Lending Club Case Study

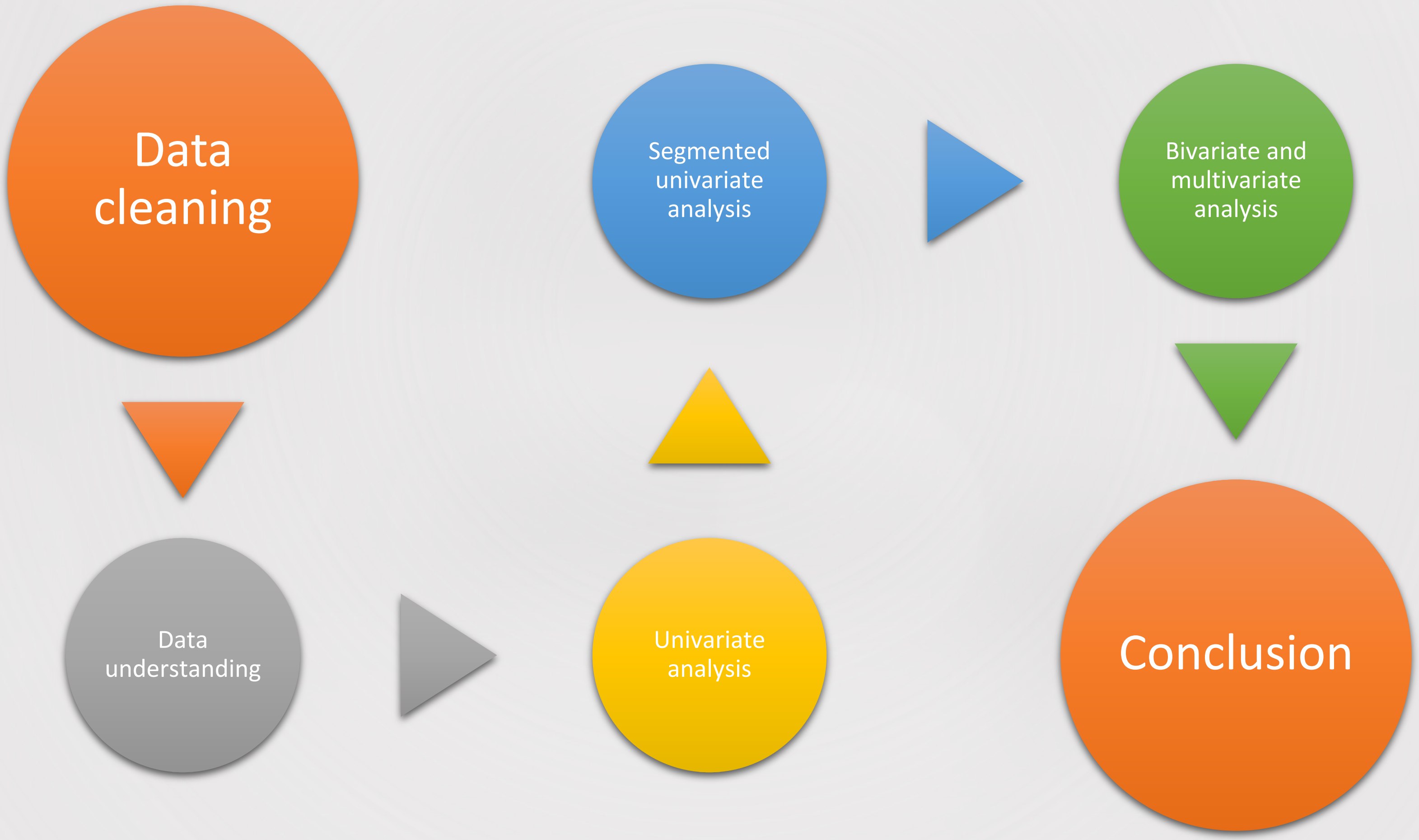
-By Arpit Lalit Jain

Problem Statement

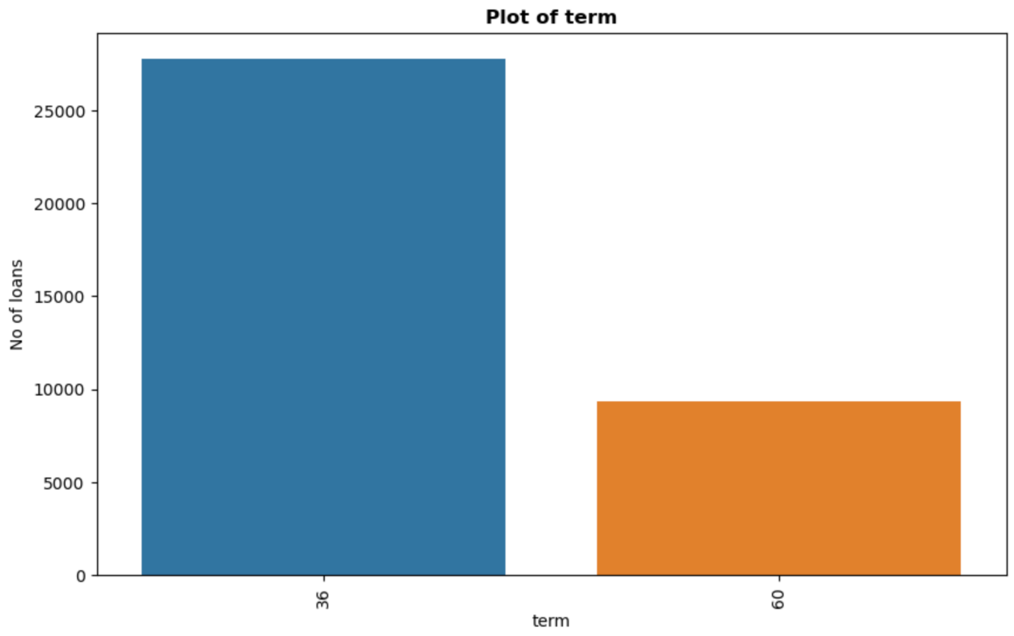
- The Objective is to analyze loan applicant data and identify key factors that can indicate whether a loan is likely to default or not. By understanding these factors, the company aims to reduce credit losses and make informed decisions when approving loans

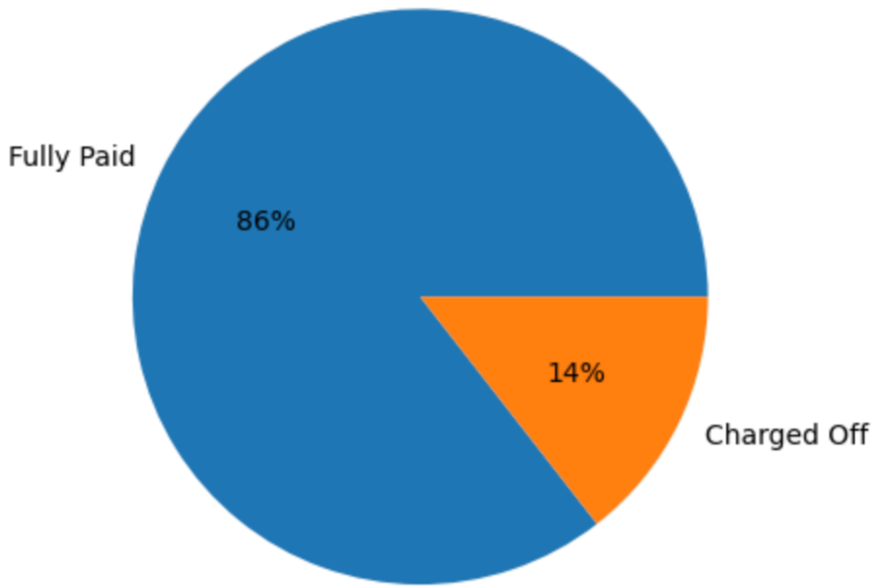
Analysis Approach:

The analysis approach involves conducting the exploratory data analysis (EDA) on the loan dataset. EDA techniques will be used to gain insights into the data and understand the relationships between difference variables and loan defaults.



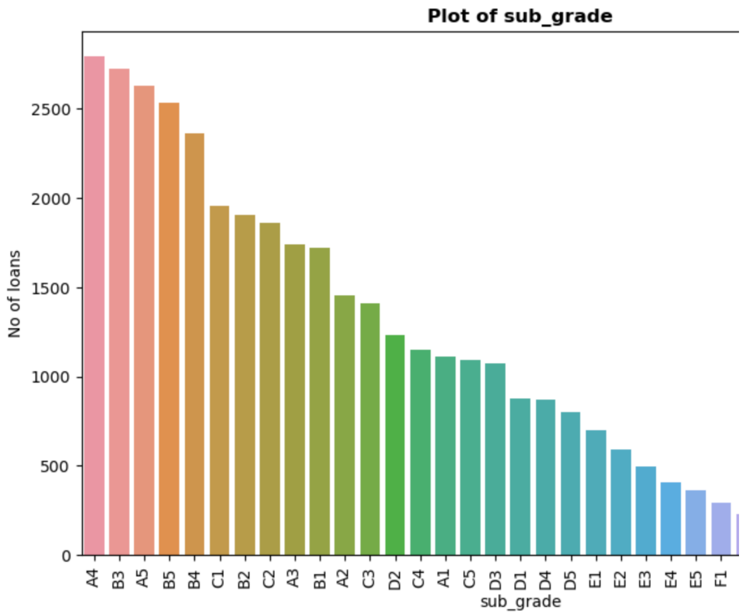
Univariate Analysis

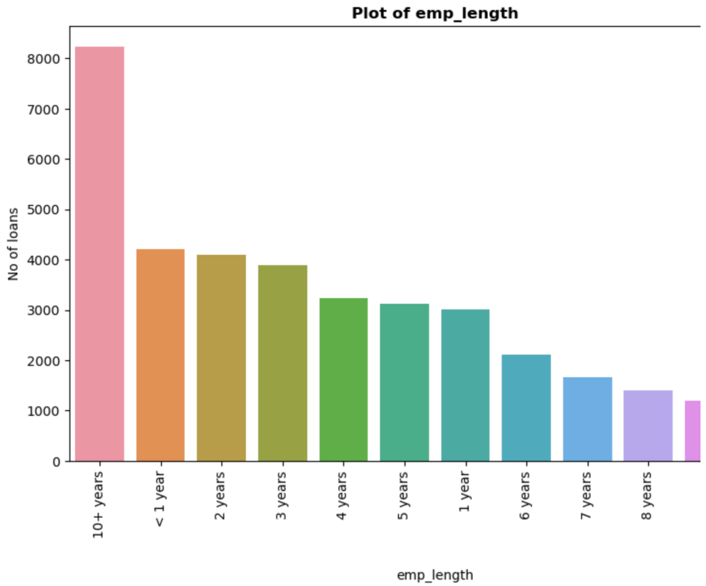


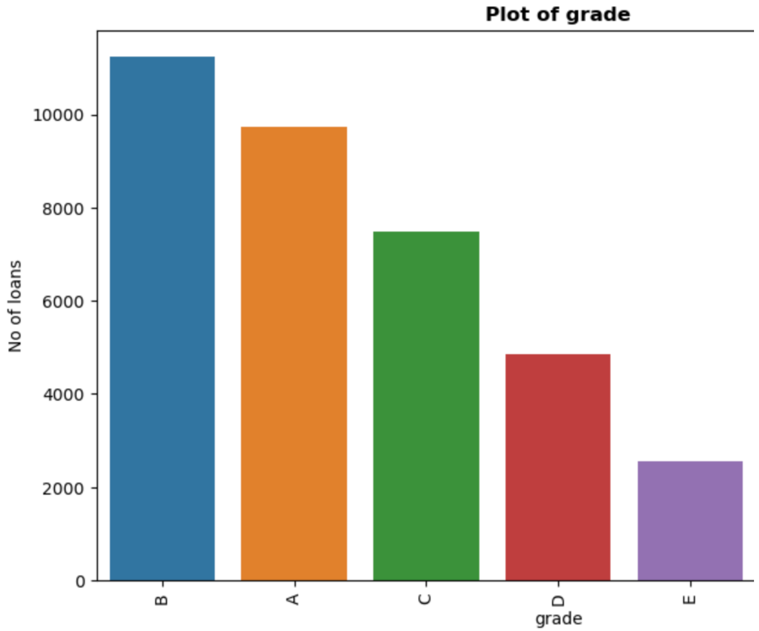


- The dataset includes two types of loans: 36-month tenure and 60-month tenure.
- The majority of applicants opt for a 36-month loan tenure.
- Approximately 14% of the applicants have defaulted on their loans.
- Around 86% of the applicants have successfully paid off their loans.

Univariate Analysis

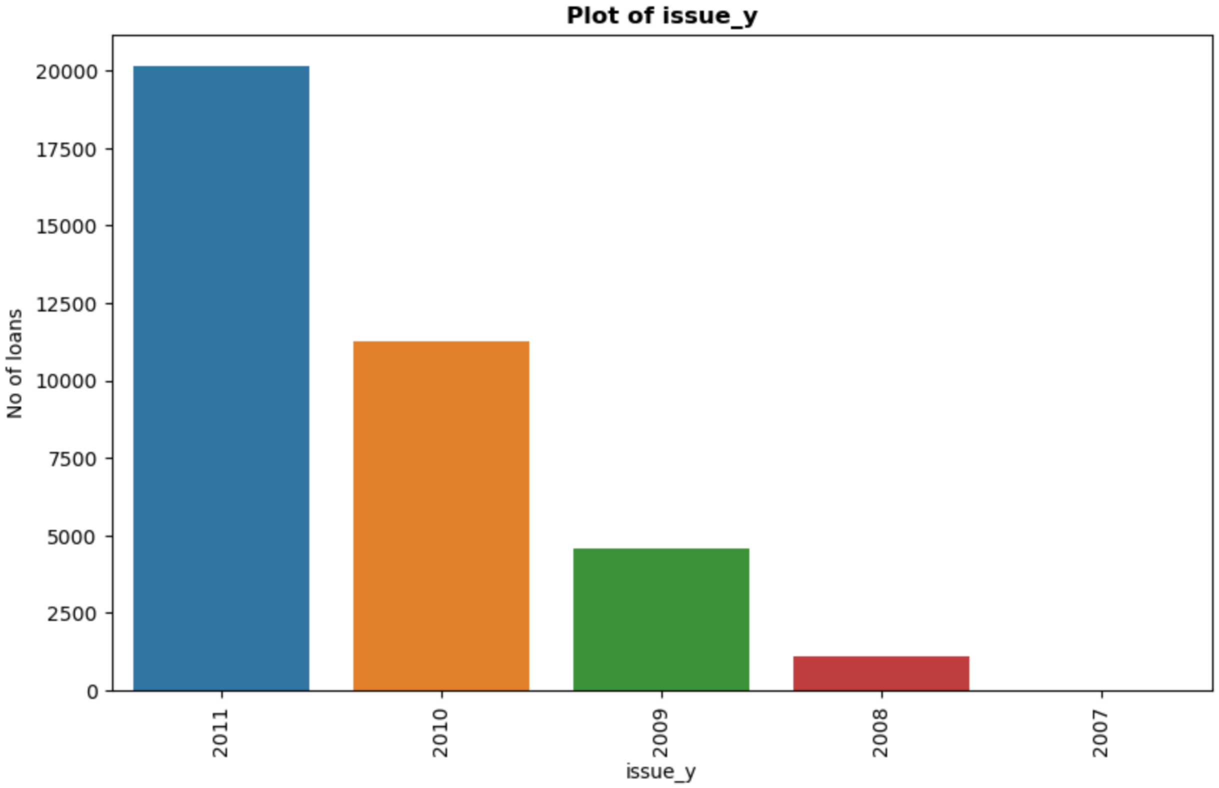


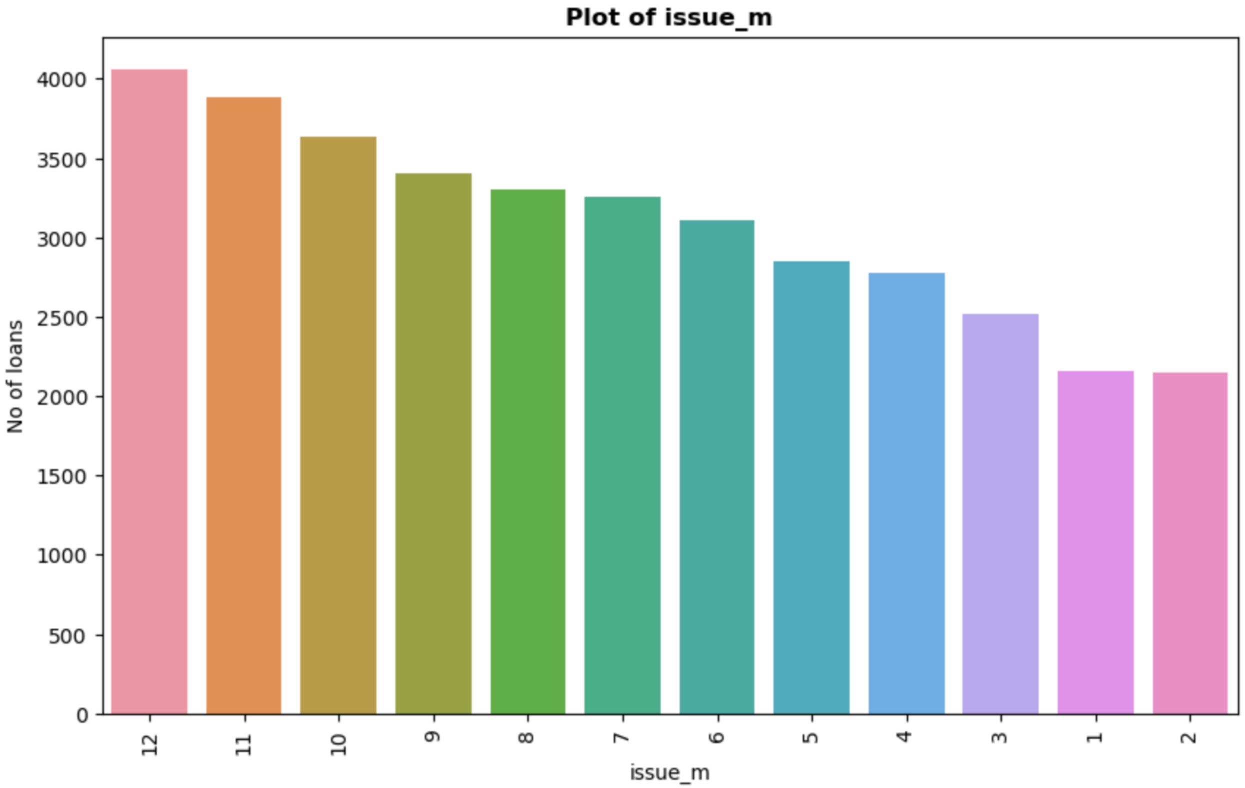




- The graph representing loan distribution by grade shows that Grade B and Grade A receive a higher number of loans compared to other grades.
- The graph illustrating loan distribution by sub-grade reveals that Sub Grades A4, B3, A5, B5, and B4 receive a greater number of loans compared to other sub-grades.
- The graph representing loan allocation based on employees' length of service demonstrates that individuals with 10 years and above of experience are granted more loans compared to those with lesser experience.

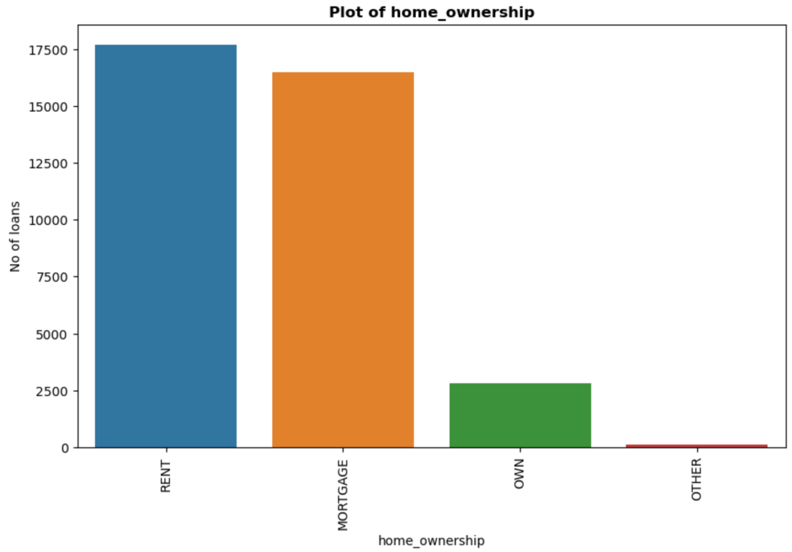
Univariate Analysis

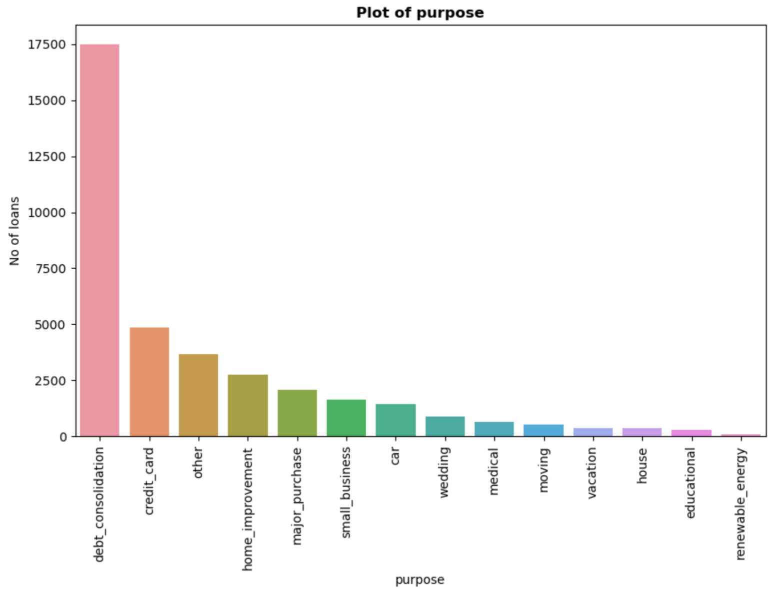


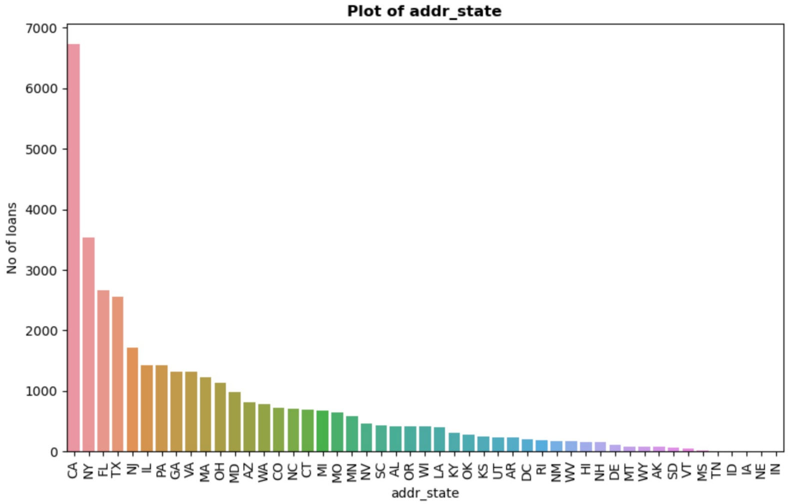


- The graph representing the issue year of loans indicates that the highest number of loans were given in the year 2011. Furthermore, there is a noticeable increasing trend in loan numbers as the years progress.
- The graph illustrating the issue month of loans reveals a consistent upward trend in the number of loans as the months increase. Particularly, the months of October, November, and December stand out with a higher number of loans being granted during these periods.

Univariate Analysis



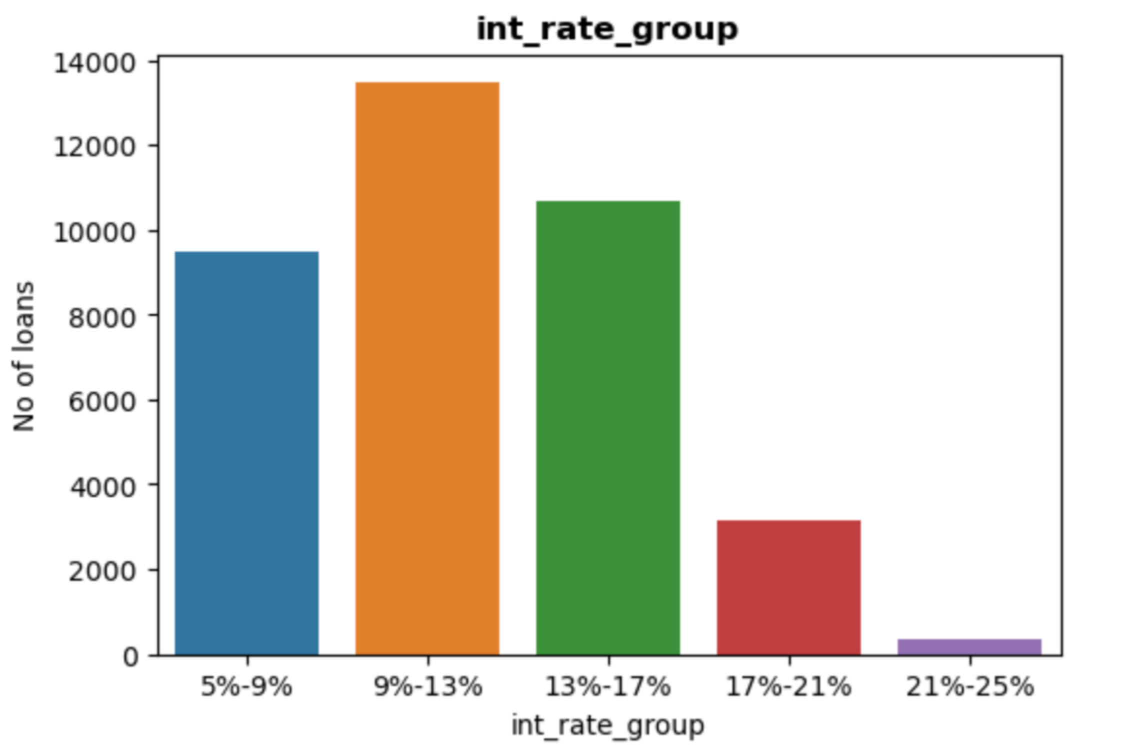




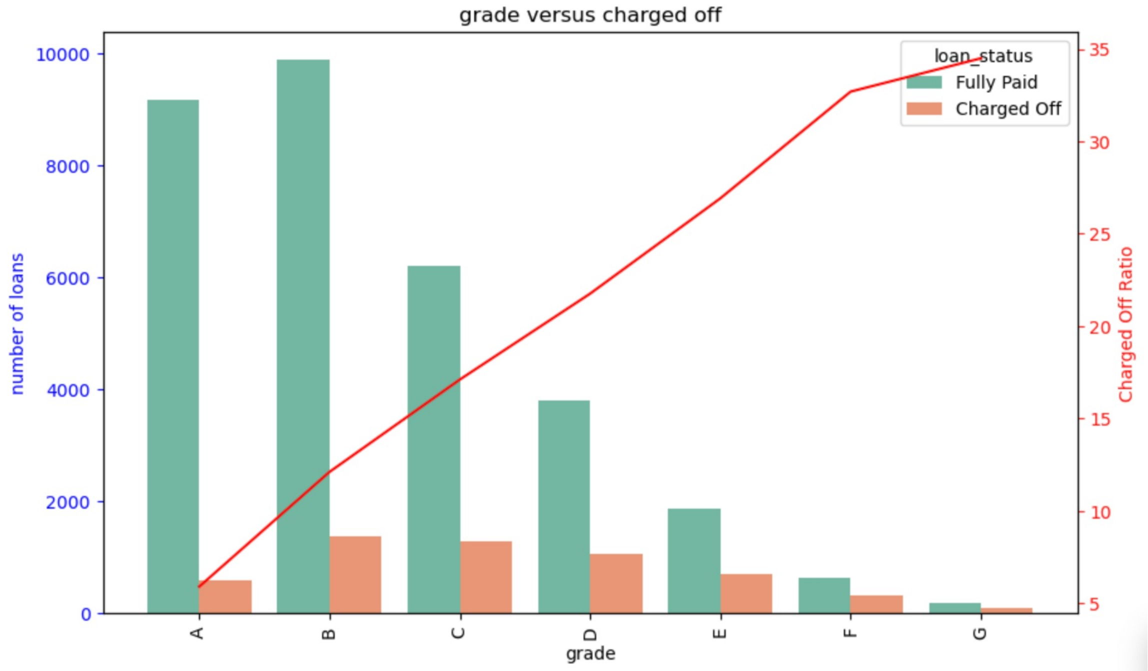
- The states with the highest number of loans taken are CA, NY, FL, and TX.
- The most common reasons for taking out loans are debt consolidation, paying off credit cards, and other purposes.
- The categories with the fewest number of loans taken are education and renewable energy.
- Individuals residing in rented houses or providing a mortgage as collateral tend to take out more loans.

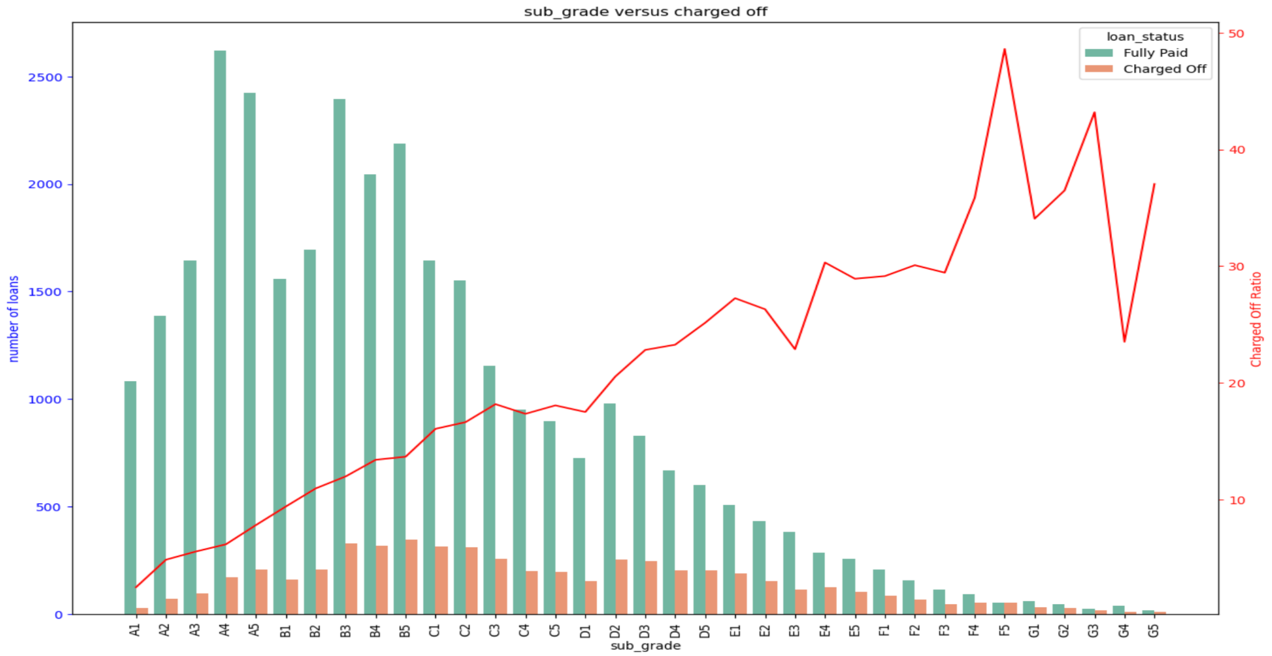
Segmented Univariate Analysis

- Maximum loan are given in the range of 9 - 13
- Minimum loans are given in range of 21-25



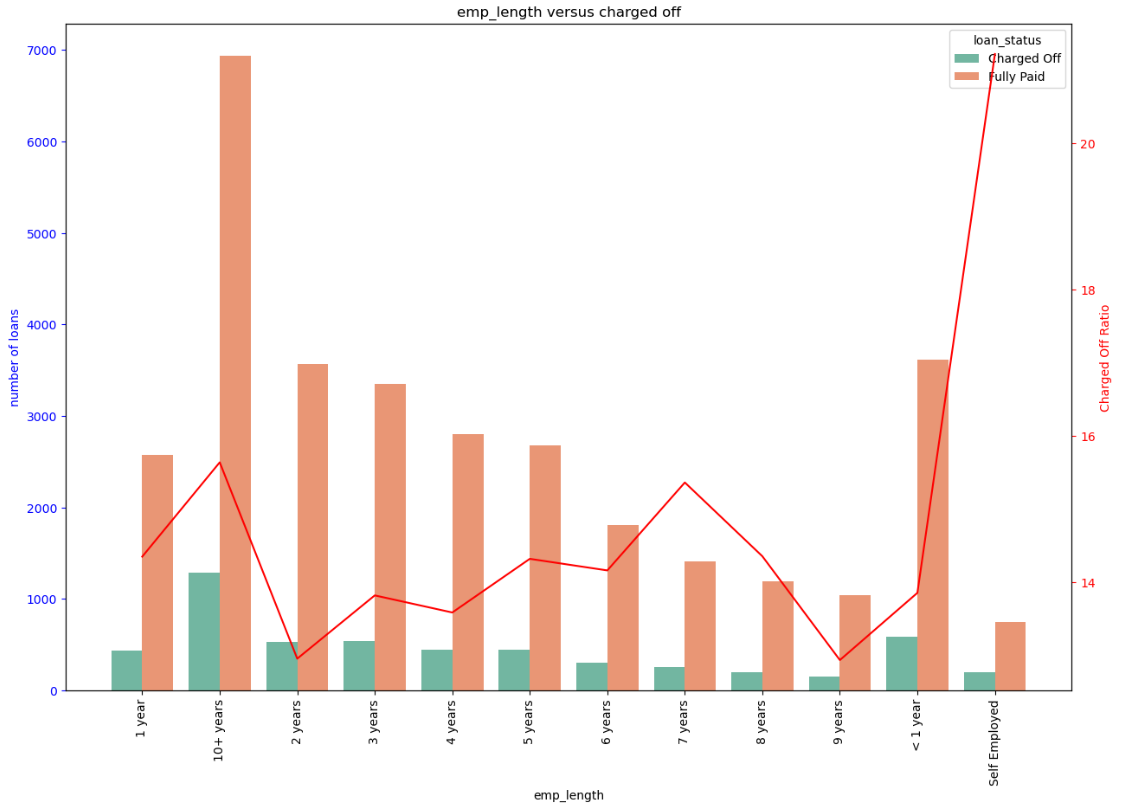
Bivariate Analysis

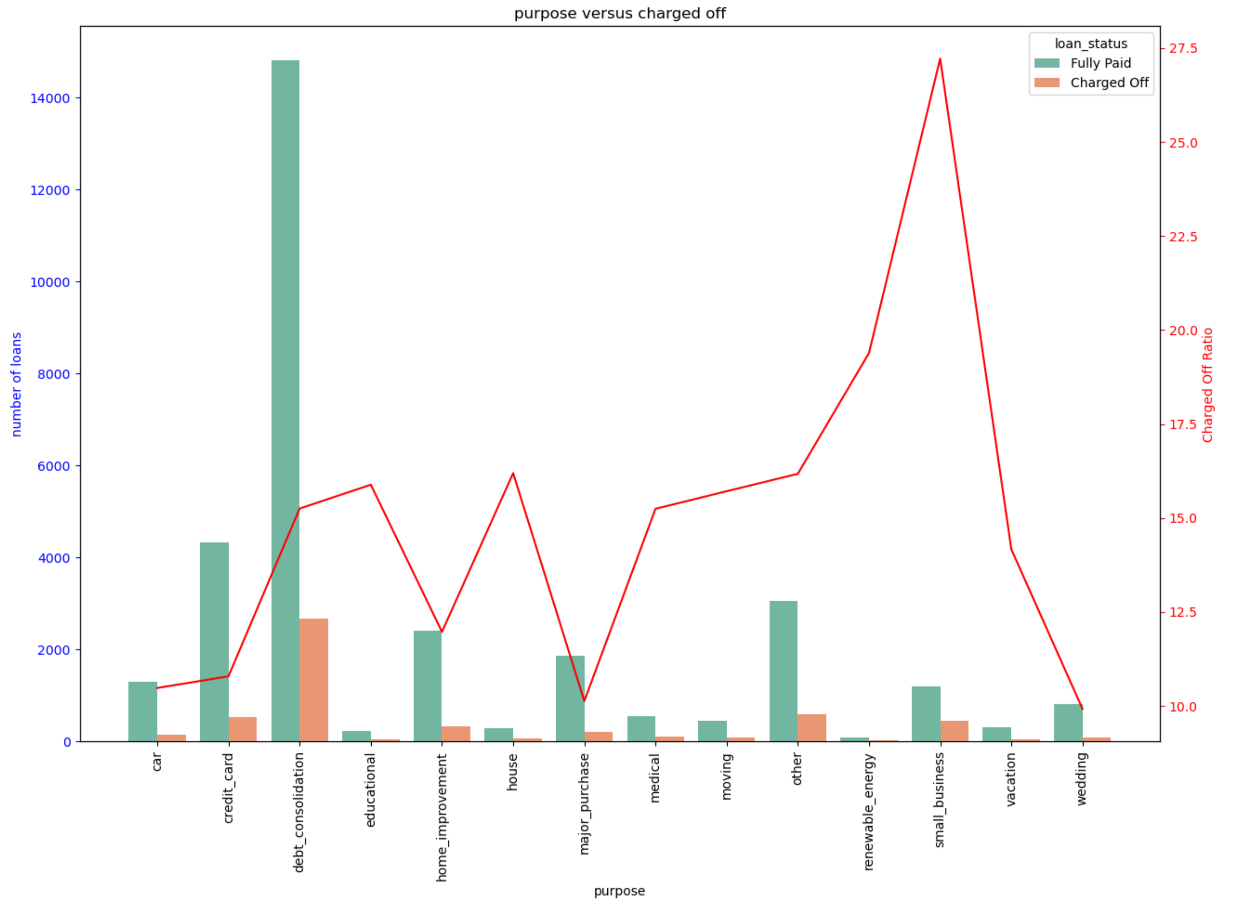




- Loan default tendency increases as the grade of the loan goes from A to G.
- Borrowers in subgrades F5, G3, and G5 exhibit the highest likelihood of loan default.

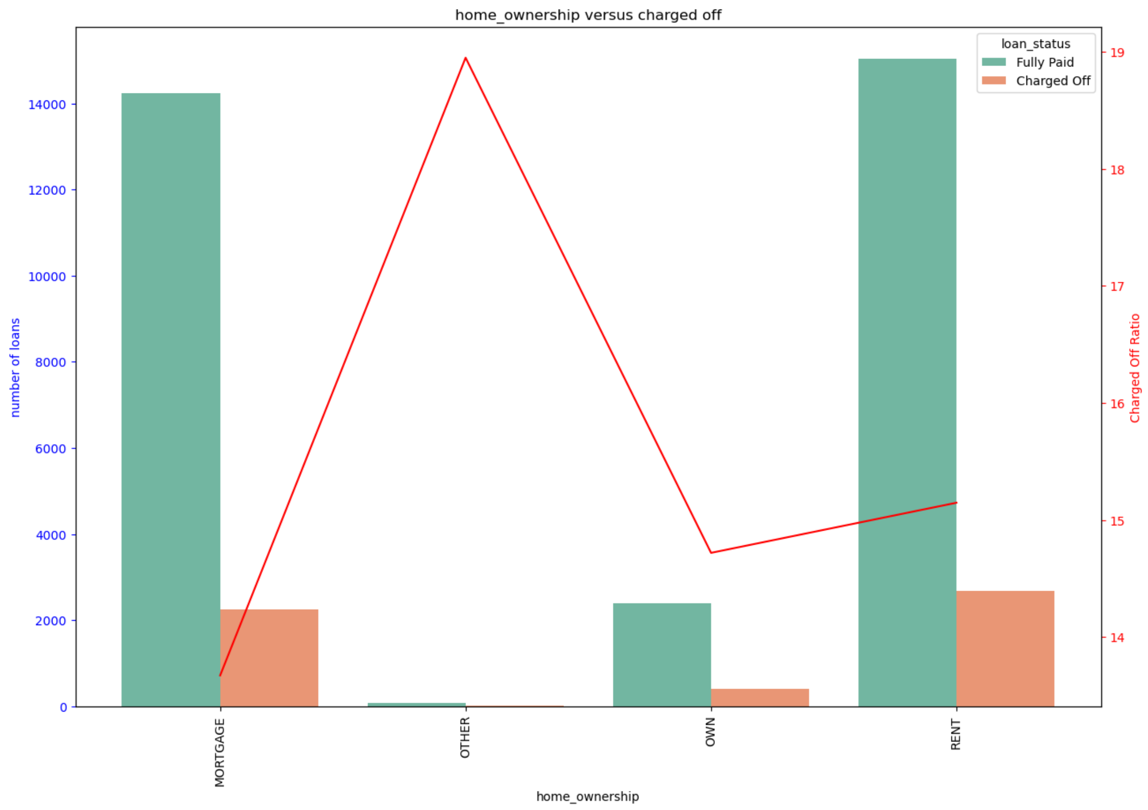
Bivariate Analysis

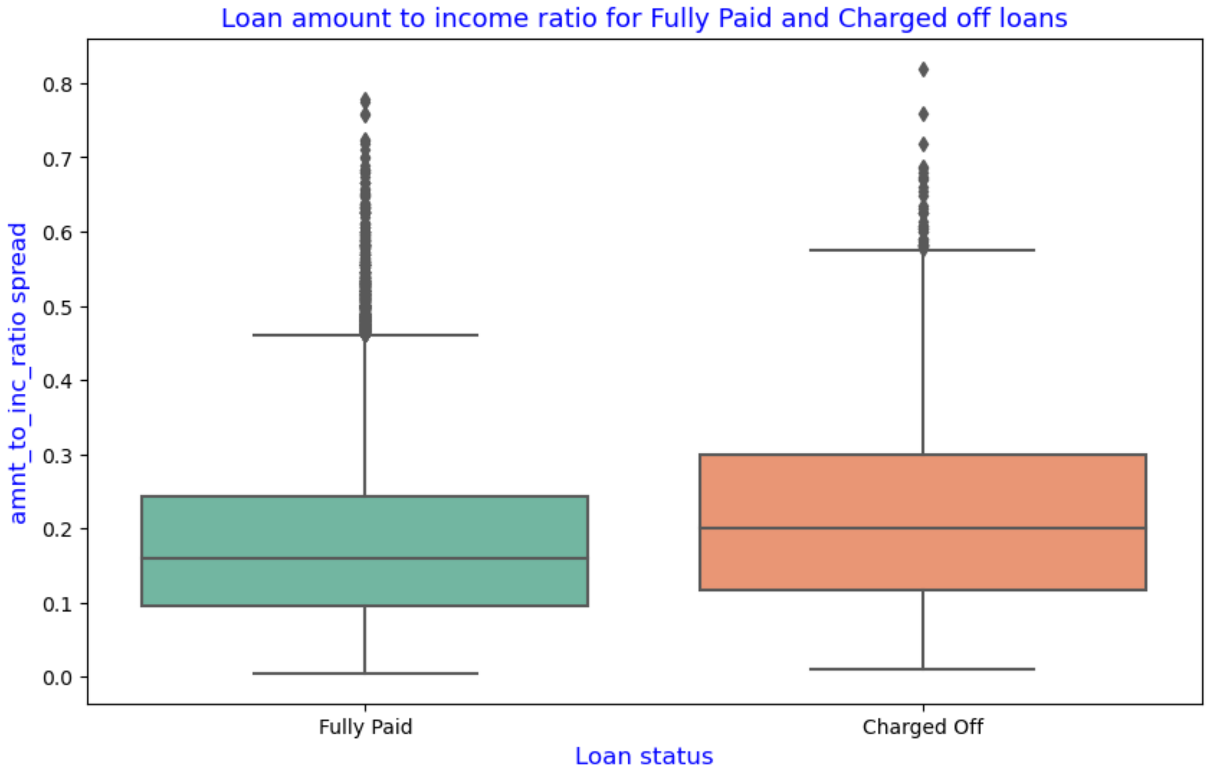




- Small business category borrowers have the highest chances of loan default.
- Self-employed borrowers, as well as those with 10 and 7 years of employment, have a higher tendency to default on loans.

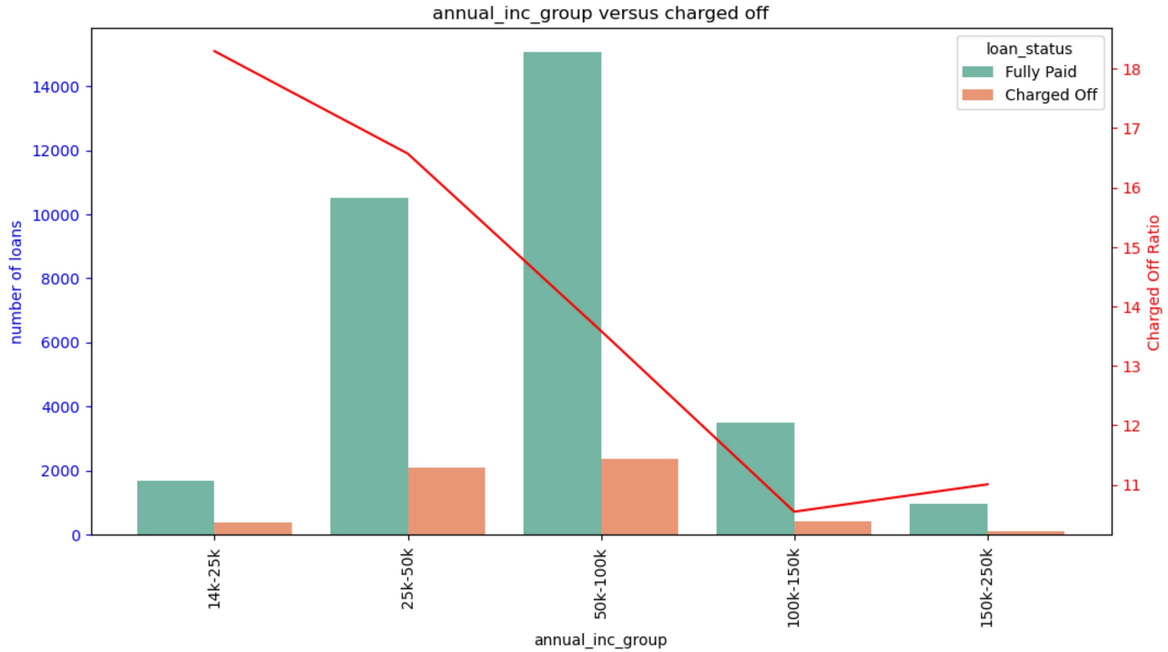
Bivariate Analysis

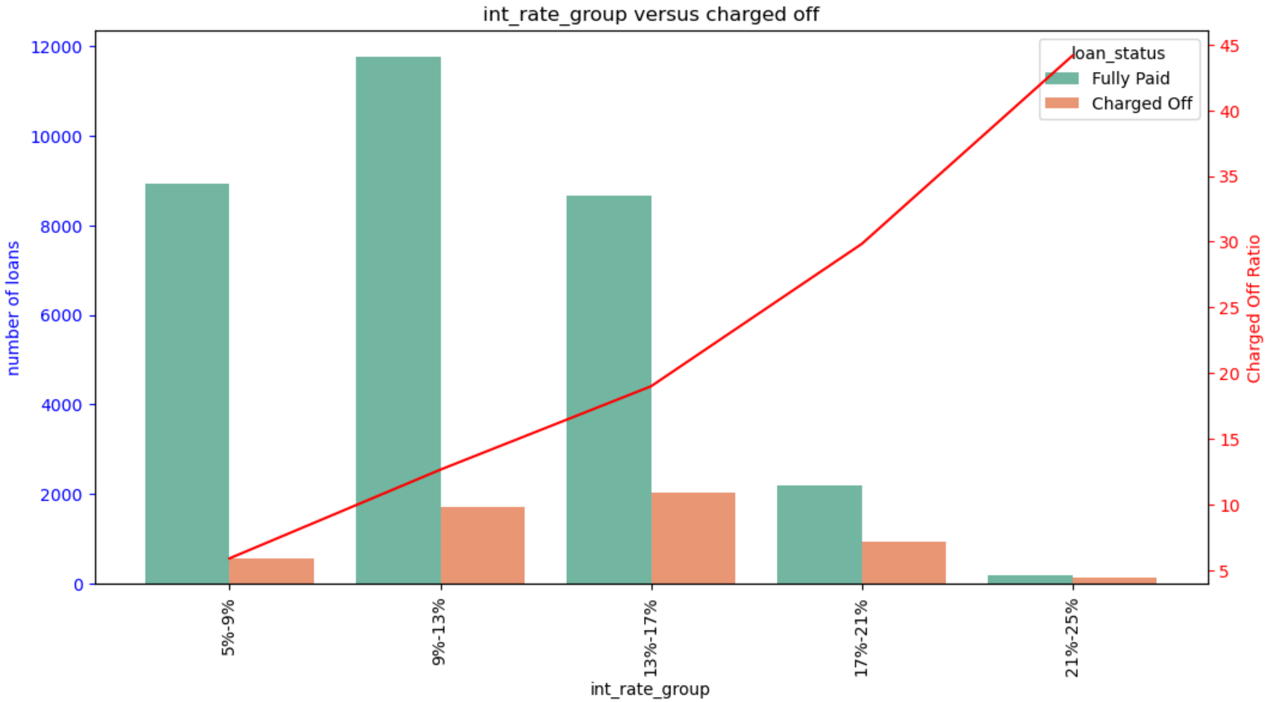




- If the loan amount is high with respect to income then the changes of Default is high.
- Other loan are more likely to be charged off
- Mortgage loan are less likely to be charged off

Bivariate Analysis

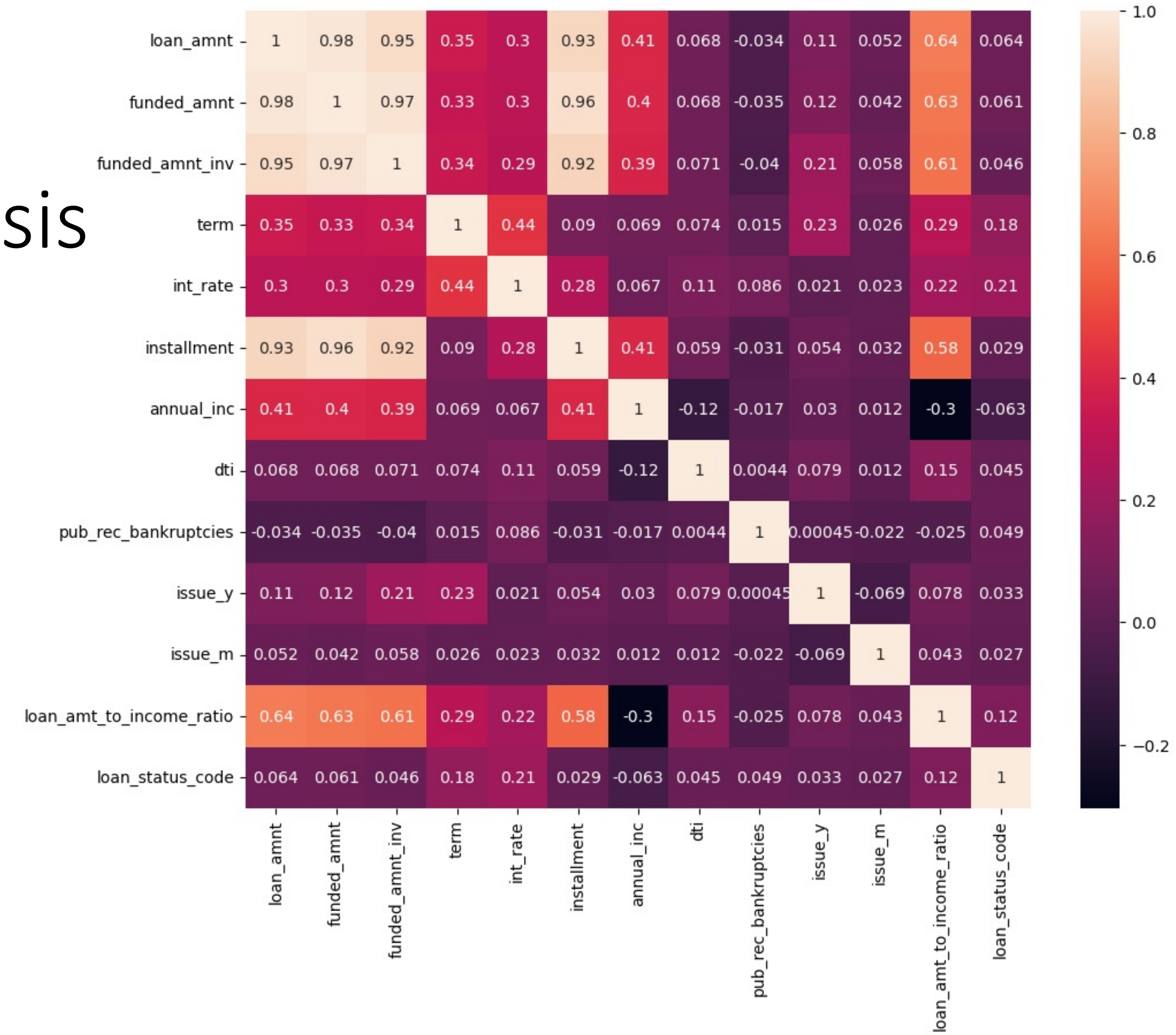




- Higher the Interest rate higher is the chance to default the loan
- Low income group are more likely to default the loan

Multivariate Analysis

- The annaul_inc, installment, issue_m, amnt_to_inc_ratio, loan_amt, funded_amt, issue_y have negative impact on the loan_status from the above correlation heatmap



Conclusion

- Loan default tendency increases as the grade of the loan goes from A to G.
- Borrowers in subgrades F5, G3, and G5 exhibit the highest likelihood of loan default.
- Self-employed borrowers, as well as those with 10 and 7 years of employment, have a higher tendency to default on loans.
- Borrowers from the states NV, TN, SD, AK, FL, and HI demonstrate the highest tendency to default on loans.
- Small business category borrowers have the highest chances of loan default.
- Borrowers classified as "Other" have the highest tendency to default on loans.
- Loan default tendency is higher among borrowers in lower income groups, decreasing as annual income increases.
- There is a positive correlation between the interest rate and loan default tendency.
- The correlation graph indicates that variables such as annual income, instalment, issue month, loan amount-to-income ratio, loan amount, funded amount, and issue year have a negative impact on the loan status variable.