LENDING CASE STUDY

BY SHWETA SONI

Business Objective Summary

Business Challenge

A consumer finance company needs to assess loan applicants and make approval decisions while balancing two key risks:

1. **Revenue Loss** – Rejecting creditworthy applicants who could have repaid the loan.
2. **Financial Loss** – Approving high-risk applicants who are likely to default.

Objective

- Identify key factors that indicate loan default risk using **Exploratory Data Analysis (EDA)**.
- Understand patterns in borrower attributes, loan characteristics, and financial history.
- Use insights to **reduce credit loss**, optimize lending policies, and improve risk assessment.

Key Decision Scenarios

- **Fully Paid:** Borrower successfully repaid the loan.
- **Current:** Loan is active with no defaults yet.
- **Charged-Off:** Borrower has defaulted.
- **Rejected:** No transactional data available as the loan was denied.

DATA PREPARATION AND ANALYSIS

➢ **META-DATA ANALYSIS :**

•All columns with **completely null values** were identified and removed.

•Remaining columns with partial null values were inspected individually to either:

- •Remove rows with nulls, or
- •Impute values for missing data

➢ **Feature Engineering**

•New features (e.g., **ratios**) were created to enhance analysis and derive actionable insights.

•**Exploratory Data Analysis (EDA)**

•**Single-variable analysis** identified patterns within individual attributes.

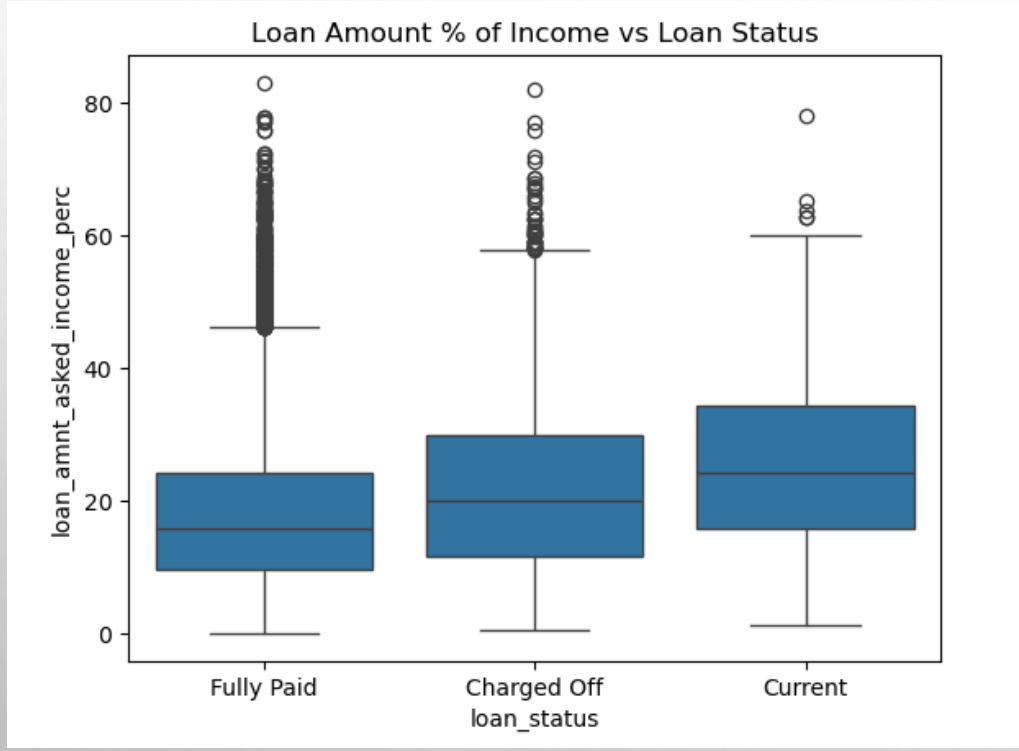
•**Multivariate analysis** explored relationships between key variables.

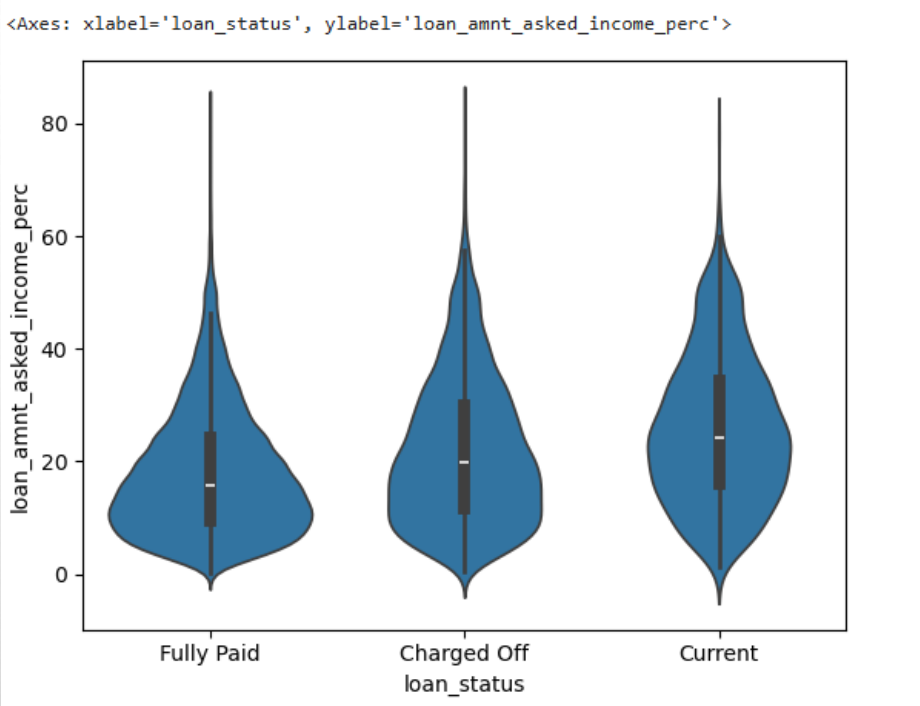
➢ **Key Insights**

Inferences were drawn to pinpoint factors contributing to loan defaults, guiding recommendations for mitigating credit risk.

SINGLE VARIABLE ANALYSIS

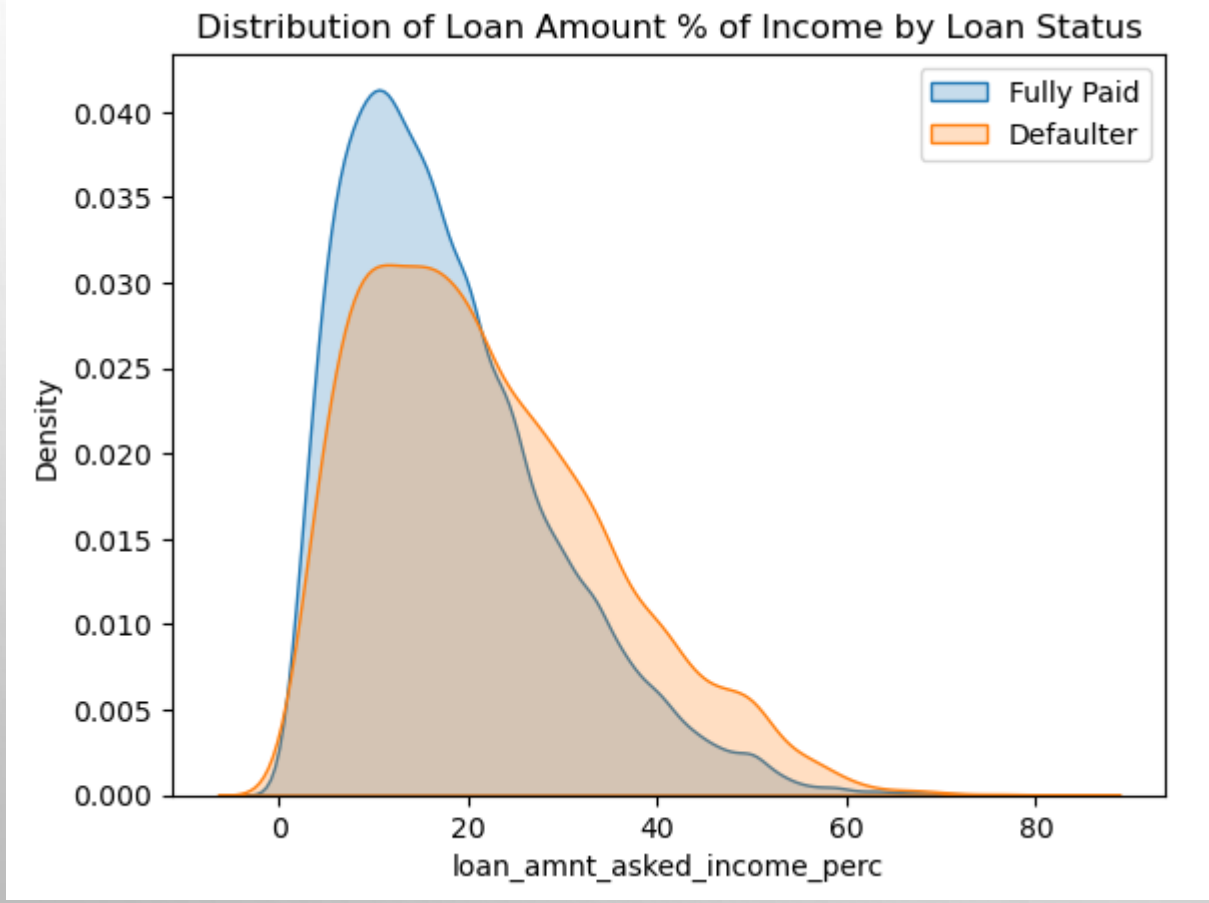
NEW FEATURE – LOAN AMOUNT AS PERCENTAGE OF INCOME





We can see that the fully paid loans are the ones with lesser loan amount per income percentage.
Loans with default status are with higher percentage.

MORE INSIGHTS…

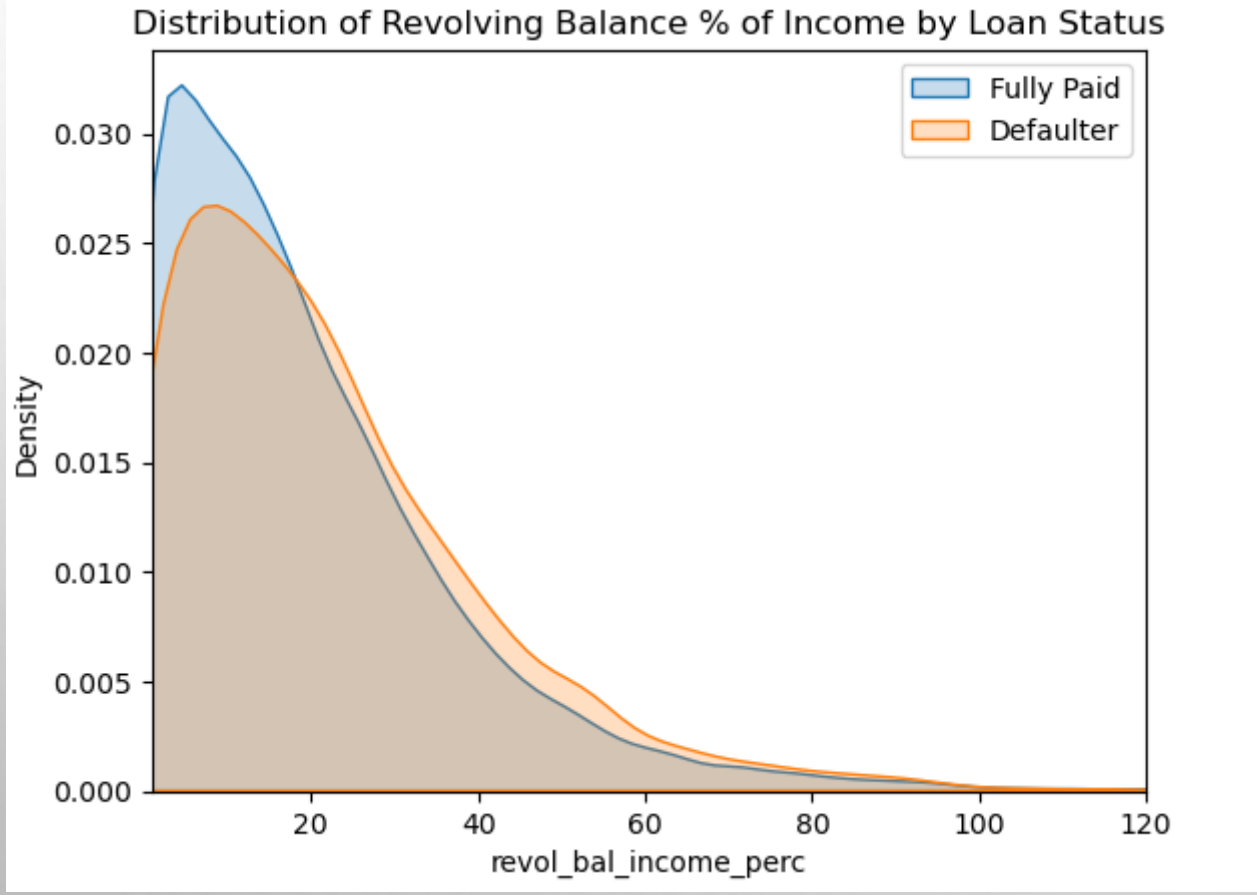


The following are observed :

i) Loan applicants asking for upto ~25 percent of their income are likely to pay off their debt

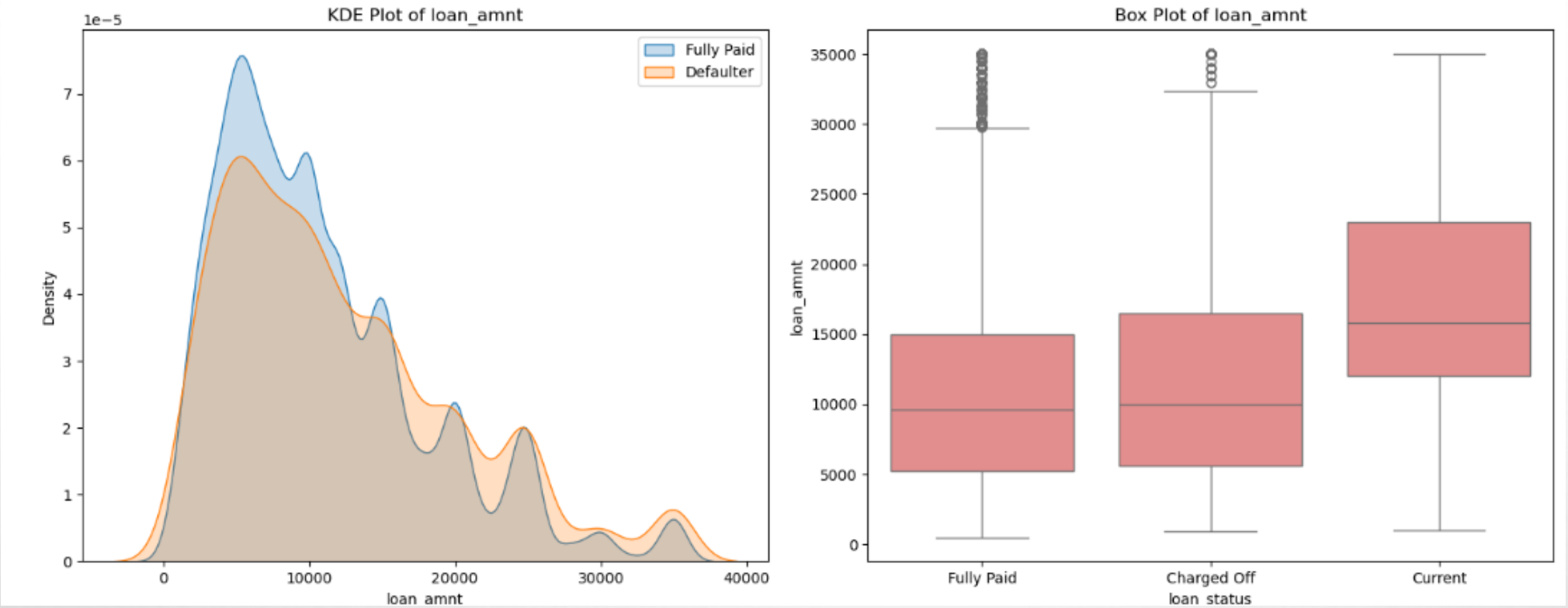
ii) Once loan applicants apply for loan that is more than ~25 percent percentage of income, they are more likely to default than pay off

NEW FEATURE - REVOLVING BALANCE % OF INCOME



We interpret that customers with revolving balance not more than 20 percent of their income are more likely to pay off the loan

SINGLE VARIABLE ANALYSIS FOR NUMERICAL COLUMNS



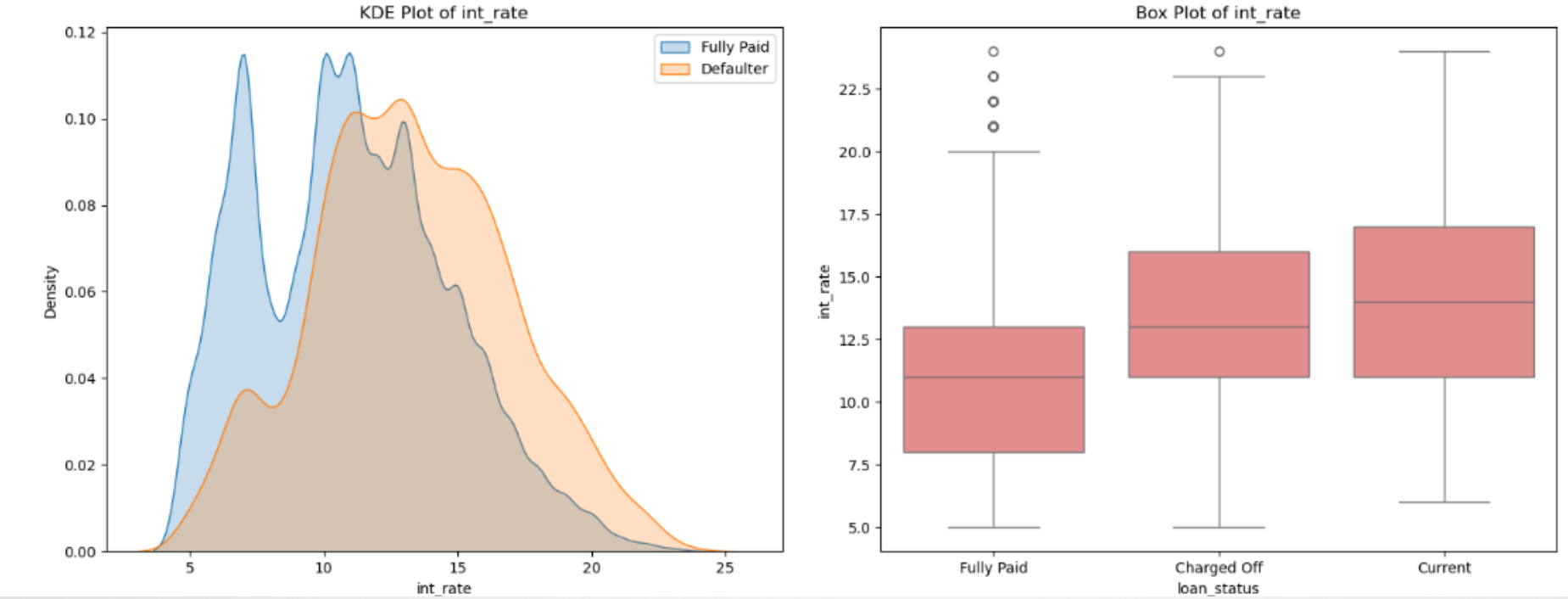
Parameter - Loan Amount

Insights from KDE Plot (Left Chart):

- **For Fully Paid Loans -** Fully paid loans are more concentrated around lower loan amounts (around $10,000), indicating that smaller loans have a higher likelihood of being successfully repaid.
- **For Defaults -** Defaulted loans tend to have a relatively higher proportion in the mid-range amounts (15,000–25,000), suggesting that larger loan sizes might increase the risk of default.

Insights from Box Plot (Right art):

- **For Fully Paid Loans -** Fully paid loans have a lower median loan amount compared to defaults, with fewer outliers, indicating that managing smaller loans is easier for borrowers.
- **For Defaults -** Defaulted loans (charged off) exhibit a wider spread with more high-value outliers, suggesting that higher loan amounts contribute to repayment difficulties and an increased risk further analysis!



Parameter - Interest Rate

Insights from KDE Plot (Left Chart):

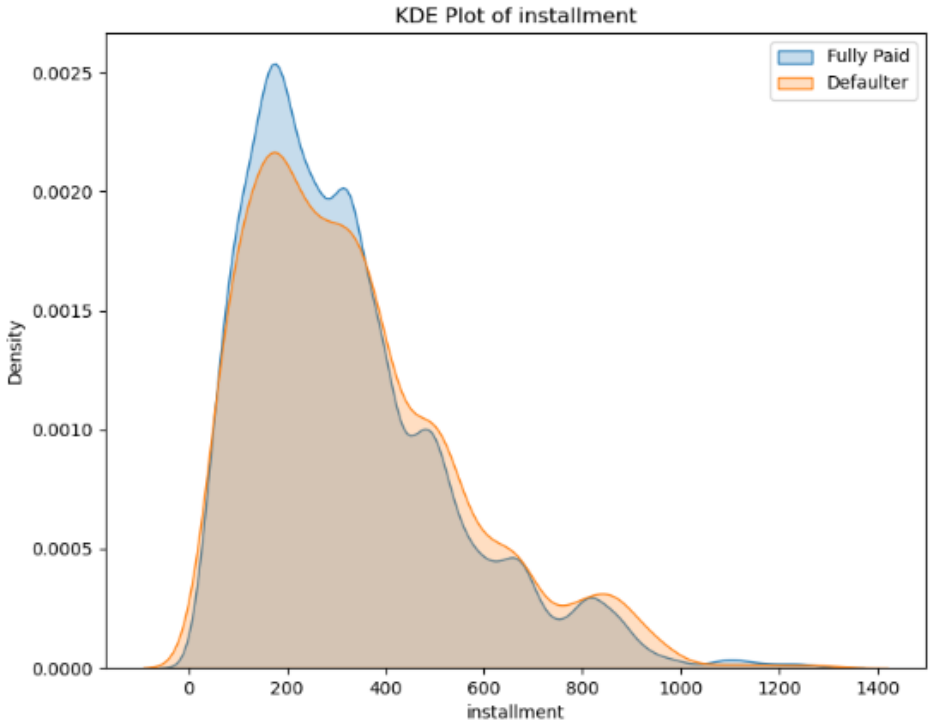
For Fully Paid Loans - Fully paid loans are more concentrated in the lower interest rate range (5% to 10%), indicating that borrowers with lower interest rates are more likely to successfully repay their loans.

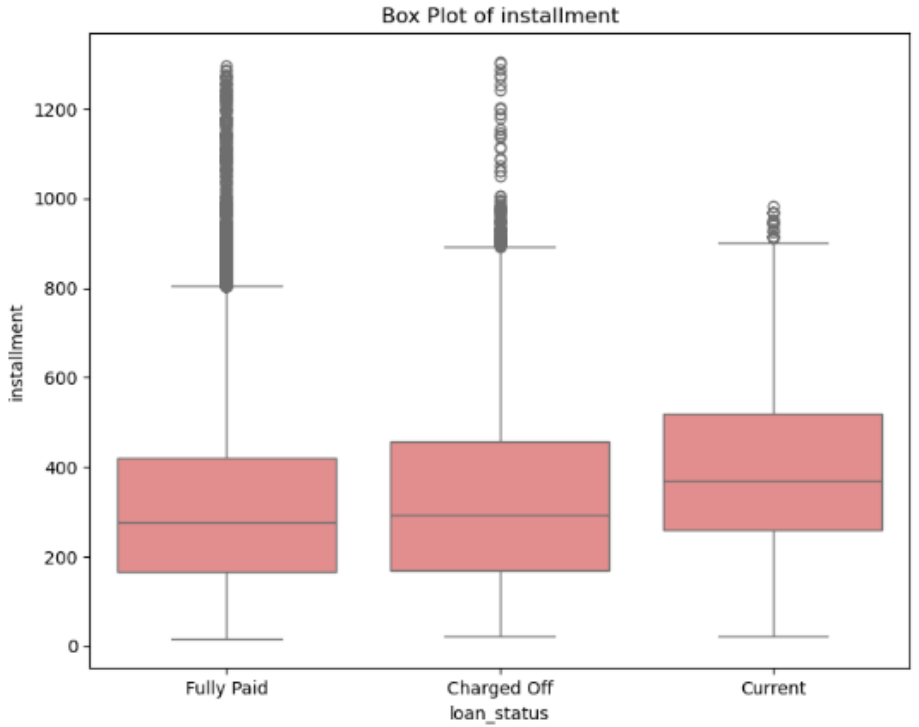
For Defaults - Defaulted loans are more prevalent at higher interest rates (above 15%), suggesting that loans with higher interest rates pose a higher risk of default.

Insights from Box Plot (Right Chart):

For Fully Paid Loans -The median interest rate for fully paid loans is lower compared to defaults, and the overall range of interest rates is narrower, indicating that lower interest rates contribute to successful repayments.

For Defaults - Defaulted loans (charged off) have a higher median interest rate with a wider spread and more outliers, suggesting that borrowers with high interest rates face greater financial burden and are more prone to default





PARAMETER - INSTALLMENT

INSIGHTS FROM KDE PLOT (LEFT CHART):

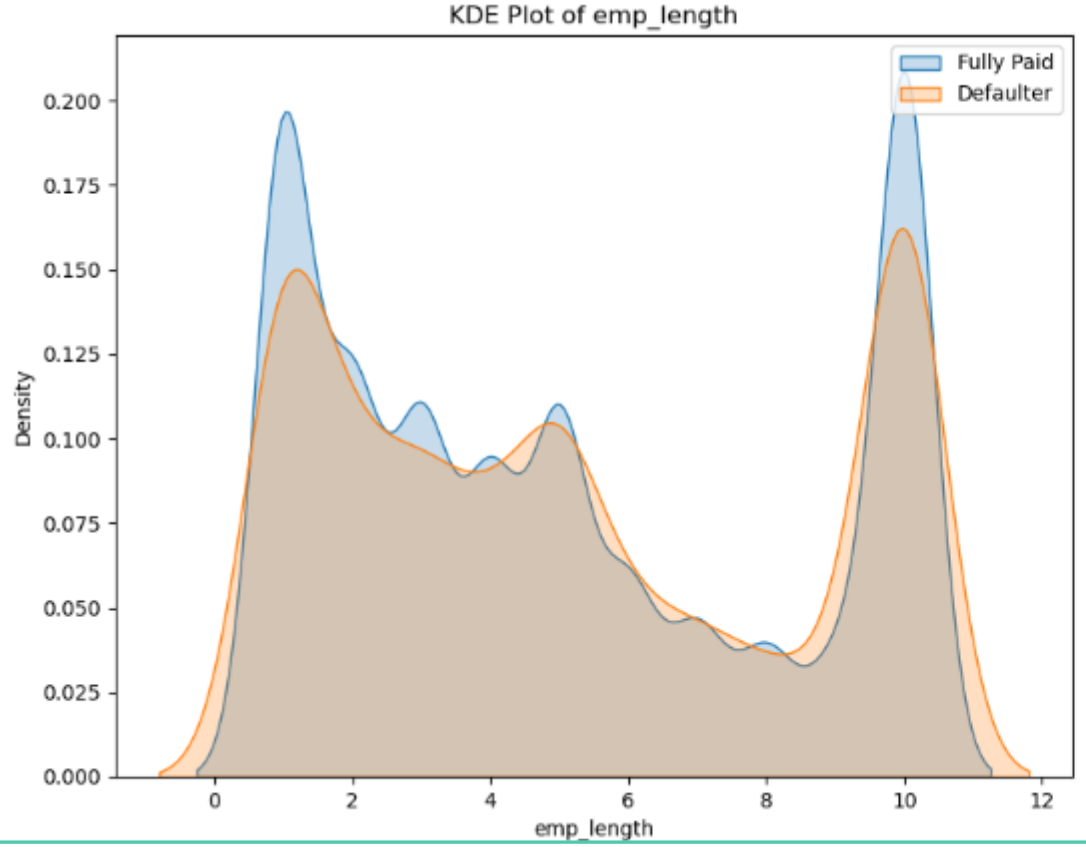
FOR FULLY PAID LOANS: THE DENSITY CURVE FOR FULLY PAID LOANS SHOWS A PEAK AROUND LOWER INSTALLMENT VALUES (APPROXIMATELY $200), INDICATING THAT BORROWERS WITH SMALLER INSTALLMENT AMOUNTS ARE MORE LIKELY TO COMPLETE THEIR PAYMENTS SUCCESSFULLY.

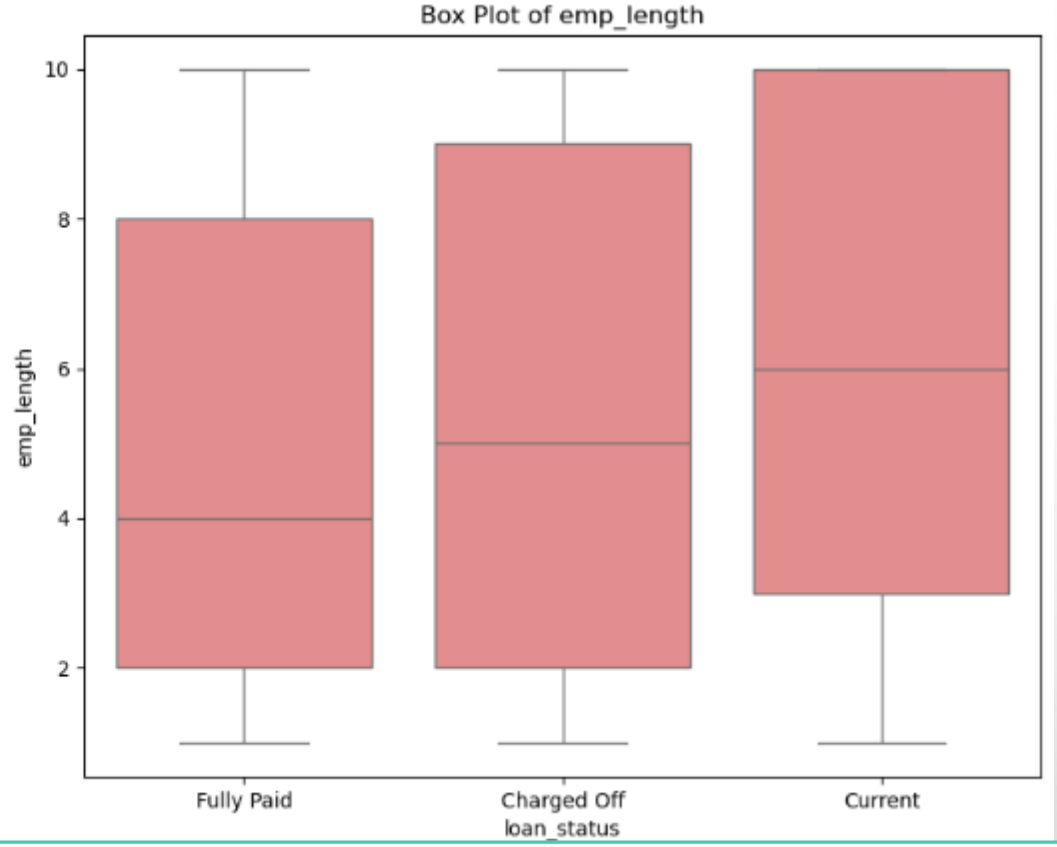
FOR DEFAULTS: DEFAULTED LOANS FOLLOW A SIMILAR DISTRIBUTION BUT WITH A SLIGHTLY LOWER PEAK, SUGGESTING THAT DEFAULTS OCCUR ACROSS A BROADER RANGE OF INSTALLMENT VALUES, ESPECIALLY IN THE MID-RANGE.

INSIGHTS FROM BOX PLOT (RIGHT CHART):

FOR FULLY PAID LOANS: THE MEDIAN INSTALLMENT AMOUNT FOR FULLY PAID LOANS IS LOWER COMPARED TO DEFAULTED AND CURRENT LOANS. THERE ARE FEWER HIGH-VALUE OUTLIERS, INDICATING THAT MANAGEABLE INSTALLMENT AMOUNTS CONTRIBUTE TO SUCCESSFUL LOAN REPAYMENT.

FOR DEFAULTS: DEFAULTED LOANS (CHARGED OFF) SHOW A WIDER DISTRIBUTION WITH NUMEROUS HIGH-VALUE OUTLIERS, SUGGESTING THAT HIGHER INSTALLMENT AMOUNTS





Parameter - Employment Length

Insights from KDE Plot (Left Chart):

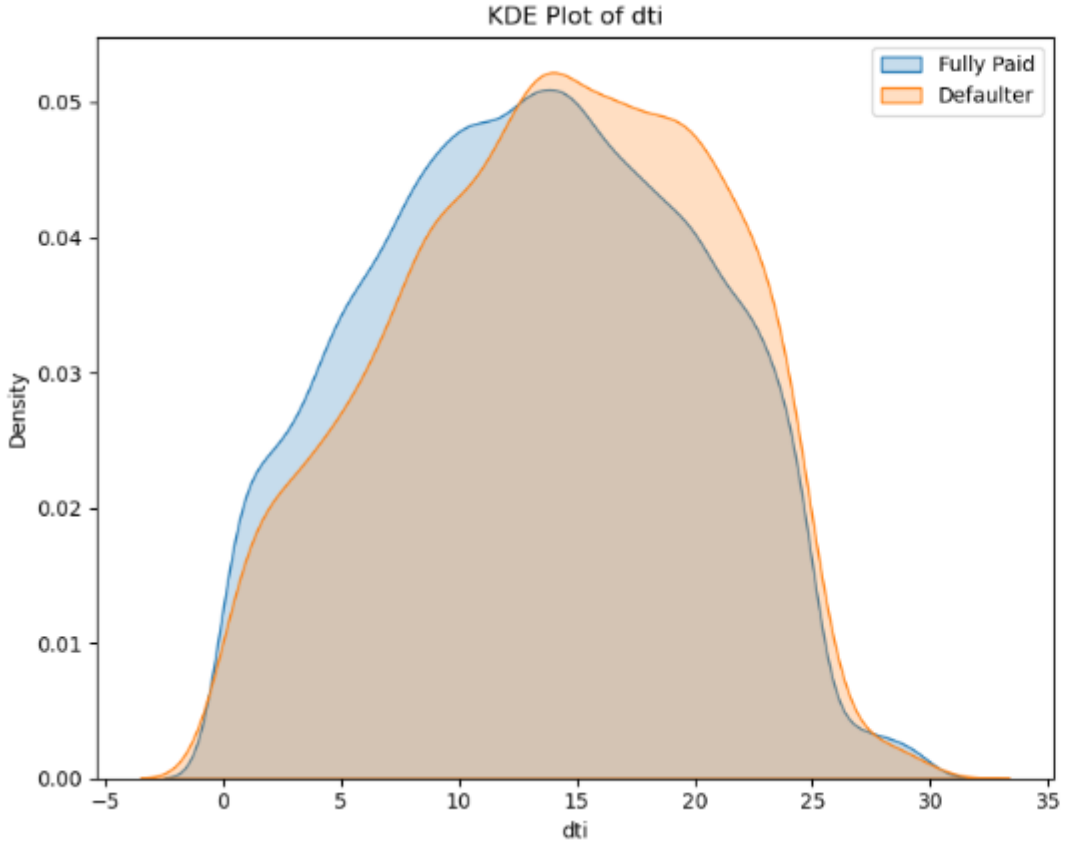
For Fully Paid Loans: Fully paid loans have two peaks, one at 0 years and another at 10 years of employment length, indicating that borrowers with either very short or very long employment tenure are more likely to repay successfully.

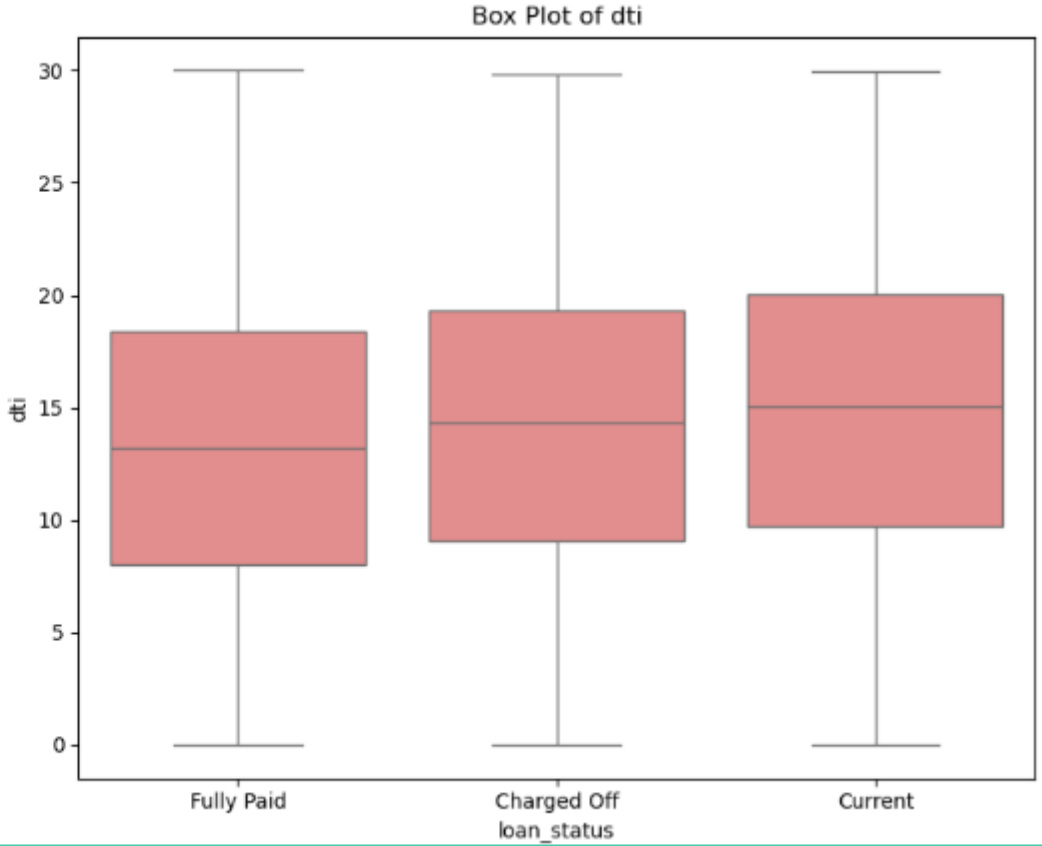
For Defaults: Defaulted loans follow a similar pattern but with slightly higher density in the mid-range (2 to 6 years), suggesting that borrowers with moderate employment experience may have higher default risks.

Insights from Box Plot (Right Chart):

For Fully Paid Loans: The median employment length for fully paid loans is lower compared to defaults, with a more concentrated distribution, indicating that borrowers with shorter employment tenure are more likely to repay successfully.

For Defaults: Charged-off loans show a wider interquartile range, implying that individuals with varied employment tenures (both short and long) face challenges in repayment. This category also has a higher median employment length, suggesting that longer employment does not always guarantee successful repayment.





Parameter - Debt-to-Income (DTI) Ratio

Insights from KDE Plot (Left Chart):

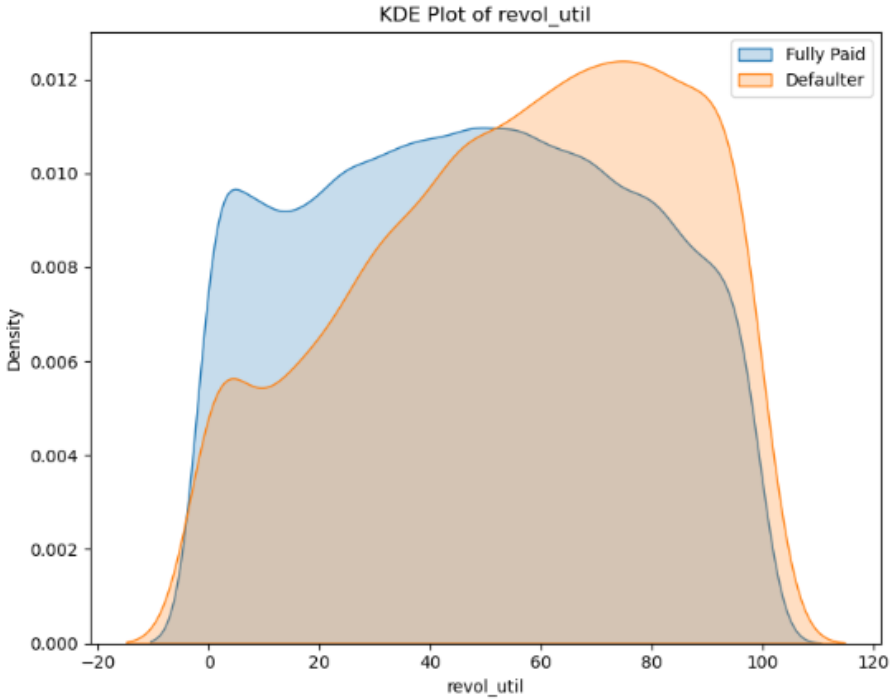
For Fully Paid Loans: The density curve for fully paid loans peaks around a DTI of 15, suggesting that borrowers with moderate debt-to-income ratios are more likely to repay their loans successfully. The distribution skews slightly to the left compared to defaulters, indicating that lower DTI values are more common among fully paid loans.

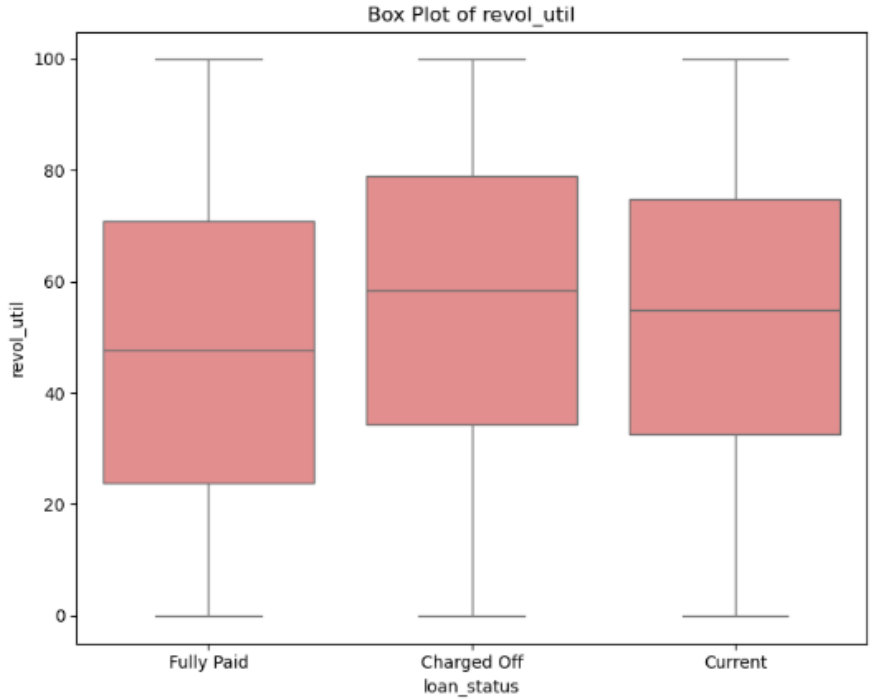
For Defaults: Defaulted loans show a similar peak around a DTI of 15 but with a slightly longer right tail, implying that higher DTI values are more associated with default risk. The overall distribution suggests that borrowers with DTIs above 20 face increased repayment challenges

Insights from Box Plot (Right Chart):

For Fully Paid Loans: Fully paid loans have a slightly lower median DTI compared to defaulted loans, indicating that borrowers with lower debt burdens relative to income are more successful in repayments.

For Defaults: Charged-off loans exhibit a similar interquartile range to fully paid loans but with a slightly higher median, suggesting that higher DTI levels correlate with increased default risk. The presence of high DTI values in the upper range could indicate financial stress leading to loan defaults.





Parameter - Revolving Utilization

Insights from KDE Plot (Left Chart):

Loans: Fully paid loans show a higher density around lower revol_util values (below 50%), suggesting that borrowers with lower credit utilization are more likely to repay successfully.

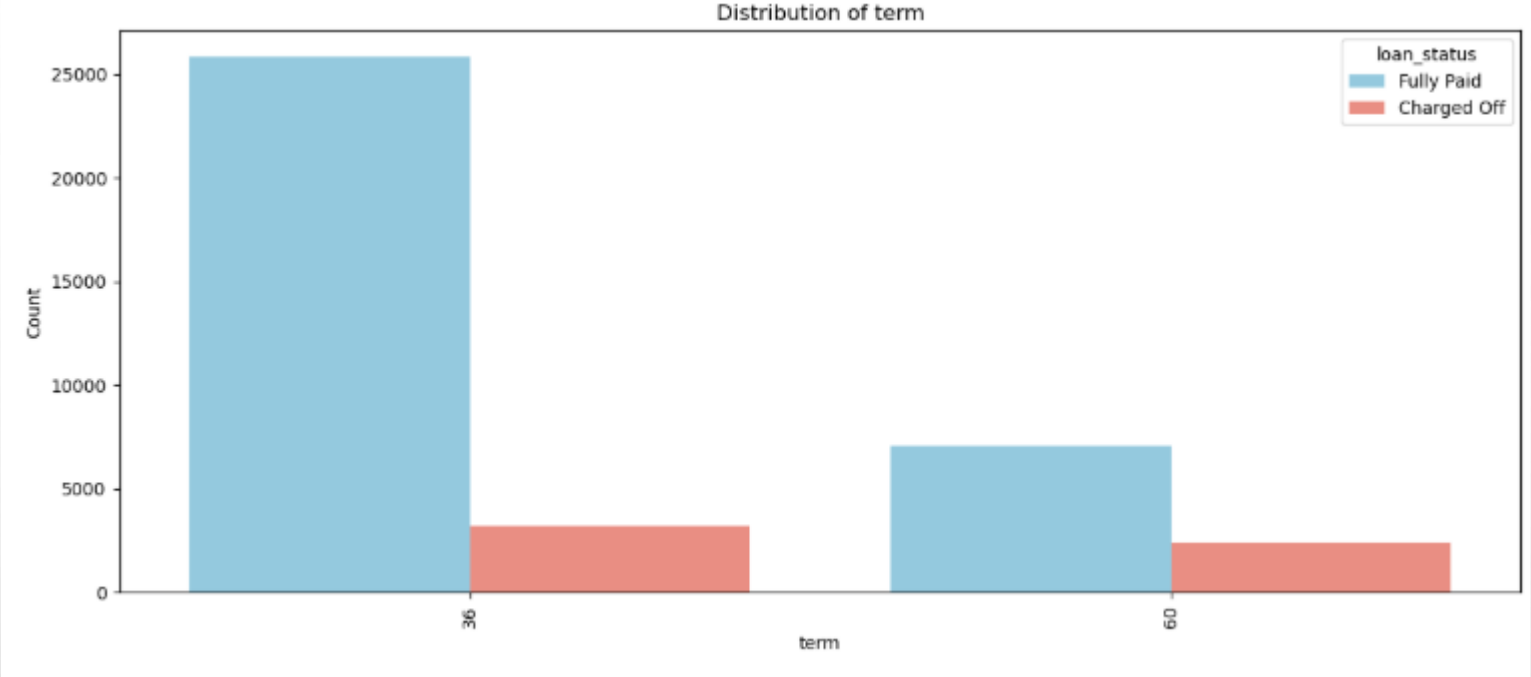
For Defaults: Defaulted loans exhibit a greater density at higher revol_util values (60% and above), indicating that higher credit utilization increases the risk of default.

Insights from Box Plot (Right Chart):

For Fully Paid Loans: Fully paid loans have a lower median revol_util value compared to defaults, with less variability, indicating that borrowers with lower and more consistent credit utilization are more likely to repay.

For Defaults: Defaulted loans (charged off) show a higher median revol_util and a wider interquartile range, suggesting that higher and more varied credit utilization may lead to repayment difficulties and an increased risk of default.

SINGLE VARIABLE ANALYSIS FOR CATEGORICAL COLUMNS



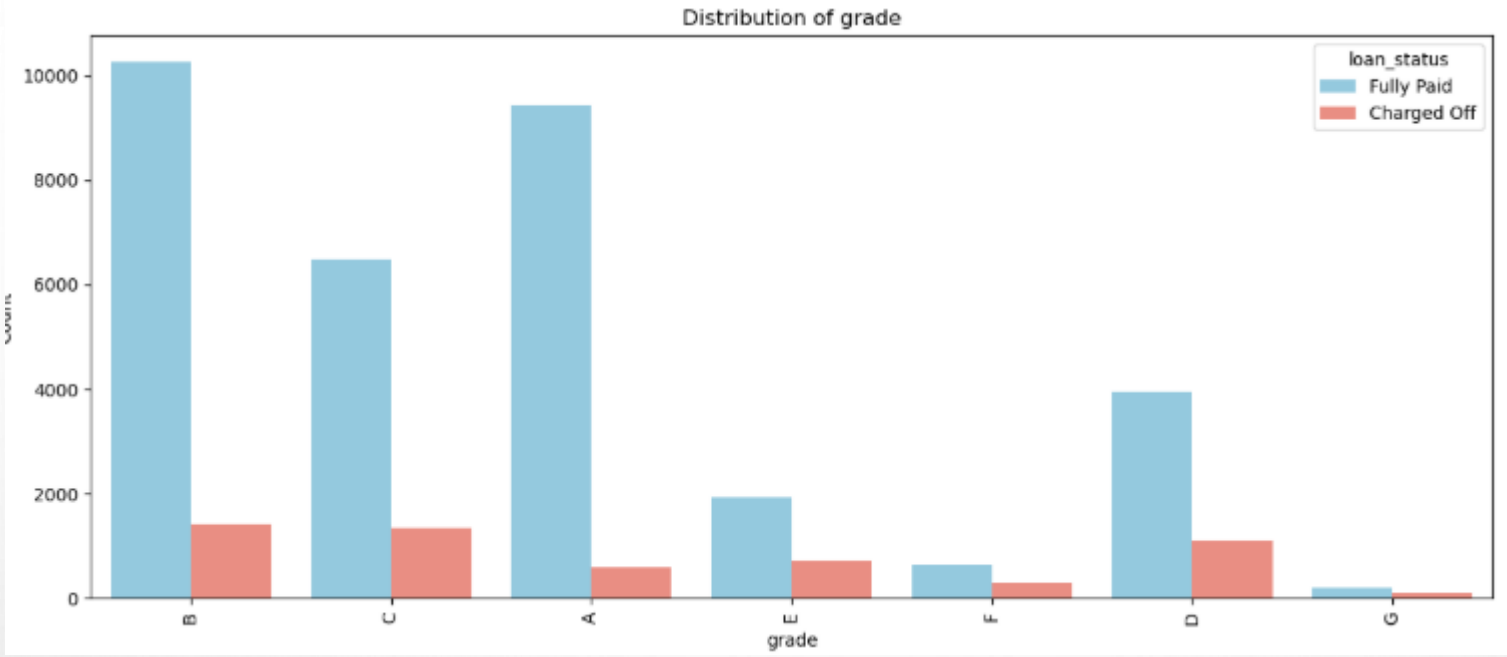
Term (Loan Duration) Analysis:

1. Fully Paid Loans:

Loans with a 36-month term have a significantly higher number of fully paid cases compared to 60-month loans, suggesting shorter loan terms might be more manageable for borrowers.

2. Charged Off Loans:

While defaults occur in both 36-month and 60-month loans, 60-month loans have a relatively higher proportion of charged-off cases, indicating longer durations may pose higher risks.



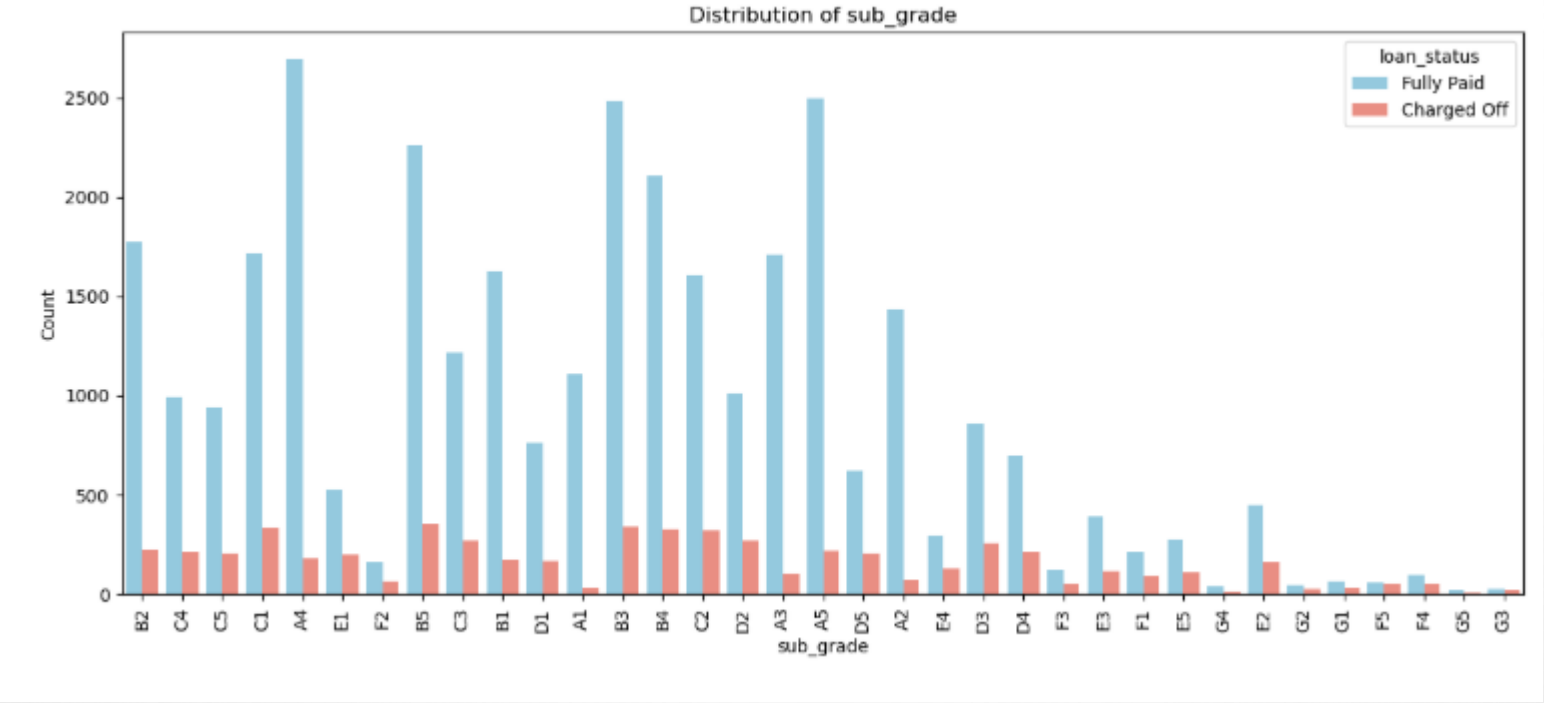
Grade Analysis:

1. Fully Paid Loans:

The majority of fully paid loans are concentrated in higher grades (A, B, C), indicating that loans of these grades are more likely to be paid successfully.

2. Charged Off Loans:

Lower grades, particularly D and below, show a higher proportion of defaults, highlighting that loans of these grades are more likely to be charged off.



Sub-Grade Analysis :

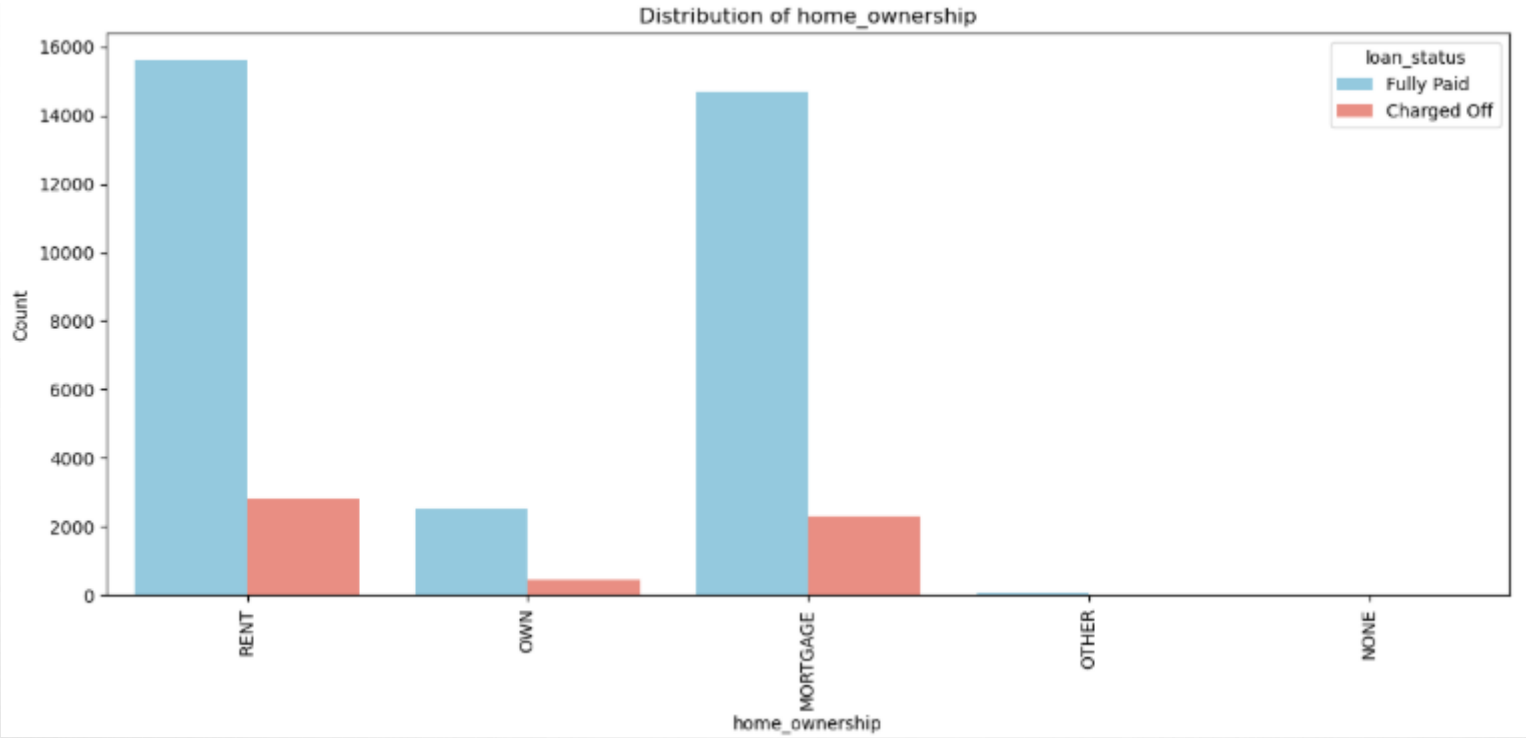
1. Fully Paid Loans:

Borrowers with sub-grades in the A and B categories (e.g., A1, A3, B1, B3) have a significantly higher count of fully paid loans, indicating that these sub-grades are associated with lower risk.

As the sub-grade moves towards lower categories (e.g., D, E, F, G), the number of fully paid loans decreases

2. Charged Off Loans:

Lower sub-grades such as E, F, and G have relatively higher proportions of defaults, indicating higher-risk borrowers. Even within the B and C categories, there is a visible portion of defaults, suggesting that borrowers in these mid-tier grades still pose some level of risk.



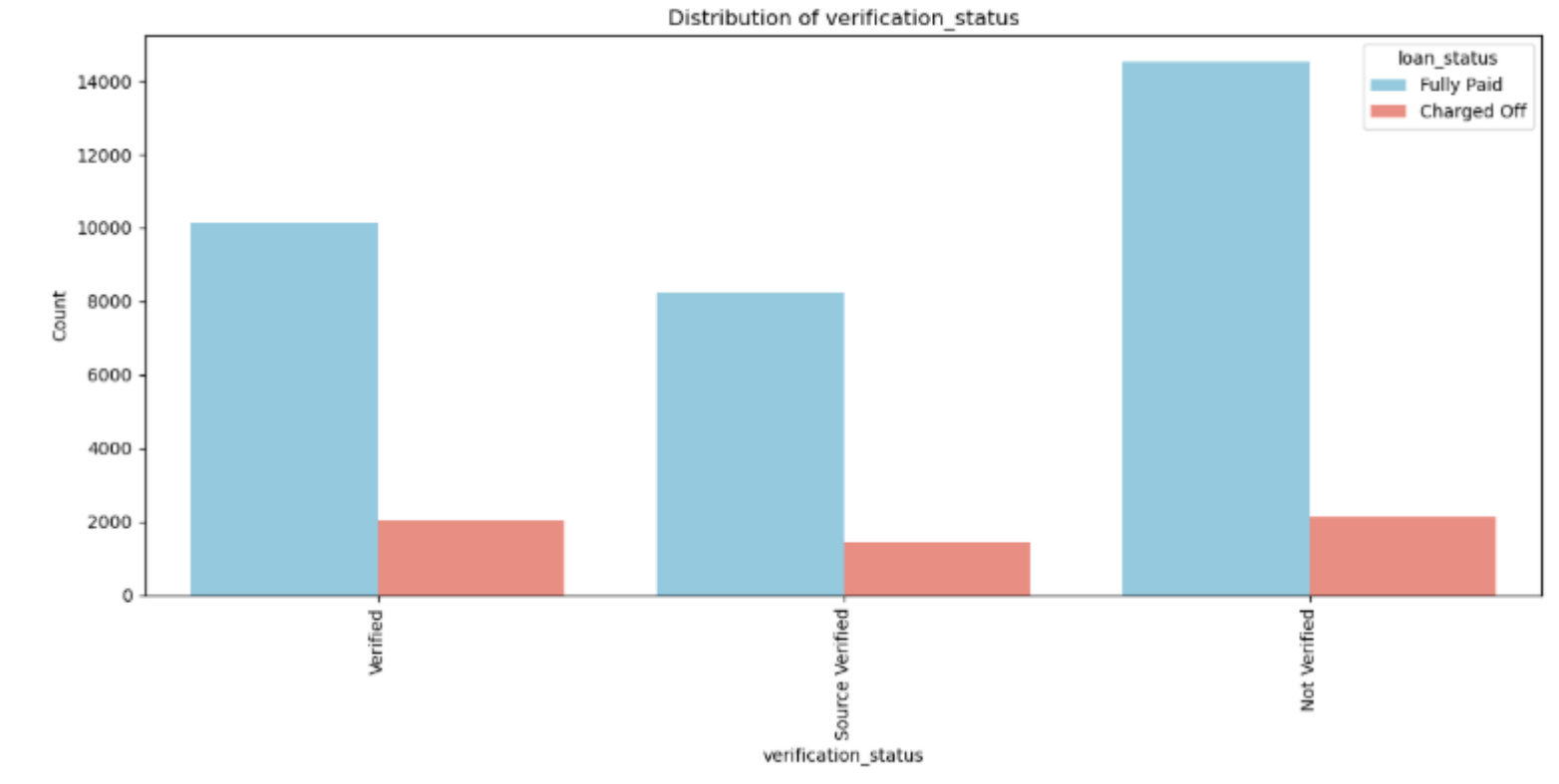
Home Ownership Analysis:

1. Fully Paid Loans:

Borrowers who own homes or have mortgages have a higher count of fully paid loans, suggesting that having property ownership contributes to financial stability and better loan repayment rates. Renters also have a large number of fully paid loans but in comparison, the proportion of charged-off loans is relatively higher.

2. Charged Off Loans:

Borrowers who have rented accomodatoin show a noticeably higher proportion of defaults compared to homeowners, indicating that lack of property ownership could be a potential risk factor for non-repayment.



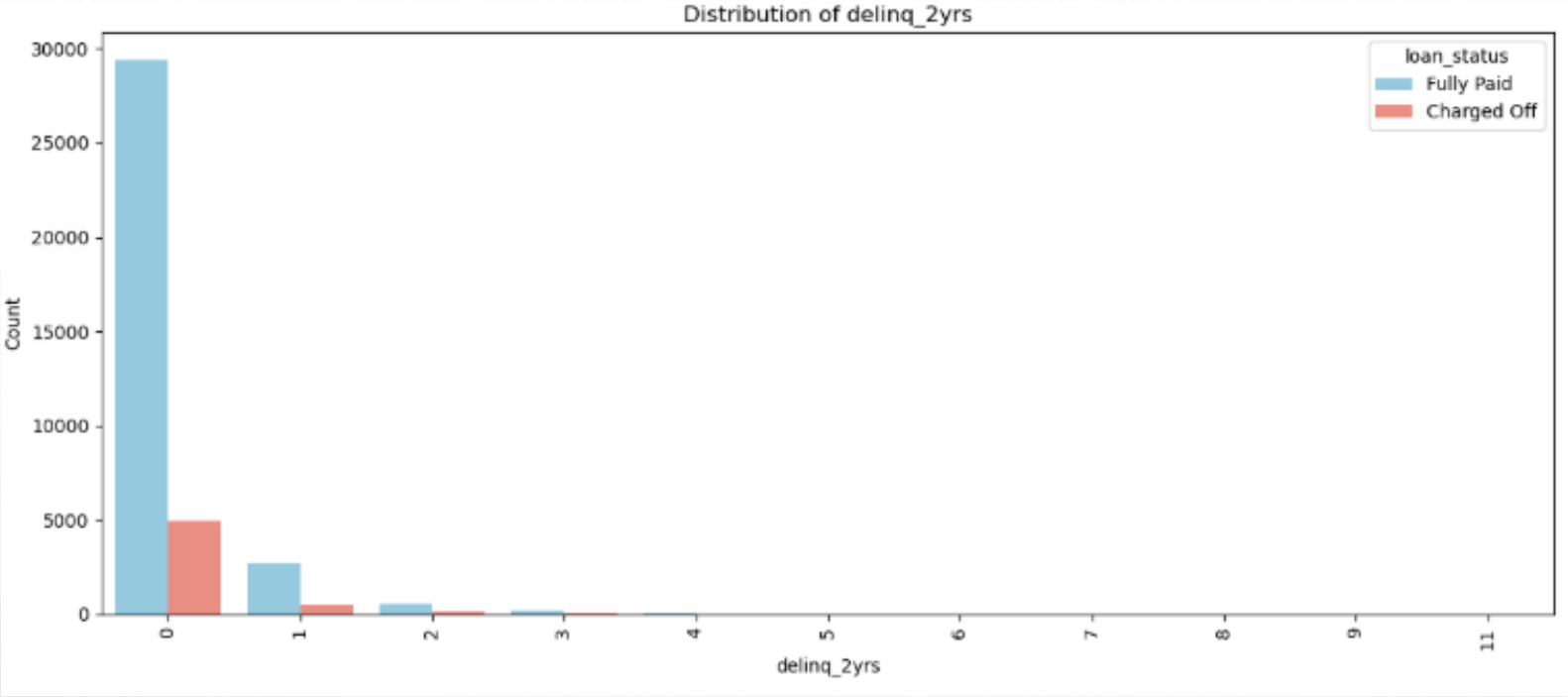
Verification Status Analysis:

•Fully Paid Loans:

- The number of fully paid loans is highest for the "Not Verified" category, followed by "Verified" and "Source Verified," indicating that loans can still be repaid without strict verification in many cases.
- However, "Verified" loans also have a strong representation among fully paid loans, suggesting that verification may contribute to better loan repayment outcomes.

•Charged Off Loans:

- Defaults are present across all verification statuses, but the proportion of defaults is higher in the "Not Verified" and "Source Verified" categories, indicating that lack of thorough verification may increase default risk.
- The "Verified" category shows relatively fewer defaults, highlighting that proper verification may help reduce credit risk.



Delinquencies in Last 2 Years (delinq_2yrs) Analysis:

•Fully Paid Loans:

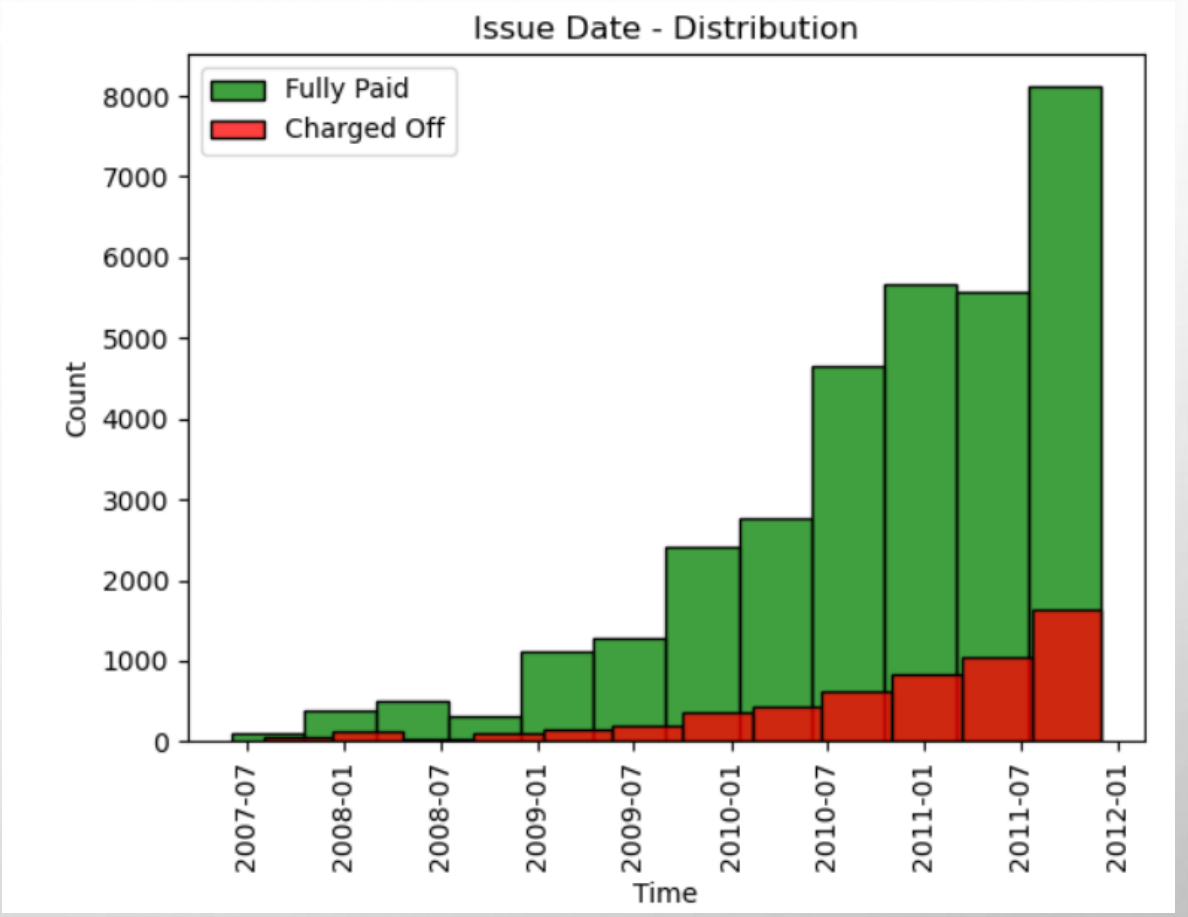
- A vast majority of fully paid loans have **zero delinquencies**, indicating that borrowers with a clean repayment history are more likely to fulfill their loan obligations.
- As the number of delinquencies increases, the number of fully paid loans drops significantly, showing a strong correlation between past delinquencies and repayment success.

•Charged Off Loans:

- A higher proportion of defaults is observed among borrowers with at least **one or more delinquencies**, reinforcing that past financial behavior is a strong predictor of loan default.
- Even at **zero delinquencies**, some defaults occur, but their proportion is relatively lower compared to those with delinquency history.

ISSUE DATE - DISTRIBUTION:

- **FULLY PAID LOANS:**
  - THE NUMBER OF FULLY PAID LOANS HAS STEADILY INCREASED OVER TIME, PEAKING AROUND 2011, WHICH MAY INDICATE GROWTH IN THE LOAN PORTFOLIO AND POTENTIALLY IMPROVED LENDING STRATEGIES.
  - THE TREND SUGGESTS THAT MORE RECENT LOANS HAVE A HIGHER CHANCE OF BEING FULLY REPAID, POSSIBLY DUE TO BETTER BORROWER SCREENING OR ECONOMIC CONDITIONS.
- **CHARGED OFF LOANS:**
  - CHARGED OFF LOANS HAVE ALSO INCREASED OVER TIME, THOUGH THEY REPRESENT A SMALLER PROPORTION COMPARED TO FULLY PAID LOANS.
  - THE RISE IN CHARGED OFF LOANS IN LATER YEARS MAY INDICATE HIGHER RISK-TAKING OR ECONOMIC DOWNTURNS AFFECTING BORROWERS' ABILITY TO REPAY.



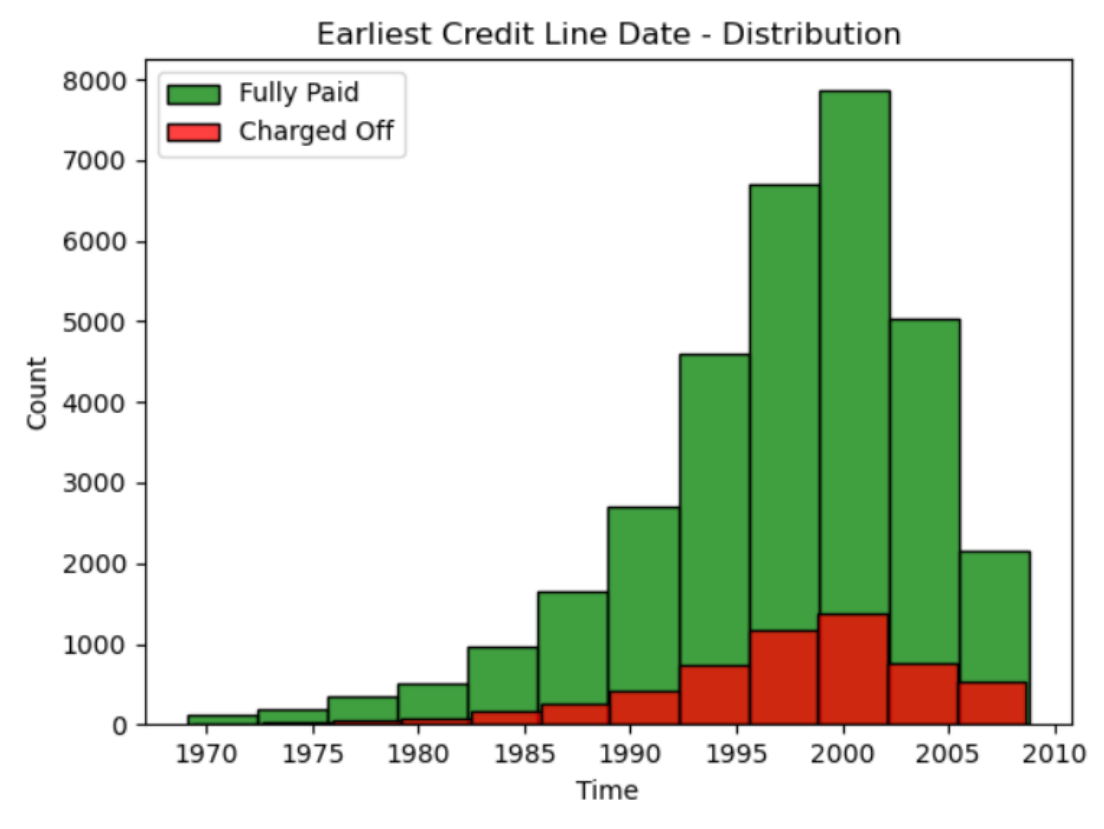
Earliest Credit Line Date - Distribution:

•Fully Paid Loans:

- Borrowers with earlier credit lines (from the 1980s and 1990s) have a higher number of fully paid loans, suggesting that individuals with longer credit histories are more reliable.
- Most fully paid loans are associated with credit lines established between 1995 and 2005, reflecting a matured credit history with potential financial stability.

•Charged Off Loans:

- Defaults are more prevalent among borrowers with shorter credit histories (post-2000), indicating that limited credit experience might contribute to higher default rates.
- However, some defaults are observed even among borrowers with older credit lines, highlighting that a long credit history does not always guarantee repayment success.



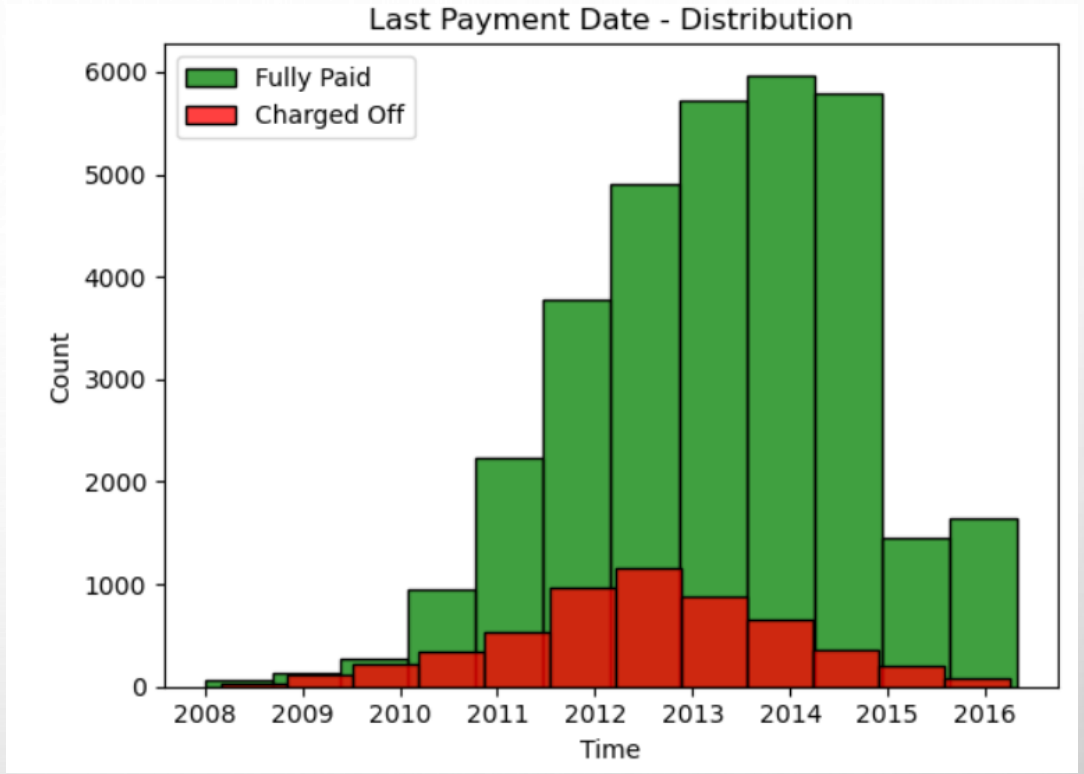
Last Payment Date Distribution

•Observation:

- The volume of loans increased steadily from 2010 to 2014, with a peak around 2013-2014.
- Loans classified as **"Fully Paid"** (green) dominate throughout the years.
- The proportion of **"Charged Off"** (red) loans appears higher in earlier years (2011-2013) and reduces in later years.

•Insights:

- The high concentration of **charged-off loans** in earlier years (2011-2013) suggests that defaults were more frequent for older loans.
- As time progresses, the number of fully paid loans surpasses the charged-off loans, indicating potential improvements in borrower behavior or credit policies.
- The tail end of the distribution in 2016 suggests fewer charged-off loans compared to earlier years.



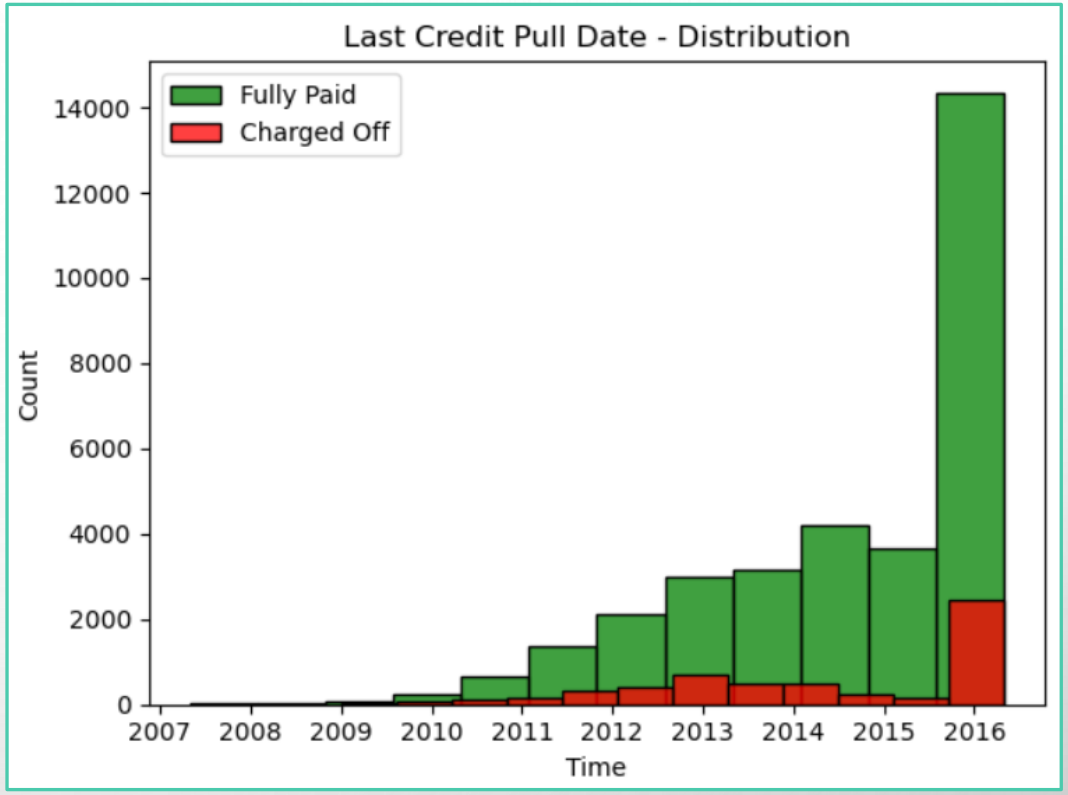
Last Credit Pull Date Distribution

•Observation:

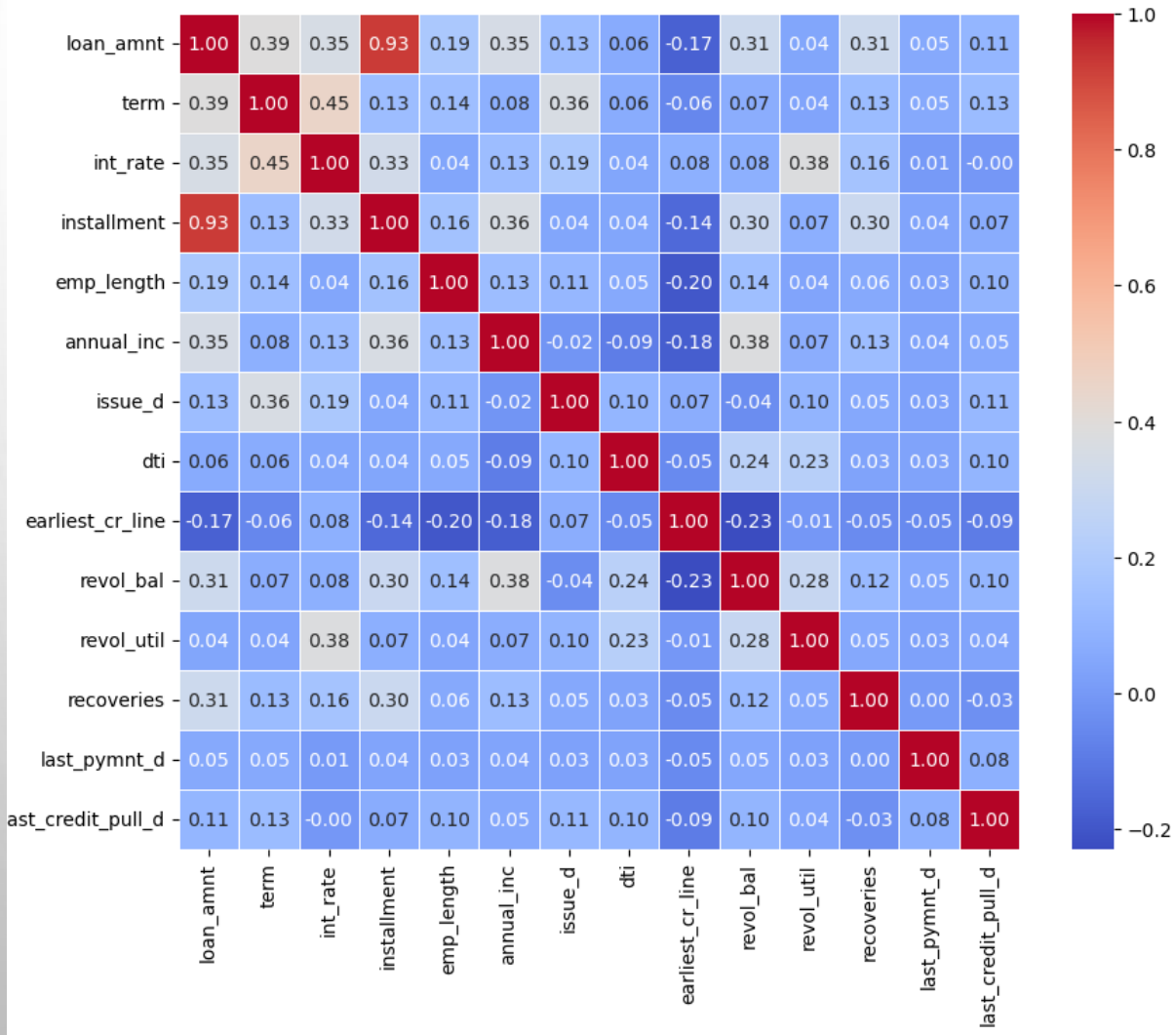
- Credit pulls increase sharply around 2015-2016, peaking at 2016.
- The share of **fully paid loans** is significantly higher than charged-off loans across all years.
- The proportion of **charged-off loans** is relatively small in recent years, indicating improved loan performance or stricter credit assessments.

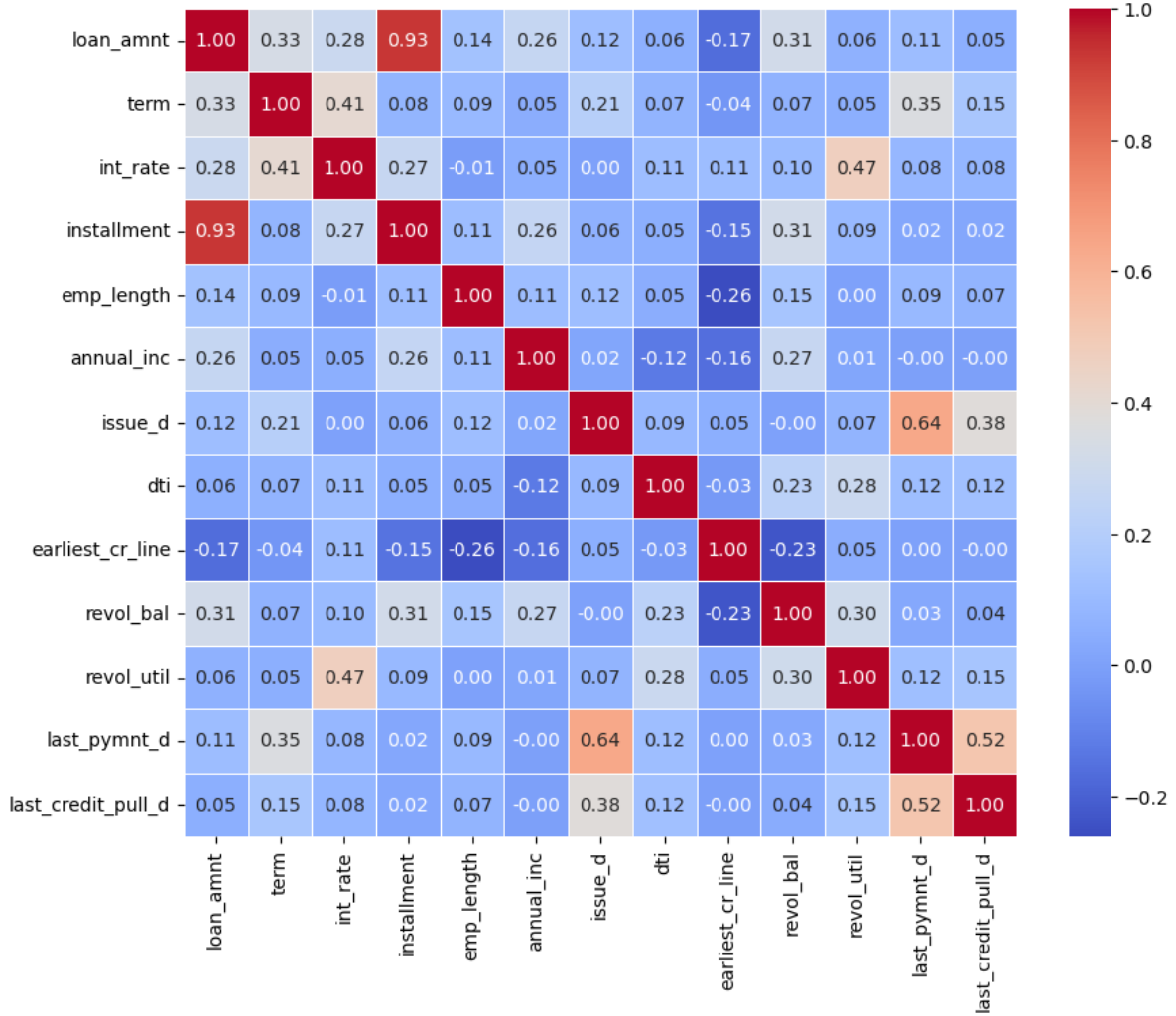
•Insights:

- A substantial number of credit pulls in 2016 might indicate a growing loan portfolio or higher customer engagement.
- The **higher number of fully paid loans** suggests borrowers were more likely to meet their obligations as the system matured.
- The charged-off proportion remains relatively low in later years compared to the last payment date chart, hinting at better monitoring and interventions.



MULTIVARIATE – HEAT MAP





Charged-off Loans

Fully-Paid Loans

KEY INSIGHTS FROM HEATMAP FOR FULLY PAID LOANS/CHARGED OFF COMPARISON

•**Loan amount and installment** show strong correlation in both **charged-off** and **fully paid** loans, indicating a consistent relationship between the borrowed amount and repayment structure.

•**Term and interest rate** are highly correlated in both cases, suggesting that longer-term loans tend to have higher interest rates regardless of repayment outcome.

•**Revolving utilization and interest rate** exhibit a stronger correlation in **charged-off** loans, implying that defaulters tend to have higher credit utilization, making them riskier borrowers.

•**Last credit pull date and last issue/payment date** are significantly more correlated in **fully paid** loans, suggesting that responsible borrowers check their credit balances more frequently

•**Term and loan amount** show higher correlation in **charged-off** loans, indicating that defaults are more common in larger, longer-term loans

•**Interest rate and loan amount** have a stronger correlation in **charged-off** loans, meaning defaulters tend to have both **higher loan amounts and higher interest rates**, reinforcing their risk profile.

EXECUTIVE SUMMARY - KEY FINDINGS AND INSIGHTS

EXECUTIVE SUMMARY - KEY INSIGHTS & FINDINGS

- **Borrower's Financial Profile & Loan Default Probability**
  - **Loan-to-Income Ratio:** Borrowers applying for **loans above 25% of their income** are significantly more likely to default.
  - **Debt-to-Income (DTI) Ratio:** Defaults increase when **DTI exceeds 20%**, indicating financial stress.
  - **Revolving Balance-to-Income Ratio:** Borrowers with a **revolving balance exceeding 20% of their income** are at a higher risk of default. Managing revolving credit utilization is critical in assessing loan repayment ability

- **Loan Parameters & Default Risk**
  - **Loan Amount:** Higher loan amounts ($15,000 - $25,000) have a greater likelihood of default.
  - **Interest Rate Suitability:**
    - Loans **fully paid** tend to have **interest rates between 5% and 10%**, indicating this is a lower-risk range.
    - Loans **with default status** mostly fall **above 15% interest rates**, suggesting high-risk borrowers tend to take costlier loans.
    - Interest rate adjustments based on risk categories can improve portfolio performance.
    - **Installments:** Higher monthly payments correlate with increased default rates, emphasizing the importance of manageable installment structures.

EXECUTIVE SUMMARY - KEY INSIGHTS & FINDINGS(CONTD.)

- **Borrower Demographics & Risk Assessment**
  - **Employment Length:** Borrowers with **shorter or very long employment histories** tend to repay better, while mid-range employment (2-6 years) has a higher default rate
  - **Home Ownership: Renters** are more likely to default than homeowners or those with mortgages
  - **Verification Status: Loans without verification** exhibit higher default risks, emphasizing the need for stronger verification protocols
  - **Delinquency History:** Borrowers with **past delinquencies** are significantly more likely to default, reinforcing its importance in credit risk assessments

- **Loan Term & Credit Grade Impact**
  - **Loan Duration: 60-month loans** have a higher default rate than **36-month loans,** suggesting shorter durations reduce risk
  - **Credit Grades:** Borrowers in lower credit grades (D, E, F, G) have disproportionately higher default rates

- **Time-Based Trends**
  - **Early vs. Recent Borrowers:** Default rates were higher in early years (2011-2013) but improved with better lending policies
  - **Credit History Length:** Longer credit histories (pre-2000) correlate with lower default risks

ACTIONABLE RECOMMENDATIONS

- **Refine loan approval criteria** by setting stricter cutoffs for Loan-to-Income, DTI, and **Revolving Balance-to-Income ratios.**
- **Optimize interest rate structures** by **limiting high-risk loans above 15% interest rates** and incentivizing borrowers with lower-risk profiles.
- **Encourage shorter loan terms (36 months)** to mitigate default risk.
- **Strengthen borrower verification** to reduce defaults, particularly among renters and lower credit grades.
- **Enhance credit risk models** by incorporating **revolving utilization, employment stability, and past delinquencies** into decision-making.