Hour 2: Power Query for ETL (Extract, Transform, Load)

Instructor's Narrative: "Welcome to Power Query, the most revolutionary tool in Excel for data preparation. If you've ever spent hours cleaning, filtering, or combining data manually, this hour will change your life. Power Query automates the 'data janitor' work. We're going to take messy, real-world data and turn it into a clean, report-ready table—and we'll do it in a way that we can repeat with a single click next month."

---

1. **The Power QPart 1: Introduction & Connecting to Data (15 mins)**

Concept: Power Query is an ETL tool built into Excel. You **Extract** data from a source, **Transform** it (clean, reshape, combine), and **Load** it into your workbook.

Exercise 1: Connecting to a CSV File

Step-by-Step:

Create Sample Data: Open Notepad and paste the following. Save the file as `Sales_Jan.csv` on your desktop.

```
Date,Region,Product,Units,Total_Sales
1/5/2024,North,Widget A,50,2500
1/12/2024,South,Widget B,35,1925
1/19/2024,East,Widget A,28,1400
1/26/2024,North,Widget C,15,900
```

2. In Excel, go to the **Data** tab.
3. Click **Get Data > From File > From Text/CSV**.
4. Navigate to and select your `Sales_Jan.csv` file. Click **Import**.
5. **Open Power Query:**

6. **uery Editor Preview:** A window opens showing a preview of your data. Click **Transform Data** to open the full Power Query Editor.

The Power Query Editor Interface Tour:

- **Ribbon:** Contains all your transformation tools (like a super-powered Find & Replace).
- **Query Settings Pane:** On the right, this shows the steps of your "recipe". This is the magic—every change is recorded and can be modified or deleted.
- **Data Preview:** The main window showing your data.

Part 2: Key Transformation Steps (30 mins)

We will use the data from the CSV. Your steps in the "Applied Steps" list should look like the ones below.

Transformation 1: Basic Cleaning & Filtering

- **Goal:** Change data types and remove a row with errors.
  - **Change Types:** Click the data type icon (`ABC123` or `1.2`) next to each column header and set them correctly: `Date` (Date), `Region` (Text), `Product` (Text), `Units` (Whole Number), `Total_Sales` (Currency).
  - **Filter Rows:** Let's say we only want orders over 30 units. Click the dropdown arrow in the `Units` column. Uncheck "Select All" and then check the box for numbers greater than 30. Click **OK**.
  - **Observe:** Look at the "Applied Steps" pane. You now have `Changed Type` and `Filtered Rows`. You can click on any previous step to see your data at that point in time!

Transformation 2: Unpivoting Data (Converting Columns to Rows)

Concept: Unpivoting converts cross-tabulated data (e.g., months as columns) into a tall, skinny, analysis-friendly format.

New Exercise: Create a new query for this.

1. **Get Data > From Table/Range**. Create this small table in Excel first:

Product	Jan	Feb	Mar
Widget A	100	110	105
Widget B	85	90	95

2. In the Power Query Editor, select the **Jan, Feb, and Mar** columns (click Jan, hold Shift, click Mar).
3. Go to the **Transform** tab and click **Unpivot Columns**.
4. **Result:** Your data now has three columns: Product, Attribute (Month), and Value (Sales). This is the perfect format for PivotTables!
5. **Rename Columns:** Double-click the "Attribute" header and rename it to "Month". Rename "Value" to "Sales".
6. **Close & Load** this to a new sheet to see the final result.

Transformation 3: Merging Queries (A Robust VLOOKUP)

Concept: Combine data from two tables, just like VLOOKUP, but more powerful and easier.

Exercise:

1. **We have our `Sales_Jan` query.** Let's create a second table for product details.
2. In an Excel sheet, create a small table:

Product	Category	Cost_Per_Unit
Widget A	Standard	40
Widget B	Premium	45
Widget C	Standard	50

3. Go back to the Power Query Editor for your `Sales_Jan` query.
4. On the **Home** tab, click **Combine > Merge Queries**.
5. In the dialog box:
    - Top table: `Sales_Jan`.
    - Select the `Product` column.
    - Bottom table: Dropdown and select your `Product_Details` table.
    - Select its `Product` column.
    - Join Kind: **Left Outer** (all from first, matching from second). Click **OK**.
6. A new column `Product_Details` appears. Click the **expand icon (↔)** on the new column's header. Uncheck "Product" and uncheck "Use original column name as prefix". Click **OK**.
7. **Result:** You've just brought in the `Category` and `Cost_Per_Unit` for each product row! The "Applied Steps" shows `Merged Queries` and `Expanded Table`.

########we just revert transformation 4 Because we can't do transformation 5.

Transformation 4: Grouping and Aggregating

- **Goal:** Find the total sales per region.
- With your `Sales_Jan` query open, select the `Region` column.
- Go to the **Home** tab and click **Group By**.
- In the dialog:
  - Group by: `Region`
  - New column name: `Total Sales by Region`
  - Operation: `Sum`
  - Column: `Total_Sales`
  - Click **OK**.
- You now have a summary table. This is a powerful way to create summaries *before* the data even hits your Excel sheet.

Transformation 5: Creating Custom Columns (M Code Basics)

- **Goal:** Add a column for "Profit" (Total_Sales - (Units * Cost_Per_Unit)).
- Go to the **Add Column** tab. Click **Custom Column**.
- In the dialog:
  - New column name: `Profit`
  - Custom column formula: `[Total_Sales] - ([Units] * [Cost_Per_Unit])`
  - **Important:** Note the square brackets `[ ]` around column names. You can double-click column names from the list on the right to add them correctly.
  - Click **OK**.
- You've just written your first line of **M code**, the language behind Power Query!

Part 3: The Power of Parameters & Dynamic Queries (10 mins)

Concept: Make your queries adaptive, so they automatically adjust to new data without you changing the steps.

Exercise: Importing from a Folder of Files

1. **Create a Folder and Files:**

- On your Desktop, create a folder named `Sales_Data_2024`.
- Save your original `Sales_Jan.csv` file inside it.
- Create a second CSV file named `Sales_Feb.csv` with this data and save it in the same folder:

```
Date,Region,Product,Units,Total_Sales
2/2/2024,South,Widget B,40,2200
2/9/2024,West,Widget A,60,3000
2/16/2024,North,Widget C,20,1200
```

2.
3. **Connect to the Folder:**
    - In Excel, go to **Data > Get Data > From File > From Folder**.
    - Navigate to your `Sales_Data_2024` folder and click **Open**.
    - A window will show a list of the files in the folder. Click **Combine & Transform Data**.
4. **Combine the Files:**
    - In the next dialog, ensure `Sales_Jan.csv` is selected as the sample file. Click **OK**.
    - Power Query now opens with the combined data from *both files*! It has automatically appended the February data under the January data.
5. **The Magic of Refresh:**
    - **Close & Load** this combined query to your workbook.
    - Now, go back to your `Sales_Data_2024` folder and create a new file, `Sales_Mar.csv`:

```
Date,Region,Product,Units,Total_Sales
3/1/2024,East,Widget A,25,1250
```

6.
    - Back in Excel, right-click your combined data table and click **Refresh**.
    - **Voilà!** The March data is now automatically included in your table. This is the power of a dynamic, repeatable process.

Capstone Exercise: Build a Monthly Report Pipeline

Objective: Combine everything into a single workflow.

Final Steps:

1. You now have a combined query from the folder (let's call it `CombinedSales`).
2. **Merge** this query with your `Product_Details` table to bring in the category and cost.
3. **Create a Custom Column** to calculate Profit.
4. **Apply a Filter** to show only "Premium" products.
5. **Group By** `Region` to see the total profit for premium products by region.
6. **Load** this final, cleaned, and aggregated table to your workbook.

Your final "Applied Steps" pane is your data cleaning recipe. Next month, all you have to do is drop the new `Sales_Mar.csv` file into your folder, refresh the query in Excel, and your entire report—from raw data to final summary—updates automatically.

Congratulations! You are no longer a data janitor. You are a data engineer. You've just learned a skill that saves countless hours and is the foundation of modern data analysis in Excel.