Lending Club Case Study

PRASANTA BHATTACHARJEE

BATCH MLC70

Background

The company faces significant financial losses due to loan defaults. By analyzing the characteristics of borrowers who default, the company aims to develop a predictive model to identify high-risk loan applicants and mitigate potential credit losses.

Objective

The primary objective of this case study is to identify the key driver variables that contribute to loan default, enabling the company to reduce credit loss and improve its risk assessment process.

Approach of Analysis

The dataset includes information on borrower characteristics, loan attributes, and loan outcomes (e.g., default or repayment).

The specific variables are Home ownership, employment length, purpose, etc.

Loan attributes: loan amount, interest rate, term, etc.

Loan outcomes: default (charged-off) or repayment (fully paid or current)

By applying EDA and statistical modeling techniques to this dataset, the company can gain valuable insights into the factors driving loan default and develop a more effective risk assessment process.

Data Understanding

Overall Default Rate is 14%

i.e., Ratio of "Charged Off" cases with respect to overall cases.

The "Current" ones, i.e., active loan application cases can be excluded from the analysis as these are neither closed or defaulter cases, and also significantly less in comparison to overall records in the dataset.
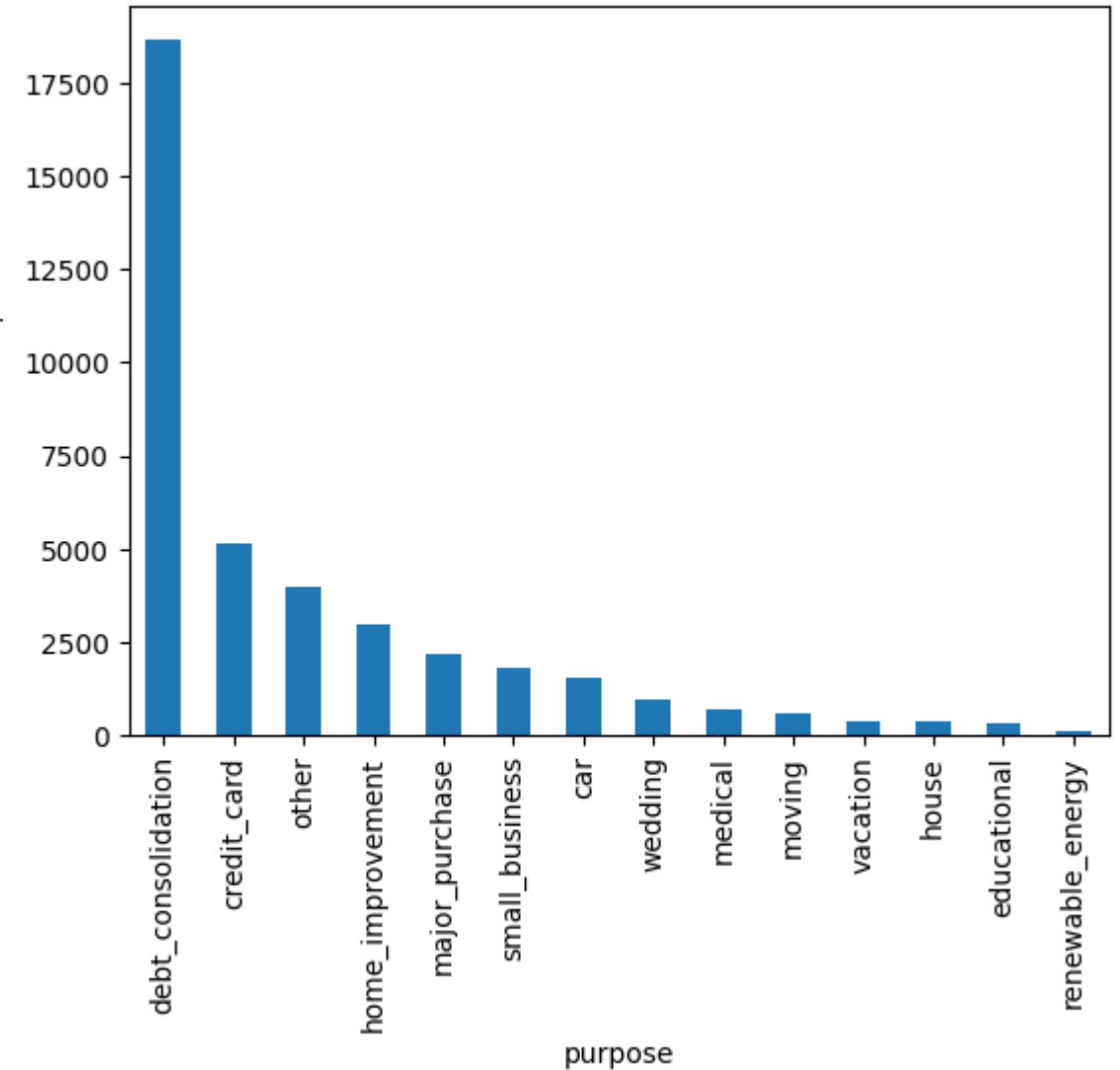
	count
loan_status	
Fully Paid	32950
Charged Off	5627
Current	1140
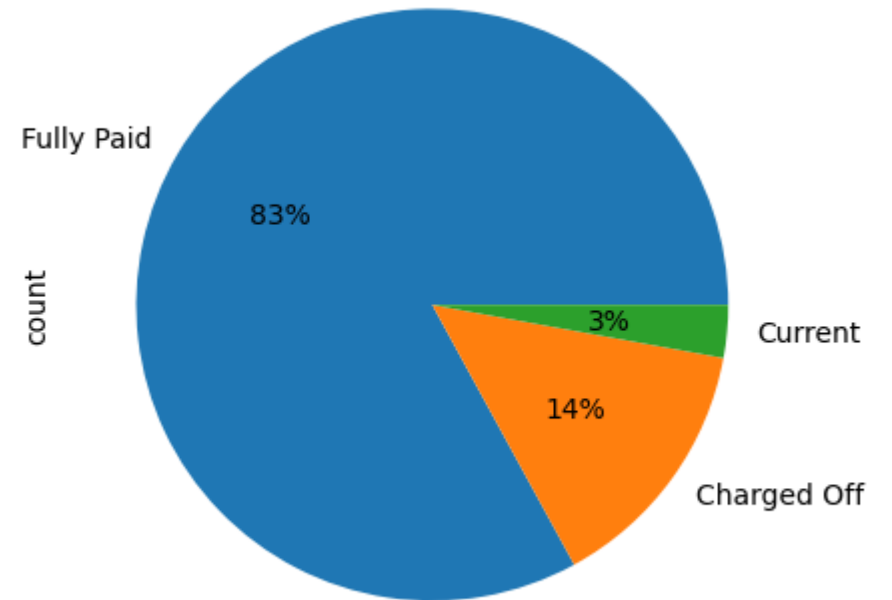
Loan Purpose Distribution

Debit consolidation most popular purpose.

It will be necessary to investigate further on default rates of each of the purposes.

Default Rates by Loan Purpose

| purpose	Default_Rate_Percent
small_business	27.08%
renewable_energy	18.63%
educational	17.23%
other	16.38%
house	16.08%
moving	15.97%
medical	15.57%
debt_consolidation	15.33%
vacation	14.13%
home_improvement	12.07%
credit_card	10.78%
car	10.67%
wedding	10.37%
major_purchase	10.33%

Applications with Small Business Purpose have the highest Default rate.
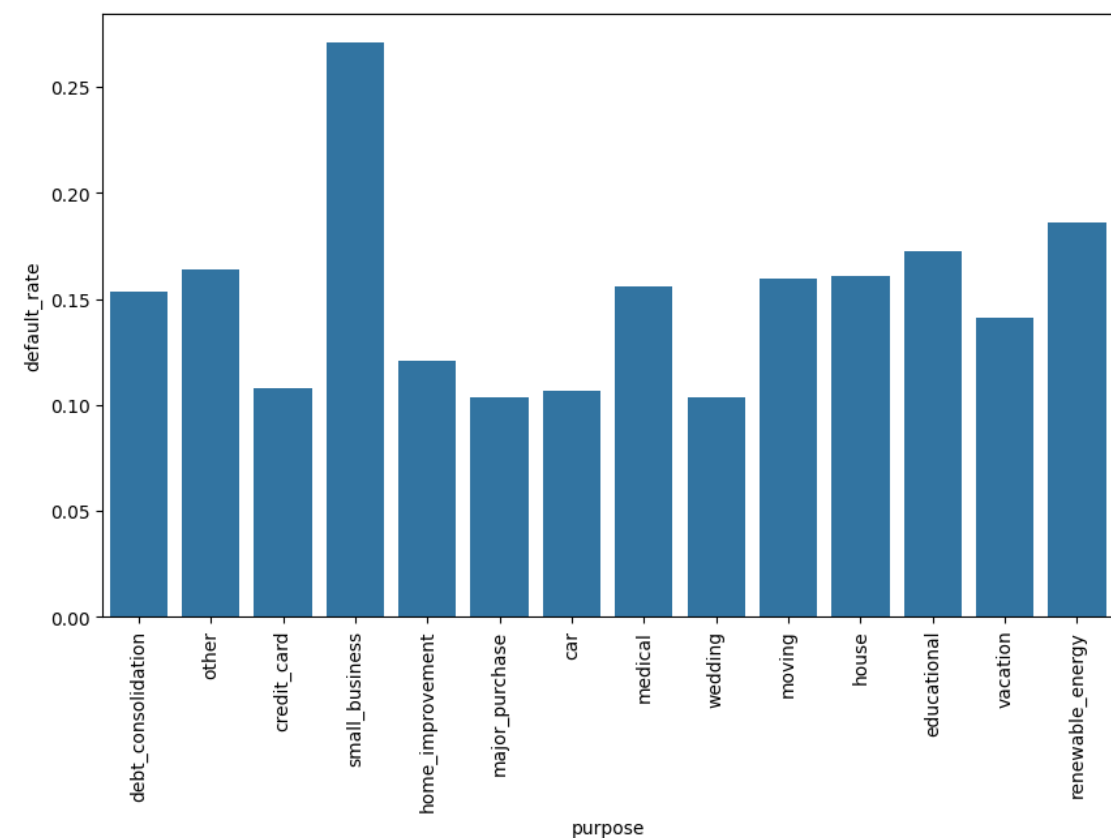
Home Ownership – not a determinant feature

It can also be observed that Home Ownership does not give any strong indication.

|	home_ownership	Default_Rate_Percent
3	OTHER	18.37
0	RENT	15.36
2	OWN	14.89
1	MORTGAGE	13.67
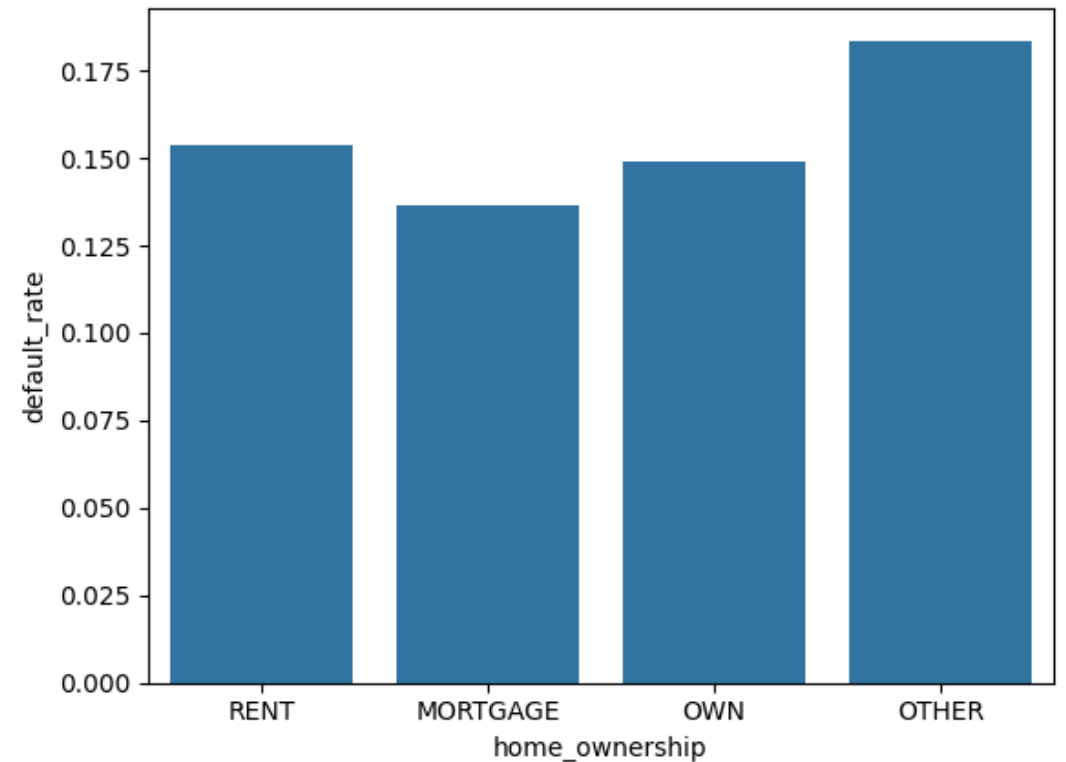
Grade & Sub Grade Distribution

It can be observed that the grades are assigned by LC on the basis of the default rates. Default rate is lowest for Grade A whereas highest for Grade G . Similar pattern can be seen for sub grades however with some exceptions as can be visualised in the chart below.
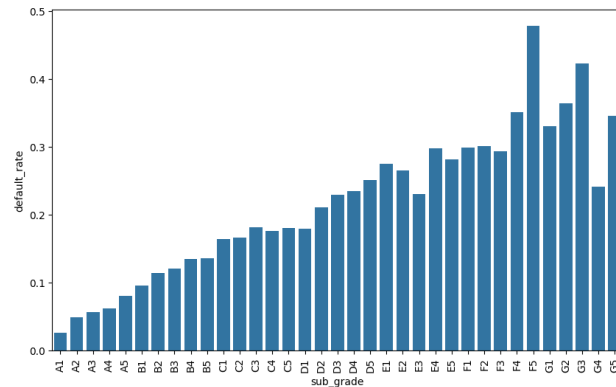
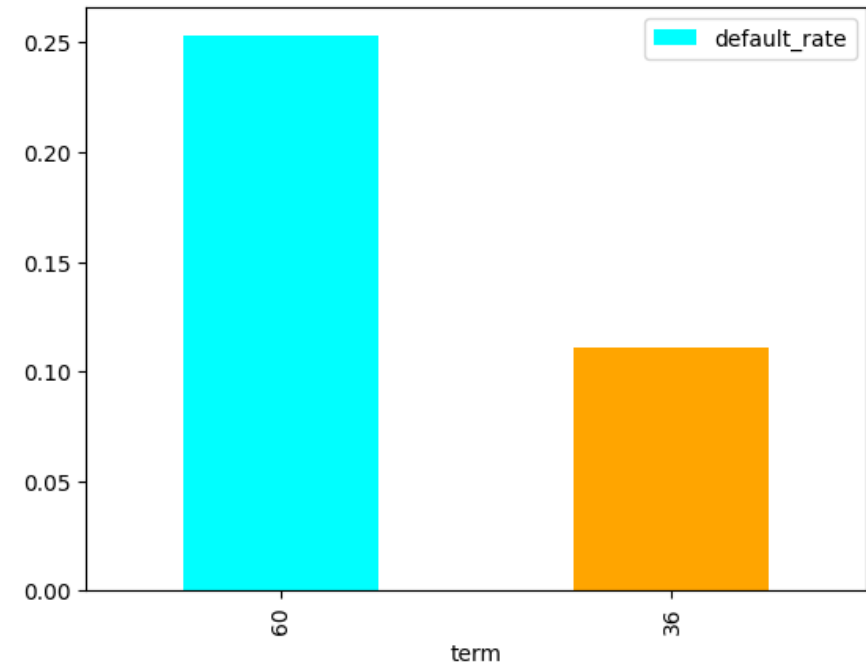| grade	default_rate_percent
A	5.99%
B	12.21%
C	17.19%
D	21.99%
E	26.85%
F	32.68%
G	33.78%
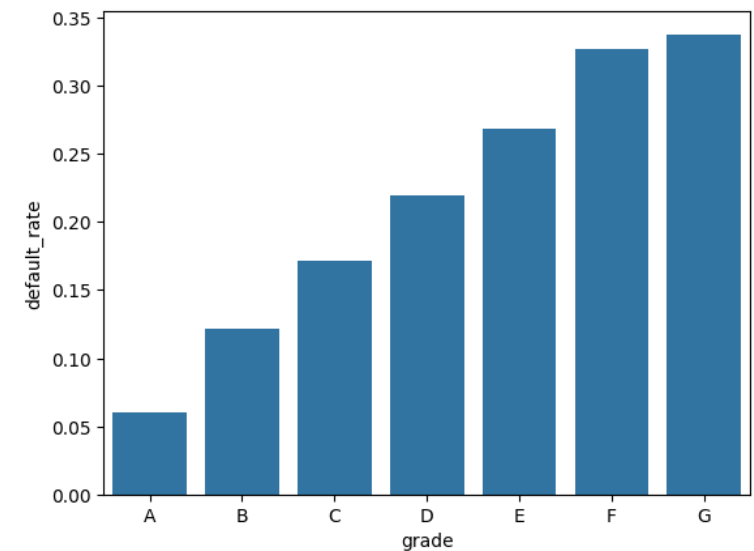
Term Feature Observations

The application terms have only 2 categories:

❖ 36 months

❖ 60 months

It can be observed that 60 months applications have higher default rate.
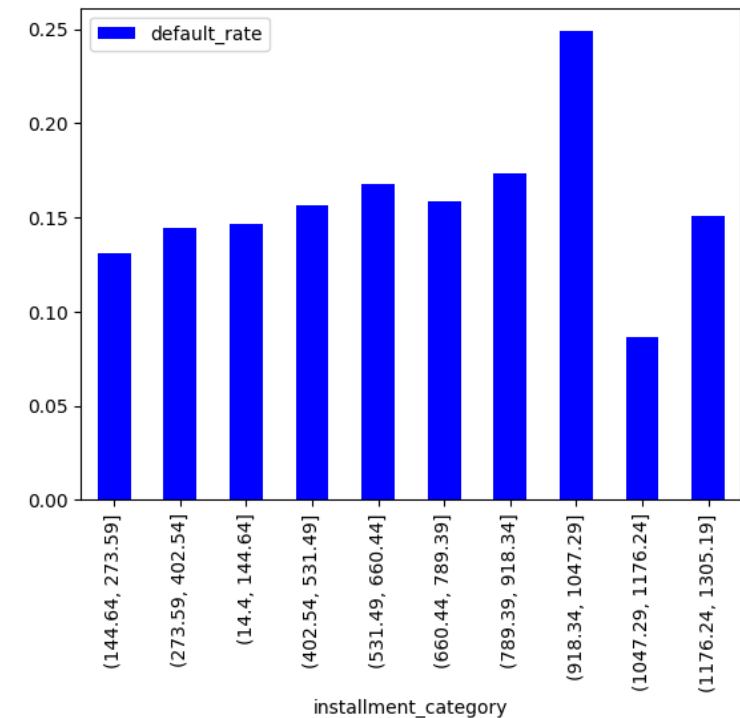
term	default_rate_percent
36	11.09%
60	25.31%

Instalment

It can be observed that there is slight increase in default rate for instalment within the range of approximately 900 to 1000. Upon drilling down further into this category it was found that interest rates and loan amounts were comparatively much higher for defaulter cases.

loan_amnt	int_rate	grade	purpose
35000	23.52	G	debt_consolidation
35000	23.13	G	home_improvement
35000	22.74	G	debt_consolidation
35000	22.74	G	debt_consolidation
35000	22.74	G	debt_consolidation
35000	22.48	G	small_business
35000	22.35	F	debt_consolidation
35000	22.35	F	debt_consolidation
35000	22.35	F	credit_card
35000	22.35	F	debt_consolidation
35000	22.06	F	small_business
35000	22.06	F	debt_consolidation
35000	22.06	F	debt_consolidation
35000	21.74	F	other
35000	21.74	F	small_business
35000	21.67	F	debt_consolidation
35000	21.36	F	debt_consolidation

installment_category	default_rate_percent
(14.4, 144.64]	14.62%
(144.64, 273.59]	13.13%
(273.59, 402.54]	14.46%
(402.54, 531.49]	15.67%
(531.49, 660.44]	16.76%
(660.44, 789.39]	15.88%
(789.39, 918.34]	17.36%
(918.34, 1047.29]	24.88%
(1047.29, 1176.24]	8.63%
(1176.24, 1305.19]	15.07%

Employment Length

Employment length is not a determinant feature.

| emp_length	default_rate_percent
0	15.65%
1	14.39%
2	13.21%
3	13.83%
4	13.82%
5	14.34%
6	14.16%
7	15.37%
8	14.15%
9	12.89%
10	15.68%
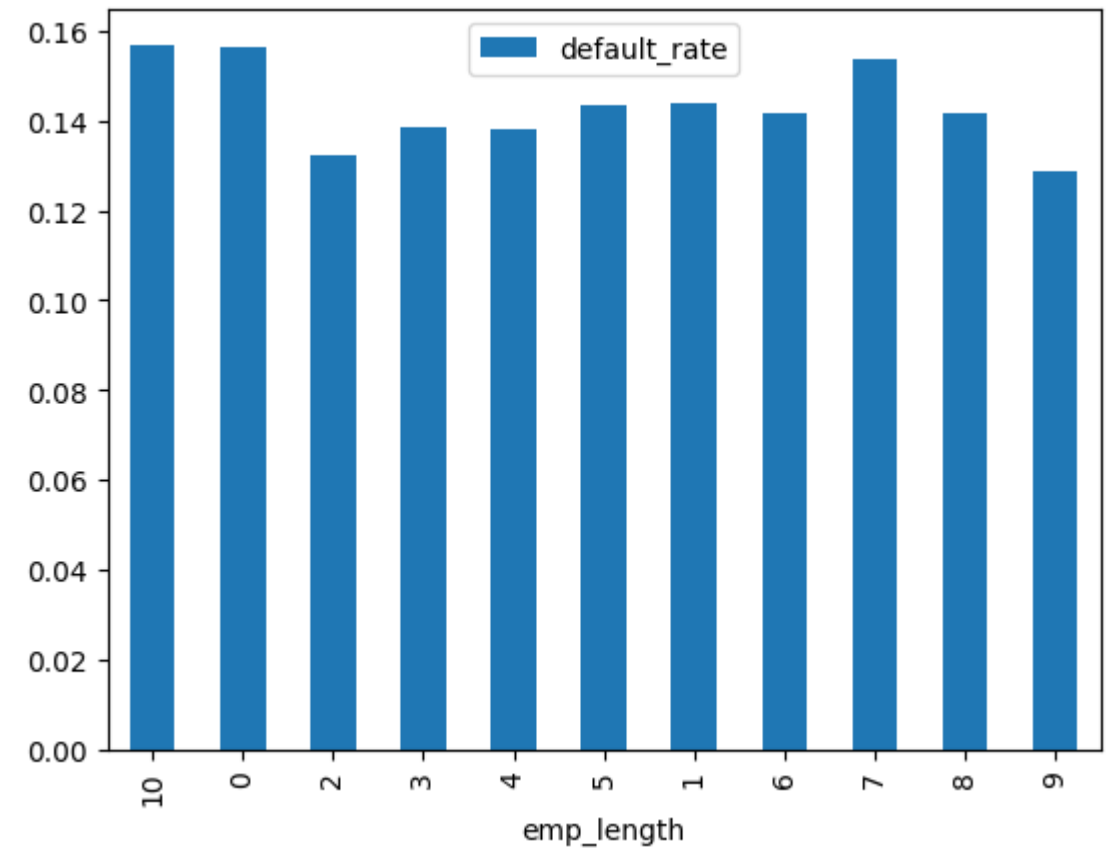
Verification Statuses

There is a possibility of some fraudulent cases because rate of defaulters is observed to be more for verified cases than non verified cases. These labels cannot be relied upon for risk analysis.

verification_status	default_rate_percent
Not Verified	12.83%
Source Verified	14.82%
Verified	16.8%
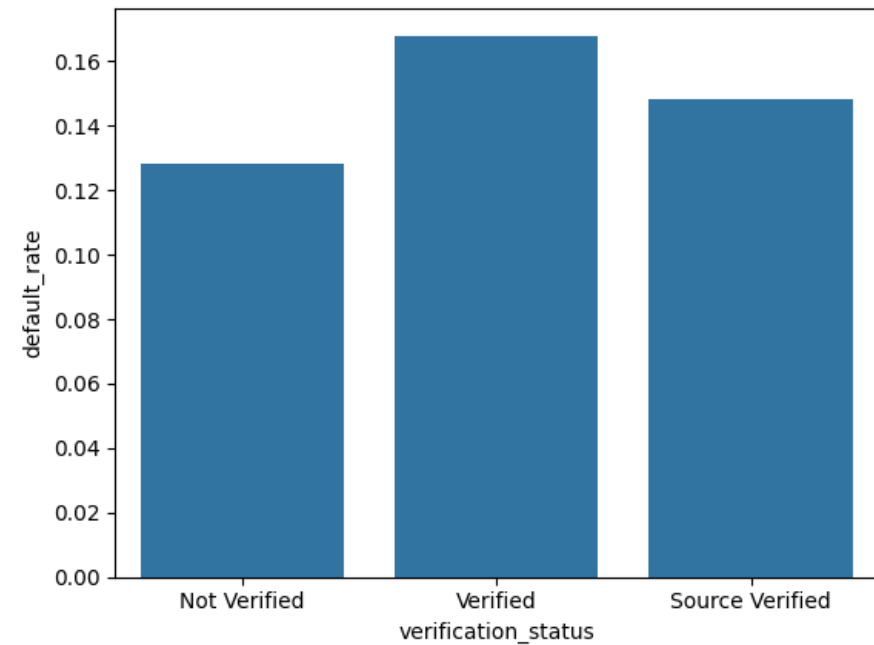
Loan amounts

There is no distinct pattern
observed with respect to the
loan amounts as well.
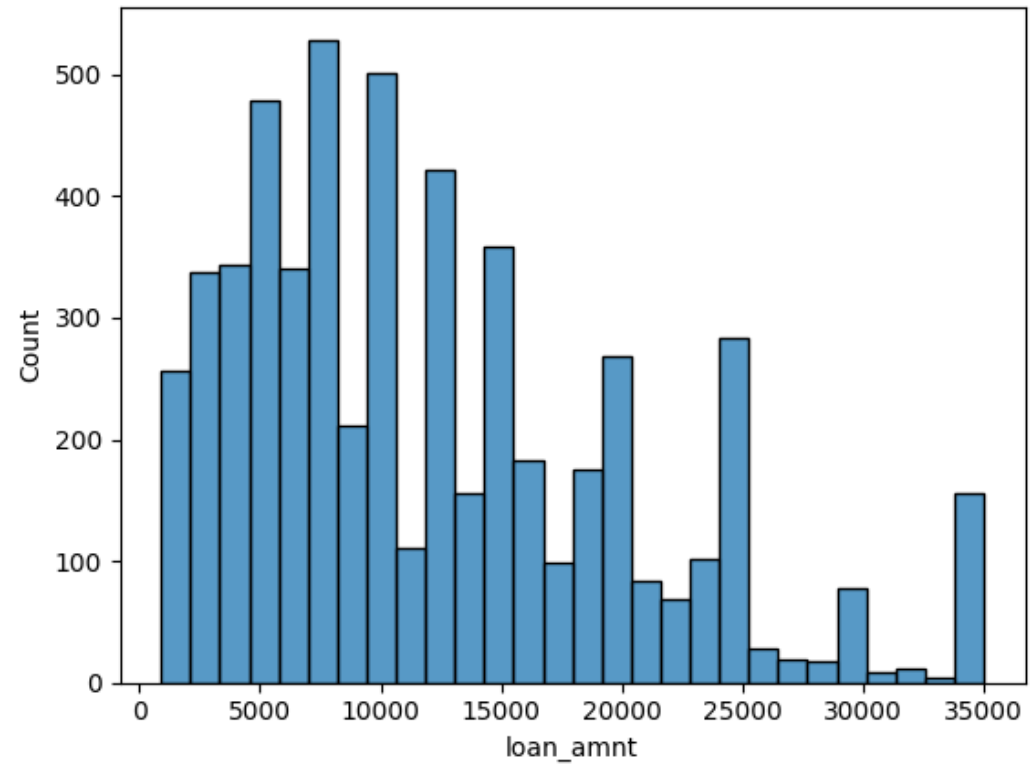
Recommendation

Based on the discussed findings and features identified such interest rate, loan amount, purpose of loan, etc., development of a predictive model will be necessary to estimate the likelihood of default based on borrower and loan characteristics.

The relevant numeric features should be used to generate the training and test datasets and then predictions can be obtained from the model.

Thank You