

Attempt 1

In Progress  
NEXT UP: Submit Assignment

Add Comment

## Unlimited Attempts Allowed

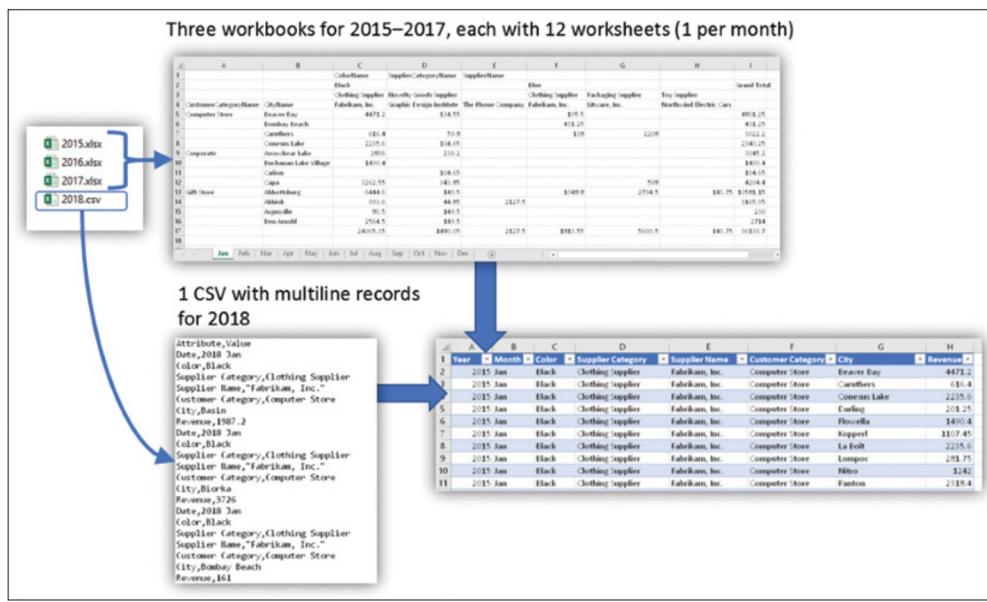
### Details

You have come a long way. In the preceding chapters of this book, you have learned how to clean messy datasets, combine multiple tables, resolve mismatched column names, unpivot and pivot data, and tackle many other data challenges. It's time to put all that knowledge to the test and apply your new data-wrangling powers on a large-scale challenge.

## Exercise 14-1: Saving the Day at Wide World Importers

Imagine that you are the new chief data officer at Wide World Importers. Last weekend, the company experienced a massive cyber-attack that was targeted at the company's data warehouse. A team of specialists is trying to assess the risks to the revenue data, and you have been asked to reconstruct a new summary report of the company's revenues from old exported data, which was not affected by the cyber-attack. Your predecessor kept reports of the company's revenues in three Excel workbooks for the years 2015–2017. Each of the workbooks has 12 worksheets—one for each month of the year. In addition, 2018 revenues were recently exported in a new challenging format.

Can you import all the data from each of the 36 total worksheets and combine it with the data of 2018 to create a single table, similar to the one in [Figure 14-1](#) (<https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14fig01>)?



**FIGURE 14-1** Your challenge is to combine multiple tables of Wide World Importers from three workbooks for 2015–2017 revenues. Each with 12 worksheets (1 per month), and an additional CSV file with multiline records for 2018 revenues.

To start the exercise, download [C14E01 - Goal.xlsx](#). (<https://mycourses.unh.edu/courses/115726/files/10027963?wrap=1>) [Download](#) ([https://mycourses.unh.edu/courses/115726/files/10027963/download?download\\_frd=1](https://mycourses.unh.edu/courses/115726/files/10027963/download?download_frd=1))

Review each of the workbooks [2015.xlsx](#) (<https://mycourses.unh.edu/courses/115726/files/10027965?wrap=1>) [Download](#) ([https://mycourses.unh.edu/courses/115726/files/10027965/download?download\\_frd=1](https://mycourses.unh.edu/courses/115726/files/10027965/download?download_frd=1)), [2016.xlsx](#) (<https://mycourses.unh.edu/courses/115726/files/10027964?wrap=1>) [Download](#) ([https://mycourses.unh.edu/courses/115726/files/10027964/download?download\\_frd=1](https://mycourses.unh.edu/courses/115726/files/10027964/download?download_frd=1)), and [2017.xlsx](#) (<https://mycourses.unh.edu/courses/115726/files/10027961?wrap=1>) [Download](#) ([https://mycourses.unh.edu/courses/115726/files/10027961/download?download\\_frd=1](https://mycourses.unh.edu/courses/115726/files/10027961/download?download_frd=1)), as well as their worksheets. Each worksheet includes a 2x3 summarized table, with the customer categories and cities in the rows and product colors, supplier categories, and supplier names in the columns.

For 2018, the revenues data is stored in a comma-separated values (CSV) file with the columns Attribute and Value, as illustrated in the bottom left of [Figure 14-1](#) (<https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14fig01>). Your goal is to combine all the data together to build a cohesive report for Wide World Importers revenues, as shown in C14E01 - Goal.xlsx.



### Note

[Previous](#)

(<https://mycourses.unh.edu/courses/115726/modules/items/2904621>)

Submit Assignment

[Next](#)

(<https://mycourses.unh.edu/courses/115726/modules/items/2904624>)

## Clues

Not sure how to proceed? You can break down the challenge into subtasks:

1. Import the folder and keep the Excel workbooks.
2. Create a new column with the 36 rows of Table objects (comprised of the 36 worksheets from the three workbooks with the 12 worksheets of revenues for each month of the years 2015–2017).
3. For each worksheet, skip the first row and remove the Grand Total column and row.
4. Apply the function *FnUnpivotSummarizedTable* on each worksheet. (Refer to [Chapter 7](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07), “[Advanced Unpivoting and Pivoting of Tables](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07)”.)
5. Include the 2018.csv file after you follow the Pivot sequence covered in [Chapter 7](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07), “[Combining Data from Multiple Sources](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07)” to refresh your memory on column name normalization.



### Note

You are encouraged to stop reading for now. Try to solve this challenge by yourself. Spend a few more hours trying it. If you need help, read [Chapter 7](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07), “[Advanced Unpivoting and Pivoting of Tables](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07)” again. If you need further help, you can follow [Exercise 14-1](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1), step by step.

### Part 1: Starting the Solution

To transform the pivoted data from the years 2015–2017 and append to it the multiline records of 2018, you can follow the steps in this exercise.

1. Download files and create C14E01 folder
2. Open a blank workbook or Power BI report:
  - In Excel:** On the Data tab, select Get Data, From File, From Folder.
  - In Power BI Desktop:** Select Get Data, File, Folder.
3. Select the folder path C:\C14E01 and click OK. In the next dialog box, click Edit. The Power Query Editor opens.
4. To filter out the file [2018.csv](https://mycourses.unh.edu/courses/115726/files/10027966?wrap=1) ([https://mycourses.unh.edu/courses/115726/files/10027966?download?download\\_frd=1](https://mycourses.unh.edu/courses/115726/files/10027966?download?download_frd=1)), which has a different format from the other files (you will handle it later), select the filter control on the Name column and select 2018.csv. Then click OK to close the Filter pane.
5. Rename the query from C14E01 to *Revenues*.
6. Click the Combine Files icon in the Content column header.
7. When the Combine Files dialog box opens, right-click Sample File Parameter1 and select Edit.
8. When Power Query Editor opens and Revenues is selected, remove the last three columns.
9. Remove the .xlsx extension from the *Source.Name* values to extract the year and rename the column *Year*. (There are multiple ways to remove the file extension. If you are not sure how to do it, revisit [Chapter 3](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch03.xhtml#ch03), “[Combining Data from Multiple Sources](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch03.xhtml#ch03)”.)
10. Rename the column from *Name* to *Month*.

In the Data column you have the Table objects, each with the summarized tables. Before you apply the unpivot sequence, you should remove the first row, which doesn't include any valuable data and will not work well if you apply the Unpivot sequence on it. You should also remove the Grand Totals at the last column and row. The following steps walk through how to accomplish these.

11. In the Add Column tab, select Custom Column. The Custom Column dialog box opens. Make the following changes:

1. Enter *Skip First Row* in the New Column Name box.
2. Set the following formula in the Custom Column Formula box and then click OK:

```
= Table.Skip([Data], 1)
```

The new column returns the table object in the Data column after the first row is skipped.

3. Remove the Data column.

12. On the Add Column tab, select Custom Column. The Custom Column dialog box opens. Make the following changes:

[Previous](#)

Submit Assignment

[Next](#)

(<https://mycourses.unh.edu/courses/115726/modules/items/2904621>) (<https://mycourses.unh.edu/courses/115726/modules/items/2904624>)

2. Set the following formula in Custom Column Formula and then click OK:

[Click here to view code image](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14_images.xhtml#e0378-01a) (https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14\_images.xhtml#e0378-01a)

```
= Table.RemoveColumns(  
    [Skip First Row], {  
        List.Last(Table.ColumnNames([Skip First Row]))  
    })
```

This formula removes the last column. You use `Table.ColumnNames` to retrieve the column names of each table object in Data column and apply `List.Last` to get the last column name.

3. Remove the Skip First Row column.

13. On the Add Column tab, select Custom Column. The Custom Column dialog box opens. Make the following changes:

1. Set New Column Name to Summarized Table.

2. Set the following formula in Custom Column Formula and then click OK:

[Click here to view code image](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14_images.xhtml#e0378-02a) (https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14\_images.xhtml#e0378-02a)

```
= Table.RemoveLastN([No Grand Total Column], 1)
```

3. Remove the No Grand Total Column column.



#### Note

An alternative way to clean the summarized tables, instead of following steps 11–13, is to use a single function. You can create a blank query with the following code:

[Click here to view code image](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14_images.xhtml#e0378-03a) (https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14\_images.xhtml#e0378-03a)

```
(inputTable) =>  
let  
    SkipFirstRow = Table.Skip(inputTable, 1),  
    NoGrandTotalColumn = Table.RemoveColumns(  
        SkipFirstRow, {  
            List.Last(Table.ColumnNames(SkipFirstRow))  
        })  
    ,  
    Result = Table.RemoveLastN(NoGrandTotalColumn, 1)  
in  
    Result
```

Name the query `FnCleanSummarizedTable` and then invoke it in the Revenues query, by selecting Invoke Custom Column on the Add Column tab and applying it on the Data column.

## Part 2: Invoking the Unpivot Function

In this portion of the [Exercise 14-1](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1) (https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1) solution, you will apply the custom function `FnUnpivotSummarizedTable`. Recall that in [Exercise 7-3](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07lev6) (https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07lev6), you created that function to unpivot summarized tables. At this point in [Exercise 14-1](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1) (https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1), you are now ready to unpivot the summarized Wide World Importers table for the years 2015–2017. Follow these steps to continue [Exercise 14-1](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1) (https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1):

14. Follow these steps to copy `FnUnpivotSummarizedTable` from [Exercise 7-3](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07lev6) (https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07lev6):

1. Save your current report and keep it open.
2. Open the workbook C07E03 - Solution.xlsx from <https://aka.ms/DataPwrBIPivot/downloads> (https://aka.ms/DataPwrBIPivot/downloads).
3. On the Data tab, select Queries & Connections.
4. In the Queries & Connections pane, right-click `FnUnpivotSummarizedTable` and select Copy.
5. Go back to your current [Exercise 14-1](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1) (https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1) workbook or report and launch the Power Query Editor.
6. Right-click anywhere in the background space of the Queries pane and select Paste to copy the `FnUnpivotSummarizedTable` function.

15. In the Queries pane, select the Revenues query.

16. On the Add Column tab, select Custom Column. The Custom Column dialog box opens. Make the following changes:

1. Enter `Unpivoted` in the New Column Name box.

[Previous](#)

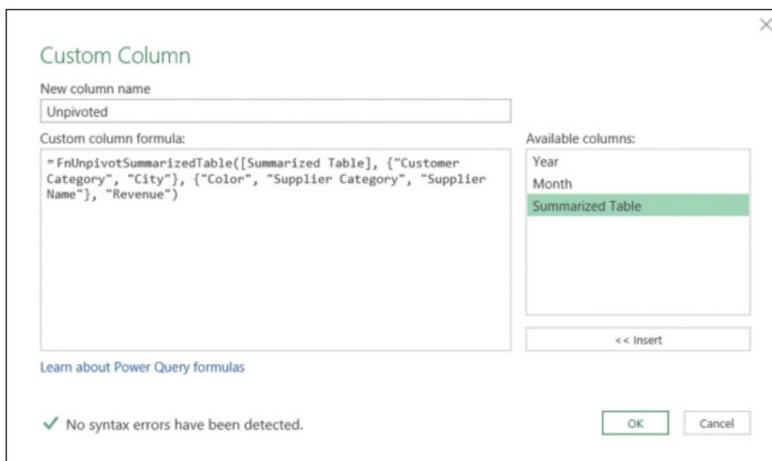
(https://mycourses.unh.edu/courses/115726/modules/items/2904621)

Submit Assignment

[Next](#)

(https://mycourses.unh.edu/courses/115726/modules/items/2904624)

```
= FnUnpivotSummarizedTable([Summarized Table], {"Customer Category", "City"}, {"Color", "Supplier Category", "Supplier Name"}, "Revenue")
```



**FIGURE 14-2** To unpivot the summarized table, invoke the *FnUnpivotSummarizedTable* function by using the Custom Column dialog box.

3. Remove the Summarized Table column.
  17. Expand the Unpivoted column. In the Expand pane, deselect Use Original Column Name As Prefix and click OK.
  - You can see in the Preview pane that all the worksheets have been unpivoted correctly. It's time to handle the 2018 revenues.
- Part 3: The Pivot Sequence on 2018 Revenues**
- Recall that in step 4 of this exercise, you filtered out the file 2018.csv. This file contains comma-separated values of attribute/value pairs, and you can assume that each multi-line revenue record in 2018.csv file starts with the Date attribute. In this part of the exercise, you need to import it and apply the Pivot sequence that you learned in [Chapter 7](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07) (<https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch07.xhtml#ch07>). Follow these steps to continue [Exercise 14-1](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1) (<https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1>):
18. Import the file 2018.csv in the Power Query Editor by selecting New Source on the Home tab. Rename the new query *2018 Revenues*.
  19. On the Transform tab, select Use First Row as Headers.
  20. On the Add Column tab, select Index Column.
  21. On the Add Column tab, select Conditional Column. When the Add Conditional Column dialog box opens, make the following changes:

1. Enter *Row ID* in the New Column Name box.
2. Set Column Name to Attribute.
3. Keep equals as the operator.
4. Enter *Date* in the Value box.
5. Set Output to Select a Column and select the Index column.
6. Enter *null* in the Otherwise box.
7. Click OK.

22. Select the Row ID column, and on the Transform tab, select Fill, Down. Delete the Index column.
  23. Select the Attribute column, and on the Transform tab, select Pivot Column. When the Pivot Column dialog box opens, make the following changes:
1. Set Values Column to Value.
  2. Expand Advanced Options.
  3. Set Aggregate Value Function to Don't Aggregate.
  4. Click OK.

24. Remove the Row ID column.

You have now completed the Pivot sequence for 2018 revenues, and it's time to combine the results.

#### Part 4: Combining the 2018 and 2015–2017 Revenues

The steps you need to take in this portion of the exercise are discussed in [Chapter 4](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch04.xhtml#ch04) (<https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch04.xhtml#ch04>), “Combining Mismatched Tables” (<https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch04.xhtml#ch04>). You now have

[◀ Previous](#)

[Submit Assignment](#)

[Next ▶](#)

(<https://mycourses.unh.edu/courses/115726/modules/items/2904621>)

(<https://mycourses.unh.edu/courses/115726/modules/items/2904624>)

26. When the Merge Columns dialog box opens, select Space as Delimiter, enter Date in the New Column Name box, and click OK.

27. On the Home tab, select Append Queries. When the Append dialog box opens, set Table to Append to 2018 Revenues and click OK.



#### Note

In step 16, you declared the column names to use for the unpivoted revenues of the years 2015–2017. If you used different names as Row fields or Column fields arguments in `FnUnpivotSummarizedTable`, you have additional mismatching column names at this stage. To fix this, use the same column names as in step 16 or rename 2018 Revenues to match the Revenues query before the append.

28. Change the type of the Revenue column to Currency or Decimal Number, and change the type of the Date column to Date.

29. Finally, load the Revenues query to your report, and disable the load for the 2018 Revenues query:

**In Excel:** On the Data tab, select Queries & Connections. Right-click 2018 Revenues and select Load To. In the Import Data dialog box that opens, select Only Create Connection and click OK.

**In Power BI Desktop:** In the Power Query Editor, right-click the 2018 Revenues query in the Queries pane and deselect Enable Load.

You can review the solution files C14E01 - Solution.xlsx and C14E01 - Solution.pbix, which are available at <https://aka.ms/DataPwrBIPivot/downloads> (<https://aka.ms/DataPwrBIPivot/downloads>).

## Exercise 14-2: Comparing Tables and Tracking the Hacker

Thanks to your success in [Exercise 14-1](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1) (<https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1>), you were able to create the revenues report, which helped your company, Wide World Importers make informed decisions to keep the business running after the cyber-attack. According to the Cybersecurity team's investigation, the attackers tampered with the revenues values in the data warehouse.

Your task now is to compare the summarized revenues from [Exercise 14-1](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1) (<https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1>) against the compromised data in [C14E02 - Compromised.xlsx](https://mycourses.unh.edu/courses/115726/files/10027962?wrap=1) (<https://mycourses.unh.edu/courses/115726/files/10027962?wrap=1>) ([https://mycourses.unh.edu/courses/115726/files/10027962/download?download\\_frd=1](https://mycourses.unh.edu/courses/115726/files/10027962/download?download_frd=1)) with the goal of identifying which records were tampered by the hacker and are compromised. Can you find the records that contain the modified revenues? Can you find the hacker's message in the data?

Try to resolve this challenge by creating a query that compares C14E01 - Solution.xlsx and C14E02 - Compromised.xlsx. While you may be able to find the differences in the tables by using PivotTables or DAX measures, focus on finding a solution using the Power Query Editor.

#### Clues

If you are not sure how to tackle this challenge, here are a few clues to help you get started:

1. The hacker made changes in the Revenue column. You can merge the two tables to find the mismatching revenues. Use all the columns except the Revenue column for the merge and then expand the Revenue column from the second table. Next, apply a filter on the two Revenue columns to find differences.
2. To find the new rows that the hacker added, you can apply Merge Columns with an anti right join between the summarized table from [Exercise 14-1](https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1) (<https://learning.oreilly.com/library/view/collect-combine-and/9781509307982/ch14.xhtml#ch14lev1>) and the compromised table. This way, only rows that exist in the compromised table will be detected during the expand step, which comes after the Merge.

#### Choose a submission type



Drag a file here, or  
Choose a file to upload

or

Webcam Photo

Canvas Files

[◀ Previous](#)

(<https://mycourses.unh.edu/courses/115726/modules/items/2904621>)

Submit Assignment

[Next ▶](#)

(<https://mycourses.unh.edu/courses/115726/modules/items/2904624>)

[« Previous](#)

(<https://mycourses.unh.edu/courses/115726/modules/items/2904621>)

[Submit Assignment](#)

[Next »](#)

(<https://mycourses.unh.edu/courses/115726/modules/items/2904624>)