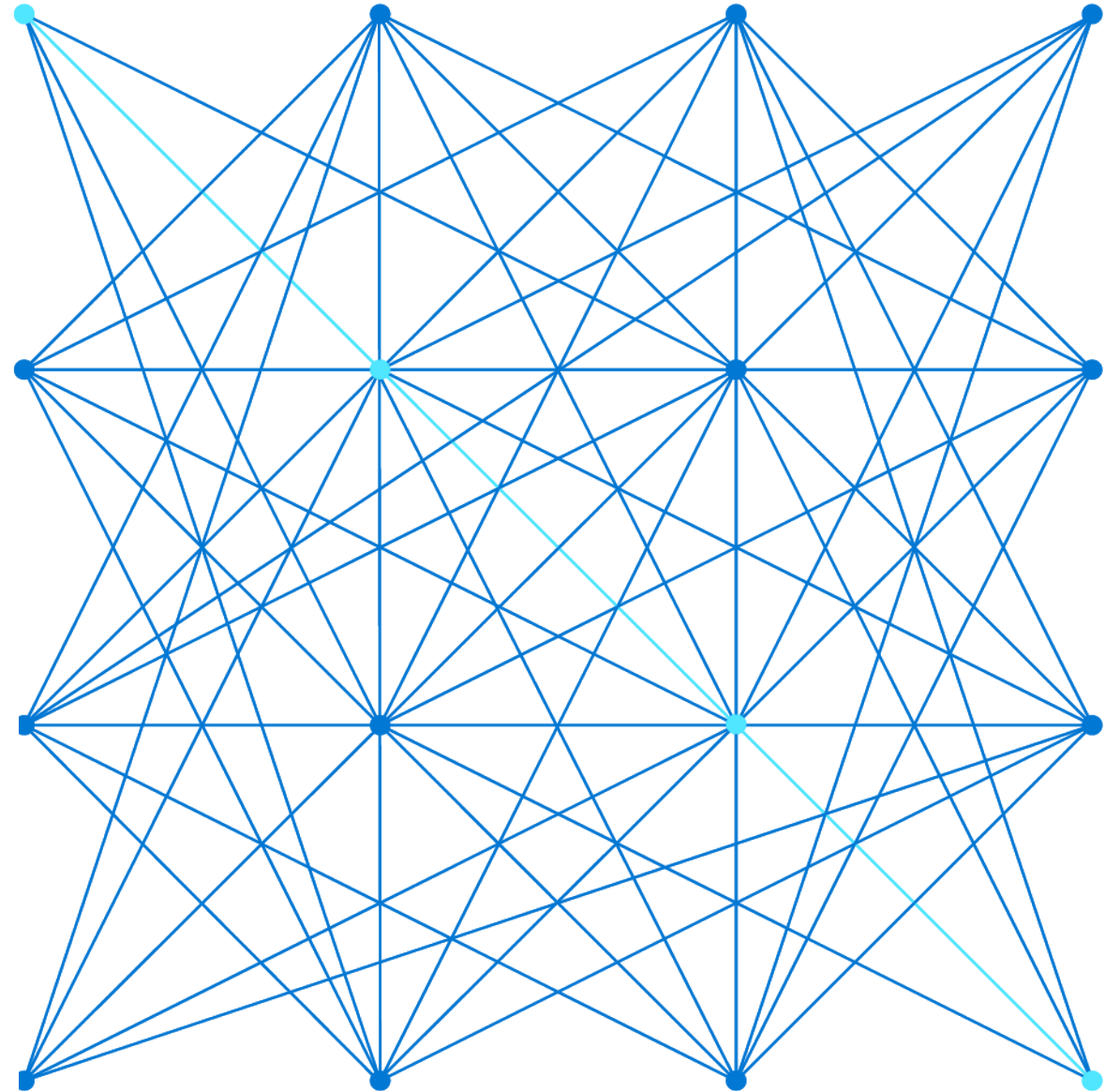


# DA-100 Analyzing Data with Power BI 3

Prof. Ernesto Lee



# **Module 3: Clean, Transform, and Load Data In Power BI**

# Learning Objectives

You will learn the following concepts:

- Profiling the Data
- Shaping the Data
- Enhancing the structure of the data





# Lesson 1: Profiling the Data



# Profiling Data and Examining Structures

The screenshot displays the Microsoft Power Query Editor interface. On the left, a diagram shows the data model with tables: Association, Product, Customer, Sales, and Calendar. The main area shows a query named 'SalesPerson' with columns: 1.2 Latitude, 1.2 Longitude, 1.2 Total Excluding Tax, 1.2 Total Including Tax, 1.2 Profit, and A. SalesPerson. The 'A. SalesPerson' column is highlighted, and its data is shown in a table view. Below the table, a 'Column statistics' and 'Value distribution' section is visible, showing the distribution of values for the 'SalesPerson' column.

**Column statistics**

Statistic	Value
Count	1000
Error	0
Empty	0
Distinct	10
Unique	0
Empty string	0
Min	Amy Trefl
Max	Taj Shand

**Value distribution**

Value	Count
Anthony Grosse	1000
Kayla Woodcock	1000
Sophia Hinton	1000
Archer Lambie	1000
Taj Shand	1000
Hudson Hollinworth	1000
Hudson Onslow	1000
Jack Potter	1000
Amy Trefl	1000
Lily Code	1000

Data profiling is understanding the state and structure of the data you are working with.

# Review Questions

- Q01 – How many rows does Power Query scan to detect the type of data in the columns?
- A01 – 1000
- Q02 – Data profiling is defined as what?
- A02 – Studying the nuances of the data
- Q03 – What is the risk of having null values in a numeric column?
- A03 – DAX expressions that AVERAGE data will be incorrect.



# Lesson 2: Shaping the Data





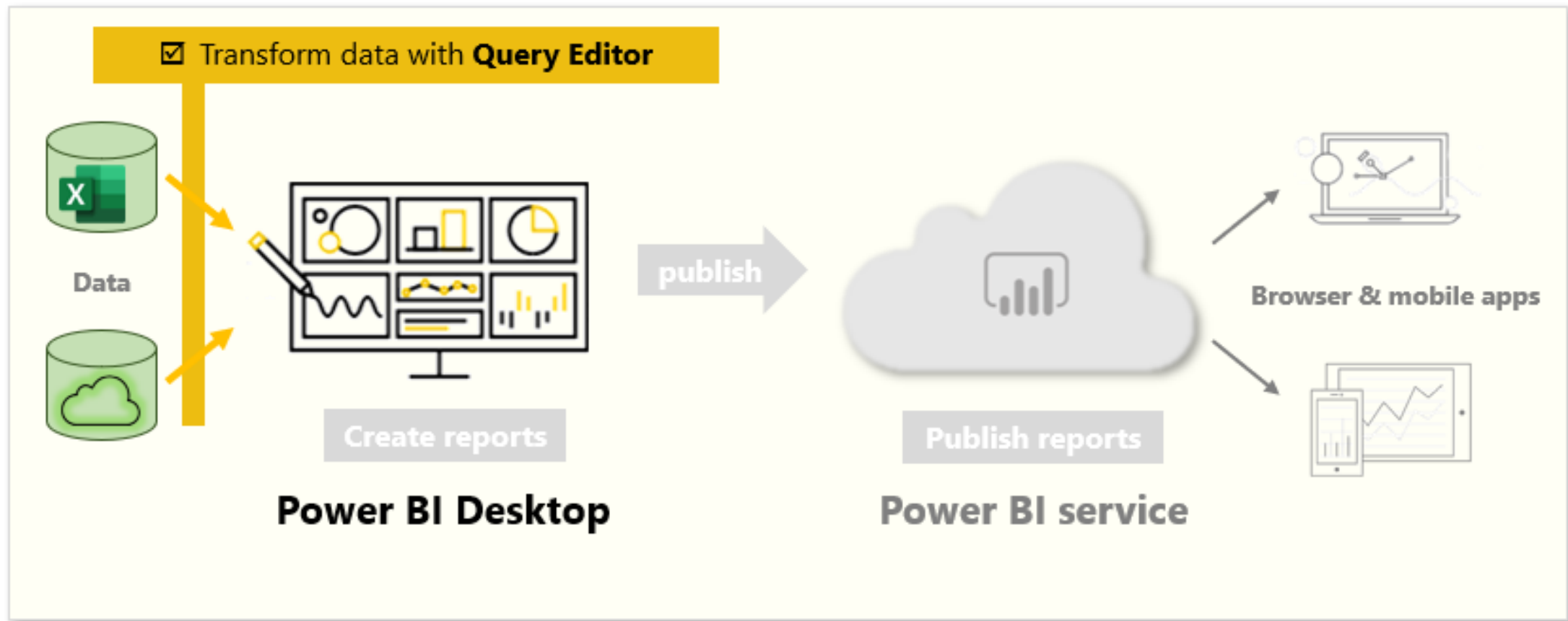
# Introduction

- Benefits of clean data:
  - More accurate results
  - Better organized tables
  - Simpler data navigation
  - Human-readable values



# Identify column headers and names

Use Power Query Editor to clean up and shape data.



# Shaping Table Structure

Shape the data to meet reporting needs.

The screenshot shows the Power Query ribbon with the 'Remove Rows' dropdown menu open. The 'Remove Top Rows' option is highlighted. The ribbon includes tabs for Home, Transform, Add Column, View, Tools, and Help. The 'Remove Rows' dropdown menu contains the following options: Remove Top Rows, Remove Bottom Rows, Remove Alternate Rows, Remove Duplicates, Remove Blank Rows, and Remove Errors.

	ProductSubcategoryID	Subcategory Name	January
1			
2	ProductSubcategoryID	Name	
3	1	Mountain Bikes	780000

The screenshot shows the Power Query ribbon with the 'Remove Columns' dropdown menu open. The 'Remove Columns' and 'Remove Other Columns' options are highlighted. The ribbon includes tabs for Properties, Advanced Editor, Manage, Query, and Sort. The 'Remove Columns' dropdown menu contains the following options: Remove Columns, Remove Other Columns, Keep Rows, and Remove Rows.

	Column13	Column14
2	November	December
	880000	890000
	9500	10000
	511000	512000
	2100	2200
	2100	2200



# Unpivot and Pivot columns

Transfer data from rows to columns, and columns to rows.

	Category Name	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

# Review Questions

- Q01 – The primary data preparation tool in Power BI is called what?
- A01 – Power Query Editor
- Q02 – The process of shaping data by converting your flat data into a table that contains an aggregation value for each unique value in a column is called what?
- A02 – Pivot (pivoting a column)
- Q03 – What can be achieved by removing unnecessary rows and columns?
- A03 – Deleting unnecessary rows and columns will reduce a dataset size and its good practice to load only necessary data into your data model.



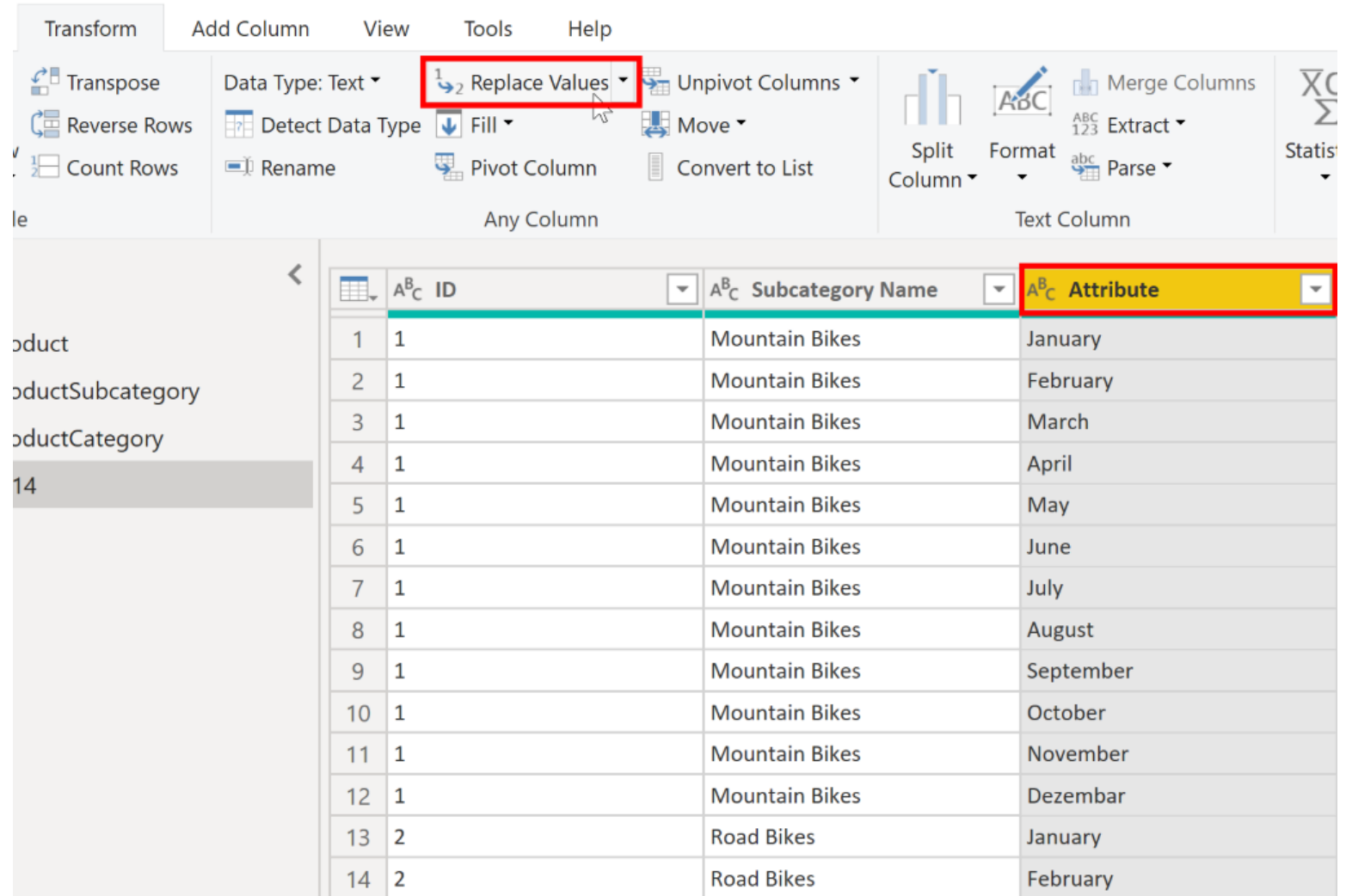
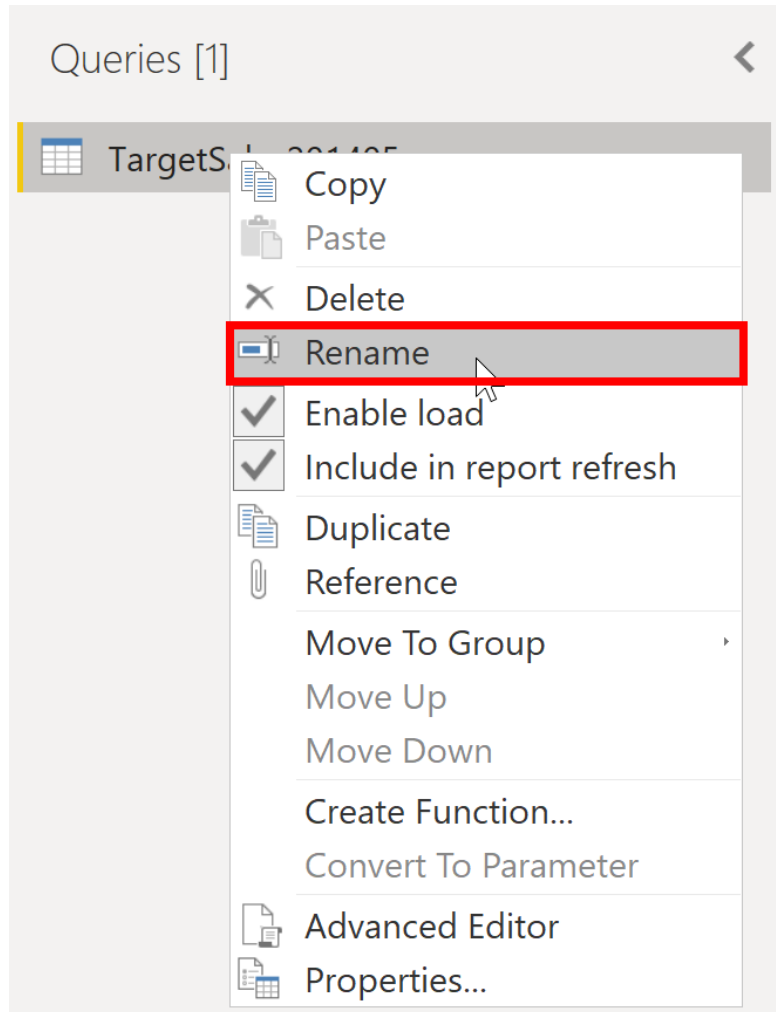


# Lesson 3: Enhance the Data Structure



# Apply user-friendly value replacements

Make information user-friendly.





# Evaluate and Change Column Data Types

	1 <sup>2</sup> <sub>3</sub> SalesOrderID	A <sup>B</sup> <sub>C</sub> OrderDate	A <sup>B</sup> <sub>C</sub> Sort_of_Sales	1 <sup>2</sup> <sub>3</sub> ProductID	1 <sup>2</sup> <sub>3</sub> OrderQty
1	52242	07/07/2013	Internet	870	1
2	52592	14/07/2013	Internet	870	1
3	52694	16/07/2013	Internet	870	1
4	52799	18/07/2013	Internet	870	1
5	53799	03/08/2013	Internet	870	1
6	54058	08/08/2013	Internet	870	1
7	54059	08/08/2013	Internet	870	1
8	54063	08/08/2013	Internet	870	1
9	54158	10/08/2013	Internet	870	1
10	54281	12/08/2013	Internet	870	1

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 Tables into a Single Table

## 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

Add >>

Tables to append

- Production Suppliers
- Sales Customers
- HR Employees

OK

Cancel

## 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 SCQ
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



# Use Advanced Editor to Modify M Code

See the code that Power Query Editor is creating with each step.

Sales Orders

Display Options ?

```
let
    Source = Sql.Database("localhost", "tsqlv4"),
    Sales_Orders = Source[[Schema="Sales",Item="Orders"]][Data],
    #"Split Column by Delimiter" = Table.SplitColumn(Sales_Orders, "shipaddress", Splitter.SplitTextByDelimiter(",", QuoteStyle.Csv), {"shipaddress.1", "shipaddress.2"}),
    #"Changed Type" = Table.TransformColumnTypes(#"Split Column by Delimiter",{{"shipaddress.1", type text}, {"shipaddress.2", type text}})
in
    #"Changed Type"
```

✓ No syntax errors have been detected.

Done

Cancel

# Review Questions

- Q01 – What is not a best practice for naming conventions in Power BI?
- A01 – Abbreviated column names.
- Q02 – What functionality lets you see the code that is generated as part of each transformation step?
- A02 – Advanced Editor
- Q03 – If you have two queries that contain different data with the same structure, and you want to combine them into one query, which operation should you perform?
- A03 - Append

# Lab: Shape and Combine Data



# Module Overview

We covered the following concepts:

- Profiling the Data
- Shaping the Data
- Enhancing the structure of the data

# References

- DA-100 Clean, transform, and load data in Power BI  
<https://docs.microsoft.com/en-us/learn/modules/clean-data-power-bi/>

