Lending Club Case Study

By Tejas Hankare & Madhur Gupta

Problem Statement :

- Like most other lending companies, lending loans to 'risky' applicants is the largest source of financial loss (called credit loss). The credit loss is the amount of money lost by the lender when the borrower refuses to pay or runs away with the money owed.
- Expected Solution:
- 1. Identify these risky loan applicants
- 2. Understand the 'Driver Factors' behind 'Loan Default'

Abstract

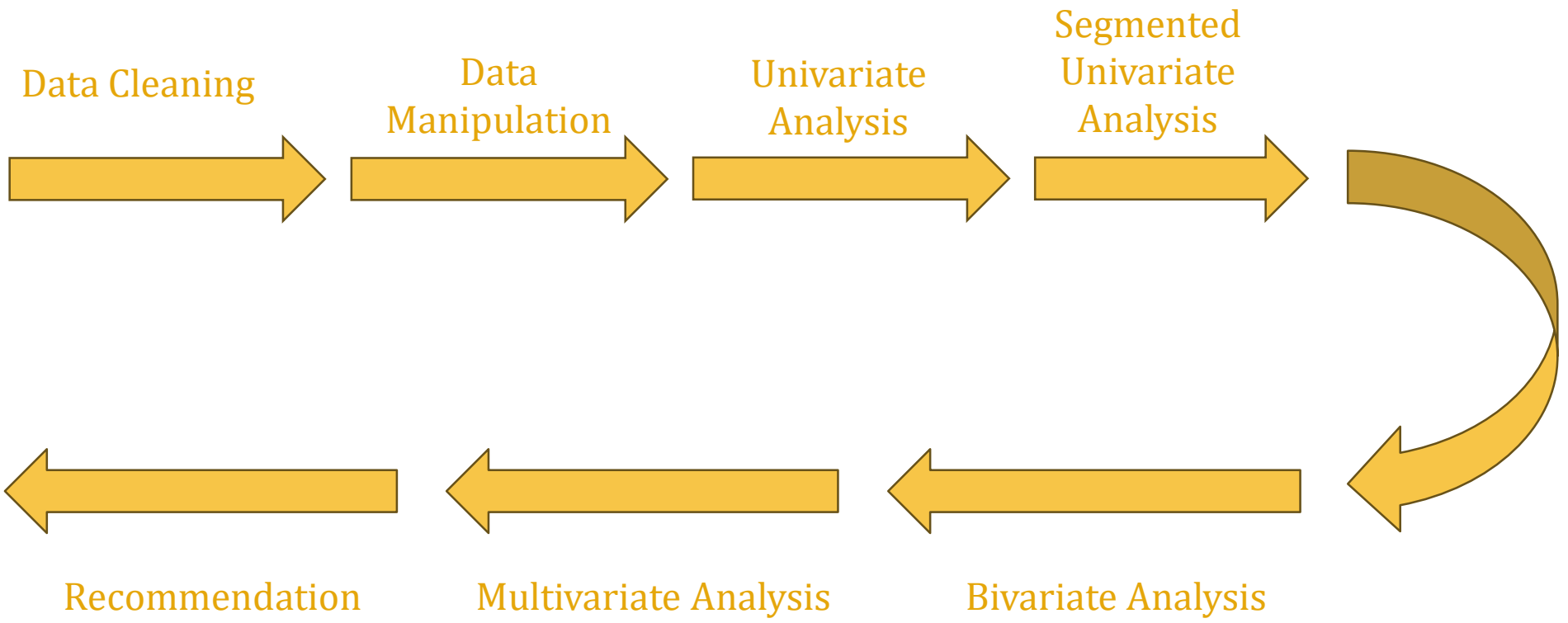
Lending Club is a marketplace for personal loans that matches borrowers who are seeking a loan with investors looking to lend money and make a return.

When the company receives a loan application, the company has to make a decision for loan approval based on the applicant's profile. Two types of risks are associated with the bank's decision:

- If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company.
- If the applicant is not likely to repay the loan, i.e. he/she is likely to default, then approving the loan may lead to a financial loss for the company.

So, company wants 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.

Problem Solving Methodology



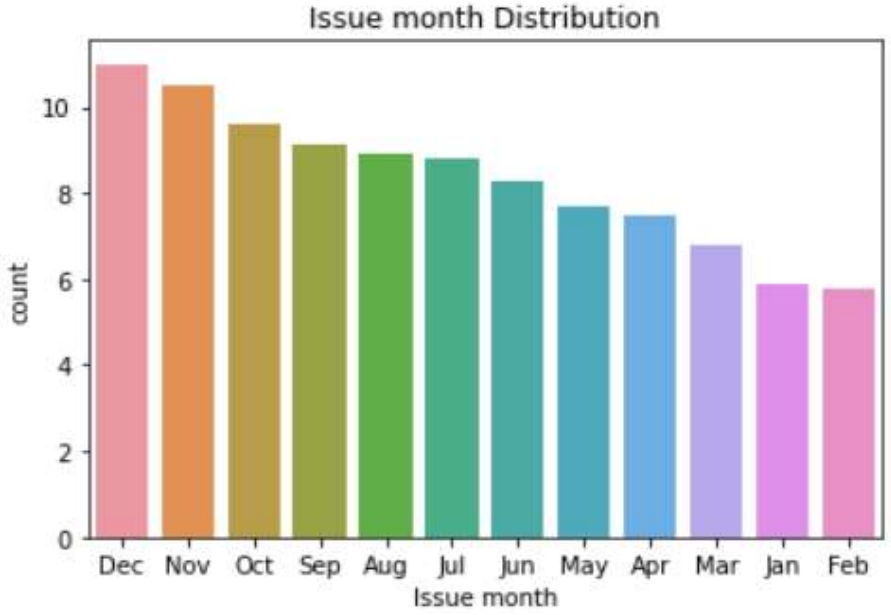
Data Cleaning

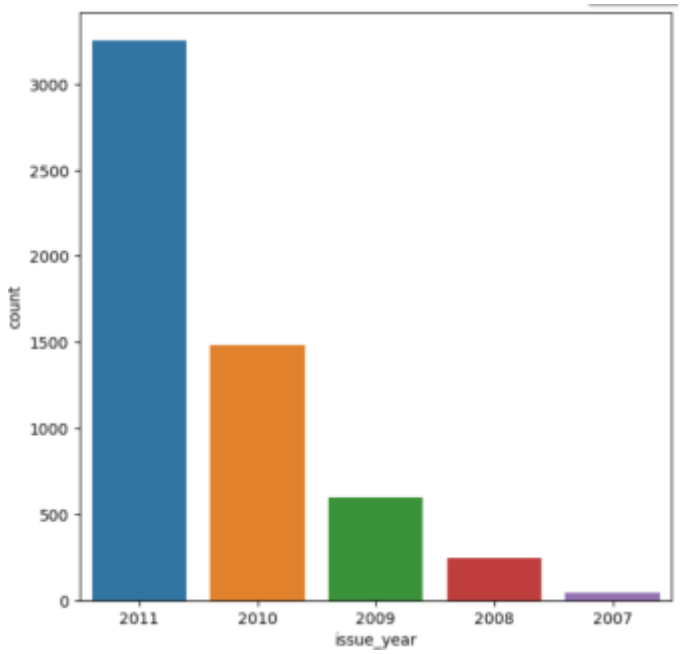
- Removed all the columns with Null values.
- Removed all the rows with majority Null values.
- Removed Columns which are not defined in the Column Definition dataset

Data Manipulation

- Data manipulation refers to the process of changing or transforming data to make it more suitable for analysis or visualization.
- Standardizing data formats, such as converting date columns to a consistent format.
- Handling missing values by either imputing them or removing rows with missing data.

Analysis





- Lending club has really expanded year by year, every year the number of loans has doubled.
- Issue of loan is also increasing every month from January to December and in the final quarter of year there are more loans issued due to vacation.

Univariate Analysis

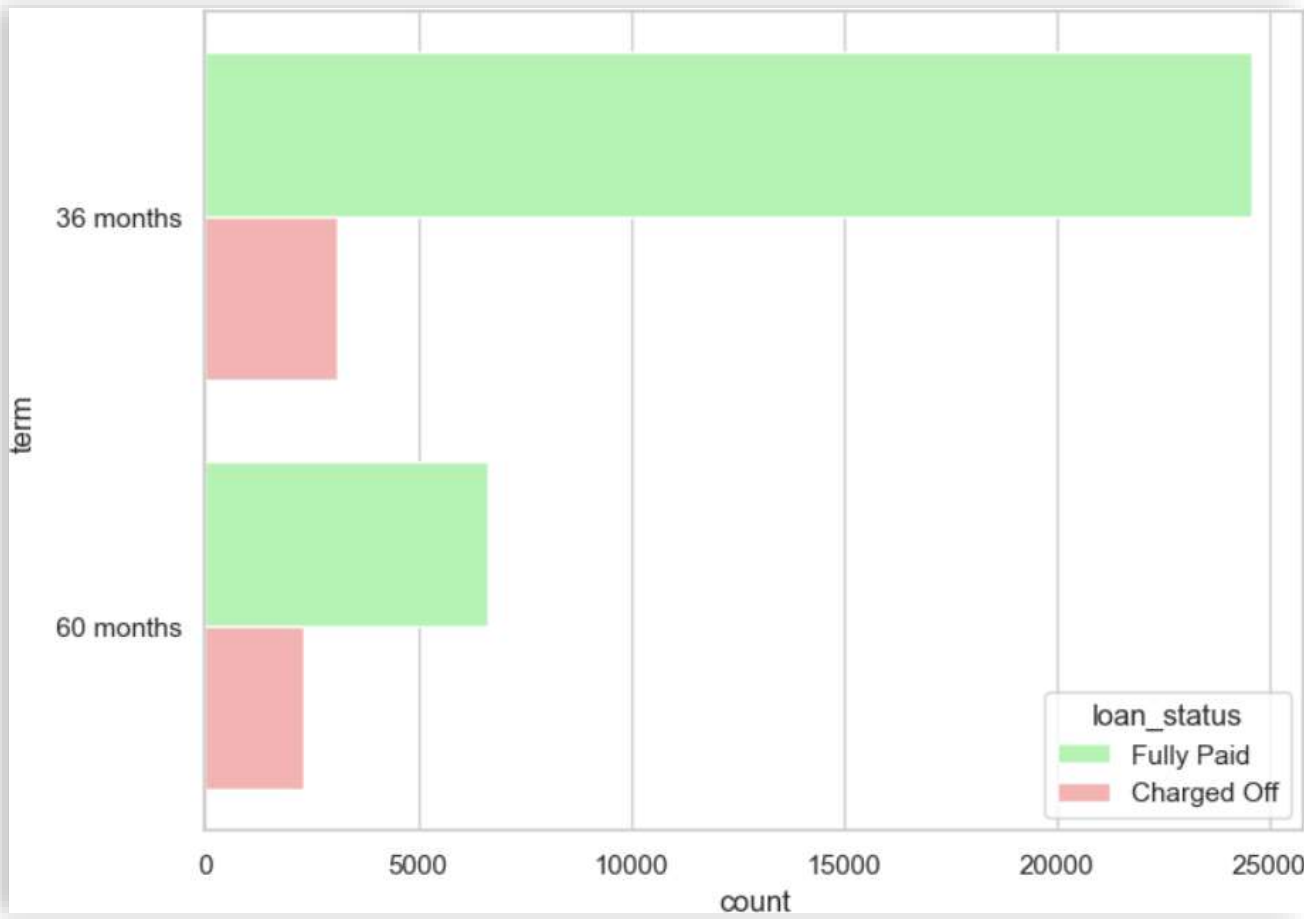
- Univariate analysis is the simplest form of analyzing data, where the data has only one variable.
- Univariate analysis helps in understanding and summarizing data by focusing on
- Univariate data consists of observations on only a single characteristic or attribute, making it a fundamental type of analysis in statistics. a single variable at a time.

Loan Status Distribution

From above box plot we can conclude that "higher the amount" will tend to "Write off".

Paying Term Loan

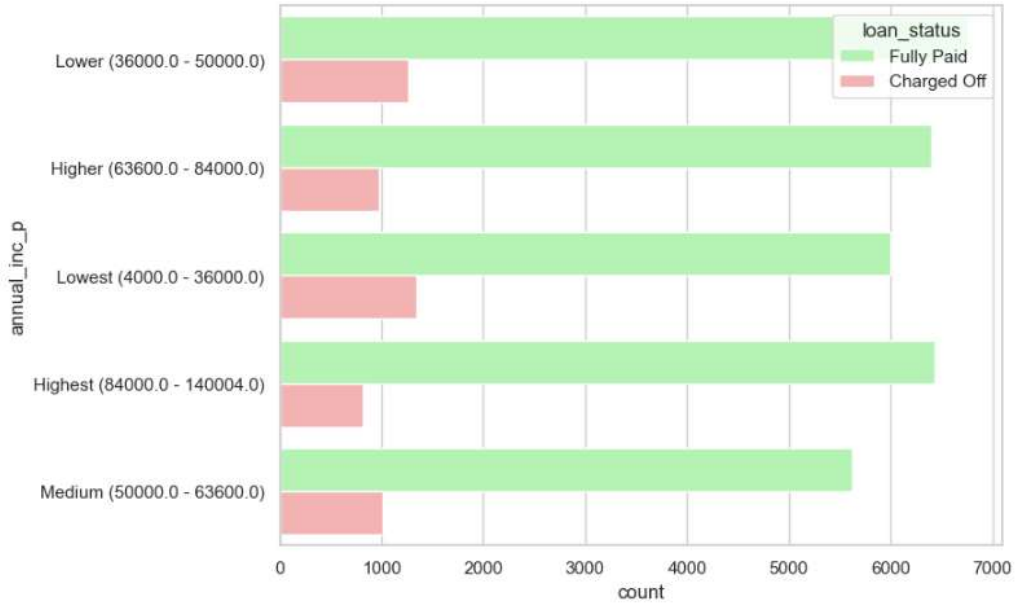
There are 2 loan term and most of the borrower took 36 months tenure. But the ratio of charged off is high in 60 months tenure.



Annual Income

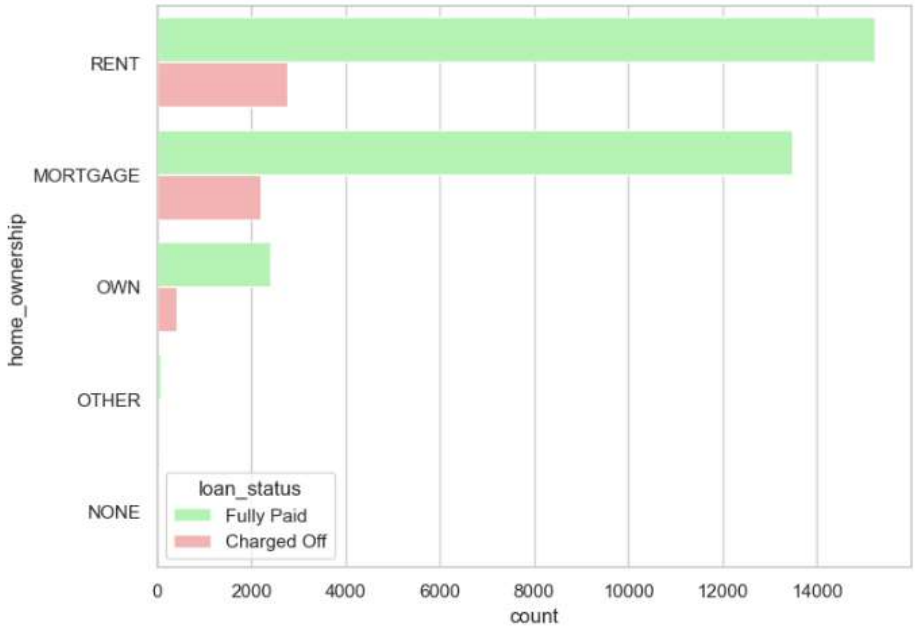
As annual income is increasing, charged off portion is decreasing.

So, we can say that lower the annual income most is the chance of getting defaulted



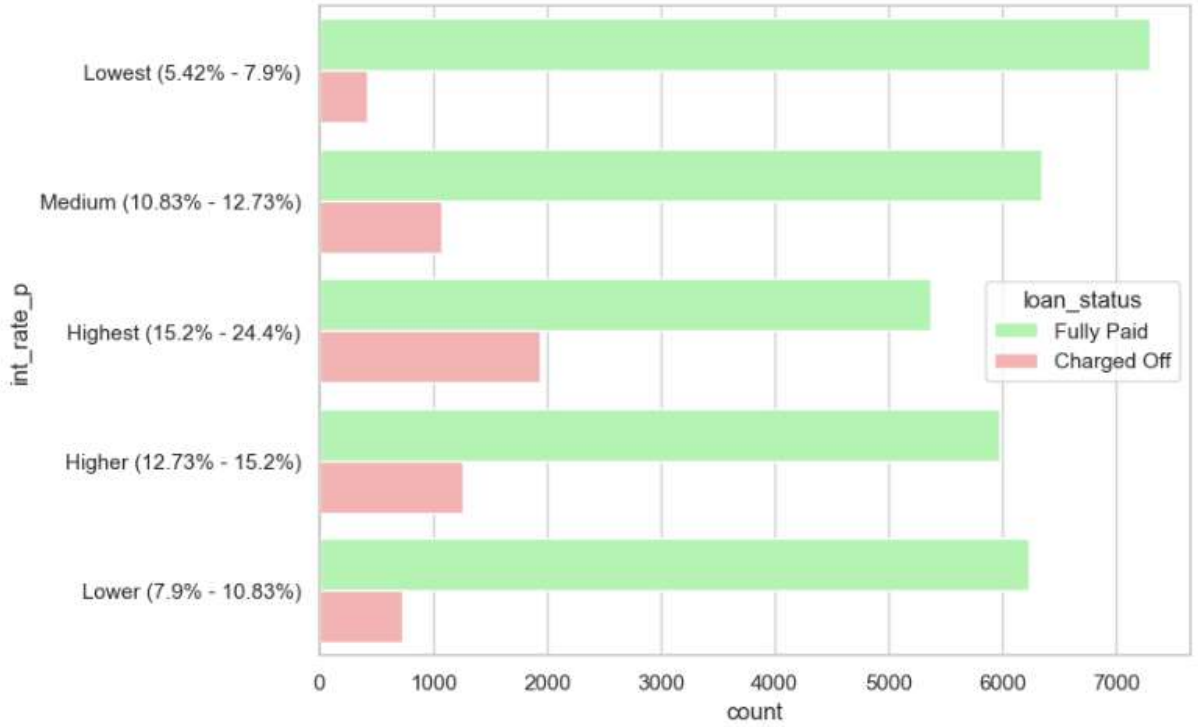
Home Ownership

According to this, we can say that more loan is taken by the borrower whose home ownership is rent and they get charged off when they take loan for the purpose of debt consolidation.



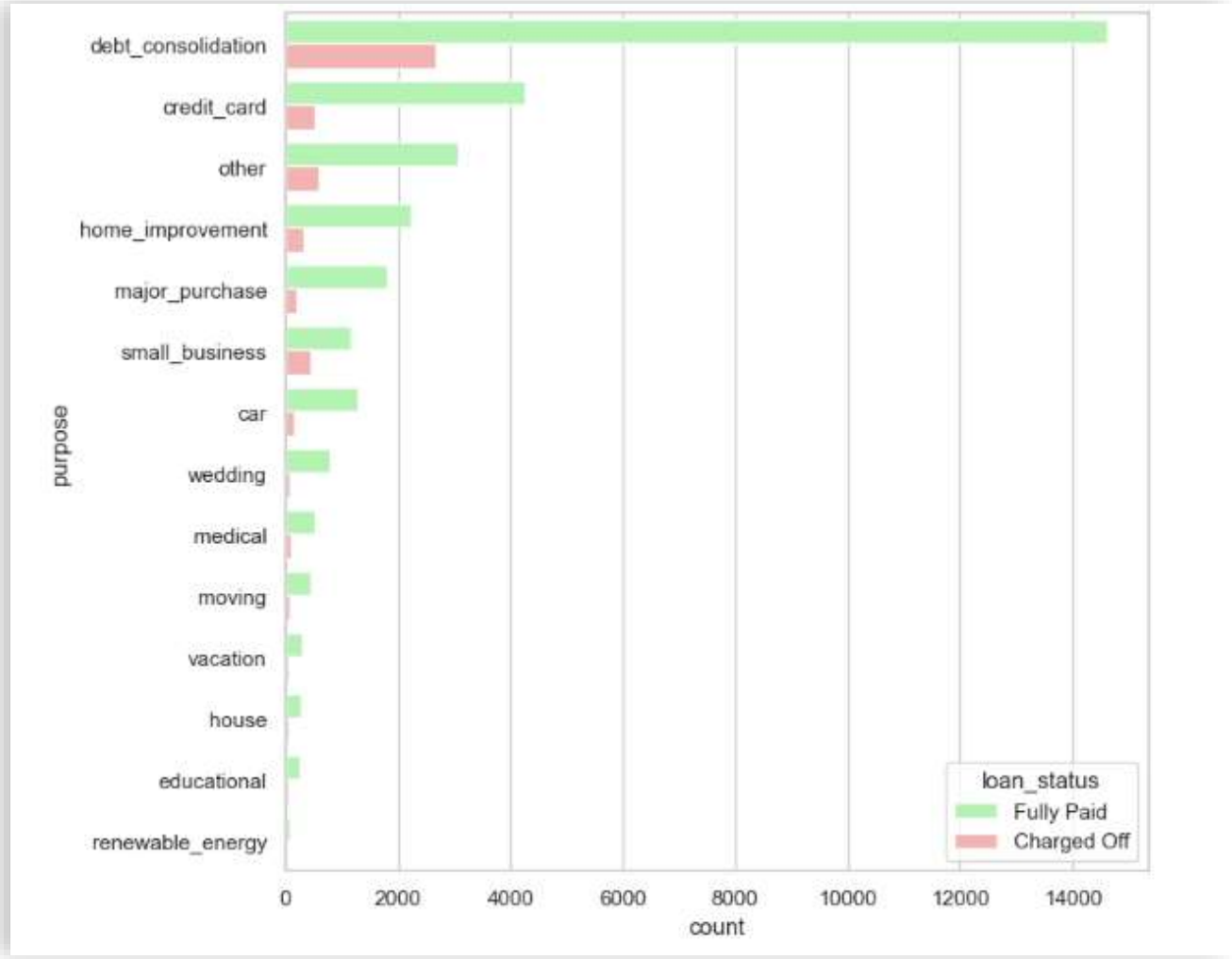
Interest Rate

Loan with 15.2% to 24.4% interest rate are mostly getting defaulted.



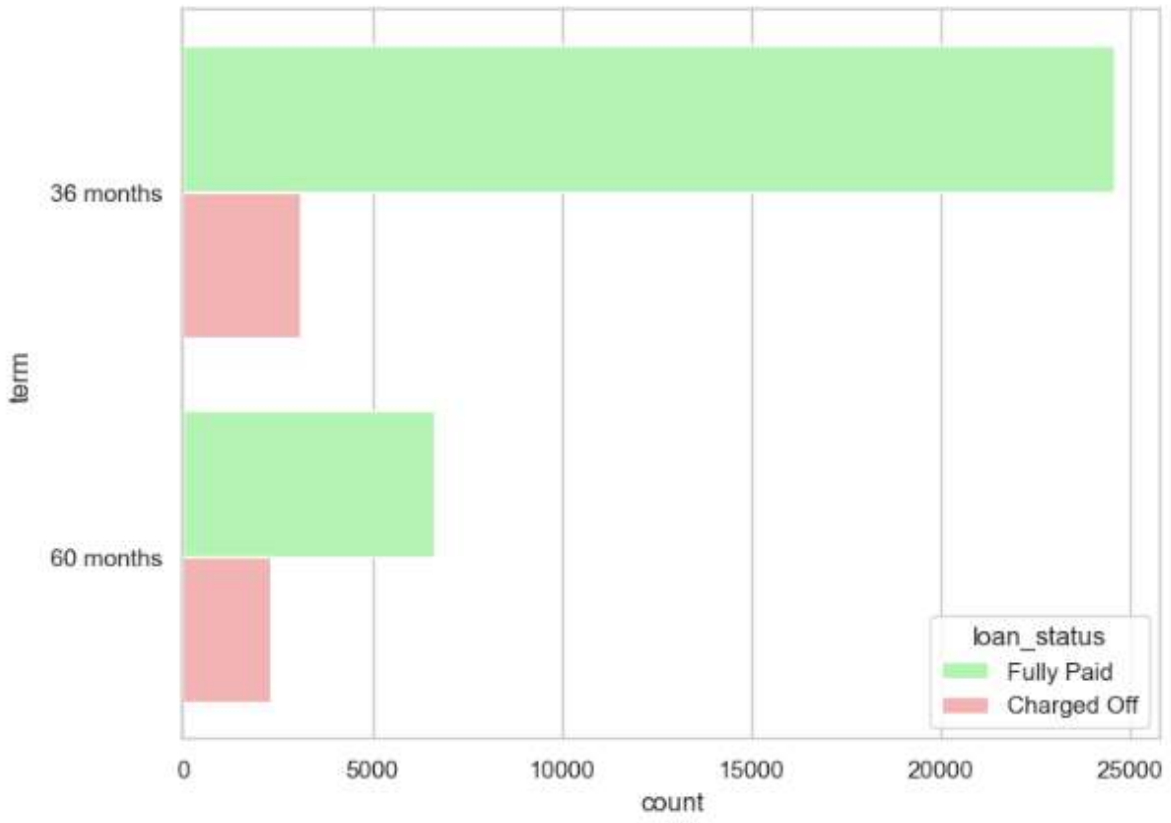
Purpose of Loan

The borrower who took loan for the purpose of small business, renewable energy and educational are mostly getting charged off.



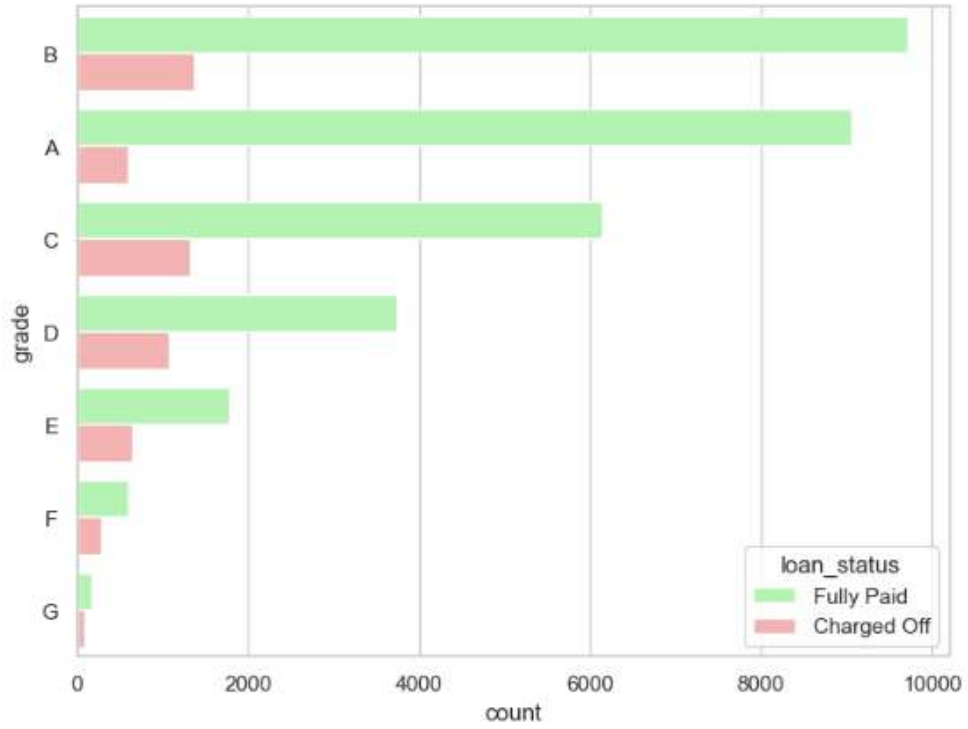
Term VS Loan Amount

The borrower who takes higher loan amount tends to choose the loan term 60 months.



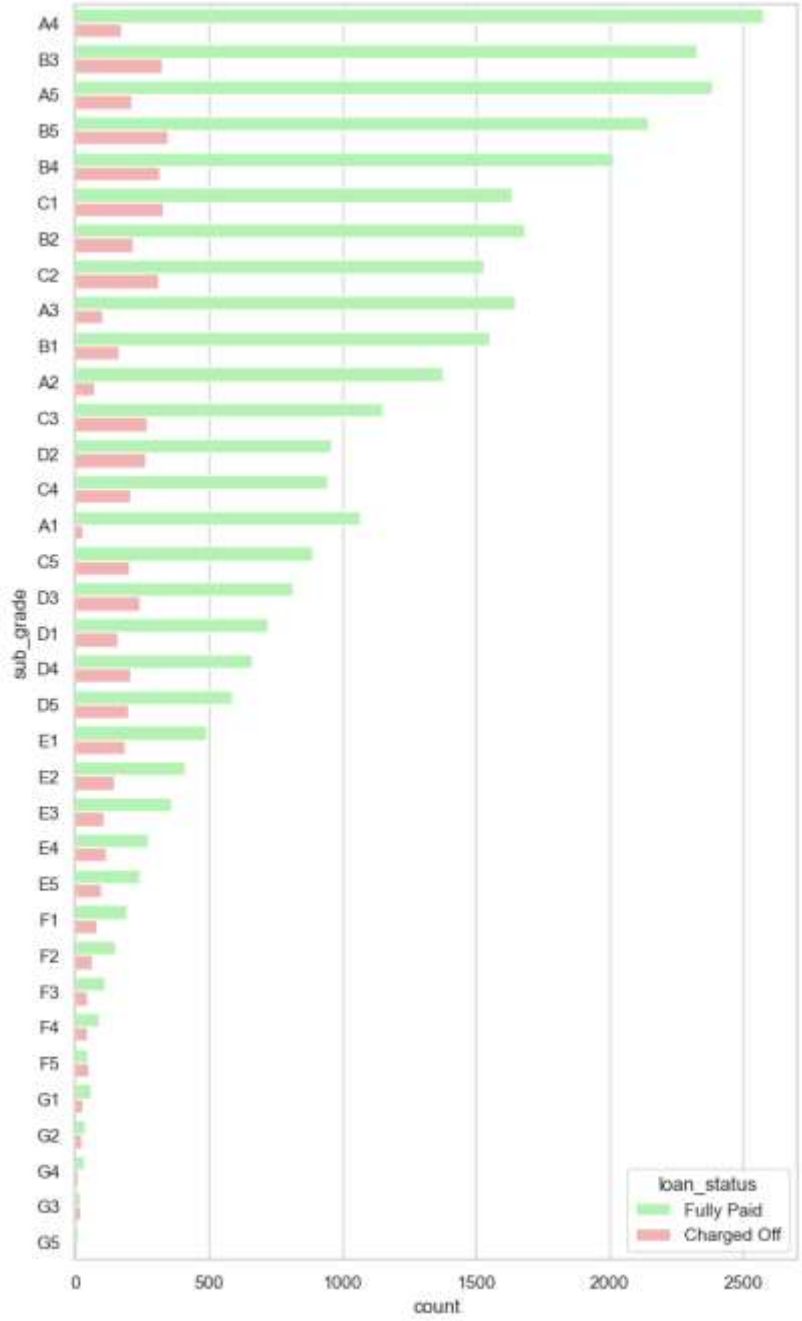
Grade Vs Charged off Proportion

We can clearly see that loan grades having highest default percentages. G, F, E and D form grades where default rate is much higher than others.



Subgrade Vs Charged off Proportion

The above table shows the loan sub-grade versus the default percentage. The G3 and F5 subcategories have above 40% default rate. This field is a clear indicator of the default percent.

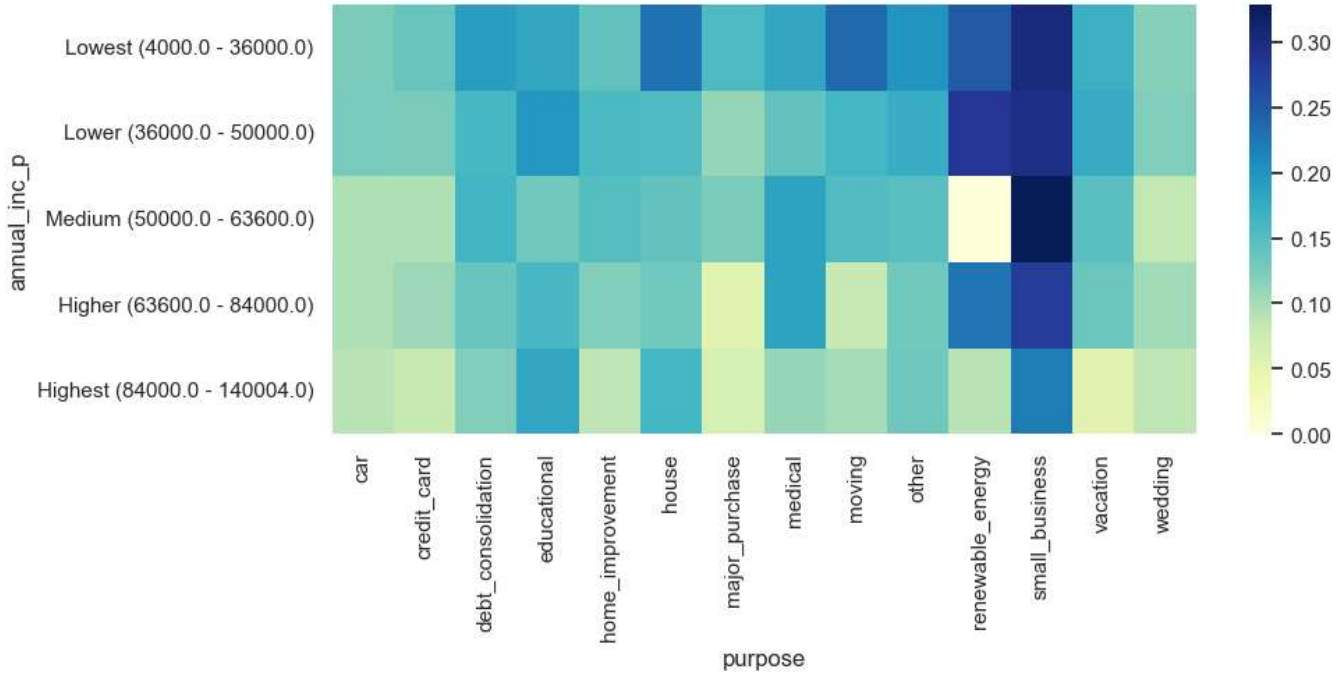


Bivariate Analysis

- Bivariate analysis examines the relationship between two variables.
- It involves analyzing two variables simultaneously to determine if there is a relationship or correlation between them.
- This analysis helps in understanding how the value of one variable changes when the value of another variable changes.

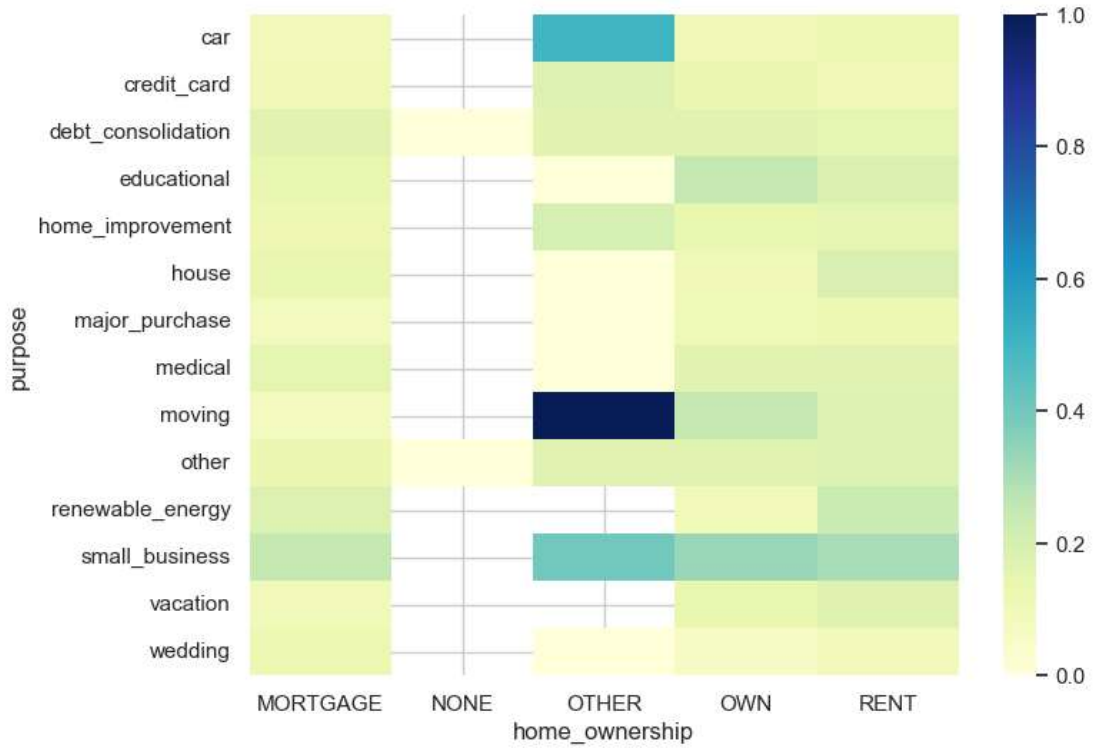
Annual Income Proportion

Plot of various income groups versus the risky purposes of loans for them. Some examples are, small business loans for lowest and medium income groups, renewable energy loans for higher income group



Purpose

Above figure shows that for higher installments for any income group have a greater number of defaults.



RECOMMENDATIONS

From above analysis we can say, that there is more probability of defaulting when

- Borrowers are taking loan for the term '60 months '.
- Borrower's whose loan status is 'Verified' as they taken high amount of loan with 60 months tenure.
- Borrower's who are having home ownership as 'Rent' and they take loan for the
  purpose of debt consolidation.
- Borrower's whose annual income is low i.e. (0-20000).
- Borrower's who takes loan amount in the range 0 to 14000.
- Borrower's who receive interest at the rate of 15-20%.
- Borrower who takes loan for the purpose of small business.
- Borrowers with lower Grade i.e. F<G.
- Borrower's whose subgrade is F5,G3,G5.

Thank You!

Presentation By Tejas Hankare &
Madhur Gupta