

Data Cleaning and Preparation Project

Project Report

This report summarizes a complete data cleaning workflow completed on real public sales order data. The project includes exploration, quality checks, cleaning, validation, feature engineering, charts, and export files for catalogue presentation.

Project Scope

Area	Work Completed
Raw data review	Loaded and inspected sales order records.
Data quality	Checked missing values, duplicates, unique order IDs, dates, numeric fields, and formulas.
Cleaning	Standardized column names, trimmed text fields, parsed dates, and converted numeric values.
Validation	Validated revenue, cost, profit, positive values, and date sequence rules.
Outputs	Exported cleaned CSV, audit log, summary metrics, data dictionary, charts, notebook, and Python script.

Final Validation Results

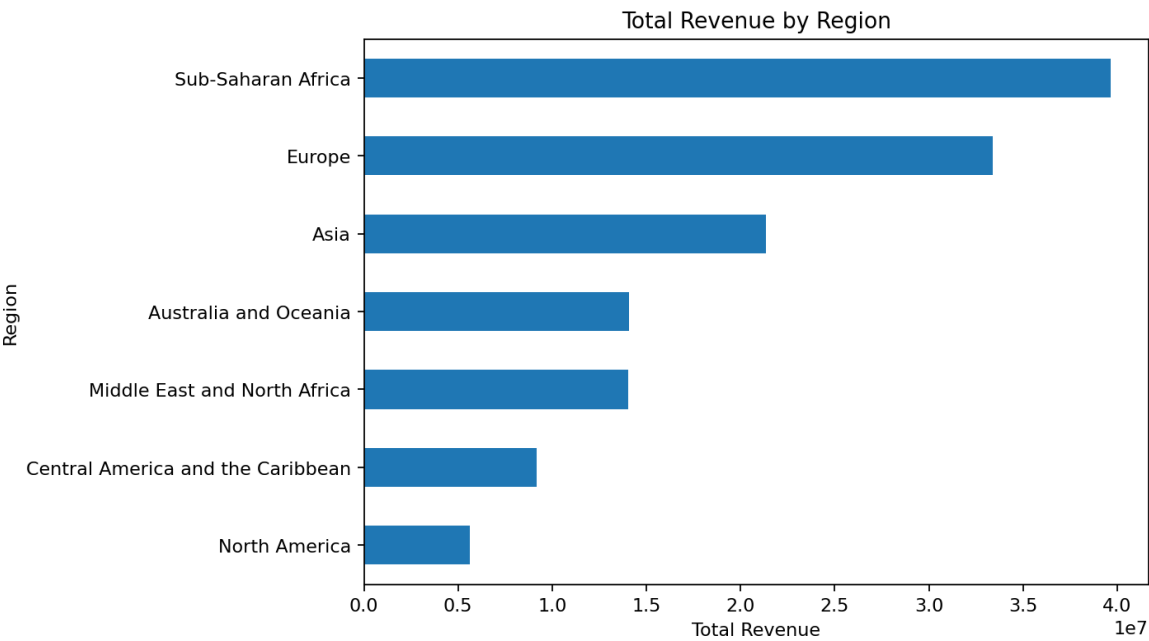
Check	Result
Missing Values Remaining	0
Duplicate Rows Remaining	0
Duplicate Order Ids Remaining	0
Date Sequence Errors	0
Revenue Formula Errors	0
Cost Formula Errors	0
Profit Formula Errors	0

Key Deliverables

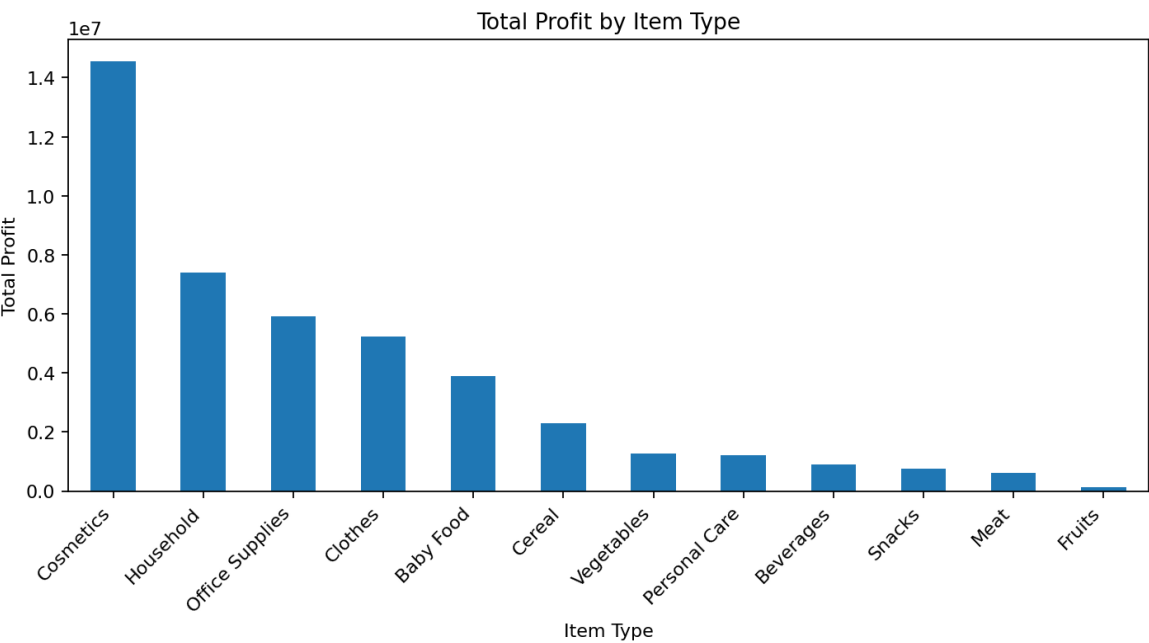
Raw CSV, cleaned CSV, Jupyter Notebook, Python script, data quality audit log, cleaning summary metrics, data dictionary, chart files, HTML notebook export, and PDF report.

Exploratory Charts

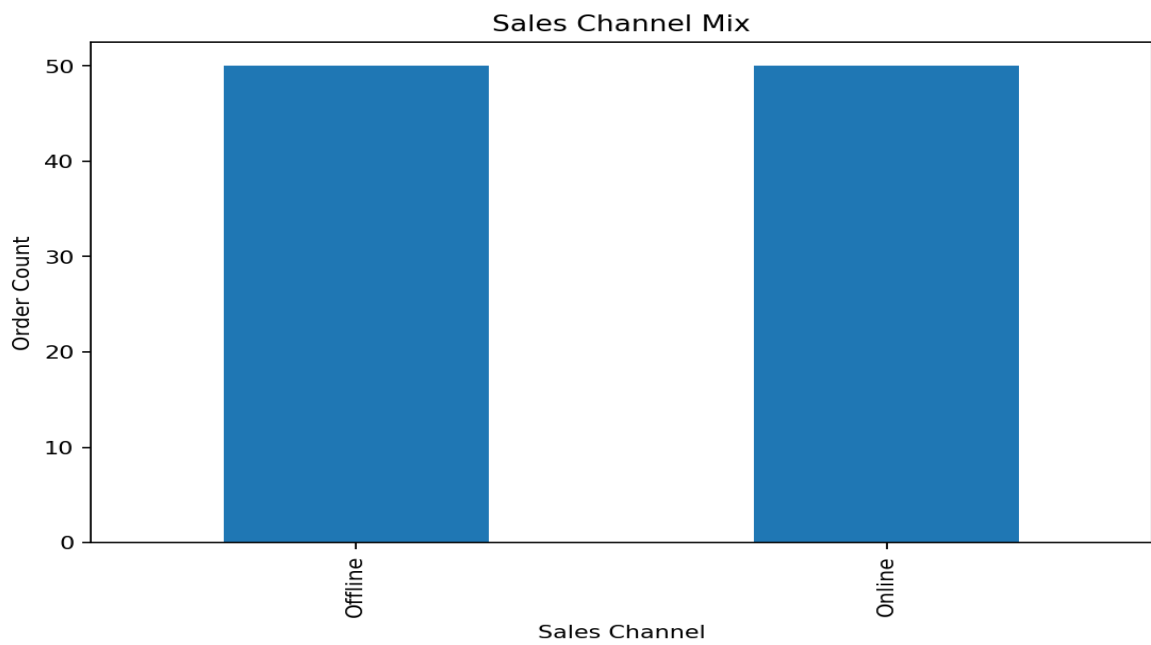
01 Total Revenue By Region



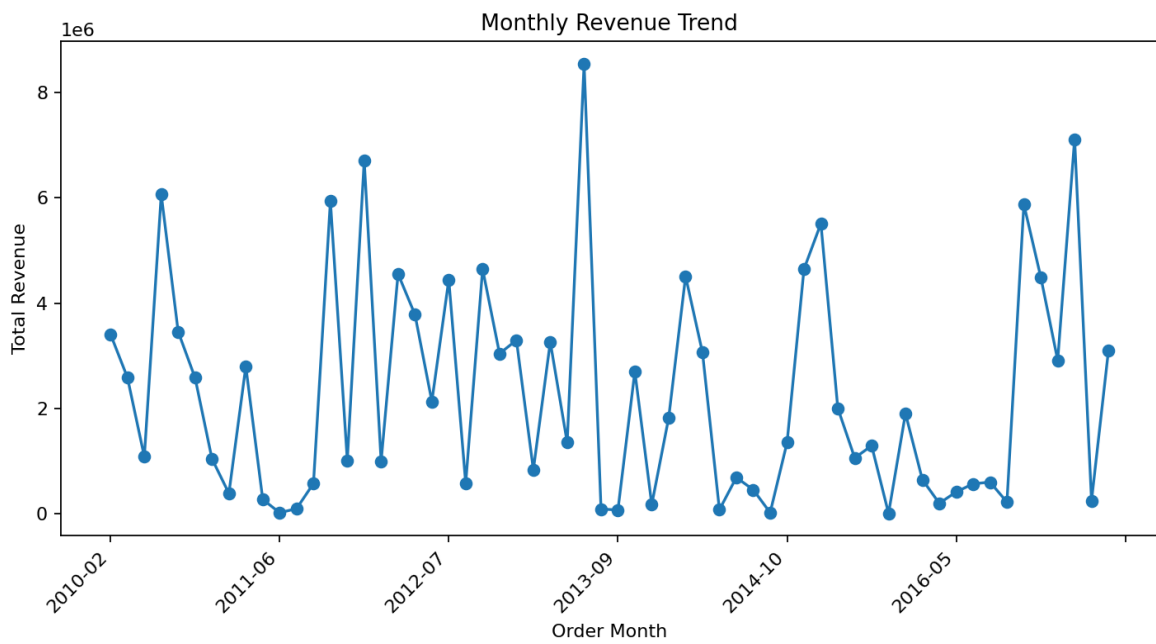
02 Total Profit By Item Type



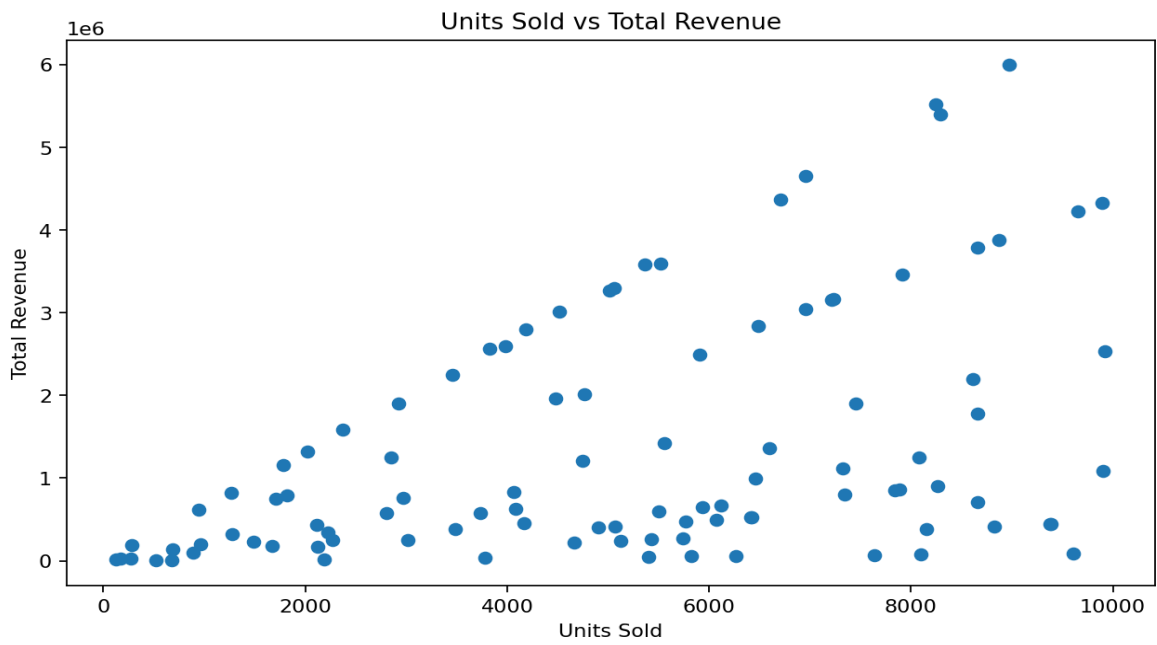
03 Sales Channel Mix



04 Monthly Revenue Trend



05 Units Vs Revenue Scatter



06 Correlation Matrix

