

EDA:Lending Case Study

Shweta Jain

10th May 2023

We want to identify the predictors of default so that at the time of loan application, we can use those variables for approval/rejection of the loan

Data understanding

- I looked at the overall structure of the data to understand the number of rows (observations) and columns (variables).
- I tabulated each variable to understand the whether the variable is measurable variable or categorical variable.
- I identified the invalid and missing values in each variable

- I examined whether there are duplicate values or not in the identity variable.
- I coded the missing values as a separate category in each variable.
- I converted the value labels of categorical variable into a discrete variable either in the form of binary variable or multi-category variable.

Data Analysis

- I plotted the histogram of the relevant variables.
- I also examined summary statistics (mean, standard deviation, minimum, maximum, three quartiles) of the continuous variables.
- I also plotted the box plot of the continuous variables to identify the outliers in the dataset.
- I also plotted the kernel density graphs to examine if distribution is skewed or not.
- We ran the bivariate regression to assess the association between the two variables.

Recommendations

- Based on the bivariate regressions, we identified the following variables as the statistically significant determinants of the likelihood of default: loan amount, funded amount, invested funded amount, duration of the loan amount, and interest rate.
- The key determinants of the default were :
- We find that the following variables statistically reduces the likelihood of default: loan amount, funded amount, invested funded amount, duration of the loan amount, and interest rate.
- We find that the following variables statistically increases the likelihood of default: earliest credit line, delinq in 2yrs, total payment, collection of recovery fee
- In order to reduce the likelihood of the default duration of loan amount should be reduced.
- Bigger amount should be provided at bigger interest rate.