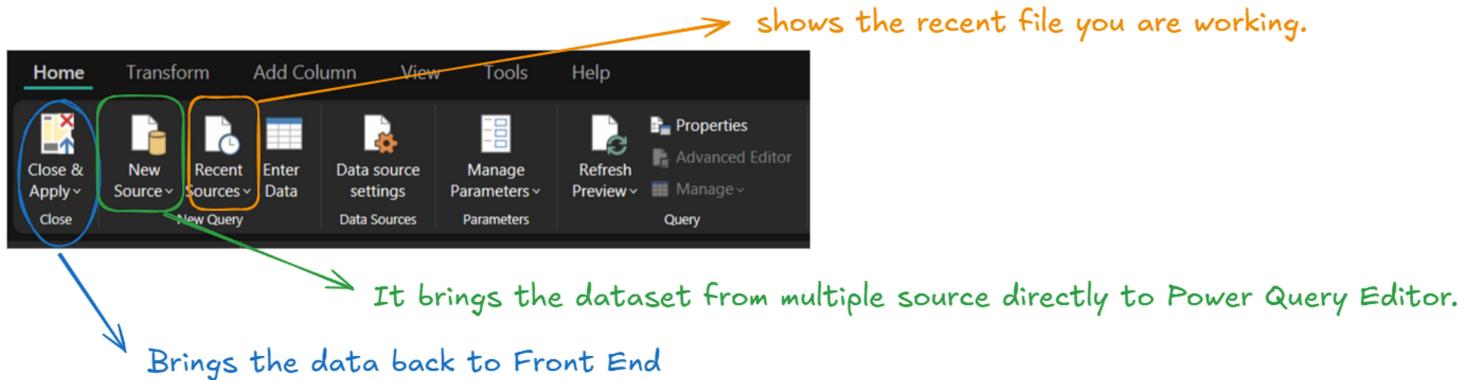


Data Cleaning with Power BI - p3

- Understanding all the features of Power BI.



	Name	Age	Email	Country
1	Krishna Madan	35	krishnamadan77@gmail.com	India
2	Vaishnavi Niljikar	24	vaish.niljikar@gmail.com	India
3	Naval Mandhana	26	naval45@gmail.com	India
4	Kesavan	35	kesavanplace@gmail.com	India
5	Mukund	35	mukund.potdar89@outlook.com	India
6	Rashmi Ranjan	25	rashmiranjan04@gmail.com	India
7	Nirvisha	26	nirvisha99@gmail.com	India

Query Properties

Name: Students Info

Description:

Enable load to report

Include in report refresh

OK Cancel

If, you uncheck the enable load to report, Basically it won't load the dataset to the frontEnd.

Data

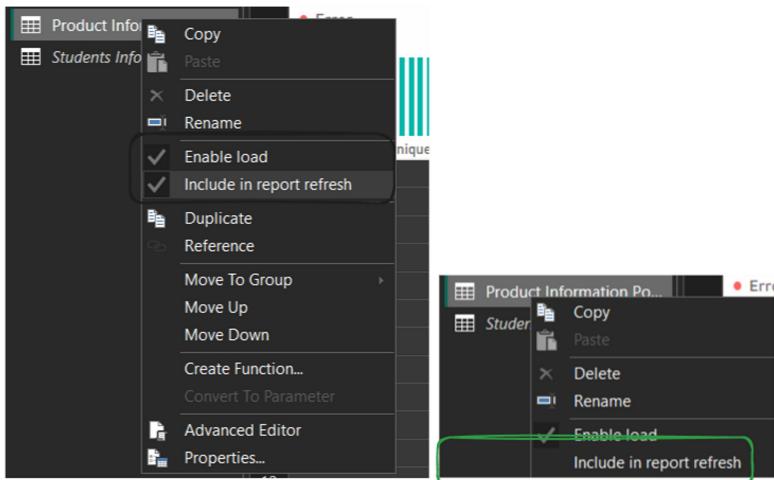
... >

Search

> Customers

> Product Information P...

> Sales



Advanced Editor

Customers

```

let
    Source = Excel.Workbook(File.Contents("C:\Users\krish\Downloads\Sales Dataset Power BI (1).xlsx"), null, true),
    Customers_Sheet = Source[[Item="Customers", Kind="Sheet"]][Data],
    #"Promoted Headers" = Table.PromoteHeaders(Customers_Sheet, [PromoteAllScalars=true]),
    #"Changed Type" = Table.TransformColumnTypes(#"Promoted Headers",{{"Customer ID", Int64.Type}, {"Customer Name", type text}})
in
    #"Changed Type"

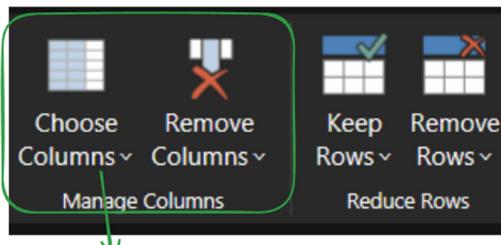
```

M-Code

APPLIED STEPS

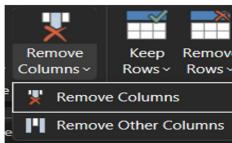
- Source
- Navigation
- Promoted Headers
- Changed Type**

M-Code - Its a system generated language used in Back end of Power BI, which is Power Query Editor, and builds the language on behalf of the applied steps.



Remove the other columns which are not picked.

Select the selected columns, mostly useful when we have a large set of columns.



Removes the selected column.
Removes the unselected column.

Keep Top Rows

Specify how many rows to keep.

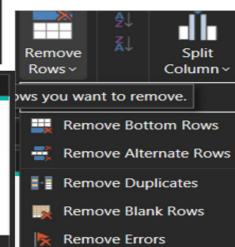
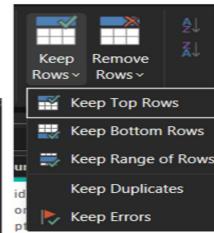
Number of rows

5

OK

Cancel

	A _B C Name	T ₁ ₂ Age	A _B C Email	A _B C Country
	<ul style="list-style-type: none"> Valid 100% Error 0% Empty 0% 	<ul style="list-style-type: none"> Valid 100% Error 0% Empty 0% 	<ul style="list-style-type: none"> Valid 100% Error 0% Empty 0% 	<ul style="list-style-type: none"> Valid 100% Error 0% Empty 0%
1	Krishna Madan		35 krishnamadan77@gmail.com	India
2	Vaishnavi Niljikar		24 vaish.niljikar@gmail.com	India
3	Naval Mandhana		26 naval45@gmail.com	India
4	Kesavan		35 kesavanplace@gmail.com	India
5	Mukund		35 mukund.potdar89@outlook.com	India



Keep Range of Rows

Specify the range of rows to keep.

First row

2

Number of rows

4

	A _B C Name	T ₁ ₂ Age	A _B C Email	A _B C Country
	<ul style="list-style-type: none"> Valid 100% Error 0% Empty 0% 	<ul style="list-style-type: none"> Valid 100% Error 0% Empty 0% 	<ul style="list-style-type: none"> Valid 100% Error 0% Empty 0% 	<ul style="list-style-type: none"> Valid 100% Error 0% Empty 0%
1	Krishna Madan		35 krishnamadan77@gmail.com	India
2	Vaishnavi Niljikar		24 vaish.niljikar@gmail.com	India
3	Naval Mandhana		26 naval45@gmail.com	India
4	Kesavan		35 kesavanplace@gmail.com	India
5	Mukund		35 mukund.potdar89@outlook.com	India
6	Rashmi Ranjan		25 rashmiranjan04@gmail.com	India
7	Nirvisha		26 nirvisha99@gmail.com	India
8	Lokesh Babu		25 LokeshBabu@hmail.com	India
9	Surya Narayan		28 Surya@gmail.com	India

Split Column

Group By

Replace Value

Data Type: Text

By Delimiter

By Number of Characters

By Positions

By Lowercase to Uppercase

By Uppercase to Lowercase

By Digit to Non-Digit

By Non-Digit to Digit

"@"

Username

krishnamadan77
mukund.potdar89
kesavanplace
Surya
nirvisha99
naval45
rashmiranjan04
LokeshBabu
vaish.niljikar

krishnaMadan

0 1 2 3 4 5 6 7 8 9 10 11

Coding Ninja

Position - 6

Ninja

By Lowercase to Uppercase

codingNinja - coding Ninja

By Uppercase to Lowercase

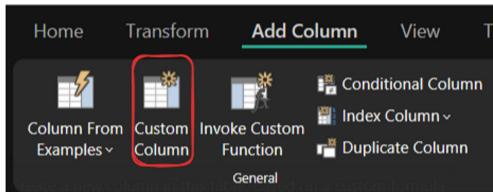
CODINGninja - CODING ninja

By Digit to Non-Digit

Krishnamadan77@gmail.com - Krishnamadn 77@gmail.com

By Non-Digit to Digit

krishnaMadanKrishna	1111\$@Krishna
vaishnaviNiljikarVaishnavi	389279Vaishnavi
navalMandhanaNaval	239237Naval



Customer ID	Product ID	Product bought	Qty bought	Price per item	Sales
Valid Error Empty	Valid Error Empty	Valid Error Empty	Valid Error Empty	Valid Error Empty	Valid Error Empty
100% 0% 0%	100% 0% 0%	100% 0% 0%	100% 0% 0%	100% 0% 0%	100% 0% 0%
20 distinct, 20 unique	5 distinct, 0 unique	5 distinct, 0 unique	20 distinct, 20 unique	10 distinct, 5 unique	20 distinct, 20 unique
1	Pen Set		62	4.95	309.38
2	Binder		29	1.99	57.71
3	Pen Set		55	12.49	686.95
4	Binder		81	19.95	1,619.19
5	Pen Set		42	23.95	1,005.90
6	Pencil		35	4.95	174.65
7	Desk		3	275.00	825.00
8	Desk		2	125.00	250.00
9	Pencil		7	1.25	9.03
10	Pen Set		16	15.95	255.84
11	Pen		76	1.99	151.24
12	Binder		28	8.95	251.72
13	Binder		57	19.99	1,139.43
14	Pen		64	8.99	575.36
15	Pencil		14	1.25	18.06
16	Pen		15	19.99	299.85
17	Binder		11	4.99	54.89
18	Pen Set		96	4.99	479.04
19	Binder		94	19.99	1,879.06

Custom Column

Add a column that is computed from the other columns.

New column name: Sales

Custom column formula: `= [Qty bought] * [Price per item]`

Available columns:

- Customer ID
- Product ID
- Product bought
- Qty bought
- Price per item

<< Insert

Learn about Power Query formulas

✓ No syntax errors have been detected.

OK Cancel



Group By - Grouping the column having distinct values. So that each occurrence of each value would be referred and the next column will always be Aggregated Columns.

Based on Each occurrence of Product Bought, How much revenue they are generating.

Revenue is nothing but sum of Sales.

Product bought	Qty bought	Price per item	Sales
Pen Set	62	4.99	309.38
Binder	29	1.99	57.71
Pen Set	55	12.49	686.95
Binder	81	19.99	1619.19
Pen Set	42	23.95	1005.9
Pencil	35	4.99	174.65
Desk	3	275.00	825
Desk	2	125.00	250

Group By

Aggregated with Sum of Sales.

Group By

Specify the column to group by and the desired output.

Basic Advanced

Product bought

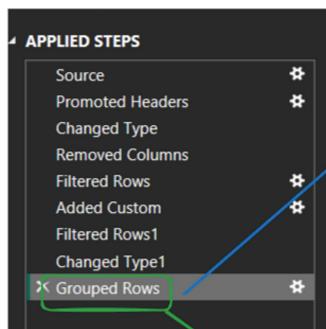
New column name: Revenue

Operation: Sum

Column: Sales

OK Cancel

A	B	C	D
Valid	100%	Valid	100%
Error	0%	Error	0%
Empty	0%	Empty	0%
5 distinct, 5 unique		5 distinct, 5 unique	
Pen Set	2737.11		
Binder	5002		
Pencil	288.17		
Desk	1075		
Pen	1026.45		



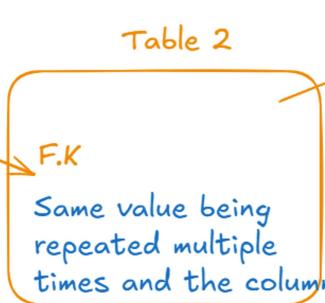
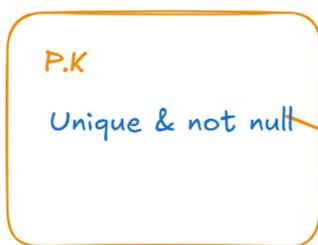
Product bought	Sum of Sales
Binder	\$5,002.00
Desk	\$1,075.00
Pen	\$1,026.45
Pen Set	\$2,737.11
Pencil	\$288.17
Total	\$10,128.73

So, Make sure to remove the Group By option from the applied steps so that you won't lose the original data.

It is over writing our raw data, We just use Group By for a reference. Largely it has a replacement in Front End. Report View - Matrix Visuals

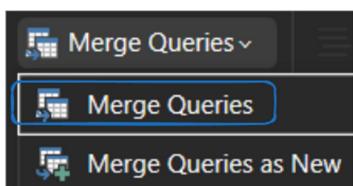
Merge Queries

Table 1



If 2 tables, consist of P.K & F.K , then only we can do merge Queries.

Your New Table would be expanding on the horizontal direction,



Merge

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

Sales

Customer ID	Customer Region	Customer Age	Sales
1	North	49	309.38
2	Unknown	10	57.71
3	North	47	686.95
4	South	11	1,619.19
5	East	17	1,005.90

Customers

Customer ID	Customer Name
1	Elvin Catron
2	Gustano Primm
3	Lacresha Whitty
4	Octavio Ricci
5	Jacinda Moffett

Join Kind

Left Outer (all from first, matching from second)

Use fuzzy matching to perform the merge

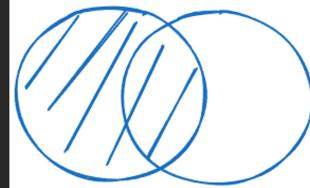
▷ Fuzzy matching options

✓ The selection matches 20 of 20 rows from the first table.

OK

Cancel

Left Join



All the parts of left table along with commons parts of right table.

Customers	Customer ID	Customer Region	Customer Age	Sales	Customers
Sales	1	North	49	309.38	Table
Product Information Po...	2	Unknown	10	57.71	Table
Students Info	3	North	47	686.95	Table
	4	South	11	1,619.19	Table
	5	East	17	1,005.90	Table
	6	Unknown	23	174.65	Table
	7	Unknown	21	825.00	Table
	8	Unknown	27	250.00	Table
	9	Unknown	44	9.03	Table
	10	Unknown	10	255.84	Table
	11	North	29	151.24	Table
	12	South	29	251.72	Table
	13	South	42	1,139.43	Table
	14	North	12	575.36	Table
	15	Unknown	45	18.06	Table
	16	Unknown	30	299.85	Table
	17	Unknown	35	54.89	Table
	18	West	50	479.04	Table
	19	Unknown	42	1,879.06	Table

Sales

Customers

Search Columns to Expand

Expand Aggregate

(Select All Columns) Customer ID Customer Name

Use original column name as prefix

OK Cancel

Properties

Name Sales

All Properties

APPLIED STEPS

Source Navigation Promoted Headers Changed Type Replaced Value Replaced Value1 Replaced Errors Merged Queries Expanded Customers

Merge Queries

Merge Queries

Merge Queries as New

Close & Apply Close New Source Recent Sources New Query Enter Data Data source settings Manage Parameters Refresh Preview Advanced Editor Properties Choose Columns Remove Columns Keep Rows Remove Rows Split Column Group By Replace Values Sort Data Type: Text Use First Row as Headers Transform Merge Queries Append Queries Combine Files Text Analytics Vision Azure Machine Learning Ai Insights

Queries [5]

Customers Sales Product Information Po... Students Info Expanded Merge Table

Customer ID Customer Region Customer Age \$ Sales Customer.Customer Name

20 distinct, 20 unique 5 distinct, 2 unique 16 distinct, 12 unique 20 distinct, 29 unique 20 distinct, 20 unique

	Customer ID	Customer Region	Customer Age	\$ Sales	Customer.Customer Name
1	1	North	49	305.38	Elvin Catron
2	2	Unknown	10	57.71	Gustano Prim
3	3	North	47	686.95	Lacresha Whitty
4	4	South	11	1,619.19	Octavio Ricci
5	5	East	17	1,005.90	Jacinda Moffett
6	6	Unknown	23	174.65	Betty scruggs
7	7	Unknown	21	825.00	Era Gassner
8	8	Unknown	27	250.00	Laticia Crimi
9	9	Unknown	44	9.03	Alberta Fabela
10	10	Unknown	10	255.84	Taylor Pogue
11	11	North	29	151.24	Grisede Mariscal
12	12	South	29	251.72	Sylvester Willbourn
13	13	South	42	1,139.47	Ricky Cheshire
14	14	North	12	575.36	Karisa Oquwndo
15	15	Unknown	45	18.06	Debbie Orone
16	16	Unknown	35	506.46	Loralee Widell

Query Settings

Properties

Name Expanded Merge Table

All Properties

APPLIED STEPS

Source Expanded Customers

Append Query

Sales 2020

500

Sales 2021

1000

Sales 2022

1500

Make sure
all the tables
need to be
identical

3000 rows

Note : Each Table should be identical means,
each table have same number of columns in same order
having same data types.

Get Data From Web - Transpose [Converting rows into column, and column into rows.]

Queries [6]

Quarterly Results

	Sales	Expenses	Operating Profit	CRM %	Other Income	Interest
1	72,229	65,151	7,078	10%	789	2,401
2	78,439	70,156	8,289	11%	139	2,381
3	71,935	69,523	2,413	3%	2,381	2,421
4	79,611	74,039	5,572	7%	1,351	2,487
5	88,489	77,668	10,800	12%	1,130	2,676
6	105,932	92,818	13,114	12%	1,453	2,642
7	102,236	85,019	13,217	13%	895	2,615
8	105,329	93,362	13,267	13%	1,357	2,652
9	110,577	95,159	15,418	14%	1,634	2,485
10	119,986	100,851	17,135	14%	1,619	2,734
11	108,648	92,263	15,785	15%	1,747	2,088
12	101,450	89,291	12,159	12%	1,647	2,034
13	113,575	100,522	13,043	11%	1,784	1,725
14						

Query Settings

PROPERTIES

APPLIED STEPS

List

2401
2381
2421
2487
2676
2642
2615
2652
2485
2234
2088
2034
1725

[2401, 2381, 2421, 2487]

Format

Merge Columns

ABC Extract

123 Parse

lowercase

UPPERCASE

Capitalize Each Word

Trim

Clean

Add Prefix

Add Suffix

Extract

Length

First Characters

Last Characters

Range

Text Before Delimiter

Text After Delimiter

Text Between Delimiters

Parse

XML

JSON

```
{
  "Name": "Krishna Madan",
  "Country": "India",
  "Languages": ["C++", "Java", "Python", "MySQL"]
}
```

Queries [7]

- Customers
- Sales
- Product Information Po...
- Students Info
- Expanded Merge Table
- Quarterly Results
- sample_json_column

= Table.PromoteHeaders(Source, [PromoteAllScalars=true])

	A ^B _C ID	A ^B _C Name	A ^B _C Details
1	1	Alice	{"age": 25, "city": "New York", "skills": ["Python", "SQL"]}
2	2	Bob	{"age": 30, "city": "San Francisco", "skills": ["Java", "Spring"]}
3	3	Charlie	{"age": 28, "city": "Los Angeles", "skills": ["JavaScript", "React"]}
4	4	David	{"age": 35, "city": "Chicago", "skills": ["C#", "Azure"]}

Home Transform Add Column View Tools Help

Transpose Reverse Rows Detect Data Type Fill Move Split Column Format Extract Parse Text XML JSON

Groups By Use First Row as Headers Count Rows Table Any Column

Queries [7]

- Customers
- Sales
- Product Information Po...
- Students Info
- Expanded Merge Table
- Quarterly Results
- sample_json_column

= Table.PromoteHeaders(Source, [PromoteAllScalars=true])

	A ^B _C ID	A ^B _C Name	A ^B _C Details
1	1	Alice	{"age": 25, "city": "New York", "skills": ["Python", "SQL"]}
2	2	Bob	{"age": 30, "city": "San Francisco", "skills": ["Java", "Spring"]}
3	3	Charlie	{"age": 28, "city": "Los Angeles", "skills": ["JavaScript", "React"]}
4	4	David	{"age": 35, "city": "Chicago", "skills": ["C#", "Azure"]}

Customers

Sales

Product Information Po...

Students Info

Expanded Merge Table

Quarterly Results

sample_json_column

	A ^B _C ID	A ^B _C Name	A ^B _C Details
1	1	Alice	Record
2	2	Bob	Record
3	3	Charlie	Record
4	4	David	Record

The screenshot shows a Power BI data flow transformation. The source table 'sample_json_column' contains four rows of JSON data. The target table has five columns: 'ABC ID', 'ABC Name', 'ABC Age', 'ABC City', and 'ABC Skills'. The 'Skills' column is highlighted with a blue border. A context menu is open for this column, showing options like 'Expand to New Rows' and 'Extract Values...'. A green arrow points from the 'Extract Values...' option to a 'Extract values from list' dialog box. This dialog box asks for a delimiter ('Comma') and has 'OK' and 'Cancel' buttons. The target table shows the transformed data where the 'Skills' column now contains lists of skills for each row.

Extract values from list

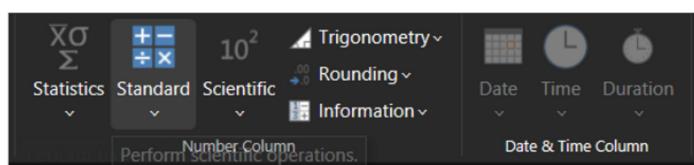
Select a delimiter to use for concatenating list values

Comma

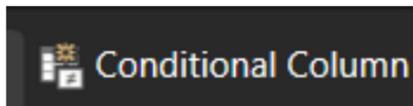
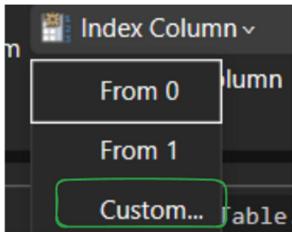
OK

Cancel

	ABC ID	ABC Name	ABC Age	ABC City	ABC Skills
	Valid 0% Error 0% Empty 0%	Valid 0% Error 0% Empty 0%	Valid 0% Error 0% Empty 0%	Valid 0% Error 0% Empty 0%	Valid 0% Error 0% Empty 0%
	4 distinct, 4 unique	4 distinct, 4 unique	4 distinct, 0 unique	4 distinct, 0 unique	4 distinct, 4 unique
1	1	Alice		25	New York
2	1	Alice		25	New York
3	2	Bob		30	San Francisco
4	2	Bob		30	San Francisco
5	3	Charlie		28	Los Angeles
6	3	Charlie		28	Los Angeles
7	4	David		35	Chicago
8	4	David		35	Chicago



Self Explore



If my order Qty ≥ 50 , So I'll create a new column name Order Type, so if its ≥ 50 - Bulk Order, Else Normal Order.

Add Conditional Column

Add a conditional column that is computed from the other columns or values.

New column name:

If Then

Else

OK **Cancel**

