Lending Club Case Study

Jyotee Ingale

Problem Statement

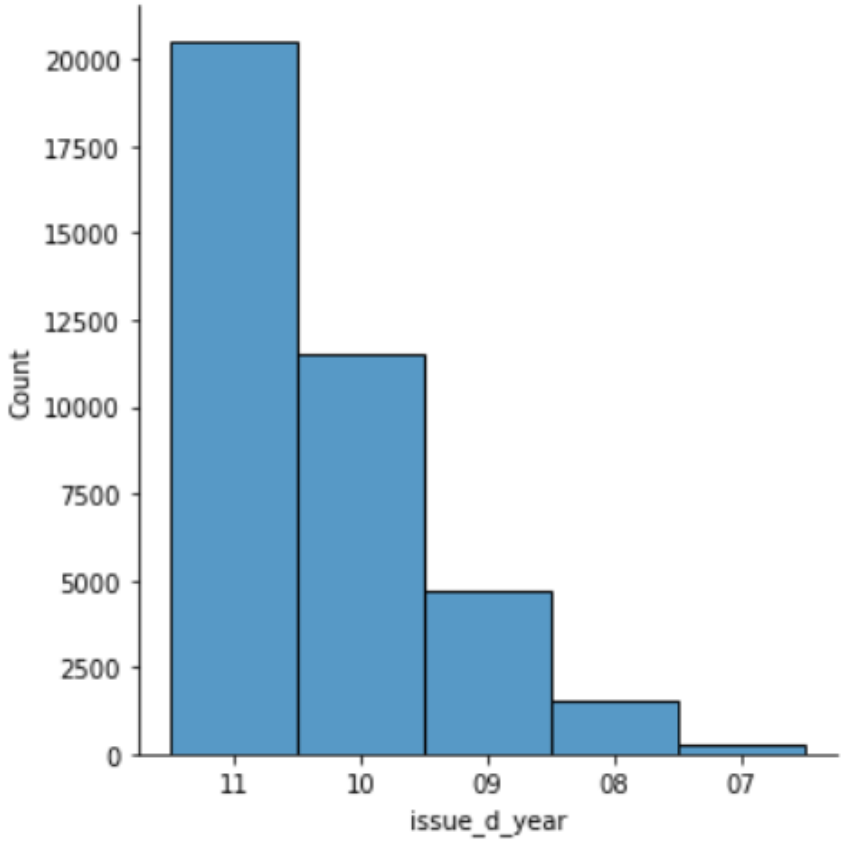
The company wants to understand the **driving factors (or driver variables)** behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

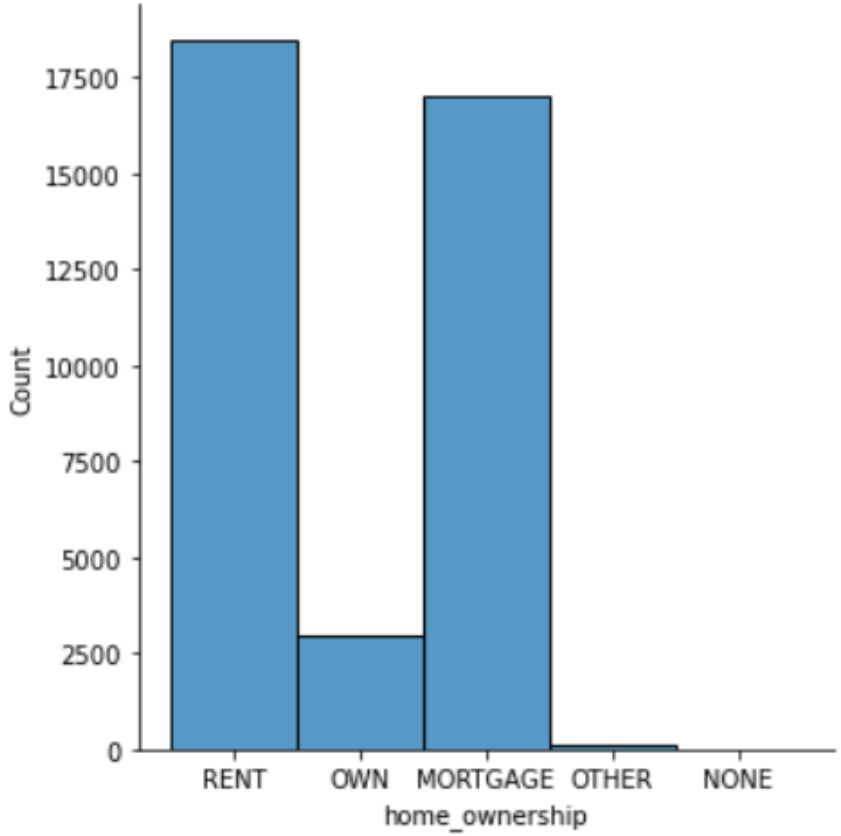
Approach

- I first understood the data and then based on understanding I removed the current loan_status rows as they don't add any value in data insights.
- I cleanse the data by removing NULL columns, rows & duplicate values
- Removed the columns related to current borrower which has not fully paid or defaulted.

Below is the Univariate variable where you can see how we can show the described data using only one variable.

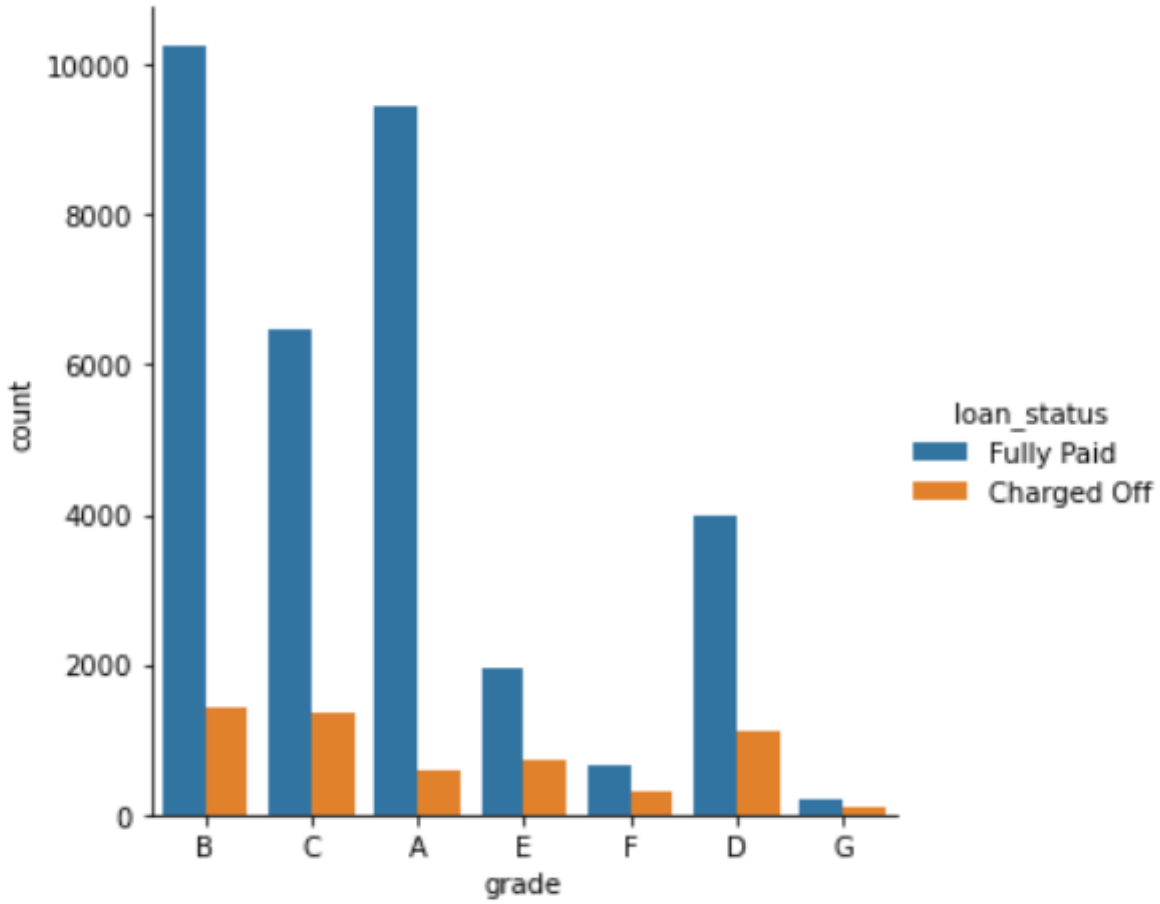
First graph shows the number of loans taken over the 5 years of period.
Second graph shows number of members with different home ownership status.

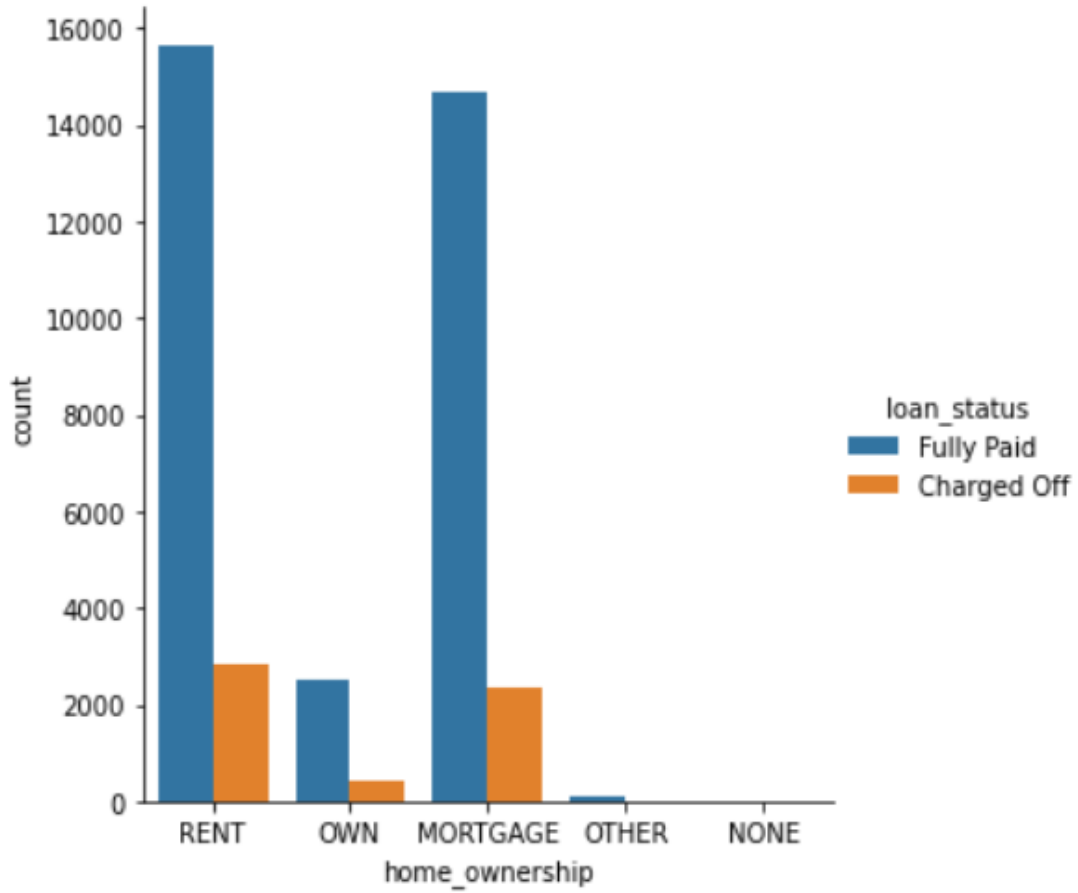




Below is the Univariate variable where you can see how we can show the described data using two variables.

First graph shows the number of members with their respective grades as fully paid or defaulted.
Second graph shows number of members with different home ownership status as wither fully paid or defaulted.





Recommendations

- While giving the loan or verifying the borrower we should consider the rejected data as well like when was the last time borrower applied for the loan. So rejected data also should be in the dataset.

- Borrower should have at least 6 months of experience and source verified which can decrease the risk of not paying the loan.