PhenoQC Quality Control Report

Source file: quality_metrics_input.csv

Summary

Imputation Strategy	Mean
Schema Validation Score	0.00%
Missing Data Score	93.75%
Mapping Success Score	100.00%
Overall Quality Score	64.58%

Imputation Settings

Global Strategy	knn
Global Params	{'n_neighbors': 5, 'weights': 'uniform'}
Tuning Enabled	True
Best Params	{'n_neighbors': 3}
Tuning Score	0.0000 (MAE)

Data Quality Scores:

Schema Validation Score: 0.00%
Missing Data Score: 93.75%
Mapping Success Score: 100.00%
Overall Quality Score: 64.58%

Schema Validation Results

Format Validation: False
Duplicate Records: No issues found.
Conflicting Records: No issues found.
Integrity Issues: 4 issues found.
Referential Integrity Issues: No issues found.
Anomalies Detected: No issues found.

Invalid Mask: 4 issues found.

Additional Quality Dimensions

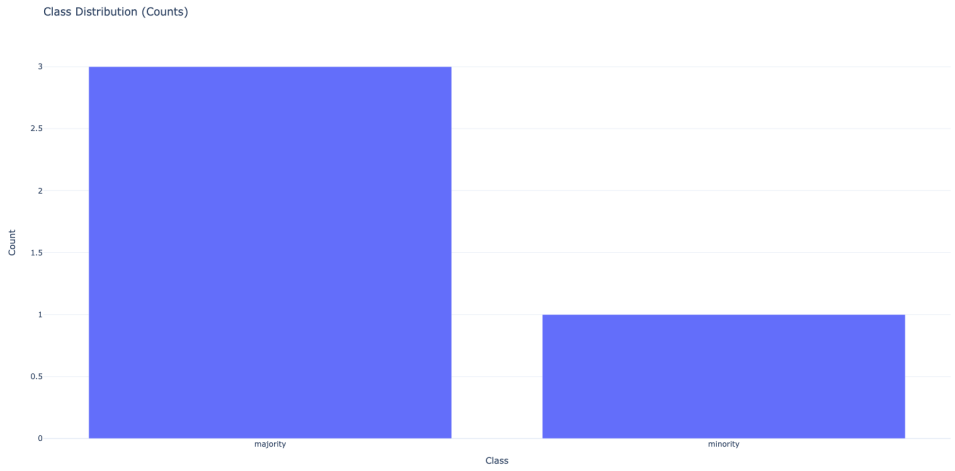
Accuracy Issues: No issues found.
Redundancy Issues: 1 issues found.

column_1	column_2	metric	value
Height_cm	Weight_kg	correlation	1.0

Traceability Issues: No issues found.
Timeliness Issues: No issues found.

Class Distribution

Class	Count	Proportion
majority	3	75.00%
minority	1	25.00%



Missing Data Summary

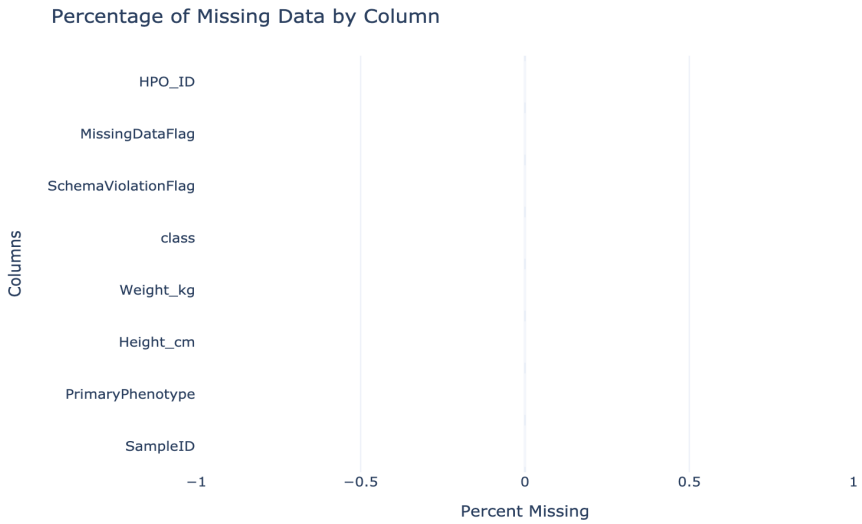
Column	Missing Count
Height_cm	1
Weight_kg	1

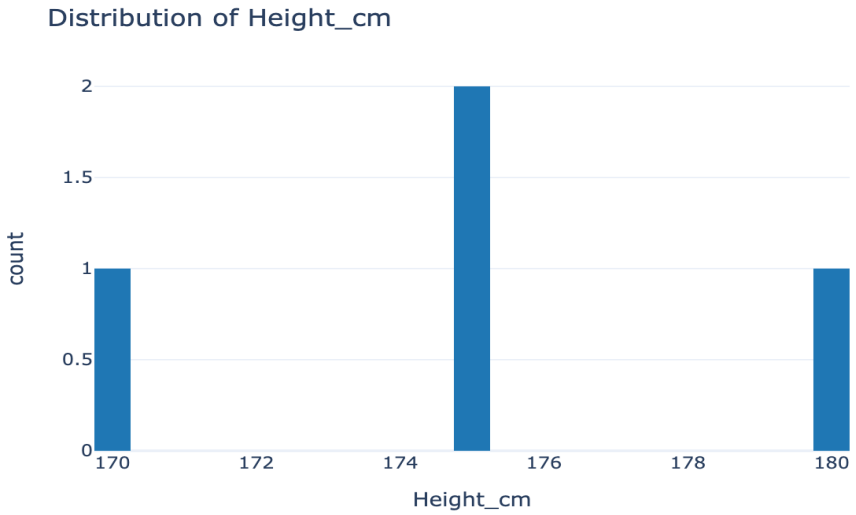
Records Flagged for Missing Data: 2

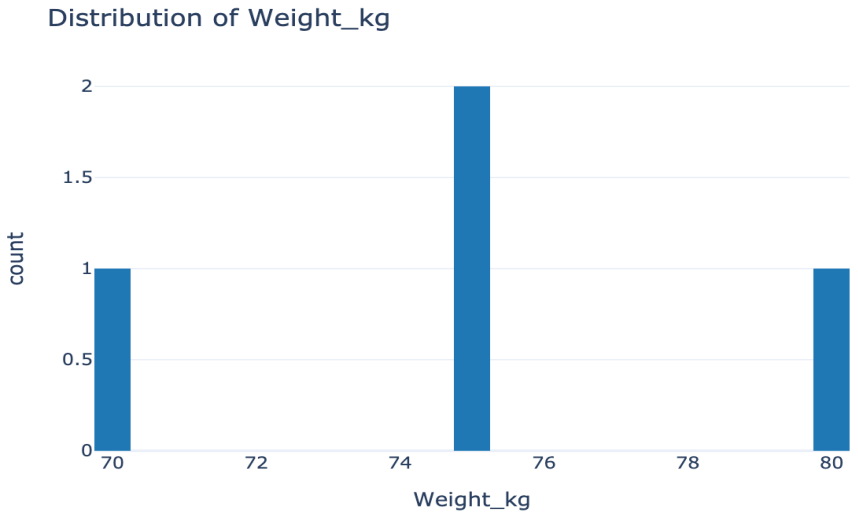
Ontology Mapping Summary

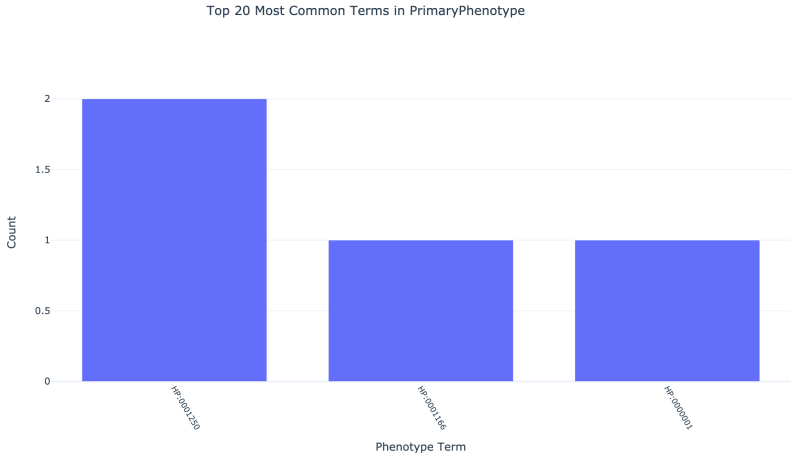
Ontology	Total Terms	Mapped	Success Rate
HPO	3	3	100.00%

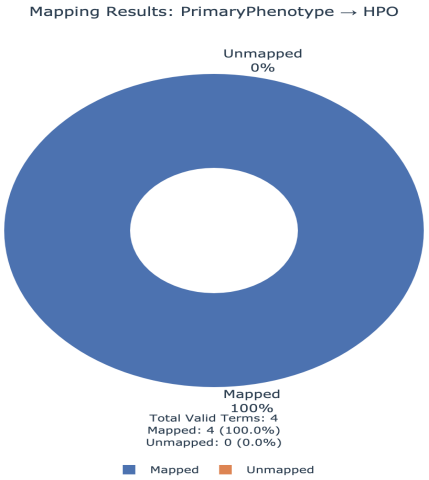
Visualizations








Top 20 Most Common Terms in PrimaryPhenotype
Count
2
1.5
1
0.5
0
HP:0001250
HP:0001166
HP:0000001
Phenotype Term


Mapping Results: PrimaryPhenotype → HPO
Unmapped
0%
Mapped
100%
Total Valid Terms: 4
Mapped: 4 (100.0%)
Unmapped: 0 (0.0%)
Mapped
Unmapped