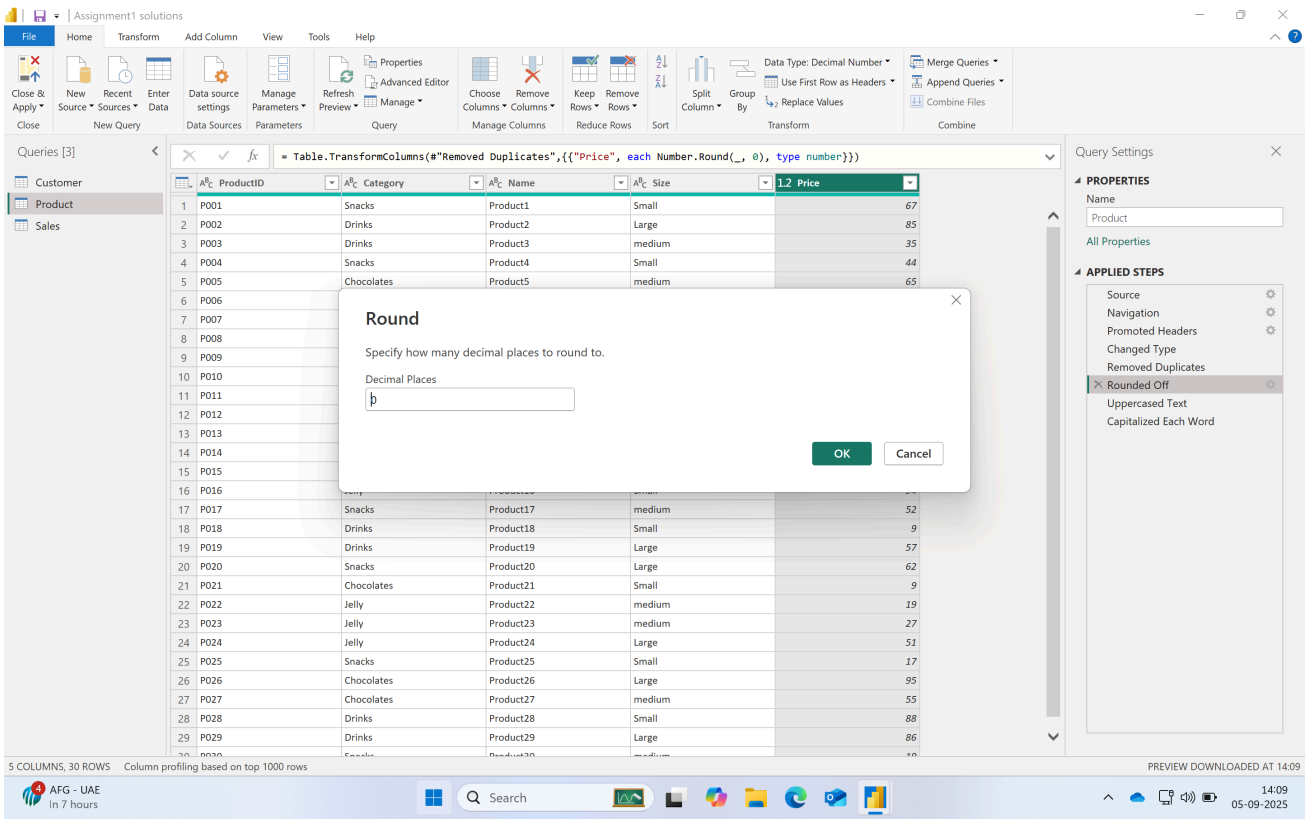
Internshala Trainings

Data Science

PowerBI Assignment 1

Prepared By,
Utkarsh Anand

Task 1 — Data Cleaning



Part 1: Round the ‘Price’ column (Product) to the nearest integer

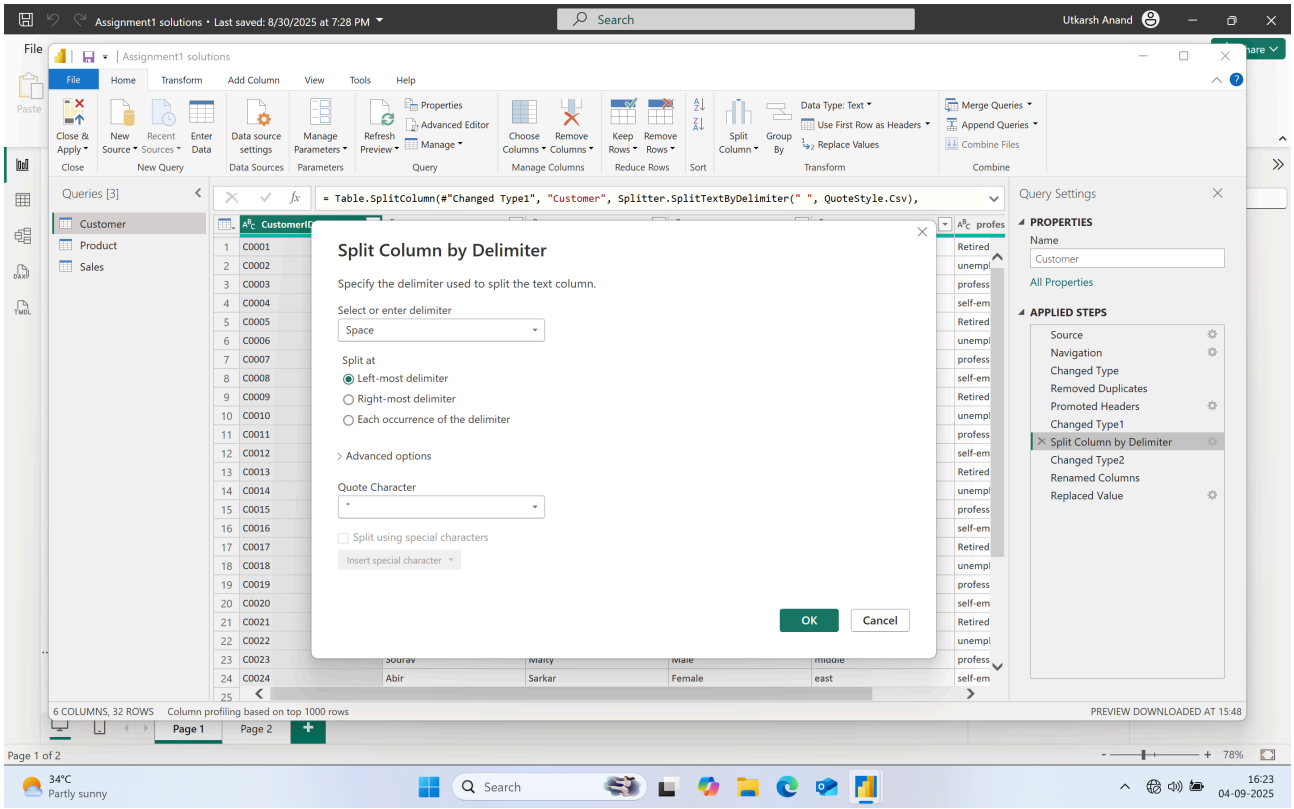
Steps taken to round the ‘Price’ column in the Product dataset to the nearest integer for simplicity:

1. In Power BI Desktop, on the Home ribbon, click Transform data to open Power Query Editor.
2. In the Queries pane (left), click the Product table.
3. Click the Price column header to select the entire column.
4. Go to the Transform tab (top) → in the Number Column group, click Rounding ▼ → choose Round.
5. If a gear icon appears next to the new step in Applied Steps (right), click it and set Decimal places = 0, then OK (this guarantees whole numbers).

6. Confirm that the Price values now display as whole numbers (no decimals).

7. Click Home (in Power Query) → Close & Apply to save changes back to the model.

Part 2: Split the ‘Customer’ column (Customer) into ‘FirstName’ and ‘LastName’



Steps taken to split the ‘Customer’ column in the Customer table into two columns — ‘FirstName’ and ‘LastName’:

1. On the Home ribbon in Power BI Desktop, click Transform data to open Power Query Editor (if not already open).

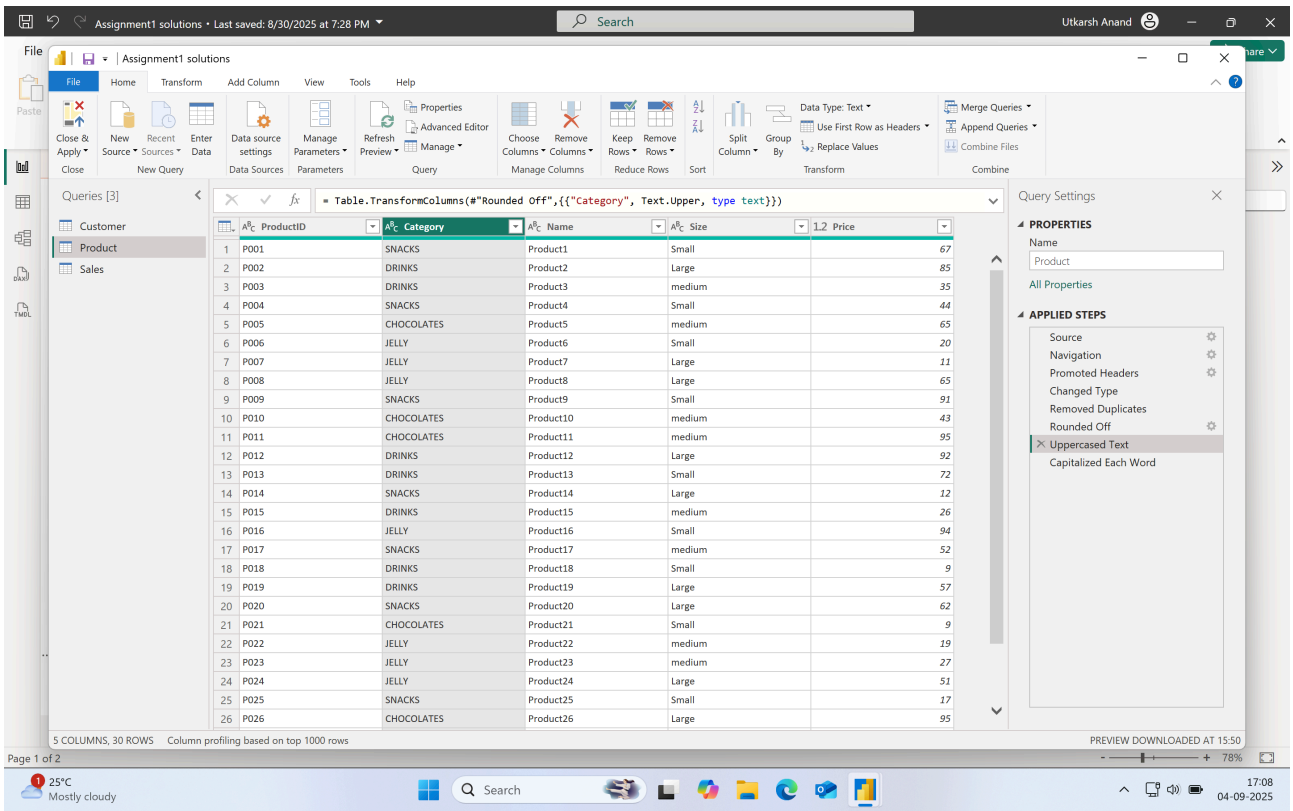
2. In Queries (left), click the Customer table.

3. Click the Customer column header.

4. Go to Home tab → Split Column ▼ → By Delimiter…

5. In the dialog:
   - Select or enter delimiter: choose Space.
   - Open Advanced options → set Split at: Left-most delimiter (creates only two parts).
   - Click OK.
6. Rename the resulting columns: double-click the headers and set them to FirstName and LastName.
7. Click Home → Close & Apply.

Task 2 — Standardization

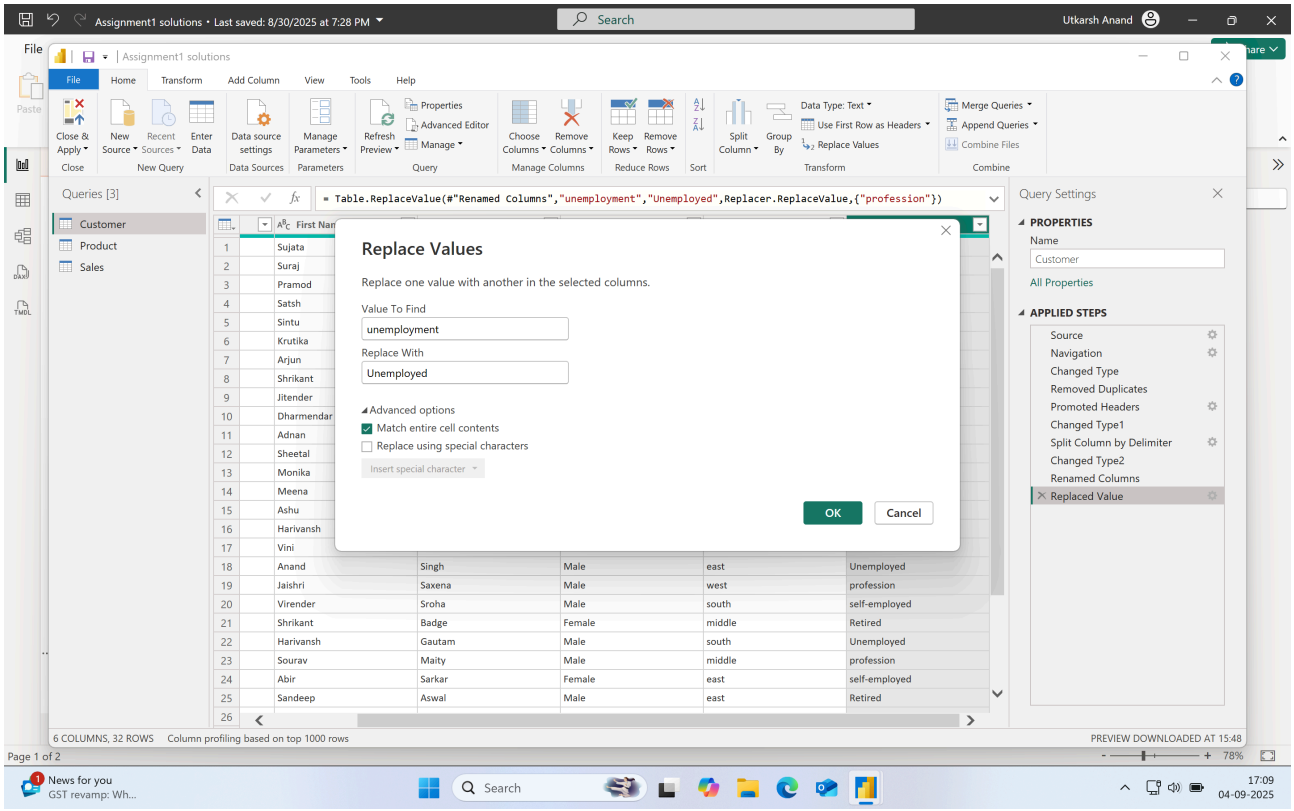


Part 1: Convert ‘Category’ (Product) to UPPERCASE

Steps taken to convert all entries in the ‘Category’ column in the Product table to uppercase:

1. Open Power Query Editor (Home → Transform data).
2. Select the Product table → click the Category column header.
3. Go to Transform tab → Format ▼ → click UPPERCASE.
4. Verify that each Category value is now fully capitalized.
5. Click Home → Close & Apply.

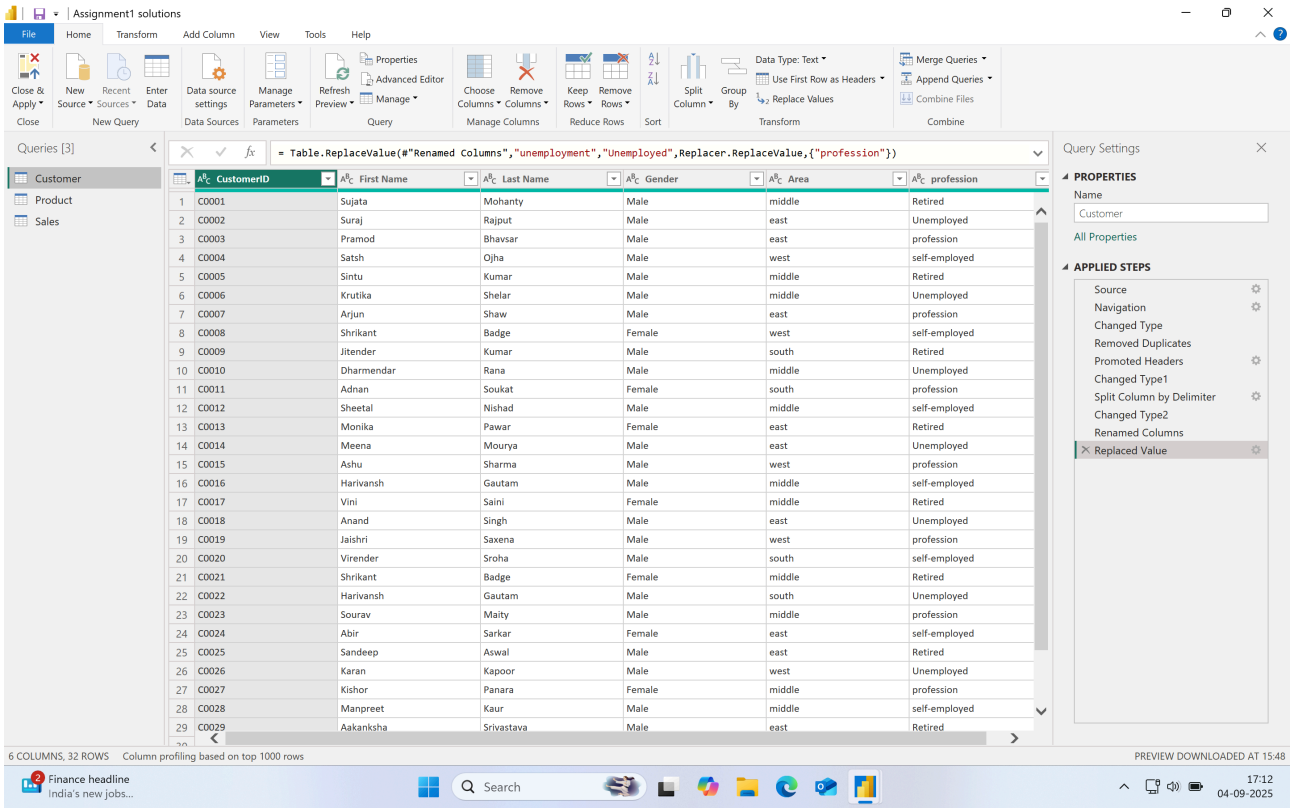
Part 2: Replace ‘unemployment’ with ‘Unemployed’ in ‘Profession’ (Customer)



Steps taken to replace all occurrences of ‘unemployment’ with ‘Unemployed’ in the ‘Profession’ column of the Customer table:

1. In Power Query Editor, select the Customer table.
2. Click the Profession column header.
3. Go to Home tab → click Replace Values.
4. In the dialog:
   - Value To Find: unemployment
   - Replace With: Unemployed
   - Click Advanced options and tick Match entire cell contents (ensures only the exact word is replaced).
   - Click OK.
5. Review a few rows to confirm the change, then Close & Apply.

Task 3 — Data Types & Consistency

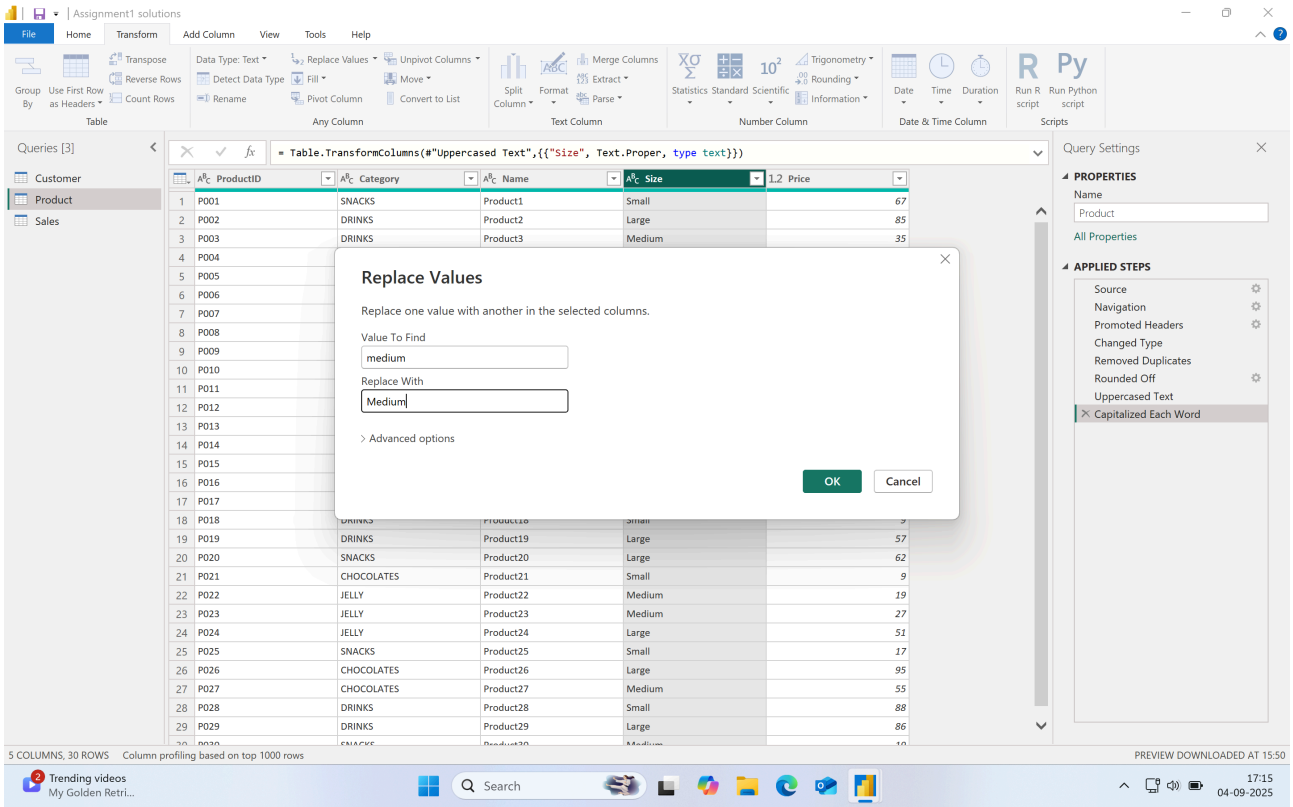


Part 1: Ensure appropriate data types (Date, Price, etc.)

Steps taken to ensure all columns in the datasets have appropriate data types (e.g., ‘Date’ as Date, ‘Price’ as Decimal/Whole Number):

1. Open Power Query Editor.
2. Sales table:
   - Click the Date column → on Home tab (Power Query), open the Data Type dropdown (icon left of column name or on ribbon) → choose Date → Replace current.
3. Product table:
   - Click Price → set Data Type to Whole Number (since we rounded to integers) or Decimal Number (if you kept decimals).
4. Repeat for any other columns (IDs as Whole Number/Text as appropriate).
5. Confirm the ABC/123/calendar icons reflect the correct types, then Close & Apply.

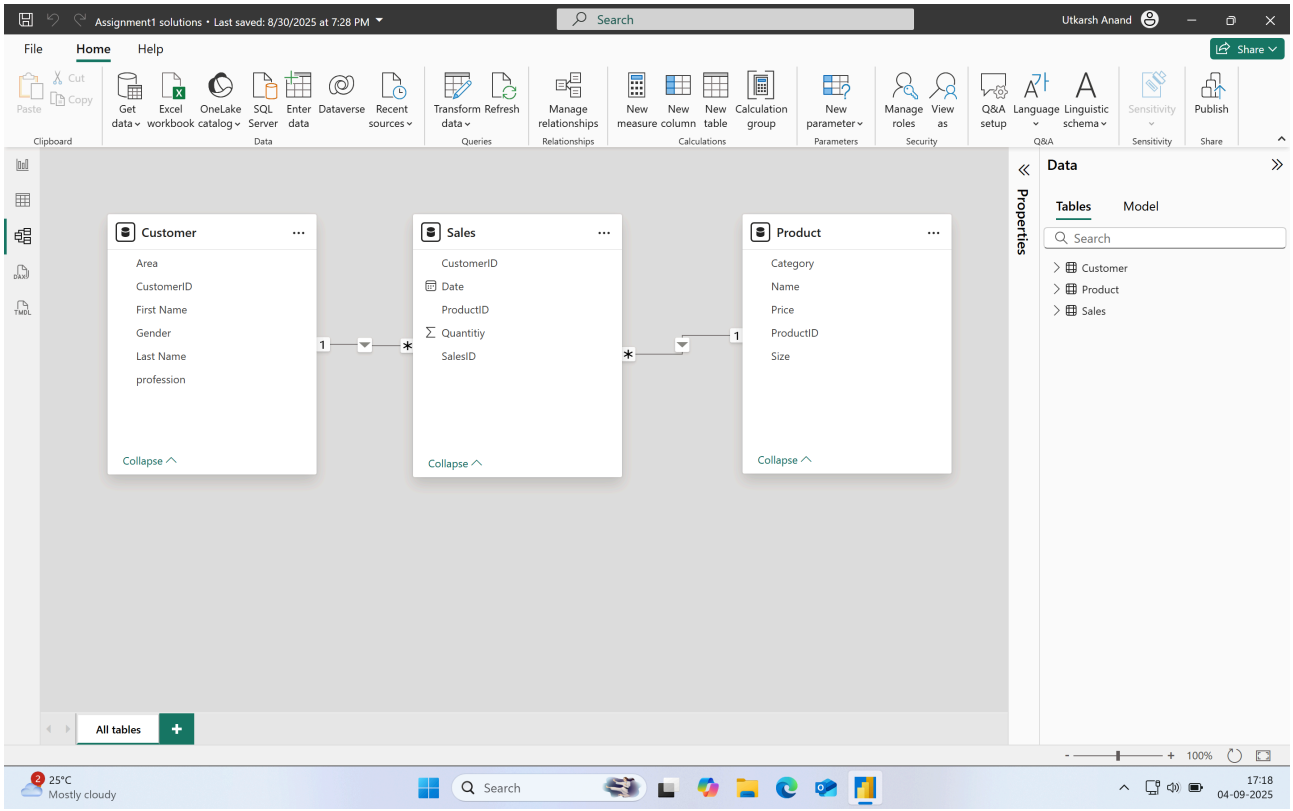
Part 2: Standardize ‘Size’ (Product) — fix inconsistent values



Steps taken to identify and replace inconsistent values in the ‘Size’ column of the Product dataset to ensure uniformity:

1. In Power Query Editor, select the Product table → click the Size column.
2. Quick normalization: Transform tab → Format ▼ → Capitalize Each Word (converts “medium”, “SMALL” to “Medium”, “Small”).
3. Targeted cleanup (if needed): Home → Replace Values to convert any remaining variants (e.g., Med → Medium, Lrg → Large).
4. Review the Value Distribution (column header profiling if enabled) to ensure only the intended forms remain.
5. Close & Apply.

Task 4 — Modeling & Deduplication

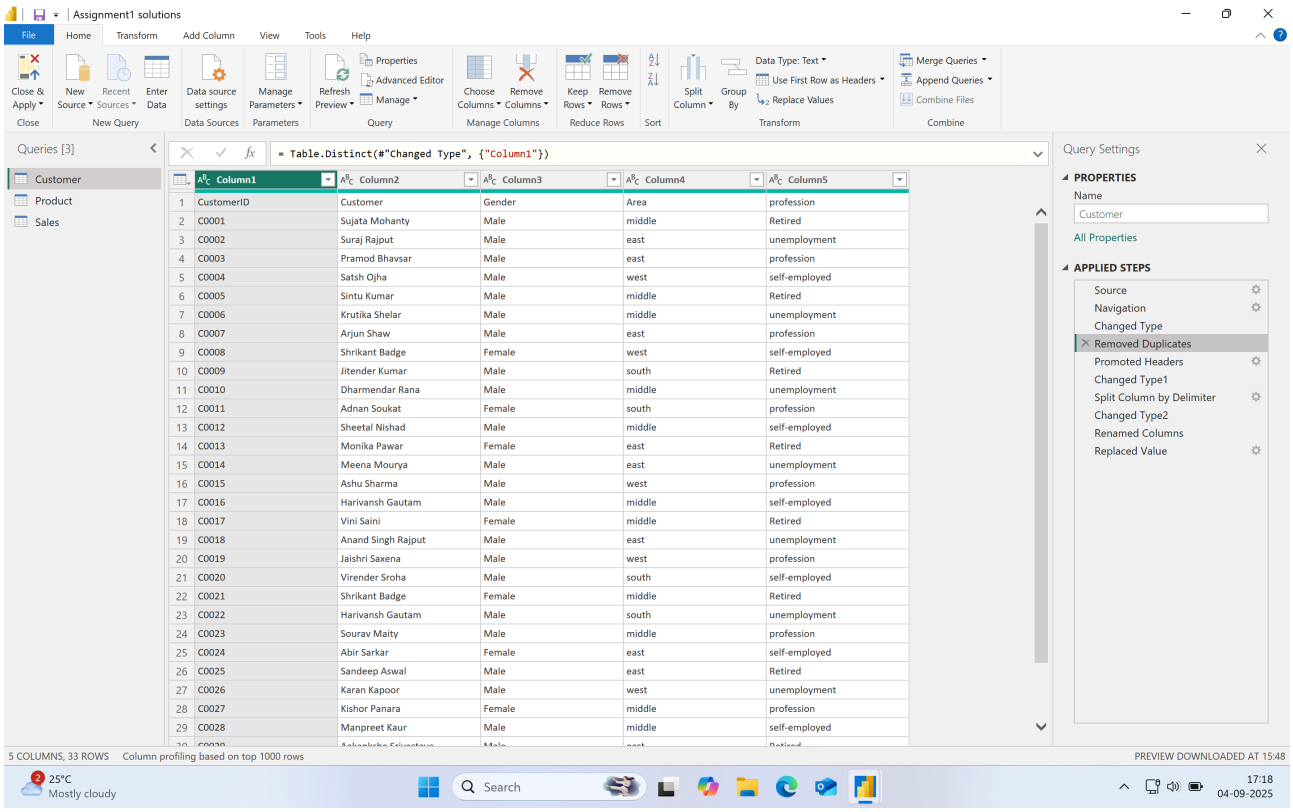


Part 1: Create relationships via ‘CustomerID’ and ‘ProductID’

Steps taken to create relationships between the tables using ‘CustomerID’ and ‘ProductID’ as keys:

1. In Power BI Desktop, click the Model view icon (left sidebar diagram).
2. Drag Customer[CustomerID] onto Sales[CustomerID] → confirm Cardinality: One to many (1:*), Cross filter: Single, Make this relationship active: On → OK.
3. Drag Product[ProductID] onto Sales[ProductID] → same settings as above → OK.
4. Verify both relationships show **1 → *** from Customer/Product to Sales.

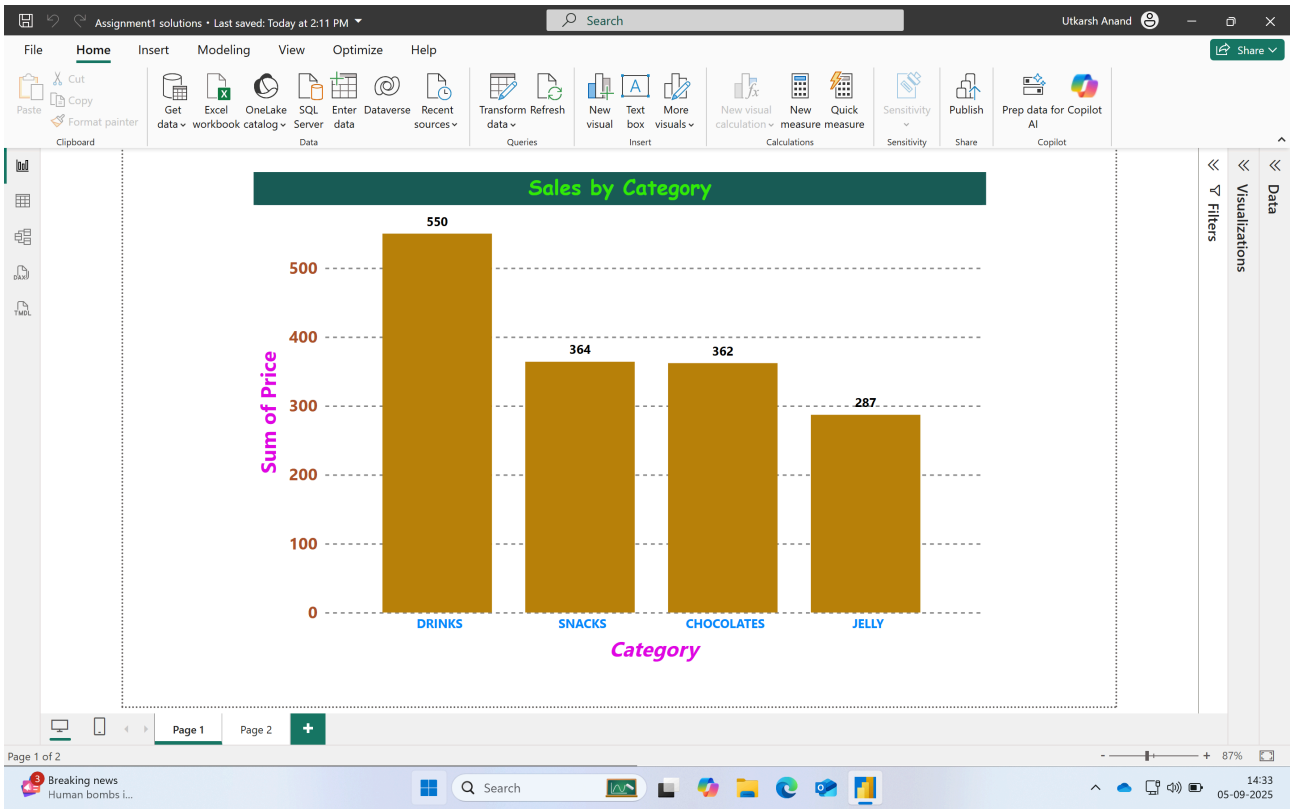
Part 2: Remove duplicates in Customer and Product



Steps taken to clean data by removing any duplicate entries in the Customer and Product tables:

1. Open Power Query Editor.
2. Customer table:
   - Select the CustomerID column (or select multiple columns if the uniqueness rule demands).
   - Home tab → Remove Rows ▼ → Remove Duplicates.
3. Product table:
   - Select ProductID → Home → Remove Rows ▼ → Remove Duplicates.
4. Confirm the Applied Steps reflect the duplicate removal, then Close & Apply.

Task 5 — Visualizations



Part 1: Sales by Category (Total Sales / TotalPrice)

Steps taken to create and format a chart showing total sales (TotalPrice) by product category, with colors, title, and data labels:

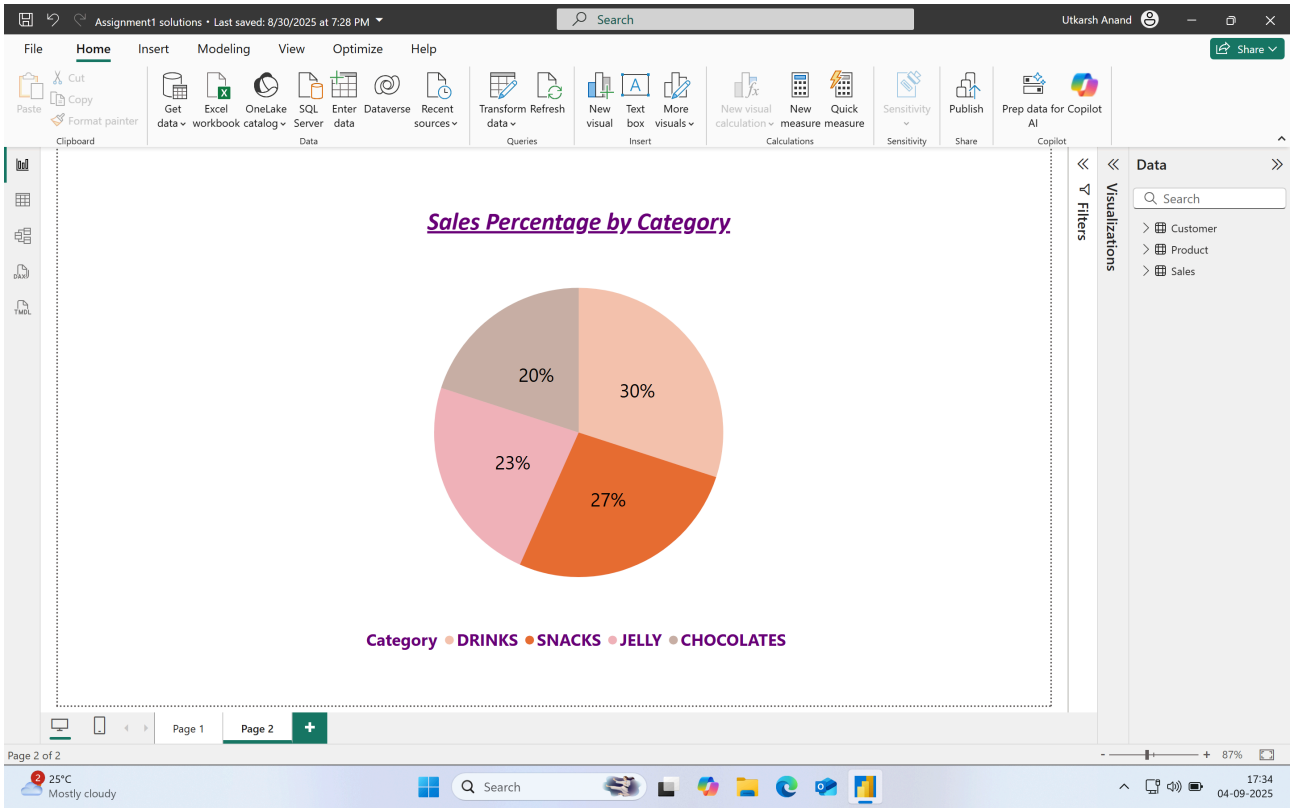
1. Switch to Report view (left sidebar report icon).
2. In the Visualizations pane, click Clustered column chart (or Stacked column chart if you prefer).
3. In Fields:
   - Drag Product[Category] to X-axis.
   - Drag Sales[TotalPrice] to Y-axis and ensure Summarization = Sum.
   - If you don't have a TotalPrice field: create a quick measure:
   - Modeling ribbon → New measure → enter

   Total Sales := SUMX(Sales, Sales[Price] * Sales[Quantity])

•Use Total Sales in Y-axis instead.

4. Format the chart: click the chart → Format (paint-roller icon) →
   - General → Title: On → Title text: "Total Sales by Category (₹)".
   - Data labels: On → Display units: None → Decimal places: 0.
   - Columns/Colors: set distinct colors for readability.
   - X-axis/Y-axis: adjust Text size and Gridlines as needed.
5. Resize and align the visual neatly on the canvas.

Part 2: Sales Percentage by Category (Price)



Steps taken to create a chart showing the percentage share of sales (Price) for each product category, kept compact and clear:

1. In Report view, choose Pie chart (or Donut chart for a compact ring layout).
2. In Fields:
   - Drag Product[Category] to Legend.
   - Drag Sales[Price] to Values (summarization: Sum).
   - (If allowed by your rubric, using TotalPrice here is usually more accurate for share of sales.)

3. Format: Format (paint-roller) →
   - Data labels: On → Label style: Data value and percent of total.
   - Legend: On → position to Bottom or Right to save space.
   - Title: On → “Sales % by Category”.
   - Detail labels: reduce Text size so the visual remains compact and easy to read.
4. Position the chart next to the bar/column chart for a clean dashboard look.

Prepared By,

Utkarsh Anand