Excel for Public Health Students

Razi N

Table of contents

Preface

This handbook is for public health professionals and researchers who want to use Excel more effectively in their work. It focuses on practical tasks like organizing data, using basic formulas, and creating simple charts without going too deep into complex features.

The aim is to keep things clear and easy to follow. Many functions are explained in a simplified way, so that even those with limited experience can understand and apply them confidently.

You can explore the handbook online at https://razinoushad.github.io/excel-handbook/, where all examples and files are available.

I hope this guide helps you use Excel more comfortably in your public health work.

1 Getting Started

1.1 Introduction

- Excel is the most used spreadsheet software and a part of Microsoft Office suite.
- Excel skill is very much required for a public health professional for their day to day activities.
- Some uses
  - Crunching numbers
  - Creating charts
  - Organizing lists
  - Manipulating text
  - Access data from other sources

1.2 Workbook and Worksheets

- A workbook is a file that you create in Excel (default file extension is `.xlsx`)
- Each workbook contains one or more worksheets
- Each worksheet contains individual cells
- There is an invisible layer on top of each worksheet which holds the charts, diagrams, pictures, etc

1.3 Data Types

- There are four basic types of data in Excel
  - A numeric value
  - Text
  - A formula
  - An error
- Numeric value can be numbers, dates, time, amount,...
- Formulas are what makes spreadsheet a spreadsheet, more on this later.

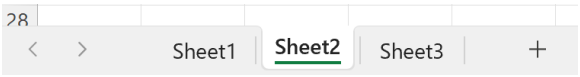
1.4 Number Formatting

- Formatting numbers make data more understandable
- Commonly used number formats:
  - General
  - Number
  - Currency
  - Date
  - Time
  - Percentage
- Number formatting can be changed by clicking Number format dropdown.

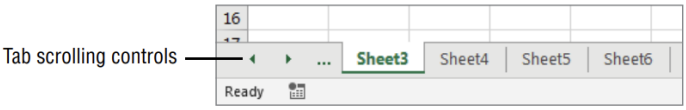
- Number of decimals can be increased or decreased from the command next to Number format dropdown.

1.5 Worksheet Operations

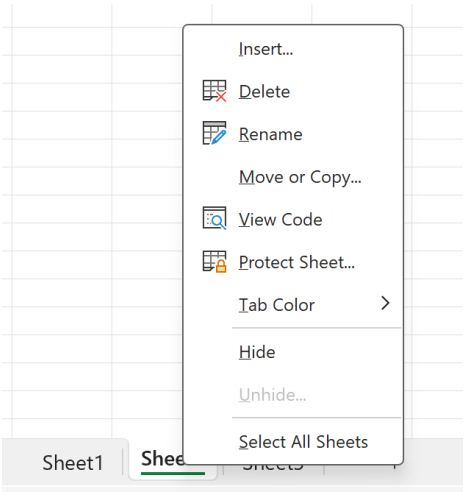
- Sheet tab in the bottom left corner lists all the worksheets(except hidden)



- If there are many worksheets in the workbook, all the sheets may not be visible. Use the tab scrolling control to see the sheet tabs.



- Right click on the sheet tab to see the available operations.



1.6 Rows and Columns

- There are exactly 1,048,576 rows and 16,384 columns in every worksheet.
- Row numbers are at the left side of the sheet.
- Column numbers are at the top of the sheet.
- To insert new row or rows, select an entire row or multiple rows by clicking the row numbers in the worksheet border. Right-click and choose Insert.
- To insert new column or columns, select an entire column or columns by clicking the column letters in the worksheet border. Right-click and choose Insert.
- Excel will insert exactly the number of rows or columns you selected.
- To delete rows or columns, select the rows or columns to remove from the worksheet border, right-click and choose Delete.

1.7 Cells and Ranges

- A cell is a single element in a worksheet that can hold a value, some text, or a formula.
- A cell is identified by its address, which consists of its column letter and row number.
- For example, cell `D9` is the cell in the 4^{th} column and the 9^{th} row.
- A group of one or more cells is called a range.
- You designate a range address by specifying its upper-left cell address and its lower-right cell address, separated by a colon.
- For example, `A1:A100` is the range which contains 1^{st} to 100^{th} row in the 1^{st} column.

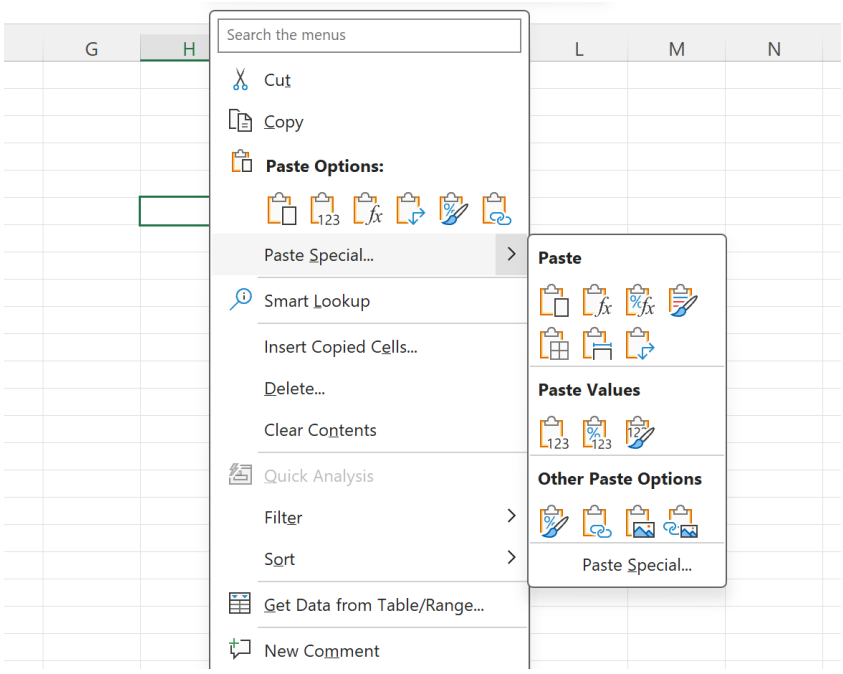
1.8 Copying and Moving Cells or Ranges

- Copying or moving consists of two steps:
  1. Select the cell or range to copy (the source range), and copy it to the Clipboard. To move the range instead of copying it, cut the range instead of copying it.
  2. Select the cell or range that will hold the copy (the destination range), and paste the Clipboard contents.

> **Caution**
>
> When you paste the information, Excel overwrites any cells that is in the destination range without any warning.

1.8.1 Pasting in special ways

- To control what is pasted into the destination range, right-click and choose `Paste Special`. There are multiple options available



Some useful paste options are as follows:

Select	To Paste
Paste	All cell contents and formatting, including linked data
Formulas	Only the formulas
Formulas & Number Formatting	Pastes formulas and number formatting only
Keep Source Formatting	All cell contents and formatting
Transpose	Reorients the content of copied cells when pasting. Data in rows is pasted into columns and vice versa
Values	Pastes the results of formulas
Values & Number Formatting	Pastes the results of formulas plus the number formatting
Values & Source Formatting	Pastes the results of formulas plus all formatting
Formatting	Pastes only the formatting of the source range
Paste Link	Creates formulas in the destination range that refer to the cells in the copied range

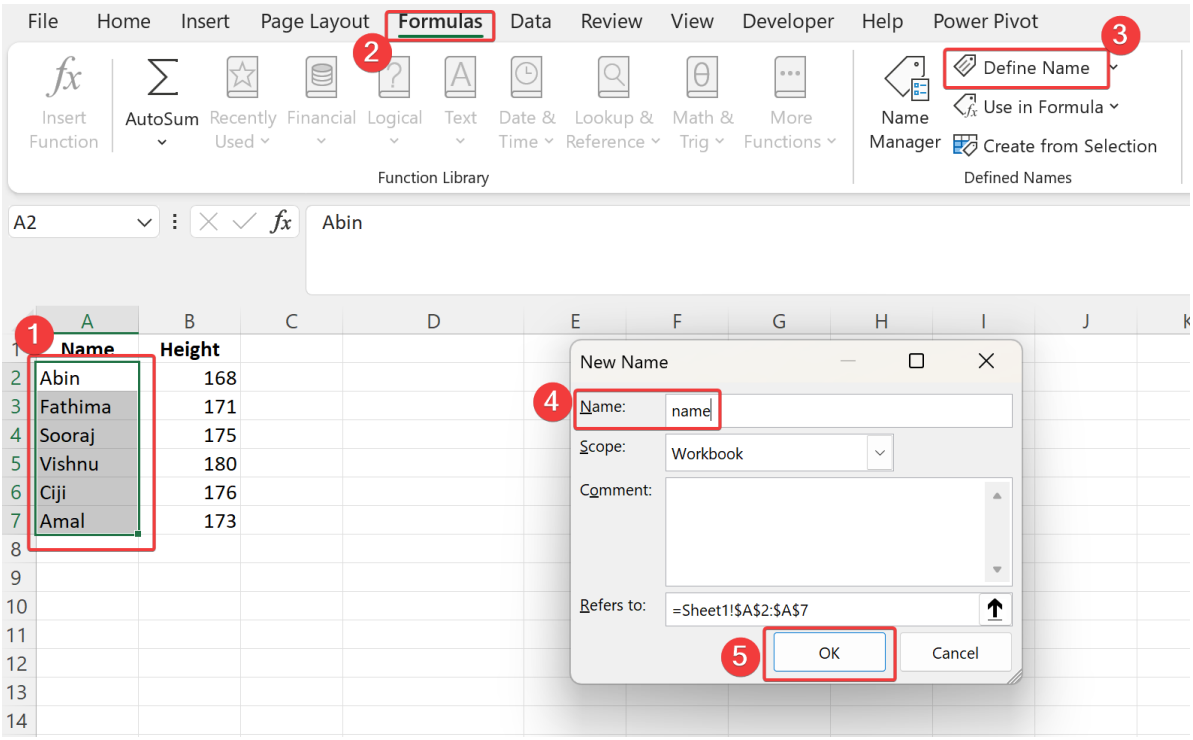
1.9 Named Ranges

Named ranges in Excel are like shortcuts for your data. They allow you to assign a meaningful name to a specific cell or range of cells, making it easier to reference them throughout your workbook. This feature not only streamlines your formulas but also enhances the readability and organization of your spreadsheets. Let's see how named ranges work and explore their practical applications through an example.

1.9.1 Creating Named Ranges

To create a named range in Excel, follow these simple steps:

1. **Select the Cells**: Highlight the cell or range of cells you want to name.
2. **Name the Range**: Go to the "Formulas" tab on the ribbon, click on "Define Name," and enter your desired name for the range.
3. **Confirm**: Click "OK" to save the named range.



Tip

The fastest way to create a name is to use the Name box (to the left of the Formula bar). Select the cell or range to name, click the Name box, and type the name. Press Enter to create the name.

1.9.2 Advantages of Named Ranges

- **Readability**: Instead of referring to cells by their coordinates (e.g., A1:B10), you can use descriptive names that convey the purpose of the data.
- **Ease of Use**: A meaningful range name (such as Hb_level) is much easier to remember than a range address
- **Flexibility**: They make your spreadsheet more dynamic as named ranges automatically adjust when you insert or delete rows and columns.

1.9.3 Example

Suppose you have the following data:

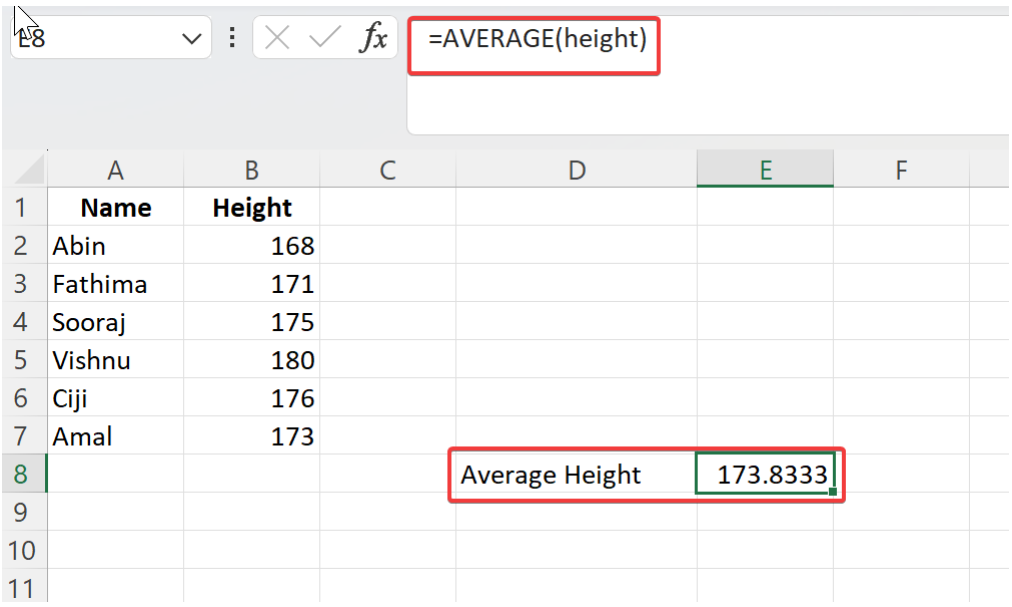
	A	B	C
1	**Name**	**Height**	
2	Abin	168	
3	Fathima	171	
4	Sooraj	175	
5	Vishnu	180	
6	Ciji	176	
7	Amal	173	
8			

Instead of referencing cell ranges directly in your formulas, you can name the ranges for better clarity. Let's name cell range A2:A7 as "names" and B2:B7 as "height"

- **Define Named Ranges**:
  - names: `=$A$2:$A$7`
  - height: `=$B$2:$B$7`

Now, to calculate the average heights, you can simply use the formula:

```
=AVERAGE(height)
```



1.10 Tables

Tables in Excel are powerful tools for organizing and analyzing data efficiently. With their flexibility and functionality, tables can significantly enhance your data management capabilities. In this section, we'll see how to add a table, sorting and filtering data within a table, and changing the appearance of tables in Excel.

1.10.1 Adding a Table:

Adding a table in Excel is a straightforward process:

1. **Select Your Data**: Highlight the range of cells containing your data. This can include headers if your data has them.
2. **Insert a Table**: Navigate to the "Insert" tab on the Excel ribbon. Click on the "Table" button. Excel will automatically detect the range of cells you selected and prompt you to confirm it. Ensure the "My table has headers" option is checked if your data includes headers.
3. **Customize Your Table**: After inserting the table, you can modify its design and structure using the "Table Design" tab that appears when the table is selected. Here you can change the table style, add or remove columns, and make other adjustments to suit your preferences.

> Tip
>
> You can use the keyboard shortcut `Ctrl + T` to insert a table from the selected range of cells.

1.10.2 Sorting and Filtering Data:

Once you've created a table, sorting and filtering data becomes incredibly efficient:

1. **Sorting Data**: To sort data within a table, simply click on the dropdown arrow in the header of the column you want to sort by. Choose either "Sort A to Z" or "Sort Z to A" based on your preference. Excel will instantly rearrange the rows according to your selection.

   Example: Suppose you have a table of data about basic details of people with columns for "Name", "Age", and "Sex". You can sort the table by "Age" in descending order to see the oldest people at the top.

2. **Filtering Data**: Filtering allows you to display only specific rows of data based on criteria you define. Click on the dropdown arrow in the header of the column you want to filter. You can then select specific values to display or use custom filter options for more complex filtering conditions.

   Example: In a table containing patients information, you can filter the "Department" column to show only patient from the "Cardiology" department.

1.10.3 Naming a Table

Assigning names to tables in Excel is a straightforward process, yet its implications are profound.

1. **Select the Table:** Begin by selecting the range of cells that constitute your table.
2. **Navigate to the 'Table Design' Tab:** Once your table is selected, navigate to the 'Table Design' tab on the Excel ribbon. Here, you'll find a `Table Name` option.
3. **Enter a Name:** Below the `Table Name` option, there is a textbox where you can enter the name for the table. Give the table a name and press "Enter".

1.10.4 Changing Appearance:

The "Table Design" tab offers a range of predefined table styles that you can apply with a single click. These styles include different colors, fonts, and formatting options. If you are not happy with any of the built-in Excel table styles, you can create your own table style from `New Table Style` option underneath the predefined styles.

Experiment with the features to discover the full potential of tables in Excel for your specific needs.

1.11 Formatting Worksheets

Formatting a worksheet in Excel is essential for enhancing readability, organizing data, and making it visually appealing. Whether you're working on a budget, report, or any other document, mastering formatting options can significantly improve the presentation of your data. In this section, we'll explore various formatting tools including font, text alignment, color and shading, borders and lines.

1.11.1 Font

Font formatting allows you to change the appearance of text within cells. Excel offers a variety of font styles, sizes, and effects to customize text according to your preference.

1. Select the range of cells you want to format.
2. Go to the 'Home' tab in the Excel ribbon.
3. In the 'Font' group, you can choose different font styles, sizes, bold, italic, underline, or other effects.

1.11.2 Text Alignment:

Text alignment determines how the content within a cell is positioned. Excel provides options for aligning text horizontally and vertically within cells.

1. Select the cells you want to align.
2. Navigate to the 'Alignment' group under the 'Home' tab.
3. Click on the alignment icons to align text left, center, or right horizontally, and top, middle, or bottom vertically.

1.11.3 Colour and Shading:

Colour and shading help to differentiate data, highlight important information, and improve visual clarity. Excel allows you to change the background colour of cells, apply shading effects, and colorize text.

1. Select the cell or range of cells you want to format.
2. Locate the 'Font' and 'Fill' options in the 'Home' tab.
3. Choose a font color to change the text color.
4. Use the 'Fill Color' tool to change the background color of cells.

1.11.4 Borders and Lines:

Borders and lines enable you to create clear boundaries around cells or ranges, making your data more organized and structured.

1. Select the cells you want to apply borders to.
2. Navigate to the 'Font' group under the 'Home' tab.
3. Click on the 'Borders' dropdown menu to choose from various border styles such as outline borders, inside borders, or diagonal borders.
4. Select the desired border style to apply it to the selected cells.

Experiment with these formatting features to create professional-looking documents tailored to your specific needs.

1.12 Conditional Formatting

Conditional formatting in Excel is a powerful tool that allows you to visually highlight and format cells based on specific conditions. Whether you're a data analyst, program manager, or a student, understanding how to utilize conditional formatting can significantly enhance the readability and analysis of your Excel spreadsheets. In this section, we'll see the various aspects of conditional formatting including graphical conditional formats and creating formula-based rules.

1.12.1 Specifying Conditional Formatting:

To specify conditional formatting in Excel, follow these simple steps:

1. **Select the Range:** Begin by selecting the range of cells you want to apply conditional formatting to.
2. **Access the Conditional Formatting Menu:** Navigate to the "Home" tab on the Excel ribbon, then click on the "Conditional Formatting" dropdown menu located in the "Styles" group.
3. **Choose a Formatting Rule:** From the dropdown menu, select the type of conditional formatting rule you want to apply. Excel offers a variety of predefined rules such as highlighting cells that contain specific text, dates, or values, as well as rules based on comparison operators.
4. **Specify Conditions:** Define the conditions under which the formatting should be applied. This could involve setting threshold values, text criteria, date ranges, or custom formulas.
5. **Set Formatting Options:** After selecting the rule type, configure the formatting options according to your preferences. This may include choosing font color, cell fill color, font style, borders, and more.
6. **Apply and Review:** Once you've set up the formatting rule, click "OK" to apply it to the selected range. Review the changes to ensure they meet your expectations.

1.12.2 Using Graphical Conditional Formats:

Excel provides several graphical conditional formats that offer visually appealing ways to represent your data. Some popular graphical formats include:

- **Data Bars:** Data bars add horizontal bars to cells proportional to their values. This makes it easy to visually compare values across a range of cells.

- **Color Scales:** Color scales apply a gradient of colors to cells based on their values. For example, you can use a green-yellow-red color scale to highlight high, medium, and low values in your dataset.
- **Icon Sets:** Icon sets display icons (such as arrows, shapes, or symbols) in cells based on predefined criteria. For instance, you can use arrow icons to indicate whether values are increasing, decreasing, or staying the same.

1.12.3 Formula-Based Rules:

Formula-based conditional formatting allows you to create custom rules using Excel formulas. This provides flexibility in defining conditions for formatting. Here's an example of how to create a formula-based rule:

Example: Highlighting Cells with Values Greater Than Average

1. **Select the Range:** Choose the range of cells you want to format.
2. **Access Conditional Formatting:** Go to the "Conditional Formatting" menu and select "New Rule."
3. **Choose a Rule Type:** Select "Use a formula to determine which cells to format."
4. **Enter the Formula:** In the formula box, enter a formula that defines the condition. For this example, the formula could be `=A1>AVERAGE(A:A)` assuming the data is in column A.
5. **Set Formatting:** Specify the formatting options you want to apply to cells that meet the condition.
6. **Apply and Review:** Click "OK" to apply the rule and review the changes in your spreadsheet.

1.13 Relative and Absolute Cell Referencing

There are two types of cell references: **relative** and **absolute**. Relative and absolute references behave differently when copied and filled to other cells. Relative references change when a formula is copied to another cell. Absolute references, on the other hand, remain constant no matter where they are copied.

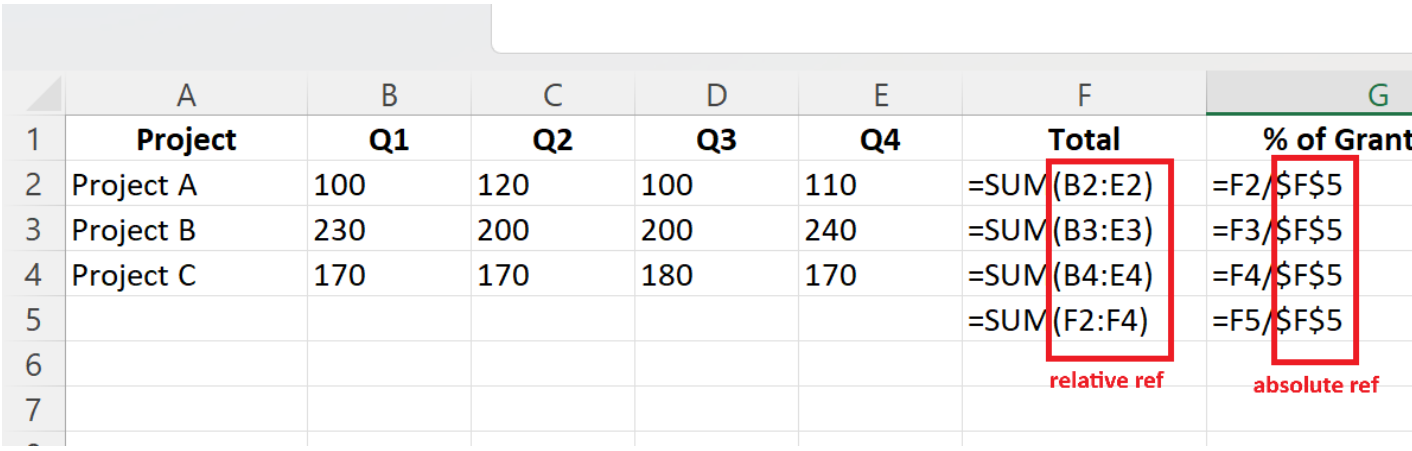
Relative: The row and column references can change when you copy the formula to another cell because the references are actually offsets from the current row and column. By default, Excel creates relative cell references in formulas.

Absolute: The row and column references don't change when you copy the formula because the reference is to an actual cell address. An absolute reference uses two dollar signs in its address: one for the column letter and one for the row number (for example, `$A$5`).

Mixed: Either the row or the column reference is relative, and the other is absolute. Only one of the address parts is absolute (for example, `$A4` or `A$4`).

The type of cell reference is important only if you plan to copy the formula to other cells.

1.13.1 Example



Relative Cell Referencing: In the given example, if we're calculating the Total for each project, the formula for Project A's Total might be `=SUM(B2:E2)`. When this formula is copied to the row below for Project B, Excel automatically adjusts the formula to `=SUM(B3:E3)`. This adjustment is because Excel sees that the formula has been moved down by one row, so it updates the row references accordingly. This behavior is known as relative cell referencing.

Absolute Cell Referencing: Now, let's say we want to calculate the percentage of each project's total compared to the grand total. To do this, we might use a formula like `=F2/$F$5` for Project A's percentage. The `$` signs before the row and column in `$F$5` make the reference absolute. When you copy this formula to other rows, the reference to cell F5 will remain unchanged. This is useful when you want a specific cell reference to stay constant, such as when referencing a grand total or a fixed value.

In summary, relative cell referencing adjusts the reference based on the position of the formula when copied, while absolute cell referencing keeps the reference constant regardless of where the formula is copied.

1.14 Creating, Opening, and Saving Workbooks

In Excel, workbooks are the foundation of organizing and manipulating data. Whether you're starting from scratch, accessing an existing file, or preserving your work, understanding how to create, open, and save workbooks is fundamental.

1.14.1 Creating a New Workbook:

1. **Launch Excel:** Start by opening Microsoft Excel on your computer. You can usually find it in your applications folder or by searching your computer's programs.
2. **Select 'Blank Workbook':** Upon launching Excel, you're typically greeted with a new blank workbook automatically. If not, you can create a new one by clicking on "Blank Workbook" from the available templates or options.
3. **Begin Your Work:** Once the new workbook is created, you're ready to start entering data, formulas, or any other content you need. Each workbook consists of individual worksheets (tabs at the bottom), where you can organize your data efficiently.

> Tip
>
> You can use the keyboard shortcut `Ctrl + N` to create a new workbook

1.14.2 Opening an Existing Workbook:

1. **Navigate to File:** To open an existing workbook, click on the "File" tab located in the top-left corner of the Excel window. This will bring up the File menu.
2. **Select 'Open':** From the File menu, select the "Open" option. Alternatively, you may see a list of recent workbooks that you can directly access without going through the 'Open' dialog.
3. **Locate the Workbook:** A dialog box will appear, allowing you to browse through your computer's files. Navigate to the location where your desired workbook is stored.
4. **Choose the Workbook:** Once you've found the workbook you wish to open, select it by clicking on its file name, then click 'Open.' Excel will load the workbook, displaying its contents in the Excel window.

> Tip
>
> You can use the keyboard shortcut `Ctrl + O` to open an existing workbook

1.14.3 Saving a Workbook:

1. **Click on File:** When you're ready to save your workbook, navigate to the "File" tab again.
2. **Select 'Save':** If you're saving the workbook for the first time, choose 'Save As.' If you've already saved it before and want to update the existing version, choose 'Save.'

3. **Choose Location and Name:** In the Save As dialog box, choose where you want to save the workbook on your computer. Additionally, give your workbook a descriptive name so you can easily identify it later.
4. **Click 'Save':** Once you've selected the location, name, and file format, click 'Save' to save your workbook. If you chose 'Save' earlier, it will overwrite the existing file with the updates you've made.

> Tip
>
> You can use the keyboard shortcut `Ctrl + S` to save a workbook

1.15 Printing Your Excel File

You may sometimes need to print the excel file for sharing, presenting, or archiving your data. Excel has a variety of printing options to ensure your document look professional and meet your specific needs. Here, we touch upon the basic printing options. There are many other options or formatting setups which you can read on your own.

1. **Review Your Spreadsheet**: Before printing, ensure your spreadsheet is formatted correctly. Check for any errors, inconsistencies, or formatting issues that may affect the printout.
2. **Set Print Area (Optional)**: If you only want to print a specific area of your spreadsheet, select the cells you want to print, then navigate to the Page Layout tab on the Excel ribbon. Click on "Print Area" and select "Set Print Area" to define the range.
3. **Adjust Page Layout**: Navigate to the Page Layout tab and click on "Page Setup." Here you can adjust settings such as orientation (portrait or landscape), paper size, margins, and scaling to fit your spreadsheet appropriately on the page.
4. **Preview Print**: Use the Print Preview function (located in the File tab under Print) to review how your spreadsheet will appear on paper. This allows you to make any final adjustments before printing.
5. **Print Your Spreadsheet**: Once satisfied with the preview, select the printer you wish to use and click "Print." Excel will then send your spreadsheet to the printer, producing a hard copy according to your specified settings.

1.16 Converting to PDF Format

Excel provides a straightforward method to convert your spreadsheet into a PDF file. Simply go to the File tab, select "Save As," and choose "PDF" from the list of available file formats. Name your file and select the destination folder, then click "Save" to generate the PDF. Converting to PDF ensures that your spreadsheet maintains its formatting, layout, and data integrity across different devices and platforms. This is particularly useful for sharing documents with others who may not have Excel installed or to preserve the original formatting for archival purposes.

1.17 Useful Keyboard Shortcuts

Shortcut	Description
Ctrl + C	Copy
Ctrl + X	Cut
Ctrl + V	Paste
Ctrl + Z	Undo
Ctrl + Y	Redo
Ctrl + S	Save
Ctrl + N	New workbook
Ctrl + O	Open workbook
Ctrl + P	Print
Ctrl + F	Find
Ctrl + H	Replace
Ctrl + A	Select all
Ctrl + B	Bold
Ctrl + I	Italic
Ctrl + U	Underline
Ctrl + Spacebar	Select entire column
Shift + Spacebar	Select entire row
F7	Spell check
Alt + =	AutoSum
Alt + Enter	Start a new line in the same cell
Alt + F4	Close Excel
Ctrl + Tab	Switch between open workbooks
Ctrl + Page Up	Move to the previous worksheet
Ctrl + Page Down	Move to the next worksheet

2 Formulas and Functions

As said earlier, formulas are what makes spreadsheet a spreadsheet. You use formulas in your Excel worksheets to calculate results from the data stored in the worksheet. When data changes, the formulas calculate updated results with no extra effort on your part. In this book, we will not be explaining each and every formula or function in detail. Instead, we try to equip you with the ability to understand a function and how to apply it by looking at the syntax and utilizing the built-in help feature in Microsoft Excel. There is little use committing every function to memory when assistance is readily accessible alongside each function. Also, Excel continually introduces newer, more efficient functions and understanding fundamental principles of functions is more important.

2.1 Formula Basics

Formulas in Excel are expressions that perform calculations or operations on values in a worksheet. They can range from simple arithmetic calculations to sophisticated logical operations and statistical analyses. Excel formulas start with an equal sign (=) followed by any of these elements:

- Mathematical operators such as + (for addition), * (for multiplication)
- Cell references (including named cells and ranges)
- Values or text
- Worksheet functions (such as `SUM` and `AVERAGE`)

After you enter the formula, excel displays the calculated result in the cell(s).

Few examples of formulas:

Example	Formula
Simple Addition	`=A1 + B1`
Multiplication	`=A1 * B1`
SUM Function	`=SUM(A1:A10)`
AVERAGE Function	`=AVERAGE(A1:A10)`
IF Function	`=IF(A1>10, "Pass", "Fail")`
Conditional Calculation	`=IF(A1>100, A1*10, A1*5)`

2.2 Operators used in formulas

Operators are symbols that indicate what mathematical (or logical) operation you want the formula to perform.

Operator	Name
+	Addition
-	Subtraction
*	Multiplication
/	Division
^	Exponentiation
&	Concatenation
=	Logical comparison (equal to)
>	Logical comparison (greater than)
<	Logical comparison (less than)
>=	Logical comparison (greater than or equal to)
<=	Logical comparison (less than or equal to)
<>	Logical comparison (not equal to)

2.3 Function Basics

A function is a predefined formula that performs calculations or task using specific values in a particular order. These functions are designed to simplify complex calculations and streamline repetitive tasks, saving you time and effort.

2.3.1 Syntax

Every Excel function follows a standardized syntax, consisting of the function name, followed by parentheses enclosing one or more arguments. Arguments are the inputs provided to the function, which determine its behavior and output. Understanding the syntax of a function is crucial for using it correctly. Let's break it down:

```
=FUNCTION_NAME(argument1, argument2, ...)
```

- **Function Name**: This is the name of the function that defines its purpose. Examples include SUM, AVERAGE, IF, VLOOKUP, etc.
- **Arguments**: These are the values or cell references provided to the function to perform the desired calculation. The number and type of arguments vary depending on the function.

2.4 Referencing Cells in Other Worksheets

To use a reference to a cell in another worksheet in the same workbook, use this format:

```
=SheetName!CellAddress
```

In other words, precede the cell address with the worksheet name followed by an exclamation point. Here's an example of a formula that uses a cell on the Sheet2 worksheet:

```
=A1*Sheet2!A1
```

This formula multiplies the value in cell A1 on the current worksheet by the value in cell A1 on Sheet2.

2.5 Referencing Data in a Table

We have discussed about Excel Tables in Section 1.10.

Tables in Excel come with predefined headers for each column. These headers provide an intuitive way to reference data within the table. You can use these headers in formulas and functions to retrieve specific information.

For instance, you have a table named "PatientData" with columns: "Name", "Age", "Sex", "Days of Hospitalization". To find the average days of hospitalization, you can use the formula:

```
=AVERAGE(PatientData[Days of Hospitalization])
```

Here `PatientData[Days of Hospitalization]` refers to "Days of Hospitalization" column in the table "PatientData".

2.6 Common Formula Errors

Sometimes when you enter a formula, Excel displays a value that begins with a hash mark (#). This is a signal that the formula is returning an error value. You have to correct the formula (or correct a cell that the formula references) to get rid of the error display.

Error Value	Explanation
#DIV/0!	The formula is trying to divide by zero. Because Excel applies a value of 0 to empty cells, this error also occurs when the formula attempts to divide by a cell that is blank or has a value of 0.
#NAME?	The formula uses a name that Excel doesn't recognize. This can happen if you delete a name that's used in the formula, if you misspell a name and then hit Enter, or if you have unmatched quotes when using text.
#N/A	The formula is referring (directly or indirectly) to a cell that uses the NA function to signal that data is not available. For instance, the following formula returns an #N/A error if the A1 is empty: =IF(A1=" ", NA(), A1) Some lookup functions (for example, VLOOKUP and MATCH) can also return #N/A when they do not find a match.
#NULL!	The formula uses an intersection of two ranges that don't intersect. (This concept is described later in the chapter.)
#NUM!	A problem with a value exists; for example, you specified a negative number as an argument where a positive number is required.
#REF!	The formula refers to a cell that isn't valid. This can happen if the cell has been deleted from the worksheet.
#VALUE!	The formula includes an argument or operand of the wrong type. (An operand is a value or cell reference that a formula uses to calculate a result.

2.7 Formula for Common Mathematical Operations

Like everyone who has to use Excel, public health students and professionals also will be asked to perform mathematical operations that provide insight into the data. We will see some examples of formulas that will help with this type of analyses.

2.8 Calculating Percentages

Understanding the proportion of individual values in relation to the total is crucial for insightful analysis. Excel simplifies this task with the following formula:

```
= (Value / Total) * 100
```

Here, "Value" represents the specific data point you're interested in, and "Total" denotes the sum of all relevant values. By multiplying the quotient with 100, you obtain the percentage representation.

For example, suppose you have a dataset containing the number of cases of a disease for different states in a country and you want to calculate the percentage of the cases in each state out of the total cases in the country. You can use the formula `=(Cases in State / Total Cases in Country) * 100` to find the percentage for each region.

Dragging this formula across the dataset will swiftly compute the percentage of cases in each state.

Analyzing variations between two data points is essential for detecting trends or anomalies. The formula for calculating percentage variance is:

```
= ((New Value - Old Value) / Old Value) * 100
```

In this formula, "New Value" signifies the updated or later data point, while "Old Value" represents the initial or previous data point. The difference between the new and old values, divided by the old value and multiplied by 100, yields the percentage variance.

Consider a scenario where you have data on the number of cases of a disease for two consecutive years. To calculate the percentage increase in cases from one year to the next, you can use the formula `=((Cases in Year 2 - Cases in Year 1) / Cases in Year 1) * 100`.

Executing this formula provides a clear percentage representation of the case variation, indicating whether cases have increased or decreased and by what percentage.

Exercise file is available here

2.9 Counting Values in a Range

Excel provides several functions to count the values in a range. They allow ypu to quickly tally up occurrences, quantify datasets, and derive valuable insights.

2.9.1 COUNT() Function

The `COUNT()` function is a fundamental tool for counting numeric values within a dataset. It disregards empty cells and cells containing text or errors.

Syntax: `=COUNT(value1, [value2], ...)`

Example:

Consider a dataset containing the number of confirmed COVID-19 cases per day in a specific area:

Day	Confirmed Cases
Day 1	350
Day 2	520
Day 3	
Day 4	430
Day 5	Error
Day 6	650

To count the the total number of days with reported cases:

```
=COUNT(A2:A7)
```

This formula will return `4`, as there are four numeric values in the specified range.

2.9.2 COUNTA() Function

While `COUNT()` tallies only numeric values, `COUNTA()` extends its capability to count all non-blank cells, including text and errors.

Syntax: `=COUNTA(value1, [value2], ...)`

Example:

Using the same dataset, let's count all non-blank cells:

```
=COUNTA(A2:A7)
```

This formula will return `5`, including both numeric values and the error.

2.9.3 COUNTIF() Function

The `COUNTIF()` function is invaluable for counting cells that meet specific criteria within a dataset.

Syntax: `=COUNTIF(range, criteria)`

Example:

Suppose we want to count the number of days confirmed cases exceeding 400:

```
=COUNTIF(A2:A7, ">400")
```

This formula will return 3, indicating there are three days in which number of daily confirmed cases were greater than 400.

2.9.4 COUNTIFS() Function

Similar to `COUNTIF()`, `COUNTIFS()` extends its functionality to count cells based on multiple criteria.

Syntax: `=COUNTIFS(criteria_range1, criteria1, [criteria_range2, criteria2], ...)`

Example:

Consider a scenario where we want to count the days with more than 100 cases in a specific region. We can use COUNTIFS() to achieve this by specifying both the region and the case count threshold.

```
=COUNTIFS(A2:A7, ">400", B2:B7, "Region A")
```

2.9.5 IF() Function

The `IF` function evaluates whether a condition is met or not and returns one value if the condition is true and another value if the condition is false.

The syntax of the `IF` function is:

```
=IF(logical_test, value_if_true, value_if_false)
```

- `logical_test`: This is the condition that you want to evaluate. It can be a logical expression, a cell reference, or another function that returns a logical value (TRUE or FALSE).
- `value_if_true`: The value that is returned if `logical_test` evaluates to TRUE.
- `value_if_false`: The value that is returned if `logical_test` evaluates to FALSE.

Suppose we have a dataset containing cholesterol levels, and we want to categorize individuals as having high cholesterol if their level is greater than 240 mg/dL. We can use the `IF` function to create a column that identifies whether each individual has high cholesterol or not.

Example: Assume cholesterol levels are in column A and start at row 2. In cell B2, enter the formula:

```
=IF(A2 > 240, "High Cholesterol", "Normal Cholesterol")
```

This formula checks if the value in cell A2 (cholesterol level) is greater than 240. If true, it returns "High Cholesterol"; otherwise, it returns "Normal Cholesterol".

2.9.6 IFS() Function

`IFS` function extends the capabilities of the `IF` function by allowing you to evaluate multiple conditions and return a value corresponding to the first TRUE condition.

The syntax of the `IFS` function is:

```
=IFS(condition1, value_if_true1, [condition2, value_if_true2], ...)
```

- `condition1`, `condition2`, etc.: These are the conditions that you want to evaluate.
- `value_if_true1`, `value_if_true2`, etc.: These are the values that are returned if the corresponding condition is true.

Let's extend the previous example to categorize cholesterol levels into different ranges (Normal, Borderline High, High) using the `IFS` function.

Example: Assume cholesterol levels are in column A and start at row 2. In cell C2, enter the formula:

```
=IFS(A2 > 240, "High Cholesterol", A2 > 200, "Borderline High Cholesterol", A2 > 0, "Normal C
```

This formula evaluates the cholesterol level in cell A2 against multiple conditions. It returns "High Cholesterol" if A2 $>$ 240, "Borderline High Cholesterol" if A2 $>$ 200 but not greater than 240, and "Normal Cholesterol" if A2 is less than or equal to 200.

2.10 Commonly Used Excel Functions Cheat Sheet

1. **SUM**: Adds up a range of cells.
   Syntax: `SUM(number1, [number2], ...)`
   Example: `=SUM(A1:A10)`
2. **COUNT**: Counts the number of cells in a range that contain numbers.
   Syntax: `COUNT(value1, [value2], ...)`
   Example: `=COUNT(A1:A10)`

3. **AVERAGE**: Calculates the average of a range of cells.
   Syntax: `AVERAGE(number1, [number2], ...)`
   Example: `=AVERAGE(A1:A10)`

4. **MIN**: Returns the smallest number in a range of cells.
   Syntax: `MIN(number1, [number2], ...)`
   Example: `=MIN(A1:A10)`

5. **MAX**: Returns the largest number in a range of cells.
   Syntax: `MAX(number1, [number2], ...)`
   Example: `=MAX(A1:A10)`

6. **IF**: Checks whether a condition is met, and returns one value if true, and another value if false.
   Syntax: `IF(logical_test, value_if_true, value_if_false)`
   Example: `=IF(A1>10, "Yes", "No")`

7. **TRIM**: Removes all spaces from text except for single spaces between words.
   Syntax: `TRIM(text)`
   Example: `=TRIM(A1)`

8. **LEN**: Returns the number of characters in a text string.
   Syntax: `LEN(text)`
   Example: `=LEN(A1)`

9. **VLOOKUP**: Looks for a value in the first column of a table and returns a value in the same row from a specified column.
   Syntax: `VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])`
   Example: `=VLOOKUP(A1, B1:C10, 2, FALSE)`

10. **XLOOKUP**: Searches a range or an array, and returns an item corresponding to the first match it finds. If a match doesn't exist, then `XLOOKUP` can return the closest (approximate) match.
    Syntax: `XLOOKUP(lookup_value, lookup_array, return_array, [if_not_found], [match_mode], [search_mode])`
    Example: `=XLOOKUP(A1, B1:B10, C1:C10, "Not Found")`

11. **INDEX**: Returns the value of a cell in a specified row and column.
    Syntax: `INDEX(array, row_num, [column_num])`
    Example: `=INDEX(A1:C10, 2, 3)`

12. **MATCH**: Searches for a specified value in a range of cells, and returns the relative position of that item.
    Syntax: `MATCH(lookup_value, lookup_array, [match_type])`
    Example: `=MATCH(A1, B1:B10, 0)`

13. **CONCATENATE**: Joins several text strings into one text string.
    Syntax: `CONCATENATE(text1, [text2], ...)`
    Example: `=CONCATENATE(A1, " ", B1)`

14. **LEFT**: Returns the first character or characters in a text string, based on the number of characters specified.
    Syntax: `LEFT(text, [num_chars])`
    Example: `=LEFT(A1, 3)`

15. **RIGHT**: Returns the last character or characters in a text string, based on the number of characters specified.
    Syntax: `RIGHT(text, [num_chars])`
    Example: `=RIGHT(A1, 3)`

16. **MID**: Returns a specific number of characters from a text string, starting at the position you specify.
    Syntax: `MID(text, start_num, num_chars)`
    Example: `=MID(A1, 2, 3)`

17. **UPPER**: Converts text to uppercase.
    Syntax: `UPPER(text)`
    Example: `=UPPER(A1)`

18. **LOWER**: Converts text to lowercase.
    Syntax: `LOWER(text)`
    Example: `=LOWER(A1)`

19. **PROPER**: Capitalizes the first letter of each word in a text string.
    Syntax: `PROPER(text)`
    Example: `=PROPER(A1)`

20. **NOW**: Returns the current date and time.
    Syntax: `NOW()`
    Example: `=NOW()`

21. **TODAY**: Returns the current date.
    Syntax: `TODAY()`
    Example: `=TODAY()`

22. **DATE**: Returns the number that represents the date in Excel date-time code.
    Syntax: `DATE(year, month, day)`
    Example: `=DATE(2021, 12, 25)`

23. **TIME**: Returns the number that represents the time in Excel date-time code.
    Syntax: `TIME(hour, minute, second)`
    Example: `=TIME(12, 0, 0)`

24. **DAYS**: Returns the number of days between two dates.
    Syntax: `DAYS(end_date, start_date)`
    Example: `=DAYS(B1, A1)`

25. **YEAR**: Returns the year of a date.
    Syntax: `YEAR(serial_number)`
    Example: `=YEAR(A1)`

26. **MONTH**: Returns the month of a date.
    Syntax: `MONTH(serial_number)`
    Example: `=MONTH(A1)`

27. **DAY**: Returns the day of a date.
    Syntax: `DAY(serial_number)`
    Example: `=DAY(A1)`

28. **HOUR**: Returns the hour from a time value.
    Syntax: `HOUR(serial_number)`
    Example: `=HOUR(A1)`

29. **MINUTE**: Returns the minute from a time value.
    Syntax: `MINUTE(serial_number)`
    Example: `=MINUTE(A1)`

30. **SECOND**: Returns the second from a time value.
    Syntax: `SECOND(serial_number)`
    Example: `=SECOND(A1)`

31. **ROUND**: Rounds a number to a specified number of digits.
    Syntax: `ROUND(number, num_digits)`
    Example: `=ROUND(A1, 2)`

32. **ROUNDUP**: Rounds a number up, away from zero.
    Syntax: `ROUNDUP(number, num_digits)`
    Example: `=ROUNDUP(A1, 2)`

33. **ROUNDDOWN**: Rounds a number down, toward zero.
    Syntax: `ROUNDDOWN(number, num_digits)`
    Example: `=ROUNDDOWN(A1, 2)`

34. **CEILING**: Rounds a number up to the nearest multiple of significance.
    Syntax: `CEILING(number, significance)`
    Example: `=CEILING(A1, 1)`

35. **FLOOR**: Rounds a number down to the nearest multiple of significance.
    Syntax: `FLOOR(number, significance)`
    Example: `=FLOOR(A1, 1)`

36. **ABS**: Returns the absolute value of a number.
    Syntax: `ABS(number)`
    Example: `=ABS(A1)`

37. **POWER**: Returns the result of a number raised to a power.
    Syntax: `POWER(number, power)`
    Example: `=POWER(A1, 2)`

38. **SQRT**: Returns the square root of a number.
    Syntax: `SQRT(number)`
    Example: `=SQRT(A1)`

39. **MOD**: Returns the remainder from division.
    Syntax: `MOD(number, divisor)`
    Example: `=MOD(A1, B1)`

40. **SUMIF**: Adds the cells specified by a given condition or criteria.
    Syntax: `SUMIF(range, criteria, [sum_range])`
    Example: `=SUMIF(A1:A10, ">10")`

41. **COUNTIF**: Counts the number of cells that meet a criterion.
    Syntax: `COUNTIF(range, criteria)`
    Example: `=COUNTIF(A1:A10, ">10")`

42. **AVERAGEIF**: Calculates the average of cells that meet a criterion.
    Syntax: `AVERAGEIF(range, criteria, [average_range])`
    Example: `=AVERAGEIF(A1:A10, ">10")`

43. **AND**: Returns TRUE if all of its arguments are TRUE.
    Syntax: `AND(logical1, [logical2], ...)`
    Example: `=AND(A1>10, B1<5)`

44. **OR**: Returns TRUE if any argument is TRUE.
    Syntax: `OR(logical1, [logical2], ...)`
    Example: `=OR(A1>10, B1<5)`

45. **NOT**: Reverses the logic of its argument.
    Syntax: `NOT(logical)`
    Example: `=NOT(A1>10)`

46. **ISNUMBER**: Returns TRUE if the value is a number.
    Syntax: `ISNUMBER(value)`
    Example: `=ISNUMBER(A1)`

47. **ISTEXT**: Returns TRUE if the value is text.
    Syntax: `ISTEXT(value)`
    Example: `=ISTEXT(A1)`

48. **ISBLANK**: Returns TRUE if the cell is empty.
    Syntax: `ISBLANK(value)`
    Example: `=ISBLANK(A1)`

3 Creating Visuals

Data visualization is a graphical representation of data. It can transform your raw data into meaningful visual stories and identify trends faster. It also helps in communicating complex information effectively.

This chapter provides a walk through of Excel's fundamental visualization capabilities. Rather than going deeply into in depth details, we aim to make you confident in navigating menu options independently. We assume that you know what you want, and just require guidance on execution. For a comprehensive understanding of visualization principles, we strongly suggest referring a good general visualization book. Our recommendation is The Truthful Art by Albert Cairo. It doesn't teach the mechanics of creating visualizations, but instead focuses on what you need to think about in order to create effective graphics.

3.1 Creating Charts

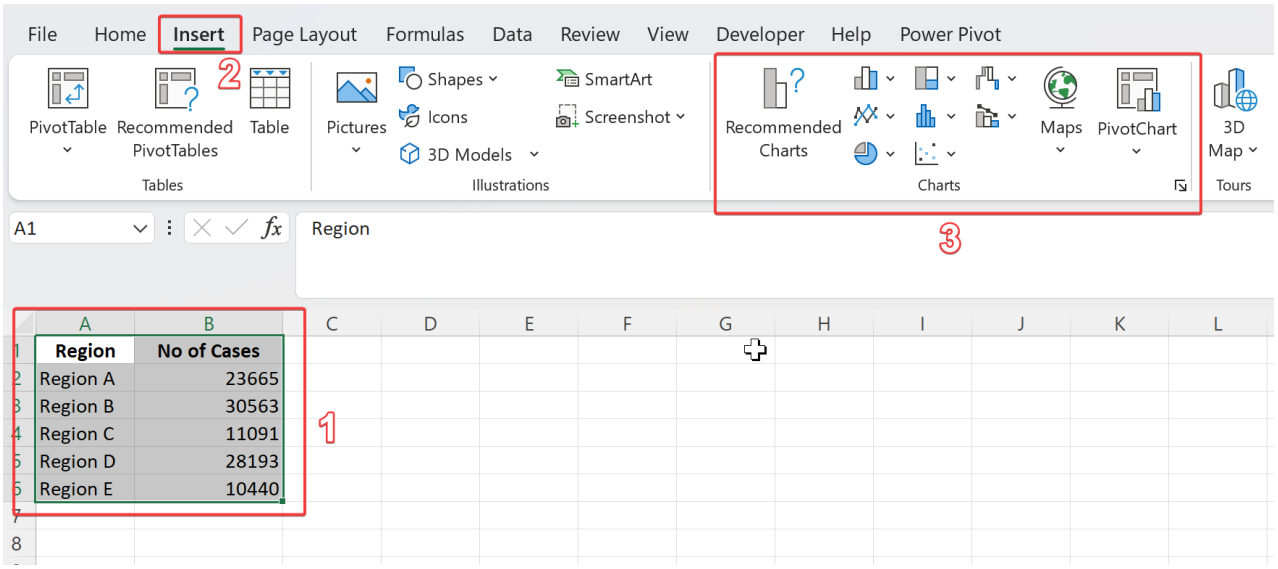
In this section, we'll walk through the steps to create charts in Excel, allowing you to convey your message with precision and impact.

1. **Prepare Your Data**

    - Ensure your data is organized in a table format with clear labels for rows and columns.
    - Select the data range you want to include in your chart, including headers if applicable.

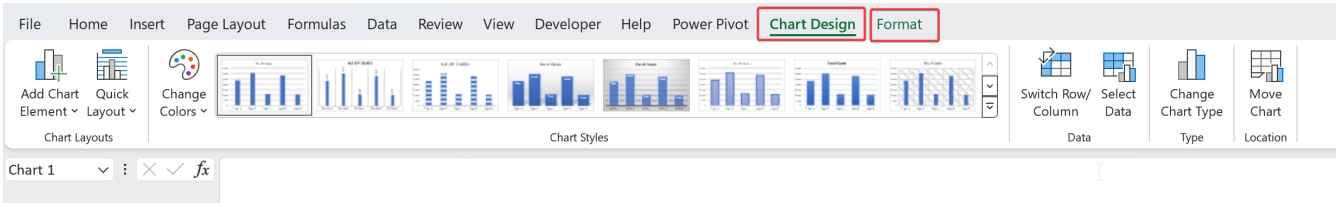
2. **Insert a Chart**

    - With your data selected, navigate to the "Insert" tab in the Excel ribbon.
    - Click on the desired chart type from the "Charts" group. Common options include Column, Line, Pie, Bar, and Scatter.
    - Selecting the desired chart type will automatically create a basic chart using your selected data.



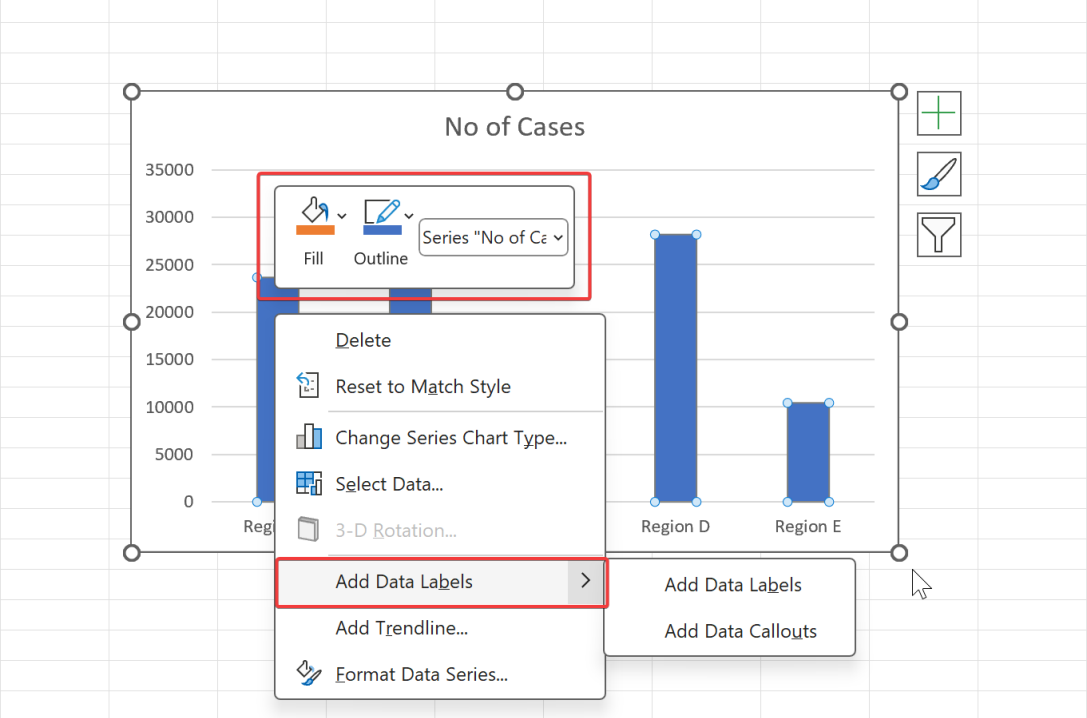
3. **Customize Your Chart**

- Once your chart is created, you can customize it to better suit your needs.
- Click on the chart to select it, and then navigate to the "Chart Design" and "Chart Format" tabs in the ribbon.
- From here, you can change chart elements such as titles, axes labels, legends, colors, and styles.



4. **Add Data Labels and Annotations (Optional)**

- To make your chart more informative, you can add data labels, annotations, or other elements.
- Right-click on the chart element you want to modify and select "Add Data Labels" or "Add Chart Element" to insert annotations, such as trendlines or error bars.



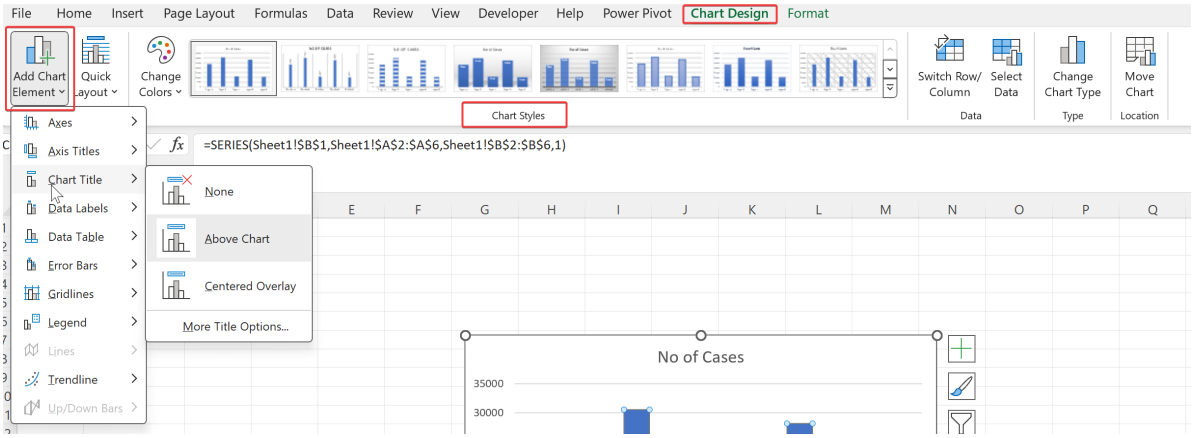
5. **Share Your Chart**

    - To share your chart with others, you can either share the entire workbook or copy and paste the chart into other documents or presentations.

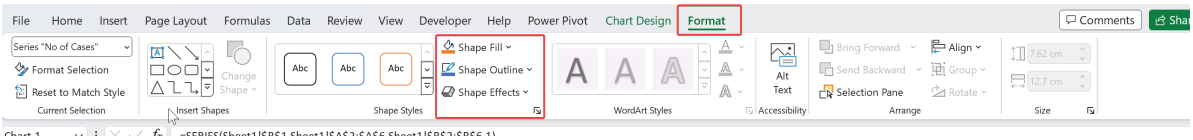
3.2 Customizing Charts

You can customize the chart created from "Chart Design" and "Format" tabs in the ribbon.

1. **Chart Elements:** Excel provides an extensive range of chart elements to enhance clarity and aesthetics. These elements include titles, axis labels, data labels, gridlines, legends, and trendlines. Each element can be customized to suit your presentation needs.
2. **Chart Styles:** Excel offers pre-designed chart styles, enabling you to quickly apply professional-looking formatting to your charts. These styles have different color schemes, line weights, and shading effects, allowing you to match your chart's appearance with your overall document or presentation theme.



3. **Formatting:** Excel provides different formatting options to customize the appearance of your charts and tailor them to your specific requirements. You can adjust colors, fonts, borders, and other visual attributes to create visually appealing and professional-looking charts.

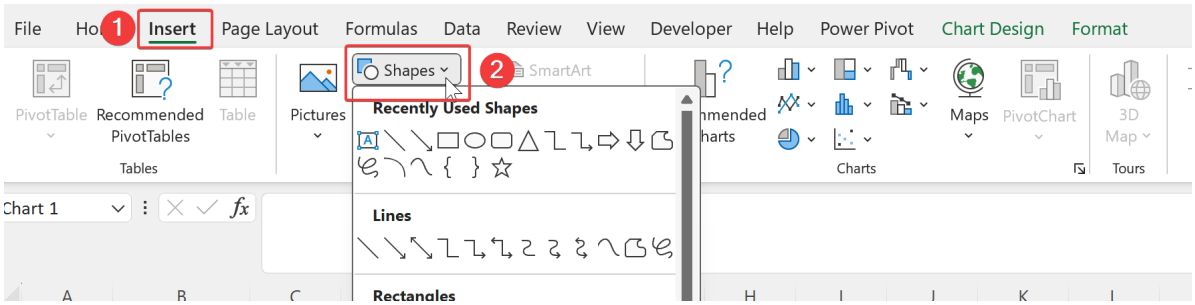


3.3 Using Shapes

Excel provides variety of customizable graphic images known as shapes. You can use them to display text or add some visual appearance to a worksheet.

3.3.1 Inserting Shapes

1. To insert a shape, navigate to the "Insert" tab on the Excel ribbon. Click on the "Shapes" button located in the "Illustrations" group. A dropdown menu will appear, offering a variety of shape categories to choose from, including basic shapes, arrows, flowchart elements, and more.

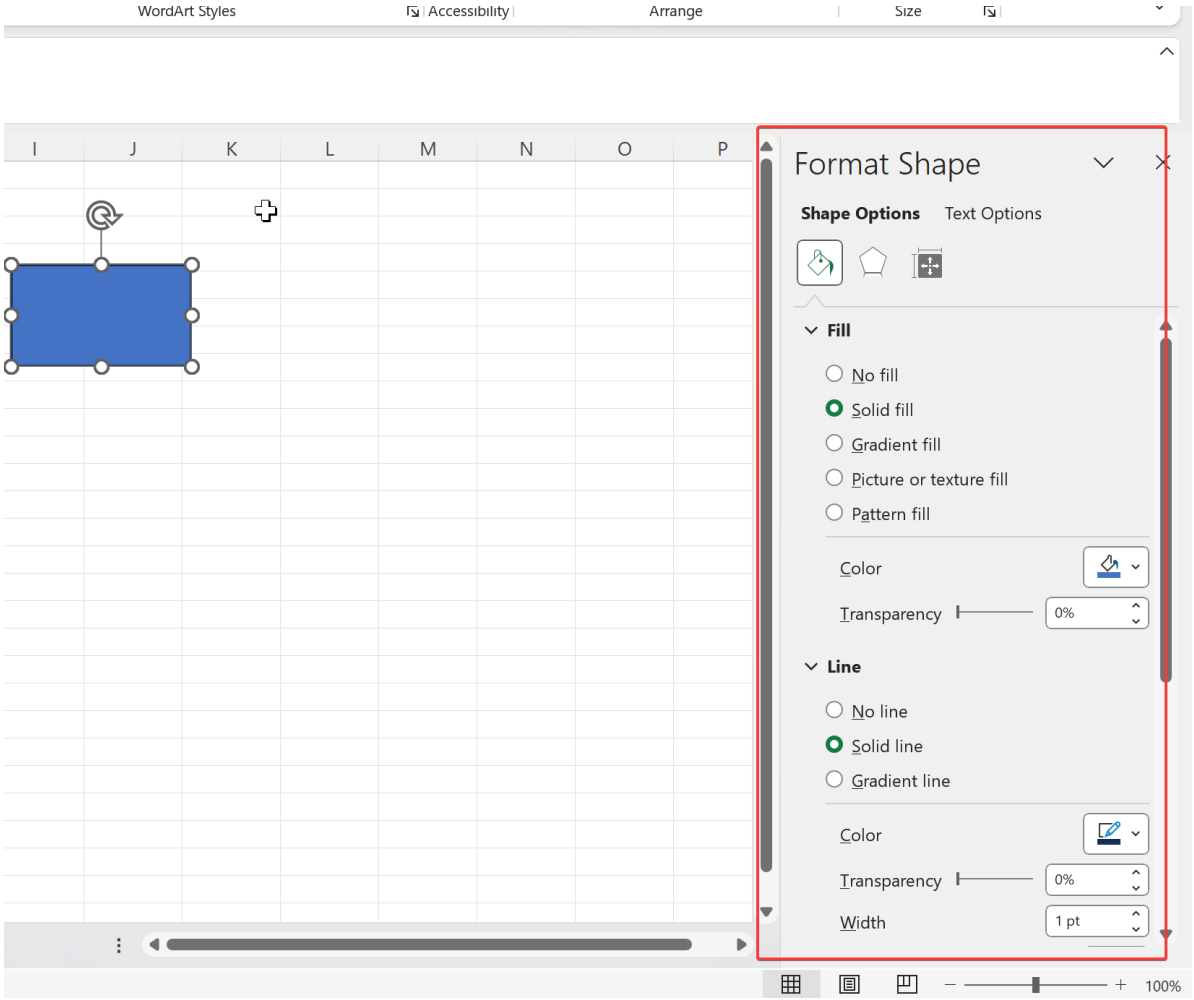


2. Once you've chosen a category, simply click on the desired shape to select it. Your cursor will now transform into a crosshair, allowing you to draw the shape directly onto your worksheet. Click and drag to define the size and proportions of the shape.

3. Excel also allows for the insertion of custom shapes. To do this, select "Edit Shape" from the dropdown menu after clicking on a shape. Here, you can modify the shape's vertices and curves to create a custom design tailored to your needs.

3.3.2 Formatting Shapes

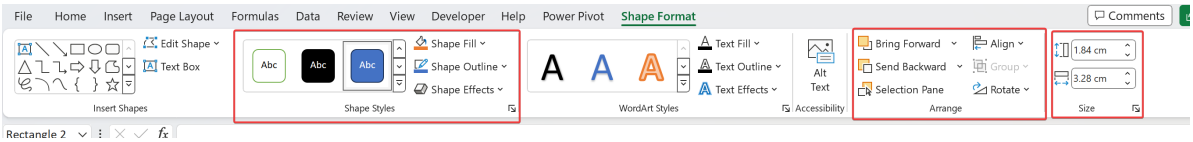
1. **Fill and Outline:** After inserting a shape, you can customize its appearance using the formatting options available in the "Format Shape" pane. To access this pane, right-click on the shape and select "Format Shape" from the context menu. Here, you can adjust the fill color, outline color, and effects such as shadows and reflections.



2. **Shape Styles:** Excel offers a range of predefined shape styles that can be applied with a single click. These styles include variations in fill color, outline weight, and effects, allowing you to quickly achieve a polished look for your shapes.

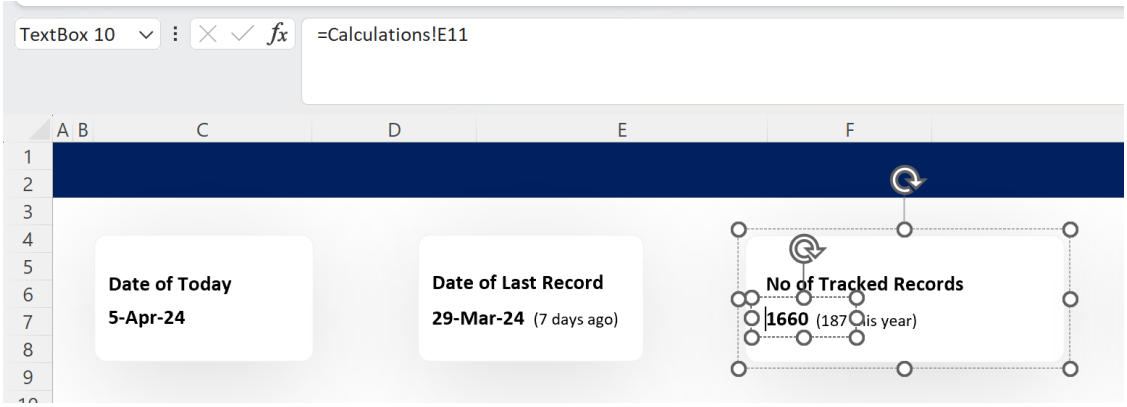
3. **Size and Position:** Precise control over a shape's size and position is essential for achieving a professional layout. Use the sizing handles to adjust the dimensions of the shape, or enter specific values in the "Size & Properties" pane. To align multiple shapes, utilize the alignment tools available in the "Format" tab.

4. **Grouping and Layering:** When working with multiple shapes, grouping them together can simplify manipulation and formatting. Select the shapes you wish to group, right-click, and choose "Group" from the context menu. Additionally, you can change the stacking order of shapes using the "Bring Forward" and "Send Backward" options.



3.3.3 Creating Dynamic Labels

Dynamic labels are the labels that changes with the changes in the underlying data. The below figure illustrates this. The selected text box shape is linked to cell E11 in Calculations worksheet. As the value of this cell changes, the text box displays updated value. To add dynamic labels, select the text box or shape and enter the cell reference in formula bar.



4 Data Cleaning and Analysis

Data analysis is the process of inspecting, cleaning, transforming, and modeling data to discover useful information, inform conclusions, and support decision-making. Excel is also a tool for data analysis that enables you to manipulate, analyze, and visualize data quickly and easily. Excel can do most of the data analysis jobs you encounter in your day to day activities. For advanced analysis other software can be used.

4.1 Cleaning Data

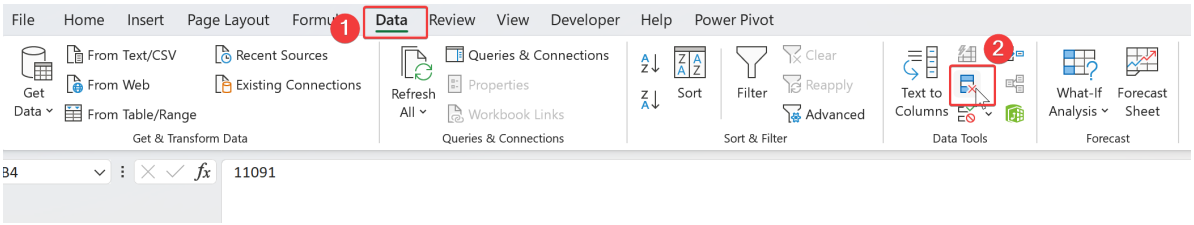
Data cleaning involves fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset. When combining various data sources, there are numerous chances for data duplication or misidentification. Inaccurate data can lead to unreliable results and outcomes. Since data cleaning processes differ from one dataset to another, there isn't a singular approach to outlining precise steps. Nonetheless, it's crucial to establish a template for your data cleaning process to ensure consistent and correct analysis everytime. In this section, we will see some data cleaning techniques which you must be familiar with.

> **!** Important
>
> Before cleaning your dataset, consider making a copy of your original dataset. This way, you can always revert to the original data if needed.

4.1.1 Remove Duplicate Rows

Duplicate rows can clutter your data and make analysis challenging. Excel provides powerful tools to identify and remove these duplicate entries efficiently.

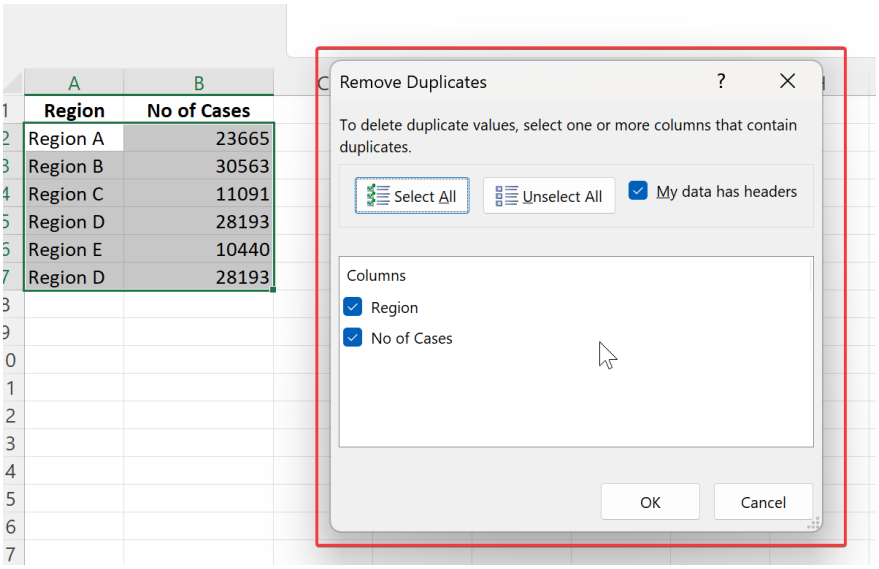
1. **Select Your Data**: Begin by selecting the range of cells or the entire table from which you want to remove duplicates.
2. **Navigate to the Data Tab**: Click on the "Data" tab in the Excel ribbon at the top of the screen.
3. **Click on Remove Duplicates**: In the Data Tools group, locate and click on "Remove Duplicates."



4. **Choose Columns**: A dialog box will appear listing all the columns in your selected range. You can choose the columns where you want Excel to check for duplicates. By default, all columns are selected.

5. **Determine Criteria**: Decide whether you want Excel to consider all columns collectively or only specific columns when identifying duplicates. Check or uncheck the boxes accordingly. For example, if your data have some unique identifier (like ID Card No) then you can select that column only.

6. **Remove Duplicates**: After setting your criteria, click "OK." Excel will process the data and remove any duplicate rows based on your specified criteria.



7. **Review Results**: Excel will display a message indicating how many duplicate values were found and removed. Click "OK" to close the message box.

> **Note**
>
> To visually identify duplicates before removing them, apply conditional formatting to highlight duplicate values within your dataset.

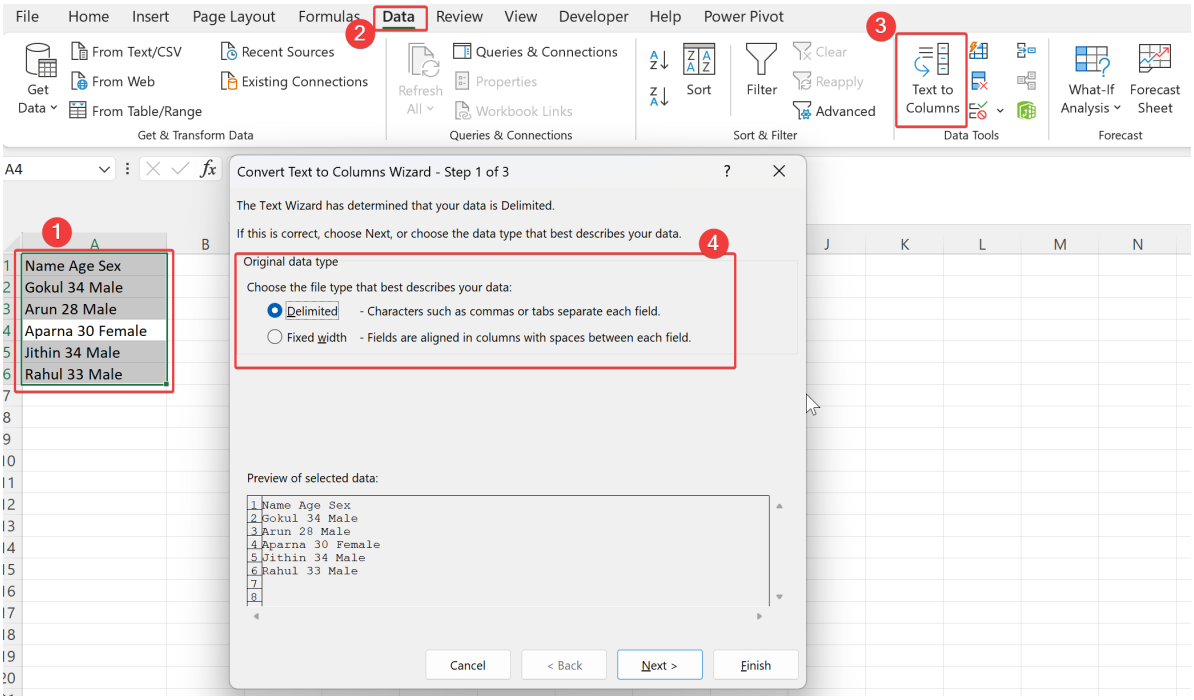
4.1.2 Splitting Text

One of the powerful features of Excel is its ability to manipulate and organize data efficiently. The "Split to Columns" function is useful when you need to separate text within cells into multiple columns based on a delimiter or a specific character. This feature can save you time and effort, especially when dealing with large datasets.

1. **Select Your Data Range:**
   First, make sure the column that contains the data to be split up has enough empty columns to the right to accommodate the extracted data. Select the cells that contain the text you want to split into columns. Ensure that each cell contains the text you wish to separate.

2. **Navigate to the Data Tab:**
   Go to the Excel ribbon and click on the "Data" tab.

3. **Choose "Text to Columns":**
   Within the "Data Tools" group, locate and click on "Text to Columns." This action will launch the "Convert Text to Columns Wizard."



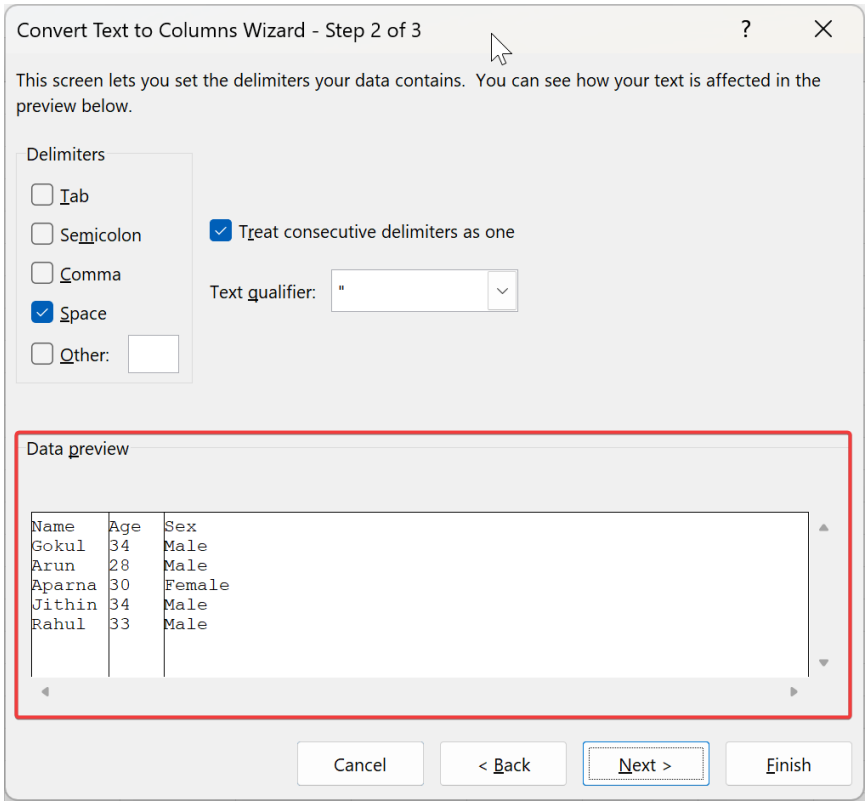
4. **Select the Delimiter:**
   In the first step of the wizard, you'll be prompted to choose how you want to split your data. Excel gives you the option to split by delimiter (such as commas, spaces,

tabs, or custom characters used to separates text items) or by fixed width. Choose the appropriate option based on your data format and click "Next."

5. **Specify Delimiter Settings (if applicable):**
   If you selected to split by delimiter, specify which delimiter Excel should use for splitting your text. You can choose from predefined delimiters like Tab, Semicolon, Comma, Space, or specify a custom delimiter. Preview the results in the Data preview window to ensure the split is as expected.

6. **Configure Column Data Format (if needed):**
   In the next step, you can format each column of data. Choose the data format (General, Text, Date, etc.) for each column or keep it as is. Click "Finish" once you are satisfied with your selections.

7. **Review Your Split Data:**
   Excel will automatically split the selected text into separate columns based on your specifications. Review the resulting columns to ensure the split was successful and meets your requirements.



> **Note**
>
> Always use the Data preview window to see how your text will be split before finalizing the operation.

4.1.3 Removing Extra Space

Sometimes, the data contains extra spaces. It is very difficult to spot a space character at the end of a text string. It can cause problems when you are doing analysis. For example, 'Male' is not same as 'Male'. When you do analysis, there will be two variable values 'Male' and 'Male', but basically both are the same.

To solve this, use the TRIM() function, which removes leading and trailing spaces while condensing multiple interior spaces into a single space.

```
=TRIM(text)
```

`text`: This is the text string from which you want to remove extra spaces.

4.1.4 Changing the case of Text

Mostly, you need to make text in columns consistent in terms of case. Unlike Microsoft Word, Excel doesn't provide a direct way to do it. But we can change the case using functions. The relevant functions are as follows:

1. The `UPPER()` function converts all letters in a text string to uppercase.

```
=UPPER(text)
```

2. The `LOWER()` function converts all letters in a text string to lowercase.

```
=LOWER(text)
```

3. The `PROPER()` function capitalizes the first letter of each word in a text string and converts all other letters to lowercase.

```
=PROPER(text)
```

4.1.5 Converting Values

The `CONVERT()` function allows you to convert a measurement from one unit to another within the same measurement category.

The `CONVERT()` function in Excel follows a specific syntax:

```
=CONVERT(number, from_unit, to_unit)
```

- **number**: This is the value you want to convert.
- **from_unit**: The current unit of the value you want to convert.
- **to_unit**: The unit to which you want to convert the value.

Imagine you have a dataset containing temperature recorded in Fahrenheit (F) in cell A1, and you need to convert these weights to Celsius (C) for analysis or reporting purposes. Let's use the `CONVERT()` function to do this conversion.

```
=CONVERT(A1,"F","C")
```

In this example:

- A1: contain the temperature reading in Fahrenheit
- "F": represent the temperature in Fahrenheit
- "C": indicates the desired unit for conversion(Celsius)

4.1.6 Classifying Values

Often, you may have values that need to be classified into a group. In Excel, you can classify values into groups using the `IFS()` function along with logical tests. The `IFS()` function allows you for multiple conditions to be evaluated, and based on whether the condition is true or false, return different values. This can be used effectively to create categories or groups for your data.

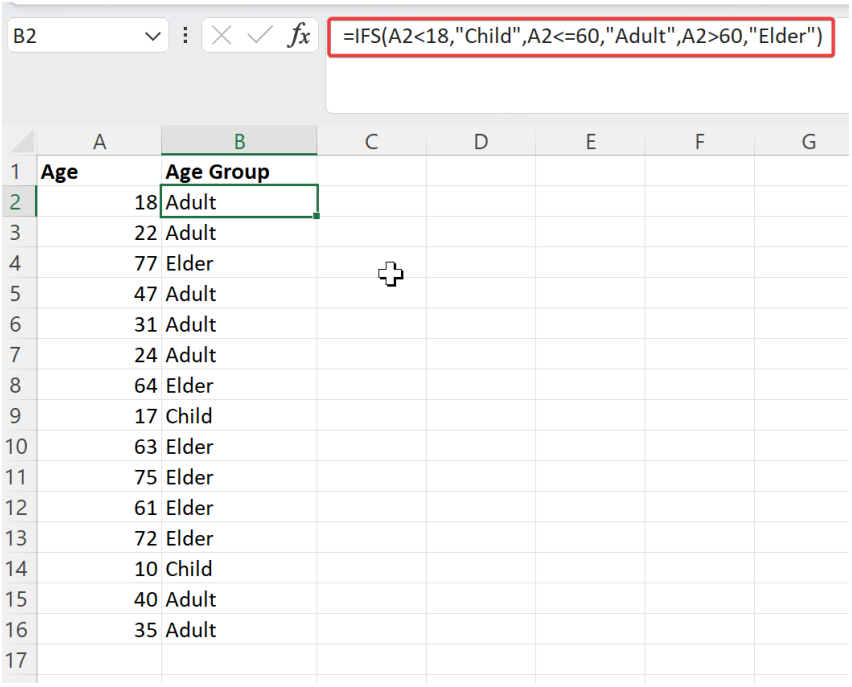
Suppose you have data on the ages of participants of a survey in column A, and you want to classify these values into three groups:

- Child: Values less than 18
- Adult: Values between 18 and 60 (inclusive)
- Elder: Values greater than 60

In an adjacent column (e.g., column B), enter the following formula in the first cell (e.g., B2) to classify the value in A2:

```
=IFS(A2<18,"Child", A2<=60,"Adult", A2>60,"Elder")
```

You can extend the formula down the column by dragging the cell handle. This will automatically apply the formula to all the cells in the column.



	A	B	C	D	E	F	G
1	**Age**	**Age Group**					
2	18	Adult					
3	22	Adult					
4	77	Elder					
5	47	Adult					
6	31	Adult					
7	24	Adult					
8	64	Elder					
9	17	Child					
10	63	Elder					
11	75	Elder					
12	61	Elder					
13	72	Elder					
14	10	Child					
15	40	Adult					
16	35	Adult					
17							

4.1.7 Joining Columns

Combining text from different columns into a single cell can be efficiently achieved using the `CONCAT()` or `TEXTJOIN()` functions.

Using CONCAT() Function

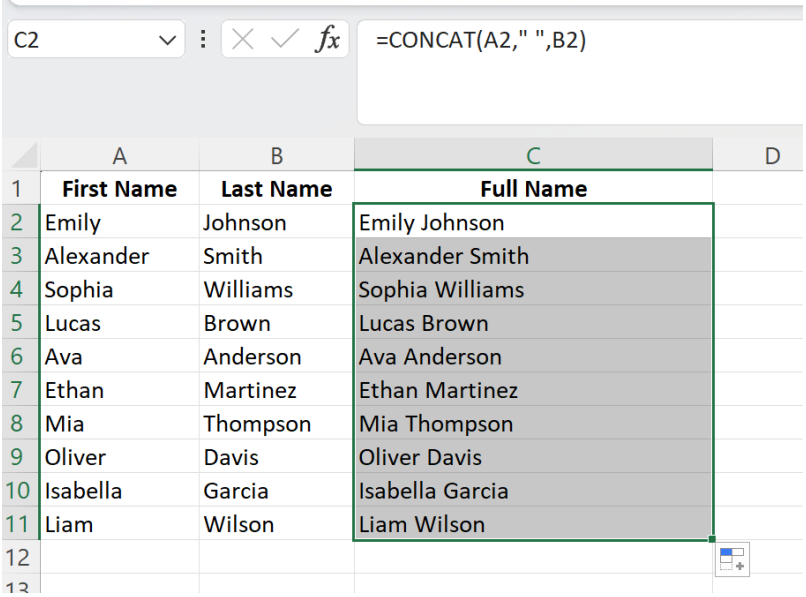
The `CONCAT()` function in Excel is used to concatenate, or join, text from different cells or ranges into one cell. It is particularly useful when you want to combine text strings without using any delimiter (a character that separates text items). You can include additional separators or formatting within the CONCAT() function to control how the text is combined (e.g., adding spaces, commas, etc.).

```
=CONCAT(text1, [text2], ...)
```

- **text1**: This is the first text or cell reference that you want to concatenate.

- [**text2**]: Additional text or cell references that you want to join.

See the example in figure below:



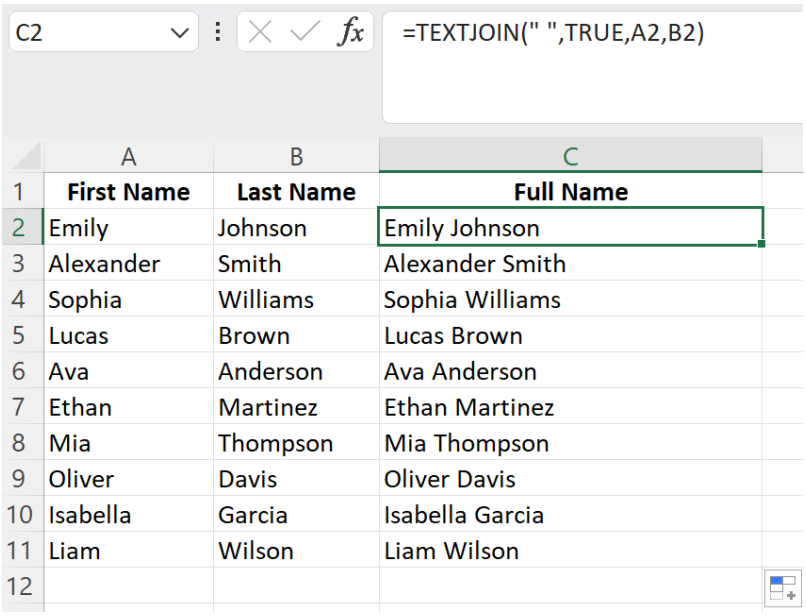
Using TEXTJOIN() Function

The TEXTJOIN() function allows you to concatenate text from a range of cells with specified delimiters (such as commas, spaces, or custom characters) between each item. This function is highly flexible and can handle situations where you need to skip empty cells or include only non-empty values.

```
=TEXTJOIN(delimiter, ignore_empty, text1, [text2], ...)
```

- **delimiter**: The delimiter to be used between each text item.
- **ignore_empty**: A logical value (TRUE or FALSE) that specifies whether to ignore empty cells.
- **text1, text2, ...**: The text values or cell ranges that you want to concatenate.

See the example in figure below:



	A	B	C
1	**First Name**	**Last Name**	**Full Name**
2	Emily	Johnson	Emily Johnson
3	Alexander	Smith	Alexander Smith
4	Sophia	Williams	Sophia Williams
5	Lucas	Brown	Lucas Brown
6	Ava	Anderson	Ava Anderson
7	Ethan	Martinez	Ethan Martinez
8	Mia	Thompson	Mia Thompson
9	Oliver	Davis	Oliver Davis
10	Isabella	Garcia	Isabella Garcia
11	Liam	Wilson	Liam Wilson
12			

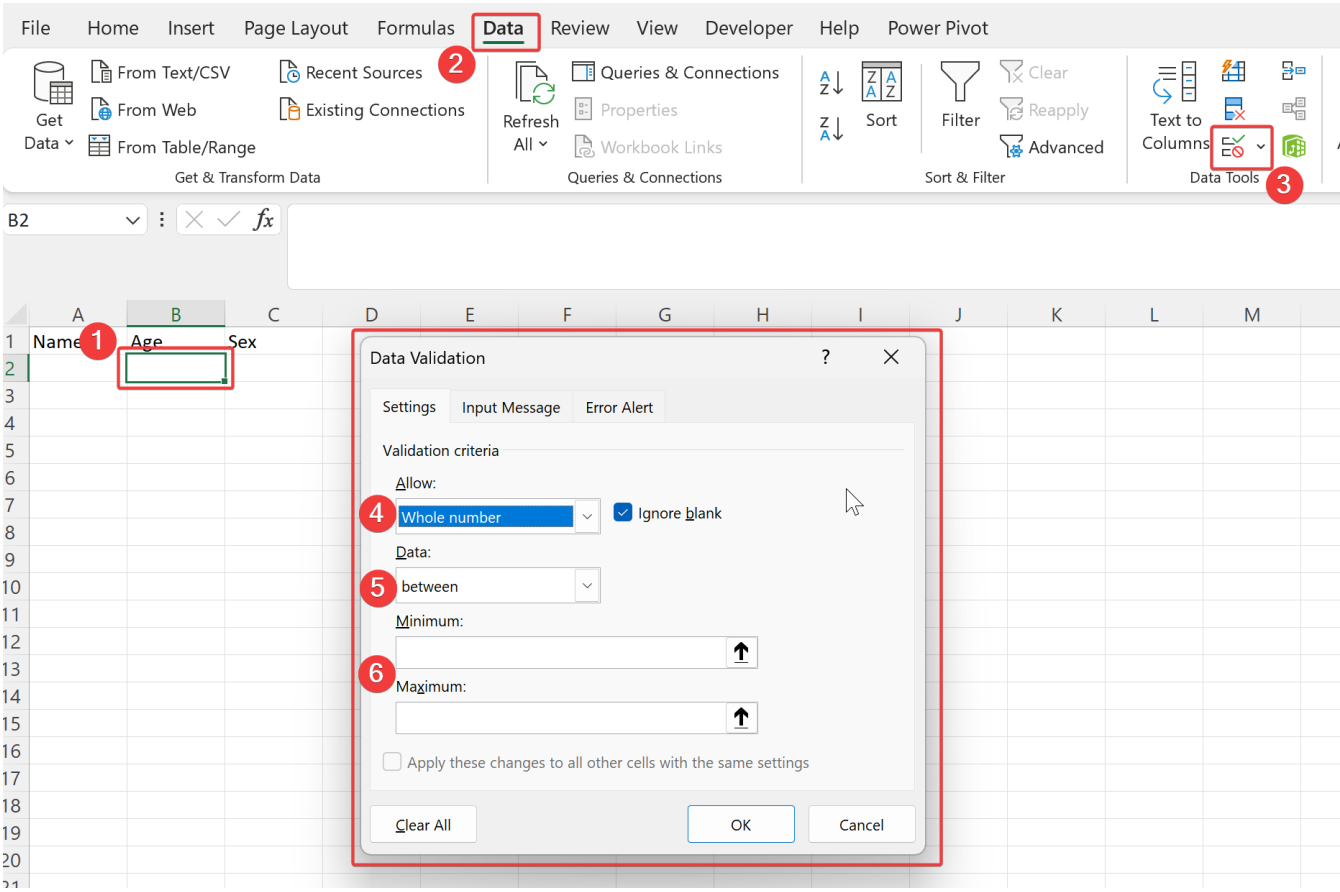
4.2 Data Validation

Data validation is a powerful feature that allows you to control what type of data can be entered into a cell. For example, you may want to limt the data entry in a particular range to whole numbers between 1 and 100. If you make an invalid entry, you can display a custom message. By setting validation criteria, you can ensure data integrity, accuracy, and consistency throughout your spreadsheets. Let's see how to use data validation effectively.

4.2.1 Specifying Validation Criteria

1. **Selecting Cells for Validation:**
    - Begin by selecting the cells where you want to apply data validation. You can select a single cell, a range of cells, or even an entire column.
2. **Accessing Data Validation:**
    - Navigate to the **Data** tab on the ribbon.
    - Click on **Data Validation** in the **Data Tools** group.
3. **Setting Validation Rules:**
    - In the Data Validation dialog box, choose the type of validation you want under the **Settings** tab.

- For example, to allow only whole numbers between 1 and 100, select **Whole number** and specify the minimum and maximum values.



4.2.2 Types of Validation Criteria

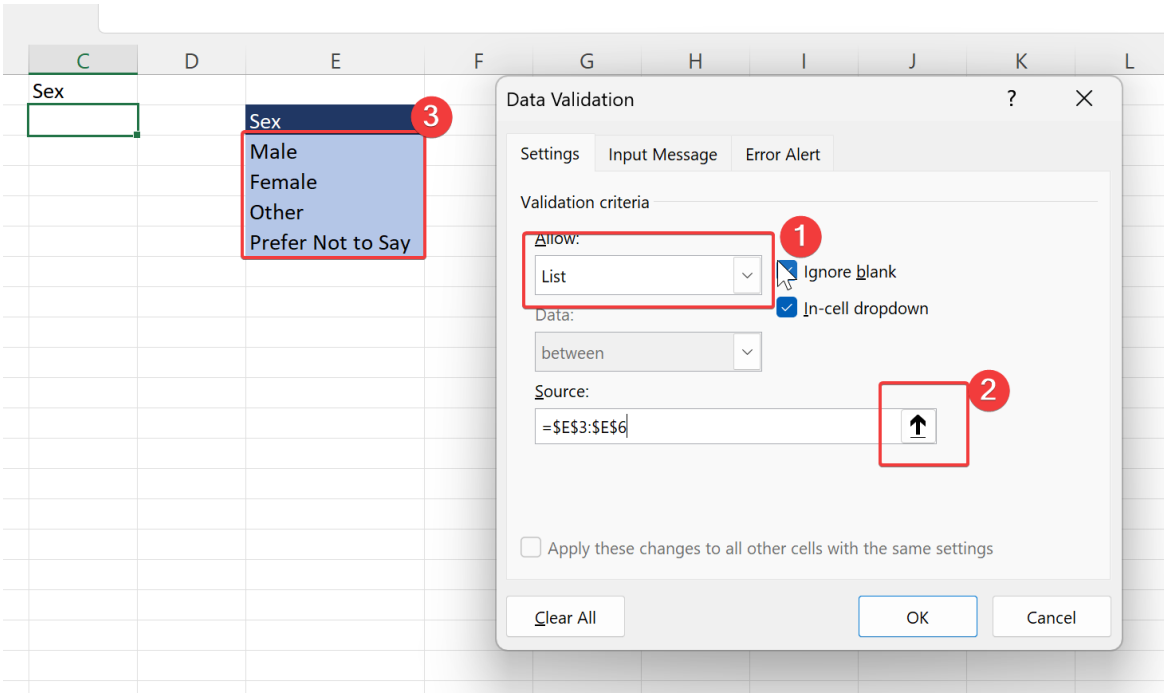
Excel offers several validation criteria options, including:

- **Whole Number:** Restricts input to whole numbers within a specified range.
- **Decimal:** Allows decimal numbers within a specified range.
- **List:** Limits input to items in a predefined list.
- **Date:** Validates input as dates within a specified period.
- **Time:** Validates input as times.
- **Text Length:** Controls the length of text allowed in a cell.
- **Custom:** Uses a custom formula to determine if input is valid.

4.2.3 Creating a Dropdown List

Dropdown lists are particularly useful for ensuring data consistency and ease of use. Here's how to set up a dropdown list using data validation:

1. **Specify List Items:**
    - Enter the items you want in the dropdown list in a separate column or range.
2. **Select Cells for Dropdown:**
    - Choose the cells where you want the dropdown list to appear.
3. **Set Data Validation:**
    - Access the Data Validation dialog box.
    - Choose **List** under **Allow**.
    - Specify the source of the list using the range selector or by typing the range directly (e.g., `=$E$3:$E$6`).
    - You can also write the list items in **Source** box, each items seperated by comma.



4. **Enable the Dropdown:**
    - Now, the selected cells will display a dropdown arrow, allowing users to choose from the predefined list.

4.3 Creating a Dependent List

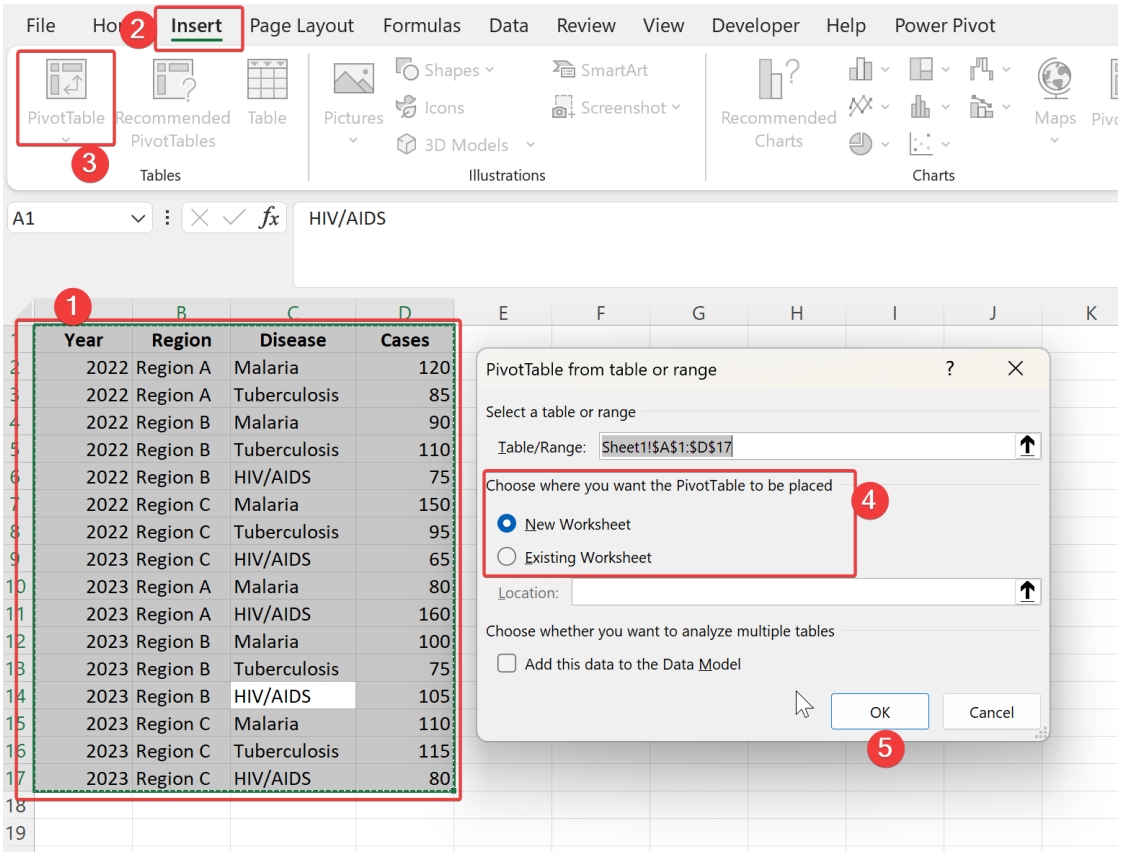
this part is yet to be written

4.4 Pivot Tables

The PivotTable feature is perhaps the most technologically sophisticated component in Excel. A pivot table can provide quick answers to questions about your table that can otherwise only be answered by complicated formulas.

4.4.1 Creating Pivot Table

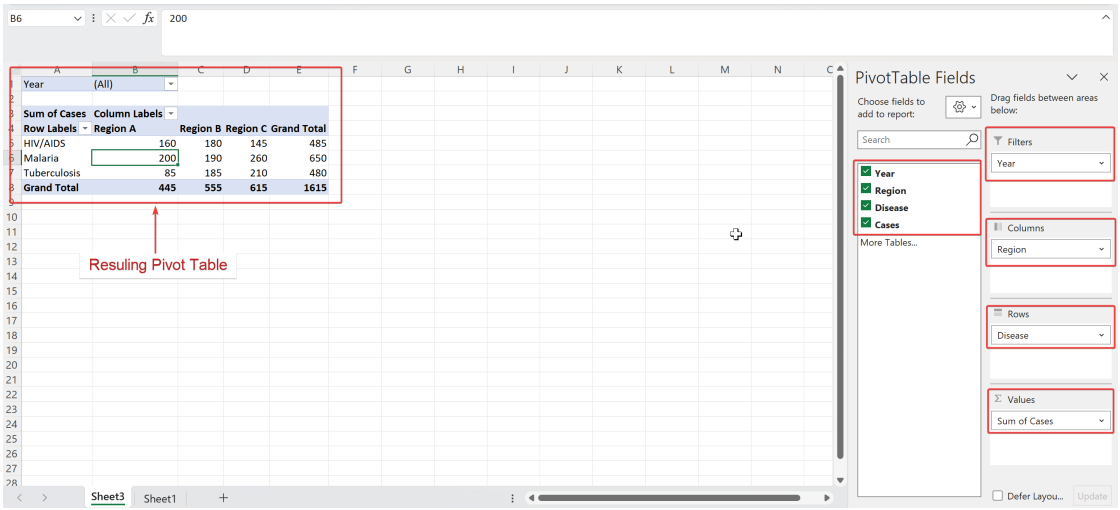
1. **Prepare Your Data** Ensure your dataset is organized with column headers and rows of data. Each column should have a clear heading, and there should be no blank rows or columns within the dataset.
2. **Select Your Data** Click anywhere inside your dataset. Alternatively, you can manually select the range of cells that contain your data.
3. **Insert PivotTable** Navigate to the "Insert" tab on the ribbon at the top of Excel. Click on "PivotTable" in the "Tables" group. This will open the Create PivotTable dialog box.
4. **Select Your Data Range** Ensure the correct range of data is selected in the "Table/Range" field of the Create PivotTable dialog box. Excel should automatically detect the range based on your selection in Step 2.
5. **Choose Where to Place the PivotTable** Select where you want your PivotTable to be placed. You can either choose to place it in a new worksheet or in an existing worksheet.
6. **Click "OK"** Once you've selected your data range and the destination for your PivotTable, click "OK". This will create a new PivotTable or insert one into the designated location.



Year	Region	Disease	Cases
2022	Region A	Malaria	120
2022	Region A	Tuberculosis	85
2022	Region B	Malaria	90
2022	Region B	Tuberculosis	110
2022	Region B	HIV/AIDS	75
2022	Region C	Malaria	150
2022	Region C	Tuberculosis	95
2023	Region C	HIV/AIDS	65
2023	Region A	Malaria	80
2023	Region A	HIV/AIDS	160
2023	Region B	Malaria	100
2023	Region B	Tuberculosis	75
2023	Region B	HIV/AIDS	105
2023	Region C	Malaria	110
2023	Region C	Tuberculosis	115
2023	Region C	HIV/AIDS	80

7. **Design Your PivotTable**

- **Fields List:** A new pane will appear on the right, called "PivotTable Fields". This pane lists all the column headers from your dataset.
- **Drag Fields to Areas:** Drag the fields you want to analyze into the different areas of the PivotTable:
  - **Values:** Drag the fields you want to perform calculations on (e.g., sums, averages) into this area.
  - **Rows:** Drag fields here to use them as row labels in your PivotTable.
  - **Columns:** Drag fields here to use them as column labels in your PivotTable.
  - **Filters:** Drag fields here to use them as filters for your PivotTable.



8. **Customize Your PivotTable**

   - **Apply Filters:** Use the dropdown arrows next to field names in the PivotTable to filter data.
   - **Change Summarization:** Right-Click on any value in the PivotTable, choose "Value Field Settings" to change the calculation type (e.g., sum, count, average).
   - **Format Your PivotTable:** Modify the PivotTable's design, style, and layout using options in the "PivotTable Analyze" tab that appears when the PivotTable is selected.

9. **Refresh Your PivotTable (if needed)** If your data changes, you can refresh your PivotTable to reflect the updated information. Simply right-click on the PivotTable and choose "Refresh".

4.4.2 Commands in the PivotTable Analyze Tab

Data

- **Refresh**: This command updates the pivot table with the latest data from the source range. It's useful when your underlying dataset has changed, and you want your pivot table to reflect those changes.
- **Change Data Source**: Use this command to modify the source data range for your pivot table. If you've added more data to your dataset or want to change the range for any reason, this option is essential.

Actions

- **Select**: Allows you to select specific elements within your pivot table, such as cells, columns, or rows, for further manipulation.
- **Move PivotTable**: You can relocate your pivot table within the same workbook or to a different worksheet using this option.
- **Clear**: Removes applied filters or customizations, or formatting from the pivot table.

Fields, Items & Sets

- **PivotTable Fields**: Opens or closes the PivotTable Fields pane, which lets you add, remove, or rearrange fields within your pivot table.
- **Show/Hide**: Customize the display of various elements in your pivot table, such as subtotals, grand totals, and field headers.

Filter

- **Insert Slicer**: When you have multiple fields in your PivotTable, inserting a slicer allows you to create visual buttons that can filter data based on the selected criteria.
- **Insert Timeline**: It allows you to filter data by specific dates using an interactive timeline control.

Calculations

- **Fields, Items & Sets**: Create calculated fields, calculated items, or calculated sets to extend the functionality of your pivot table with custom calculations.

The Pivot Table in itself can be written in one whole chapter. But here we are only giving you only the basics of the pivot table. You can learn more about pivot tables from other sources. Consider this section as an introduction to pivot tables and try mastering it since pivot table is one of the most amazing feature in Excel.

Happy Pivoting!!!

5 Power Query

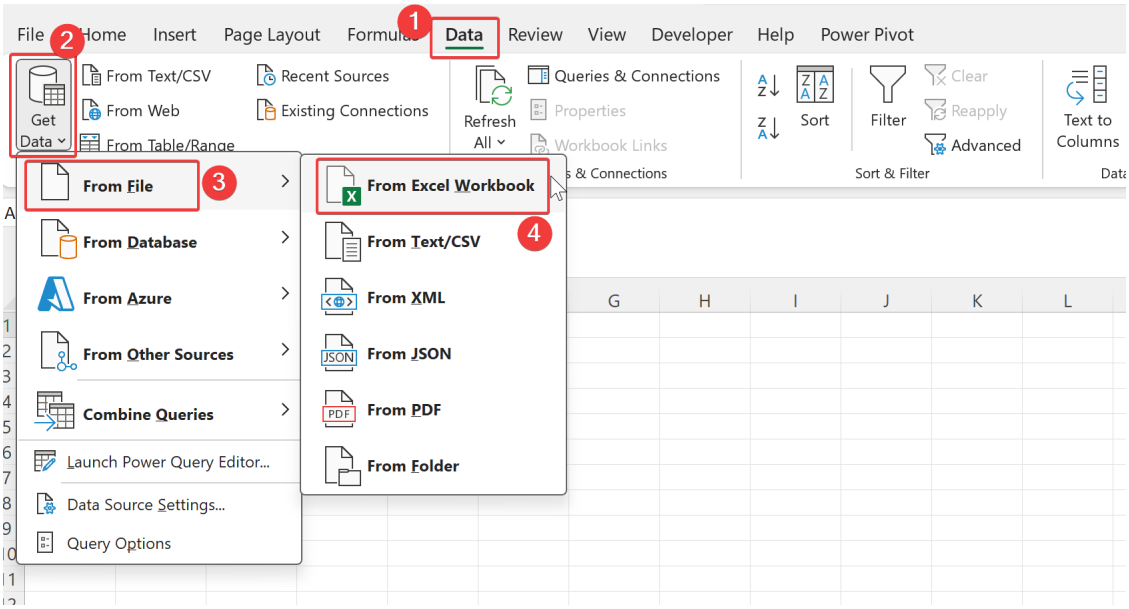
Power Query enhances the experience pf extracting data from a wide variety of sources, perform complex transformations on that data, and then load the data into a workbook. In your work, repetitive data preparation tasks can be time-consuming. Power Query enables automation of these tasks. By creating reusable query templates, you can streamline workflows and update analyses with minimal effort when new data becomes available.

5.1 Using Power Query to Load and Transform Data

5.1.1 Importing Data

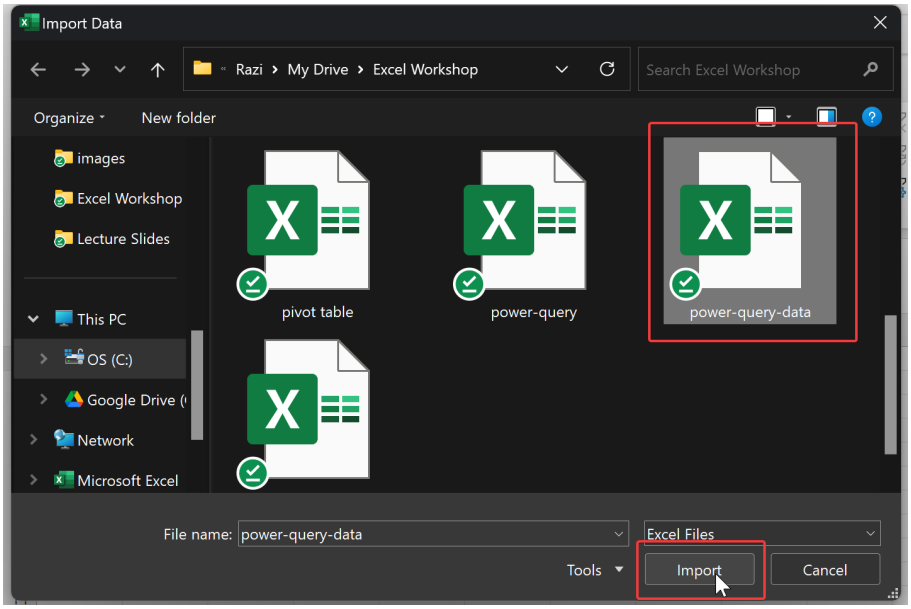
1. **Accessing Power Query**:

   - To begin using Power Query, navigate to the "Data" tab on the ribbon. From there, select "Get Data". You'll find various options like files, databases, or online services. For this example, we'll import data from an Excel workbook.

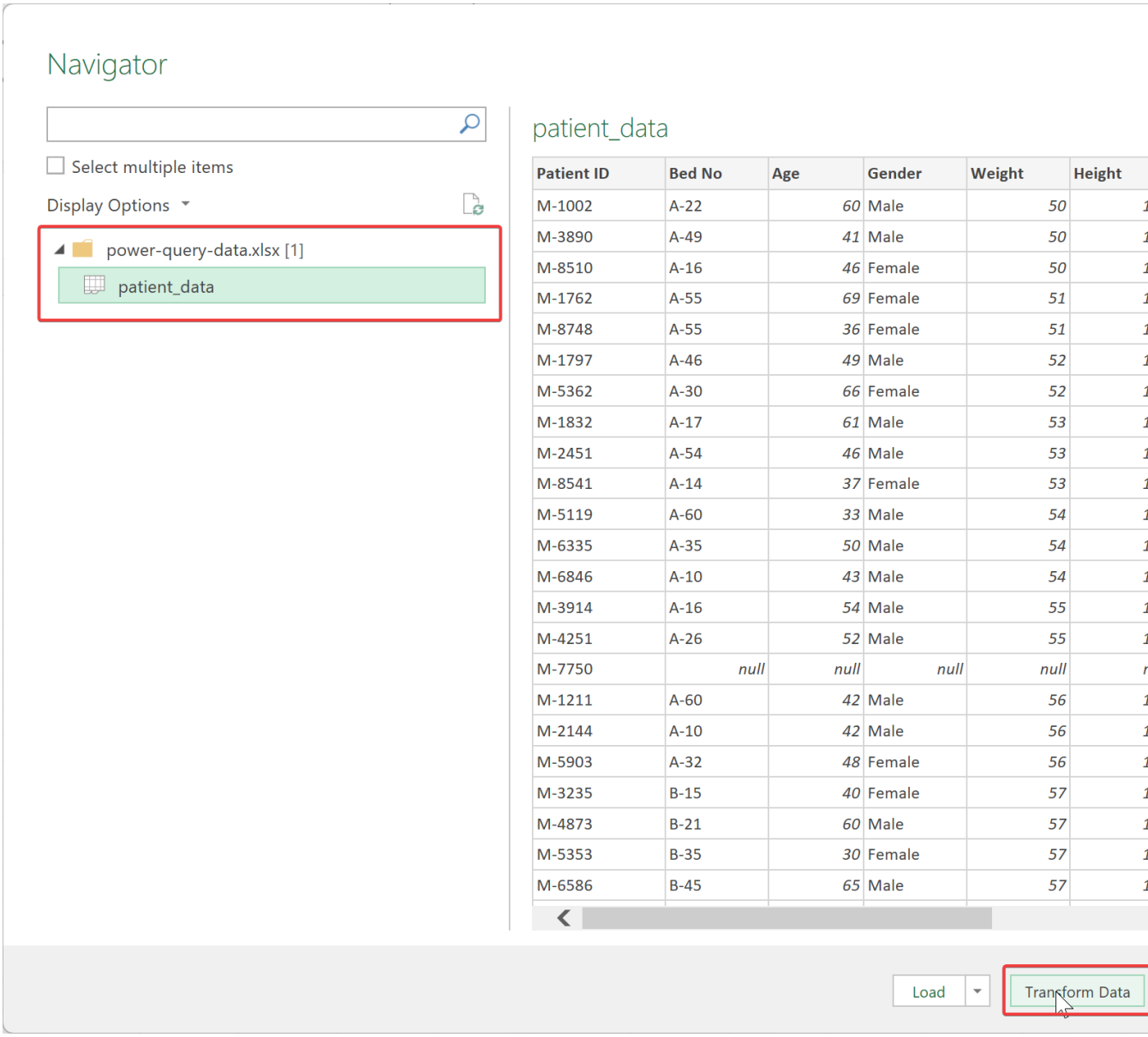


2. **Importing Data**:

- Click on “From Excel Workbook” and browse to select the file containing your data. Click “Import”.



- The Navigator pane will appear as shown in the figure. Choose the specific worksheet or named range that you want to import. When you click on each table, you can see a preview of the data.
- Click on “Transform Data”. It will open Power Query Editor window, where you can transform the data (clean, add more columns, filter unwanted data, etc) before loading into Excel. If you don’t wish to tranform the data or your data is already cleaned, then you can click on the “Load” button.



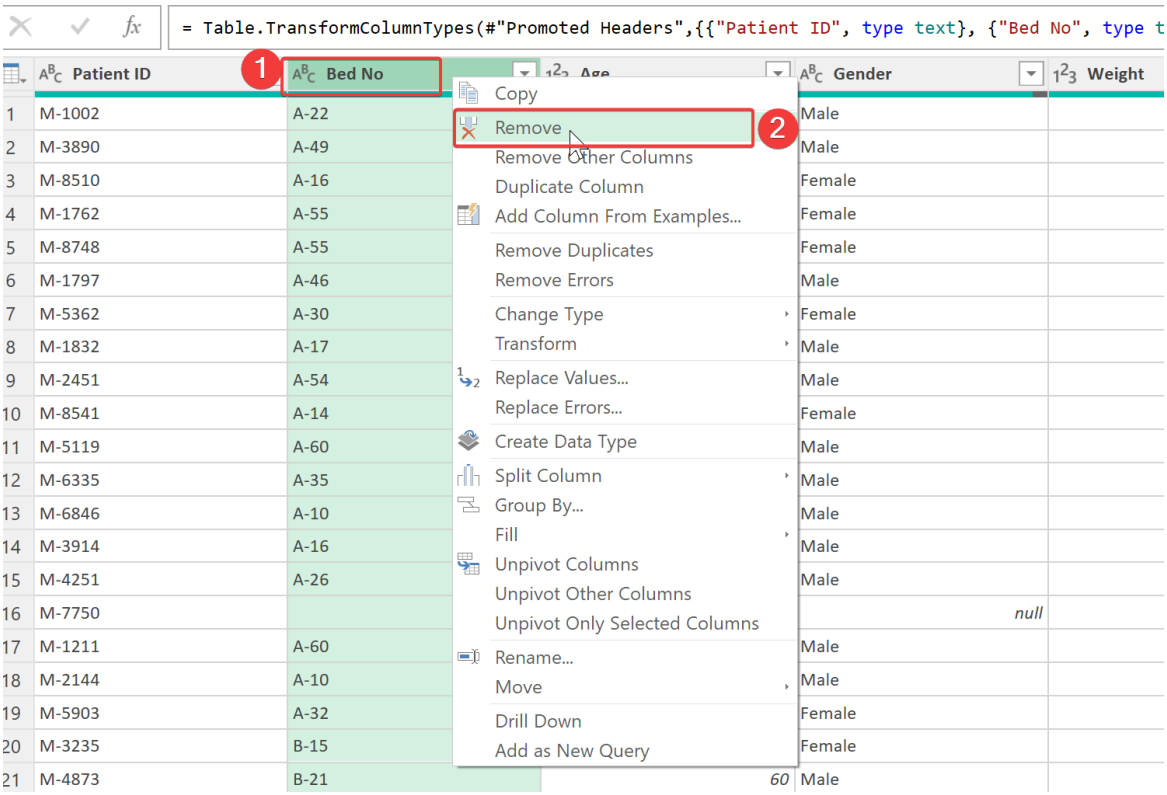
Patient ID	Bed No	Age	Gender	Weight	Height
M-1002	A-22	60	Male	50	1
M-3890	A-49	41	Male	50	1
M-8510	A-16	46	Female	50	1
M-1762	A-55	69	Female	51	1
M-8748	A-55	36	Female	51	1
M-1797	A-46	49	Male	52	1
M-5362	A-30	66	Female	52	1
M-1832	A-17	61	Male	53	1
M-2451	A-54	46	Male	53	1
M-8541	A-14	37	Female	53	1
M-5119	A-60	33	Male	54	1
M-6335	A-35	50	Male	54	1
M-6846	A-10	43	Male	54	1
M-3914	A-16	54	Male	55	1
M-4251	A-26	52	Male	55	1
M-7750	null	null	null	null	r
M-1211	A-60	42	Male	56	1
M-2144	A-10	42	Male	56	1
M-5903	A-32	48	Female	56	1
M-3235	B-15	40	Female	57	1
M-4873	B-21	60	Male	57	1
M-5353	B-35	30	Female	57	1
M-6586	B-45	65	Male	57	1

5.1.2 Simple Data Transformation in Power Query

Once you've imported your data into Power Query, you can perform several transformations to clean and refine it form the Power Query Editor Window. Here, we are mentioning only a few of them.

1. **Removing Unnecessary Columns**:

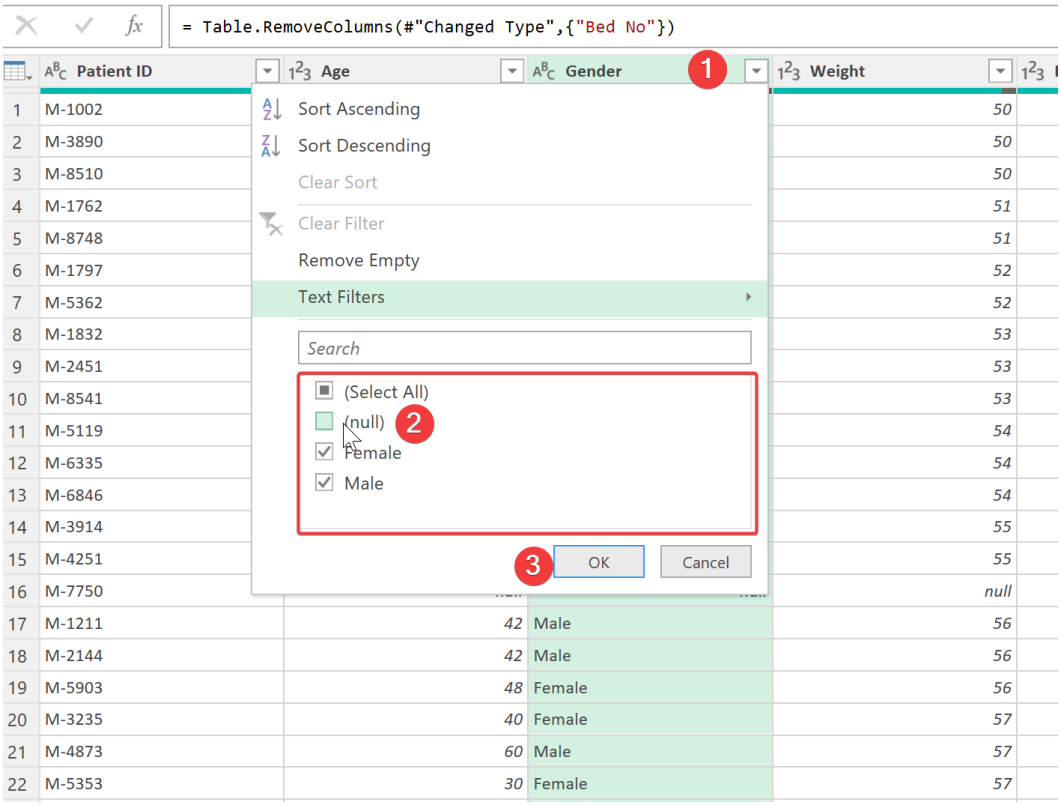
- In the Power Query Editor window, review the data in the "Queries" pane on the left.
- Click on a query to see its data preview in the main window.
- To remove columns, select the column header(s), right-click, and choose "Remove" or "Remove Other Columns" to keep only the desired data.



	Patient ID	Bed No	Age	Gender	Weight
1	M-1002	A-22		Male	
2	M-3890	A-49		Male	
3	M-8510	A-16		Female	
4	M-1762	A-55		Female	
5	M-8748	A-55		Female	
6	M-1797	A-46		Male	
7	M-5362	A-30		Female	
8	M-1832	A-17		Male	
9	M-2451	A-54		Male	
10	M-8541	A-14		Female	
11	M-5119	A-60		Male	
12	M-6335	A-35		Male	
13	M-6846	A-10		Male	
14	M-3914	A-16		Male	
15	M-4251	A-26		Male	
16	M-7750			null	
17	M-1211	A-60		Male	
18	M-2144	A-10		Male	
19	M-5903	A-32		Female	
20	M-3235	B-15		Female	
21	M-4873	B-21	60	Male	

2. **Filtering Rows**:

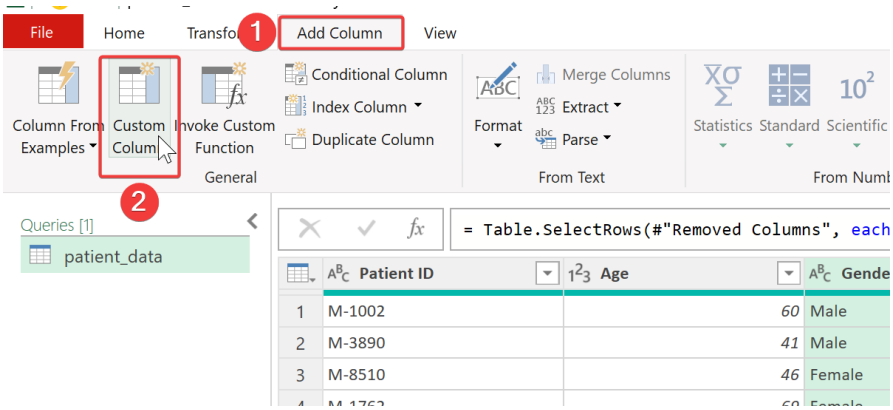
- Use filters to exclude rows based on specific criteria.
- Click on the filter icon in the column header and define your filter conditions (e.g., only include rows where a certain column's value is less than a particular number).
- You can also filter to remove blank rows.



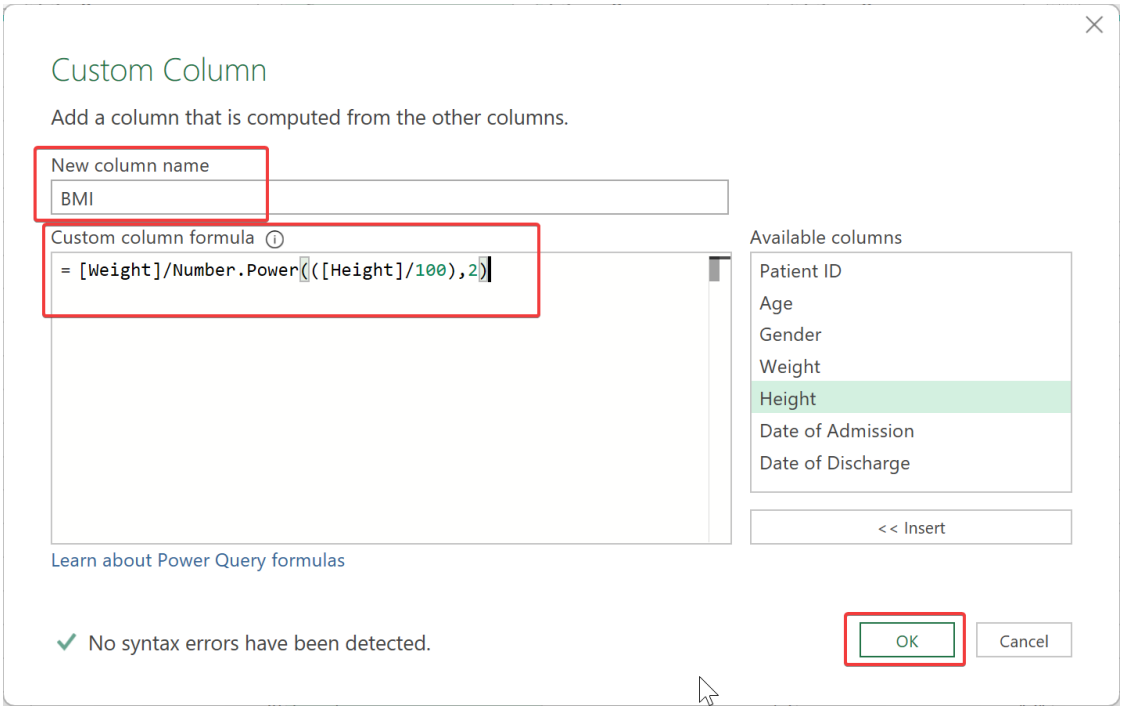
	Patient ID	Age	Gender	Weight
1	M-1002			50
2	M-3890			50
3	M-8510			50
4	M-1762			51
5	M-8748			51
6	M-1797			52
7	M-5362			52
8	M-1832			53
9	M-2451			53
10	M-8541			53
11	M-5119			54
12	M-6335			54
13	M-6846			54
14	M-3914			55
15	M-4251			55
16	M-7750			null
17	M-1211	42	Male	56
18	M-2144	42	Male	56
19	M-5903	48	Female	56
20	M-3235	40	Female	57
21	M-4873	60	Male	57
22	M-5353	30	Female	57

3. **Adding Custom Columns**:

- Sometimes you need to derive new information from existing data.
- Click on "Add Column" in the ribbon and choose "Custom Column."

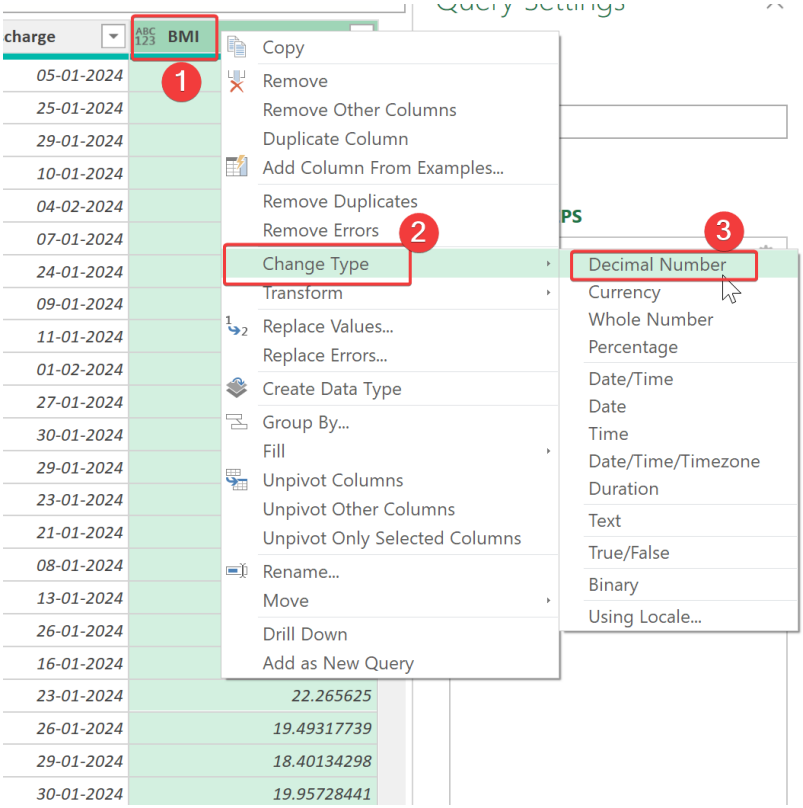


- Write a formula to create a new calculated column based on your requirements.

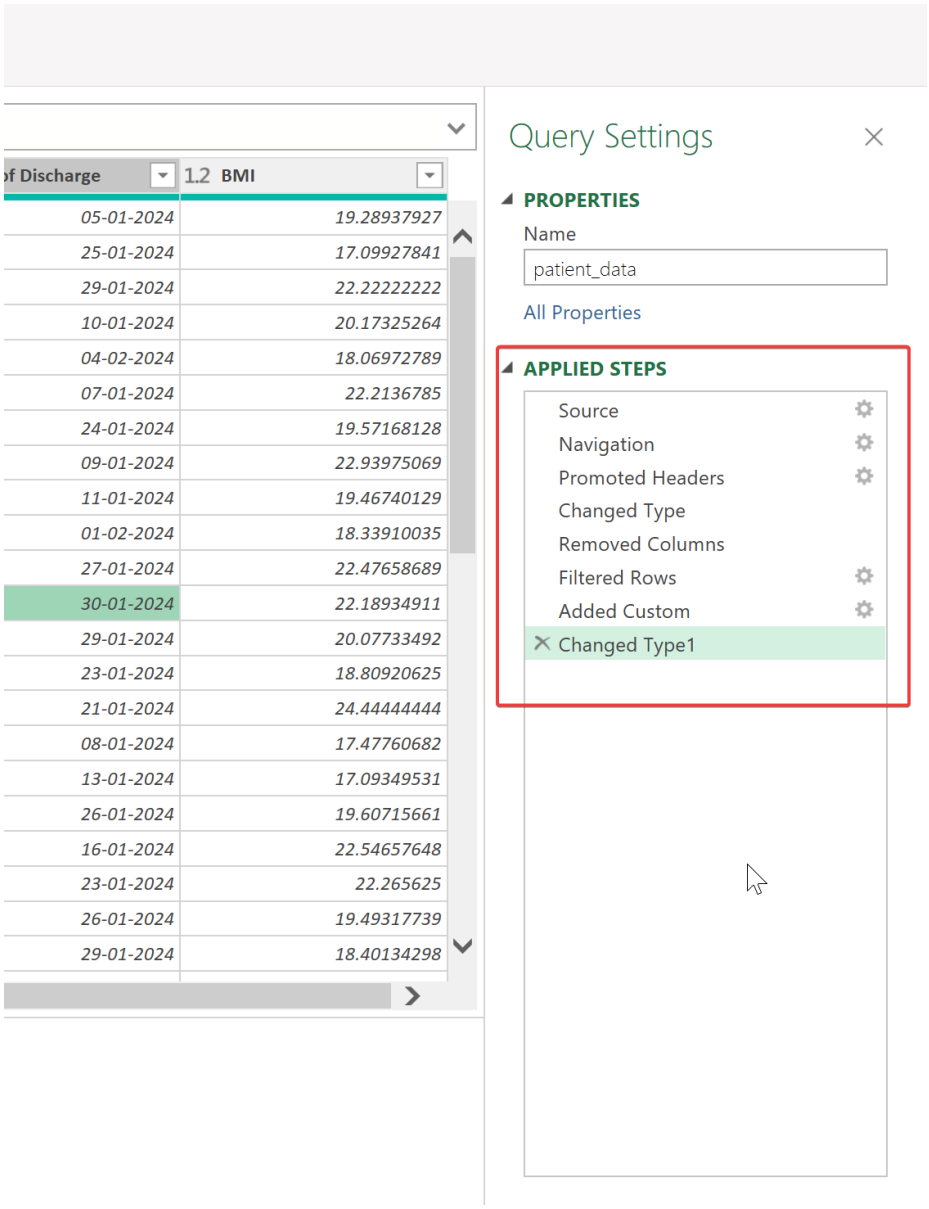


4. **Data Type and Format Adjustments**:

   - Ensure that each column is assigned the correct data type (e.g., text, number, date) to avoid issues during analysis.
   - Use the "Transform" tab to adjust formats or convert data types.
   - You can also right-click on the column header and choose "Change Type" to convert the data type.



The Applied Steps pane displays a sequential list of all the transformation steps that have been applied to your data. Each step represents a specific action or transformation applied to the dataset.



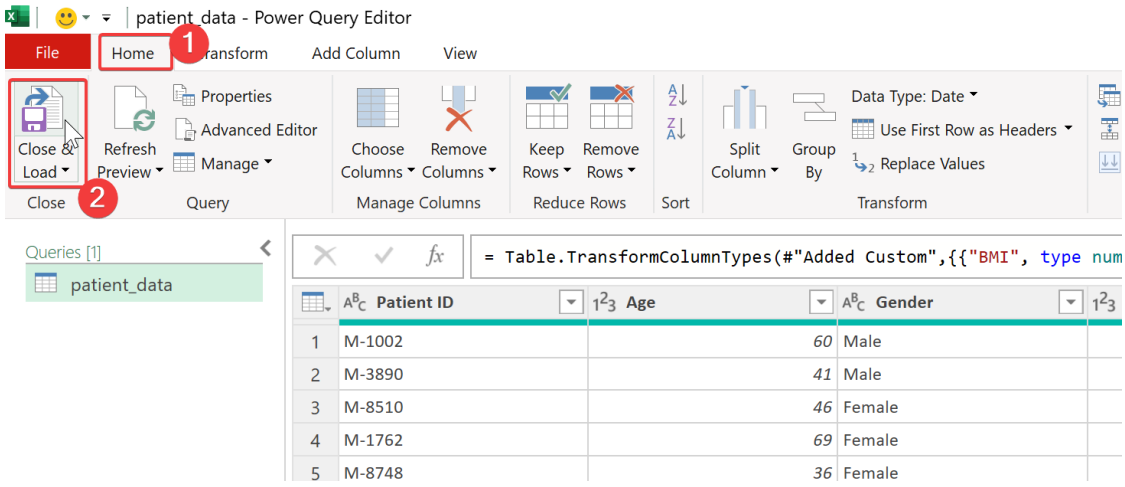
5.1.3 Applying Changes and Loading Data

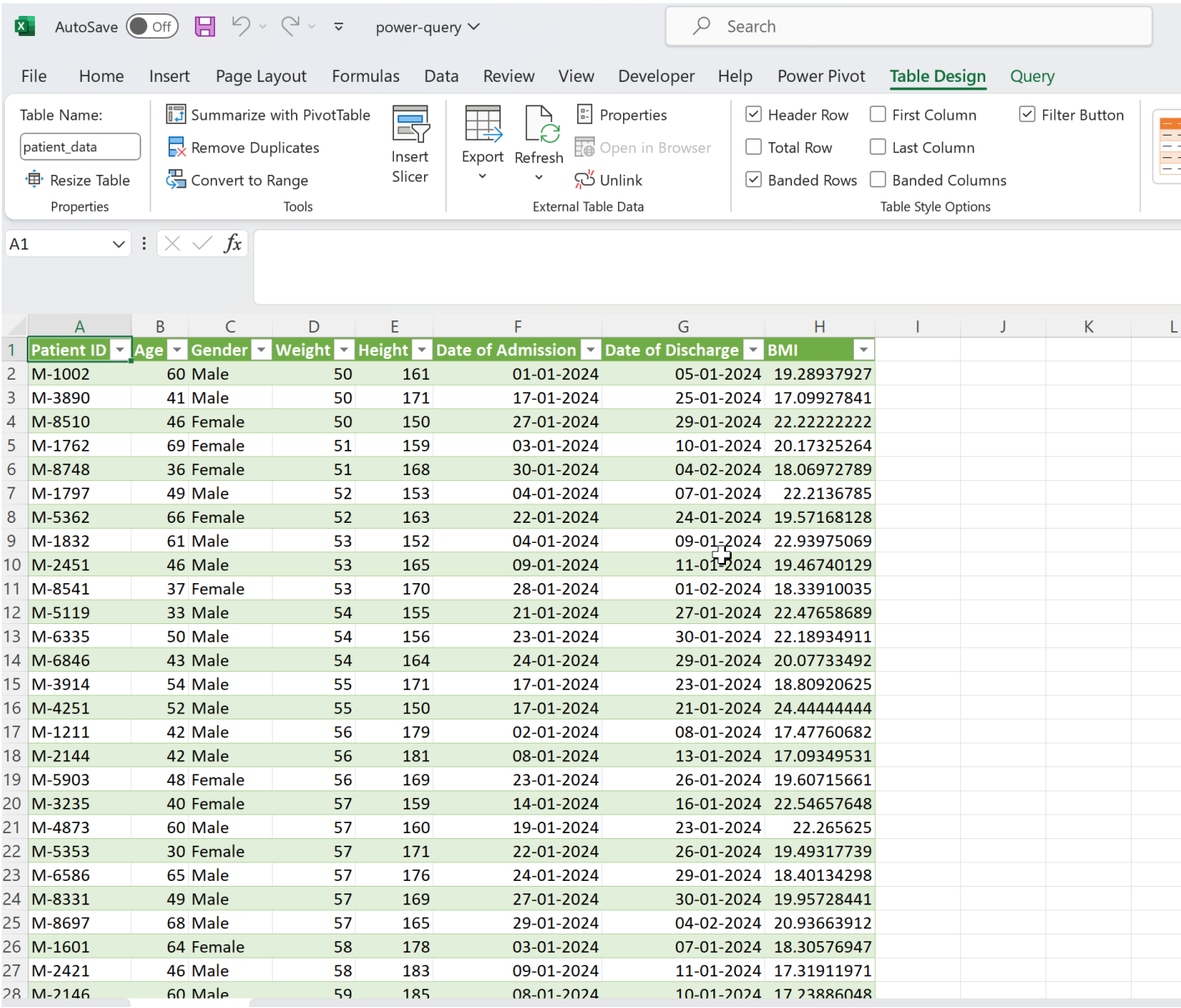
After applying the necessary transformations:

1. **Preview Changes**:
    - Review the updated data in the Power Query Editor to ensure it meets your needs.

2. **Applying Changes**:

   - Click on "Close & Load" to bring the cleaned data back into Excel.





Patient ID	Age	Gender	Weight	Height	Date of Admission	Date of Discharge	BMI
M-1002	60	Male	50	161	01-01-2024	05-01-2024	19.28937927
M-3890	41	Male	50	171	17-01-2024	25-01-2024	17.09927841
M-8510	46	Female	50	150	27-01-2024	29-01-2024	22.22222222
M-1762	69	Female	51	159	03-01-2024	10-01-2024	20.17325264
M-8748	36	Female	51	168	30-01-2024	04-02-2024	18.06972789
M-1797	49	Male	52	153	04-01-2024	07-01-2024	22.2136785
M-5362	66	Female	52	163	22-01-2024	24-01-2024	19.57168128
M-1832	61	Male	53	152	04-01-2024	09-01-2024	22.93975069
M-2451	46	Male	53	165	09-01-2024	11-01-2024	19.46740129
M-8541	37	Female	53	170	28-01-2024	01-02-2024	18.33910035
M-5119	33	Male	54	155	21-01-2024	27-01-2024	22.47658689
M-6335	50	Male	54	156	23-01-2024	30-01-2024	22.18934911
M-6846	43	Male	54	164	24-01-2024	29-01-2024	20.07733492
M-3914	54	Male	55	171	17-01-2024	23-01-2024	18.80920625
M-4251	52	Male	55	150	17-01-2024	21-01-2024	24.44444444
M-1211	42	Male	56	179	02-01-2024	08-01-2024	17.47760682
M-2144	42	Male	56	181	08-01-2024	13-01-2024	17.09349531
M-5903	48	Female	56	169	23-01-2024	26-01-2024	19.60715661
M-3235	40	Female	57	159	14-01-2024	16-01-2024	22.54657648
M-4873	60	Male	57	160	19-01-2024	23-01-2024	22.265625
M-5353	30	Female	57	171	22-01-2024	26-01-2024	19.49317739
M-6586	65	Male	57	176	24-01-2024	29-01-2024	18.40134298
M-8331	49	Male	57	169	27-01-2024	30-01-2024	19.95728441
M-8697	68	Male	57	165	29-01-2024	04-02-2024	20.93663912
M-1601	64	Female	58	178	03-01-2024	07-01-2024	18.30576947
M-2421	46	Male	58	183	09-01-2024	11-01-2024	17.31911971
M-2146	60	Male	59	185	08-01-2024	10-01-2024	17.23886048

This chapter is currently under development, and additional content will be added shortly.