LENDING CLUB
CASE STUDY – ELANGATHIR G

- Objective
- Deciphering the Data
- Data Cleansing
- Data Analysis
- Recommendations based on analysis
- Summary

OBJECTIVE

- The objective is to analyze historical data collected on over 100 parameters to derive observations, insights, findings, correlations, and actionable recommendations for Lending Club. The goal is to enhance the loan approval process, ensuring that prospective loan accounts are assessed in a way that minimizes default risk, improves portfolio performance, and optimizes decision-making for better financial outcomes.

DECIPHERING THE DATA

- Deciphering the historical data involves a analysis of various borrower attributes, loan characteristics, and payment behaviors across a large dataset. By examining over 100 parameters, such as loan amount, interest rates, employment history, income levels, credit utilization, and debt-to-income ratios, we target uncovering patterns and trends that influence loan default risk.

DATA CLEANSING

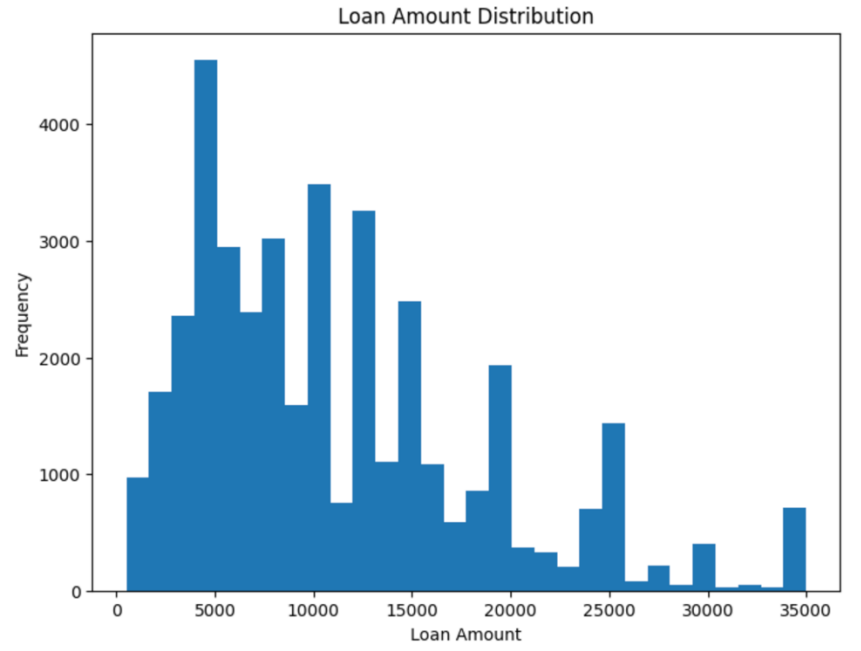
- Data cleansing and handling involve identifying and addressing issues like missing values, outliers, and inconsistencies in the dataset to ensure accurate and reliable analysis. This process typically includes:

1. **Handling Missing Data**: Imputing missing values through methods such as mean/median imputation for numerical columns or mode imputation for categorical columns were discussed. In most cases, rows with excessive missing data are removed.
2. **Outlier Detection**: Identifying and handling outliers using statistical methods to prevent them from distorting analysis results.
3. **Data Transformation**: Converting categorical variables to numerical forms and normalizing or scaling
4. **Dealing with Duplicates**: Removing duplicate entries to maintain the integrity of the dataset.

This thorough approach ensured that the dataset is clean, accurate, and ready for meaningful analysis, minimizing the risk of erroneous insights or conclusions.

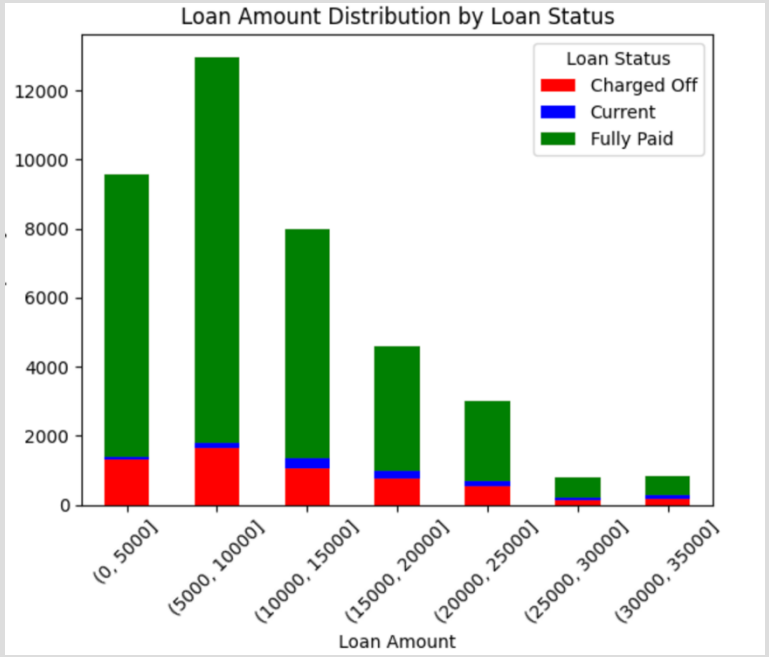
DATA ANALYSIS

- Exploratory Data Analysis (EDA) was conducted using the following approaches:
- **Segmented Univariate Analysis**: Breaking down individual variables based on key segments or groups (e.g., loan status, income brackets) to identify patterns or differences that might not be apparent in the overall data.
- **Bivariate Analysis**: Analyzing the relationships between pairs of variables to uncover correlations, dependencies, and trends. This includes using scatter plots, correlation matrices, and other relevant visualizations to examine interactions between key variables.
- Recommendations was derived based on the insights gathered from the above analyses.



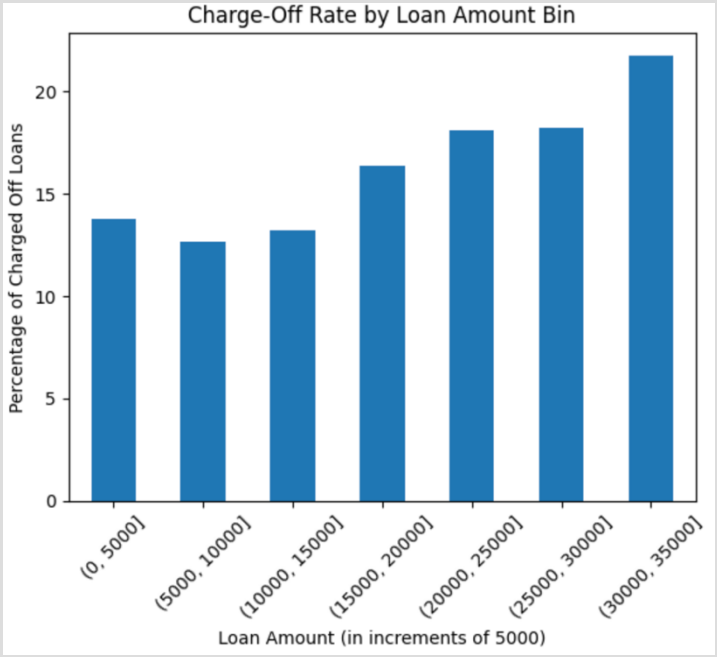
Loan Amount Frequency

- The frequency of the Loan amount ticket from low to high is plotted
- We can directly say the maximum loan amount/per ticket is in the range of $5,000



Loan Amount Distribution

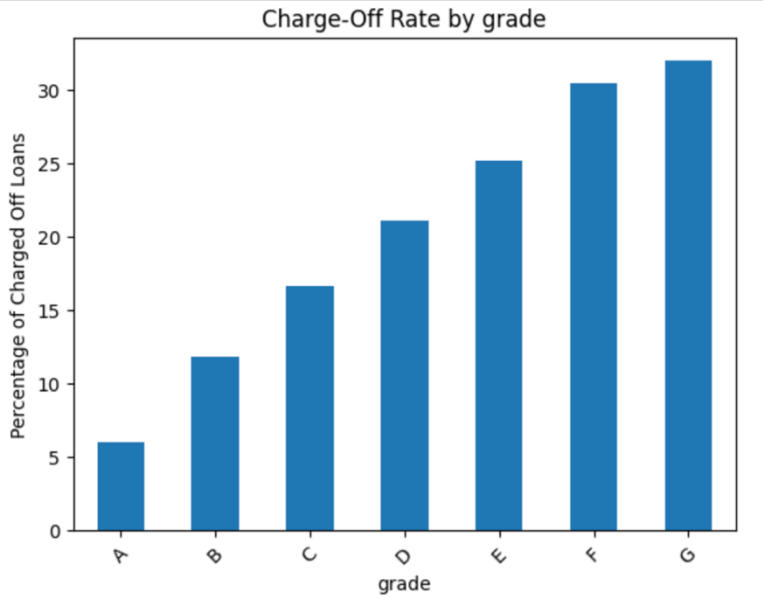
- by Status



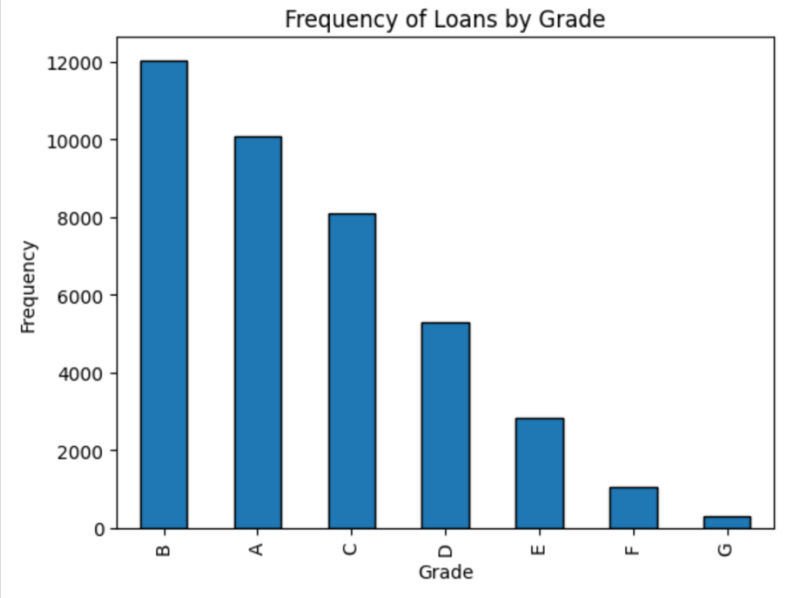
Loan Default rate

Vs Loan Amount

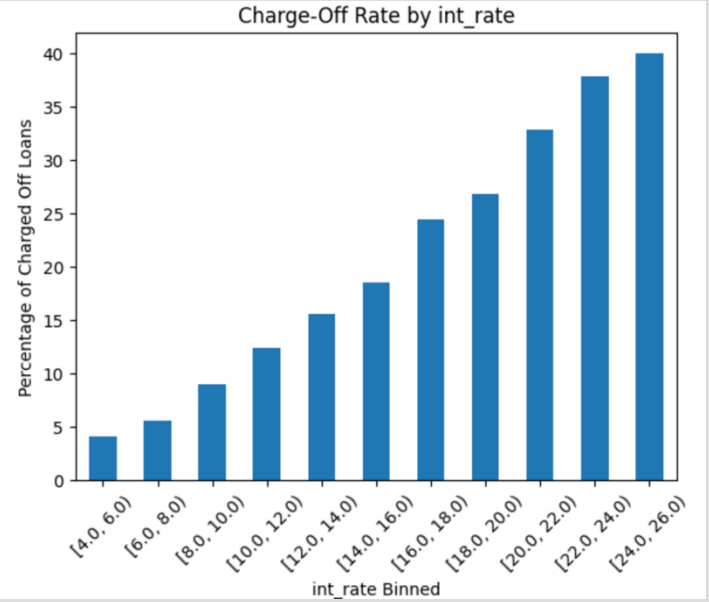
- The maximum loan amount/per ticket is in the range of $5,000 - $10,000
- Also the default in general increases proportional to the increase in Loan amount
- Not strictly linear also not very significant to provide any recommendations



Default vs Grade Profile of customers

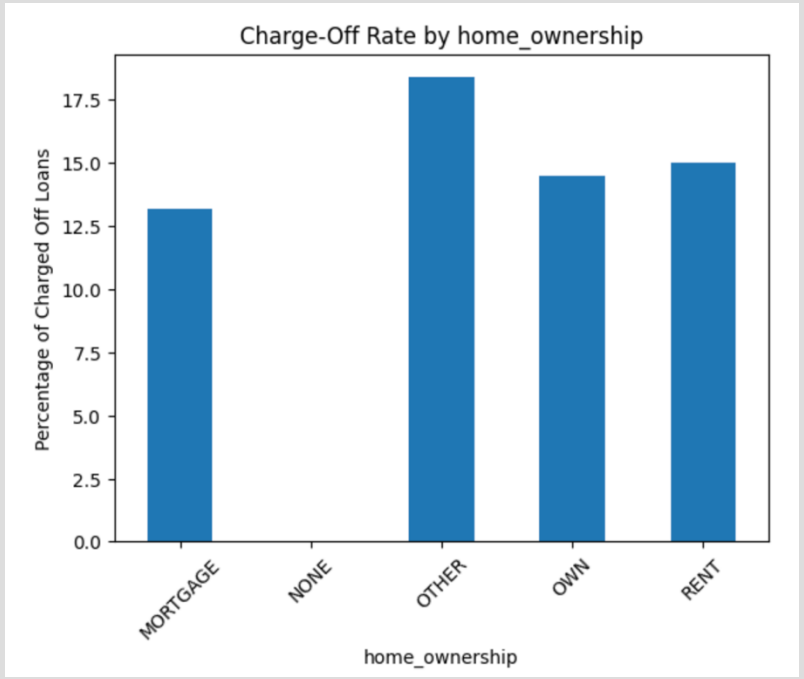


Loan Disbursal against Grade profile

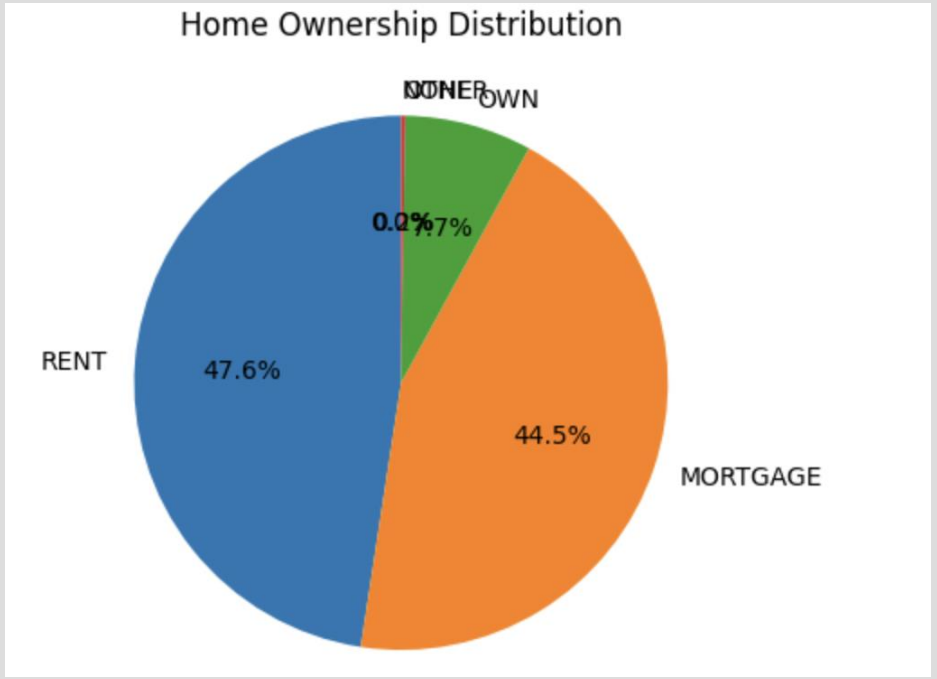


Loan Default vs Interest rate

- Strict linear corelation exists between Grade Profile and Loan default rate
- Also there is similar corelation exists between Grade Profile and Loan disbursal
- We also found (not captured here) direct corelation between Grade and Interest rate
- So the risk is factored in by charging higher interest for lower Grade profiles
- Already the co-relation is taken into account and no recommendation required

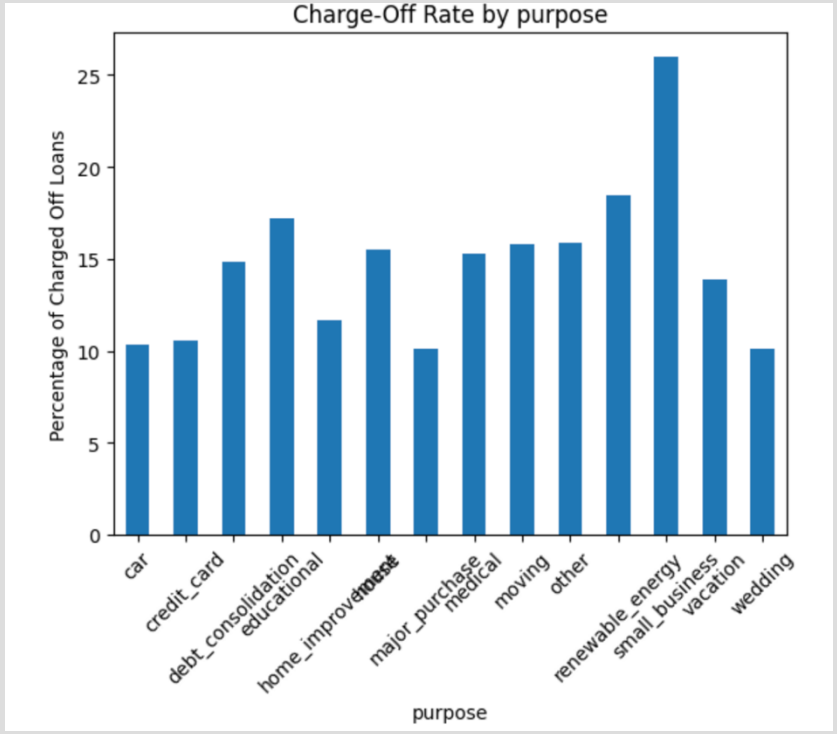


Default vs Home Ownership

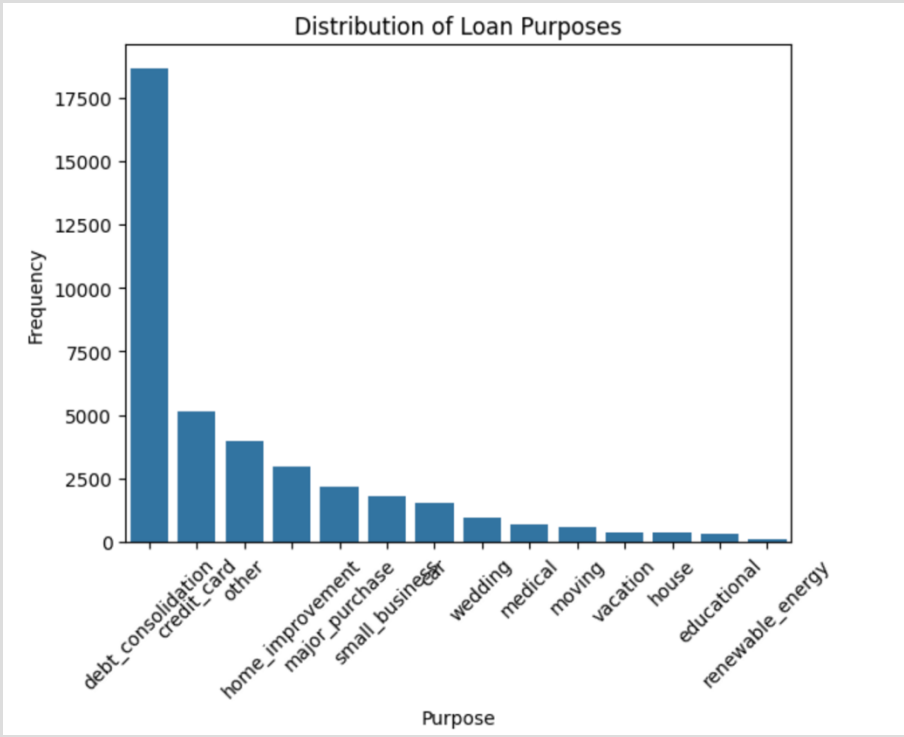


Loan Disbursal against
Home Ownership

- Mortgage, Owning and Renting Home have similar default rates (slightly increasing)
- Other category of Home Ownership has elevated default rates
- **We can recommend to study further on the Other category of Home Ownership**

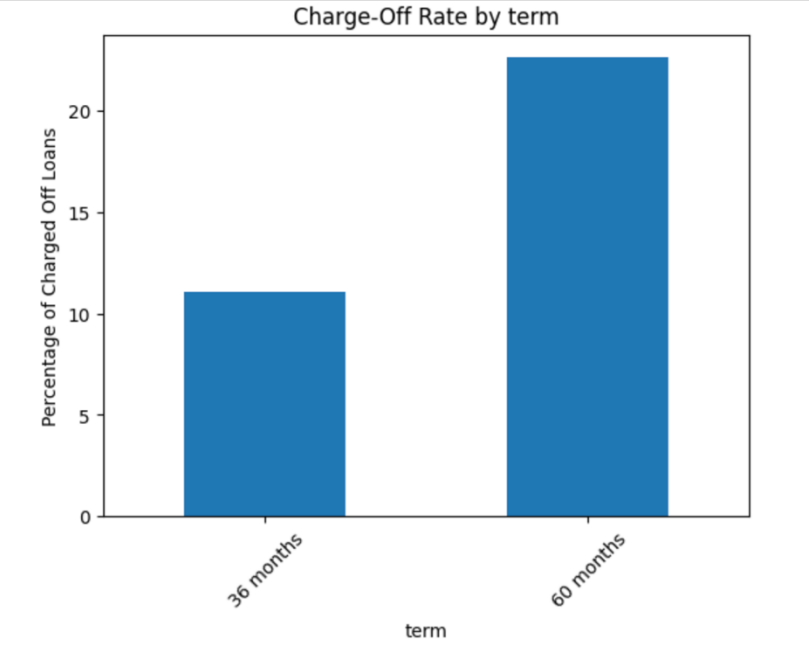


Default vs Loan Purpose



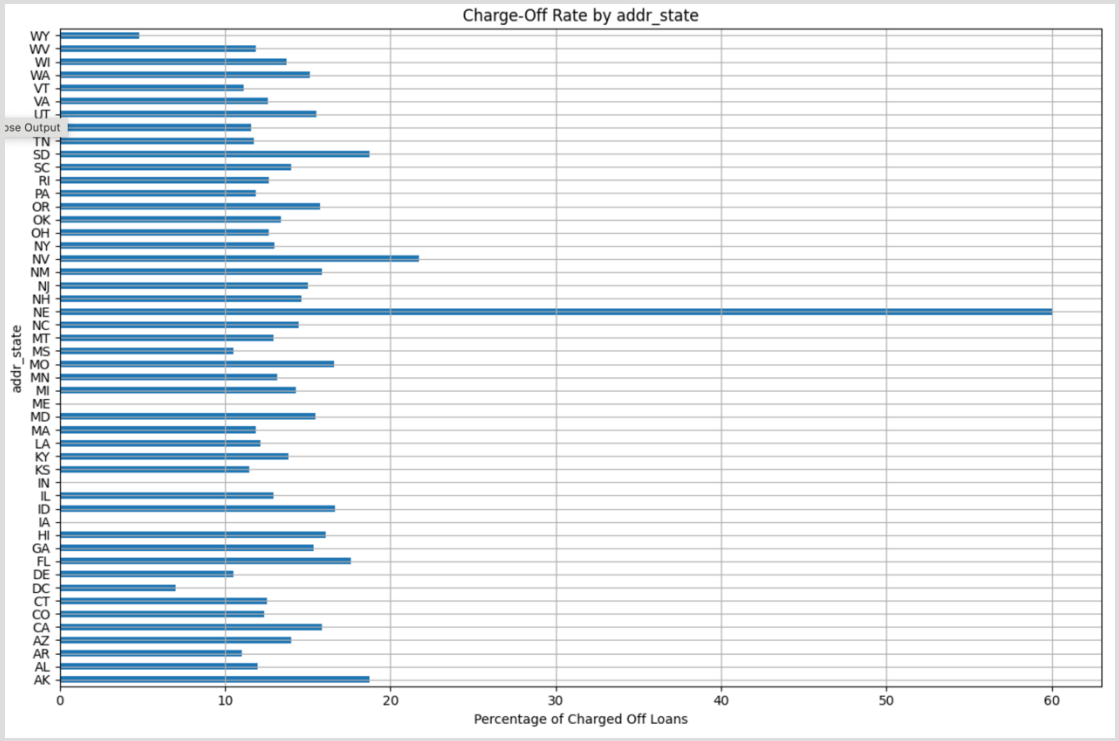
Loan Disbursal against Loan Purpose

- **Small business, Educational and Renewal Energy are riskier purposes for Loan disbursal**
- **We can recommend to increase loan disbursal for wedding and home improvement**



Default vs Repayment Terms

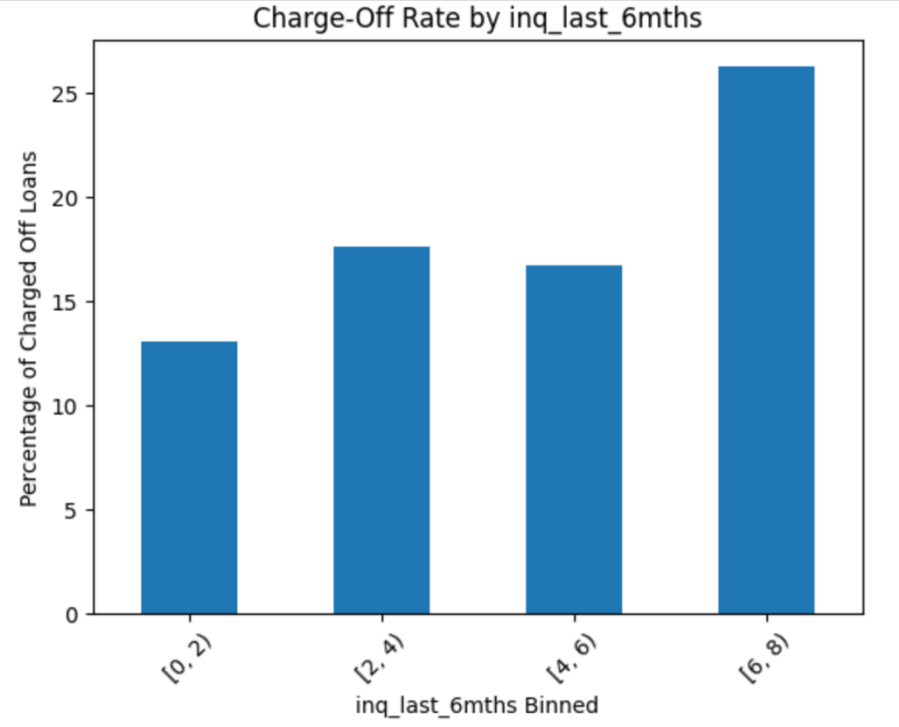
- Strong relation between longer repayment term and defaulting
- **We can strongly observe higher default risk associated with longer repayment duration**



Default vs States in the US

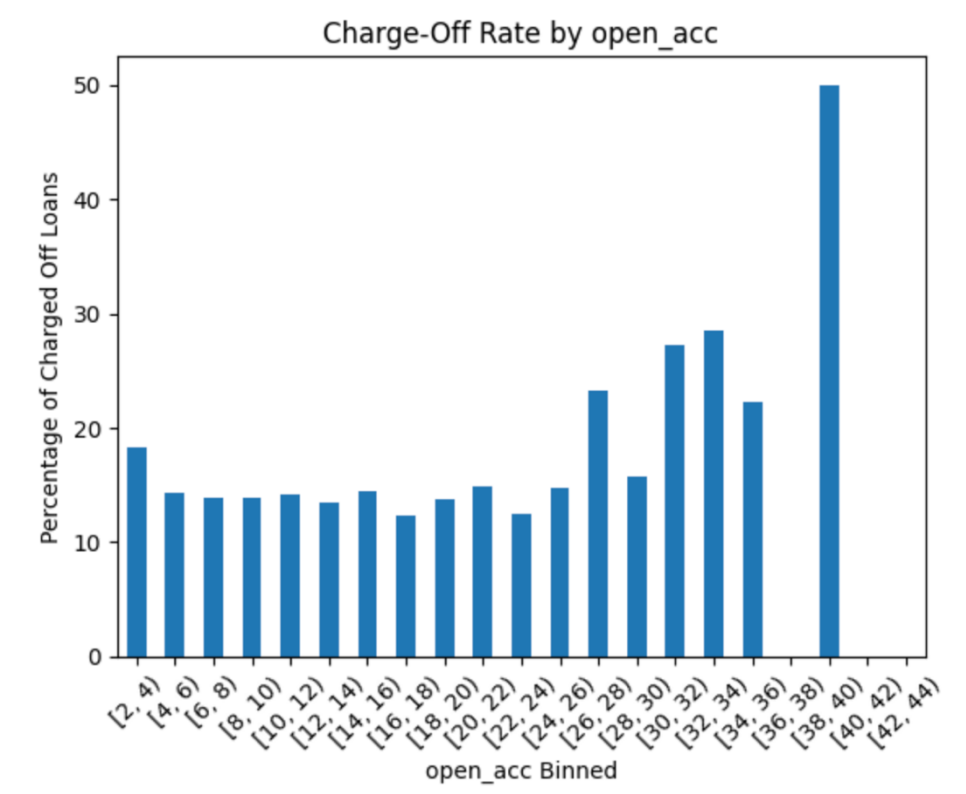
- Nebraska has highly skewed default rate but has only 6 loan disbursals, so it can be considered an outlier

➢ **We observe the following states having relatively higher defaulting rates**



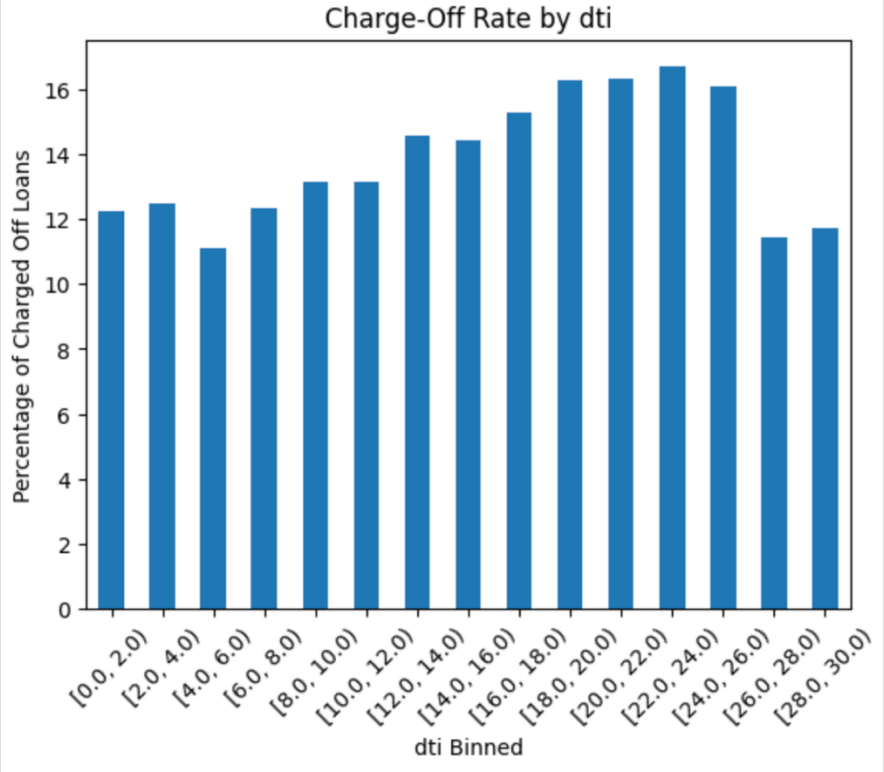
Default vs Loan Enquiries

- **The higher the loan enquiries especially at the higher end of greater than 6 enquiries default rates is higher**



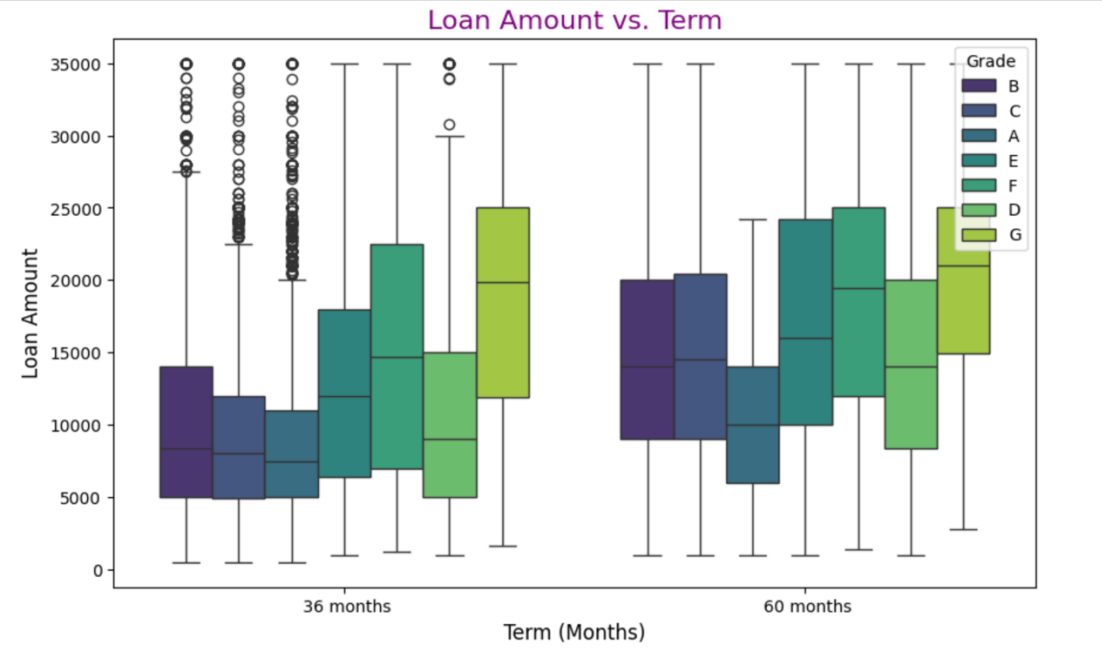
Default vs No. of Open Credit Lines

- No strong correlations can be found against default considering the extreme values
- But in context with other criteria we can use this insight



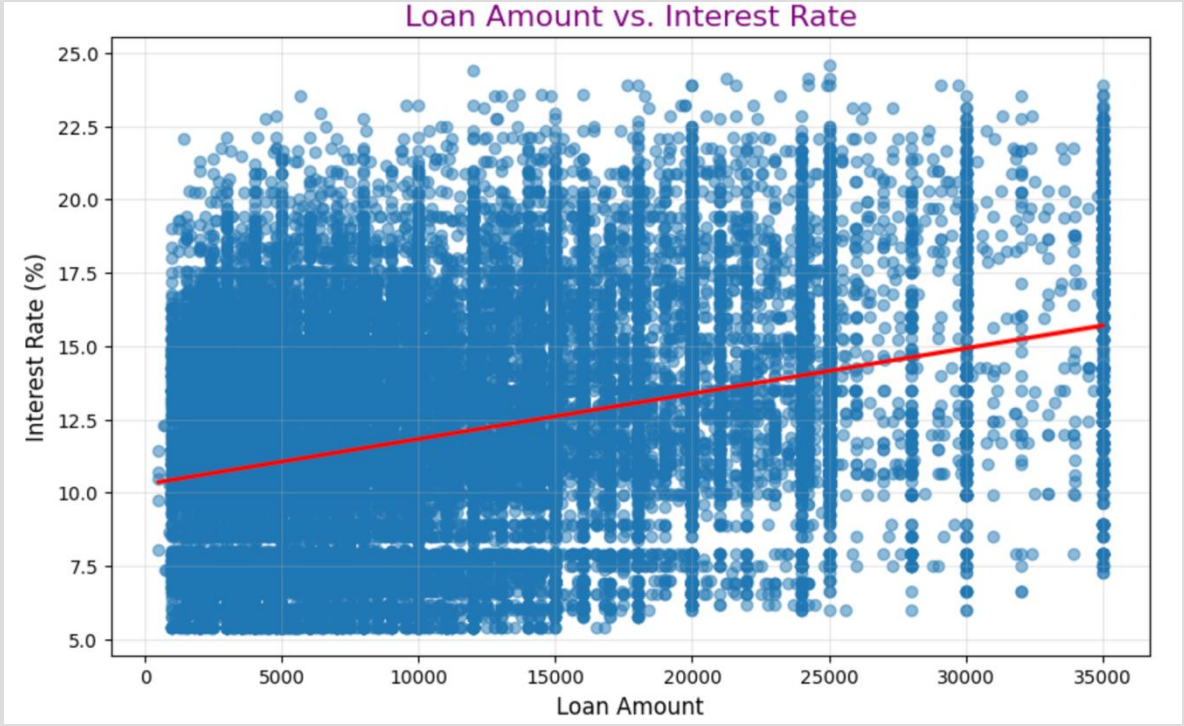
Default vs Ratio of Debt to Income

- **Leaving out the low and high extremes there is a linear correlation between Default rate and Ratio of Debt to Income**



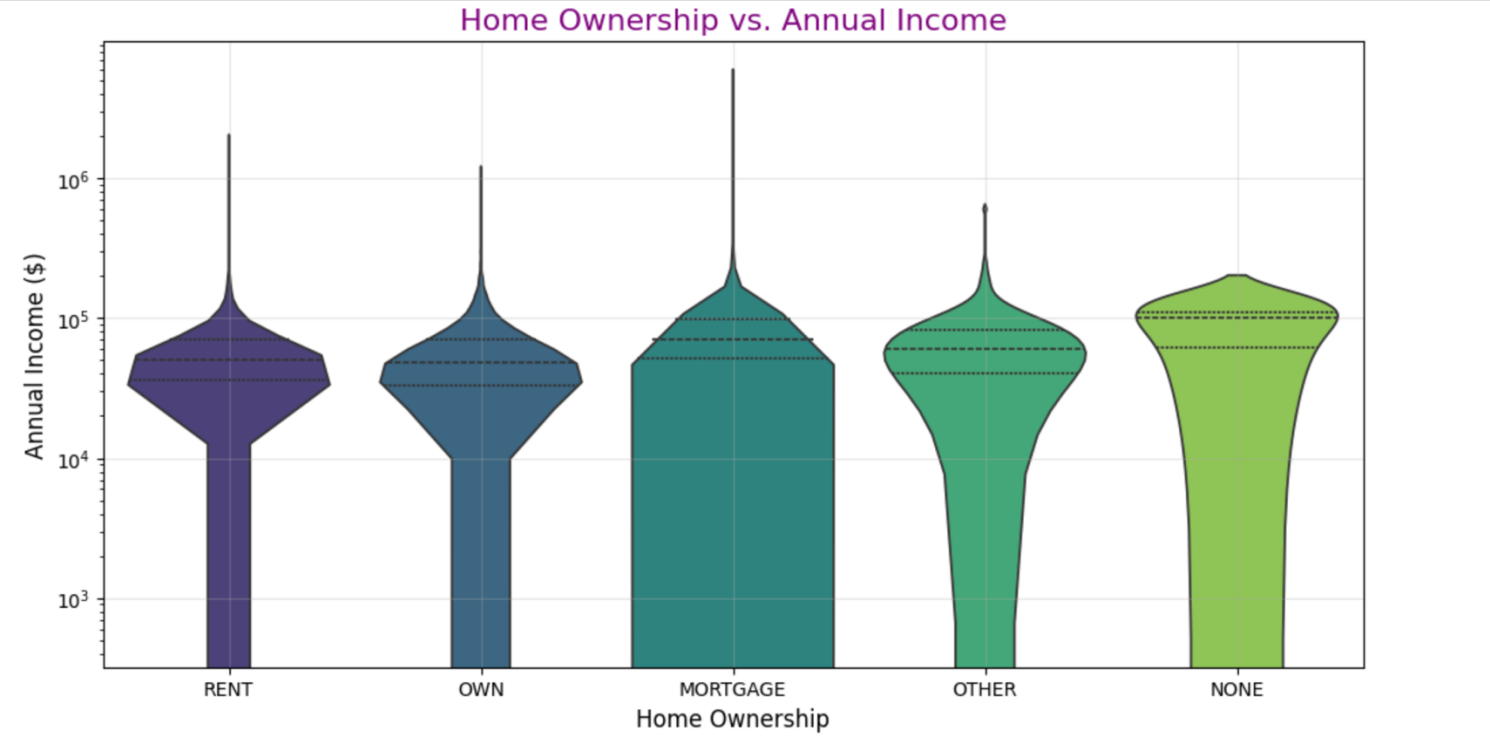
Loan Amount Vs Term

- Loan-Related Parameters - Objective: Check if higher loan amounts are associated with longer terms (e.g., 60 months).
- Strong association exists as postulated



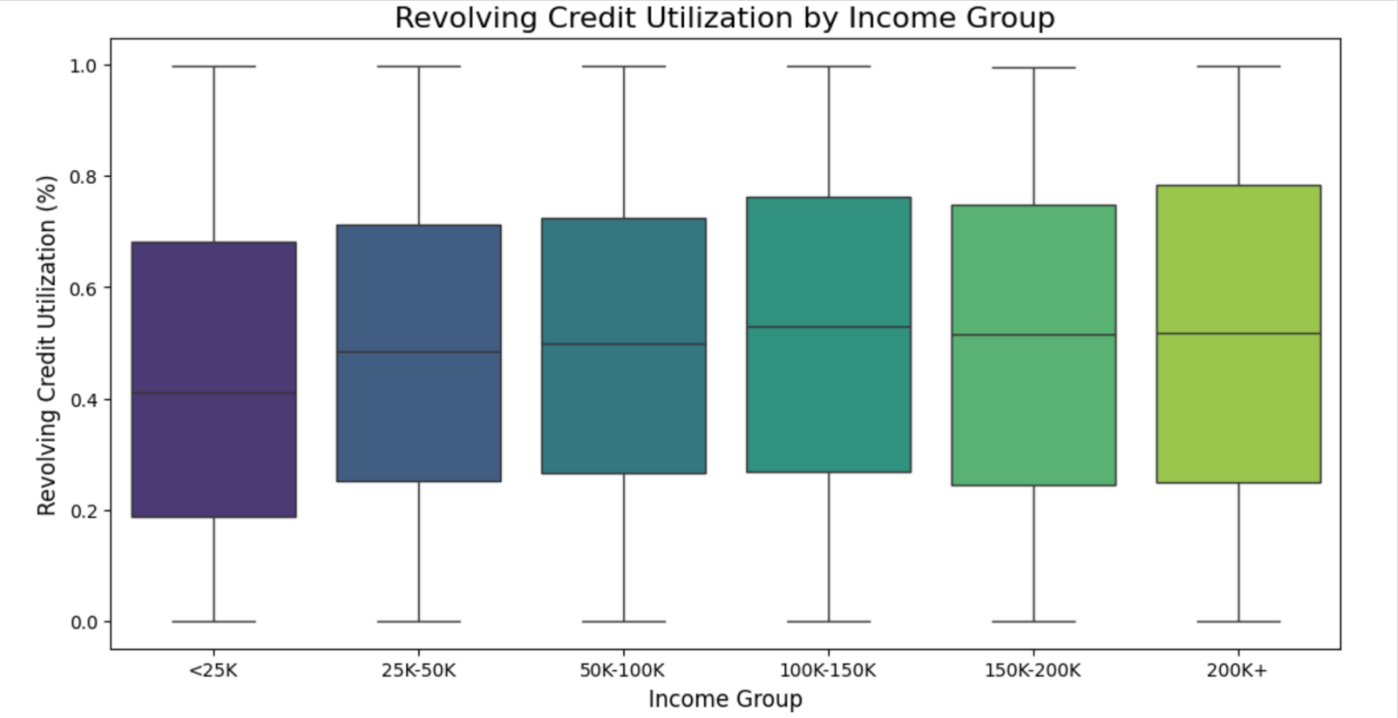
Default vs Home Ownership

- Loan-Related Parameters - Objective: to check whether higher loan amounts tend to have higher interest rates
- Again as postulated Scatter plot with regression line indicates Strong association



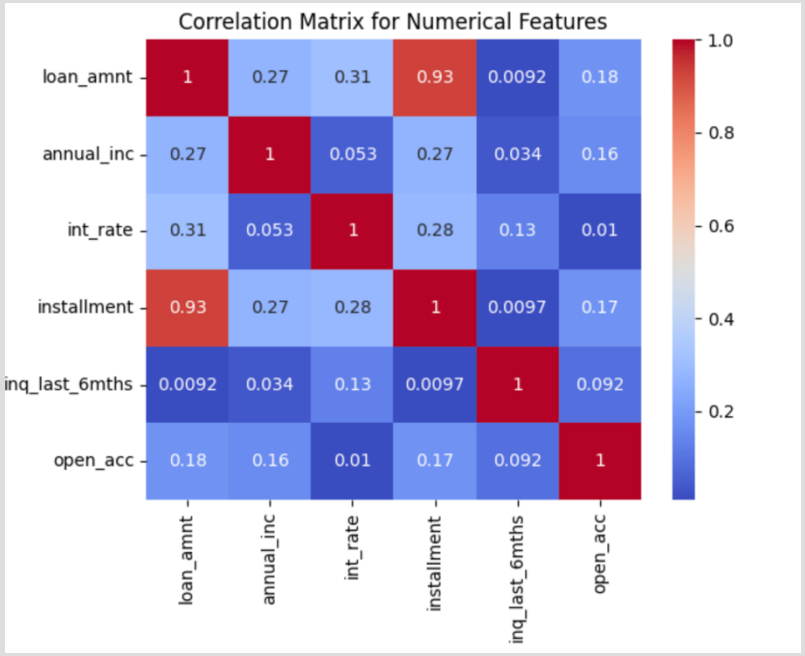
Home Ownership vs Annual Income

- Borrower Demographics - Objective: to Objective: Compare the annual income levels across different home ownership statuses
- Group violin plot does not indicate any significant associations



Revolving Credit Utilization by Income Group

- Debt and Financial Behavior - Objective: To verify whether borrowers with higher income utilize less revolving credit.
- Contrary to the postulate the high income group utilizes greater revolving credite



Correlation Matrix for Numerical Parameters

- Higher the Loan amount naturally the repayment term will be longer but the default risk is also higher
- **We are also seeing a correlation existing between interest rate and installment duration, so the default risk is countered with higher interest rates**

- Heat Map: Selected Variables

(loan_amnt, int_rate, annual_inc, dti, inq_last_6mths, open_acc, revol_bal, term)

Strong relation between the following parameters

- Interest rate and Term Duration
- Debt to Income Ratio and No. of Open Credit Lines
- Credit Revolving Balance and No. of Open Credit Lines
- Debt to Income Ratio and Credit Revolving Balance

Observations

- **With higher interest rate loans we recommend approving shorter repayment term**