

Lending Club Case Study

KIRAN SUMIT DALMIYA

Contents

- Problem Statement
- Objective of Analysis
- analysis approach
- Business insight with data visualisation
- Recommendations
- conclusion

Problem Statement

This assignment is about using Exploratory Data Analysis (EDA) to solve a real business challenge in a consumer finance company.

The goal is to understand and minimize the risk of losing money while lending to customers. In the dataset provided, which includes information about past loan applicants, we aim to identify patterns that indicate if a person is likely to default on a loan.

Key points include:

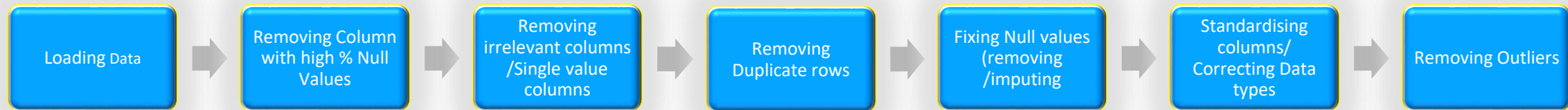
- The company faces risks in loan decisions:
 - losing business if a likely-to-repay applicant is rejected
 - financial loss if a likely-to-default applicant is approved.
- Through EDA, we want to find patterns that help make informed decisions, such as denying loans or adjusting interest rates for risky applicants.
- The dataset includes loan acceptance scenarios (fully paid, current, charged-off) and rejected loans.
- The focus is on understanding how consumer and loan attributes influence the tendency of default, ultimately optimizing lending decisions.

Objective

- This case study aims to empower the company a leading online loan marketplace with insights into loan default drivers, thereby enabling proactive measures to minimize credit loss and strengthen risk management strategies.
- Leveraging Exploratory Data Analysis (EDA), the focus is on understanding key indicators of default empowering the company to optimize risk assessment and portfolio management.

Analysis Approach

Data Cleaning and Preprocessing:



Analysis Approach

Univariate Analysis:

Explore the distribution of key variables like loan status, loan amount, and other attributes.

Bivariate Analysis:

Investigate relationships between variables, especially focusing on:

- Correlation between numerical features.
- Impact of categorical variables on loan status.

Identify key factors influencing loan default.

Multivariate Analysis:

Examine interactions between multiple variables to uncover complex patterns.

Consider the combined effect of attributes on the likelihood of default.

Visualization:

Create visual summaries to aid in conveying insights effectively.

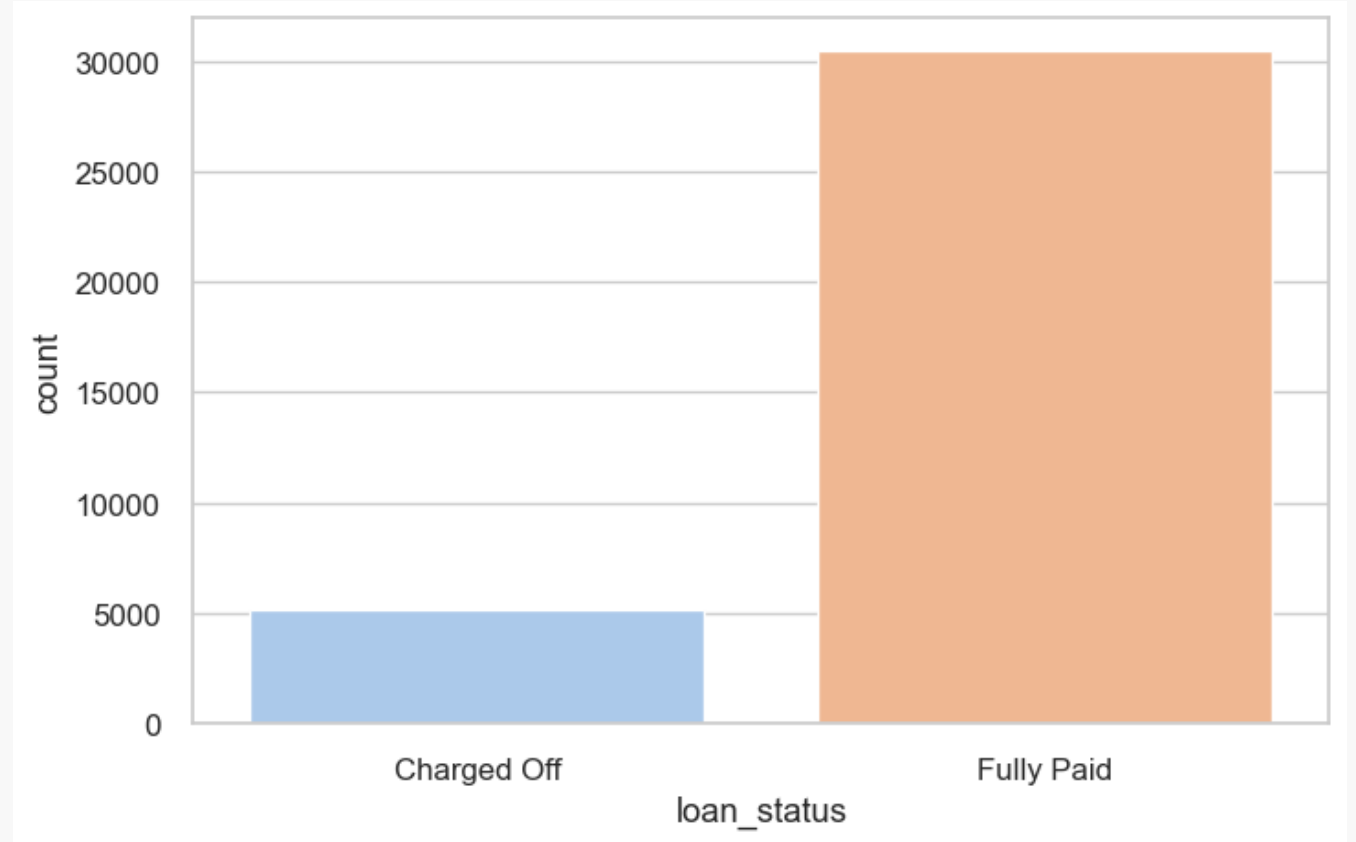
Business Interpretation:

- Translate findings into actionable insights for the business.
- Recommend strategies for risk mitigation and optimizing lending decisions.

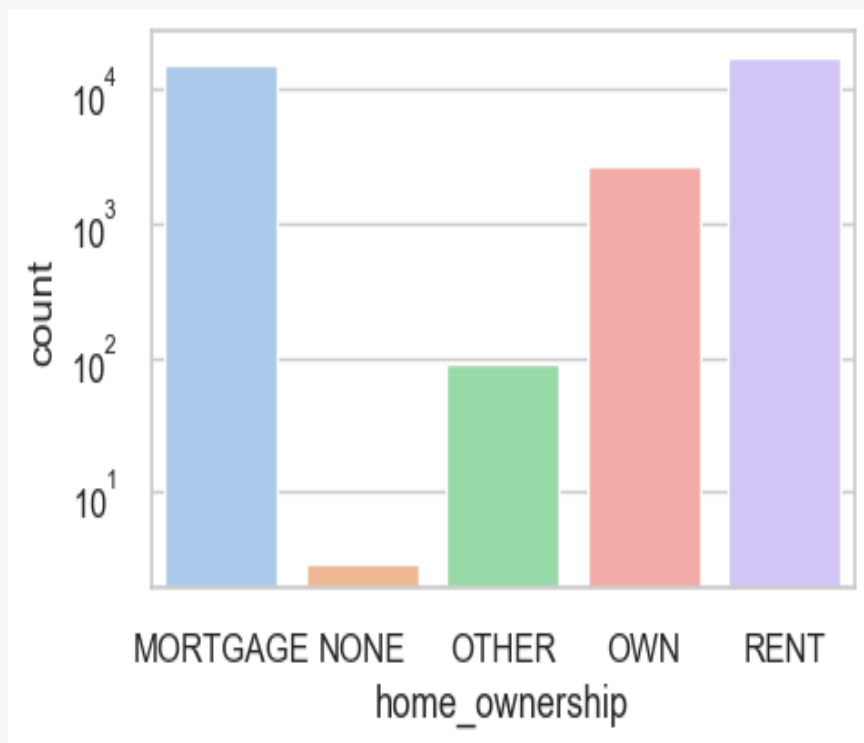
Data Visualisation using Univariate , Bivariate and Multivariate Data Analysis

UDA on loan status column

- Charged off loans are less compare to fully paid loans



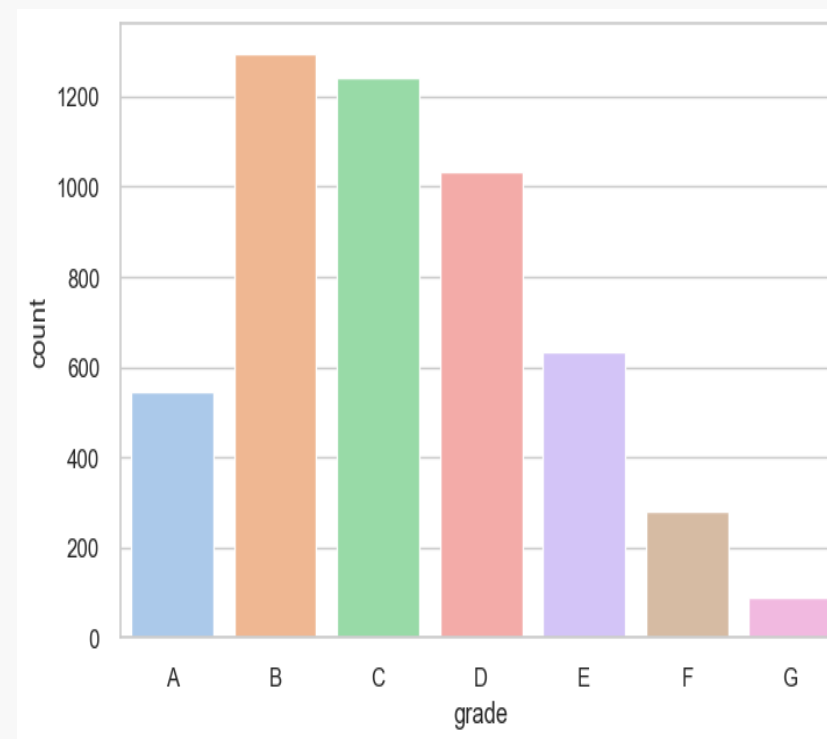
Home Ownership Frequency plot



Observation:

Most borrowers do not own property and instead are in a mortgage or rental arrangement.

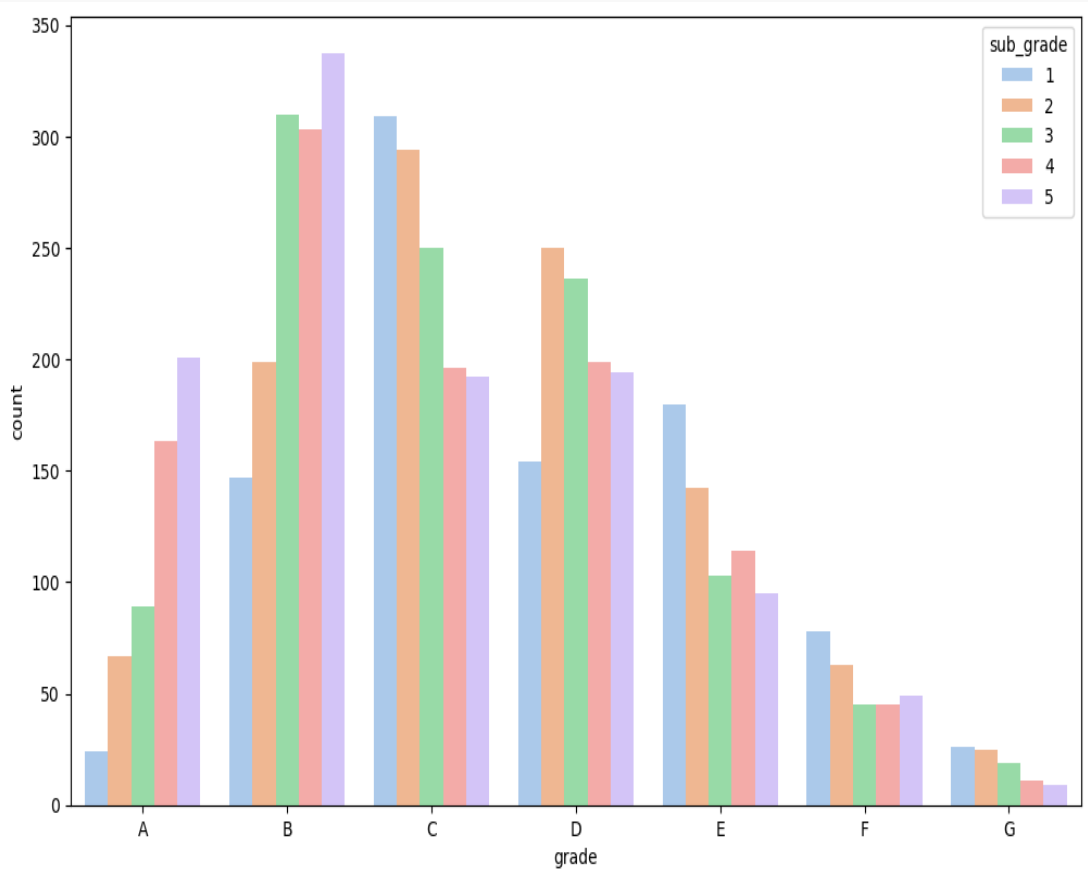
Grade for Charged off loans Frequency plot :



Observation:

A large amount of loans are with grade B,C and D

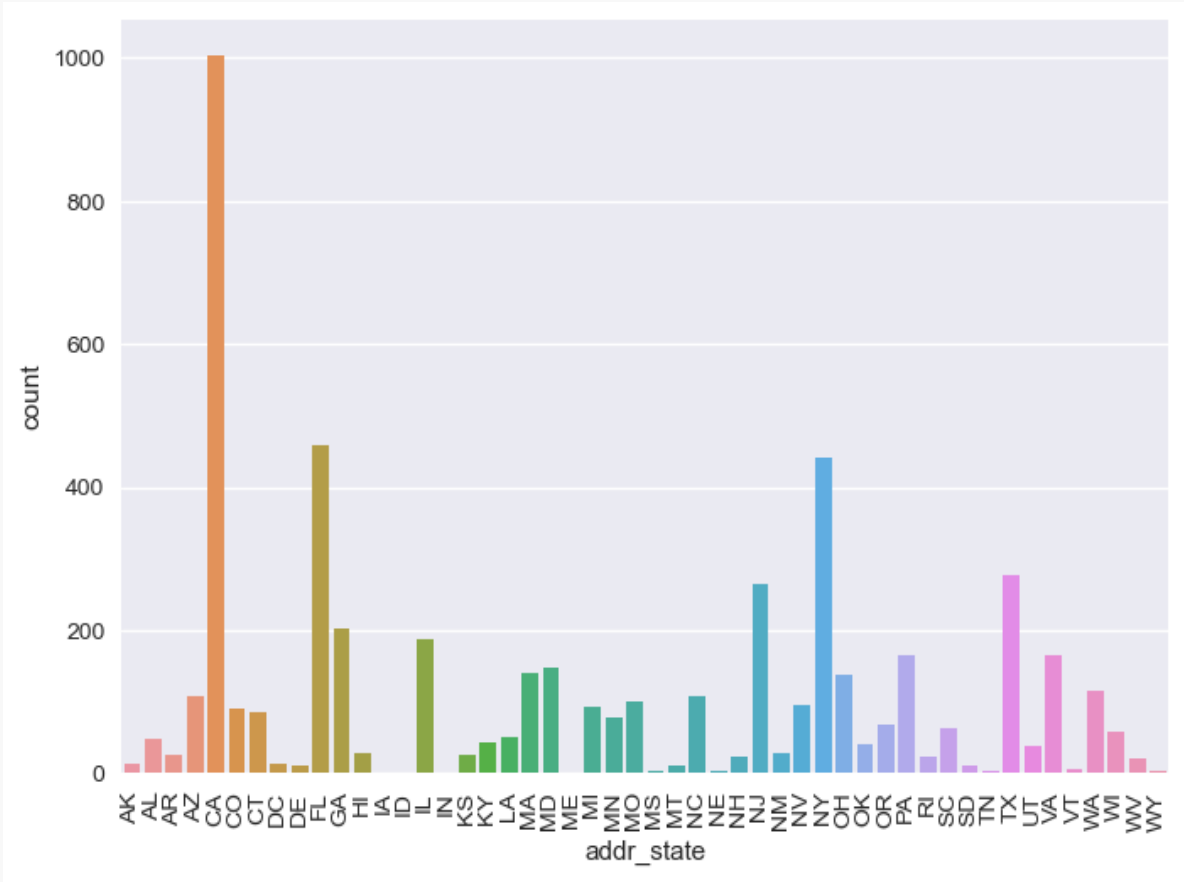
Sub grade for charged-off loans Frequency plot



Observation:

A large amount of loans are with grade B5,B3,C1

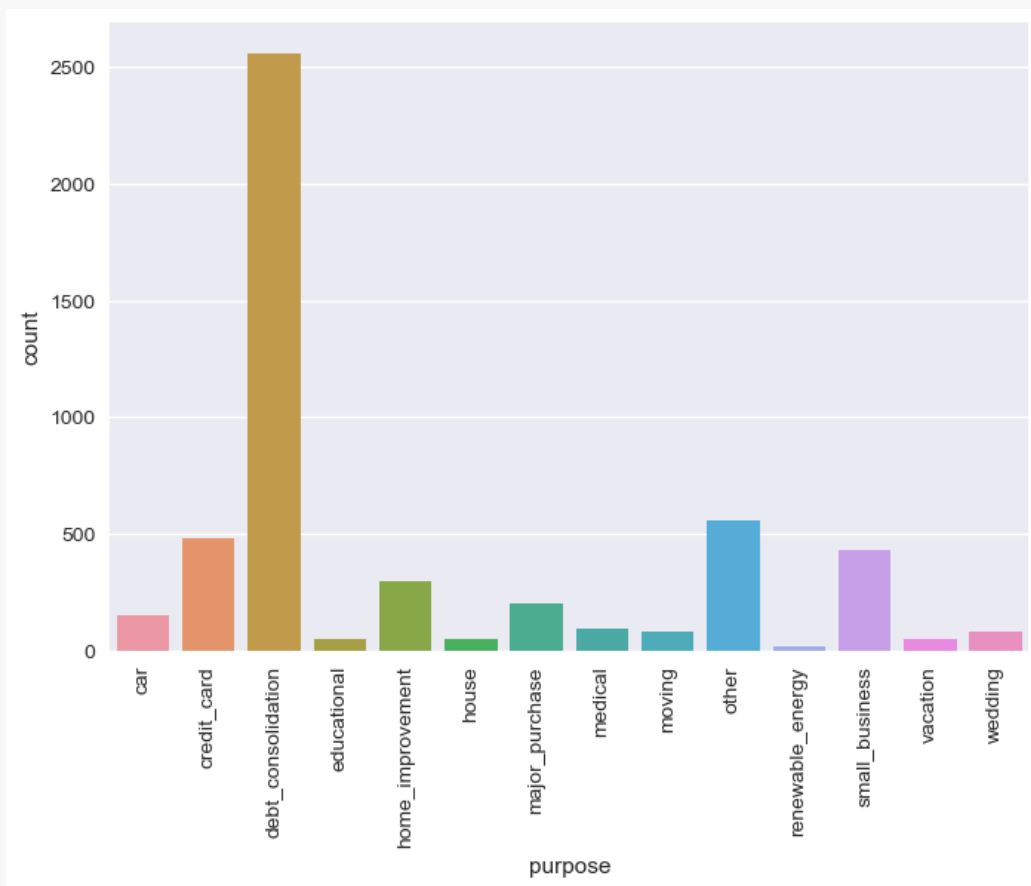
Grade for Charged off loans Frequency plot :



Observation:

A large amount of loans are with grade B,C and D

Purpose for charged-off loans Frequency plot

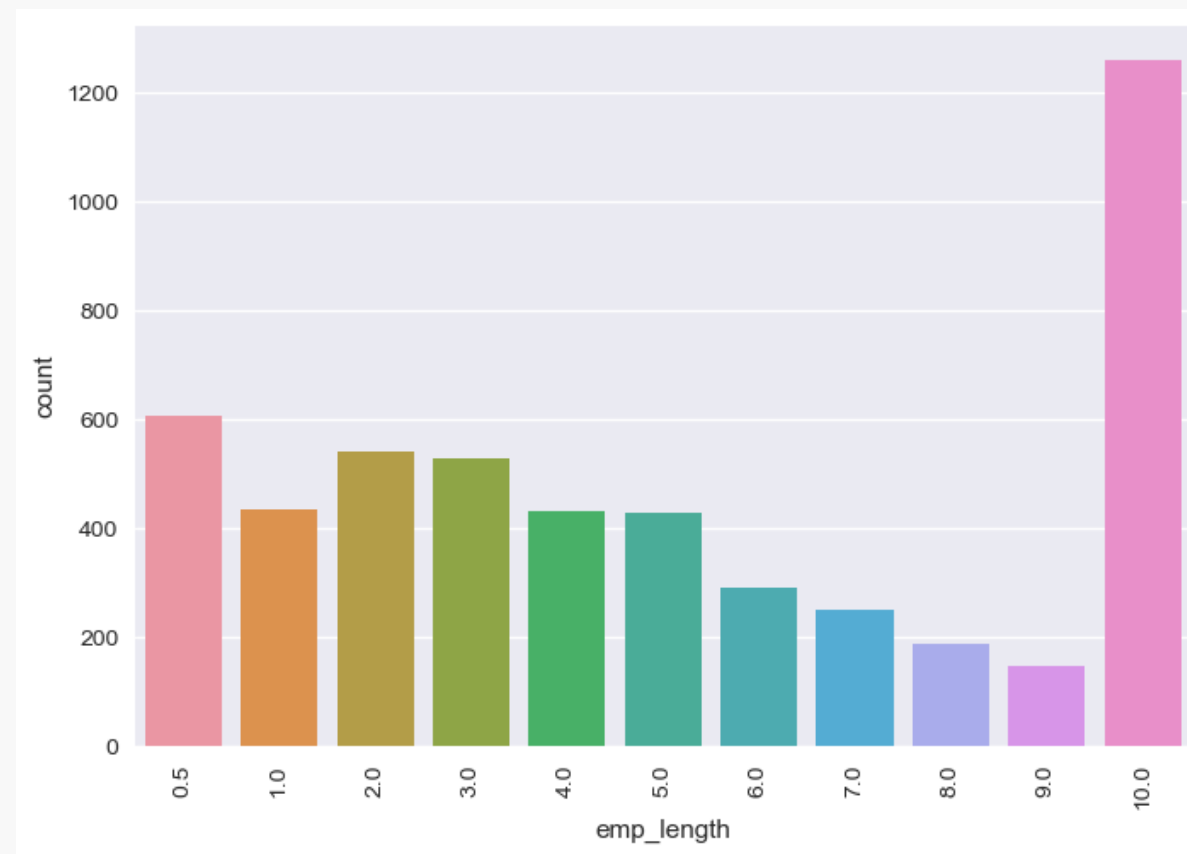


Observation:

Debt consolidation was the top reason for "Charged off" loans, then debit card, small business.

The lending company should be careful when approving loans for debt consolidation, as it led to the most defaults.

Emp_length for Charged off loans Frequency plot :

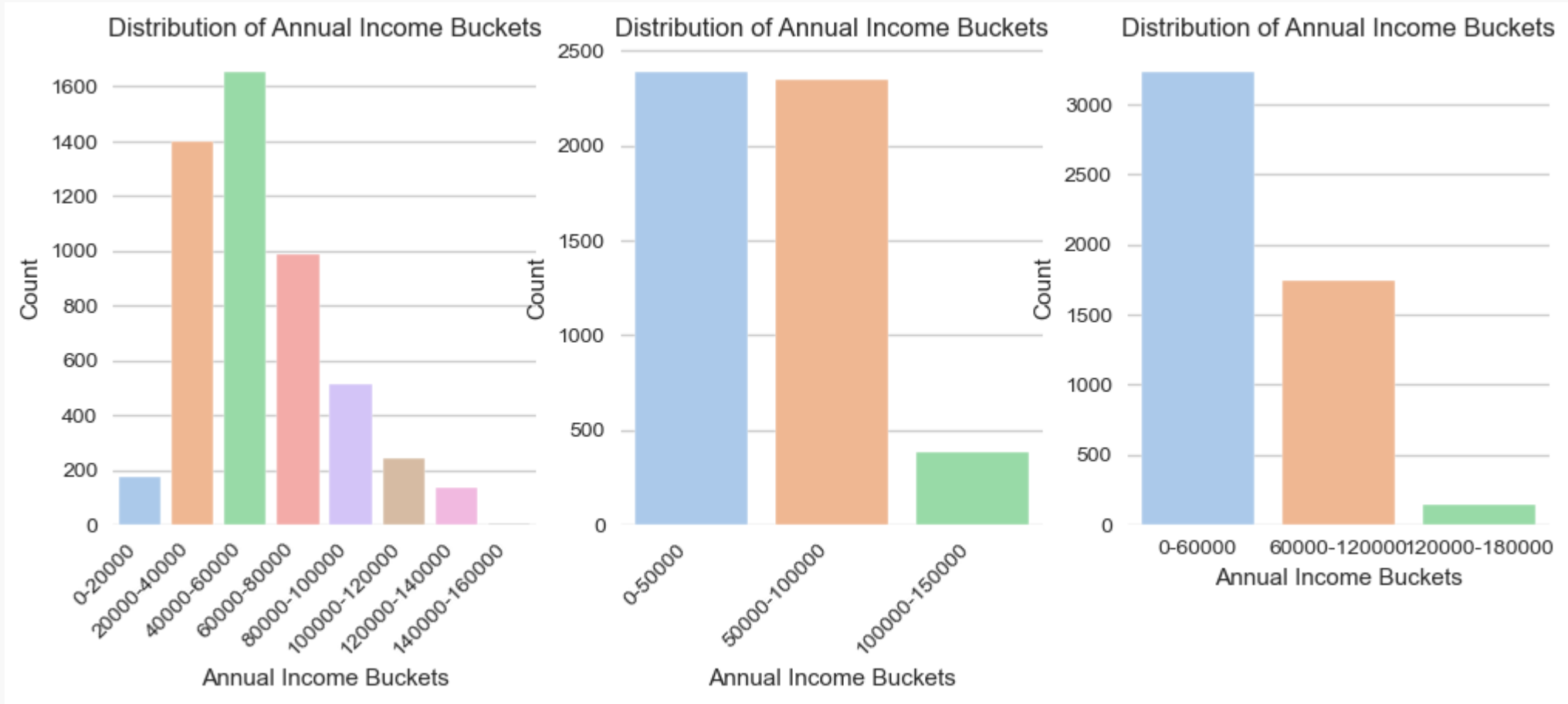


Observation:

Applicants with over 10 years of work experience had the highest number of "Charged off" loans.

This suggests that even with a long employment history, successful loan repayment wasn't guaranteed.

Annual Income for charged-off loans Frequency plot



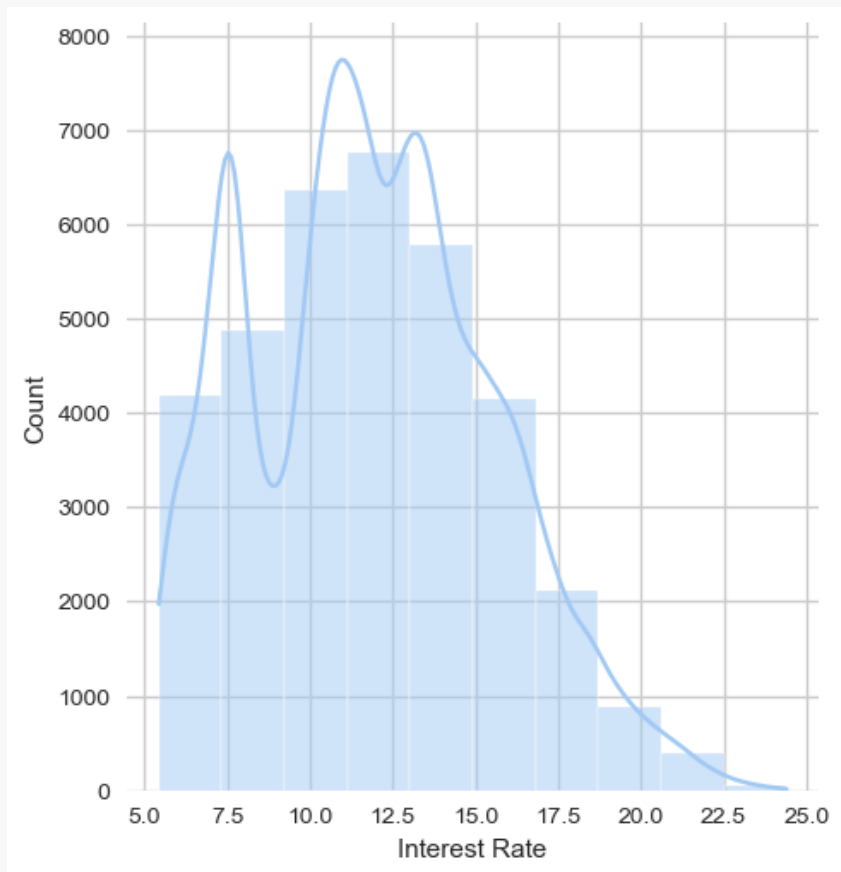
Observation:

Loan applicants who faced repayment issues ("charged off") typically earned less than 60,000 USD more specifically earn between 20000 to 600000 annually. (This can be because people with less then 20000 annul income are hardly taking any bigger loan.)

For individuals within this income range, it is advisable for the lending company to exercise prudence by implementing stringent income verification processes and conducting thorough assessments of repayment capacity.

Later we will see bivariate analysis of defaulters by income buckets and loan amounts .

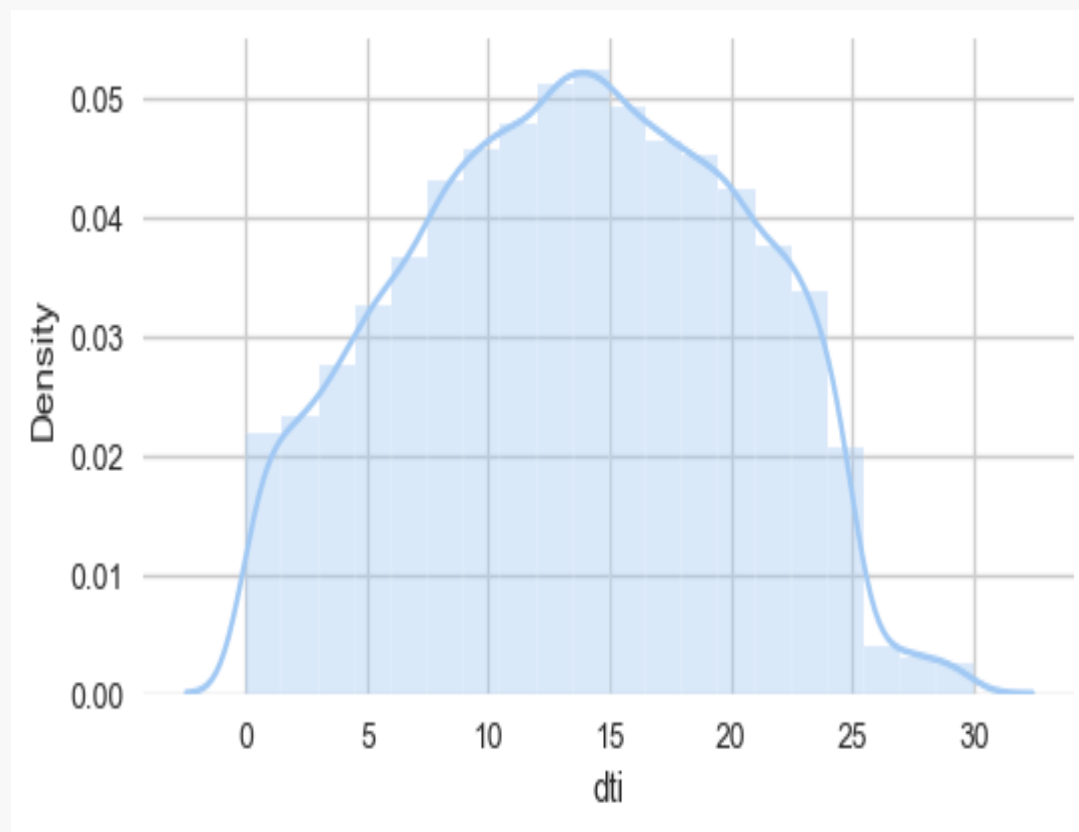
Interest for charged-off loans Frequency plot



Observation:

Defaulter decreases as the interest rate increases after 15.0 % of interest rate...
for lower interest rate it keep increasing ; only exception is at around 10.0% interest rate where defaulter rate fell.

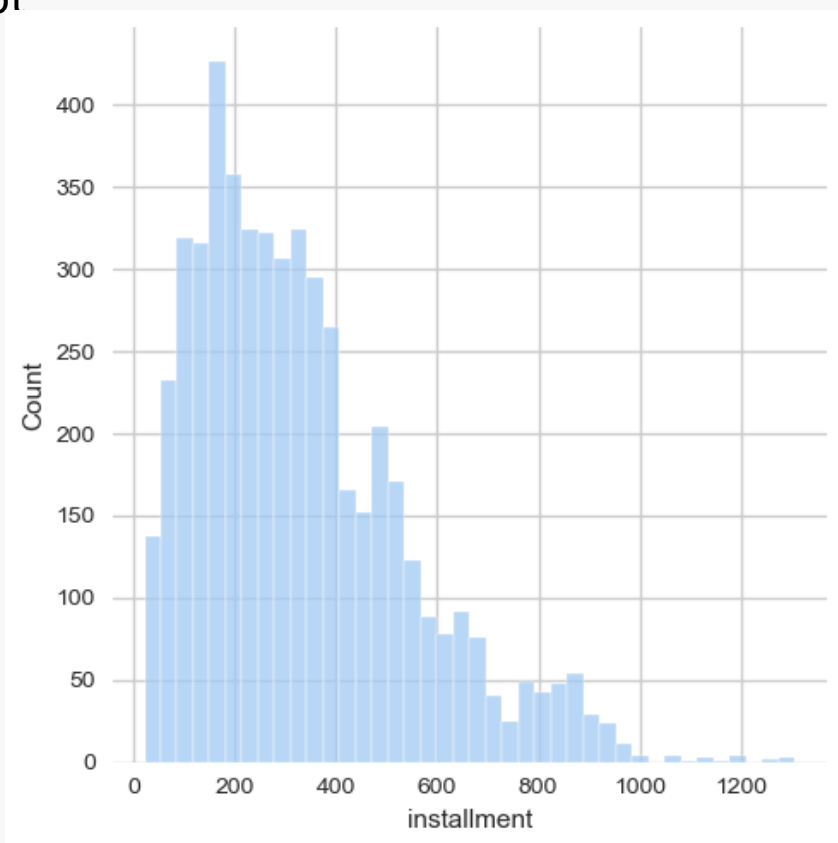
dti for Charged off loans Frequency plot :



Observation:

Most borrowers have a big difference between their debt and income, especially in the 10-15 DTI range.

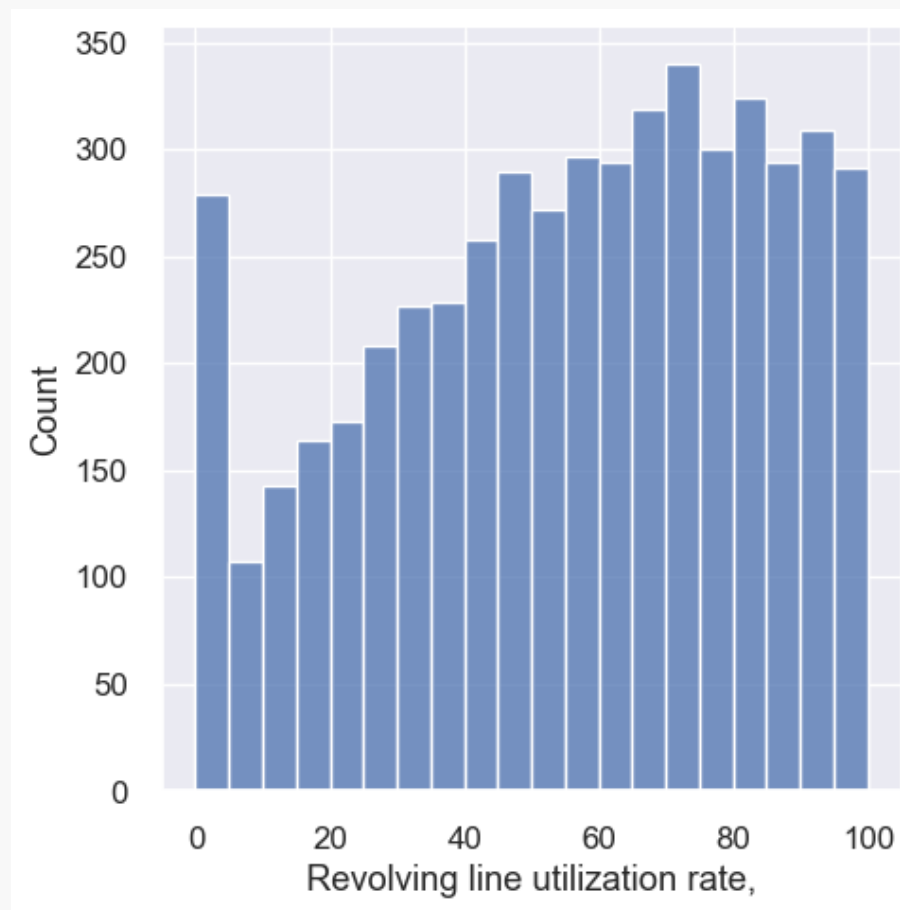
Installment for charged-off loans Frequency plot



Observation:

Loan participants who defaulted commonly had monthly installment amounts ranging from 150 to 450 USD. To minimize the risk of loan defaults, the lending company should carefully monitor and evaluate applicants with similar installment amounts.

Revol_util for Charged off loans Frequency plot :

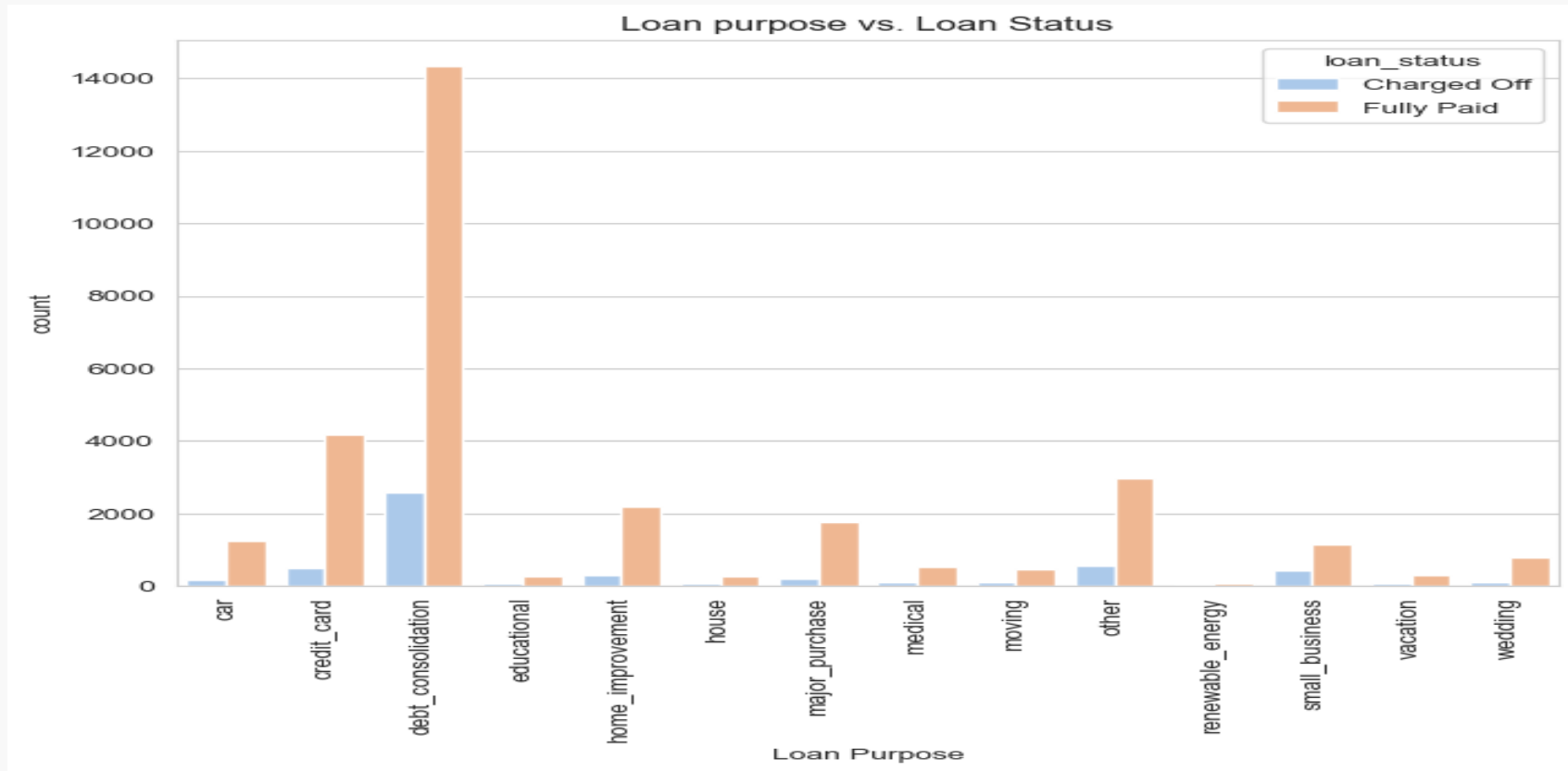


Observation:

The average utilization rate is around 48% Average revolving Utilisation rate is seen to be higher for Charged Off loans than the Fully Paid loans.

Bivariate Data Analysis

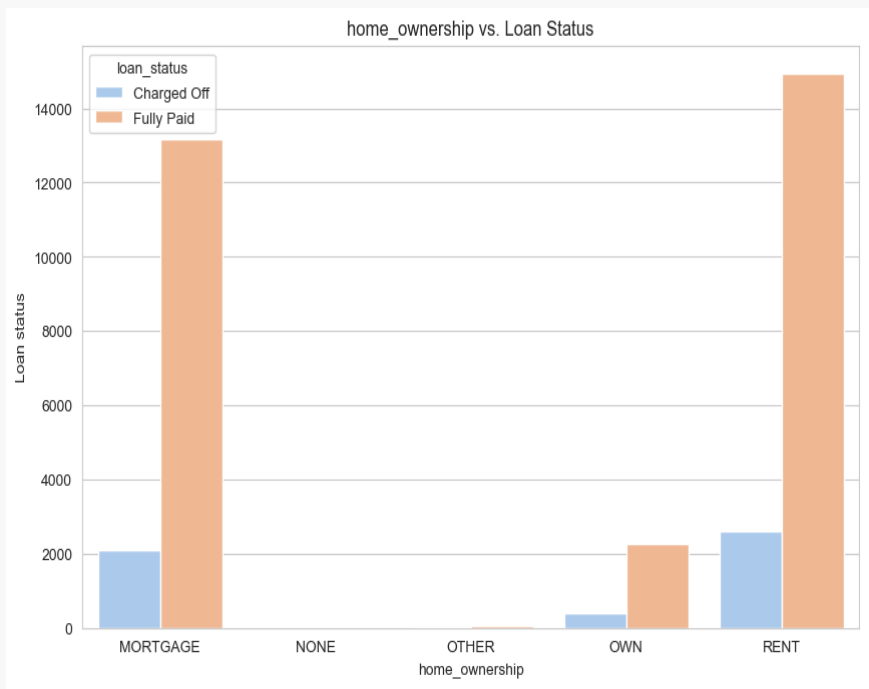
Purpose Vs Loan Status



Observation:

Debt consolidation was the top reason for "Charged off" loans followed by small business. The lending company should be careful when approving loans for debt consolidation as well as small business as it led to the most defaults..

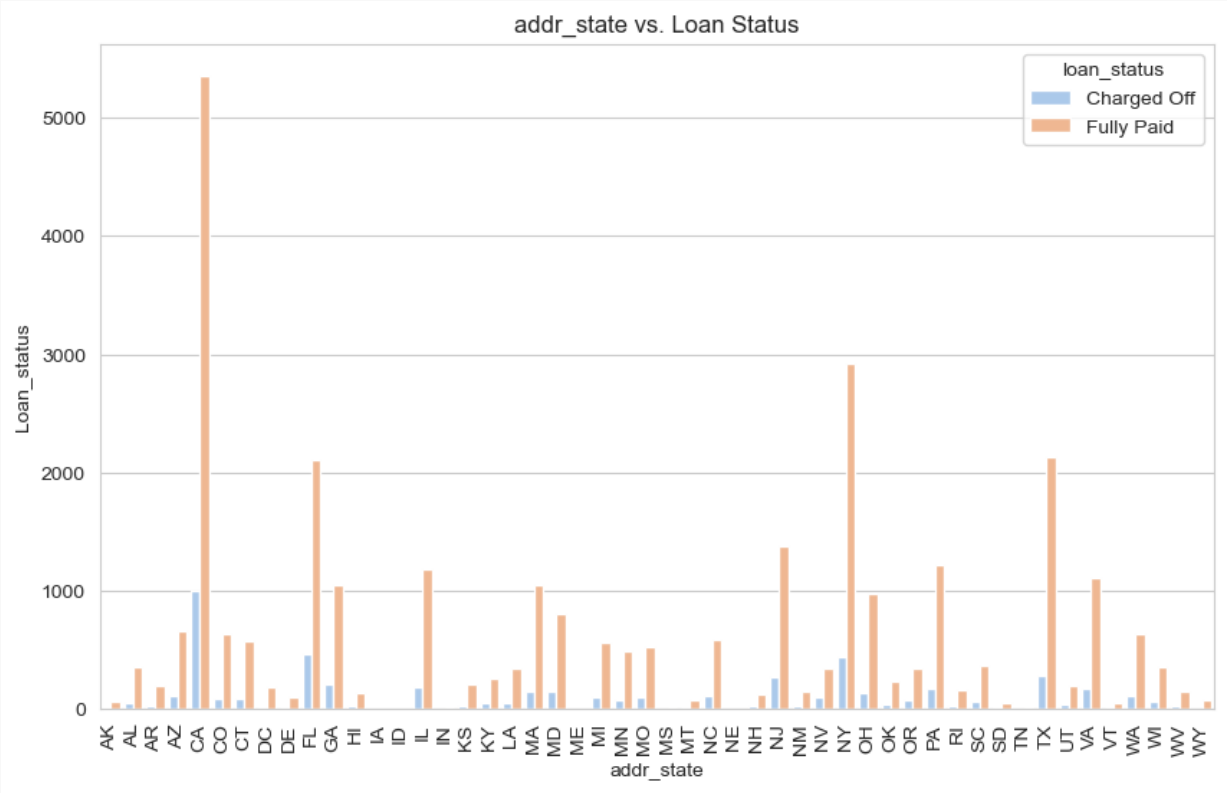
Home Ownership Vs Loan Status



Observation:

Most borrowers who lives in mortgage or rental arrangement are more likey to default.

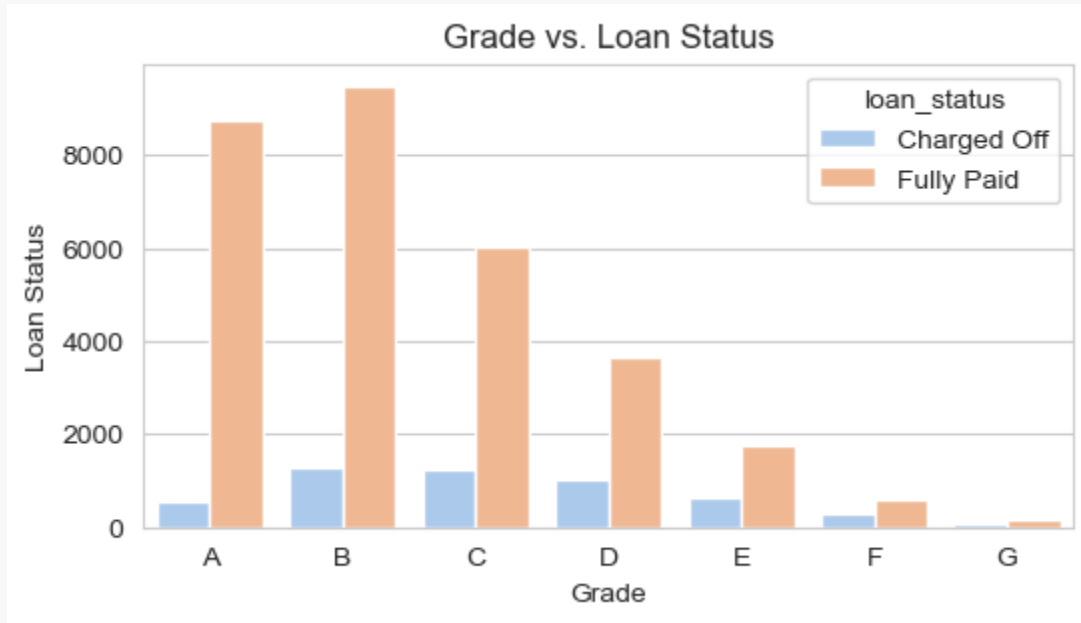
address state Vs Loan Status :



Observation:

Most number of loan are taken from CA state then florida,new york and texas respectively
California had the most "Charged off" loan applicants. To address this, the lending company should consider stricter eligibility criteria or better credit assessments for applicants from California.

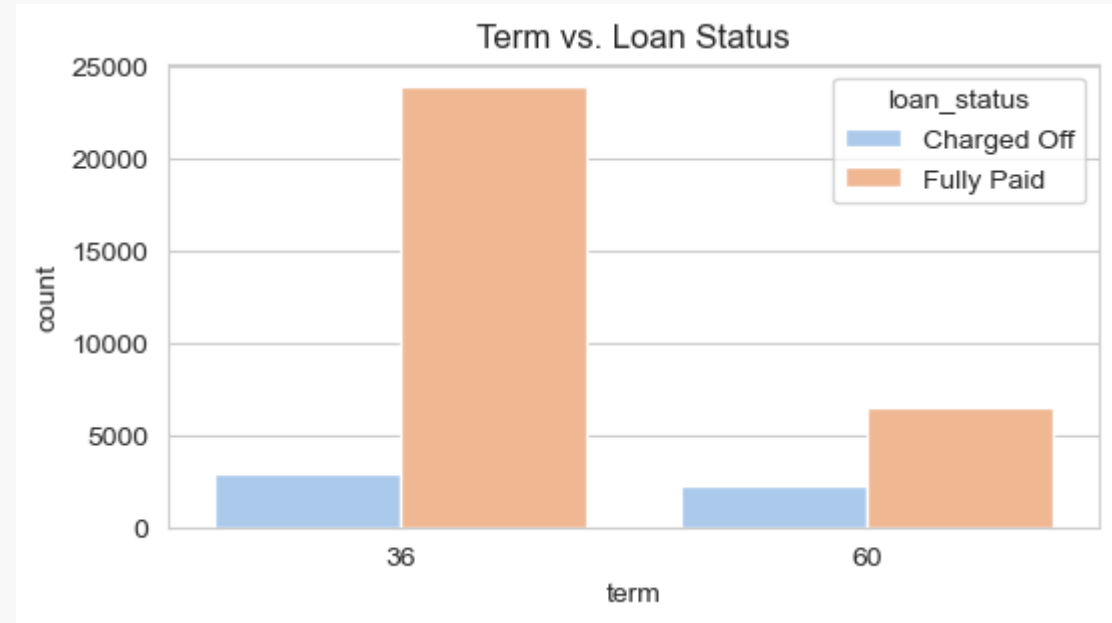
Grade vs Loan status



Observation:

Borrower who are in grade B,C and D are tend to default more

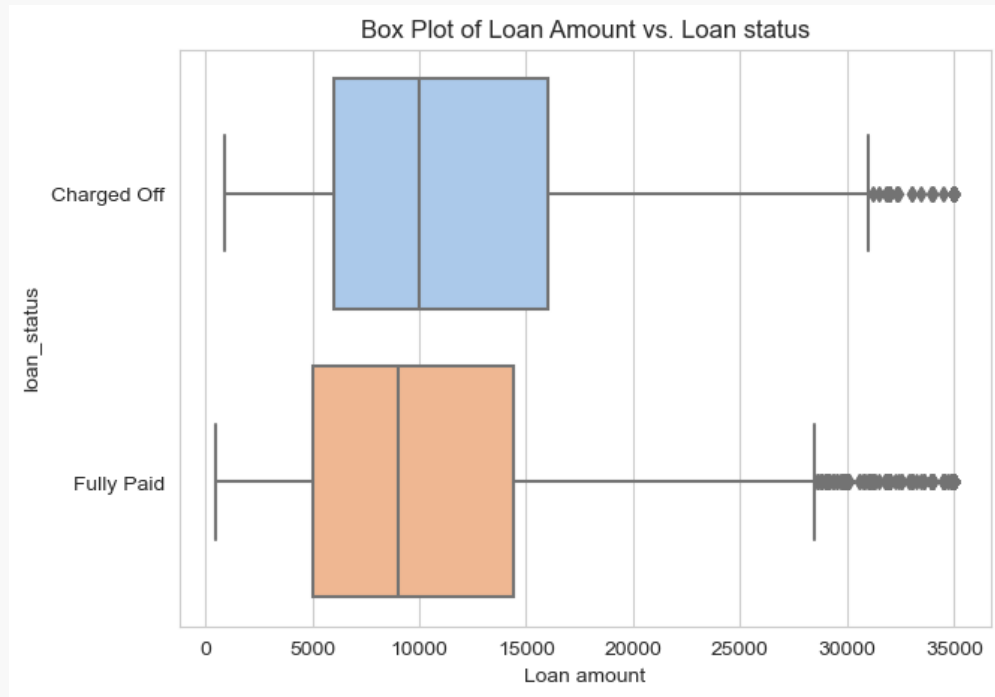
term Vs Loan status:



Observation:

The majority of "Charged off" applicants favored short-term loans lasting 36 months, indicating that a significant number of borrowers facing loan default opted for quicker repayment periods.

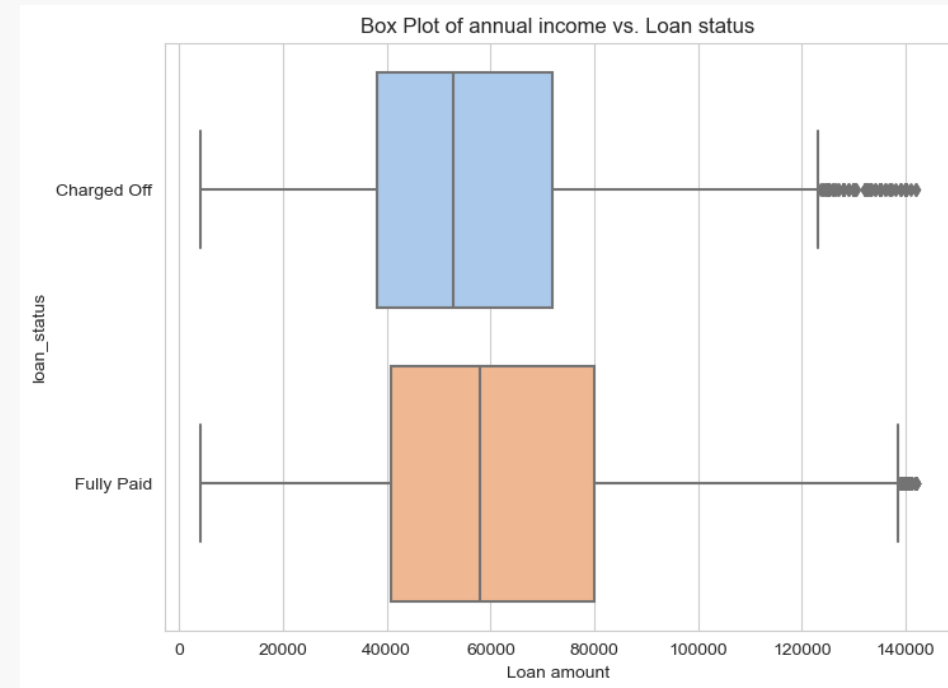
Loan amount vs Loan status



Observation:

The majority of "Charged off" applicants takes high amount loans. Whenever providing high amount loan, company should do more checks different attributes.

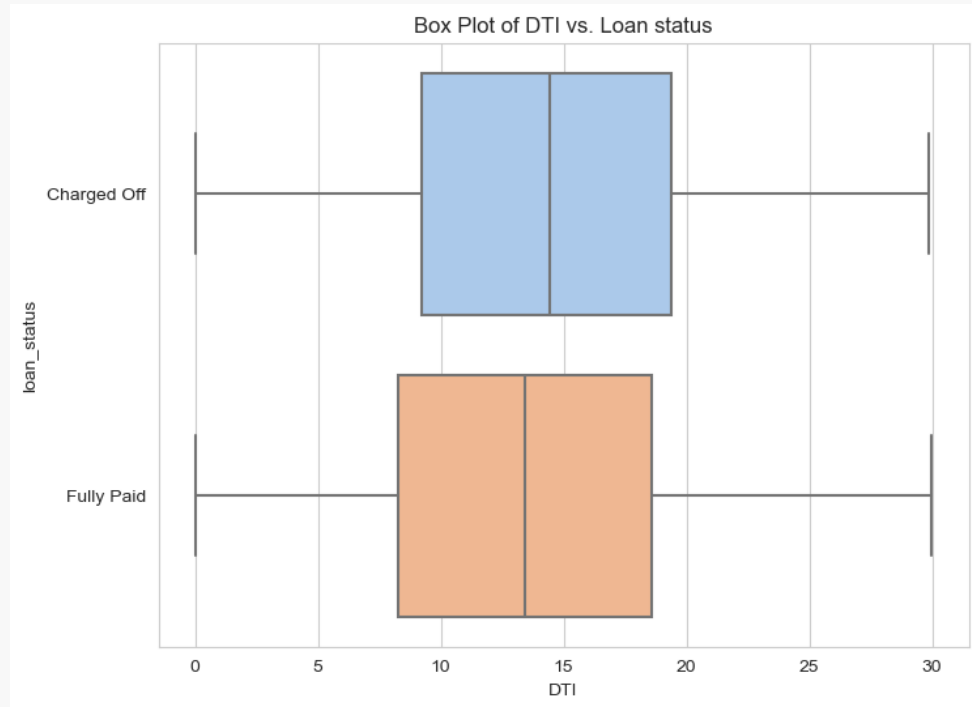
Annul Income Vs Loan status:



Observation:

The majority of "Charged off" applicants have annual income between 38,000 to 73,000.

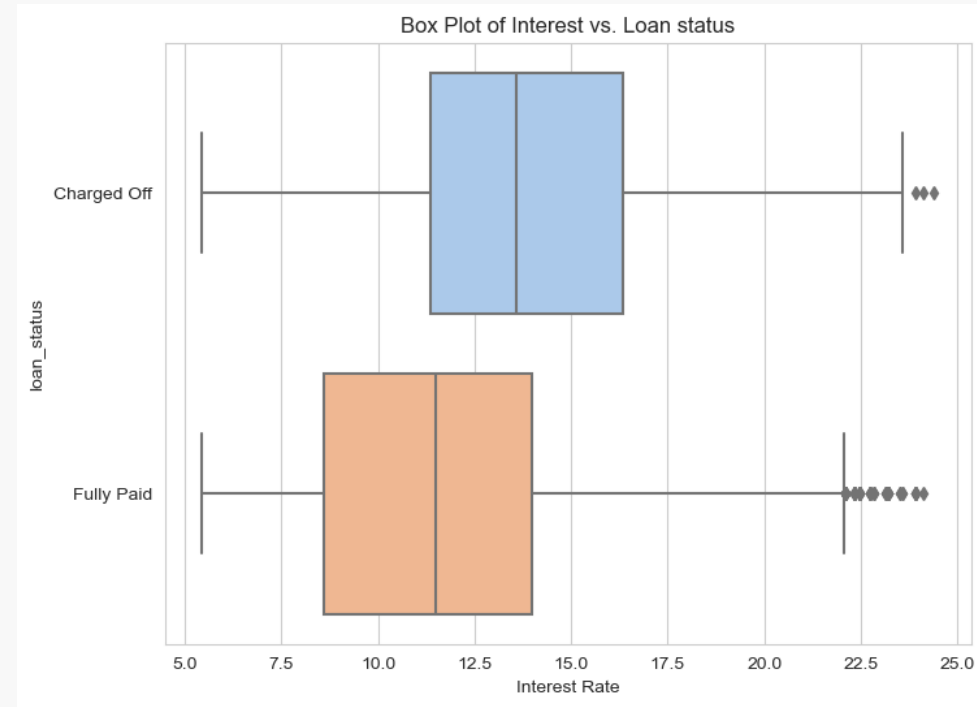
DTI vs Loan status



Observation:

The majority of loan applicants who charged off had significantly high Debt-to-Income (DTI) ratios.

Int_rate Vs Loan status:



Observation:

A significant portion of loan applicants who defaulted received loans with interest rates falling within the range of 12% to 16%.

loan_amnt vs grade



Observation:

Bigger loans usually get lower grades. For instance, a grade G loan has a median amount almost \$10,000 more than grades A, B, or C loans.

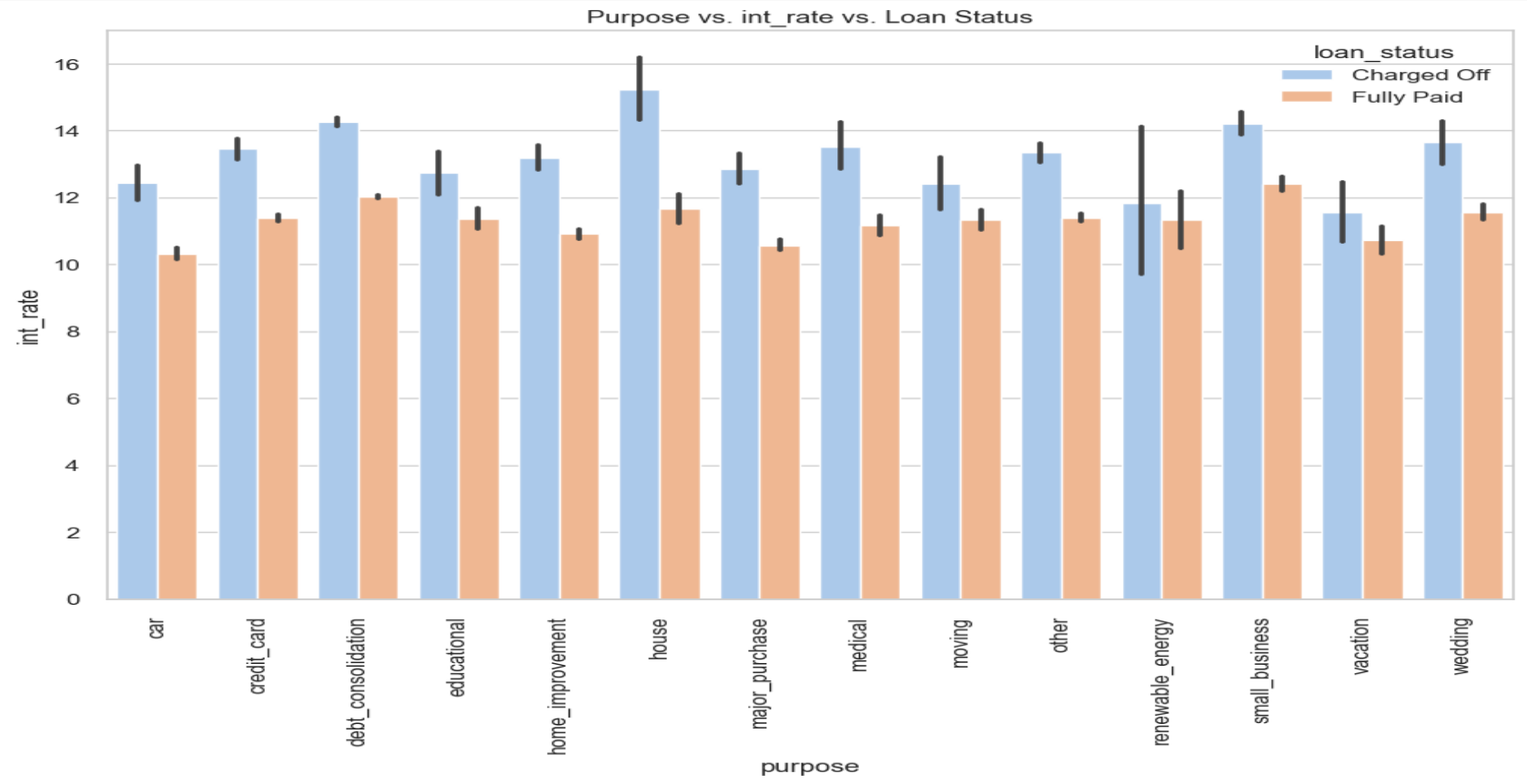
Annual income with Grade



Observation:

High earning people usually get lower grades; By comparing this plot with loan amount vs grade plot we can observe that the people with higher income borrow high amount of loan

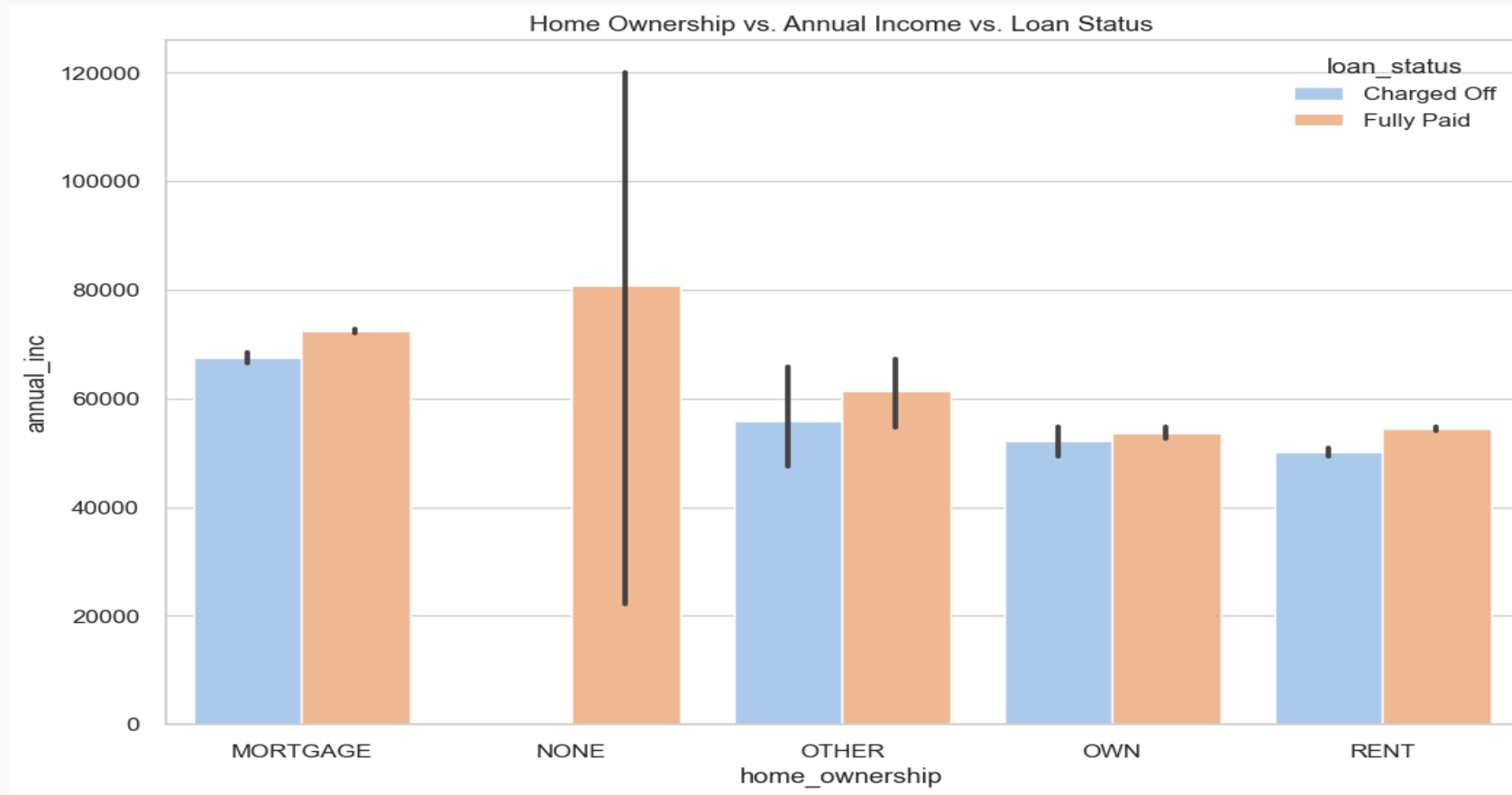
Verification status Vs int_rate Vs Loan status



Observation

When Loan taken on higher rate and for small business,house and debt consolidation are more likely to get defaulted

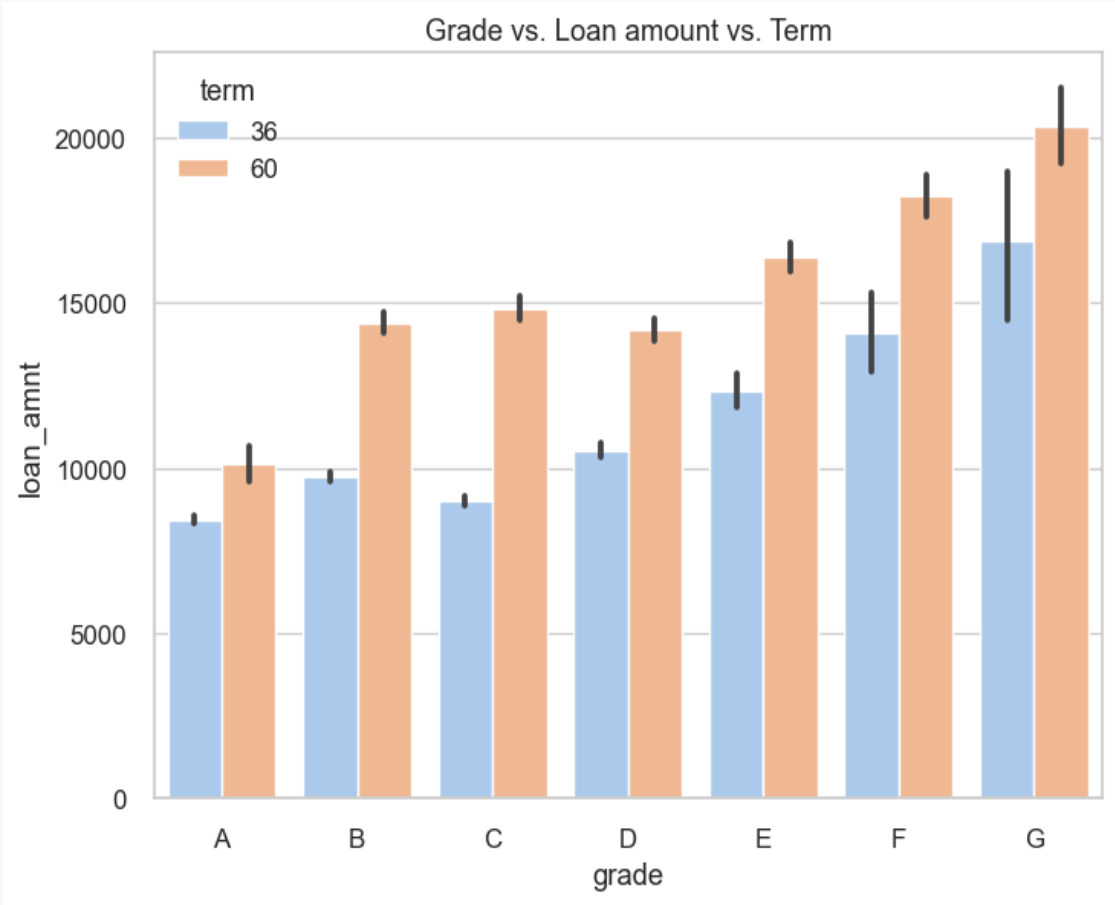
Home Ownership vs. Annual Income vs. Loan Sta



Observation:

Borrower with low annual income and living in mortgage home is more likely to be a defaulter

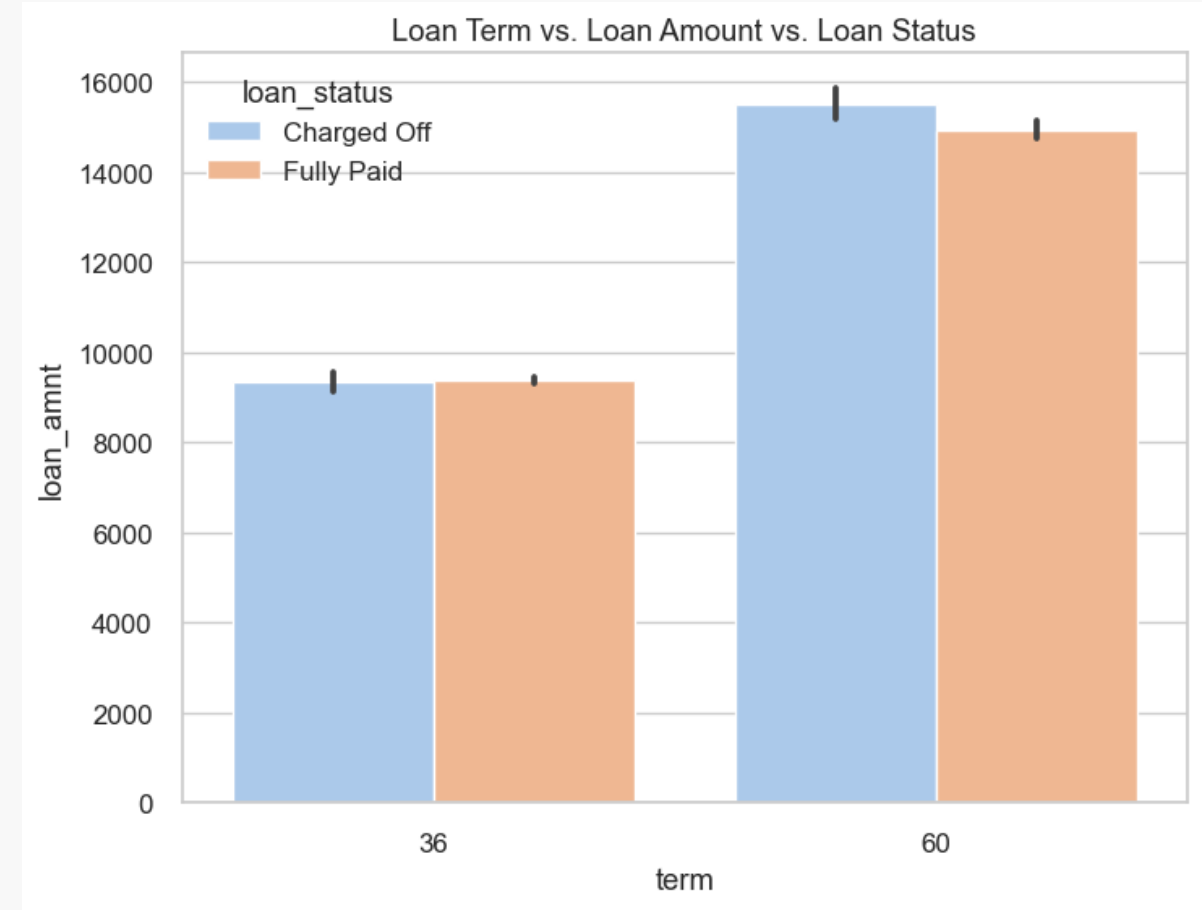
Grade vs. Loan amount vs. Term



Observation:

High earning people usually gets lower grade & borrow high amount of loan and for long term here say 60 months

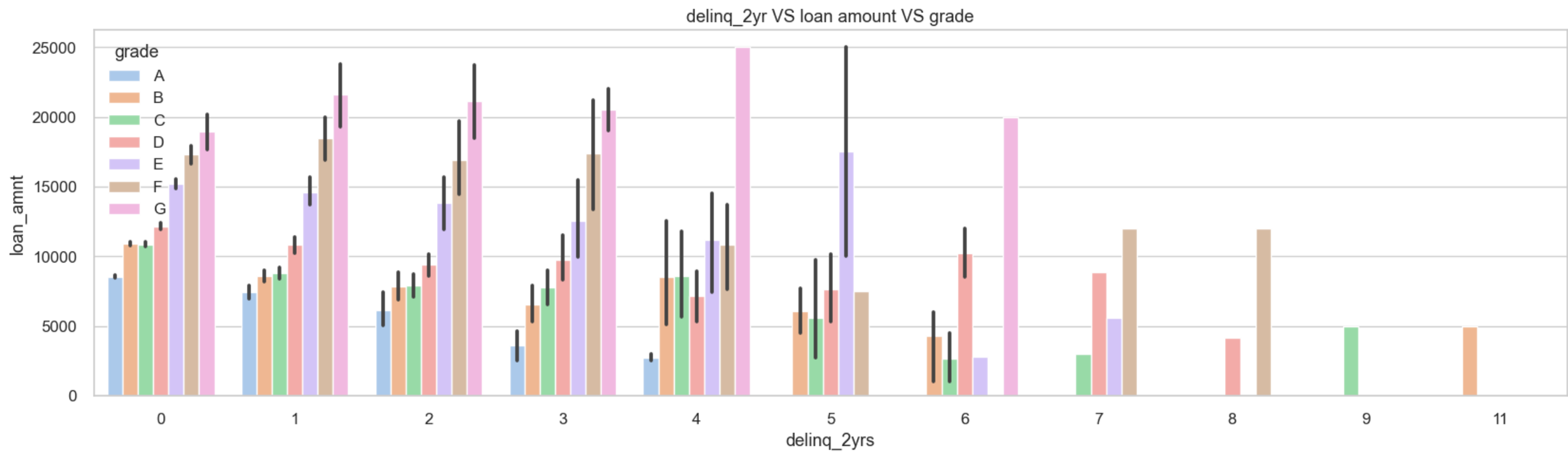
Loan Term vs. Loan Amount vs. Loan Status



Observation:

The trend observed in our univariate analysis becomes more apparent in this plot. It clearly shows that higher loan amounts tend to be associated with longer terms and exhibit a higher occurrence of Charged Offs.

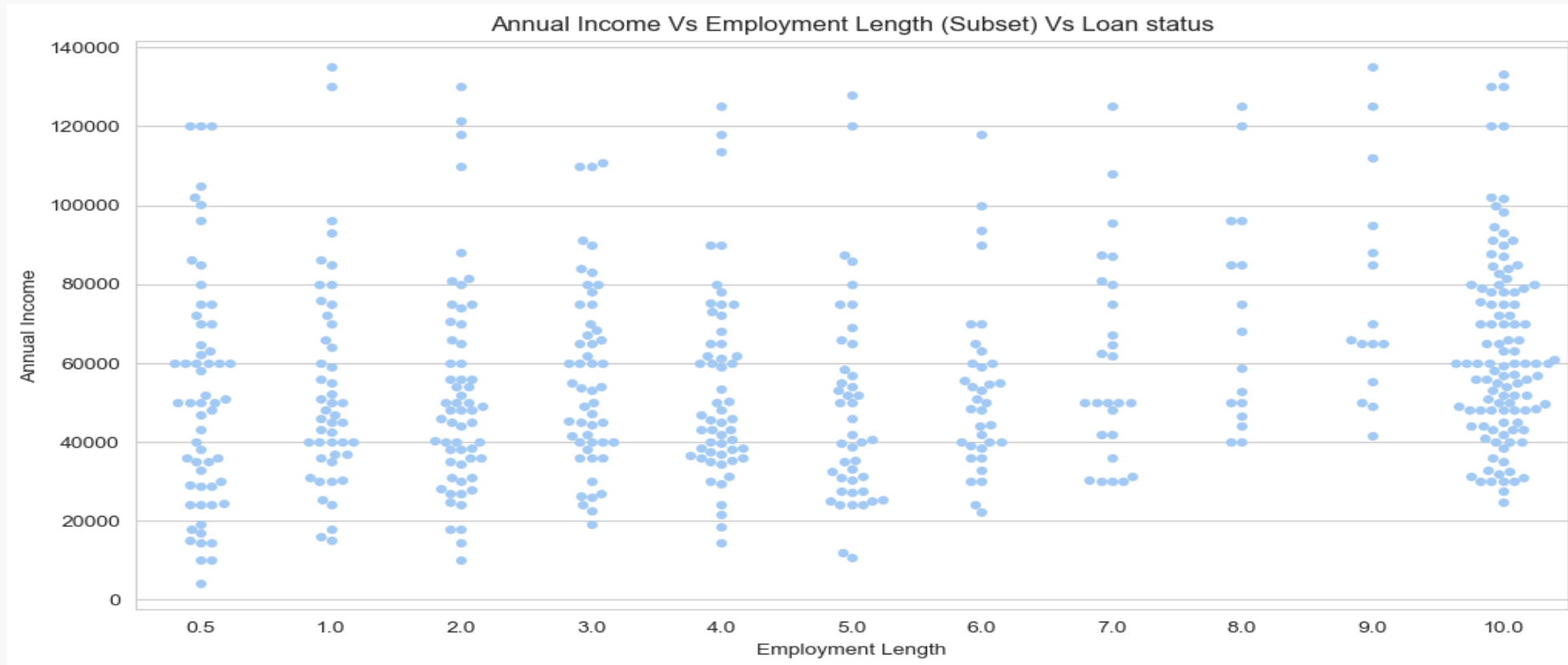
delinq_2yr Vs loan amount Vs grade



Observation:

Loans with many missed payments don't get much investment, even if they ask for less money. They're seen as risky, so Lending Club might want to limit investing in them. Also, people with at least two missed payments are getting loans averaging over \$15,000.

Annual Income Vs Employment Length (Subset) Vs Loan status



Observation:

Borrower with high employment length despite of having high annual income tends to default more compare to the ones having low employment length. The reason may be short period of employment left(they are reaching their retirement in few years)

Correlation Heatmap analysis

Key Findings from Correlation Analysis:

Strong Correlation:

The monthly installment amount (installment) shows a robust correlation with loan amount.

The loan term (term) exhibits a strong correlation with the interest rate.

Weak Strong Positive Association:

The annual income (annual_inc) is strongly correlated with the loan amount.

Weak Correlation:

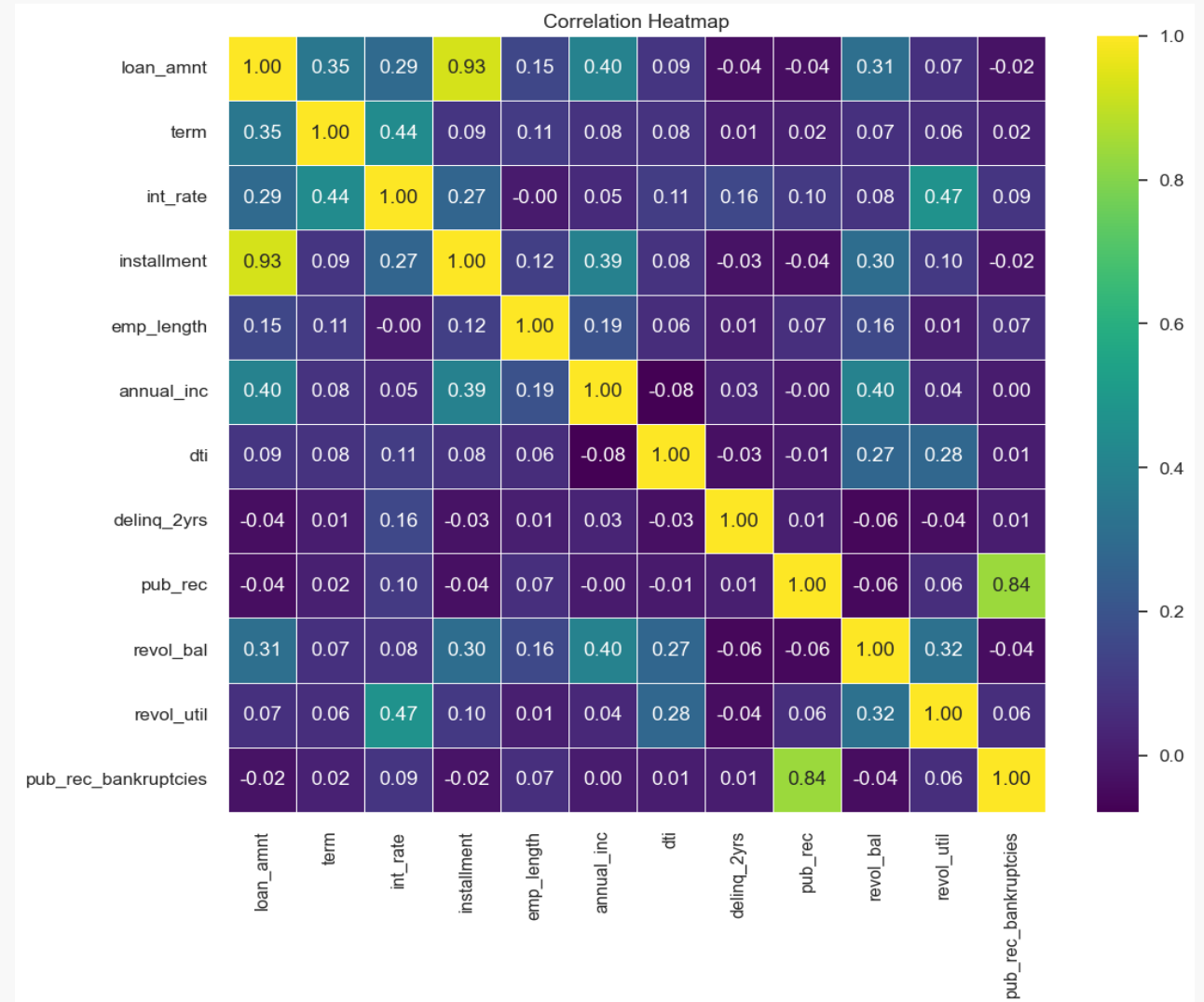
The debt-to-income ratio (dti) demonstrates a weak correlation with most of the observed fields.

Employment length (emp_length) displays a weak correlation with various attributes.

Negative Correlation:

The presence of public record bankruptcies (pub_rec_bankruptcies) shows a negative correlation with almost every observed field.

Annual income (annual_inc) demonstrates a negative correlation with the debt-to-income ratio (dti).



Recommendations from EDA of the Lending Club

1 Loan applicants from states like California (CA), Florida (FL), and New York (NY) are more likely to default.

The company should monitor regional risk trends and adjust lending strategies or rates accordingly in these areas.

2. Given that a significant portion of "Charged Off" loans comes from applicants in Grades B, C, and D, it is advisable for the company to enhance its risk assessment and underwriting criteria specifically for individuals in these grades.

This measure aims to mitigate the occurrence of defaulted loans and strengthen the overall quality of the lending portfolio.

3. Loan applicants with ten or more years of experience are more likely to default.

This suggests that experience alone may not be a reliable indicator of credit worthiness. The company should use a more comprehensive credit scoring system that factors in other risk-related attributes.

4. Lending Club should exercise caution when dealing with loans intended for Small Business purposes, as they exhibit the highest charge-off rate. Loans designated for Weddings, major purchases, car financing, and credit card payments demonstrate favorable repayment trends, making them advisable for acceptance.

5. Higher the loan amount, the higher the chances of loan being charged off. Therefore Lending Company should consider accepting loans of lower amount. Hence the risk factor is low for lending club.

Recommendations from EDA of the Lending Club

6. Lending Club should consider accepting more loans from applicants whose annual income is greater than 1000000 as their probability of charge off is minimum.

7. Lending Club should consider accepting more loans where interest rate is less than 7.5% as their probability of charge off is minimum.

8. Lending Club may find it beneficial to extend loans to applicants with both a substantial income and a moderate number of years of employment.

9. Lending Club should consider accepting more loans from people who own a house.

10. Lending Club should accept more loans for the term of 36 months as the % of charged off loans is less and the no. of loan applicants are more. accepting more loans from applicants whose annual income is greater than 1000000 as their probability of charge off is minimum.