

Data Cleaning and Preparation Techniques

- TRIM
- PROPER
- UPPER
- LOWER
- LEFT
- MID
- RIGHT
- CONCATENATE
- TEXTJOIN
- SPLIT
- IF
- IFERROR

Excel spreadsheet showing text cleaning functions. The spreadsheet has columns A through Q and rows 1 through 20. The data in the spreadsheet is as follows:

| Raw Text | |
|----------|-------|
| alice | alice |
| BOB | |
| CHARlie | |
| DAVID | |

Handwritten notes and diagrams:

- Text Cleaning**
 - TRIM → Removes all spaces
 - TRIM (Text)
 - Diagram: → (with arrows pointing to the spaces being removed)
- Raw data** → **Clean/Pre processed data**

Excel spreadsheet showing text cleaning functions. The spreadsheet has columns A through Q and rows 1 through 20. The data in the spreadsheet is as follows:

| Raw Text | | | |
|----------|---------|-------|-------|
| alice | alice | Alice | ALICE |
| BOB | | | |
| CHARlie | charlie | | |
| DAVID | | | |

Handwritten notes and diagrams:

- Text Cleaning**
 - Diagram: (TEXT) → uppercase
 - List of functions:
 - TRIM
 - PROPER
 - UPPER
 - LOWER

Excel spreadsheet showing text manipulation functions. The spreadsheet has columns A through R and rows 1 through 20. The data in the spreadsheet is as follows:

| Full Name | |
|---------------|--|
| Alice Johnson | |
| Bob Smith | |
| Charlie Brown | |
| David Lee | |

Handwritten notes and diagrams:

- TEXT MANIPULATION**
 - LEFT
- Diagram: **LEFT (text, 5)** with an arrow pointing to the first 5 characters of "Alice Johnson" (Alice).

Excel screenshot showing a table with Full Name and First Name columns. The formula bar shows `=SEARCH("L",A5)`.

| Full Name | First Name |
|---------------|------------|
| Alice Johnson | Alice |
| Bob Smith | |
| Charlie Brown | |
| David Lee | |

Handwritten notes:

- SEARCH** - (,)
str (what) cell (where)
- TEXT MANIPULATION**
LEFT
MID

Excel screenshot showing the same table. The formula bar shows `=LEFT(A2,5)`. The cell B2 contains the value 7.

| Full Name | First Name |
|---------------|------------|
| Alice Johnson | Alice |
| Bob Smith | |
| Charlie Brown | |
| David Lee | |

Handwritten notes:

- SEARCH** - (,)
str (what) cell (where)
- TEXT MANIPULATION**
LEFT
MID
- LEFT(A2,5)**
Constant variable
" "
- SEARCH(" ", A2)**
ABHI RAM
1 → 5

Excel screenshot showing the formula bar with `=LEFT(A2,SEARCH(" ",A2)-1)`. The cell B2 contains the value 7.

| Full Name | First Name |
|---------------|------------|
| Alice Johnson | Alice |
| Bob Smith | |
| Charlie Brown | |
| David Lee | |

Excel screenshot showing a table with Full Name, First Name, and Last Name columns. The formula bar shows `=MID(A2,SEARCH(" ",A2)+1,LEN(A2))`. The cell C2 contains the value 7.

| Full Name | First Name | Last Name |
|---------------|------------|-----------|
| Alice Johnson | Alice | Johnson |
| Bob Smith | Bob | Smith |
| Charlie Brown | Charlie | Brown |
| David Lee | David | Lee |

Handwritten notes:

- LEN(TEXT)**
TEXT start no. of chars
- MID** - (,)
TEXT start no. of chars

Excel Formula: =TEXTJOIN(" ",TRUE,A2:B2)

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R |
|---------------|-----------|---------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| First Name | Last Name | Full Name | | | | | | | | | | | | | | | |
| Alice | Johnson | Alice Johnson | | | | | | | | | | | | | | | |
| Bob | Smith | Bob Smith | | | | | | | | | | | | | | | |
| Charlie | Brown | Charlie Brown | | | | | | | | | | | | | | | |
| David | Lee | David Lee | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | |
| Alice-Johnson | | | | | | | | | | | | | | | | | |
| Bob-Smith | | | | | | | | | | | | | | | | | |
| Charlie-Brown | | | | | | | | | | | | | | | | | |
| David-Lee | | | | | | | | | | | | | | | | | |

Handwritten Notes:

- CONCAT** -- (A," ",B)
- RAW** (Empty spaces)
- Delimiters** (Comma, space, etc.)
- TEXT** (Text to be joined)
- EMPTY CELL** (Empty cell in the range)
- TEXT Combining**
 - Concatenate
 - TEXT JOIN

Excel Spreadsheet:

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
|---|---|-------|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | | Sales | Cost | | | | | | | | | | | | | | | | |
| 2 | | 200 | 150 A | | | | | | | | | | | | | | | | |
| 3 | | 100 | 100 B | | | | | | | | | | | | | | | | |
| 4 | | 300 | 250 A | | | | | | | | | | | | | | | | |
| 5 | | | | | | | | | | | | | | | | | | | |

Handwritten Notes:

- $SP \leq CP$
- Logic: $Sales > Cost \rightarrow A$
 $Sales < Cost \rightarrow B$
- Conditional Logic Diagram:
 $3 > 4 \rightarrow \begin{cases} T \rightarrow A \\ F \rightarrow B \end{cases} \rightarrow IF$
- IF (True, False, Condⁿ)
- Conditional logic $\rightarrow IF$