Lending Club Case Study: Assignment Submission

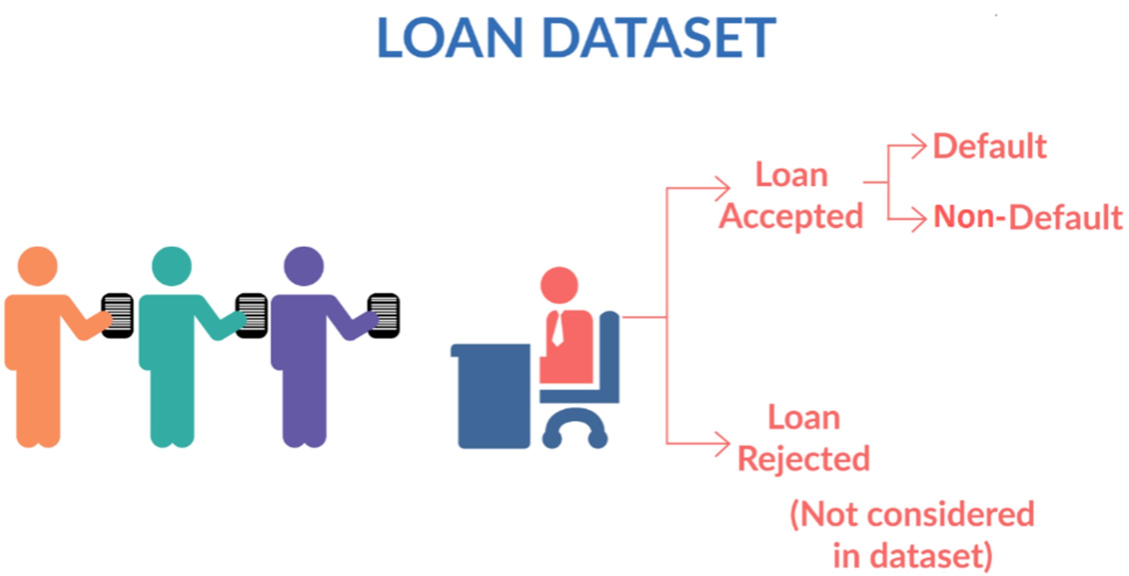
Sandeep Pathania/Chandan Kumar Singh

Vision and Business Objective

Vision: The online lending organization want to reduce the financial loss by identifying the Risky applicants and either rejecting the lending to such customer or providing loans on higher rates based on the risk analysis.

Business Objective: 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.

upGrad



Fully paid: Applicant has fully paid the loan (the principal and the interest rate)

Current: Applicant is in the process of paying the instalments, i.e. the tenure of the loan is not yet completed. These candidates are not labelled as 'defaulted'.

Charged-off: Applicant has not paid the instalments in due time for a long period of time, i.e. he/she has defaulted on the loan

Note: Fully paid and Current are not defaults. Charged-off are Loan defaulters.

Lending Club: EDA Case Study

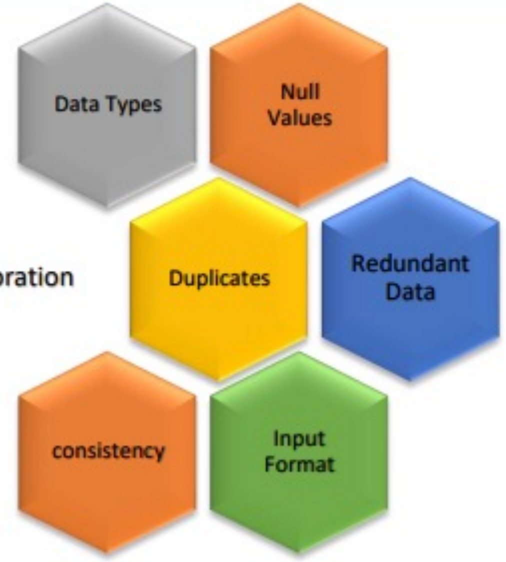
upGrad

Problem Solving Methodology

Input

- Loan Data
- Data Dictionary
- Mapping

Data Exploration



Data Formatting → Remove Duplicate Rows → Remove Duplicate Columns → Standardize Values → Convert Data Types → Filter Data

Data Analysis

Identify Unique Values	Loan Data Analysis	Loan Applicant Analysis	Identify Primary Identifier	Univariate \Bivariate Analysis	Plot Data
Load Data	Identify Outliers	Select Loan Variables	Split Category	Charged-off customers	Representative Variables
Identify Unique Values	Describe Data	Group by Experience	Merge Mapping	Experience wise defaulters	Most defaulters as per experience
Merge Data	Representative Value	Loan types	Map loan types to other variables	Top defaulters as per loan type	Loans defaults for debt_consolidation

Loan Data Analysis

Goal:

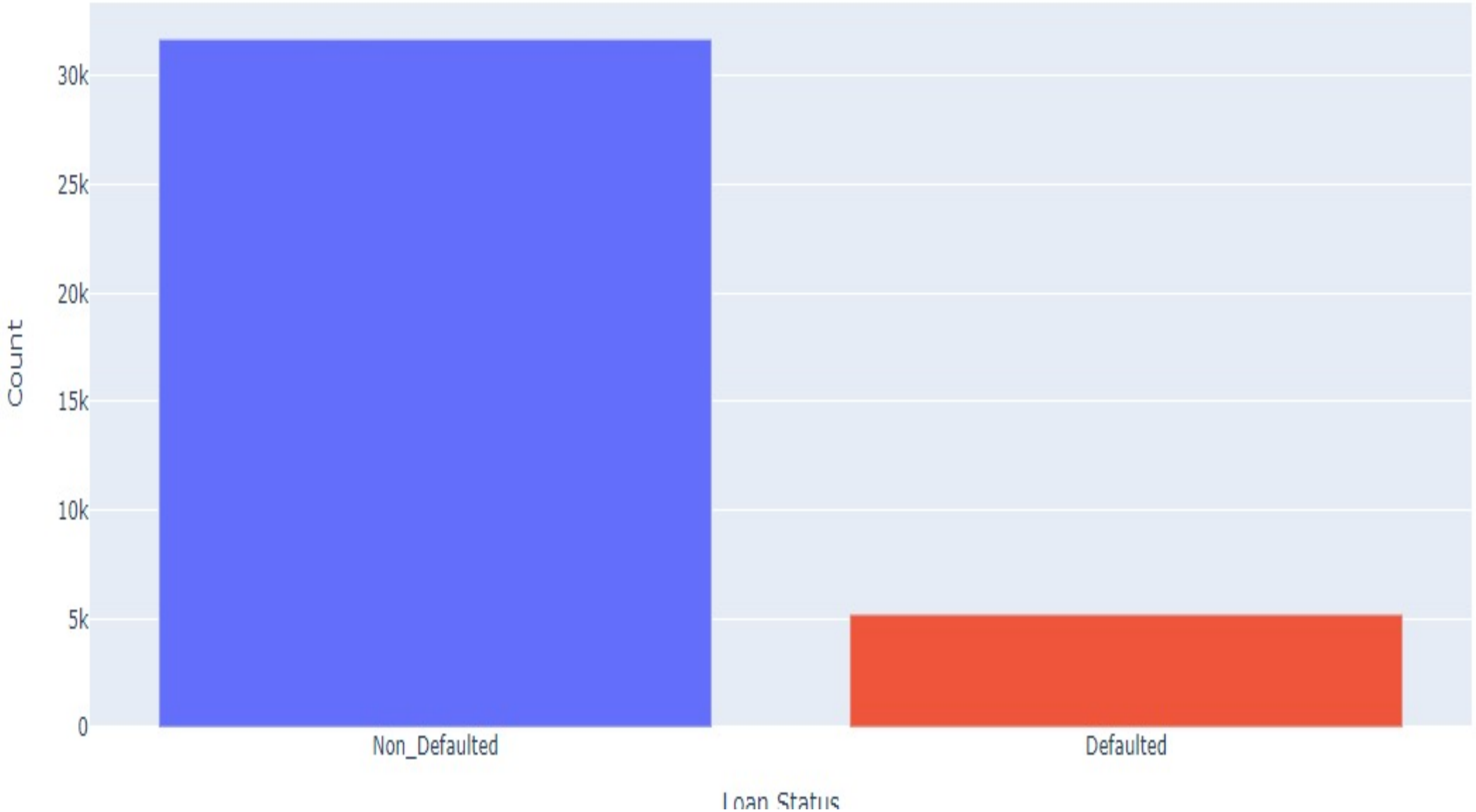
1. Identify Loan Defaulters statistics.
2. Characterize loan defaulters based on various variables (Grade, Employment Tenure, Housing etc.)
3. Focus on risk analysis based on different variables.

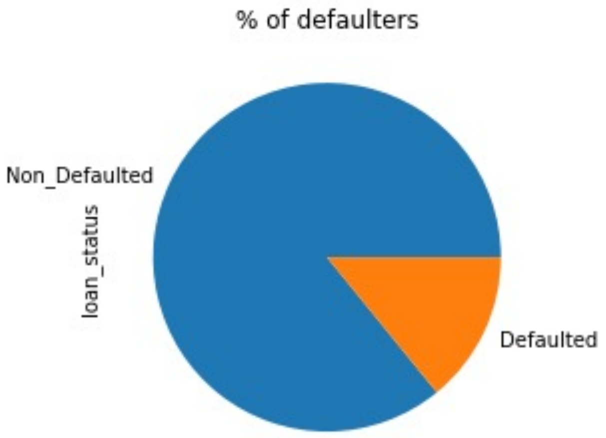
Analysis :

1. Calculated count & %age of loan defaulters under various categories and variables.
2. Pivoted data as per loan type, loan amount and home ownership.
3. Identified max values of loan defaulters under different categories.

upGrad

Defaulters vs Non-Defaulters





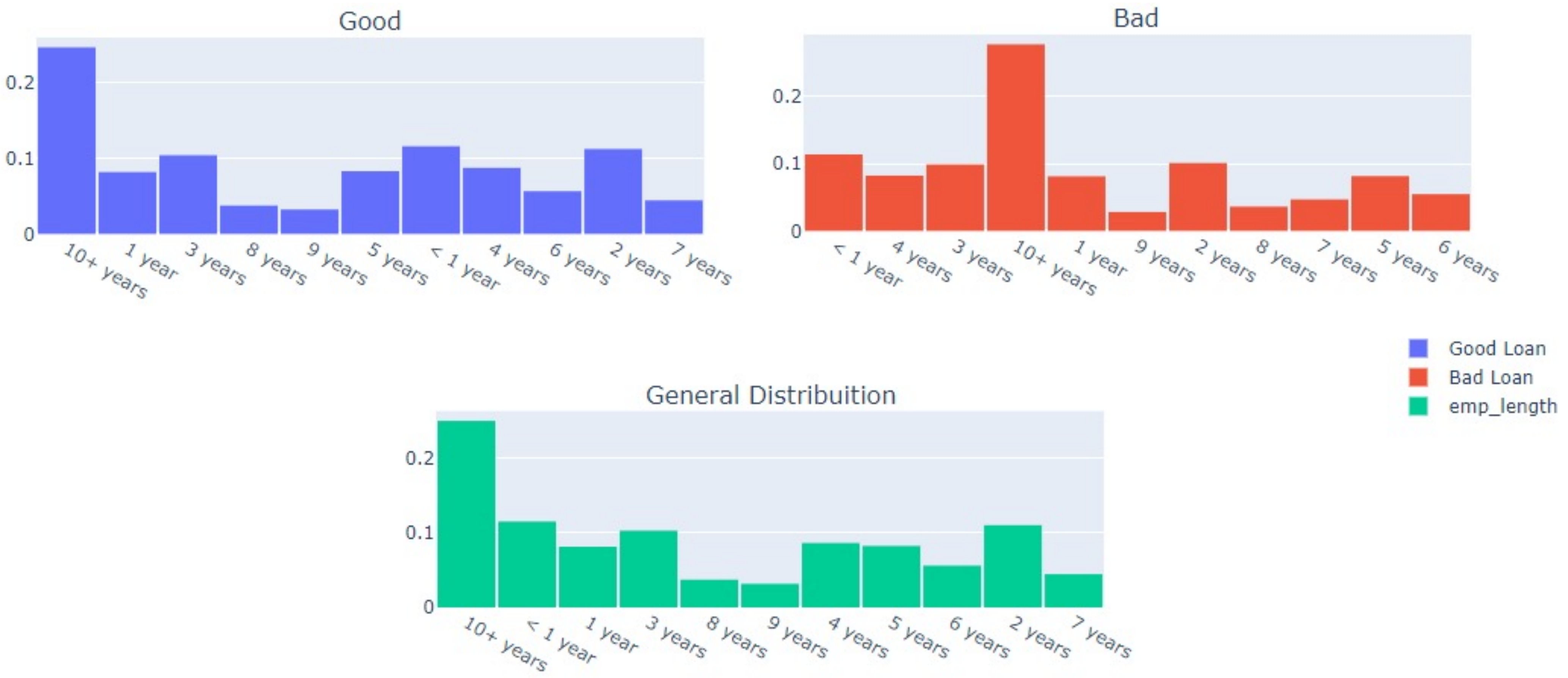
Distribution based on Loan Term: Loans with 36 months tenure have more defaulters.

Lending Club: EDA Case Study

upGrad

Distribution based on Grades: Grade B,D and C are topmost candidates of Bad loans . Out of this Grade C,D gives more Bad loans than Good Loans .

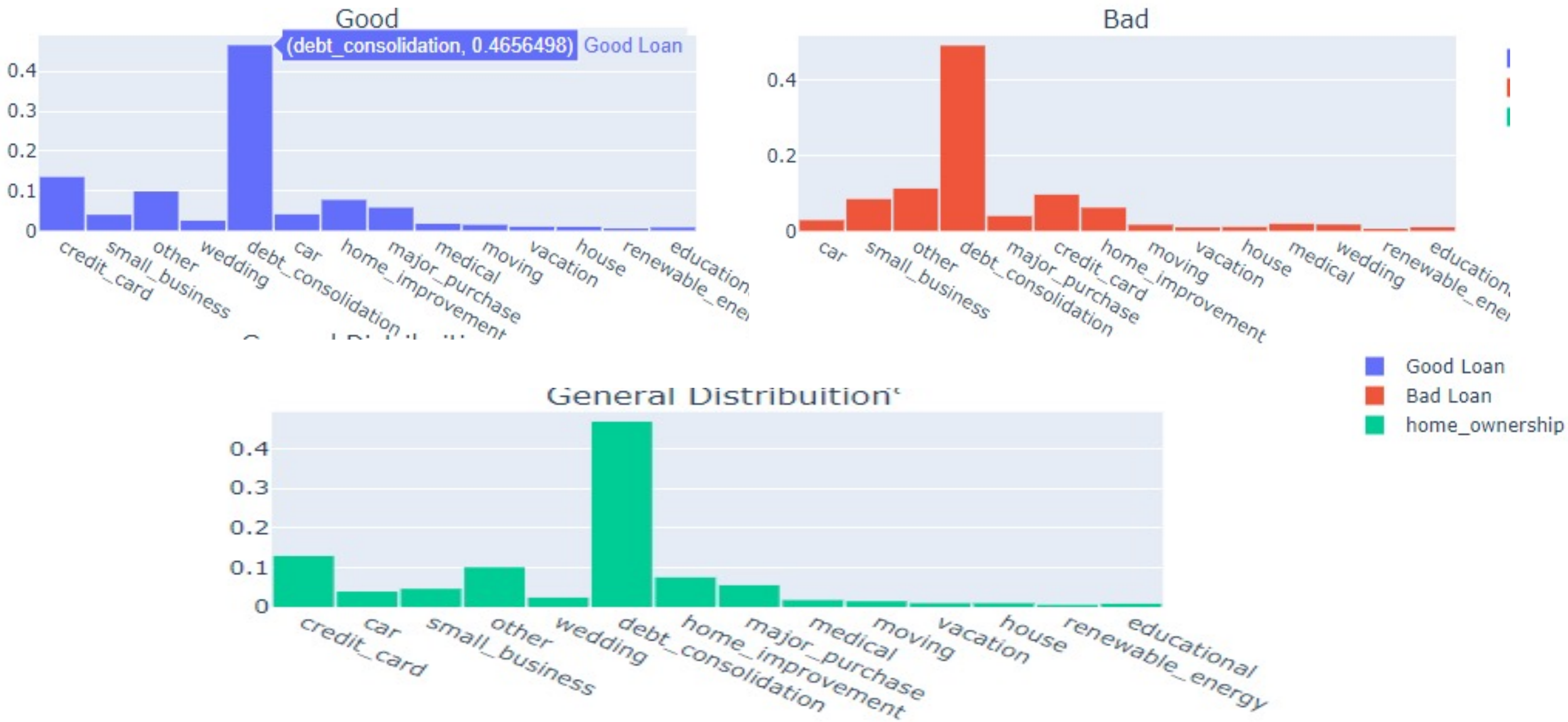
Distribution based on Experience/Employment length: Too much experience and freshers top bad loans . Employee with 10+ yrs , <=2 years top the bad loans



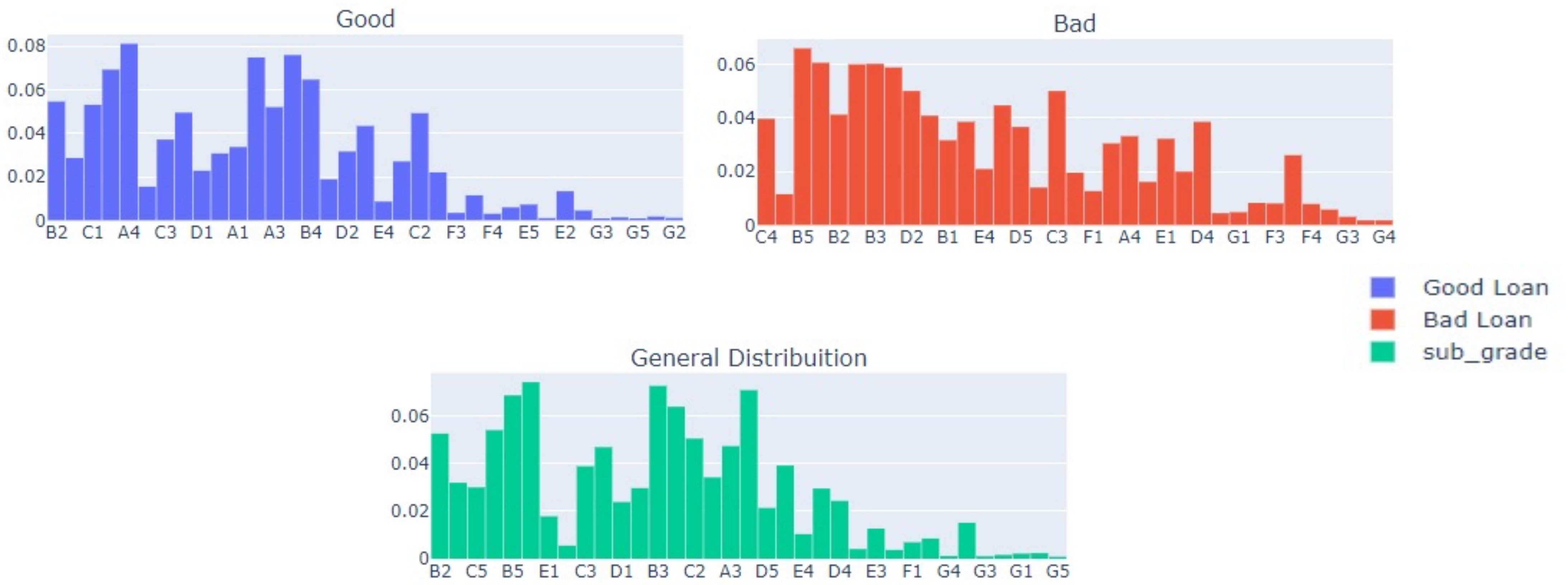
Lending Club: EDA Case Study

Distribution based on Loan Amount: High chances of Bad loans fall when loans granted is <10k.

Loans allocated for debt_consolidation peaks bad loans and House Ownership

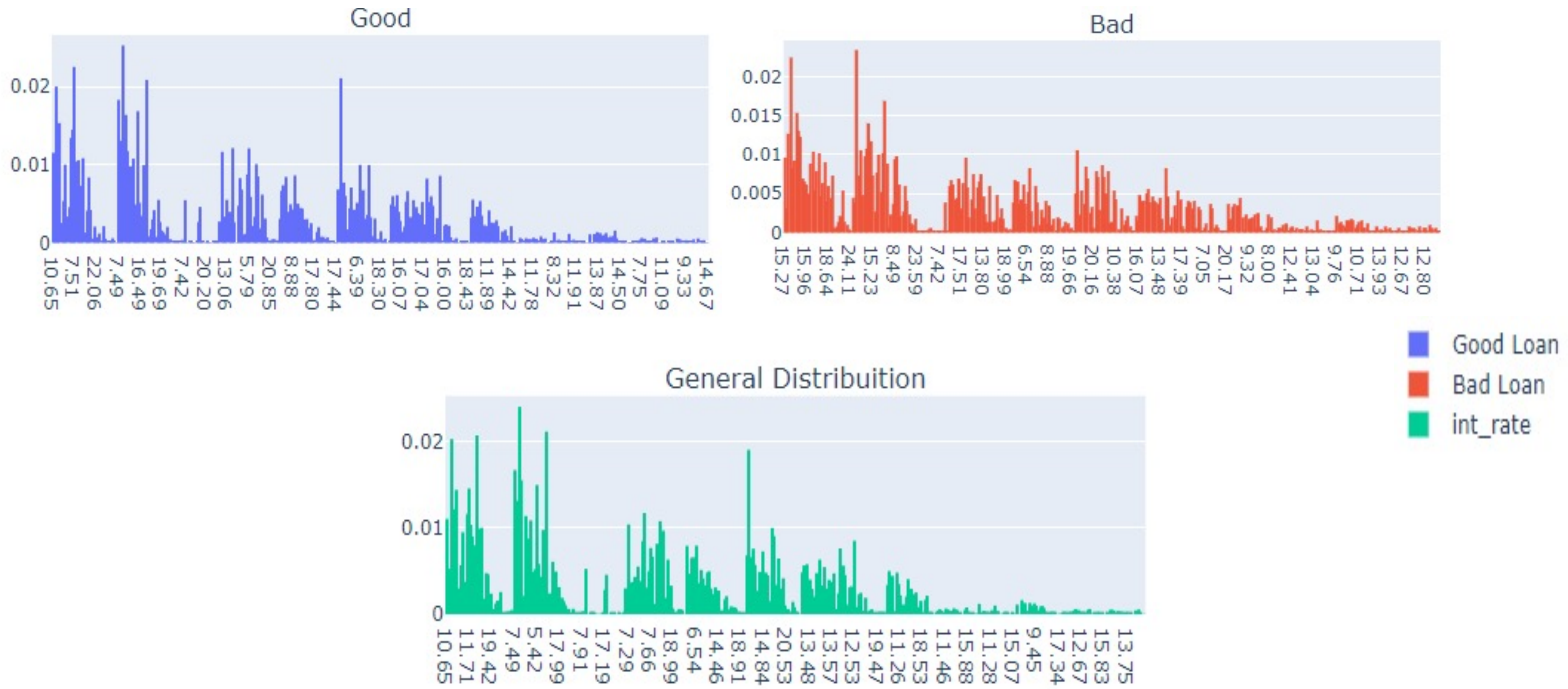


Distribution based on Sub-Grades



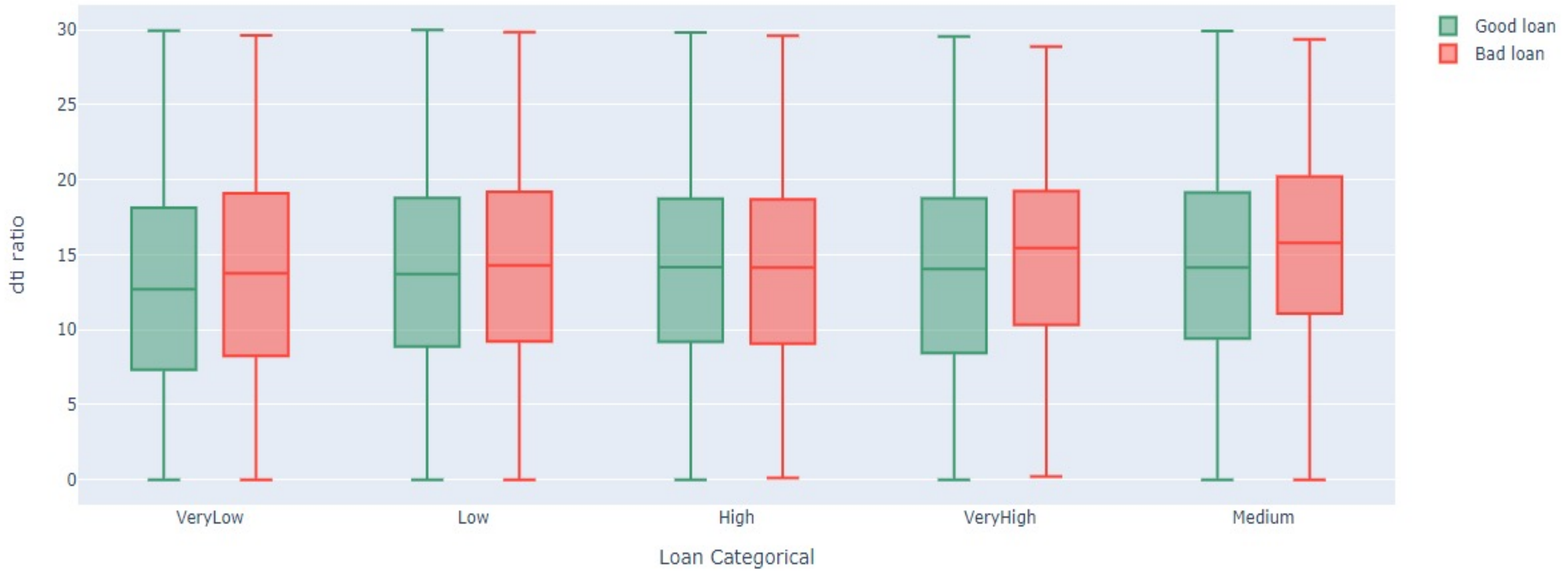
upGrad

Distribution based on Interest Rates

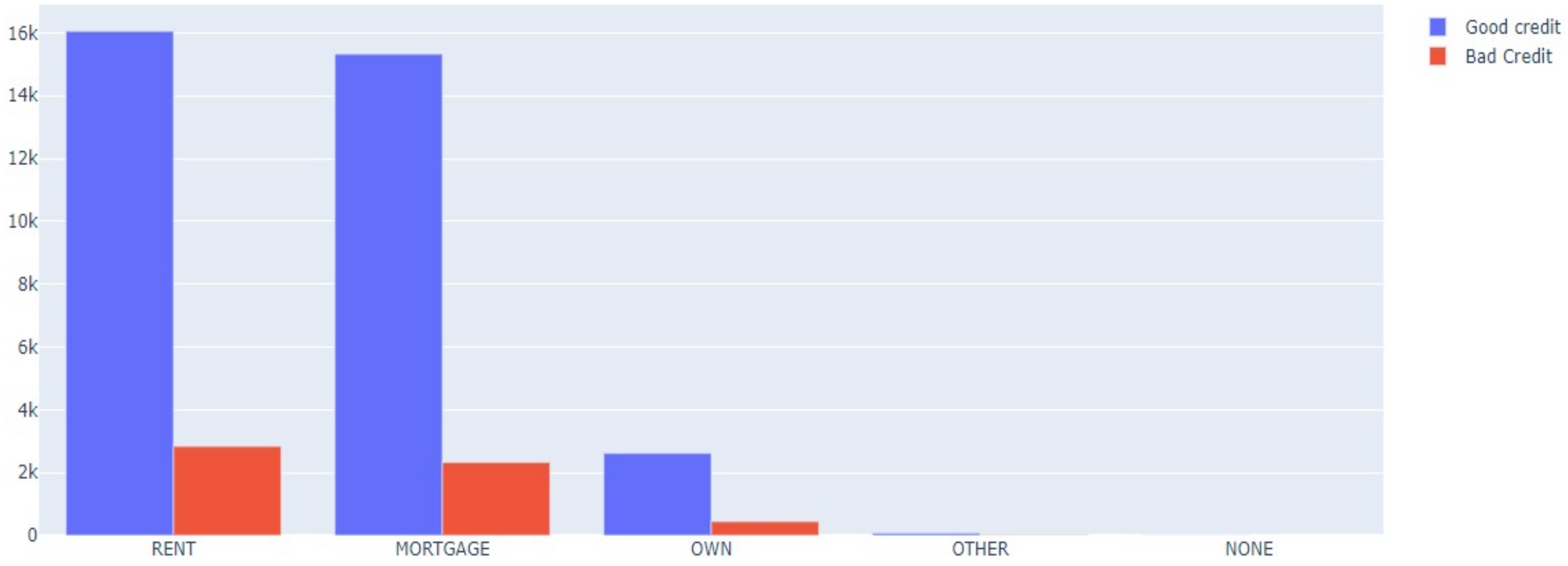


upGrad

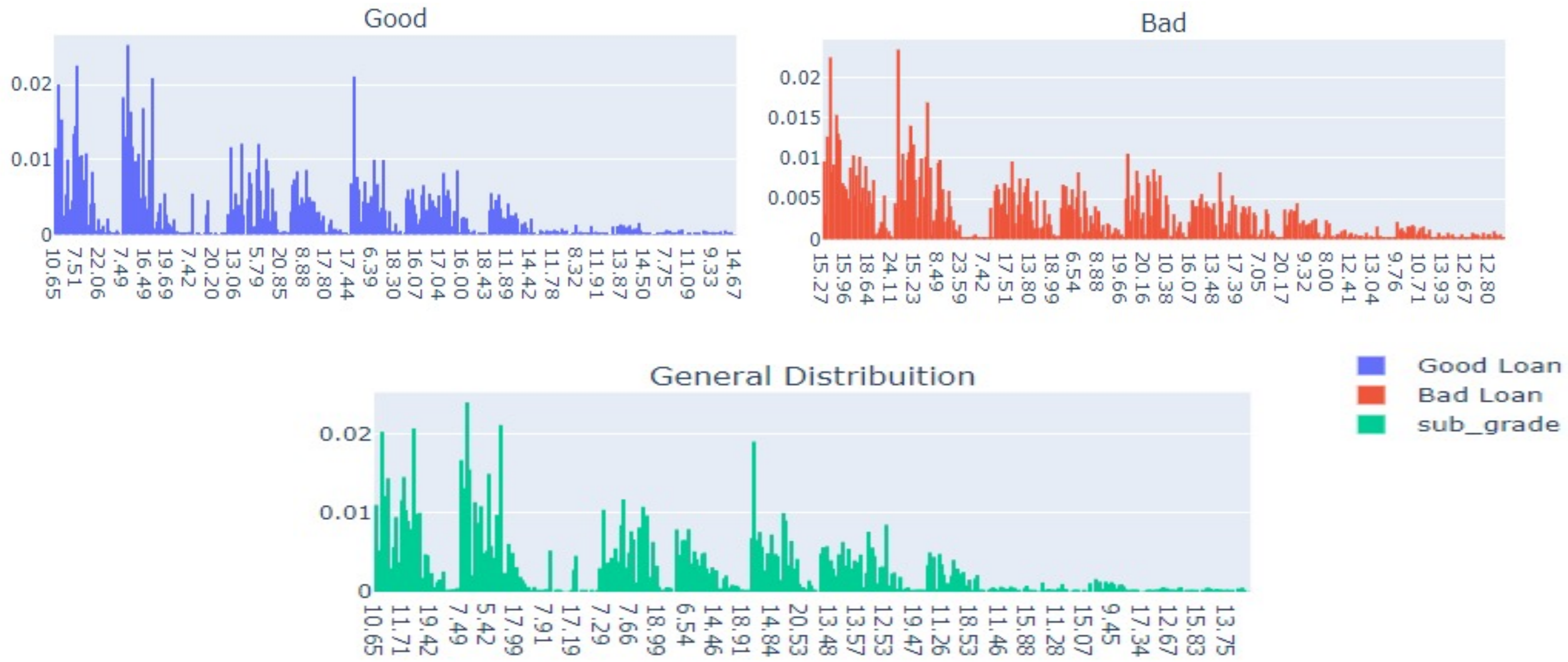
Very Low Loan amount [1000-10000] has most of the bad Loans



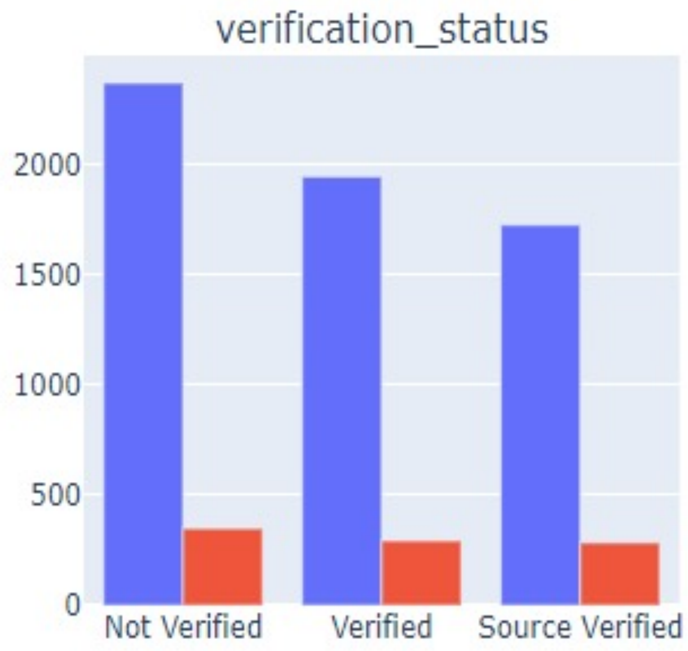
Distribution based on Housing: People having Rent and Mortgaged houses have max Bad loans

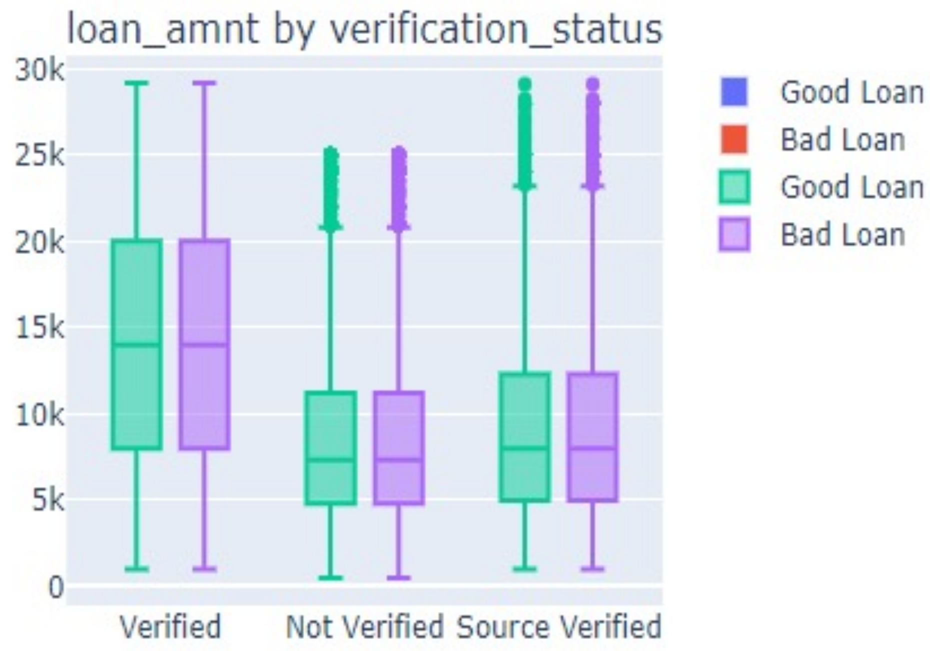


Distribution based on Interest: People having interest around [10-15]% have max defaulters , we see peak around 11



Distribution based Verification Status





Conclusion and Observations:

- Higher number of defaulters in case the loan term is 36 months. .
- Loans to applicants with Grades C and D makes losses as they give more bad loans to good loans . Among these , People with subgrades C[C1,C2] , B[B3,B4,B5],D[D2] have highest defaulters.
- Experienced and very new employees fail to repay the loans . Mostly they take loans for enjoyment purpose and end up not paying back. Reason may be health conditions for elder peoples and carelessness with younger ones.
- Loans where purpose is debt_consolidation has maximum number of defaulters.
- People having interest around [10-15]% have max defaulters. We observed the peak around11
- People having DTI ratio between [8-19] result in bad loans, more specifically with smaller loan.
- People having Rent and Mortgaged houses have max Bad loans
- During Year_end maximum number of loans turn bad, with highest in December followed by Nov , Oct
- People who apply for less loan amount and are not verified have higher chances to default. They might have thought it will be a loss to authority to search for defaulters with very less loan amount sanctioned
- Frequently low earning people take small loans .
- Grade C and D give more bad loans then good .

Recommendations

Cautious approach for below Scenarios:

- High Risk months needs extra cautiousness in granting loans , especially small loans [around 5k] in Dec. Year-end months is followed by year-start month [Jan]
- Lower Grant of loans to Newly joined employees who belong to C and B grades with tenure less than 2 yrs, if they apply for loans during above Risk months, have less salary and/or live on Rent .
- Same should apply for experienced employees and also for lesser loan amounts.
- In case, loan is approved, they should be given more time to pay the loan at slightly lower interest , if possible .

Increased Scrutiny for below scenarios:

- Increase scrutiny in allocating loans with - less amount loans for short term and applicants under grades C .
- This should be followed by grade B and D, People living on rent especially in above grade .
- People who belong to Grade- C,B and D and are living on Rent should not be given short team loans for 36 months during High-Risk months with an exception if they have “Verified_Status”.