BFSI - Credit Risk Assignment

- By:
Arnab Basu
Rohit Mishra
Rohan Saboo

BFSI–CREDIT CARD ASSIGNMENT

OBJECTIVE:

The objective is to build a statistical model to estimate borrowers' Loss
Given Default (LGD)

$$LGD = \frac{Loan\ Amount - (Collateral\ value + Sum\ of\ Repayments)}{Loan\_Amount}$$

BACKGROUND

Credit risk analytics in the context of the banking sector and model a common metric used for estimating the expected credit loss (ECL)

ECL method is used for provisioning the capital buffer to protect banks against possible default of the customers.

Expected credit loss = Exposure at default x

Probability of Default x

•Loss given default

The loss given default (LGD) is a measure of the amount of loss that a bank is expected to incur in the event of a default by a borrower.

DATA SOURCES

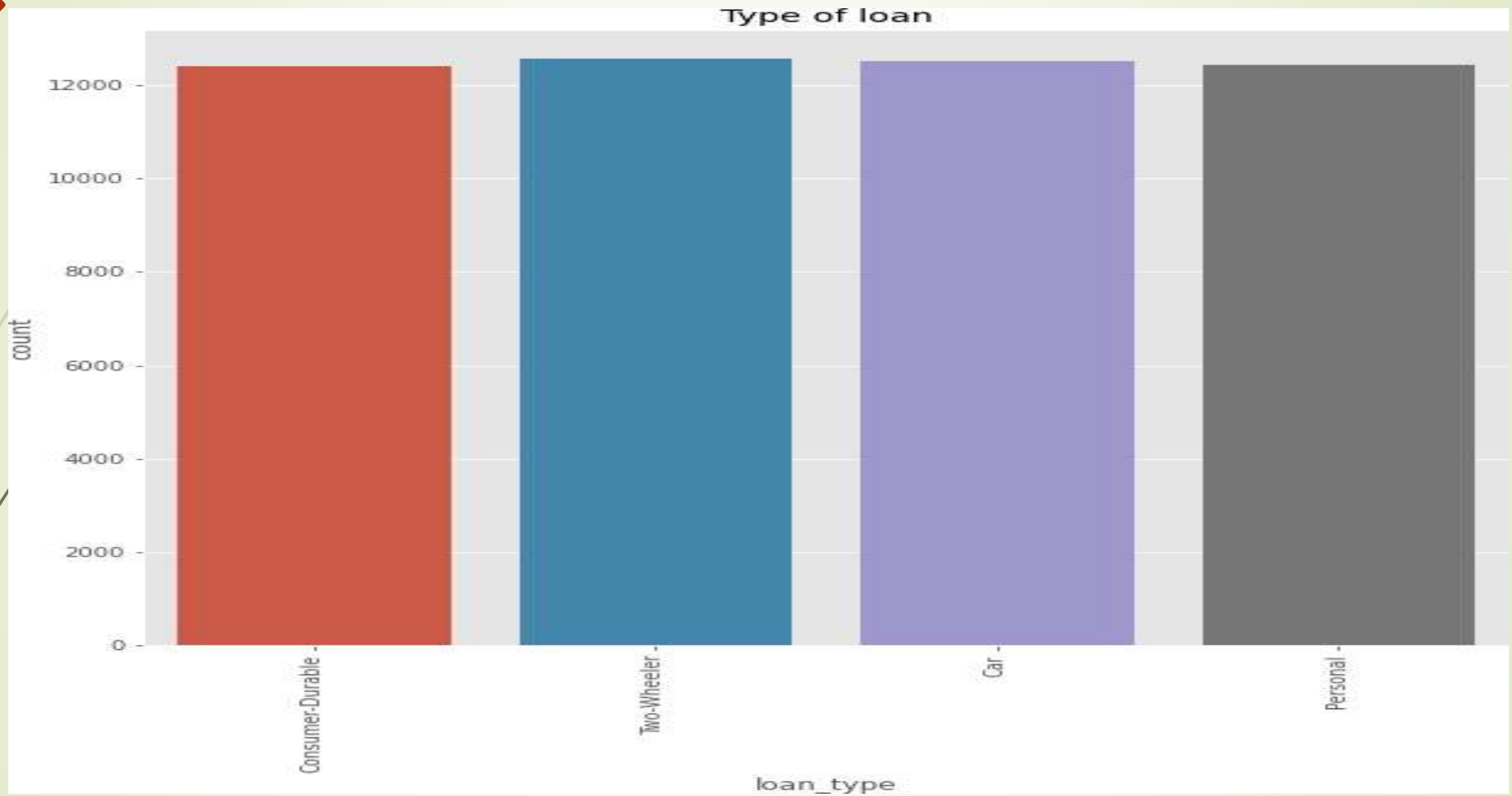
Used 3 Data sets for model Building:

- The main_loan_base data set contains information about loan accounts and other relevant information for the corresponding borrowers.
- The repayment base data set contains information about the repayments received by the banks in the form of EMIs or through other collection efforts
- The monthly_balance_base contains the information pertaining to the monthly balance statements in the borrower’s accounts

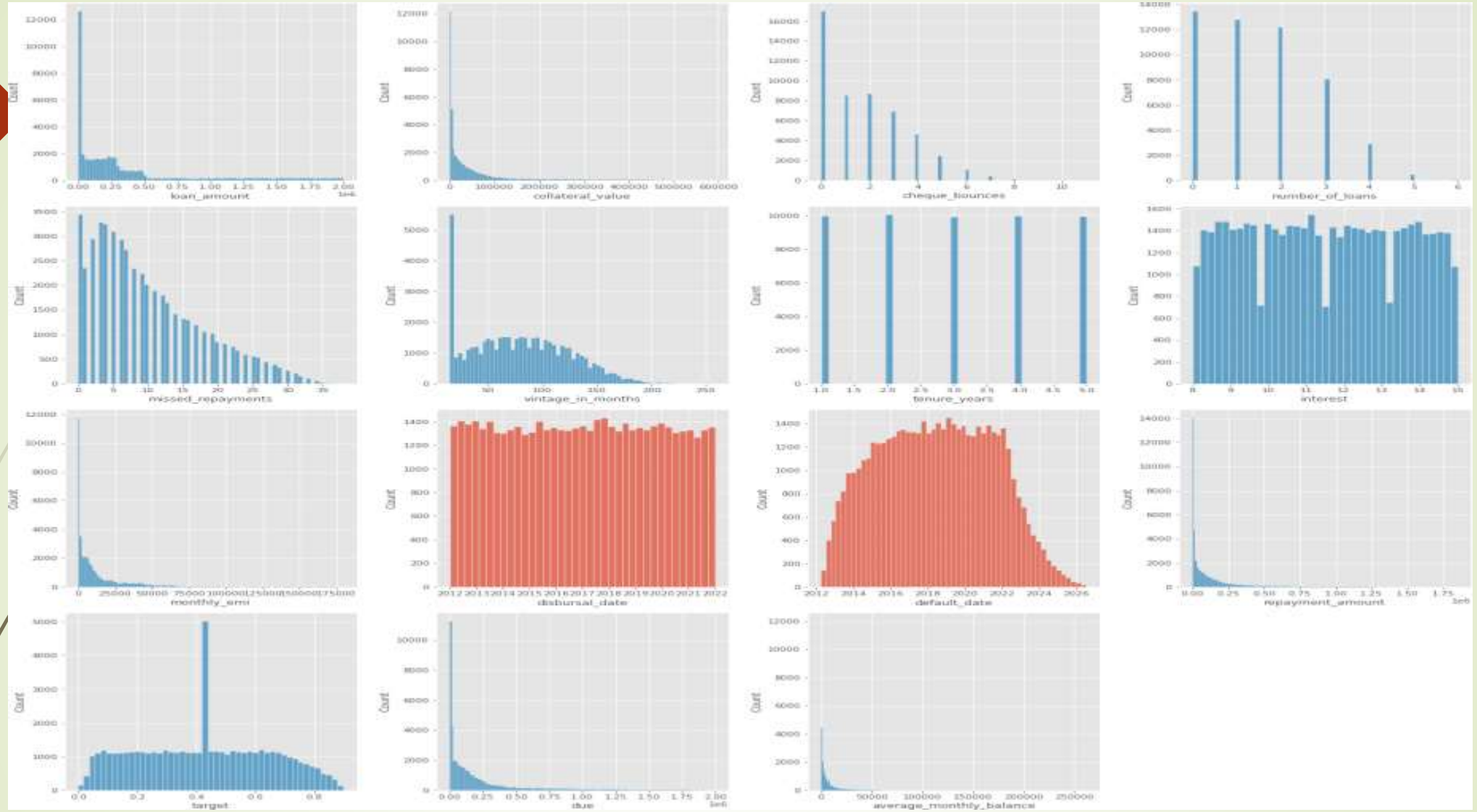
PROCESSING OF DATA

- For each data set converted Data types if necessary
- Merging the data sets and created target variable(LGD)
- Exploratory Data Analysis has been performed
- Variable Transformation
- Dummy Encoding Scaling
- using Standard Scale

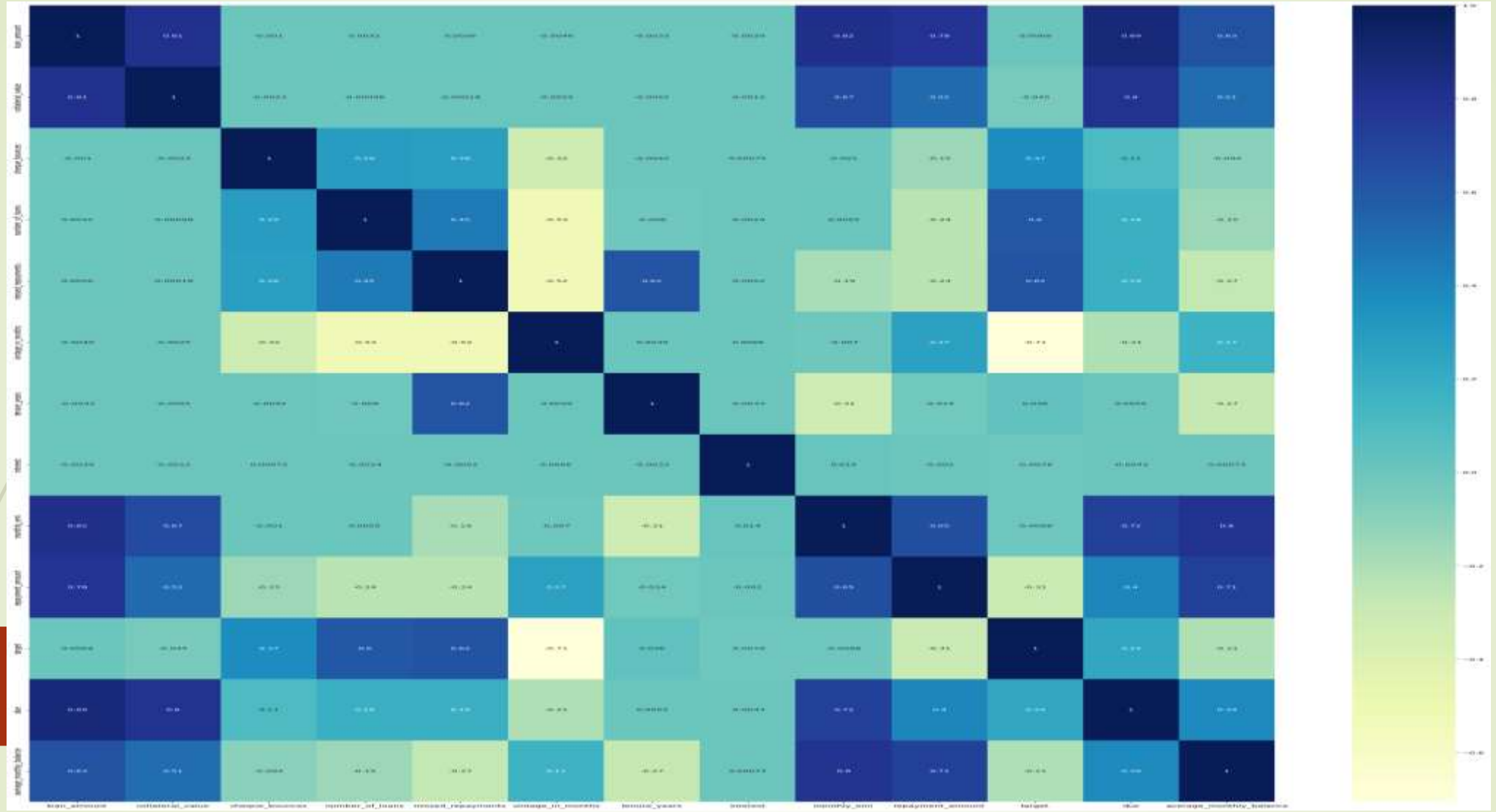
EDA



Monthly EMI also car loan is much higher compared to other loans.



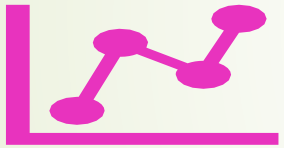
Plotted Histograms for the numerical columns to understand the distribution of data.



Created the Heatmap to understand the correlation between the variables

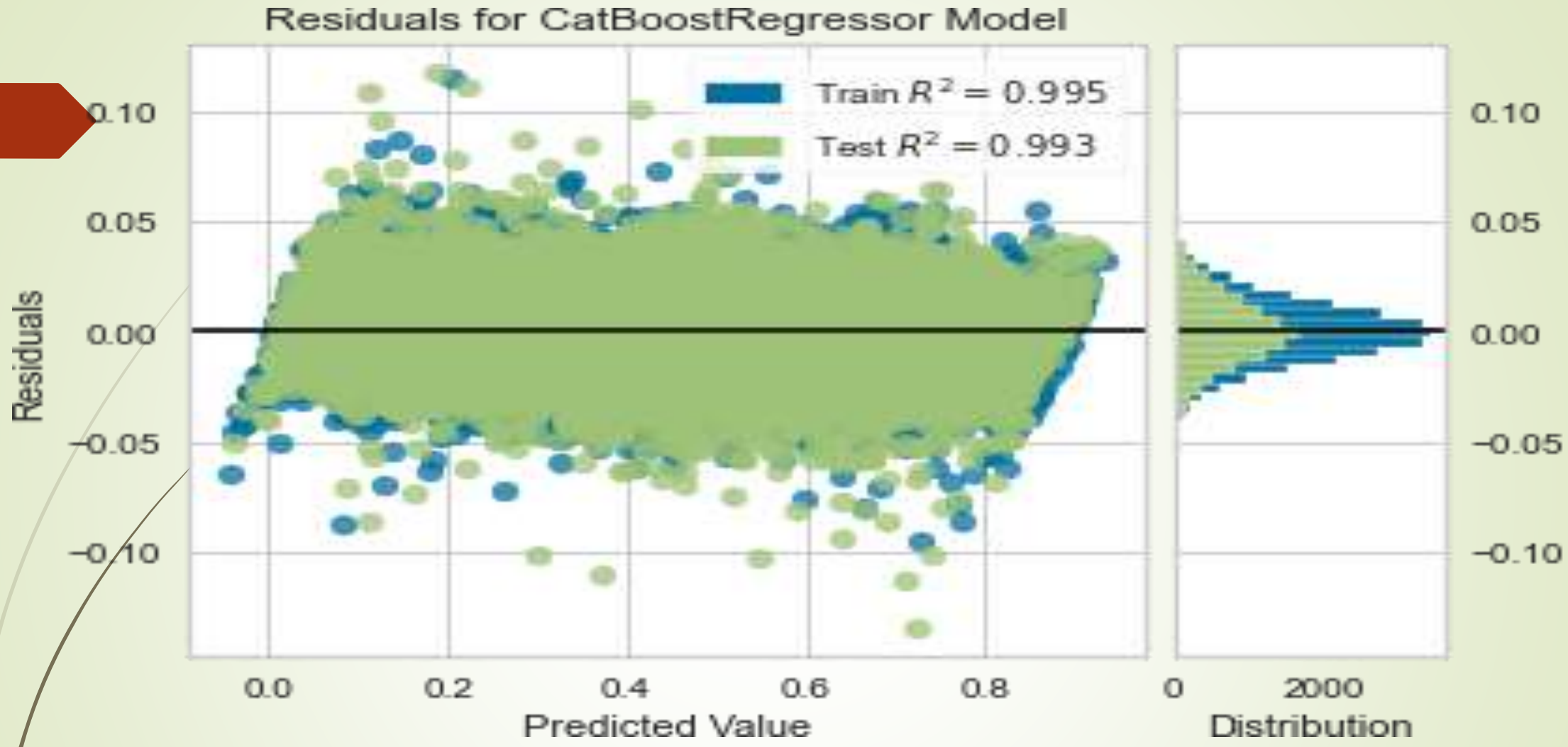
MODEL BUILDING

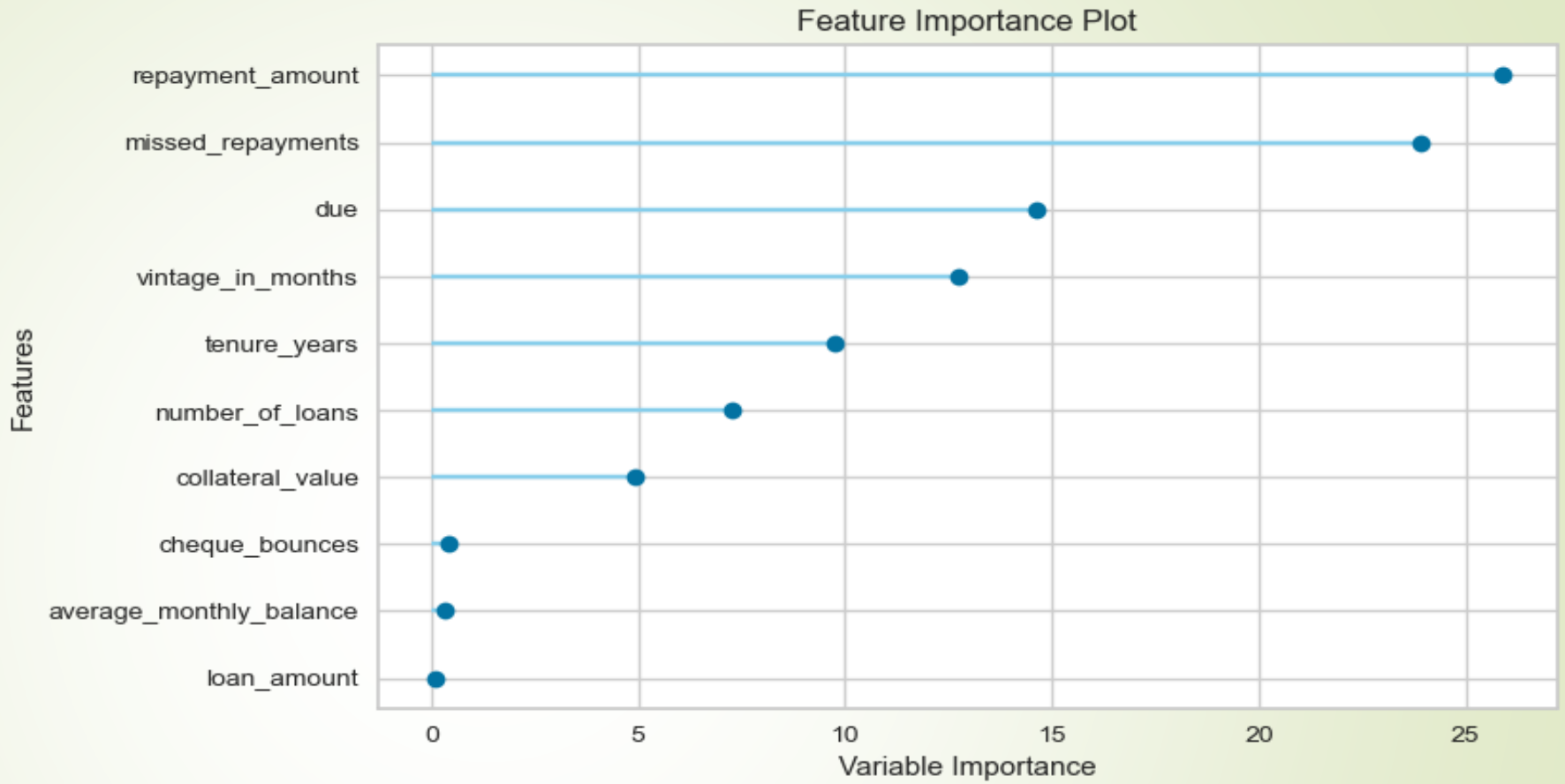
Used various models like Multiple Linear Regression, Random Forest Regressor, Gradient Boosting Regressor, XGBoost Regressor, Adaboost Regressor, ElasticNet : Hybrid Regularized Model, LightGBM for model building.



Used R Squared as a performance metrics.

XGBoost has given us 99.5% R squared on test data across the models.


Residuals for CatBoostRegressor Model
Train $R^2 = 0.995$
Test $R^2 = 0.993$
Residuals
0.10
0.05
0.00
−0.05
−0.10
0.0
0.2
0.4
0.6
0.8
Predicted Value
0
2000
Distribution



REGRESSION INTERPRETATION

RECOMMENDATIONS

We should focus more on Car and

We should focus more on car and two-wheeler

Missed Repayment customers with high repayment amount should be highlighted

Customer’s due factors and tenure are another subset of influencers to predict the LGD of the customers