

# INSIGHTS

## BFSI: CREDIT RISK

---

- 
- A large, abstract graphic in the background consists of a grid of small red dots arranged in a circular, wave-like pattern, creating a sense of motion and depth.
1. POOJA KUMARI
  2. RAGUL

# OBJECTIVE

---

- CREDIT RISK ANALYTICS PLAYS A PIVOTAL ROLE IN THE BANKING SECTOR, WHERE ASSESSING AND MANAGING THE POTENTIAL CREDIT RISK ASSOCIATED WITH BORROWERS IS CRUCIAL FOR MAINTAINING FINANCIAL STABILITY.
- THE GOAL IS TO EMPLOY SOPHISTICATED ANALYTICAL TECHNIQUES TO PREDICT AND QUANTIFY 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.



# GOAL

---

- The goal is to develop a statistical model that can predict the Loss Given Default (LGD) for borrowers.
  - The LGD value for a loan, given its collateral and assuming that the customer has already made some repayments.
-

# ABOUT DATASET

---

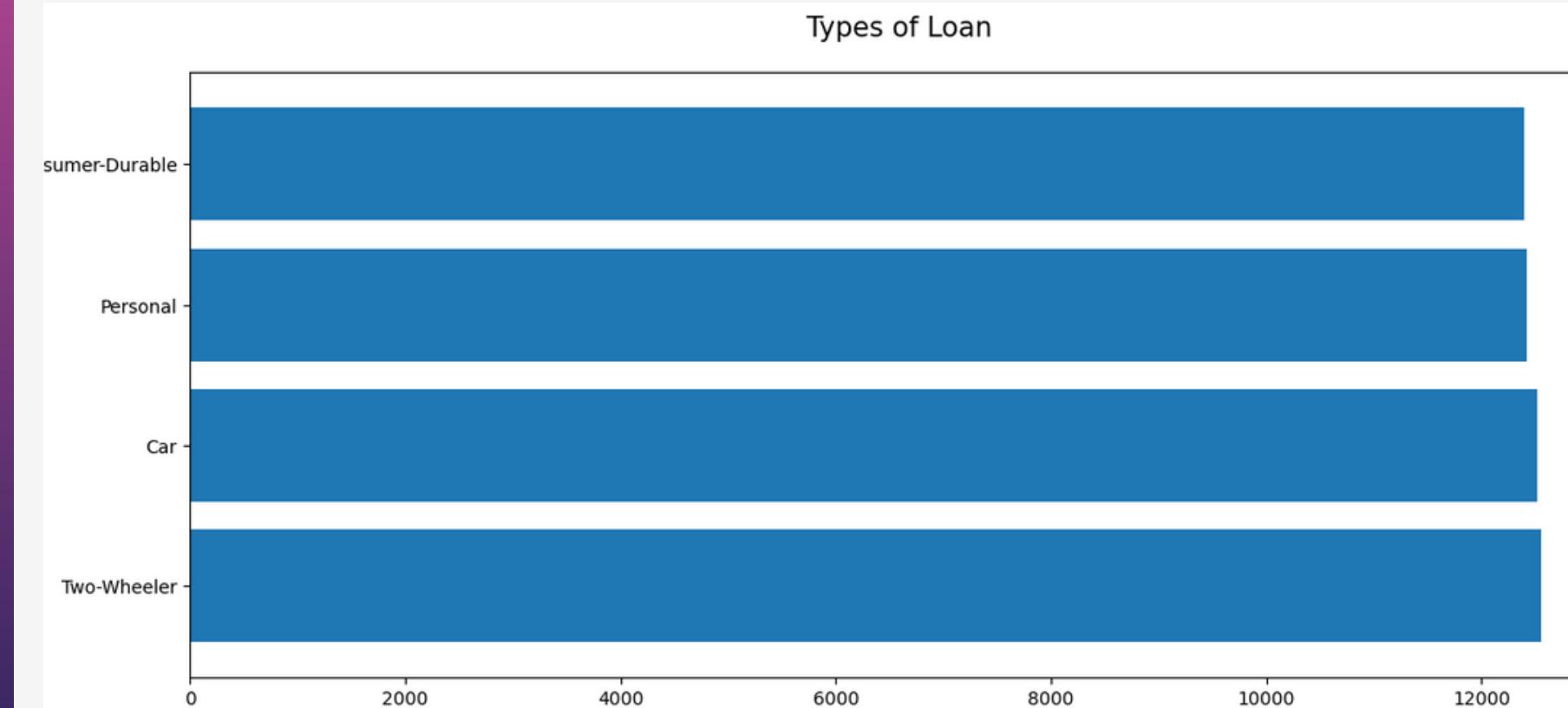
- The `main_loan_base` data set contains information about loan accounts and other relevant information for the corresponding borrowers.
- The `repayment_base` data set contains information about the repayments received by the banks in the form of EMIs or through other collection efforts pertaining to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

# Data processing steps

- CONVERTED DATA TYPES FOR EACH NECESSARY DATASET.
- HANDLED NULL VALUES USING DELETION AND IMPUTATION TECHNIQUES TO EITHER REMOVE INCOMPLETE DATA OR FILL IN MISSING VALUES.
- REMOVED DUPLICATE VALUES FROM THE DATASETS TO ENSURE DATA INTEGRITY AND PREVENT REDUNDANCY.
- MERGED DATASETS AND CREATED A TARGET VARIABLE (LGD), WHICH IS LIKELY TO BE USED FOR PREDICTIVE MODELING OR ANALYSIS.
- CONDUCTED EXPLORATORY DATA ANALYSIS TO GAIN INSIGHTS INTO THE CHARACTERISTICS AND PATTERNS WITHIN THE DATASETS.
- DUMMY ENCODING
- SCALING USING STANDARD SCALER

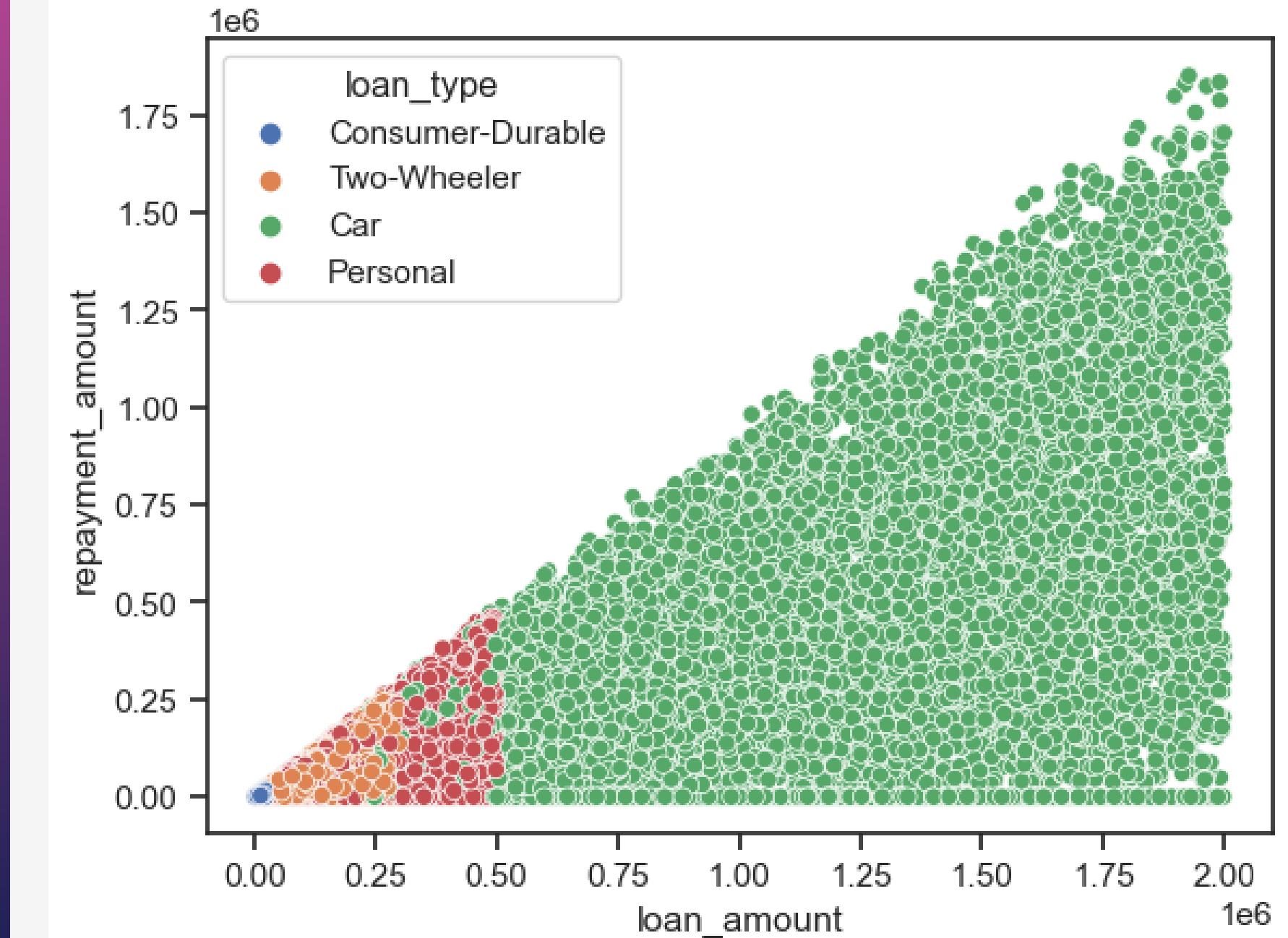
# EDA

**THIS OBSERVATION INDICATES  
A HIGHER DEMAND OR  
PREVALENCE OF LOANS  
ASSOCIATED WITH  
TWO WHEELER ACQUISITIONS  
COMPARED TO OTHER  
CATEGORIES IN THE DATASET.**



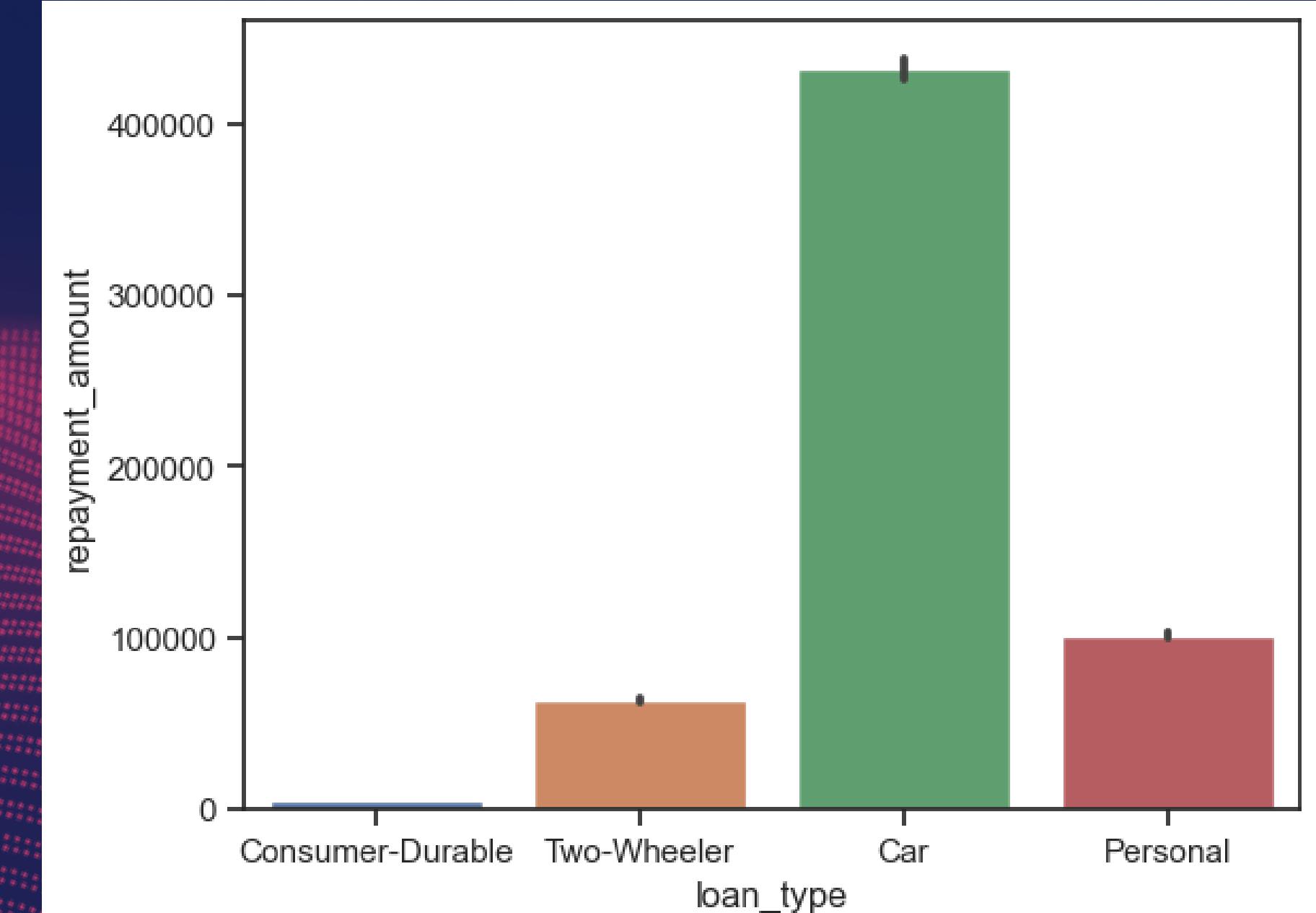
# EDA

DESPITE THE HIGHER NUMBER OF LOANS FOR TWO-WHEELER VEHICLES, CAR LOANS HAVE THE HIGHEST LOAN AMOUNTS AMONG ALL VEHICLE TYPES. THIS SUGGESTS THAT CUSTOMERS SEEKING CAR LOANS ARE WILLING TO BORROW LARGER AMOUNTS, POSSIBLY DUE TO THE HIGHER VALUE OF CARS COMPARED TO TWO-WHEELERS.



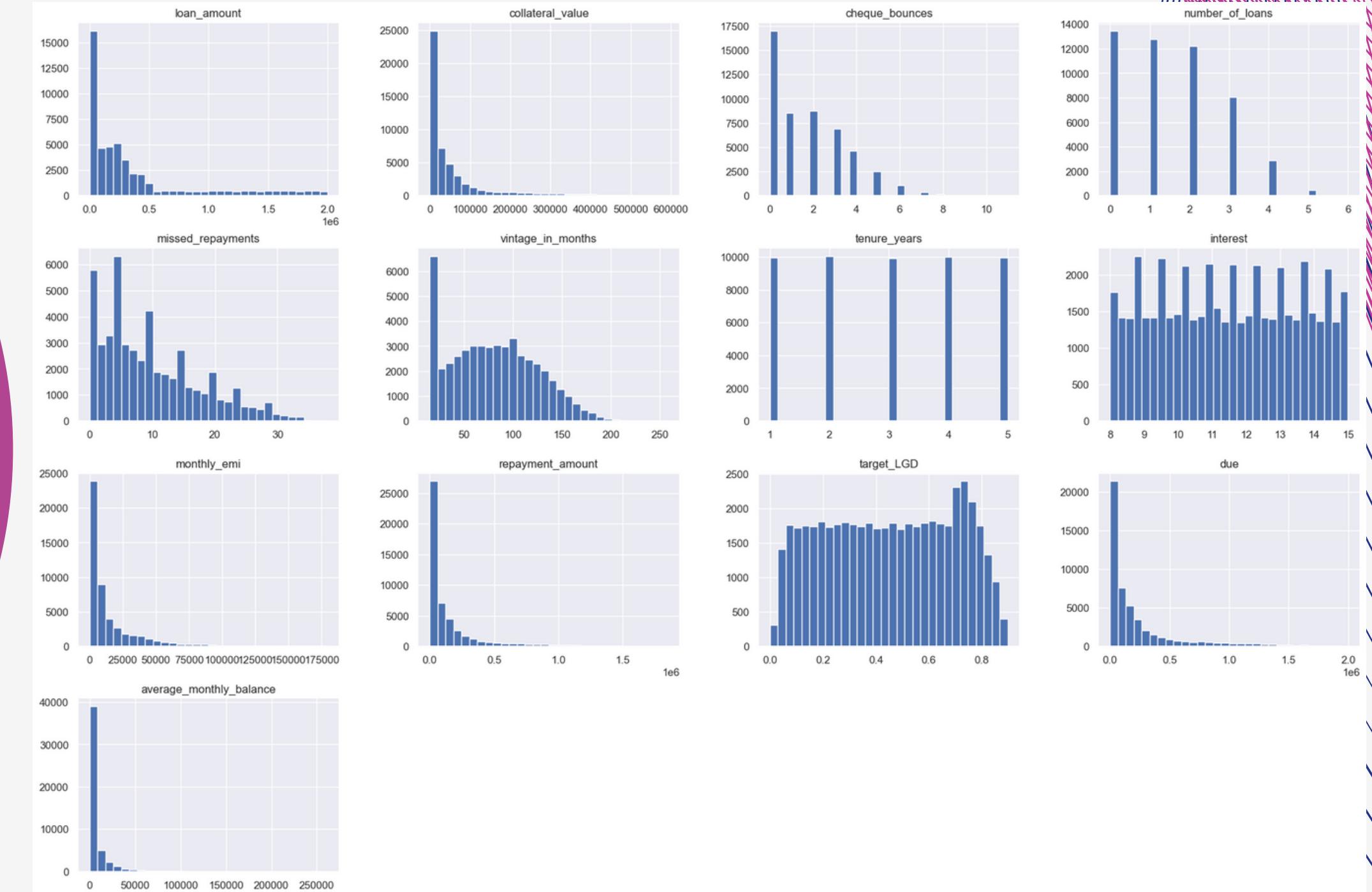
## Income raise

CAR LOANS HAVE SIGNIFICANTLY HIGHER MONTHLY EQUATED MONTHLY INSTALLMENTS (EMIS) COMPARED TO OTHER TYPES OF LOANS. THIS COULD BE ATTRIBUTED TO THE LARGER LOAN AMOUNTS ASSOCIATED WITH CAR LOANS AND POTENTIALLY LONGER REPAYMENT TENURES.



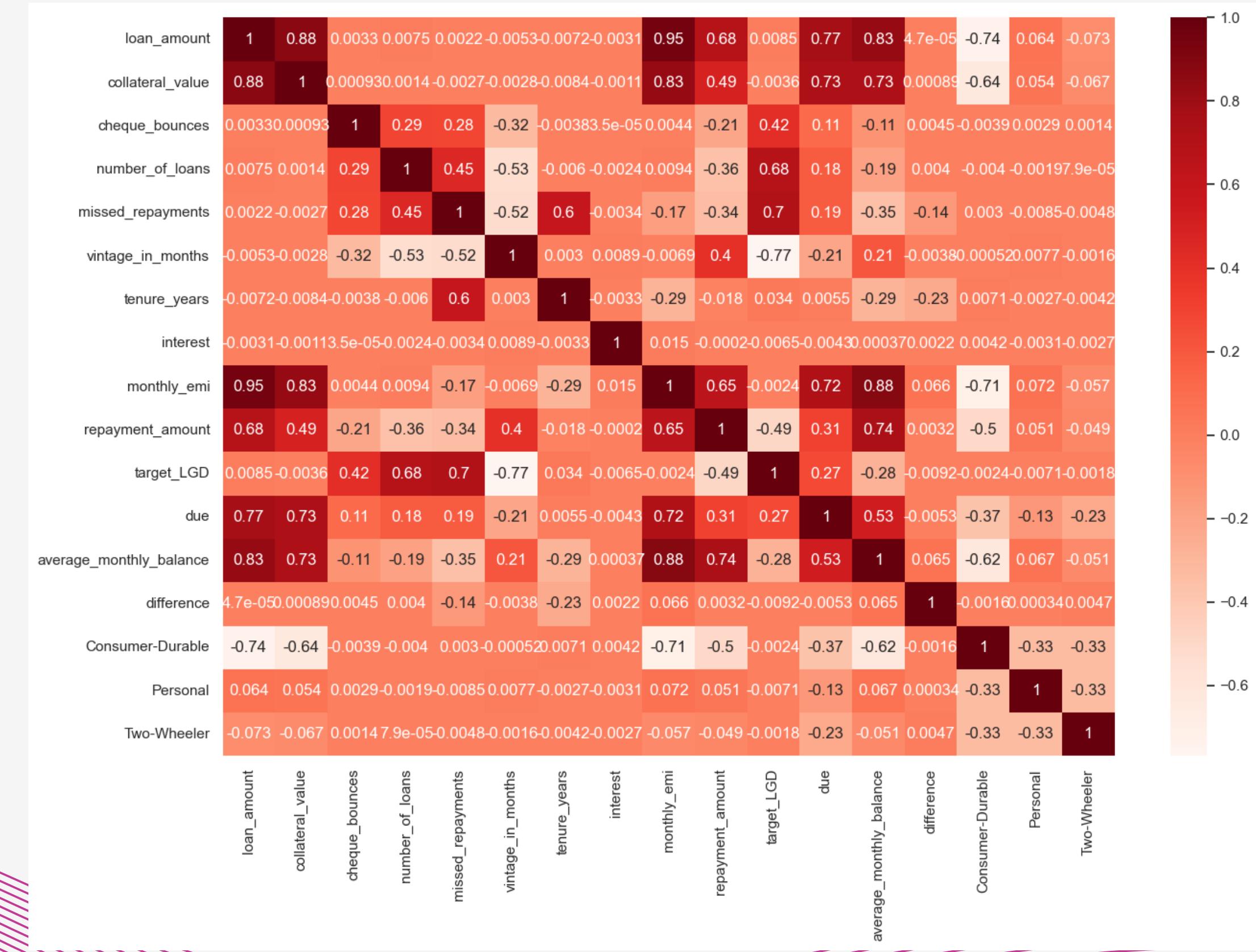
# DATA DISTRIBUTION

CREATED HISTOGRAMS FOR THE NUMERICAL COLUMNS TO GAIN INSIGHTS INTO THE DISTRIBUTION OF DATA.



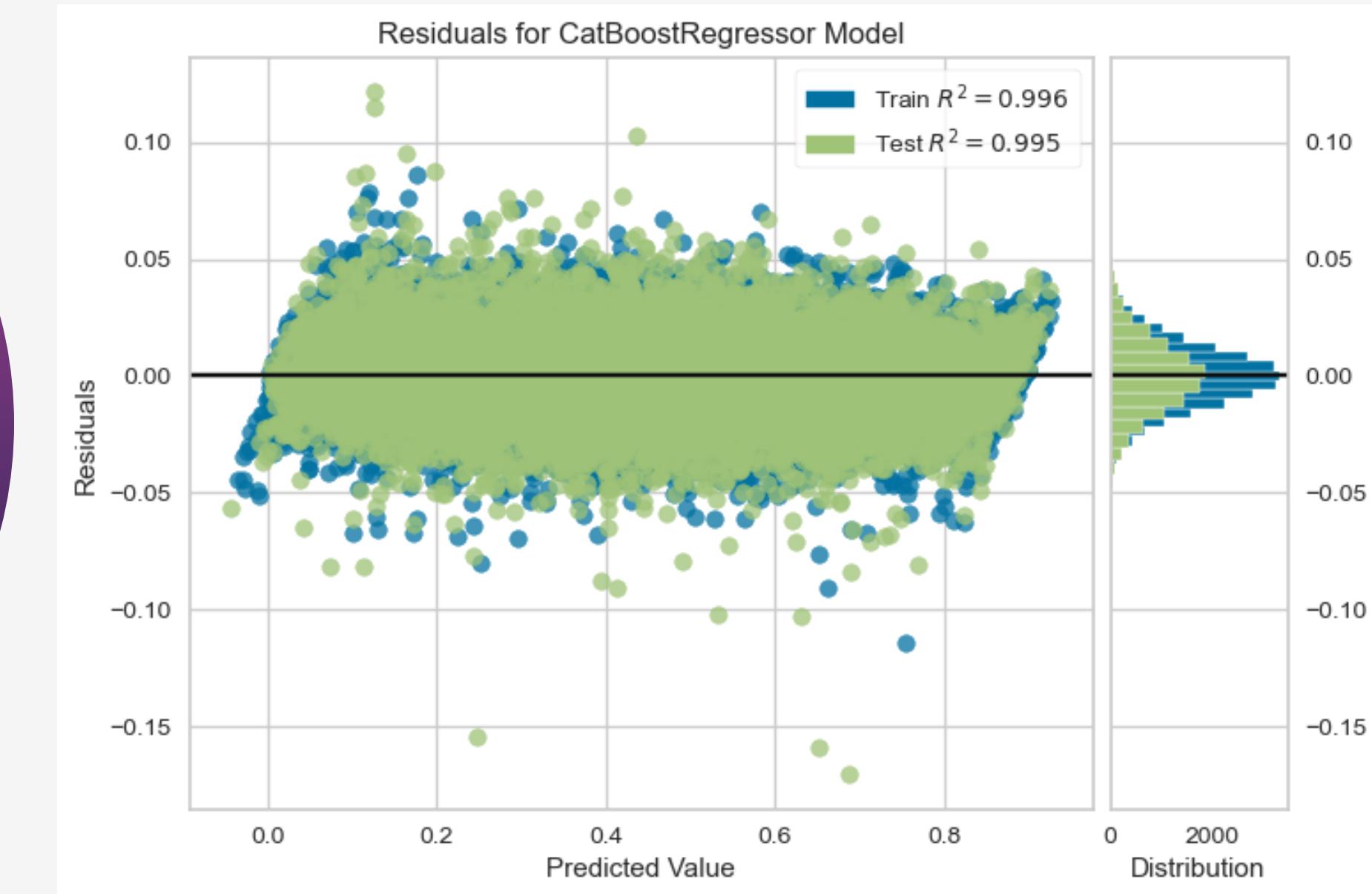
# CORRELATION

- OBSERVED A NOTABLE HIGH CORRELATION AMONG CERTAIN VARIABLES THROUGH THE USE OF A HEATMAP.



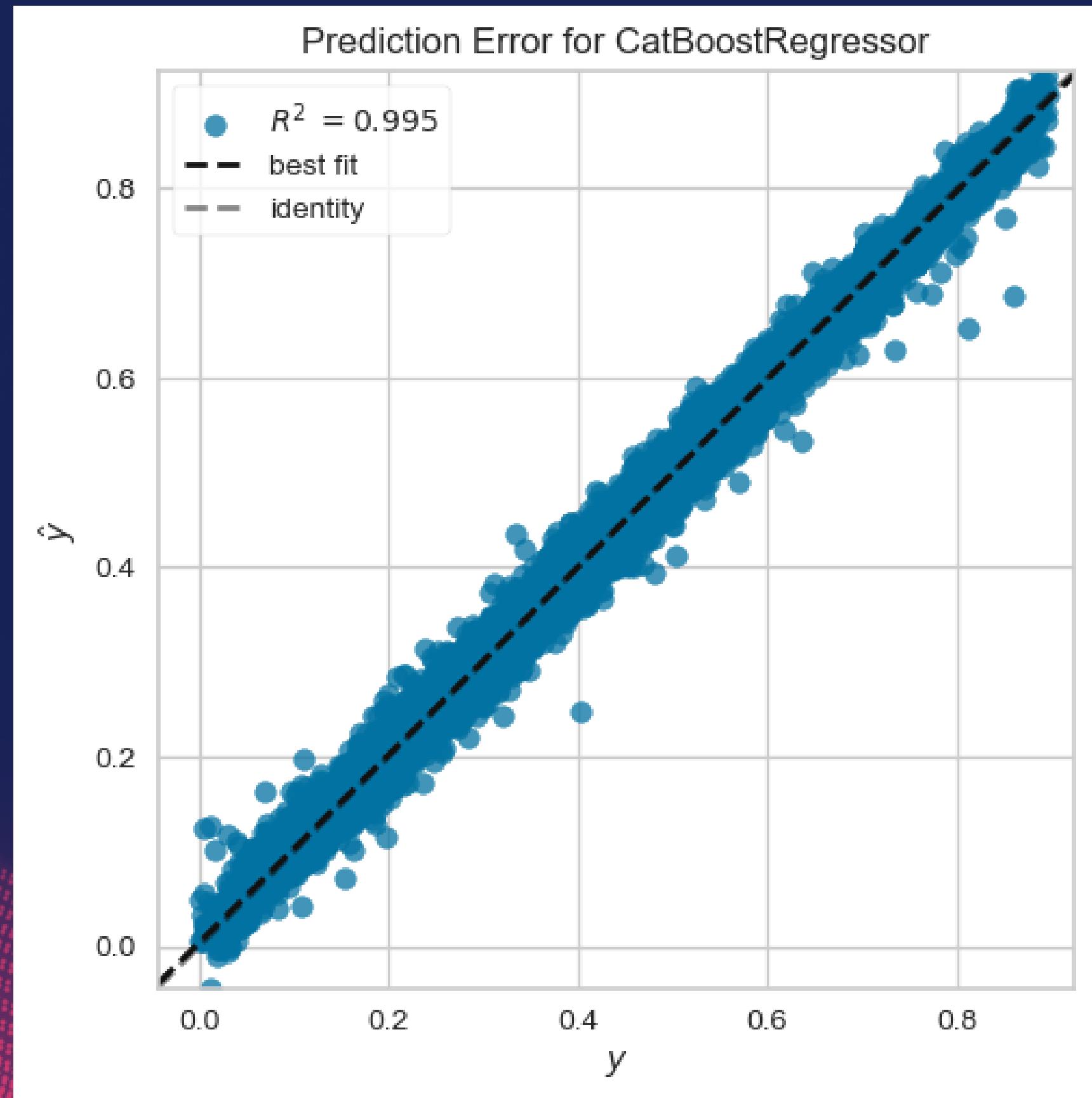
# INTERPRETATIONS

**THE TEST AND  
TRAIN VALUE OF R<sup>2</sup>  
IS SAME**



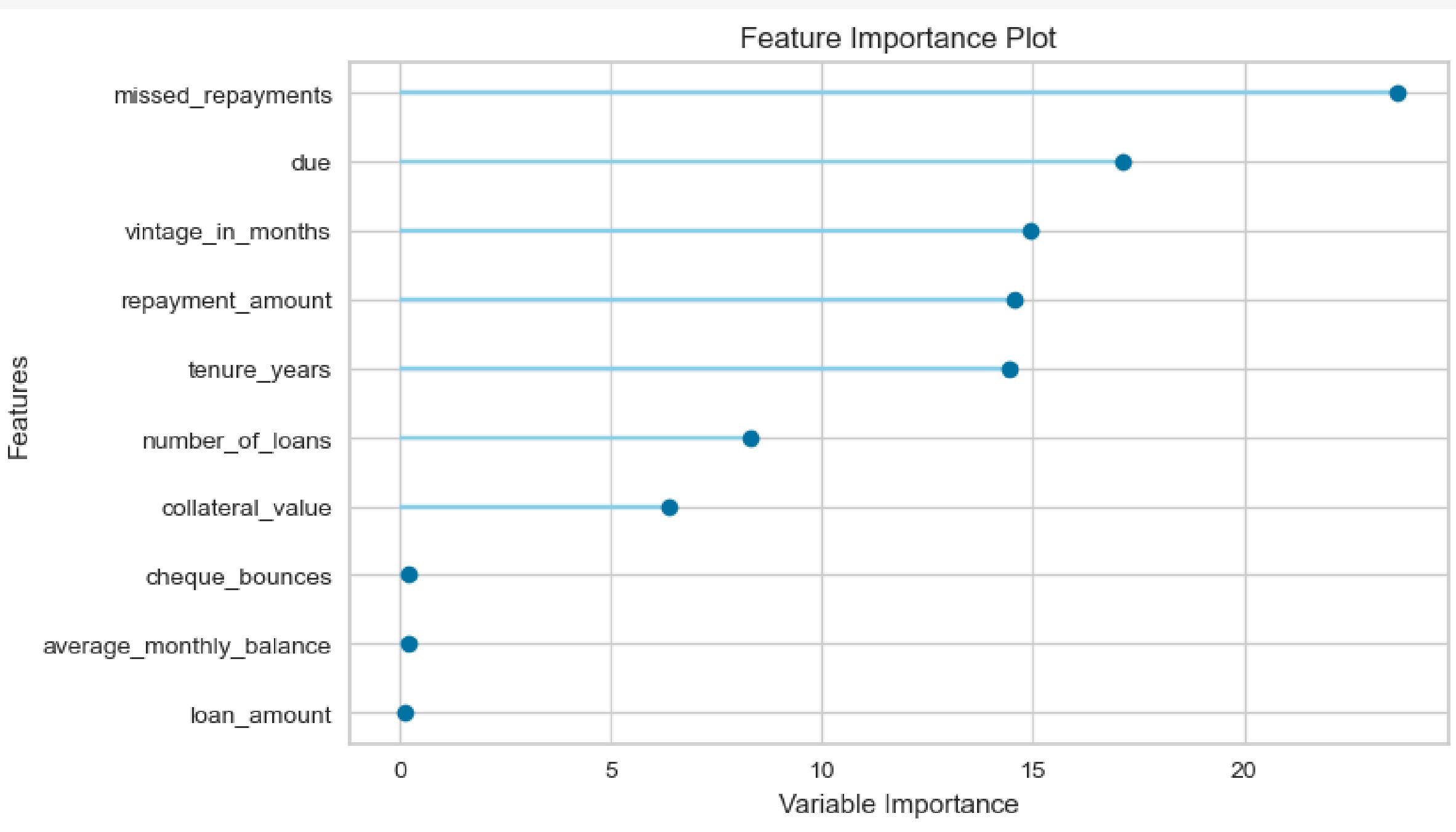
# Interpretations

**THE VALUE OF R<sup>2</sup> IS  
REVOLVING AROUND  
THE BEST FIT LINE**



# INTERPRETATIONS

## FEATURE IMPORTANCE



# INSIGHTS

---

TWO-WHEELER LOANS DOMINATE DEMAND, WHILE CAR LOANS EXHIBIT THE HIGHEST LOAN AMOUNTS. CAR LOANS ALSO INCUR SIGNIFICANTLY HIGHER EMIS, INDICATING POTENTIAL PREFERENCES FOR LARGER PURCHASES.

UNDERSTANDING CORRELATIONS AMONG VARIABLES AIDS IN IDENTIFYING KEY RELATIONSHIPS.

ROBUST PREDICTIVE MODELS, PARTICULARLY XGBOOST, YIELD HIGH ACCURACY.

PRIORITIZING CAR AND TWO-WHEELER LOANS ALONGSIDE CUSTOMER REPAYMENT BEHAVIOR ENHANCES RISK MANAGEMENT STRATEGIES.

# Recommendations

- It is recommended to prioritize our attention on Personal and Two-Wheeler loan types. These loan categories appear to have significant relevance and impact on operations.
- It is crucial to pay special attention to customers with missed repayments, especially those with high repayment amounts.
- Customers due factors and tenure emerges as a significant subset influencing the prediction of Loss Given Default (LGD)

**THANK YOU**