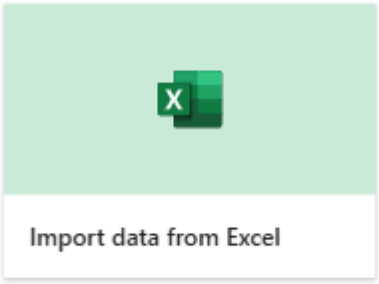
Hasti Gohel (PDS25_06009)

Section 1: Data Loading and Initial Cleanup

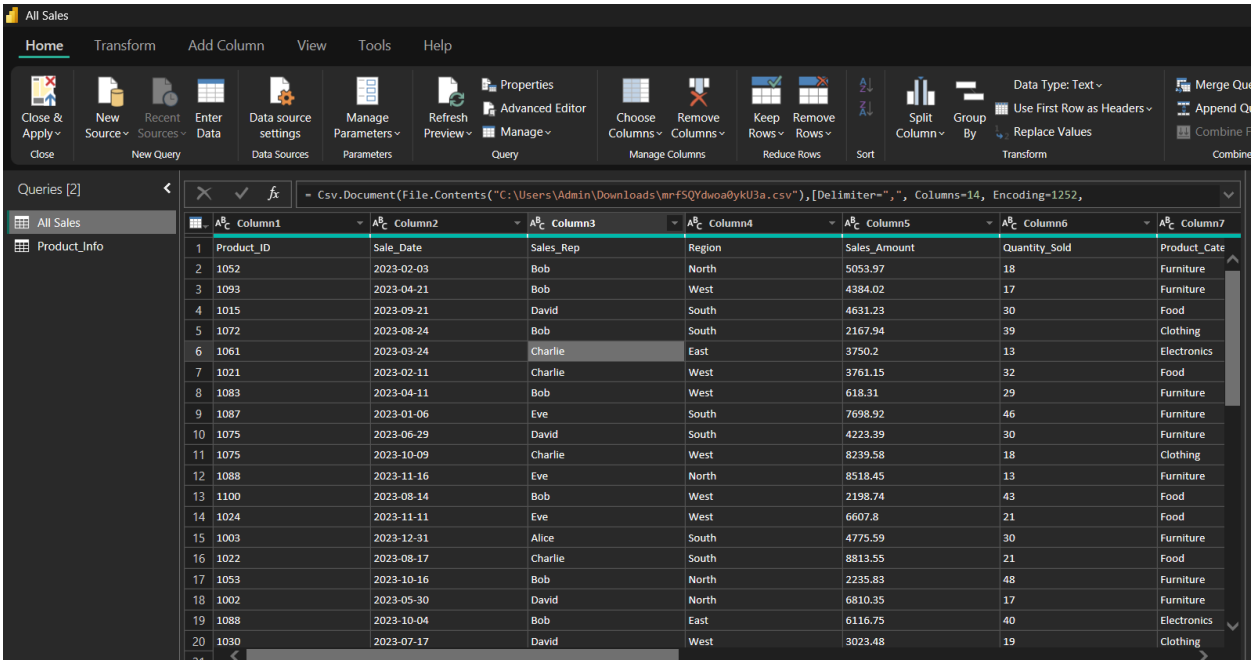
1. **Load the dataset into Power BI.**
   - **Import the dataset into Power BI Desktop using "Get Data".**

The Excel file was imported using the "Get Data" option in Power BI Desktop. This made the dataset available for transformation and modeling



2. **Open Power Query Editor.**
   - **From the "Home" tab, click "Transform Data" to enter the Power Query Editor.**

By selecting "Transform Data" from the Home tab, the dataset was opened in Power Query Editor, where all data cleaning and shaping operations were performed.

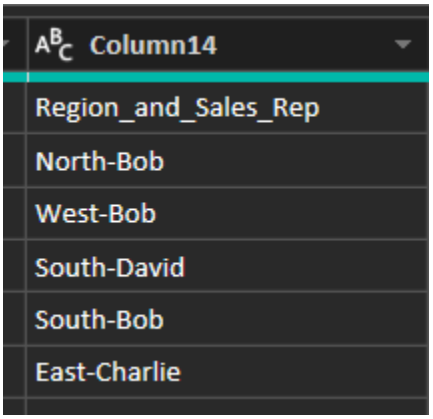


	Column1	Column2	Column3	Column4	Column5	Column6	Column7
1	Product_ID	Sale_Date	Sales_Rep	Region	Sales_Amount	Quantity_Sold	Product_Cate
2	1052	2023-02-03	Bob	North	5053.97	18	Furniture
3	1093	2023-04-21	Bob	West	4384.02	17	Furniture
4	1015	2023-09-21	David	South	4631.23	30	Food
5	1072	2023-08-24	Bob	South	2167.94	39	Clothing
6	1061	2023-03-24	Charlie	East	3750.2	13	Electronics
7	1021	2023-02-11	Charlie	West	3761.15	32	Food
8	1083	2023-04-11	Bob	West	618.31	29	Furniture
9	1087	2023-01-06	Eve	South	7698.92	46	Furniture
10	1075	2023-06-29	David	South	4223.39	30	Furniture
11	1075	2023-10-09	Charlie	West	8239.58	18	Clothing
12	1088	2023-11-16	Eve	North	8518.45	13	Furniture
13	1100	2023-08-14	Bob	West	2198.74	43	Food
14	1024	2023-11-11	Eve	West	6607.8	21	Food
15	1003	2023-12-31	Alice	South	4775.59	30	Furniture
16	1022	2023-08-17	Charlie	South	8813.55	21	Food
17	1053	2023-10-16	Bob	North	2235.83	48	Furniture
18	1002	2023-05-30	David	North	6810.35	17	Furniture
19	1088	2023-10-04	Bob	East	6116.75	40	Electronics
20	1030	2023-07-17	David	West	3023.48	19	Clothing

Section 2: Data Cleaning Tasks

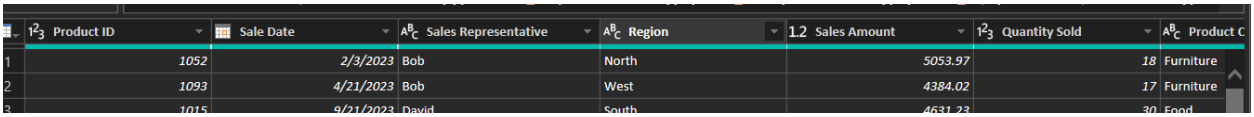
3. **Remove unnecessary columns.**
    - **Remove the Region_and_Sales_Rep column.**

The column Region_and_Sales_Rep was removed because it was redundant or irrelevant to the analysis goals.



4. **Rename columns for better readability.**
    - **Rename the following:**
        - **Product_ID → Product ID**
        - **Sale_Date → Sale Date**
        - **Sales_Rep → Sales Representative**
        - **Product_Category → Product Category**
        - **Payment_Method → Payment Method**
        - **Sales_Channel → Sales Channel**
        - **Customer_Type → Customer Type**

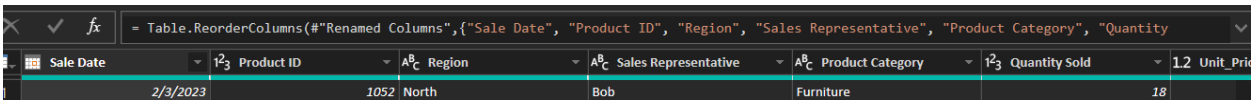
Several columns were renamed to be more readable and user-friendly, aligning with reporting standards and easier interpretation. For example, Sale_Date became Sale Date, and Sales_Rep became Sales Representative.



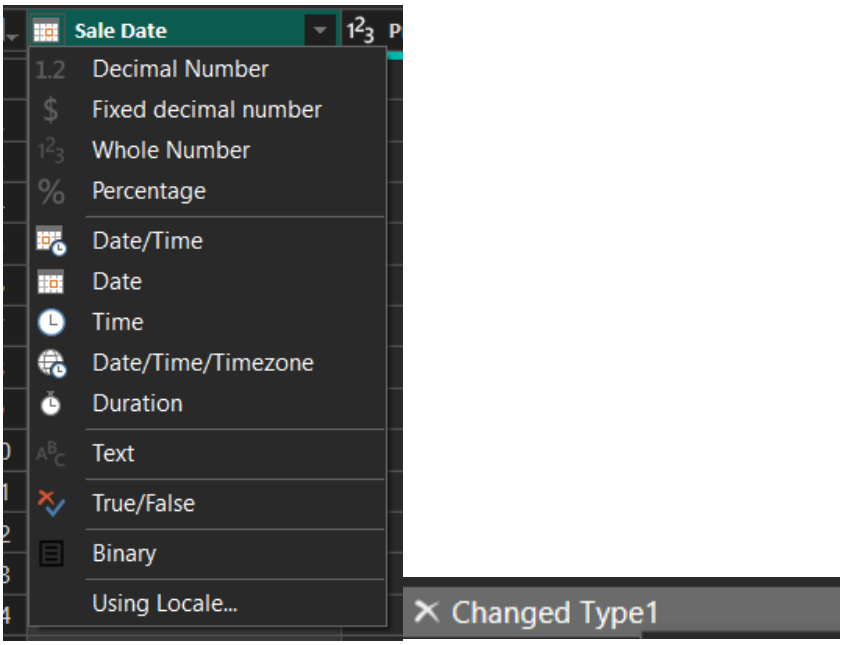
	Product ID	Sale Date	Sales Representative	Region	Sales Amount	Quantity Sold	Product C
1	1052	2/3/2023	Bob	North	5053.97	18	Furniture
2	1093	4/21/2023	Bob	West	4384.02	17	Furniture
3	1015	9/21/2023	David	South	4631.23	30	Food

5. **Change column order to improve logical flow.**
    - **Reorder the columns to:**
        - **Sale Date, Product ID, Region, Sales Representative, Product Category, Quantity Sold, Unit Price, Unit Cost, Sales Amount, Discount, Customer Type, Payment Method, Sales Channel**

Columns were reordered to follow a logical sequence, starting with Sale Date and ending with Sales Channel, improving navigation and consistency during analysis.



6. **Correct data types.**
   - **Ensure appropriate data types:**
     - **Sale Date: Date**
     - **Quantity Sold: Whole Number**
     - **Sales Amount, Unit Cost, Unit Price, Discount: Decimal Number**
     - **Categorical columns: Text**

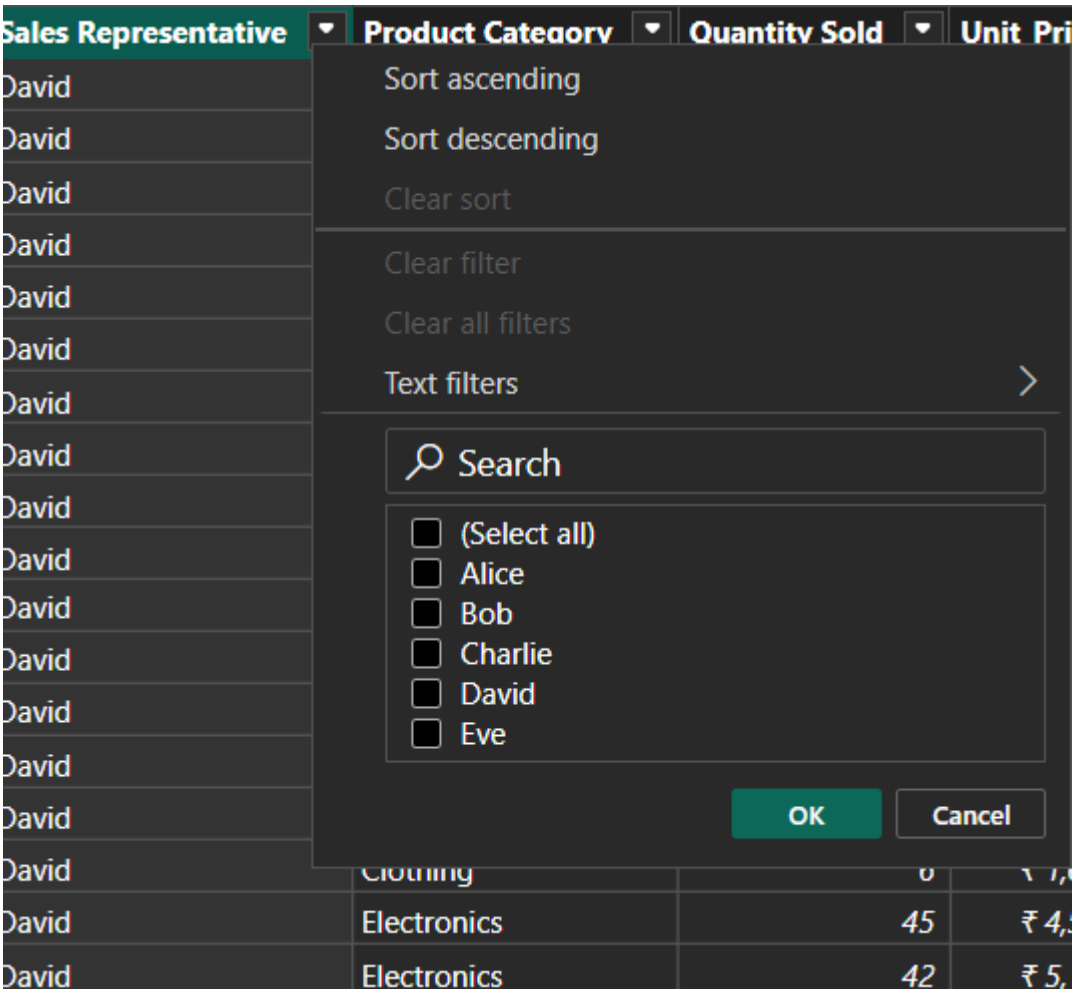


7. **Format currency fields.**
   - **Format Sales Amount, Unit Price, Unit Cost as currency.**

Unit_Price	Unit_Cost	Sales Amount
₹ 3,007.47	₹ 2,565.30	₹ 5,238.42
₹ 4,186.98	₹ 3,851.45	₹ 7,751.92
₹ 4,061.04	₹ 3,741.30	₹ 7,724.57
₹ 4,361.70	₹ 4,138.41	₹ 8,090.84
₹ 2,147.14	₹ 1,833.95	₹ 9,631.41
₹ 3,108.30	₹ 2,873.53	₹ 6,773.89
₹ 1,955.56	₹ 1,531.20	₹ 7,952.11
₹ 4,784.28	₹ 4,646.55	₹ 5,842.88
₹ 1,709.71	₹ 1,281.65	₹ 6,966.82
₹ 1,482.88	₹ 1,434.20	₹ 6,581.04

8. **Remove rows with null or blank Sales Representative values.**
    - **Filter and remove any rows where the Sales Representative field is empty or null.**

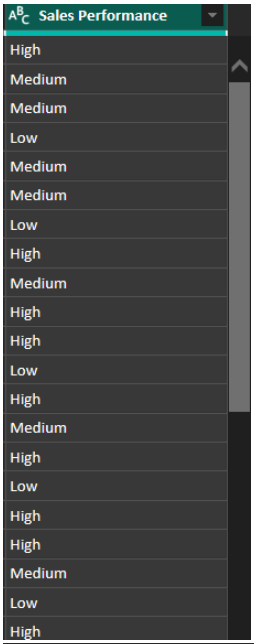
Here we don’t have any null value so I’m not able to display it.

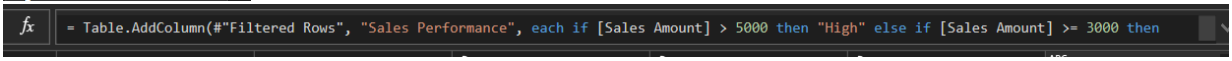


Section 3: Data Transformation

9. Add a new conditional column.

- **Create a column Sales Performance:**
  - **"High" if Sales Amount > 5000**
  - **"Medium" if between 3000–5000**
  - **"Low" otherwise**





10. Create a calculated column for Total Profit.

- **Total Profit = (Unit Price - Unit Cost) * Quantity Sold**

```
Total Profit = ([Unit_Price] - [Unit_Cost]) * [Quantity Sold]
```

11. Create a column for Final Sale Price after discount.

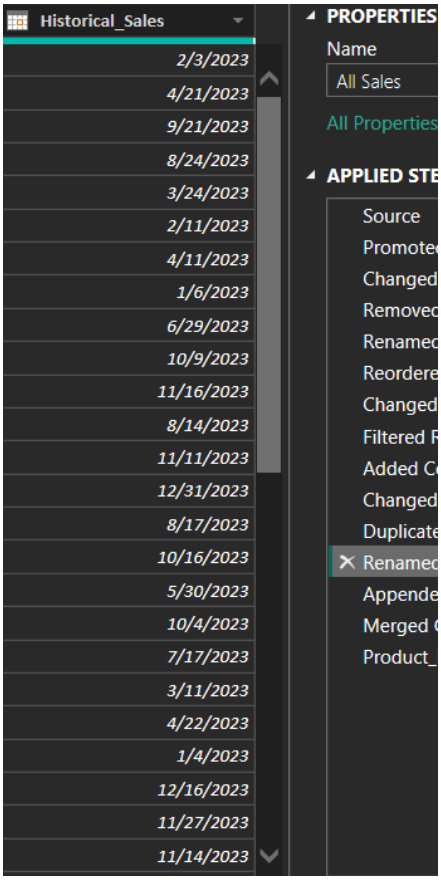
- **Formula: Final Sale Price = Sales Amount * (1 - Discount)**

```
Final Sale Price = [Sales Amount] * (1 - [Discount])
```

Section 4: Appending & Merging

12. Duplicate the current query and name it Historical_Sales.

- **Filter for records with Sale Date before 01-Jan-2023.**



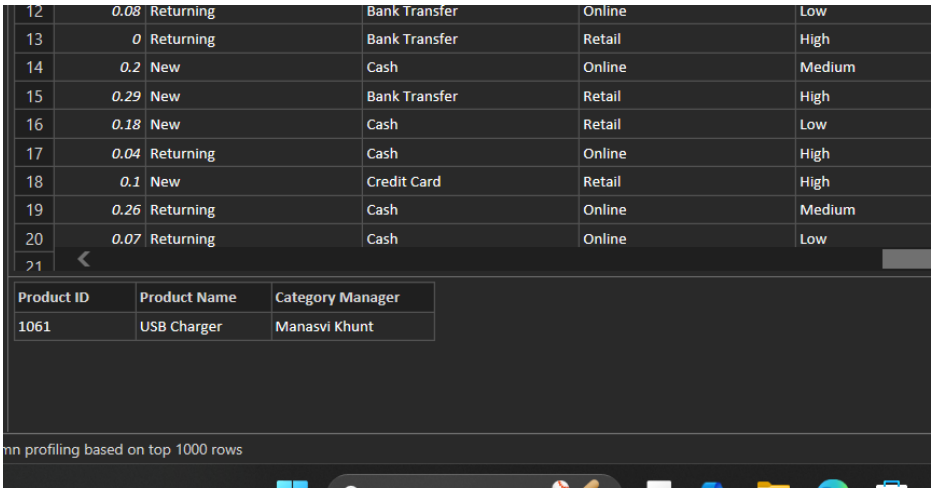
13. Append Historical_Sales to the original table.

- **Create a new table All_Sales by appending both tables.**

14. Merge external data.

- **Create or simulate a Product_Info table:**
- **Columns: Product ID, Product Name, Category Manager**
- **Merge it with All_Sales on Product ID and expand the merged columns.**

Product ID	Product Name	Category Manager
1015	Smartwatch	Hasti Gohel
1061	USB Charger	Manasvi Khunt
1021	Laptop Sleeve	Unnati Pandya
1002	Wireless Mouse	Aditi Shah
1038	Mouse	Pathik Patel
P6	KeyBoard	Ankit Chotaliya



12	0.08	Returning	Bank Transfer	Online	Low
13	0	Returning	Bank Transfer	Retail	High
14	0.2	New	Cash	Online	Medium
15	0.29	New	Bank Transfer	Retail	High
16	0.18	New	Cash	Retail	Low
17	0.04	Returning	Cash	Online	High
18	0.1	New	Credit Card	Retail	High
19	0.26	Returning	Cash	Online	Medium
20	0.07	Returning	Cash	Online	Low
21					

Product ID	Product Name	Category Manager
1061	USB Charger	Manasvi Khunt

Section 5: Final Steps

15. Apply all transformations.

- **Click “Close & Apply” to load cleaned and transformed data back into Power BI for report building.**