Lending Club Case Study Analysis

Submitted by:
Amandeep Singh Mallan

Date: 04-01-2023

Problem Statement:-

- For a largest online loan marketplace, which specialises in lending various types of loans to urban customers at lower interest rate through a fast online interface, analysis is required to understand the driving factors behind loan default.

- The data provided contains information about past loan applicants and whether they ‘defaulted’ or not. Considering that the analysis is done.

Approach Taken:-

- Data Sanity Checks were done first to understand the data and see if there are any missing values. The columns where missing values were high, were dropped.

- Missing Value Imputation was done to do the replacement of missing values for the important columns required for further analysis.

- Univariate Analysis carried out on cleaned columns in a segment manner for charged off & fully paid applicants to understand the distribution of various categorical and continuous columns. Accordingly outliers were removed & insights were derived

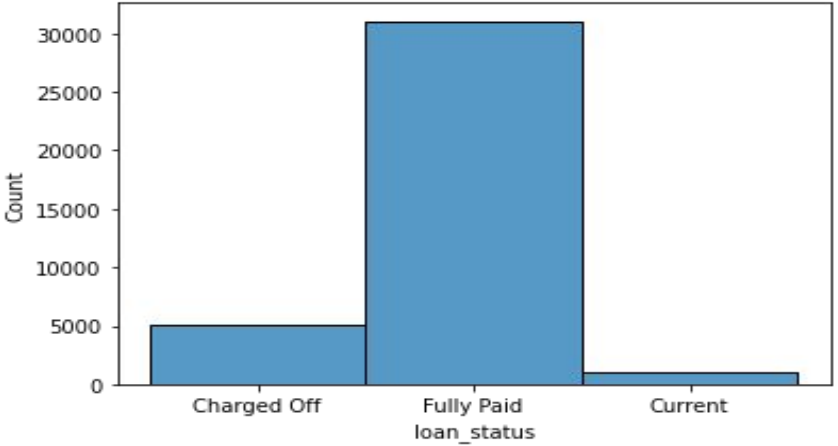
- Bivariate Analysis was carried out between continuous columns & variations in continuous column values across categorical column were observed in a segment manner.

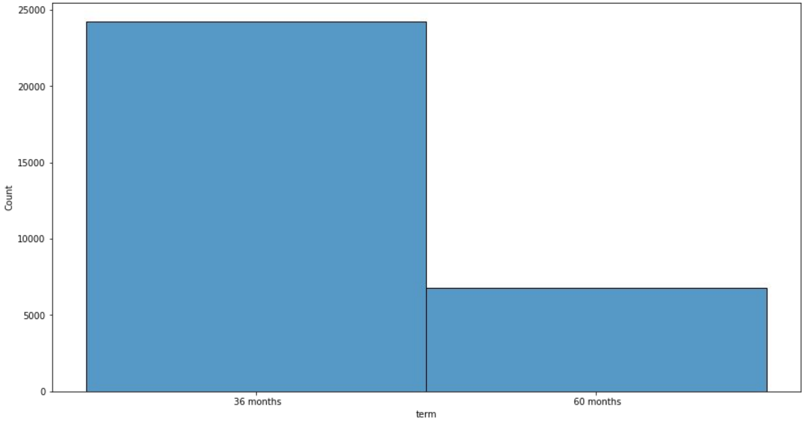
- Considering the findings of Univariate & Bi-variate Analysis, suggestions were derived for the problem statement.

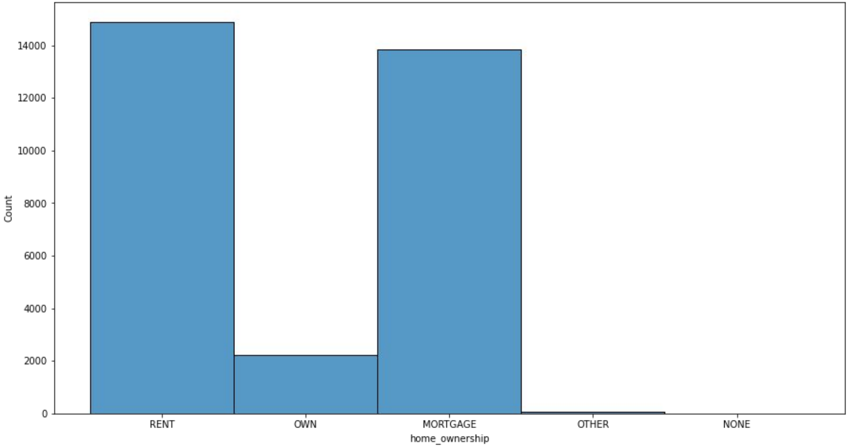
Findings from Univariate Analysis-

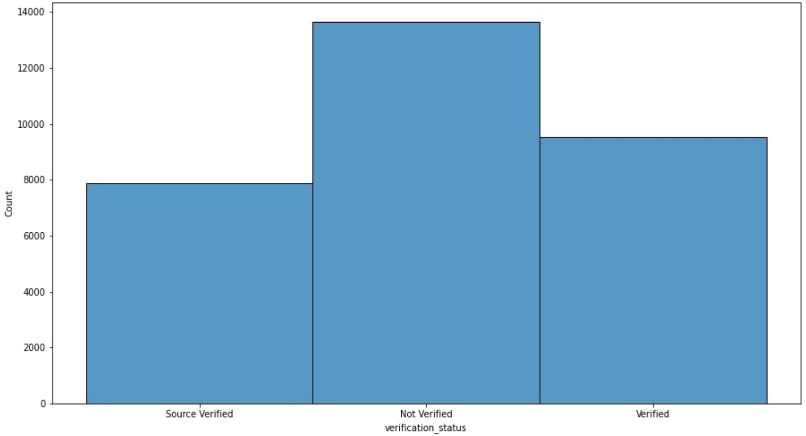
- The term of Loan is mostly 36 months & most of them had the employment history of 10+ years
- Majority of them live on Rent or have Mortgage and the income verification status was not verified for most of them.
- A good difference between the average monthly income of the applicants that were charged off and the ones that had fully paid the loan amount was observed.
- The mean dti ratio value, which is the ratio of 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, is fairly high for the charged off applicants as compared to the applicants that had fully paid the loan amount. This can also be one of the reasons why the applicants would have found it difficult to pay off their loans.
- The median value for Total Credit Revolving balance is fairly high for applicants that had charged off as compared to the applicants that didn't defaulted.

Supporting Visuals for Univariate Analysis-





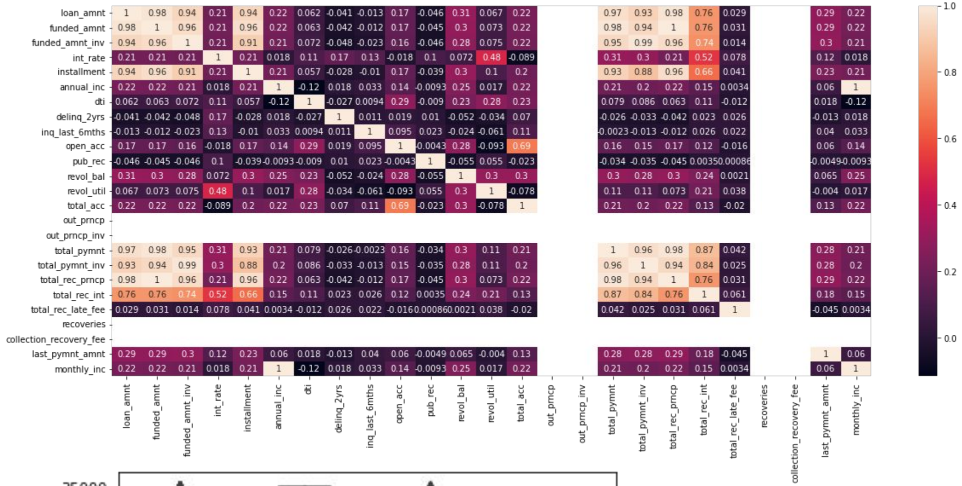


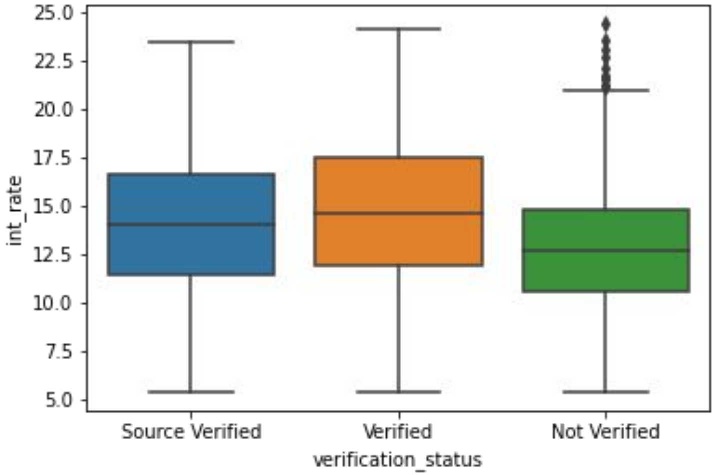


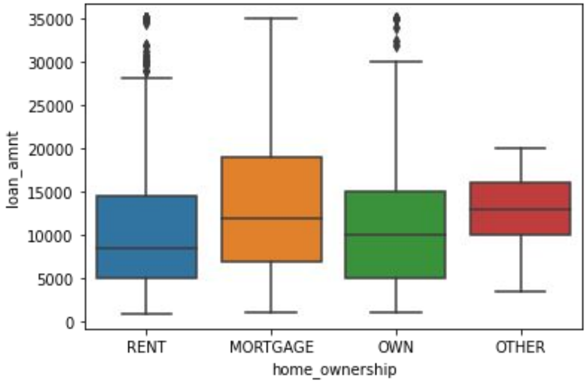
Findings from Bi-Variate Analysis-

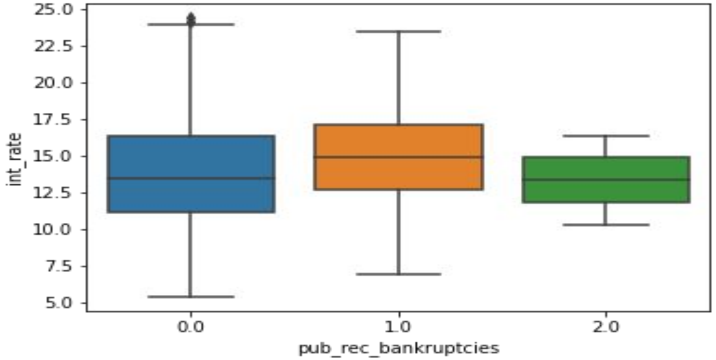
1. Strong positive correlation was observed between the loan amount parameter and the funded amount & funded amount investment parameters for the applicants that had fully paid the amount.
2. Strong positive correlation was observed between total recurring principle which represents the principal received to date and the funded & loan amount for the applicants that had fully paid the amount.
3. A weak negative correlation was observed between annual income and dti ratio for charged off applicants.
4. The loan amount, funded amount, funded amount by investors, interest rate, installment, dti ratio tends to be higher for higher term (60 months) loans.
5. For Grade F Loans, the loan amount is the highest whereas for Grade A Loans, it is the least. Moreover, for Grade G loans the interest rate is seen to be the highest whereas for Grade A Loans it is the lowest
6. The loan amount of applicants who have home ownership as Mortgage is higher as compared to other categories
7. The loan amount for the applicants whose verification status is verified is higher as compared to the applicants who are not verified. Moreover, the interest rate & installment for Not verified applicants is less
8. Even the applicants for whom the Number of public record bankruptcies is 2, for them the interest rate is fairly less.

Supporting Visuals for Bivariate Analysis-









Recommendations to curtail the risk of credit default

- For the applicants for which the number of public record bankruptcies is 2 or more, the interest rate of the loan should be kept as high to limit the loss from the risk of defaulting, as it was observed that for charged off applicants, the interest rate was less whereas for the applicants who fully paid the loans, it was high.
- The mean dti ratio value was less for the applicants who fully paid the loan as compared to the ones who didn't. Hence, to limit the risk of issuing the loans to the defaulters, the bank can benchmark it against the mean value of the dti for applicants who fully paid the loan and accordingly issue the loans to the applicants who meet that benchmark value.
- The interest rate for charged off applicants whose income verification status is not verified is fairly less as compared to the verified ones. The bank can increase the interest rate for the applicants whose income verification status is not verified to curtail the risks of loss.
- The median value for Total Credit Revolving balance was observed to be fairly high for applicants that had charged off as compared to the applicants that didn't. Hence, to limit the risk, the bank can benchmark it against the median value of the Total Credit Revolving balance for applicants who fully paid the loan and accordingly issue the loans to the applicants who meet that benchmark value.
- It was noticed that there is a good difference between the average monthly income of the applicants that charged off and the ones that had fully paid the loan amount. The bank can benchmark this parameter value against the mean monthly income value of applicants who fully paid the loan and accordingly issue the loans to the applicants who meet that benchmark value.

THANK YOU