



Prepare data in Power BI Desktop

aka.ms/PL300-2



Agenda

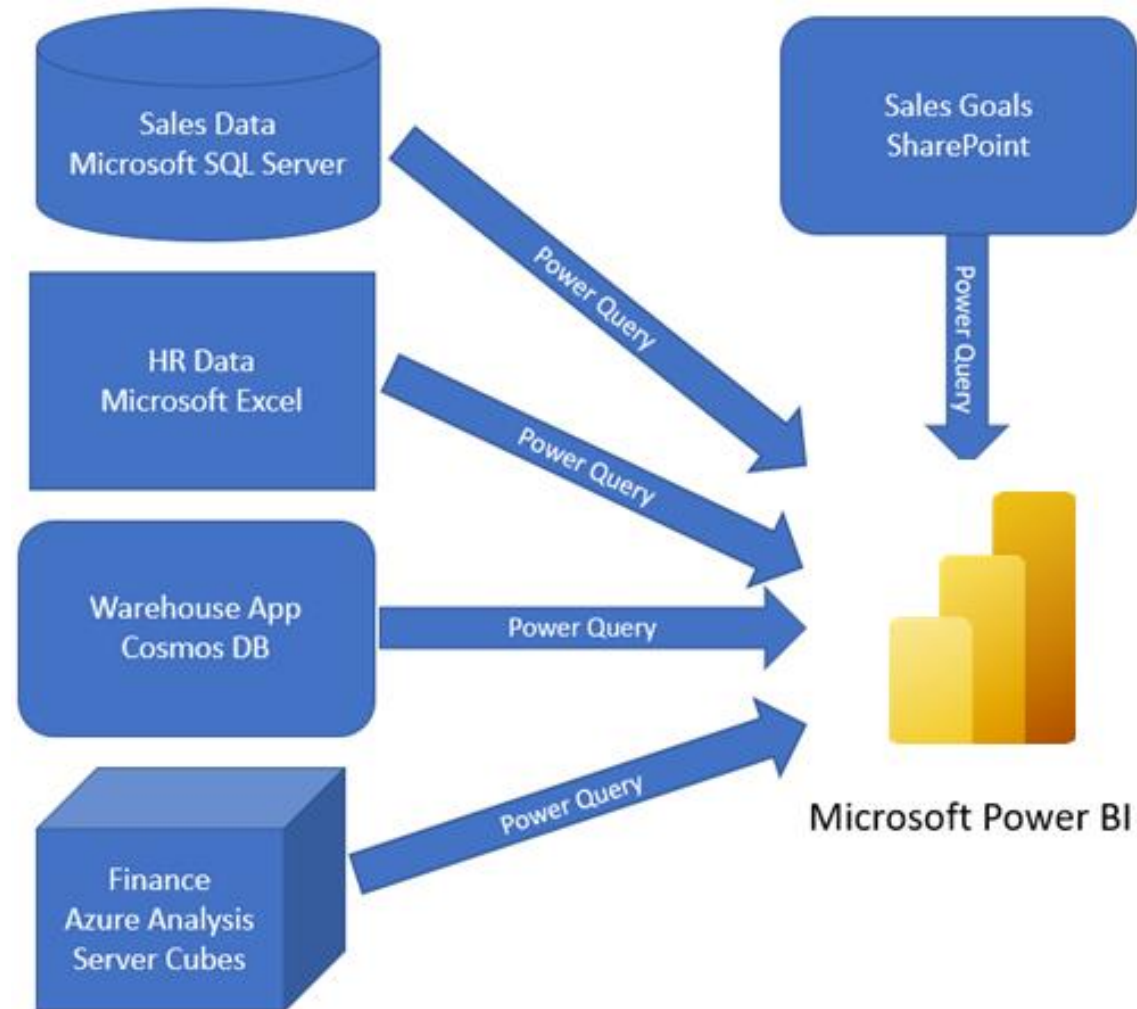


- Get data in Power BI
- Clean, transform, and load data in Power BI

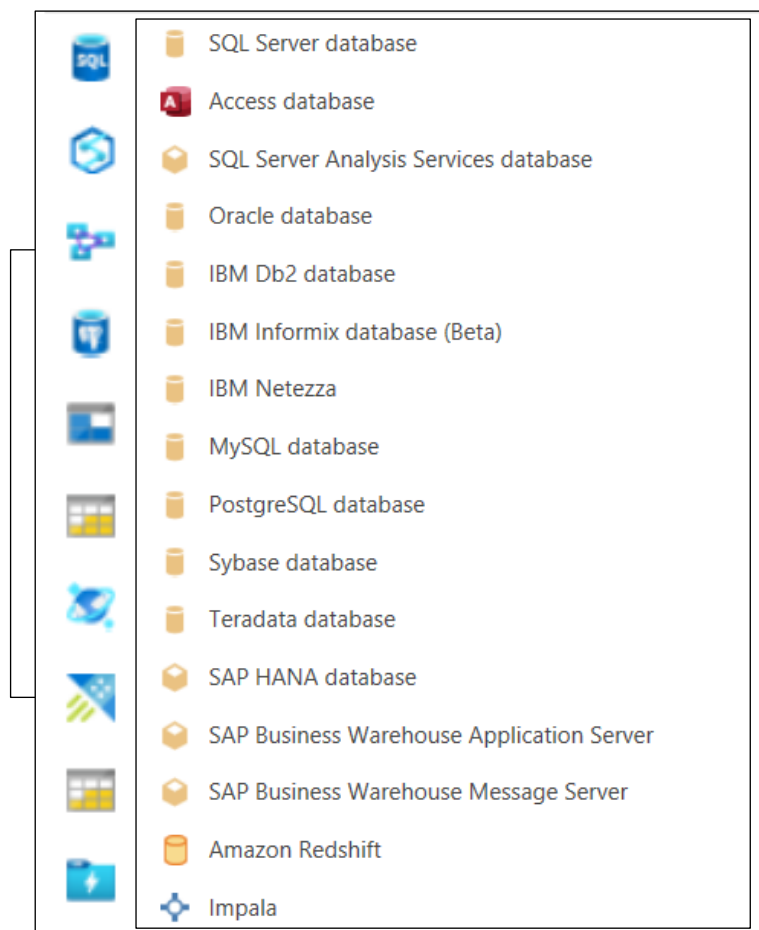
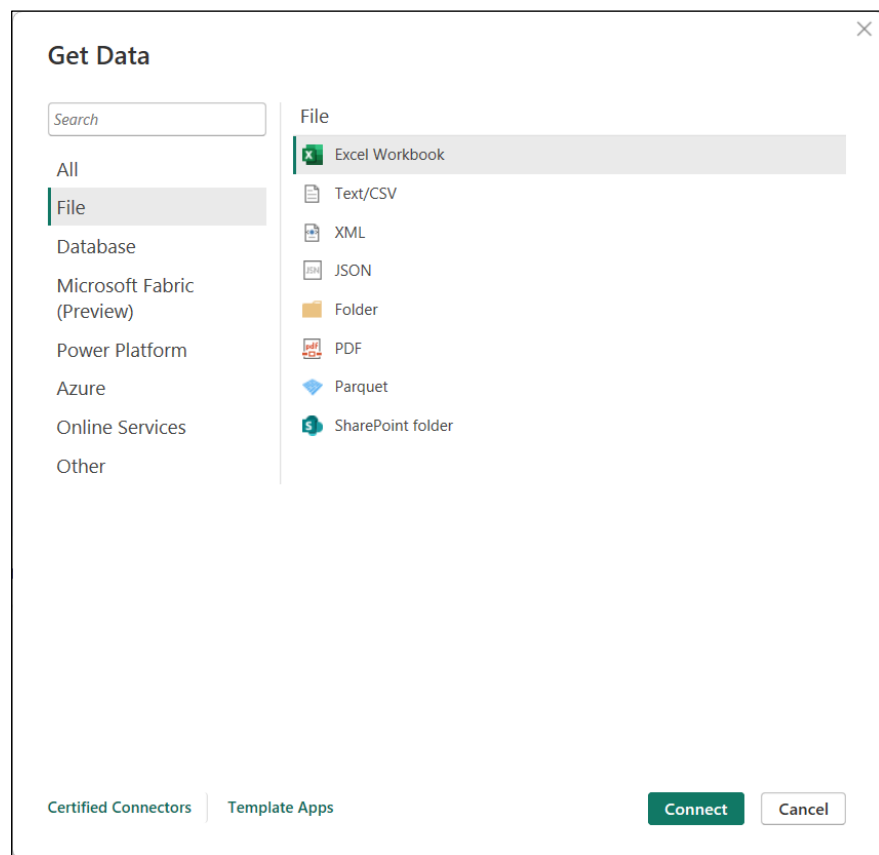
Get data in Power BI



Combine all data into a single semantic model



Get data from a wide variety of sources

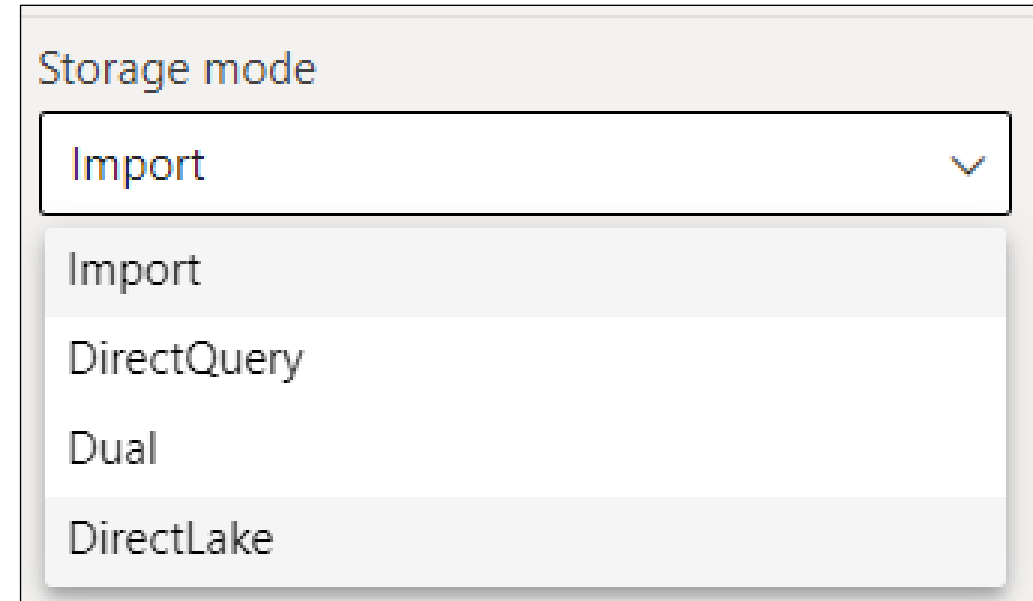


Select a storage mode

Storage mode affects

- Available transformations
- Report performance

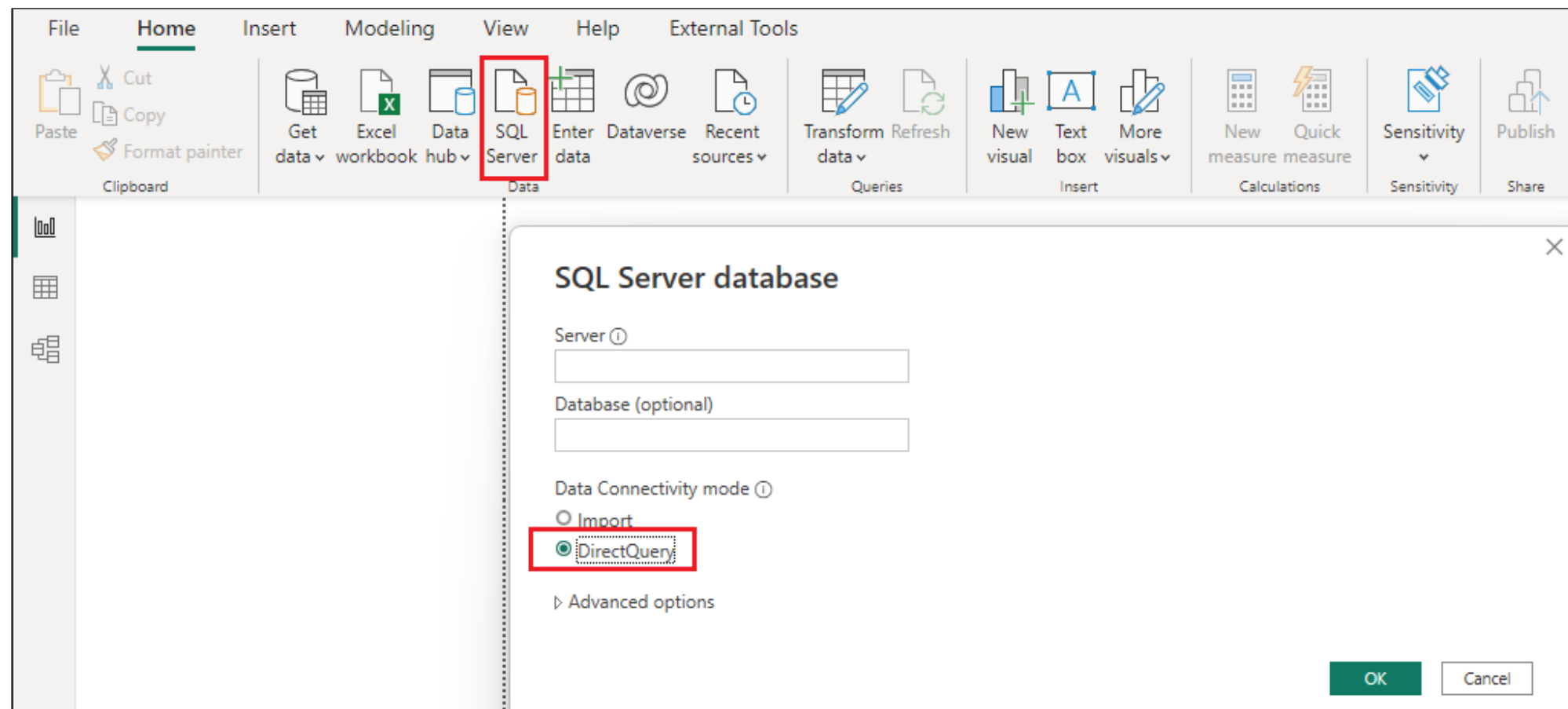
**Not all sources support all modes*



The image shows a user interface element titled "Storage mode". It features a dropdown menu with a light gray header bar containing the title. The dropdown is currently open, displaying four options: "Import", "DirectQuery", "Dual", and "DirectLake". The "Import" option is highlighted with a light gray background, and a small blue checkmark is visible in the top right corner of the dropdown box.

Storage mode
Import
DirectQuery
Dual
DirectLake

Introduction to DirectQuery



Implications of using DirectQuery

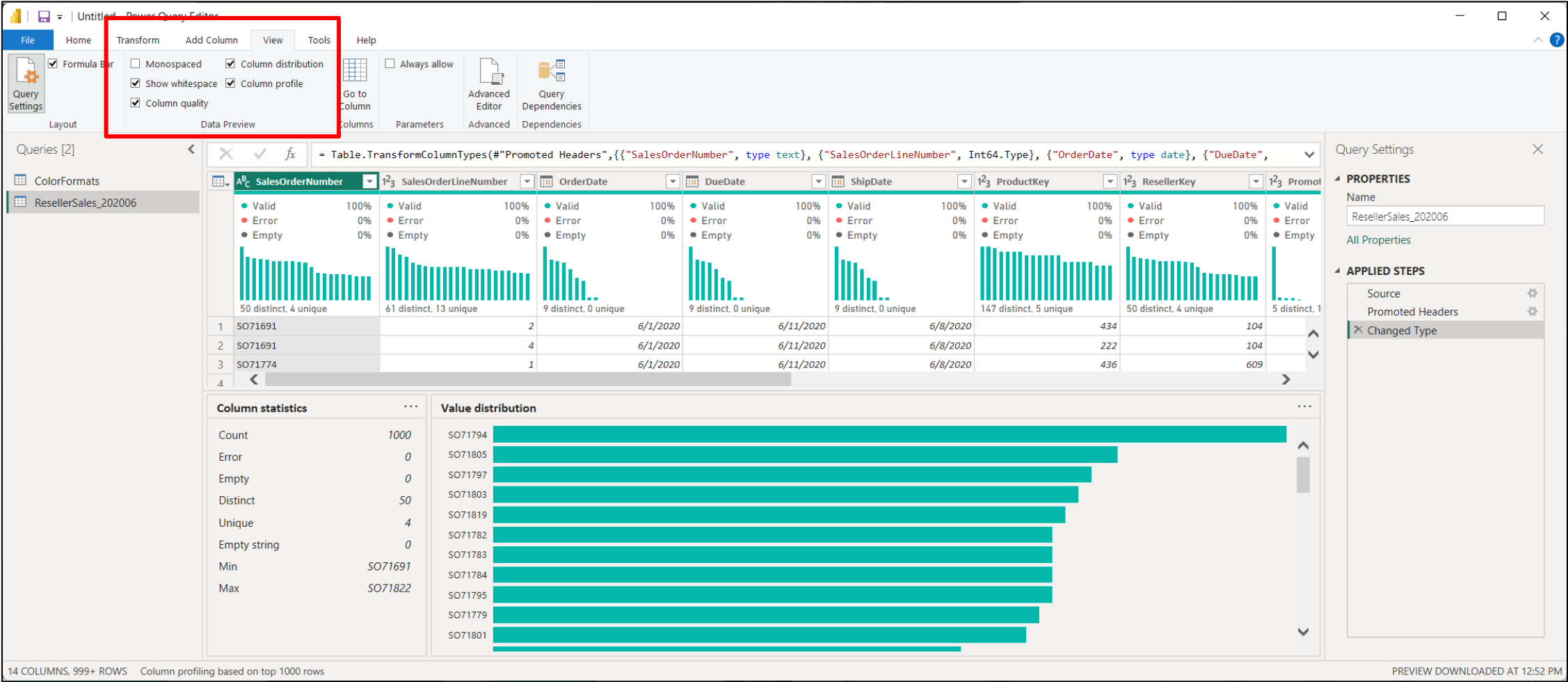
Benefits

- Frequently changing data
- Need near real-time
- Large data volumes
- Multi-dimensional data

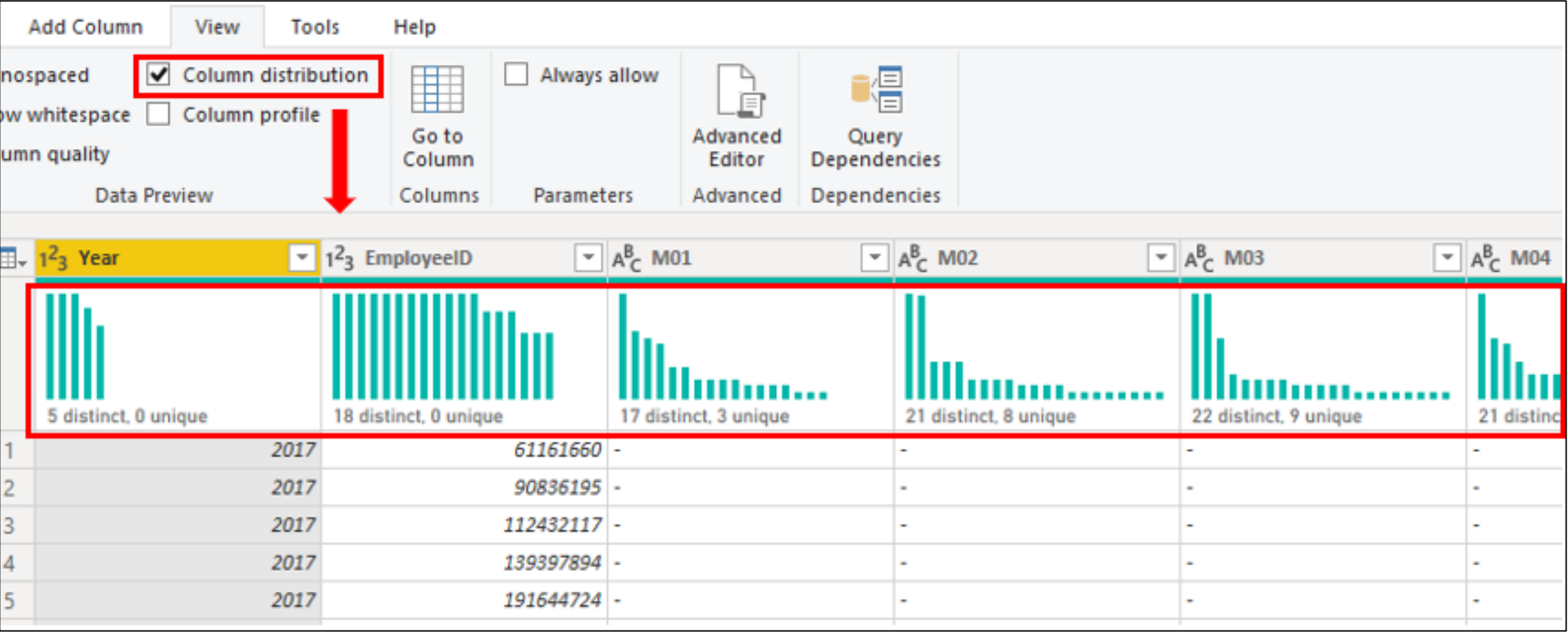
Limitations

- Dependent on data source performance
- Security between source and destination
- Limited modeling capabilities
- Limited transformation features

Data profiling options in Power Query Editor

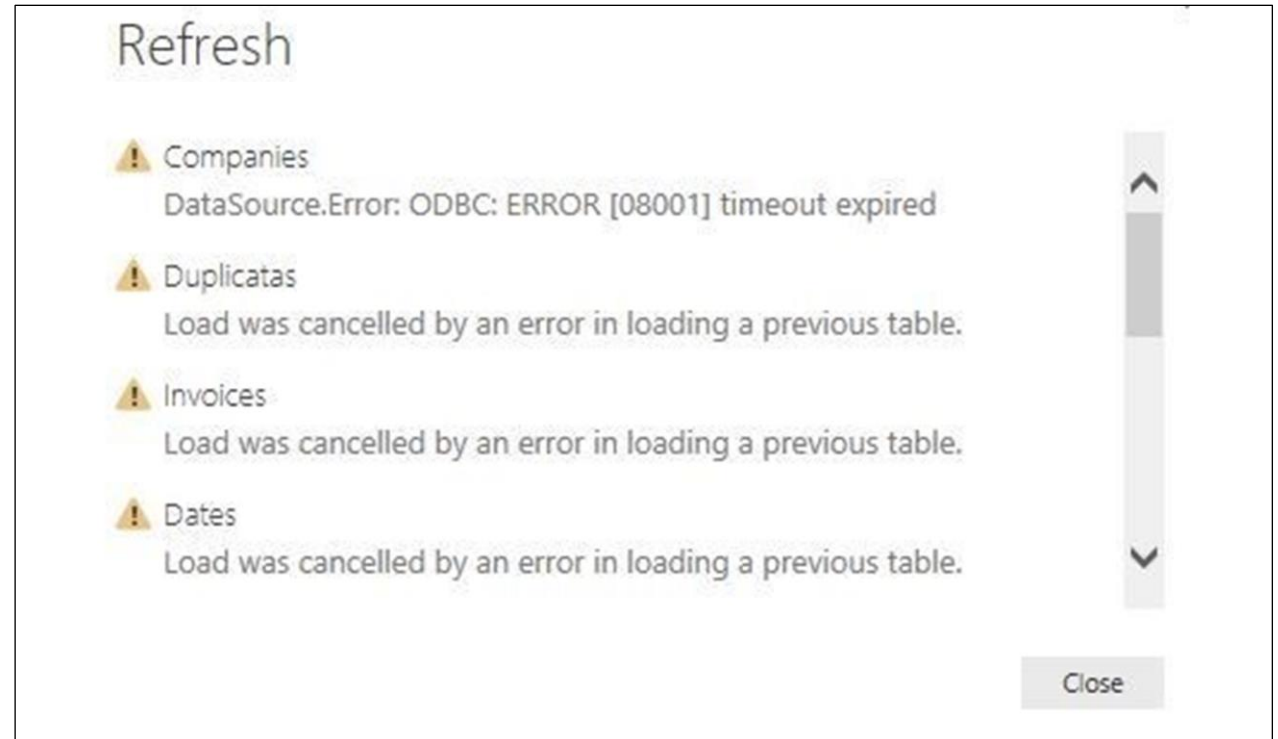


Determine cardinality of a column



Data import errors

- Possible data load errors:
 - Query Timeout
 - Couldn't find data formatted as a table
 - Could not find file
 - Data type errors



Lab: Prepare data in Power BI Desktop (30 minutes)



Prepare Data | GitHub Exercise

This lab is designed to introduce you to Power BI Desktop application and how to connect to data and how to use data preview techniques to understand the characteristics and quality of the source data.

- Open Power BI Desktop
- Connect to source data
- Preview source data
- Use data profile tools

Knowledge check: get data



When connecting to a SQL Server database to get data, what language should you use to extract data?

- ☐ DAX
- ☒ T-SQL
- ☐ MDX

How many rows are sampled in Power Query Editor by default?

- ☐ All
- ☒ 1000
- ☐ 100

What can you do to improve performance when you're getting data in Power BI?

- ☒ Remove unnecessary columns and rows
- ☐ Export database files to CSV to load
- ☐ Combine date and time columns into a single column

Clean, transform, and load data



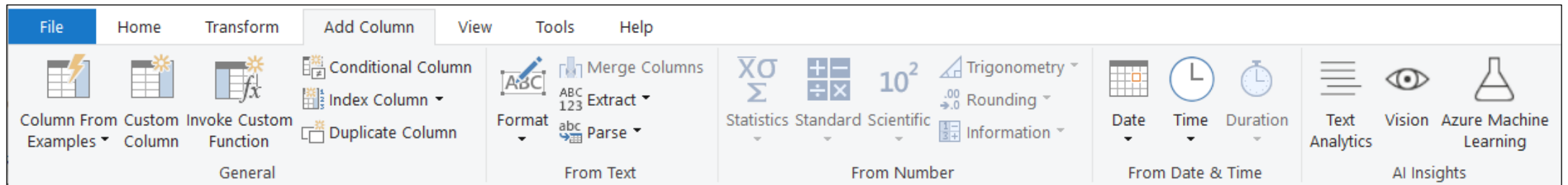
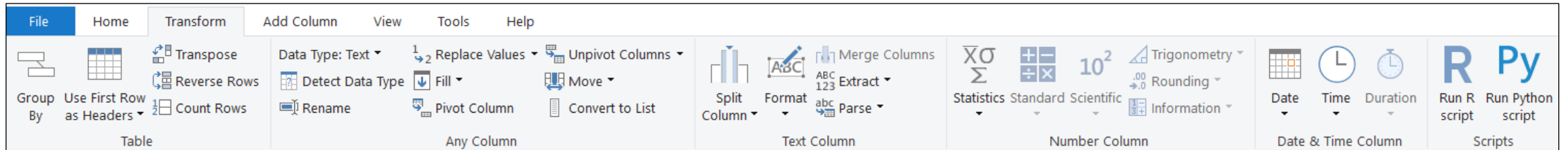
Transform data with Power Query Editor

The screenshot displays the Power Query Editor window titled "Untitled - Power Query Editor". The ribbon at the top includes tabs for File, Home, Transform, Add Column, View, Tools, and Help. The Transform tab is active, showing options like Merge Queries, Append Queries, Combine Files, Text Analytics, Vision, and Azure Machine Learning. The main area shows a data table with the following columns: SalesOrderNumber, SalesOrderLineNumber, OrderDate, DueDate, ShipDate, and ProductKey. The table contains 17 rows of data. The right-hand pane shows the Query Settings for "ResellerSales_202006", including the Name, All Properties, and Applied Steps (Source, Promoted Headers, Changed Type). The status bar at the bottom indicates "14 COLUMNS, 999+ ROWS" and "Column profiling based on top 1000 rows".

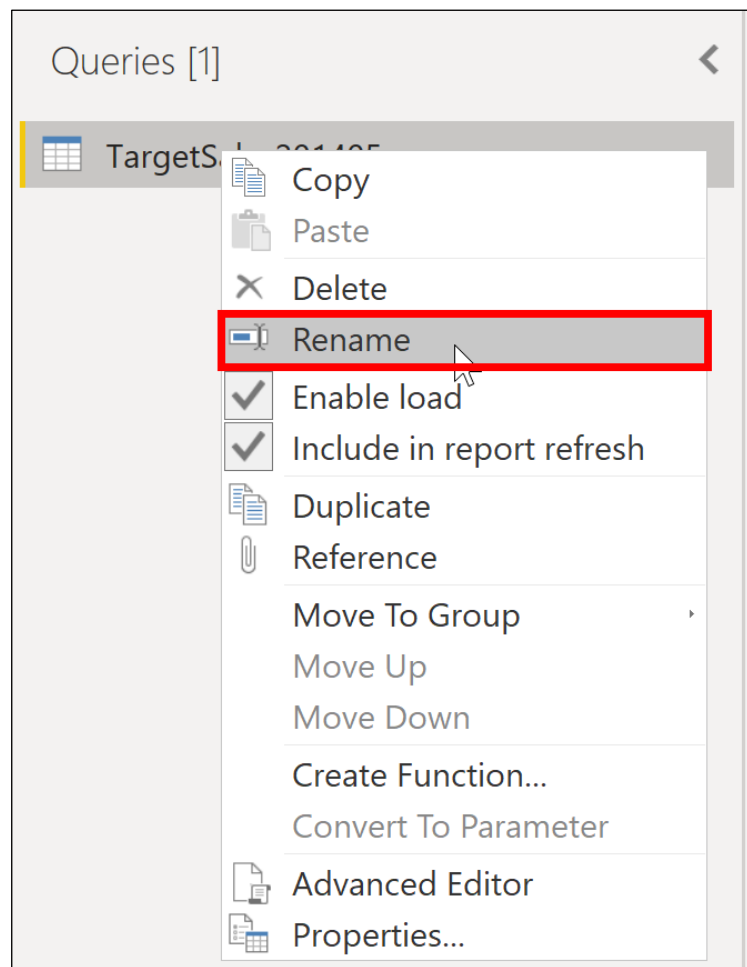
	SalesOrderNumber	SalesOrderLineNumber	OrderDate	DueDate	ShipDate	ProductKey
1	SO71691	2	6/1/2020	6/11/2020	6/8/2020	434
2	SO71691	4	6/1/2020	6/11/2020	6/8/2020	222
3	SO71774	1	6/1/2020	6/11/2020	6/8/2020	436
4	SO71774	2	6/1/2020	6/11/2020	6/8/2020	418
5	SO71775	1	6/1/2020	6/11/2020	6/8/2020	573
6	SO71775	2	6/1/2020	6/11/2020	6/8/2020	555
7	SO71775	3	6/1/2020	6/11/2020	6/8/2020	490
8	SO71776	1	6/2/2020	6/12/2020	6/9/2020	514
9	SO71777	1	6/2/2020	6/12/2020	6/9/2020	408
10	SO71777	2	6/2/2020	6/12/2020	6/9/2020	436
11	SO71778	1	6/2/2020	6/12/2020	6/9/2020	467
12	SO71778	2	6/2/2020	6/12/2020	6/9/2020	566
13	SO71778	3	6/2/2020	6/12/2020	6/9/2020	554
14	SO71778	4	6/2/2020	6/12/2020	6/9/2020	497
15	SO71779	1	6/2/2020	6/12/2020	6/9/2020	589
16	SO71779	2	6/2/2020	6/12/2020	6/9/2020	543

Common transformations

Transform columns, add new, split, extract, and more



Choose user-friendly values



A screenshot of the Power BI Transform ribbon. The 'Replace Values' button is highlighted with a red rectangle. Below the ribbon, a table is displayed with columns 'ID', 'Subcategory Name', and 'Attribute'. The 'Attribute' column header is highlighted with a red rectangle. The table contains 14 rows of data.

	ID	Subcategory Name	Attribute
1	1	Mountain Bikes	January
2	1	Mountain Bikes	February
3	1	Mountain Bikes	March
4	1	Mountain Bikes	April
5	1	Mountain Bikes	May
6	1	Mountain Bikes	June
7	1	Mountain Bikes	July
8	1	Mountain Bikes	August
9	1	Mountain Bikes	September
10	1	Mountain Bikes	October
11	1	Mountain Bikes	November
12	1	Mountain Bikes	Dezember
13	2	Road Bikes	January
14	2	Road Bikes	February

Shaping table structure

Advanced Editor

Manage ▾

Query

Choose Columns ▾ Remove Columns ▾

Keep Rows ▾ Remove Rows ▾

Reduce Rows ▾ Sort ▾

Remove Columns

Remove Other Columns

	A ^B _C Column13	A ^B _C Column14
	November	December
	880000	890000
	9500	10000
	511000	512000

Properties

Advanced Editor

Manage ▾

Query

Choose Columns ▾ Remove Columns ▾

Keep Rows ▾ Remove Rows ▾

Remove Top Rows

Remove Bottom Rows

Remove Alternate Rows

Remove Duplicates

Remove Blank Rows

Remove Errors

	A ^B _C Column13	A ^B _C Column14
	November	December
	880000	890000
	9500	10000
	511000	512000

Evaluate and change column data types

ABC SalesOrderNumber	ABC OrderDate	ABC TotalCost
SO43659	7/1/2017	2024.99
SO43659	7/1/2017	6074.97
SO43659	7/1/2017	2024.99
SO43659	7/1/2017	2039.99
SO43659	7/1/2017	2039.99
SO43659	7/1/2017	4079.98
SO43659	7/1/2017	2039.99
SO43659	7/1/2017	86.52
SO43659	7/1/2017	28.84
SO43659	7/1/2017	34.2

Couldn't load the data for this visual

MdxScript(Model) (19, 40) Calculation error in measure
'Sales'[Quantity of Orders YTD]: A column specified in the call to
function 'TOTALYTD' is not of type DATE. This is not supported.

[Copy details](#)

Send a Frown

Close

Combine multiple queries into one

Append

Append

Concatenate rows from three or more tables into a single table.

☐ Two tables ☒ Three or more tables

Available tables

Production Suppliers

Sales Customers

HR Employees

Tables to append

Production Suppliers

Sales Customers

HR Employees

Add >>

OK

Merge

Merge

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

Sales Orders

orderid	custid	empid	orderdate	requireddate	shippeddate	shipperid	freight	shipname
10248	85	5	7/4/2014	8/1/2014	7/16/2014	3	32.38	Ship to 85-B
10249	79	6	7/5/2014	8/16/2014	7/10/2014	1	11.61	Ship to 79-C
10250	34	4	7/8/2014	8/5/2014	7/12/2014	2	65.83	Destination SCC
10251	84	3	7/8/2014	8/5/2014	7/15/2014	1	41.34	Ship to 84-A

Sales OrderDetails

orderid	productid	unitprice	qty	discount
10248	11	14.00	12	0
10248	42	9.80	10	0
10248	72	34.80	5	0
10249	14	18.60	9	0
10249	51	42.40	40	0

Join Kind

Left Outer (all from first, matching from second)

☐ Use fuzzy matching to perform the merge

Fuzzy matching options

✓ The selection matches 830 of 830 rows from the first table.

OK Cancel

Unpivot or pivot columns

Add or remove table structure to meet your aggregation needs.

	A ^B C Category Name	A ^B C Subcategory Name
1	Bikes	Mountain Bikes
2	Bikes	Road Bikes
3	Bikes	Touring Bikes
4	Clothing	Bib-Shorts
5	Clothing	Caps
6	Clothing	Gloves
7	Clothing	Jerseys
8	Clothing	Shorts
9	Clothing	Socks
10	Clothing	Tights
11	Clothing	Vests
12	Accessories	Bike Racks
13	Accessories	Bike Stands
14	Accessories	Bottles and Cages

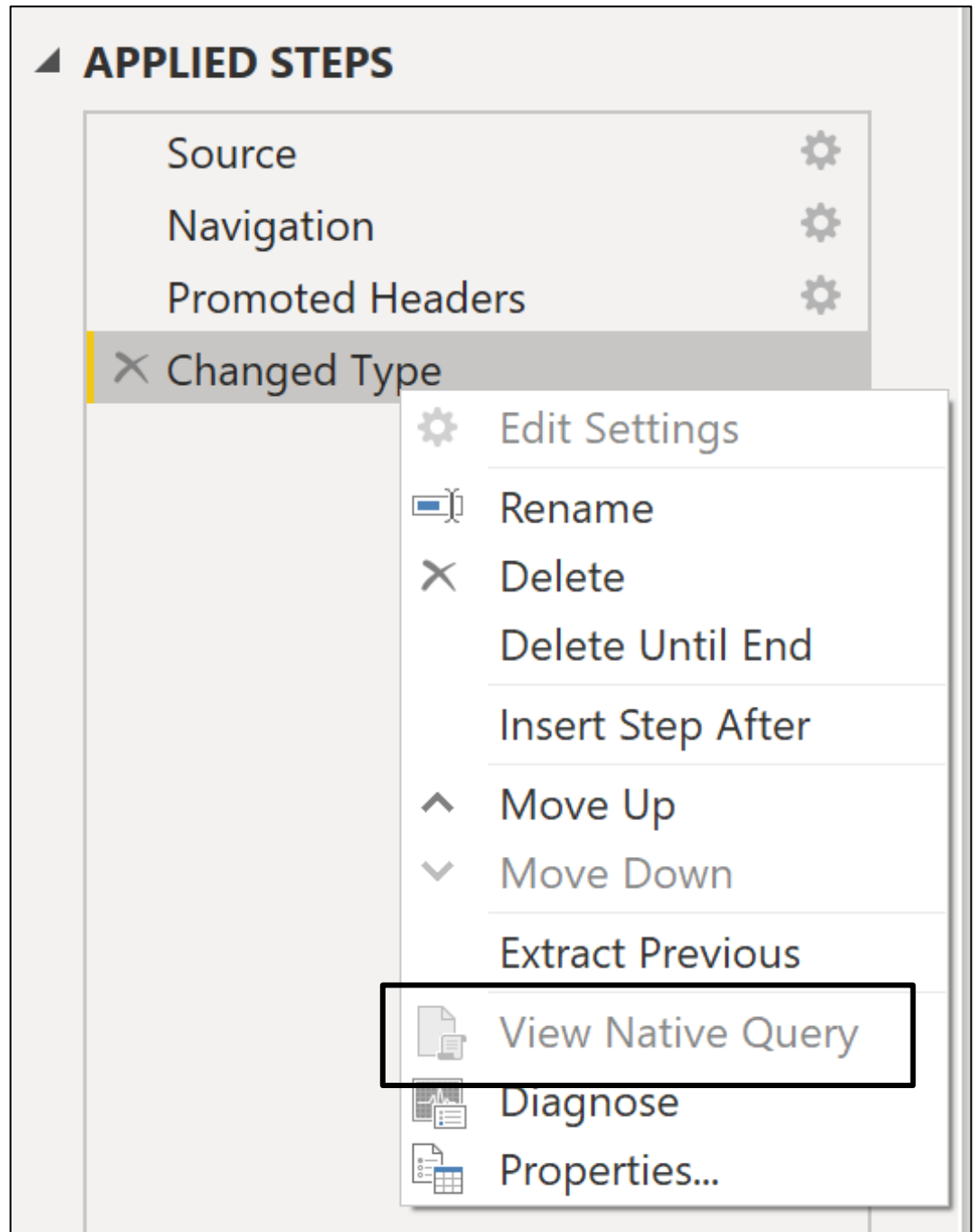
	1.2 Bikes	1.2 Components	1.2 Clothing	1.2 Accessories
1	3	14	8	12

Query folding

Pushes data transformations to the source for better performance and efficiency.

Supported sources

- Relational databases
- OData feeds
- Active Directory



Dynamic reports with parameters

Manage Parameters

New

A^BC SalesPerson

Name

SalesPerson

Description

Required

Type

Text

Suggested Values

Any value

Current Value

279

	A ^B C SalesOrderNumber	1 ² 3 SalesOrderID	1 ² 3 SalesPersonID
1	SO43659	43659	279
2	SO43660	43660	279
3	SO43681	43681	279
4	SO43684	43684	279
5	SO43685	43685	279
6	SO43694	43694	279
7	SO43695	43695	279
8	SO43696	43696	279

Performance recommendations

- Only keep necessary data
- Check data types
- Reduce cardinality
- Disable query load
- Use parameters

Lab: Load data in Power BI Desktop (45 minutes)



Load Data | GitHub Exercise

In this lab, you'll apply transformations to each of the queries created in the previous lab. You'll then apply the queries to load each as a table to the data model.

In this lab you learn how to:

- Apply various transformations
- Apply queries to load them to the data model

Knowledge check: clean and transform data



What is a risk of having null values in a numeric column?

- ☐ That function SUM of data are incorrect.
- ☐ That function MAX of data will be incorrect.
- ☒ That function AVERAGE of data will be incorrect.

If you have two queries that have different data but the same column headers, and you want to combine both tables into one query with all the combined rows, which operation should you perform?

- ☒ Append
- ☐ Merge
- ☐ Combine

Recap

In this section, we covered:

- Connecting to various data sources.
- Determining the right storage mode.
- Profiling data for accuracy and insights.
- Cleaning, transforming, and loading data.
- Using Query Folding for appropriate sources.
- Recommendations for more performant reports.



Thanks

Resources

[Prepare data for analysis](#)

[Clean, transform, and load data in Power BI](#)