Lending Club Case Study and Report

Presented by: Susheel Kumar Singh
Batch:ML 57

Problem Statement

The objective is to identify trends within the historical loan data set that indicate the probability of a borrower defaulting. This information could be leveraged to make informed decisions, including loan rejection, reduced loan amounts, or lending to high-risk applicants at a premium interest rate.

Approach

- Importing, reading the file and doing basics(head(), info(),etc.)
- Data cleaning:missing data and dropping the data:
- Finding useful data
- Univariant analysis
- Bivariant analysis

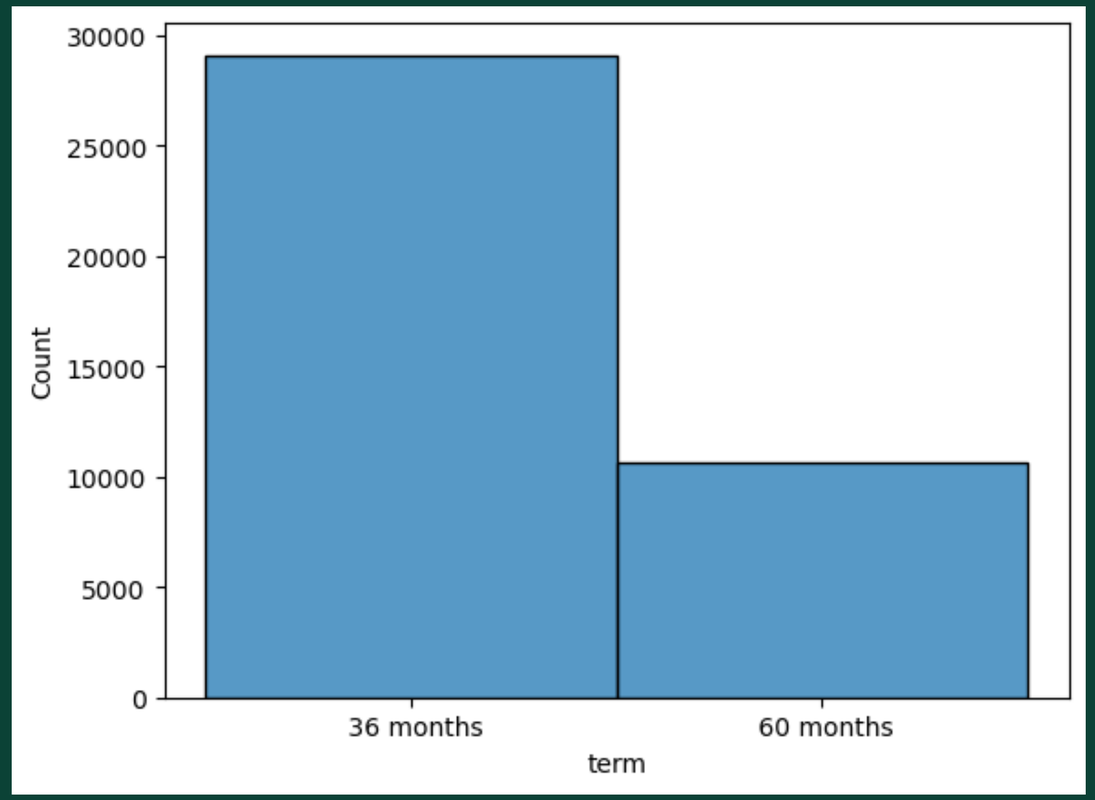
Data Cleaning

- To check for null values, use this code:100*lending.isnull().mean()
- Establish a threshold and remove any unnecessary columns.

Inputting Data

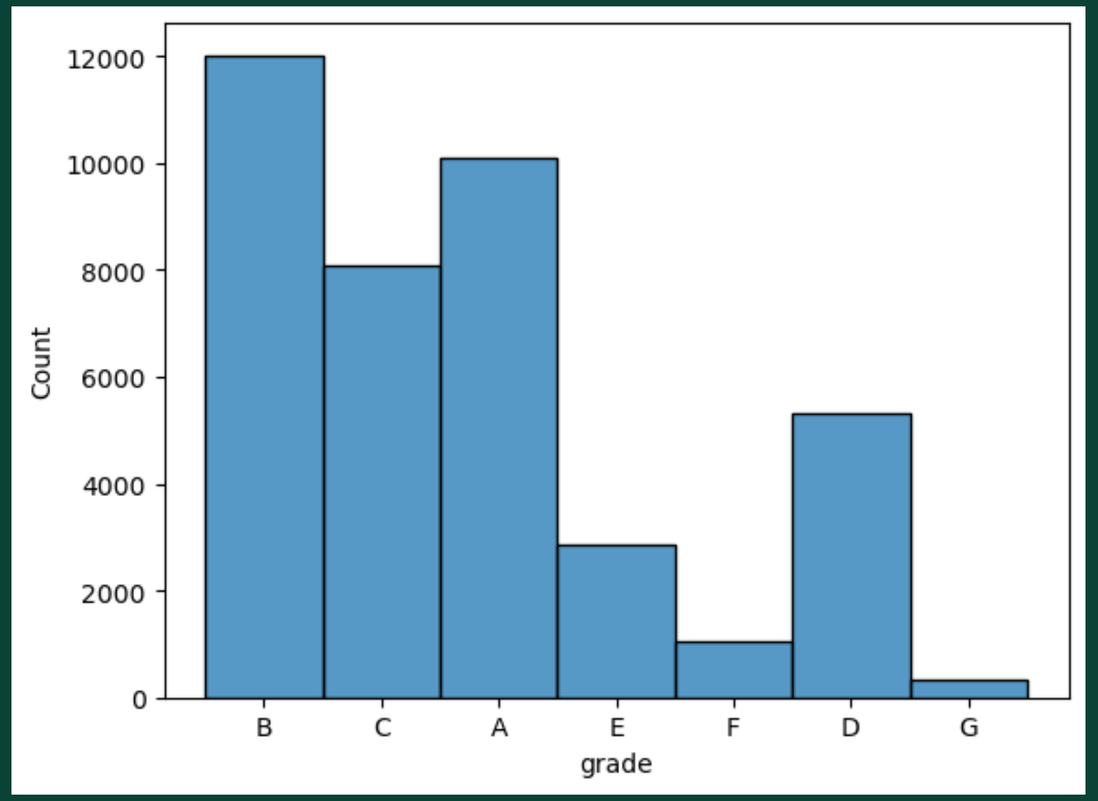
- Computing the Median for a Numeric Value
- Replace NaN values with the calculated median.

Univariant For Categorical Data



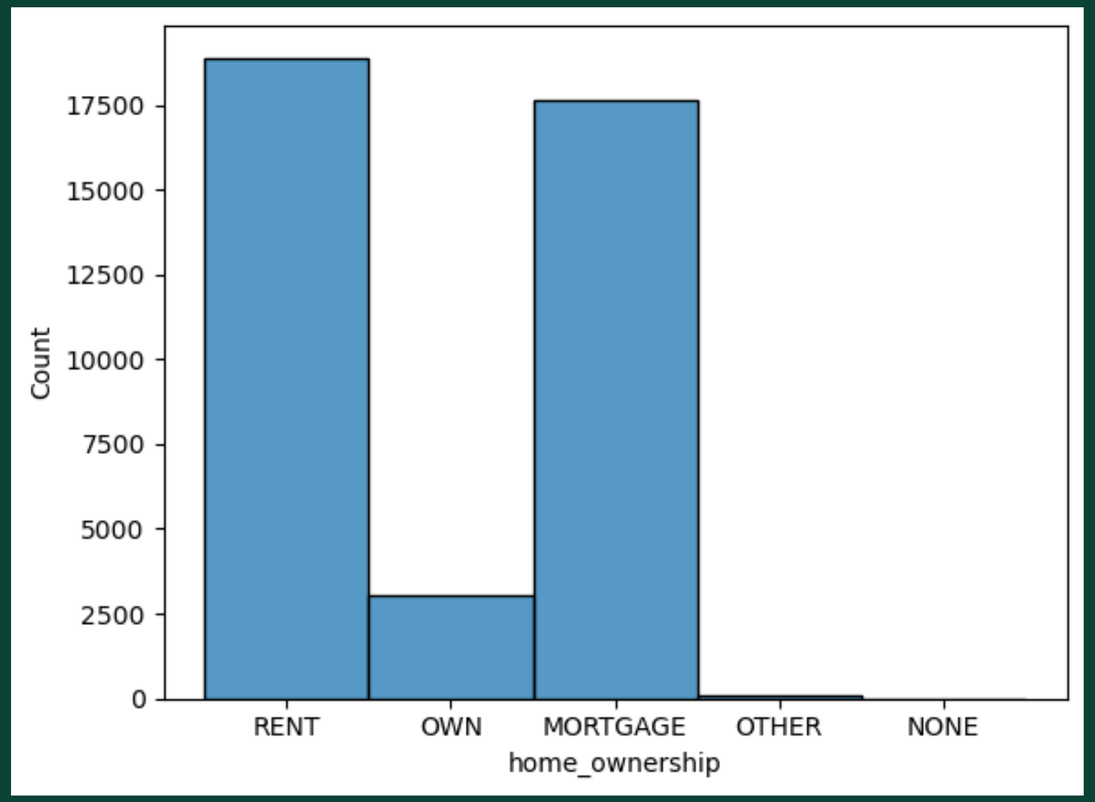
The majority of individuals opted for a 36-month loan, while a minority cho
a 60-month loan.

Univariant For Categorical Data



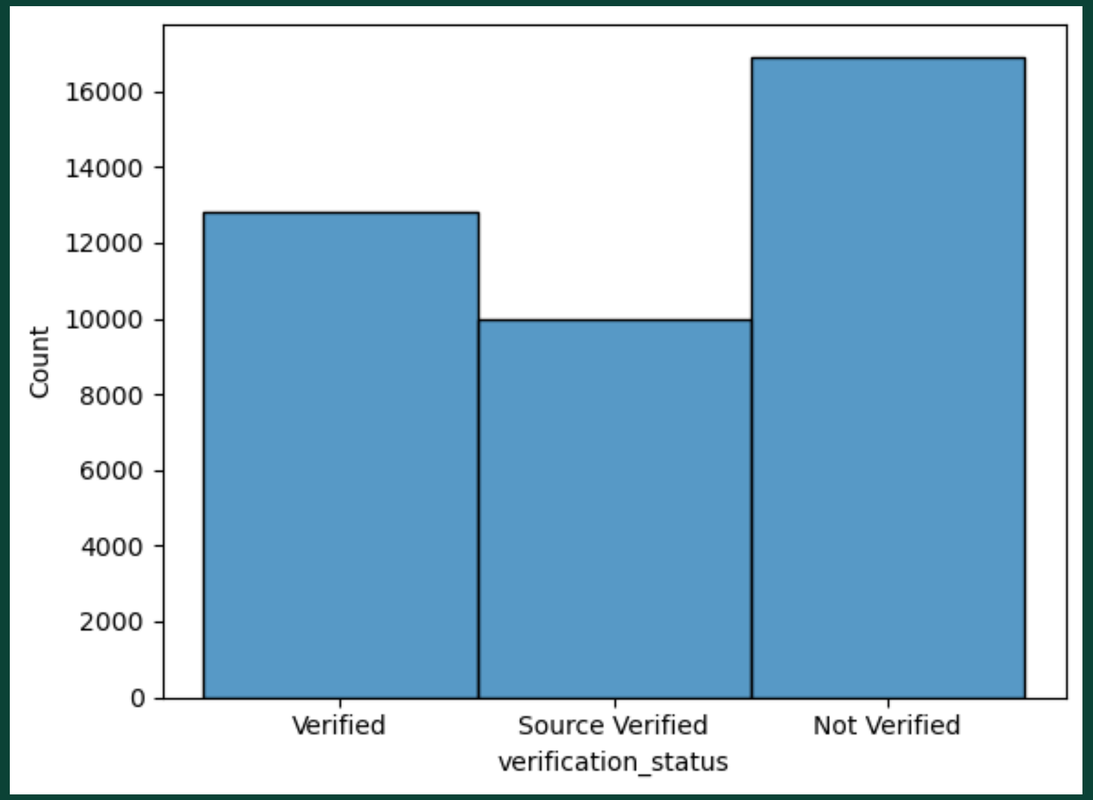
When it comes to loan recipients, the majority are from grade B, followed by grade A, and then grade C.

Univariant For Categorical Data



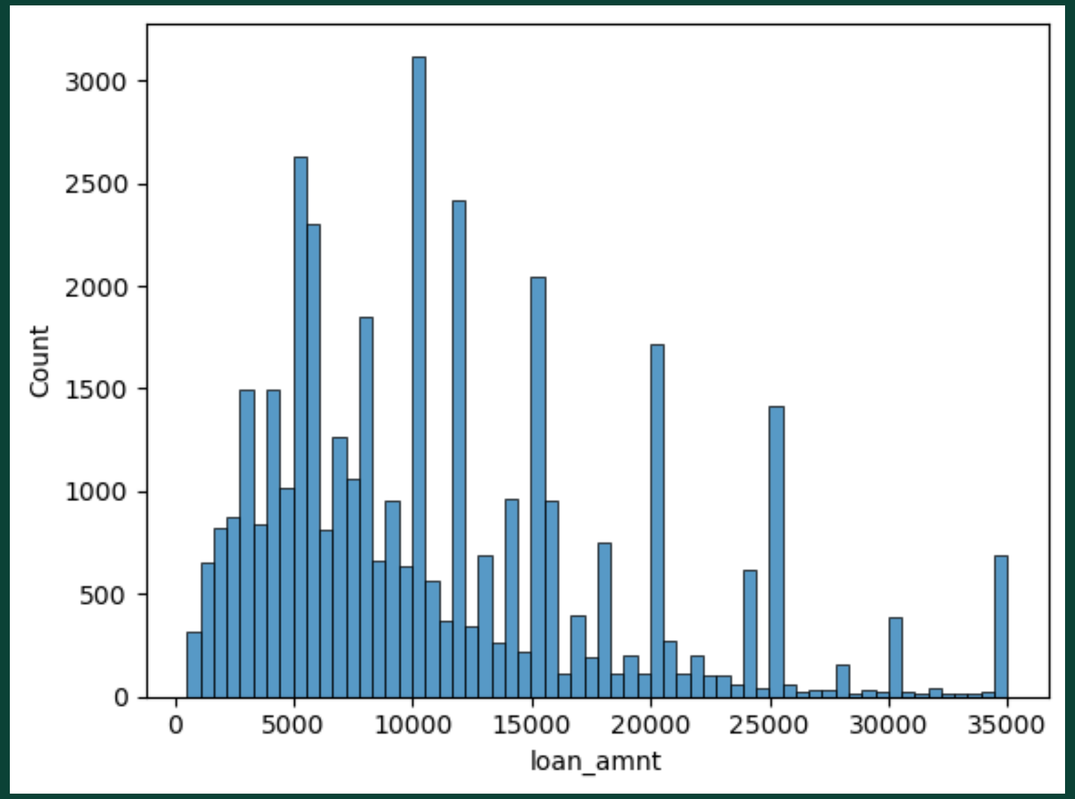
The majority of those who borrowed were residing in rental properties, followed by those with mortgages, while only a small percentage owned their own homes.

Univariant For Categorical Data



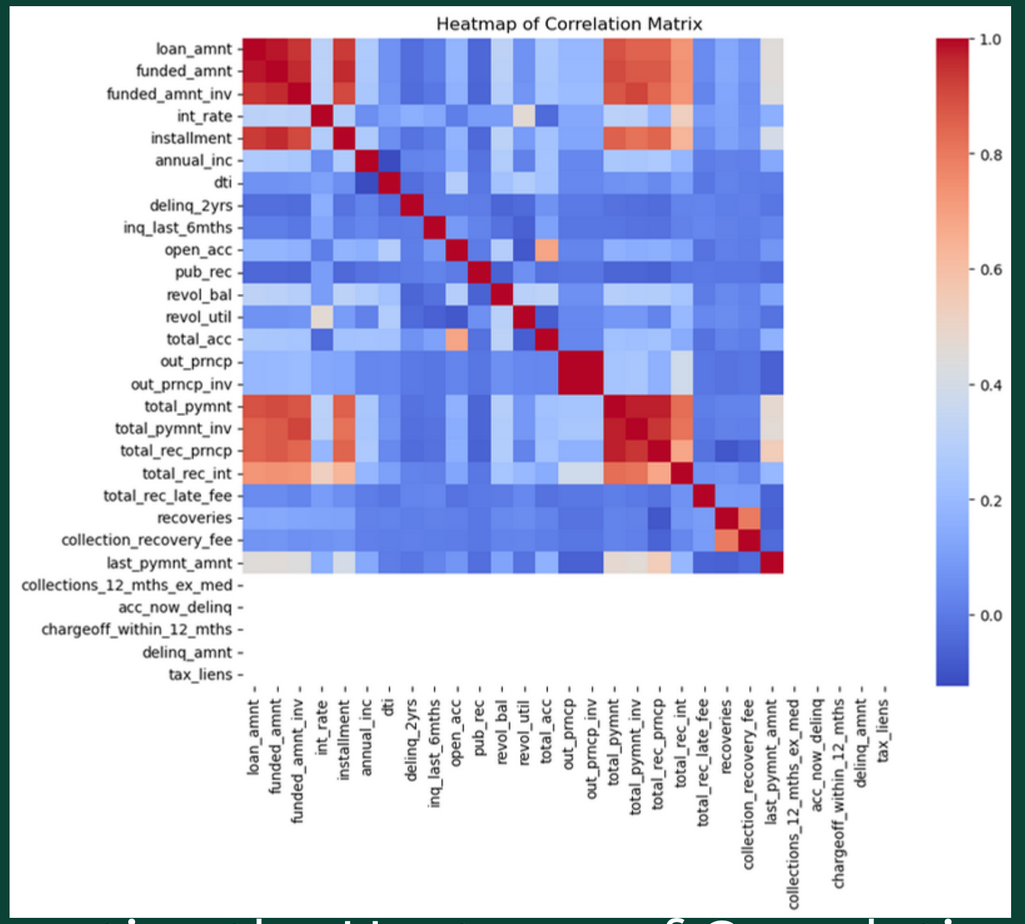
- The least amount of persons were confirmed, and for many, source verification came after verification.

Univariant For Numerical Data



Distribution of Loan Amounts:
Based on the data, the most frequent loan amounts tend to hover around 10,000.

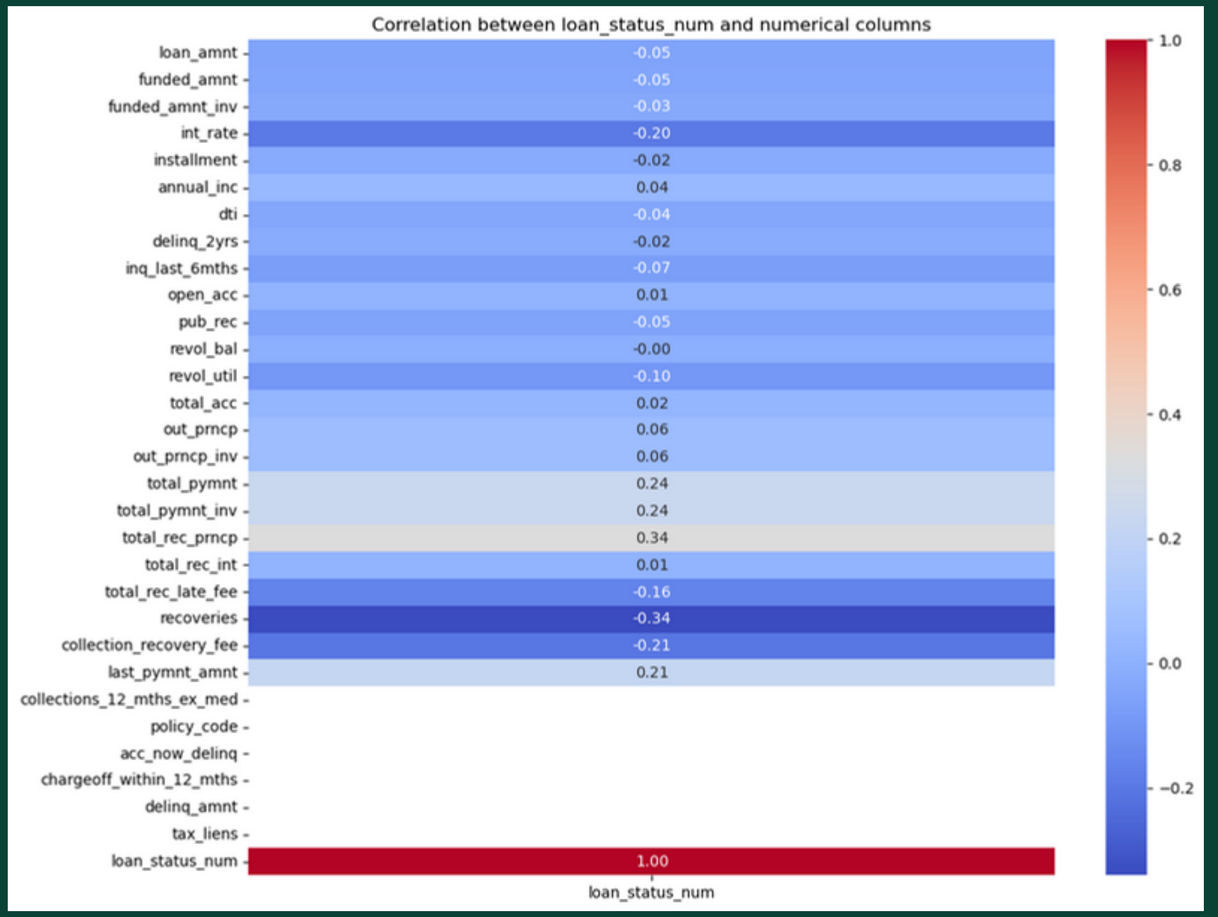
Bivariant For Numerical Data



Interpreting the Heatmap of Correlation Coefficients

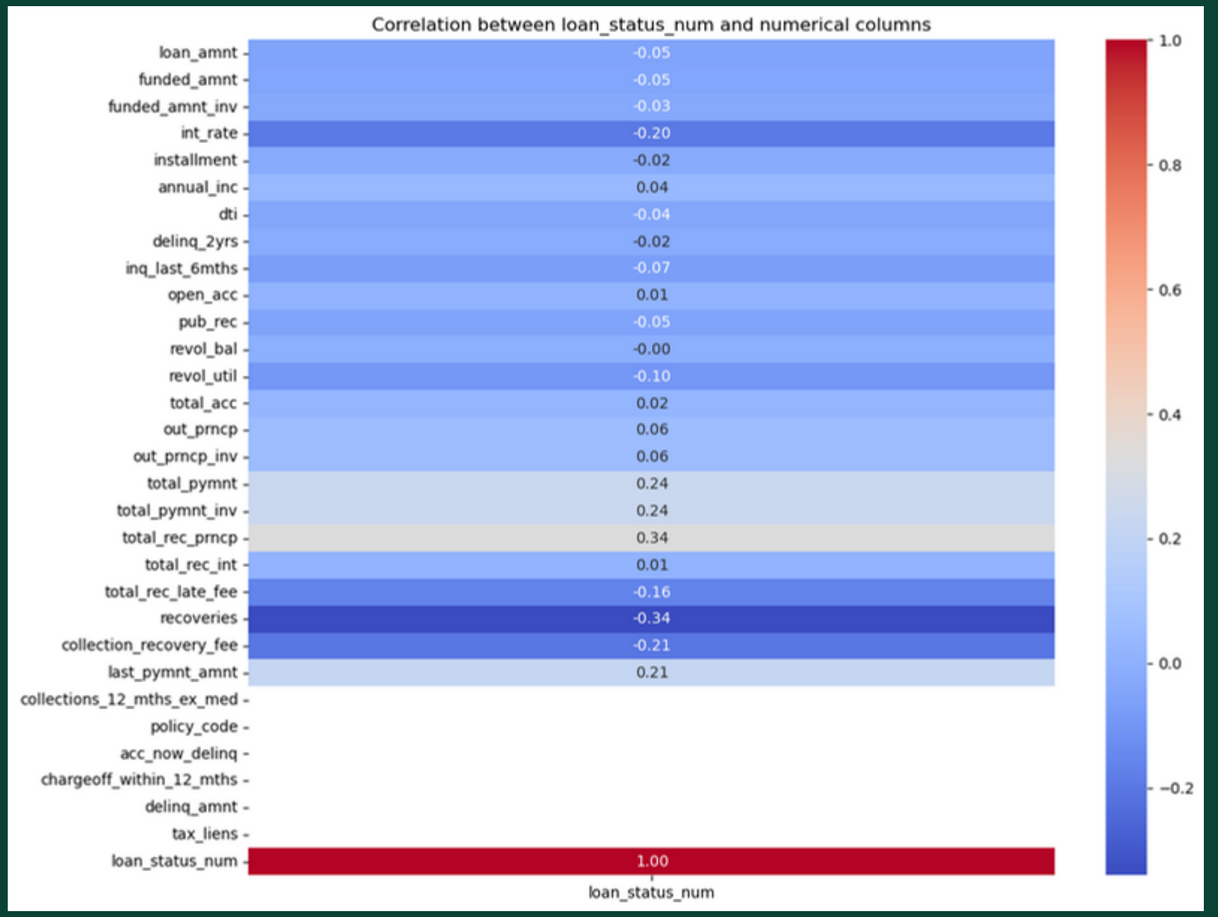
The heatmap in the dataset illustrates the correlation coefficients between various numerical variables. The darker shades in the heatmap indicate a stronger correlation between the variables

Loan Status Correlation



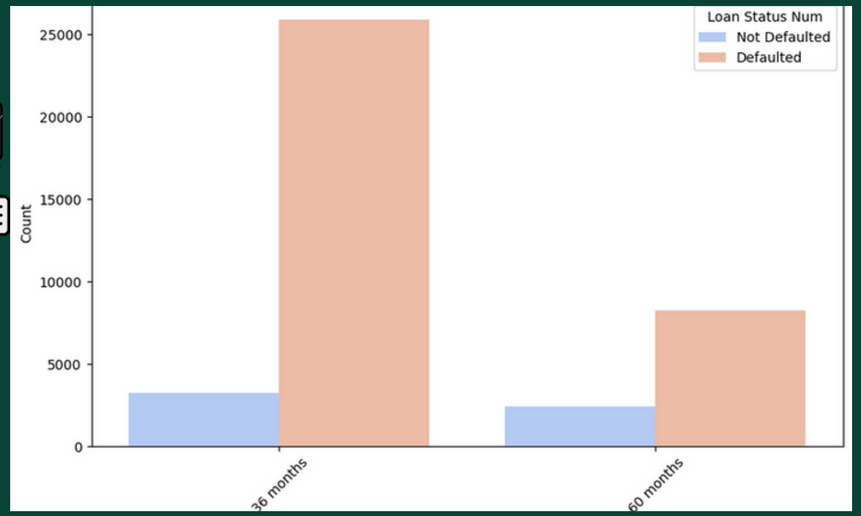
The correlation coefficient ranges from -1 to 1. If the value is close to 1, it means that there is a strong positive correlation between the two variables. When it is close to -1, the variables have a strong negative correlation

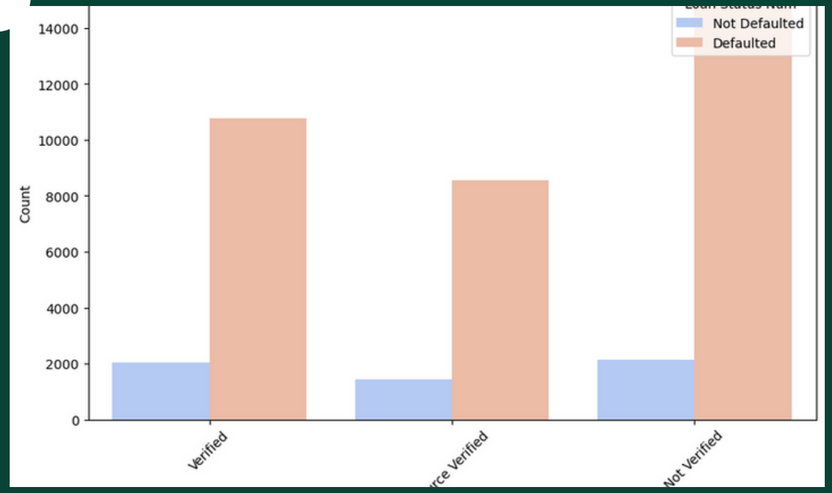
Loan Status Correlation

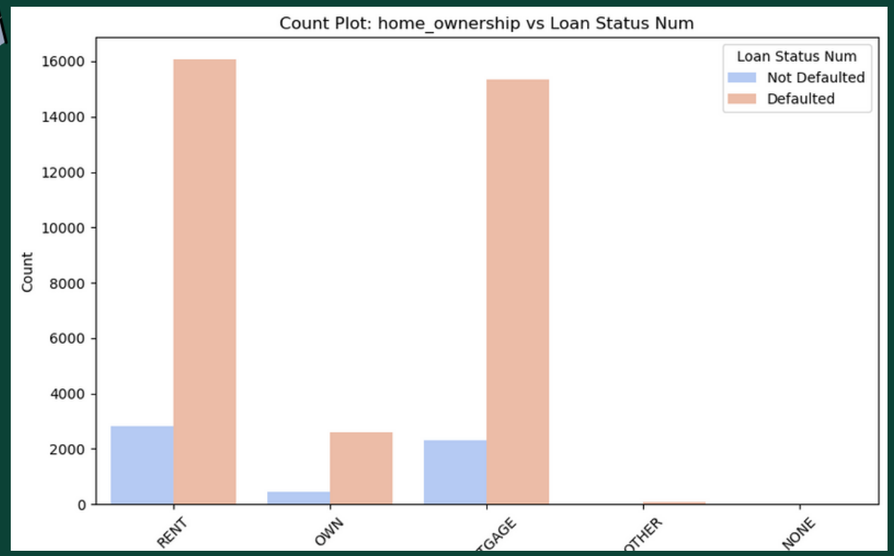


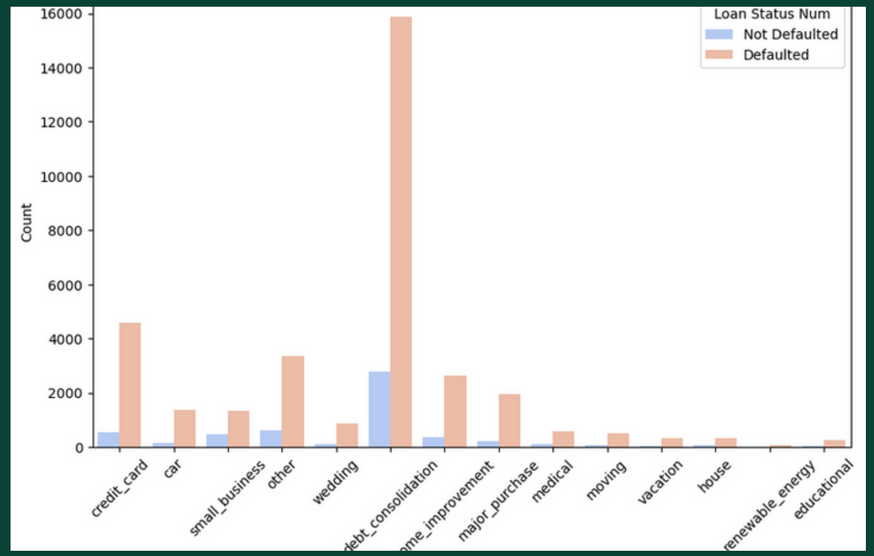
The correlation coefficient ranges from -1 to 1. If the value is close to 1, it means that there is a strong positive correlation between the two variables. When it is close to -1, the variables have a strong negative correlation

Loan Status Vs Categorical Data

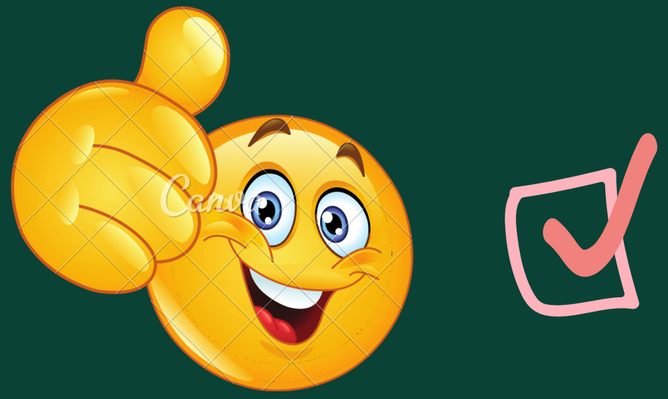








Conclusion:Understanding Loan Default Predictors



To sum up, the strong correlations between loan success and borrower behavior – such as repaying a significant portion of the principal, making larger total payments, paying higher sums to investors, and completing payments with a large final amount – indicate that these individuals possess financial stability and responsibility. From a lending perspective, these borrowers present lower risk.

Interpreting Negative Correlations in Loan Default Data
In your dataset, negative correlations have been observed between loan default rates and several factors, including lower recoveries, higher interest rates, larger loan amounts, higher credit utilization, recent credit inquiries, public records, and higher debt-to-income ratios. These factors are considered high-risk from a lending perspective, meaning borrowers exhibiting these characteristics are more likely to default on their loans.

Conclusion:Understanding Loan Default Predictors

- Loans with longer terms and lower grades (E, F, G) have a higher proportion of being charged off or defaulted.
- Borrowers who rent or own a home also have a higher proportion of charged off or defaulted loans.
- Additionally, loans where the income is not verified have a higher proportion of being charged off or defaulted.