PhenoQC Quality Control Report

Source file: data.csv

Summary

Imputation Strategy	Mean
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': 3}
Tuning Enabled	True
Best Params	{'n_neighbors': 5}
Tuning Score	52.2368 (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	0.9996491791289479

Imputation Stability & Bias

Stability (repeatability)

Imputation Stability (top variables): 5 issues found.

column	repeats	metric	mean_error	sd_error	cv_error
feat1	5	RMSE	0.5945	0.07	0.1178
outcome	5	RMSE	0.2936	0.0295	0.1005
feat2	5	RMSE	1.1523	0.1071	0.093
SampleID	5	RMSE	274.6487	23.827	0.0868
feat3	5	RMSE	3.3351	0.2525	0.0757

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

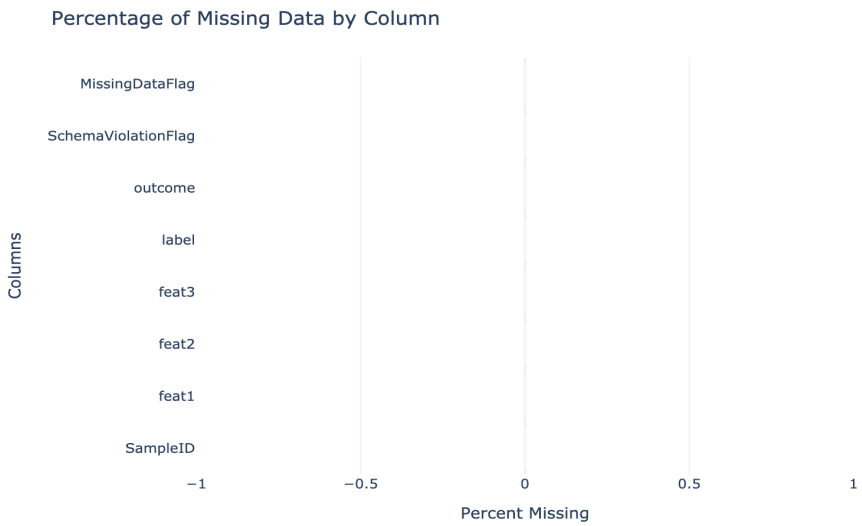
Column	Missing Count
feat1	160
feat2	160
feat3	160

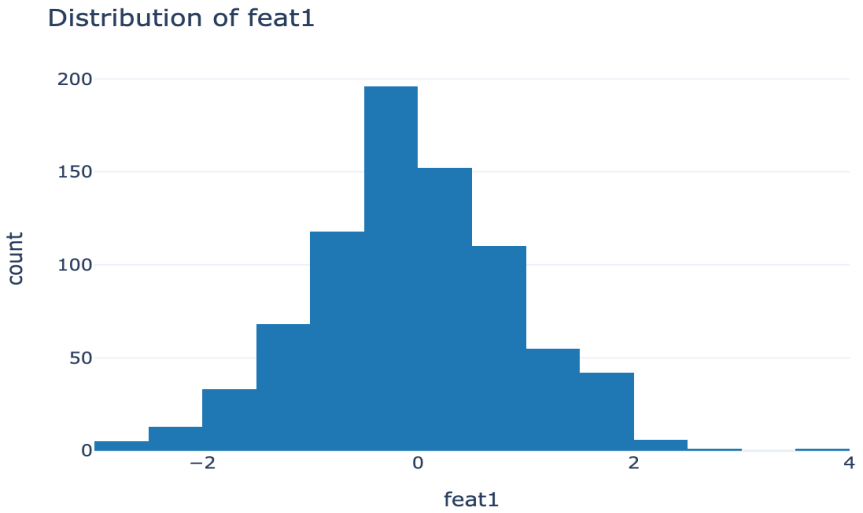
Records Flagged for Missing Data: 384

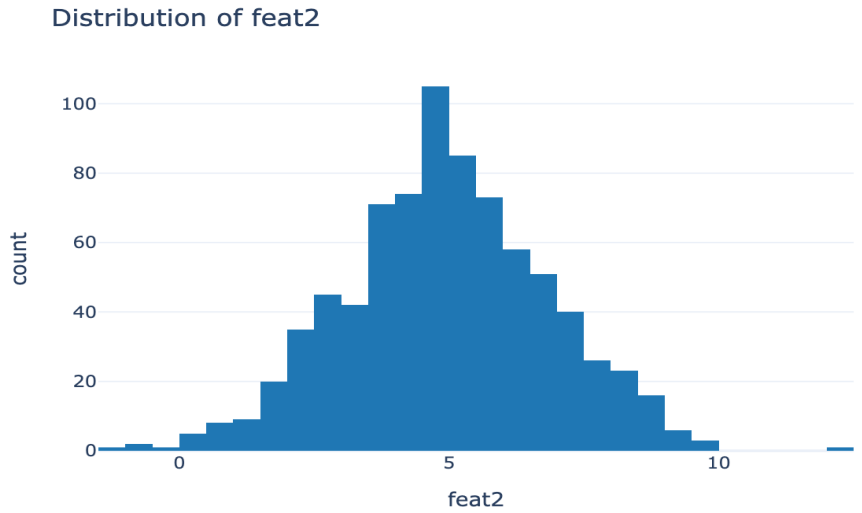
Ontology Mapping Summary

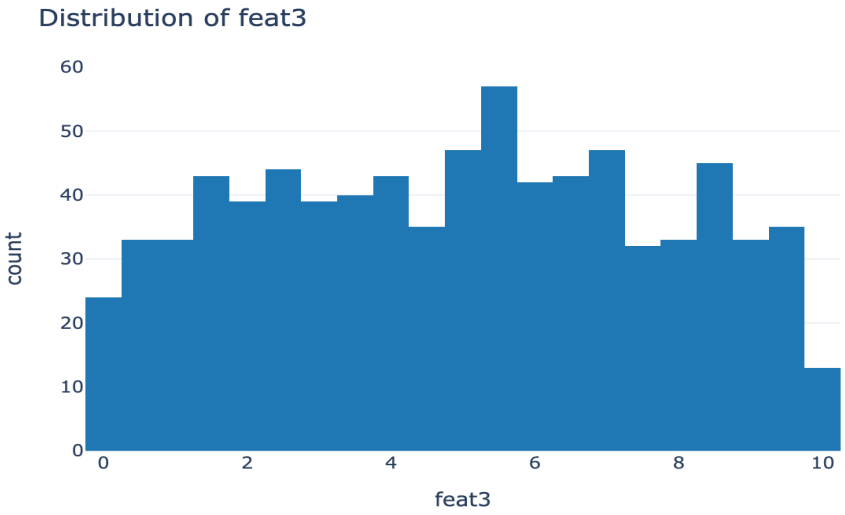
Ontology	Total Terms	Mapped	Success Rate
HPO	0	0	0.00%
DO	0	0	0.00%
MPO	0	0	0.00%

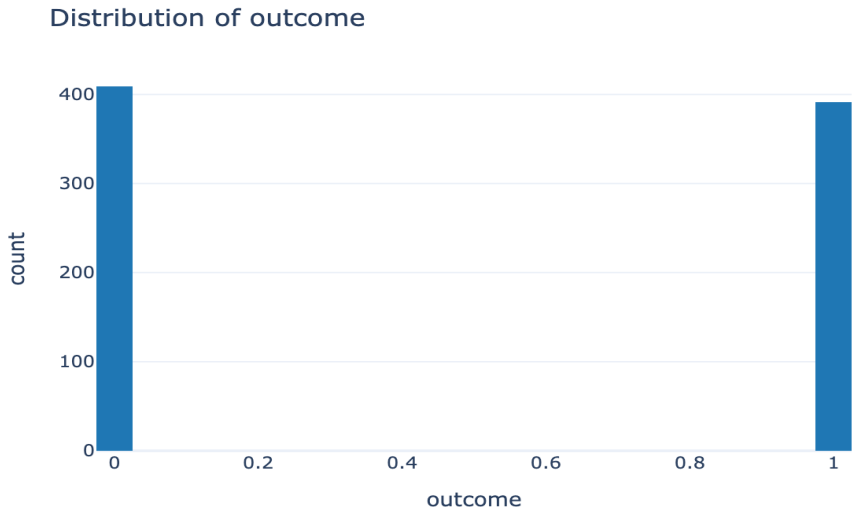
Visualizations








Distribution of feat3
count
feat3
0
10
20
30
40
50
60
0
2
4
6
8
10


Distribution of outcome
count
outcome
0
100
200
300
400
0
0.2
0.4
0.6
0.8
1