Lending Club Case Study

Anand Maitrey
Satyam Singh

ML C66

Lending Club Case Study

Background

We are given the assignment of the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures.

Problem Statement

- As lending loans to ‘risky’ applicants is the largest source of financial loss, we need to identify the risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.
- Identification of such applicants using EDA is the aim of this case study.

Problem Solving Methodology

- **Data Understanding**

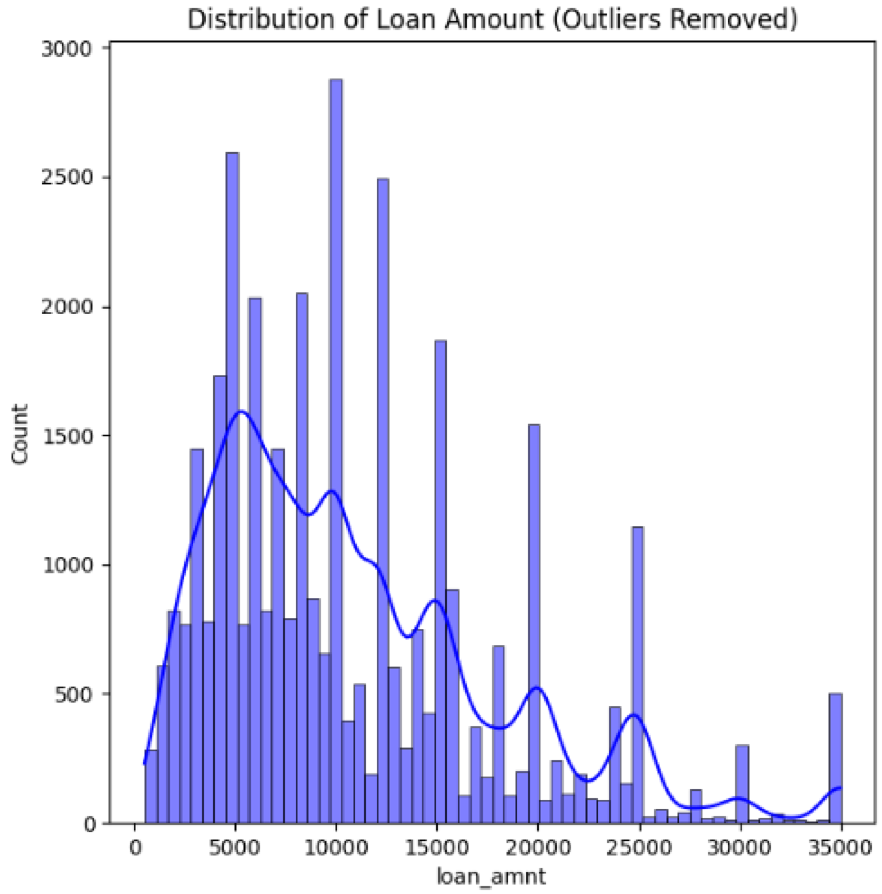
  Information of columns was read from data dictionary. Then data summary was perceived based on data description and columns datatype.

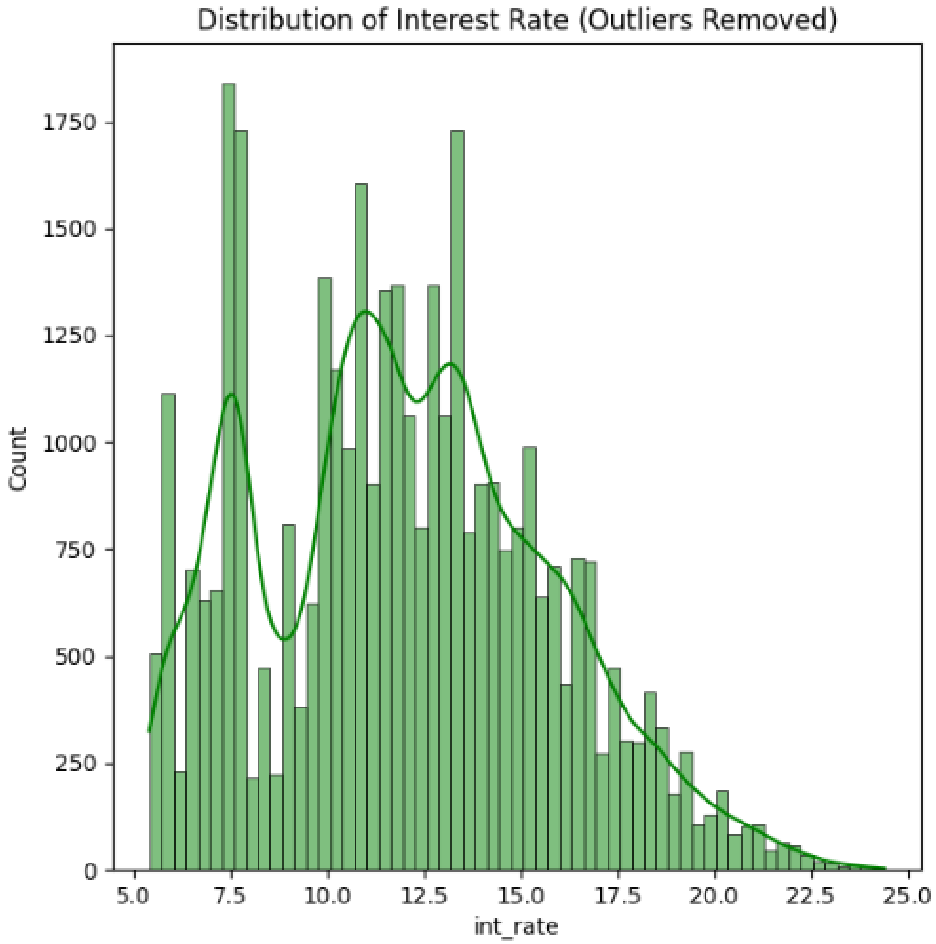
- **Data Cleaning & Manipulation**
  Irrelevant columns and rows are dropped. Columns with null values> 70% or columns with only one repeated values are dropped.
  Only columns relevant to the business data are kept for analysis.
  Outliers are removed for each numerical column.

- **Data Analysis –** Univariate, Segmented univariate & Bivariate analysis is done on the data.

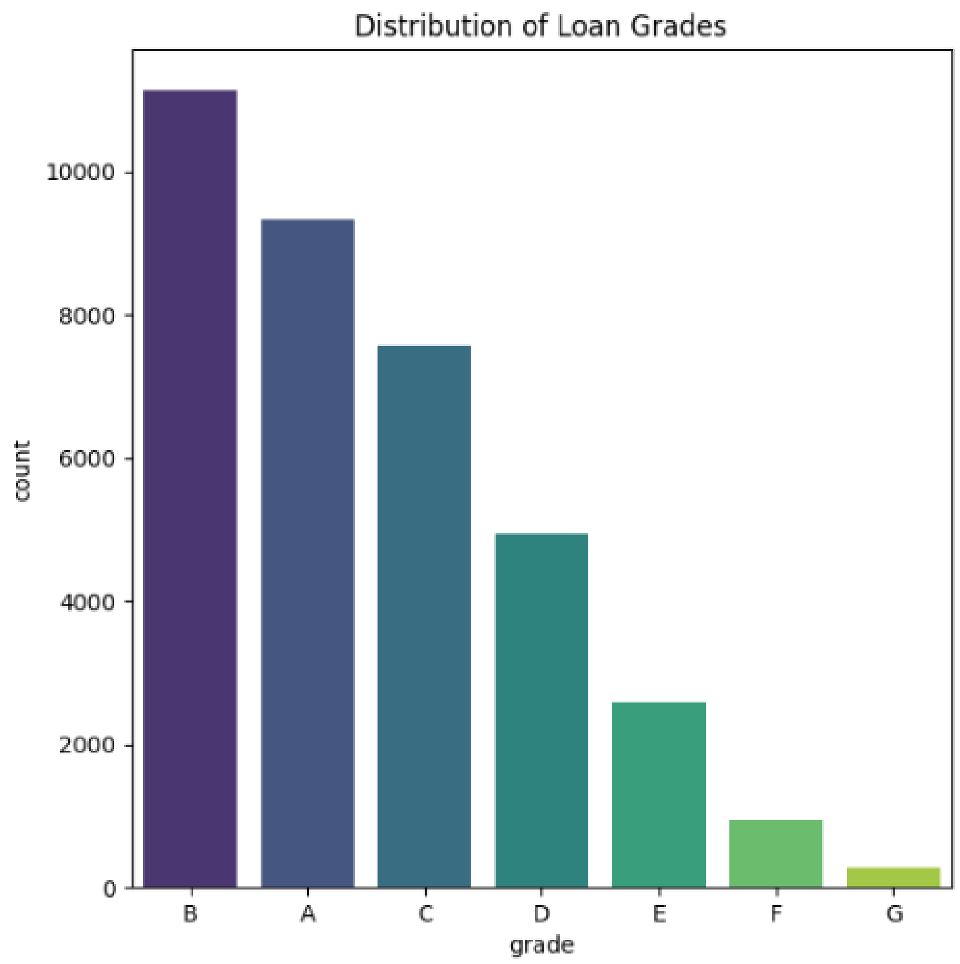
- **Recommendations** are provided in the end based on above analysis.

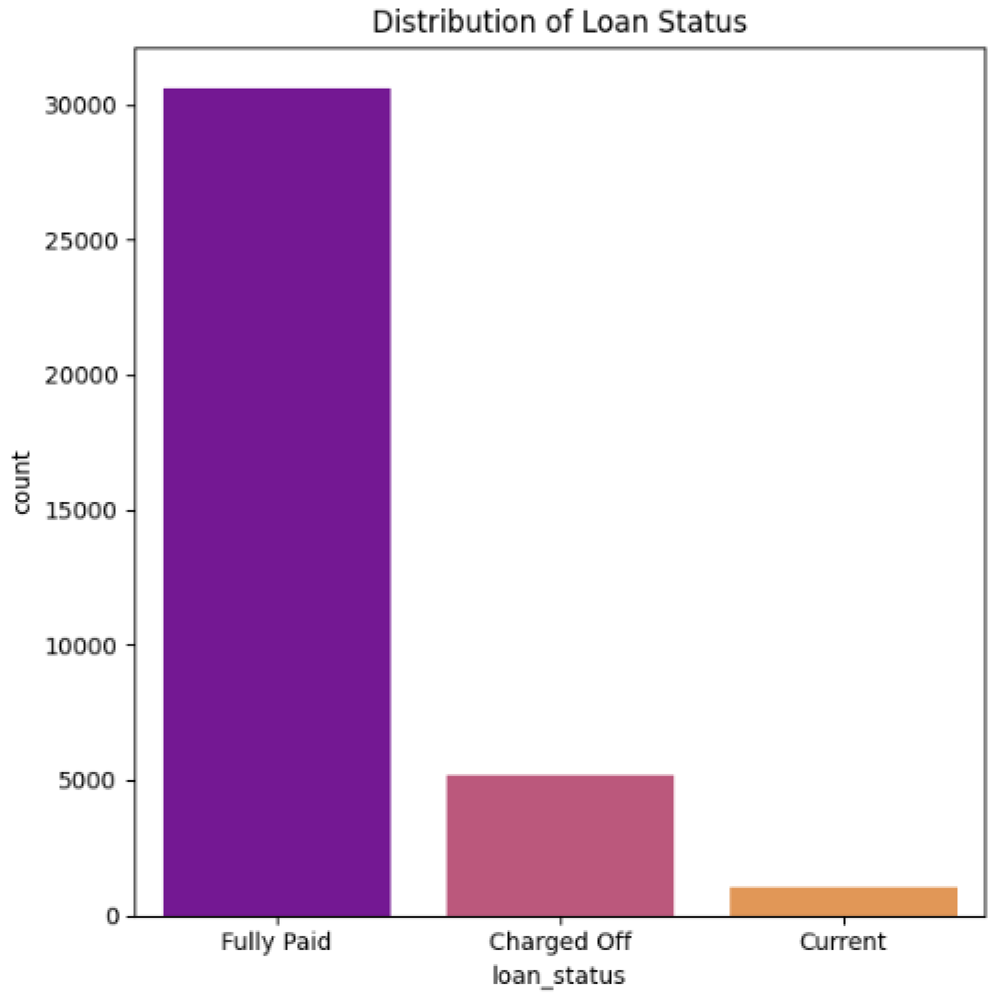
Data Analysis – Univariate Analysis



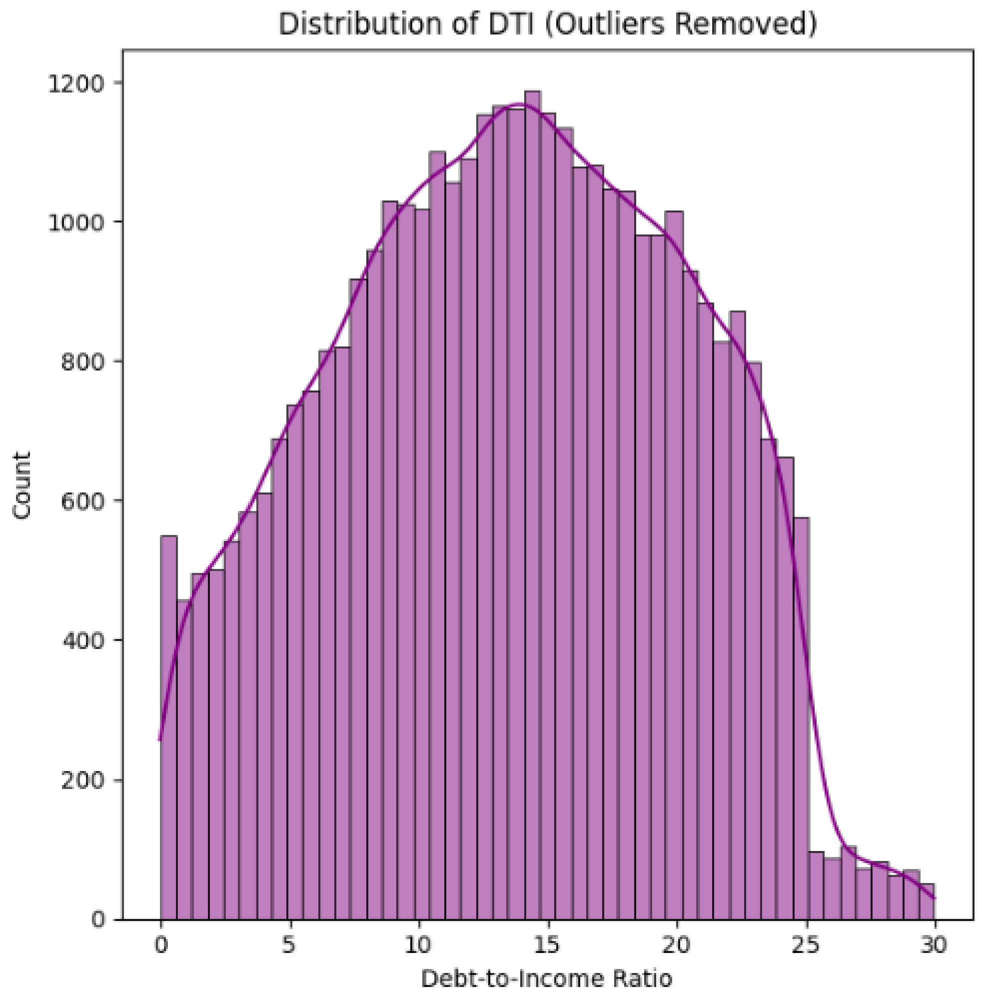


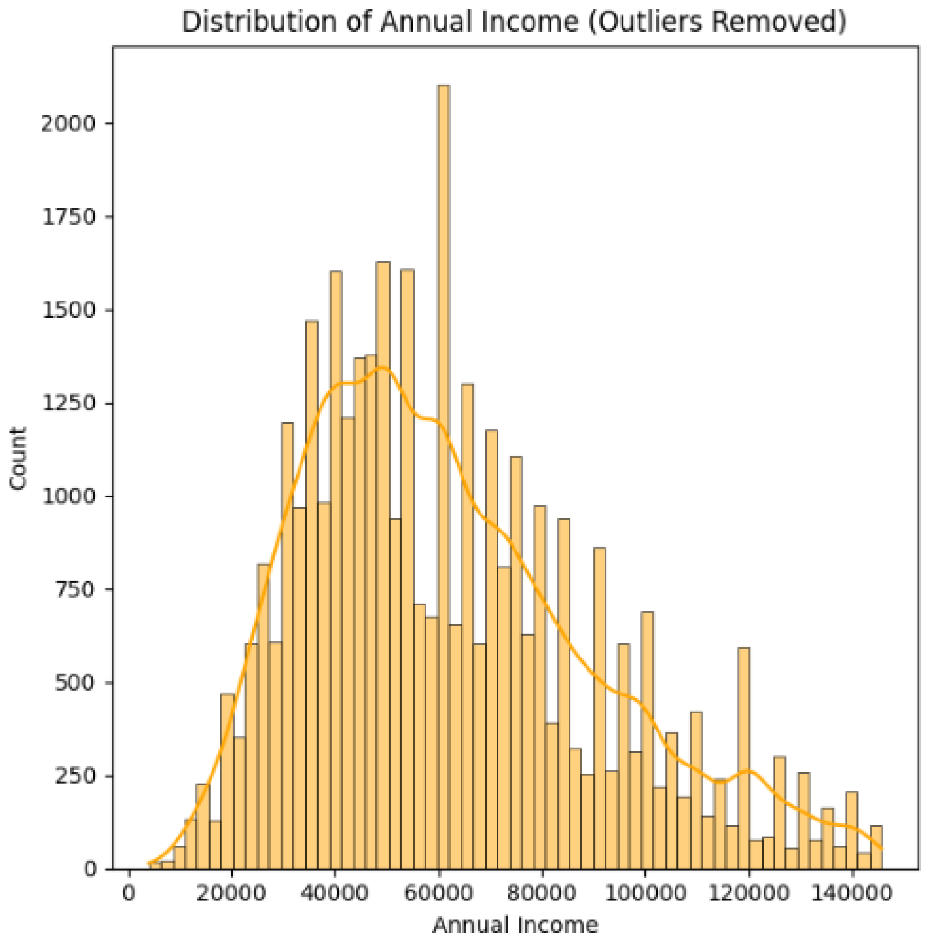
Data Analysis – Univariate Analysis





Data Analysis – Univariate Analysis



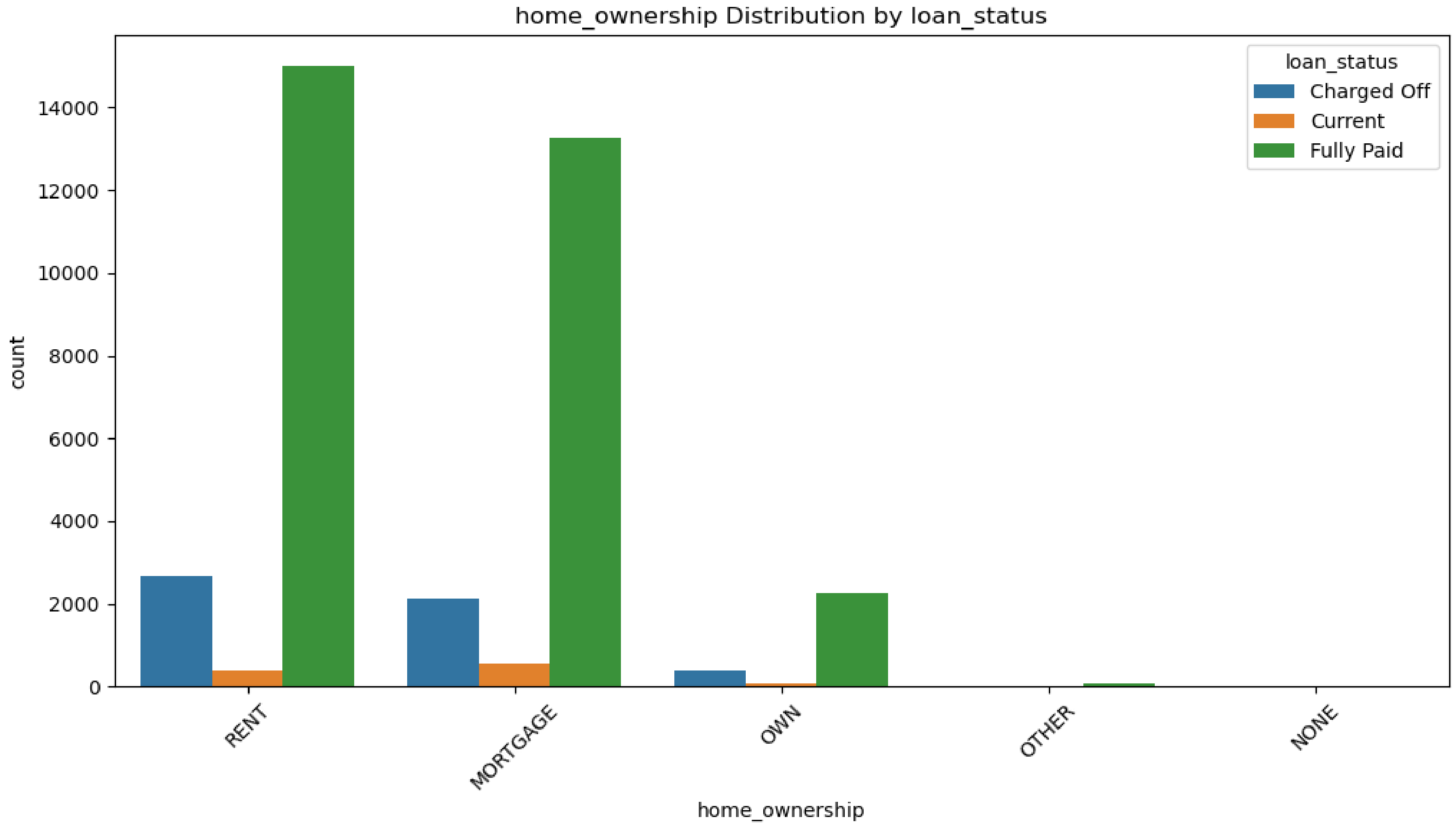


Data Analysis – Univariate Analysis

Observation:

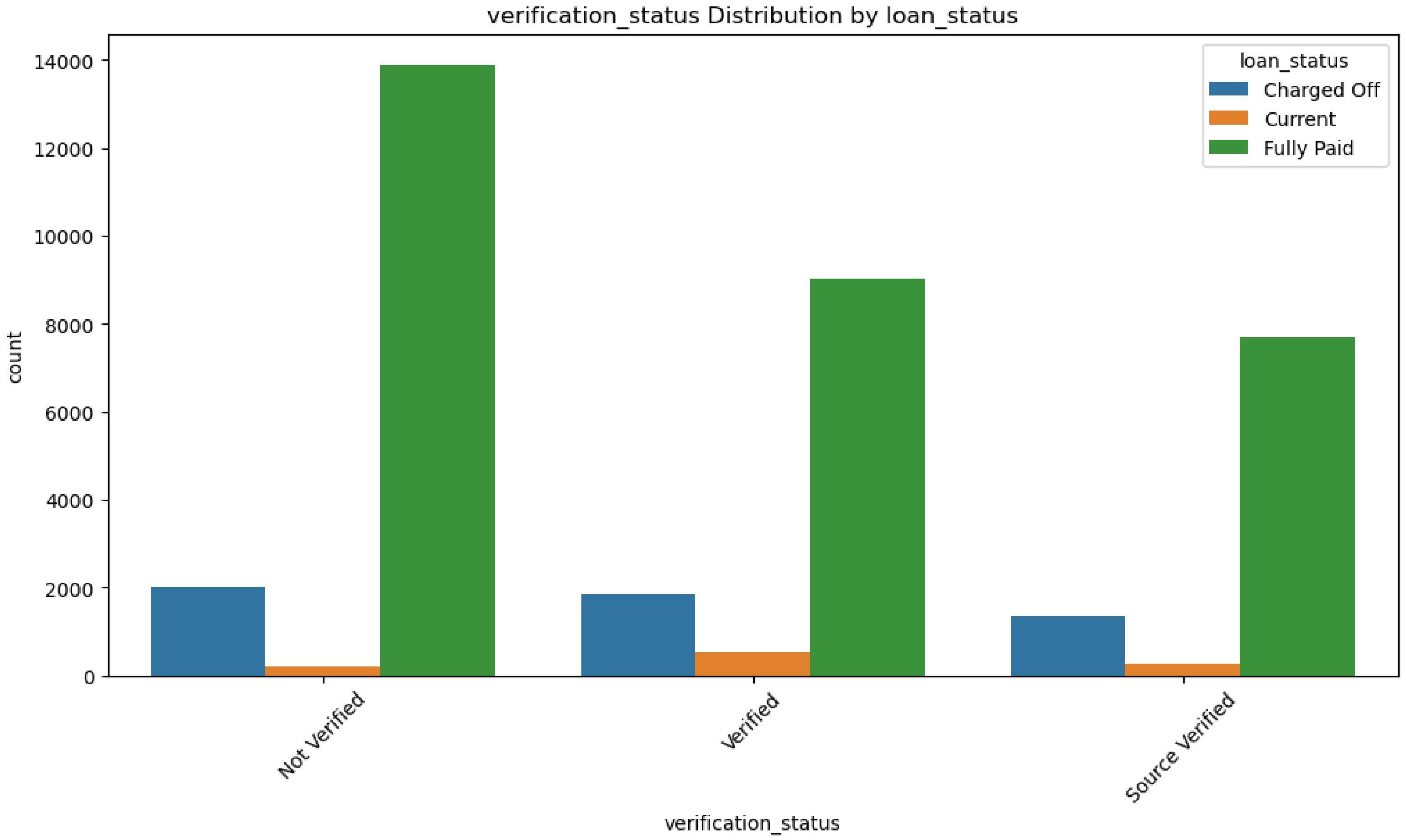
1. Loan Grades: The majority of loans fall into categories B and C, which denotes a moderate level of credit risk.
2. Loan Status: Fewer loans are in default. Most are fully paid off or charged off.
3. Interest Rate: Interest rates have a peak of approximately 10 to 15 percent and a range of approximately 5percent to 25 percent.
4. Loan Amount: The majority of loans range from USD 5000 to USD 15000.
5. Annual Income: While some outliers have substantially greater salaries, the majority of borrowers make less than $100,000 annually.
6. Debt-to-income ratio: A greater ratio has a longer tail, with most DTI ratios falling below 20%.

Data Analysis – Segmented Univariate Analysis



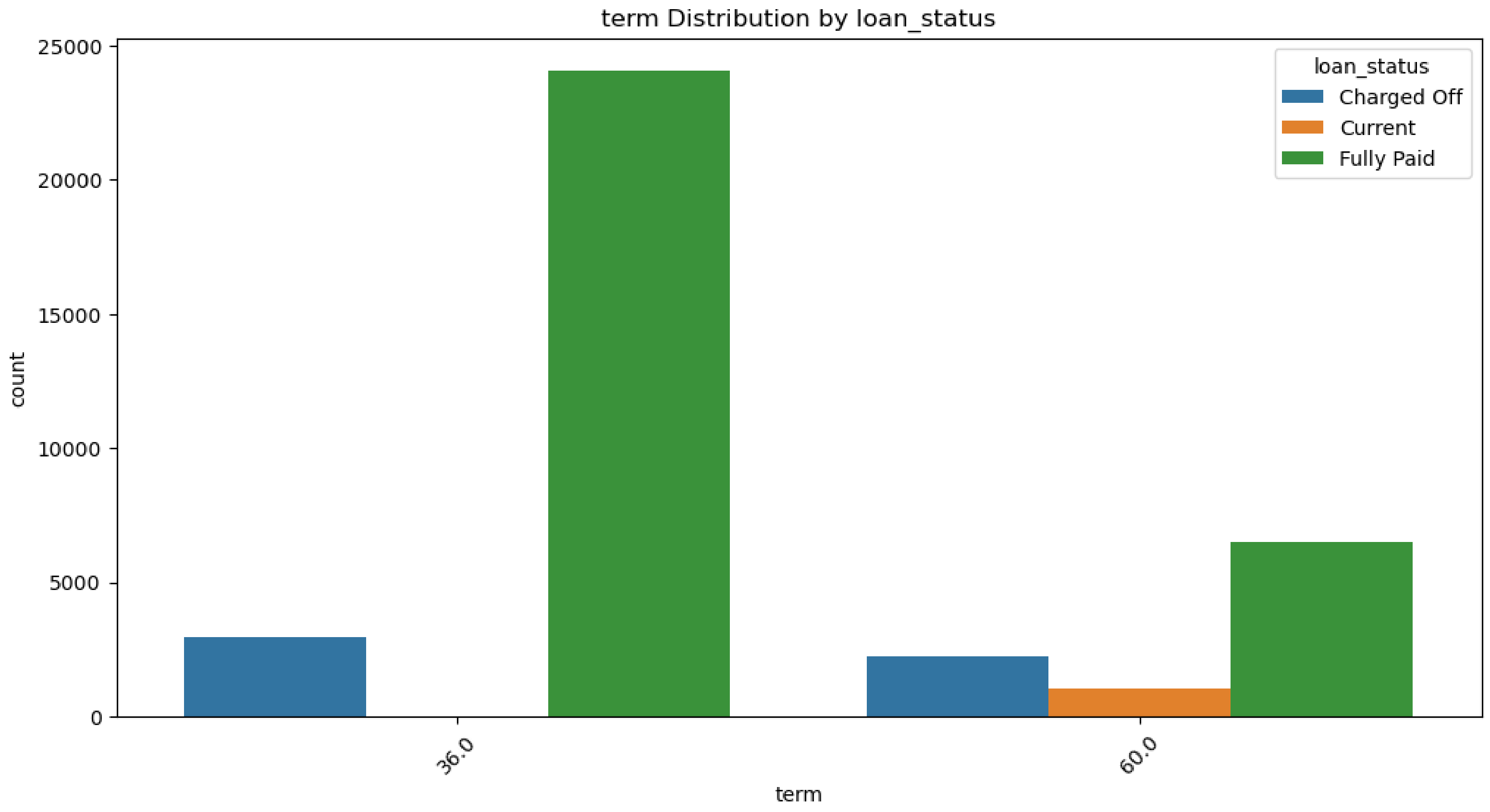
- **Observation**: Compared to homeowners, renters and mortgage holders have a larger percentage of loans that are charged off.

Data Analysis – Segmented Univariate Analysis



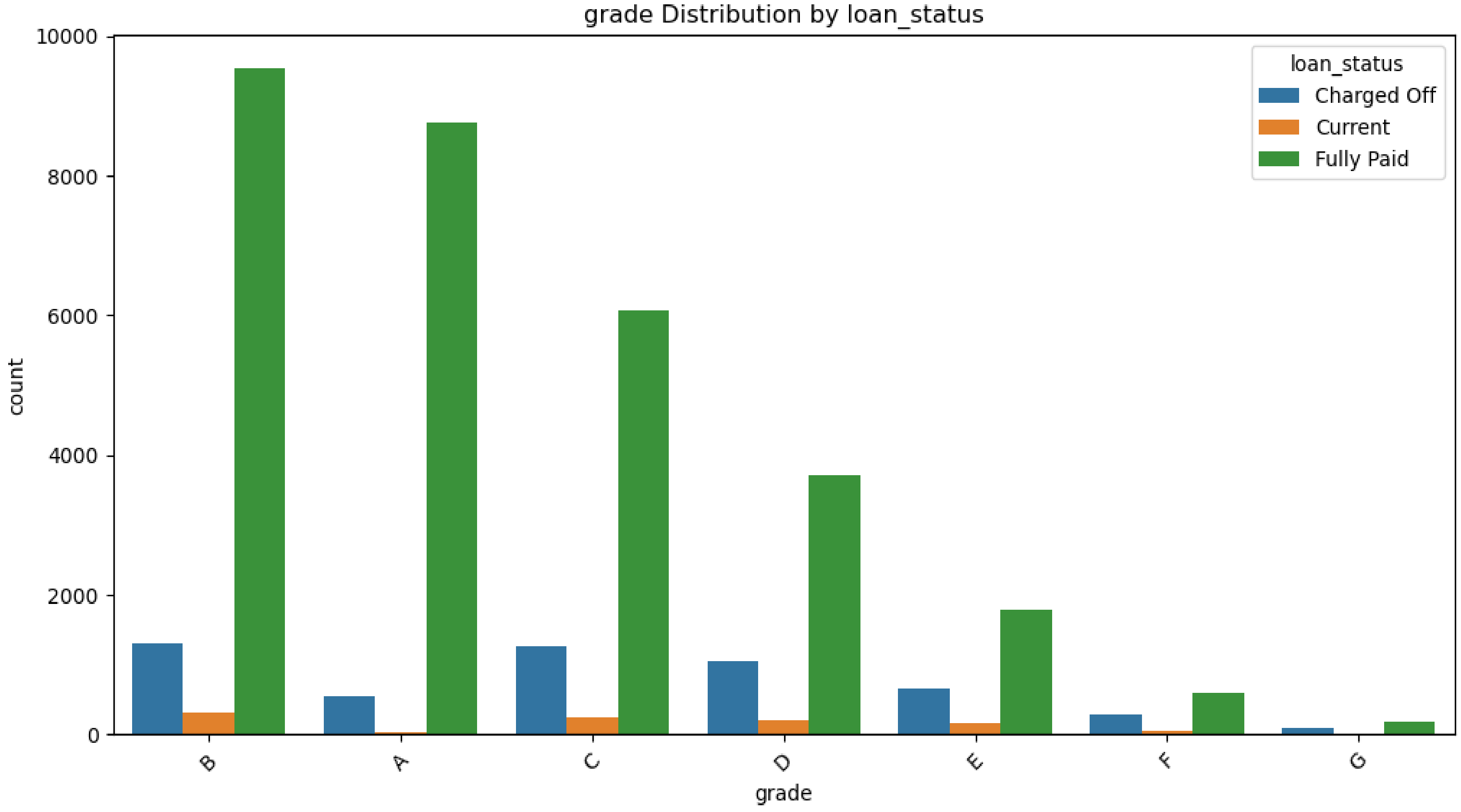
- **Observation**: Charged off loans are minimum in case of income source verified.
- Fully paid loans are common in all segments.

Data Analysis – Segmented Univariate Analysis



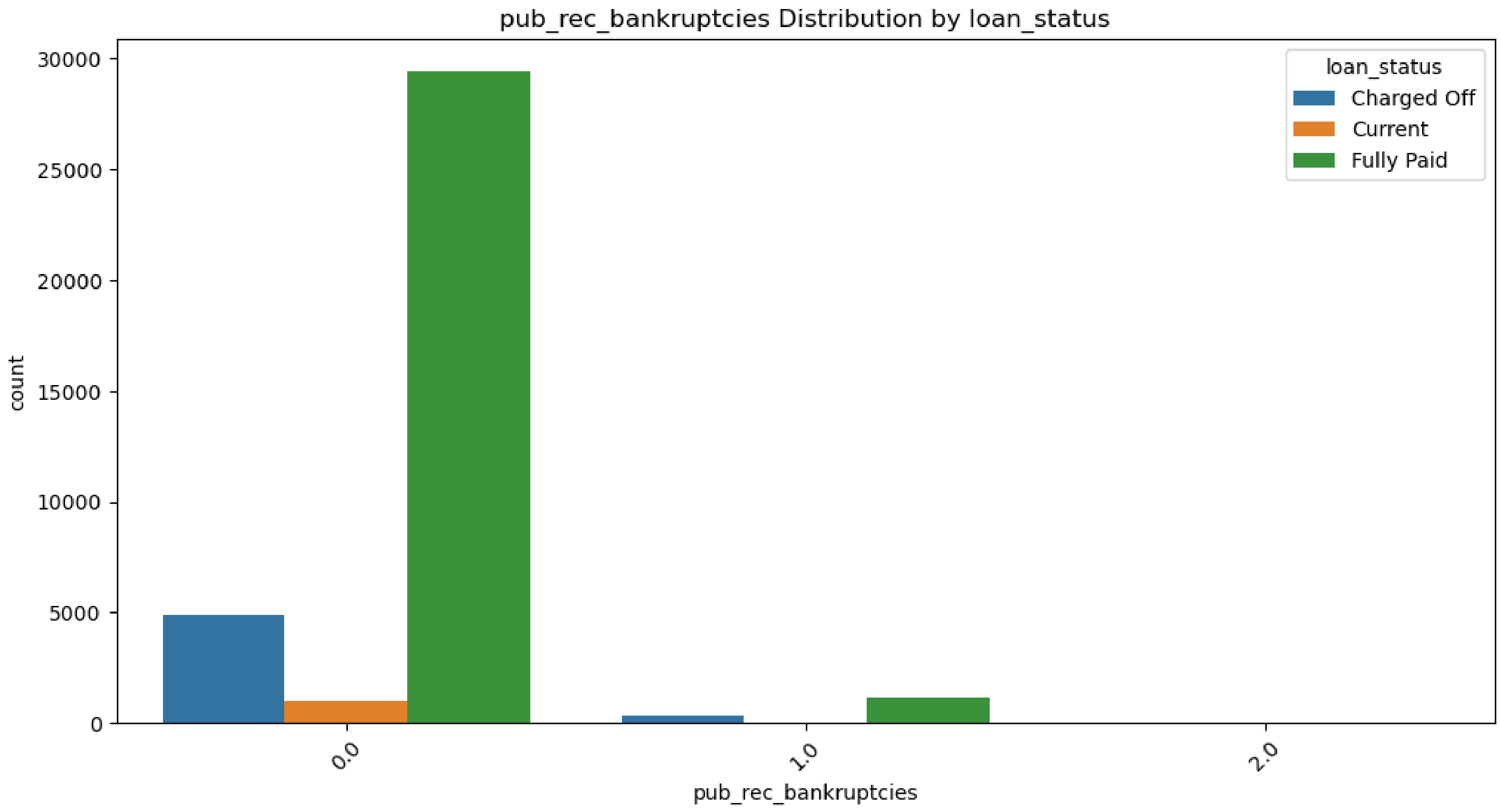
- **Observation**: Loans with 36 months term are more paid off compared to 60 months term.

Data Analysis – Segmented Univariate Analysis



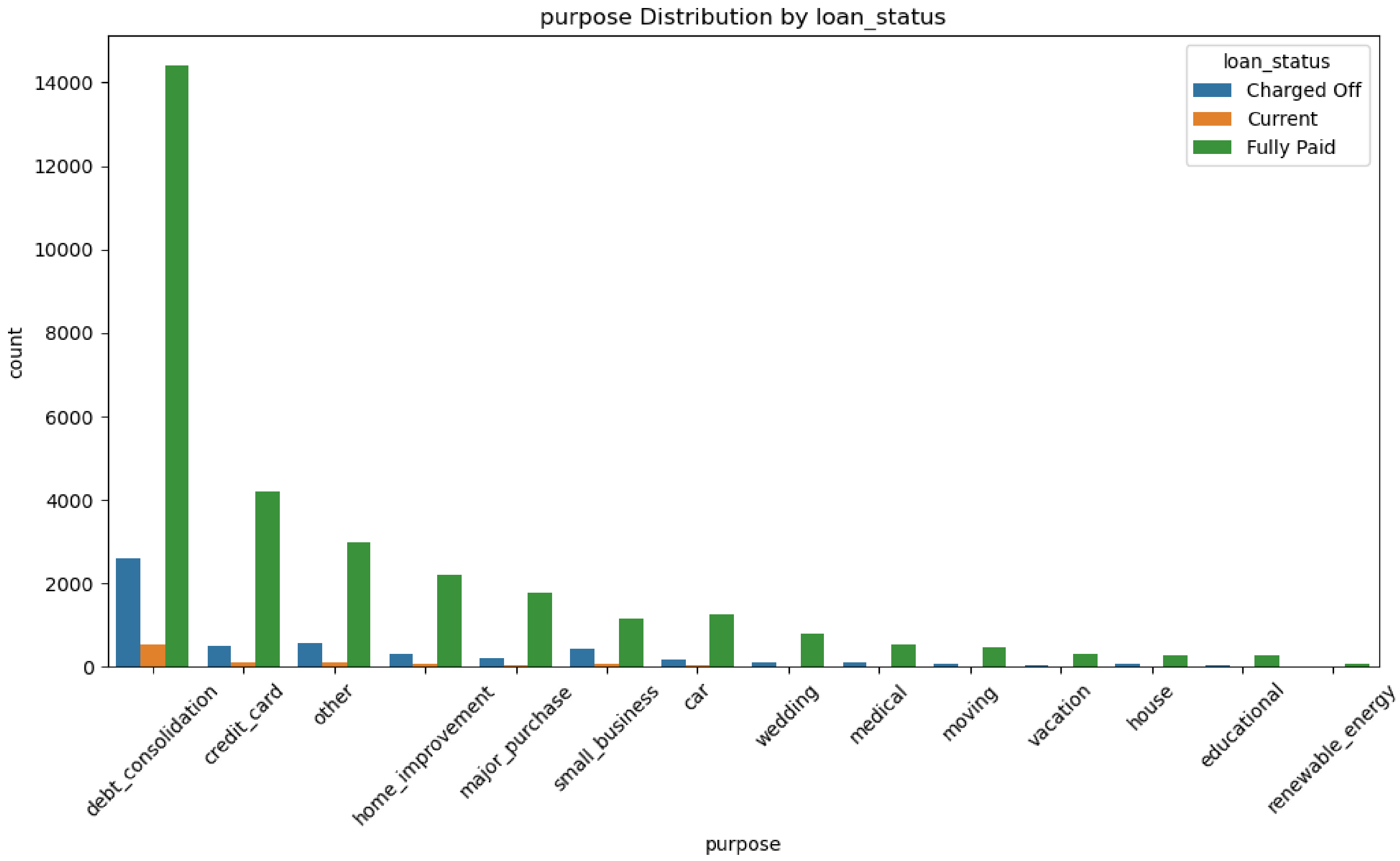
- **Observation**: Higher-grade loans (A, B) have more fully paid loans, whereas lower-grade loans (C, D, and E) have higher rates of being charged off.

Data Analysis – Segmented Univariate Analysis



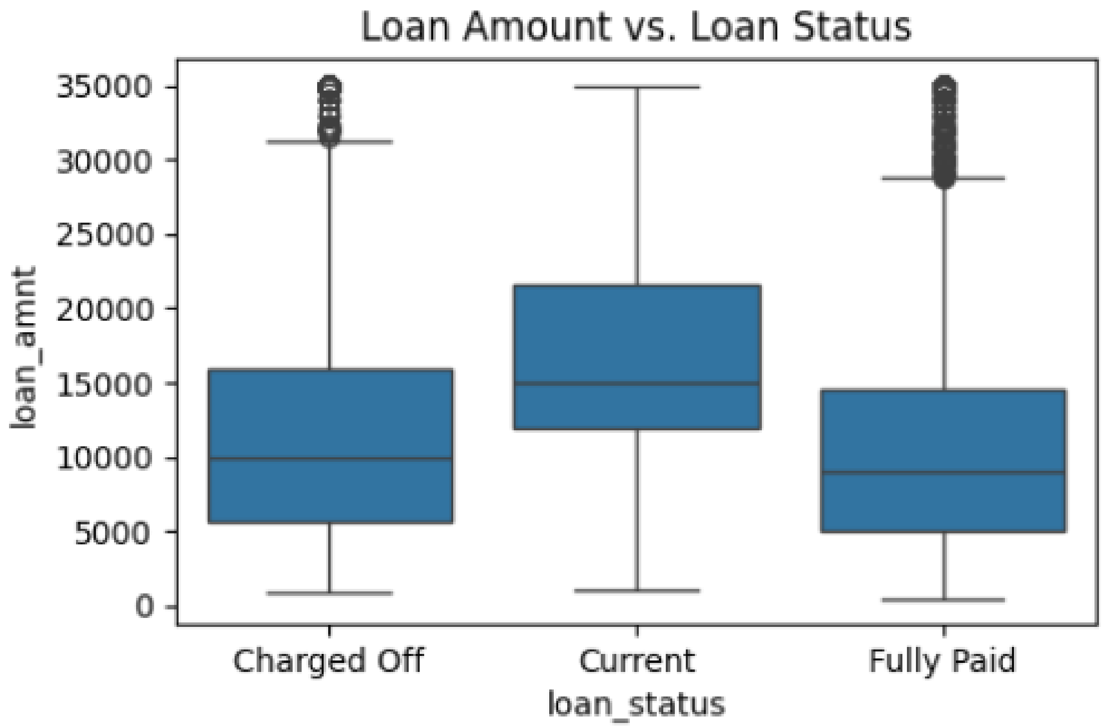
- **Observation**: Loans with no bankruptcy have considerably higher chance of being paid fully.

Data Analysis – Segmented Univariate Analysis



- **Observation**: Debt consolidation charged-off loans, Small business loans and credit card refinancing also indicate a higher risk.

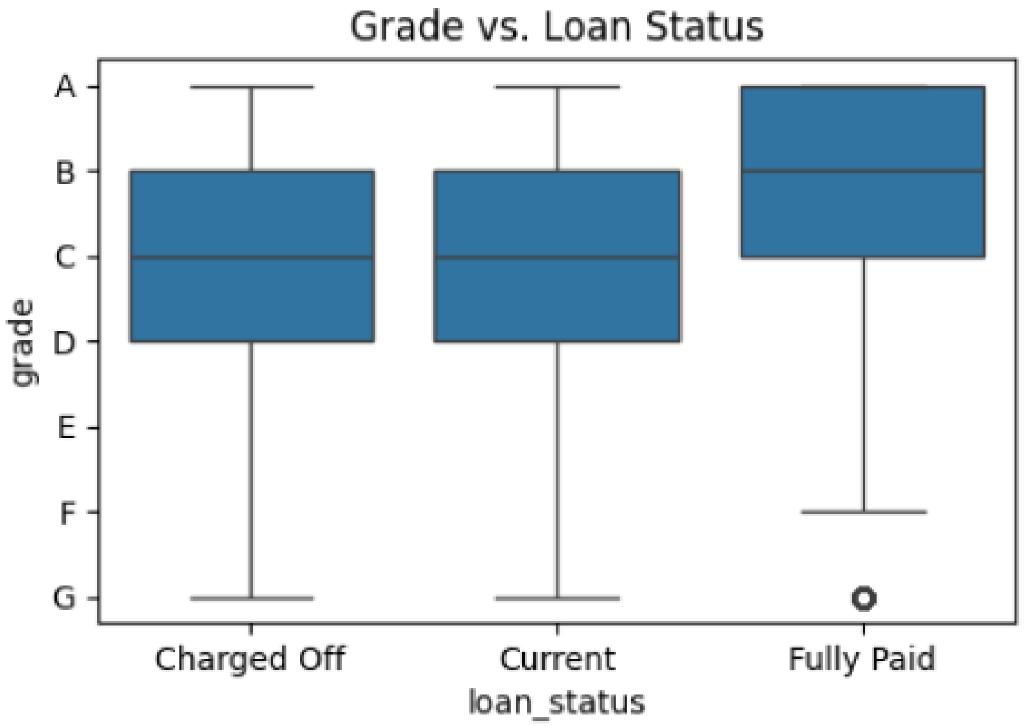
Data Analysis – Bivariate Analysis
Loan Amount vs Loan Status



- **Observation**: Higher loan amounts tend to be associated with higher risks.

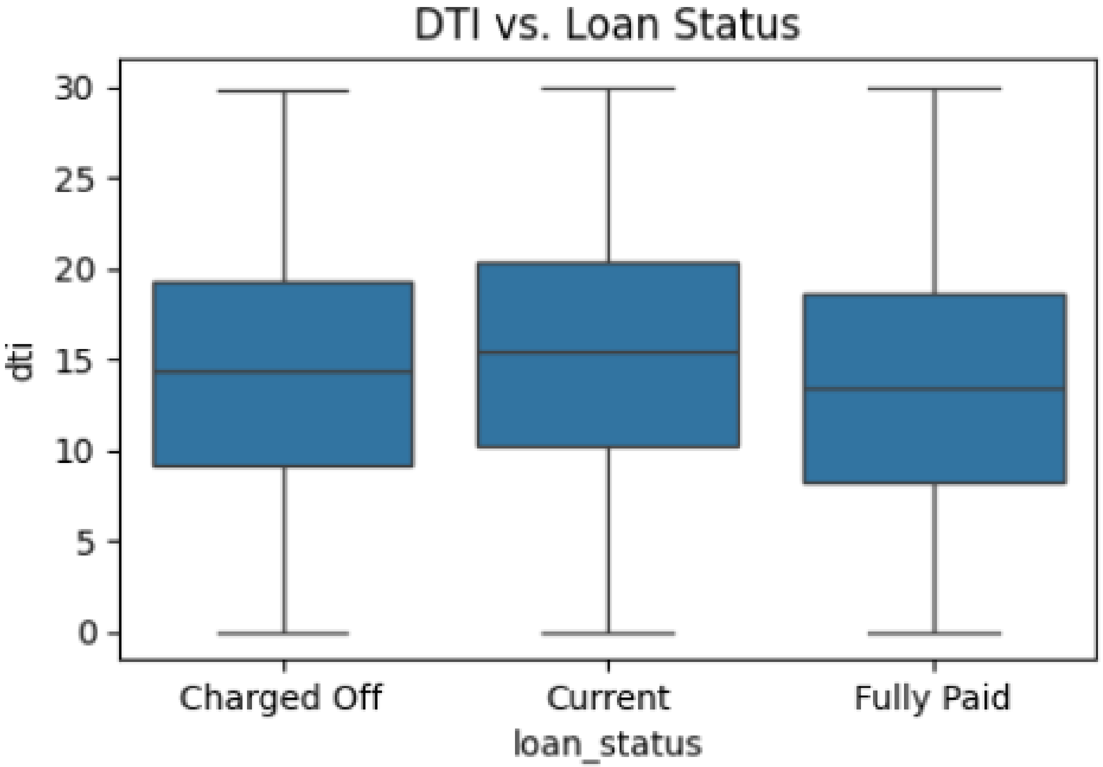
Data Analysis – Bivariate Analysis

Grade vs Loan Status



- **Observation**: Higher grades (A, B) are less likely to be defaulted, while lower grades (C, D, E, etc.) are associated with higher charge off.

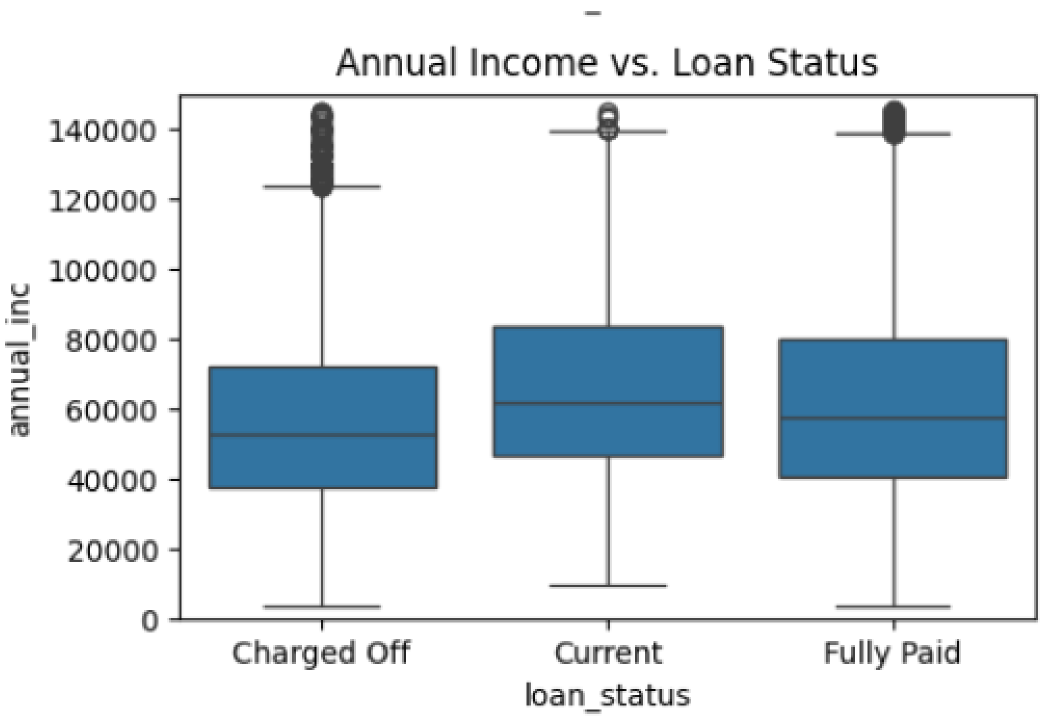
Data Analysis – Bivariate Analysis
DTI vs Loan Status



- **Observation**: Though there is not a significant correlation in DTI vs loans status, but borrowers with fully paid loans have slightly lower debt to income ratio.

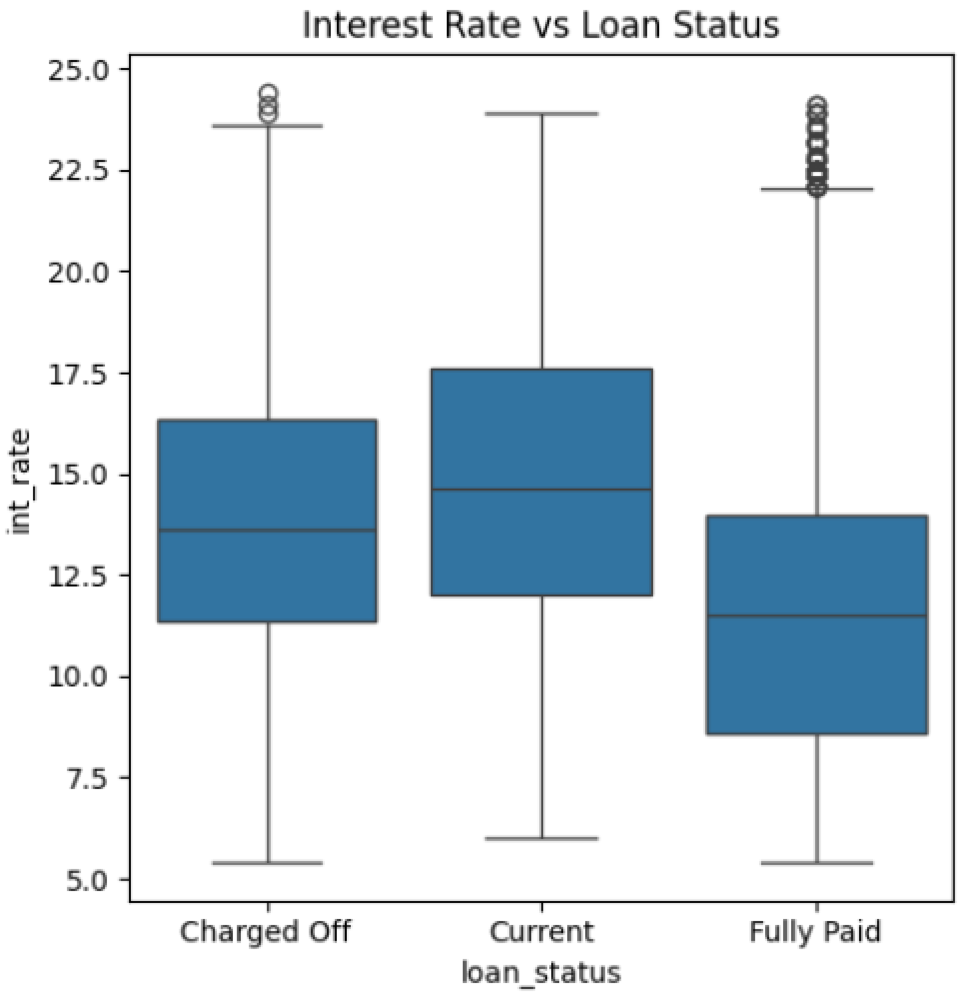
Data Analysis – Bivariate Analysis

Annual Income vs Loan Status



- **Observation**: Lower income is associated with higher charged off.

Data Analysis – Bivariate Analysis
Interest Rate vs Loan Status



- **Observation**:
  Loans with lower interest rates are more likely to be fully paid.

Recommendations

- We need to be cautious while approving higher loan amounts, particularly for borrowers with higher risk profiles (e.g., low credit grades, high DTI, low income etc.).
- As higher grades are less likely to be defaulted, prefer customer with grade A,B.
- Prefer customers who are already home-owners.
- Customers with higher & stable income can be prioritized.
- Loans with 36 months term have more chances to be paid back compared to 60 months.
- Loans where applicant’s source of income is verified are more reliable.
- Loan applicants with history of bankruptcy are less preferable.
- Loans with interest rates between 7.5% to 11% are more preferable.