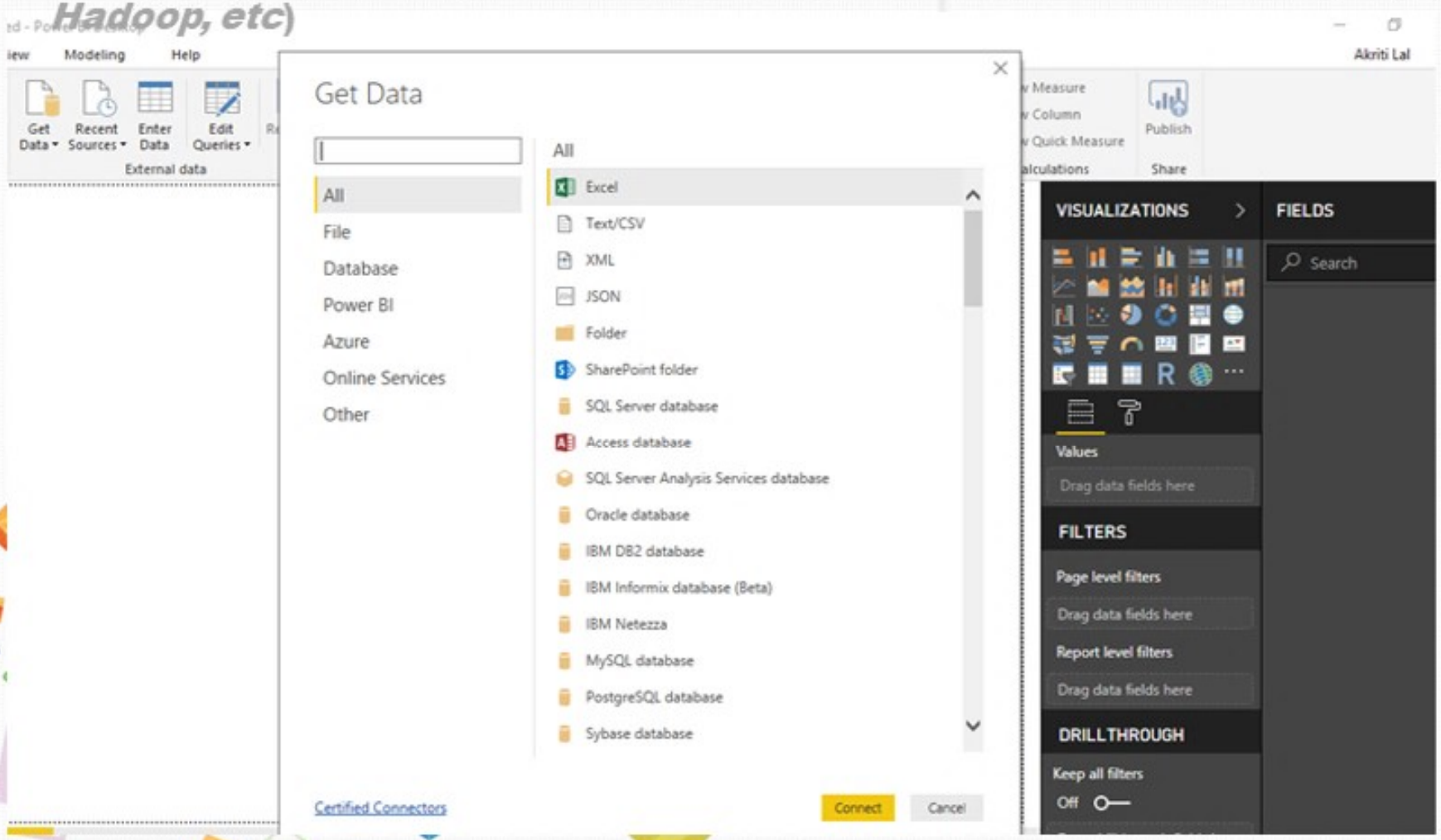
Power BI – Data Discovery : Connecting & Shaping Data

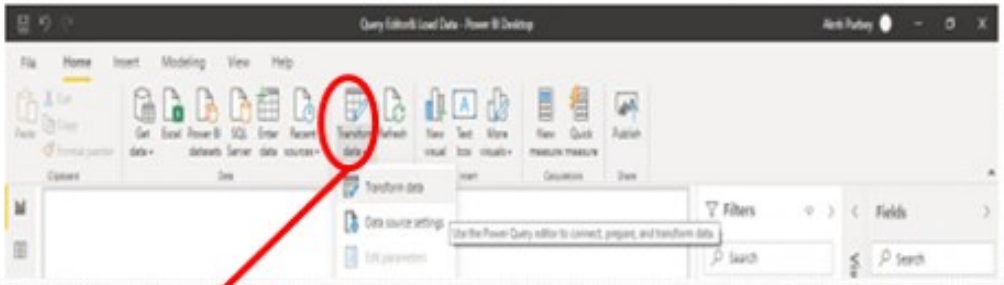
TYPES OF DATA CONNECTORS

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

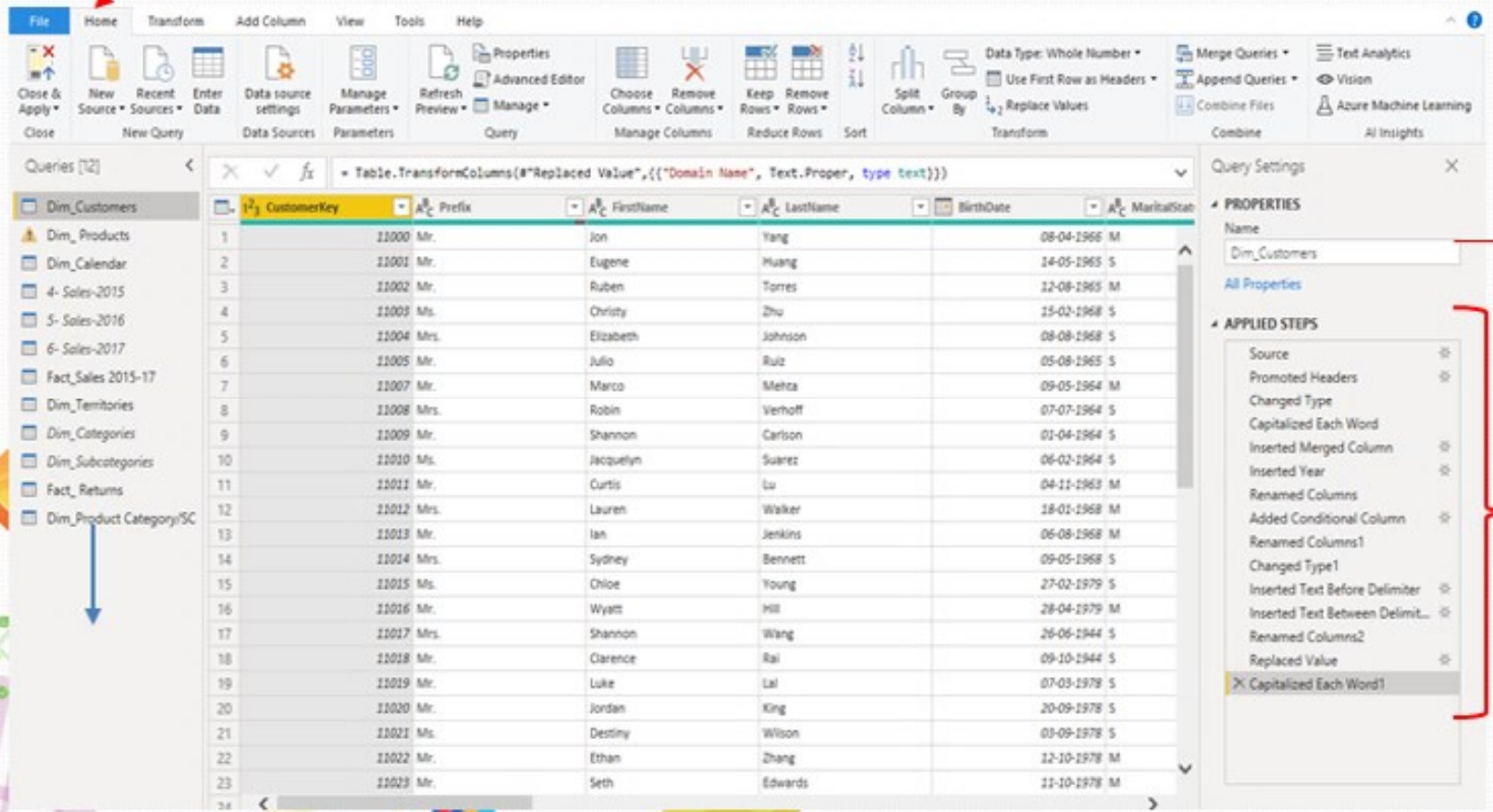
- **Flat files & Folders** (*csv, text, xls, json*), **Databases** (*SQL Server, Access, Oracle, IBM, Azure, etc*), **Online Services** (*Sharepoint, GitHub, Dynamics 365, Google Analytics, Salesforce, Power BI Service,*), **Others** (*Web feeds, R scripts, Spark, Hadoop, etc*)



THE QUERY EDITOR

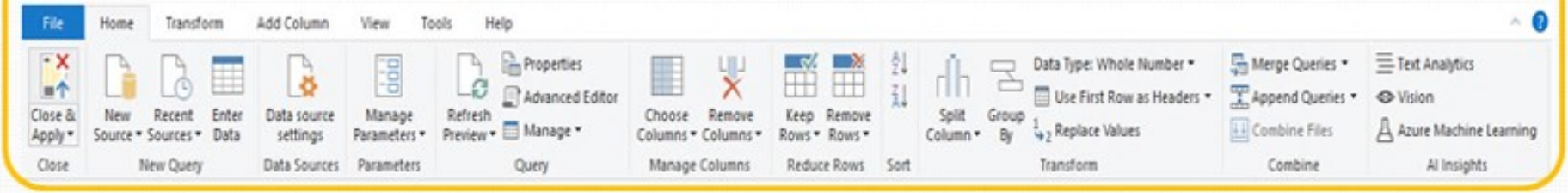


When you click on Transform Data, this opens a new window for you. This is Query Editor.

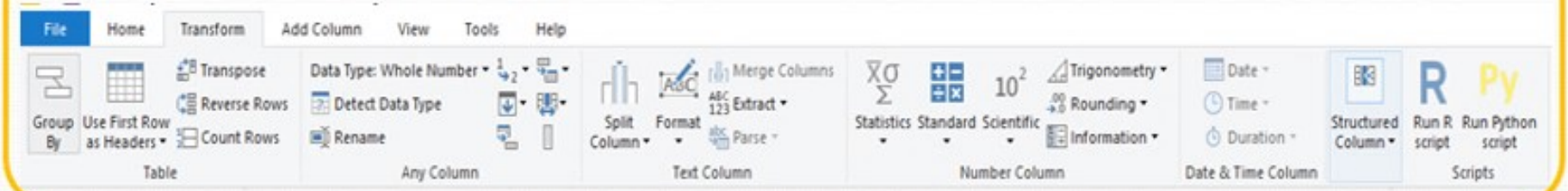


QUERY EDITING TOOLS

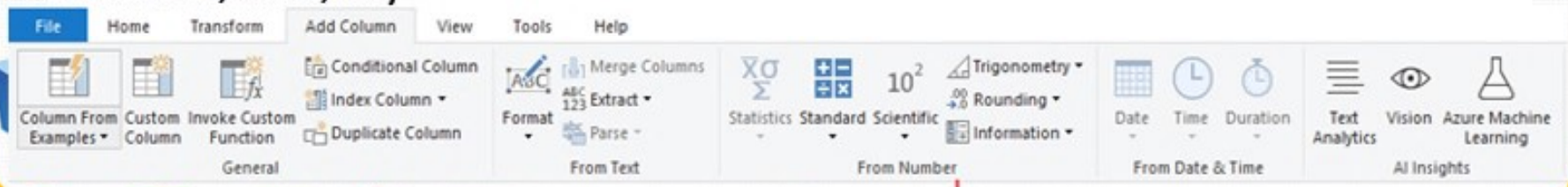
The HOME tab includes general settings and common table transformation tools



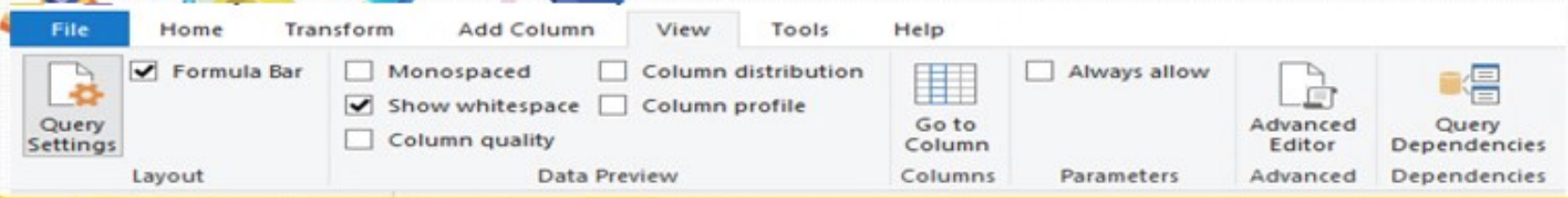
The TRANSFORM tab includes tools to modify existing columns (splitting/grouping, transposing, extracting text, etc)



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

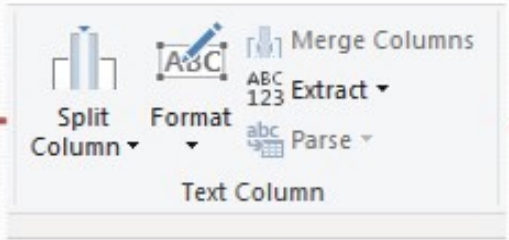


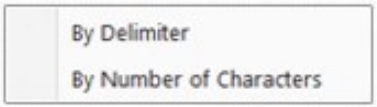
The VIEW tab on the ribbon is used to toggle whether certain panes or windows are displayed.



TEXT-SPECIFIC TOOLS

In Transform & Add Column tabs we have Text specific tools

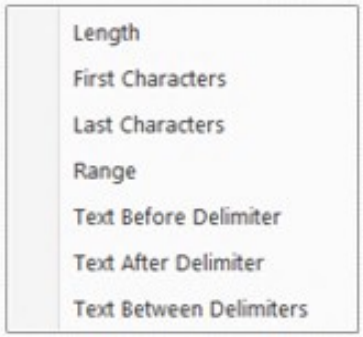




Split a text column based on either a specific delimiter or a number of characters

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 characters*



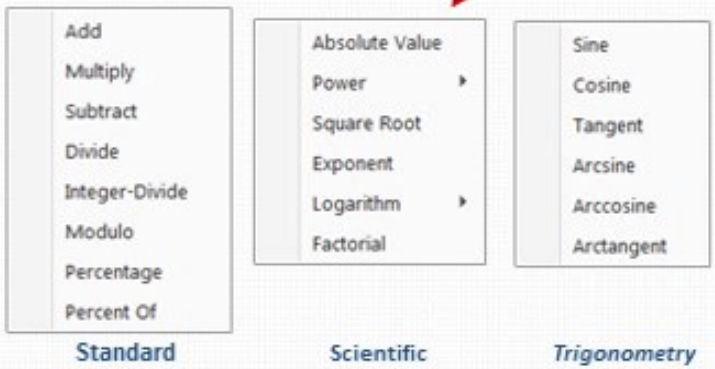
Extract characters from a text column based on fixed lengths, first/last, ranges or delimiters

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

NUMBER-SPECIFIC TOOLS

Statistics is available only in Transform tab. We do not get this option in Add Column tab

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

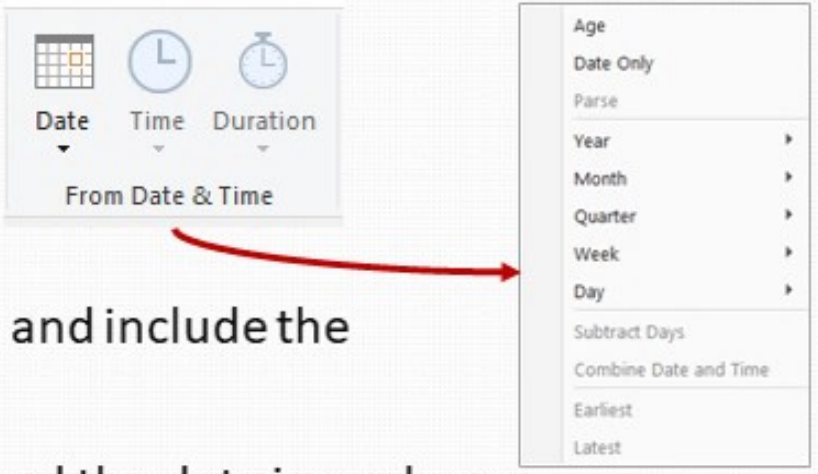


Standard, Scientific and **Trigonometry** tools allow you to apply standard operations (addition, multiplication, division, etc.) or more advanced calculations (power, logarithm, sine, tangent, etc) to each value in a column

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

In Add Column we have special time series, where can work with Date, Time & Duration

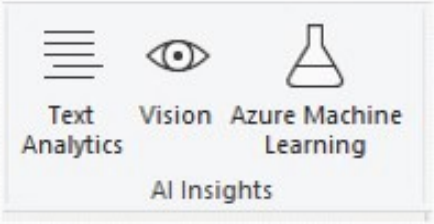


Date & Time tools are relatively straight-forward, and include the following options:

- **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)

Artificial Intelligence Insights

In Add Column we have the new AI Insight features of Power BI

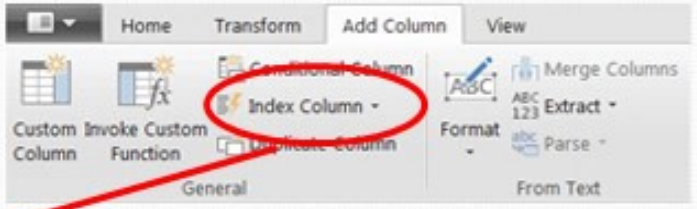


You can use AI Insights to gain access to a collection of pre-trained machine learning models that enhance your data preparation efforts. AI Insights is accessed in the **Power Query editor**, and its associated features and functions are accessed through the **Home** and **Add Column** tabs in **Power Query editor**.

ADDING INDEX COLUMNS

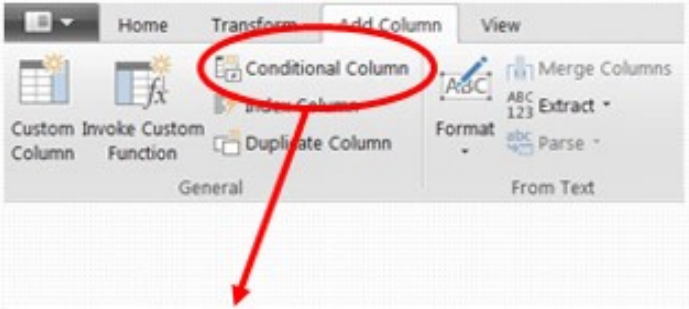
Index Columns contain a list of sequential values that can be used to identify each unique row in a table (*typically starting from 0 or 1*)

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

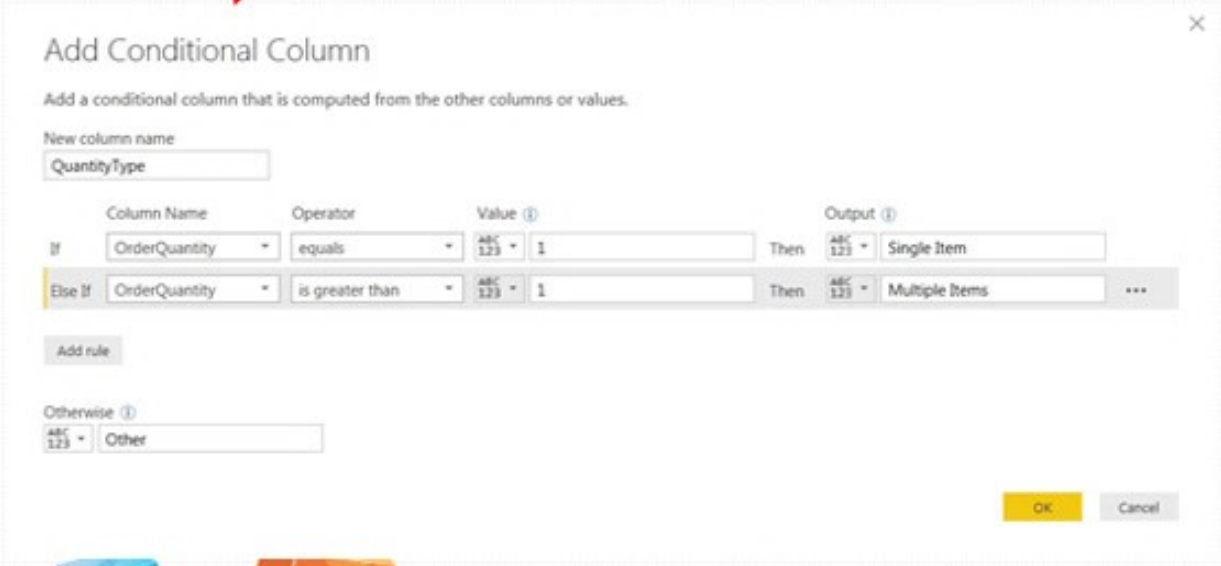


	Index	OrderDate	StockDate	OrderNumber	ProductKey	CustomerKey
1	1	1/1/2015	9/21/2001	SO45080	332	14657
2	2	1/1/2015	12/5/2001	SO45079	312	29255
3	3	1/1/2015	10/29/2001	SO45082	350	11455
4	4	1/1/2015	11/16/2001	SO45081	338	26782
5	5	1/2/2015	12/15/2001	SO45083	312	14947
6	6	1/2/2015	10/12/2001	SO45084	310	29143
7	7	1/2/2015	12/18/2001	SO45086	314	18747
8	8	1/2/2015	10/9/2001	SO45085	312	18746
9	9	1/3/2015	10/3/2001	SO45093	312	18906
10	10	1/3/2015	9/29/2001	SO45090	310	29170
11	11	1/3/2015	12/11/2001	SO45088	345	11398
12	12	1/3/2015	10/24/2001	SO45092	313	18899
13	13	1/3/2015	12/16/2001	SO45089	351	25977
14	14	1/3/2015	10/26/2001	SO45091	314	18909
15	15	1/3/2015	9/11/2001	SO45087	350	11388
16	16	1/3/2015	9/11/2001	SO45094	310	22785
17	17	1/4/2015	10/30/2001	SO45096	312	12483
18	18	1/4/2015	10/30/2001	SO45097	313	29151

ADDING CONDITIONAL COLUMNS



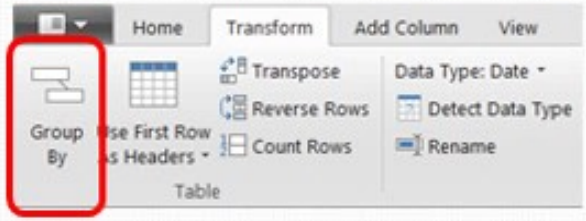
Conditional Columns allow you to define new fields based on logical rules and conditions (*IF/THEN statements*)



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

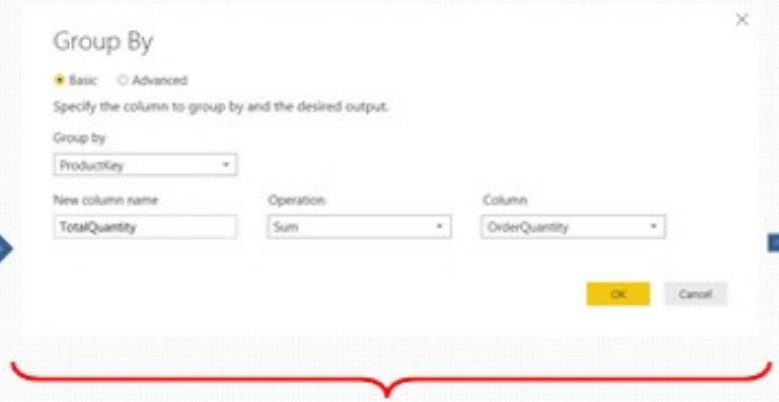
- ***If OrderQuantity =1, QuantityType = "Single Item"***
- ***If OrderQuantity >1, QuantityType = "Multiple Items"***
- ***Otherwise QuantityType = "Other"***

GROUPING & AGGREGATING DATA



Group By allows you to aggregate your data at a different level
(*i.e. transform daily data into monthly, roll up transaction-level data by store, etc*)

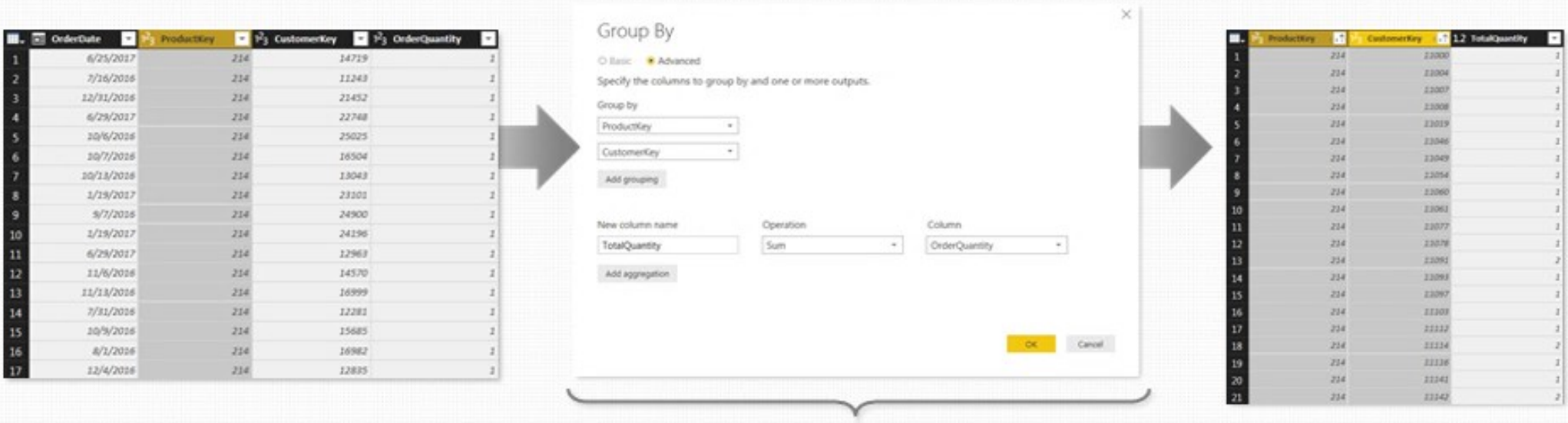
	OrderDate	ProductKey	CustomerKey	OrderQuantity
1	6/25/2017	214	14719	1
2	7/16/2016	214	11243	1
3	12/31/2016	214	21452	1
4	6/29/2017	214	22748	1
5	10/6/2016	214	25025	1
6	10/7/2016	214	16504	1
7	10/13/2016	214	13043	1
8	1/19/2017	214	23101	1
9	9/7/2016	214	24900	1
10	1/19/2017	214	24196	1
11	6/29/2017	214	12963	1
12	11/6/2016	214	14570	1
13	11/13/2016	214	16999	1
14	7/31/2016	214	12281	1
15	10/9/2016	214	15685	1
16	8/1/2016	214	16982	1
17	12/4/2016	214	12835	1



	ProductKey	1.2 TotalQuantity
1	214	2099
2	217	1940
3	222	1995
4	225	4151
5	228	392
6	231	408
7	234	424
8	237	381
9	310	169
10	311	139
11	312	179
12	313	168
13	314	157
14	320	10
15	321	55
16	322	5
17	323	34

In this case we're transforming a daily, transaction-level table into a summary of "TotalQuantity" rolled up by "ProductKey". Here You get a Table with only the Columns specified in Group BY. Those that are not a part of Group By are lost.

GROUPING & AGGREGATING DATA (ADVANCED)



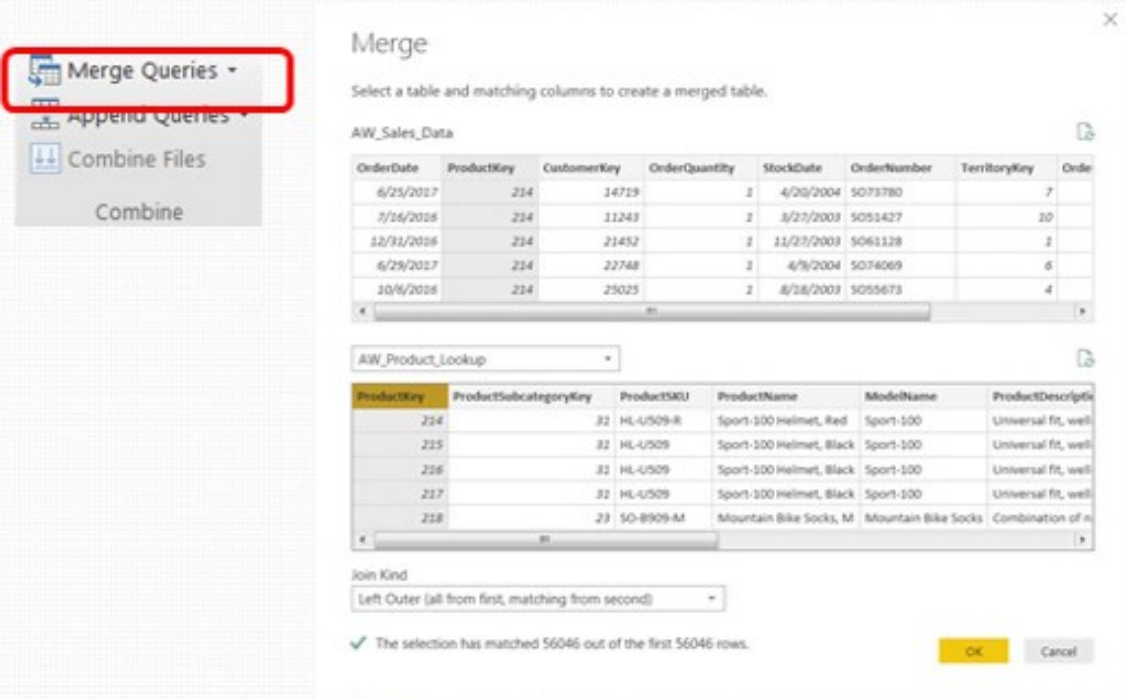
	OrderDate	ProductKey	CustomerKey	OrderQuantity
1	6/25/2017	214	14719	1
2	7/16/2016	214	11243	1
3	12/31/2016	214	21452	1
4	6/29/2017	214	22748	1
5	10/6/2016	214	25025	1
6	10/7/2016	214	16504	1
7	10/13/2016	214	13043	1
8	1/19/2017	214	23101	1
9	9/7/2016	214	24900	1
10	1/19/2017	214	24196	1
11	6/29/2017	214	12963	1
12	11/6/2016	214	14570	1
13	11/13/2016	214	16999	1
14	7/31/2016	214	12281	1
15	10/9/2016	214	15685	1
16	8/1/2016	214	16982	1
17	12/4/2016	214	12835	1

	ProductKey	CustomerKey	TotalQuantity
1	214	11000	1
2	214	11004	1
3	214	11007	1
4	214	11008	1
5	214	11019	1
6	214	11046	1
7	214	11049	1
8	214	11054	1
9	214	11060	1
10	214	11061	1
11	214	11077	1
12	214	11078	1
13	214	11091	2
14	214	11093	1
15	214	11097	1
16	214	11103	1
17	214	11112	1
18	214	11114	2
19	214	11116	1
20	214	11141	1
21	214	11142	2

*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)*

***NOTE:** This is similar to creating a PivotTable in Excel and pulling in "**Sum of OrderQuantity**" with **ProductKey** and **CustomerKey** as row labels*

MERGING QUERIES

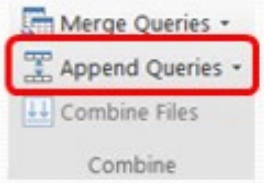


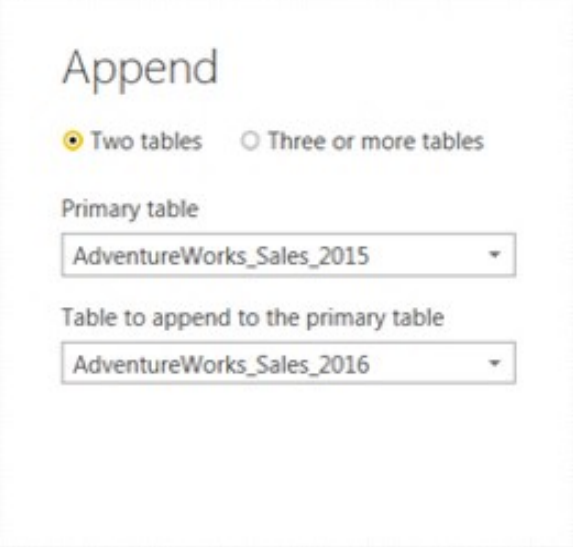
Merging queries allows you to **join tables** based on a common column (like VLOOKUP)

In this case we're merging the **Dim_Product** table with the **Dim_Product Sub Category** table, which share a common "*ProductKey*" column

merging ***adds columns*** to an existing table. This makes the table wider by bringing new columns to one of the table.

APPENDING QUERIES



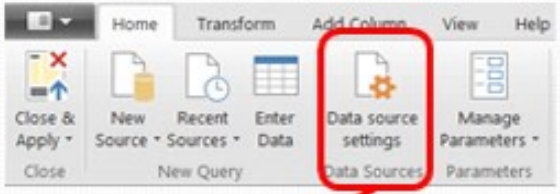


Appending queries allows you to **combine** (or ***stack***) tables that share the exact same column structure and data types

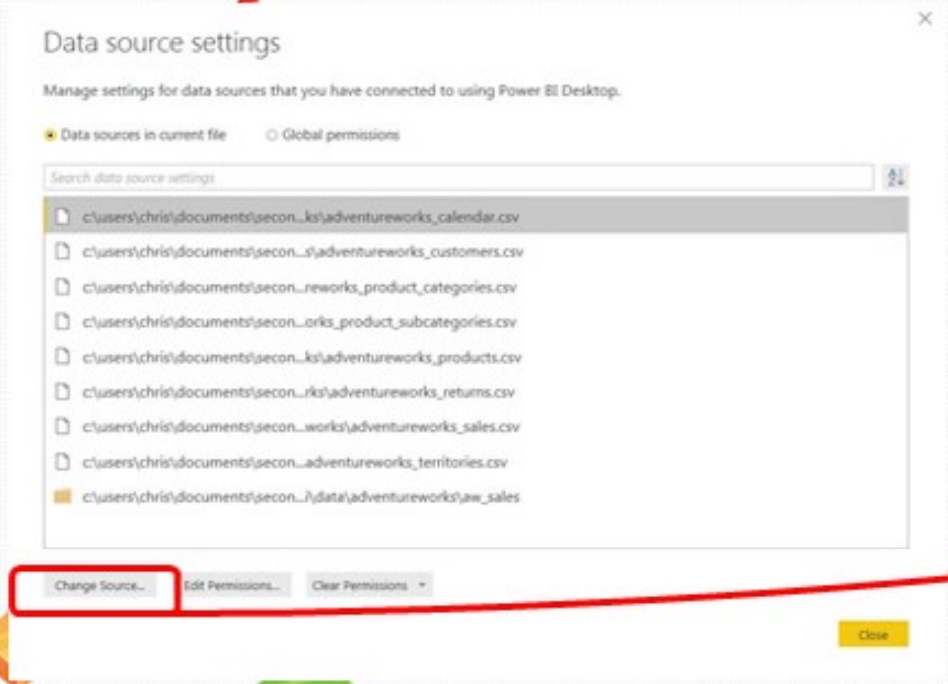
In this case we're appending the **Fact_Sales2015** table to the **Fact_Sales2016 & Fact_Sales2017** tables, as these have the same structure. They are just for different years.

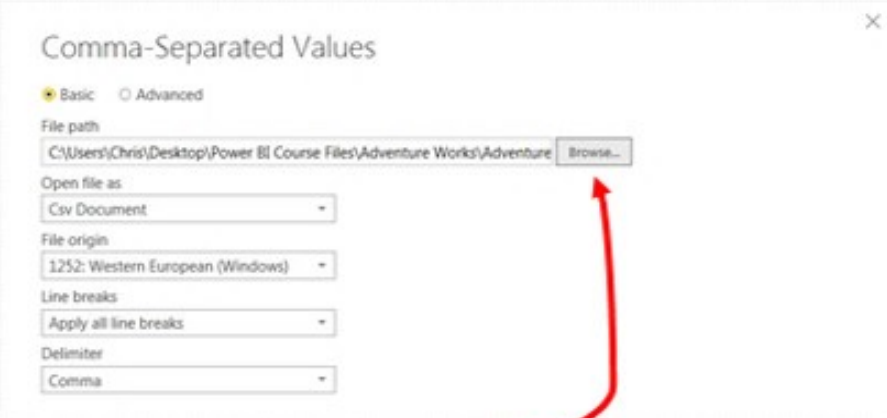
Appending ***adds rows*** to an existing table. These increase the length of the table.

DATA SOURCE SETTINGS

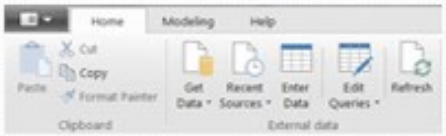


The **Data Source Settings** in the Query Editor allow you to manage data connections and permissions



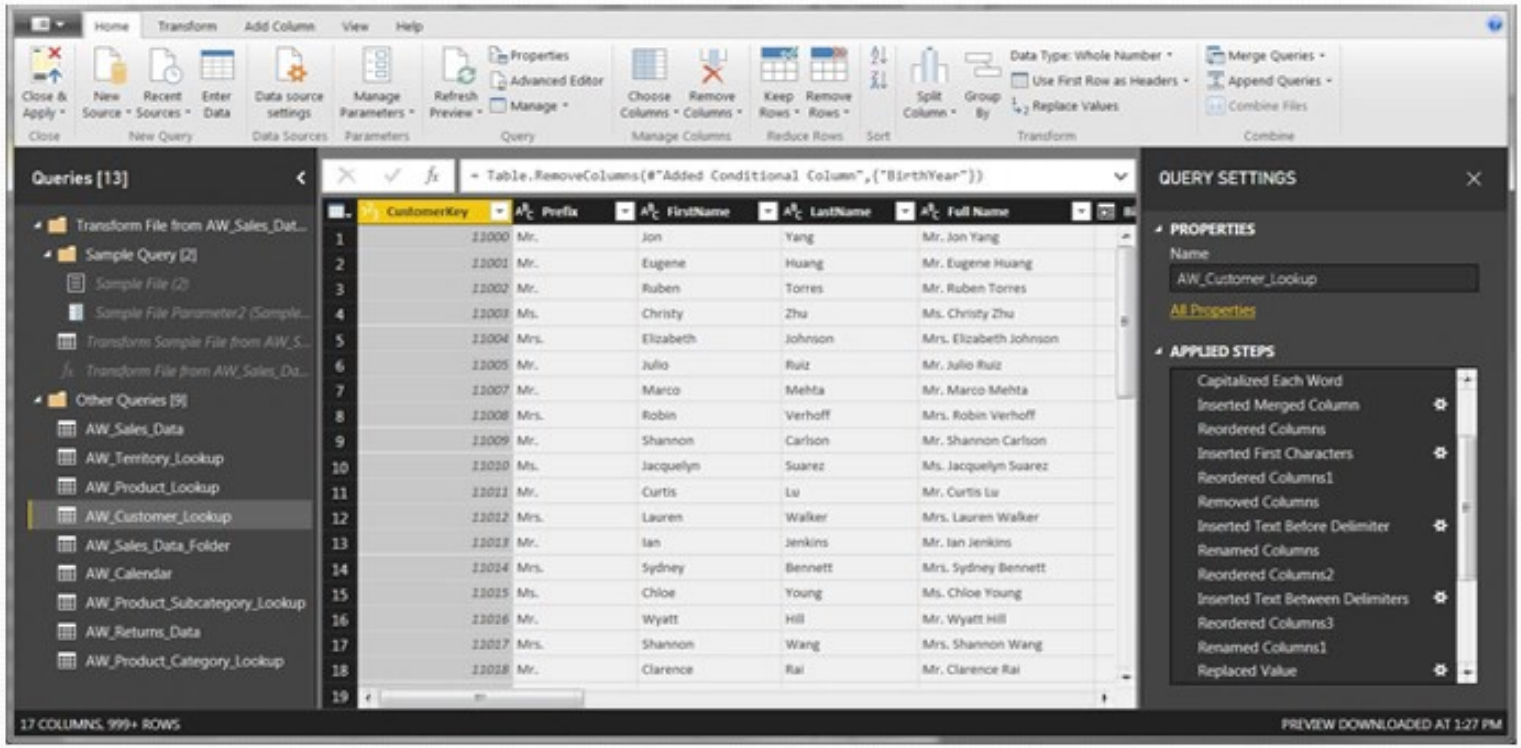


MODIFYING QUERIES



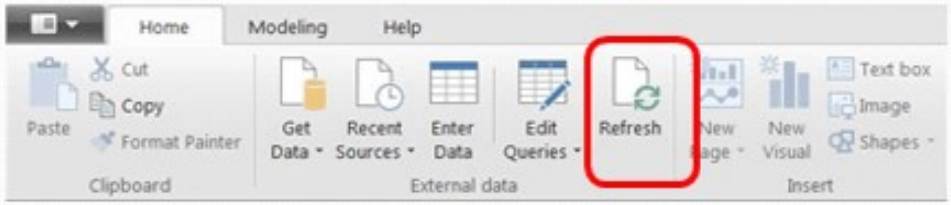
Select "**Edit Queries**" from the **Home** tab to launch the Query Editor

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