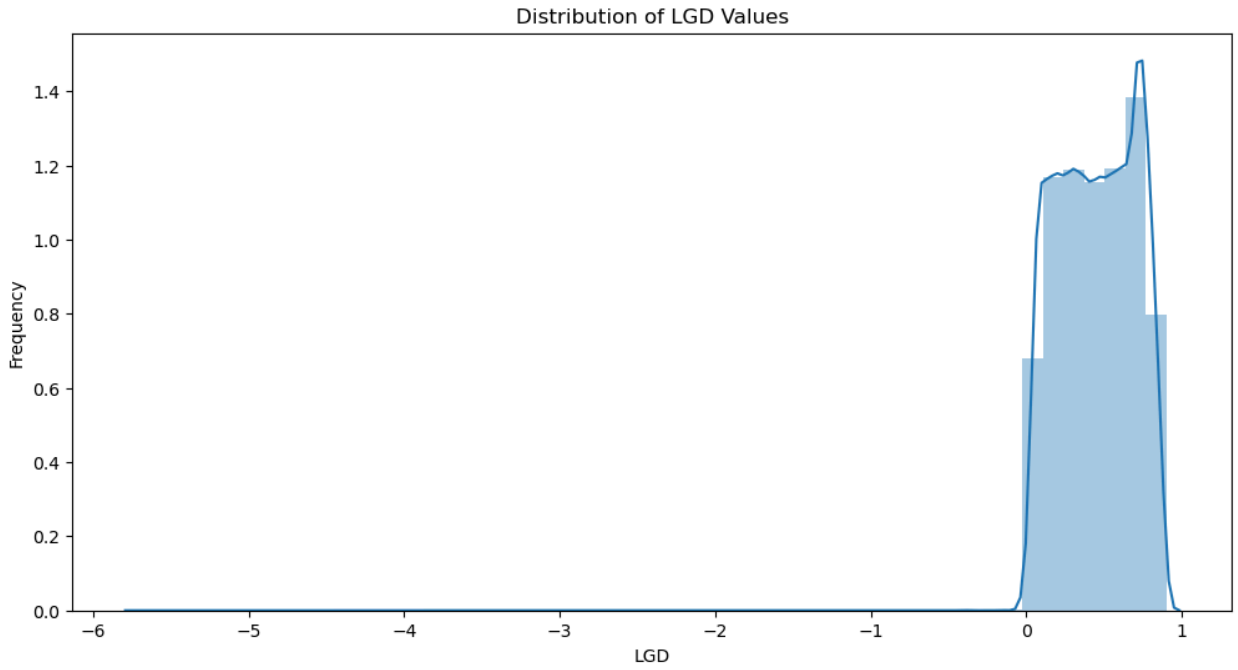
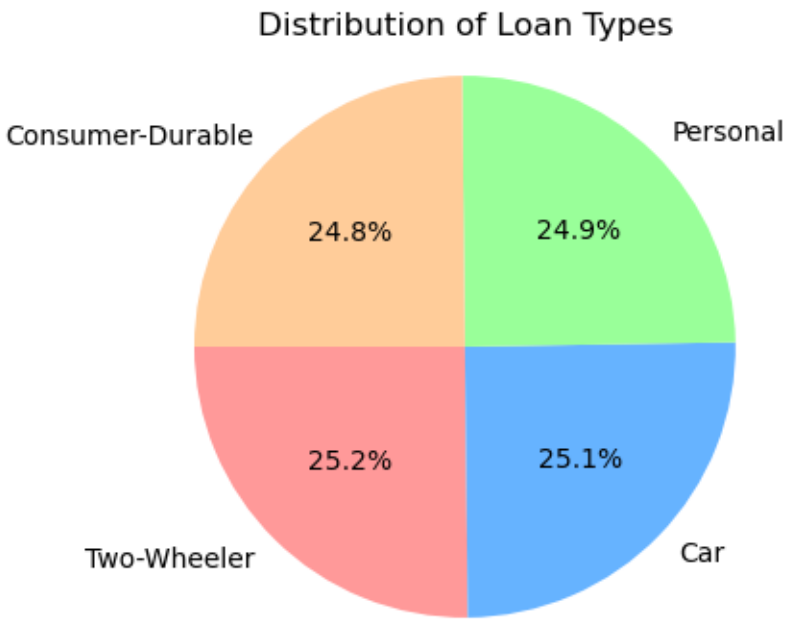
Key Insights

As per distribution of LGD values for the entire Dataset, we can see all values are ranging between 0 to 1, & distribution is as per above plot

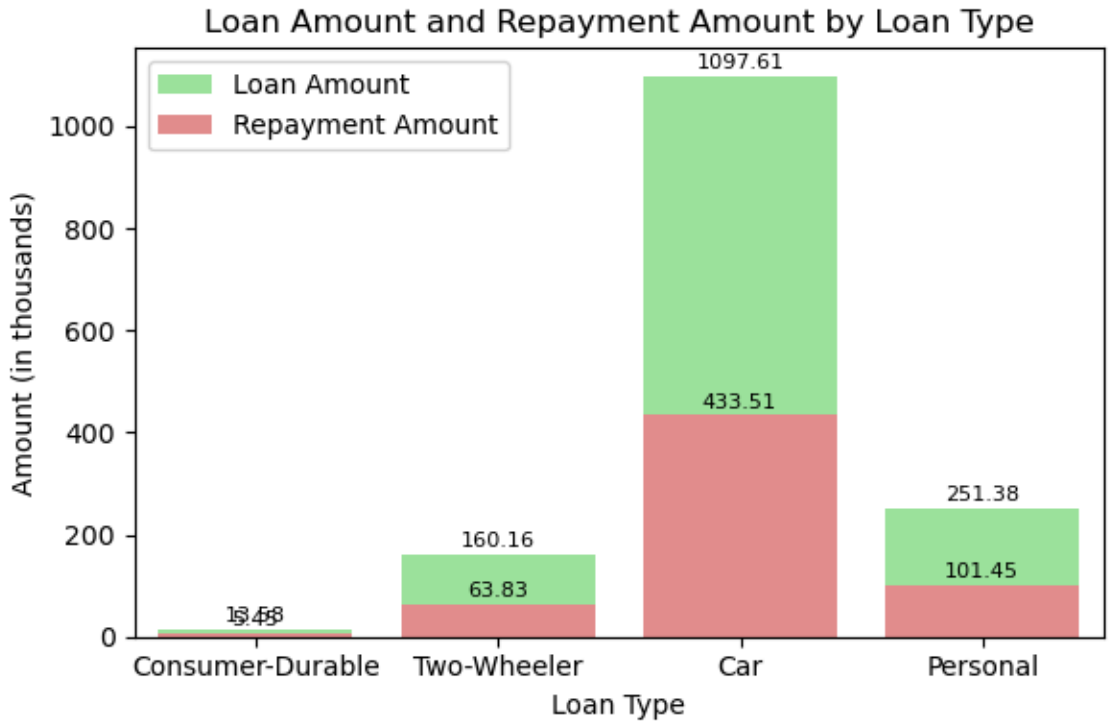




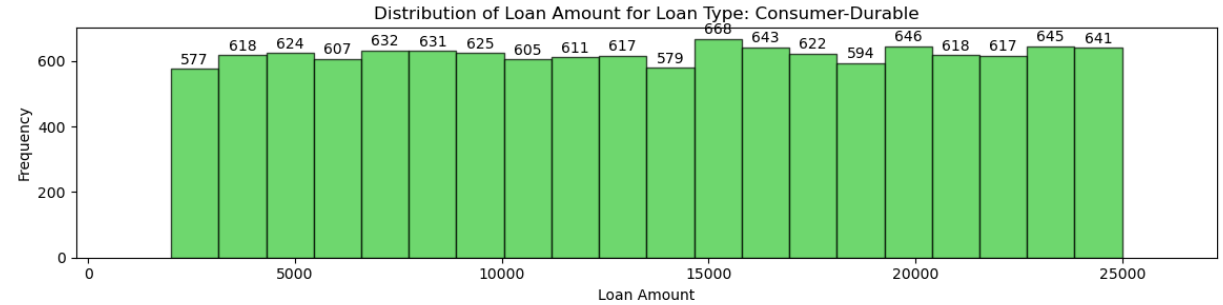
As pie chart shows all the 4 types of loan categories are mostly equally distributed

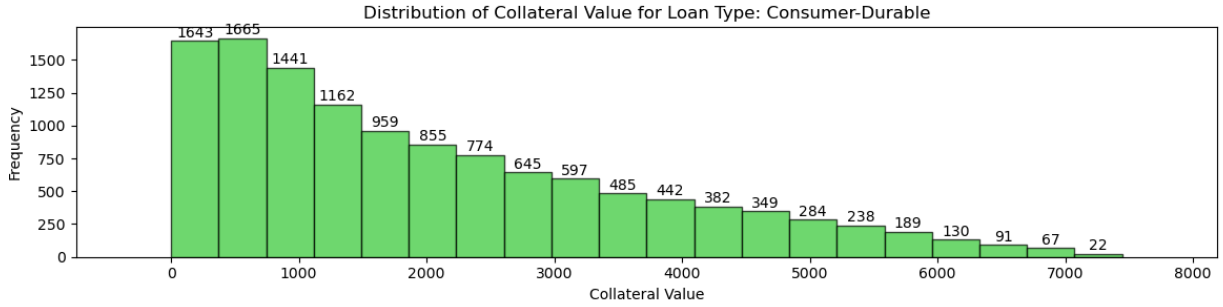
EDA Insights

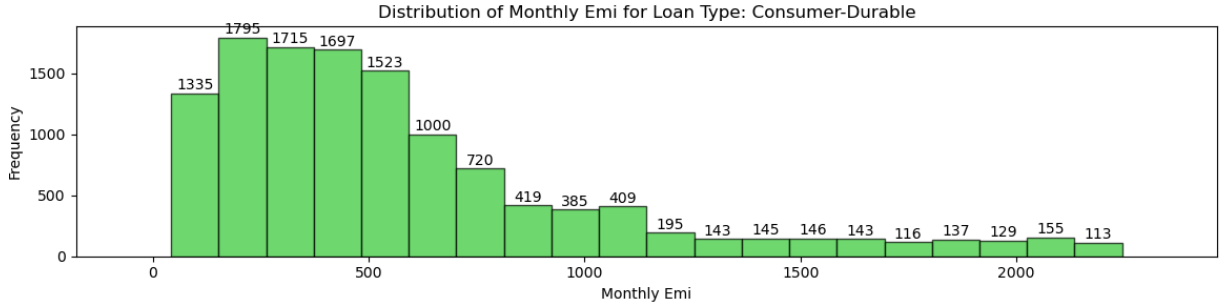
As the bar graph below shows, most of the loans are for car while the least are for consumer-durables.

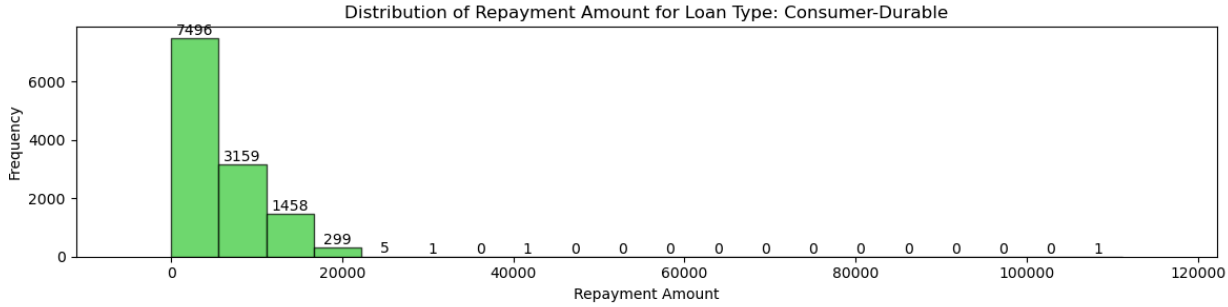


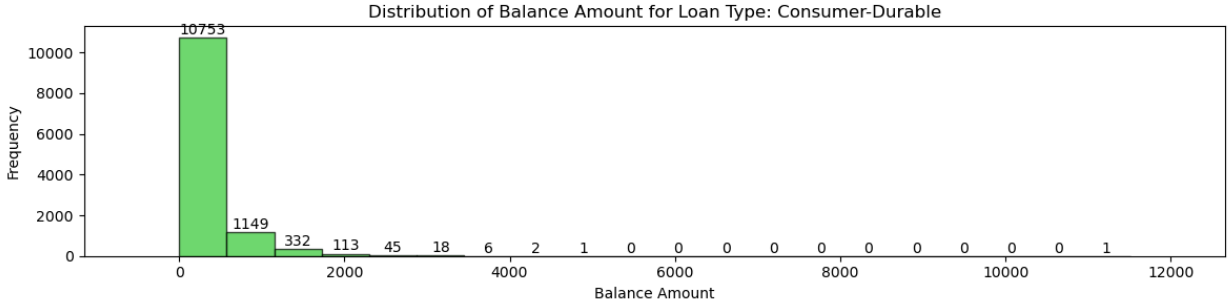
Below histograms describe the distribution of different parameters present in the dataset

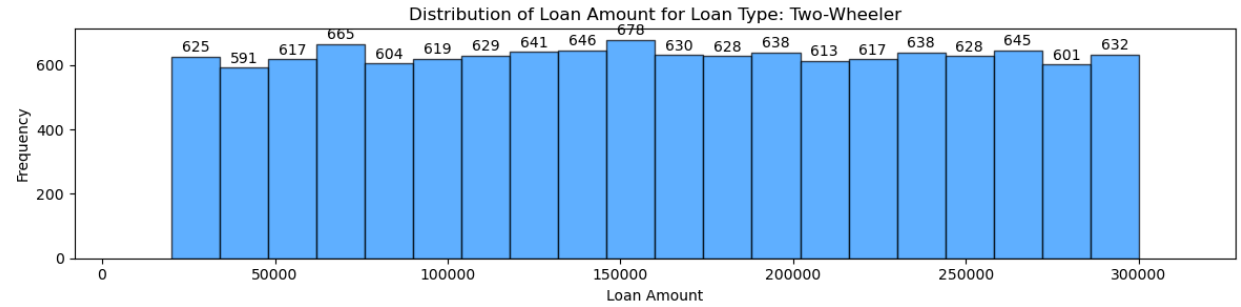


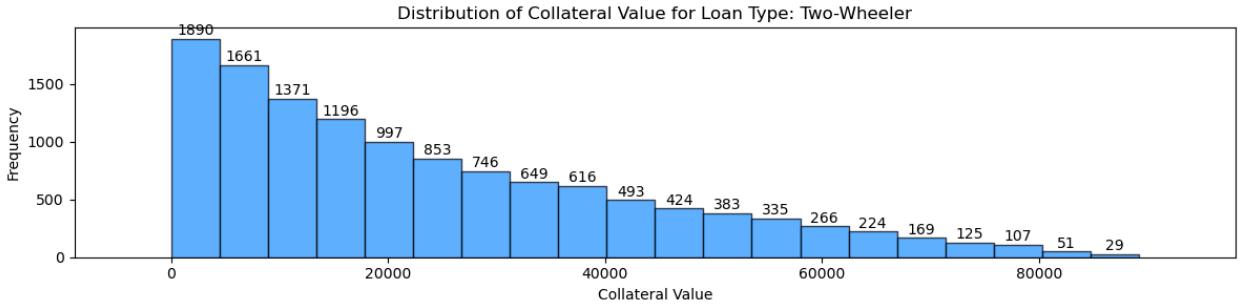


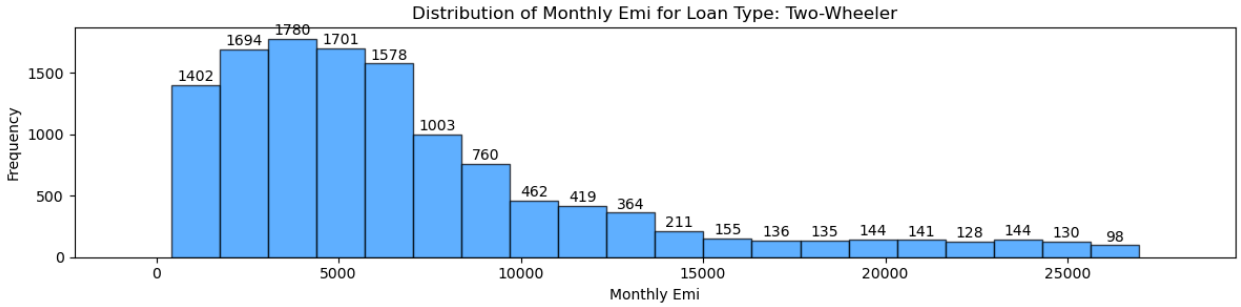


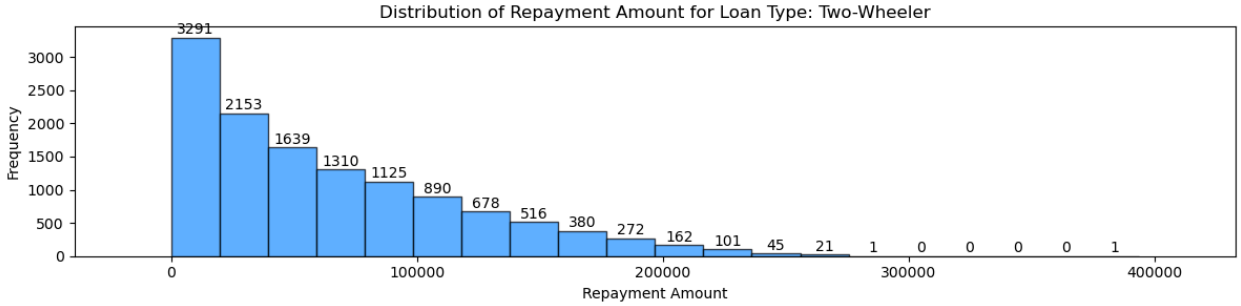


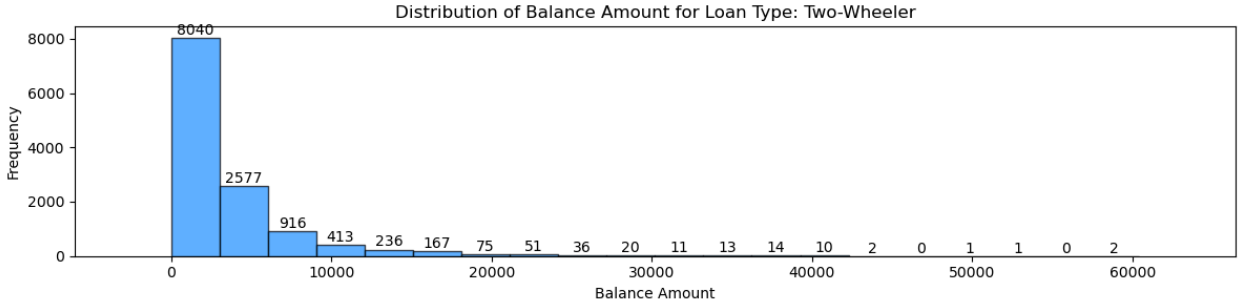


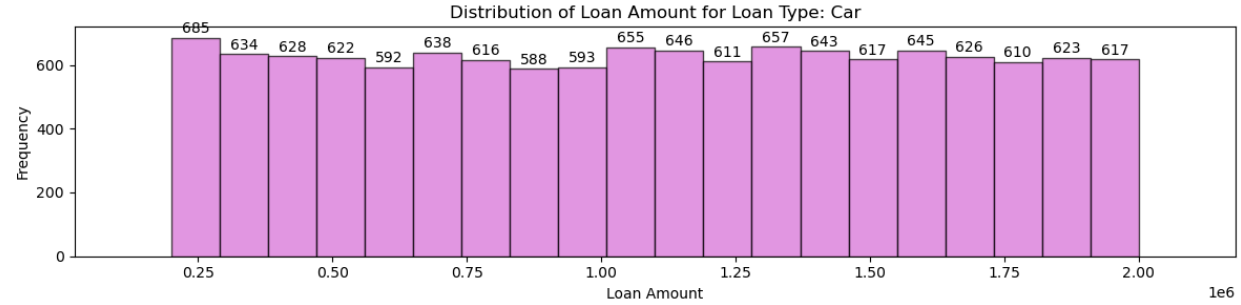

Distribution of Loan Amount for Loan Type: Two-Wheeler
Frequency
Loan Amount
625 591 617 665 604 619 629 641 646 678 630 628 638 613 617 638 628 645 601 632
0 50000 100000 150000 200000 250000 300000
0 200 400 600

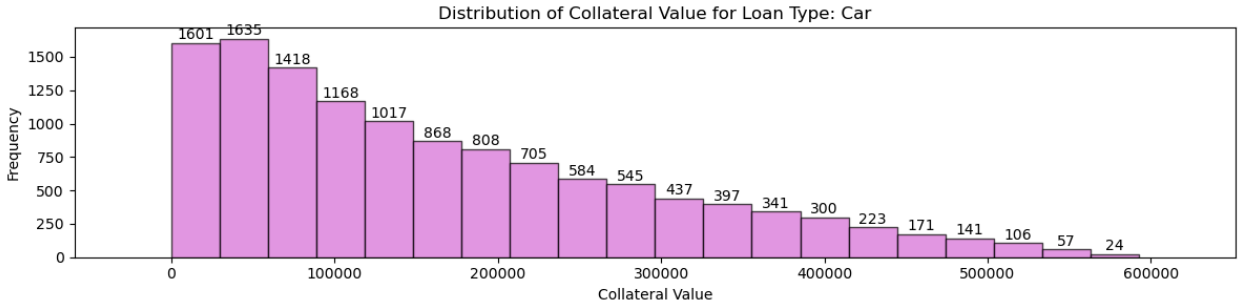

Distribution of Collateral Value for Loan Type: Two-Wheeler
Frequency
Collateral Value
1890 1661 1371 1196 997 853 746 649 616 493 424 383 335 266 224 169 125 107 51 29
0 20000 40000 60000 80000
0 500 1000 1500

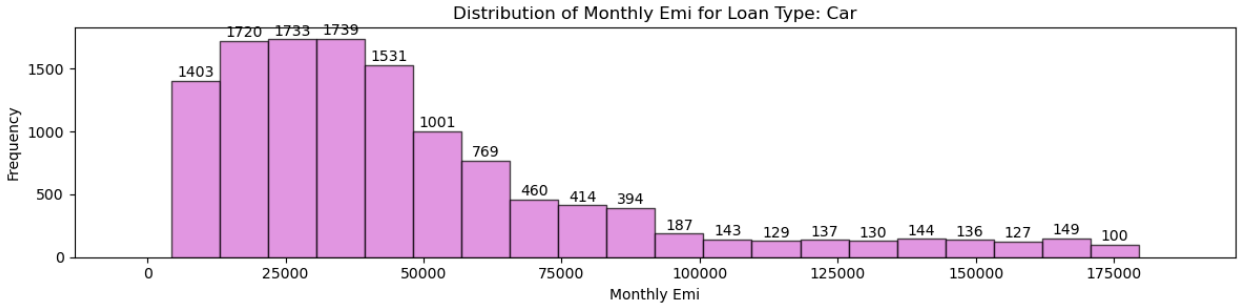

Distribution of Monthly Emi for Loan Type: Two-Wheeler
Frequency
Monthly Emi
1402 1694 1780 1701 1578 1003 760 462 419 364 211 155 136 135 144 141 128 144 130 98
0 5000 10000 15000 20000 25000
0 500 1000 1500

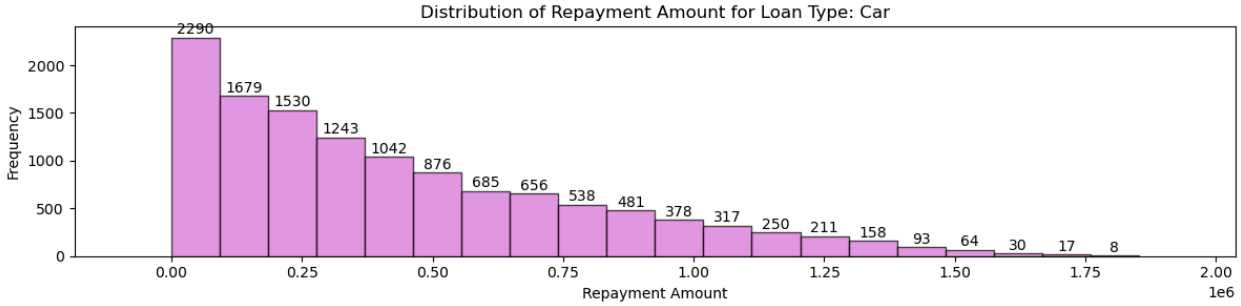

Distribution of Repayment Amount for Loan Type: Two-Wheeler
Frequency
Repayment Amount
3291 2153 1639 1310 1125 890 678 516 380 272 162 101 45 21 1 0 0 0 0 1
0 100000 200000 300000 400000
0 500 1000 1500 2000 2500 3000

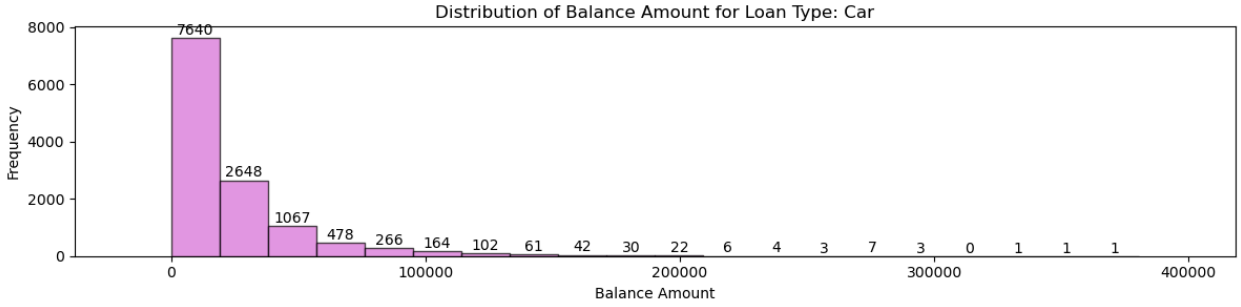

Distribution of Balance Amount for Loan Type: Two-Wheeler
Frequency
Balance Amount
8040 2577 916 413 236 167 75 51 36 20 11 13 14 10 2 0 1 1 0 2
0 10000 20000 30000 40000 50000 60000
0 2000 4000 6000 8000

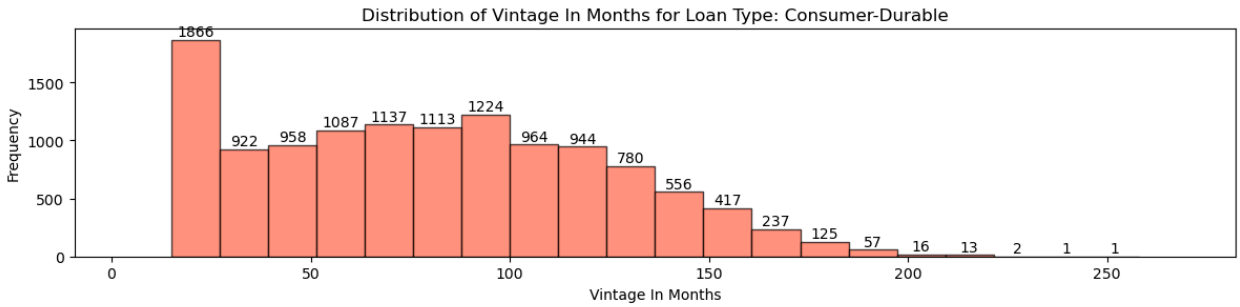

Distribution of Loan Amount for Loan Type: Car
Frequency
Loan Amount
1e6
685 634 628 622 592 638 616 588 593 655 646 611 657 643 617 645 626 610 623 617

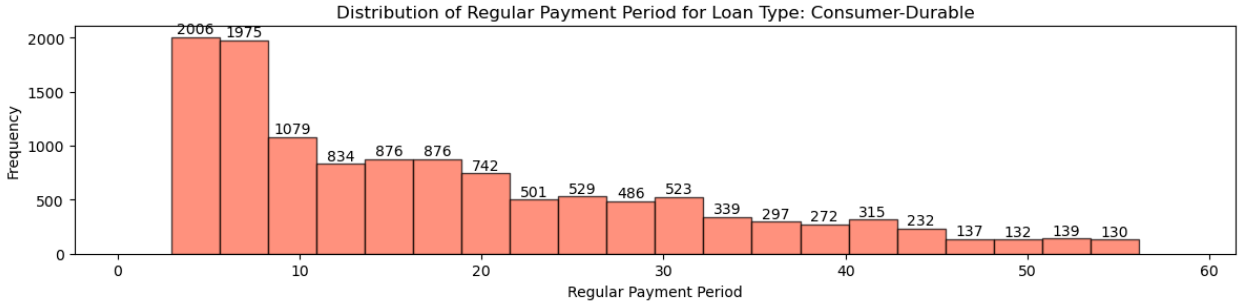

Distribution of Collateral Value for Loan Type: Car
Frequency
Collateral Value
1601 1635 1418 1168 1017 868 808 705 584 545 437 397 341 300 223 171 141 106 57 24

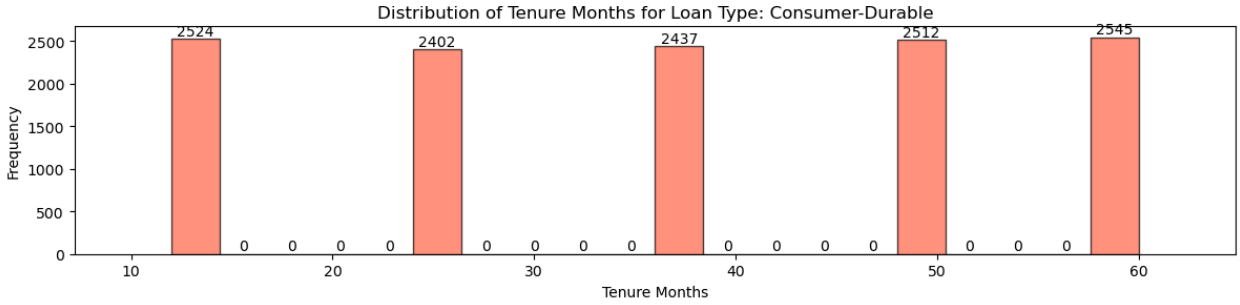

Distribution of Monthly Emi for Loan Type: Car
Frequency
Monthly Emi
1403 1720 1733 1739 1531 1001 769 460 414 394 187 143 129 137 130 144 136 127 149 100

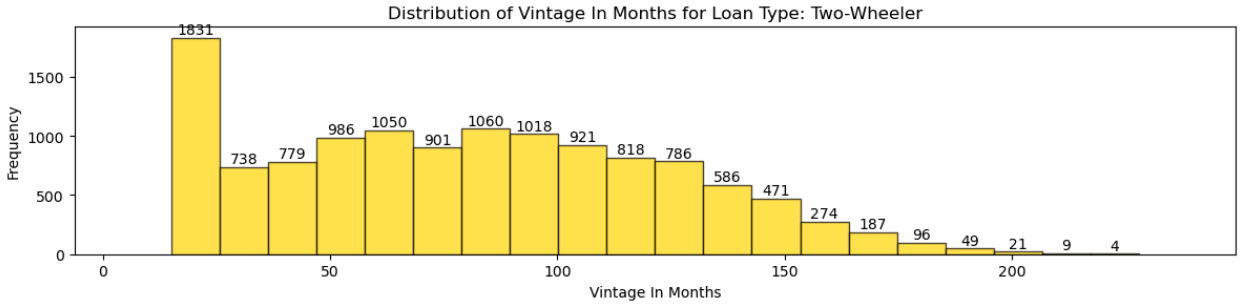

Distribution of Repayment Amount for Loan Type: Car
Frequency
Repayment Amount
1e6
2290 1679 1530 1243 1042 876 685 656 538 481 378 317 250 211 158 93 64 30 17 8

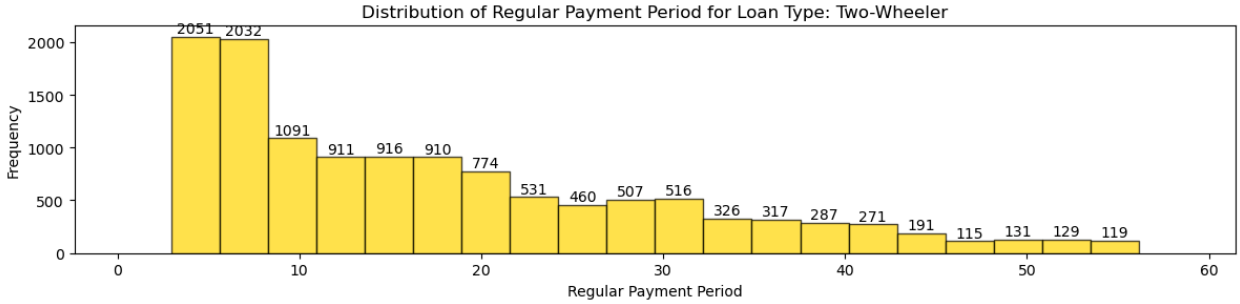

Distribution of Balance Amount for Loan Type: Car
Frequency
Balance Amount
7640 2648 1067 478 266 164 102 61 42 30 22 6 4 3 7 3 0 1 1 1

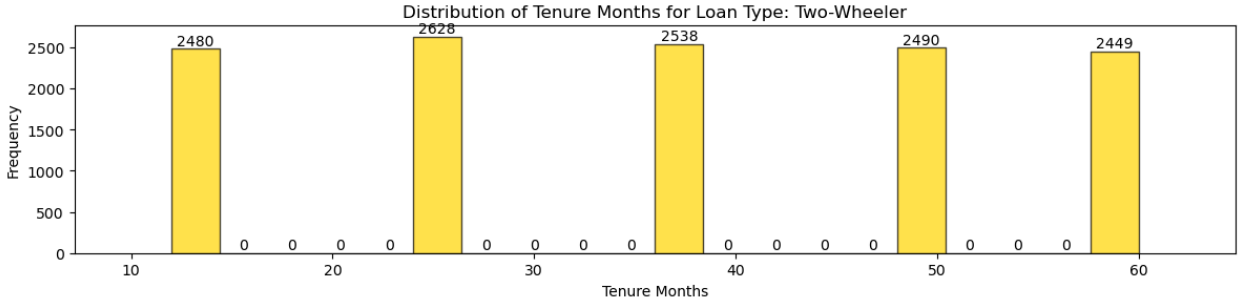

Distribution of Vintage In Months for Loan Type: Consumer-Durable
Frequency
Vintage In Months
1866
922
958
1087
1137
1113
1224
964
944
780
556
417
237
125
57
16
13
2
1
1

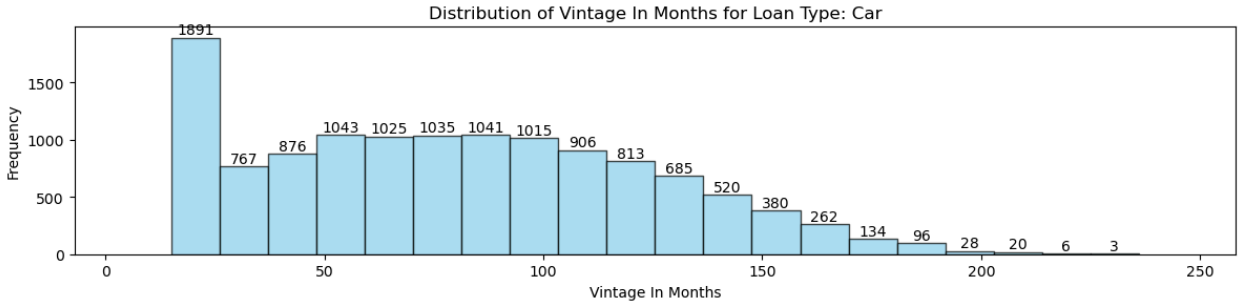

Distribution of Regular Payment Period for Loan Type: Consumer-Durable
Frequency
Regular Payment Period
2006
1975
1079
834
876
876
742
501
529
486
523
339
297
272
315
232
137
132
139
130

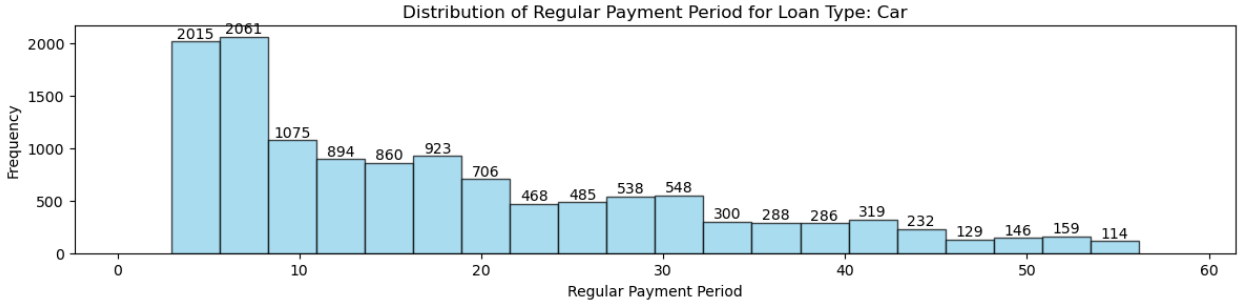

Distribution of Tenure Months for Loan Type: Consumer-Durable
Frequency
Tenure Months
2524
2402
2437
2512
2545

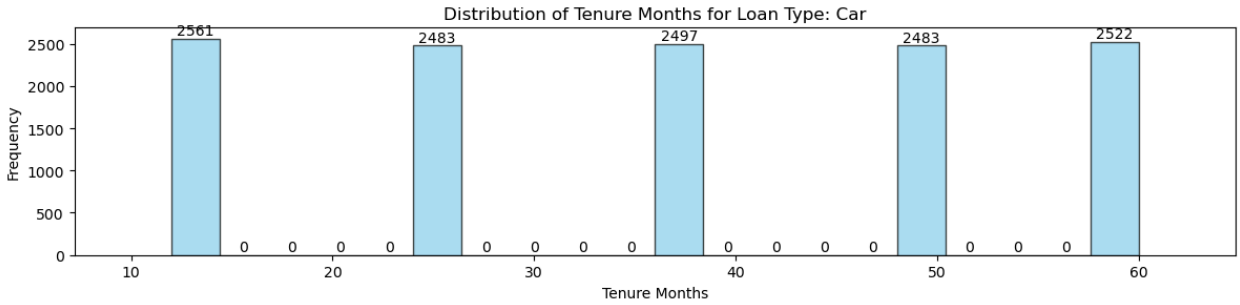

Distribution of Vintage In Months for Loan Type: Two-Wheeler
Frequency
Vintage In Months
1831
738
779
986
1050
901
1060
1018
921
818
786
586
471
274
187
96
49
21
9
4

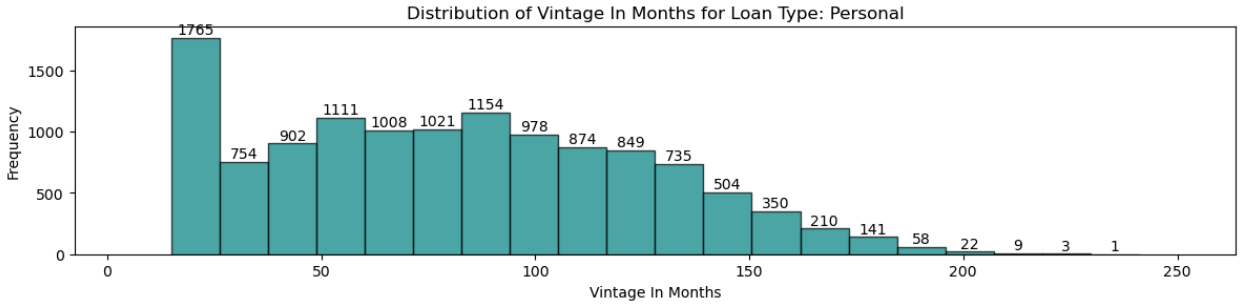

Distribution of Regular Payment Period for Loan Type: Two-Wheeler
Frequency
Regular Payment Period
2051
2032
1091
911
916
910
774
531
460
507
516
326
317
287
271
191
115
131
129
119

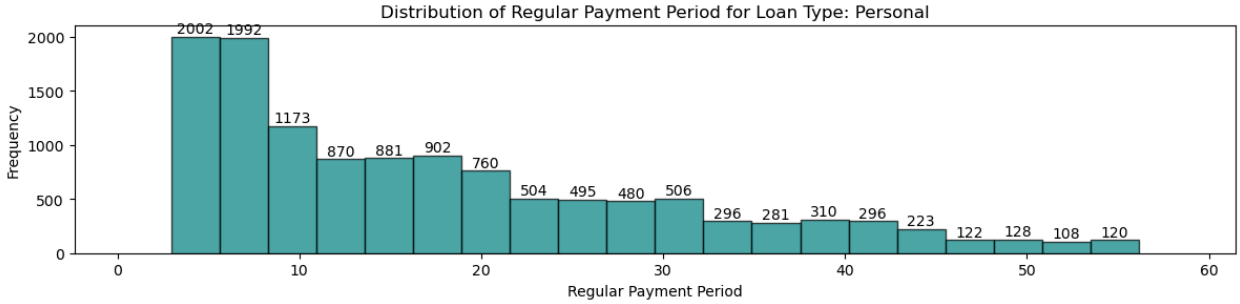

Distribution of Tenure Months for Loan Type: Two-Wheeler
Frequency
Tenure Months
2480
2628
2538
2490
2449

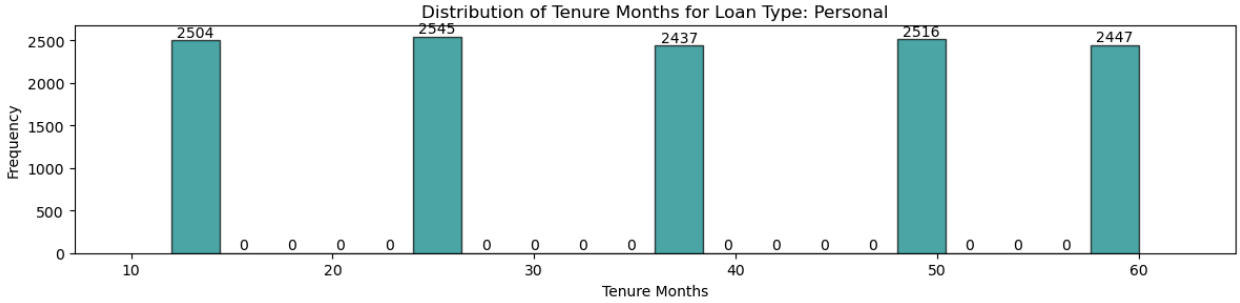

Distribution of Vintage In Months for Loan Type: Car
Frequency
Vintage In Months
1891
767
876
1043
1025
1035
1041
1015
906
813
685
520
380
262
134
96
28
20
6
3

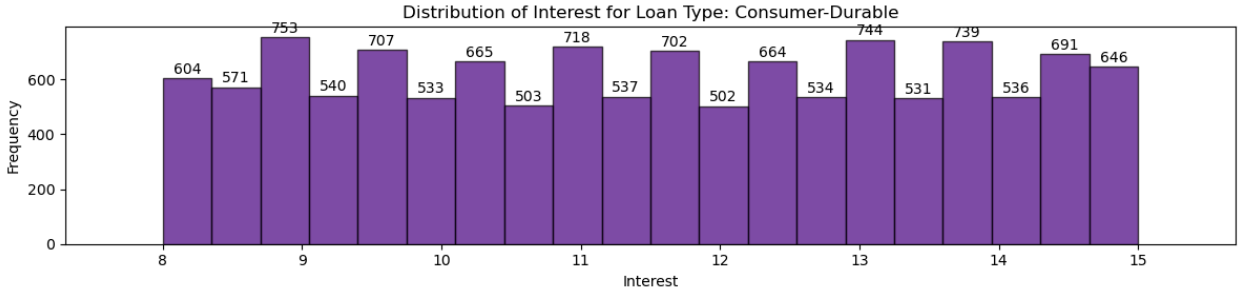

Distribution of Regular Payment Period for Loan Type: Car
Frequency
Regular Payment Period
2015
2061
1075
894
860
923
706
468
485
538
548
300
288
286
319
232
129
146
159
114

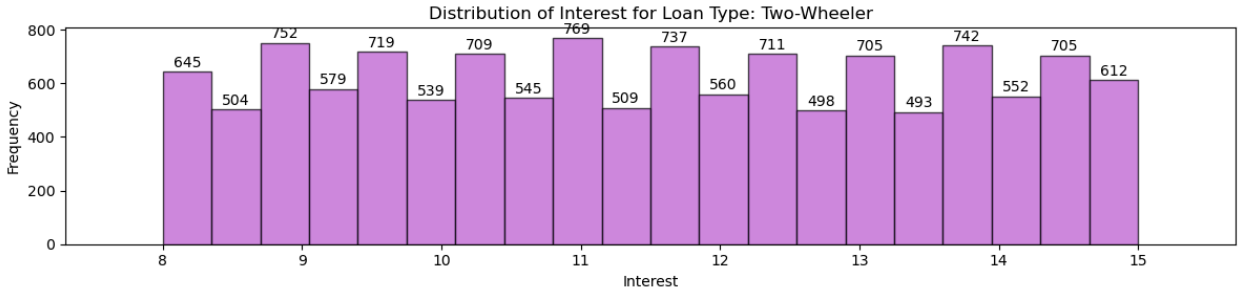

Distribution of Tenure Months for Loan Type: Car
Frequency
Tenure Months
2561
2483
2497
2483
2522

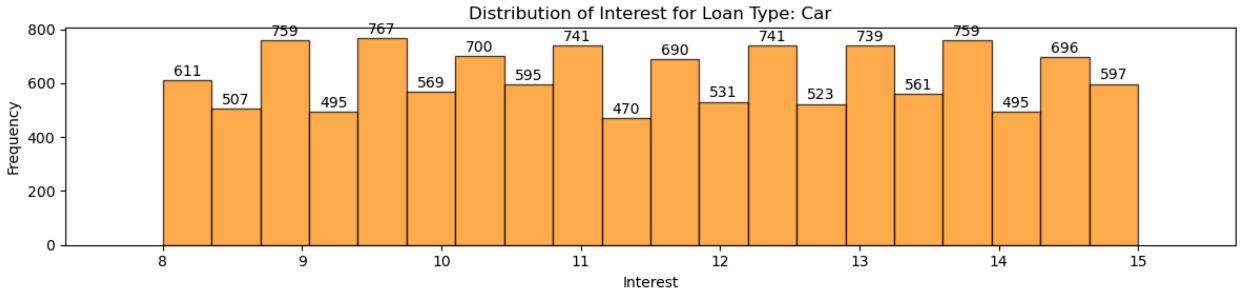

Distribution of Vintage In Months for Loan Type: Personal
Frequency
Vintage In Months
1765
754
902
1111
1008
1021
1154
978
874
849
735
504
350
210
141
58
22
9
3
1

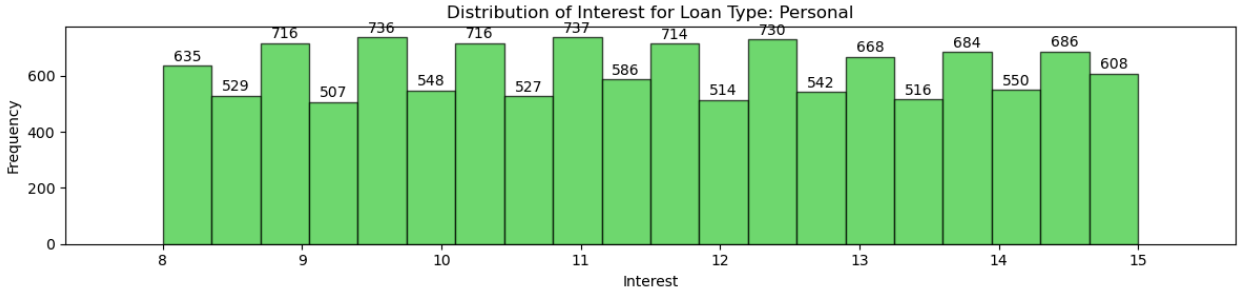

Distribution of Regular Payment Period for Loan Type: Personal
Frequency
Regular Payment Period
2002
1992
1173
870
881
902
760
504
495
480
506
296
281
310
296
223
122
128
108
120

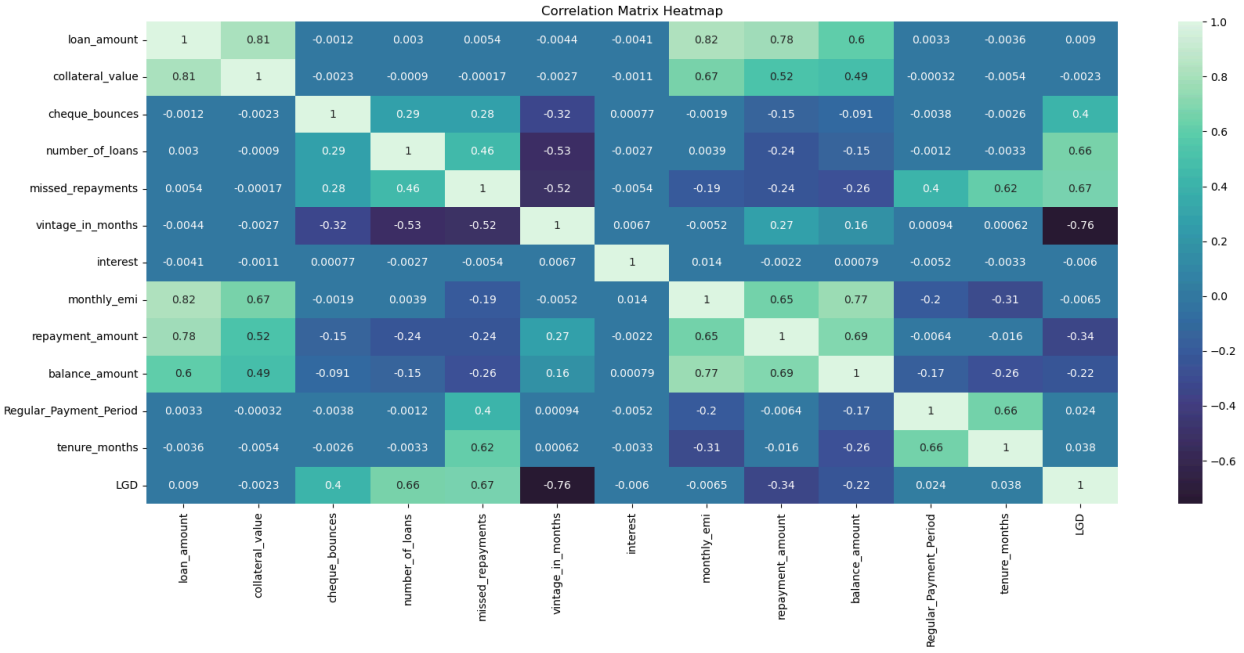

Distribution of Tenure Months for Loan Type: Personal
Frequency
Tenure Months
2504
2545
2437
2516
2447

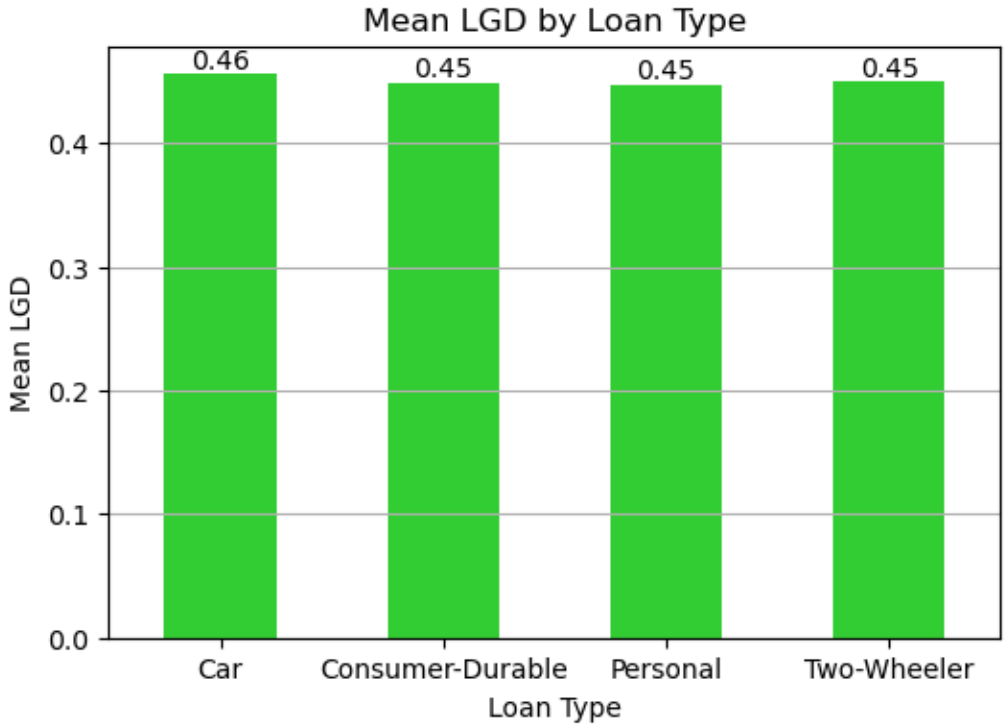

Distribution of Interest for Loan Type: Consumer-Durable
Frequency
Interest
604
571
753
540
707
533
665
503
718
537
702
502
664
534
744
531
739
536
691
646


Distribution of Interest for Loan Type: Two-Wheeler
Frequency
Interest
0
200
400
600
800
8
9
10
11
12
13
14
15
645
504
752
579
719
539
709
545
769
509
737
560
711
498
705
493
742
552
705
612

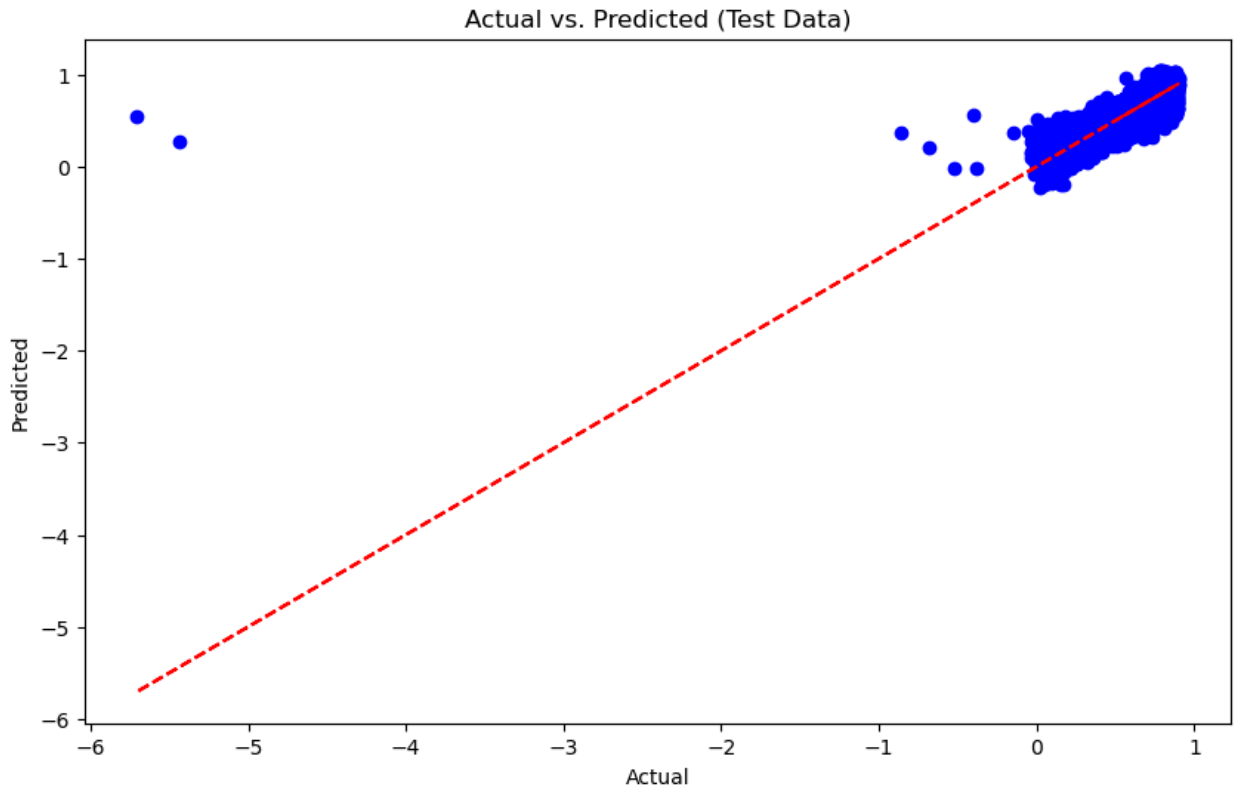

Distribution of Interest for Loan Type: Car
Frequency
Interest
0
200
400
600
800
8
9
10
11
12
13
14
15
611
507
759
495
767
569
700
595
741
470
690
531
741
523
739
561
759
495
696
597

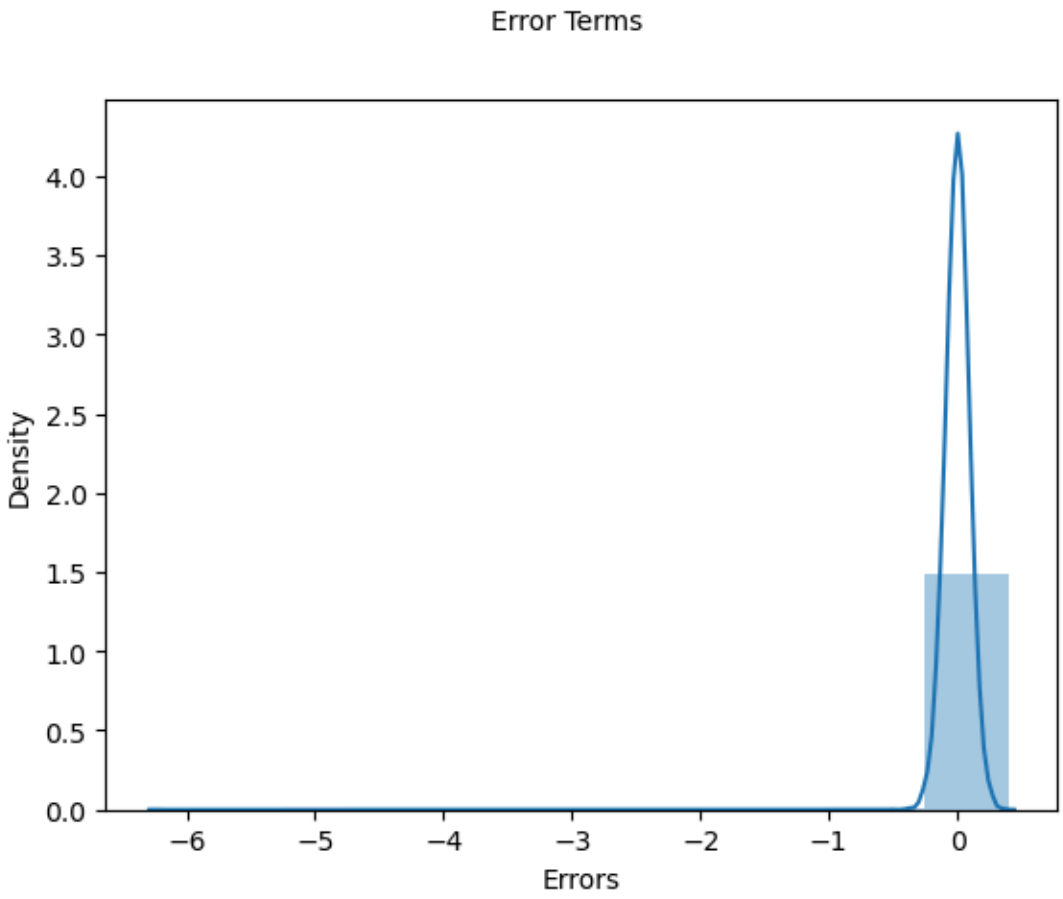

Distribution of Interest for Loan Type: Personal
Frequency
Interest
0
200
400
600
8
9
10
11
12
13
14
15
635
529
716
507
736
548
716
527
737
586
714
514
730
542
668
516
684
550
686
608





As we can see, all loan types contribute equally to the overall distribution. Therefore, this categorical input variable will exhibit multicollinearity during model building. Consequently, we will delete this colum

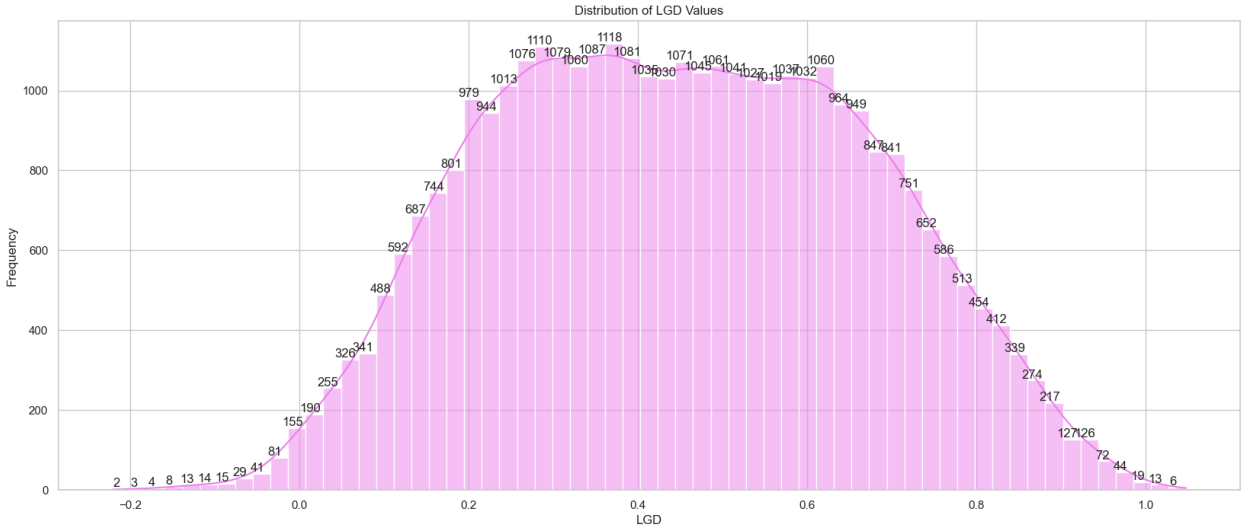

Actual vs. Predicted (Test Data)
Predicted
Actual
1
0
−1
−2
−3
−4
−5
−6
−6
−5
−4
−3
−2
−1
0
1

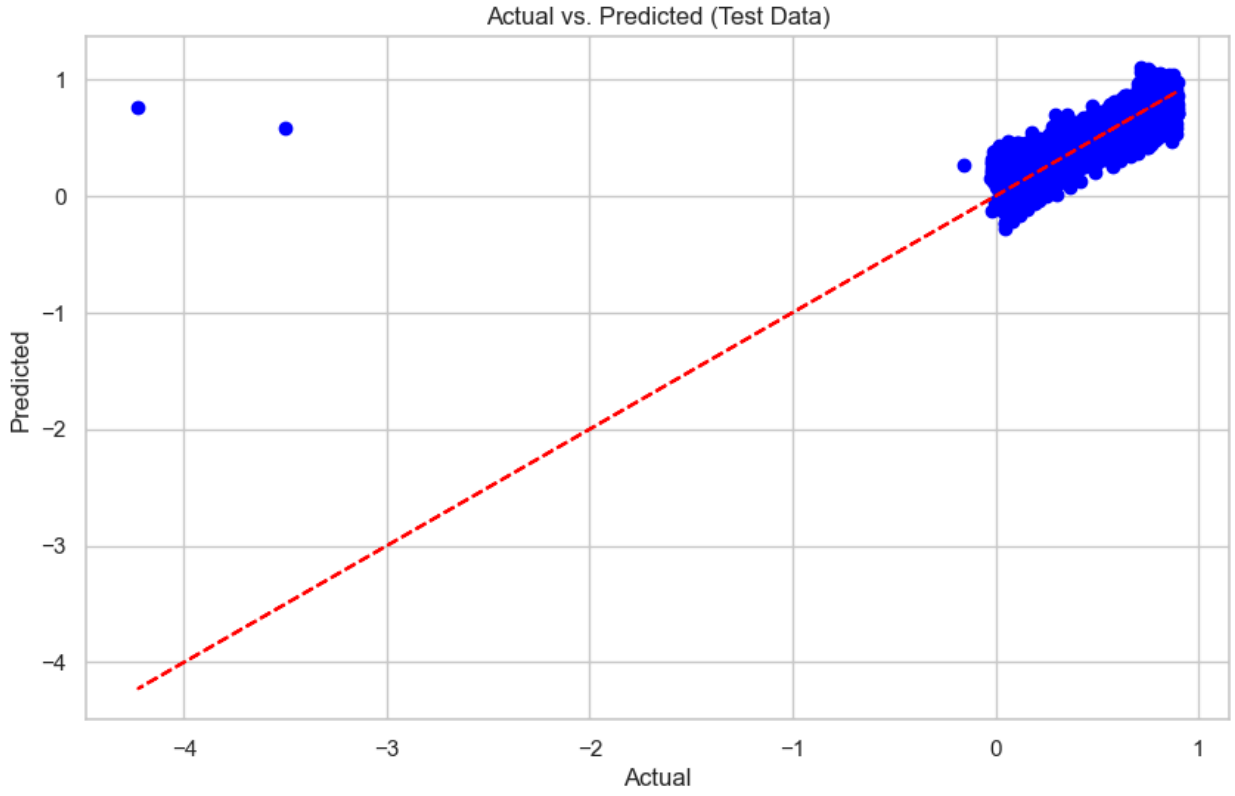


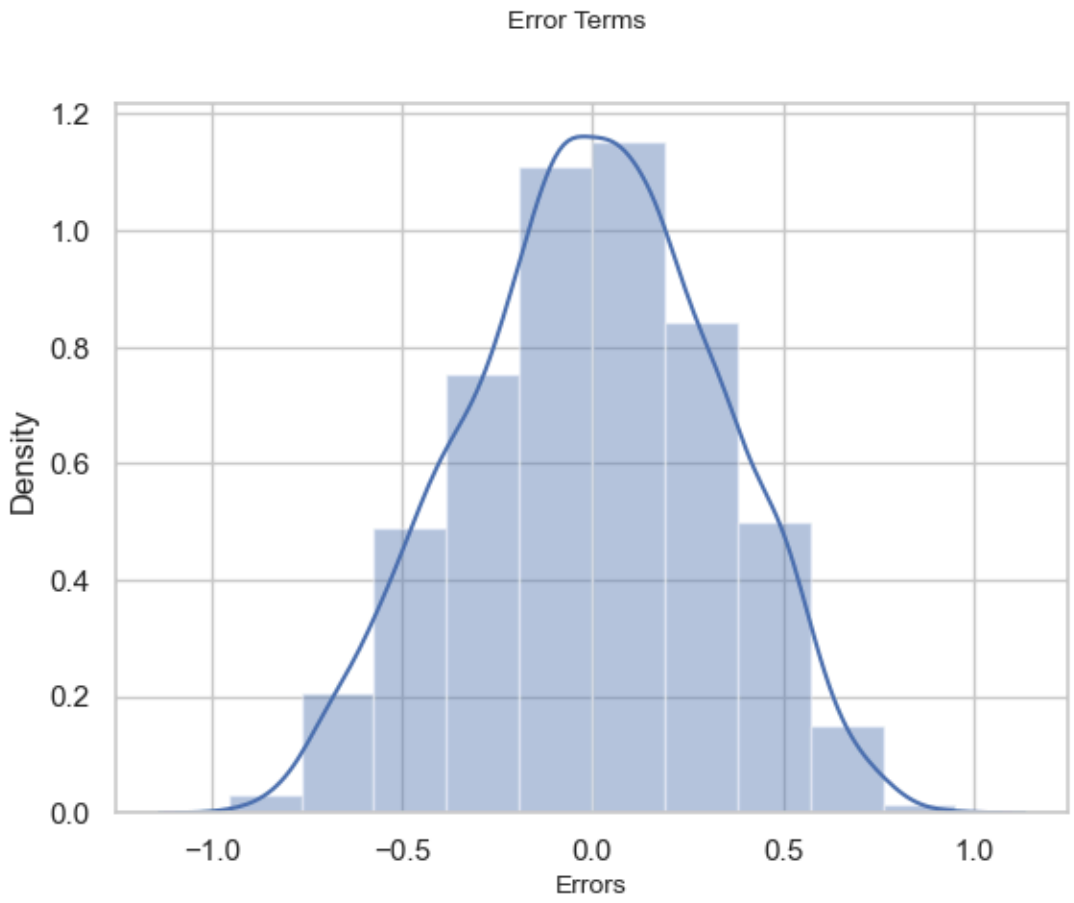
Model Fit: The model explains approximately 81.19% of the variability in LGD (Loss Given Default), indicating a good fit

Prediction Accuracy: The model's predictions are quite accurate, with the mean squared error accounting for about 18.81% of the maximum possible variance in LGD

Error Magnitude: On average, the model's predictions deviate by only about 1.15% from the actual LGD values, relative to the range of LGD values


Distribution of LGD Values
Frequency
LGD
0
200
400
600
800
1000
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
2 3 4 8 13 14 15 29 41 81 155 190 255 326 341 488 592 687 744 801 944 979 1013 1076 1079 1060 1087 1081 1118 1110 1035 1030 1071 1045 1061 1041 1027 1019 1037 1032 1060 964 949 847 841 751 652 586 513 454 412 339 274 217 127 126 72 44 19 13 6


Actual vs. Predicted (Test Data)
Predicted
Actual
1
0
−1
−2
−3
−4
−4
−3
−2
−1
0
1

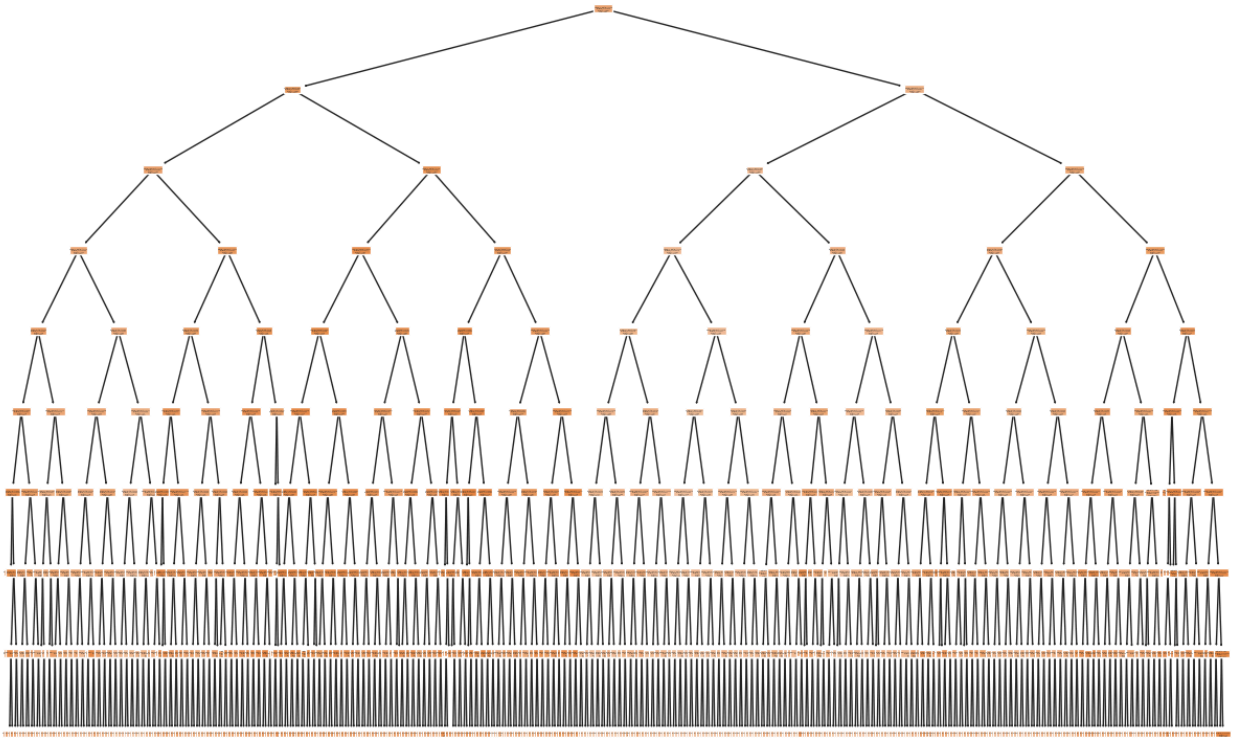


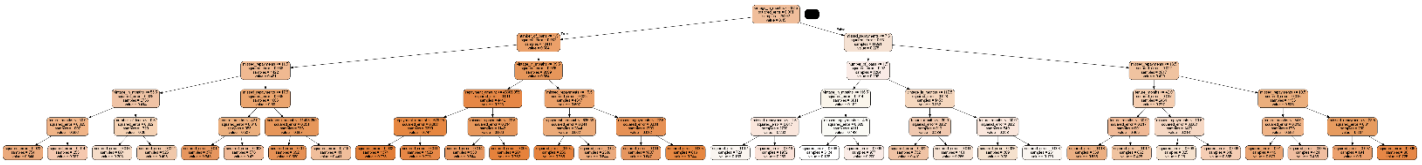
Model Fit: The model explains approximately 79.74% of the variability in LGD (Loss Given Default), indicating a reasonably good fit

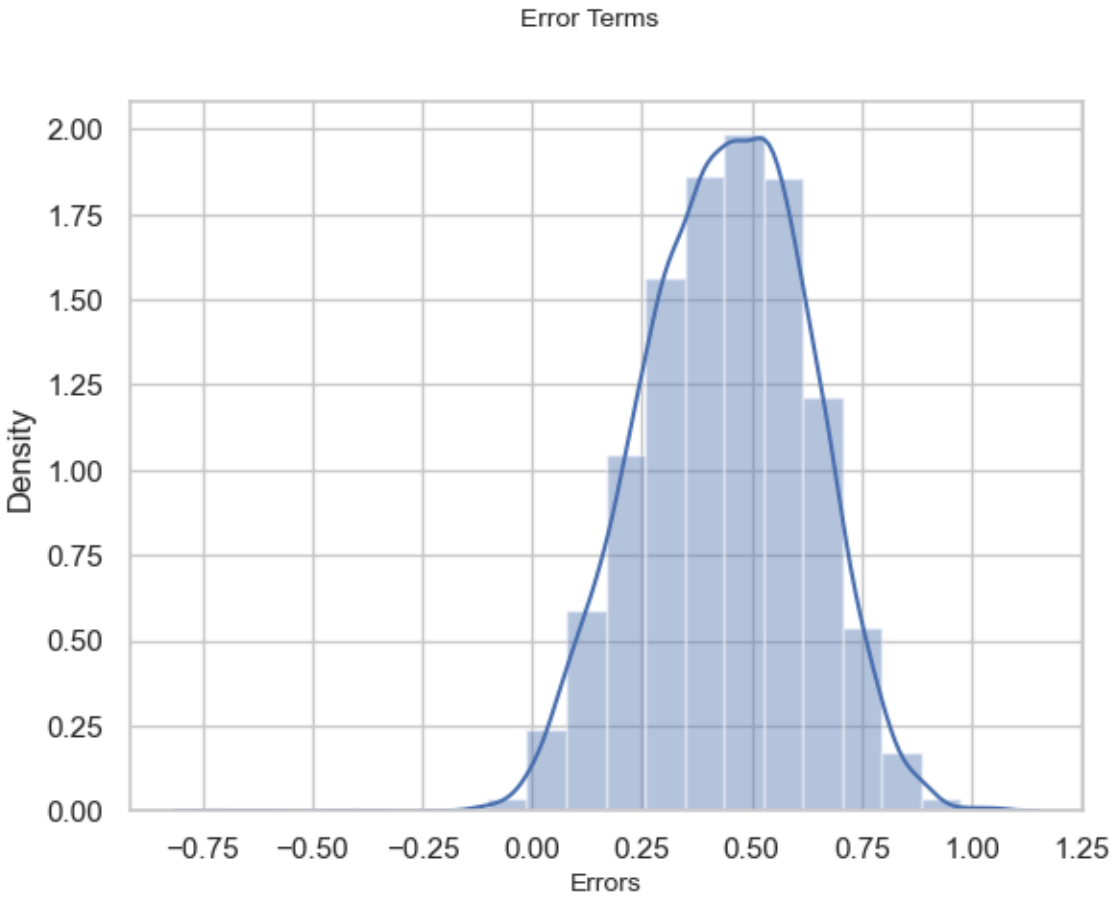
Prediction Accuracy: The model's predictions have a mean squared error of about 20.25%, indicating moderate accuracy in predicting LGD values

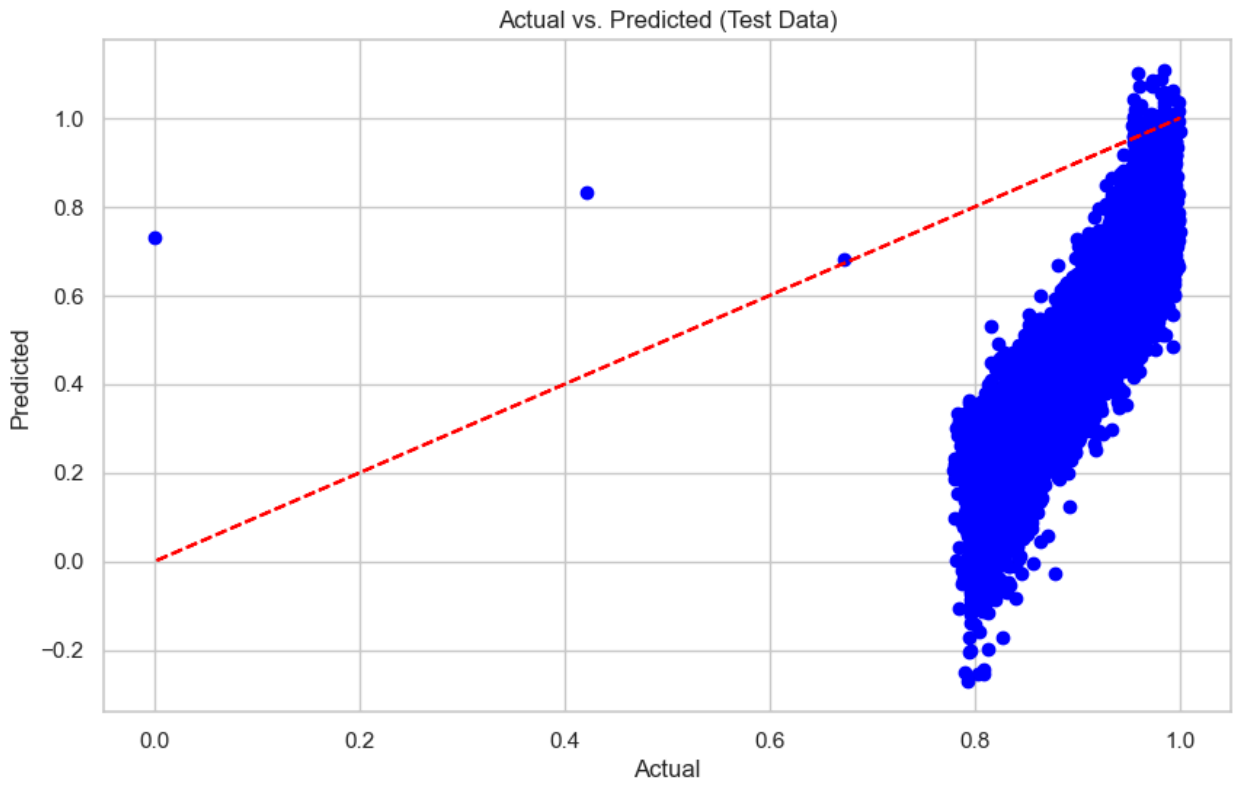
Error Magnitude: On average, the model's predictions deviate by approximately 1.51% from the actual LGD values, relative to the range of LGD values

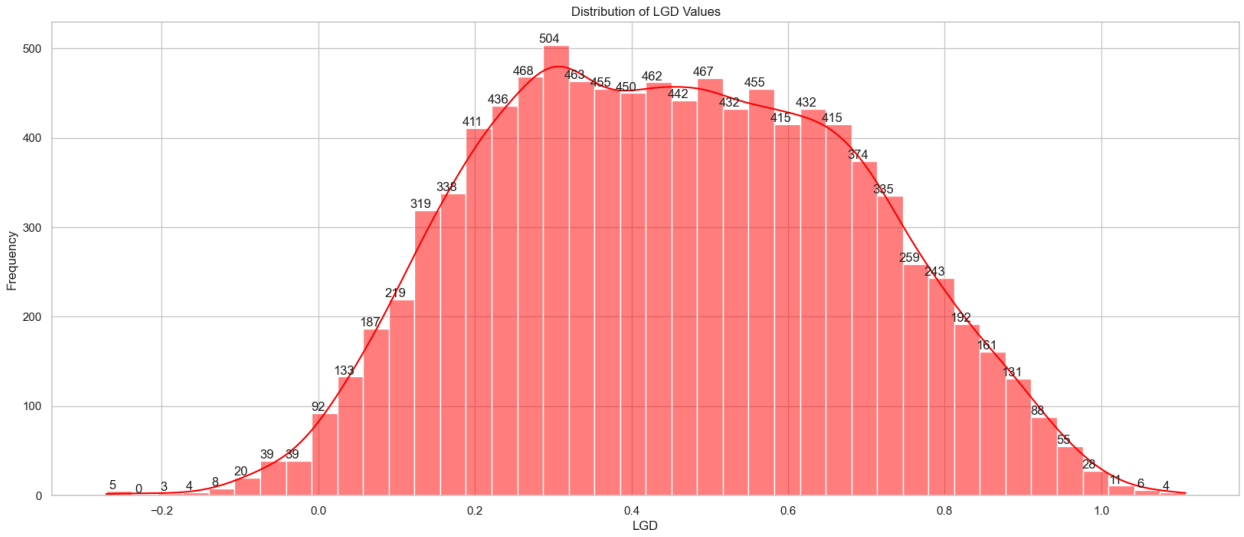
Decision Tree






Error Terms
Density
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
−0.75
−0.50
−0.25
0.00
0.25
0.50
0.75
1.00
1.25
Errors


Actual vs. Predicted (Test Data)
Predicted
Actual
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0.0
0.2
0.4
0.6
0.8
1.0


Distribution of LGD Values
Frequency
LGD
0
100
200
300
400
500
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
5
0
3
4
8
20
39
39
92
133
187
219
319
338
411
436
468
504
463
455
450
462
442
467
432
455
415
432
415
374
335
259
243
192
161
131
88
55
28
11
6
4

Loss Given Default (LGD) Equation

The Loss Given Default (LGD) equation is as follows:LGD=0.411+(−0.008)⋅collateral_value+0.101⋅cheque_bounces+0.211⋅number_of_loans+0.706⋅missed_repayments−0.360⋅vintage_in_months−0.004⋅interest+0.272⋅monthly_emi−0.163⋅repayment_amount−0.642⋅balance_amount−0.212⋅tenure_monthsLGD=0.411+(−0.008)⋅collateral_value+0.101⋅cheque_bounces+0.211⋅number_of_loans+0.706⋅missed_repayments−0.360⋅vintage_in_months−0.004⋅interest+0.272⋅monthly_emi−0.163⋅repayment_amount−0.642⋅balance_amount−0.212⋅tenure_months

Business Insights from the LGD Model

- **Collateral Value (Coefficient: -0.008):** A decrease in collateral value is associated with an increase in LGD. This implies that lower collateral value leads to higher losses in the event of default. Lending institutions should consider the quality and value of collateral when assessing the risk of potential defaults

- **Cheque Bounces (Coefficient: 0.101):** A higher incidence of cheque bounces correlates with increased LGD. Borrowers with a history of bounced cheques are more likely to default, resulting in higher losses for lenders. Monitoring and addressing cheque bounce occurrences can help mitigate default risk

- **Number of Loans (Coefficient: 0.211):** More loans held by a borrower contribute to higher LGD. This suggests that borrowers with multiple outstanding loans are at a greater risk of default, potentially increasing the lender's losses. Lending institutions should assess borrowers' existing debt burden and repayment capacity to manage risk effectively

- **Missed Repayments (Coefficient: 0.706):** The strong positive coefficient indicates that missed repayments significantly elevate LGD. Borrowers with a history of missed repayments pose a substantial default risk, leading to increased losses for lenders. Implementing proactive measures to address missed payments and monitor repayment behavior is crucial for risk mitigation

- **Vintage in Months (Coefficient: -0.360):** Longer vintage (age) of loans is associated with lower LGD. Loans with a longer repayment history may have demonstrated better creditworthiness and repayment behavior, resulting in reduced losses upon default. Lenders should consider vintage when assessing default risk and pricing loans accordingly

- **Interest (Coefficient: -0.004):** The interest rate has a minor negative impact on LGD. Higher interest rates are associated with slightly lower LGD, indicating that loans with higher interest may result in lower losses upon default. However, the effect is relatively small compared to other factors

- **Monthly EMI (Coefficient: 0.272):** Higher monthly EMIs contribute to increased LGD. Borrowers with higher EMIs may face financial strain, increasing the likelihood of default and subsequent losses for lenders. Lending institutions should assess borrowers' debt-to-income ratio to ensure affordability and reduce default risk

- **Repayment Amount (Coefficient: -0.163):** Higher repayment amounts are associated with lower LGD. Borrowers who repay larger amounts demonstrate better repayment capacity, resulting in reduced losses for lenders in the event of default. Lenders should encourage higher repayment amounts to mitigate default risk

- **Balance Amount (Coefficient: -0.642):** A higher balance amount is strongly correlated with lower LGD. Loans with lower outstanding balances at default lead to reduced losses for lenders. Lending institutions should focus on minimizing outstanding balances to mitigate potential losses upon default

- **Tenure in Months (Coefficient: -0.212):** Longer loan tenures are associated with lower LGD. Loans with longer repayment periods may have lower monthly payments, reducing the impact of default on lenders' losses. Lenders should consider offering longer tenures for risk mitigation purposes.

Conclusion

By understanding the influence of these factors on LGD, lending institutions can refine their risk assessment models, develop targeted risk management strategies, and make informed lending decisions to minimize losses and enhance profitability.