

Lending Club Case Study

EDA by

Sridhar G
Sudhindra S

Contents

- Business Objectives
- Data Overview
- Understanding the data
- Data Cleaning
- Data Analysis
 - Univariate Analysis
 - Bivariate Analysis
 - Multivariate Analysis/ Derived Metrics
- Recommendations

Business Objectives

- **Company Overview:**

- Largest online loan marketplace for personal, business, and medical loans.
- Offers fast online interface for easy access to lower interest rate loans.

- **Challenge: Credit Loss from Risky Applicants:**

- Significant financial loss occurs when borrowers default ('charged-off' customers).
- Lending to 'risky' applicants increases credit loss.

- **Aim: Identify Risky Loan Applicants Using EDA:**

- Utilize Exploratory Data Analysis to detect borrowers likely to default.
- Reduce risky loans to minimize credit loss.

- **Objective: Understand Drivers of Loan Default:**

- Determine key variables that strongly indicate default.
- Use insights for improved portfolio and risk assessment.

Data Overview

- Loan Data is provided by Lending Club
- Data is mostly for the cities in US
- Matrix of data is around 111 columns and 39717 rows
- Loan details and few factors of loan finance is given
- Data dictionary comprising the explanation of each column is given

Understanding the Data

- Loan Status(`loan_status`) is the column that is considered as target
- Loan Status is classified into 3 main categories
 - Charged Off : Indicates the loan applicant is unable to repay his loan, considered as Defaulter
 - Fully Paid: Indicates the loan applicant has successfully repayed the loan along with interest and principal
 - Current: Indicates loan applicant is still pursuing his repayment of the loan

Understanding the Data

- Customer attributes
 - Employment year (emp_length)
 - State (addr_state)
 - Annual Income (annual_inc)
 - Home ownership (home_ownership)
 - Debt to Income (dti)
 - The month the borrower's earliest reported credit line was opened (earliest_cr_line)
- Loan attributes
 - Loan amount (loan_amnt)
 - Loan date (issue_d)
 - Grade (grade)
 - Loan Term (term)
 - Interest rate (int_rate)
 - Loan purpose (purpose)
 - Public record bankruptcies (pub_rec_bankruptcies)
 - Loan Amount sanctioned (funded_amnt)

Data Cleaning

- Modifying values
 - Removing 'months' from column *term*
 - % removed from the interest rate column *int_rate*
- Modifying emp_length column by removing string years and converting 10+ to 10 and converting < 1 to 0
- Deriving columns next_pymnt_year, emp_length_normalized, issue_month, issue_year, loan_amnt_bin, earliest_cr_line_year from various columns
- Removing outliers
 - home_ownership has NONE as 3 rows this can be removed , outlier

Data cleaning

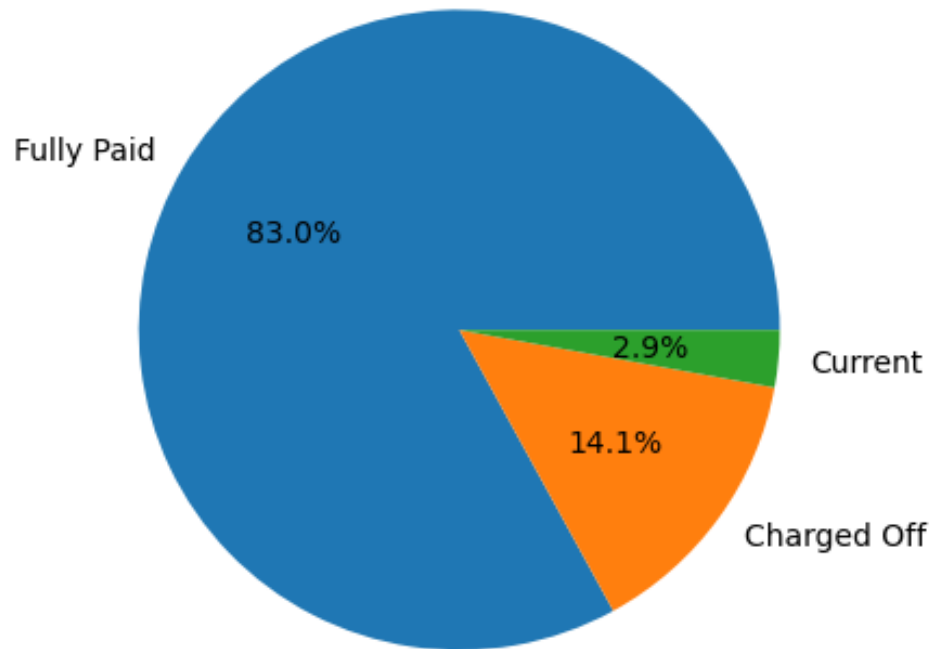
- Categorizing the data
 - Loan amount (loan_amnt) can be classified into 'Very Low', 'Low', 'Medium', 'High', 'Very High', 'Super High']
 - Annual income(annual_inc) can be classified into 'Aspirers', 'Middle class', 'Upper middle class' and 'Rich'
 - Loan interest rate (int_rate) can be classified into 'Very Low', 'Low', 'Medium', 'High', 'Very High' and 'Super High'

Analysis of Data

- Categorical = grade, emp_length, home_ownership, verification_status, loan_status, purpose, addr_state
 - Also some of the columns has be divided into its respective bin sized for analysis
- Numerical = loan_amnt, funded_amnt, funded_amnt_inv , term, int_rate , installment, annual_inc, dti
- Date/Time = earliest_cr_line , earliest_cr_line_year
- Target Variable = loan_status
- Funded amount is having marginal difference on loan amount hence loan amount is considered

Current status of the Bank w.r.t Loan Payment Status

Payment Status Type

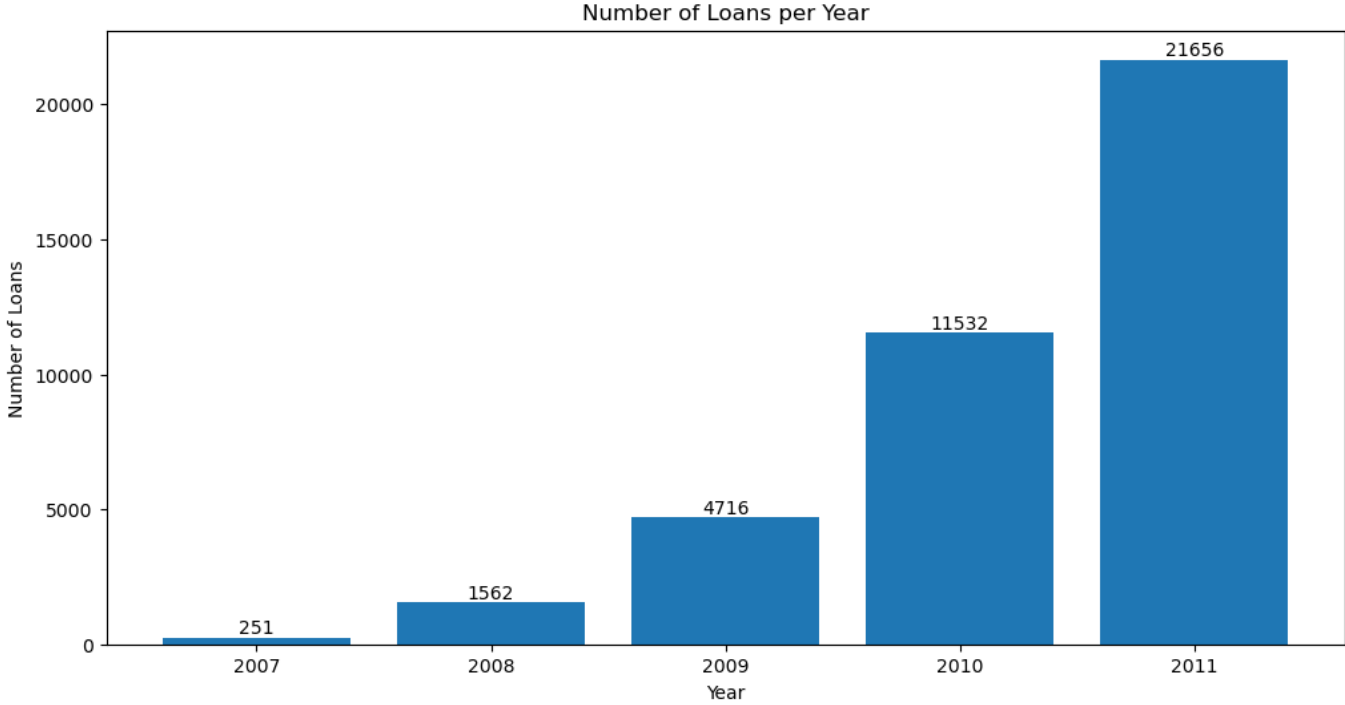


Currently this is where Bank is standing w.r.t to its loans

1. 83% members have paid their loan completely
2. 14.1 % members are defaulters , these members are not reachable for payments, or in another words they have run away
3. 2.9% members are still paying their loan
4. Each member is having only 1 loan not multiple loans*
5. Chart shown here represented correctly w.r.t each category

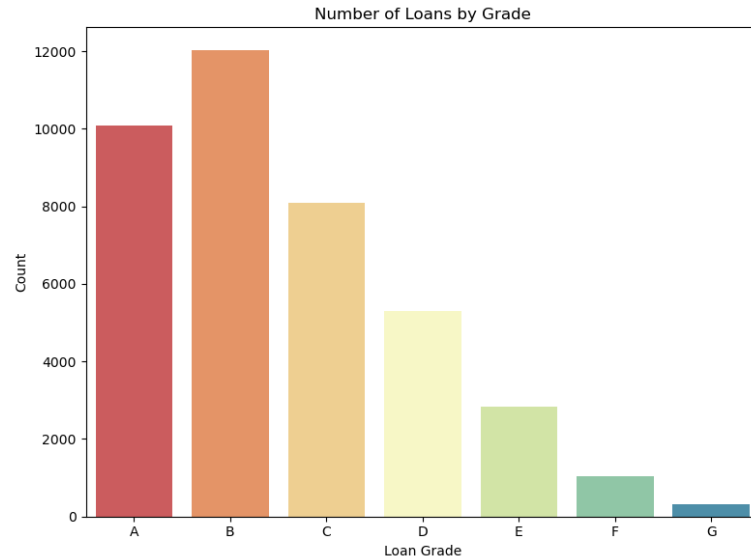
* Its worked upon using dataframes

Loan applications over the Years

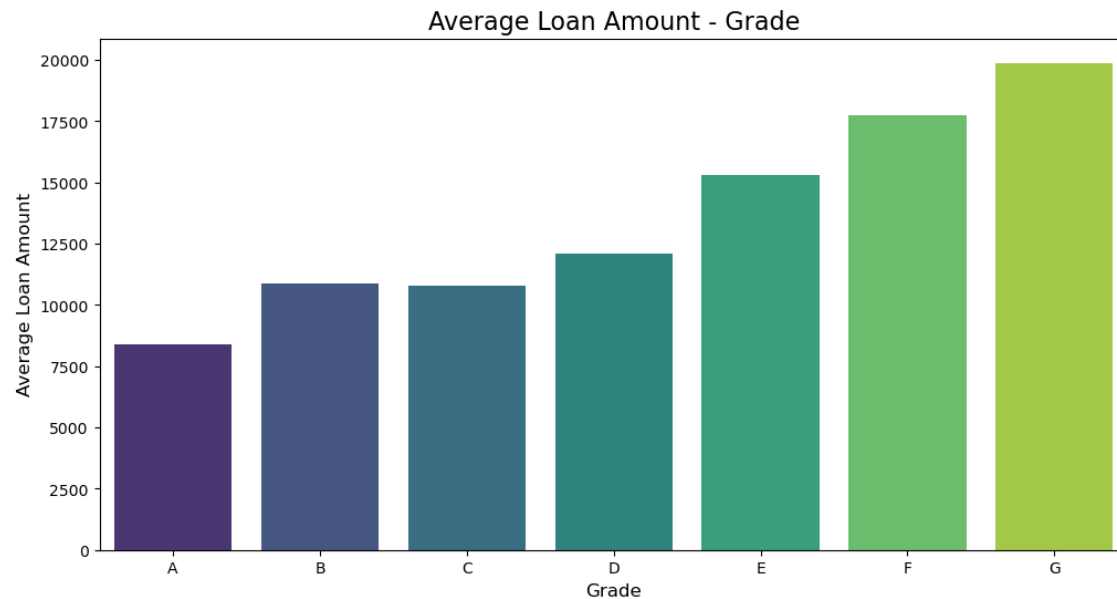


Number of loans issued is exponential in year 2010 and 2011

Number of Grade counts

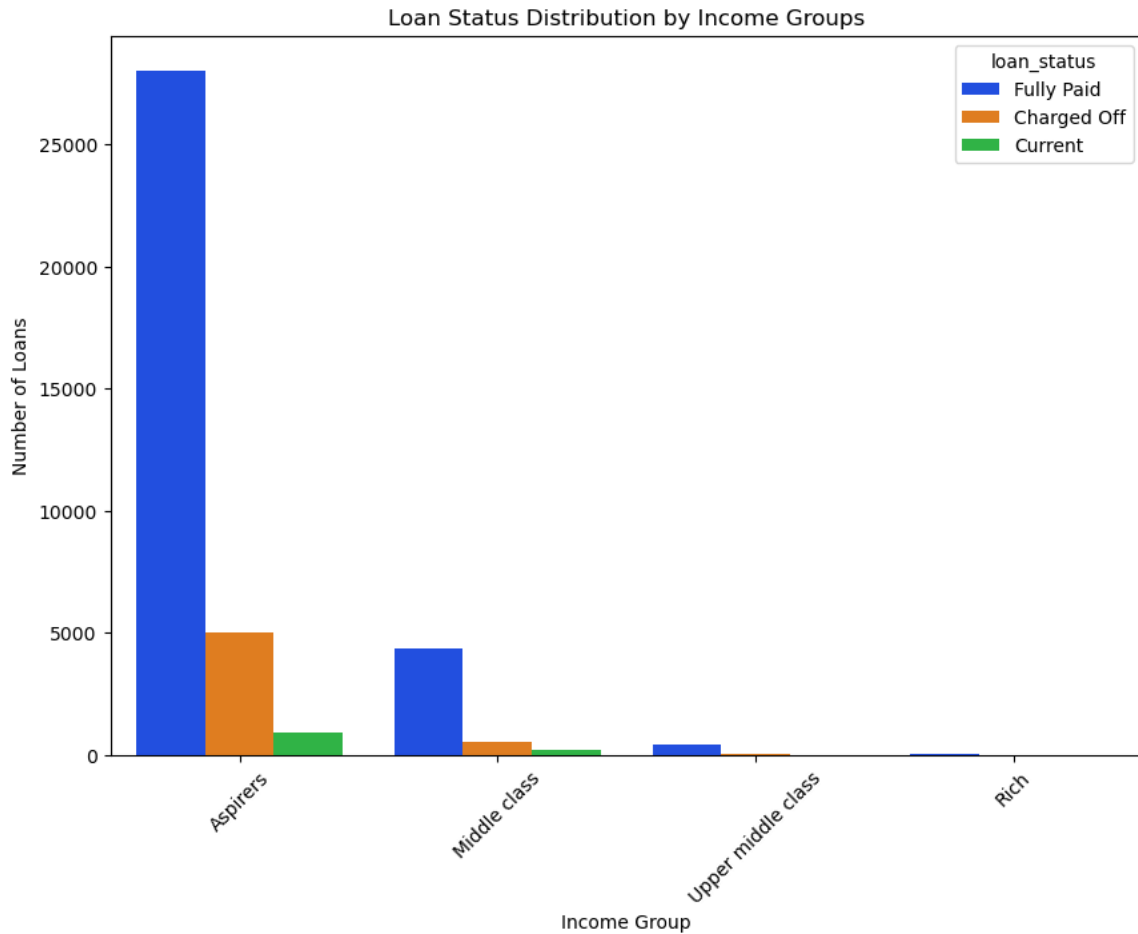


Loan Grades are mostly in Grade 'B'



A grade loans average loan amount is the least of all grades. If grade is higher, lower the loan amount is issued

Income segment of Defaulters



Loan Status w.r.t Income group of Defaulters

The defaulters mostly are the group of low income or Aspirers

Income brackets across each category

'Aspirers' : \$0 - \$100000

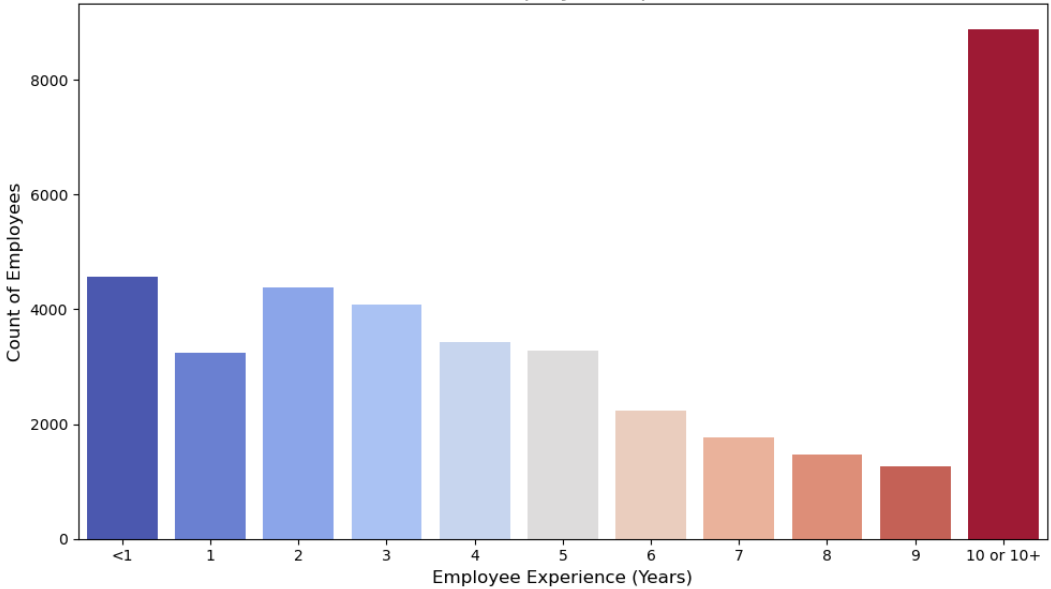
'Middle class': \$100000 - \$200000

'Upper middle class': \$200000 - \$500000

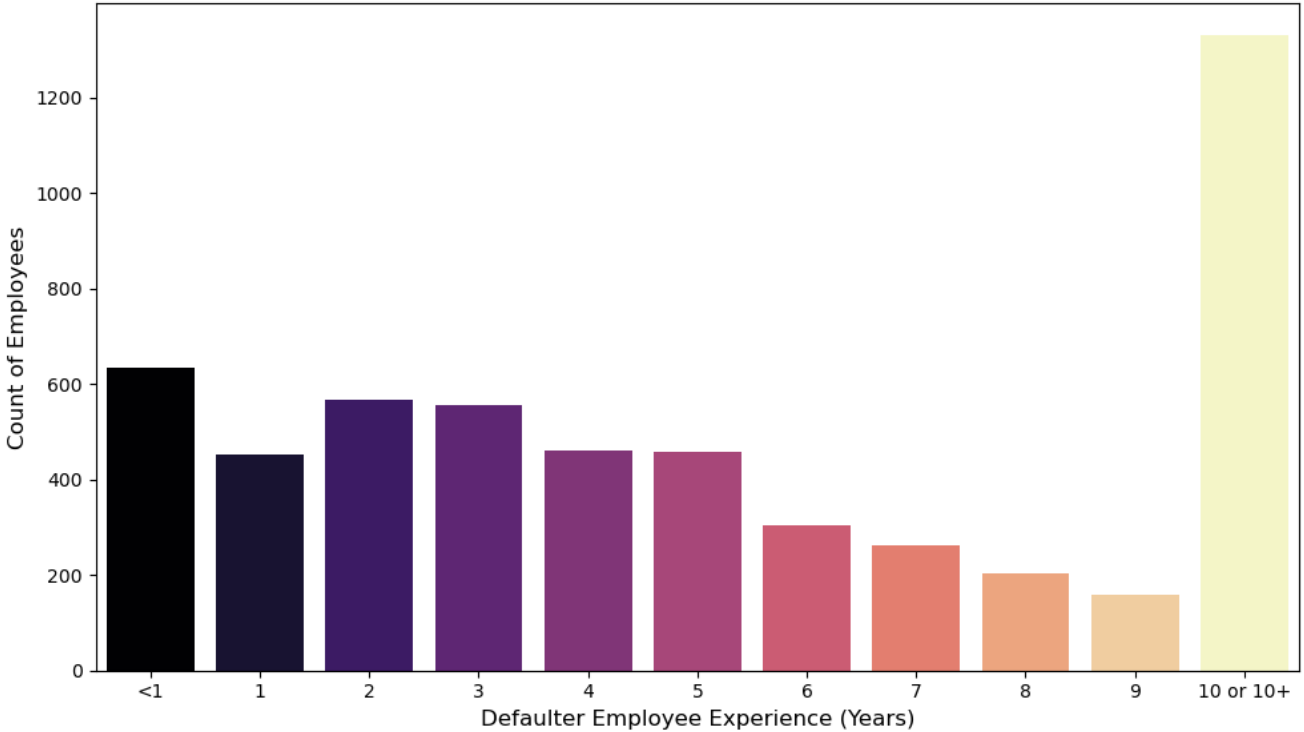
'Rich': \$500000+

Experience level of employee for everyone and Defaulters

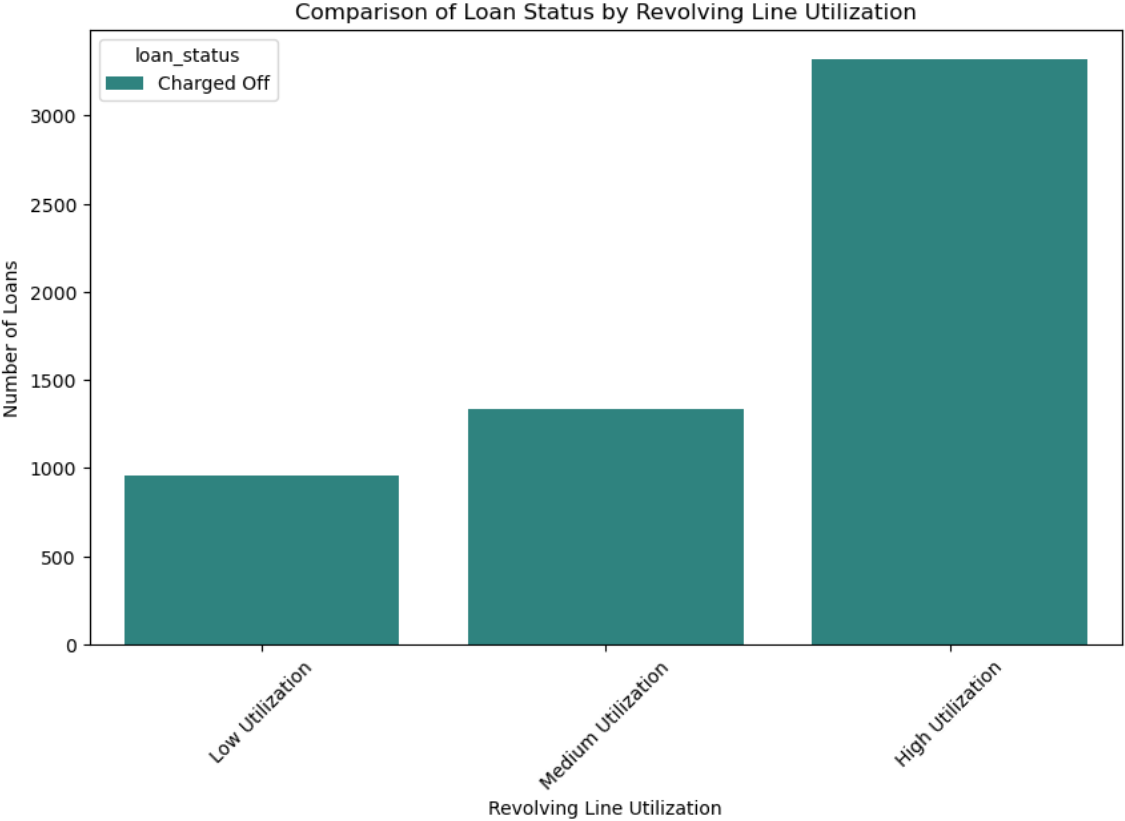
Count of Each Employee Experience Level



Count of Each Defaulter Employee Experience Level



Revolving line utilization of Defaulters



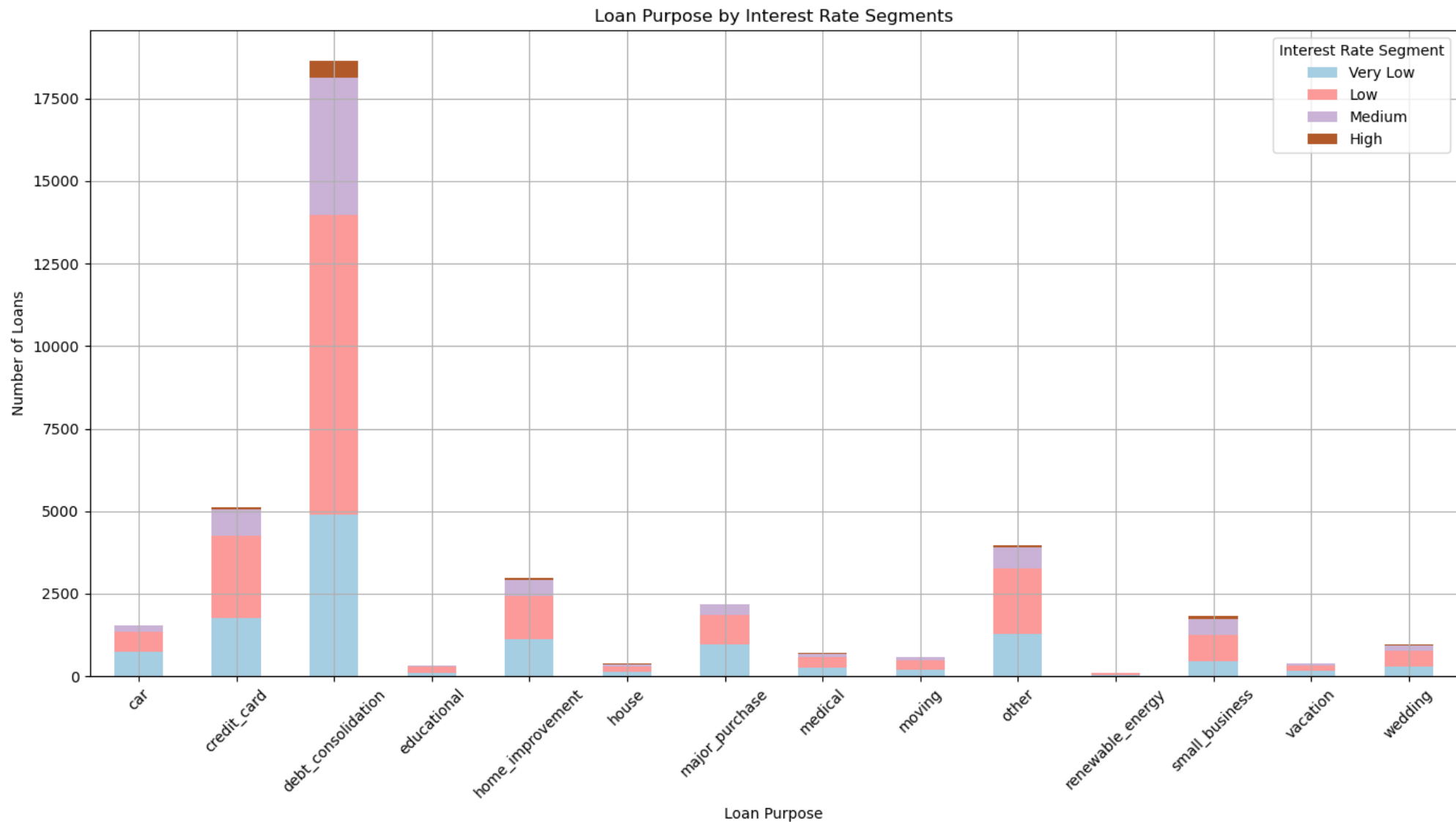
Low Revolving Line Utilization rate should be less
Lower the rate better for repayment

As seen, higher the rate higher is the defaulters/Charged off

[https://www.nationalfunding.com/blog/what-is-revolving-utilization/#:~:text=All%20major%20credit%20bureaus%20\(Equifax,credit%20card%20balance%20every%20month.](https://www.nationalfunding.com/blog/what-is-revolving-utilization/#:~:text=All%20major%20credit%20bureaus%20(Equifax,credit%20card%20balance%20every%20month.)

Utilization lower the rate, the better

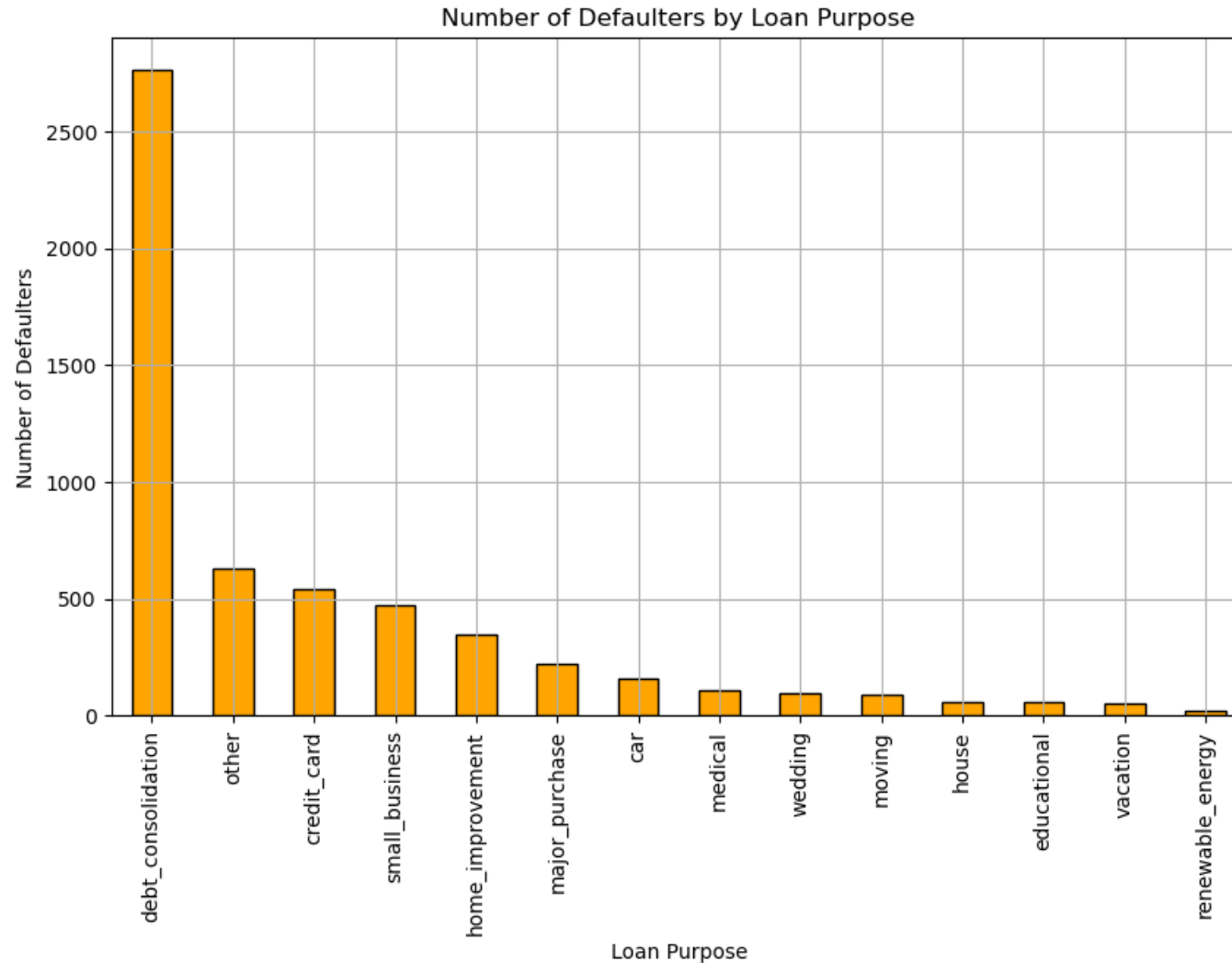
Current status of the Bank w.r.t Loan interest rate on loan purpose in general



Observation

- Currently bank is keeping the loan rate high on the category Debt consolidation loans
- Also Debt consolidation loan numbers are high in nature
- Number of loans towards debt consolidation bank should be cautious
- Also next in the segment is Credit card purpose loans.
- Top 2 are Debt consolidation and Credit card loans; these should always be checked
- Bank should further increase the interest rate on this Debt consolidation category for the members who have tendency to be defaulters

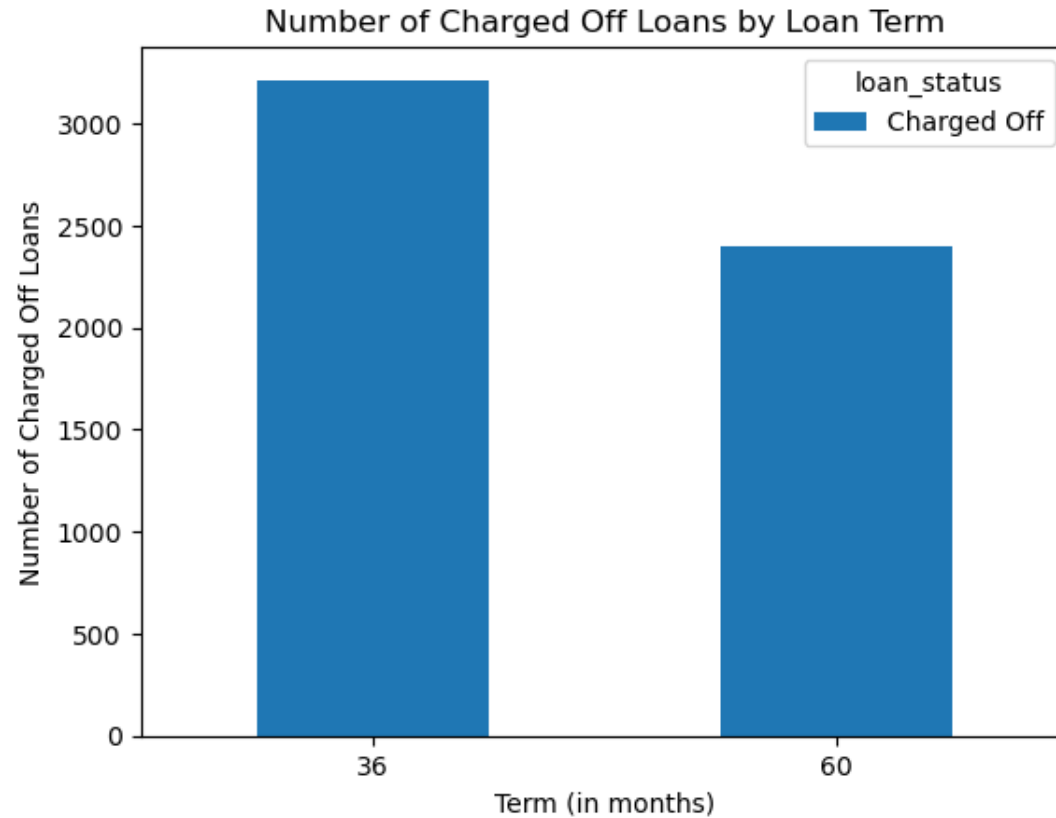
Current status of the Bank w.r.t to loan purpose for defaulters



Observation

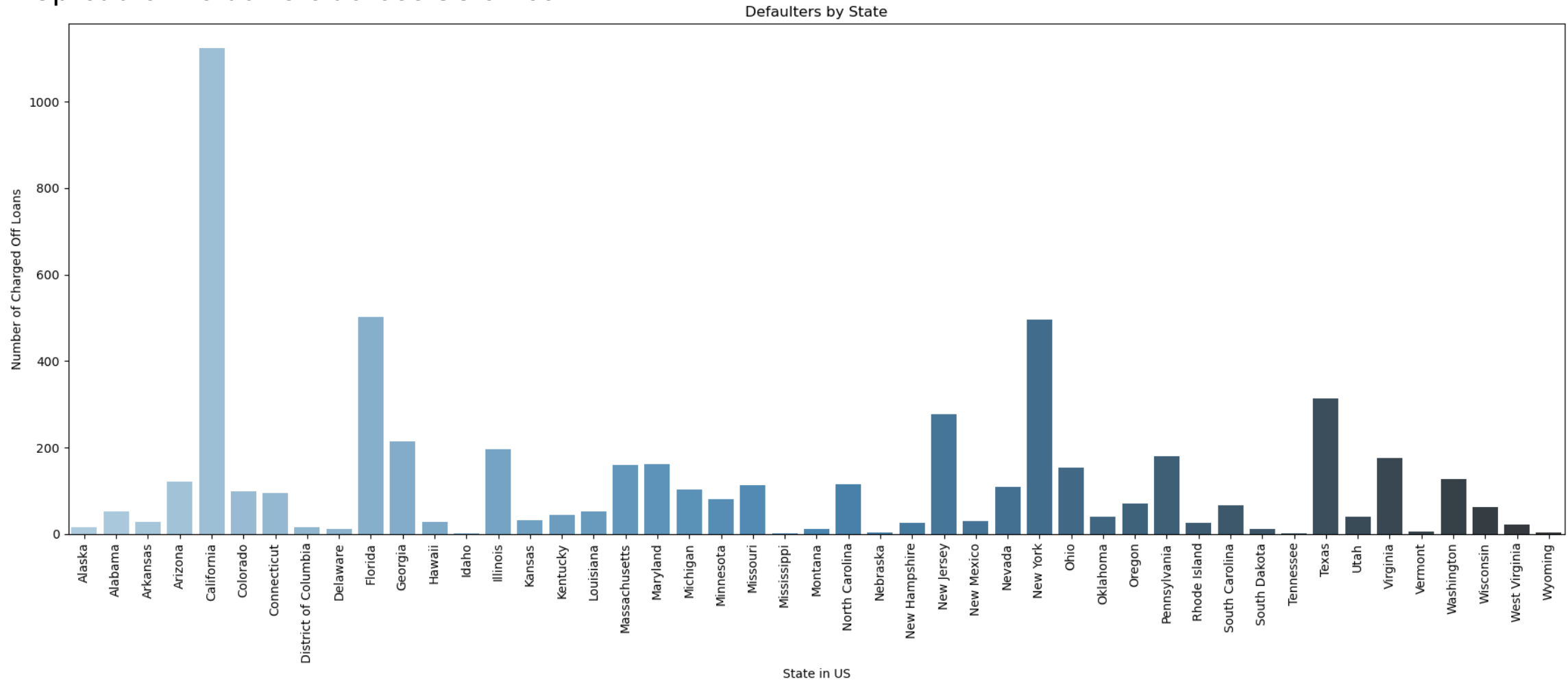
- The loan purpose with the maximum number of defaulters is 'debt_consolidation' with 2762 defaulters.
- Also an observation , 'Other' category should be clearly defined. Bank should categorized the loan type
- In general for any loan w.r.t Debt consolidation should be cautious
- Also next in the segment is Credit card purpose loans after 'Others'.
- Top 3 are Debt consolidation , Others and Credit card loans; these should always be checked

Defaulters on various categories



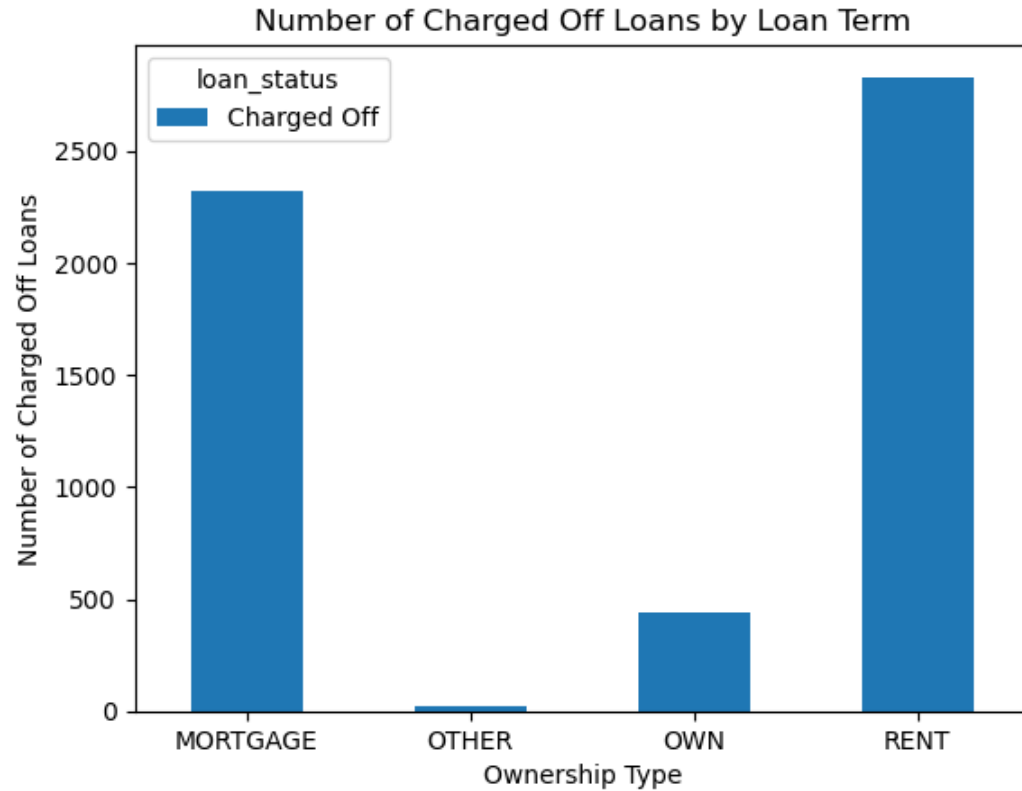
Number of defaulter have tendency who take 36 months than 60 months term

Spread of Defaulters across US cities



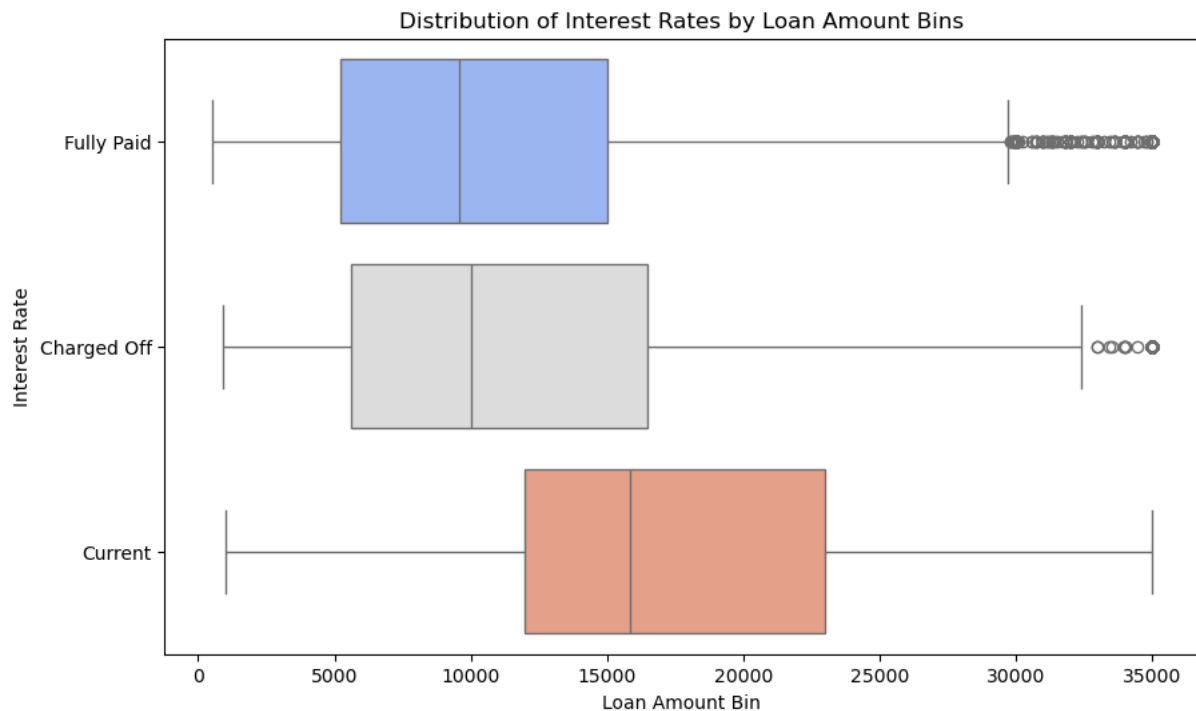
California state has more number of defaulters when compared to other states. Likewise we can see the defaulters are more in other major cities like New York and Florida.

Ownership type of Defaulters



Defaulters are usually high for the ownership type who are renting a house or done a mortgage

Distribution of Interest Rates by Loan Amount



Median Interest Rates:

- Fully Paid loans have the lowest median interest rate
- Charged Off loans have a higher median interest rate than Fully Paid loans
- Current loans have the highest median interest rate

Loan Amounts:

- The majority of loans across all statuses fall between about 5,000 and 20,000
- There are some outliers in the Fully Paid category for larger loan amounts (around 30,000-35,000)

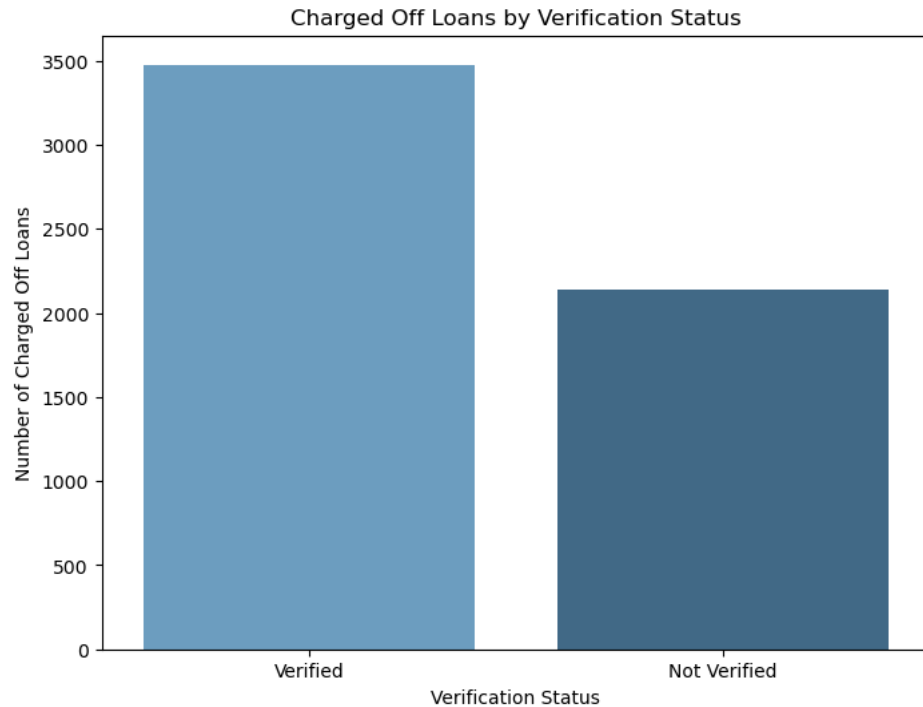
Risk Assessment:

- The higher interest rates for Charged Off loans suggest that riskier borrowers were charged higher rates
- Current loans having the highest rates might indicate recent trends in interest rate increases or that newer loans are considered higher risk

Outliers:

- Fully Paid loans have several high-amount outliers with relatively low interest rates, possibly indicating good credit borrowers taking larger loans
- Charged Off loans have a few outliers with very high interest rates, potentially representing very high-risk loans that ended up defaulting

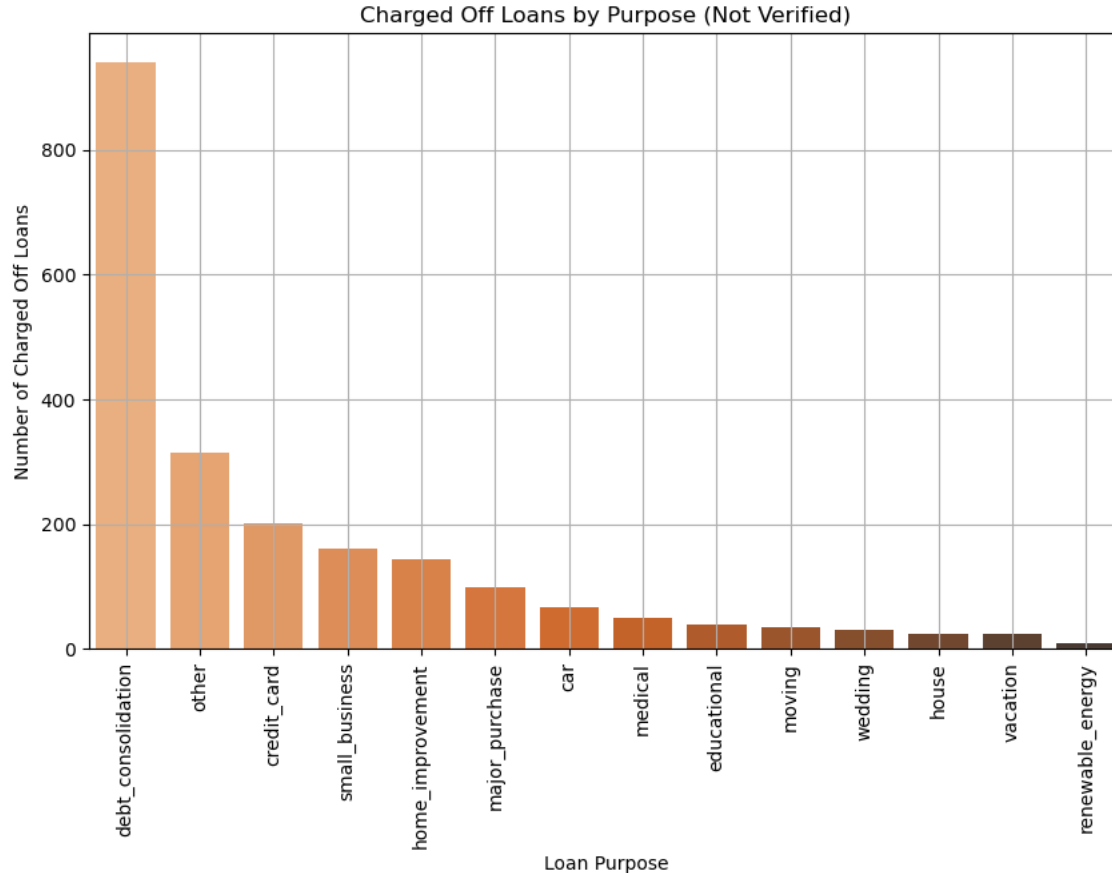
Verification Status on Defaulters



- It's a strange behavior that number of defaulters Income aren't verified.
- This is one of the prime reason they turned to be Defaulters
- Numbers are more than 2000

This needs a urgent attention from the bank

Verification Status on Defaulters based on purpose

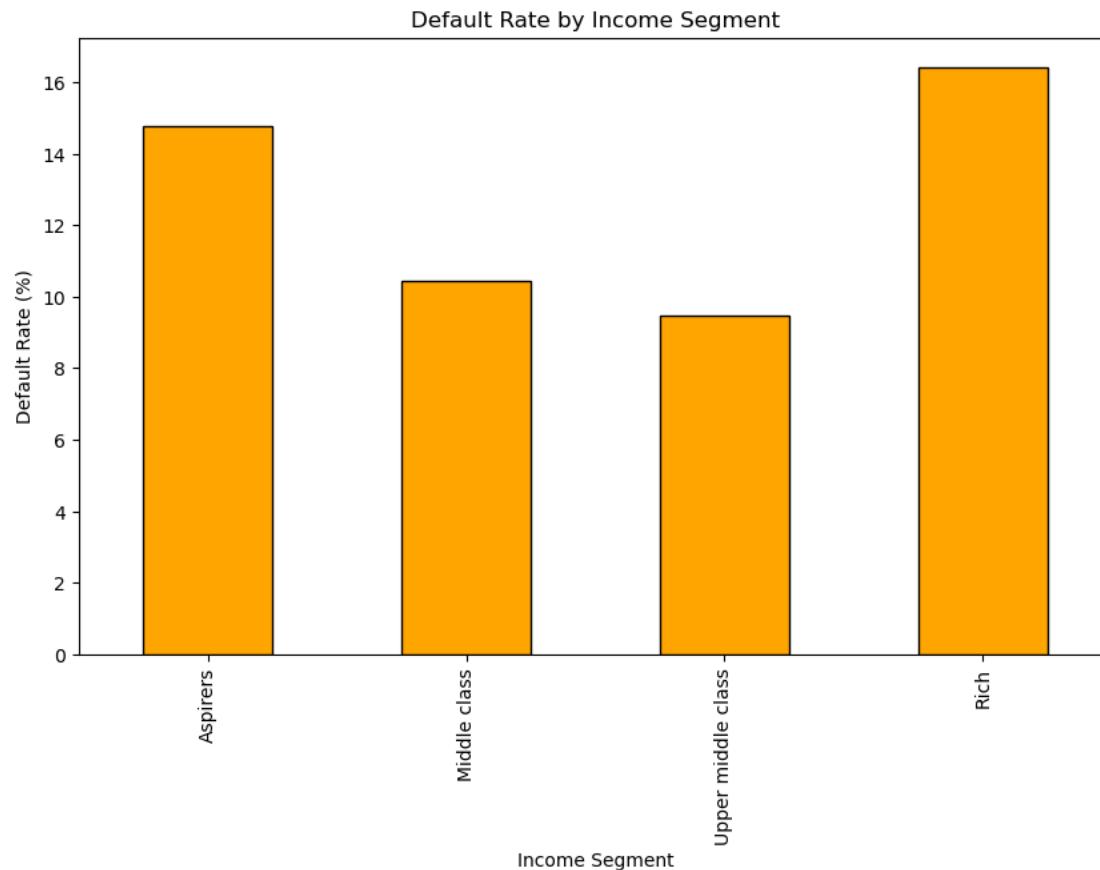


Continuing on analysis on Purpose, here mostly members who do close multiple loans by taking Debt Consolidation loan haven't got their income verified by the bank. Again Debt consolidation category is highest here for all the defaulters

Bank should think on increasing interest rate on the loans given for Debt consolidation "With income verification". The number of Debt consolidation is also very high when compared to other categories. Some issue in loan dept who are granting Debt consolidation as

- Income source of the member is not verified for this category
- Number of unverified debt consolidation loans are high for this category

Income segment of Defaulters



Now if the income is verified

1. Rich people have tendency to become a defaulter
2. Aspirers also have tendency to become a defaulter
3. Upper Middle are better in repaying the loans when compared to other categories

Bank has to keep eye on Rich and Aspirers class of income sectors while giving loan

Income brackets across each category

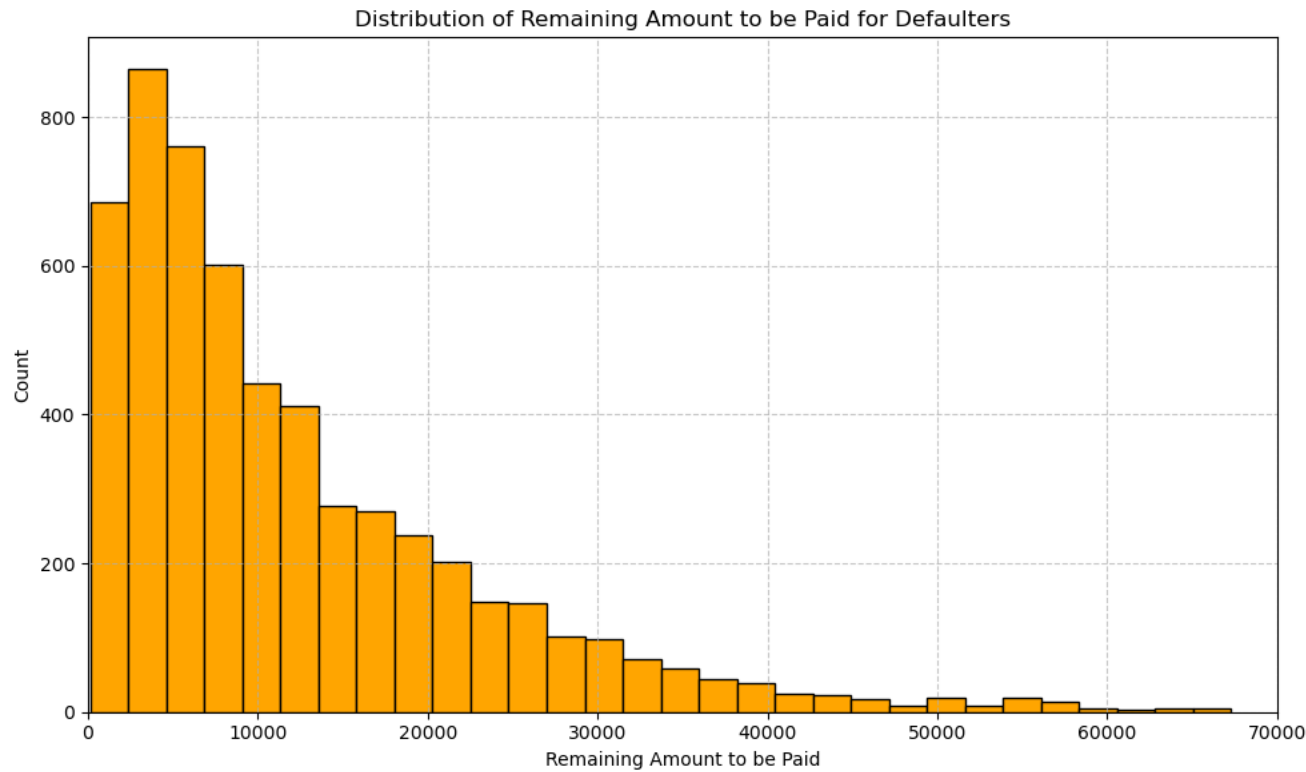
'Aspirers' : \$ 0 - \$100000

'Middle class': \$100000 - \$200000

'Upper middle class': \$200000 - \$500000

'Rich': \$500000+

Amount to be recovered from defaulters



Amount to be paid by defaulters less than or equal to 10000 are more in number

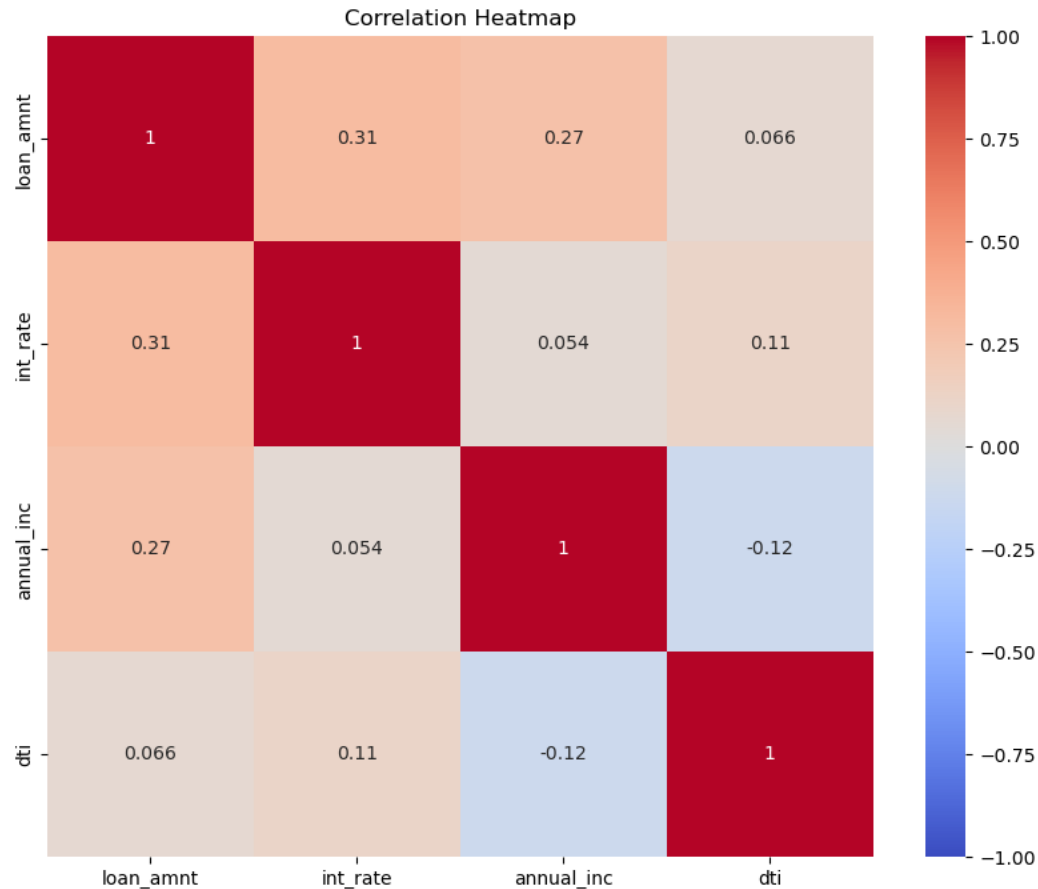
There is 64% loss in case of defaulters amount recovery

Total amount funded for defaulters is \$106,895,884

Total amount to be recovered from defaulters is \$68,457,204

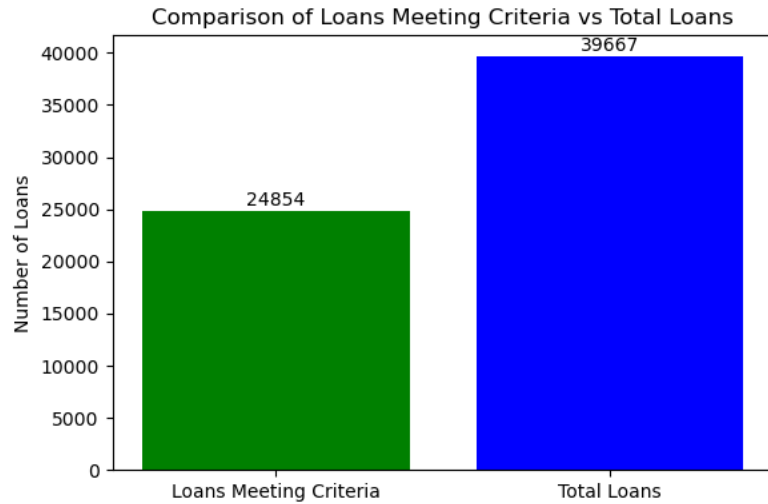
Total loss is 64%

Co-relation map

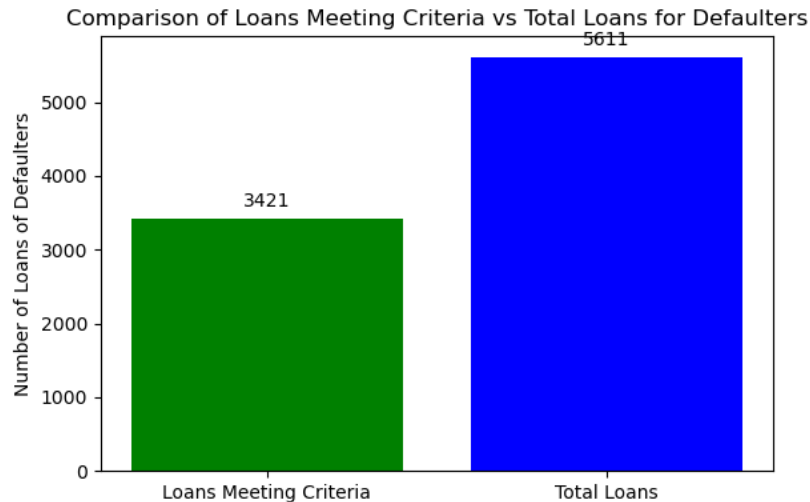


1. **Loan Amount and Interest Rate (0.31 correlation):**
There's a moderate positive correlation between the loan amount and interest rate. This suggests that as loan amounts increase, interest rates tend to increase slightly as well. This could indicate that lenders charge higher rates for larger loans, possibly due to increased risk.
2. **Loan Amount and Annual Income (0.27 correlation):**
The positive correlation here implies that borrowers with higher annual incomes tend to take out larger loans. This makes sense, as people with higher incomes may qualify for and be able to afford larger loan amounts.
3. **Interest Rate and Annual Income (0.054 correlation):**
There's a very weak positive correlation, suggesting that annual income doesn't have a strong direct influence on the interest rate offered. Other factors likely play a more significant role in determining interest rates.
4. **Interest Rate and DTI (Debt-to-Income Ratio) (0.11 correlation):**
The weak positive correlation indicates that as a borrower's debt-to-income ratio increases, there's a slight tendency for the interest rate to increase. This aligns with lending practices, as higher DTI ratios may be seen as riskier, potentially resulting in slightly higher interest rates.

Loan criteria based on DTI - Observation



Number of members who match Loan criteria are 24,854 and total given is 39,667



Number of members who match 3421 and total given is 5611

Surprisingly DTI greater 36% and older credit line were Defaulters we need to consider other factors as well