Lending Club Case Study:

This project is a data science project that uses the lending club data set to predict whether a loan will be defaulted or not.

Group Members:

Vijay Sekhar Lattala

Srikanth Kaspe

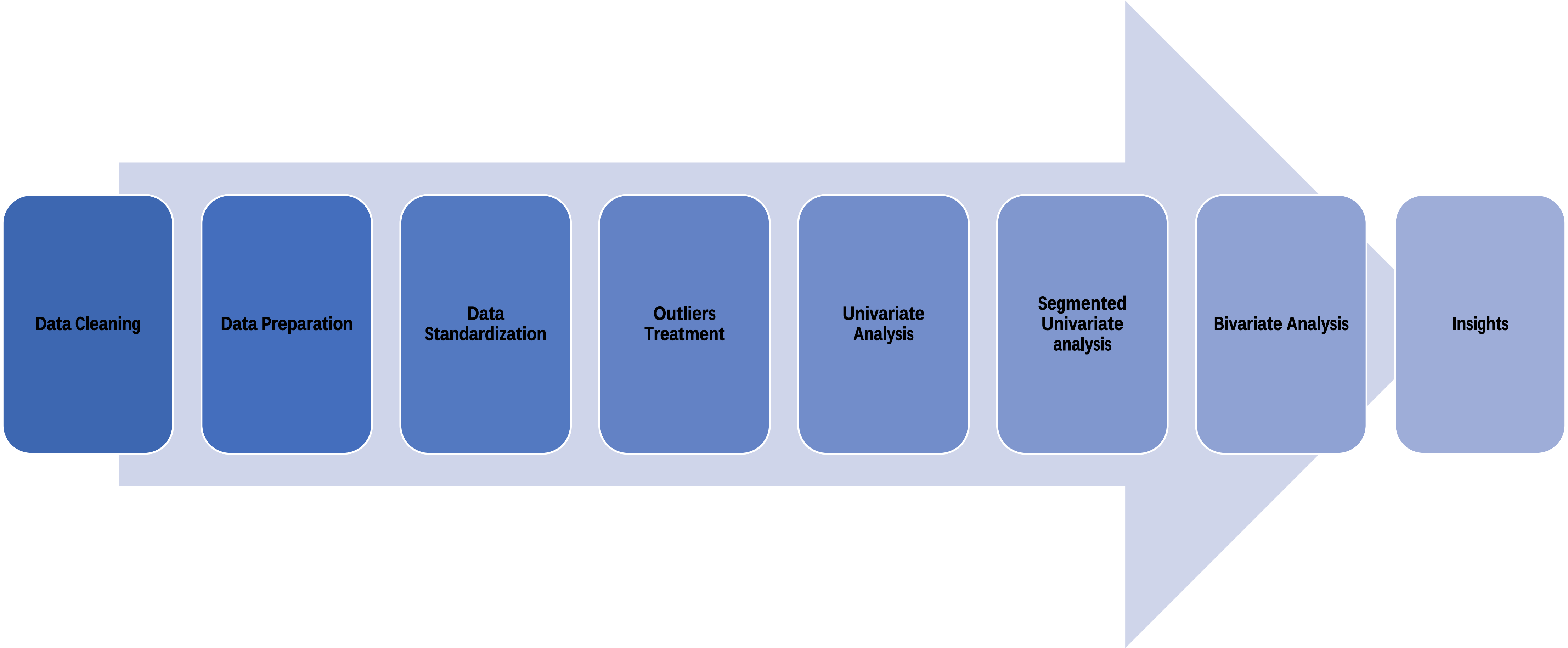
Problem Statement

- Lending Club is a consumer finance marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.
- It specializes in lending various types of loans to urban customers. When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile.

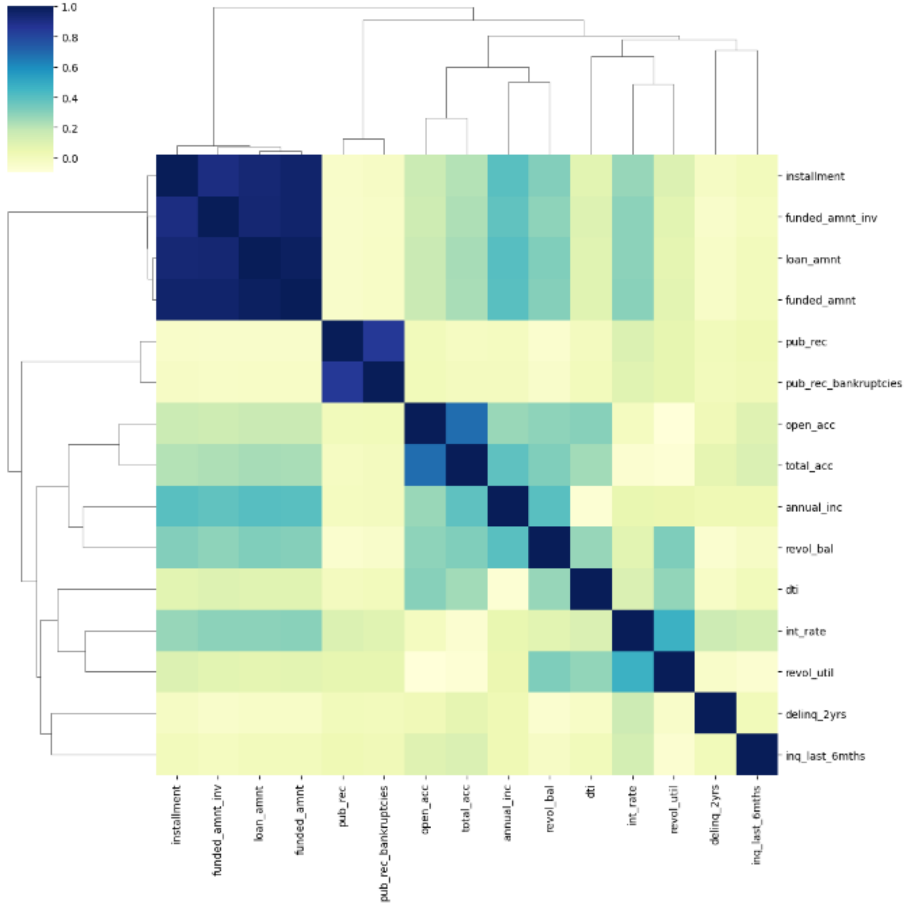
Objective

- ➢The goal is to identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss.
- ➢The objective is to use EDA to 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.

Process

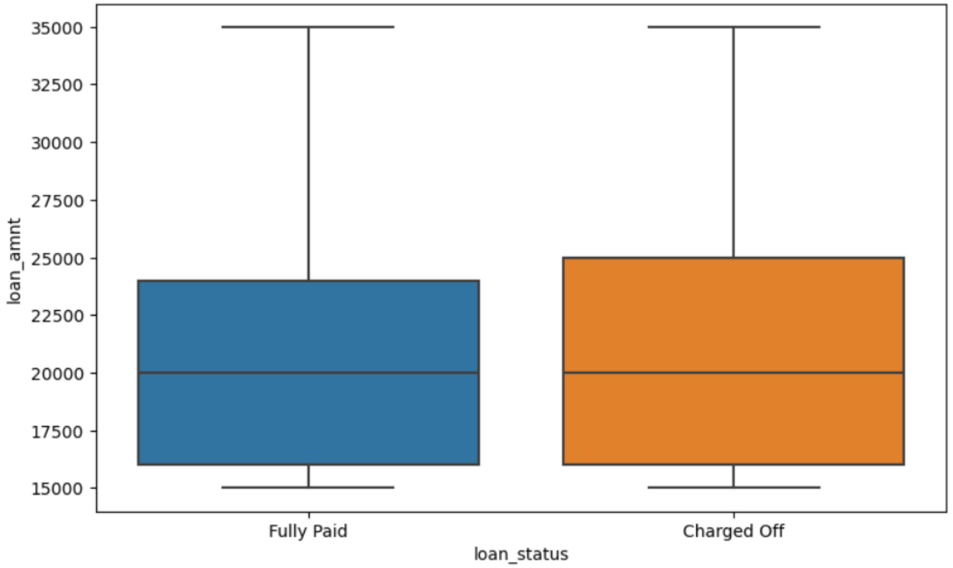

Data Cleaning
Data Preparation
Data Standardization
Outliers Treatment
Univariate Analysis
Segmented Univariate analysis
Bivariate Analysis
Insights

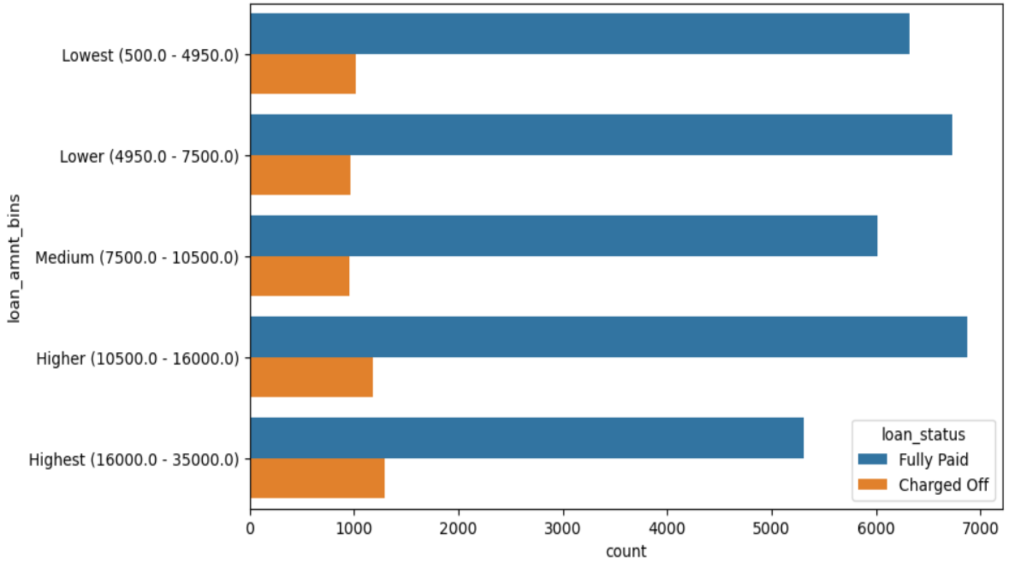
Correlation between numerical fields



- ➢Here, we can clearly see loan_amnt, funded_amnt, funded_amnt_inv and installment have huge correlation as we know that darker the value heigher the correlation.
- ➢Next, the fields pub_rec & pub_rec_bankrupcies and number of accounts related fields open_acc & total_acc are correlated.

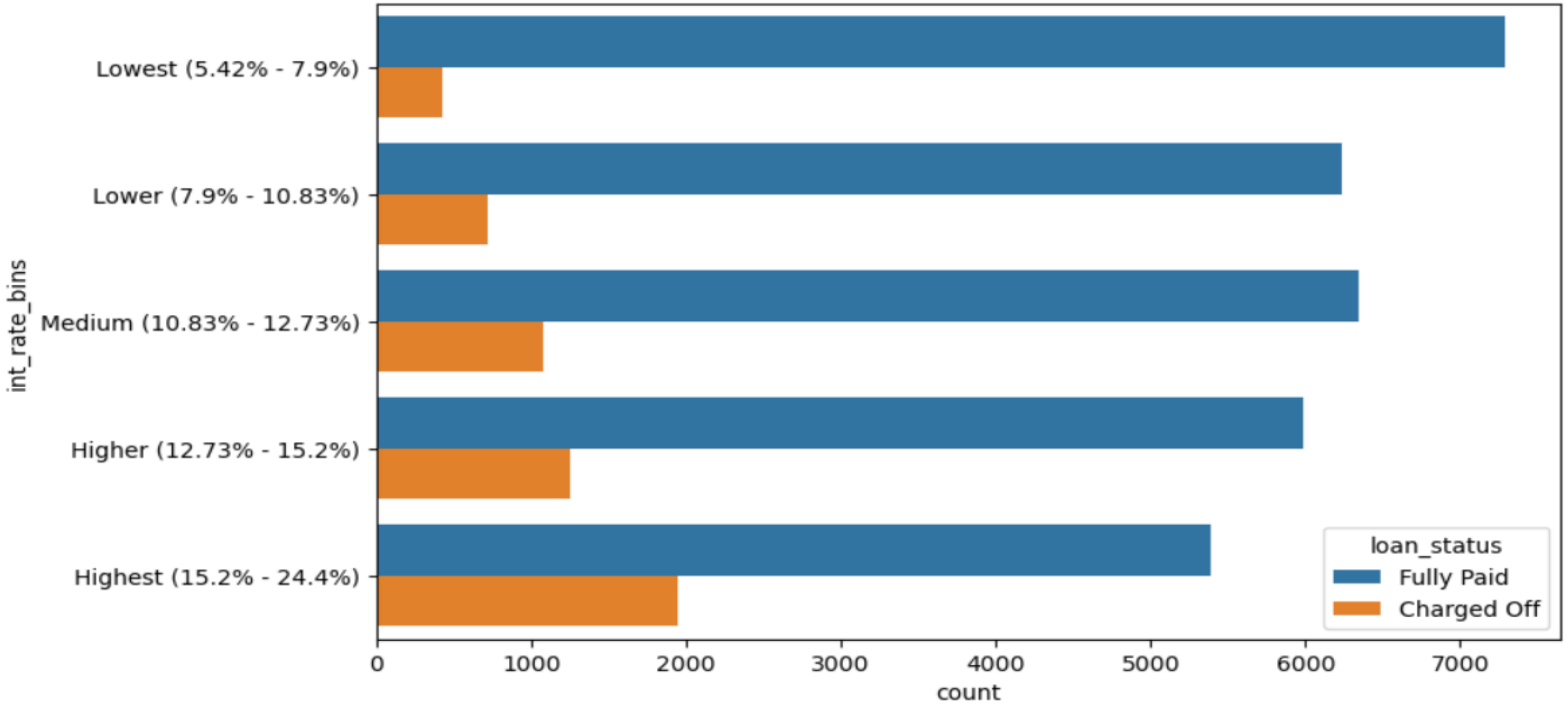
Loan Amount





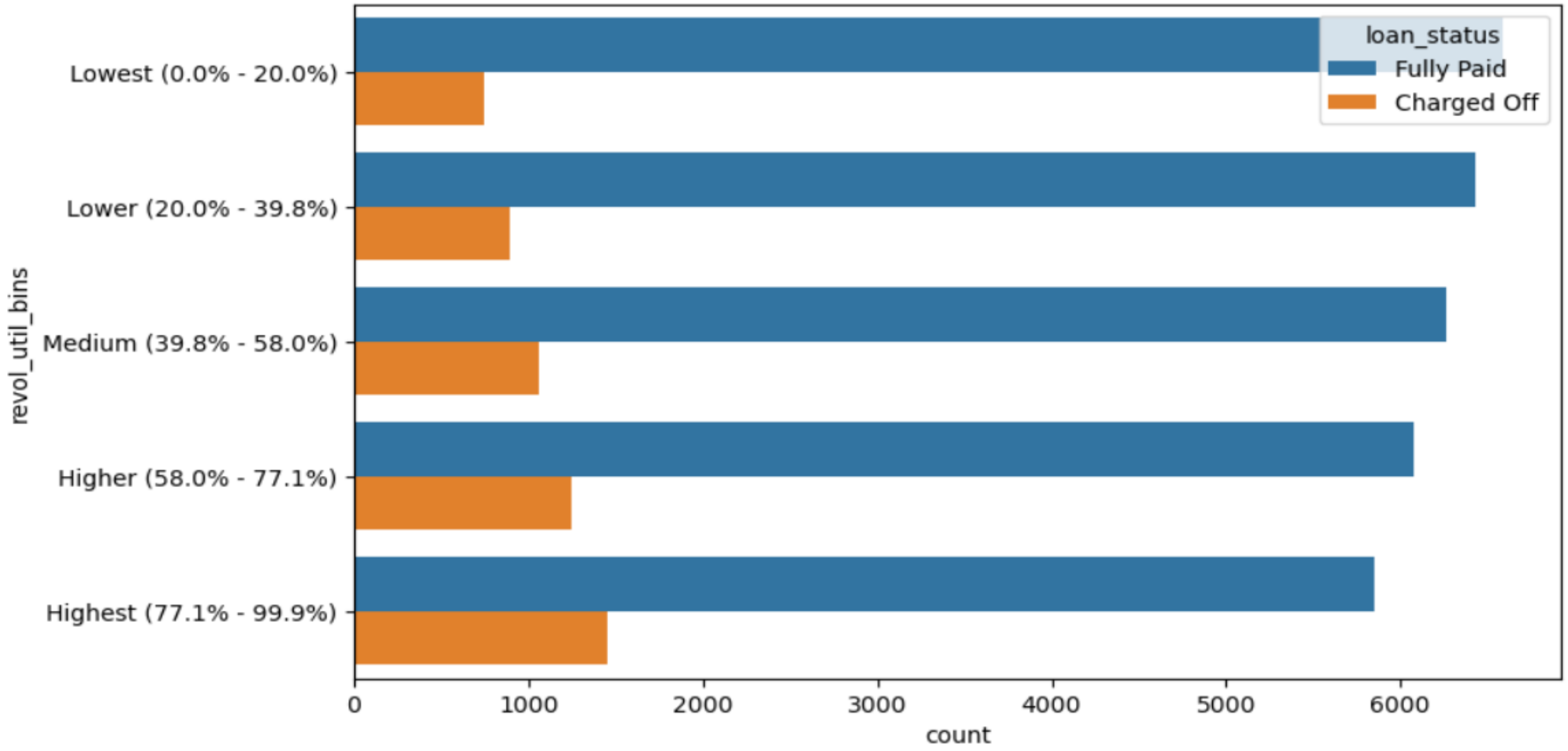
- Higher the loan amount (above 16K) will lead to higher default.

Interest Rate



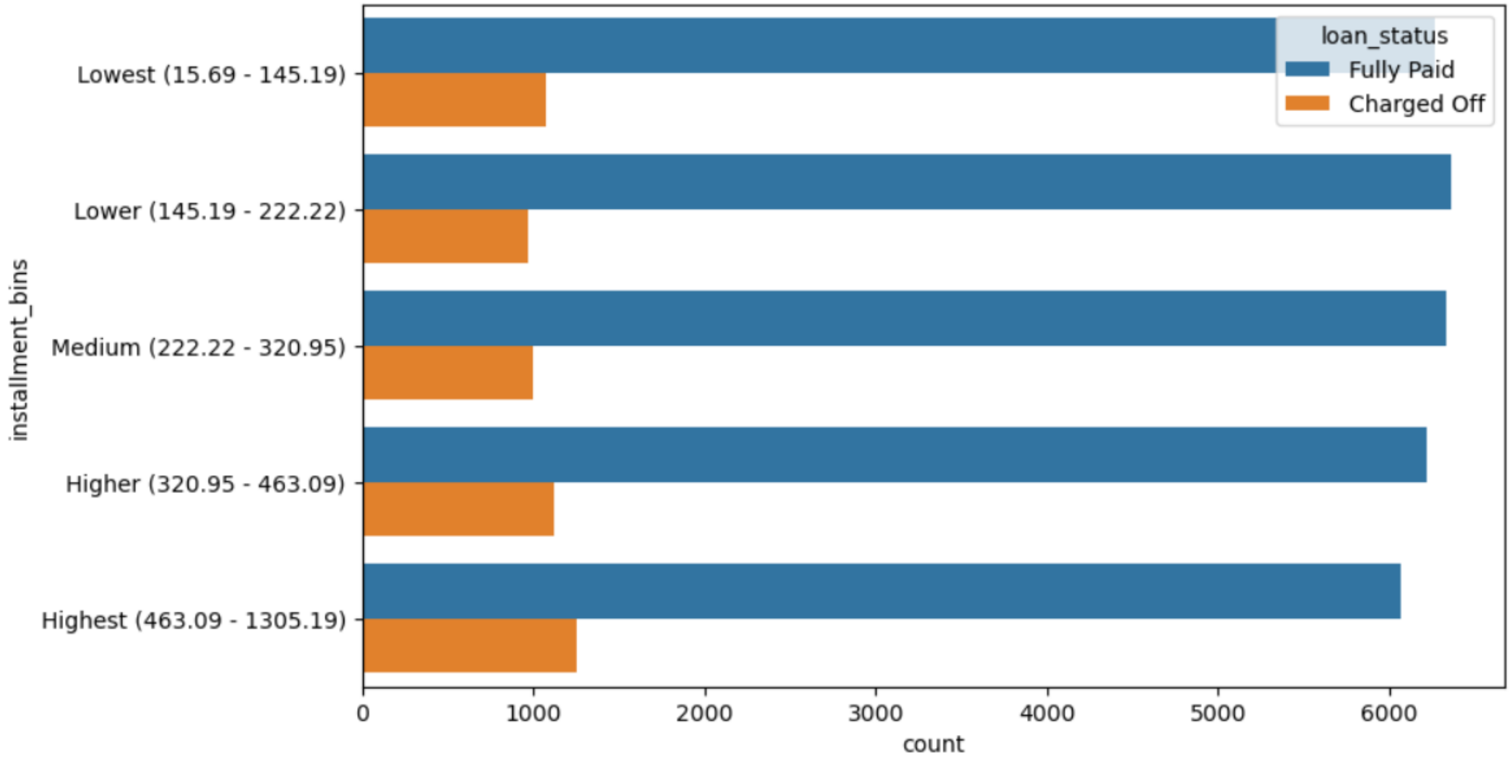
- Higher the interest rate above 13% will lead to higher default.

Revolving Utilization Rate



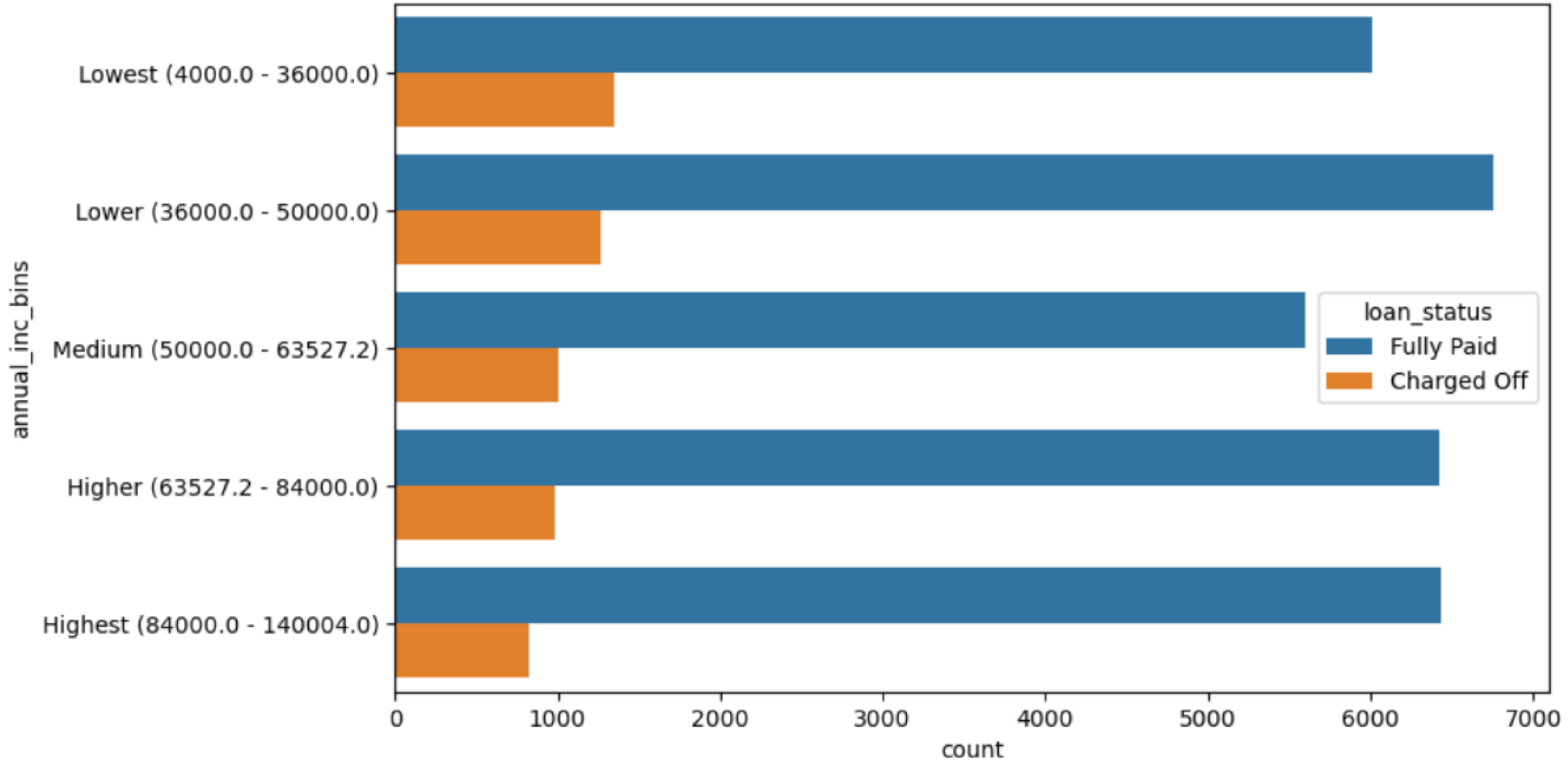
- ➢ Higher the revolving utilization rate above 58% will lead to higher default.

Instalment



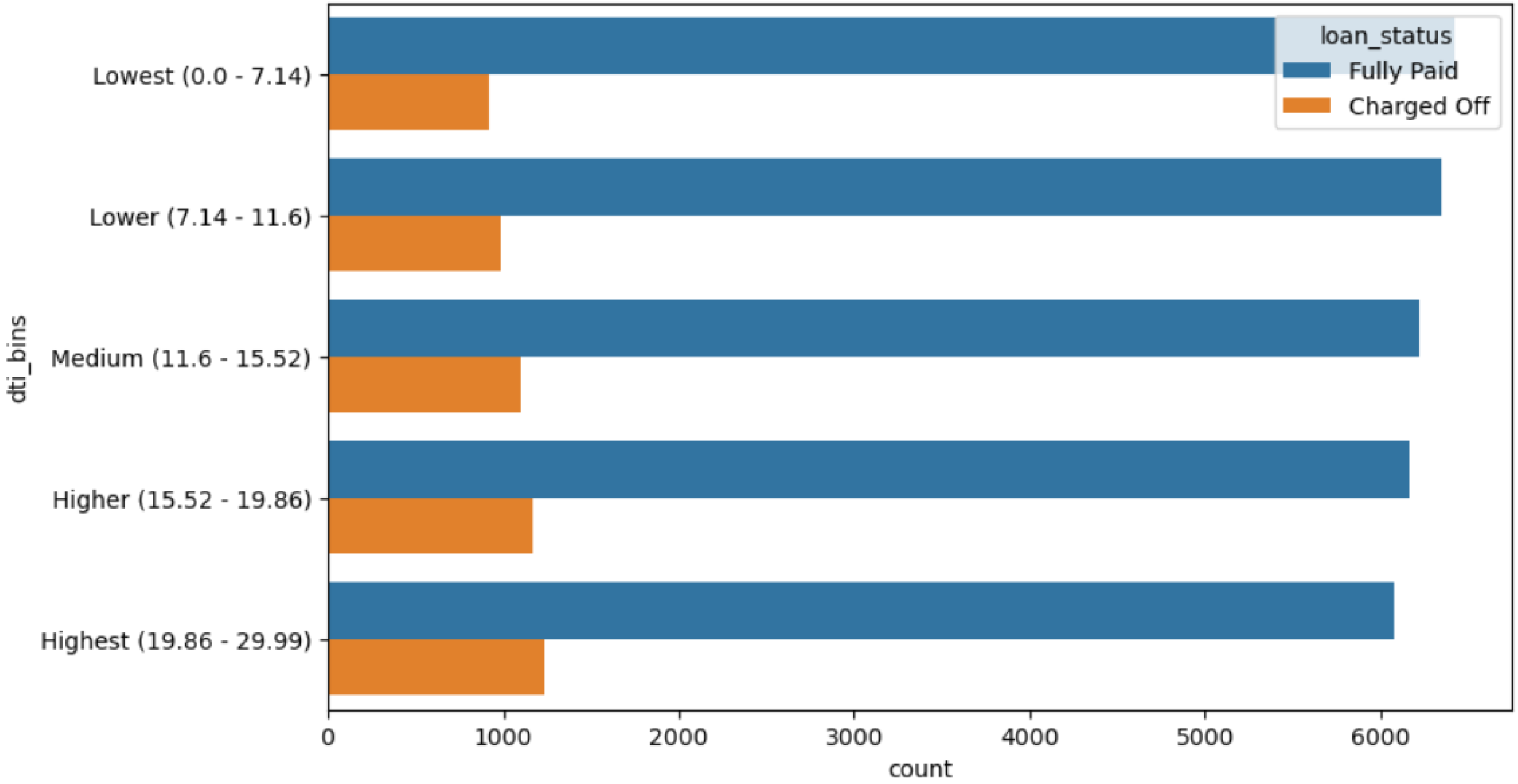
- ➢ Higher the installment amount (above 327) will lead to higher default.

Annual Income



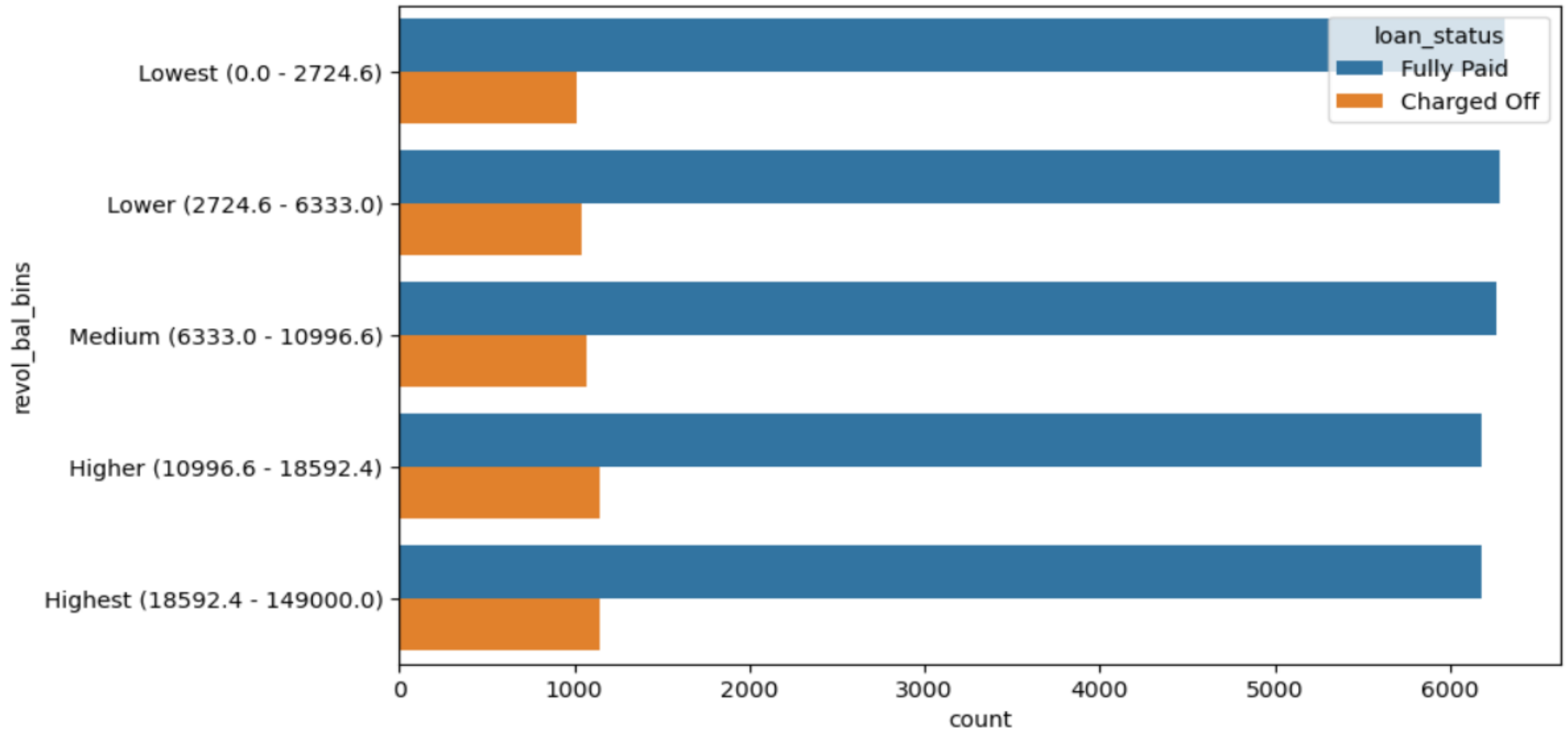
- Lower annual income (below 37K) will lead to higher default.

Debt to Income Ratio



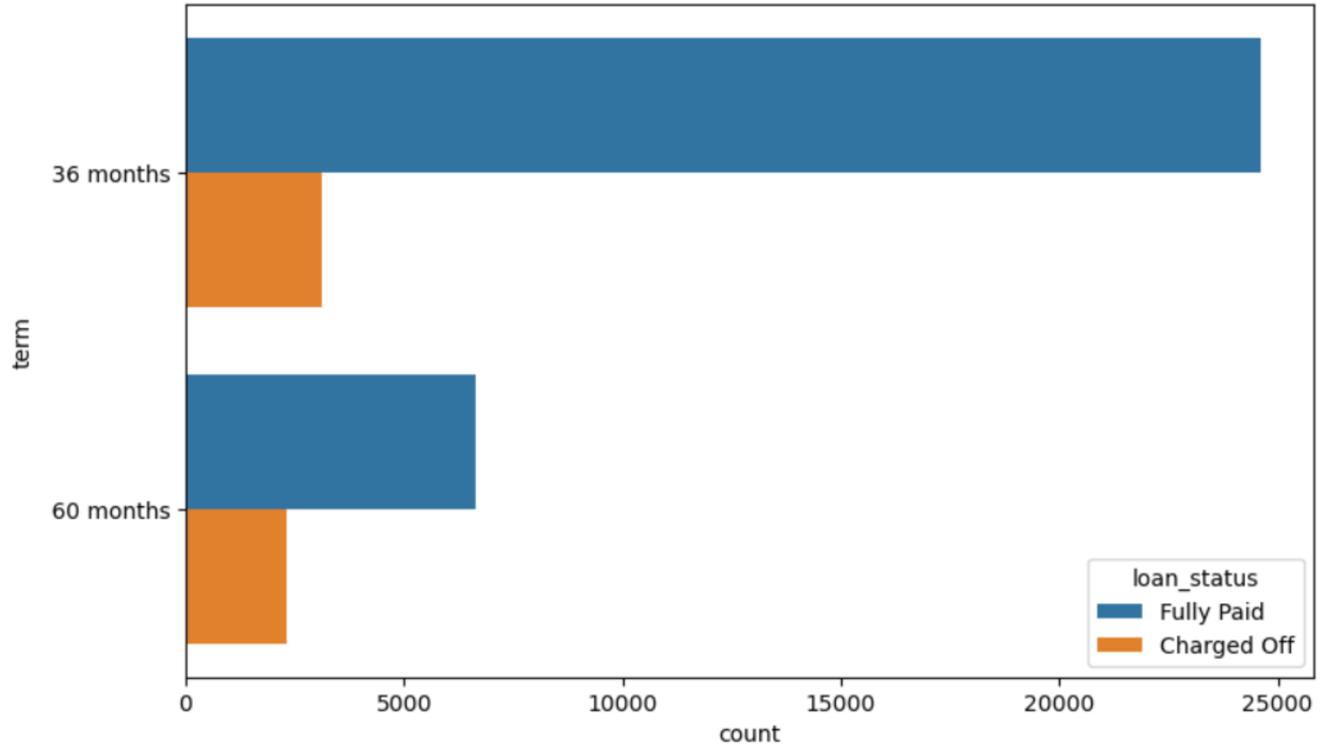
- Higher debt to income ratio (above 15%) will lead to higher default.

Revolving Balance



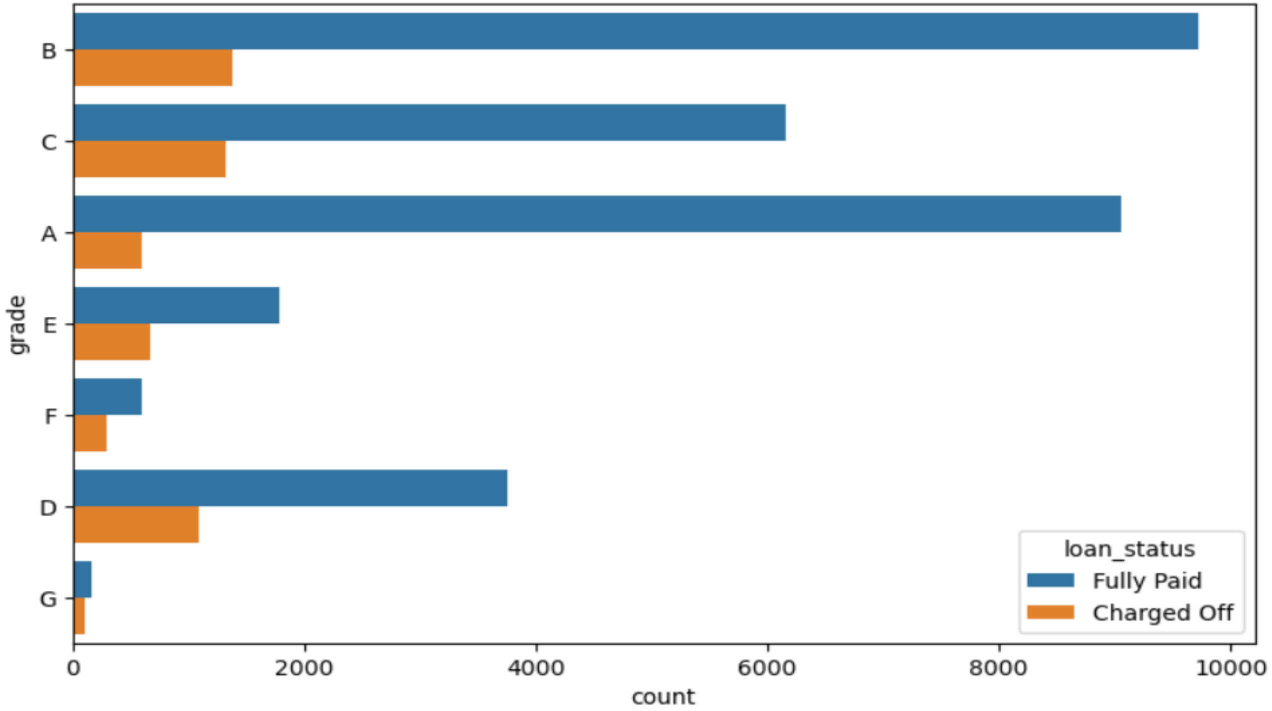
- ➢ Higher the revolving balance above 10k, bigger the chance of the loan getting defaulted.

Repayment Term



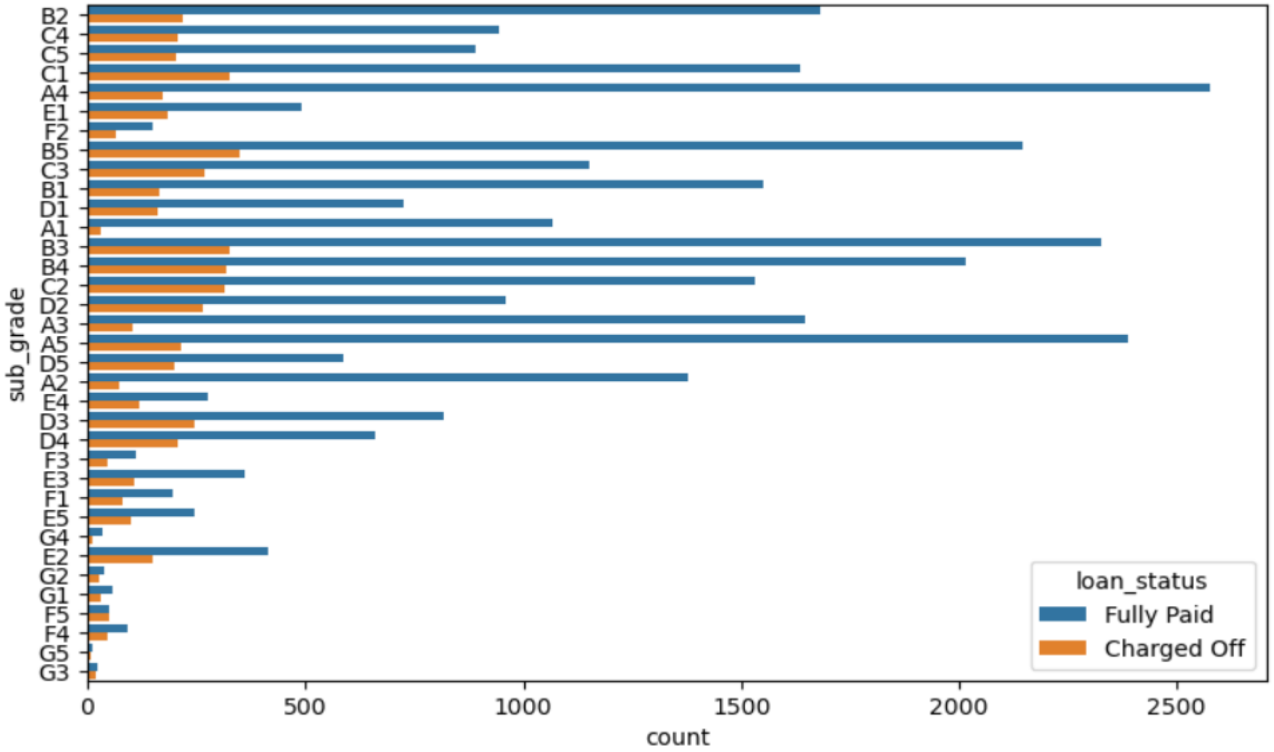
- Loan Repayment term (5 years) will lead to the loan getting defaulted.

Grade



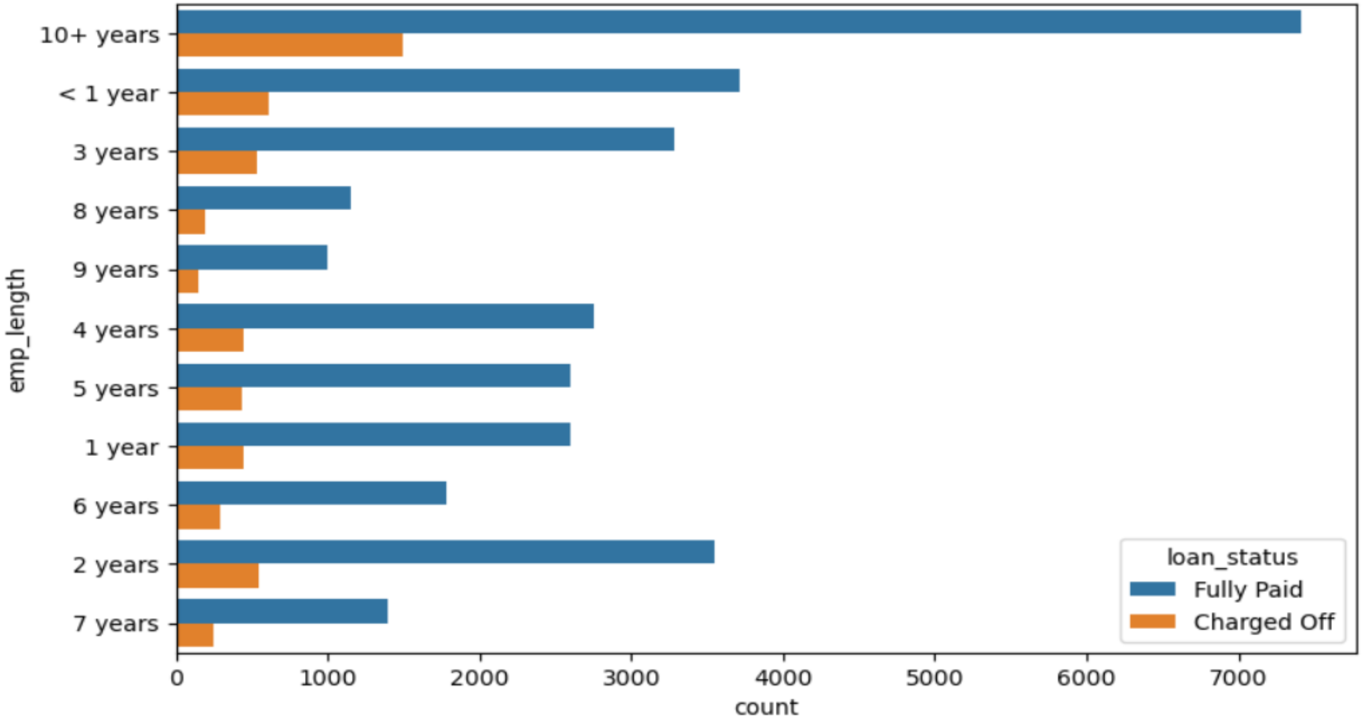
- Loan Grades from D, E, F and G when compared to others will lead to the loan getting defaulted.

Sub grade



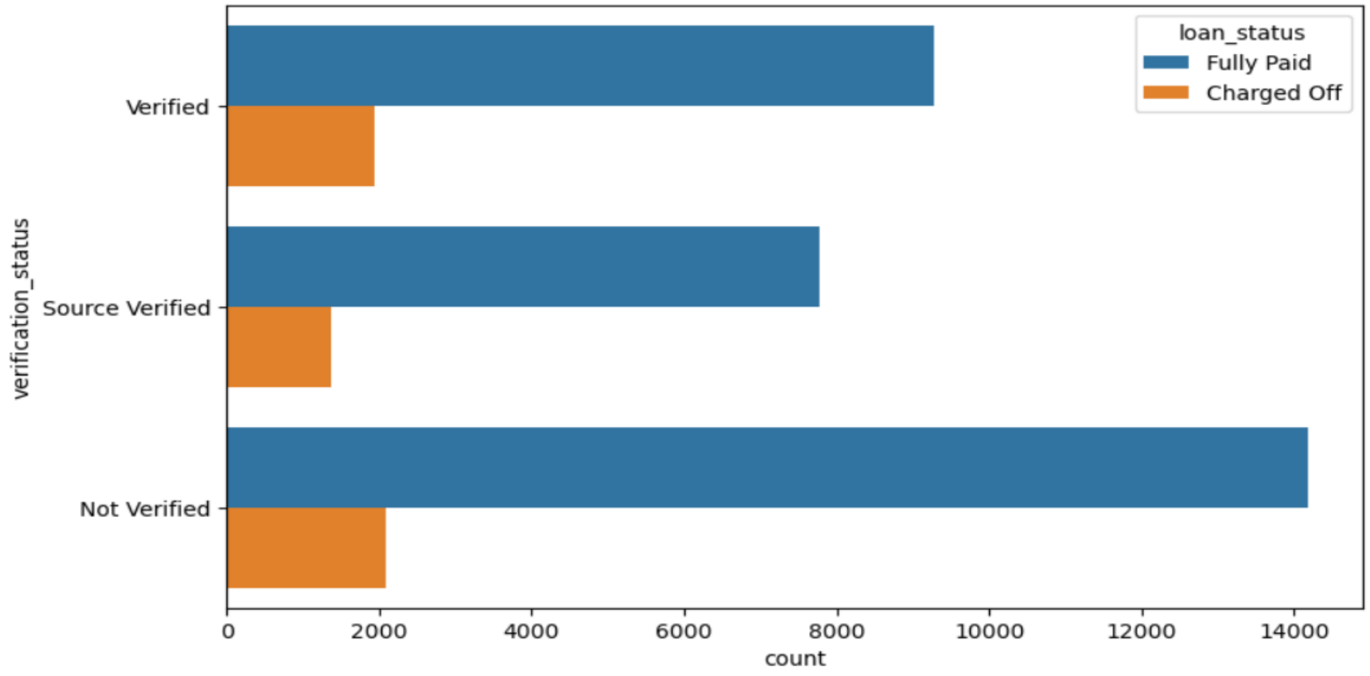
➢ Loan sub grades G3 and F5 when compared to others will lead to the loan getting defaulted.

Employment Length



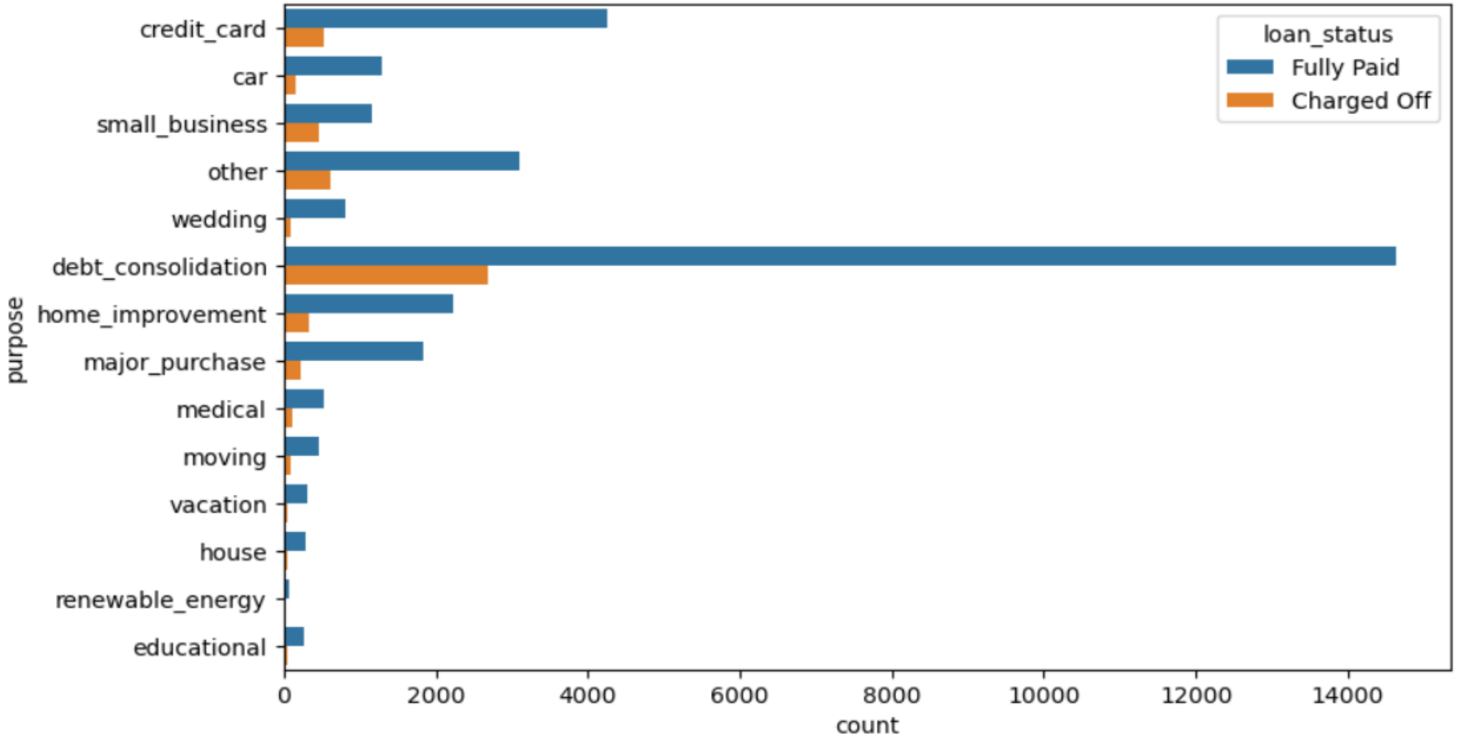
- Although there is nothing much to conclude from above data, we can see having no employment will lead to Defaulted payment.

Verification Status



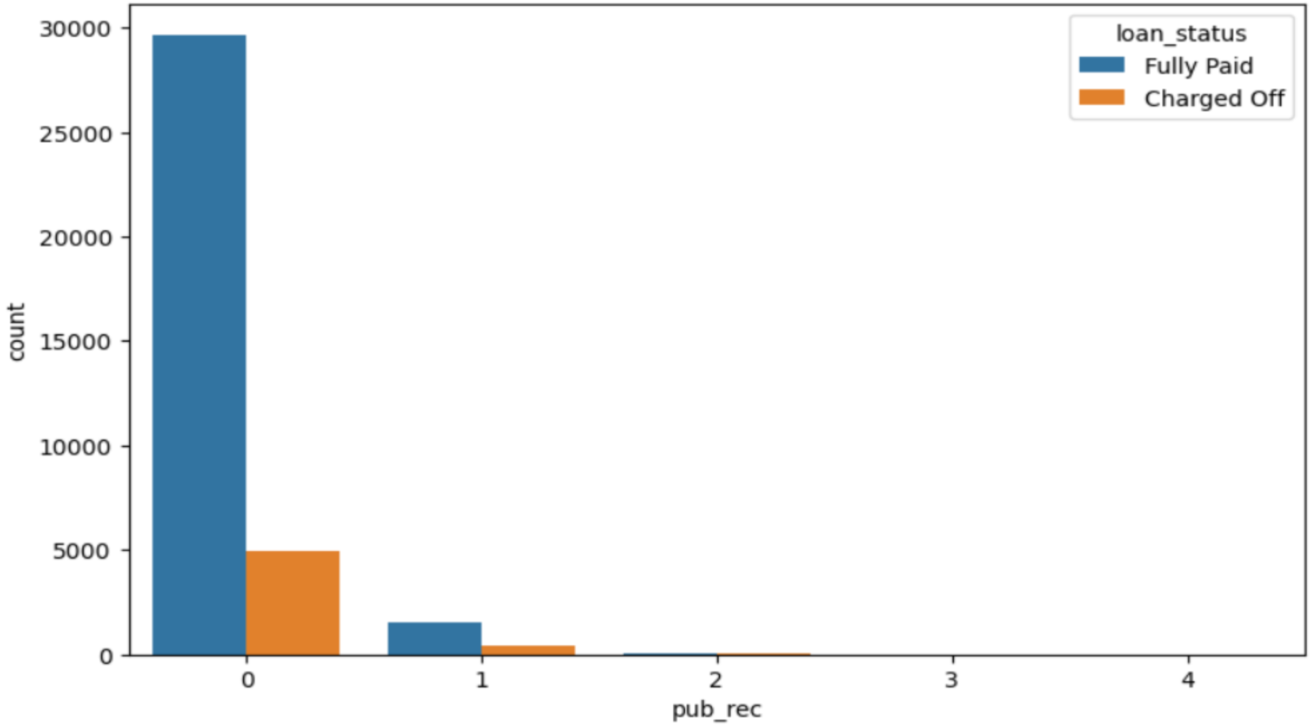
- Interestingly verified applicants have more defaulted which means the verification process is not correct.

Purpose



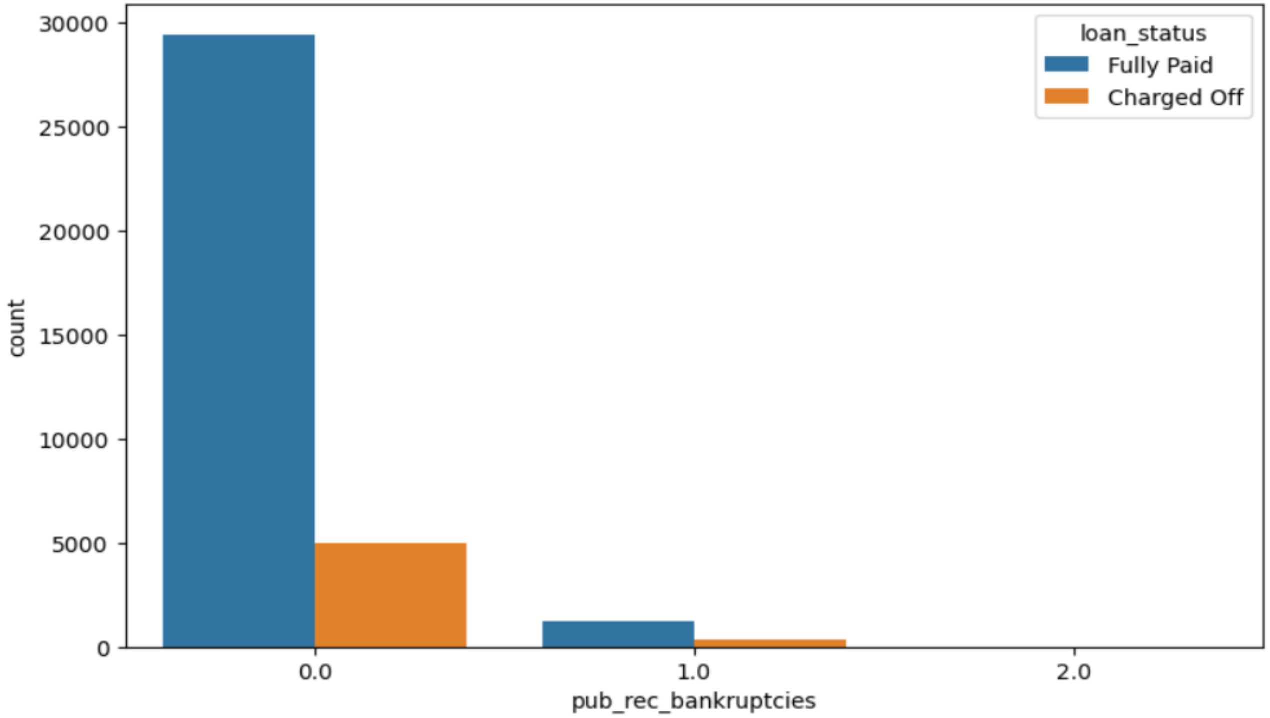
➢ Loans taken for purpose like small business, renewable energy, educational are the risky ones.

Derogatory public records



➢ Derogatory public records with (1 or 2) have the chance of defaulting the loan.

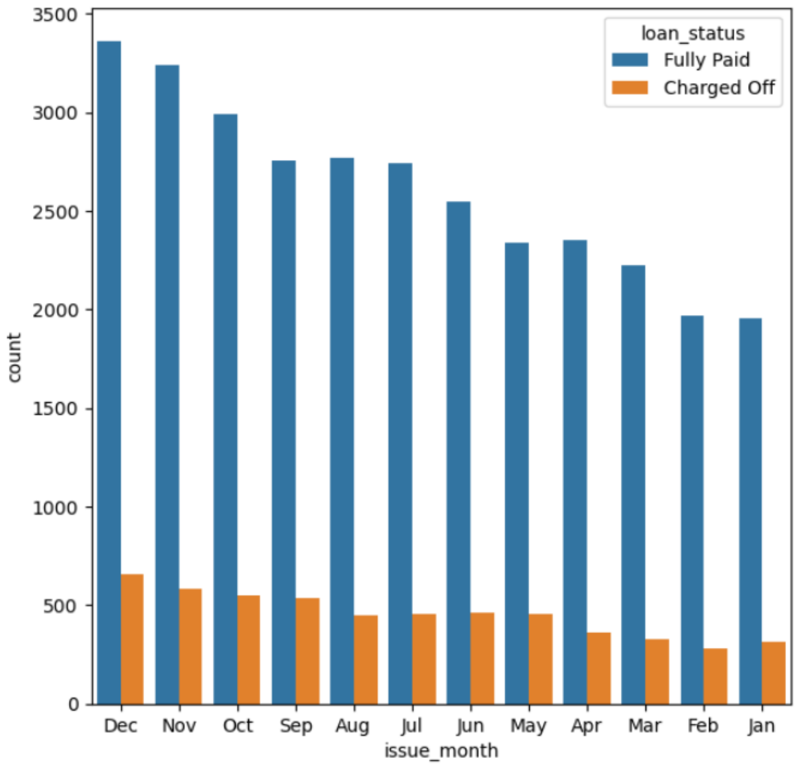
Public bankruptcy records

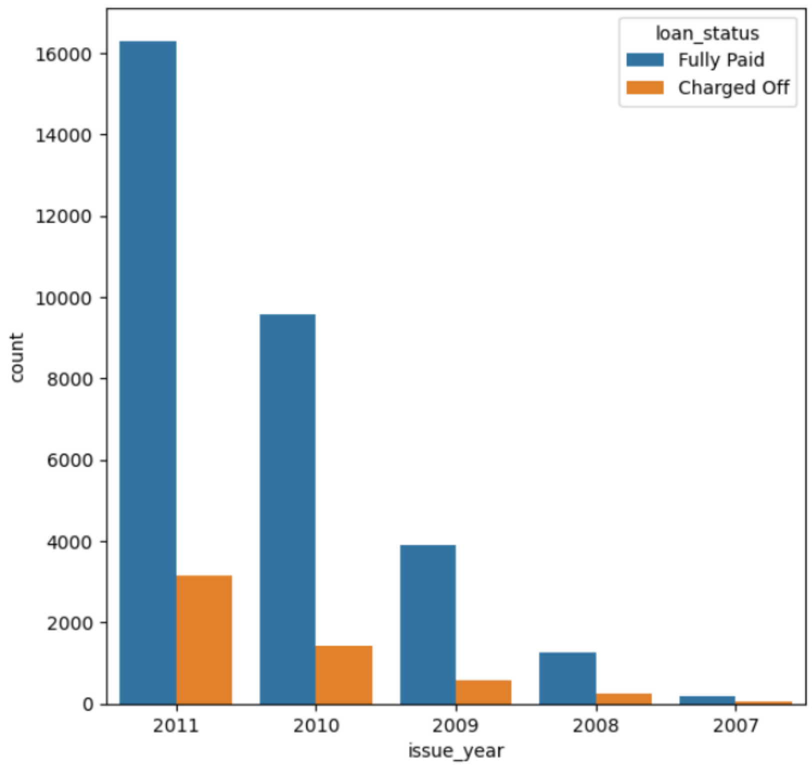


- Higher the number of public bankruptcy records, bigger the chance of defaulting the loan.

Issue Month and Issue Year

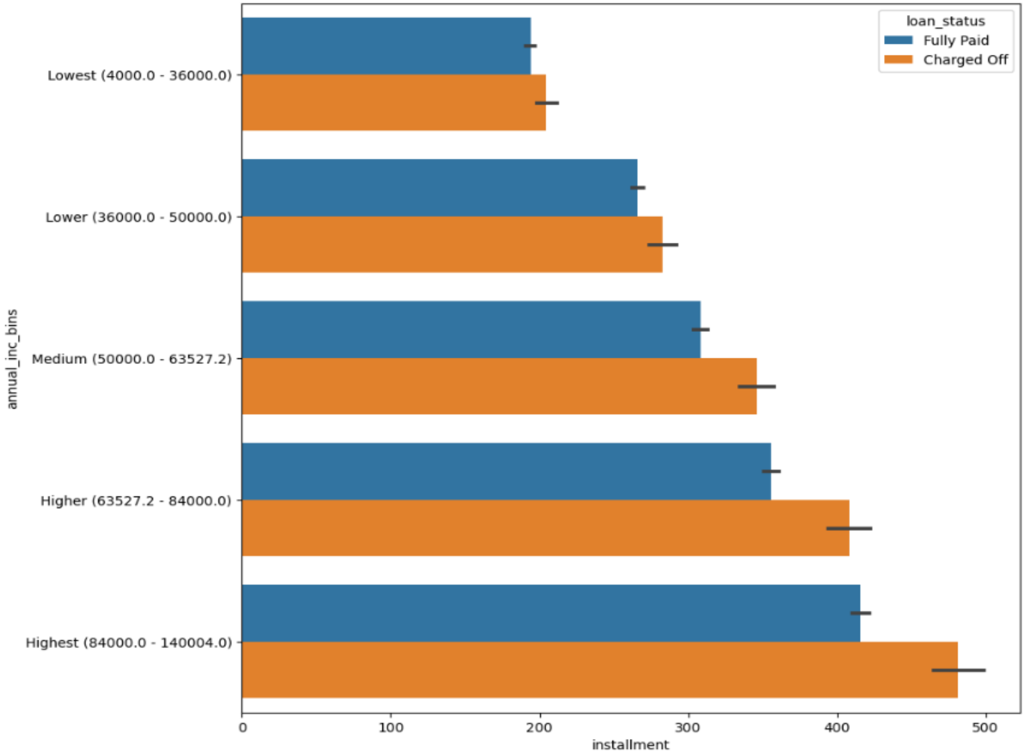
<Axes: xlabel='issue_year', ylabel='count'>





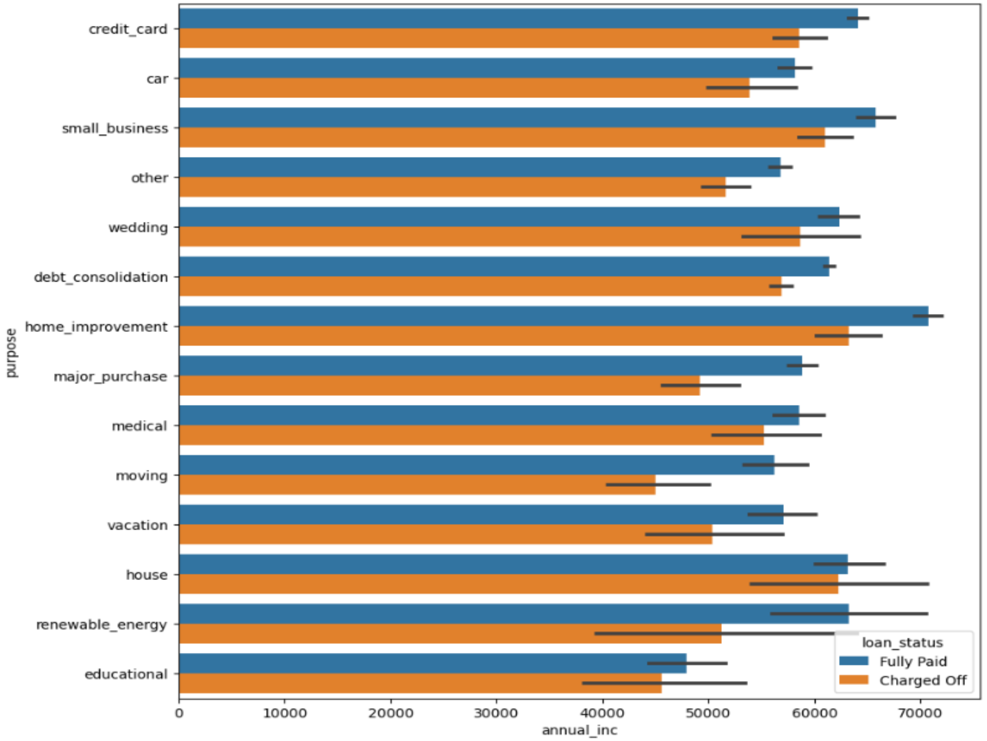
➢ Loan issue month in Dec, May, Sep have the biggest default ratio.

Instalment vs Annual Income



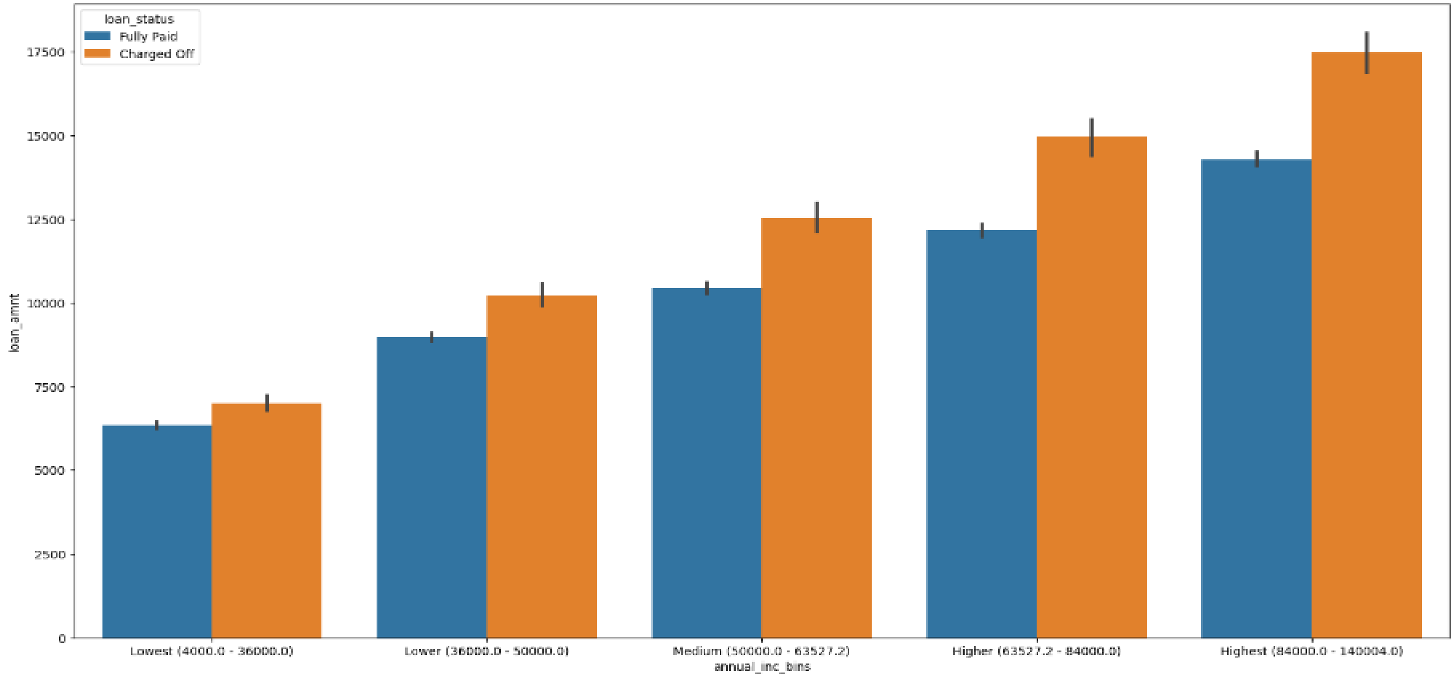
➢ Higher instalments for any income group have more number of defaults.

Annual Income vs Purpose



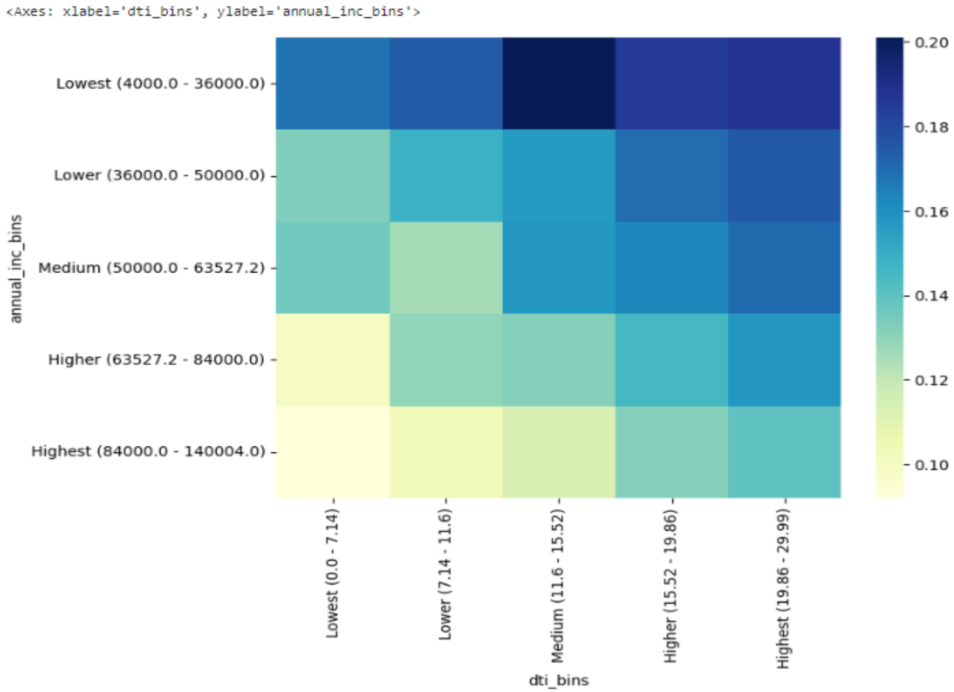
- ➢ Though the number of loans applied and defaulted are the highest in number for "debt_consolation", the annual income of those who applied isn't the highest.

Annual Income vs Loan Amount



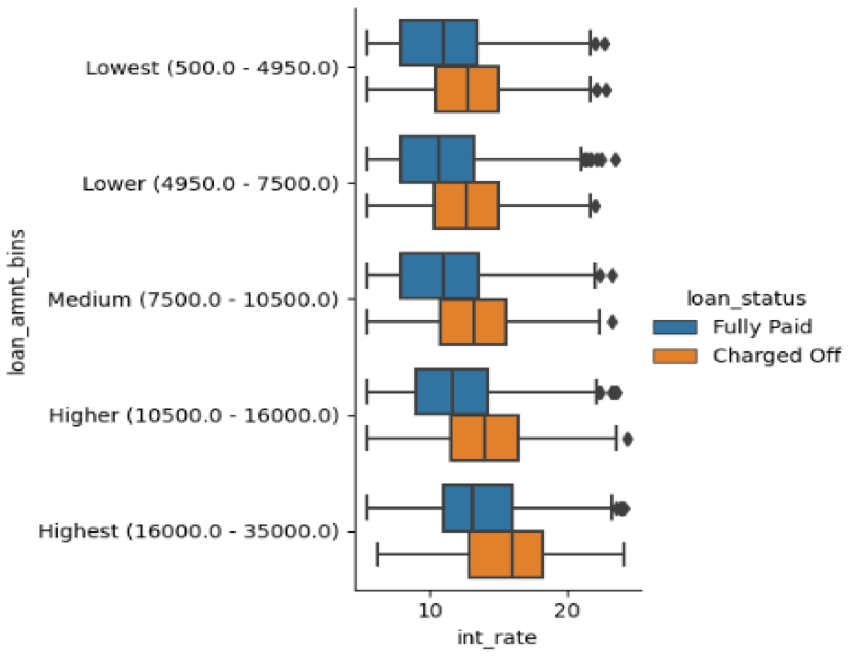
- Across all the income groups, the loan amount is higher for people who defaulted.

Annual Income vs Debt to Income



- ➢ Medium debt-to-income group in the lowest income range is the most risky when it comes to loan repayment.

Interest Rate vs Loan Amount



- ➢ Interest rate for charged off loans is pretty high than that of fully paid loans in all the loan amount groups.

Insights

According to our observations, below are the most impactful driving factors (or driver variables) behind loan default.

- Higher the loan amount (above 16K) will lead to higher default.
- Higher the installment amount (above 327) will lead to higher default.
- Higher the interest rate above 13% will lead to higher default.
- Lower annual income (below 37K) will lead to higher default.
- Higher debt to income ratio (above 15%) will lead to higher default.
- Loan issue month is in Dec, May, Sep.
- Higher the revolving utilization rate above 58%, bigger the chance of the loan getting defaulted.

(continued on next page)

Insights

- Higher the revolving balance above 10k, bigger the chance of the loan getting defaulted.
- Repayment term (5 years).
- Loan Grades from D, E, F and G when compared to others.
- Loan sub grades G3 and F5 when compared to others.
- Missing employment record will lead to default.
- Loan purpose like small business, renewable energy, educational.

(continued on next page)

Insights

- Derogatory public records with (1 or 2) have the chance of defaulting the loan.
- Higher the number of public bankruptcy records, bigger the chance of defaulting the loan.
- Higher installments for any income group have more number of defaults.
- Across all the income groups, the loan amount is higher for people who defaulted.
- Medium debt-to-income group in the lowest income range is the most risky when it comes to loan repayment.
- Interest rate for charged off loans is pretty high than that of fully paid loans in all the loan amount groups.

Thank you