Domain Oriented Case Study
BFSI Credit Risk Assignment

Team Members:
- Nirajkumar Pithva
- Harshal Naik

Understanding Credit Risk

The Basel norms, also known as the Basel Accords or Basel Regulations, are a set of international regulatory standards for the banking industry.

The history of the Basel norms can be traced back to the late 1970s and early 1980s when the banking industry was facing a series of crises and failures. These crises were caused by a combination of factors, including insufficient capital and liquidity, inadequate risk management and weak supervisory oversight. In response to these crises, the Basel Committee began to develop a set of international standards for bank capital and risk management to strengthen the resilience of the global banking system and reduce the risk of bank failures.

The Basel norms are used to **ensure that banks maintain sufficient levels of capital and liquidity to withstand financial shocks and reduce the risk of bank failures**.

To comply with the regulatory norms, a bank needs to **provision funds**. Provisioning refers to the process of setting aside funds to cover potential losses from defaulted loans. Therefore, **provisioning is an important part of a bank's risk management strategy.**

Brief about the Domain

Expected credit loss (ECL) computation is a method used in credit risk management to determine the amount of loss a bank is expected to incur in the event a borrower defaults on their loan.

The formula for ECL typically used in practice is as follows:

ECL = EAD x PD x LGD

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

Probability of default (PD) is a measure of the likelihood that a borrower will default on their loan. It is calculated based on the borrower's creditworthiness along with other factors such as their income, their timely repayments, number of loans, cheque bounces and debt-to-income ratio.

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. In a dataset containing the historical data of loans defaulted, such as the value of the collateral (if any), loan tenure, number of missed repayments, etc. the bank tries to estimate the approximate amount that it stands to lose if a borrower defaults.

The LGD value for a loan, given its collateral and assuming that the customer has already made some repayments, is given by:

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

Exposure at default (EAD) is a measure of the amount of credit extended to a borrower at the time of default. The EAD is the outstanding loan amount after the repayments received until the time of default are deducted. The EAD is a term that measures the absolute amount and does not need any statistical estimation such as PD and LGD. It is simply the loan amount subtracted from the total repayments received.

Business Objective

For this assignment, we will focus on the **Loss Given Default (LGD)** component of the ECL computation. The objective is to build a statistical model to estimate borrowers' LGD.

As a business analyst working for a bank, we have developed a model that can estimate borrowers' LGD. To develop this model, we have used relevant data sets that include information about defaulted accounts and the amount of money that has been retrieved from them using collaterals and other collection methods.

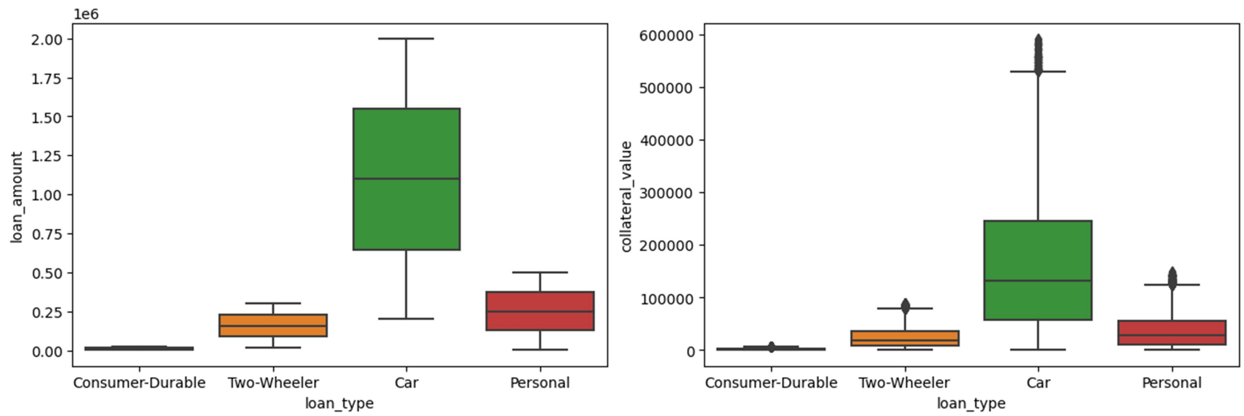
Steps followed to develop model

Following steps have been performed to understand the dataset and develop model to estimate LGD. Please refer this together with commented python program for better understanding.

- Reading and understanding data
- Cleaning and Preparing the dataset
- Exploratory Data Analysis using categorical variables
- Feature engineering
- Exploratory Data Analysis using Numerical variables
- Develop and test different linear regression models
- Finalize the model
- Reading the Test database
- Cleaning and preparing Test database
- Apply model and predict LGD

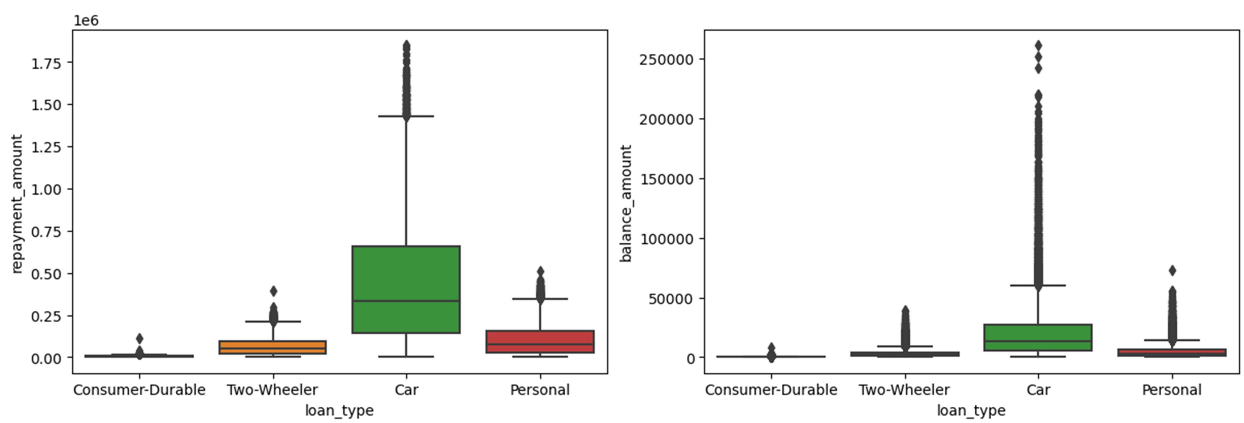
Findings from EDA

Analysis of Loan amount and Collateral value for different loan type



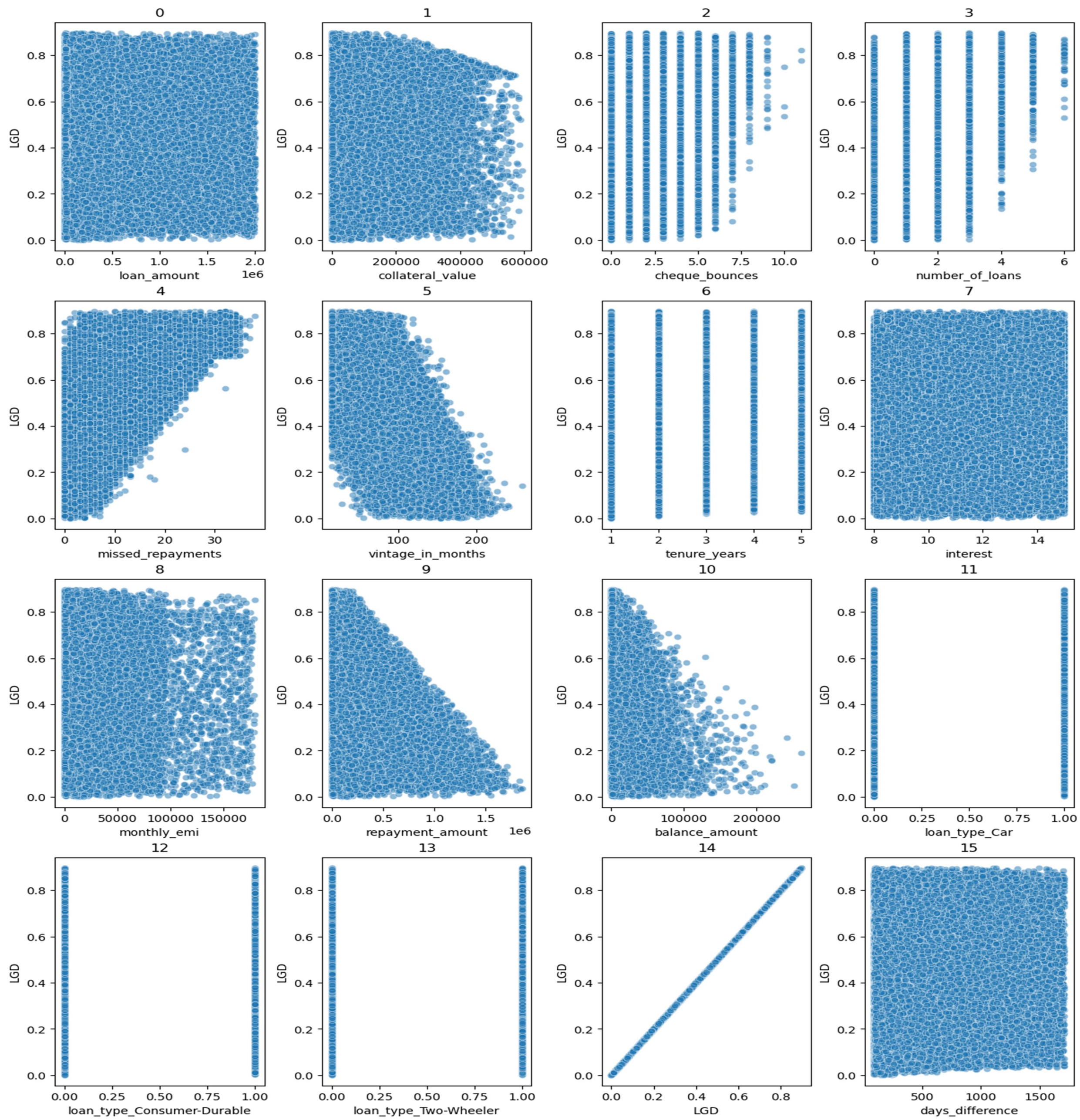
- The loan amount seems to be appropriate as line with loan_type
- Also, collateral_value shows the similar pattern.
- Though, there are some outliers, these seems within limits and it may due short loan tenure or other reason.

Analysis of repayment and balance amount for different loan type

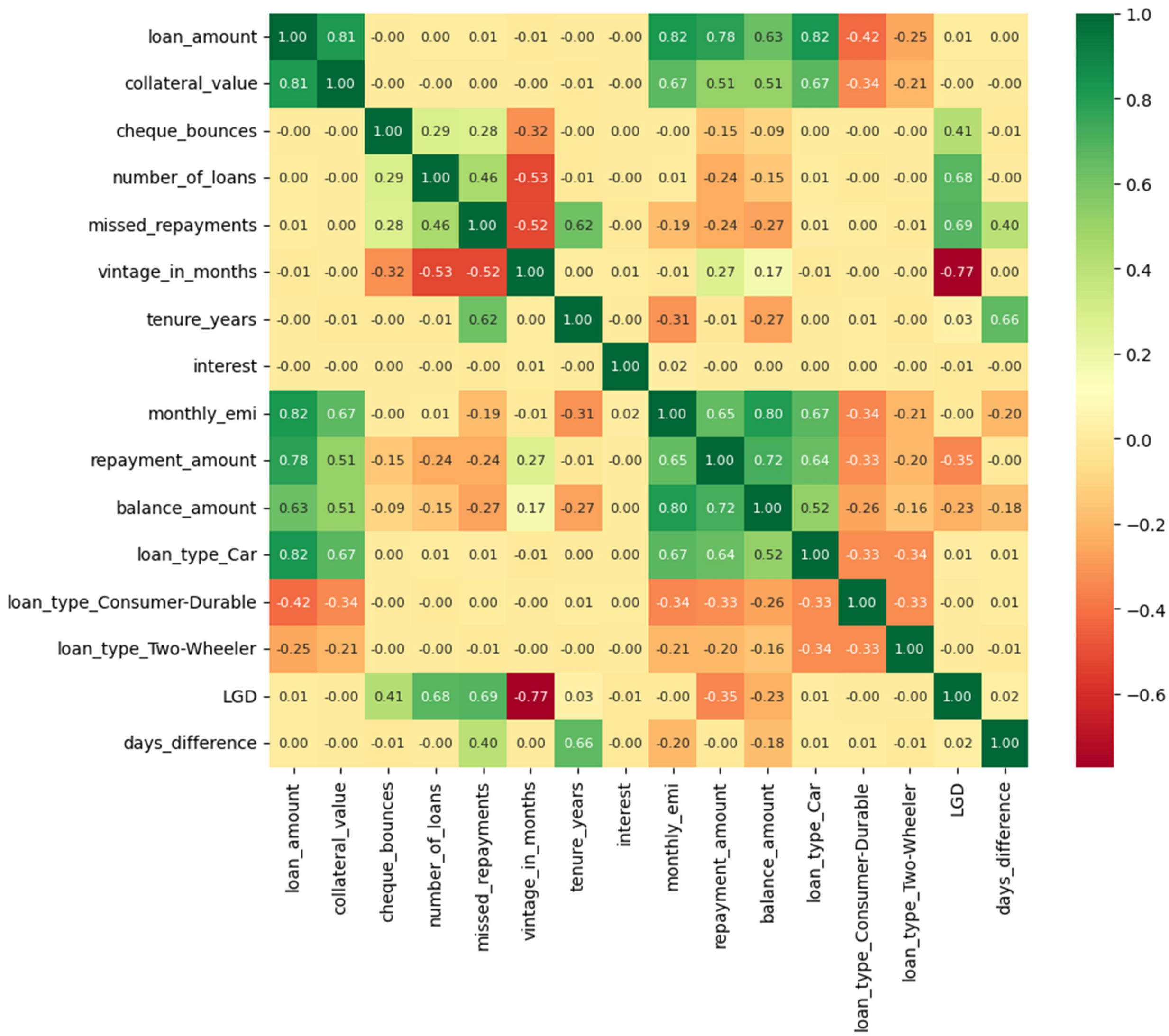


- The repayment _amount and balance_amount is in line with loan_type
- Here also, there are some outliers but that may be due to loan tenure, vintage in months or other reason.
- It does not necessarily show abnormality.

Analysis of calculated LGD with respect to other numeric variables



- LGD is highly related to collateral amount and repayment amount. This is very obvious as the equation of LGD has these variables.
- As the cheque_bounces goes beyond 6, LGD starts increasing linearly.
- As the number of loans goes 4 and beyond, LGD increases
- Number of missed repayments is also highly related to higher LGD
- As Vintage in months grow, LGD comes down.
- Also as balance amount increases, LGD is decreasing
- Loan amount, loan type, monthly emi and interest does not seem to be impacting LGD.



LGD relationship from heatmap

- LGD is negatively correlated with vintage in months with correlation factor of -0.77
- LGD is highly correlated with number of loans and missed repayments with factor of 0.68 and 0.69 respectively.
- LGD is moderately correlated with cheque bounces with factor of 0.41

Model outcome & Performance

OLS Regression Results

Dep. Variable:	LGD	R-squared:	0.824
Model:	OLS	Adj. R-squared:	0.824
Method:	Least Squares	F-statistic:	4.052e+04
Date:	Tue, 02 Jan 2024	Prob (F-statistic):	0.00
Time:	16:59:22	Log-Likelihood:	30489.
No. Observations:	34709	AIC:	-6.097e+04
Df Residuals:	34704	BIC:	-6.093e+04
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.4519	0.001	837.458	0.000	0.451	0.453
number_of_loans	0.0504	0.001	73.905	0.000	0.049	0.052
missed_repayments	0.1553	0.001	158.006	0.000	0.153	0.157
vintage_in_months	-0.0771	0.001	-104.192	0.000	-0.079	-0.076
tenure_years	-0.0878	0.001	-107.952	0.000	-0.089	-0.086

Omnibus:	78.603	Durbin-Watson:	2.011
Prob(Omnibus):	0.000	Jarque-Bera (JB):	90.882
Skew:	-0.062	Prob(JB):	1.84e-20
Kurtosis:	3.218	Cond. No.	3.47

Area where model performed well:

- The model accuracy is 82.4%, which suggests a good model.
- The p-value is also within 5%.
- The target variable LGD is explained by four factors, i.e. number_of_loans, missed_repayments, vintage_in_months and tenure_years
- This is also in line with the analysis based on scatter plot and heatmap above.
- The model also performed well on the test data with accuracy of 81.9%.

Area where model can improve:

- The good percentage of 45% is attributed to const, which implies 45% of LGD values is explained by a simple constant and could not be controlled.
- The rather better model will have more variables with more weightage explaining the LGD. This will give the leverage for bank to estimate and control LGD by controlling variables to the possible extent.

Final Recommendations

1) Bank need not have any restriction on loan type as it is not affecting the LGD values.
2) The car and personal loan may have high loan amount, as compared to other two wheeler and consumer durable loan. This will require higher provisioning; hence bank should sanction loan size according to possible provisioning.
3) The bank should consider missed repayment and cheque bounce as high risk indicators. Bank should introduce stricter measure for the loan accounts that cross five cheque bounce or missed repayments.
4) The bank should restrict loan count to maximum three at a time.
5) Bank should be target customers having higher average balance to restrict LGD.
6) In case of limited resource for recovery, bank can reduce vigilance on loan accounts with high vintage and divert it to one who defaults early during their loan tenure.
7) Bank should prefer issuing loan with higher tenure.