

Data Quality Issues – Report Section

1. Invalid Dates

- Several records contain impossible dates such as:
- *2025-02-30*
2023-13-01

Multiple date formats appear in the same column (e.g., *2025-02-30*, *8/2/2024*, *2/10/2020*).

→ These issues prevent accurate time-based analysis.

2. Inconsistent Date Formats

Dates are written using different styles (YYYY-MM-DD and MM/DD/YYYY).

→ The date column must be standardized before processing.

3. Negative or Unrealistic Quantity Values

Some transactions include negative quantities (e.g., -5, -7, -1).

→ Negative quantities are illogical in sales data and should be corrected or removed.

- **4. Invalid or Inconsistent Price Values**
- Presence of negative prices (e.g., -445.34, -763).

Some prices include currency symbols such as "\$420.21" .

Mixed numeric formats across the column.

→ The price column needs to be cleaned and converted to numeric values only.

5. Inconsistent Payment Method Labels

Payment methods appear with different spellings, such as:

"credit card"

"PayPal"

→ Standardization is required to unify labels into consistent categories.

6. Inconsistent or Misspelled Transaction Status

Variations include:

"completed" , "Completed" , "complete"

"pending" , "Pending"

→ These should be normalized to a fixed set (e.g., *Completed, Pending, Failed*)

7. Product Name Inconsistencies

Misspelled or inconsistent product names, such as:

Tab" instead of "Tablet" .

Extra spaces or different capitalization.

→ Product names require cleaning to maintain consistent categories.

8. Missing or Duplicated Customer IDs

Some Customer_ID values may be missing, incorrect, or repeated.

→ This affects customer-level tracking and reliability of analysis.

9. Non-Numeric Values in Numeric Columns

Text or symbols appear in numeric fields such as Price or Quantity.

→ These must be cleaned to ensure proper statistical calculations.

Possible Duplicate Transactions

9. The dataset may contain repeated Transaction_ID values or duplicate rows.

→ **Duplicate detection and removal are necessary for accurate reporting.**