



Data Quality Evidence Report

1,545

RECORDS ANALYSED

1

HIGH PRIORITY ISSUES

10

TOTAL ISSUES

EXECUTIVE SUMMARY

The 1,545 records of 'data_with_issues' indicate critical data quality issues. The two worst issues were identifier integrity failures and systematic data gaps across dataset. Overall, the analysis showed 1 high priority, 9 medium priority and 0 low priority issues.

Data quality issues on uniqueness and completeness represent a present risk to effective data insights and an opportunity to improve data standards throughout the organisation. Given these issues, suggestions are to implement mandatory field validation and duplicate detection systems with systematic cleanup of existing data.



Data Quality Evidence Report

DATA QUALITY ASSESSMENT

HIGH PRIORITY

1

- Primary key field "id" contains 118 duplicate values, revealing opportunities to enhance identifier assignment processes.

MEDIUM PRIORITY

9

- Dataset examination reveals 17.3% completion with gaps that may warrant attention to maintain data utility.

- Field completeness reveals gaps in columns such as "OrderDate" (68.8% complete), "Country" (69.6% complete), "Phone" (74.0% complete) that may benefit from attention to maintain utility.

- Email addresses in "email" show 72 format variations that may limit communication effectiveness.

- Phone numbers in "phone" show 86 format variations that may limit communication effectiveness.

- Age data in "age" shows 44 negative values that may warrant process review and correction.

- Numeric column "purchaseamount" shows 61 non-numeric entries that may limit analytics and business use.

- Domain verification identifies 179 suspicious test-like addresses in "email", highlighting potential for domain verification improvements.

- Record integrity finds 154 incomplete rows (10.0%) representing normal operational variance worth monitoring.



Data Quality Evidence Report

RECOMMENDED NEXT STEPS

Uniqueness Dimension:

- Deploy automated duplicate detection and prevention systems
- Establish unique identifier validation rules at point of entry
- Implement regular deduplication processes for existing data

Completeness Dimension:

- Implement mandatory field validation for critical data elements
- Review data entry processes to identify systematic collection gaps
- Establish completeness monitoring with automated alerts for key fields

Organisational Strategy:

Based on the worst-performing data quality dimensions, a targeted improvement program is justified. Focus resources on the identified priority areas before expanding to other dimensions.