



Library Project - Data Engineering Dashboard



Data Quality Summary

Books Dataset

Original Records

114

Cleaned Records

20

Rows Removed

95

Customers Dataset

Original Records

9

Cleaned Records

10

Customers Added

2

Data Quality Issues

Invalid Dates Fixed

2

Overdue Loans Detected

5

Missing Return Dates

0



Data Quality Metrics

Loan Period (days)

2 weeks

Avg Days Borrowed

6.2

Max Days Overdue

16

Unknown Customers

2

Overdue Loans Details

	Id	Books	Customer ID	days
5	6	Little Women		1
11	12	Little Women		1
15	16	Dark Tales		2
18	19	Dracula		10
19	20	Frankenstein		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
0	1	Catcher in the Rye	20/02/2023	25/02/2023	21/02/2023
1	2	Lord of the rings the two towers	24/03/2023	21/03/2023	21/03/2023
2	3	Lord of the rings the return of the	29/03/2023	25/03/2023	21/03/2023
3	4	The hobbit	02/04/2023	25/03/2023	21/03/2023
4	5	Dune	02/04/2023	25/03/2023	21/03/2023
5	6	Little Women	02/04/2023	01/05/2023	21/03/2023
6	7	IT	10/04/2023	03/04/2023	21/03/2023
7	8	Misery	15/04/2023	03/04/2023	21/03/2023
8	9	Catch 22	15/04/2023	16/04/2023	21/03/2023
9	10	Animal Farm	20/04/2023	24/04/2023	21/03/2023

Showing 10 of 20 records

Customers Data

	Customer ID	Customer Name
1	1	Jane Doe
2	2	John Smith
3	3	Dan Reeves
4	4	Unknown Customer
5	5	William Holden
6	6	Jaztyn Forest
7	7	Jackie Irving
8	8	Matthew Stirling
9	9	Emory Ted
10	10	Unknown Customer

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.