Lending Club Case Study

Ajay V M & Harish Babu

Project Description

The most significant contributor to financial loss for lenders is the act of granting loans to applicants deemed 'risky,' which results in what is known as credit loss. Credit loss represents the monetary value that lenders lose when borrowers either fail to repay their loans or abscond with the owed funds.

The primary goal is to have the capability to recognize these high-risk loan applicants as accurately as possible. By achieving this, it becomes possible to decrease the number of such loans, ultimately reducing the total credit loss.

The case study primarily emphasizes Exploratory Data Analysis (EDA) with the objective of identifying the key parameters that significantly influence the prediction of whether a customer will default on a loan. The data set used is from Lending Club for past loans and various parameters of those loans, including their final settlement (Paid vs Defaulted). A presentation was created to visually illustrate the crucial factors to take into account when granting loans, including their respective data distributions. Finally, recommendations are made to Lending Club to prevent future defaults.

This analysis was performed by Ajay V M & Harish Babu from the August 2023 Batch of Executive PG Programme in Machine Learning & AI at IIITB

Approach & Method

Steps of the analysis

I — Understand Data

- Eyeball the data and use the data dictionary to understand the nature of the data and quickly make a mental map of the data

II — Data Cleanup

- Fix Missing Values
- Remove Null Values
- Numeric to Categorical Conversion
- Transform & Standardize Values
- Derive new fields that could be relevant.

III — Univariate Analysis

- Summarize single variables using key statistical measures
- Visualize using plots

IV — Segmented Univariate Analysis

- Looking at columns and categorizing them in to various groups
- Derive insights based on the categories

V — Bivariate Analysis

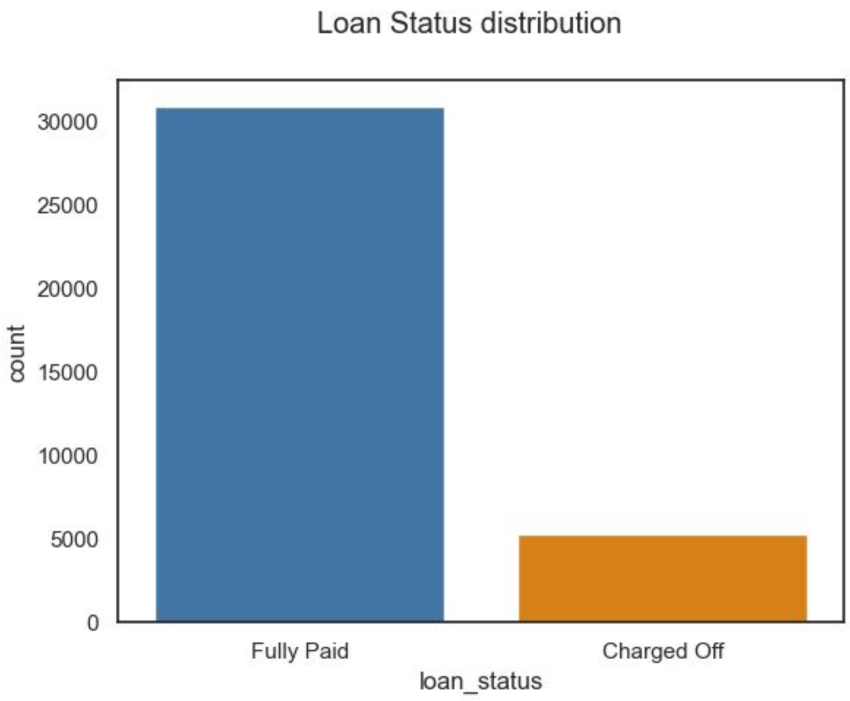
- Look at the impact of multiple variables on the key outcomes (e.g. income vs interest to defaults)

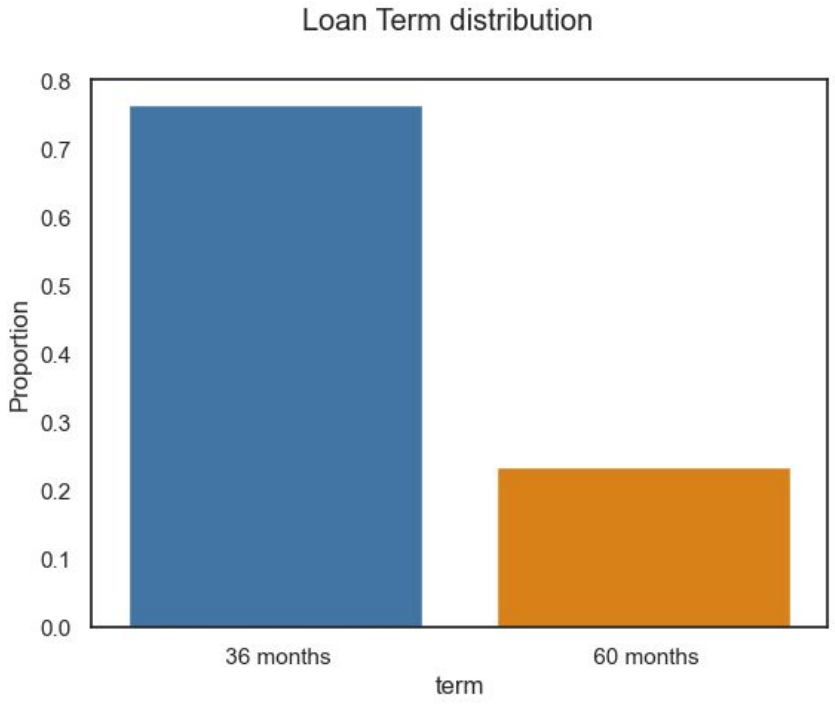
VI — Insights

- Make observations in each of the analysis
- Combine the observations to derive insights and make recommendations

Univariate Analysis

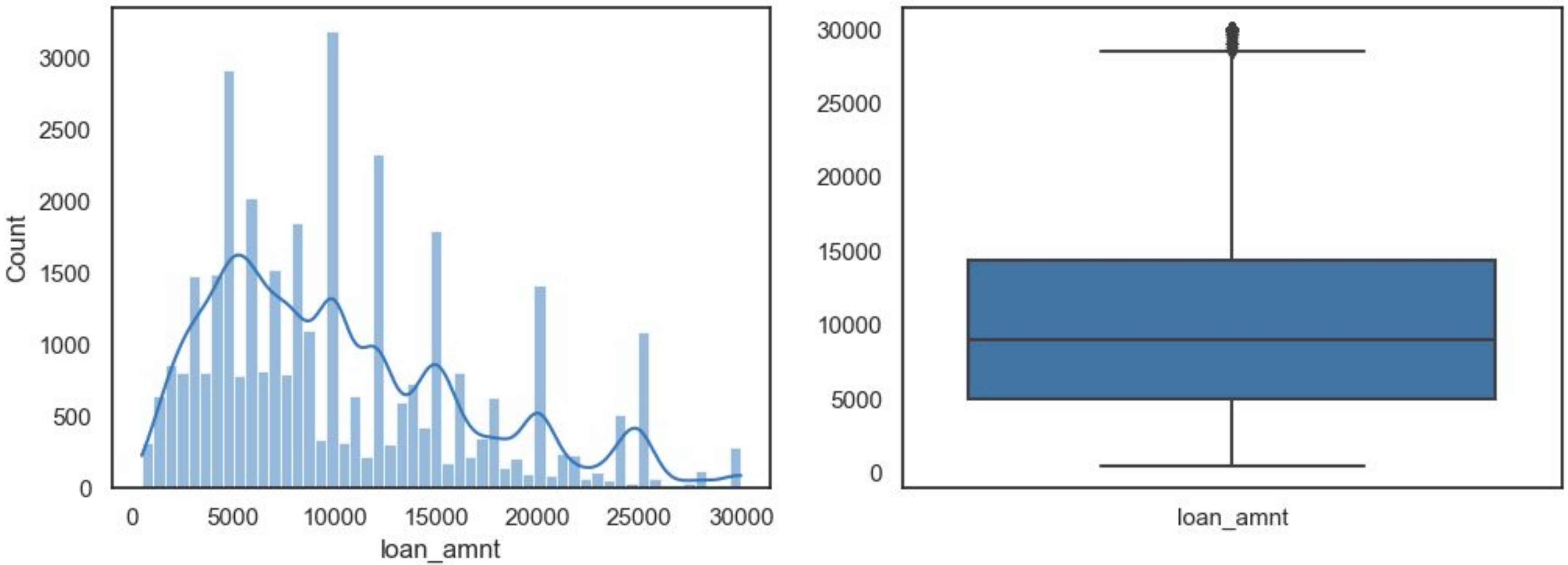
Loan Status Distribution





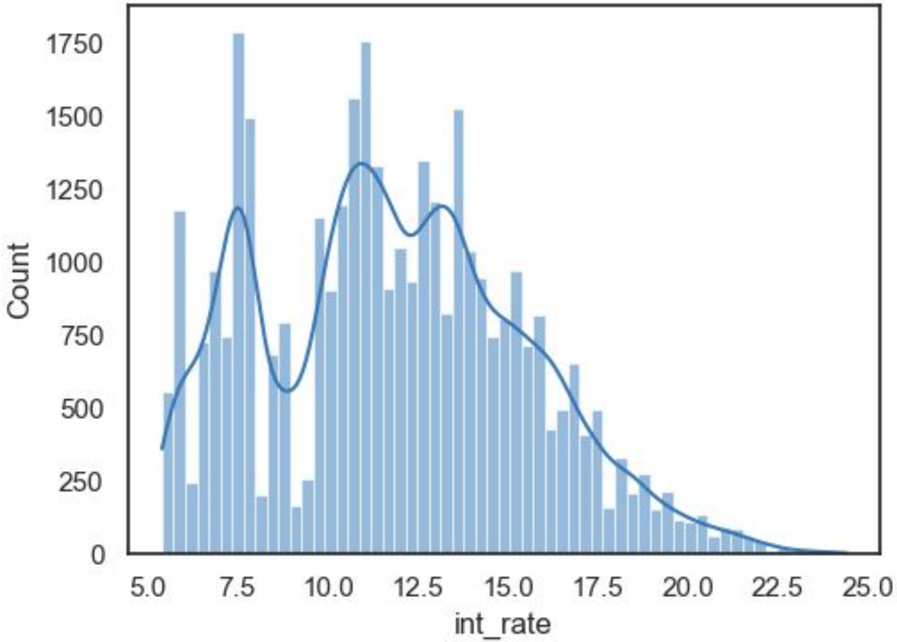
- 85% of borrowers have successfully paid off their loans, while the remaining 15% are the defaulters.
- Majority of loans (76%) have a 36-month term

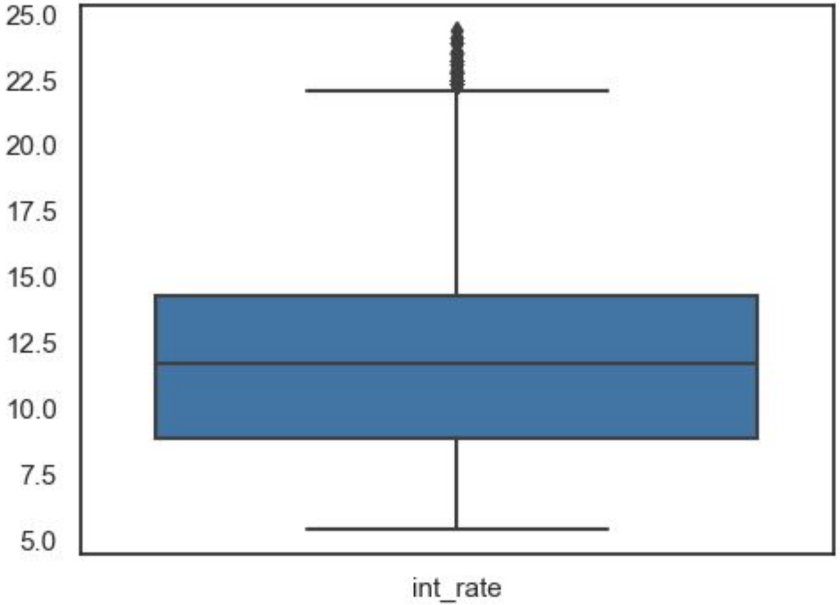
Loan Amount Distribution



- The average loan amount is around 10,000, with a median loan amount of 9,000.

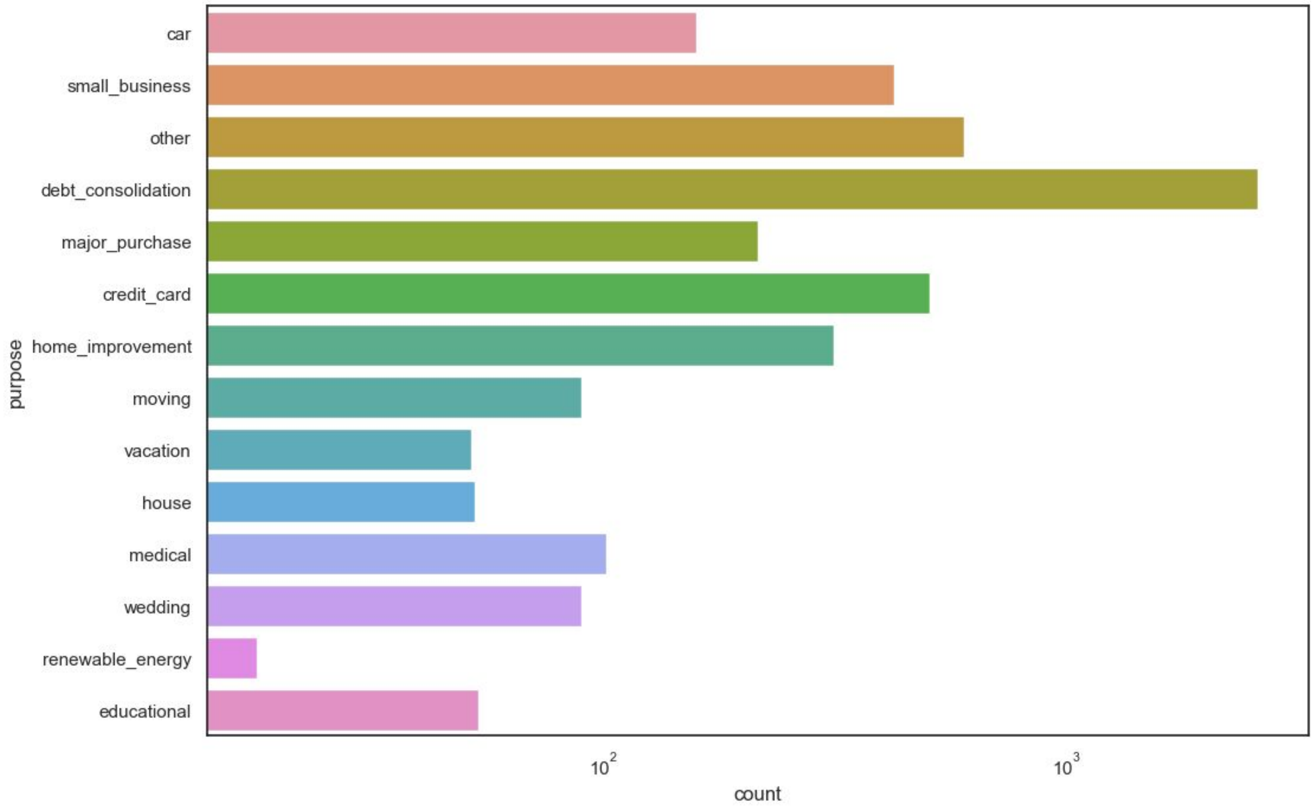
Interest Rate Distribution





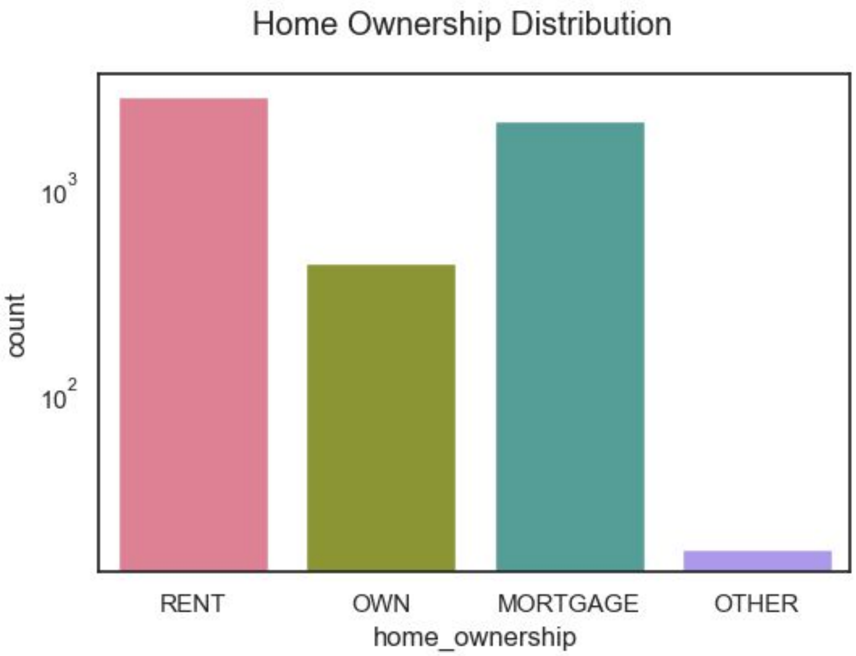
- Most of the loans interest rate lies between 8% to 14%.

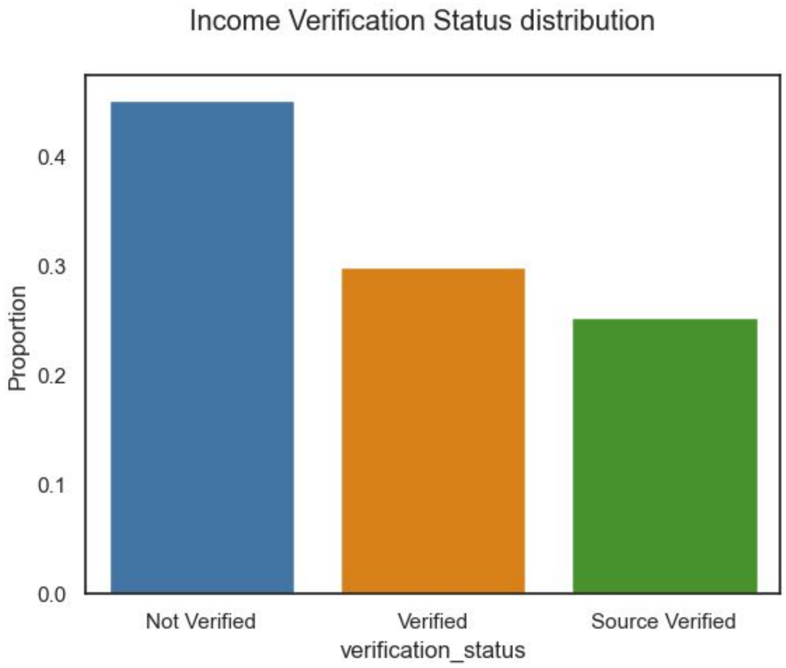
Loan Purpose Distribution



- More number of people took loan for debt consolidation and few took for renewable energy.

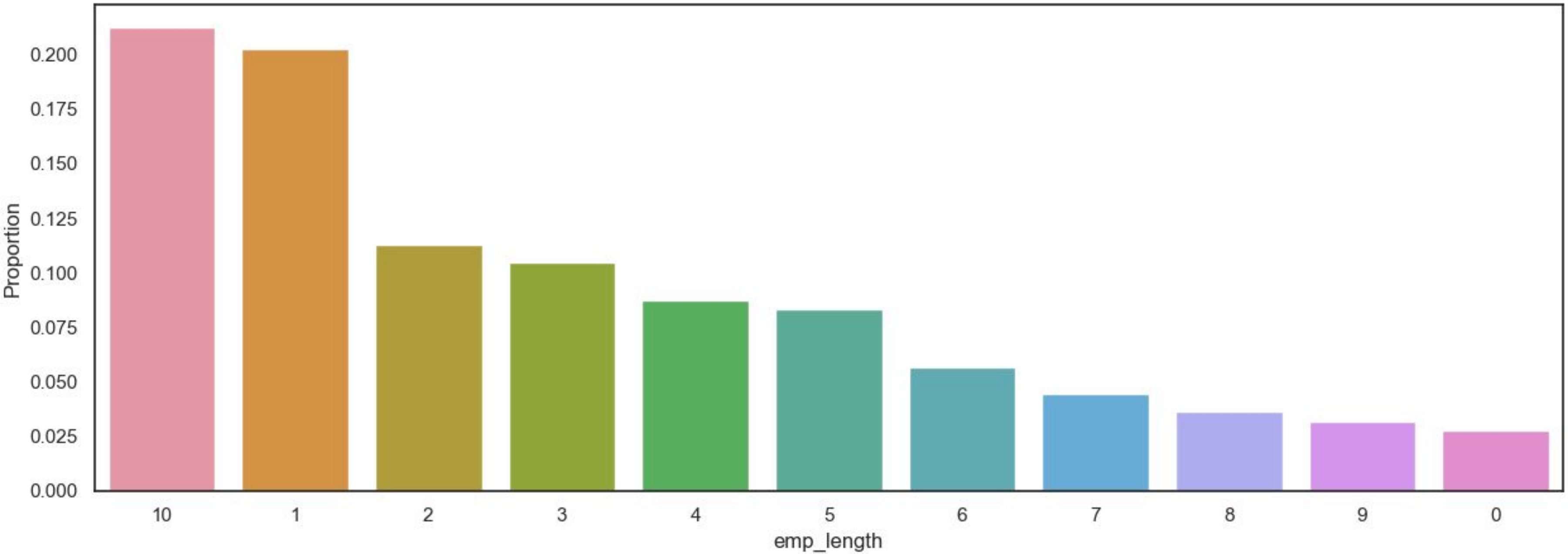
Home Ownership , Verification Status Distribution





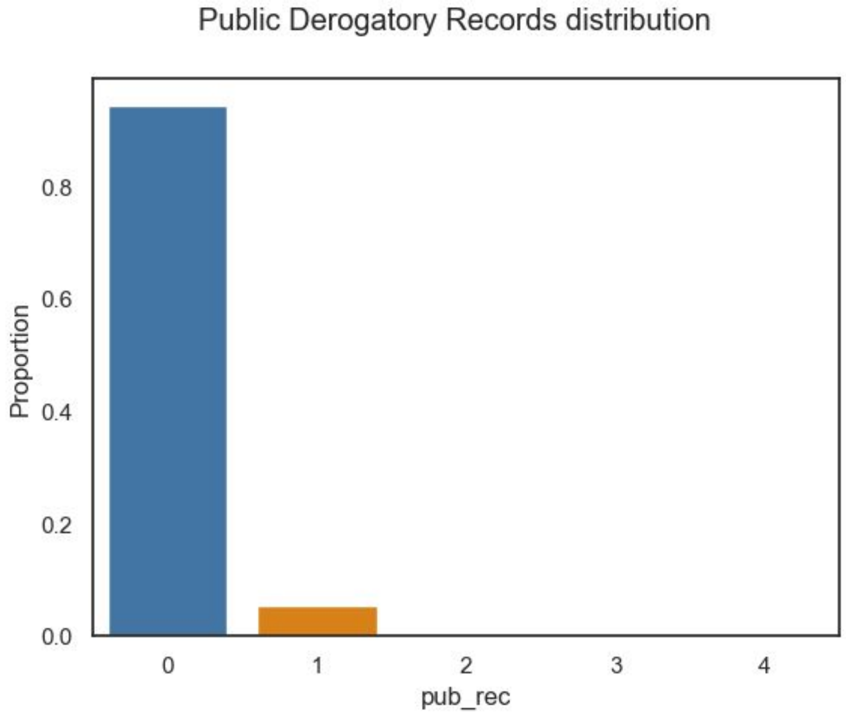
- Loan borrowers have mostly have rented or mortgaged houses.
- The majority of loans were disbursed without verifying the applicants.

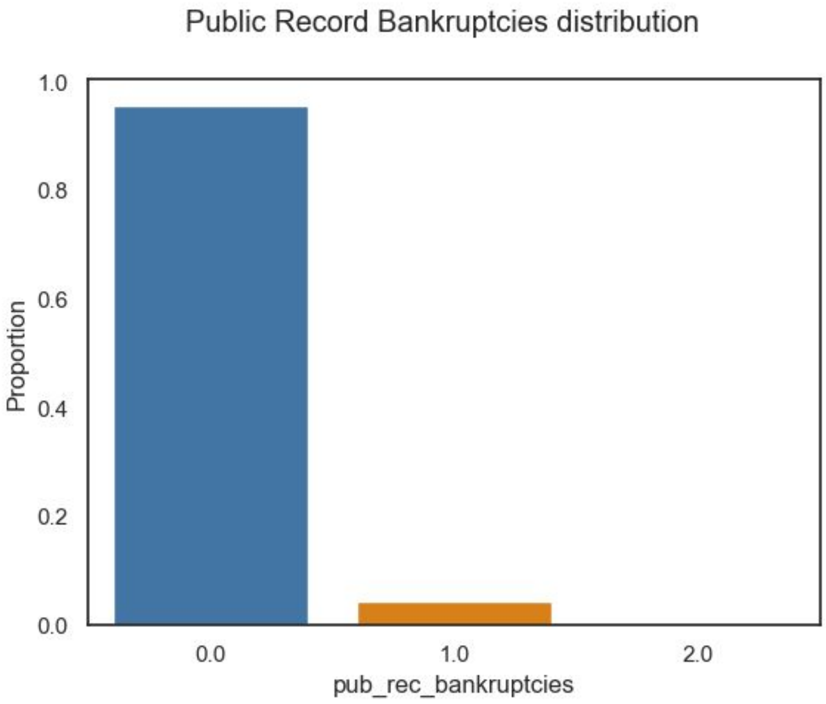
Employment Length Distribution



- Most borrowers have over 10 years of experience, with the second-largest group having less than 1 year.

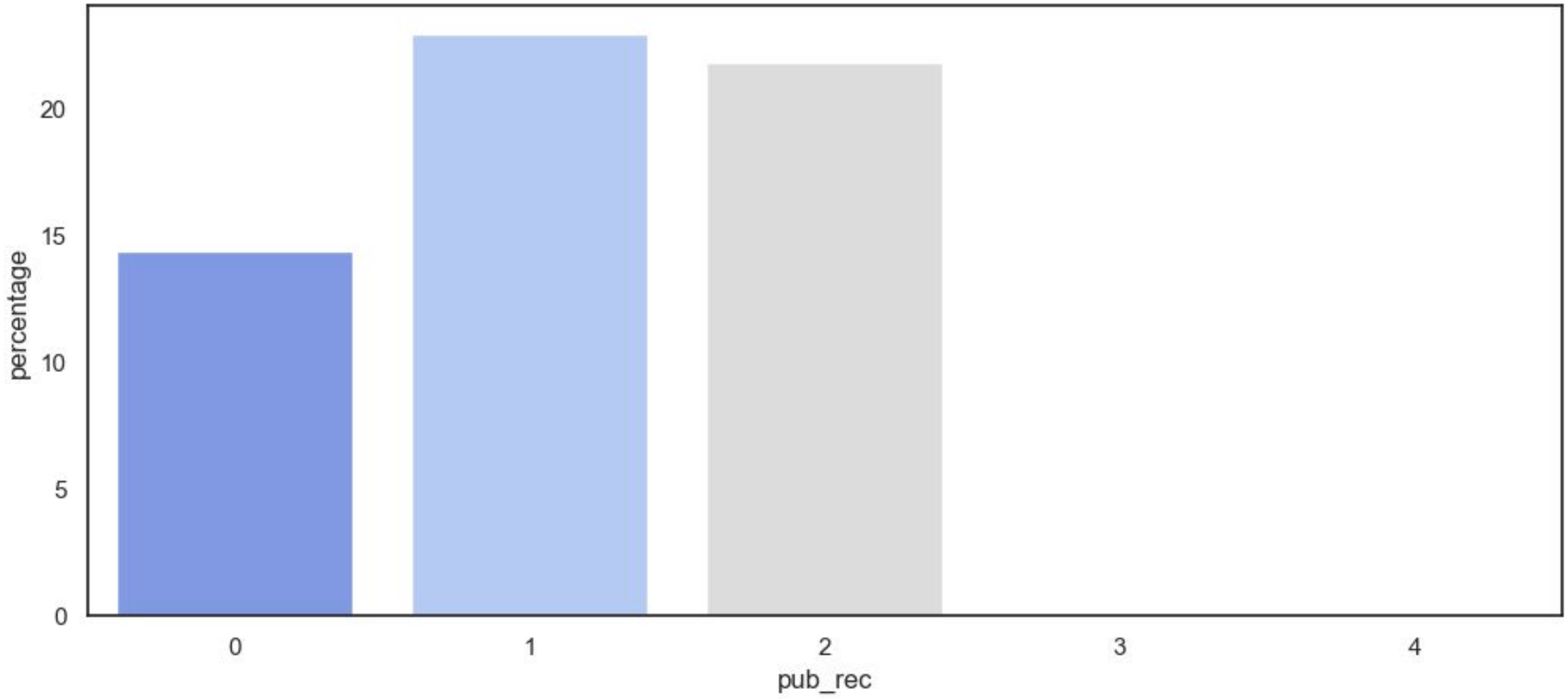
Public records , Bankruptcies Distributions





- Around 94% borrowers have no public derogatory records.
- Nearly 96% of borrowers have no recorded instances of Public Record Bankruptcies.

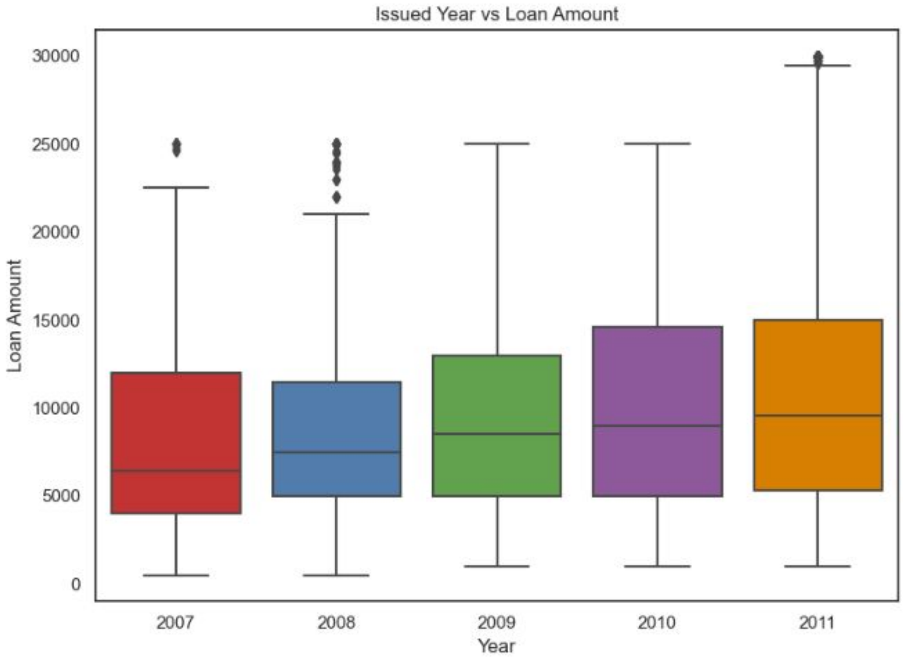
Public records vs Defaults

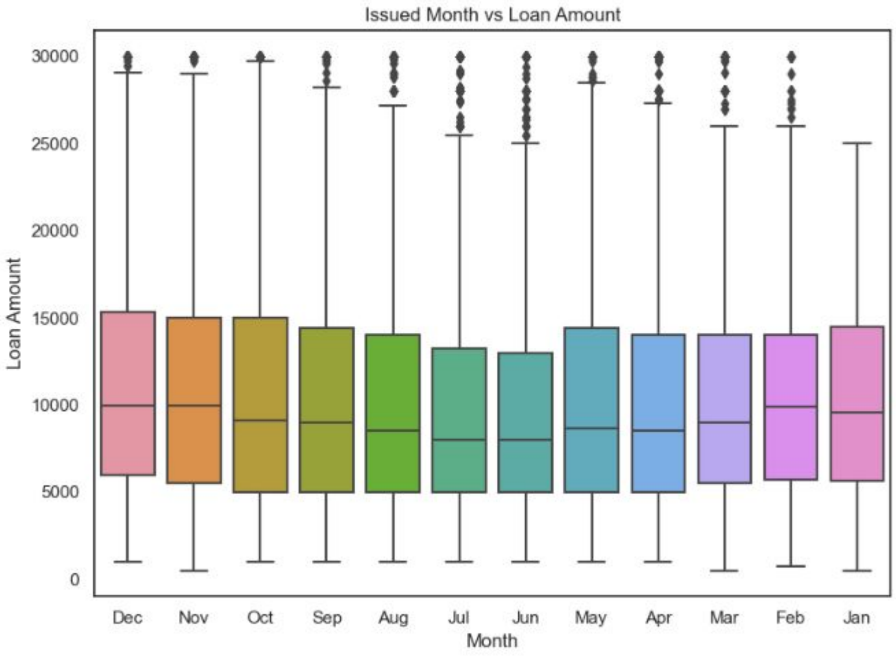


- Percentage of defaults for people with 1 & 2 public records is 1.5X of those without any public derogatory records
- Surprisingly, there aren't any defaults for people with high (3+) public derogatory records

Segmented Analysis

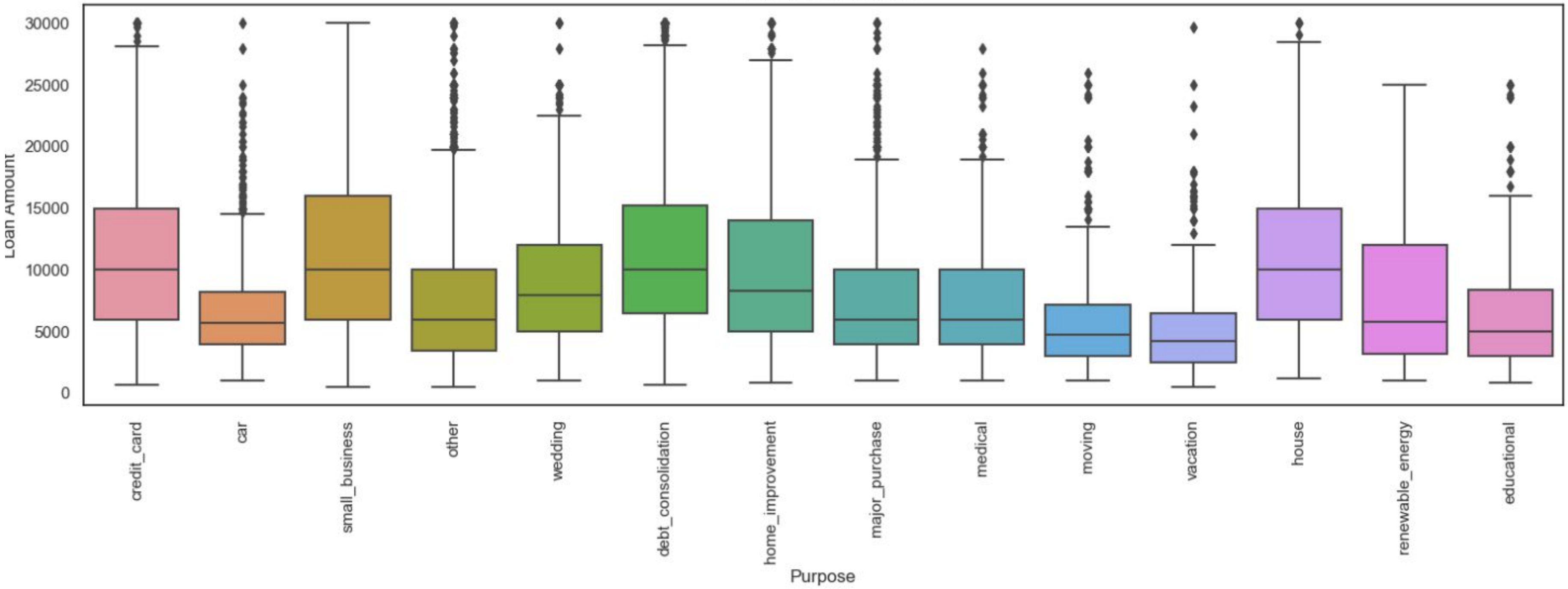
Year , Month vs Loan Amount





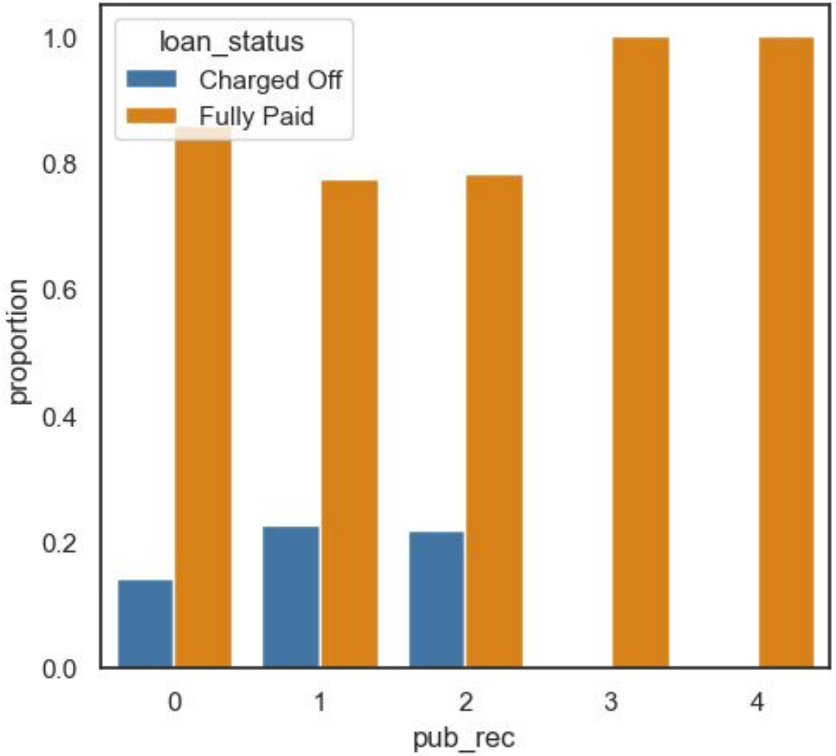
- The average loan amount has been steadily rising year by year.
- The average loan amount appears to be slightly higher for loans disbursed in December.

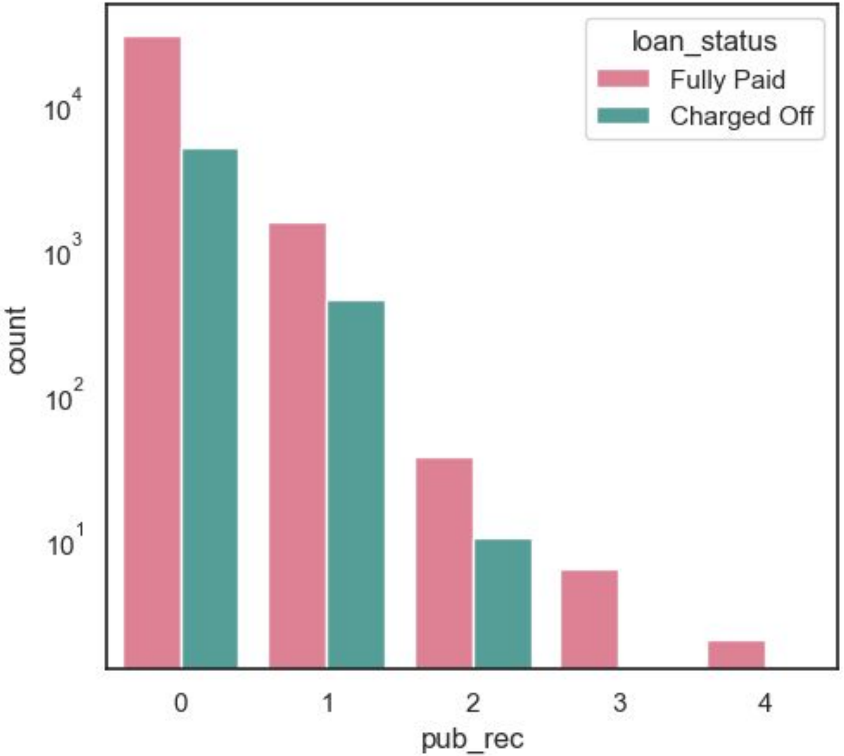
Loan Purpose vs Loan Amount



- The majority of charged-off loans are small business loans.

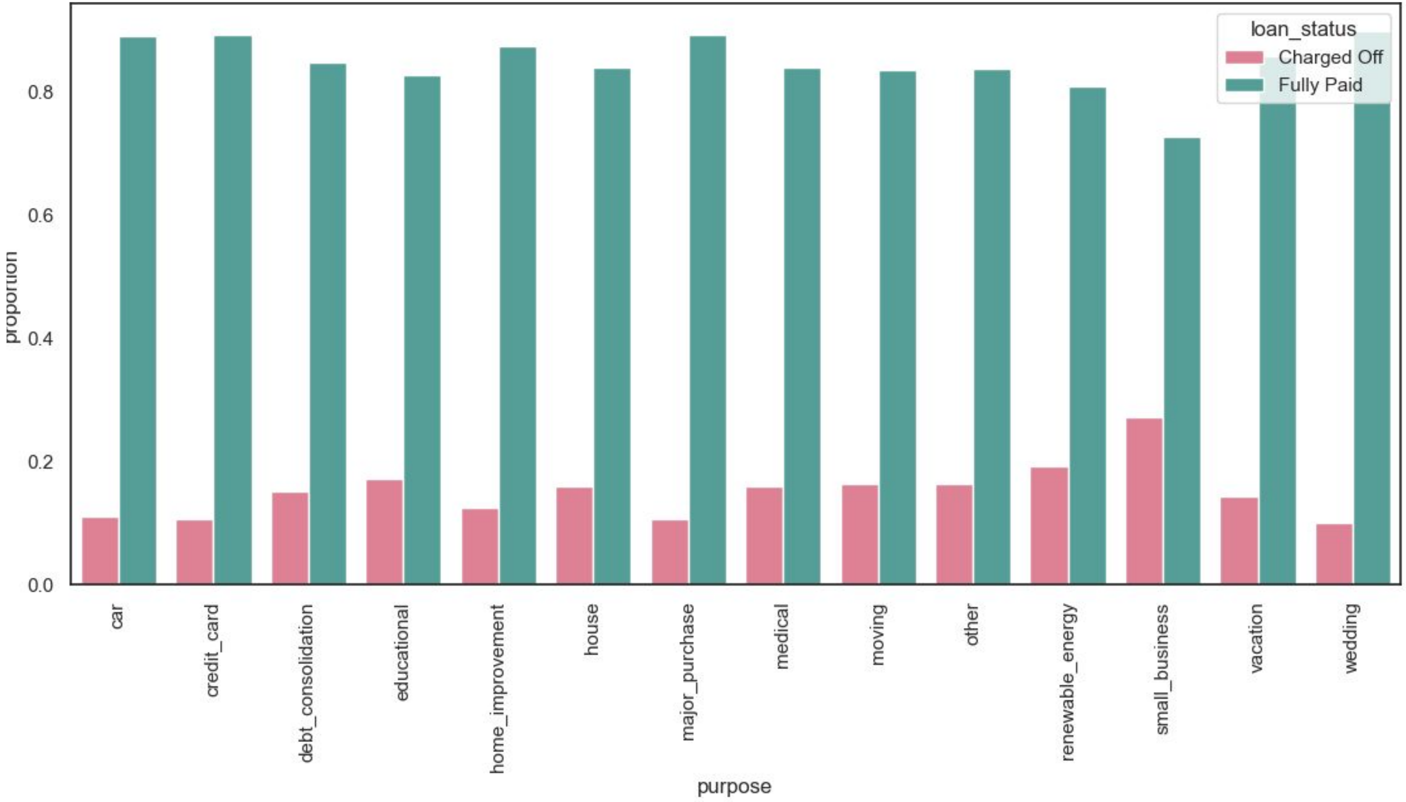
Derogatory public records vs Defaults





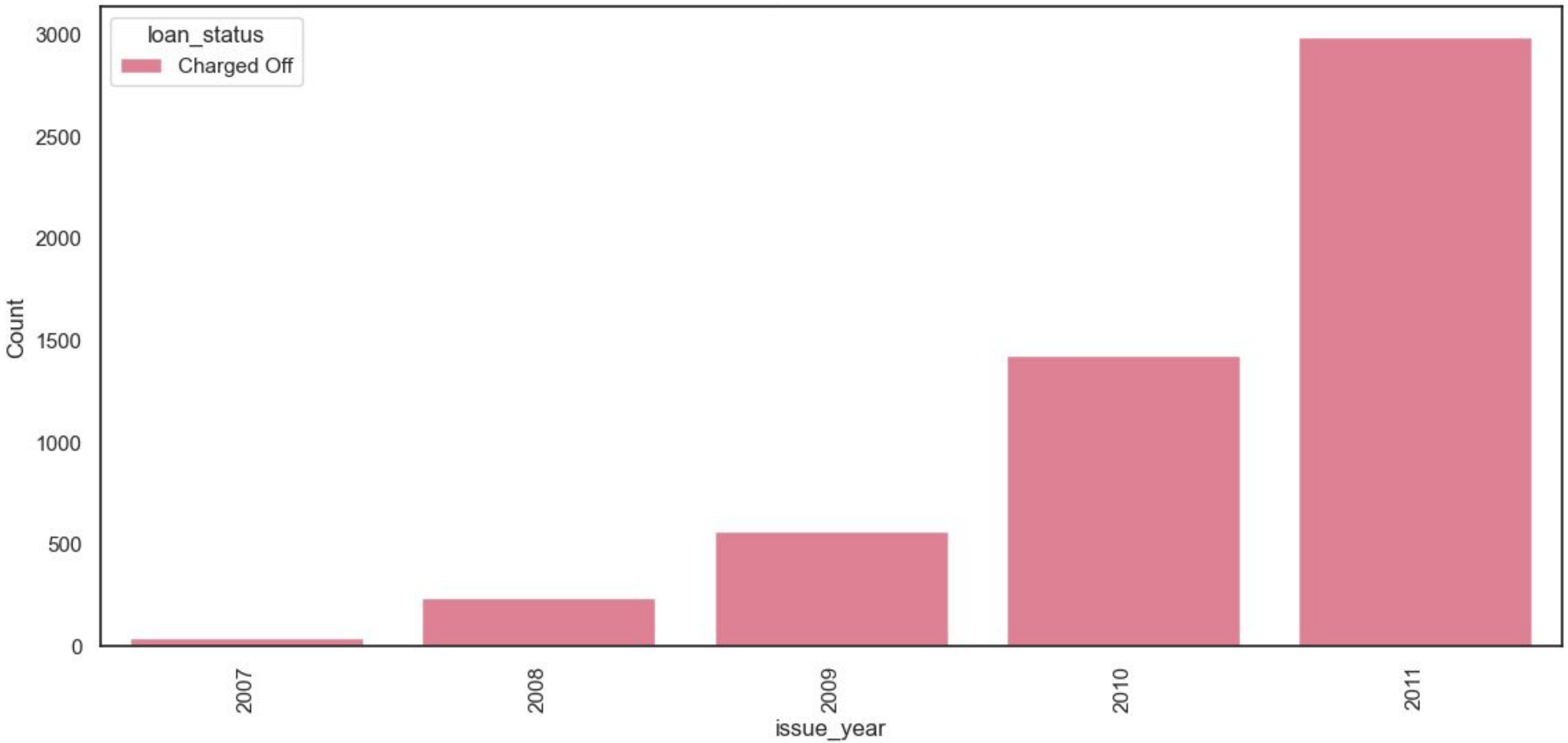
- Most of the charged-off loans are from borrowers with no derogatory public records.

Loan Purpose vs Loan Status



- The majority of charged-off loans are small business loans.

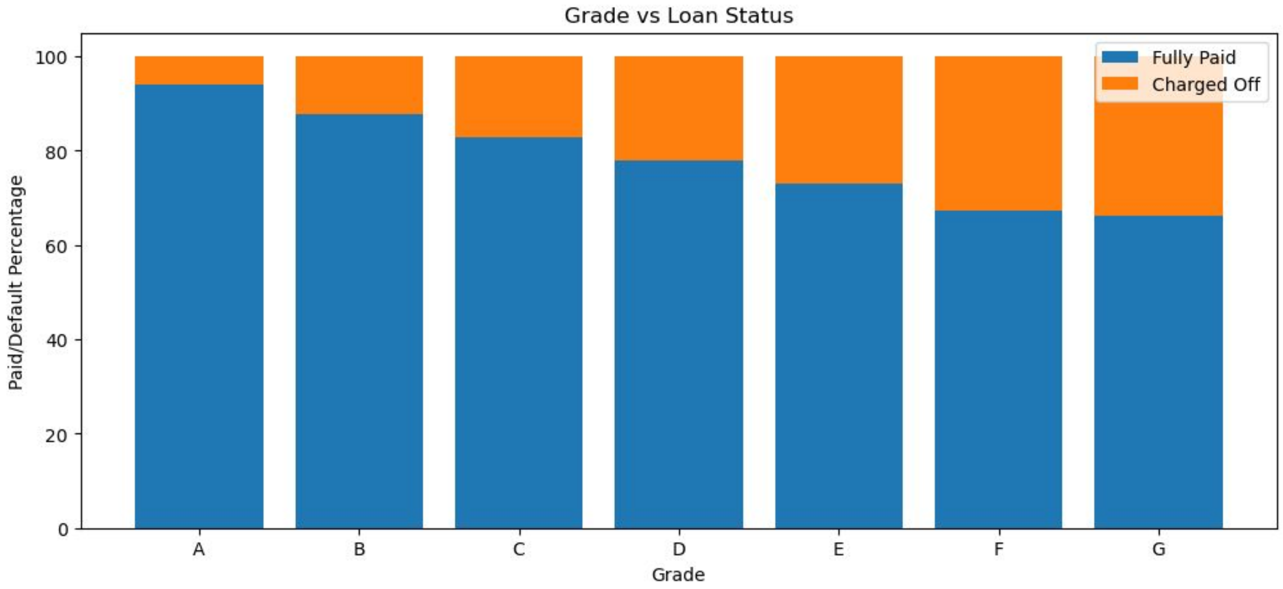
Loan Issued Year vs Defaults



- Each year, there is a substantial rise in the count of charged-off loans.

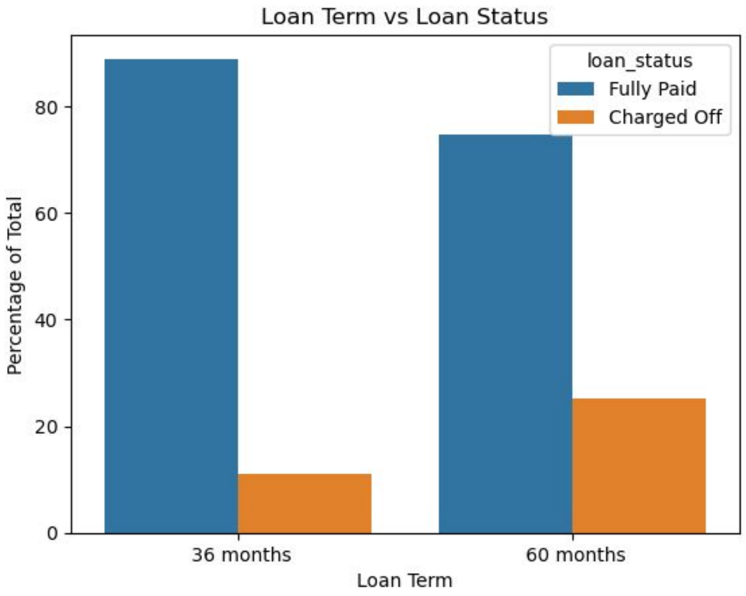
Bivariate Analysis

Loan Grade vs Defaults

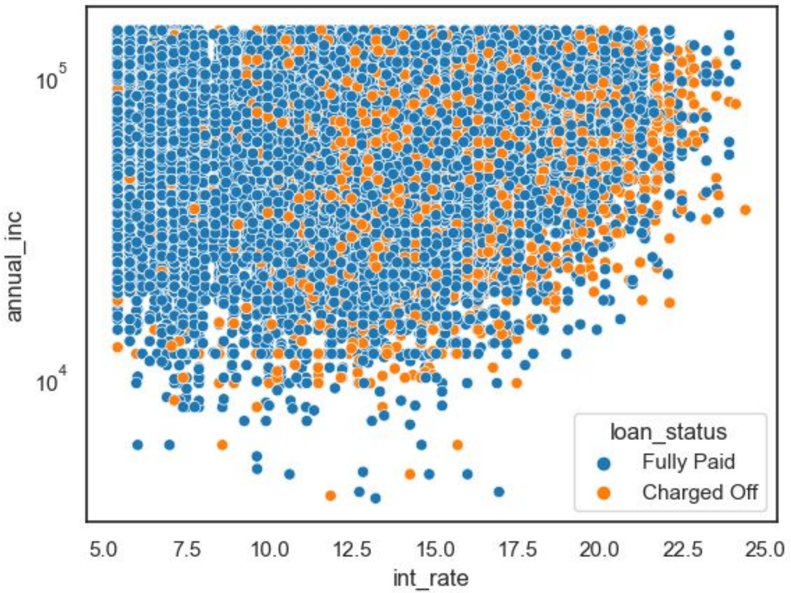


- The lower the grade of the loan, the higher the chance of default.

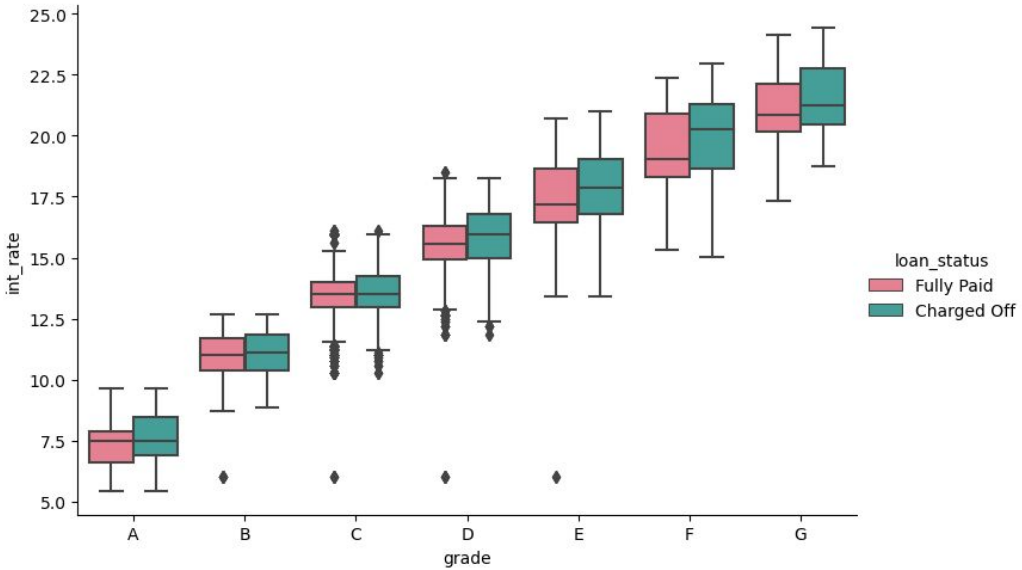
Term, Interest Rate, Annual Income vs Defaults



- 60 Month loan term has a higher rate of default.
- Irrespective of annual income, higher interest rates show higher defaults.

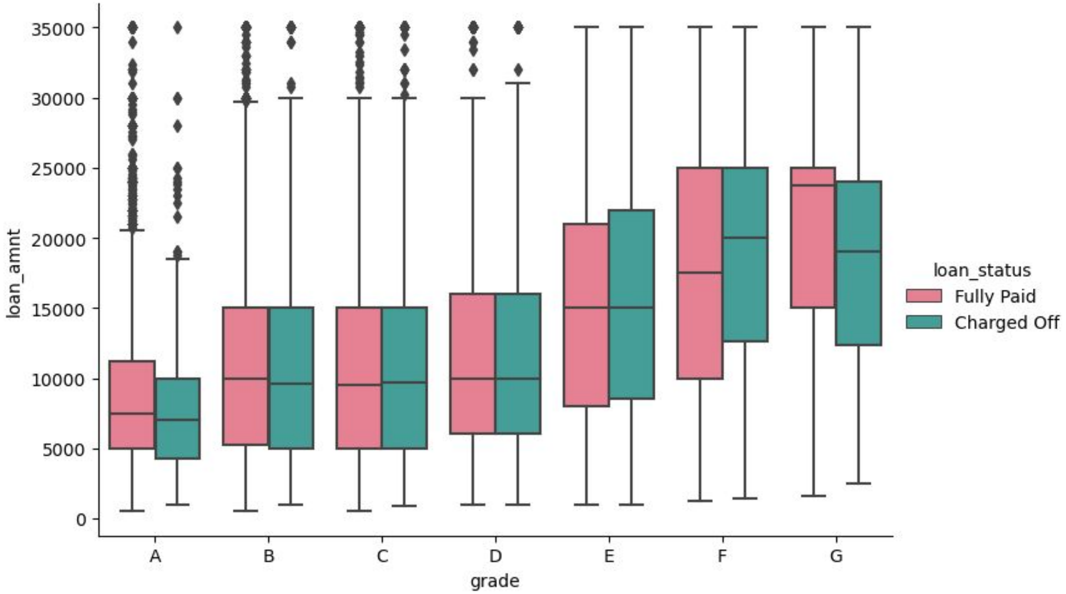


Loan Grade vs Defaults



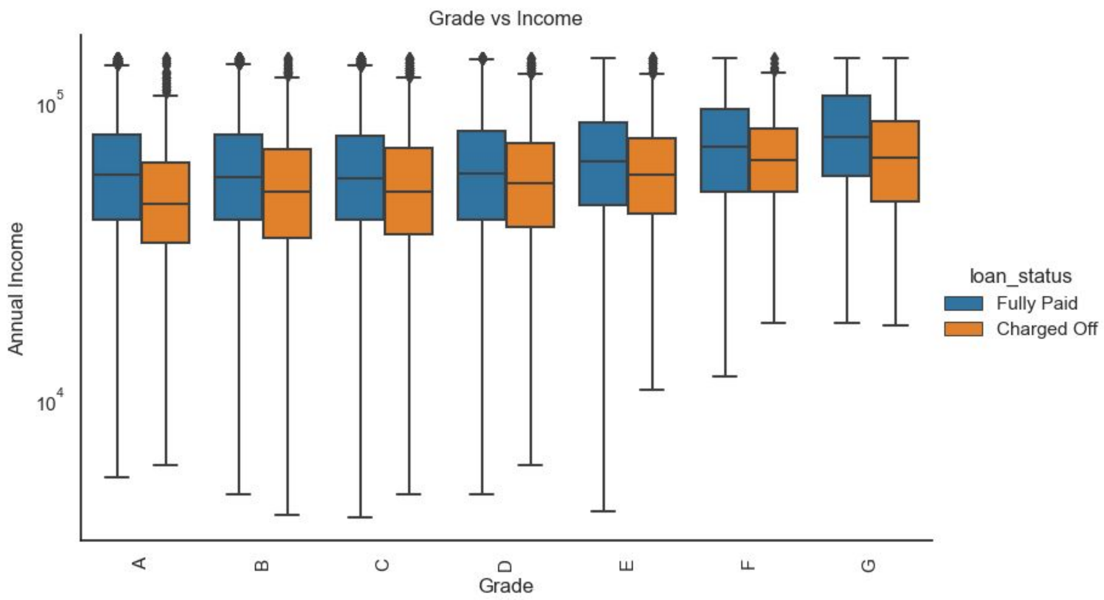
- As grade decreases, you see a significant difference in median interest rates between 'Charged Off' loans vs 'Fully Paid' loans

Loan Grade, Loan Amount vs Defaults



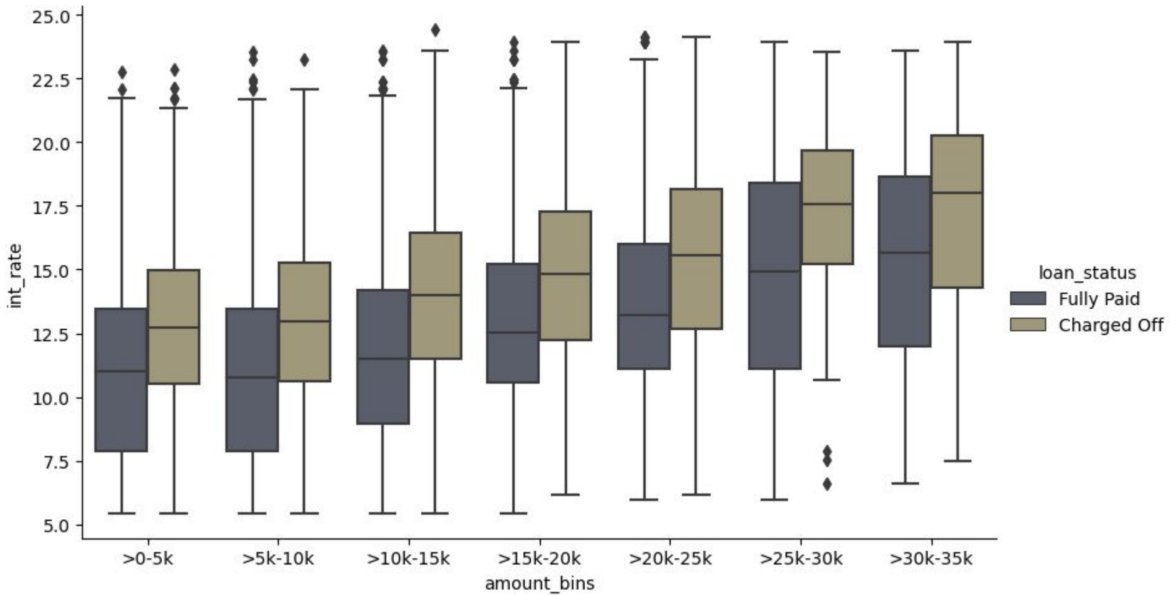
- Loan amounts and grades do not show any specific pattern here. Median values are pretty much the same across the grades.

Grade vs Income vs Defaults



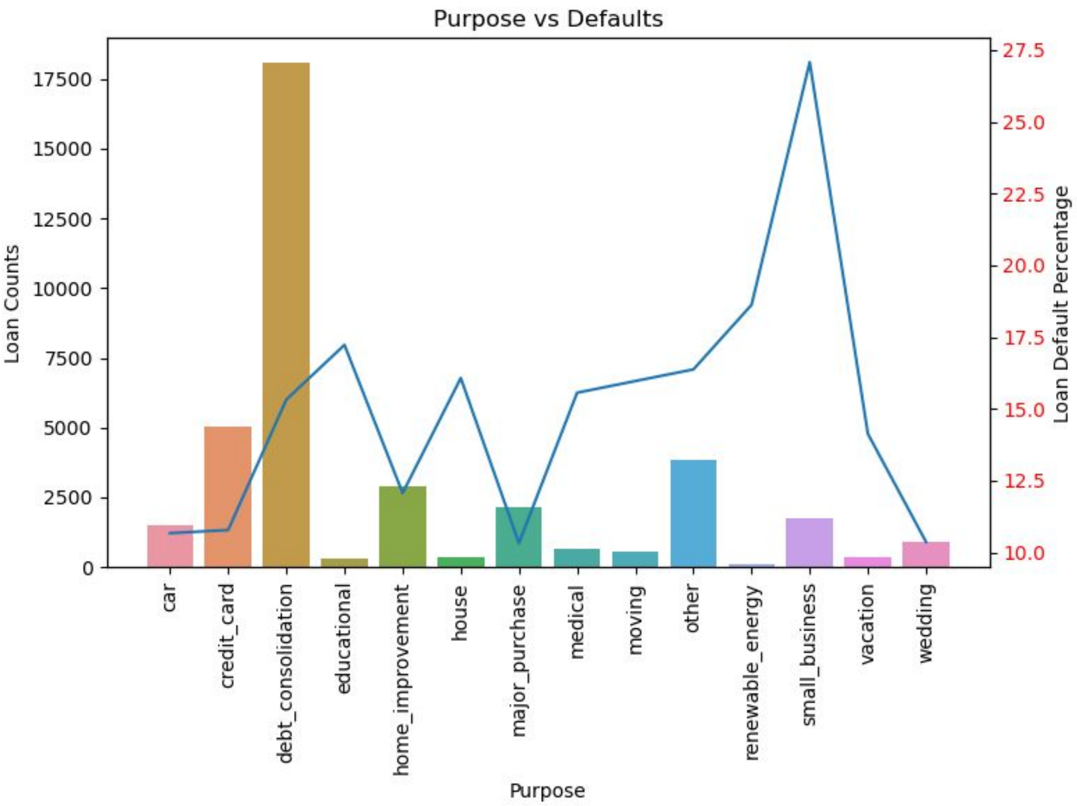
- Across grades, lower income has shown higher rates of defaults.

Interest Rate vs Defaults



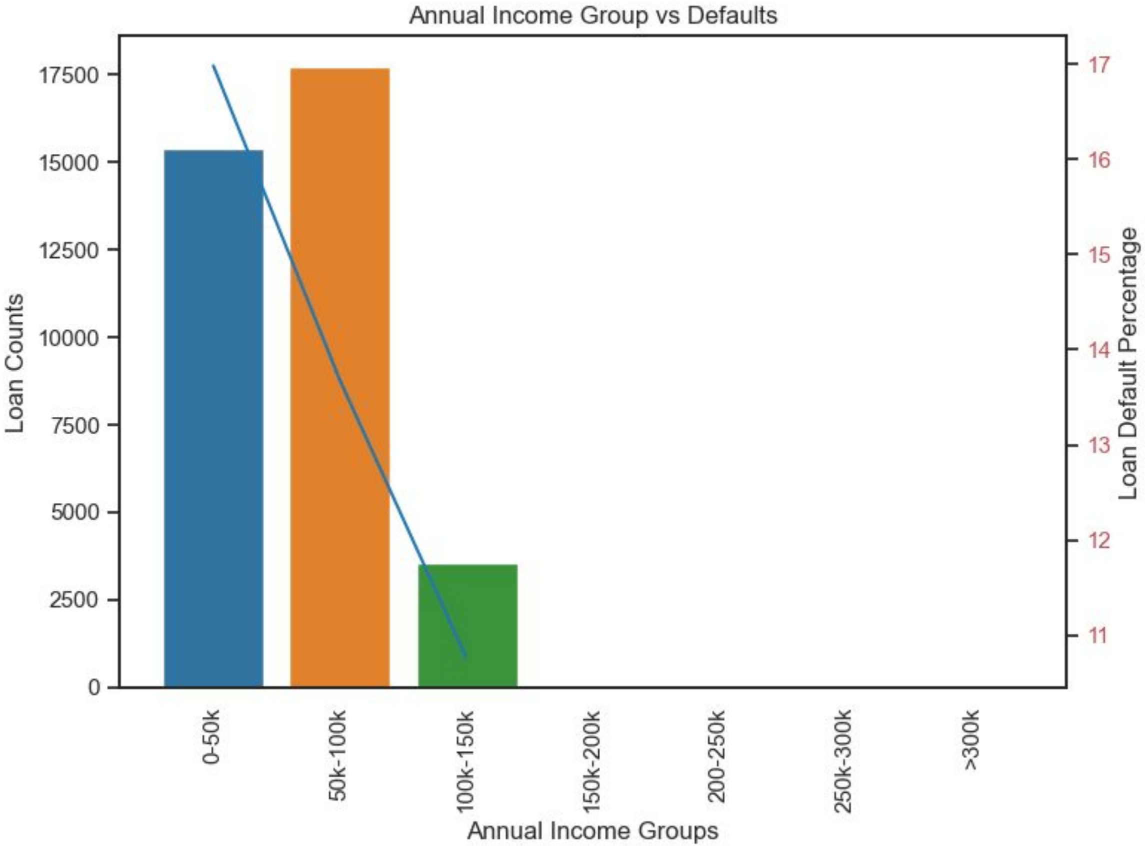
- Irrespective of loan amount, higher interest rate shows increased risk

Purpose of Loan vs Defaults



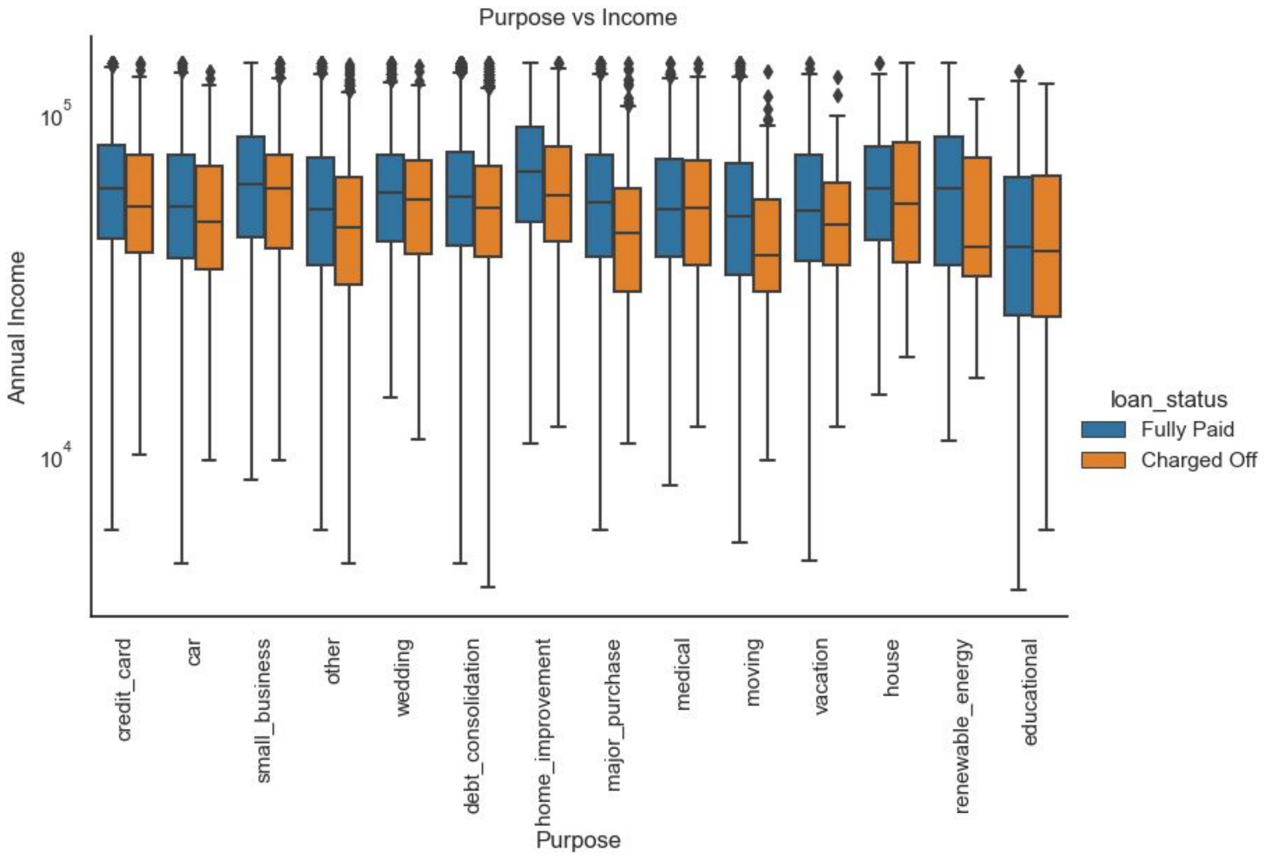
- Debt consolidation is the main reason for people to take a loan.
- While debt consolidation has a high default rate, the highest default rate is for small businesses.

Annual Income Group vs Defaults



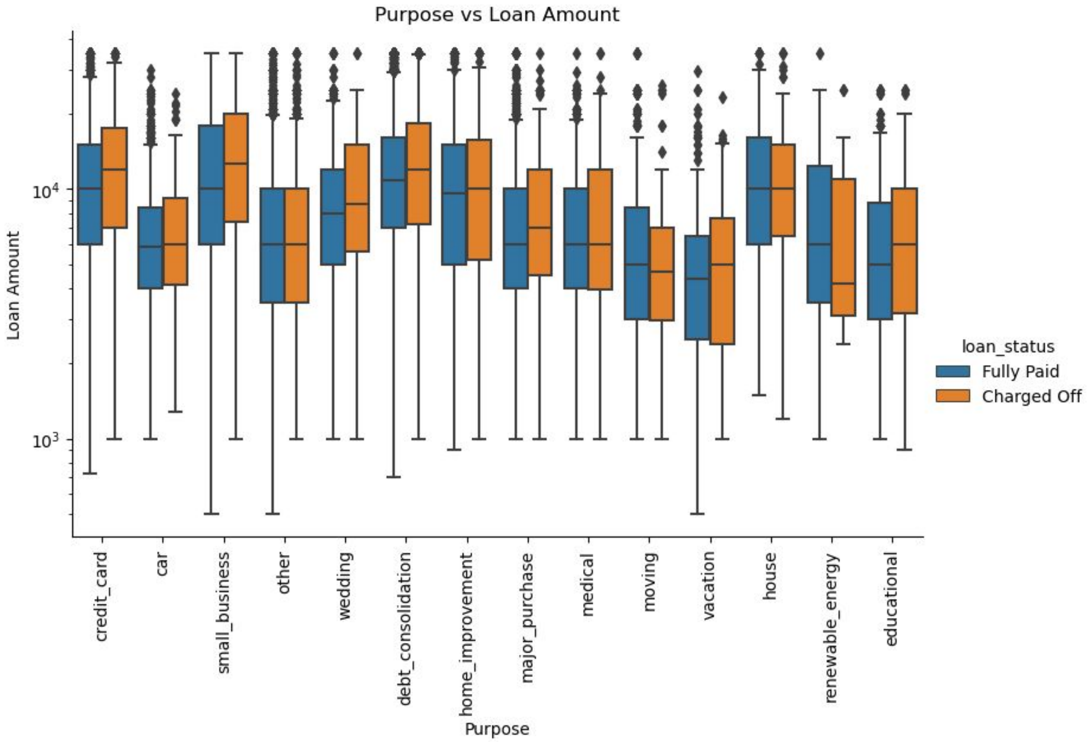
- Most loans are given to borrowers with annual income <100K.
- Some of the highest default rates are also for this group of borrowers.

Purpose vs Income



- Across loan purposes, median annual income of defaulters is less than those who pay fully.
- The difference is stark in case of loans taken for Renewable Energy, Home Improvement & Major purchase.

Purpose vs Loan Amount



- The median loan amount for those loans that are defaulted, is the highest in case of loans taken by small businesses.
- Loan taken for the purpose “Credit Card” also show a high median loan amount for those that are defaulted.

Recommendations to Lending Club

TO REDUCE THE DEFAULTS IN THE FUTURE, HERE ARE THIS PROJECT GROUP'S RECOMMENDATION TO LENDING CLUB

- Given the correlation between lower loan grades (Grades D - G) and higher chances of default, Lending Club should continue to refine its risk assessment and grading system. Also, since lower loan grades (Grades D-G) have higher chance of default, they should scrutinize these loans more.
- Lending Club could consider imposing limits on the loan amount for certain high-risk loan purposes, especially for small businesses and credit card-related loans.
- Lending Club should closely monitor and manage loan portfolios related to small businesses, credit card purposes, and other high-default groups.
- Since 60-month loan terms have a higher rate of default, Lending Club might want to review the feasibility of extending such long-term loans, especially to borrowers with a history of financial instability.
- Given that a significant number of loans are extended to borrowers with annual incomes below 100k, Lending Club should have improve its income verification processes
- Public Derogatory Records should be checked to see if people have some -ve records. Moderate number of records (1-2), seems to significantly increase the chance of default. Exercise caution when lending to such borrowers.