LENDING CLUB CASE STUDY

-By Lekhana BC and Jayant

Business Understanding

->The company specializes in lending various types of loans to urban customers.

->When evaluating a loan application, two primary risks arise:-

Loss of Business: If a creditworthy applicant is rejected.

Financial Loss: If a risky applicant who defaults is approved.

->The objective is to use historical loan data to identify patterns and key factors influencing loan defaults (Charged-off). These insights will help the company:

->Minimize credit losses.

->Optimize loan approval decisions (e.g., deny loans, adjust loan amounts, or increase interest rates for risky applicants)

Data Understanding:

Dataset Overview:

- Data contains loan applicant details (demographics, financial status, loan attributes, etc.) and loan performance metrics.
- Period: Loans issued between 2007 and 2011.
- Label: Loan Status (e.g., Fully Paid, Current, Charged-off).
- Key Variable Categories : Borrower Information, Loan Attributes, Payment behavior.

Goal:

Identify the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default

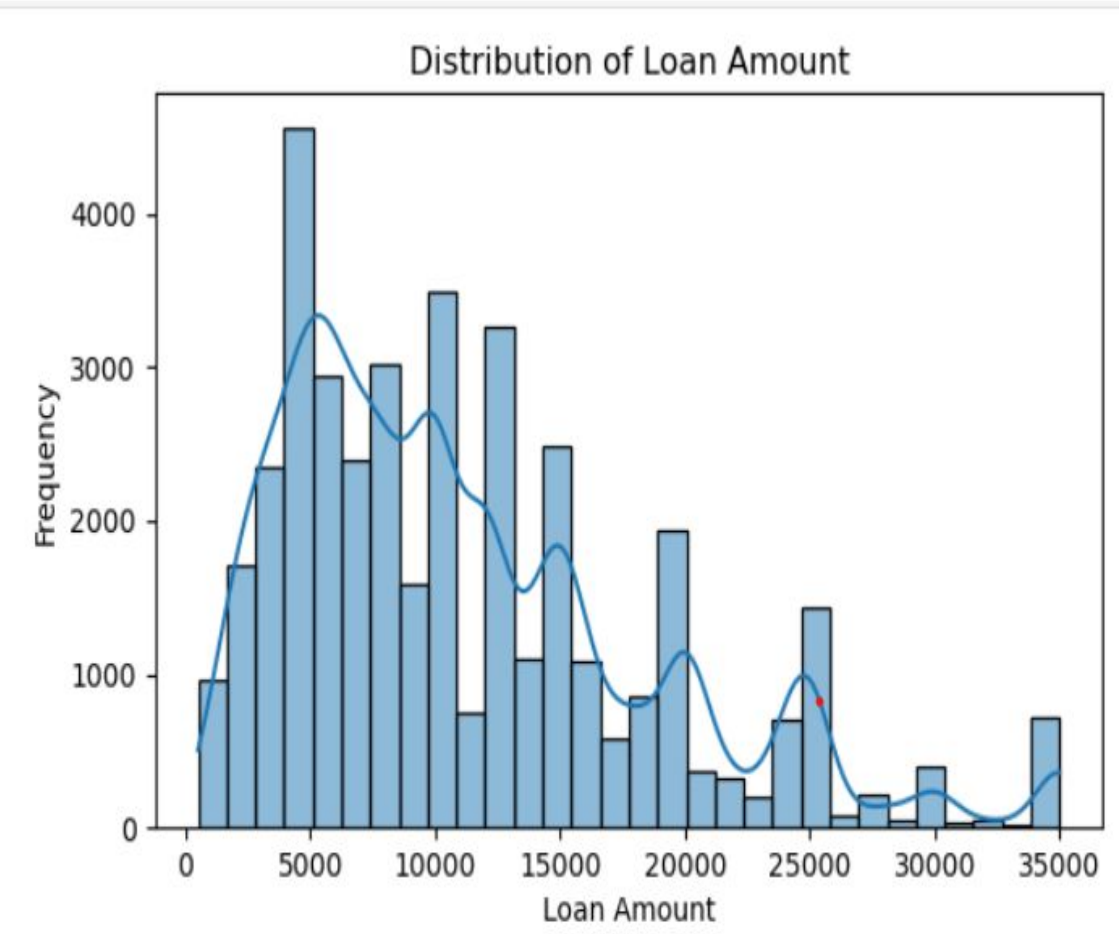
Data Cleaning Summary:

- Whitespace Removal: Stripped extra whitespace from column names and string values.
- Duplicate Removal**:** Eliminated duplicate rows to ensure data integrity.
- Data Type Conversion**:** Converted columns to appropriate types (e.g., datetime, category, numeric).
- Missing Value Handling**:**
  - Filled missing values using mean, median, or mode, or nan based on the column
  - Dropped columns with over 60% missing data.
- Formatting**:** Standardized columns (e.g., dates, currency) for consistency.

DATA ANALYSIS

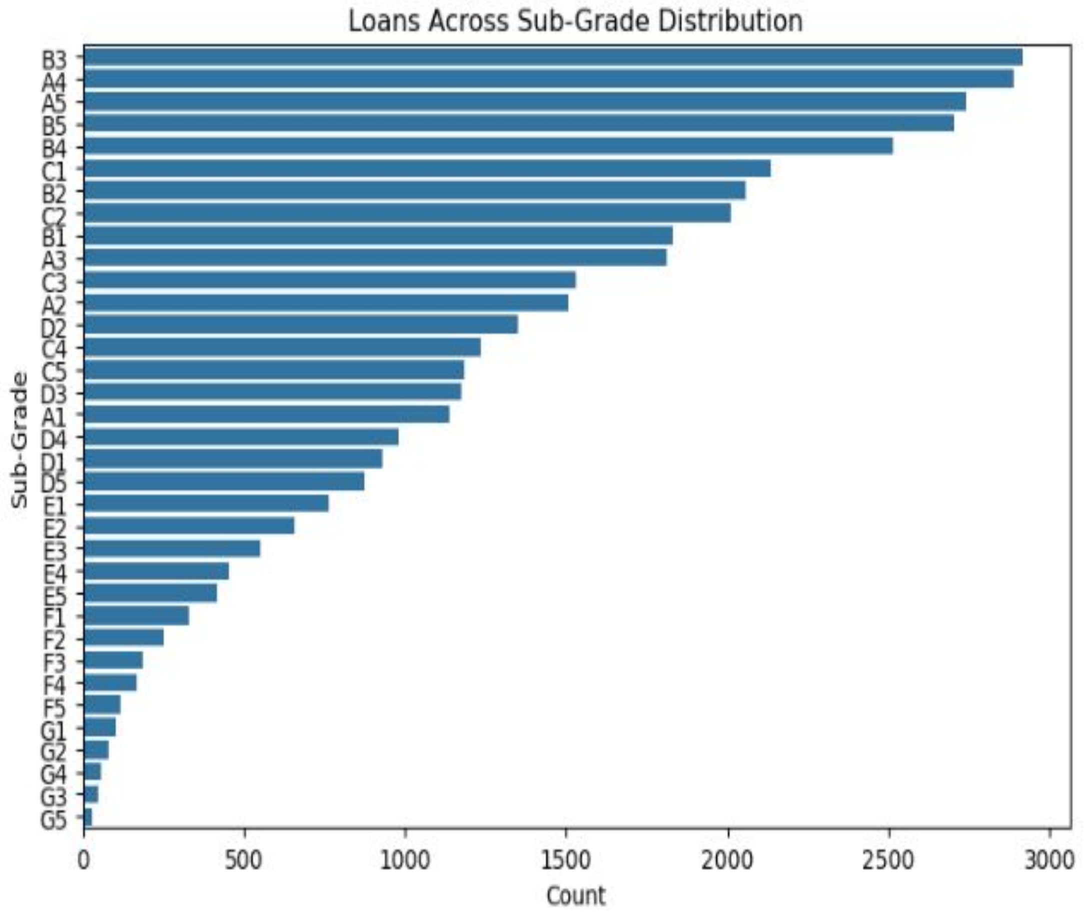
EDA-Univariate Analysis

Distribution of loan amounts across borrowers



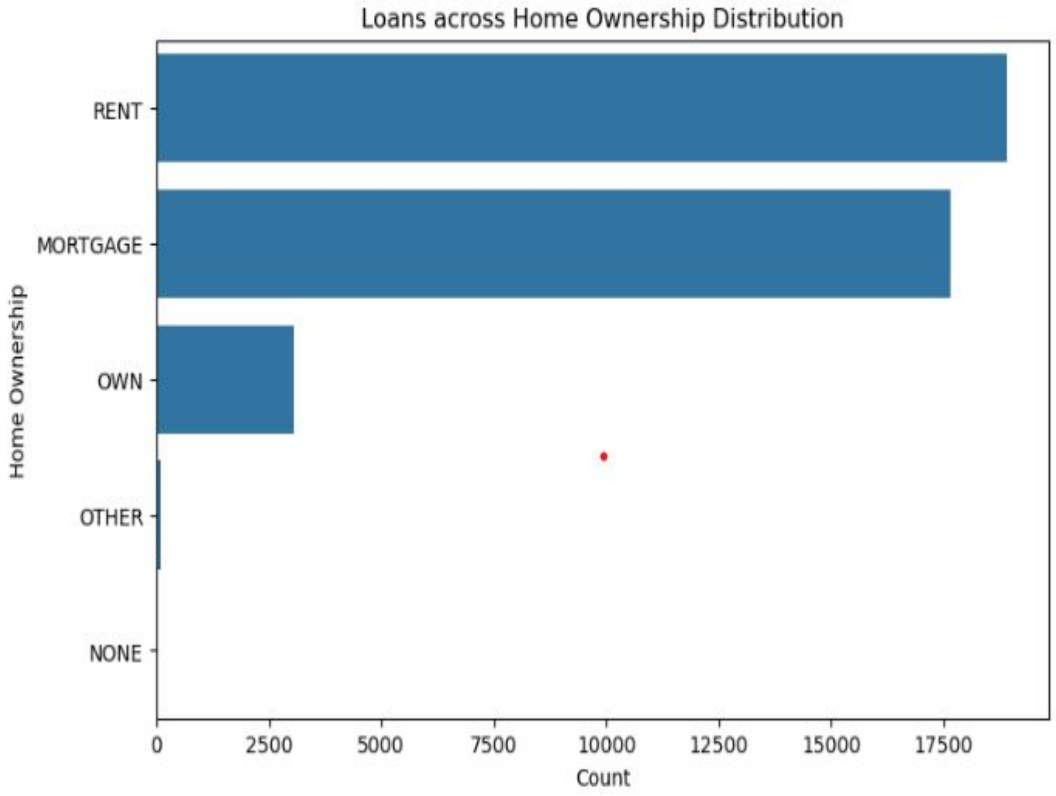
- Most loans fall in the range of $5,000 to $15,000, with peak around $5,000.
- Smaller loans dominate, indicating lower-risk borrowers

Loan Across Sub-Grade Distribution



- Lenders prioritize issuing loans to higher subgrade A and B.
- Borrowers in F/G sub-grades have weaker financial profiles, lenders enforce stricter credit policies for F/G subgrades

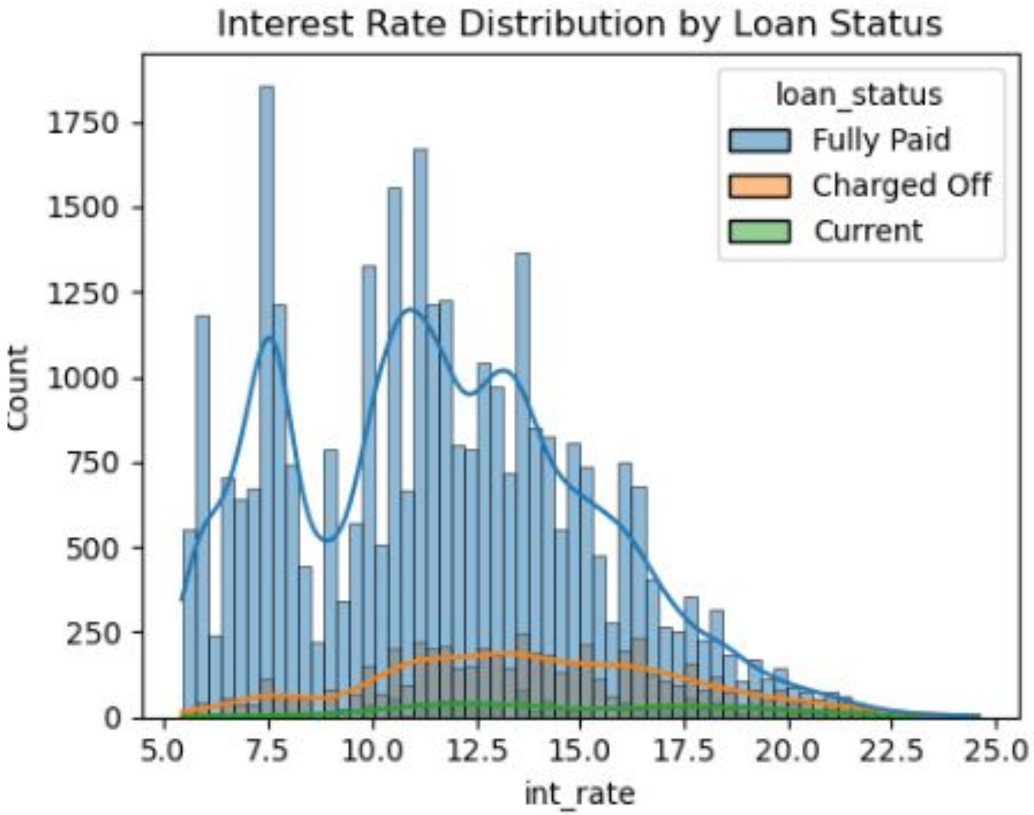
Distribution of loans by home ownership status



- Borrowers with ‘Rent’ and ‘Mortgage’ status account for the majority of loans.
- Borrowers who ‘Own’ their homes constitute a smaller proportion.
- Categories like ‘Other’ and ‘None’ are almost negligible in comparison.

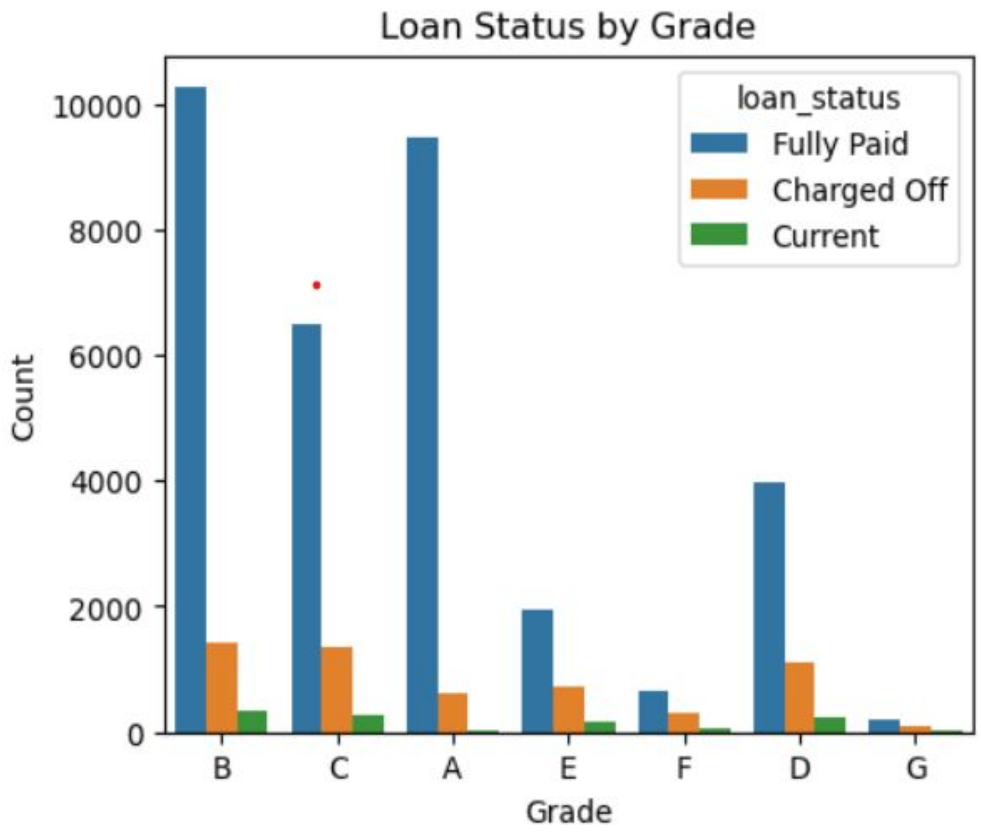
EDA-Bivariate Analysis

Interest Rate Distribution by Loan Status



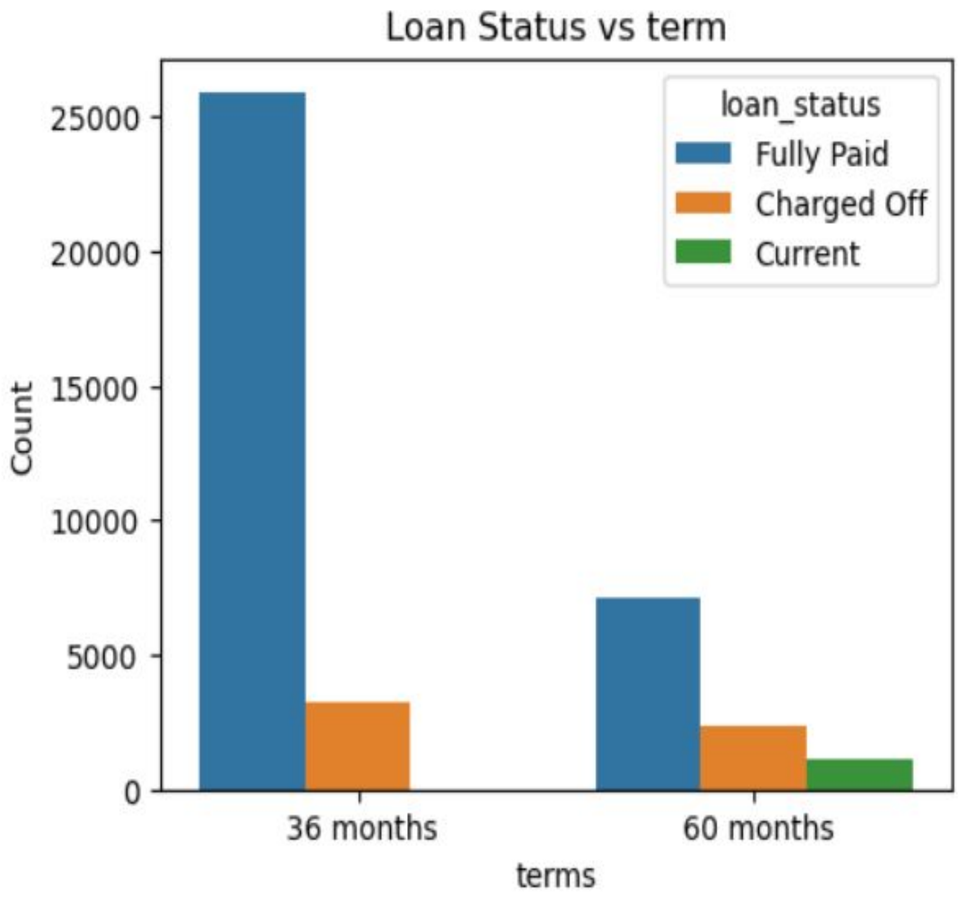
- Higher interest rates are strongly associated with loan defaults, reinforcing the notion that with higher rates are more likely to result in non-payment.

Loan Status by Grade



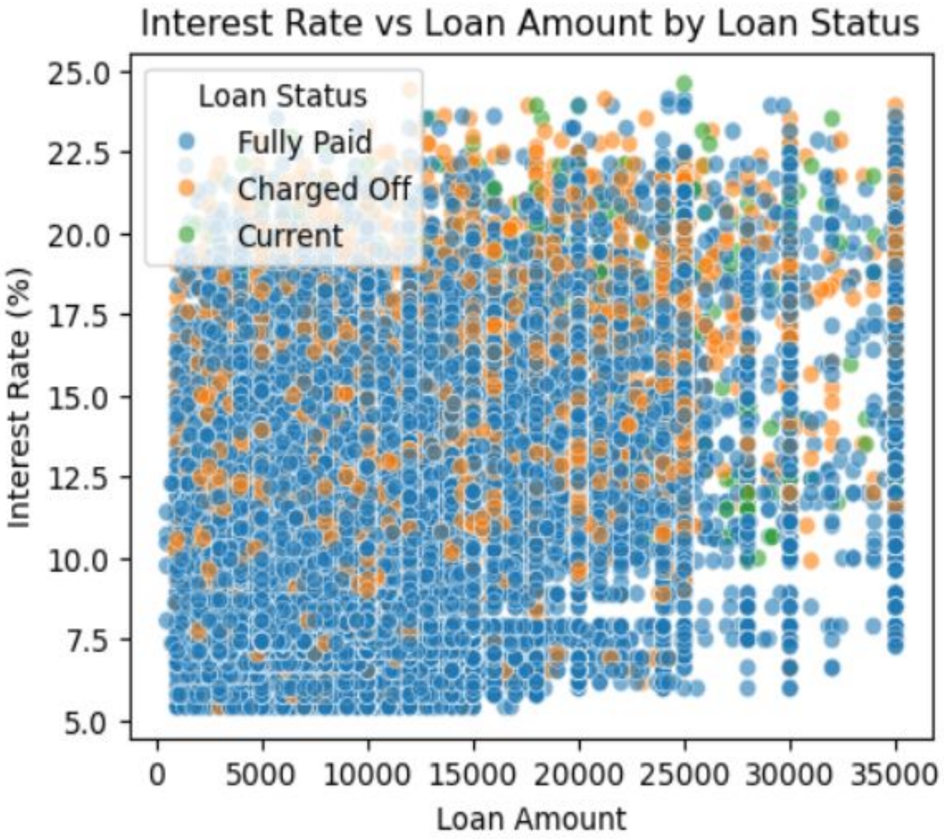
- Borrowers with **lower credit grades (D, E, F, G)** are at a significantly higher risk of defaulting (Charged Off).
- Borrowers in **higher grades (A, B, C)** demonstrate better repayment behavior, highlighting their financial reliability.

Loan Status by Term



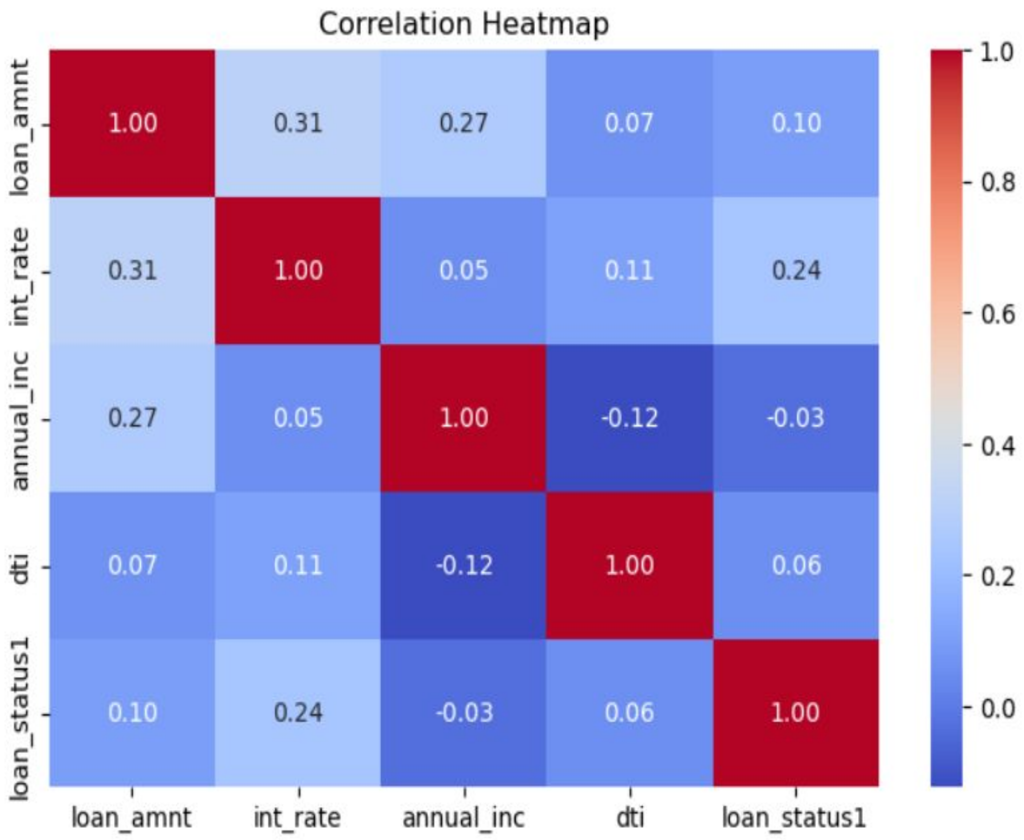
- 36 months, result in a higher proportion of fully paid loans, this could indicate several advantages for the business.

Interest Rate vs Loan Amount by Loan Status



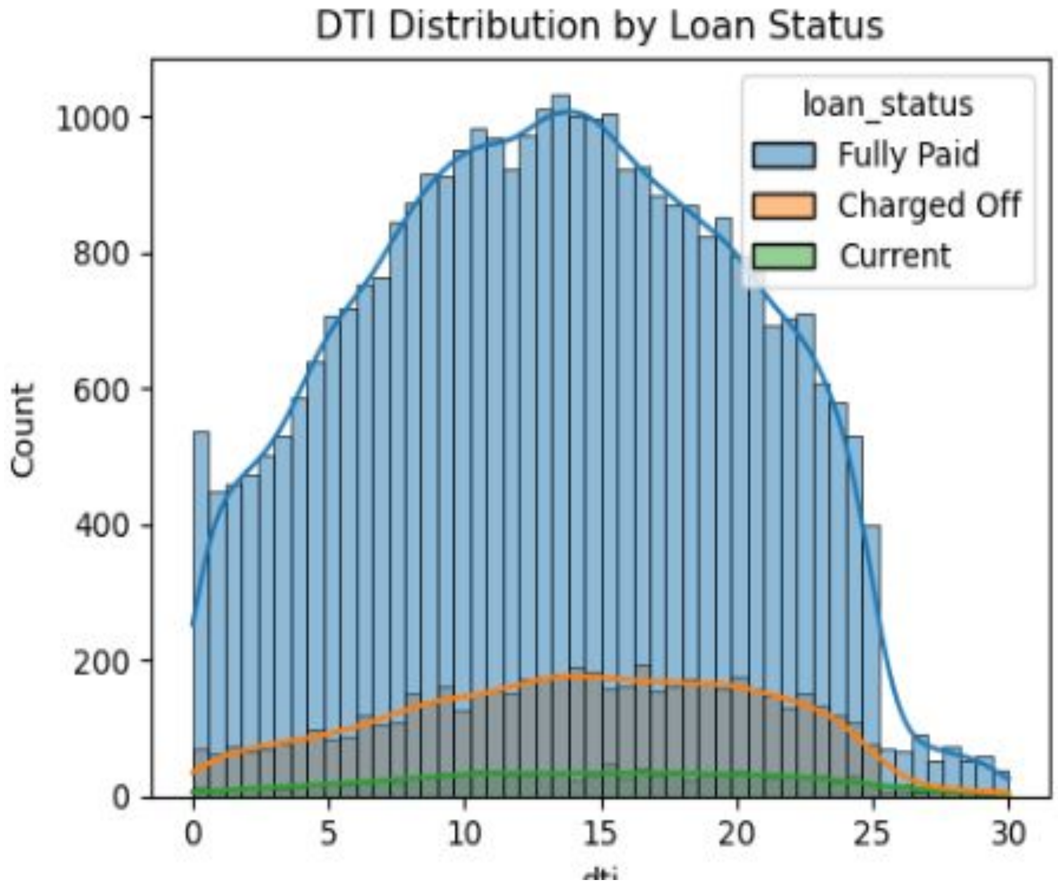
- Charged Off loans (orange dots) are more concentrated at higher interest rates (above ~15%), suggesting that borrowers with higher interest rates are more likely to default.
- Fully paid loan tend to cluster in the lower range (below ~15%), indicating better repayment behaviour.

Correlation Heatmap



- loan amount has Positive correlation with loan_status1 (~0.10): Higher loan amounts are slightly more likely to be associated with default.
- Interest Rate (int_rate) has moderate positive correlation with loan_status1 (~0.24): Indicates that higher interest rates are associated with an increased likelihood of default.
- Annual Income (annual_inc) has weak negative correlation with loan_status1 (~-0.03): Suggests that borrowers with higher annual income are slightly less likely to default.
- Debt-to-Income Ratio (dti) has weak positive correlation with loan_status1 (~0.06): Borrowers with higher debt-to-income ratios are marginally more likely to default.

DTI Distribution by Loan Status



- Majority of borrowers with fully paid loans have a DTI between 10% and 20%.
- Charged-off loans appear more evenly distributed across higher DTI ranges ,indicating that borrowers with higher DTIs are more prone to default.

Key Drivers of Loan Defaults

1. **Loan Amount-** Higher amounts increase default probability.
2. **Interest Rates-** Defaults are more likely with rates >15%.
3. **Income-** Borrowers with lower incomes have a higher likelihood of defaulting.
4. **DTI Ratio-** High ratios indicate financial strain and default risk.
5. **Loan Grades-** Grades D,E,F,G are risk indicators.
6. **Home Ownership-** Renters and mortgage holders are more vulnerable to defaults.
7. **Term-** 36 months, result in a higher proportion of fully paid loans, this could indicate several advantages for the business. Default rate is high for 60 months term

Recommendations

- Tighten approval criteria for high-risk borrowers - e.g., DTI >30%, low grades.
- Adjust loan amounts and interest rates based on borrower profiles
- Offer financial counseling to borrowers with high DTI or low income.
- Use identified patterns to prioritize reliable borrowers and reduce default risks.