Lending Club Case Study

Team Members:

Merine George (Group Facilitator)
Yashaswini B

Problem Statement

You work for a consumer finance company which specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company
The aim is to identify patterns which indicate if a person is likely to default, which may be used for taking actions such as denying the loan, reducing the amount of loan, lending (to risky applicants) at a higher interest rate, etc.

Use EDA to understand how consumer attributes and loan attributes influence the tendency of default.

Solving the above problem using univariate, segmented univariate and bivariate analysis.

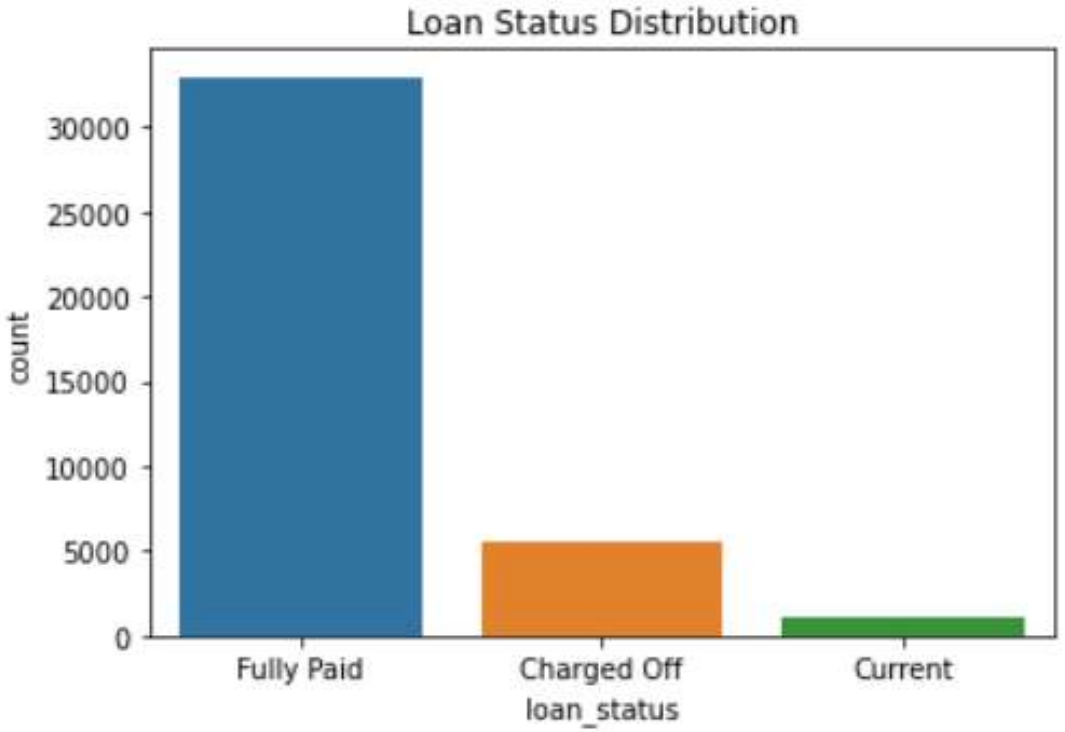
Dataset Summary

- **Number of loans started with: 39717**
- **After cleaning the number of loans analyzed: 37878**
- **Key attributes: Loan Amount, Term, Interest Rate, Employment Length, Annual Income, etc.**
- **Target variable: Loan Status (Fully Paid, Charged-Off, Current) /**

Univariate Analysis

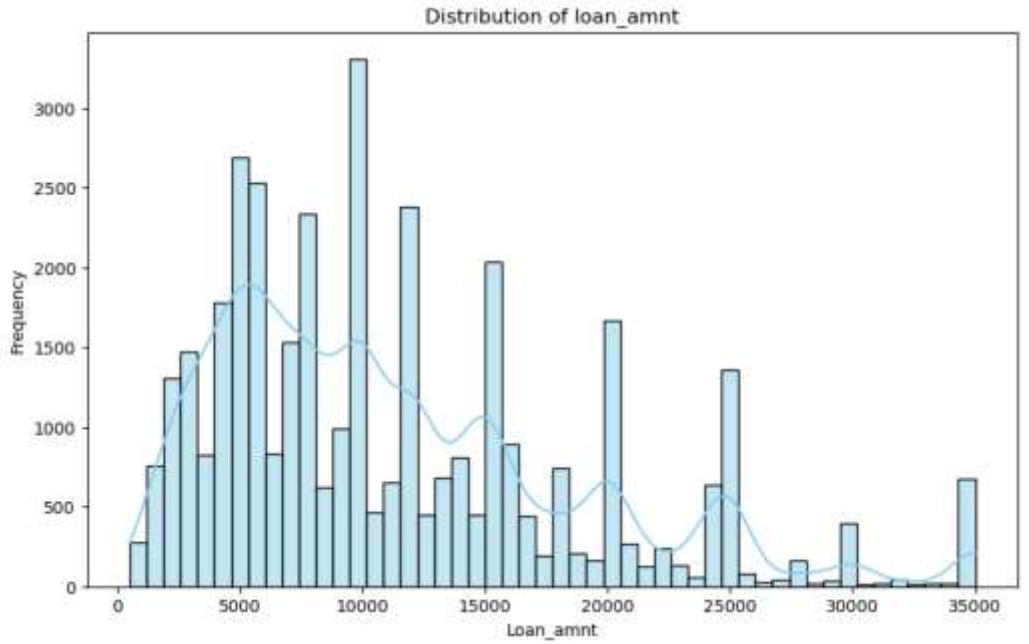
Distribution of Loan Status:

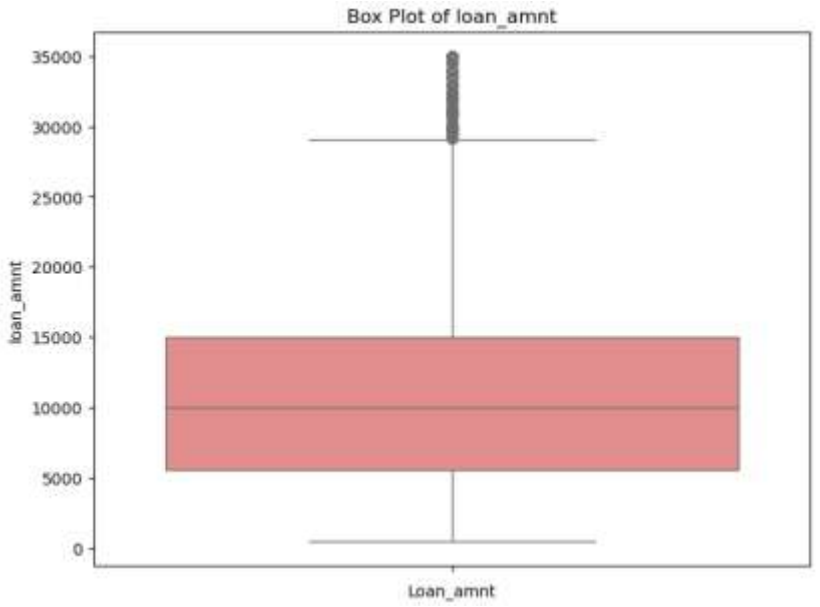
The graphs show that most borrowers have successfully repaid their debts, but a few have defaulted, giving us an understanding of the bank's potential losses. Since the number of charged-off loans is relatively low, we can infer that the bank faces minimal risk. However, it is still important to explore strategies to minimize even the small losses.



Distribution of Loan Amount :

The graphs indicate that most borrowers have taken smaller loan amounts. However, the box plot also reveals a few outliers with significantly higher loan amounts, suggesting a higher level of risk.

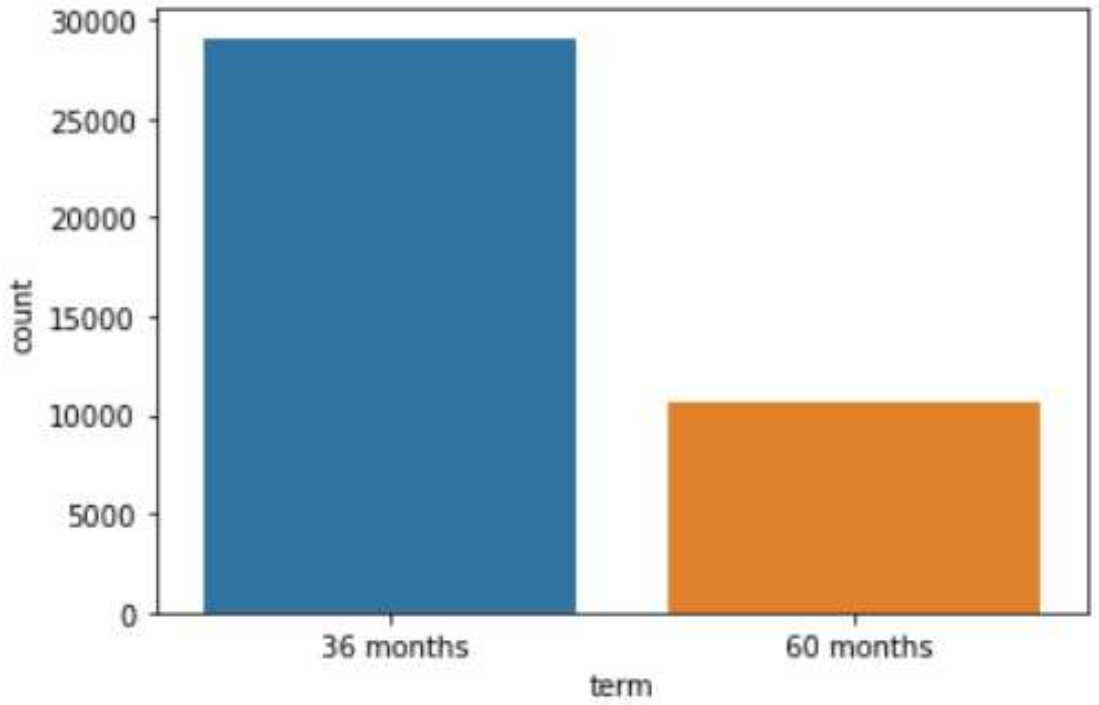




Distribution of Loan Term:

Term: Refers to the number of payments on the loan.

The graph illustrates that a larger number of borrowers are choosing loans with shorter terms compared to those opting for longer-term loans.

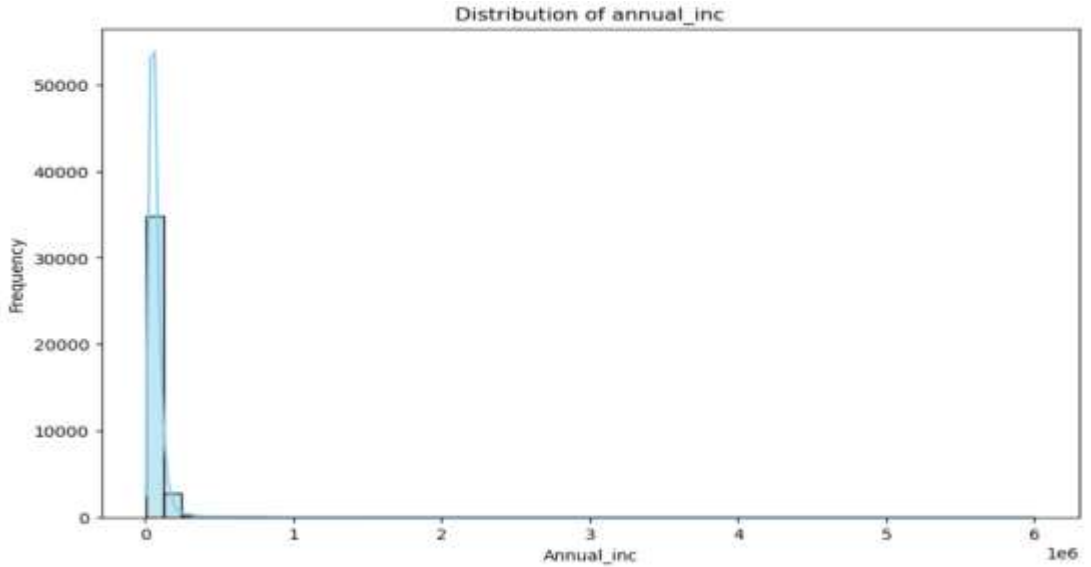


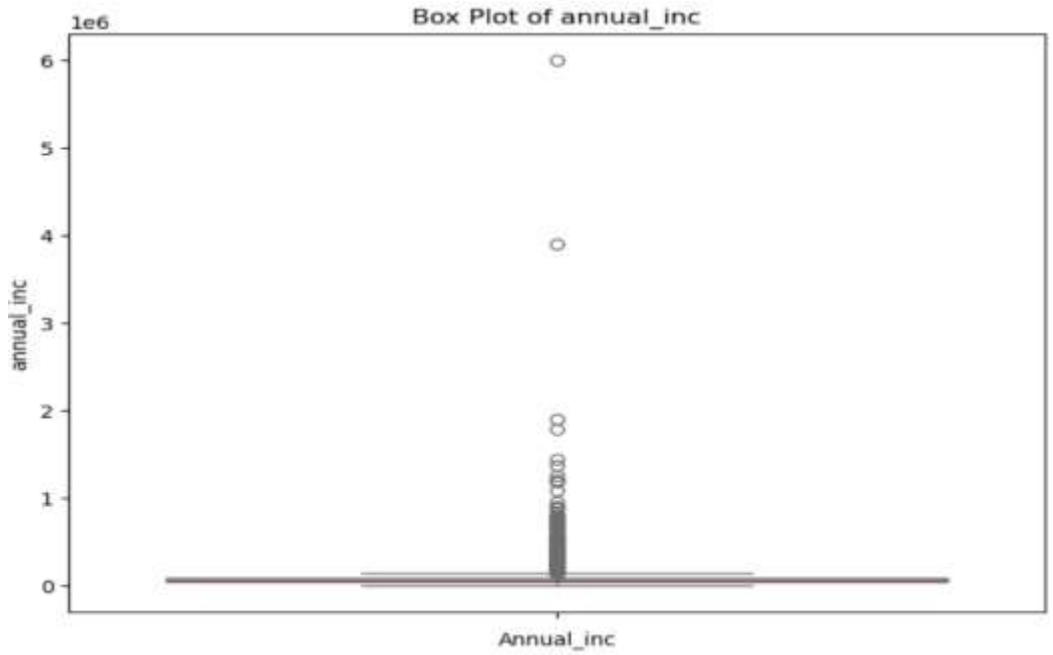
Distribution of Annual Income:

Annual Income: The self-reported annual income provided by the borrower during registration.

•The minimum value is 4000 and max value is 6 million , seems to be extreme outliers, Need to check whether it is a mistake or valid data values

•The distribution of annual income is right skewed (postive-skewed),there are smaller number of high salary individuals

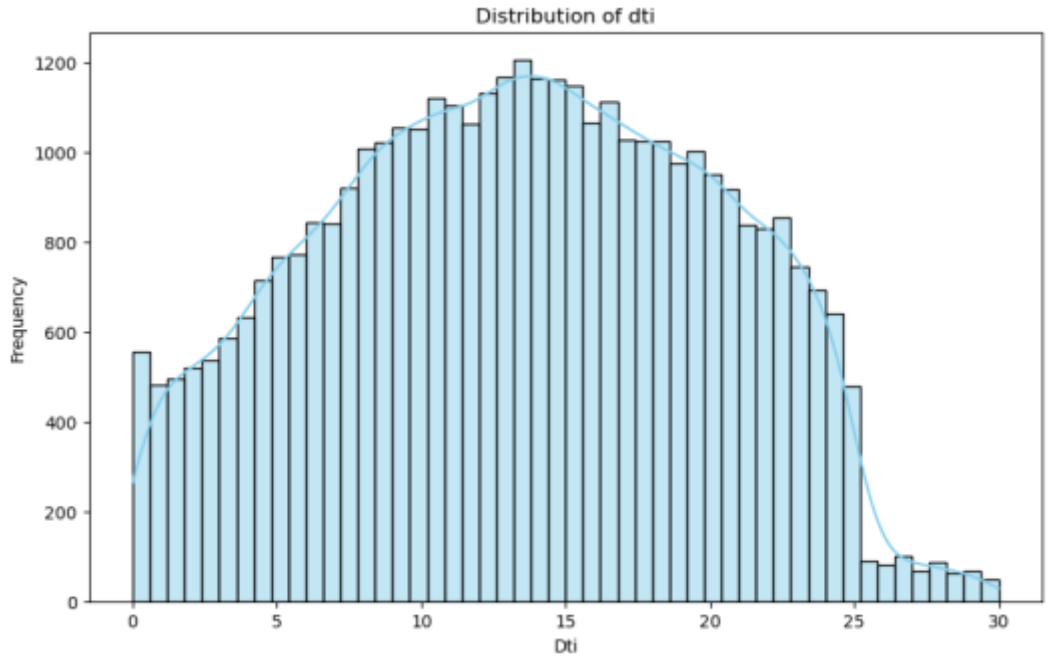


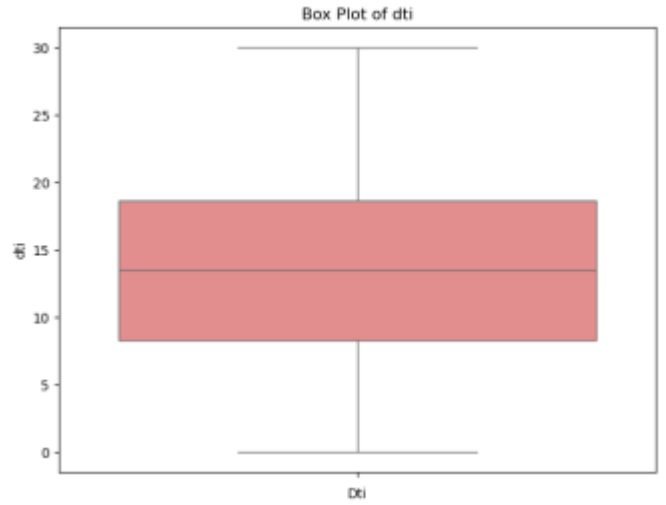


Distribution of DTI :

Debt-to-Income is used by lenders to assess an individual's ability to manage monthly debt payments.

- The majority of values are clustered around the mean (center).
- Borrowers with higher DTI ratios (closer to the maximum) may face an increased risk of default.
- Borrowers with lower DTI ratios are generally less likely to default.

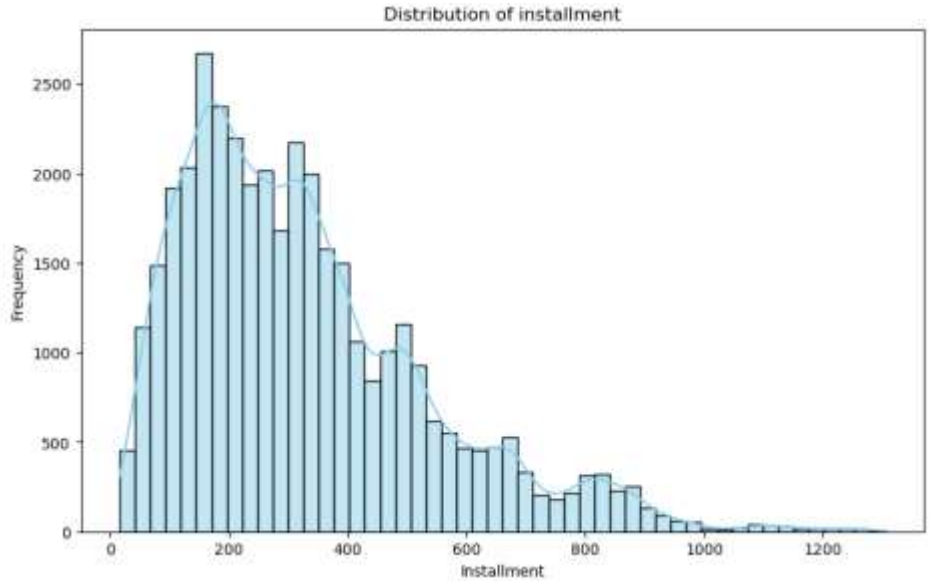


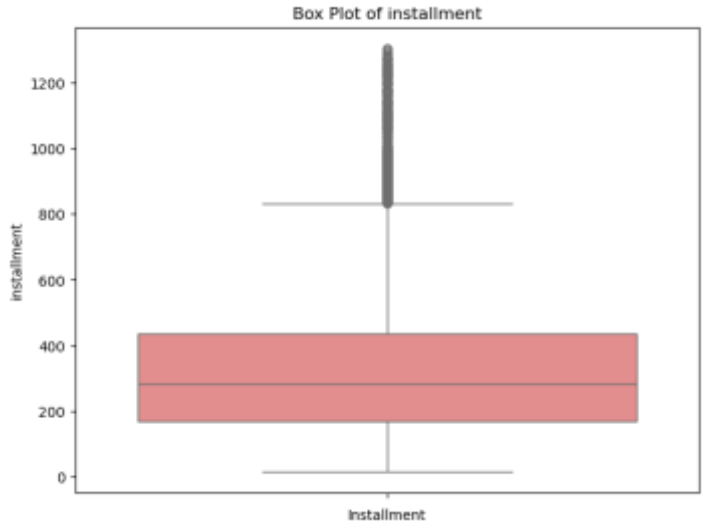


Distribution of Instalments:

Instalment: The monthly payment owed by the borrower if the loan originates

- The minimum value is 326.95 and max value is 1305.19
- The distribution of funded amount is right skewed (postive-skewed), few loans with high installments
- Majority of applicants between 25% -75% range.

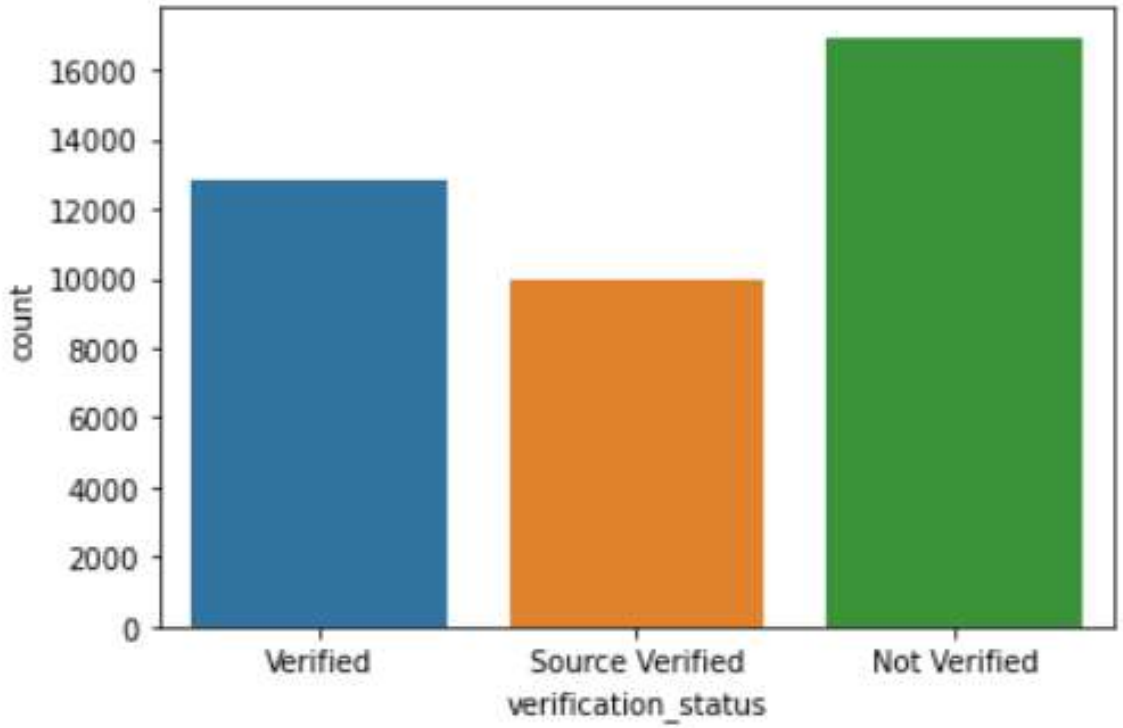




Distribution of Loan Verification Status:

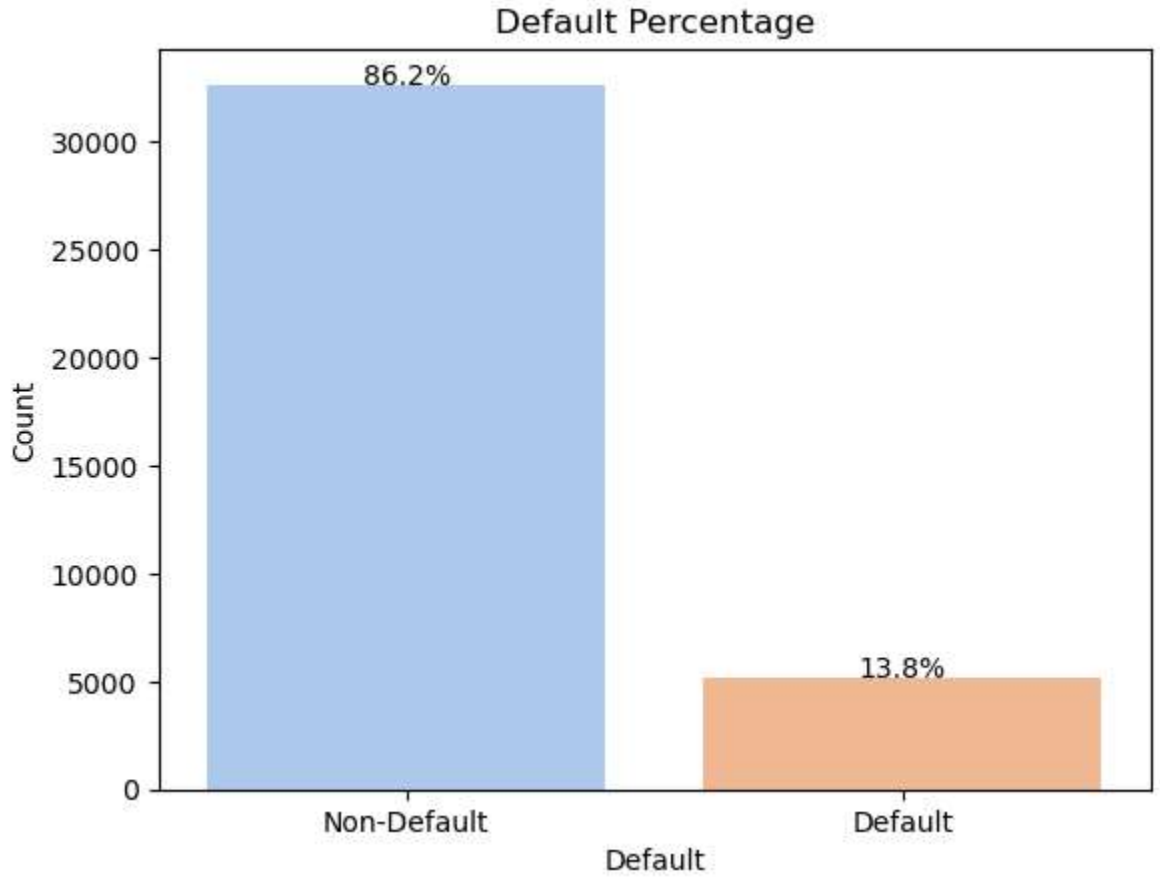
Verification Status: Indicates whether the borrower's income was verified by LC, not verified, or if the income source was confirmed.

The graph reveals that a significant portion of borrowers have not had their income verified, which raises concerns for the bank regarding potential risks.



Loan Default Rate:

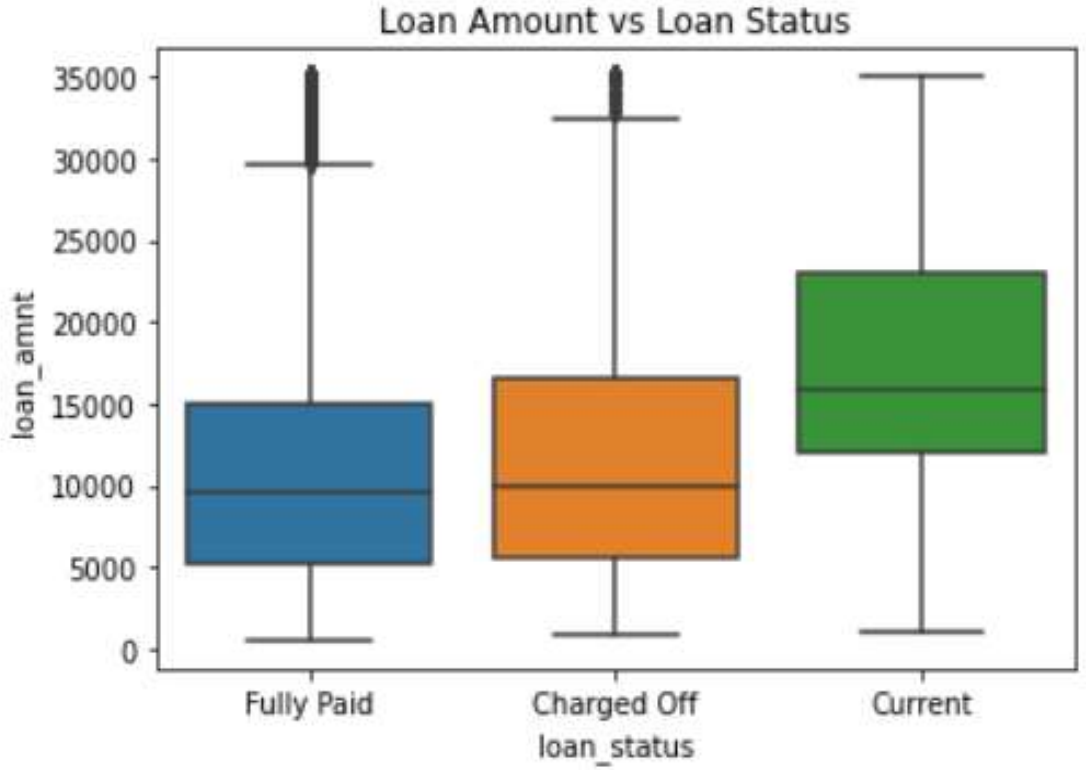
Once data is cleaned and Upon calculating the default rate, it is observed that approximately 13.8% of borrowers have been marked as "charged-off," indicating they have defaulted on their loans.



Bivariate Analysis

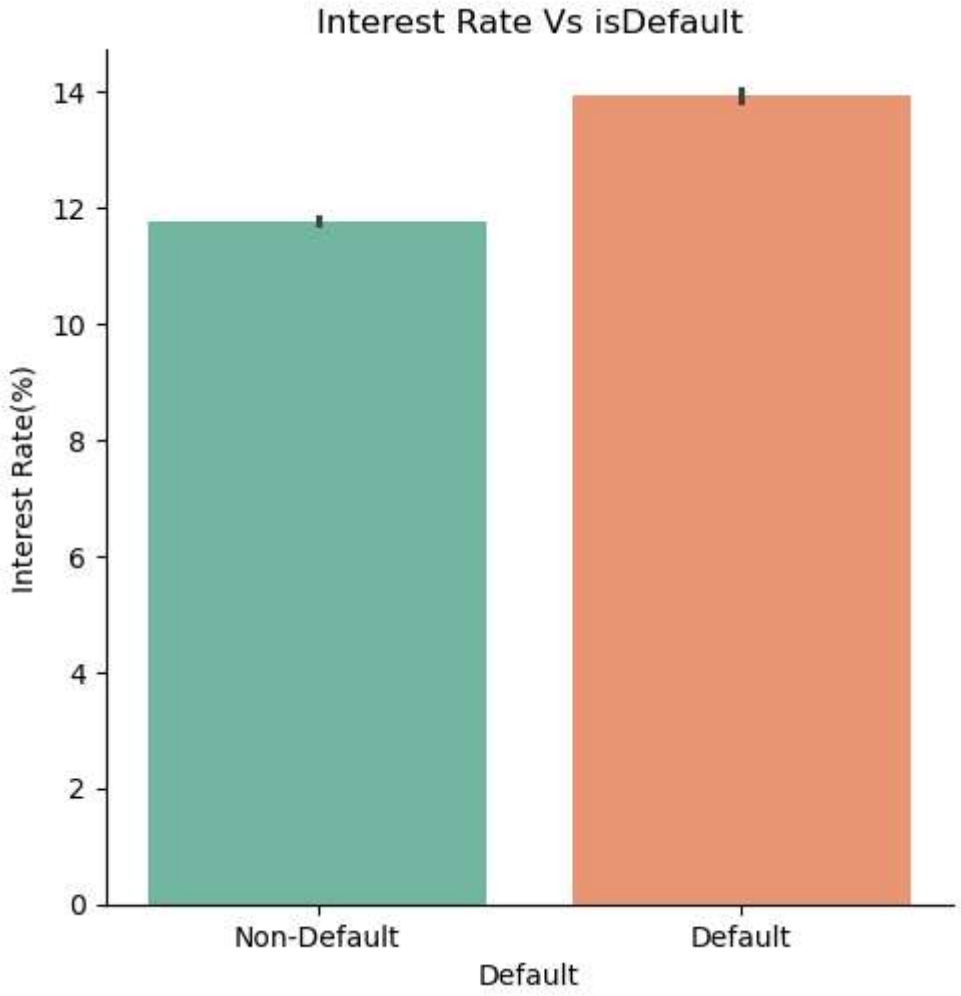
Loan Status vs Loan Amount:

- Higher loan amount is taken by people in the current status
- The range of loan amount for the defaulters is highest with some outliers taking high loan amount
- Maximum outliers are present for fully paid status



Interest Rate vs Default:

This graph clearly demonstrates a correlation between interest rates and the likelihood of borrower default. As the interest rate increases, there is a noticeable rise in the probability of borrowers defaulting on their loans. Higher interest rates lead to higher monthly payments, which may strain borrowers' ability to repay. This trend indicates that borrowers facing higher interest rates are more likely to experience financial difficulties, making them more susceptible to default. The graph underscores the importance of managing interest rates to mitigate the risk of loan defaults.

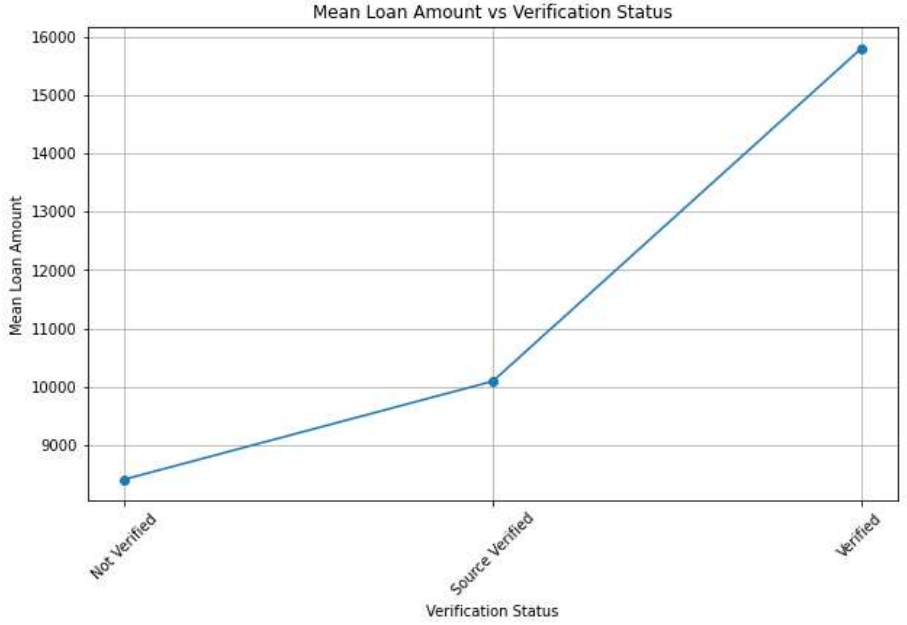


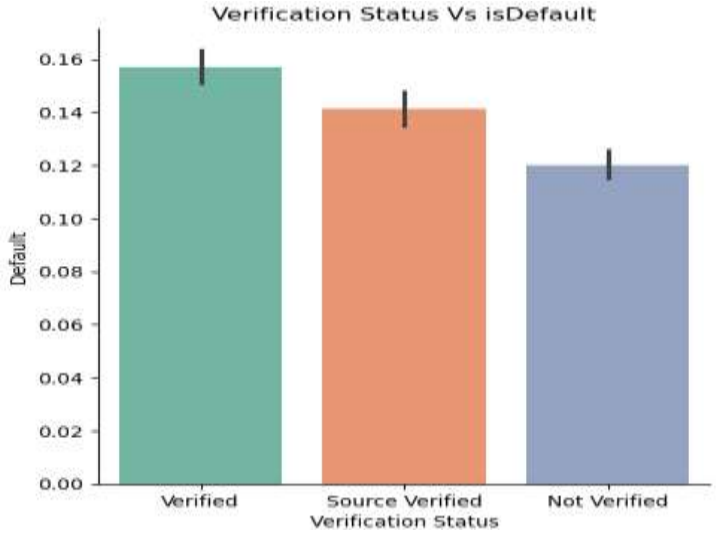
Impact of Verification Status on Borrower Default Risk

Verification Status and Borrower Default: The data from these graphs suggest that verification status does not have a strong influence on whether a borrower will default on their loan.

Graph 1: This graph shows that a large proportion of borrowers have had their income verified, indicating that the verification process is widely used by the bank to assess borrowers' financial stability.

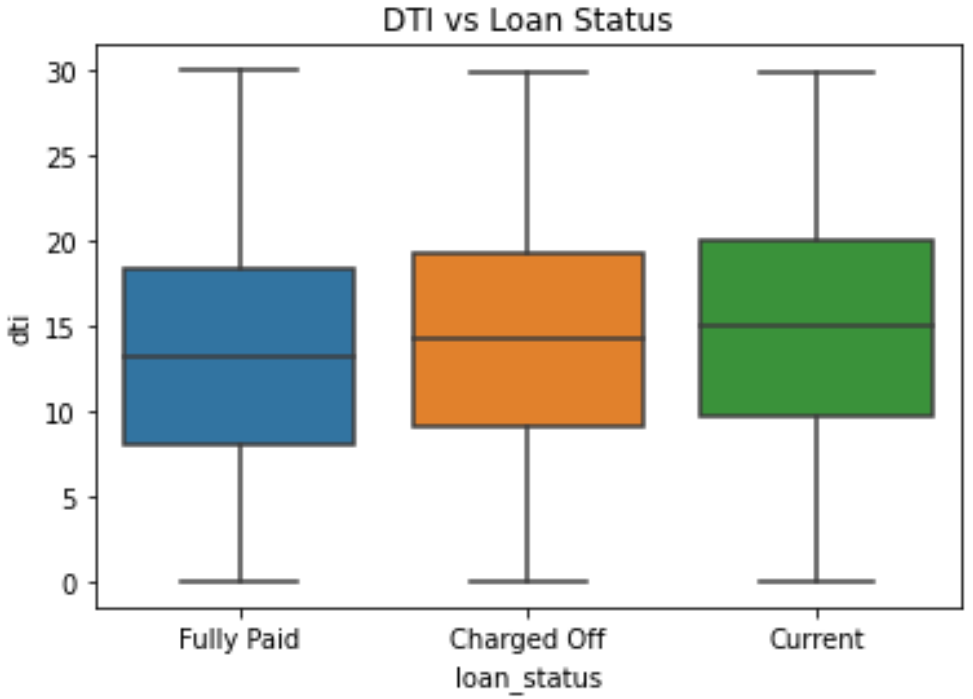
Graph 2: However, the second graph presents a concerning trend: despite the high number of verified borrowers, a significant portion of the loan defaulters also falls within the verified group. This suggests that even income verification may not be a sufficient indicator to predict the likelihood of default.





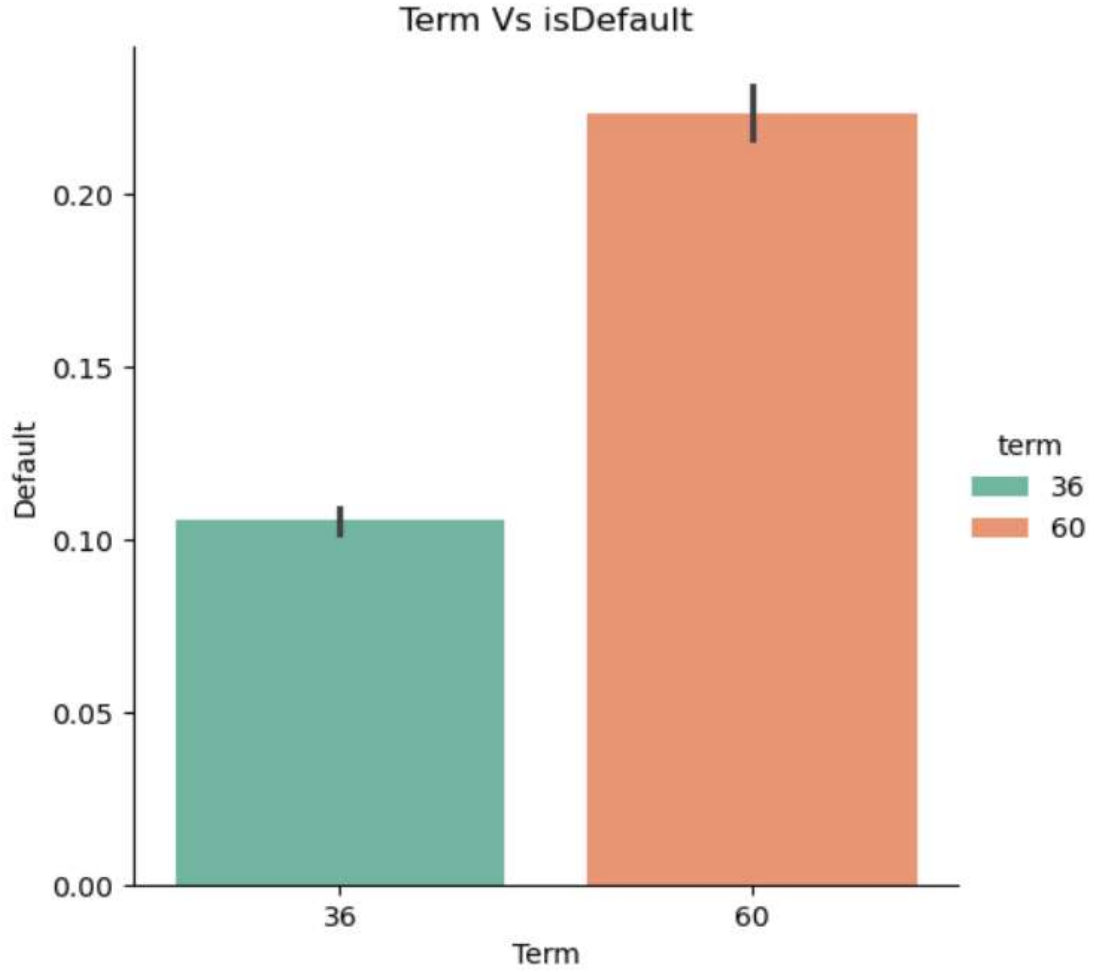
DTI vs Loan Status:

This box plot illustrates that the majority of **Debt-to-Income** (DTI) ratios are concentrated around the middle value range, regardless of the loan status. This suggests that a typical borrower tends to have a moderate DTI, irrespective of whether their loan is performing or has defaulted. The plot highlights the general trend of borrowers having balanced income-to-debt ratios, with outliers being less common.



Loan Default Rate vs Term:

•36 months have lower default rate and 60 months terms have higher default rate
•This suggests lenders might need stricter approval criteria for 60 months loans.

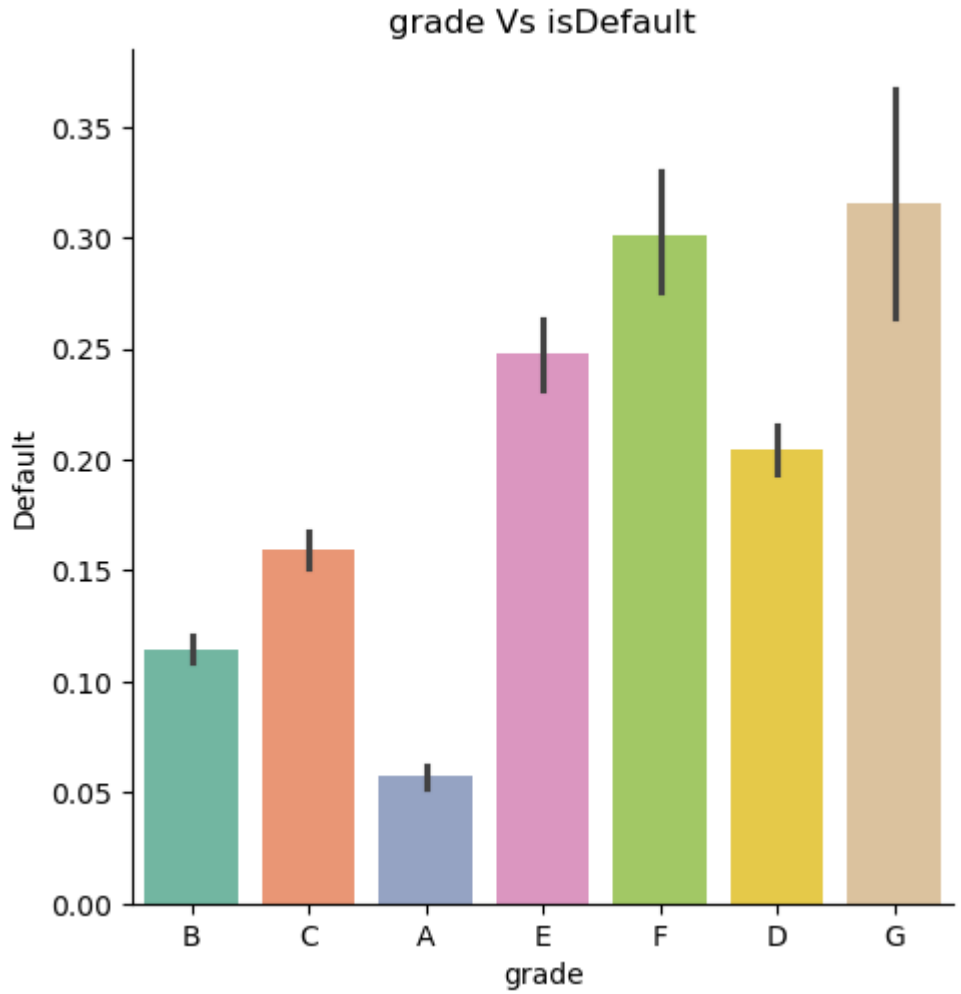


Loan Default Rate vs Grade:

- A - 5.7%
- B - 11.4%
- C - 15.9%
- D - 20.5%
- E - 24.75%
- F - 30.10%
- G - 31.51%

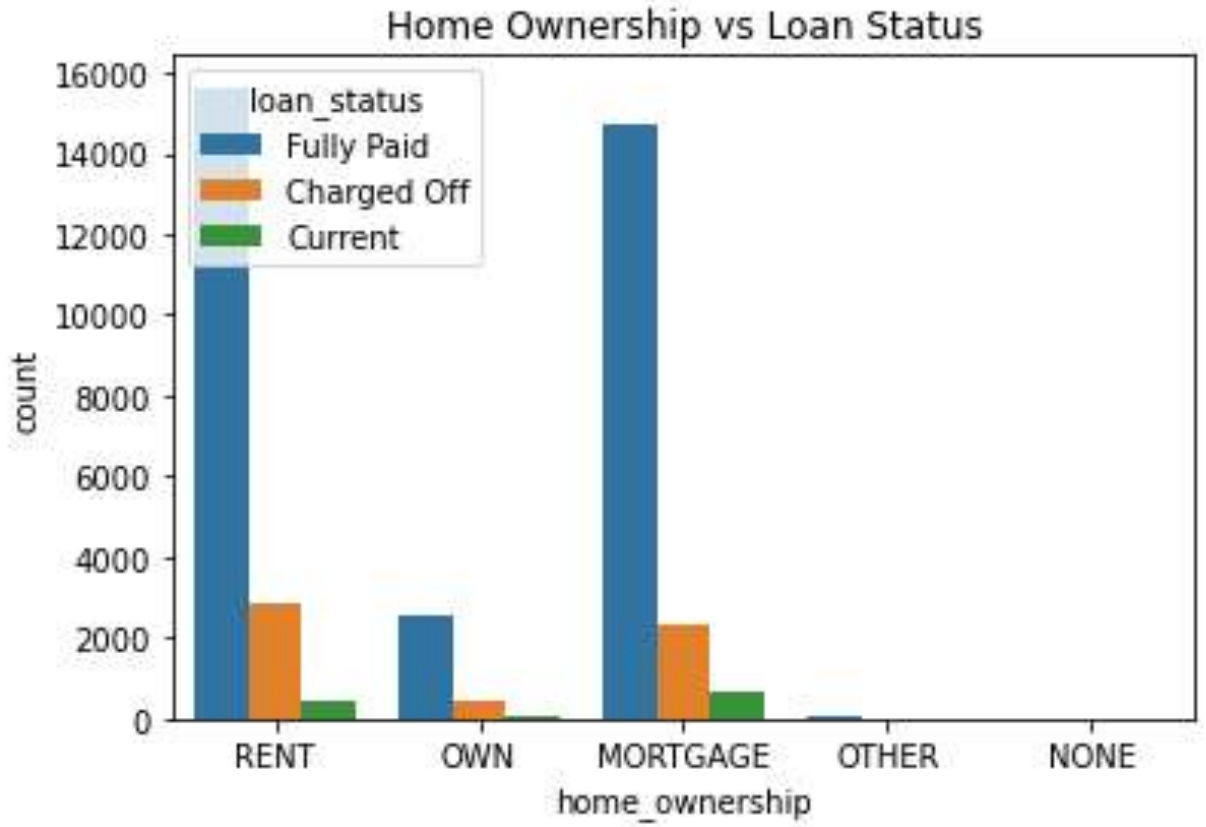
Upon analyzing the graph, it is evident that as Grades move from A to G, the default rate is increasing.

- Grade G has the highest default rate & high risk, while Grade A is low risk
- Institutions can increase interest rates for Grade D-G borrowers to compensate for the higher risk.



Home Ownership vs Loan Status:

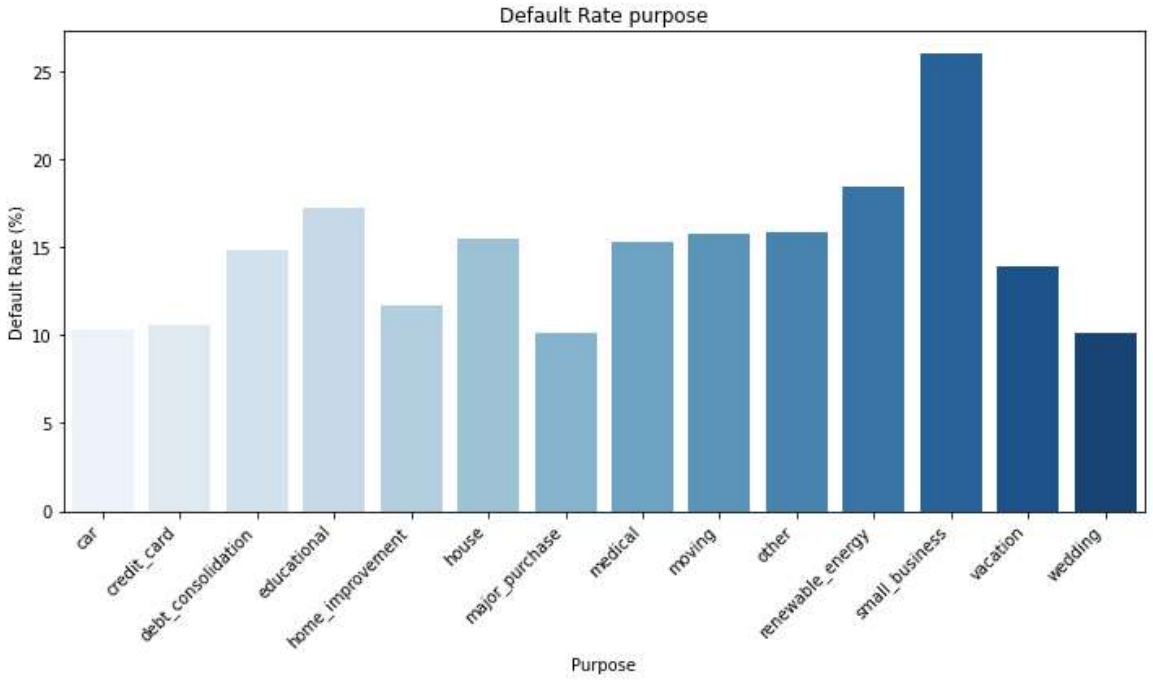
This graph reveals that, regardless of home ownership status, the majority of borrowers have successfully paid off their loans. However, the data also shows that the lowest number of defaulters is among homeowners ("Own"), while renters ("Rent") account for the highest proportion of defaults. Additionally, it is evident that the number of loans taken by homeowners is relatively small, which may suggest that many individuals take out loans specifically for purchasing homes, rather than for other purposes. This trend highlights the potential connection between homeownership and financial stability.



Loan Default Rate vs Purpose:

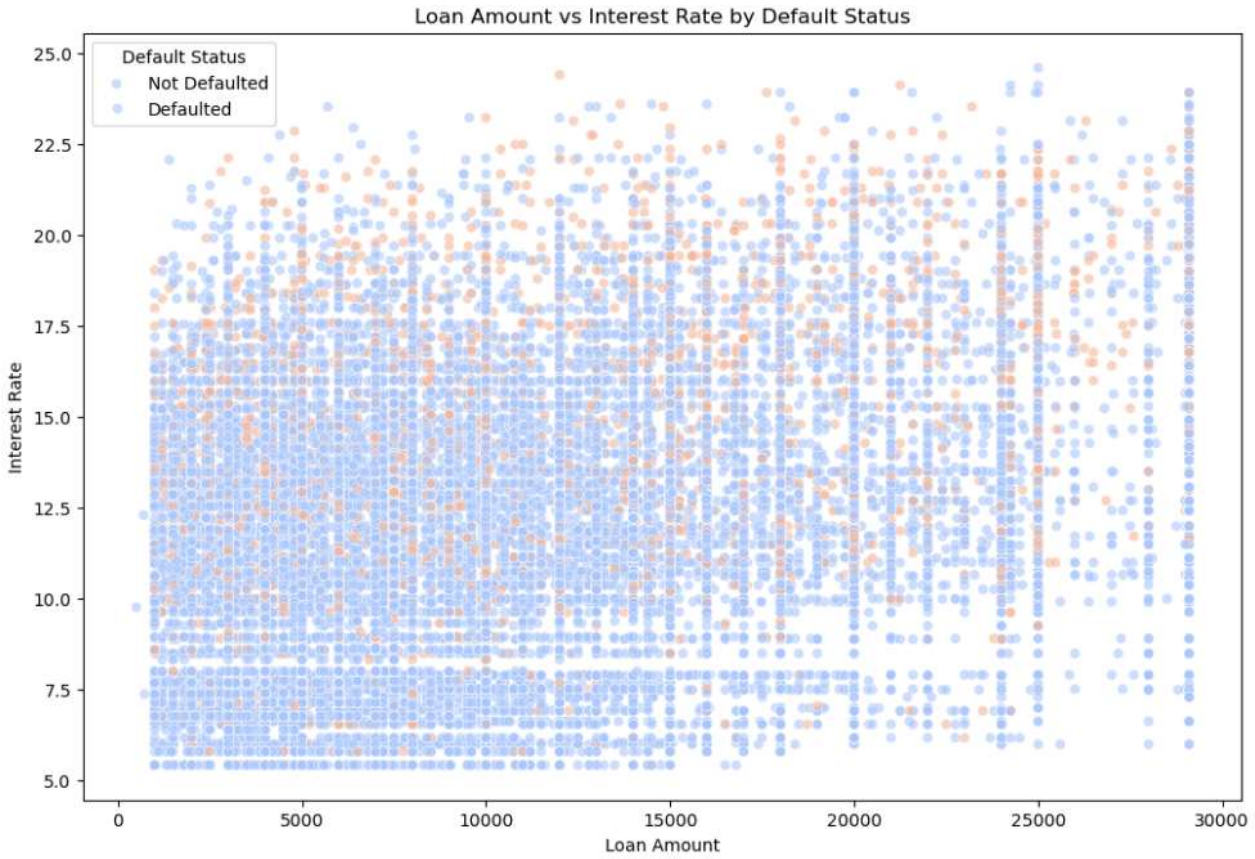
- Small business: 25.3%
- Renewable energy: 18%
- Educational: 15.824916

Upon analyzing the graph, it is evident that the highest number of defaulters are borrowers who took out loans for small business purposes. This group accounts for the largest share of defaults, followed by those who borrowed for renewable energy projects and, in third place, for educational expenses. This trend suggests that borrowers seeking loans for small businesses are more likely to default, possibly due to the inherent risks and financial challenges associated with starting or running a business. The data also indicates that, while defaults are notable in renewable energy and education, the default rate is comparatively lower in these areas.



Loan Amount vs Default by Status:

- Higher interest rate tend to default more from the graph
- Combination of high loan amount and high interest rate increase the default rate
- Too much clustered blue dots around the lower/moderate interest rates and lower loan amount. This shows they are more tend to pay on time and considered low risk
- There are some extreme outlier which are defaulted, they are very high risk. They have to be checked more

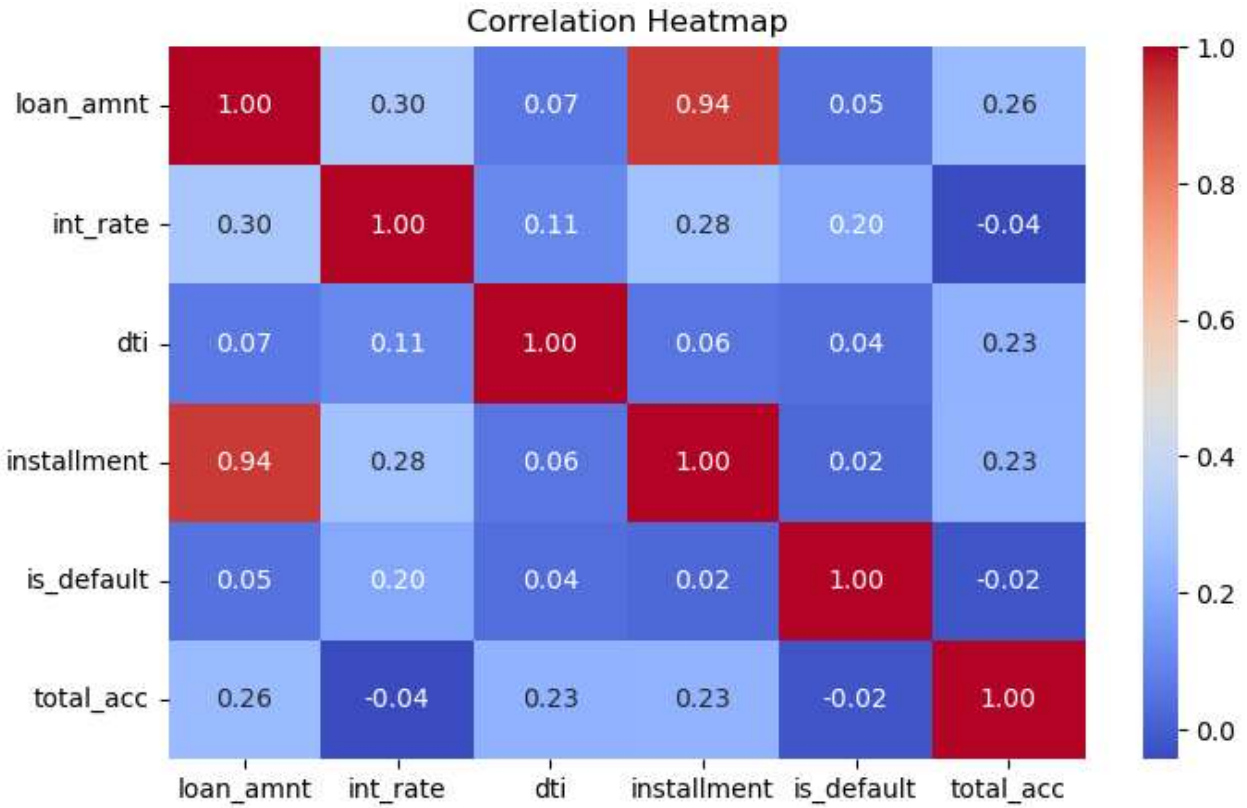


Insights on Heatmap:

- Loan Amount (loan_amnt) and Interest Rate (int_rate): Positive correlation. The Large loan amount can be associated to high interest rate. Possible chance to default.

- Loan Amount (loan_amnt) and Monthly Installment (installment): Strong Positive correlation as 0.94 is the value. When loan amount increase, installment also increases. Higher chance to default.

- The rest of the correlations are moderate.



Top Key Indicators for Default

As per the analyses we found the following key factors to be considered by the bank before approving the loan to avoid credit loss:

- Longer Loan Terms
- Debt-To-Income Ratio
- Grade (Grade between B to G is more likely to default)
- Higher Interest Rates
- Higher Loan Amount
- High Instalments
- Loan Purpose such as Small business, Renewable energy

Recommendations

After analysing we think the consumer finance company should do the below to reduce the defaulters:

- Start by tightening more on the criteria for approving the loan
- Additional Documentation for high risk loans