

Lesson:



Introduction to MS Excel - 3



In our previous session, we covered

- Basic Excel operators
- Sorting and Filtering Data

In this session, we will be covering

- Cell References
- Introduction to a large dataset
- Writing your first Excel Function

Cell References

A cell is an intersection between a row and a column. In a spreadsheet, values are stored in a cell. Cell reference is like an address for a cell. If a cell falls under column A, and row 10, the reference for that cell is A10. Cell references are very useful in Excel. Now let's start our first spreadsheet operation. Let's go back to our Bank Dataset. You can download this dataset from pwskills.com for free.

Let's perform an addition on cell D2. Any operation you do on a cell will always need to start with an '=' sign. Let's say you simply want to add 24485 and 1754, the corresponding values in B2 and C2. The most basic way to perform this operation in D2 is to select D2 and enter

=24485+1754

	A	B	C	D	E
1	Customer Number	Amount Deposited	Interest Amount		
2	Customer 1	24485	1754	=24485+1754	

This will give you the required sum - 26239 in D2. But this is no different from performing this action on a calculator. This doesn't save any time. This is where references come into play. Instead of entering the numbers, you can enter the cell references corresponding to the numbers. This will make the new formula -

=B2+C2

D2
X
✓
fx
=
=B2+C2

	A	B	C	D
1	Customer Number	Amount Deposited	Interest Amount	
2	Customer 1	24485	1754	=B2+C2

When you are typing a formula, you can even use arrow keys to select the cell you want and the reference will automatically be populated in the cell. Now even this is not particularly time saving versus a calculator, but at least, if you change the value in B2 or C2, the value in D2 will automatically change.

Now to understand the real use of spreadsheets. Copy paste the formula in D2 to D3. Now check the formula in D3 – it will be

=B3+C3

This is the genius of spreadsheets – you do not need to keep re entering the same formula – you can just copy paste the formula and the software will replace the references in the formula with the corresponding references where the formula is pasted.

If you had pasted the formula to cell D20, the formula would have become

=B20+C20

If you had posted it to cell E20, it would have been

=C20+D20

Try this with different cells and see how the formula is being modified.

When you select a cell and hover your mouse over the bottom right corner of the cell, a '+' sign will appear. You can click on the corner and drag it downwards or to the right. This performs the same action as copy pasting the same formula across a large area. You can also select a number of cells and use the paste option to paste the formula to the entire selection.

If you have a large number of rows where you need to copy paste a formula, you can double click the '+' sign and it will automatically paste to the entire range below the cell (till the first blank row).

Absolute vs Relative References

All the references we have been discussing so far, have been Relative References. This means that when the formula

=A2+B2

Is copied to the cell below, it automatically becomes

=A3+B3

But if I put dollar signs in the formula,

=A2+\$B\$2

And copy it to the cell below, it becomes

=A3+\$B\$2

Here, \$B\$2 is called an Absolute Reference. This reference does not change when the formula is moved across cells.

There is another reference as well, called a Mixed Reference.

Let's assume you have the formula

=A2+B\$2

In cell C2. When you copy it to cell C3, the formula becomes

=A3+B\$2

But if you copy it to cell D2, the formula becomes

=B2+C\$2

Similarly a column lock would be \$B2 - here, if you have a formula in C2

=A2+B\$2

And you copy it to cell D2, the formula becomes

=B2+C\$2

But if you copy it to C3, it becomes

=A3+B\$2

The part of the reference that comes immediately after the \$ sign - be it row or column, will not change if the formula is copied to another cell.

Applying references

While you can always apply references by manually typing in \$ signs, Excel has a shortcut.

While entering a cell reference, you can press F4 key to convert it into an absolute reference (\$A\$4). Pressing F4 again will convert it into a mixed reference with row number locked (A\$4). Pressing F4 again will convert it into a mixed reference with column number locked (\$A4). Pressing F4 will convert it back into a relative reference.

Getting familiar with large datasets

Typical data that you have to use as a business analyst will be have a large number of rows. Let's familiarize ourselves with the dataset we will be using during this Excel course. While we will use smaller datasets to practice basic examples, we will be using this dataset throughout the course to ensure that you understand the range of operations that can be performed on data with Excel.

You can access this dataset on pwskills.com

Now let's understand this dataset. This was uploaded on Kaggle.com by DEEPAK SIROHIWAL. Kaggle.com is a great website where you can find many resources related to data - including datasets, projects, and more. This dataset contains information of around 10,000 Indian companies. This information was taken from ambitionbox.com.

This is what the first row looks like -

	A	B	C	D	E	F	G	H	I
1	Serial Number	Name	Rating	Reviews	Company_type	Headquarter	Age	No. of Employees	Tags
2	0	TCS	3.9	47000	Public	Mumbai,Maharashtra + 267 more	54	1 Lakh+ Employees (India)	BPO / Call Centre
3	1	Accenture	4.2	31200	Public	Dublin + 140 more	33	1 Lakh+ Employees (India)	IT Services & Consulting

Column A – Serial Number

Column B – Name of the company

Column C – Average rating of the company by users out of 5

Column D – Number of reviews

Column E – Company type – is it Publically Listed on the stock market or is it privately owned – accepted values are Public, Private, Partnership, LLC, LLP, Government, Govt Organization, NGO/NPO, JV, Central Public Sector Enterprise, Conglomerate, Proprietorship, One person company, Sole Proprietorship, Other

Column F – Location of headquarters

Column G – Age of the company

Column H – Number of employees in the company

Column I – Tags to explain the type of company

As you can see, this dataset gives the average rating of a company, along with some information on its location, size, type etc. You can open the file on Excel by downloading it and double clicking on it if you have Excel installed on your computer. If you do not have Excel, you can open this in Google Sheets by going to sheets.google.com, opening a blank worksheet, clicking on File – Open – Upload, and opening this file.

Excel Functions

One of the most important features of spreadsheets is inbuilt functions. Every function will have a specific operation it will perform. It will also have a syntax in which it must be entered for it to work. If it is not entered correctly, you will get an ERROR.

There are hundreds of functions on Excel which serve different purposes. We will be covering some common functions that are used most commonly in our course.

Writing Functions

Like Operators, functions also start with “=” sign. For the purpose of our example, let's look at the UPPER function. This is a text function on Excel that converts any text to uppercase (all capital letters). Let's enter some random text in a cell.

	A	B	C
1			
2			
3		Pwskills	
4			

Now let's convert this to uppercase using the UPPER function. In another cell, let's start with "="

	A	B	C
1			
2			
3		Pwskills	=
4			

Now let's start typing the function name

	A	B	C	D	E	F	G	H	I
1									
2									
3		Pwskills	=u						
4									
5									
6									
7									
8									

You can see that as soon as you enter the first letter, Excel will start recommending functions to you. It will also show a description of what each function does. Here, you can see UPPER at the bottom of the list. I can use arrow keys to go select UPPER in the list -

	A	B	C	D	E	F	G
1							
2							
3		Pwskills	=u				
4							
5							
6							
7							
8							

Now, you can just press the Tab button, and the function will appear in the cell -

	A	B	C	D
1				
2				
3		Pwskills	=UPPER(
4				UPPER(text)
5				

You can see that it will also show you the syntax – the input required for the function to work.

Here, the text that is required to be converted to uppercase is the input required.

Some advanced functions can have both mandatory inputs and optional inputs. We will see some of those later.

Here, since the text to be converted is in B3, I can enter

=UPPER(B3)

and press Enter.

C3				
	A	B	C	D
1				
2				
3	Pwskills	PWSKILLS		

As you can see, it has been converted to uppercase.

Key things to remember about writing functions

- Start with “=” sign
- Excel will suggest functions for you based on the letter you enter (like we entered U for upper)
- Excel will also explain the operation the function performs
- You can select a function with Tab. At that point Excel will show you the Syntax of that function.

Common Text Functions

- **UPPER**

Syntax: =UPPER(text)

Converts text into uppercase.

Exercise – Convert all the company names in our dataset to uppercase

- **LOWER**

Syntax: =LOWER(text)

Converts text into lowercase

Exercise – convert all tags in our dataset to lowercase

- **PROPER**

Syntax: =PROPER(text)

Capitalizes the first letter in each word of a text value

Exercise – Convert all the Headquarters in our dataset to Proper case

- **LEN**

Syntax: =LEN(text)

Counts the number of characters (alphabets, special symbols etc) in text. Please note, even space, full stop, etc. will be counted as characters.

Exercise – Find the top 10 companies with the longest name in our dataset. Use LEN and sorting.

- **LEFT**

Syntax: =LEFT(text, number of characters)

Extract the left most characters from text. The number of characters extracted is specified within the function.

- **LEFT**

Syntax: =LEFT(text, number of characters)

Extract the left most characters from text. The number of characters extracted is specified within the function.

Exercise – Extract the left 3 characters from all companies in the dataset and arrange them in alphabetical order

- **RIGHT**

Syntax: =RIGHT(text, number of characters)

Extracts the right most characters from text. The number of characters extracted is specified within the function.

Exercise – Extract the right 2 characters from all companies in the dataset and arrange them in alphabetical order.

- **MID**

Syntax: =MID(text, start number, number of characters)

Extracts characters from the middle of text. Start number signifies the first letter that will be extracted, and number of characters specifies how many characters are extracted.

For example, =MID("PWskills, 2,4) will give the output - Wski

As you can see, the 2nd character (start number) is also extracted along with the next characters.

Exercise – Extract the middle 4 characters of all the company names, starting from 3rd character, and arrange them in reverse alphabetical order (Z to A)

- **TRIM**

Syntax: =TRIM(text)

This is an interesting function – During data entry, there will always be some cases where someone enters a space before or after the text by mistake. This can cause problems in cases where you are looking for something specific – let's say "PWSkills", but the entry in the sheet is "PWskills " or " PWskills".

The trim function removes additional spaces before and after a text entry. It will keep spaces within the text entry intact – for example, =TRIM(" PW Skills") will give you the output – "PW Skills".

Exercise – Trim all the company names and headquarters in the dataset.

- **FIND**

Syntax: =FIND(text to find, text to be searched in, [start number])

Here, start number is within brackets because it is optional.

The find function helps returns the starting position of a text within another text. Let's see this with an example –

If a cell A2 has "PWskills" in it, and we want to find where the string "skill" starts in it, =FIND("skill",A2) gives 3 – as "skill" starts in the third position.

This function is case sensitive – which means, if I search for "Skill" in PWskills, it will give an error. (as PWskills contains "skill", and not "Skill")

If you input the start number as well (let's assume 4), the formula starts to check for text to find only from the 4th letter. In the previous case, =FIND("skill",A2,3) will still return 3 – as "skill" starts from the 3rd character.

But =FIND("skill",A2,4) will return an error because Excel will start searching from the 4th character in PWskills – which is "k".

- **SEARCH**

Syntax: =SEARCH(text to find, text to be searched in, [start number])

This is the same as FIND, but is not case sensitive. This means that

=SEARCH("skill",A2) will give 3 as the output, and

=SEARCH("Skill",A2) will also give 3 as the output – whereas FIND would have given an error.

Exercise – In our dataset, find where the letter S starts in each company name

- **SUBSTITUTE**

Syntax: =SUBSTITUTE(full text, text to be replaced, replacement text, instance number)

SUBSTITUTE takes the full text, finds the text to be replaced, deletes that, and puts in replacement text in its place. If the text to be replaced appears multiple times, every instance of it is replaced, unless you specify the instance number.

For example, let's say A2 has the text "PWskills is the best place to gather skills"

=SUBSTITUTE(A2,"skill","great")

Will return the text "PWgreats is the best place to gather greats"

=SUBSTITUTE(A2,"skill","great",2)

Will return "PWskills is the best place to gather greats"

SUBSTITUTE function is also case-sensitive, like FIND function.

This is useful in some cases where the data may have certain unwanted characters in it. For example, if a phone number is recorded in the data as "999-999-0099", and we want to convert it to "9999990099", we can use

=SUBSTITUTE("999-999-0099","-","")

Here, all instances of "-" are identified and replaced with nothing (as there is nothing in the replacement text field).

Exercise - Replace all "IT" in tags column with "Tech" in the dataset

- **REPLACE**

Syntax: =REPLACE (full text, starting position, number of characters to be removed, new text to be added)

This function takes the full text, goes to the starting position, deletes the number of characters to be removed, and adds the new text to be added.

For example, if A2 has the text "PWskills is a good platform to learn Excel",

=REPLACE(A2,15, 4, "great")

Will give you "PWskills is a great platform to learn Excel"

Exercise - In the company names, replace the 3rd and 4th characters with "PWskills"

- **CONCATENATE**

Syntax: =CONCATENATE(text1, text2, text3...)

This function takes text1, adds text2 at the end of it, then adds text3 at the end of text2, and so on.

This is useful in multiple cases. One usecase is when you want to add some text before or after an existing text string.

If A2 contains "PWskills",

=CONCATENATE(A2," is great") will give the output

"PWskills is great"

Another example is if one cell contains the first name of a person, and the other cell contains the second name -

Let's say A2 contains "Ranveer" and A3 contains "Singh",

=CONCATENATE(A2, " ",A3)

Will give you "Ranveer Singh" - You can see we have used " " as text2 to add a space in between to get the right output.

Exercise - Use CONCATENATE to add the phrase "is great" to all the company names in our sheet ("TCS" should becomes "TCS is great")

This covers most of our text functions.

Exercise

Now let's take a new copy of our dataset and perform the following operations on it

- Convert Name, Company_type, Headquarter, and Tags to Proper case
- TRIM any extra space in these fields
- Remove the string "Employees (India)" in No. of Employees (replace it with "")
- Substitute "Govt" in type Company_type "Government"
- Extract the first 3 letters of all company names
- Identify the longest Company name in the dataset

Additional tips – Shortcuts

Ctrl + F4- It is used to close Excel.

Ctrl + Z- It is used to undo the applied/ previous changes when the file has not been closed.

Ctrl + Y- It is used to redo the applied/ previous changes when the file has not been closed.

Ctrl + W- It is used to close a workbook.

Ctrl + F- It is used to launch a 'Find and Replace' dialog where the find tab is a default selected tab.