Data Quality Report

timestamp: 2026-01-22T22:12:31.572942

dataset_info:

  rows: 10100

  columns: 6

  memory_usage_mb: 1.97

schema:

  has_required_columns: True

  missing_columns: []

  extra_columns: ['event_type', 'extra_column']

  total_columns: 6

missing_values:

  total_missing: 100

  columns_with_missing: {'rating': 100}

  missing_percentages: {'rating': 0.99}

  rows_with_any_missing: 10100

duplicates:

  total_duplicates: 295

  duplicate_percentage: 2.92

  unique_rows: 10049

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: 2.9777

  rating_std: 1.4193675247859667

quality_score: 80.89