

Library Project - Data Engineering Dashboard

Data Quality Summary

Books Dataset	Customers Dataset	Data Quality Issues
Original Records 114	Original Records 9	Invalid Dates Fixed 2
Cleaned Records 20	Cleaned Records 10	Overdue Loans Detected 5
Rows Removed 95	Customers Added 2	Missing Return Dates 0

Data Quality Metrics

Loan Period (days)	Avg Days Borrowed	Max Days Overdue	Unknown Customers
2 weeks	6.2	16	2

Overdue Loans Details

ID	Book Title	Customer ID	Days Overdue
5	Little Women	1	1
11	Little Women	1	1
15	Dark Tales	2	2
18	Dracula	10	10
19	Frankenstein	2	2

Data Cleaning Process Log

[Full Log](#) [Errors Only](#) [Warnings Only](#)

Complete Cleaning Log

Full log output

```
2026-01-09 14:10:07,593 - INFO - =====
2026-01-09 14:10:07,594 - INFO - Library System Data Cleaning & Validation Started
2026-01-09 14:10:07,594 - INFO - =====
2026-01-09 14:10:07,594 - INFO - Loading data from:
2026-01-09 14:10:07,594 - INFO - Books: 03_Library Systembook.csv
2026-01-09 14:10:07,594 - INFO - Customers: 03_Library SystemCustomers.csv
2026-01-09 14:10:07,599 - INFO - Data loaded successfully: 114 books records, 9 customers records
2026-01-09 14:10:07,600 - INFO - Starting data quality analysis
2026-01-09 14:10:07,609 - INFO - Books data: Total rows: 114
2026-01-09 14:10:07,611 - WARNING - Books - Rows with NaN in Id: 93
2026-01-09 14:10:07,613 - WARNING - Books - Rows with NaN in Books: 94
2026-01-09 14:10:07,613 - WARNING - Books - Rows with NaN in Customer ID: 94
2026-01-09 14:10:07,614 - WARNING - Books - Completely empty rows: 93
2026-01-09 14:10:07,615 - ERROR - Invalid date found - Row 6: 10/04/2063 (year 2063 in future)
2026-01-09 14:10:07,615 - ERROR - Invalid date found - Row 16: 32/05/2023 (day 32 invalid)
2026-01-09 14:10:07,632 - INFO - Customers data: Total rows: 9
```

Cleaned Data Preview

Books Data

	Id	Books	Book checkout	Book Returned	Date	Customer ID	Customer Name
0	1	Catcher in the Rye	20/02/2023	25/02/2023	2023-02-25	0	1 Jane Doe
1	2	Lord of the rings the two towers	24/03/2023	21/03/2023	2023-03-21	1	2 John Smith
2	3	Lord of the rings the return of the	29/03/2023	25/03/2023	2023-03-25	2	3 Dan Reeves
3	4	The hobbit	02/04/2023	25/03/2023	2023-03-25	3	4 Unknown Custo
4	5	Dune	02/04/2023	25/03/2023	2023-03-25	4	5 William Holden
5	6	Little Women	02/04/2023	01/05/2023	2023-04-01	5	6 Jaztyn Forest
6	7	IT	10/04/2023	03/04/2023	2023-04-03	6	7 Jackie Irving
7	8	Misery	15/04/2023	03/04/2023	2023-04-03	7	8 Matthew Stirling
8	9	Catch 22	15/04/2023	16/04/2023	2023-04-16	8	9 Emory Ted
9	10	Animal Farm	20/04/2023	24/04/2023	2023-04-24	9	10 Unknown Custo

Showing 10 of 20 records

Customers Data

Showing 10 of 10 records

>About This Dashboard

This dashboard visualizes the results of the library system data cleaning process. It shows:

- **Data Quality Issues Found:** Missing values, invalid dates, referential integrity problems
- **Cleaning Actions Taken:** Row removal, date fixing, customer additions
- **Final Data Quality:** Metrics on overdue loans, data completeness, and validity
- **Audit Trail:** Complete log of all cleaning steps for transparency and reproducibility

For more information, check the `library_cleaning.log` file or re-run the cleaning script with the `--save-to-db` flag to save data to SQLite.