Lending Club Case Study

Submitted by:
- **Viswa Teja Bura**
- **Munish Gupta**

OBJECTIVE

The Objective of this case study is to understand the **driving factors/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.

Benefits of the case study:

➢ Identification of risky loan applicants, can prevent such loans thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

➢ The business objective is to take a decision whenever they receive a loan application whether to reject or approve based on certain variables.

DATASET Details

Customer's historical data is provided for this case study, it contains information about past credit history of customers. Data contains information about past loan applicants and whether they 'defaulted' or not.
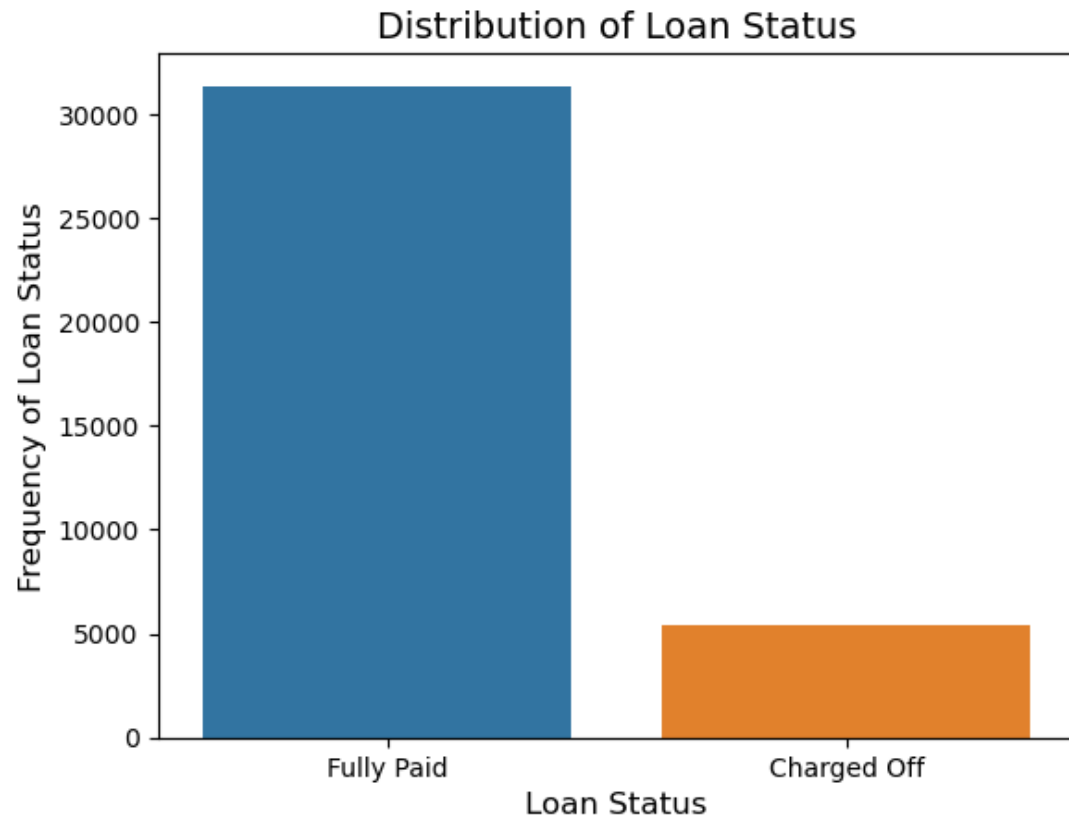
Data cleaning STEPS

- ➢ Importing the Data
- ➢ Remove null value columns
- ➢ Remove Duplicate data
- ➢ Remove columns which are not useful for analysis
- ➢ Fix null values in columns
- ➢ Correct data types of columns
- ➢ Remove outliers

Distribution of loan

Distribution of Loan Status

Insights

Charged off loans are lesser compared to Fully Paid loans. 14.5% loans are Charged off which is high percentage of defaults.
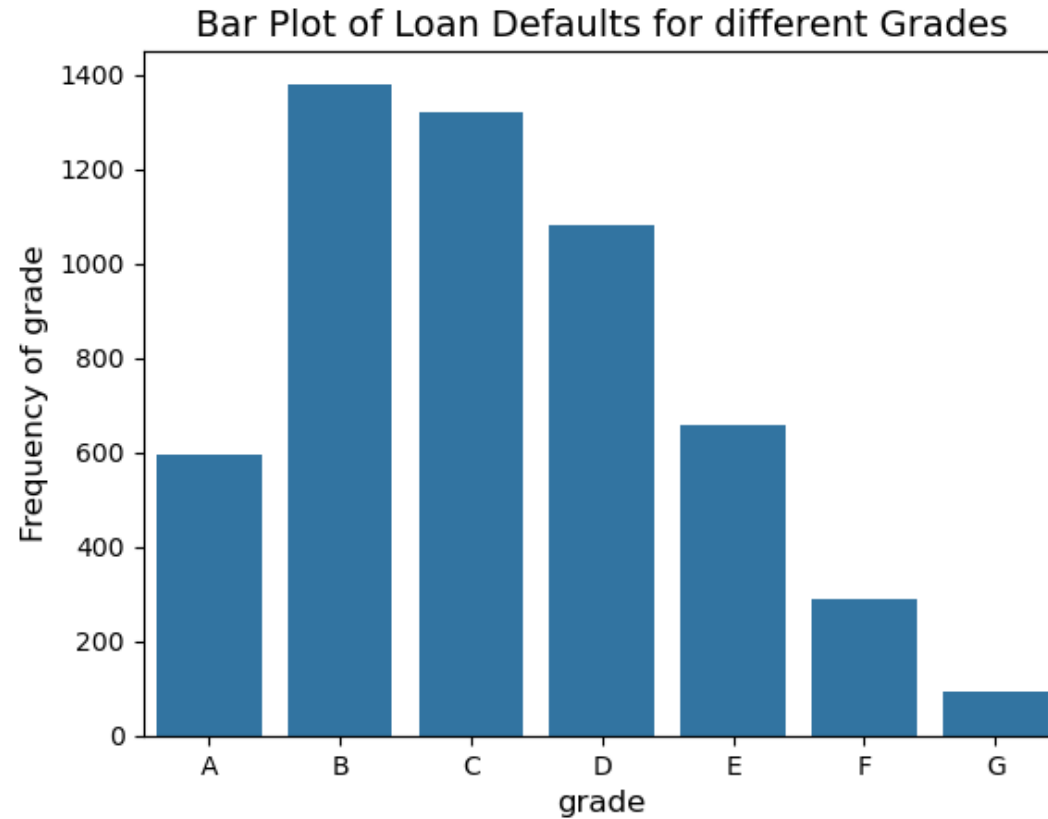
Univariate Analysis

Ordered categorical data

- ➢ 1. Sub grade (sub_grade)
- ➢ 2. Grade (grade)
- ➢ 3. Term (36 / 60 months) (term)
- ➢ 4. Employment length (emp_length)
- ➢ 5. Issue year (issue_y)
- ➢ 6. Issue month (issue_m)

Bar Plot of Loan Defaults for different Grades
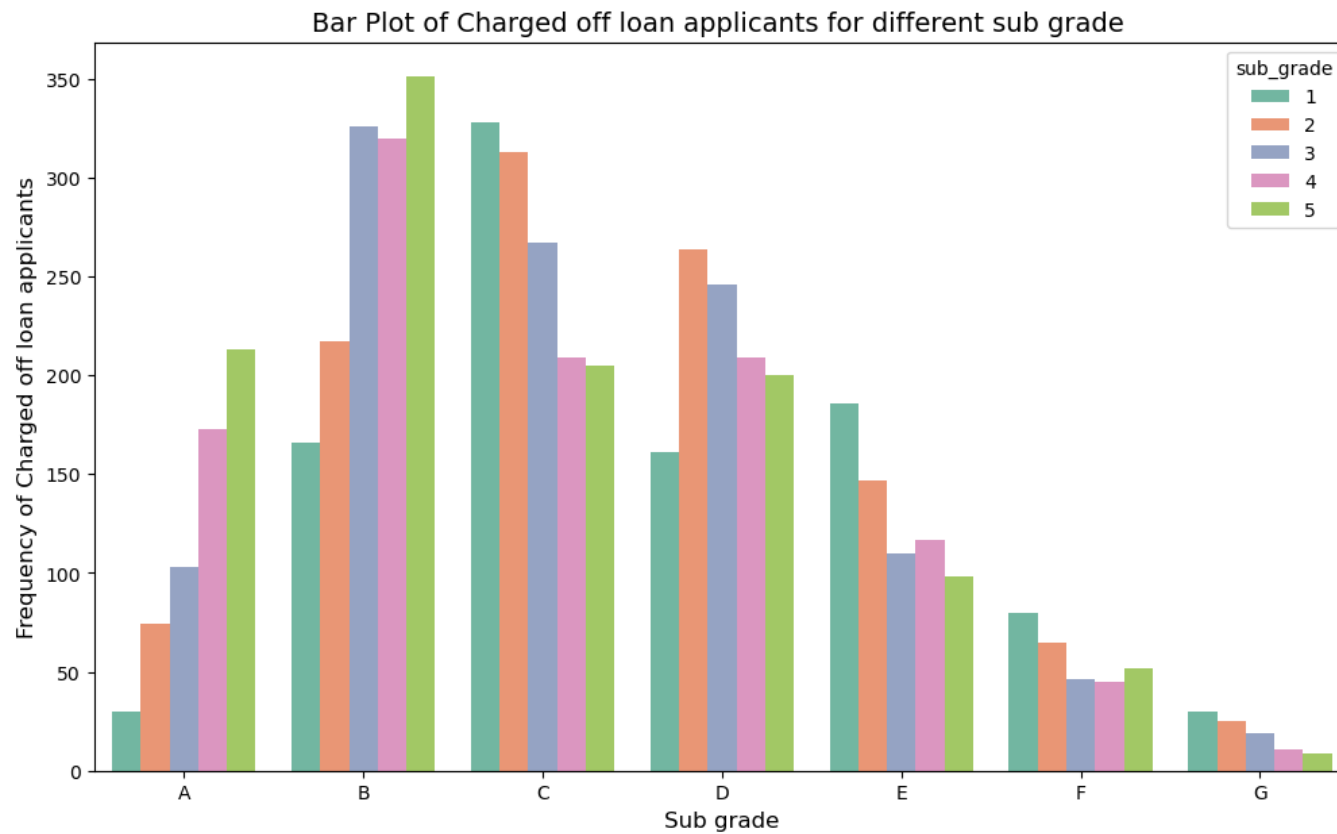

Bar Plot of Loan Defaults for different Grades

Insights

Grades B,C,D had the highest number of "Charged off" loan applicants, indicating that applicants with this credit grade faced challenges in repaying their loans.

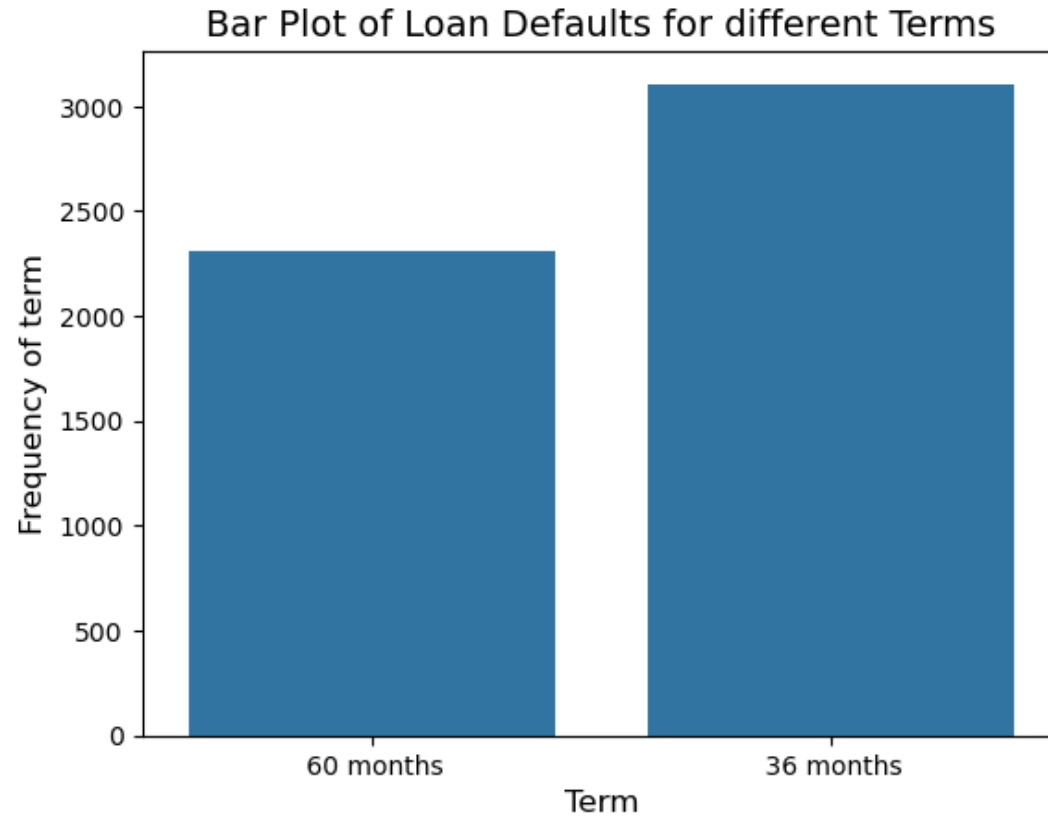# Bar Plot of Charged off loan applicants for different sub grade

Bar Plot of Charged off loan applicants for different sub grade

Insights

Borrowers from sub grade B3, B4 and B5 have maximum tendency to default.

# Bar Plot of Loan Defaults for different Terms

Bar Plot of Loan Defaults for different Terms

Insights

Short-term loans with a duration of 36 months were more likely to Loan Default. This suggests that a significant portion of applicants who experienced loan default chose shorter repayment terms.

# Distribution of Employment length for Charged off loan applicants
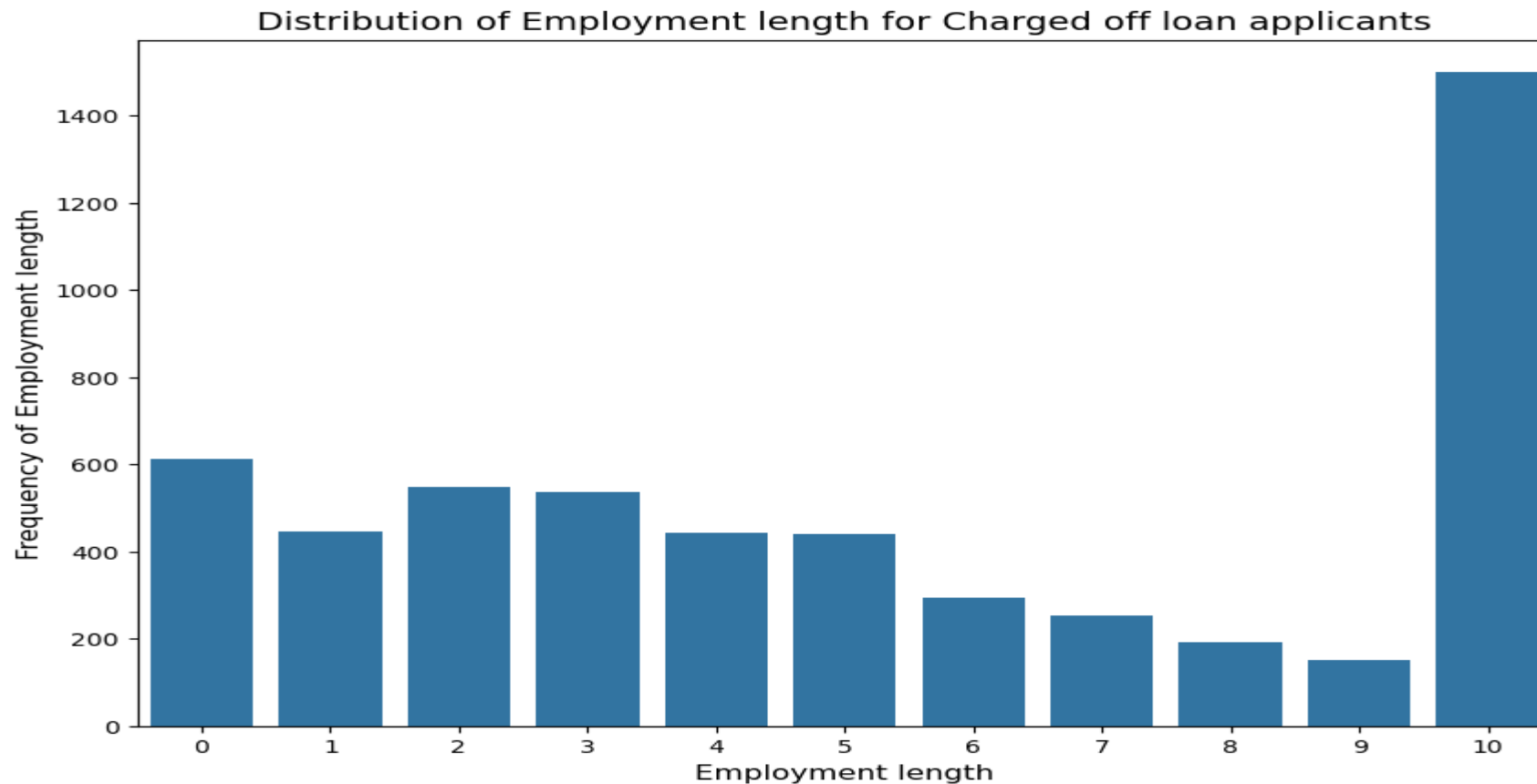
Insights

Applicants who had been employed for more than 10 years accounted for the highest number of "Charged off" loans. This indicates that long-term employment history did not necessarily guarantee successful loan repayment.

Bar Plot of Charged off loan applicants across years

Bar Plot of Charged off loan applicants across years
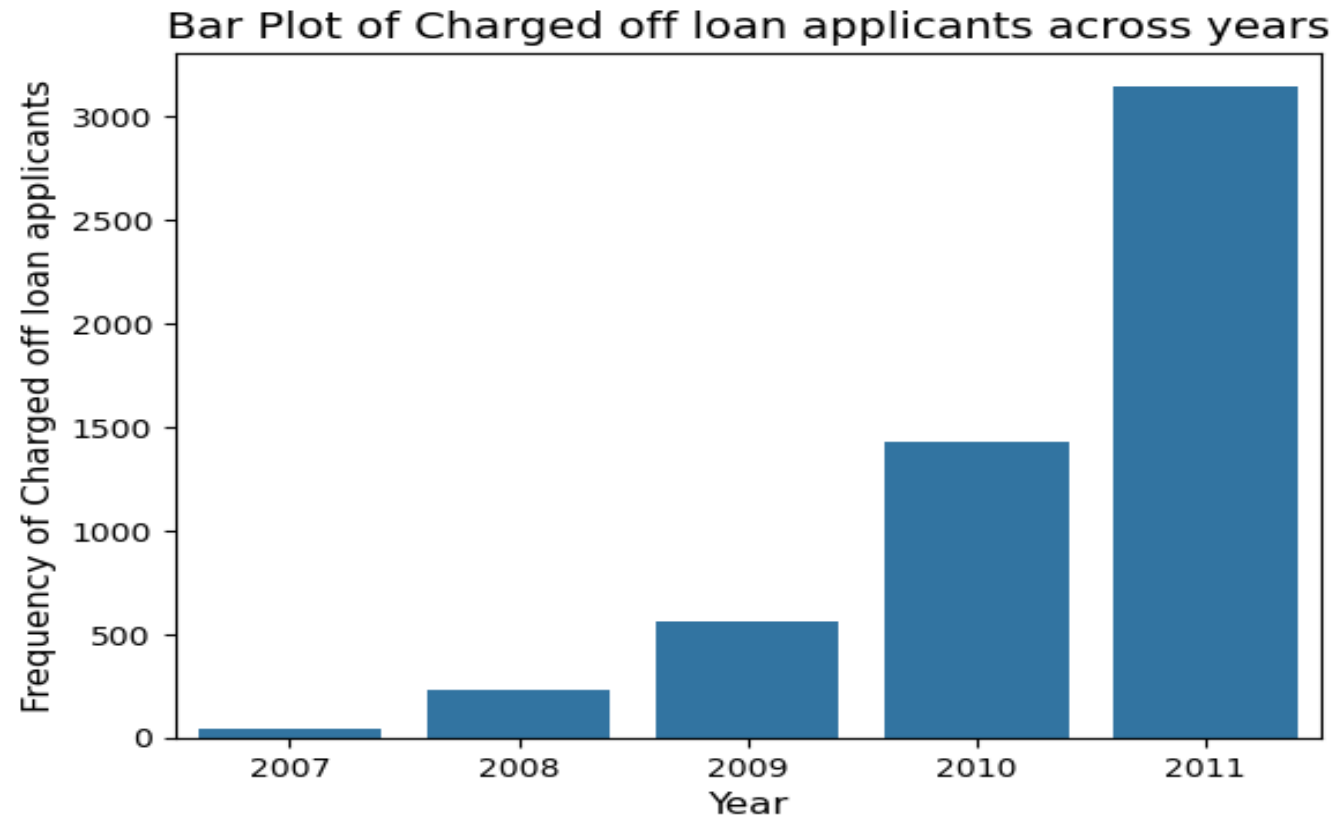
Insights

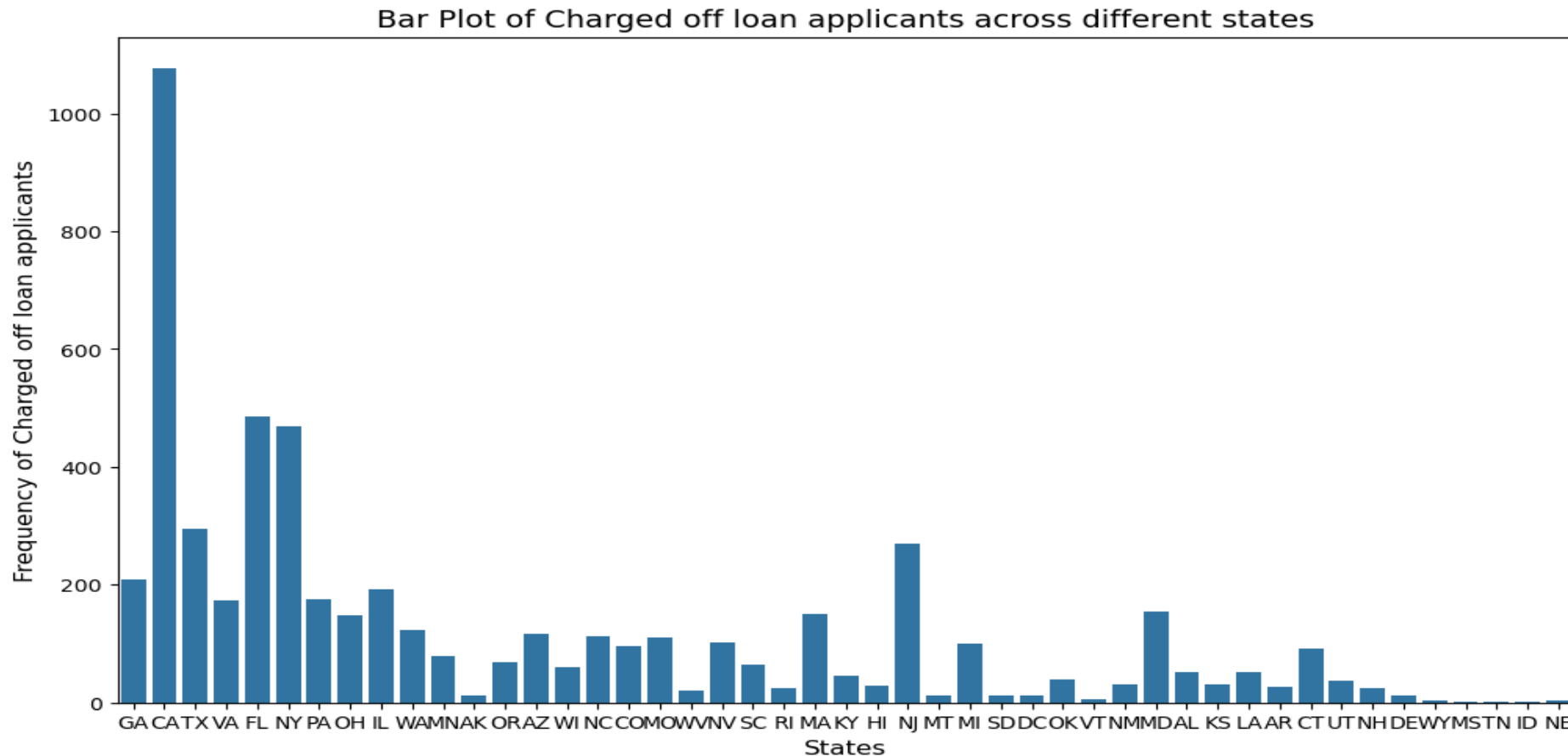The high number of loan defaults happened in 2011 could be due to some financial crisis.

Univariate Analysis

Unordered categorical data

- ➤ Address State (addr_state)
- ➤ Loan purpose (purpose)
- ➤ Home Ownership (home_ownership)

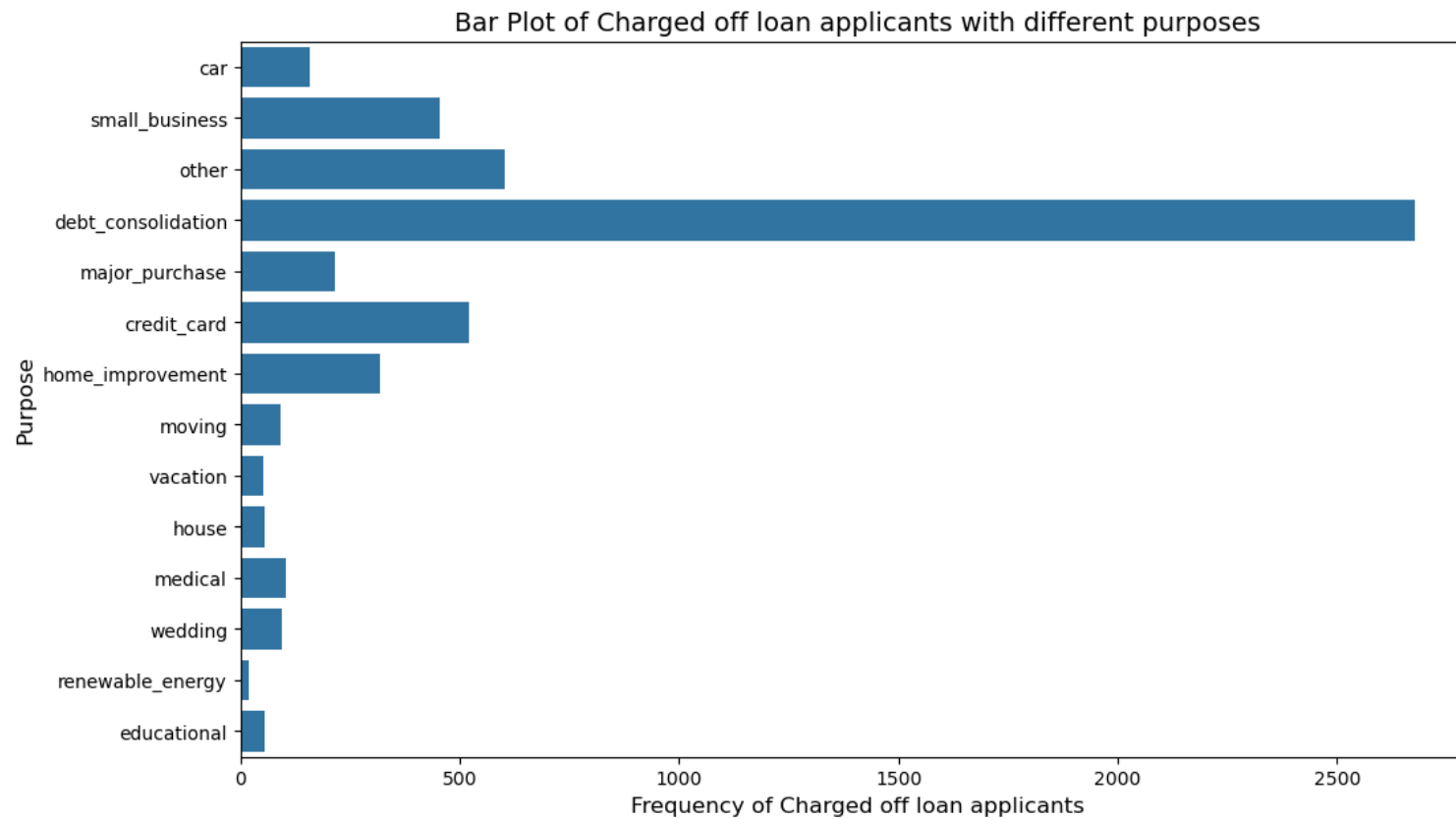# Bar Plot of Charged off loan applicants across different states

Bar Plot of Charged off loan applicants across different states

Insights

Borrowers from states CA, FL, NJ have maximum tendency to default the loan.

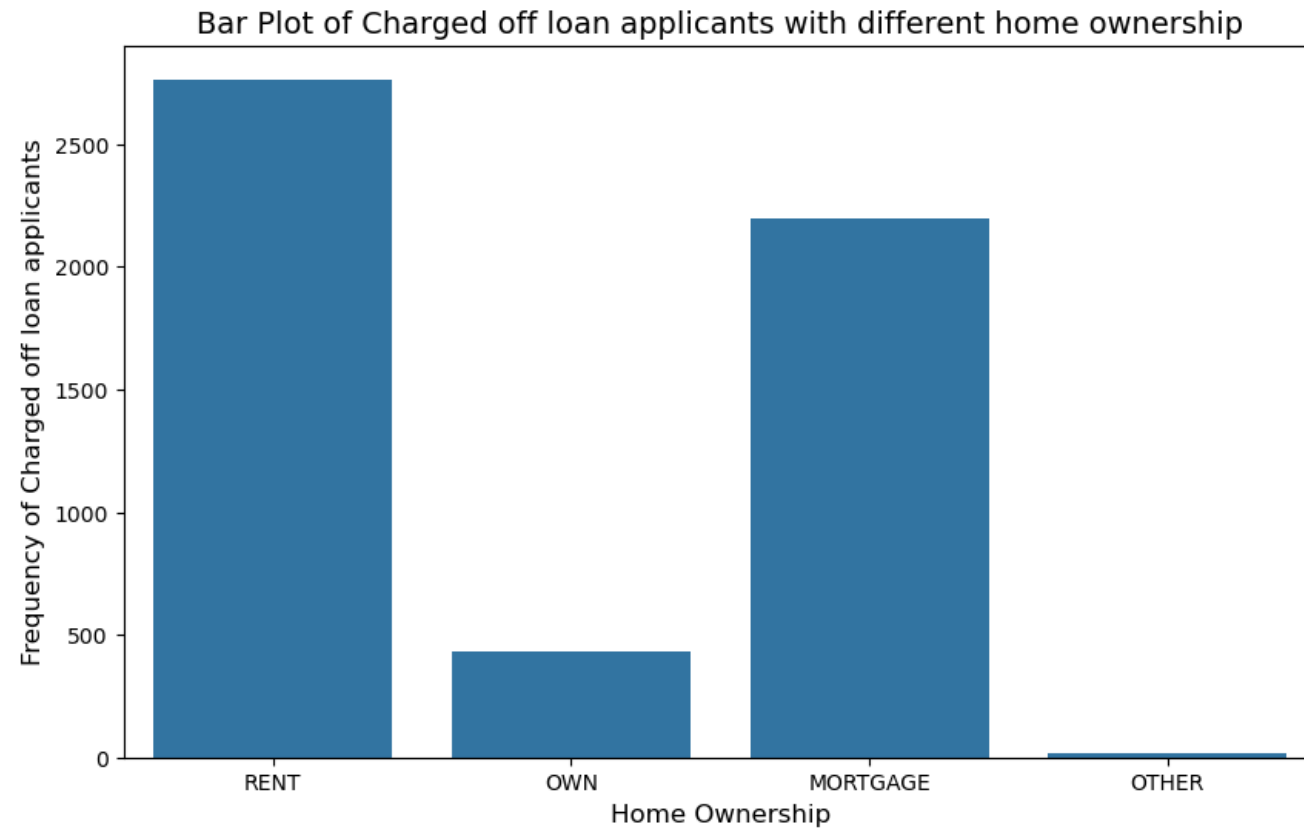# Bar Plot of Charged off loan applicants with different purposes

Bar Plot of Charged off loan applicants with different purposes

Insights

Debt consolidation was the primary loan purpose for most "Charged off" loan applicant.

Bar Plot of Charged off loan applicants with different home ownership

Bar Plot of Charged off loan applicants with different home ownership

Insights

The majority of "Charged off" loan participants lived in rented houses or mortgage.

Univariate Analysis
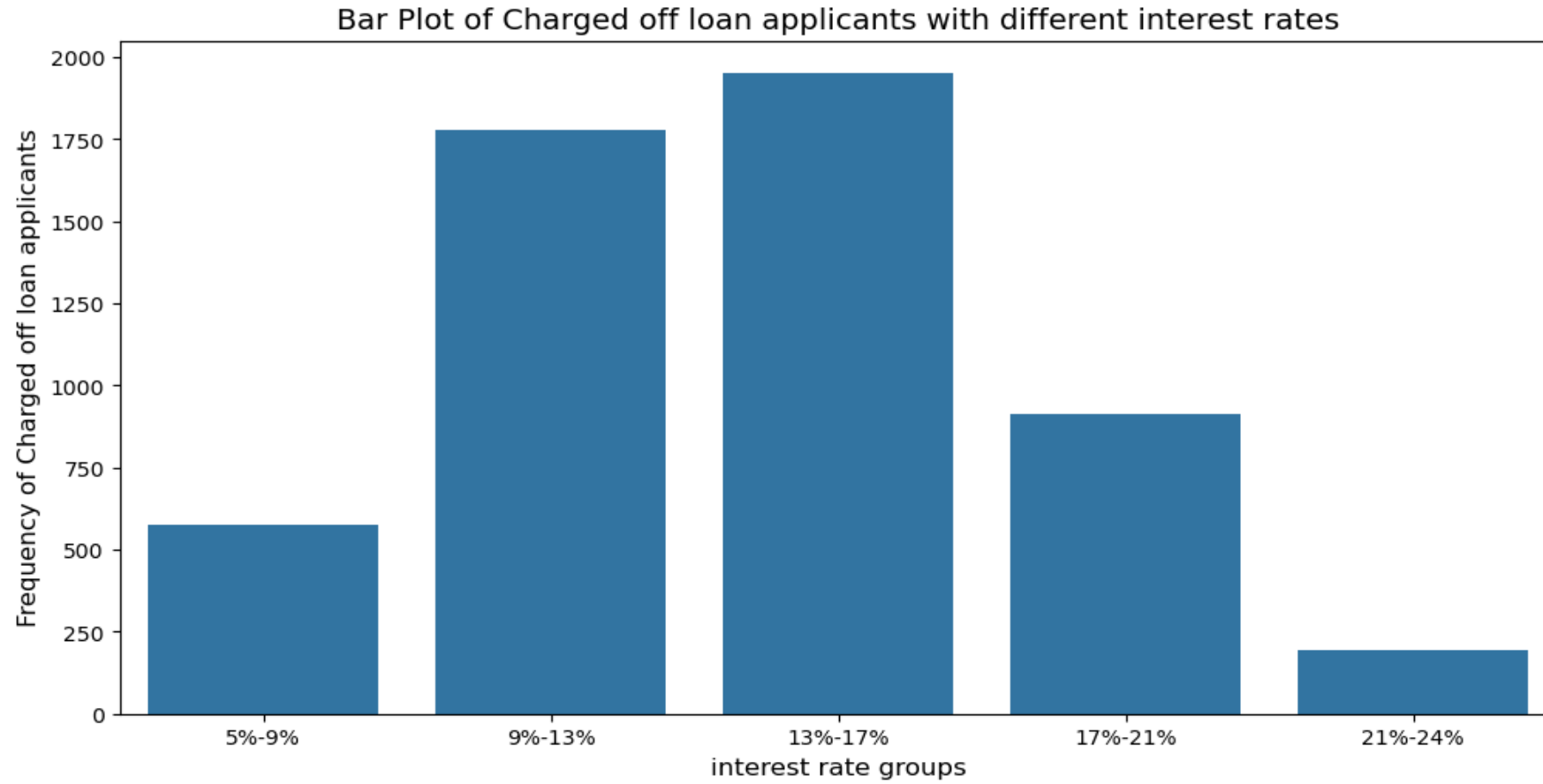
Quantitative variables

- ➢ Interest rate bucket (int_rate_groups)
- ➢ Annual income bucket (annual_inc_groups)
- ➢ Loan amount bucket (loan_amnt_groups)
- ➢ Debt to Income rate (DTI)
- ➢ Funded amount bucket (funded_amnt_groups)

Bar Plot of Charged off loan applicants with different interest rates
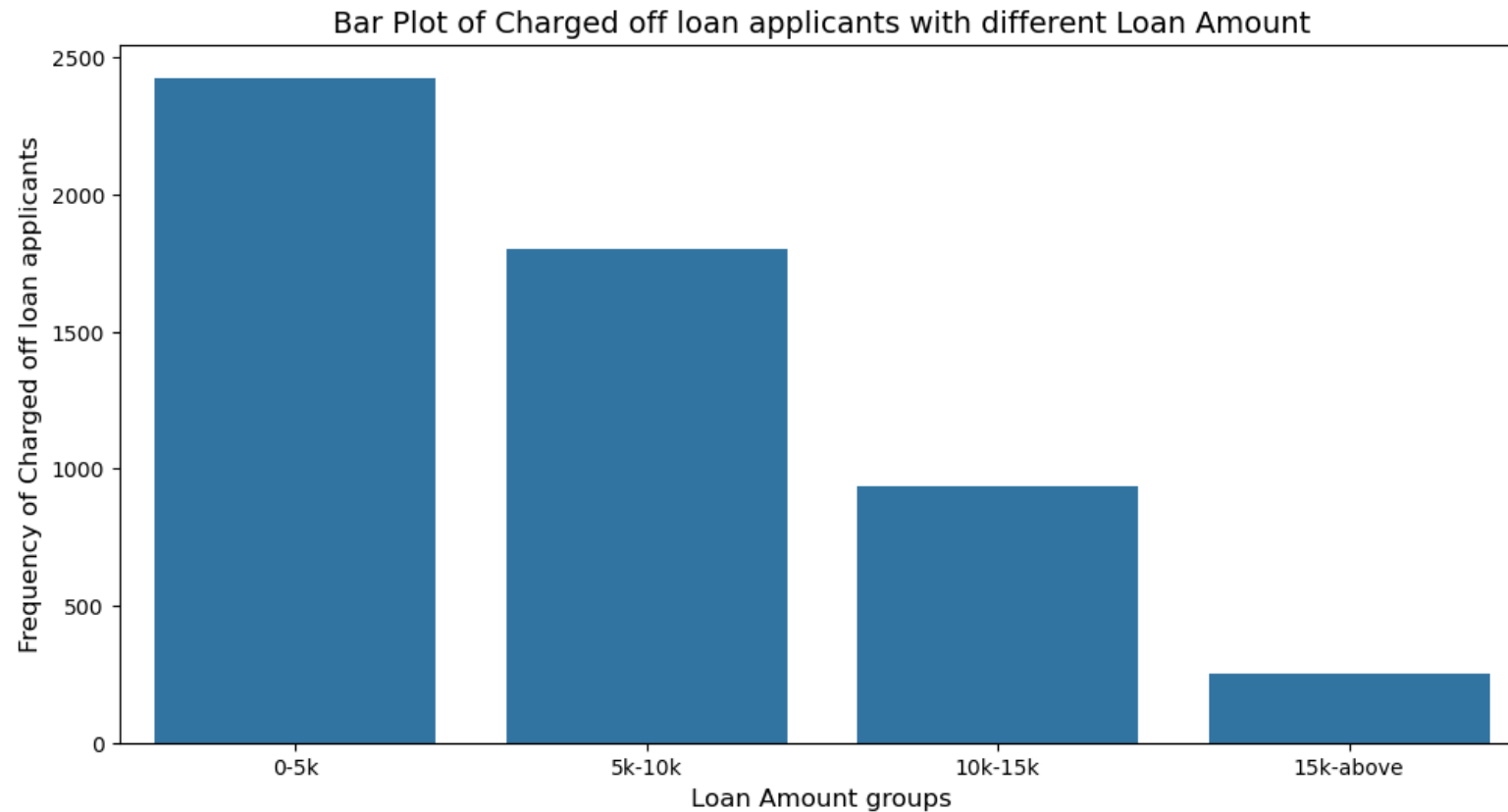

Bar Plot of Charged off loan applicants with different interest rates

Insights

Most of the loan participants who charged off has Dti is between 13-17

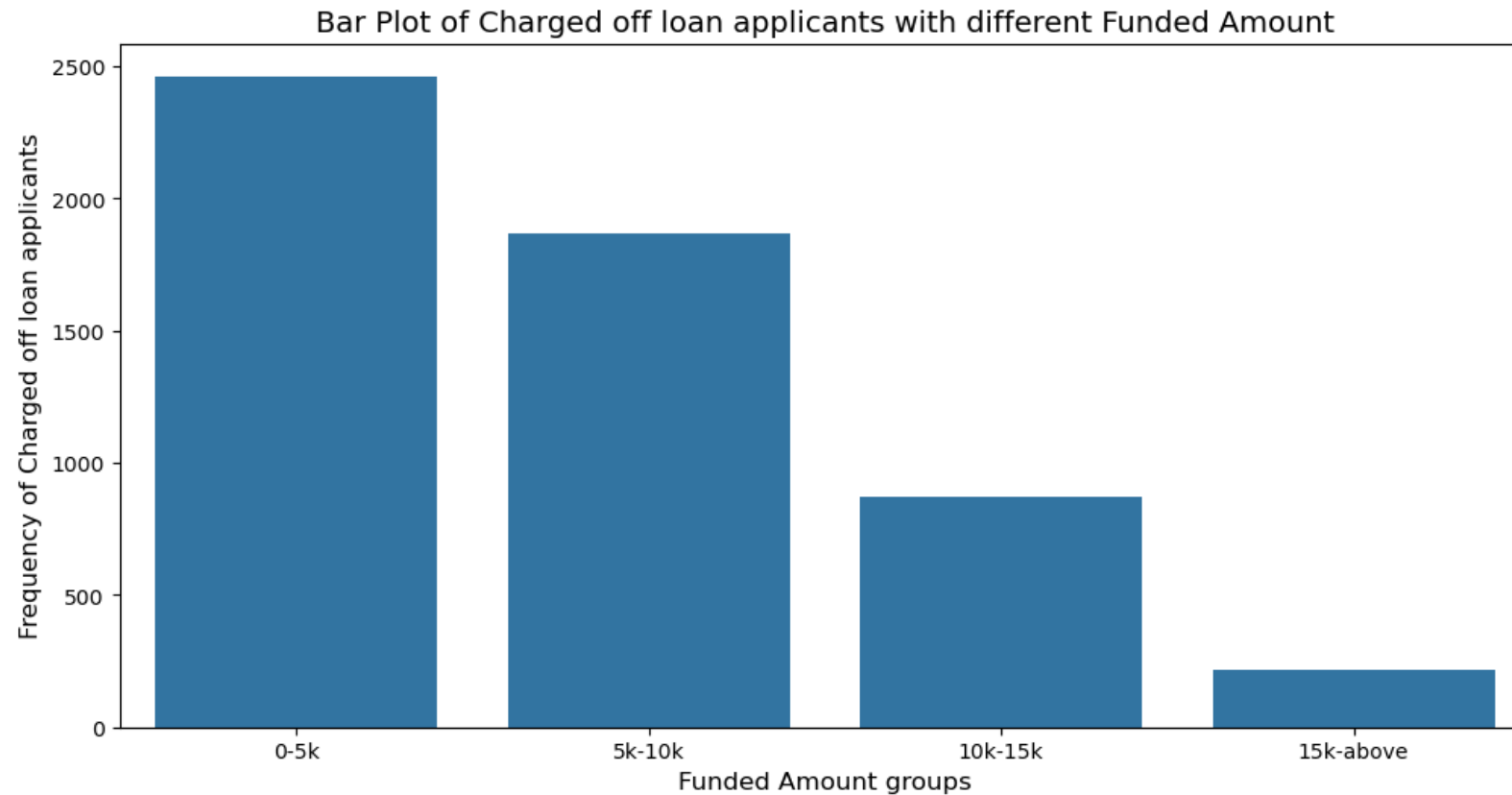# Bar Plot of Charged off loan applicants with different Loan Amount

Bar Plot of Charged off loan applicants with different Loan Amount

Insights

Most of the loan participants who charged off received loan amounts are below 10,000 USD.

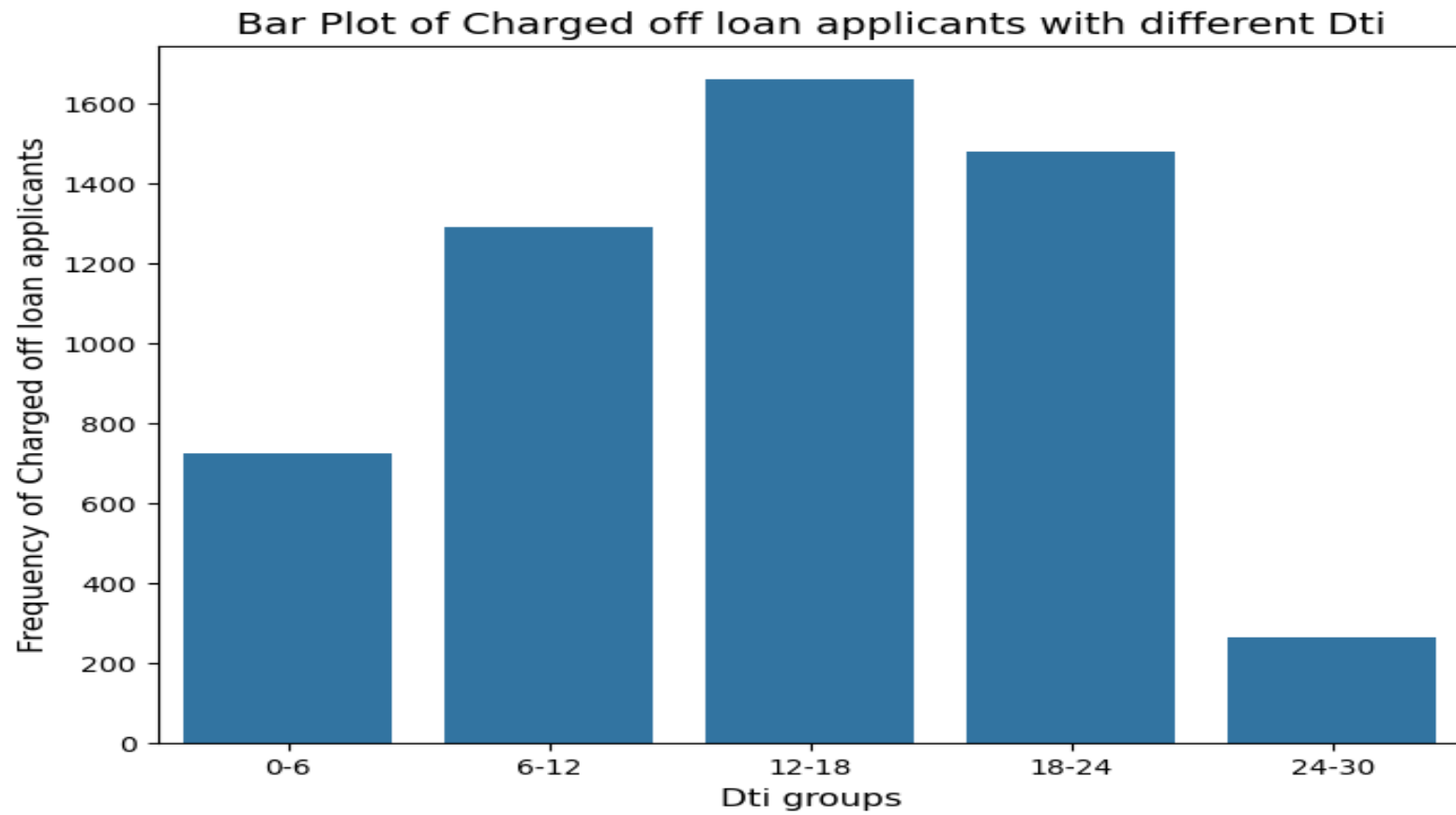# Bar Plot of Charged off loan applicants with different Funded Amount

Bar Plot of Charged off loan applicants with different Funded Amount

Insights

Most of the loan participants who charged off received funded amounts are below 10,000 USD .

Bar Plot of Charged off loan applicants with different Dti

Insights

The majority of "Charged off" loan participants lived in rented houses or mortgage.

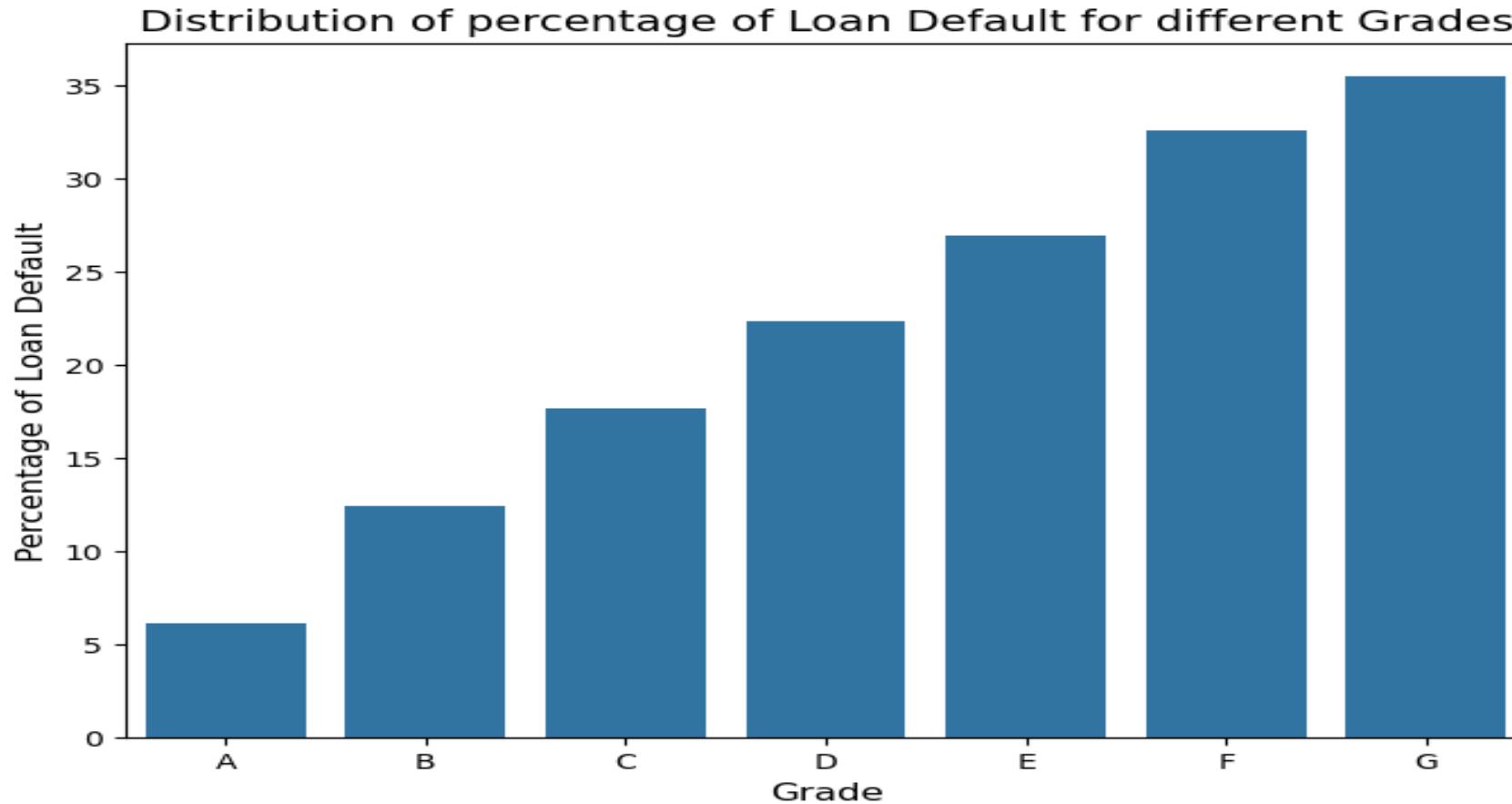# Distribution of Annual Income for Charged off loan applicants

Distribution of Annual Income for Charged off loan applicants
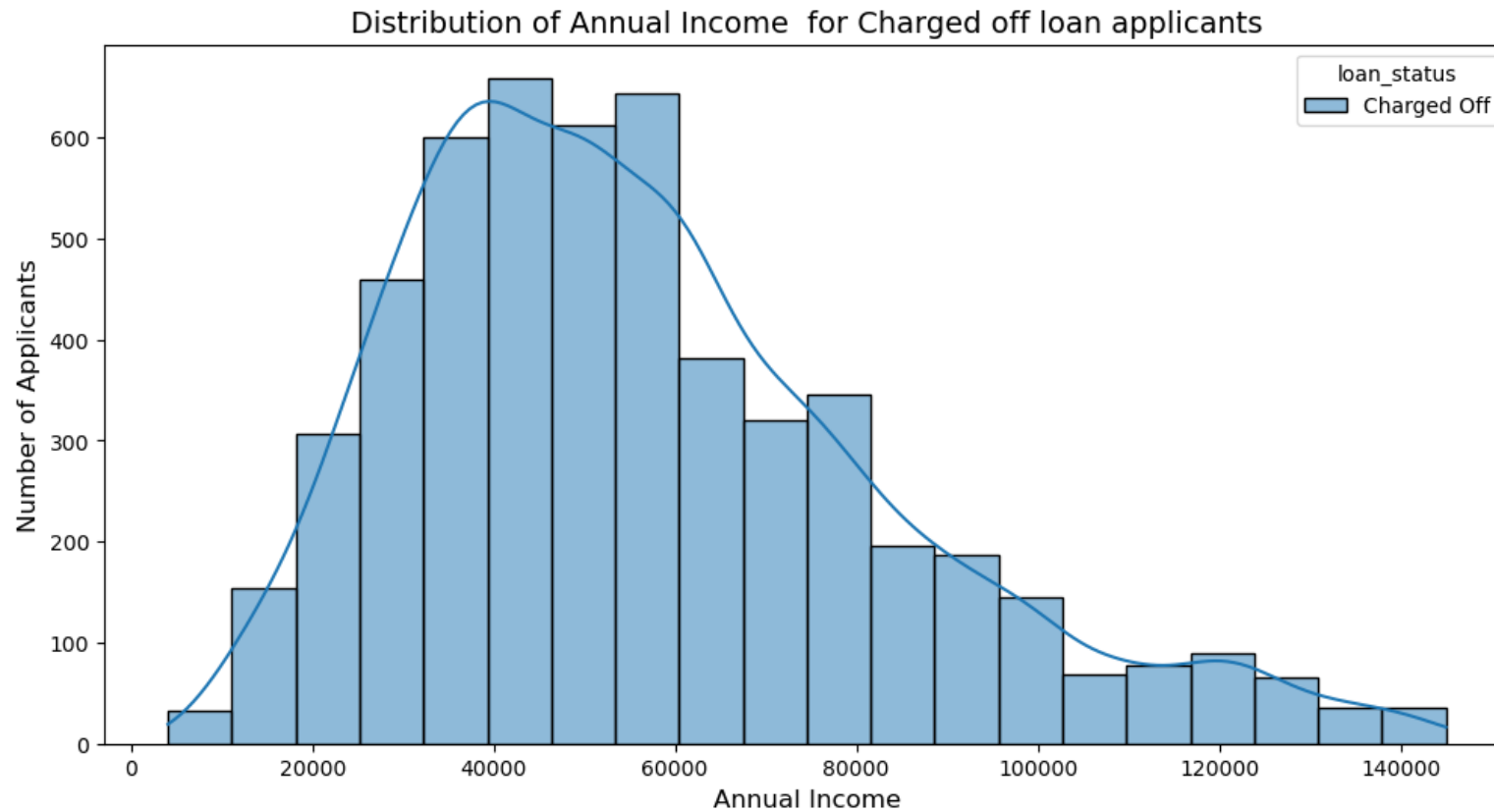
Insights

Most of the charged off loan applicants had annual salaries between 35,000 to 60,000 USD.
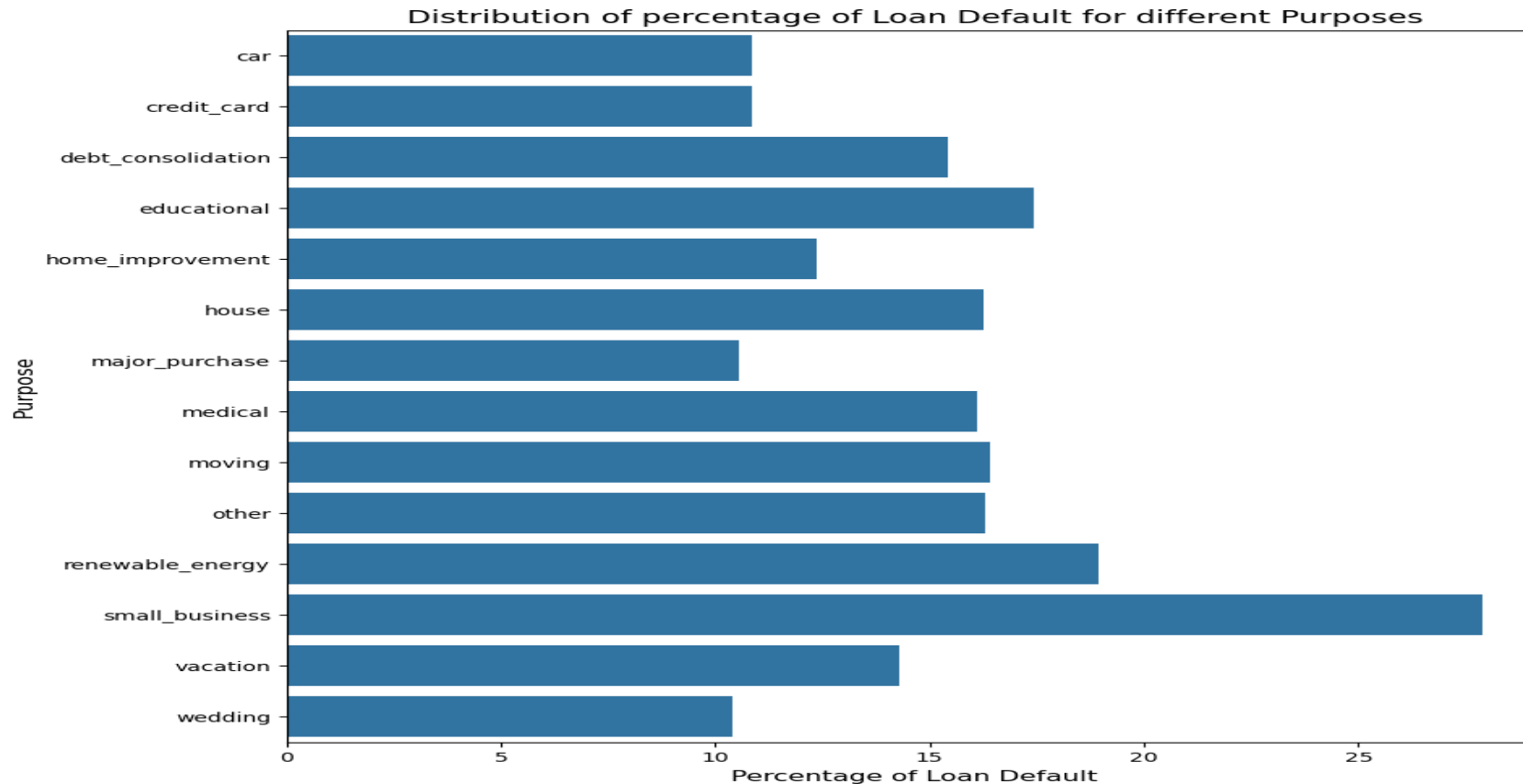
.

Bivariate Analysis and Multivariate Analysis

Distribution of percentage of Loan Default for different Grades

Distribution of percentage of Loan Default for different Grades

Insights

The Percentage of loan defaults increase with grade ,Grade E,F,G contributes to more percentage of Loan Defaults.

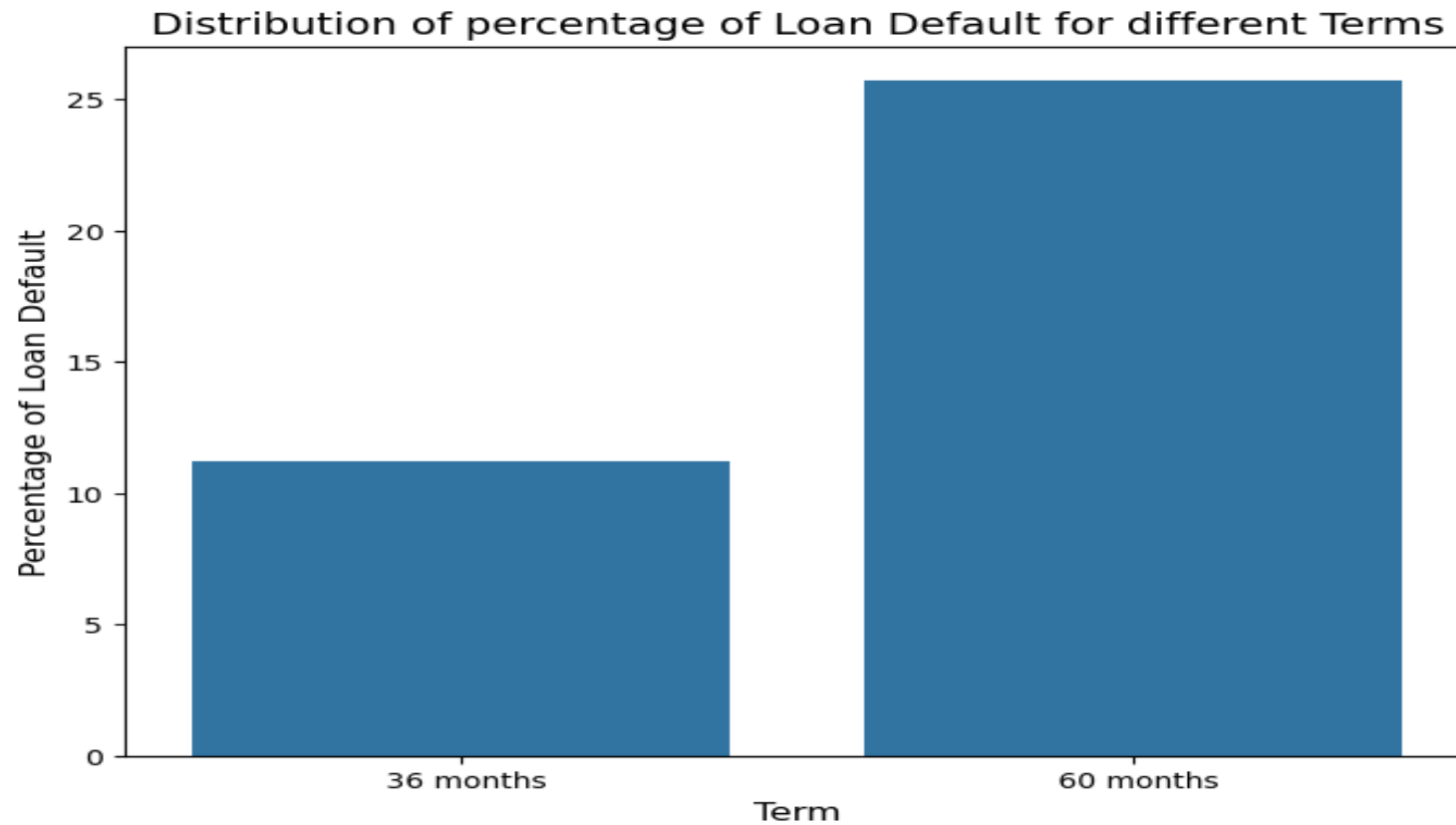# Distribution of percentage of Loan Default for different Purposes

Distribution of percentage of Loan Default for different Purposes

Insights

Small Business is the category where the maximum percentage of loans are defaulted.

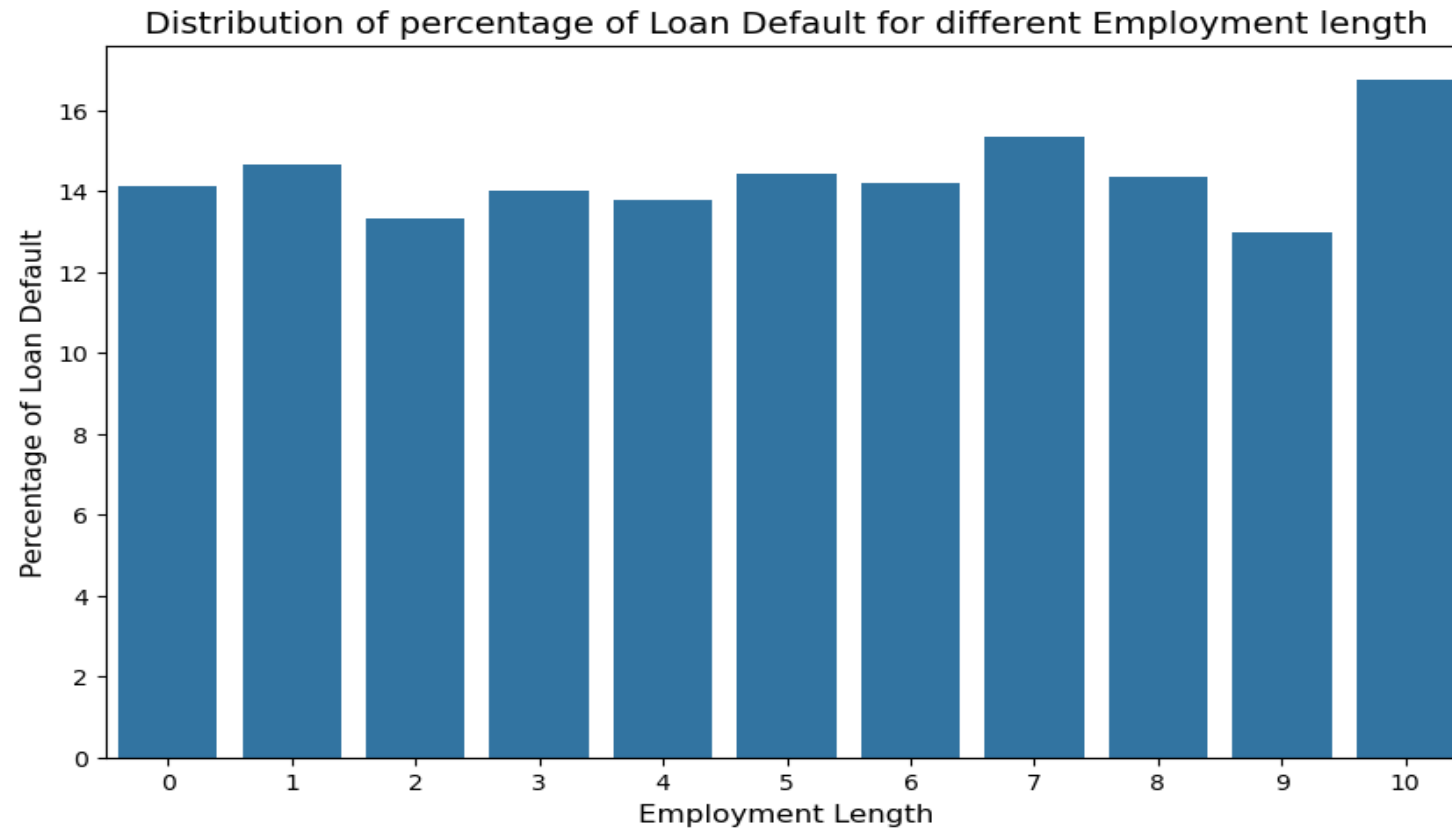# Distribution of percentage of Loan Default for different Terms

Distribution of percentage of Loan Default for different Terms

Insights

Loan applicants applying loan for 60 months are likely to default more than the one taking loan for 36 months.

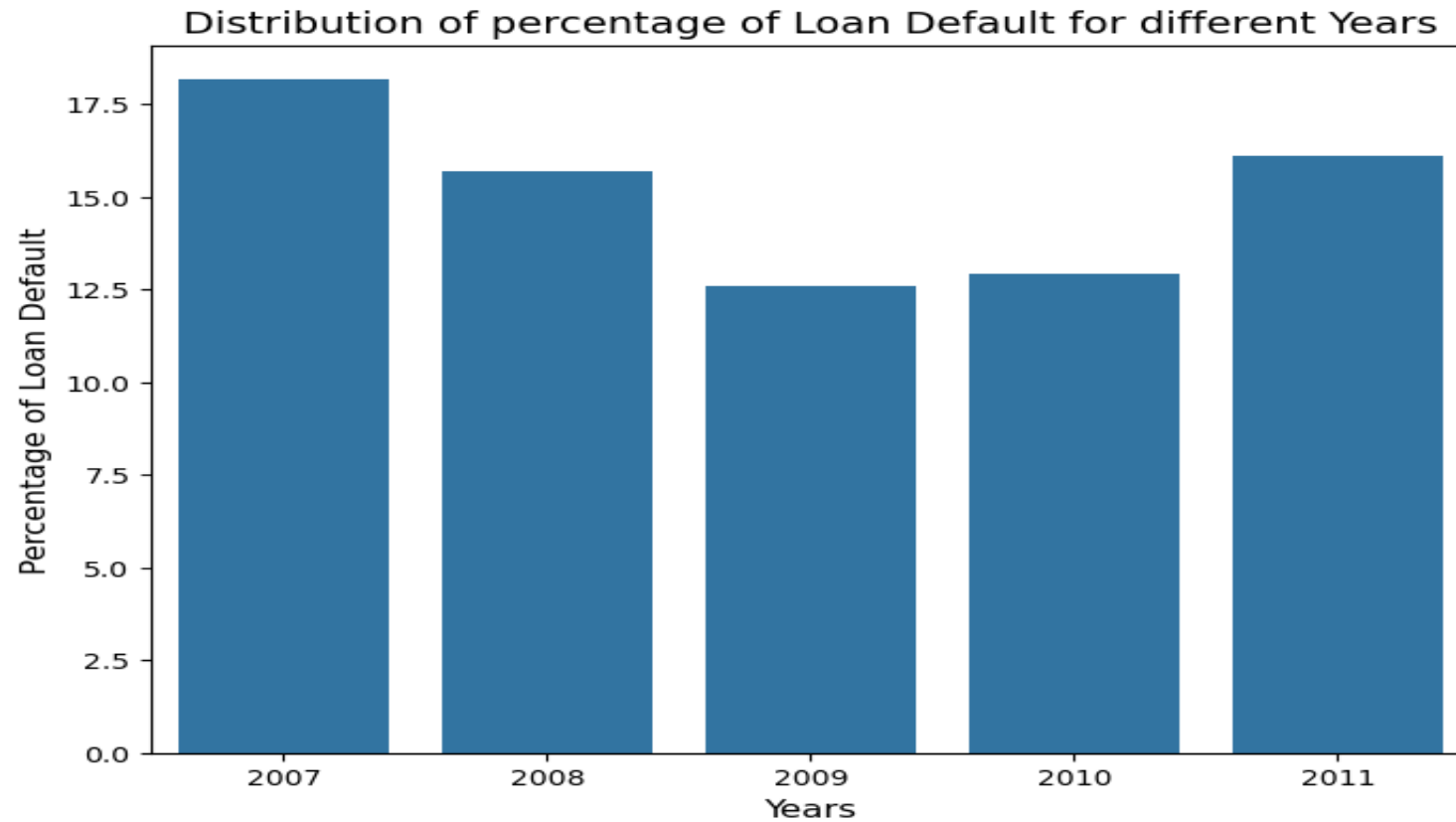# Distribution of percentage of Loan Default for different Employment length

Distribution of percentage of Loan Default for different Employment length

Insights

Most loan applicants have ten or more years of experience, and they are also the most likely to default.

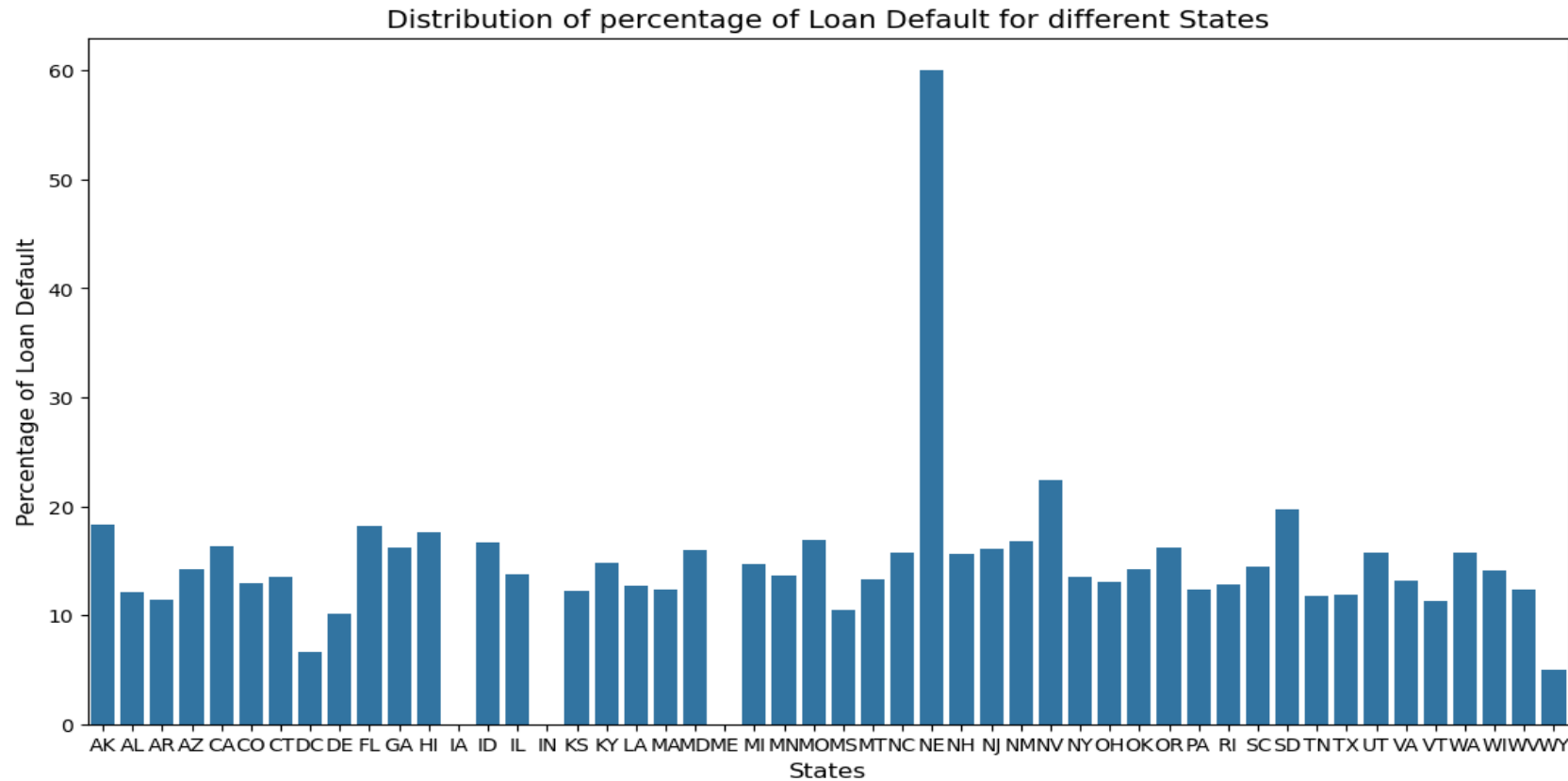# Distribution of percentage of Loan Default for different Years

Distribution of percentage of Loan Default for different Years

Insights

The percentage of loan Defaults is in reducing pattern from 2007 to 2009 and it increased in 2011 drastically.

Distribution of percentage of Loan Default for different States
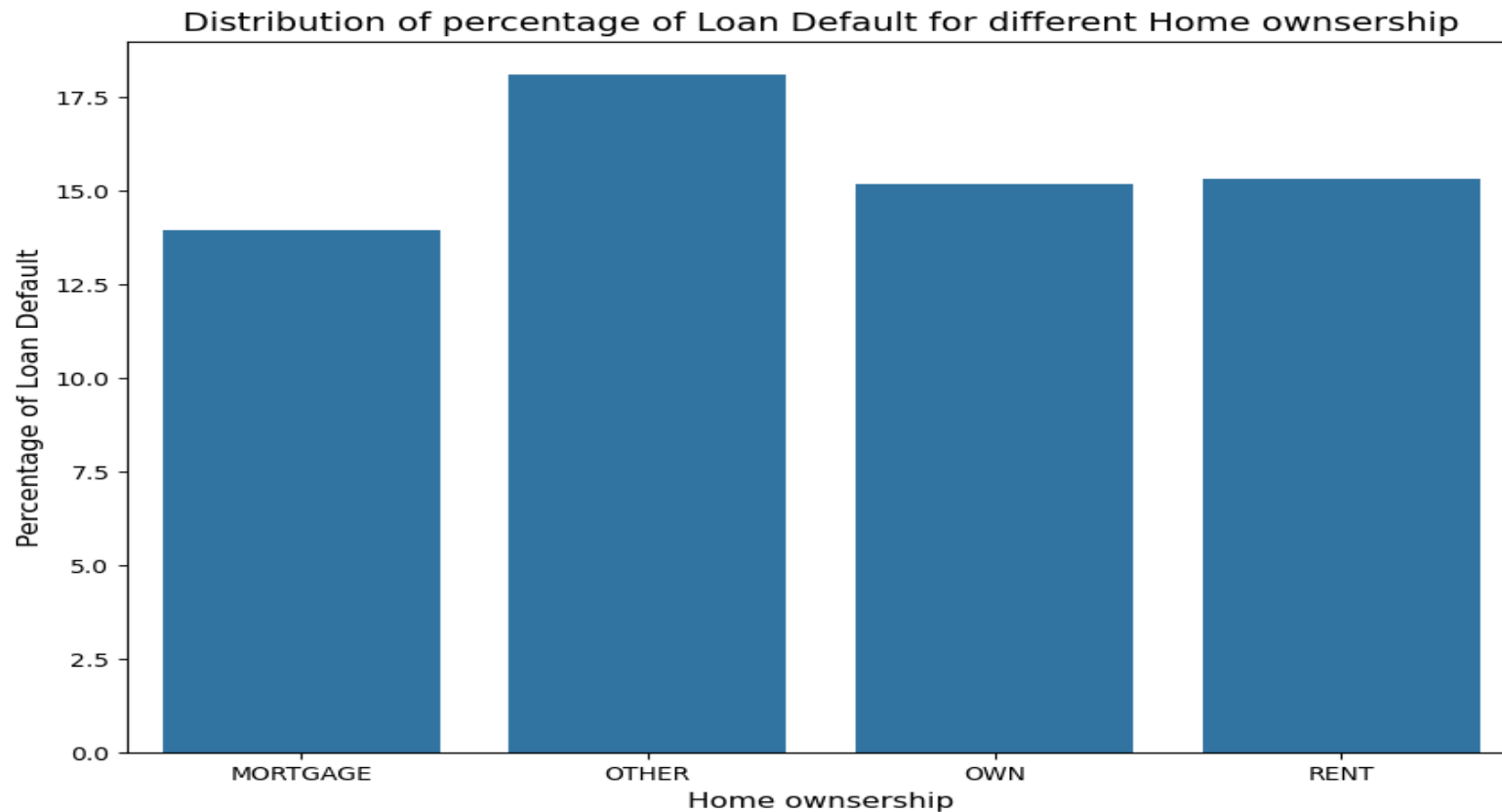

Distribution of percentage of Loan Default for different States

Insights

Loan applicants from the state of Nebraska(NE) are most likely to default.

Distribution of percentage of Loan Default for different Home ownership
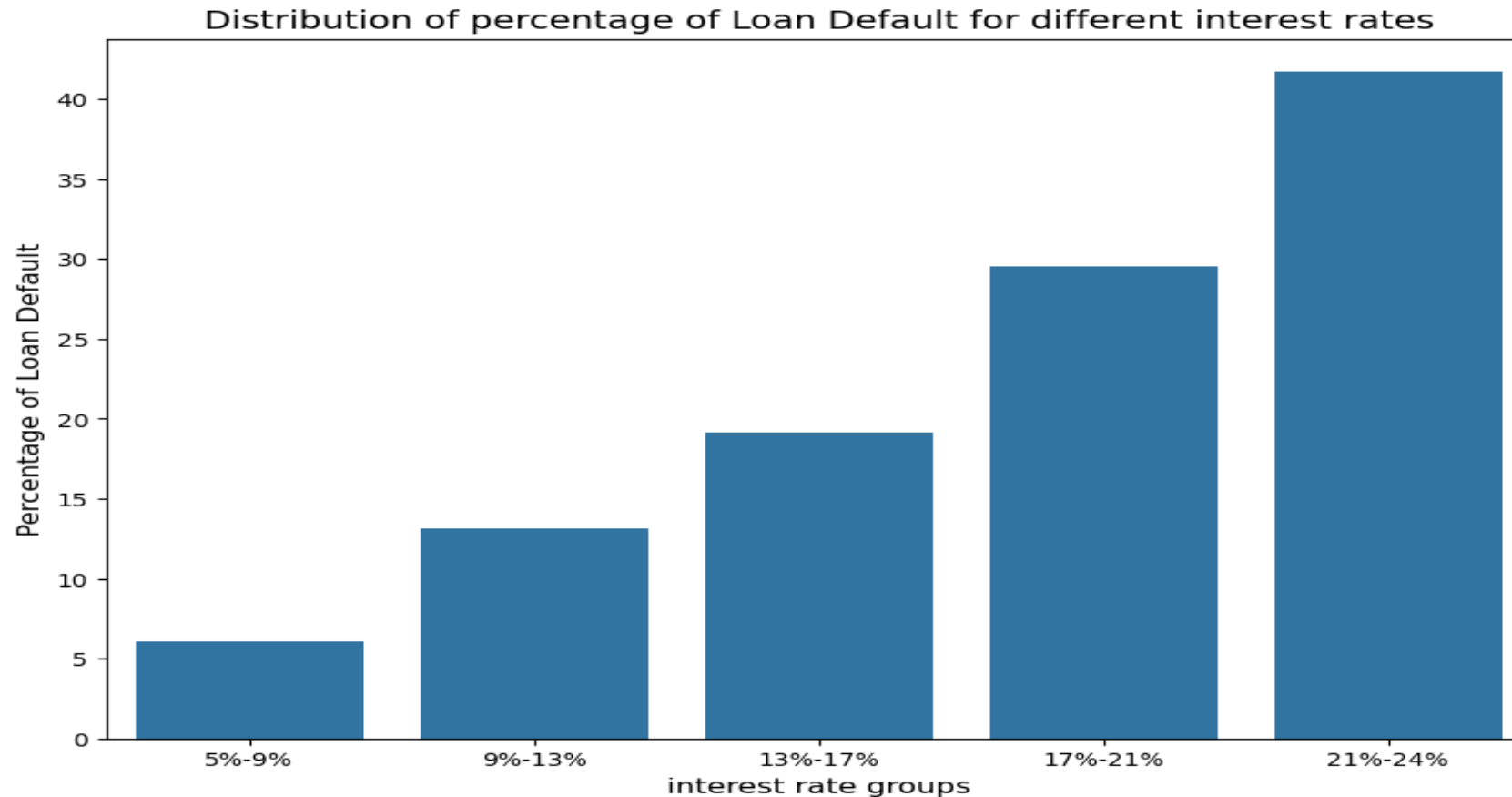

Distribution of percentage of Loan Default for different Home ownership

Insights

Loan applicants who mentioned other as home ownership are more likely to default.

Distribution of percentage of Loan Default for different interest rates
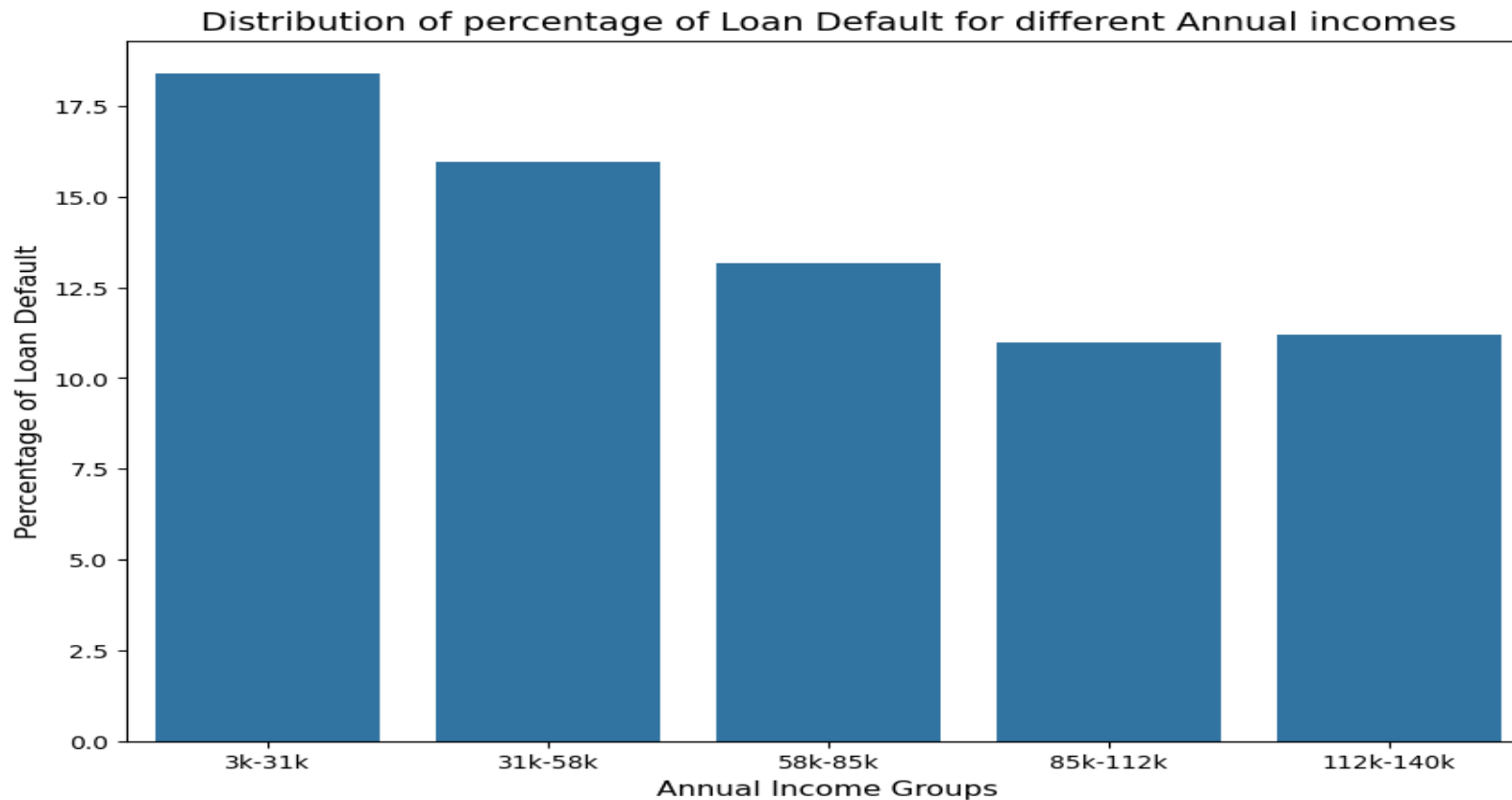

Distribution of percentage of Loan Default for different interest rates

Insights

The percentage of loan Defaults is in increasing pattern with increase in interest rates,21%-24% range are mostly associated with defaults.

Distribution of percentage of Loan Default for different Annual incomes

Distribution of percentage of Loan Default for different Annual incomes

Insights

A majority of the loan applicants who charged off reported an annual income of less than $58,000.

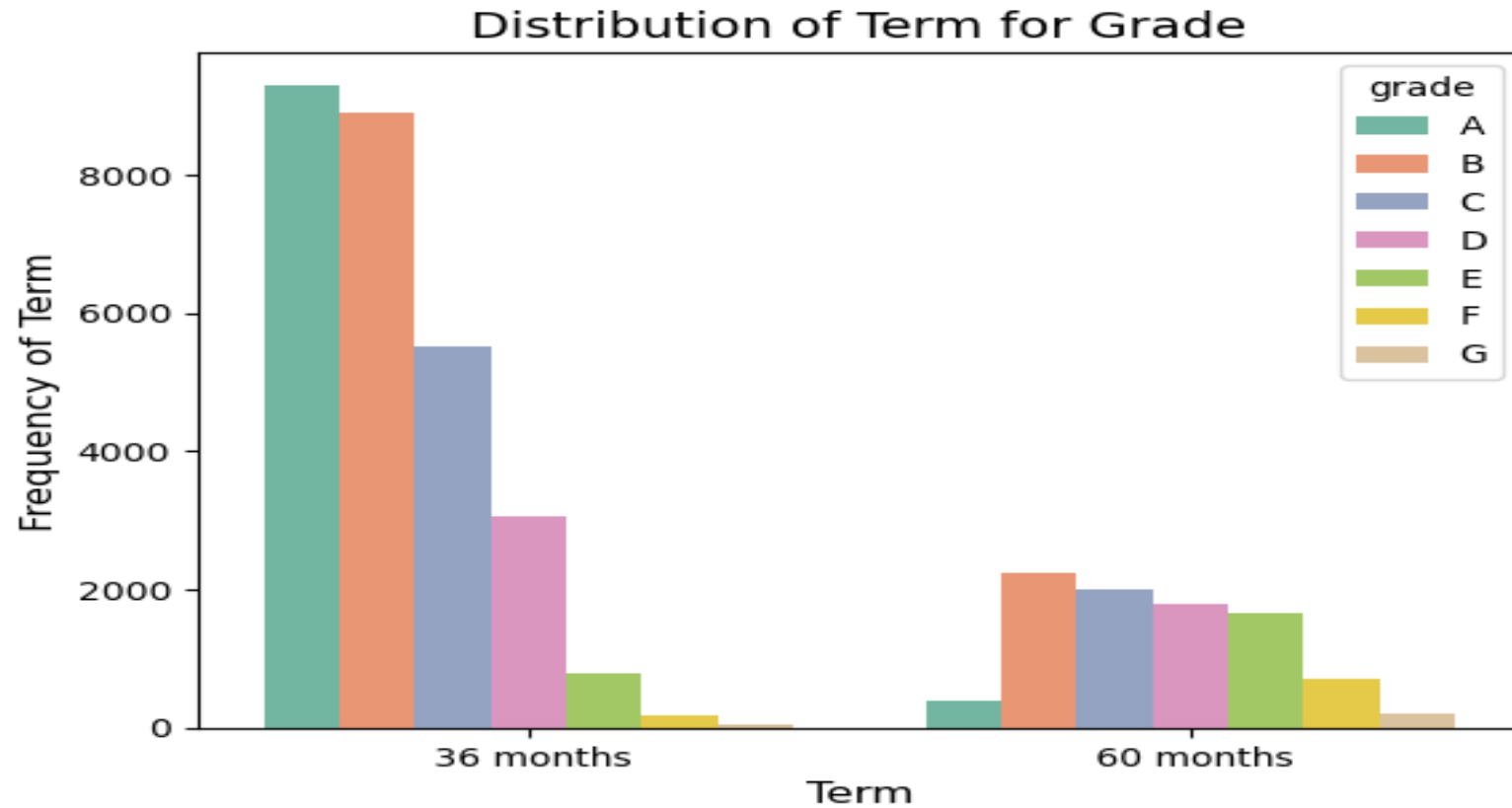# Distribution of percentage of Loan Default for different Loan Amounts

Distribution of percentage of Loan Default for different Loan Amounts
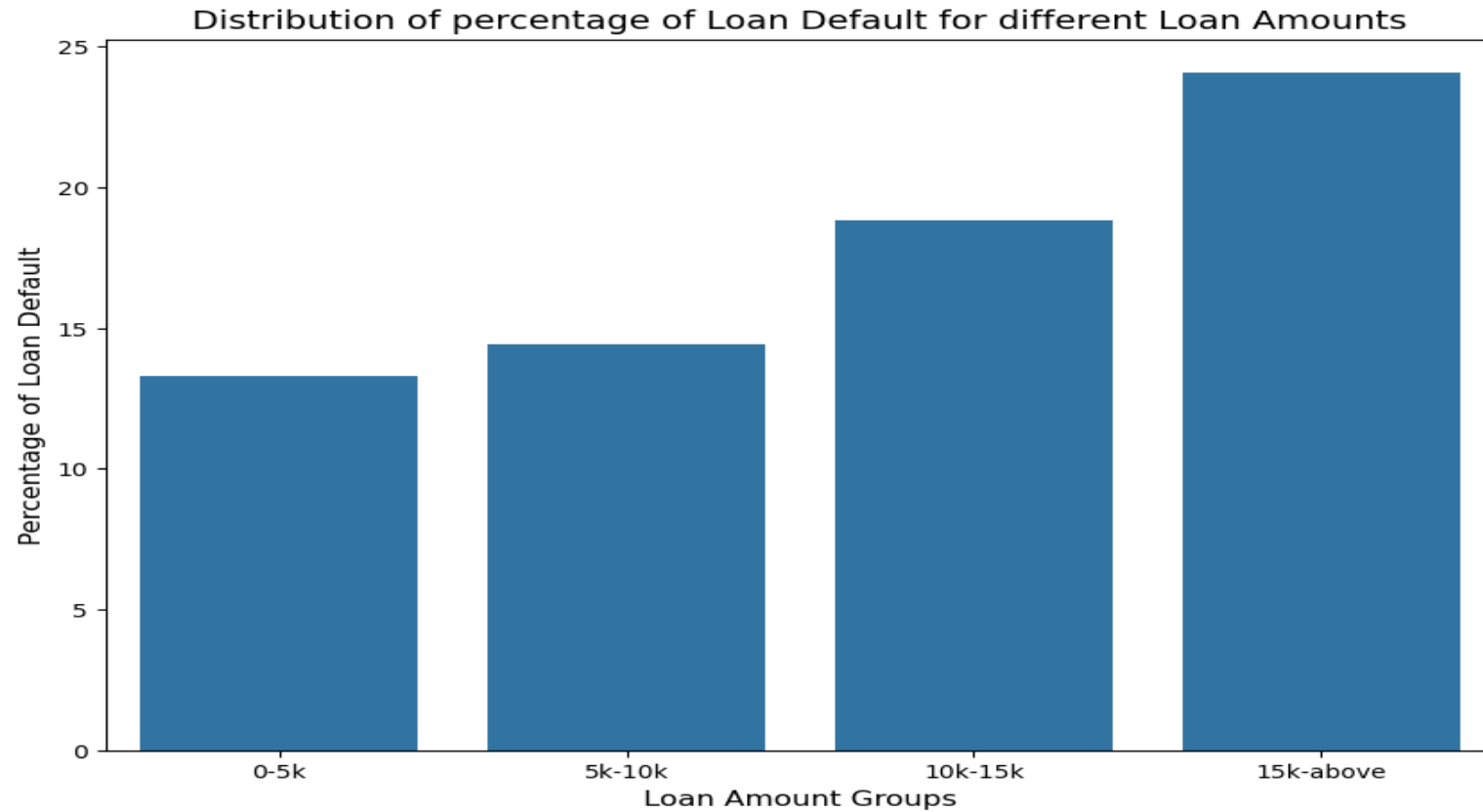
Insights

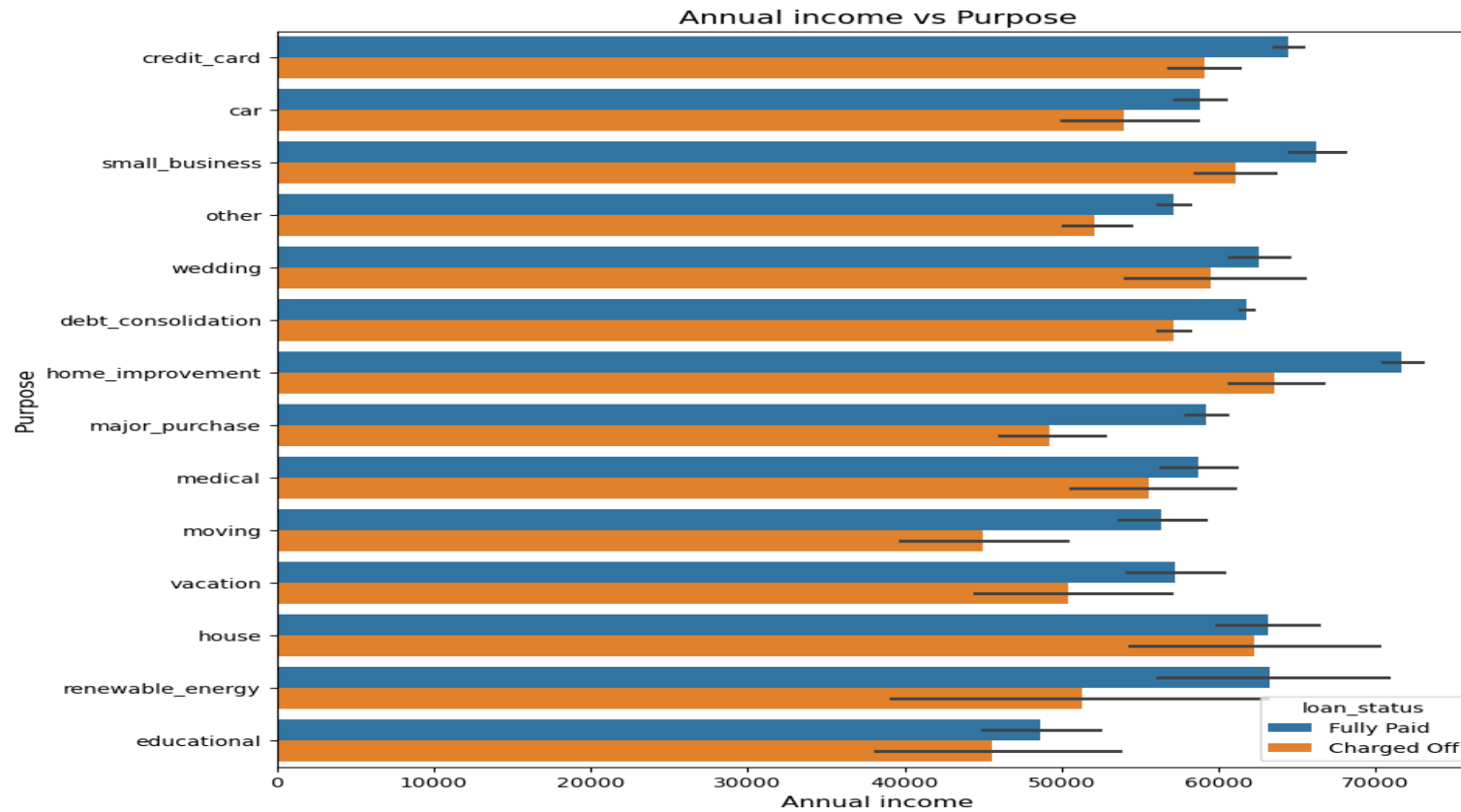Majority of the loan applicants who defaulted received loan amounts of $15,000 or higher.

Distribution of Term for Grade

Distribution of Term for Grade

Insights

Grade A and B for short term loan has high chance of defaulting.
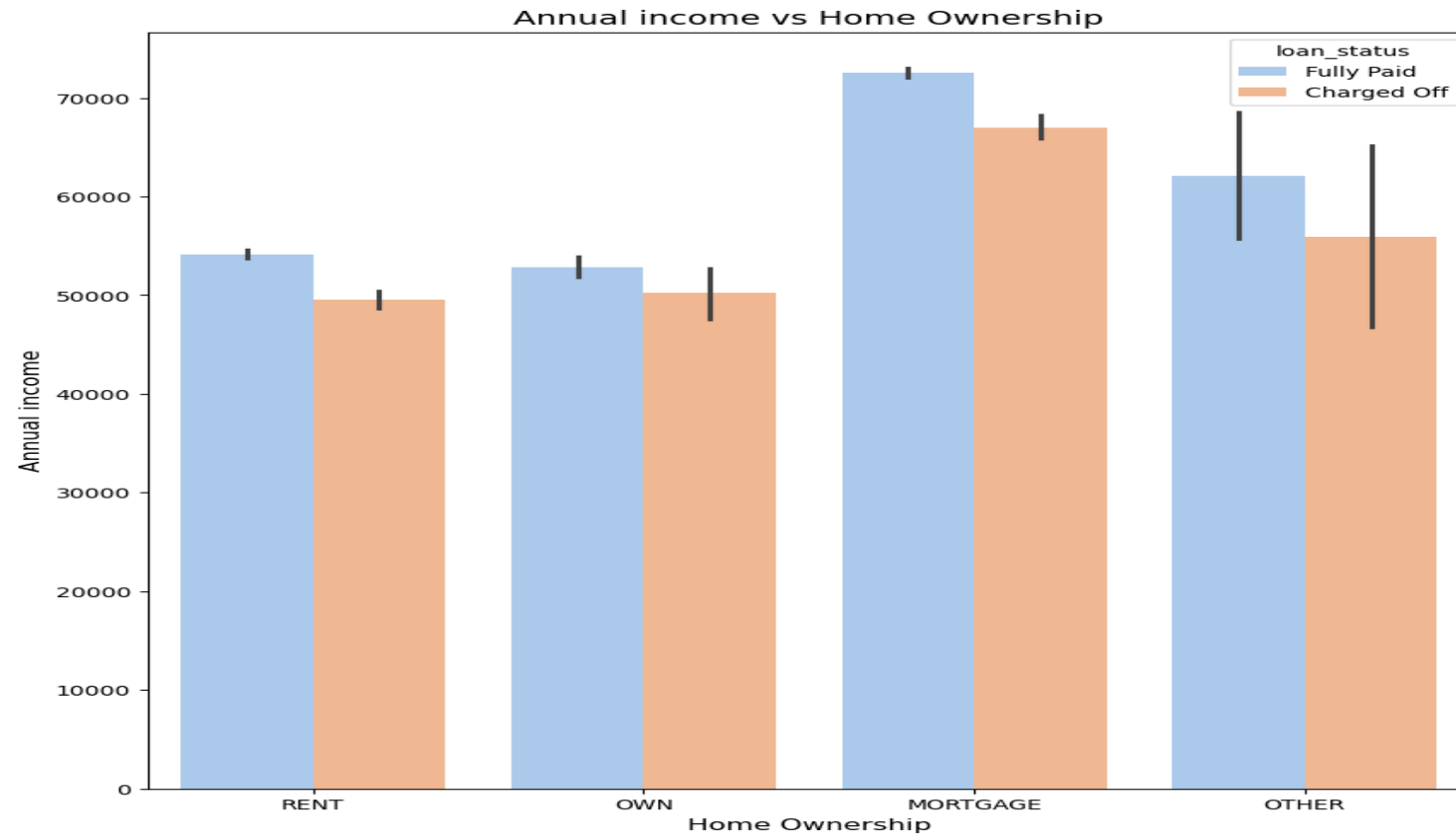
Annual income vs Purpose

Annual income vs Purpose

Insights

Applicants taking loan for 'home improvement' and have income of 60k -70k, default most.
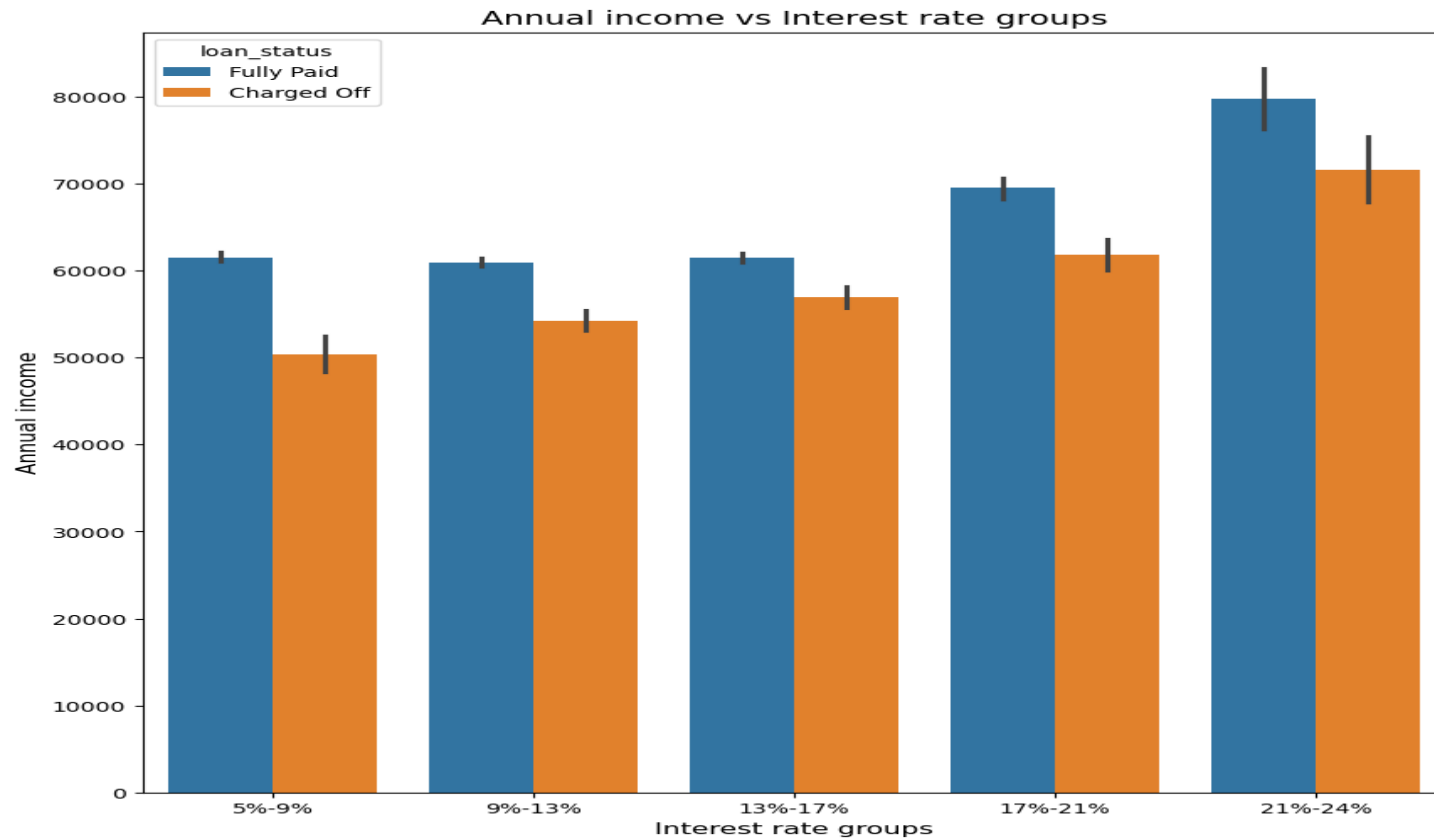
Annual income vs Home Ownership

Annual income vs Home Ownership

Insights

Applicants whose home ownership is 'MORTGAGE and have income of 60-70k default the most.
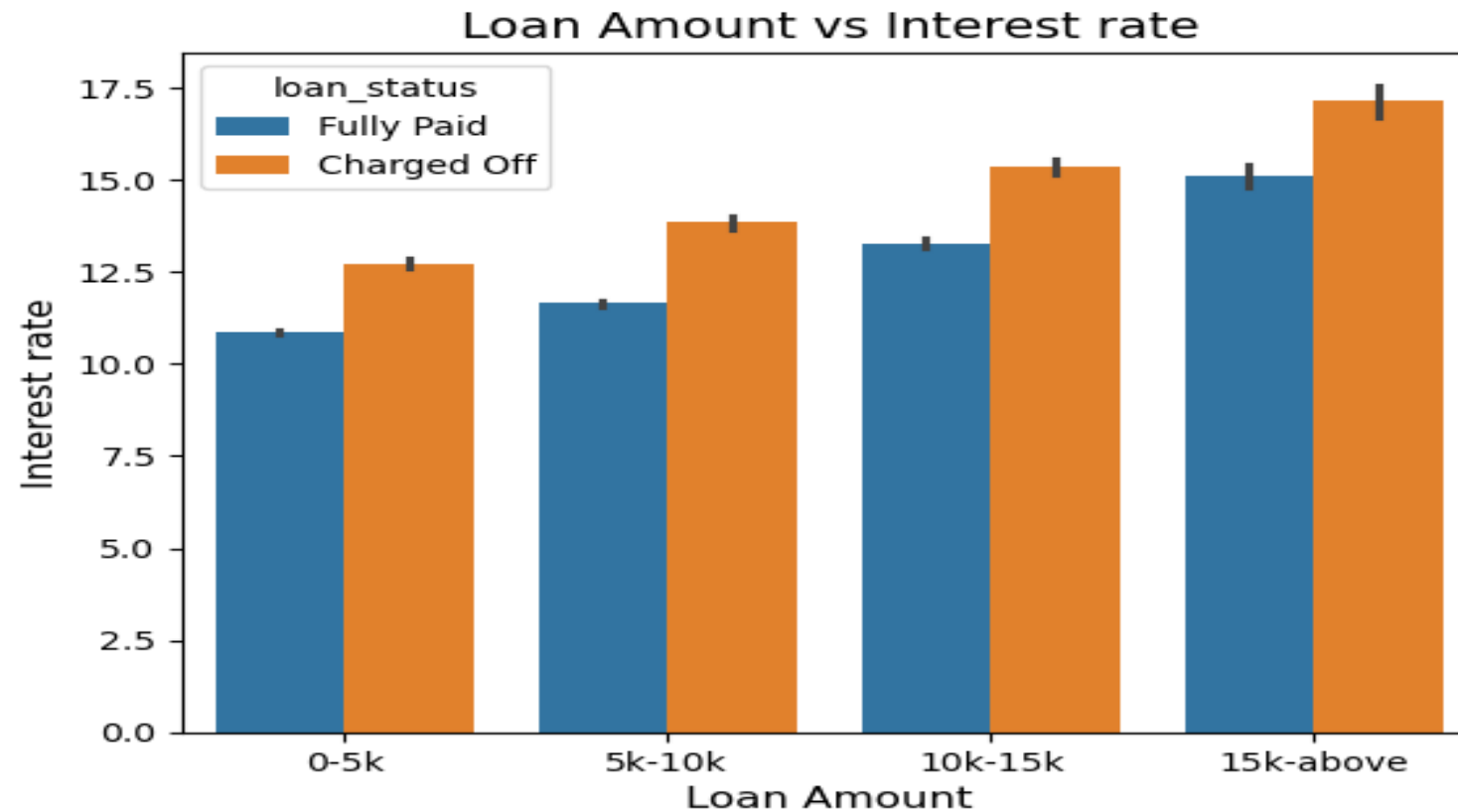
Annual income vs Interest rate groups

Annual income vs Interest rate groups

Insights

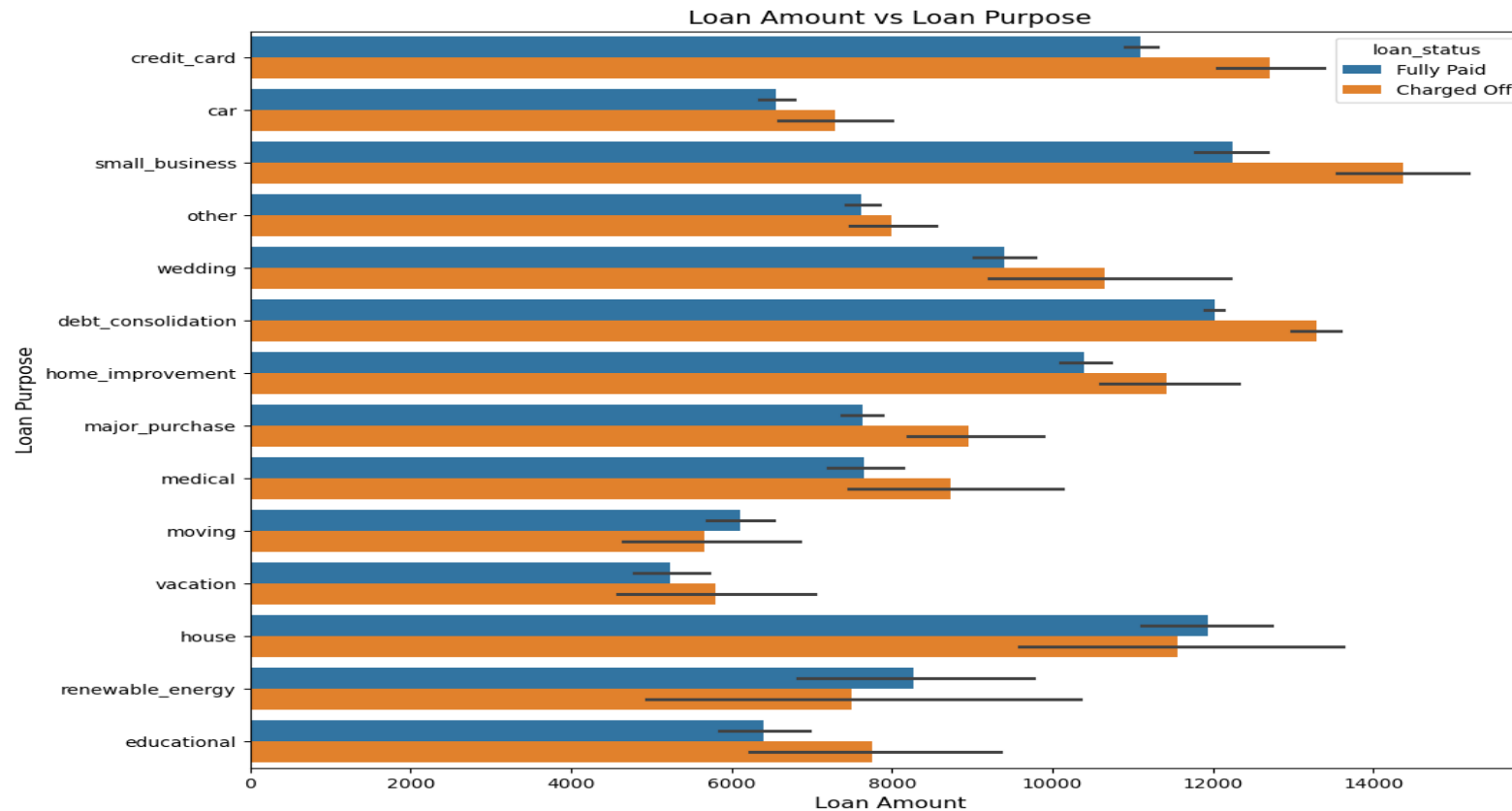Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k, default most.

Loan Amount vs Interest rate

Insights

Applicants who have taken a loan in the range 15k -above and are charged interest rate of 15-17.5 % Default most.
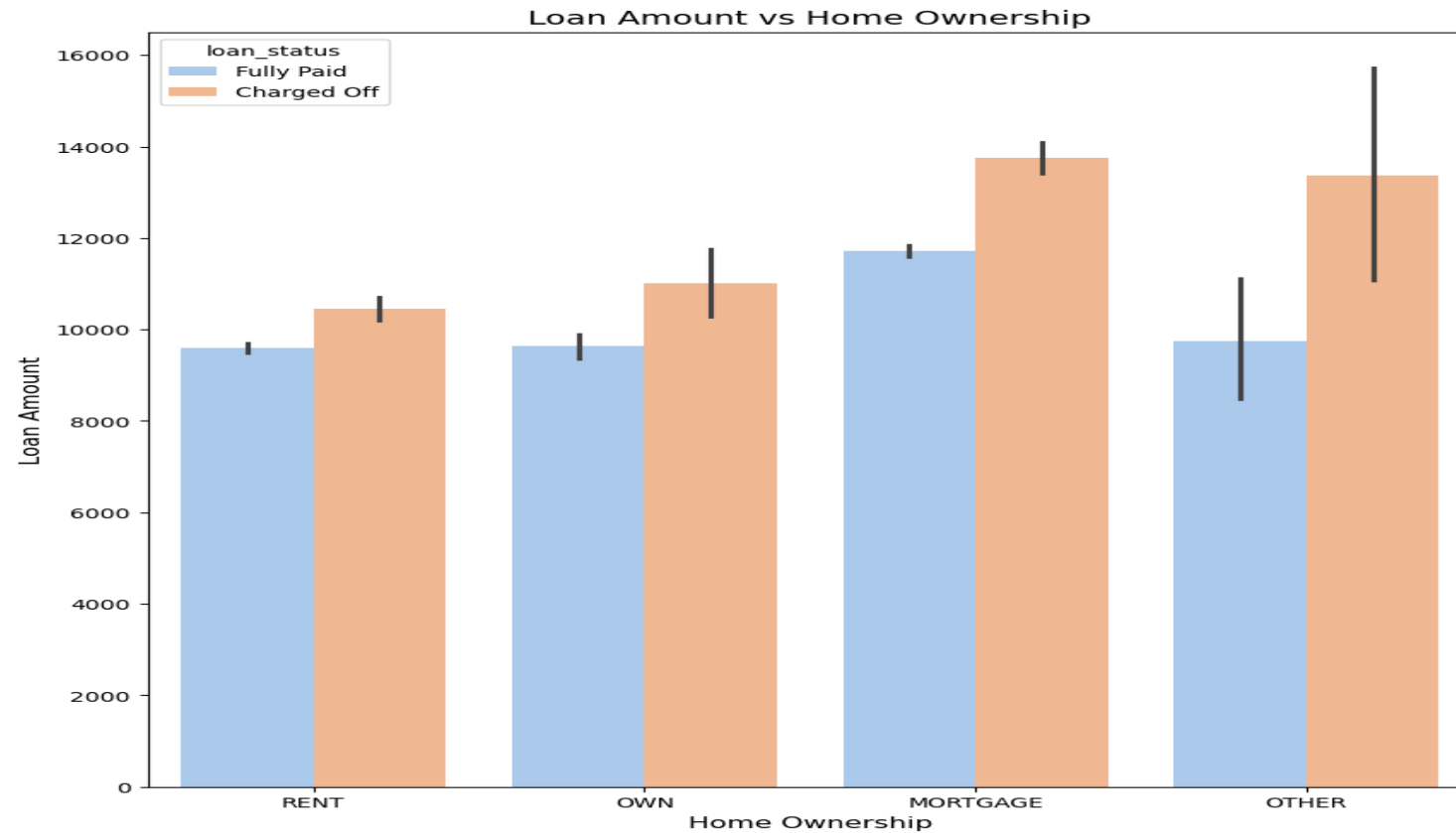
Loan Amount vs Loan Purpose

Insights

Applicants who have taken a loan for small business and the loan amount is greater than 14k default most.
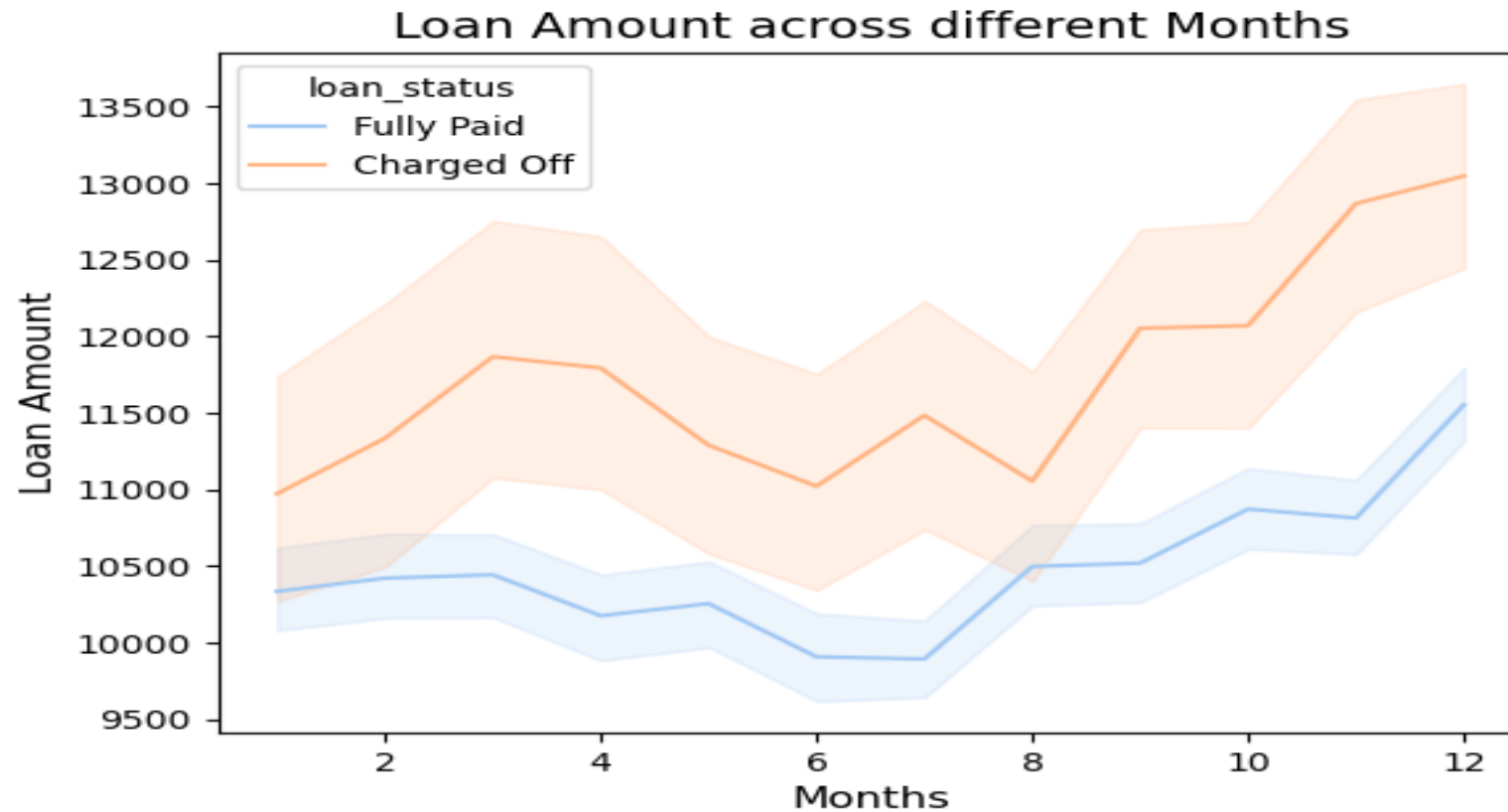
Loan Amount vs Home Ownership

Loan Amount vs Home Ownership

Insights

Applicants whose home ownership is 'MORTGAGE and have loan of 14-16k default most.
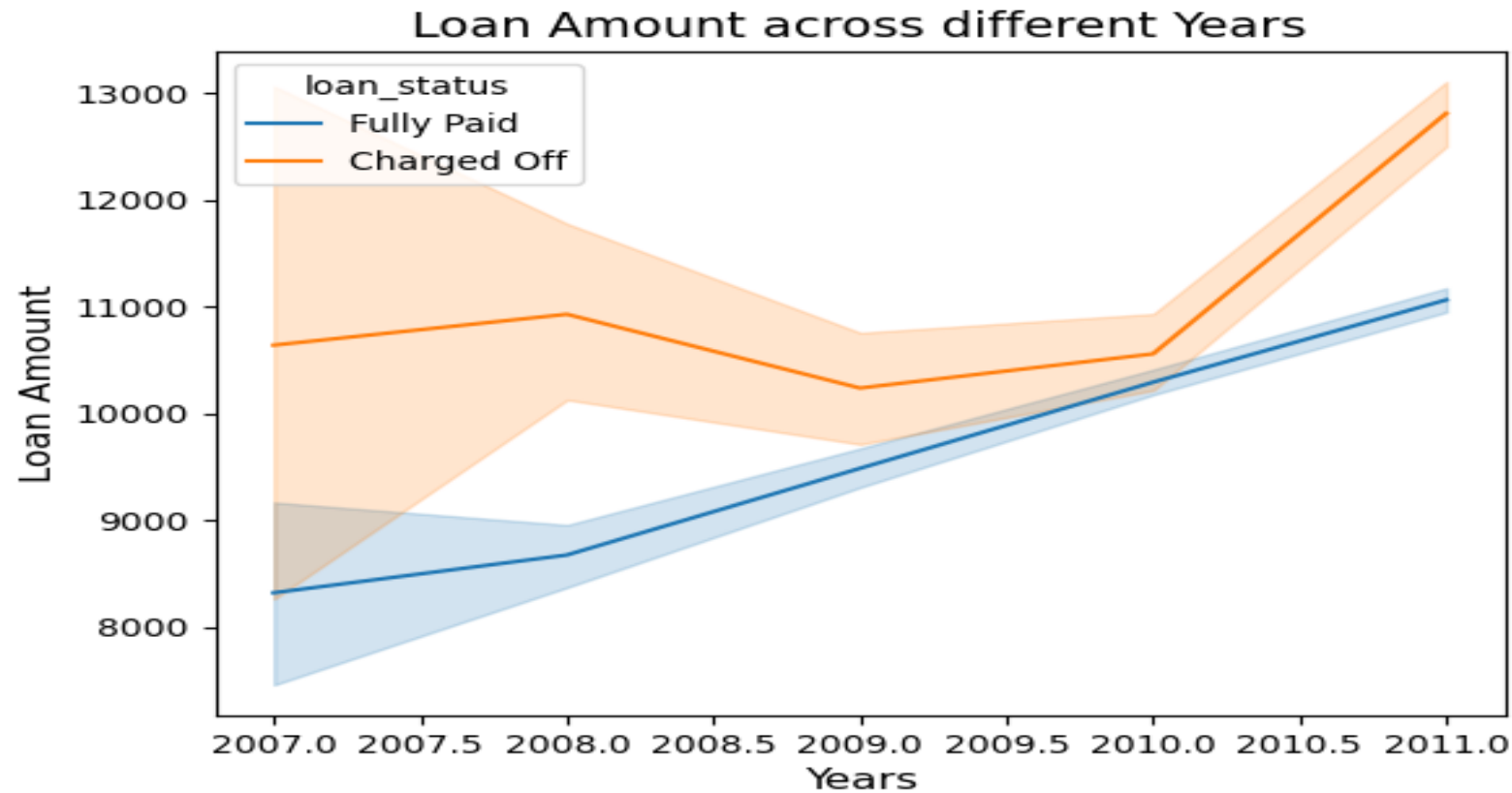
Loan Amount across different Months

Insights

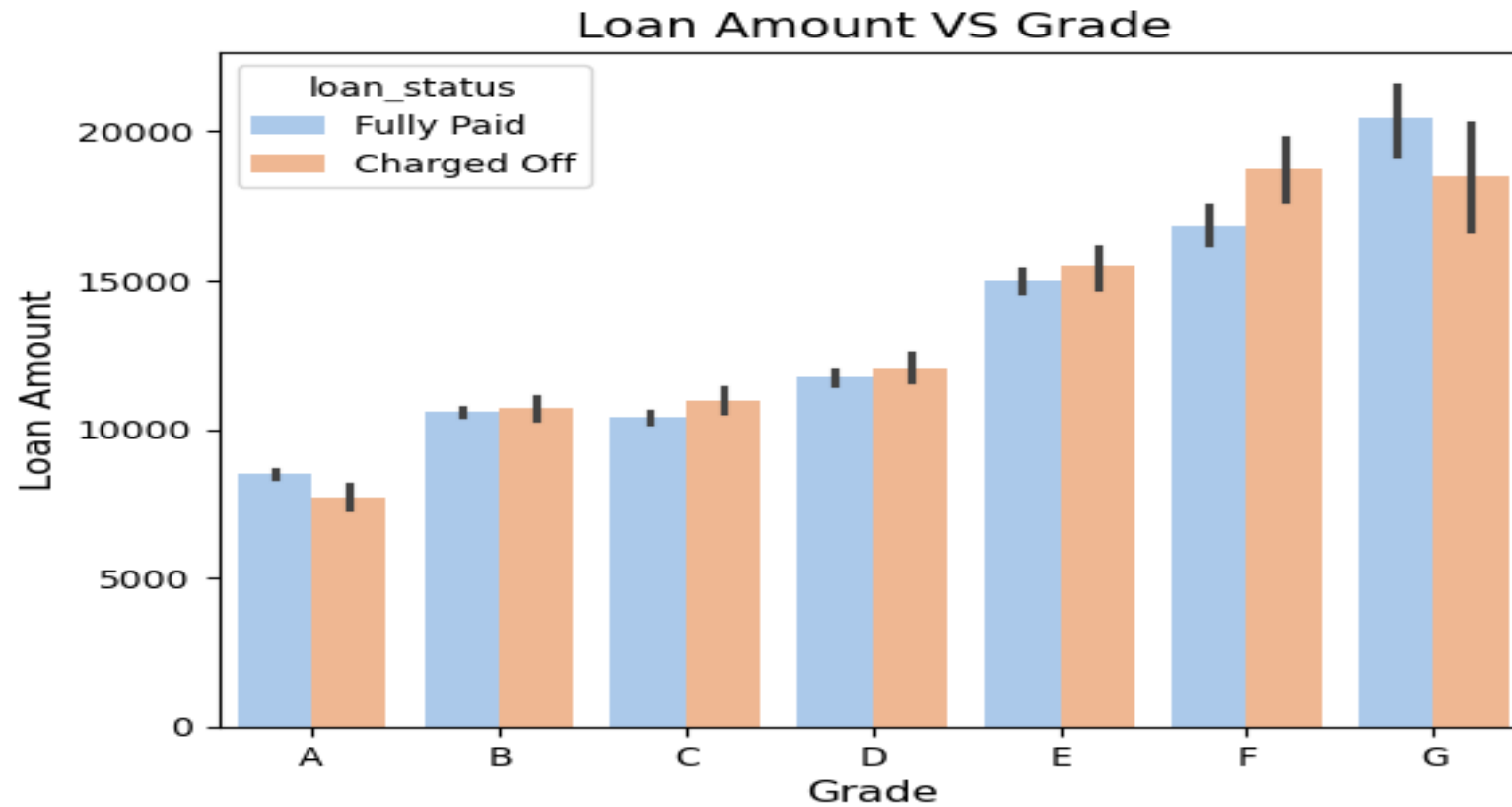Loan taken in December had highest defaults.

Loan Amount across different Years

Insights
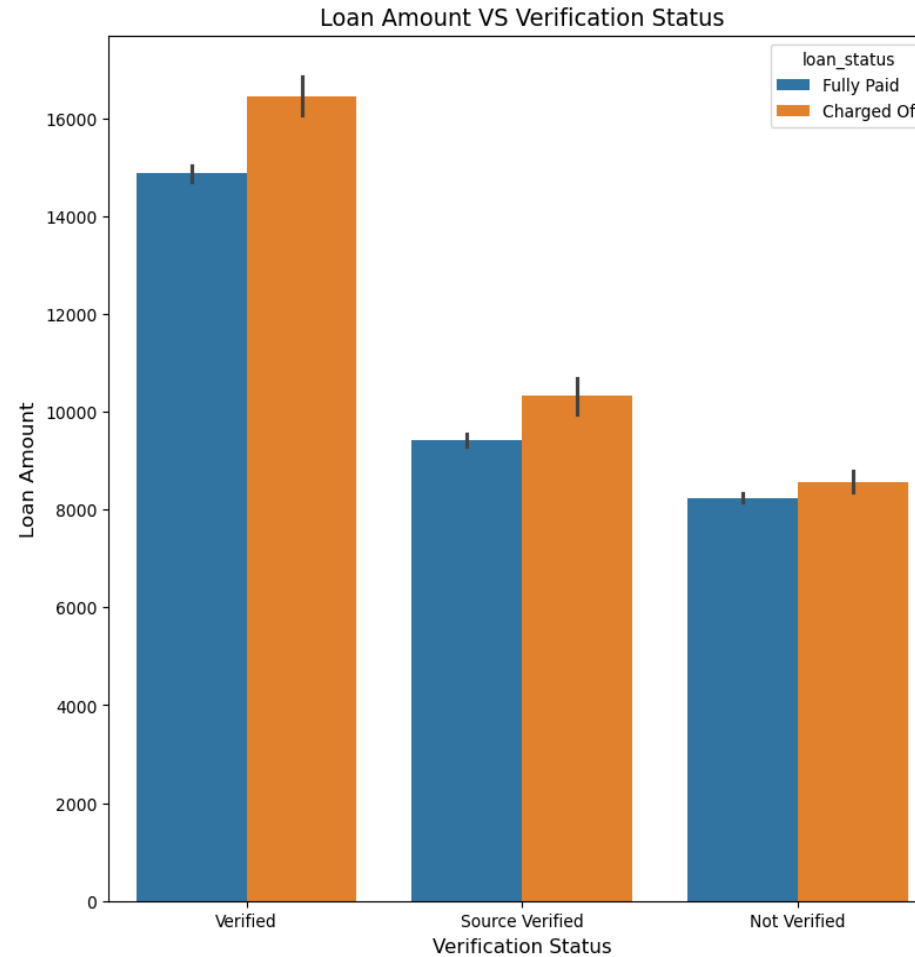
2011 had most defaults dues to some crisis.

Loan Amount VS Grade

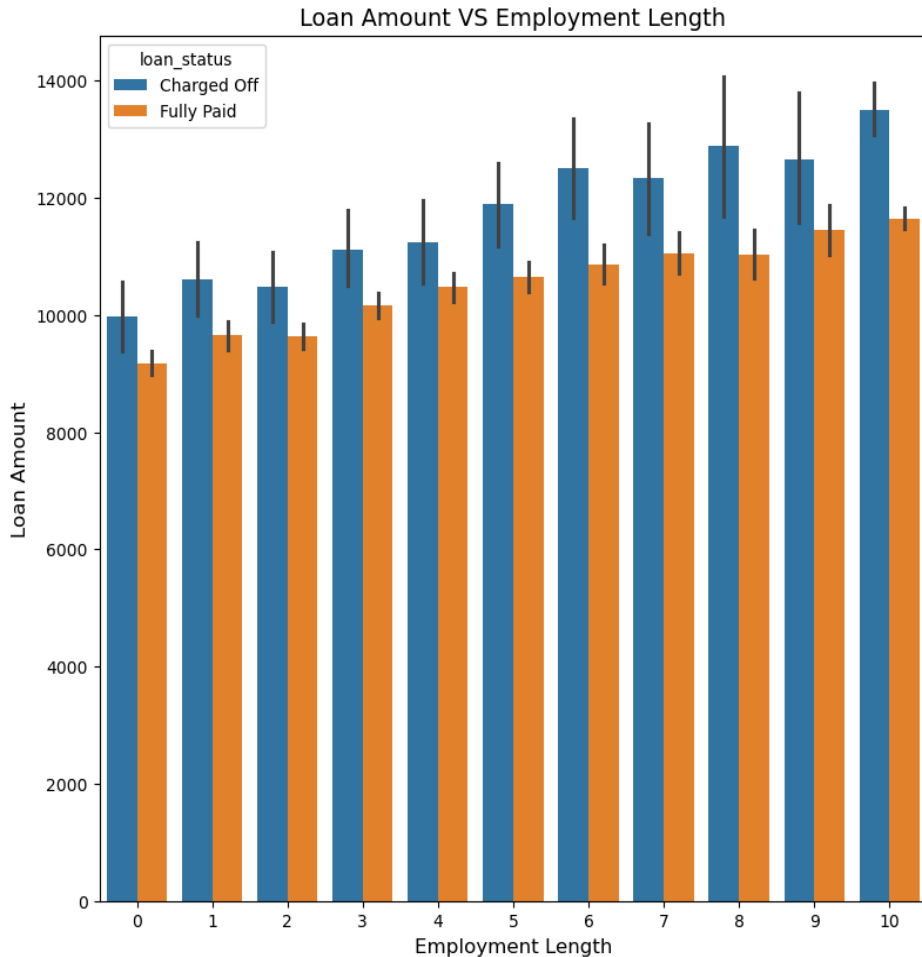**Insights**

When grade is F,G and loan amount is between 15k-20k Had most defaults.

Loan Amount VS Employment Length Loan Amount VS Verification Status
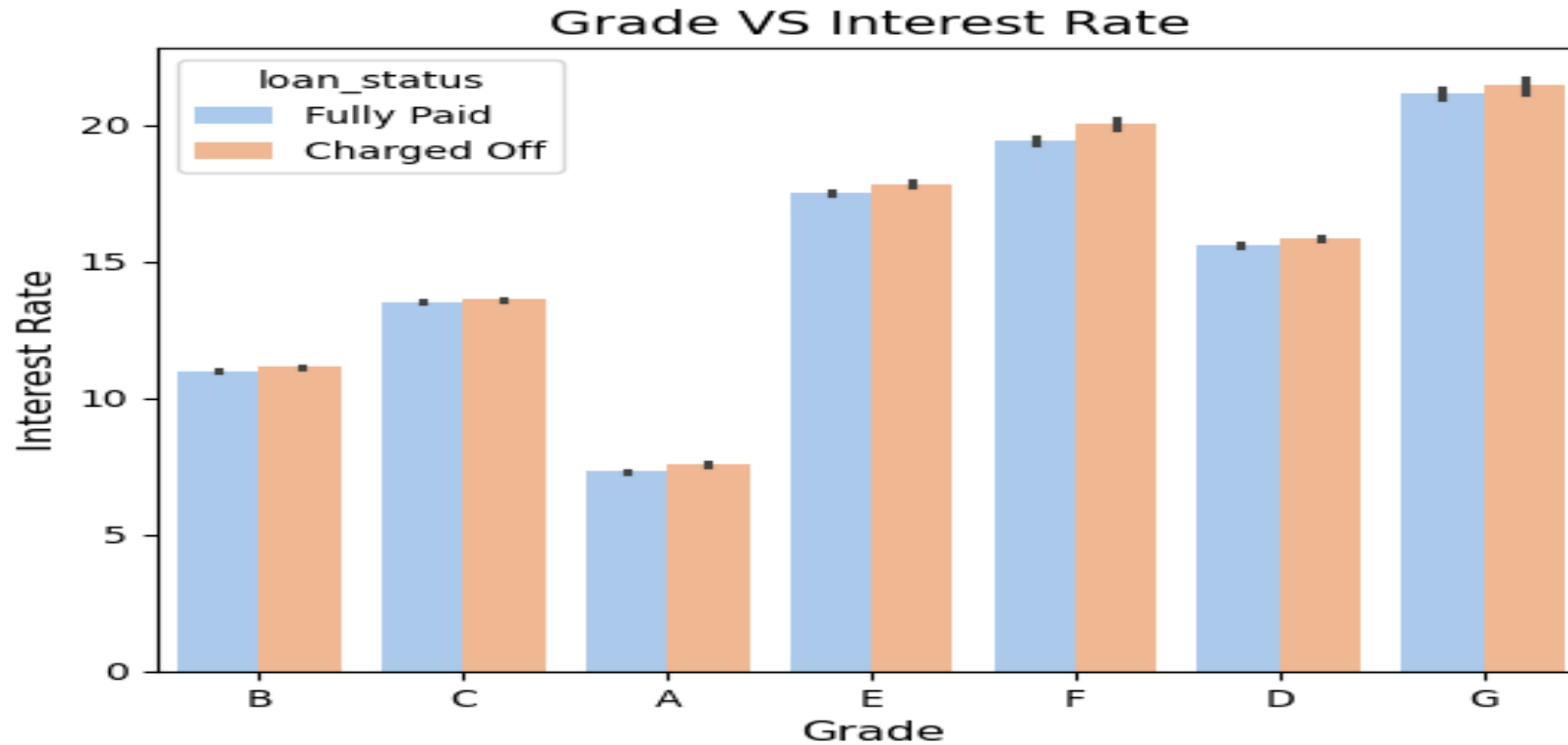

Loan Amount VS Employment Length

Loan Amount VS Verification Status

Insights

➢ When employment length is 10yrs and loan amount is 12k-14k had most default.

➢ When the loan is verified and loan amount is above 16K had most default.

Grade VS Interest Rate

Insights

Grade G and interest rate above 20% had max default.

Recommendations

- ➢ Loan applicants applying loan for 60 months are likely to default more than the one taking loan for 36 months.
- ➢ Applicants who had been employed for more than 10 years accounted for the highest number of "Charged off" loans. This indicates that long-term employment history did not necessarily guarantee successful loan repayment.
- ➢ The majority of "Charged off" loan participants lived in rented houses or mortgage.
- ➢ Most of the charged off loan applicants had annual salaries between 35,000 to 60,000 USD
- ➢ Applicants taking loan for 'home improvement' and have income of 60k -70k likely to default
- ➢ Applicants whose home ownership is 'MORTGAGE and have income of 60-70k likely to default
- ➢ Applicants who receive interest at the rate of 21-24% and have an income of 70k-80k are likely default
- ➢ When grade is F,G and loan amount is between 15k-20k likely to default
- ➢ When the loan is verified and loan amount is above 16K likely to default
- ➢ For grade G and interest rate above 20% likely to default
- ➢ Loan applicants from the state of Nebraska(NE) have high percentage of default.
- ➢ Borrowers from states CA, FL, NJ have maximum number of loan defaults

Recommendations Cont.

➢ Most of the loan participants who charged off has Dti is between 13-17
➢ Small Business is the category where the maximum percentage of loans are defaulted.
➢ A majority of the loan applicants who charged off reported an annual income of less than $58,000.
➢ The percentage of loan Defaults is in increasing pattern with increase in interest rates,21%-24% range are mostly associated with defaults.
➢ Majority of the loan applicants who defaulted received loan amounts of $15,000 or higher.
➢ The Percentage of loan defaults increase with grade ,Grade E,F,G contributes to more percentage of Loan Defaults.