**LAB MANUAL**

**Chapter 6 | ARTIFICIAL INTELLEGENCE**

**Lab – 25**



A close up of a sign

Description automatically generated

**Power Query for Data Transformation**

**Power BI installation**

**Objective:**

* To Consolidate Data: Combine sales data from multiple sources into a single, cohesive dataset.
* To Clean Data: Remove duplicates, handle missing values, and correct errors in the data.
* To Transform Data: Reshape the data to fit the analysis requirements.

**Equipment Required:**

The following list provides the minimum requirements to run Power BI Desktop:

**Important**

* Power BI Desktop is no longer supported on Windows 7.
* Windows 8.1 or Windows Server 2012 R2 or later.
* .NET 4.7.2 or later.
* Microsoft Edge browser (Internet Explorer is no longer supported)
* Memory (RAM): At least 2 GB available, 4 GB or more recommended.
* Display: At least 1440x900 or 1600x900 (16:9) required. Lower resolutions such as 1024x768 or 1280x800 aren't supported because some controls (such as closing the startup screens) display beyond those resolutions.
* Windows display settings: If you set your display to change the size of text, apps, and other items to more than 100%, you won't see some dialogs that you must interact with to continue using Power BI Desktop. If you encounter this issue, check your display settings in Windows by going to Settings > System > Display, and use the slider to return display settings to 100%.
* CPU: 1 gigahertz (GHz) 64-bit (x64) processor or better recommended.
* WebView2: If WebView2 wasn't automatically installed with Power BI Desktop or if it was uninstalled, [download and run the installer for WebView2](https://go.microsoft.com/fwlink/p/?LinkId=2124703).

**Note**

We recommend using a client version of Windows, such as Windows 10, instead of Windows Server. Power BI Desktop doesn't support Internet Explorer Enhanced Security Configuration because it will stop Power BI Desktop from signing in to the Power BI service.

**Prerequisites:**

1. Internet skill to access the Power BI site
2. Computer OS skill to install the software

**Problem Statement:**

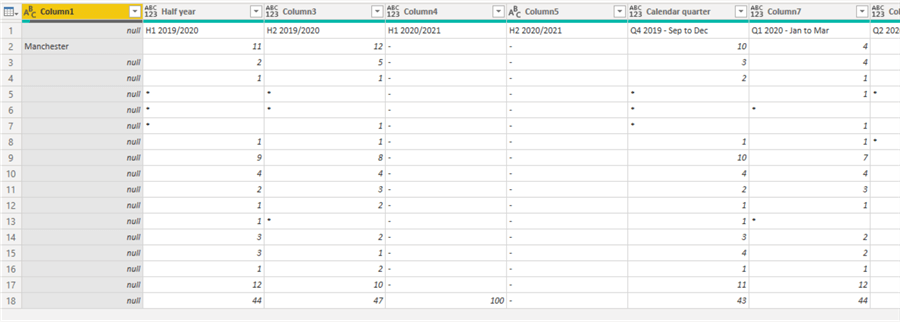
Sometimes we get data from data sources (SQL Server, Excel, CSV file, OData, etc.) in formats which are not in a format that can be easily transformed using simple methods in Microsoft Power BI. This might be due to the complexity and structure of the dataset. Thus, it becomes a hectic task to write simple DAX calculations from the datasets or even use it in any form in Power BI development.

**Solution:**

To demonstrate this process, we are going to transform the dataset as shown in the worksheet below from the current format into an easy to read format which would then make writing calculations and performing data analysis on it easier.

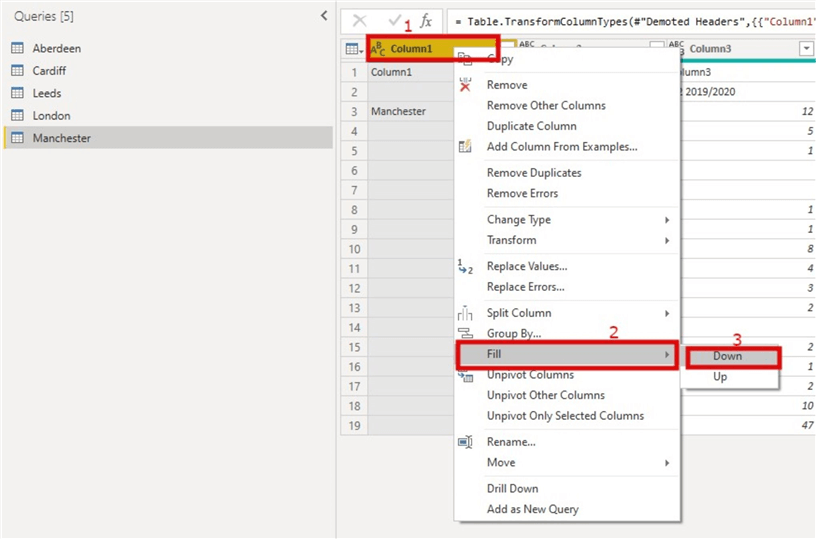
**STEP 1: Import the datasets into Power BI**

Once data importing is complete, you should be able to see the dataset like the screenshot below.

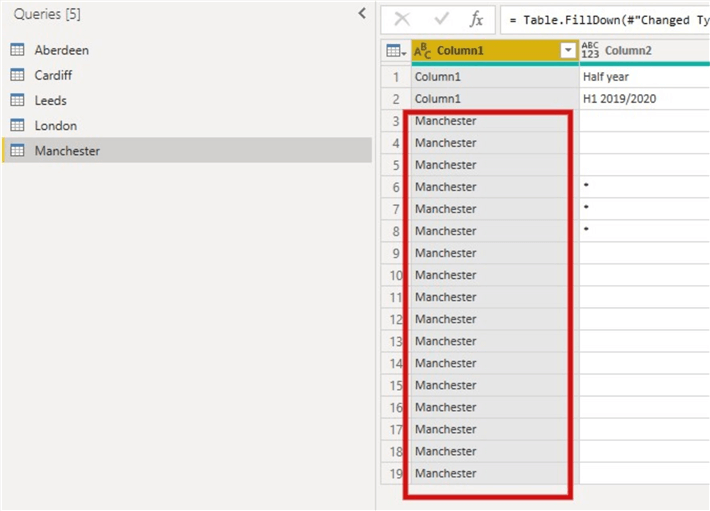


**STEP 2: Fill Down the City names**

As can be seen in the diagram above, the city name has a lot of "nulls" which would make the table a bit messy later, so we need to do a fill down to exclude the nulls. To do this, we need to right click on the "Column1" and then click on "Fill" and then "Down" as shown in the diagram below.



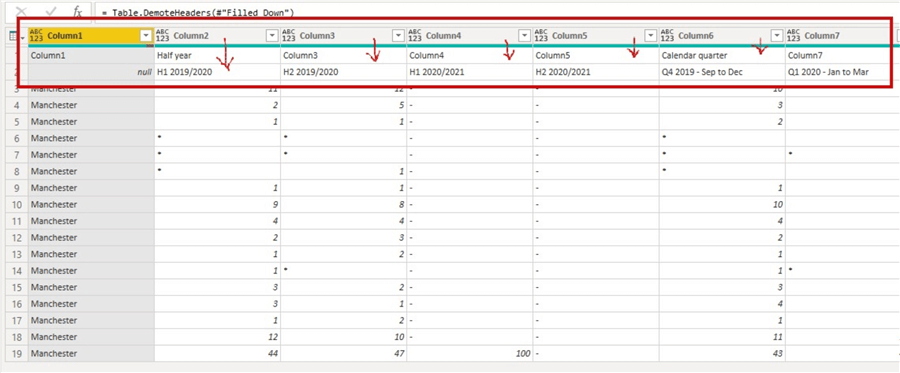
This would ensure each row with "nulls" under the city name is populated with the city name.



**STEP 3: Move headers into first rows**

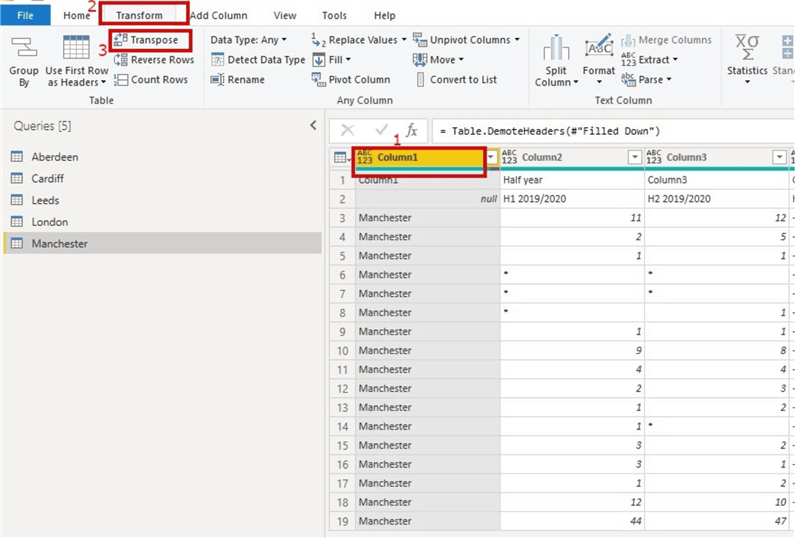
As you can see in the dataset, we have two rows that contain possible headers, we need to ensure this is in a less complex format by moving the "Half Year" and "Calendar Quarter" into columns. We also need to ensure the first-row values are in Columns too.

To do this, we first need to move the headers into first rows as seen in the diagram below. Within the "Home" ribbon click on the drop down on "Use First Rows as Headers" and select "Use Headers as First Row"

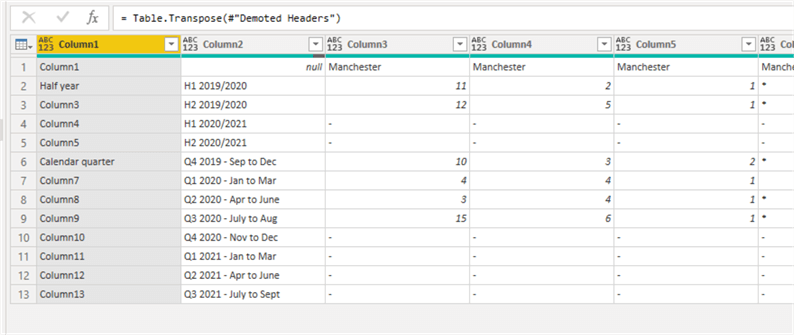


**STEP 4: Transpose the table**

To do this click on "Column1", then click on the "Transform" tab, then click on "Transpose" as seen in the diagram below.



You should now be able to see the table part as seen in the snapshot below.

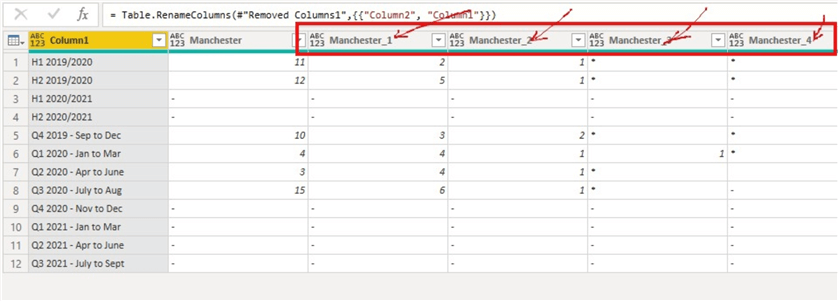


And since we may not need "Column1" after all, we remove columns at this point.

**STEP 5: Promote first rows as headers**

Next, we need to promote the first rows as headers as seen in the diagram below.

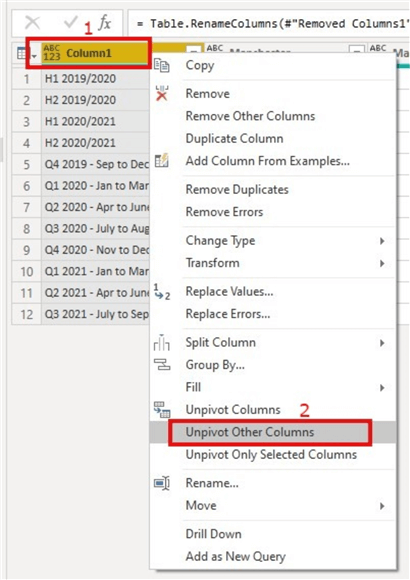
To do this, we need go to the "Home" tab ribbon and select "Use First rows as Headers".



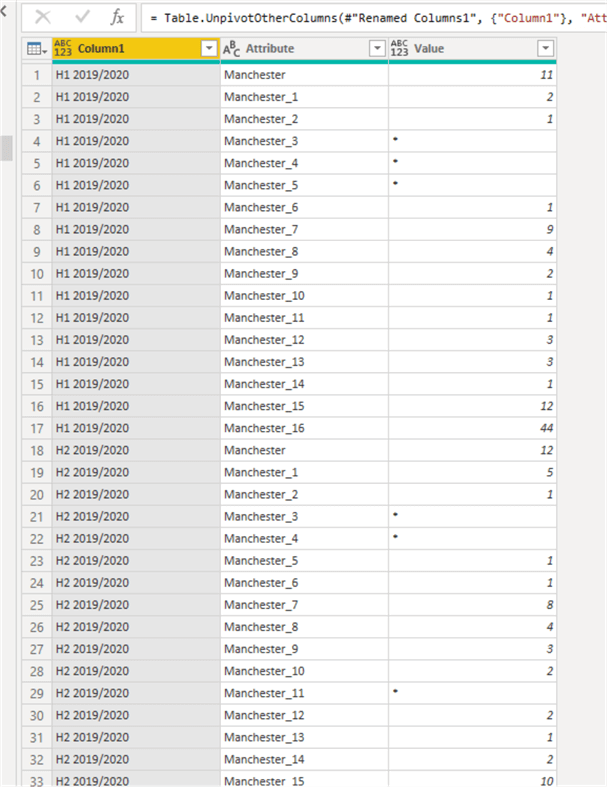
As you can see, there is some of the column headers city names having an attached underscore and a number. We will deal with this later.

**STEP 6: Unpivot columns**

To do this, we need to right click on "Column1" and then select the "Unpivot Other Columns" as seen in the diagram below. Note that this column doesn’t contain numerical values.



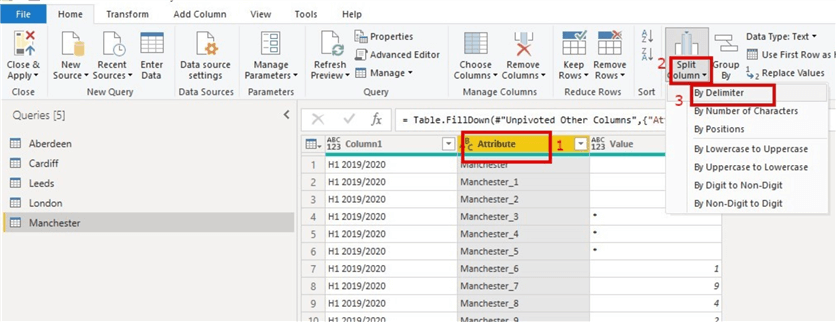
This should transform the table into a more simplified three-column table as seen in the diagram below.



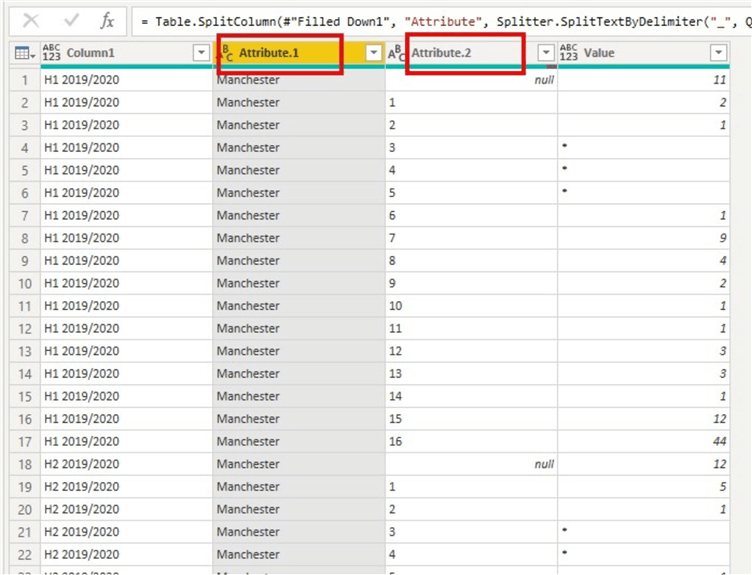
**STEP 7: Cleanse the City Name Column**

The column holding city names is currently having some values concatenated with underscore and a number as seen in the "Attribute" column above. We need to eliminate the underscore and number from the City Name, just leaving "Manchester" alone in this case.

To do this, we need to click on the "Attribute" column, then on the "Home" tab click on the dropdown on "Split Column", then select "By Delimiter" as shown in the diagram below.

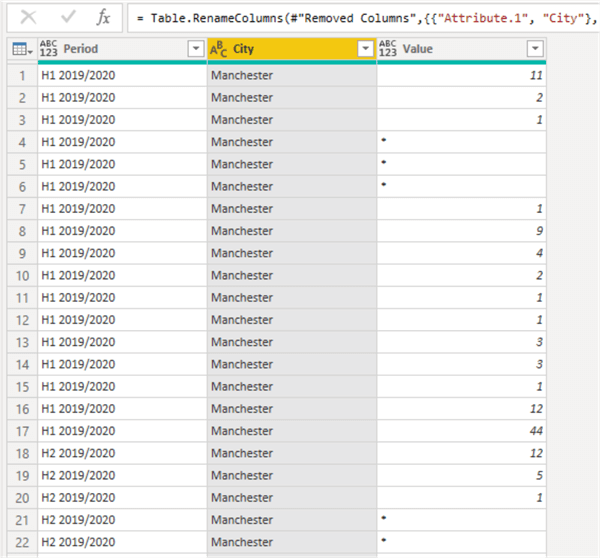


Once this is done, "Attribute" column would split into two columns "Attribute.1" and "Attribute.2" as shown in the diagram below, we then need to delete "Attribute.2" as we don’t need it.



**STEP 8: Rename columns appropriately**

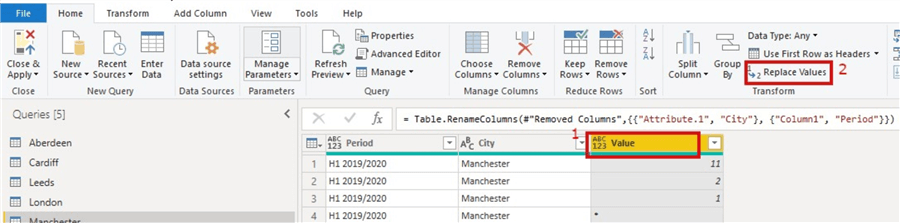
As you would observe the table has column names like "Column1" and "Attribute" which are not appropriate column names. So, we would rename "Column1" as "Period" and "Attribute.1" as "City" as shown in the diagram below.



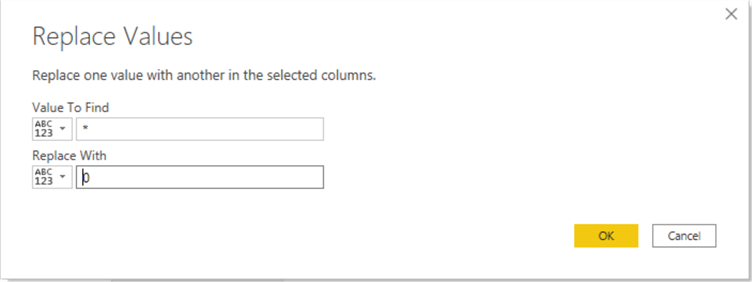
**STEP 9: Replace any special character values within the "Values" column with "0"**

As can be seen in the "Values" column, we have some special characters like "\*" and "- "both of which will throw errors on the cell when we try to correct the datatype of the "Values" column. To ensure this is corrected, we need to replace these special character values with a numerical value such as zero ("0").

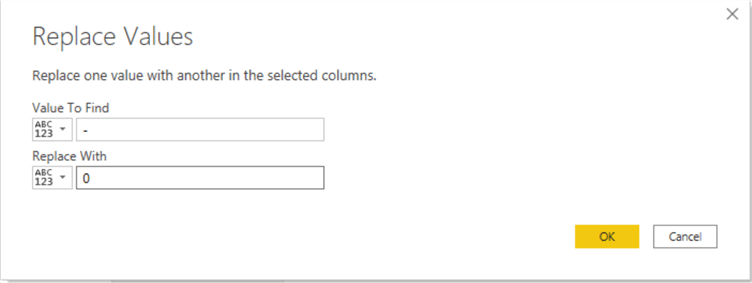
To do this, we click on the "Values" column, then within the "Home" tab and the transform section select "Replace Values" as seen in the diagram below.



The below dialog box will open, then just enter the values as seen in the diagram below.



Repeat same actions for replacing "- "with "0" as seen in the diagram below.



Then after these we can now change the datatypes of the columns appropriately too. The product of the first transformation is as seen below.

