Data Analysis with Power BI

Module 5 (Part 1)

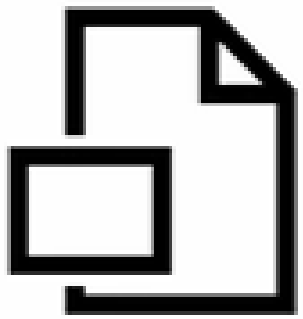
Structure

- Part 1: Extract, Transform and Load (ETL) – identify, explain and configure multiple data sources in Power BI; clean and transform data using Power Query.
- Part 2: Data Modeling – identify and create appropriate model relationships; configuring your table and column properties; data analysis expressions (DAX) to configure and optimize your models
- Part 3: Data Analysis and Visualization – add visualizations to reports; format visuals

Part 1

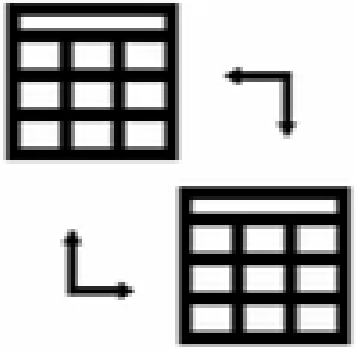
Extract, Transform and Load (ETL)

Data Sources



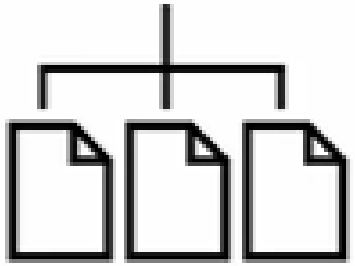
Flat files

CSV, TXT, XLSX



Relational data sources

SQL, mySQL, Oracle



NoSQL databases

MongoDB, Cassandra

PowerBI has the flexibility to connect to a wide range of data sources

Data Sources (cont.)

Data source	Examples
File	XML, JSON, PDF, Excel workbook (Limited to a maximum file size of 1GB)
Database	MySQL, SQL Server, Access, PostgreSQL, Google BigQuery
Power Platform	Microsoft Power Platform Power BI datasets, Datamarts, Dataverse, Dataflows
Azure	Azure SQL Database, Azure Synapse Analytics SQL, Azure Data Explorer
Online Services	GitHub, QuickBooks Online, Stripe, Dynamics 365, Salesforce
Other	R scripts, Python scripts and Active Directory

Dataset vs Data Source

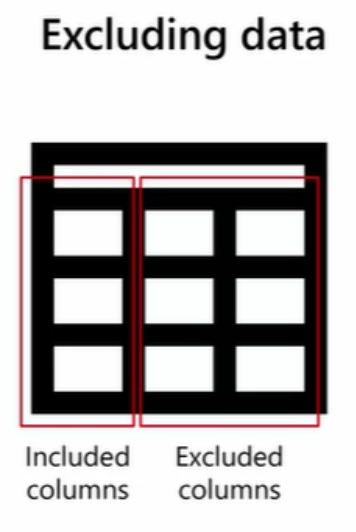
A **dataset** is what you get when you use the Get Data feature to bring in data from a file, template app, or live source. It includes information about the data source, credentials, and a portion of the data itself. When you create reports and dashboards, you review the data from the dataset.

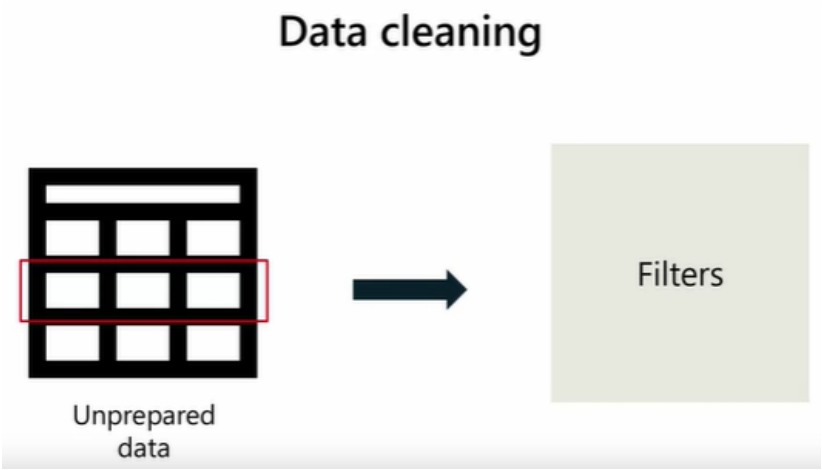
However, the data in the dataset comes from a **data source**. A data source can be an online service, a cloud database, or a local file or server. The data source is where the data comes from. An example of a data source could be a cloud database like Azure SQL, or a file on your computer.

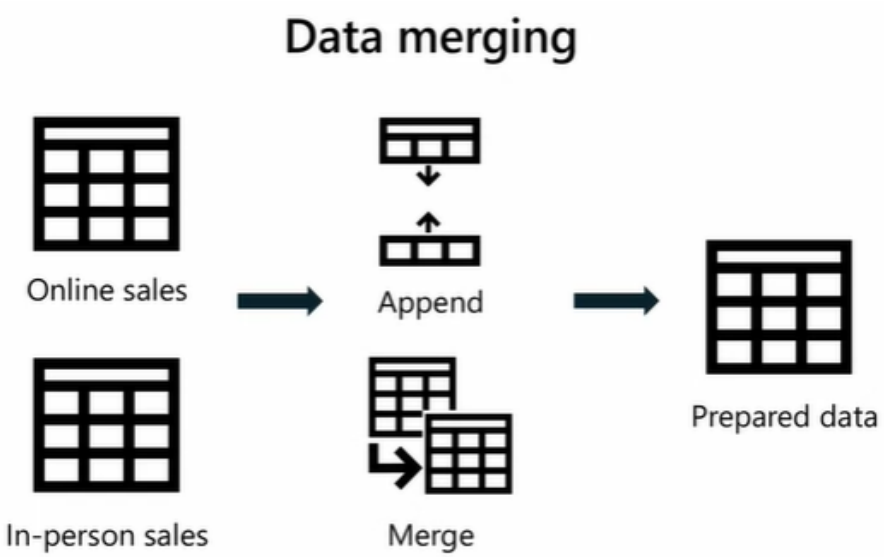
Exercise: Setting up an Excel data source

- Choose the required data source from the **Get Data** drop-down list.
- Navigate to the location where your Excel file (*SalesOrderDetail.xlsx*) is saved.
- Choose the worksheet(s) and table(s) you want to import in the Navigator window.
- Click Load. This will import your Excel data into Power BI.

Data Transformation

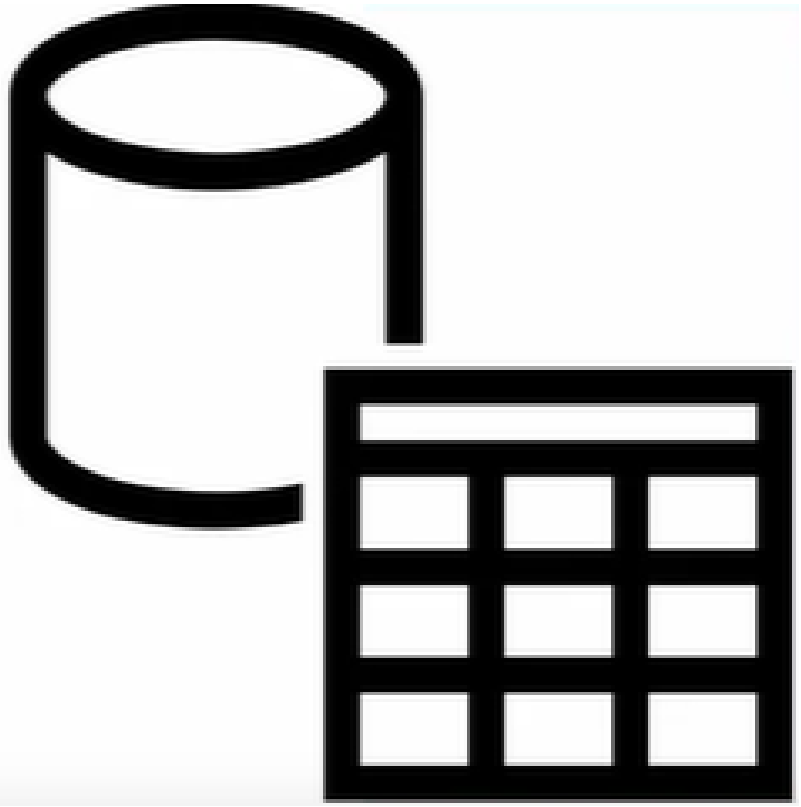




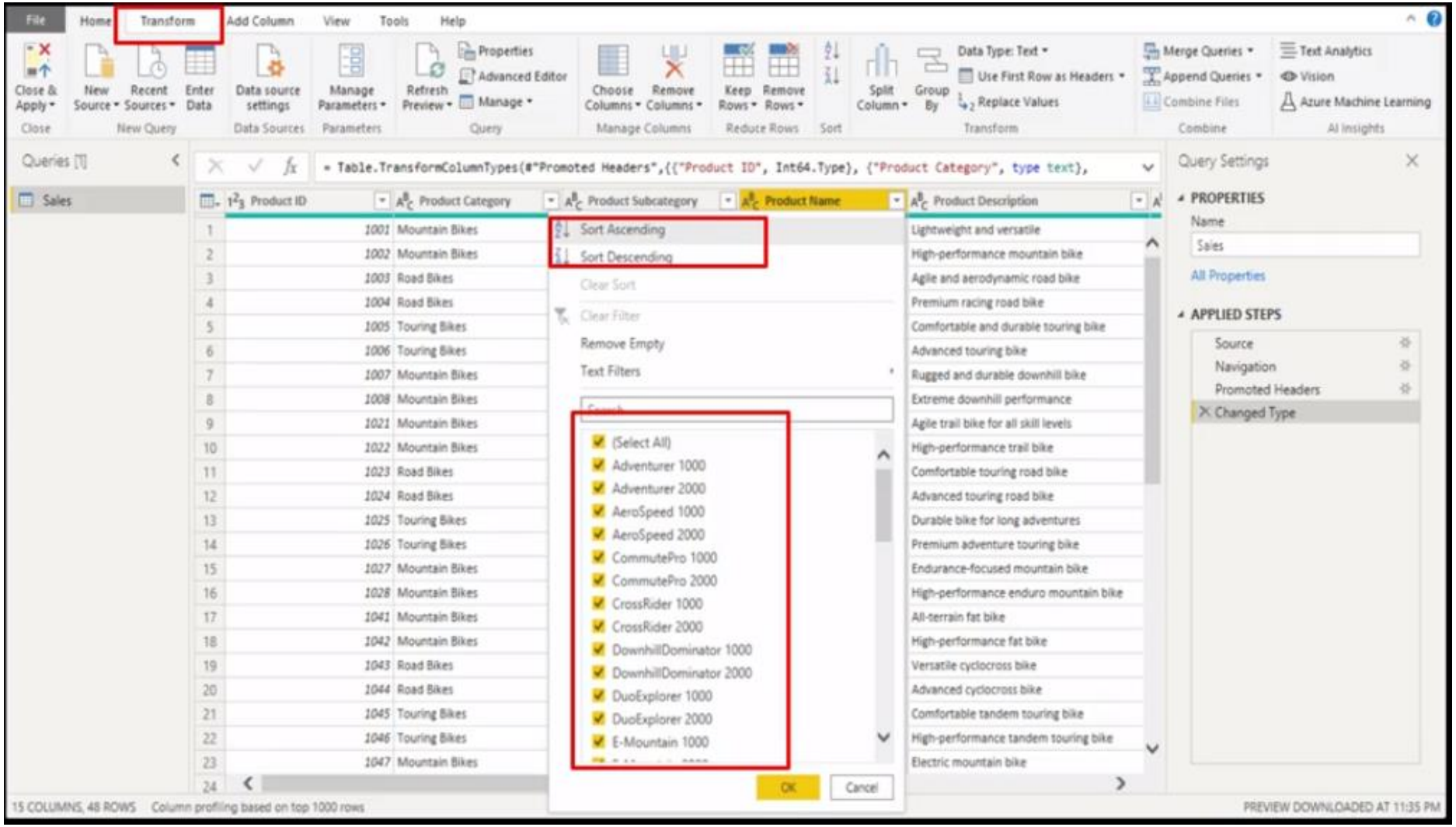


Power Query

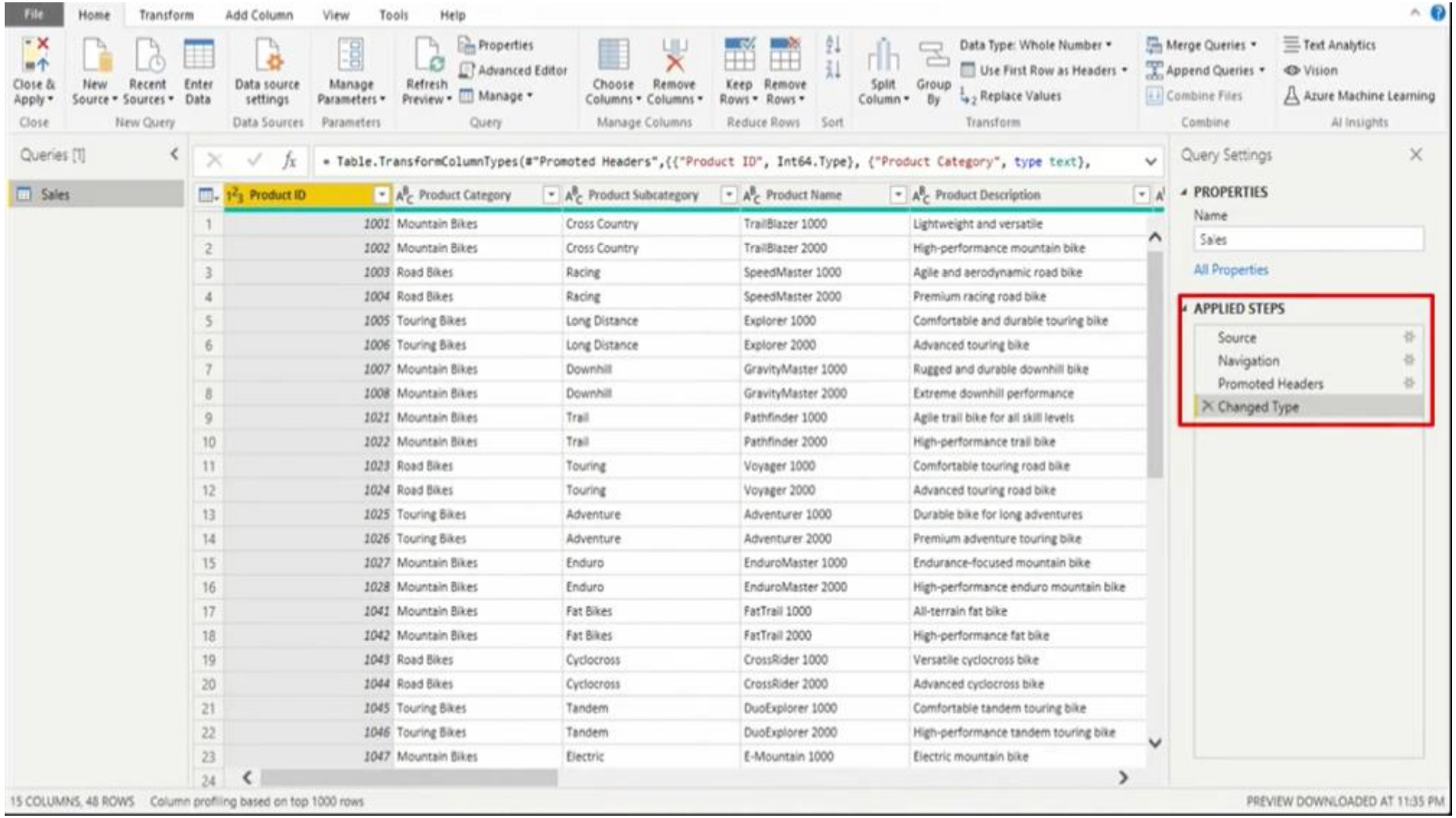
Data transformation
and preparation tool.



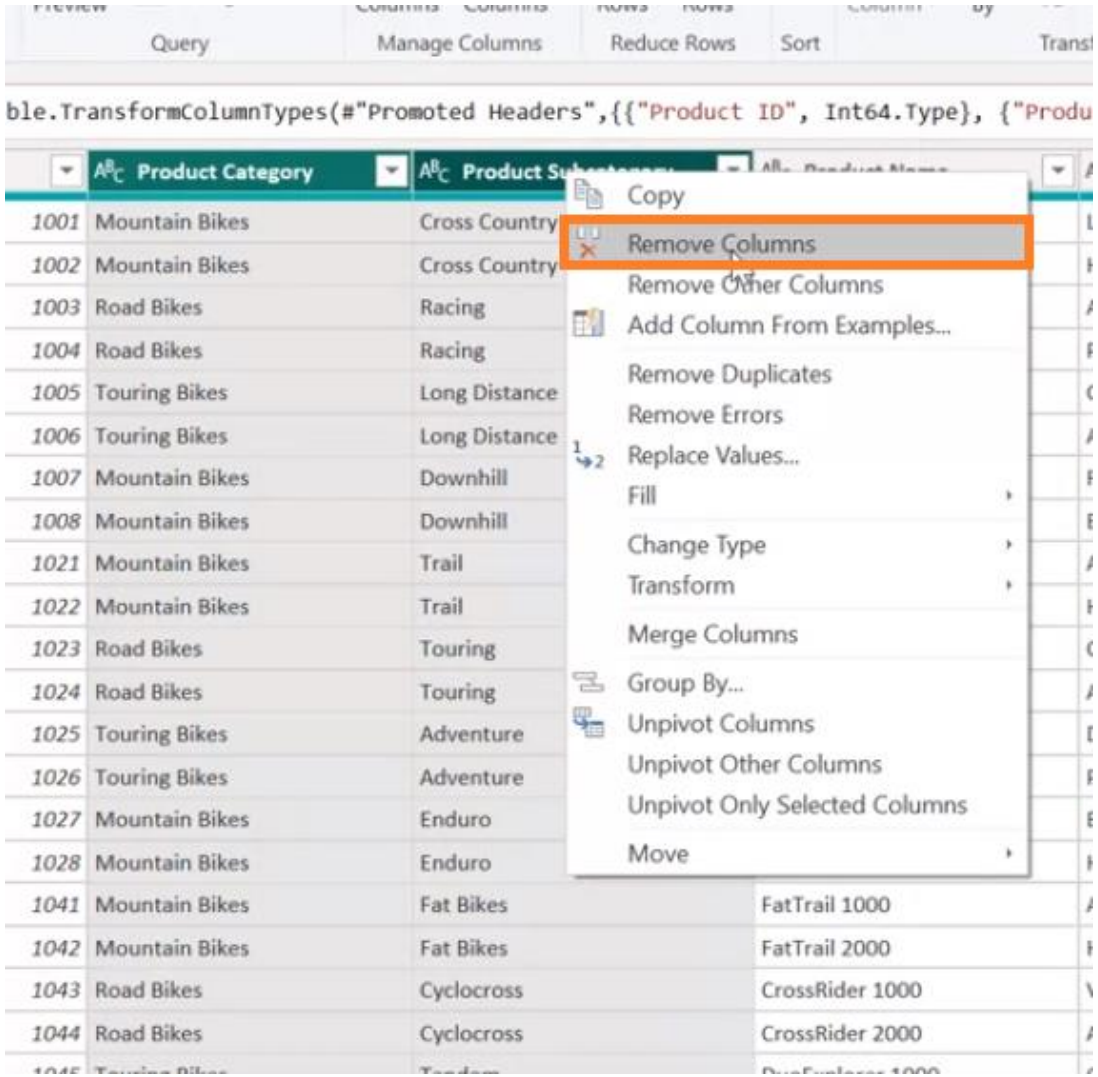
Power Query – Data Extraction and Transformation



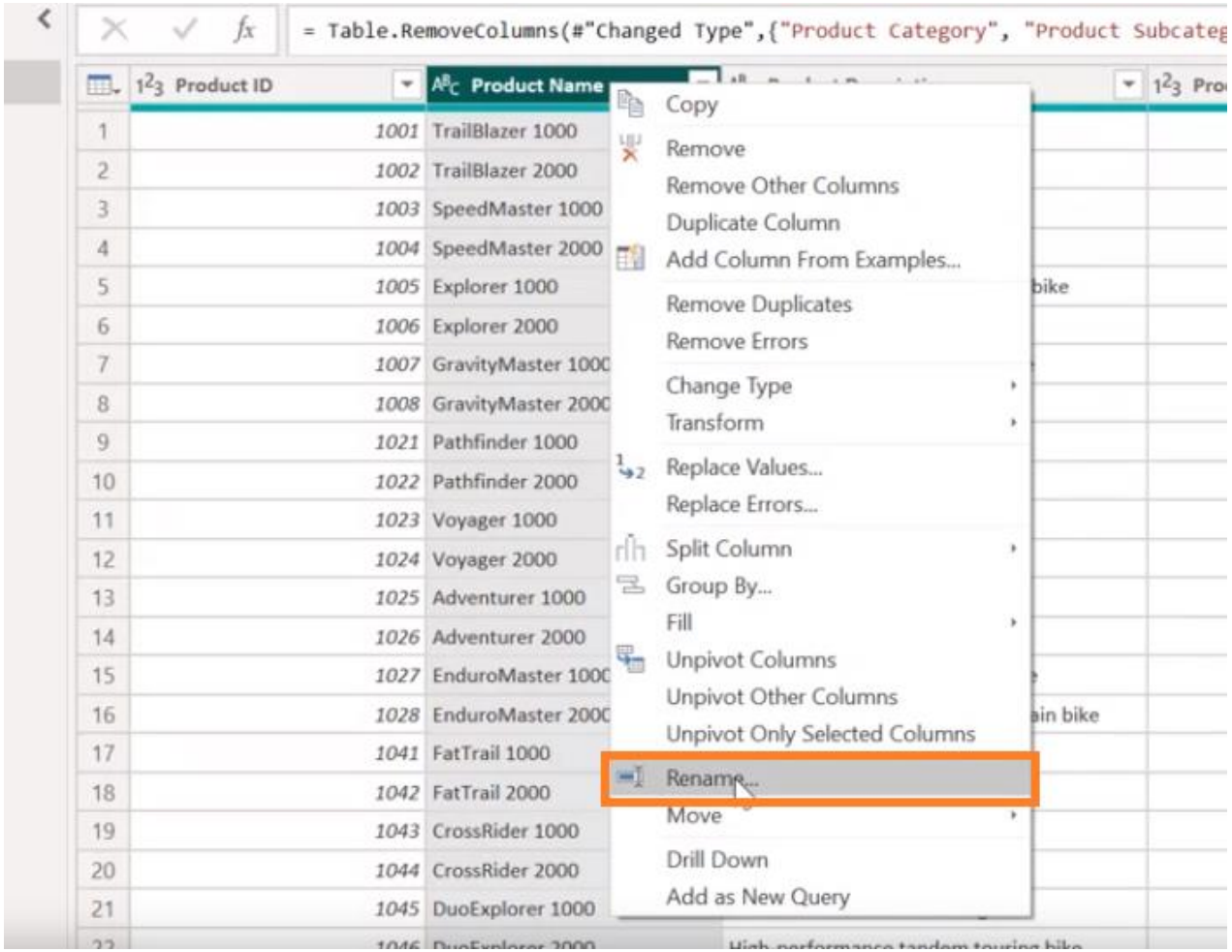
Power Query – Query Reuseability



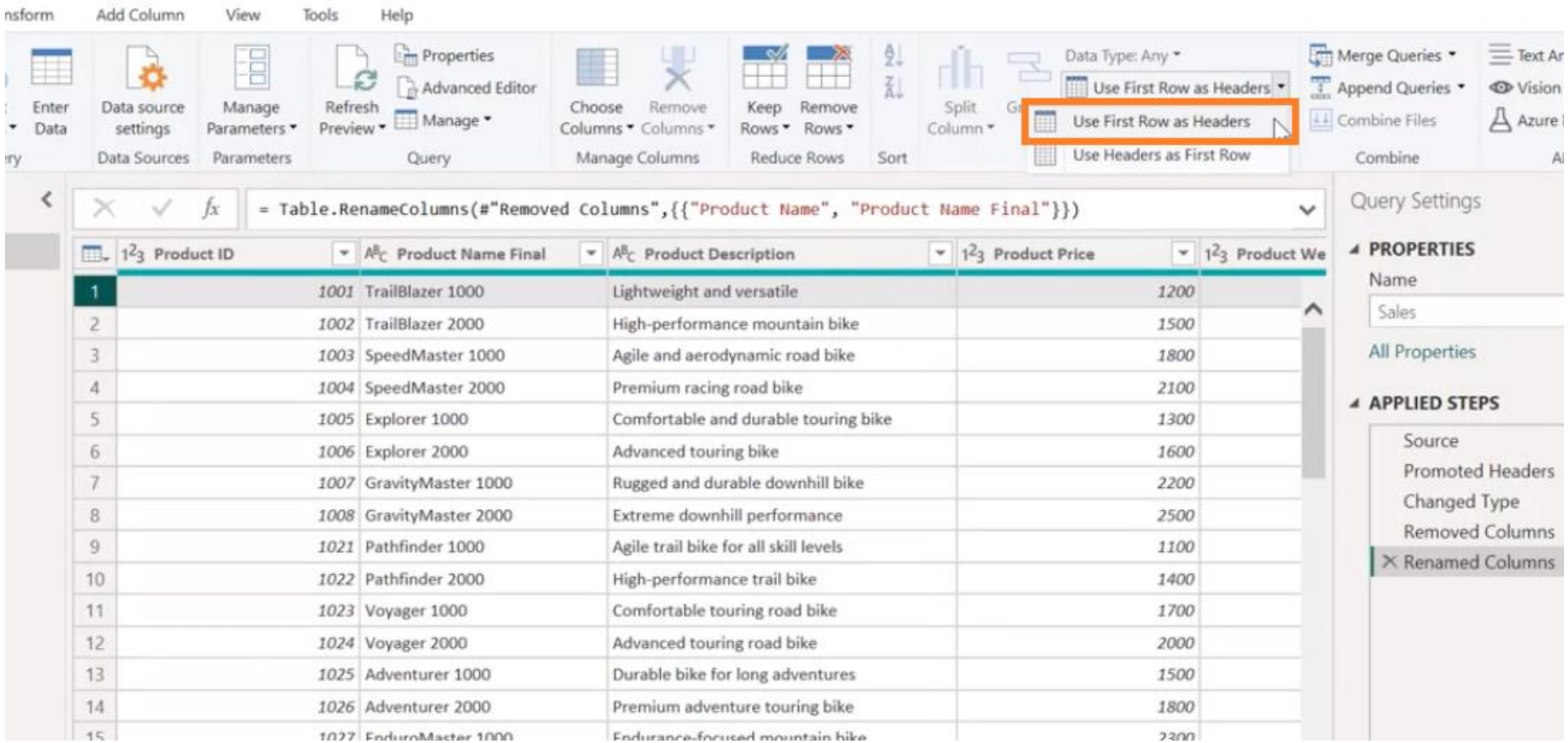
Working with columns



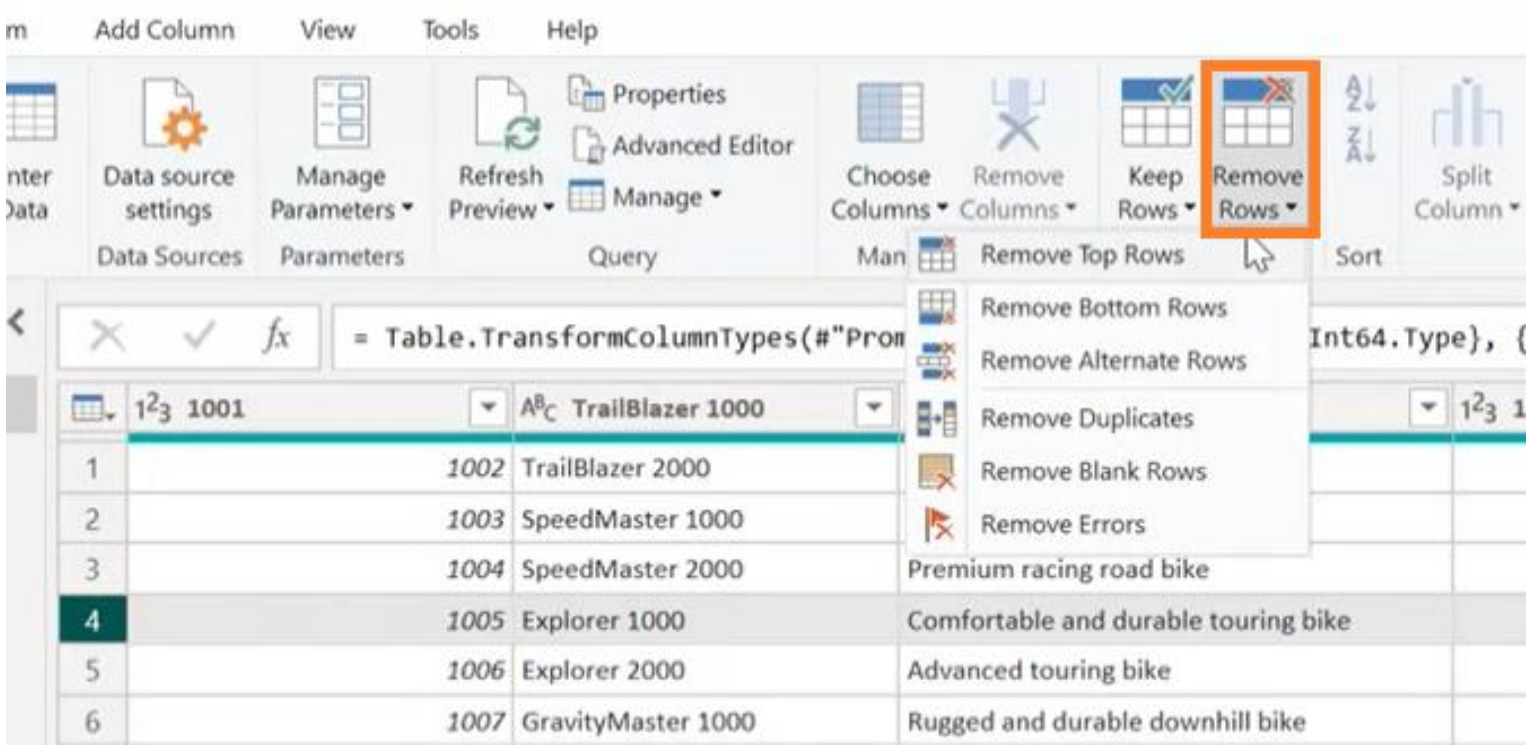
Working with columns (cont.)



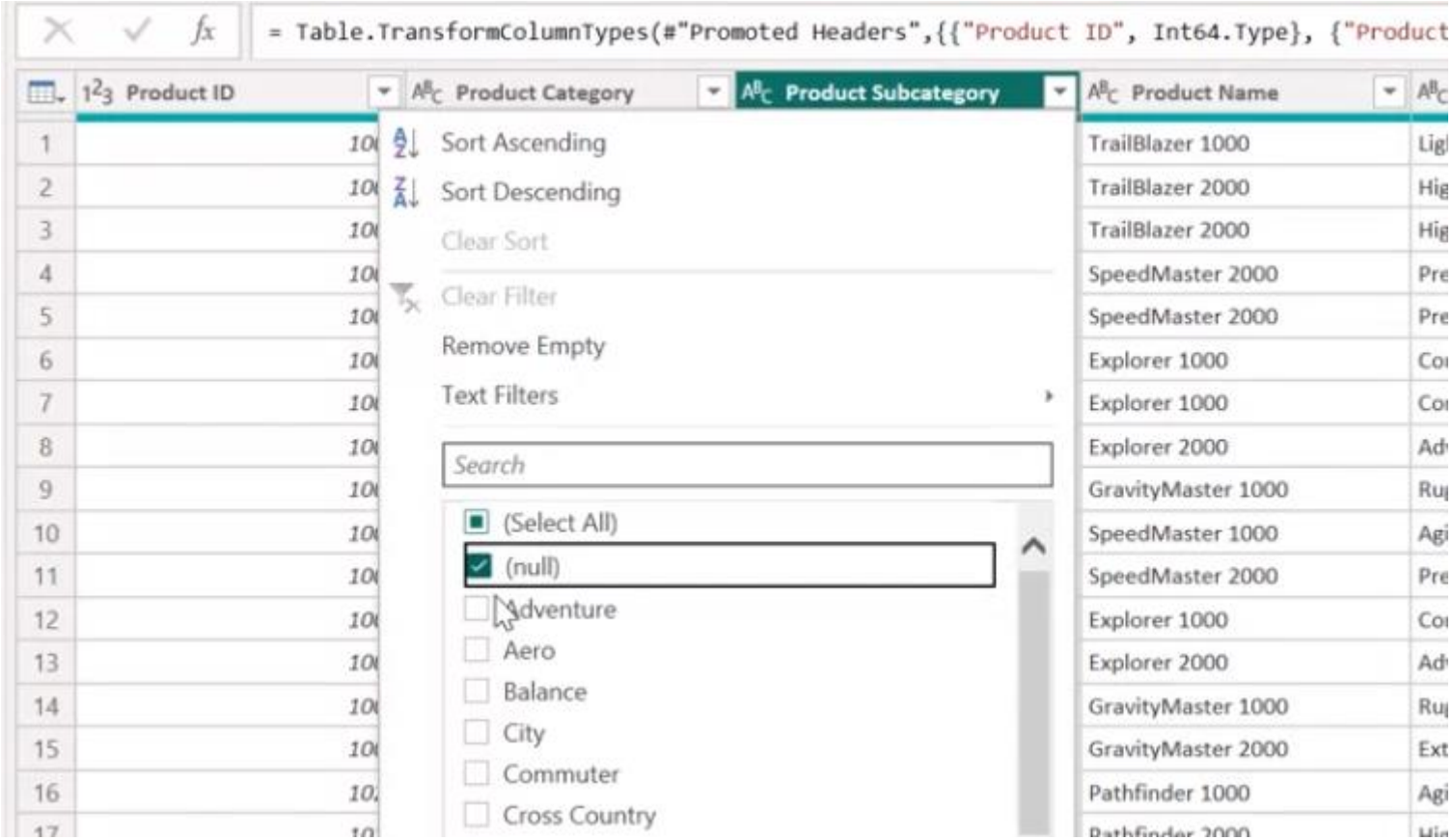
Working with columns (cont.)



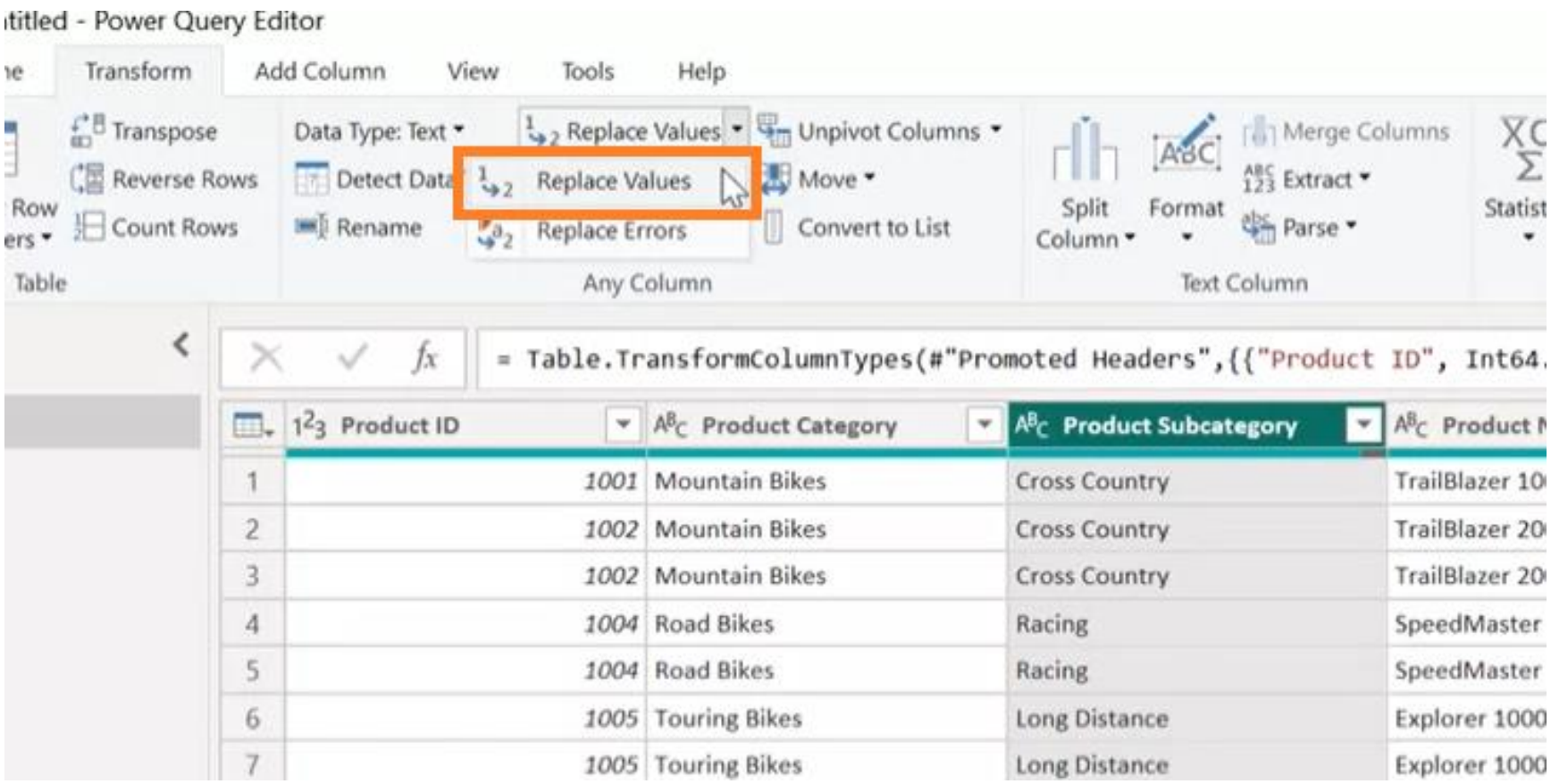
Working with columns (cont.)



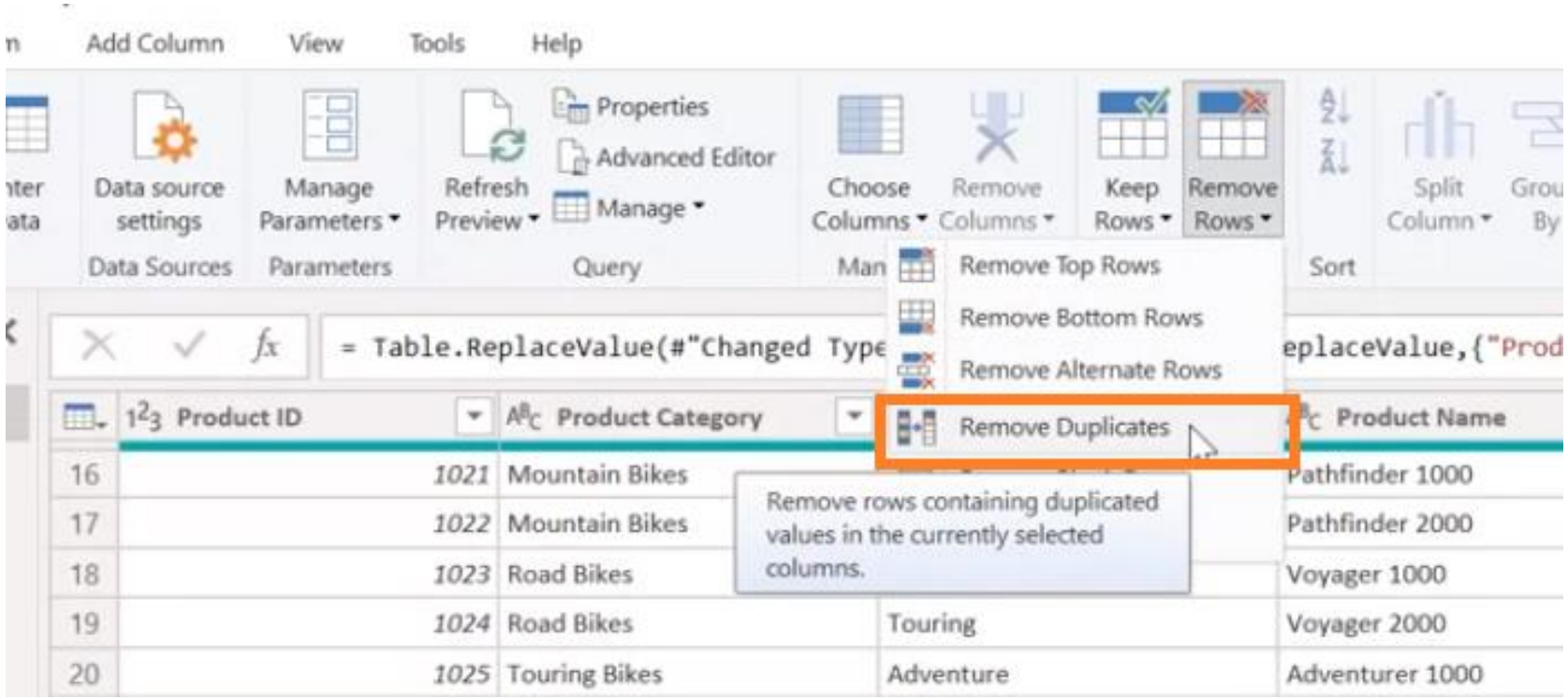
Dealing with errors in Power Query



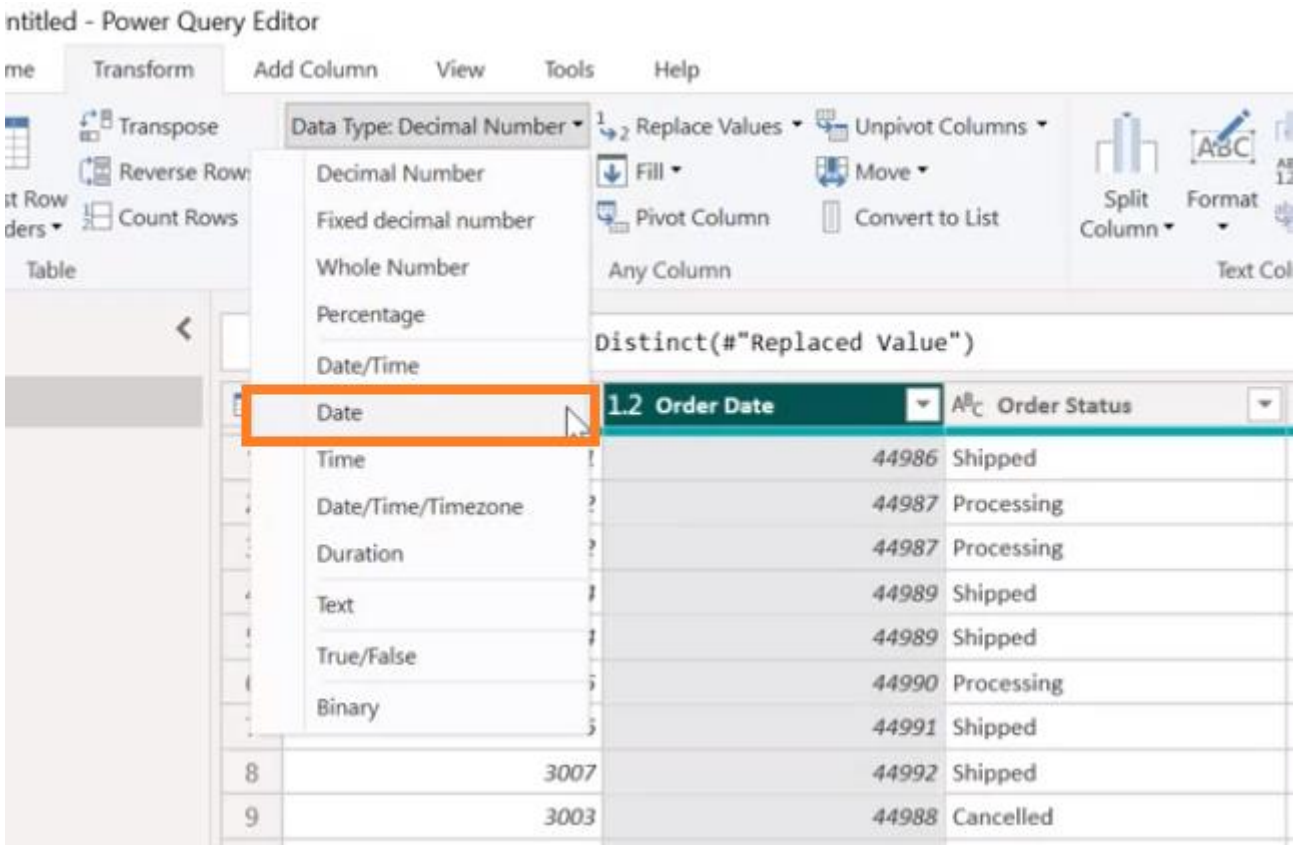
Dealing with errors in Power Query (cont.)



Dealing with errors in Power Query (cont.)



Dealing with errors in Power Query (cont.)



Exercise: Preparing a Dataset

Load the workbook

1. Download the Microsoft Excel workbook **SalesFile.xlsx.**
2. Import the **SalesFile.xlsx** Excel file as your dataset in Power BI.

Open the Power Query Editor

Click on the Transform Data button to open the Power Query Editor.

Exercise: Preparing a Dataset (cont.)

Address missing values

1. Locate and select the **Units Sold** column.
2. Identify all **null** values within the column and replace them with a value of **0**.
3. Repeat this task for the **Sale Price**, **Sales**, and **Profit** columns.

Exercise: Preparing a Dataset (cont.)

Clean the Manufacturing Price and Sale Price columns

1. Locate and select the **Manufacturing Price** and **Sale Price** columns.
2. Change the data type for both columns to **Decimal Number**.
3. Repeat this task for the **Sales** and **Profit** columns.

Exercise: Preparing a Dataset (cont.)

Clean the Discount Band Column

1. Select the **Discount Band** column.
2. Locate each instance of value **1** in the column. Replace each instance of this value with **None**.
3. Then change the data type of the column to **Text**.

Exercise: Preparing a Dataset (cont.)

Clean the Units Sold column

1. Select the **Units Sold** column. Search for and locate all instances of the text value **six hundred**.
2. Replace each instance of this text value with the numerical value **600**.
3. Then change the column's data type to **Whole Number**.

Exercise: Preparing a Dataset (cont.)

Address inconsistencies in the Date column

1. Select the **Date** column. Ensure that the column’s data type is **Date**.
2. The column also contains several null values. Replace all null values with the default date of **March 03rd 2023**.
3. Next, select the **Month Number** column. Change the column's data type to **Whole Number**.

Exercise: Preparing a Dataset (cont.)

Drop records with errors

1. Select the **Manufacturing Price** column. The column contains errors in rows **6** and **38**. Drop these rows.
2. Repeat the same steps for the errors in the **Sales** and **Profit** columns.

Drop duplicate rows

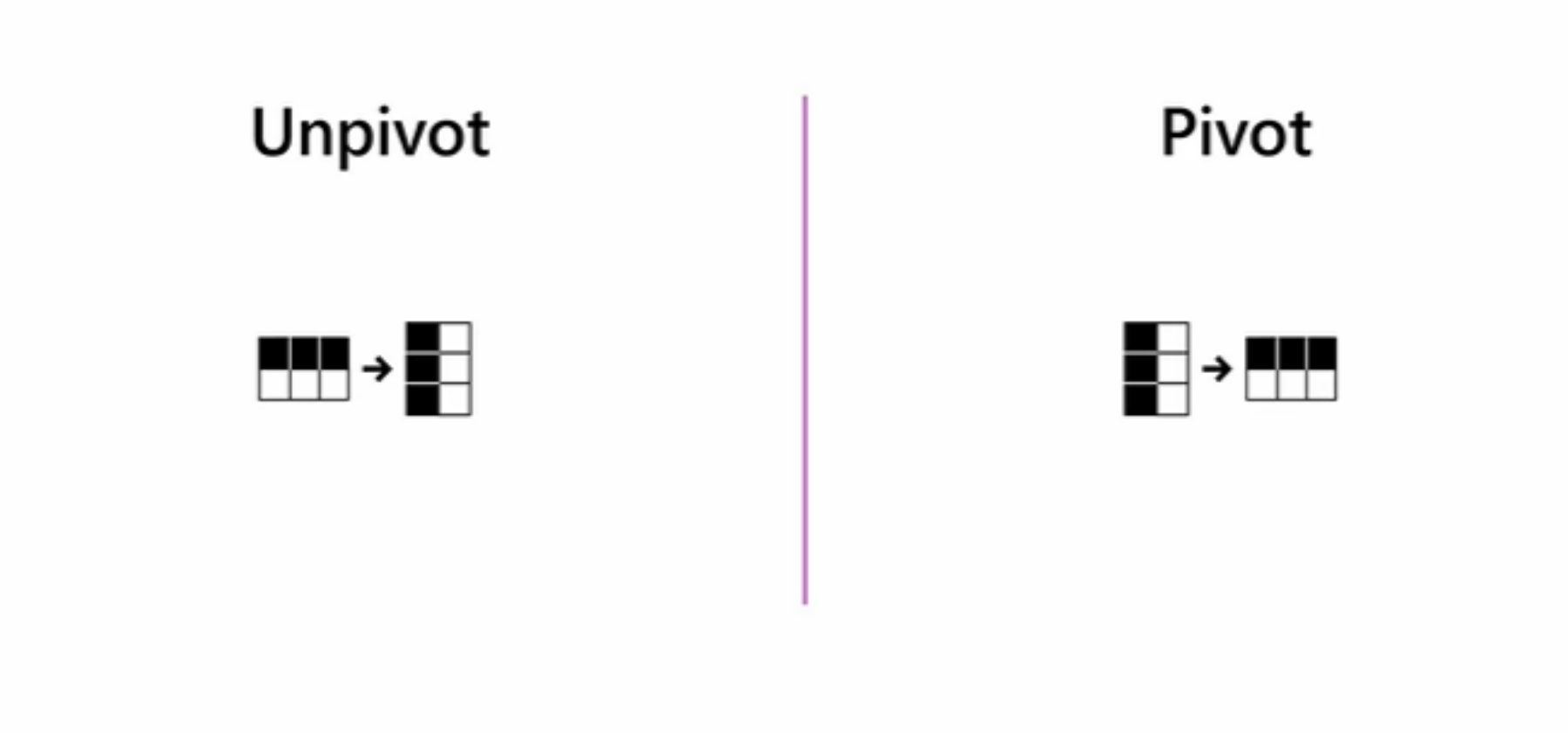
Use the **Drop Duplicates** feature to remove duplicate rows.

Exercise: Preparing a Dataset (cont.)

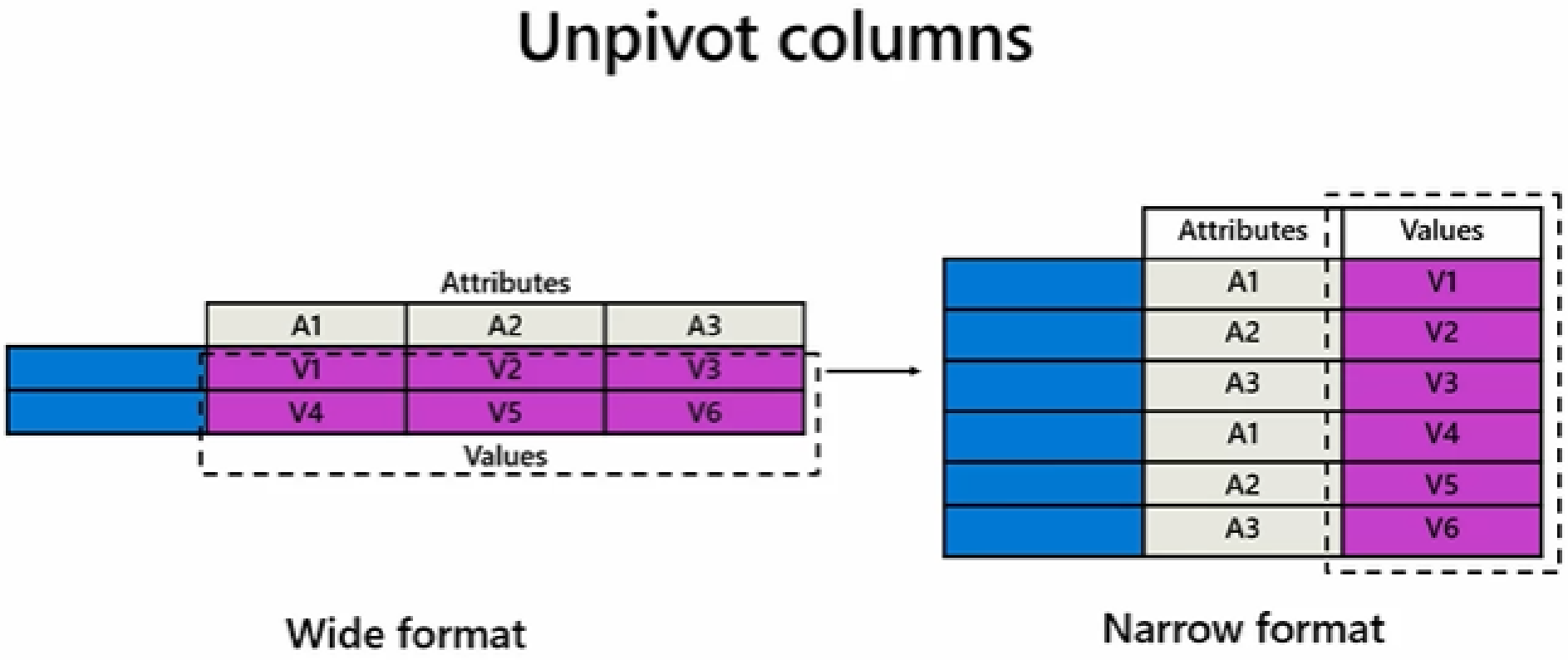
Apply the data transformations

Once you have completed all the above data cleaning steps, select the **Close & Apply** button on the top left.

Un-pivot and Pivot Columns



Un-pivot and Pivot Columns (cont.)



Un-pivot and Pivot Columns (cont.)

Unpivot columns

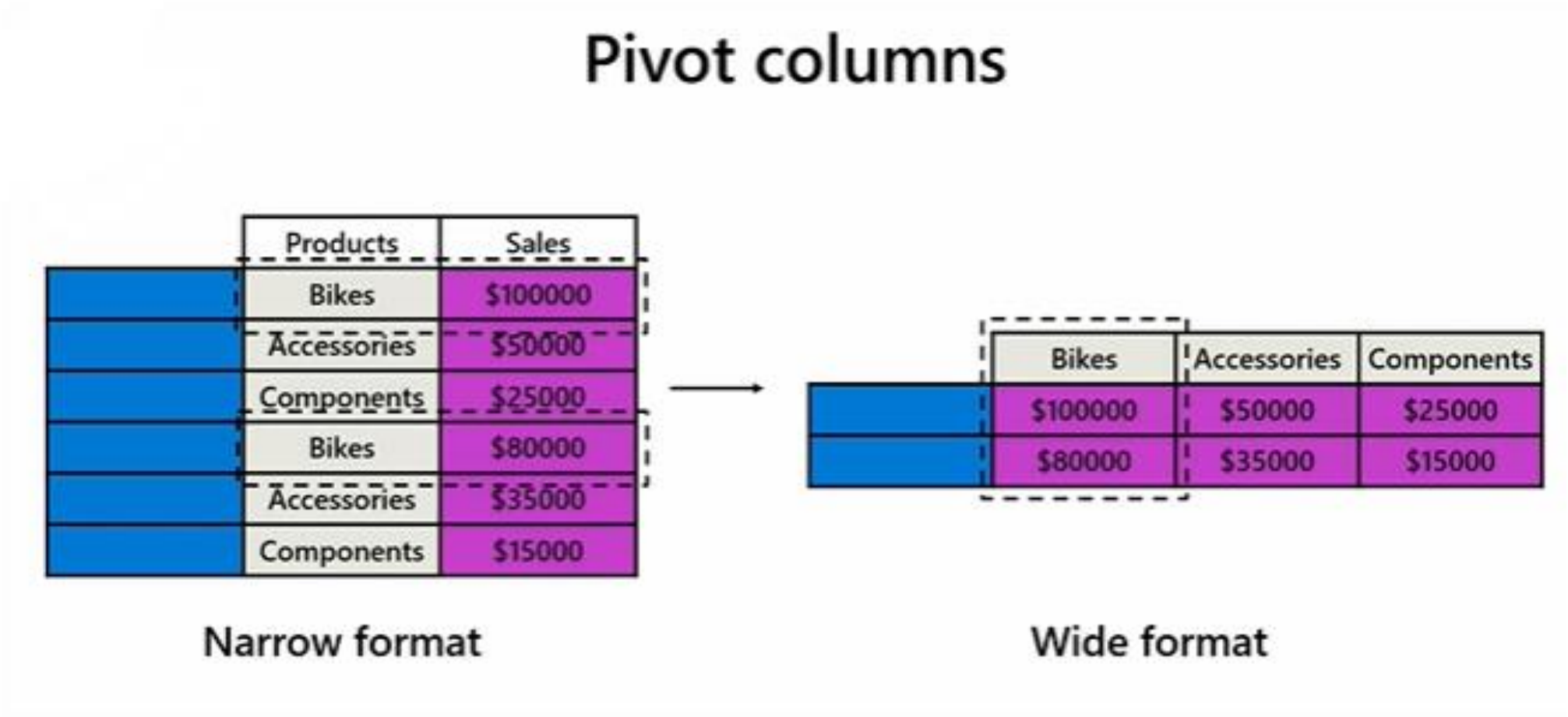
	Washington	Minnesota	California
Bikes	$100000	$80000	$120000
Accessories	$50000	$35000	$65000

Wide format

	Regions	Sales
Bikes	Washington	$100000
Bikes	Minnesota	$80000
Bikes	California	$120000
Accessories	Washington	$50000
Accessories	Minnesota	$35000
Accessories	California	$65000

Narrow format

Un-pivot and Pivot Columns (cont.)



Activity: Apply a pivot

Select the data source type

1. Open **Power BI Desktop**.
2. On the Home ribbon tab, inside the Data group, select the **Get Data** down arrow followed by **Excel** to find *Product-Color-Model.xlsx.*

Import Excel data

1. Import the **Excel data** to add the **Color Model** query to the Queries pane.
2. Observe the 3 columns in the table: **Product Name, Color** and **Model**.
3. Remove the **Product Name** column.

Activity: Apply a pivot (cont.)

Pivot columns

1. To pivot the table columns, select the **Color Model** query on the left menu.
2. Select the **Transform ribbon** tab, followed by **Pivot Column**.
3. On the **Pivot Column** window that displays, select **Model** as the Values Column.
4. Expand the **Advanced options** and select option **Count (All)** from the **Aggregate Value Function** dropdown list, and then select **OK**.

Quiz

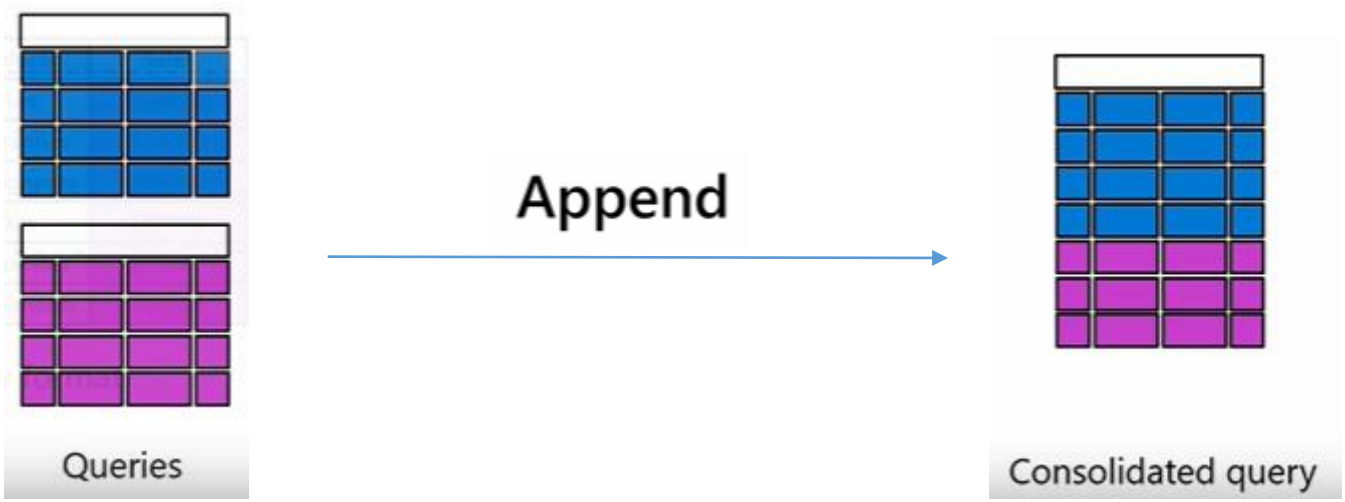
- Which of the following operations are steps in the data transformation process? Select all that apply.
  a) Creating insights from data
  b) Shaping data
  c) Removing data
  d) Cleaning data
- Which of the following data types are part of the number type group? Select all that apply.
  a) Fixed decimal number
  b) Binary
  c) Whole number
  d) Text

Quiz (cont.)

- Which one of the following features are used to track, re-order or delete the steps completed in Power Query?
  a) Applied Steps
  b) Queries
  c) New Source
  d) Properties

- Which of the following options can be used for Power Query Optimization? Select all that apply.
  a) Filter rows in the queries.
  b) Choose only the columns that you will use in the data model.
  c) Choose the right data types for columns.

Combining tables with append



Exercise: Appending two tables

Download the Excel files

Download the *AdventureWorksSales.xlsx* and *OtherSales.xlsx* files, which you will use in this exercise.

Open the Power Query Editor

Open the Power Query editor and import your datasets – *AdventureWorksSales* and *OtherSales*.

Exercise: Appending two tables (cont.)

Format Excel files

1. You have to append *OtherSales* data to *AdventureWorksSales* data. So, you will use *AdventureWorksSales* data as the first table and *OtherSales* data as second table.
2. For this reason, format the *OtherSales* data and rename the column names, using the *AdventureWorksSales* data, for example, **Quantity** to **OrderQty**, **Name** to **ProductName**, and **Total** to **LineTotal**.

Exercise: Appending two tables (cont.)

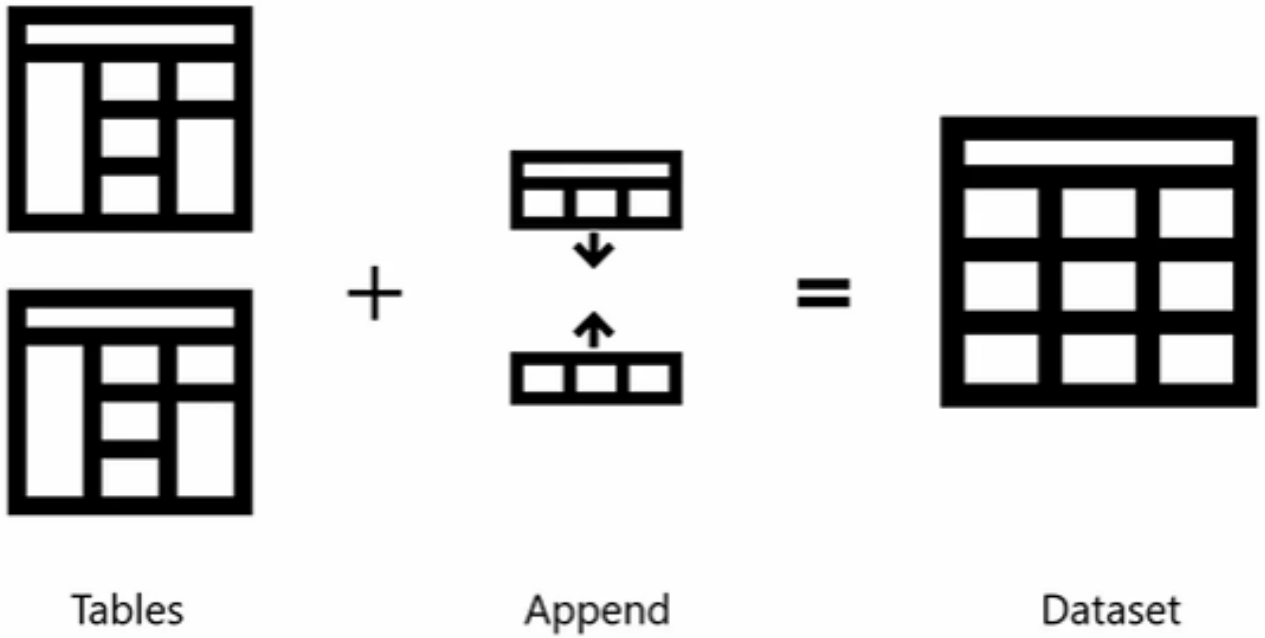
Append queries

Append queries in a new master table using the **Append Queries** button in the **Home** ribbon. In the newly created query, check the column names, row number and the values appended. Make sure that the operation has been completed successfully.

Rename new query

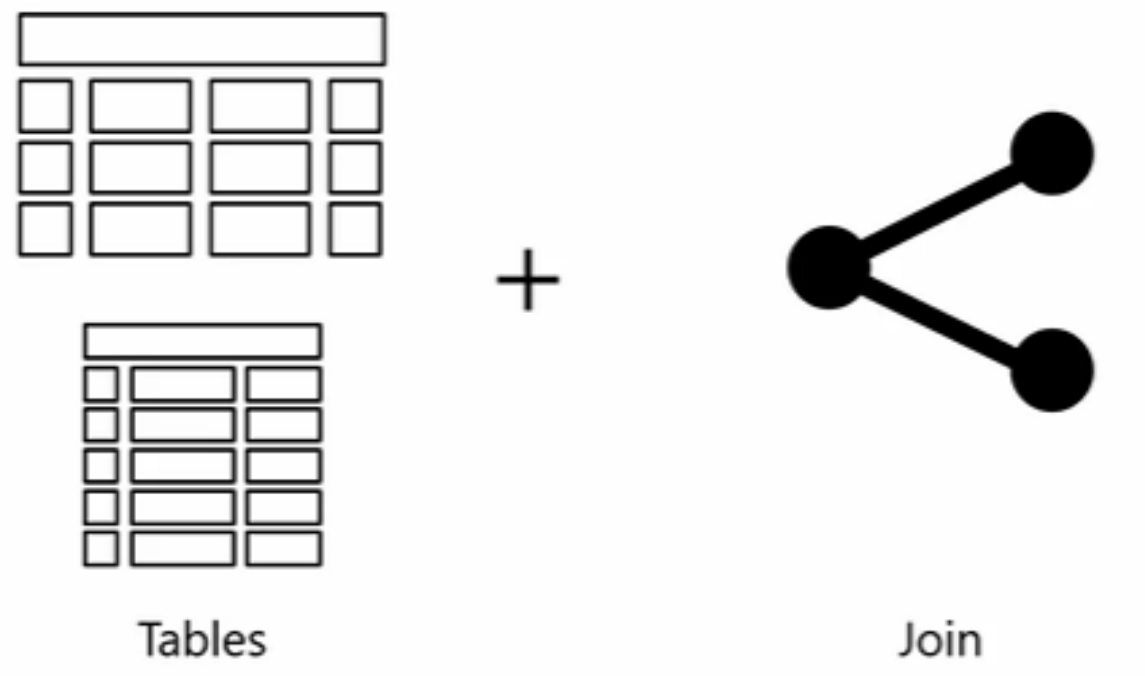
In the left menu, select the new query and change its name to *Consolidated Sales* and select **Enter** on the right pane, named **Properties**.

What is a join?



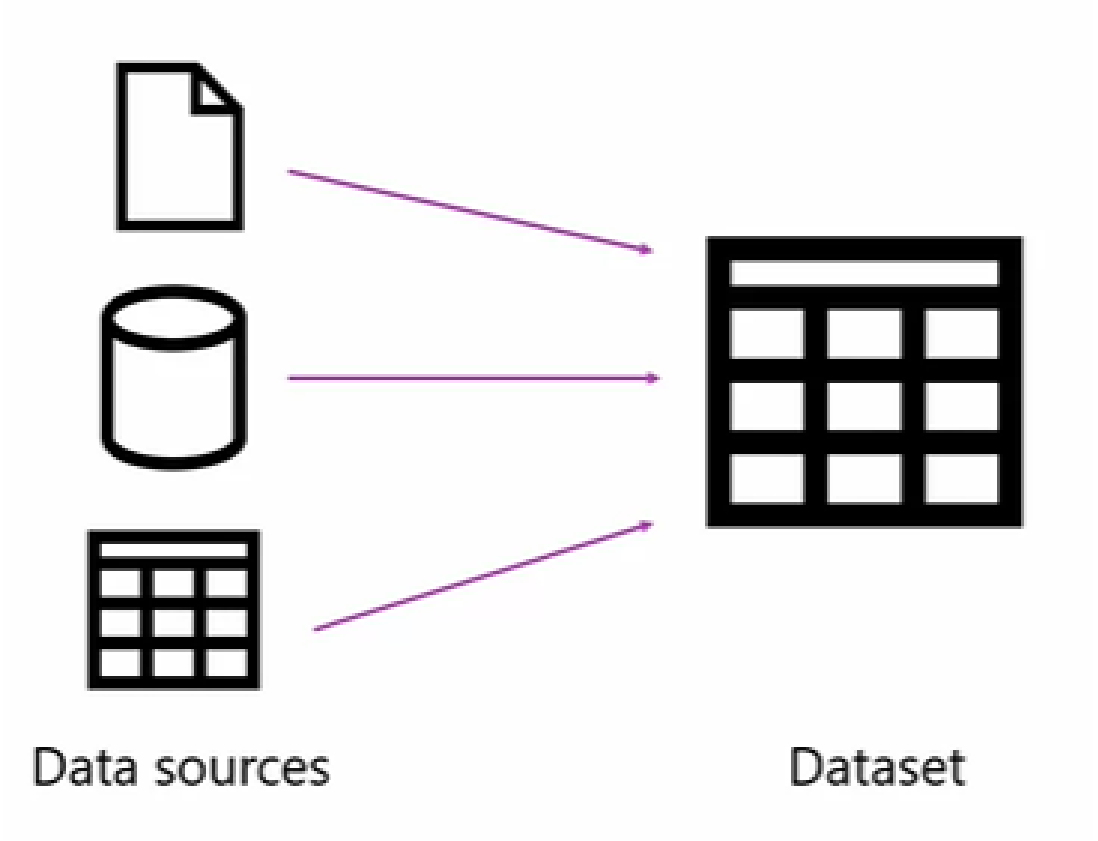
When you have two tables with the same structure, merging them is straightforward.

What is a join? (cont.)



When you have two tables with different structures, you need to specify the method of joining them.

What is a join? (cont.)



‘Join’ is when you combine multiple data sources to get a bigger dataset.

What is a join? (cont.)

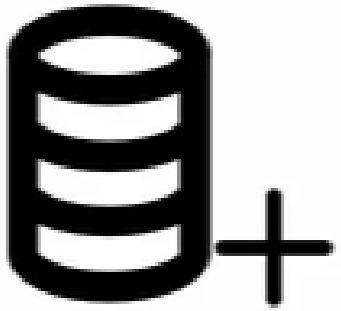
Products	
CategoryKey	
1	
2	

Categories	
CategoryKey	CategoryName
1	Bikes

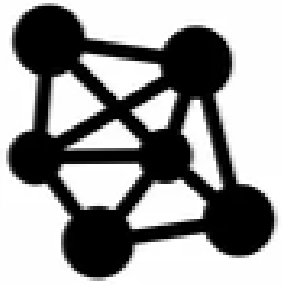
Here, you have two tables with the same column, but the column has different distributions in the two tables. Why is that?

What is a join? (cont.)

Match related data



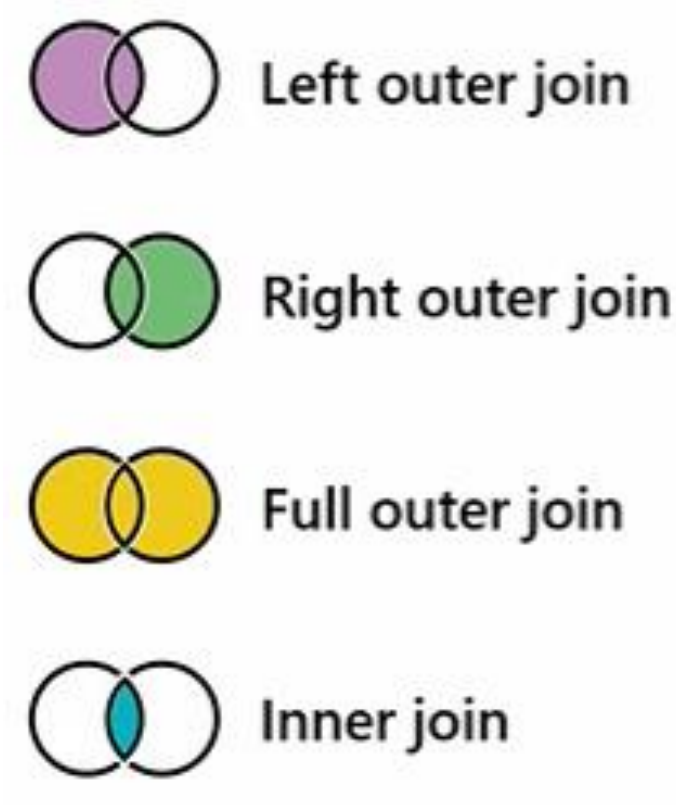
Integrate data



Relationships

When you merge data with join, it allows you to match and integrate related data.
It also allows you explore relationships between tables.

Join Types



Join Types: Left Outer

Left Table

Date	CountryID	Units
1/1/2020	1	40
1/2/2020	1	25
1/3/2020	3	30
1/4/2020	4	35

Right Table

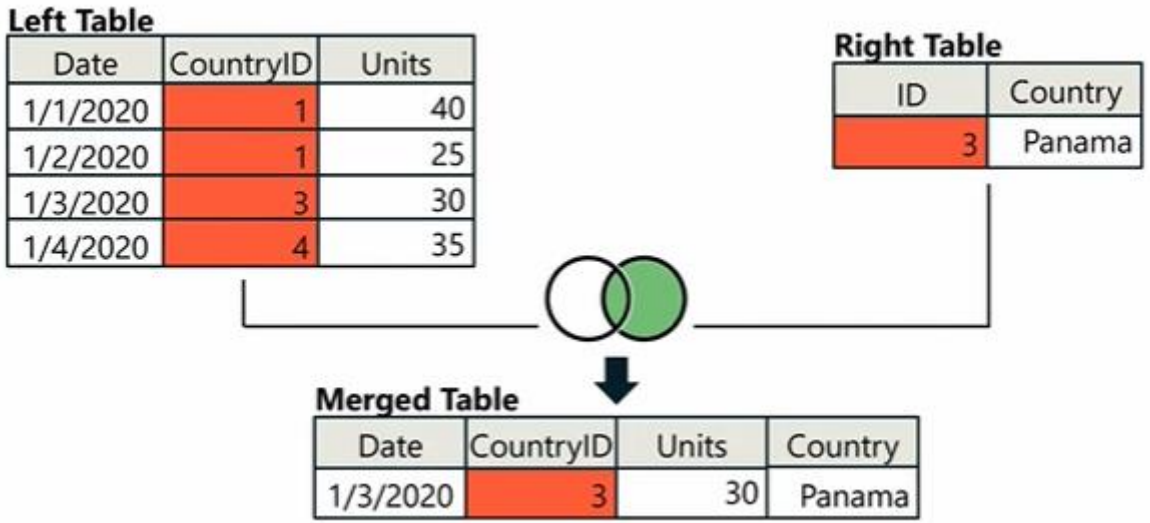
ID	Country
1	USA
2	Canada
3	Panama

Merged Table

Date	CountryID	Units	Country
1/1/2020	1	40	USA
1/2/2020	1	25	USA
1/3/2020	3	30	Panama
1/4/2020	4	35	*null*

If there is no match between the tables, default/null values will be used.

Join Types: Right Outer



If there is no match between the tables, default/null values will be used.

Join Types: Full Outer

Left Table

Date	CountryID	Units
1/1/2020	1	40
1/2/2020	1	25
1/3/2020	3	30
1/4/2020	4	35

Right Table

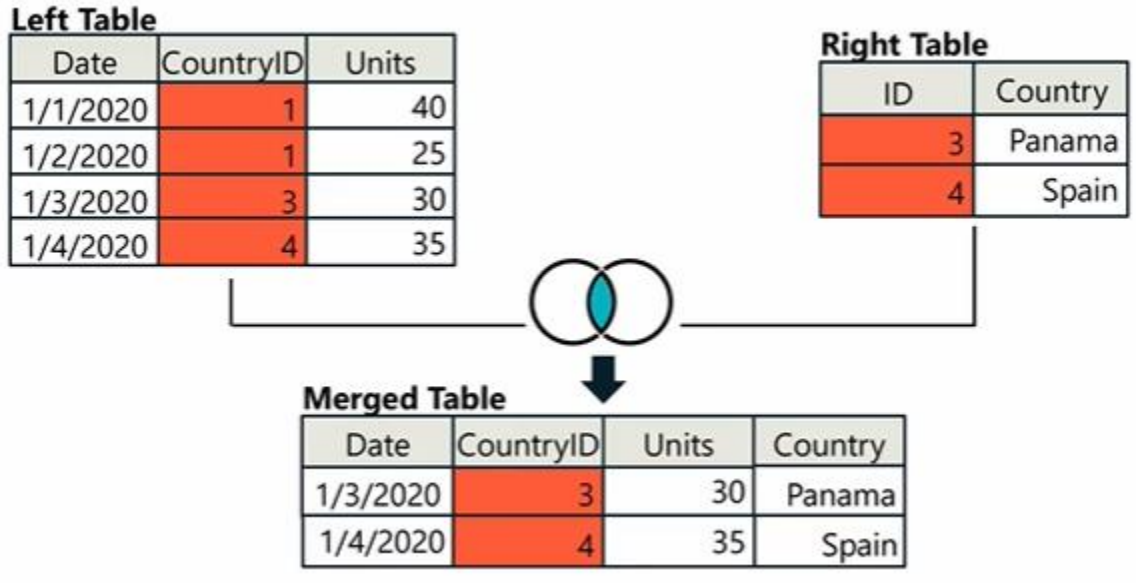
ID	Country
1	USA
2	Canada
3	Panama

Merged Table

Date	CountryID	Units	ID	Country
1/1/2020	1	40	1	USA
1/2/2020	1	25	1	USA
1/3/2020	3	30	3	Panama
1/4/2020	4	35	*null*	*null*
null	*null*	*null*	2	Canada

The full outer join is used when you want to retrieve all records from both tables, regardless of whether they have matching values in the join condition.

Join Types: Inner



For inner join, only matching rows from both left and right tables are merged together.

Exercise: Merging two data sources

Download the Excel files

Download the *Sales.xlsx* and *Product.xlsx* files, which will be used in this exercise.

Create a Power BI project

1. Create a Power BI project and open the Power Query editor.
2. Import your datasets, *Sales* and *Product*.

Exercise: Merging two data sources (cont.)

Merge queries

1. After selecting the *Sales* data in the **Queries** pane, select **Merge Queries**.
2. In the opened window, the *Sales* table will be shown automatically in the upper section of the window.
3. Choose the next table for merging, which is *Product*.
4. **ProductKey** is the common column between the tables, so click on the **ProductKey** columns in each table.
5. For the **Join Kind** dropdown, choose the join type **Left Outer Join**, which selects all records from the left table and matching records from the right table.

Exercise: Merging two data sources (cont.)

Select column(s) from Product

1. After you merged the tables, a new column, named **Product** is added to the right side of the *Sales* data. This allows you to choose columns from the *Product* table.
2. Select the column named **Product** from the *Product* table.

Exercise: Merging two data sources (cont.)

Choose and reorder columns from Sales

1. After you add the new column, **Product,** it is added to the *Sales* query as **Product.Product**. You must rename this column as **Product** to avoid confusion.
2. Move the **Product** field from right to left.
3. Remove the unwanted columns, **Product Key** (name of product is added by merge, so you will not need the key value of product), **Reseller Key**, **Employee Key** and **Sales Territory Key** columns.
4. Reorder the final list as indicated in your task to **Sales Order Number**, **Order Date**, **Product**, **Quantity** and **UnitPrice**.

Quiz

- Which feature allows you to combine related data between differently structured data sources in Power Query?
  a) Appending
  b) Merging
  c) Grouping

- Which of the following can be considered as a purpose of merging data with joins? Select all that apply:
  a) Integrating Data
  b) Exploring Relationships
  c) Expanding Data
  d) Matching Related Data

Quiz (cont.)

- The full outer join is useful when you want to retrieve all the records from both tables, regardless of whether they have matching values in the join condition. True or False?

- You import two Microsoft Excel tables named *Product* and *Categories* into Power Query. There are 319 rows in the *Product* table. Nine of the total rows in the *Product* table do not have Categories data, so the **CategoryKey** of these rows has **NULL** values. Your manager asked you to list Product data by showing their category names including the rows which have NULL values in **CategoryKey** column. What should you do to accomplish this task?
  a) Merge *Product* and *Categories* tables based on **ResellerKey** column.
  b) Merge *Product* and *Categories* tables based on **CategoryKey** column by choosing **Inner Join** in the join kind dropdown.
  c) Merge *Product* and *Categories* tables based on **CategoryKey** column by choosing **Left Outer Join** in the join kind dropdown.