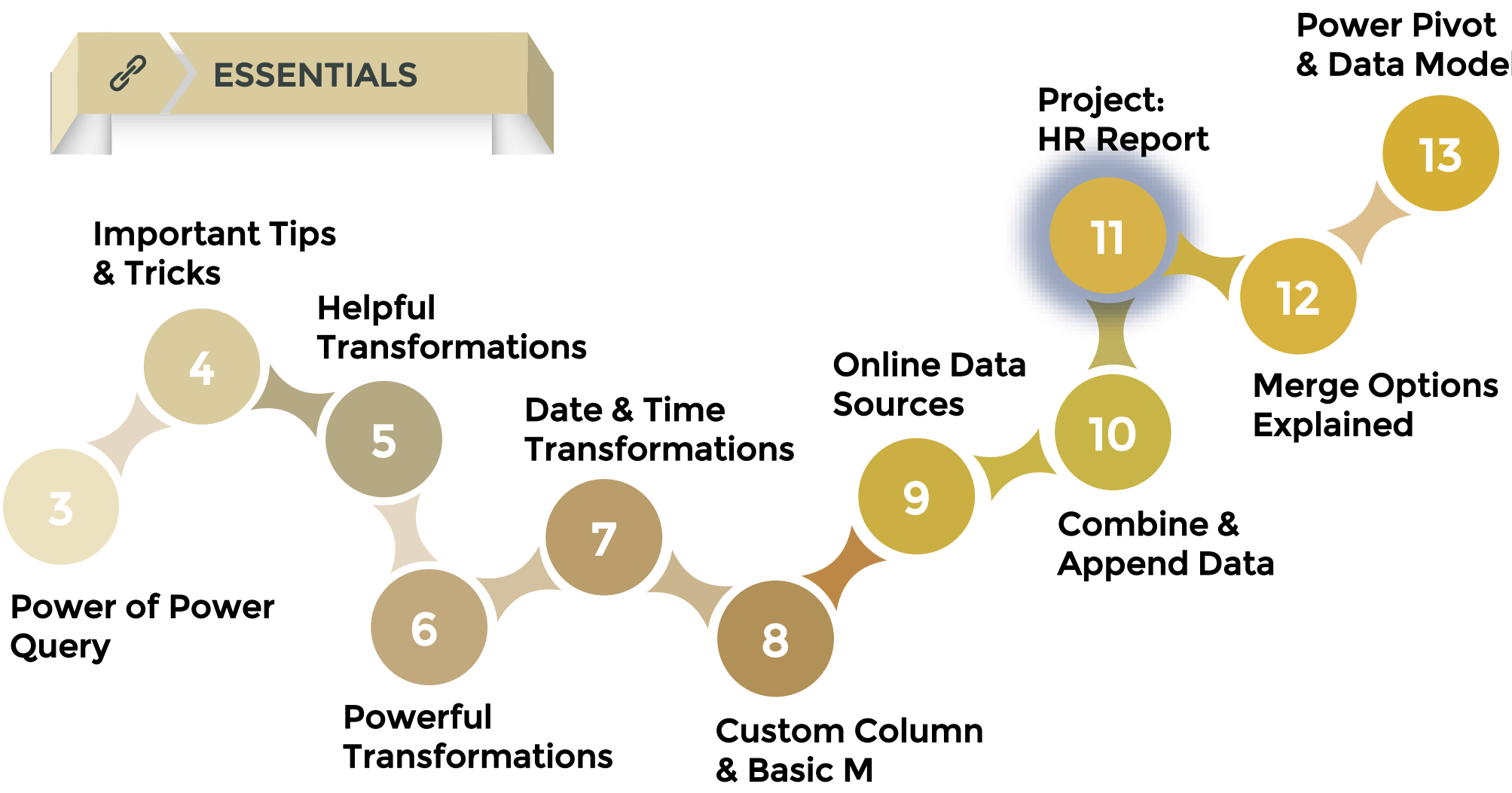
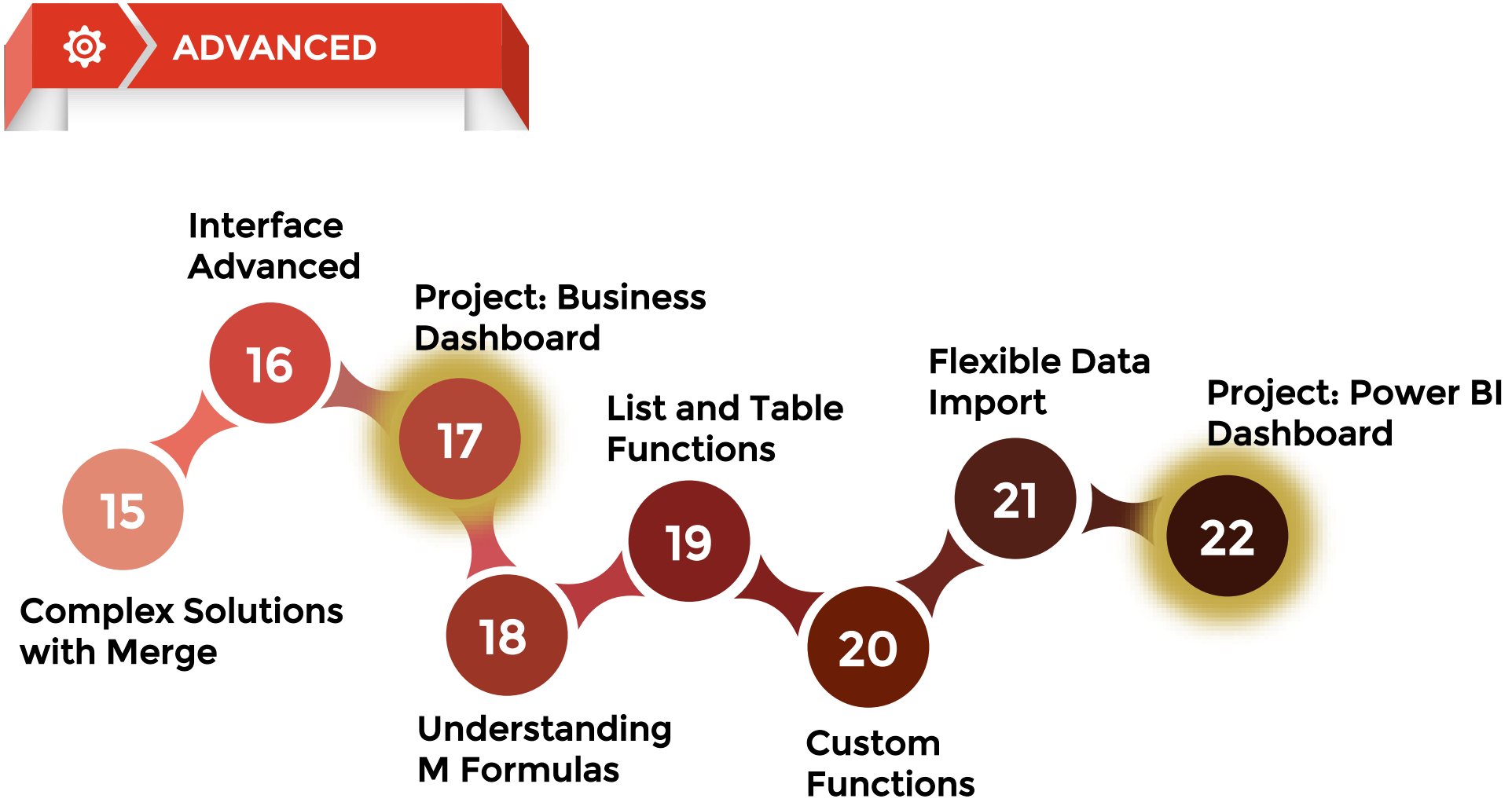
Course Roadmap




ADVANCED
Interface Advanced
16
Project: Business Dashboard
15
17
Flexible Data Import
List and Table Functions
Project: Power BI Dashboard
21
19
22
Complex Solutions with Merge
18
20
Understanding M Formulas
Custom Functions

What is a Proper Data Set?

A proper data set is a table of data
In the top row is has unique column headers (field names) followed by values (records)
There are no gaps or empty columns in between
Each column has a specific data type (e.g. numbers, text, date)
There are no subtotals in rows

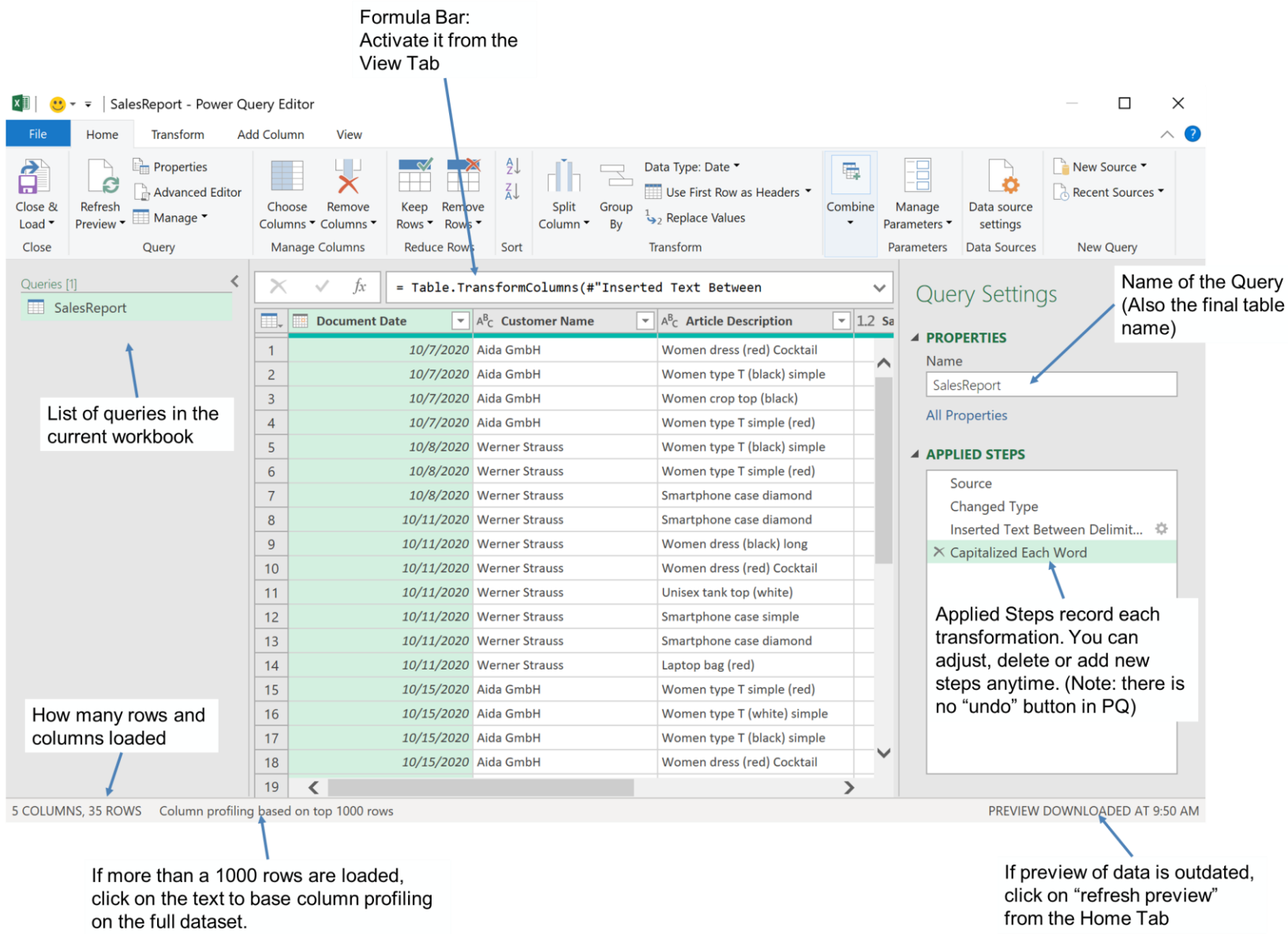
Not Proper

		Quantity	Price
2019	Product A	20	10.50
	Product B	30	15.00
Total		**50**	
2020	Product A	40	11.50
	Product B	20	16.00
Total		**60**	

Proper

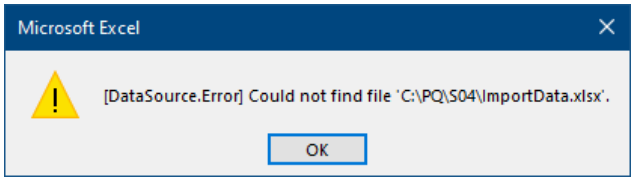
Year	Product	Quantity	Price
2019	Product A	20	10.50
2019	Product B	30	15.00
2020	Product A	40	11.50
2020	Product B	20	16.00

PQ Editor


Formula Bar: Activate it from the View Tab
SalesReport - Power Query Editor
File
Home
Transform
Add Column
View
Close & Load
Refresh Preview
Properties
Advanced Editor
Manage
Close
Query
Choose Columns
Remove Columns
Manage Columns
Keep Rows
Remove Rows
Reduce Rows
Sort
Split Column
Group By
Data Type: Date
Use First Row as Headers
Replace Values
Transform
Combine
Manage Parameters
Parameters
Data source settings
Data Sources
New Source
Recent Sources
New Query
Queries [1]
SalesReport
List of queries in the current workbook
= Table.TransformColumns(#"Inserted Text Between
Document Date
Customer Name
Article Description
1 10/7/2020 Aida GmbH Women dress (red) Cocktail
2 10/7/2020 Aida GmbH Women type T (black) simple
3 10/7/2020 Aida GmbH Women crop top (black)
4 10/7/2020 Aida GmbH Women type T simple (red)
5 10/8/2020 Werner Strauss Women type T (black) simple
6 10/8/2020 Werner Strauss Women type T simple (red)
7 10/8/2020 Werner Strauss Smartphone case diamond
8 10/11/2020 Werner Strauss Smartphone case diamond
9 10/11/2020 Werner Strauss Women dress (black) long
10 10/11/2020 Werner Strauss Women dress (red) Cocktail
11 10/11/2020 Werner Strauss Unisex tank top (white)
12 10/11/2020 Werner Strauss Smartphone case simple
13 10/11/2020 Werner Strauss Smartphone case diamond
14 10/11/2020 Werner Strauss Laptop bag (red)
15 10/15/2020 Aida GmbH Women type T simple (red)
16 10/15/2020 Aida GmbH Women type T (white) simple
17 10/15/2020 Aida GmbH Women type T (black) simple
18 10/15/2020 Aida GmbH Women dress (red) Cocktail
19
Query Settings
PROPERTIES
Name
SalesReport
All Properties
Name of the Query (Also the final table name)
APPLIED STEPS
Source
Changed Type
Inserted Text Between Delimit...
Capitalized Each Word
Applied Steps record each transformation. You can adjust, delete or add new steps anytime. (Note: there is no “undo” button in PQ)
How many rows and columns loaded
5 COLUMNS, 35 ROWS
Column profiling based on top 1000 rows
PREVIEW DOWNLOADED AT 9:50 AM
If more than a 1000 rows are loaded, click on the text to base column profiling on the full dataset.
If preview of data is outdated, click on “refresh preview” from the Home Tab

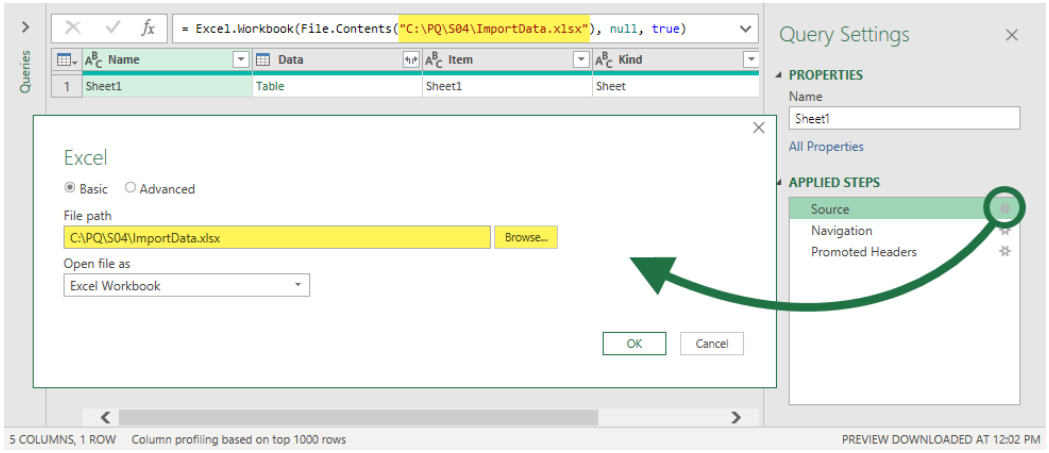
Updating the Source Location / Name

If the connection to the source data can't be resolved, you will get an error message



To fix this you need to:

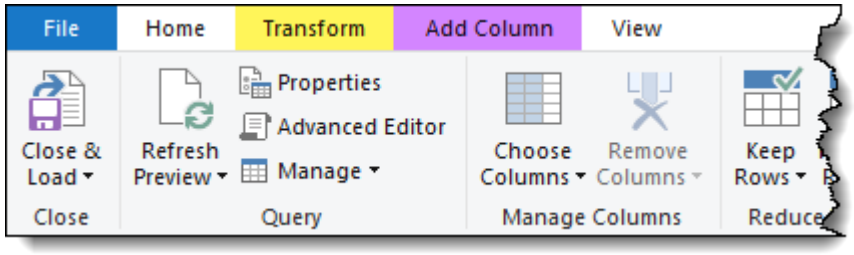
1. Open the query in Power Query
2. Update the file location/name reference by clicking the gear icon next to the Source step and edit the location/name



The advantage to the gear icon is that you are presented with a user-friendly way to browse to the new location, thus avoiding typographic errors.

Transform Versus Add Column

The tabs labeled **Transform** and **Add Column** have many of the same features. This can be confusing for beginners as to which is the correct feature version to use.



Transform will replace the original data with the transformed version of the data.

Add Column will create a new column to hold the results of the transformation.

Data Types

Data Types	Description
Decimal Number	Max 15 digits
Currency	4 digits to the right of decimal
Whole Number	Integer value (no decimals)
Percentage	Shown as decimal type when loaded to the workbook.
Date/Time	Date & Time in one column (PQ stores this as decimal number type)
Date	Dates from 1900 to 9999 are supported
Time	Time only (stored as decimals)
Date/Time/Timezone	UTC Date/Time
Duration	Length of time shown as days, hours, minutes & seconds (stored as decimals)
Text	Text, can be numbers as well
True/False	Boolean value
Binary	Sequence of bytes (e.g. when loading from a folder)
Using Locale…	Important if you're importing data from sources that have different regional settings

Correcting Date Formats from Other Regions

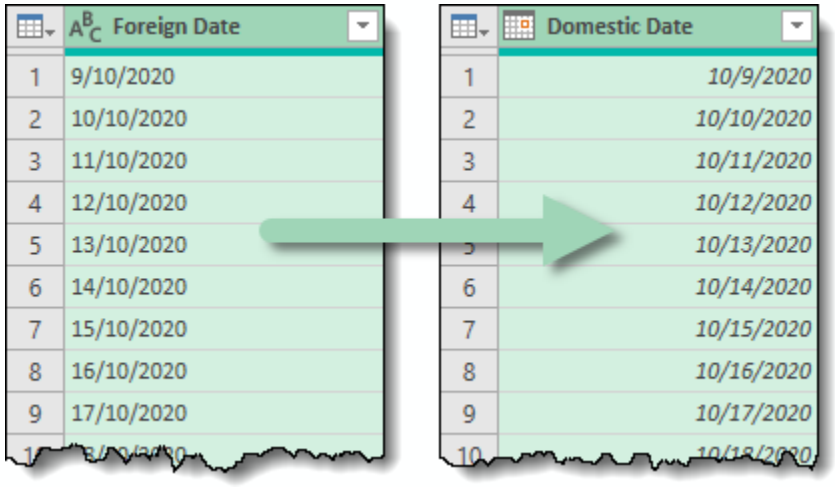
Date and number errors could be due to different regional settings of the source file and your Power Query settings.

To account for this:

1. Select the Foreign Date column.
2. Select the button next to the “Foreign Date” heading and select “Using Locale…”
3. In the Change Type with Locale dialog box, set the Data Type to “Date” and the Locale to choice that best represents the country the data originated, such as Germany (Germany), and click OK.

Alternative: Update PQ Options

Go to **Data > Get Data > Query Options**. Under Current Workbook, update the regional settings to match the one from the imported file.



Duplicate or Reference a Query?

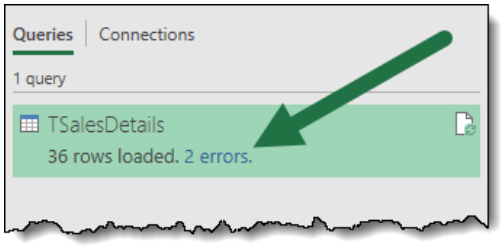
Duplicate creates a second copy of your existing query independent to the existing query.

Reference creates a new query that's dependent on the existing query.

Reference when...	Duplicate when...
You'd like the starting point of your new query to be the ending step of your existing query	You'd like to create a separate view of the existing query.

Error Handling

In case your query loads but with errors:



Check the Power Query Error Report or use a transformation to check for errors:

Home > Keep Rows > Keep Errors (Analyze and then remove the step)

To remove the errors:

Correct errors in source file

Remove error lines in Power Query Editor: **Home > Remove Rows > Remove Errors**

Replace errors with another value: **Transform > Replace > Replace Errors**

Manage Queries

Copy & Paste Queries

You can copy a query from the Queries & Connections box (right-mouse click and copy). Open a new Workbook, go to **Data > Queries & Connections**, right-mouse click on the pane and select paste.

Group Queries

Select the queries you want to group together from the **Queries & Connections** box by holding down the Control key and clicking on the query, right-mouse click and select **New Group**. Give the group a name. The remaining queries will be automatically grouped in an “Other Queries” Folder.
To ungroup the queries, you can right-mouse click and select ungroup.

Backing up Your Query Results

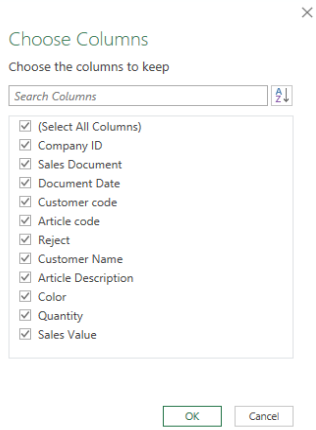
After loading the data as a table, right-mouse click and delete the query. The results table can no longer be refreshed. To remove all queries at once, save the file under a different name first, then go to **File > Info > Check for Issues > Inspect Document > Custom XML Data > Remove All**. This will delete all the queries and connections in the current file.

Power Query Shortcuts

Shortcut	Task
F2	Edit the name of a column or query
Arrow keys (L & R)	Navigate left or right through columns
CTRL key	Select multiple, non-contiguous columns
Shift key	Select contiguous columns
CTRL-A	Select ALL columns
CTRL-Space	Select the entire column of a selected cell
ALT (while opening Excel)	Open a second, unrelated instance of Excel

Common Time Saving Features

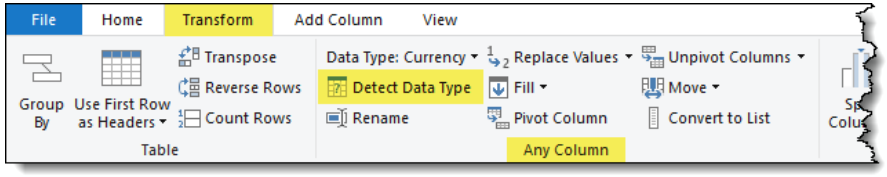
Selecting / Deselecting Columns



To select or deselect multiple columns, a selectable list of columns can be displayed by clicking **Home (tab) → Manage Columns (group) → Choose Columns**.

Detecting Data Types

It's not uncommon to delete the automatically applied type detection step from the Applied Steps list in Power Query.



When it comes time to perform data type detection, you can select the column(s) you want to data type and click **Transform (tab) → Any Column (group) → Detect Data Type**.

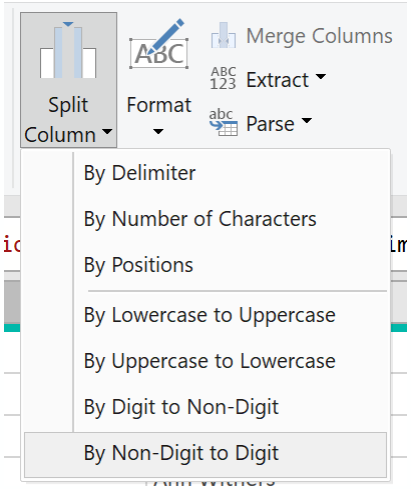
PQ Best Practice

1. Source data is in the structure you need.
2. Name the Query (avoid spaces).
3. Use column profiling on the entire data to define data types and check for errors.
4. Update step names and add descriptions to steps (good for documentation).
5. Activate the formula bar and keep your eye on hard-coded field names.
6. Test different versions of the solution by duplicating the query.
7. Take backup of results by deleting the Query connection and saving a copy of the file.
8. Ensure you have correct data types before loading to workbook.
9. In case you need to leave the query, but you'd like to save it, load it as "connection only". Later you can go back, adjust and load it to another destination.
10. Keep an open communication with the provider of your source data. You need to be informed in case something changes on the source side.

Text Transformation

In Power Query we get access to useful text transformations such as:

- Split text by delimiter
- By Number of characters,
- By Positions
- By Lowercase to Uppercase & vice versa
- By Digit to Non-Digit & vice versa



This saves us from having to write complex formulas to transform our data.

Number Transformation

You have an extensive set of number transformations in Power Query. Here's what you need to watch out for:

Null Result → If you add or multiply two columns with one another and one column includes null values, the result will be null. If you'd like to treat null values as zeros, you can either:

1. Replace null with 0 values
2. Add a third dummy column and apply your transformation. The moment you have three or more columns, Power Query will use an inbuilt function instead of the mathematical operator. For example, it will use List.Sum instead of + and List.Product instead of *. The functions will ignore null values.

Useful Time Transformations

Feature	Use
Time Only	Extract the Time component from the Date/Time values
Local Time	Return the Date/Time/Timezone values
Parse	Return the Time value parsed from the text
Hour	Extract the Hour component from the Date/Time values
Start of Hour	Return the start of the hour corresponding to each Time value
End of Hour	Return the end of the hour corresponding to each Time value
Minute	Extract the Minute component from the Date/Time values
Second	Extract the Second component from the Date/Time values
Subtract	Contains the duration between the values in the first and second selected columns
Combine Date and Time	Merge the selected columns into a new column containing both Date and Time data
Earliest	Return the earliest Time value
Latest	Return the latest Time value

Useful Date Transformations

Feature	Use
Age	Return the duration between the current local time and the values in the selected columns
Date Only	Extract the Date component from the Date/Time values
Year	Extract the Year component from the Date/Time values
Start of Year	Return the first day if the year corresponding to each Date/Time value
End of Year	Return the last day if the year corresponding to each Date/Time value
Month	Extract the Month component from the Date/Time values
Start of Month	Return the first day if the month corresponding to each Date/Time value
End of Month	Return the last day if the month corresponding to each Date/Time value
Days in Month	Return the number of days in the month corresponding to each Date/Time value
Name of Month	Return the name of the month corresponding to each Date/Time value
Quarter of Year	Return the quarter corresponding to each Date/Time value
Start of Quarter	Return the start of the quarter corresponding to each Date/Time value
End of Quarter	Return the end of the quarter corresponding to each Date/Time value
Week of Year	Return the week of the year corresponding to each Date/Time value
Week of Month	Return the week of the month corresponding to each Date/Time value
Start of Week	Return the start of the week corresponding to each Date/Time value
End of Week	Return the end of the week corresponding to each Date/Time value
Day	Extract the Day component from the Date/Time values
Day of Week	Return the day of the week corresponding to each Date/Time value
Day of Year	Return the day of the year corresponding to each Date/Time value
Start of Day	Return the start of the day corresponding to each Date/Time value
End of Day	Return the end of the day corresponding to each Date/Time value
Name of Day	Return the name of the day corresponding to each Date/Time value
Subtract Days	Return the number of days between the values in the first and second selected columns
Combine Date and Time	Merge the selected columns into a new column containing both Date and Time data
Earliest	Return the earliest Date value
Latest	Return the latest Date value

Useful Duration Features

Feature	Use
Days	Return the days component corresponding to each Duration value
Hours	Return the hours component corresponding to each Duration value
Minutes	Return the minutes component corresponding to each Duration value
Seconds	Return the seconds component corresponding to each Duration value
Total Years	Return the total number of years in each Duration value
Total Days	Return the total number of days in each Duration value
Total Hours	Return the total number of hours in each Duration value
Total Minutes	Return the total number of minutes in each Duration value
Total Seconds	Return the total number of seconds in each Duration value
Subtract	Contains the duration between the values in the first and second selected columns
Multiple	Multiplies each value in the selected columns by a specified value
Divide	Divides each value in the selected columns by a specified value
Statistics (Sum, Min, Max, Median, Average)	Sum, Min, Max, Median, or Average of all the Durations

Unpivot Columns

Create rows from columns
Bring numbers that track a specific Attribute together in one column

	January	February
Laptop Bag	62,461	81,661
Men Dress Shirt	58,652	88,354
Men Shorts	37,694	15,193
Men Type T Simple	40,502	12,542

Product	Month	Value
Laptop Bag	January	62,461
Laptop Bag	February	81,661
Men Dress Shirt	January	58,652
Men Dress Shirt	February	88,354
Men Shorts	January	37,694
Men Shorts	February	15,193
Men Type T Simple	January	40,502
Men Type T Simple	February	12,542

UnPivot Data with Multiple Headers

Steps to Unpivot Data with Multiple Column & Row Headers:

1. Identify anchor columns – these are the columns you want to keep
2. Fill down the anchor columns in case there are gaps
3. Merge anchor columns to create one column
4. Transpose the Table
5. Fill down the columns that have gaps
6. Promote first row to header
7. Select the new anchor columns and Unpivot Other Columns
8. Split the attribute column by delimiter
9. Continue with anything else that needs cleaning – for example filtering out total values

Pivot Column

Create columns from names in rows
Pivot Column allows for data aggregation

Product	Month	Value
Laptop Bag	January	62,461
Laptop Bag	February	81,661
Men Dress Shirt	January	58,652
Men Dress Shirt	February	88,354
Men Shorts	January	37,694
Men Shorts	February	15,193
Men Type T Simple	January	40,502
Men Type T Simple	February	12,542

Product	January	February
Laptop Bag	62,461	81,661
Men Dress Shirt	58,652	88,354
Men Shorts	37,694	15,193
Men Type T Simple	40,502	12,542

Pivot Flat Data into Multiple Columns

Steps to Pivot Flat Data into Multiple Columns:

1. Add an index column.
2. Think of a logic to separate each group of data the belongs to the same row:

a) If you have a consistent number of columns, you can use a transformation on the index column and use integer divide. The value will be the number of columns.

b) If you have an inconsistent number of columns, you can use a conditional column and extract the index number based on the first column (this assumes the first column is always present). Then Fill Down and remove the original index column.

3. Pivot the column that has the column headers. Select Don't Aggregate from Advanced Options.
4. Remove the column you created in step 2.

Merge Queries

Merging Queries is like VLOOKUP in Excel

The columns matched should be of the same type

InvoiceLineID	CustomerID
144	10
5180	20
5181	20
5184	124
5185	124
5186	591
5187	591
5188	935
5189	935
5190	935
5191	413
5192	413
5193	26
5194	26
5195	26
5196	26
5197	142
5198	142

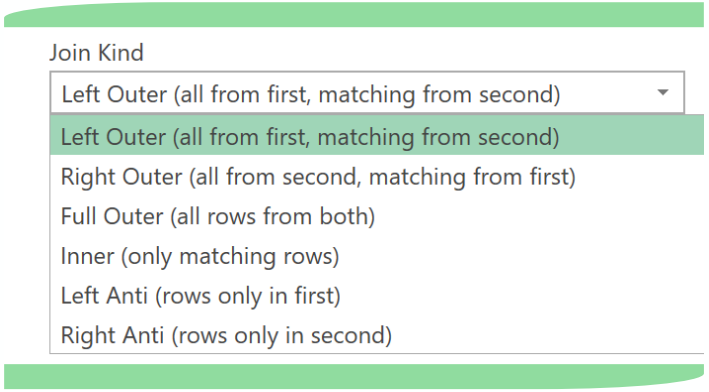
CustomerID	CustomerName
1	Tailspin Toys (Head Office)
2	Tailspin Toys (Sylvanite, MT)
3	Tailspin Toys (Peeples Valley, AZ)
4	Tailspin Toys (Medicine Lodge, KS)
5	Tailspin Toys (Gasport, NY)
6	Tailspin Toys (Jessie, ND)
7	Tailspin Toys (Frankewing, TN)
8	Tailspin Toys (Bow Mar, CO)
9	Tailspin Toys (Netcong, NJ)
10	Tailspin Toys (Wimbledon, ND)
11	Tailspin Toys (Devault, PA)
12	Tailspin Toys (Biscay, MN)
13	Tailspin Toys (Stonefort, IL)
14	Tailspin Toys (Long Meadow, MD)
15	Tailspin Toys (Batson, TX)
16	Tailspin Toys (Coney Island, MO)
17	Tailspin Toys (East Fultonham, OH)
18	Tailspin Toys (Goffstown, NH)
19	Tailspin Toys (Lemeta, AK)
20	Tailspin Toys (College Place, WA)

Overview of Merge Options

Advantage of merging in PQ:

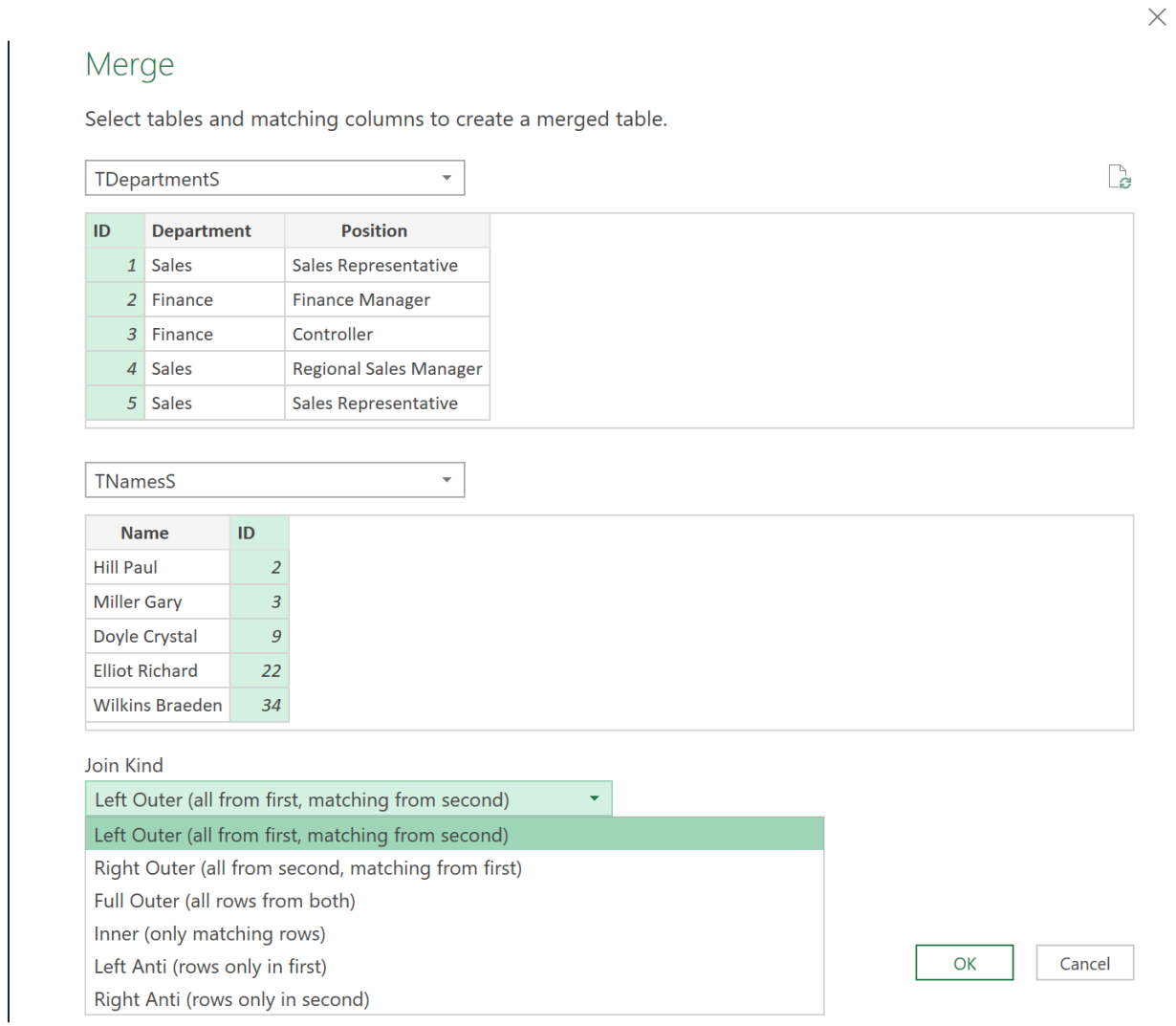
- Can merge data from external sources
- Is faster on large data sets
- No need to write formulas

Aside from the Default we get access to the below merging options:



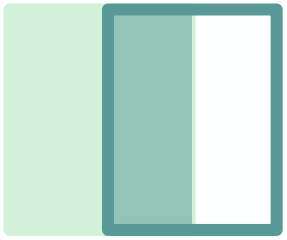
Merge Join Options

Instead of
Left & Right
we have
First & Second

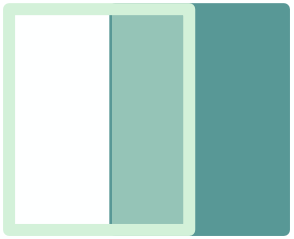


Join Kinds Explained

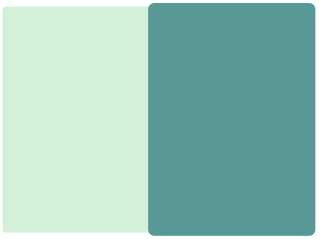
Left Outer
(All from first, matching from second)



Right Outer
(All from second, matching from first)



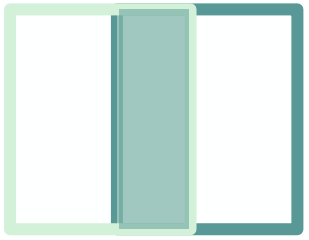
Full Outer
(All rows from both)



Left Anti
(Rows only in first)

Right Anti
(Rows only in second)

Inner
(Only matching rows)



Examples of Join Kinds

Registered

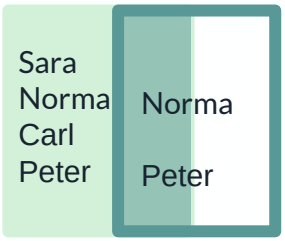
Sara
Norma
Carl
Peter

Attended

Norma
Peter
Tom

(Left Outer – all from 1st matching from 2nd)

Everyone who registered matched to attended



(Right Outer – all from 2nd matching from 1st)

Everyone who attended matched to registered

(Left Anti – rows only in 1st)

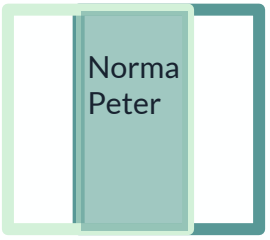
Registered but did not attend

(Right Anti – rows only in 2nd)

Attended but did not register

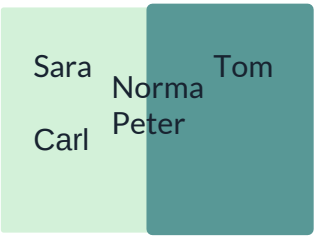
(Inner Join – only matching rows)

Registered AND attended?



(Full Outer – all rows from both)

Registered OR attended



Appending Data

When Appending files remember:

1. All column headers should be identical in both spelling and letter case.
2. The order of the columns does not matter – it's the column headers that are matched.
3. It is good practice to apply a change type step in the appended version and remove the change type step from the individual queries (unless you need the right type to apply transformations before appending).

Appending a Few Files

To append a Few Files:

1. From **Data > Get Data > From Workbook** (or any other source as needed)
2. Apply individual transformations for each file as needed
3. For context (to know which row of data is from which query) add a custom column with information as required.
4. Load each query to the workbook by only creating a connection to the files.
5. From Excel, right-mouse click on any query and select **Append**.
6. Select the first and then the second query. If you have more, select “Three or more Tables”.
7. Give the new appended query a name and load to the workbook either as table or as Pivot Table.

Appending All Sheets from a File

To append All Sheets from a Workbook:

To append all sheets from an external workbook and create a consolidated table or Pivot Table, you need to follow these steps:

1. From **Data > Get Data > From File > From Workbook**
2. Select the file and import. To import individual sheets, you can enable multi select and place a check mark beside the sheets you'd like to import. To import all sheets, select the folder icon and then **Transform**.
3. Keep the column metadata you need. Common selections are Name & Data columns.
4. When appending sheets, we no longer get the "combine" button and a sample query that's automatically created. Instead we get the "expand" button which automatically appends the content of each sheet below one another. This means we can do the cleaning up after appending. [Once we learn about Power Query functions in the advanced section, we'll also learn how to create our own custom function that transforms and prepares the data before appending.]
5. Transform the data by filtering out null values and the headers.

Remember: Double-check your steps to ensure you don't have any hard-coded values in the column headers. If yes, try and apply other steps to see if there are ways around this.

Appending All Sheets from a File

To append All Sheets from the Current Workbook:

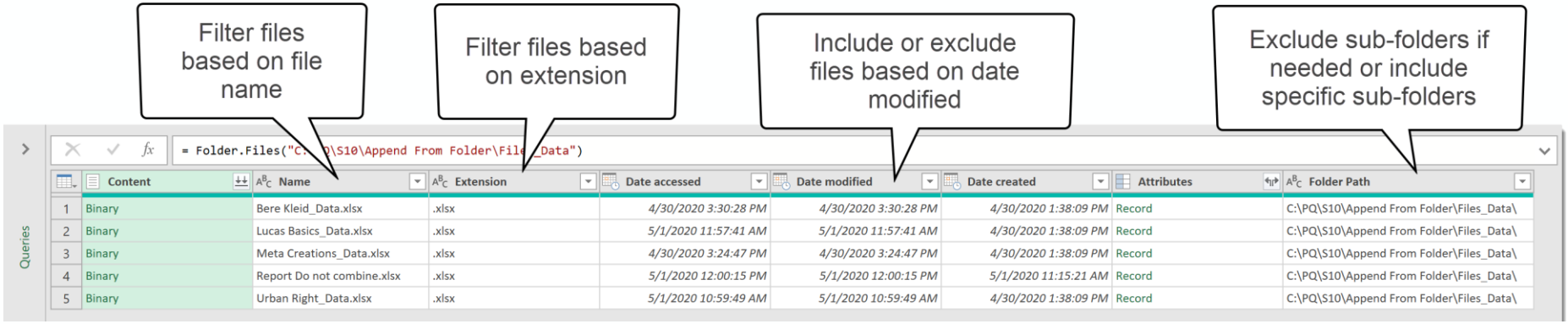
Follow these steps To append all sheets in the current workbook and create a consolidated table or Pivot table:

1. Use tables to collect the data in each sheet.
2. Instead of creating a connection to each table individually, you can set it all up at once. Go to Data > Get Data > From Other Sources > Blank Query.
3. In the formula bar type in = Excel.CurrentWorkbook()
4. You see all the objects (tables and named ranges) inside the workbook.
5. Add a filter to exclude or include only specific table names
6. Expand the content column and apply any transformations needed
7. Load as a table or Pivot Table.

Note: If your end result is a table, make sure you exclude the final query from your source list, otherwise the data will be included in the append process every time your refresh.

Appending Files from a Folder

Append All Files in Folder:
From **Data > Get Data > From File > From Folder**
Power Query creates two queries. One is a “sample query” that is applied to each file and the other is final appended result.



	Content	Name	Extension	Date accessed	Date modified	Date created	Attributes	Folder Path
1	Binary	Bere Kleid_Data.xlsx	.xlsx	4/30/2020 3:30:28 PM	4/30/2020 3:30:28 PM	4/30/2020 1:38:09 PM	Record	C:\PQ\S10\Append From Folder\Files_Data\
2	Binary	Lucas Basics_Data.xlsx	.xlsx	5/1/2020 11:57:41 AM	5/1/2020 11:57:41 AM	4/30/2020 1:38:09 PM	Record	C:\PQ\S10\Append From Folder\Files_Data\
3	Binary	Meta Creations_Data.xlsx	.xlsx	4/30/2020 3:24:47 PM	4/30/2020 3:24:47 PM	4/30/2020 1:38:09 PM	Record	C:\PQ\S10\Append From Folder\Files_Data\
4	Binary	Report Do not combine.xlsx	.xlsx	5/1/2020 12:00:15 PM	5/1/2020 12:00:15 PM	5/1/2020 11:15:21 AM	Record	C:\PQ\S10\Append From Folder\Files_Data\
5	Binary	Urban Right_Data.xlsx	.xlsx	5/1/2020 10:59:49 AM	5/1/2020 10:59:49 AM	4/30/2020 1:38:09 PM	Record	C:\PQ\S10\Append From Folder\Files_Data\

When transforming the query, think about which transformations can apply to each single file before appending and apply these directly to the sample query.

Remember - When appending from a folder, make sure:

- You only include the files you need.
- The content of sub folders will also be included unless the sub folder name is excluded by filtering the name from the folder path column.
- The final columns that will be included depend on the sample file (first file in the folder).

The Excel Data Model

When do you need the data model?

Use Excel's data model when your final report is based on multiple lookup tables.

Do you need to use Power Query before?

You don't need to use Power Query to load your data to the data model if your data is in Excel tables and if they don't need any extra transformation steps before you create the relationships. You can directly go to the relationships button in the data tab and select your tables. The moment you create a relationship between tables in your current workbook, they are automatically added to the data model.
Use Power Query if you need to transform the data before you load to the data model and if you're importing data from different sources.

How to create relationships

Relationships between tables or queries are done from **Data > Relationships** or from **Data > Manage Data Model > Diagram View**.

Power Query Merge Vs. Data Model

Can you use Power Query Merge instead of Data Model?

Yes.

If you just have two tables that need to be imported, transformed, and then merged, you can also stick with Power Query for the entire process.

The difference between Power Query Pivot Table and a data model Pivot Table is:

- With Power Query you have all the fields you need in one place.
- With the Data Model you have the fields in multiple tables.

3 Important Power Query Rules

1. Power Query is Case Sensitive

2. Power Query is Type Sensitive

3. Power Query is zero-based (Position count starts at zero)

Query's Result:
A Query's result is what you see in the last step of the query. This is what the query returns.

Step Identifiers:
If step names have spaces in the name you get the hash sign and quotation marks around the name. For example #"Changed Type". If step names don't have a space in the name, you can directly reference the name: ChangedType.

Brackets in Power Query

Brackets ( )
Square brackets [ ]
Curly brackets { }

Functions Use Brackets:

Text.Contains("Finance Manager","Manager")

Square Brackets:

Column Selection → Text.Contains([Department],"Manager")
Field Selection (Records) → [Employee Name = "West Kim"]
Projection Operator → [[Employee Name],[Position]]

Curly Brackets:

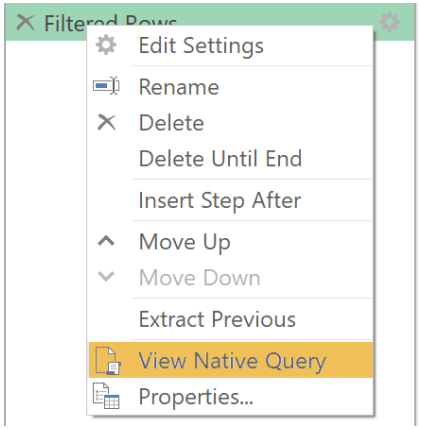
Selection / positional operator → #"Changed Type"{2}
Holds list of values (one column & can be of different types) → {"Department", "Position"}

Query Folding

Whenever possible, Power Query translates transformation steps to the native language of the data source system.

Processing work is sent back to source.

Query folding applies to relational databases, OData feed or exchange.
Query folding does not apply to flat files like csv files or Excel files.



Does Not support query folding

Pivot Columns
Adding Index columns
Complex custom functions

Power BI

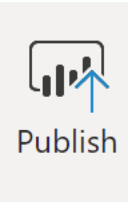
Power BI Desktop

App

Power Query: Get to connect to different data sources

Power Pivot: Get to create relationships and create data models

Interactive Reports: Create state of the art reports and visualizations



Publish: Share the reports with others

You can download Power BI Desktop for free from *PowerBi.Microsoft.com*