LENDING CLUB CASE STUDY

Team:
- HARSHAL PANDIT
- SURIN SURESH

Introduction

The dataset provided contains historical information about individuals who applied for loans and whether they experienced a 'default' event.

The objective is to discern recurring trends that can help predict the likelihood of loan default. This insight can guide decision-making processes, including loan approval, loan amount adjustments, and the application of higher interest rates for riskier candidates.

Steps Involved

1. Comprehending the Loan Dataset
2. Data Cleansing
3. Data Correction
4. Univariate Analysis
5. Segmented Univariate Analysis
6. Bivariate Analysis
7. Variable Correlation Assessment
8. Conclusions

Understanding the Loan Dataset

The data given below contains the information about past loan applicants and whether they ‘defaulted’ or not.

Loan Status - Key Leading Attribute (*loan_status*). The column has three distinct values (Fully paid, current and charged off)

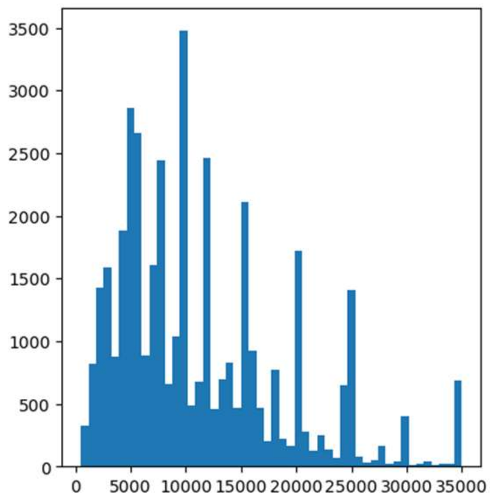
Following are the major Loan Attributes for analysis- Loan Amount (loan_amt) ,Grade (grade), Term (term), Loan Date (issue_date), Purpose of Loan (purpose), Verification Status (verification_status), Interest Rate (int_rate), Installment (installment)

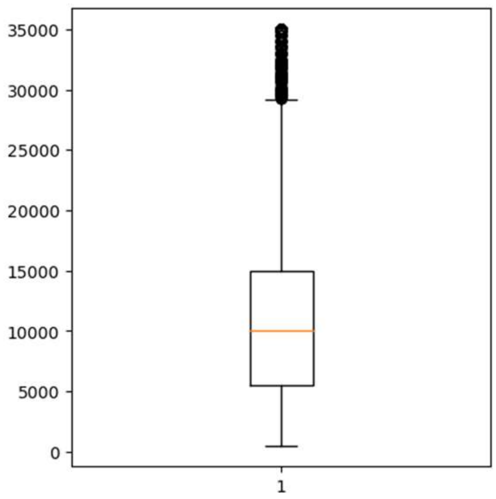
Data Cleansing and Correction

Following columns will be dropped
- Columns with missing values more than 75%
- Columns which have only one unique value
- Columns which are behaviour factors
- Columns which are generated after loan
- Columns which are irrelevant for loan status analysis

Missing and outliers values has to be handled

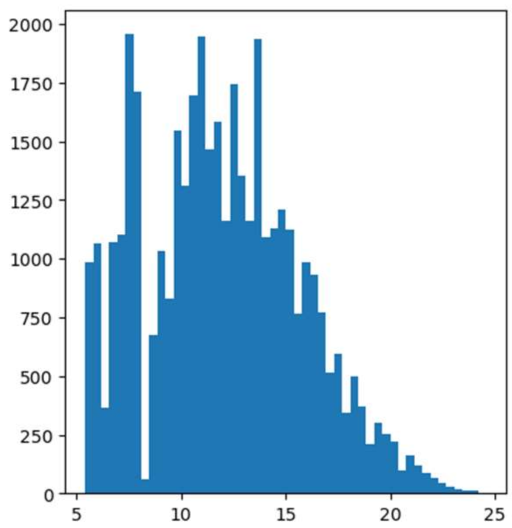
Univariant Analysis (loan amount)

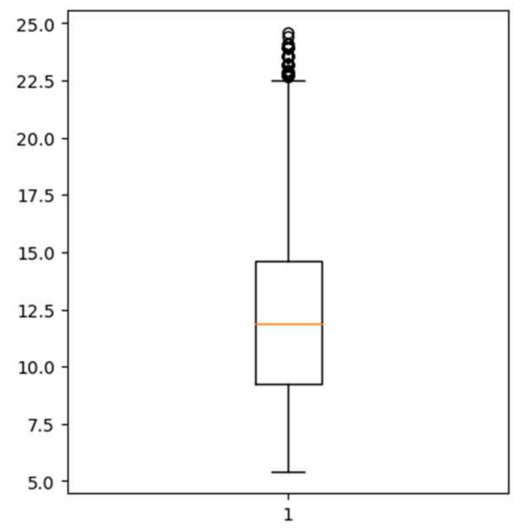




The analysis of loan amounts reveals a notable trend: a significant proportion of loans fall within the range of 5000 to 15000. The peak loan amount is at 10000, and an evident pattern emerges with a pronounced spike in loan amounts that are multiples of 5000. This observation suggests a borrower preference for seeking loans in increments of 5000.

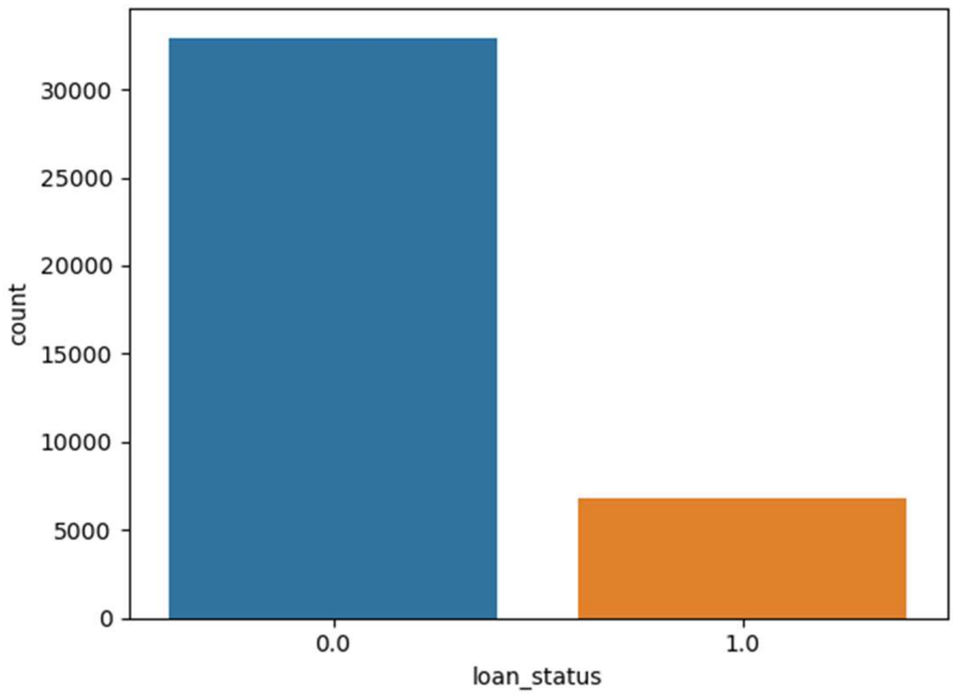
Univariant Analysis (interest Rate)





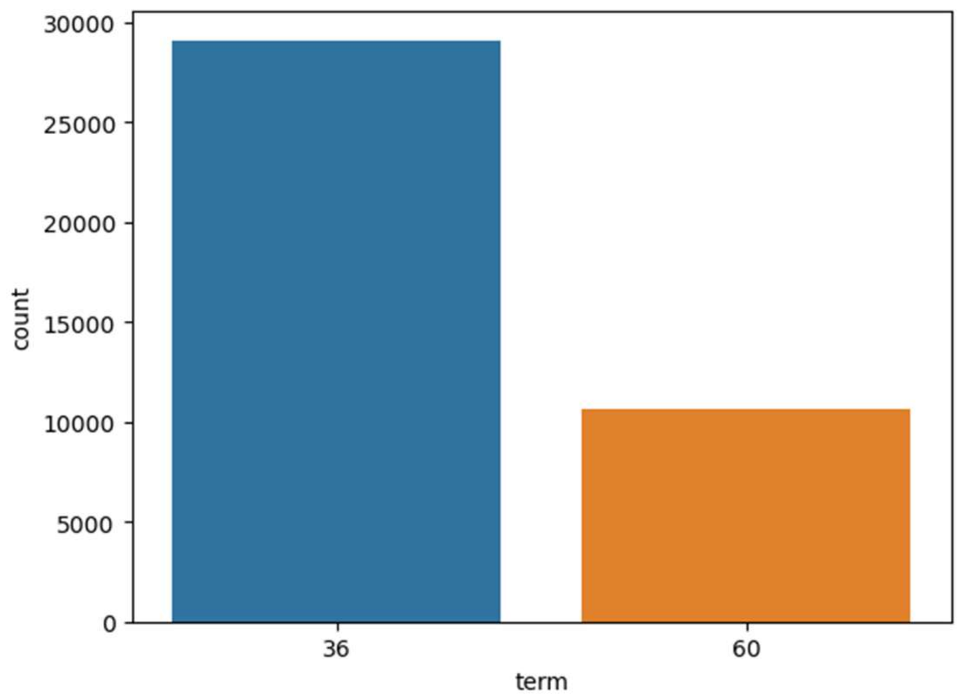
The examination of interest rates reveals several insightful patterns:
- The average interest rate across loans is approximately 12%. This serves as a benchmark for assessing individual interest rates.
- Notably, there are distinct spikes in interest rates at around 8.5% and 12%. These figures emerge as the most prevalent interest rates based on the type of loan.

Univariant Analysis (loan_status)



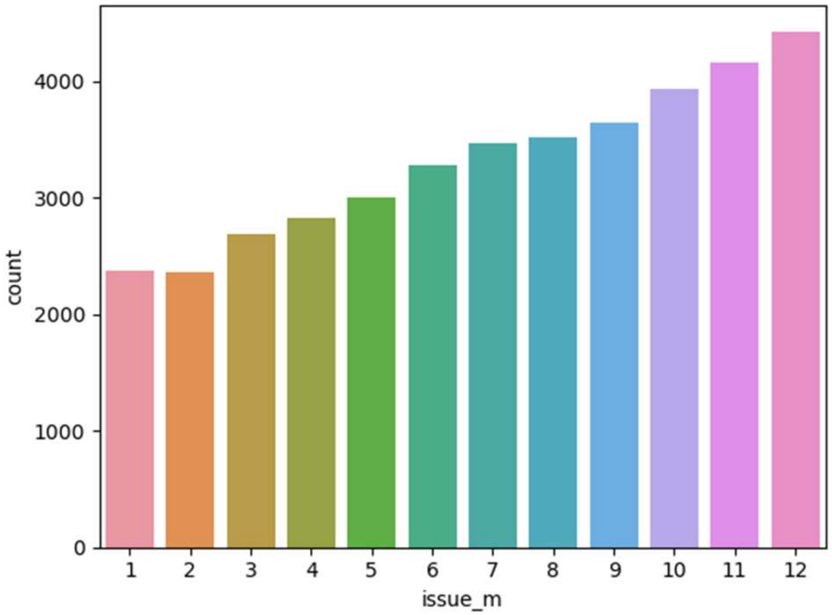
Evidently, the percentage of individuals who have defaulted on their loans is quite low.

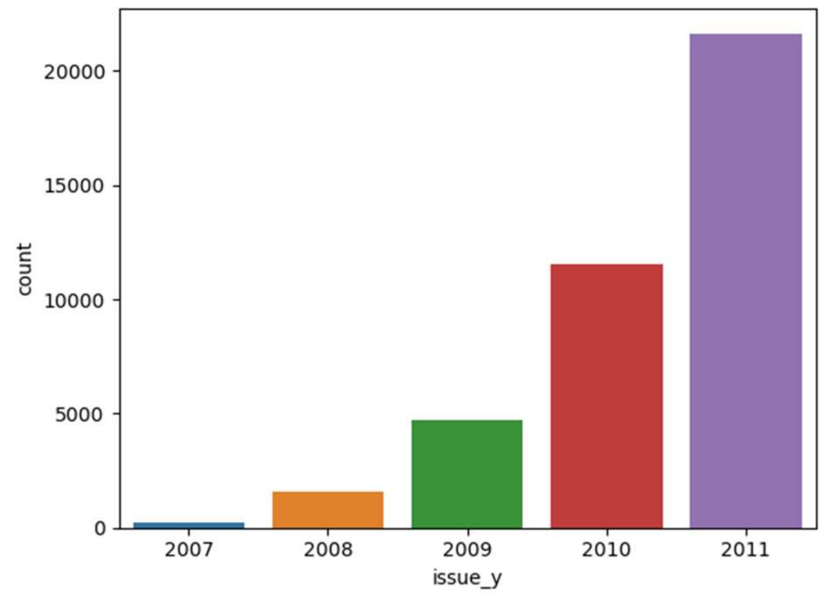
Univariant Analysis (term)



There is a noticeable trend wherein a higher number of borrowers opt for a loan duration of 36 months compared to the 60-month term.

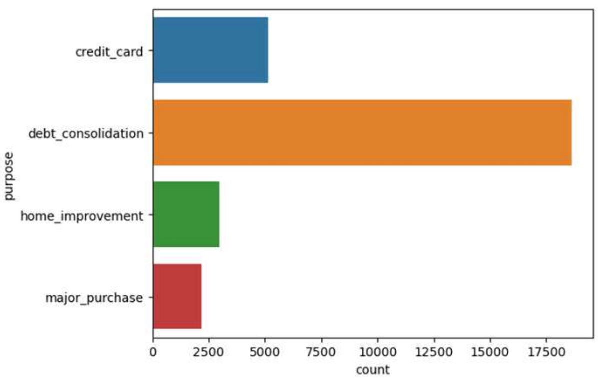
Derived Variable Analysis (Issue Month and Year)

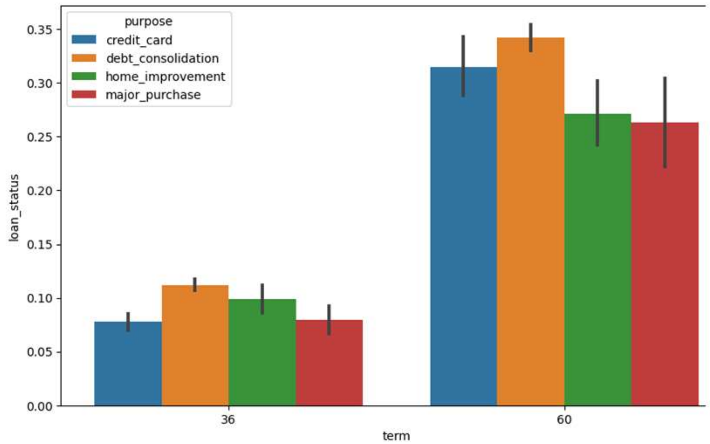


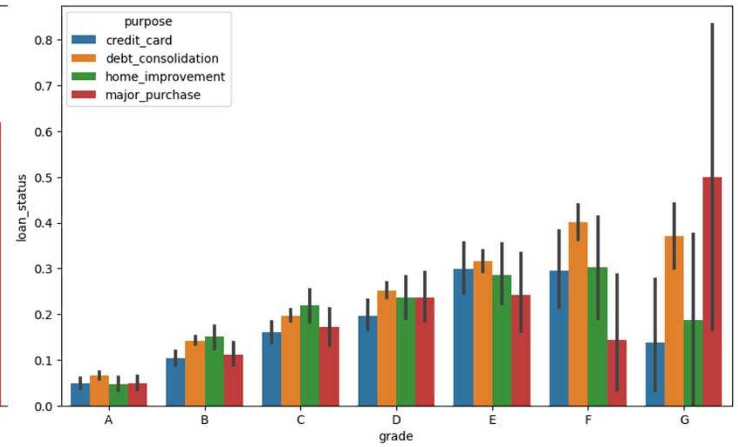


A discernible pattern emerges as the approved loan amounts steadily increase from January to December, also exhibiting an exponential growth trajectory over the years.

Segmented Univariant Analysis

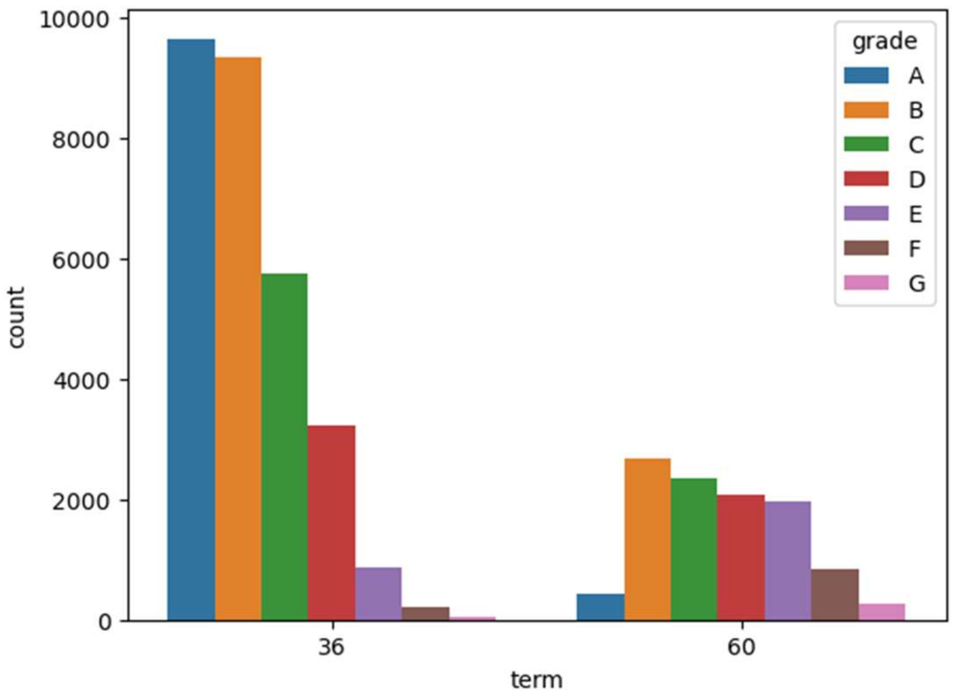






When Loan purpose was analysed with other variables, it was observed that debt consolidation loans tend to exhibit the highest default rates

Bivariant Analysis

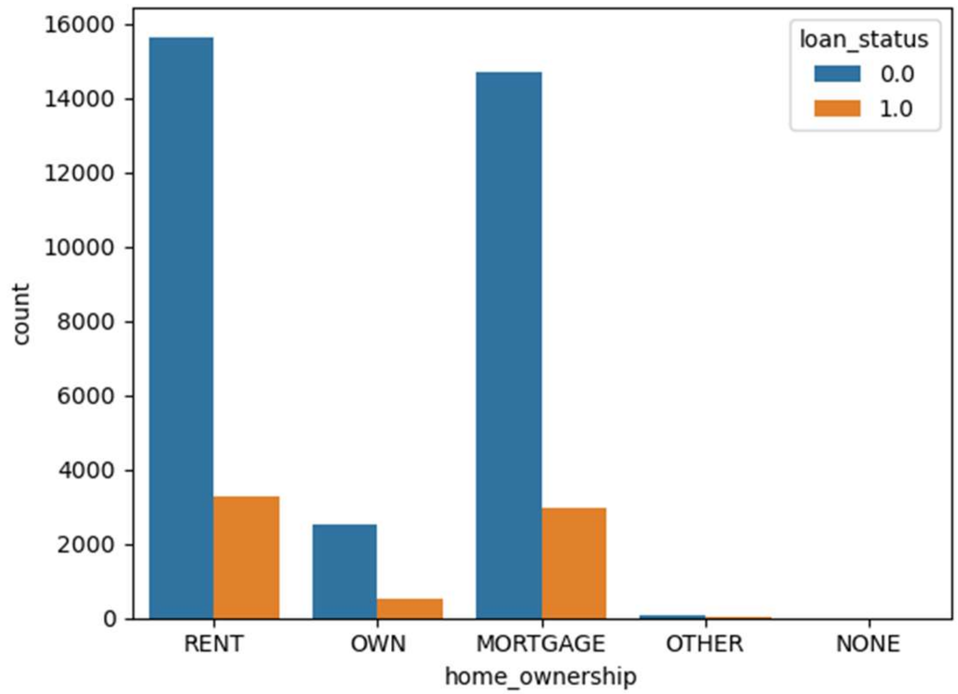


Loan Duration and Credit Grade Analysis

A distinct pattern emerges when considering the relationship between loan duration and credit grades:

- Loans with a duration of 36 months are predominantly favoured by individuals with credit grades A and B.
- In contrast, the 60-month loan term is more commonly selected by borrowers with credit grades B, C, and D.

Bivariant Analysis



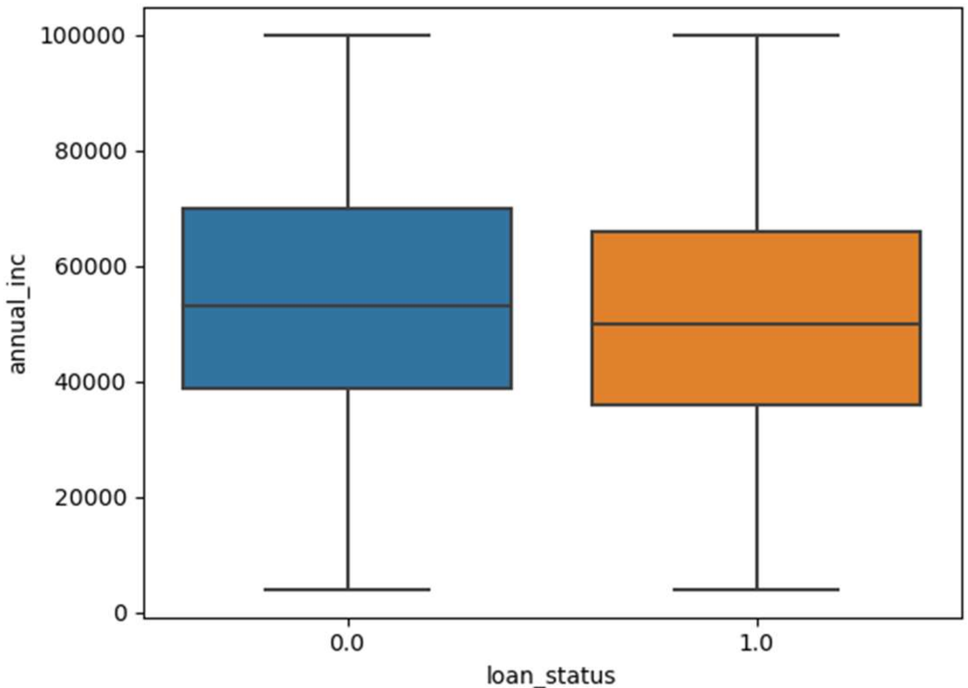
Default Rate and Housing Status

An intriguing observation emerges regarding the default rate based on housing status:

- Individuals who are renting homes or have a mortgage exhibit a relatively higher default rate.
- Conversely, those who own their homes tend to have a lower default rate.

- Loan_status = 0.0 = Fully Paid
- Loan_status = 1.0 = Default

Bivariant Analysis

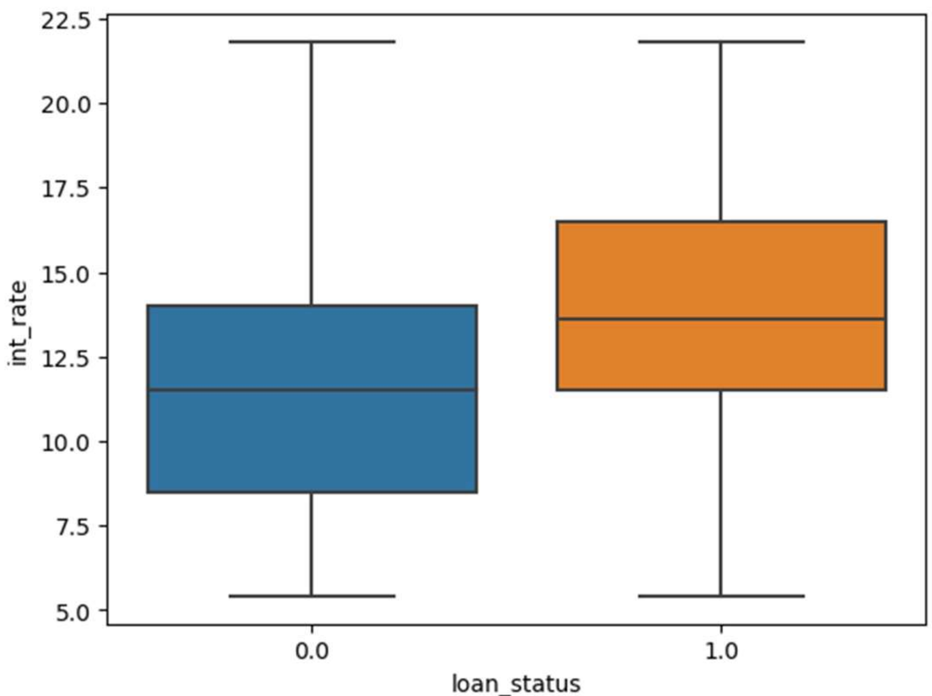


Income and Default Probability

A clear association is evident: individuals with lower income levels have a higher likelihood of defaulting on their loans

- Loan_status = 0.0 = Fully Paid
- Loan_status = 1.0 = Default

Bivariant Analysis

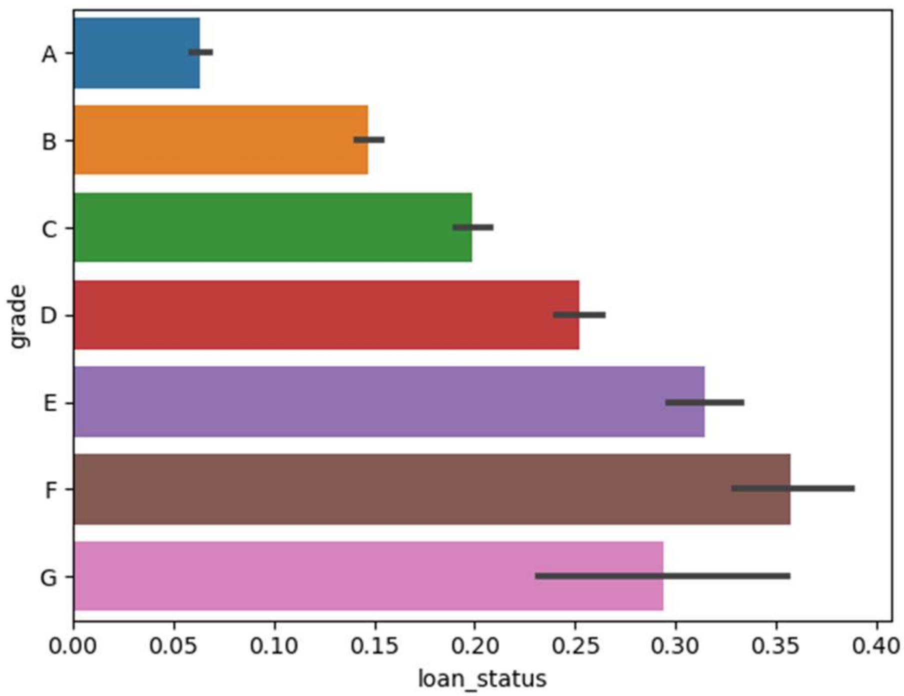


Interest Rate and Default Relationship

A noticeable trend emerges: borrowers who are subject to higher interest rates are more prone to default on their loans.

- Loan_status = 0.0 = Fully Paid
- Loan_status = 1.0 = Default

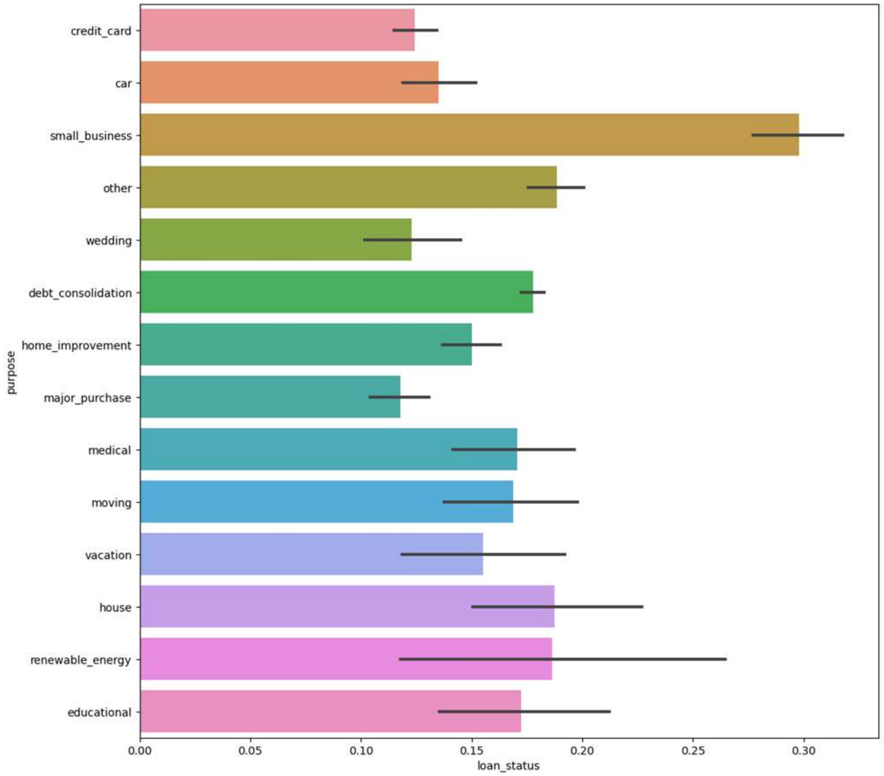
Bivariant Analysis



Credit Grade and Default Risk

A distinct pattern emerges: there is a significantly higher probability of default for borrowers classified under credit grades E, F, and G.

Bivariant Analysis



- Small business loan seems to be defaulting more compared to other lending categories

Conclusion

● To mitigate the risk of loan defaults, Lending Club ought to consider lowering the interest rates on loans with a 60-month tenure that currently carry high rates.

● Utilizing grades as a reliable indicator, Lending Club could enhance its default prediction by requesting additional information from potential borrowers falling within the lower grade range (G to A) prior to loan approval.

● Given the higher default rate associated with small business loans, Lending Club should contemplate either reducing or ceasing the issuance of loans to this category.

● Lending Club should exercise caution when granting loans exceeding $12,000 to borrowers who possess mortgage home ownership, as they appear to be both requesting larger loan amounts and exhibiting a higher propensity for default.