

Clean-to-Report

Automated Data Cleaning & Reporting Pipeline



Project Snapshot

Problem: Messy business data + manual reporting

Solution: Automated cleaning pipeline + live dashboard

Outcome: Clean data, faster insights, zero manual effort

Role: Data Engineer

Tools: Python, Pandas, SQL, Power BI / Google Data Studio

Objective

Build an **end-to-end data pipeline** that:

- Cleans raw business data
- Automates processing
- Delivers business-ready reports

Designed for **small businesses** and **non-technical users**.

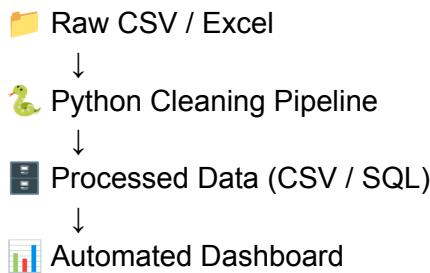
Business Problem

- Duplicate and missing records
- Inconsistent formats (dates, names, prices)
- Manual Excel reporting every month
- No reliable KPIs

Solution Overview

- ✓ Automated data cleaning
- ✓ Standardized datasets
- ✓ Scheduled processing
- ✓ Auto-updating dashboard

System Architecture



Tech Stack

Layer	Technology
Language	Python
Data Processing	Pandas
Storage	CSV / SQLite
Automation	Cron / Task Scheduler
Visualization	Power BI / Google Data Studio
Version Control	Git & GitHub

Dataset

Simulated Small Business Sales Data

- Order ID
- Date
- Product
- Category
- Quantity
- Price
- Revenue

 Raw data intentionally includes errors to mimic real-world data.



Data Cleaning Pipeline

- ✓ Remove duplicate rows
 - ✓ Standardize date formats
 - ✓ Handle missing values
 - ✓ Normalize text fields
 - ✓ Validate numeric columns
 - ✓ Create derived metrics
-



Automation Flow

- ⌚ Scheduled pipeline execution
- 📦 New data triggers processing
- 🔄 Dashboard refreshes automatically

No manual steps required after setup.



Dashboard Highlights

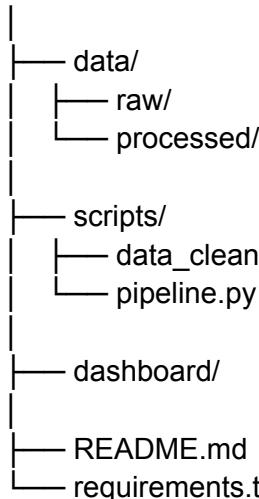
- 📈 Monthly sales trends
- 🏆 Top-selling products
- 💰 Total revenue
- 📦 Category-wise performance

Designed for quick business decision-making.



Project Structure

clean-to-report-pipeline/





Results & Impact

- Reduced manual reporting time
 - Fewer data errors
 - Reliable KPIs
 - Reusable pipeline for new datasets
-



Skills Demonstrated

- Data pipeline design
 - Data cleaning & validation
 - Automation & scheduling
 - Business-focused reporting
 - Documentation & version control
-



Future Enhancements

- Web upload interface
 - API integrations
 - Cloud deployment
 - Data quality alerts
-



Conclusion

This project demonstrates **real-world data engineering skills** by transforming messy business data into clean, automated, decision-ready insights.

Riadh Bouhamed