

Tasks 1. Data Quality

a. Example of Poor Quality Structured Data

An example of poor-quality structured data is a **sales database** containing entries with invalid dates (e.g., "2023-02-30"), missing customer IDs, and duplicate transaction records (Kimball & Caserta, 2004).

b. Recognizing Poor Quality in Structured Data

This data fails to meet the criteria of good quality data:

1. **Validity:** The date "2023-02-30" is invalid, as February 2023 only had 28 days (ISO 8000, 2023).
2. **Completeness:** Missing customer IDs create gaps in critical fields, hindering analysis.
3. **Consistency:** Duplicate transactions distort revenue calculations and violate uniqueness constraints (Wang & Strong, 1996).

c. Example of Poor Quality Unstructured Data

A collection of **customer feedback emails** with unstructured text containing slang, spelling errors, and incomplete sentences (Batini et al., 2009).

d. Challenges in Assessing Unstructured Data

This unstructured data lacks key quality attributes:

1. **Standardization:** Free-form text lacks a consistent format, making automated analysis difficult (IBM, 2021).
2. **Accuracy:** Spelling errors and ambiguous phrases (e.g., "It's gr8!") obscure true sentiment.
3. **Relevance:** Many emails discuss unrelated topics, diluting actionable insights (Kimball & Caserta, 2004).

References

1. Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, 41(3), 1–52. <https://doi.org/10.1145/1541880.1541883>
2. IBM. (2021). *What is data quality?* IBM Cloud Learn Hub. <https://www.ibm.com/cloud/learn/data-quality>
3. ISO 8000. (2023). *Data quality — Part 1: Overview*. International Organization for Standardization.
4. Kimball, R., & Caserta, J. (2004). *The data warehouse ETL toolkit*. Wiley.
5. Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–34. <https://doi.org/10.1080/07421222.1996.11518099>