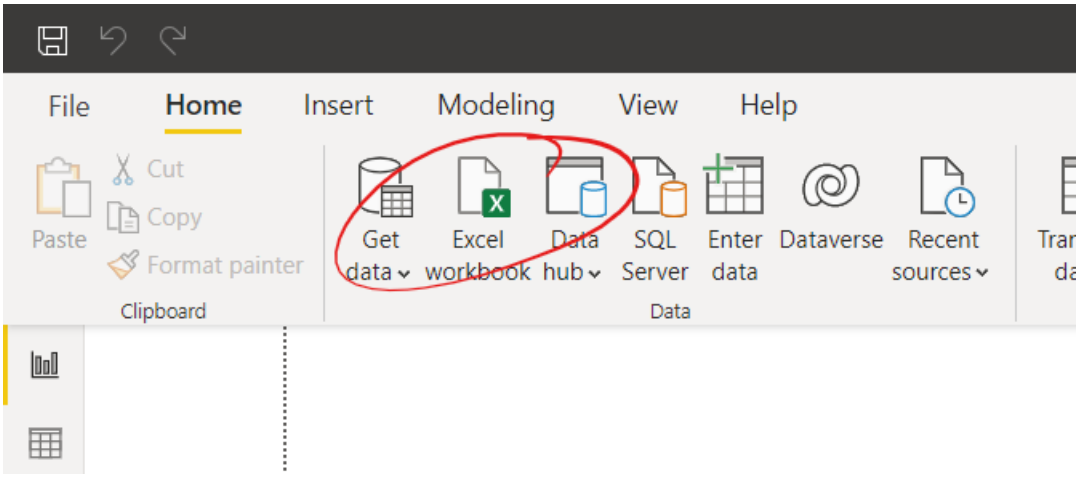
Tutorial

Introduction to PowerBI

1. Load Data

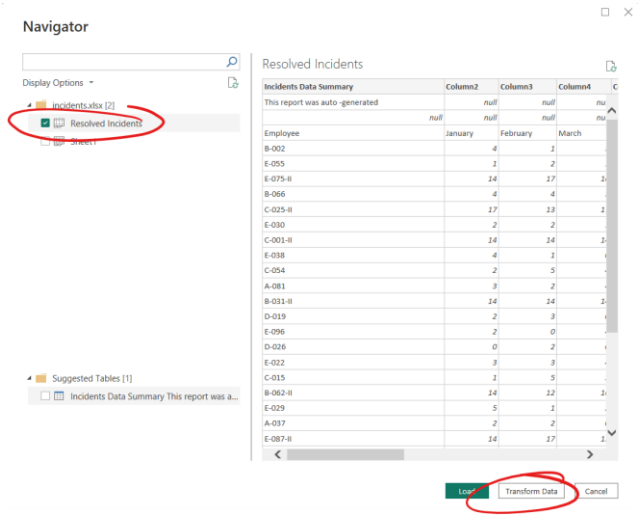
1. Open PowerBI
2. Click on Excel Workbook to load data from Excel



3. Open the file **incidents.xlsx**

Note that Power BI automatically suggests a the correct range to load, but for the sake of this exercise, we'll load the Resolved Incidents sheet.

4. Select **Resolved Incidents** and click **Transform Data**

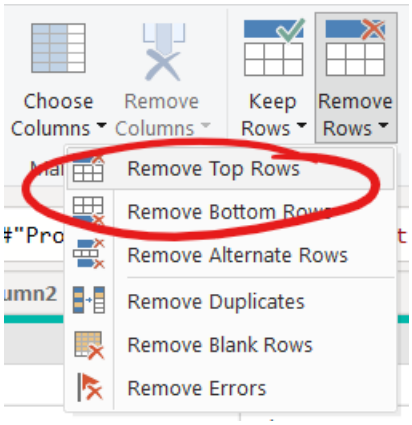


2. Remove irrelevant rows

We often need to remove some of the top rows, for example, if they are blank or if they contain data that you do not need in your reports.

You will notice that the two top rows need to be removed.

1. Click **Remove Rows** -> **Remove Top Rows**



2. Remove 2 rows

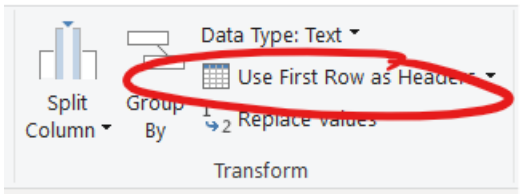
While we're removing rows, let's also remove the bottom rows.

3. Remove 4 bottom rows

3. Promote headers

You will notice that the column names appear in the first row.

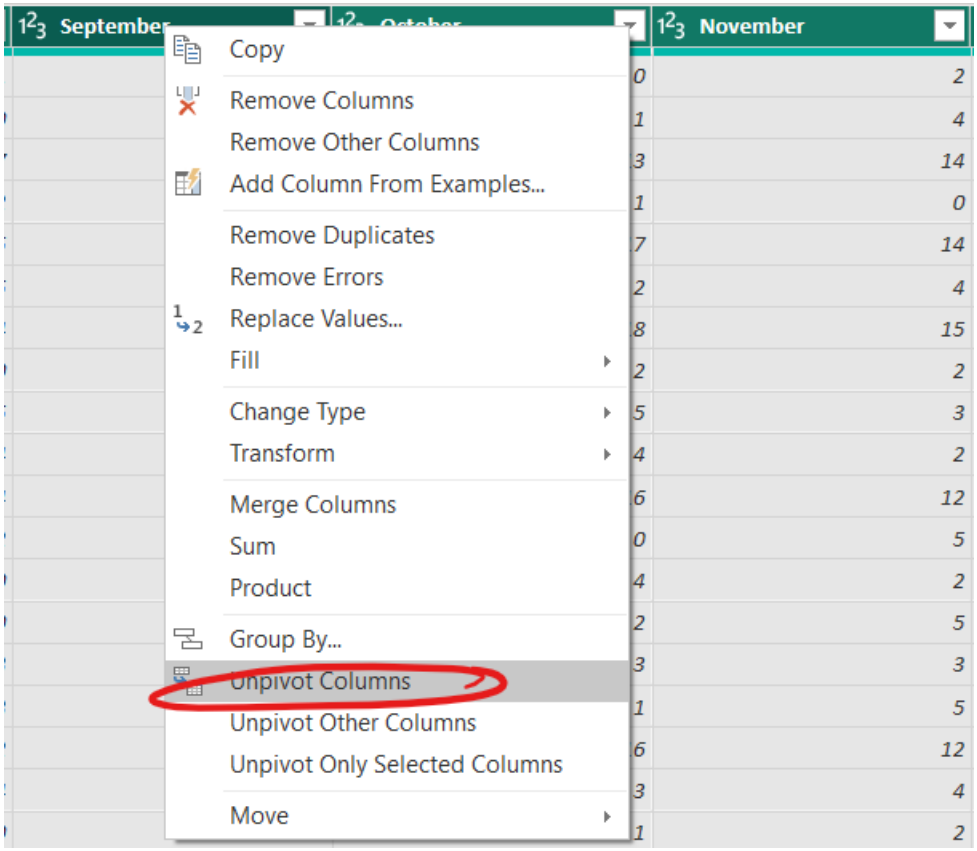
1. Click on **Use First Row as Headers** to use the correct column names.



4. Unpivot columns

The data is not in a good format to be analysed, because multiple observations are split across columns (January, February, etc). It's easier to analyse if data related to different times appears in separate rows.

1. Select all columns except the first
2. Right-click on a selected column header and select **Unpivot columns**



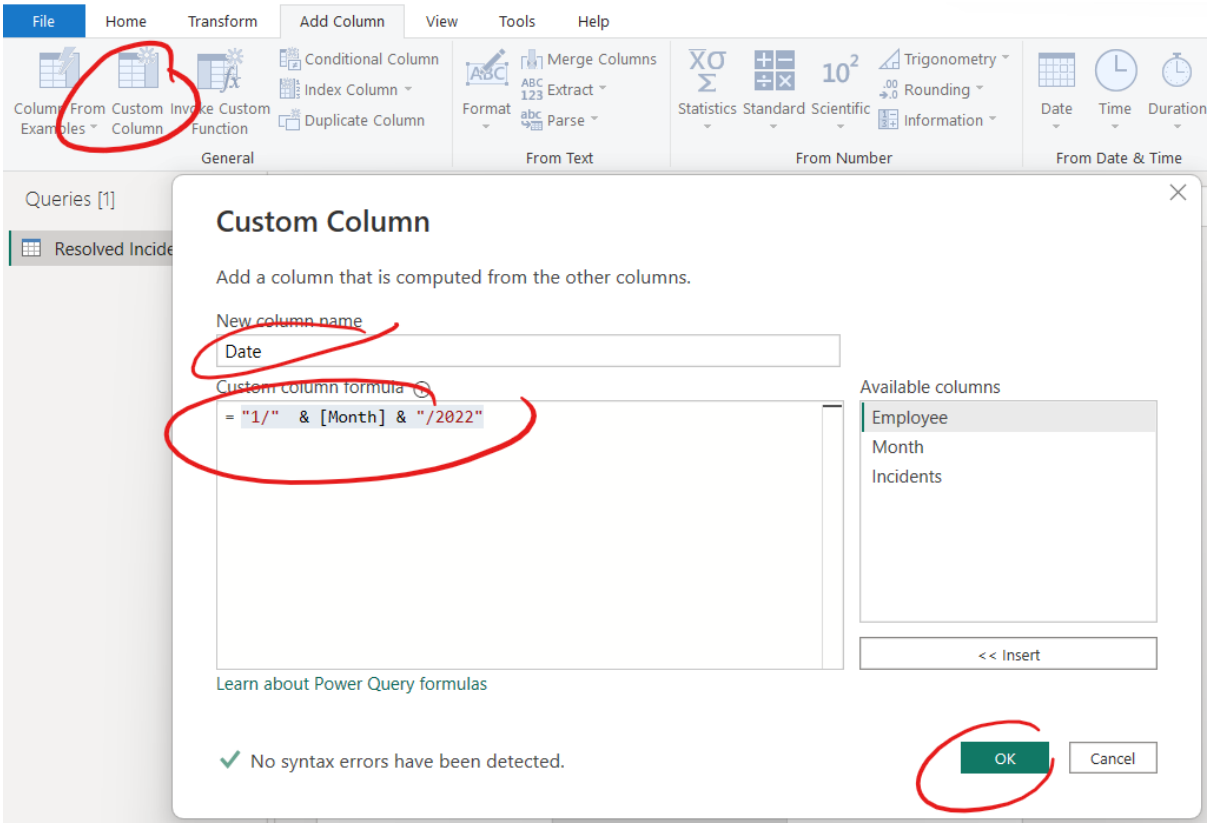
3. Rename the **Attribute** column to **Month** (right-click the column header and select **rename**)
4. Rename the **Value** column to **Incidents**

5. Sorting columns

The Month column is text column. Power BI sorts text columns alphabetically, which is obviously not working in this example.

To fix this, we're creating a custom column which contains a date.

1. Click on **Custom Column** (in the **Add Column** tab)
2. Call the new column **Date** and use this formula
   = "1/" & [Month] & "/2022"
3. Click **OK**



4. Right-click the **Date** column header and select **Change Type** -> **Date**

We now have a proper date column. Now, we can create a column **Month Number**.

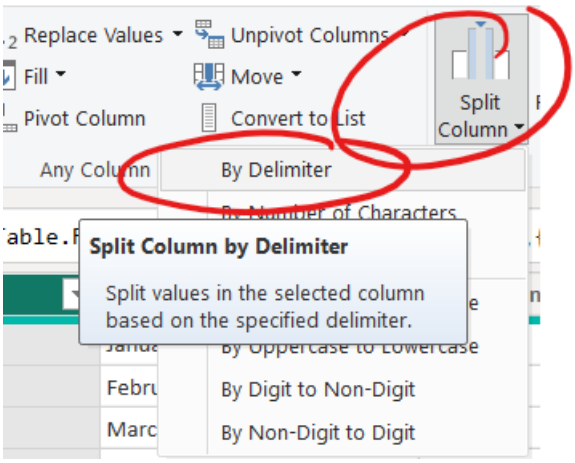
5. Click on **Date**, select **Month** -> **Month**
6. Rename the newly created column to **Month Number**
7. Remove the **Date** column (we don't need it anymore)

We will use this column later to order the month column.

6. Split columns

The Employee column contains three pieces of information: The employee location, employee Id and whether the employee is a Tier 2 employee. We need to split it into three columns to be able to analyse it.

1. Select the **Employee** column and click **Split Column** -> **By delimiter** (on the **Transform** tab)



2. Click **OK**

Power BI created three columns. We need to rename them accordingly.

1. Rename **Employee.1** to **Location**
2. Rename **Employee.2** to **Employee Id**
3. Rename **Employee.3** to **Tier**

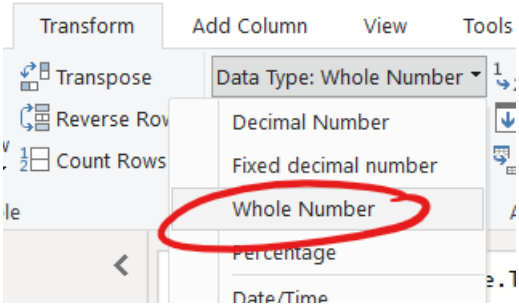
7. Replace values

Finally, we want to replace the values in the **Tier** column

1. Right the **Tier column** header and select **Replace values**
2. Replace **II** with **2**
3. Click **Replace values** again
4. Replace **null** with **1**

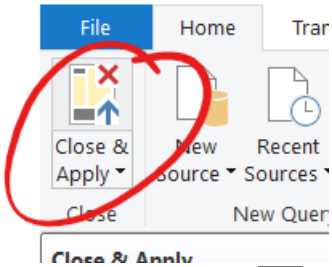
Now that the column only contains number, we should change its data type.

5. Change the data type to **Whole Number**



8. Apply changes and visualise data

1. Click **Close & Apply** to apply the changes and load the data into Power BI

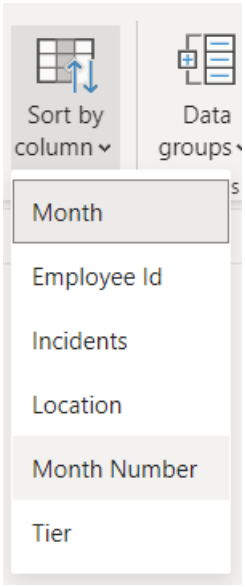


2. Add an **Area Chart** ( )
3. Drag **Month** to X-axis, **Incidents** to Y-Axis and **Tier** to Legend

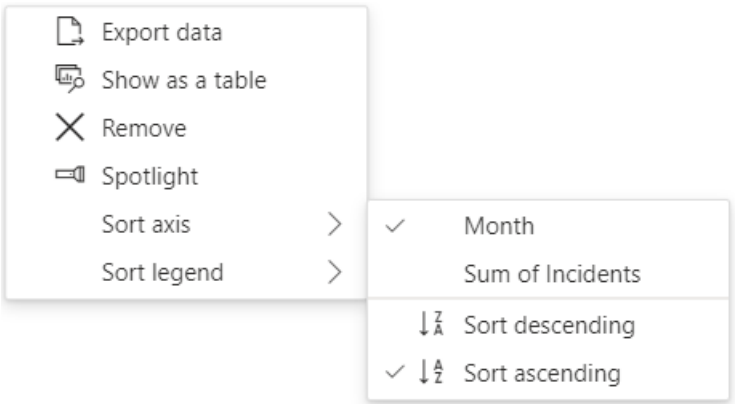
You'll see that the months are ordered more or less randomly.

4. Change to the Data View

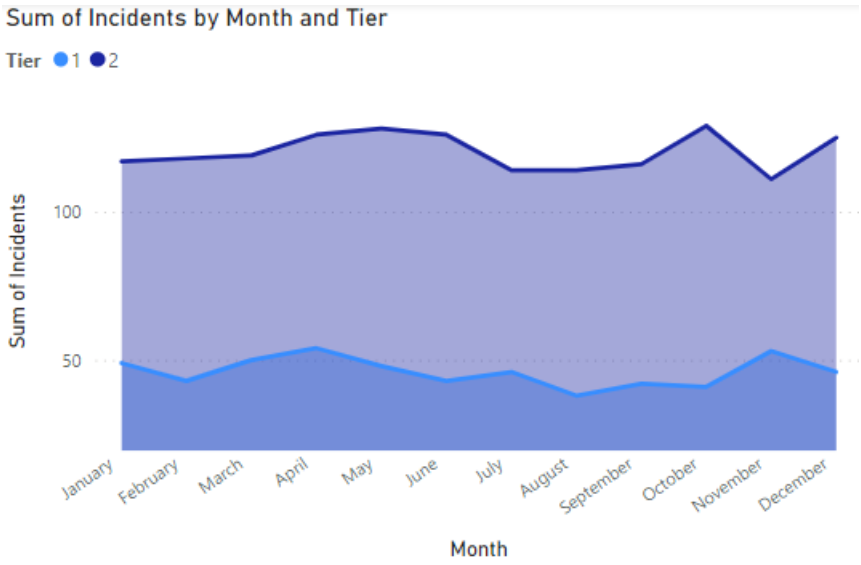
5. Select the **Month** column
6. Click on **Sort by column** -> **Month Number**



7. Go back to the **Report View** ( )
8. Click the ellipsis (···) and sort the chart by Month



Your chart should look similar to the one below.



9. Append multiple tables

Data is sometimes stored in multiple tables. Let's consider the example in **Sales.xlsx**, where we have sales data from London and Manchester.

London

Id	Sales	Date
3085	160.5	03/01/2023
3540	55.85	03/01/2023
6475	23.6	04/01/2023
8089	205.75	05/01/2023

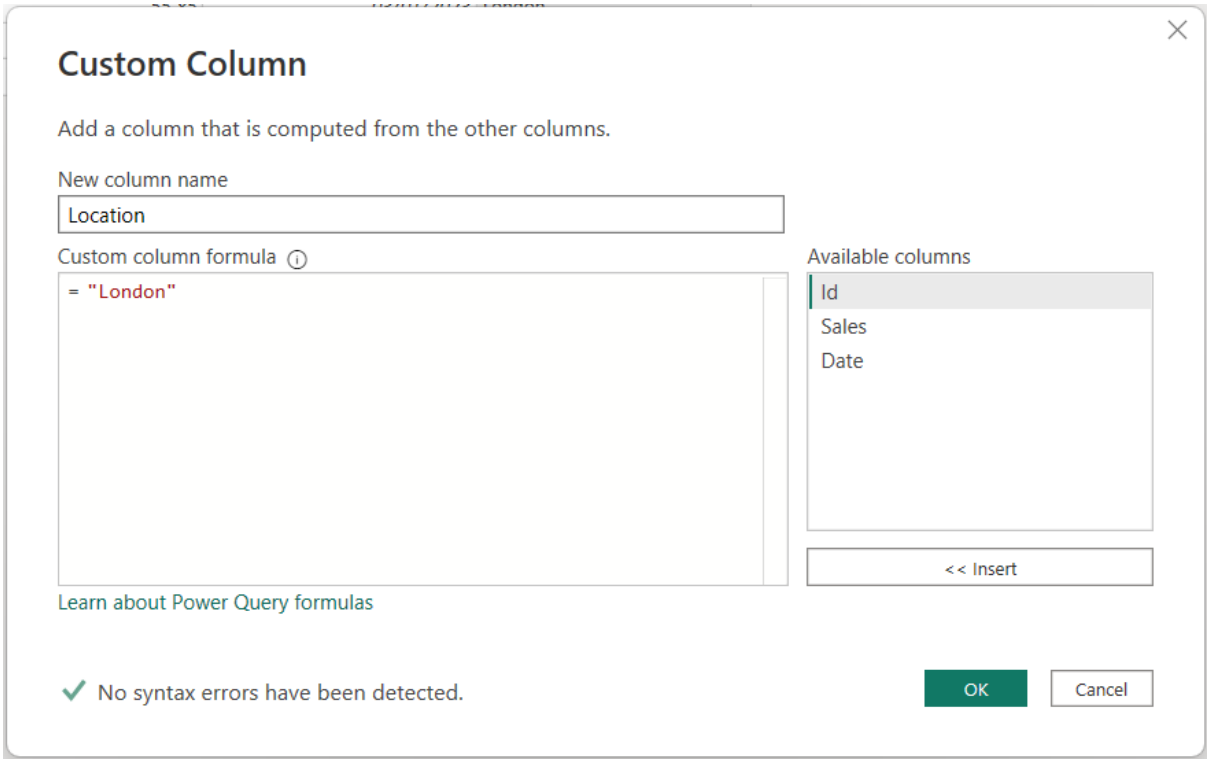
Manchester

Id	Sales	Date
354	15.85	03/01/2023
1654	539.45	04/01/2023
3248	20	04/01/2023
6498	35.8	05/01/2023
9566	12.05	05/01/2023

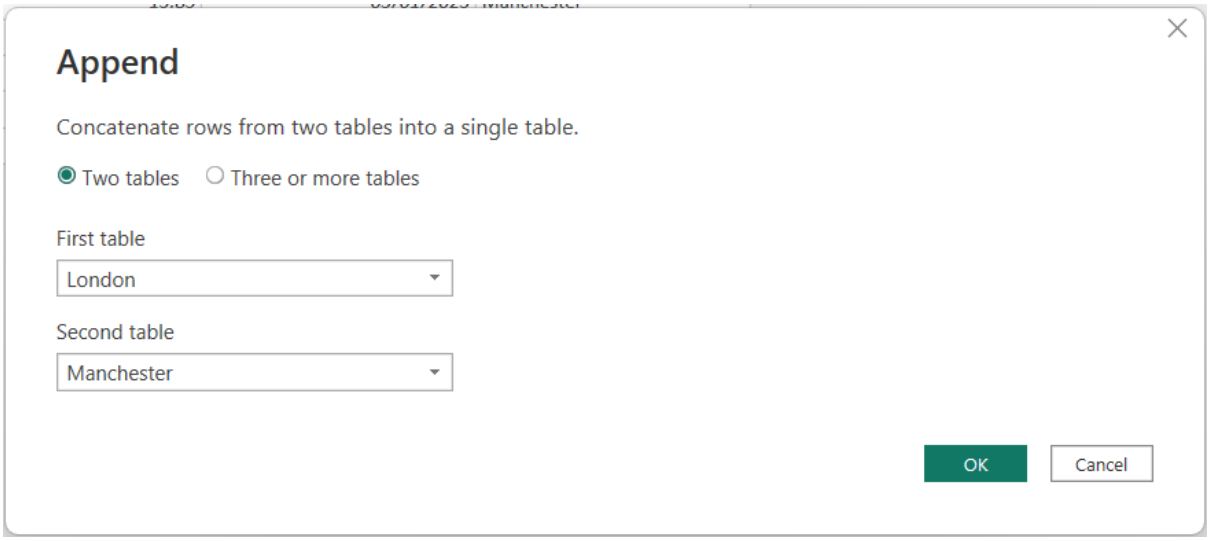
We want to create one table with data from London and Manchester.

1. Load the Excel Workbook **Sales.xlsx**

2. Tick all three tables (London, Manchester & Seller) and click on **Transform Data**
3. Select the **London** table
4. Add a custom column with the name **Location** and the formula = "London"



5. Do the same for the **Manchester** query
6. Click on **Append Queries** -> **Append Queries as New** (on the **Home** tab in the **Combine** group)
7. Select **London** and **Manchester**



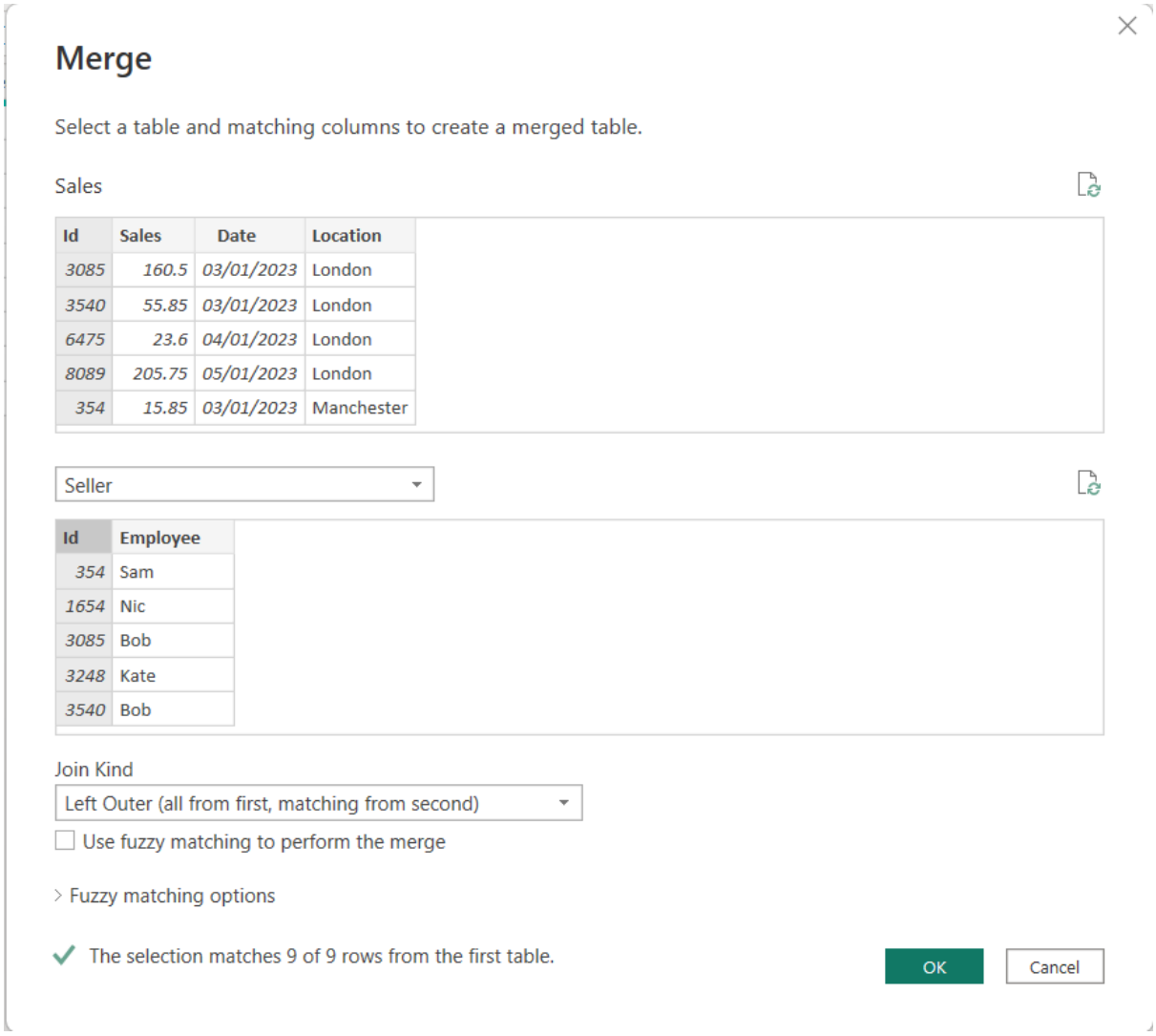
8. Rename the **Append.1** table to **Sales**

10. Merge tables

Another common operation you might need is to merge tables. This is similar to a SQL-join.

In this example, we want to merge the Seller and Sales tables.

1. Select the **Sales** table
2. Click **Merge Queries** -> **Merge Queries**
3. Select the **Seller** table as the second table
4. Select both **id** columns



You can select the join kind:

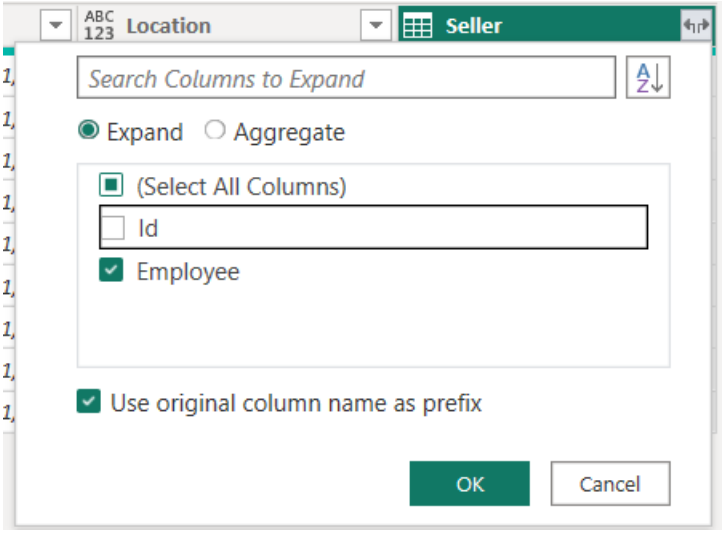
- **Left Outer** - Displays all rows from the first table and only the matching rows from the second.
- **Full Outer** - Displays all rows from both tables.
- **Inner** - Displays the matched rows between the two tables.

In our case, Left Outer is appropriate.

5. Click OK

Finally, we need to expand the new column.

6. Click the **Expand** icon and select the **Employee** column



7. Rename the new column to **Employee**

Let's visualise our data.

8. Click Close & Apply
9. Create a clustered column chart
10. From the **Sales** table, add **Location** and **Employee** to X-Axis
11. Add **Sales** to Y-Axis

You should see a chart like this:

11. Find data anomalies and data statistics

The Power Query editor provides some neat features to get some quick indications regarding the data quality.

1. Load the Excel Workbook **Global Superstore.xlsx**
2. Select **Orders** and click on **Transform data**
3. Go to the **View** tab
4. Tick **Column quality**, **Column distribution** and **Column profile**

You should see the following: