

LENDING CLUB CASE STUDY

Prepared By
Sree Satya Latha
Srivatchala

Problem Statement

This case study is for a consumer finance marketplace which provides personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

Borrowers who **default** cause the largest amount of **loss to the lenders**. In this case, the customers labeled as *'charged-off' are the 'defaulters'*.

The core objective of the exercise is to **help the company minimize the credit loss**.

There are two potential sources of **credit loss**:

1. Applicant **likely to repay the loan**, such an applicant will bring in profit to the company with interest rates.** Rejecting such applicants will result in loss of business**.
2. Applicant **not likely to repay** the loan, i.e. and will potentially default, then approving the loan may lead to a financial loss* for the company

Objectives

The goal is 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 using the given dataset, is the aim of this case study.

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.

Approach:

➤ **Clean Data**

- ☐ Drop the columns having all null values and the columns having more null values
- ☐ Removing columns which are single valued as they do not contribute to analysis
- ☐ Modify the datatypes of columns and reformat the data wherever required.

➤ **Remove Outliers**

- ☐ Check for any outliers present for the numerical columns and remove them

➤ **Univariate analysis**

- ☐ Check distributions and frequencies of various numerical and categorical variables
- ☐ Create derived variables where applicable

➤ **Bivariate and Multivariate Analysis**

- ☐ Do correlation analysis, check how one variable effects other variable or third variable
- ☐ Analyze the join distributions across different variables

➤ **Summarize results**

- ☐ Provide the summary of analysis results, insights and observations

Driver variables

Below variables are pre-approval variables and are helpful for the analysis.

Loan Attributes:

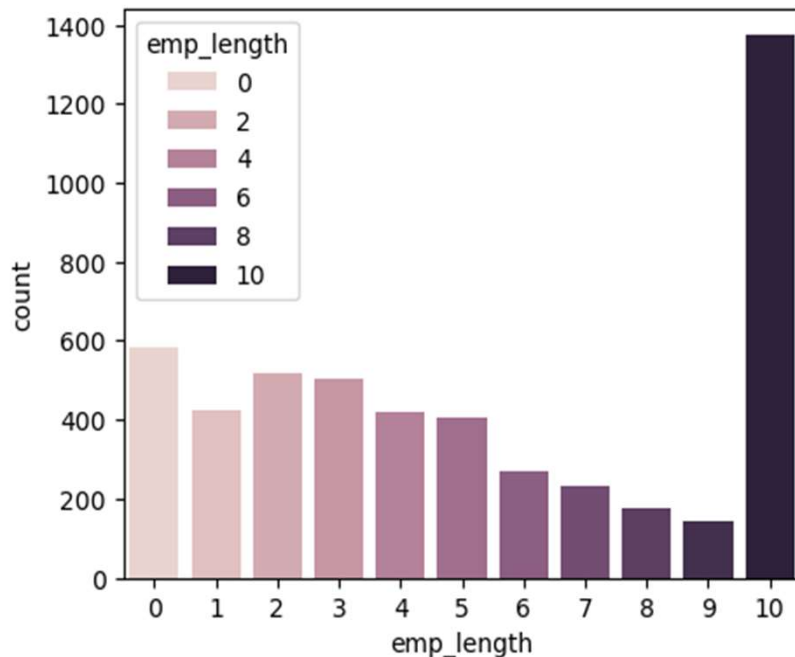
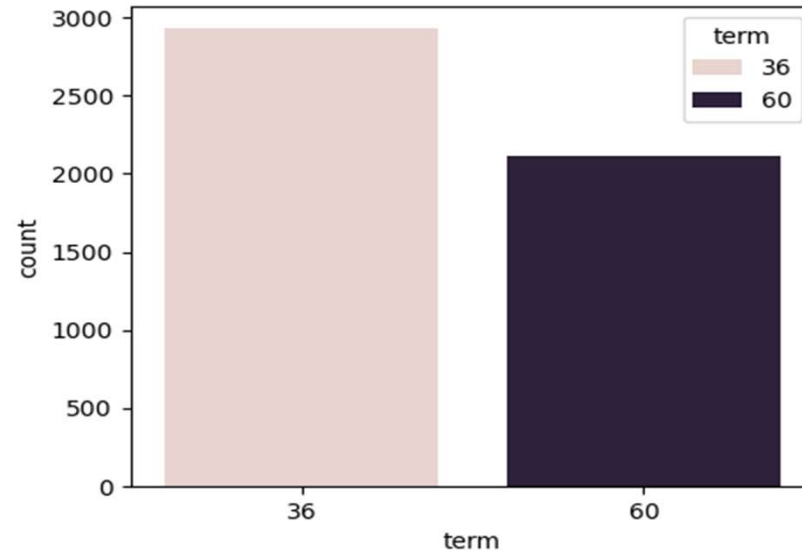
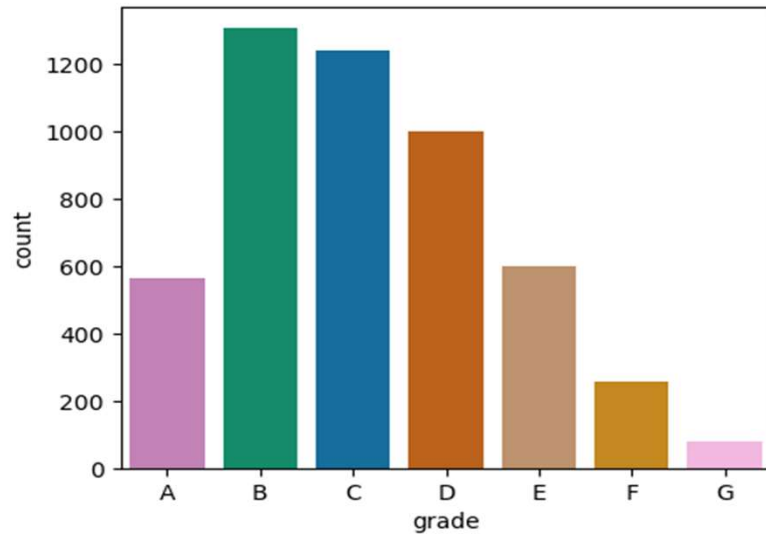
- Loan Ammount (loan_amt)
- Grade (grade)
- Term (term)
- Purpose of Loan (purpose)
- Loan Status
- Interest Rate (int_rate)

Customer Demographics

- Annual Income (annual_inc) - Annual income of the customer.
- Home Ownership(home_ownership) - Whether the customer owns a home or stays rented.
- Employment Length (emp_length) - Employment tenure of a customer
- Debt to Income (DTI) - The percentage of the salary which goes towards paying loan.
- State(addr_state) - Location of the customer.

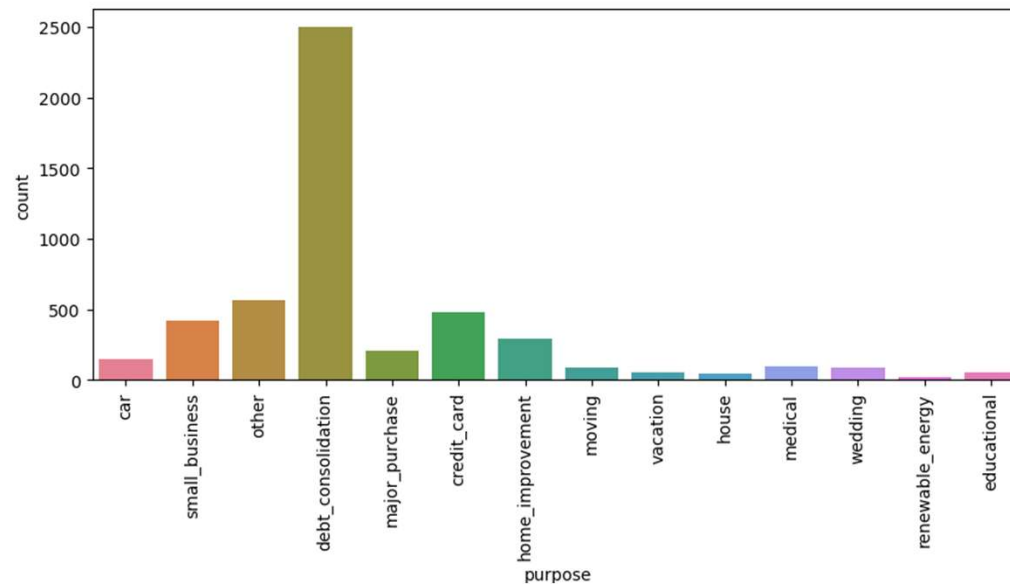
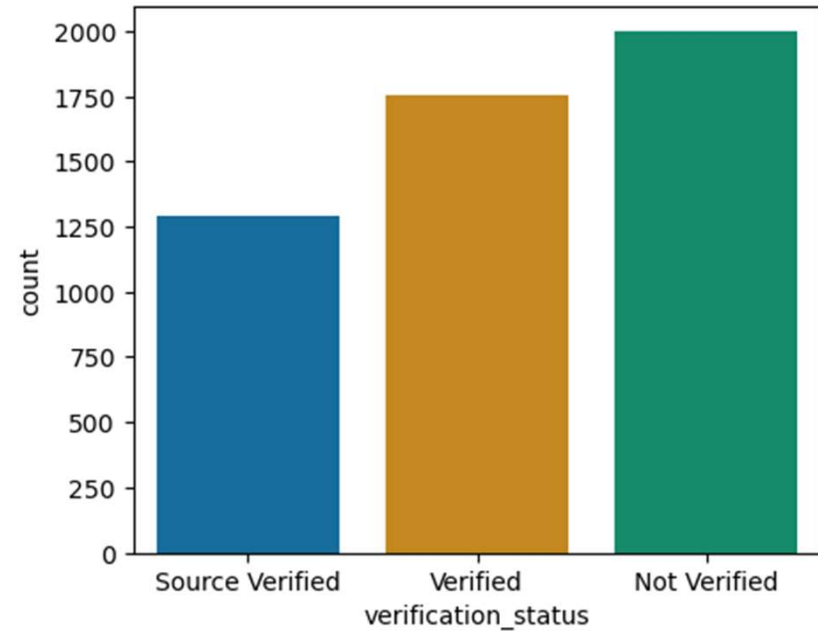
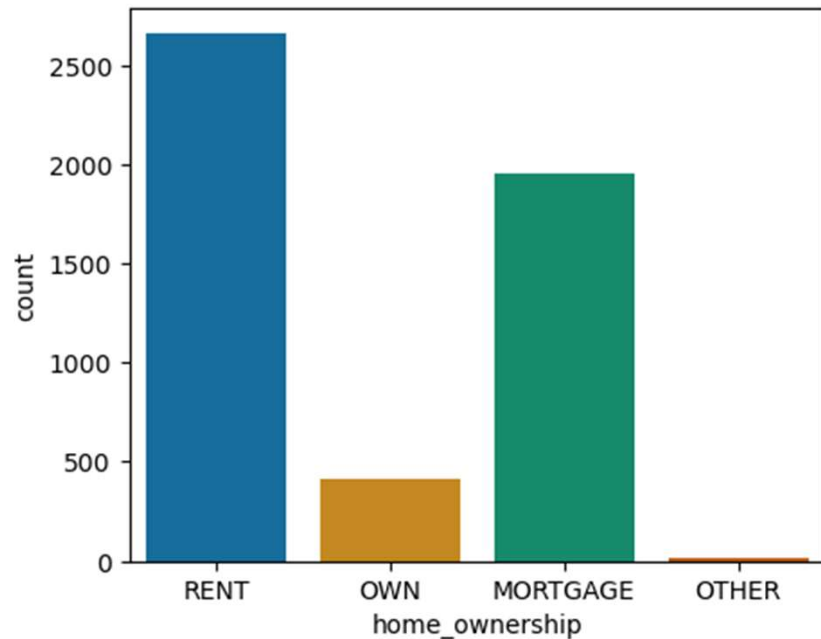
Univariate Analysis:

Analysis of Grade, emp_length, term:

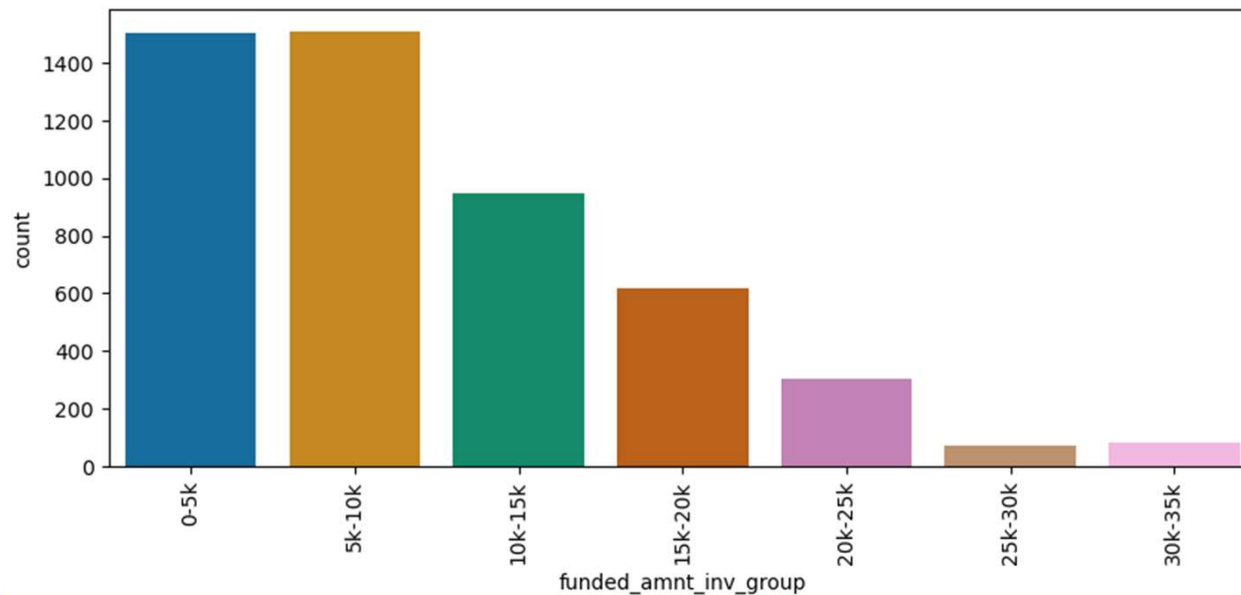
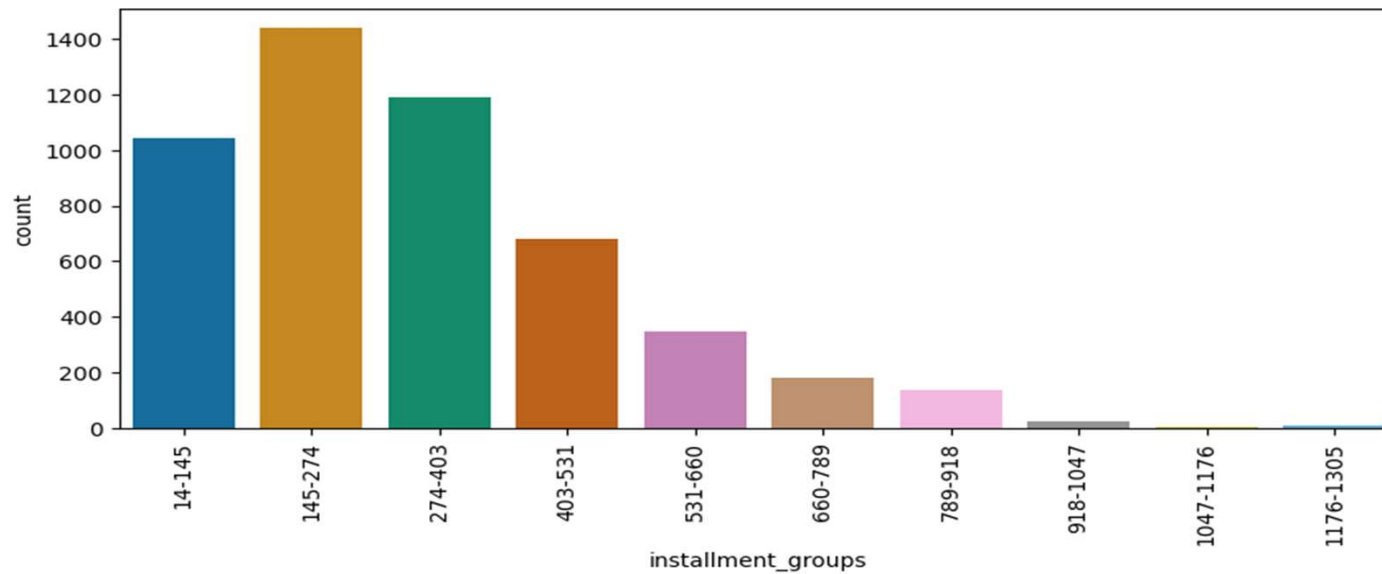


- Out of all the grades, the chances of loan defaults are more for B.
- Loan defaults are more for 36 months tenure compared to 60 months.
- Employees with 10 years employment are more in loan defaulters category

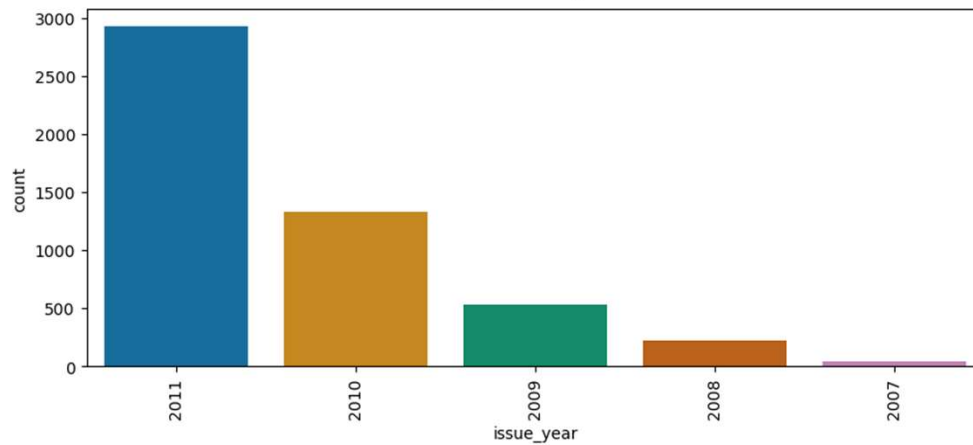
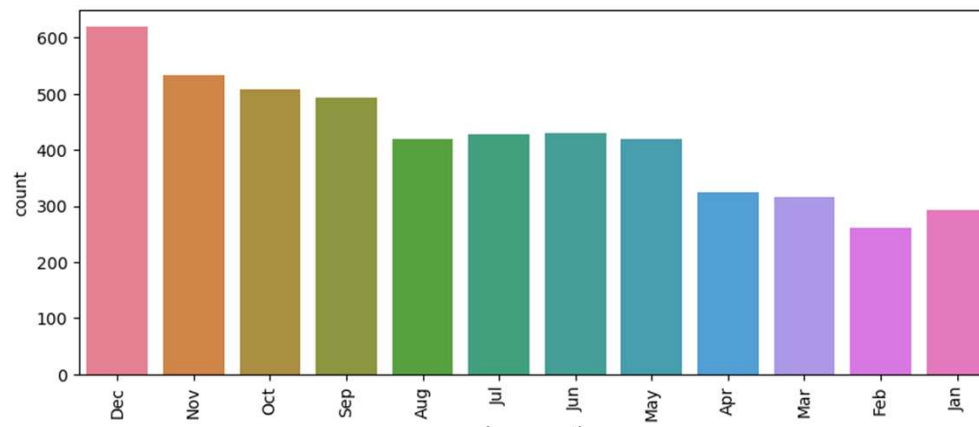
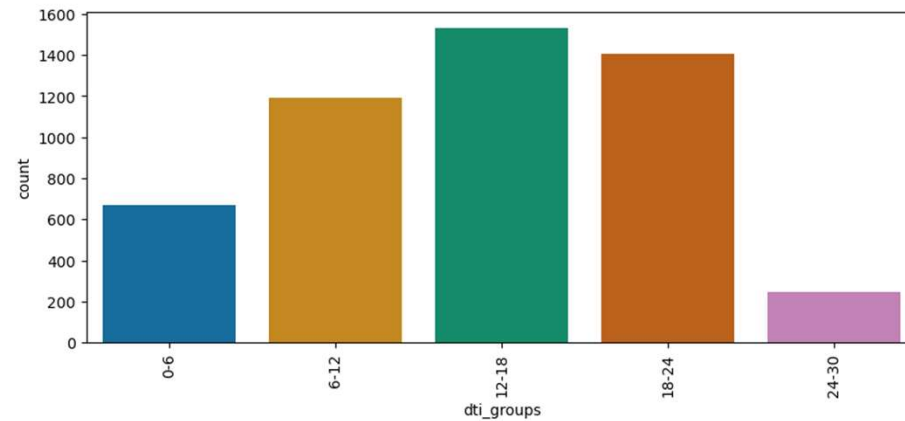
Analysis on home_ownership, verification_status, purpose:



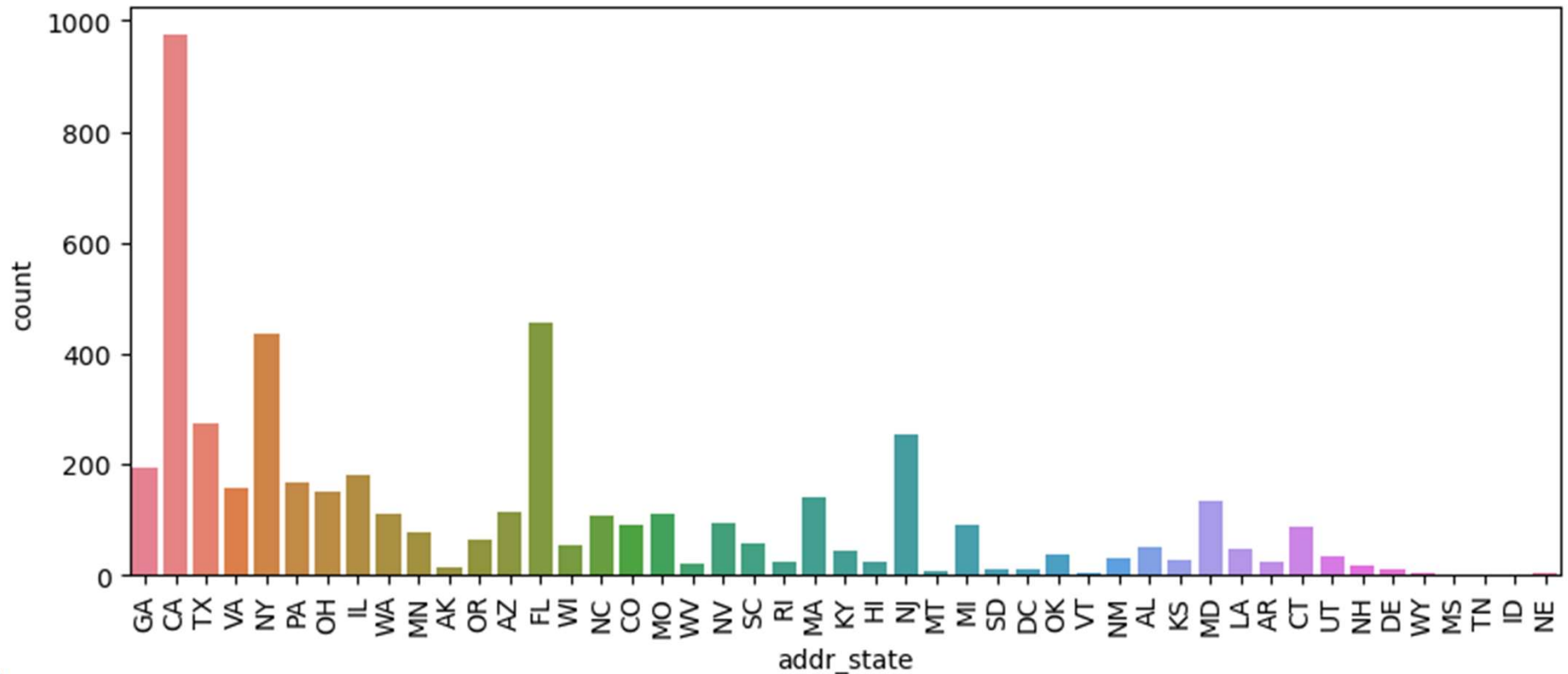
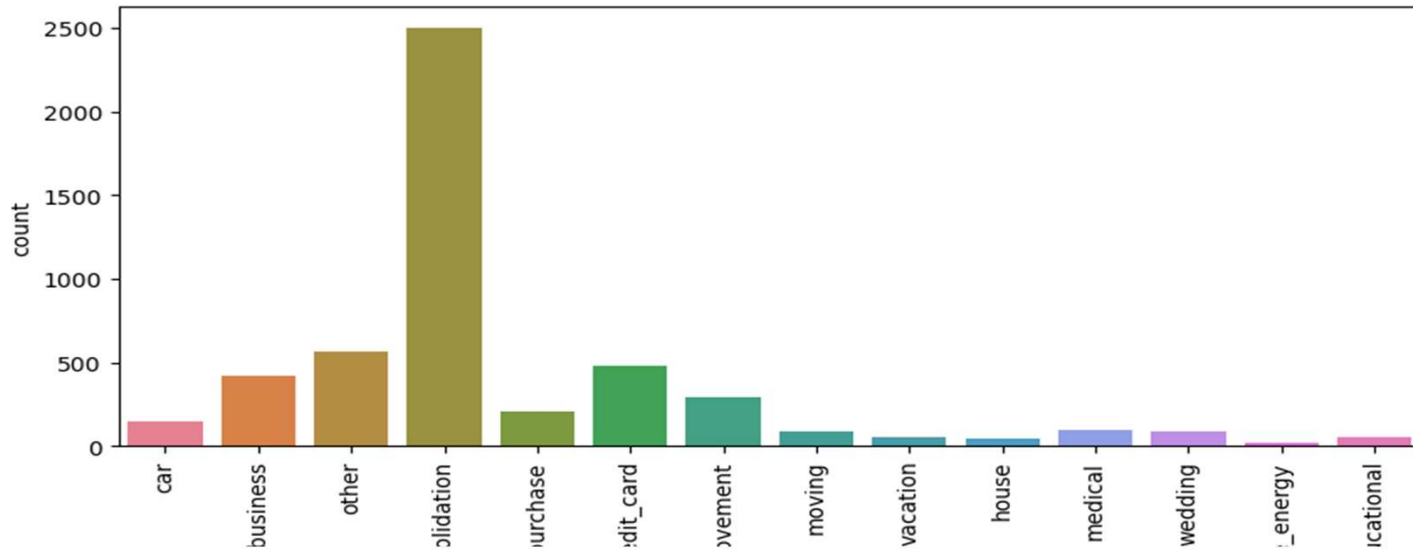
Analysis on derived columns: Installment_groups, funded_amnt_inv_group:



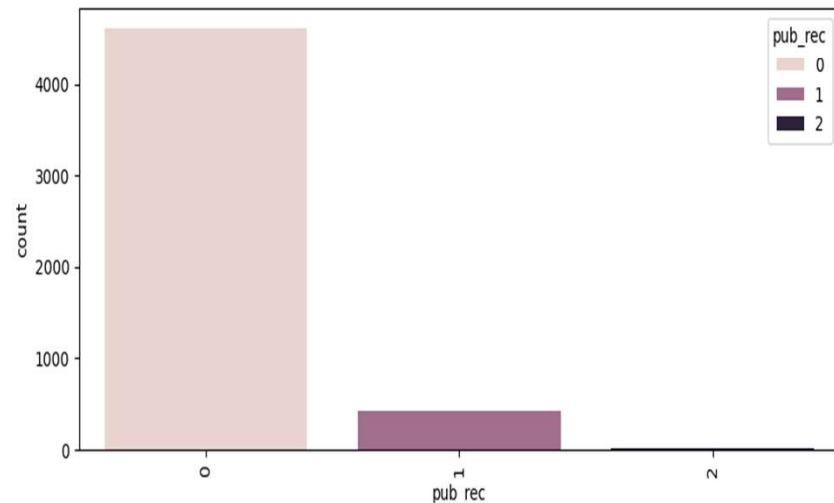
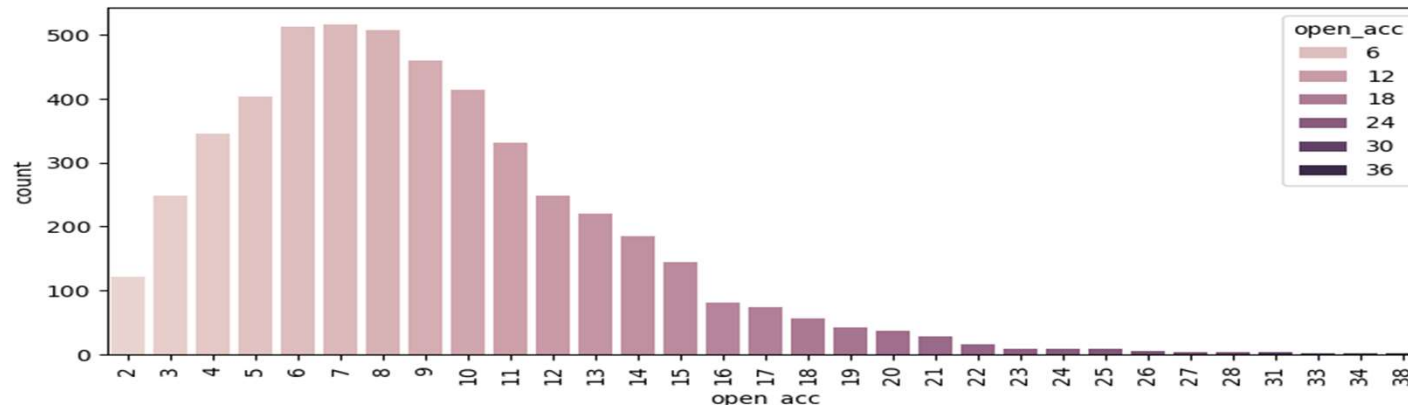
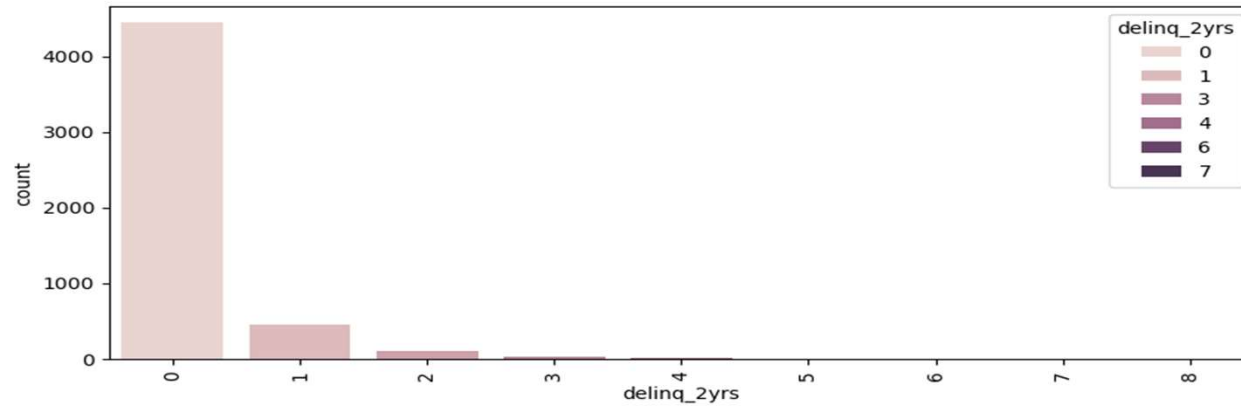
Analysis on dti groups, issue_month, issue_year:



Analysis on Loan Purpose, Loan state:



Analysis on delinq_years, open_acc, pub_rec:



Observations post Univariate analysis: Chances of loan defaulting is more:

- ✓ For loans with Grade B
- ✓ For loans with lesser tenure i.e 36 months
- ✓ When the employment tenure is more
- ✓ applicants with RENT house than with Own house
- ✓ For the loans disbursed without verification
- ✓ With installment range of 145-274
- ✓ When the funded amount is low
- ✓ For lower loan_amounts than higher loan_amounts
- ✓ For middle dti range of 12-18
- ✓ Most of the loan defaults happened during Year ends
- ✓ In 2011, there are highest defaults compared to other years
- ✓ Debt consolidation loans are more prone to defaulting
- ✓ In the state of CA, there are more defaulters

Multi Variate Analysis - I

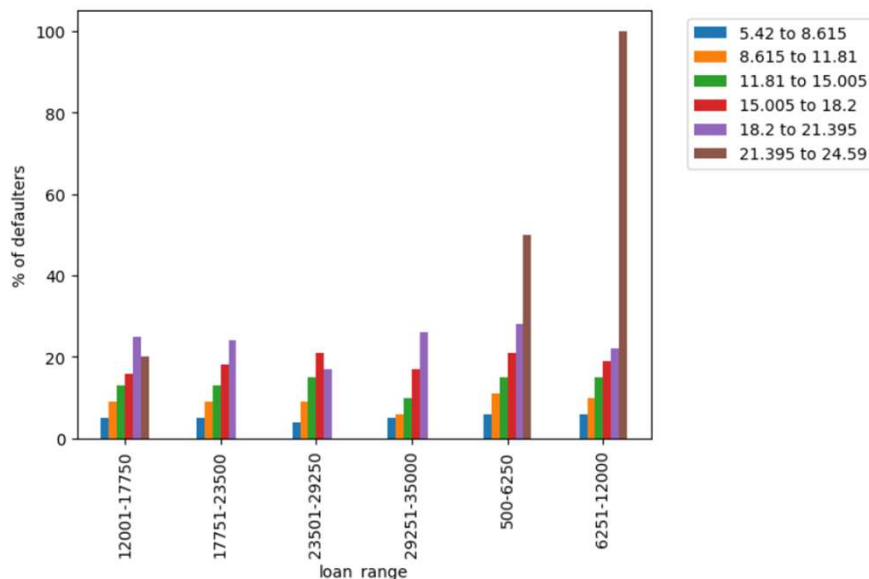
1. Identifying pattern of defaulters by combination of different loan amount range and interest rate range for term of 36 months and 60 months

For term of 36 months and any loan amount range, % of defaulters increases for increasing interest rates except for following two cases

#1. For loan range of 12000 to 17750, % of defaulters is high for 18.2% to 21.395% than 21.39% to 24.59% (applicants more likely to default for 18.2% to 21.395%)

#2. For loan range of 23501 to 29250, % of defaulters is high for 15.005% to 18.2% than 18.2% to 21.395% (applicants more likely to default for 15.005% to 18.2%)

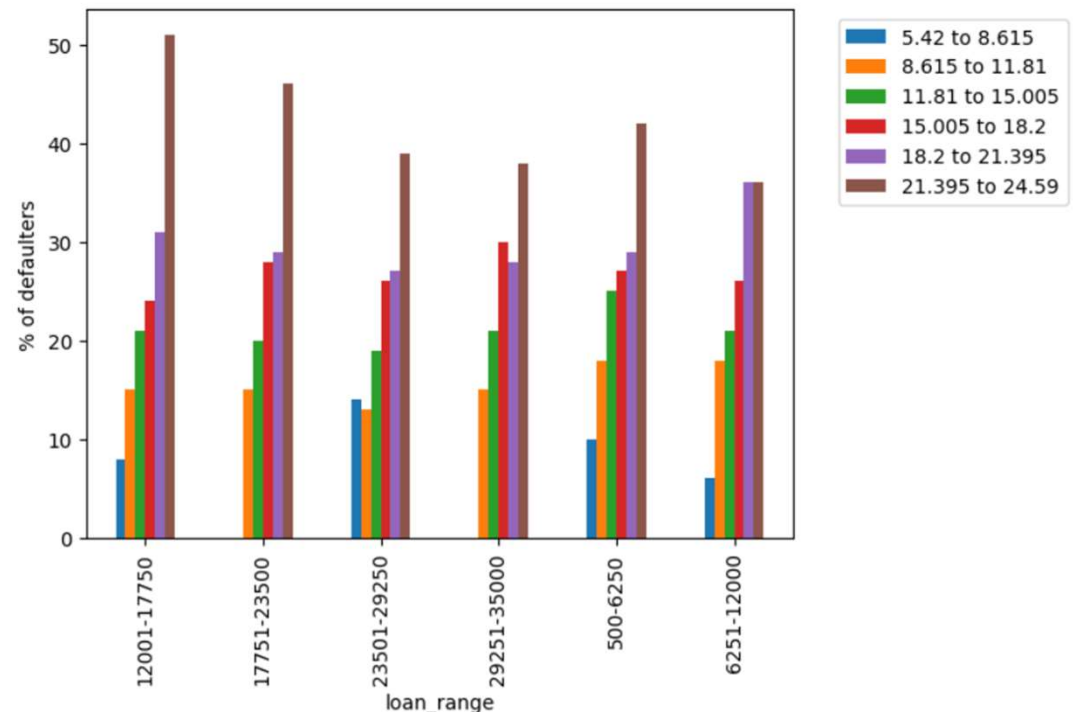
% Defaulters by loan amount range and interest rate range for term of 36 months



On comparing % of defaulters from term of 60 months with 36 months, we can see that for all interest rate except interest rate of > 21.395, term 36 months defaulters were near to 20% while for 60 months, all loan ranges with interest rate of >= 15% have more than 25% of defaulters.

Term 60 months have higher % of defaulters for higher interest rates

% Defaulters by loan amount range and interest rate range for term of 60 months

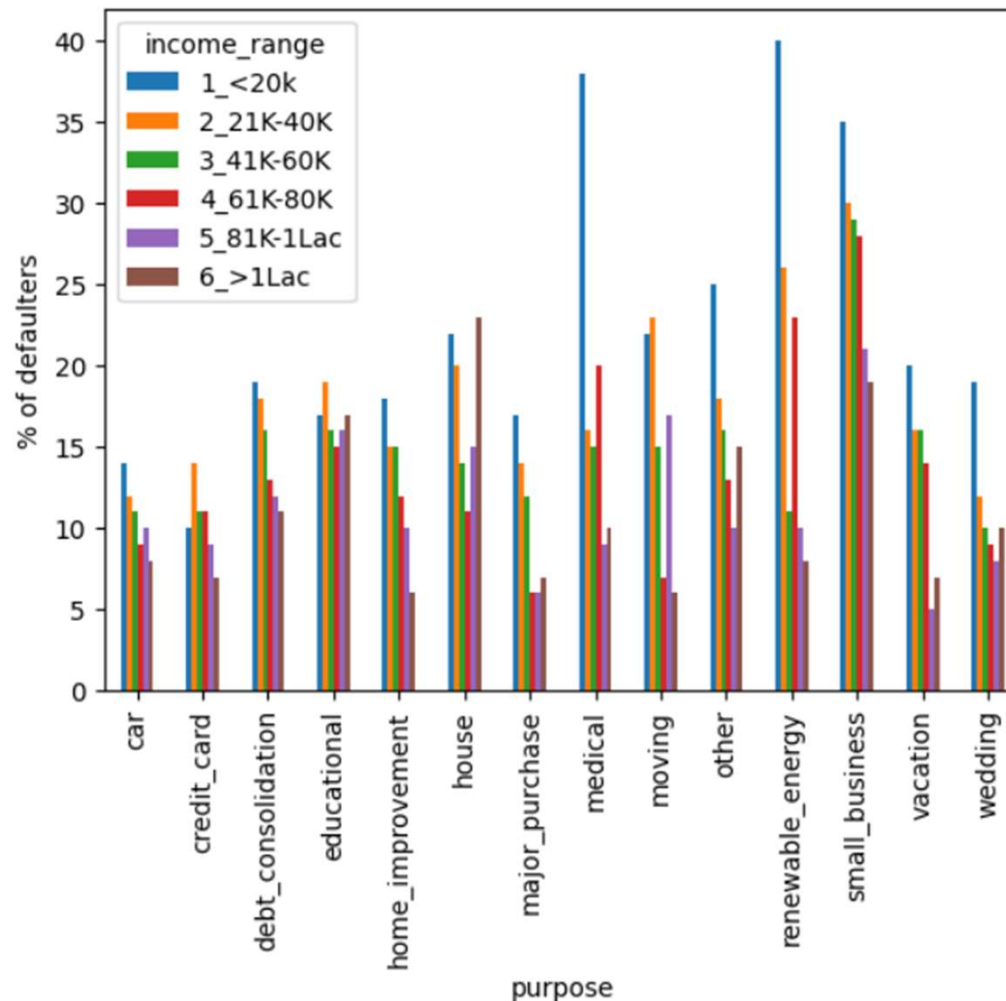


Multi Variate Analysis - 2

1. Identifying pattern of defaulters by combination of purpose for which the loan was bought and their annual income

When the **purpose of loan is for house**, applicants with income of **>1 lac** have defaulted(**23%**) more than **<20K range**(22%). For other purposes, <20K income range applicants have higher % of defaulters. Especially for purpose of renewable energy(40%), medical(38%) and small business (35%) defaulters from <20K range

% Defaulters by purpose of loan and their annual income

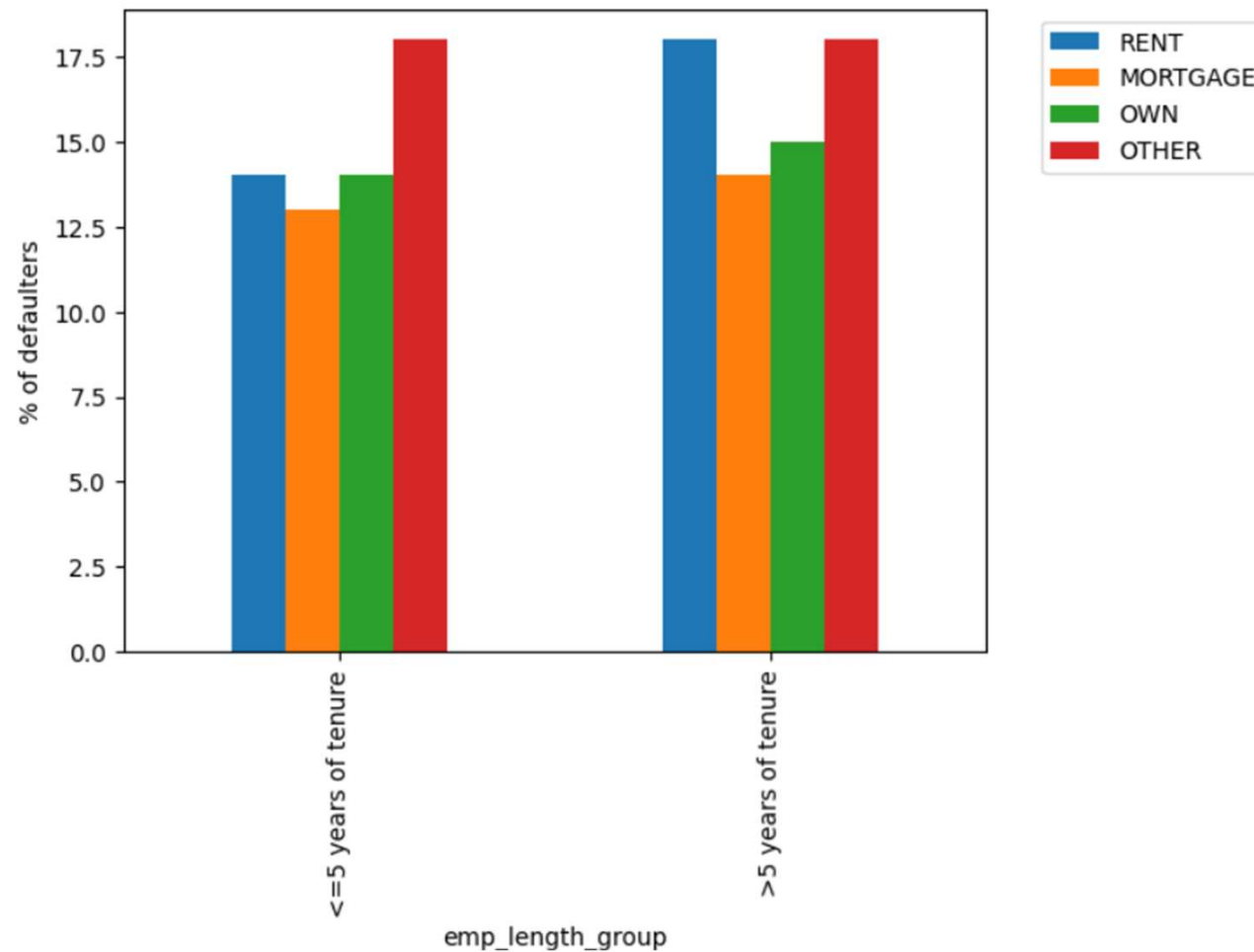


Multi Variate Analysis - 3

1. Identifying pattern of defaulters by combination of employment tenure and home ownership

Applicants with **employment tenure of more than 5 years and staying in a rented house have higher % of defaulters** compared to employee with tenure of less than or equal to 5 years. For other home ownerships of own, mortgage and others , both the tenures have similar % of defaulters

% Defaulters by employment tenure and home ownership



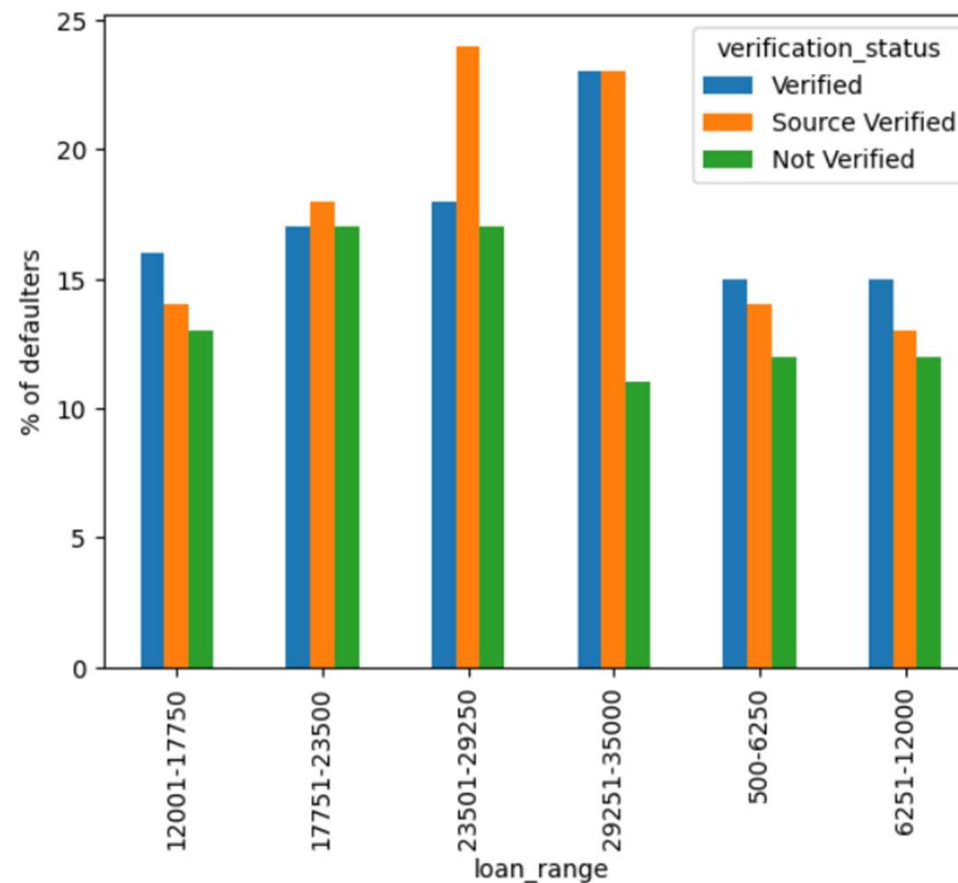
Multi Variate Analysis - 4

1. Identifying pattern of defaulters by combination of loan amount range and verification status

Applicants who **took loan amount in the range > 17750** have higher % of defaulters from **source verified** compared to other statuses

Applicants who **took loan amount in the range < 17750** have higher % of defaulters from **verified status** compared to other statuses

% Defaulters by loan range and verification status



Thank you