How I use Excel as a Data Analytics Student

Excel is a really valuable tool for just about anyone, regardless of your field or profession. I have used excel as an undergrad, as a professional IT business analyst, but also in my day-to-day work from planning my budget, making reading lists to organizing my books collection. It has a ton of cool features but most of us end up using very basic and limited number of features.

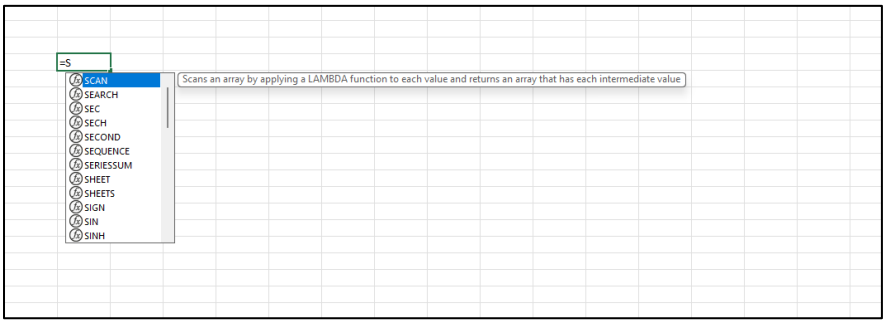
When I started learning data analytics, I realized early on that Excel is one tool that I should master. From then onwards I started researching and self-learning things that a data analyst should know. I opted for the 80-20 rule here where I started looking for the 20% Excel features that are been used to complete 80% of data analytics tasks. This article is all about how I self-learned Excel and reached a point where I was comfortable and confident enough to tackle data analytics related tasks using Excel. Please note that this is just my journey and not a comprehensive learning guide. There's a vast area of knowledge that I'm yet to acquire, and this is but just a little bit of knowledge that I managed to gain along the way.

Hope it helps...

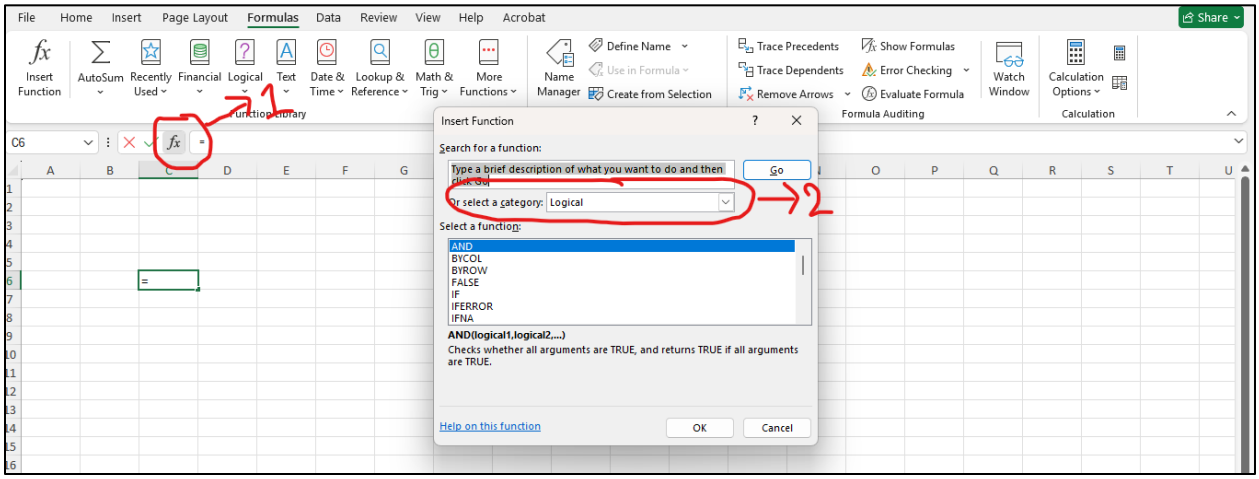
1.1 Functions

Excel has more than 450 built in functions, ranging from simple arithmetic operations like SUM to complex functions like VLOOKUP. These functions cover different categories like text, logic, math, stats, time functions etc. but most frequently we use simple functions like SUM, AVERAGE, IF or SUMIF among others.

Usually, you start a function by tying equal sign in a cell, followed by the function name and the arguments.



This doesn't mean that one should remember all the functions by name to use it. As shown in the below screenshot, you can click on the functions button to open up the insert function popup. You can select the required category from the dropdown and view the available functions under that category.



However, like I said before, you may not even need to use all these functions for your work. Below I've listed the most frequently used excel functions that are worth remembering.

Function	Formula	Use Case
SUM	`=SUM(A1:A10)`	Adds up values
AVERAGE	`=AVERAGE(A1:A10)`	Finds mean value
IF	`=IF(A1>50, "Pass", "Fail")`	Conditional logic
VLOOKUP	`=VLOOKUP(101, A2:D10, 2, FALSE)`	Searches for a value in a table
HLOOKUP	`=HLOOKUP(101, A1:D4, 2, FALSE)`	Searches for a value **horizontally**
INDEX	`=INDEX(A1:C3,2,3)`	Returns a value from a table
MATCH	`=MATCH(50, A1:A10, 0)`	Finds the position of a value
LEN	`=LEN(A1)`	Counts characters in a cell
TRIM	`=TRIM(A1)`	Removes extra spaces from text
LEFT/RIGHT/MID	`=LEFT(A1,3)`, `=RIGHT(A1,3)`, `=MID(A1,2,3)`	Extracts part of a text
CONCATENATE	`=A1 & " " & B1`	Joins text

For data analytics, I've found the below functions to be most useful.

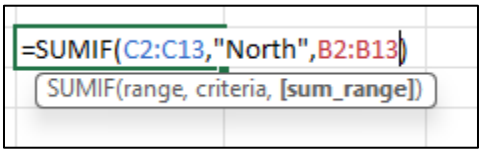
Function	Formula	Use Case
SUMIF	`=SUMIF(A:A, "Region1", B:B)`	Conditional sum
COUNTIF	`=COUNTIF(A:A, "CategoryA")`	Count matching values
IF	`=IF(A2>100, "High", "Low")`	Conditional logic
VLOOKUP	`=VLOOKUP(101, A2:D10, 2, FALSE)`	Lookup data
INDEX-MATCH	`=INDEX(B:B, MATCH(101, A:A, 0))`	Advanced lookup
TEXT()	`=TEXT(TODAY(), "DD-MMM-YYYY")`	Date formatting
LEN()	`=LEN(A1)`	Count characters

Let's do a practice task:

- Create a sales dataset with columns: **Product, Sales, Region**.
- Use **SUMIF** to get total sales for a specific region.
- Use **COUNTIF** to count how items have >$10000 sales.

Solution:

Product	Sales($)	Region
Notebook	12000	North
Blue pen	8000	North
Black pen	7000	South
Pencil	10000	South
Eraser	5000	West
Ruler	9000	West
Notebook	10000	East
Blue pen	8000	East
Black pen	7000	North
Pencil	8000	North
Eraser	3000	South
Ruler	5000	South



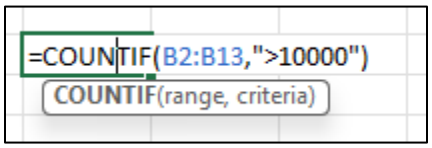
Here C2:C3 check the region column,

"North" is the criteria/ condition to check

B2:B13 check the sales column to get the sum of sales in the North region.

Answer should be $35000

COUNTIF solution:

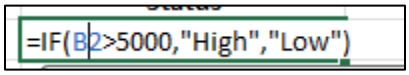


Here the range is the column where the condition is checked. “>10000” is the condition we are checking.

Below I have recorded few other functions essential for data analytics,

1.1.1 IF

I’m taking the same dataset as above. I’m creating a new column called status. The value for that column should be if the sales amount is more than 5000, status is high. If the sales amount is less than 5000, then the status is low. I’m going to use the IF function for that.



After I get the result for the first row, I’m just going to drag it up to the end, which gave me the required result as shown below.

Product	Sales($)	Region	Status
Notebook	12000	North	High
Blue pen	8000	North	High
Black pen	7000	South	High
Pencil	10000	South	High
Eraser	5000	West	Low
Ruler	9000	West	High
Notebook	10000	East	High
Blue pen	8000	East	High
Black pen	7000	North	High
Pencil	8000	North	High
Eraser	3000	South	Low
Ruler	5000	South	Low

1.1.2 VLOOKUP

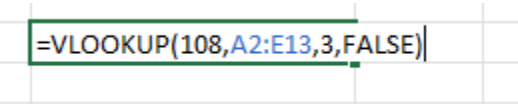
This function is used to look for a value in the first column of a table and returns a value from another column. I’m going to add a new column to our dataset called Product ID

Product ID	Product	Sales($)	Region	Status
100	Notebook	12000	North	High
101	Blue pen	8000	North	High
102	Black pen	7000	South	High
103	Pencil	10000	South	High
104	Eraser	5000	West	Low
105	Ruler	9000	West	High
106	Notebook	10000	East	High
107	Blue pen	8000	East	High
108	Black pen	7000	North	High
109	Pencil	8000	North	High
110	Eraser	3000	South	Low
111	Ruler	5000	South	Low

Task:

Find the total sales amount of product id 108.

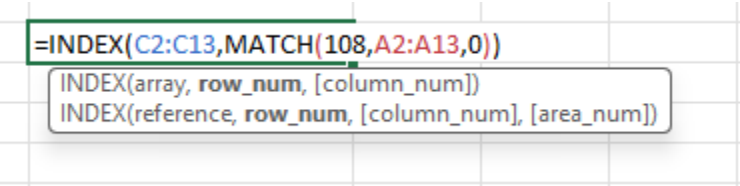
Solution:



Here 108 is the value to look for in the first column, next up is the search range which is usually the range of the table, 3 is the index number of the column that the return value is (in this case the sales column since we are looking for the sales amount of product 108), followed by TRUE/ FALSE. Here TRUE means approximate match and FALSE means exact match.

1.1.3 INDEX-MATCH

This function can be used as an alternative to VLOOKUP. If we take the same task as above,



INDEX – here the range is of the sales column.

MATCH – this checks for the value 108 in the product id column. 0 means it is looking for the exact value.

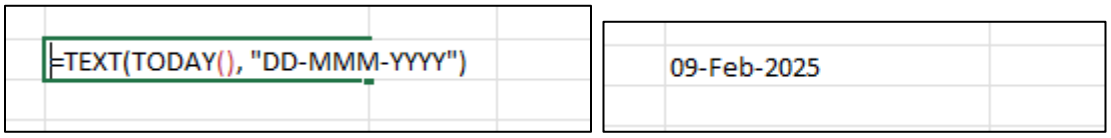
1.1.4 TEXT()

The TEXT() function formats numbers, dates, and text in a specific way.

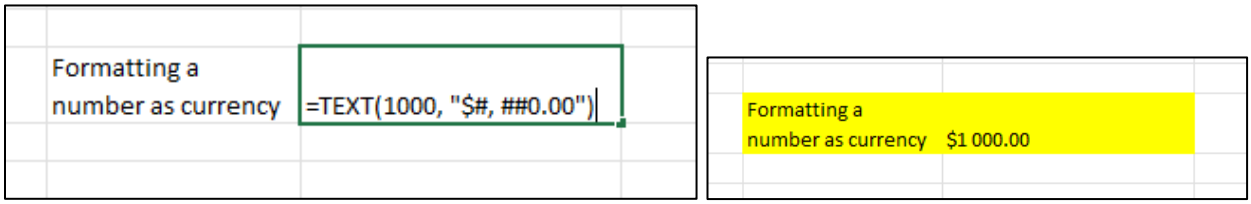
Task:

Format todays date in the “DD-MMM-YYYY” format

Solution:

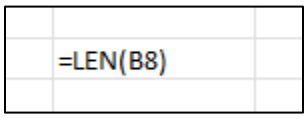


Another example is formatting a number as currency,



1.1.5 LEN()

The LEN() function counts how many characters (including spaces) are in a cell.



I did a ton of practice tasks for these functions to get a better hang of it and to be quick and efficient in using these functions. Once I got familiar with these, I started just trying out other functions as well. Functions are one of the most important features in Excel to make your life easier, specially for data analysts. So it’s a really handy skill to have.

1.2 Data Cleaning and Formatting

Data cleaning is a critical step in data analysis to remove errors, inconsistencies, and missing values. It ensures that datasets are accurate, reliable, and ready for analysis.

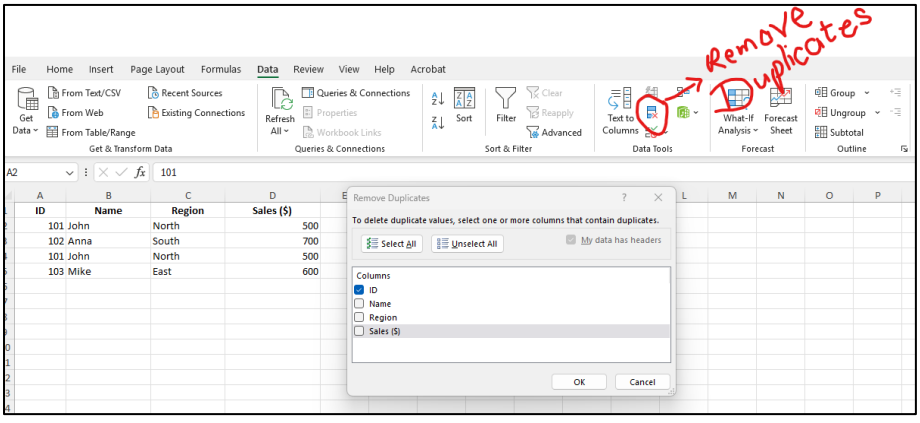
Following are some common data cleaning tasks I've found useful,

- Removing duplicates
- Fixing Inconsistent Formatting (Text, Numbers, Dates)
- Handling Missing Data (Blanks, NAs, Errors)
- Splitting & Combining Data (Text to Columns, CONCATENATE)
- Trimming Spaces & Removing Non-Printable Characters
- Standardizing Case (UPPER, LOWER, PROPER)

1.2.1 Removing Duplicates

ID	Name	Region	Sales ($)
101	John	North	500
102	Anna	South	700
101	John	North	500
103	Mike	East	600

- First select the dataset
- Go to DATA > REMOVE DUPLICATES
- Select which columns to check
- Click ok



ID	Name	Region	Sales ($)
101	John	North	500
102	Anna	South	700
103	Mike	East	600

1.2.2 Fixing Inconsistent Formatting (Text, Numbers, Dates)

- Select the data.
- Go to Home → Number Format and choose the correct format (Currency, Date, Number).

Here you can utilize your knowledge on functions (like TEXT()) to better format the data.

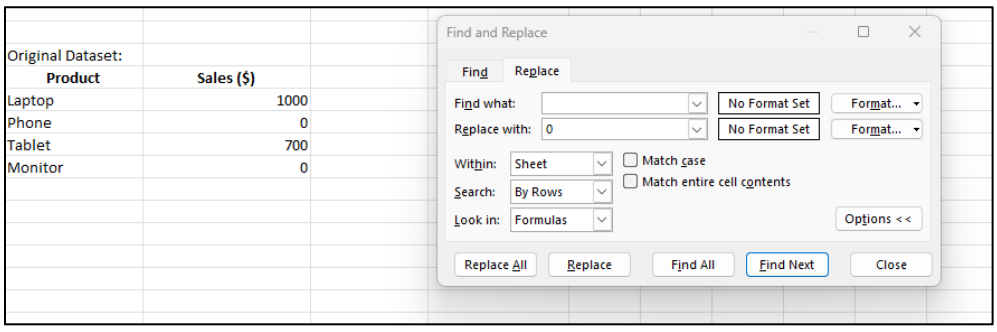
1.2.3 Handling Missing Data (Blanks, NAs, Errors)

We can use find and replace (CRTL+H) to find empty cells and assign a value like “N/A” to those cells.

IFERROR() function can be used to handle errors.

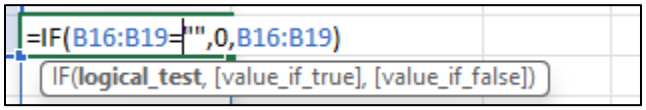
Product	Sales ($)
Laptop	1000
Phone	#VALUE!
Tablet	700
Monitor	

If we take the above dataset, we are going to use find and replace to replace the empty values with 0.

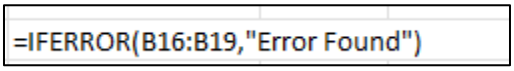


If you want to find empty cells then leaven the find what field empty.

Alternatively, we can use IF function to accomplish this task.



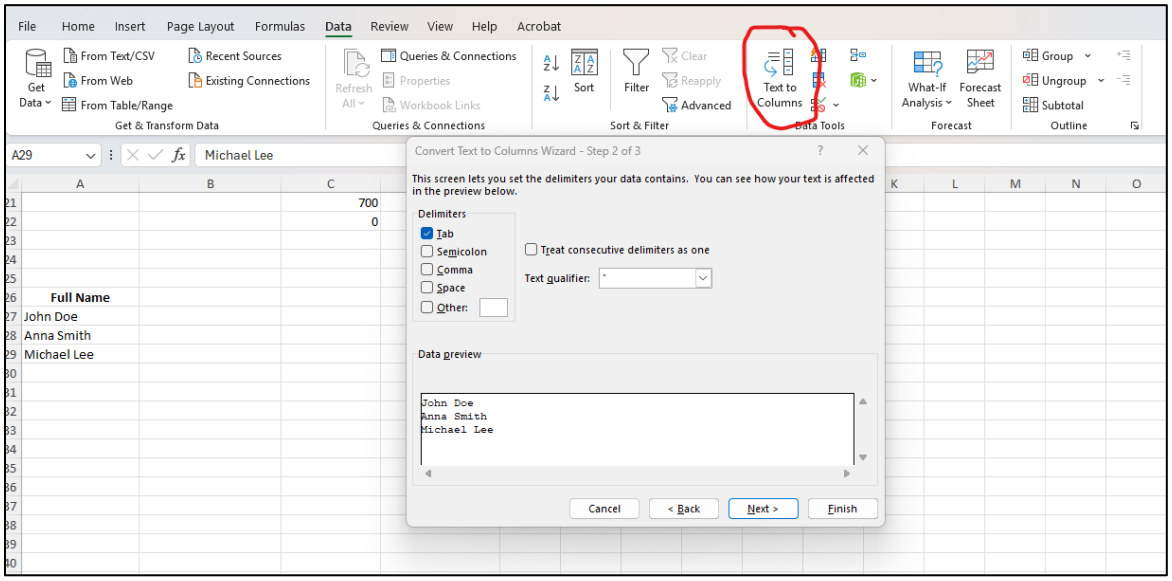
This will replace all empty cells with 0



1.2.4 Splitting & Combining Data (Text to Columns, CONCATENATE)

Full Name
John Doe
Anna Smith
Michael Lee

Splitting is a really useful feature in data analytics. Let's see how we can split the above dataset into two columns, first name and last name.



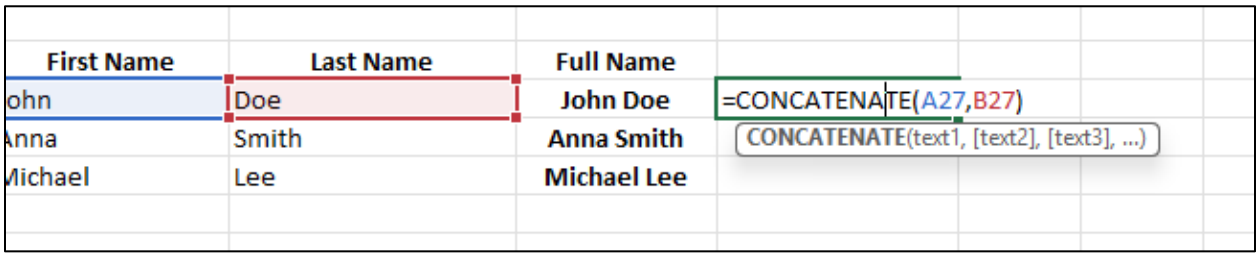
First Name	Last Name
John	Doe
Anna	Smith
Michael	Lee

Similarly, you can combine data using the following features

- &
- CONCATENATE() function

Let's create a new column named full name and combine the first name and last name.

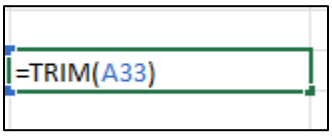
First Name	Last Name	Full Name
John	Doe	=A27 & " " &B27
Anna	Smith	
Michael	Lee	



1.2.5 Trimming Spaces & Removing Non-Printable Characters

Name	
John	Doe
Anna	Smith

As you can see in the above dataset there are quite a lot of unwanted spaces. We can use the TRIM() function to remove these unwanted spaces.



Name	After Trimming
John Doe	John Doe
Anna Smith	Anna Smith

Non-printing characters are symbols that are not visible when printed or displayed on a screen. They are used to control the formatting, spacing, and processing of text and data. If your dataset includes such characters, you might want to remove those before analyzing the data.

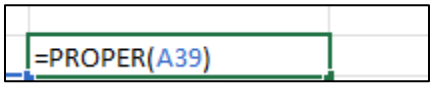
You can use the CLEAN() function to remove such non printable characters.

1.2.6 Standardizing Case (UPPER, LOWER, PROPER)

- UPPER() – converts text to uppercase
- LOWER() – converts text to lowercase
- PROPER() – converts text to proper case

The first two functions are pretty straightforward. The PROPER() function works like this. Lets take this dataset into consideration,

Name
john doe
ANNA SMITH



By using the PROPER() function, the second name is formatted into proper writing.

Name	After changing
john doe	John Doe
ANNA SMITH	Anna Smith

These are the very basics of data cleaning with Excel. If you are comfortable with these features, then you can spend some time on Excel and figure out what other amazing features

are there that can help with data analytics. Personally, I find these features to be the most valuable for a beginner data analyst like me.

1.3 Pivot Tables and Data Summarization

When I started learning excel as a beginner, I haven't heard of pivot tables. As I started deep diving into the world of Excel, I found out that pivot tables not only are useful for data analysts handling very large datasets, but also useful for someone who is trying to manage their finances at home or trying to make sense of the data they've collected from their small home business.

Pivot tables let you quickly summarize, analyze and visualize large datasets. They will help you answer questions like

- Total sales by the region?
- Average salary per department?
- What are your 5 top performing products?

So, what is a pivot table?

A pivot table is a dynamic summary of a dataset that helps analyze trends and relationships.

Common uses of pivot tables,

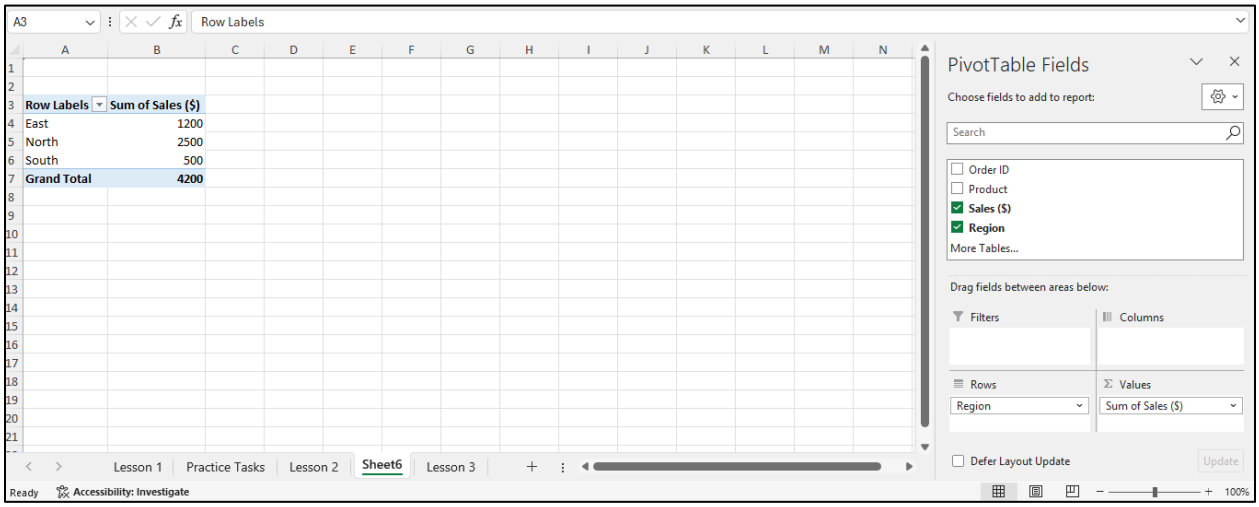
- Sum, count, or average data quickly.
- Group and filter data dynamically.
- Compare categories (e.g., Sales by Region).

How to create a pivot table?

Order ID	Product	Sales ($)	Region
101	Laptop	1000	North
102	Phone	500	South
103	Tablet	700	North
104	Laptop	1200	East
105	Phone	800	North

I have the above dataset and I'm going to create a pivot table for it below.

- Select the dataset
- Insert > pivot table
- Choose new worksheet
- Click ok
- Drag and drop fields as needed
- Click ok



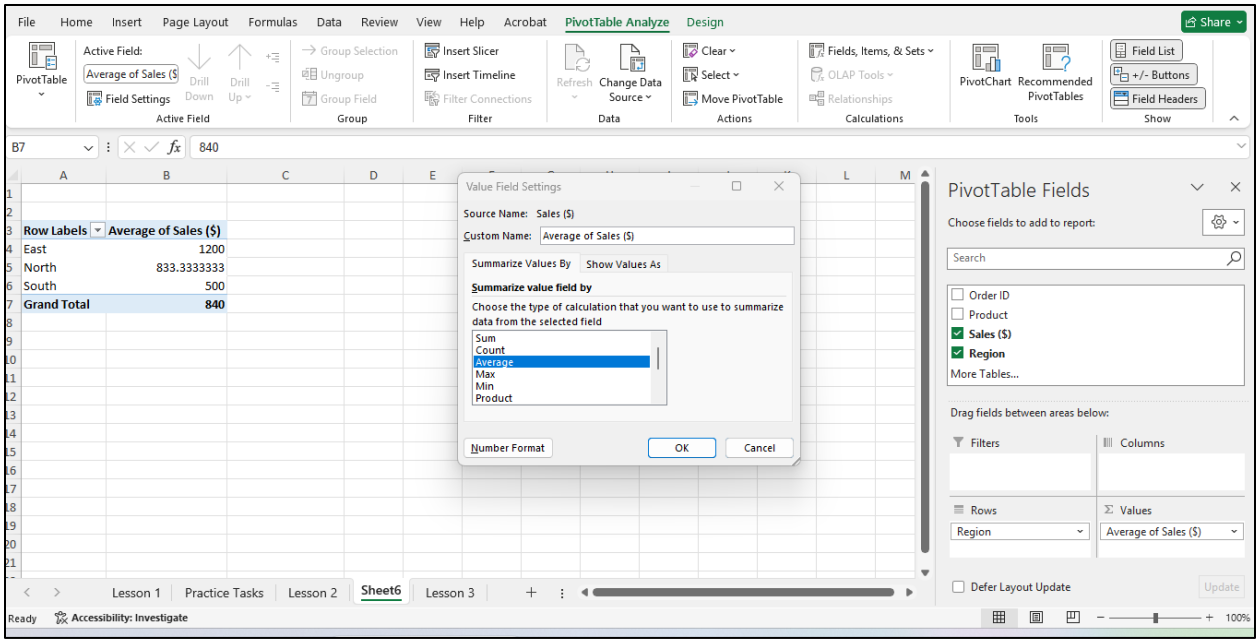
This pivot table has given us the total sales by region.

1.3.1 Pivot Table Features

As discussed above, pivot tables have certain features. Below we are going to look at the different features and how to use them to better analyze your data.

The first feature is **summarizing your data.** We already saw how to get the sum of a particular column using pivot tables. In addition to that we can get the average and the count. Let's see below how we can get these different values.

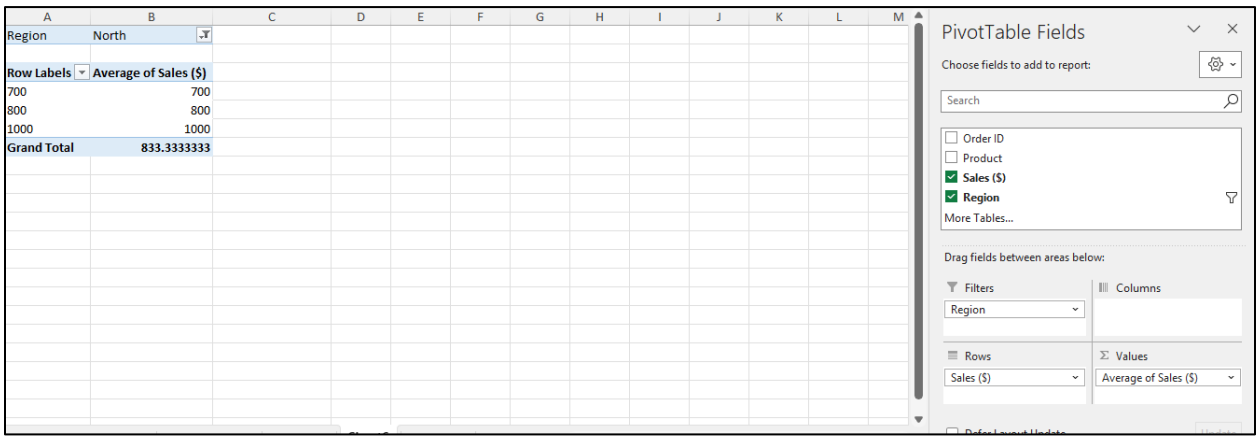
- Click on the pivot table
- Go to value field settings
- Choose sum, count or average



Here when I select average from the field settings, I get the average of sales.

Second feature is **filtering data.** As you may have already seen, there's an area called filters in the pivot table fields section. Let's say I wanted to see the records of only the North region.

I'm going to drag region into the filters area and select only North.



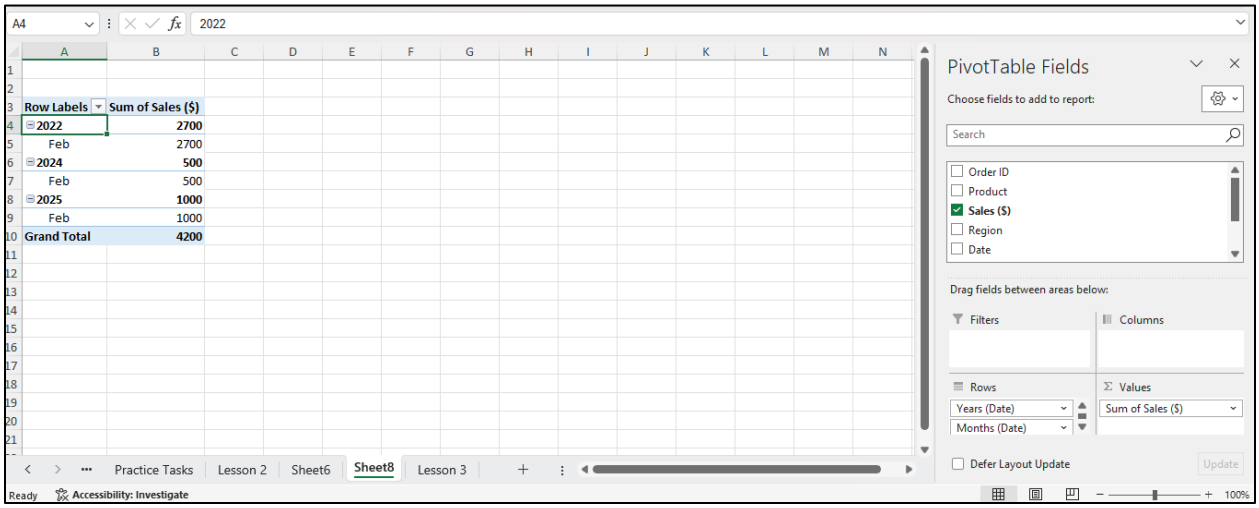
Now I can see the records only where the region is North.

The third feature is **grouping data.** If your dataset has dates then you can group sales by month or year. I've added a new column to our dataset called date.

Order ID	Product	Sales ($)	Region	Date
101	Laptop	1000	North	01/02/2025
102	Phone	500	South	02/02/2024
103	Tablet	700	North	03/02/2022
104	Laptop	1200	East	04/02/2022
105	Phone	800	North	05/02/2022

What we are trying to do here is to group our dataset according to date, it could be by year, month or week or even quarterly.

- First select the entire dataset and create a pivot table
- Drag date into the Rows area
- Drag sales into the values area
- Right click on any date on the pivot table and select group
- Select months/ years



Next up we are going to look at pivot charts, which is a method of data visualization. We are going to talk in detail on data visualization in the next section, but I thought I should introduce pivot charts now since we are already on the subject.

Basically, pivot charts help you visualize pivot table data.

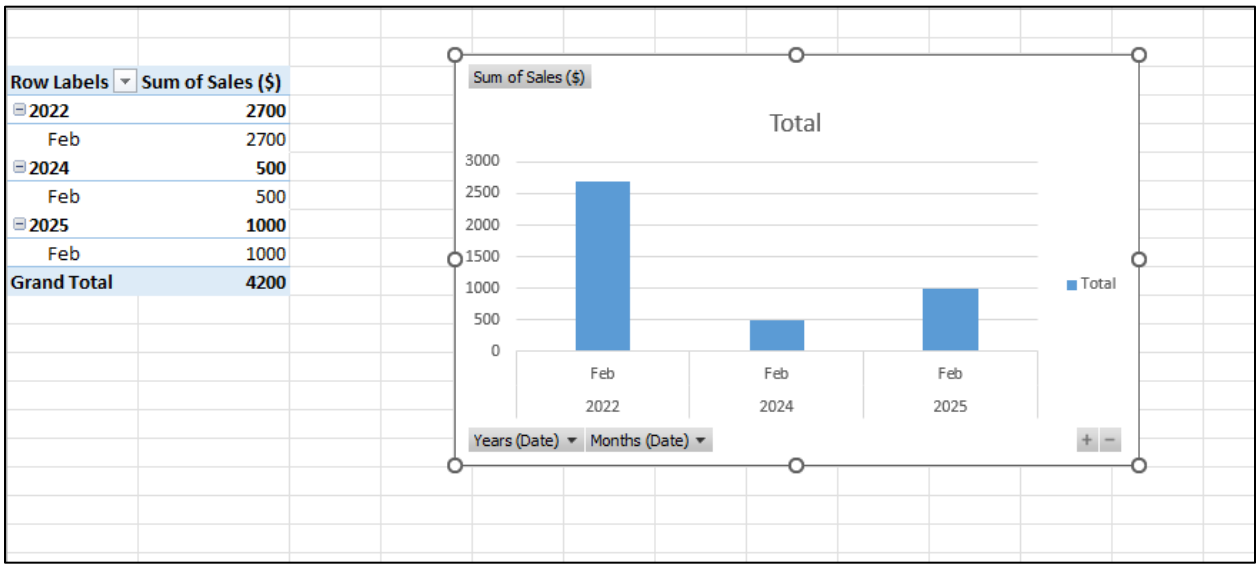
It has,

- Column charts
- Line charts

- Pie charts

Here's how you can insert a pivot chart,

- Click inside the pivot table
- Go to insert > pivot chart
- Select column chart
- Click ok



Here you have the total sales by the year and month.

1.4 Data Visualization

Data visualization helps present data in a clear, visually appealing way for better understanding and decision-making. Charts transform raw numbers into insights, making it easier to spot trends, comparisons, and patterns.

Why should we use charts?

- Compare data
- Show trends
- Highlight differences
- Identify patterns

Charts help us to communicate and express our findings to a wider audience in an effective way.

There are several types of charts in Excel.

Choosing the correct type of chart makes a huge difference in presenting your data. Here I have a summary of commonly used charts and when to use them.

Chart Type	Best For
Column Chart	Comparing categories (e.g., Sales by Product)
Line Chart	Showing trends over time (e.g., Website Traffic)
Pie Chart	Displaying percentage distribution (e.g., Market Share)
Bar Chart	Similar to column but horizontal (e.g., Survey Results)
Scatter Plot	Showing relationships between two variables (e.g., Age vs Salary)

Let's look at an example. I'm going to use the following dataset.

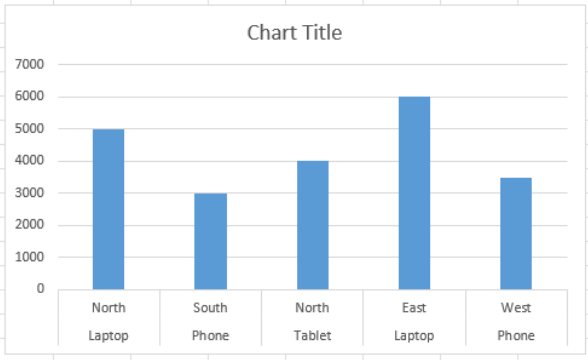
Product	Region	Sales ($)
Laptop	North	5000
Phone	South	3000
Tablet	North	4000
Laptop	East	6000
Phone	West	3500

Im going to create a bar chart for this dataset,

- Select the data range
- Go to insert > choose a chart type
- Click ok

You can customize the chart, add a title, modify the appearance once the chart is created.

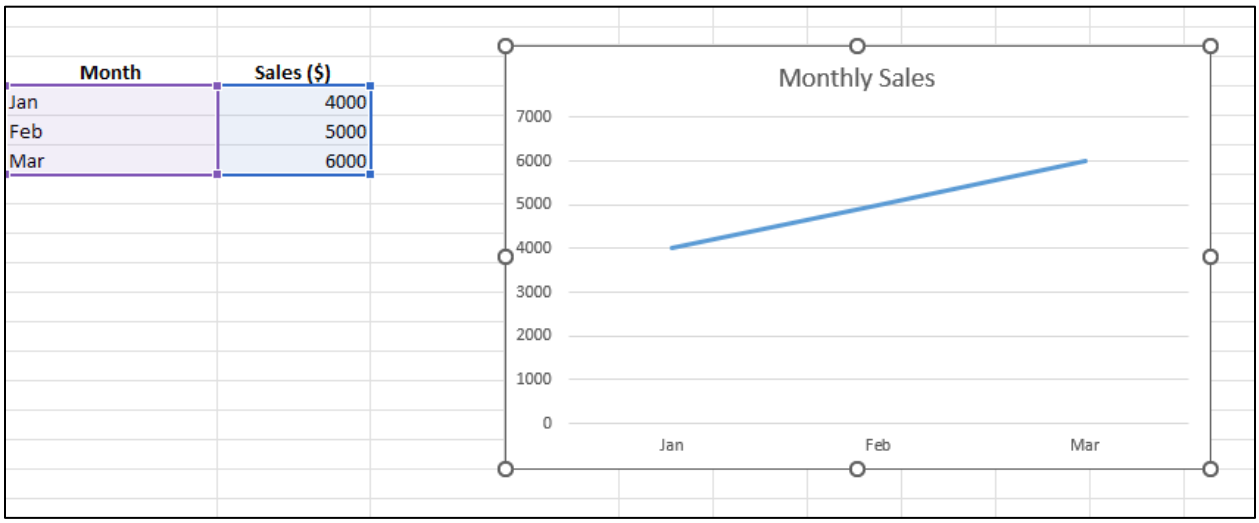
Product	Region	Sales ($)
Laptop	North	5000
Phone	South	3000
Tablet	North	4000
Laptop	East	6000
Phone	West	3500



Column charts are best for comparing categories. For example, sales by product, revenue by region etc.

Another type of chart is line chart which is best for showing trends over time. Lets take the below dataset,

Month	Sales ($)
Jan	4000
Feb	5000
Mar	6000

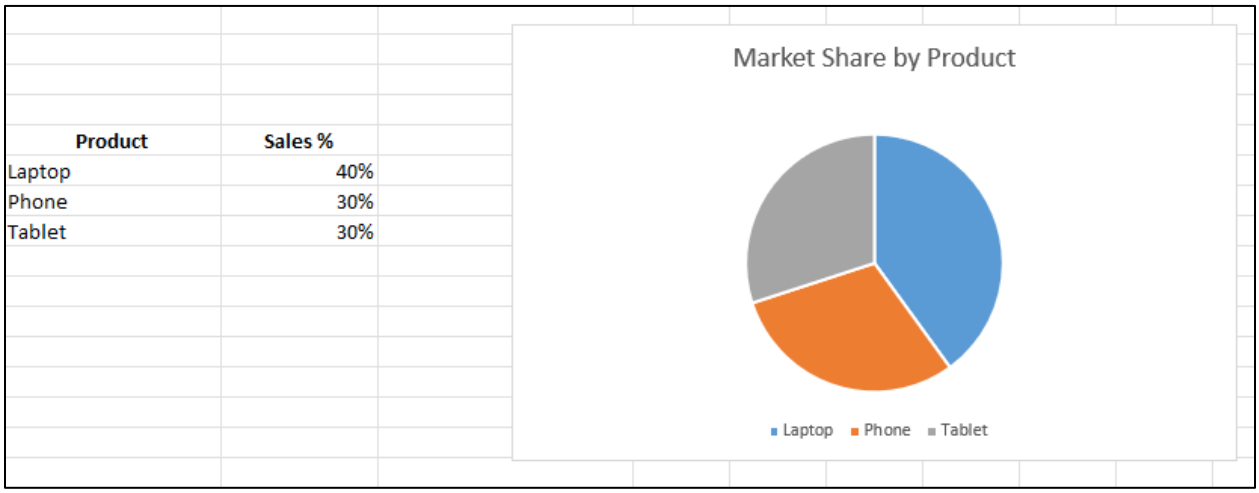


A line chart helps to visualize the trend in sales as the months progresses. In shows a steady incline which means the sales have increased over the months. Now imagine this been applied to a large data set with years and years of data. By visualizing the data, it helps to understands patterns quickly and accurately.

Another type of chart is pie chart. This is best to show percentage breakdowns such as market share, sales contribution etc.

Lets consider the below dataset,

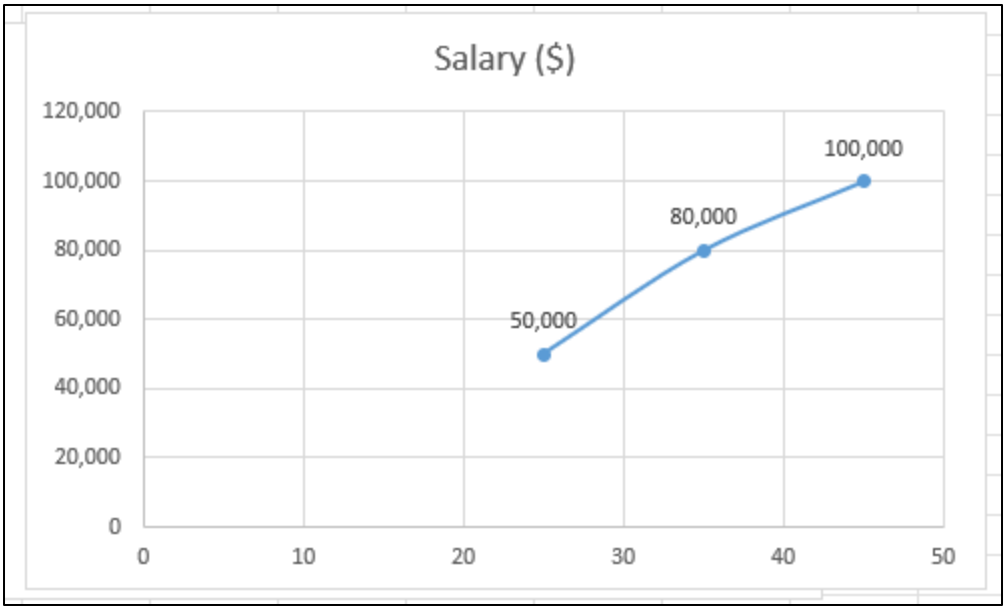
Product	Sales %
Laptop	40%
Phone	30%
Tablet	30%



Another type of chart is scatter plot. I think this chart is best for comparing two variables to find a relationship. Let's take the below dataset,

Age	Salary ($)
25	50,000
35	80,000
45	100,000

- Select the x and y axic values (in the above dataset there are only two columns so you can choose those two)
- Click insert > Scatter (select the required chart)
- Click ok



1.5 Summary

In this article we looked at the basics of Excel that is useful for data analytics. We have covered the most basics such as,

- Understanding cells, rows, columns
- Keyboard shortcuts for faster workflows
- Formatting data

One of the most important features covered is the different types of functions in Excel. Going a bit beyond the basic functions like SUM(), AVERAGE(), we looked at more useful functions like SUMIF, COUNTIF, VLOOKUP etc. that are most useful for a data analyst.

Moving on to more data related features in Excel, we looked at pivot tables and the features of pivot tables which is valuable in summarizing, filtering and grouping data quickly. Then we looked at data cleaning and formatting using various features in Excel. The more common use cases of this been standardizing data, removing extra spaces and fixing messy text. Basically, these features help you to prepare the data before doing proper analysis.

Then we looked at data visualization using Excel charts. As we saw earlier, there are many types of different charts in Excel and choosing the correct chart to show your data is a vital step in successfully visualizing your data.

I hope this article will help many people to start learning the basics of Excel. Not only for data analysts, this tool is valuable to everyone as its truly a handy tool that can help you with your day-to-day tasks as well as complex business-related tasks.