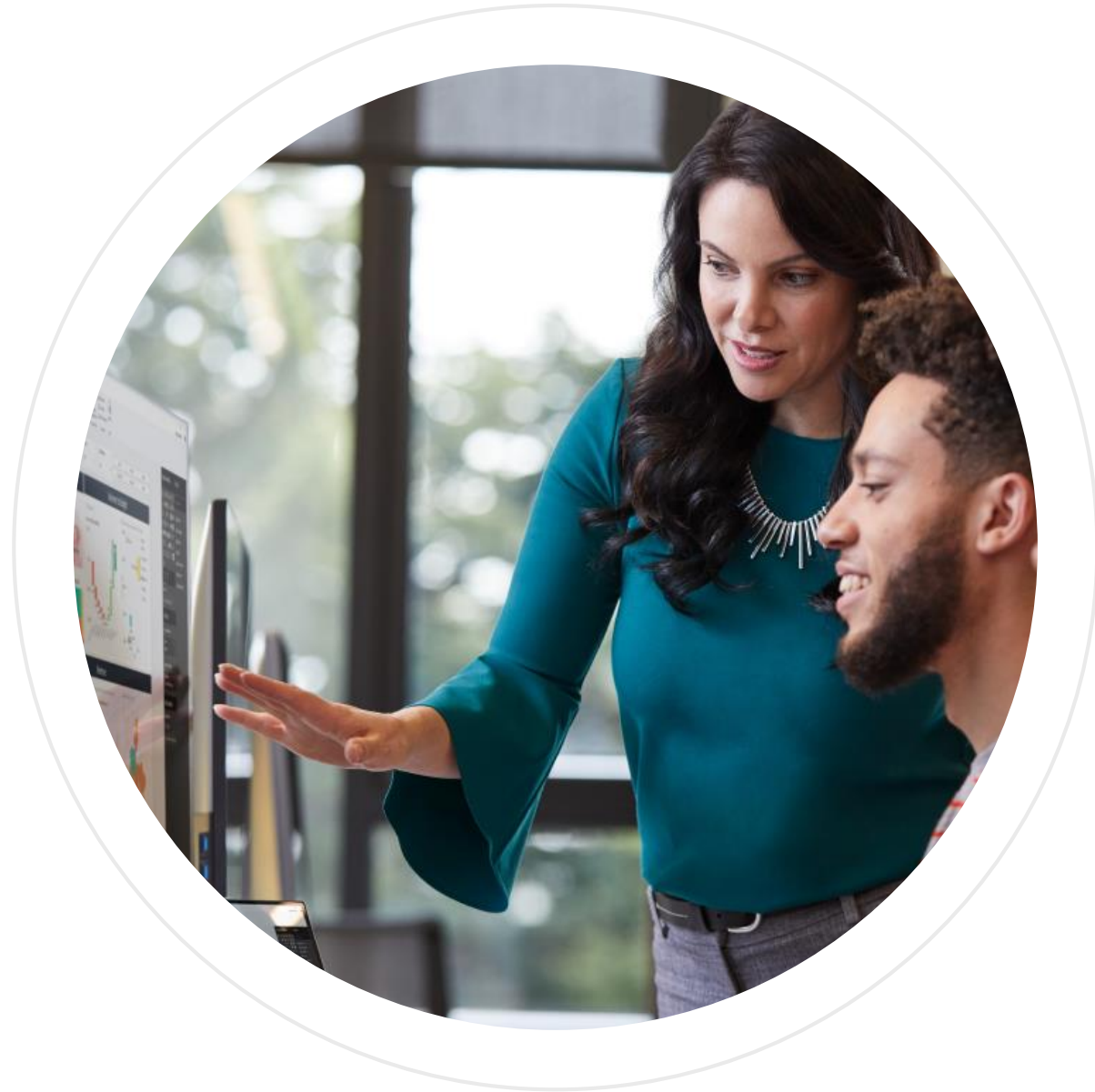


M03: Cleaning, Transforming, and Loading Data

Mohammed Arif



Module Agenda



Shaping the Data



Profiling the Data



Enhance the Data Structure

Learning Objectives

You will learn the following concepts:

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

Lesson 1: Shaping the Data

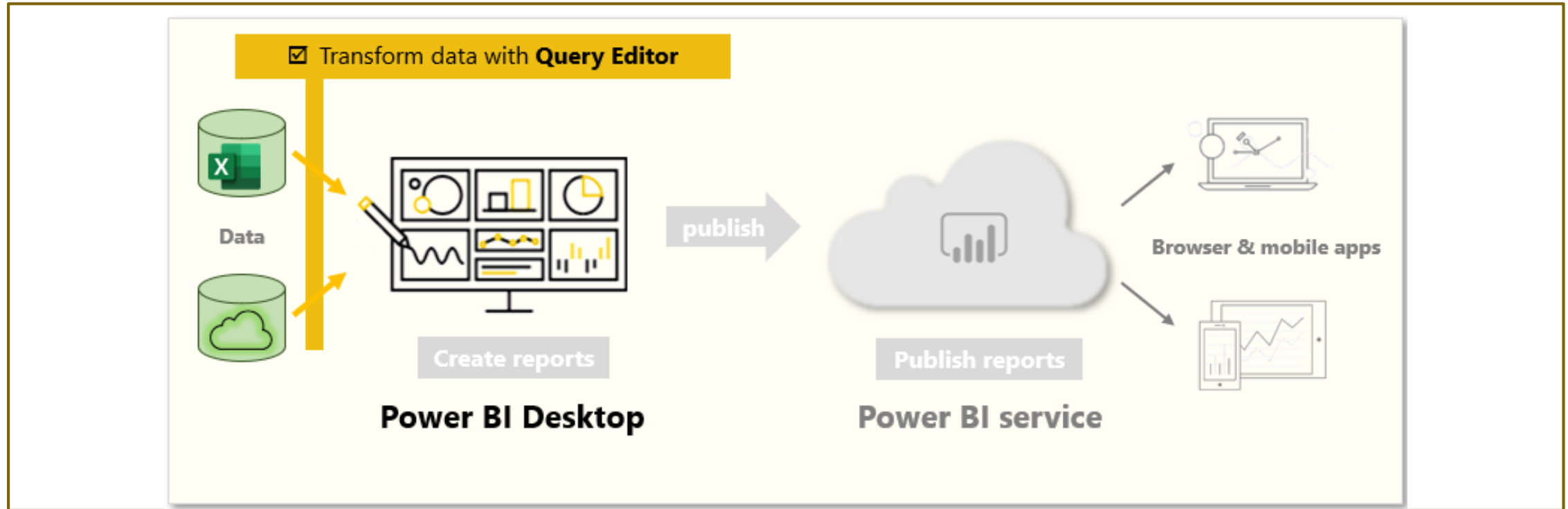


Learning Objectives

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

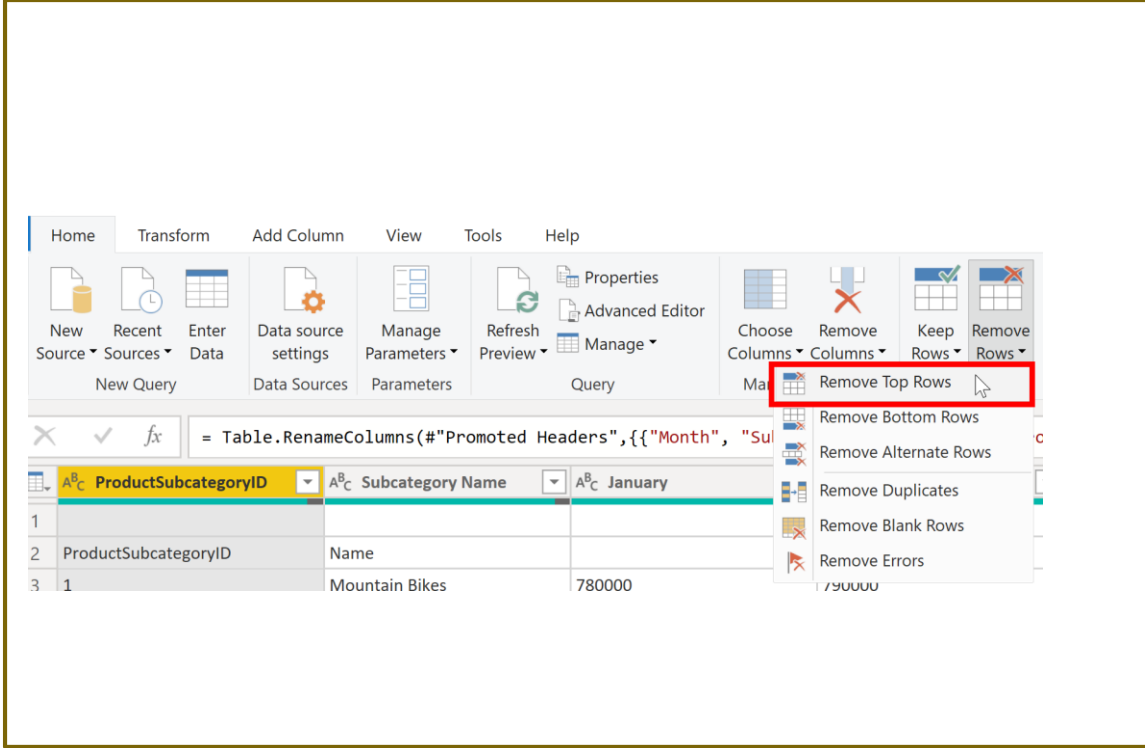
Identify column headers and names

Use Power Query Editor to clean up and shape data.

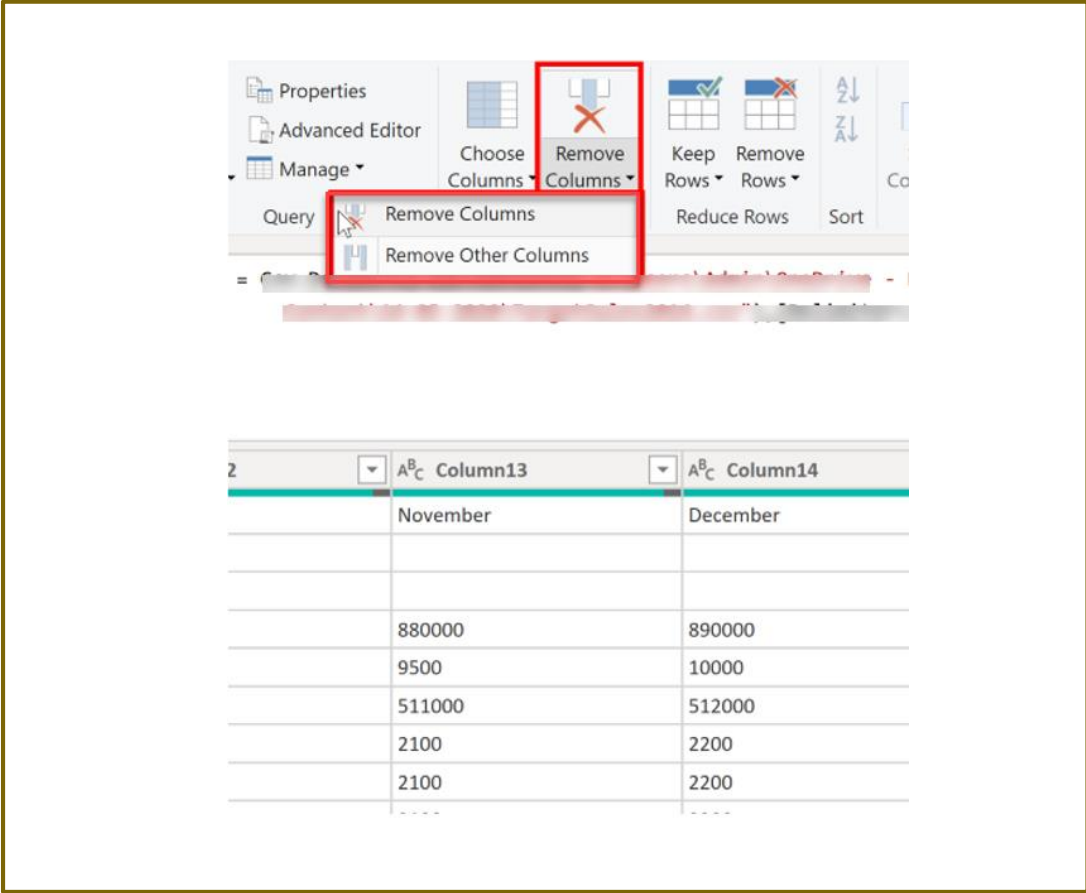


Shaping Table Structure

Shape the data to meet reporting needs.



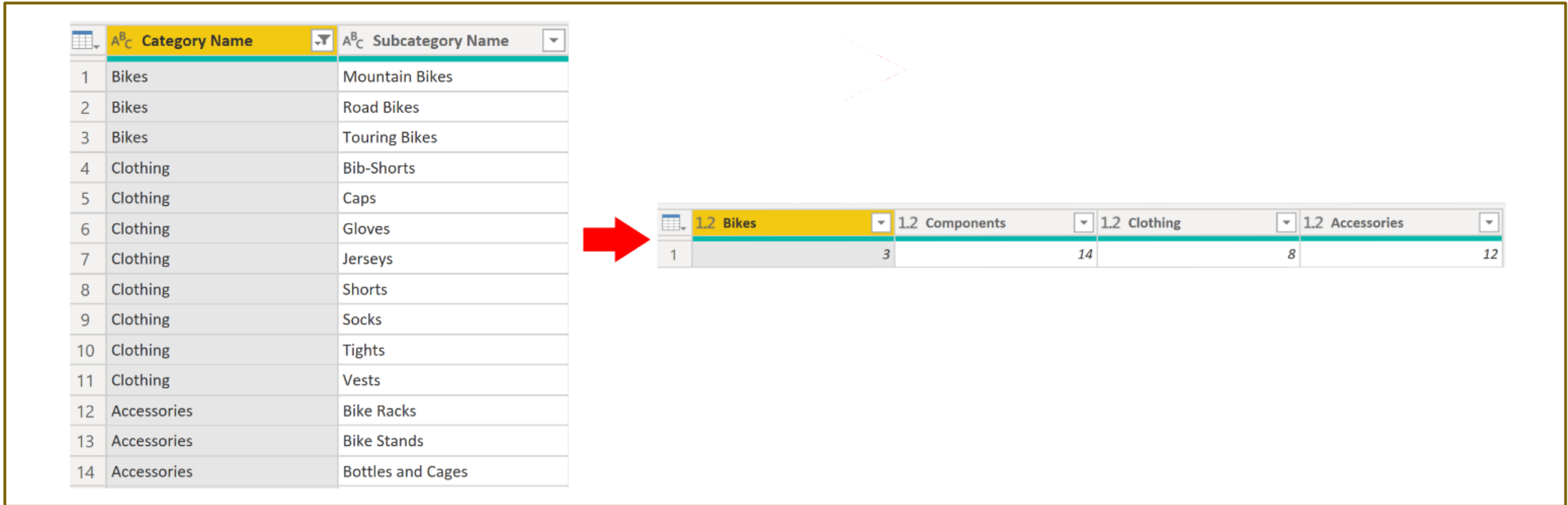
The screenshot shows the Power Query ribbon with the 'Remove Top Rows' option highlighted in a red box. The ribbon includes tabs for Home, Transform, Add Column, View, Tools, and Help. The 'Remove Top Rows' option is located in the 'Remove Rows' group. Below the ribbon, a data table is visible with columns 'ProductSubcategoryID', 'Subcategory Name', and 'January'. The formula bar shows a partial M formula: `= Table.RenameColumns("#Promoted Headers", {"Month", "Su`.



The screenshot shows the Power Query ribbon with the 'Remove Columns' and 'Remove Other Columns' options highlighted in red boxes. The ribbon includes tabs for Properties, Advanced Editor, Manage, Query, and Sort. The 'Remove Columns' option is located in the 'Remove Columns' group. Below the ribbon, a data table is visible with columns 'Column13' and 'Column14'. The table contains data for months like November and December, with values such as 880000, 890000, 9500, 10000, 511000, 512000, 2100, and 2200.

Unpivot and Pivot columns

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



The diagram illustrates the process of pivoting data. On the left, a table with two columns, 'Category Name' and 'Subcategory Name', is shown. A red arrow points from this table to a pivoted table on the right. The pivoted table has four columns: 'Bikes', 'Components', 'Clothing', and 'Accessories', and one row of data.

Category Name	Subcategory Name
Bikes	Mountain Bikes
Bikes	Road Bikes
Bikes	Touring Bikes
Clothing	Bib-Shorts
Clothing	Caps
Clothing	Gloves
Clothing	Jerseys
Clothing	Shorts
Clothing	Socks
Clothing	Tights
Clothing	Vests
Accessories	Bike Racks
Accessories	Bike Stands
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 2: Profiling the Data



Profiling Data and Examining Structures

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

The screenshot shows the Microsoft Power BI Desktop interface. The main area displays a data model with several tables: Assortment, Product, Customer, Sales, and Assortment Product. The Customer table is selected, and its properties are shown in the Properties pane on the right. The Properties pane includes fields for Name, Description, and Synonyms. The Synonyms field contains the value 'customer'.

The screenshot shows the Microsoft Power Query Editor interface. The main area displays a data preview of a query with columns for Latitude, Longitude, Total Excluding Tax, Total Including Tax, Profit, and SalesPerson. A red box highlights the Column statistics and Value distribution sections. The Column statistics section shows the following data:

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

The Value distribution section shows a horizontal bar chart with the following data:

Value	Count
Anthony Grosse	1000
Kayla Woodcock	1000
Sophia Hinton	1000
Empty	0
Archer Lambie	10
Taj Shand	0
Hudson Hollinworth	0
Amy Trefl	0
Jack Potter	0
Amy Trefl	0
Lily Code	0

Review Questions

Q01 – How many rows does Power Query scan to detect the type of data in the columns?

- A01 – 1000

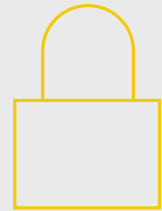
Q02 – Data profiling is defined as what?

- A02 – Studying the nuances of the data

Q03 – What is the risk of having null values in a numeric column?

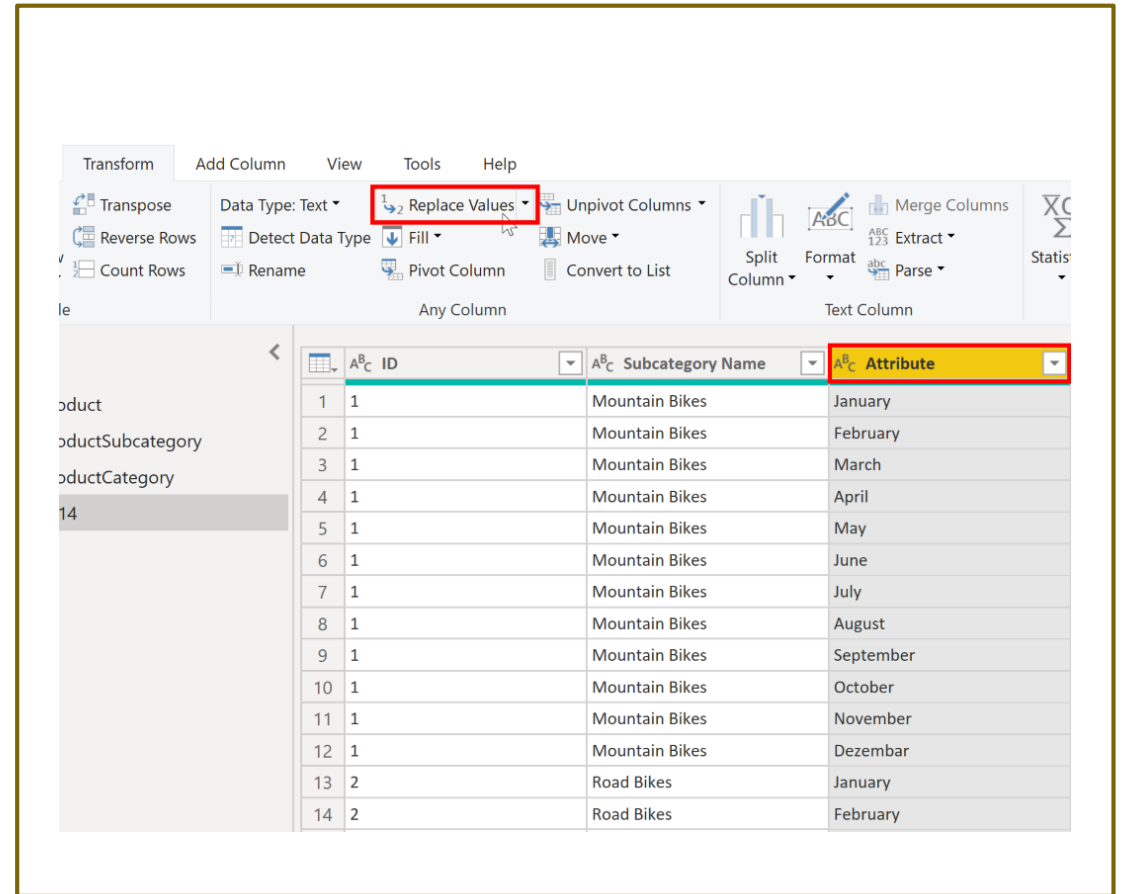
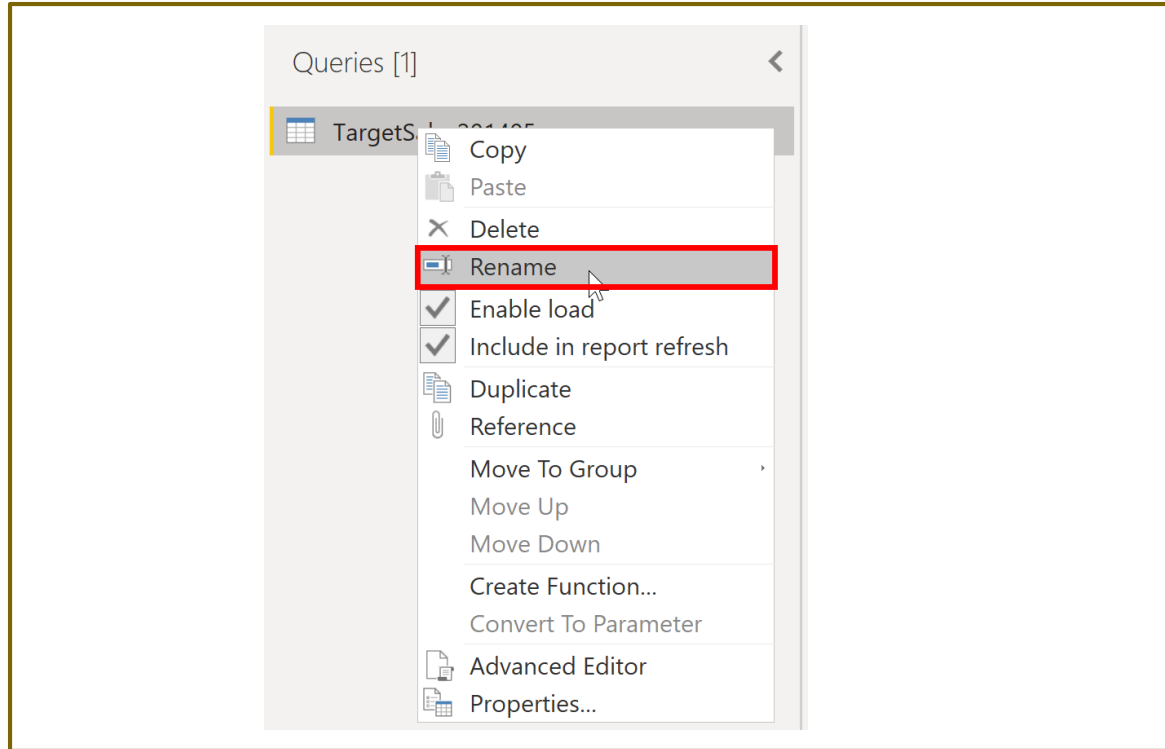
- A03 – DAX expressions that AVERAGE data will be incorrect.

Lesson 3: Enhance the Data Structure



Apply user-friendly value replacements

Make information user-friendly.



Evaluate and Change Column Data Types

	123 SalesOrderID	A8C OrderDate	A8C Sort_of_Sales	123 ProductID	123 OrderQty
1	52242	07/07/2013	Internet		870
2	52592	14/07/2013	Internet		870
3	52694	16/07/2013	Internet		870
4	52799	18/07/2013	Internet		870
5	53799	03/08/2013	Internet		870
6	54058	08/08/2013	Internet		870
7	54059	08/08/2013	Internet		870
8	54063	08/08/2013	Internet		870
9	54158	10/08/2013	Internet		870
10	54281	12/08/2013	Internet		870

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

Tables to append

- Production Suppliers
- Sales Customers
- HR Employees

Add >>

OK Cancel

Merge

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

Sales Orders

orderid	custid	empid	orderdate	requiredate	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

Review Questions

Q01 – What is not a best practice for naming conventions in Power BI?

- A01 – Abbreviated column names

Q02 – 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?

- A02 - Append

Lab: Load Data in Power BI Desktop

Lab: Load Data in Power BI
Desktop



Module Overview

We covered the following concepts:

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

References

PL-300 Clean, transform, and load data in Power BI

<https://docs.microsoft.com/en-us/learn/modules/clean-data-power-bi/>

