PhenoQC Quality Control Report

Source file: data.csv

Summary

Imputation Strategy	Knn
Schema Validation Score	100.00%
Missing Data Score	92.50%
Mapping Success Score	0.00%
Overall Quality Score	64.17%

Imputation Settings

Global Strategy	knn
Global Params	{'n_neighbors': 5}
Tuning Enabled	True
Best Params	{'n_neighbors': 5}
Tuning Score	55.1553 (MAE)

Data Quality Scores:

Schema Validation Score: 100.00%
Missing Data Score: 92.50%
Mapping Success Score: 0.00%
Overall Quality Score: 64.17%

Schema Validation Results

Format Validation: True
Duplicate Records: No issues found.
Conflicting Records: No issues found.
Integrity Issues: No issues found.
Referential Integrity Issues: No issues found.
Anomalies Detected: 2 issues found.

Invalid Mask: 800 issues found.

Additional Quality Dimensions

Redundancy Issues: 1 issues found.

column_1	column_2	metric	value
feat1	feat2	correlation	0.9997319157158756

Imputation Stability & Bias

Stability (repeatability)

Imputation Stability (top variables): 5 issues found.

column	repeats	metric	mean_error	sd_error	cv_error
outcome	10	MAE	0.1729	0.0288	0.1667
feat1	10	MAE	0.3839	0.06	0.1563
feat2	10	MAE	0.7928	0.1018	0.1284
feat3	10	MAE	2.6613	0.2126	0.0799
SampleID	10	MAE	216.9366	12.6886	0.0585

Warning rules: SMD≥0.1 | Var-ratio∉ [0.5,2.0] | KS p<0.05

Variables evaluated: 3
: 3 issues found.

Variable	n_obs	n_imp	SMD	Var-ratio	KS p	Triggers	Warn
feat1	640	160	-0.017	0.861	0.0	KS p<0.05	True
feat2	640	160	0.026	0.886	0.008	KS p<0.05	True
feat3	640	160	-0.021	0.842	0.0	KS p<0.05	True

Variable	Status	Triggers
feat2	WARN	KS p<0.05
feat1	WARN	KS p<0.05

feat3	WARN	KS p<0.05

Missing Data Summary

Column	Missing Count
feat1	160
feat2	160
feat3	160

Records Flagged for Missing Data: 384

Ontology Mapping Summary

Ontology	Total Terms	Mapped	Success Rate
HPO	0	0	0.00%
DO	0	0	0.00%
MPO	0	0	0.00%

Visualizations

Missingness Heatmap

Percentage of Missing Data by Column

Distribution of feat1

Distribution of feat2
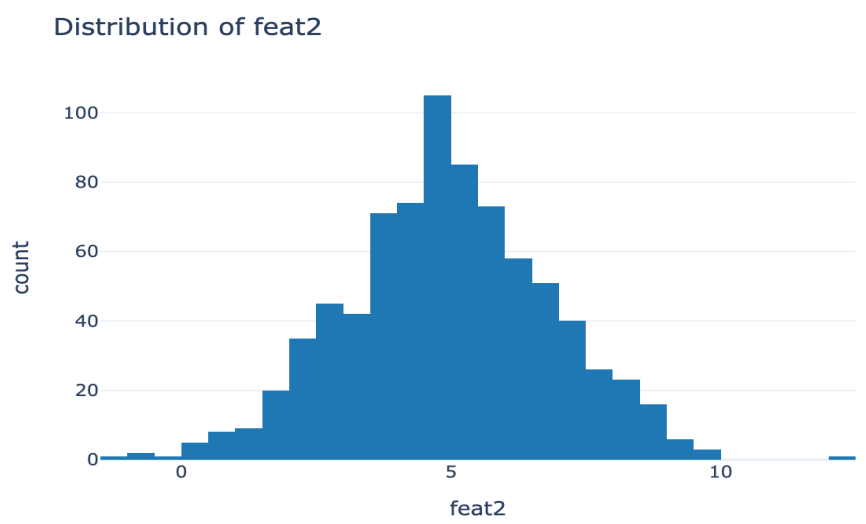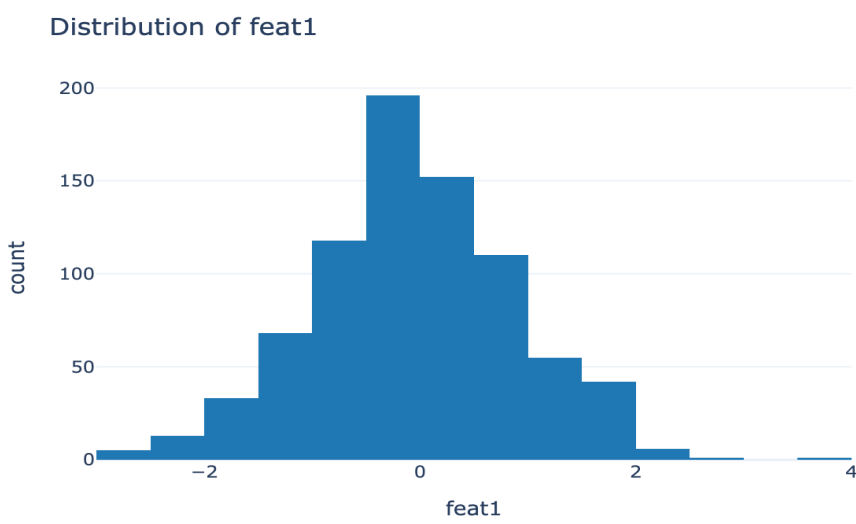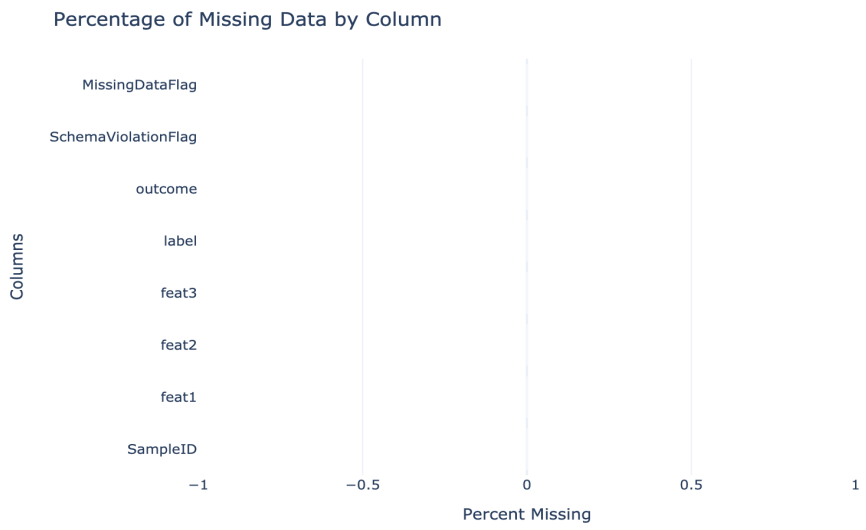
Distribution of feat3
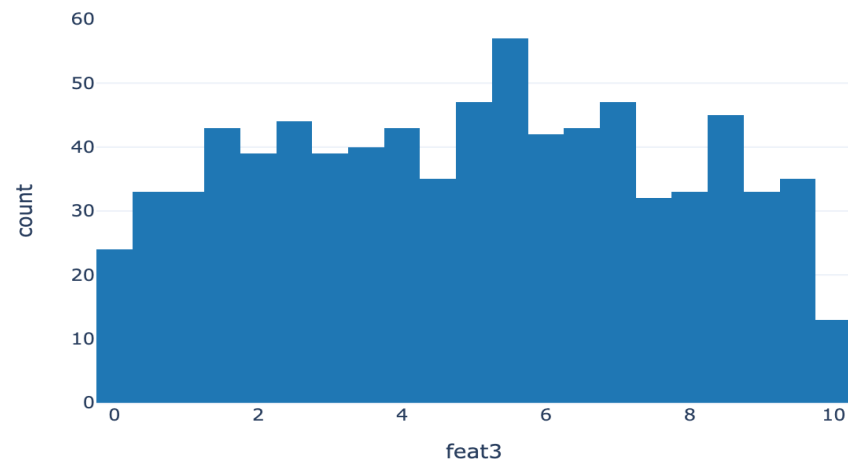
Distribution of outcome