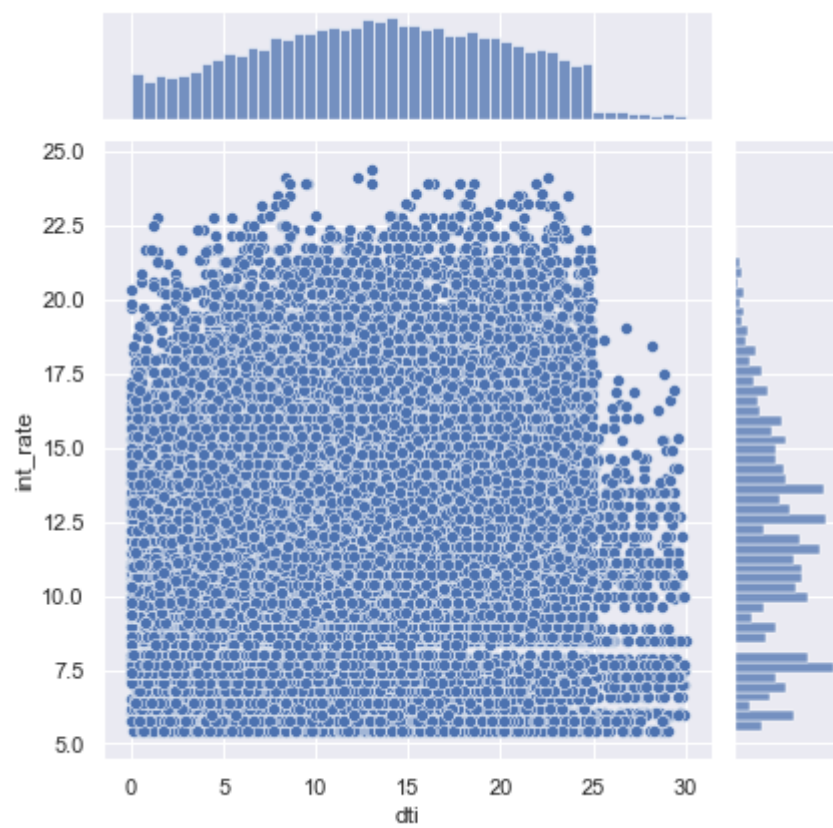
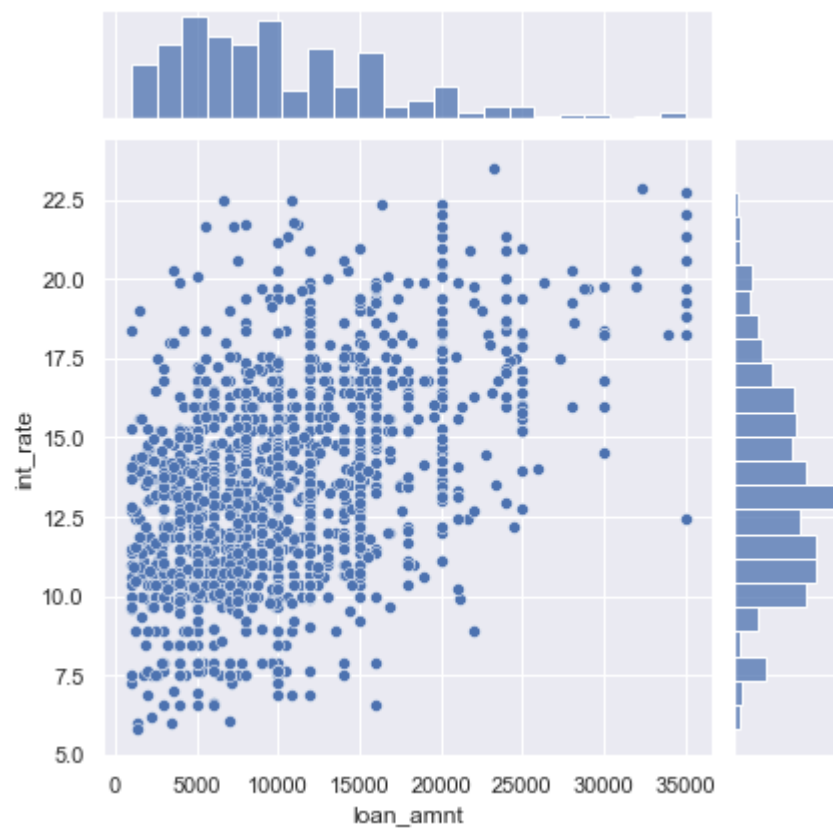


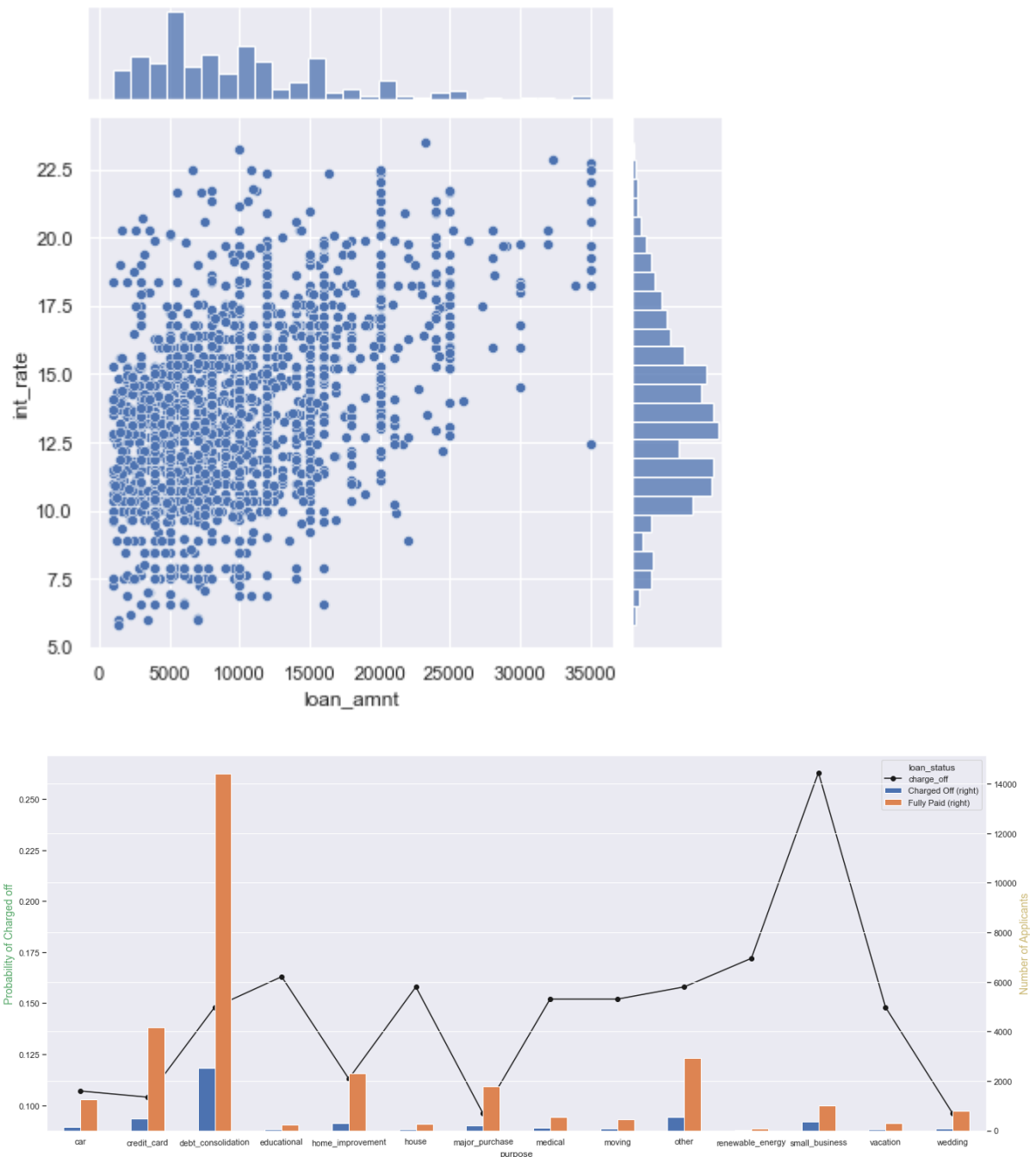


- Loans with higher interest rate are more likely to Charged off









- Applicants who has taken the Loan for 'wedding' has the lowest probability of charge off of less than 1%, which guarantees the loan will be payed off where as on the other hand it is observed that the 'small_business' has the highest probability of charge off with 26% which clearly states that banks should be extra careful when approving loans for the purpose of small business and after verifying the guarantee assets the loan should be approved and proper credit approval process should be followed for this purpose.

Recommendations:

Final Take away with analysis

- Probability of charging off loans with respect to the analysis is

1. If applicant taking loan for 'home improvement' and having higher income bracket of 60-70k
2. Loan amount range 30-35k with interest charged rate of 15-17.5%
3. Home ownership is 'Mortgage 'and with loan of 14-16k
4. Employment length is 10 years with loan amount 12-14k
5. Grade G with higher interest rate of 20%