MLC64- Lending Club Case Study

Study Group :

Wasim Feroz Ali , Zaid Ahmed

Problem Statement

- Lending Club faces a critical challenge in managing its loan approval process. Evaluating loan applications requires making sound decisions to minimize financial losses, primarily stemming from loans extended to risky applicants.

- The primary objective is to assist Lending Club in mitigating credit losses, focusing on two key scenarios:

1. Identifying applicants likely to repay loans is crucial for generating profits through interest payments. Rejecting such applicants would result in a loss of potential business.

2. Approving loans for applicants unlikely to repay and at risk of default can lead to substantial financial losses for the company.

WHAT AM I LOOKING FOR?

- The aim is to look for patterns which indicate if a person is likely to default.
- & Based on the findings, necessary steps can be taken which includes loan being denied, reducing loan amount, increasing interest rate, guarantees etc.

Handling the Data

- The data has <u>39717</u> Rows and <u>111</u> columns.

- For proper Data Analysis, we performed Data Cleaning by taking care of null values, odd values, dropping the columns which does not correspond to the loan being charged off.

- Since loan Defaulting can be found using only 'fully paid' or 'charged off' entries. We discarded 'current' status entries from our dataframe.

- Next Step in our Analysis is Finding Missing Values and Fixing them either by removing the values not needed or entering most suitable value based on the column.

- After that, outliners were treated.

Univariate Analysis

- The majority of loans in the dataset are fully paid, indicating that most borrowers successfully repay their loans.

- A significantly smaller proportion of loans are charged off, suggesting that default rates are relatively low compared to the total number of loans.
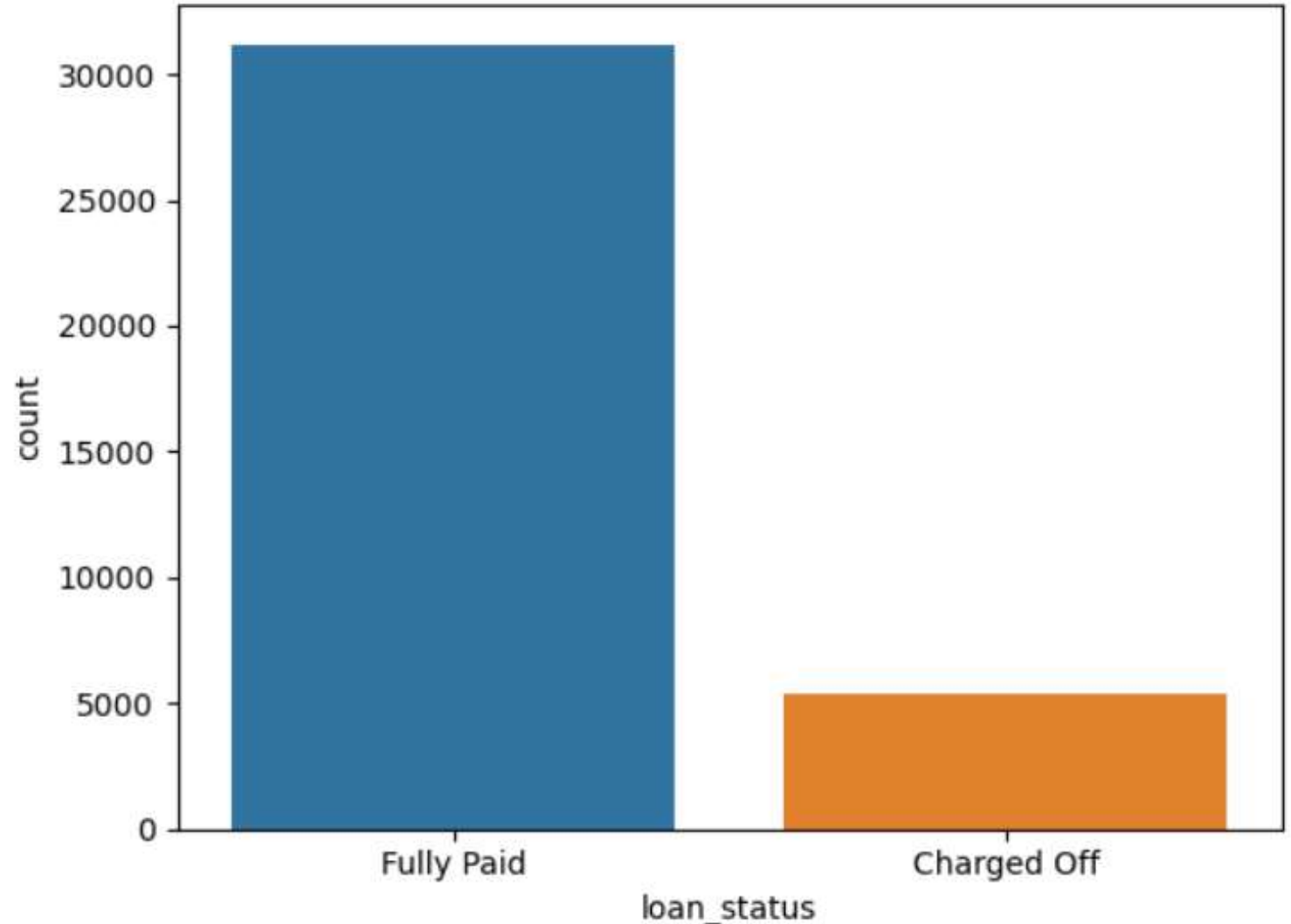
Univariate Analysis – LC Assigned Loan Grade

- From LC assigned Loan grade we can conclude that,

- Loans with grades B and C are more prone to being charged off, indicating that borrowers in these categories might pose a higher risk

- The lower count of charged off loans in grades A and the higher grades suggest that these loans are less risky
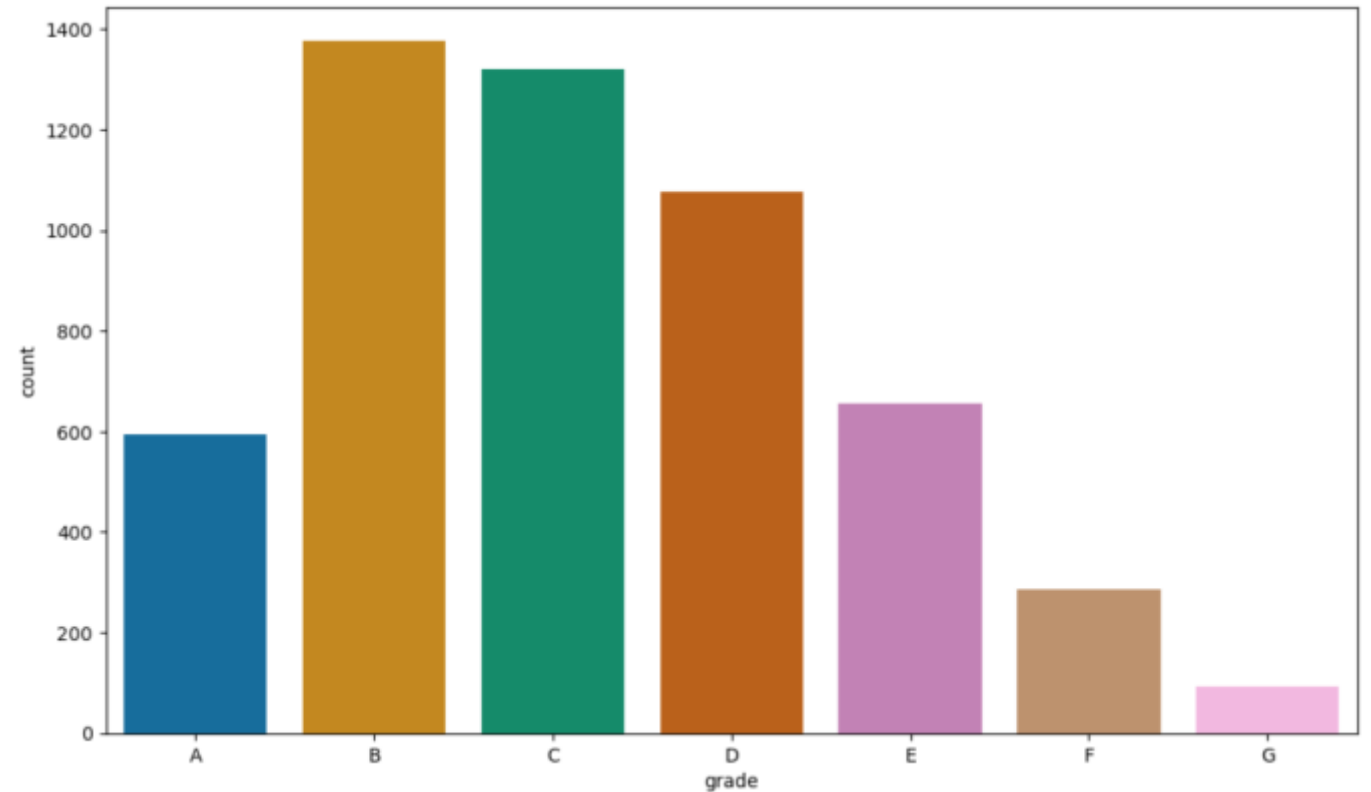
The data suggests a higher risk associated with loans graded B and C by Lending Club, while loans graded A, F, and G are less likely to be charged off. This could inform risk reducing strategies and loan approval criteria

Univariate Analysis – Home Ownership

- Borrowers who rent or have a mortgage are more likely to default on their loans compared to those who own their homes outright.

- The high count of charged-off loans for renters suggests they might be more financially unstable or have less disposable income compared to homeowners.

- Borrowers with mortgages still pose a significant risk, potentially due to the financial burden of both the mortgage and additional loans.
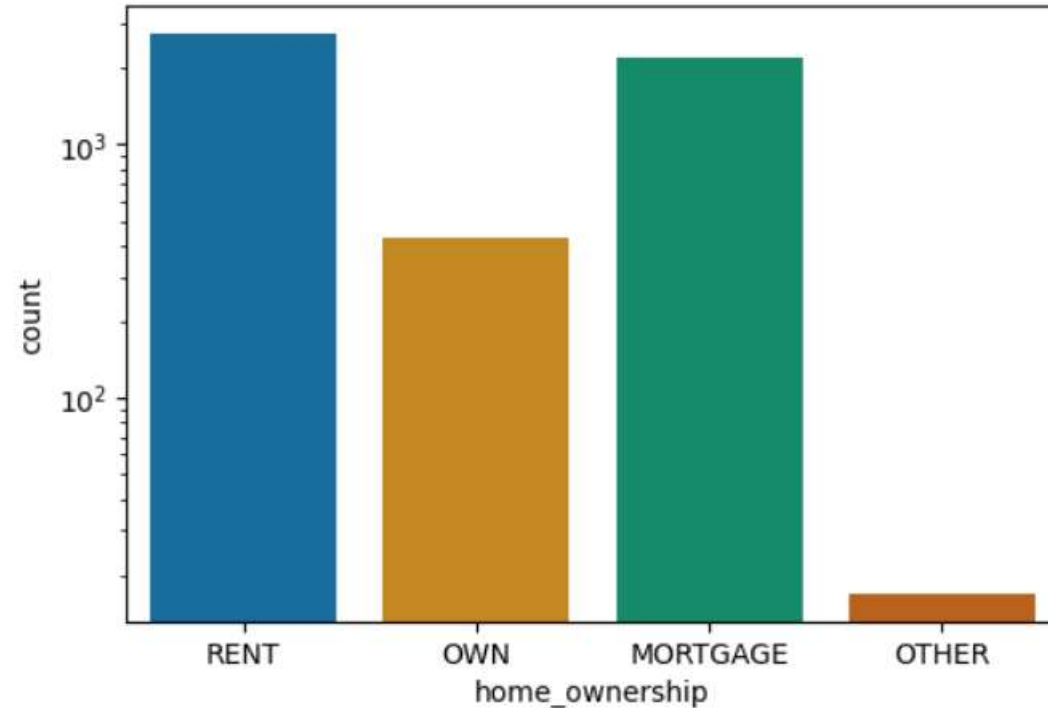

Home ownership graph suggests that home ownership status is a significant indicator of loan default risk. Borrowers who rent or have mortgages are at a higher risk of default, whereas those who own their homes outright are less likely to default on their loans. This information could be valuable for lenders in assessing risk and making lending decisions

# Univariate Analysis – Purpose - Interest Rate
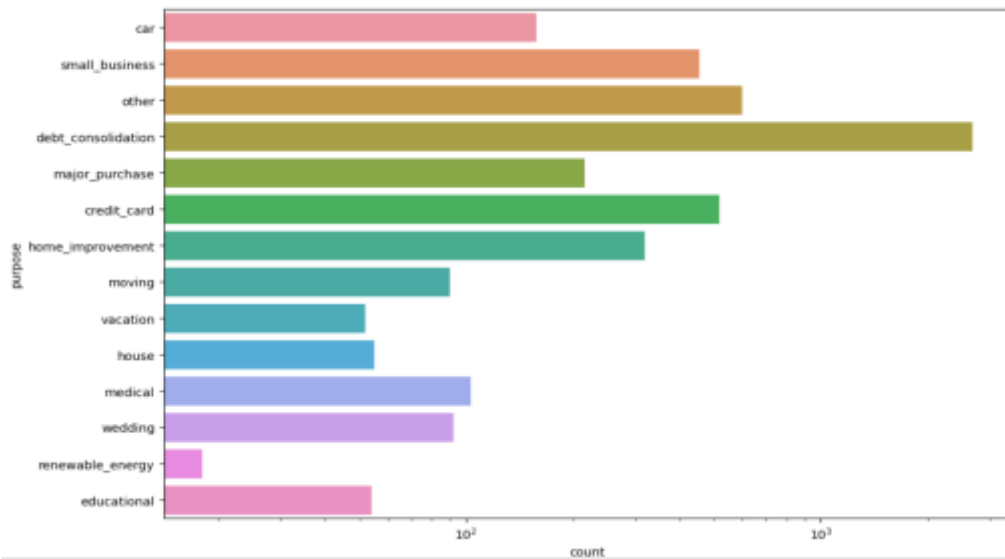
Purpose of the loan shows that loans taken out for debt management purposes (debt consolidation and credit card refinancing) have the highest default rates. This could indicate that borrowers using loans to manage existing debts may be struggling financially. Lenders might consider implementing stricter criteria for these types of loans to mitigate risk. Conversely, loans for educational and renewable energy purposes appear to be lower risk, which could inform lending strategies.
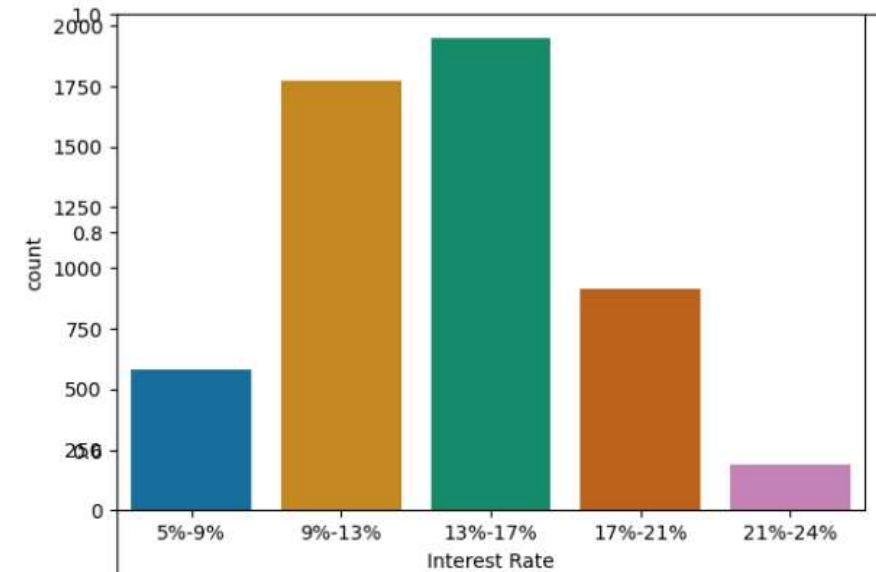
The interest rate information here suggests that mid-range interest rates (particularly 13%-17%) are associated with the highest number of loan defaults. Lenders might focus on these interest rate ranges for risk assessment and consider additional measures to mitigate default risks for loans issued within these ranges. Conversely, loans with lower interest rates (5%-9%) and very high rates (21%-24%) seem to be less prone to default, which could inform lending strategies and risk management practices

Univariate Analysis – Employment Length & Annual Income
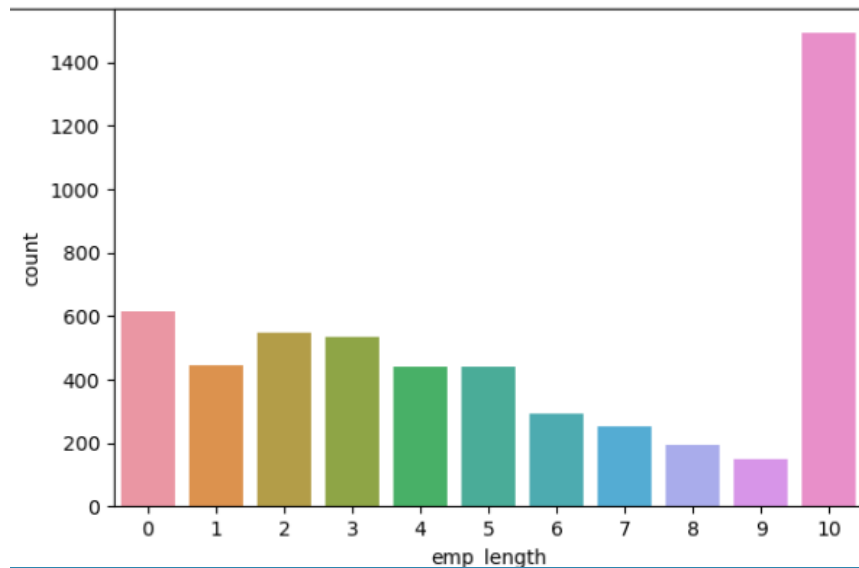
Employment terms here suggests that both very long and very short employment lengths are associated with higher counts of charged-off loans. Lenders might need to consider additional factors beyond employment length when assessing borrower risk, as longevity in employment alone does not guarantee a lower risk of default. More comprehensive risk assessments that include employment stability, income, and other financial factors could be beneficial in mitigating loan defaults
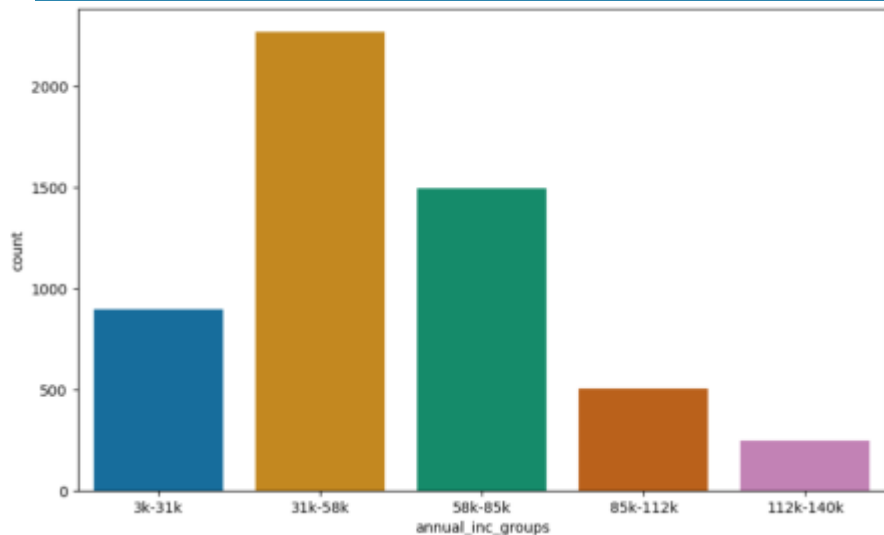

The income data here indicates that middle-income borrowers (31k-85k) have the highest risk of defaulting on their loans. Lenders might consider implementing additional risk mitigation strategies for these income groups. Conversely, higher income borrowers (above 112k) appear to be lower risk, which could inform lending practices and risk assessments. This visualization helps in identifying the correlation between income levels and loan default rates, providing valuable insights for financial decision-making

# Univariate Analysis – Credit Line / Revolving Accounts

This suggests that borrowers with a moderate number of open accounts (2-19) are at a higher risk of default. Lenders might consider closely evaluating the credit profiles of borrowers in this range to mitigate the risk of defaults. Borrowers with a very high number of open credit lines (27-44) are less likely to default

The revolving utilization rate here is a strong indicator of loan default risk. Borrowers with higher utilization rates (60-100%) are significantly more likely to default, while those with lower utilization rates (0-40%) pose a lower risk. Lenders might consider closely monitoring the revolving utilization rates of borrowers to assess default risk and make informed lending decisions.

The total credit line data shows that borrowers with a moderate to low number of total credit lines (2-37) are at a higher risk of default. Lenders might consider closely evaluating the credit profiles of borrowers in this range to mitigate the risk of defaults. Borrowers with a very high number of total credit lines (55-90) are less likely to default
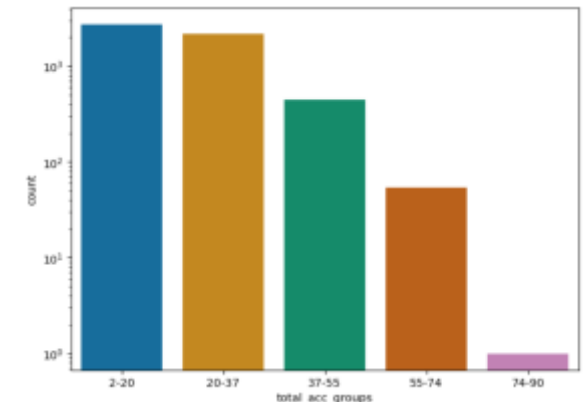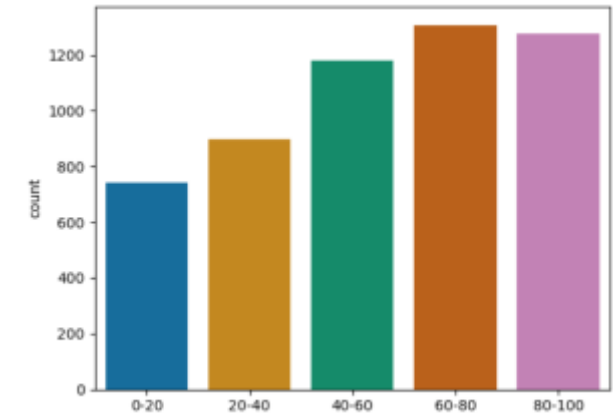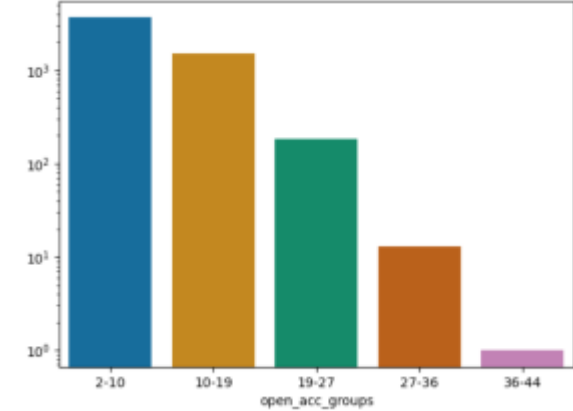
# Univariate Analysis – Duration – Income verification
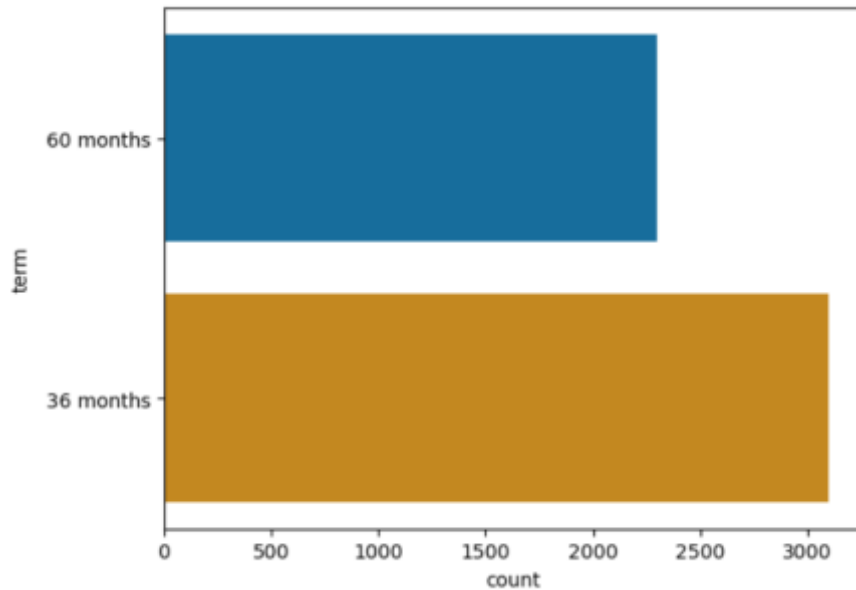
This graph suggests that loan term is an important factor in assessing default risk. Loans with a 36-month term have a higher default count, which could inform lending strategies and risk assessment. Lenders might consider implementing stricter criteria or additional support for borrowers opting for shorter-term loans to mitigate the risk of defaults
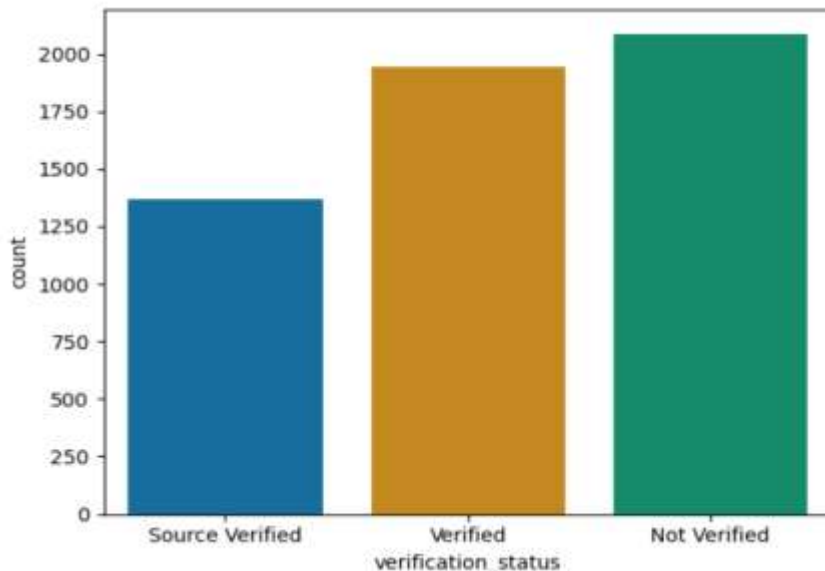

From this, we can draw conclusions that income verification status is an important factor in assessing loan default risk. Loans where income is not verified have the highest default counts, indicating that lenders should consider stricter criteria for such cases. Thorough verification processes, like source verification, are associated with lower default rates

# Univariate Analysis – Public Record – Month of Issue
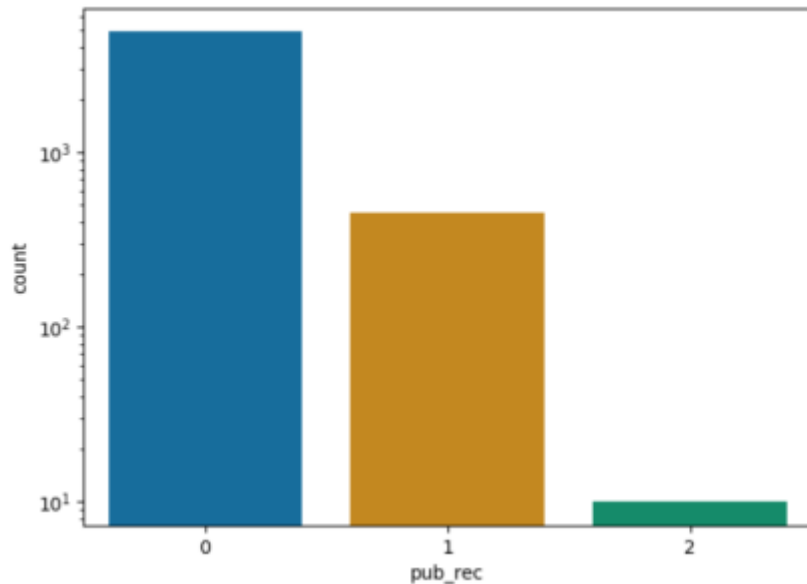
The number of derogatory public records is an important factor in assessing loan default risk. Interestingly, most defaults occur among borrowers with no derogatory public records, indicating that these borrowers might have other underlying financial issues not captured by public records. Lenders should consider a comprehensive risk assessment that includes other financial metrics alongside public records to better predict and mitigate the risk of loan defaults.

This month data suggests that the loan issue month is an important factor in assessing the risk of loan defaults. Loans issued in December and the surrounding months (October and November) have the highest default counts, indicating a higher risk during this period. Lenders should be aware of these seasonal trends and consider implementing additional risk measures for loans issued towards the end of the year.
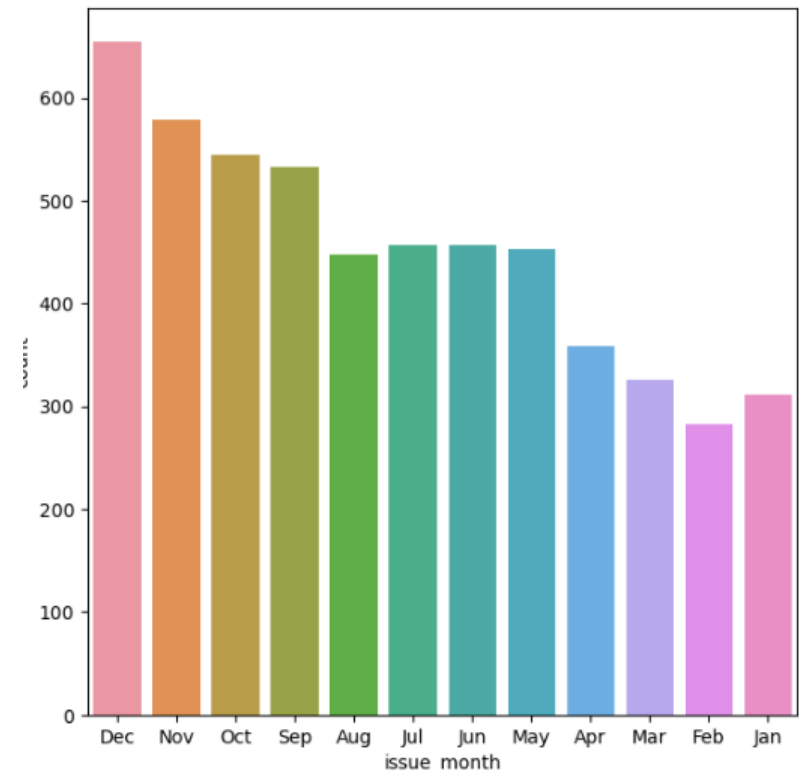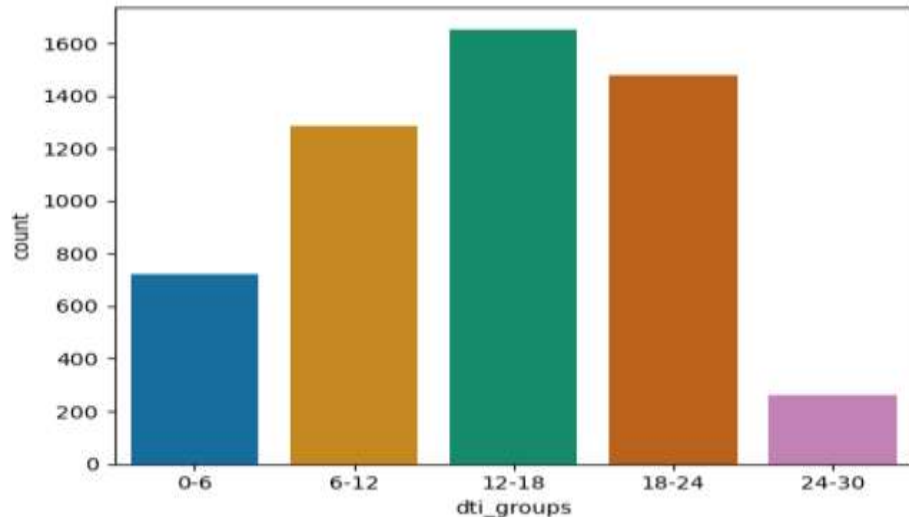
Univariate Analysis – Loan Amount – DTI
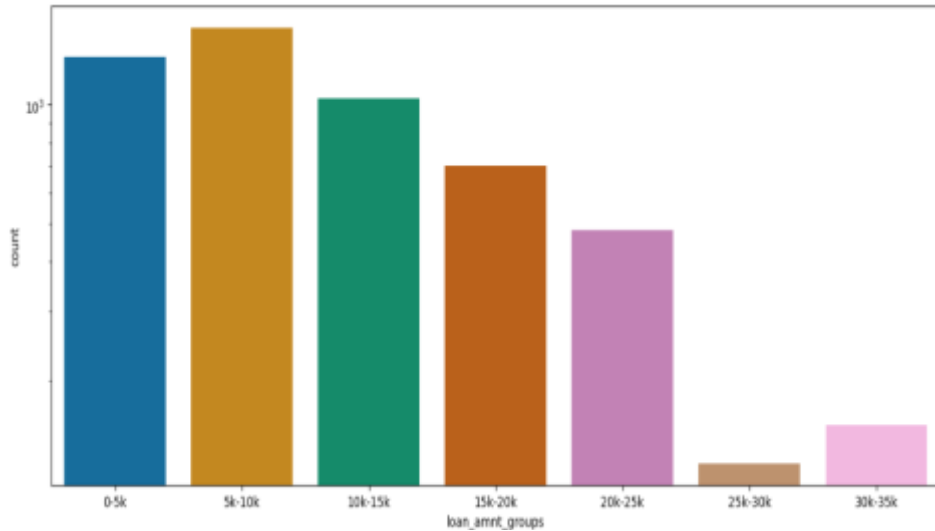
The highest count of charged-off loans is associated with the loan amount group **$5k-$10k**, indicating that this loan size is the most prone to defaults Conversely, larger loans ($25k-$35k) show the lowest default rates, suggesting that these loans are less risky or better managed

The topmost count of charged-off loans is associated with the DTI group **12-18**, indicating that borrowers within this DTI range are the most prone to default. The next highest count of defaults is for the DTI group **18-24**, followed closely by the **6-12** group. This proves that borrowers with DTI ratios between **12-24** are at the highest risk of default

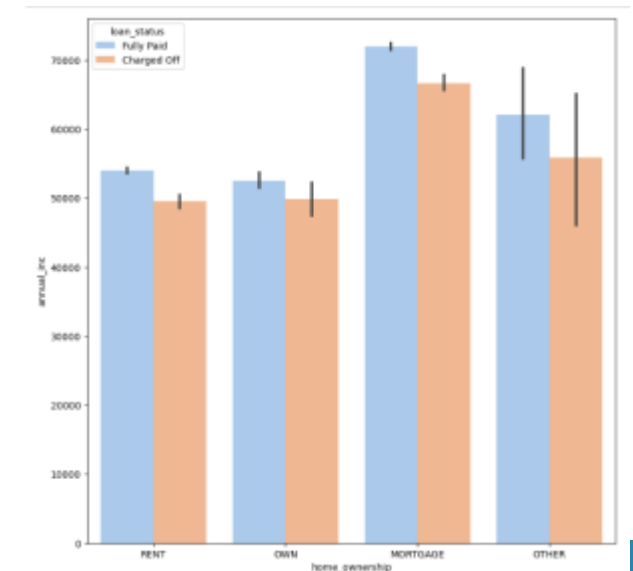# Bivariate Analysis - Annual income vs loan purpose & Home ownership
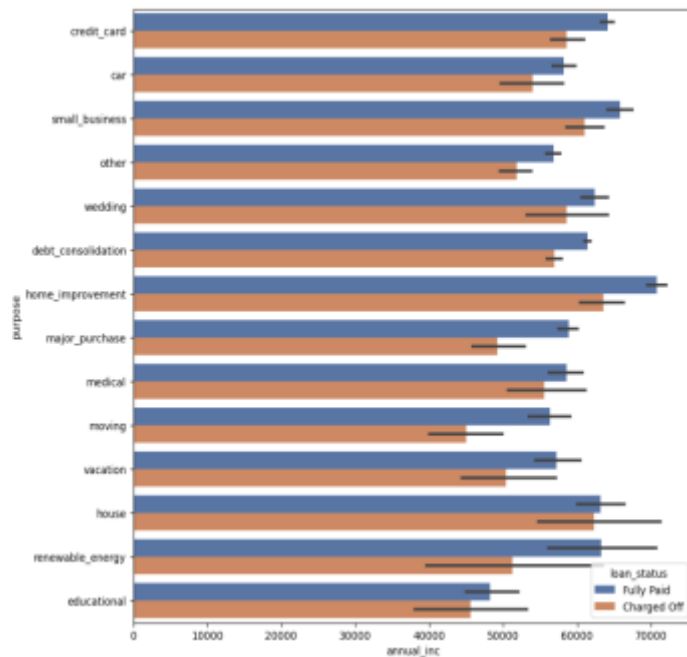
Across most loan purposes, borrowers who have fully paid their loans tend to have higher annual incomes compared to those who have charged off (defaulted) loans. This trend is consistent, showing that higher income levels are generally associated with lower default rates.

This bivariate analysis highlights the importance of annual income as a predictor of loan repayment success. Lenders should consider incorporating income levels more prominently in their risk assessment models, especially for loan purposes that show a higher sensitivity to income variations. Tailored strategies might be needed for different loan purposes to mitigate default risks effectively

Annual income is a significant predictor here of the loan repayment success across different types of home ownership. Higher income levels are consistently associated with fully paid loans, indicating lower default rates

Lenders should consider incorporating income levels prominently in their risk assessment models. The significant income differences, especially in the mortgage category, suggest that lenders might need to pay closer attention to income when evaluating mortgage loan applications
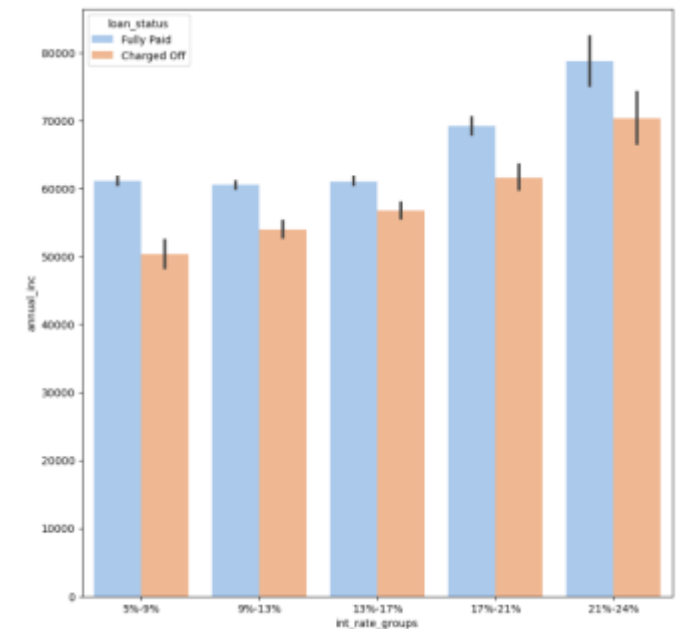
# Bivariate Analysis - Annual income vs Loan Amount & Interest Rate
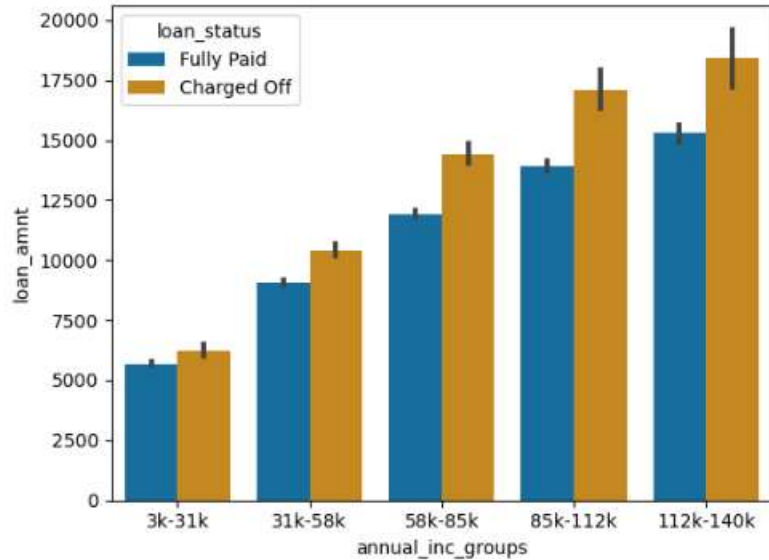
Across all annual income groups, borrowers who have fully paid their loans generally have slightly lower loan amounts compared to those who have charged off (defaulted) loans. The trend shows that higher loan amounts are associated with higher default rates across different income groups.

Larger loan amounts are associated with higher default rates across all income groups, suggesting that lenders should carefully consider the loan amount in their risk assessment models. The increasing difference in loan amounts between fully paid and charged off loans in higher income groups highlights the need for tailored risk management strategies for different income brackets

The sensitivity to income levels is more pronounced at higher interest rates (17%-24%), indicating that income plays a crucial role in the ability to manage higher interest rate loans
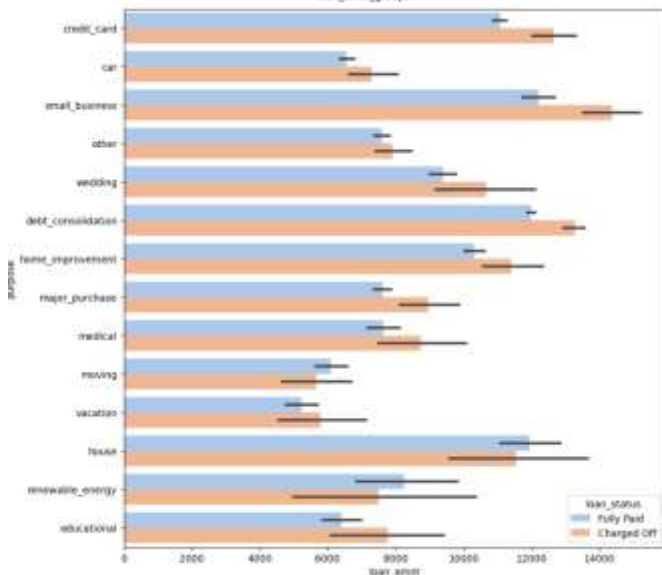
Bivariate Analysis - Loan Amount vs Interest Rate & Purpose
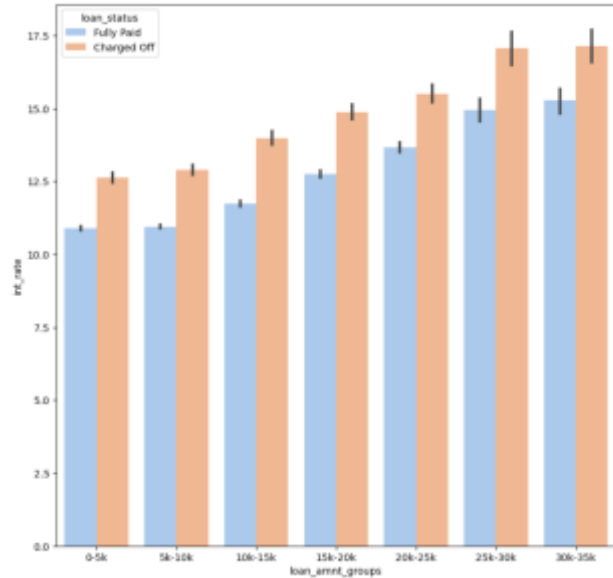
Across all loan amount groups, borrowers who have charged off (defaulted) loans generally have higher interest rates compared to those who have fully paid their loans. This trend shows that higher interest rates are associated with higher default rates across different loan amount groups.

The increasing difference in interest rates between fully paid and charged off loans in higher loan amount groups highlights the need for specified risk management strategies to adapt to avoid default risks effectively


Across various loan purposes, borrowers who have charged off (defaulted) loans generally have higher loan amounts compared to those who have fully paid their loans

Certain loan purposes like credit card refinancing, debt consolidation, small business, and home improvement show larger differences in loan amounts between fully paid and charged off loans.

Bivariate Analysis – Loan Amount vs Issue MM/YYYY & Employment Length and Verification Status
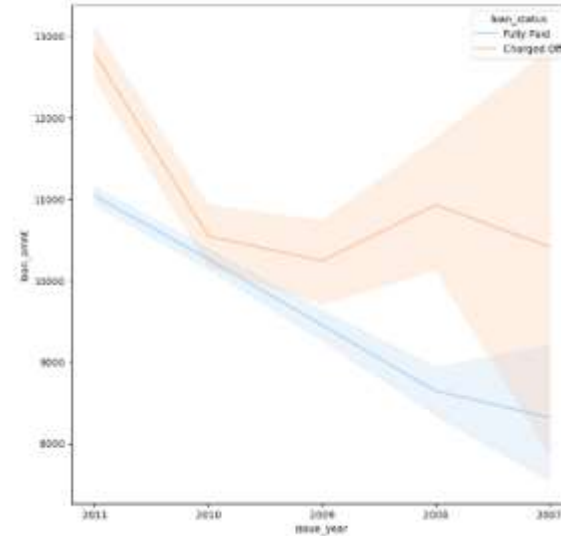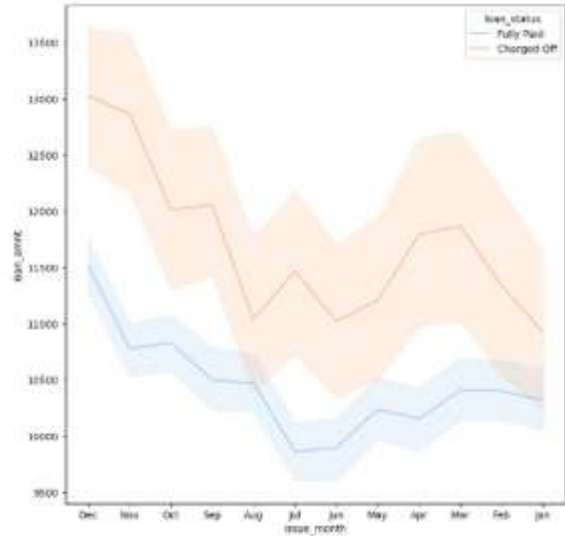
The loan amounts for both fully paid and charged off loans exhibit some seasonal variation throughout the year. The seasonal variations in loan amounts suggest that certain times of the year, particularly around December, are associated with higher loan amounts and potentially higher default rates

The declining trend in loan amounts over the years suggests that loan amounts have been more conservative in recent years, possibly due to stricter lending criteria or probably changes in economic conditions
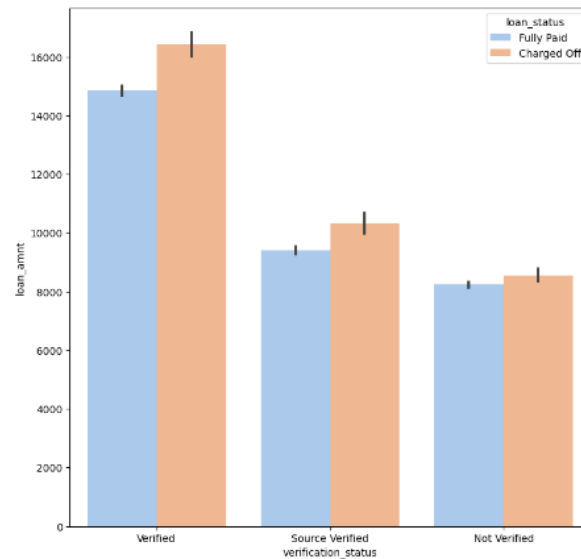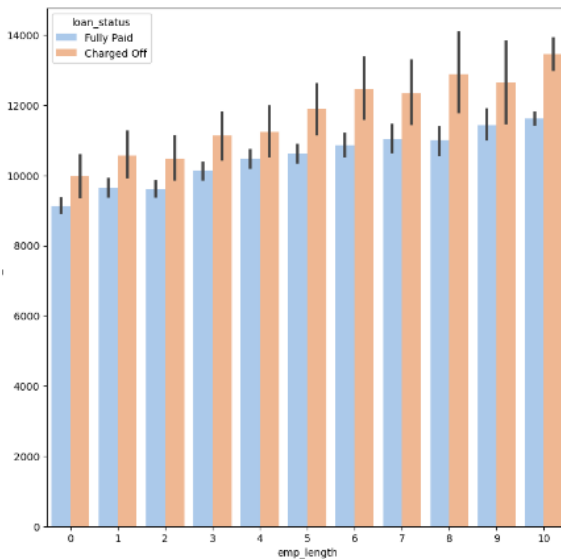
The difference in loan amounts between fully paid and charged off loans becomes more pronounced as employment length increases, suggesting that longer employment length might be associated with higher loan amounts and potentially higher default risks

The verification status also plays a significant role, with verified loans showing the largest gap between fully paid and charged off loan amounts. This suggests that verification might be a critical factor in risk assessment.

# Bivariate Analysis –Interest Rate vs Grades & Loan Amount
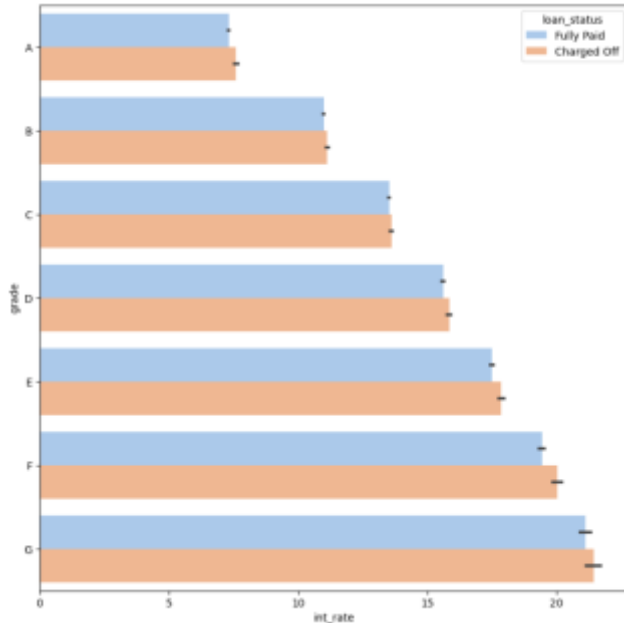
Across different loan grades, charged off (defaulted) loans generally have higher interest rates compared to fully paid loans.

The trend indicates that higher interest rates are associated with higher default rates across different loan grades

Higher interest rates are consistently associated with higher default rates across all loan grades. This indicates that higher interest rates pose a greater risk of default, regardless of the grade

Higher interest rates are consistently associated with higher default rates across all loan amount groups. This indicates that higher interest rates pose a greater risk of default, regardless of the loan amount
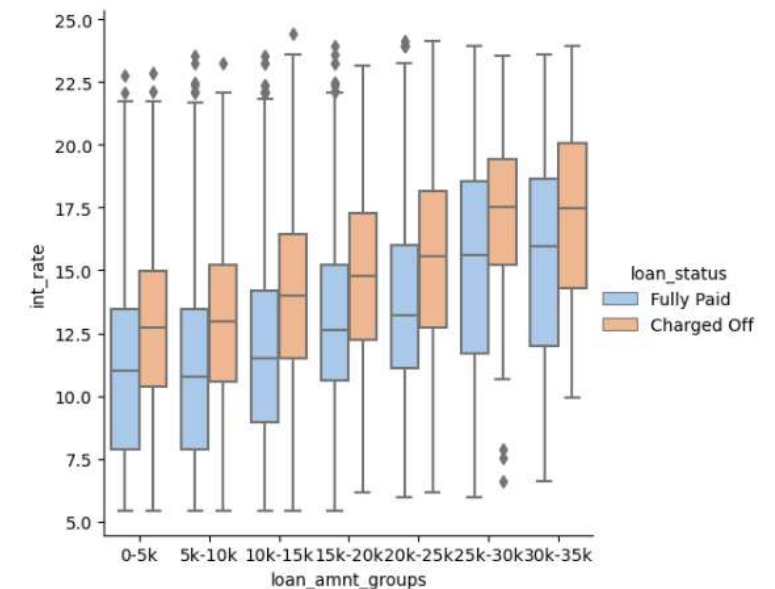
SUGGESTIONS:

- Risk Assessment:
  - The analysis indicates that higher loan amounts are more likely to be charged off, as evidenced by the data at the higher end of the loan amount spectrum. This information could be useful for risk assessment and loan approval processes.
- Loan Amount Limits:
  - Lenders might consider implementing stricter criteria for higher loan amounts to mitigate the risk of defaults, given that higher loan amounts have shown a tendency to be charged off.
- Policy Adjustments:
  - Adjusting lending policies for loans above a certain threshold could be beneficial in reducing the overall default rate. This could include more stringent credit checks, or additional collateral requirements for higher loan amounts, lower than 12 DTI etc.
- Verifications:
  - Process of income verification should be reinstated. From the data, it can be seen that even verified income has high number of charged off.
- Misc:
  - The purpose of the loan and the time of the year during which loan is requested should be considered with reference to the past data.
  - Thorough review of credit utilization rate.