Lending Club Case Study

Case Study Objective

The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc

Python Libraries

To perform the EDA on the given dataset we are using python libraries like

- numpy
- pandas
- datetime

For plotting the extracted data on different graphs we will using below libraries

- seaborn
- matplotlib
- plotly

Data Loading

- The given dataset contains the loan details of the lending company from 2007 to 2011.
- In data loading we are reading csv file into a data frame for analysis.

Data Understanding

Shape

Data set has 39717 rows and 111 features.

Details

Dataset has important features which will be considered in this analysis.
Loan amount, Approved amount, Investor Amount (actual given loan amount), Loan term, Interest rate, Monthly installment amount, Loan Grade and Subgrades, Borrowers employment title, Employment experience, Home ownership (Own, Rent, Mortgage), Annual Income, Income source verification status, Loan issue date, Loan status, Purpose of loan, Title of loan, Address state code, Debt to income ratio.
Apart from above features this dataset also contains the behavioral features like delinquency, Charged Off with in 12 months etc. These features are not available at the time of loan approval, so these behavioral features are not considered in this analysis.

Data Cleaning

Missing Values

All the features which has only null values has no significance in this analysis, so need to be dropped from the dataset.
Also features which has missing values above 60% can create biased results. So these features also need to be dropped from the dataset.

Unique Values

The features which has unique values do not have any correlation with other features. So those features need to be dropped from the dataset.
Zip code has encrypted values so need to be dropped from dataset.
Desc feature requires NLP to derive useful metrics, it is dropped from dataset for time being.
Url feature has loan_id as meaning full information, but it is also unique value, so dropped.

Missing Value Imputation

The missing values in emp_title are filled with Untitled.
The missing values in emp_length & title are filled with Unknown.

Data Correction

Integer Values

Term feature converted into data type integer by removing months keyword from data values.

Float Values

The interest rate feature converted into float data type by removing `%` sign from the data values.

Data Deduplication

There are no duplicate rows found in the dataset.

Derived Metrics

Monthly Installment Amount Percentage

This derived metric will be used to analyze whether this higher percentage is increasing the count of default applications.

Code :

```
loan_df['monthly_inst_percentage'] = (loan_df['installment']/(loan_df['annual_inc']/12))*100
```

Binning

Continuous variable like dti, int_rate & annual_inc are binned to analyze trend with other features.

Code:

```
loan_df['dti_bin'], cut_bin = pd.qcut(loan_df['dti'], q = 15, retbins = True)
loan_df['ann_inc_bin'], cut_bin = pd.qcut(loan_df['annual_inc'], q = 20, retbins = True)
loan_df['int_rate_bin'], cut_bin = pd.qcut(loan_df['int_rate'], q = 20, retbins = True)
```

The feature issue_d converted into continuous feature like month and year.

```
loan_df['issue_month'] = loan_df.issue_d.apply(lambda x: int(datetime.strptime(x.split('-')[0],
```

Derived Metrics

Create Continuous Metric

The feature issue_d converted into continuous feature like month and year.

Code:

```
loan_df['issue_month'] = loan_df.issue_d.apply(lambda x: int(datetime.strptime(x.split('-')[0], '%b').month))
loan_df['issue_year'] = loan_df.issue_d.apply(lambda x: int(x.split('-')[1]))
```

Outliers

Invester Amount

The invested amount above 28000 is outlier.

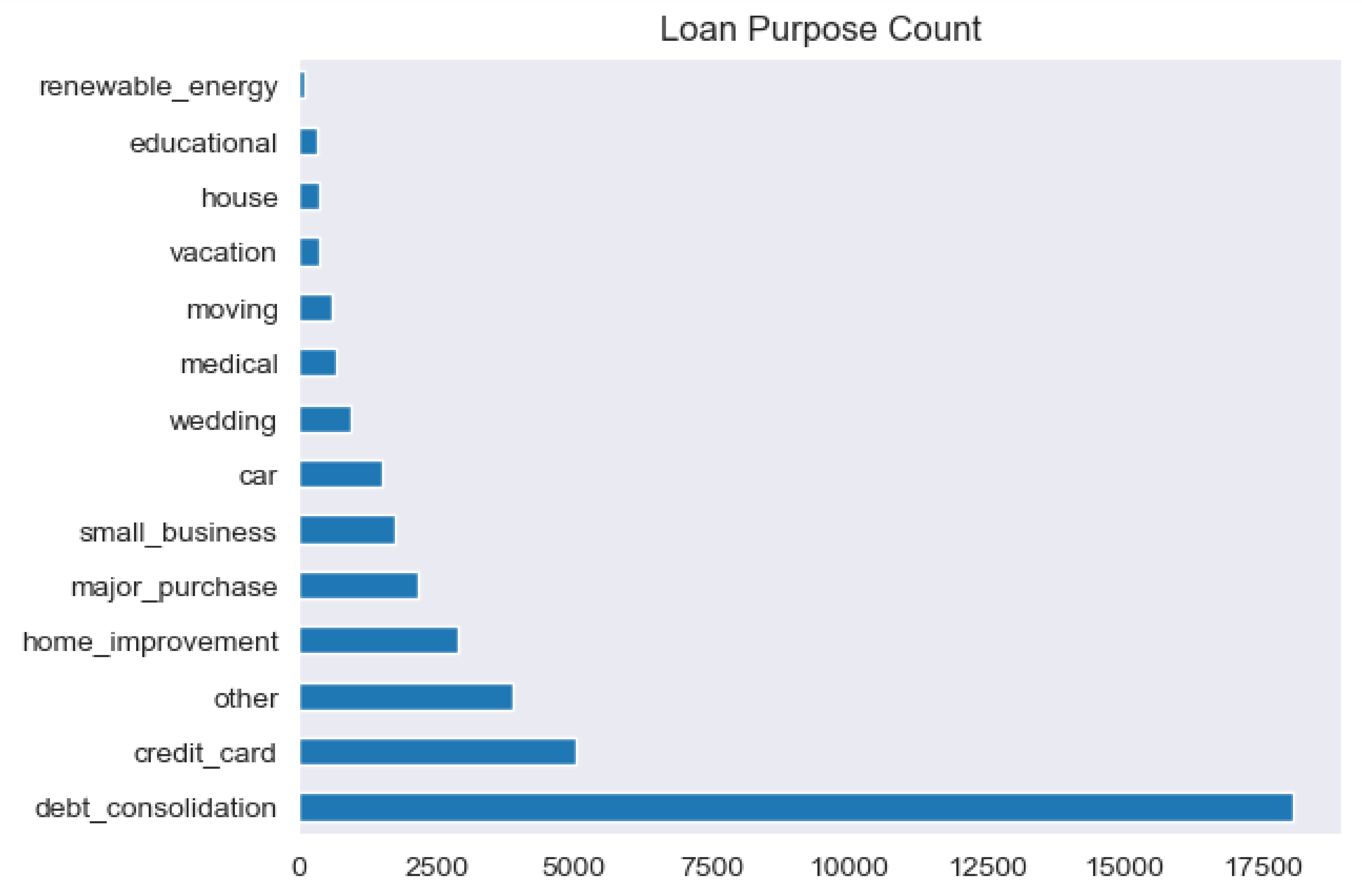
Data Filtering

Loan status:

Loan status has 3 values as 'Fully Paid', 'Current', 'Charged Off'.
Application with loan status as 'current' can turn into any category between 'Fully Paid' or 'Charged Off'.
So we are filtering out only those two loan status which has certain understanding. And not considering account having loan status as 'Current' in this analysis.

Data Analysis & Plotting

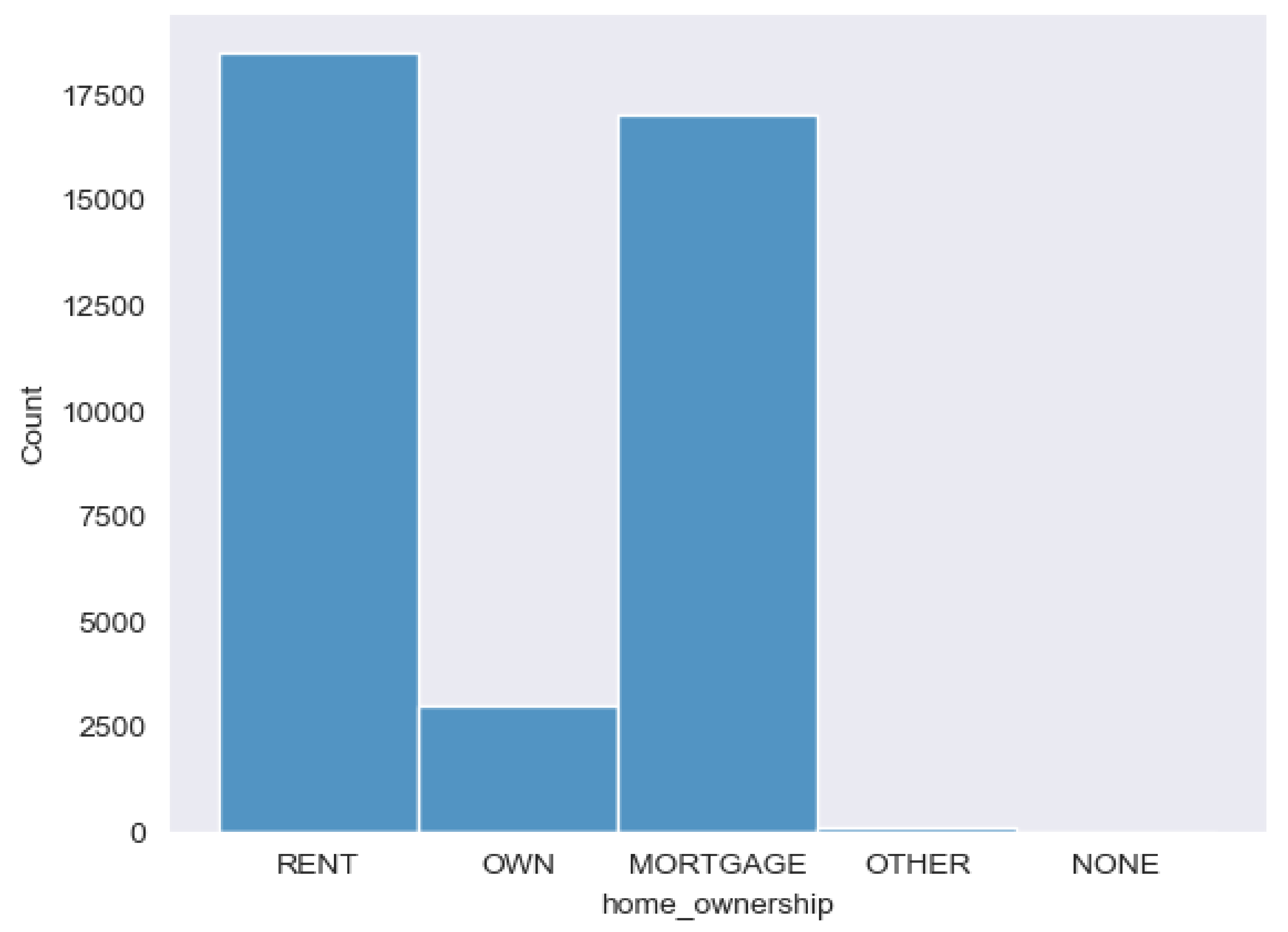
Loan distributed in different types



Observations:

Most of the loans are given for debt consolidation

Home ownership wise loans



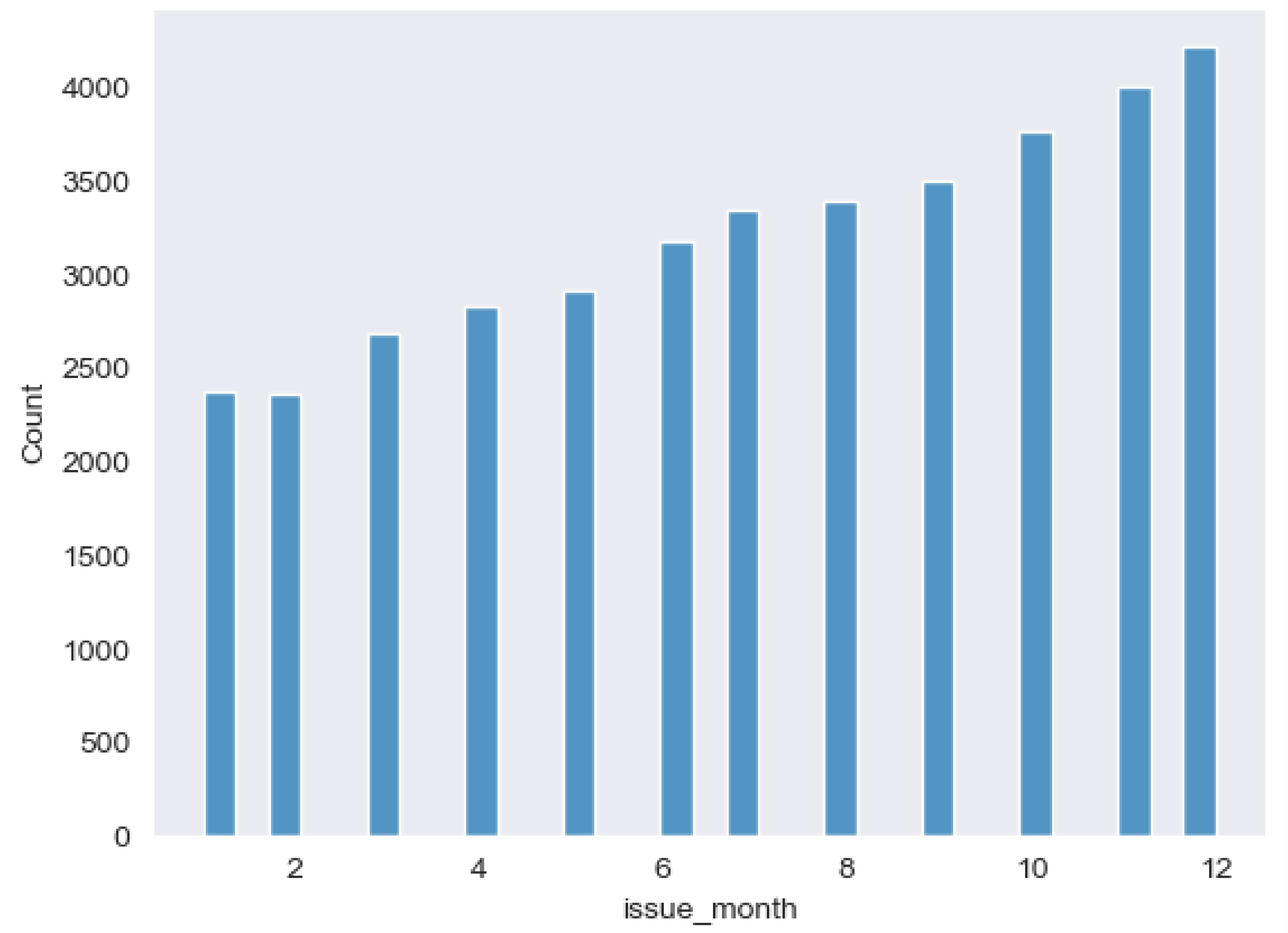
Observations:

People having home ownership are less likely to apply for loans.

Recommendations:

Promote these people to take loans by offering less interest rates. These will be secured loans.

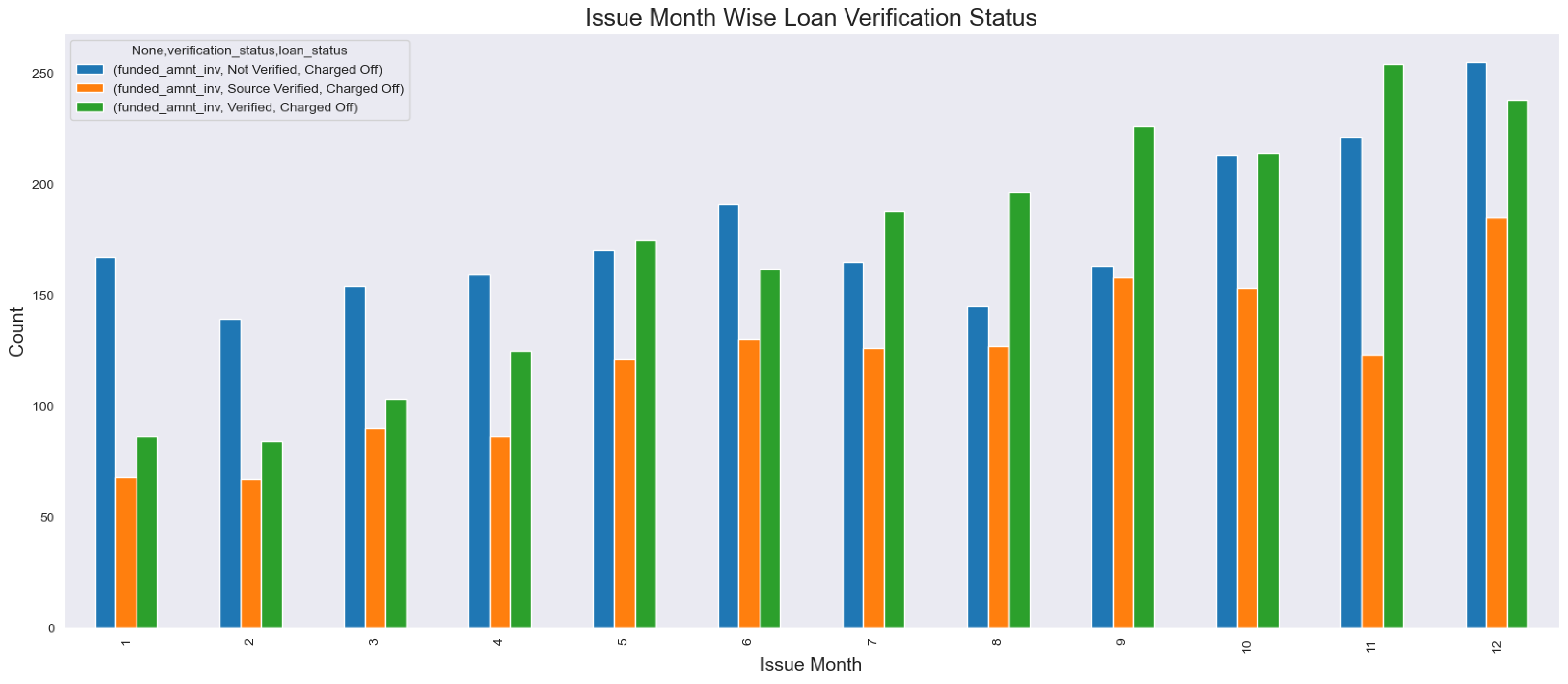
Monthwise Loans



Observations:

Maximum loans are given in the second half of the year.

Monthwise Verified Loans



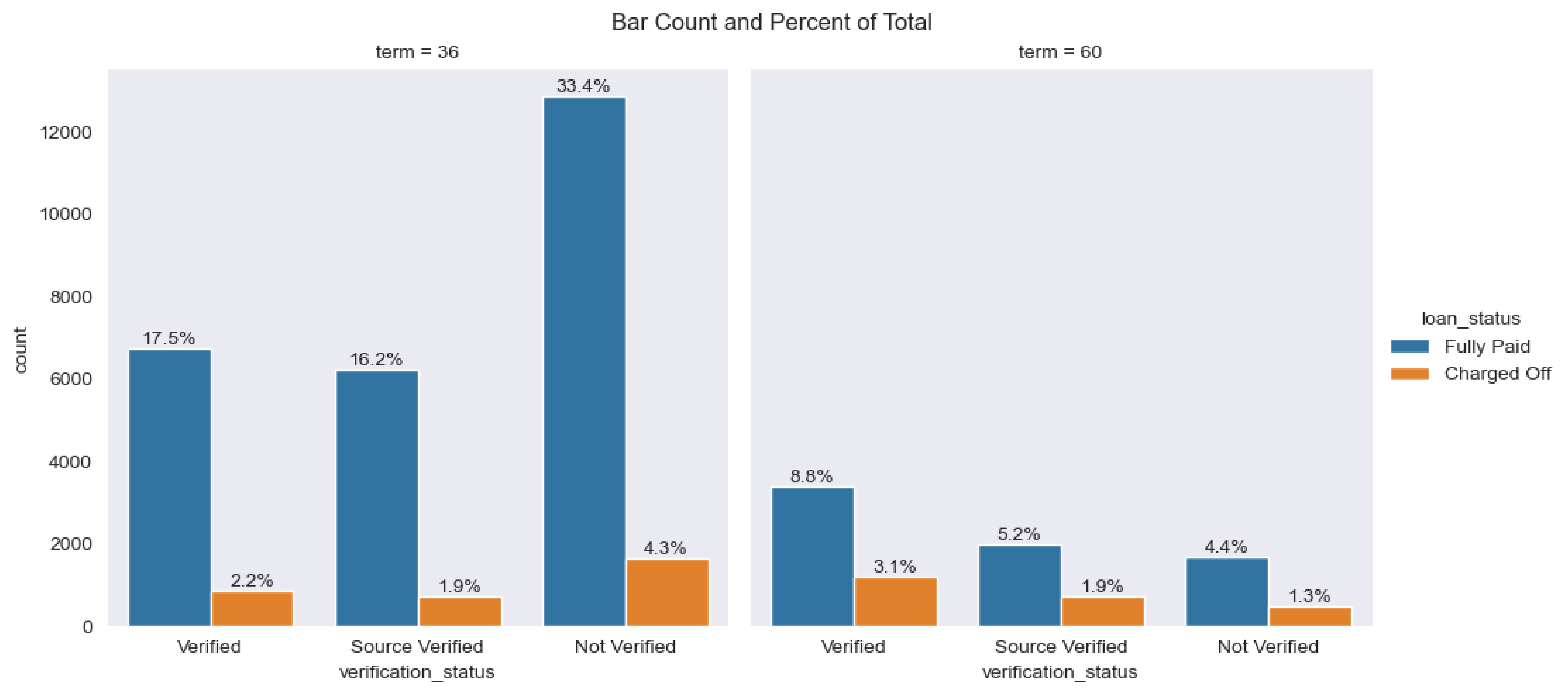
Observations:

In the month of Dec, Jan, Feb, Mar, June the loan applications are not verified most of the times, which results in the default.

Recommendations:

Verification of application should be mandatory in all the observed months.

Termwise Verified Loans



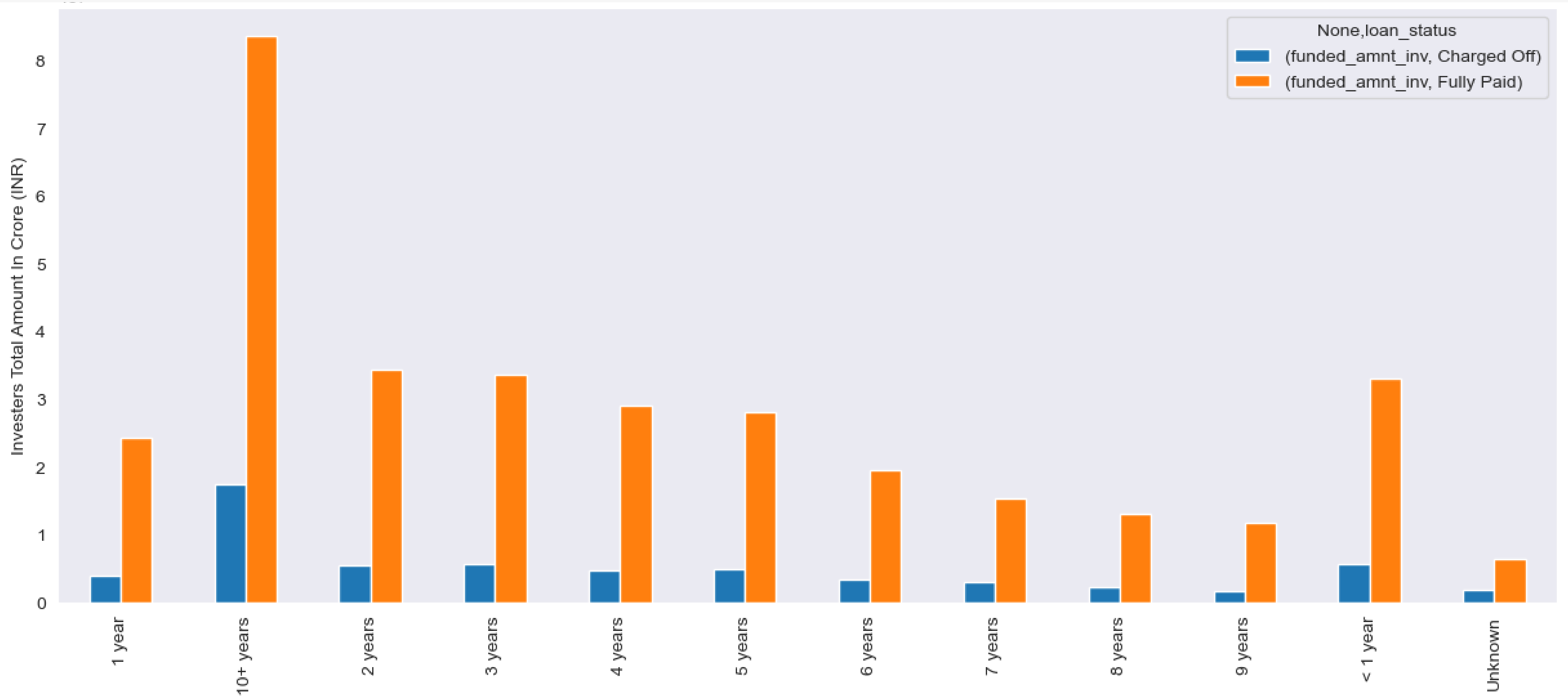
Observations:

Short term (36 months) loans are more profit generating.

Recommendations:

Verification of application should be mandatory for long term (60 months) loans.

Experience wise Invested Amt

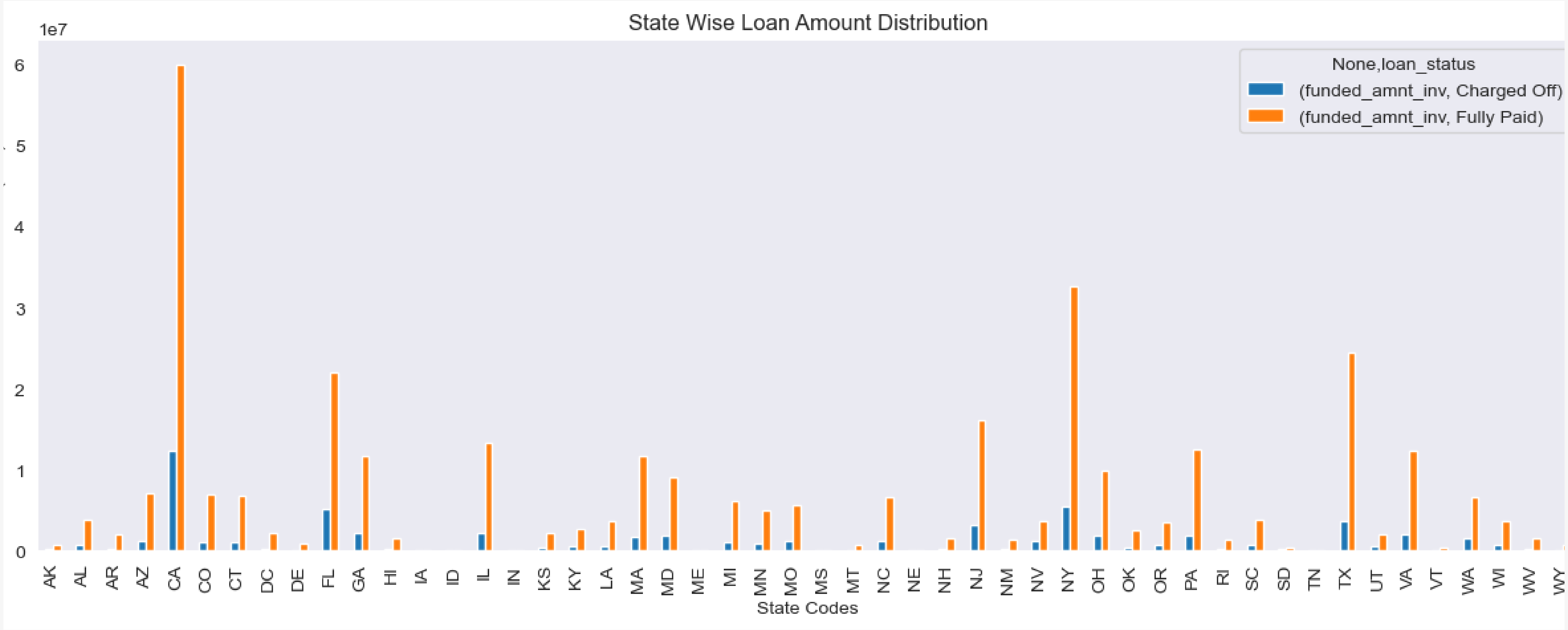


Observations:

Maximum loan amount is given to borrowers having 10+ years experience.

Maximum loan amount is successfully re-paid by 10+ year experience borrowers.

State wise Investment Loans



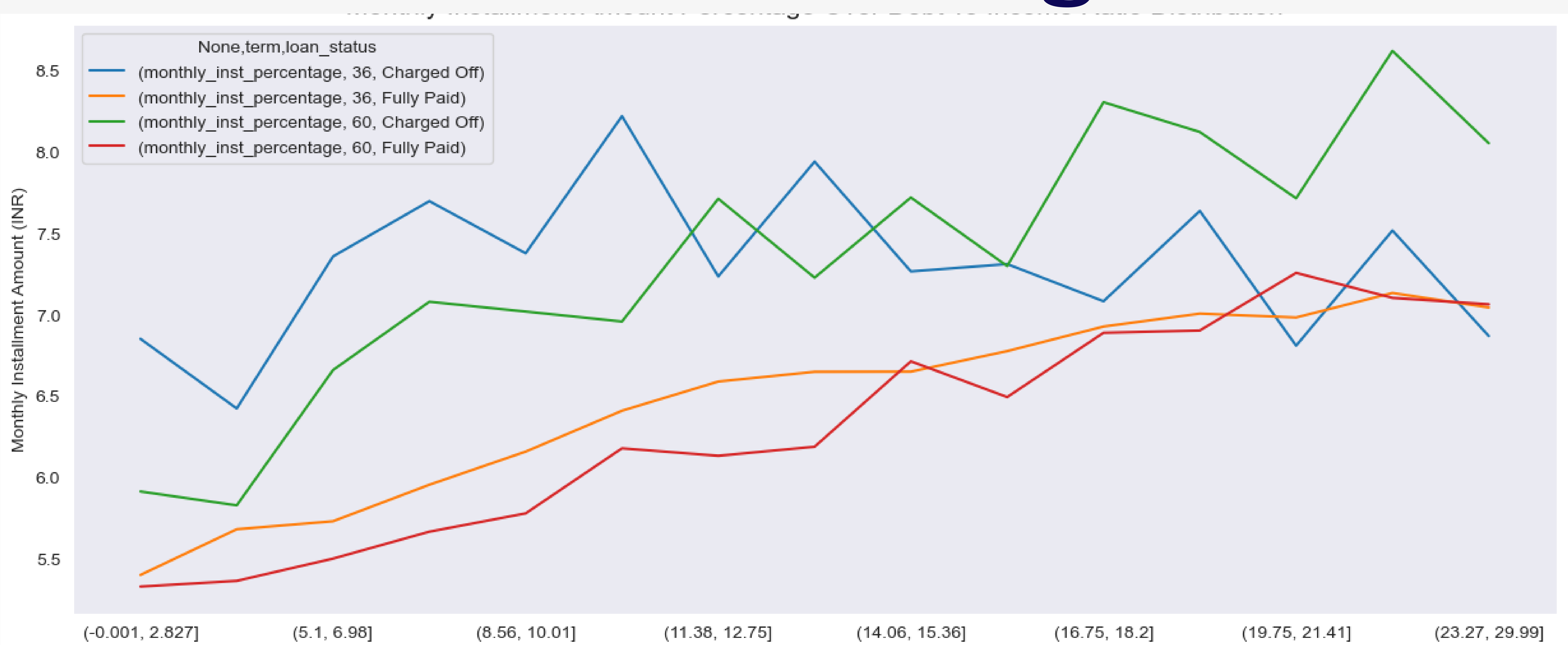
Observations:

The most successful loan distribution is in CA state, followed by NY, TX, FL.

Recommendations:

We can prioritize the loan approval as per successful repayment history of states.

Installment Percentage Vs Dti



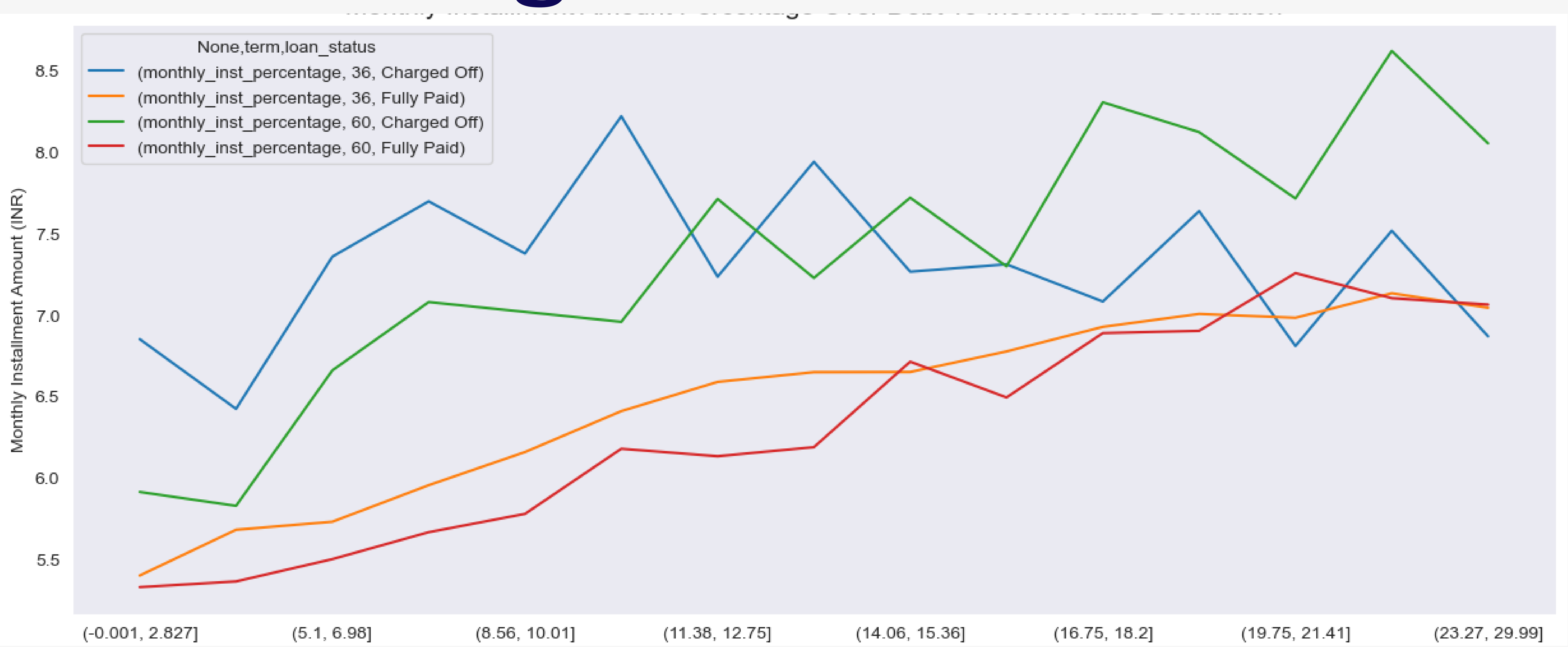
Observations:

For short term (36 months) loans credit amount is high for low dti ratio.
for long term (60 months) loans credit amount is high for higher dti ratio

Recommendations:

Decrease the invested amount with respect to dti ratio for both term loan.

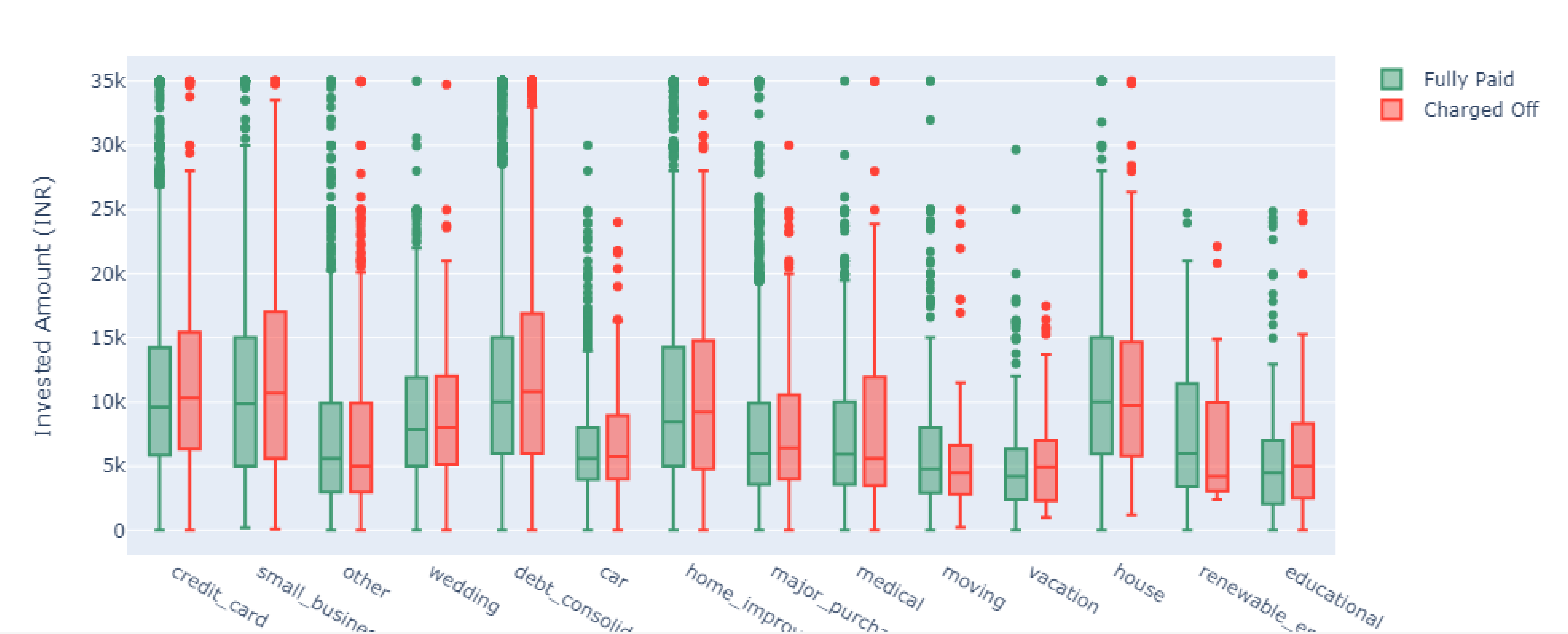
Loan Subgrade And Investment



Observations:

A & B & c grade loans are more successfully repaid & generated profit.

Loan purpose & Investment Amount



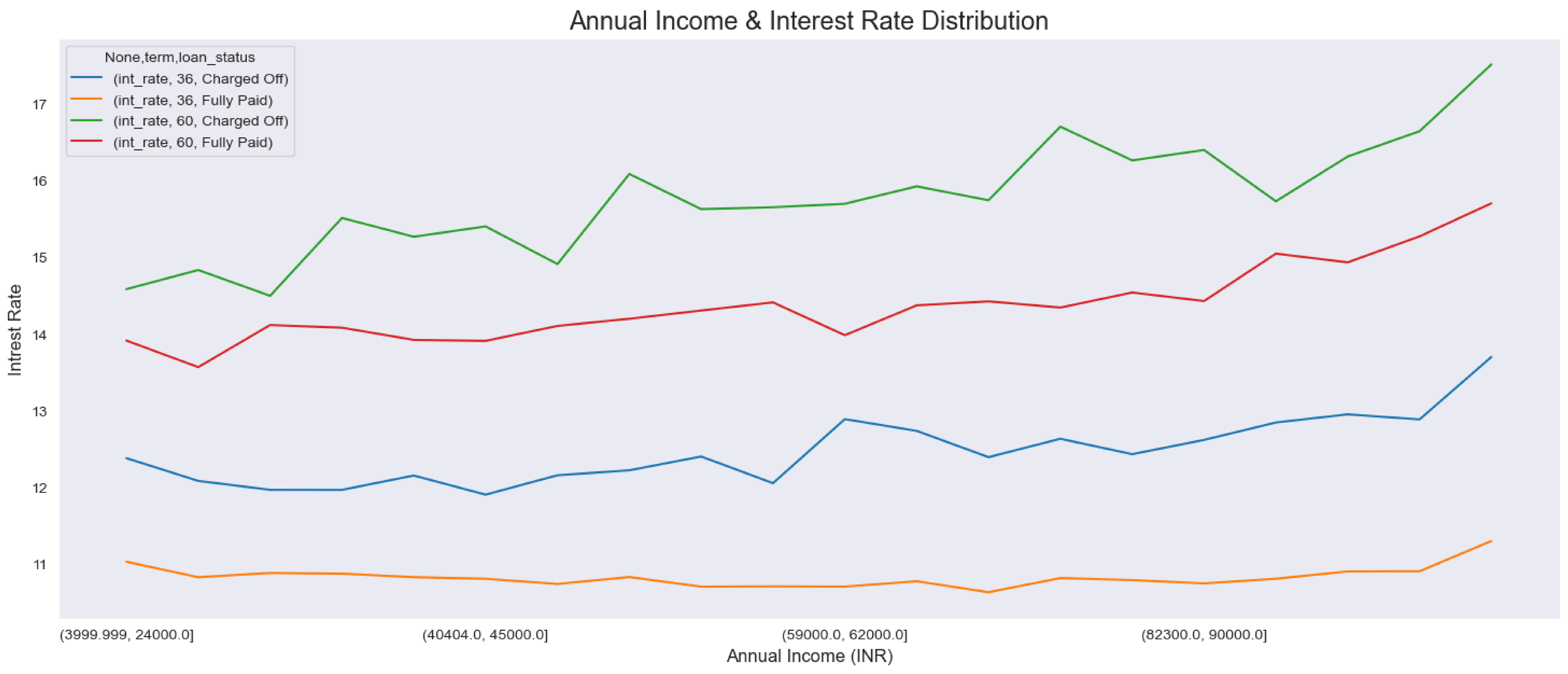
Observations:

debt_consolidation & small_business with high credit amount are more like to default.

Recommendations:

Minimize high credit amount to debt_consolidation & small_business.

Annual Income & Interest Rate



Observations:

high interest rate loan accounts are more likely to default.

Recommendations:

Minimize high credit amount for high interest rates.