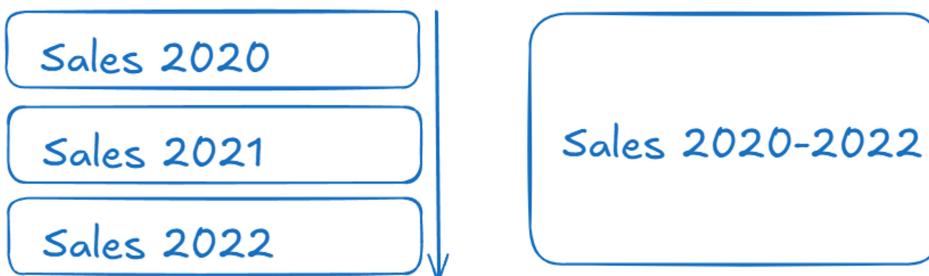
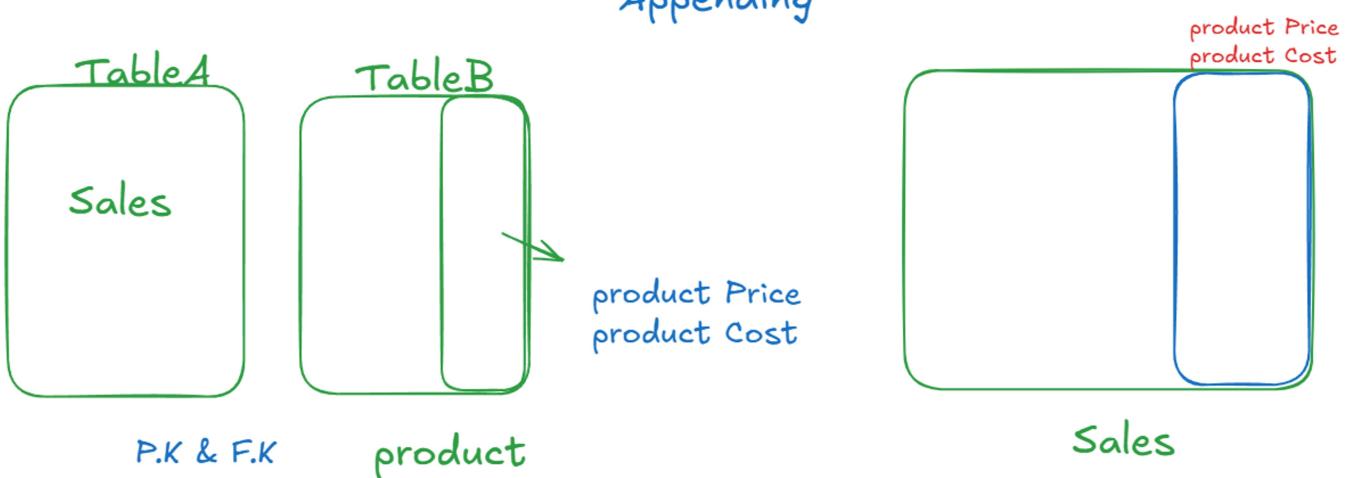


Power BI - Data Transformation - p1

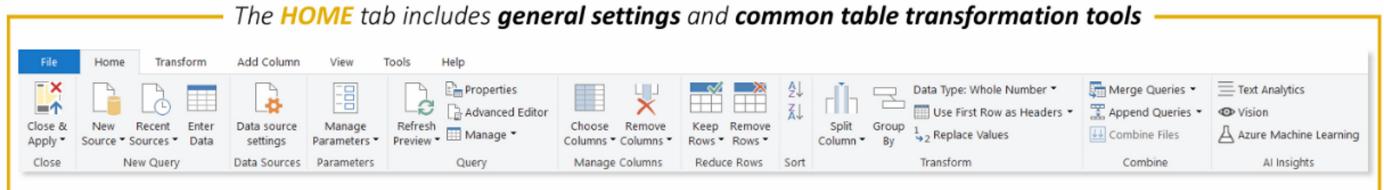


In order to append multiple table they should follow same structure & having same format.

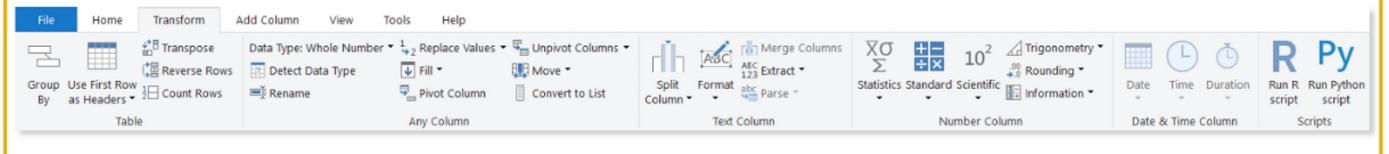
Same Structure to Merge



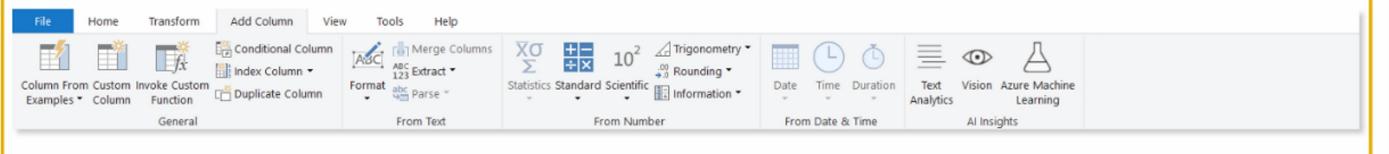
QUERY EDITING TOOLS



The **TRANSFORM** tab includes tools to **modify existing columns** (splitting/grouping, transposing, extracting text, etc.)



The **ADD COLUMN** tools **create new columns** (based on conditional rules, text operations, calculations, dates, etc.)



BASIC TABLE TRANSFORMATIONS

Sort values (A-Z, Low-High, etc.)

Change data type (date, \$, %, text, etc.)

Promote headers

Choose or remove columns

Tip: use the "Remove Other Columns" option if you always want a specific set

Remove Columns

Remove Other Columns

Keep or remove rows

Tip: use "Remove Duplicates" to create new lookup tables from scratch

Remove Top Rows

Remove Bottom Rows

Remove Alternate Rows

Remove Duplicates

Remove Blank Rows

Remove Errors

Duplicate, move or rename columns

Tip: Right-click column headers to access common tools

Transform ribbon tools

Manage Columns

Remove Rows

Sort

Data Type

Use First Row as Headers

Replace Values

Split Column

Group By

Merge Queries

Append Queries

Combine Files

Text Analytics

Vision

Azure Machine Learning

AI Insights

Order context menu

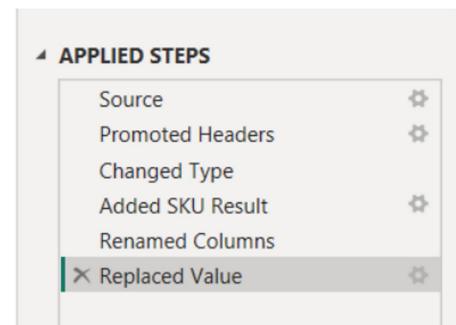
- Copy
- Remove
- Remove Other Columns
- Duplicate Column
- Add Column From Examples...
- Remove Duplicates
- Remove Errors
- Change Type
- Transform
- Replace Values...
- Replace Errors...
- Group By...
- Fill
- Unpivot Columns
- Unpivot Other Columns
- Unpivot Only Selected Columns
- Rename...
- Move
- Drill Down
- Add as New Query

Assignment : Transforming Data

1. Create queries to connect to the two new .csv files
2. Name your queries Product Category Lookup and Product Subcategory Lookup
3. Confirm that column headers have been promoted and that all data types are correct
4. Add a new column to extract all characters before the dash ("") in the Product SKU column, and name it "SKU Type"
5. Update the SKU Type calculation above to return all characters before second dash, instead of the first
6. Replace zeros (0) in the Product Style column with "NA"
7. Close and load to your data model



HL-U509-R



Text Before Delimiter

Enter the delimiter that marks the end of what you would like to extract.

Delimiter

Advanced options

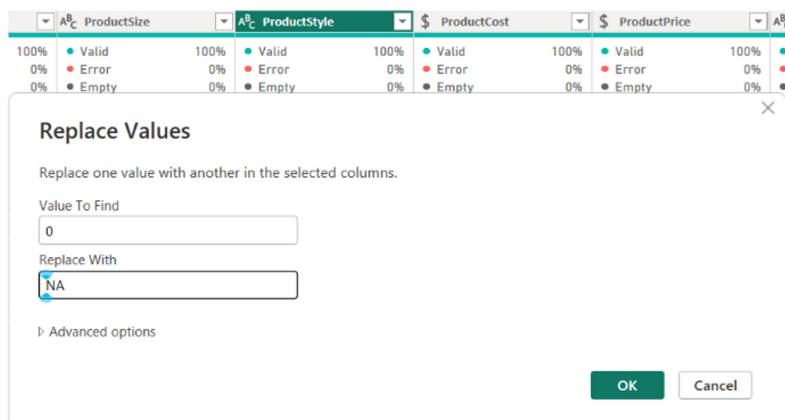
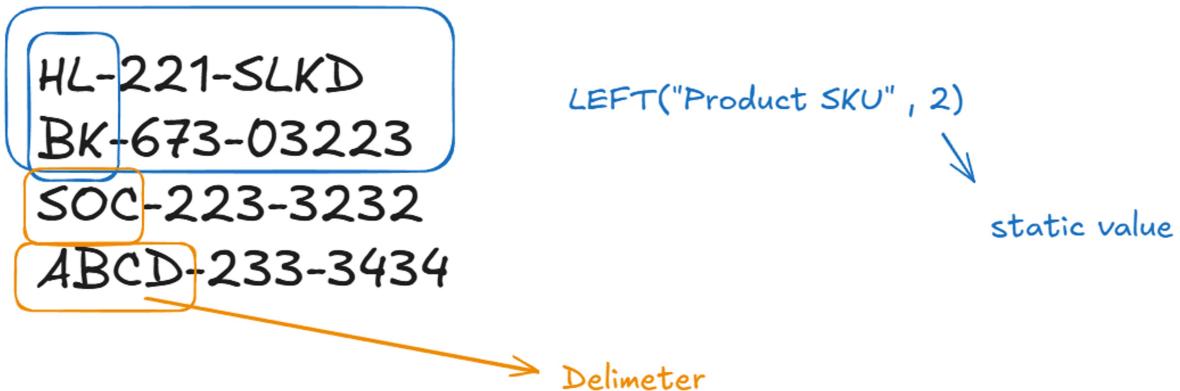
Scan for the delimiter

From the start of the input

Number of delimiters to skip (1)

skipping the first occurrence of "-" 

- Step 1 : Select the column you need to fetch the SKU Type.
 Step 2 : Option Add Column > Extract > Text Before Delimiter
 Step 3 : Insert "-" in delimiter and click ok.
 Step 4 : Rename the new column to SKU Type.



1	8	3	1	5
9	2	6	4	2
7	3	4	5	1

- 1 - 4 times
 2 - 2 times
 3 - 2 times
 4 - 2 times
 5 - 3 times
 6 - 2 times
 7 - 1 time
 8 - 1 time
 9 - 1 time

1 or more
 distinct
 count = 9

1 only
 Unique
 count = 3

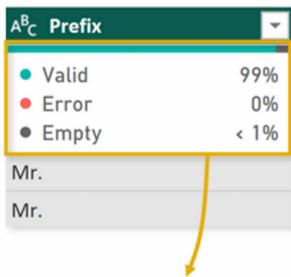
EXTRACTING DATA FROM THE WEB



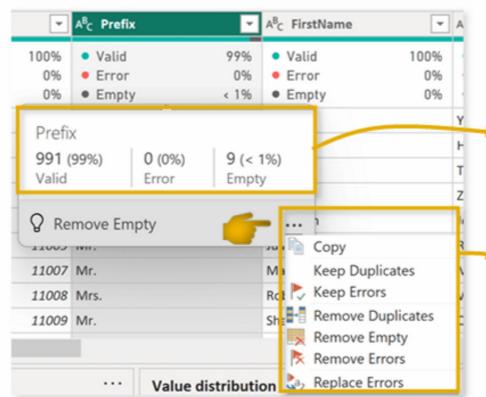
- Power Query includes a native Web connector for importing web-hosted files (csv, xlsx, etc.) or scraping URLs for anything that Power Query can identify as a structured table.

DATA PROFILING: COLUMN QUALITY

Profiling tools like column quality, column distribution, and column profile allow you to explore the quality, composition, and distribution of your data before loading it into the Power BI front-end.



Column quality shows the percentage of values within a column that are valid, contain errors, or are empty



Hover over the column quality box to see the number of records in each category

Click the options menu to remove duplicates, errors or empty values

Profiling tools are a great way to quickly find and address common data quality issues in one place, instead of having to manually apply multiple tools or filters

COLUMN DISTRIBUTION

Column distribution provides a sample distribution of the data in a column

The screenshot shows two columns in a Power BI data view:

- Product Cost:** A histogram showing 108 distinct values and 47 unique values. The average is 13.09 and the minimum is 12.03.
- Order Date:** A list of dates from 1/1/2022 to 10/21/2022. It shows 911 distinct values and 3 unique values. A tooltip indicates 911 (1%) Distinct and 3 (< 1%) Unique.

Annotations provide additional details:

- An arrow points to the Product Cost histogram with the text: "Suggested action based on column distribution results".
- A callout box over the Order Date list says: "Hover over the column quality box to see the number of distinct & unique records".
- A callout box over the options menu for the Order Date list says: "click the options menu to remove duplicates, errors or empty values".

COLUMN PROFILE

Column statistics provide more detailed profiling metrics, including:

Count = 293
(total number of values in column)

Distinct Count = 119
(total number of distinct values, whether they appear once or multiple times)

Unique = 82
(total number of values that appear exactly once)

Min & Max
(lowest and highest observed values)
Note: Typically only useful for numerical values

Column profile provides a more holistic view of the data in a column, including a sample distribution and profiling statistics

The screenshot shows a column profile for the **Model Name** column:

- Column statistics:** Shows a count of 293, 119 distinct values, and 82 unique values.
- Value distribution:** A bar chart showing the frequency of different model names. The top values are LL Road Frame (Road-650), Mountain-500, Touring-3000, LL Touring Frame, LL Mountain Frame, HL Mountain Frame, HL Road Frame (Road-250), Mountain-100, Touring-1000, HL Touring Frame (Mountain-200), ML Road Frame (ML Road Frame-W), ML Mountain Frame-W (Road-450), and Road-150.
- Transformation context:** A tooltip for the first bar (LL Road Frame) shows suggested transformations: Equals, Does Not Equal, Copy, Text Filters, and Replace Values... Hovering over the bars provides these options.

TEXT TOOLS

The screenshot shows the Power BI ribbon with the 'Transform' tab selected. A green callout box highlights the 'Text Column' section under the 'Tools' dropdown. Three arrows point from this callout to three separate boxes containing specific transformation options:

- Left Box (Split Column):** Options include 'By Delimiter', 'By Number of Characters', 'By Positions', 'By Lowercase to Uppercase', 'By Uppercase to Lowercase', 'By Digit to Non-Digit', and 'By Non-Digit to Digit'.
Description: Split a text column based on a specific delimiter, number of characters, or other attributes.
- Middle Box (Format Column):** Options include 'lowercase', 'UPPERCASE', 'Capitalize Each Word', 'Trim', 'Clean', 'Add Prefix', and 'Add Suffix'.
Description: Format a text column to upper, lower or proper case, or add a prefix or suffix.
- Right Box (Text Column):** Options include 'Length', 'First Characters', 'Last Characters', 'Range', 'Text Before Delimiter', 'Text After Delimiter', and 'Text Between Delimiters'.
Description: Extract characters from text based on fixed lengths, first/last characters, ranges or delimiters.

Tip: Use "Trim" to eliminate leading & trailing spaces, or "Clean" to remove non-printable characters

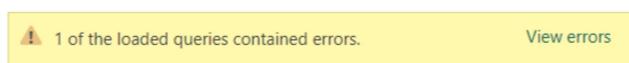
You can access many tools from both the "Transform" and "Add Column" menus -the difference is whether you want to ADD a new column or OVERWRITE an existing one

Text Tool : Assignment

1. Duplicate the email address column and name it "Domain Name"
2. In the new column, remove all text/characters except for the domain name
3. Use transformation steps to clean up and capitalize the domain names (i.e. "Adventure Works")
4. Save & Apply changes

If you find error while loading customer lookup back to front end. This is changing data type error.

Load



Customer Lookup
18,154 rows loaded. 5 errors.

How you can resolve???

- Select the customer key , right click.
- Click on option remove error.

1.

CustomerKey	EmailAddress	Domain Name
1000 distinct, 100	1000 distinct, 100	1000 distinct, 1000 unique
jon24@adventure	jon24@adventure	jon24@adventure-works.com
eugene10@adven	eugene10@adven	eugene10@adventure-works.com
ruben35@advent	ruben35@advent	ruben35@adventure-works.com
christy12@adven	christy12@adven	
elizabeth5@adve	elizabeth5@adve	
julio1@adventure	julio1@adventure	
marco14@advent	marco14@advent	
rob4@adventure-	rob4@adventure-	
shannon38@adve	shannon38@adve	
jacquelyn20@adv	jacquelyn20@adv	
curtis9@adventur	curtis9@adventur	
lauren41@advent	lauren41@advent	
ian47@adventure	ian47@adventure	
sydney23@adven	sydney23@adven	
chloe23@adventu	chloe23@adventu	

jon24@adventure-works.com

username@domainName

Extract - Text after "@"

2.

Text After Delimiter

Enter the delimiter that marks the beginning of what you would like to extract.

Delimiter
@

Advanced options

OK Cancel

adventure-works.com

this also needs to be deleted

Text Before Delimiter

Enter the delimiter that marks the end of what you would like to extract.

Delimiter
-

Advanced options

adventure-works

Replace - "-" with " " [space]

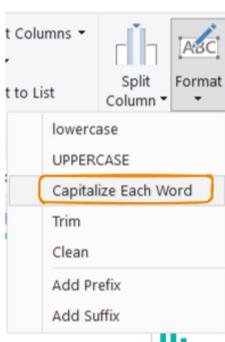
Replace Values

Replace one value with another in the selected columns.

Value To Find
-
Replace With

adventureworks

3.



Adventure Works

4. Close & Apply.