LENDING CLUB CASE STUDY

PRESENTED BY

SHOBHIT SRIVASTAVA

ABOUT THE COMPANY

- The Lending company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

PROBLEM STATEMENT

- The company wants to minimize its credit loss
- They want to identify if the applicant is risky and is likely to default the loan
- Want to understand the driving factor behind the loan default
- The company can utilize this knowledge for its portfolio and risk assessment.

ANALYSIS

- The approach used to analyze the problem is :

1. Try to understand trend of loan deffault
2. Perform univariate analysis on
    1. Loan Amount
    2. Interest Rate
    3. Loan Term
    4. Loan Grade etc.
3. Perform Bivariate analysis on
    1. Loan Amount vs. Loan Grade
    2. Credit utilization vs Loan Status

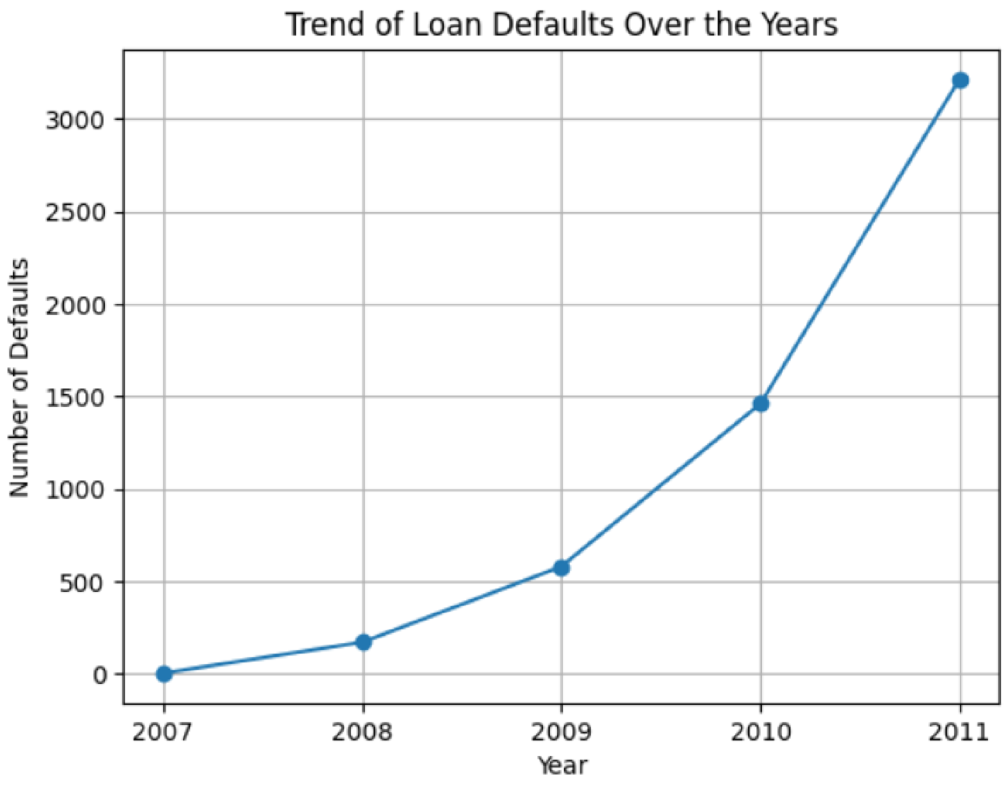
DATA CLEANING AND IMPUTATION

- Before cleaning, there were 39717 records with 111 columns
- 54 columns contain NA values only, and these columns were removed.
- Emp_title and desc is filled with "Not Available in case if the data is missing
- Columns with single value and not relevant in EDA were removed
- Deriving value from url will duplicate the I column and thus, url not considered
- Columns with loan_status as Current has not been considered.
- Removed single record for revol_util whose value was null
- Year and month were derived from the date columns
- revol_util, last_pymnt_d, last_credit_pull_d, pub_rec_bankruptcies – only non-null records considered

- % sign was removed from int_rate and revol_util
- Rounded off float type columns to 2 decimal places
- Employment length was mapped as:

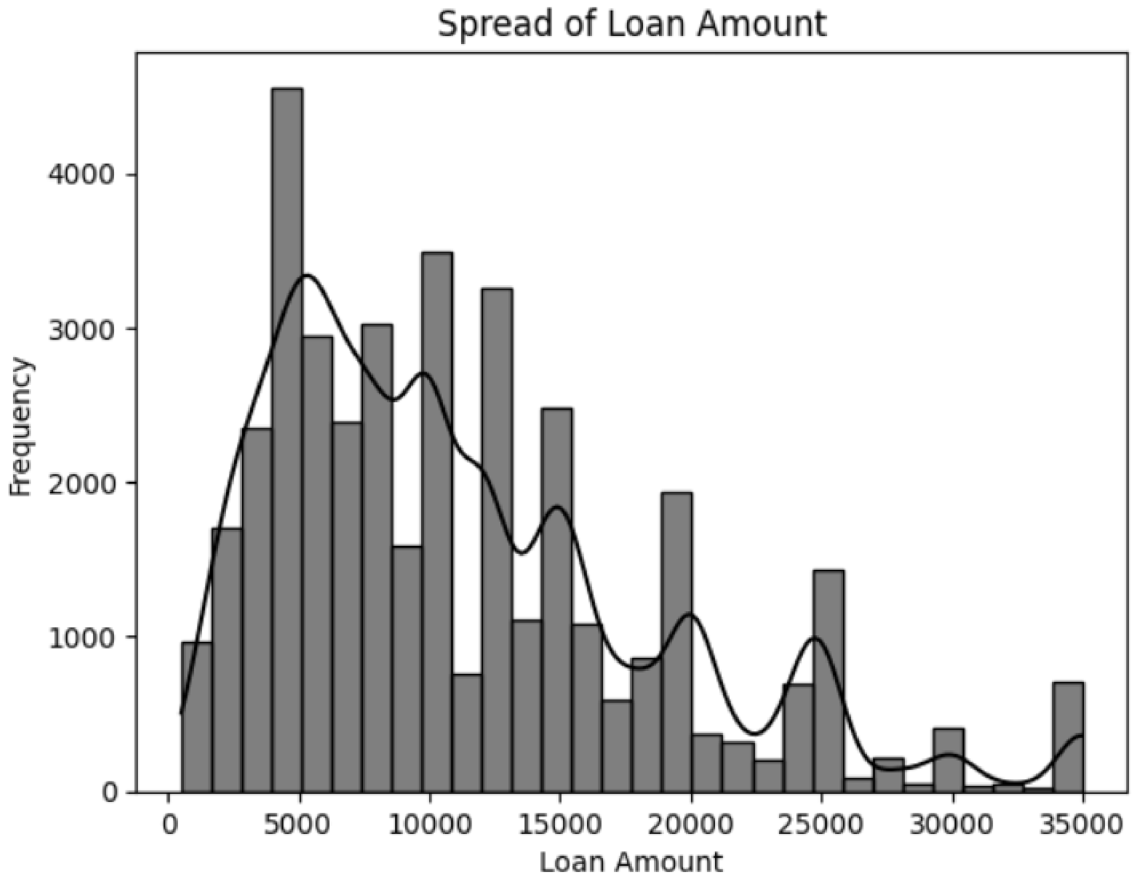
```
'< 1 year': 0,
'1 year': 1,
'2 years': 2,
'3 years': 3,
'4 years': 4,
'5 years': 5,
'6 years': 6,
'7 years': 7,
'8 years': 8,
'9 years': 9,
'10+ years': 10}
```

LOAN DEFAULT TREND



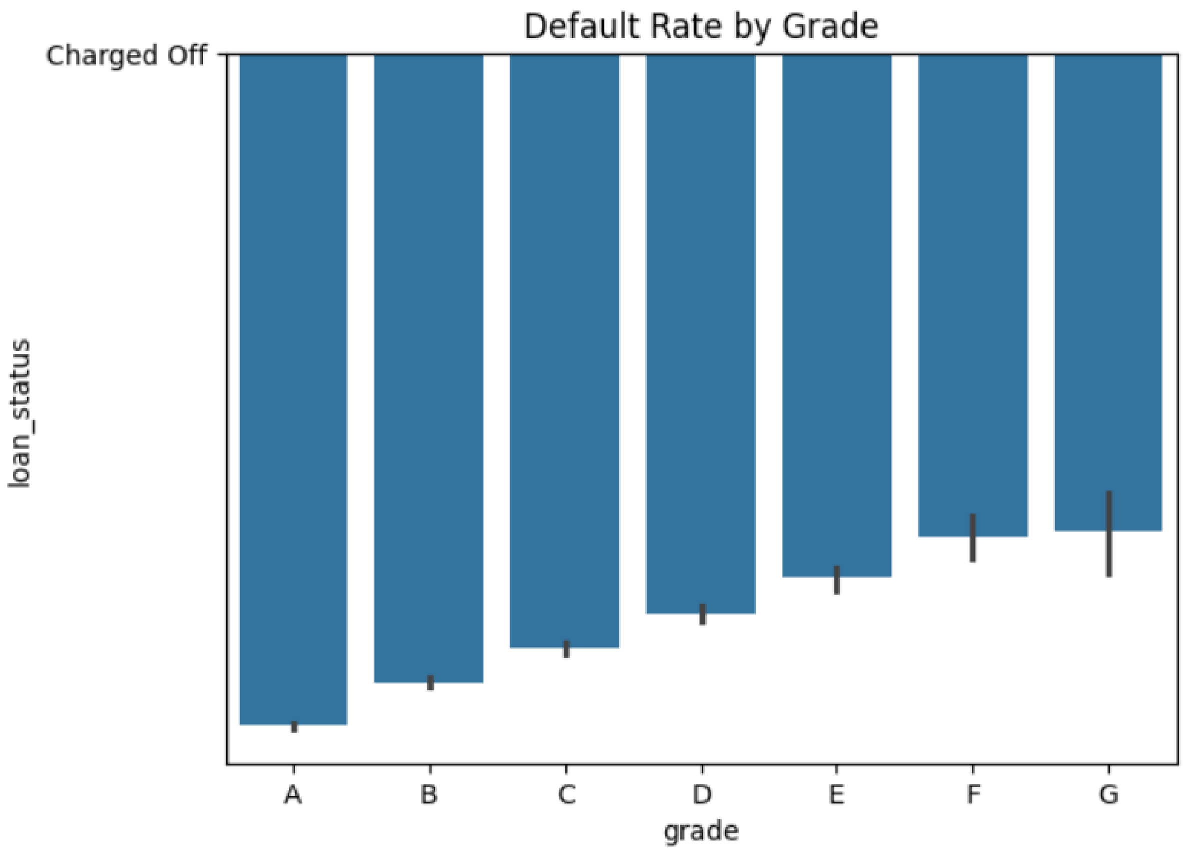
- Over the period of 5 years, the count of defaulted has increased exponentially
- It has tripled from year 2009 to 2010 and increased ~2.5 times between the year 2010 and 2011

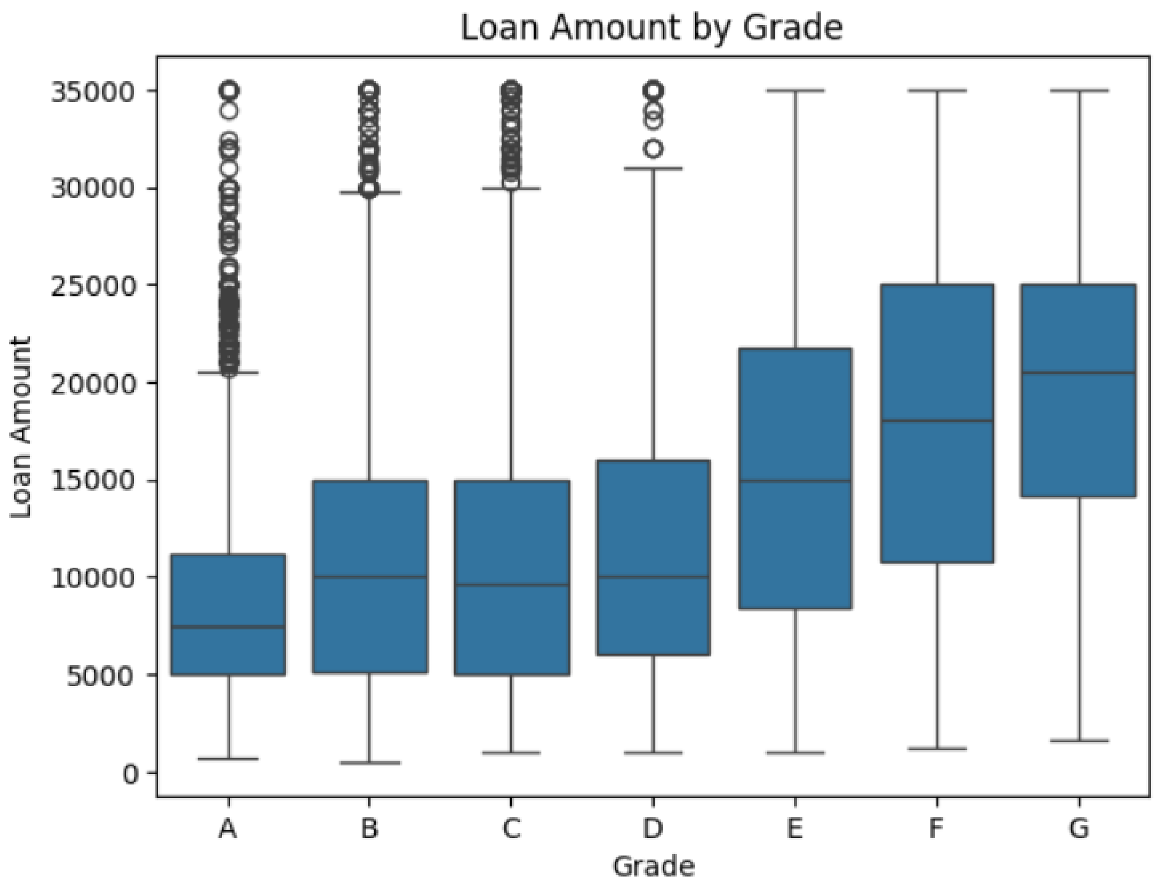
SPREAD OF LOAN AMOUNT



- Lower count of loan as we move towards high loan amount
- Most of the loan amount is towards lower value and much availed by the borrowers.

GARDE VS LOAN AMOUNT VS LOAN STATUS



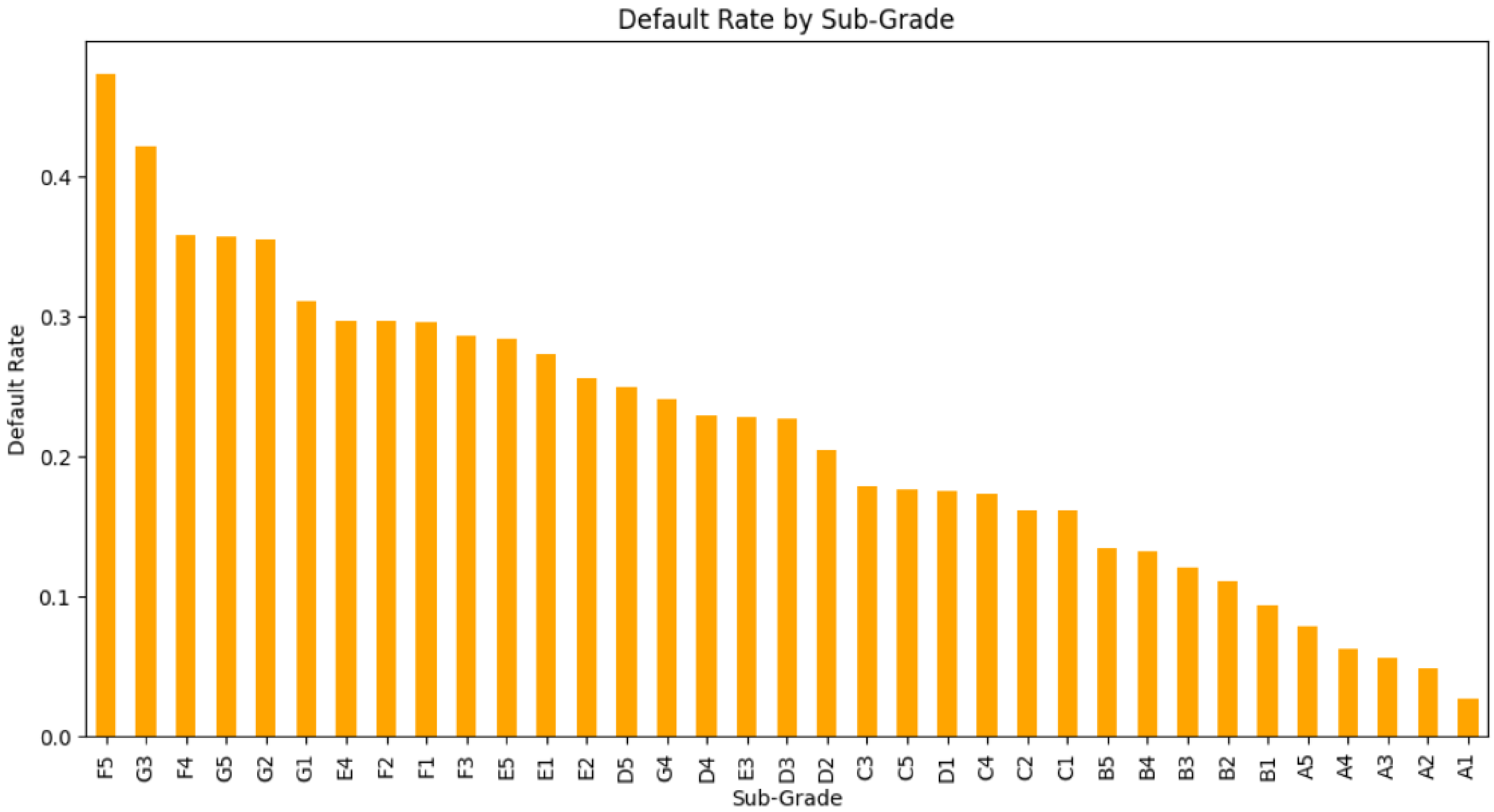


CONTD…

- Grade A: Loans with Grade A have the lowest default rate, indicating that these loans are the least risky.
- Grade B: Loans with Grade B have a slightly higher default rate than Grade A but are still relatively low risk.
- Grade C: The default rate increases for Grade C loans, suggesting a moderate risk level.
- Grades D to G: The default rate continues to increase from Grade D to Grade G, with Grade G having the highest default rate. This indicates that loans with lower grades (closer to G) are significantly riskier

- The IQR is wider for lower grades (D to G) compared to higher grades (A to C). This indicates more variability in loan amounts for lower-grade loans
- The higher loan amounts and greater variability in **lower-grade loans suggest that these loans might carry more risk**.

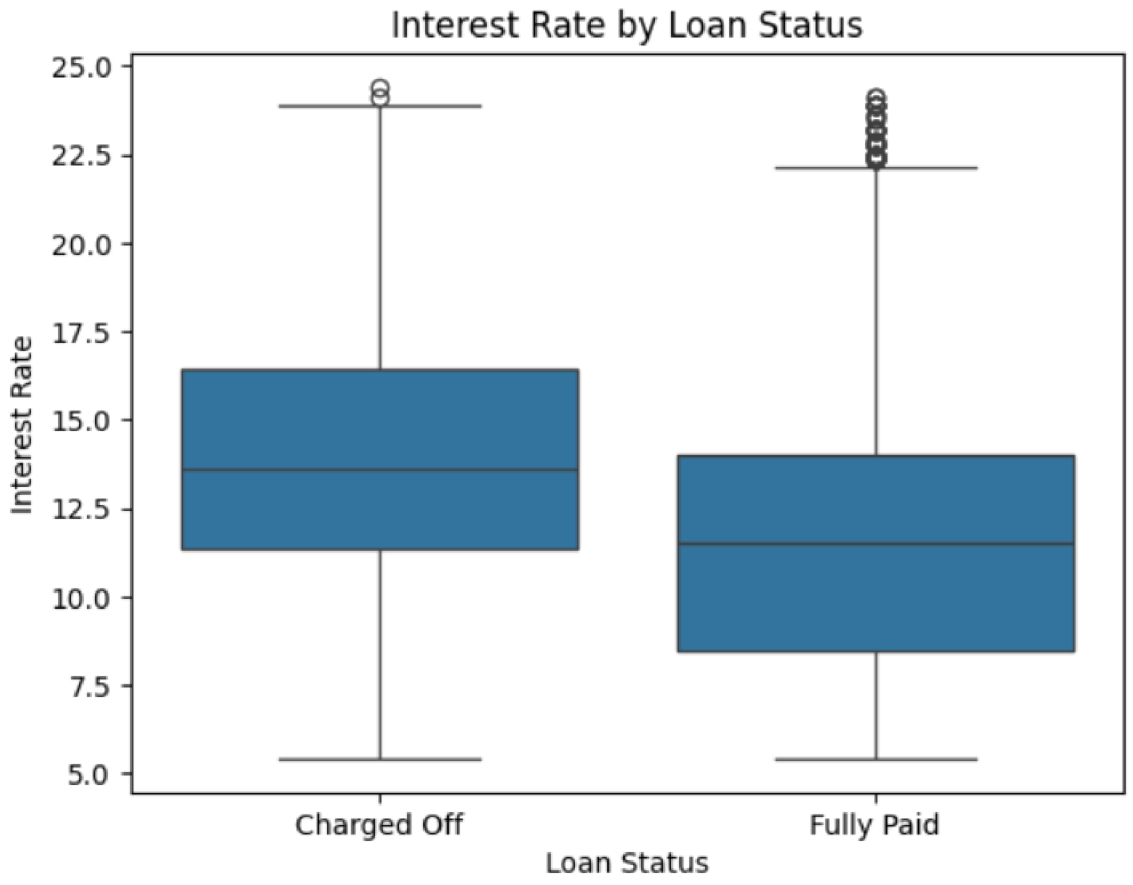
LOAN DEFAULTS BY SUB GRADE



Among the lower grades – D to G which have the most likelihood to default, the top 5 sub grades with > 0.3 default rate are:

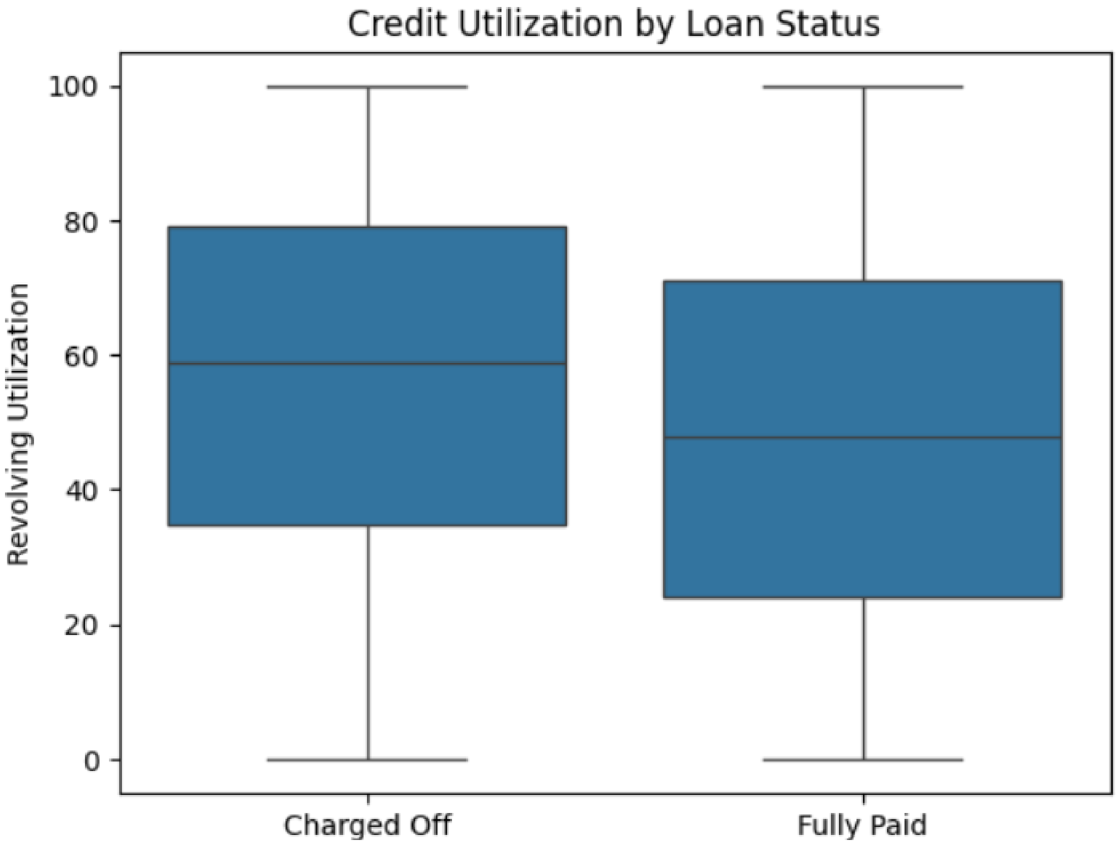
1. F5
2. G3
3. F4
4. G5
5. G2

INTEREST RATE VS LOAN STATUS



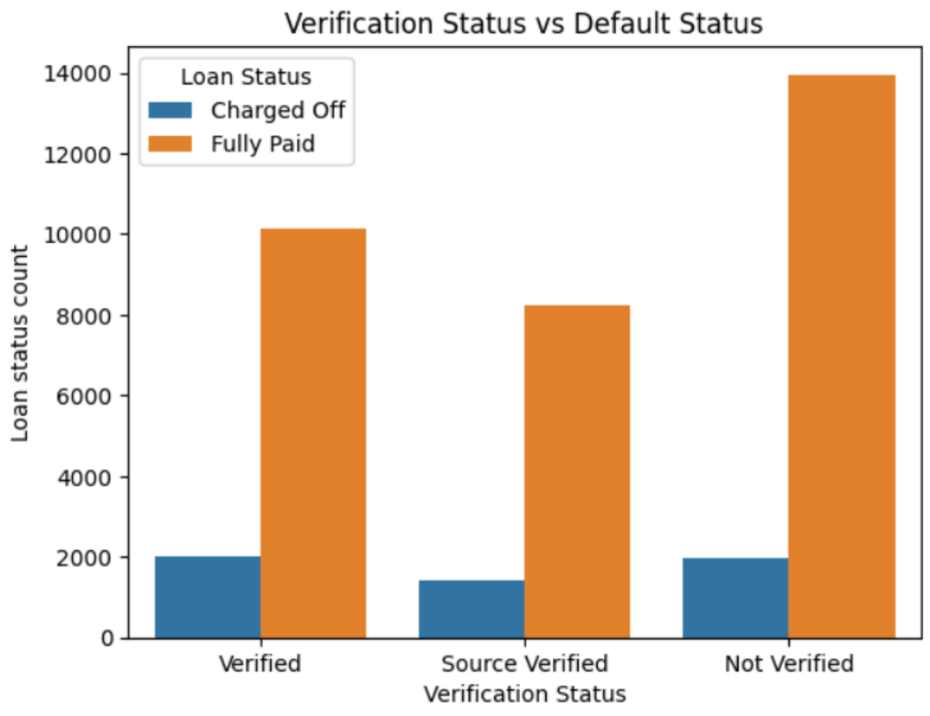
- The 50 percentile for Charged off loan is higher than the fully paid loans and suggests that loans with higher interest rates are more likely to default.
- The company should monitor such cases closely

CREDIT UTILIZATION VS LOAN STATUS



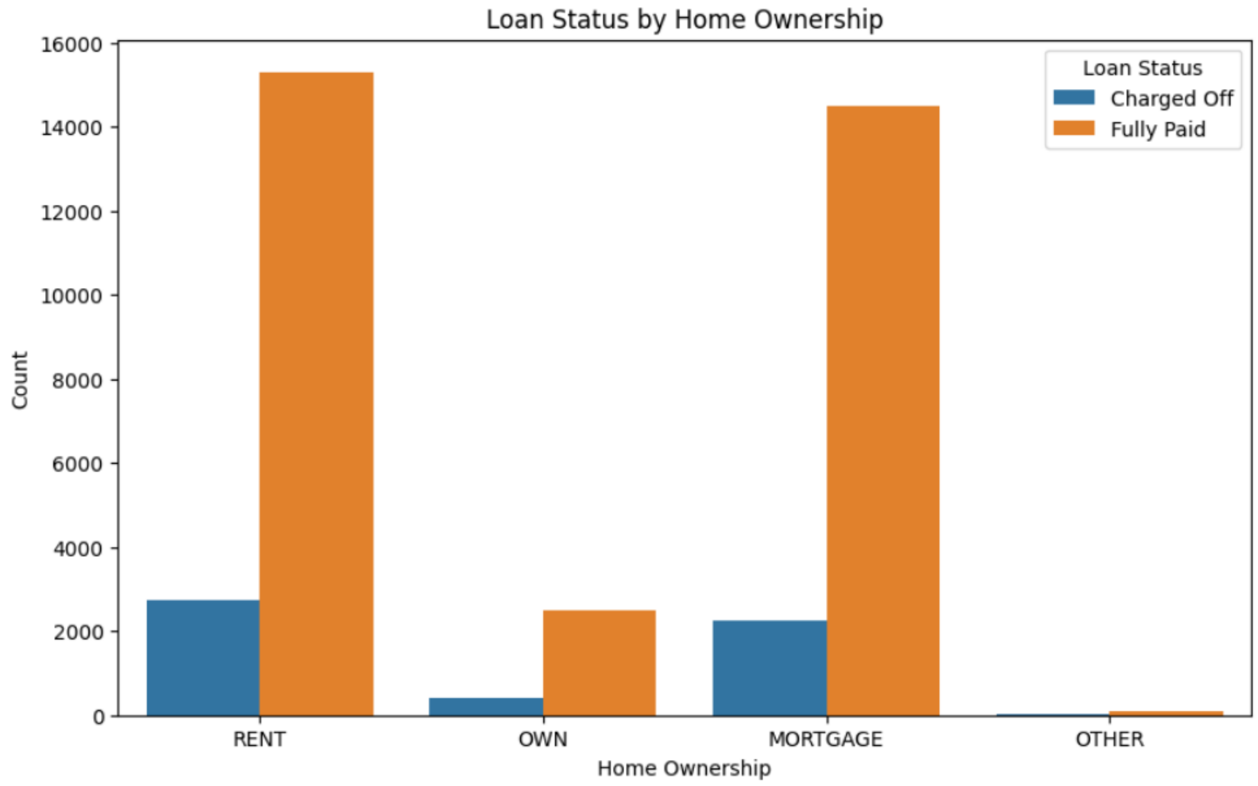
- Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
- A high rate suggests potential risk of default
- The 50 percentile is higher for Charged Off loans compared to Fully Paid loans. This suggests that borrowers with higher credit utilization are more likely to default on their loans

VERIFICATION VS LOAN DEFAULTS



- Verification Status ndicates if income was verified by LC, not verified, or if the income source was verified
- We see that even for the verified loans and source verified loans, there are defaults happening and thus, this should be looked in the lending company.
- If any third party is doing the verification, they may consider changing it.

LOAN STATUS BY OWNERSHIP



- Borrowers with a MORTGAGE or on RENT have the highest number of loans
- Borrowers who are on RENT or have a MORTGAGE tend to have a higher count of charged off loans compared to those who are having their OWN house.

SUMMARY

- Loans with ‘Not Verified’ status have the highest count of both fully paid and charged off loans
- Higher interest rates are associated with a higher likelihood of loan default. The median interest rate for charged off loans is higher compared to fully paid loans.
- Higher credit utilization is associated with a higher likelihood of loan default. Borrowers with higher revolving utilization are more likely to default.
- **Lower-grade loans (closer to G) have higher default rates. The default rate increases as the grade decreases from A to G. This has the highest impact on the loan default**
- Renters and those have a mortgage have a significant number of charged off loans