Data Quality Report – DM4ML Assignment

Generated on: 2026-01-27 06:50:01

1. Interactions Dataset

Source File: data_lake/raw/interactions/csv\2026\01\27\interactions_20260127_052410.csv
Total Records: 20010

Null Values:
user_id: 0
item_id: 0
event_type: 0
rating: 15094
timestamp: 0
device: 0
session_id: 0

Validation Results:
invalid_ratings: 349
invalid_events: 0
missing_user_or_item: 0
duplicate_rows: 10

2. Products Dataset

Source File: data_lake/raw/products/api\2026\01\27\products_20260127_055214.json
Total Records: 200

Validation Results:
missing_item_id: 0
missing_name: 0
missing_category: 0
invalid_price: 0
invalid_rating_avg: 0
invalid_popularity_score: 0

3. Conclusion

The data quality checks were successfully executed on both the interaction and product datasets. Each PDF report is automatically versioned using a timestamp, ensuring full auditability and traceability of data quality across pipeline runs.