



Loan Default Analysis for Lending Club Case Study

SYED RAMEEZ &
RANGANATH SHEERYA

Case Study approach

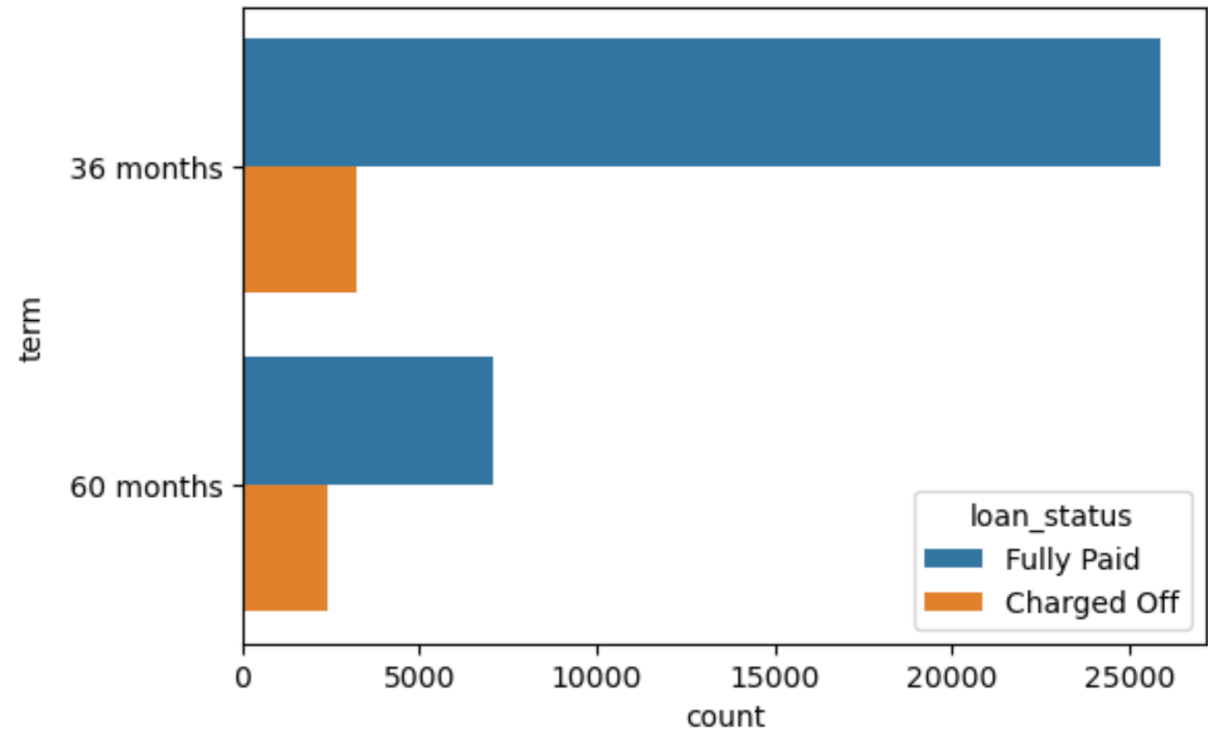
| ASK | DATA CLEANING | DATA ANALYSIS USING UNIVARIATE AND BIVARIATE SEGMENTS | CONCLUSION |
|--|--|---|---|
| Lending Club is the largest Consumer Finance company which facilitates personal loans, business loans and financing medical procedures | Working with the Data Dictionary and getting knowledge of all the columns and their domain specific uses of banking. | Analysing each column, plotting the distributions of each column. | Whom to approve/reject the loan decision depends on a few factors |
| Lending Club has two types of risks to mitigate. Provide easily accessible lower interest rates through a fast online interface | Removing the unnecessary Variables, null valued columns and removing the respective rows/columns. | Analysing the two variable behaviour like term and loan status with respect to loan amount. | How these deductions are made based on data analysis |

Data Overview

- ❖ Dataset contains 39,717 records and 111 columns.
- ❖ Types of data: Loan details, borrower information, and loan status.
- ❖ Columns with more than 10,000 missing values were removed
- ❖ The dataset was reduced from 111 to 53 columns by removing these columns.
- ❖ Dropped columns with unique values and those not relevant to loan approval and the dataset was further reduced to 39717 rows and 28 columns.
- ❖ Loan status 'Current' was removed as it does not help in decision-making.
- ❖ Added new columns for issue year and month by splitting the issue date.

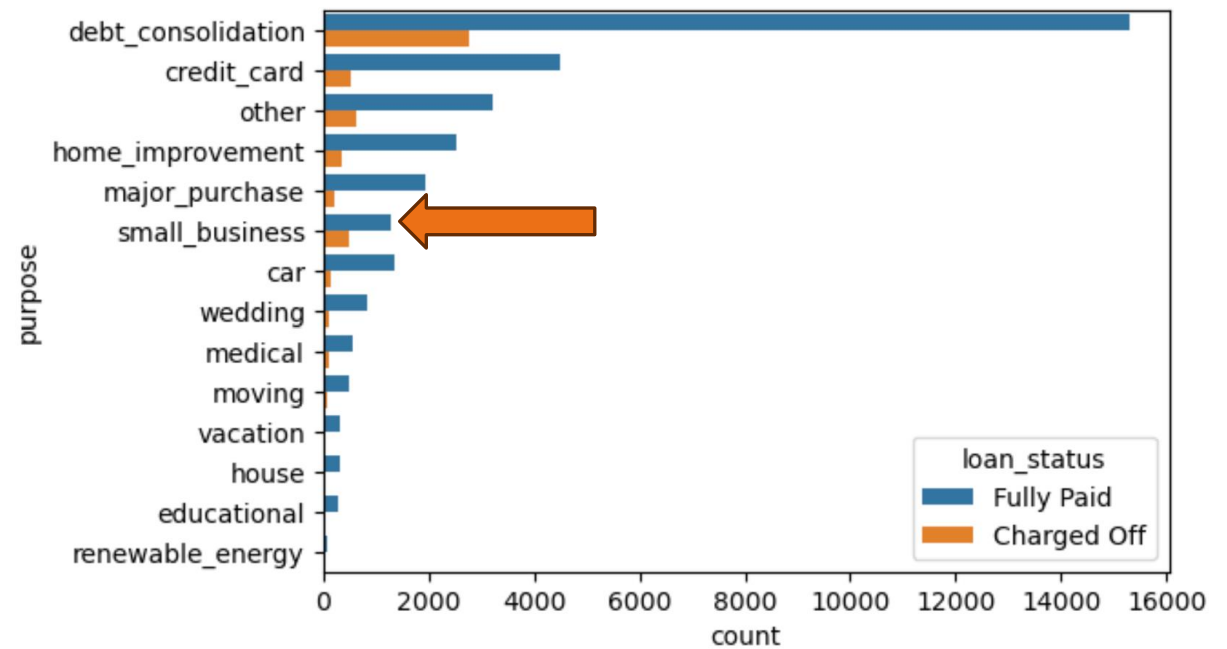
Univariate Analysis

❖ From the bar plot graph we can conclude that the ratio of charged-off is higher if the loan tenure is of 60months.



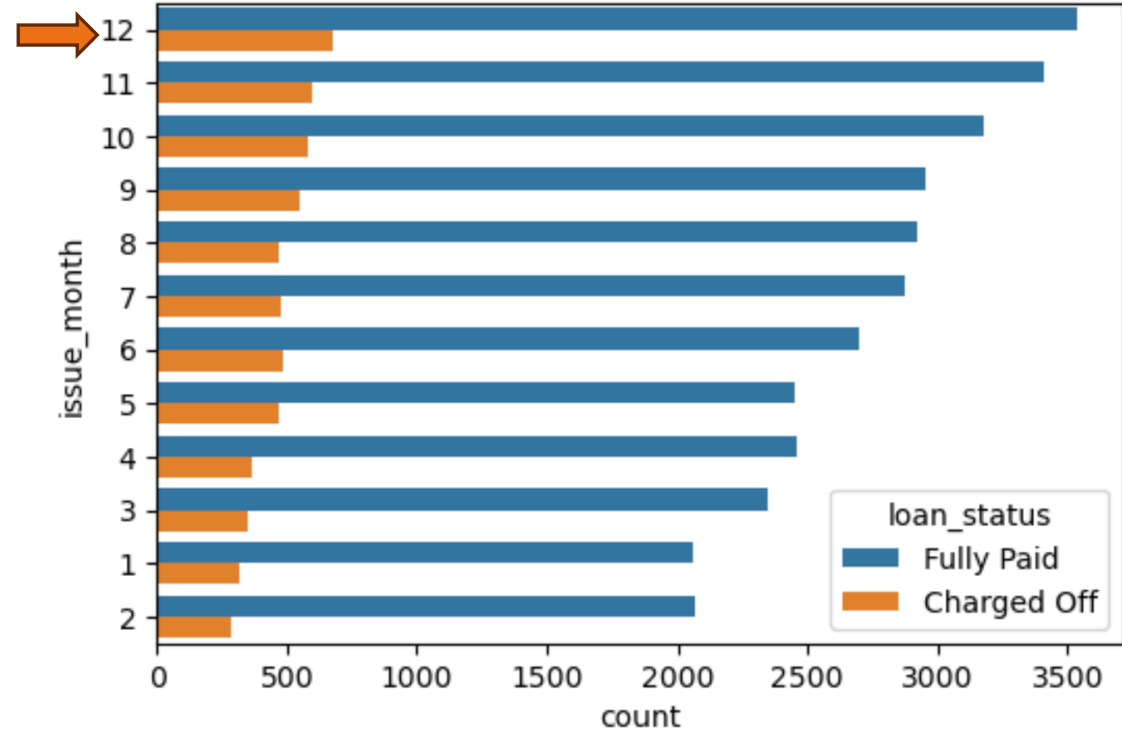
Univariate Analysis

❖ Approving loans to small business is turning out to be more risk.



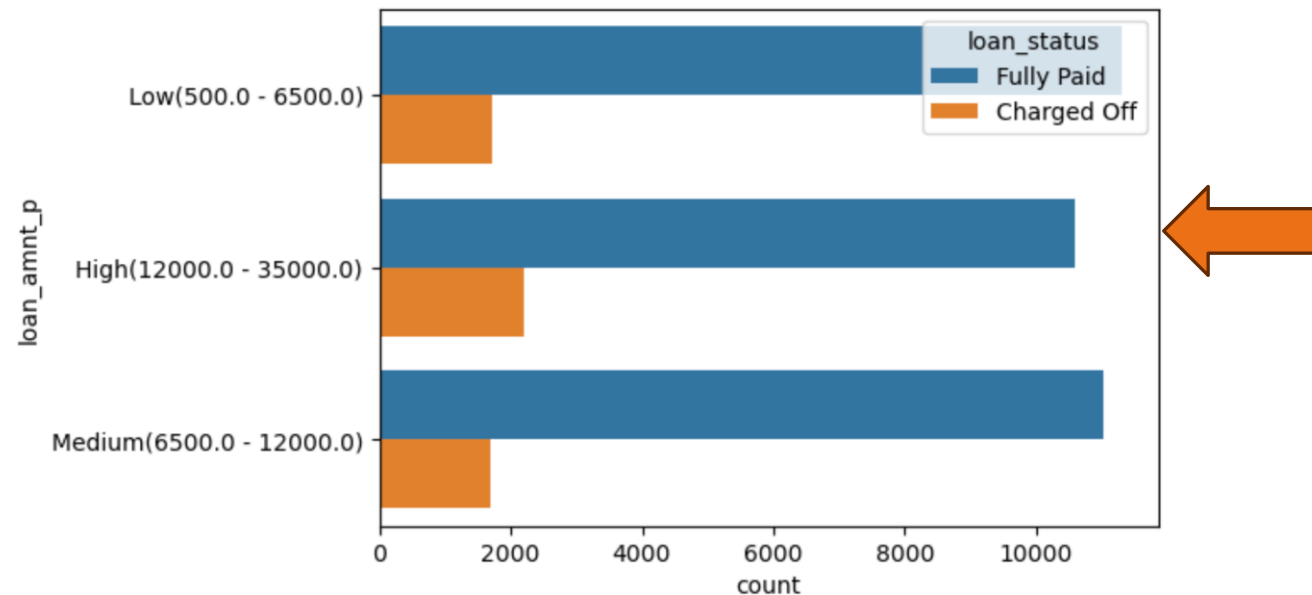
Univariate Analysis:

The higher the loan amount, the greater the chance of the loan getting defaulted.



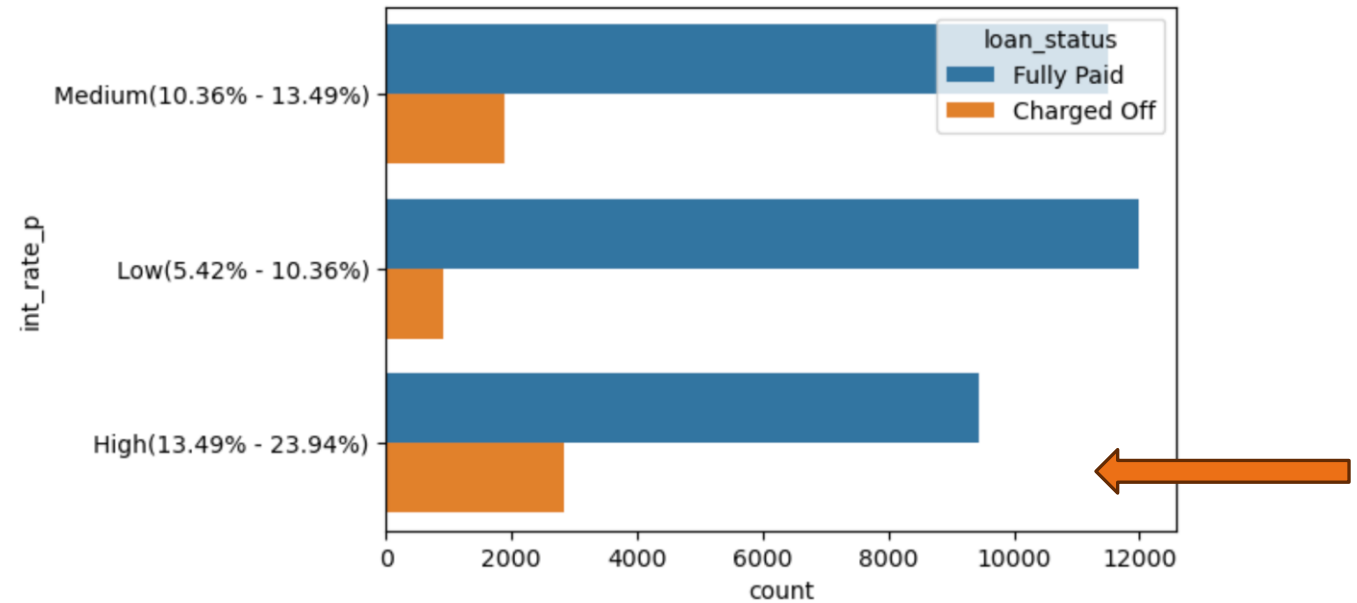
Univariate Analysis:

This data shows that the higher the loan amount higher are the chances of charged-off



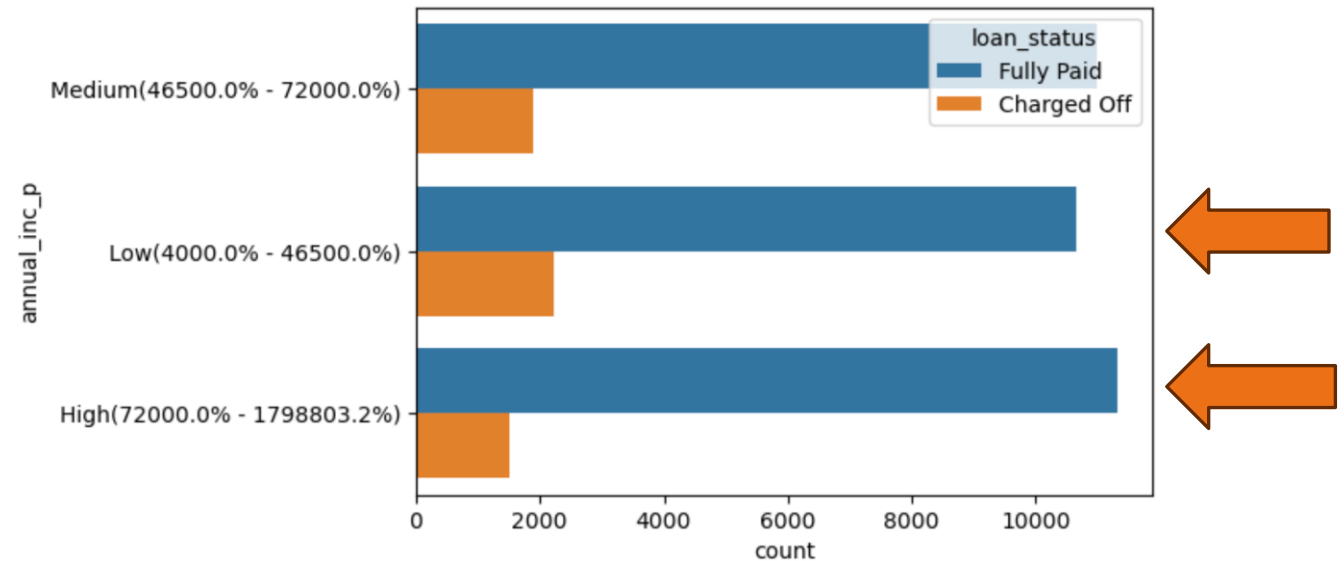
Univariate Analysis:

Higher interest rate leads to higher charged off or defaulter



Univariate Analysis:

The higher the income higher the repayment percentage and lower the income higher the defaulters



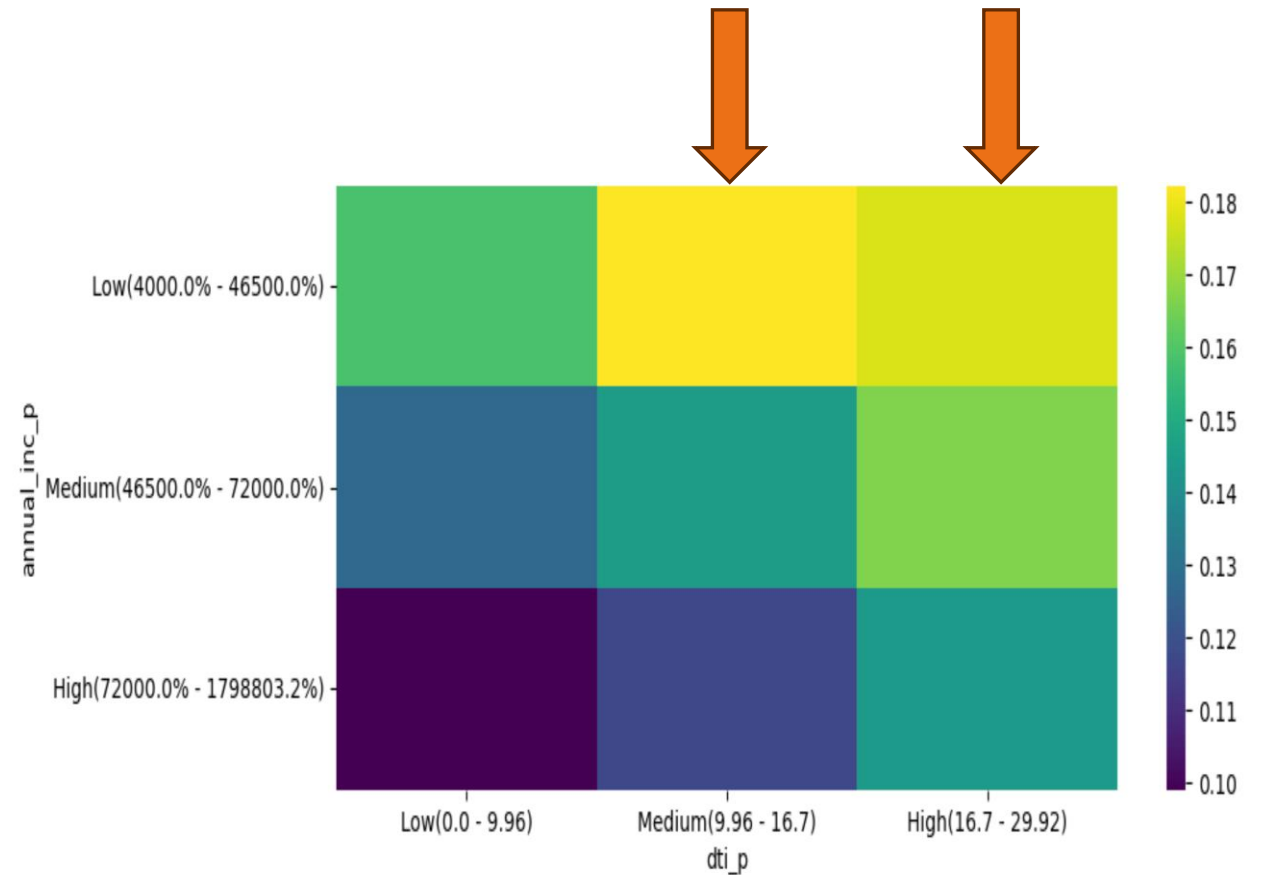
Bi-variate Analysis:

- ❖ The intersection of home ownership (None) and purpose (Other) has a darker color and represents riskier
- ❖ The intersection of home ownership (None) and purpose (debt_consolidation) has the darker color representing riskier
- ❖ OTHER with small business ,moving, car purpose is riskier



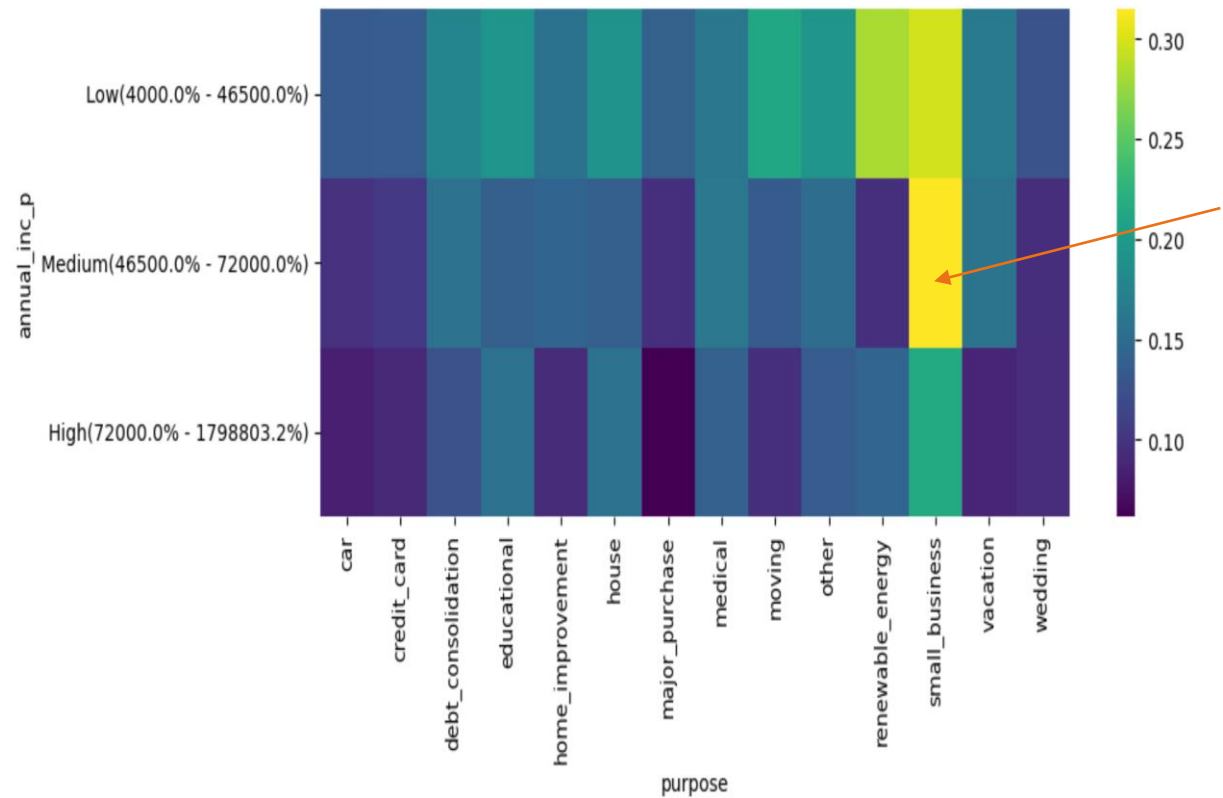
Bivariate Analysis:

Medium to High debt to income vs
lower the annual income risk is
higher



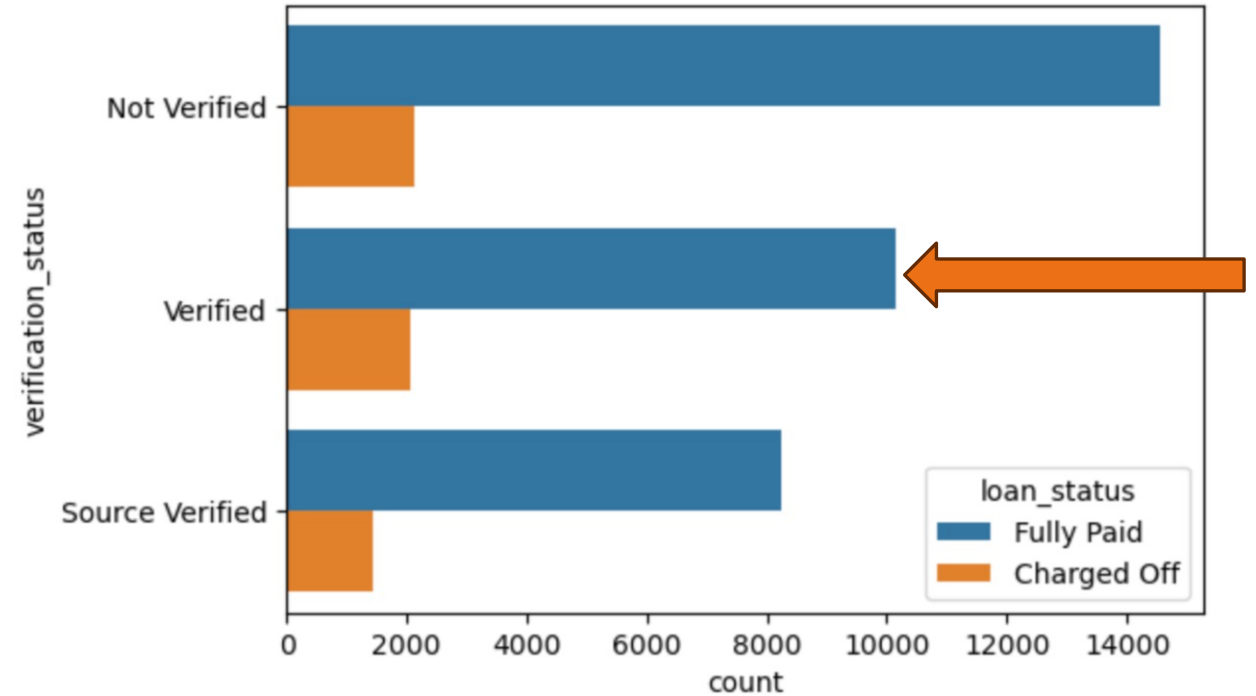
Bi-variate Analysis:

- ❖ Applicant with medium income applying loan for small business purpose is very risky.
- ❖ Applicant with high income applying loan for car, credit_card, major_purchase, home_improvement, vacation, wedding purpose less risky.



Recommendation:

- ❖ Defaulters are more in verified applicants as well which signifies verification process needs improvement.



Conclusion-1,Key Findings

1. The higher the loan amount, the greater the chance of the loan getting defaulted.
2. Higher interest rate leads to higher charged-off or defaulter
3. The higher the installment amounts show the higher the default percentages.

Conclusion-2, Recommendations

1. Verified applicants are more defaulters which means the verification process is not accurate. This process needs to be revamped.
2. People with more public derogatory records have more chance of filing a bankruptcy. The lending club should make sure there are no public derogatory records for the borrower.
3. Borrowers with mortgage homeownership are taking higher loans and defaulting. The lending club should stop giving loans to this category when the loan amount requested is more than 12000.