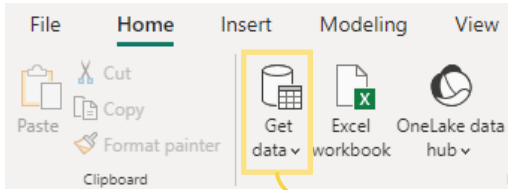
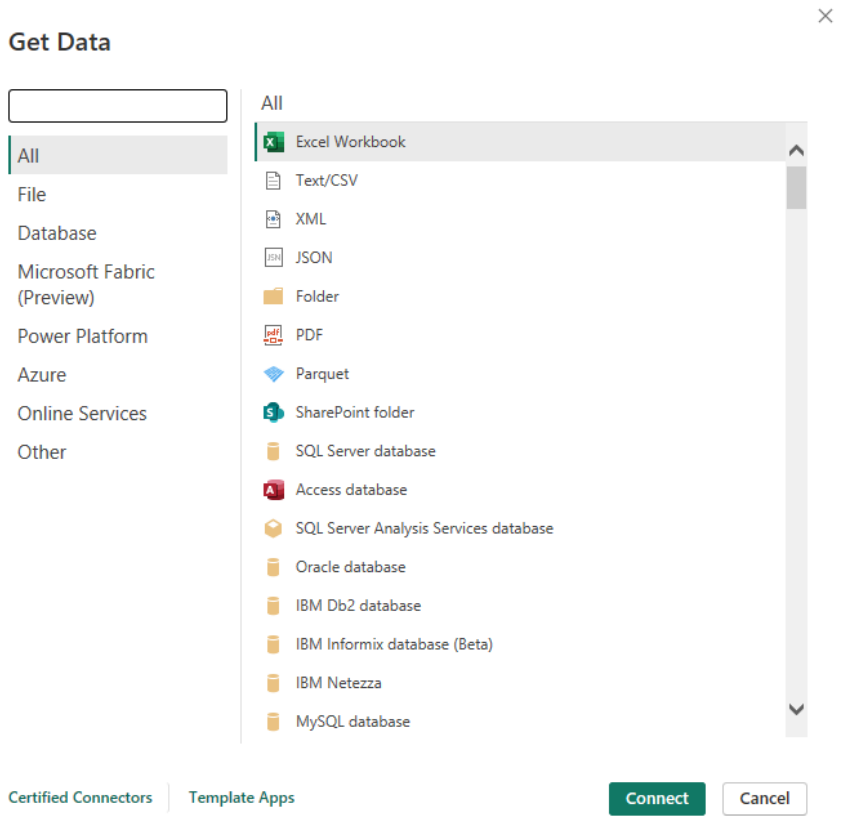
CONNECTING & SHAPING DATA

Type of Data Connectors

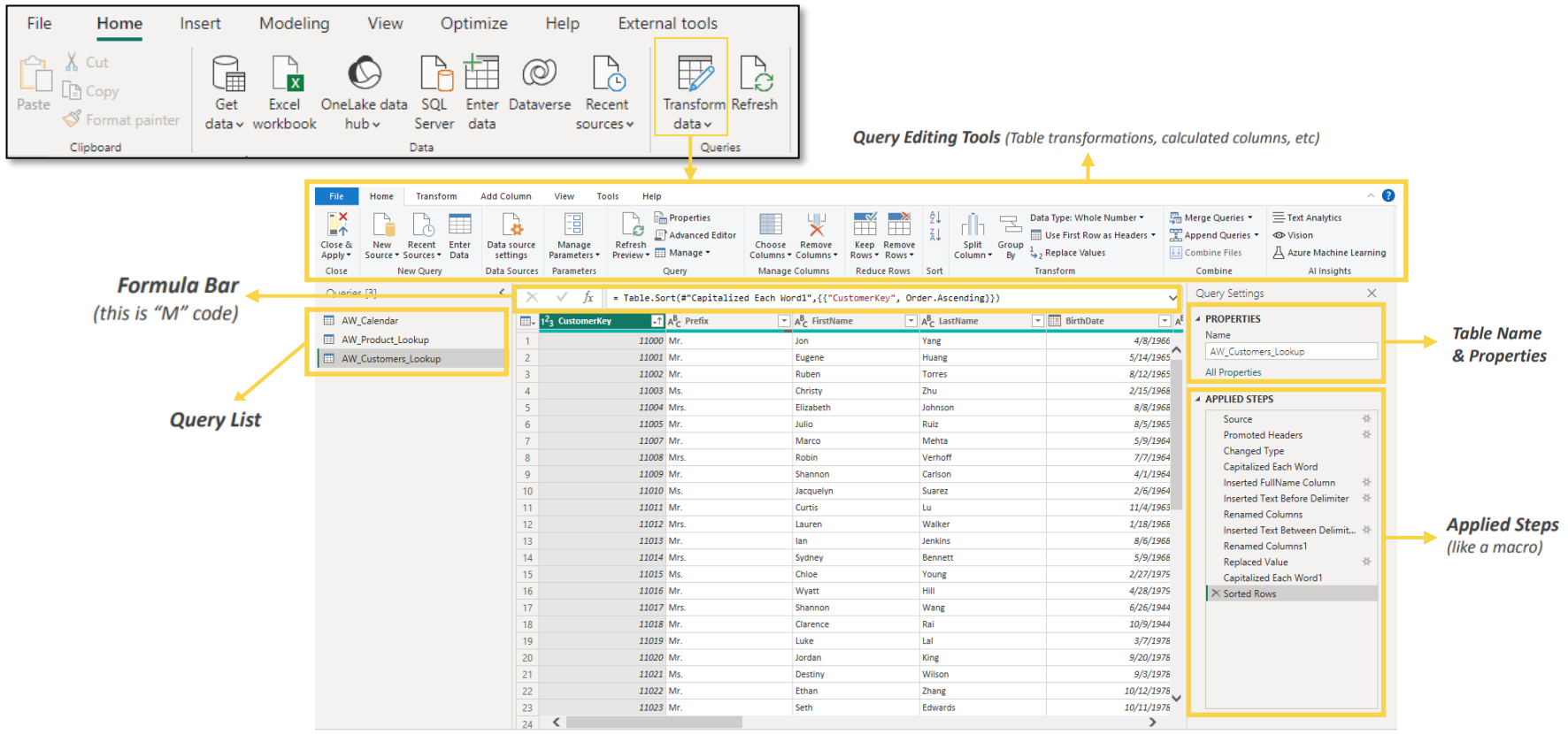




Power BI can connect to virtually **any** type of source data, including (*but not limited to*):

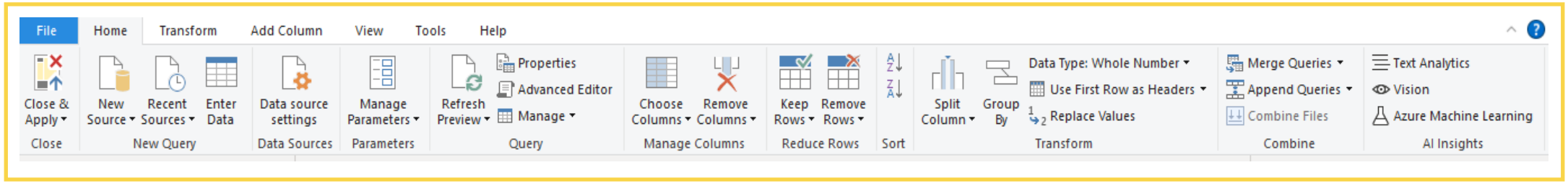
- **Flat files & Folders** (*csv, text, xls, etc*)
- **Databases** (*SQL, Access, Oracle, IBM, Azure, etc*)
- **Online Services** (*Sharepoint, GitHub, Dynamics 365, Google Analytics, Salesforce, Power BI Service, etc*)
- **Others** (*Web feeds, R scripts, Spark, Hadoop, etc*)

The Transform Data

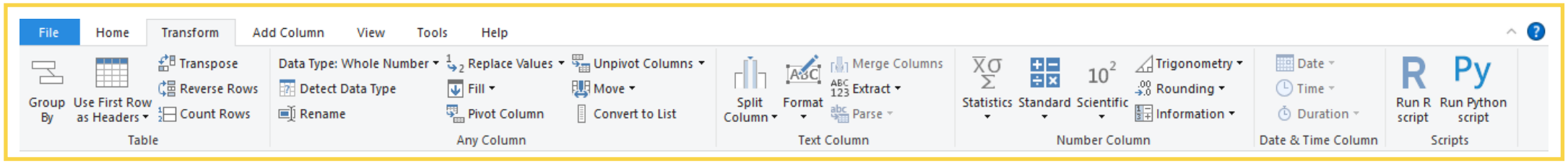


Query Editing Tools

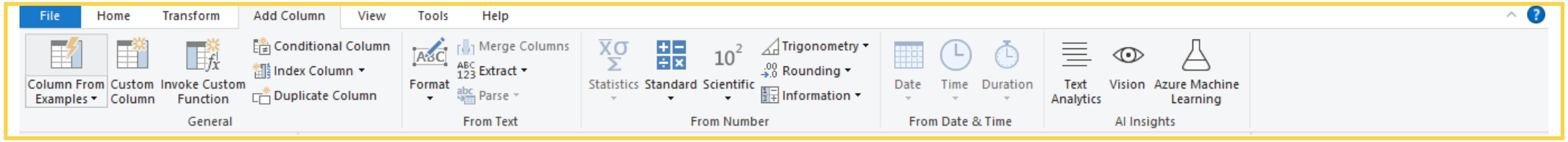
*The **HOME** tab includes **general settings** and **common table transformation tools***



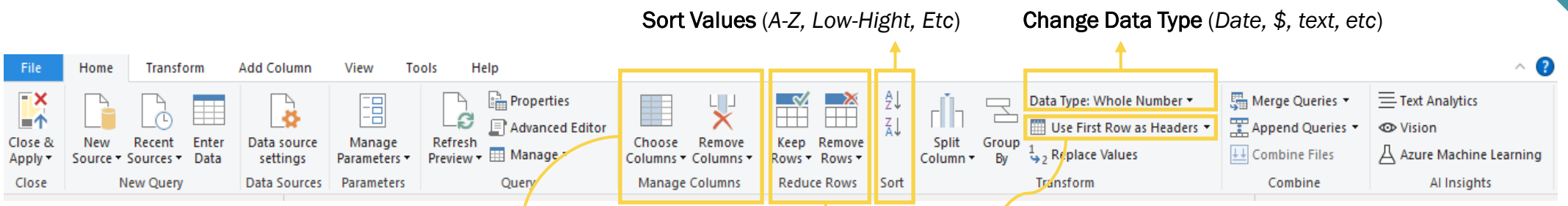
*The **TRANSFORM** tab includes tools to **modify existing columns** (splitting/grouping, transposing, extracting text, etc)*



*The **ADD COLUMN** tools **create new columns** (based on conditional rules, text operations, calculations, dates, etc)*

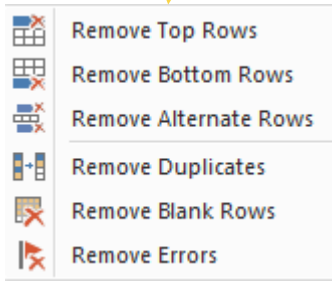


Basic Table Transformations



Chose or Remove Column

Tip: use the "Remove Other Column" option if you always want a specific set

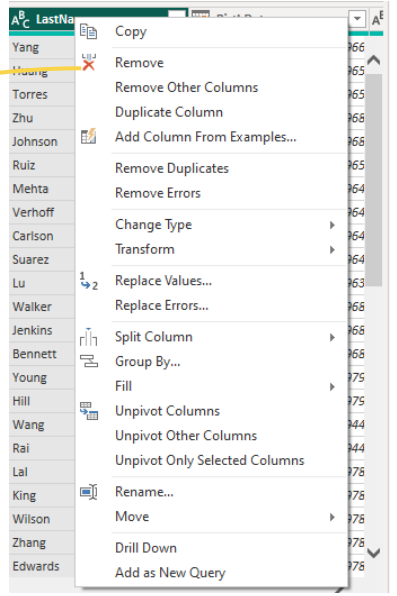


Keep or remove rows

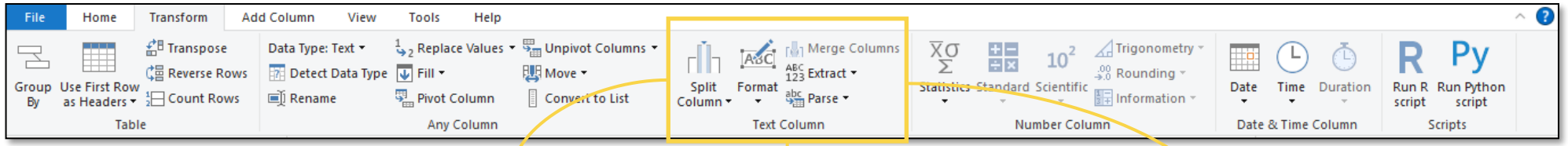
Promote header row

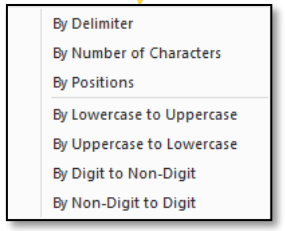
Duplicate, move & rename columns

Tip: Right-click the column header to access common tools.

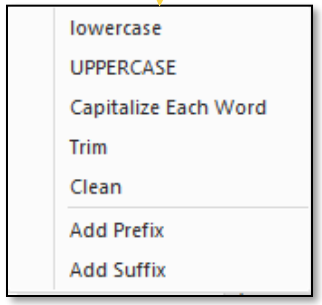


Text-specific Tools



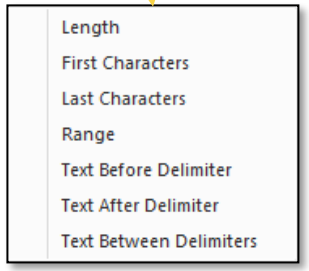


Split a text column based on either a specific delimiter or number of character



Format a text column to upper, lower or proper case, or add a prefix or suffix.

Tip: *use "Trim" to eliminate leading & trailing spaces, or clean to remove non-printable character*



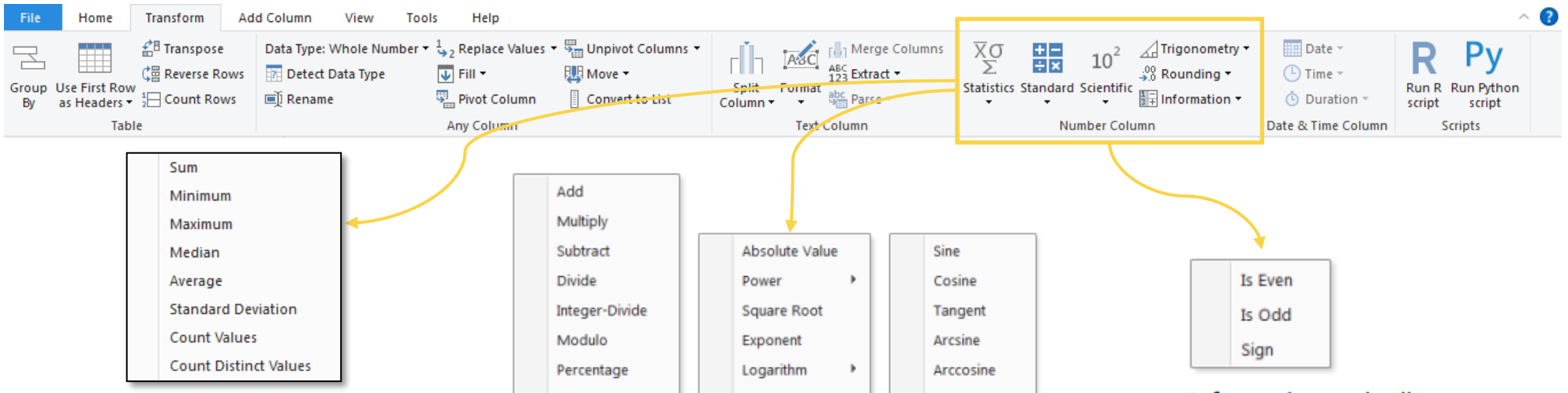
Extract character from a text column
Based on fixed lengths, first/last, ranges or delimiters

Tip: *select two or more columns to* ***merge*** *(or* ***concatenate****) fields.*

HEY THIS IS IMPORTANT!

You can access many of these tools in both the "**Transform**" and "**Add Column**" menus -- the difference is whether you want to ***add a new column*** or ***modify an existing one***

Number-specific Tools



Statistics functions allow you to evaluate basic stats for the selected column (sum, min/max, average, count, countdistinct, etc)

Note: *These tools return a SINGLE value, and are commonly used to explore a table rather than prepare it for loading*

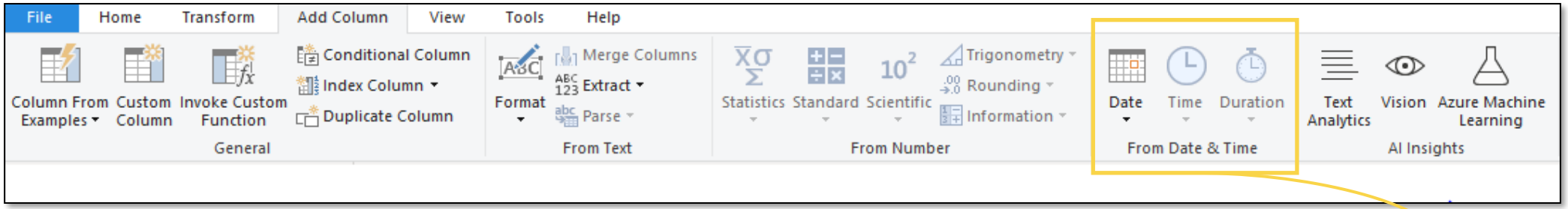
Standard *Scientific* *Trigonometry*

Standard, Scientific and **Trigonometry** tools allow you to apply standard operations (addition, multiplication, division, etc.) or more advanced calculations (power, logarithm, sine, tangent, etc) to each value in a column

Note: *Unlike the Statistics options, these tools are applied to each individual row in the table*

Information tools allow you to define binary flags (*TRUE/FALSE or 1/0*) to mark each row in a column as even, odd, positive or negative

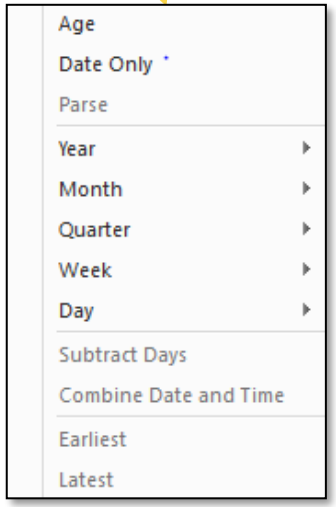
Date-specific Tools



Date & Time tools are relatively straight-forward, and include the following options:

- **Age**: Difference between the current time and the date in each row
- **Date Only:** Removes the time component of a date/time field
- **Year/Month/Quarter/Week/Day**: Extracts individual components from a date field (Time-specific options include Hour, Minute, Second, etc.)
- **Earliest/Latest:** Evaluates the earliest or latest date from a column as a single value (can only be accessed from the "Transform" menu)

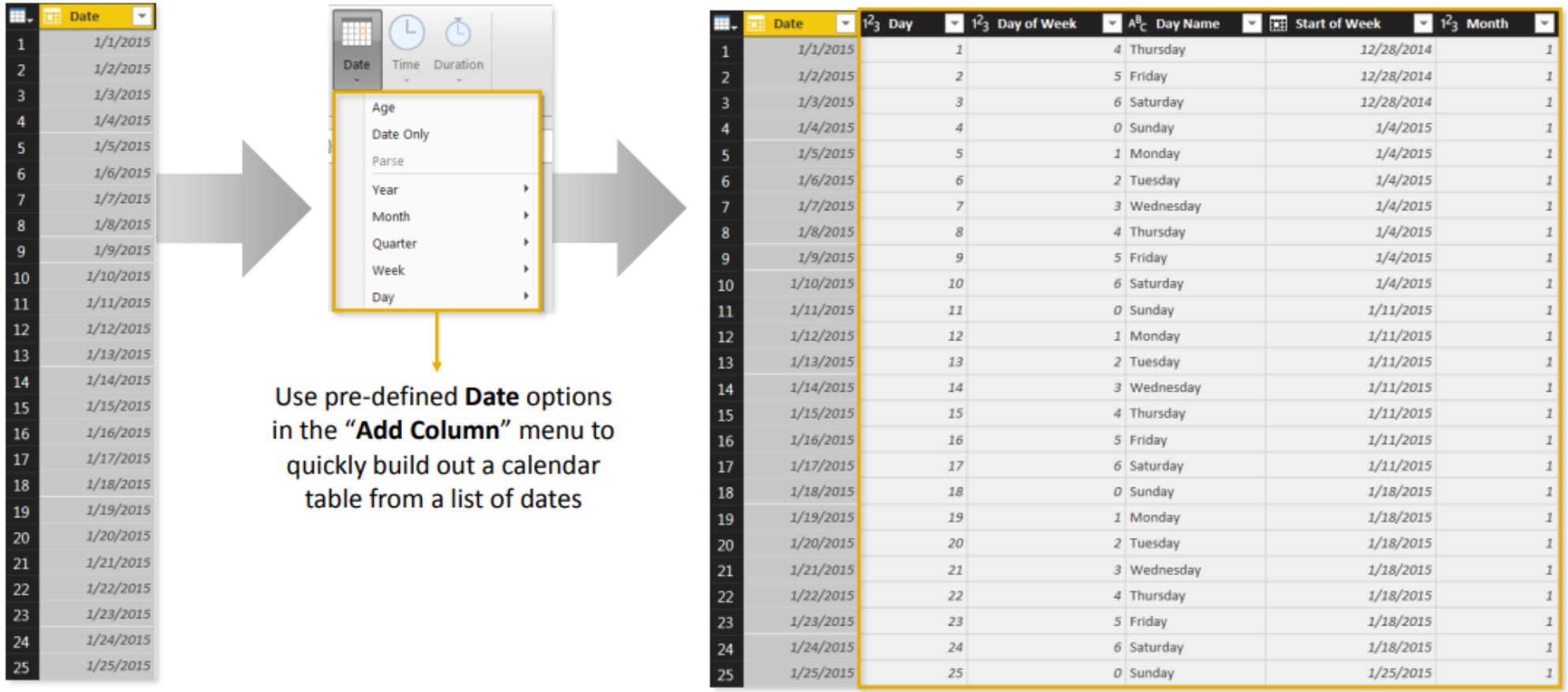
Note: *You will almost always want to perform these operations from the "Add Column" menu to build out new fields, rather than transforming an individual date/time column*



PRO TIP:

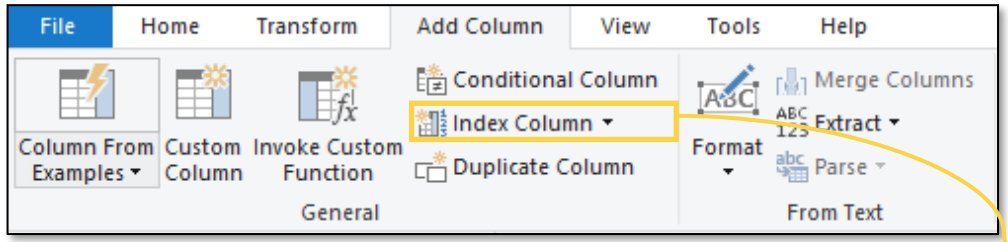
*Load up a table containing a **single date column** and use Date tools to build out an **entire calendar table***

Creating a basic Callender Table



	Date	Day	Day of Week	Day Name	Start of Week	Month
1	1/1/2015	1	4	Thursday	12/28/2014	1
2	1/2/2015	2	5	Friday	12/28/2014	1
3	1/3/2015	3	6	Saturday	12/28/2014	1
4	1/4/2015	4	0	Sunday	1/4/2015	1
5	1/5/2015	5	1	Monday	1/4/2015	1
6	1/6/2015	6	2	Tuesday	1/4/2015	1
7	1/7/2015	7	3	Wednesday	1/4/2015	1
8	1/8/2015	8	4	Thursday	1/4/2015	1
9	1/9/2015	9	5	Friday	1/4/2015	1
10	1/10/2015	10	6	Saturday	1/4/2015	1
11	1/11/2015	11	0	Sunday	1/11/2015	1
12	1/12/2015	12	1	Monday	1/11/2015	1
13	1/13/2015	13	2	Tuesday	1/11/2015	1
14	1/14/2015	14	3	Wednesday	1/11/2015	1
15	1/15/2015	15	4	Thursday	1/11/2015	1
16	1/16/2015	16	5	Friday	1/11/2015	1
17	1/17/2015	17	6	Saturday	1/11/2015	1
18	1/18/2015	18	0	Sunday	1/18/2015	1
19	1/19/2015	19	1	Monday	1/18/2015	1
20	1/20/2015	20	2	Tuesday	1/18/2015	1
21	1/21/2015	21	3	Wednesday	1/18/2015	1
22	1/22/2015	22	4	Thursday	1/18/2015	1
23	1/23/2015	23	5	Friday	1/18/2015	1
24	1/24/2015	24	6	Saturday	1/18/2015	1
25	1/25/2015	25	0	Sunday	1/25/2015	1

Adding Index Column

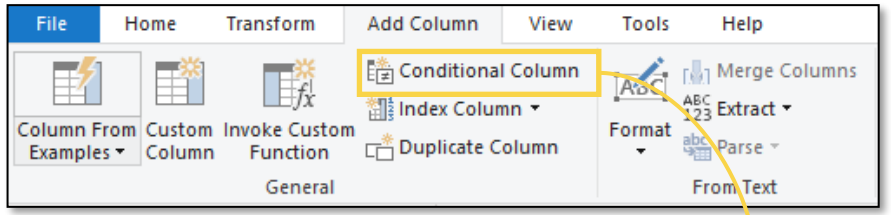


Index Columns contain a list of sequential values that can be used to identify each unique row in a table (*typically starting from 0 or 1*)

These columns are often used to create **unique IDs** that can be used to form relationships between tables (*more on that later!*)

	1²3 Index	Date
1	1	1/1/2015
2	2	1/2/2015
3	3	1/3/2015
4	4	1/4/2015
5	5	1/5/2015
6	6	1/6/2015
7	7	1/7/2015
8	8	1/8/2015
9	9	1/9/2015
10	10	1/10/2015
11	11	1/11/2015
12	12	1/12/2015
13	13	1/13/2015
14	14	1/14/2015
15	15	1/15/2015
16	16	1/16/2015
17	17	1/17/2015
18	18	1/18/2015
19	19	1/19/2015
20	20	1/20/2015
21	21	1/21/2015
22	22	1/22/2015
23	23	1/23/2015
24	24	1/24/2015

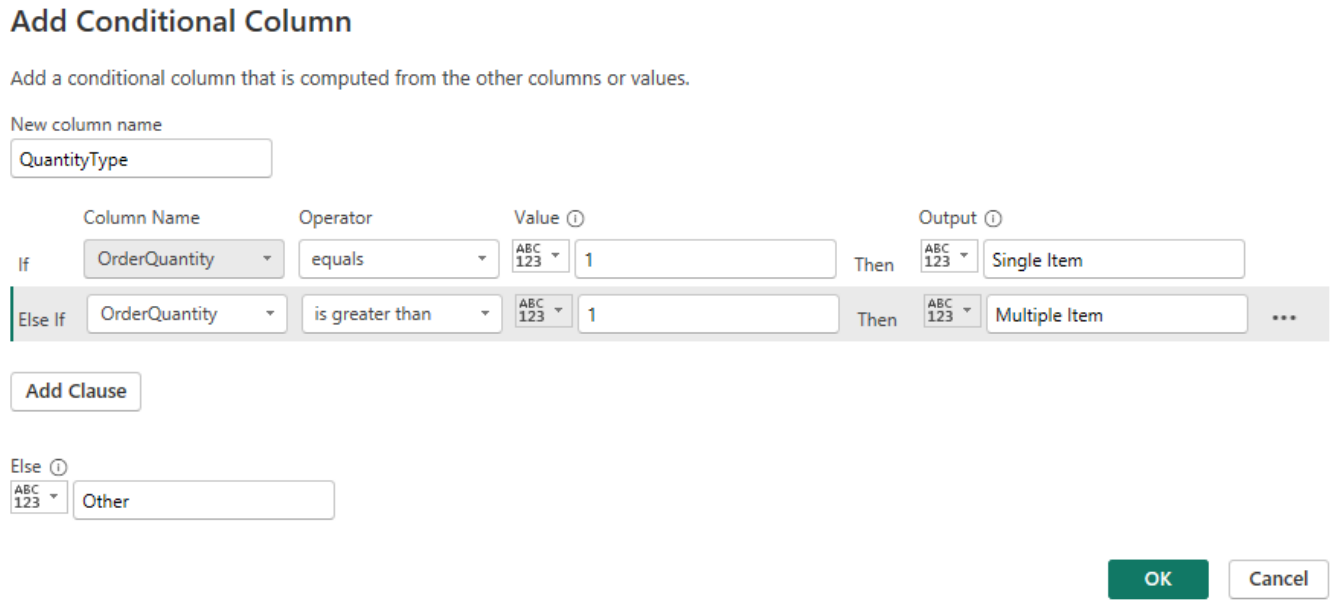
Adding Conditional Column



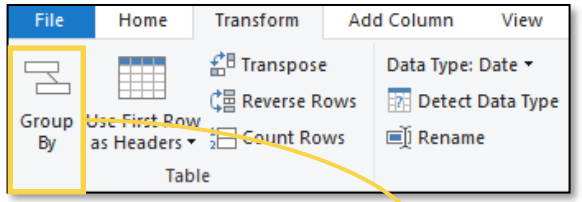
Conditional Column allow you to define new fields based on logical rules and conditions (IF/THEN Statement)

*In this case we're creating a new conditional column called "**QuantityType**", which depends on the values in the "**OrderQuantity**" column, as follows:*

- *If OrderQuantity **=1**, QuantityType = "**Single Item**"*
- *If OrderQuantity **>1**, QuantityType = "**Multiple Items**"*
- *Otherwise QuantityType = "**Other**"*

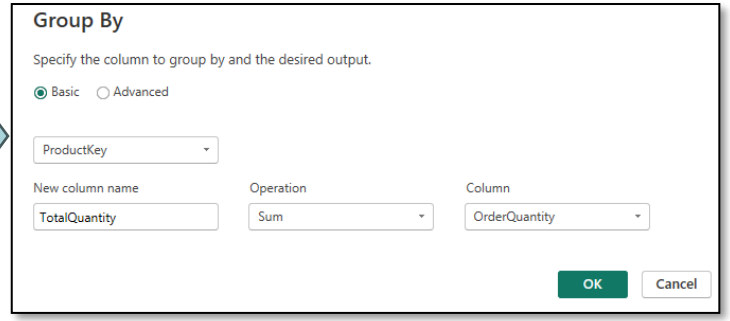


Grouping & Aggregating Data



Group By allows you to aggregate your data at a different level
(example: transform daily data into monthly, roll up transaction-level data by store)

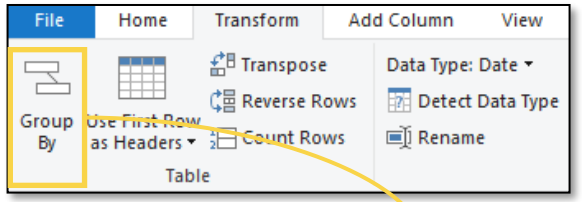
	OrderDate	StockDate	OrderNumber	ProductKey	CustomerKey
1	1/1/2017	12/13/2003	SO61285	529	23791
2	1/1/2017	9/24/2003	SO61285	214	23791
3	1/1/2017	9/4/2003	SO61285	540	23791
4	1/1/2017	9/28/2003	SO61301	529	16747
5	1/1/2017	10/21/2003	SO61301	377	16747
6	1/1/2017	10/23/2003	SO61301	540	16747
7	1/1/2017	9/4/2003	SO61269	215	11792
8	1/1/2017	10/21/2003	SO61269	229	11792
9	1/1/2017	10/24/2003	SO61286	528	11530
10	1/1/2017	9/27/2003	SO61286	536	11530
11	1/1/2017	10/23/2003	SO61298	530	18155
12	1/1/2017	12/2/2003	SO61298	214	18155
13	1/1/2017	12/15/2003	SO61298	223	18155
14	1/1/2017	10/1/2003	SO61310	538	13541
15	1/1/2017	11/8/2003	SO61310	584	13541
16	1/1/2017	12/6/2003	SO61270	485	18255
17	1/1/2017	10/31/2003	SO61289	528	23694
18	1/1/2017	11/16/2003	SO61289	536	23694
19	1/1/2017	11/20/2003	SO61289	215	23694
20	1/1/2017	10/12/2003	SO61278	477	22308
21	1/1/2017	9/28/2003	SO61278	479	22308
22	1/1/2017	12/10/2003	SO61278	488	22308
23	1/1/2017	9/25/2003	SO61278	580	22308
24					



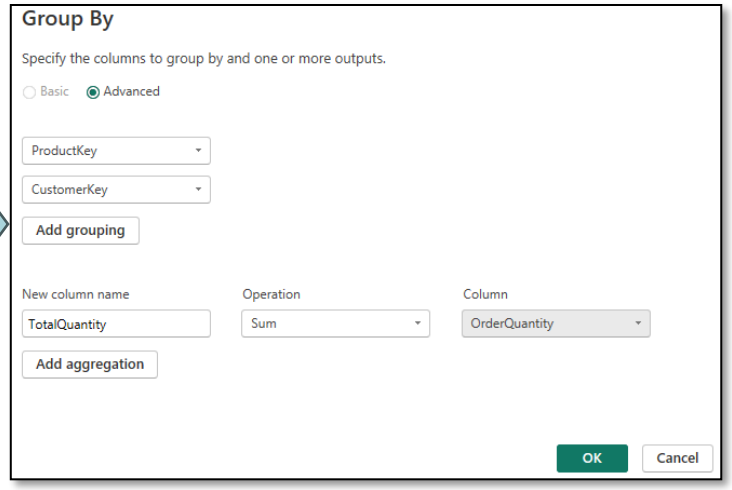
In this case we are transforming a daily, transaction-level table into a summary of **"TotalQuantity"** rolled up by **"ProductKey"**

	ProductKey	TotalQuantity
1	529	2379
2	214	1218
3	540	406
4	377	57
5	215	1074
6	229	243
7	528	3121
8	536	1132
9	530	1572
10	223	2394
11	538	1069
12	584	188
13	485	2222
14	477	4565
15	479	1943
16	488	216
17	580	154
18	583	150

Grouping & Aggregating Data (Advanced)



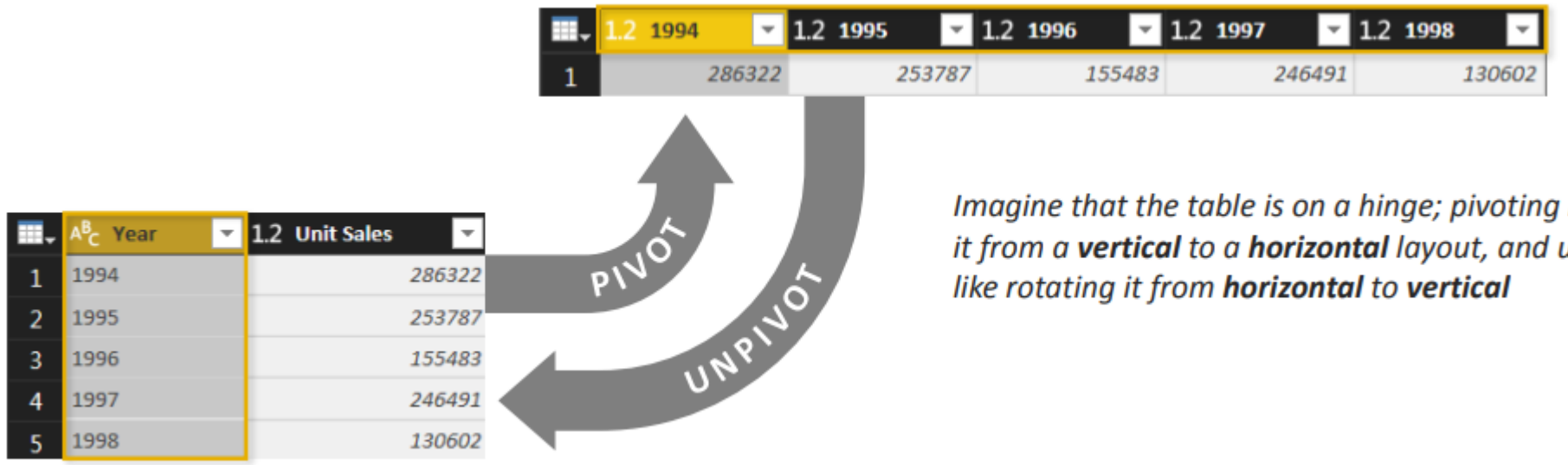
	OrderDate	StockDate	OrderNumber	ProductKey	CustomerKey	
1	1/1/2017	12/13/2003	SO61285	529	23791	
2	1/1/2017	9/24/2003	SO61285	214	23791	
3	1/1/2017	9/4/2003	SO61285	540	23791	
4	1/1/2017	9/28/2003	SO61301	529	16747	
5	1/1/2017	10/21/2003	SO61301	377	16747	
6	1/1/2017	10/23/2003	SO61301	540	16747	
7	1/1/2017	9/4/2003	SO61269	215	11792	
8	1/1/2017	10/21/2003	SO61269	229	11792	
9	1/1/2017	10/24/2003	SO61286	528	11530	
10	1/1/2017	9/27/2003	SO61286	536	11530	
11	1/1/2017	10/23/2003	SO61298	530	18155	
12	1/1/2017	12/2/2003	SO61298	214	18155	
13	1/1/2017	12/15/2003	SO61298	223	18155	
14	1/1/2017	10/1/2003	SO61310	538	13541	
15	1/1/2017	11/8/2003	SO61310	584	13541	
16	1/1/2017	12/6/2003	SO61270	485	18255	
17	1/1/2017	10/31/2003	SO61289	528	23694	
18	1/1/2017	11/16/2003	SO61289	536	23694	
19	1/1/2017	11/20/2003	SO61289	215	23694	
20	1/1/2017	10/12/2003	SO61278	477	22308	
21	1/1/2017	9/28/2003	SO61278	479	22308	
22	1/1/2017	12/10/2003	SO61278	488	22308	
23	1/1/2017	9/25/2003	SO61278	580	22308	
24						



	ProductKey	CustomerKey	TotalQuantity
1	529	23791	2
2	214	23791	1
3	540	23791	1
4	529	16747	2
5	377	16747	1
6	540	16747	1
7	215	11792	1
8	229	11792	1
9	528	11530	2
10	536	11530	2
11	530	18155	2
12	214	18155	1
13	223	18155	1
14	538	13541	2
15	584	13541	1
16	485	18259	2
17	528	23694	2
18	536	23694	2
19	215	23694	1
20	477	22308	2
21	479	22308	1
22	488	22308	1
23	580	22308	1
24	583	20236	1

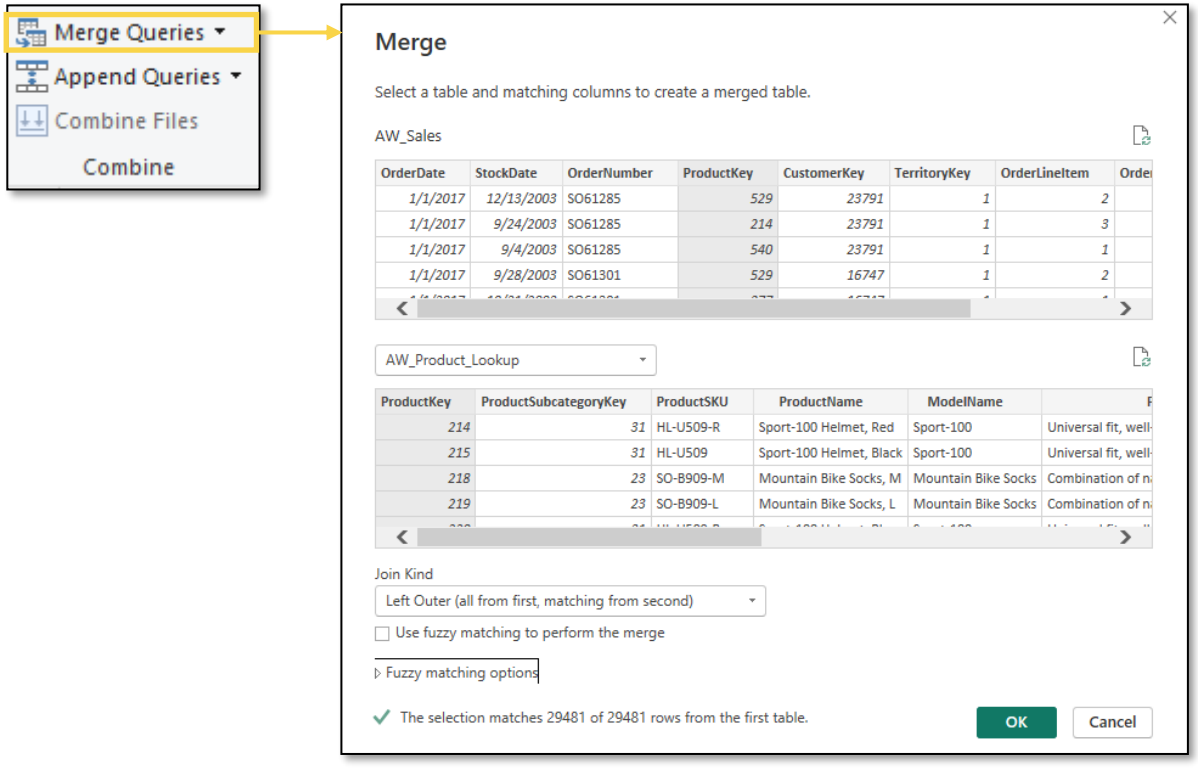
*This time we're transforming the daily, transaction-level table into a summary of "**TotalQuantity**" aggregated by both "**ProductKey**" and "**CustomerKey**" (using the advanced option in the dialog box)*

Pivoting & Unpivoting



*Imagine that the table is on a hinge; pivoting is like rotating it from a **vertical** to a **horizontal** layout, and unpivoting is like rotating it from **horizontal** to **vertical***

Merging Queries

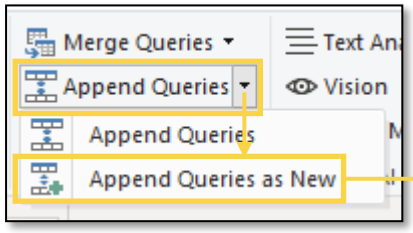


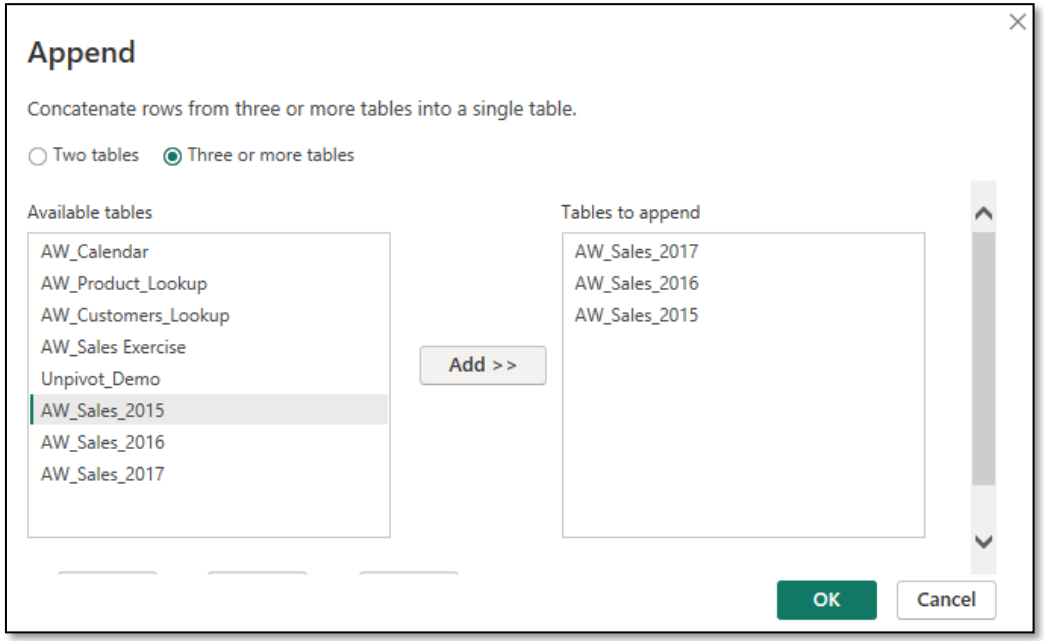
Merging queries allows you to **join tables** based on a common column (like VLOOKUP)

In this case we're merging the **AW_Sales_Data** table with the **AW_Product_Lookup** table, which share a common "*ProductKey*" column

NOTE: Merging ***adds columns*** to an existing table

Appending Queries



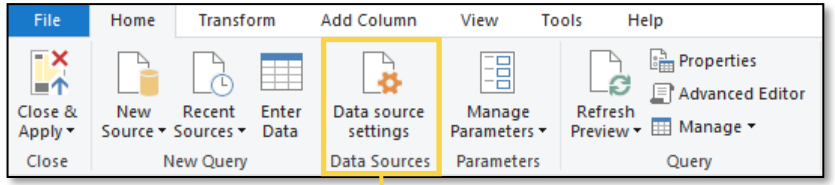


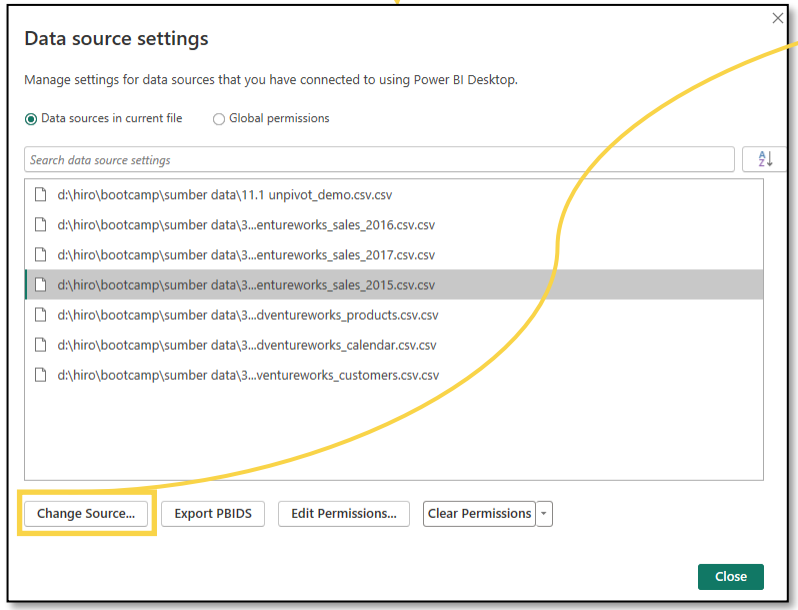
Appending queries allows you to **combine** (or ***stack***) tables that share the exact same column structure and data types

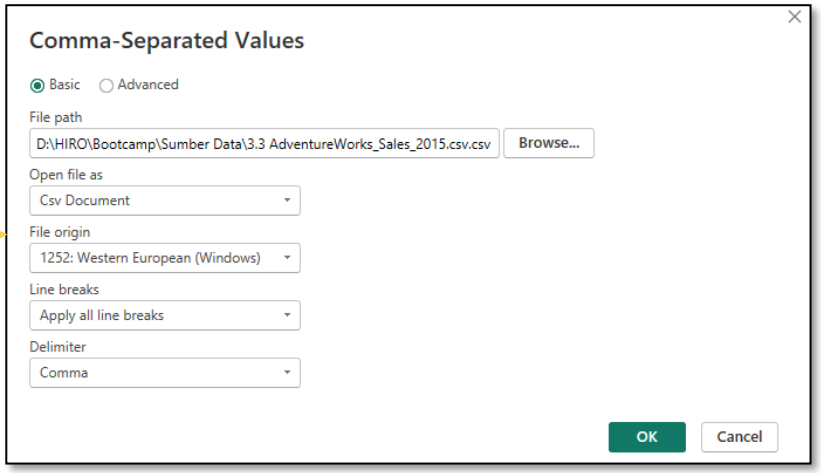
In this case we're appending the **AdventureWorks_Sales_2015** table to the **AdventureWorks_Sales_2016** table, which is valid since they share identical table structures

NOTE: Appending ***adds rows*** to an existing table

Data Source Settings







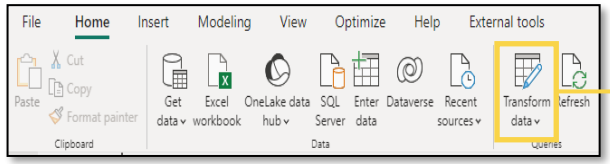
The **Data Source Settings** in the Query Editor allow you to manage data connections and permissions

HEY THIS IS IMPORTANT!

Connections to local files reference the *exact* path

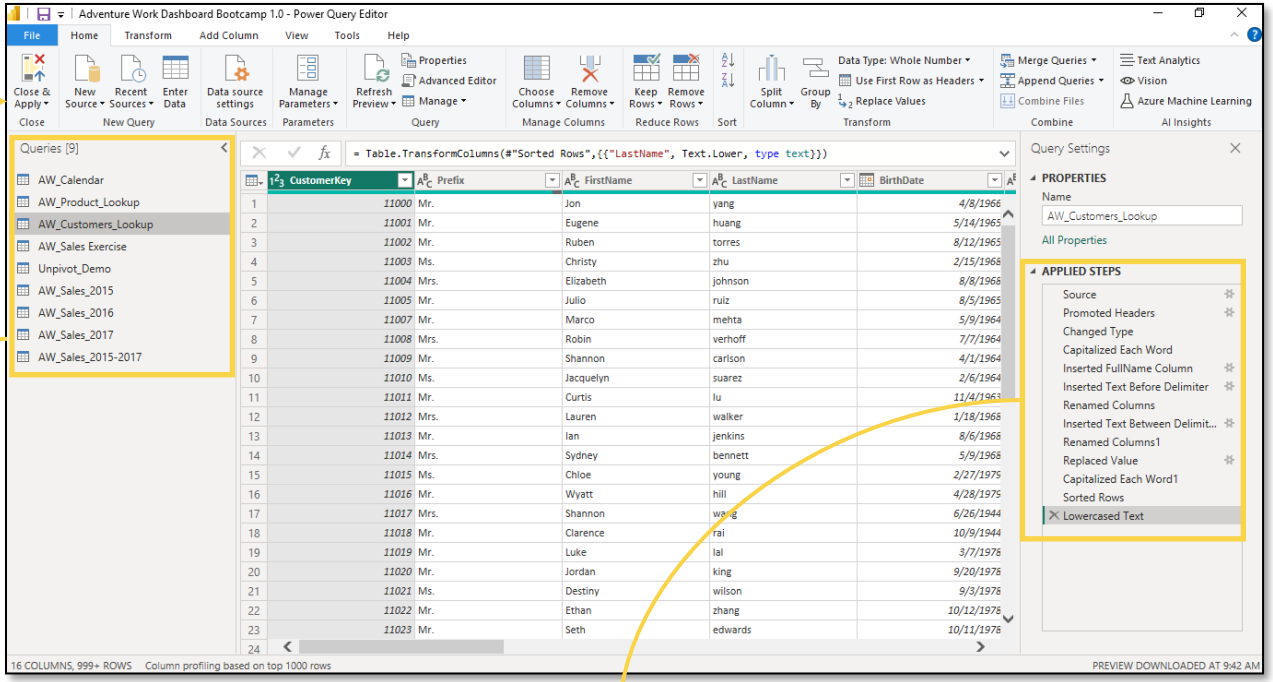
If the file name or location changes, **you will need to change the source and browse to the current version**

Modifying Queries



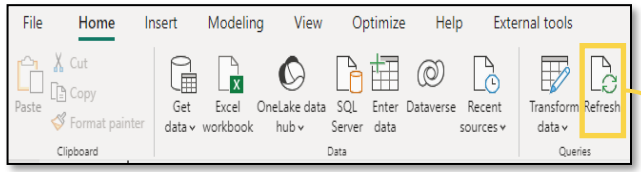
Select **Transform Data** from the Home tab to launch the query editor

Within the editor, view or modify existing queries in the “Queries” pane



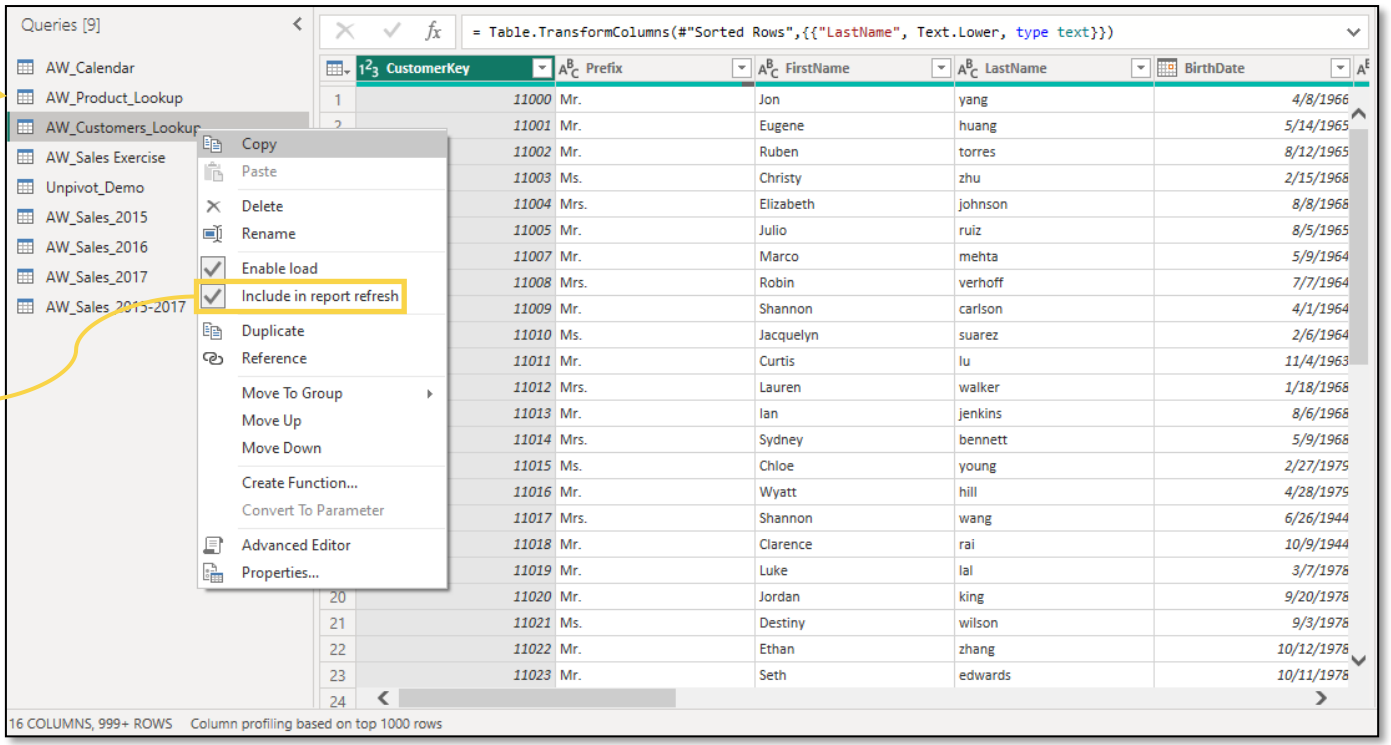
Within each query, you can click each item within the **“Applied Steps”** pane to view each stage of the transformation, add new steps or delete existing ones, or modify individual steps by clicking the gear icons

Refreshing Queries



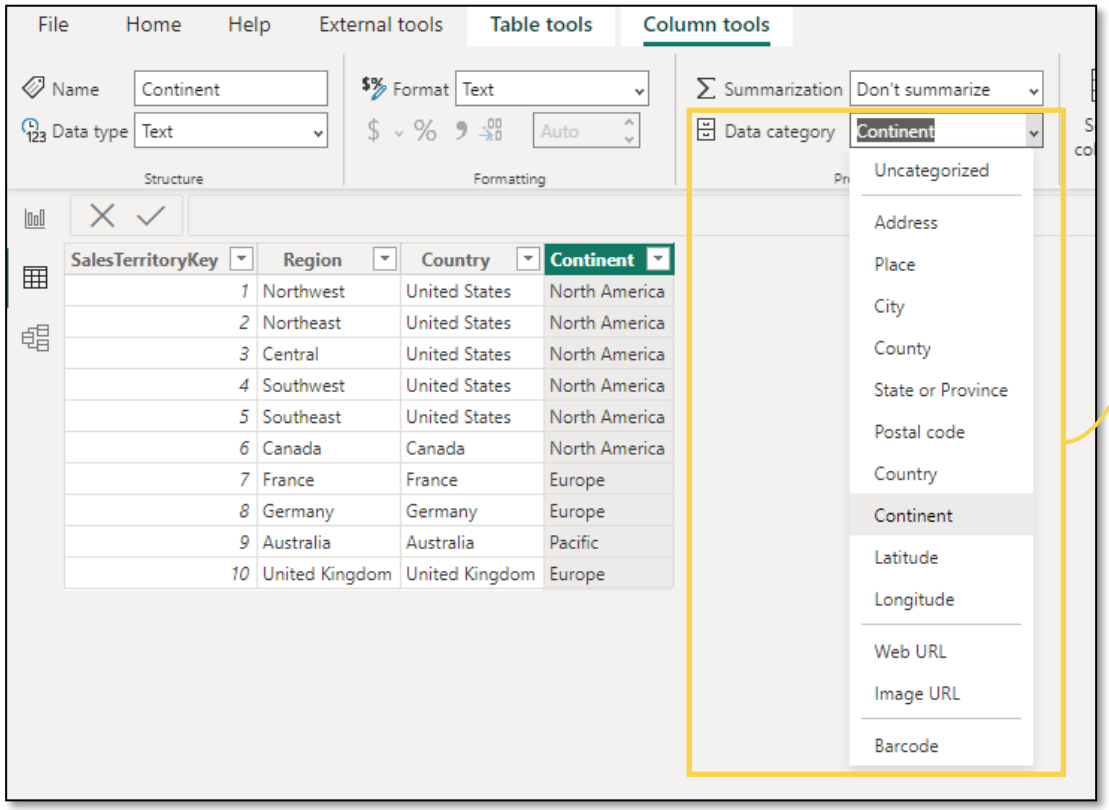
By default, **ALL** queries in the model will refresh when you use the "Refresh" command from the **Home** tab

From the Query Editor, uncheck **"Include in report refresh"** to exclude individual queries from the refresh



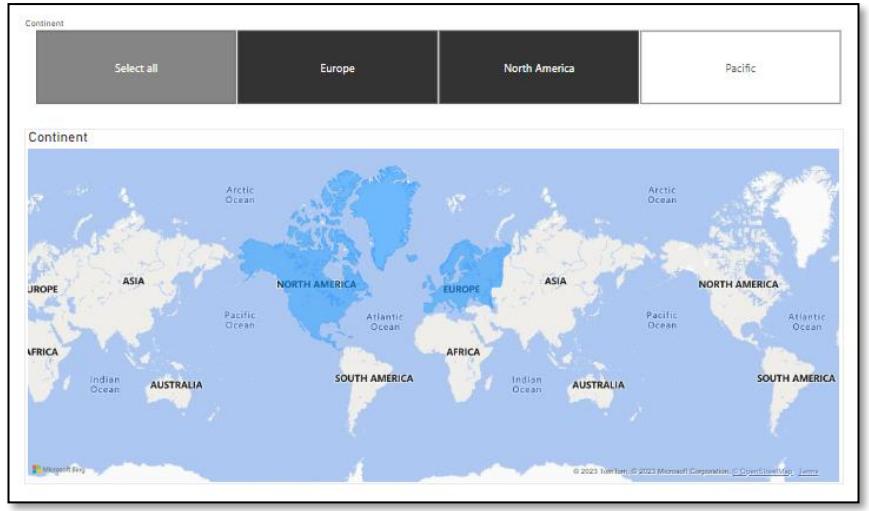
TIP: Exclude queries that don't change often, like lookups or static data tables

Defining Data Categories



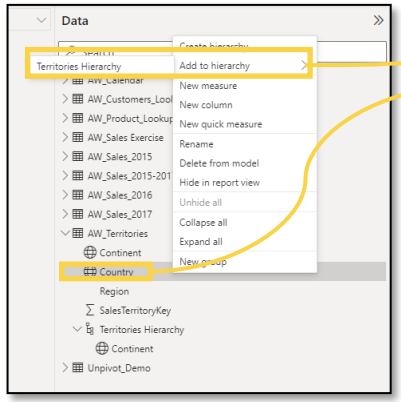
From the **"Column Tools"** tab in the **Data Category**, you can edit field properties to define specific categories

This is commonly used to help Power BI accurately map location-based fields like addresses, countries, cities, latitude/longitude coordinates, zip codes, etc

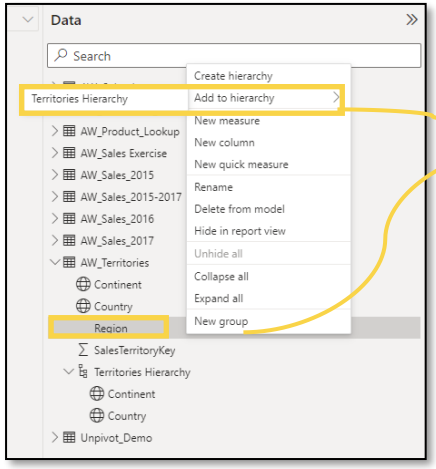


Defining Hierarchies

1) From within the **Data view**, right-click a field (or click the ellipsis) and select **"New hierarchy"** (here we've selected **"Continent"**)



2) This creates a hierarchy field containing **"Continent"**, which we've renamed **"Territories Hierarchy"**



3) Right-click other fields (like **"Region"**) and select **"Add to Hierarchy"**