

Data Quality Plan

1 Handling strategies for data irregularities

Feature	Data Quality Issue	Handling Strategy
MSinceOldestTradeOpen	negative values (2%)	imputation
MSinceMostRecentInqexcl7days	negative values (25 %)	imputation
NetFractionRevolvingBurden	negative values (2%)	imputation
NetFractionInstallBurden	negative values (35%)	imputation
NumRevolvingTradesWBalance	negative values (2%)	imputation
NumInstallTradesWBalance	negative values (10%)	imputation
NumBank2NatlTradesWHighUtilization	negative values (5%)	imputation
MSinceMostRecentDelq	negative values (47%)	drop feature
NumTrades60Ever2DerogPubRec	many 0s	do nothing
NumTrades90Ever2DerogPubRec	many 0s	do nothing
NumTradesOpeninLast12M	many values in [0-1]	do nothing
NumInqLast6M	many 0s	do nothing
NumInqLast6Mexcl7days	many 0s	do nothing
MSinceMostRecentTradeOpen	outliers	do nothing
AverageMInFile	outliers	do nothing
NumSatisfactoryTrades	outliers	do nothing
NumTotalTrades	outliers	do nothing

2 Cleaned data results

2.1 Continuous features

	count	mean	std	min	25%	50%	75%	max
ExternalRiskEstimate	942.0	72.150743	10.039486	37.0	65.0	72.0	80.0	93.0
MSinceOldestTradeOpen	942.0	196.823779	101.320551	0.0	134.0	183.5	252.0	528.0
MSinceMostRecentTradeOpen	942.0	9.325902	11.255838	0.0	3.0	6.0	12.0	145.0
AverageMlnFile	942.0	78.936306	32.789643	5.0	58.0	75.5	98.0	256.0
NumSatisfactoryTrades	942.0	21.399151	11.398674	1.0	13.0	20.0	28.0	70.0
NumTrades60Ever2DerogPubRec	942.0	0.587049	1.202354	0.0	0.0	0.0	1.0	12.0
NumTrades90Ever2DerogPubRec	942.0	0.384289	0.905109	0.0	0.0	0.0	0.0	9.0
PercentTradesNeverDelq	942.0	91.898089	12.294483	0.0	88.0	97.0	100.0	100.0
NumTotalTrades	942.0	23.289809	13.362273	0.0	14.0	21.0	31.0	85.0
NumTradesOpeninLast12M	942.0	1.817410	1.731313	0.0	0.0	1.0	3.0	11.0
PercentInstallTrades	942.0	35.298301	17.510804	0.0	22.0	33.0	46.0	100.0
MSinceMostRecentInqexcl7days	942.0	1.829087	4.286435	0.0	0.0	0.0	1.0	24.0
NumInqLast6M	942.0	1.450106	1.871299	0.0	0.0	1.0	2.0	16.0
NumInqLast6Mexcl7days	942.0	1.397028	1.834697	0.0	0.0	1.0	2.0	16.0
NetFractionRevolvingBurden	942.0	33.378981	28.939596	0.0	7.0	27.0	55.0	112.0
NetFractionInstallBurden	942.0	43.814225	38.054613	0.0	0.0	49.5	79.0	142.0
NumRevolvingTradesWBalance	942.0	3.968153	3.000008	0.0	2.0	3.0	5.0	20.0
NumInstallTradesWBalance	942.0	2.317410	1.800370	0.0	1.0	2.0	3.0	14.0
NumBank2NatlTradesWHighUtilization	942.0	1.046709	1.574216	0.0	0.0	0.5	2.0	18.0
PercentTradesWBalance	942.0	66.690021	22.626372	0.0	50.0	67.0	83.0	100.0

2.2 Categorical features

	count	unique	top	freq
RiskPerformance	942	2	Bad	490
MaxDelq2PublicRecLast12M	942	8	7	388
MaxDelqEver	942	7	8	416