Lesson 3

Topic: Data Transformation with Power Query (Part 1)
Prerequisites: Download Customer_Orders.txt, Orders.txt

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

   The "Applied Steps" panel in Power Query records each transformation you apply to your data step by step. Every time you perform an action—like filtering rows, renaming columns, removing duplicates, or changing data types-Power Query logs that action as a separate step.
   - Key purposes:

- Step-by-step tracking of your data transformation process
- Easy debugging – you can click on any step to see the intermediate result
- Reusability and modification – you can go back, modify, delete, or rearrange steps
- Transparency – makes your data preparation process understandable and reproducible
  - Example:

If you filter rows -> rename columns -> change data type, each will appear as:

- Filtered Rows
- Renamed Columns
- Changed Type

.pbit

```
let
   Source = Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI HOMWORK\LESSON-3\Customer_Orders.txt"),[Delimiter=",", Columns=3, Encoding=1252, QuoteStyle=QuoteStyle.None]),
   #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
   #"Merged Queries" = Table.NestedJoin(#"Promoted Headers", {"CustID"}, Orders, {"CustID"}, "Orders", JoinKind.LeftOuter),
   #"Expanded Orders" = Table.ExpandTableColumn(#"Merged Queries", "Orders", {"CustID", "Name", "OrderDate", "Product", "Quantity", "Price"}, {"Orders.CustID", "Orders.Name", "Orders.OrderDate", "Orders.Product", "Orders.Quantity", "Orders.Price"}),
   #"Added Custom" = Table.AddColumn(#"Expanded Orders", "Total Sales", each [Orders.Quantity] * [Orders.Price])
in
   #"Added Custom"
```

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

Select Columns to Check for Duplicates
If you want to remove duplicates based on all columns, you don’t need to select anything.
If you want to remove duplicates based on specific columns, hold Ctrl and click each column header.
Remove Duplicates
Go to the Home tab.
Click Remove Rows → Remove Duplicates
.pbit

```
let
   Source = Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI
HOMWORK\LESSON-3\Customer_Orders.txt"),[Delimiter=",", Columns=3,
Encoding=1252, QuoteStyle=QuoteStyle.None]),
   #"Removed Duplicates" = Table.Distinct(Source, {"Column1"}),
   #"Promoted Headers" = Table.PromoteHeaders(#"Removed Duplicates",
[PromoteAllScalars=true]),
   #"Merged Queries" = Table.NestedJoin(#"Promoted Headers", {"CustID"}, Orders,
{"CustID"}, "Orders", JoinKind.LeftOuter),
   #"Expanded Orders" = Table.ExpandTableColumn(#"Merged Queries", "Orders",
{"CustID", "Name", "OrderDate", "Product", "Quantity", "Price"}, {"Orders.CustID",
"Orders.Name", "Orders.OrderDate", "Orders.Product", "Orders.Quantity",
"Orders.Price"}),
   #"Added Custom" = Table.AddColumn(#"Expanded Orders", "Total Sales", each
[Orders.Quantity] * [Orders.Price]),
   #"Removed Duplicates1" = Table.Distinct(#"Added Custom", {"CustID"})
in
   #"Removed Duplicates1"
```

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

The Filter icon in Power Query allows you to include or exclude specific rows based on the values in a column.

You can use it to:

- Show only rows with certain values (e.g., only orders from "Alice")
- Exclude null or empty values
- Filter by date ranges, number ranges, or text conditions

This operation creates a step called "Filtered Rows" in the Applied Steps panel, making it part of your transformation process.

.pbit

```
let
  Source = Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI HOMWORK\LESSON-3\Orders.txt"),[Delimiter=",", Columns=6, Encoding=1252, QuoteStyle=QuoteStyle.None]),
  #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
  #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"OrderDate", type date}, {"Price", Currency.Type}, {"Quantity", Int64.Type}}),
  #"Filtered Rows" = Table.SelectRows(#"Changed Type", each [OrderDate] > #date(2023, 1, 15))
in
  #"Filtered Rows"
```

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

   To rename a column in Power Query:

   1. Open the Power Query Editor
   2. Right-click the column header CustID
   3. Choose Rename
   4. Type in the new name: CustomerID

   This creates a step called "Renamed Columns" in the Applied Steps panel.

```
let
  Source = Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI HOMWORK\LESSON-3\Customer_Orders.txt"),[Delimiter=",", Columns=3, Encoding=1252, QuoteStyle=QuoteStyle.None]),
  #"Removed Duplicates" = Table.Distinct(Source, {"Column1"}),
```

```
    #"Promoted Headers" = Table.PromoteHeaders(#"Removed Duplicates",
[PromoteAllScalars=true]),

    #"Merged Queries" = Table.NestedJoin(#"Promoted Headers", {"CustID"},
Orders, {"CustID"}, "Orders", JoinKind.LeftOuter),

    #"Expanded Orders" = Table.ExpandTableColumn(#"Merged Queries",
"Orders", {"CustID", "Name", "OrderDate", "Product", "Quantity", "Price"},
{"Orders.CustID", "Orders.Name", "Orders.OrderDate", "Orders.Product",
"Orders.Quantity", "Orders.Price"}),

    #"Added Custom" = Table.AddColumn(#"Expanded Orders", "Total Sales",
each [Orders.Quantity] * [Orders.Price]),

    #"Removed Duplicates1" = Table.Distinct(#"Added Custom", {"CustID"}),

    #"Renamed Columns" = Table.RenameColumns(#"Removed
Duplicates1",{{"CustID", "CustomerID"}})

in

    #"Renamed Columns"
```

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

   When you click "Close & Apply" in Power Query:

   - Power BI applies all the transformations you made in Power Query (filtering, merging, renaming, calculated columns, etc.)
   - Then, it loads the transformed data into the Power BI data model
   - The Power Query Editor closes
   - You can now use the cleaned data for creating visuals, DAX measures, and reports

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

   To remove all rows where the Quantity is less than 2 in Power Query:

   1. Open the Power Query Editor
   2. Find the Quantity column
   3. Click the filter icon next to it
   4. Select Number Filters → Greater Than or Equal To
   5. Enter the value 2

6. Click OK

   This will keep only the rows where Quantity >= 2.
   Power Query will create a step called "Filtered Rows" in the Applied Steps pane.

.pbit

```
let
  Source = Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI HOMWORK\LESSON-3\Orders.txt"),[Delimiter=",", Columns=6, Encoding=1252, QuoteStyle=QuoteStyle.None]),
  #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
  #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"OrderDate", type date}, {"Price", Currency.Type}, {"Quantity", Int64.Type}}),
  #"Filtered Rows" = Table.SelectRows(#"Changed Type", each [Quantity] >= 2)
in
  #"Filtered Rows"
```

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

   Steps:

1. Open Power Query Editor
2. Select the OrderDate column
3. Go to the Add Column tab
4. Click on:

- Date -> Year -> Year
- Date -> Month -> Month
- Date -> Day -> Day

Each of these actions will create a new column:

- OrderDate.Year
- OrderDate.Month
- OrderDate.Day

You can rename them to: Year, Month, Day (optional)

.pbit

```
let
    Source = Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI HOMWORK\LESSON-3\Orders.txt"),[Delimiter=",", Columns=6, Encoding=1252, QuoteStyle=QuoteStyle.None]),
    #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"OrderDate", type date}, {"Price", Currency.Type}, {"Quantity", Int64.Type}}),
    #"Split Column by Delimiter" = Table.SplitColumn(Table.TransformColumnTypes(#"Changed Type", {{"OrderDate", type text}}, "en-US"), "OrderDate", Splitter.SplitTextByDelimiter("/", QuoteStyle.Csv), {"OrderDate.1", "OrderDate.2", "OrderDate.3"}),
    #"Renamed Columns" = Table.RenameColumns(#"Split Column by Delimiter",{{"OrderDate.2", "OrderDate.Day"}, {"OrderDate.1", "OrderDate.Month"}, {"OrderDate.3", "OrderDate.Year"}})
in
    #"Renamed Columns"
```

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

   Open Power Query Editor

   Select the Product column

   Right-click on the column → choose Replace Values

   In the popup:

   - Value To Find: Mouse
   - Replace With: Computer Mouse

   Click OK

This creates a new step called **"Replaced Value"** in the **Applied Steps** pane.

.pbit

```
let

    Source = Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI HOMWORK\LESSON-3\Orders.txt"),[Delimiter=",", Columns=6, Encoding=1252, QuoteStyle=QuoteStyle.None]),

    #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),

    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"OrderDate", type date}, {"Price", Currency.Type}, {"Quantity", Int64.Type}}),

    #"Split Column by Delimiter" = Table.SplitColumn(Table.TransformColumnTypes(#"Changed Type", {{"OrderDate", type text}}, "en-US"), "OrderDate", Splitter.SplitTextByDelimiter("/", QuoteStyle.Csv), {"OrderDate.1", "OrderDate.2", "OrderDate.3"}),

    #"Renamed Columns" = Table.RenameColumns(#"Split Column by Delimiter",{{"OrderDate.2", "OrderDate.Day"}, {"OrderDate.1", "OrderDate.Month"}, {"OrderDate.3", "OrderDate.Year"}}),

    #"Replaced Value" = Table.ReplaceValue(#"Renamed Columns","Mouse","Computer Mouse",Replacer.ReplaceText,{"Product"})

in

    #"Replaced Value"
```

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

    Open **Power Query Editor**

    Find the OrderDate column

    Click the **dropdown arrow** next to OrderDate

    Select **Sort Descending**

This creates a step called **"Sorted Rows"** in the **Applied Steps** panel.

.pbit

```
let
```

```
    Source = Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI
HOMWORK\LESSON-3\Orders.txt"),[Delimiter=",", Columns=6, Encoding=1252,
QuoteStyle=QuoteStyle.None]),

    #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),

    #"Changed Type" = Table.TransformColumnTypes(#"Promoted
Headers",{{"OrderDate", type date}, {"Price", Currency.Type}, {"Quantity",
Int64.Type}}),

    #"Sorted Rows" = Table.Sort(#"Changed Type",{{"OrderDate", Order.Descending}})

in

    #"Sorted Rows"
```

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

1. Replace nulls with a default value (e.g., 0):

- Select the Price column
- Go to Transform tab → Click Replace Values
- Replace null with 0

.pbit

```
let

    Source =
Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI
HOMWORK\LESSON-3\Orders.txt"),[Delimiter=",", Columns=6,
Encoding=1252, QuoteStyle=QuoteStyle.None]),

    #"Promoted Headers" = Table.PromoteHeaders(Source,
[PromoteAllScalars=true]),

    #"Changed Type" = Table.TransformColumnTypes(#"Promoted
Headers",{{"OrderDate", type date}, {"Price", Currency.Type}, {"Quantity",
Int64.Type}}),

    #"Sorted Rows" = Table.Sort(#"Changed Type",{{"OrderDate",
Order.Descending}}),

    #"Replaced Value" = Table.ReplaceValue(#"Sorted
Rows",null,0,Replacer.ReplaceValue,{"Price"})

in

    #"Replaced Value"
```

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

.pbit

```
let
  Source = Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI HOMWORK\LESSON-3\Orders.txt"),[Delimiter=",", Columns=6, Encoding=1252, QuoteStyle=QuoteStyle.None]),
  #"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),
  #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"OrderDate", type date}, {"Price", Currency.Type}, {"Quantity", Int64.Type}}),
  #"Sorted Rows" = Table.Sort(#"Changed Type",{{"OrderDate", Order.Descending}}),
  #"Replaced Value" = Table.ReplaceValue(#"Sorted Rows",null,0,Replacer.ReplaceValue,{"Price"}),
  #"TotalSpent" = Table.AddColumn(#"Replaced Value", "TotalSpent", each [Quantity] * [Price])
in
 #"TotalSpent"
```

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

- Open Power Query Editor
- Make sure you already have a TotalSpent column (from previous step)
- Select the CustID column
- Go to the Home tab → Click Group By

In the popup window:

- Group by: CustID
- New column name: CustomerTotal
- Operation: Sum
- Column: TotalSpent

- Click OK

.pbit

```
let
    Source =
Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI
HOMWORK\LESSON-3\Orders.txt"),[Delimiter=",", Columns=6,
Encoding=1252, QuoteStyle=QuoteStyle.None]),
    #"Promoted Headers" = Table.PromoteHeaders(Source,
[PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted
Headers",{{"OrderDate", type date}, {"Price", Currency.Type}, {"Quantity",
Int64.Type}}),
    #"Sorted Rows" = Table.Sort(#"Changed Type",{{"OrderDate",
Order.Descending}}),
    #"Replaced Value" = Table.ReplaceValue(#"Sorted
Rows",null,0,Replacer.ReplaceValue,{"Price"}),
    #"TotalSpent" = Table.AddColumn(#"Replaced Value", "TotalSpent", each
[Quantity] * [Price]),
    #"Grouped Rows" = Table.Group(TotalSpent, {"CustID"}, {{"CustomerTotal",
each List.Sum([TotalSpent]), type number}})

in
 #"Grouped Rows"
```

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

- Open **Power Query Editor**
- Select the OrderDate column
- Go to the **Transform** tab → Click **Data Type** → Choose **Date**
- Power Query will automatically try to **convert text values to proper date format**

Optional: Split and Rebuild (for custom parsing)

If conversion fails or misinterprets dates, you can:

- Step 1: Change column type to Text
- Step 2: Use Split Column → By Delimiter (/ or -)
- Step 3: Rearrange parts (Year, Month, Day)
- Step 4: Use Add Column → Custom Column:

m

Copy

Edit

= #date(Number.FromText([Year]), Number.FromText([Month]), Number.FromText([Day]))

- Step 5: Change new column type to Date

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

Open Power Query Editor

Select your table.

Add Conditional Column

Go to the "Add Column" tab → Click "Conditional Column".

Set Up the Rule

New column name: OrderValue

Column name: Price

Operator: is greater than

Value: 100

Output: "High Value"

Else: "Standard" (or leave blank for null)

Click OK → "Close & Apply".
.pbit

let

Source = Csv.Document(File.Contents("C:\Users\ХУМОИДДИН\Desktop\POWERBI HOMWORK\LESSON-3\Orders.txt"),[Delimiter=",", Columns=6, Encoding=1252, QuoteStyle=QuoteStyle.None]),

#"Promoted Headers" = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),

```
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted
Headers",{{"OrderDate", type date}, {"Price", Currency.Type}, {"Quantity",
Int64.Type}}),

    #"Sorted Rows" = Table.Sort(#"Changed Type",{{"OrderDate", Order.Descending}}),

    #"Replaced Value" = Table.ReplaceValue(#"Sorted
Rows",null,0,Replacer.ReplaceValue,{"Price"}),

    #"TotalSpent" = Table.AddColumn(#"Replaced Value", "TotalSpent", each [Quantity]
* [Price]),

    #"Changed Type1" = Table.TransformColumnTypes(TotalSpent,{{"OrderDate", type
date}}),

    #"Added Conditional Column" = Table.AddColumn(#"Changed Type1", "
OrderValue", each if [Price] > 100 then "High Value" else """Standard"" ")

in

    #"Added Conditional Column"
```

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

To optimize your Power Query and reduce refresh time, apply the following best practices:

1. Remove Unused Columns Early

- Unused columns consume memory and processing power.
- As early as possible, remove unnecessary columns using:
    - Home → Choose Columns or
    - Remove Columns

M Code Example:

m

CopyEdit

= Table.SelectColumns(#"Previous Step", {"CustID", "OrderDate", "Product", "Price", "Quantity"})

---

2. Filter Early

- Apply filters (e.g., dates, statuses) early in the query to minimize data volume.

  Example:

  m

  CopyEdit

  = Table.SelectRows(#"Previous Step", each [Quantity] >= 2)

---

3. Avoid Changing Data Types Multiple Times

- Set correct data types once, after necessary transformations.

---

4. Disable Load on Intermediate Queries

- Right-click unnecessary intermediate queries → Enable Load ❌ (uncheck it)

---

5. Use Buffering When Needed

- If referencing a query multiple times, buffer it:

  m

  CopyEdit

  = Table.Buffer(Source)

---

6. Avoid Complex Steps in Source Queries

- Keep source queries clean — do transformations in separate query steps.

  let

  // Step 1: Load data from Orders.txt

  Source = Csv.Document(File.Contents("C:\YourPath\Orders.txt"), [Delimiter=",", Columns=6, Encoding=65001, QuoteStyle=QuoteStyle.None]),

  PromotedHeaders = Table.PromoteHeaders(Source, [PromoteAllScalars=true]),

```
    // Step 2: Remove unused columns early (assume we only need CustID, OrderDate, Quantity, Price)
    RemovedOtherColumns = Table.SelectColumns(PromotedHeaders, {"CustID", "OrderDate", "Quantity", "Price"}),

    // Step 3: Convert data types only once
    ChangedTypes = Table.TransformColumnTypes(RemovedOtherColumns, {
        {"CustID", Int64.Type},
        {"OrderDate", type date},
        {"Quantity", Int64.Type},
        {"Price", type number}
    }),

    // Step 4: Buffer the table for performance
    BufferedTable = Table.Buffer(ChangedTypes),

    // Step 5: Add TotalSpent column (Quantity * Price)
    AddedTotalSpent = Table.AddColumn(BufferedTable, "TotalSpent", each [Quantity] * [Price], type number)

in
    AddedTotalSpent
```