

Office 2016– Excel Basics 25

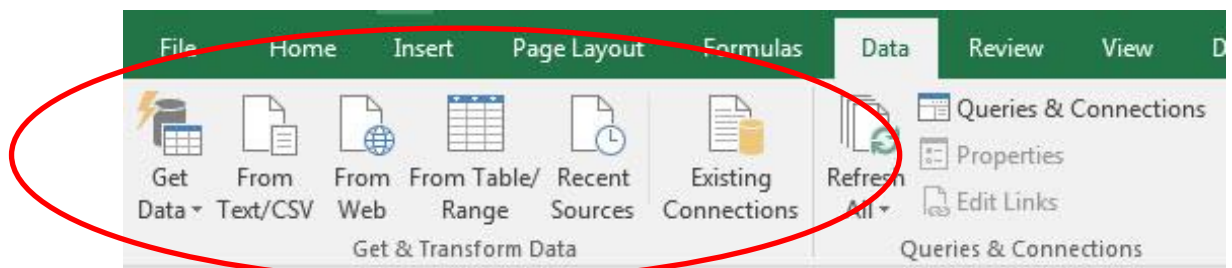
Video/Class Project #37

Excel Basics 25: Power Query (Get & Transform Data) to Convert “Bad Data” into Proper Data Set

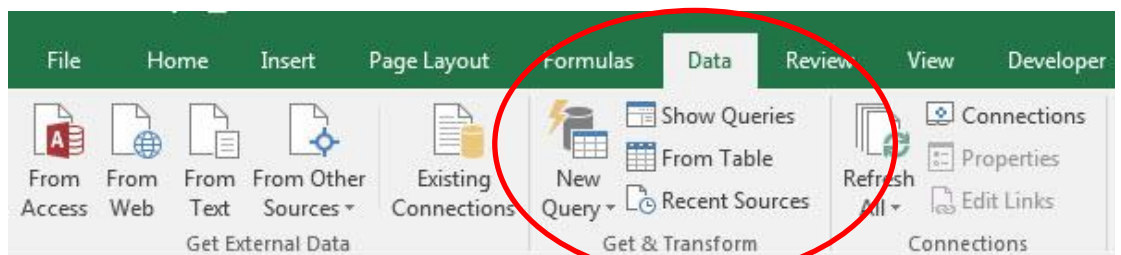
Goal in video # 25: Learn about how to use the Get & Transform group in the Data Ribbon Tab to import, clean and transform data.

Topics Covered in Video:

- 1) Data Analysis:
 - i. Converting Raw Data into Useful Information
- 2) Proper Data Set:
 - i. Field Names / Column Headers in first row
 - ii. Records in each row (sales transactions, employee records and so on)
 - iii. Empty cells all around data set (or Excel row or column headers)
- 3) “Bad Data”
 - i. “Bad Data” = Raw Data that is not in a Proper Data Set form.
 - ii. If we don’t have a Proper Data Set than we can use Excel Data Analysis features.
 - iii. So when we get “Bad Data”, we have to clean and transform the data so it becomes data that is stored in a Proper Data Set.
 - iv. Power Query (Get & Transform) is an Excel Feature that can clean and transform data.
- 4) Power Query (Get & Transform Data)
 - i. Power Query history:
 1. Power Query has been around since Excel 2010
 2. In Excel 2010 & 2013 Power Query was an add-in that you had to download and install
 3. In Excel 2016 Power Query become the “Get & Transform” group in the Data Ribbon Tab.
 4. Depending on the version of Excel that you have, the Get & Transform group looks different.
 - i. If you have Office 365, your “Get and Transform Data” group may look like this:



- ii. If you do NOT have Office 365, your “Get and Transform” group may look like this:

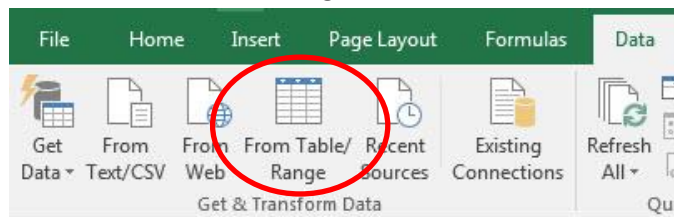


ii. What does Power Query do:

1. Import Data from External sources
 - i. External sources such as text files, Excel workbooks, databases and more
2. Clean Data
 - i. Like split data in a column into three columns
3. Transform Data
 - i. Like combining multiple tables into one Proper Data Set

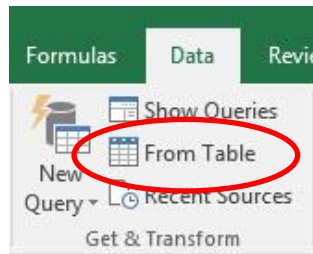
iii. If your data is in an Excel sheet, you must convert the data to an Excel Table using the Excel Table feature.

1. This is to ensure that if you add new data, the output from Power Query will update when new data is added.
2. Then you click the “From Table” button in the Get & Transform group in the Data Ribbon Tab to put the Excel Data into the Power Query Editor:
 - i. In Office 365 the “From Table/Range” button looks like this:



1.

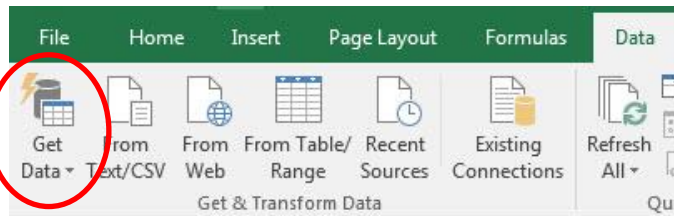
ii. In a version that is NOT Office 365 the “From Table” button looks like this:



1.

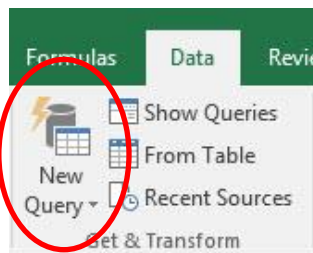
iv. If you have data from an external source, we can begin the process of importing the data by clicking the following button:

i. In Office 365 to get external data click the “Get Data” button:



1.

ii. In a version that is NOT Office 365 to get external data click the “New Query” button:



1.

- v. Clean Data with Power Query (Get & Transform), Flash Fill or formulas?
 1. Use Flash Fill when the data cleaning is a one-time event and you don't need the solution to update when the source data changes.
 2. Use Formulas when the data cleaning solution needs immediately update when the source data changes.
 3. Use Power Query when you need the solution to update after you click the Refresh button.
- vi. Other reasons to use Power Query rather than Flash Fill, Formulas and other methods of cleaning and transforming data:
 1. Power Query can be significantly easier when the cleaning is complicated or has multiple steps, especially as compared to formulas
 2. If the data is coming from an external source, Power Query can be significantly easier than any other method.
 3. If you have multiple files that you need to transform into a single Proper Data Set, like in the example in our video, Power Query is significantly easier than any other method.
- vii. Every time we use Power Query, we create a new "Query" of the Start Data Set.
 1. The word "Query" means to ask a question.
 2. When we start a "Query" based on a Start or Begin Data Set, we are asking a question of the data set.
 3. For example:
 - i. If our Start or Begin Data Set looks like this:

Description	Amount
Carlota / 05/05/2016 / West	485
 - ii. We might want to ask the question or "Query": "Can I see the Description Column Split into three separate columns?", where the answer to the "Query" would look like this:

Product	Date	Region	Sales
Carlota	5/5/2016	West	485
- viii. Power Query uses "Data Types" for each Column / Field
 1. "Data Type" indicates what sort of data can go in a particular Field or Column.
 - i. For example, we may designate a column as one of these Data Types:
 1. **Text**: Where the column can only contain Text values
 2. **Date**: Where the column can only contain Date values
 3. **Whole Number**: Where the column can only contain Whole Numbers with no decimals
 4. **Currency**: Where column can contain Currency amounts with upto 4 decimals
 5. There are more Data Types also...
 2. When you designate a Column / Field as a specific data type, you are building a more robust data storage system, because you are assuring that the right kind of data goes into the column.
 3. Power Query and Access both use "Data Types".
 4. **Important**: If you do not choose the correct data type for each column, the data analysis features in Excel, like a PivotTable may not work correctly.
 - i. For example:
 1. A Date Field without the Date Data Type cannot group in a PivotTable.
 2. A Sales Column without a number Data Type may not be able to be summed.

5) Delimiter?

- i. A delimiter is a character or set of characters that separate data that should be broken apart into separate columns
 1. For example for this data: Carlota / 05/05/2016 / West
 - i. The delimiter is " / " (Space, Forward Slash, Space).
 - ii. The delimiter tells us how the data should be split into separate columns.
 - iii. The resultant data should be:

Description	Amount
Carlota / 05/05/2016 / West	485

Product	Date	Region	Sales
Carlota	5/5/2016	West	485

6) Why store data in Text Files?

- i. Because there are so many different systems for storing data, the agreed upon file type "Text Files" is often used to transfer data from one system to another.
- ii. Examples of Text File extensions:
 1. ".txt"
 2. ".csv"
- iii. The Delimiters used for different text files:
 1. ".txt" uses a Tab as the delimiter
 2. ".csv" uses comma as the delimiter

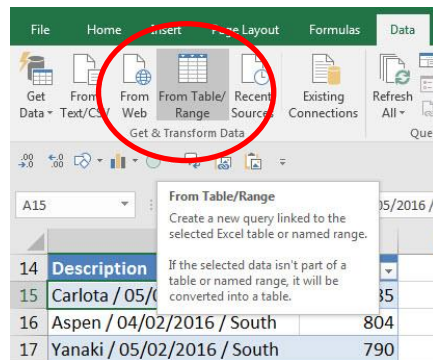
Example #1: Convert "NOT a Proper Data Set" into a Proper Data Set & Make PivotTable Report

- 1) On the sheet named "Ex(1)" there is a table with two Columns/Fields named "Description" and "Amount". The Description column is "Bad Data" because we can't use the Product, Date or Region in a PivotTable Report unless we break it apart into three columns. This is what the data looks like:

	A	B	
14	Description	Amount	
15	Carlota / 05/05/2016 / West	485	
16	Aspen / 04/02/2016 / South	804	
17	Yanaki / 05/02/2016 / South	790	
18	FlatTop / 04/11/2016 / South	965	

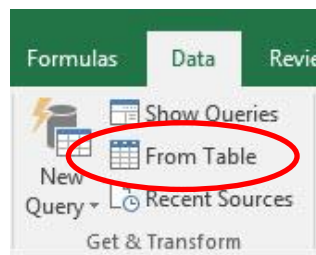
- 2) Notice that what separates the Product, Date and Region columns is a "delimiter". The delimiter for this column is " / ", that is a space, forward slash and a space.
- 3) In order to use data from an Excel Sheet we must convert the data to an Excel Table. We can convert the table of data to an Excel Table by clicking in a single cell and using the keyboard Ctrl + T. In the Properties group in the Table Tools Design Ribbon Tab we can name the table "StartSalesTable".
- 4) To bring the Excel Table into the "Power Query Editor Window", click in one cell in the Excel Table and in the Data Ribbon Tab, in the "Get & Transform" group, click the "From Table" button:

- i. In Office 365 it looks like this:



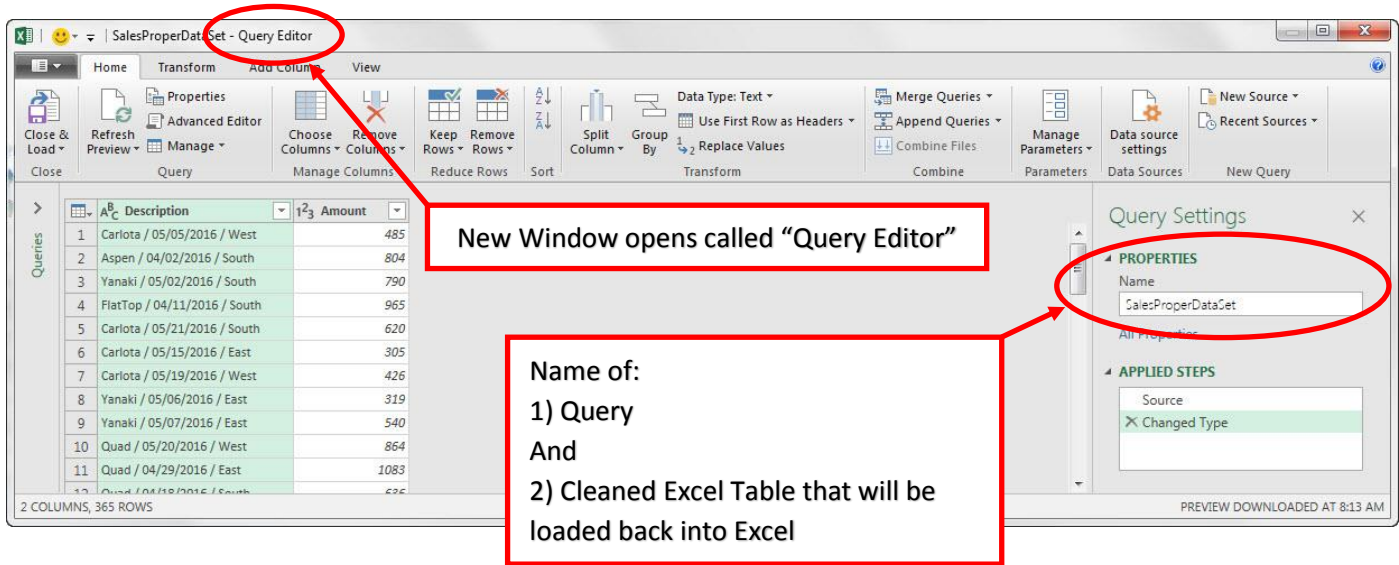
1.

- ii. In a version that is NOT Office 365, it looks like this:

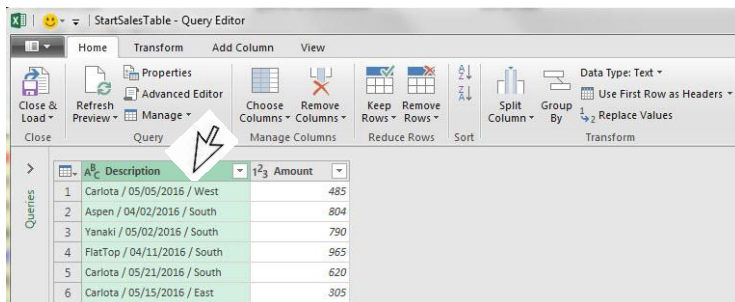


1.

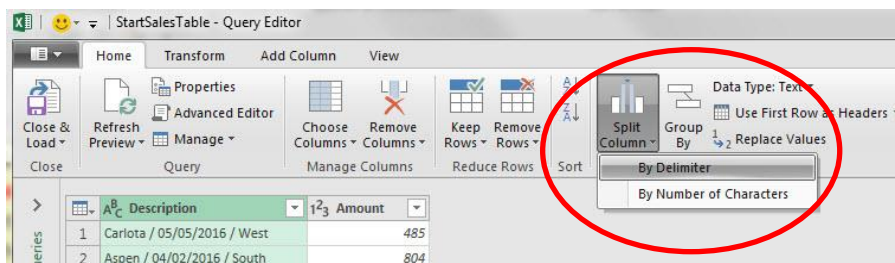
- 5) A New window opens up that reads “Query Editor” in title Bar. On the Right, we want to click in Query Settings Task Pane Name textbox and name the Query “SalesProperDataSet”. This name is the name of the Query and the name of the “Cleaned” Excel Table that will be loaded back into Excel.



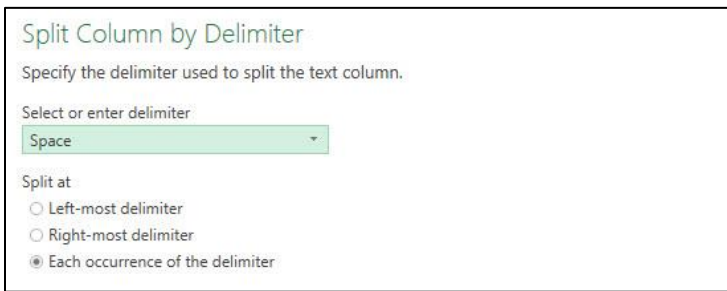
- 6) Click on the “Description” Column Header to select the whole column



- 7) Notice that in the Description column, what separates the Product, Date and Region columns is a “delimiter”. The delimiter for this column is “ / ”, that is a space, forward slash and a space. Now we need to split the Description column into three separate columns.
- 8) In the Query Editor Home Ribbon Tab, in the Transform group, click on the Split button and then click on “By Delimiter”:



9) The Split By Delimiter dialog box pops up and looks like this:



Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter

Space

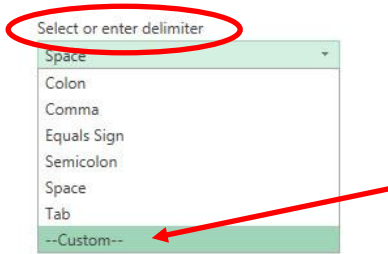
Split at

☐ Left-most delimiter

☐ Right-most delimiter

☒ Each occurrence of the delimiter

10) From the “Select or enter delimiter” drop-down, select “Custom” like this:



Select or enter delimiter

Space

Colon

Comma

Equals Sign

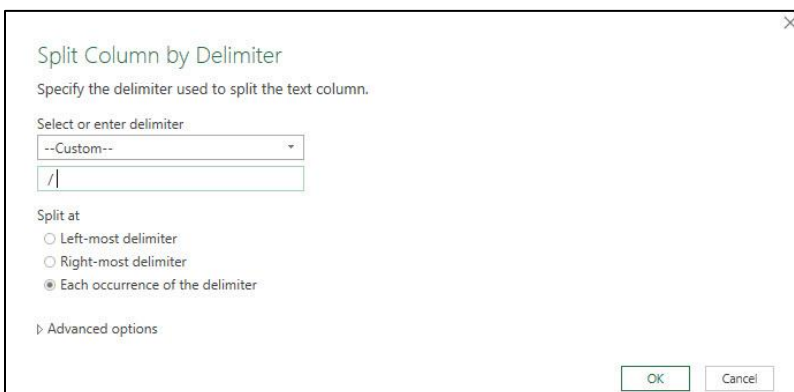
Semicolon

Space

Tab

--Custom--

11) Type a space, forward slash and then space, like this:



Split Column by Delimiter

Specify the delimiter used to split the text column.

Select or enter delimiter

--Custom--

/ /

Split at

☐ Left-most delimiter

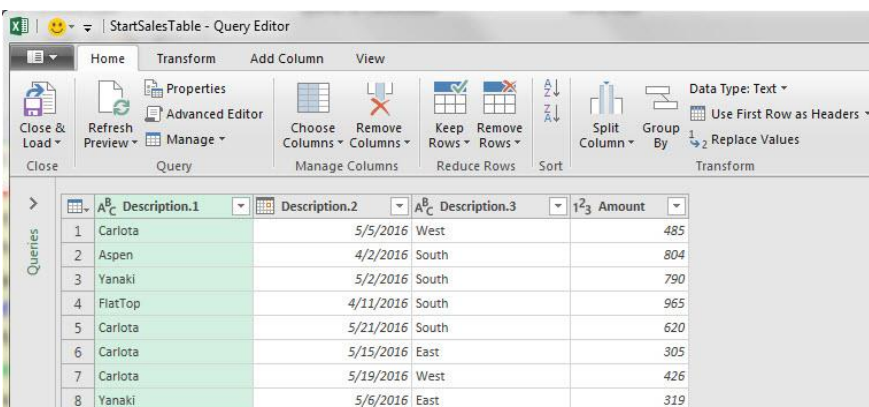
☐ Right-most delimiter

☒ Each occurrence of the delimiter

Advanced options

OK Cancel

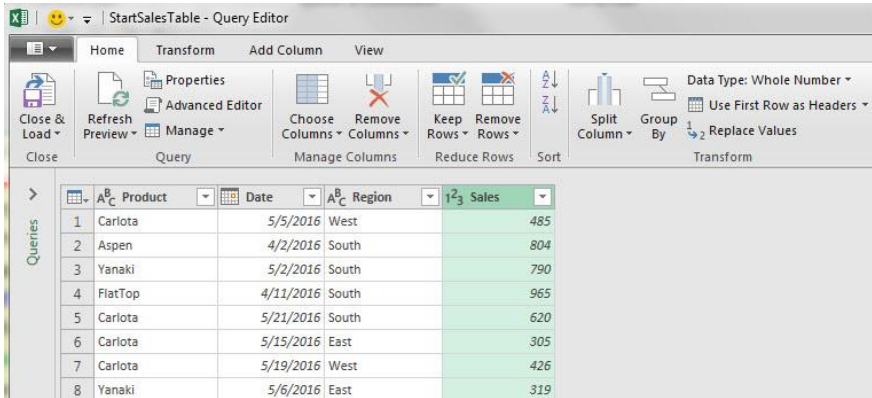
12) After you click OK, the new “split” data set looks like this:



	A _C Description.1	Description.2	A _C Description.3	1 ₂ Amount
1	Carlota	5/5/2016	West	485
2	Aspen	4/2/2016	South	804
3	Yanaki	5/2/2016	South	790
4	FlatTop	4/11/2016	South	965
5	Carlota	5/21/2016	South	620
6	Carlota	5/15/2016	East	305
7	Carlota	5/19/2016	West	426
8	Yanaki	5/6/2016	East	319

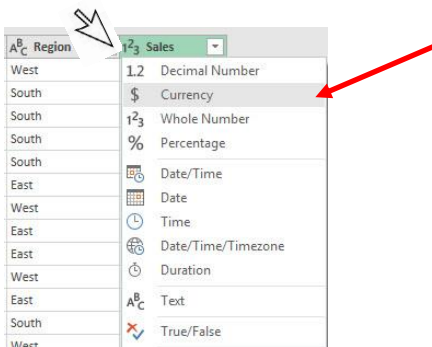
13) Notice that the Column Header / Field Names are not useful names.

- 14) To rename each Column Header / Field, double click the Column Header / Field Name, type name and hit Enter.
The renamed columns should look like this:

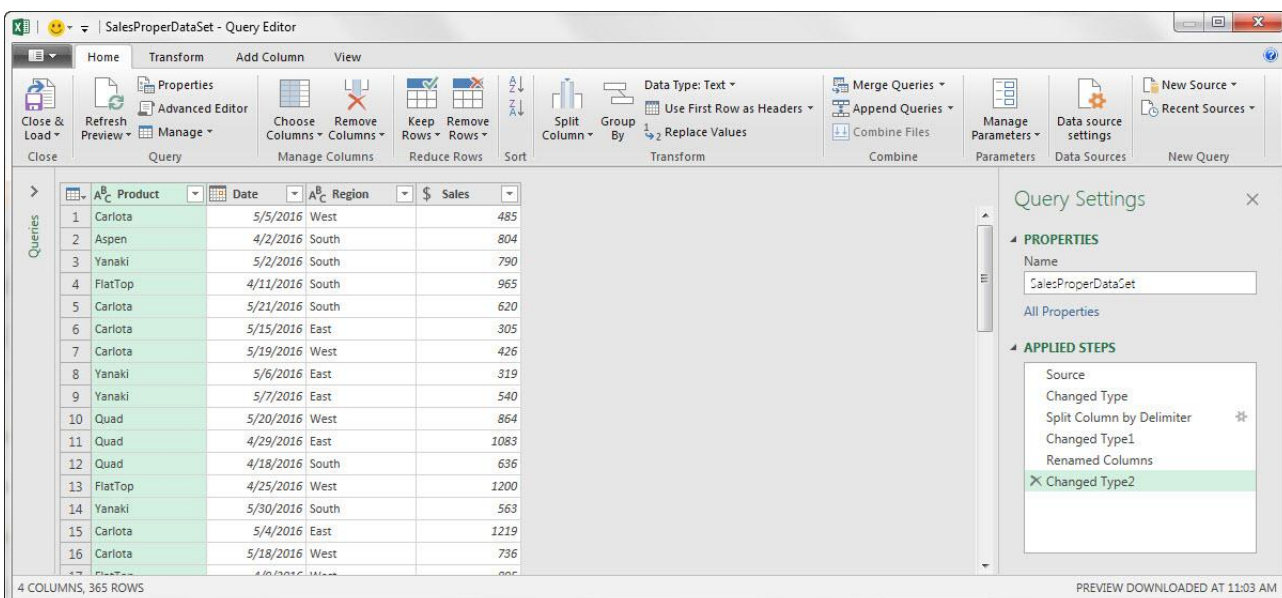


	Product	Date	Region	123 Sales
1	Carlota	5/5/2016	West	485
2	Aspen	4/2/2016	South	804
3	Yanaki	5/2/2016	South	790
4	FlatTop	4/11/2016	South	965
5	Carlota	5/21/2016	South	620
6	Carlota	5/15/2016	East	305
7	Carlota	5/19/2016	West	426
8	Yanaki	5/6/2016	East	319

- 15) Now we want to take a closer look at each Column Header / Field Name and notice the icons like “ABC”, Calendar Icon and “123”. These are the Data Types for each column. These Data Types assure that the right kind of data goes into the column. The Product, Date and Region all have the correct Data Type. But the Sales Field needs to have the Data Type changed to “Currency” Data Type. To do this, click the “123” icon on the left side of the Sales Column Header and then click on Currency, like this:



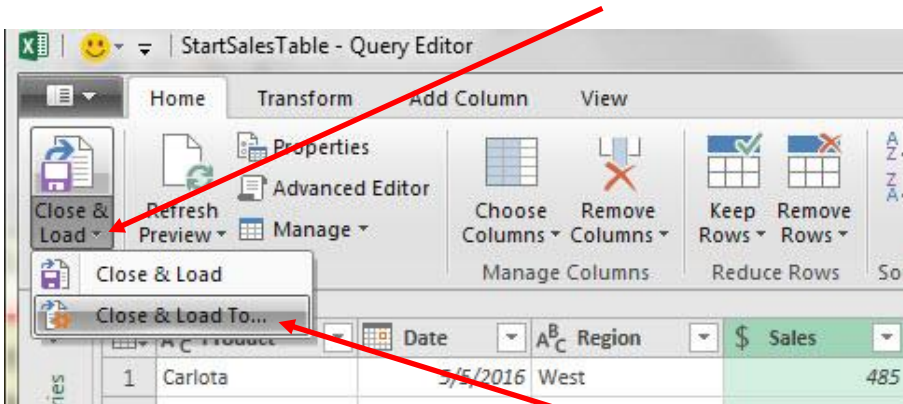
- 16) Now our Query is finished. Notice that the steps of the query are listed on the right. These steps can be viewed by clicking on them and can be edited if necessary.



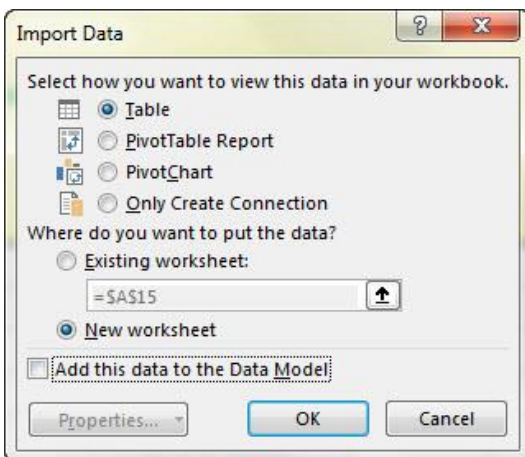
	Product	Date	Region	\$ Sales
1	Carlota	5/5/2016	West	485
2	Aspen	4/2/2016	South	804
3	Yanaki	5/2/2016	South	790
4	FlatTop	4/11/2016	South	965
5	Carlota	5/21/2016	South	620
6	Carlota	5/15/2016	East	305
7	Carlota	5/19/2016	West	426
8	Yanaki	5/6/2016	East	319
9	Yanaki	5/7/2016	East	540
10	Quad	5/20/2016	West	864
11	Quad	4/29/2016	East	1083
12	Quad	4/18/2016	South	636
13	FlatTop	4/25/2016	West	1200
14	Yanaki	5/30/2016	South	563
15	Carlota	5/4/2016	East	1219
16	Carlota	5/18/2016	West	736

Query Settings
PROPERTIES
Name: SalesProperDataSet
APPLIED STEPS
Source
Changed Type
Split Column by Delimiter
Changed Type1
Renamed Columns
X Changed Type2

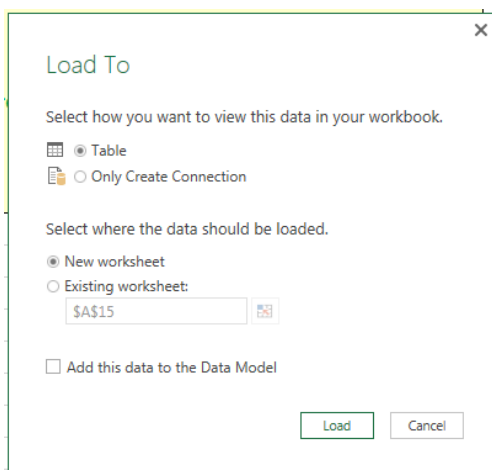
- 17) To load the Cleaned & Transformed Proper Data Set to an Excel Sheet, in the Home Ribbon Tab, in the Close group (all the way on the left), click the Close & Load drop-down arrow and then click on the “Close & Load To...” option



- 18) The next dialog box that pops-up depends on what version you have:
- i. If you have Office 365, then your dialog box looks like this and is named “Import Data”:



- ii. If you do NOT have Office 365, then your dialog box looks like this and is named “Load To”:



- iii. In either dialog box you now select “Table” and “New Worksheet” dialog buttons.
- iv. By selecting the “Table” option we are create a new Excel Table on a new sheet. This Excel Table will be have the same name as the Query and will be the Power Query Output or Result.

19) After the Cleaned & Transformed Proper Data Set is loaded to an Excel Sheet, we can see that the name of the Query and the name of the new Excel Table are the same. They are both named SalesProperDataSet “. After the data is loaded, be sure to name the sheet “Cleaned Data”. Notice that the Query Pane shows that 365 records have been loaded.

Table Name: SalesProperDataSet

Queries & Connections

1 query

SalesProperDataSet
365 rows loaded.

Same Name for both:
1) Query
And
2) Cleaned Excel Table that is loaded to a new sheet
** It is simultaneously both an Excel Table and the
Output from Power Query

Be sure to name sheet
“Cleaned Data”

This is called the
Query Pane

	Product	Date	Region	Sales
1	Carlota	5/5/2016	West	485
2	Aspen	4/2/2016	South	804
3	Yanaki	5/2/2016	South	790
4	FlatTop	4/11/2016	South	965
5	Carlota	5/21/2016	South	620
6	Carlota	5/15/2016	East	305
7	Carlota	5/19/2016	West	426
8	Yanaki	5/6/2016	East	319
9	Yanaki	5/7/2016	East	540
10	Quad	5/20/2016	West	864
11	Quad	4/29/2016	East	1083
12	Quad	4/18/2016	South	636
13	FlatTop	4/25/2016	West	1200
14	Yanaki	5/30/2016	South	563

20) Now we can build a PivotTable Report from our cleaned data:

21) Back on the sheet “Ex(1)”, select the new data (new records) in the range AA15:AB130, like in this picture:

	Z	AA	AB	AC
10				
11		New Records:		
12				
13		Description	Amount	
14				
15		Yanaki / 07/03/2016 / South	1071	
16		Quad / 06/12/2016 / East	606	
17		Quad / 06/07/2016 / East	842	
18		Yanaki / 06/29/2016 / East	1232	
19		Yanaki / 07/20/2016 / West	1165	
20		Aspen / 07/24/2016 / South	1165	
21		Quad / 07/24/2016 / South	1142	
22		Sunset / 06/02/2016 / South	405	
23		FlatTop / 07/31/2016 / South	403	
24		Quad / 06/21/2016 / West	1194	
25		Carlota / 07/14/2016 / South	588	
26		Carlota / 07/14/2016 / South	1143	

22) Select cell A380 below the original data set, the Excel Table named, “StartSalesTable”, like in this picture:

	A	B	C
377	Quad / 04/08/2016 / East	432	
378	FlatTop / 04/14/2016 / West	484	
379	FlatTop / 05/11/2016 / East	903	
380			
381			

23) Paste the copied new records so they become part of the Excel Table named StartSalesTable”, like this:

	Description	Amount
7	Quad / 04/08/2016 / East	432
8	FlatTop / 04/14/2016 / West	484
9	FlatTop / 05/11/2016 / East	903
0	Yanaki / 07/03/2016 / South	1071
1	Quad / 06/12/2016 / East	606
2	Quad / 06/07/2016 / East	842
3	Yanaki / 06/29/2016 / East	1232
4	Yanaki / 07/20/2016 / West	1165
5	Aspen / 07/24/2016 / South	1165
6	Quad / 07/24/2016 / South	1142
7	Sunset / 06/02/2016 / South	405
8	FlatTop / 07/31/2016 / South	403
9	Quad / 06/21/2016 / West	1194
0	Carlota / 07/14/2016 / South	588

24) Go back to the Query Output, the Excel Table “SalesProperDataSet” with the cleaned data, and then right-click and click on the “Refresh” option. This will update the Power Query Output, which means it will run through all the listed steps in the Power Query window and re-load the complete new data set.

	A	B
1	Product	Date
2	Carlota	5/5/2
3	Aspen	4/2/2
4	Yanaki	5/2/2
5	FlatTop	4/11/2
6	Carlota	5/21/2
7	Carlota	5/15/2
8	Carlota	5/19/2
9	Yanaki	5/6/2
10	Yanaki	5/7/2

25) Then in the PivotTable, right-click and click on the “Refresh” option.

A screenshot of a PivotTable context menu. The menu is open, showing options like Copy, Format Cells..., Refresh, Sort, and Filter. The 'Refresh' option is highlighted. The background shows a PivotTable with columns for Product, Region, and Sales, and rows for months (Apr, May, Jun, Jul).

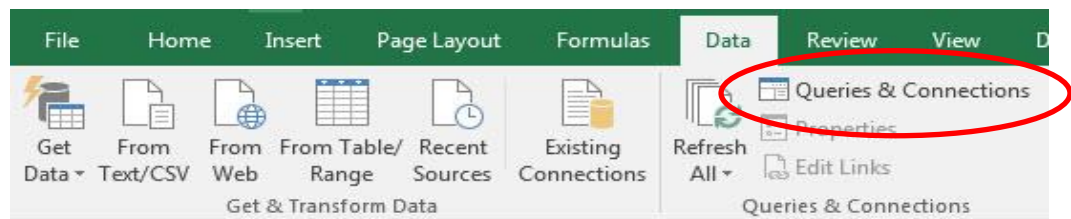
26) The completed project should look like this:

A screenshot of the completed project. It shows a PivotTable with columns for Product, Region, and Sales, and rows for months (Apr, May, Jun, Jul). The PivotTable Fields task pane is visible on the right, showing the fields for the PivotTable. The PivotTable data is as follows:

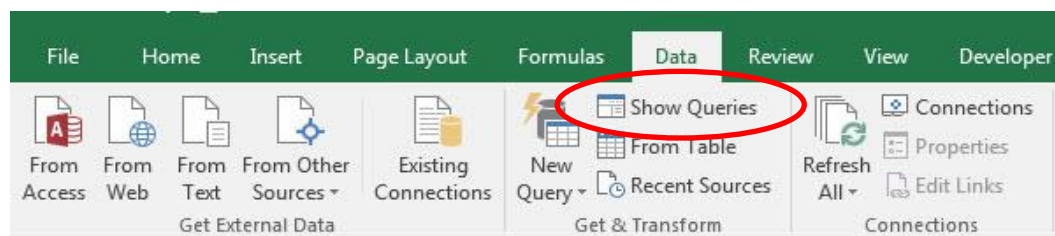
Month	Product	Region	Sum of Sales
Apr	Aspen	East	\$7,913.00
	Carlota	East	\$11,219.00
	FlatTop	East	\$9,171.00
	Quad	East	\$28,303.00
	Yanaki	East	\$21,829.00
May	Aspen	East	\$8,791.00
	Carlota	East	\$7,547.00
	FlatTop	East	\$10,616.00
	Quad	East	\$26,954.00
	Yanaki	East	\$17,058.00
Jun	Aspen	East	\$8,407.00
	Carlota	East	\$6,012.00
	FlatTop	East	\$2,639.00
	Quad	East	\$14,653.00
	Yanaki	East	\$18,077.00
Jul	Aspen	East	\$5,098.00
	Carlota	East	\$5,942.00
	FlatTop	East	\$3,613.00
	Quad	East	\$14,653.00
	Yanaki	East	\$18,077.00

27) If you ever need to edit, or delete or look at your query, you must open the Query Pane. The button to click to open the Query Pane is in a different location depending on whether or not you have Office 365:

- i. In Office 365, you click the “Queries and Connections” button in the Queries and Connections group in the Data Ribbon Tab:



- 1.
- ii. In a version that is NOT Office 365, you click the “Show Queries” button in the Get & Transform group in the Data Ribbon Tab:



- 1.
- iii. If you ever need to edit or look at your query, you can double-click the query in the Query Pane. You can right-click the query and point to edit also.
- iv. If you need to delete your query and start over, you can delete the query in the Queries Pane.

Example #2: Import Multiple Text Files & Transform into Proper Data Set

- 1) After you download the zipped folder named “EB25FileDownloads” to the desktop (or some other location) and then unzip the folder, you will see that there are two text files and one folder named “Start” inside, like in this picture:



- 2) If you look inside the Start Folder, you will see two Text Files. This is the folder where we store our Text File data and we only put “.txt” files in this folder. Each file contains the Sales Data for a single month. As seen in this picture:



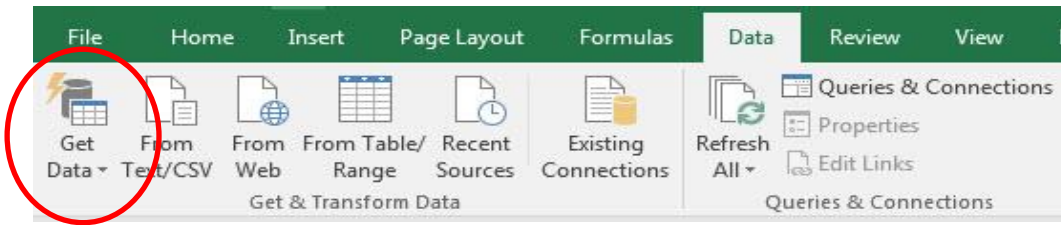
Start Folder is where we store or monthly sales data. We only store “.txt” files in this folder.

- 3) If you right-click the Text File named “01-April.txt” and click on “Open With”, then click on “Word” you will see this:

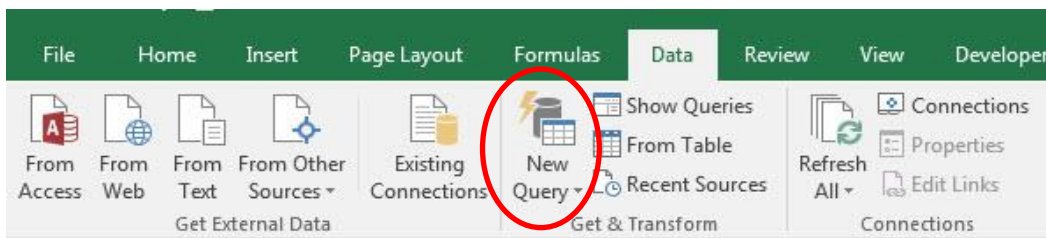
```
Product → Date → Region → Sales
Aspen→4/1/2016 → West→ 471
Aspen→4/1/2016 → West→1007
Quad→ 4/1/2016 → West→697
Carlota → 4/1/2016 → West→286
Aspen→4/1/2016 → West→450
Yanaki → 4/2/2016 → West→551
Aspen→4/2/2016 → South→804
Carlota → 4/2/2016 → East→889
Sunset → 4/3/2016 → South→789
Quad→ 4/3/2016 → South→1140
Sunset → 4/3/2016 → South→1096
Quad→ 4/3/2016 → South→1064
Aspen→4/3/2016 → South→652
Sunset → 4/3/2016 → East→391
Quad→ 4/3/2016 → East→447
FlatTop → 4/4/2016 → East→368
Quad→ 4/4/2016 → West→385
```

- 4) Each one of the Text Files has four columns of data, where each column is separated by a Tab delimiter. This means that the Text File contains the columns Product, Date, Region and Sales and each column is separated by a Tab. Tab delimited data is a common way that different systems can share data. This data came from a data storage system that exported monthly data in a Tab Delimited form. This is good for us because Power Query can easily interpret Tab Delimited data and can easily combine the different Text Files into a single Proper Data Set.

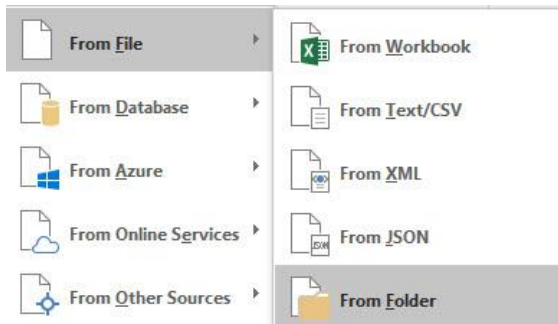
- 5) Our goal is to combine (sometimes referred to as “append”) the text files into a single Proper Data Set.
- 6) Go to the sheet named “Ex(2)” and select cell A11. Then in the Data Ribbon Tab, in the Get & Transform Group start a new query to import the Text Files from a Folder. How we start a new query to import from a folder depends on the version you have:
 - i. In Office 365, you click the “Get Data” button in the Get & Transform Data group in the Data Ribbon Tab:



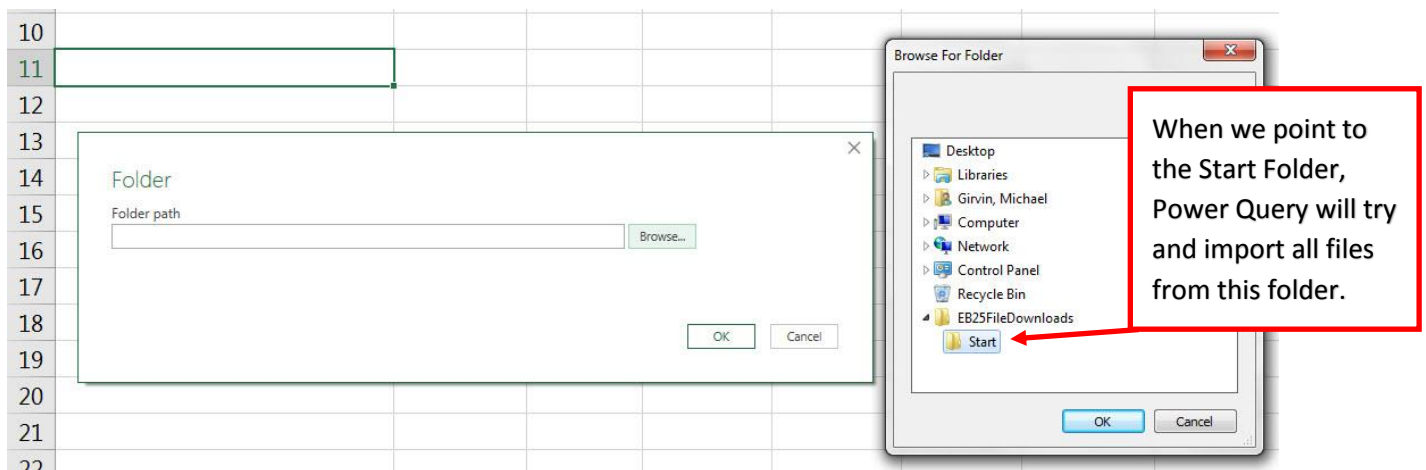
- ii. In a version that is NOT Office 365, you click the “New Queries” button in the Get & Transform group in the Data Ribbon Tab:



- 7) From the drop-down, you select “From File” and then click on “From Folder”, like in this picture:

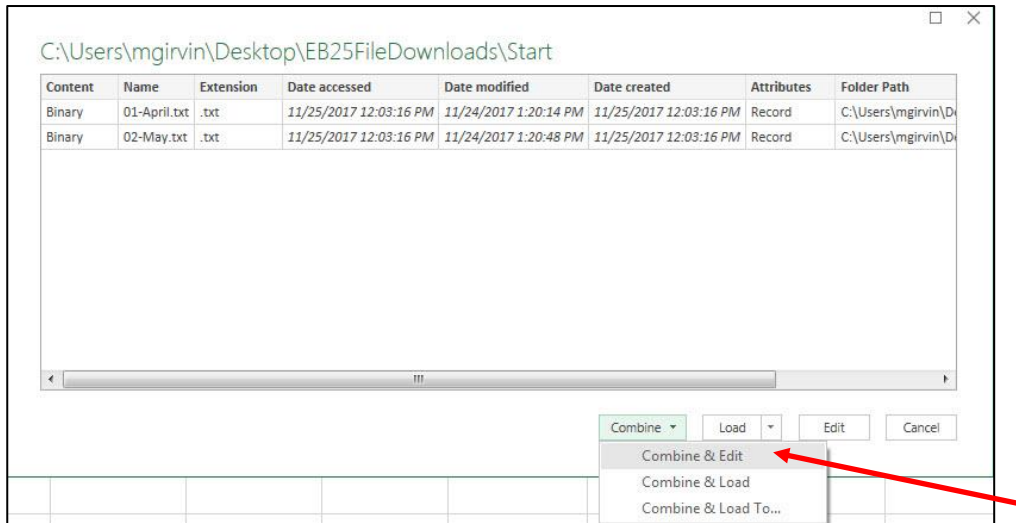


- 8) Click in the Browse textbox, then click on the Browse button, and navigate to the Start Folder, like in this picture:

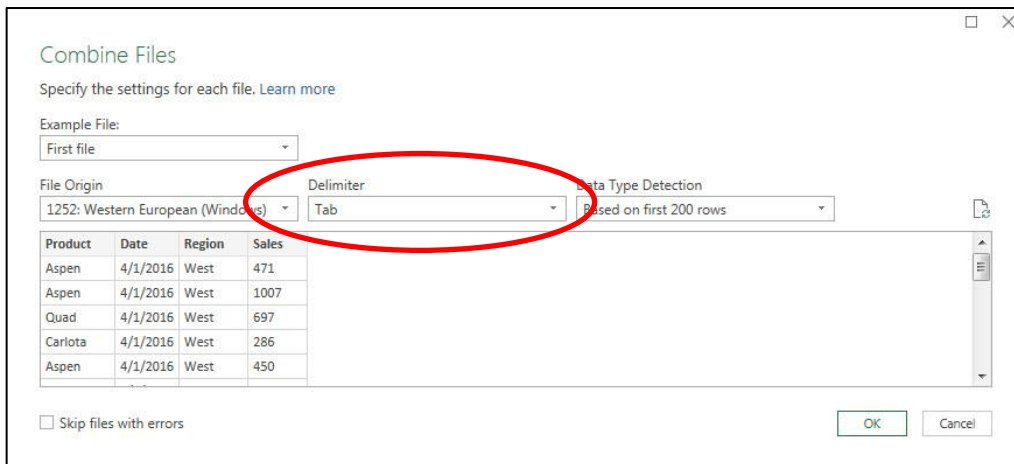


- 9) Click OK in the “Browse For Folder” dialog box. Click OK in the “Folder” dialog box.

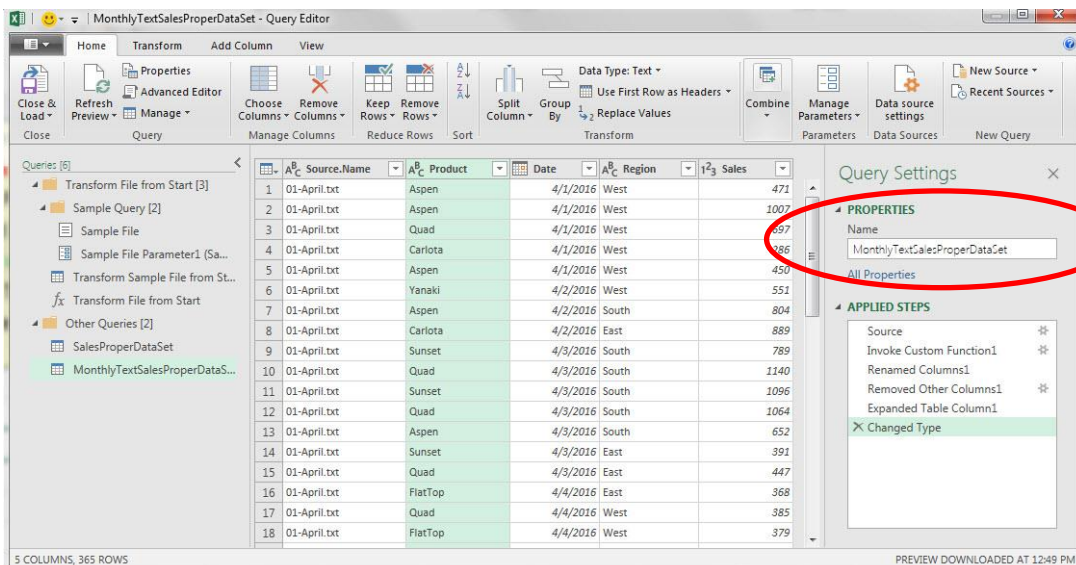
10) In the next dialog box, select the Combine button drop-down and then select “Combine and Edit”, like in this picture:



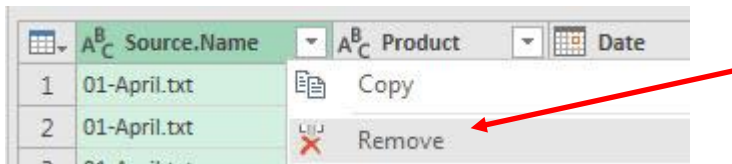
11) In the next dialog box, make sure that the Delimiter is Tab. If it is, click OK.



12) When the Query Editor comes up, name the query “MonthlyTextSalesProperDataSet”, as seen here:



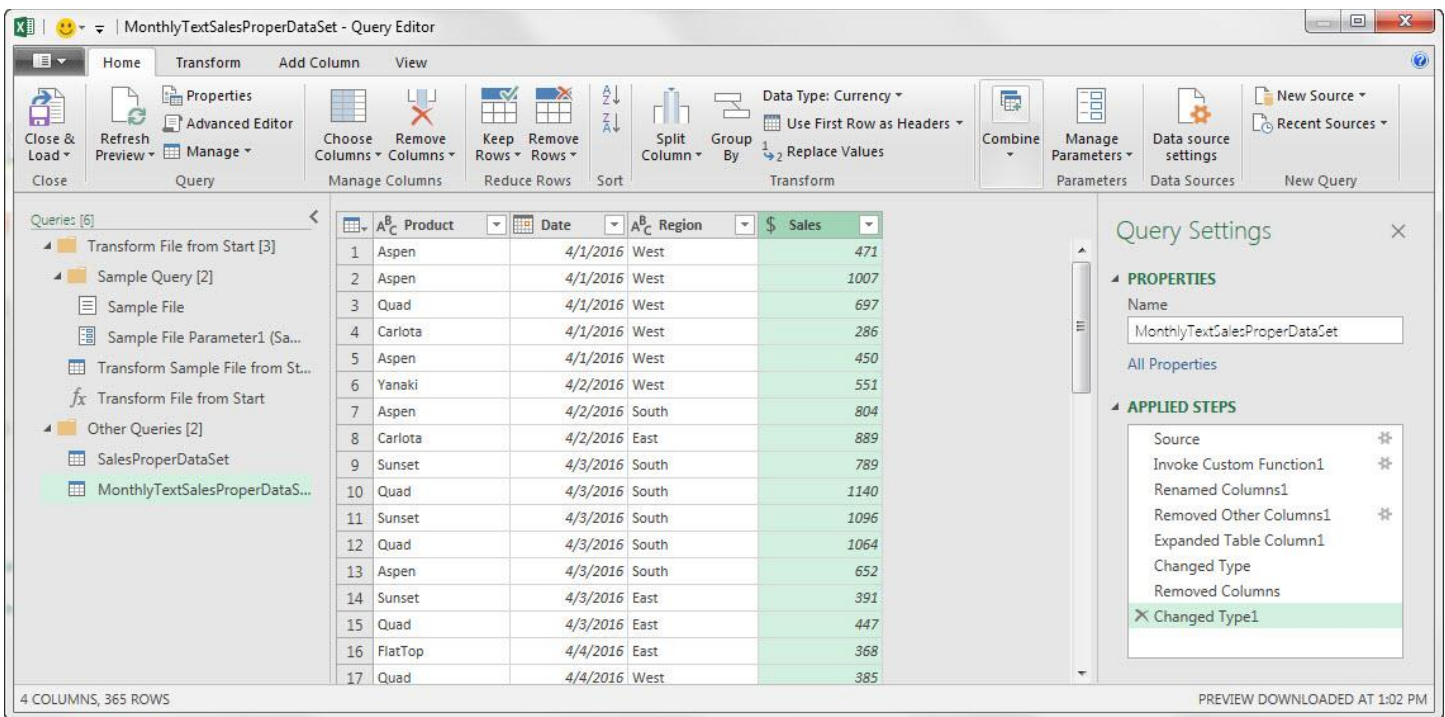
13) Right-click Source.Name Column Header and then click on Remove, like seen here:



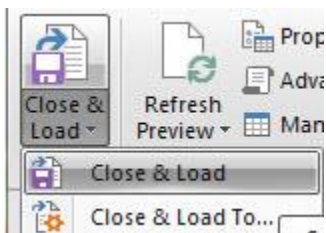
14) Change the Data Type for the Sales Field to Currency, like in this picture:



15) The finished query should look like this:



16) From the Home Ribbon Tab, in the Close group, click the Close & Load drop-down and then click on Close & Load option, like seen here:



- 17) When the new table loads, be sure to name the sheet “CleanedTextData”. In the Queries Pane you can see a lot of steps that were automatically created to combine the text files. At the bottom of the Queries Pane you can see the two queries that we created and you can see the number of rows that were loaded.

Product Date Region Sales

Aspen	4/1/2016	West	471
Aspen	4/1/2016	West	1007
Quad	4/1/2016	West	697
Carlota	4/1/2016	West	286
Aspen	4/1/2016	West	450
Yanaki	4/2/2016	West	551
Aspen	4/2/2016	South	804
Carlota	4/2/2016	East	889
Sunset	4/3/2016	South	789
Quad	4/3/2016	South	1140
Sunset	4/3/2016	South	1096
Quad	4/3/2016	South	1064
Aspen	4/3/2016	South	652
Sunset	4/3/2016	East	391
Quad	4/3/2016	East	447
FlatTop	4/4/2016	East	368
Quad	4/4/2016	West	385
FlatTop	4/4/2016	West	379
Aspen	4/4/2016	West	733
Carlota	4/4/2016	East	597
Aspen	4/4/2016	South	1205
Aspen	4/5/2016	South	858
Sunset	4/5/2016	West	399
Sunset	4/5/2016	East	1108
Sunset	4/5/2016	South	560
Quad	4/5/2016	South	1037
Carlota	4/5/2016	East	935

Queries & Connections

Queries | Connections

6 queries

- Transform File from Start [3]
- Sample Query [2]
 - Sample File Parameter1 (Samp... Connection only.
 - Sample File Connection only.
- Transform Sample File from Start Connection only.
- Transform File from Start Connection only.
- Other Queries [2]
 - SalesProperDataSet 481 rows loaded.
 - MonthlyTextSalesProperDataSet 365 rows loaded.

Ex(1) CleanedTextData Ex(2)

Ready

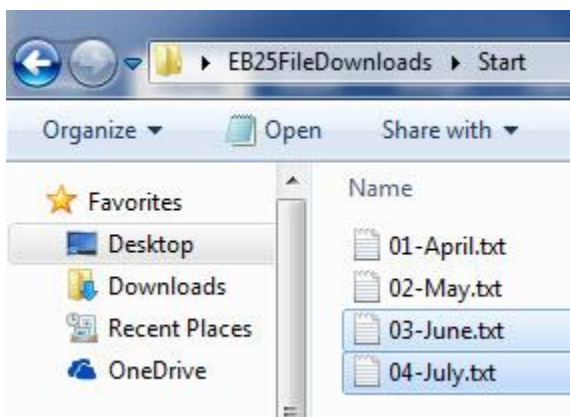
100%

Lots of steps automatically created by Power Query to combine Text Files.

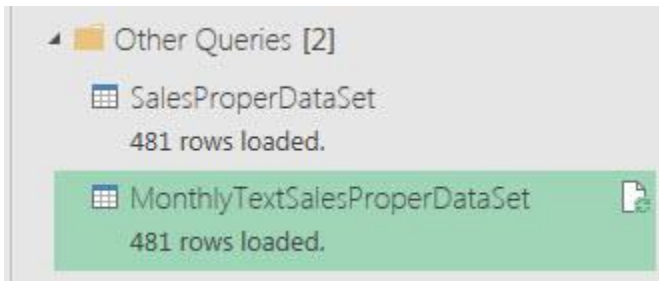
Name the sheet “CleanedTextData”

These are the two queries that we created

- 18) Now go back to the “EB25FileDownloads” folder and copy the June and July text files and then paste them into the “Start” folder, like seen here:



19) Now go back to the “MonthlyTextSalesProperDataSet” Power Query Output and right-click and then click on the “Refresh” option. In the Queries Pane you will see that the new Text Files for June and July have been included in the Table:



20) Revised Definition of Data Analysis:

- i. Define Data Analysis / Business Intelligence:
 1. Convert the raw data into useful information.
 - i. If the data is not in a Proper Data Set and we need it to be (like for PivotTables and Filtering), then the full definition is: Import, Clean and Transform Data into a Proper Data Set and then convert the raw data into useful information.
 2. Here is a picture:

Define: Data Analysis / Business Intelligence:

