

Optimizing Sales Data for Profit Insights

A Power BI Data Cleaning & Profitability Analysis
Project

Filename: Final_task.pbix

By: [Muthuraj.M]

Table of Contents

1. Introduction
2. Preliminary Data Cleaning
 - 2.1 Row Removal
 - 2.2 Date Column Cleaning and Standardization
 - 2.3 Text Capitalization
3. Main Workflow
 - 3.1 Clean Numeric Columns
 - 3.2 Calculate Unit Values per Product
 - 3.3 Fill Missing Values in Sales, Quantity, Profit, Discount
 - 3.4 Rounding Outputs
 - 3.5 Profit Percentage Calculations
- 4.1 New filtered sheet creation
- 4.2 Category Hierarchy
- 4.3 Dollar symbols added.
5. Why These Steps Matter
6. Final Notes
 - 7.1 Relationship between sheets.
 - 7.2 Final cleaned sheets.
8. Summary of Visuals Used in the Report.

1. Introduction

This project focuses on cleaning and optimizing sales data to enable accurate profit analysis and visualization in Power BI. The dataset contained missing values, inconsistent formats, and text mismatches that

required structured processing. An giving the clear visualisation for the analysis insight also for making good decision for upcoming sales

2. Preliminary Data Cleaning

2.1 Row Removal

Removed irrelevant, duplicate, or erroneous rows to ensure the dataset contains only valid data. This improves overall data quality and prevents analysis errors from bad data.

2.2 Date Column Cleaning and Standardization

Addressed inconsistent date formats such as mixed formats (10/06/2013), dates with time (01-08-2014 00:00:00), and invalid years. Used Power BI date filters and transformations to convert all dates to a uniform proper date format without time components. Ensured dates are recognized as valid date data types to enable accurate filtering, sorting, and time-based analysis.

2.3 Text Capitalization

Applied 'Capitalize Each Word' transformation on text columns like Category and Sub-Category. Standardized casing to avoid mismatches during grouping and aggregation (e.g., avoiding 'furniture' vs. 'Furniture' issues)

2.1 Row Removal

First Delete the null values on order id Because Order is important for the sales data.

Order ID
(Select All)
(null)
CA-2012-AD10180140-41231
CA-2014-AH10120140-41992
CA-2015-AD10180140-42322
CD-2014-SM1032025-41684
EG-2013-JR567038-41363
ES-2013-KN1639048-41352
ES-2014-AG1090045-41801
ES-2014-QJ1925564-41793

2.2 Date Column Cleaning and Standardization

Date rows are contained with time also so I try to remove and make tables to text type to date type

Screenshot of Power BI Data Editor showing the 'Split Column by Delimiter' dialog and the transformed data table.

Split Column by Delimiter Dialog:

- Specify the delimiter used to split the text column.
- Select or enter delimiter: Space
- Split at: Each occurrence of the delimiter (radio button selected)
- Quote Character: None
- Advanced options: Split using special characters (checkbox unchecked), Insert special character (button)

Transformed Data Table:

Order Date.1	Order Date.2	Order Date.3	Ship Date.1	Ship Date.2
01-05-2012 00:00:00	02-05-2012 00:00:00	Jason Klamczynski	China	Eastern Asia
27-02-2014 00:00:00	01-03-2014 00:00:00	Laurel Beltran	United Kingdom	Northern Europe
31-07-2015 00:00:00	01-08-2015 00:00:00	Naresj Patel	Mexico	Central America
05-09-2015 00:00:00	06-09-2015 00:00:00	Valarie Dominguez	El Salvador	Central America

Power BI Data Editor Context Menu:

- Copy
- Remove
- Remove Other Columns
- Duplicate Column
- Add Column From Examples...
- Remove Duplicates
- Remove Errors
- Change Type
- Transform
- Replace Values...
- Replace Errors...
- Split Column
- Group By...
- Fill
- Unpivot Columns
- Unpivot Other Columns
- Unpivot Only Selected Columns
- Rename...
- Move
- Drill Down
- Add as New Query

After removed unwanted column from date table some dates it more than 2105 so I use to change those year by creatin new calculated column also change it to date format.

Custom Column

Add a column that is computed from the other columns.

New column name: ship date

Custom column formula:

```
= try Date.FromText([Ship Date.1]) otherwise null
```

Available columns:

- Order ID
- Order Date.1
- Ship Date.1
- Customer Name
- Country
- Region
- Category

Learn about Power Query formulas

OK Cancel

Filter date up to 2015 only

Filter Rows

Apply one or more filter conditions to the rows in this table.

Basic Advanced

Keep rows where 'Order Date'

is before 31-12-2015

And Or

Enter or select a value...

OK Cancel

05-12-2015

06-12-2015

07-12-2015

08-12-2015

09-12-2015

10-12-2015

13-12-2015

17-12-2015

19-12-2015

20-12-2015

22-12-2015

24-12-2015

25-12-2015

29-12-2015

30-12-2015

31-12-2015

27-09-2102

26-02-2103

⚠ List may be incomplete. Load more

OK Cancel

Main Workflow

3.1 Clean Numeric Columns

Created clean versions of columns (CleanSales, CleanQuantity, CleanProfit, CleanDiscount) by:

- Removing invalid text entries such as 'missing' or 'unknown'.
- Converting text to numeric types to enable calculations.

3.2 Calculate Unit Values per Product

Created measures to compute average per-unit values from historical data with no missing values:

- UnitSalesPerProduct
- UnitProfitPerProduct
- UnitDiscountPerProduct

These provide normalized sales, profit, and discount values per single product unit.

3.3 Fill Missing Values in Sales, Quantity, Profit, Discount

Created filled columns (FilledSales, FilledQuantity, FilledProfit, FilledDiscount) where missing values are replaced by calculated estimates based on unit values and related fields.

Retained original values where present.

Applied conditional logic to avoid incorrect or circular calculations.

3.4 Profit Percentage Calculations

Created a calculated column ProfitPercentagePerOrder:

Formula: $(\text{FilledProfit} / \text{FilledSales}) * 100$

Rounded to 3 decimals.

Includes checks to handle division by zero or blank values safely.

Created a measure UnitProfitPercentagePerProduct:

Formula: $(\text{UnitProfitPerProduct} / \text{UnitSalesPerProduct}) * 100$

Rounded to 3 decimals.

Gives average profit % per product category for summary and insights.

3.5 Rounding Outputs

Rounded all filled columns to 3 decimals using ROUND(..., 3) for professional and consistent data presentation.

3.1 Clean Numeric Columns

Created clean versions of columns (CleanSales, CleanQuantity, CleanProfit, CleanDiscount) by:

```
1 CleanDiscount =           1 CleanProfit =
2 IF(                         2 IF(
3   ISBLANK(Orders[Discount]) 3   ISBLANK(Orders[Profit])
4   || Orders[Discount] & "" = "missing" 4   || Orders[Profit] & "" = "missing"
5   || Orders[Discount] & "" = "Unknown", 5   || Orders[Profit] & "" = "Unknown",
6   BLANK(),                  6   BLANK(),
7   VALUE(Orders[Discount])    7   VALUE(Orders[Profit])
8 )                           8 )
9                               9
10
```

```
1 CleanQuantity =
2 IF(
3   ISBLANK(Orders[Quantity]) || Orders[Quantity] = "missing" || Orders[Quantity] = "Unknown",
4   BLANK(),
5   VALUE(Orders[Quantity])
6 )
7
```

```
CleanSales =
IF(
  ISBLANK(Orders[Sales]) || Orders[Sales] = "missing" || Orders[Sales] = "Unknown",
  BLANK(),
  VALUE(Orders[Sales])
)
```

Final output clean data calculated columns

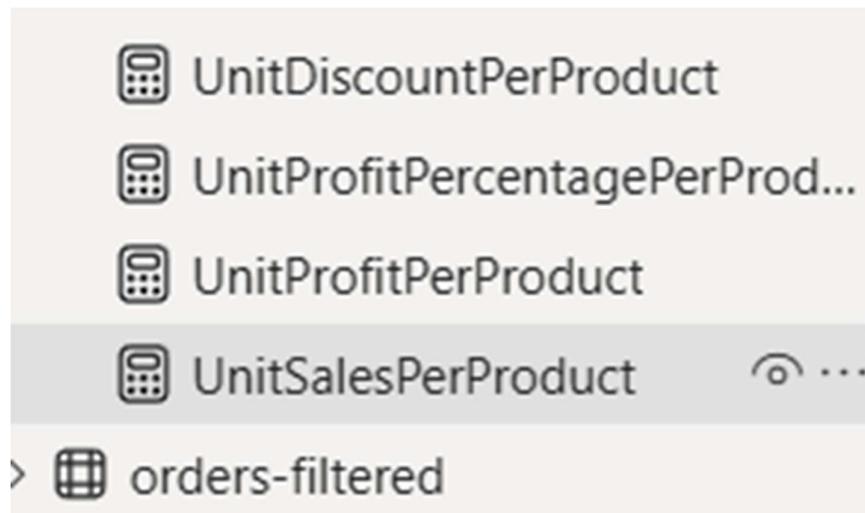
CleanSales	CleanQuantity	CleanProfit	CleanDiscount
323.10	2	61.389	0
18.54	2	8.71379999999999	0
33.40		16.032	0
20.96	2	5.24	0
30.40	5	15.2	0
	3	17.0268	0
7.64	1		0
52.98	2	14.8344	0
35.40	5	13.452	0
9.24	3	4.4352	0

3.2 Calculate Unit Values per Product

Created measures to compute average per-unit values from historical data with no missing values:

```
1 UnitDiscountPerProduct =
2 CALCULATE(
3     DIVIDE(
4         SUM(Orders[CleanDiscount]),
5         SUM(Orders[CleanQuantity])
6     ),
7     ALLEXCEPT(Orders, Orders[Sub-Category]) )
8 )
9
10 UnitSalesPerProduct =
11 CALCULATE(
12     DIVIDE(
13         SUM(Orders[CleanSales]),
14         SUM(Orders[CleanQuantity])
15     ),
16     ALLEXCEPT(Orders, Orders[Sub-Category])
17 )
```

Measures are created in sheet:



3.3 Fill Missing Values in Sales, Quantity, Profit, Discount

Created filled columns (FilledSales, FilledQuantity, FilledProfit, FilledDiscount) where missing values are replaced by calculated estimates based on unit values and related fields.

```
FilledQuantity =
ROUND(
    IF(
        ISBLANK(Orders[CleanQuantity]) && NOT(ISBLANK(Orders[CleanSales])),
        DIVIDE(Orders[CleanSales], [UnitSalesPerProduct], BLANK()),
        Orders[CleanQuantity]
    ),
    3
)

FilledSales =
ROUND(
    IF(
        ISBLANK(Orders[CleanSales]) && NOT(ISBLANK(Orders[CleanQuantity])),
        Orders[CleanQuantity] * [UnitSalesPerProduct],
        Orders[CleanSales]
    ),
    3
)

FilledProfit =
ROUND(
    IF(
        ISBLANK(Orders[CleanProfit]) && NOT(ISBLANK(Orders[CleanQuantity])),
        Orders[CleanQuantity] * [UnitProfitPerProduct],
        Orders[CleanProfit]
    ),
    3
)

FilledDiscount =
ROUND(
    IF(
        ISBLANK(Orders[CleanDiscount]) && NOT(ISBLANK(Orders[CleanQuantity])),
        Orders[CleanQuantity] * [UnitDiscountPerProduct],
        Orders[CleanDiscount]
    ),
    3
)
```

From these try to fill the missing values also roundoff the digit values with three decimals.

3.4 Profit Percentage Calculations

Created a calculated column ProfitPercentagePerOrder:

Formula: (FilledProfit / FilledSales) * 100

Rounded to 3 decimals.

Includes checks to handle division by zero or blank values safely.

Created a measure UnitProfitPercentagePerProduct:

Formula: (UnitProfitPerProduct / UnitSalesPerProduct) * 100

Rounded to 3 decimals.

Gives average profit % per product category for summary and insights.

These calculated column and measures are created for my analysis purpose.

```
UnitProfitPercentagePerProduct =  
IF(  
    NOT(ISBLANK([UnitSalesPerProduct])) && [UnitSalesPerProduct] <> 0,  
    ROUND(  
        DIVIDE([UnitProfitPerProduct], [UnitSalesPerProduct]) * 100,  
        3  
    ),  
    BLANK()  
)  
  
. ProfitPercentagePerOrder =  
IF(  
    NOT(ISBLANK(Orders[FilledSales])) && Orders[FilledSales] <> 0,  
    ROUND(  
        DIVIDE(Orders[FilledProfit], Orders[FilledSales]) * 100,  
        3  
    ),  
    BLANK()  
)
```

4.1 New filtered sheet creation

After this entire filtration my current have more column so I create a new sheet and put the entire column which only we need for my proper analysis. Also, still some of the columns are not satisfied so create some filtered column again in my new filtered sheet. DO this entire thing using **DAX Formulas**.

```

FilledDiscount_fromOrders =
ROUND(
  COALESCE(
    AVERAGEX(
      FILTER(
        orders,
        orders[Order ID] = 'orders-filtered'[Order ID]
        && NOT(ISBLANK(orders[FilledDiscount])))
      ),
      orders[FilledDiscount]
    ),
    0
  ),
  2
)

FilledSales_fromOrders =
ROUND(
  COALESCE(
    AVERAGEX(
      FILTER(
        orders,
        orders[Order ID] = 'orders-filtered'[Order ID]
        && NOT(ISBLANK(orders[FilledSales])))
      ),
      orders[FilledSales]
    ),
    0
  ),
  2
)

FilledQuantity_fromOrders =
ROUND(
  COALESCE(
    AVERAGEX(
      FILTER(
        orders,
        orders[Order ID] = 'orders-filtered'[Order ID]
        && NOT(ISBLANK(orders[FilledQuantity])))
      ),
      orders[FilledQuantity]
    ),
    0
  ),
  2
)

FilledProfit_fromOrders =
ROUND(
  COALESCE(
    AVERAGEX(
      FILTER(
        orders,
        orders[Order ID] = 'orders-filtered'[Order ID]
        && NOT(ISBLANK(orders[FilledProfit])))
      ),
      orders[FilledProfit]
    ),
    0
  ),
  2
)

```

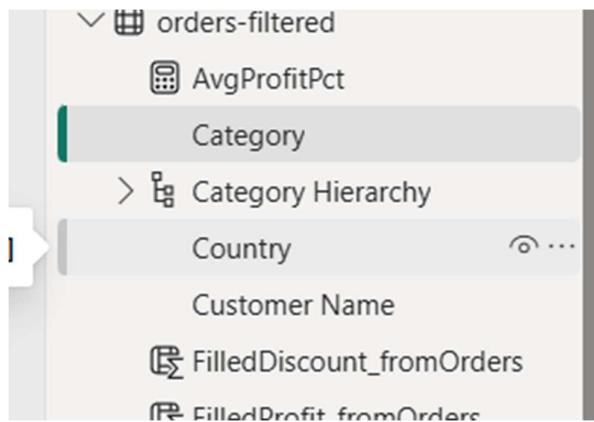
After this cleaning process I found my shipping date are more relevant to the current data so delete the column of shipping date column also my order date column is not in date type so using the filter setting to change it.

	Order Date	Customer Name	Country
41831	11 July 2014	Allen Armold	United
41247	04 December 2012	Brendan Sween	United
41068	08 June 2012	Brosina Hoffman	United
41754	25 April 2014	Craig Leslie	United
41251	08 December 2012	Cyma Kinney	United
41269	26 December 2012	Cynthia Delaney	United
41811	21 June 2014	Dave Poirier	United
41159	07 September 2012	Dennis Pardue	United
0936	28 January 2012	Ed Jacobs	United
0936	28 January 2012	Ed Jacobs	United
41299	25 January 2013	Elizabeth Moffitt	United
41171	19 September 2012	Erica Smith	United
41076	16 June 2012	Gary Hwang	United
41614	06 December 2013	Gary Hwang	United
41614	06 December 2013	Gary Hwang	United
-41038	09 May 2012	Greg Matthias	United
41269	26 December 2012	Jeremy Farry	United

4.2 Category Hierarchy

| **Category Hierarchy** so that we can:

- **Drill down** from broader categories → to more detailed sub-categories → to products (if needed)
- Make reports and visuals more **interactive** and **organized**
- Analyse sales trends at multiple levels of granularity → e.g.:
 - Total Sales by **Category**
 - Sales by **Sub-Category** within each Category
 - Possibly even by **Product Name**



4.3 Dollar symbols added.

Adding dollar symbols to the column which have currency detailed values. For easy visualisation.

FilledSales_fromOrders	FilledQuantity_fromOrders	FilledProfit_fromOrders	Fill
\$16.77	2	\$1.47	
\$640.50	6	\$87.15	
\$397.92	4.43	\$35.52	
\$448.14	3.67	\$50.81	
\$34.68	6	\$16.99	
\$33.16	2.33	\$10.25	
\$5.16	2	\$1.34	
\$25.21	2.46	\$10.17	
\$46.84	2.5	\$13.17	
\$46.84	2.5	\$13.17	
\$182.72	8	\$84.05	
\$15.84	3	\$7.61	
\$69.75	3	\$532.10	
\$117.78	5.67	(\$14.09)	

5. Why These Steps Matter

- Row Removal improves data integrity by excluding invalid or duplicate entries.
- Date Cleaning and Standardization are crucial to avoid errors in date-based analysis.
- Text Capitalization avoids aggregation errors caused by inconsistent text cases.
- Cleaning Numeric Columns avoids calculation errors from invalid text or mixed data types.
- Unit Value Calculation accounts for the relationship between sales, quantity, profit, and discount.
- Missing Value Filling ensures no gaps in data, improving completeness and reliability.
- Decimal Rounding provides consistent and clean presentation of numeric data.
- Profit Percentage Calculation offers key business metrics for individual orders and overall product profitability.

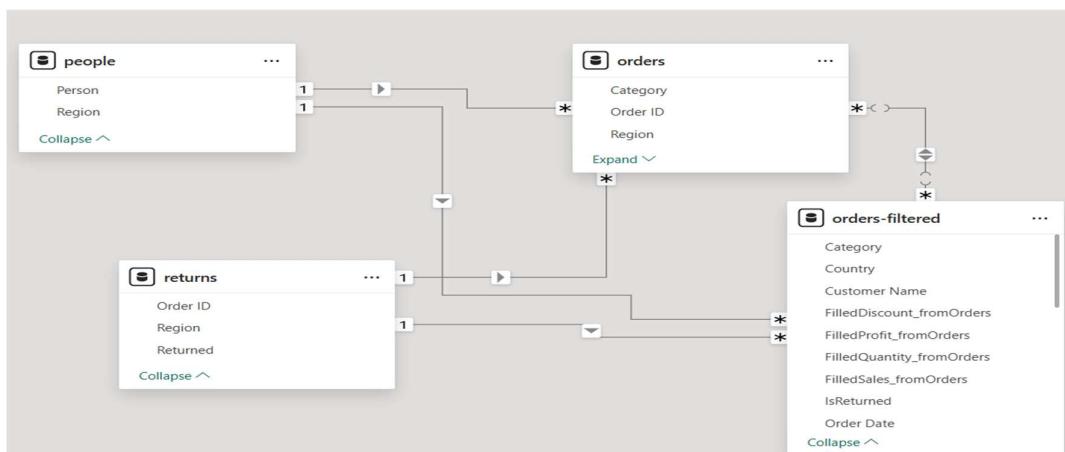
6. Final Notes

After Cleaning:

- Data ready for visualization
- Clean, consistent numeric columns
- Missing value logic applied
- Profit % metrics available for analysis
- Next step: Visualize insights!

7.1 Relationship between sheets.

In **Power BI**, relationships between tables are essential for data integration, filtering, and accurate analysis. An **active relationship** ensures that the correct connection is used when filtering or aggregating data. Only one relationship between tables can be active at a time, while inactive relationships must be activated using functions like **USERELATIONSHIP ()**.



7.2 Final cleaned sheets.

After so much filtration my sheets are completely ready for visualisation

- Cleaning ensures reliable analysis.
- Missing value filling enhances completeness.
- Profit % calculation supports business decisions.
- Structured workflow is reproducible for future projects.

8. Summary of Visuals Used in the Report.

- Implemented **Category Filters** using interactive **button slicers**.
- Displayed key business metrics (Total Orders, Total Profit, Total Returns, Total Sales) using **Card visuals**.
- Visualized **Sales by Category and Sub-Category** using a **Clustered Column Chart**.
- Analysed **overall return rates** through a **Pie Chart**.
- Showed **returned orders by Category** using a **Donut Chart**.
- Highlighted **profit contribution by Sub-Category** using a **Treemap**.
- Mapped **geographic sales distribution** using a **Map visual** (Country-wise Sales).
- Detailed **Sales by Category & Sub-Category** using a **Table visual** for clear reporting.
- Analysed **Profit by Person** using a **Clustered Bar Chart**.

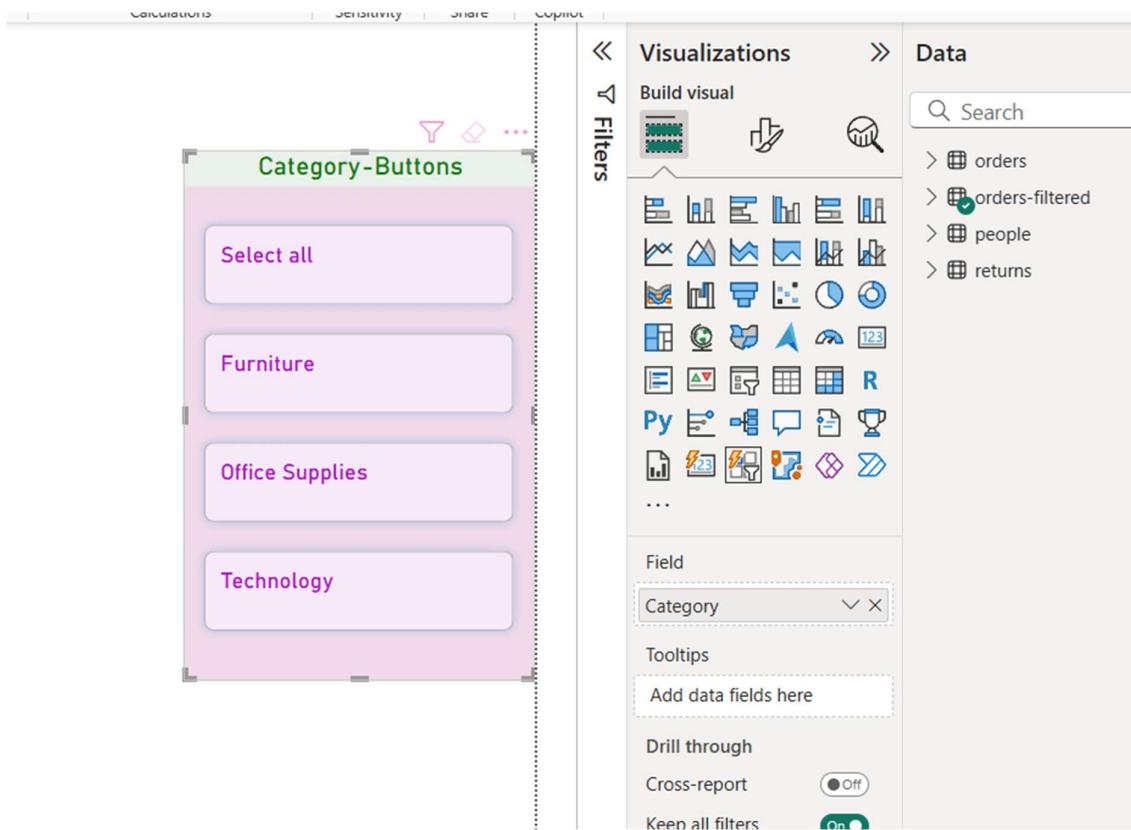
Together, these visuals create a comprehensive, interactive view of **Sales, Returns & Profit Analysis**, Profit helping to monitor performance and drive insights.

Category Buttons

Visual Type: Slicer (Button Style or Horizontal Orientation)

- 👉 This slicer allows quick filtering of data based on Category (Furniture, Office Supplies, Technology).
- 👉 Buttons are user-friendly for interactive filtering across the dashboard.

Screenshot: Category Buttons Slicer



Total Values KPIs (Total Orders, Total Profit, Total Returns, Total Sales)

Visual Type: 4 separate Card visuals

- 👉 These cards display the key KPIs of the business at a glance, with large fonts and clear values.
- 👉 Helps the user instantly understand overall order volume, profit, returns, and sales totals.

Screenshot: KPI Cards

The screenshot shows the Power BI visualization builder interface. On the left, there is a preview of four KPI cards with the following data:

- TotalOrders**: 15K
- TotalProfit**: 710.94K
- TotalReturns**: 1K
- TotalSales**: 6.12M

The central area is titled "Build visual" and contains a grid of visualization icons. Below this is a "Data" section where four measures are selected: TotalOrders, TotalProfit, TotalReturns, and TotalSales. The right side of the screen shows the "Data" pane with a search bar and a list of available fields, many of which have checkboxes next to them. Several checkboxes are checked, including TotalOrders, TotalProfit, TotalReturns, and TotalSales.

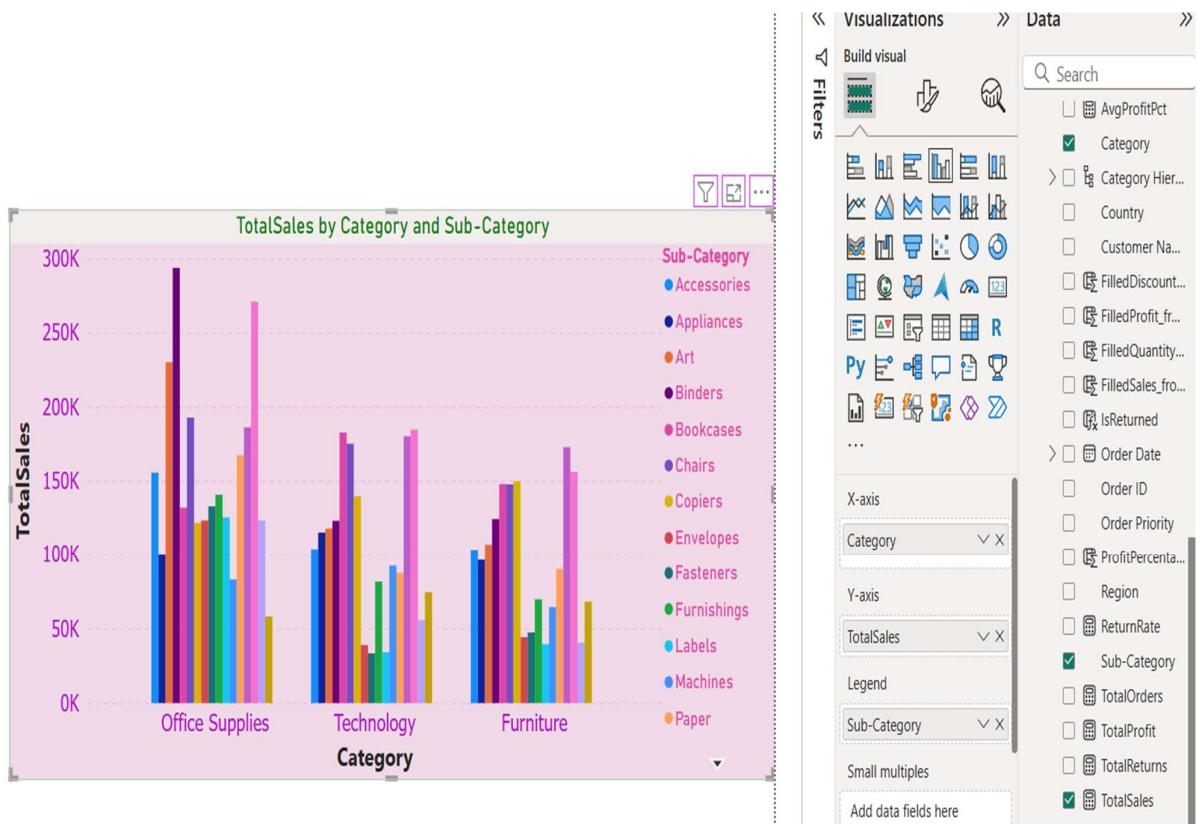
Category	Field	Status
AvgProfitPct		
Category		
Category Hier...		
Country		
Customer Na...		
FilledDiscount...		
FilledProfit_fr...		
FilledQuantity...		
FilledSales_fro...		
IsReturned		
Order Date		
Order ID		
Order Priority		
ProfitPercenta...		
Region		
ReturnRate		
Sub-Category		
TotalOrders		Selected
TotalProfit		Selected
TotalReturns		Selected
TotalSales		Selected

Total Sales by Category and Sub-Category

Visual Type: Clustered Column Chart

- 👉 This visual shows sales breakdown across Categories and their respective Sub-Categories.
- 👉 Useful for identifying which sub-categories drive revenue within each main category.

Screenshot: Clustered Column Chart - Sales by Category and Sub-Category

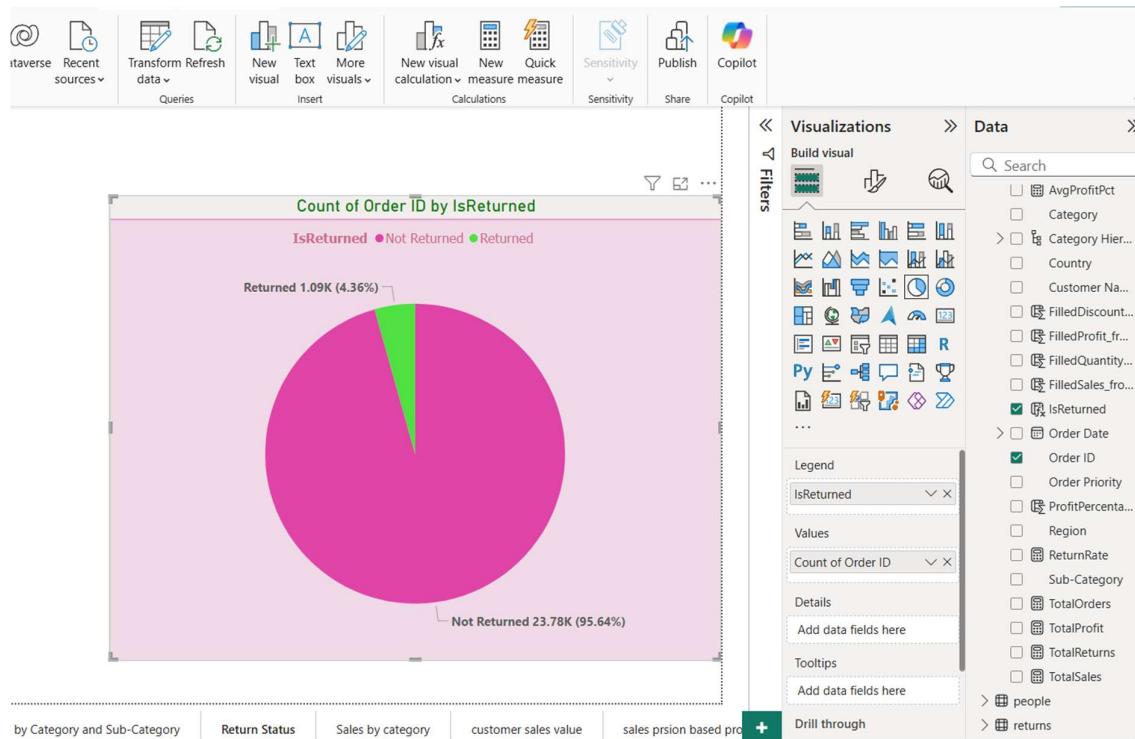


Count of Order ID by IsReturned

Visual Type: Pie Chart

- 👉 The pie chart shows the proportion of orders that were returned vs. not returned.
- 👉 Offers a quick understanding of return rates visually.

Screenshot: Pie Chart - Order Returns Count

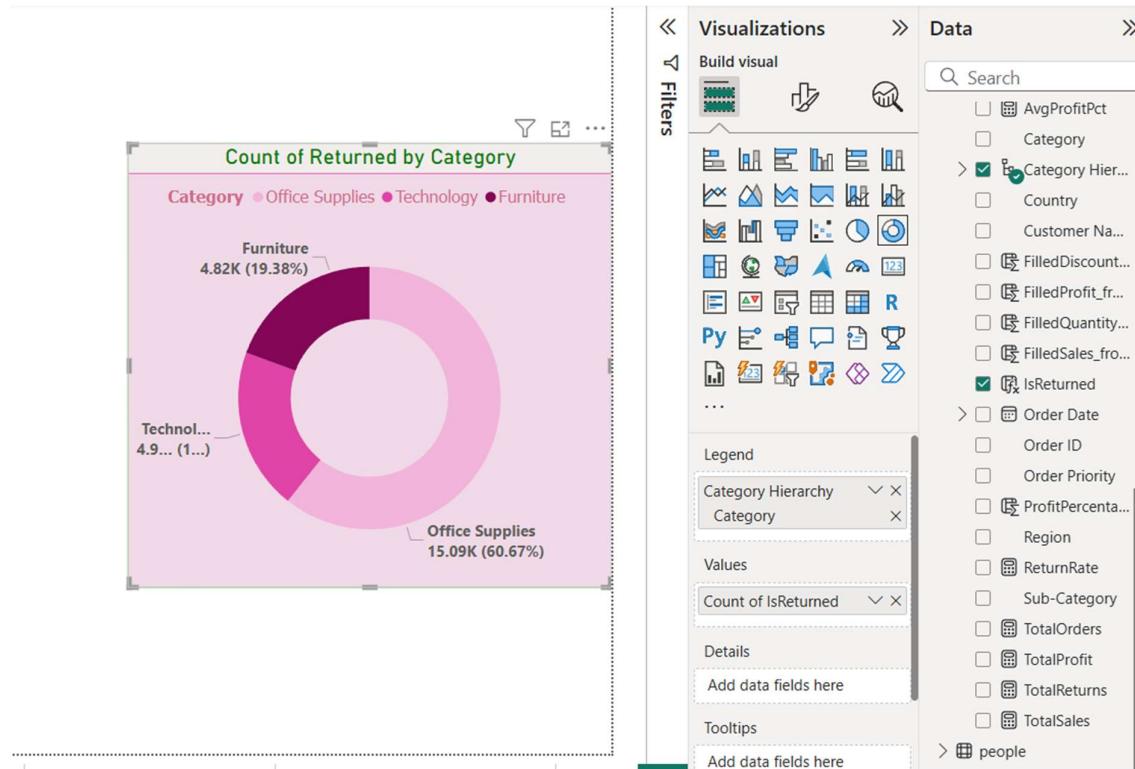


Count of Returned by Category

Visual Type: Donut Chart

- 👉 This chart displays how returned orders are distributed across different Categories.
- 👉 Helps analyse which product categories have higher return rates.

Screenshot: *Donut Chart - Returned by Category*

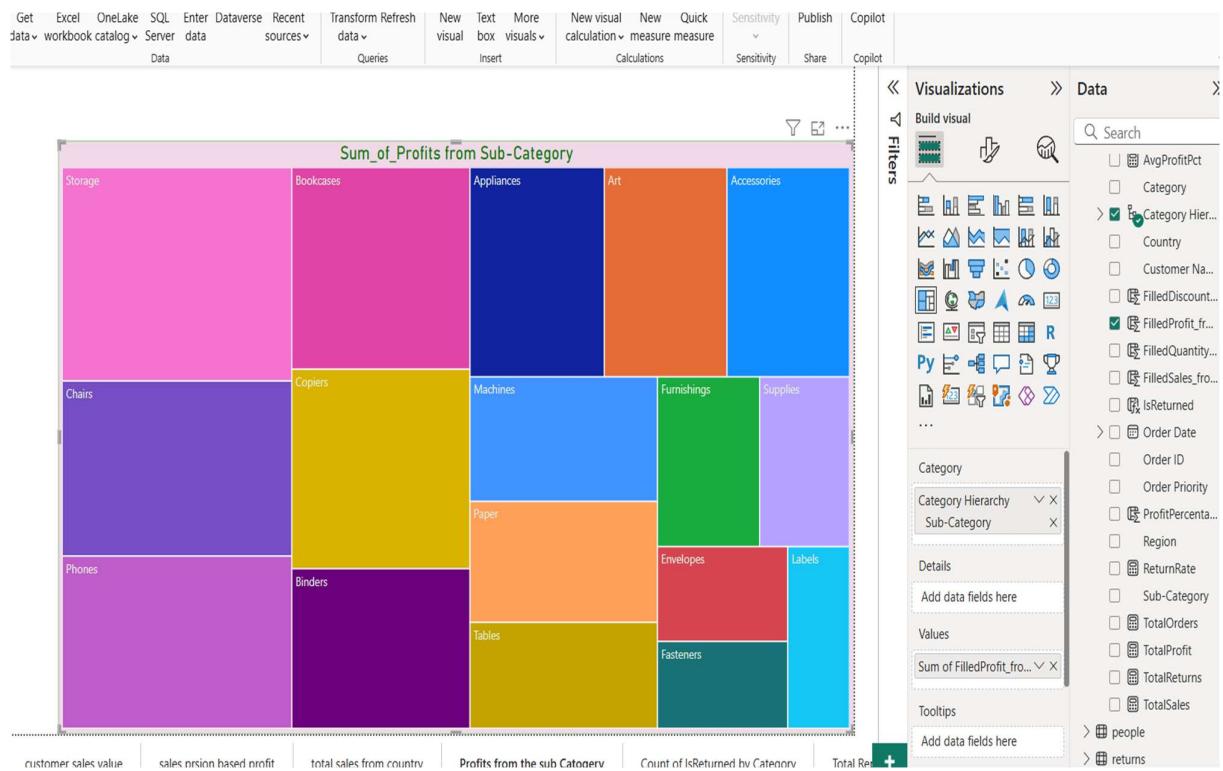


Sum of Profits from Sub-Category

Visual Type: Treemap

- 👉 The treemap visualizes profit contribution by Sub-Category, with larger areas for higher profit items.
- 👉 Great for spotting top and low profit segments visually.

Screenshot: Treemap - Profits by Sub-Category

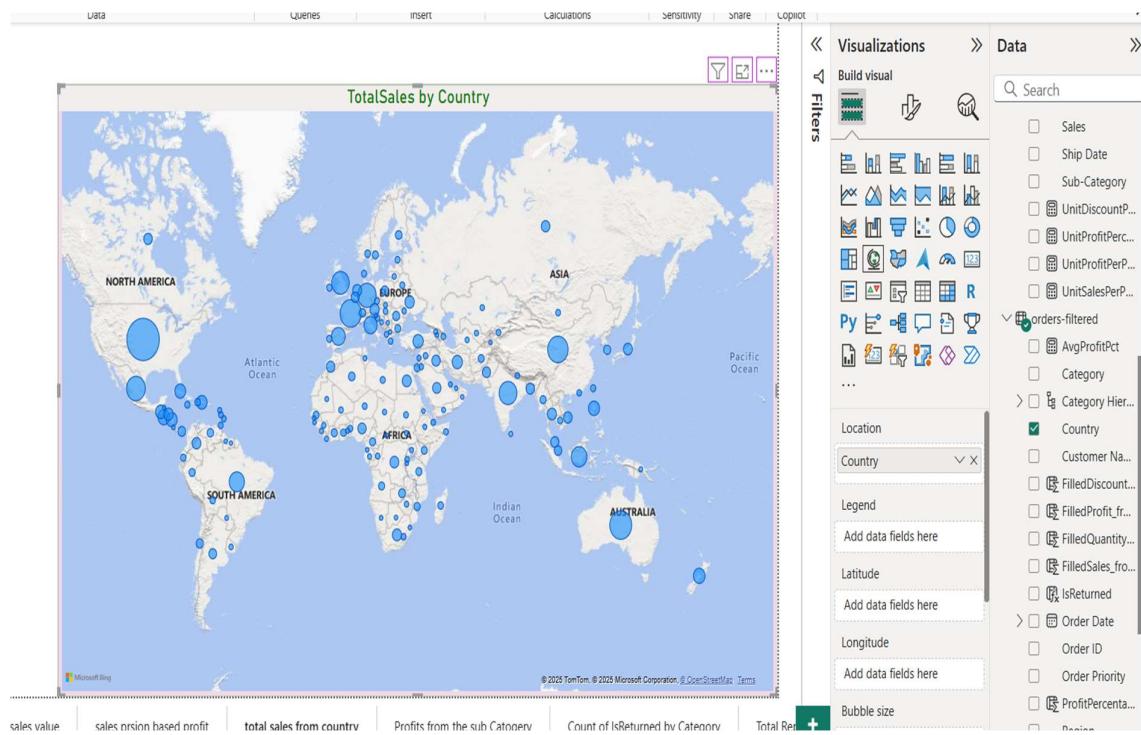


Total Sales by Country

Visual Type: Map Visual (Azure Map or Bing Map)

- 👉 This map displays sales across different geographic regions based on Country.
- 👉 Helps visualize market presence and regional performance globally.

Screenshot: Map Visual - Sales by Country



Sales by Category (with Sub-Category)

Visual Type: Table Visual

- 👉 The table provides detailed sales numbers for each Category and Sub-Category combination.
- 👉 Good for detailed reporting, export, and supporting data review.

Screenshot: Table Visual - Sales by Category & Sub-Category

The screenshot shows a Table Visual titled "Sales by category". The table has three columns: "Category", "Sub-Category", and "TotalSales". The data includes various categories like Office Supplies, Technology, Furniture, and their respective sub-categories and total sales values. A filter pane on the right side shows selected filters for "Category" and "Sub-Category".

Sales by category		
Category	Sub-Category	TotalSales
Office Supplies	Binders	2,93,819.30
Office Supplies	Storage	2,70,963.38
Office Supplies	Art	2,30,100.55
Office Supplies	Chairs	1,92,658.13
Office Supplies	Phones	1,86,068.52
Technology	Storage	1,84,487.29
Technology	Bookcases	1,82,620.45
Technology	Phones	1,80,056.62
Technology	Chairs	1,75,019.77
Furniture	Phones	1,72,726.48
Office Supplies	Paper	1,67,236.55
Furniture	Storage	1,55,962.39
Office Supplies	Accessories	1,55,488.17
Furniture	Copiers	1,49,677.23
Furniture	Bookcases	1,47,786.41
Furniture	Chairs	1,47,603.33
Office Supplies	Furnishings	1,40,566.89
Technology	Copiers	1,39,539.94
Office Supplies	Fasteners	1,32,692.19
Office Supplies	Bookcases	1,31,827.61
Total		61,19,517.95

Filters

Build visual

Search

Category

Sub-Category

Columns

Category

Sub-Category

TotalSales

Drill through

Cross-report

Keep all filters

Add drill-through fields here

AvgProfitPct

Category Hier..

Country

Customer Na...

FilledDiscount...

FilledProfit.fr...

FilledQuantity..

FilledSales_fro...

IsReturned

Order Date

Order ID

Order Priority

ProfitPercenta...

Region

ReturnRate

TotalOrders

TotalProfit

TotalReturns

TotalSales

people

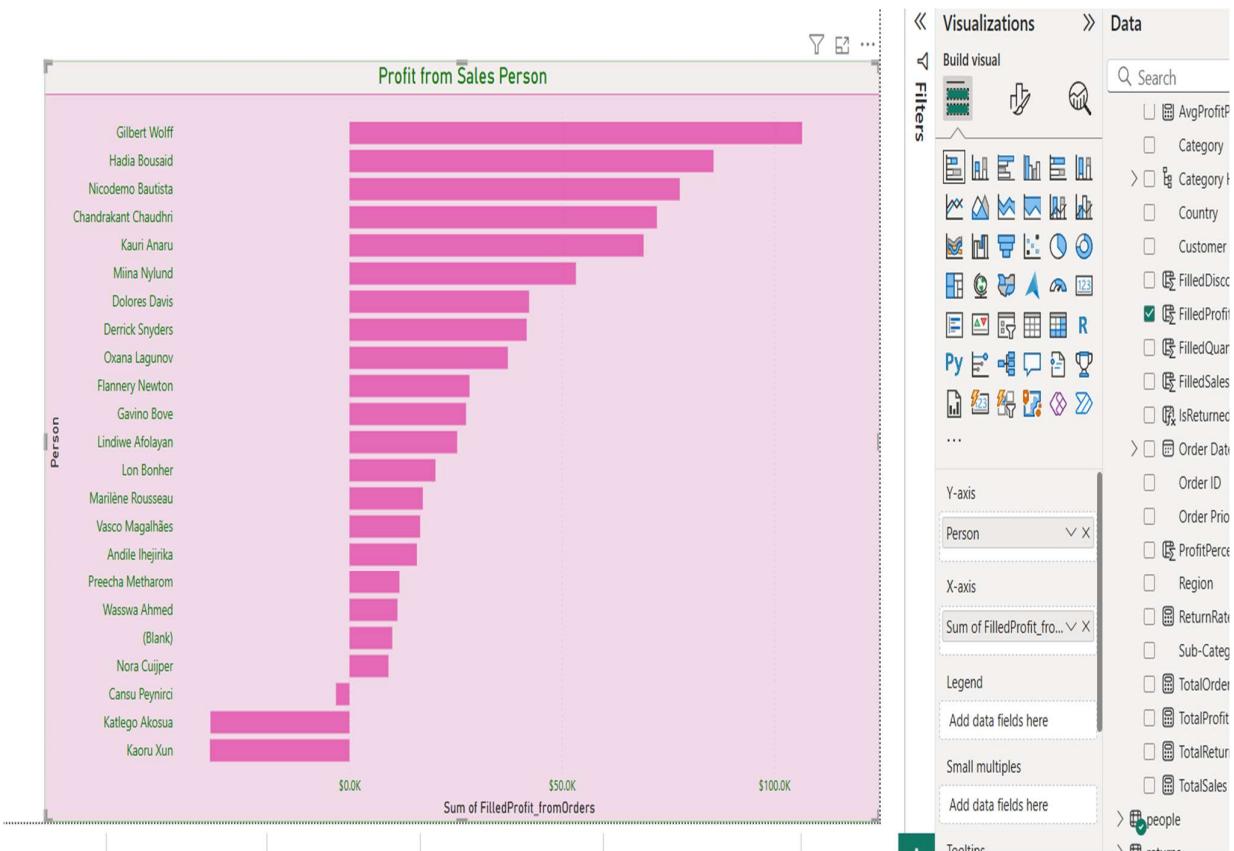
returns

Profit from Person

Visual Type: Clustered Bar Chart

- 👉 This bar chart shows how much profit each Person (likely salesperson or region head) generated.
- 👉 Useful for evaluating individual or regional performance.

Screenshot: Clustered Bar Chart - Profit from Person



Conclusion

- **✓ Comprehensive Dashboard:** The Sales & Returns Analysis Dashboard provides an interactive view of sales, profits, and returns across categories, sub-categories, countries, and individual contributors.
- **✓ Rich Visual Insights:** Utilized a wide range of Power BI visuals — KPI cards, clustered column charts, pie and donut charts, treemaps, geographic maps, tables, and slicers — to deliver actionable business insights.
- **✓ Systematic Data Preparation:** Followed a structured workflow including row removal, date standardization, numeric cleaning, missing value handling, and profit percentage calculation to ensure data accuracy and reliability.
- **✓ Interactive User Experience:** Enhanced analysis through category filters, hierarchy drill-downs, and responsive slicers for improved stakeholder engagement.
- **✓ Business Impact:** Enabled stakeholders to:
 - Understand high-level KPIs and trends
 - Analyse product and category performance
 - Monitor returns and profitability
 - Assess regional and individual contributions
- **✓ Scalable Workflow:** The methodology applied is scalable and adaptable for future projects, ensuring ongoing support for continuous business improvement and data-driven strategic planning.

Screenshot: Final Report

