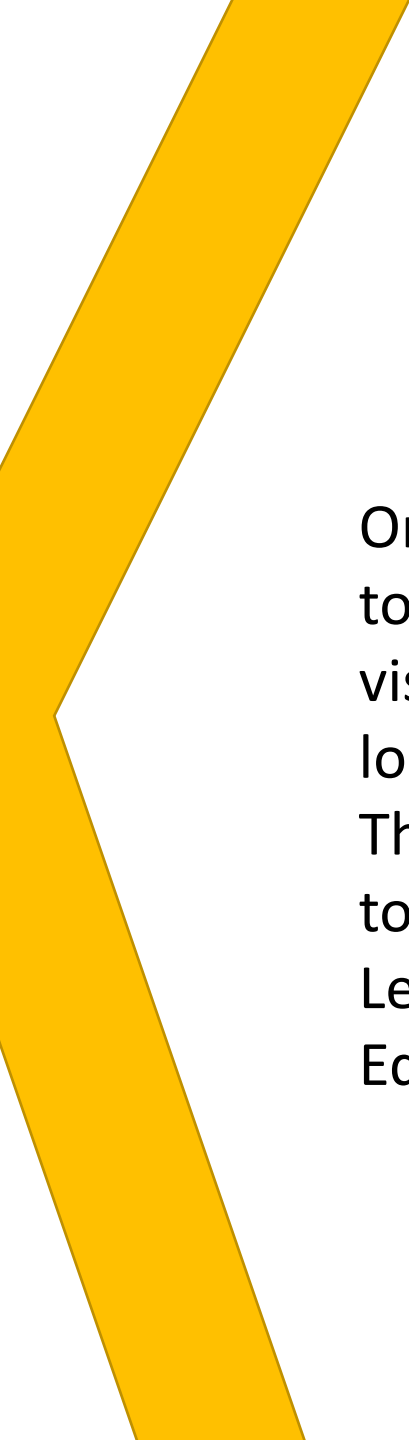


MICROSOFT POWER-BI

SERIES 3


DATA TRANSFORMATION IN POWER QUERY

- MAYURI .D.

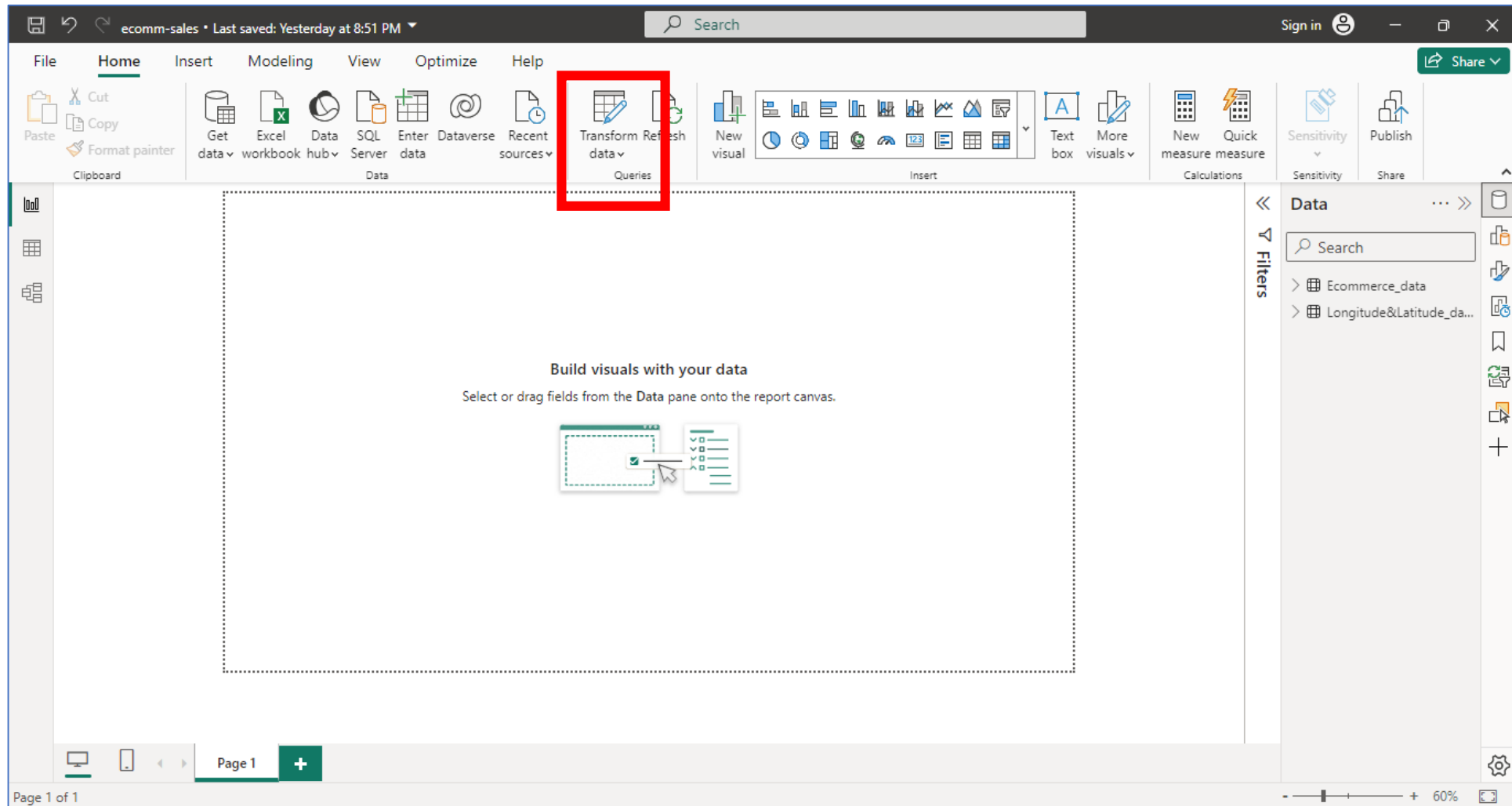
A thick yellow diagonal bar runs from the top-left corner towards the bottom-left corner of the slide.

Once you've connected to your data source in Power BI, the next step is to transform the data into a format that can be easily analyzed and visualized. This process is often known as ETL (extract, transform, load).

The Power Query Editor provides a user-friendly interface that allows users to perform a series of data transformations through a set of tabs. Lets go through some basic data transformation steps in Power Query Editor---

A yellow semi-circle is located in the bottom-right corner of the slide.

To open Power Query Editor, click on “Transform data” under Home tab.



1. FORMAT TABLE HEADER

Sometimes, power query include headers in data as a first row so to set that first row as header click on “Use First Row as Headers” under Transform tab.

The screenshot shows the Microsoft Power Query Editor interface. The 'Transform' tab is active, and the 'Use First Row as Headers' option is highlighted with a red box. The data table below has the following columns and data:

	customer_id	customer_first_name	customer_last_name	category_name	product_name
1	C_ID_45866	Mary	Fuller	Office Supplies	Xerox 1913
2	C_ID_44932	Alan	Edelman	Office Supplies	#6 3/4 Gummed Flap White Envelope
3	C_ID_70880	Mary	Gauman	Office Supplies	Belkin 8 Outlet Surge Protector

2. CHANGING DATA TYPE OF COLUMN

Check the data type of column shown beside column name. If found any wrong data type select particular column, click on “Detect Data Type” under Transform tab to automatically set data type OR click on data type beside column name and select respected type.

The screenshot displays the Microsoft Excel interface, specifically the 'Transform' tab. The 'Detect Data Type' button is highlighted with a red box. Below the ribbon, the 'Queries [2]' pane shows 'Ecommerce_data' and 'Longitude&Latitude_data'. The main data table has columns: 'customer_id', 'customer_first_name', 'customer_last_name', 'category_name', and 'product_name'. The 'customer_id' column is selected, and a dropdown menu is open, showing various data types like 'Decimal Number', 'Fixed decimal number', 'Whole Number', 'Percentage', 'Date/Time', 'Date', 'Time', 'Date/Time/Timezone', 'Duration', 'Text', 'True/False', 'Binary', and 'Using Locale...'. The 'Text' option is highlighted with a red box. The formula bar shows the formula: `= Table.TransformColumnTypes("#Promoted Headers",{{"customer_id", type text}}, {"customer_first_name",`

	customer_id	customer_first_name	customer_last_name	category_name	product_name
1	C_ID_45866			Office Supplies	Xerox 1913
2	C_ID_44932			Office Supplies	#6 3/4 Gummed Flap
3	C_ID_70880			Office Supplies	Belkin 8 Outlet Surge
4	C_ID_33157			Office Supplies	GBC VeloBinder Manu
5	C_ID_58303			Furniture	Eldon Pizzaz Desk Acc
6	C_ID_34556			Office Supplies	Hoover Shoulder Vac
7	C_ID_56882			Office Supplies	Rogers Handheld Barr
8	C_ID_51638			Office Supplies	Dixon Prang Watercol
9	C_ID_60578			Office Supplies	Ibico Hi-Tech Manual

3. RENAMING COLUMN NAME

Columns from raw data can be difficult to read or meaningless. Renaming the columns in your query to a meaningful name will make it easier for you to understand your data. Either double-click on column name and rename it OR simply select the column and click on “Rename” under Transform tab.

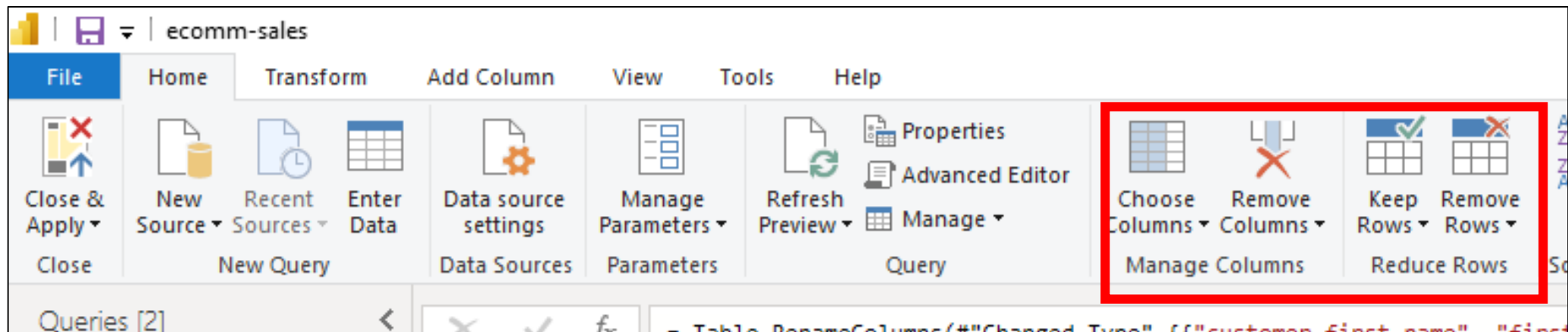
The screenshot displays the Power Query Editor interface for a query named 'ecommerce-sales'. The 'Transform' tab is active, showing various data transformation options. The 'Rename' button is highlighted with a red box. The formula bar shows the following formula: `= Table.RenameColumns(#"Changed Type", {"customer_first_name", "first_name"})`. The column headers in the data view are 'customer_id', 'first_name', 'customer_last_name', 'category_name', and 'product_name'. The 'customer_last_name' column is highlighted with a red box. The data view shows a table with columns and their respective data types and validity percentages.

customer_id	first_name	customer_last_name	category_name	product_name
Valid 100%	Valid 100%	Valid 100%	Valid 100%	Valid 100%
Error 0%	Error 0%	Error 0%	Error 0%	Error 0%
Empty 0%	Empty 0%	Empty 0%	Empty 0%	Empty 0%

4. REMOVE COLUMNS/ROWS

Use this transformation to clean up your dataset by removing unnecessary or redundant columns, which can improve query and report performance.

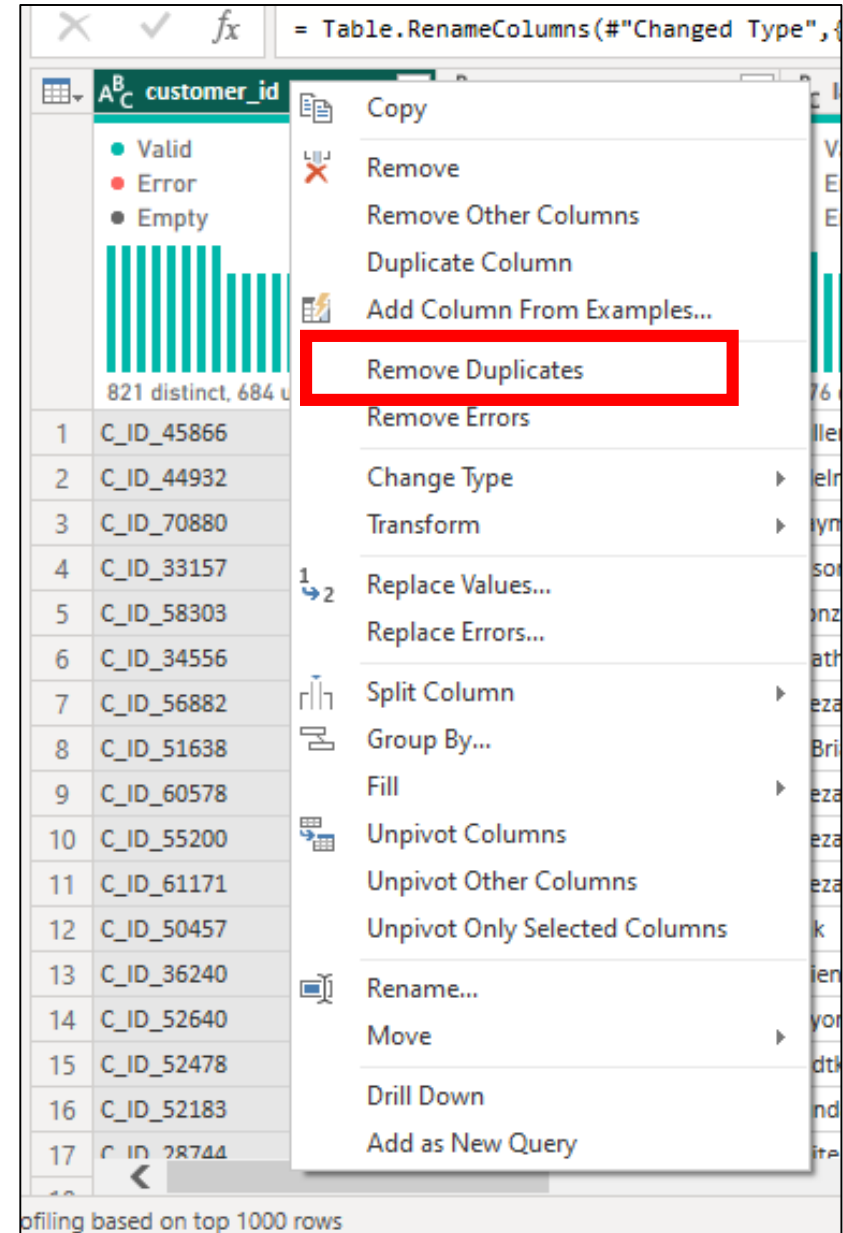
Simply select the column OR row and click on “Remove column” or “Remove row” under Home tab.



5. REMOVE DUPLICATES

It allows you to eliminate duplicate rows from your dataset based on the values in one or more columns. It helps you maintain data integrity by keeping only unique Records.

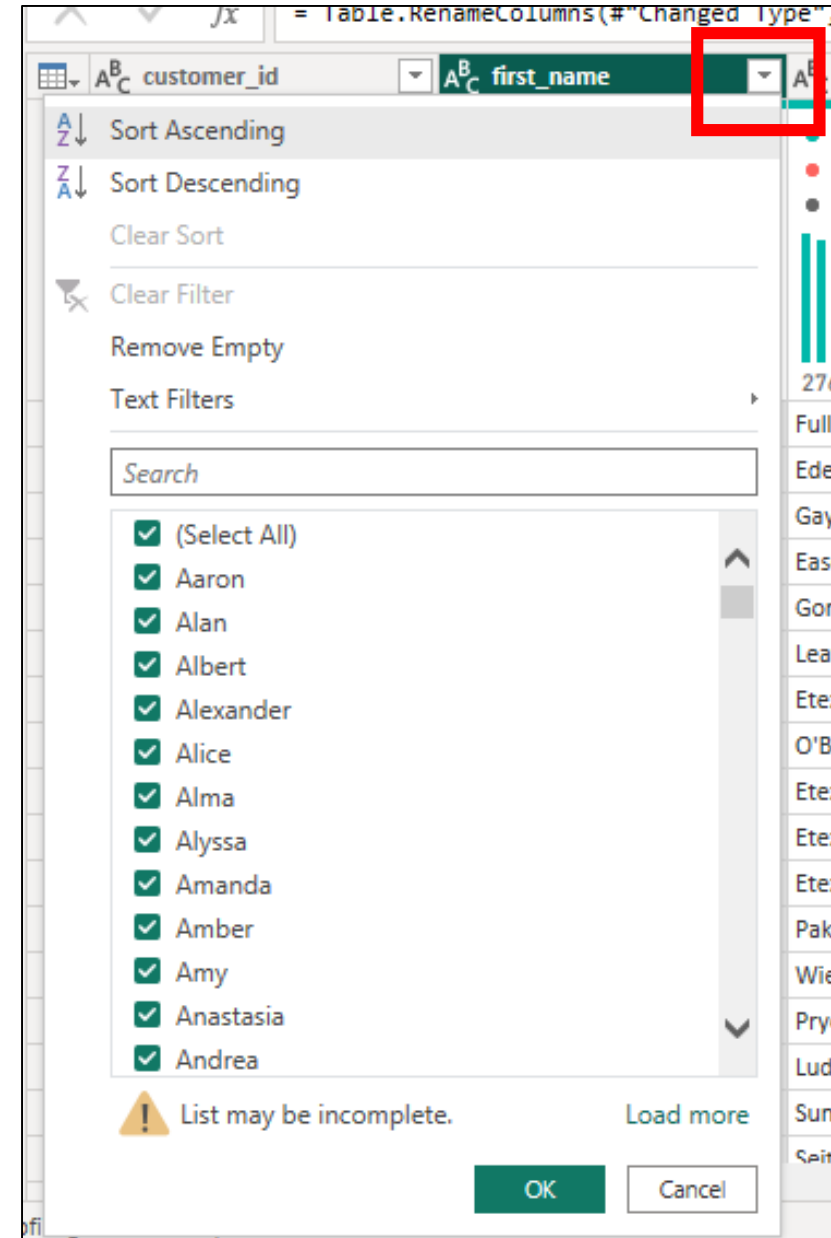
Simply right-click on the column and click on “Remove duplicates”.



6. FILTER ROWS

It enables you to include or exclude rows from your dataset based on specific criteria or conditions.

Click on the filter arrow besides column name and check the distinct values- if there are any misspell words, formats, missing values, etc



7. MERGE COLUMN

It lets you combine the values from multiple columns into a single new column. You can choose the delimiter or separator to use between merged values.

Select 2 columns that you need to concatenate and click on “Merge columns” under Transform tab.

The screenshot shows the 'Merge Columns' dialog box in the Power Query interface. The dialog is titled 'Merge Columns' and has a 'Text Column' section. The 'Delimiter' is set to 'Space' and the 'Separator' is set to 'None'. The 'Columns to merge' list contains 'first_name' and 'last_name'. The 'Merge Columns' button is highlighted with a red box.

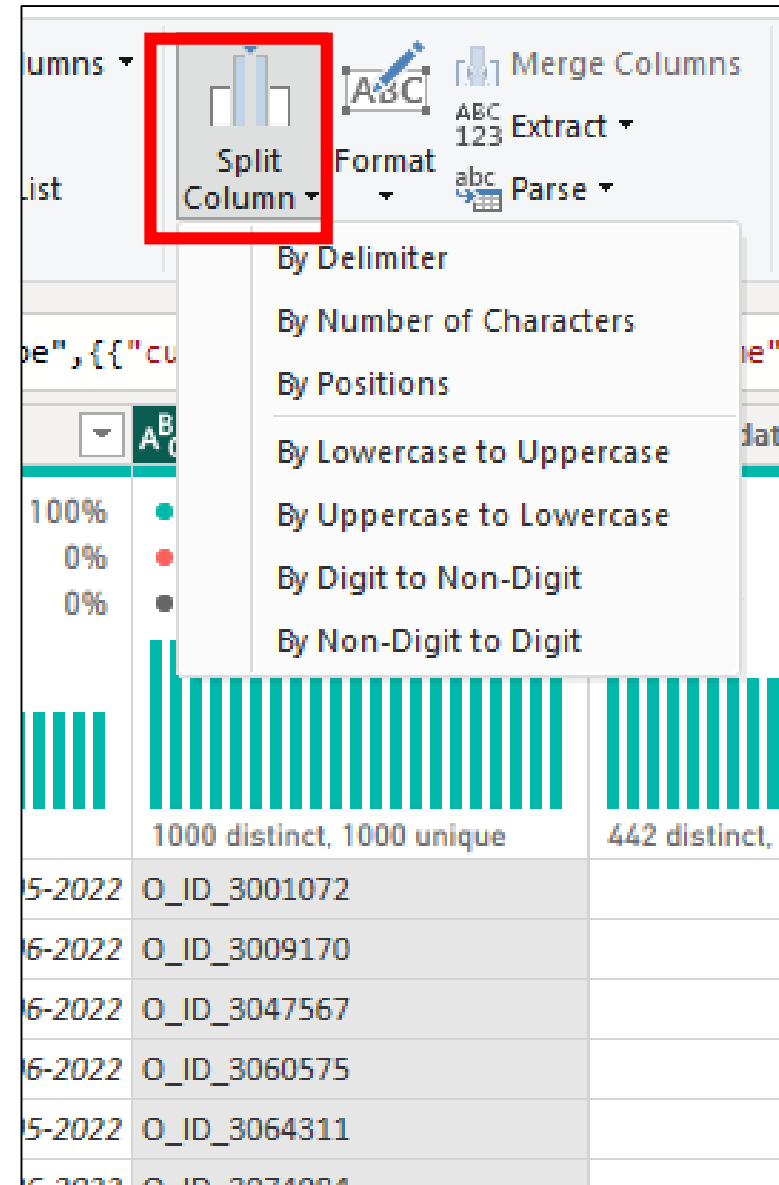
Below the dialog, the resulting data table is shown. The table has three columns: 'first_name', 'last_name', and 'office'. The 'first_name' column has 232 distinct values and 93 unique values. The 'last_name' column has 276 distinct values and 89 unique values. The 'office' column has 3 distinct values. The table contains 8 rows of data.

first_name	last_name	office
Mary	Fuller	Office
Alan	Edelman	Office
Mary	Gayman	Office
Raymond	Eason	Office
Mary	Gonzalez	Furn
Joe	Leatherbury	Office
Mary	Etezadi	Office

8. SPLIT COLUMN

It allows you to divide a single column into multiple columns based on a specified delimiter or pattern.

Select the column and click on “split column” and select the respective option as per selected column.



9. FORMAT COLUMN

It enables you to standardized the data like trim, lowercase, uppercase, etc

Select the column and click on “format” and select respected option.

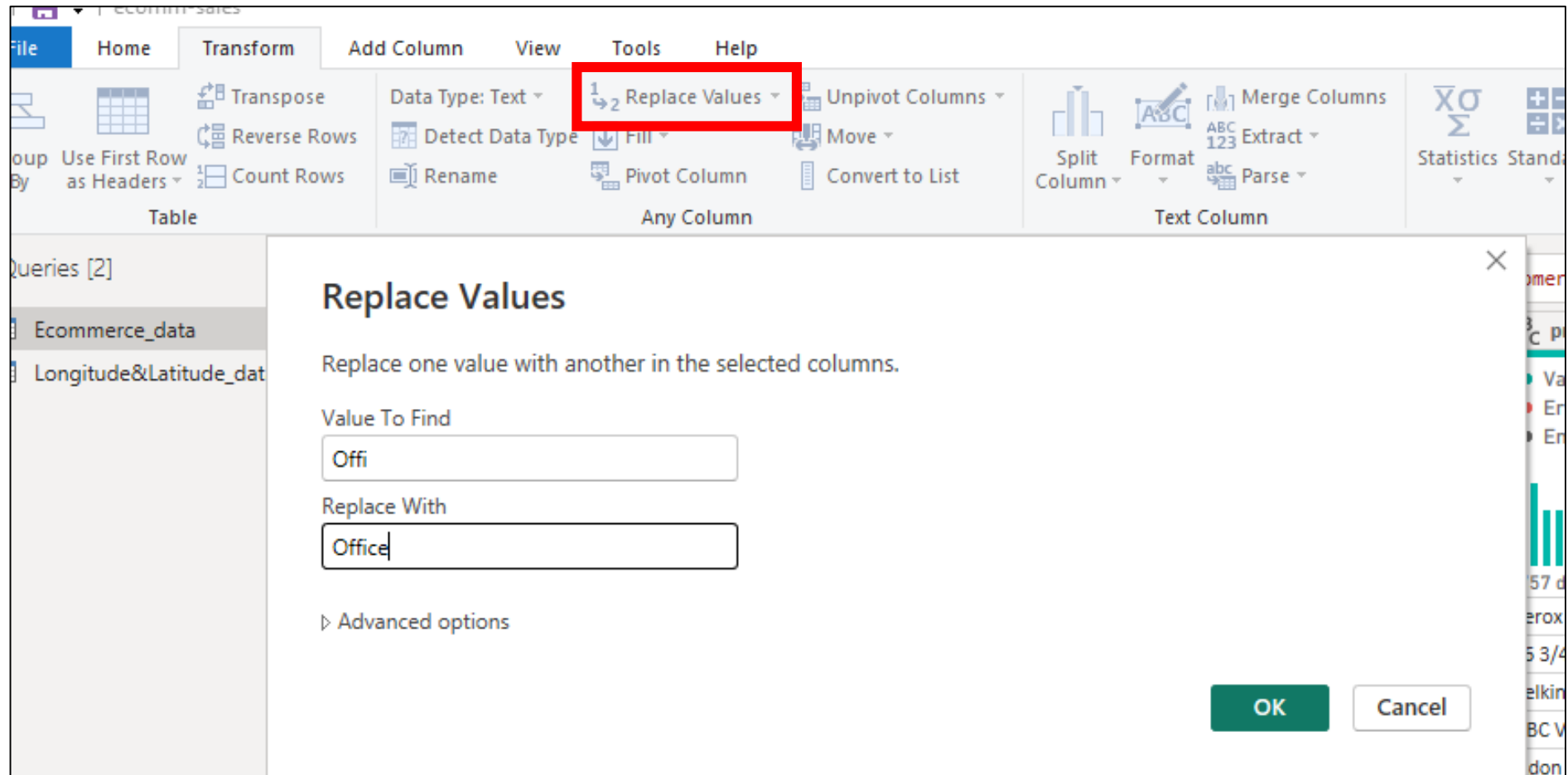
The screenshot shows a data tool interface. At the top, there's a 'Columns' dropdown and a 'Split Column' button. A red box highlights the 'Format' button, which has a dropdown menu open. The menu options are: lowercase, UPPERCASE, Capitalize Each Word, Trim, Clean, Add Prefix, and Add Suffix. Below the menu, a table is visible with a column named 'last_name'. The table has a legend for 'Valid' (green dot), 'Error' (red dot), and 'Empty' (grey dot). Below the legend, there are two bar charts. The first chart shows '276 distinct, 89 unique' values. The second chart shows '3 distinct, 0 unique' values. The table rows show 'Fuller' and 'Edelman' in the first column, and 'Office Supplies' in the second column.

last_name	
Fuller	Office Supplies
Edelman	Office Supplies

10. REPLACE VALUES

It enables you to format misspell words.

Click on “Replace values” under Transform tab and specify the required words.

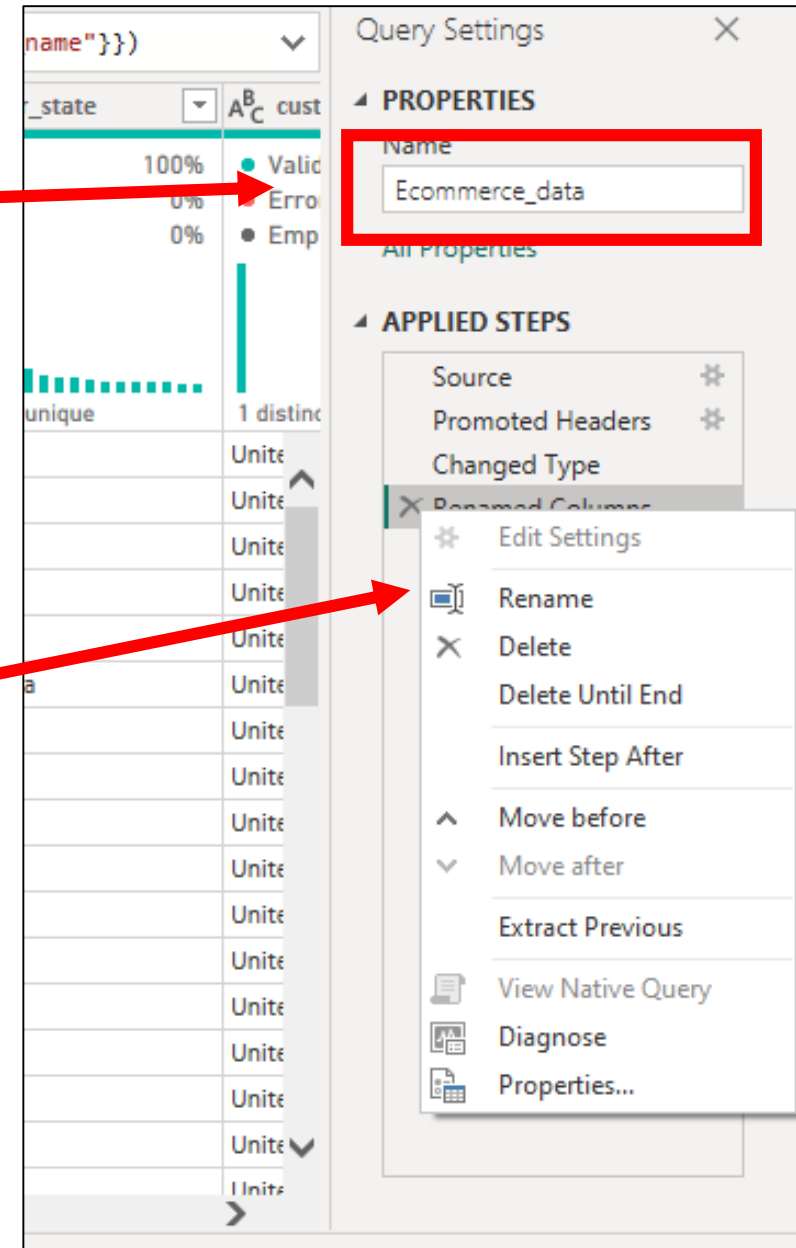


11. EDIT TABLE NAME AND RENAME APPLIED STEPS

You can edit your table name here.

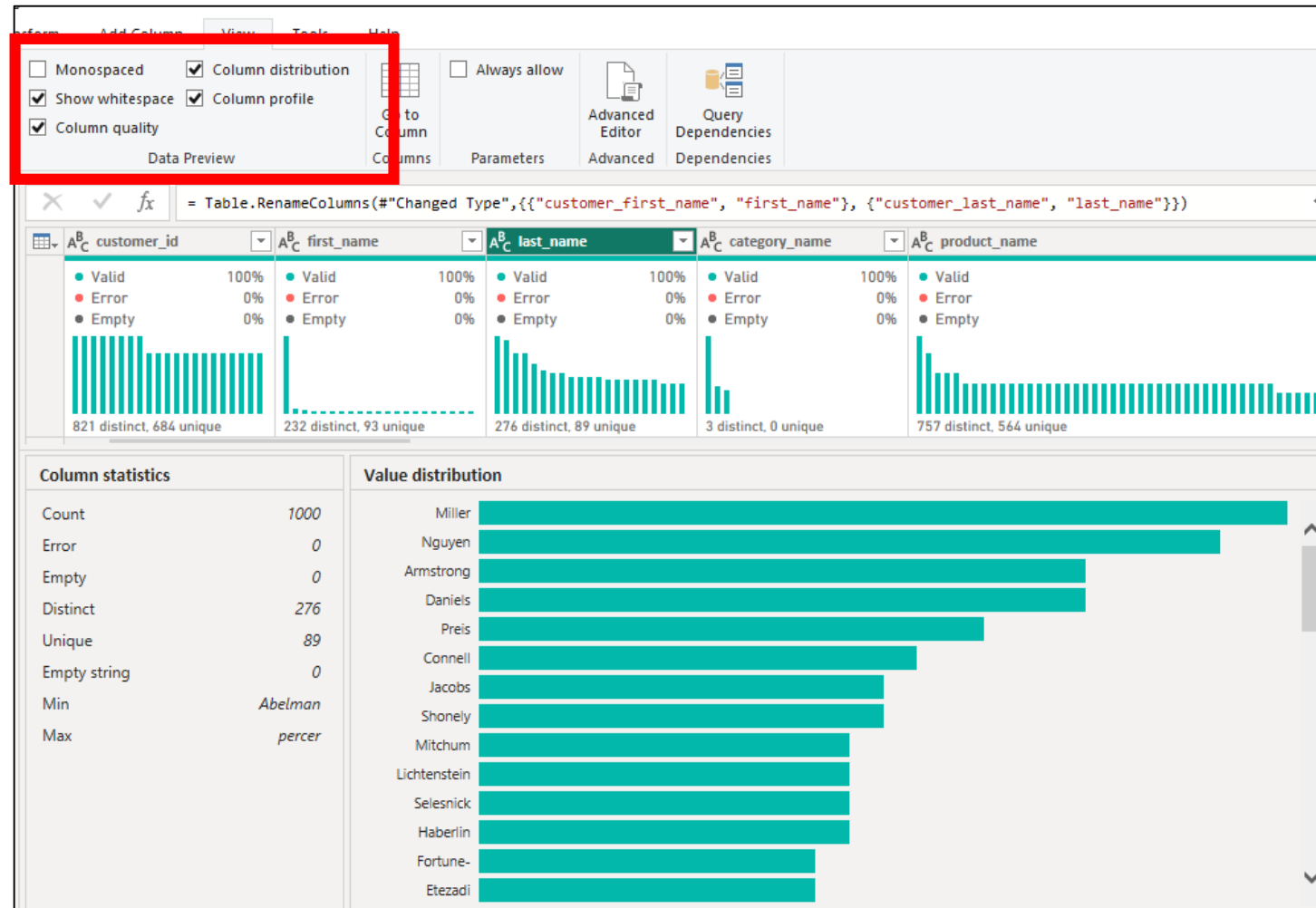
Rename the steps that you performed while transforming the data so that when you revisit the project after days you know what steps were performed.

Right-click on particular step and select rename.



12. DATA PREVIEW

While working on large datasets, it is necessary to get an basic idea of the data in a seconds, for that turn ON following options under View tab – Column distribution, Column profile, Column quality.



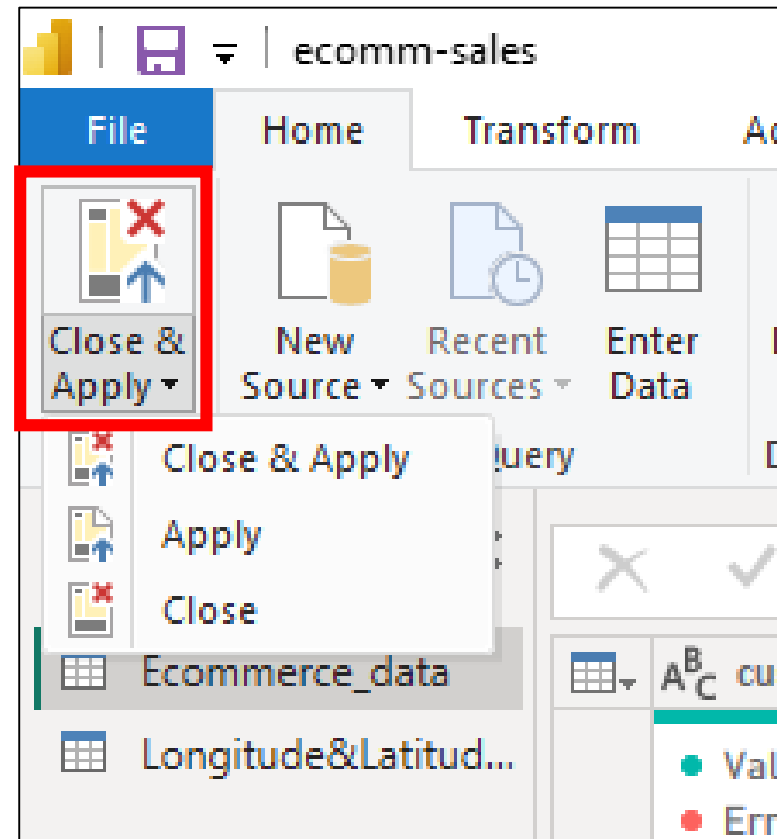
LOADING DATA TO POWER-BI DESKTOP FROM POWER QUERY EDITOR

After transforming your data in Power Query, you can load it into Power BI Desktop to create your reports and visualizations.

Under home tab Click on “close & Apply” to save changes and close power query.

Selecting “Apply ” will just save the changes.

Selecting “close” will close the power query without saving the changes.





THANK YOU

- MAYURI .D.