LENDING CLUB CASE STUDY

- Rezon Chakraborty
- Ayush Saxena

AGENDA

- Project Background
- Problem Statement
- Data Classification & Identification
- Data Observation
- Data Preprocessing
- Univariate Analysis
- Bi-Variate Analysis
- Conclusion

PROJECT BACKGROUND

- Lending Club Is a consumer finance company. It is the first peer-to-peer lending fintech company.
- LC provides a platform to lenders and borrowers where;
  - lenders can invest from the loan listings and select loans they want to invest after a thorough research on the information provided by the borrower.
  - The borrowers can request for unsecured loans from LC's platform for various purposes with a tenure

PROBLEM STATEMENT

"Using Exploratory Data Analysis technique perform analysis on the data set of Lending Club, to find strong indicators and patterns to identify potential defaulters."

- Lending Club provides unsecured loans to borrowers and requires insights or strong indicators which enable them to make business decisions that does not lead to a Bad Business:
  - NOT reject a loan application of borrower who can Fully Repay the loan.
  - NOT approve loan application of borrower who'll be a Defaulter.

DATA CLASSIFICATION & IDENTIFICATION

- The columns that were key drivers for our analysis are as follows:
  - Term - The number of payments on the loan. Values are in months and can be either 36 or 60.
  - Revol_Util - Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
  - Loan_status - Current status of the loan.
  - Purpose - A category provided by the borrower for the loan request.
  - Grade – Lending Club assigned loan grade.
  - Sub_grade - Lending Club assigned loan subgrade
  - Revol_util_categories – Segmented the revol_util column into 5 categories i.e., Extremely Good, Good, Moderate, Risky, Extremely Risky.

DATA CLASSIFICATION & IDENTIFICATION

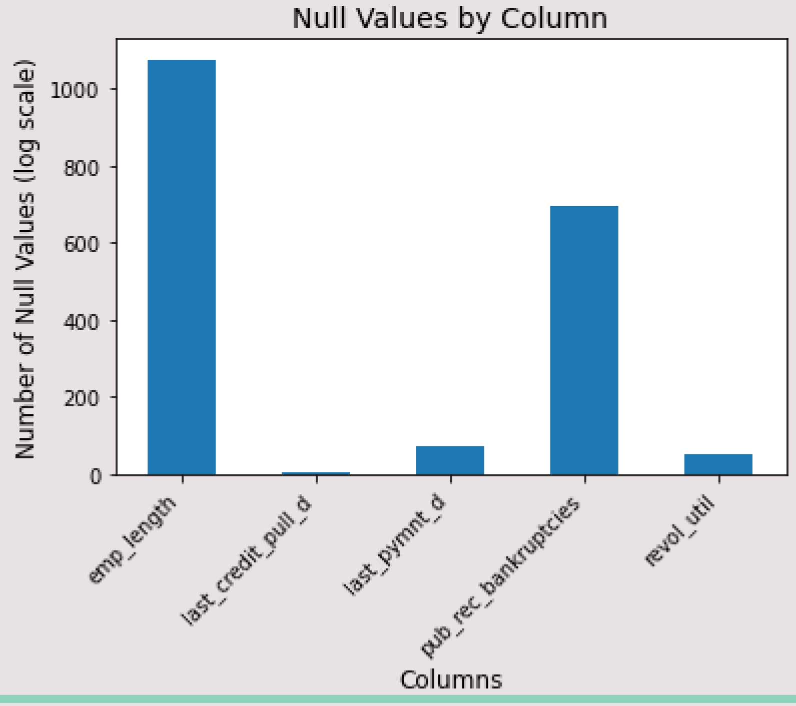
- The columns that were key drivers for our analysis are as follows:
  - Home_ownership - The home ownership status provided by the borrower during registration. The values are RENT, OWN, MORTGAGE, OTHER.
  - Int_rate - Interest Rate on the loan
  - Int_rate_cats – Segmented the column int_rates into 5 segments i.e., “0%-5%”, “5%-10%”, “10%-13%”, “12.5%-15%”, and “15%+”
  - Dti - A ratio calculated using the borrower’s total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower’s self-reported monthly income.
  - Dti_categorized – Segmented the dti column into three categories i.e., Good, Concerned, Risky

DATA OBSERVATION

The high-level observation of the data set yields the following information about the kind of data available:

- No Co-Borrower is present for all the loan accounts indicating all loan applications are for INDIVIDUALS
- All account have no payment plans
- The lenders in this data set are listed under **fractional loans**
- All account have **no payment plans**
- No tax liens on the assets of the borrower is found in the data set
- The Fully Paid ratio to Charged off is **4:1** respectively

DATA PREPROCESSING



- Remove all columns that are null:
  - **Out of 111 Columns, 54 Columns contains only null value**
- Remove summary columns and descriptive columns.
- Remove columns with highest null values
  - mths_since_last_delinq
  - mths_since_last_record
  - next_pymnt_d
  - desc
  - emp_title
- Columns that remain after removing null values columns are 39 and 39717 rows

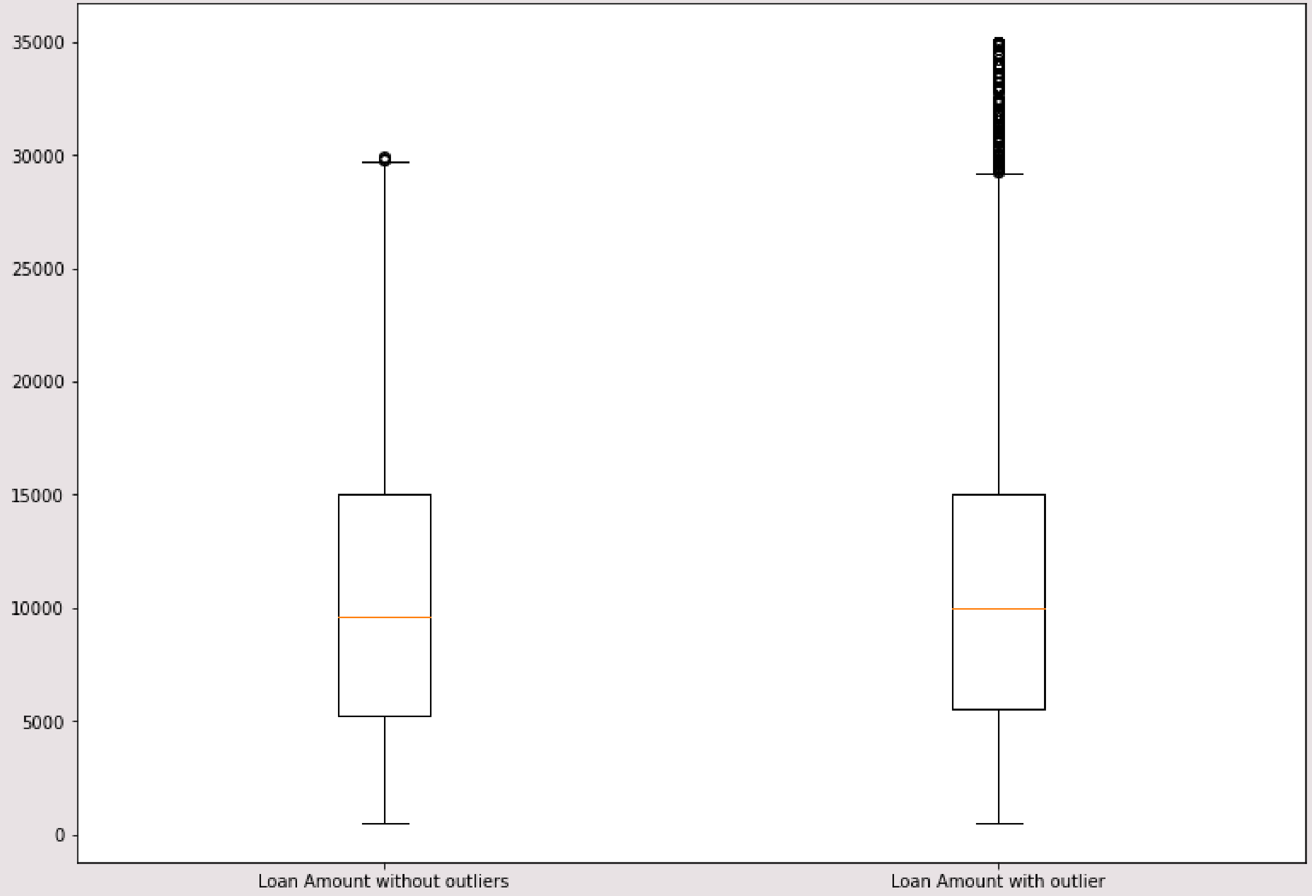
DATA PREPROCESSING

- Remove "**None**" from the **home_ownership** columns as it indicates that the borrower is either **homeless** or the value is **unrealistic** in the data set.

- Updated all the "**N.A.**" rows in the **emp_length** to 0 which translates to all the people who have no experience or do not have a job.

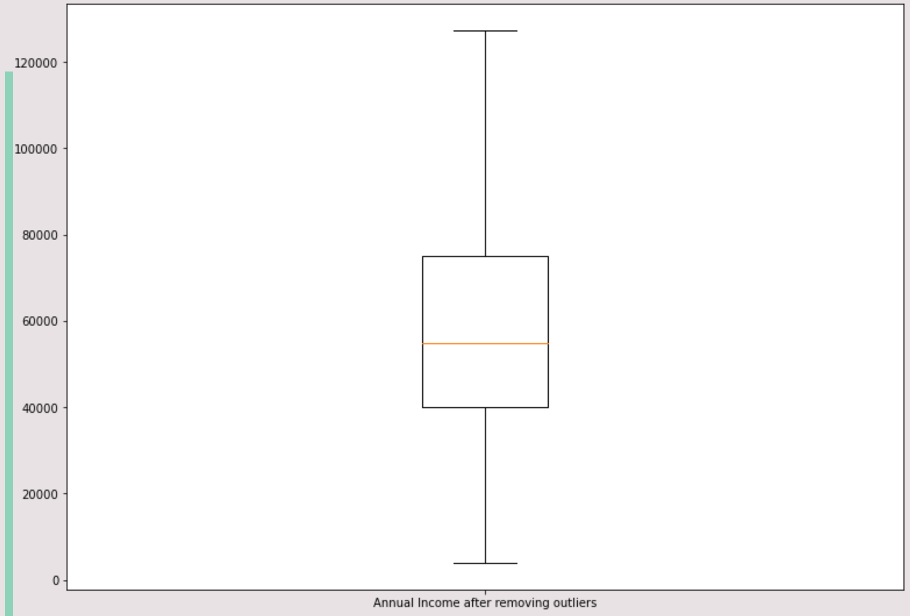
DATA PREPROCESSING

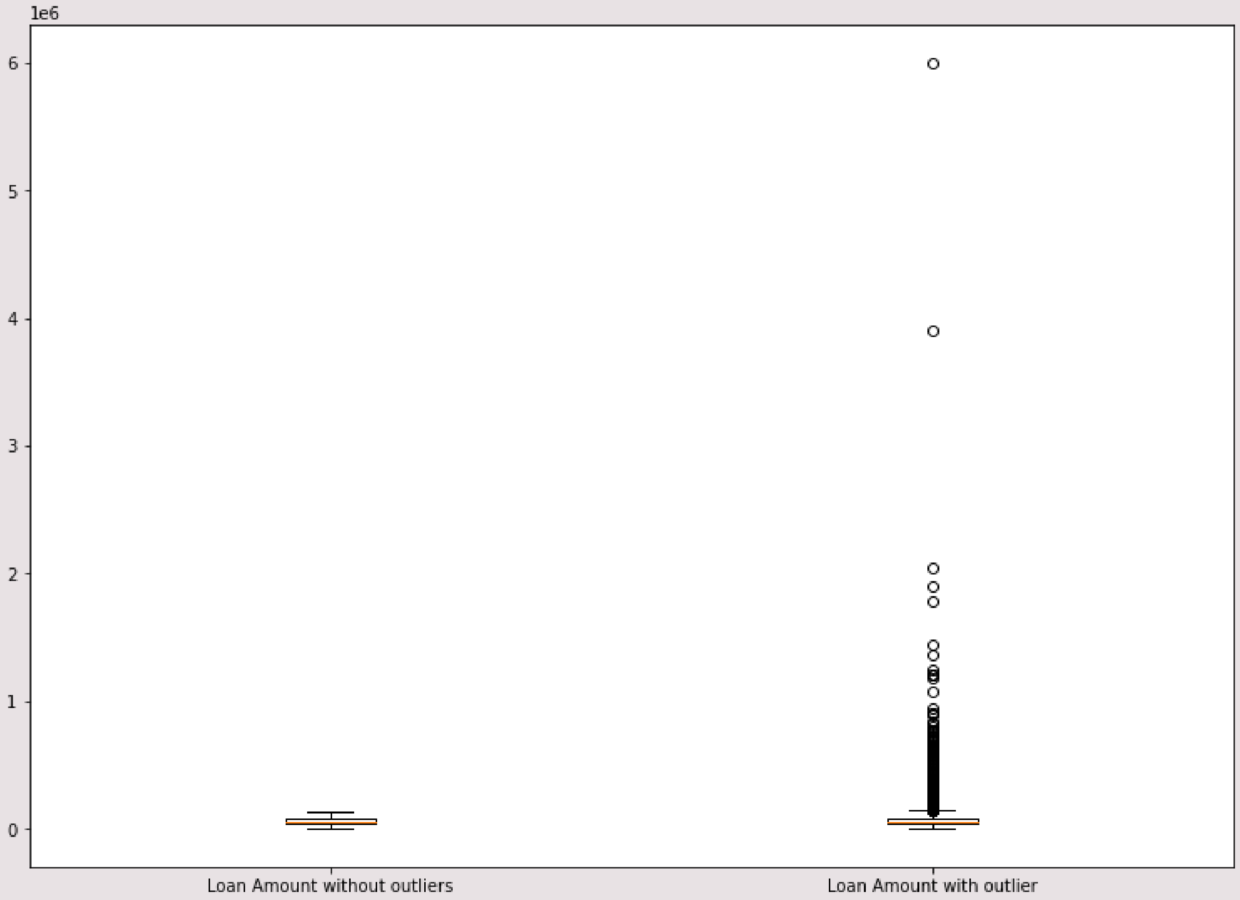
- Removed outliers from the column **"loan_amount".**



DATA PREPROCESSING

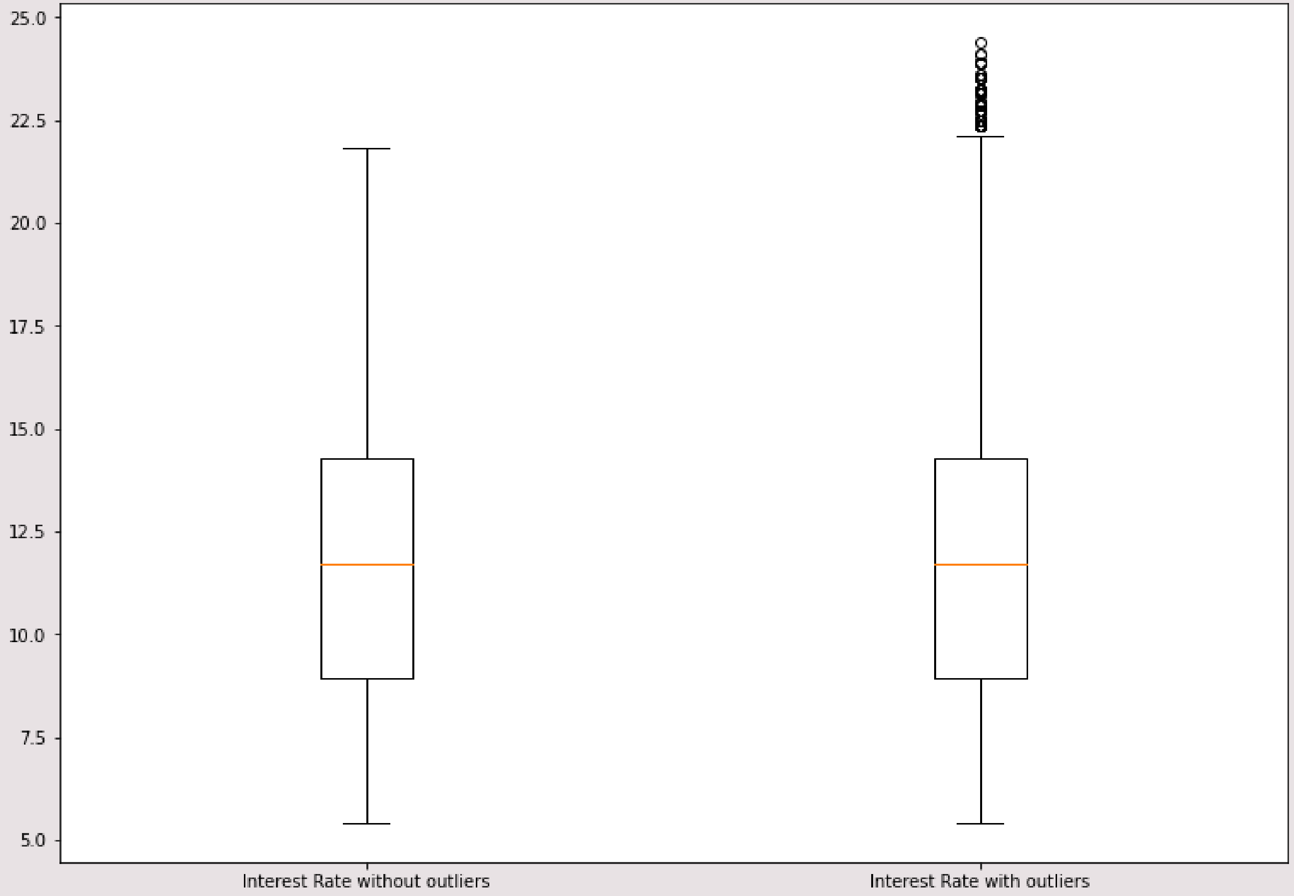
- Removed outliers from the column **“annual_inc”.**





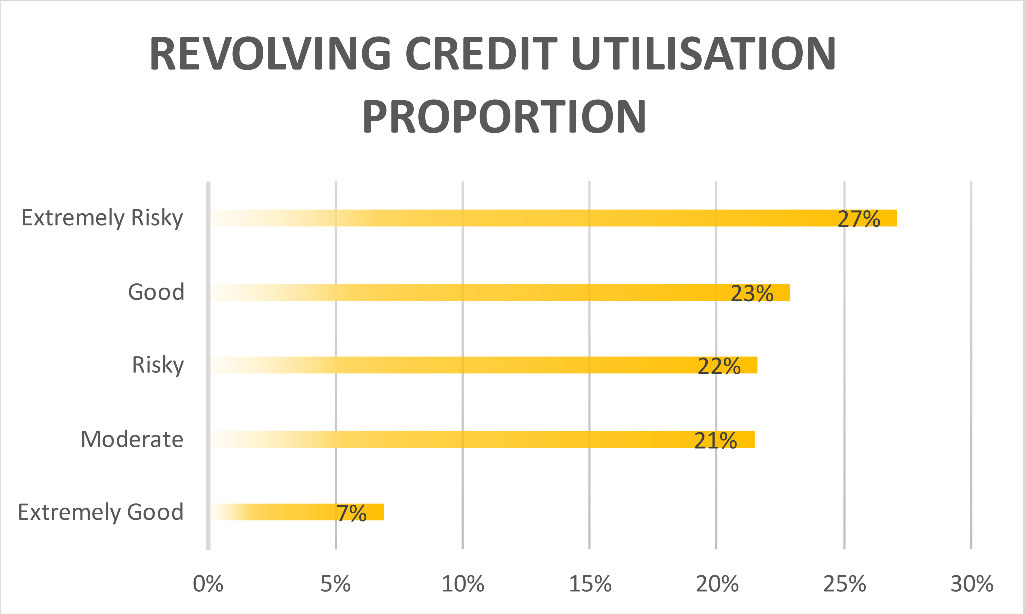
DATA PREPROCESSING

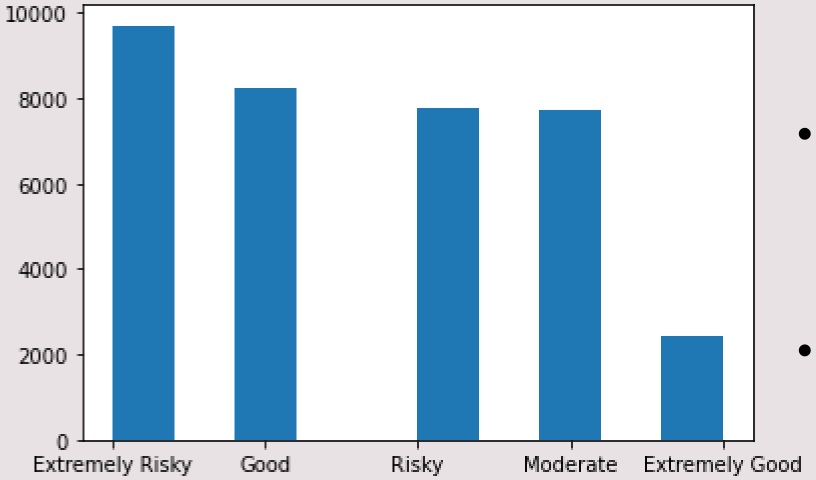
- Removed outliers from the column **“int_rate”.**



UNIVARIATE ANALYSIS

- Analysis on the **revol_util** column after segmenting into 5 categories reveals that LC has highest loan accounts having **“Extremely Risky”** revolving credit utilization

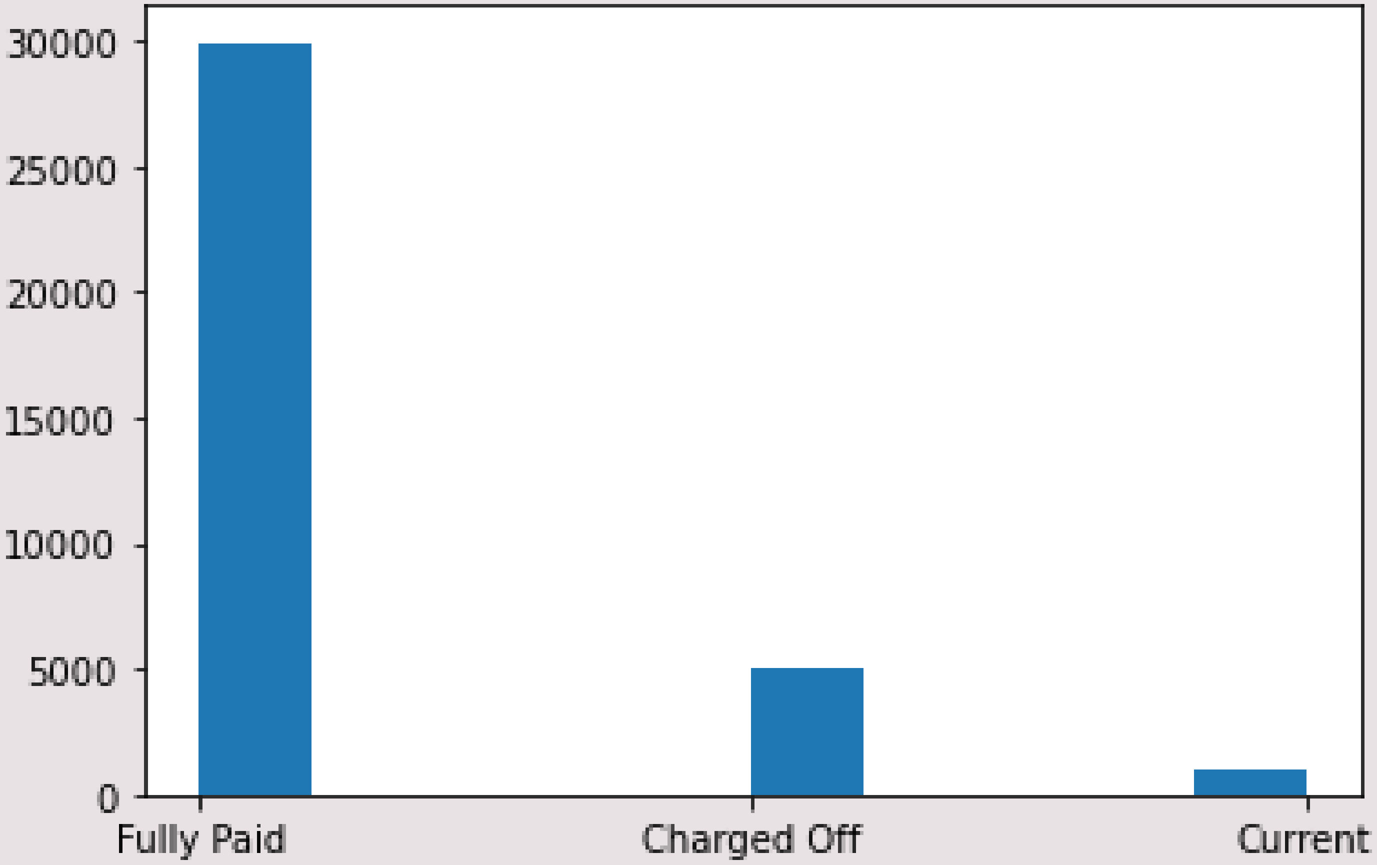


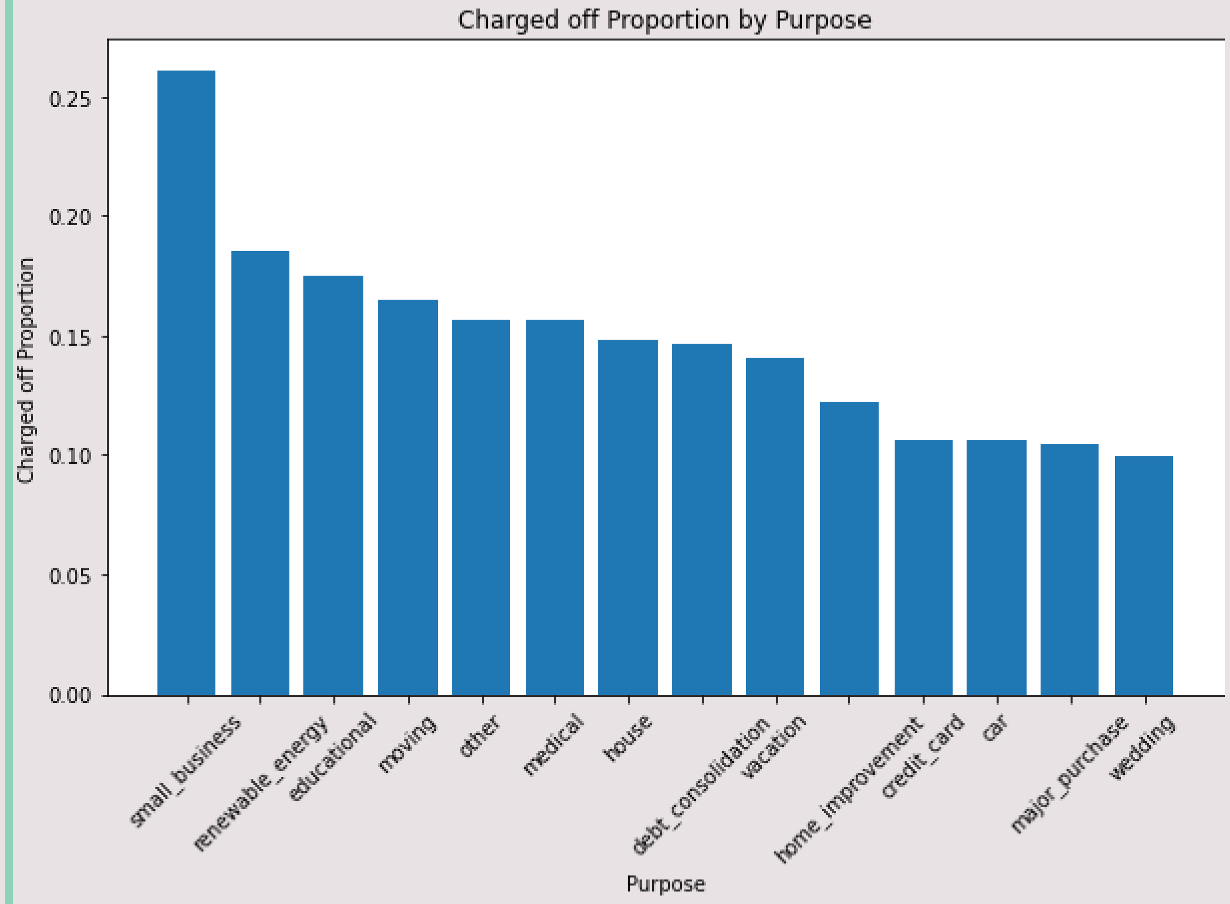


- The data set reveals that loan accounts falling into the categories, **Good, Risky and Moderate** have less significant difference (i.e., Credit utilisation between **5% to 70%**).
- The major borrower segment in need of a personal loan have a credit utilization between **30% to 70% .**

UNIVARIATE ANALYSIS

- The analysis on the loan_status column indicates that the current data set consists of borrower accounts have high counts of Fully Paid status and a significantly less Charged Off accounts.
- The data set is biased towards Fully Paid accounts, or it can be concluded that Lending Club accounts have borrows that always pay up the loan.



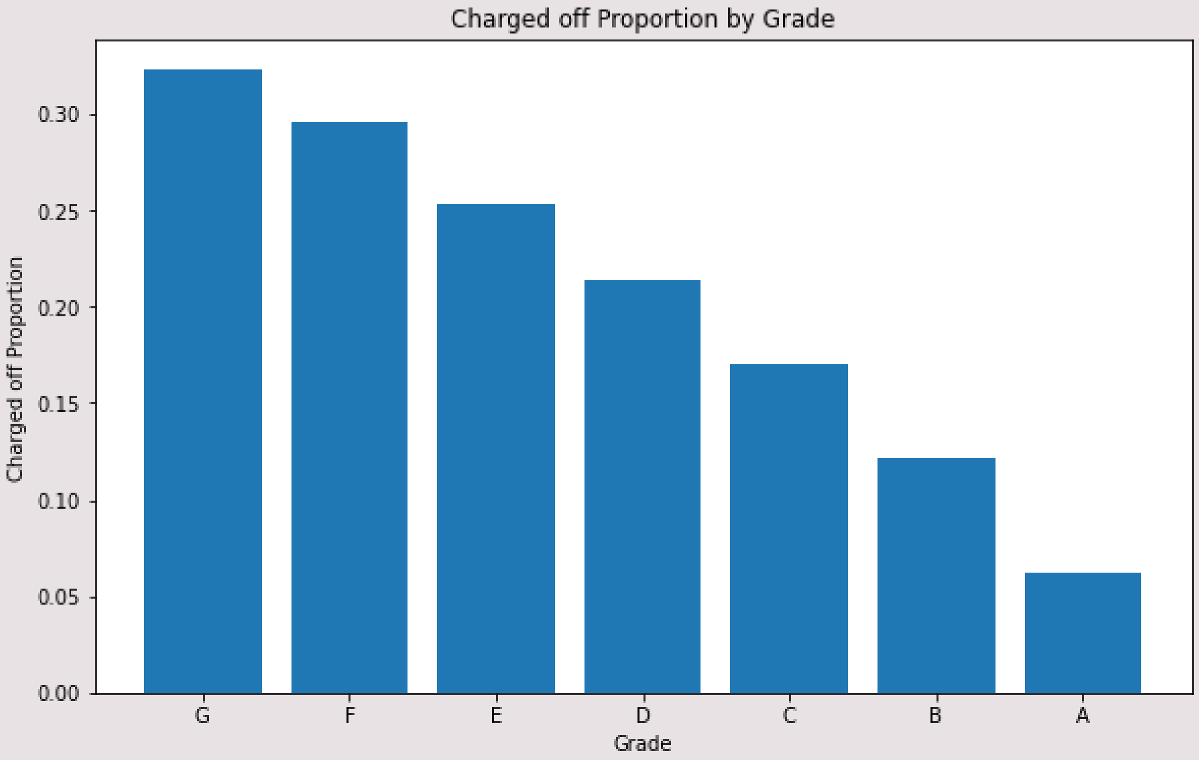


BIVARIATE ANALYSIS

The analysis of **loan_status vs purpose** has the following outcome.

- The loans taken for **"Small Business"** are the most to get defaulted.
- This indicates that small business do not survive in the market and have **high risk of failing** and be **Charged Off**.
- When compared to other sectors borrower for **"Renewable Energy"** is the second highest purpose to be a defaulter

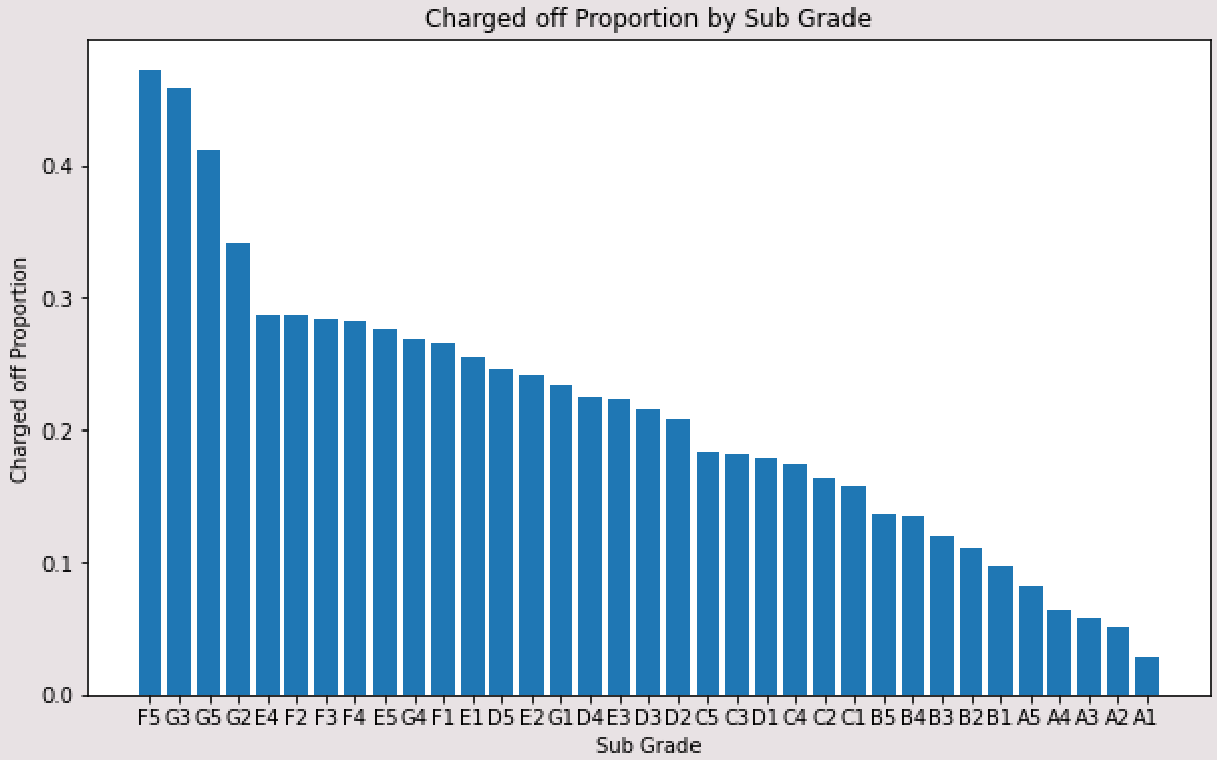
BIVARIATE ANALYSIS



The analysis of **loan_status vs grade** has the following outcome.

- Borrower's whose account has "**A**" grade are less likely to be a defaulter.
- The graph indicates that borrower with "**G**" and "**F**" are highly likely to default or be **Charged Off**.
- As we move from **A to G** the likelihood of a borrower increases significantly.

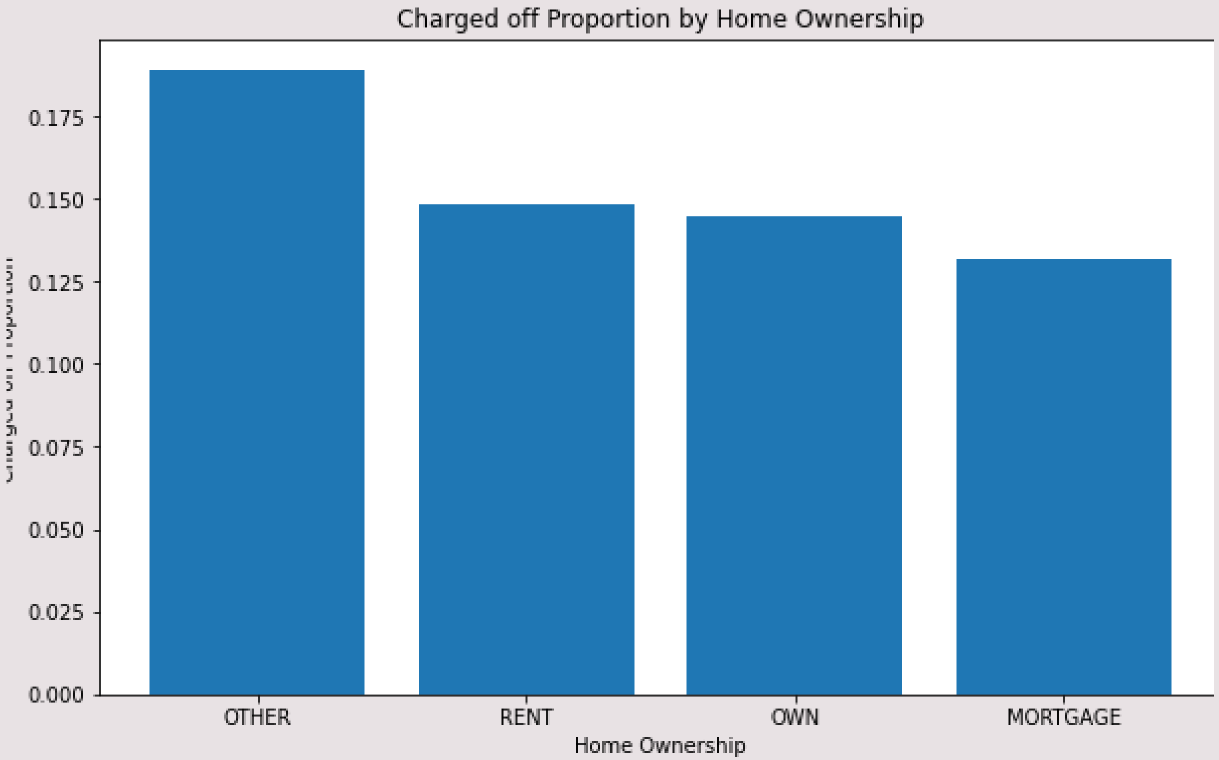
BIVARIATE ANALYSIS



The analysis of **loan_status vs sub_grade** has the following outcome.

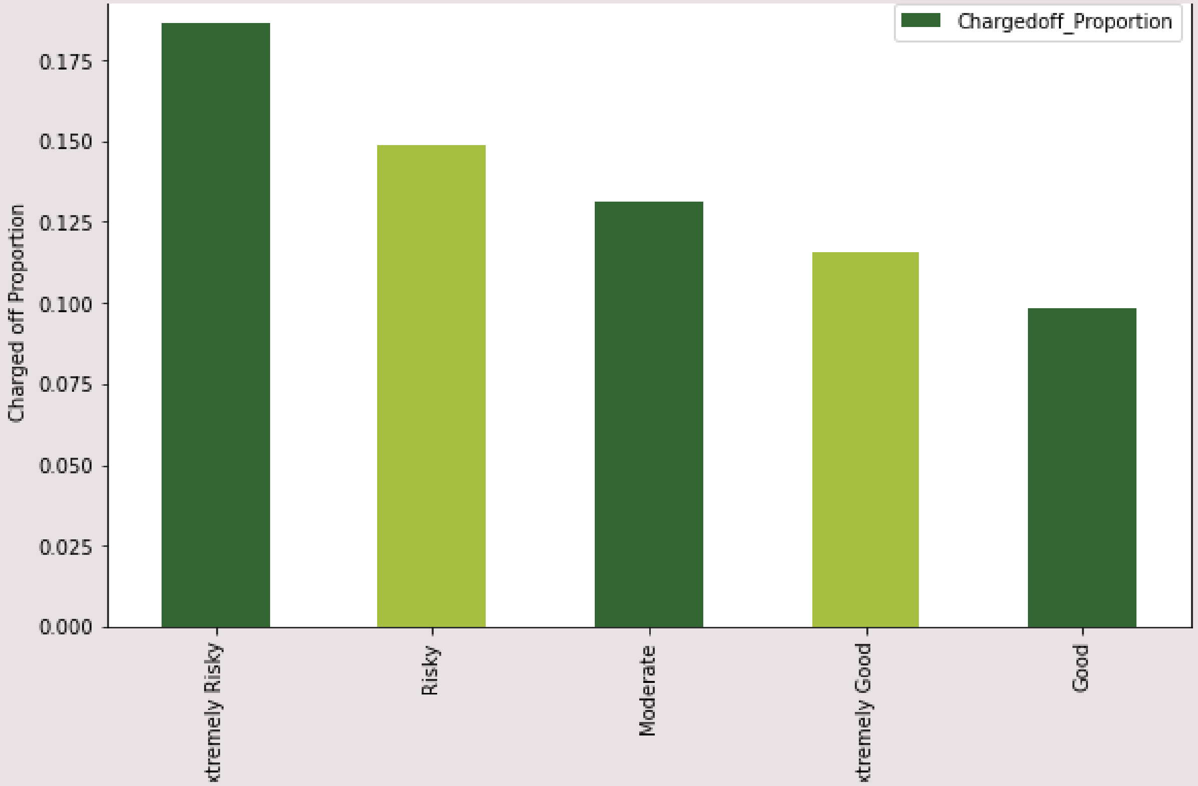
- Borrower's whose account has a sub grade of "**F**" and "**G**" are **Significantly** more likely to default or be **Charged Off.**
- The sub grade of "**A**" have very less likelihood of defaulting their loan.

BIVARIATE ANALYSIS



The analysis of **loan_status vs home_ownership** has the following outcome.

- Borrower who have stated their home ownership as **“OTHER”** are more likely to default their loan.
- This indicates that borrowers who do not state their home ownership are not financially stable to pay off their credit dues.

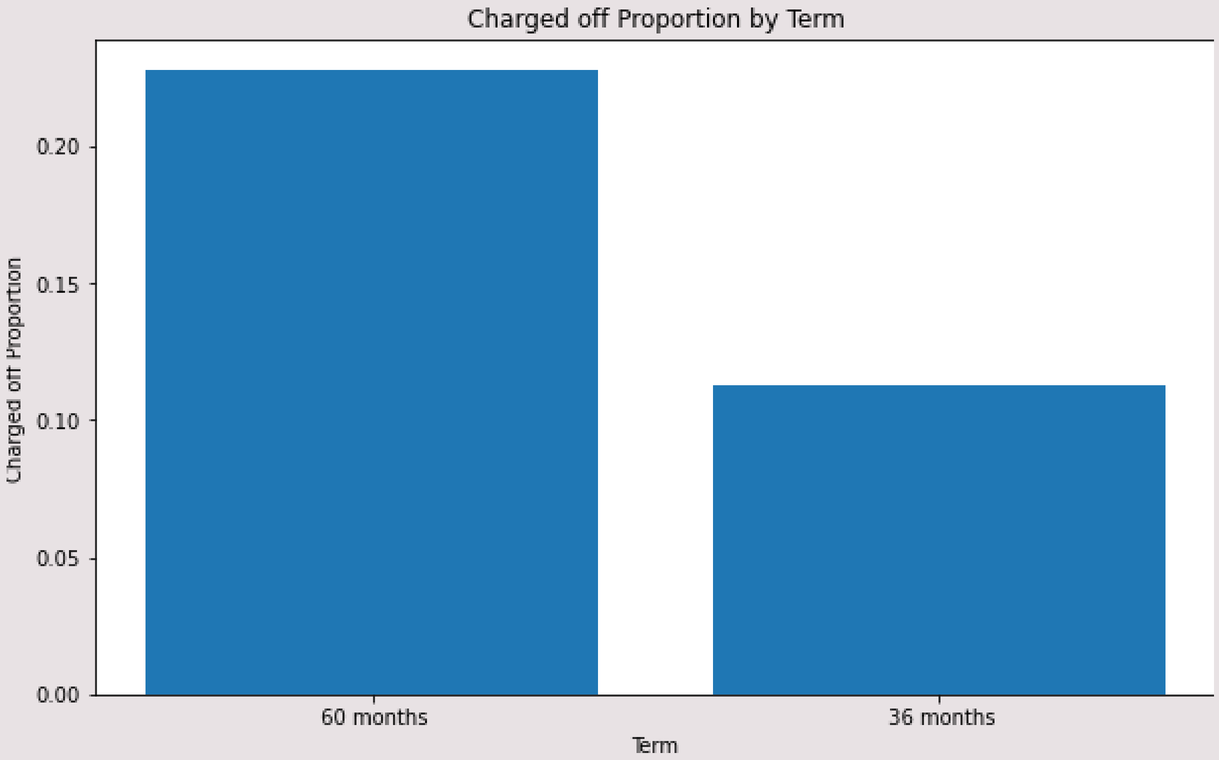


BIVARIATE ANALYSIS

The analysis of **loan_status vs revol_util_categories** has the following outcome.

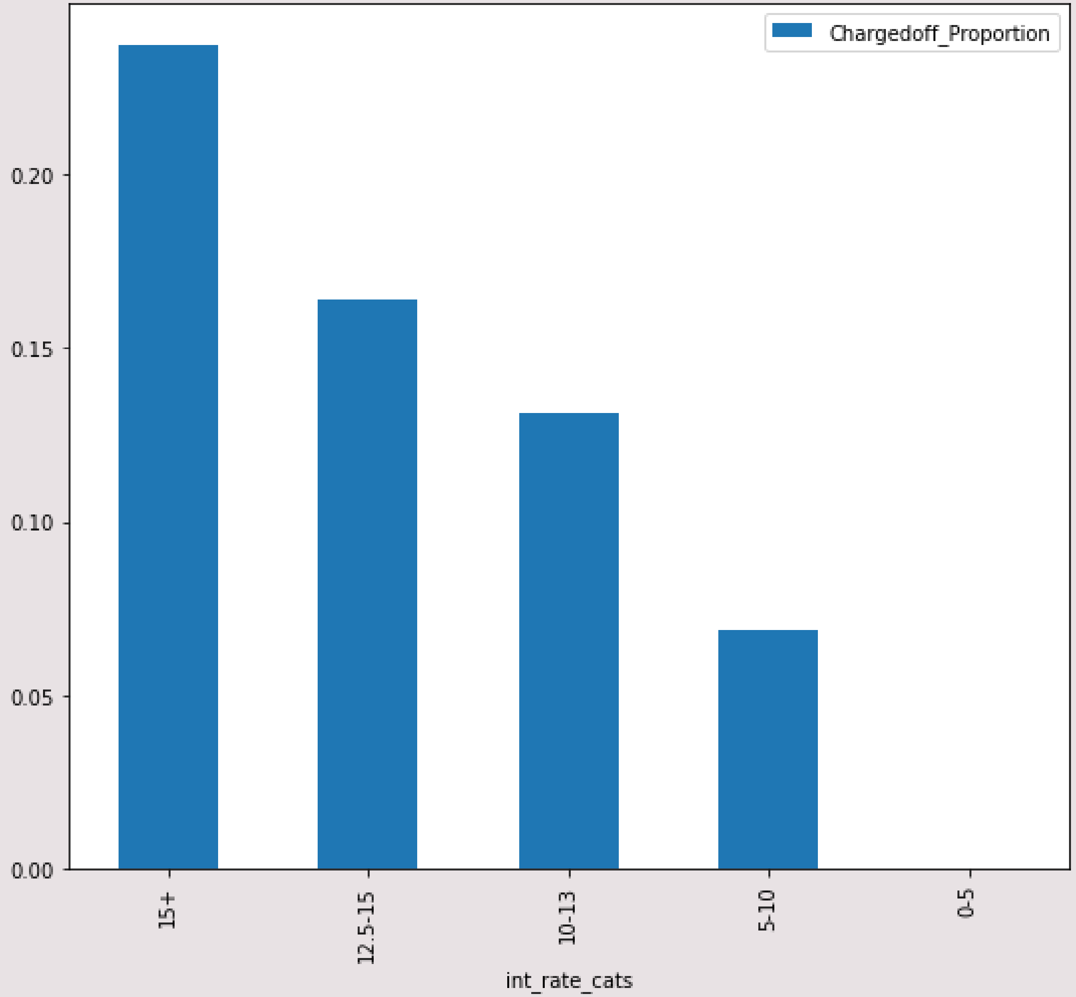
- Borrower's who fall under the **"Extremely Risky"** category are significantly more likely to default on their loan payment.
- The graph shows that those borrowers who fall in **"Extremely Good"** and **"Good"** category are considerably less likely to be defaulters

BIVARIATE ANALYSIS



The analysis of **loan_status vs term** has the following outcome.

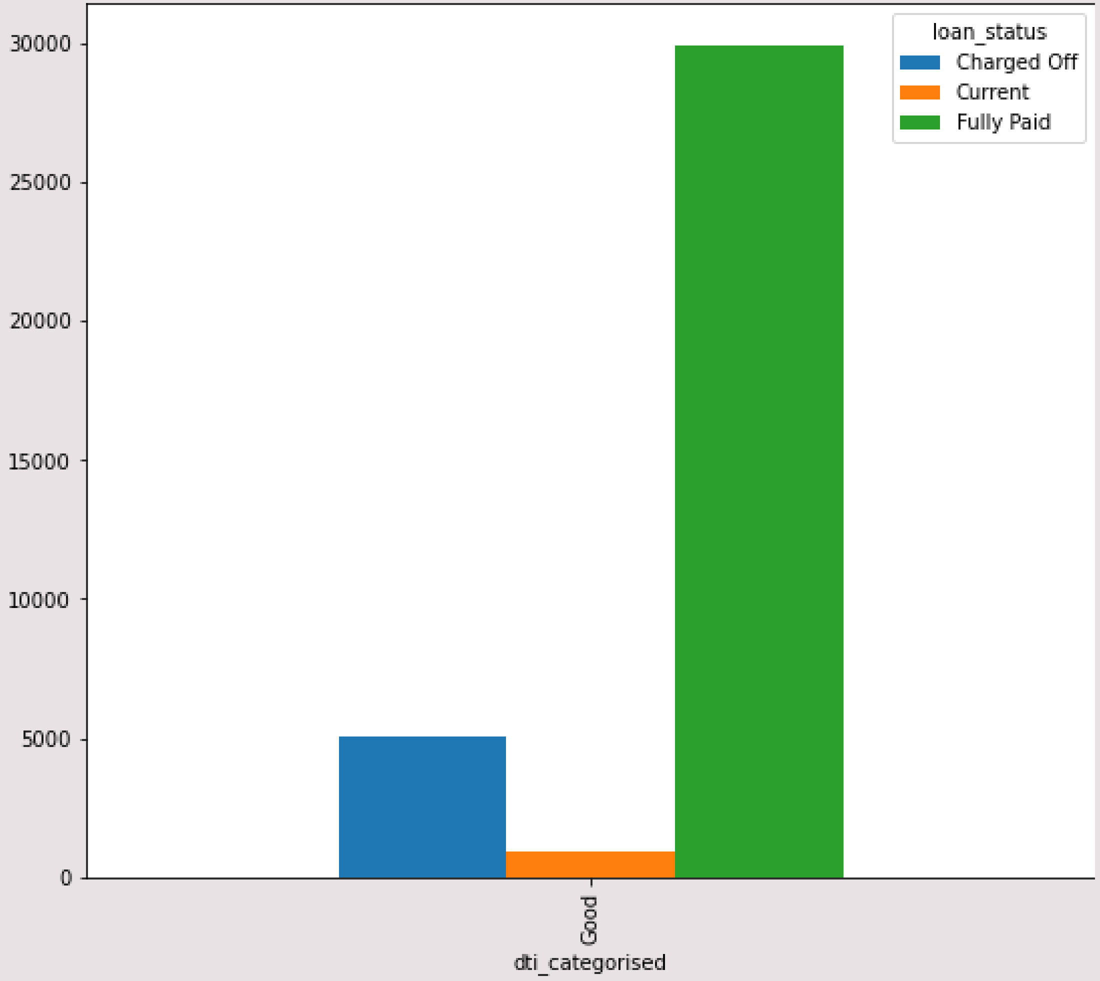
- Most borrower's who have a loan sanctioned for **60 Months** or more have higher likelihood to default on their loan or be **Charged Off**.



BIVARIATE ANALYSIS

The analysis of **loan_status vs int_rate** has the following outcome.

- Most borrower's that have defaulted had loans with interest rates **greater than 15%.**
- The loan accounts with interest rates between **5% to 10%** have less likelihood to default on their loans.



BIVARIATE ANALYSIS

The analysis of **loan_status vs dti** has the following outcome.

- As seen the graph, borrowers with **30% or less "Debt-to-Income"** are **more likely to fully pay** off their loans and less likely to be charged off

CONCLUSION

Borrowers are like to have a high probability to default their loan if:

- The borrower requests for a loan for **"Small Business"** and **"Renewable Energy"**
- The borrower has a grade of **"G"** and **"F"**
- The borrowers who state their house ownership as "OTHER" are less capable to pay off their loans. Indicating financial instability.
- Borrower having credit utilization between 50% to 70%, or 70%+ are the most likely to default. These borrowers can be given loans for a longer duration and lesser EMI's or reject the loan application if credit is over 70%.
- Borrowers with **"Extremely High"** revolving credit utilization and having grade **"F"** and house ownership as "OTHERS" and loan duration of 60 Months, with 15%+ interest rates are highly likely default on their loan and be Charged Off.