Business Objectives

This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Like most other lending companies, lending loans to ‘risky’ applicants is the largest source of financial loss (called credit loss). Credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed. In other words, borrowers who default cause the largest amount of loss to the lenders. In this case, the customers labelled as 'charged-off' are the 'defaulters'.

If one is able to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, the company wants to understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

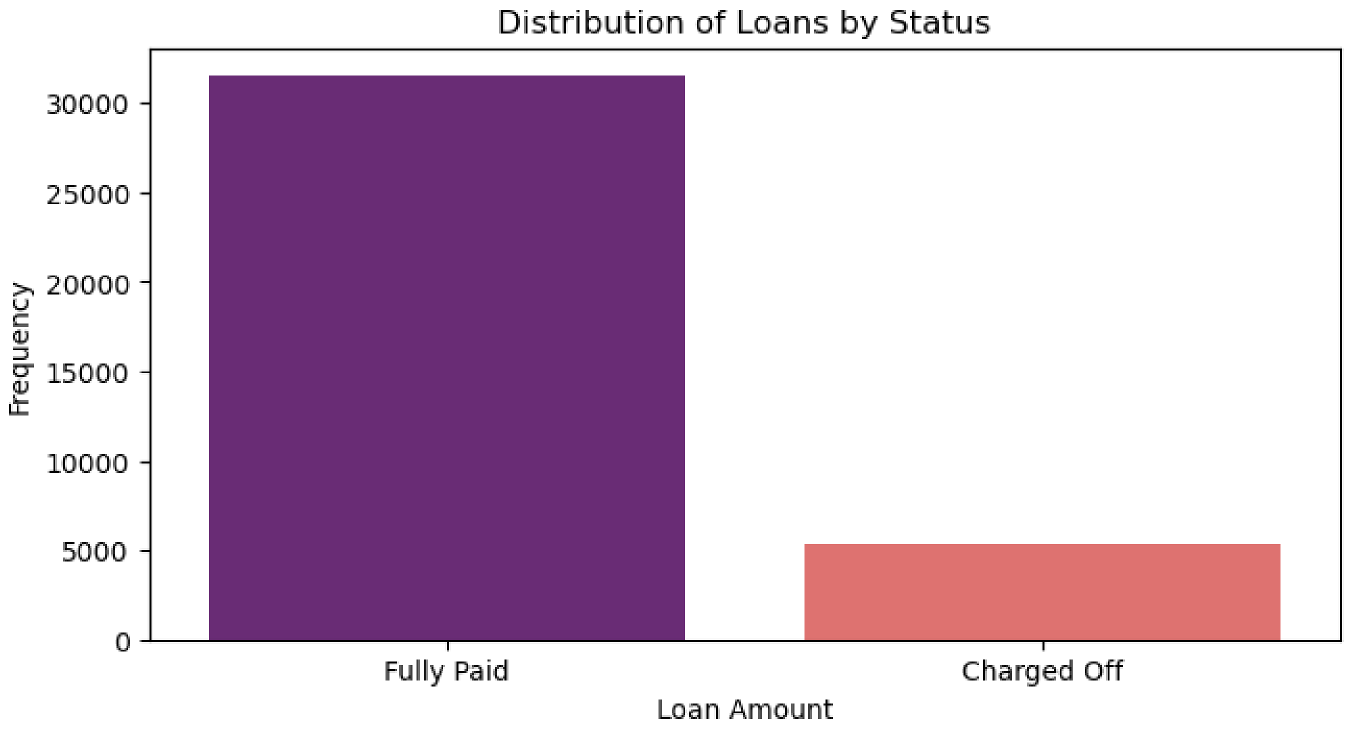
To develop your understanding of the domain, you are advised to independently research a little about risk analytics (understanding the types of variables and their significance should be enough).

Prepared By:

Chetan Rajhans
Ashutosh Kumar

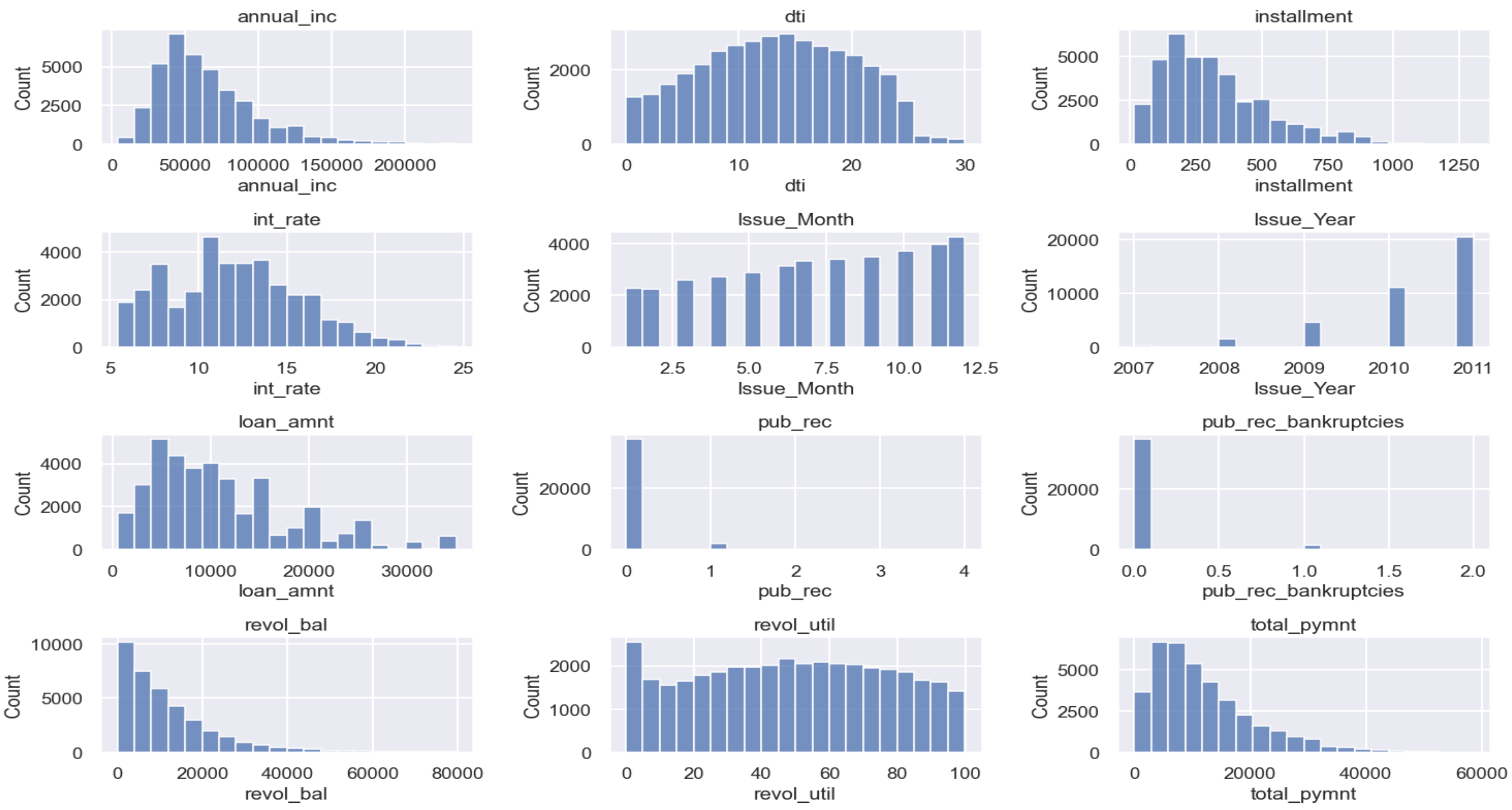
Univariate Analysis

Distribution of Loan Status

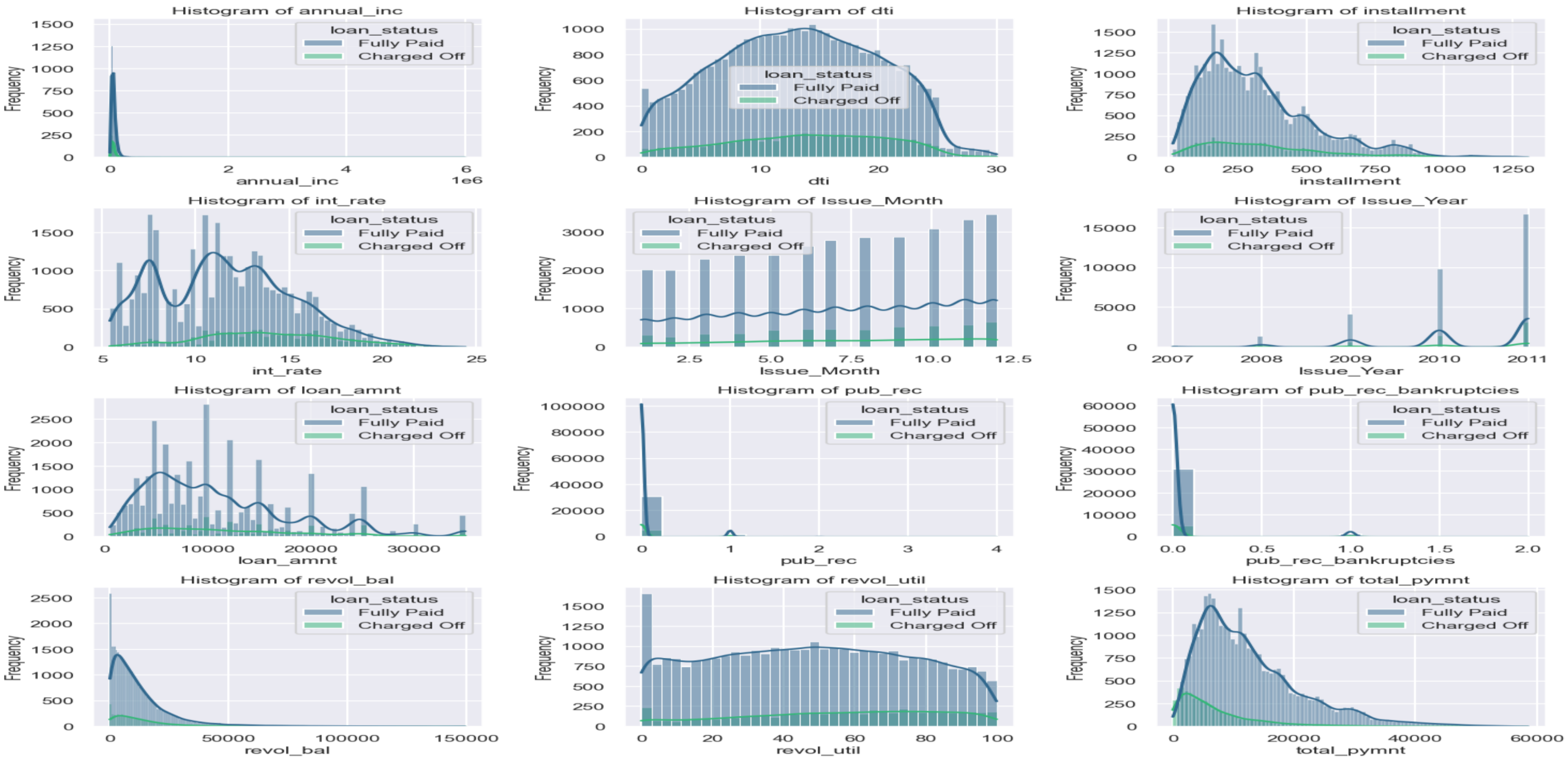


Around 15% of loan have gotten charged of based on data we have

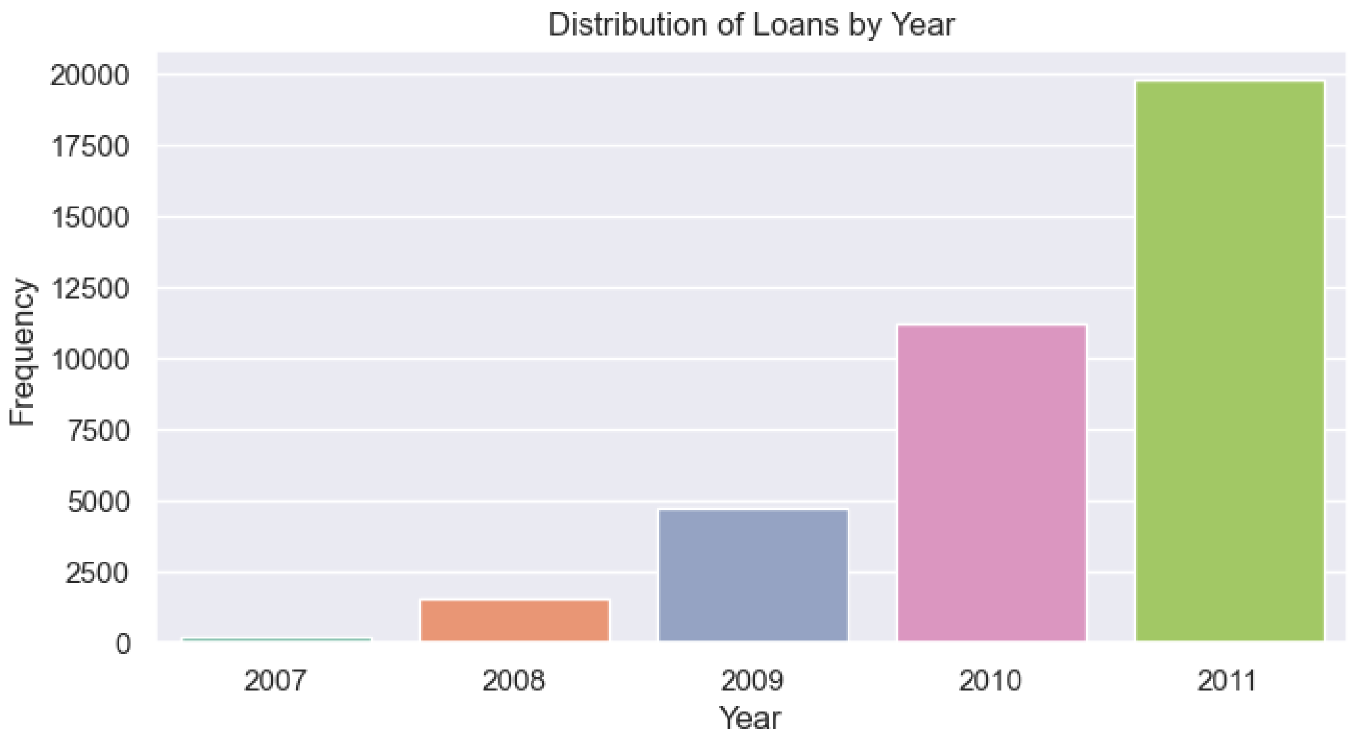
Univariate distribution of loans across key metrics



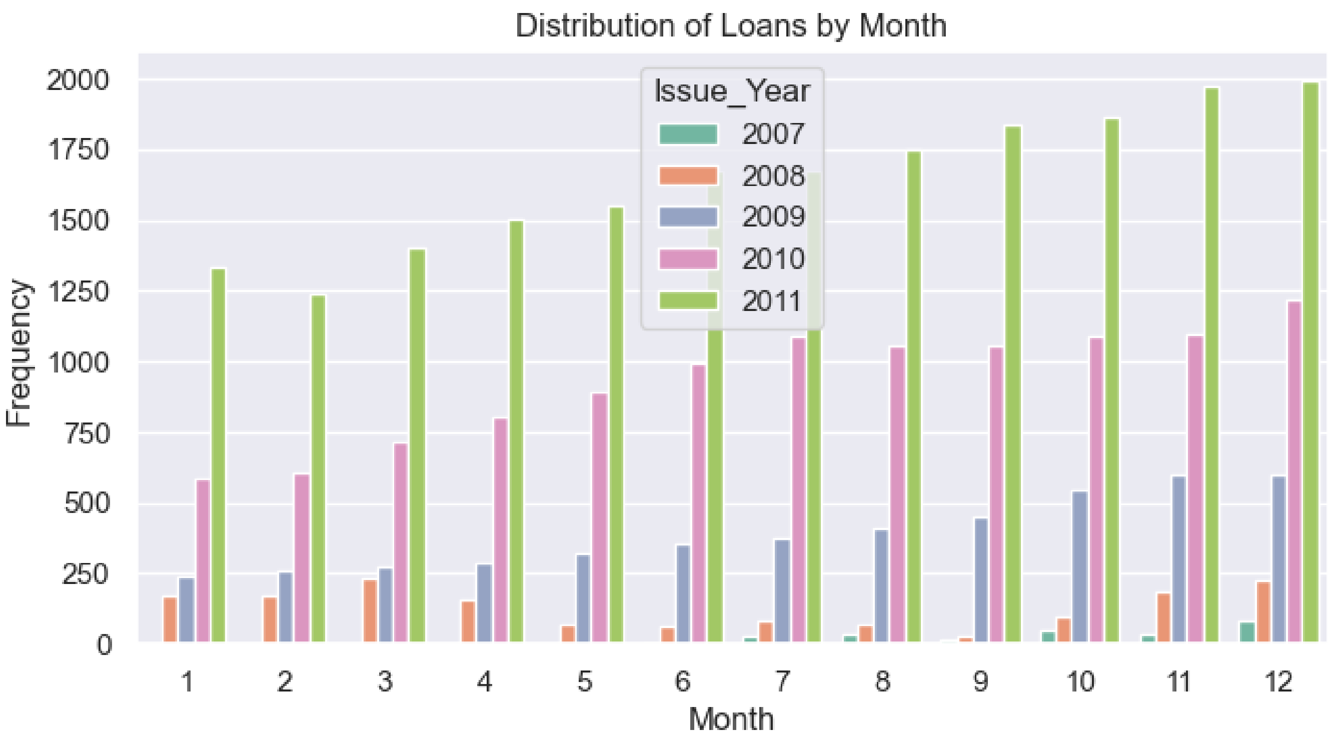
Loan disbursement and loan default across key metrics



Number loans have increased year on year

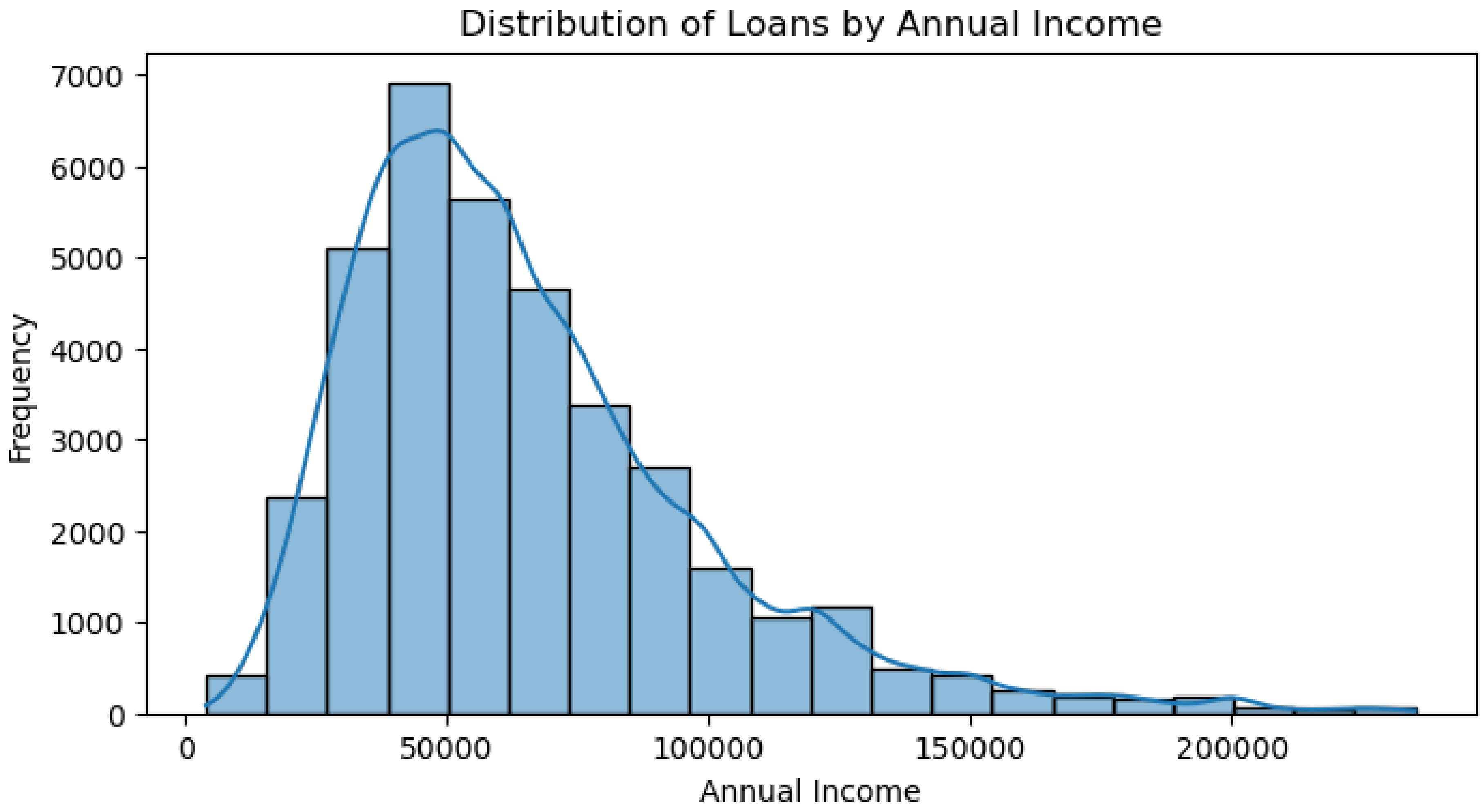


The number of loans increases month-over-month, with the exception of the year 2008. This coincide with the period of financial crisis

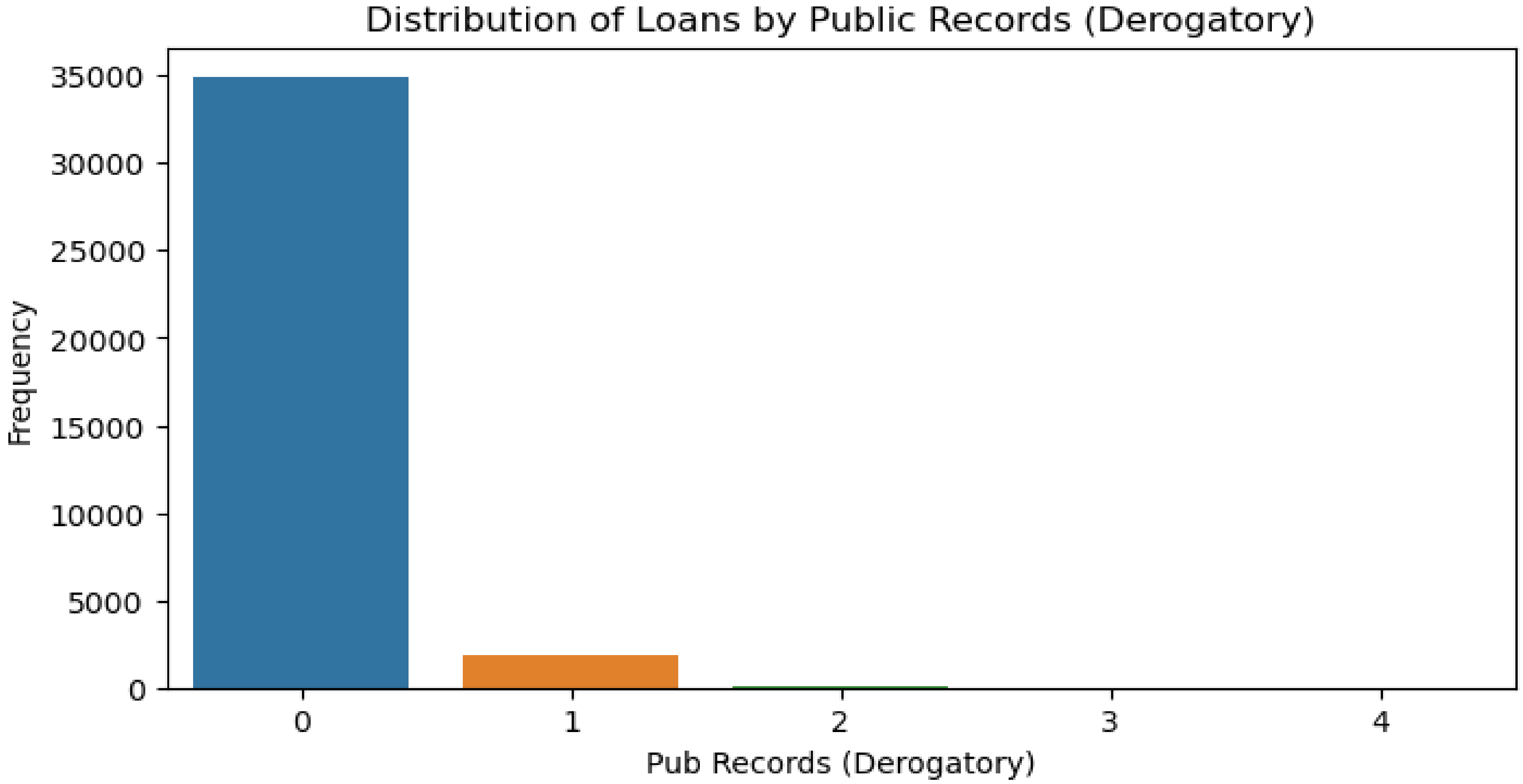


Both loan amount and loan default increase over the year with highest recorded in December of every year signifying increasing lending activities around Christmas holidays

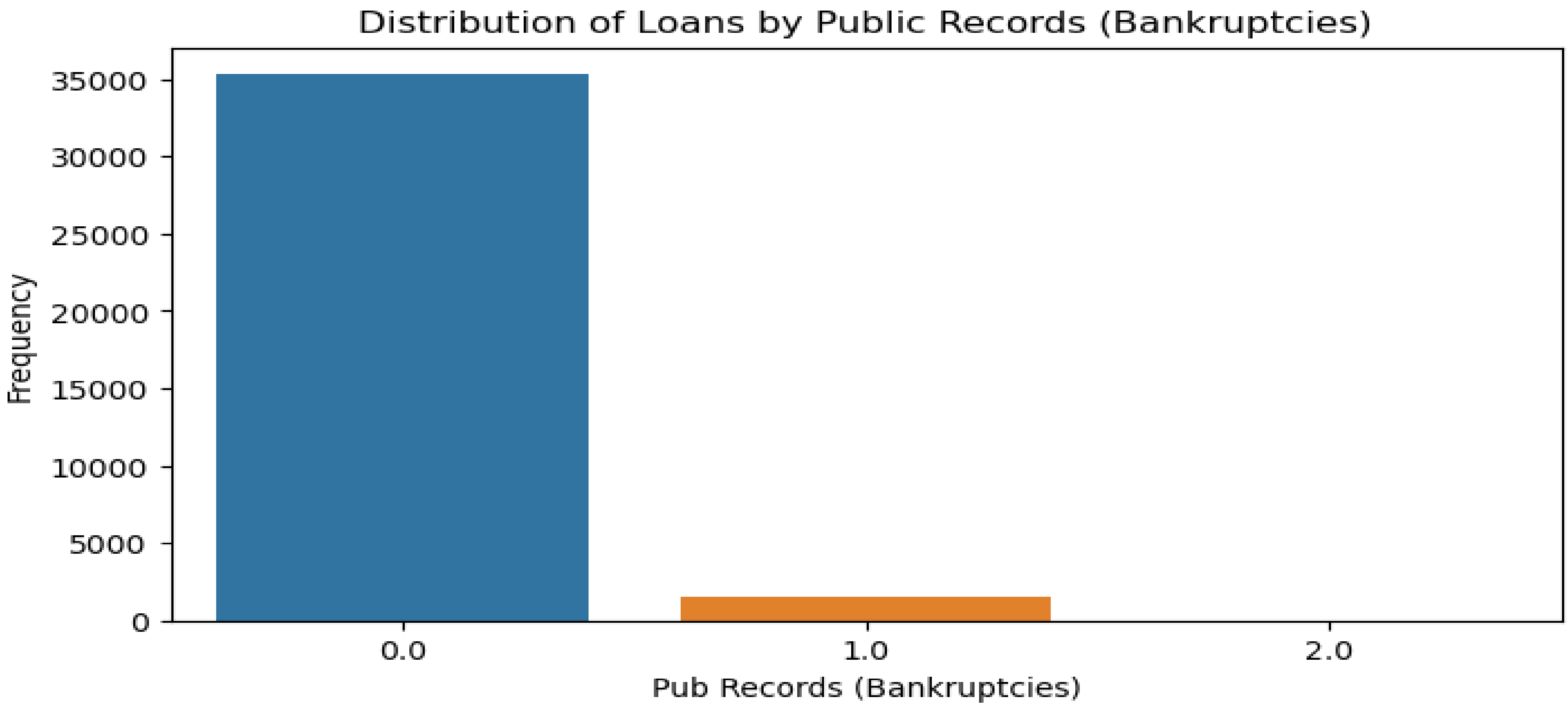
The majority of loans are from individuals within the low income range of less than 60,000.



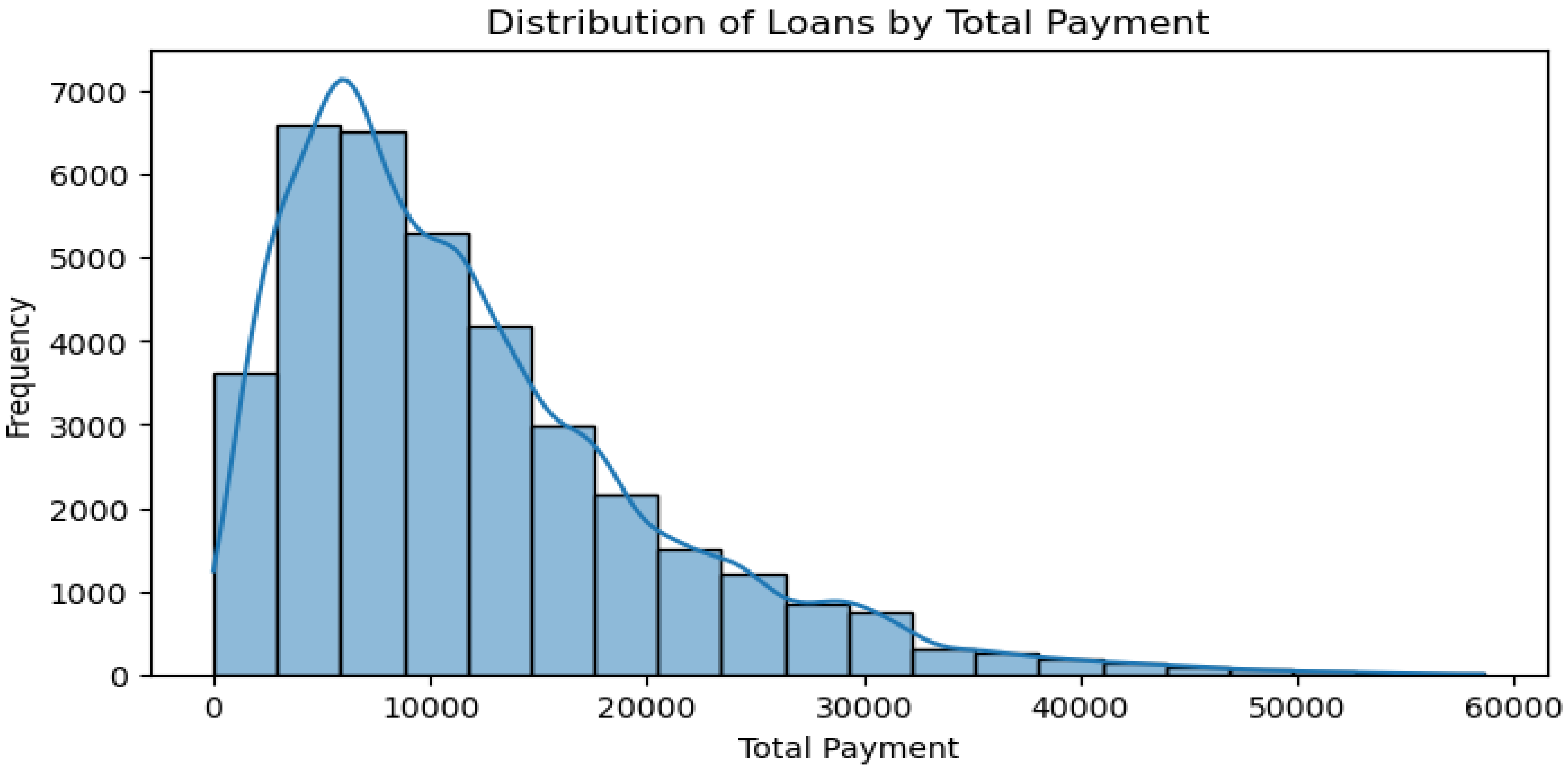
The majority of borrowers do not have public records of derogatory.



The majority of borrowers do not have public records of bankruptcies.



Peak of total payments is <10000

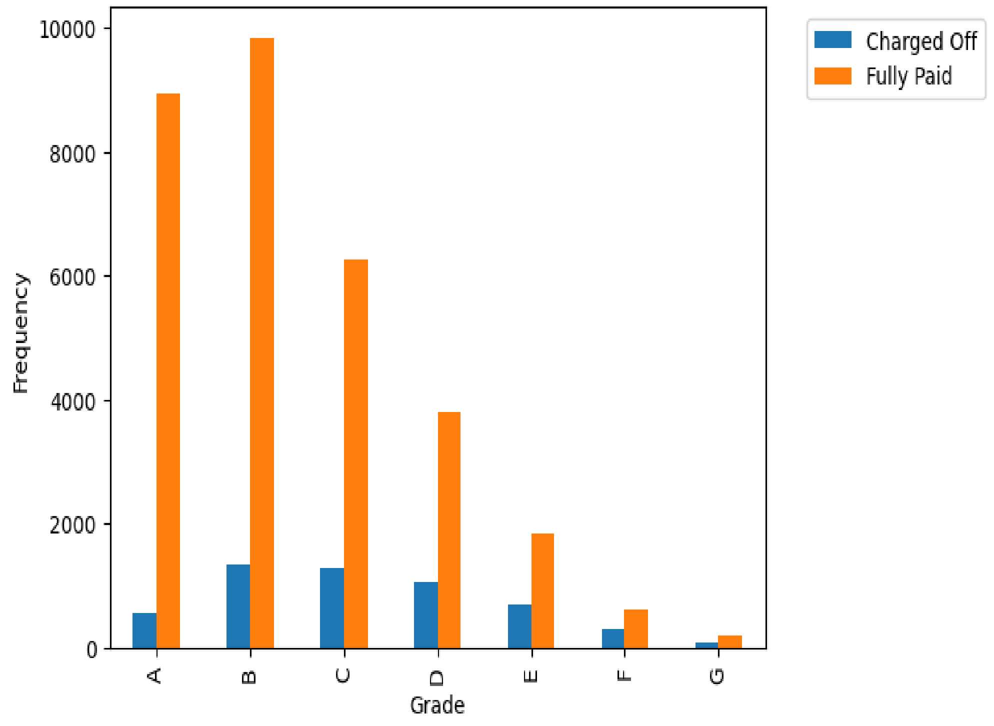


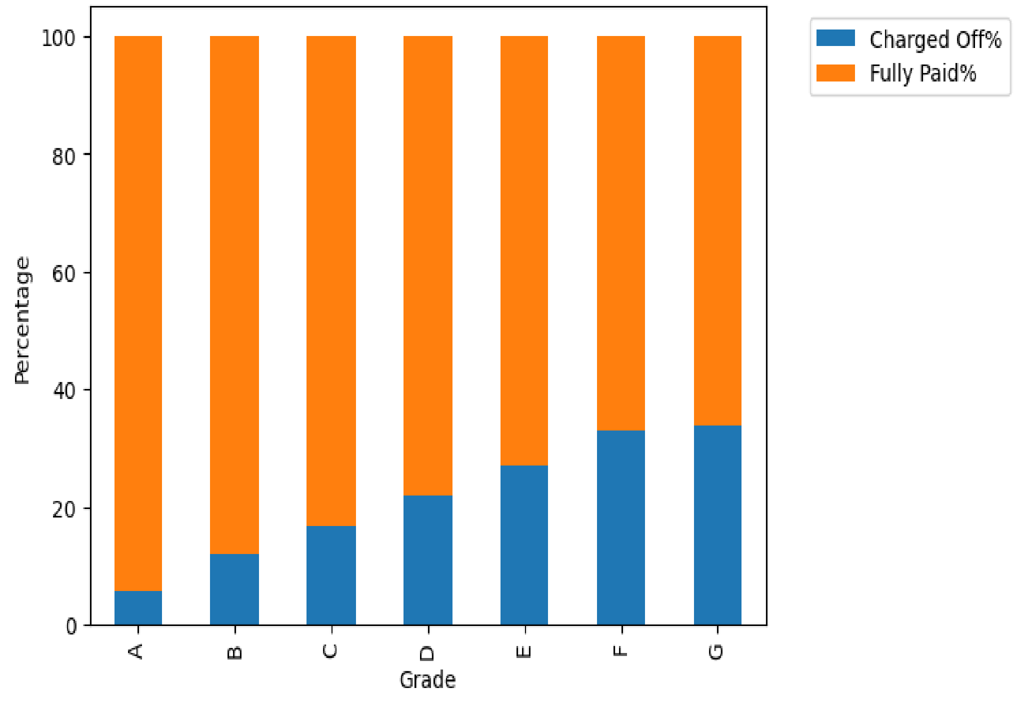
Univariate Analysis Summary

- The majority of loans are fully repaid.
- Most loans fall within the 5,000-15,000 range.
- Interest rates typically range between 5-7.5% and 10-15%.
- Individuals with lower incomes tend to have more loans.
- Loans are generally granted to individuals with no public record incidents.
- The number of loans has increased year over year.
- There has been a month-over-month increase in loans, with exceptions in 2007 and a slight dip in 2010.
- A large number of loans are taken out by individuals in grades A and B.
- A short loan tenure of 36 months is preferred.
- Employees with over 10 years of experience tend to have more loans.
- Most borrowers do not own a home.
- People residing in larger cities tend to have more loans.

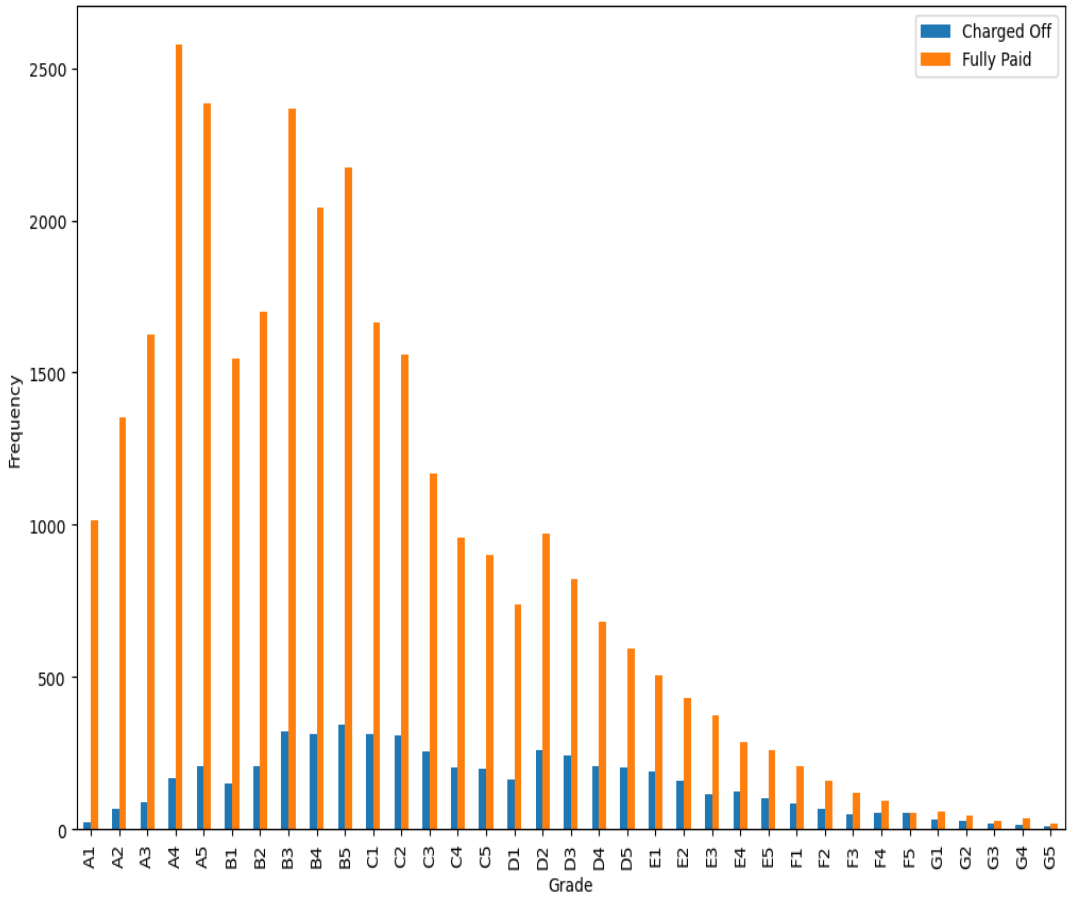
Bivariate Analysis

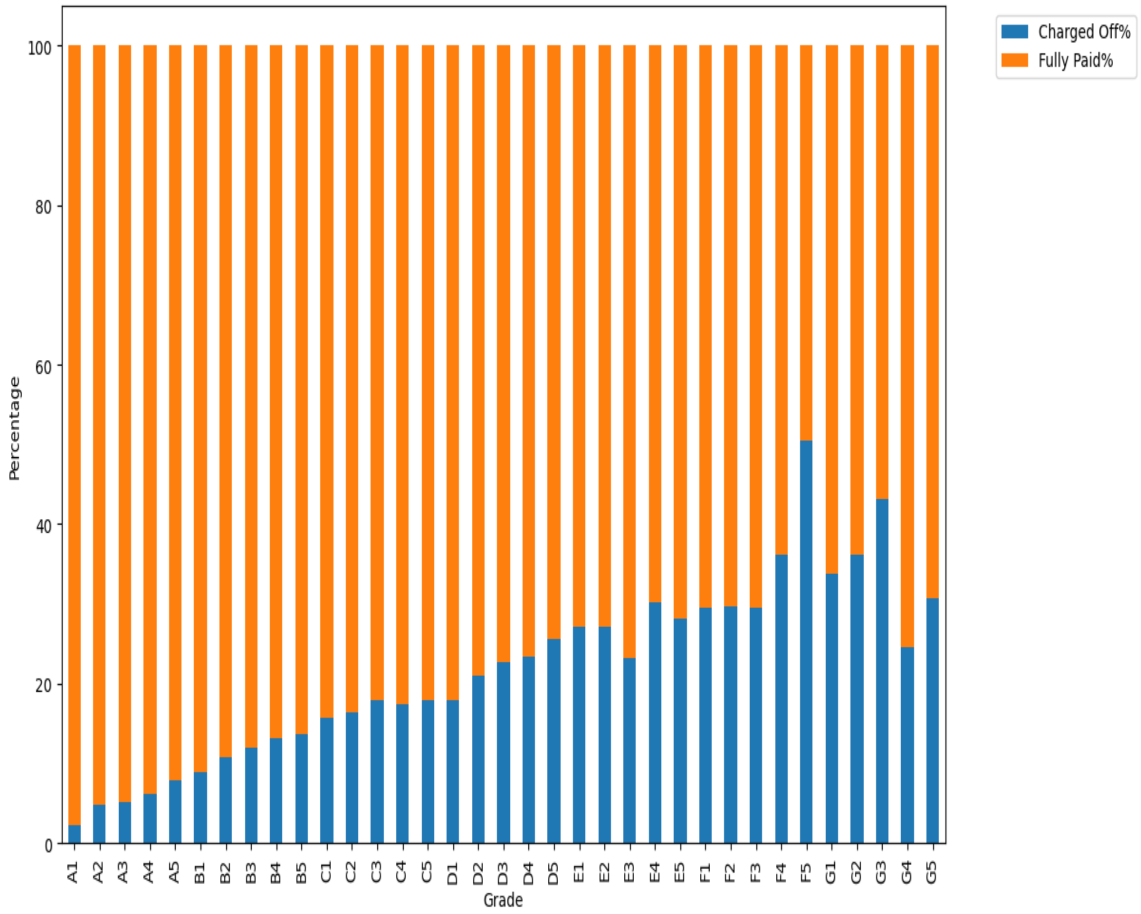
Majority of loans are in Garde A and B but default rates are higher for grades G and F



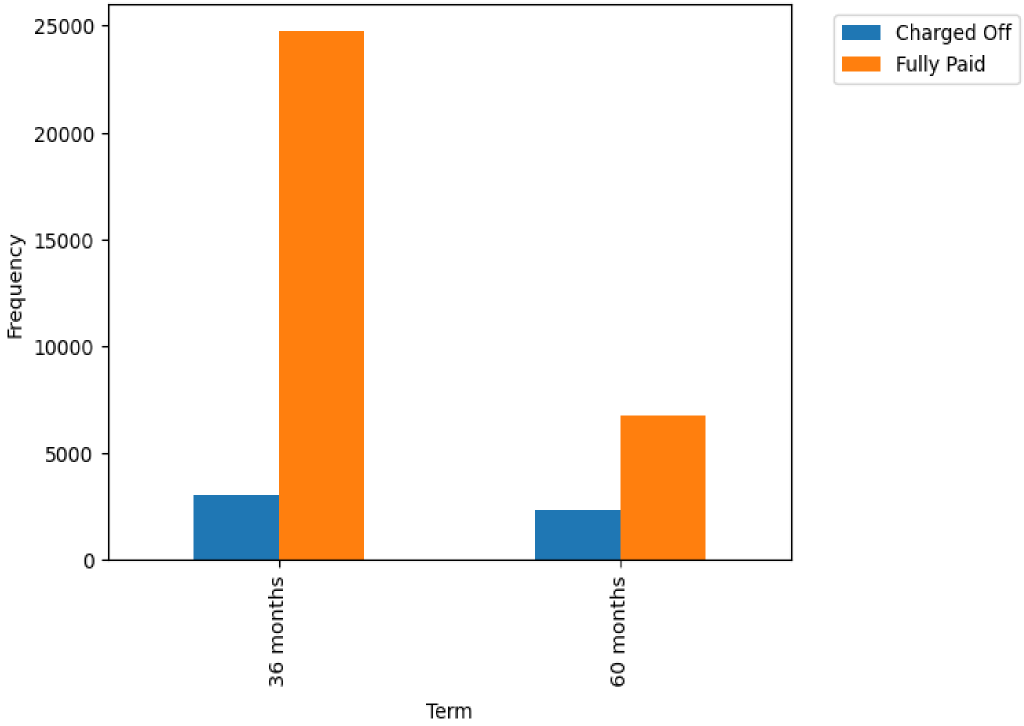


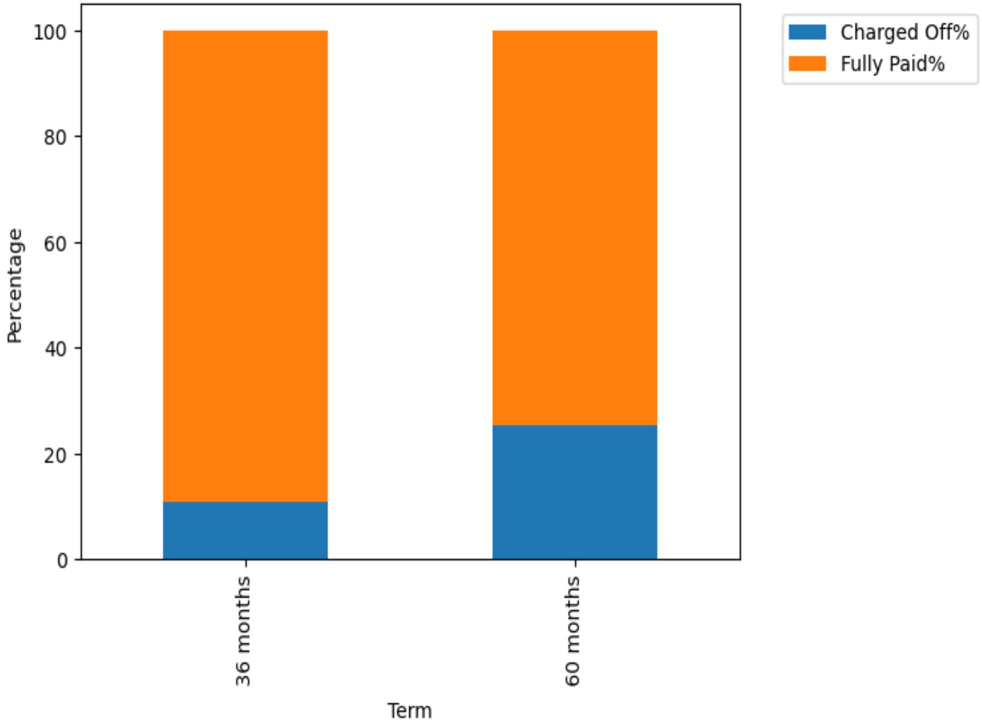
Among sub grades F4 and G3 have the highest default rates



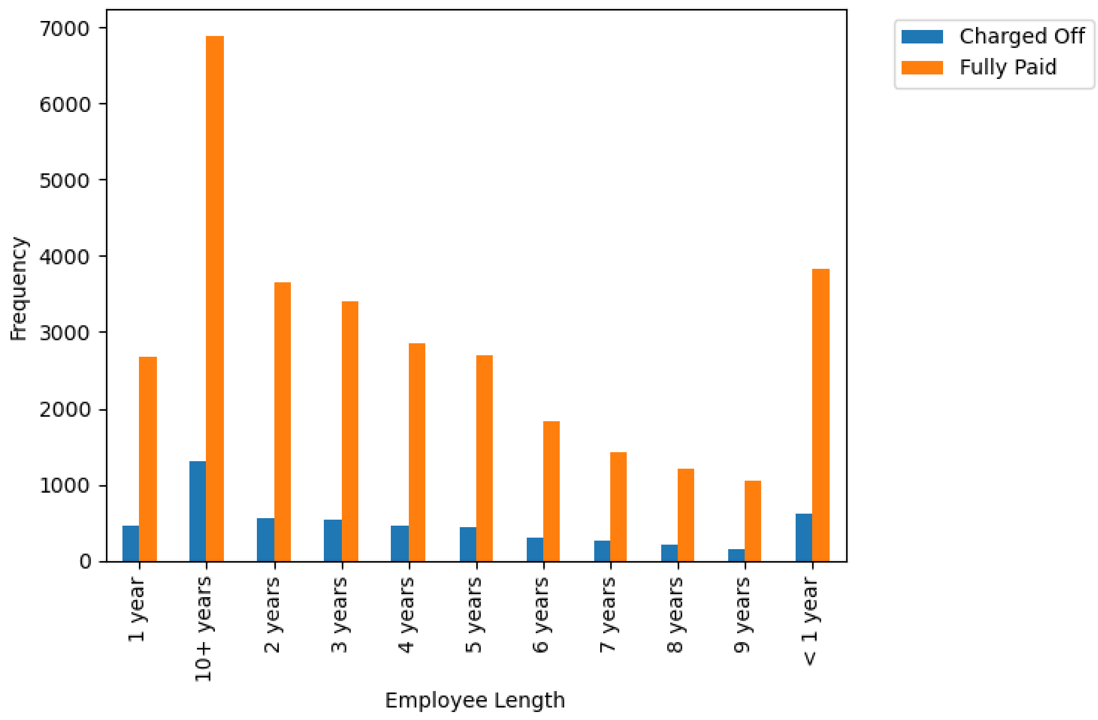


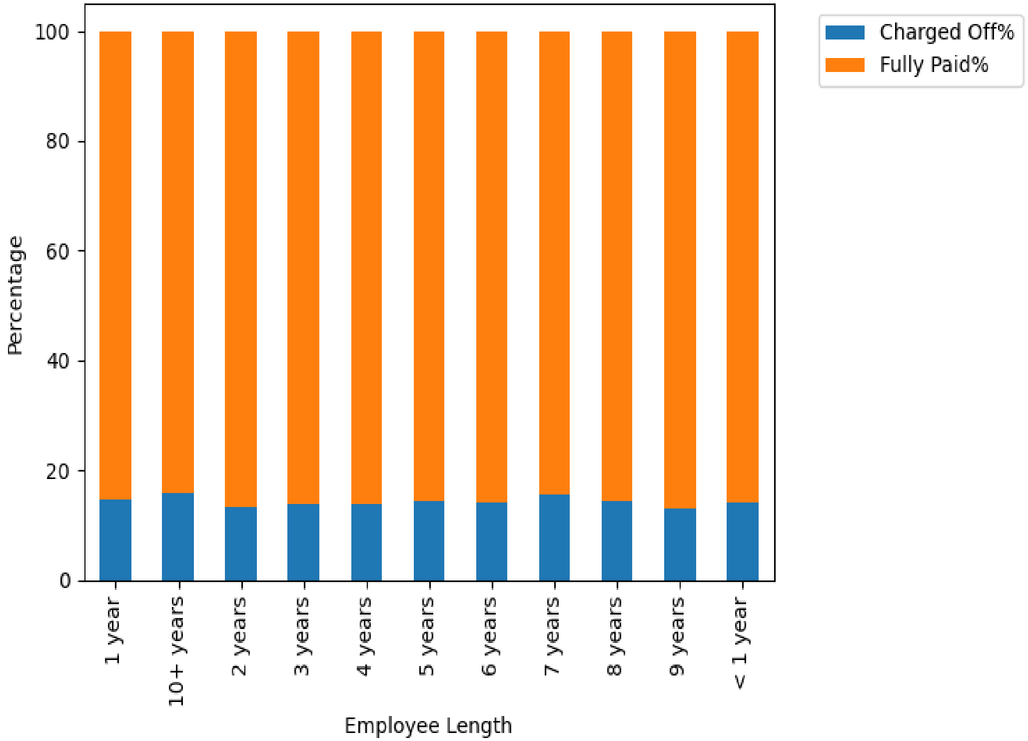
Loans are borrowed mostly for 36 months but the default rate is higher for 60 months duration



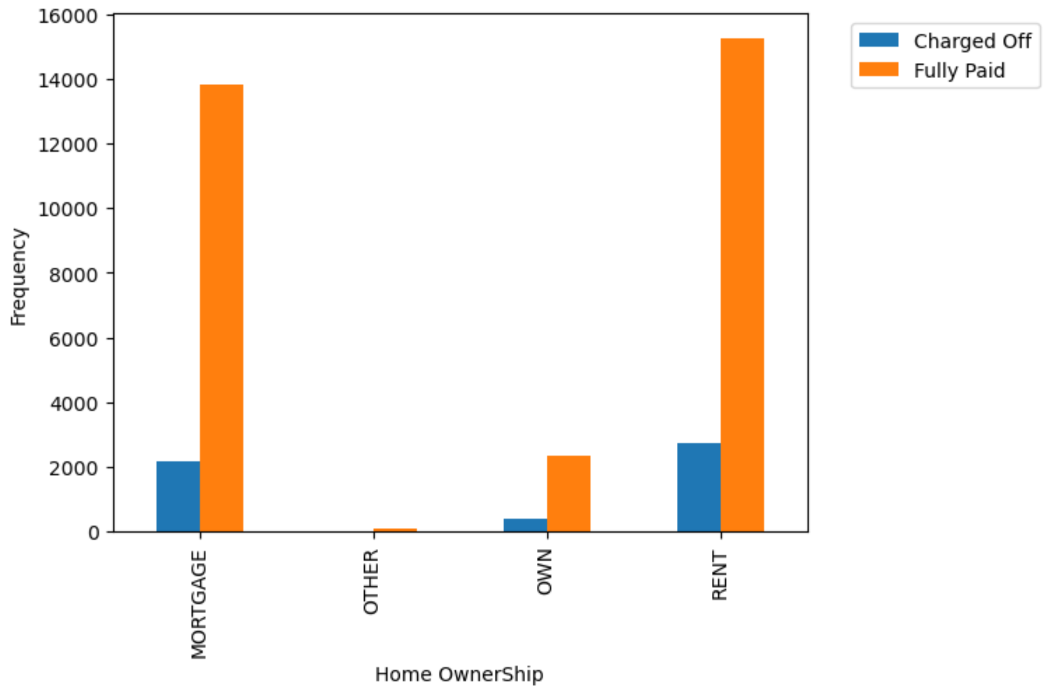


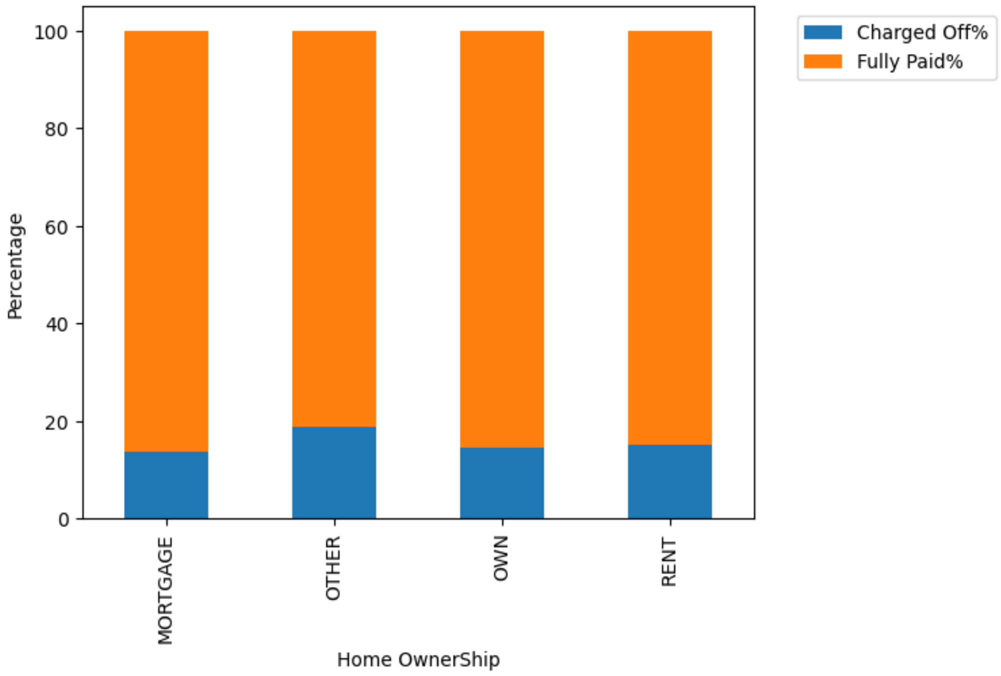
Employment with length of 10+years have more loans and default rate high too



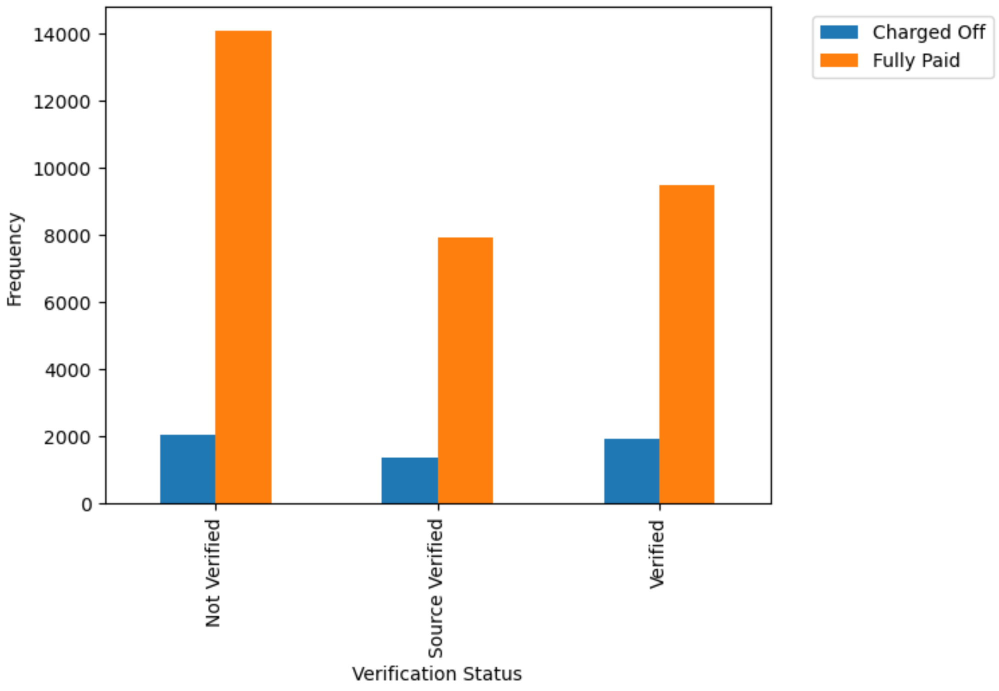


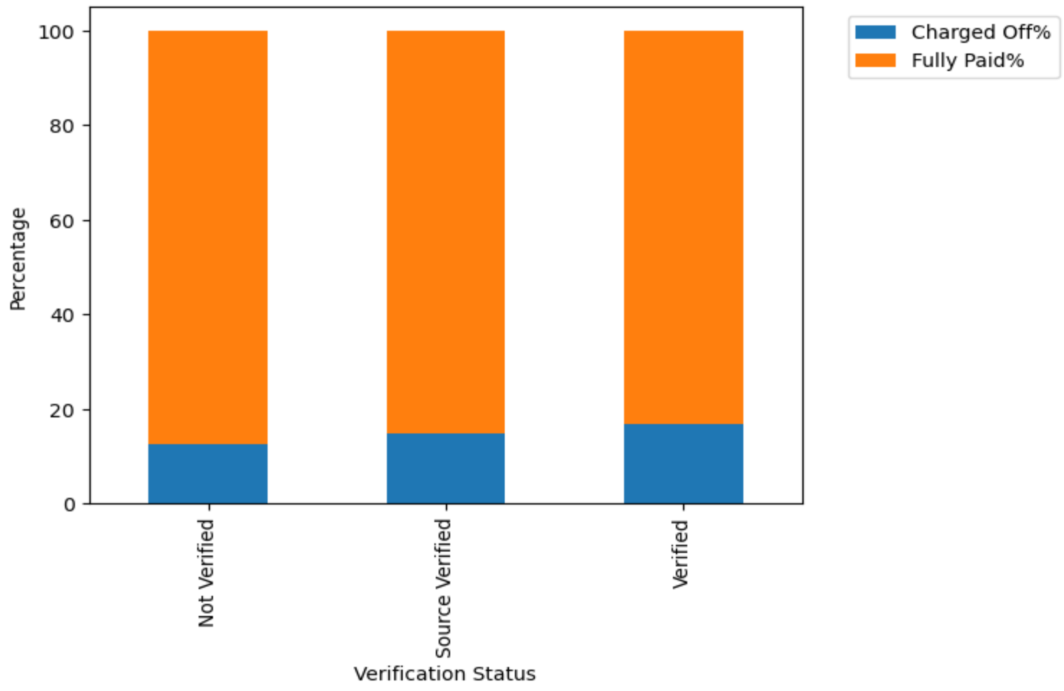
The majority of borrowers borrow it for rent but the default is highest for Mortgage



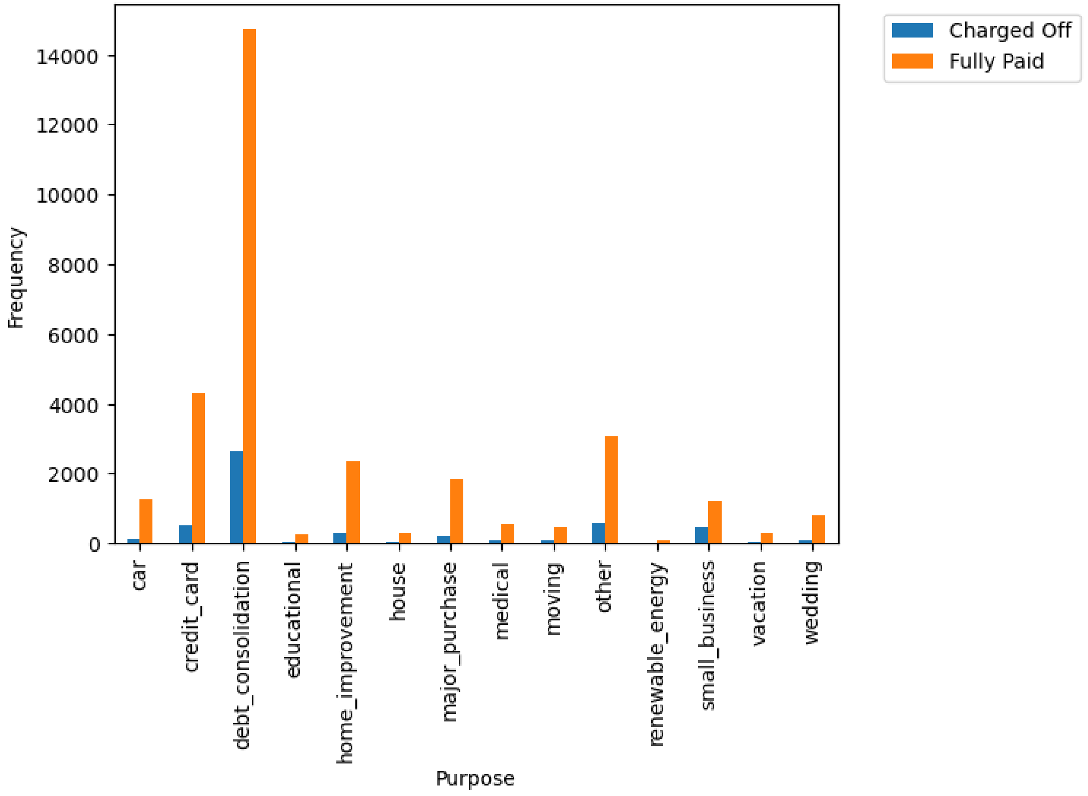


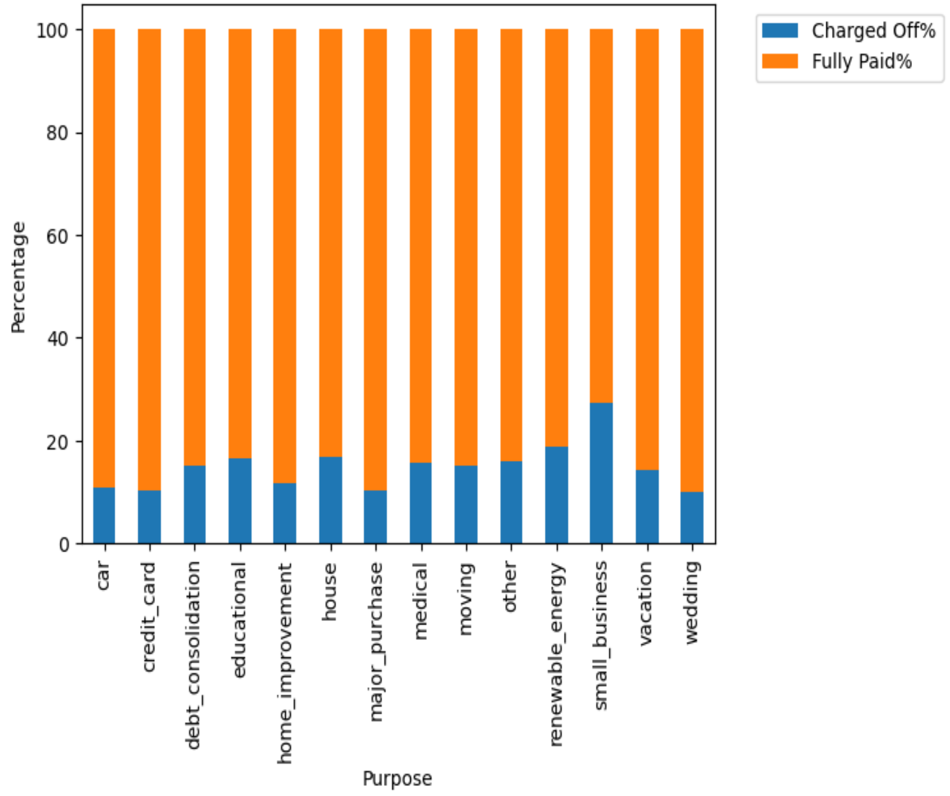
The majority of loans are Not verified either by the company or the source but the verified ones have a higher default rate



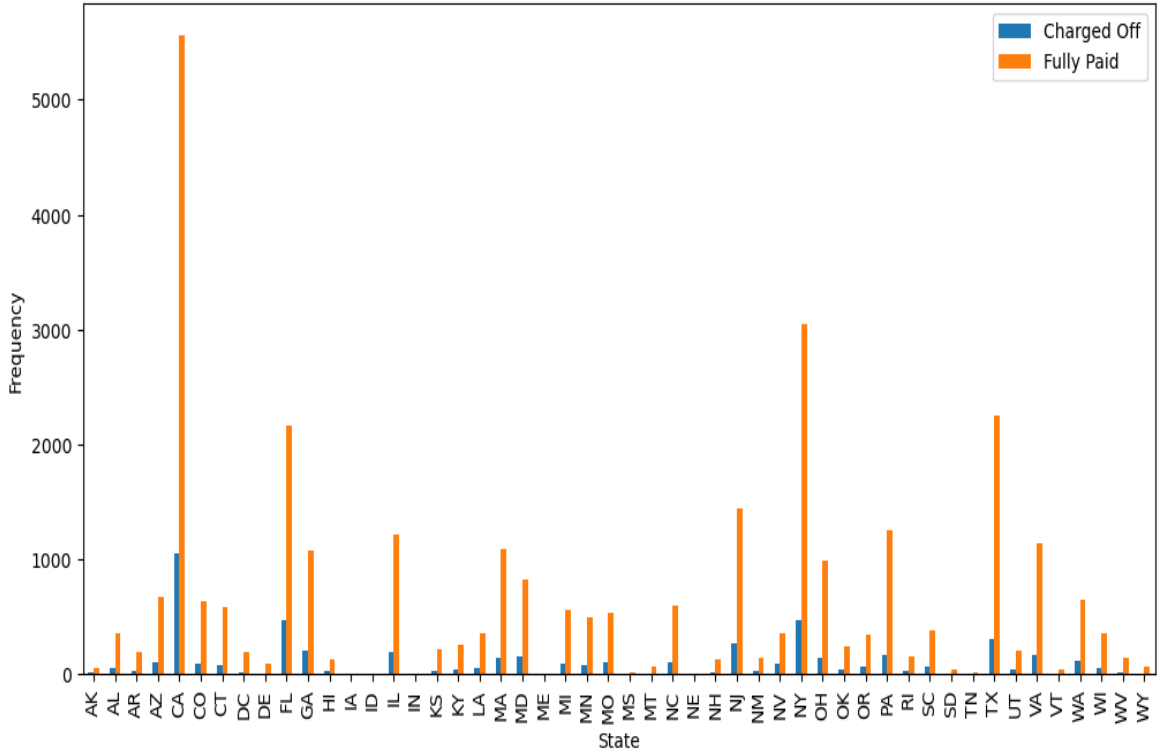


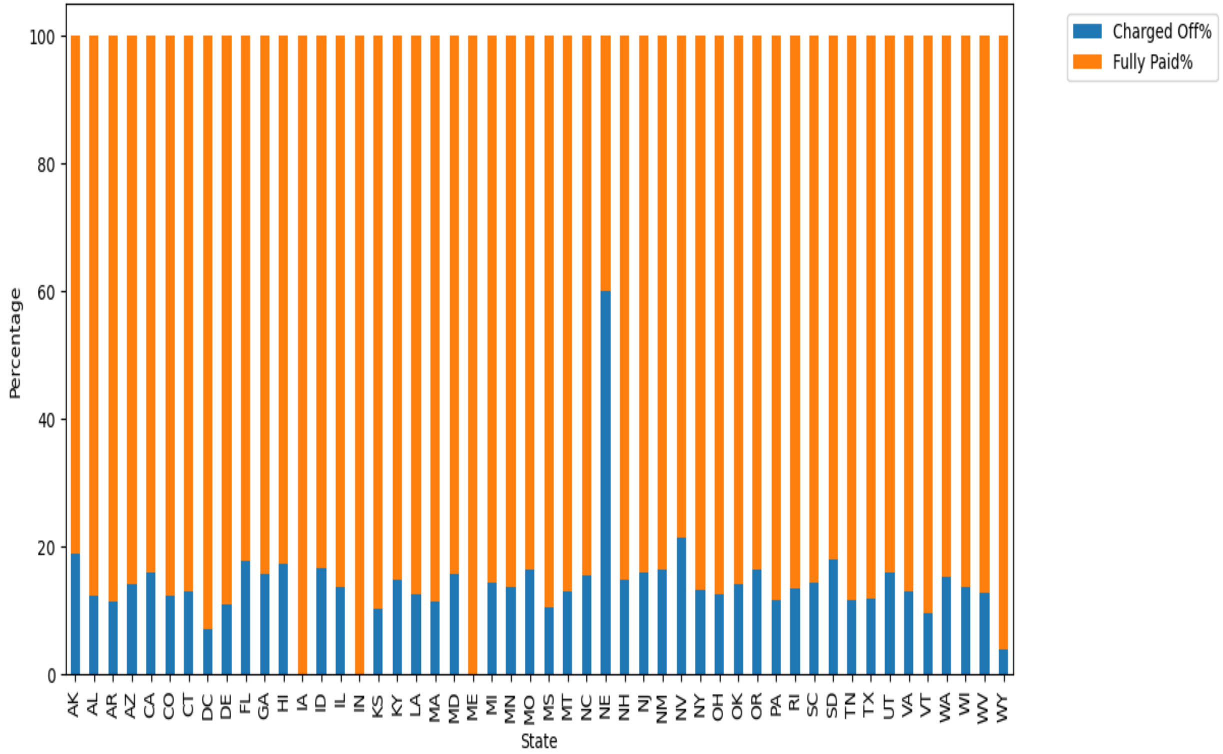
Top reason for loan are debt consolidation and credit card but renewable energy and small_business have a higher default rate



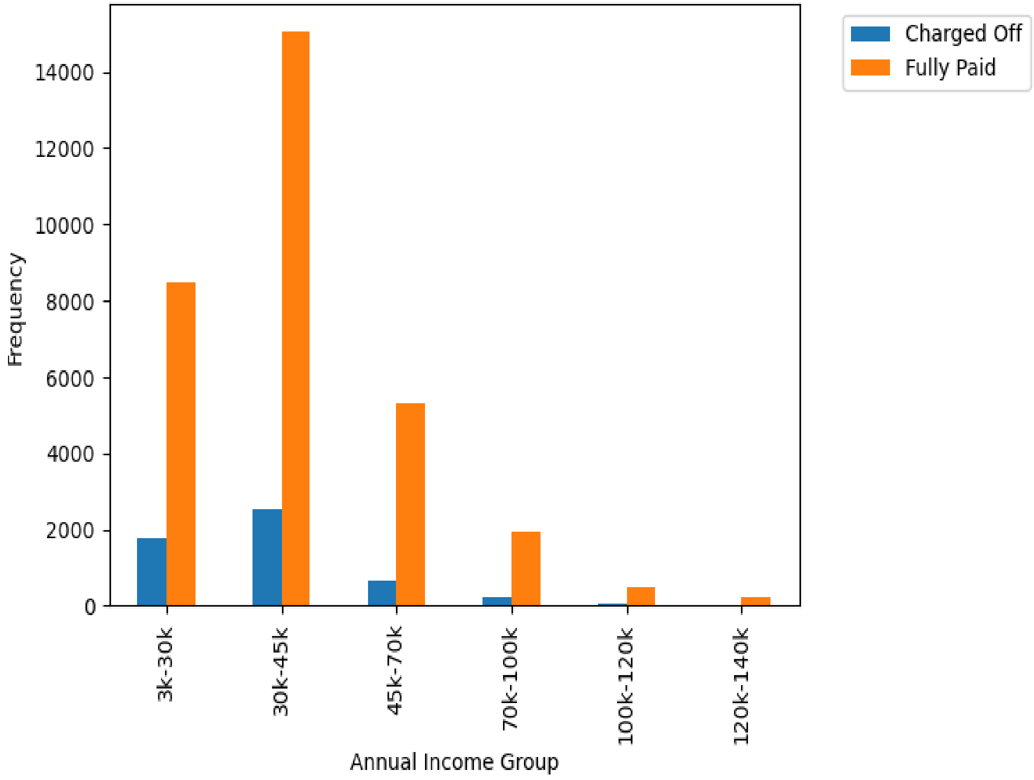


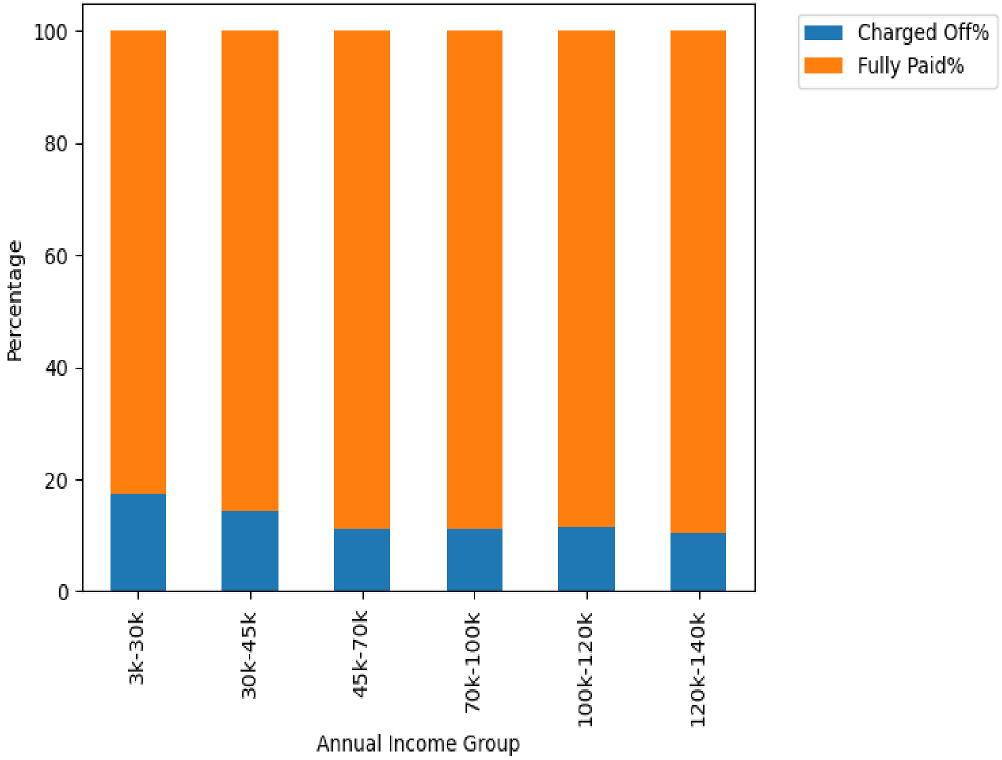
CA, NY, FL, TX reports higher loan with the default rate highest for NE



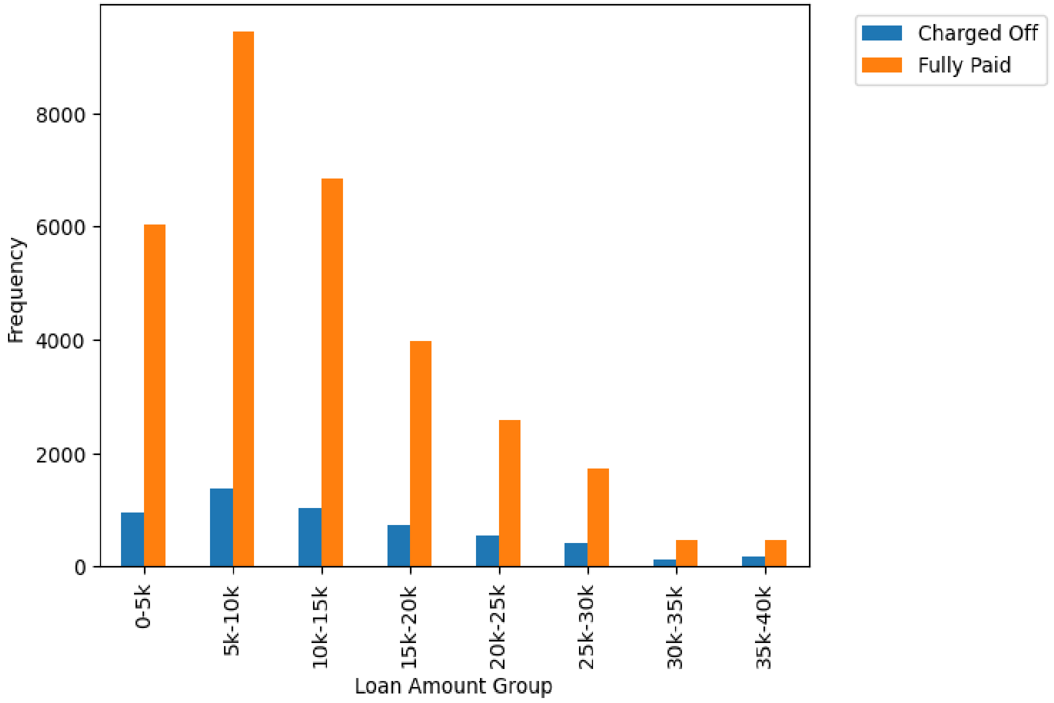


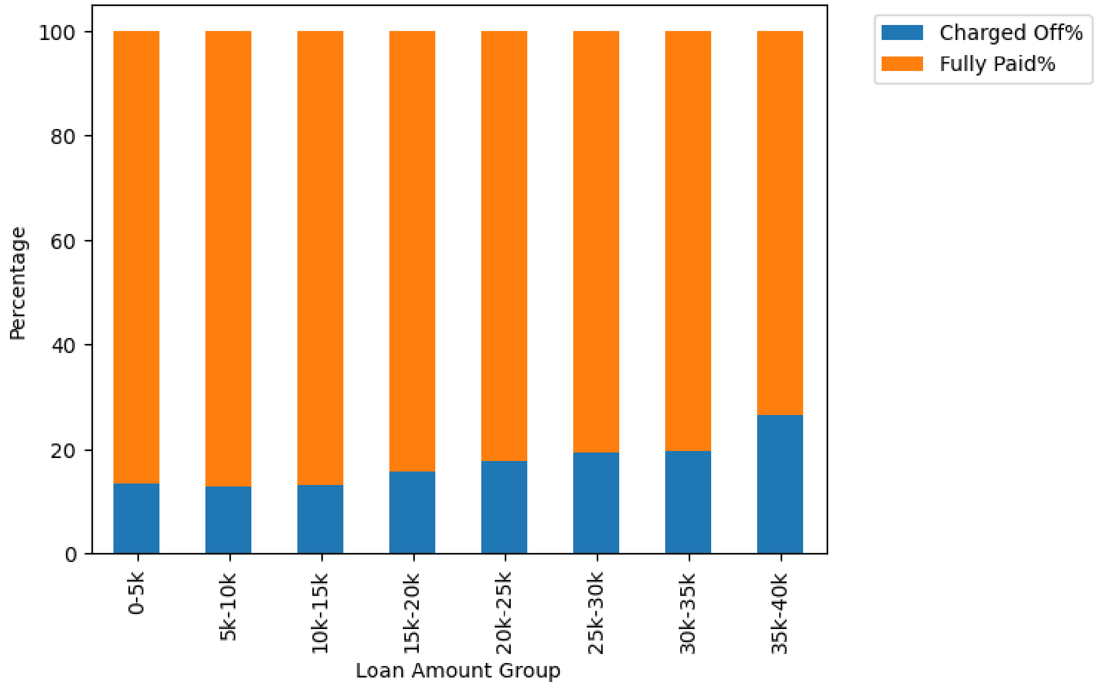
30 to 45k annual group receives most loans but default rate is highest for 3-30k



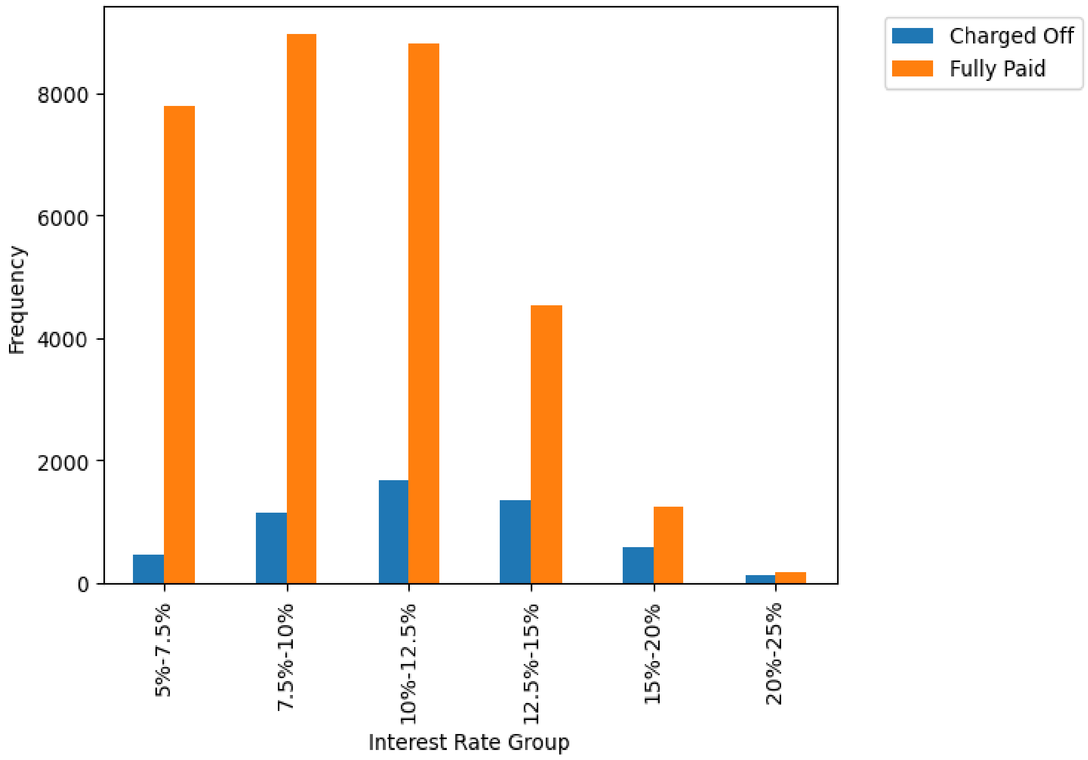


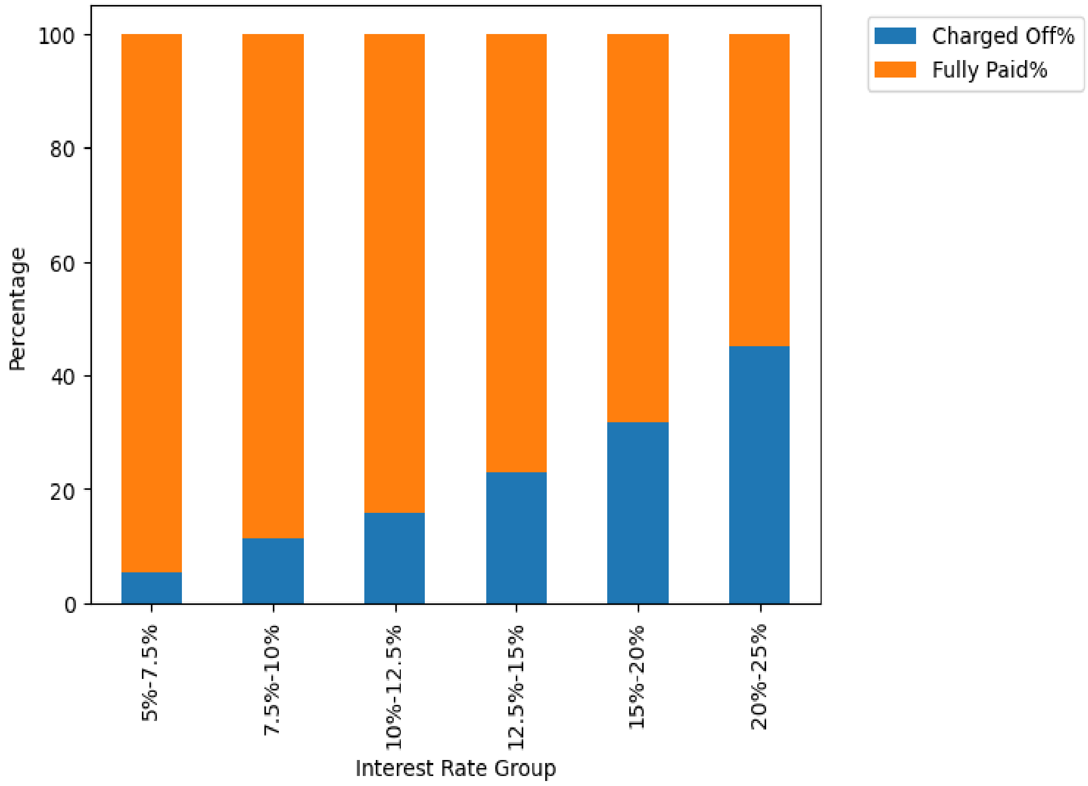
5-10k loan amount category is highest but default rate is highest for 35-40k





Lending rates of 10-12.5% is highest but default rate is seen highest for 20-25%





Bivariate Analysis- Summary

- The income group earning 30-45K has a higher number of loans, while the 3-30K income group has a higher percentage of defaults.
- Loans in the range of 5K-10K are more common, but the 25-35K loan amount group has a higher percentage of defaults.
- Loans with interest rates between 7.5-12.5% are more frequent, and the percentage of defaults increases with the interest rate.
- A loan term of 60 months is associated with a higher percentage of defaults.
- Grade B borrowers have a higher number of loans compared to Grade A, whereas Grade G has a higher percentage of defaults.
- The subgrade for G has a higher percentage of defaults.
- Property owners apply for fewer loans, and those without property have a higher rate of loan defaults.
- Employment with over 10 years of experience have a higher number of loans and a higher percentage of defaults.
- The majority of loans are taken out for debt consolidation, whereas loans for small business, home, and education have a higher default rate.

Final Recommendation

- There is a more probability of defaulting when
  - Applicants taking loan for 'home improvement' or 'Educational’
  - Applicants who does not won a property
  - Applicants who receive interest at the rate of 7.5-10.5%
  - Applicants who have taken a loan in the range 25-35K
  - When grade is G
  - When employment length is 10yrs
  - When the loan amount is above 25K
  - When term of loan is 60 months