

Data Cleaning with Power BI - p2

- Understanding Features of Power Query Editor.
- Get data from the Web
- Create Your Own Data
- Invoke Custom Function - To build Rolling Calendar
- Parse Json data.

	Product ID	Product Name	Product Rank	+
1	1	Pen	2	
2	2	Pencil	3	
3	3	Pen Set	1	
4	4	Desk	5	
5	5	Binder	4	
+				

Github Link:

<https://github.com/KrishnaMentor/CN01-PowerBI-Notes>

```
List.Dates(  
    Source,  
    Number.From(DateTime.LocalNow()) - Number.From(Source),  
    #duration(1, 0, 0, 0)  
)
```

- List.Dates(start, count, step)

Get the Data From the Web

→ Cricbuzz / IPL website

1. Fetch the url Link where you want to scrap.

- <https://www.cricbuzz.com/live-cricket-scorecard/112462/ind-vs-aus-1st-semi-final-a1-v-b2-icc-champions-trophy-2025>

Navigator

Display Options ▾

- Suggested Tables [10]
 - Table 1
 - Table 2
 - Table 3
 - Table 4
 - Table 5
 - Table 6
 - Table 7
 - Table 8
 - Table 9
 - Table 10
- Text [2]
 - HTML Code
 - Displayed Text

No items selected for preview

Add Table Using Examples

Load Transform Data Cancel

APPLIED STEPS

- Source
- Extracted Table From Html
- Renamed Columns

Customers
Sales
Product Information
Manual Data
Batsman Record
Bowlers Record

	A ^b _c Batsman	A ^b _c Wicket Reason	A ^b _c Runs	A ^b _c Balls	A ^b _c Fours
	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%	● Valid 100% ● Error 0% ● Empty 0%
	19 distinct, 19 unique	17 distinct, 16 unique	17 distinct, 15 unique	16 distinct, 13 unique	7 distinct, 2 unique
1	Head	c Shubman Gill b Varun Chakravarthy	39	33	5
2	Cooper Connolly	c Rahul b Shami	0	9	0
3	Steven Smith (c)	b Shami	73	96	4
4	Labuschagne	lbw b Ravindra Jadeja	29	36	2
5	Josh Inglis (wk)	c Kohli b Ravindra Jadeja	11	12	0
6	Alex Carey	run out (Shreyas Iyer)	61	57	8
7	Maxwell	b Axar	7	5	0
8	Dwarshuis	c Shreyas Iyer b Varun Chakravarthy	19	29	1
9	Zampa	b Hardik Pandya	7	12	0
10	Nathan Ellis	c Kohli b Shami	10	7	0
11	T Sangha	not out	1	1	0
12	Rohit Sharma (c)	lbw b Cooper Connolly	28	29	3
13	Shubman Gill	b Dwarshuis	8	11	1
14	Virat Kohli	c Dwarshuis b Zampa	84	98	5
15	Shreyas Iyer	b Zampa	45	62	3
16	Axar Patel	b Nathan Ellis	27	30	1
17	KL Rahul (wk)	not out	42	34	2
18	Hardik Pandya	c Maxwell b Nathan Ellis	28	24	1
19	Ravindra Jadeja	not out	2	1	0

Queries [6]

Customers
Sales
Product Information
Manual Data
Batsman Record
Bowlers Record

= Table.ReorderColumns(#"Changed Type", {"Bowler Name", "Overs", "Wickets", "No Balls", "Maiden", "Runs", "Economy"})

	A ^b _c Bowler Name	1.2 Overs	A ^b _c Wickets	A ^b _c No Balls	A ^b _c Maiden	A ^b _c Runs	1.2 Economy
	● Valid 100% ● Error 0% ● Empty 0%						
	13 distinct, 13 unique	7 distinct, 5 unique	4 distinct, 1 unique	1 distinct, 0 unique	2 distinct, 0 unique	11 distinct, 9 unique	11 distinct, 9 unique
1	Shami	10	3	0	0	48	48
2	Hardik Pandya	5.3	1	0	0	40	40
3	Kuldeep Yadav	8	0	0	0	44	44
4	Varun Chakravarthy	10	2	0	0	49	49
5	Axar	8	1	0	1	43	43
6	Ravindra Jadeja	8	2	0	1	40	40
7	Ben Dwarshuis	7	1	0	0	39	39
8	Nathan Ellis	10	2	0	0	49	49
9	Cooper Connolly	8	1	0	0	37	37
10	Adam Zampa	10	2	0	0	60	60
11	Tanveer Sangha	6	0	0	0	41	41
12	Glenn Maxwell	6.1	0	0	0	35	35
13	Travis Head	1	0	0	0	6	6

Query Settings

PROPERTIES

Name: Bowlers Record

All Properties

APPLIED STEPS

- Source
- Extracted Table From Html
- Removed Duplicates
- Promoted Headers
- Renamed Columns
- Changed Type
- Reordered Columns

Web Scrapping - Tata Motors - Screener

- <https://www.screener.in/company/TATAMOTORS/consolidated/>

Navigator

The Navigator interface shows a sidebar with a tree view of tables and a main area with a preview of Table 2.

- Display Options** dropdown.
- HTML Tables [12]** section:
 - Table 1
 - Table 2 (highlighted with an orange box)
 - Table 3
 - Table 4
 - Table 5
 - Table 6
 - Table 7
 - Table 8 (highlighted with an orange box)
 - Table 9
 - Table 10
 - Table 11
 - Table 12
- Suggested Tables [4]** section:
 - Table 13
 - Table 14
 - Table 15
 - Table 16
- Text [2]** section:
 - HTML Code

Table View tab is selected. The preview shows Table 2 with columns: Column1, Column2, Column3, Column4, Column5, Column6, Column7. The data includes financial metrics like Sales+, Expenses+, Operating Profit, etc., with values and percentages.

Buttons at the bottom: **Add Table Using Examples**, **Load**, **Transform Data** (highlighted with an orange box), and **Cancel**.

The transformed data table has columns for each row of the original table, labeled A through G. The first row contains validation statistics for each column.

A	B	C	D	E	F	G
Valid: 92%						
Error: 0%						
Empty: 8%						
13 distinct, 13 unique						
Sales +	Dec 2021	Mar 2022	Jun 2022	Sep 2022	Dec 2022	Mar 2023
Expenses +	72,229	78,439	71,935	79,611	88,489	105,932
Operating Profit	65,151	70,156	69,522	74,039	77,668	92,818
OPM %	7,078	8,283	2,413	5,572	10,820	13,114
Other Income +	10%	11%	3%	7%	12%	12%
Interest	789	189	2,381	1,351	1,130	1,453
Depreciation	2,401	2,381	2,421	2,487	2,676	2,642
Profit before tax	6,078	6,432	5,841	5,897	6,072	7,050
Tax %	-612	-341	-3,468	-1,461	3,203	4,875
Net Profit +	119%	222%	44%	-31%	8%	-13%
EPS in Rs	-1,451	-992	-4,951	-898	3,043	5,496
Raw PDF	-4.57	-3.11	-15.08	-2.84	8.91	16.28

Transpose

Converting Row into columns and vice - versa.

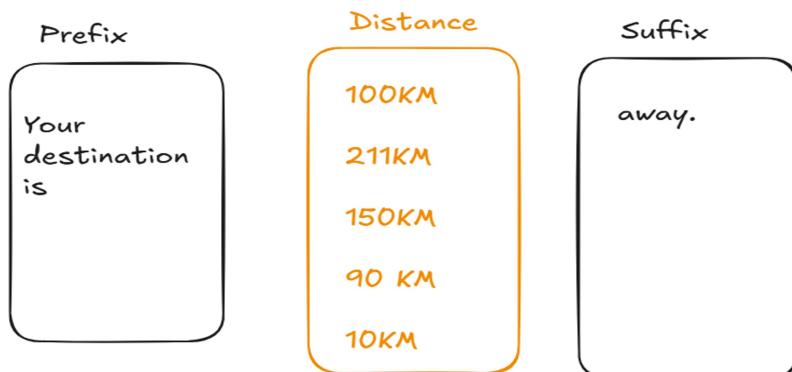
PivotTable Data:

	Sales	Expenses	Operating Profit	OPM %	Other Income	Interest
Dec 2021	72,229	65,151	7,078	10%	789	2,401
Mar 2022	78,439	70,156	8,283	11%	189	2,381
Jun 2022	71,935	69,522	2,413	3%	2,381	2,421
Sep 2022	79,611	74,039	5,572	7%	1,351	2,487
Dec 2022	88,489	77,668	10,820	12%	1,130	2,676
Mar 2023	105,932	92,818	13,114	12%	1,453	2,642
Jun 2023	102,236	89,019	13,217	13%	895	2,615
Sep 2023	105,129	91,362	13,767	13%	1,557	2,652
Dec 2023	110,577	95,159	15,418	14%	1,604	2,485
Mar 2024	119,986	102,851	17,135	14%	1,619	2,234
Jun 2024	108,048	92,263	15,785	15%	1,747	2,088
Sep 2024	101,450	89,291	12,159	12%	1,647	2,034
Dec 2024	113,575	100,532	13,043	11%	1,764	1,725

Text Transformation Options:

- lowercase
- UPPERCASE
- Capitalize Each Word
- Trim
- Clean
- Add Prefix
- Add Suffix

Trimming is like removing the leading or trailing whitespace. On the other hand, Clean is more likely to remove hidden data.



Extract Step Options:

- Length
- First Characters
- Last Characters
- Range
- Text Before Delimiter
- Text After Delimiter
- Text Between Delimiters

space - delimiter: CodingNinjas — → 13

Coding - First 6 Character: Coding

Ninjas - Last 6 Characters: Ninjas

Range (5,10) - g Nin: g Nin

It will prompt you 2 delimeters

A screenshot of the Microsoft Power Query ribbon. The 'Parse' dropdown menu is open, showing two options: 'XML' and 'JSON'. The 'XML' option is highlighted.

Json - Key : Value

Details
{"age": 25, "city": "New York", "skills": ["Python", "SQL"]}
{"age": 30, "city": "San Francisco", "skills": ["Java", "Spring"]}
{"age": 28, "city": "Los Angeles", "skills": ["JavaScript", "React"]}
{"age": 35, "city": "Chicago", "skills": ["C#", "Azure"]}



Age	City	Skills
25		
30		
28		
35		

