The background of the slide is a dense, 3D-rendered field of numbers (0-9) in various shades of blue and white. The numbers are of different sizes and are arranged in a way that creates a sense of depth and perspective, with some numbers appearing to rise from the surface.

BFSI : Credit Risk Assignment

TANVI SRIVASTAVA

NARENDRA KUMAR

MYTREYA YECHURI

Table of Contents

- ◊ Objective
- ◊ Background
- ◊ Data Analysis
- ◊ Pre-Processing of Data
- ◊ Exploratory Data Analysis (EDA)
- ◊ Model Building
- ◊ Results & Insights
- ◊ Recommendations

Process

Objective

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

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

- ❑ 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.

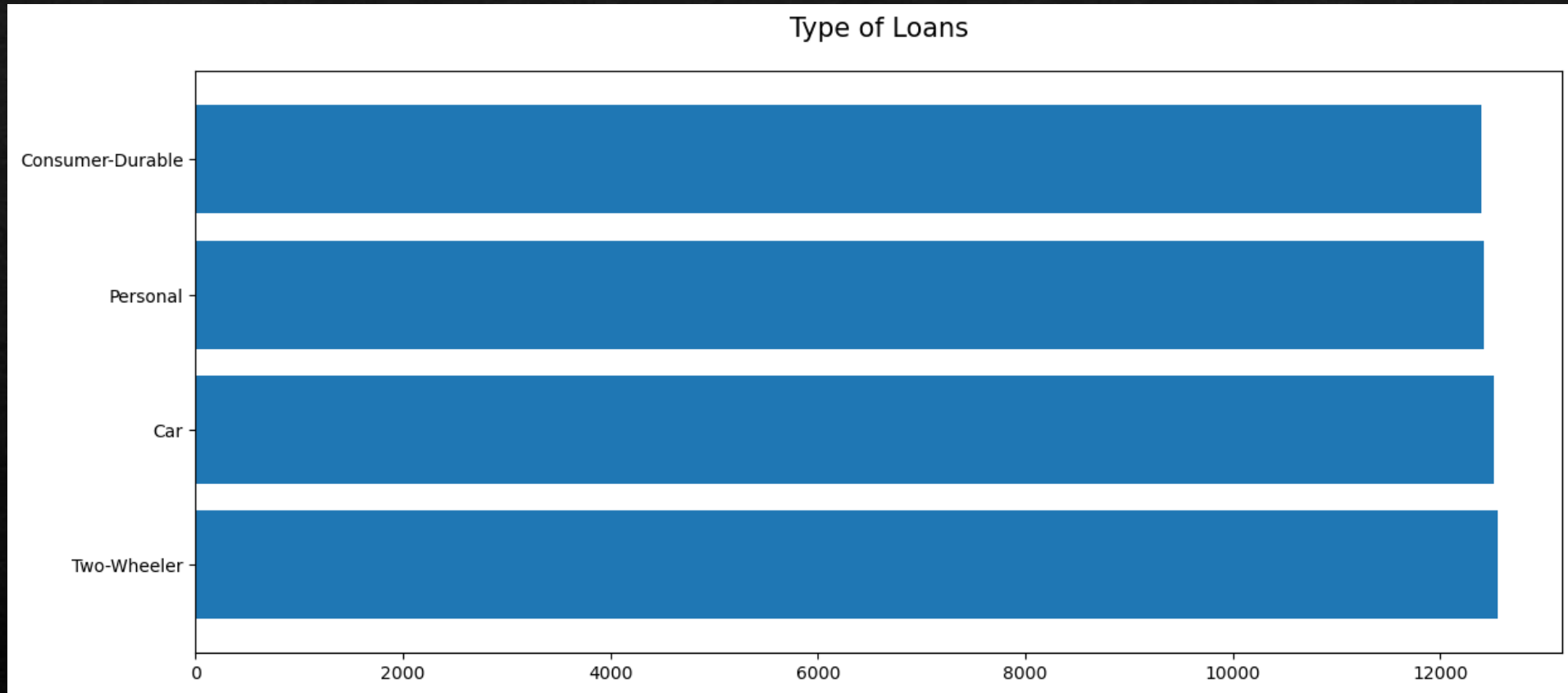
Background

- ❑ 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**

Data Analysis

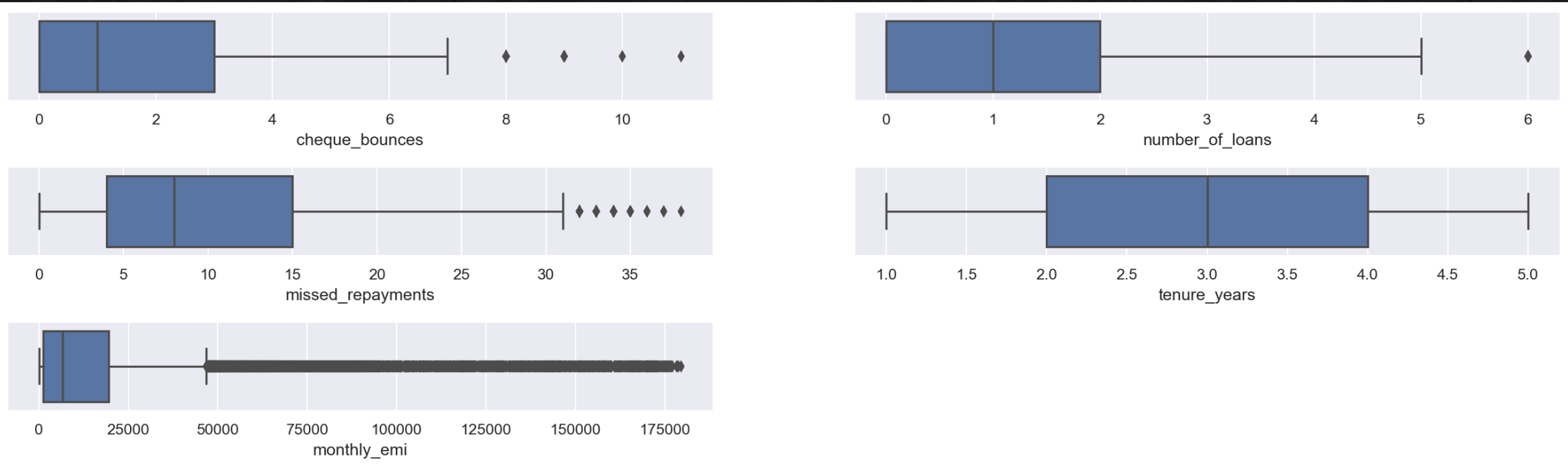
- ❑ Used 3 Datasets for model building
- ❑ The main_loan_base dataset.
- ❑ The repayment_base dataset.
- ❑ The monthly_balance_base dataset.

Two Wheeler Loans appear to be Highest



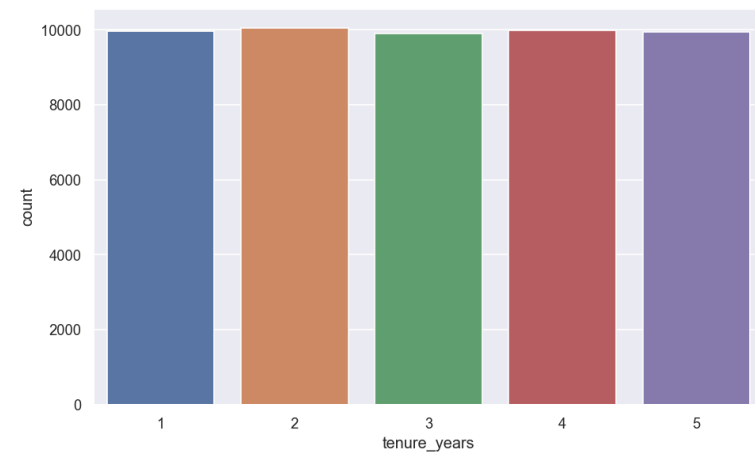
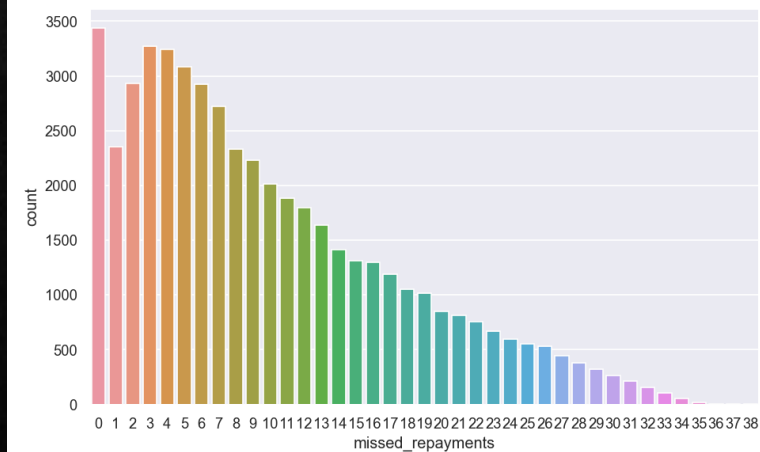
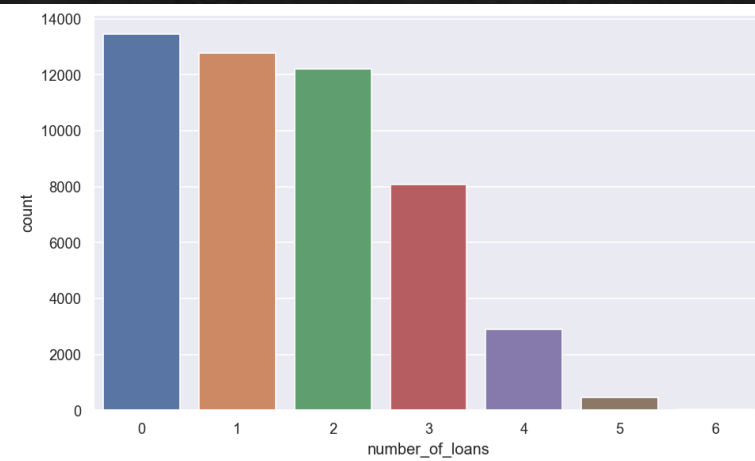
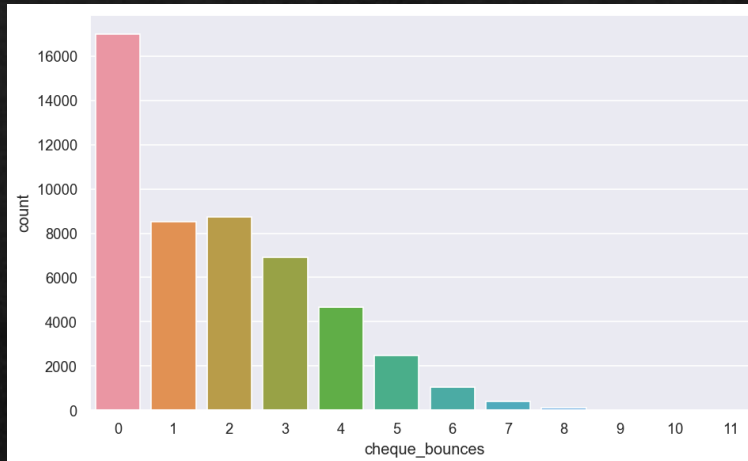
Observations:

- Monthly EMI has enormous outliers. Though it will not effect our analysis
- There are few loan accounts who have missed the repayments more than 30 times
- Generally people are taking one loan at a time
- Generally the loan tenure is 2-4 years
- There are few loan account whose cheque has bounced more than 8 times

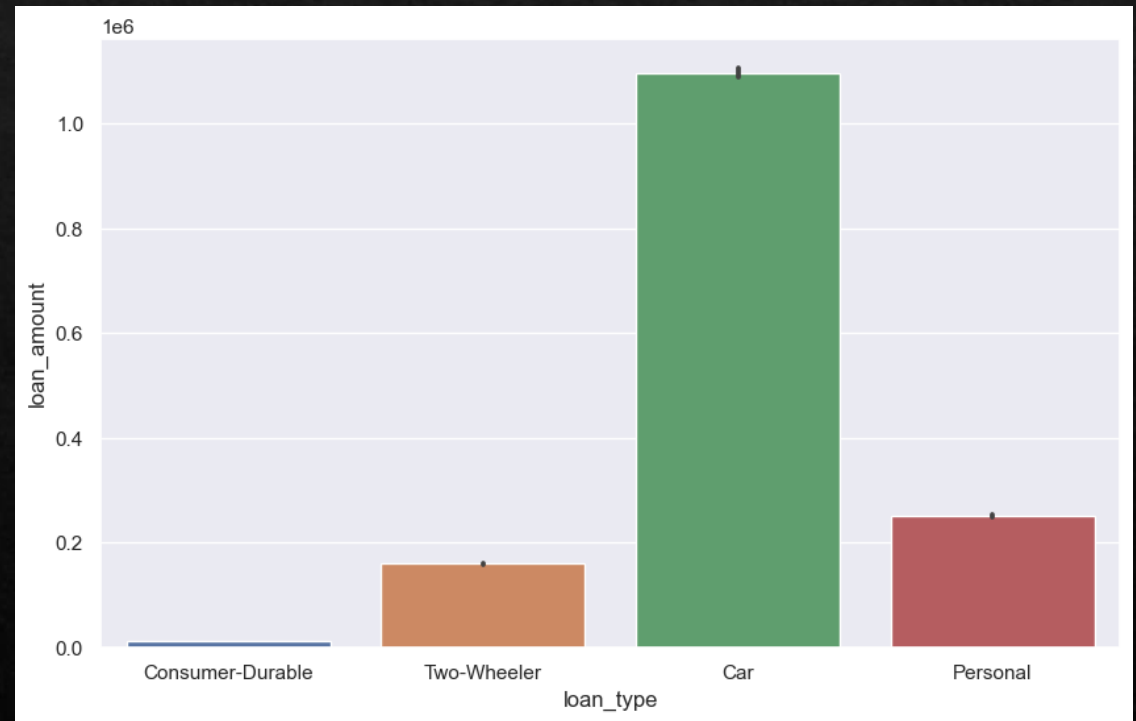
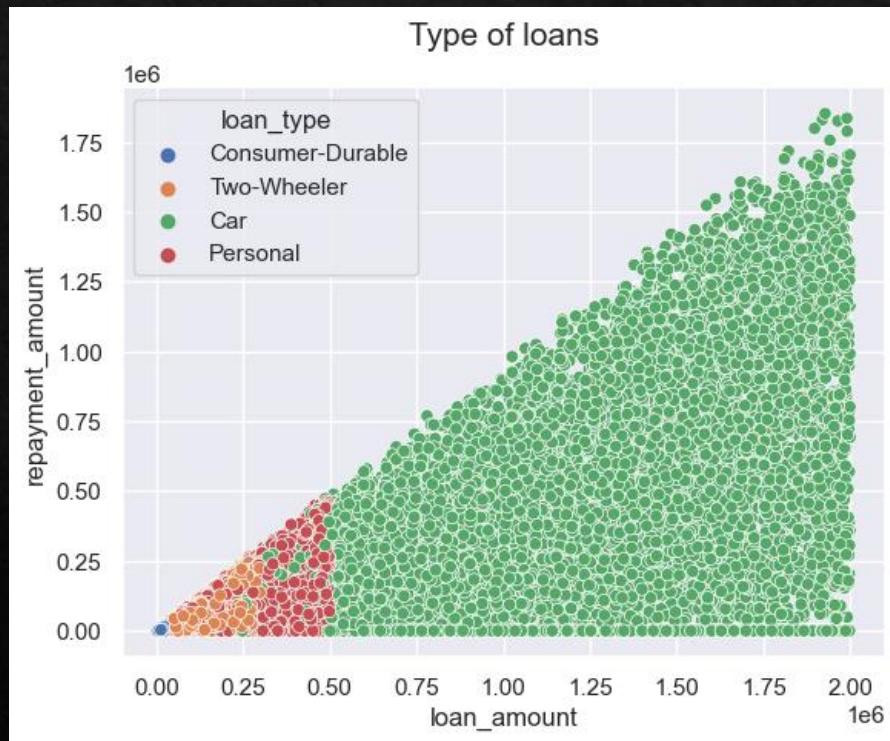


Observation:

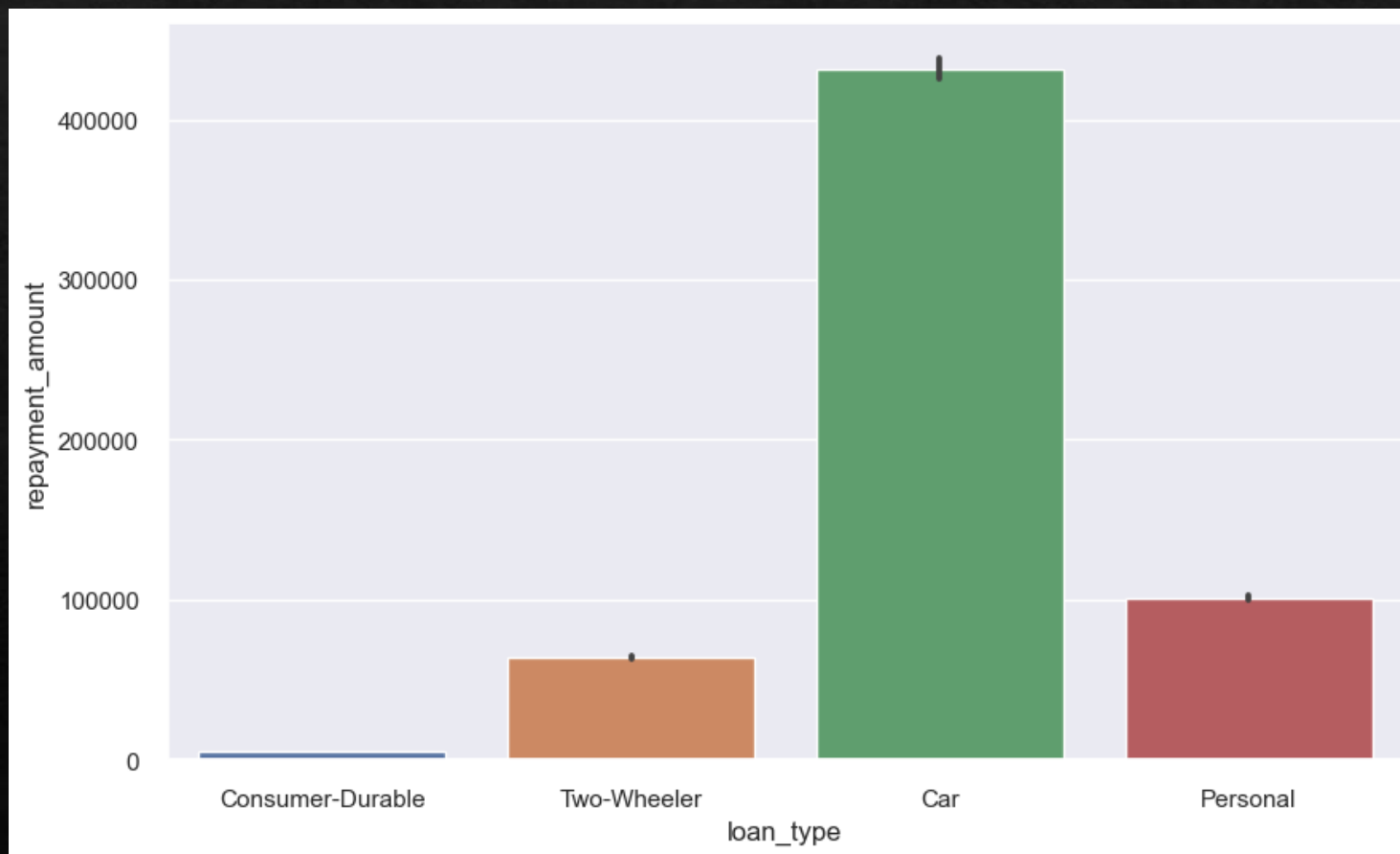
1. The cheques of over 16,000 loan accounts have bounced.
2. There is decreasing trend found on missed repayment of loan amount
3. Equal number of people are taking one and two loan at a time



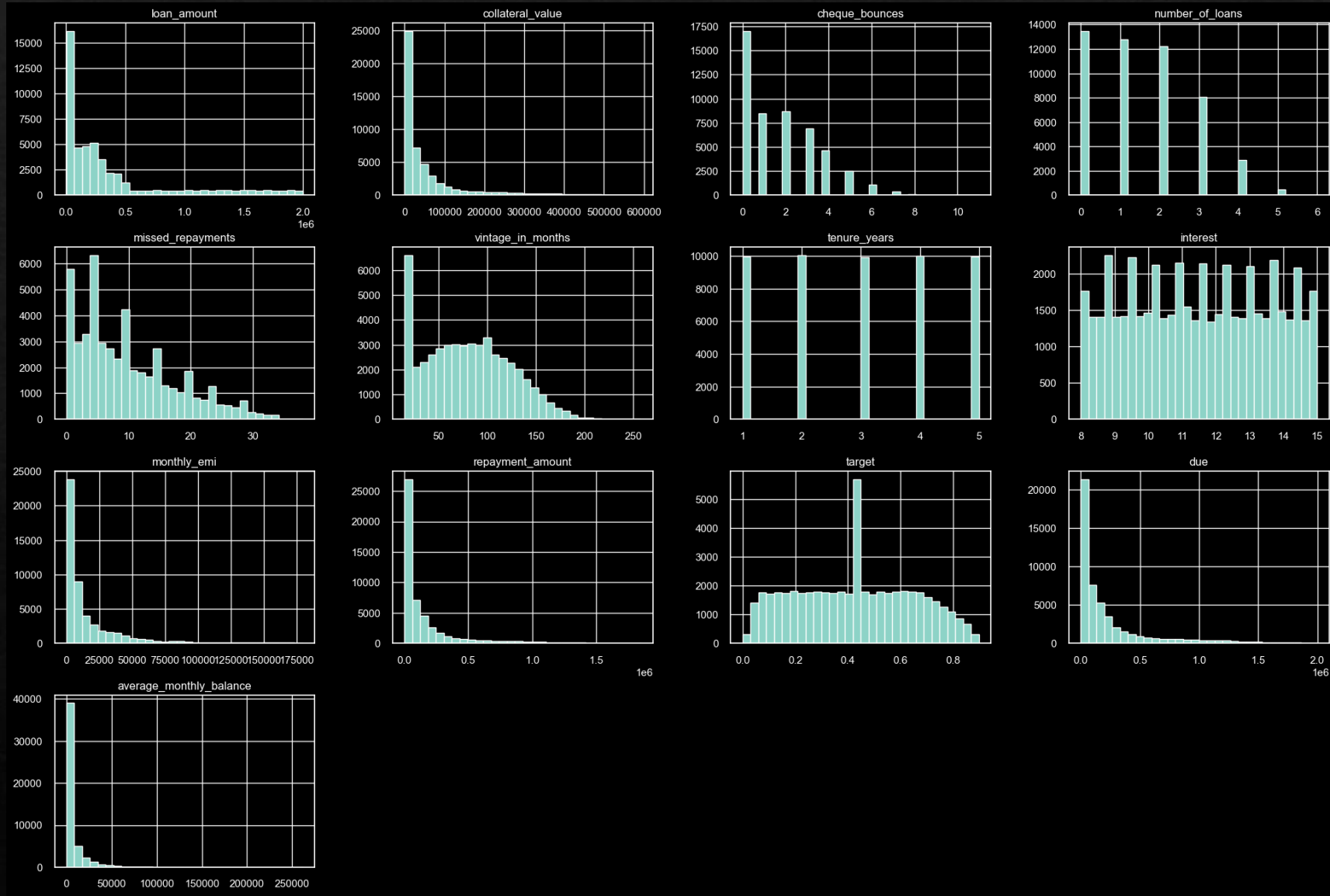
Observation:
Number to Two wheeler loans are maximum
Amount of loan disbursed for Car Loans are maximum



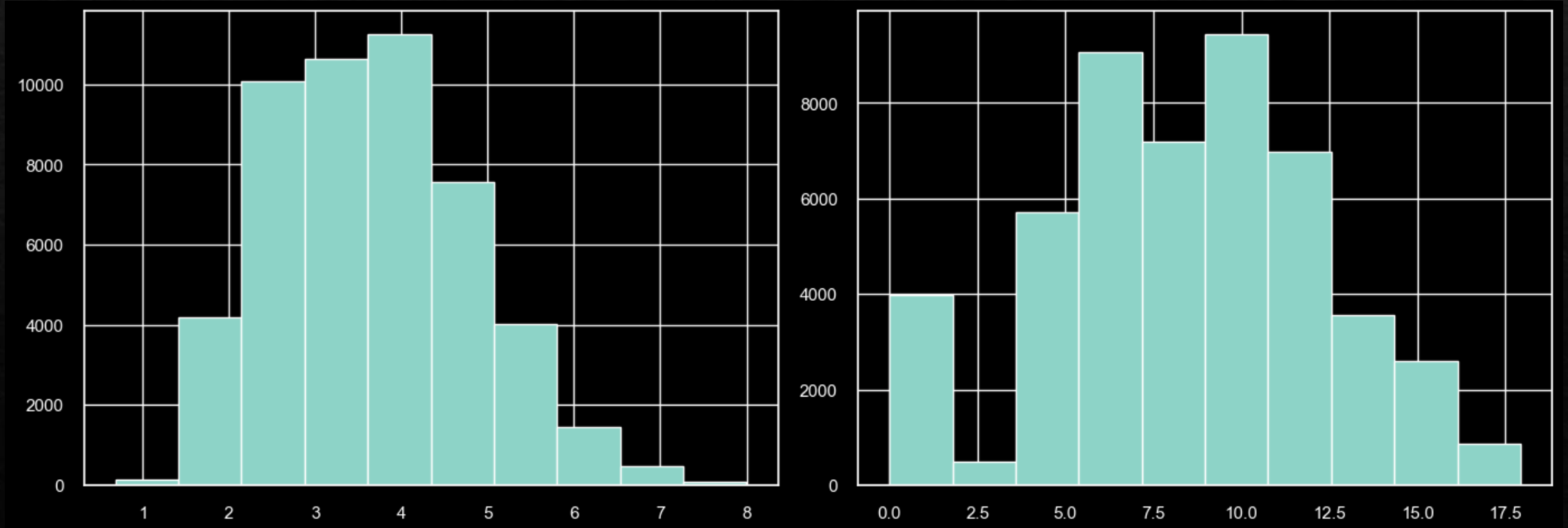
Observation:
Repayment amount & EMI for Car Loans are higher than other loan types



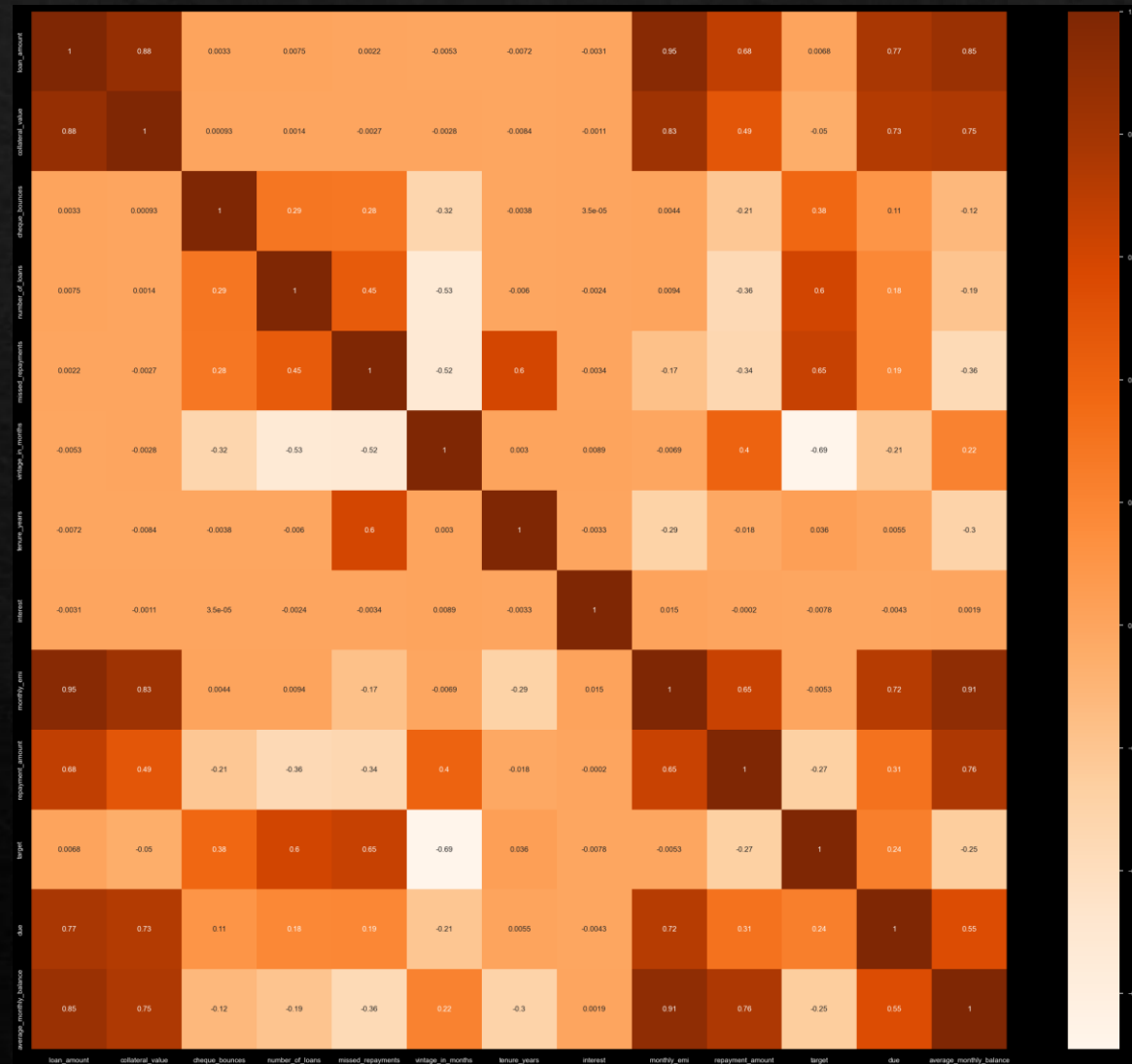
Data Preparation: variable transformation, feature engineering



Power Transformation was used here with respect to the linear regression assumption that all the independent features should have normal distribution



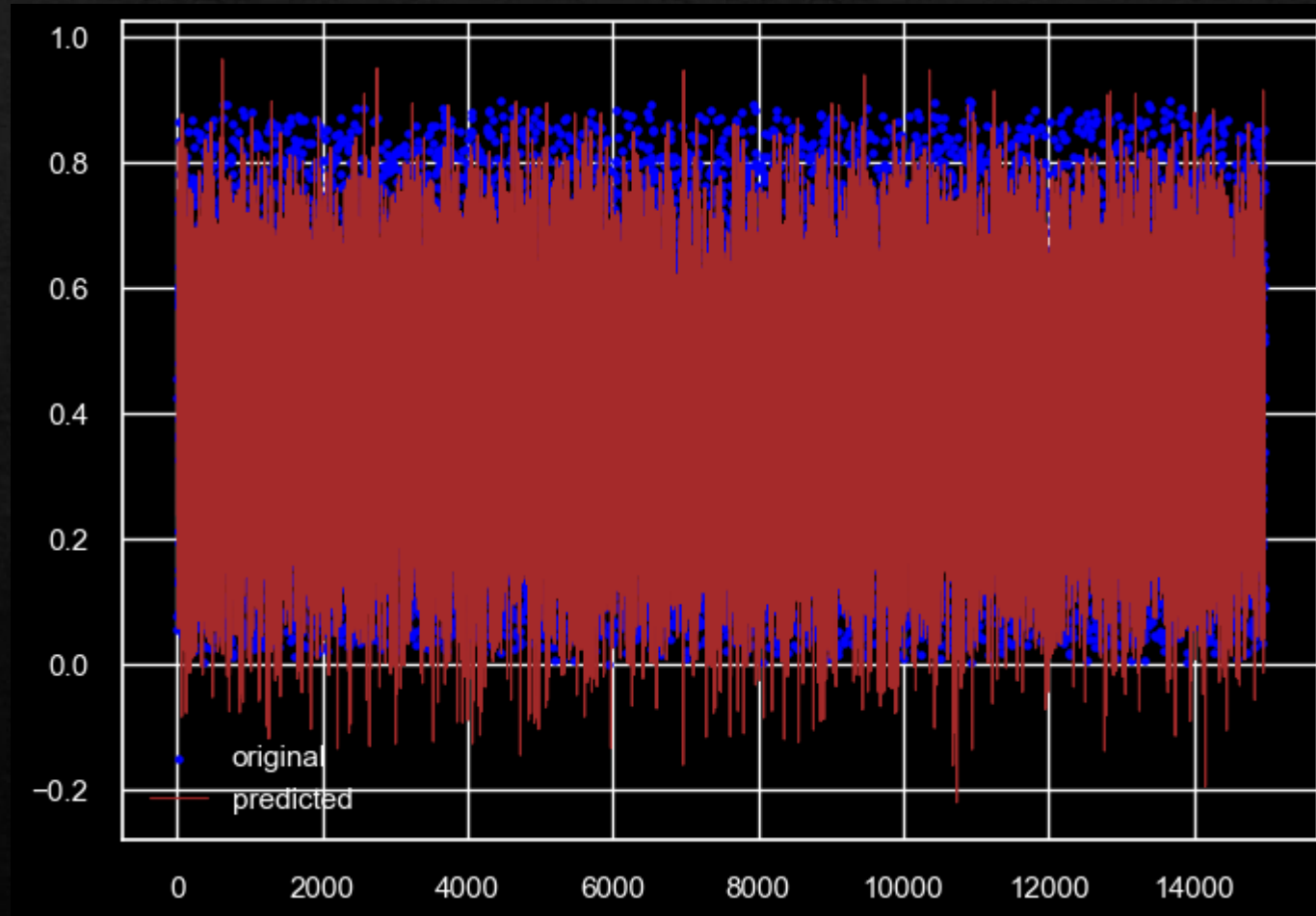
Plotted a Heatmap to understand the correlation between different variables



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.

Original vs Predicted Scatter



Recommendations

- ❖ We should focus more on Car and Two-wheeler loan types.
- ❖ Missed repayment customers with high repayment amount should be highlighted.
- ❖ Customer's due factors and tenure are another subset of influencers to predict the Loss Given Default of the customers.
- ❖ Recommendations made are subjected to the insights obtained from data set after thorough analysis.

The background of the slide is a dense field of 3D-rendered numbers in various sizes and orientations, creating a sense of depth and movement. The numbers are light blue and set against a darker blue gradient background. A solid black horizontal bar is positioned across the middle of the slide, containing the text 'Thank You !'. Below this bar, a smaller black rectangular box contains the names of the individuals being thanked.

Thank You !

TANVI SRIVASTAVA
NARENDRA KUMAR
MYTREYA YECHURI