BFSI : CREDIT RISK ASSIGNMENT

BY :

SHASHWAT BHANSALI

INDEX:

OBJECTIVE:

The objective is to build a statistical model to estimate borrowers' LossGivenDefault(LGD)

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

BACKGROUND:

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

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

3) Expected credit loss=Exposure at default x Probability of Default x Loss given default

4)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.

DATASOURCES:

Used 3 Datasets for model Building

1)Themain_loan_base dataset contains information about loan accounts and other relevant information for the corresponding borrowers.

2)The repayment_base dataset contains information about there payments received by the banks in the form of EMIs or through other collection efforts.

3)The monthly_balance_base contains the information pertaining to the monthly balance statements in the borrower's accounts.

PREPROCESSINGOFDATA:

1) For each dataset converted Data types if necessary.

2) Null values are handled using deletion and imputation techniques. As well duplicate values are removed from the datasets.

3) Merging the datasets and created target variable(LGD)

4) Exploratory Data Analysis has been performed.

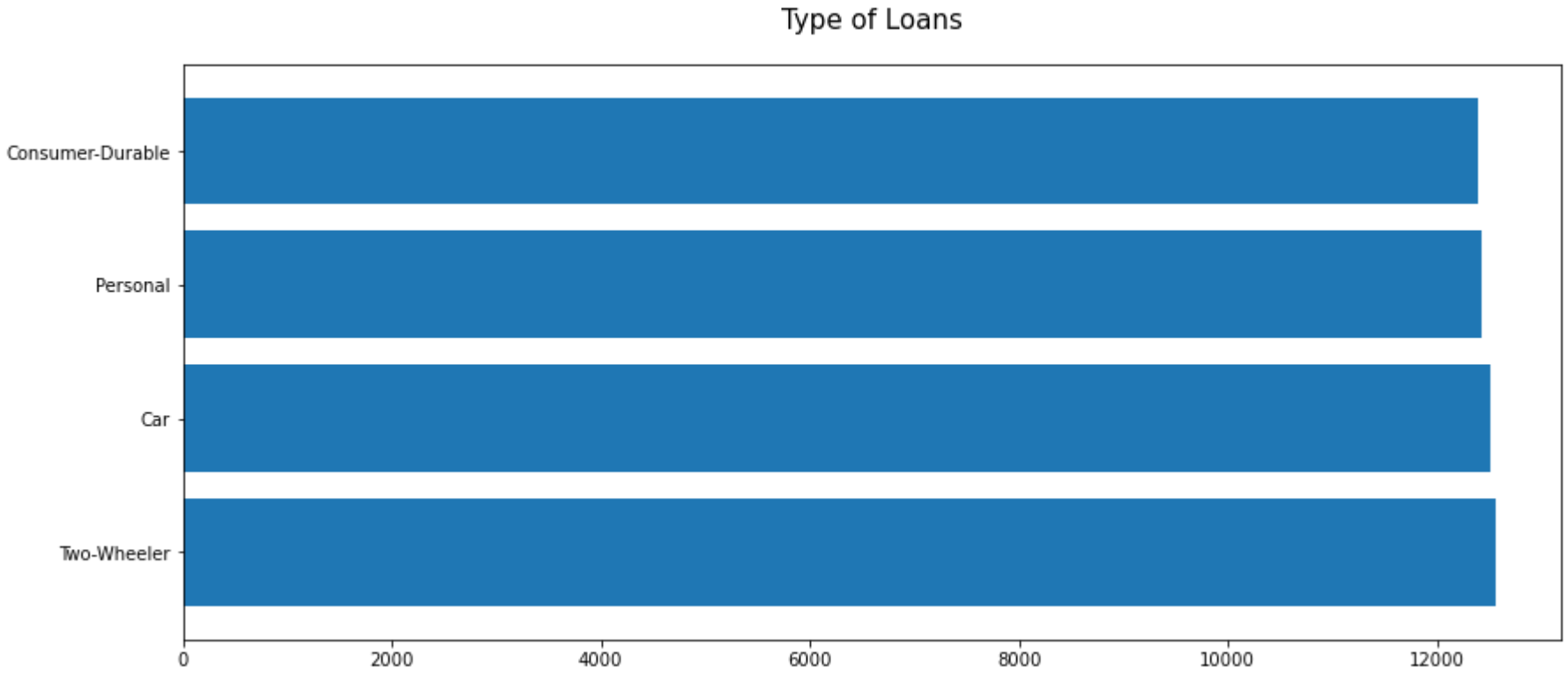
5) Variable Transformation.

6) Dummy Encoding.

7) Scaling using Standard Scaler.

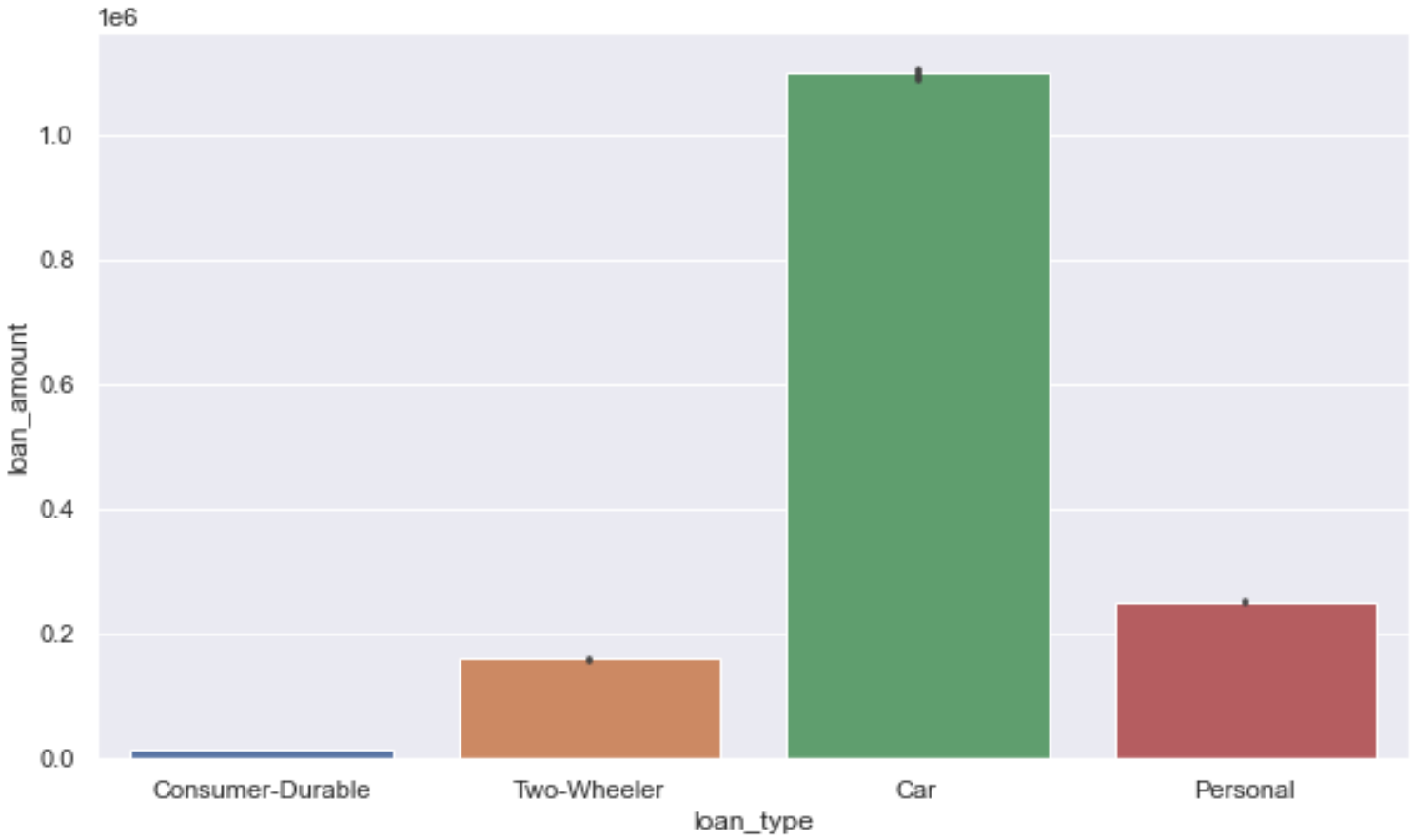
EDA:

Number of loans in Two-wheeler is higher than all others.



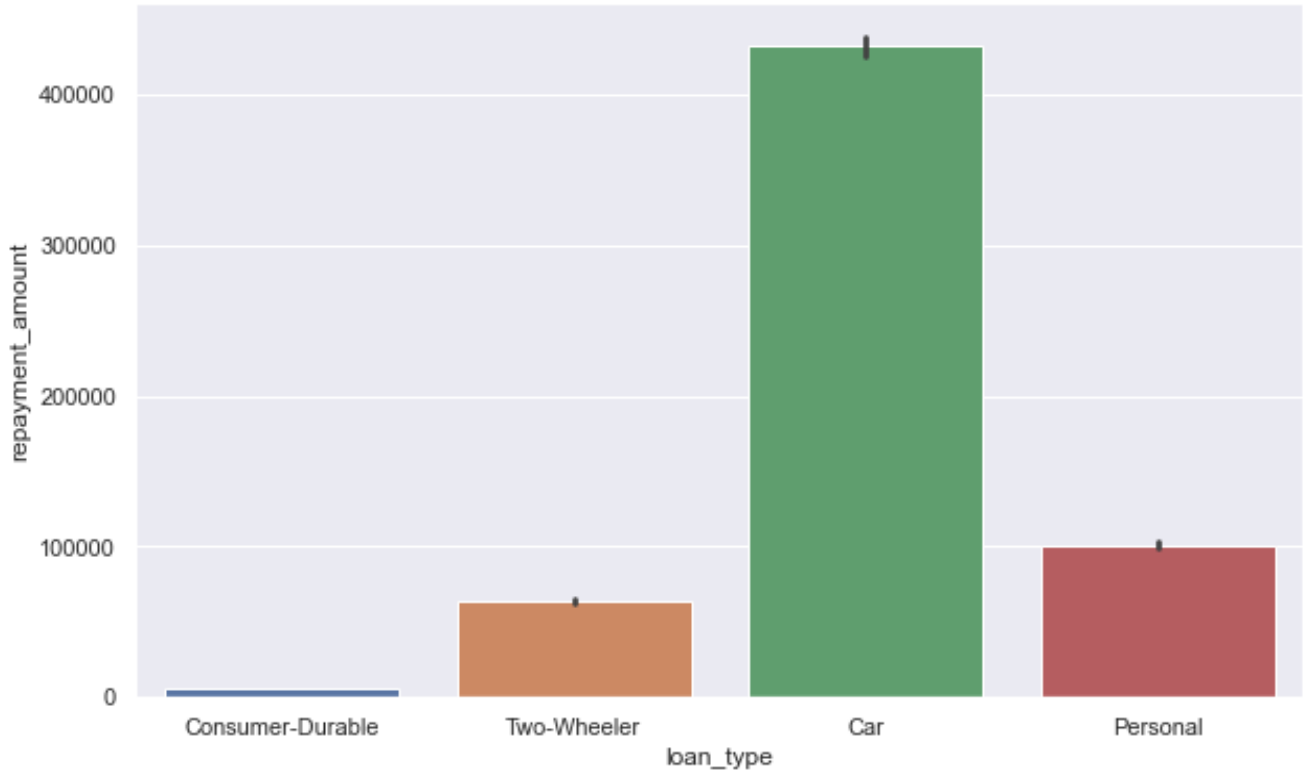
EDA:

But,the loan amount of car loan is the highest.



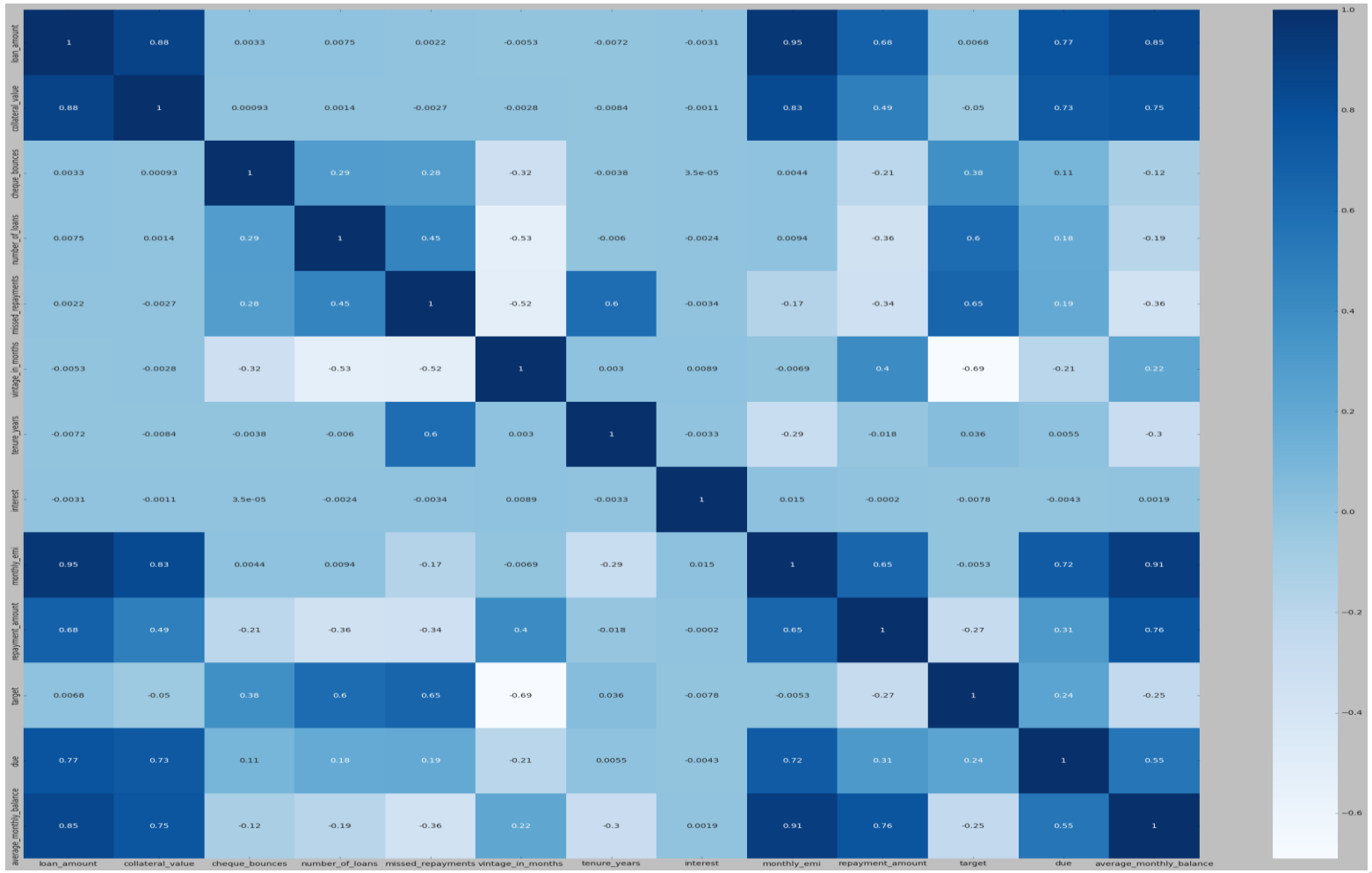
Monthly EMI:

Car loan is much higher compared to other loans.



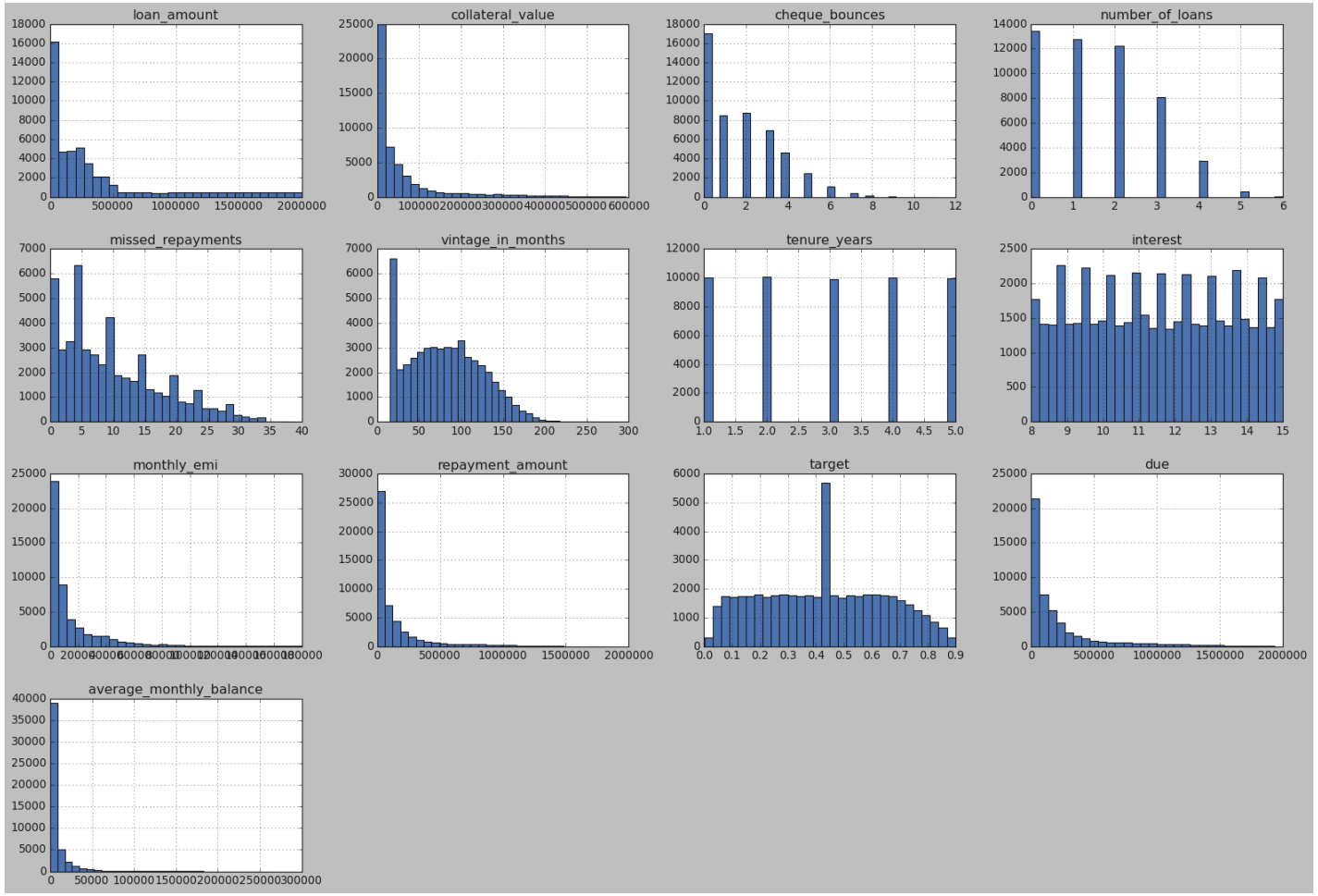
HEATMAP PRSENTATION:

Created Heatmap to understand the correlation between the variables.



GRAPHICAL PRSENTATION:

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



STEPSPERFORMED:

1) ***Used Power transformation to make numerical variables Normally distributed.***

2) Dropped unnecessary columns for model building.

3) Used One-hot encoding technique and created dummy variable for necessary categorical variables.

MODELBUILDING:

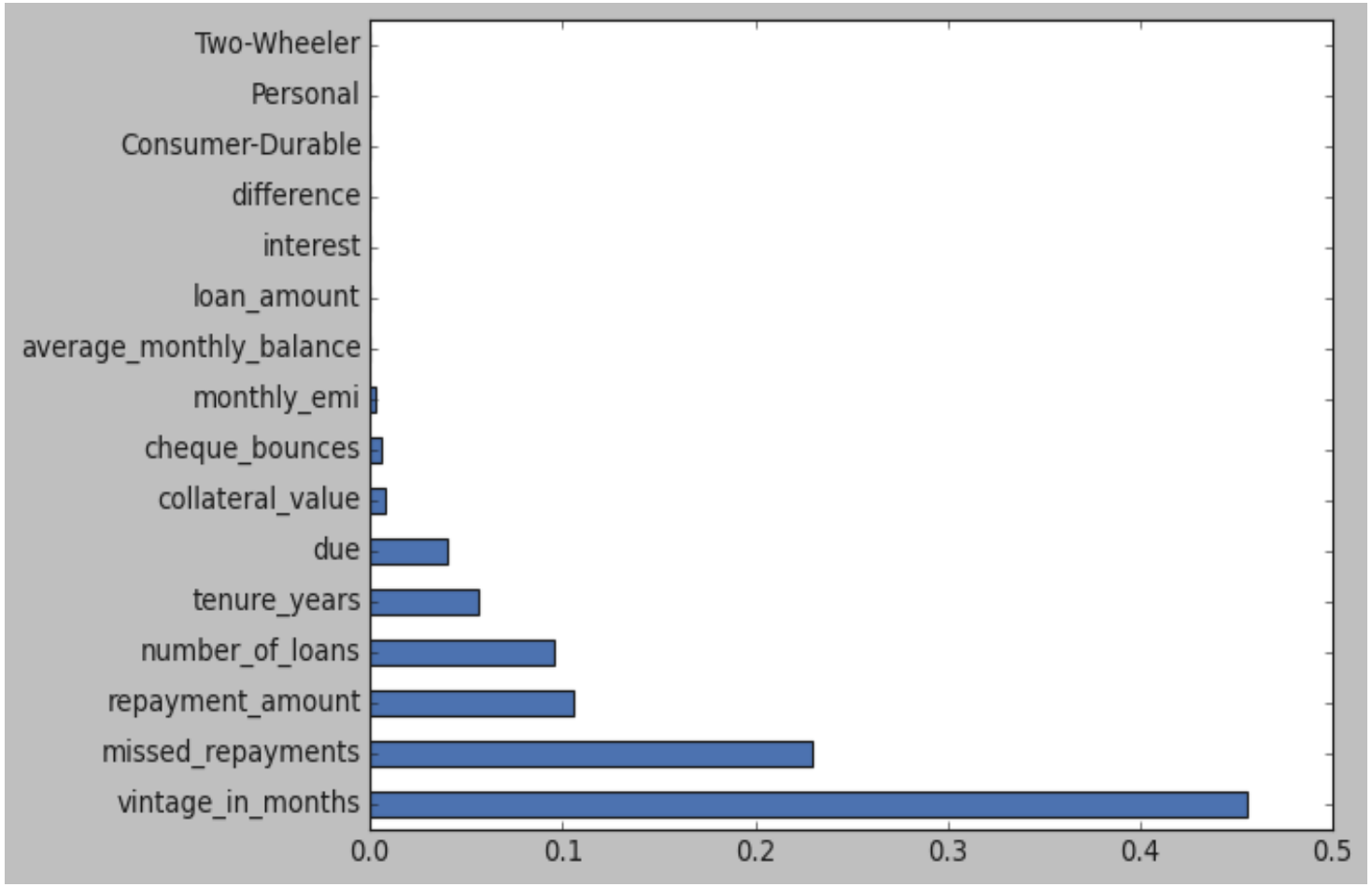
1) *Used various models like Multiple Linear Regression , Random Forest Regressor , Gradient BoostingRegressor, XGBoost Regressor, Adaboost Regressor , ElasticNet:Hybrid Regularized Model , Light GBM for model building.*

2) Used R -Squared as a performance metrics.

3)XGBoost has given us 99.5% R-squared on test data across the models

REGRESSION INTERPRETATION:

feature importance



RECOMMENDATIONS:

1) ***We should focus more on Car and Two-wheeler loan types.***

2) Missed Repayment customers with high repayment amount should be highlighted.

3) Customer's due factors and tenure are another subset of influencers to predict the Loss Given Default of the customers.

THANK YOU