

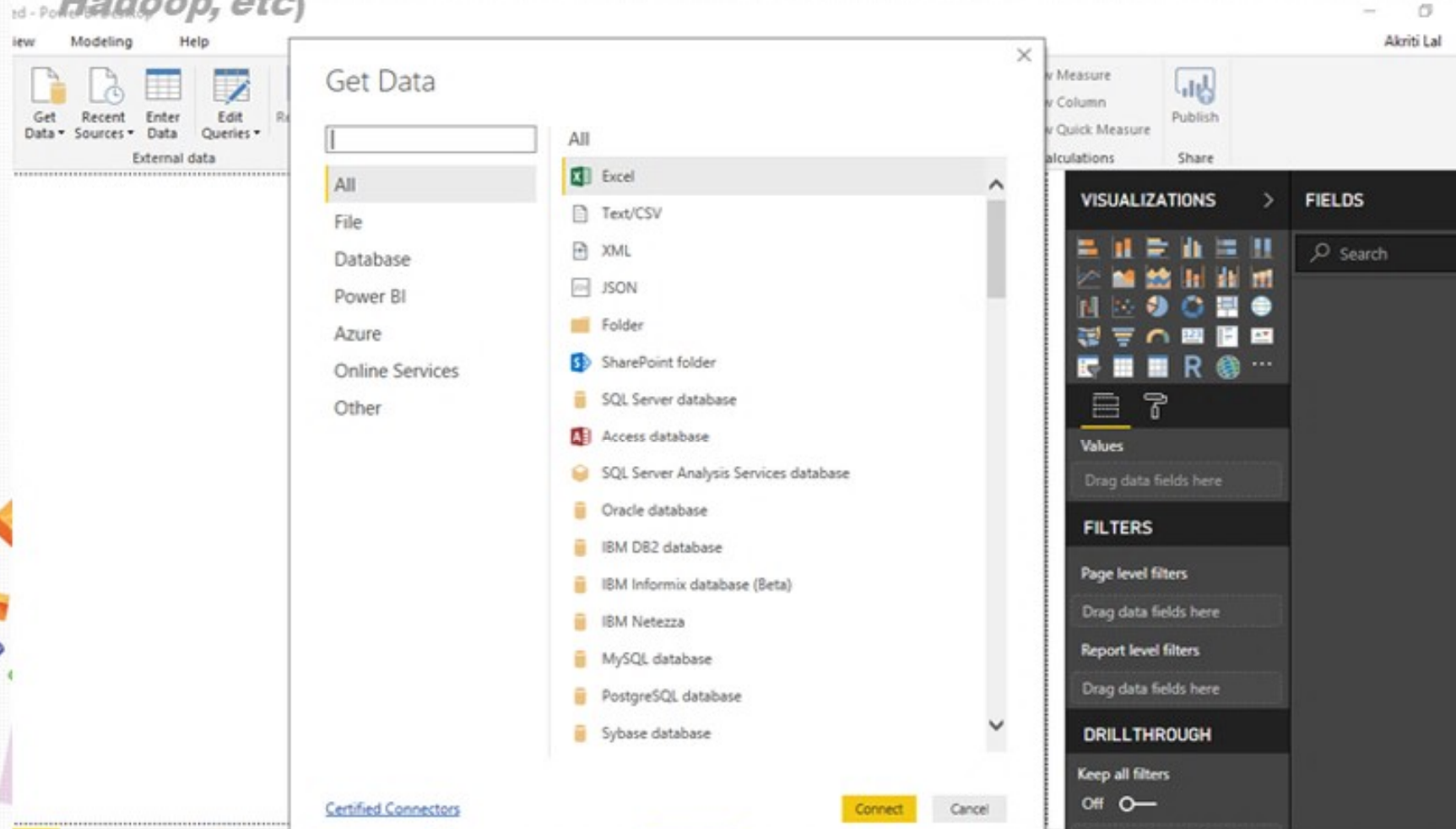
# Power BI - Data Discovery : Connecting & Shaping Data



# TYPES OF DATA CONNECTORS

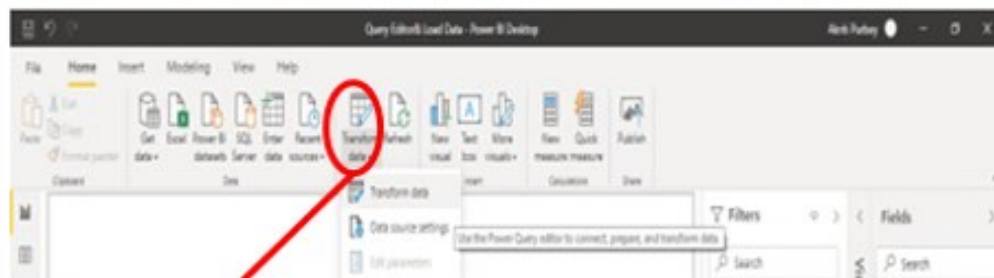
**Power BI can connect to virtually any type of source data, including (*but not limited to*):**

- **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.

A screenshot of the Query Editor interface. The main area displays a table with columns: CustomerKey, Prefix, FirstName, LastName, BirthDate, and MaritalStat. The table contains 23 rows of data. On the right, the 'APPLIED STEPS' pane is visible, listing various transformations applied to the data. A red bracket highlights the 'Capitalized Each Word1' step, with a red arrow pointing to a text box explaining its function. The 'Table Name' field at the top right of the pane is also highlighted with a red arrow.

Table Name

Any step that you do in Query Editor is saved here; like if you change the data type, or merge columns etc. You can even redo steps by clicking on X from here.

CustomerKey	Prefix	FirstName	LastName	BirthDate	MaritalStat
1	11000	Mr.	Jon	Yang	08-04-1966 M
2	11001	Mr.	Eugene	Huang	14-05-1965 S
3	11002	Mr.	Ruben	Torres	12-08-1965 M
4	11003	Ms.	Christy	Zhu	15-02-1968 S
5	11004	Mrs.	Elizabeth	Johnson	08-08-1968 S
6	11005	Mr.	Julio	Ruiz	05-08-1965 S
7	11007	Mr.	Marco	Mehra	09-05-1964 M
8	11008	Mrs.	Robin	Verhoff	07-07-1964 S
9	11009	Mr.	Shannon	Carlson	01-04-1964 S
10	11010	Ms.	Jacquelyn	Swartz	06-02-1964 S
11	11011	Mr.	Curtis	Lu	04-11-1963 M
12	11012	Mrs.	Lauren	Walker	18-01-1968 M
13	11013	Mr.	Ian	Jenkins	06-08-1968 M
14	11014	Mrs.	Sydney	Bennett	09-05-1968 S
15	11015	Ms.	Chloe	Young	27-02-1979 S
16	11016	Mr.	Wyatt	Hill	28-04-1979 M
17	11017	Mrs.	Shannon	Wang	26-06-1944 S
18	11018	Mr.	Clarence	Rai	09-10-1944 S
19	11019	Mr.	Luke	Lal	07-03-1978 S
20	11020	Mr.	Jordan	King	20-09-1978 S
21	11021	Ms.	Destiny	Wilson	08-09-1978 S
22	11022	Mr.	Ethan	Zhang	12-10-1978 M
23	11023	Mr.	Seth	Edwards	11-10-1978 M

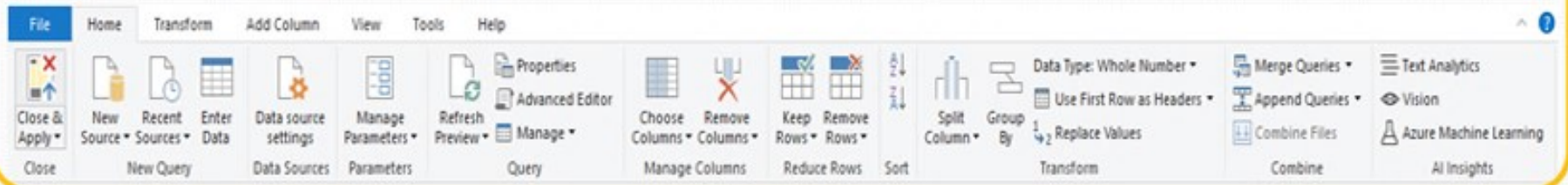
APPLIED STEPS

- Source
- Promoted Headers
- Changed Type
- Capitalized Each Word
- Inserted Merged Column
- Inserted Year
- Renamed Columns
- Added Conditional Column
- Renamed Columns1
- Changed Type1
- Inserted Text Before Delimiter
- Inserted Text Between Delimit...
- Renamed Columns2
- Replaced Value
- X Capitalized Each Word1

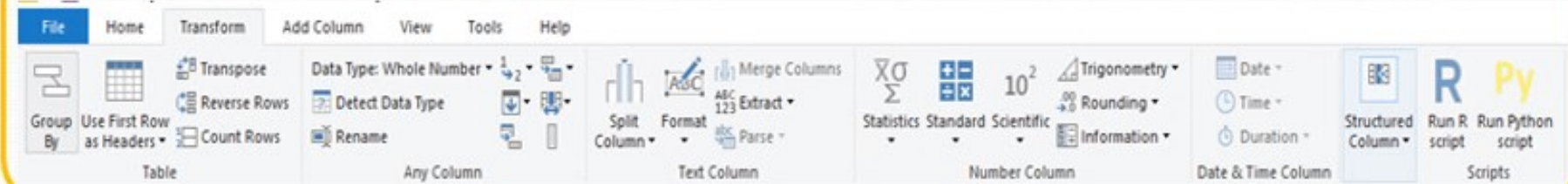


# QUERY EDITING TOOLS

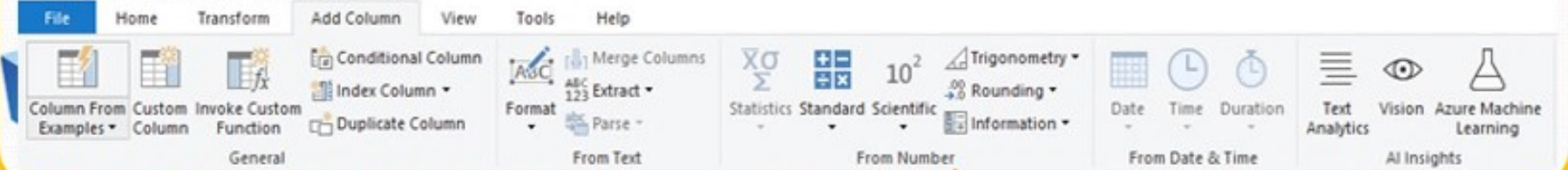
The **HOME** tab includes general settings and common table transformation tools



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)



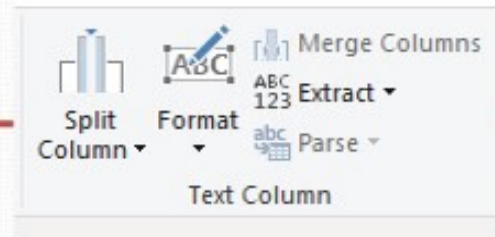
The **VIEW** tab on the ribbon is used to toggle whether certain panes or windows are displayed.





# TEXT-SPECIFIC TOOLS

In Transform & Add Column tabs we have Text specific tools



By Delimiter  
By Number of Characters

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

lowercase  
UPPERCASE  
Capitalize Each Word  
Trim  
Clean  
Add Prefix  
Add Suffix

**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*

Length  
First Characters  
Last Characters  
Range  
Text Before Delimiter  
Text After Delimiter  
Text Between Delimiters

**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



Sum
Minimum
Maximum
Median
Average
Standard Deviation
Count Values
Count Distinct Values

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

Add
Multiply
Subtract
Divide
Integer-Divide
Modulo
Percentage
Percent Of

Standard

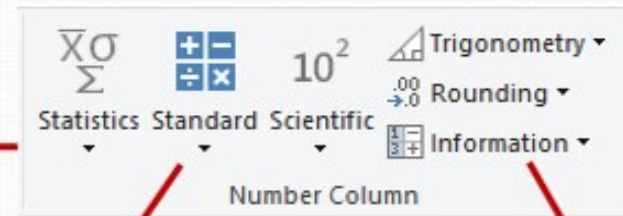
Absolute Value
Power
Square Root
Exponent
Logarithm
Factorial

Scientific

Sine
Cosine
Tangent
Arcsine
Arccosine
Arctangent

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



Is Even
Is Odd
Sign

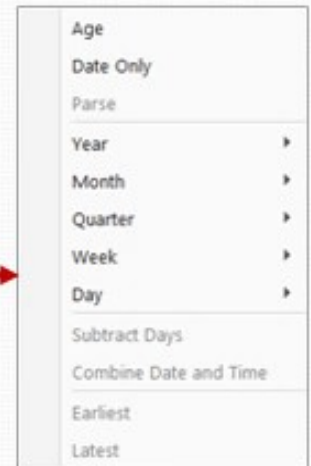
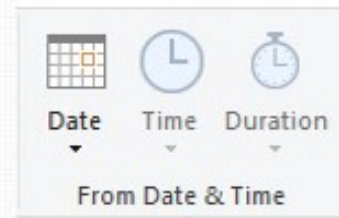
**Information** tools allow you to define binary flags (*TRUE/FALSE* or *1/0*) to mark each row in a column as even, odd, positive or negative





# DATE-SPECIFIC TOOLS

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

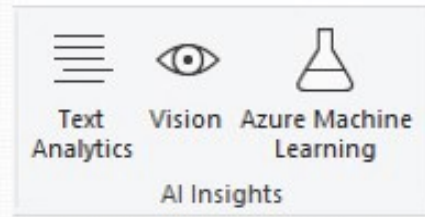


**Date & Time** tools are relatively straight-forward, and include the following options:

- **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**.

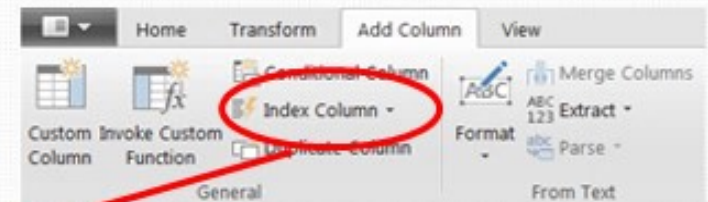




# ADDING INDEX COLUMNS

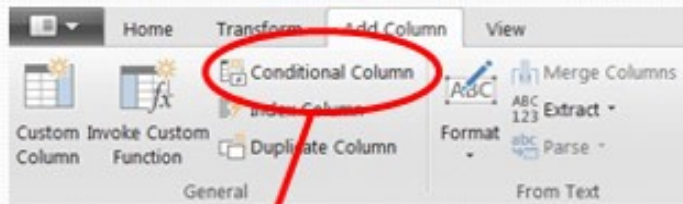
Index Columns contain a list of sequential values that can be used to identify each unique row in a table (*typically starting from 0 or 1*)

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*)

**Add Conditional Column**

Add a conditional column that is computed from the other columns or values.

New column name  
QuantityType

	Column Name	Operator	Value		Output
If	OrderQuantity	equals	1	Then	Single Item
Else If	OrderQuantity	is greater than	1	Then	Multiple Items

Add rule

Otherwise  
Other

OK Cancel

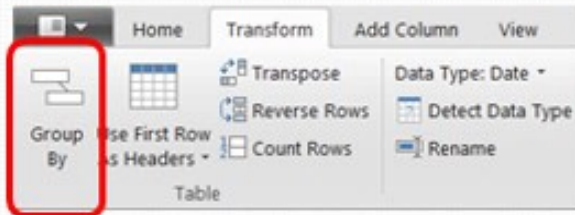
*In this case we're creating a new conditional column called "QuantityType", which depends on the values in the "OrderQuantity" column, as follows:*

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



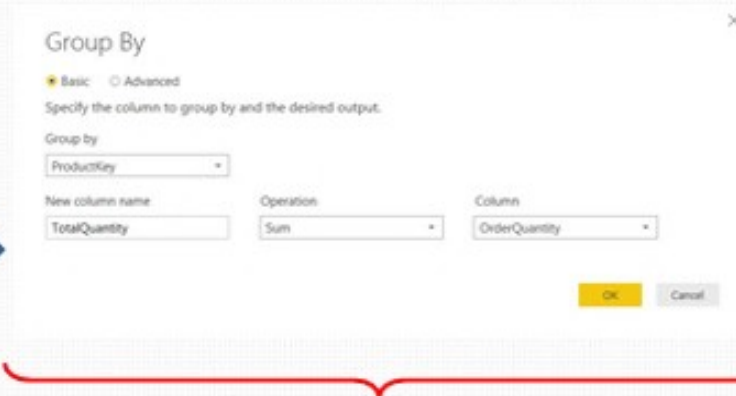


# GROUPING & AGGREGATING DATA



**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	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	21432	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

### Group By

☐ Basic ☒ Advanced

Specify the columns to group by and one or more outputs.

Group by

ProductKey

CustomerKey

Add grouping

New column name

TotalQuantity

Operation

Sum

Column

OrderQuantity

Add aggregation

OK Cancel

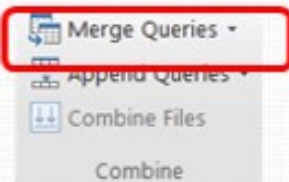
	ProductKey	CustomerKey	TotalQuantity
1	214	11000	1
2	214	11004	1
3	214	11007	1
4	214	11008	1
5	214	11019	1
6	214	11046	1
7	214	11049	1
8	214	11054	1
9	214	11060	1
10	214	11061	1
11	214	11077	1
12	214	11078	1
13	214	11091	2
14	214	11091	1
15	214	11097	1
16	214	11101	1
17	214	11112	1
18	214	11114	2
19	214	11118	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



Merge

Select a table and matching columns to create a merged table.

AW\_Sales\_Data

OrderDate	ProductKey	CustomerKey	OrderQuantity	StockDate	OrderNumber	TerritoryKey	Order
6/21/2017	214	14719	2	4/20/2004	SO73780	7	
7/16/2016	214	11243	1	3/27/2003	SO51427	10	
12/31/2016	214	21452	1	11/27/2003	SO61128	1	
6/29/2017	214	22748	1	4/9/2004	SO74069	6	
10/6/2016	214	25025	1	8/18/2003	SO55673	4	

AW\_Product\_Lookup

ProductKey	ProductSubcategoryKey	ProductSKU	ProductName	ModelName	ProductDescription
214	32	HL-U509-R	Sport-100 Helmet, Red	Sport-100	Universal fit, well
215	32	HL-U509	Sport-100 Helmet, Black	Sport-100	Universal fit, well
216	32	HL-U509	Sport-100 Helmet, Black	Sport-100	Universal fit, well
217	32	HL-U509	Sport-100 Helmet, Black	Sport-100	Universal fit, well
218	29	SO-B909-M	Mountain Bike Socks, M	Mountain Bike Socks	Combination of n

Join Kind

Left Outer (all from first, matching from second)

✓ The selection has matched 56046 out of the first 56046 rows.

OK Cancel

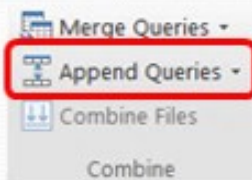
Merging queries allows you to **join tables** based on a common column (like VLOOKUP)

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



## Append

☒ Two tables ☐ Three or more tables

Primary table

AdventureWorks\_Sales\_2015

Table to append to the primary table

AdventureWorks\_Sales\_2016

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

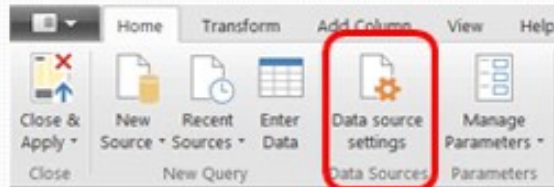
In this case we're appending the **Fact\_Sales2015** table to the **Fact\_Sales2016 & Fact\_Sales2017** tables, as these have the same structure. They are just for different years.

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

## Data source settings

Manage settings for data sources that you have connected to using Power BI Desktop.

☒ Data sources in current file ☐ Global permissions

Search data source settings

- c:\users\christ\documents\secon...ks\adventureworks\_calendar.csv
- c:\users\christ\documents\secon...s\adventureworks\_customers.csv
- c:\users\christ\documents\secon...eworks\_product\_categories.csv
- c:\users\christ\documents\secon...orks\_product\_subcategories.csv
- c:\users\christ\documents\secon...ks\adventureworks\_products.csv
- c:\users\christ\documents\secon...rks\adventureworks\_returns.csv
- c:\users\christ\documents\secon...works\adventureworks\_sales.csv
- c:\users\christ\documents\secon...adventureworks\_territories.csv
- c:\users\christ\documents\secon...data\adventureworks\aw\_sales

## Comma-Separated Values

☒ Basic ☐ Advanced

File path

C:\Users\Chris\Desktop\Power BI Course Files\Adventure Works\Adventure

Open file as

Csv Document

File origin

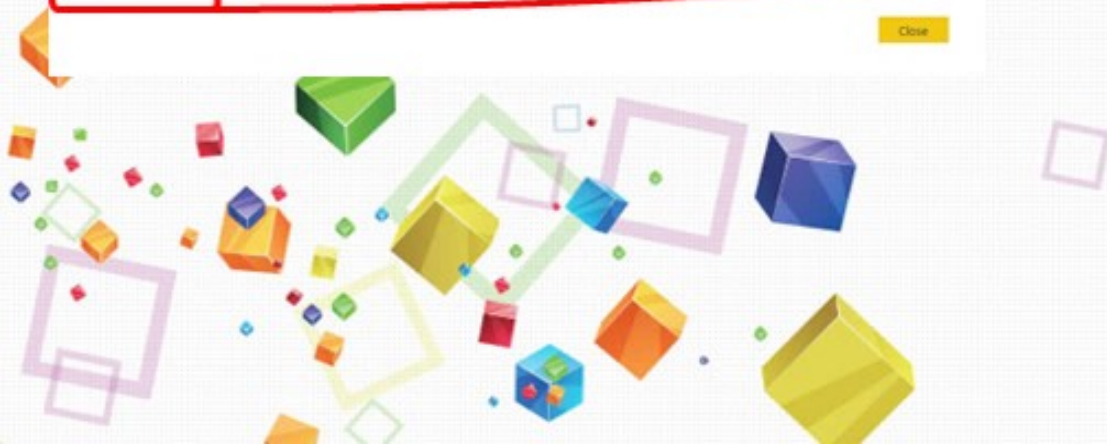
1252: Western European (Windows)

Line breaks

Apply all line breaks

Delimiter

Comma



# 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

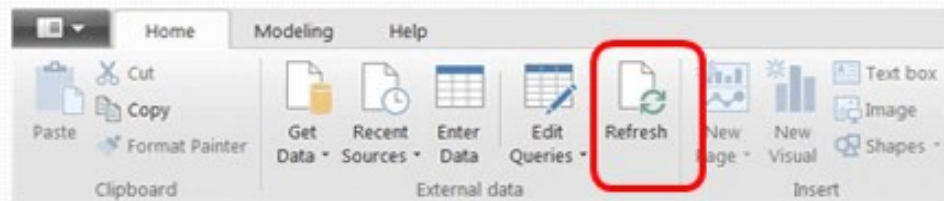
A screenshot of the Power BI Query Editor interface. On the left is the 'Queries' pane showing a list of queries, with 'AW\_Customer\_Lookup' selected. The main area displays a data preview for the selected query, showing columns: CustomerKey, Prefix, FirstName, LastName, and Full Name. On the right is the 'QUERY SETTINGS' pane, which includes a 'PROPERTIES' section for the query name and an 'APPLIED STEPS' section listing various transformations like 'Capitalized Each Word', 'Inserted Merged Column', etc. The bottom status bar indicates '17 COLUMNS, 999+ ROWS' and 'PREVIEW DOWNLOADED AT 1:27 PM'.

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

