Practical 7

Gather and transform data from multiple sources

Objective: 1. To gather the data from different sources

2. To transform the gather data

Theory:

Power Query can be of great help when you want to combine multiple workbooks into one single workbook.

For example, suppose you have the sales data for different regions (East, West, North, and South). You can combine this data from different workbooks into a single worksheet using Power Query.

If you have these workbooks in different locations/folders, it's a good idea to move all these into a single folder (or create a copy and put that workbook copy in the same folder).

So to begin with, I have four workbooks in a folder (as shown below).
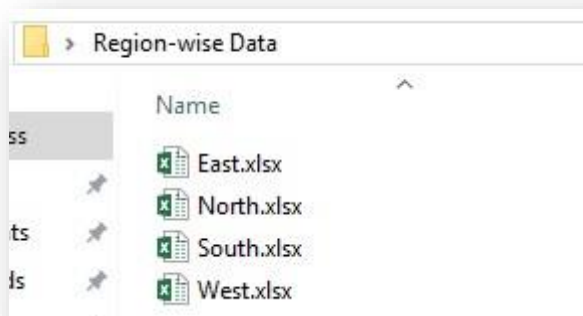
Now, in this tutorial, I am covering three scenarios where you can combine the data from different workbooks using Power Query:

- Each workbook has the data in an [Excel Table](), and all the table names are same.

- Each workbook has the data with the same worksheet name. This can be the case when there is sheet named 'summary' or 'data' in all the workbooks, and you want to combine all these.

- Each workbook has many sheets and tables, and you want to combine specific tables/sheets. This method can also be helpful when you want to combine table/sheets that don't have a consistent name.

Let's see how to combine data from these workbooks in each case.

This Tutorial Covers:

Each workbook has the data in an Excel Table with the same structure

The below technique would work when your Excel Tables has been structured the same way (same column names).

The number of rows in each table can vary.

Don't worry if some of the Excel Tables have additional columns. You can choose one of the Tables as the template (or as the 'key' as Power Query calls it), and Power Query would use it to combine all the other Excel Tables with it.

In case there are additional columns in other tables, those will be ignored and only the ones specified in the template/key would be combined. For example, if the template/key table that you select has 5 columns, and one of the tables in some other workbook has 2 additional columns, those additional columns would be ignored.

Now I have four workbooks in a folder that I want to combine.

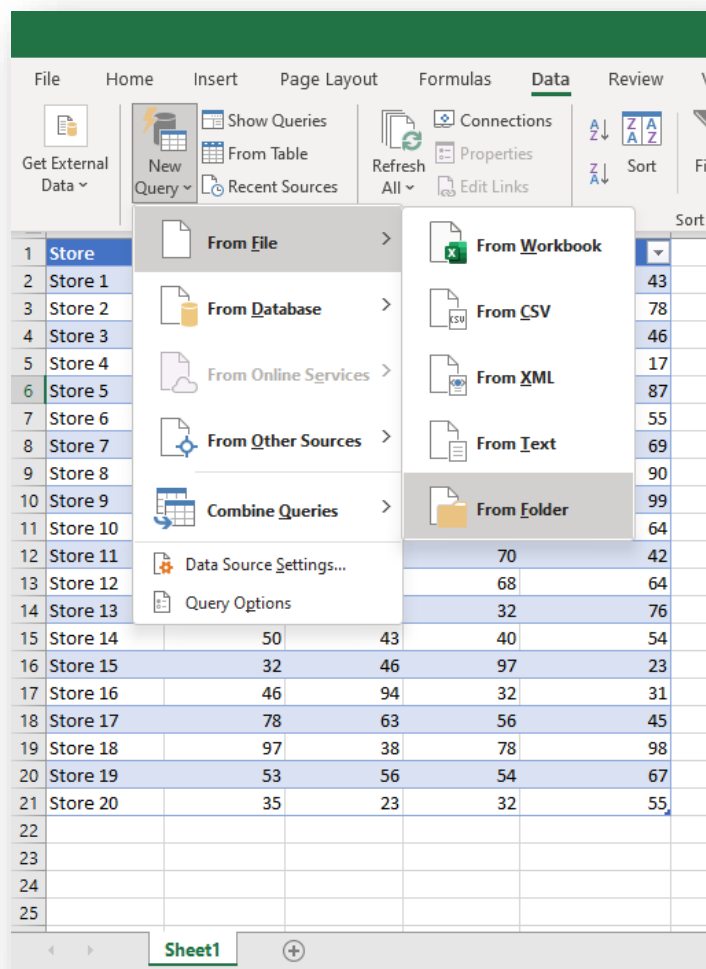Below is a snapshot of the table I have in one of the workbooks.

Here are the steps to combine the data from these workbooks into a single workbook (as a single table).
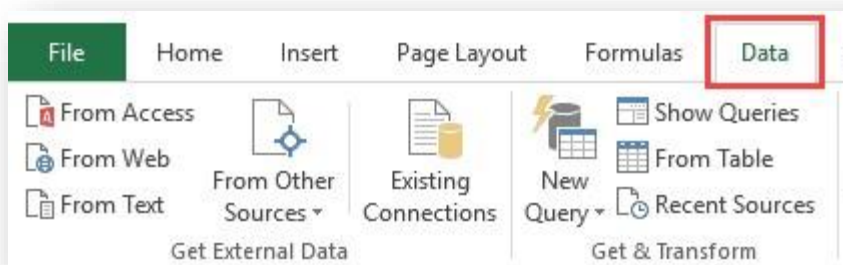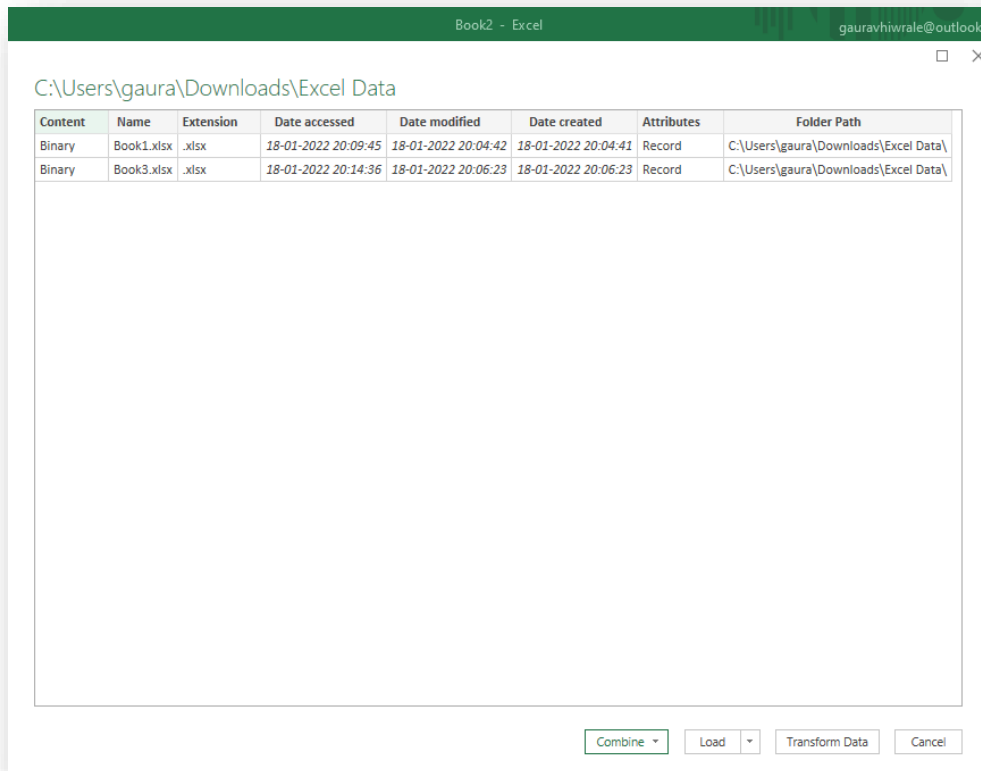

1. Go to the Data Tab.
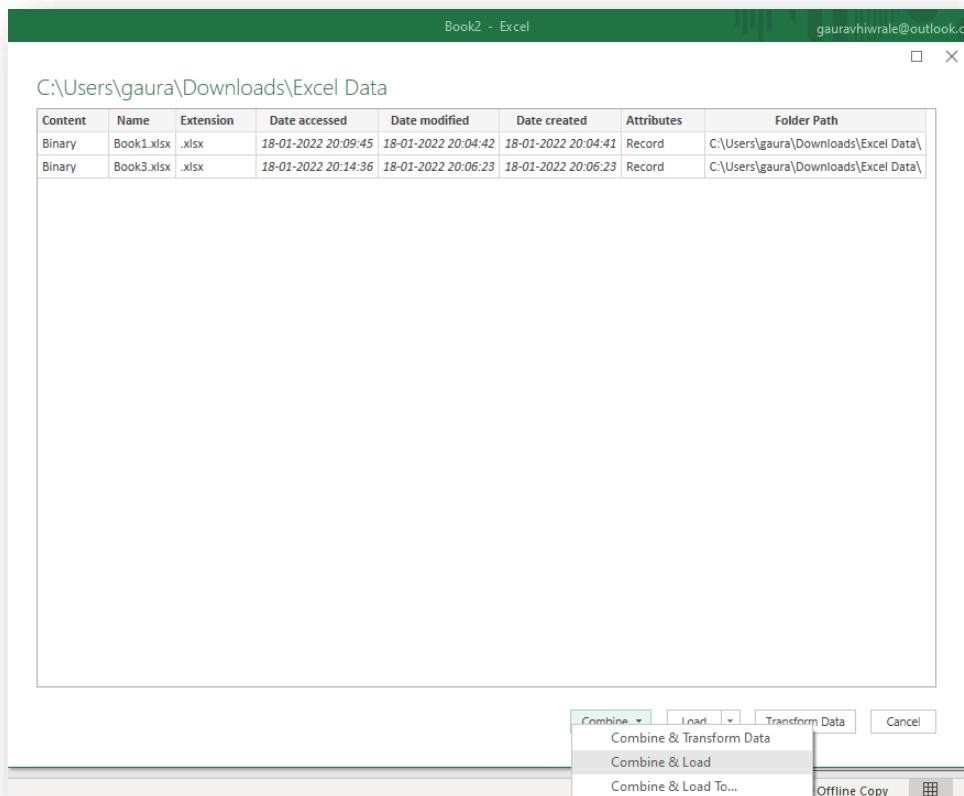2. In the Get & Transform group, click on the New Query drop down.

3. Hover your cursor on 'From File' and click on 'From Folder'.

4. In the Folder dialog box, enter the file path of the folder that has the files, or click on Browse and locate the folder.
5. Click OK.
6. In the dialog box that opens, click on the combine button.

7. Click on 'Combine & Loa.

In the 'Combine Files' dialog box that opens, select the Table in the left pane. Note that Power Query shows you the Table from the first file. This file would act as the template (or the key) to combine other files. Power Query would now look for 'Table 1' in other Workbooks.

8. And combine it with this one.
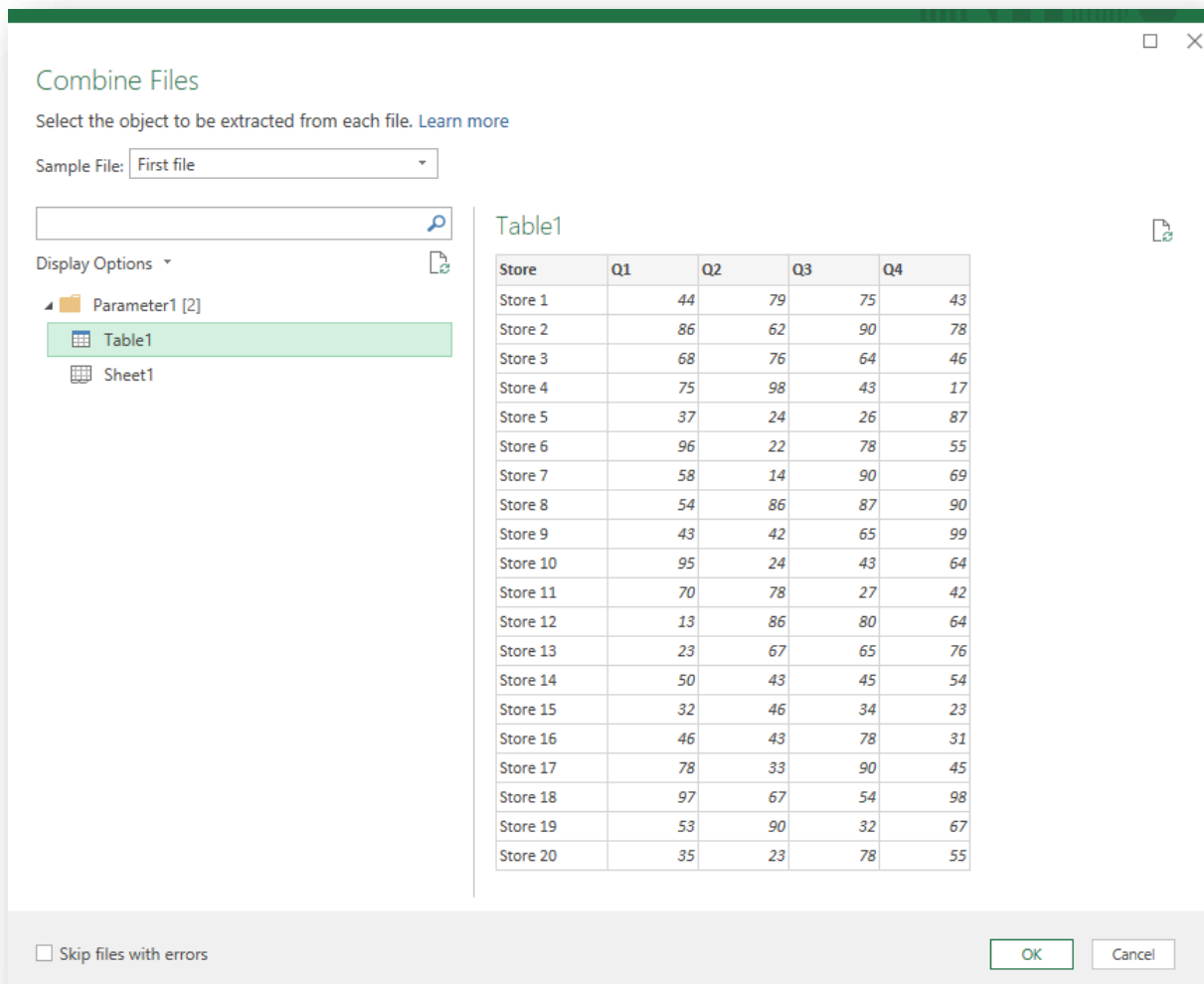
9. Click OK.

This will load the final result (combined data) into your active worksheet.

Note that along with the data, Power Query automatically adds the workbook name as the first column of the combined data. This helps in keeping track of what data came from which workbook.

In case you want to first Edit the data before loading it into Excel, in Step 6, select 'Combine and Edit'. This will open the final result in the Power Query editor where you can edit the data.
A few things to know:

- If you select an Excel Table as the template (in Step 7), Power Query will use the column names in this Table to combine the data from other Tables. If other Tables have additional columns, those will be ignored. In case those other Tables don't have a column, which is there in your Template Table, Power Query would just put 'null' for it.

- The columns don't need to be in the same order as Power Query uses column headers to map columns.

- Since you have selected Table1 as the key, Power Query will look for Table1 in all the workbooks, and combine all these. In case it doesn't find an Excel Table with the same name (Table1 in this example), Power Query will give you an error.

Adding New Files to the Folder

Now let's take a minute and understand what we did with the above steps (which only took us a few seconds).

We combined the data from four different workbooks in one single table in a few seconds without even opening any of the workbooks.

But that's not all.

The real POWER of Power Query is that now when you add more files to the folder, you don't need to repeat any of these steps.

All you need to do move the new workbook in the folder, refresh the query, and it will automatically combine the data from all the workbooks in that folder.
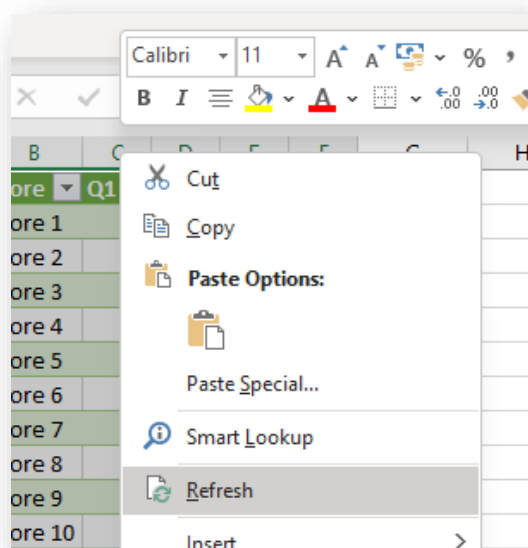
For example, in the above example, if I add a new workbook – *'Mid-West.xlsx'* to the folder, and refresh the query, it will instantly give me the new combined dataset.
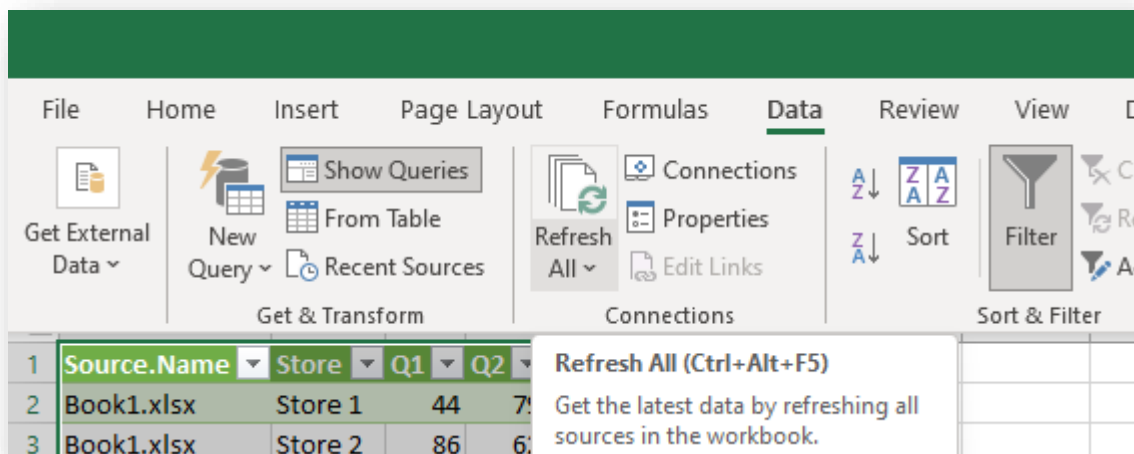
Here is how you refresh a query:

- Right-click on the Excel Table that you loaded in the worksheet and click Refresh.
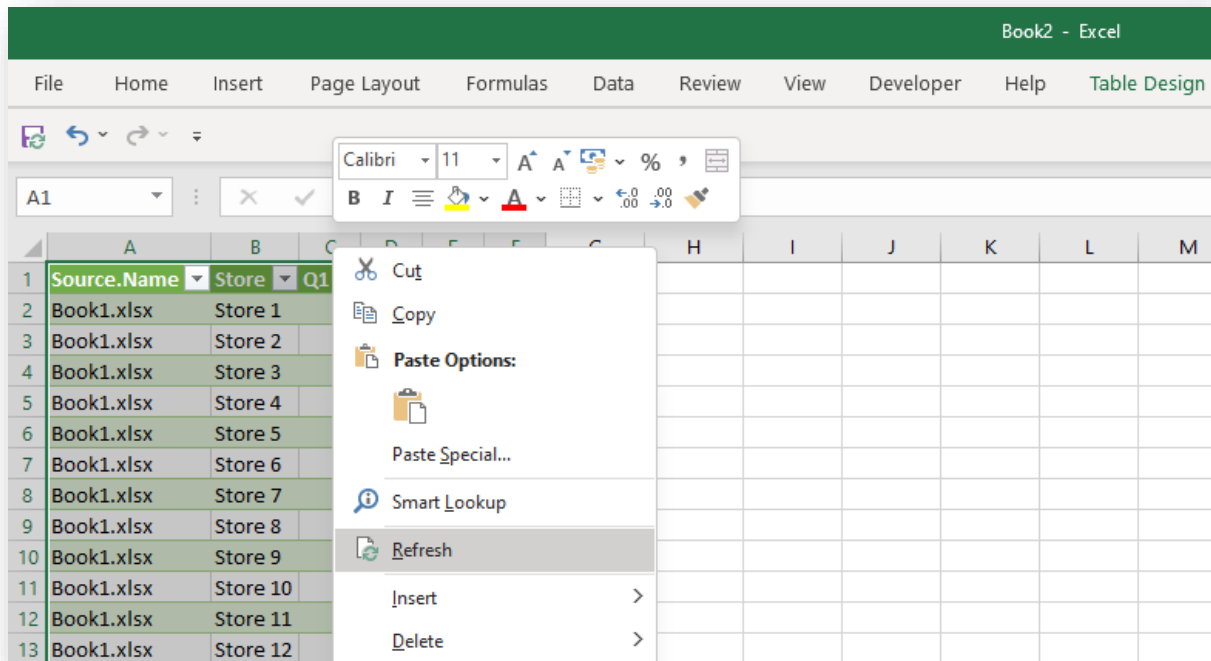
- Right-click on the Query in the 'Workbook Query' pane and click Refresh

- Go to the Data tab and click on Refresh.

## Each workbook has the data with the same worksheet name

In case you don't have the data in an Excel Table, but all the sheet names (from which you want to combine the data) are the same, then you can use the method shown in this section.

There are a few things you need to be cautious about when it's just tabular data and not an Excel Table.

- The worksheet names should be the same. This will help Power Query to go through your workbooks and combine the data from the worksheets that have the same name in each workbook.

- Power Query is case sensitive. This means a worksheet named 'data' and 'Data' are considered different. Similarly, a column with the header 'Store' and one with 'store' are considered different.

- While it's important to have the same column headers, it's not important to have the same order. If column 2 in the 'East.xlsx' is column 4 in 'West.xlsx', Power Query will match it correctly by mapping the headers.

Now let's see how to quickly combine data from different workbooks where the worksheet name is the same.

In this example, I have a folder with four files.
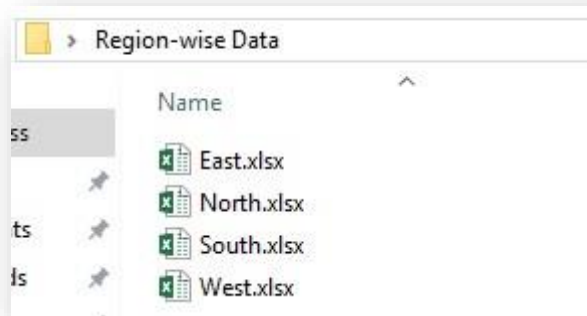
In each workbook, I have a worksheet with the name 'Data' that contains the data in the following format (note that this is not an Excel Table).

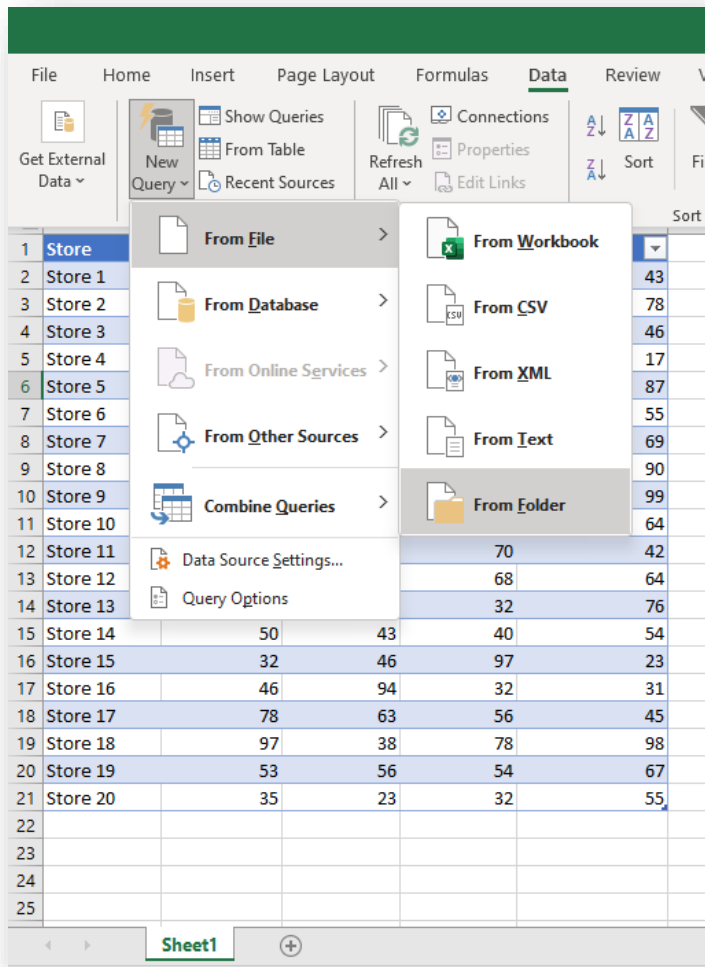|	Store	Q1	Q2	Q3	Q4	
2	Store 1	44	79	75	43	
3	Store 2	86	62	90	78	
4	Store 3	68	76	64	46	
5	Store 4	75	98	43	17	
6	Store 5	37	24	26	87	
7	Store 6	96	22	78	55	
8	Store 7	58	14	90	69	
9	Store 8	54	86	87	90	
10	Store 9	43	42	65	99	
11	Store 10	95	24	43	64	
12	Store 11	70	78	27	42	
13	Store 12	13	86	80	64	
14	Store 13	23	67	65	76	
15	Store 14	50	43	45	54	
16	Store 15	32	46	34	23	
17	Store 16	46	43	78	31	
18	Store 17	78	33	90	45	
19	Store 18	97	67	54	98	
20	Store 19	53	90	32	67	
21	Store 20	35	23	78	55	

Here are the steps to combine data from multiple workbooks into one single worksheet:
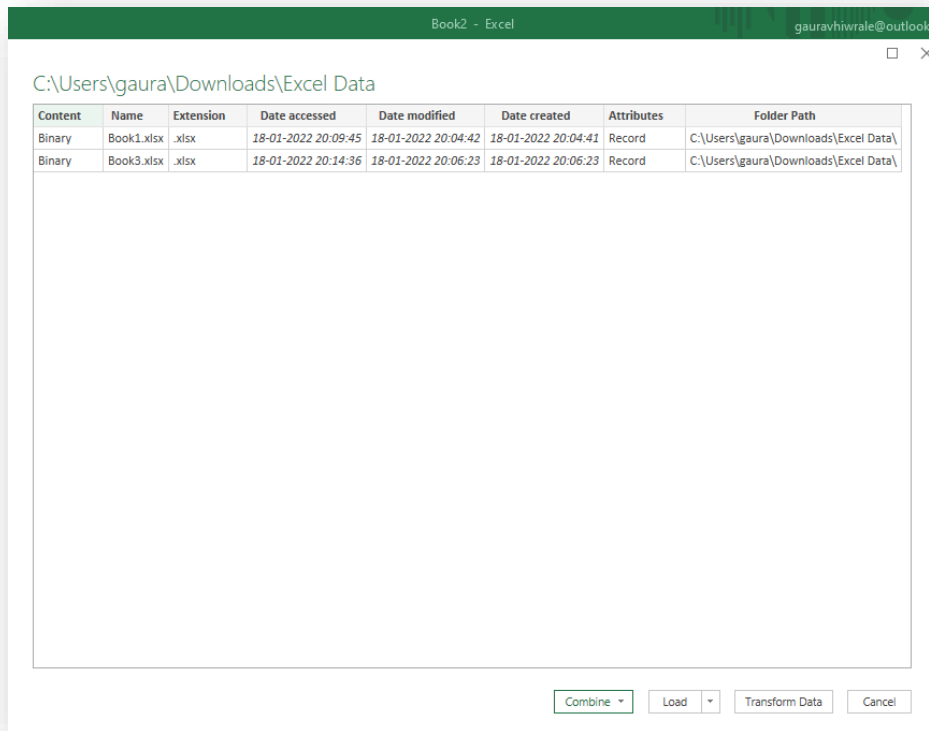
1. Go to the Data tab
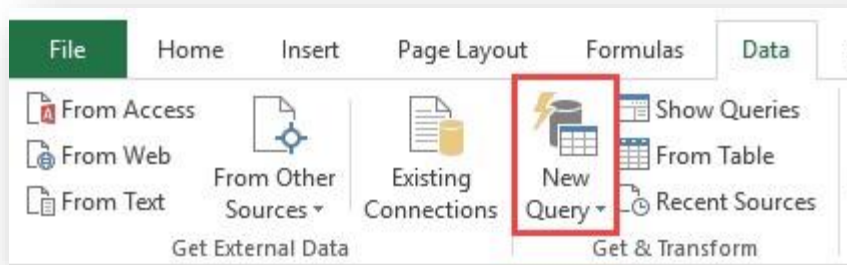
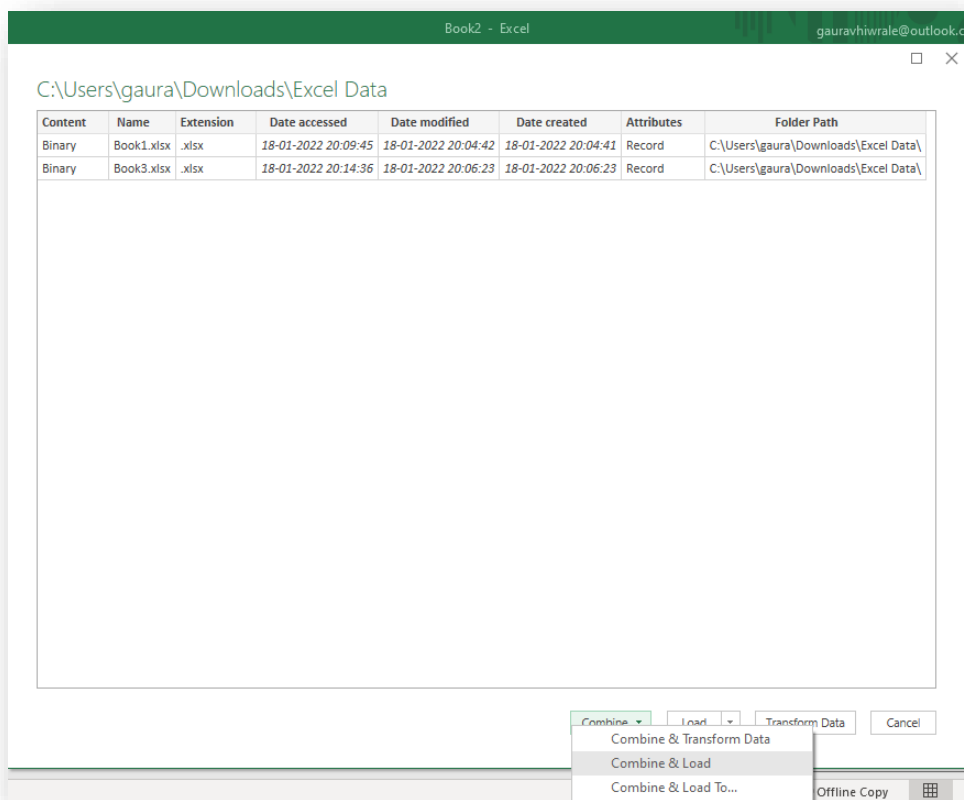2. In the Get & Transform group, click on the New Query drop down.

3. Hover your cursor on 'From File' and click on 'From Folder'.

4. In the Folder dialog box, enter the file path of the folder that has the files, or click on Browse and locate the folder.
5. Click OK.
6. In the dialog box that opens, click on the combine button.

C:\Users\gaura\Downloads\Excel Data

Content	Name	Extension	Date accessed	Date modified	Date created	Attributes	Folder Path
Binary	Book1.xlsx	.xlsx	18-01-2022 20:09:45	18-01-2022 20:04:42	18-01-2022 20:04:41	Record	C:\Users\gaura\Downloads\Excel Data\
Binary	Book3.xlsx	.xlsx	18-01-2022 20:14:36	18-01-2022 20:06:23	18-01-2022 20:06:23	Record	C:\Users\gaura\Downloads\Excel Data\

Combine ▾   Load ▾   Transform Data   Cancel

7. Click on 'Combine & Load'.

C:\Users\gaura\Downloads\Excel Data

Content	Name	Extension	Date accessed	Date modified	Date created	Attributes	Folder Path
Binary	Book1.xlsx	.xlsx	18-01-2022 20:09:45	18-01-2022 20:04:42	18-01-2022 20:04:41	Record	C:\Users\gaura\Downloads\Excel Data\
Binary	Book3.xlsx	.xlsx	18-01-2022 20:14:36	18-01-2022 20:06:23	18-01-2022 20:06:23	Record	C:\Users\gaura\Downloads\Excel Data\

Combine ▾   Load ▾   Transform Data   Cancel

Combine & Transform Data
Combine & Load
Combine & Load To...

Offline Copy

8. In the 'Combine Files' dialog box that opens, select 'Data' in the left pane. Note that Power Query shows you the worksheet name from the first file. This file would act as the key/template to combine other files. Power Query will go through each workbook, find the sheet named 'Data, and combine all these.
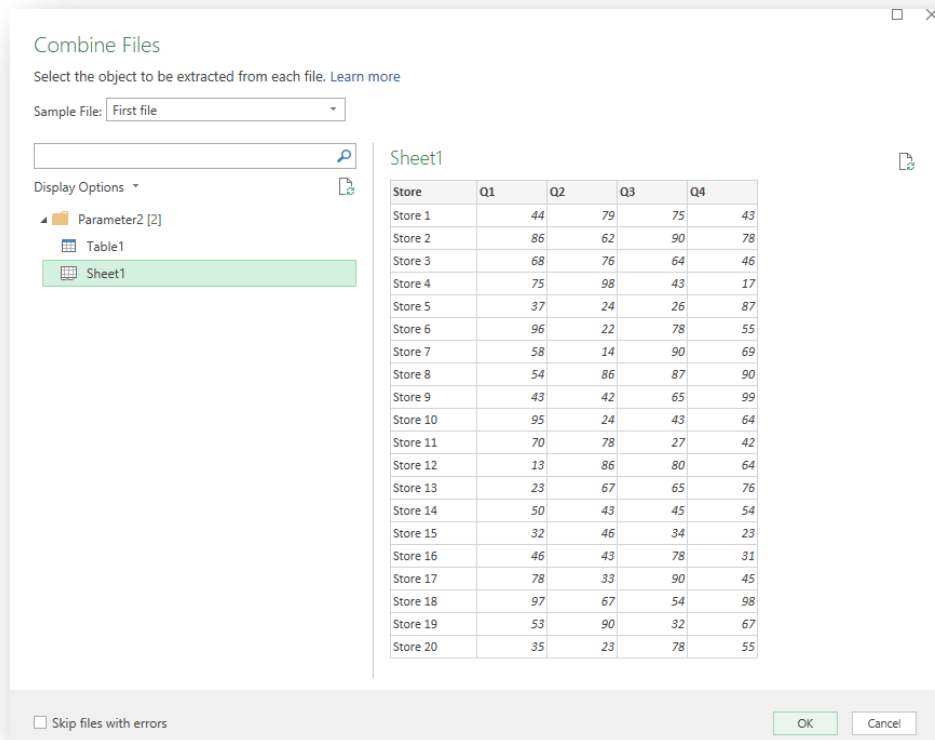
9. Click OK. Now Power Query will go through each workbook, look for the worksheet named 'Sheet' in it, and then combine all these datasets.

This will load the final result (combined data) into your active worksheet.

In case you want to first Edit the data before loading it into Excel, in Step 6, select 'Combine and Edit'. This will open the final result in the Power Query editor where you can edit the data.

Each Workbook has the data with Different Table names or Sheet Names

Sometimes, you may not get structured and consistent data (such as Tables with same name or worksheet with the same name).

For example, suppose you get the data from someone who created these datasets but named the worksheets as East Data, West Data, North Data, and South Data.

Or, the person may have created Excel tables, but with different names.

In such cases, you can still use Power Query, but you need to do it with a couple of additional steps.
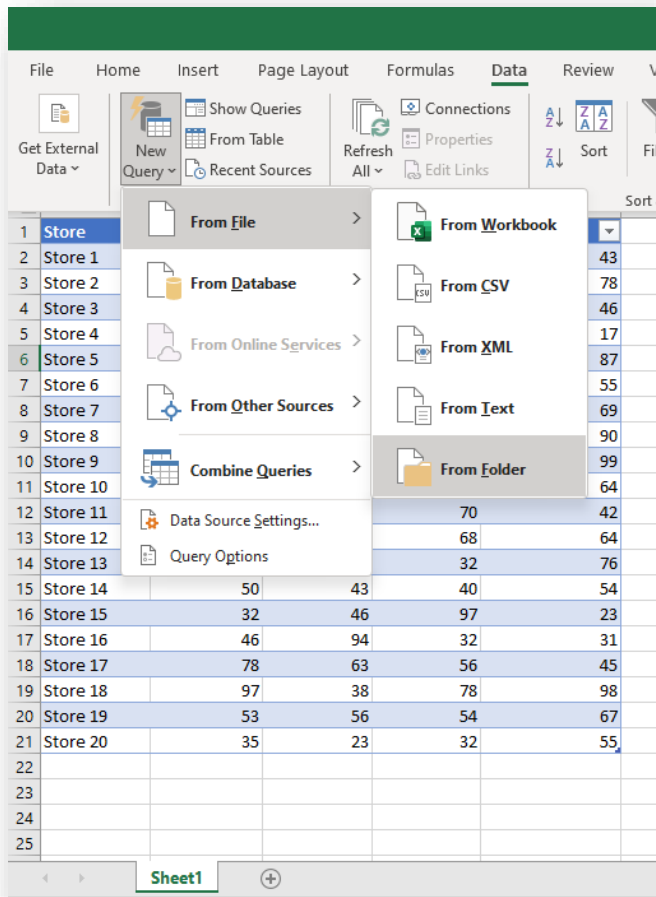
1. Go to Data tab.
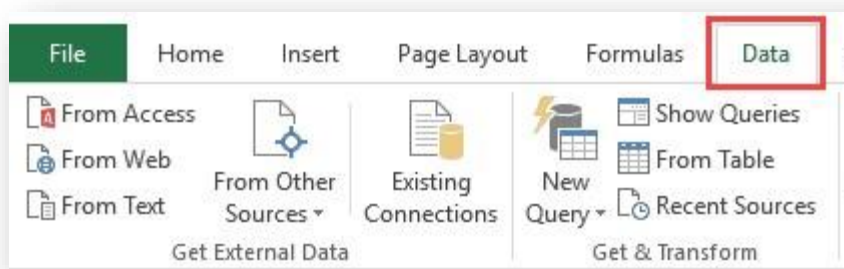
2. In Get & Transform group, click on the New Query drop down.
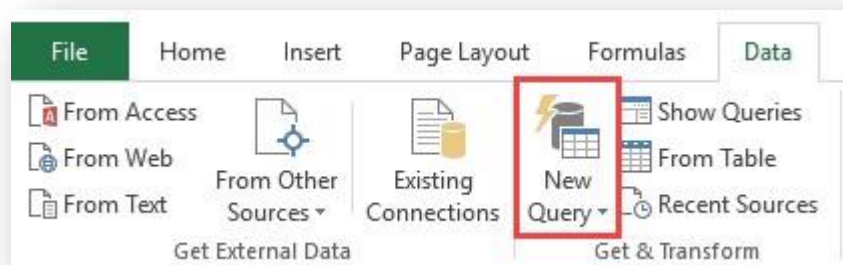

3. Hover your cursor on 'From File' and click on 'From Folder'.

4. In the Folder dialog box, enter the file path of the folder that has the files, or click on Browse and locate the folder.
5. Click OK.
6. In the dialog box that opens, click on the Edit button. This will open the Power Query editor where you will see the details of all the files in the folder.
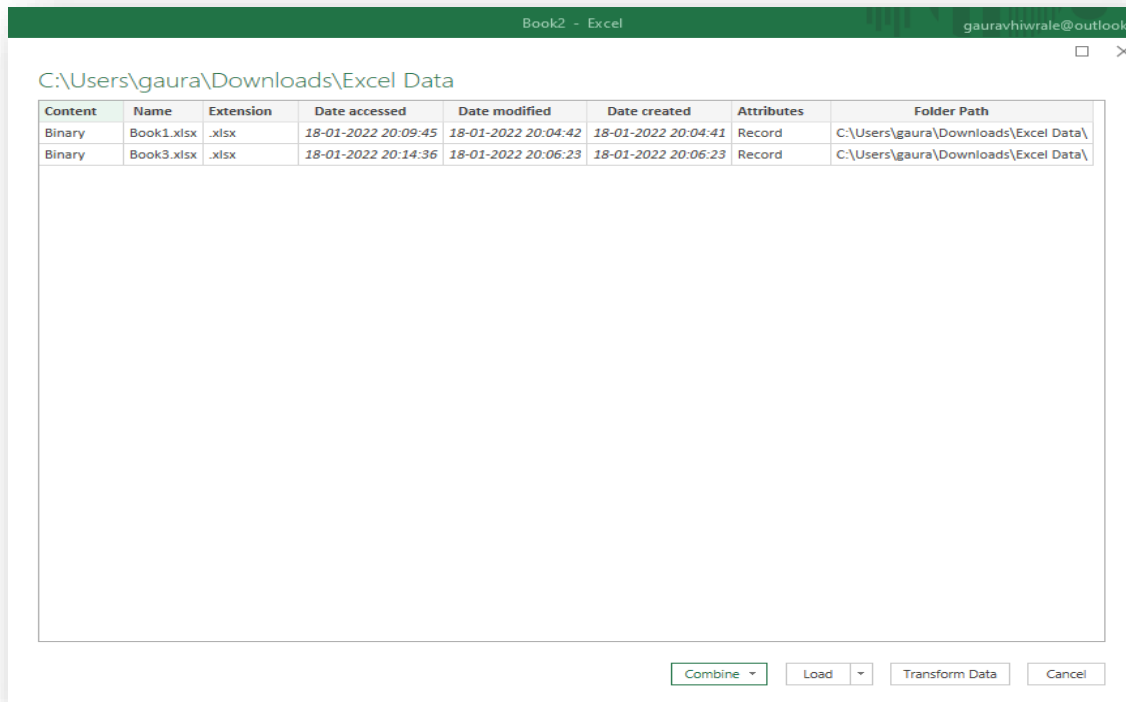
7. Hold the Control key and select the 'Content' and 'Name' columns, right-click and select 'Remove Other Columns'. This will remove all the other columns except the selected columns.
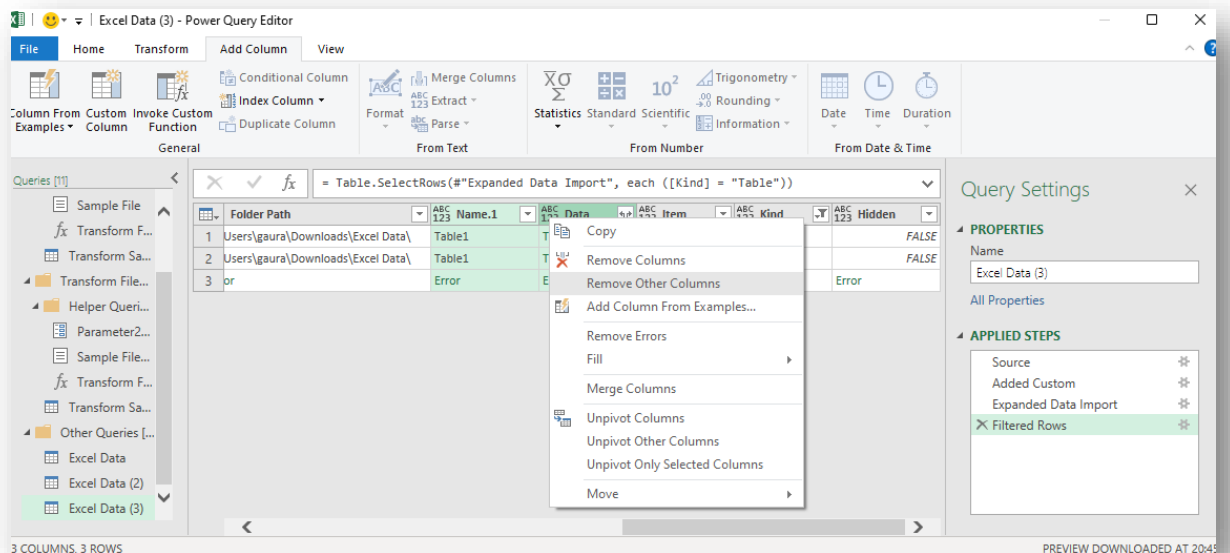
8. In the Query Editor ribbon, click on 'Add column' and then click on 'Custom Column'.
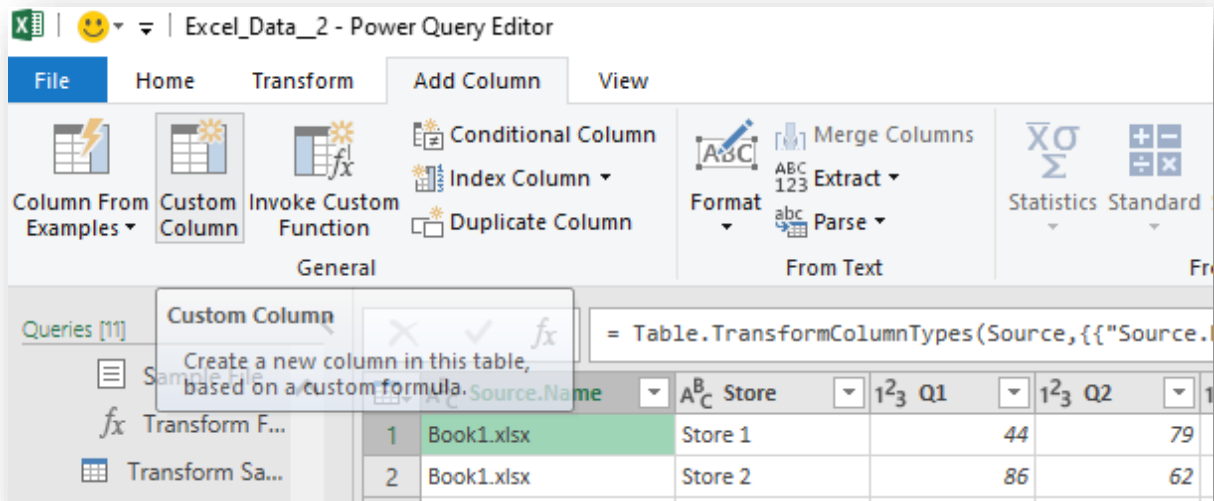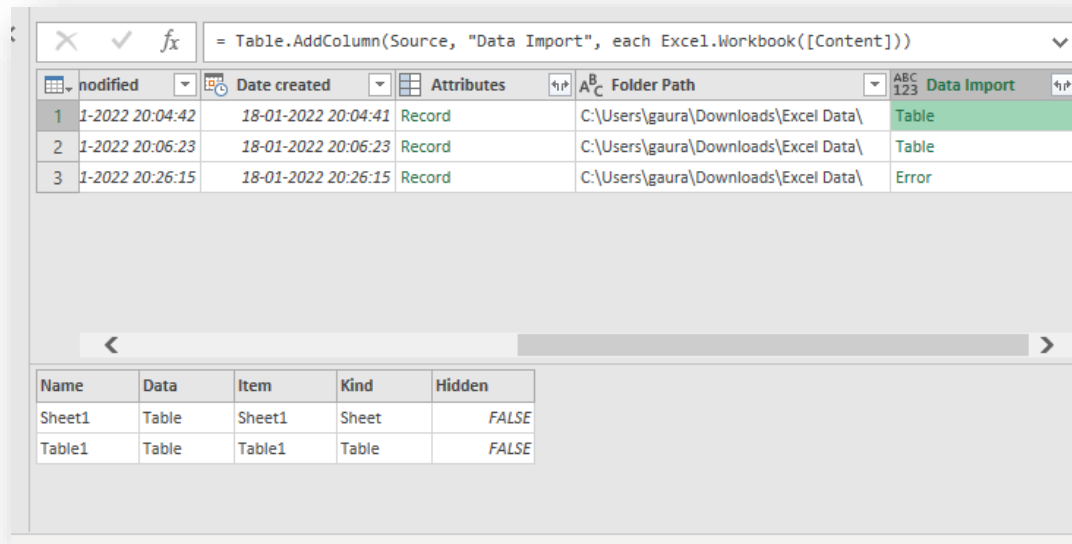
9. In the Add Custom Column dialog box, name the new column as 'Data Import' and use the following formula **=Excel.Workbook([CONTENT])**. Note that this formula is case sensitive and you need to enter it exactly the way I have shown here.
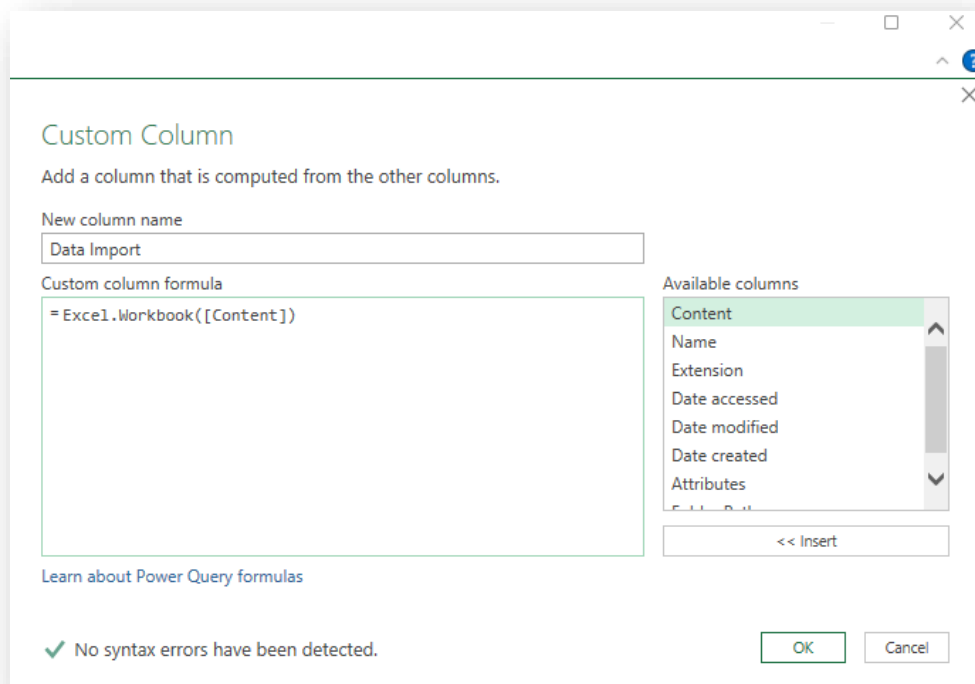
10. Now you will see a new column that has Table written in it. Now let me explain what happened here. You provided Power Query the names of the workbooks, and Power Query has fetched the objects such as worksheets, tables, and
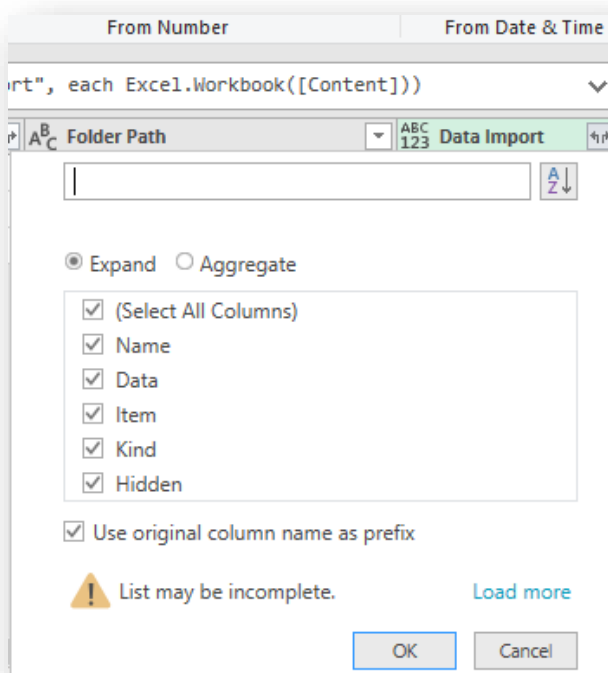
named ranges from each workbook (which resides in the Table cell as of now). You can click on the white space next to the text Table and you would see the information at the bottom. In this case, since we only have one table and one worksheet in each workbook, you can see only two.

11. Click on the double-arrow icon at the top of the 'Data Import' column.
12. In the column data box that opens, uncheck the 'Use original column as prefix', and then click OK.

13. Now you will see an expanded table where you see one row for each object in the table. In this case, for each workbook, the sheet object and the table object are listed separately.
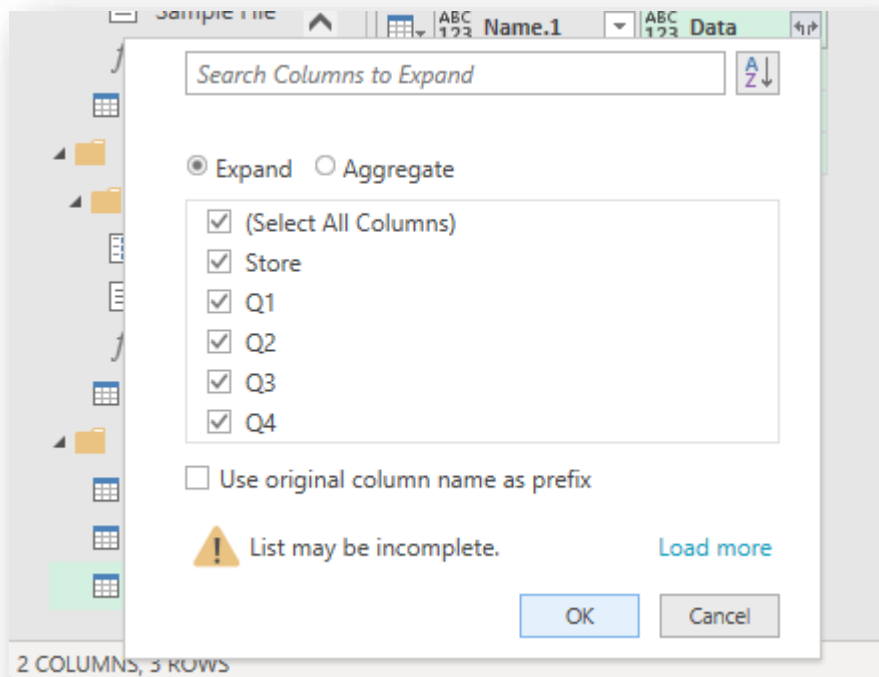14. In the Kind column, filter the list to only show the Table.
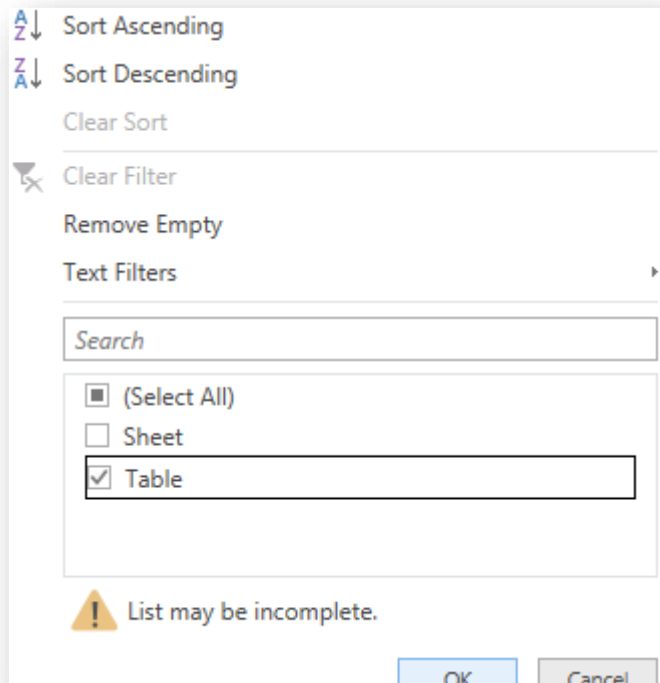
15. Hold the control key and select the Name and Data column. Now, right-click and remove all the other columns.
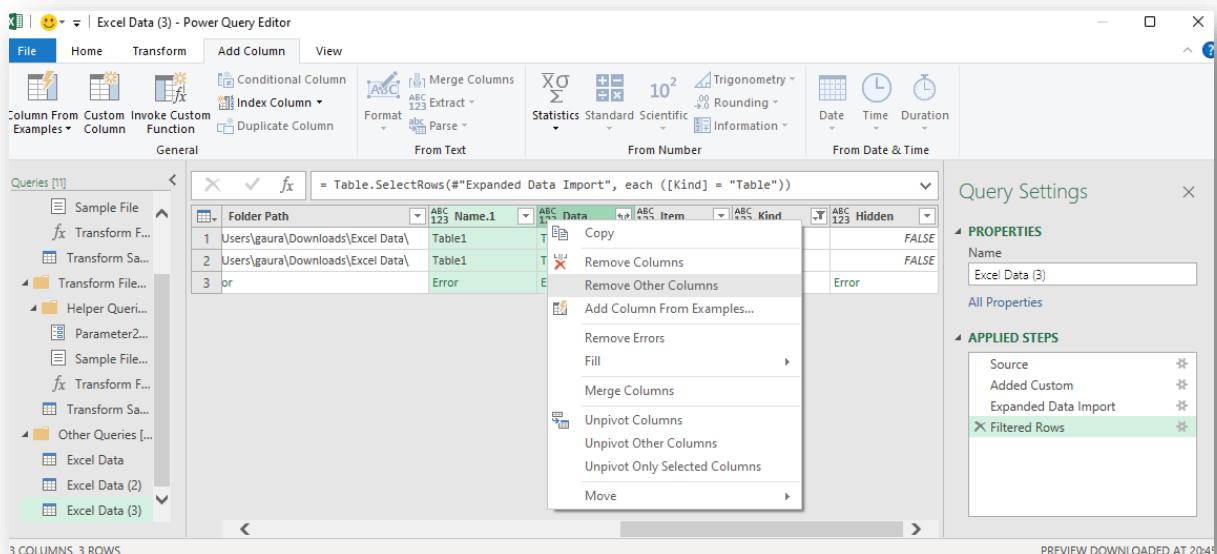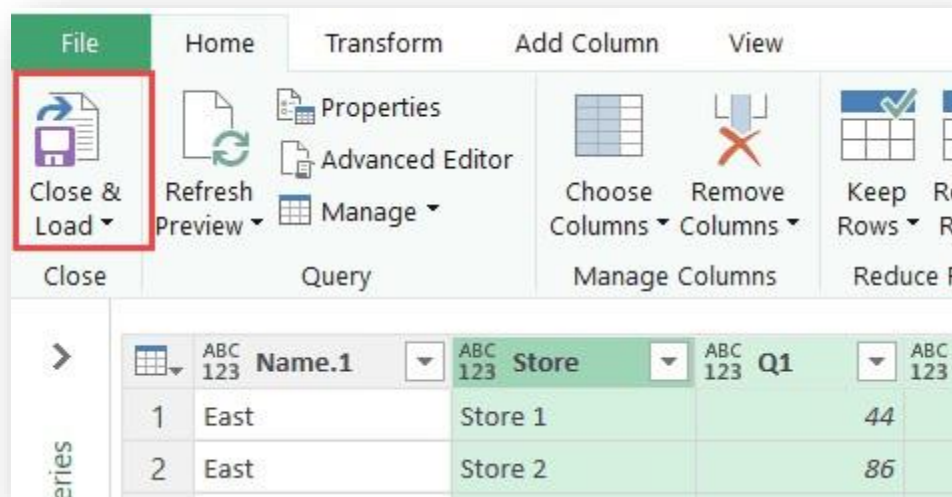

16. In the Data column, click on the double-arrow icon at the top right of the Data.

17. In the column data box that opens, click OK. This will combine the data in all the tables and show in Power Query.

18. Now you can make any transformation you need, and then go to the home tab and then go to home tab and click close & and load.

Conclusion:

In this practical we have learn the concert of gathering and transforming the data from multiple source using power query

Assessment Details:

Date of Lab Conducted	Date of Evaluation	Grade	Sign & Remark (If any)