

# PDF to Excel Data Extraction

## Real-World Case Study

Automated Multi-Page PDF Conversion Using Excel & Power Query

Prepared By  
Izharul Haque

# MONTHLY SALES PERFORMANCE REPORT

Region: All Zones  
Prepared By: Sales Department

## Executive Summary:

This report contains raw transactional sales data extracted from operational systems. The data includes product details, quantities, pricing, and regional performance. Formatting inconsistencies exist across pages.

## Detailed Transaction Data (Page 1)

Order Date	Product Name	Category	Region	Quantity	Sales	Profit
2024-12-31	Printer	Office	North	4	3640	348.93
2022-11-27	Mouse	Accessories	East	7	1197	106.53
2022-05-11	Tablet	Electronics	South	5	5865	502.73
2024-03-16	Mouse	Accessories	South	2	786	202.87
2022-09-10	Mouse	Accessories	West	1	509	103.28
2023-12-01	Camera	Electronics	West	1	524	106.35
2023-10-09	Headphones	Accessories	North	7	6167	1027.98
2022-01-14	Camera	Electronics	South	7	3059	873.5
2022-04-02	Smartwatch	Electronics	East	9	5526	595.28
2024-10-22	Printer	Office	South	8	672	186.37
2023-12-04	Monitor	Accessories	South	6	7074	1357.68
2022-09-27	Smartwatch	Electronics	East	2	502	137.4
2023-12-21	Tablet	Electronics	North	7	7462	2166.17
2024-02-03	Smartphone	Electronics	West	5	4205	585.29
2023-10-10	Camera	Electronics	West	1	570	138.54

## SALES DATA CONTINUED

Order Date	Product Name	Category	Region	Quantity	Sales	Profit
2022-10-16	Headphones	Accessories	East	1	441	102.42
2022-05-15	Mouse	Accessories	North	7	7469	589.81
2023-02-11	Tablet	Electronics	South	2	2212	145.75
2023-12-27	Tablet	Electronics	East	5	740	146.39
2023-10-05	Headphones	Accessories	North	1	662	193.92
2024-03-05	Mouse	Accessories	North	7	5712	651.27
2023-07-18	Smartwatch	Electronics	North	3	2739	572.73
2023-06-25	Keyboard	Accessories	West	5	1400	213.63
2023-11-13	Printer	Office	North	3	3477	939.2
2022-07-09	Camera	Electronics	North	5	3475	570.36
2023-12-11	Laptop	Electronics	West	9	6138	1562.14
2023-03-18	Camera	Electronics	West	5	3665	406.93
2024-08-09	Printer	Office	North	2	2372	206.42
2024-05-04	Laptop	Electronics	North	2	1382	211.11
2022-09-04	Laptop	Electronics	West	1	1126	194.08

## Notes & Observations

- Some records may contain missing values.
- Date formats may vary across departments.
- Duplicate transaction IDs may require validation.

### Additional Raw Data Snapshot:

Order Date	Product Name	Category	Region	Quantity	Sales	Profit
2023-05-14	Laptop	Electronics	West	9	9702	2771.49
2023-09-04	Keyboard	Accessories	East	8	5888	354.27
2023-09-20	Tablet	Electronics	East	4	1768	345.33
2022-06-01	Headphones	Accessories	North	2	1022	76.04
2023-12-03	Tablet	Electronics	North	3	2883	671.91
2023-02-15	Monitor	Accessories	North	4	4064	771.16
2022-09-18	Monitor	Accessories	East	1	1117	119.89
2022-04-12	Laptop	Electronics	South	4	260	66.02
2022-01-18	Printer	Office	South	3	222	50.28
2024-11-01	Smartwatch	Electronics	East	9	10530	1777.32

## Identified Data Challenges

- Multi-page PDF with inconsistent table formatting
- Repeated headers across pages
- Date formats not standardized
- Some missing values in key columns
- Mixed category naming conventions
- Manual extraction would require significant time and effort

## Power Query Automation Workflow

**Step 1:** Imported multi-page PDF directly into Excel using Power Query.

**Step 2:** Combined all pages into a single structured table.

**Step 3:** Removed repeated headers and blank rows.

**Step 4:** Standardized column names and corrected data types.

**Step 5:** Cleaned date formats and ensured numeric accuracy.

**Step 6:** Removed duplicates and validated transaction consistency.

**Step 7:** Loaded clean, analysis-ready dataset into Excel Data Model.

# Navigator

Navigator



Select multiple items

Display Options 

- └─  Messy\_Sales\_Report.pdf [7]
  -  Table001 (Page 2)
  -  Table002 (Page 3)
  -  Table003 (Page 4)
  -  Page001
  -  Page002
  -  Page003
  -  Page004

No item selected for preview

Load  Transform Data Cancel

## Append (Three or more tables selected)

Append

Concatenate rows from three or more tables into a single table.

Two tables  Three or more tables

Available tables

- Table001 (Page 2)
- Table003 (Page 4)
- Table002 (Page 3)

Tables to append

- Table001 (Page 2)

Add >>

^

x

v

OK Cancel

# Tables added and ready to append

Append ✕

Concatenate rows from three or more tables into a single table.

Two tables    Three or more tables

Available tables

Table001 (Page 2)
Table003 (Page 4)
Table002 (Page 3)

Add >>

Tables to append

Table001 (Page 2)
Table002 (Page 3)
Table003 (Page 4)

▲  
✕  
▼

OK   Cancel

# Final Cleaned & Standardized Dataset in Power Query

fx = Table.Combine({#"Table001 (Page 2)", #"Table002 (Page 3)", #"Table003 (Page 4)"})

	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	14-05-2023	Laptop	Electronics	West	9	9702	
32	04-09-2023	Keyboard	Accessories	East	8	5888	