Lesson 4

Topic: Data Transformation with Power Query (Part 2)

1. **What is the difference between "Merge" and "Append" in Power Query?**

- **Merge** = Like a **JOIN** in SQL. Combines two tables **based on a key column**.
- **Append** = Stacks tables **vertically**, adding rows (tables must have identical columns).

2. **How do you split a "Full Name" column into "First Name" and "Last Name"?**

Go to:
Home → Split Column → By Delimiter

- Choose **Space** as the delimiter
- We'll get two columns: First Name and Last Name

3. **What is "Pivot Columns" used for?**

- It **transforms row values into column headers**.
- Example: Pivoting Product to show Quantity per product in separate columns.

4. **How do you undo a step in Power Query?**

- In the **Applied Steps** pane (right side), click the **X** next to the step you want to remove.

5. **What is the purpose of "Reference" vs. "Duplicate" in queries?**

- **Duplicate**: Creates a **copy** of the entire query. It's **independent**.
- **Reference**: Creates a **linked query** that still depends on the original one.

6. Merge Orders.csv and Customers.xlsx on CustID (inner join).

- Go to **Home → Merge Queries**
- Select Orders.csv and Customers.xlsx
- Choose **CustID** column from both
- Set **Join kind** to **Inner**
- Click OK → Expand the merged column to choose needed fields

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

- Select the Product column
- Go to **Transform → Pivot Column**
- For "Values Column", choose **Quantity**
- Aggregation = **Sum**

8. Append two tables with identical columns (e.g., Orders_Jan.csv + Orders_Feb.csv).

- Go to **Home → Append Queries**
- Select both tables (Orders_Jan, Orders_Feb)
- Click OK – rows from both will be combined

9. Use "Fill Down" to replace nulls in the Email column with the previous value.

- Select Email column
- Go to **Transform → Fill → Down**

10. Extract the domain (e.g., "example.com") from the Email column.

Go to: **Add Column → Column From Examples**
Type:

- For john@example.com, write: **example.com**
  Power Query will extract the domain automatically.

Or use M-code:

= Table.AddColumn(PreviousStep, "Domain", each Text.AfterDelimiter([Email], "@"))

11. Write M-code to merge queries dynamically based on a parameter (e.g., JoinType = "Inner").

= Table.NestedJoin(Orders, {"CustID"}, Customers, {"CustID"}, "NewTable", JoinKind.Inner)

You can replace JoinKind.Inner with the parameter value:

= Table.NestedJoin(Orders, {"CustID"}, Customers, {"CustID"}, "NewTable", JoinType)

12. Unpivot a table with columns like "Jan_Sales," "Feb_Sales" into a "Month" and "Sales" format.

- Select the Jan_Sales, Feb_Sales columns
- Go to **Transform → Unpivot Columns**

13. Handle errors in a custom column (e.g., division by zero) using try...otherwise.

Use try ... otherwise:

= Table.AddColumn(PreviousStep, "SafeDivision", each try [Revenue] / [Units] otherwise 0)

14. Create a function in Power Query to clean phone numbers (e.g., remove dashes).

Create a function:

(phone as text) =>

Text.Select(phone, {"0".."9"})

Then apply it with **Transform → Invoke Custom Function**

15. Optimize a query with 10+ steps—identify bottlenecks and simplify.

- **Remove unused columns early** (minimize memory use)
- Combine multiple steps into one (e.g., merging and removing columns together)
- Avoid complex conditional logic if not needed
- Disable **"Enable Load"** for intermediate queries