

# Microsoft Power Tools for Data Analysis #04:

## Power Query: Import Multiple Excel Files & Combine (Append) into Proper Data Set

### Notes from Video:

#### Table of Contents:

1.	Goal of Video.....	2
2.	Main Difficulty When Importing Multiple Excel Files .....	2
4)	Types of objects that we may encounter when we import an Excel Workbook file using Power Query.....	3
3.	Import Multiple Excel Files From Folder .....	4
6)	Click Transform Data!! .....	5
8)	Change Query Options for Data Types.....	6
12)	To filter out non-Excel files .....	7
13)	Clean File Name .....	8
17)	Custom Column using Excel.Workbook Function .....	9
22)	The Expanded Table with each Excel Files objects looks like this.....	11
23)	Headers Have Been Promoted.....	11
24)	Filter to import only Sheet Objects.....	12
25)	Filter to remove Sheet objects that Begin with "Sheet" .....	12
30)	Final Append with all Proper Data Sets.....	13
32)	Load to Excel Sheet .....	14
4.	Change Default PivotTable Layout & Options.....	15
5.	Definition of a PivotTable.....	16
6.	AND Logical Test .....	16
7.	Add New Excel Workbook Files to the Folder & Refresh.....	17
8.	Edit Query when Folder Path Changes .....	17

1. **Goal of Video:**

- 1) Import multiple Excel Workbook files with multiple tables, extract Excel Workbook file and sheet tab names as new columns, append all tables into one proper data set, and build PivotTable report that can be refreshed when new files arrive. Here is a summary picture:

**Import Many Excel Files, Append, Build Report**

The screenshot illustrates the workflow for importing multiple Excel files. On the left, a list of files is shown: Oakland.xlsx, Portland.xlsm, SanFrancisco.xlsm, Seattle.xlsx, and Tacoma.xlsx. These files are imported into a single Excel spreadsheet. The spreadsheet has columns for Date, Product, Units, Sales, City, and SalesRep. Below the spreadsheet, a PivotTable is shown with 'Sum of Sales (\$)' as the value field, 'Product' as the row field, and 'SalesRep' as the column field. The PivotTable shows sales data for various products like Quad, Sunshin, and Tri Fly, broken down by sales representative. To the right of the PivotTable, a PivotTable task pane is visible, showing the list of products and the selected filters.

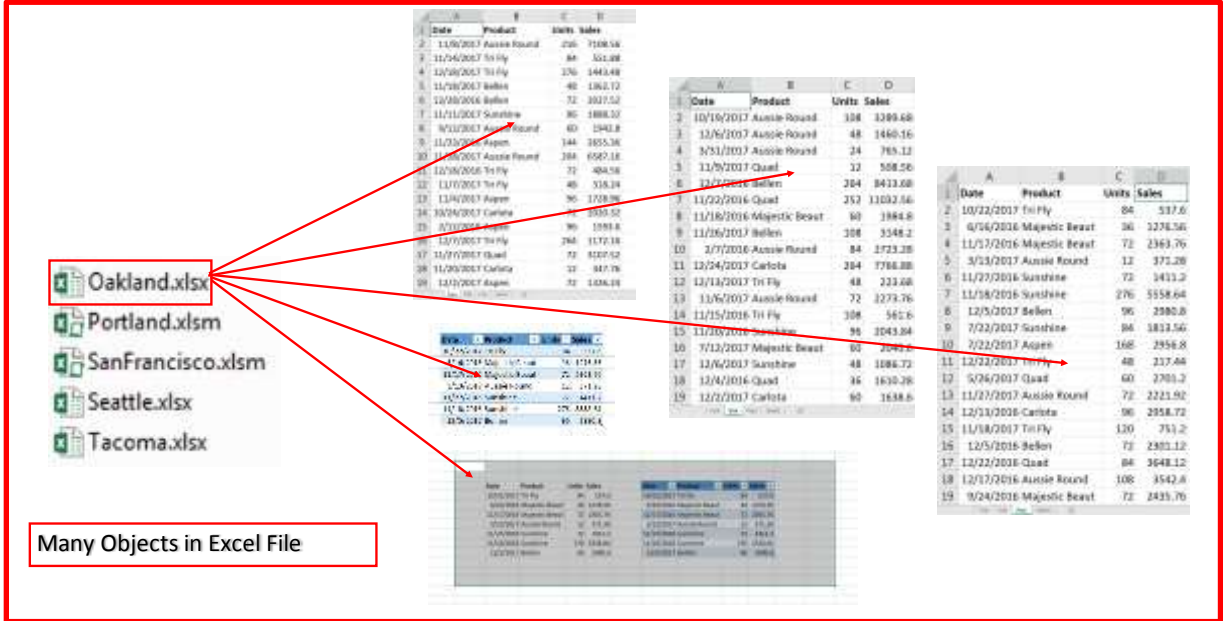
2. **Main Difficulty When Importing Multiple Excel Files** is that each Excel file may have multiple objects that need to be included or excluded in the query.

- 1) In last video, when we tried to import multiple text files, we did not have the issue of multiple objects in a single file because each text file contained only one Proper Data Set, as seen in this picture:

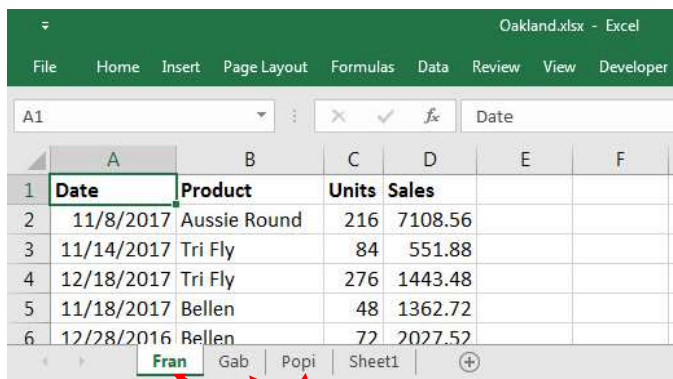
**Just One Object in Text File**

The screenshot shows a list of text files: 2015Sales.txt, 2016Sales.txt, 2017Sales.TXT, 2018Sales.txt, and 2019Sales.txt. A red box highlights the text 'Just One Object in Text File' with an arrow pointing to the first file. To the right, a sample of the text file content is shown, displaying a list of transactions with columns for TransactionID, Date, Website, Product, Quantity, RevenueDiscount, SetStandardCost, and CountryCode. Each line represents a single transaction record.

- 2) When we try to import multiple Excel files, each individual Excel file can have one or more objects in an Excel File. For example, in this picture the Excel file named “Oakland.xlsx” contains many objects such as sheets with Proper Data Sets, Excel Tables and Print Ranges, as seen in this picture:



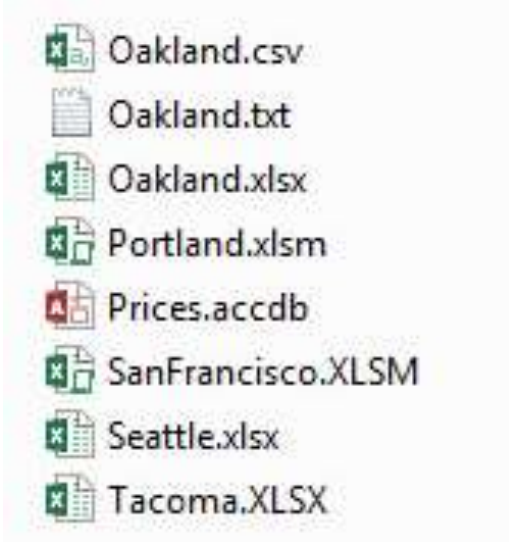
- 3) When we build our query to go and get the data from multiple Excel workbook files, we will have to build the query so that it imports only the objects that contain the data we want.
- 4) **Types of objects that we may encounter when we import an Excel Workbook file using Power Query** are:
- An Excel Worksheet with all its data
  - An Excel Table
  - A Defined Name
  - A Print Range
  - A Table that has had the Filter Feature Used
  - Automatic Defined Name Created when you use the Advanced Filter feature and you use a Criteria Range
  - Automatic Defined Name Created when you use the Advanced Filter feature and you use an Extract Range
- 5) Our Goal is to import only objects that are Sheets and have a SalesRep Sheet Tab Name, as seen in this picture of the “Oakland.xlsx” Workbook File:



Import Only Excel Sheets with a SalesRep Name

### 3. Import Multiple Excel Files From Folder.

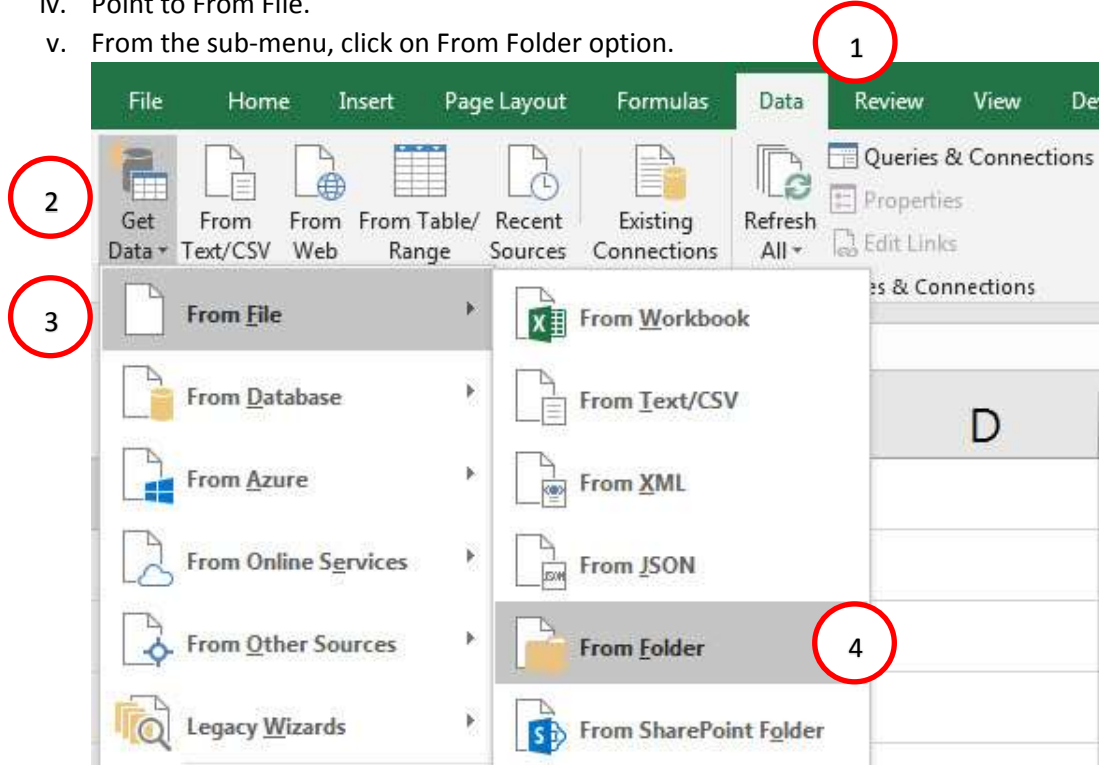
1) For our example these are the files that are sitting in the folder named “004-MSPTDA-ExcelFiles”:



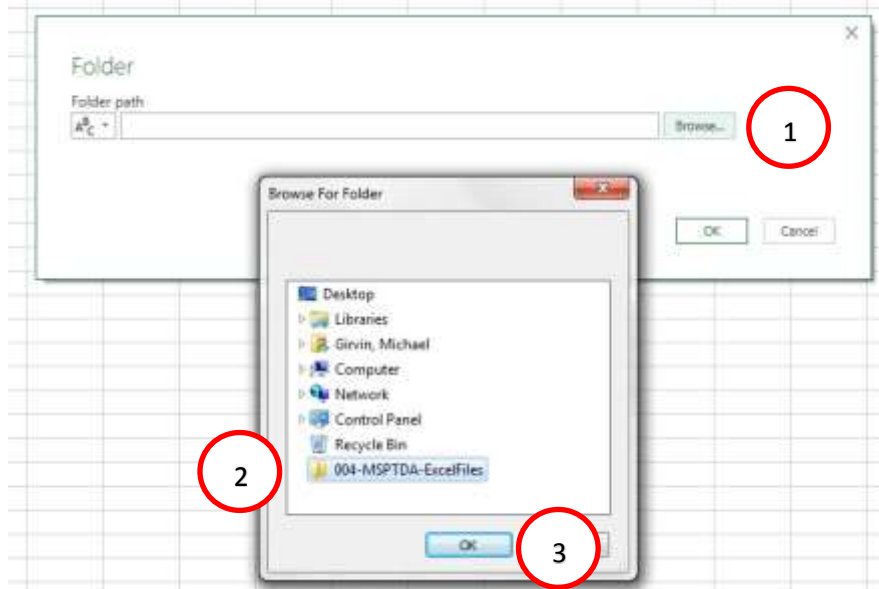
2) In the above list of files, we only want files that have an extension that contain “.xls”, and because some of the files names have extensions in capital letter and some have lower case letters, we will have to convert them all to lower case letters before importing them.

3) In Excel to import files from a folder, we can:

- i. Click on the Data Ribbon Tab.
- ii. Go to the Get & Transform group.
- iii. Click on the Get Data dropdown arrow.
- iv. Point to From File.
- v. From the sub-menu, click on From Folder option.



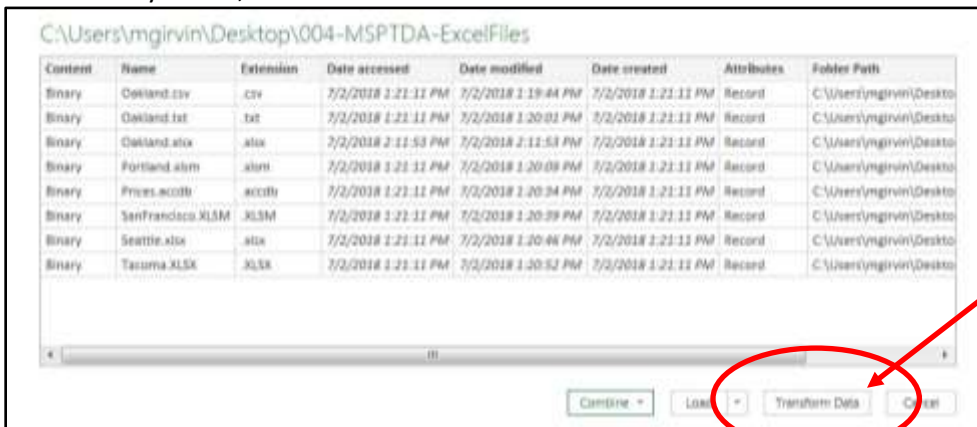
- 4) To point Power Query to the folder that contains the files we want to import, we:
  - i. In the Folder dialog box, click the Browse button.
  - ii. In the Browse for Folder dialog box, navigate to the folder named “004-MSPTDA-ExcelFiles”, then click on that folder. This is the folder that Power Query is point to in order to find files to import.
  - iii. Click OK in the Browse for Folder dialog box.



- 5) The Folder dialog box now contains the folder path that Power Query will use when we later need to refresh our query. If the folder path ever changes, we are allowed to come back and edit this folder path. To point Power Query to this folder path, click the OK button. This picture shows the Folder dialog box:



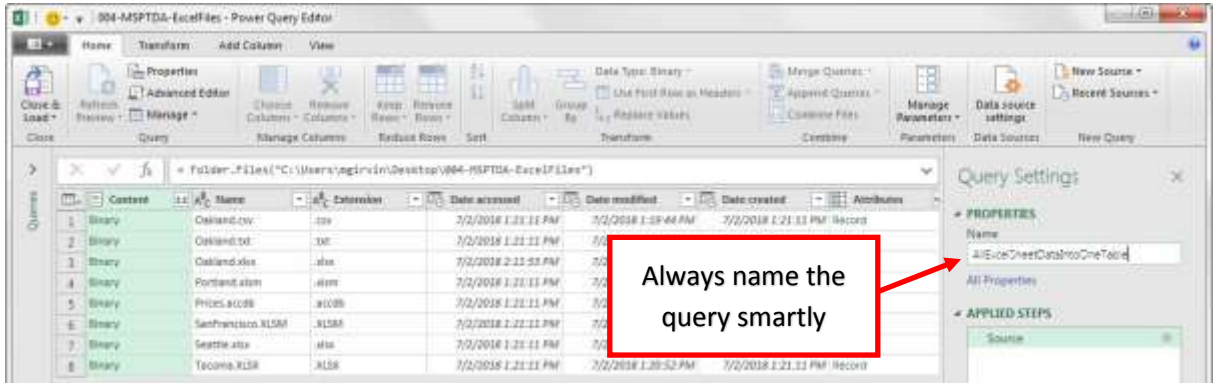
- 6) **Click Transform Data!!** In the next step click the Transform Data button to open the query in the Power Query Editor, as seen here:



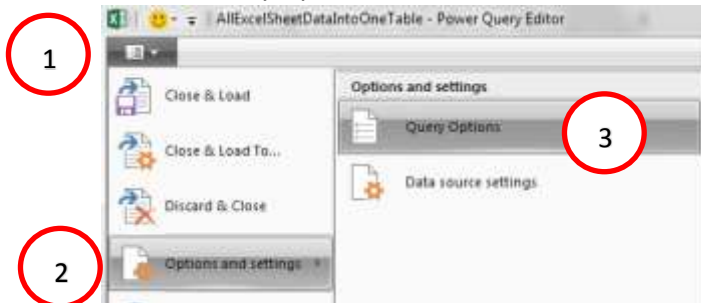
Click the **Transform Data** button, not the Combine or Load buttons.



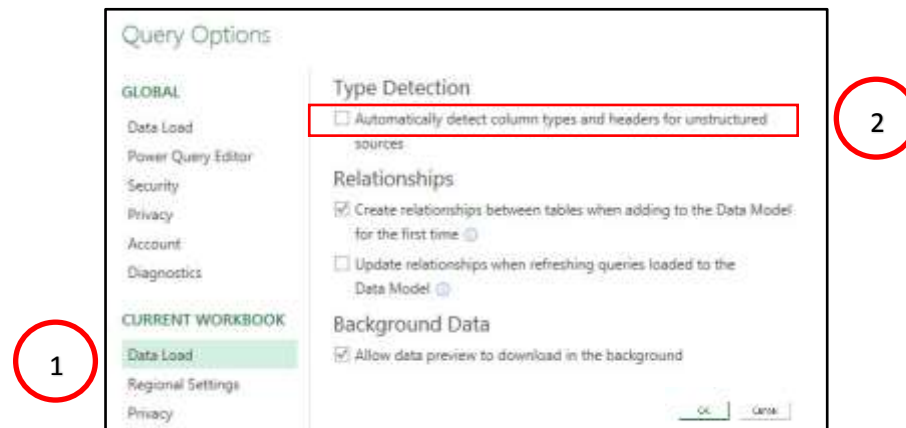
- 7) In the Power Query Editor, we can see the file name column is the second column and that there the subsequent columns are attributes about the file. The first column is named Content and is the column that contains the file. The first step for any query is always to give the Query a smart name. As seen here, we named it “AllExcelSheetDataIntoOneTable”:



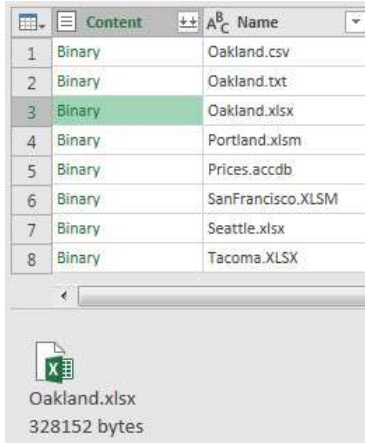
- 8) **Change Query Options for Data Types.** Before we start transforming the data using Power Query, we can turn off the default option to automatically change the Data Type for columns with unstructured data (not obvious what sort of data type the column contains, or the data source does not specify a Data Type) by:
- i. Clicking the File Menu Dropdown Arrow.
  - ii. Click on Options and settings.
  - iii. Click on Query Options.



- 9) Then in the Query Options dialog box:
- i. On the Left, click Data Load.
  - ii. Uncheck the check box for “Automatically detect column types and headers for unstructured sources”.

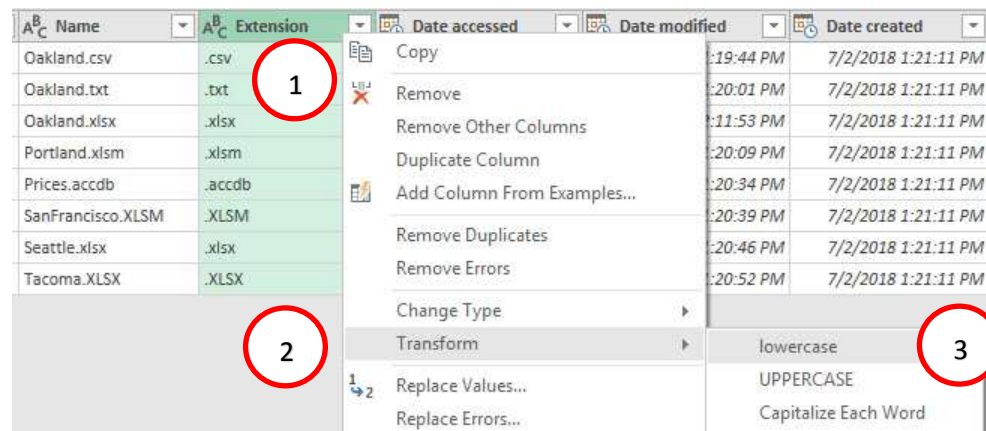


10) In the Content column, if you click to the right of the word Binary in the third row, you can see the “Oakland.xlsx” in the bottom of the query, as seen here:

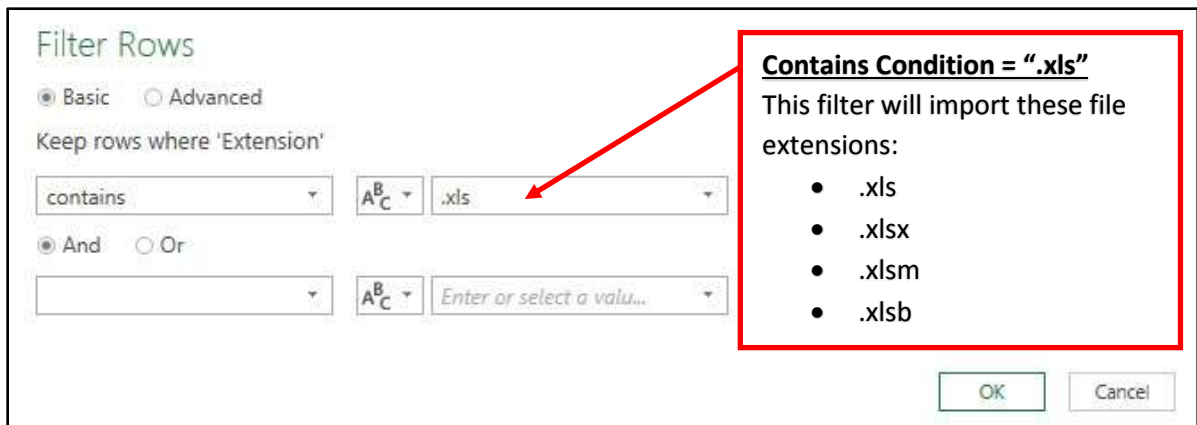


11) Because Power Query is case sensitive, and we have some extensions listed with capital letters and some with lowercase letters, we can “clean” the extensions to become all lowercase by:

- i. Right-clicking the Extension column.
- ii. Point to Transform.
- iii. Click on lowercase.



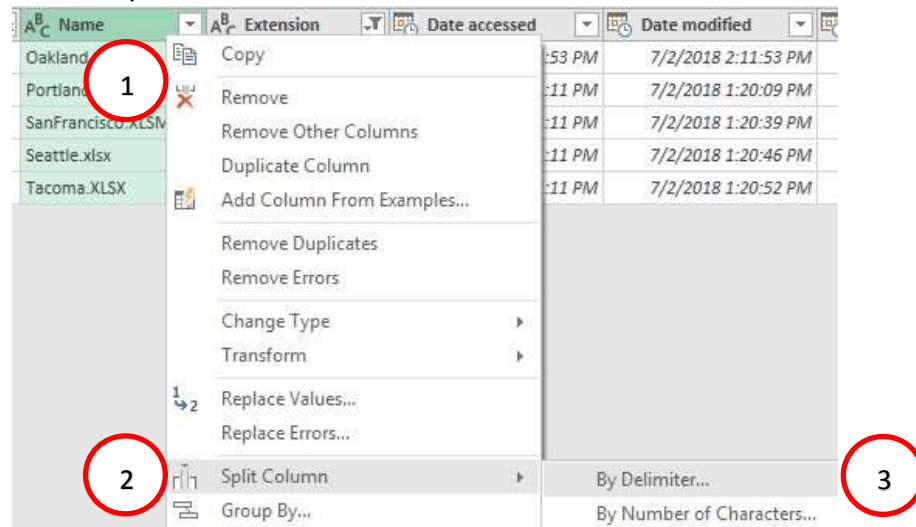
12) **To filter out non-Excel files**, go to the Extension Field, click the Filter Dropdown Arrow, point to the Text Filter option, then click on “Contains” option. Then in the Filter Rows contains text box, type “.xls”, as seen here:



13) **Clean File Name.** Next, we need to “clean” the Name column so that only the City File Name remains.

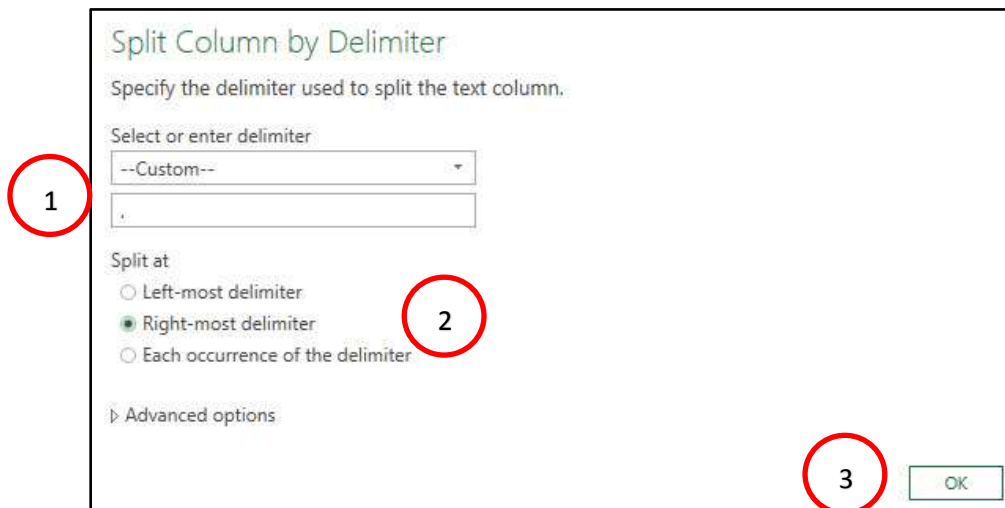
First, we must:

- i. Right-clicking the Name column.
- ii. Point to Split Column.
- iii. Click on By Delimiter.

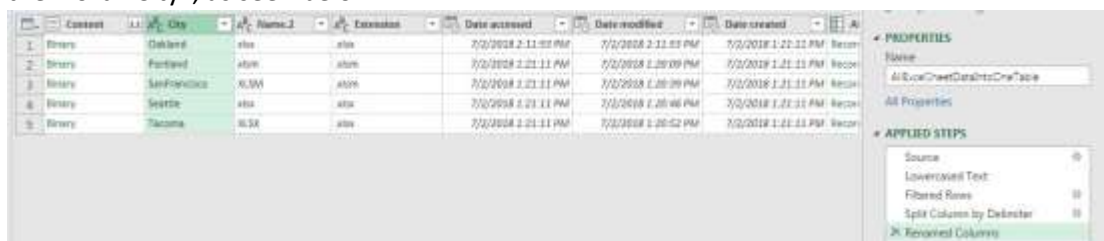


14) When the Split Column by Delimiter dialog box pops up:

- i. Type the Delimiter period (dot).
- ii. Click dialog button for “Right-most delimiter”.
- iii. Click OK.

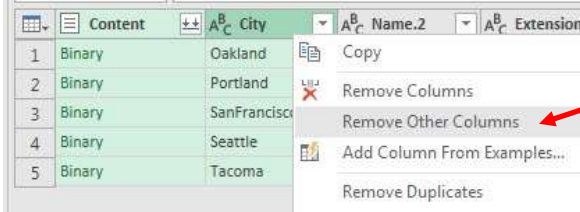


15) This will give us the file that we want to import and transform, as seen in the picture below. As soon as this filtered list appears, rename the Name.1 column by double clicking the column header and typing the word “City”, as seen below:

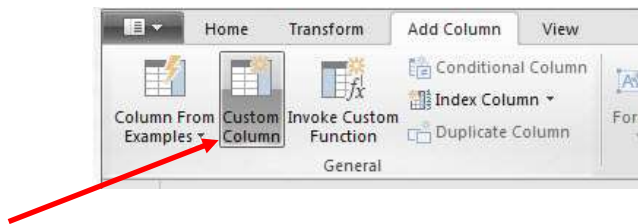




16) Next, we need to remove all columns except the Content and City columns. To do this we click on the Content Column, then hold the Ctrl Key and click on the City Column. Then with both columns selected, right-click one of the two columns and point to Remove Other Columns, as seen here:



17) **Custom Column using Excel.Workbook Function.** In order to extract the correct objects from each Excel File, we need to add a Custom Column. Custom Columns allow us to use formulas (like we would in Excel) to accomplish a task that is not available in the Ribbon Tabs. When we add our Custom Column, we will use the Power Query Function (M Code) called Excel.Workbook. To add a Custom Column, go to the Add Column Ribbon Tab, then in the General group click the Custom Column button, as seen here:

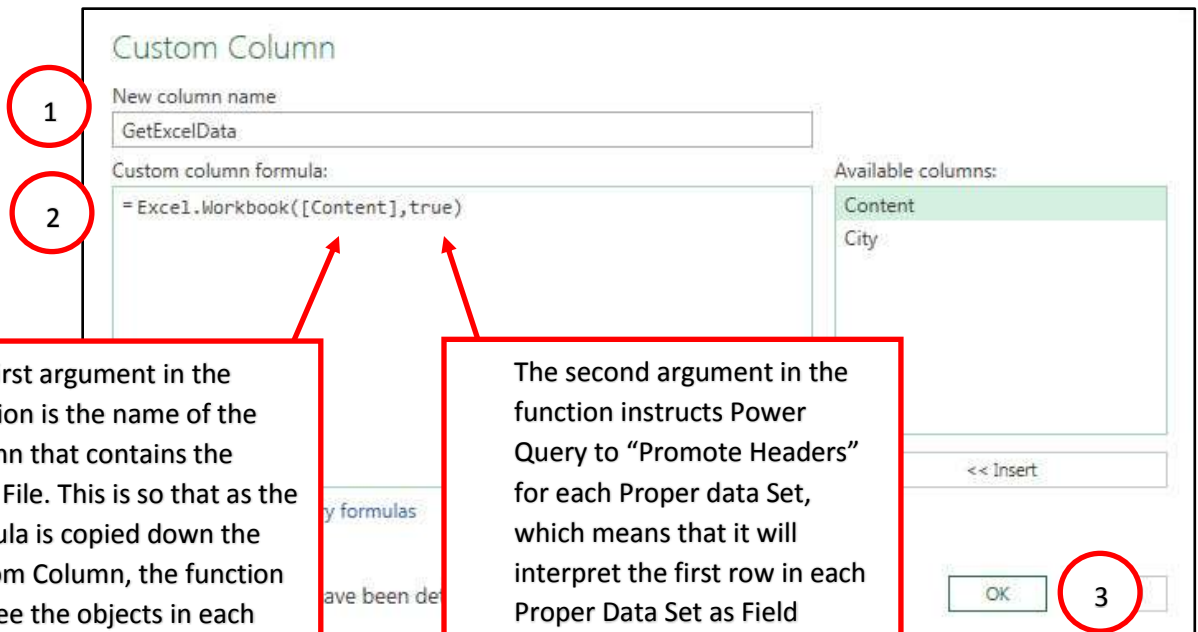


18) Then in the Custom Column dialog box:

- i. In the New column name textbox, type the name: "GetExcelObjects".
- ii. In the Custom column formula textbox, type the formula:

**= Excel.Workbook([Content],true)**

iii. Then click OK.



The first argument in the function is the name of the column that contains the Excel File. This is so that as the formula is copied down the Custom Column, the function can see the objects in each Excel File.

The second argument in the function instructs Power Query to "Promote Headers" for each Proper data Set, which means that it will interpret the first row in each Proper Data Set as Field Names for each table.

19) The result of the Custom Column is seen below.

- i. The Custom Column named GetExcelData delivers a table for each row that gives us information about what kind of objects are in each Excel File.
- ii. If you click to the right of the word table in the first cell of the GetExcelData column, you can see a table appear in the lower part of the query.
- iii. The table in the lower part of the query has five columns that give us information about what type of objects are in the Excel File. For example, in the “Kind” column we can see that some of the objects are Excel Worksheets (Sheet), one is an Excel Table (Table) and some are Defined Names (DefinedName). This information will help us to filter the objects and import only object that have data that we want.

Name	Data	Item	Kind	Hidden
Fran	Table	Fran	Sheet	FALSE
Gab	Table	Gab	Sheet	FALSE
Popi	Table	Popi	Sheet	FALSE
Sheet1	Table	Sheet1	Sheet	FALSE
SmallTable02	Table	SmallTable02	Table	FALSE
_xlnm_FilterDatabase	Table	Sheet1!_xlnm_FilterDatabase	DefinedName	TRUE
_xlnm.Criteria	Table	Sheet1!_xlnm.Criteria	DefinedName	FALSE
_xlnm.Extract	Table	Sheet1!_xlnm.Extract	DefinedName	FALSE
_xlnm.Print_Area	Table	Sheet1!_xlnm.Print_Area	DefinedName	FALSE

20) The next step is to remove the Content column, so we right-click the Content column and click on Remove.

21) To expand the tables in each row and repeat the City name for each row associated with that file:

- i. Click Expand button.
- ii. Uncheck “Use original column name as prefix”.
- iii. Click OK.

**1) Click the Expand button**

**2**

**3**

22) **The Expanded Table with each Excel Files objects looks like this:**

- i. The City column repeats the City File Name for each row that belongs to the specified file.
- ii. The Name and Item columns are similar in that they contain the name of the object. The Item column has slight more descriptive name for the Defined Names because the name also lists the Excel Worksheet Name as part of the full name.
- iii. The Data column contains the data for each object.
- iv. The Kind column tells us what type or “Kind” of object contains the data from the Data column. We can use this column, for example, to filter and then import on Sheet objects.
- v. The Hidden column tells us whether or not the object is hidden. For example, the Defined Name for the table that had the Filter Feature Applied (row 6) shows TRUE in the Hidden column because this is a Hidden Defined Name in an Excel Workbook File.

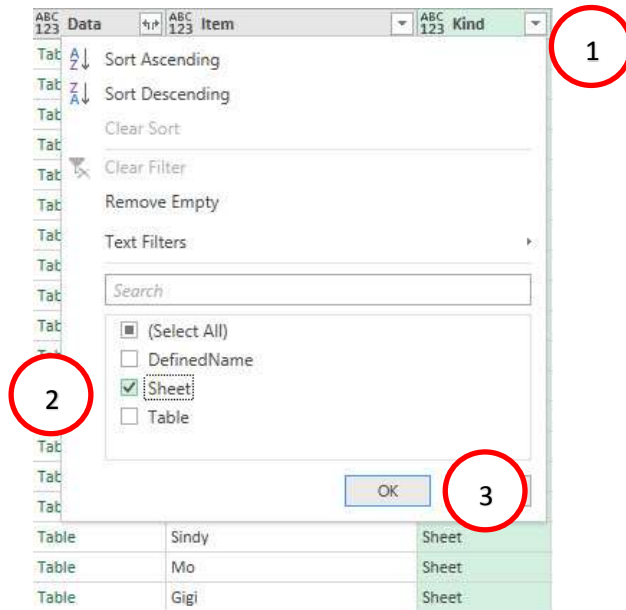
	City	Name	Data	Item	Kind	Hidden
1	Oakland	Fran	Table	Fran	Sheet	FALSE
2	Oakland	Gab	Table	Gab	Sheet	FALSE
3	Oakland	Popi	Table	Popi	Sheet	FALSE
4	Oakland	Sheet1	Table	Sheet1	Sheet	FALSE
5	Oakland	SmallTable02	Table	SmallTable02	Table	FALSE
6	Oakland	_xlnm_FilterDatabase	Table	Sheet1!_xlnm_FilterDatabase	DefinedName	TRUE
7	Oakland	_xlnm.Criteria	Table	Sheet1!_xlnm.Criteria	DefinedName	FALSE
8	Oakland	_xlnm.Extract	Table	Sheet1!_xlnm.Extract	DefinedName	FALSE
9	Oakland	_xlnm.Print_Area	Table	Sheet1!_xlnm.Print_Area	DefinedName	FALSE
10	Oakland	SmallTable	Table	SmallTable	DefinedName	FALSE
11	Portland	Sioux	Table	Sioux	Sheet	FALSE
12	Portland	Chin	Table	Chin	Sheet	FALSE
13	Portland	Tyrone	Table	Tyrone	Sheet	FALSE
14	SanFrancisco	Miki	Table	Miki	Sheet	FALSE

23) **Headers Have Been Promoted.** To prove to ourselves that the second argument in the Excel.Workbook Function actually did promoter the first row of each table, so they became Field Names, we can click to the right of the word “Table” in the second row of the Data column. The table that appears in the lower part of the Query Editor, shows the Proper Data Set with Field Names.

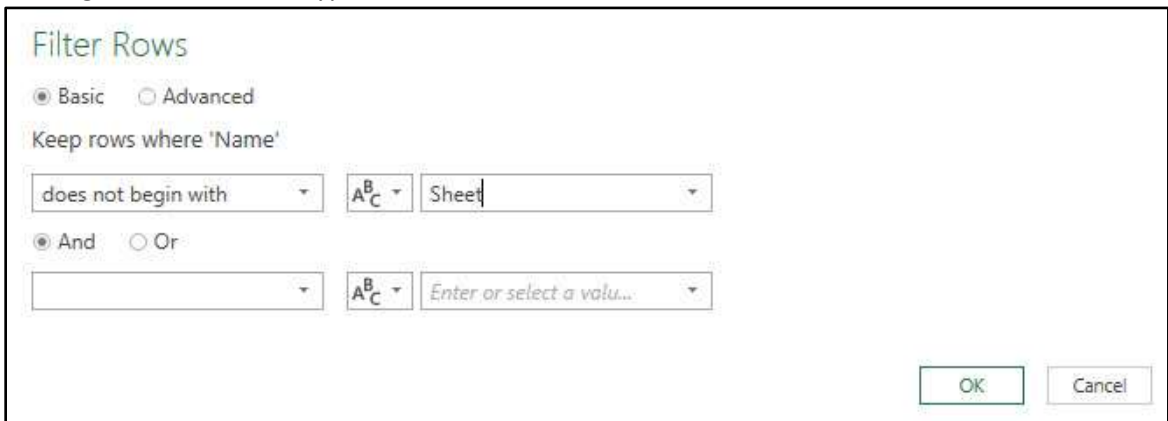
Date	Product	Units	Sales
10/19/2017	Aussie Round	108	\$289.68
12/6/2017	Aussie Round	48	\$460.16
3/31/2017	Aussie Round	24	\$95.12
11/9/2017	Quat	12	\$08.56

24) **Filter to import only Sheet Objects.** With the goal of filtering out objects that are not Sheet Objects, we can:

- i. Click the Filter Arrow in the Kind Column
- ii. From the Unique List of items, uncheck everything except for “Sheet”
- iii. Click OK.



25) **Filter to remove Sheet objects that Begin with “Sheet”.** To remove Sheet Objects that have not been properly named with a SalesRep name, go to the Name Field, click the Filter Dropdown Arrow, point to the Text Filter option, then click on “Does not begin with” option. Then in the Filter Rows “does not begin with” text box, type “Sheet”, as seen here:



26) Now we are left with only rows that contain Proper Data Sets where the object is a Sheet object and the Sheet Name does not begin with the word “Sheet” (has a SalesRep Name), as seen here:

	ABC 123 City	ABC 123 Name	ABC 123 Data	ABC 123 Item	ABC 123 Kind	ABC 123 Hidden
1	Oakland	Fran	Table	Fran	Sheet	FALSE
2	Oakland	Gab	Table	Gab	Sheet	FALSE
3	Oakland	Popi	Table	Popi	Sheet	FALSE
4	Oakland	Sheet1	Table	Sheet1	Sheet	FALSE
5	Portland	Sioux	Table	Sioux	Sheet	FALSE

27) In the above picture, the “Name” column does not have a useful name. In order to change the name, double click and change the column name to “SalesRep.”

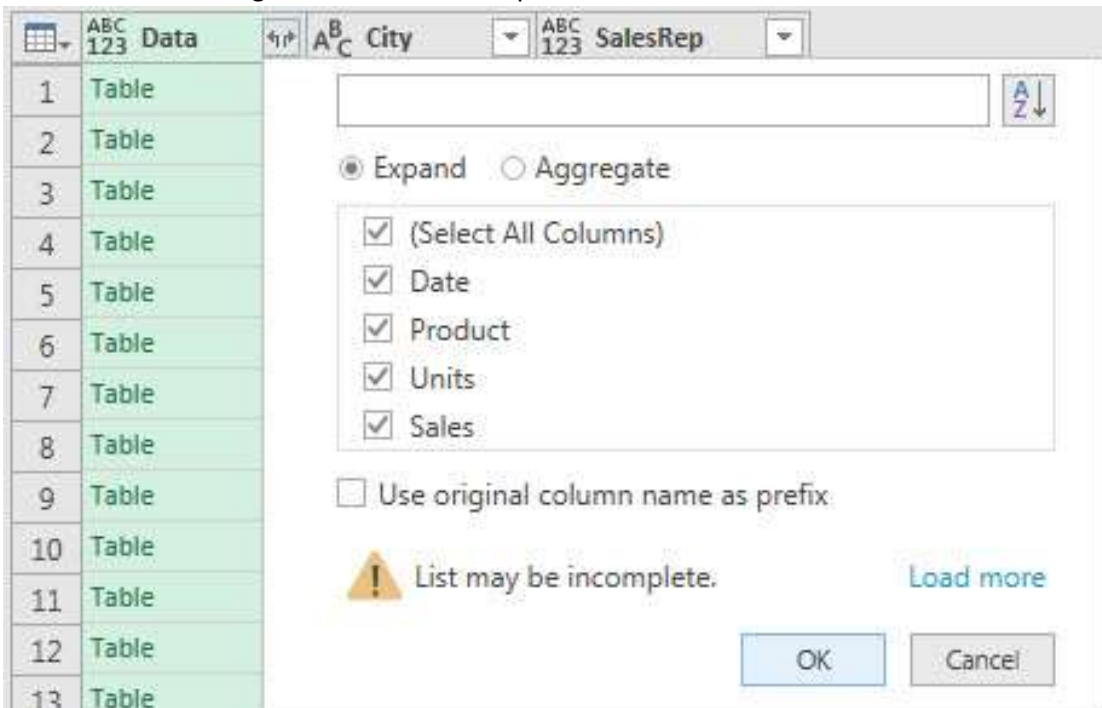
28) The final columns we need for our Appended Proper Data Set are:

- i. Data (contains the correct Proper Data Sets)
- ii. City (contains the correct City name)
- iii. SalesRep (contains the SalesRep name).

29) To remove the unwanted columns, select the columns using the Ctrl Key by first clicking on the Data column, then the City column, and finally the SalesRep column. With the columns selected, right-click any one of the columns and then click on Remove Other Columns, as seen in the picture below:



30) **Final Append with all Proper Data Sets.** To complete our final append process to combine all the Proper Data Sets into a single table with six columns, click on the expand button in the Data column, uncheck the “Use original column name as prefix” and then click OK.





31) Here is a picture of the final Proper Data Set, including the last Power Query transformation stop on adding the correct Data Types for each column. There should be 15 steps in the Import, Extract, Clean, Transformation and Load Query, as seen in the below picture:

	Date	Product	Units	Sales	City	SalesRep
1	11/8/2017	Aussie Round	216	7108.56	Oakland	Fran
2	11/14/2017	Tri Fly	84	551.88	Oakland	Fran
3	12/18/2017	Tri Fly	276	1443.48	Oakland	Fran
4	11/18/2017	Bellen	48	1362.72	Oakland	Fran
5	12/28/2016	Bellen	72	2027.52	Oakland	Fran
6	11/11/2017	Sunshine	96	1888.32	Oakland	Fran
7	9/12/2017	Aussie Round	60	1942.8	Oakland	Fran
8	11/23/2016	Aspen	144	2655.36	Oakland	Fran
9	11/28/2017	Aussie Round	204	6587.16	Oakland	Fran
10	12/18/2016	Tri Fly	72	484.56	Oakland	Fran
11	11/7/2017	Tri Fly	48	318.24	Oakland	Fran
12	11/4/2017	Aspen	96	1728.96	Oakland	Fran
13	10/24/2017	Carlota	72	2020.32	Oakland	Fran
14	2/12/2016	Aspen	96	1593.6	Oakland	Fran
15	12/7/2017	Tri Fly	264	1172.16	Oakland	Fran
16	11/27/2017	Quad	72	3107.52	Oakland	Fran
17	11/20/2017	Carlota	12	347.76	Oakland	Fran
18	12/2/2017	Aspen	72	1326.24	Oakland	Fran
19	11/1/2016	Quad	72	3053.52	Oakland	Fran
20	10/18/2017	Aspen	132	2195.16	Oakland	Fran
21	12/18/2016	Quad	108	4453.92	Oakland	Fran
22	12/19/2017	Aussie Round	72	2298.96	Oakland	Fran

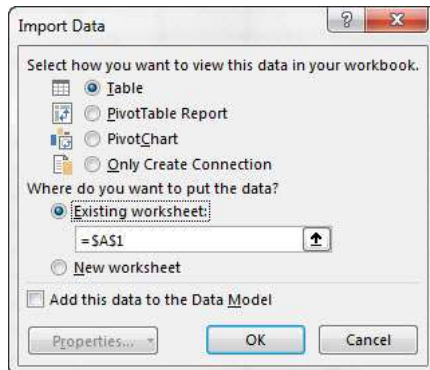
**PROPERTIES**

Name: AllExcelSheetDataIntoOneTable

**APPLIED STEPS**

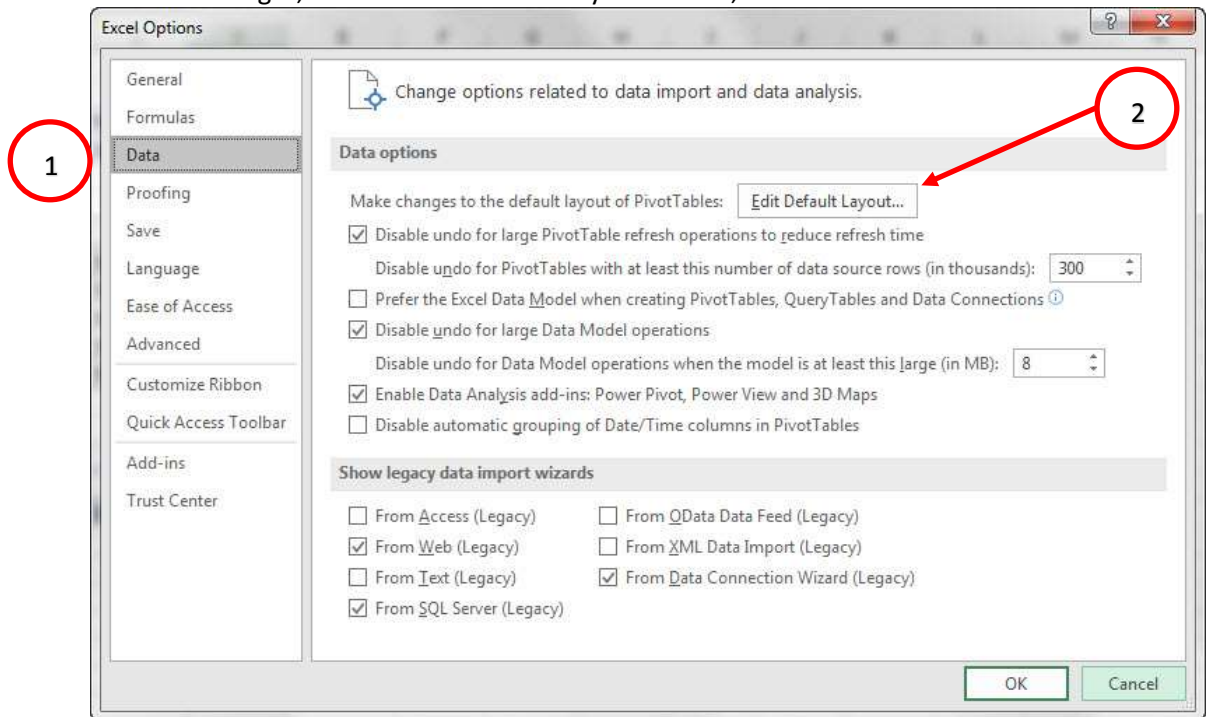
- Source \*
- Lowercased Text \*
- Filtered Rows \*
- Split Column by Delimiter \*
- Renamed Columns \*
- Removed Other Columns \*
- Added Custom \*
- Removed Columns \*
- Expanded GetExcelData \*
- Filtered Rows1 \*
- Filtered Rows2 \*
- Renamed Columns1 \*
- Removed Other Columns1 \*
- Expanded Data \*
- Changed Type \***

32) **Load to Excel Sheet.** Because we only have about 60,000 rows of data (not a lot of data), and our calculations can be done with a Standard PivotTable, and we would like to see the data in a sheet (so we can filter and sort), we will load the data to an Excel Worksheet. To Load to the Excel Sheet named data, we use the Close & Load dropdown in the Home Ribbon Tab, then select the “Close and Load To...” option and load the data as a Table in cell A1 on the Data Sheet. The Import Data dialog box is seen here:

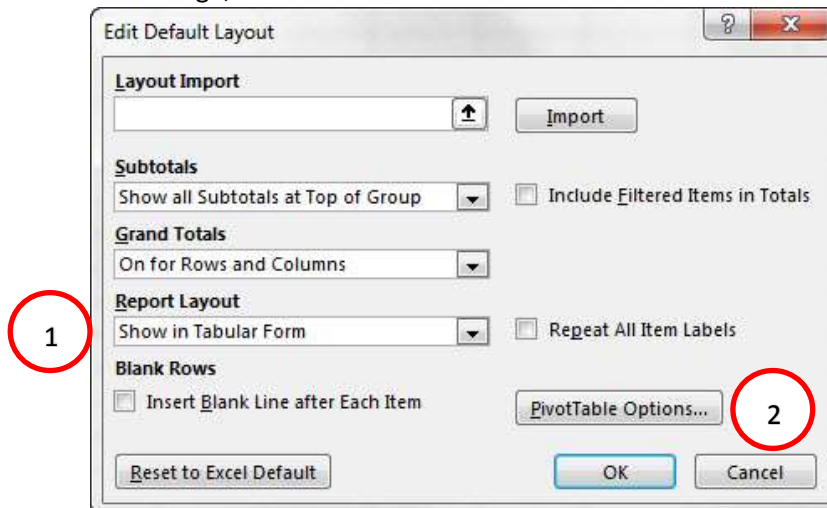




4. **Change Default PivotTable Layout & Options.** If we would like our PivotTables (both Standard and Data Model PivotTables) to have the Tabular Layout by default so that Field Names are used rather than generic labels in a PivotTable, we can change it by going to Options in the File Menu, then:
- On the Left, click on the Data Tab.
  - Then on the right, click the Edit Default Layout button, as seen here:



- 2) In the Edit Default Layout dialog box, we can change a number of default settings including:
- Using the dropdown in the Report Layout textbox to select "Show in Tabular Form". This setting will force PivotTables to use Field Names as labels in a PivotTable report rather than the generic, "Row labels" and "Column labels".
  - The PivotTable Options button opens the PivotTable Options dialog box with many settings you can change, as seen here:



3) We built a PivotTable Report:

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2												
3	Sum of Sales		Years									
4	Product	SalesRep	2016	2017	Grand Total							
5	Carlota	Xan	\$774,669	\$869,892	\$1,644,562							
6		Bo	\$749,155	\$716,638	\$1,465,793							
7		Quin	\$932,335	\$684,593	\$1,616,928							
8	Carlota Total		\$2,456,159	\$2,271,123	\$4,727,283							
9	Quad	Xan	\$1,112,996	\$1,153,815	\$2,266,811							
10		Bo	\$1,036,595	\$1,071,473	\$2,108,067							
11		Quin	\$1,185,168	\$1,335,792	\$2,520,960							
12	Quad Total		\$3,334,758	\$3,561,080	\$6,895,838							
13	Tri Fly	Xan	\$166,513	\$185,744	\$352,258							
14		Bo	\$118,796	\$125,959	\$244,755							
15		Quin	\$146,584	\$130,234	\$276,817							
16	Tri Fly Total		\$431,893	\$441,937	\$873,830							
17	Grand Total		\$6,222,810	\$6,274,140	\$12,496,951							
18												

5. **Definition of a PivotTable.** In the video we reminded ourselves of what we learned in the prerequisite class: the essence of a PivotTable is that it makes calculations with conditions or criteria. In the picture below, we see that the selected cell is making a sum calculation with four conditions or criteria or filters. We also reminded ourselves that the definition of a PivotTable is: **Summary Reports with Calculations based on One or More Conditions or Criteria.**

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2												
3	Sum of Sales		Years									
4	Product	SalesRep	2016	2017	Grand Total							
5	Aspen	Fran	\$467,009	\$412,105	\$879,114							
6		Gab	\$417,710	\$417,904	\$835,615							
7		Popi	\$451,154	\$501,077	\$952,231							
8	Aspen Total		\$1,335,873	\$1,331,087	\$2,666,960							
9	Grand Total		\$1,335,873	\$1,331,087	\$2,666,960							
10												
11												
12												
13												
14	Sum Calculation = adding											
15	Product = Aspen = Criteria 1											
16	SalesRep = Gab = Criteria 2											
17	Year = 2016 = Criteria 3											
18	City = Oakland = Criteria 4											
19												
20												
21												
22												
23												
24												
25												

6. **AND Logical Test.** We also reminded ourselves that an AND Logical Test means that in order for the record to be included in the calculation, all conditions or criteria must be met.

7. **Add New Excel Workbook Files to the Folder & Refresh.** As we saw in the video, when we added new files to the folder that met all the conditions of the query, and we refreshed both the Query and the PivotTable (two separate refreshes), the data from the new file was incorporated into the Query Load To Output and the PivotTable report.
  - i. We can refresh the Query in a few different places:
    1. In the Queries & Connections Pane, we can right-click the Query and click on Refresh.
    2. In the Queries & Connections Pane, we click the refresh icon in the upper right corner of the query.
    3. We can use the Refresh or Refresh All button in the Data Ribbon Tab in the Queries & Connections group.
    4. We can right-click the Query Load To Output in the Excel Worksheet and click on Refresh.
  - ii. We can refresh the PivotTable
    1. We can right-click the PivotTable and click on Refresh.
    2. We can use the Refresh or Refresh All button in the PivotTable Tools Analyze Ribbon Tab in the Data group.
  - iii. If we Load the Power Query Transformation to an Excel Sheet, we have to perform two separate Refreshes:
    1. First Refresh the Query.
    2. Second Refresh the PivotTable
  - iv. If we Load the Power Query Transformation to a PivotTable (option in the Import Data dialog box), all we have to do is refresh the PivotTable one time, and then both the Query and PivotTable report are refreshed.
8. **Edit Query when Folder Path Changes.** If the Folder Path changes, the you must edit the first step of the query and change the incorrect folder path to the correct folder path.