Lesson-03

Topic: Data Transformation with Power Query (Part 1)

Prerequisites: Customer_Orders.txt, Orders.txt

1. What is the purpose of the "Applied Steps" pane in Power Query?

It records every transformation step you perform (like filtering rows, renaming columns, or changing types) in chronological order. You can click on any previous step to see how the data looked at that point, or delete a step to undo it.

2. How do you remove duplicate rows in Power Query?

Select the columns that define a unique row (or the whole table), right-click the column header, and select "Remove Duplicates". Alternatively, go to the Home ribbon > Remove Rows > Remove Duplicates.

3. What does the "Filter" icon do in Power Query?

The filter icon (a small arrow/funnel on the column header) allows you to include or exclude specific values. You can uncheck specific items or use logic filters like "Text Filters" (e.g., Contains...) or "Number Filters" (e.g., Greater Than...).

4. How would you rename a column from "CustID" to "CustomerID"?

Double-click the column header "CustID", type "CustomerID", and press Enter.

5. What happens if you click "Close & Apply" in Power Query?

It saves all your transformation steps, closes the Power Query Editor window, and loads the cleaned data into the Power BI Desktop report view, making it ready for visualization.

6. Remove all rows where Quantity is less than 2.

Action: Click the arrow next to the Quantity header. Select Number Filters > Greater Than or Equal To. Enter 2. Click OK.

Result: The rows for "Laptop" and "Monitor" will disappear. Only "Mouse" and "Keyboard" remain.

7. Split the OrderDate column into separate "Year," "Month," and "Day" columns.

Action: Select the OrderDate column. Go to the Add Column ribbon (or Transform ribbon). Click Date > Year > Year. Repeat for Month and Day.

8. Replace all "Mouse" entries in the Product column with "Computer Mouse."

Action: Right-click the Product column. Select Replace Values.

Find Value: Mouse

Replace With: Computer Mouse

9. Sort the table by OrderDate (newest first).

Action: Click the arrow next to the OrderDate header. Select Sort Descending.

10. How would you handle null values in the Price column?

Remove: Filter out rows where Price is null.

Replace: Right-click > Replace Values > Replace null with 0.

Fill: Use "Fill Down" or "Fill Up" if the value can be inferred from adjacent rows.

11. Write custom M-code to add a column calculating TotalSpent = Quantity * Price.

Go to Add Column > Custom Column. In the formula box, enter:

= [Quantity] * [Price]

Note: Column names must be exact and case-sensitive inside square brackets.

12. Group the table by CustID to show total spending per customer.

Action: Select CustID. Click Group By on the Home ribbon.

New column name: Total Spent

Operation: Sum

Column: Price (or the TotalSpent column created in Q11).

Expected Result (based on full file):

101: 1360

102: 75

103: 300

13. Fix inconsistent date formats (e.g., 01/10/2023 vs. 2023-01-10) in OrderDate.

If Power BI doesn't recognize the format automatically:

Right-click the column > Change Type > Using Locale...

Select Date as the type and choose the locale (e.g., English (UK) or English (USA)) that matches the source format of the text.

14. Create a conditional column: Label orders as "High Value" if Price > 100.

Action: Go to Add Column > Conditional Column.

Rule: If Price is greater than 100 then Output "High Value" else Output "Low Value".

15. Optimize the query to reduce refresh time (e.g., remove unused columns early).

Strategy: Always remove unnecessary columns as the very first step (or as early as possible). This reduces the amount of memory Power BI needs to process for all subsequent steps (sorting, filtering, calculating).