

## Handling Strategies for Data Irregularities

Feature	Data Quality Issue	Handling Strategy
NetFractionInstallBurden	>30% N/A values	Feature dropped
ExternalRiskEstimate	N/A values (<30%)	Imputation
MSinceOldestTradeOpen	N/A values (<30%)	Imputation
MSinceMostRecentDelq	N/A values (<30%)	Imputation
MSinceMostRecentInqexcl7days	N/A values (<30%)	Imputation
NetFractionRevolvingBurden	N/A values (<30%)	Imputation
NumRevolvingTradesWBalance	N/A values (<30%)	Imputation
NumInstallTradesWBalance	N/A values (<30%)	Imputation
NumBank2NatlTradesWHighUtilization	N/A values (<30%)	Imputation
PercentTradesWBalance	N/A values (<30%)	Imputation
MSinceMostRecentDelq	-7 values	Imputation then binning
MSinceMostRecentInqexcl7days	-7 values	Imputation then binning
ExternalRiskEstimate	Outliers	Do nothing
MSinceOldestTradeOpen	Outliers	Do nothing
MSinceMostRecentTradeOpen	Outliers	Do nothing
AverageMInFile	Outliers	Do nothing
NumSatisfactoryTrades	Outliers	Do nothing
NumTrades60Ever2DerogPubRec	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
NumInqLast6M	Outliers	Do nothing
NumRevolvingTradesWBalance	Outliers	Do nothing
NumInstallTradesWBalance	Outliers	Do nothing
NumBank2NatlTradesWHighUtilization	Outliers	Do nothing
PercentTradesWBalance	Outliers	Do nothing
NumInqLast6M	Inconsistency	Do nothing
MSinceMostRecentInqexcl7days	Inconsistency	Do nothing
PercentTradesNeverDelq	Inconsistency	Do nothing
MSinceMostRecentDelq	Inconsistency	Do nothing
MaxDelq2PublicRecLast12M	Ambiguous value meaning	Meaning clarified
MaxDelqEver	Ambiguous value meaning	Meaning clarified

## Updated Descriptive Statistics for Continuous Features

	count	mean	std	min	25%	50%	75%	max
ExternalRiskEstimate	941.0	71.989373	9.797627	40.0	65.0	72.0	80.0	92.0
MSinceOldestTradeOpen	923.0	200.458288	97.507601	7.0	132.0	185.0	254.5	590.0
MSinceMostRecentTradeOpen	943.0	9.667020	13.337026	0.0	3.0	6.0	12.0	184.0
AverageMInFile	943.0	79.463415	36.269141	7.0	57.0	75.0	96.0	258.0
NumSatisfactoryTrades	943.0	20.923648	11.026435	1.0	13.0	20.0	27.0	74.0
NumTrades60Ever2DerogPubRec	943.0	0.617179	1.306850	0.0	0.0	0.0	1.0	12.0
NumTrades90Ever2DerogPubRec	943.0	0.379639	1.009556	0.0	0.0	0.0	0.0	12.0
PercentTradesNeverDelq	943.0	92.217391	11.887317	20.0	88.5	97.0	100.0	100.0
NumTotalTrades	943.0	22.686108	12.637905	0.0	14.0	21.0	29.0	78.0
NumTradesOpeninLast12M	943.0	1.889714	1.891014	0.0	0.0	1.0	3.0	17.0
PercentInstallTrades	943.0	34.397667	17.641026	0.0	21.0	33.0	45.5	100.0
NumInqLast6M	943.0	1.650053	3.009630	0.0	0.0	1.0	2.0	66.0
NetFractionRevolvingBurden	928.0	35.101293	28.927549	0.0	8.0	30.0	58.0	104.0
NumRevolvingTradesWBalance	932.0	3.969957	2.844929	0.0	2.0	3.0	5.0	29.0
NumInstallTradesWBalance	867.0	2.498270	1.794509	1.0	1.0	2.0	3.0	23.0
NumBank2NatlTradesWHighUtilization	898.0	1.031180	1.366557	0.0	0.0	1.0	2.0	13.0
PercentTradesWBalance	941.0	66.151966	22.143806	0.0	50.0	67.0	83.0	100.0

## Updated Descriptive Statistics for Categorical Features

	count	unique	top	freq
RiskPerformance	943	2	Bad	486
MSinceMostRecentDelq	913	10	NoDelq	446
MaxDelq2PublicRecLast12M	943	8	7	408
MaxDelqEver	943	7	8	436
MSinceMostRecentInqexcl7days	474	6	0-5	180