This comprehensive guide will walk you through cleaning and transforming a messy employee dataset using Power Query in Excel. By the end, you'll be comfortable with Power Query's essential features and understand how to handle common data cleaning challenges.

**Topics you will learn in this lesson:**

* Data cleaning and transformations with Power Query (PQ)
* Adding columns with PQ
* How to load and refresh data?
* Working with business data (HR Staff example)
* Understanding “column quality”
* Using advanced transformations
* Creating columns based on business rules
* Writing M code (simple example)
* Splitting values
* Updating data and refreshing the query

**Getting Started**

1. **Open your Excel file with the employee dataset**, sample-staff-data.xlsx
   * Click on the Data tab in the ribbon
   * Select "Get Data" → "From Table/Range"
   * If your data is already formatted as a table, select any cell within it
   * If not, select the entire data range including headers
   * Click "OK"
2. **The Power Query Editor window will open**
   * Take a moment to familiarize yourself with the interface
   * The main area shows your data preview
   * The ribbon contains transformation tools
   * The right panel shows Query Settings and Applied Steps
   * The left panel shows available queries (if you have multiple)

**Cleaning Text Data**

**Step 1: Trim Whitespace from the Name Column**

1. **Click on the Name column header** to select it
2. **Right-click** and select "Transform" from the context menu
3. **Select "Trim"** to remove leading and trailing spaces
4. **Note in the Applied Steps** panel on the right that a new step called "Trimmed Text" has been added

**Handling Missing Data**

**Step 2: Missing Values in the Gender Column**

1. **Click on the Gender column header**
2. **Look at the column quality indicator** (green bar below the column name)
   * Hover over it to see data quality statistics
   * Notice it shows some values are null (missing)
3. **Right-click on the Gender column** header
4. **Select "Replace Values"**
5. **In the dialog box:**
   * Value To Find: Leave empty (this represents NULL)
   * Replace With: Type "Missing"
   * Click OK
6. **Observe that** all NULL values have been replaced with "Missing"
7. **Note:** When your data is loaded to Excel, you can now filter and identify these missing values easily

**Step 3: Modifying Previous Steps (optional)**

1. **If you want to change the replacement text:**
   * Look at the Applied Steps panel on the right
   * Find the step that replaced NULL values (likely named "Replaced Value")
   * Click on it
   * Click the settings gear icon ⚙️ next to it
   * Change "Missing" to any other text you prefer
   * Click OK
2. **To remove a step:**
   * Right-click on the step in the Applied Steps panel
   * Select "Delete"
   * Confirm deletion if prompted

**Replacing Specific Values**

**Step 4: Fix Department Column with Question Marks, ???**

1. **Click on the Department column header**
2. **Right-click** and select "Replace Values"
3. **In the dialog box:**
   * Value To Find: Type "???"
   * Replace With: Type "Engineering"
   * Click OK
4. **Verify** that all instances of "???" are now "Engineering"

**Filtering Data**

**Step 5: Remove Employees with Null or Zero Salaries**

1. **Click the filter button** (▼) on the Salary column header
2. **Uncheck "null"** and **uncheck "0"**
3. **Click OK**
4. **Note in the Formula Bar** at the top that a filter expression has been applied in M language

= Table.SelectRows(#"Replaced Value1", each ([Salary] <> null and [Salary] <> 0))

1. **Observe in the Applied Steps** panel that a new "Filtered Rows" step has been added

**Working with Dates**

**Step 6: Format the StartDate Column Properly**

1. **Click on the StartDate column header**
2. **Right-click** and select "Change Type"
3. **Select "Date"** from the menu
4. **Notice the calendar icon** appears in the column header, indicating it's now formatted as a date
5. **Check a few values** to confirm they display as proper dates

**Splitting Text Data**

**Step 7: Split the Name Column into First and Last Names**

1. **Click on the Name column header** (which you trimmed earlier)
2. **Go to the Transform tab** in the ribbon
3. **Click "Split Column"** → "By Delimiter"
4. **In the dialog box:**
   * Select "Space" as the delimiter
   * Select "At the left-most delimiter" option
   * Click OK
5. **Observe** that the Name column has been split into "Name.1" and "Name.2" columns

**Step 8: Rename the Split Columns**

1. **Right-click on the "Name.1" column header**
2. **Select "Rename"**
3. **Type "FirstName"** and press Enter
4. **Right-click on the "Name.2" column header**
5. **Select "Rename"**
6. **Type "LastName"** and press Enter

**Step 9: Handle Names with Middle Names**

1. **Examine the LastName column** to see if any entries contain middle names
2. **If middle names are present:**
   * Click on the LastName column
   * Right-click and select "Trim" to remove any extra spaces
   * Click on the LastName column again
   * Go to Transform tab → "Split Column" → "By Delimiter"
   * Select "Space" as the delimiter
   * This time, choose "At the right-most delimiter" option
   * Click OK
   * This will split into "LastName.1" (which might contain middle names) and "LastName.2" (the actual last name)
   * Rename these columns as "MiddleName" and "LastName"
3. **If you want to remove the middle name column:**
   * Right-click on the "MiddleName" column header
   * Select "Remove"

**Creating Conditional Columns**

**Step 10: Categorize Salaries into Ranges**

1. **Click on the Salary column header**
2. **Right-click** and select "Change Type" → "Decimal Number" to ensure proper formatting
3. **Go to the Add Column tab** in the ribbon
4. **Click "Conditional Column"**
5. **In the dialog box:**
   * New column name: "SalaryRange"
   * Add your conditions:
     + If "Salary" is less than 50000, then value is "Below 50K"
     + Add new rule: If "Salary" is less than 100000, then value is "50K-100K"
     + Otherwise (else): "Above 100K"
   * Click OK
6. **Verify** the new SalaryRange column contains the correct categories

**Step 11: Reorder Columns**

1. **Click and hold** on the SalaryRange column header
2. **Drag it** to your preferred position (for example, next to the Salary column)
3. **Release** to drop it in the new position

**Calculating Date-Based Values**

**Step 12: Calculate Employee Tenure in Years**

1. **Click on the StartDate column header**
2. **Go to the Add Column tab** in the ribbon
3. **Click on the "Date" button** to see date-related options
4. **Select "Age"**
5. **“Age” is calculated in days from current date**
   * Choose the current date or a specific reference date
   * Click OK

= Table.AddColumn(#"Reordered Columns", "Age", each Date.From(DateTime.LocalNow()) - [Start Date], type duration)

1. **This creates a duration column**
2. **To convert to years:**
   * Click on the new column
   * Go to Transform tab
   * Click "Duration" → "Total Years"
3. **Rename this column** to "YearsOfService" by right-clicking and selecting "Rename"

**Creating Custom Columns with Logic**

**Step 13: Add an Employee Type Column Based on FTE**

1. **Go to the Add Column tab** in the ribbon
2. **Click "Custom Column"**
3. **In the dialog box:**
   * New column name: "Employee Type"
   * Custom column formula: if [FTE] = 1 then "Full-Time" else "Part-Time"
   * Click OK
4. **Observe** the new EmployeeType column containing "Full-Time" or "Part-Time" based on the FTE value

**Step 14: Rename a Column**

1. **Right-click on the "EmployeeType" column header**
2. **Select "Rename"**
3. **Type "WorkType"** and press Enter
4. **Alternatively**, you can find the step in the Applied Steps panel, click the gear icon, and change the name there

**Finalizing and Loading the Data**

**Step 15: Load Your Transformed Data to Excel**

1. **In the Power Query Editor, click the Home tab**
2. **Click "Close & Load"** to load the data to a new worksheet
3. **Or click "Close & Load To..."** for more options:
   * Load to Table: Creates a formatted Excel table
   * Load to PivotTable: Creates a PivotTable with your data
   * Load to PivotChart: Creates a PivotChart
   * Load to Data Model only: Adds to the Data Model without creating a worksheet
4. **Select your preferred option** and click Load

**Analyzing HR Data Using Pivot Tables**

**Question 1: Salary Distribution Analysis**

**Which departments have the highest representation in the "Above 100k" salary bucket, and what percentage of employees in these departments work remotely?**

Step 1: Create the basic pivot table

1. Select your entire data table including headers
2. Go to Insert tab → PivotTable
3. In the dialog box, ensure your data range is correct and select "New Worksheet"
4. Click OK

Step 2: Set up the pivot table fields

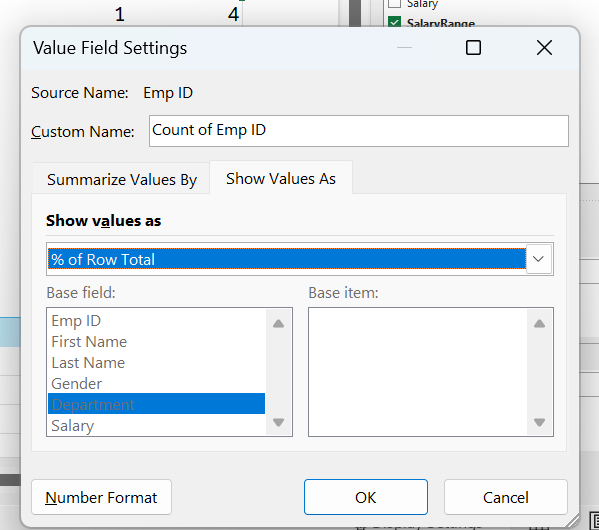
1. In the PivotTable Fields pane:
   * Drag "Department" to the ROWS area
   * Drag "Salary Range" to the FILTERS area
   * Drag "Emp ID" to the VALUES area (it will count the IDs)

Step 3: Filter for "Above 100k" salary bucket

1. Click the filter dropdown for "Salary Range"
2. Select only "Above 100k"
3. Click OK

Step 4: Add remote worker calculation

1. Drag "Work location" to the COLUMNS area
2. You'll now see a breakdown of each department by work location
3. Right-click on the Grand Total column → Value Field Settings → Show Values As → % of Row



Step 5: Format and finalize

1. Highlight all values in the "Remote" columns
2. Right-click → Number Format → Percentage → 2 decimal places
3. Add Conditional Formatting → Color Scales → Green White Color Scale
4. Now you can see which departments have the most "Above 100k" employees and what percentage work remotely

**Question 2: Part-Time vs. Full-Time Tenure Comparison**

**What is the average tenure (in years) for part-time employees compared to full-time employees, and is there a significant difference between these groups?**

Step 1: Create a new pivot table

1. Select your entire data table including headers
2. Go to Insert tab → PivotTable
3. Choose New Worksheet → OK

Step 2: Set up the pivot table fields

1. In the PivotTable Fields pane:
   * Drag "Work Type" to the ROWS area
   * Drag "Years of Service" to the VALUES area (it will automatically sum)
   * Change the value calculation: Right-click on "Sum of Years of Service" → Value Field Settings → Average → OK
   * Drag "Emp ID" to the VALUES area (to count employees)

Step 3: Format the results

1. Right-click on the average values → Number Format → Number → 2 decimal places
2. Right-click on the employee count → Value Field Settings → Number Format → Number → 0 decimal places

Step 4: Calculate the difference (optional)

1. Insert a new column to the right of your pivot table
2. Label it "Difference in Years"
3. Use a formula to calculate the difference between full-time and part-time values
4. Format as a number with 2 decimal places

Step 5: Visualize the results (optional)

1. Select the average tenure values
2. Insert tab → Column Chart
3. Format as needed

**Question 3: Gender Distribution Across Salary Buckets**

**How does gender representation vary across the different salary buckets, and which salary range shows the greatest gender imbalance?**

Step 1: Create a new pivot table

1. Select your entire data table including headers
2. Go to Insert tab → PivotTable
3. Choose New Worksheet → OK

Step 2: Set up the basic pivot table

1. In the PivotTable Fields pane:
   * Drag "Salary Bucket" to the ROWS area
   * Drag "Gender" to the COLUMNS area
   * Drag "Employee ID" to the VALUES area (it will count IDs)

Step 3: Calculate percentages

1. Right-click on any value in the Count of Employee ID area
2. Select Value Field Settings → Show values as → % of Row
3. Click OK
4. Format as percentages: Right-click → Number Format → Percentage → 2 decimal places

Step 4: Calculate gender imbalance

1. Add a column to the right of your pivot table labeled "Gender Imbalance"
2. Assuming your pivot table has Male and Female columns, create a formula:
   * =ABS(C3-D3) (adjust cell references to match your data)
   * This calculates the absolute difference between male and female percentages
3. Format as percentage

Step 5: Identify the salary bucket with the greatest gender imbalance

1. Click on your "Gender Imbalance" column
2. Home tab → Sort & Filter → Sort Largest to Smallest
3. The salary bucket at the top now shows the greatest gender imbalance

Step 6: Visualize the data (optional)

1. Select your pivot table data
2. Insert tab → Recommended Charts
3. Choose a 100% Stacked Bar chart → OK
4. Add appropriate titles and labels

Additional Tips for All Pivot Tables:

1. Add slicers for interactive filtering:
   * Select your pivot table
   * PivotTable Tools → Analyze tab → Insert Slicer
   * Select fields like Department, Work location, etc.
   * Arrange slicers on your worksheet for easy filtering
2. Create a dashboard:
   * Create all three pivot tables on separate sheets
   * Add an additional sheet named "Dashboard"
   * Insert → PivotChart → Create PivotChart & PivotTable
   * Copy and arrange your charts on the dashboard
   * Add slicers that connect to multiple pivot tables
3. **Add a title for the first chart section**
   * Click cell A4 and type "Gender Distribution by Salary Bucket"
   * Format as heading: Home tab → Cell Styles → Heading 2
   * Adjust column widths if necessary
4. **Copy charts from other Sheets** 
   * Repeat the copy process for charts on other Sheets
   * Paste in appropriate locations on the dashboard:
   * Add appropriate section headings above each chart

**Step 3: Add Interactive Slicers for Filtering**

1. **Create department slicer**
   * Go to any sheet with a pivot table)
   * Click anywhere in the pivot table
   * PivotTable Tools → Analyze tab → Insert Slicer
   * Check "Department" → OK
   * Copy the slicer (Ctrl+C)
   * Go to Dashboard sheet and paste (Ctrl+V)
   * Position in the upper right section
2. **Create Work Type slicer**
   * Repeat the process but select "Work Type"
   * Position below the Department slicer
3. **Create Work Location slicer**
   * Repeat the process but select "Work Location"
   * Position below the Work Type slicer
4. **Connect slicers to all pivot tables/charts**
   * Right-click on the Department slicer → Slicer Settings
   * Click "Report Connections"
   * Check all your pivot tables from other Sheets
   * Click OK
   * Repeat for all slicers

**Step 4: Add Key Metrics Summary Section**

1. **Create metrics summary area**
   * In cell, type "Key HR Metrics"
   * Format as heading: Home tab → Cell Styles → Heading 2
2. **Add metrics from your pivot tables**
   * Create a small table with labels and values:
     + "Total Employees"
     + "Avg. Tenure (Years)"
     + "Remote Workers %"
     + "Gender Balance Index"
   * Format cells I16

with borders: Home tab → Borders → All Borders

1. **Link metrics to source pivot tables**
   * In cell I16, link to the employee count from one of your pivot tables:
     + Type = then click on the cell in your pivot table with the total count
     + If not directly available, create a formula like =COUNTA(Raw\_Data[Employee ID])
     + Format as number with 0 decimal places
   * Repeat for other metrics, linking to appropriate pivot table cells

**Step 5: Add Data Insights Section**

1. **Create an insights box**
   * In cell A25, type "Key Insights"
   * Format as heading: Home tab → Cell Styles → Heading 2
   * Insert a text box: Insert tab → Text → Text Box
   * Draw it across cells A26
   * Type 3-4 bullet points highlighting the main findings from your analysis
2. **Format the insights box**
   * Right-click text box → Format Shape
   * Fill: Light yellow
   * Line: Thin gray
   * Add bullet points using Alt+7 (from number pad) or Insert → Symbol → Bullets

**Step 6: Polish the Dashboard Design**

1. **Add corporate branding (optional)**
   * Insert tab → Pictures → Browse to your company logo
   * Place in the header area
   * Resize appropriately
2. **Create a consistent color scheme**
   * Right-click on each chart → Format Chart Area
   * Use consistent colors across all charts
   * Match colors to your company brand if applicable
3. **Add gridlines or separators**
   * Insert tab → Shapes → Line
   * Draw horizontal lines between dashboard sections
   * Format lines: thin gray (0.75pt)
4. **Format slicer styles**
   * Select all slicers (hold Ctrl and click each one)
   * Slicer Tools → Options tab → Slicer Styles
   * Choose a style that matches your dashboard color scheme
5. **Add navigation buttons (optional)**
   * Insert tab → Shapes → Rounded Rectangle
   * Create 3-4 small buttons
   * Right-click → Edit Text → Add labels "Detail 1", "Detail 2", etc.
   * Right-click → Assign Macro → Record a macro that moves to the corresponding sheet
   * Alternatively, add hyperlinks: Right-click → Hyperlink → Place in This Document → select sheet

**Step 7: Add Last Update Information and Print Settings**

1. **Add update information**
   * In cell K2, type "Last Updated:"
   * In cell K3, enter formula =NOW() or manually enter date
   * Format as date and time: Right-click → Format Cells → Custom → "mm/dd/yyyy hh”
2. **Set up print area and settings**
   * Page Layout tab → Print Area → Set Print Area
   * Page Layout tab → Scale to Fit → Width: 1 page, Height: 1 page
   * Page Setup → Header/Footer → Custom Footer → Add file name and page numbers
3. **Protect the dashboard (optional)**
   * Review tab → Protect Sheet
   * Allow users to select locked/unlocked cells (to use slicers)
   * Add a password if necessary
   * Click OK

**Step 8: Test the Dashboard Functionality**

1. **Check slicer connections**
   * Click different options in each slicer
   * Verify that all charts update correctly
2. **Test any buttons or navigation elements**
   * Click each button to ensure it leads to the correct sheet
3. **Check data refresh**
   * Data tab → Refresh All
   * Verify that all dashboard elements update correctly
4. **Create final documentation**
   * Right-click on the worksheet tabs → Insert → New sheet
   * Name it "Documentation"
   * Add details about:
     + Data sources
     + Last update
     + How to refresh data
     + How to interpret each chart