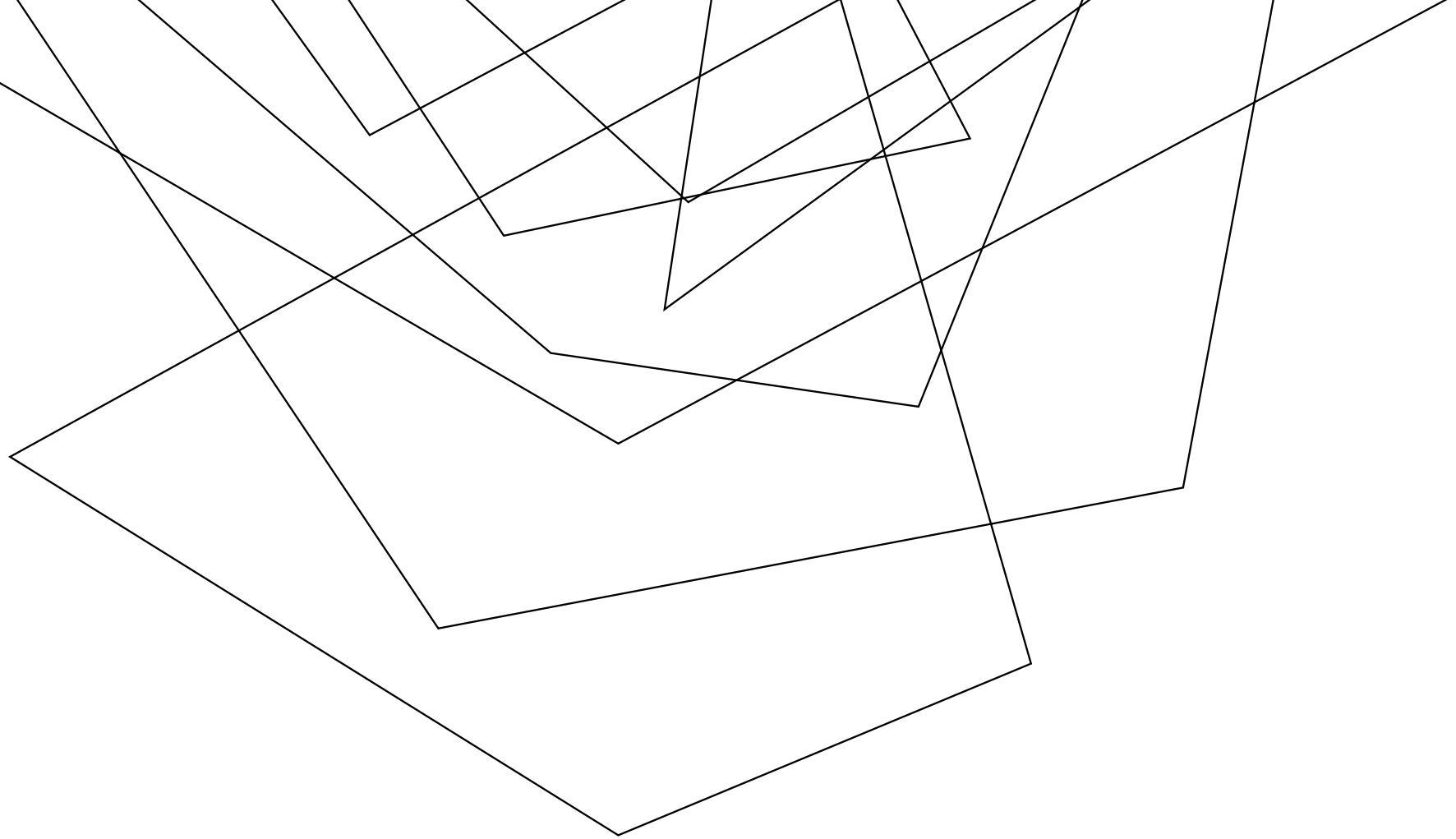


LENDING CLUB CASE STUDY

Group Member: Ritesh M Sanghavi
August 2023

BUSINESS UNDERSTANDING

This company is the largest online loan marketplace, facilitating personal loans, business loans, and financing of medical procedures. Borrowers can easily access lower interest rate loans through a fast online interface.

Like any other lending company, the success of the business is in successfully identifying the risk of lending before approving the loan using certain borrower related parameters that defines his/her behavior.

Two types of risks are associated with the bank's decision:

1. If the applicant is likely to repay the loan, then not approving the loan results in a loss of business to the company
2. 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

PROBLEM STATEMENT

To identify these risky loan applicants, then such loans can be reduced thereby cutting down the amount of credit loss. Identification of such applicants using EDA is the aim of this case study.

In other words, the 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.

APPROACH TO SOLUTION

UNDERSTAND DATA

- Loading data
- Understanding the structure and data dictionary
- Sanity Checks

DATA CLEANING

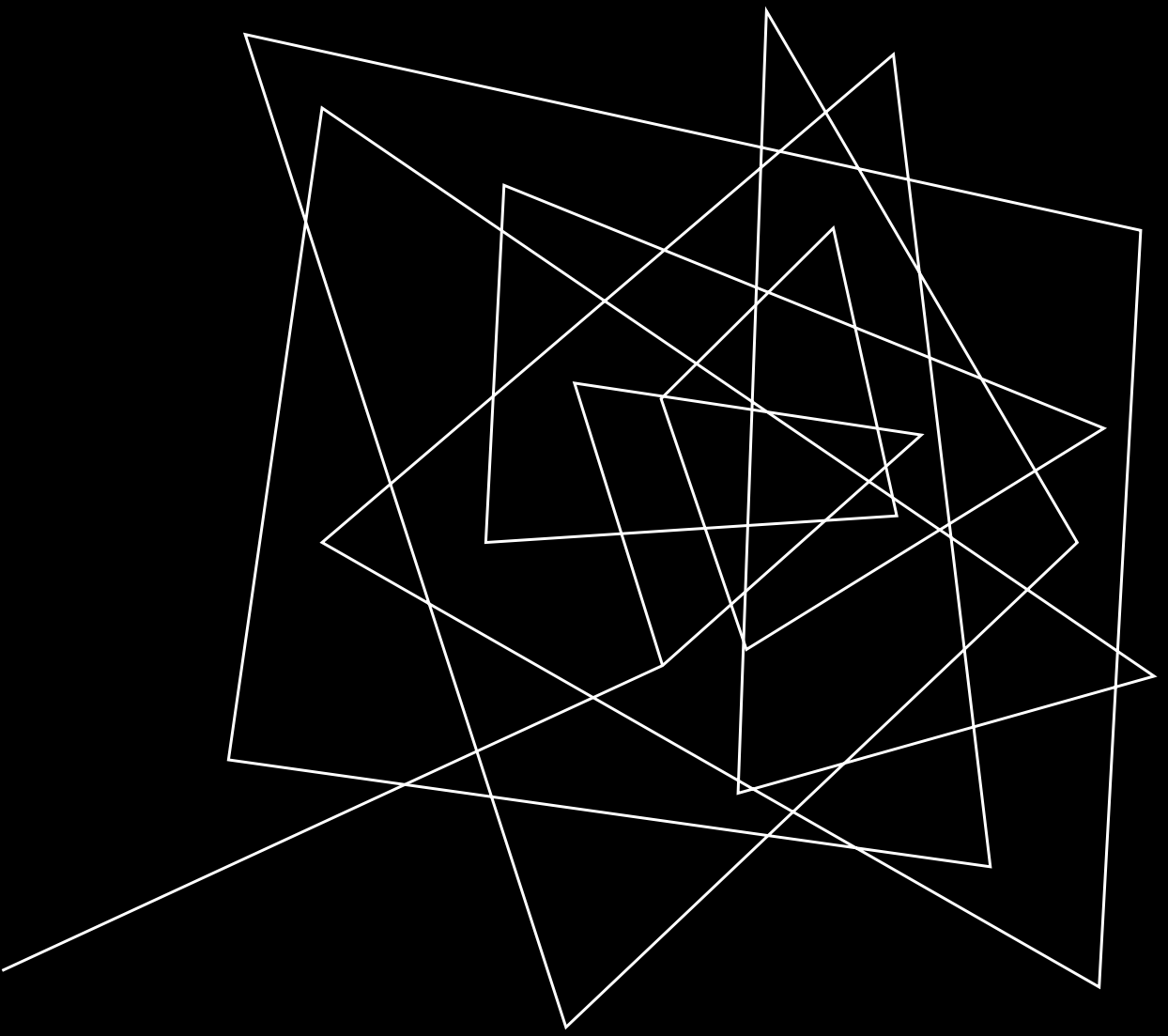
- Data cleanup
- Fill missing data
- Derive columns
- Correct the format, Drop unwanted columns

DATA ANALYSIS

- Analyze data for each individual column
- Analyze its correlation with other columns
- Identify outliers and create meaningful relation between variables by applying domain knowledge

RECOMMENDATIONS

- Based on our observations and correlations between variables we can study the user behavior and patterns
- This can be used to flag any risk for future approvals of loan with similar patterns and minimize the loss



UNDERSTAND DATA

Step 1

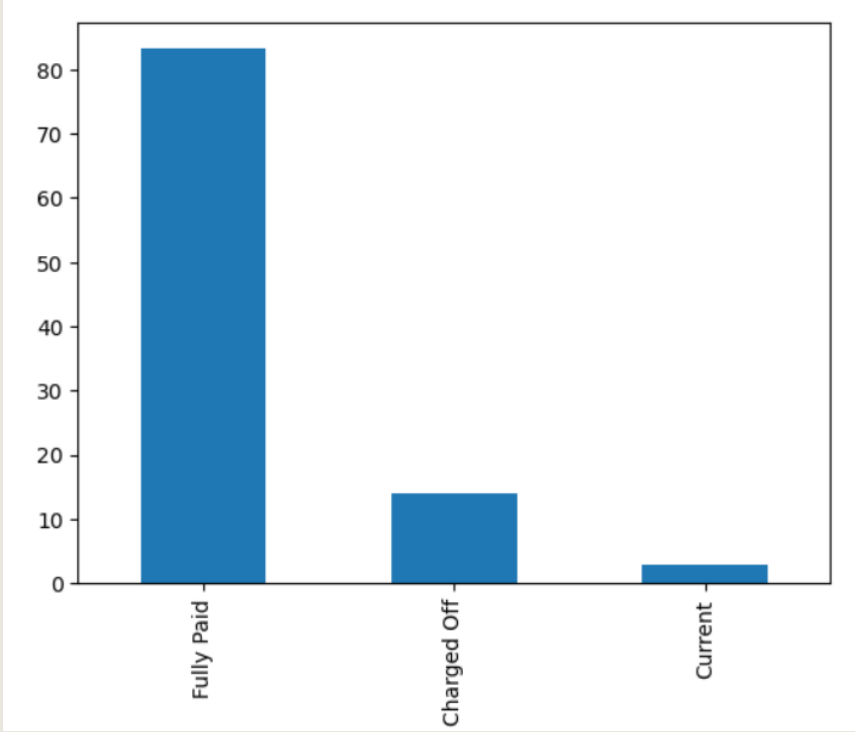
DATA UNDERSTANDING

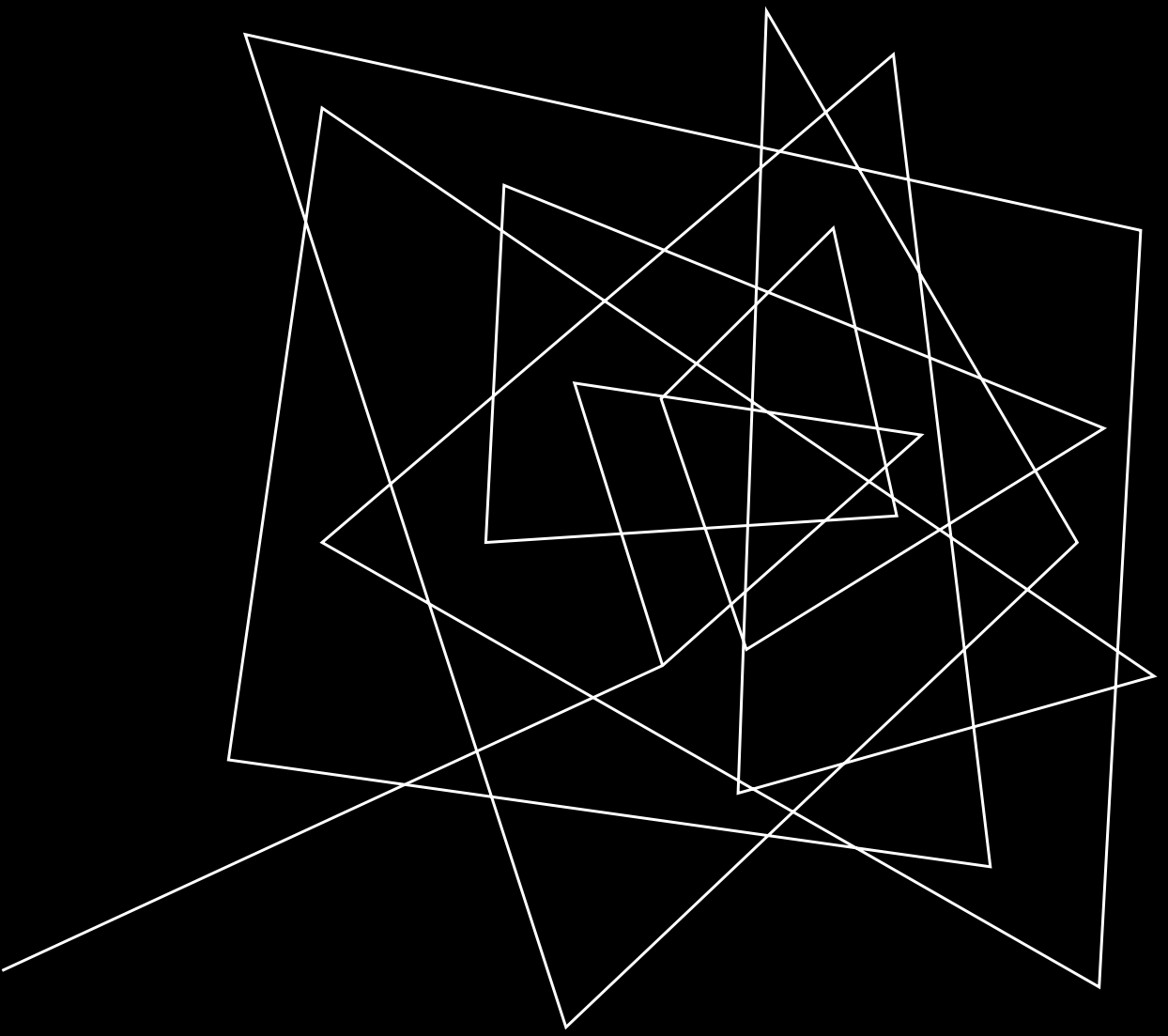
- Load required libraries like numpy, pandas, matplotlib, seaborn, etc
- Load data from CSV file into python variable for ease of manipulation
- Identify empty rows and columns
- Understand Data Dictionary to understand meaning of each row to identify the purpose of each column so that we can establish correlation

INITIAL ASSESSMENT OF LOAN STATUS

- 83.19% is Fully Paid
- 13.97% is Charged Off(defaulted)
- 2.84% is Current (in progress loans)

For the data provided between year 2007-2011, 13.97% borrowers defaulted on their loans



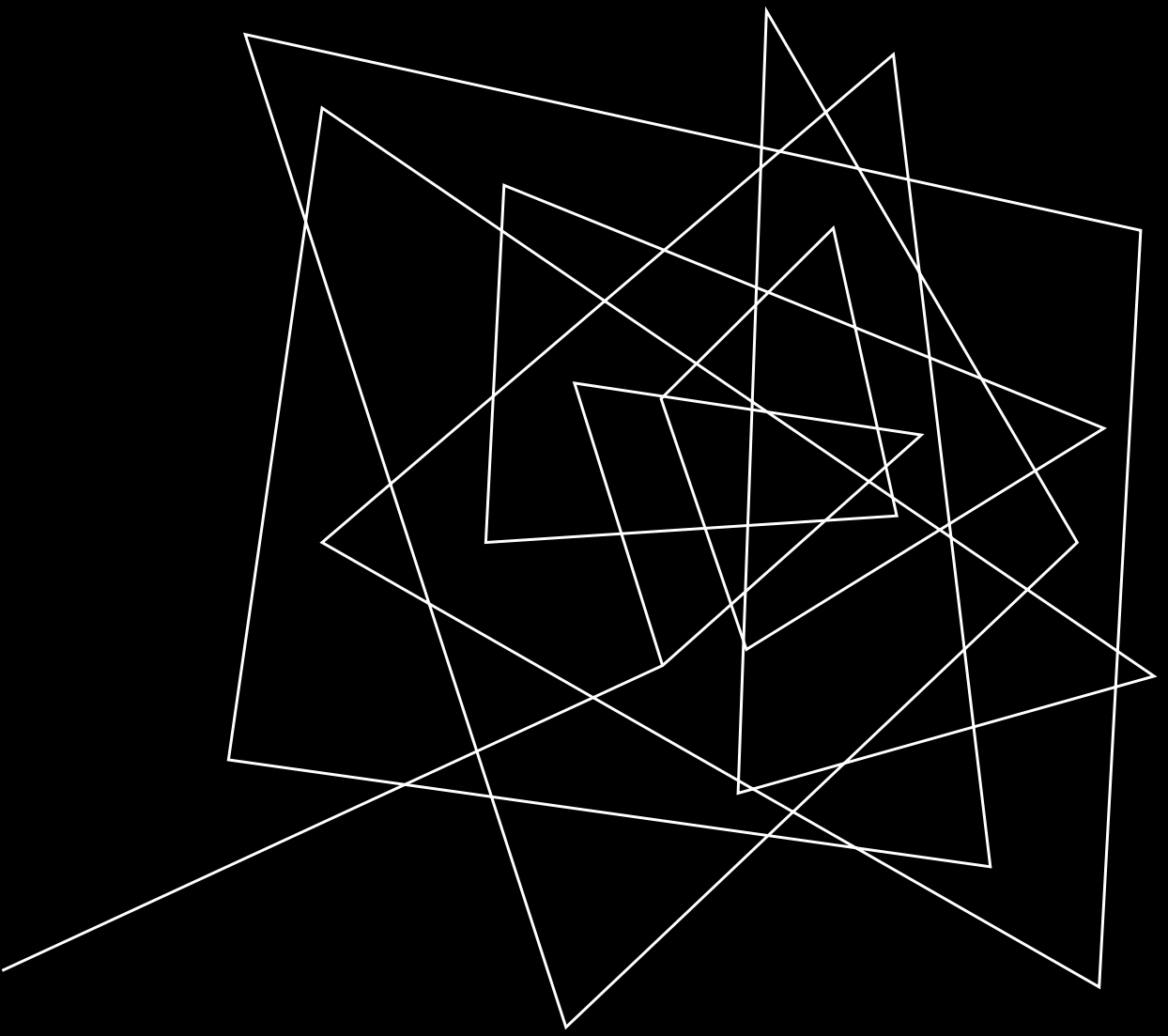


DATA CLEANING

Step 2

DATA CLEANING

- Identify columns with missing values. Delete these columns as they don’t help with analysis
- Repeat the process for columns with NULL or no values.
- Drop the rows with loans that are in progress as the data they have do not help us understand if the parameters of a borrower are of positive or negative impact
- Remove columns with customer behavior as it is not available at the time of loan application. Hence, it bears no impact on credit approval
- Go through the remaining columns available and you will find few who have no impact on loans outcome. Drop these columns
- Standardize the format of column values. If required, change them from object to INT or FLOAT as per the need.



DATA ANALYSIS

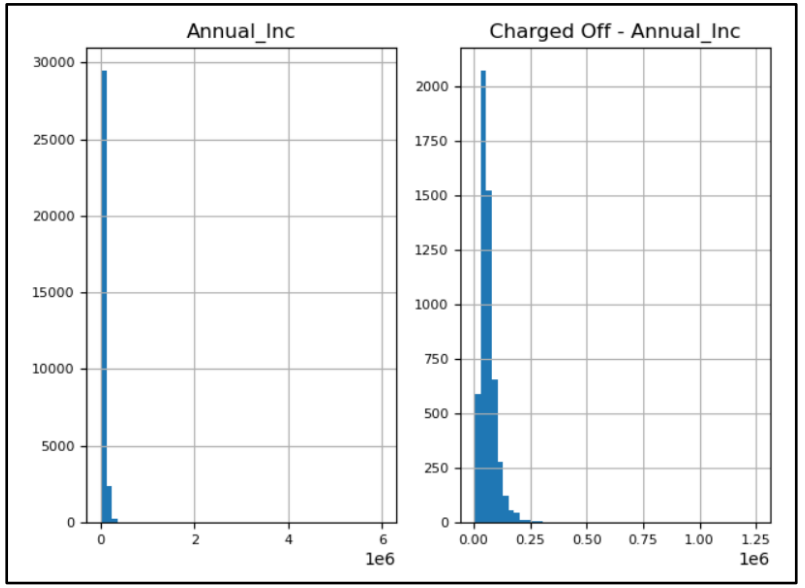
Step 3

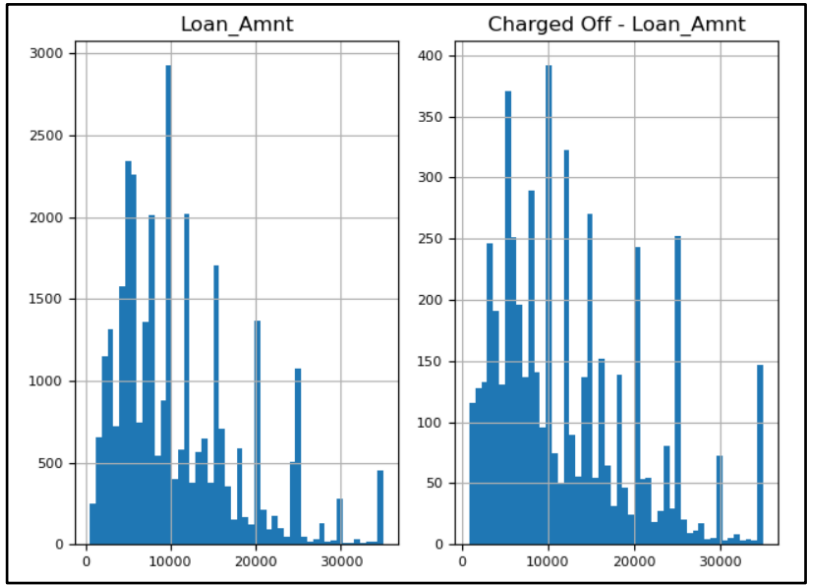
Loan Amount: Loan amount ranges between 500-35k across categories

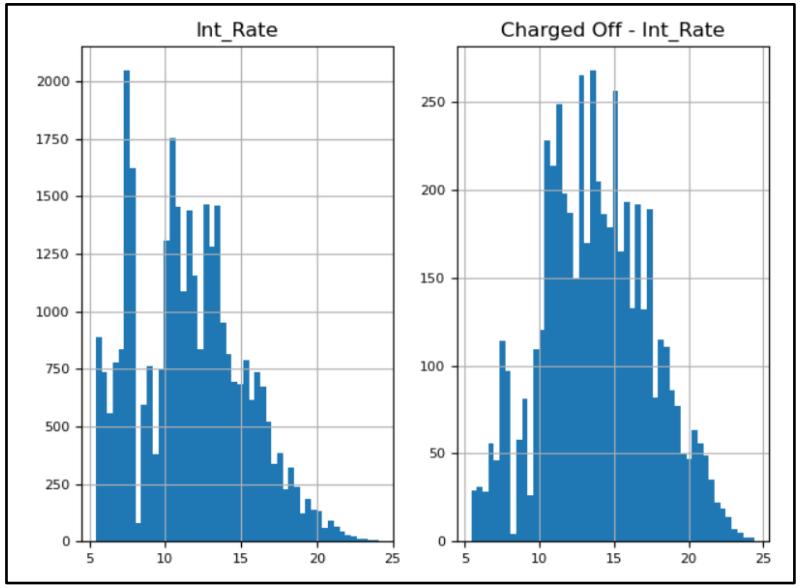
Term: Loan term is either 36 months or 60 months

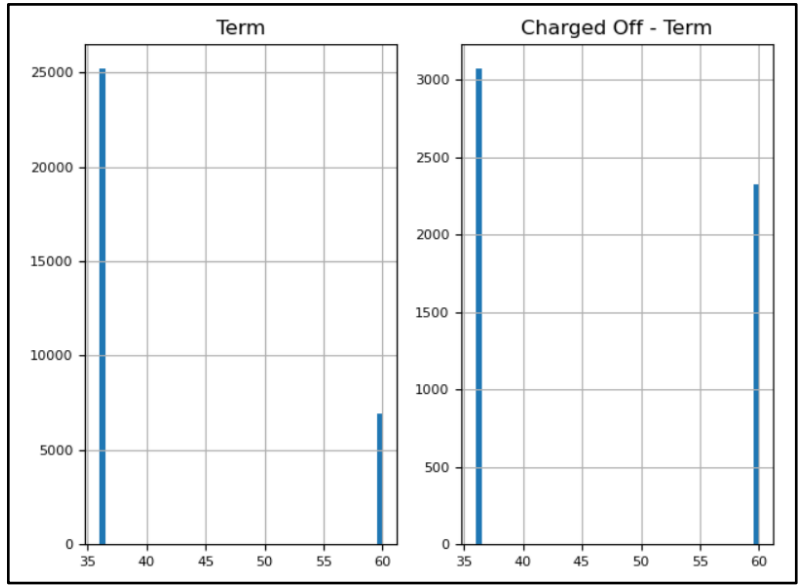
Interest Rate: Interest rates ranges between ~5-25%

Annual income: Annual income is on lower side for most charged off loans







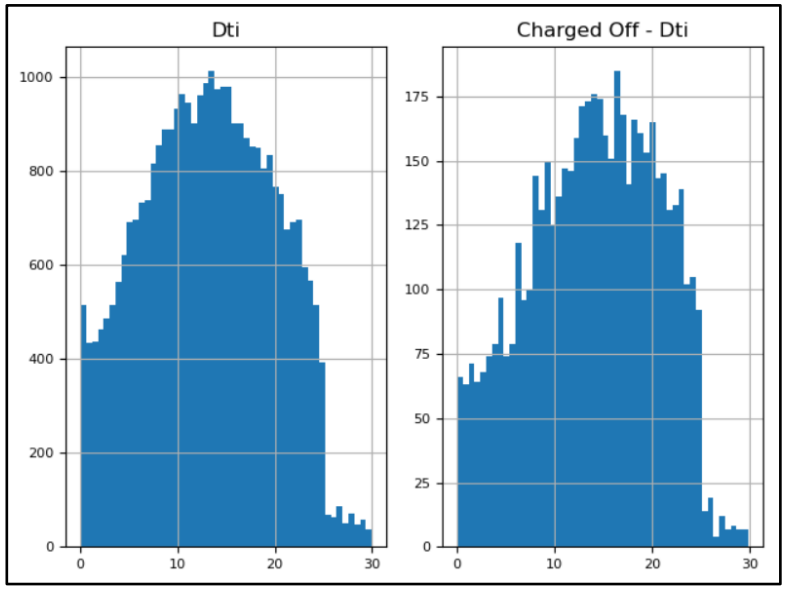


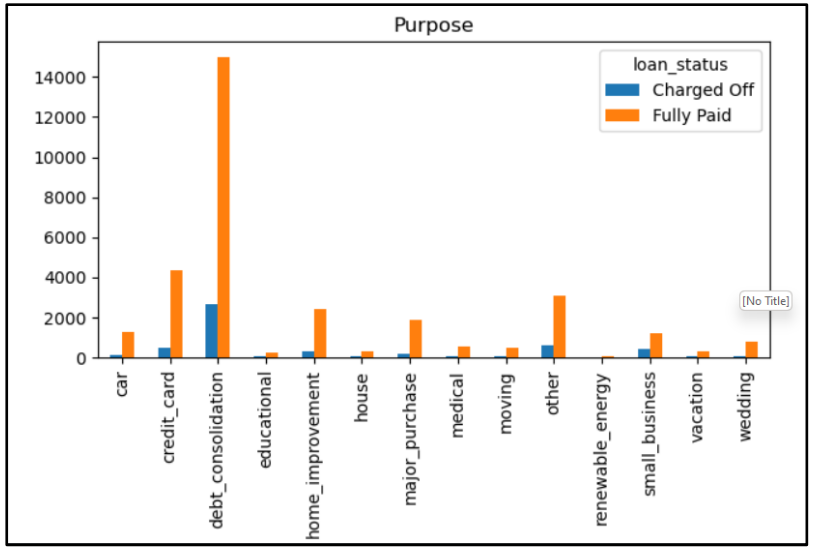
DTI Ratio: Debt to Interest ratio is low for default loans

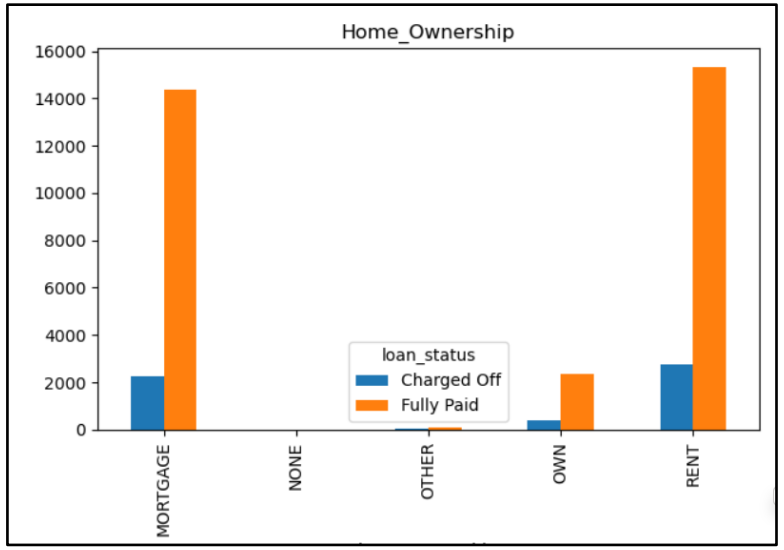
Grade: Loans are approved for grades A, B and C. Default loans have higher proportion of low-grade loans

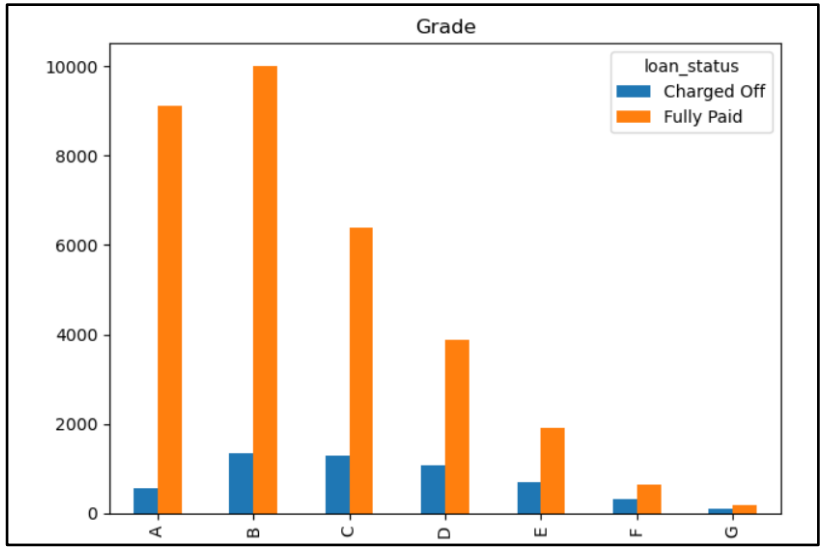
Home ownership: Home ownership is either renter, mortgaged or owned for majority of loans

Purpose: The purpose of loan is majorly for debt consolidation, credit card of other reasons





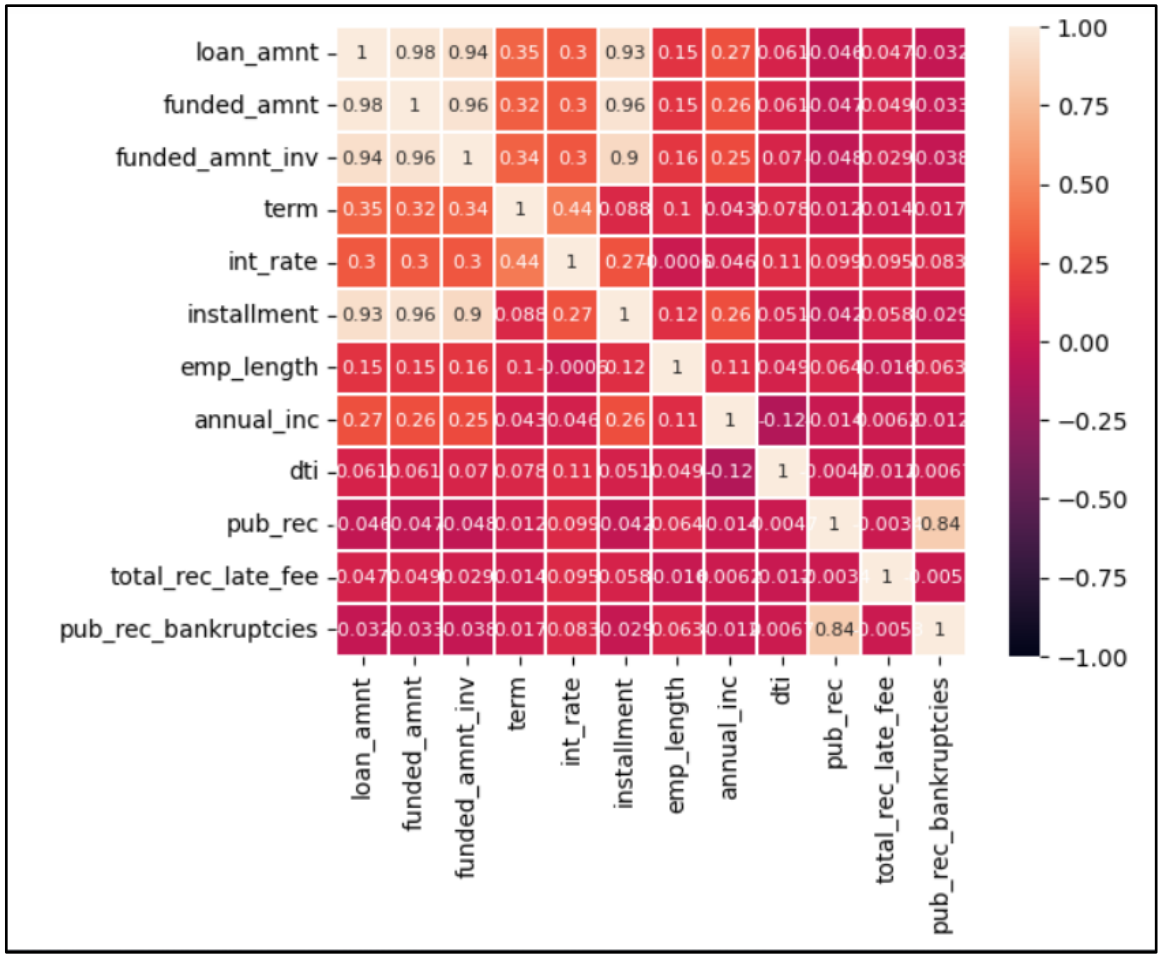




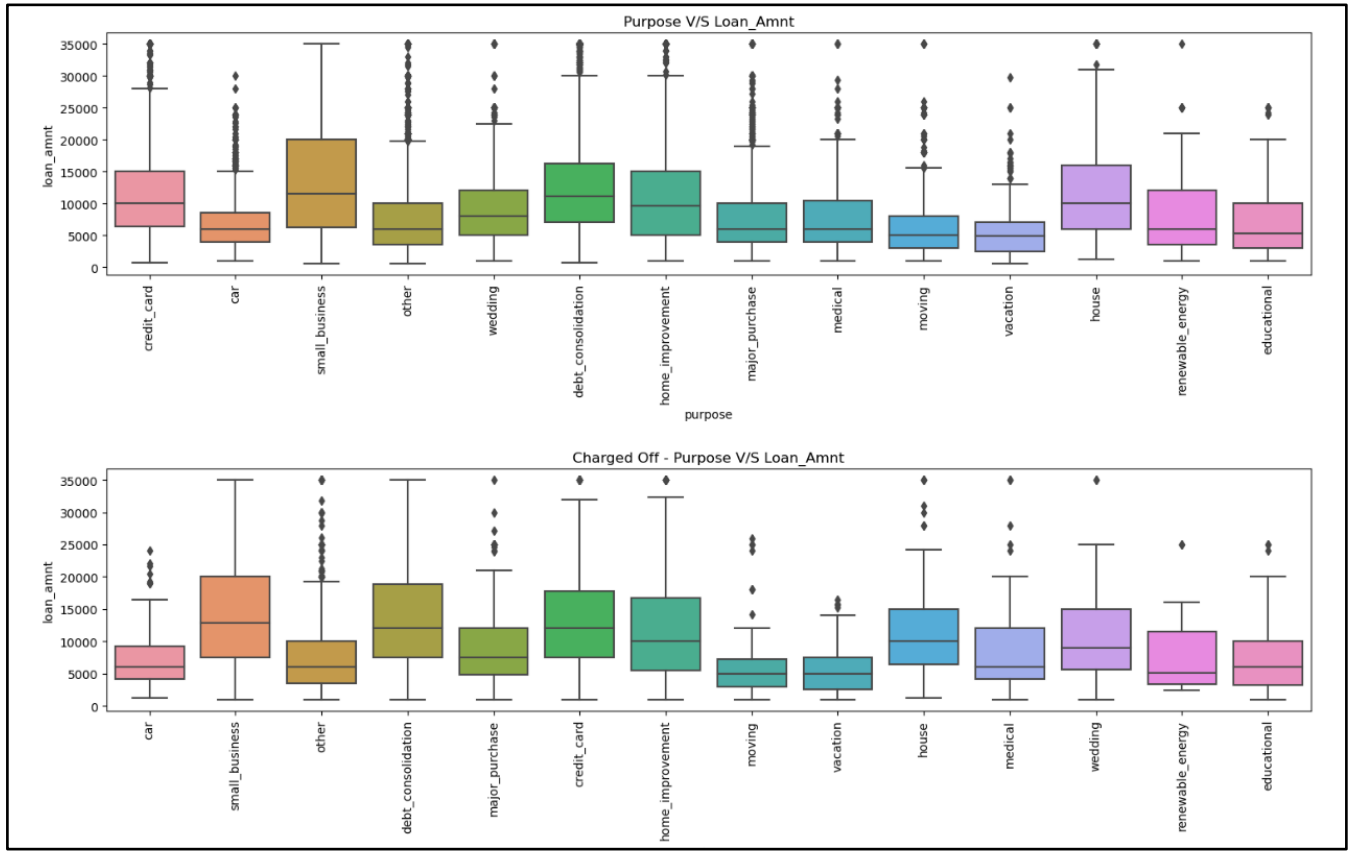
BIVARIATE ANALYSIS

Correlation Matrix:

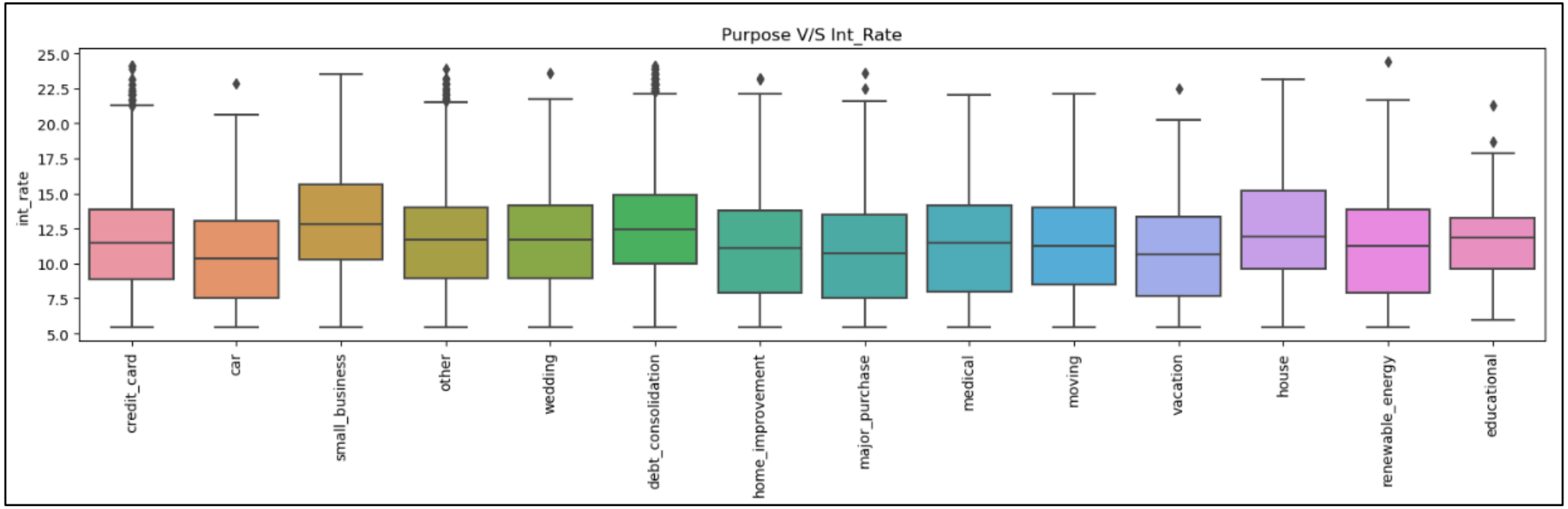
- Correlated parameters: Loan amount, committed Funded amount, committed funded amount by investors
- Negatively Correlated parameters:
  - DTI vs Annual income: Debt to income ratio is high if the income is low and vice versa
  - Employee Length vs Interest Rates



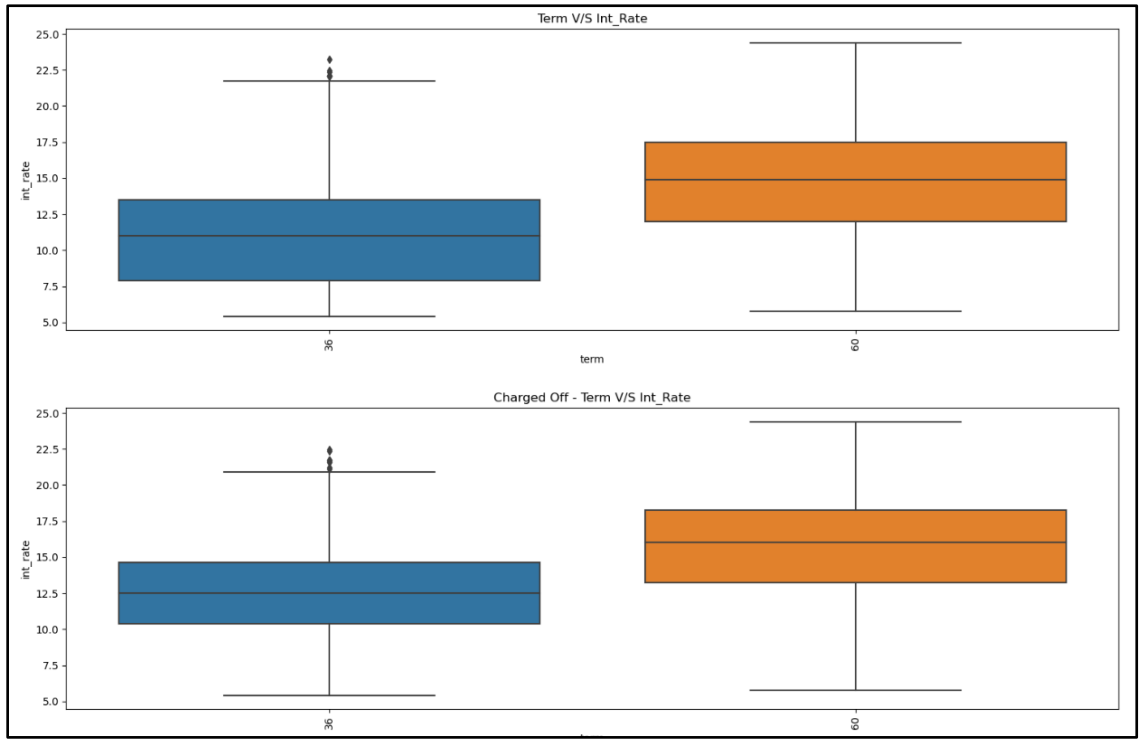
- Amount borrowed is highest for small business
- Small business is also a reason for loan with highest amount getting default
- Other reasons for charged off loans are – Debt consolidation, credit card and home improvements

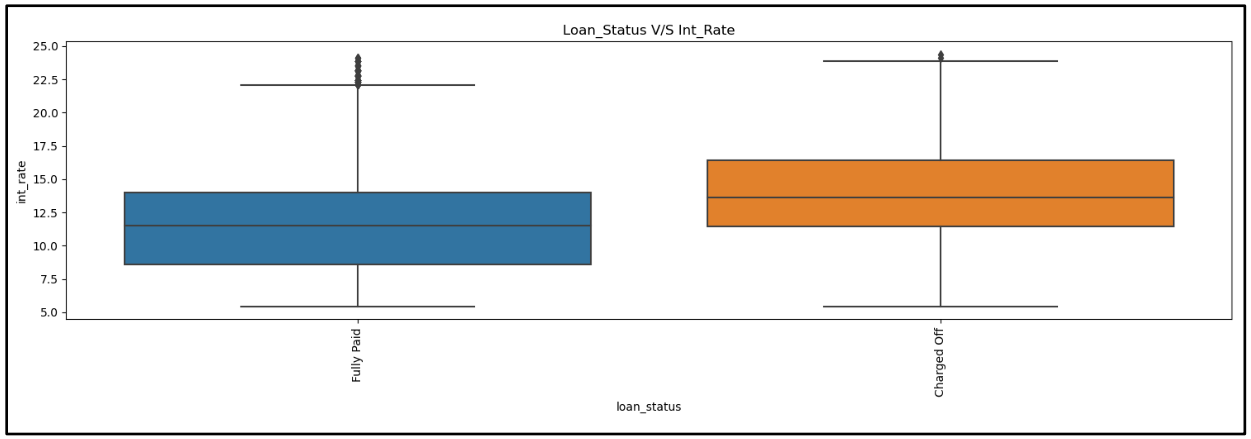


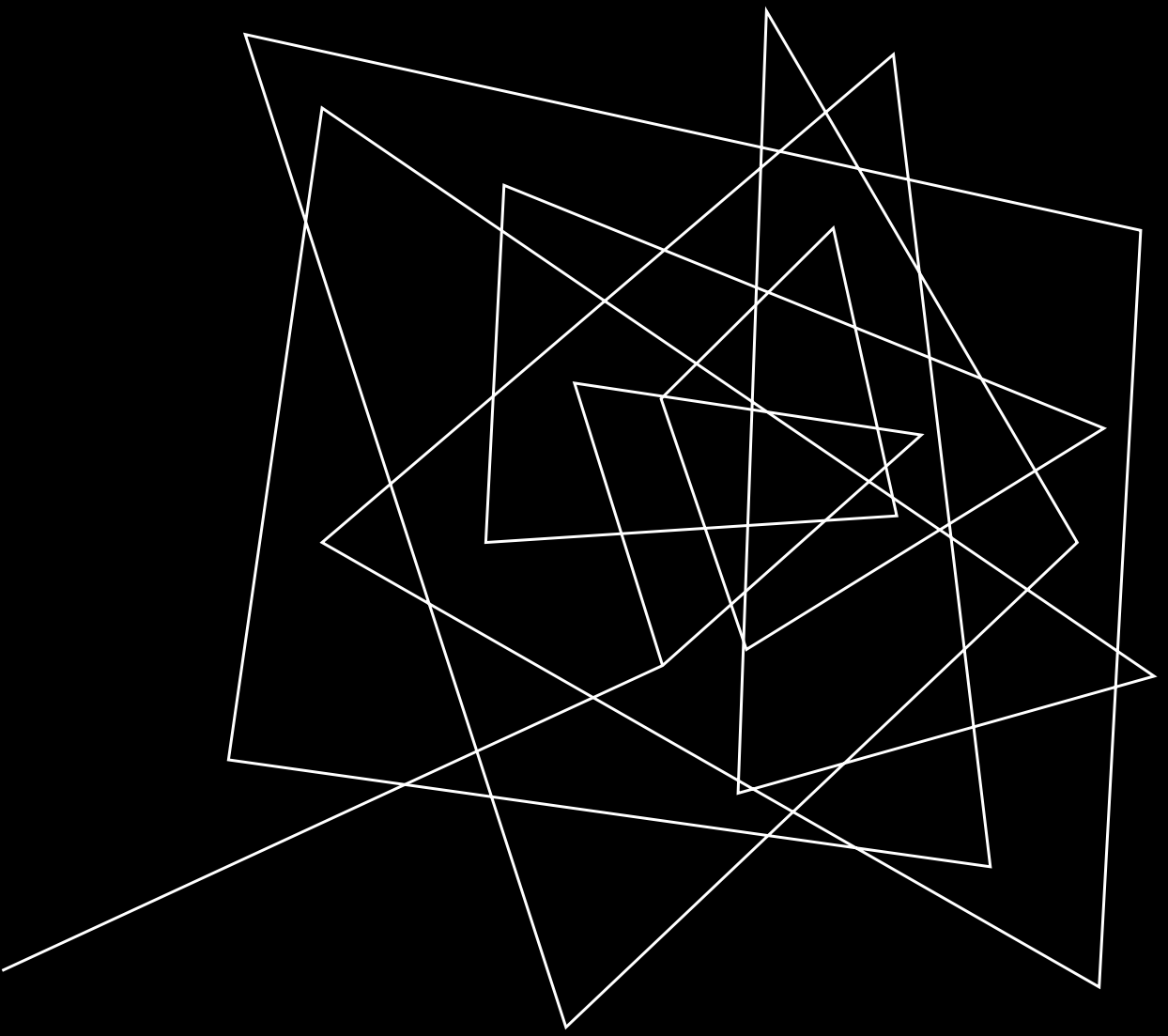
- Interest rates are highest for small business. These explains the reason for Small business loans getting charged off the most



- Interest rates are higher for loan with longer tenure. In our case, loan with 60 months tenure
- Majority of default loans have interest rate in higher range





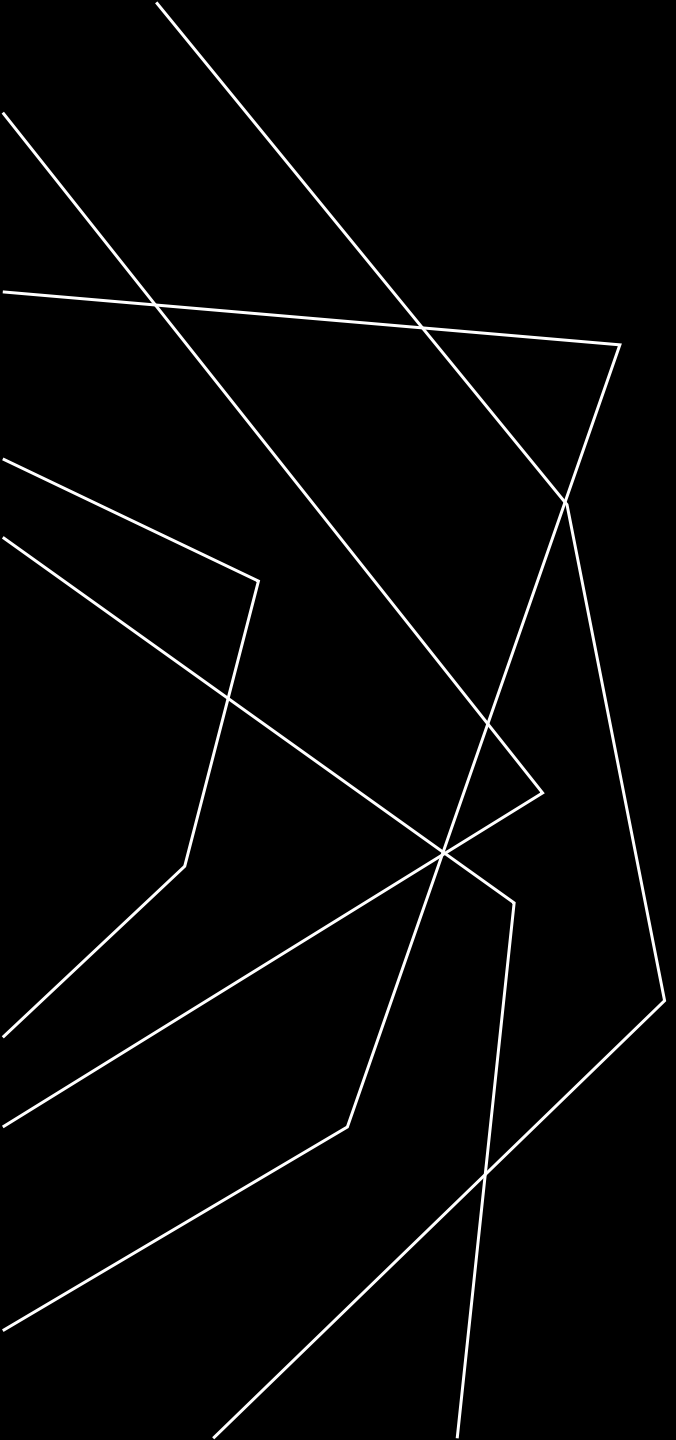


RECOMMENDATION

Final Step

SUMMARY

- Following variables are driving factors in loan defaults: Grade, Interest Rate, Term, Home Ownership, Purpose, Year of approval
- **Grade** – Higher the grade better the chance of successful closure of loans
- **Interest Rate** – Loan with higher interest rates are ones with higher risk
- **Home ownership** – Data shows that rented or mortgaged home ownership is directly related to loan defaults
- **Purpose** – Small business tends to fail to repay the loans
- **Year of Approval** – Lender got aggressive with passing years and the number of defaults also went up



THANK YOU

Ritesh Sanghavi