Lending Club Case Study

Submitted by
Kamarudheen P
Akanksha Kumari

Contents

Problem Statement & Objective

Lending Club, a consumer finance marketplace that specializes in providing a diverse range of loans to urban customers, faces a significant challenge in optimizing its loan approval process. When assessing loan applications, the company must make informed decisions to minimize financial losses, particularly those arising from loans extended to applicants deemed "risky."

These financial losses, known as credit losses, occur when borrowers fail to repay their loans or default on their obligations. In essence, borrowers classified as "charged-off" contribute the most substantial losses to the company.

OBJECTIVE:

- understand the driving factors (or driver variables) behind loan default, i.e. the variables which are strong indicators of default. The company can utilize this knowledge for its portfolio and risk assessment.

Data Understanding

- **The data set (loan.csv) provided has 39717 rows and 111 columns which include multiple attributes related to the "customer" and “Loan"information”**
- **Since our goal is to find the features which influence default of loan, "loan_status" is our target variable.**

- Loan Status:
    - Fully Paid: Loan successfully repaid by the borrower.
    - Charged Off: Borrower defaulted, and lender doesn’t expect further recovery.
    - Current: Ongoing loans still being paid (excluded from analysis).

Major Customer and Loan Atributes:

Customer Information	Loan Information
Annual Income (annual_inc)	Loan Amount (loan_amt)
Verification Status	Interest Rate (int_rate)
Employment Title (emp_title)	Loan Purpose (purpose)
Employment Length (emp_length)	Term
Home Ownership	
Debt-to-Income Ratio (DTI)	
State (addr_state)	

Data Cleaning and Manipulation

Missing Value Analysis:

- ➢ 54 column contains 100% null values.
- ➢ 2 columns contain over 90% null values.
- ➢ 1 column contains more than 60% null values.
- ➢ 43 columns have no null values.

Treatment:

- ✓ Columns with >50% null values: Dropped, resulting in 51.35% of the columns removed.
- ✓ emp_title (6% missing): Imputed with 'Unknown' instead of the mode ('US Army') as it doesn’t logically fit all missing values.
- ✓ emp_length: Missing values imputed with the mode ('10+ years'), a reasonable assumption for employment duration.
- ✓ pub_rec_bankruptcies: Rows with missing values were dropped, as imputing with the mode ('0’) seem’s inappropriate.
- ✓ title: Missing values imputed with the mode ('Debt Consolidation'), the most common title in the dataset.

Dropping Non-Contributory Columns for Improved Analysis

Dropped Post-Approval Columns**:**

Since our goal is to identify features that influence whether a customer is likely to default on a loan, Certain columns capture information only available after loan approval, which does not help in predicting loan defaults at the application stage.

These columns include:
Examples: delinq_2yrs, total_pymnt, total_rec_prncp, total_rec_int, total_rec_late_fee

✓Total Dropped: 18 columns related to post-loan approval events.

Dropped Irrelevant Customer Attributes**:**

Some attributes do not influence loan default decisions and are therefore excluded. For example, id (a unique loan listing identifier) has no impact on the likelihood of default.

Examples: id, member_id, desc, url, zip_code
✓Total Dropped: 5 columns deemed irrelevant.

Outlier Treatment

1. Annual Income:

- High values like 6,000,000, 3,900,000, 1,900,000 etc were identified as outliers.
- Treatement: High-income borrowers may still need loans, so these values are not treated as outliers.

2. Interest Rate:

- Rates like 24.59%, 24.4%, 24.11% etc were identified as outliers.
- Treatment: High rates are realistic for riskier borrowers, so these values are not treated as outliers.

3. Loan Amount:

- Values like 35,000, 34,800, 34,675 etc were identified as outliers.
- Treatment: The company may offer high loan amounts, so these values are not treated as outliers.

Derived Columns (Total: 7 New Columns derived based on existing data.)

1.issue_m and issue_y:

- Extracted the month and year from the loan issuance date (issue_d).

2.annual_inc_bucket:

- Categorized annual income into buckets:

['0 - 40k', '40k - 50k', '50k - 60k', '60k - 70k', '70k - 80k', '80k - above’].

3. int_rate_bucket:

- Grouped interest rates into categories:

['0%-5%', '5%-9%','9%-13%','13%-17%','17%-21%','21%-25%’].

4. loan_amnt_bucket:

- Classified loan amounts into buckets:

['0- 5K', '5K - 10K', '10K - 15K', '15K - Above’].

5. funded_amnt_bucket:

- Grouped funded amounts into categories:

['0 - 5k', '5k - 10k', '10k - 15k', '15k - Above’].

6. dti_bucket:

- Segmented Debt-to-Income ratios into:

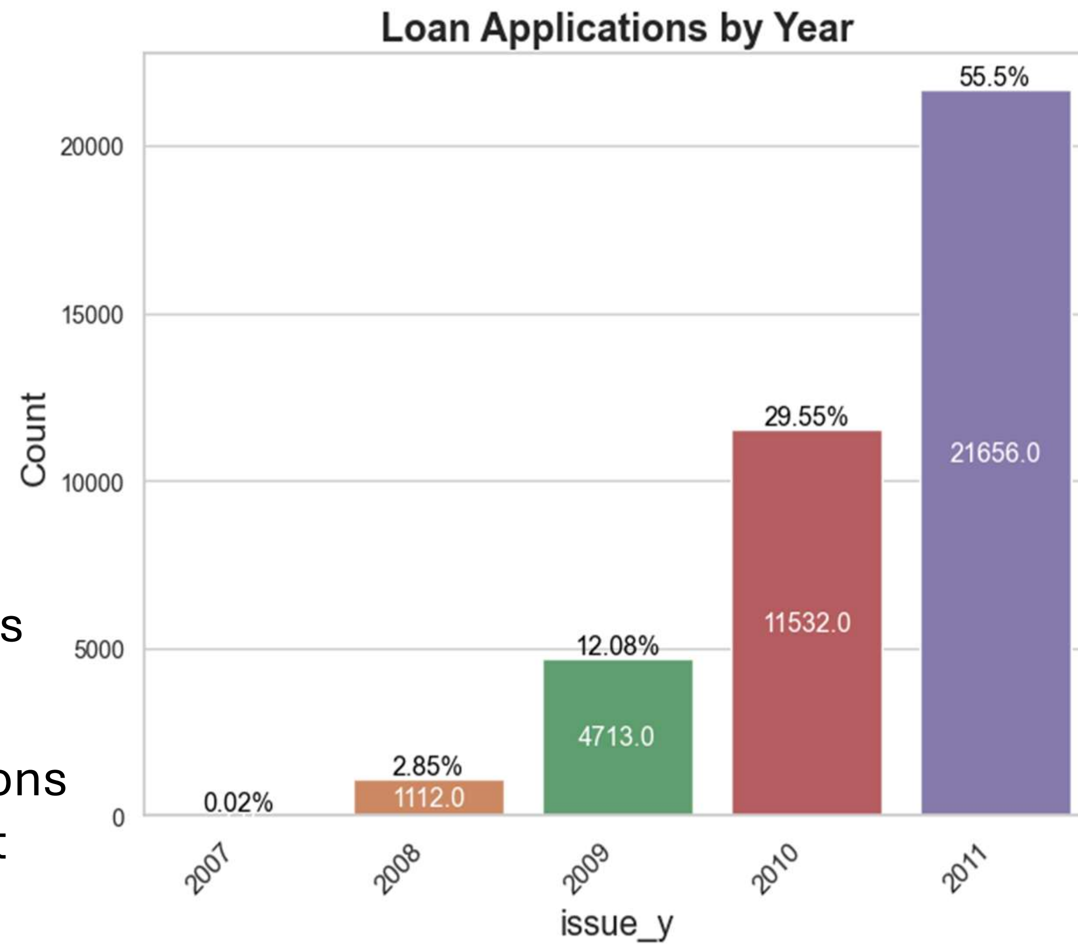
['Very Low', 'Low', 'Medium', 'High', 'Very High', 'Extremely High'].

Data Type Conversions

1. Converted to int64:
   - funded_amnt_inv, annual_inc, pub_rec_bankruptcies.
2. Term Column:
   - Removed "months" and converted to int.
3. Issue Date:
   - Converted issue_d to date format.
4. Interest Rate:
   - Removed "%" and converted to float.
5. Verification Status:
   - Replaced "Source Verified" with "Verified".

UNIVARIATE ANALYSIS

- Total Loan Application Analysis

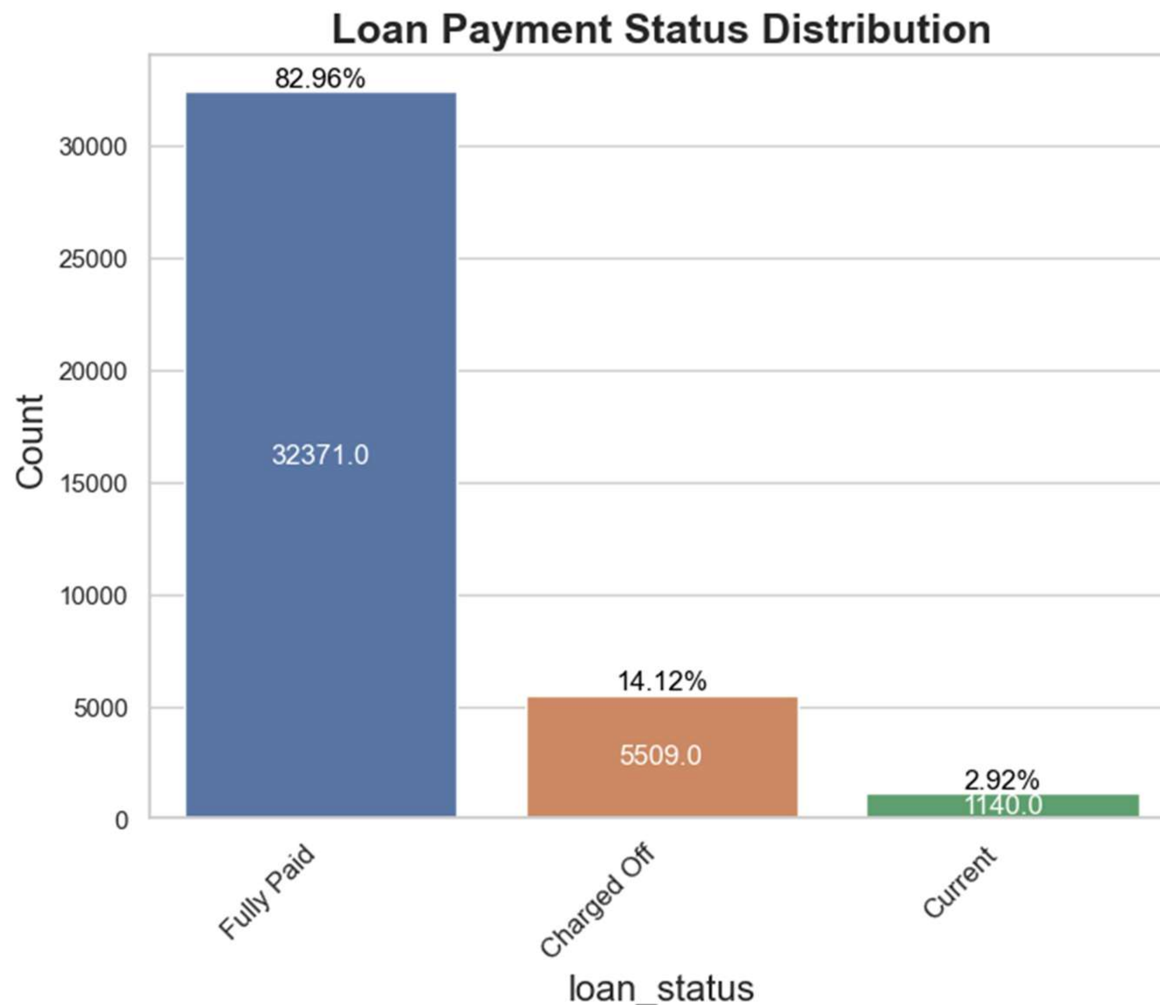


❖ Significant increase in loan applications from 2007 to 2011

❖ 2011 saw an 88% increase in applications compared to 2010, marking the highest growth during this period.

UNIVARIATE ANALYSIS

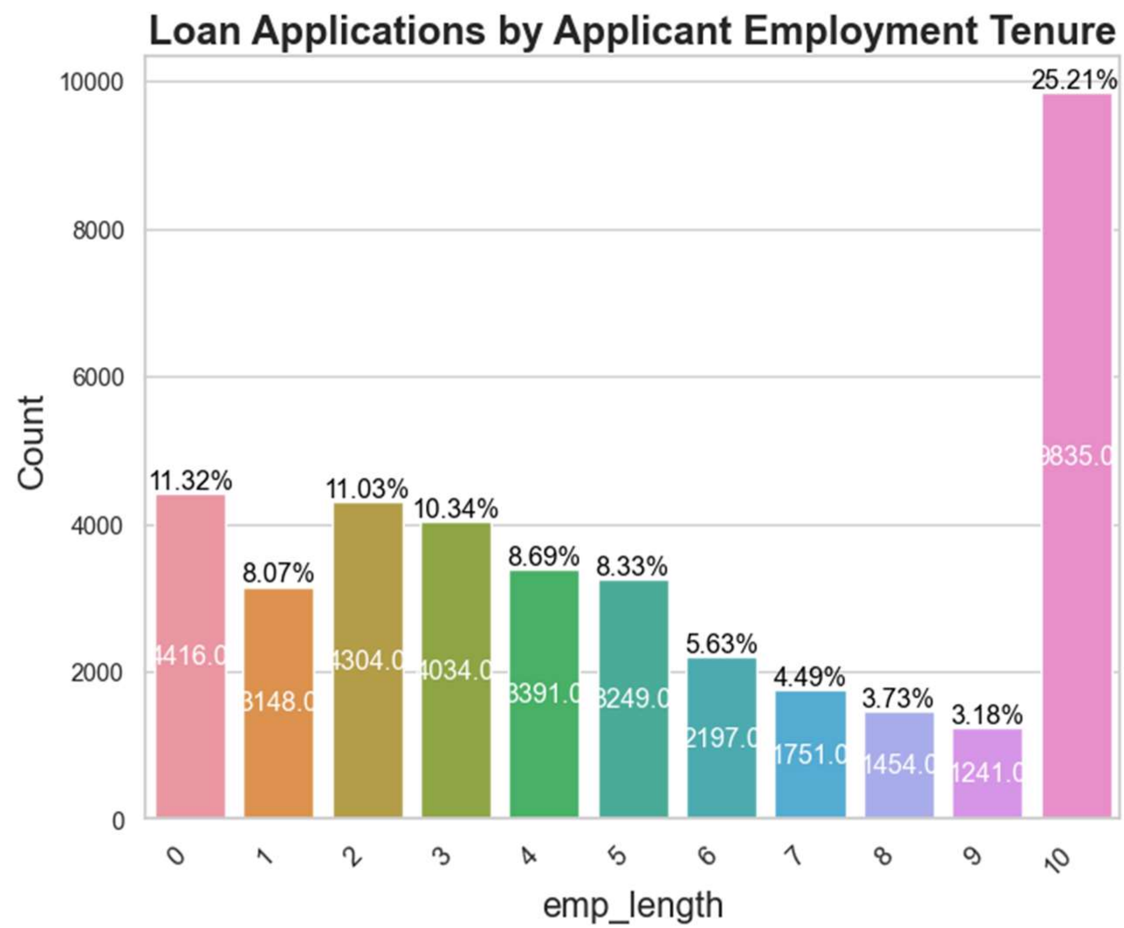
- Loan Payment Status



❖ Approximately **14%** of the total issued loans are defaulting.

UNIVARIATE ANALYSIS

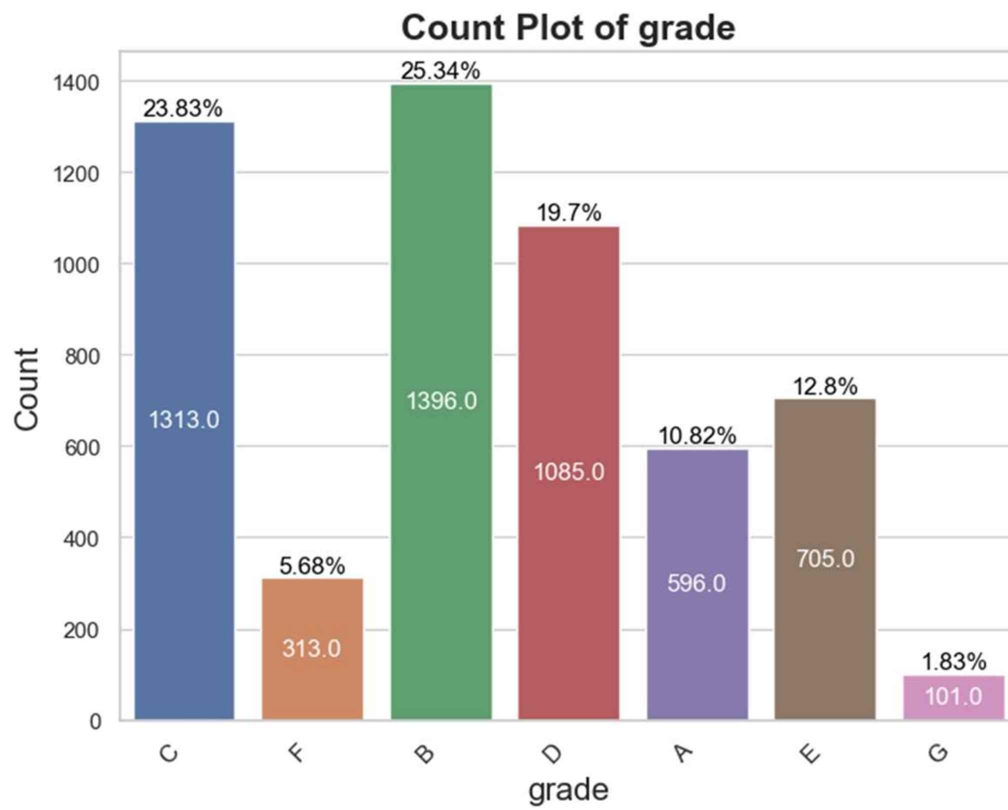
- Employment Tenure



❖ Approximately **25%** of loan applicants have **10+ years** of employment tenure.

UNIVARIATE ANALYSIS (charged off loans only)

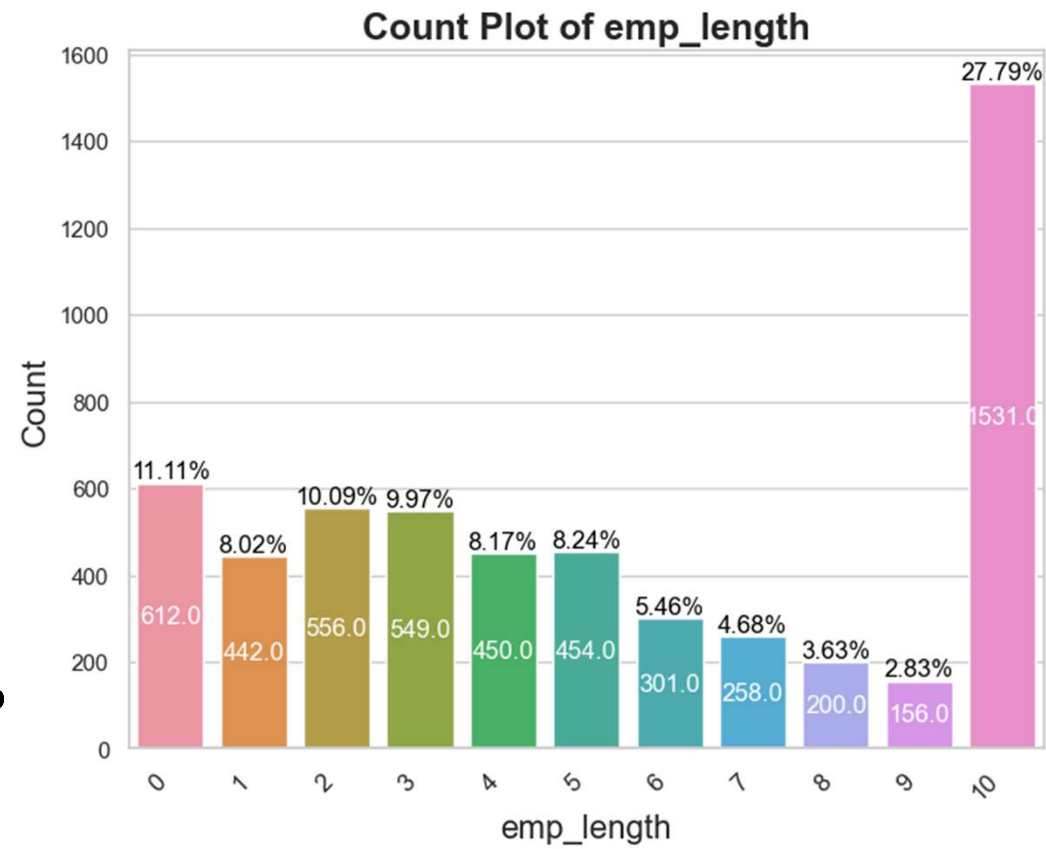
- Applicant's Grade



❖ Among the charged-off loan applicants, Grade 'B' stands out with the highest count, comprising 25.34% of the total at 1,396 applicants.

UNIVARIATE ANALYSIS (charged off loans only)

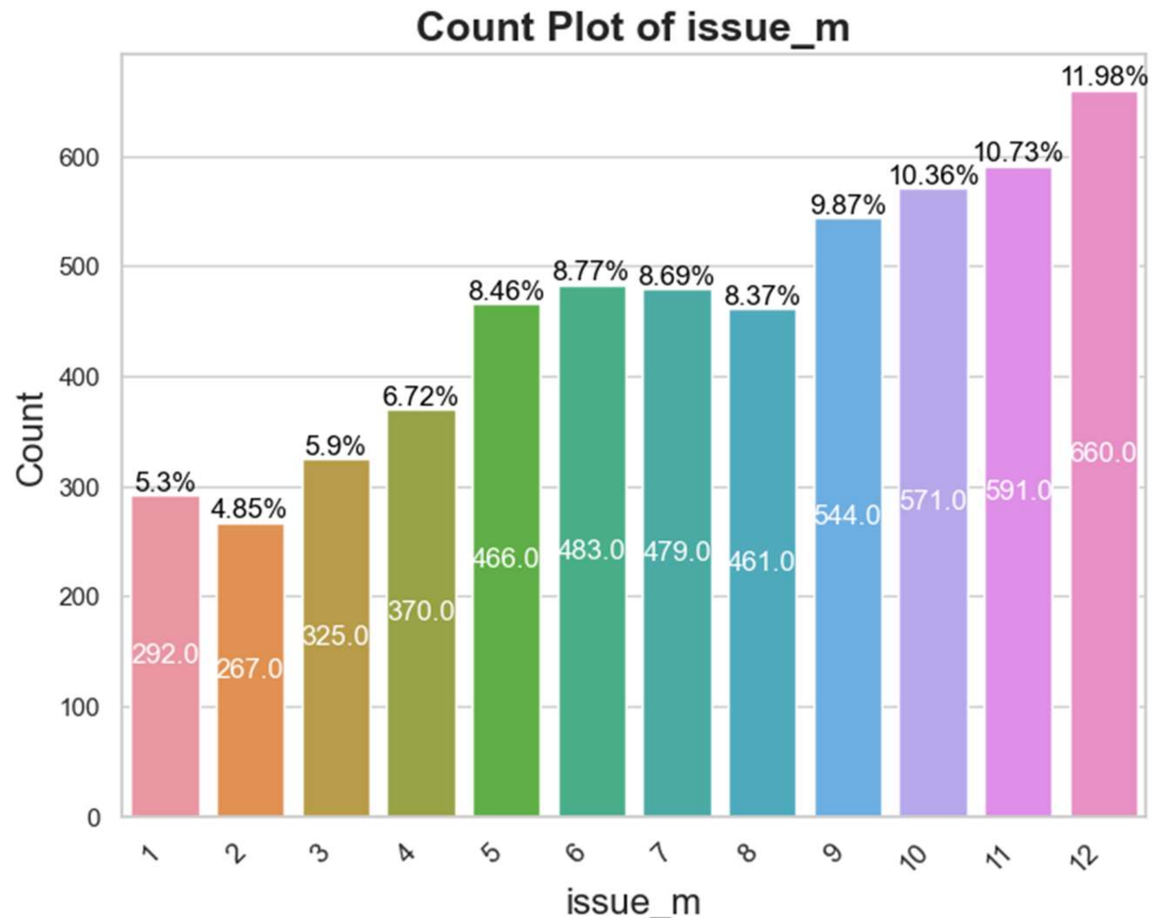
- Applicant's Employment Tenure



❖ Applicants with over 10 years of employment account for the largest share of charged-off loans, representing 27.79% or 1,431 cases.

UNIVARIATE ANALYSIS (charged off loans only)

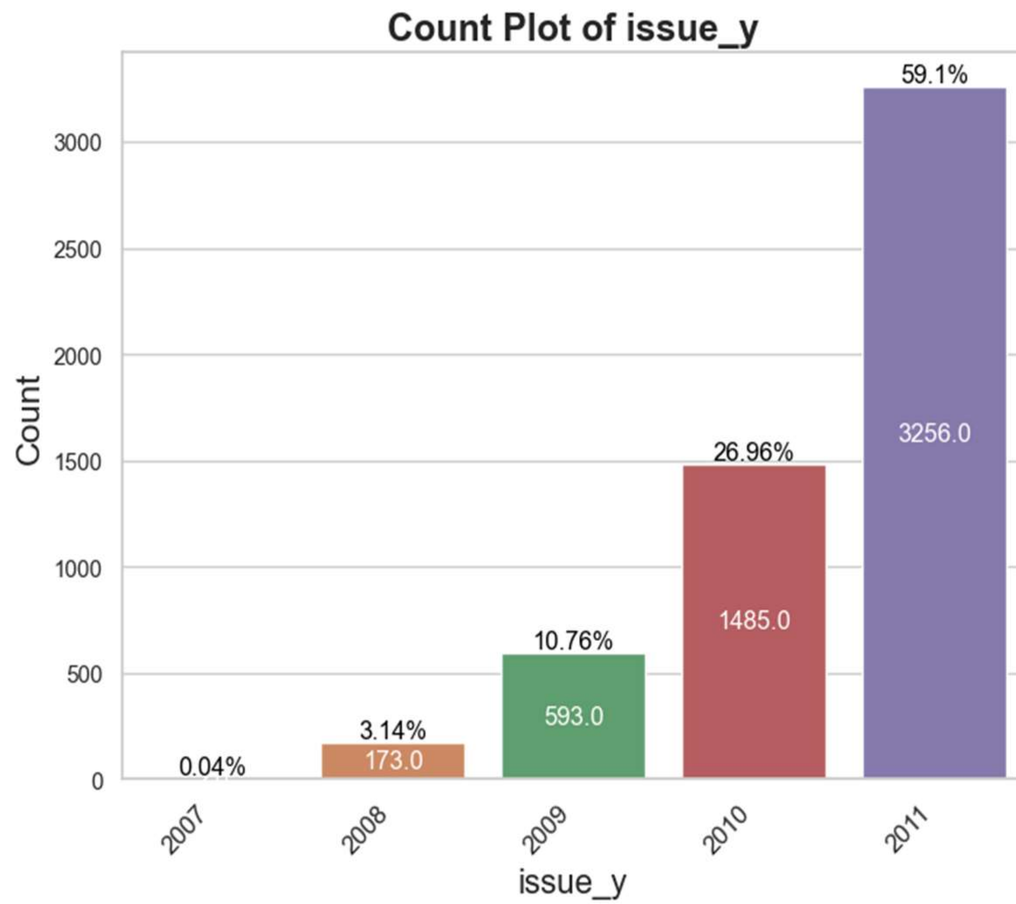
- Loan issue Month



❖ Majority of charged-off loans were originated in December

UNIVARIATE ANALYSIS (charged off loans only)

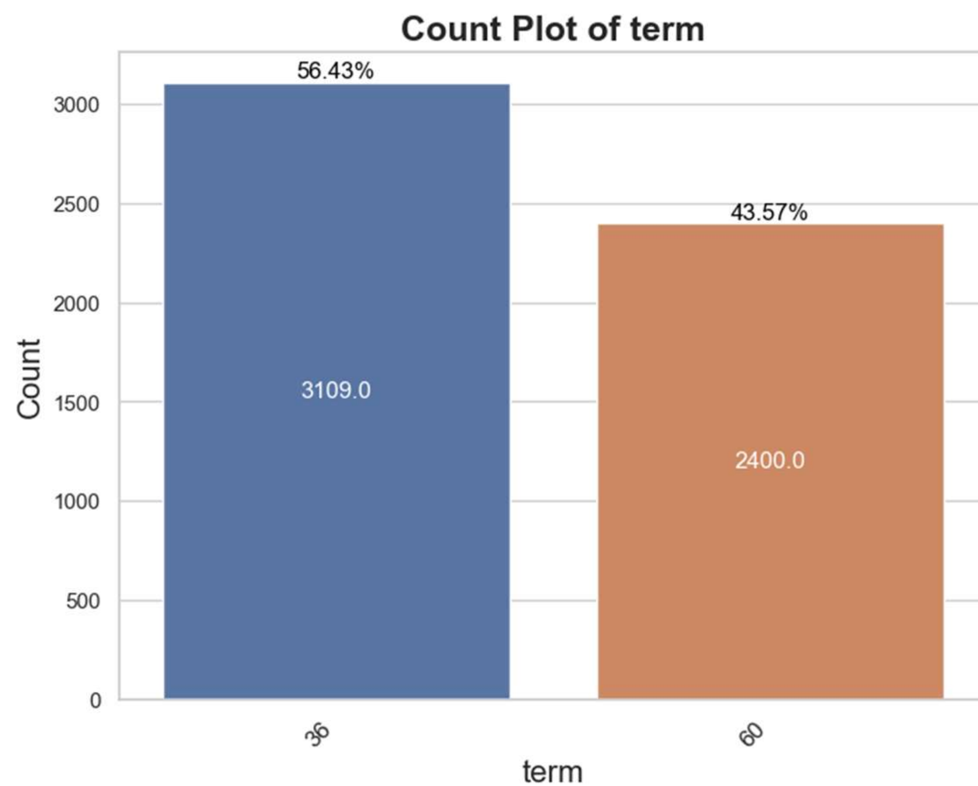
- Loan Issue Year



❖ 2011 marked a peak in charged-off loans, contributing 59.1% of the total with 3,256 applications.

UNIVARIATE ANALYSIS (charged off loans only)

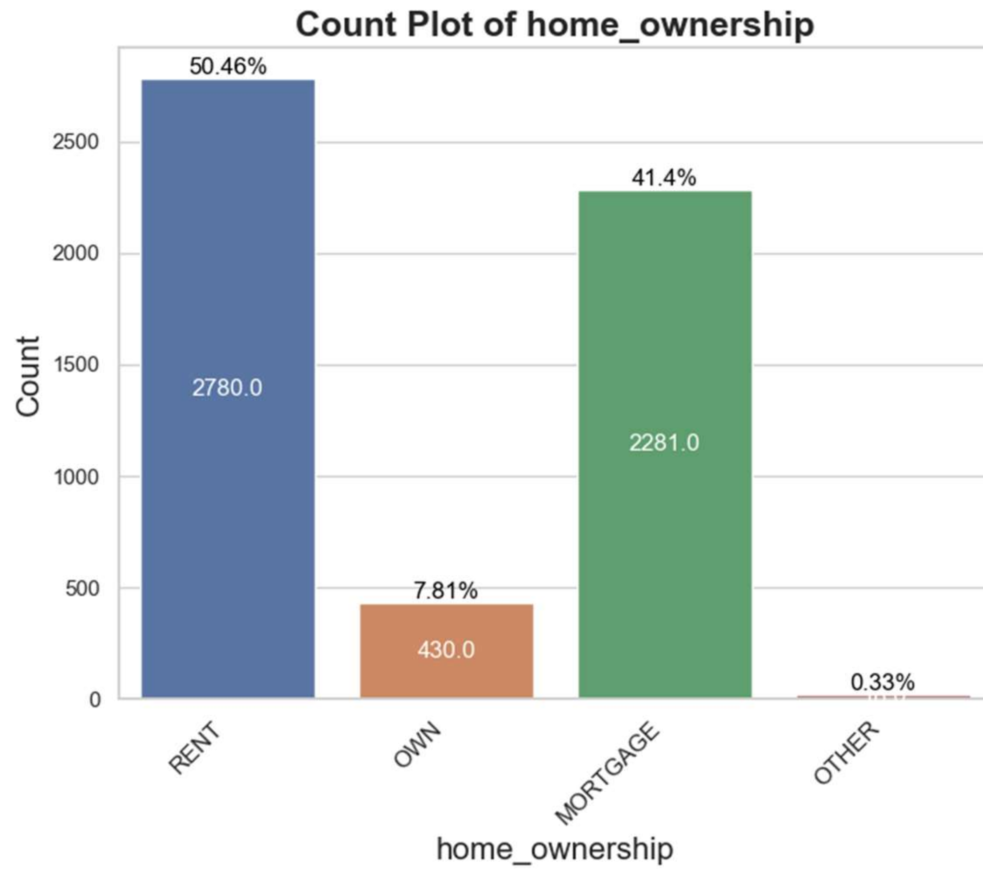
- Loan Term



❖ Short-term loans with a duration of 36 months were the most popular among Charged Off applicants.

UNIVARIATE ANALYSIS (charged off loans only)

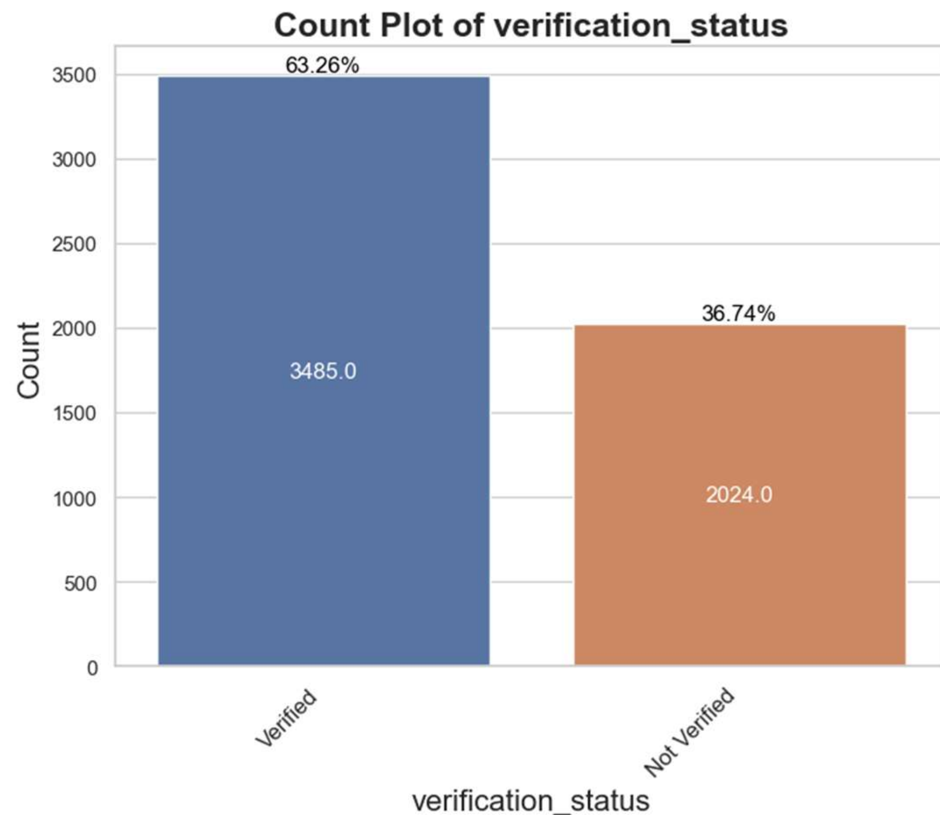
- Applicant's Home Ownership



❖ 50% of charged-off loan participants, totaling 2,780 individuals, reside in rented accommodations

UNIVARIATE ANALYSIS (charged off loans only)

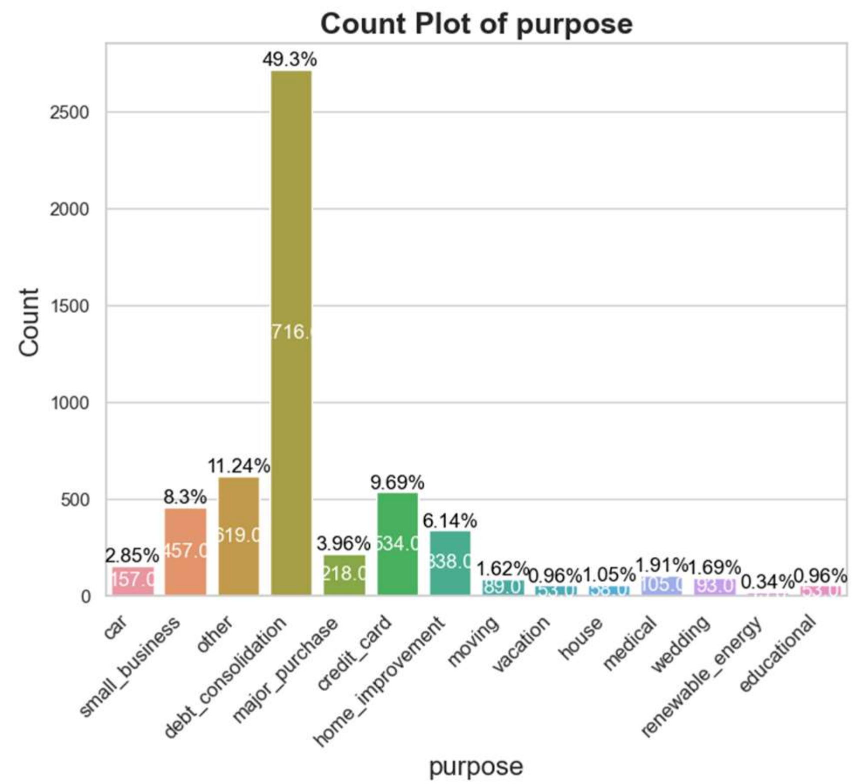
- Applicant's Income Source Verification



❖ 63% of defaulters have verified income sources, heightening the risk of loan defaults

UNIVARIATE ANALYSIS (charged off loans only)

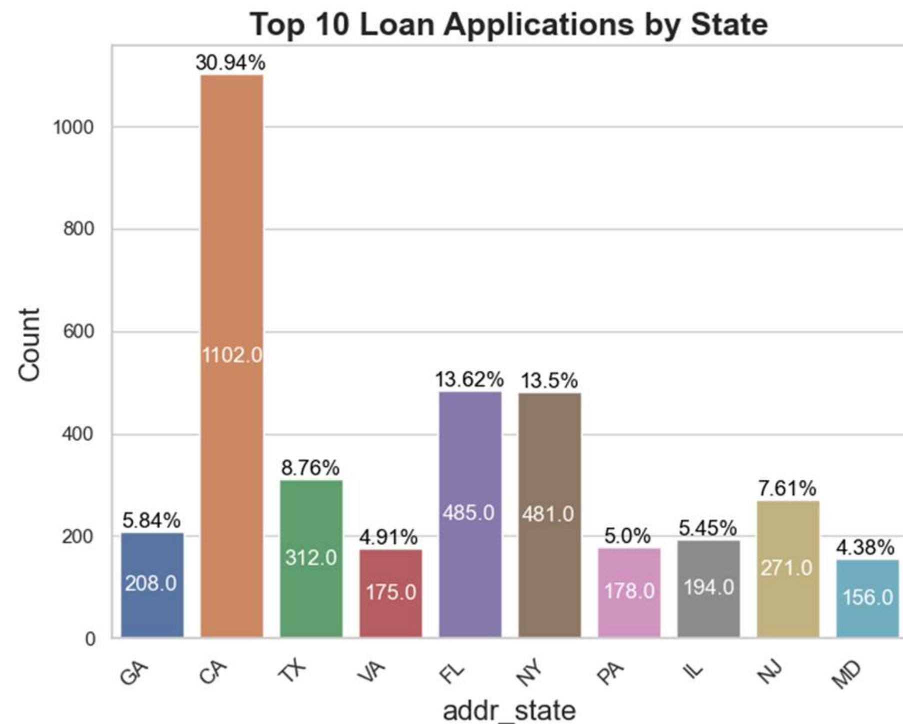
- Loan Purpose



❖ 49.3% of charged-off loan applicants indicated that debt consolidation was their primary loan purpose.

UNIVARIATE ANALYSIS (charged off loans only)

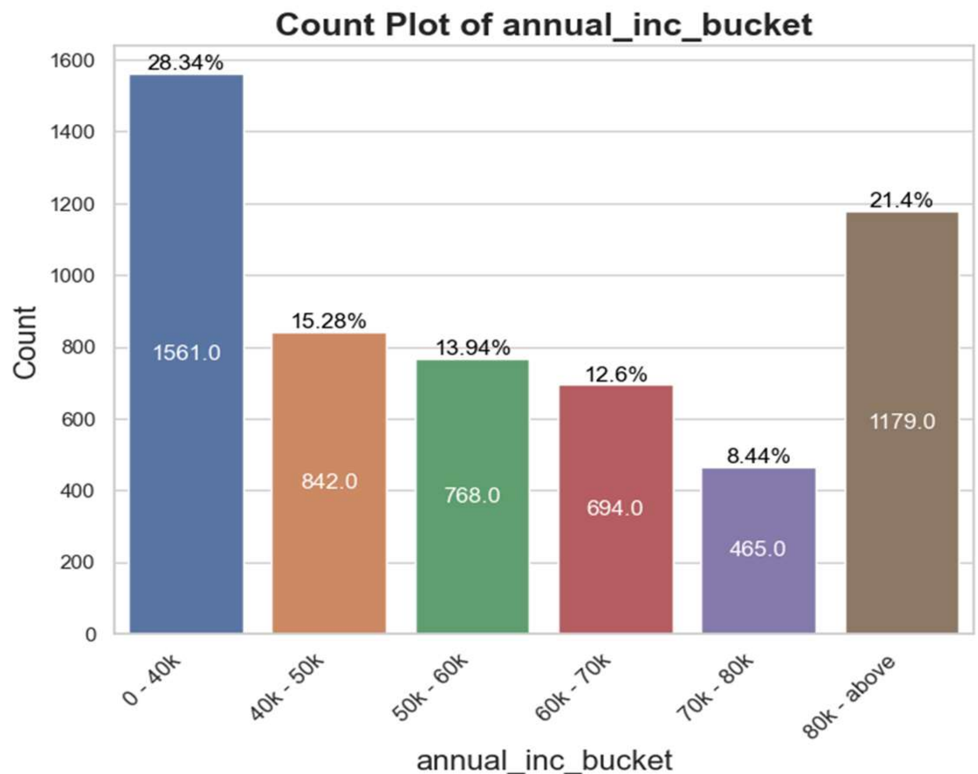
- Applicant's State of Residence



❖ California stands out with the highest number of charged-off loan applicants, totaling 1,102 individuals

UNIVARIATE ANALYSIS (charged off loans only)

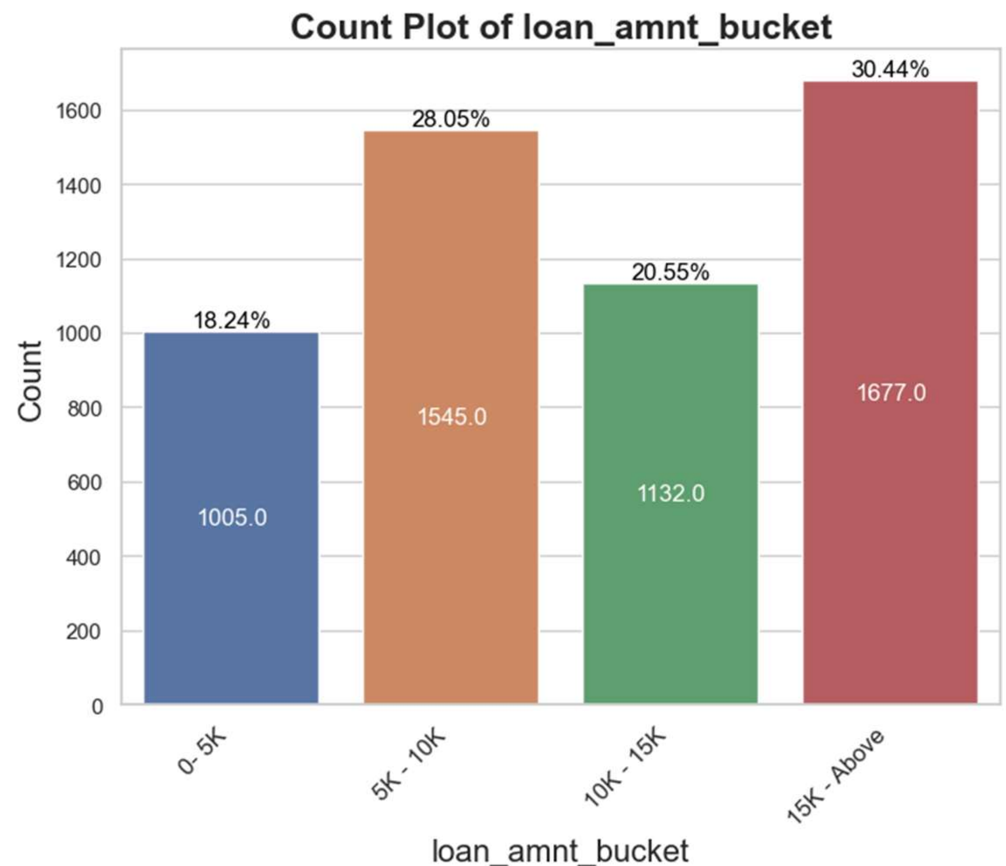
- Applicant's Annual Income



❖ 28.34% of loan applicants who defaulted had annual salaries below 40,000

UNIVARIATE ANALYSIS (charged off loans only)

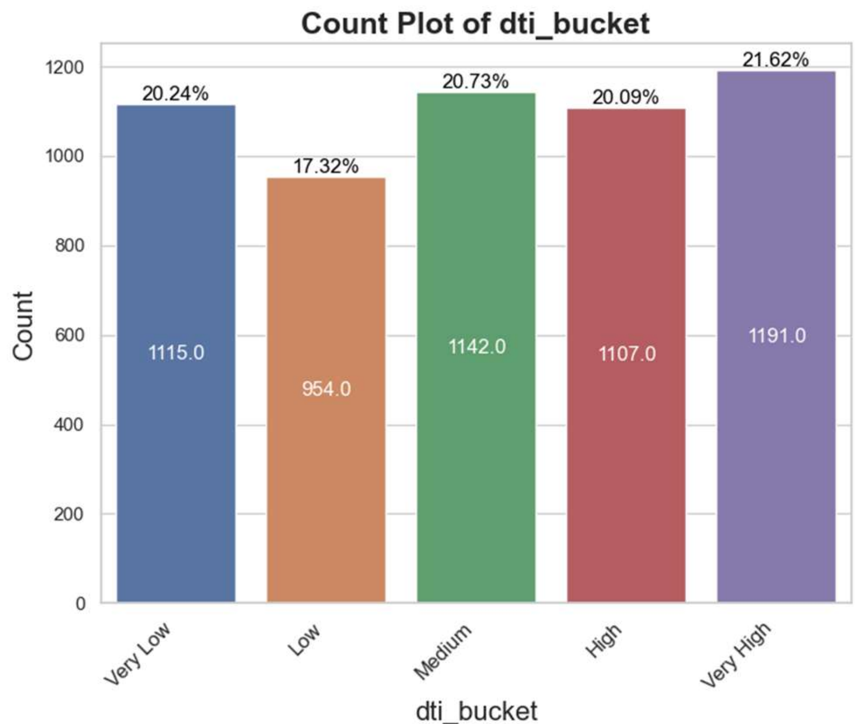
- Loan Amount



❖ 30.44% of charged-off loan applicants, totaling 1,677 individuals, received loan amounts of 15,000 or more

UNIVARIATE ANALYSIS (charged off loans only)

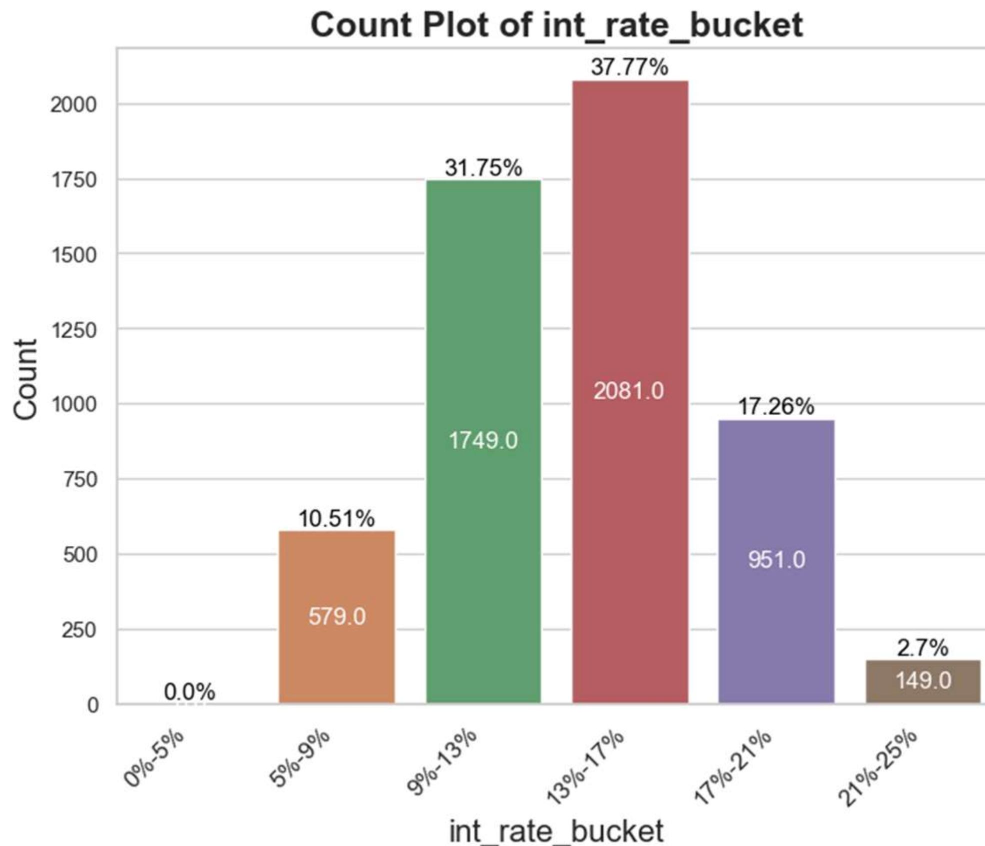
- Debt to Income Ratio



❖ Around 21.62 % a loan participants who charged off, had very high debt-to-income ratio

UNIVARIATE ANALYSIS (charged off loans only)

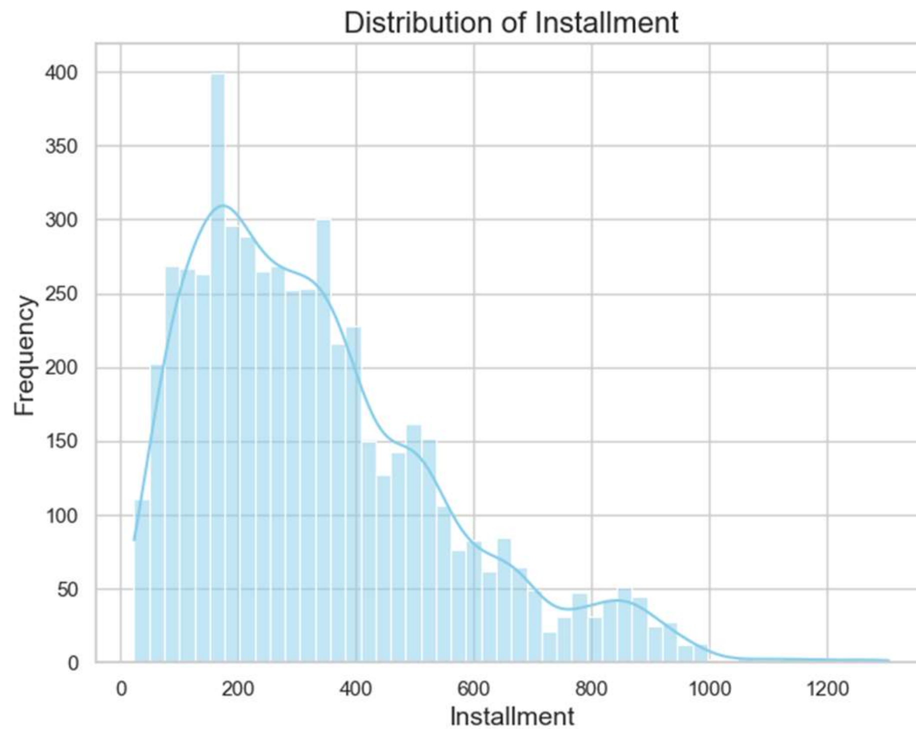
- Interest Rate



❖ Around 37% of loan participants who defaulted, totaling 2,081 individuals, had interest rates between 13% and 17%

UNIVARIATE ANALYSIS (charged off loans only)

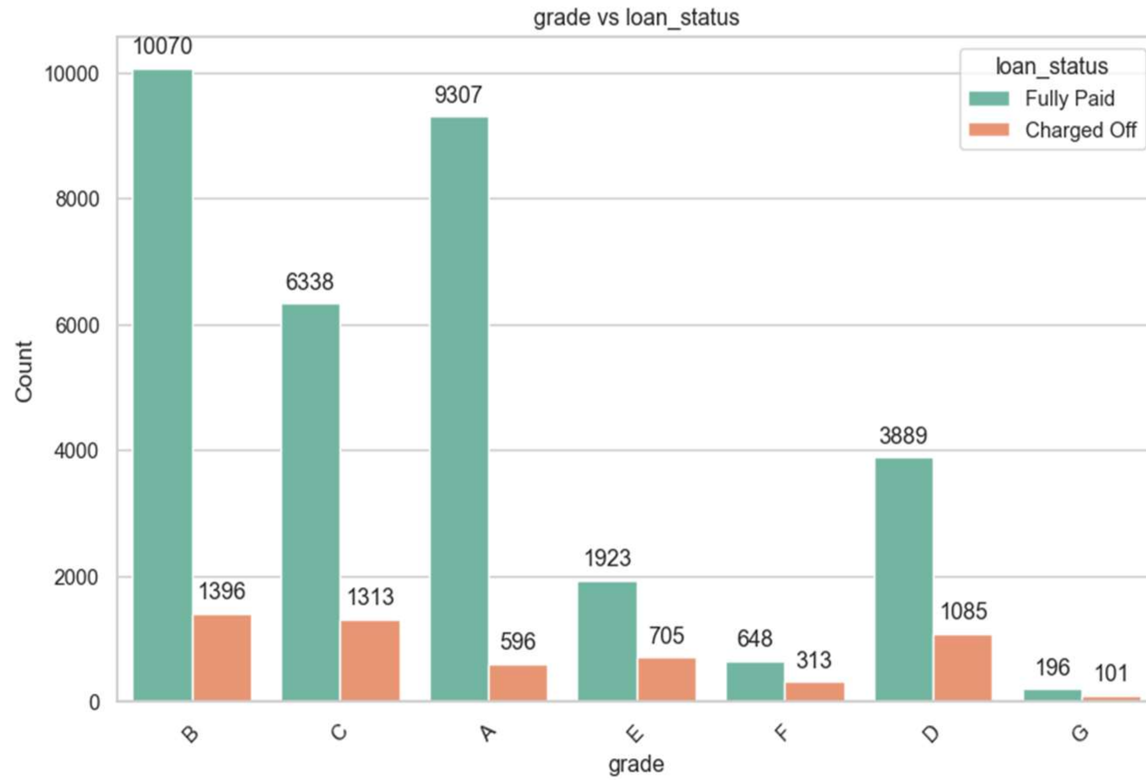
- Installment Amount



❖ Most charged-off loan participants had monthly installments ranging from 180 to 420

BIVARIATE ANALYSIS

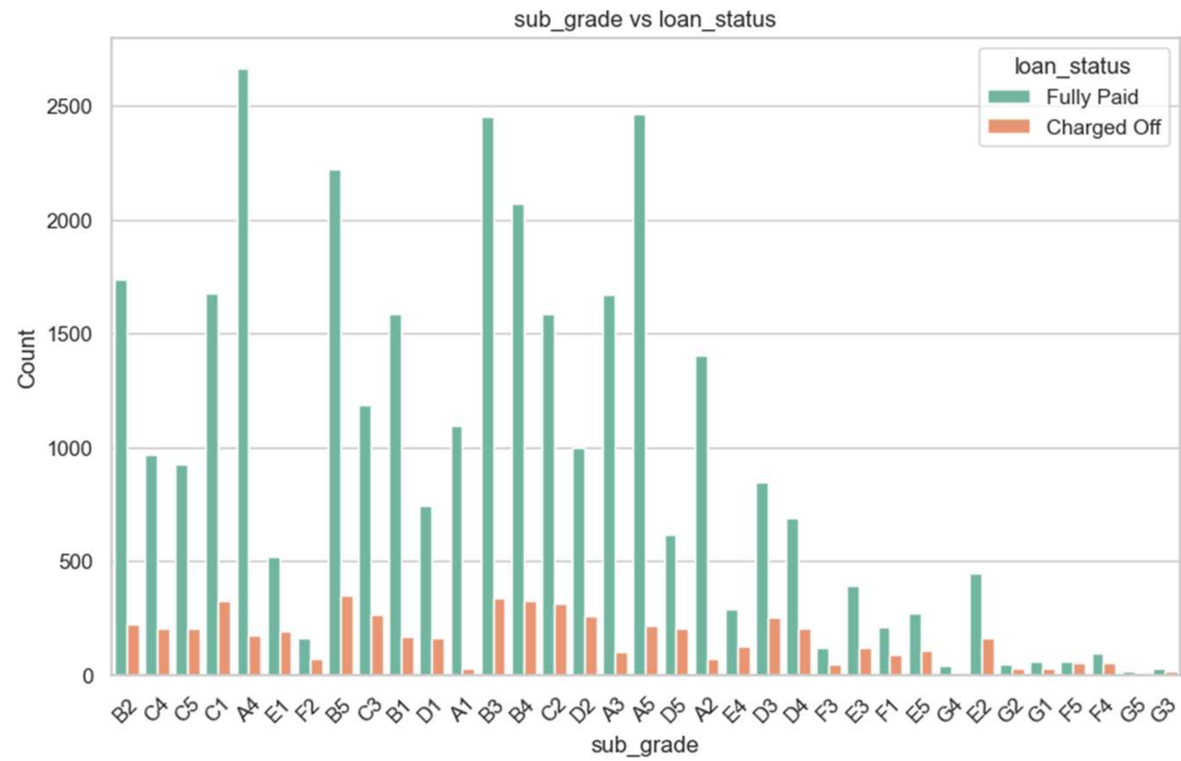
- Applicant's Grade vs Loan Status



❖ The majority of "Charged Off" loans are attributed to loan applicants belonging to Grades B, C, and D.

BIVARIATE ANALYSIS

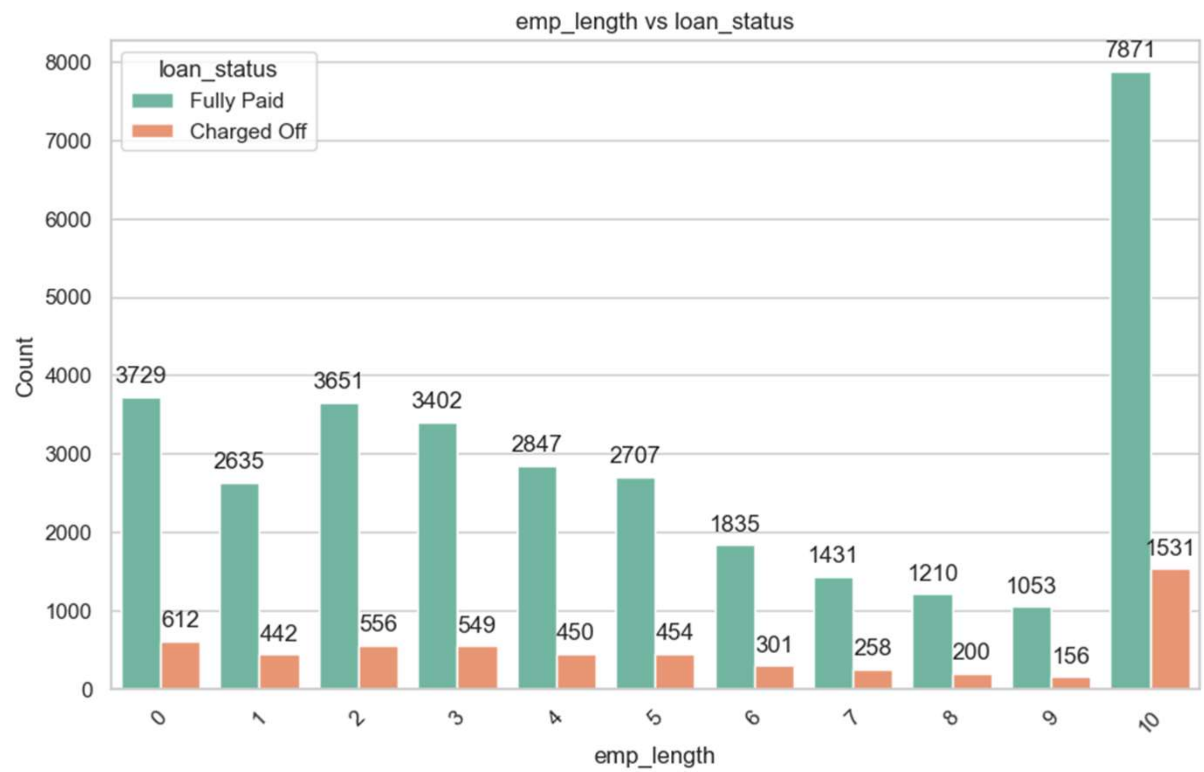
- Applicant's Sub Grade vs Loan Status



❖ Within the B grade category, Sub-Grades B3, B4, and B5 show a higher likelihood of "Charged Off" loans.

BIVARIATE ANALYSIS

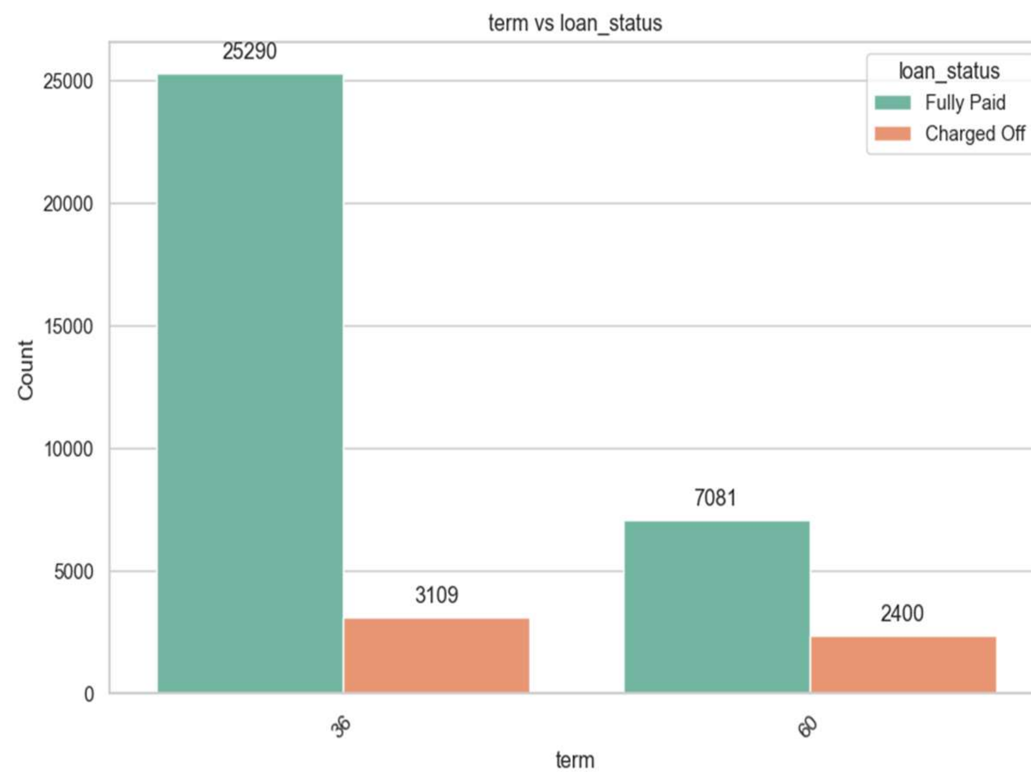
- Applicant's Employment Tenure vs Loan Status



❖ A large proportion of loan applicants have 10 or more years of experience, and interestingly, this group also has the highest rate of loan defaults.

BIVARIATE ANALYSIS

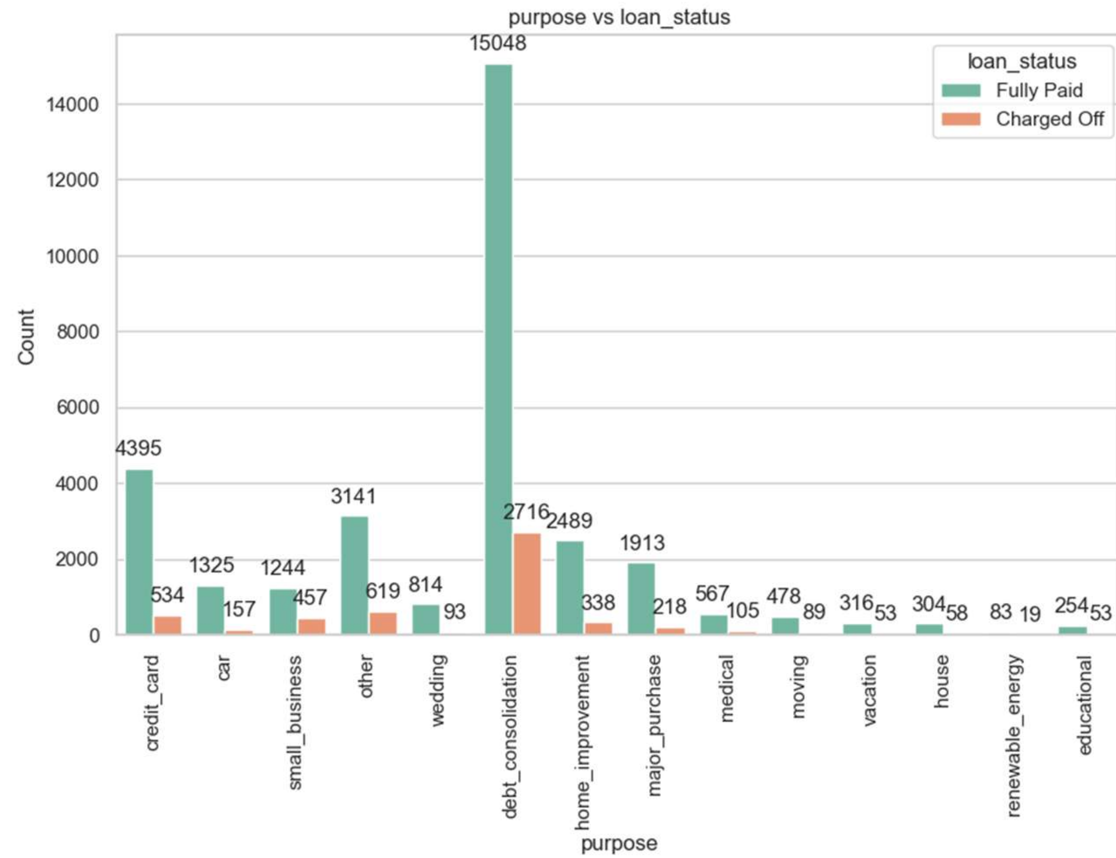
- Loan Term vs Loan Status



❖ Loan applicants opting for 60-month loan terms have a significantly higher likelihood of defaulting compared to those choosing 36-month terms.

BIVARIATE ANALYSIS

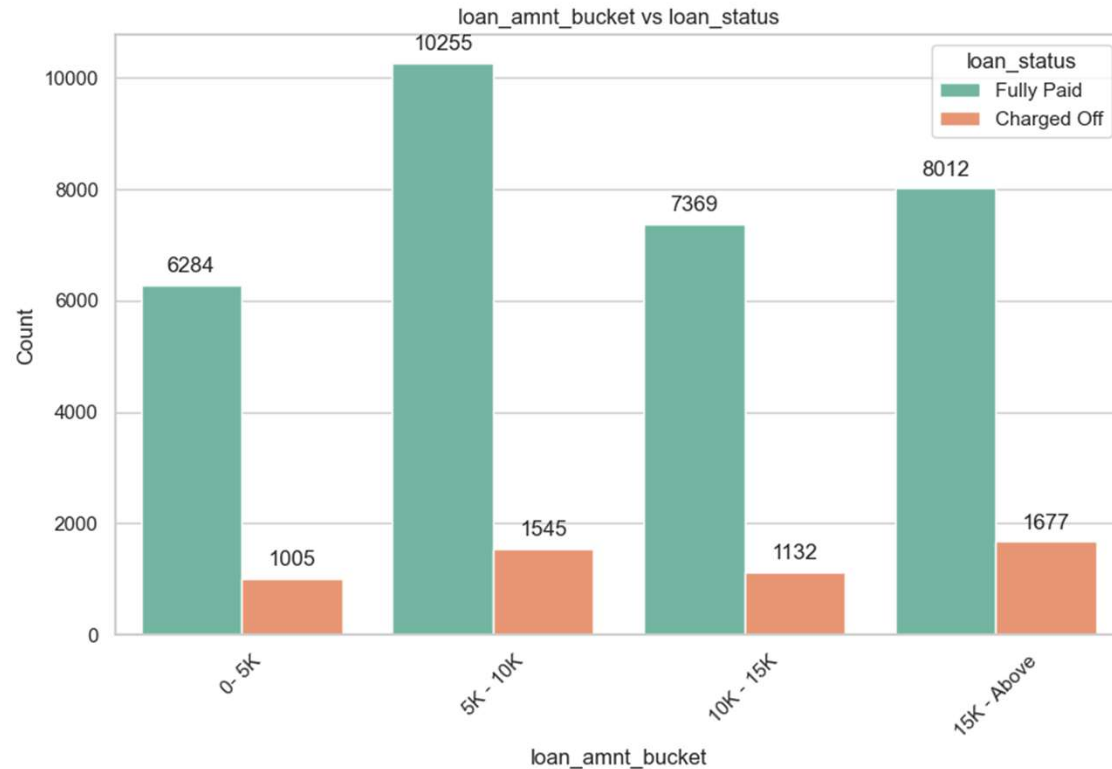
- Loan Purpose vs Loan Status



❖ Debt consolidation loans have the most borrowers and the highest default rates.

BIVARIATE ANALYSIS

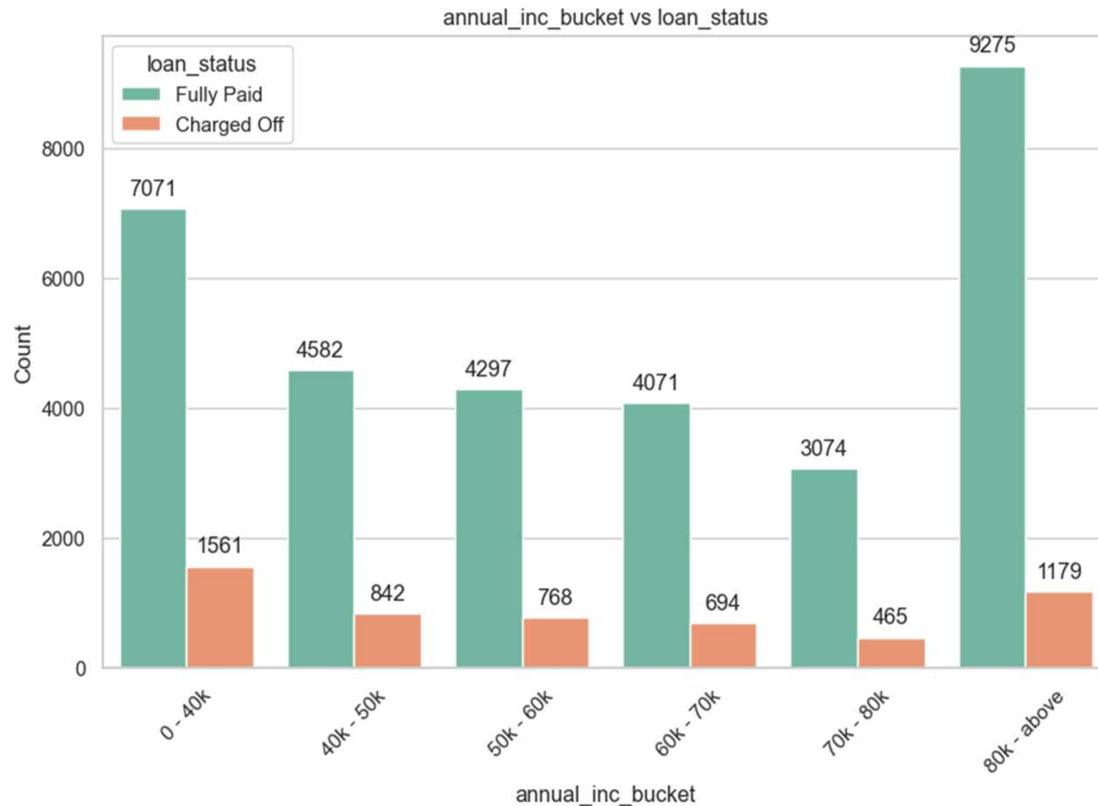
- Loan Amount vs Loan Status



❖ A large number of applicants who defaulted took loans of 15,000 or more.

BIVARIATE ANALYSIS

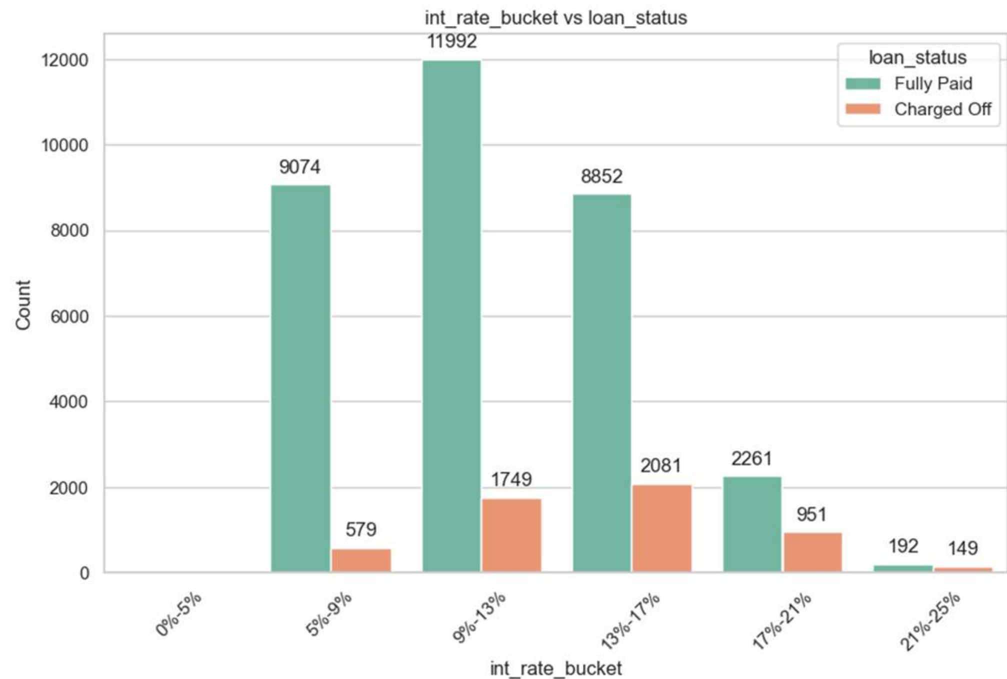
- Applicant's Annual Income vs Loan Status



❖ Most applicants who defaulted on their loans reported an annual income below 40,000.

BIVARIATE ANALYSIS

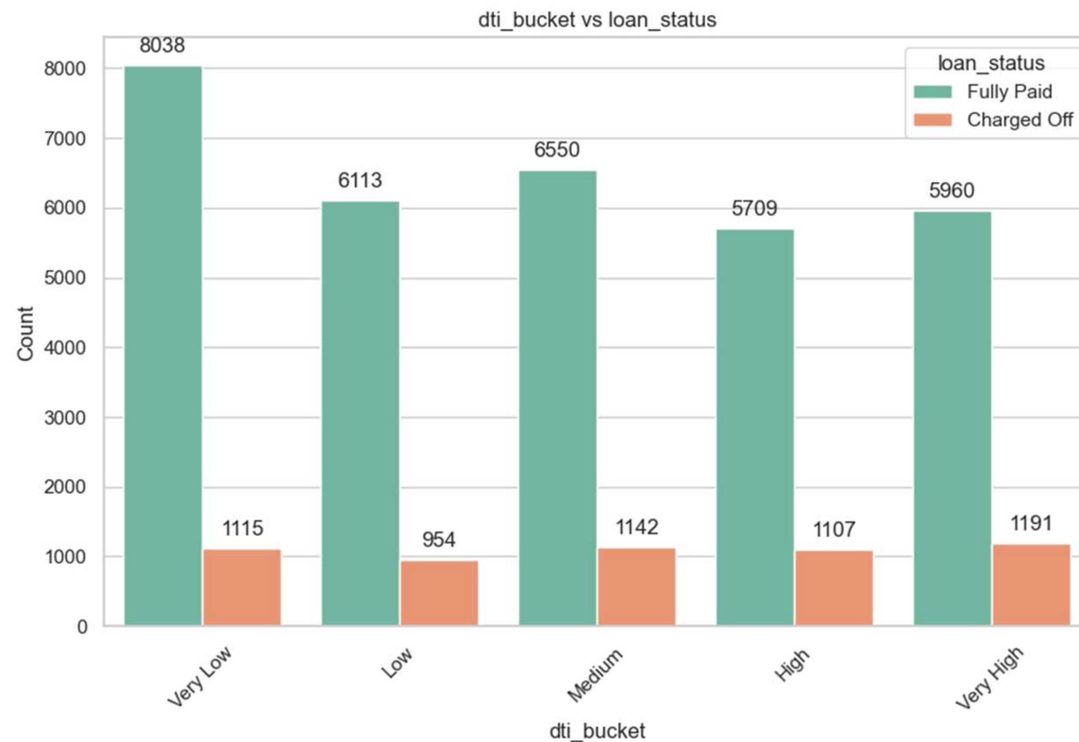
- Interest Rate vs Loan Status



❖ Many defaulted borrowers had loans with interest rates between 13% and 17%.

BIVARIATE ANALYSIS

- Debt to income Ratio vs Loan Status



❖ Borrowers with significantly high Debt-to-Income (DTI) ratios were more likely to default.

Multivariate ANALYSIS

- Loan Amount & Installment: Strong positive correlation; higher loan amounts lead to increased installment payments.
- Loan Term & Interest Rate: Strong positive correlation; longer loan terms are often linked to higher interest rates.
- Annual Income & DTI Ratio: Negative correlation; higher incomes typically result in lower DTI ratios.

Findings

- Majority of "Charged Off" loans are attributed to loan applicants belonging to Grades B, C, and D. These grades represent a significant risk for loan defaults, suggesting that borrowers with lower creditworthiness (as indicated by their grade) are more likely to default.

- Within the B grade category, Sub-Grades B3, B4, and B5 show a higher likelihood of "Charged Off" loans. This indicates that even small variations within the same grade can reflect different levels of risk, with B3, B4, and B5 applicants being more probable to default compared to others in the B grade.

- A large proportion of loan applicants have 10 or more years of experience, and interestingly, this group also has the highest rate of loan defaults. This trend suggests that long-term work experience does not necessarily correlate with lower default risk, possibly due to other factors like loan size or financial commitments.

- The month of December emerges as the most popular time for taking loans, likely driven by the holiday season and increased consumer spending. However, this trend could also be linked to a higher risk of defaults.

- The highest volume of loan applications occurred in 2011, and correspondingly, 2011 also saw the highest number of "Charged Off" loans. This could be reflective of Global economic crisis prevailed at that time, with a surge in loan demand followed by higher-than-average defaults during that period.

- Loan applicants opting for 60-month loan terms have a significantly higher likelihood of defaulting compared to those choosing 36-month terms. Longer loan terms may result in financial strain over time, leading to a higher chance of non-repayment..

Findings

- Applicants who live in rented or mortgaged homes are more likely to default on loans. This insight can help the company consider housing stability when assessing someone's ability to repay.

- Verified applicants are defaulting more than those who aren't verified. The company should revisit its verification process to ensure it's accurately assessing creditworthiness and look for ways to improve it.

- Debt consolidation loans have the most borrowers and the highest default rates. The company should be more cautious when approving these loans.

- Borrowers from California, Florida, and New York are more likely to default. The company should keep an eye on these states and consider adjusting lending strategies or rates to reduce risk.

- Most applicants who defaulted on their loans reported an annual income below 40,000. This indicates that lower-income borrowers may struggle with repayment, and income level could be a key factor in assessing risk.

- A large number of applicants who defaulted took loans of 15,000 or more. This suggests that higher loan amounts could be riskier, and the company may want to adjust its approval criteria or interest rates for large loans.

- Many defaulted borrowers had loans with interest rates between 13% and 17%. This suggests that loans in this rate range might pose higher risks, and the company could explore offering lower rates for better-qualified borrowers or tightening approval for those in this range.

- Borrowers with significantly high Debt-to-Income (DTI) ratios were more likely to default. This shows that DTI is a strong indicator of potential default, and the company should consider more stringent DTI limits to reduce default risk.

Recommendations

- ✓ Enhance Verification: Strengthen verification processes to ensure accurate creditworthiness assessment.
- ✓ Improve Credit Scoring: Make grading criteria to better identify risk within grades (e.g., B3, B4, B5).
- ✓ Stricter Loan Approval: Tighten criteria for applicants with incomes below $40,000 and for loans over $15,000.
- ✓ Adjust Interest Rates: Lower interest rates for responsible borrowers and reconsider rates for debt consolidation loans.
- ✓ Consider Housing Status: Assess housing status, as renters may have higher default risk.
- ✓ Monitor Employment Trends: Check why long-term employees are defaulting and adjust criteria accordingly.
- ✓ Implement DTI Limits: Set stricter DTI ratio thresholds to identify higher-risk borrowers.