Case Study: Professional Data Cleaning and Dataset Validation (Retail Sales)

From Raw, Inconsistent Data to a Clean, Analysis-Ready Dataset

Tools Used: Microsoft Excel

Project Overview

This project demonstrates a complete professional data cleaning and validation workflow using a retail sales dataset. The objective was to audit data quality, correct errors, and produce a fully reliable, analysis-ready dataset for reporting and business decision making.

High-quality data cleaning is the foundation of any successful analytics or reporting project. This case study reflects the exact process applied in real client engagements.

Business Problem

Raw business datasets often contain missing values, duplicates, inconsistent formats, and logical errors. If these issues are not addressed, all downstream analysis becomes unreliable. The goal of this project was to systematically identify and eliminate data quality risks before any business decisions are made.

Raw Dataset Audit

The project started with a full audit of the raw dataset. Data structure, column definitions, value ranges, missing fields, duplicates, and formatting inconsistencies were carefully reviewed. Key business fields such as dates, product attributes, quantities, and sales values were validated for logical correctness.

Data Cleaning and Standardization Process

Based on the audit findings, a structured cleaning process was applied using Excel. This included removing duplicate records, correcting invalid data types, standardizing text fields, handling missing values, and validating numerical ranges. All transformations were documented for transparency.

Quality Control and Reporting

All cleaning steps and data quality improvements were summarized in a final quality report. This report serves as an audit trail that documents what issues were found, how they were corrected, and what level of data reliability was achieved.

Final Dataset Validation

After cleaning, the dataset was re-validated to confirm consistency, completeness, and analytical readiness. Final checks ensured that the cleaned data could be safely used for reporting, dashboards, and further data analysis without risk of distortion.

Skills Demonstrated

- Data Quality Auditing
- Duplicate Detection and Removal
- Missing Value Treatment
- Data Type and Format Standardization
- Logical Data Validation
- Excel-Based Data Cleaning

- Data Quality Reporting

- Structured and Reproducible Cleaning Workflows

Client Value

- Reliable, analysis-ready dataset

- Reduced risk of incorrect business decisions

- Full transparency of data corrections

- Improved reporting and dashboard accuracy

- Scalable cleaning workflow for future data updates

Summary

This case study showcases how raw, imperfect business data can be transformed into a clean, trusted, decision-ready dataset using a structured and auditable Excel-based cleaning process. It reflects the professional standards applied when preparing client data for reporting, analytics, and business use.