

Handling Strategies for Data Irregularities

Feature	Data Quality Issue	Handling Strategy
MSinceMostRecentInqexcl7days	-7 special value	Set -7s to NaN, and record -7 presence in new binary column
MSinceMostRecentDelq	-7 special value	Set -7s to NaN, and record -7 presence in new binary column
MSinceOldestTradeOpen	-8 special value	Imputation
MSinceMostRecentDelq	-8 special value	Imputation
MSinceMostRecentInqexcl7days	-8 special value	Imputation
NetFractionRevolvingBurden	-8 special value	Imputation
NetFractionInstallBurden	-8 special value	Imputation
NumRevolvingTradesWBalance	-8 special value	Imputation
NumInstallTradesWBalance	-8 special value	Imputation
NumBank2NatlTradesWHighUtilization	-8 special value	Imputation
PercentTradesWBalance	-8 special value	Imputation
All features except RiskPerformance target feature	-9 special values	Remove affected rows
NumSatisfactoryTrades	Higher values than NumTotalTrades	Remove affected rows
MaxDelq2PublicRecLast12M	Special values with same meaning	Replace value
NumTrades60Ever2DerogPubRec	Redundancy with NumTrades90Ever2DerogPubRec	Remove column
NumInqLast6M	Rendundancy with NumInqLast6Mexcl7days	Remove column
MSinceOldestTradeOpen	Outliers	Do nothing
MSinceMostRecentTradeOpen	Outliers	Do nothing
AverageMInFile	Outliers	Do nothing
NumSatisfactoryTrades	Outliers	Do nothing
NumTrades90Ever2DerogPubRec	Outliers	Do nothing
PercentTradesNeverDelq	Outliers	Do nothing
MSinceMostRecentDelq	Outliers	Do nothing
NumTotalTrades	Outliers	Do nothing
NumTradesOpeninLast12M	Outliers	Do nothing
PercentInstallTrades	Outliers	Do nothing
MSinceMostRecentInqexcl7days	Outliers	Do nothing
NumInqLast6Mexcl7days	Outliers	Do nothing
NetFractionRevolvingBurden	Outliers	Do nothing
NumRevolvingTradesWBalance	Outliers	Do nothing
NumInstallTradesWBalance	Outliers	Do nothing
NumBank2NatlTradesWHighUtilization	Outliers	Do nothing
PercentTradesWBalance	Outliers	Do nothing

Cleaned Data Results:

Continuous Features

	count	mean	std	min	25%	50%	75%	max
ExternalRiskEstimate	887.0	71.644870	9.781076	44.0	64.0	71.0	80.0	92.0
MSinceOldestTradeOpen	887.0	195.778815	89.158499	7.0	137.5	185.0	250.0	538.0
MSinceMostRecentTradeOpen	887.0	9.202931	12.199687	0.0	3.0	6.0	11.0	184.0
AverageMInFile	887.0	76.714769	32.486615	6.0	56.5	75.0	93.0	219.0
NumSatisfactoryTrades	887.0	20.835400	11.204843	1.0	13.0	19.0	27.0	59.0
NumTrades90Ever2DerogPubRec	887.0	0.400225	1.073443	0.0	0.0	0.0	0.0	16.0
PercentTradesNeverDelq	887.0	92.298760	11.973062	0.0	89.0	97.0	100.0	100.0
MSinceMostRecentDelq	467.0	21.501071	20.924386	0.0	5.0	13.0	33.0	81.0
NumTotalTrades	887.0	23.554679	12.445416	1.0	14.5	22.0	31.0	77.0
NumTradesOpeninLast12M	887.0	1.928974	1.848227	0.0	1.0	2.0	3.0	14.0
PercentInstallTrades	887.0	34.572717	17.627608	0.0	22.0	33.0	44.0	100.0
MSinceMostRecentInqexcl7days	734.0	2.282016	4.703287	0.0	0.0	0.0	2.0	23.0
NumInqLast6Mexcl7days	887.0	1.393461	2.168855	0.0	0.0	1.0	2.0	29.0
NetFractionRevolvingBurden	887.0	35.186020	28.160663	0.0	9.0	30.0	56.0	131.0
NetFractionInstallBurden	887.0	66.449831	21.108756	1.0	58.0	58.0	84.0	165.0
NumRevolvingTradesWBalance	887.0	4.122886	2.874659	0.0	2.0	4.0	5.0	18.0
NumInstallTradesWBalance	887.0	2.410372	1.518369	1.0	1.0	2.0	3.0	14.0
NumBank2NatlTradesWHighUtilization	887.0	1.076663	1.430397	0.0	0.0	1.0	2.0	9.0
PercentTradesWBalance	887.0	66.795941	21.603707	0.0	50.0	67.0	83.0	100.0

Categorical Features

	count	unique	top	freq
RiskPerformance	887	2	Bad	473
MaxDelq2PublicRecLast12M	887	7	7	382
MaxDelqEver	887	7	8	409
ConditionNotMet	887	2	True	502