

## Handling Strategies for Data Irregularities.

Feature	Data Quality Issue	Handling Strategy
age	Missing Values (30%)	Imputation
handsetAge	Negative Value (1 row)	Make value positive
currentHandsetPrice	Missing Values (50%)	Remove column
occupation	Missing Values (>50%)	Remove column
regionType	Missing Values (~50%)	Remove column
marriageStatus	Missing Values (40%)	Remove column
customer	No useful information	Remove column
creditCard	Cardinality	Merge synonymous cardinalities
numHandsets	Outliers	Do nothing
handsetAge	Outliers	Do nothing
avgBill	Outliers	Do nothing
avgMins	Outliers	Do nothing
avgrecurringCharge	Outliers	Do nothing
avgOverBundleMins	Outliers	Do nothing
avgRoamCalls	Outliers	Do nothing
callMinutesChangePct	Outliers	Do nothing
billAmountChangePct	Outliers	Do nothing
avgReceivedMins	Outliers	Do nothing
avgOutCalls	Outliers	Do nothing
avgInCalls	Outliers	Do nothing
peakOffPeakRatio	Outliers	Do nothing
avgDroppedCalls	Outliers	Do nothing
peakOffPeakRatioChangePct	Outliers	Do nothing
lifeTime	Outliers	Do nothing
lastMonthCustomerCareCalls	Outliers	Do nothing
numRetentionCalls	Outliers	Do nothing
numRetentionOffersAccepted	Outliers	Do nothing
newFrequentNumbers	Outliers	Do nothing

## Cleaned data results:

### Continuous Features

	count	mean	std	min	25%	50%	75%	max
age	1000.0	40.664000	11.580207	18.000000	34.000000	34.000000	48.000000	86.000000
numHandsets	1000.0	1.769000	1.365102	1.000000	1.000000	1.000000	2.000000	13.000000
handsetAge	1000.0	401.155000	259.400470	2.000000	218.000000	344.500000	542.000000	1812.000000
avgBill	1000.0	57.347800	44.519769	0.000000	33.247500	49.155000	69.175000	469.670000
avgMins	1000.0	501.303570	530.272984	0.000000	147.500000	347.000000	663.875000	4598.750000
avgrecurringCharge	1000.0	46.328200	24.246087	0.000000	30.000000	44.990000	59.990000	299.990000
avgOverBundleMins	1000.0	38.888760	94.715038	0.000000	0.000000	4.000000	42.250000	1389.000000
avgRoamCalls	1000.0	0.916220	4.940806	0.000000	0.000000	0.000000	0.275000	131.040000
callMinutesChangePct	1000.0	-0.537772	5.197712	-44.655000	-1.856250	-0.175000	0.990000	28.050000
billAmountChangePct	1000.0	-0.010483	0.927263	-5.763000	-0.136300	-0.005000	0.036900	17.911400
avgReceivedMins	1000.0	109.156550	163.394282	0.000000	5.787500	49.705000	150.755000	1549.930000
avgOutCalls	1000.0	25.387990	36.119986	0.000000	2.330000	12.835000	36.000000	332.000000
avgInCalls	1000.0	8.439290	17.036525	0.000000	0.000000	2.000000	9.415000	233.000000
peakOffPeakRatio	1000.0	2.137873	3.465256	0.000000	0.722636	1.380353	2.446368	74.774775
peakOffPeakRatioChangePct	1000.0	0.273071	9.620523	-28.154016	-6.099596	0.056048	6.636753	37.779743
avgDroppedCalls	1000.0	9.699970	14.972256	0.000000	1.330000	5.000000	11.415000	159.330000
lifeTime	1000.0	18.747000	9.499433	6.000000	11.000000	17.000000	24.000000	60.000000
lastMonthCustomerCareCalls	1000.0	1.641900	4.167139	0.000000	0.000000	0.000000	1.330000	42.000000
numRetentionCalls	1000.0	0.051000	0.233353	0.000000	0.000000	0.000000	0.000000	2.000000
numRetentionOffersAccepted	1000.0	0.024000	0.159529	0.000000	0.000000	0.000000	0.000000	2.000000
newFrequentNumbers	1000.0	0.204000	0.637802	0.000000	0.000000	0.000000	0.000000	3.000000

Categorical Features

	count	unique	top	freq
children	1000	2	False	765
income	1000	10	0	268
smartPhone	1000	2	True	891
creditRating	1000	7	B	389
homeOwner	1000	2	False	652
creditCard	1000	2	true	669
churn	1000	2	True	532