

PowerQuery Essentials - I

Satyam Govila

PowerQuery : Recap

- Power Query is a business intelligence tool available in Excel that allows you to import data from many different sources and then clean, transform and reshape your data as needed.
- It allows you to set up a query once and then reuse it with a simple refresh.
- It's also pretty powerful as it can import and clean millions of rows into the data model for analysis after.
- The user interface is intuitive and well laid out so it's really easy to pick up.
- It's an incredibly short learning curve when compared to other Excel tools like formulas or VBA.
- The best part about it, is you don't need to learn or use any code to do any of it. The power query editor records all your transformations step by step and converts them for you!

Data Ingestion in PowerQuery

Importing your data with **Power Query** is simple. Excel provides many common data connections that are accessible from the **Data** tab and can be found from the **Get Data** command.

- Get data from a **single file** such as an Excel workbook, Text or CSV file, XML and JSON files. You can also import multiple files from within a given folder.
- Get data from **various databases** such as SQL Server, Microsoft Access, Analysis Services, SQL Server Analysis Server, Oracle, IBM DB2, MySQL, PostgreSQL, Sybase, Teradata and SAP HANA databases.
- Get data from **Microsoft Azure**
- Get data from **online services** like Sharepoint, Microsoft Exchange, Dynamics 365, Facebook and Salesforce.
- Get data from **other sources** like a table or range inside the current workbook, from the web, a Microsoft Query, Hadoop, OData feed, ODBC and OLEDB.
- We can **merge two queries together** similar to joining two queries in SQL.
- We can **append a query to another query** similar to a union of two queries in SQL.

Data Ingestion in PowerQuery

There are a couple of the more common query types available in the top level of the ribbon commands found in the **Get & Transform** section of the **Data** tab. From here we can easily access the **From Text/CSV**, **From Web** and **From Table/Range** queries.

Data Ingestion in PowerQuery : Example

Let's suppose you want to import some data from an Excel workbook **Office Supply Sales Data.xlsx**

It contains sales data on one sheet called **Sales Data** and customer data on another sheet called **Customer Data**.

Both sheets of data start in cell A1 and the first row of the data contains column headers.

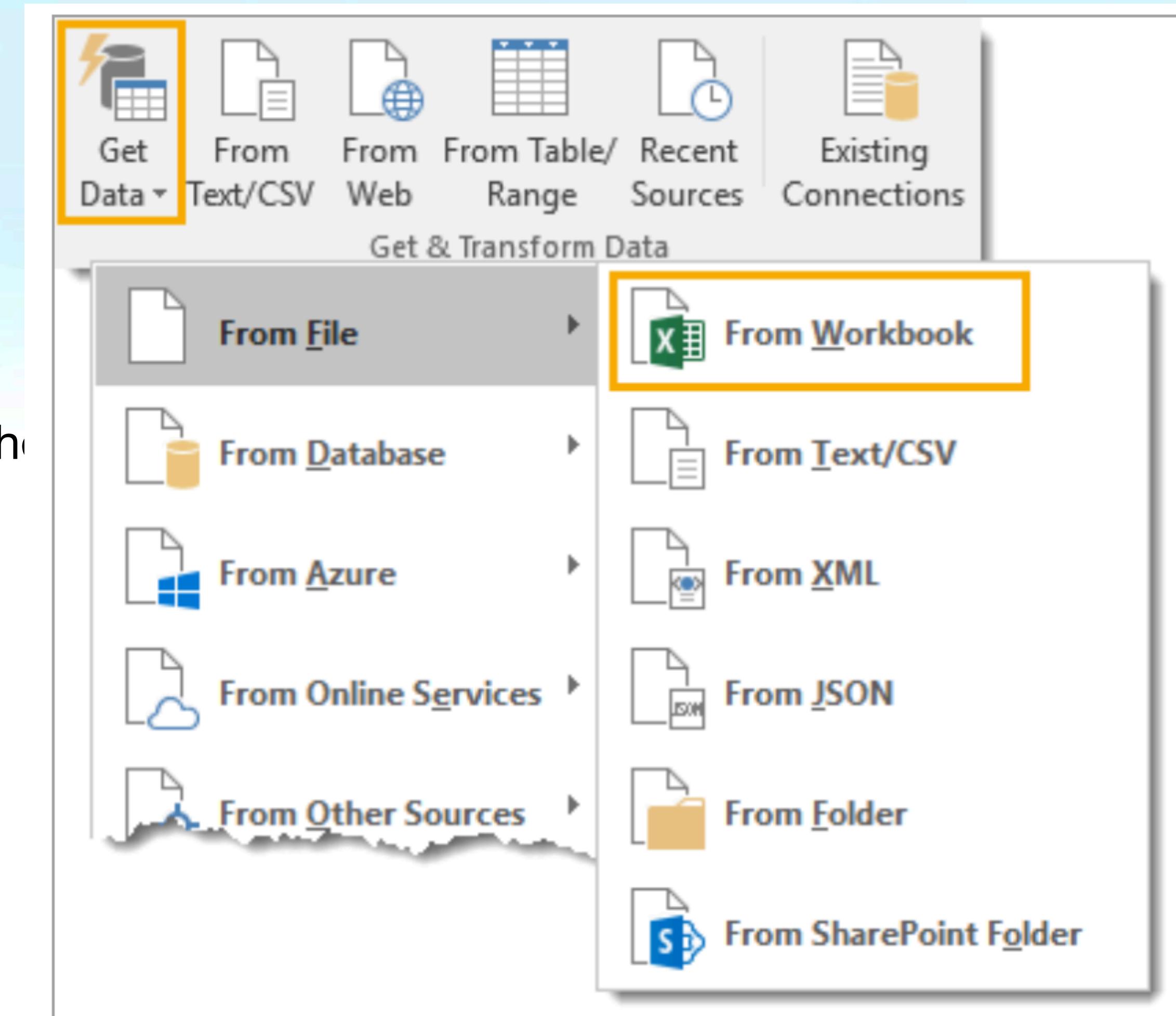
Data Ingestion in PowerQuery : Example

Steps to load data from WorkBook:-

Go to the **Data** tab and select the **Get Data** command in the **Get & Transform Data** section.

Then go to **From File** and choose **From Workbook**.

This will open a **file picker** menu where you can navigate to the file and press the **Import** button.



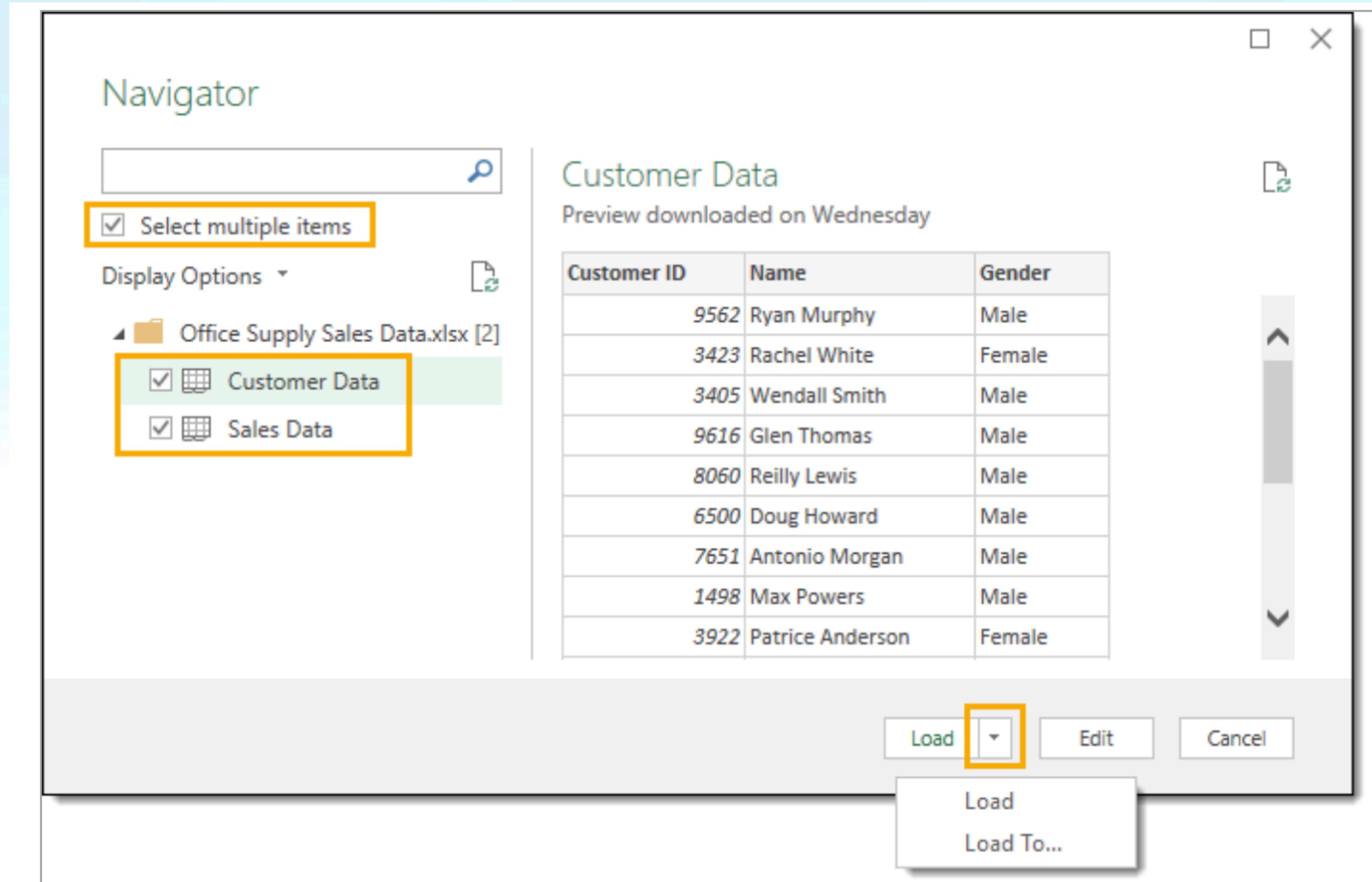
Data Ingestion in PowerQuery : Example

After selecting the file you want to import,
the data preview **Navigator** window will open.

This will give you a list of all the objects
available to import from the workbook.

Check the box to **Select multiple items**
since we will be importing data from two
different sheets.

Now we can check both the **Customer Data**
and **Sales Data**.



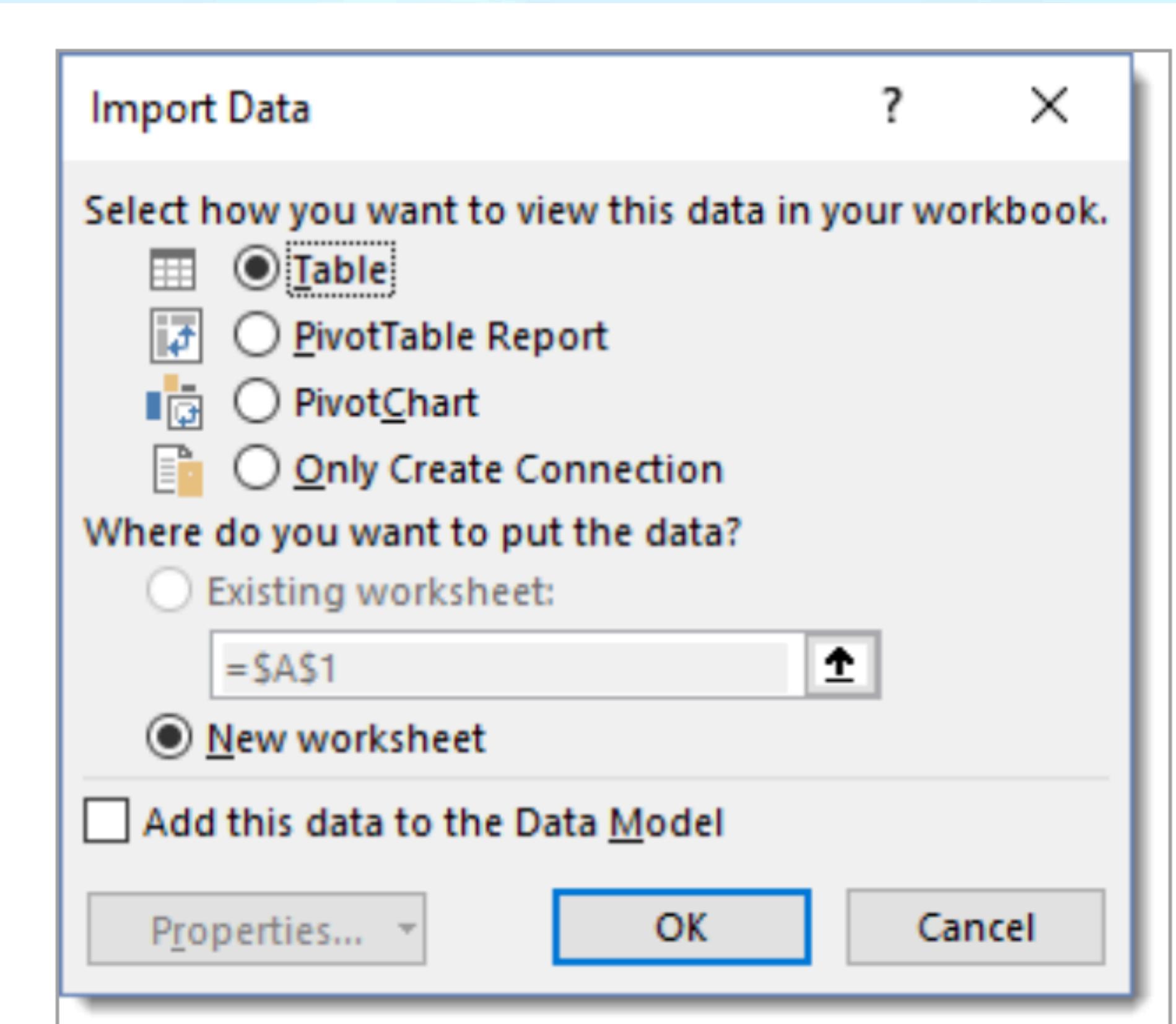
Data Ingestion in PowerQuery : Example

In this simple example, we will bypass the editor

and go straight to loading the data into Excel.

Press the small arrow next to the Load button

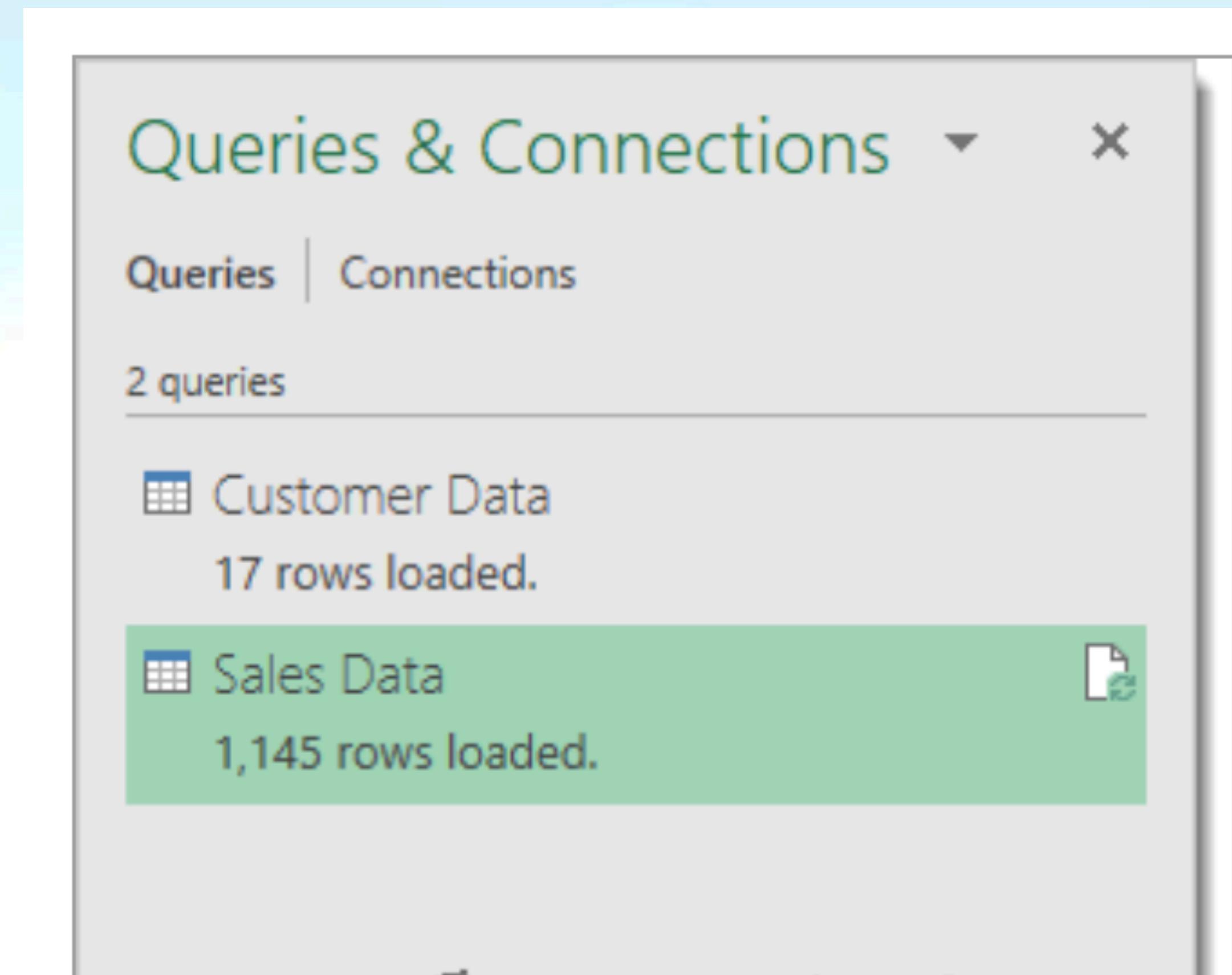
to access the **Load To** options.



Data Ingestion in PowerQuery : Example

Now the tables are loaded into new sheets in Excel and

we also have two queries which can
quickly be refreshed if the data in the
original workbook is ever updated.



The Query Editor

The screenshot illustrates the Microsoft Power Query Editor interface with various features highlighted by numbered circles:

- 1** The ribbon bar at the top, specifically the **Home** tab.
- 2** The **Queries [2]** pane on the left, listing the current queries: **Sales_Data** and **Customer_Data**.
- 3** The main data grid displaying a table of sales data with columns: Customer ID, State, Product, Order Date, Cost, and Price.
- 4** The formula bar above the grid showing the query formula: `= Table.TransformColumnTypes(#"Promoted Headers",{{"Customer ID", type text}, {"State",`
- 5** The **Properties** pane on the right, showing the query name **Sales_Data**.
- 6** The **Applied Steps** pane on the right, listing the steps taken: Source, Navigation, Promoted Headers, and Changed Type.

Below the grid, the status bar shows **6 COLUMNS, 999+ ROWS** and **PREVIEW DOWNLOADED AT 1:05 AM**.

| Customer ID | State | Product | Order Date | Cost | Price |
|-------------|----------------|------------|------------|--------|-------|
| 3405 | Arizona | Paperclips | 2017-05-29 | 2.15 | 2.79 |
| 3651 | Virginia | Pencil | 2016-02-28 | 4.46 | 5.65 |
| 7687 | Arizona | Notepad | 2016-04-22 | 1.19 | 1.59 |
| 1904 | Nevada | Pen | 2017-08-31 | 4.79 | 5.95 |
| 600 | Alabama | Eraser | 2017-03-01 | 3.04 | 3.75 |
| 87 | Arizona | Paperclips | 2017-06-23 | 2.15 | 2.79 |
| 3198 | North Carolina | Notebook | 2016-03-07 | 1.76 | 2.35 |
| 9616 | California | Pencil | 2017-06-17 | 4.46 | 5.65 |
| 3423 | South Carolina | Pen | 2016-02-13 | 4.79 | 5.95 |
| 9616 | California | Chair | 2017-09-11 | 60.38 | 115 |
| 9616 | California | Pen | 2016-09-18 | 4.79 | 5.95 |
| 6500 | Alabama | Pen | 2016-02-03 | 4.79 | 5.95 |
| 8060 | Nevada | Paperclips | 2016-06-07 | 2.15 | 2.79 |
| 3198 | North Carolina | Notepad | 2017-01-13 | 1.19 | 1.59 |
| 3198 | North Carolina | Chair | 2017-04-29 | 60.38 | 115 |
| 3880 | Arizona | Desk | 2017-02-01 | 109.45 | 199 |
| 3880 | Arizona | Notebook | 2017-01-21 | 1.76 | 2.35 |
| 1498 | California | Chair | 2017-02-15 | 60.38 | 115 |
| 3880 | Arizona | Pen | 2016-02-01 | 4.79 | 5.95 |
| 9616 | California | Eraser | 2017-01-10 | 3.04 | 3.75 |
| 3651 | Virginia | Eraser | 2017-03-17 | 3.04 | 3.75 |

The Query Editor

The screenshot illustrates the Microsoft Power Query Editor interface with various features highlighted by numbered circles:

- 1** The ribbon bar at the top, specifically the Home tab.
- 2** The Queries list pane on the left, showing two queries: Sales_Data and Customer_Data.
- 3** The main data grid area displaying a table of sales data.
- 4** The formula bar at the top of the data grid.
- 5** The Properties pane on the right, showing the query name "Sales_Data".
- 6** The Applied Steps pane on the right, listing the steps taken to process the query, including "Promoted Headers" and "Changed Type".

Below the main area, the status bar shows "6 COLUMNS, 999+ ROWS" and "PREVIEW DOWNLOADED AT 1:05 AM".

| Customer ID | State | Product | Order Date | Cost | Price |
|-------------|----------------|------------|------------|--------|-------|
| 3405 | Arizona | Paperclips | 2017-05-29 | 2.15 | 2.79 |
| 3651 | Virginia | Pencil | 2016-02-28 | 4.46 | 5.65 |
| 7687 | Arizona | Notepad | 2016-04-22 | 1.19 | 1.59 |
| 1904 | Nevada | Pen | 2017-08-31 | 4.79 | 5.95 |
| 600 | Alabama | Eraser | 2017-03-01 | 3.04 | 3.75 |
| 87 | Arizona | Paperclips | 2017-06-23 | 2.15 | 2.79 |
| 3198 | North Carolina | Notebook | 2016-03-07 | 1.76 | 2.35 |
| 9616 | California | Pencil | 2017-06-17 | 4.46 | 5.65 |
| 3423 | South Carolina | Pen | 2016-02-13 | 4.79 | 5.95 |
| 9616 | California | Chair | 2017-09-11 | 60.38 | 115 |
| 9616 | California | Pen | 2016-09-18 | 4.79 | 5.95 |
| 6500 | Alabama | Pen | 2016-02-03 | 4.79 | 5.95 |
| 8060 | Nevada | Paperclips | 2016-06-07 | 2.15 | 2.79 |
| 3198 | North Carolina | Notepad | 2017-01-13 | 1.19 | 1.59 |
| 3198 | North Carolina | Chair | 2017-04-29 | 60.38 | 115 |
| 3880 | Arizona | Desk | 2017-02-01 | 109.45 | 199 |
| 3880 | Arizona | Notebook | 2017-01-21 | 1.76 | 2.35 |
| 1498 | California | Chair | 2017-02-15 | 60.38 | 115 |
| 3880 | Arizona | Pen | 2016-02-01 | 4.79 | 5.95 |
| 9616 | California | Eraser | 2017-01-10 | 3.04 | 3.75 |
| 3651 | Virginia | Eraser | 2017-03-17 | 3.04 | 3.75 |

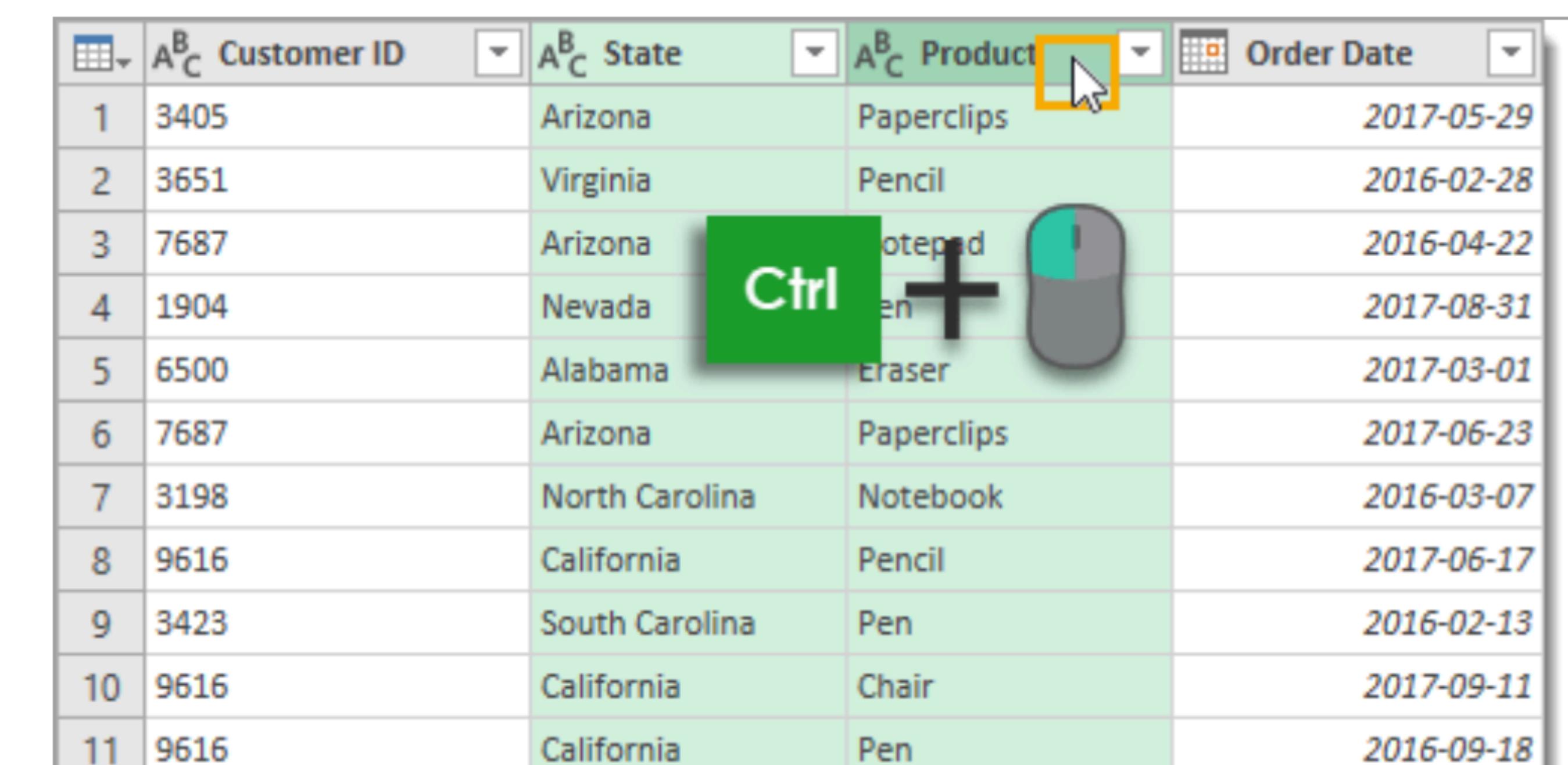
Data Preview

The main job of the data preview area is to apply

transformation steps to your data

and show a preview of these steps

you're applying.



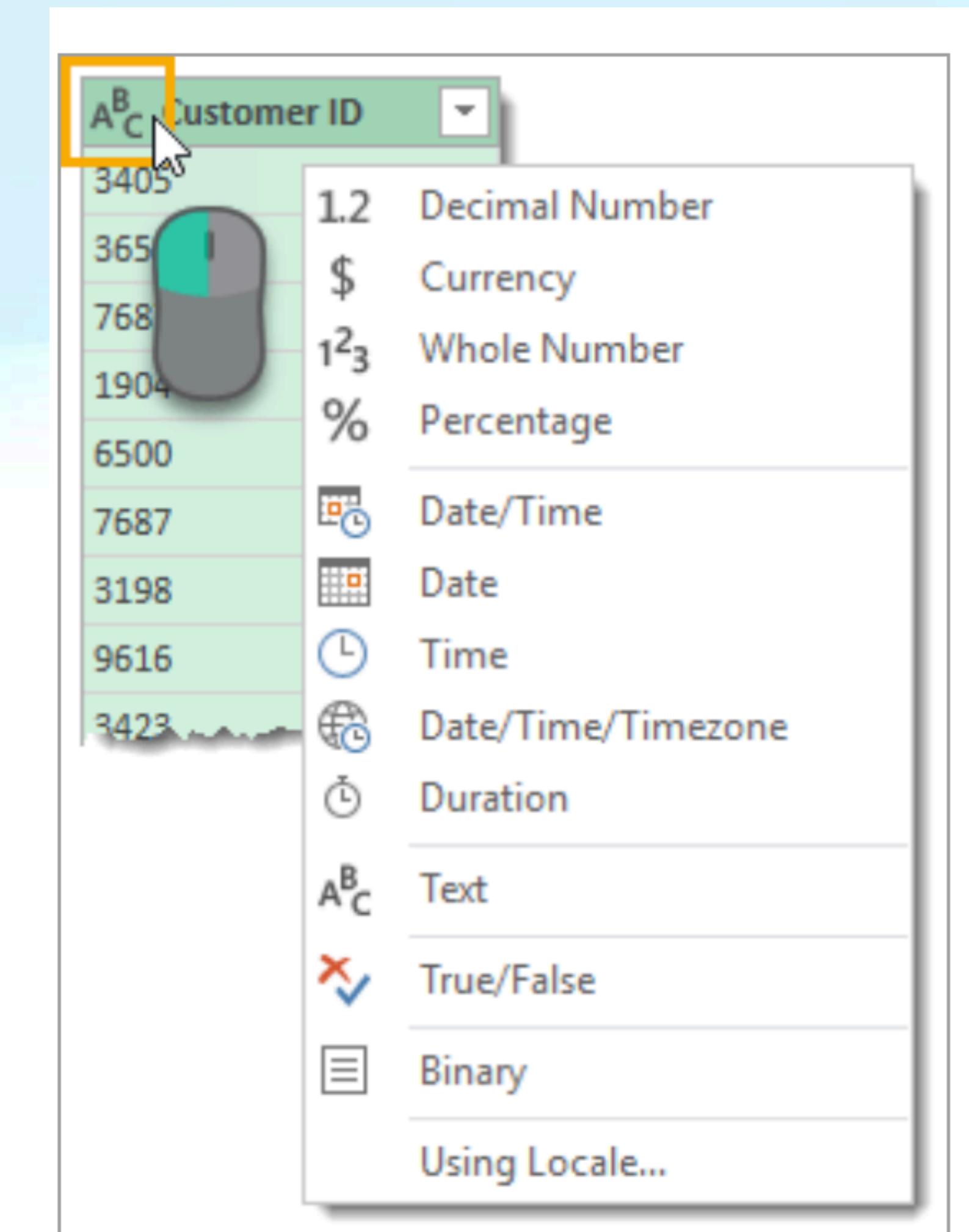
| | Customer ID | State | Product | Order Date |
|----|-------------|----------------|------------|------------|
| 1 | 3405 | Arizona | Paperclips | 2017-05-29 |
| 2 | 3651 | Virginia | Pencil | 2016-02-28 |
| 3 | 7687 | Arizona | Notepad | 2016-04-22 |
| 4 | 1904 | Nevada | Pen | 2017-08-31 |
| 5 | 6500 | Alabama | Eraser | 2017-03-01 |
| 6 | 7687 | Arizona | Paperclips | 2017-06-23 |
| 7 | 3198 | North Carolina | Notebook | 2016-03-07 |
| 8 | 9616 | California | Pencil | 2017-06-17 |
| 9 | 3423 | South Carolina | Pen | 2016-02-13 |
| 10 | 9616 | California | Chair | 2017-09-11 |
| 11 | 9616 | California | Pen | 2016-09-18 |

Data Preview

1. Select a single column with a **left click** on the column heading.
2. Select multiple adjacent columns with a **left click** on the first column heading, then hold  and **left click** on the last column heading.
3. Select multiple non-adjacent columns by holding  then **left click** on any column headings you want to select.

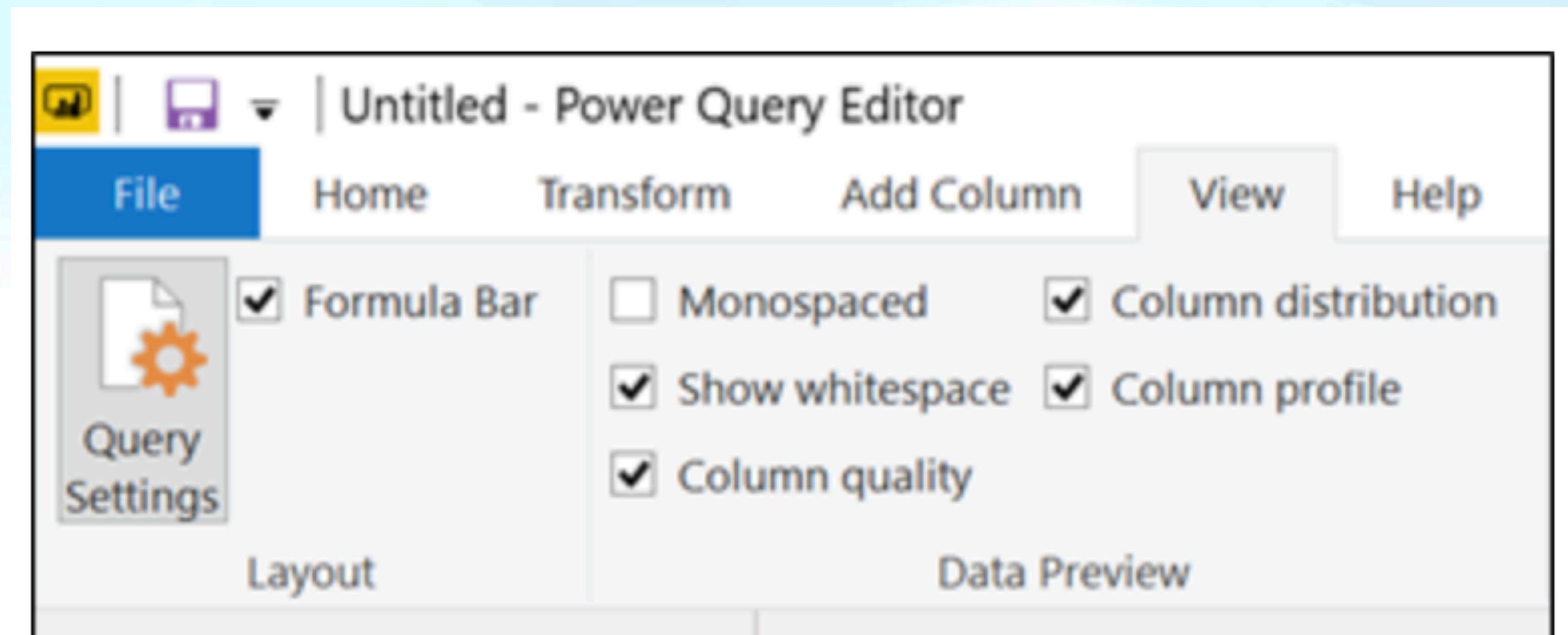
Data Preview

- Each column has a data type icon on the left hand of the column heading. You can **left click** on it to change the data type of the column.
- You can choose from decimal numbers, currency, whole numbers, percentages, date and time, dates, times, timezone, duration, text, Boolean, and binary.

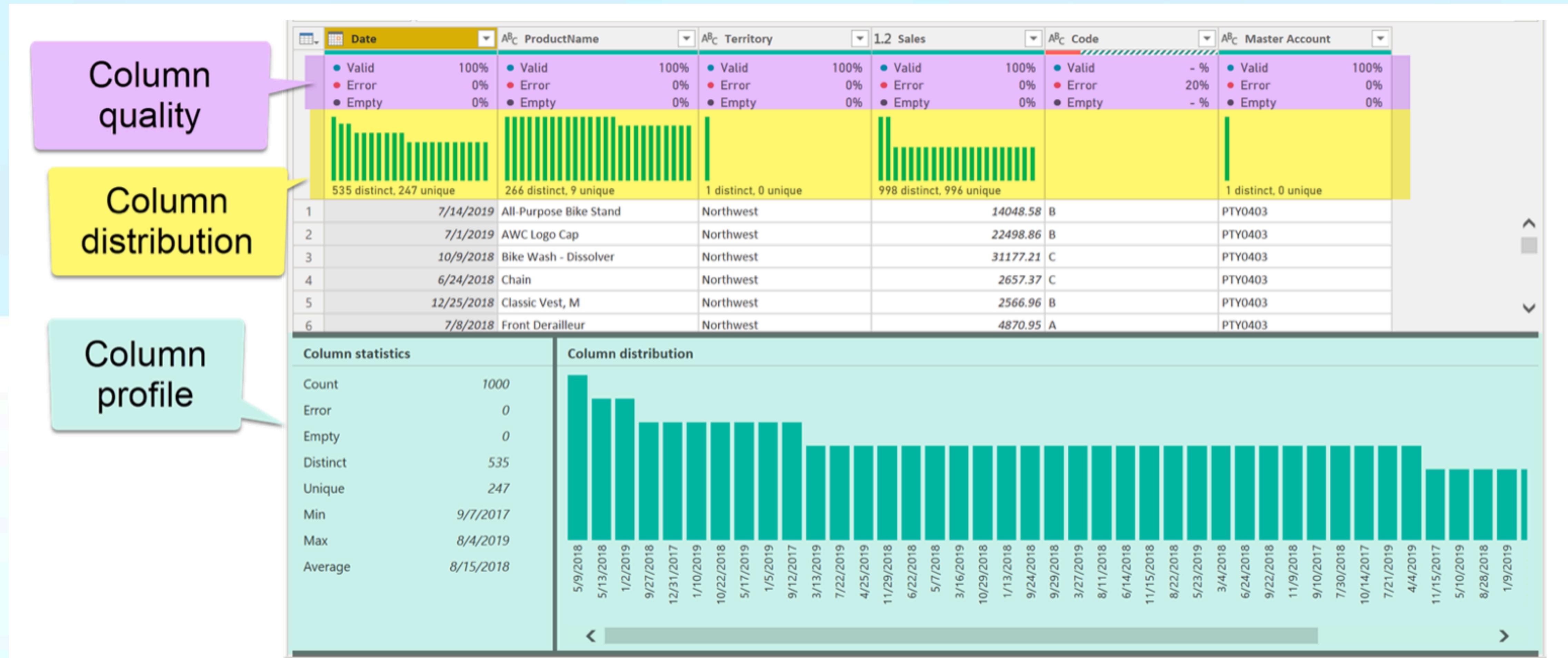


Data Quality Checks in Power Query

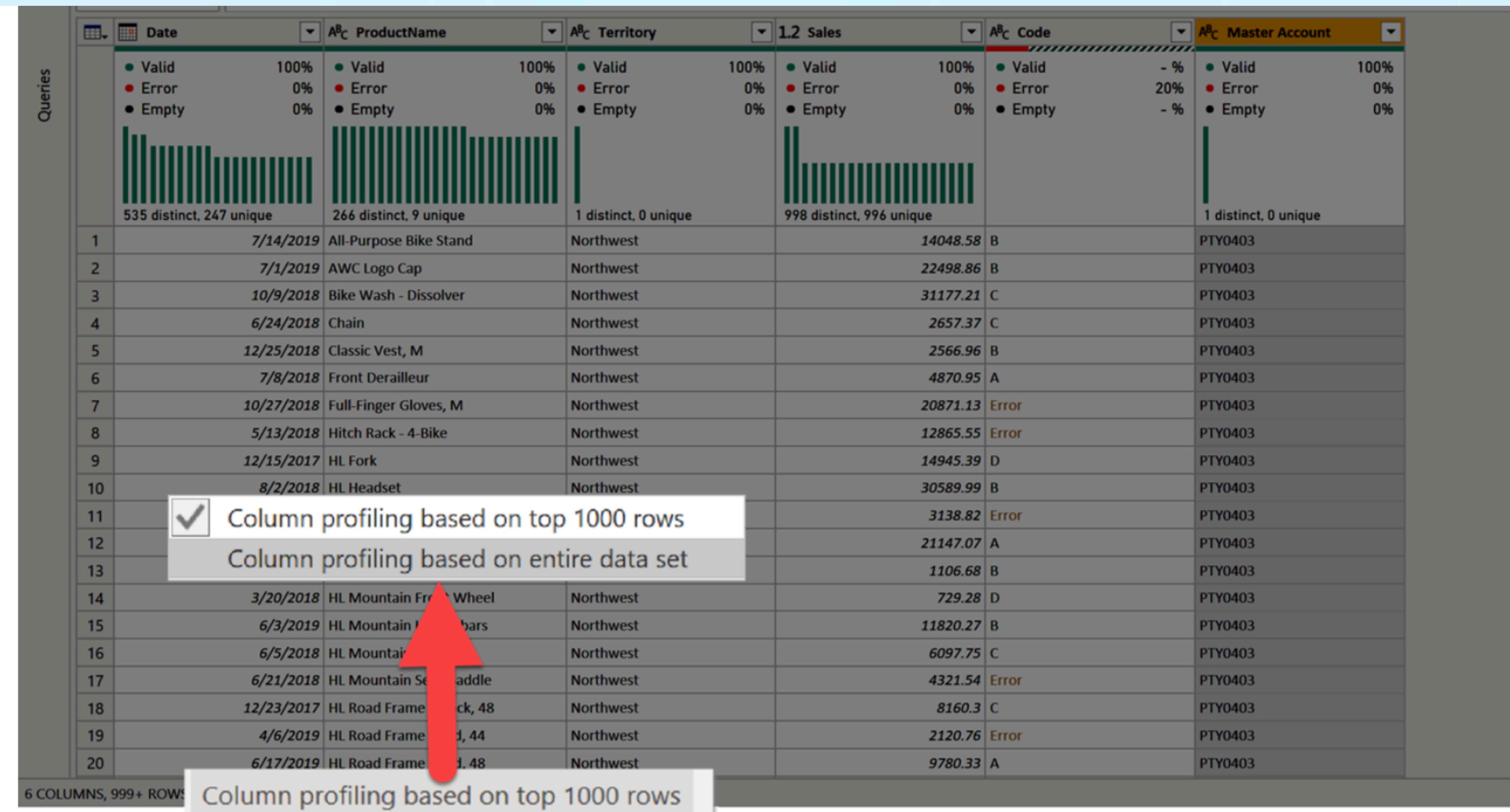
- To activate: To enable these features, you need to go to the View tab → Data Preview Group → Check the following:
 - Column quality
 - Column distribution
 - Column profile



Data Quality Checks in Power Query



Data Quality Checks in Power Query



Text Processing using Power Query

Case Study

Assume you are the newly appointed Marketing Manager of a 4-star SPA hotel.

You want to analyze reviews about the hotel in sites like [booking.com](#) and [tripadvisor.com](#) to understand the major service issues. With the help of Power Query, have to extract, clean and shape data from those sites to receive a nice tablular format.

Preparing Data for Analysis

- Load data into Power Query
- Clean punctuation symbols like commas, semicolons, dashes, brackets, etc., as well as replace capital letters with lower cased letters. Why?
- The cleaning can easily be done with tools like Replace Values, Trim, and Lowercase.
- First, duplicate the Opinions column so you can have original text.
- Then, replace all punctuation symbols with a space.
- Later we will use a period a separator to split opinions into sentences. And since a sentence can finish with exclamation mark or a question mark it is useful to replace ‘!’ and ‘?’ with ‘.’
- Since this is a hotel, clients may use the symbol ‘*’ in their opinions (i.e. ‘nice 4* hotel’ or ‘definitely not a 4* hotel’). So don’t replace this symbol.

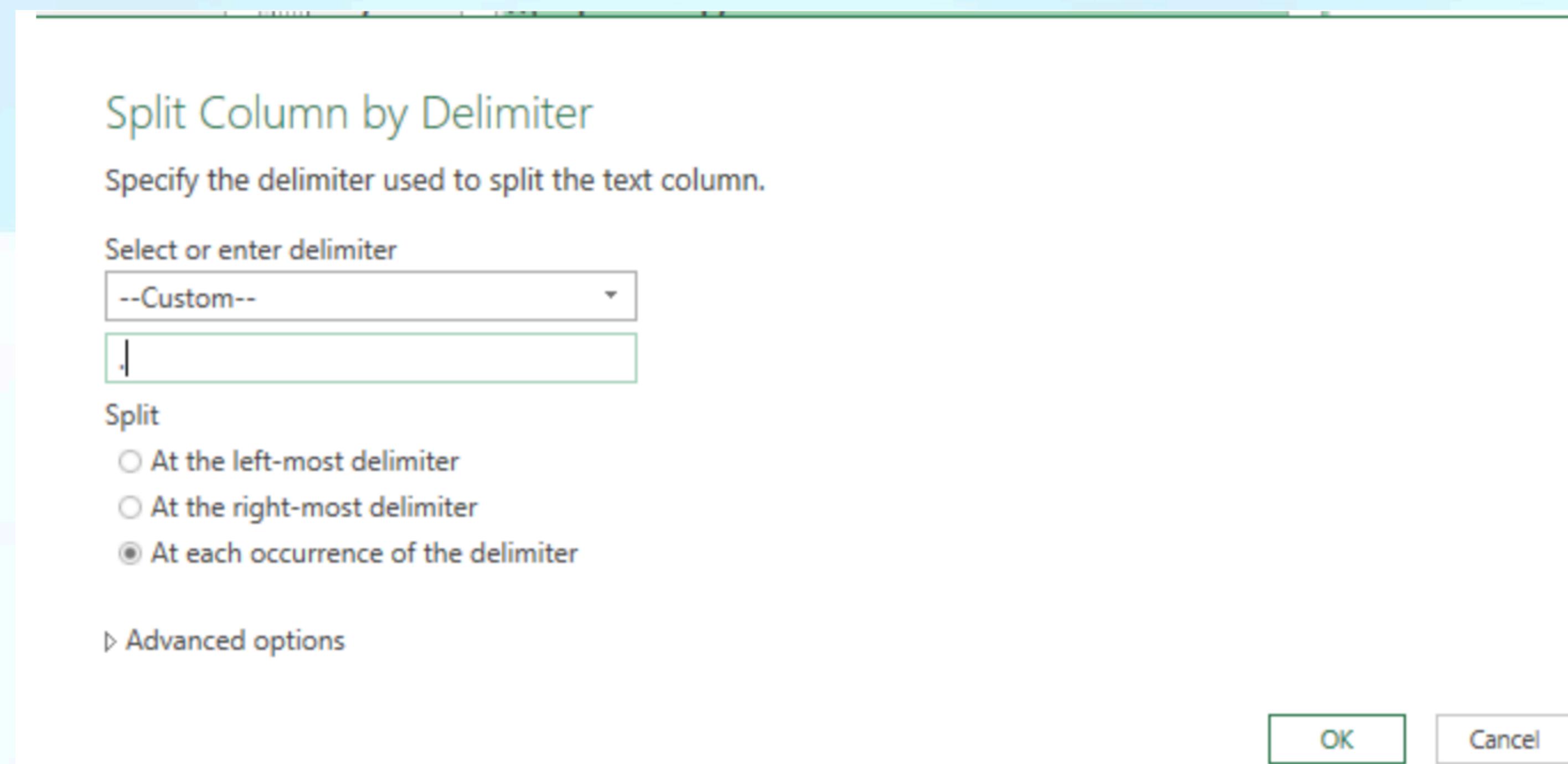
Preparing Data for Analysis

- Here are the steps applied so far...

| APPLIED STEPS | |
|-------------------|---|
| Source | |
| Changed Type | * |
| Duplicated Column | * |
| Lowercased Text | * |
| Replaced , | * |
| Replaced ? | * |
| Replaced ! | * |
| Replaced " | * |
| Replaced (| * |
| Replaced) | * |
| Replaced : | * |
| Replaced + | * |
| Replaced - | * |
| Replaced ' | * |
| Replaced / | * |
| Replaced \ | * |
| Trimmed Text | |
| Cleaned Text | * |

Preparing Data for Analysis

- Separating each comment =>



Preparing Data for Analysis

The screenshot shows a Microsoft Excel spreadsheet titled "Opinion" with 16 rows of text data. A context menu is open over the first two rows, listing various data manipulation options. The option "Unpivot Other Columns" is highlighted with a red box.

Opinion

| | |
|----|---|
| 1 | Super noisy. You cannot sleep, you cannot rest. |
| 2 | On Saturday afternoon has started a birthday party |
| 3 | We just came back a few days ago and we are gene |
| 4 | Anyway,we inform the staff and they tried to blam |
| 5 | The breakfast was good,nothing special. |
| 6 | If the management make some effort to renovate |
| 7 | A pleasant place but can't say the same about their |
| 8 | The hotel itself is pretty good, the breakfast was go |
| 9 | Nice big pool, but this time... |
| 10 | + The outdoor pool is really big and nice with hot ja |
| 11 | + The breakfast was very good. |
| 12 | + The bed was comfortable. |
| 13 | - The sink in the room was very slow and hotel staf |
| 14 | - The tv could not decrease the volume from the re |
| 15 | - The room was facing the external AC bodies and i |
| 16 | Fantastic pools, great room, beautiful grounds... |

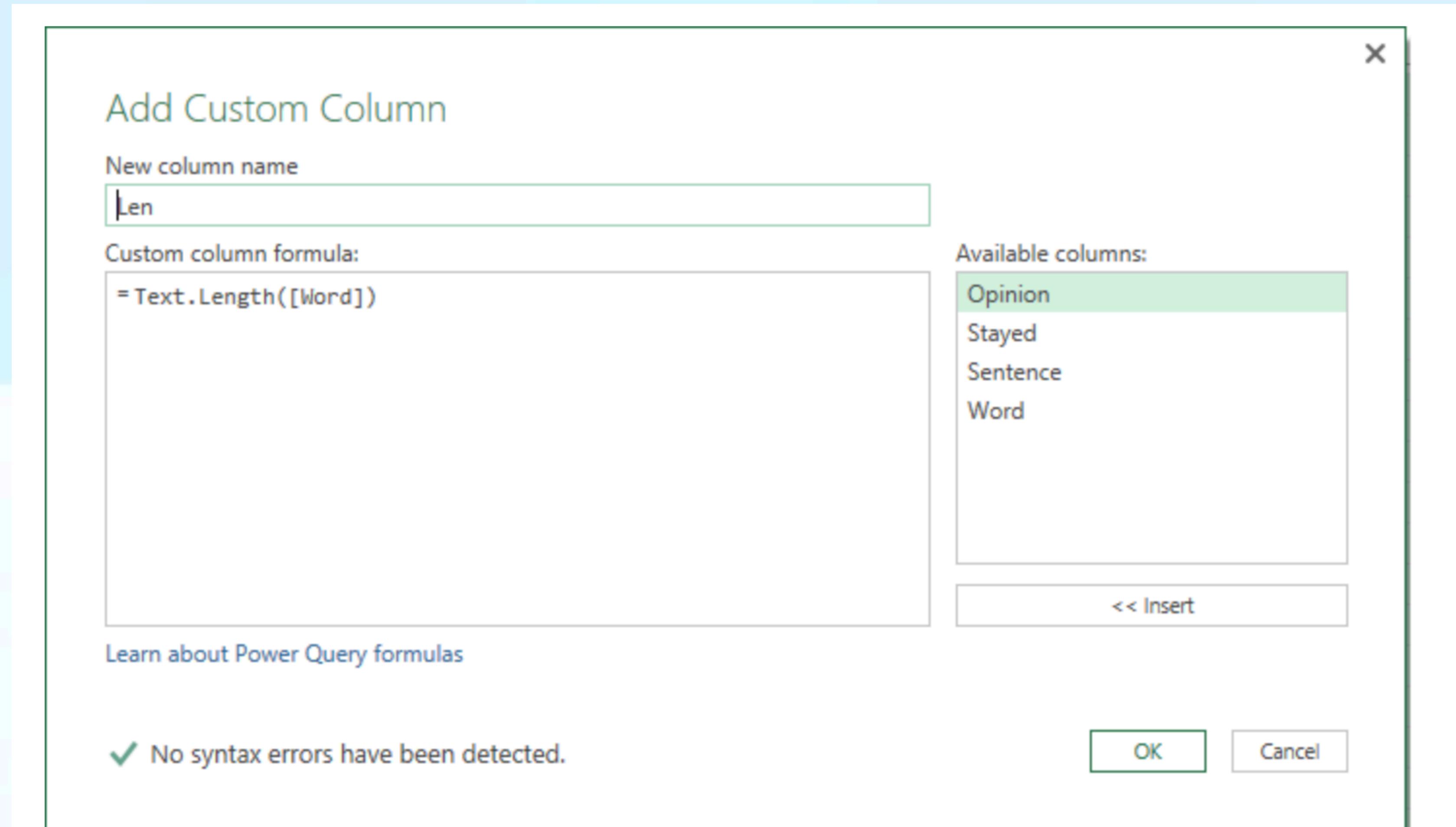
Copy
Remove Columns
Remove Other Columns
Remove Duplicates
Remove Errors
Replace Values...
Fill
Change Type
Merge Columns
Group By...
Unpivot Columns
Unpivot Other Columns
Move

01.06.2018

Preparing Data for Analysis

| A ^B _C Opinion | Stayed | A ^B _C Sentence |
|---|------------|---|
| 1 Super noisy. You cannot sleep, you cannot rest. | 01.08.2018 | super noisy |
| 2 Super noisy. You cannot sleep, you cannot rest. | 01.08.2018 | you cannot sleep you cannot rest |
| 3 On Saturday afternoon has started a birthday party if hotel's owner 12... | 01.08.2018 | on saturday afternoon has started a birthday party if hotel s owner 12 ... |
| 4 On Saturday afternoon has started a birthday party if hotel's owner 12... | 01.08.2018 | theyâ€™ve woke us up at the afternoon and the noise has continued u... |
| 5 On Saturday afternoon has started a birthday party if hotel's owner 12... | 01.08.2018 | what kind of spa hotel is that with all that noise |
| 6 On Saturday afternoon has started a birthday party if hotel's owner 12... | 01.08.2018 | iâ€™ve expected a relaxing weekend that has been spoiled by the hot... |
| 7 We just came back a few days ago and we are generally happy with ou... | 01.08.2018 | we just came back a few days ago and we are generally happy with ou... |
| 8 We just came back a few days ago and we are generally happy with ou... | 01.08.2018 | the pools both indoor and outdoor were great with mineral water |
| 9 We just came back a few days ago and we are generally happy with ou... | 01.08.2018 | green space around the hotel was great with activities for kids |
| 10 We just came back a few days ago and we are generally happy with ou... | 01.08.2018 | but our big disapointment was the room and especially the bathroom |
| 11 We just came back a few days ago and we are generally happy with ou... | 01.08.2018 | once we entered the room we noticed there are some problem becou... |
| 12 We just came back a few days ago and we are generally happy with ou... | 01.08.2018 | when we take shower the water was coming under the bathtub |
| 13 We just came back a few days ago and we are generally happy with ou... | 01.08.2018 | it is not very nice and safe when you have small kids and they may fall ... |
| 14 Anyway,we inform the staff and they tried to blame us.They send som... | 01.08.2018 | anyway we inform the staff and they tried to blame us |
| 15 Anyway,we inform the staff and they tried to blame us.They send som... | 01.08.2018 | they send someone to fix it but the problem in not minor |
| 16 Anyway,we inform the staff and they tried to blame us.They send som... | 01.08.2018 | so we need to stay there and to clean every time we take shower |
| 17 Anyway,we inform the staff and they tried to blame us.They send som... | 01.08.2018 | my husband was witness to another complain from other customer ab... |
| 18 Anyway,we inform the staff and they tried to blame us.They send som... | 01.08.2018 | so they sell the rooms absolutely aware about this leek and don t care ... |
| 19 The breakfast was good,nothing special. | 01.08.2018 | the breakfast was good nothing special |
| 20 If the management make some edford to renovate the rooms ,we will ... | 01.08.2018 | if the management make some edford to renovate the rooms we will ... |
| 21 | 01.08.2018 | |

Remove unnecessary short words('a', 'the', 'is', 'or', 'but' etc.)



Remove unnecessary short words('a', 'the', 'is', 'or', 'but' etc.)

| | A ^B _C Opinion | Stayed | A ^B _C Sentence | ABC 123 Word | 1 ² ₃ Len |
|----|---|------------|---|-------------------|---------------------------------|
| 1 | Super noisy. You cannot sleep, you cannot rest. | 01.08.2018 | super noisy | super | 5 |
| 2 | Super noisy. You cannot sleep, you cannot rest. | 01.08.2018 | super noisy | noisy | 5 |
| 3 | Super noisy. You cannot sleep, you cannot rest. | 01.08.2018 | you cannot sleep you cannot rest | cannot | 6 |
| 4 | Super noisy. You cannot sleep, you cannot rest. | 01.08.2018 | you cannot sleep you cannot rest | sleep | 5 |
| 5 | Super noisy. You cannot sleep, you cannot rest. | 01.08.2018 | you cannot sleep you cannot rest | cannot | 6 |
| 6 | Super noisy. You cannot sleep, you cannot rest. | 01.08.2018 | you cannot sleep you cannot rest | rest | 4 |
| 7 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | saturday | 8 |
| 8 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | afternoon | 9 |
| 9 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | started | 7 |
| 10 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | birthday | 8 |
| 11 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | party | 5 |
| 12 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | hotel | 5 |
| 13 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | owner | 5 |
| 14 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | years | 5 |
| 15 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | granddaughter | 13 |
| 16 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | pool | 4 |
| 17 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | with | 4 |
| 18 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | live | 4 |
| 19 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | concert | 7 |
| 20 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | pavel&ventzivents | 17 |
| 21 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | fireworks | 9 |
| 22 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | singing | 7 |
| 23 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | on saturday afternoon has started a birthd... | screaming | 9 |
| 24 | On Saturday afternoon has started a birthday party if ... | 01.08.2018 | they've woke us up at the afternoon a... | they've | 9 |

Finding Lookup words

The screenshot shows the Power BI desktop interface with a query named "Opinions" selected. A context menu is open, and the "Duplicate" option is highlighted with a red box.

Group By
Specify the columns to group by.

Group by: Word

New column name: Count

Operation: Count Rows

Column: (empty)

OK **Cancel**

File **Home** **Transform** **Add Column** **Vi**
Close & Load **Refresh Preview** **Properties** **Advanced Editor**
Close **Query** **Choose Columns** **Ren**
Manage Colu **Manage Colu**

Queries [1] **X** **✓** **fx** **=**

Opinions

- Copy**
- Paste**
- Delete**
- Rename**
- Duplicate**
- Reference**
- Move To Group**
- Move Up**
- Move Down**
- Advanced Editor**
- Properties...**

13 On Saturday afternoons

Finding Lookup words

| 1 | Word | Count |
|----|------------|-------|
| 2 | hotel | 151 |
| 3 | very | 116 |
| 4 | room | 104 |
| 5 | with | 103 |
| 6 | there | 97 |
| 7 | good | 93 |
| 8 | pool | 80 |
| 9 | were | 67 |
| 10 | this | 63 |
| 11 | staff | 60 |
| 12 | that | 57 |
| 13 | nice | 56 |
| 14 | breakfast | 54 |
| 15 | have | 52 |
| 16 | restaurant | 48 |
| 17 | which | 46 |
| 18 | they | 45 |
| 19 | clean | 44 |
| 20 | pools | 43 |
| 21 | rooms | 43 |
| 22 | from | 42 |

Filtering Lookup words

| LookupWords | ▼ |
|-------------|---|
| staff | |
| breakfast | |
| room | |
| food | |
| pool | |
| clean | |

Filter Rows

Basic Advanced

Show rows where: Word

equals staff

And Or

OK Cancel

Text Preprocessing Techniques

- Expand Contractions
- Lower Case
- Remove Punctuations
- Remove words and digits containing digits
- Remove Stopwords
- Rephrase Text
- Stemming and Lemmatization
- Remove White spaces