Power Query Cleaning Log

Retail Budget, Sales and Profit Forecasting Model

Ernest-Cyril Divine Onyeze | September 2026

Developed as a professional portfolio project to demonstrate practical data analytics skills to recruiters and hiring managers.

A record of the source preparation, implemented transformations and reported validation checks.

Source register

File	Query	Recorded rows
sales.csv	Sales_Raw	829,262
products.csv	Products_Raw	35
stores.csv	Stores_Raw	50
inventory.csv	Inventory_Raw	1,593
calendar.csv	Calendar_Raw	638
data_dictionary.csv	DataDictionary_Raw	18

Raw queries are recorded as connection-only sources. The query-list screenshot shows 12 queries: six raw, five clean, and Monthly_Category_Actuals. DataDictionary_Raw provides field definitions and is not an analytical dimension.

Preparation workflow

CSV files were imported through Data > Get Data > From Text/CSV, with Transform Data used to open Power Query. Headers and field definitions were checked. Dates were assigned Date types, identifiers/counts appropriate whole-number types, and product cost and price numeric types suitable for financial calculations. Currency text required conversion to numeric values before subtraction or multiplication could work.

Products_Clean visibly records Source, Trimmed Text, Cleaned Text, Changed Type and custom-column steps for Unit_Profit and Unit_Margin. Sales_Clean visibly records a product merge, expansion, four calculated columns and type changes. Calendar_Clean visibly records year, month, month name, quarter, month start and year-month attributes.

Permanent transformation log

Query	Transformation and purpose
Products_Clean	Reference Products_Raw; trim and clean names/categories; ensure numeric cost/price; add Unit_Profit and Unit_Margin.
Stores_Clean	Reference Stores_Raw; store attributes supply store, city and location reporting. Prior guidance included text cleaning and opening year; exact applied-step sequence is not shown in the supplied screenshots.
Calendar_Clean	Reference Calendar_Raw; add Year, Month_Number, Month_Name, Quarter_Number, Month_Start, Year_Month and Year_Month_Sort. Screenshot shows 638 rows.
Sales_Clean	Reference Sales_Raw; merge product cost/price; add Sales_Revenue, Total_Cost, Gross_Profit and Profit_Margin. Screenshot confirms transaction-level rows remain.
Inventory_Clean	Product-cost valuation supplies Inventory Value reporting. The model confirms the inventory table; the complete inventory applied-step sequence was not separately captured.
Monthly_Category_Actuals	Summarise sales by month and category. tblMonthlyActuals contains units, revenue, cost, profit, transaction lines and grouped margin across 105 rows.

Join and aggregation choices

Left Outer merges retain the sales or inventory source rows while bringing in matching product fields. Unique product keys prevent row multiplication. Left Anti joins were used as temporary audit tools to identify unmatched keys. The monthly aggregation is separate from the transaction-level Sales_Clean table so detailed store and product analysis remains possible.

Financial values are multiplied or subtracted before aggregation. Business margin is total gross profit divided by total sales, rather than the average of row-level percentages.

Recorded data-quality checks

Check	Historical result / evidence
Sales required fields and valid units	0 flagged rows; author reported
Products required fields and numeric prices/costs	0 flagged rows; author reported
Stores required fields and opening dates	0 flagged rows; author reported
Inventory required fields and nonnegative stock	0 flagged rows; author reported
Calendar blank dates, repeated dates and gaps	0 for each check; author reported
Duplicate Sale_ID, Product_ID and Store_ID	0 for each key; author reported
Duplicate inventory Store_ID + Product_ID	0 duplicate composite keys; author reported
Sales Product_ID / Store_ID absent from dimensions	0 for each Left Anti check; author reported
Inventory Product_ID / Store_ID absent from dimensions	0 for each Left Anti check; author reported
Sales dates absent from calendar	0 unmatched dates; author reported
Data dictionary	Visually reviewed; headers/field descriptions checked

These checks were performed during the project, according to the author. Their temporary reference queries are no longer present and were not rerun for this documentation. The workbook Raw Data note records a narrower subset and explicitly leaves other checks unconfirmed there. This log preserves the broader chat-confirmed results with their evidence status.

Independent saved-output reconciliation

tblMonthlyActuals totals 1,090,565 units and 829,262 transaction lines, matching Analysis headline counts. Its sales, cost and profit totals match the Analysis totals within less than $0.01. Sales $14,444,572.35 = cost $10,430,543.35 + gross profit $4,014,029.00. All 45 budget rows matched the examined sales-budget and variance arithmetic; all 15 forecast cost rows matched the stated cost formula.

Problems addressed and lessons learned

Issue	Resolution or lesson from the work history
Power Query access in older Excel	The project moved to Excel 2019 and used its Data/Get Data workflow.
Errors in Unit_Profit	Cost and price needed numeric types before subtraction; fixed-decimal financial types were used.
Merge dialog could not proceed	Matching join columns and compatible ID types were checked; key-type differences were addressed during the workflow.
Filter list reached 1,000 values	The filter dropdown is not a full-dataset quality audit. Dedicated check queries and row counts were used instead.
Temporary checks deleted	Raw queries remained. Historical results were retained, but future source changes require recreation and rerunning.
Overall margin interpretation	Use total profit / total sales; do not average transaction margins.
Excel 2019 lacks XLOOKUP	INDEX/MATCH was used for scenario and category target lookups.
Forecast chart refresh	Scenario-driven formulas recalculate, but the PivotTable-based chart must be refreshed separately.

These items describe the documented learning and implementation process. They do not imply that every proposed transformation was executed exactly as originally suggested. The final workbook and screenshots determine the implementation claims.

Evidence: supplied completed XLSX workbook, 16 project screenshots and project-work conversation history. Source attribution: Mexico Toy Sales, Maven Analytics (https://mavenanalytics.io/data-playground). Dollar symbols follow the supplied workbook and source price formatting; this documentation does not independently establish a currency code.