Lesson 4
Topic: Data Transformation with Power Query (Part 2)
Prerequisites: Download Customers.xlsx, Orders.csv

1. Difference between "Merge" and "Append" in Power Query
Merge joins two tables based on a matching column (like SQL JOIN), combining columns from both tables into one.

Append stacks tables on top of each other, combining rows (like UNION in SQL). Used when tables have identical columns.

2. Split "Full Name" into "First Name" and "Last Name"
Select the "Full Name" column.

Go to the Home or Transform tab → Split Column → By Delimiter (choose space as delimiter).

This will create two columns: typically "First Name" and "Last Name".

3. What is "Pivot Columns" used for?
Pivot Columns transforms unique values in a column into separate columns, summarizing values (e.g., sum, count) based on another column.

Useful for turning row data into a matrix.

4. Undo a Step in Power Query
In the Applied Steps pane (on the right), right-click the step you want to remove and select Delete.

There is no classic "undo," but you can delete or reorder steps.

5. Purpose of "Reference" vs. "Duplicate" in Queries
Duplicate: Makes a complete copy (clone) of a query; changes do not affect the original.

Reference: Creates a new query that references the result of the original; changes in the source flow to the reference.

6. Merge Orders.csv and Customers.xlsx on CustID (Inner Join)
Load both tables in Power Query.

Select Orders.csv → Home → Merge Queries.

Choose Customers.xlsx and select CustID in both tables.

Select Inner join.

Click OK.

7. Pivot Product column to show total Quantity per product
Select the table.

Go to Transform → Pivot Column.

Select the Product column to pivot.

Set Quantity as the values column, and aggregation as Sum.

8. Append Two Tables with Identical Columns
Load both tables.

Home → Append Queries → Select both tables → Click OK.

9. Use "Fill Down" to Replace Nulls in the Email Column
Select the Email column.

Transform tab → Fill → Down.

This fills empty cells with the last non-null value above.

10. Extract the Domain from the Email Column
Add a Custom Column:

m
Копировать
Редактировать
Text.AfterDelimiter([Email], "@")
This returns the text after "@", i.e., the domain.

11. M-code: Merge Queries Dynamically (e.g., JoinType = "Inner")
m
Копировать
Редактировать

```
let
    JoinType = "Inner", // or "LeftOuter", etc.
    Merged = Table.NestedJoin(
        Orders,
        {"CustID"},
        Customers,
        {"CustID"},
```

"NewColumn",
JoinKind = if JoinType = "Inner" then JoinKind.Inner else JoinKind.LeftOuter
)
in
Merged

12. Unpivot Table with "Jan_Sales", "Feb_Sales" into "Month" and "Sales"
Select "Jan_Sales", "Feb_Sales", etc.

Transform tab → Unpivot Columns.

Result: two columns "Attribute" (month) and "Value" (sales).

13. Handle Errors in a Custom Column using try...otherwise
m
Копировать
Редактировать
= try [Value1] / [Value2] otherwise null
This returns null if there is an error (e.g., division by zero).

14. Function to Clean Phone Numbers (Remove Dashes)
Create a custom function:

m
Копировать
Редактировать
(phone as text) => Text.Select(phone, {"0".."9"})
Or, simply: Text.Replace([Phone], "-", "")

15. Optimize a Query with 10+ Steps (Identify Bottlenecks & Simplify)
Remove unused columns early in the process.

Filter rows early.

Combine related transformations into fewer steps where possible.

Disable loading for intermediate queries.

Review each step in the Applied Steps pane; merge or delete unnecessary ones.