Lending Analysis

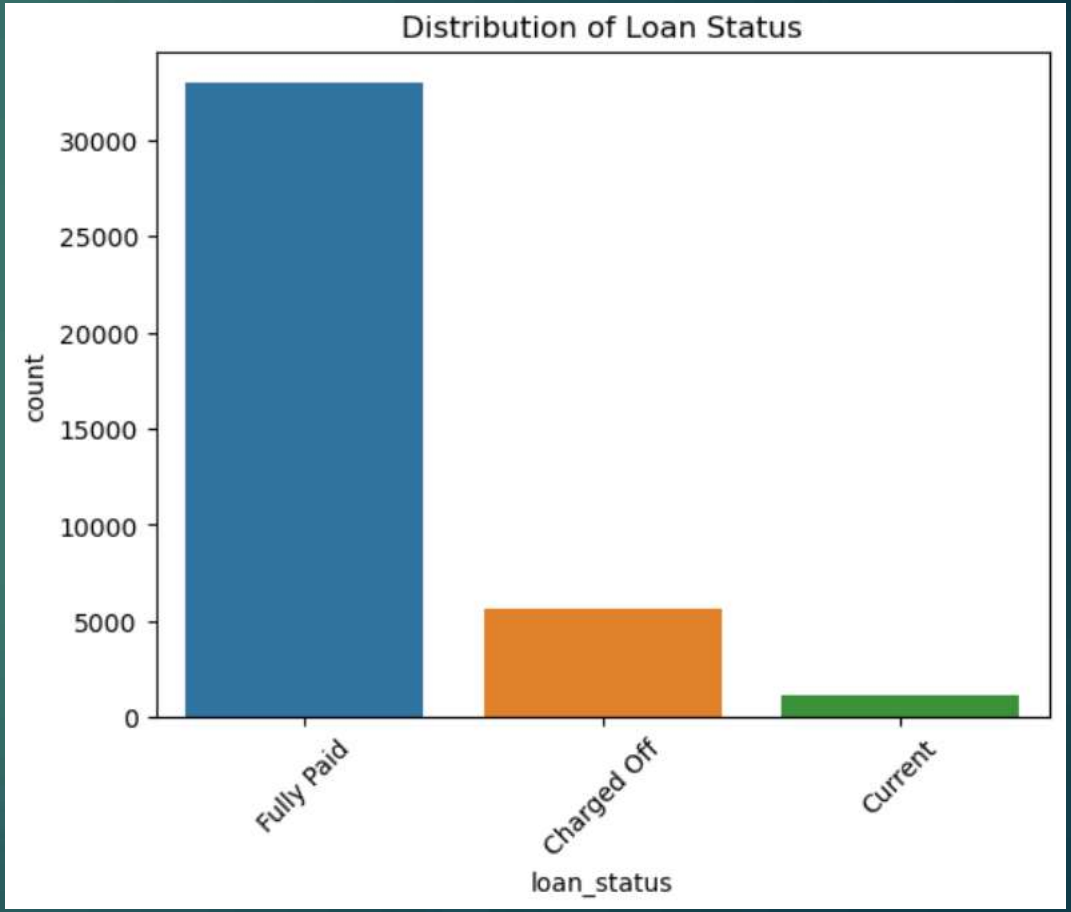
Dataset Description

- The dataset consists of 39,717 entries with 111 columns, including various features related to loan applications, such as loan amount, interest rate, term, grade, and more. There are a mix of numerical and categorical data types.
- We will take the following steps here to analyse our data:

1. Data Cleaning: Handle missing values, inconsistent data types, and irrelevant columns.
2. Exploratory Data Analysis **(EDA)**: Identify key features that may influence loan default.
3. Insights and Observations: Summarize findings that indicate the likelihood of default.

Loan Status Analysis of the given data

- On going through the loan status of the current borrowers through the given graph,the following observations can be made:



- **Fully Paid**: 82.96% of the loans were fully paid, which indicates that the majority of borrowers successfully repaid their loans.
- **Charged Off**: 14.17% of the loans were charged off, meaning these borrowers defaulted on their loans. This represents the key area of concern for the company, as these defaults contribute to credit loss.
- **Current**: 2.87% of the loans are currently in the process of being repaid, meaning these borrowers are neither fully paid nor defaulted yet.

The Following Conclusions can be drawn through the initial analysis:

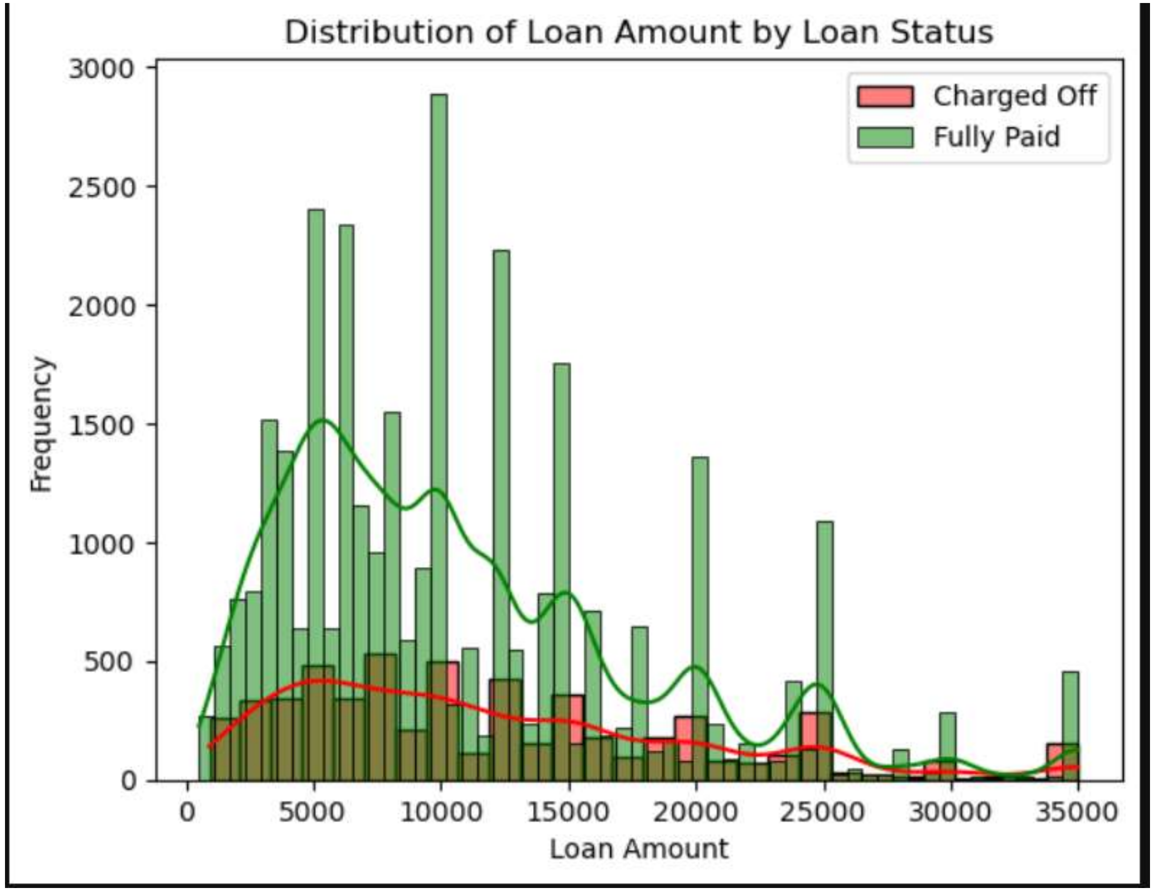
- **High Repayment Rate**: With 82.96% of loans being fully paid, the company has a relatively strong repayment rate. This suggests that the company's lending policies are generally effective for most borrowers.

- **Significant Default Rate**: The 14.17% default rate is non-negligible and represents a significant potential loss for the company. This indicates that there is room for improvement in identifying high-risk borrowers before loans are issued.

- **Current Loans**: Since 2.87% of loans are currently active, it’s important to continue monitoring these to anticipate potential defaults. The performance of these loans in the near future could slightly adjust the current statistics

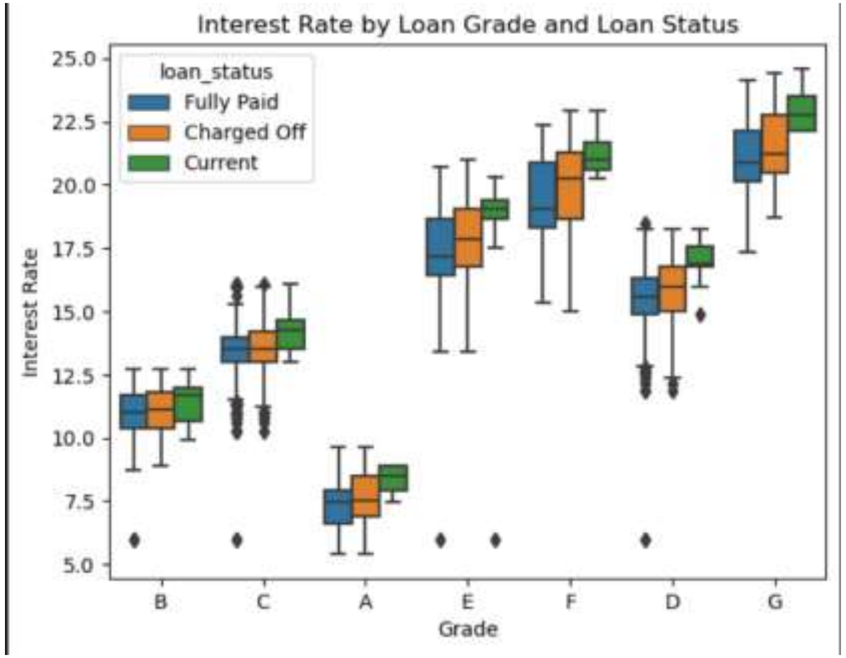
Loan Amount vs Status

- **For smaller amounts the frequency of payoff is higher than for bigger amounts not conclusive as seen by the given graphs, the data is inconclusive to make an accurate prediction**



Bivariate Analysis

- We can deduce the loan grade and interest rate are key factors while determining defaults in the loans,the plot gives further information regarding this relationship in regards to loan status



Observations:

Interest Rates by Loan Grade:

- As the loan grade progresses from A to G, the interest rates increase. This is expected since higher grades (A, B, C) are associated with lower risk, and therefore, lower interest rates. Conversely, lower grades (D, E, F, G) indicate higher risk, leading to higher interest rates.

Loan Status Distribution:

- For each loan grade, you can observe three distinct groups corresponding to Fully Paid, Charged Off, and Current loans. These indicate how borrowers within each grade have repaid their loans:
  - **Fully Paid (Blue)**: These borrowers have completed their loan payments.
  - **Charged Off (Orange)**: These are the defaulters who have not repaid the loan.
  - **Current (Green)**: These borrowers are still in the process of repayment.

Patterns of Default (Charged Off):

- **Grade A**: Low interest rates, with few defaults (most loans are Fully Paid).
- **Grade B and C**: A mix of Fully Paid and Charged Off, with a slight increase in the proportion of defaults as interest rates increase.
- **Grades D to G**: As grades decline, there's a noticeable increase in Charged Off loans. Especially in Grades F and G, where the interest rates are high, the proportion of defaults increases significantly.

Risk Indicators:

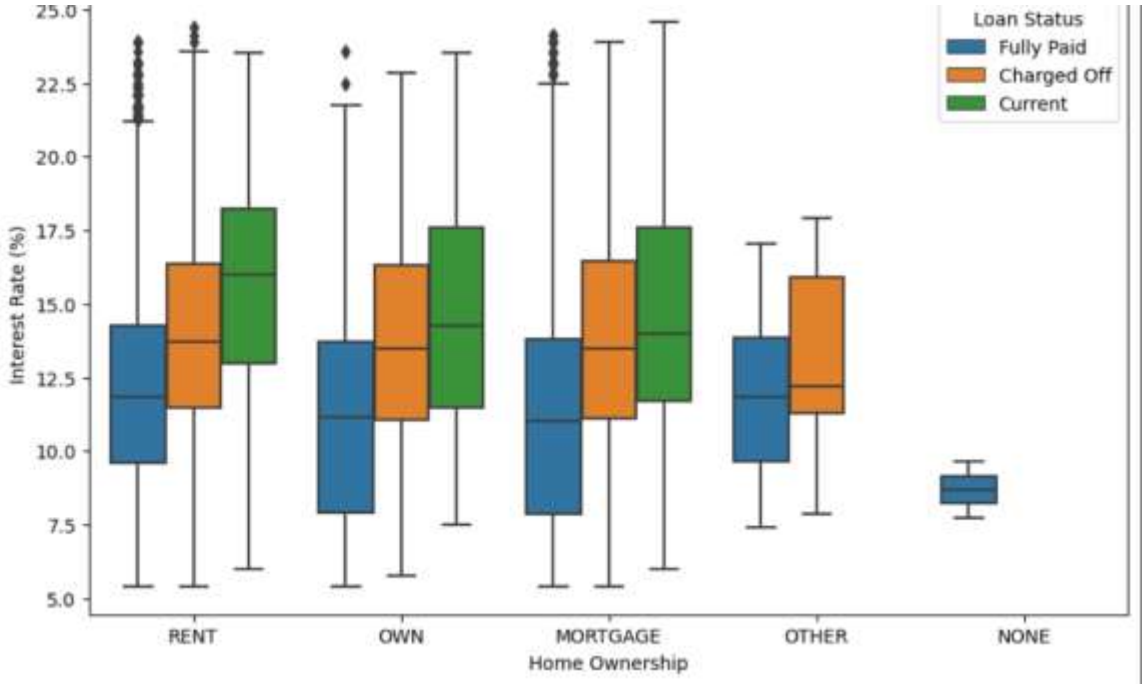
- **High Default Rates in Lower Grades**: The box plot clearly shows that as the grade gets worse (D, E, F, G), the interest rates are higher, and the proportion of Charged Off loans increases. This suggests that borrowers in these grades are riskier.
- **Interest Rate Variation**: Even within a grade, there is a variation in interest rates. For example, in Grade C, there are some Charged Off loans at higher interest rates, while Fully Paid loans tend to have slightly lower rates. This could indicate that borrowers with higher interest rates within a grade are more likely to default.

Conclusions

1. **Interest Rate as a Predictor**:
   - Interest rate appears to be a strong predictor of loan default. Higher interest rates, especially within lower grades (e.g., D, E, F, G), are associated with a higher likelihood of default.
2. **Grade as a Risk Indicator**:
   - Loan grade is another important predictor. As the grade worsens (from A to G), the risk of default increases, which is reflected in the increased proportion of Charged Off loans.
3. **Combined Effect**:
   - The combination of loan grade and interest rate provides a clear risk profile. Borrowers with lower grades and higher interest rates are significantly more likely to default, which can be used to adjust loan approval criteria or pricing strategies

Bivariate Analysis

- Home ownership status plays a key role in loan defaulting or repayment,here we can see how it affects borrowers generally with respect to loan amount and interest rates



Observations

1. **Renters**:
   - **Charged Off**: Renters who defaulted (Charged Off) tend to have a slightly higher median interest rate compared to those who fully repaid their loans (Fully Paid).
   - **Fully Paid**: The median interest rate for loans that were fully paid by renters is lower, indicating that lower interest rates might be associated with successful loan repayments.
   - **Current**: The interest rate distribution for current loans is on the higher side, similar to Charged Off.
2. **Homeowners**:
   - **Own**: Borrowers who own their homes without a mortgage show a pattern where the interest rates for Charged Off and Fully Paid loans are close, but those who defaulted have a slightly higher median interest rate.
   - **Mortgage**: Borrowers with mortgages also exhibit a similar pattern, with Charged Off loans generally having a slightly higher interest rate compared to Fully Paid loans.
   - **Current**: For both Own and Mortgage, current loans tend to have higher interest rates.

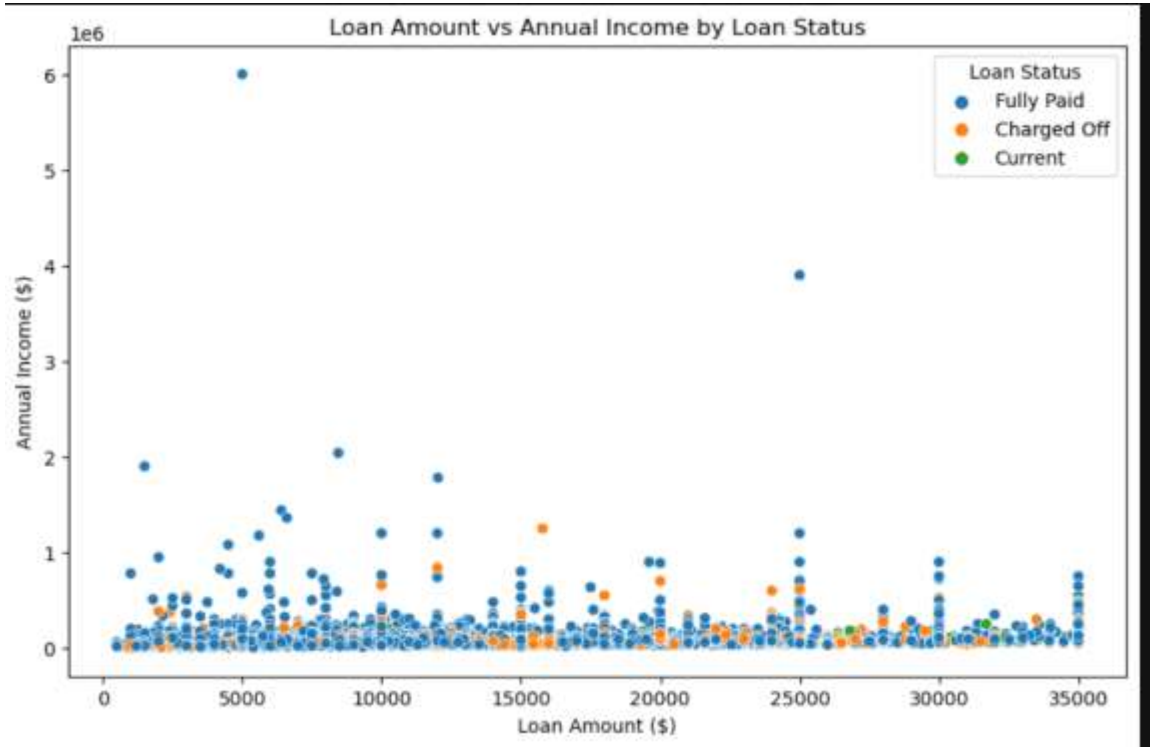
Conclusions

- **Interest Rate as a Risk Indicator**: Higher interest rates are associated with a greater likelihood of loan default across all categories of home ownership. This is particularly evident in the Rent, Own, and Mortgage groups where Charged Off loans generally have higher median interest rates than Fully Paid loans.
- **Home Ownership Status**: Renters and those in the Other category appear to be at higher risk, as indicated by the higher median interest rates for Charged Off loans. Homeowners with or without a mortgage have lower median interest rates and a somewhat lower likelihood of default.
- **Policy Implications**: These insights suggest that interest rates could be used as a predictive factor for loan defaults. Adjusting interest rates based on home ownership status could help in better managing risk.

This analysis can inform lending practices by highlighting how the interaction between interest rates, home ownership status, and loan outcomes can predict default risk, helping to mitigate potential financial losses.

Bivariate Analysis

- Another important factor is the annual income when it comes to repayment
- Let us look at the factors that effect loan status with respect to annual income and total loan amount



Observations

1. **Overall Distribution**:
   - The majority of the data points are concentrated in the lower ranges of both Loan Amount and Annual Income.
   - There are a few outliers with significantly higher Annual Incomes, but these are mostly in the Fully Paid category.
2. **Loan Status and Loan Amount**:
   - **Charged Off** (orange dots) and **Fully Paid** (blue dots) statuses seem to be spread across all loan amounts.
   - The Current status (green dots) appears more often with lower loan amounts.
3. **Loan Status and Annual Income**:
   - Most of the Charged Off loans occur with borrowers who have relatively low annual incomes. This suggests that borrowers with lower incomes might be more likely to default.
   - Higher incomes are associated almost exclusively with Fully Paid loans, indicating that borrowers with higher incomes are better able to manage their loan payments.

Conclusions

1. **Risk Assessment Based on Income**:
   - Borrowers with lower annual incomes are more likely to default on their loans (Charged Off status). This indicates that income is a critical factor in predicting loan defaults. Loans to individuals with lower incomes may need stricter scrutiny or higher interest rates to compensate for the increased risk.
2. **Loan Amount Considerations**:
   - While loan defaults occur across various loan amounts, there doesn't seem to be a strong concentration of defaults at higher loan amounts. This suggests that the loan amount itself might not be a strong standalone predictor of default without considering the borrower's income.
3. **Targeting for Financial Products**:
   - Financial institutions might consider offering different products based on income brackets. For example, lower-income borrowers could be offered smaller loans with more flexible terms to reduce default risk.

Summarising Conclusions

- **Risk-Based Loan Pricing**:
  •**Solution**: Adjust interest rates based on a combination of loan grade, homeownership status, and income levels. For riskier borrowers (lower grades, renters, low income), higher interest rates or stricter loan terms should be applied.
  •**Benefit**: This would compensate for the higher risk of default and protect the lender from financial losses.
- **Income-Based Lending Limits**:
  •**Solution**: Set stricter loan amount limits for borrowers with lower annual incomes. Alternatively, offer smaller loans or require more collateral from these borrowers.
  •**Benefit**: Reducing the loan amount for lower-income borrowers can lower the default rate, thus minimizing credit losses

- **Targeted Financial Products**:
  - **Solution**: Develop specific loan products for different segments. For example, create lower-interest, smaller loans for low-income renters with strict repayment terms, or offer premium products to high-income borrowers.
  - **Benefit**: Tailoring loan products to borrower profiles can reduce risk and improve customer satisfaction by meeting their financial capabilities.
- **Enhanced Risk Assessment**:
  - **Solution**: Incorporate additional factors such as credit scores, debt-to-income ratios, and employment stability into the risk assessment process to better identify potential defaulters.
  - **Benefit**: A more holistic risk assessment can lead to more informed lending decisions and lower default rates.