

Data Cleaning with Power BI

- Quick Revision of the last lectures.
- Exploring the features of Power Query Editor.
- Difference Between Transform & Add Columns.
- Transforming the data by scraping it from Web.
- Filling up the manual Data.

Power Query Editor

→ Get Data > Excel Workbook / .CSV format.

[Make sure you have downloaded the dataset shared in the chat]

The screenshot shows the Power Query Editor interface. On the left, the 'Common data sources' pane is open, displaying various options like 'Excel workbook', 'Text/CSV', and 'Web'. The 'Text/CSV' option is highlighted with a green box. In the center, the 'Navigator' pane lists datasets: 'Sales Dataset Po...', 'Customers', and 'Sales', with 'Customers' and 'Sales' checked. To the right, the 'Preview' pane displays a table titled 'Sales' with four columns: Column1, Column2, Column3, and Column4 (Sales). The table contains 20 rows of data. At the bottom right of the preview pane are three buttons: 'Load', 'Transform Data' (which is highlighted with a green box), and 'Cancel'.

Column1	Column2	Column3	Column4
Customer ID	Customer Region	Customer Age	Sales
1	North	49	309.38
2		10	57.71
3	North	47	686.95
4	South	11	1619.19
5	East	17	1005.9
6	null	23	174.65
7	null	21	825
8		27	250
9		44	9.03
10	null	10	255.84
11	North	abc	151.24
12	South	29	251.72
13	South	42	1139.43
14	North	12	575.36
15	null	45	18.06
16	null	30	299.85
17	null	35	54.89
18	West	50	479.04
19	null	42	1879.06
20		44	86.43

It stores all the tables.

Step 1 : Promote the Headers, so that when you change the data-type , your columns header won't make it heterogenous column.

Step 2 : Changing the data type of each column.

Step 3 : Do remember to activate the Column Quality, Column Distribution, Column Profile from view tab to get the understanding of your dataset, how much it's clean and how much you need to process.

Transform / Add Column

Transform - When you have to change your existing dataset without adding any other column.
It's changing your original table.

Add Columns - Its being added based on some condition or because of some calculation.

Drawback - Increases the file size.

The screenshot shows the Power BI ribbon with the 'Transform' tab selected. A context menu is open over a table named 'Customer Region'. The 'Replace Values' option is highlighted with a yellow box. Below the ribbon, two 'Replace Values' dialog boxes are displayed side-by-side. Both dialogs have the following fields: 'Value To Find' (empty), 'Replace With' ('Unknown'), and an 'OK' button at the bottom right. The left dialog has a note above it: 'add a space to find an empty cell'.

The screenshot shows the Power BI ribbon with the 'Transform' tab selected. A context menu is open over a table named 'Customer Age'. The 'Statistics' dropdown is open, with 'Average' highlighted with a green box. Below the ribbon, two tables are shown side-by-side. The left table has a green box around its 'Customer Age' column header. The right table shows the results of the transformation, with a note below it: '[Once the column being cleaned, you will see the column distribution]'.

$$x + 0 = x.$$

[0 is used in addition as an identity property]

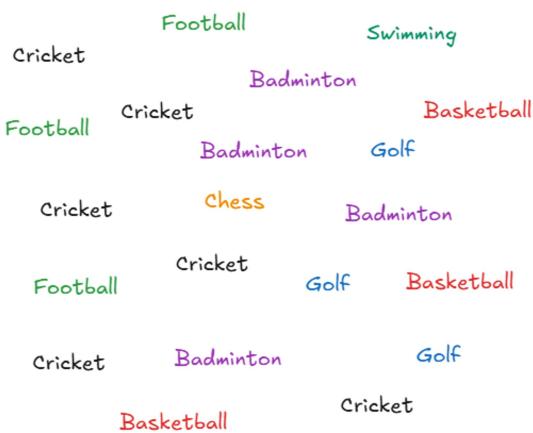
$$x * 1 = x.$$

[1 is used in multiplication as an identity property]

Distinct / Unique.

	Customer ID	Customer Region	Customer Age	Sales
	Valid: 100% Error: 0% Empty: 0%			
	20 distinct, 20 unique	5 distinct, 2 unique	17 distinct, 14 unique	20 distinct, 20 unique
1	1 North	1	49	309.38
2	2 Unknown	2	10	57.71
3	3 North	3	47	686.95
4	4 South	4	11	1,619.19
5	5 East	5	17	1,005.90
6	6 Unknown	6	23	174.65
7	7 Unknown	7	21	825.00
8	8 Unknown	8	27	250.00
9	9 Unknown	9	44	9.03
10	10 Unknown	10	10	255.84
11	11 North	11	29	151.24
12	12 South	12	29	251.72
13	13 South	13	42	1,139.43
14	14 North	14	12	575.36
15	15 Unknown	15	45	18.06
16	16 Unknown	16	30	299.85
17	17 Unknown	17	35	54.89
18	18 West	18	50	479.04
19	19 Unknown	19	42	1,879.06
20	20 Unknown	20	44	86.43

Sports Category



Distinct - 7

[Multiple Occurrence,
consider it only once.]

Unique - 2

[Only 1 Occurrence]

Cricket - 6

Football - 3

Badminton - 4

Basketball - 3

Golf - 3

Swimming - 1

Chess - 1

Swimming - 1

Chess - 1

Product Information - .CSV File Cleaning

i23 Customer ID	i23 Product ID	A8c Product bought	i23 Qty bought	\$ Price per item
Valid	47%	Valid	47%	Valid 47%
Error	0%	Error	0%	Error 0%
Empty	53%	Empty	53%	Empty 53%
6 distinct, 0 unique				
Customer ID	20 (47%)	23 (53%)	21 distinct, 20 unique	11 distinct, 5 unique
Valid	0 (0%)	Empty		
...				
3	3	1 Pen Set	62	4.99
4	4	2 Binder	29	1.99
5	5	1 Pen Set	55	12.49
6	6	2 Binder	81	19.99
7	7	1 Pen Set	42	23.95
8	8	3 Pencil	35	4.99
		4 Desk	3	275.00
		4 Desk	2	125.00

Customers

- Sort Ascending
- Sort Descending
- Clear Sort
- Clear Filter
- Remove Empty
- Number Filters
- Search
 - (Select All)
 - (null)
 - 1

After you done with all cleaning,
load the data to front End and
Save the file.

Home Transform Add Column View Tools Help

Close & Apply New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Properties Advanced Editor Choose Columns Remove Columns Keep Rows Remove Rows Split Column Group By Data Type: Whole Number Merge Queries Append Queries Combine Files Text Analytics Vision Azure Machine Learning

Queries [3]

Customer ID Product ID Product bought Qty bought Price per item

i23 Customer ID	i23 Product ID	A8c Product bought	i23 Qty bought	\$ Price per item
Valid	100%	Valid	100%	Valid 100%
Error	0%	Error	0%	Error 0%
Empty	0%	Empty	0%	Empty 0%
20 distinct, 20 unique				
1	1	1 Pen Set	62	4.99
2	2	2 Binder	29	1.99
3	3	1 Pen Set	55	12.49
4	4	2 Binder	81	19.99
5	5	1 Pen Set	42	23.95
6	6	3 Pencil	35	4.99
7	7	4 Desk	3	275.00
8	8	4 Desk	2	125.00
9	9	4 Pencil	7	1.29
10	10	1 Pen Set	16	15.99
11	11	5 Pen	76	1.99
12	12	2 Binder	28	8.99
13	13	2 Binder	57	19.99

Query Settings

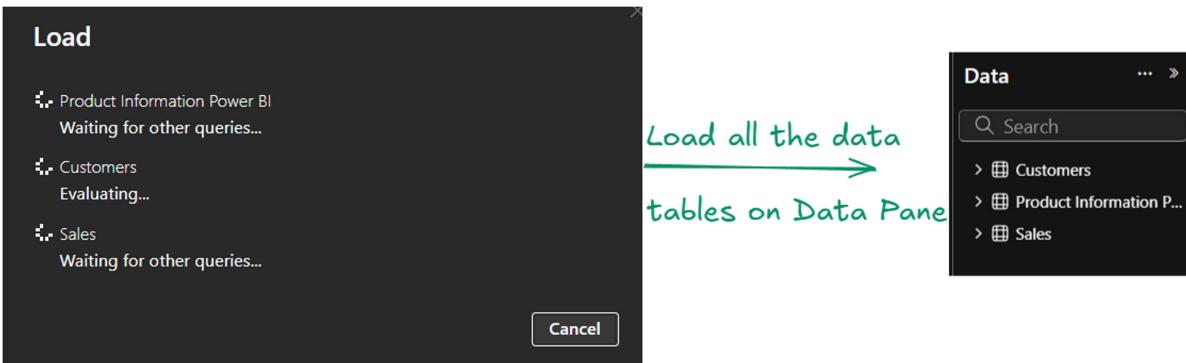
PROPERTIES

Name: Product Information Power BI

All Properties

APPLIED STEPS

Source: Promoted Headers: Changed Type: Removed Columns: Filtered Rows



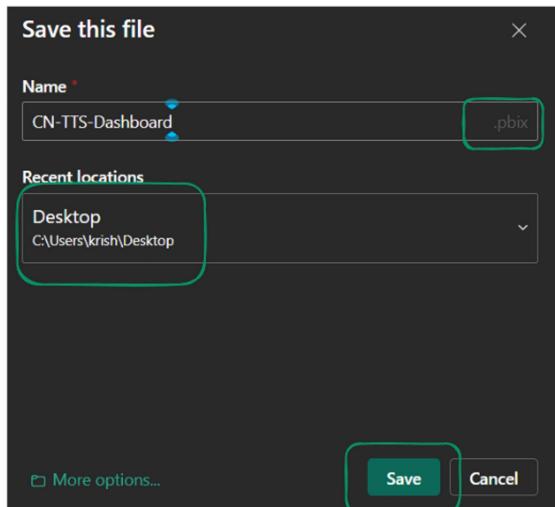
File Home Help Table tools Column tools

Name: Price per item Format: Currency

Data type: Fixed decimal num... \$ %

Customer ID	Product ID	Product bought	Qty bought	Price per item
1	1	Pen Set	62	\$4.99
2	2	Binder	29	\$1.99
3	1	Pen Set	55	\$12.49
4	2	Binder	81	\$19.99
5	1	Pen Set	42	\$23.95
6	3	Pencil	35	\$4.99
7	4	Desk	3	\$275.00
8	4	Desk	2	\$125.00
9	4	Pencil	7	\$1.29
10	1	Pen Set	16	\$15.99
11	5	Pen	76	\$1.99
12	2	Binder	28	\$8.99
13	2	Binder	57	\$19.99
14	5	Pen	64	\$8.99
15	3	Pencil	14	\$1.29
16	5	Pen	15	\$19.99
17	2	Binder	11	\$4.99
18	1	Pen Set	96	\$4.99
19	2	Binder	94	\$19.99
20	3	Pencil	67	\$1.29

Transform the currency as per your need.
By default it chooses the currency of that country which is being stored as Keyboard[Country].



Primary Key Vs Foreign Key

Swiggy Schema

