

Handling Strategies for Data Irregularities.

Feature	Data Quality Issue	Handling Strategy
RiskPerformance	Outliers	Imputation
ExternalRiskEstimate	Outliers	Do nothing
MSinceOldestTradeOpen	Outliers	Do nothing
MSinceMostRecentTradeOpen	Outliers	Do nothing
AverageMInFile	Outliers	Do nothing
NumSatisfactoryTrades	Outliers	Do nothing
NumTrades60Ever2DerogPubRec	Repeat information	Remove column
NumTrades90Ever2DerogPubRec	Outliers	Do nothing
PercentTradesNeverDelq	Outliers	Do nothing
MSinceMostRecentDelq	Missing Values (~50%)	Remove column
MaxDelq2PublicRecLast12M	Outliers	Do nothing
MaxDelqEver	Outliers	Do nothing
NumTotalTrades	Outliers	Do nothing
NumTradesOpeninLast12M	Outliers	Do nothing
PercentInstallTrades	Outliers	Do nothing
MSinceMostRecentInqexcl7days	Repeat information	Remove column
NumInqLast6M	Outliers	Do nothing
NumInqLast6Mexcl7days	Outliers	Do nothing
NetFractionRevolvingBurden	Outliers	Do nothing
NetFractionInstallBurden	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	953.0	72.801679	9.857361	40.0	65.0	73.0	81.0	94.0
MSinceOldestTradeOpen	953.0	196.194124	96.798799	4.0	129.0	179.0	248.0	589.0
MSinceMostRecentTradeOpen	953.0	9.951731	13.678780	0.0	3.0	6.0	12.0	152.0
AverageMInFile	953.0	78.655824	34.232385	4.0	56.0	76.0	97.0	245.0
NumSatisfactoryTrades	953.0	21.218258	11.955037	1.0	13.0	20.0	28.0	78.0
NumTrades90Ever2DerogPubRec	953.0	0.356768	0.917099	0.0	0.0	0.0	0.0	10.0
PercentTradesNeverDelq	953.0	92.961175	11.562834	0.0	90.0	100.0	100.0	100.0
NumTotalTrades	953.0	22.547744	13.612744	0.0	13.0	21.0	30.0	100.0
NumTradesOpeninLast12M	953.0	1.982162	2.024715	0.0	0.0	2.0	3.0	19.0
PercentInstallTrades	953.0	34.273872	18.170918	0.0	21.0	33.0	45.0	100.0
NumInqLast6M	953.0	1.493179	2.284448	0.0	0.0	1.0	2.0	29.0
NetFractionRevolvingBurden	953.0	33.578174	28.330070	0.0	8.0	27.0	55.0	115.0
NetFractionInstallBurden	953.0	59.537251	22.541906	0.0	46.0	46.0	79.0	132.0
NumRevolvingTradesWBalance	953.0	4.089192	3.078698	0.0	2.0	3.0	5.0	26.0
NumInstallTradesWBalance	953.0	2.432319	1.690894	1.0	1.0	2.0	3.0	19.0
NumBank2NatlTradesWHighUtilization	953.0	0.993704	1.531175	0.0	0.0	0.0	1.0	12.0
PercentTradesWBalance	953.0	66.311647	21.983150	0.0	50.0	67.0	83.0	100.0

Categorical Features

	count	unique	top	freq
RiskPerformance	953	2	Bad	493
MaxDelq2PublicRecLast12M	953	8	7	440
MaxDelqEver	953	7	8	466