

# Data Cleaning & Structuring Using Power Query

## From Raw Combined Data to Analysis-Ready Structured Dataset

This case study demonstrates how raw appended data extracted from multi-page PDF reports was systematically cleaned, standardized, and transformed into a structured, analysis-ready dataset using Microsoft Excel Power Query.

**The objective was to ensure consistency, accuracy, and business usability for reporting and dashboard development.**

# Data Cleaning & Structuring Using Power Query

## Data Quality Issues Identified After Append

After combining multiple PDF tables into a single dataset, the following data quality challenges were identified:

- Inconsistent column names
- Incorrect data types (dates stored as text)
- Repeated header rows within data
- Blank rows and null values
- Mixed numeric and text formatting
- Potential duplicate transactions

# Power Query Cleaning Workflow

**Step 1:** Promoted first row as headers to standardize column naming.

**Step 2:** Renamed columns to business-friendly names

(Order Date, Product Name, Category, Region, Quantity, Sales, Profit).

**Step 3:** Corrected data types

- Date → Date format
- Quantity → Whole Number
- Sales & Profit → Decimal Number

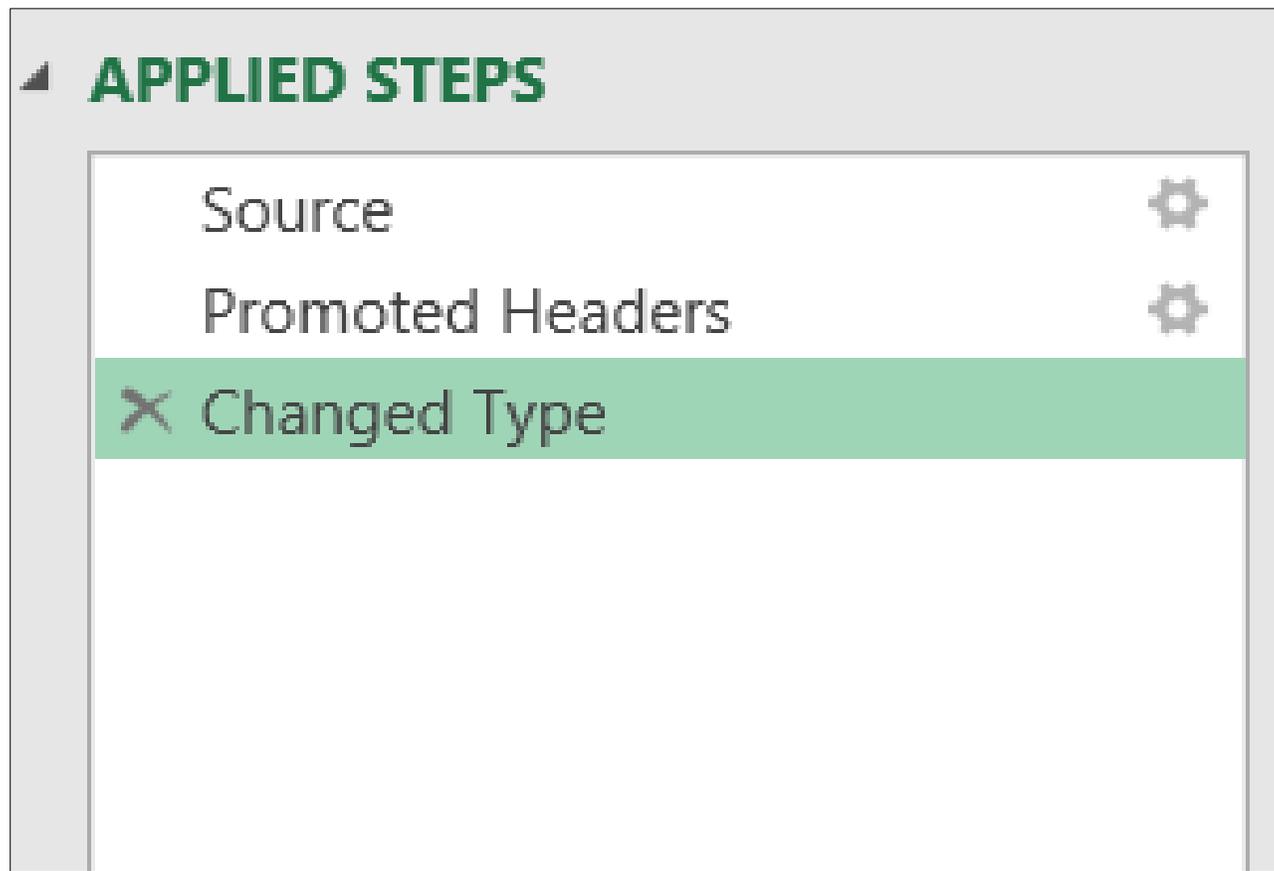
**Step 4:** Removed blank rows and null records.

**Step 5:** Removed duplicate transactions based on key columns.

**Step 6:** Validated numeric consistency and ensured no conversion errors.

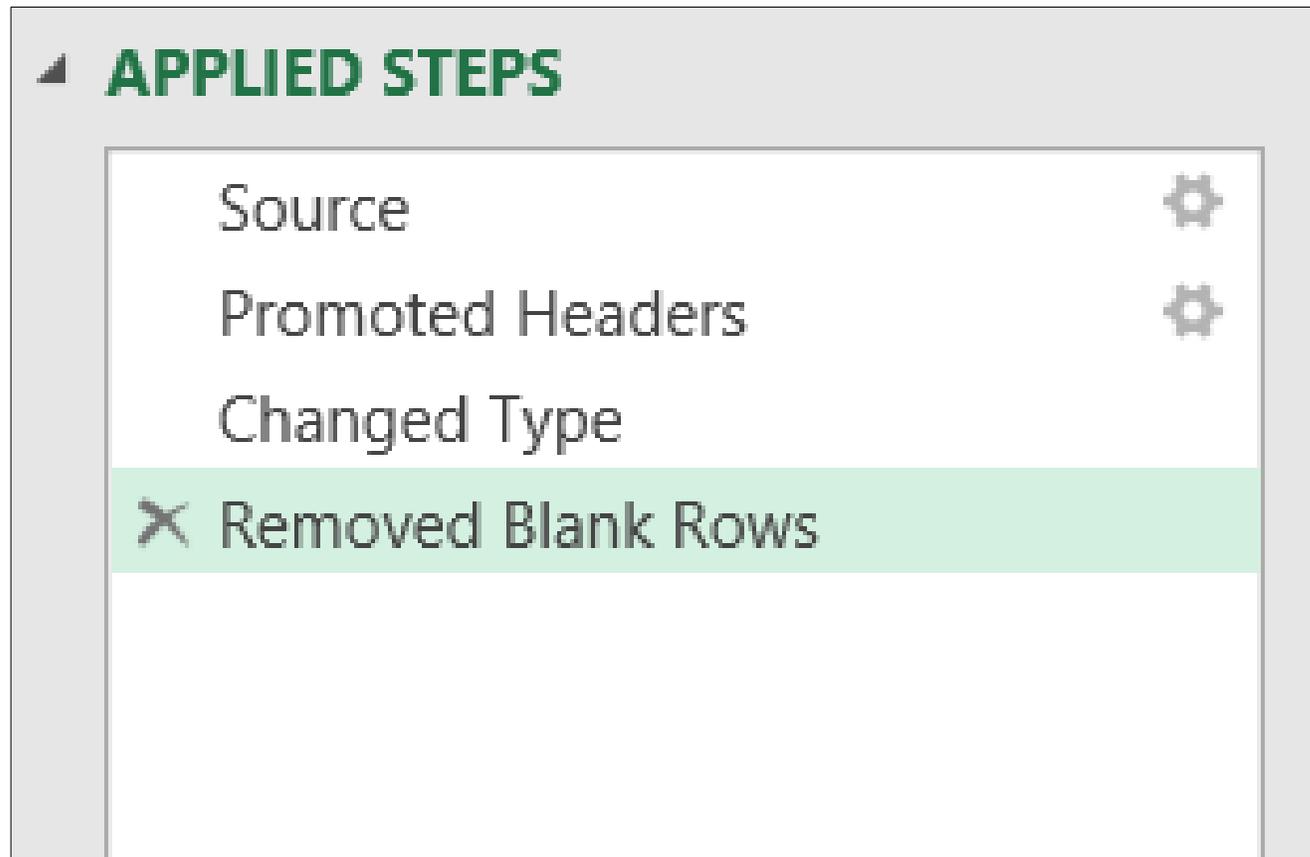
**Step 7:** Final structured dataset loaded into Excel Data Model.

Data types were standardized to ensure correct formatting for dates, text fields, and numeric columns.



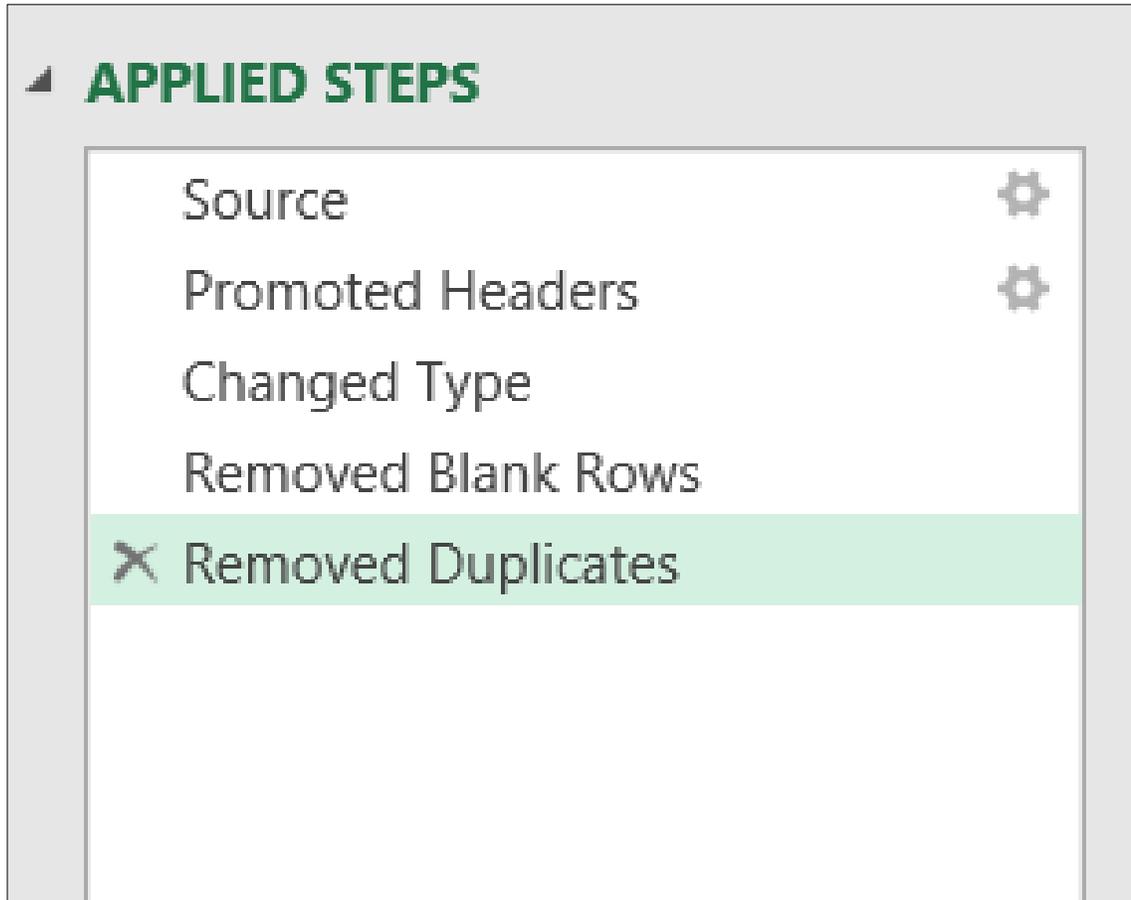
Data types converted using Power Query automatic type detection and manual correction.

**Blank and unnecessary rows were removed to improve data structure and reporting accuracy.**



Blank and null rows removed using Remove Rows → Remove Blank Rows function

**Duplicate records were identified and removed to maintain data integrity and consistency.**



Duplicates removed based on key transaction fields to ensure data integrity.

# Final cleaned and structured dataset prepared for reporting, dashboard creation, and analysis.

	Order Date	Product Name	Category	Region	Quantity	Sales	Profit
1	31-12-2024	Printer	Office	North	4	3640	
2	27-11-2022	Mouse	Accessories	East	7	1197	
3	11-05-2022	Tablet	Electronics	South	5	5865	
4	16-03-2024	Mouse	Accessories	South	2	786	
5	10-09-2022	Mouse	Accessories	West	1	509	
6	01-12-2023	Camera	Electronics	West	1	524	
7	09-10-2023	Headphones	Accessories	North	7	6167	
8	14-01-2022	Camera	Electronics	South	7	3059	
9	02-04-2022	Smartwatch	Electronics	East	9	5526	
10	22-10-2024	Printer	Office	South	8	672	
11	04-12-2023	Monitor	Accessories	South	6	7074	
12	27-09-2022	Smartwatch	Electronics	East	2	502	
13	21-12-2023	Tablet	Electronics	North	7	7462	
14	03-02-2024	Smartphone	Electronics	West	5	4205	
15	10-10-2023	Camera	Electronics	West	1	570	
16	16-10-2022	Headphones	Accessories	East	1	441	
17	15-05-2022	Mouse	Accessories	North	7	7469	
18	11-02-2023	Tablet	Electronics	South	2	2212	
19	27-12-2023	Tablet	Electronics	East	5	740	
20	05-10-2023	Headphones	Accessories	North	1	662	
21	05-03-2024	Mouse	Accessories	North	7	5712	
22	18-07-2023	Smartwatch	Electronics	North	3	2739	
23	25-06-2023	Keyboard	Accessories	West	5	1400	
24	13-11-2023	Printer	Office	North	3	3477	
25	09-07-2022	Camera	Electronics	North	5	3475	
26	11-12-2023	Laptop	Electronics	West	9	6138	
27	18-03-2023	Camera	Electronics	West	5	3665	
28	09-08-2024	Printer	Office	North	2	2372	
29	04-05-2024	Laptop	Electronics	North	2	1382	
30	04-09-2022	Laptop	Electronics	West	1	1126	
31	21-10-2022	Printer	Office	West	8	1376	
32	22-01-2024	Mouse	Accessories	West	4	988	

## Business Impact:

- 100% standardized structure
- Error-free numeric formatting
- Ready for pivot tables, dashboards, and BI tools