

Data Quality Report

timestamp: 2026-01-22T21:17:01.350327

dataset_info:

rows: 10050

columns: 5

memory_usage_mb: 1.5

schema:

has_required_columns: True

missing_columns: []

extra_columns: ['extra_column']

total_columns: 5

missing_values:

total_missing: 100

columns_with_missing: {'rating': 100}

missing_percentages: {'rating': 1.0}

rows_with_any_missing: 100

duplicates:

total_duplicates: 278

duplicate_percentage: 2.77

unique_rows: 10000

rating_range:

column: rating

expected_range: [1, 5]

actual_min: 0.0

actual_max: 6.0

out_of_range_count: 20

out_of_range_percentage: 0.2

data_types:

type_mismatches: {'rating': {'expected': 'int', 'actual': 'float64'}}

all_types_correct: False

statistics:

unique_users: 1000

unique_items: 500

avg_rating: 2.9785929648241205

rating_std: 1.418557998392824

quality_score: 81.03