SOFTWARE PACKAGES FOR ECONOMIC ANALYSIS (SPREADSHEET)

POSTGRADUATE PROGRAMME ECONOMICS

M23EC01AC

Self Learning Material

SREENARAYANAGURU OPEN UNIVERSITY

SREENARAYANAGURU OPEN UNIVERSITY

The State University for Education, Training and Research in Blended Format, Kerala

Vision

To increase access of potential learners of all categories to higher education, research and training, and ensure equity through delivery of high quality processes and outcomes fostering inclusive educational empowerment for social advancement.

Mission

To be benchmarked as a model for conservation and dissemination of knowledge and skill on blended and virtual mode in education, training and research for normal, continuing, and adult learners.

Pathway

Access and Quality define Equity.

Software Packages for Economics Analysis (Spreadsheet)

Course Code: M23EC01AC

Semester-I

Ability Enhancement Compulsory Course
Master of Arts Economics
Self Learning Material

SREENARAYANAGURU
OPEN UNIVERSITY

SREENARAYANAGURU OPEN UNIVERSITY

The State University for Education, Training and Research in Blended Format, Kerala

Documentation

M23EC01AC
Software Packages for Economics Analysis (Spreadsheet)
Semester -I

SREENARAYANAGURU
OPEN UNIVERSITY

www.sgou.ac.in

ISBN 978-81-967121-4-3

Academic Committee

Dr. Anitha V
Mr. Santhosh T Varghese
Dr. B. Pradeepkumar
Dr. Christabella P.J.
Dr. Aparna Das
Dr. S. Jayasree
Dr. Prasad A.K.
Dr. C.C. Babu
Dr. Sindhu Prathap
Dr. Mothi George

Development of the content

Yedu T. Dharan

Review

Content: Santhosh T. Varghese
Format: Dr. I.G. Shibi
Linguistic: Dr. Gopakumar C.

Edit

Santhosh T. Varghese

Scrutiny

Yedu T. Dharan, Soumya V. D., Dr. Suchithra K.R., Soumya V.D., Hima Chandran J., Muneer K.

Co-ordination

Dr. I.G. Shibi and Team SLM

Design Control

Azeem Babu T.A.

Production

December 2023

Copyright

Message from Vice Chancellor

Dear

I greet all of you with deep delight and great excitement. I welcome you to the Sreenarayanaguru Open University.

Sreenarayanaguru Open University was established in September 2020 as a state initiative for fostering higher education in open and distance mode. We shaped our dreams through a pathway defined by a dictum 'access and quality define equity'. It provides all reasons to us for the celebration of quality in the process of education. I am overwhelmed to let you know that we have resolved not to become ourselves a reason or cause a reason for the dissemination of inferior education. It sets the pace as well as the destination. The name of the University centers around the aura of Sreenarayanaguru, the great renaissance thinker of modern India. His name is a reminder for us to ensure quality in the delivery of all academic endeavors.

Sreenarayanaguru Open University rests on the practical framework of the popularly known "blended format". Learner on distance mode obviously has limitations in getting exposed to the full potential of classroom learning experience. Our pedagogical basket has three entities viz Self Learning Material, Classroom Counselling and Virtual modes. This combination is expected to provide high voltage in learning as well as teaching experiences. Care has been taken to ensure quality endeavours across all the entities.

The university is committed to provide you stimulating learning experience. The PG programme in Economics is conceived to be a continuum of the UG programme in Economics as it has organic linkage with the content and the form of treatment. In fact is a progression of the finer aspects of theories and practices. The discussions are meant to arouse interest among the learners in understanding the discipline in the real context and therefore, the examples are drawn heavily from the real life experiences. The provision for empirical evidences integrated endeavour of the academic content makes this programme special and relevant. We assure you that the university student support services will closely stay with you for the redressal of your grievances during your studentship.

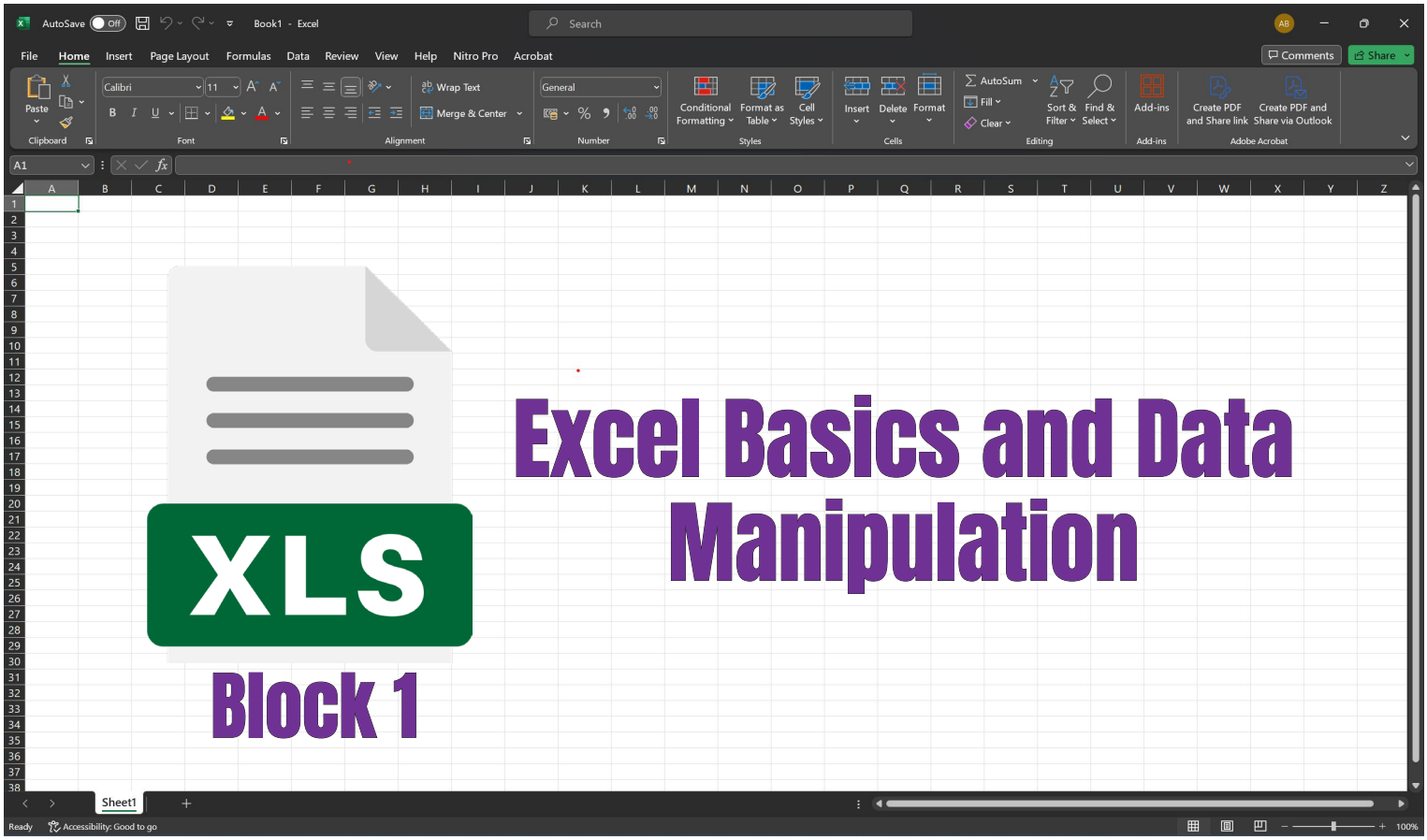
Feel free to write to us about anything that you feel relevant regarding the academic programme.

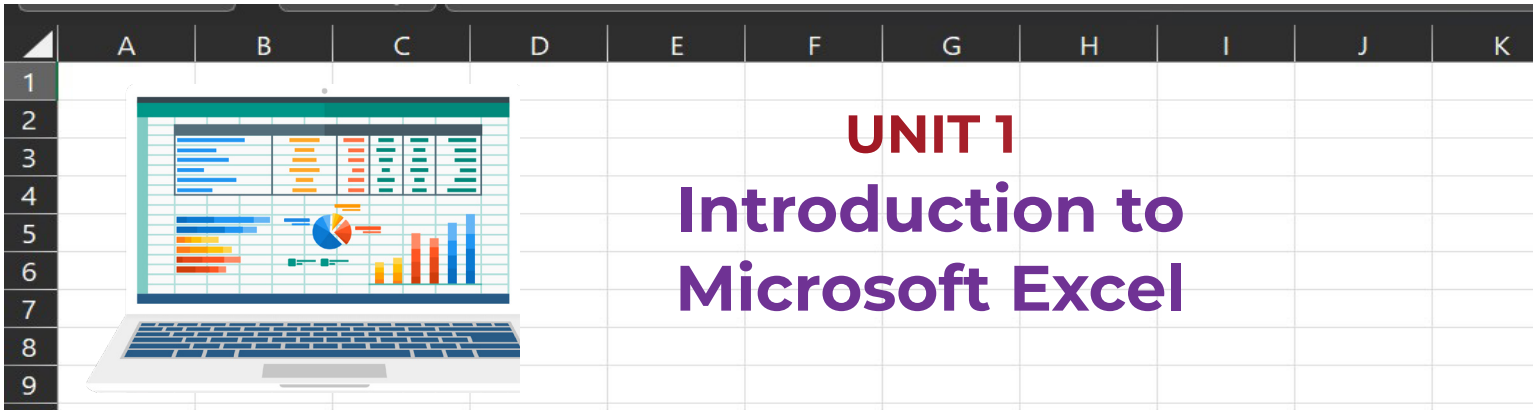
Wish you the best.

Dr. P.M. Mubarak Pasha

01.12.2023

Contents


AutoSave
Book1 - Excel
Search
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Help
Nitro Pro
Acrobat
Comments
Share
Paste
Calibri
11
Wrap Text
Merge & Center
General
Conditional Formatting
Format as Table
Cell Styles
Insert
Delete
Format
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Add-ins
Create PDF and Share link
Create PDF and Share via Outlook
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
Add-ins
Adobe Acrobat
A1
XLS
Excel Basics and Data Manipulation
Block 1
Sheet1
Ready
Accessibility: Good to go
100%



UNIT 1
Introduction to Microsoft Excel

Learning Outcomes

After the completion of this unit, the learner will be able to

- identify the main features of Excel
- navigate and select data in Excel
- enter data and formulas in cells accurately
- format cells, rows, and columns appropriately
- apply basic formatting options to enhance your data presentation

Background

Data refers to raw facts, observations, or measurements that are collected and documented. It may be expressed in a number of ways, such as text, photos, audio, and video. Data can be both structured and unstructured. Structured data is organised in a predefined format, typically in tables or spreadsheets, with clearly defined categories and relationships between the data elements. Unstructured data, on the other hand, lacks a specific format and may include things like text documents, social media posts, or multimedia files. Variables are categorised based on their level of measurement. There are four main levels: nominal (like gender or ethnicity, representing categories with no specific order), ordinal (such as educational levels or customer satisfaction ratings, where there is a meaningful order), interval (like temperature or IQ scores, with consistent intervals), and ratio (including variables like height or income, which have a true zero point).

Data is the foundation for generating information and knowledge. It can be collected through various methods, such as surveys, experiments, sensors, or online sources. Once collected, data can be processed, analysed, and interpreted to extract insights, identify patterns, make informed decisions, or support research in different domains. Data is a valuable resource for businesses, organisations, governments, and research-

ers. It can be used for purposes like market research, customer analysis, forecasting, risk assessment, policy-making, scientific discoveries, and much more. Effective management, analysis, and interpretation of data are crucial for leveraging its potential and deriving meaningful outcomes.

Excel serves as a powerful tool for enabling efficient data management. It offers essential features and functionalities that facilitate various aspects of data management, including organisation, analysis, and visualisation. With Excel, users can create structured tables, sort, and filter data, and perform calculations using built-in formulas and functions. The software also provides charting and graphing capabilities for visualising data trends and patterns. Importing and exporting data from different sources is seamless, and Excel includes options for data validation and security. Overall, Excel empowers users to effectively organise, analyse, and derive insights from data, making it an invaluable asset for data management tasks.

Keywords

Workbook, Worksheet, Data Management, Data Formatting, Data Navigation

Discussion

1.1.1 Introduction to Microsoft Excel

- Spreadsheet software that supports creation, editing, and analysis of data

Introduced by Microsoft in 1985 for the Macintosh computer, Microsoft Excel is an exceptional spreadsheet software that supports creation, editing, and analysis of data in various arrangements. Excel was code-named “Odyssey” during its development stage. As a pioneer of spreadsheet programs, Excel set itself apart by incorporating pull-down menus and mouse clicks into the interface, which was then graphical. It gained its market leadership following the release of Excel 2.0 for Windows in 1987 that boasted an image-intensive interface as well as numerous new features. Subsequently, updates were released regularly, introducing additional compatibility, security, and functionality. There exist numerous versions of Microsoft’s Excel, ranging from Excel 95 to Excel 2021. Also, the Microsoft Office suite offers the mobile and web browser option for Excel.

- Excel supports data creation, editing, and analysis.

Excel offers an abundance of powerful features for analysing and visualising data. These capabilities include creating different types of charts, such as pie, bar, line, scatter plots, histograms, and more. Pivot tables and pivot charts provide a tool to aggregate large amounts of data by filtering it according to various criteria. Additionally, users can perform advanced analysis using Solver, Goal Seek, Scenario Manager and other tools like Data Table. With Power Query or Forecast Sheet users can work with large volumes of data promptly.

Microsoft Excel offers a wide range of uses and applications across various industries and disciplines. Here are some detailed explanations of the key uses of Microsoft Excel:

- Data Organisation and Analysis

- Excel provides an efficient and structured platform for organising and analysing data. It allows users to input, store, and manipulate large amounts of data in a tabular format. With features such as sorting, filtering, and data validation, Excel enables users to effectively manage and make sense of their data.

- Mathematical and Statistical Calculations

- A wide variety of mathematical and statistical functions are included in Excel. Arithmetic operations, statistical analysis, forecasting, and trend analysis are just a few of the complicated tasks that users may carry out using Excel. Excel's formula functionality allows users to create customised formulas and perform calculations on data sets effortlessly.

- Financial Management and Planning

- Excel is widely used in finance and accounting for budgeting, financial analysis, and planning. It offers financial functions such as compound interest calculations, loan amortisation, and financial modelling. Excel's ability to create dynamic charts and graphs facilitates visual representation of financial data for effective decision-making.

- Reporting and Visualisation

- Excel provides powerful tools for creating visually appealing reports and dashboards. Users can summarise and present data using tables, charts, graphs, and pivot tables. Excel's formatting options allow users to customise the appearance of their reports, making them more engaging and understandable for stakeholders.

- Project Management and Tracking

- Excel is often utilised for project management tasks, such as creating Gantt charts, tracking project timelines, and monitoring progress. With features like conditional formatting and data validation, Excel enables users to set up automated alerts and reminders, ensuring project deadlines and milestones are met.

- Data Import and Integration

- Excel seamlessly integrates with other data sources and applications, allowing users to import data from various formats such as CSV, text files, databases, and web sources. This integration enables data consolidation, analysis, and reporting in a single platform, streamlining workflows, and enhancing productivity.

- Collaboration and Sharing

- Spreadsheet collaboration and sharing among team members are supported by Excel. The ability for several users to interact on the same file at once facilitates decision-making, report creation, and data analysis. Excel also offers security features to protect sensitive data and control user access.

- Automation and Macros

- Excel allows users to automate repetitive tasks using macros and Visual Basic for Applications (VBA). Macros enable users to record a series of actions and play them back, saving time and effort. VBA programming enables advanced automation, customisation, and the creation of user-defined functions.

These are just a few of the uses of Microsoft Excel. Its versatility, flexibility, and extensive features make it an indispensable tool for data management, analysis, reporting, and decision-making across industries and professions.

- Empowers economists for data analysis, financial modelling, forecasting, and decision-making.

Excel is an essential tool for economists, offering a range of features that facilitate data analysis, financial modelling, and forecasting. Economists can efficiently organise and manipulate large datasets, perform complex calculations using formulas and functions, and visualise data through graphs and charts. Excel enables economists to identify patterns, trends, and relationships within economic data, making it easier to draw insights and communicate findings effectively. It also supports financial modelling and scenario analysis, allowing economists to evaluate investment projects, simulate economic scenarios, and assess policy impacts. With Excel's collaborative capabilities, economists can work together, share insights, and enhance research quality. Overall, Excel is a vital companion for economists, providing the computational power and tools necessary for their data-driven analysis, decision-making, and economic understanding.

Example: Given below is a sample data set. The given data set represents the scores of 10 students in different subjects. Each student has a unique Student ID, Student Name, and scores in Accountancy, Economics, Business Studies, Maths, and English.

Table 1.1.1 Student's Marklist

Sl. No.	Name of the Student	Accountancy	Economics	Business Studies	Mathematics	English
1	Aryan	85	90	92	88	89
2	Riya	92	88	95	87	91
3	Rohit	78	80	85	79	83
4	Nisha	95	92	93	96	90
5	Deepak	86	88	82	85	90
6	Pooja	90	92	87	89	88
7	Mohit	83	85	78	80	79
8	Sneha	92	95	90	91	93
9	Rahul	88	90	92	85	87
10	Komal	85	87	90	88	86

- Datasets can be organised for analysis

The above given data set can be entered into Excel and can be organised in a professional way for effective calculations and analysis. Excel provides a structured platform to input and organise data in rows and columns. This makes it easy to manage and retrieve information effectively.

1.1.2 Overview of Excel Interface and Features

In Excel, understanding the concepts of workbook, worksheet, and cells is fundamental to effectively working with data. So let us start from the basics of Excel.

1.1.2.1 Starting Excel

To open Excel in computer, follow the steps given below:

- Click on “Start menu”
- Then look for “All Programs”
- Double-click on the “Excel” icon to launch the application

After launching Excel, the Excel Start Screen will appear as shown in the Fig 1.1.1. In this screen, you can see options - to create a new workbook, choose an already existing template, and access the workbooks recently saved and edited by the user.

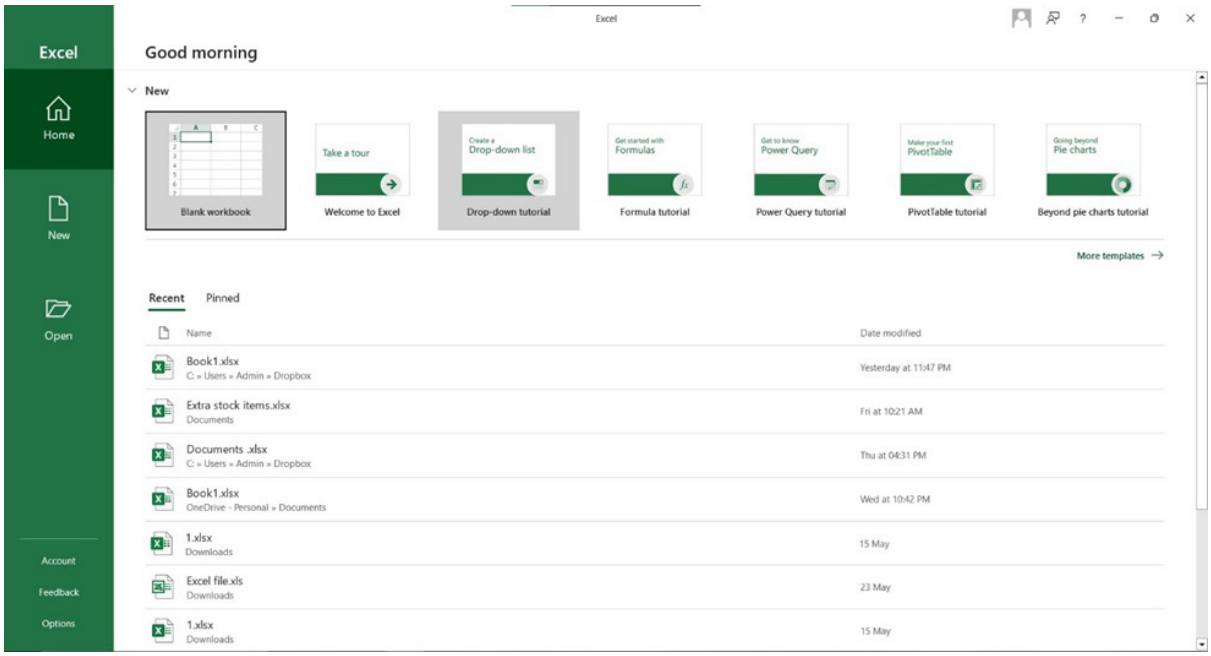


Fig 1.1.1 Excel Starter Screen

From the Start Screen, select “Blank workbook” to open the Excel interface and to start working on a new Excel work-sheet. This is as shown in the Fig 1.1.2. The worksheet starter screen is the initial interface that appears when you open the Excel application.

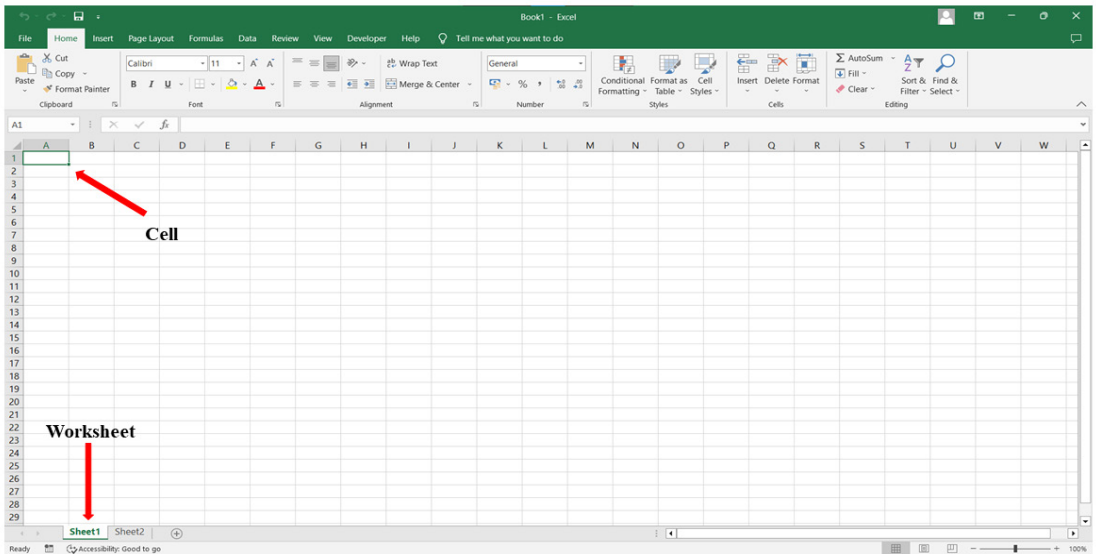


Fig 1.1.2 Workbook

1.1.2.2 Understanding Workbook, Worksheet and Cells

- Provides structured framework for efficient data management

A **Workbook** consists of a number of distinct worksheets, each of which can store data. Worksheets are grids made up of columns and rows, with cells where data is entered. Workbooks allow for organising related data, applying formatting, adding charts and images, and performing calculations using formulas and functions. They can be saved as files, shared with others, and provide a structured framework for managing and analysing data in Excel. A workbook will be as shown in the Fig 1.1.2.

- Facilitates data organisation, complex calculations, formatting, and data analysis

A **Worksheet** is a collection of rows and columns. Excel worksheets offer a range of features and functionalities, including the ability to perform complex calculations, create charts and graphs, and apply formatting to cells. You can easily adjust the size of rows and columns to accommodate different types of data. Worksheets provide a clear and organised structure for data analysis and reporting. With the ability to switch between worksheets within a workbook, you can work with multiple sets of data or organise information into different categories. By default, a new workbook contains one worksheet, but you can add or delete worksheets as needed. Each worksheet is identified by a name, usually displayed at the bottom of the Excel window as shown as Sheet1, Sheet2 in the Fig 1.1.2. A row and a column intersect to form a cell.

- Basic units of data entry and storage

Cells are the basic units where data is entered and stored within a worksheet. A unique column letter and row number is used to identify each cell. They can contain different types of information, such as text, numbers, formulas, or functions. Cells serve as the building blocks for organising, manipulating, and analysing data in Excel. Cells offer a range of functionalities and customisation options. You can perform calculations by using formulas or functions that reference other cells, allowing for efficient data analysis. Cells can be formatted to display data in various formats, such as dates or currency, and can be customised with different fonts, colours, and alignments. They can also be copied, moved, or formatted based on specific conditions or data values

1.1.2.3 Excel Interface and Features

The Excel interface consists of various components that facilitate data management, analysis, and visualisation. When you start the excel application the basic window of Excel will be as shown in the Fig 1.1.3.

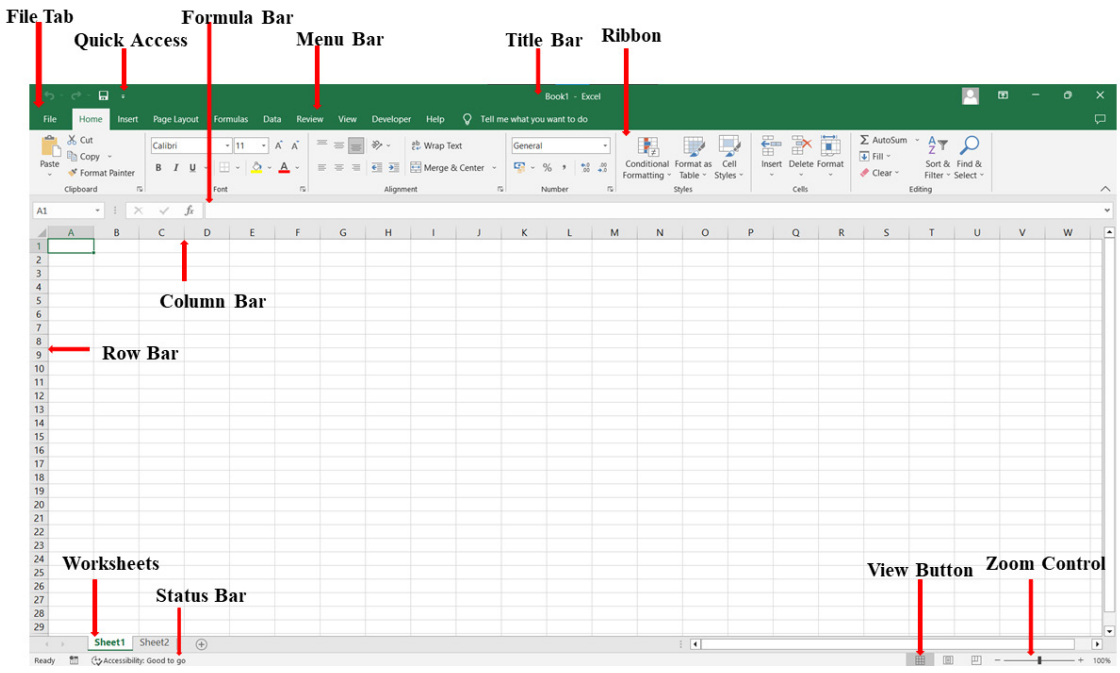


Fig 1.1.3 Components of Excel Sheet

The major components of an Excel sheet are as follows :

File Tab: The Office button from Excel 2007 has been replaced with the File tab. You can click it to go to the Backstage view, where you can find options to open or save files, create new sheets, print a sheet, and do other file-related operations.

Quick Access Toolbar: The Quick Access Toolbar is available in the upper left corner. As the name suggests it included functions that you want to use quickly and is simple to customise the options available in that.

Title bar: It is available at the top-middle part of the window. Title bar shows titles of the program and the working sheet.

Ribbon: It contains all the main tabs and its subsequent features. Each tab is divided into sub-groups, for example the Home tab contains Clipboard, Font, Alignment, Number and Styles, etc. Each group has also a triangular icon in the lower right corner and clicking on this will bring up additional options.

Formula bar: This is the area where you can enter the formula.

Row Bar: The numbering of Rows starts from 1 and keeps on increasing as you enter data. Maximum limit is 1,048,576 rows.

- Provide efficient access to file operations, customisation options, formula input, management capabilities

Column Bar: Columns are numbered from A and keeps on increasing as you keep entering data. After Z, it will start the series of AA, AB and so on. Maximum limit is 16,384 columns.

Status Bar: This shows the status of the active cell in the worksheet. A cell can be in one of the four states (i) Ready mode, (ii) Edit mode, (iii) Enter mode and (iv) Point mode.

Zoom Control: You may zoom in or out by moving the slider on the zoom control to the left or right. The + buttons can be clicked to increase the zoom factor.

View Buttons: You can switch between sheet views by pressing any of the three of buttons that are situated to the left of the Zoom control, towards the bottom of the screen.

1.1.3 Creating a New Workbook

Creating a new workbook in Excel involves starting a new file that serves as a container for multiple worksheets, allowing you to work with different sets of data and perform various tasks within a single file. There are many ways to start working with a workbook in Excel. This is shown in the Fig 1.1.4

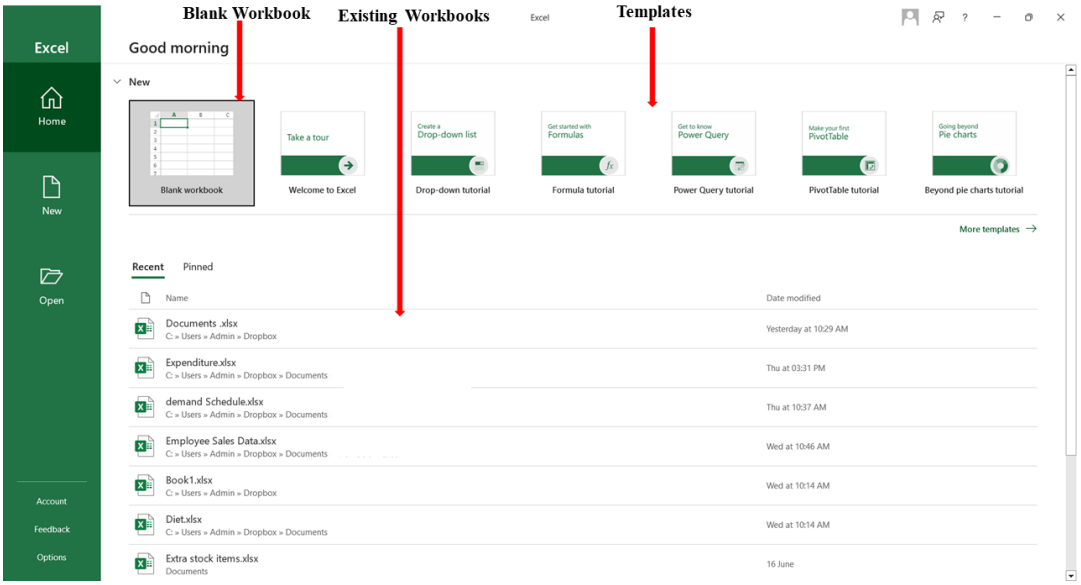


Fig 1.1.4 Creating a New Workbook

You can either choose to create a new workbook or select an already existing template or open an already existing workbook.

1.1.3.1 Creating New Worksheet

When you open Excel for the first time, it will show a starter screen and you can either create a new workbook or select an already existing file or choose from any existing templated as shown in the Fig 1.1.4. Once you select the "blank workbook" option a blank screen will open with all the options to start working. This is shown in the Fig 1.1.5

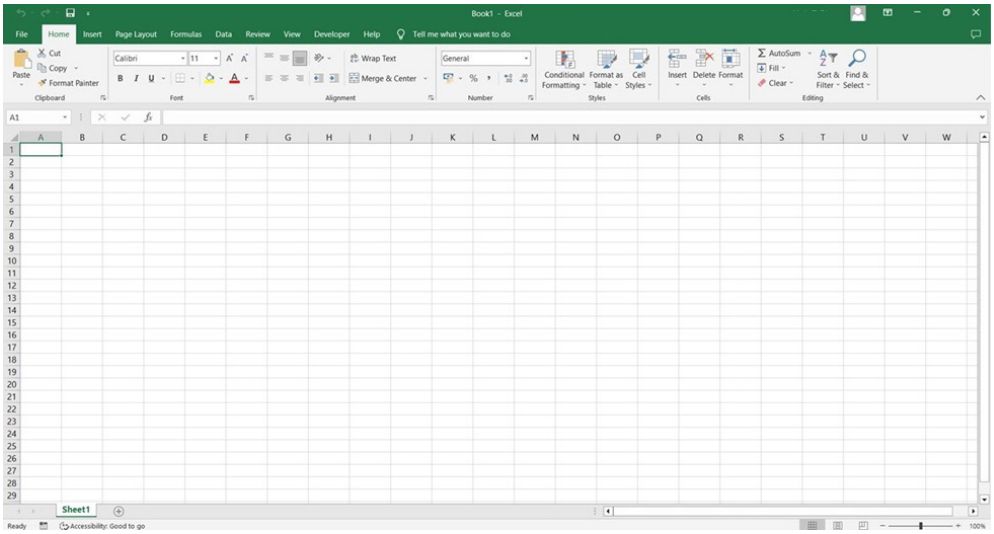


Fig 1.1.5 Blank Workbook

If you want to create a new workbook apart from the one you have already opened, click on the File tab, and then click on the 'New' in the left column and then select 'Blank Workbook" (Refer Fig 1.1.4).

Worksheets are often given default names like "Sheet1," "Sheet2," etc. There are 1,048,576 rows and 16,384 columns in a spreadsheet, which equals to 17,179,869,184 cells. If you already have an existing workbook open in Excel, you have a couple of options to create a new worksheet such as :

1. Right-click on the existing worksheet tab (e.g., "Sheet1") at the bottom of the window.
   - In the context menu that appears, select "Insert" to open a submenu as shown in the Fig 1.1.6
   - Choose "Worksheet" from the submenu to create a new worksheet. Excel will insert the new worksheet to the left of the currently selected worksheet, as shown in the Fig 1.1.7

- A workbook has multiple worksheets

2. Alternatively, you can use the plus (+) button located at the end of the worksheet title tabs as shown in the Fig 1.1.6.
   - Click on the plus button to create a new worksheet. Excel will add the new worksheet to the right of the existing worksheets.

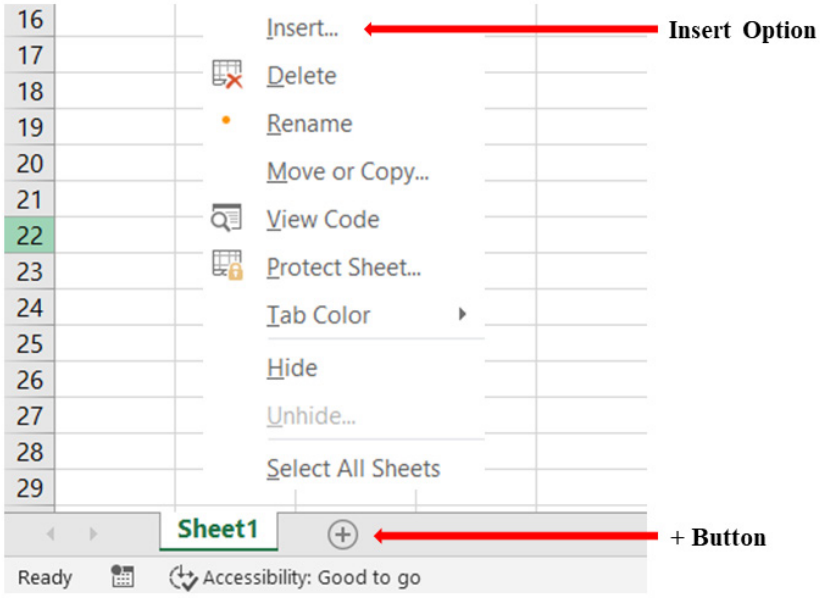


Fig 1.1.6 Creating New Worksheet

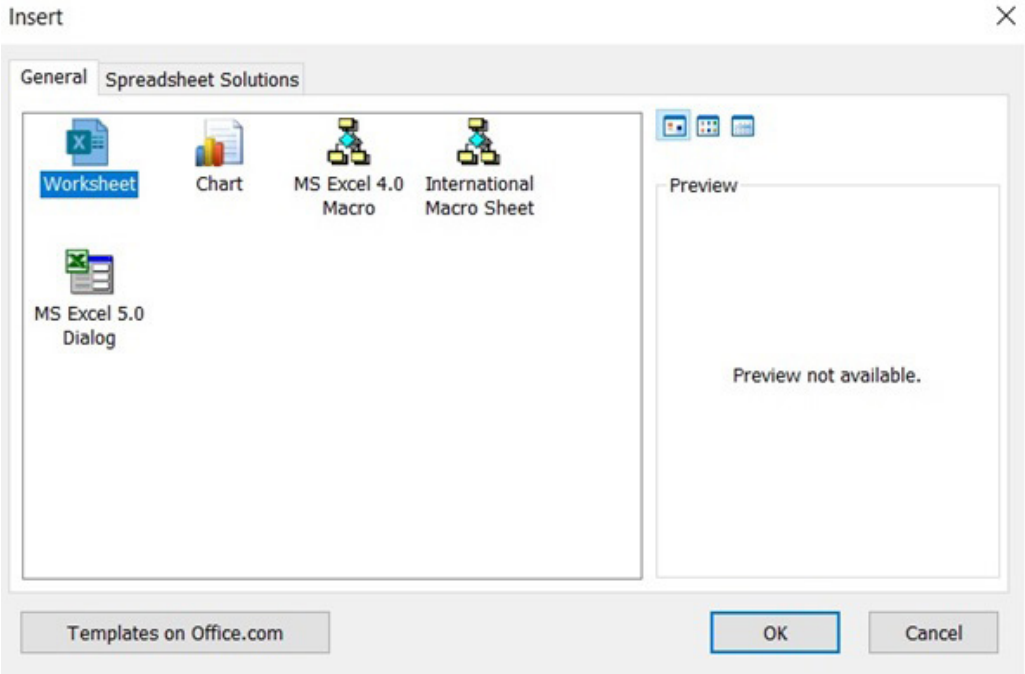


Fig 1.1.7 Inserting Worksheet

1.1.3.2 Entering Data in Excel

Sheet area is the where you type the desired text. The insertion point is the area where text will appear as you type and is represented by the flashing vertical bar. When you click on any cell in the sheet area then that cell will get highlighted. When you double click the cell, the flashing vertical bar appears and then you can start entering your data.

There are various methods to enter data in Excel. They are as follows:

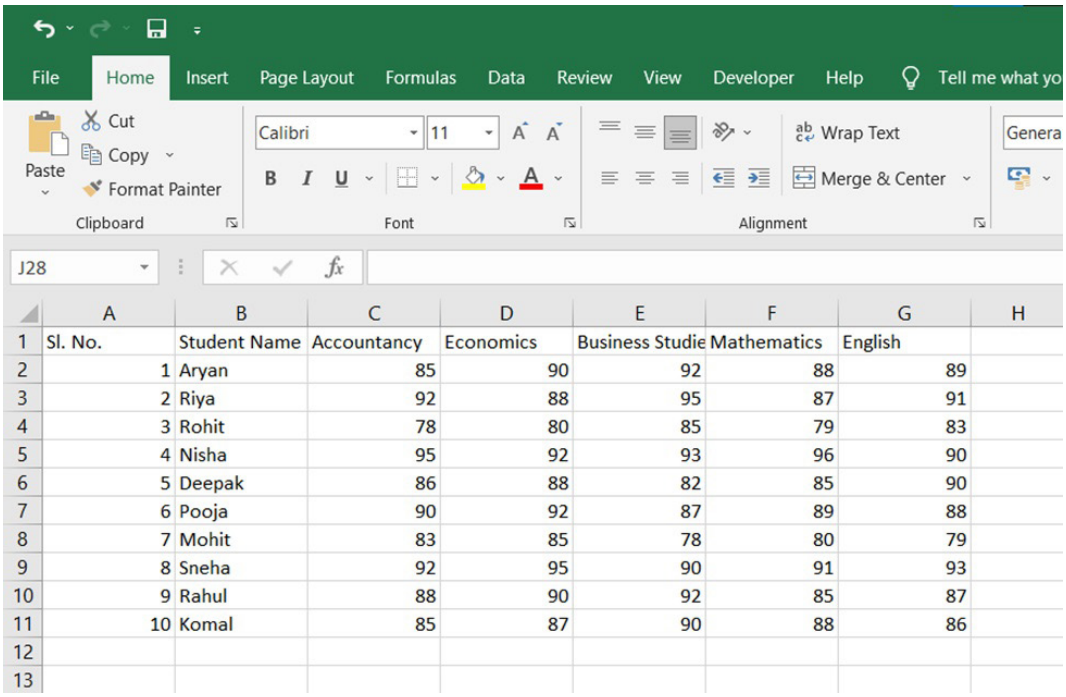
a. **Directly Typing Data into the Cells**

	A	B	C
1	Hello world	Hi 123	
2	1	123	
3			
4			
5			
6			
7			
8			

Fig 1.1.8 Typing Data

The most straightforward method is to directly type the data into the desired cell. Click on the cell where you want to enter the data, and start typing. Press Tab to move to the next cell. Here in the Fig 1.1.8, we have typed in text, numbers, and combination of both.

b. Copying and Pasting from Other Sources



	A	B	C	D	E	F	G	H
1	Sl. No.	Student Name	Accountancy	Economics	Business Studie	Mathematics	English	
2	1	Aryan	85	90	92	88	89	
3	2	Riya	92	88	95	87	91	
4	3	Rohit	78	80	85	79	83	
5	4	Nisha	95	92	93	96	90	
6	5	Deepak	86	88	82	85	90	
7	6	Pooja	90	92	87	89	88	
8	7	Mohit	83	85	78	80	79	
9	8	Sneha	92	95	90	91	93	
10	9	Rahul	88	90	92	85	87	
11	10	Komal	85	87	90	88	86	
12								
13								

Fig 1.1.9 Copy Pasting Data into Excel

If you have data in another source, such as a website, document, or another Excel file, you can copy the data and paste it into Excel. Select the source data, right-click, choose “Copy,” then navigate to the Excel worksheet, click on the destination cell, right-click, and choose “Paste.” Here in the Fig 1.1.9, we have pasted the data from the Table 1.1.1 into the worksheet.

c. Importing Data from External Sources

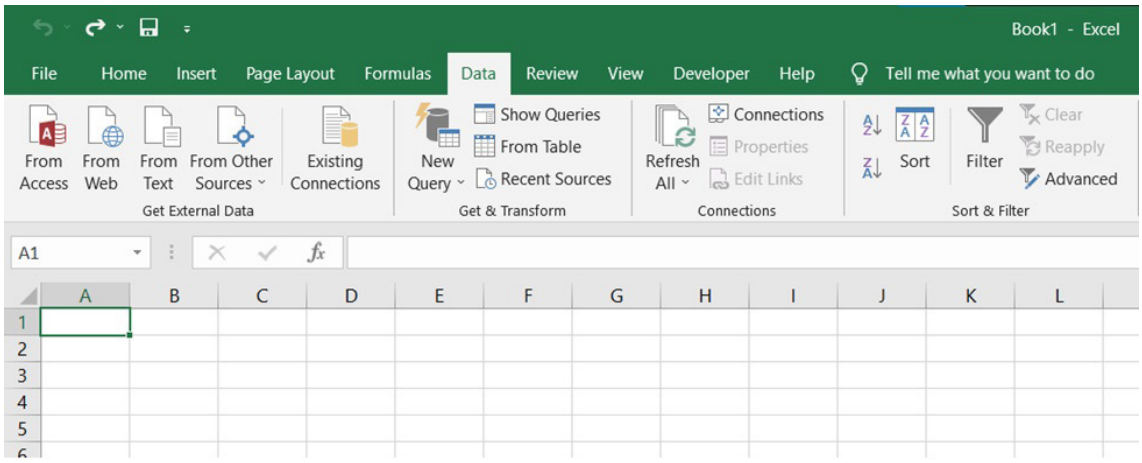


Fig 1.1.10 Data Ribbon for Importing Data

- Entering data - typing into cells, copying, and pasting from or importing data from external sources.

Excel provides various options to import data from external sources like text files, CSV(Comma-Separated Values) files, databases, and other formats. You can use the import features, available under the "Data" tab in the Excel ribbon (as shown in the Fig1.1.10), to import data from external sources.

Entering various kinds of data in Excel involves using different techniques based on the type of data you are working with.

1.1.3.3 Entering Formulas in Excel

Formulas are equations that help with calculations, fetch information, change what is in other cells, check conditions etc. They follow a set of instructions to do calculations using specific values in a certain order. Excel have formulas for quickly finding things like sum, average, count, maximum value, and minimum value for a range of cells. They are versatile tools for handling numbers and data in various ways.

To enter a formula in Excel, follow these steps:

1. Select the cell where you want to enter the formula.
2. Begin the formula by typing the equal sign (=). This means that you are entering a formula.
3. Enter the formula using cell references, operators, and functions. For example, to add the values in cells A1 and B1, you can enter "=A1+B1".
4. Use parentheses or round brackets ( ) to control the order of calculations if needed. For example, "(A1+B1)*C1" will add the values in cells A1 and B1, and then multiply

the result by the value in cell C1.

5. Press Enter to complete the formula. Excel will calculate the result based on the formula you entered and display it in the selected cell.

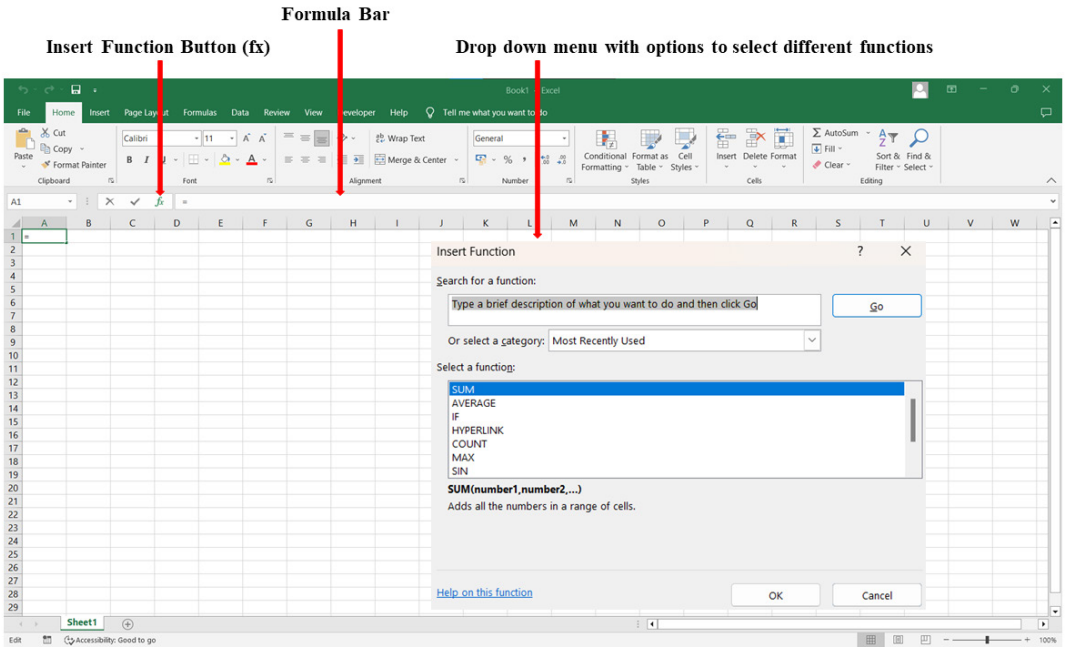


Fig 1.1.11 Entering Formula in Excel

- Use Insert Function button (fx) to insert built-in functions

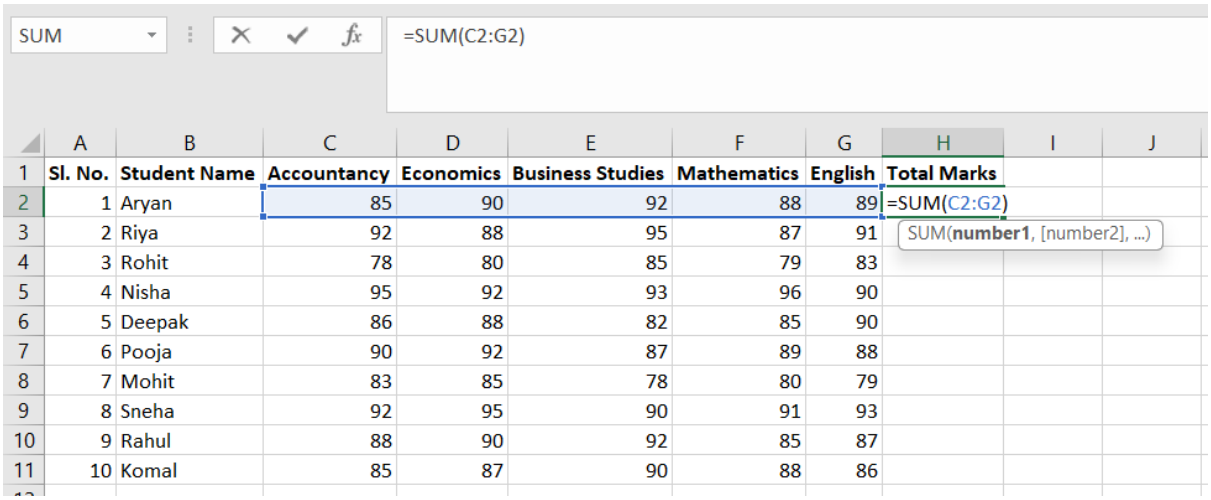
You can also use the Insert Function button (fx) on the formula bar to access a list of built-in functions and easily insert them into your formula (Refer Fig 1.1.11). Simply click on the Insert Function button, select the function you want to use, and follow the prompts to enter the necessary arguments. The detailed explanation and usage of different functions is given in subsequent chapters.

Example:- Now we can try calculating the total marks for the data we have copy pasted to Excel in the Fig 1.1.9. For that we have added another column titled "Total Marks" to the Existing data set. To calculate the total marks let us follow these stes:-

1. Click on the cell where you want the result
2. Type the sum formula i.e., "=SUM"
3. Type open parentheses (
4. Press control key and select the cells for which we need the sum for
5. Type close parentheses)
6. And press Enter

- 'Parentheses' refers to round brackets ( )

The result will be as shown in Fig 1.1.12



	A	B	C	D	E	F	G	H	I	J
1	Sl. No.	Student Name	Accountancy	Economics	Business Studies	Mathematics	English	Total Marks		
2	1	Aryan	85	90	92	88	89	=SUM(C2:G2)		
3	2	Riya	92	88	95	87	91			
4	3	Rohit	78	80	85	79	83			
5	4	Nisha	95	92	93	96	90			
6	5	Deepak	86	88	82	85	90			
7	6	Pooja	90	92	87	89	88			
8	7	Mohit	83	85	78	80	79			
9	8	Sneha	92	95	90	91	93			
10	9	Rahul	88	90	92	85	87			
11	10	Komal	85	87	90	88	86			

Fig 1.1.12 Result of Entering Formula

The same can be done using the fx option in the formula ribbon. We will study about this in detail in coming units. The marks for remaining students can be done in the same manner or by using drag option or autofill options.

1.1.3.4 Saving a Workbook

Saving a workbook in Excel ensures that your data and changes are preserved for future use.

The different methods or ways by which an Excel file can be saved are as follows:

1. Using "Save" Option

a. Click on the "File" tab located in the top-left corner of the Excel window. This will open the Backstage view (or starter screen) as shown in the Fig 1.1.1.

b. In the Backstage view, you will see various options. Choose "Save" or "Save As" (Refer Fig 1.1.13) from the list on the left side. The specific option may vary depending on your version of Excel.

c. Choose the location where you want to save the file.

d. Enter a name for the file in the "File Name" field.

e. Choose the desired format (e.g., Excel Workbook (*.xlsx)).

f. Click the "Save" button.

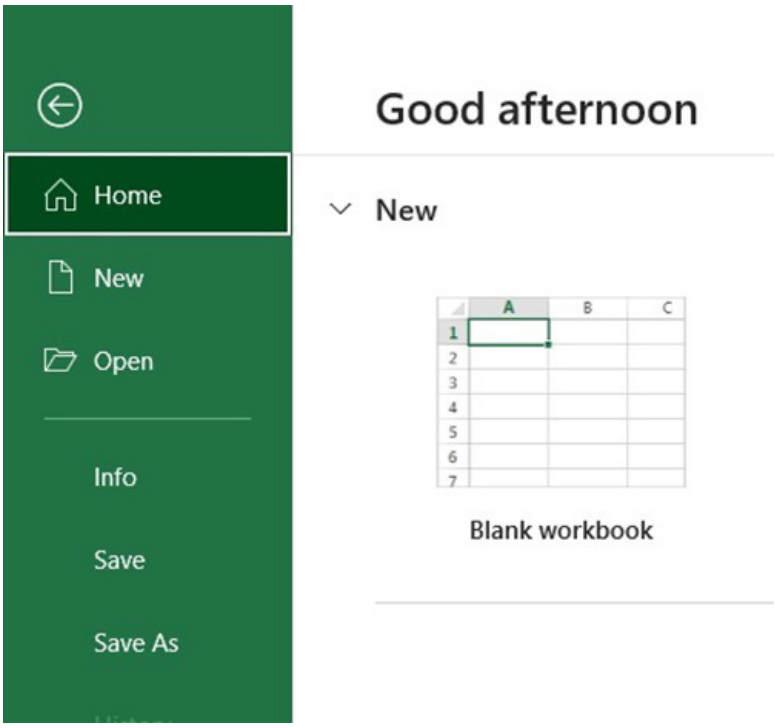


Fig 1.1.13 Saving an Excel File

If the file is already saved once, you can use the Save option to save the recently made changes. Excel will save the workbook using the same name and file location it had previously been saved with.

2. Using " Save As"

If you select "Save As", Excel will open a dialog box where you can specify the name, file location, and format of the saved workbook using " Save As"

a. If you chose "Save As", follow these steps:

b. Choose the desired location where you want to save the workbook. This can be a local folder on your computer or a cloud storage service like OneDrive or Dropbox.

c. In the "File name" field, enter a name for the workbook. You can use alphanumeric characters and spaces, but special characters may not be allowed.

d. Select the appropriate file format from the "Save as type" drop-down menu. Excel offers various formats like .xlsx, .xlsm, .xlsb, .csv, etc. Choose the format that best suits your needs.

e. Optionally, you can choose to save the workbook as a

template by selecting the "Excel Template (*.xltx)" format.

f. Click the "Save" button to save the workbook with the specified name, location, and format.

If you chose "Save", Excel will save the workbook with the same name and file format it had previously been saved with, and the changes will be saved to the existing file.

3. Using Keyboard Shortcuts

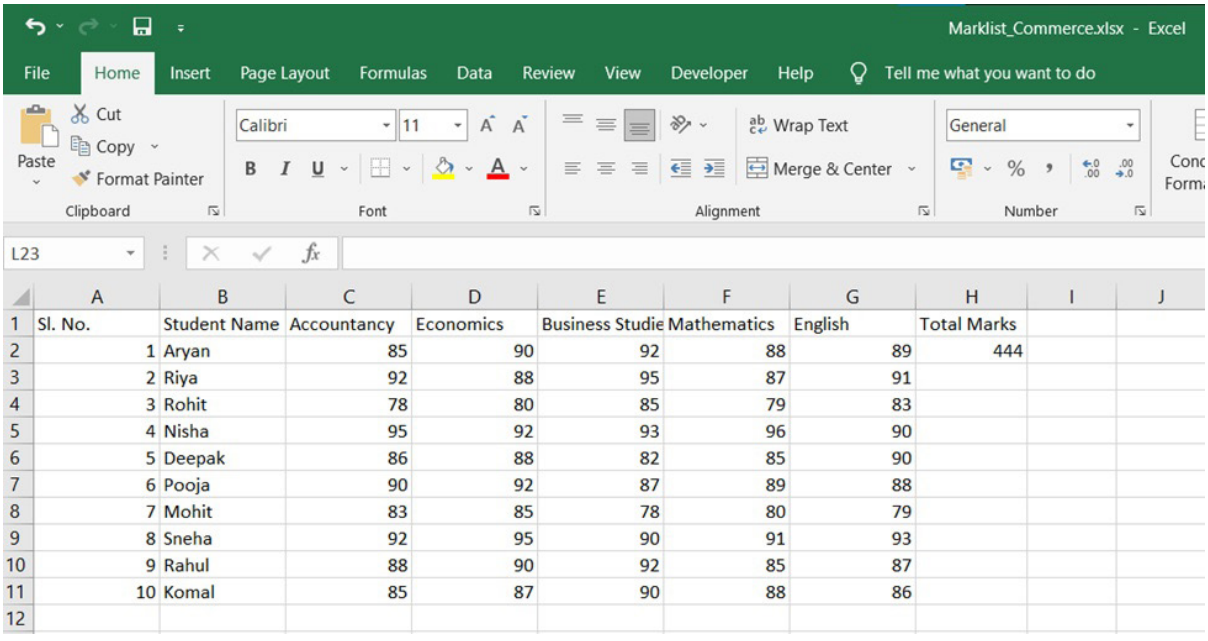
Excel also provides keyboard shortcuts for saving workbooks such as:

- Use keyboard shortcuts like Ctrl + S or F12 to save Excel files

a. To quickly save a workbook, you can use the Ctrl + S keyboard shortcut.

b. To save a workbook with a new name or in a different location, you can use the F12 key to open the "Save As" dialog box directly.

Now let us save the list of students and their mark list with the file name "Marklist_Commerce".



	A	B	C	D	E	F	G	H	I	J
1	Sl. No.	Student Name	Accountancy	Economics	Business Studie	Mathematics	English	Total Marks		
2	1	Aryan	85	90	92	88	89	444		
3	2	Riya	92	88	95	87	91			
4	3	Rohit	78	80	85	79	83			
5	4	Nisha	95	92	93	96	90			
6	5	Deepak	86	88	82	85	90			
7	6	Pooja	90	92	87	89	88			
8	7	Mohit	83	85	78	80	79			
9	8	Sneha	92	95	90	91	93			
10	9	Rahul	88	90	92	85	87			
11	10	Komal	85	87	90	88	86			
12										

Fig 1.1.14 Saving Excel Workbook

After we have saved the file, we can see that the name in the title bar has also changed. Saving files in Excel is a crucial step to ensure that your work is preserved and can be accessed later.

1.1.4 Navigating and Selecting Data in Excel

Navigating and selecting data in Excel refers to the process of moving through the worksheets and selecting specific cells, rows, or columns to work with. You can navigate within a worksheet by using the arrow keys on the keyboard, the scroll bars, or by clicking on the desired cell.

1.1.4.1 Moving within the Worksheet:

- *To move between cells:* You can use the arrow keys on your keyboard to navigate up, down, left, or right within the worksheet. Each press of an arrow key will move the selection to the adjacent cell in that direction. Press Tab key to go to next the column and Press Enter key to go to next row
- *To move to the beginning or end of a row or column:* You can use the Home and End keys on your keyboard. Pressing the Home key will move the selection to the first cell of the row or column, while pressing the End key will move it to the last cell.
- *To move to the edge of the data range:* You can hold down the Ctrl key on your keyboard and press an arrow key. For example, Ctrl+Down Arrow will take you to the last filled cell in the column, Ctrl+Up Arrow will take you to the topmost filled cell, Ctrl+Right Arrow will take you to the last filled cell in the row, and Ctrl+Left Arrow will take you to the leftmost filled cell.
- Press the right Alt + Enter to enter a new line in the same cell.

- Navigate cells using arrow keys, Home/ End keys, Ctrl key with arrow keys and right Alt + Enter keys.

1.1.4.2 Selecting Cells, Rows, and Columns:

- *To select a single cell*: Simply click on the desired cell with the mouse cursor, and it will be selected. The selected cell is indicated by a thick border.
- *To select a range of cells:* Click and hold the mouse button on the starting cell, then drag the cursor to the ending cell of the desired range. Release the mouse button, and all the cells within that range will be selected. The selected cells will be highlighted.
- *To select an entire row or column:* Click on the row number or column letter to select the entire row or column. For example, clicking on the row number "3" will select the entire row 3, and clicking on the column letter "D" will select the entire column D.
- *To select multiple non-adjacent cells or ranges*: Hold

- To select cells, click on a single cell or drag the mouse to select a range

down the Ctrl key on your keyboard while selecting individual cells or ranges. Each selected cell or range will be added to the existing selection.

1.1.4.3 Selecting Data with Keyboard Shortcuts:

- *To select adjacent cells:* Click on the starting cell, hold down the Shift key on your keyboard, and then click on the ending cell. All the cells between the starting and ending cells will be selected, and the entire range will be highlighted.
- *To select entire rows or columns:* Use the Shift key in combination with the arrow keys. For example, to select multiple rows, click on the first row you want to select, hold down the Shift key, and press the Down Arrow key to extend the selection to multiple rows.
- *To select the entire worksheet:* Press the Ctrl + A keys together. This keyboard shortcut selects the entire worksheet, highlighting all cells, rows, and columns.

- Use Shift + click, Shift + arrow keys, and Ctrl + A to select data

1.1.5 Formatting Cells

Formatting cells, rows, and columns in Excel is a important aspect of data presentation and customisation. By applying formatting options, you can enhance the visual appearance of your data, emphasise important information, and make it more readable and professional. An Excel cell can hold different types of data like Numbers, Currency, Dates, etc.

	A	B	C	D	E	F	G	H	I
1	**General**	**Number**	**Currency**	**Date**	**Time**	**Percentage**	**Fraction**	**Scientific**	**Text**
2	abcd1234	1234567890.00	₹ 12,34,56,789.00	01-01-2023	12:00:00	100000%	1/2	1.23E+09	abcd

Fig 1.1.15 Examples of Cell Formats

The various cell formats are as follows;

1. **General:** The default cell format that displays the cell content as it is.
2. **Number:** Formats the cell content as a number with appropriate separators.
3. **Currency:** Formats the cell content as currency, including the currency sign.
4. **Accounting**: Similar to the currency format, used specifically for accounting purposes.
5. **Date**: Formats the cell content as a date, with various date

formats available.

6. **Time**: Formats the cell content as a time, with different time formats to choose from.
7. **Percentage**: Formats the cell content as a percentage with decimal places.
8. **Fraction**: Formats the cell content as a fraction, such as 1/4 or 1/2.
9. **Scientific**: Formats the cell content in scientific notation, such as 5.6E+01.
10. **Text**: Treats the cell content as normal text without applying any formatting.
11. **Special**: Provides special formats for specific data types, such as zip codes or phone numbers.
12. **Custom**: Allows you to create a custom format for the cell content based on your specific requirements.

- Formatting enhance the visual appearance of data

To access the formatting options for cells, rows, and columns in Excel, follow these steps given below:

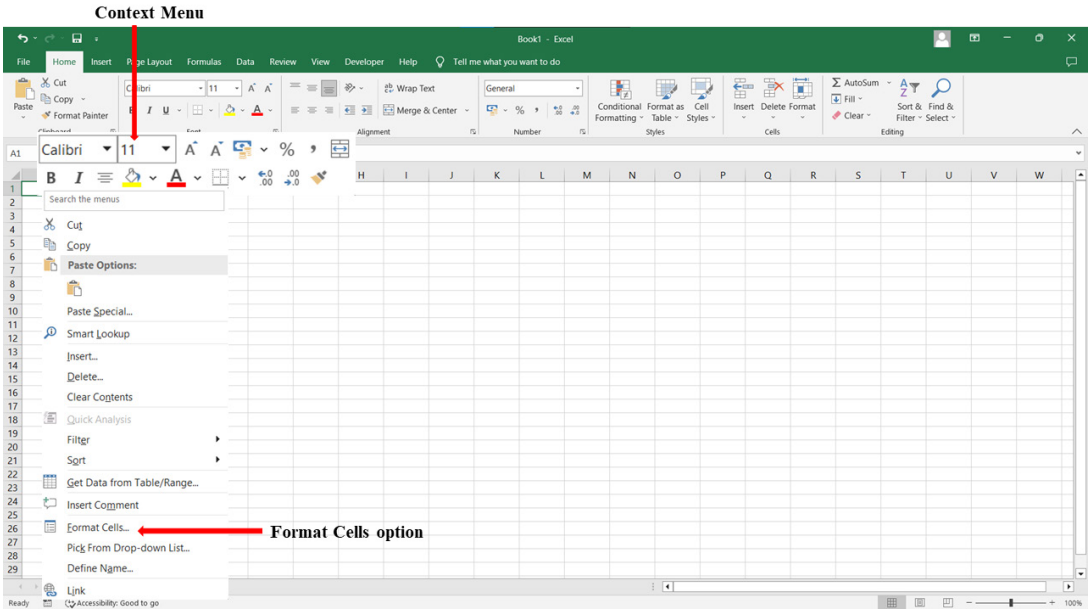


Fig 1.1.16 Formatting Cells

1. Select the cell(s), row(s), or column(s) that you want to format. You can do this by clicking and dragging the mouse cursor over the desired cells, rows, or columns, or by clicking on a cell and then holding down the Shift key while selecting additional cells, rows, or columns.

2. Once the desired cells, rows, or columns are selected, you can access the formatting options through various methods:

    - *Format Cells Dialog Box:* Right-click on the selected cells, rows, or columns and choose "Format Cells" from the context menu (Refer Fig 1.1.16). Alternatively, you can go to the Home tab in the Excel ribbon, click on the "Format" drop-down arrow in the Cells group, and select "Format Cells." (Refer Fig 1.1.17)
    - *Ribbon Options:* In the Home tab of the Excel ribbon, you will find different formatting options in the Cells group. These options include Font, Alignment, Number Format, Cell Styles, and more. Click on the appropriate button for the formatting option you want to apply. (Refer Fig 1.1.17)
    - *Shortcut Menus*: Right-click on the selected cells, rows, or columns to open a shortcut menu. The options in the menu will depend on the type of selection, but you can find formatting options such as Format Cells, Insert, Delete, and more. (Refer Fig 1.1.16)
    - *Keyboard Shortcuts:* Excel also provides keyboard shortcuts for formatting. For example, pressing Ctrl+1 opens the Format Cells dialog box, Ctrl+B applies bold formatting, Ctrl+I apply italic formatting, and Ctrl+U underlines the text.

- Select the desired range and access formatting options

Once you have accessed the formatting options, you can modify various aspects of the selected cells, rows, or columns, such as font style, font size, font colour, cell borders, cell background colour, number formatting, alignment, and more. Make the desired changes in the formatting options dialog box or ribbon, and then click OK or apply the formatting button to confirm the changes.

By using these methods, you can easily access and apply formatting options to cells, rows, and columns in Excel, allowing you to customise the appearance and style of your data.

Excel provides a range of formatting tools that allow you to customise the appearance of your data. These formatting options help you improve readability, emphasise important information, and create a visually organised worksheet.

The key formatting tools available in Excel are:

1. **Font Formatting:** Excel allows you to modify the font type, size, colour, and style of the text in your cells. You can choose from a variety of fonts to suit your preferences and apply formatting options like bold, italic, underline,

and strikethrough. Adjusting the font formatting helps in highlighting specific data.

2. **Cell Formatting:** In addition to font formatting, Excel provides options to format the cells themselves. You can adjust the alignment of cell contents (left, center, right) to improve readability and presentation. Excel also allows you to apply different formats to cells, such as number, currency, percentage, date, or custom formats. This ensures that the data is displayed in a consistent and appropriate manner.

3. **Borders and Gridlines**: Excel enables you to add borders around cells or ranges. Borders can be customised with different line styles, colours, and thickness to create visual separation and structure within your worksheet. Additionally, gridlines help in displaying the boundaries of cells, making it easier to navigate and read the data.

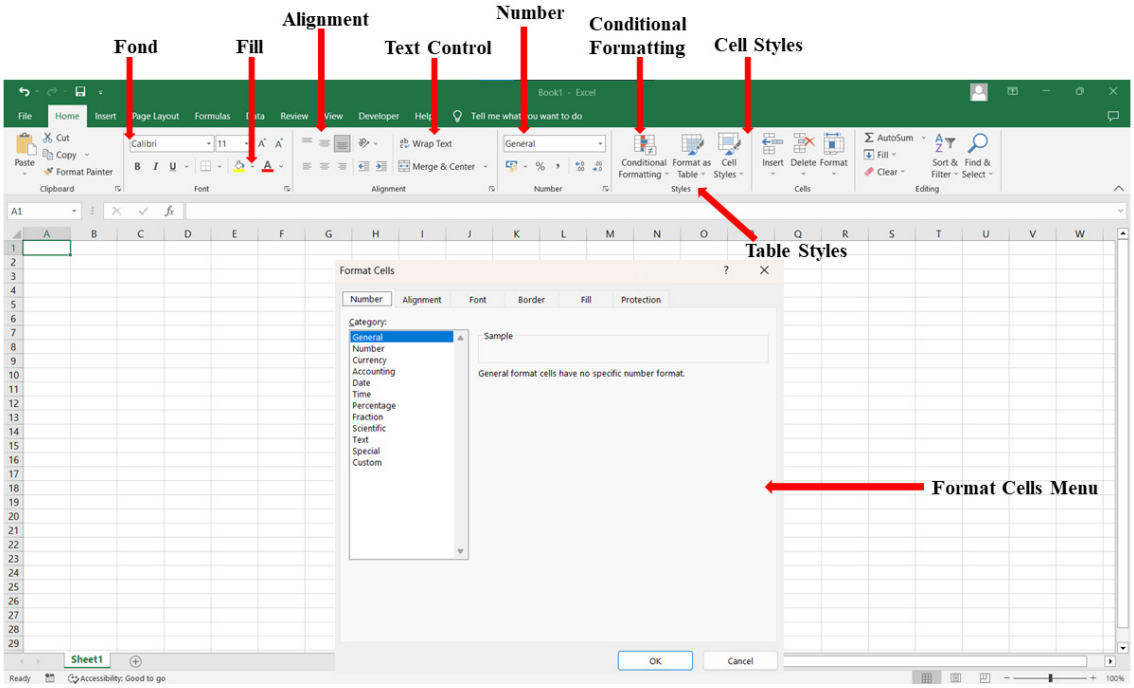


Fig 1.1.17 Options in Formatting of Cells

4. **Fill Colours:** Excel provides a wide range of fill colours that can be applied to cells or cell ranges. You can fill cells with solid colours, gradients, or patterns to visually distinguish specific data or highlight important information. Using fill colours helps in organ-

ising and categorising data effectively.

5. **Conditional Formatting**: One of the powerful features in Excel is conditional formatting. This tool allows you to dynamically format cells based on specific conditions or rules. You can highlight cells based on data values, apply colour scales, data bars, or icon sets to visualise patterns or trends in your data. Conditional formatting helps in drawing attention to significant data points and making it easier to interpret the information.

- Various formatting options are available

6. **Number Formatting:** Excel offers various options to format numbers. You can specify the number of decimal places, choose thousand separators, apply currency symbols, or display percentages. Number formatting ensures that numeric data is presented in a desired and consistent format across your worksheet.
7. **Cell Styles**: Excel provides predefined cell styles that combine font formatting, number formatting, borders, and background colours. Cell styles offer a quick way to apply consistent formatting across your worksheet. You can choose from various built-in styles or create custom styles to suit your specific requirements.
8. **Column Width and Row Height**: Excel allows you to adjust the width of columns and the height of rows. You can resise columns and rows to accommodate the content within them and improve the visual layout of your data. Adjusting column widths and row heights ensures that the data is displayed clearly and neatly.

- Enables customisation, readability, and visual appeal in presenting and organising data.

9. **Merging Cells**: Excel allows you to merge multiple cells together to create a single larger cell. This is useful for creating headings or labels that span across multiple columns or rows. Merging cells helps in improving the organisation and structure of your worksheet.
10. **Themes:** Excel provides built-in themes that allow you to change the overall look and feel of your workbook. Themes include predefined combinations of fonts, colours, and effects that you can apply with a single click. Using themes helps in creating a consistent and visually appealing presentation of your data.

1.1.6 Basic Formatting Options

The "Home" tab in Microsoft Excel offers a set of basic formatting options to help you organise and enhance the visual presentation of your data. It is as shown in the Fig 1.1.17. Let us explore some of the key formatting options available in the "Home" tab:

1.1.6.1 Font Group

The Font Group in the Home tab of Excel is where you can make your text stand out. They are as follows:

1. **Font Style:**
   - **Bold, Italic, Underline:** Easily emphasise or stylise text using these options.
   - **Strikethrough**: Adds a horizontal line through the middle of selected text.
2. **Font Size:**
   - **Increase Font Size, Decrease Font Size**: Quickly adjust the size of the font in your cells.
   - **Font Colour:** Change the colour of the text to emphasise or differentiate information.
3. **Cell Fill:**
   - **Fill Colour:** Modify the background colour of cells for better visibility and organisation.

- Basic formatting options

1.1.6.2 Alignment Group

The Alignment Group is all about how your text is positioned within a cell. They are as given below:

1. **Horizontal Alignment:**
   - **Left, Center, Right:** Align text or numbers horizontally within cells.
   - **General**: Default alignment based on data type.
2. **Vertical Alignment:**
   - **Top, Middle, Bottom**: Align content vertically within cells.
3. **Text Control:**
   - **Wrap Text:** Allow text to wrap within a cell for better readability.
   - **Shrink to Fit:** Automatically reduce font size to fit cell content.

4. **Orientation:**

   - **Angle Clockwise/Angle Counterclockwise:** Rotate text within cells.
   - **Vertical Text:** Display text vertically.

5. **Indentation:**

   - **Increase Indent/Decrease Indent:** Adjust the indentation of cell content.

6. **Merge and Center:**

   - **Merge & Center:** Combine selected cells into a single, centered cell.

1.1.6.3 Editing Group

The Editing Group provides tools for modifying and managing your data. They are as follows:

1. **Fill:**
   - **Fill Colour:** Apply background colour to selected cells.
   - **Clear:** Remove cell content, formatting, or both.

2. **Clear:**

   - **Clear All**: Remove all cell content, formatting, and comments.
   - **Clear Formats:** Remove cell formatting.

3. **Sort & Filter:**

   - **Sort Ascending/Sort Descending**: Organise data in ascending or descending order.
   - **Filter:** Apply or remove filters to facilitate data analysis.

1.1.6.4 Styles Group

The Styles Group lets you add a professional touch to your spreadsheet. They are as given below:

1. **Cell Styles:**
   - Apply predefined styles to cells, such as titles, headings, and data.

2. **Conditional Formatting:**

   - Apply formatting based on specified conditions to highlight trends or outliers in your data.

1.1.6.5 Cells Group

The Cells Group offers various tools for managing the overall structure of your worksheet. They are as follows:

1. **Format:**
   - **Format Cells**: Access advanced formatting options through the Format Cells dialog box.
   - **Format Painter:** Copy formatting from one cell and apply it to another.
2. **Insert/Delete:**
   - **Insert:** Add new cells, rows, or columns.
   - **Delete:** Remove selected cells, rows, or columns.
3. **Format as Table:**
   - Convert a selected data range into a table for improved data management.
4. **Name Manager:**
   - Manage named ranges in the workbook.

1.1.7 Freeze Panes

The "Freeze Panes" feature, is a tool for keeping specific rows and columns visible while scrolling through large sets of data. It is useful when you want to keep headers or labels in view, providing context as you navigate through your spreadsheet. Let us see how to use the option in detail. The "Freeze Panes" feature, located in the "View" menu under the Ribbon as shown in the figure given below. Freeze Panes feature is particularly useful when dealing with extensive datasets where

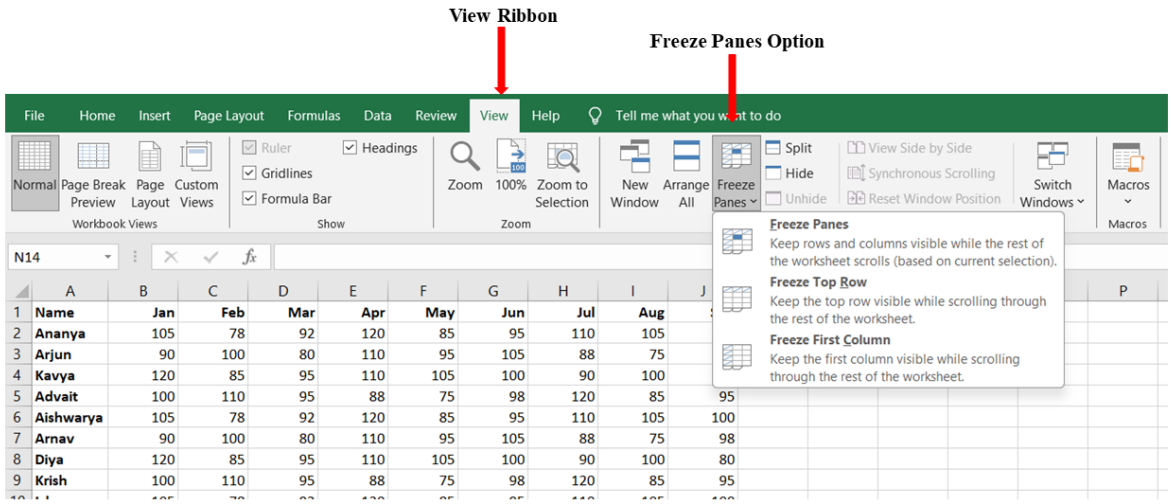


Fig 1.1.18 Freeze Panes

headers or labels need to remain in view for context.

The "Freeze Panes" feature in Excel allows you to lock specific rows or columns, making them always visible while scrolling through large datasets. There are three main options under "Freeze Panes," each serving a different purpose as shown in the figure above. They are:

1. **Freeze Panes:** The "Freeze Panes" option in Excel is a powerful feature that allows you to lock both rows and columns based on the cell you have selected. This means that the rows above and the columns to the left of the chosen cell will remain visible while scrolling through a large dataset. It is particularly useful when you want to keep both row and column headers in view for constant reference.
2. **Freeze Top Row:** "Freeze Top Row" is a specific option under "Freeze Panes" that focuses on locking only the top row of a worksheet. By selecting any cell in the row below the one you want to freeze and choosing this option, you ensure that the top row stays fixed at the top of the worksheet. This is beneficial when dealing with datasets where the header row provides essential information that needs to be visible at all times.

- Freeze Panes enhance the user experience

3. **Freeze First Column:** The "Freeze First Column" option in Excel is designed to lock only the first column of a worksheet. By selecting any cell in the column to the right of the one you want to freeze and choosing this option, you ensure that the first column remains visible on the left side of the worksheet. This is particularly helpful when dealing with datasets where the first column contains labels or identifiers crucial for understanding the data.

These three options, collectively known as "Freeze Panes," enhance the user experience when working with extensive datasets by providing a way to keep specific rows or columns visible while navigating through the data.

The steps to use the freeze pane option is as follows:

1. **Select the Cell Below and to the Right:**
   - Click on the cell below the row(s) you want to freeze.
   - Click on the cell to the right of the column(s) you want to freeze.
2. **Access the "Freeze Panes" Option:**
   - Go to the "View" tab on the ribbon.

3. **Freeze Panes Options:**
   - In the "Window" group, find and click on "Freeze Panes."

You have three options:

- *Freeze Panes:* Freezes both rows and columns based on the cell you selected.
- *Freeze Top Row*: Only the top row remains visible while scrolling.
- *Freeze First Column:* Only the first column remains visible while scrolling.

4. **Unfreeze Panes:**
   - If you want to remove the frozen panes, go back to the "Freeze Panes" option and choose "Unfreeze Panes."

1.1.8 Printing in Excel

Printing in Excel is an important aspect of sharing and presenting data, providing an easily shareable format for your spreadsheet. Whether you need a hard copy or a digital representation in the form of a PDF, understanding the printing process is essential for ensuring that your data is effectively communicated. Excel provides a range of features and settings that allow you to customise the printout, controlling everything from layout and orientation to the inclusion of specific elements like headers, footers, and gridlines.

- Printing involves configuring various settings

Printing in Excel involves configuring various settings to ensure that the data is presented on paper or as a PDF in a way that is both visually appealing and functionally effective. Let us now learn how to print sheet in Excel:

1. **Page Layout:** Before printing, it is essential to consider the page layout. Decide whether you want to print in portrait (vertical) or landscape (horizontal) orientation. You can set this in the "Page Layout" tab.
2. **Size and Margins:** Adjust the page size and margins using the "Size" and "Margins" options in the "Page Layout" tab. This step ensures that your content fits well on the printed page.
3. **Print Titles:** If specific rows or columns need to repeat on each printed page, such as headers, use the "Print Titles" command. Go to the "Page Layout" tab, click on "Print Titles," and set the rows or columns to repeat.
4. **Print Area**: Define a specific print area to include only the selected range in your worksheet. Select the range you

want to print, go to the "Page Layout" tab, and click on "Print Area," then select "Set Print Area."

5. **Print Preview**: Before printing, use the "Print Preview" option to see how your document will look. Click on "File" and then select "Print." The right panel will show a preview of your document, allowing you to make adjustments if needed.
6. **Page Breaks:** Adjust page breaks to control where Excel starts a new page. Go to the "View" tab and select "Page Break Preview." Drag the page break lines to adjust as necessary.
7. **Headers and Footers:** Customise headers and footers to add information such as page numbers, file name, and date. Go to the "Insert" tab and select "Header & Footer." Add elements as needed for your document.
8. **Scaling Options:** Adjust the size of the printout using scaling options. In the "Print" menu, go to "Settings" and choose options like "Fit Sheet on One Page" to optimise the document for printing.
9. **Print Settings:** Configure additional print settings for quality and speed. In the "Print" menu, go to "Settings" and select the desired options, such as print quality and paper type. A shown in the figure given below
10. **Print to PDF:** Save your Excel file as a PDF for easy sharing and viewing. In the "File" menu, choose "Save As," select PDF as the format, and save your document.
11. **Print:** Once you have configured your settings, click on the "Print" button. Choose the printer or print to PDF, and click "Print."

- Following these steps, ensure that printouts are both visually appealing and convey the necessary information

By following these steps, you can optimise your Excel printouts, ensuring that they are both visually appealing and convey the necessary information effectively.

Once you have decided to print you can follow the given steps to print an Excel worksheet

1. Click on the "File" tab in the top-left corner of the Excel window.
2. Select "Print" from the menu as given in the Fig 1.1.19
3. Adjust print settings such as printer, number of copies,

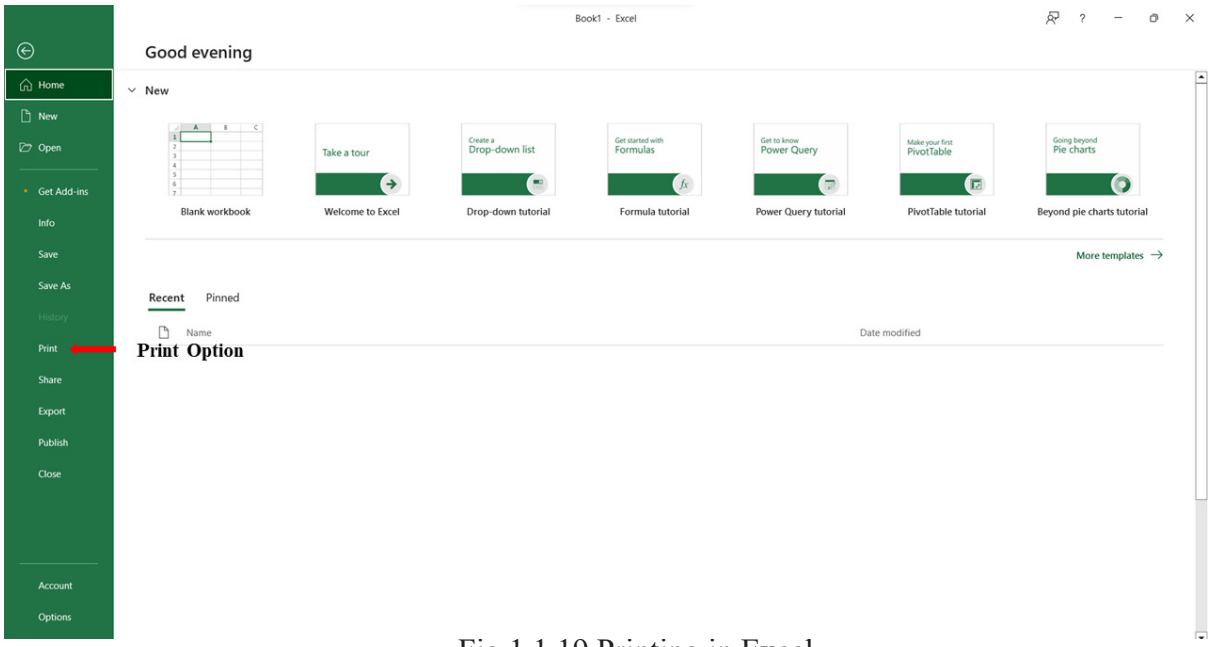


Fig 1.1.19 Printing in Excel

and print range if needed.

4. Preview the print layout on the right side to ensure it looks as expected.
5. Click the "Print" button to send the document to the printer.
6. Make sure your printer is properly set up and connected before attempting to print.

By following these steps, you can print your Excel worksheet, ensuring a clear and accurate representation of your data.

Summarised Overview

Microsoft Excel is a powerful spreadsheet software that supports data creation, editing, and analysis in various arrangements. Excel gained market leadership with the release of Excel 2.0 for Windows in 1987, offering an image-intensive interface and new features. It has since seen regular updates and compatibility improvements. Excel is a valuable tool for economists, providing features for data analysis, financial modelling, forecasting, and economic understanding. The software allows you to create workbooks containing multiple worksheets and cells, where data can be organised, manipulated, and analysed. The Excel interface consists of various components that provide efficient access to file operations, customisation options, and formula input. Creating a new worksheet involves starting a new workbook, and data can be entered directly into cells in the sheet area.

We can navigate within a worksheet using arrow keys, scroll bars, or by clicking on desired cells. They can move between cells using arrow keys, go to the beginning or end of a row or column using the Home and End keys, or move to the edge of the data range by holding Ctrl and pressing an arrow key. You can access formatting options through the Format Cells dialog box, Ribbon options, shortcut menus, or keyboard shortcuts. Excel offers font formatting, cell formatting, borders and gridlines, fill colours, conditional formatting, number formatting, cell styles, column width and row height adjustments, merging cells, and themes to enhance the visual appearance and readability of data.

The Freeze Panes feature is helpful for maintaining visibility of specific rows and columns while navigating through large datasets. This feature is beneficial for keeping headers or labels in view, ensuring a seamless experience when working with extensive datasets. Printing is a vital aspect of data sharing and presentation, offering both physical copies and digital formats like PDFs. This process involves configuring settings for optimal visual appeal and functionality. The steps include considering page layout, adjusting size and margins, defining print areas, and using features like print titles and headers and footers.

Self-Assessment

1. Explain the difference between a workbook and a worksheet in Excel. How can you navigate between multiple worksheets within a workbook?
2. Describe the basic components of the Excel interface. Explain the purpose and functionality of the Ribbon, Formula Bar, and Name Box.
3. Discuss the steps involved in formatting cells, rows, and columns in Excel. Provide examples of formatting options such as changing font styles, adjusting alignment, and applying borders.
4. Explain the concept of cell references in Excel formulas. What are the different types of cell references (relative, absolute, mixed), and how do they impact the formula calculations?
5. What is the purpose of the AutoFill feature in Excel? How can you use it to quickly populate a series of data or formulas in a column or row?

Assignments

1. Using the dataset given below complete the assignment questions given below

Student Name	Age	Marks	Attendance (in %)
Anoop	15	78	80
Bhavana	16	92	95
Chaitanya	14	85	82
Devika	17	93	96
Eshaan	15	79	81
Fathima	16	88	88
Gopika	14	77	80
Harsha	17	90	99
Indrajith	15	81	92
Jyothi	16	95	91

a. Create a new workbook and enter the following data into that workbook.

b. Save the workbook and give the title as ‘Student Details.’

c. Apply basic formatting options such as font styles, font sizes, and font co lours to each column.

d. Save the workbook, close it, and reopen it to ensure that the data and for matting changes are retained.

e. Print the worksheet in PDFformat

Reference

1. Alexander, M., Kusleika, R., & Walkenbach, J. (2018). *Excel 2019 Bible.* John Wiley & Sons.

2. Frye, C. (2018). *Microsoft Excel 2019 step by step*. Microsoft Press.

3. Harvey, G. (2018). *Excel 2019 for dummies.* John Wiley & Sons.

4. Bluttman, K. (2013). *Excel formulas and functions for dummies*. John Wiley & Sons.

Suggested Reading

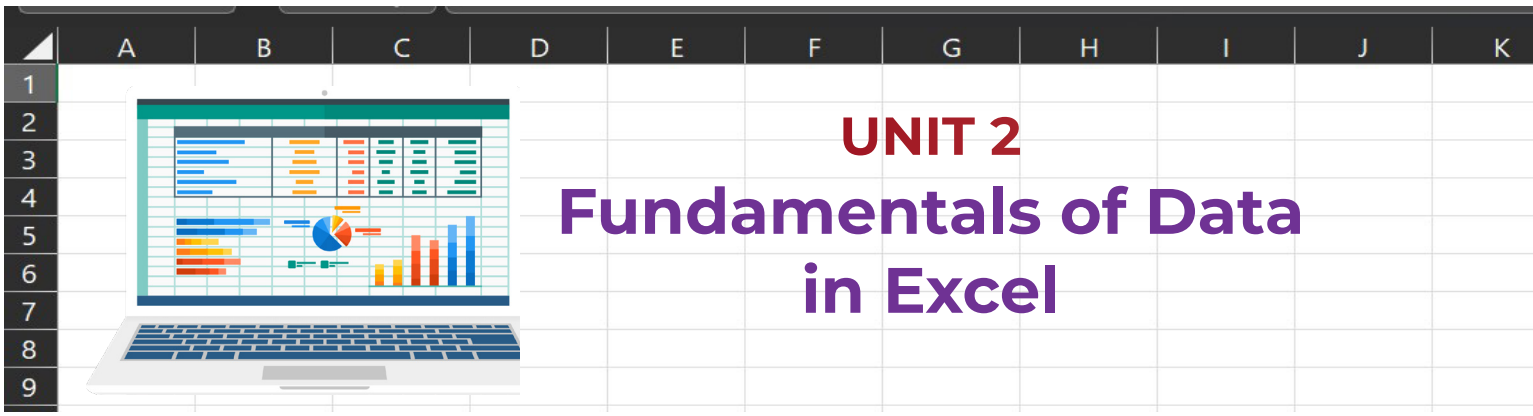
1. McFedries, P., & Harvey, G. (2021). *Excel All-in-one for Dummies.* John Wiley & Sons.

2. Jelen, B., & Juhasz, S. (2015). *MrExcel XL: The 40 Greatest Excel Tips of All Time*. Tickling Keys, Inc.

3. Jelen, B. (2015). *Power Excel with MrExcel: Master Pivot Tables*, Subtotals, Charts, VLOOKUP, IF, *Data Analysis in Excel 2010–2013*. Tickling Keys, Inc.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.



UNIT 2
Fundamentals of Data in Excel

Learning Outcomes

After the completion of this unit, the learner will be able to

- apply sorting techniques to organise and prioritise data
- understand the concept of filtering data
- learn how to set up data validation rules
- gain an understanding of basic mathematical operators and their use in Excel formulas
- apply absolute and relative references to create flexible and scalable formulas

Background

Data validation is a vital feature in Excel that helps maintain data accuracy and integrity. It allows users to control the input that is entered into cells, ensuring that the data adheres to specified criteria or rules. Data validation helps prevent incorrect or invalid data from being entered, reducing errors and maintaining consistency. By setting up data validation rules, users can restrict input to specific formats, numeric ranges, or predefined list values. They can also provide custom error messages to guide users towards accurate and appropriate data entry. Data validation is an essential tool for improving data quality and ensuring reliable analysis.

Understanding and applying data management techniques, such as sorting and data validation, empower users to effectively manage and manipulate their data in Excel. These skills enhance data accuracy, improve productivity, and enable users to derive meaningful insights from their datasets. Excel functionalities are crucial for successful data management and analysis. Mastery of these techniques enables users to optimise their data workflows, make informed decisions, and unlock the full potential of Excel as a data management tool.

Keywords

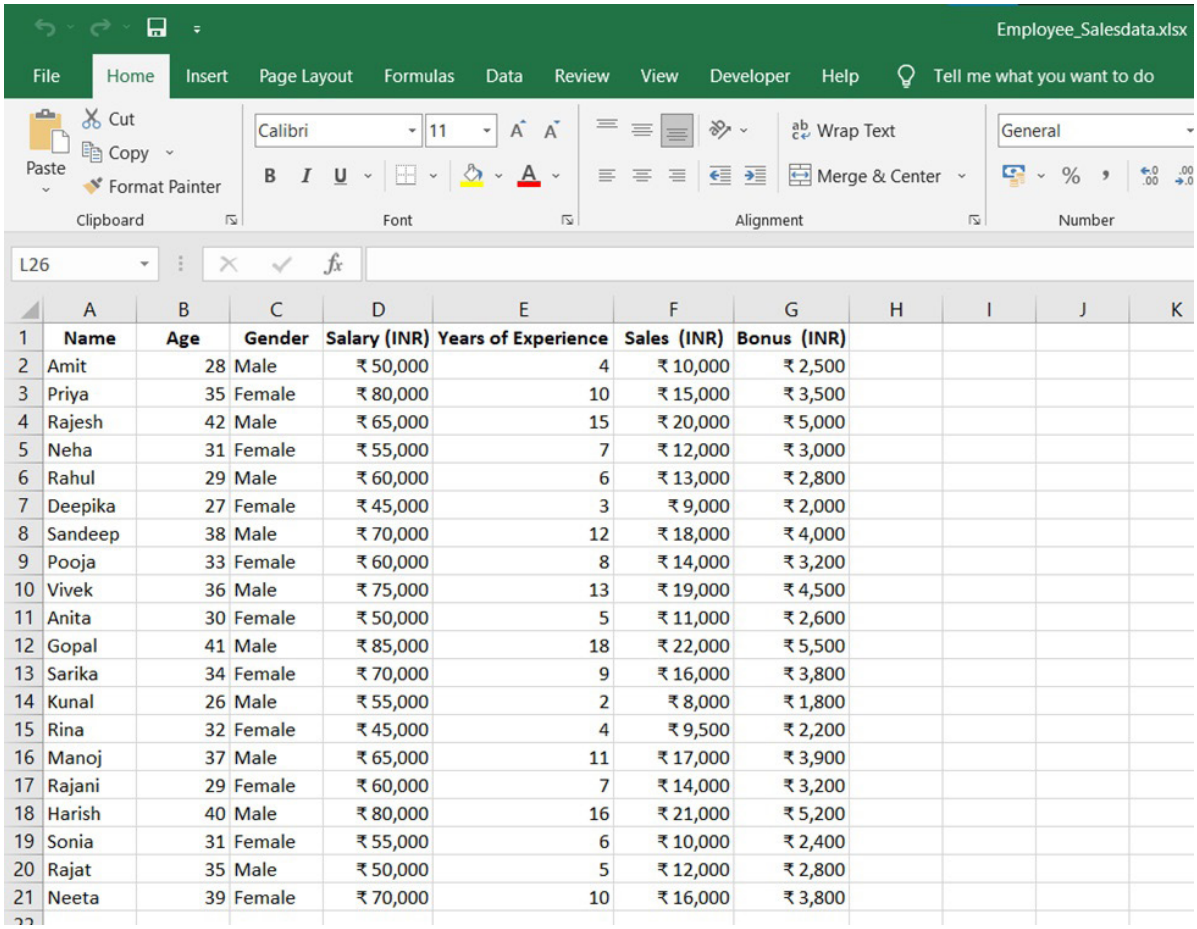
Sorting, Filtering, Data Validation, Functions, Cell References

Discussion

1.2.1 Sorting Data in Excel

Sorting data in Excel refers to the process of rearranging the rows in a selected range of cells based on a specific criteria or column(s). It allows you to organise your data in a more structured and meaningful way, making it easier to analyse and understand.

Let us start working on a new dataset titled "Employee_Sales-data"



	A	B	C	D	E	F	G
1	Name	Age	Gender	Salary (INR)	Years of Experience	Sales (INR)	Bonus (INR)
2	Amit	28	Male	₹ 50,000	4	₹ 10,000	₹ 2,500
3	Priya	35	Female	₹ 80,000	10	₹ 15,000	₹ 3,500
4	Rajesh	42	Male	₹ 65,000	15	₹ 20,000	₹ 5,000
5	Neha	31	Female	₹ 55,000	7	₹ 12,000	₹ 3,000
6	Rahul	29	Male	₹ 60,000	6	₹ 13,000	₹ 2,800
7	Deepika	27	Female	₹ 45,000	3	₹ 9,000	₹ 2,000
8	Sandeep	38	Male	₹ 70,000	12	₹ 18,000	₹ 4,000
9	Pooja	33	Female	₹ 60,000	8	₹ 14,000	₹ 3,200
10	Vivek	36	Male	₹ 75,000	13	₹ 19,000	₹ 4,500
11	Anita	30	Female	₹ 50,000	5	₹ 11,000	₹ 2,600
12	Gopal	41	Male	₹ 85,000	18	₹ 22,000	₹ 5,500
13	Sarika	34	Female	₹ 70,000	9	₹ 16,000	₹ 3,800
14	Kunal	26	Male	₹ 55,000	2	₹ 8,000	₹ 1,800
15	Rina	32	Female	₹ 45,000	4	₹ 9,500	₹ 2,200
16	Manoj	37	Male	₹ 65,000	11	₹ 17,000	₹ 3,900
17	Rajani	29	Female	₹ 60,000	7	₹ 14,000	₹ 3,200
18	Harish	40	Male	₹ 80,000	16	₹ 21,000	₹ 5,200
19	Sonia	31	Female	₹ 55,000	6	₹ 10,000	₹ 2,400
20	Rajat	35	Male	₹ 50,000	5	₹ 12,000	₹ 2,800
21	Neeta	39	Female	₹ 70,000	10	₹ 16,000	₹ 3,800

Fig 1.2.1 Employee Sales Data

- Rearranging rows based on specific criteria or column values

The Fig 1.2.1 displays the dataset titled "Employee Sales Data," containing columns with the following titles: Name, Age, Gender, Salary, Years of Experience, Sales, and Bonus. We will utilise this dataset for further analysis.

When you sort data, Excel compares the values in the specified column(s) or the chosen criteria (such as cell color, font color, or cell icon) and reorders the rows accordingly. The sorting can be done in either ascending (A to Z, smallest to largest) or descending (Z to A, largest to smallest) order.

By sorting data, you can achieve the following benefits:

- **Identify Patterns**: Sorting data helps you identify patterns, trends, or outliers within your dataset. It allows you to group similar values together, making it easier to spot similarities or differences.
- **Organise Information:** Sorting enables you to organise your data in a logical manner. For example, you can sort a list of names alphabetically, sort sales data by date, or arrange items by their numerical values.
- **Facilitate Analysis:** Sorting data makes it more convenient to perform calculations, create charts, or apply other data analysis techniques. It provides a structured format that enhances data analysis capabilities.
- **Filter and Find Data:** Sorting data can work in conjunction with filtering. Once your data is sorted, you can easily apply filters to focus on specific subsets of data or quickly find specific values within the sorted range.

- Enables easy filtering and finding of specific data

Excel offers various sorting options, including sorting by a single column, sorting with multiple levels, sorting by color or icon, and more. These features provide flexibility and customisation when arranging your data according to your specific requirements.

Overall, sorting data is a tool for organising and manipulating data, enabling you to make proper decisions based on data analysis.

To sort data in Excel, you can follow these steps:

1. Select the range of cells that you want to sort. You can click and drag to select multiple cells, or use keyboard shortcuts such as Ctrl+A to select the entire dataset.
2. Once the range is selected, go to the "Data" tab on the Ribbon at the top of the Excel window.

- Organises dataset by rearranging rows based on specific criteria

3. In the "Sort & Filter" group, you will find the "Sort" button. Click on it to open the Sort dialog box.

In the Sort dialog box, you can specify the sorting options:

- *Sort by:* Choose the column or columns you want to sort by. If you have selected a range of cells, Excel will automatically detect the headers and suggest sorting by those columns.
- *Sort on:* Select whether you want to sort based on values, font color, cell color or cell icon.
- *Order:* Choose whether you want to sort in ascending or descending order.
- *Add Level:* If you want to sort by multiple columns, you can add additional sorting levels by clicking on this button.

4. After selecting the desired sorting options, click on the "OK" button to sort the data. Excel will rearrange the selected range of cells based on your specified criteria.

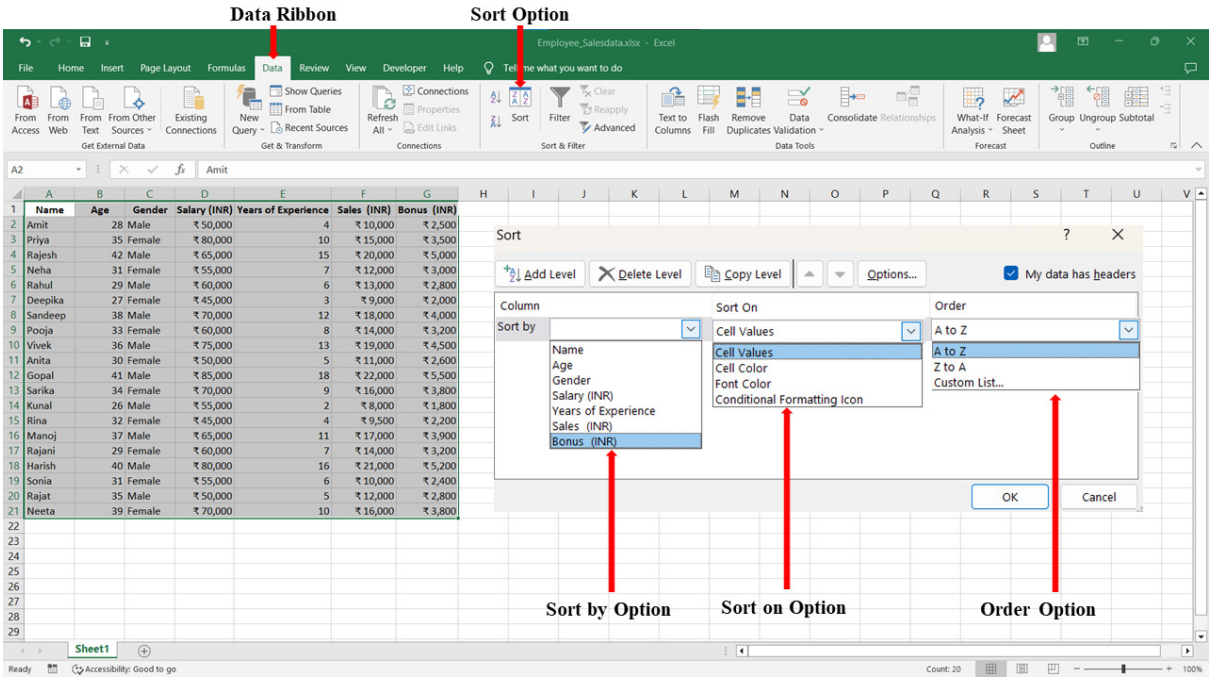


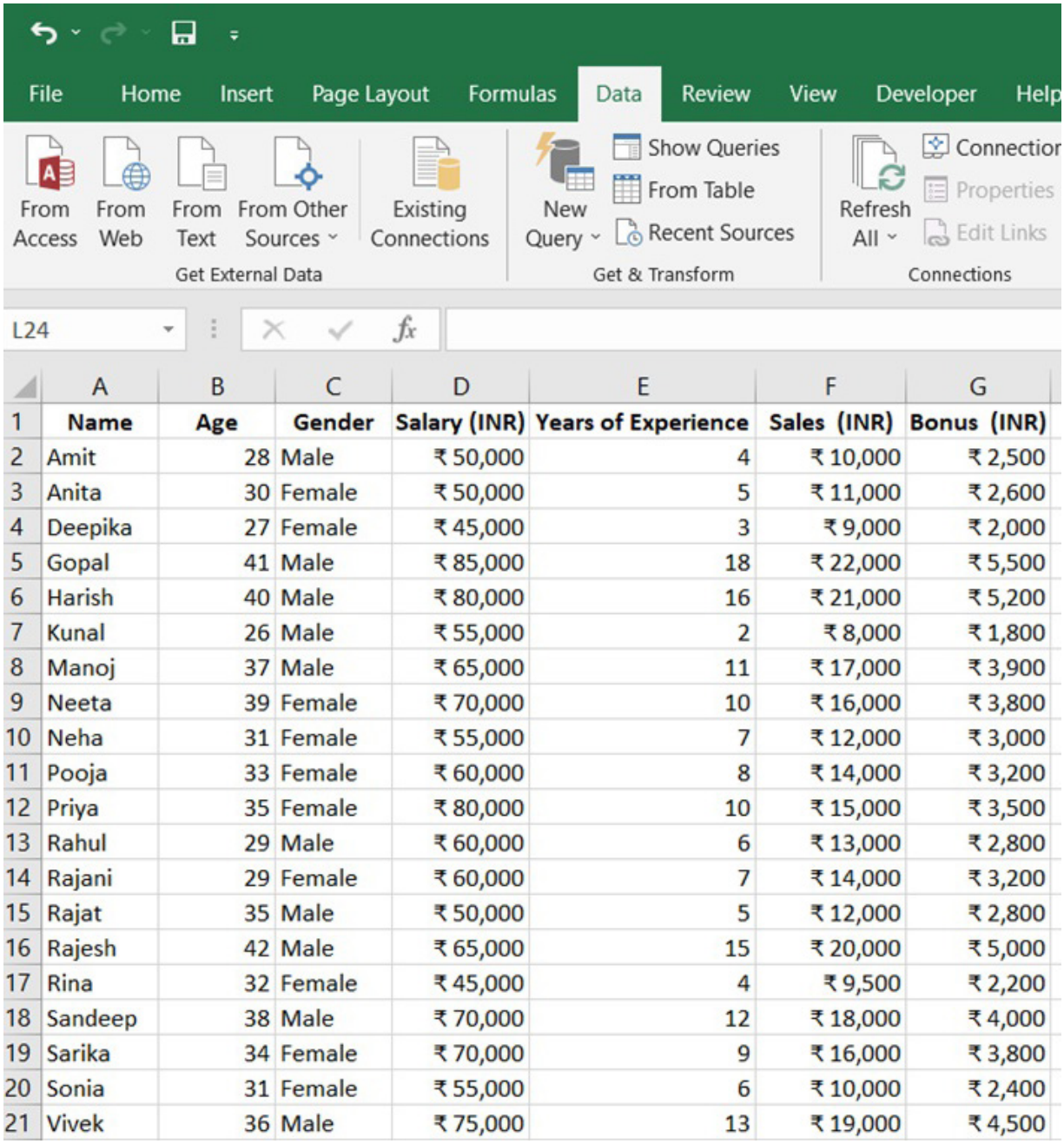
Fig 1.2.2 Name List in Ascending Order

Let us try sorting the "Employee_Salesdata" in Name wise – ascending order

As shown in the Fig 1.2.2, first the data is selected and to sort the whole list in name wise ascending order must follow the following steps;

1. Select the whole data.

2. Choose Sort option from Data Ribbon
3. Select Name from Sort by option
4. Select cell values from Sort on option
5. Since we want the result in ascending order select A to Z in the Order option.
6. Click on OK
7. The result will be as shown in the Fig 1.2.3



	A	B	C	D	E	F	G
1	**Name**	**Age**	**Gender**	**Salary (INR)**	**Years of Experience**	**Sales (INR)**	**Bonus (INR)**
2	Amit	28	Male	₹ 50,000	4	₹ 10,000	₹ 2,500
3	Anita	30	Female	₹ 50,000	5	₹ 11,000	₹ 2,600
4	Deepika	27	Female	₹ 45,000	3	₹ 9,000	₹ 2,000
5	Gopal	41	Male	₹ 85,000	18	₹ 22,000	₹ 5,500
6	Harish	40	Male	₹ 80,000	16	₹ 21,000	₹ 5,200
7	Kunal	26	Male	₹ 55,000	2	₹ 8,000	₹ 1,800
8	Manoj	37	Male	₹ 65,000	11	₹ 17,000	₹ 3,900
9	Neeta	39	Female	₹ 70,000	10	₹ 16,000	₹ 3,800
10	Neha	31	Female	₹ 55,000	7	₹ 12,000	₹ 3,000
11	Pooja	33	Female	₹ 60,000	8	₹ 14,000	₹ 3,200
12	Priya	35	Female	₹ 80,000	10	₹ 15,000	₹ 3,500
13	Rahul	29	Male	₹ 60,000	6	₹ 13,000	₹ 2,800
14	Rajani	29	Female	₹ 60,000	7	₹ 14,000	₹ 3,200
15	Rajat	35	Male	₹ 50,000	5	₹ 12,000	₹ 2,800
16	Rajesh	42	Male	₹ 65,000	15	₹ 20,000	₹ 5,000
17	Rina	32	Female	₹ 45,000	4	₹ 9,500	₹ 2,200
18	Sandeep	38	Male	₹ 70,000	12	₹ 18,000	₹ 4,000
19	Sarika	34	Female	₹ 70,000	9	₹ 16,000	₹ 3,800
20	Sonia	31	Female	₹ 55,000	6	₹ 10,000	₹ 2,400
21	Vivek	36	Male	₹ 75,000	13	₹ 19,000	₹ 4,500

Fig 1.2.3 Name List in Ascending Order - Result

It is important to note that when sorting data, Excel will rearrange the entire rows based on the sorted column(s), so make sure your data is properly selected and all related information in each row stays together.

Additionally, Excel may prompt you to expand the selection if it detects that the selected range does not include all the related data you want to sort. You can choose to expand the selection or sort only the selected range.

Excel provides several types of sorting options to organise data. The main types of sorting available in Excel are:

- **Sort by Values:** This is the most common type of sorting where you sort data based on the values in one or more columns. You can sort in ascending or descending order.
- **Sort by Cell Color:** Excel allows you to sort data based on the background color of cells. You can choose to sort in a specific color order or by the default color scheme.
- **Sort by Font Color:** Similar to sorting by cell color, you can sort data based on the font color of cells. This option is useful when you use different font colors to represent specific categories.
- **Sort by Cell Icon:** If you have applied conditional formatting with icons (such as arrows or symbols) to your data, you can sort based on these icons. You can specify the order of the icons or sort them by the default order.
- **Sort by Cell Format**: This option allows you to sort data based on the cell format, such as number format, date format, or text format.
- **Sort by Column Order:** In this type of sorting, you can sort data based on the order of columns within your dataset. It is useful when you want to restore the original order of rows based on a specific column.
- **Custom Sort:** Custom sorting provides advanced options to sort data based on multiple criteria or sort in a specific order. You can add multiple levels of sorting and specify the sort order for each level.

- Sorting by values, cell color, font color, cell icon, cell format, column order, and custom sorting

These sorting options give you flexibility and control over how your data is arranged. You can choose the most appropriate sorting method based on your data type and sorting requirements.

1.2.2 Filtering in Excel

Filtering allows you to selectively display specific data in a range based on certain criteria. It helps you narrow down a large dataset to focus on the data that meets specific conditions or requirements. Filtering does not change the order of the data; it only hides the rows that do not meet the filter criteria.

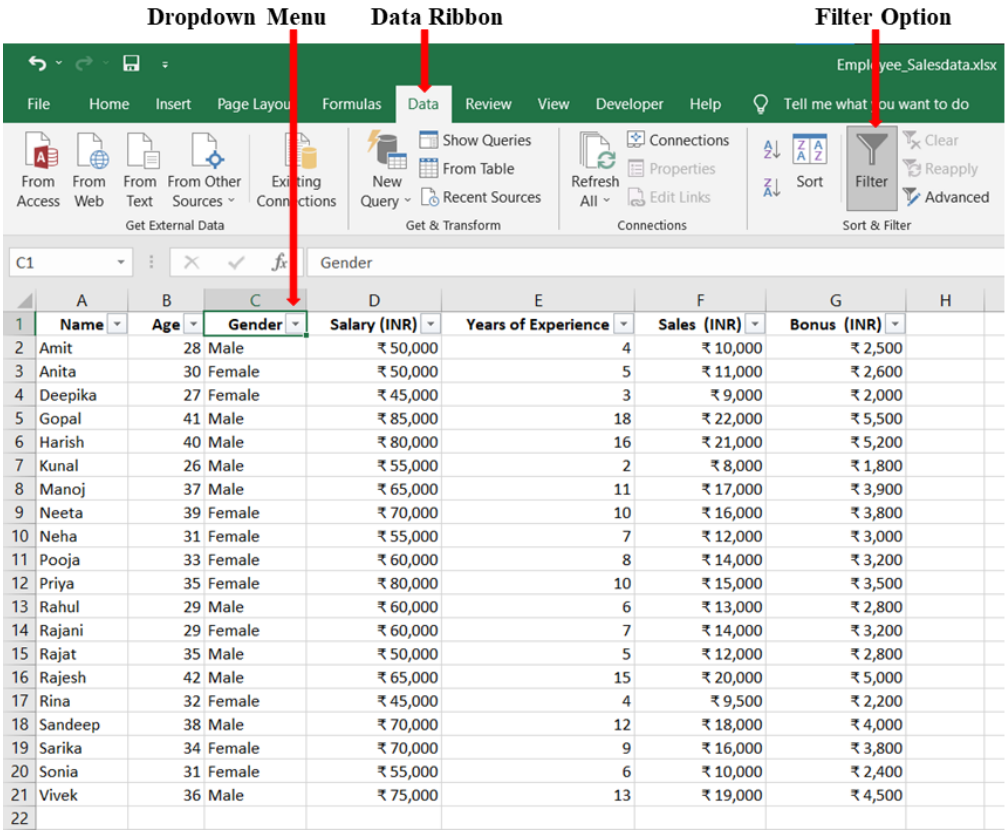
The steps to follow while using filtering in Excel are as given below :

1. Select the range of cells that you want to filter. This range should contain headers in the first row, indicating the column names.
2. Go to the "Data" tab on the Ribbon.
3. In the "Sort & Filter" group, click on the "Filter" button. This will add filter drop-down arrows to each column header in the selected range.
4. Click on the filter drop-down arrow in the column you want to filter. A menu will appear with filtering options specific to the data type in that column.
5. Select the filter criteria you want to apply. You can choose from various options such as text filters (e.g., equals, contains, begins with), number filters (e.g., greater than, less than), date filters, and more.
6. Excel will filter the data based on your chosen criteria, hiding the rows that do not match. You can apply filters to multiple columns simultaneously to further refine your data.
7. To remove the filters and display all the data again, go back to the "Data" tab and click on the "Filter" button to toggle off the filtering.

- Selective display data based on specific criteria

Let us try filtering the Employee_Salesdata based on Gender category. In order to do that follow steps as mentioned below;

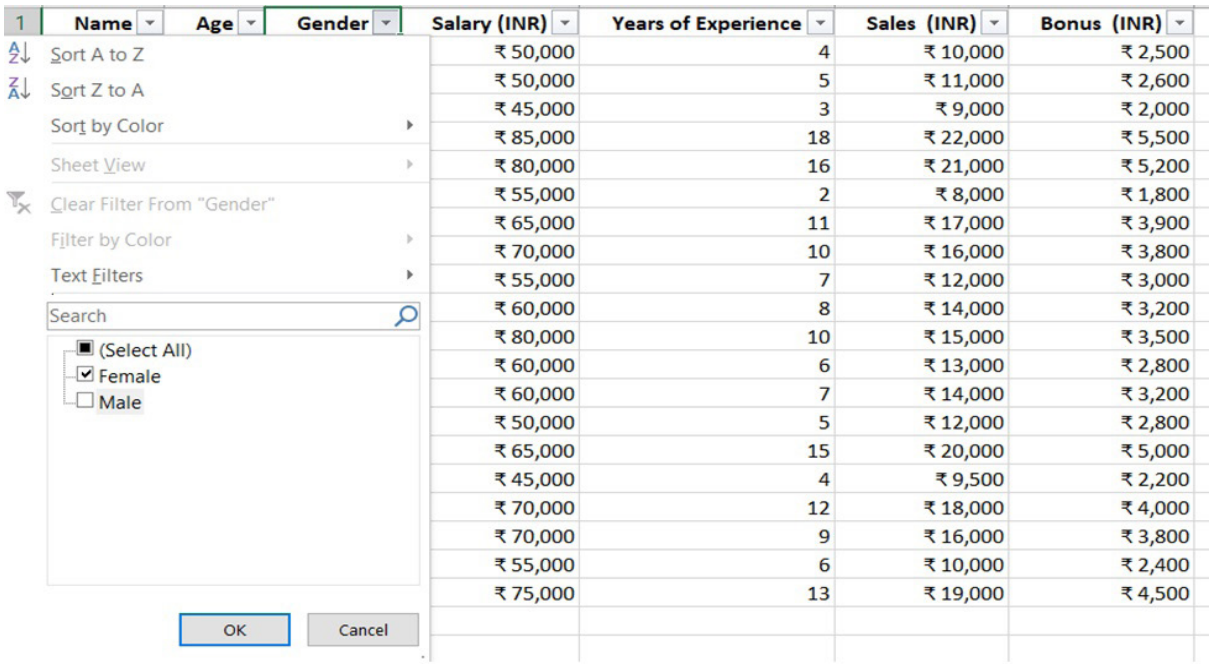
1. Select the Gender column heading as we want to filter based on Gender category.
2. Click on the Data Ribbon and Select Filter option from that
3. There will be a drop-down menu on the column we want to filter. It will be as shown in the Fig 1.2.4
4. When we click on the drop-down new menu will open and deselect the category you want to eliminate.



	A	B	C	D	E	F	G	H
1	Name	Age	Gender	Salary (INR)	Years of Experience	Sales (INR)	Bonus (INR)	
2	Amit	28	Male	₹ 50,000	4	₹ 10,000	₹ 2,500	
3	Anita	30	Female	₹ 50,000	5	₹ 11,000	₹ 2,600	
4	Deepika	27	Female	₹ 45,000	3	₹ 9,000	₹ 2,000	
5	Gopal	41	Male	₹ 85,000	18	₹ 22,000	₹ 5,500	
6	Harish	40	Male	₹ 80,000	16	₹ 21,000	₹ 5,200	
7	Kunal	26	Male	₹ 55,000	2	₹ 8,000	₹ 1,800	
8	Manoj	37	Male	₹ 65,000	11	₹ 17,000	₹ 3,900	
9	Neeta	39	Female	₹ 70,000	10	₹ 16,000	₹ 3,800	
10	Neha	31	Female	₹ 55,000	7	₹ 12,000	₹ 3,000	
11	Pooja	33	Female	₹ 60,000	8	₹ 14,000	₹ 3,200	
12	Priya	35	Female	₹ 80,000	10	₹ 15,000	₹ 3,500	
13	Rahul	29	Male	₹ 60,000	6	₹ 13,000	₹ 2,800	
14	Rajani	29	Female	₹ 60,000	7	₹ 14,000	₹ 3,200	
15	Rajat	35	Male	₹ 50,000	5	₹ 12,000	₹ 2,800	
16	Rajesh	42	Male	₹ 65,000	15	₹ 20,000	₹ 5,000	
17	Rina	32	Female	₹ 45,000	4	₹ 9,500	₹ 2,200	
18	Sandeep	38	Male	₹ 70,000	12	₹ 18,000	₹ 4,000	
19	Sarika	34	Female	₹ 70,000	9	₹ 16,000	₹ 3,800	
20	Sonia	31	Female	₹ 55,000	6	₹ 10,000	₹ 2,400	
21	Vivek	36	Male	₹ 75,000	13	₹ 19,000	₹ 4,500	
22								

Fig 1.2.4 Filtering Data

5. Here we are deselecting Male to display Female Category. It will be as shown in the Fig 1.2.5



1	Name	Age	Gender	Salary (INR)	Years of Experience	Sales (INR)	Bonus (INR)
				₹ 50,000	4	₹ 10,000	₹ 2,500
				₹ 50,000	5	₹ 11,000	₹ 2,600
				₹ 45,000	3	₹ 9,000	₹ 2,000
				₹ 85,000	18	₹ 22,000	₹ 5,500
				₹ 80,000	16	₹ 21,000	₹ 5,200
				₹ 55,000	2	₹ 8,000	₹ 1,800
				₹ 65,000	11	₹ 17,000	₹ 3,900
				₹ 70,000	10	₹ 16,000	₹ 3,800
				₹ 55,000	7	₹ 12,000	₹ 3,000
				₹ 60,000	8	₹ 14,000	₹ 3,200
				₹ 80,000	10	₹ 15,000	₹ 3,500
				₹ 60,000	6	₹ 13,000	₹ 2,800
				₹ 60,000	7	₹ 14,000	₹ 3,200
				₹ 50,000	5	₹ 12,000	₹ 2,800
				₹ 65,000	15	₹ 20,000	₹ 5,000
				₹ 45,000	4	₹ 9,500	₹ 2,200
				₹ 70,000	12	₹ 18,000	₹ 4,000
				₹ 70,000	9	₹ 16,000	₹ 3,800
				₹ 55,000	6	₹ 10,000	₹ 2,400
				₹ 75,000	13	₹ 19,000	₹ 4,500

Fig 1.2.5 Deselecting Categories

6. Then click OK

1	Name	Age	Gender	Salary (INR)	Years of Experience	Sales (INR)	Bonus (INR)
3	Anita	30	Female	₹ 50,000	5	₹ 11,000	₹ 2,600
4	Deepika	27	Female	₹ 45,000	3	₹ 9,000	₹ 2,000
9	Neeta	39	Female	₹ 70,000	10	₹ 16,000	₹ 3,800
10	Neha	31	Female	₹ 55,000	7	₹ 12,000	₹ 3,000
11	Pooja	33	Female	₹ 60,000	8	₹ 14,000	₹ 3,200
12	Priya	35	Female	₹ 80,000	10	₹ 15,000	₹ 3,500
14	Rajani	29	Female	₹ 60,000	7	₹ 14,000	₹ 3,200
17	Rina	32	Female	₹ 45,000	4	₹ 9,500	₹ 2,200
19	Sarika	34	Female	₹ 70,000	9	₹ 16,000	₹ 3,800
20	Sonia	31	Female	₹ 55,000	6	₹ 10,000	₹ 2,400

Fig 1.2.6 Filtered Category

Now the result will be as shown in the Fig 1.2.6

Filtering in Excel is a feature that allows users to display specific subsets of data from a larger dataset based on specified criteria. It helps in focusing on relevant information and analysing data more efficiently.In Excel, there are different types of filtering options available to help you selectively display data based on specific criteria. The main types of filtering in Excel are as follows:

1. **AutoFilter:** AutoFilter is the basic filtering option in Excel. It allows you to filter data by selecting options from drop-down lists that appear in the column headers. You can filter data based on text, numbers, dates, or custom criteria. AutoFilter provides options like Equals, Contains, Begins With, Ends With, Greater Than, Less Than, and more.
2. **Filter by Selection:** This type of filtering allows you to filter data based on a specific cell value. You can select a cell or range of cells that contain the value you want to filter, right-click, and choose "Filter" > "Filter by Selected Cell's Value" from the context menu. Excel will filter the data to show only the rows that match the selected cell value.
3. **Filter by Color:** If you have applied color formatting to cells in your data range, you can use the "Filter by Color" option to filter data based on cell color. You can choose to filter by cell background color, font color, or both.
4. **Advanced Filter**: Advanced Filter provides more advanced filtering options compared to AutoFilter. It allows

you to define complex filtering criteria using formulas. You set up a separate criteria range where you define the filtering conditions based on Excel's formula syntax. Advanced Filter provides greater flexibility and allows you to perform complex logical operations and wildcard matching.

- Various filtering options available in Excel

5. **Filter by Values:** This option allows you to filter data based on specific values in a column. You can select the desired values from a list of unique values that appear when you click on the filter drop-down arrow in the column header.
6. **Top/Bottom Filters:** These filters allow you to filter data based on top or bottom values in a column. You can choose to display the top or bottom percentage, number, or specific number of items.
7. **Date Filters**: Excel provides various options to filter data based on dates. You can filter by specific dates, date ranges, time periods (such as months or years), or relative dates (such as the current month or next week).
8. **Custom Filters**: Custom Filters provide more advanced filtering options by allowing you to create your own custom filtering criteria. You can define conditions based on comparison operators, logical operators (AND, OR), and wildcard characters.

1.2.3 Advanced Filtering in Excel

Advanced Filtering is a more powerful filtering feature that allows you to apply complex filter criteria using formulas. It provides greater control and flexibility in defining the filtering conditions.

The steps to use advanced filtering in Excel are :

1. Create a separate range where you define the filtering conditions. This range should have column headers identical to the original data range.
2. In the "Data" tab, click on the "Advanced" button in the "Sort & Filter" group to open the Advanced Filter dialog box.
3. In the Advanced Filter dialog box, select the following

options:

- Select the “Copy to another location” option
- Specify the original data range in the “List range” field
- Specify the criteria range you set up in the “Criteria range” field
- Specify the destination range where you want the fil tered data to be copied

4. Click on the “OK” button, and Excel will apply the advanced filter based on the criteria you specified.

Let us use our “Employee Salesdata” and find about sales persons who have experience of more than 10 years. For that let us follow the following steps

1. First Copy the column titles to another place
2. Type in the condition- here our condition is under Years of Experience category and the condition is greater than 10. So under that category we typed in >10

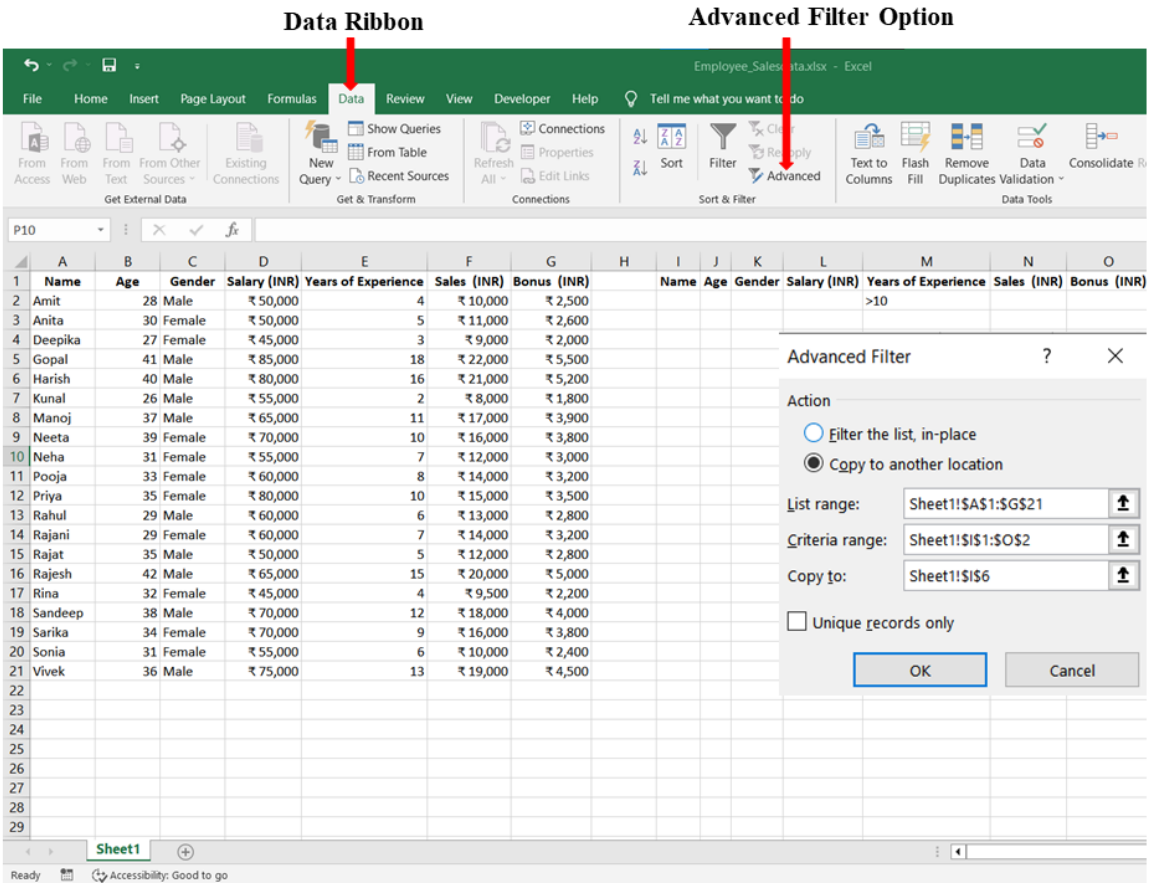


Fig 1.2.7 Advanced Filtering

3. Now Click on advanced option in the Data ribbon tab
4. A new dialogue box will open as shown in the Fig 1.2.7
5. Select Copy to another location in order to select the destination of the result
6. List range is the whole data set. Click on the blank area in list range, click on the arrow button next to the blank area and select the whole dataset
7. Criteria range are the conditions. So, select the copied column headings and the criteria we want for this operation i.e., Years of Experience greater than 10. So, we will select that cell also the same we selected in the above step
8. Copy to is the destination cell. So, we have selected a cell where we want the result
9. The result will be as shown in the Fig 1.2.8

The advanced filtering option will display the details of sales persons with more than 10 years of experience

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Name	Age	Gender	Salary (INR)	Years of Experience	Sales (INR)	Bonus (INR)		Name	Age	Gender	Salary (INR)	Years of Experience	Sales (INR)	Bonus (INR)
2	Amit	28	Male	₹ 50,000	4	₹ 10,000	₹ 2,500						>10		
3	Anita	30	Female	₹ 50,000	5	₹ 11,000	₹ 2,600								
4	Deepika	27	Female	₹ 45,000	3	₹ 9,000	₹ 2,000								
5	Gopal	41	Male	₹ 85,000	18	₹ 22,000	₹ 5,500								
6	Harish	40	Male	₹ 80,000	16	₹ 21,000	₹ 5,200		Name	Age	Gender	Salary (INR)	Years of Experience	Sales (INR)	Bonus (INR)
7	Kunal	26	Male	₹ 55,000	2	₹ 8,000	₹ 1,800		Gopal	41	Male	₹ 85,000	18	₹ 22,000	₹ 5,500
8	Manoj	37	Male	₹ 65,000	11	₹ 17,000	₹ 3,900		Harish	40	Male	₹ 80,000	16	₹ 21,000	₹ 5,200
9	Neeta	39	Female	₹ 70,000	10	₹ 16,000	₹ 3,800		Manoj	37	Male	₹ 65,000	11	₹ 17,000	₹ 3,900
10	Neha	31	Female	₹ 55,000	7	₹ 12,000	₹ 3,000		Rajesh	42	Male	₹ 65,000	15	₹ 20,000	₹ 5,000
11	Pooja	33	Female	₹ 60,000	8	₹ 14,000	₹ 3,200		Sandee	38	Male	₹ 70,000	12	₹ 18,000	₹ 4,000
12	Priya	35	Female	₹ 80,000	10	₹ 15,000	₹ 3,500		Vivek	36	Male	₹ 75,000	13	₹ 19,000	₹ 4,500
13	Rahul	29	Male	₹ 60,000	6	₹ 13,000	₹ 2,800								
14	Rajani	29	Female	₹ 60,000	7	₹ 14,000	₹ 3,200								
15	Rajat	35	Male	₹ 50,000	5	₹ 12,000	₹ 2,800								
16	Rajesh	42	Male	₹ 65,000	15	₹ 20,000	₹ 5,000								
17	Rina	32	Female	₹ 45,000	4	₹ 9,500	₹ 2,200								
18	Sandeep	38	Male	₹ 70,000	12	₹ 18,000	₹ 4,000								
19	Sarika	34	Female	₹ 70,000	9	₹ 16,000	₹ 3,800								
20	Sonia	31	Female	₹ 55,000	6	₹ 10,000	₹ 2,400								
21	Vivek	36	Male	₹ 75,000	13	₹ 19,000	₹ 4,500								
22															
23															

Fig 1.2.8 Advanced Filtering Results

Unlike basic filtering, which applies simple criteria to filter data, advanced filtering provides more advanced options, including multiple criteria, logical conditions, and wildcard characters.

Advanced Filtering allows you to use complex criteria, combine multiple criteria using logical operators (AND, OR), use wildcard characters, and more. It provides a more versatile way to filter data when the standard filtering options are not sufficient. By using filtering and advanced filtering, you can quickly

analyse specific subsets of data, identify patterns, and extract the information you need from large datasets. In Excel, advanced filtering refers to the use of more complex criteria and formulas to filter data.The types of advanced filtering techniques available in Excel are:

- Use "Advanced" option in the "Sort & Filter" group for complex filtering

- **Filter with Formulas:** With advanced filtering, you can filter data using formulas to define complex criteria. This allows you to perform calculations and logical operations to determine which rows should be displayed. You can use various Excel functions, logical operators (AND, OR), comparison operators (=, >, <, etc.), and wildcard characters to create custom filtering conditions.
- **Multiple Criteria Filtering:** Advanced filtering allows you to apply multiple criteria simultaneously. You can specify different conditions for multiple columns using logical operators like AND or OR. This way, you can create more complex rules to filter data based on specific combinations of criteria.
- **Complex Logical Operators:** Advanced filtering provides the ability to use complex logical operators like NOT, XOR, and nested logical functions. These operators allow you to create intricate filtering conditions that go beyond simple AND or OR operations.
- **Wildcard Filtering:** You can use wildcard characters like asterisks (*) or question marks (?) in your advanced filtering criteria. Wildcards help match patterns or parts of text values, giving you greater flexibility in filtering data based on specific patterns or partial matches.
- **Filtering with Named Ranges**: Advanced filtering can also be applied to named ranges. By defining named ranges for your data, you can use those names in your advanced filtering criteria instead of cell references. This makes the filtering process more intuitive and easier to manage, especially for large and complex datasets.

- Filters extracts specific subsets of data from large datasets

- **Filtering Unique Records:** Advanced filtering allows you to filter unique records from a dataset. You can specify criteria to display only the unique values or remove duplicates based on one or more columns. This is particularly useful when you want to extract distinct values from a dataset.
- **Advanced Filter Copy to Another Location**: In addition to filtering data within the same range, advanced filtering also provides the option to copy the filtered results to another location. You can specify a different range to

copy the filtered data, allowing you to separate the filtered data from the original dataset.

By utilising these advanced filtering techniques in Excel, you can apply more sophisticated and customised filters to your data, enabling you to extract specific subsets of information and gain deeper insights from your datasets.

1.2.4 Data Validation in Excel

Data validation allows you to define rules and constraints on the cells where data is entered. It helps control input and ensures that only valid and accurate data is entered into your worksheets. Data validation provides various methods and uses to enforce data integrity and improve data accuracy.

1.2.4.1 Uses of Data Validation:

The main uses of data validation process are as follows:

- Limit data types, set range limits, create drop-down lists, and customise error messages

1. **Limiting Data Types:** It allows you to limit the type of data that can be entered into a cell. For example, you can specify that only whole numbers, decimal numbers, dates, or times are accepted, preventing users from entering invalid or incorrect data types.
2. **Setting Data Range Limits**: You can define minimum and maximum values or ranges for cells. This ensures that entered data falls within specific limits, such as minimum and maximum dates, allowable price ranges, or a specified range of percentage values.
3. **Creating Drop-down Lists**: Data validation enables you to create drop-down lists in cells. This restricts users to selecting values from a predefined list, ensuring consistency, and preventing input errors.
4. **Customising Error Messages**: It enables you to display custom error messages when users enter invalid/inaccurate data. This provides immediate feedback to users, explaining why their input is not valid and helping them correct it.

Data validation in Excel is a powerful tool that helps enforce data integrity, improve accuracy, and provide guidance to users when entering data. It offers various methods to control input and ensures that data conforms to specific rules, resulting in reliable and consistent data analysis and reporting.

1.2.4.2 Methods of Data Validation in Excel

In Excel, there are several methods of data validation that you can use to control input and enforce data integrity. Here are the main methods of data validation available in Excel:

1. **Whole Number:** This method allows you to restrict input to whole numbers only. You can specify the minimum and maximum values that are allowed for the input.
2. **Decimal:** This method restricts input to decimal numbers. You can choose the lowest and maximum values as well as the number of decimal places allowed.
3. **Date:** With this method, you can enforce the input to be in a valid date format. You can specify the allowable date range, minimum, and maximum dates.

- Includes whole numbers, decimals, dates, times, text length, custom formulas, and drop-down lists

4. **Time:** This method validates the input as a valid time value. You can set the time format, minimum and maximum time values, and define the time increments if needed.
5. **Text Length:** This method allows you to limit the length of the entered text. You can specify the maximum number of characters allowed.
6. **Custom:** The custom method provides the most flexibility by allowing you to define your own validation criteria using Excel formulas. You can create custom formulas based on specific conditions and logic to validate the input. This method offers advanced validation options based on your specific requirements.
7. **List:** This method enables you to create a drop-down list in a cell. You can specify a list of valid options, and users can select from the predefined options. This helps ensure consistent and accurate data entry.

These methods of data validation provide a range of options to control input and enforce data integrity in Excel. You can choose the most appropriate method based on the type of data you are working with and the specific validation requirements you have for your worksheet.

1.2.4.3 Using Data Validation to Control Input

The steps to control input using data validation method is as follows:

1. Select the cell or cells where you want to apply data validation.
2. Go to the "Data" tab. As shown in the Fig 1.2.9
3. In the "Data Tools" group, click on the "Data Validation" button. This will open the Data Validation dialog box.
4. In the Settings tab of the Data Validation dialog box, you can specify the validation criteria:
   - Allow: Choose the type of data you want to allow in the selected cell(s). Options include whole numbers, dates, decimal numbers, times, custom formulas, text length, and more.
   - Data: Specify the specific range, values, or formula that the entered data should meet. For example, you can set a range of valid dates, a list of acceptable text values, or a custom formula that validates the input.
   - Input Message: Optionally, you can provide an input message that appears when the user selects the validated cell. The message can provide instructions or additional information about the expected input.
   - Error Alert: You can configure an error message that appears when the user enters invalid data. You can choose the style of the error alert (Stop, Warning, or Information), set the error message text, and specify the error title.
5. Once you have set up the desired data validation criteria, click on the "OK" button to apply the data validation to the selected cell(s).
6. Test the data validation by entering data into the validated cell(s). Excel will validate the input based on the specified criteria. If the data entered does not meet the criteria, an error message will appear. The data validation tool options are as shown in the Fig 1.2.9.

- Displays error messages for invalid input

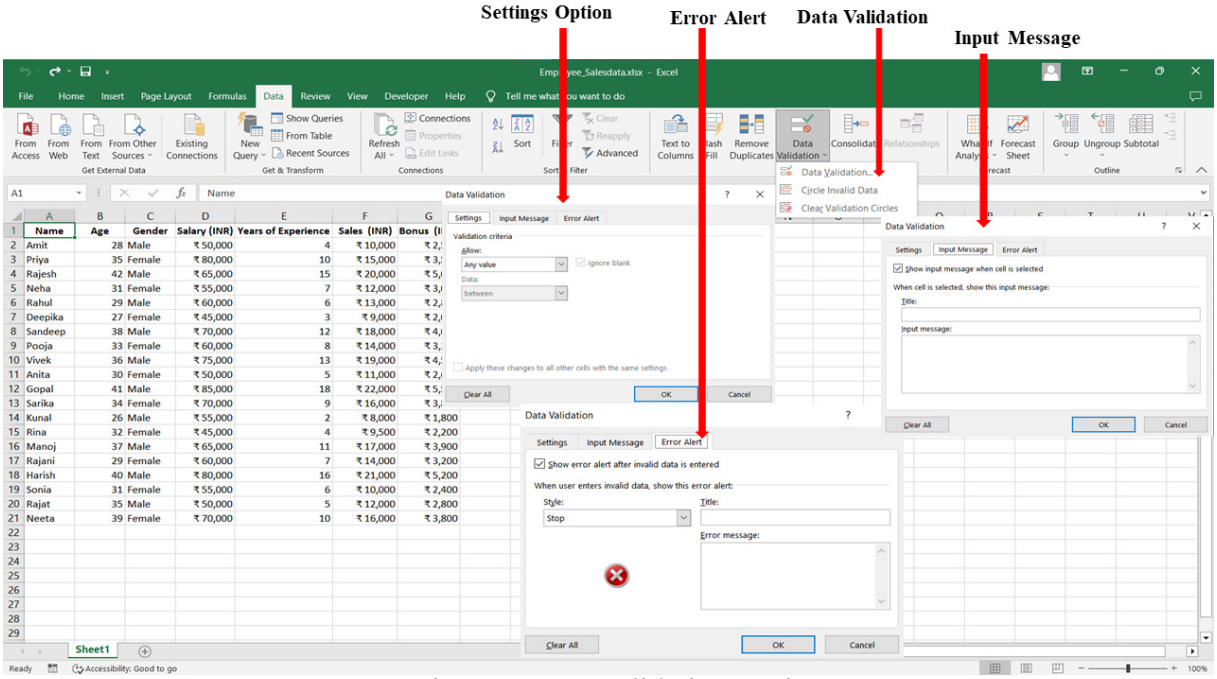


Fig 1.2.9 DataValidation Options

The steps to perform data validation is as follows:

1. Select Data validation from Data ribbon
2. Then click on data validation option
3. A pop-up menu will come and it will have 3 options
4. Select the settings you want to validate
5. Select the input settings and
6. Finally select the error alert you want to display.

By using data validation, you can control the input and ensure that only valid data is entered into your worksheets. This helps maintain data integrity, prevents data entry errors, and enforces specific constraints or rules on the data being entered.

- Rules and constraints on data, ensure data integrity and accuracy

1.2.5 Understanding Basic Formulas and Operators in Excel

In Excel, formulas are expressions that perform calculations, manipulate data, and return results. They are used to automate calculations and perform various operations on the data in your Excel worksheets. Formulas in Excel consist of a combination of values, cell references, functions, and operators. They are written in specific syntax, starting with an equal sign (=), followed by the elements that make up the formula.

The components commonly used in Excel formulas are:

1. **Values:** These are specific numbers, text, or dates that you can directly include in a formula. For example, "=5+3" or "= "Hello World""
2. **Cell References:** Excel allows you to refer to cells by their column and row reference within a formula. Cell references help dynamically include the values stored in those cells in your calculations. For example, "=A1+B1" adds the values in cells A1 and B1.
3. **Functions:** There are several built-in functions in Excel that carry out particular operations or computations. Functions can be used to manipulate data, perform mathematical calculations, analyse data, and more. Examples of functions include SUM, AVERAGE, IF, COUNT, and many others. Functions are typically written with the function name followed by parentheses that may contain additional arguments. For example, "=SUM(A1:A5)" calculates the sum of the values in the range A1 to A5.

- Consists of values, cell references, functions, and operators

4. **Operators:** Operators are symbols or characters that perform specific operations on values or cell references. They include arithmetic operators (+, -, *, /), comparison operators (=, <>, >, <, >=, <=), logical operators (AND, OR, NOT), concatenation operator (&), and more. Operators are used to define the type of operation you want to perform within a formula.

Excel formulas can be as simple as a single value or cell reference, or they can be complex, involving multiple values, cell references, functions, and operators. Formulas can be entered directly into cells, or they can be used in various functions and features throughout Excel to perform calculations, data analysis, conditional formatting, and more.

To use Formulas in Excel, let us follow these steps given below :

1. Select the cell where you want the formula result to display.
2. Start the formula by typing an equal sign (=). This tells Excel that you are entering a formula.
3. Enter all the elements of the formula, which can include cell references, values, functions, and operators. You

can type them directly or use the mouse to select cells or functions from the toolbar.

4. As you enter the formula, Excel provides suggestions and auto-complete options to help you with syntax and function names. You can also refer to the Formula Bar at the top of the Excel window to see and edit the formula.
5. Press Enter key once you have completed the formula. Excel will calculate the formula and display the result in the selected cell.

Operators allow you to perform mathematical calculations, comparisons, concatenations, and logical operations. Here are the main types of operators in Excel:

1. **Arithmetic Operators:**

   - Addition (+): Adds values together. For example, "=A1 + B1" adds the values in cells A1 and B1.
   - Subtraction (-): Subtracts one value from another. For example, "=A1 - B1" subtracts the value in cell B1 from the value in cell A1.
   - Multiplication (*): Multiplies values together. For example, "=A1 * B1" multiplies the values in cells A1 and B1.
   - Division (/): Divides one value by another. For example, "=A1 / B1" divides the value in cell A1 by the value in cell B1.
   - Exponentiation (^): Raises a value to the power of another value. For example, "=A1 ^ B1" raises the value in cell A1 to the power of the value in cell B1.

- Operators specify the type of calculation that we want to perform on elements in a formula

2. **Comparison Operators:**

   - Equal to (=): Checks if two values are equal. For example, "=A1 = B1" returns TRUE if the values in cells A1 and B1 are equal.
   - Not equal to (<>): Checks if two values are not equal. For example, "=A1 <> B1" returns TRUE if the values in cells A1 and B1 are not equal.
   - Greater than (>): Checks if one value is greater than another value. For example, "=A1 > B1" returns TRUE if the value in cell A1 is greater than the value in cell B1.
   - Less than (<): Checks if one value is less than another value. For example, "=A1 < B1" returns TRUE if the value in cell A1 is less than the value in cell B1.

- Greater than or equal to (>=): Checks if one value is greater than or equal to another value.
- Less than or equal to (<=): Checks if one value is less than or equal to another value.

3. **Concatenation Operator:**

- Ampersand (&): Joins two or more text strings together. For example, "=A1 & " " & B1" concatenates the values in cells A1 and B1 with a space in between.

4. **Logical Operators:**

- AND: Checks if all specified conditions are TRUE.
- OR: Checks if at least one of the specified conditions is TRUE.
- NOT: Negates the logical value of a condition.

- Symbols or characters used to perform specific operations on values or cell references within formulas.

These operators, when combined with values, cell references, and functions, allow you to create formulas that perform calculations, comparisons, text manipulations, and logical operations in Excel.

We will learn about the complex functions and operations in coming units. Now let us try some basic functions on our "Employee Salesdata" dataset. Let us try to find the total benefits derived by an employee. Total benefits is the sum of salary and bonus. It is illustrated in the Fig 1.2.10.

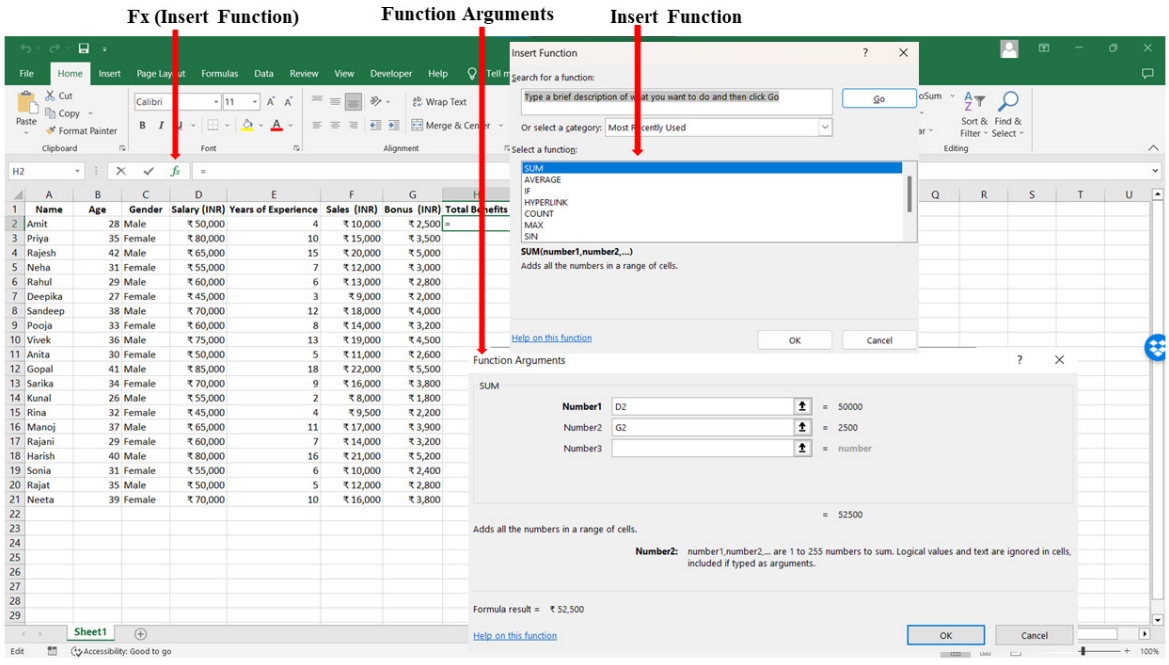


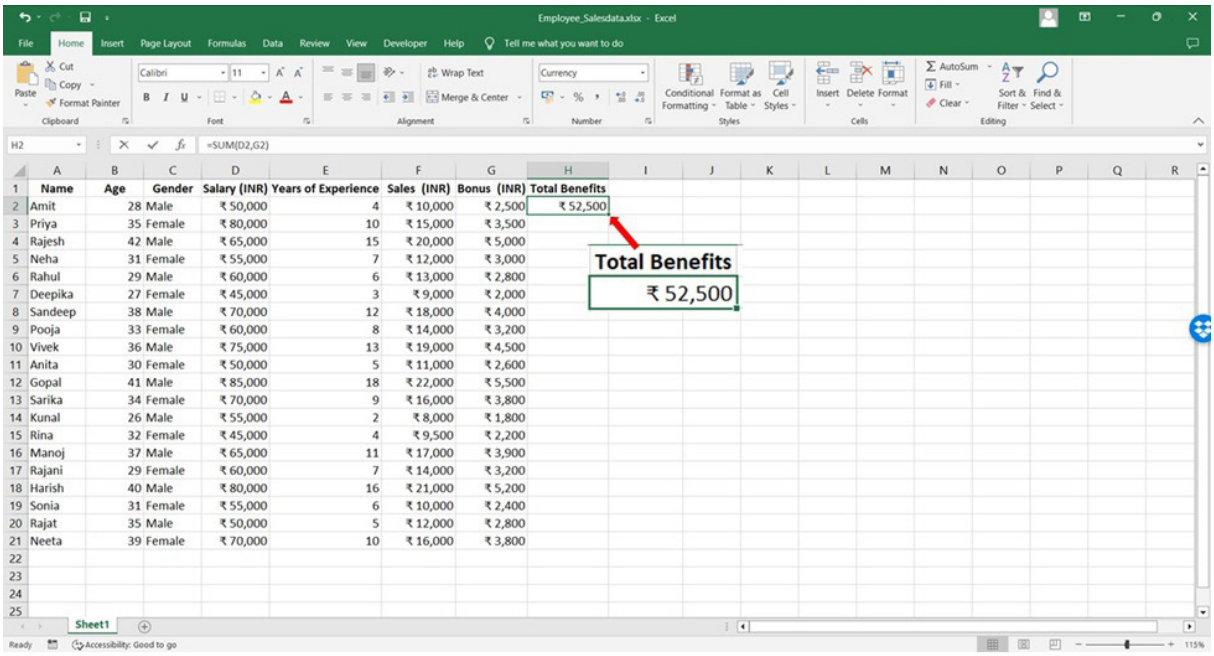
Fig 1.2.10 Using Insert Function

- Perform specific operations on values or cell references within formulas.

The steps to find out the Total Benefits are as follows;

1. Create additional column titled “Total Benefits”
2. Click on the cell where you want the result to be published.
3. Click on Insert Function or fx
4. A dialogue box titled insert function will appear as shown in Fig 1.2.10
5. Select sum from that list
6. Click on the Number 1 option and select the first cell you want to add
7. Click on the Number 2 option and select the second cell to be added
8. You can add more cells if you have any
9. Click on OK

The result will be as shown in Fig 1.2.11



	Name	Age	Gender	Salary (INR)	Years of Experience	Sales (INR)	Bonus (INR)	Total Benefits
2	Amit	28	Male	₹ 50,000	4	₹ 10,000	₹ 2,500	₹ 52,500
3	Priya	35	Female	₹ 80,000	10	₹ 15,000	₹ 3,500	
4	Rajesh	42	Male	₹ 65,000	15	₹ 20,000	₹ 5,000	
5	Neha	31	Female	₹ 55,000	7	₹ 12,000	₹ 3,000	
6	Rahul	29	Male	₹ 60,000	6	₹ 13,000	₹ 2,800	
7	Deepika	27	Female	₹ 45,000	3	₹ 9,000	₹ 2,000	
8	Sandeep	38	Male	₹ 70,000	12	₹ 18,000	₹ 4,000	
9	Pooja	33	Female	₹ 60,000	8	₹ 14,000	₹ 3,200	
10	Vivek	36	Male	₹ 75,000	13	₹ 19,000	₹ 4,500	
11	Anita	30	Female	₹ 50,000	5	₹ 11,000	₹ 2,600	
12	Gopal	41	Male	₹ 85,000	18	₹ 22,000	₹ 5,500	
13	Sarika	34	Female	₹ 70,000	9	₹ 16,000	₹ 3,800	
14	Kunal	26	Male	₹ 55,000	2	₹ 8,000	₹ 1,800	
15	Rina	32	Female	₹ 45,000	4	₹ 9,500	₹ 2,200	
16	Manoj	37	Male	₹ 65,000	11	₹ 17,000	₹ 3,900	
17	Rajani	29	Female	₹ 60,000	7	₹ 14,000	₹ 3,200	
18	Harish	40	Male	₹ 80,000	16	₹ 21,000	₹ 5,200	
19	Sonia	31	Female	₹ 55,000	6	₹ 10,000	₹ 2,400	
20	Rajat	35	Male	₹ 50,000	5	₹ 12,000	₹ 2,800	
21	Neeta	39	Female	₹ 70,000	10	₹ 16,000	₹ 3,800	

Fig 1.2.11 Results of Insert Function

The result of the insert function will be as shown in the Fig 1.2.11. Here we have found the Total Benefits for the first employee. To find the same for other employees we can either use Autofill option or use Drag and fill option as given below.

- Click on the right bottom corner of result cell
- Then there will be a + sign that will appear and
- Drag till you want the results.

The result will be as shown in Fig 1.2.12

	A	B	C	D	E	F	G	H
1	Name	Age	Gender	Salary (INR)	Years of Experience	Sales (INR)	Bonus (INR)	Total Benefits
2	Amit	28	Male	₹ 50,000	4	₹ 10,000	₹ 2,500	₹ 52,500
3	Priya	35	Female	₹ 80,000	10	₹ 15,000	₹ 3,500	₹ 83,500
4	Rajesh	42	Male	₹ 65,000	15	₹ 20,000	₹ 5,000	₹ 70,000
5	Neha	31	Female	₹ 55,000	7	₹ 12,000	₹ 3,000	₹ 58,000
6	Rahul	29	Male	₹ 60,000	6	₹ 13,000	₹ 2,800	₹ 62,800
7	Deepika	27	Female	₹ 45,000	3	₹ 9,000	₹ 2,000	₹ 47,000
8	Sandeep	38	Male	₹ 70,000	12	₹ 18,000	₹ 4,000	₹ 74,000
9	Pooja	33	Female	₹ 60,000	8	₹ 14,000	₹ 3,200	₹ 63,200
10	Vivek	36	Male	₹ 75,000	13	₹ 19,000	₹ 4,500	₹ 79,500
11	Anita	30	Female	₹ 50,000	5	₹ 11,000	₹ 2,600	₹ 52,600
12	Gopal	41	Male	₹ 85,000	18	₹ 22,000	₹ 5,500	₹ 90,500
13	Sarika	34	Female	₹ 70,000	9	₹ 16,000	₹ 3,800	₹ 73,800
14	Kunal	26	Male	₹ 55,000	2	₹ 8,000	₹ 1,800	₹ 56,800
15	Rina	32	Female	₹ 45,000	4	₹ 9,500	₹ 2,200	₹ 47,200
16	Manoj	37	Male	₹ 65,000	11	₹ 17,000	₹ 3,900	₹ 68,900
17	Rajani	29	Female	₹ 60,000	7	₹ 14,000	₹ 3,200	₹ 63,200
18	Harish	40	Male	₹ 80,000	16	₹ 21,000	₹ 5,200	₹ 85,200
19	Sonia	31	Female	₹ 55,000	6	₹ 10,000	₹ 2,400	₹ 57,400
20	Rajat	35	Male	₹ 50,000	5	₹ 12,000	₹ 2,800	₹ 52,800
21	Neeta	39	Female	₹ 70,000	10	₹ 16,000	₹ 3,800	₹ 73,800

Fig 1.2.12 Results After Drag and Autofill Options

By using formulas in Excel, you can perform calculations, analyse data, automate tasks, and manipulate data efficiently.

1.2.6 Working with Built in Functions

Excel provides a wide range of built-in functions that perform specific operations and calculations. These functions are categorised into different groups based on their purpose. Some of the main categories of built-in functions in Excel are as follows:

1. **Mathematical Functions:** These functions perform various mathematical calculations.
   - SUM: Adds up values in a range.
   - AVERAGE: Calculates the average of values in a range.
   - MIN: Returns the minimum value from a range.
   - MAX: Returns the maximum value from a range.
   - COUNT: Counts the number of cells that contain numbers.
2. **Text Functions:** These functions manipulate and analyse text.

- CONCATENATE: Joins multiple text strings together.
- LEN: Returns the length of a text string.
- LEFT: Extracts a specified number of characters from the beginning of a text string.
- RIGHT: Extracts a specified number of characters from the end of a text string.
- MID: Extracts a specified number of characters from a text string, starting at a specified position.

3. **Logical Functions:** These functions evaluate logical conditions and return logical results.

- IF: Performs a logical test and returns different values based on the result.
- AND: Checks if all specified conditions are TRUE.
- OR: Checks if at least one of the specified conditions is TRUE.
- NOT: Negates the logical value of a condition.

4. **Date and Time Functions:** These functions manipulate and perform calculations with dates and times.

- Built-in functions - mathematical, text, logical, date and time, and financial functions

- TODAY: Returns the current date.
- NOW: Returns the current date and time.
- DATE: Returns a date based on the provided year, month, and day.
- EOMONTH: Returns the last day of the month before or after a specified number of months.
- DATEDIF:Calculates thedifference between two dates in various units (days, months, years, etc.).

5. **Financial Functions:** These functions are used for fi nancial calculations.

- PMT: Calculates the payment for a loan based on con stant payments and a constant interest rate.
- PV: Calculates the present value of an investment or loan.
- FV: Calculates the future value of an investment or loan.
- RATE: Calculates the interest rate per period of an in vestment or loan.

These are just a few of the many built-in functions available in Excel. Excel offers a number of functions that cover a wide range of calculations, statistical analysis, conditional formatting, data manipulation, and more. You can access

these functions through the Formulas tab in Excel or by typing the function name directly into a cell and providing the required arguments.

1.2.7 Using Absolute and Relative Cell References

A cell reference is a way to identify and refer to a specific cell or range of cells within a worksheet. It is used in formulas, functions, and other operations to perform calculations, manipulate data, and retrieve values from cells. The cell reference is a combination of column letter and the row number; the two components uniquely identify a cell within the worksheet. For example, the cell reference "A1" refers to the cell in column A and row 1.

There are three types of cell references in Excel:

1. **Relative Cell Reference:** A relative cell reference adjusts when copied or filled to other cells. For example, if you have a formula "=A1+B1" in cell C1 and copy it to cell C2, the formula in C2 will automatically adjust to "=A2+B2". The references A1 and B1 change relative to the new position of the formula.
2. **Absolute Cell Reference:** An absolute cell reference remains fixed when copied or filled to other cells. To create an absolute reference, you add a dollar sign ($) before the column letter and/or row number. For example, the formula "=A$1+B$1" will always refer to cell A1 and B1, regardless of the new position of the formula.
3. **Mixed Cell Reference:** A mixed cell reference combines aspects of both absolute and relative references. You can fix either the column or the row while allowing the other part to adjust when copied. For example, the formula "=A$1+B2" will always refer to cell A1, but the row reference adjusts to B2 when copied to other cells.

- Cell references enable creation of dynamic formulas and functions

Cell references can be used in various ways in Excel. They are primarily used in formulas and functions to perform calculations based on the values in different cells. By referencing cells, you can create dynamic and flexible formulas that update automatically when the referenced cells change. This allows you to perform complex calculations, analyse data, and create dynamic reports and dashboards within your Excel worksheets.

Cell references are an essential aspect of Excel because they provide flexibility, automation, and dynamic capabilities in various spreadsheet tasks. The main reasons why we use cell references in Excel are:

1. **Perform Calculations:** Cell references allow us to perform calculations and mathematical operations in Excel. By referencing cells containing values or formulas, we can create complex formulas that dynamically update based on the values in the referenced cells. This enables us to automate calculations and perform calculations across multiple cells or ranges.
2. **Dynamic Data Analysis:** Cell references are crucial for data analysis in Excel. By referencing cells or ranges of data, we can perform various analysis operations such as sorting, filtering, conditional formatting, and creating charts. Cell references enable us to analyse and manipulate data dynamically, making it easier to derive insights and make informed decisions.
3. **Report Generation:** Cell references are widely used in creating reports in Excel. By referencing cells containing data, we can generate dynamic reports that update automatically as the underlying data changes. Cell references allow us to pull data from different cells or ranges, perform calculations, and present the information in a summarised or organised format. This makes it easier to generate accurate and up-to-date reports.
4. **Data Validation**: Cell references play a vital role in data validation in Excel. By referencing cells, we can define rules and conditions that restrict the type of data or values that can be entered. Cell references enable us to validate and verify data entered by users, ensuring data accuracy, consistency, and adherence to predefined rules.
5. **Conditional Formatting:** Cell references are used extensively in conditional formatting in Excel. By referencing cells, we can define formatting rules that dynamically change the appearance of cells based on specified conditions. Cell references allow us to highlight data, apply color schemes, icon sets, or data bars based on the values or conditions in the referenced cells. This helps in visualising data, identifying patterns, and drawing attention to important information.
6. **Data Integration:** Cell references are used when integrating Excel with other applications or data sources.

- Enable calculations, data analysis, report generation, data validation, conditional formatting, and data integration

By referencing cells in formulas, we can retrieve or link data from external sources, such as databases, web services, or other Excel workbooks. Cell references enable us to establish connections and pull data from external sources into our Excel worksheets, allowing for data consolidation and analysis.

Overall, cell references in Excel provide the ability to create dynamic and automated spreadsheets, perform calculations, analyse data, generate reports, and integrate with external sources. They enhance the efficiency, accuracy, and flexibility of our Excel tasks and allow for efficient data manipulation and analysis.

To create a **relative cell reference** in Excel, you simply enter the cell reference without any dollar signs ($). Here is how you can do it:

- Allows the reference to adjust automatically when copied or filled to other cells

1. Select the cell where you want to enter the formula or value that contains the relative cell reference.
2. Begin typing the formula or value in the cell, starting with an equal sign (=) to indicate that you are entering a formula.
3. To create a relative cell reference, enter the column letter followed by the row number of the cell. For example, to reference cell A1, you would enter "A1".
4. After entering the relative cell reference, you can continue entering the rest of the formula or value.
5. Press Enter to confirm the formula or value, and Excel will use the relative cell reference.

When you copy or fill a formula or value that contains a relative cell reference to other cells, the referenced column(s) and row(s) will adjust automatically relative to the new position of the formula. For example, if you copy a formula with a relative cell reference from cell A1 to cell B1, the reference will adjust to B1. Similarly, if you copy the formula to cell A2, the reference will adjust to A2.

Relative cell references are useful when you want the formula or value to adapt to different locations in the worksheet. They allow you to perform calculations based on the relative position of cells, making it easier to apply the same logic to multiple cells or ranges.

To explain relative cell reference let us use a data sheet as given in Fig 1.2.13

	A	B	C	D	E	F
1	**Salesman**	**Quarter 1 (₹)**	**Quarter 2 (₹)**	**Quarter 3 (₹)**	**Quarter 4 (₹)**	**Total Sales (₹)**
2	Ravi	₹ 15,000	₹ 18,000	₹ 20,000	₹ 22,000	
3	Asha	₹ 12,000	₹ 16,000	₹ 14,000	₹ 18,000	
4	Rajesh	₹ 18,000	₹ 22,000	₹ 19,000	₹ 21,000	
5	Meena	₹ 14,000	₹ 16,000	₹ 17,000	₹ 19,000	
6	Mohan	₹ 16,000	₹ 18,000	₹ 15,000	₹ 20,000	
7						

Fig 1.2.13 Quarterwise Sales Dataset

So in order to find the Total Sales we can use multiple options like either using "= sign" or "using fx insert option" or " Alt+ = sign and then enter option". Then the result will be as shown in the Fig 1.2.14

	A	B	C	D	E	F
1	**Salesman**	**Quarter 1 (₹)**	**Quarter 2 (₹)**	**Quarter 3 (₹)**	**Quarter 4 (₹)**	**Total Sales (₹)**
2	Ravi	₹ 15,000	₹ 18,000	₹ 20,000	₹ 22,000	**₹ 75,000**
3	Asha	₹ 12,000	₹ 16,000	₹ 14,000	₹ 18,000	
4	Rajesh	₹ 18,000	₹ 22,000	₹ 19,000	₹ 21,000	
5	Meena	₹ 14,000	₹ 16,000	₹ 17,000	₹ 19,000	
6	Mohan	₹ 16,000	₹ 18,000	₹ 15,000	₹ 20,000	
7						

Fig 1.2.14 Finding Total Sales in One Row

Here we found total sales for one person, so to find the total sales for rest of the salesmen we can use drag and drop option or auto fill option.

When we do this the sum of all the for quarters will be added to the total sales column and shown as in the Fig 1.2.15

	A	B	C	D	E	F
1	**Salesman**	**Quarter 1 (₹)**	**Quarter 2 (₹)**	**Quarter 3 (₹)**	**Quarter 4 (₹)**	**Total Sales (₹)**
2	Ravi	₹ 15,000	₹ 18,000	₹ 20,000	₹ 22,000	**₹ 75,000**
3	Asha	₹ 12,000	₹ 16,000	₹ 14,000	₹ 18,000	**₹ 60,000**
4	Rajesh	₹ 18,000	₹ 22,000	₹ 19,000	₹ 21,000	**₹ 80,000**
5	Meena	₹ 14,000	₹ 16,000	₹ 17,000	₹ 19,000	**₹ 66,000**
6	Mohan	₹ 16,000	₹ 18,000	₹ 15,000	₹ 20,000	**₹ 69,000**

Fig 1.2.15 Total Sales of all Salesmen

This is relative cell reference. By default, all cell references are relative cell references.

To create an **absolute cell reference,** you need to add a dollar sign ($) before the column letter and/or row number of the cell reference. The steps to create an absolute cell reference are as follows:

- Identifies and refer to specific cells or ranges within a worksheet.

1. Select the cell where you want to enter the formula or value that contains the absolute cell reference.
2. Begin typing the formula or value in the cell, starting with an equal sign (=) to indicate that you are entering a formula.
3. To create an absolute cell reference, you need to add a dollar sign ($) before the column letter and/or row number of the cell reference. Here are the different ways to use absolute cell references:
4. To make both the column and row absolute, place a dollar sign before both the column letter and row number. For example, to create an absolute reference to cell A1, you would enter “A1”.
5. To make only the column absolute, place a dollar sign before the column letter but not before the row number. For example, to create an absolute reference to column A but allow the row number to adjust, you would enter “$A1”.
6. To make only the row absolute, place a dollar sign before the row number but not before the column letter. For example, to create an absolute reference to row 1 but allow the column letter to adjust, you would enter “A$1”.
7. After adding the dollar sign(s) to the appropriate location(s), you can continue entering the rest of the formula or value.
8. Press Enter to confirm the formula or value, and Excel will use the absolute cell reference.

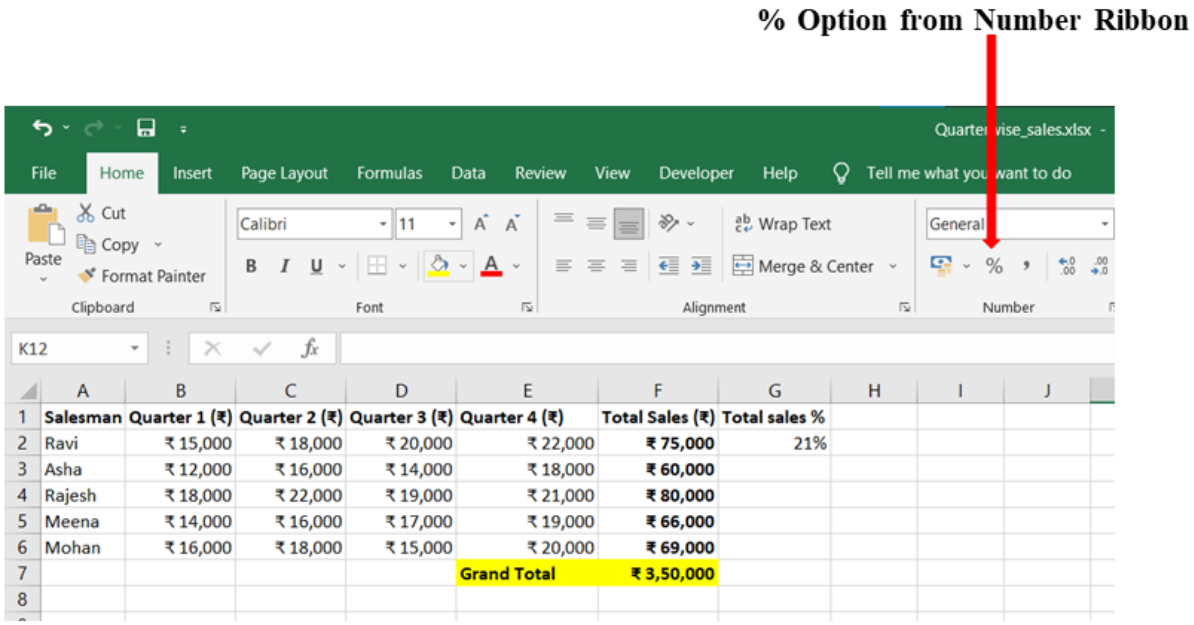
When you copy or fill a formula or value that contains an absolute cell reference to other cells, the referenced column(s) and/or row(s) will remain fixed, and they will not adjust relative to the new position of the formula. This allows you to refer to specific cells or ranges consistently in your formulas, even when copied to different locations in the worksheet.

Let us work on the same Quarterwise Sales Dataset for understanding absolute cell reference. For that we have taken the total sales in that year and added one more column titled Total Sales %. It is as shown in the Fig 1.2.16.

	A	B	C	D	E	F	G
1	**Salesman**	**Quarter 1 (₹)**	**Quarter 2 (₹)**	**Quarter 3 (₹)**	**Quarter 4 (₹)**	**Total Sales (₹)**	**Total sales %**
2	Ravi	₹ 15,000	₹ 18,000	₹ 20,000	₹ 22,000	**₹ 75,000**	
3	Asha	₹ 12,000	₹ 16,000	₹ 14,000	₹ 18,000	**₹ 60,000**	
4	Rajesh	₹ 18,000	₹ 22,000	₹ 19,000	₹ 21,000	**₹ 80,000**	
5	Meena	₹ 14,000	₹ 16,000	₹ 17,000	₹ 19,000	**₹ 66,000**	
6	Mohan	₹ 16,000	₹ 18,000	₹ 15,000	₹ 20,000	**₹ 69,000**	
7					**Grand Total**	**₹ 3,50,000**	
8							

Fig 1.2.16 Quarterwise Sales Dataset with Total Sales Percentage

To find the total sales percentage of Ravi we will use the formula Total sales of Ravi divided by Grand Total. And then, choose percentage option from the Number ribbon. The result will be as shown in the Fig 1.2.17.



	A	B	C	D	E	F	G	H	I	J
1	**Salesman**	**Quarter 1 (₹)**	**Quarter 2 (₹)**	**Quarter 3 (₹)**	**Quarter 4 (₹)**	**Total Sales (₹)**	**Total sales %**			
2	Ravi	₹ 15,000	₹ 18,000	₹ 20,000	₹ 22,000	**₹ 75,000**	21%			
3	Asha	₹ 12,000	₹ 16,000	₹ 14,000	₹ 18,000	**₹ 60,000**				
4	Rajesh	₹ 18,000	₹ 22,000	₹ 19,000	₹ 21,000	**₹ 80,000**				
5	Meena	₹ 14,000	₹ 16,000	₹ 17,000	₹ 19,000	**₹ 66,000**				
6	Mohan	₹ 16,000	₹ 18,000	₹ 15,000	₹ 20,000	**₹ 69,000**				
7					**Grand Total**	**₹ 3,50,000**				
8										

Fig 1.2.17 Sales Percentage of Ravi

If we try to drag and drop or auto fill the result will be as shown in Fig 1.2.18

	A	B	C	D	E	F	G
1	**Salesman**	**Quarter 1 (₹)**	**Quarter 2 (₹)**	**Quarter 3 (₹)**	**Quarter 4 (₹)**	**Total Sales (₹)**	**Total sales %**
2	Ravi	₹ 15,000	₹ 18,000	₹ 20,000	₹ 22,000	**₹ 75,000**	21%
3	Asha	₹ 12,000	₹ 16,000	₹ 14,000	₹ 18,000	**₹ 60,000**	#DIV/0!
4	Rajesh	₹ 18,000	₹ 22,000	₹ 19,000	₹ 21,000	**₹ 80,000**	#DIV/0!
5	Meena	₹ 14,000	₹ 16,000	₹ 17,000	₹ 19,000	**₹ 66,000**	#DIV/0!
6	Mohan	₹ 16,000	₹ 18,000	₹ 15,000	₹ 20,000	**₹ 69,000**	#DIV/0!
7					**Grand Total**	**₹ 3,50,000**	

Fig 1.2.18 Error Message During Auto Fill

This is where we use absolute cell reference. By using absolute references, you ensure that the formula always refers to a specific cell, regardless of its location when copied.

The distinction between absolute and relative cell references in Excel is pivotal for effective spreadsheet design and data analysis. Whether choosing absolute, relative, or a combination of both in the form of mixed references, understanding these concepts helps the users to create Excel formulas that meet the needs of their data analysis, offering a balance between stability and adaptability in spreadsheet applications.

Summarised Overview

Sorting data allows you to rearrange rows based on specific criteria, such as alphabetical order or numerical values, facilitating pattern identification, data organisation, and analysis. Excel offers various sorting options, including sorting by values, cell color, font color, cell icon, and more. Filtering, on the other hand, allows you to selectively display specific data based on criteria, narrowing down large datasets to focus on relevant information. Excel provides different filtering options such as AutoFilter, Filter by Selection, Filter by Color, Advanced Filter, and more. These options enable you to filter data based on text, numbers, dates, cell color, font color, or custom criteria. Advanced filtering in Excel provides even more control and flexibility by allowing the use of complex filter criteria using formulas. It enables you to define intricate conditions, apply multiple criteria simultaneously, use logical operators, and wildcard characters. Advanced filtering is particularly useful when standard filtering options are not sufficient for the analysis.

Data validation 1 is a feature that allows you to set rules and restrictions on the data entered into cells. It helps ensure the accuracy and validity of the data by controlling the input. Excel provides various methods of data validation, such as whole number, decimal, date, time, text length, custom formulas, and list validation. Formulas and operators are essential in Excel for performing calculations, manipulating data, and obtaining results. Formulas consist of values, cell references, functions, and operators. Operators are symbols or characters that perform specific operations on values or cell references, such as arithmetic operations, comparison operations, concatenation, and logical operations. By using operators within formulas, you can create complex calculations and manipulate data effectively.

Cell references are a way to identify and refer to specific cells or ranges within a worksheet. They consist of a column letter and a row number, and there are three types: relative, absolute, and mixed. Cell references are essential for performing calculations, analysing data, generating reports, validating data, applying conditional formatting, integrating with external sources, and creating dynamic and automated spreadsheets. Absolute references remain fixed when copied, relative references adjust automatically, and mixed references allow for fixing either the column or the row. By utilising cell references, users can efficiently manipulate and analyse data.

Self-Assessment

1. Explain the purpose and benefits of sorting data in Excel. How can you sort data in ascending or descending order based on a specific column?
2. Describe the steps involved in applying filters to a data set in Excel. How can you filter data to display specific criteria or conditions?
3. Describe the process of using advanced filtering in Excel. How does advanced filtering differ from regular filtering, and what additional functionalities does it offer?
4. Explain the concept of absolute and relative cell references in Excel formulas. How do these references affect the formula calculations when copied to different cells?
5. Provide a step-by-step guide on how to use the SUM function in Excel. Include examples of using the SUM function with both absolute and relative cell references.

Assignments

1. Using the dataset given below and complete the assignment questions.

Student ID	Name	Age	Mathematics	Science
1	Aarav	16	85	90
2	Nisha	15	92	88
3	Arjun	17	78	92
4	Riya	16	88	94
5	Aryan	15	90	85
6	Maya	17	95	96
7	Dhruv	15	82	80
8	Ishita	16	90	92
9	Rohan	17	88	85
10	Aanya	15	92	90

a. Sort the student data based on their mathematics scores in descending order.

b. Use advanced filtering to extract the records of students who scored above 90 in both mathematics and science.

c. Create a new column to calculate the average score for each student using a formula.

d. Use the SUM function to calculate the total maths score for the class

e. Use absolute and relative cell reference and find total marks for both the subjects for all the students.

f. Calculate the average Mathematics score for the entire class using an absolute cell reference for the range.

g. Compute the average Mathematics score for each student using a relative cell reference for the range.

Reference

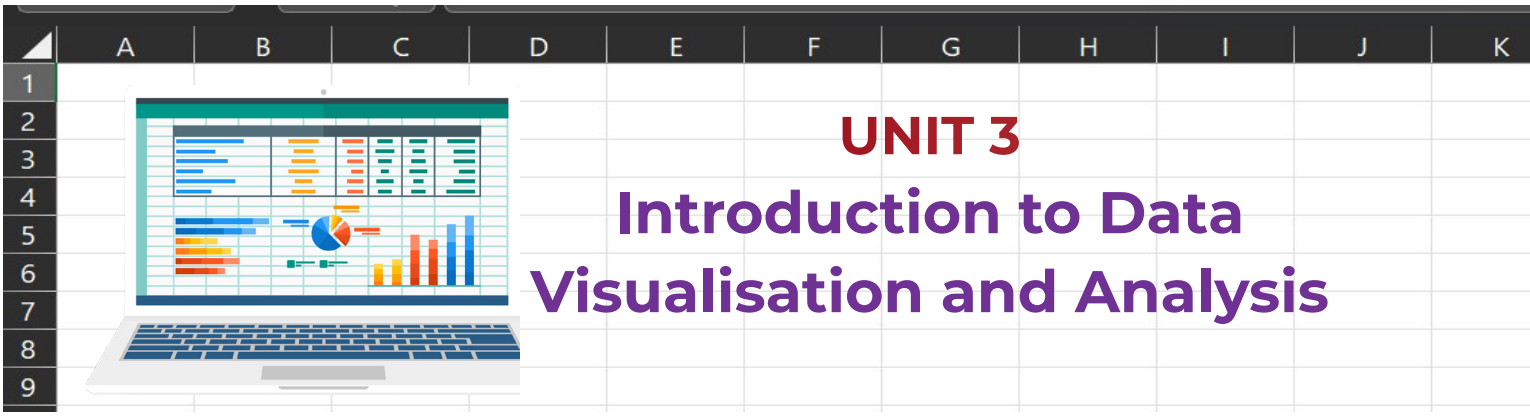
1. Alexander, M., Kusleika, R., & Walkenbach, J. (2018). *Excel 2019 Bible*. John Wiley & Sons.
2. Frye, C. (2018). *Microsoft Excel 2019 step by step*. Microsoft Press.
3. Harvey, G. (2018). *Excel 2019 for dummies*. John Wiley & Sons.
4. Bluttman, K. (2013). *Excel formulas and functions for dummies.* John Wiley & Sons.

Suggested Reading

1. McFedries, P., & Harvey, G. (2021). *Excel All-in-one for Dummies*. John Wiley & Sons.
2. Jelen, B., & Juhasz, S. (2015). *MrExcel XL: The 40 Greatest Excel Tips of All Time*. Tickling Keys, Inc.
3. Jelen, B. (2015). *Power Excel with MrExcel: Master Pivot Tables, Subtotals, Charts, VLOOKUP, IF, Data Analysis in Excel 2010–2013*. Tickling Keys, Inc.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.



UNIT 3
Introduction to Data Visualisation and Analysis

Learning Outcomes

After the completion of this unit, the learner will be able to

- understanding data visualisation techniques
- proficiently use conditional formatting for data visualisation
- create and modify different types charts
- create pivot tables and pivot charts

Background

Data visualisation is the process of presenting data visually using charts, graphs, and tables to facilitate understanding and analysis. With a variety of chart types available, including column, bar, line, and pie charts, Excel empowers users to transform raw data into visually appealing representations. These charts can be customised with formatting options, such as colours, labels, and titles, to effectively communicate data insights. Additionally, conditional formatting allows users to apply formatting rules to cells based on specific conditions, enabling the highlighting of data patterns and trends. This feature enhances data visualisation by providing visual cues that aid in the interpretation and analysis of the data.

Excel goes beyond individual charts by offering the capability to create dynamic dashboards and reports. Users can combine multiple visualisations, charts, and tables on a single worksheet, creating interactive dashboards that provide a comprehensive view of the data. Slicers and filters further enhance the interactive experience, allowing users to drill down into details and dynamically update the visualisations based on changing data. These dynamic dashboards and reports enable users to explore the data more effectively and derive meaningful insights.

Keywords

Conditional Formatting, Data Visualisation, Charts, Pivot Tables, Pivot Chart

Discussion

1.3.1 Conditional Formatting

Conditional formatting in allows the user to apply formatting to cells based on specific conditions or criteria. It provides a way to visually highlight, emphasise, or colourise cells that meet certain conditions, making it easier to interpret and analyse data.

- Conditional formatting enables users to apply formatting to cells based on conditions

Conditional formatting plays a significant role in enhancing data visualisation by providing visual cues that highlight specific data patterns, trends, or outliers. It helps users to focus on key insights and draw attention to important aspects of the data.

1.3.1.1 Using Conditional Formatting for Data Visualisation

Conditional formatting helps in data visualisation in the following ways:

1. **Highlighting Data Points:** You can use conditional formatting to highlight specific values or data points that meet certain conditions. For example, you can apply a bold font or a different background colour to cells that contain sales figures above a certain threshold, making it easier to identify high-performing products or regions.
2. **Colour-Scale Formatting:** Excel provides colour-scale formatting options where the colour of cells is determined by their relative values within a range. For example, you can apply a colour scale that ranges from green to red, where higher values appear in shades of green and lower values in shades of red. This visual gradient helps you quickly identify variations and patterns in the data.
3. **Data Bars:** Data bars are horizontal bars that are added to cells to represent the magnitude or value of the data. The length of the data bar corresponds to the value, allowing you to compare values visually. This is useful when comparing sales figures, budgets, or any other numeric data.
4. **Icon Sets:** Excel provides built-in icon sets that you can

apply based on specific conditions. For example, you can use arrow icons to indicate whether a value is increasing, decreasing, or staying the same. This makes it easy to spot trends or changes in data at a glance.

5. **Highlighting Duplicates or Unique Values:** Conditional formatting can be used to highlight duplicate values or unique values within a dataset. This is useful for identifying duplicate entries in a list or finding unique items in a large dataset.
6. **Data Validation:** Although not strictly conditional formatting, data validation can be used to control input and highlight invalid data. By setting validation rules, you can restrict the type of data that can be entered in a cell and display an error message if the input does not meet the specified criteria.

- Enhances the visual appeal, interpretation, and pattern identification of data

By applying these visual formatting techniques, conditional formatting helps to make your data more visually appealing, easier to interpret, and facilitates the identification of patterns and trends. It allows you to focus on the most important aspects of your data and present it in a way that engages and informs the audience.

In Excel, conditional formatting offers a range of options to customise the formatting of cells based on specific conditions. The some commonly used conditional formatting options are:

a. Highlight Cells Rules:

- Greater Than: Highlights cells that are greater than a specified value.
- Less Than: Highlights cells that are less than a specified value.
- Between: Highlights cells that are between two specified values.
- Equal To: Highlights cells that are equal to a specified value.
- Text That Contains: Highlights cells that contain specific text.
- Duplicate Values: Highlights duplicate values within a range.
- Unique Values: Highlights unique values within a range.

b. Top/Bottom Rules:

- Top 10 Items: Highlights the top values in a range (e.g., top 10 sales figures).
- Bottom 10 Items: Highlights the bottom values in a range.

- Above Average: Highlights values that are above the average of a range.
- Below Average: Highlights values that are below the average of a range.

c. Data Bars:

- Solid Fill: Adds a horizontal bar with solid fill colour proportional to the cell's value.
- Gradient Fill: Adds a horizontal bar with a gradient fill based on the cell's value.
- Colour Scales: Applies a gradient of colours to cells based on their values, creating a visual scale.
- Icon Sets: Adds icons to cells based on their values, such as arrows, traffic lights, or symbols.

- Enables flexible and customisable formatting of cells based on specific conditions

d. Colour Scales:

- Two-Colour Scale: Applies a two-colour gradient to cells based on their values.
- Three-Color Scale: Applies a three-color gradient to cells based on their values.
- Data Bars: Similar to the Data Bars option, adds horizontal bars with gradient fill based on the cell's value.
- Icon Sets: Similar to the Icon Sets option, adds icons based on the cell's value.

e. New Rule (Custom Rules):

- Formula-based Rule: Allows you to create custom rules using formulas. You can define conditions based on your specific criteria using Excel formulas. For example, you can highlight cells that meet a specific condition by writing a formula that evaluates to TRUE or FALSE.

These options provide flexibility in defining conditions and formatting styles to suit your specific data and requirements. You can mix and match different conditional formatting rules and combine them for more complex formatting scenarios. Excel also allows you to manage and modify conditional formatting rules, reorder them, and adjust their settings as needed.

To use conditional formatting in Excel follow the steps given below :

1. Select the cells or range of cells that you want to apply conditional formatting to.
2. Go to the "Home" tab in the ribbon.
3. Click on the "Conditional Formatting" button, usually

found under the “Styles” group.Refer Fig 1.3.2

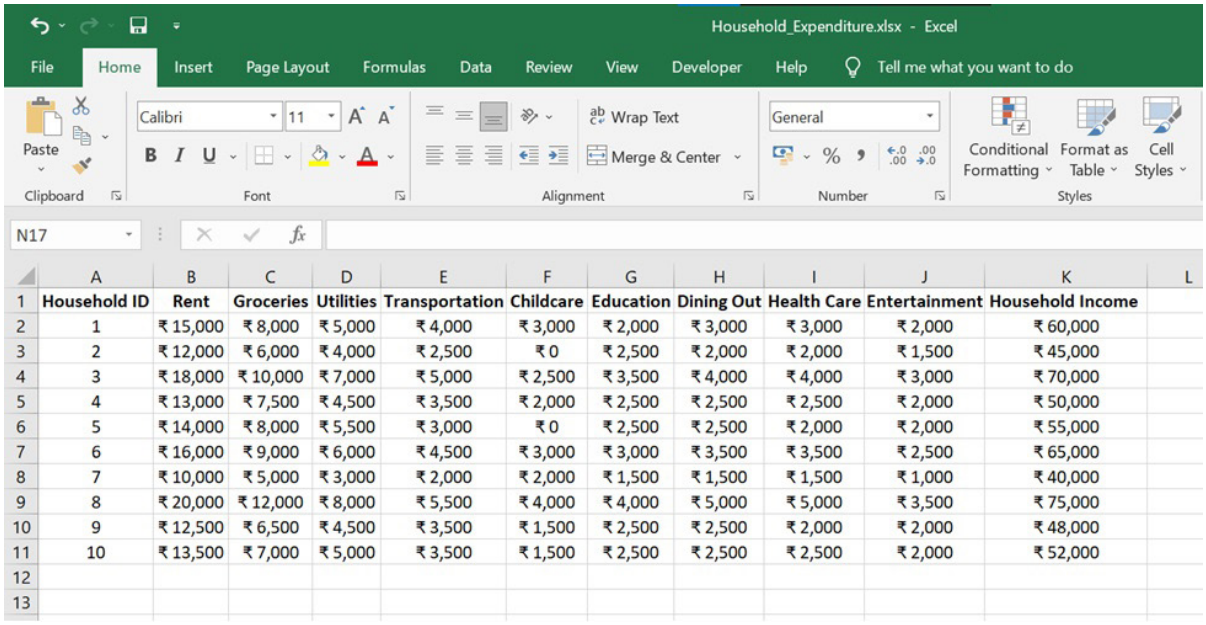
4. A dropdown menu will appear with various conditional formatting options as shown in Fig 1.3.2. Let us explore some of the common pre-defined options:
   - Highlight Cells Rules
   - Top/Bottom Rules
   - Data Bars
   - Colour Scales
   - Icon Sets
5. If you select one of the pre-defined options, Excel will prompt you to enter the criteria or values for applying the formatting. For example, if you choose the “Greater Than” option, you need to specify the value that cells should be greater than to trigger the formatting.
6. Additionally, you can select “New Rule” from the dropdown menu to create custom formatting rules using formulas. This provides more flexibility in defining the conditions for formatting. When you select “New Rule,” the “New Formatting Rule” dialog box will appear.
7. In the “New Formatting Rule” dialog box, you have the option to choose the rule type, such as Format only cells that contain, Format only top or bottom ranked values, Format only values that are above or below average, and more. Each rule type has its own set of options for defining the conditions.
8. For custom rules, you can enter a formula or expression in the “Format values where this formula is true” field. The formula should evaluate to TRUE or FALSE for each cell, where TRUE indicates that the formatting should be applied.
9. After defining the formatting rule, you can click on the “Format” button to specify the formatting style. This includes options such as font color, background color, number format, borders, and more.
10. Once you have set up the formatting rule and style, click “OK” to apply the conditional formatting to the selected cells.

- Choose a pre-defined option or create a custom rule, define the conditions, specify the formatting style

Excel will then automatically apply the formatting to the cells that meet the specified conditions. The formatting will dynamically update as you modify the values in the cells, allowing you to visually identify trends, patterns, outliers, or specific conditions within your data.

You can manage and modify the conditional formatting rules by selecting the cells and clicking on the "Conditional Formatting" button again. From the dropdown menu, you can choose "Manage Rules" to view, edit, or delete existing rules.

For working on conditional formatting let us create a new dataset titled "Household_Expenditure". The Fig 1.3.1 shown below explains the data set. It includes the total household income for a month and various expenditures incurred by 10 households in that month. Let us use conditional formatting and explain the household expenditure dataset.



	A	B	C	D	E	F	G	H	I	J	K	L
1	Household ID	Rent	Groceries	Utilities	Transportation	Childcare	Education	Dining Out	Health Care	Entertainment	Household Income	
2	1	₹15,000	₹8,000	₹5,000	₹4,000	₹3,000	₹2,000	₹3,000	₹3,000	₹2,000	₹60,000	
3	2	₹12,000	₹6,000	₹4,000	₹2,500	₹0	₹2,500	₹2,000	₹2,000	₹1,500	₹45,000	
4	3	₹18,000	₹10,000	₹7,000	₹5,000	₹2,500	₹3,500	₹4,000	₹4,000	₹3,000	₹70,000	
5	4	₹13,000	₹7,500	₹4,500	₹3,500	₹2,000	₹2,500	₹2,500	₹2,500	₹2,000	₹50,000	
6	5	₹14,000	₹8,000	₹5,500	₹3,000	₹0	₹2,500	₹2,500	₹2,000	₹2,000	₹55,000	
7	6	₹16,000	₹9,000	₹6,000	₹4,500	₹3,000	₹3,000	₹3,500	₹3,500	₹2,500	₹65,000	
8	7	₹10,000	₹5,000	₹3,000	₹2,000	₹2,000	₹1,500	₹1,500	₹1,500	₹1,000	₹40,000	
9	8	₹20,000	₹12,000	₹8,000	₹5,500	₹4,000	₹4,000	₹5,000	₹5,000	₹3,500	₹75,000	
10	9	₹12,500	₹6,500	₹4,500	₹3,500	₹1,500	₹2,500	₹2,500	₹2,000	₹2,000	₹48,000	
11	10	₹13,500	₹7,000	₹5,000	₹3,500	₹1,500	₹2,500	₹2,500	₹2,500	₹2,000	₹52,000	
12												
13												

Fig 1.3.1 Household Expenditure Dataset

Using conditional formatting let us find all the expenses higher than ₹2,000 in that month for Household ID 5.

The steps to do that is as follows and is shown in Fig1.3.2:

1. Select the cells which you want to use for conditional formatting.
2. Click on Conditional formatting Option from Ribbon
3. Here we are going to use Greater than option and it is found under highlight cell rules.

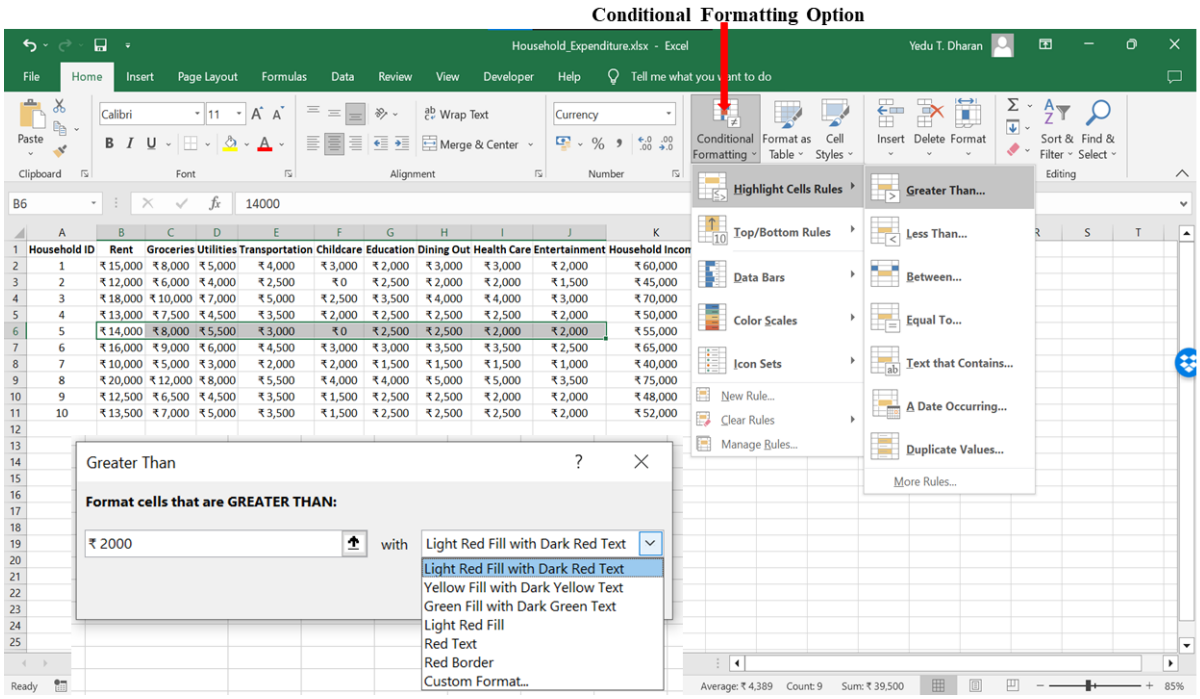


Fig 1.3.2 Using Greater than Condition

4. Click on Highlight cell rules and then on Greater than option.
5. A menu will pop up and type in your condition in the first cell. Here our condition is ₹2000 so, type that in
6. Select whichever option you want from the next cell. Here we have chosen Light red with dark red text for conditional formatting.
7. Then click OK and the result will be as shown in the Fig 1.3.3

	A	B	C	D	E	F	G	H	I	J	K
1	Household ID	Rent	Groceries	Utilities	Transportation	Childcare	Education	Dining Out	Health Care	Entertainment	Household Income
2	1	₹ 15,000	₹ 8,000	₹ 5,000	₹ 4,000	₹ 3,000	₹ 2,000	₹ 3,000	₹ 3,000	₹ 2,000	₹ 60,000
3	2	₹ 12,000	₹ 6,000	₹ 4,000	₹ 2,500	₹ 0	₹ 2,500	₹ 2,000	₹ 2,000	₹ 1,500	₹ 45,000
4	3	₹ 18,000	₹ 10,000	₹ 7,000	₹ 5,000	₹ 2,500	₹ 3,500	₹ 4,000	₹ 4,000	₹ 3,000	₹ 70,000
5	4	₹ 13,000	₹ 7,500	₹ 4,500	₹ 3,500	₹ 2,000	₹ 2,500	₹ 2,500	₹ 2,500	₹ 2,000	₹ 50,000
6	5	₹ 14,000	₹ 8,000	₹ 5,500	₹ 3,000	₹ 0	₹ 2,500	₹ 2,500	₹ 2,000	₹ 2,000	₹ 55,000
7	6	₹ 16,000	₹ 9,000	₹ 6,000	₹ 4,500	₹ 3,000	₹ 3,000	₹ 3,500	₹ 3,500	₹ 2,500	₹ 65,000
8	7	₹ 10,000	₹ 5,000	₹ 3,000	₹ 2,000	₹ 2,000	₹ 1,500	₹ 1,500	₹ 1,500	₹ 1,000	₹ 40,000
9	8	₹ 20,000	₹ 12,000	₹ 8,000	₹ 5,500	₹ 4,000	₹ 4,000	₹ 5,000	₹ 5,000	₹ 3,500	₹ 75,000
10	9	₹ 12,500	₹ 6,500	₹ 4,500	₹ 3,500	₹ 1,500	₹ 2,500	₹ 2,500	₹ 2,000	₹ 2,000	₹ 48,000
11	10	₹ 13,500	₹ 7,000	₹ 5,000	₹ 3,500	₹ 1,500	₹ 2,500	₹ 2,500	₹ 2,500	₹ 2,000	₹ 52,000
12											

Fig 1.3.3 Result of Conditional Formatting

- Enables effective data visualisation and analysis by highlighting cells that meet conditions

Here we have all the expenditure of Household 5 that is more than ₹2000 in that month. In the same way we can find out and format according to the conditions we provide. That is why this method is called conditional formatting and it is a great data visualisation tool.

1.3.2 Creating and Modifying Charts

Charts in Excel are graphical representations of data that allow you to visualise and analyse information more effectively. They provide a visual way to present data, making it easier to understand patterns, trends, and comparisons. Excel offers a wide variety of chart types, each suited for different data types and analysis purposes.

1.3.2.1 Types of Charts in Excel:

Excel provides various types of charts for specific needs. We have the flexibility to choose a suitable chart type depending on the nature of data. Moreover, Excel allows us to modify the chart type at a later stage to better align with our requirements.

- Flexibility to choose a suitable chart type depending on the nature of data

a. Column Charts:

- Clustered Column Chart: Displays vertical bars side by side to compare values across categories.
- Stacked Column Chart: Shows the relationship of individual items to the whole, with values stacked on top of each other.
- 100% Stacked Column Chart: Similar to the stacked column chart, but the stacked columns represent percentages of the whole.

b. Bar Charts:

- Clustered Bar Chart: Similar to the clustered column chart, but with horizontal bars.
- Stacked Bar Chart: Like the stacked column chart, but with horizontal bars.
- 100% Stacked Bar Chart: Represents each value as a percentage of the whole using horizontal bars.

c. Line Charts:

- Line Chart: Displays trends over time or ordered categories by connecting data points with lines.
- Stacked Line Chart: Similar to the line chart, but with multiple lines stacked on top of each other.

d. Pie Charts:

- Pie Chart: Represents proportions of a whole by dividing a circle into slices.

- Exploded Pie Chart: Similar to the pie chart, but with one or more slices pulled away for emphasis.
- 3D Pie Chart: Displays a three-dimensional view of the pie chart.

e. Area Charts:

- Area Chart: Shows the magnitude of change over time and emphasises the total value across a trend.
- Stacked Area Chart: Similar to the area chart, but with multiple areas stacked on top of each other.

f. Scatter Charts:

- Scatter Chart: Displays relationships between two sets of values as individual data points on a graph.
- Bubble Chart: Extends the scatter chart by adding a third variable that determines the size of the data points.

- Charts represent data in various formats

g. Stock Charts:

- High-Low-Close Chart: Used to analyse stock market data, showing the high, low, and closing prices.
- Open-High-Low-Close Chart: Similar to the high-low-close chart, but includes the opening price.

h. Surface Charts:

- 3D Surface Chart: Represents data in a three-dimensional format, where the X and Y axes represent input values, and the Z axis represents the dependent variable.

i. Radar Charts:

- Radar Chart: A two-dimensional chart that has multiple axes extending from a central point that represents multivariate data.

j. Combo Charts:

- Combo Chart: Combines different chart types, allowing you to display multiple data series using different chart types on the same graph.

These are the main chart types available in Excel. Each chart type has its own specific purpose and is suited for different data presentations and analysis requirements. Excel provides a range of customisation options for each chart type, allowing you to format and personalise the charts to meet your specific needs.

1.3.2.2 Creating a Chart in Excel

To create a chart in Excel, follow these steps:

1. Select the data you want to turn into a chart, including both the labels and values.
2. Click on the "Insert" tab in the ribbon.
3. Click on the desired chart type in the "Charts" group. Excel offers various chart types, such as column charts, bar charts, pie charts, bar charts, line charts, scatter plots, area charts etc. This is shown below in the Fig 1.3.4

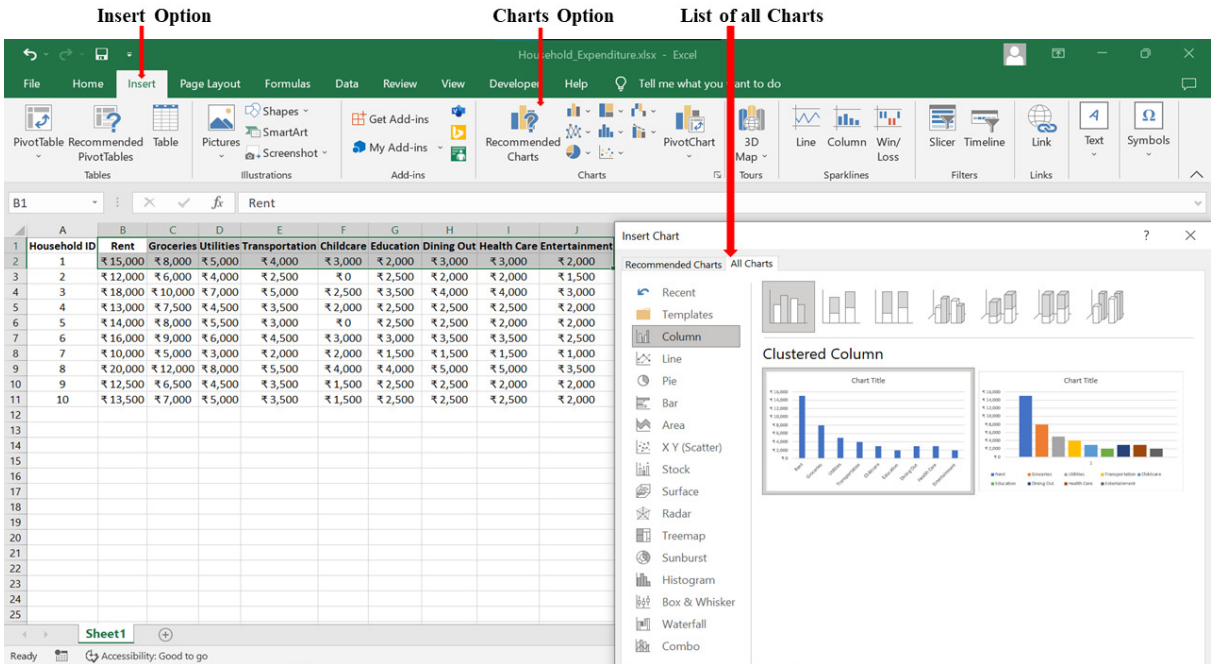


Fig 1.3.4. Creating Chart

Excel will generate a basic chart using the selected data. You can further customise the chart by modifying its design, lay-out, and data series.

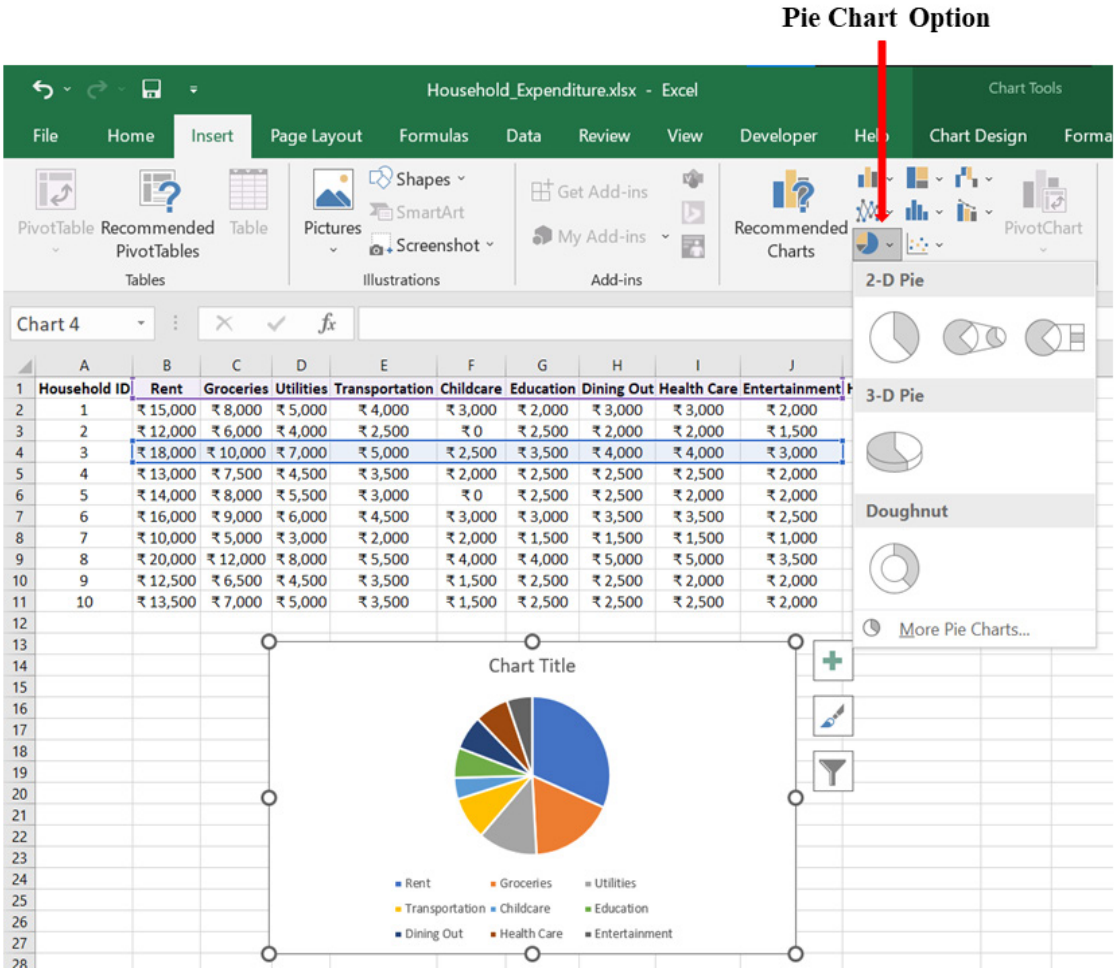
Let us use the "Household_Expenditure" dataset and create a pie chart to analyse the expenditure distribution of Household ID 3.

The steps to find the result are as follows;

1. Select the titles for creating the chart, here they are the expenditure column headings.

- Select the data go to the “Insert” tab, choose the desired chart type from the “Charts” group

2. Press Control key and select the data for creating the chart
3. Click on Insert ribbon and either select Pie Chart op tion from that or
4. Click on recommended charts and select Pie Chart from that
5. Here we have opted for the basic 2D option.
6. The result will be as shown in the Fig 1.3.5



Household ID	Rent	Groceries	Utilities	Transportation	Childcare	Education	Dining Out	Health Care	Entertainment
1	₹ 15,000	₹ 8,000	₹ 5,000	₹ 4,000	₹ 3,000	₹ 2,000	₹ 3,000	₹ 3,000	₹ 2,000
2	₹ 12,000	₹ 6,000	₹ 4,000	₹ 2,500	₹ 0	₹ 2,500	₹ 2,000	₹ 2,000	₹ 1,500
3	₹ 18,000	₹ 10,000	₹ 7,000	₹ 5,000	₹ 2,500	₹ 3,500	₹ 4,000	₹ 4,000	₹ 3,000
4	₹ 13,000	₹ 7,500	₹ 4,500	₹ 3,500	₹ 2,000	₹ 2,500	₹ 2,500	₹ 2,500	₹ 2,000
5	₹ 14,000	₹ 8,000	₹ 5,500	₹ 3,000	₹ 0	₹ 2,500	₹ 2,500	₹ 2,000	₹ 2,000
6	₹ 16,000	₹ 9,000	₹ 6,000	₹ 4,500	₹ 3,000	₹ 3,000	₹ 3,500	₹ 3,500	₹ 2,500
7	₹ 10,000	₹ 5,000	₹ 3,000	₹ 2,000	₹ 2,000	₹ 1,500	₹ 1,500	₹ 1,500	₹ 1,000
8	₹ 20,000	₹ 12,000	₹ 8,000	₹ 5,500	₹ 4,000	₹ 4,000	₹ 5,000	₹ 5,000	₹ 3,500
9	₹ 12,500	₹ 6,500	₹ 4,500	₹ 3,500	₹ 1,500	₹ 2,500	₹ 2,500	₹ 2,000	₹ 2,000
10	₹ 13,500	₹ 7,000	₹ 5,000	₹ 3,500	₹ 1,500	₹ 2,500	₹ 2,500	₹ 2,500	₹ 2,000

Fig 1.3.5 Pie chart

These are the steps to create a chart in Excel. The user-friendly interface allows for the seamless selection and customisation of various chart types, enabling effective communication of trends, patterns, and comparisons within datasets.

1.3.2.3 Alternative Method of Creating Charts

When creating a chart in Excel without initially selecting data, you can follow an alternative method that involves inserting a chart and then selecting the series and data. Let us now learn the step-by-step process of doing the same

1. **Insert a Blank Chart:**
   - Open your Excel worksheet and navigate to the "Insert" tab in the Ribbon.
   - Click on "Chart" in the Charts group.
   - Choose a chart type from the options presented (e.g., Bar, Line, Pie) or select "More Options" to explore additional types.
2. **A Blank Chart Appears:**
   - Excel will insert a blank chart into your worksheet.
3. **Add Data to the Chart:**
   - Click on the chart to select it.
   - Go to the "Chart Design" or "Chart Tools Design" tab that appears on the Ribbon.
   - Click on "Select Data" in the Data group.

- Alternative method for creating charts

4. **Edit Series and Categories:**
   - In the "Select Data Source" dialog box, you can see the current data series on the left.
   - Click on the "Add" button to add a new series. Enter the series name and select the data range.
   - You can also edit existing series or categories by selecting them and clicking "Edit."
5. **Adjust Axis Labels:**
   - In the same "Select Data Source" dialog box, you can modify axis labels by clicking on the "Edit" button next to Horizontal (Category) Axis Labels or Vertical (Value) Axis Labels.
6. **Finalise and Confirm:**
   - Once you have added, edited, or adjusted the series and data, click "OK" to close the "Select Data Source" dialog box.
7. **Format the Chart:**
   - With the chart still selected, use the Chart Tools tabs (e.g., "Chart Design," "Chart Format") to format your chart as needed. Adjust titles, colours, and other elements.
8. **Additional Editing:**
   - Right-click on different chart elements (e.g., data

points, axis labels) to access additional formatting and editing options.

This alternative method allows you to start with a blank chart and then define or modify your data series, providing flexibility in chart creation even if you have not pre-selected the data.

1.3.2.4 Customising Charts in Excel

- Inserting chart without selecting data

Excel charts can be made interactive by adding features like data labels, data tables, and chart filters. These interactive elements enable users to explore the data further, drill down into specific values, or filter the chart to focus on specific data subsets.Once a chart is created, you can easily update the underlying data. Simply select the chart, go to the "Design" or "Chart Tools" tab, and use the "Select Data" option to modify the data range. In Excel, you can make various changes and modifications to charts to customise their appearance and enhance their visual impact. The common actions to modify charts are as follows :

1. **Selecting and Editing Charts:**
   a. Click on the chart to select it. The chart elements will be highlighted, and you can make changes specific to the selected element.
   b. Double-click on a specific element, such as a data series or axis, to make changes to that element directly.
   c. Right-click on the chart to access a context menu with additional options for editing and formatting.
2. **Changing Chart Type:**
   a. Select the chart and go to the "Design" or "Chart Design" tab in the Excel ribbon.
   b. Click on the "Change Chart Type" button to choose a different chart type for your data.
   c. Select the desired chart type and customise its settings as needed.
3. **Modifying Chart Data:**
   a. To update the data range used in a chart, click on the chart, and then click on the "Select Data" button in the Excel ribbon.
   b. In the "Select Data Source" dialog box, you can modify the series names, category (X) axis labels, and the data range for each series.

4. **Formatting Chart Elements:**
   a. Select a specific chart element, such as data series, axis, title, or legend, and right-click to access formatting options.
   b. You can change colours, fonts, borders, sizes, and other formatting properties to customise the appearance of the chart.

- Common modifications made in a chart

5. **Adding and Removing Chart Elements:**
   a. Use the "Chart Elements" or "Add Chart Element" button in the Excel ribbon to add or remove specific chart elements like data labels, axis titles, gridlines, and trendlines.
   b. Check or uncheck the desired elements to show or hide them in the chart.
6. **Adjusting Chart Layout:**
   a. Use the "Chart Layouts" or "Quick Layout" button in the Excel ribbon to apply predefined layouts to your chart, including different combinations of titles, legends, and axes.
   b. Customise the layout further by dragging elements within the chart to reposition them.
7. **Applying Chart Styles:**
   a. Select the chart and go to the "Design" or "Chart Design" tab in the Excel ribbon.
   b. Use the "Chart Styles" or "Quick Styles" button to apply different predefined styles to your chart, which can change colours, fills, and other visual properties.
8. **Adding Chart Titles and Labels:**
   a. Click on the chart and go to the "Layout" or "Chart Layout" tab in the Excel ribbon.
   b. Use the "Chart Title" or "Axis Titles" button to add or edit titles for the chart, horizontal axis, and vertical axis.
   c. You can also add data labels to the chart to display specific values or other information.

- Customise Charts by selecting and editing chart elements

These are just some of the ways you can make changes and modifications to charts in Excel. The specific options available may vary depending on the version of Excel you are using, but the general principles remain the same. Experiment with dif-

ferent customisation options to create charts that effectively communicate your data and insights

In the Fig 1.3.5 we can see that it is difficult for a person to analyse the pie chart without checking the dataset. So, in order to make the pie chart more appealing we make use of the modifying options available in Excel. Let us add a new chart title and add data labels to the existing Pie Chart in the Fig 1.3.5

The steps for that are as follows:

- Click on the Chart title available inside the Pie Chart and edit it. We are renaming it as "Expenditure _HH3"
- Click on the ✚ sign. This is the Chart elements option.
- Select the Data labels option from that. In our example, Chart title and Legend are already selected by default. But in some cases, we will have to select those options too.

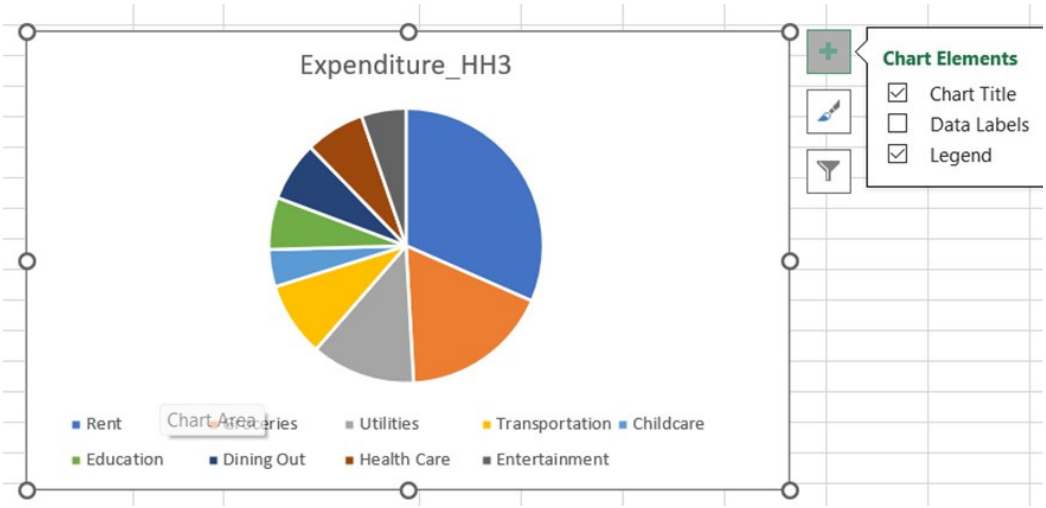


Fig 1.3.6 Modifying Data labels

- Now we can see that data labels are visible on the Pie Chart
- The results will be as shown in the Fig 1.3.7

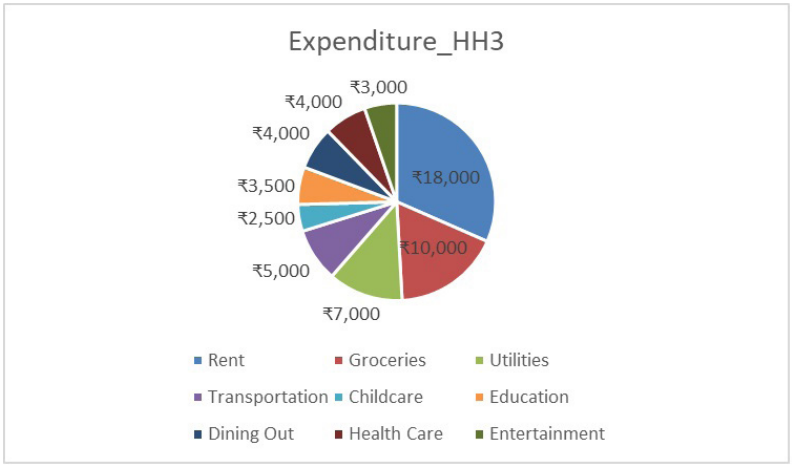


Fig 1.3.7 Modified Pie Chart

Creating an effective chart in Excel goes beyond just plotting data points; it involves enhancing the visual presentation through customisation. Legends and data labels play crucial roles in explaining information and facilitating comprehension. Additionally, features like leader lines, value labels in percentage, and connecting lines for scatter plots contribute to the clarity and communicative power of the chart. Understanding how to manipulate and customise these elements is crucial for tailoring charts to specific needs and optimising the visual impact of the data. They are:

- Enables editing elements

1. **Legend:** The legend serves as a key to deciphering the various data series or categories represented in a chart. Positioned at the top, bottom, left, or right of the chart, the legend is customizable in terms of location, font size, and style. By right-clicking on the legend, one can easily modify its appearance or remove it altogether, allowing for a clean and focused presentation of the data.
2. **Data Labels:** Data labels provide specific information about individual data points on the chart, such as precise values or percentages. Adding data labels is a straightforward process: click on the chart, go to the "Chart Elements" button, and check the "Data Labels" box. Customization options abound, accessible by right-clicking on a data label and selecting "Format Data Labels." Here, one can tailor the display to showcase values, percentages, or both, adjusting font size, color, and positioning for optimal clarity.
3. **Leader Lines:** Leader lines are invaluable for connecting data labels to their corresponding points, particularly when labels are situated outside data markers. By right-clicking on a data label, navigating to "Format Data Labels," and enabling the "Leader Lines" option, one can enhance the readability of the chart. The length and style of leader lines can be adjusted to suit the aesthetic preferences and visual hierarchy of the chart.
4. **Value Labels in Percentage:** For charts involving percentage data, directly displaying these values can offer valuable insights. To achieve this, right-click on a data label after adding them, choose "Format Data Labels," and select the option for displaying percentages. This feature provides a clear representation of proportional contributions within the data set, aiding in the audience's comprehension.
5. **Positioning Data Labels:** Ensuring the strategic place-

ment of data labels is crucial for chart readability. Customization options, available through the "Format Data Labels" menu, enable adjustments in positioning, rotation, and precise label placement. This flexibility allows for a tailored approach, ensuring that data labels are optimally situated for clarity and visual appeal.

6. **Connecting Lines for Scatter Plots:** In scatter plots, the addition of connecting lines between data points aids in visualizing relationships or trends. Achieve this by clicking on the chart, navigating to "Chart Elements," checking "Data Labels," and selecting "Connected Lines." This feature is particularly useful for highlighting patterns or sequences within the data set, contributing to a more comprehensive understanding.

- Customisation of legends and data labels

7. **Format Painter for Consistency:** Maintaining consistency across data labels is crucial for a polished and professional chart. The Format Painter tool facilitates this by allowing users to copy the formatting of one label and apply it to others. By clicking on a data label with the desired format, selecting the Format Painter tool, and clicking on other labels, a uniform appearance is ensured, contributing to the overall coherence of the chart.

The customisation options available in Excel helps to make the data visualisation easier and more appealing and understanding even to the common man.

1.3.3 Introducing Pivot Tables and Pivot Charts

Pivot tables and pivot charts are powerful tools that allow you to analyse and summarise large sets of data quickly and efficiently. They help you gain insights and make data-driven decisions.

1.3.3.1 Pivot Tables

Pivot tables provide a way to summarise, analyse, and manipulate data from a larger dataset. With a pivot table, you can extract key information, such as totals, averages, counts, or percentages, from a dataset based on different criteria.

Pivot tables have four main components:

1. **Rows:** These represent the categories or groups by which you want to organise your data.
2. **Columns:** These define the different data attributes or variables you want to analyse.
3. **Values:** These are the data points that you want to sum-

marise or calculate using functions like sum, count, average, etc.

4. **Filters:** These allow you to apply specific criteria to include or exclude data from the pivot table.

- Flexible and interactive way to summarise and analyse data

Pivot tables are interactive, and you can easily modify their structure and layout by dragging and dropping fields to different areas. They also support filtering, sorting, and drilling down into details for more in-depth analysis.

1.3.3.2 Pivot Charts

Pivot charts are visual representations of data from a pivot table. When you create a pivot chart, it is linked to the underlying pivot table, and any changes you make to the pivot table will be reflected in the chart. Pivot charts provide a dynamic and interactive way to visualise and explore your data. You can choose from various chart types, such as column charts, line charts, pie charts, etc., to represent your data in a meaningful way. Similar to regular charts in Excel, you can customise the appearance, formatting, and layout of pivot charts to suit your needs.

- Linked to pivot tables and provide dynamic and interactive visual representations of data

Pivot charts also support filtering and drilling down into details, allowing you to interactively explore the data behind the chart

The main uses and benefits of pivot tables and pivot charts in Excel are as follows:

a. **Data Summarisation and Analysis:**

- Pivot tables allows summarising large datasets by grouping and aggregating data based on different criteria, such as categories, dates, or regions. This provides a quick overview of the data and allows for easy identification of patterns and trends.
- Pivot tables enable you to perform calculations and analysis on the summarised data, such as calculating totals, averages, percentages, or other custom formulas.
- Pivot charts visually represent the summarised data, making it easier to understand and interpret complex information. They help in identifying key insights and presenting data in a visually appealing manner.

b. **Data Exploration and Visualisation:**

- Pivot tables and pivot charts provide interactive tools for exploring and analysing data. You can easily change the arrangement of fields, apply filters, and drill down

into details to gain a deeper understanding of the data.

- Pivot charts offer a variety of chart types, such as column charts, pie charts, line charts, etc., allowing you to visualise data in different ways and identify patterns or anomalies more effectively.
- Interactivity features in pivot tables and pivot charts enable you to dynamically update and manipulate data, providing real-time insights into changing datasets.

c. Report Generation and Presentation:

- Pivot tables and Pivot charts are powerful tools for creating reports and presentations. They allow you to summarise and analyse data from various angles, making it easier to communicate findings and present information to stakeholders.
- With Pivot tables, you can generate cross-tabulated reports that provide a comprehensive view of the data, allowing users to analyse data based on different dimensions and variables.
- Pivot charts enhance the visual appeal of reports and presentations, enabling stakeholders to grasp information quickly and make data-driven decisions.

- Generates comprehensive reports and presentations

d. Data Cleansing and Quality Assessment:

- Pivot tables can be used to assess data quality and identify inconsistencies, missing values, or outliers in datasets.
- By summarising data and performing calculations, you can identify data discrepancies or anomalies that require further investigation and data cleansing.

e. Business Intelligence and Decision Making:

- Pivot tables and pivot charts are valuable for business intelligence purposes. They help in identifying key performance indicators (KPIs), tracking business metrics, and monitoring progress toward goals.
- The ability to analyse data from different perspectives and dimensions allows for deeper insights and informed decision making.
- Pivot tables and pivot charts provide a user-friendly and interactive way to explore data, empowering individuals, and organisations to make data-driven decisions.

Overall, Pivot tables and Pivot charts are essential tools for data analysis, exploration, and visualisation. They facilitate efficient data summarisation, provide valuable insights, and

assist in effective decision making.

The step-by-step guide on how to create a pivot table in Excel are as follows

Step 1: Prepare Your Data

- Ensure that your data is organised in a tabular format with column headers.
- Include a header row with labels for each column.
- Avoid blank rows or columns within your data range.

Step 2: Select Your Data Range

- Click anywhere within your data range.
- Drag and select the entire range of your data, including the header row.

Step 3: Insert the Pivot Table

- Go to the "Insert" tab in the Excel ribbon.
- Click on the "PivotTable" button. A dialog box will appear.

Step 4: Choose the Data Range

- In the PivotTable dialog box, make sure the "Select a table or range" option is selected.
- Verify that the correct data range is displayed in the "Table/Range" field. Excel usually auto-detects the range based on your selection in Step 2.
- If needed, you can manually adjust the data range by typing the range or using the "Select Table" button.

Step 5: Select the Destination

- Choose whether you want to place the pivot table on a new worksheet or an existing worksheet.
- If you select an existing worksheet, click on the cell where you want the top-left corner of the pivot table to be placed.

Step 6: Design Your Pivot Table

- The PivotTable Field List will appear on the right side of the Excel window.
- This list shows the column headers from your data range.
- Drag and drop the desired field names into the four areas of the PivotTable Field List: Rows, Columns, Values, and Filters.

Step 7: Define the Pivot Table Structure

- Drag the field names from the field list into the desired areas of the PivotTable Field List:
- Rows: Determines the rows or categories in your pivot table.
- Columns: Defines the columns or subcategories in your pivot table.
- Values: Specifies the data you want to summarise or analyse using functions like sum, count, average, etc.
- Filters: Allows you to apply specific criteria to include or exclude data from the pivot table.

Step 8: Customise the Pivot Table

- You can modify the pivot table by dragging and dropping fields between areas in the PivotTable Field List.
- To change the summary function for a value field, click on the drop-down arrow next to the field name in the "Values" area and choose the desired calculation.
- You can also format the pivot table by right-clicking on elements within the pivot table, accessing the context menu, and selecting formatting options.

- Highly customisable, allowing you to explore your data from different angles and gain valuable insights

Step 9: Refresh the Pivot Table

- If your data changes or you add new data, you need to refresh the pivot table to update it.
- Right-click anywhere within the pivot table and choose "Refresh" from the context menu, or go to the "Pivot Table Tools" tab and click on the "Refresh" button.

By following these steps, you can create a pivot table in Excel and begin summarising and analysing your data in a dynamic and interactive way.

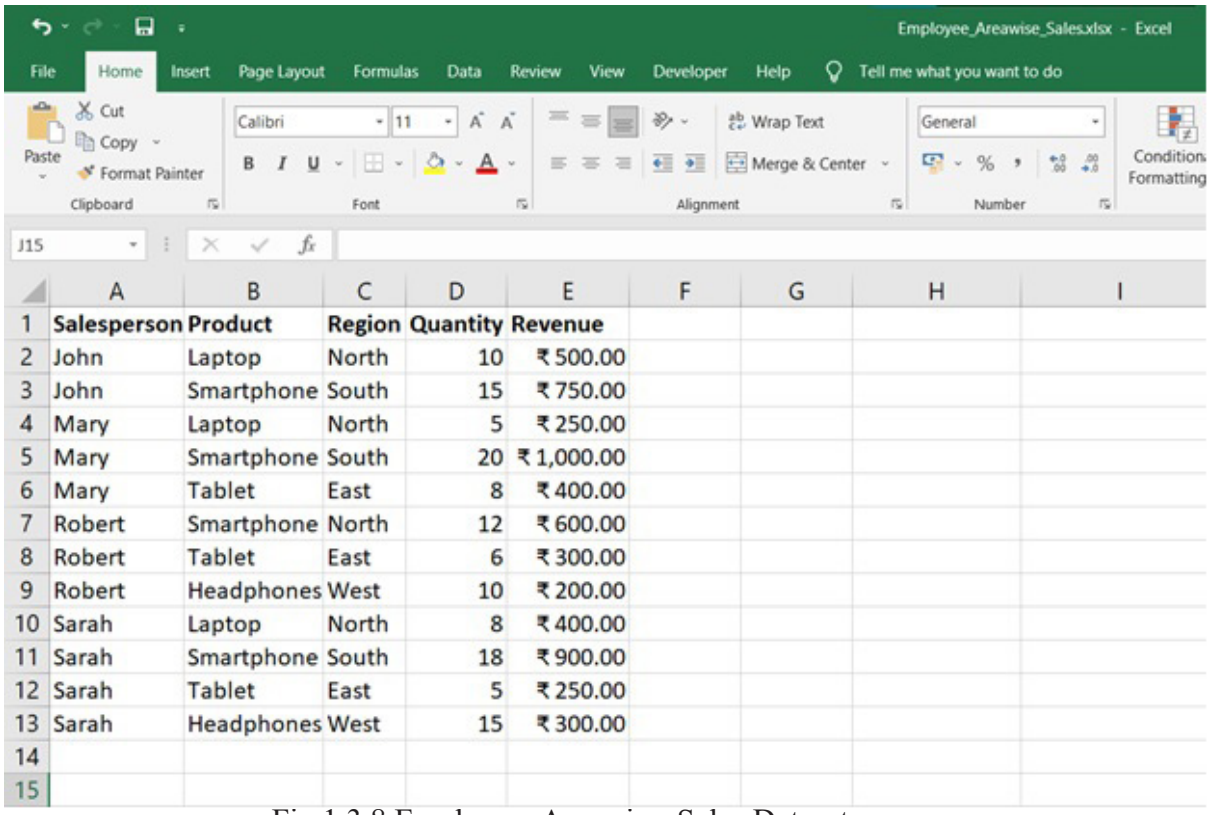
The pivot chart provides a visual representation of the summarised pivot table data. Here is an example of a data set that you can use to create pivot tables. In this data set, we have information about salesperson, the products they sold, the region of the sale, the quantity sold, and the revenue generated.

Table1.3.1 Dataset for Pivot table

Salesperson	Product	Region	Quantity	Revenue
John	Laptop	North	10	₹ 500.00
John	Smartphone	South	15	₹ 750.00
Mary	Laptop	North	5	₹ 250.00
Mary	Smartphone	South	20	₹ 1,000.00
Mary	Tablet	East	8	₹ 400.00
Robert	Smartphone	North	12	₹ 600.00
Robert	Tablet	East	6	₹ 300.00
Robert	Headphones	West	10	₹ 200.00
Sarah	Laptop	North	8	₹ 400.00
Sarah	Smartphone	South	18	₹ 900.00
Sarah	Tablet	East	5	₹ 250.00
Sarah	Headphones	West	15	₹ 300.00

The steps to create a Pivot Table are as follows:

1. Enter the following dataset into excel sheet and save it as "Employee_Areawise_Sales"
2. It will be as shown in the Fig 1.3.8
3. Click on Insert ribbon
4. Select Pivot tables and choose the From Table/Range option as shown in the Fig 1.3.9



	A	B	C	D	E	F	G	H	I
1	**Salesperson**	**Product**	**Region**	**Quantity**	**Revenue**				
2	John	Laptop	North	10	₹ 500.00				
3	John	Smartphone	South	15	₹ 750.00				
4	Mary	Laptop	North	5	₹ 250.00				
5	Mary	Smartphone	South	20	₹ 1,000.00				
6	Mary	Tablet	East	8	₹ 400.00				
7	Robert	Smartphone	North	12	₹ 600.00				
8	Robert	Tablet	East	6	₹ 300.00				
9	Robert	Headphones	West	10	₹ 200.00				
10	Sarah	Laptop	North	8	₹ 400.00				
11	Sarah	Smartphone	South	18	₹ 900.00				
12	Sarah	Tablet	East	5	₹ 250.00				
13	Sarah	Headphones	West	15	₹ 300.00				
14									
15									

Fig 1.3.8 Employee_Areawise_Sales Dataset

5. A dialogue box will pop up and in the select a table or range option

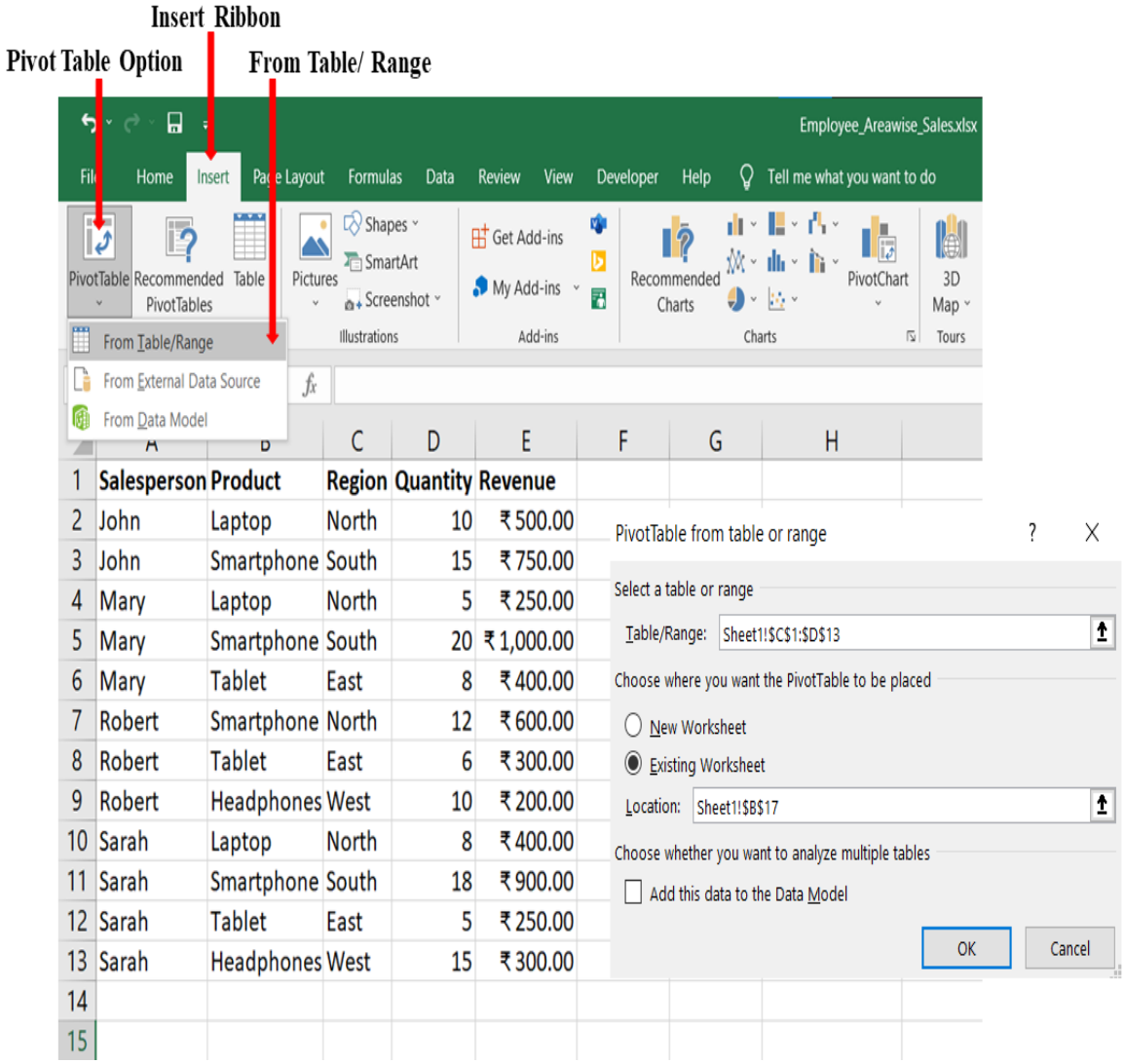


Fig 1.3.9 Making Pivot Table

6. To select data or range, press the control key and select the necessary columns together.
7. Here we are trying to find Region wise quantity sale. So, here we have selected Region and Quantity columns.
8. In the next option we choose the Existing worksheet option
9. And location is specified as B17 cell in this worksheet itself. In order to select the location, click on the blank area near location option and click on desired cell
10. Then click OK

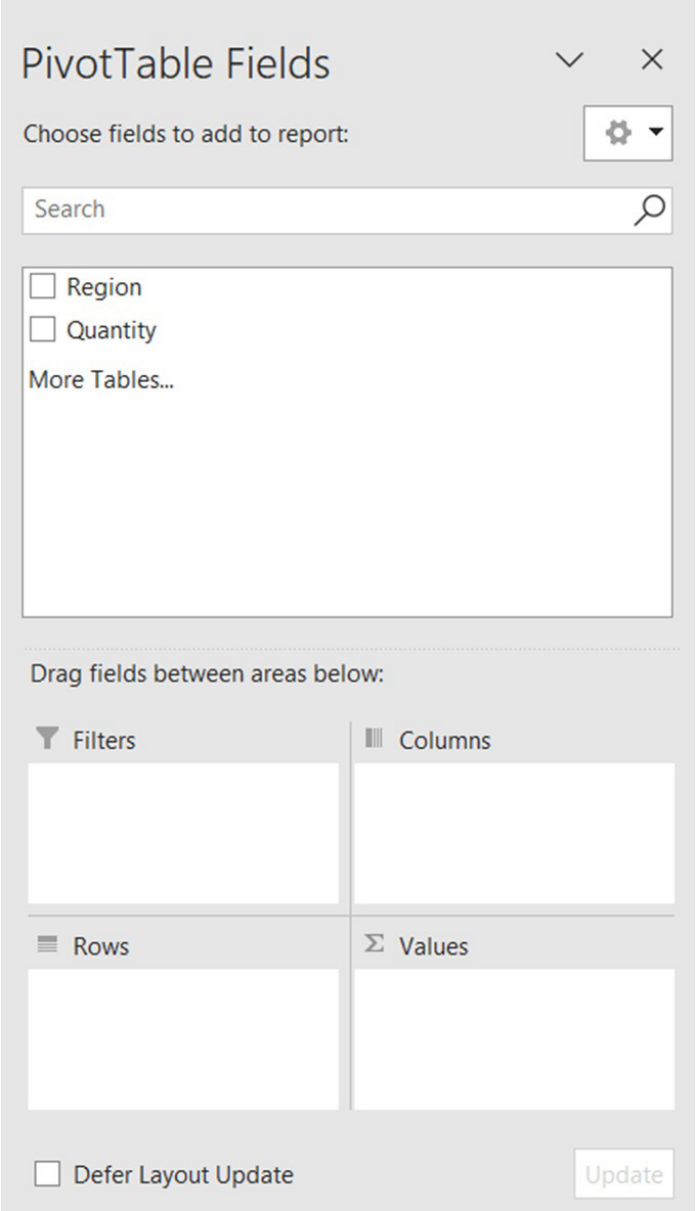


Fig 1.3.10 Pivot Table Fields

11. Then a PivotTable Fields will open up on the right-hand side as shown in the Fig 1.3.10
12. Select Region and Quantity from the options
13. A new table will appear on B17 and that is the Pivot Table

It will be as shown in the Fig 1.3.11

Row Labels	Sum of Quantity
East	19
North	35
South	53
West	25
Grand Total	**132**

Fig 1.3.11 Pivot Table

Multiple Tables can be created with the same set of data.

For example, total revenue by salesperson, total revenue by region and product, total quantity sold by product and region, average revenue per sale by salesperson and product etc can be calculated using the same data set.

Now let us see the steps to create a Pivot Chart from a Pivot Table :

Step 1: Create a Pivot Table

- Ensure that you have already created a pivot table in Excel with the desired data summarised and organised. If you have not created a pivot table yet, refer to the previous explanation for creating a pivot table.

Step 2: Select the Pivot Table

- Click anywhere inside the pivot table to select it. This indicates that you want to create a pivot chart based on the data in the pivot table.

Step 3: Insert the Pivot Chart

- With the pivot table selected, go to the "Insert" tab on the Excel ribbon at the top of the window.

Step 4: Choose the Chart Type

- In the "Charts" group, you will find various chart types. Click on the desired chart type that best represents your data. For example, you can choose a column chart, bar chart, line chart, pie chart, etc. A dropdown menu will appear with different chart subtype options.

Step 5: Select the Specific Chart Subtype

- From the dropdown menu, select the specific subtype of the chart you want to create by clicking on it. Excel will

generate a default pivot chart based on your selection and insert it into the worksheet alongside the pivot table. The chart will be linked to the pivot table, so any changes you make in the pivot table will automatically update the chart.

Step 6: Customise the Pivot Chart

- You can modify the pivot chart's appearance, formatting, and layout using the chart tools available in Excel. The chart tools will appear on the ribbon when the pivot chart is selected.
- Customise the chart title, axis labels, formatting, data series, legends, data labels, trendlines, and other chart options using the options provided in the chart tools.

- To create a pivot chart from a pivot table

Step 7: Refresh the Pivot Chart

- If your data or the underlying pivot table changes, you need to refresh the pivot chart to reflect those changes.
- Right-click anywhere within the pivot chart and choose "Refresh" from the context menu, or go to the "PivotChart Tools" tab and click on the "Refresh" button.

By following these steps, you should be able to create a pivot chart from a pivot table. The pivot chart will dynamically update as you make changes to the underlying pivot table, allowing you to visualise and analyse your data effectively.

With the same example of "Employee_Areawise_Sales" dataset and the pivot table we created, we can now create a pivot chart. The steps to create a Pivot Chart from a Pivot Table are as follows :

1. After creating the Pivot table, when we click on it, then a new Ribbon titled Pivot Table Analyse will appear as shown in the Fig 1.3.12
2. Click on Pivot Chart from Pivot Table Analyse ribbon
3. Now a list of all available charts will appear and select the suitable chart from the list.

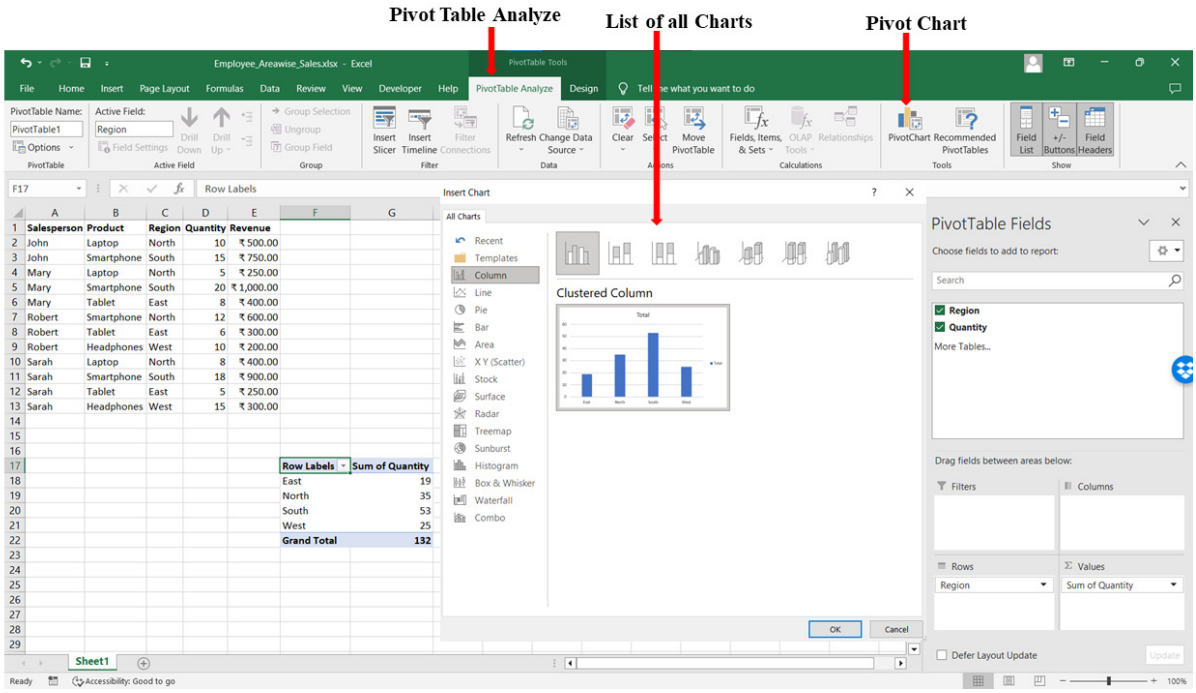


Fig 1.3.12 Inserting Pivot Chart

4.. Here we opted for column chart for visualisation

5. The result will be as shown in the Fig 1.3.13

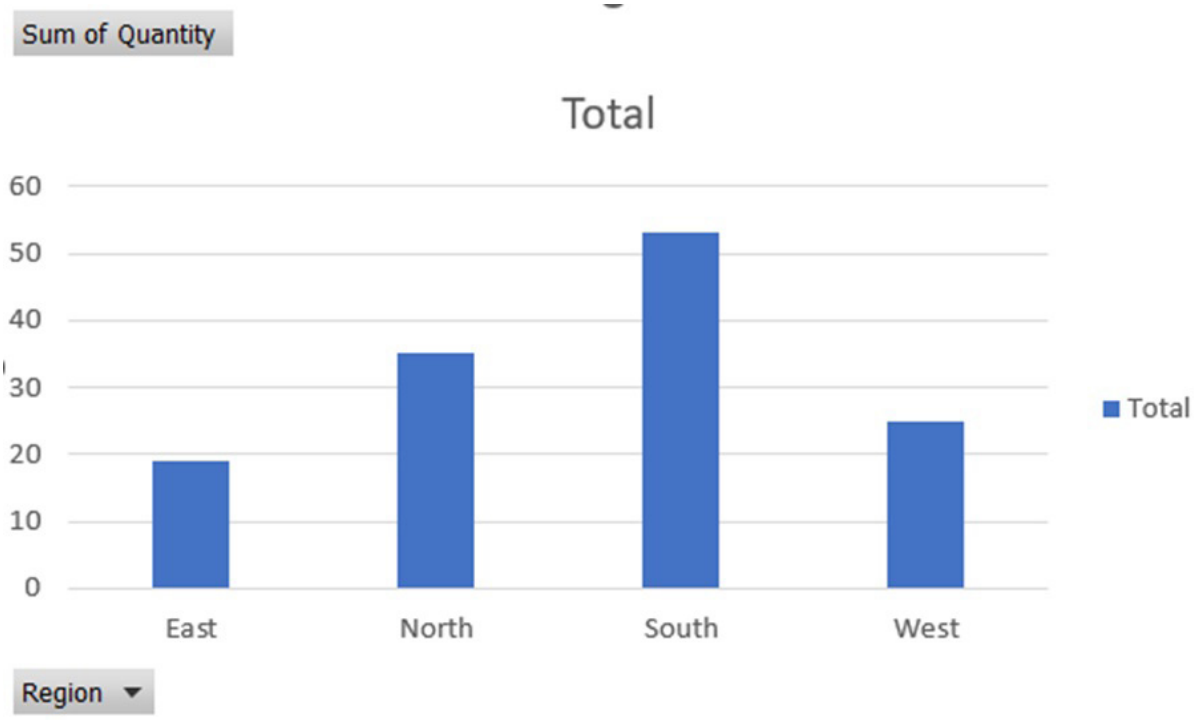


Fig 1.3.13 Pivot chart

Here we have more options to customise the chart and more details can be understood from pivot chart.

The combination of Pivot Tables and Pivot Charts provides a powerful toolkit for data analysis and visualisation. Pivot Tables offer a structured approach to organising and summarising data, while Pivot Charts convert this data into visually appealing and informative charts.

Summarised Overview

Conditional formatting in Excel is a powerful feature that allows you to apply formatting to cells based on specific conditions. It enhances data visualisation by highlighting data points, applying color scales, adding data bars and icon sets, and identifying duplicates or unique values. Conditional formatting helps to make data more visually appealing, easier to interpret, and facilitates the identification of patterns and trends. It provides flexibility in defining conditions and formatting styles, allowing users to customise and manage conditional formatting rules.

Charts in Excel are graphical representations of data that enable effective visualisation and analysis. Excel offers various chart types, such as column charts, line charts, pie charts, and scatter charts, each suited for different data types and analysis purposes. Charts help to present data in a visually engaging manner, making it easier to understand patterns, trends, and comparisons. Excel provides customisation options for charts, allowing users to modify their design, layout, and data series. By using charts, complex data can be presented in a visually appealing and easily understandable format, aiding effective communication of insights and analysis.

Pivot tables in Excel are powerful data summarisation and analysis tools that allow users to quickly analyse and extract meaningful information from large datasets. With pivot tables, users can easily rearrange, summarise, and manipulate data by dragging and dropping fields, creating custom calculations, and applying filters. Pivot tables enable users to gain insights, identify patterns, and perform data analysis tasks such as grouping, sorting, and calculating totals or averages. They provide a flexible and interactive way to explore data and generate reports without the need for complex formulas or coding.

Self-Assessment

1. Explain the concept of conditional formatting in data visualisation. Give an example of how it can be used to highlight specific data points in a spreadsheet.
2. Describe the steps involved in applying conditional formatting to a range of cells in Microsoft Excel. Provide an example of a scenario where you would use conditional formatting to visualise data effectively.
3. Discuss the process of creating and modifying charts in spreadsheet applications. Explain the key elements involved in creating a chart, and how you can customise its appearance to enhance data visualisation.
4. Define a pivot table and explain its purpose in data analysis. Provide an example of how pivot tables can help summarise and analyse large datasets effectively.
5. How can pivot charts enhance data visualisation when working with pivot tables? Explain the relationship between pivot tables and pivot charts, and provide an example of how a pivot chart can be created and customised.

Assignments

Using the dataset given below complete the assignment question.

Product	Category	Quarter	Sales
Smartphone	Electronics	Q1	100
T-Shirt	Clothing	Q1	150
Laptop	Electronics	Q1	80
Smartphone	Electronics	Q2	120
Jeans	Clothing	Q2	180
Headphones	Electronics	Q2	90
Smartphone	Electronics	Q3	80
Jacket	Clothing	Q3	200
Smartwatch	Electronics	Q3	70
Smartphone	Electronics	Q4	130

Dress	Clothing	Q4	160
Tablet	Electronics	Q4	110

a. Create a column chart to visualise the total sales for each quarter. Include the appropriate chart title, axis labels, and legend.

b. Apply conditional formatting to highlight cells with sales above 150 units

c. Create a pivot table to summarise the total sales by category and quarter.

d. Calculate the total sales for each category and determine the top three categories with the highest sales overall.

e. Based on the pivot table created in question c, create a pivot chart (e.g., bar chart, line chart, etc.) to visualise the sales performance of each category across quarters. Include an appropriate chart title, axis labels, and legend.

Reference

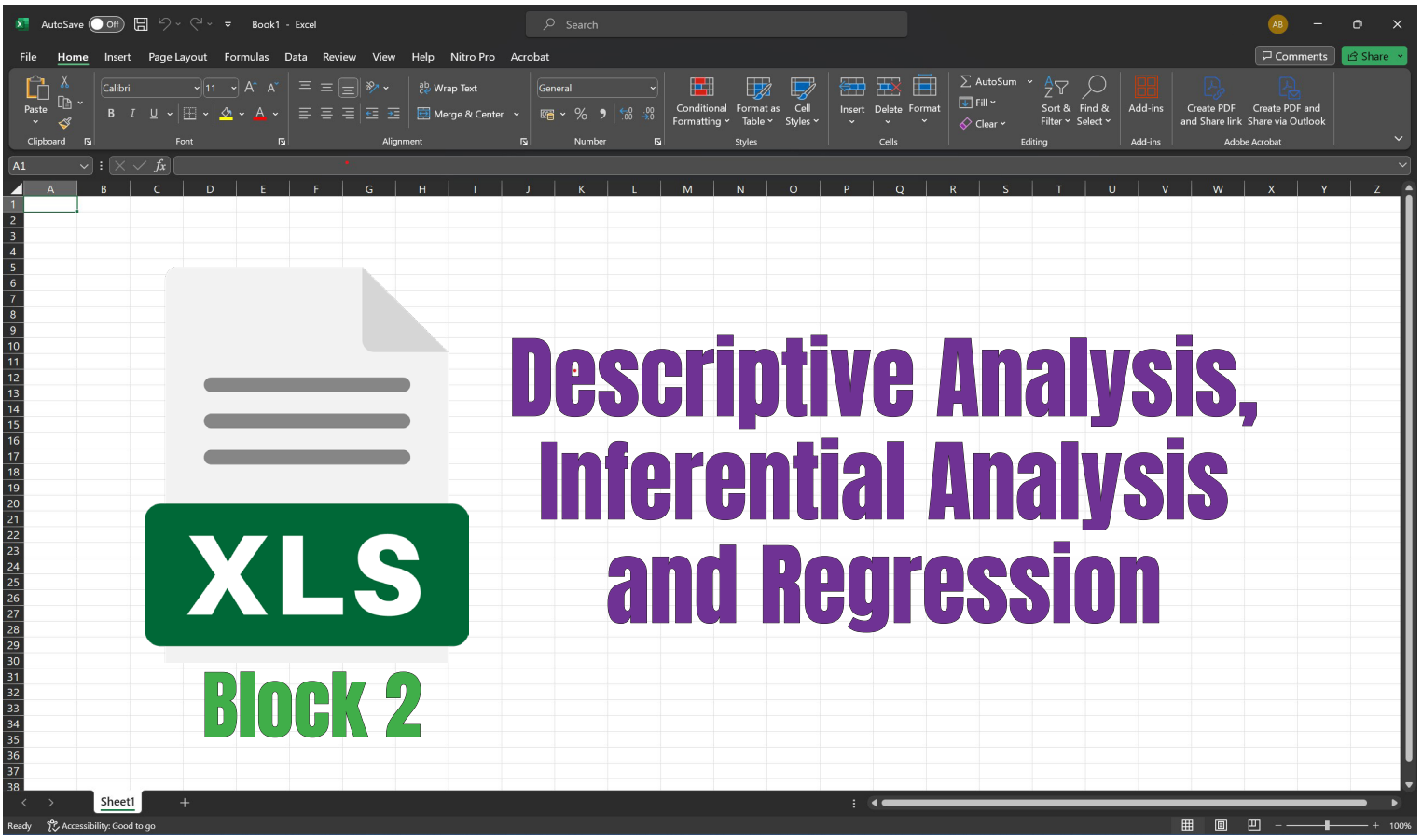
1. Alexander, M., Kusleika, R., & Walkenbach, J. (2018). *Excel 2019 Bible.* John Wiley & Sons.
2. Frye, C. (2018). *Microsoft Excel 2019 step by step.* Microsoft Press.
3. Harvey, G. (2018). *Excel 2019 for dummies.* John Wiley & Sons.
4. Bluttman, K. (2013). *Excel formulas and functions for dummies.* John Wiley & Sons.

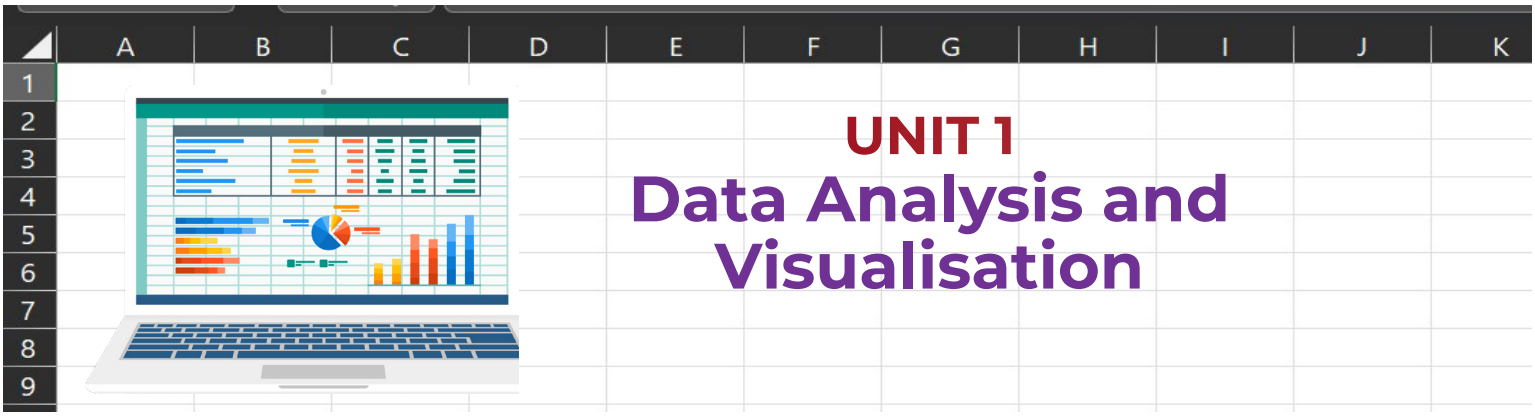
Suggested Reading

1. McFedries, P., & Harvey, G. (2021). *Excel All-in-one for Dummies*. John Wiley & Sons.
2. Jelen, B., & Juhasz, S. (2015). *MrExcel XL: The 40 Greatest Excel Tips of All Time.* Tickling Keys, Inc.
3. Jelen, B. (2015). *Power Excel with MrExcel: Master Pivot Tables, Subtotals, Charts, VLOOKUP, IF, Data Analysis in Excel 2010–2013*. Tickling Keys, Inc.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.


AutoSave
Book1 - Excel
Search
File
Home
Insert
Page Layout
Formulas
Data
Review
View
Help
Nitro Pro
Acrobat
Comments
Share
Paste
Calibri
11
Wrap Text
Merge & Center
General
Conditional Formatting
Format as Table
Cell Styles
Insert
Delete
Format
AutoSum
Fill
Clear
Sort & Filter
Find & Select
Add-ins
Create PDF and Share link
Create PDF and Share via Outlook
Clipboard
Font
Alignment
Number
Styles
Cells
Editing
Add-ins
Adobe Acrobat
A1
XLS
Block 2
Descriptive Analysis, Inferential Analysis and Regression
Sheet1
Ready
Accessibility: Good to go
100%



UNIT 1
Data Analysis and Visualisation

Learning Outcomes

After the completion of this unit, the learner will be able to

- generate descriptive statistics
- create advanced charts and graphs
- implement formatting options
- interpret and analyse data visually

Background

Data analysis is a fundamental process in the field of statistics and decision-making that involves the examination, interpretation, and transformation of raw data into meaningful insights and actionable information. In various industries and domains, data analysis plays a pivotal role in understanding patterns, trends, and relationships within datasets, aiding in informed decision-making and problem-solving. By harnessing the power of data analysis, organisations can make data-driven decisions, identify opportunities for improvement, and gain a competitive edge in today's data-driven world. When we analyse data, we use two main types of tools: descriptive statistics and inferential statistics. Descriptive statistics help us understand what the data looks like by summarising its main features, like the average or spread. On the other hand, inferential statistics explore how different things in the data are connected. Descriptive statistics describe the data, while inferential statistics help us make predictions and understand relationships between things we are looking at.

Data analysis is a powerful tool that empowers individuals and organisations to make evidence-based decisions, solve real-world problems, and uncover valuable insights from data. As data continues to grow in volume and complexity, the importance of data analysis will only continue to grow, enabling advancements in various fields and shaping the way we understand and interact with the world around us.

Keywords

Descriptive Statistics, Advanced Charts, Data Analysis, Formatting

Discussion

2.1.1 Descriptive Statistics

- Analyse data to provide a comprehensive summary

Descriptive statistics is a fundamental branch of statistics that encompasses various processes, starting from the collection and organisation of data to its analysis, interpretation, and presentation. The ultimate goal of descriptive statistics is to provide a comprehensive summary of the main features of a dataset, allowing analysts and researchers to gain valuable insights and understand the underlying patterns and characteristics of the data. The process of descriptive statistics begins with data collection, where relevant data is gathered from various sources or experiments. Once the data is collected, it undergoes organisation, which involves arranging it in a structured format that facilitates analysis. Organising data may include sorting, grouping, and creating data tables or frequency distributions. Next comes the analysis stage, where various statistical measures are computed to summarise the dataset. Graphical representations are an integral part of descriptive statistics. Visualisations, such as histograms, bar charts, line graphs, and scatter plots, offer a powerful way to present data in a more intuitive and easily interpretable manner.

- Provide a clear overview for initial data exploration

Descriptive statistics is particularly useful during the initial stages of data exploration, as it provides analysts with a clear overview of the dataset before engaging in more complex analyses or hypothesis testing. It allows researchers to identify patterns, trends, outliers, and any unusual characteristics present in the data. Importantly, descriptive statistics should not be confused with inferential statistics. While descriptive statistics summarises and describes the observed data, inferential statistics involves drawing conclusions or making predictions about a larger population based on a sample. Descriptive statistics focuses solely on the sample data and does not involve making any generalisations or drawing inferences beyond the analysed dataset. Descriptive statistics can be divided into three main categories: measures of central tendency, measures of variability, and measures of shape.

Measures of Central Tendency

- Describes central value of a dataset

Measures of central tendency describe the typical or central value of a dataset. The three commonly used measures of central tendency are:

- **Mean:** The mean, also known as the average, is calculated by summing up all the values in the dataset and dividing the sum by the total number of values. It represents the balance point of the data and is sensitive to extreme values.
- **Median**: The median is the middle value in a dataset when it is arranged in ascending or descending order. If there is an even number of values, the median is the average of the two middle values. The median is often used when there are outliers or when the data is not symmetrically distributed.
- **Mode:** The mode is the value that occurs most frequently in a dataset. A dataset can have multiple modes (multimodal) or no mode if all values occur with equal frequency. The mode is useful for categorical or discrete data.

Measures of Variability

- Describe how data points spread from the central tendency

Measures of variability, also known as measures of dispersion, describe the spread or variability of the data points. They provide information about how much the values differ from the central tendency. Some common measures of variability include:

- **Range:** The range is the difference between the maximum and minimum values in a dataset. It provides a simple measure of the spread but can be influenced by extreme values.
- **Variance:** The variance measures the average squared deviation from the mean. It considers all values in the dataset and gives an indication of how much the data points are spread out.
- **Standard Deviation:** The standard deviation is the square root of the variance. It represents the average distance between each data point and the mean. It is a widely used measure of dispersion and is helpful in understanding the spread of data.

Measures of Shape

Measures of Shape refer to statistical indicators that describe the shape or form of a probability distribution or dataset. These

measures help us understand the pattern and characteristics of the data distribution, providing insights into its symmetry or asymmetry. Measures of shape include:

- **Skewness:** Skewness is a measure of the asymmetry of the data distribution. It indicates the direction and degree to which the data deviates from a symmetric distribution.
- **Kurtosis:** Kurtosis is a measure of the "tailedness" of the distribution, indicating the presence of outliers or extreme values in the dataset. It assesses how heavily the data is concentrated in the tails relative to the center of the distribution. Kurtosis can provide insights into whether the dataset contains outliers or extreme observations.

- Describe the symmetry and tail behavior of data distributions

Overall, descriptive statistics allow us to summarise and describe the essential characteristics of a dataset, providing insights and understanding of the data at hand. It forms the foundation for further statistical analysis and helps in making informed decisions in various fields, such as research, business, and public policy.

2.1.2 Generating Descriptive Statistics using Excel Functions

Generating descriptive statistics using Excel is a straightforward process. Excel provides a variety of functions that allow you to quickly calculate basic descriptive statistics for a set of data. There are various ways to generate descriptive statistics in Excel such as :

2.1.2.1 Excel Formulas

The step-by-step guide on generating descriptive statistics using Excel formulas is as follows:

1. Start by entering your data into an Excel spreadsheet. Ensure that each column represents a variable and each row represents an observation or data point.
2. Click and drag to select the range of data you want to analyse.
3. There are several functions in Excel that can help you calculate different descriptive statistics. Here are the most commonly used functions:
4. In an empty cell, enter the appropriate function for the

Table 2.1.1 Descriptive Statistics Functions & Formula

Descriptive Statistics	Excel Formula
Mean: The average of the data points.	AVERAGE(range)
Median: The middle value of the data when arranged in ascending or descending order.	MEDIAN(range)
Mode: The most frequently occurring value in the data.	MODE(range)
Standard deviation: A measure of how spread out the data is from the mean.	STDEV(range)
Variance: The square of the standard deviation.	VAR(range)
Minimum: The smallest value in the data.	MIN(range)
Maximum: The largest value in the data.	MAX(range)
Range	MAX(range)–MIN(range)
Count: The number of data points in the range	COUNT(range)
Skewness: Measures the asymmetry of the data distribution.	SKEW(range)
Kurtosis: Measures the peakedness or flatness of the data distribution compared to a normal distribution.	KURT(range)

statistic you want to calculate, replacing "range" with the actual data range you selected. Press Enter to get the result.

5. If you want to calculate multiple descriptive statistics, repeat steps each function.

Let us try to find descriptive analysis using an example of age of 20 employees of a company. The steps are as follows:
Age of 20 Employees - 25,32, 28, 23, 19, 37, 24, 29, 35, 25, 30, 27, 31, 26, 33, 22, 25, 23, 20, 22

1. Enter the following data into Excel worksheet.

2. Type in the titles of descriptive statistics we want to cal culate. This is for making the process easier and simpler

3. Click on the cell adjacent to the descriptive statistics we want. Here we are first trying to find the mean, so click on the cell adjacent to the title Mean and type the formula.

	A	B	C
1	**Age**		
2	25		Mean
3	32		Median
4	28		Mode
5	23		Standard Deviation
6	19		Variance
7	37		Minimum
8	24		Maximum
9	29		Range
10	35		Count
11	25		Skewness
12	30		Kurtosis
13	27		
14	31		
15	26		
16	33		
17	22		
18	25		
19	23		
20	20		
21	22		
22			

Fig 2.1.1 Dataset and Descriptive Statistics Titles

4. Start typing the formula by typing = sign and then use the formula for mean i.e., AVERAGE. And select the range and click enter. So the formula for the given dataset will be as follows =AVERAGE(A2:A21).

	A	B	C	D	E
1	**Age**				
2	25		Mean	=AVERAGE(A2:A21)	
3	32		Median		
4	28		Mode		
5	23		Standard Deviation		
6	19		Variance		
7	37		Minimum		
8	24		Maximum		
9	29		Range		
10	35		Count		
11	25		Skewness		
12	30		Kurtosis		
13	27				
14	31				
15	26				
16	33				
17	22				
18	25				
19	23				
20	20				
21	22				

Fig 2.1.2 Formula for Finding Mean

5. Click Enter and the result will be as follows:

- Generating descriptive statistics using built-in functions

Mean	26.8
Median	
Mode	
Standard deviation	
Variance	
Minimum	
Maximum	
Range	
Count	
Skewness	
Kurtosis	

Fig 2.1.3 Finding Mean

6. Similarly let us generate descriptive statistics for the whole dataset. The result will be as follows.

Mean	26.8
Median	25.5
Mode	25
Standard deviation	4.97996
Variance	24.8
Minimum	19
Maximum	37
Range	18
Count	20
Skewness	0.450248
Kurtosis	-0.55497

Fig 2.1.4 Results of Descriptive Statistics

The same results can be obtained by using Data Analysis Toolpak. Instead of repeating each function again and again, the descriptive statistics for the whole data set can be analysed in one click using the Data analysis Toolpak

2.1.2.2 Data Analysis Toolpak

The Data Analysis ToolPak is an add-in for Microsoft Excel that provides a collection of powerful data analysis tools and functions. It is designed to assist users in performing various statistical and data analysis tasks directly within Excel, without the need for complex programming or additional software. The Data Analysis ToolPak enhances the spreadsheet software's capabilities by providing a comprehensive set of advanced data analysis and statistical tools. This add-in is designed to cater to the needs of data analysts, researchers, students, and professionals who require powerful statistical functions and tools within the familiar Excel environment.

The Data Analysis ToolPak includes a wide range of statistical functions and tools that allow users to perform tasks such as:

1. **Descriptive Statistics**: Calculate measures such as mean, median, standard deviation, variance, quartiles, and more for a dataset.
2. **Histograms**: Create frequency distributions and histograms to visualise the distribution of data.
3. **Random Number Generation**: Generate random samples from different probability distributions, such as normal, uniform, and exponential distributions.
4. **Sampling**: Perform random sampling from a dataset to estimate population parameters.
5. **Correlation**: Calculate correlation coefficients, such as Pearson's correlation and Spearman's rank correlation, to measure the relationship between variables.
6. **Regression Analysis**: Conduct linear regression analysis to examine the relationship between independent and dependent variables.
7. **ANOVA**: Perform analysis of variance (ANOVA) to compare means across multiple groups.
8. **t-Tests**: Conduct various types of t-tests, such as one-sample t-test, independent samples t-test, and paired samples t-test.
9. **Fourier Analysis**: Perform Fourier transforms to analyse periodic data or signals.

10. **Covariance**: Calculate the covariance between two sets of data.

- Offers a range of statistical functions and tools for performing data analysis

The Data Analysis ToolPak provides a user-friendly interface that makes it easy for Excel users, even those without extensive statistical knowledge, to perform a variety of data analysis tasks. By installing the ToolPak, users gain access to these powerful analytical tools, expanding Excel's capabilities beyond its basic spreadsheet functions.

To install the Data Analysis ToolPak in Excel, follow these steps:

1. Open Excel and go to the "File" tab in the top-left corner.

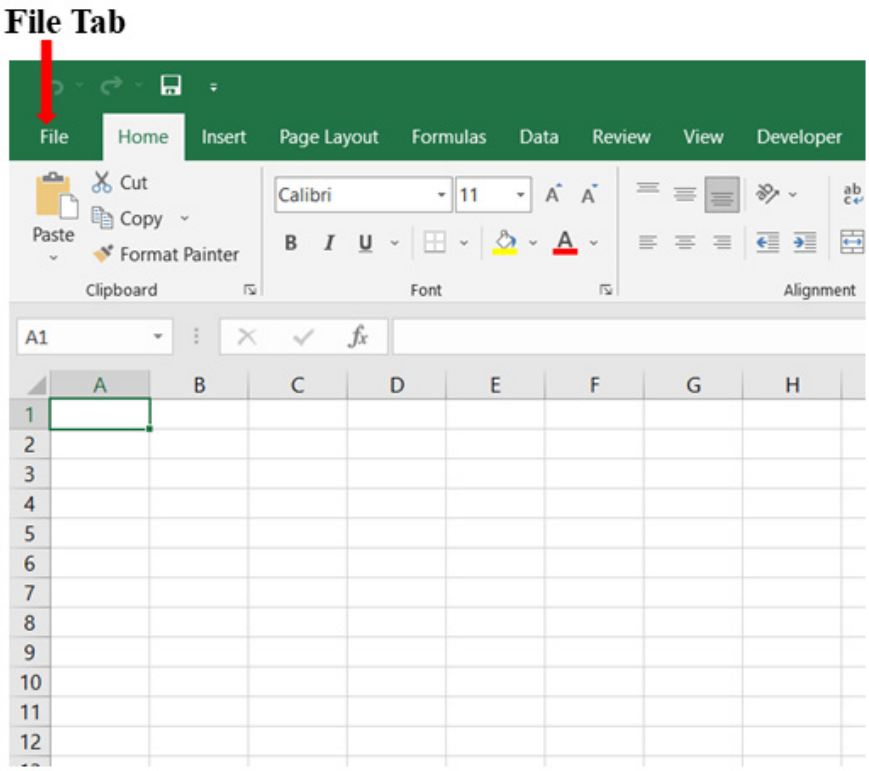


Fig 2.1.5 Selecting File Tab

2. Click on "Options" at the bottom of the navigation pane. This will open the Excel Options window.

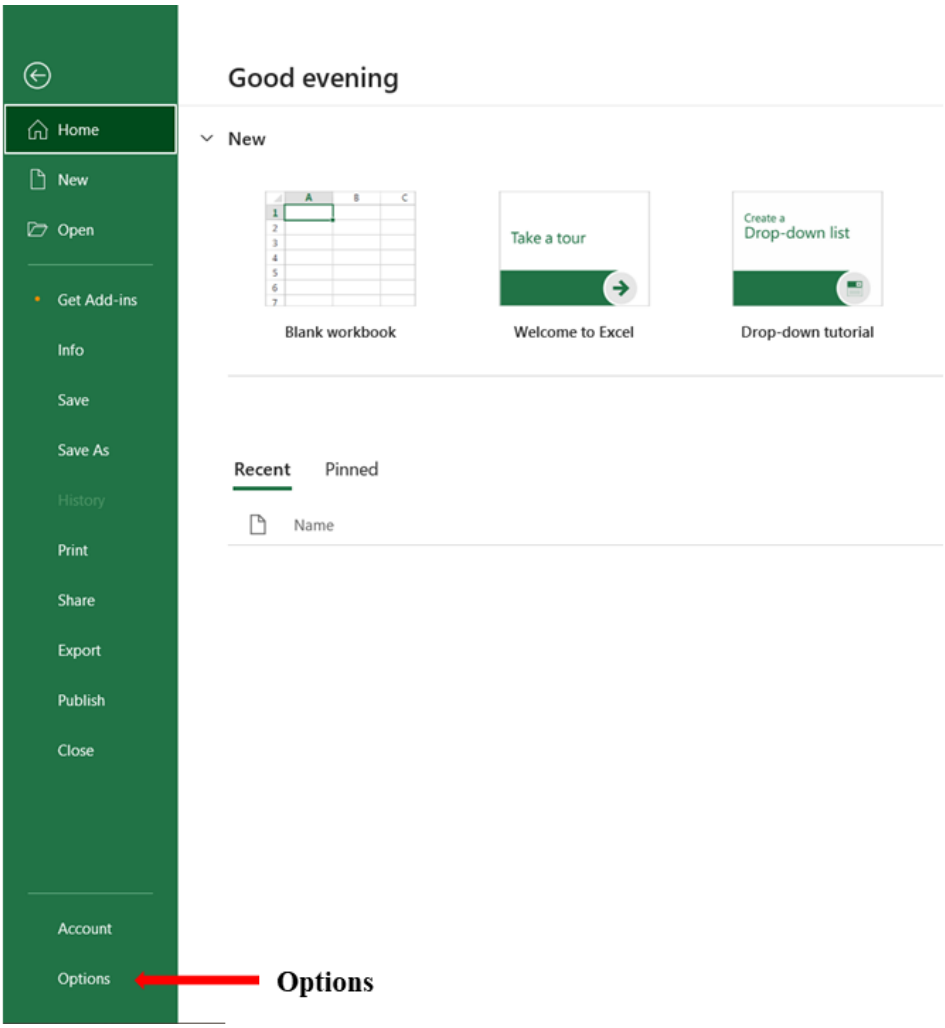


Fig 2.1.6 Selecting Options

3. In the Excel Options window, select "Add-Ins" from the left-hand side menu.

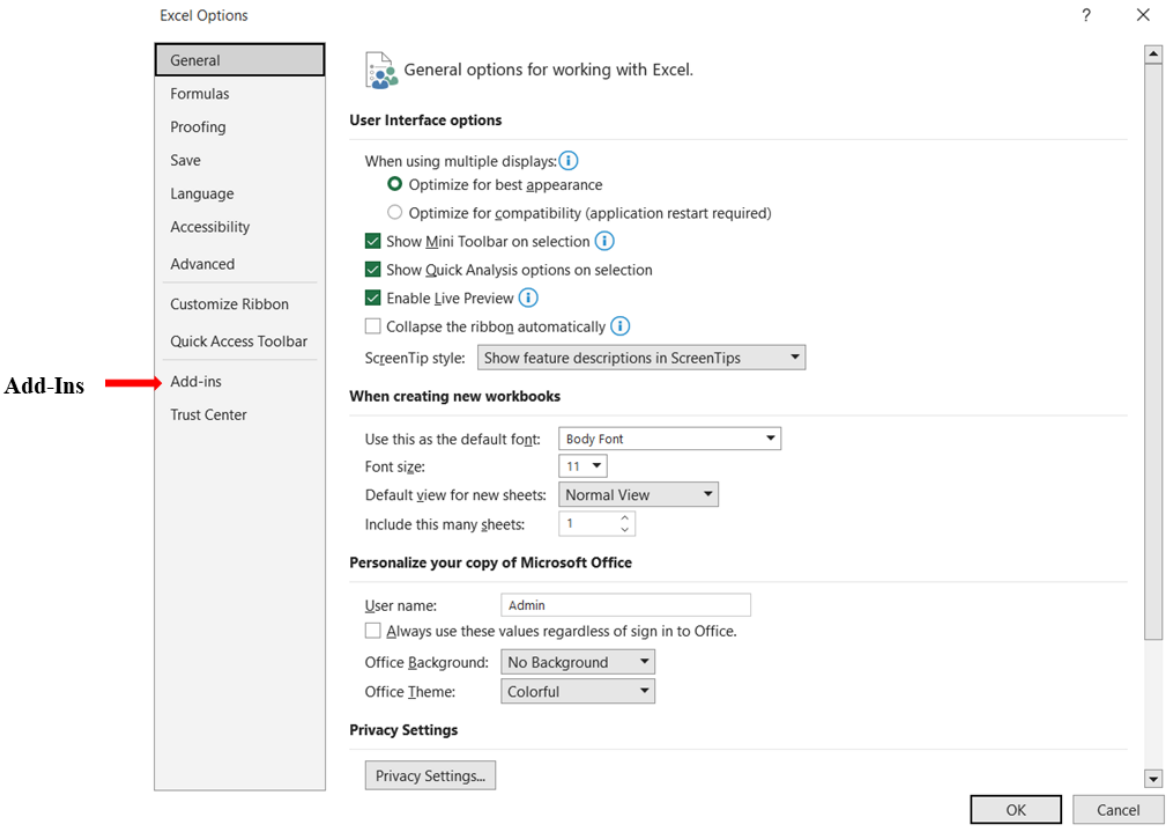


Fig 2.1.7 Selecting Add-ins

4. In the "Manage" dropdown list at the bottom of the window, select "Excel Add-ins," and then click on the "Go" button.

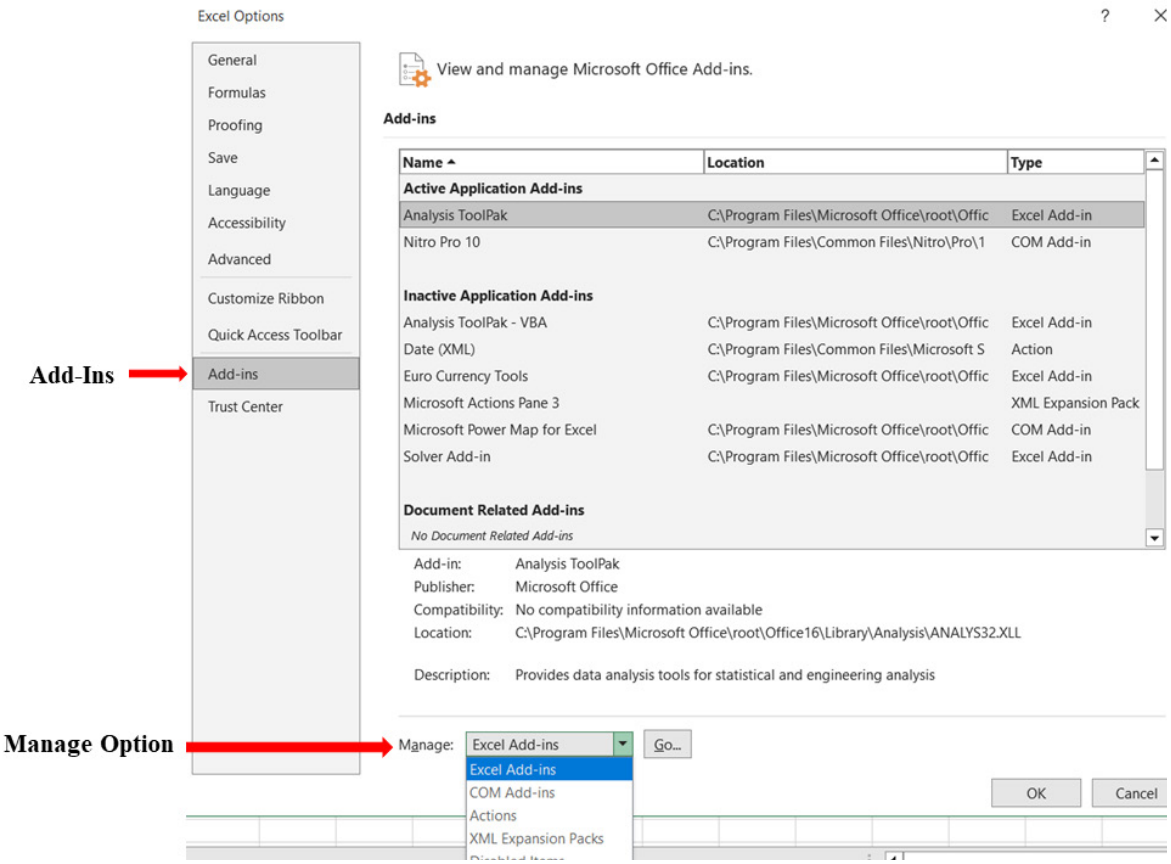


Fig 2.1.8 Choosing Excel Add-ins

5. In the Add-Ins window that appears, check the box next to "Analysis ToolPak" and then click "OK."
6. Excel will now install the Data Analysis ToolPak, and you should see the Data Analysis option under the "Data" tab in the Excel ribbon.
7. To use the Data Analysis ToolPak, go to the "Data" tab in the Excel ribbon, click on "Data Analysis" in the Analysis group, and select the tool or function you want to use.

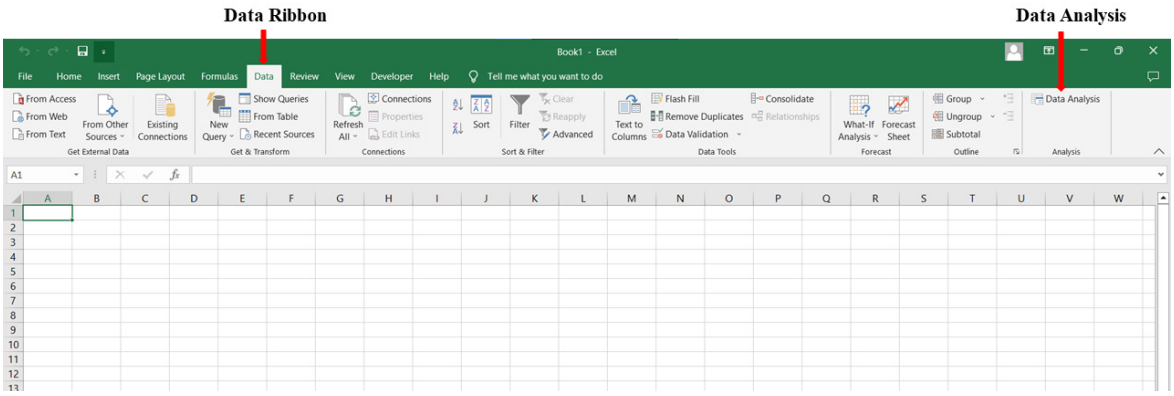


Fig 2.1.9 Data Analysis Tool

If you do not see the "Data Analysis" option under the "Data" tab after completing the steps above, it is possible that the ToolPak was not installed correctly. Try closing and reopening Excel to see if the option appears. If not, repeat the installation steps.

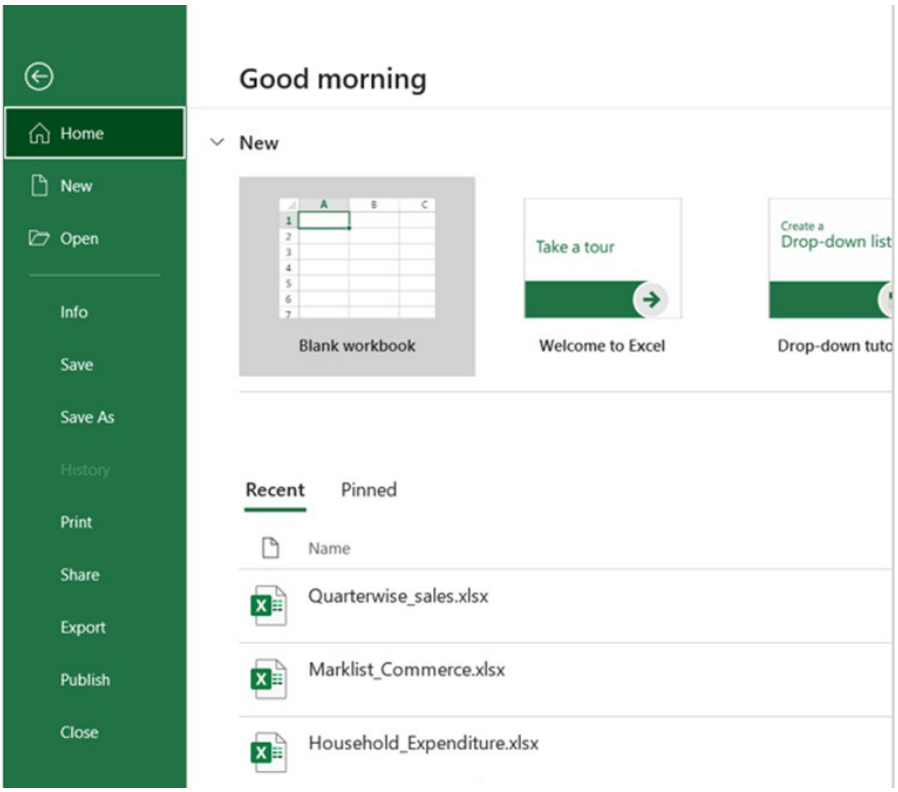


Fig 2.1.10 Options under Data Analysis

The major options available under Data Analysis ToolPak are listed out in Fig 2.1.6. here in our chapter, we will be using them.

To use the Data Analysis ToolPak and perform descriptive statistics in Excel, follow these steps:

1. Enter your dataset into an Excel spreadsheet. Ensure that each column represents a variable, and each row represents an observation or data point.
2. Go to the "Data" tab in the Excel ribbon. Look for the "Data Analysis" button in the "Analysis" group. If you do not see it, close, and reopen Excel, or check if you successfully enabled the ToolPak in the previous step.
3. Click on "Data Analysis."
4. Choose the Descriptive Statistics option:
5. In the "Data Analysis" dialog box, select "Descriptive Statistics" from the list. Click "OK."

6. In the "Descriptive Statistics" dialog box, click on the "Input Range" box, and Select the range of data you want to analyse. Refer Fig2.1.12
7. Optionally, check the "Labels in first row" box if your data has column headers.
8. Choose the output options like "Summary statistics" and "New Worksheet Ply" if you want the results to appear in a new sheet. Or select choose output range and click on a desired output location and click "OK."

Excel will calculate the descriptive statistics based on the selected data range and display the results either in a new worksheet or in the designated location you specified.

Let us try to find descriptive analysis using an example of age of 20 employees of a company. The steps are all follows

1. Let us use the same data set give in Fig 2.1.7 or enter another data set
2. Click on Data Ribbon and Click on Data Analysis as shown in Fig 2.1.5
3. A new window will open up and Select Descriptive Statistics from that and click OK, as shown below.

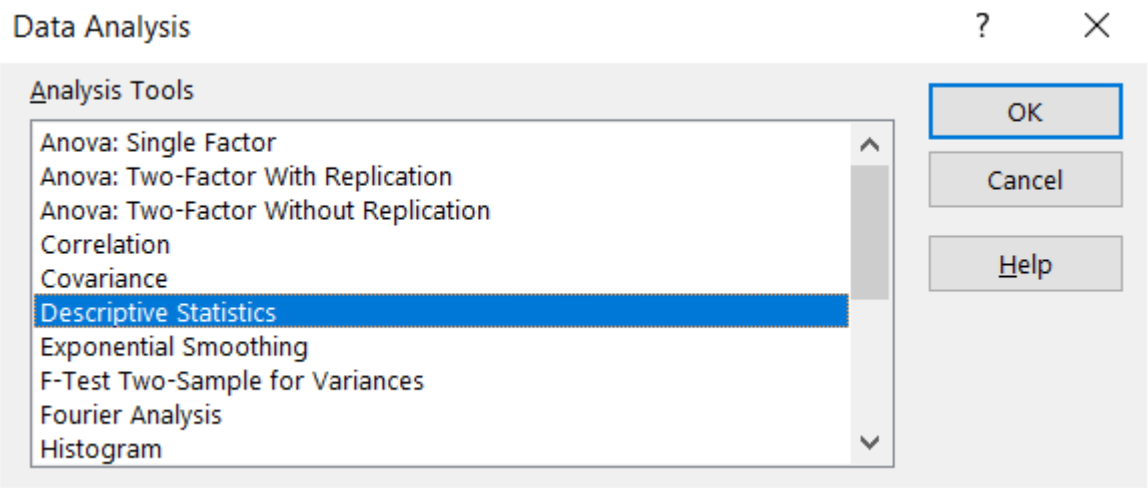


Fig 2.1.11 Selecting Descriptive Statistics Option

4. A new window will pop up and select the entire dataset including the column label. Select the "Label' in first row option if you are selecting the column label , it will give a title to the descriptive statistics we are formulating.
5. Give the output range and then select "Summary statistics", and then click OK. It is shown below.

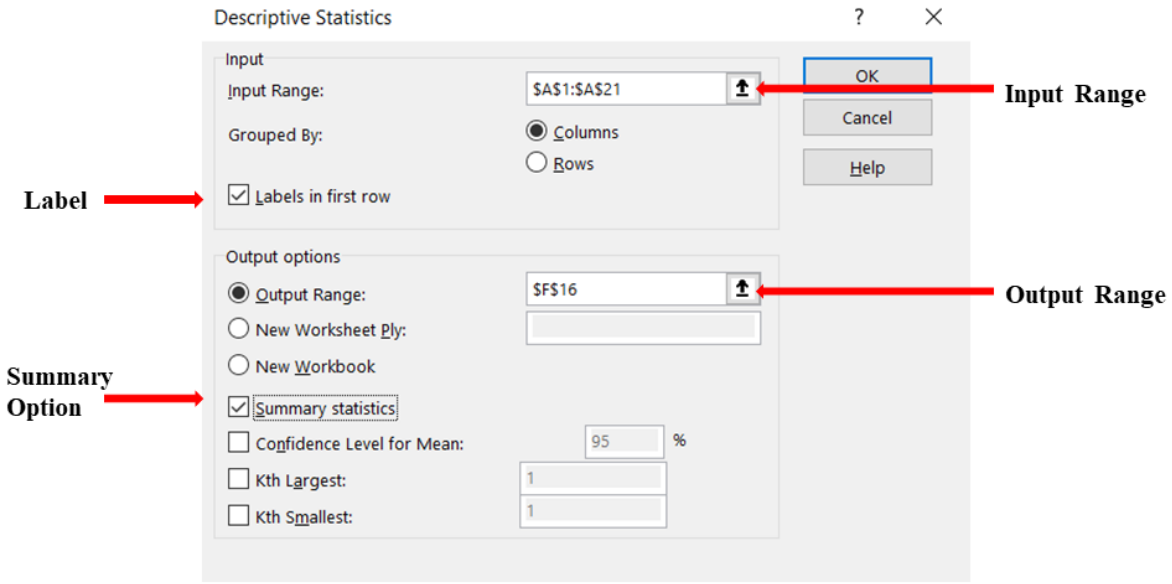


Fig 2.1.12 Working with Data analysis Toolpak

5. The results will be generated as follows

Age	
Mean	26.8
Standard Error	1.113553
Median	25.5
Mode	25
Standard Deviation	4.97996
Sample Variance	24.8
Kurtosis	-0.55497
Skewness	0.450248
Range	18
Minimum	19
Maximum	37
Sum	536
Count	20

Fig 2.1.13 Results of Descriptive statistics

- ToolPak generates the Descriptive Statistics effortlessly

Here we can see that the entire descriptive statistics is generated in one click instead of typing in many formulas.

Remember that Excel is a powerful tool for basic statistical analysis, but for more advanced statistical methods and data analysis, you might need to consider using dedicated statistical softwares.

2.1.4 Creating Advanced Charts and Graphs in Excel

Excel provides several advanced chart types that offer unique ways to visualise data. Here are some of the advanced charts available in Excel and their common uses:

2.1.4.1 Waterfall Chart

A waterfall chart displays the cumulative effect of sequentially introduced positive and negative values. It is useful for illustrating financial statements, budget analysis, and showing the cumulative impact of different factors on a total.

Steps to insert a waterfall chart are as follows:

- Enter the data set into Excel, including the "Month" and numeric columns representing the sales and sales growth. It will be as shown in the Table 2.1.1

Table 2.1.2 Dataset with Sales & Sales Growth

Month	Sales	Sales Growth (%)
January	₹ 10,000	0%
February	₹ 12,000	20%
March	₹ 10,800	-10%
April	₹ 13,500	25%
May	₹ 15,000	11.10%
June	₹ 14,500	-3.30%
July	₹ 16,000	10.30%
August	₹ 17,000	6.30%
September	₹ 16,800	-1.20%
October	₹ 18,500	10.10%
November	₹ 18,000	-2.70%
December	₹ 20,000	11.10%

- Select the Month & Sales Growth columns, including the headers and values.
- Go to the “Insert” tab in the Excel ribbon as shown in the Fig 2.1.14
- Click on Recommended charts option and Select All charts options from the tab as shown in the Fig 2.1.14
- Click on the “Waterfall Chart” option as shown in the Fig 2.1.14 and click on OK

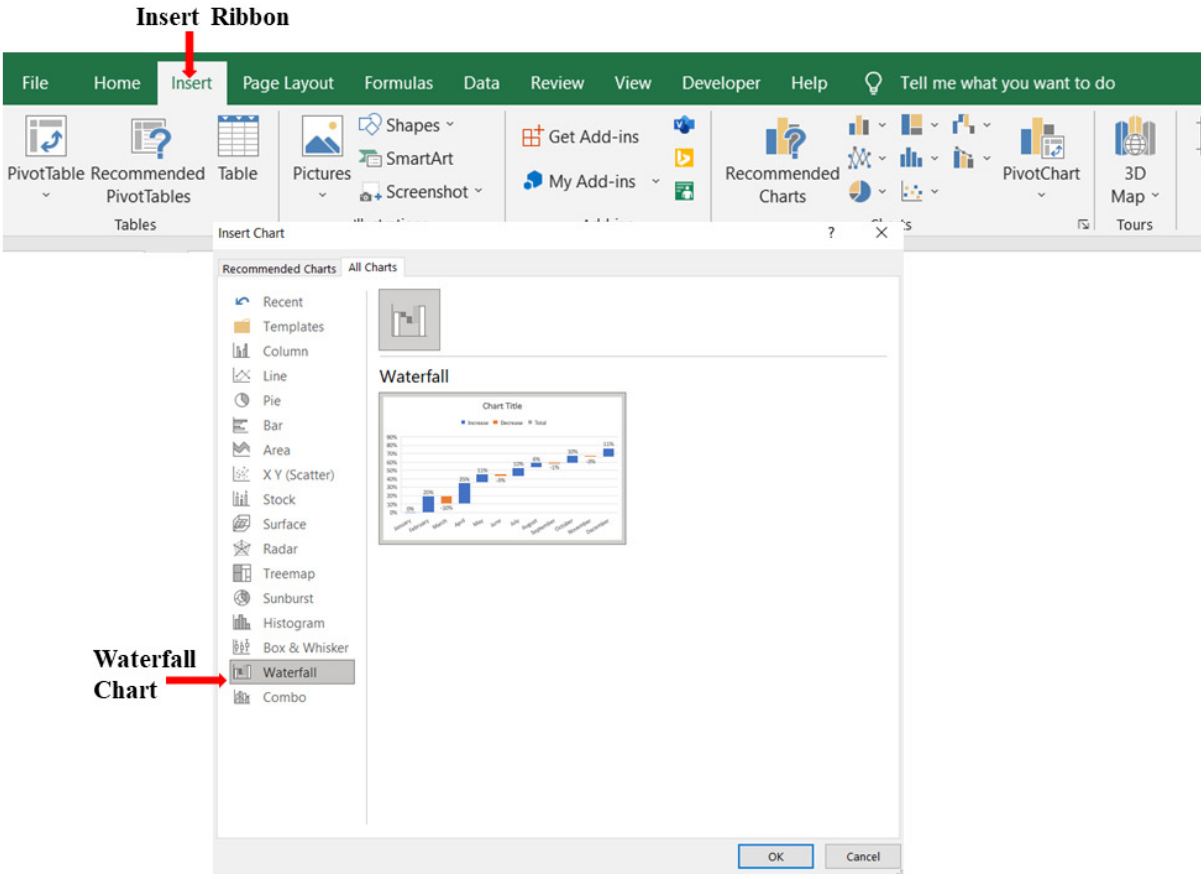


Fig 2.1.14 Inserting Waterfall Chart

- Excel will generate the Waterfall Chart using the selected dataset as shown in the Fig 2.1.15

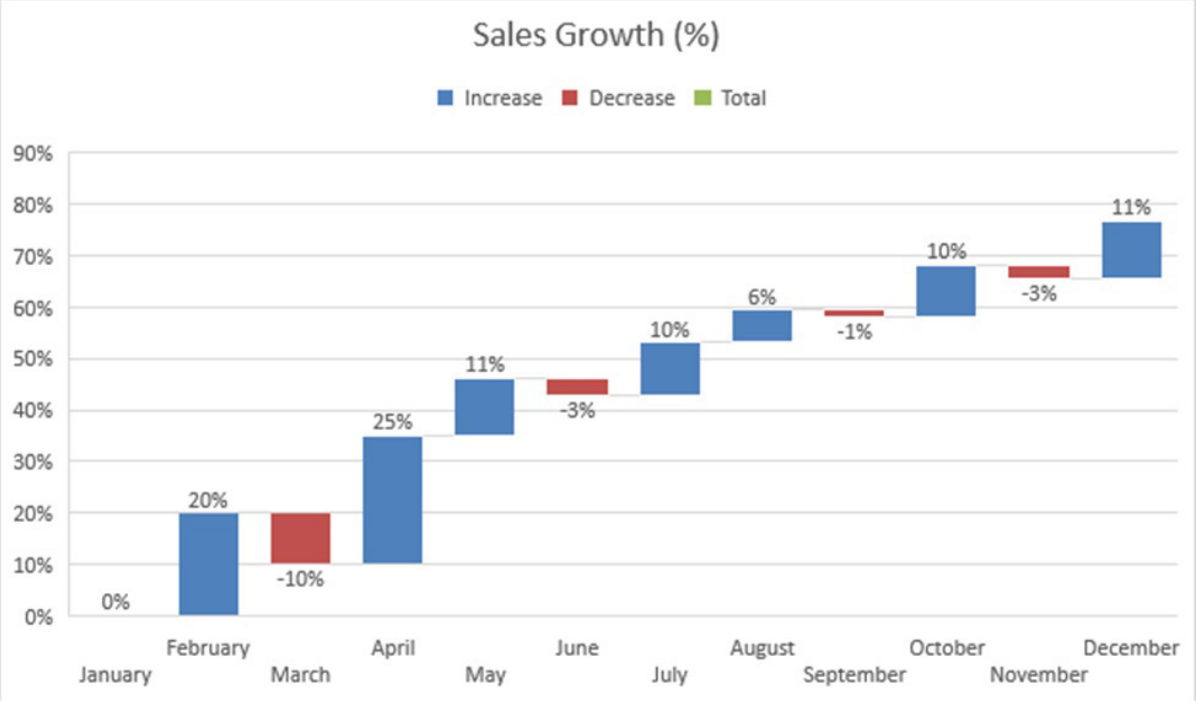


Fig 2.1.15 Waterfall Chart

- Waterfall chart shows cumulative impact of positive and negative values over time

Waterfall Chart shows the cumulative effect of sequentially introduced positive or negative values over time or any other ordered dimension. It is commonly used in finance, project management, and other fields to demonstrate how a starting value changes incrementally through different factors or steps.

2.1.4.2 Radar Chart

A radar chart, also known as a spider chart, displays multivariate data on multiple axes originating from a common point. It is useful for comparing multiple variables across different categories, such as evaluating the performance of different products across various dimensions.

Let us use the data from the Table 2.1.2 for generating radar chart. Steps to insert a radar chart are as follows:

- Enter the data set into Excel, including the "Month" and numeric columns representing the sales. It will be as shown in Table 2.1.2
- Select the Month & Sales columns, including the headers and values.
- Click on Recommended charts option and Select All charts options from the tab as shown in the Fig 2.1.16
- Click on the "Radar Chart" option as shown in the Fig 2.1.16 and click on OK

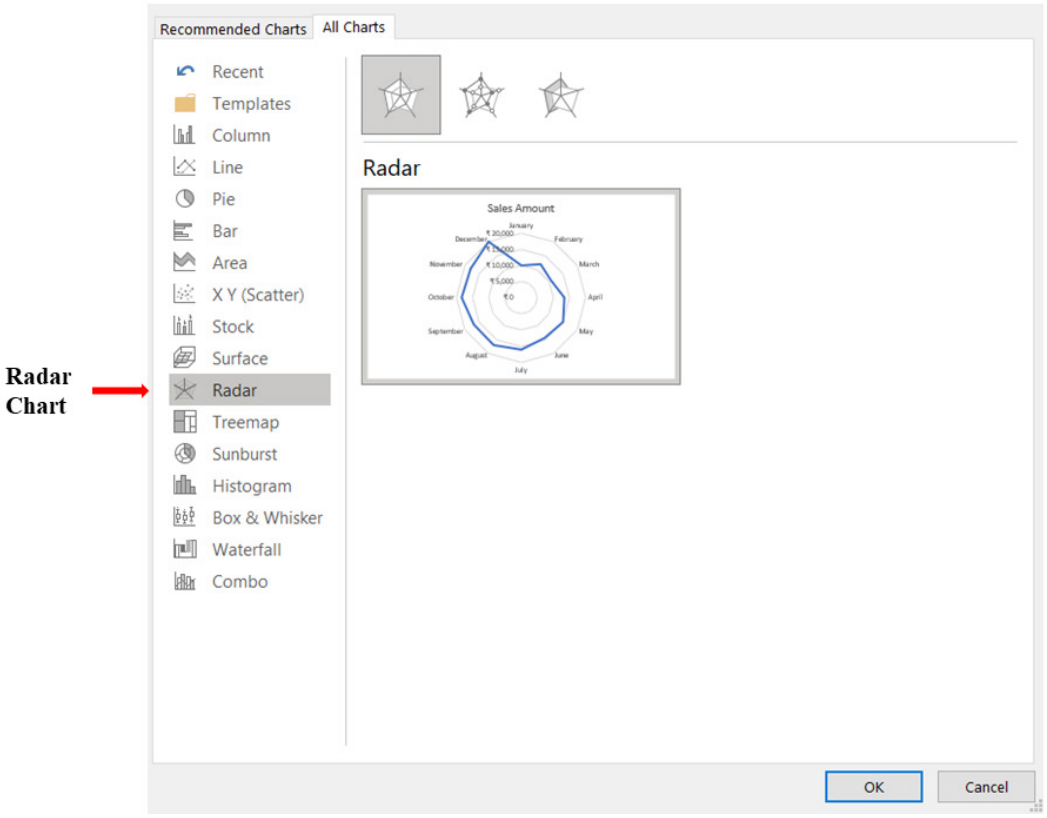


Fig 2.1.16 Inserting Radar Chart

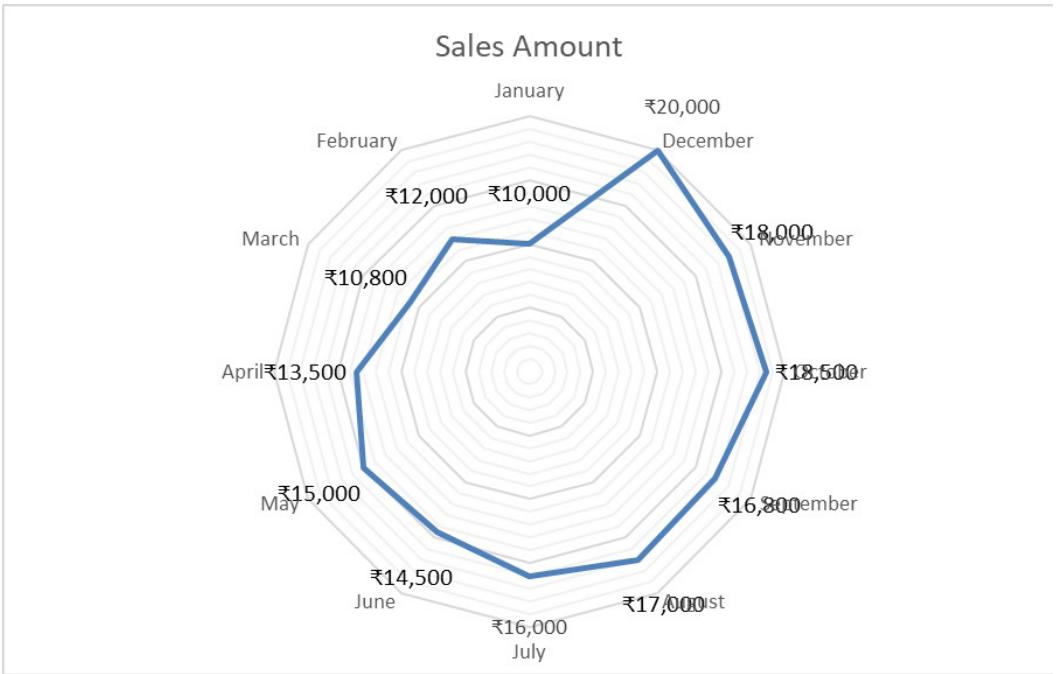


Fig 2.1.17 Radar Chart

- Radar chart displays multivariate data useful for comparing variables across categories

Excel will generate a Radar Chart as shown in Fig 2.1.17. You can customise the chart by right-clicking on it and selecting the "Format Data Series" option. Modify the axis labels, add data labels, adjust the chart title, and other formatting options as desired.

Radar Chart is particularly useful for comparing the relative performance or characteristics of different entities across the variables. It is commonly used in various fields such as market research, sports analytics, and performance evaluations to visualise and analyse how different entities score or perform across multiple attributes simultaneously.

2.1.4.3 Treemap Chart

A treemap chart represents hierarchical data using nested rectangles. Each rectangle's size and colour represent a particular dimension of the data, making it useful for visualising hierarchical structures, such as file directories, market share, or organisational structures.

Steps to insert a Treemap chart are as follows:

- Enter the data set from the Table 2.1.3 into Excel, including the "Month" and "Sales amount" columns.

Table 2.1.3 Sales

Month	Sales Amount
January	₹ 10,000
February	₹ 12,000
March	₹ 10,800
April	₹ 13,500
May	₹ 15,000
June	₹ 14,500
July	₹ 16,000
August	₹ 17,000
September	₹ 16,800
October	₹ 18,500
November	₹ 18,000
December	₹ 20,000

- Select the range of data, including the headers and values.

- Go to the “Insert” tab in the Excel ribbon and click on the “Recommended Charts” option.
- In the “Insert Chart” dialog box, select the “All Charts” tab, and choose the “Treemap” chart type, as shown in the Fig 2.1.18, and click OK.

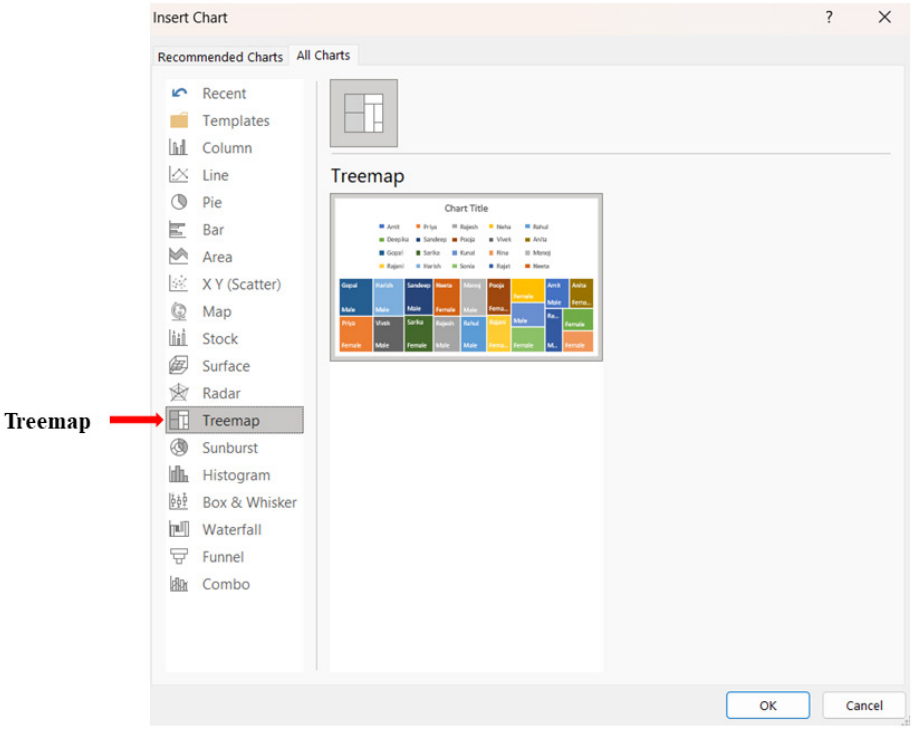


Fig 2.1.18 Inserting Treemap Chart

Excel will generate a default Treemap Chart as shown in the Fig 2.1.19.

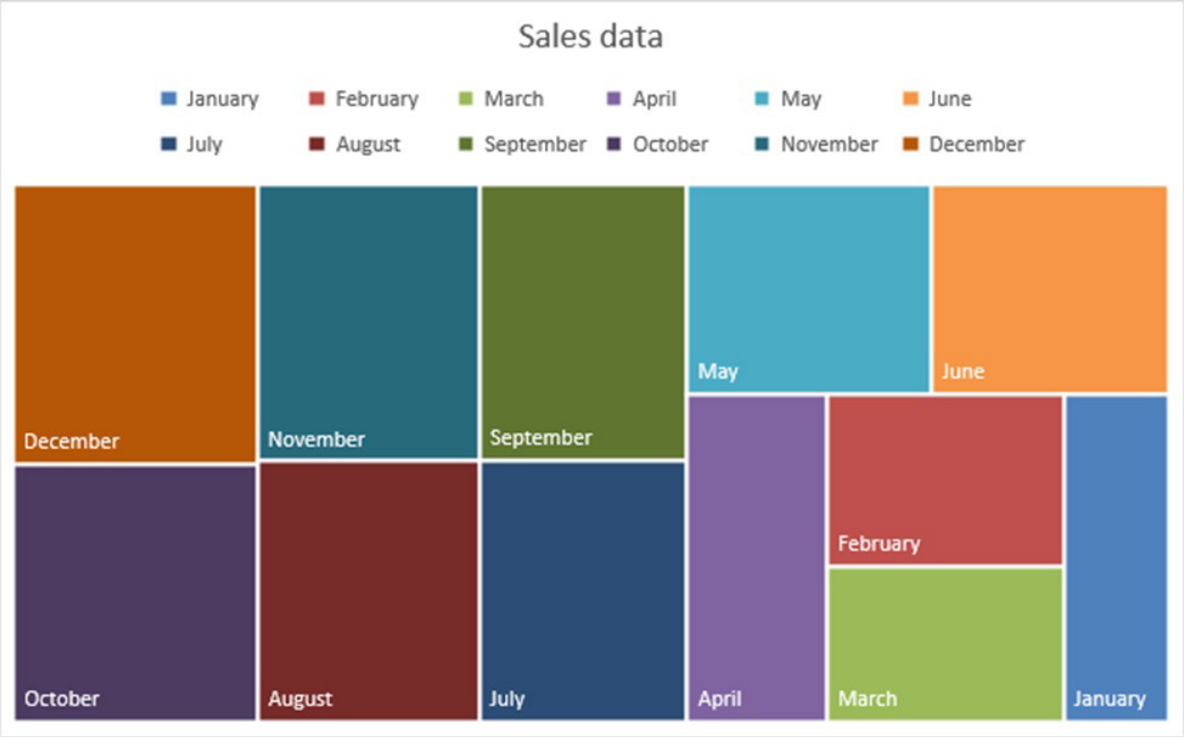


Fig 2.1.19 Treemap Chart

- Treemap chart illustrates hierarchical structures and market share data

You can customise the chart by right-clicking on it and selecting the "Format Data Series" option to adjust the colour scheme, labels, and other settings.

The size of each rectangle represents a quantitative value while the nesting of the rectangles represents different levels of a hierarchy. The largest rectangles represent the largest values, and the smaller rectangles within them represent smaller values within those categories.

2.1.4.4 Pareto Chart

A Pareto chart combines both a column chart and a line graph. It displays data in descending order to identify the most significant factors contributing to a problem. It is commonly used in quality management and decision-making to prioritise actions based on the 80/20 rule (Pareto principle).

Steps to insert a Pareto chart are as follows:

- Here we are using the same table we used for creating Treemap Chart. So, use the data from the Table 2.1.3.
- Select the data including all the headers and values.
- Go to the "Insert" tab in the Excel ribbon and click on the "Histogram Chart" option. Select the 2nd option i.e., the "Pareto " chart type as shown below, and click OK.

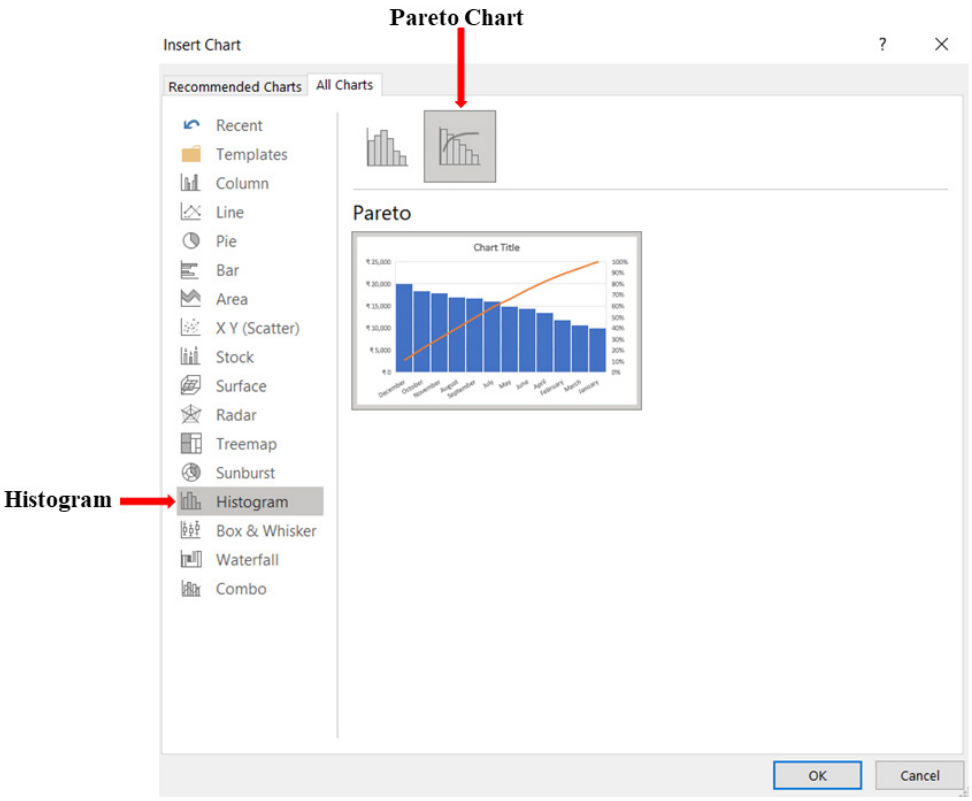


Fig 2.1.20 Inserting Pareto Chart

Excel will generate a default Pareto Chart as shown in Fig 2.1.21.

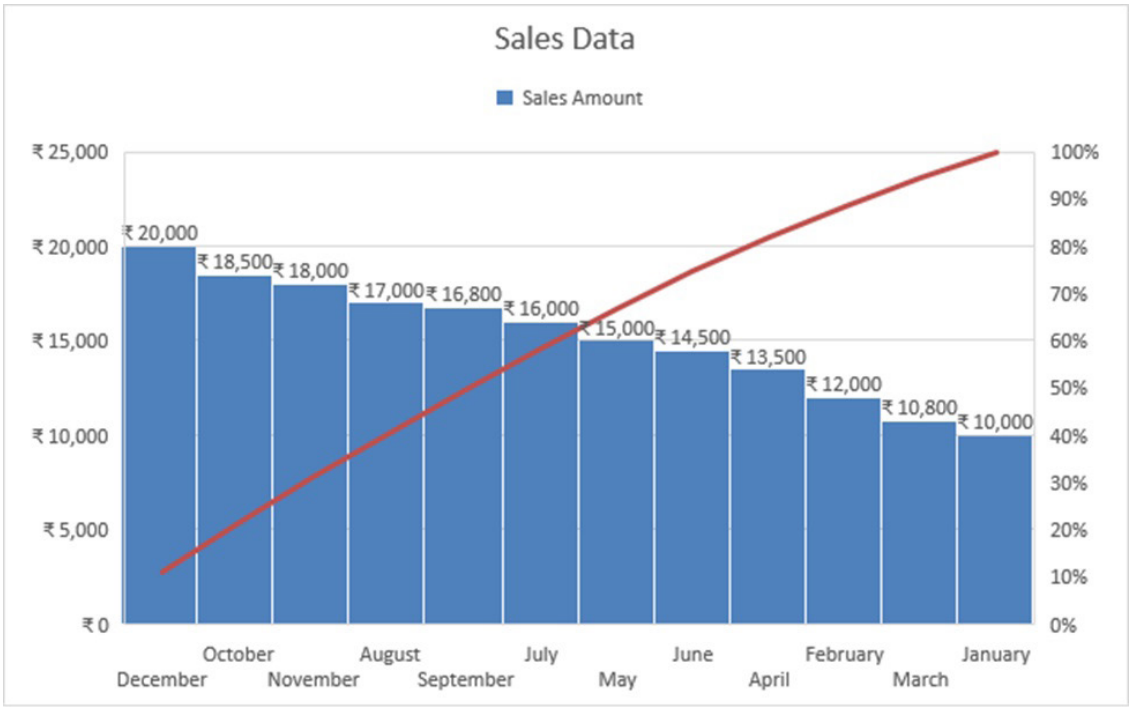


Fig 2.1.21 Pareto Chart

Customise the Pareto chart as desired. You can modify the chart title, axis labels, formatting, and other elements by right-clicking on them and selecting the appropriate options.

- Pareto chart display data in descending order to identify the most significant factors

A Pareto chart displays the frequency or occurrence of different categories in descending order on the vertical axis, represented by the height of the bars. Additionally, a line graph is overlaid on the chart, representing the cumulative percentage of the categories.

The Pareto principle, also known as the 80/20 rule, suggests that a significant majority of effects (approximately 80%) come from a small number of causes (approximately 20%). By analysing a Pareto chart, you can quickly identify and prioritise the most important factors or issues that need attention. This helps focus resources and efforts on the most impactful areas for improvement.

Pareto charts are commonly used in quality management, project management, and problem-solving processes to identify and address the most critical issues for improvement or optimisation.

2.1.4.5 Heat Map

A heat map uses colours to represent data values across a matrix. It is helpful for visualising patterns, relationships, and identifying areas of concentration or intensity. Heat maps are commonly used in data analysis, risk assessment, and geographic data representation.

Steps to insert a Heat map are as follows:

- We are using the data set from the Table 2.1.4 for making Heat Map. It is as given below

Month	Rent	Utilities	Salaries	Advertising	Other Expenses
January	₹ 2,00,000	₹ 50,000	₹ 3,00,000	₹ 1,00,000	₹ 70,000
February	₹ 2,00,000	₹ 55,000	₹ 3,00,000	₹ 1,20,000	₹ 80,000
March	₹ 2,00,000	₹ 60,000	₹ 3,00,000	₹ 90,000	₹ 75,000
April	₹ 2,00,000	₹ 60,000	₹ 3,20,000	₹ 1,00,000	₹ 70,000
May	₹ 2,00,000	₹ 55,000	₹ 3,20,000	₹ 1,10,000	₹ 75,000
June	₹ 2,00,000	₹ 50,000	₹ 3,30,000	₹ 1,00,000	₹ 80,000
July	₹ 2,00,000	₹ 55,000	₹ 3,30,000	₹ 90,000	₹ 70,000
August	₹ 2,00,000	₹ 60,000	₹ 3,50,000	₹ 1,20,000	₹ 85,000
September	₹ 2,00,000	₹ 60,000	₹ 3,50,000	₹ 1,30,000	₹ 80,000
October	₹ 2,00,000	₹ 55,000	₹ 3,60,000	₹ 1,20,000	₹ 70,000
November	₹ 2,00,000	₹ 55,000	₹ 3,60,000	₹ 1,00,000	₹ 75,000
December	₹ 2,00,000	₹ 60,000	₹ 3,70,000	₹ 1,10,000	₹ 80,000

Table 2.1.4 Company Expenses Dataset

- Select the range of data, including the headers and values as shown in the Fig 2.1.22.

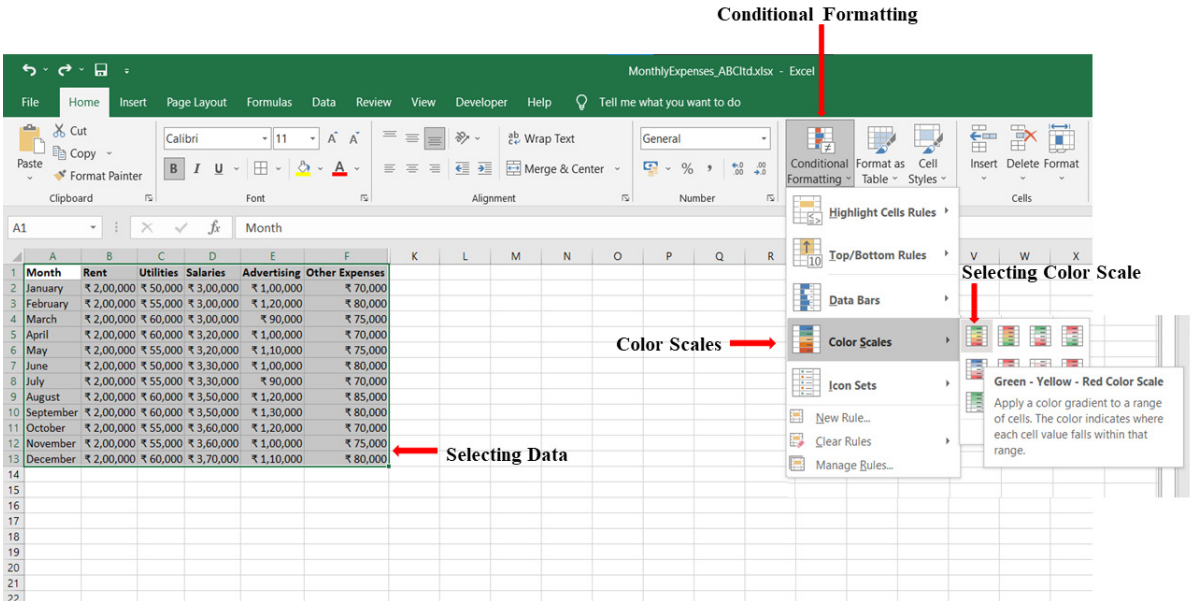


Fig 2.1.22 Selecting Color Scales

- Go to the “Home” tab in the Excel ribbon and click on the “Conditional Formatting” option. This is shown in the Fig 2.1.22
- Choose “Colour Scales” and select the desired heat map colour scale. Here we choose the first option – Green – Yellow – Red Color scale.
- Excel will apply the selected colour scale to the data set, creating a Heat Map.
- The Heat Map will be as shown in the Fig 2.1.23

	A	B	C	D	E	F
1	**Month**	**Rent**	**Utilities**	**Salaries**	**Advertising**	**Other Expenses**
2	January	₹ 2,00,000	₹ 50,000	₹ 3,00,000	₹ 1,00,000	₹ 70,000
3	February	₹ 2,00,000	₹ 55,000	₹ 3,00,000	₹ 1,20,000	₹ 80,000
4	March	₹ 2,00,000	₹ 60,000	₹ 3,00,000	₹ 90,000	₹ 75,000
5	April	₹ 2,00,000	₹ 60,000	₹ 3,20,000	₹ 1,00,000	₹ 70,000
6	May	₹ 2,00,000	₹ 55,000	₹ 3,20,000	₹ 1,10,000	₹ 75,000
7	June	₹ 2,00,000	₹ 50,000	₹ 3,30,000	₹ 1,00,000	₹ 80,000
8	July	₹ 2,00,000	₹ 55,000	₹ 3,30,000	₹ 90,000	₹ 70,000
9	August	₹ 2,00,000	₹ 60,000	₹ 3,50,000	₹ 1,20,000	₹ 85,000
10	September	₹ 2,00,000	₹ 60,000	₹ 3,50,000	₹ 1,30,000	₹ 80,000
11	October	₹ 2,00,000	₹ 55,000	₹ 3,60,000	₹ 1,20,000	₹ 70,000
12	November	₹ 2,00,000	₹ 55,000	₹ 3,60,000	₹ 1,00,000	₹ 75,000
13	December	₹ 2,00,000	₹ 60,000	₹ 3,70,000	₹ 1,10,000	₹ 80,000
14						

Fig 2.1.23 Heat Map

- You can further customise the Heat Map by adjusting the colour scale, conditional formatting rules, and other settings.

- Heat map represents data values using colours

Heatmaps are useful for visualising patterns, relationships, or distributions in large datasets. Heatmaps facilitate comparisons between different categories or groups. By scanning the colours across rows or columns, you can compare the relative values or performance of various variables or entities. Heatmaps make it easier to spot outliers or unusual values. Cells with distinct or contrasting colours can indicate exceptional or unexpected occurrences in the data

2.1.4.6 Sparkline

A sparkline is a small, condensed chart that fits within a single cell. It provides a quick visual representation of trends or pat-

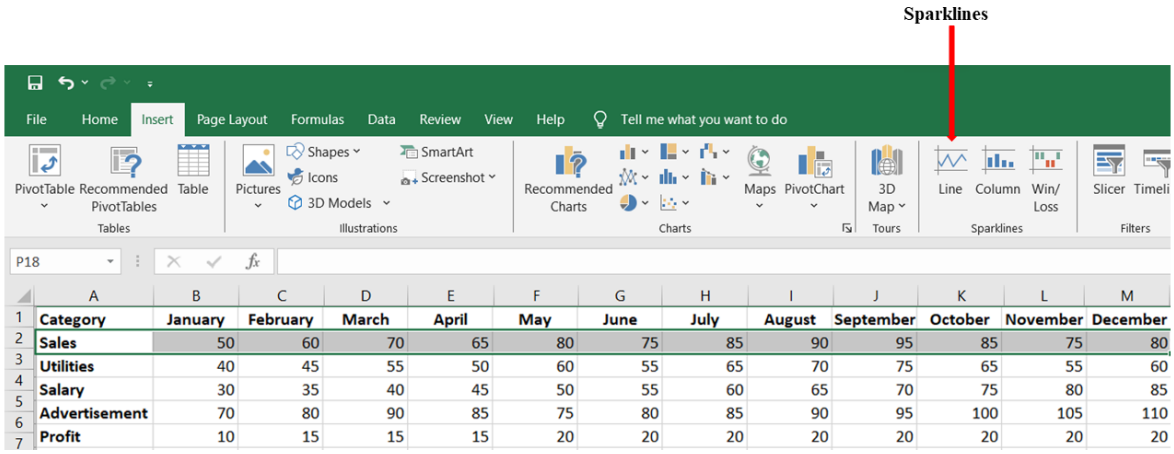
terns in data over time. Sparklines are useful for embedding charts within tables or conveying a high-level overview of data. Steps to insert a Sparkline are as follows:

- Enter the following data set into Excel, including the "Month" and a numeric column representing a trend or data series.

Table 2.1.5 Monthly Dataset

Category	January	February	March	April	May	June	July	August	September	October	November	December
Sales	50	60	70	65	80	75	85	90	95	85	75	80
Utilities	40	45	55	50	60	55	65	70	75	65	55	60
Salary	30	35	40	45	50	55	60	65	70	75	80	85
Advertisement	70	80	90	85	75	80	85	90	95	100	105	110
Profit	10	15	15	15	20	20	20	20	20	20	20	20

- Select a range of cells where you want to insert the sparkline chart. Here we are selecting the second row. i.e. the sales row. It is as shown in the Fig 2.1.24.
- Go to the "Insert" tab in the Excel ribbon and click on the "Line" or "Column" Sparkline option, as shown in the Fig 2.1.24.



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Category	January	February	March	April	May	June	July	August	September	October	November	December
2	Sales	50	60	70	65	80	75	85	90	95	85	75	80
3	Utilities	40	45	55	50	60	55	65	70	75	65	55	60
4	Salary	30	35	40	45	50	55	60	65	70	75	80	85
5	Advertisement	70	80	90	85	75	80	85	90	95	100	105	110
6	Profit	10	15	15	15	20	20	20	20	20	20	20	20
7													

Fig 2.1.24 Selecting Sparklines

- Here we are selecting the Line option from the Sparklines option. There are column option and win/loss option.

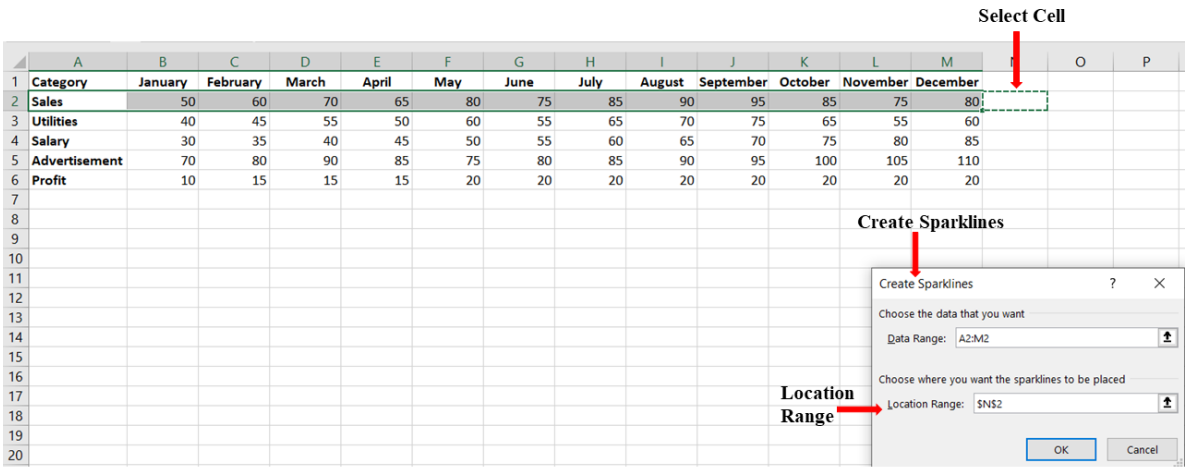


Fig 2.1.25 Inserting Sparkline

- When you click the Line option a new window will pop up as shown in Fig 2.1.26.
- Click on the location range option in Fig 2.1.26, and select where we want the result to be shown. Here we have selected cell N2.

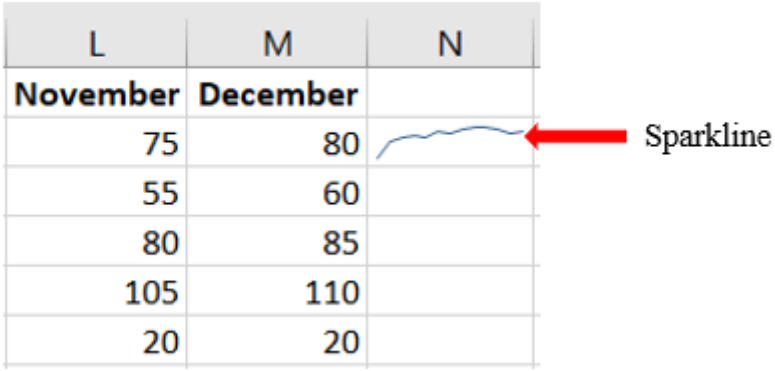
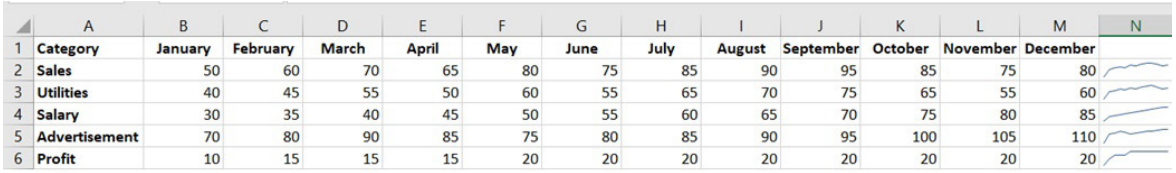


Fig 2.1.26 Sparkline for Sales

- Excel will generate a default Sparkline chart within the selected range of cells. It will be as shown in Fig 2.1.27.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Category	January	February	March	April	May	June	July	August	September	October	November	December	
2	Sales	50	60	70	65	80	75	85	90	95	85	75	80	
3	Utilities	40	45	55	50	60	55	65	70	75	65	55	60	
4	Salary	30	35	40	45	50	55	60	65	70	75	80	85	
5	Advertisement	70	80	90	85	75	80	85	90	95	100	105	110	
6	Profit	10	15	15	15	20	20	20	20	20	20	20	20	

Fig 2.1.27 Sparkline

- You can similarly insert sparklines for other row headings such as utilities, salary etc. The sparklines for the rest of the rows will be as shown in Fig 2.1.27.
- You can customise the chart by right-clicking on it and selecting the "Sparkline Design" or "Type" options to

modify the appearance, colour, and other settings.

- Sparklines are condensed charts that fit within a single cell

Sparklines typically represent data series such as time series, trends, or comparisons. They can be used to show the overall pattern, fluctuations, or changes in the data over a specific period. Unlike traditional charts, sparklines do not have axis or labels, but instead, they rely on the visual representation of data points and their relative positions within the chart.

2.1.4.7 Bubble Chart

A bubble chart displays three dimensions of data using bubbles of different sizes and colours. It is helpful for illustrating the relationships between variables and is often used in financial analysis, population studies, and market research.

- Here we are using the same table we used for creating Treemap Chart. So, use the data from Table 2.1.4
- Select the range of data, including the headers and values. Here we are selecting only the first and last columns. i.e., month and other expenses.
- Go to the “Insert” tab in the Excel ribbon and Select "Recommended charts"
- Click on the “X Y Scatter” option All charts option and click on bubble chart option from that, as shown below.

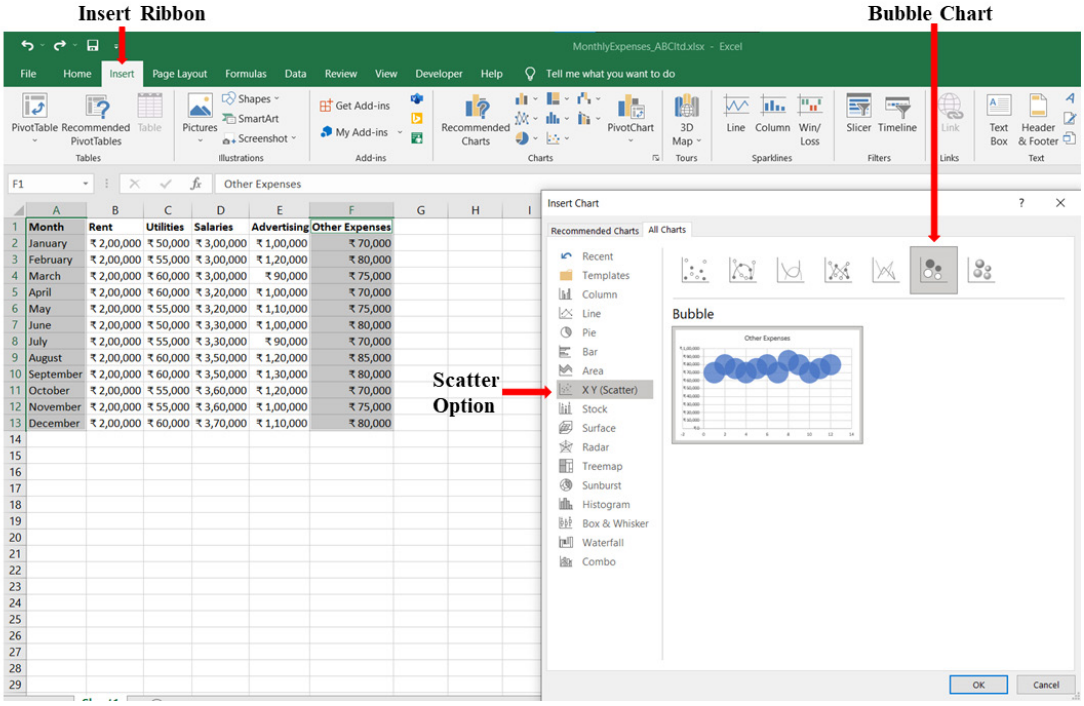


Fig 2.1.28 Inserting Bubble Chart

- Select the desired bubble chart type, such as “Bubble with 3-D effect” or “Bubble with straight lines.” Here we have selected the normal Bubble chart
- Excel will generate a Bubble Chart using the selected data set as shown below

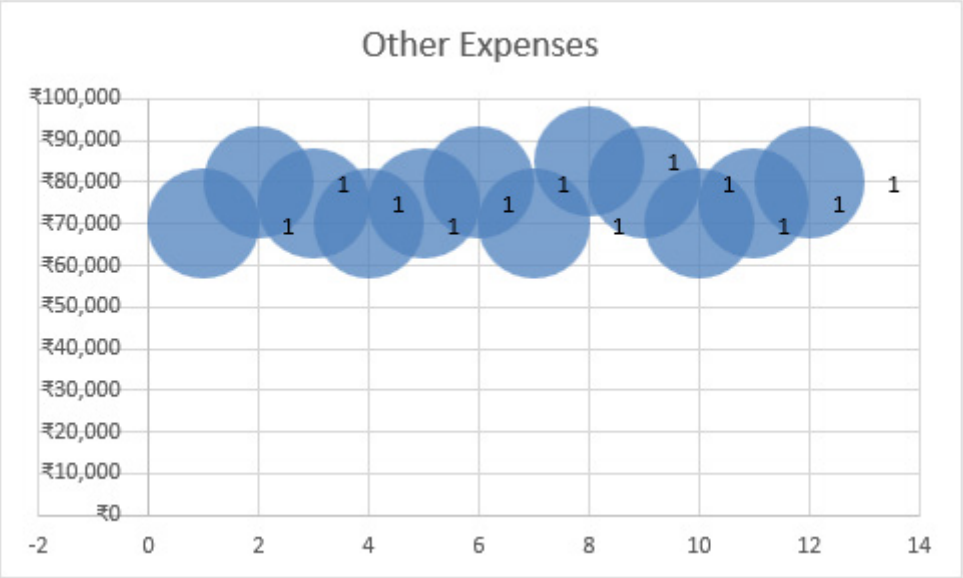


Fig 2.1.29 Bubble Chart

You can customise the chart by right-clicking on it and selecting the “Format Chart Area” option to modify the chart title, axis labels, formatting, and other elements.

- Bubble chart displays three dimensions of data using bubbles of different sizes and colours

Bubble chart displays three dimensions of data in a two-dimensional chart. It is similar to a scatter plot but introduces a third dimension by varying the size of the data points (bubbles), in addition to their x and y positions.

2.1.5 Data Analysis Tools Using Advanced Charts

Data analysis using advanced charts can be a powerful way to gain insights and present information in a visually appealing and informative manner. Advanced charts go beyond the standard bar, line, and pie charts and often involve visualising complex relationships or multidimensional data. Here are some popular advanced charts and how they can be used in data analysis.

2.1.5.1 Histogram

A histogram is a graphical representation of the distribution of data, showing the frequency of values within specific intervals (or bins). It is commonly used to visualise the underlying

distribution of a continuous variable or to identify patterns and trends in the data. The x-axis represents the ranges of values, and the y-axis shows the frequency or count of values falling within those ranges.

Here are the steps to create a Histogram chart in Excel

- Enter the following dataset into Excel

	A	B
1	Student Name	Hours Studied
2	Priya	2
3	Gupta	1
4	Das	3
5	Nisha	4
6	Akash	3
7	Hari	2
8	Sanjay	5
9	Nikhil	1
10	Rupin	3
11	Huzam	2

Fig 2.1.30 Student Dataset

- Select the Hours Studied row and click on "Insert" Ribbon and Select "Recommended charts"

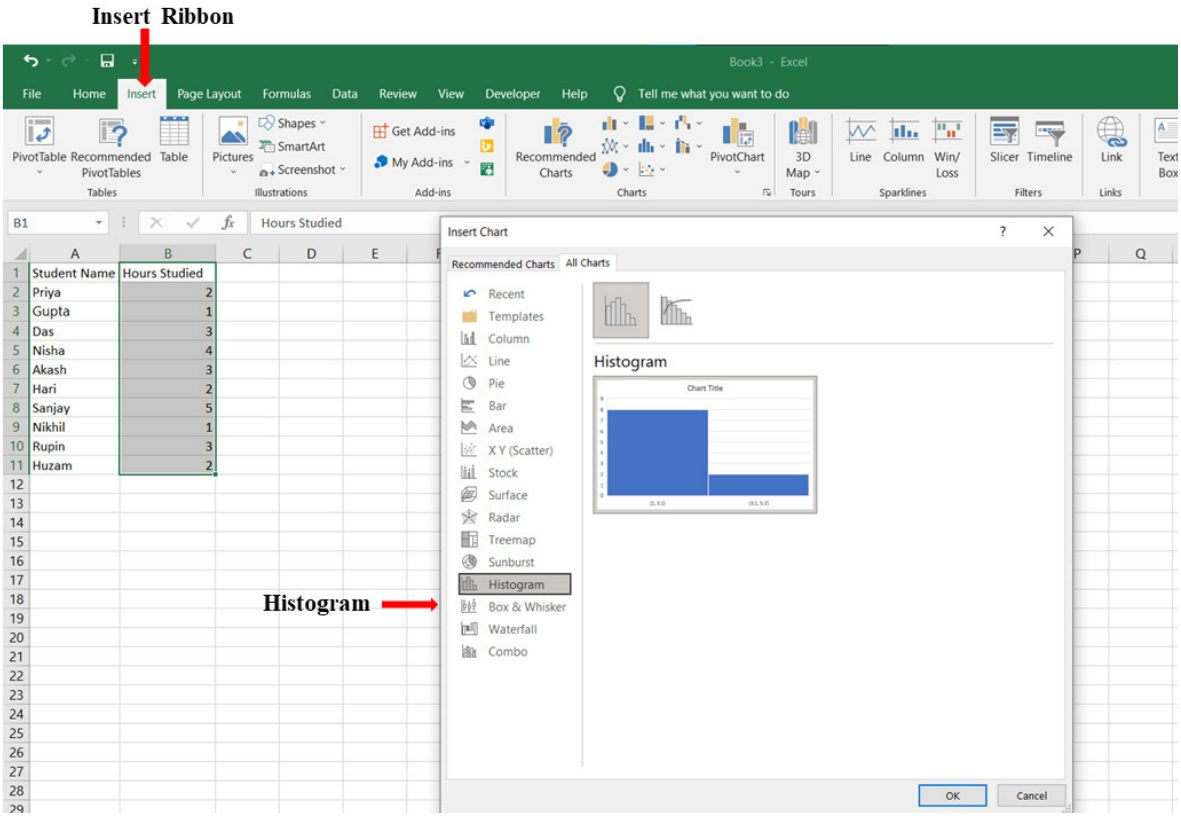


Fig 2.1.31 Inserting Histogram

- Select "Histogram" from that list, click on the first option and click OK
- Excel will generate a Histogram within the selected range of cells. It will be as shown below

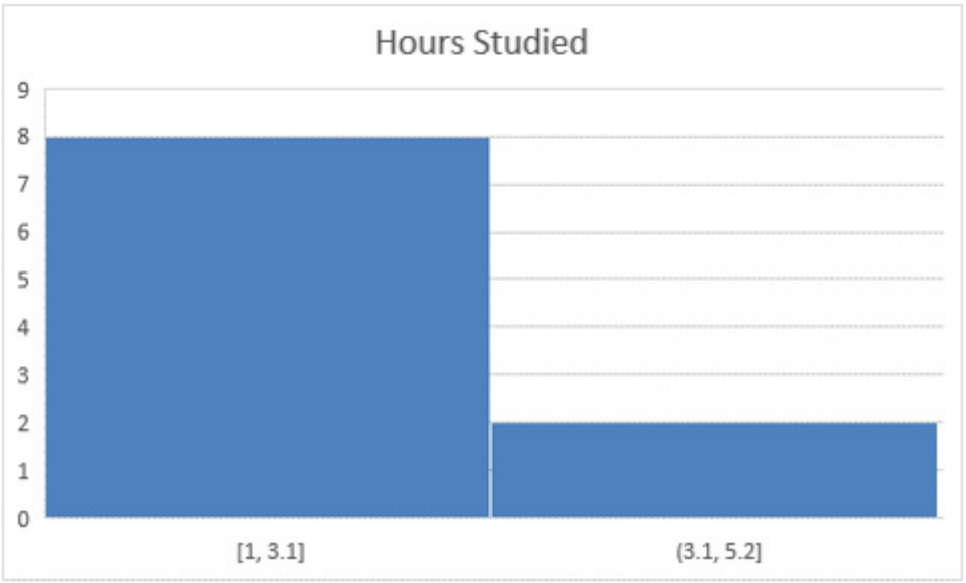


Fig 2.1.32 Histogram

- Histogram visually represents the distribution of data within specific intervals

A histogram displays the distribution of data by showing the frequency or count of values falling within specific intervals, often referred to as "bins." It is a graphical representation of the underlying frequency distribution of a continuous variable. The main purpose of a histogram is to visualise the pattern or shape of the data, which can provide insights into its central tendency, spread, and potential outliers.

2.1.5.2 Scatter Plot

A scatter plot is a graph that displays individual data points along two axes, typically used to visualise the relationship between two continuous variables. Each data point is represented by a dot on the graph, and the position of the dot depends on the values of the variables being plotted. Scatter plots are useful for identifying correlations or patterns between two variables. A scatter plot is also called a scatter chart, scattergram, or scatter plot, XY graph.

Here are the steps to create a Scatterplot chart in Excel:

- Let us create a dataset with two sets of numerical data in two different columns that you can use to create a scatter plot in Excel. For example, we will create a dataset representing the relationship between "Hours Studied" and "Test Scores" for a group of students. It is as follows:

	A	B	C
1	**Student Name**	**Hours Studied**	**Test Scores**
2	Priya	2	70
3	Gupta	1	65
4	Das	3	75
5	Nisha	4	85
6	Akash	3	80
7	Hari	2	73
8	Sanjay	5	88
9	Nikhil	1	66
10	Rupin	3	72
11	Huzam	2	69
12			

Fig 2.1.33 Test Score & Hours Studied Dataset

- Select both sets of numerical data (both columns).
- Click on Insert Ribbon and Select Recommended charts in the “Charts” group, click on the “Scatter” option as shown below

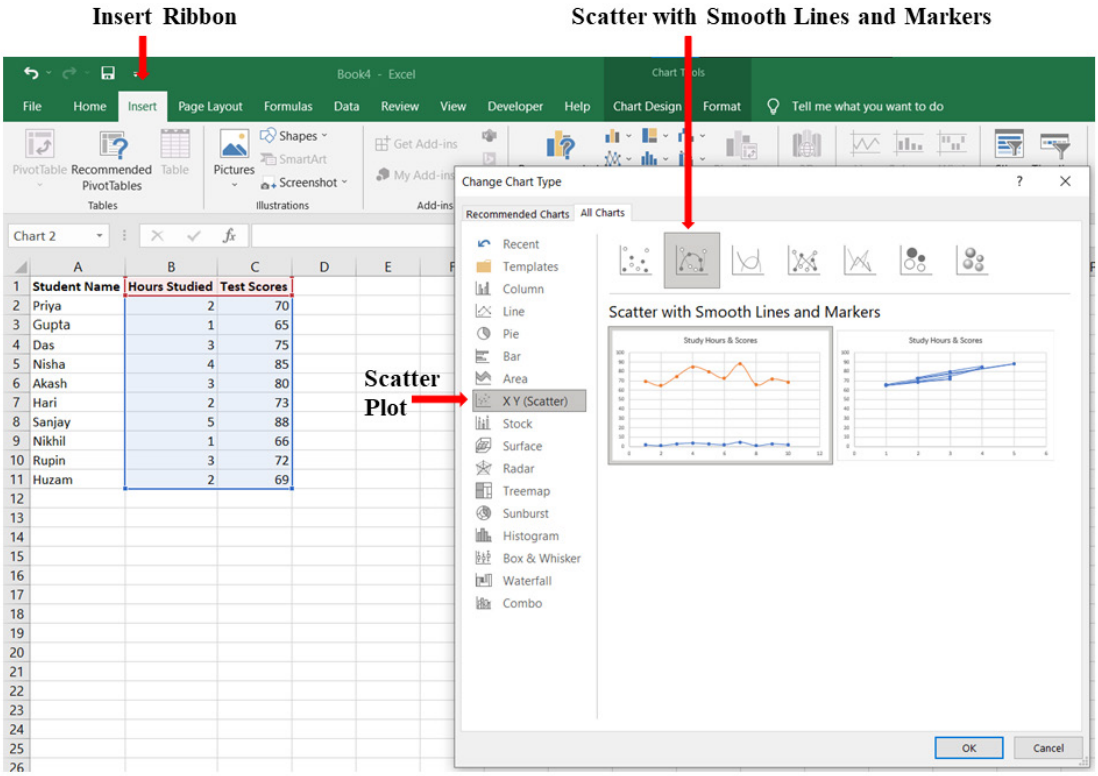


Fig 2.1.34 Inserting Scatterplot

- Select the type of scatter plot you prefer (e.g., with or without connecting lines between data points). Here we have selected the second option to make the scatter plot look for understandable. It is shown in the figure above.
- Excel will create the scatter plot on the same sheet as shown below

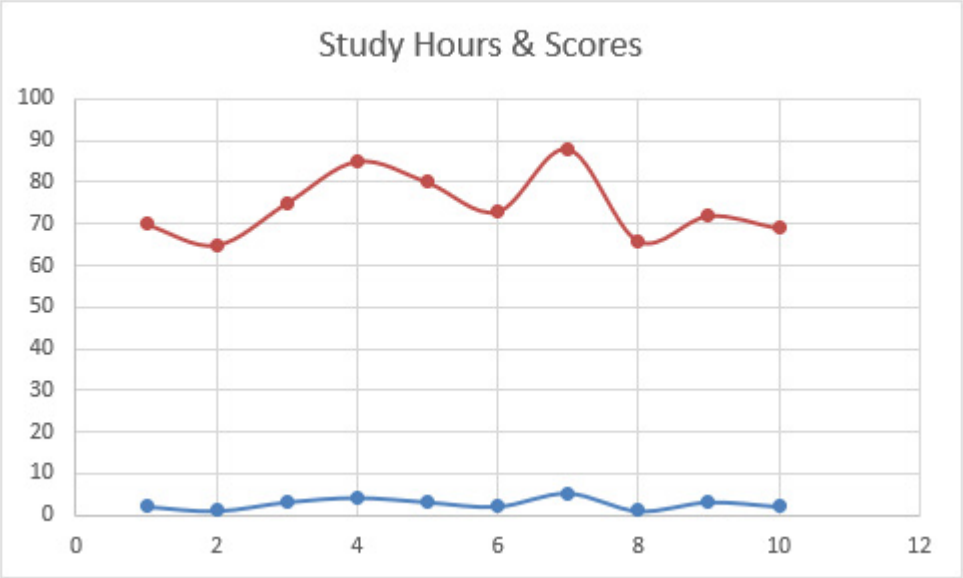


Fig 2.1.35 Scatterplot

- Scatter plot visually represents the relationship between two continuous variables

You should now have a scatter plot that displays the relationship between the two variables. If you want to add labels, customise the chart, or modify any other aspect of the scatter plot, you can use the chart tools available in Excel.

- Trendlines in scatterplot

A trendline in a scatter plot chart is a line that visually represents the general direction and pattern of the data points. It helps in identifying trends and making predictions based on the observed data. Trendlines are particularly useful in analysing relationships between variables and understanding the overall pattern or tendency in the data. To add a trendline to a scatter plot chart in Excel, begin by creating the scatter plot with the desired data points. After selecting the scatter plot, right-click on one of the data points and choose "Add Trendline" from the context menu. In the Format Trendline pane, opt for the type of trendline you want (e.g., linear, exponential) and adjust formatting options if needed. You can choose to display the regression equation on the chart by checking the "Display Equation on chart" option.

Displaying the equation on the chart, often referred to as the regression equation, provides a mathematical representation of the trendline. This equation summarises the relationship between the variables and allows for quantitative analysis. It typically takes the form of $y = mx + b$, where "y" is the dependent variable, "x" is the independent variable, "m" is the slope

- Displaying the regression equation

of the line, and "b" is the y-intercept. If you initially didn't enable the equation display, you can still show it by right-clicking on the trendline, selecting "Add Trendline," checking the "Display Equation on chart" option in the Format Trendline pane, and closing the pane.

The trendline and its equation are valuable tools for making predictions about future data points based on the established pattern. They also helps in understanding the strength and direction of the relationship between variables.

2.1.5.3 Stock Chart

A stock chart is specifically designed to visualise financial data, including the open, high, low, and closing prices of a stock or other financial instrument. It helps analyse stock trends, patterns, and volatility. The steps to create a stock chart are as follows:

- Enter the stock market data into Excel, including the date, open price, high price, low price, and closing price columns. It is as shown below.

	A	B	C	D	E
1	**Date**	**Open Price**	**High Price**	**Low Price**	**Closing Price**
2					
3	01-07-2023	150.5	155.2	149.8	153.75
4	02-07-2023	153.8	156.5	152.2	154.9
5	03-07-2023	155	158.3	152.5	157.2
6	04-07-2023	157.3	159.7	155.8	158.5
7	05-07-2023	158.6	160.4	156.9	159.2
8	06-07-2023	159.5	162.1	158.4	161.5
9	07-07-2023	161.8	164.2	160.7	163.9
10	08-07-2023	163.7	166.8	161.5	165.2
11	09-07-2023	166.3	169	164.8	167.5
12	10-07-2023	167.8	170.5	166.2	169.8
13					

Fig 2.1.36 Stock Market Dataset

- Select the range of data, including the headers and values.
- Go to the “Insert” tab in the Excel ribbon and click on the “Stock Chart” option, from the "All Charts" option in "Charts" group, as shown below

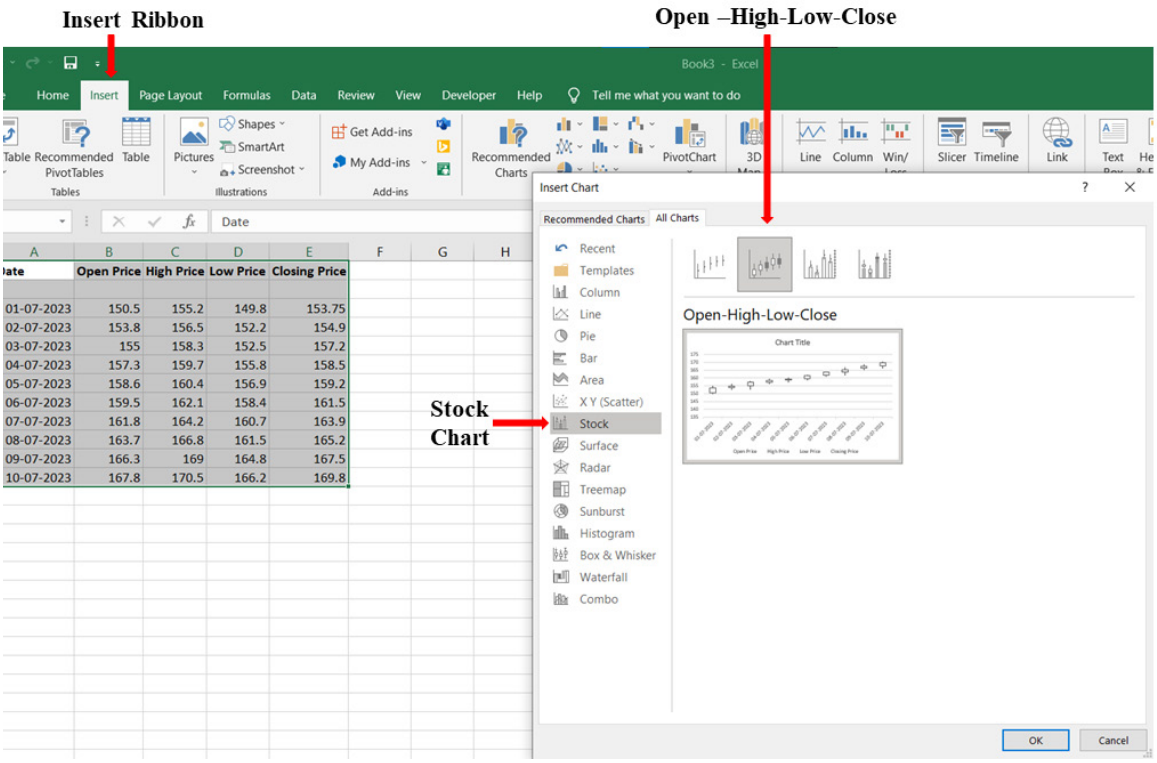


Fig 2.1.37 Inserting Stock Chart

- Select the desired stock chart type, such as “High-Low-Close” or “Open-High-Low-Close.” Here we have selected “Open- High-Low-Close” since we have all the 4 sets of data.
- Excel will generate a Stock Chart using the selected data set as shown below.

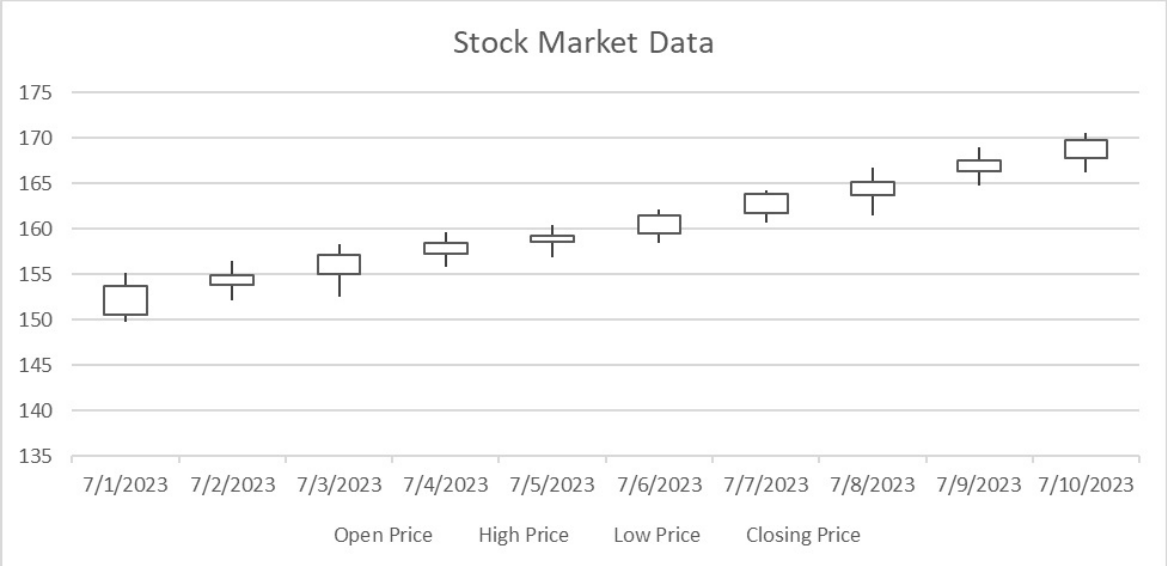


Fig 2.1.38 Stock Chart

You can customise the chart by right-clicking on it and selecting the "Format Chart Area" option to modify the chart title, axes labels, formatting, and other elements.

- Stock chart visualises financial data

Stock charts display valuable information about the performance and behaviour of a stock in the financial markets.

2.1.5.4 Box and Whisker Plot

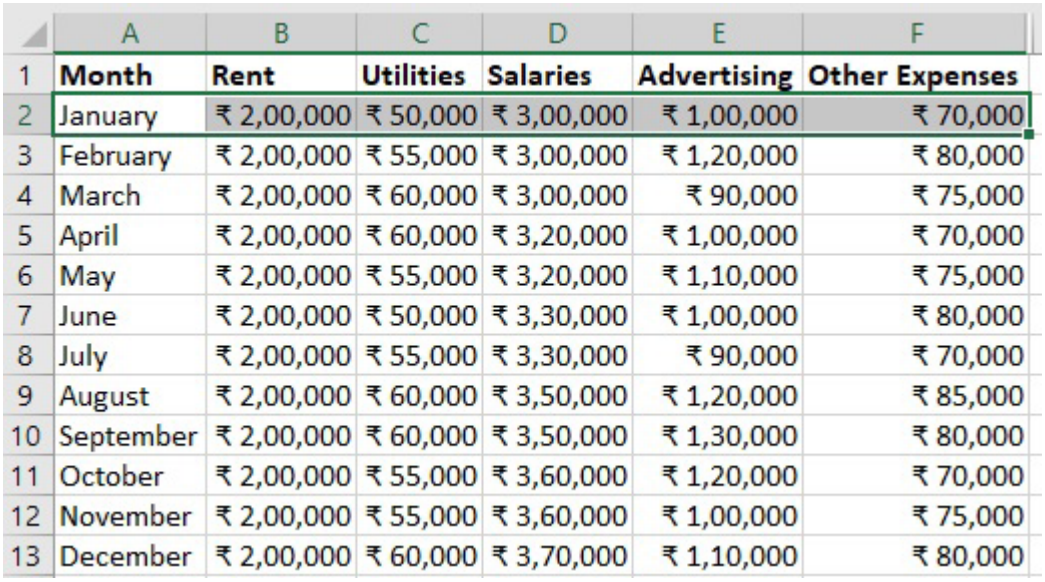
A box and whisker plot is used to display the distribution of a dataset and to identify outliers. It provides insights into the range, median, quartiles, and potential skewness of the data.

Steps to insert a Box & Whisker chart are as follows:

- Enter the data set given below in the Table 2.1.6 into Excel
- Select the column range of data, including the headers and values. Here we are selecting the second row "January" as shown in the Fig 2.1.39

Table 2.1.6 Company Expense Dataset

Month	Rent	Utilities	Salaries	Advertising	Other Expenses
January	₹ 2,00,000	₹ 50,000	₹ 3,00,000	₹ 1,00,000	₹ 70,000
February	₹ 2,00,000	₹ 55,000	₹ 3,00,000	₹ 1,20,000	₹ 80,000
March	₹ 2,00,000	₹ 60,000	₹ 3,00,000	₹ 90,000	₹ 75,000
April	₹ 2,00,000	₹ 60,000	₹ 3,20,000	₹ 1,00,000	₹ 70,000
May	₹ 2,00,000	₹ 55,000	₹ 3,20,000	₹ 1,10,000	₹ 75,000
June	₹ 2,00,000	₹ 50,000	₹ 3,30,000	₹ 1,00,000	₹ 80,000
July	₹ 2,00,000	₹ 55,000	₹ 3,30,000	₹ 90,000	₹ 70,000
August	₹ 2,00,000	₹ 60,000	₹ 3,50,000	₹ 1,20,000	₹ 85,000
September	₹ 2,00,000	₹ 60,000	₹ 3,50,000	₹ 1,30,000	₹ 80,000
October	₹ 2,00,000	₹ 55,000	₹ 3,60,000	₹ 1,20,000	₹ 70,000
November	₹ 2,00,000	₹ 55,000	₹ 3,60,000	₹ 1,00,000	₹ 75,000
December	₹ 2,00,000	₹ 60,000	₹ 3,70,000	₹ 1,10,000	₹ 80,000



	A	B	C	D	E	F
1	Month	Rent	Utilities	Salaries	Advertising	Other Expenses
2	January	₹ 2,00,000	₹ 50,000	₹ 3,00,000	₹ 1,00,000	₹ 70,000
3	February	₹ 2,00,000	₹ 55,000	₹ 3,00,000	₹ 1,20,000	₹ 80,000
4	March	₹ 2,00,000	₹ 60,000	₹ 3,00,000	₹ 90,000	₹ 75,000
5	April	₹ 2,00,000	₹ 60,000	₹ 3,20,000	₹ 1,00,000	₹ 70,000
6	May	₹ 2,00,000	₹ 55,000	₹ 3,20,000	₹ 1,10,000	₹ 75,000
7	June	₹ 2,00,000	₹ 50,000	₹ 3,30,000	₹ 1,00,000	₹ 80,000
8	July	₹ 2,00,000	₹ 55,000	₹ 3,30,000	₹ 90,000	₹ 70,000
9	August	₹ 2,00,000	₹ 60,000	₹ 3,50,000	₹ 1,20,000	₹ 85,000
10	September	₹ 2,00,000	₹ 60,000	₹ 3,50,000	₹ 1,30,000	₹ 80,000
11	October	₹ 2,00,000	₹ 55,000	₹ 3,60,000	₹ 1,20,000	₹ 70,000
12	November	₹ 2,00,000	₹ 55,000	₹ 3,60,000	₹ 1,00,000	₹ 75,000
13	December	₹ 2,00,000	₹ 60,000	₹ 3,70,000	₹ 1,10,000	₹ 80,000

Fig 2.1.39 Selecting Expense Data

- Go to the “Insert” tab in the Excel ribbon and click on the “Box and Whisker” chart option, from the "All Charts" option in "Charts" group, as shown in figure given below

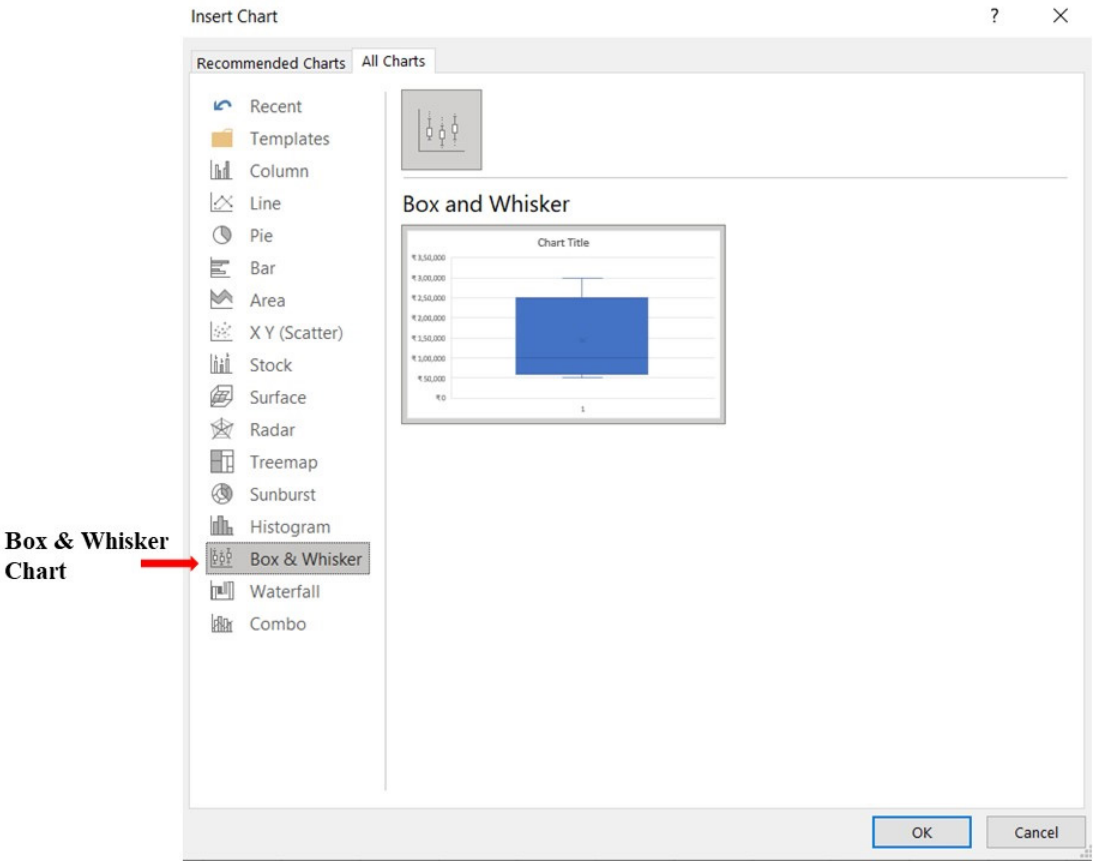


Fig 2.1.40 Inserting Box & Whisker Chart

- Excel will generate a default Box and Whisker chart as shown in Fig 2.1.13.

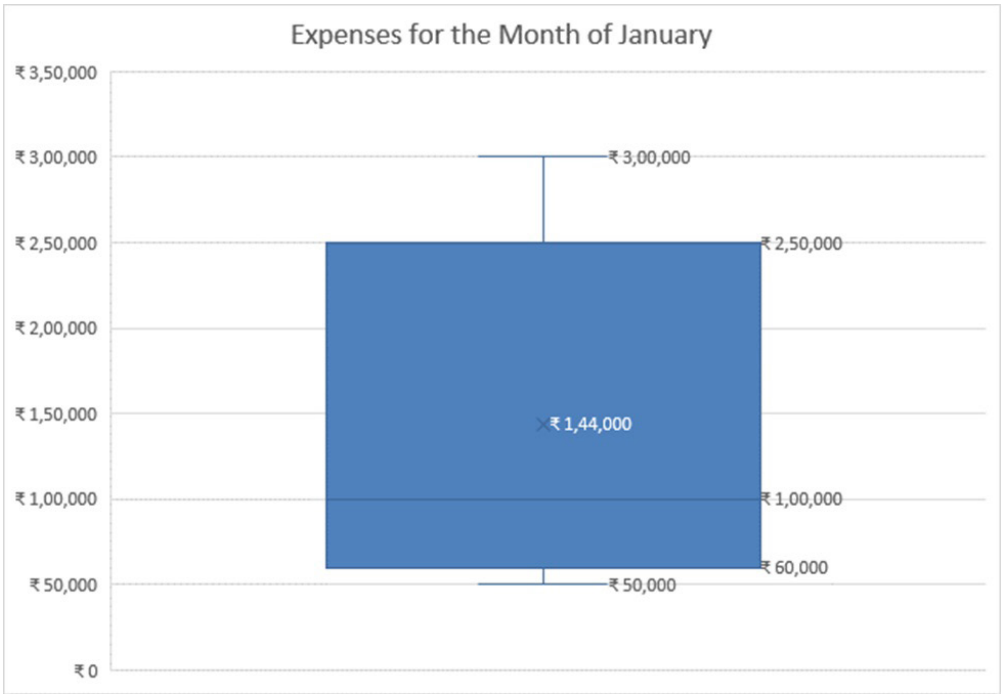


Fig 2.1.41 Box & Whisker Chart

- Box and Whisker plot represents the distribution of a dataset

- You can customise the chart by right-clicking on it and selecting the "Format Data Series" option. Modify the axis labels, add data labels, adjust the chart title, and other formatting options as desired.
- The Box & Whisker chart displays the following information:
    - Minimum: ₹ 50,000
    - Maximum: ₹ 3,00,000
    - Mean: ₹ 1,44,000
    - Median: ₹ 1,00,000
    - Lower Quartile (Q1): ₹ 60,000
    - Upper Quartile (Q3): ₹ 2,50,000
- The data can be generated for all the other months in the same manner.

A box and whisker chart shows distribution of data into quartiles, highlighting the mean and outliers. The boxes may have lines extending vertically called "whiskers". These lines indicate variability outside the upper and lower quartiles, and any point outside those lines or whiskers is considered an outlier.

2.1.6 Applying formatting options in Excel to enhance Visual Representation

Excel offers a wide range of formatting options to enhance the visual representation of data, making it more attractive, readable, and insightful. Let us dive into the key formatting options available in Excel and how to use them:

1. **Number Formatting:** Number formatting allows you to change the display format of numerical data. You can apply decimal places, thousands of separators, currency symbols, and percentage formatting. To access number formatting options, select the cells or data range you want to format, right-click, and choose "Format Cells." In the "Format Cells" dialog box, go to the "Number" tab to select the desired formatting options.
2. **Font and Alignment:** Customising font styles, sizes, colors, and alignments can significantly impact the appearance of your data. You can use bold, italic, underline, and change font colors to highlight important information. Aligning data to the left, right, or center can improve readability. The font and alignment options are available on the Home tab of the Excel ribbon.

3. **Conditional Formatting:** Conditional formatting allows you to apply different formatting styles based on specific conditions or rules. For example, you can highlight cells that meet a certain criterion, create colour scales to visualise data ranges, or apply data bars to show data proportions. Conditional formatting can be accessed through the "Conditional Formatting" button on the Home tab.
4. **Cell Borders and Fill Colours:** Adding borders and fill colours to cells can help separate data and make it easier to read. You can choose from various border styles, such as solid lines, dashed lines, or double lines, and apply different fill colours to cells. These options are available on the Home tab under the "Border" and "Fill colour" buttons.
5. **Chart Styles and Colours:** Excel offers a variety of pre-defined chart styles and colour schemes that you can apply to your charts. These styles can help make your charts look more professional and consistent with the overall theme of your data presentation. To change chart styles and colours, right-click on the chart and select "Change Chart Type" or "Format Chart Area."
6. **Data Labels and Data Point Markers:** Data labels are text annotations that display the values of data points directly on the chart. You can customise data labels to show various information, such as the actual values, percentages, or custom labels. Additionally, you can use data point markers, such as circles or squares, to make individual data points stand out and improve the readability of your chart.
7. **Axis Formatting:** Formatting options for chart axes allow you to adjust the appearance of the X and Y axes. You can change the axis labels, add a title, adjust the scale, and customise the line style and colour. Properly formatted axes can help provide context to your data and make it easier for viewers to understand the data distribution and trends.
8. **Legend and Data Series Formatting:** The legend is an important component of a chart that identifies the different data series represented in the chart. You can format the legend to change its position, font, size, and colour to ensure it complements the overall look of the chart. Additionally, you can format each data series individually, such as changing the colour, line style, and marker shape, to distinguish them from one another and highlight specific data.

- Formatting enhance readability and convey information effectively

9. **Chart Titles and Labels:** Adding a descriptive title to your chart and labelling axes is essential for clear communication of the data's context and meaning. You can format chart titles and axis labels by changing the font, size, and alignment. Well-formatted titles and labels provide context and make the chart more accessible to your audience.
10. **Chart Area and Background:** Formatting options for the chart area and background allow you to customise the look and feel of the entire chart. You can set different background colors or add images to make the chart visually appealing and fit seamlessly into your overall data visualisation design.
11. **Sparklines**: Sparklines are mini-charts that can be inserted into individual cells. They are useful for displaying trends and patterns in a compact format. Excel offers various formatting options for sparklines, such as line style, marker shape, and axis formatting.
12. **Gridlines and Tick Marks:** Gridlines and tick marks provide reference lines on the chart, making it easier to read and interpret data points. You can format gridlines by changing their colour, style, and visibility to improve the chart's visual clarity.
13. **Themes and Styles**: Excel provides pre-defined themes and styles that can be applied to your entire workbook, including charts and tables. Themes allow you to quickly change the colour scheme, fonts, and effects across all your visual elements, creating a consistent look and feel.
14. **Data Bars and Icon Sets:** Data bars and icon sets are options within conditional formatting that allow you to visually represent data using coloured bars or icons based on cell values. Data bars are useful for comparing data within a range, while icon sets can display symbols (e.g., arrows, traffic lights) to represent data variations.
15. **Shapes and Text Boxes:** Excel allows you to insert shapes and text boxes to annotate or draw attention to specific data points or sections within your worksheet. You can format these shapes by changing their colours, sizes, and styles.

- Formatting options include chart styles, axis formatting, data labels, legends etc

Remember to balance aesthetics with clarity when formatting your charts. Ensure that the formatting choices you make enhance the readability of the data and assist in conveying your message effectively. Experiment with different formatting options to find the best combination that suits your data and enhances its visual impact.

Summarised Overview

Data analysis is a critical process in statistics and decision-making, involving the examination, interpretation, and transformation of data into meaningful insights. It plays a pivotal role in various industries, helping organisations make data-driven decisions, identify opportunities, and gain a competitive edge. Through statistical and computational techniques, data analysis allows us to explore datasets, draw conclusions, and solve real-world problems. Descriptive statistics is a fundamental branch of statistics that involves collecting, organising, analysing, and presenting data to provide a comprehensive summary of its main features. It encompasses measures of central tendency (mean, median, mode), measures of variability (range, variance, standard deviation), and measures of shape (skewness, kurtosis). Graphical representations, such as histograms and scatter plots, aid in visualising the data. The Data Analysis ToolPak in Microsoft Excel is an add-in that offers a wide range of statistical functions and tools for data analysis. It simplifies the process of generating descriptive statistics, making it accessible to users within the Excel environment. With the ToolPak, users can easily calculate various statistical measures without complex programming, aiding in data exploration and decision-making.

In Excel, you can create advanced charts like the Waterfall Chart, useful for showing cumulative effects of positive and negative values over time; Radar Chart, suitable for comparing multiple variables across categories; Treemap Chart, effective for illustrating hierarchical data using nested rectangles; Pareto Chart, combining a column chart and a line graph to identify significant factors contributing to a problem; Heat Map, which uses colours to represent data values across a matrix; Sparkline, a condensed chart fitting within a cell to show trends over time; Bubble Chart, displaying three dimensions of data using bubble sizes and colours. Formatting options include number formatting, font and alignment customisation, conditional formatting, chart styles and colours, data labels, axis formatting, legend and data series formatting, chart titles and labels, chart area and background formatting, and more. Remember to balance aesthetics and clarity when formatting charts to enhance readability and convey information effectively.

Self-Assessment

1. Explain the purpose and importance of generating descriptive statistics using Excel functions in data analysis. Provide examples of at least two common Excel functions used for this purpose.
2. Describe the process of creating a scatter plot and its applications in real-world data analysis.
3. Explain how applying formatting options can enhance the visual repre sentation of data in charts and graphs.
4. Discuss the advantages and limitations of creating advanced charts and graphs in data visualisation.
5. What is data analysis toolpak in Excel, and how do they help in data analysis in Excel?

Assignments

1. Using the dataset representing the test scores of two groups of students (Group A and Group B) given below, and answer the following questions.
   a. Calculate the mean, standard deviation, and variance of each group's test scores.
   b. Compare the variability between the two groups and discuss which group has more consistent performance.

Group A Test Scores:	90, 85, 78, 92, 88, 80, 95, 87, 89, 86
Group B Test Scores:	70, 75, 72, 68, 80, 73, 78, 74, 77, 72

2. Using data for three regions over the past year given below, and answer the following questions.
   a. Calculate the total sales for each region and create a 3D clustered column chart to compare their performance visually.
   b. Include a trendline in the chart and analyse the growth patterns.

Month	Region A	Region B	Region C
Jan	50	60	70
Feb	55	65	75
Mar	60	70	80
Apr	65	75	85
May	70	80	90
Jun	75	85	95
Jul	80	90	100
Aug	85	95	105
Sep	90	100	110
Oct	95	105	115
Nov	100	110	120
Dec	105	115	125

3. Using the dataset given below, containing the monthly sales data (in INR) for a retail store for the past year, Calculate descriptive statistics on this dataset by
   a. Excel formulas
   b. Data analysis toolpak

Month	Sales (INR)
Jan	1000
Feb	1200
Mar	900
Apr	1500
May	1800
Jun	1350
Jul	2000
Aug	1750
Sep	1600
Oct	1300
Nov	1100
Dec	1900

4. Consider the dataset of employee salaries in a company given below and answer the following questions.

a. Create a frequency distribution of salary ranges.

b. Apply appropriate formatting options to enhance the visual representation of he histogram.

c. Analyse the distribution and comment on the salary ranges where most em ployees lie.

Employee ID	Salary (INR)
1	45000
2	52000
3	48000
4	50000
5	54000
6	60000
7	58000
8	55000
9	49000
10	53000
11	47000
12	51000
13	57000
14	52000
15	50000
16	54000
17	52000
18	49000
19	46000
20	51000

5. Using the data collected on customer satisfaction for a retail store is given be low, and answer the following questions.

a. Design a radar chart in Excel to display the ratings of five different aspects of customer experience (e.g., product quality, customer service, pricing, etc.).

b. Describe how this radar chart can help identify the store's strengths and areas for improvement.

Aspect	**Rating (out of 5)**
Product Quality	4.2
Customer Service	3.8
Pricing	3.5
Store Ambiance	4
Product Variety	4.5

Reference

1. Schwabish, J. (2023). *Data Visualisation in Excel: A Guide for Beginners, Intermediates, and Wonks.* CRC Press.
2. Walkenbach, J. (2019). *Excel 2019 Bible.* John Wiley & Sons, Inc.
3. Held, B., Moriarty, B., & Richardson, T. (2019). *Microsoft Excel Functions and Formulas with Excel 2019/Office 365*. Mercury Learning and Information.
4. Frye, C. (2018). *Microsoft Excel 2019 step by step*. Microsoft Press.
5. Wilke, C. O. (2019). *Fundamentals of data visualisation: a primer on making informative and compelling figures.* O'Reilly Media.

Suggested Reading

1. Rudraksh, N., & McDonald, A. (2020). *Data Analysis Using Microsoft Excel: Updated for Office 365 (2nd ed.).* Cambridge University Press

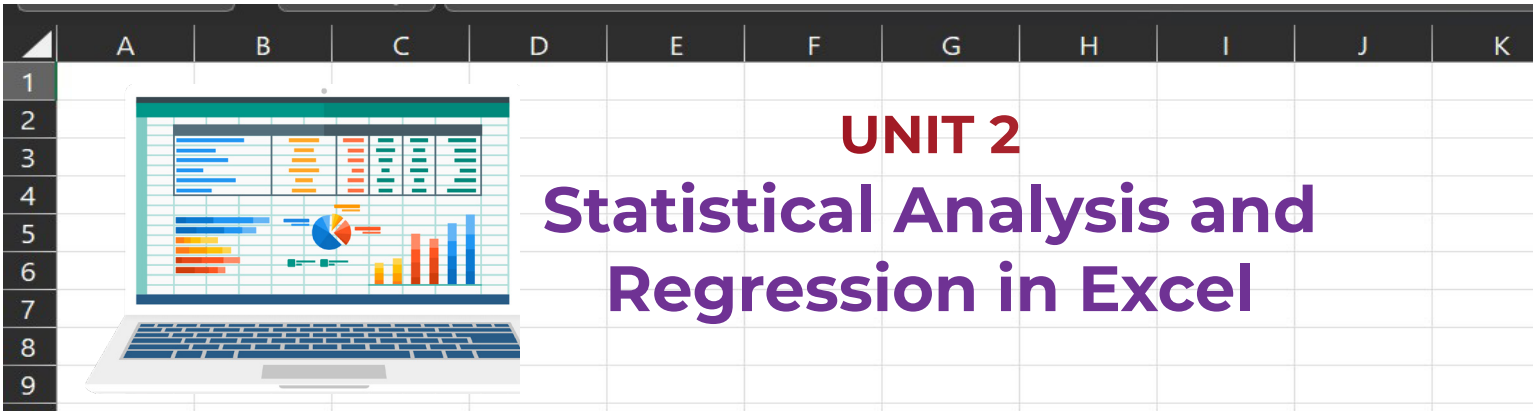
2. Dutt, R. (2020). E*xcel Charts and Graphs: A Simple Guide to Excel Data Visualisation*. Apress

3. Chambers, J. M., Cleveland, W. S., Kleiner, B., & Tukey, P. A. (1983). *Graphical Methods for Data Analysis.* Wadsworth & Brooks/Cole.

4. Few, S. (2012). *Show Me the Numbers: Designing Tables and Graphs to Enlighten.* Analytics Pres

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.



UNIT 2
Statistical Analysis and Regression in Excel

Learning Outcomes

After the completion of this unit, the learner will be able to

- understand and perform hypothesis testing
- conduct t-tests for means comparison
- perform chi-square tests for categorical data
- build regression models
- analyse regression results and interpret coefficients

Background

Statistical analysis is a systematic and objective method of examining data to uncover patterns, relationships, and insights that can inform decision-making and problem-solving. It involves a series of techniques and procedures designed to transform raw data into meaningful information. The process typically begins with data collection, where relevant information is gathered and organised for analysis. Next, data cleaning and preparation are conducted to address errors, missing values, and inconsistencies, ensuring the data is accurate and reliable. Once the data is prepared, descriptive statistics are employed to summarise and describe the main features of the dataset. Descriptive statistics provide a clear overview of the data's characteristics, allowing researchers to understand its distribution and gain initial insights.

Beyond descriptive statistics, statistical analysis includes inferential statistics, which involve drawing conclusions or making predictions about a larger population based on a sample of data. Hypothesis testing is a fundamental aspect of inferential statistics, where researchers formulate hypotheses and use statistical tests to determine wheth-

er there is enough evidence to support or reject these hypotheses. Various methods, such as t-tests, ANOVA, regression analysis, and chi-square tests, are applied to assess relationships between variables, compare groups, and measure the significance of findings.

Statistical analysis is widely used across diverse fields, including science, social sciences, finance, healthcare, and business, to name a few. It plays a crucial role in research, allowing scientists to validate theories and draw generalisations from their data. In business and industry, statistical analysis helps optimise processes, identify market trends, and make data-driven decisions to gain a competitive edge. As data continues to grow in complexity and volume, the importance of statistical analysis will only continue to rise, making it an indispensable tool for extracting valuable insights and knowledge from the vast sea of information in today's data-driven world.

Keywords

Hypothesis Testing, t-Test, Chi-Square Test, Regression Analysis

Discussion

2.2.1 Introduction to Hypothesis Testing

Hypothesis testing is a statistical method used to make inferences or draw conclusions about a population based on sample data. It involves comparing the observed data with a hypothesis or claim about the population to determine if there is enough evidence to support or reject that hypothesis. Here is a detailed explanation of the key steps involved in hypothesis testing:

Step 1: State the null and alternative hypotheses:

The first step is to clearly state the null hypothesis (H_0) and the alternative hypothesis (H_a). The null hypothesis represents the assumption or claim that we want to test, while the alternative hypothesis represents the opposite or alternative to the null hypothesis. For example, if we want to test whether a new drug is effective in treating a disease, the null hypothesis (H_0) could be "The drug has no effect," while the alternative hypothesis (H_a) could be "The drug is effective."

Step 2 : Set the significance level (α):

The significance level, denoted by α (alpha), determines the threshold for rejecting the null hypothesis. It represents the maximum acceptable probability of making a Type I error, which is the error of rejecting the null hypothesis when it is actually true. Commonly used values for α are 0.05 (5%) or 0.01 (1%), but the choice of significance level depends on the specific context and the consequences of making a Type I error.

Step 3: Gather sample data:

To perform hypothesis testing, we need a representative sample from the population of interest. The sample should be collected using appropriate sampling methods to ensure it accurately represents the population.

Step 4: Choose an appropriate test statistic:

The choice of a test statistic depends on the nature of the data and the hypothesis being tested. Commonly used test statistics include the t-statistic, z-statistic, and chi-square statistic. For example, if we are comparing the means of two independent samples, the appropriate test statistic could be the t-statistic. If we are testing the proportion of a categorical variable, the z-statistic might be used.

Step 5: Conduct the statistical test:

Using the sample data and the chosen test statistic, we calculate the test statistic value. The specific calculation depends on the chosen test statistic and the nature of the hypothesis being tested.

Once we have the test statistic value, we compare it to the critical value(s) or calculate the p-value.

Step 6: Determine the critical value or calculate the p-value:

The critical value is a threshold value that determines the region of rejection for the null hypothesis. If the test statistic falls in the region of rejection, we reject the null hypothesis in favor of the alternative hypothesis. Alternatively, we can calculate the p-value, which is the probability of obtaining a test statistic as extreme as the one observed, assuming the null hypothesis is true. The p-value represents the strength of the evidence against the null hypothesis.

Step 7: Interpret the results:

Based on the critical value or the p-value, we interpret the results. If the test statistic falls in the region of rejection or the

p-value is less than the significance level (α), we reject the null hypothesis. This suggests that there is enough evidence to support the alternative hypothesis. If the test statistic does not fall in the region of rejection or the p-value is greater than α, we fail to reject the null hypothesis. This means that there is not enough evidence to support the alternative hypothesis, and we do not have sufficient grounds to conclude that the null hypothesis is false.

- Statistical method used to determine if there is enough evidence to support or reject a hypothesis

Step 8: Draw conclusions:

Based on the results of the hypothesis test, we draw conclusions about the hypothesis being tested. If we reject the null hypothesis, we conclude that there is evidence to support the alternative hypothesis. If we fail to reject the null hypothesis, we conclude that there is not enough evidence to support the alternative hypothesis. It is important to note that hypothesis testing provides evidence for or against a hypothesis, but it does not prove or disprove it conclusively. The conclusions are based on the observed data and the statistical analysis conducted.

Hypothesis testing can be performed using various statistical tests in Excel, such as the T.TEST for comparing means, Z.TEST for comparing proportions, and CHISQ.TEST for chi-square tests. These Excel functions simplify the calculation of test statistics and p-values, making hypothesis testing more accessible.

2.2.2 Parametric and Non-Parametric Tests

Parametric and non-parametric tests are two broad categories of statistical tests used in hypothesis testing and data analysis. They differ in their assumptions about the underlying distribution of the data. They are explained as follows :

Parametric Tests:

- Parametric tests rely on parameters such as mean and variance

Parametric tests are based on specific assumptions about the distribution of the data, usually assuming a known or specified probability distribution, such as the normal (Gaussian) distribution. These tests rely on parameters, such as means and variances, to make inferences about the population. Parametric tests are powerful when the data meet the assumptions of the specified distribution. Some commonly used parametric tests include t-tests, ANOVA, linear regression, and parametric correlation tests. These tests generally require that the data are continuous and have a specific distri-

bution, and they provide precise estimates of parameters and statistical significance if the assumptions are met.

Non-Parametric Tests:

Non-parametric tests, also known as distribution-free tests, make fewer assumptions about the underlying distribution of the data. These tests are designed to analyse data that may not adhere to specific distributional assumptions or have unknown distributions. Non-parametric tests are often used when the data are ordinal, skewed, or have outliers. They rely on fewer assumptions and instead focus on the ranks or order of the data values. Non-parametric tests are robust to violations of distributional assumptions and provide more general statistical inference. Examples of non-parametric tests include the Mann-Whitney U test, Wilcoxon signed-rank test, Kruskal-Wallis test, and Spearman's rank correlation test. These tests compare medians, ranks, or relationships between variables without assuming a specific distribution.

- Non-parametric tests focus on ranks or order of data values

2.2.3 Common Types of Hypothesis Tests

The major types of hypothesis testing are as follows :

Tests for Comparing Means

- One-sample z-test
- Two-sample z-test
- One-sample t-test
- Two-sample t-test
- Paired t-test

Tests for Comparing Proportions

- One-sample proportion test
- Two-sample proportion test

Tests for Comparing Variance

- Chi-square test for variance
- F-test for variance

Other Common Tests

- Goodness of fit test
- Testing for independence of two attributes (Contingency Tables
- ANOVA (Analysis of Variance)

Fig 2.2.1 Types of Hypothesis Tests

- One-sample tests compare a single sample to a known or hypothesised population parameter

Let us explain the differences between one-sample, two-sample, and more than two-sample tests:

1. **One-Sample Test:** A one-sample test compares the data from a single sample to a known or hypothesised population parameter. The goal is to determine if the sample data significantly differ from the population or a specified value. The most common one-sample test is the one-sample t-test, which compares the sample mean to a known population mean. It assesses if there is a significant difference between the two means.
2. **Two-Sample Test:** A two-sample test compares the data from two independent samples to assess if there is a significant difference between their means or other relevant parameters. The two samples can be from different populations, different groups, or represent different conditions. When it comes to two-sample tests, we have two main categories: paired tests and unpaired tests.

   - **Paired Tests:** Paired tests, also known as dependent samples or matched pairs tests, are used when the data in the two samples are not independent but are related or paired in some way. The pairing could arise from a before-and-after scenario, repeated measures on the same individuals, or matching subjects based on specific criteria. In a paired test, the differences between the paired observations are analysed. The focus is on the change or the relationship within each pair rather than comparing the absolute values of the two groups. Paired tests increase the statistical power by accounting for the within-pair variation and reducing the effect of individual differences. Common examples of paired tests are the paired samples t-test, which compares the means of two paired samples, and the Wilcoxon signed-rank test, a non-parametric alternative for comparing the medians of paired samples.
   - **Unpaired Tests:** Unpaired tests, also known as independent samples tests, are used when the data in the two samples are independent of each other. The samples are from different groups or conditions, and there is no inherent pairing or relationship between the observations. In an unpaired test, the focus is on comparing the means or other relevant parameters between the two independent groups. These tests assess if there is a significant difference in the distributions or characteristics of the two groups. Common exam-

ples of unpaired tests are the independent samples t-test, which compares the means of two independent groups, and the Mann-Whitney U test, a non-parametric alternative for comparing the medians of two independent groups. The choice between paired and unpaired tests depends on the study design, the relationship between the observations, and the research question. Paired tests provide greater sensitivity when there is within-pair correlation or dependency, while unpaired tests are suitable when the samples are independent.

- Two-sample tests compare two independent samples

3. **More Than Two-Sample Test:** A more than two-sample test extends the comparison to three or more independent samples. It assesses if there are significant differences among the means or other parameters of the multiple groups. The most common test in this category is the analysis of variance (ANOVA), which compares the means of multiple groups simultaneously. ANOVA determines if there are significant differences among the means and is typically followed by post-hoc tests to identify specific group differences.

- More than two-sample tests extend the comparison to three or more independent samples.

When choosing the appropriate test, consider the number of groups you want to compare and the nature of the research question. One-sample tests are suitable when comparing a sample to a known or hypothesised value. Two-sample tests are used when comparing two independent groups. More than two-sample tests, such as ANOVA, are employed when comparing the means or parameters of three or more independent groups.

It is important to select the correct test based on the specific research question, the type of data, and the assumptions associated with each test.

Let us dive into each test in detail:

1. **Z-Test:** The Z-test is a statistical test used to compare a sample mean to a known population mean when the population standard deviation is known. It assumes that the sample data are independent and approximately normally distributed. The test statistic is calculated as (sample mean - population mean) / (population standard deviation / √sample size). The Z-test is appropriate for large sample sizes (typically, greater than 30) and is based on

- Z-Test compares a sample mean to a known population mean

the standard normal distribution (Z-distribution).

- Student's t-Test compares means of two independent or paired samples when the population standard deviation is unknown

2. **Student's t-Test:** The Student's t-test, or simply t-test, is a statistical test used to compare the means of two independent samples or paired samples when the population standard deviation is unknown. It assumes that the sample data are independent and approximately normally distributed. The t-test calculates the t-statistic, which measures the difference between the means relative to the variability within the samples. The t-test uses the t-distribution to determine statistical significance and is more suitable for small sample sizes (typically, less than 30).

- F-Test compare variances between groups in ANOVA

3. **F-Test:** The F-test is a statistical test used to compare the variances of two or more groups or to test the overall significance of a regression model. In ANOVA (Analysis of Variance), the F-test is used to compare the variances between groups, determining if there are significant differences in the means. The test statistic follows the F-distribution. In regression analysis, the F-test assesses the overall significance of the regression model by comparing the variability explained by the model to the residual variability.

- Chi-Square Test examine the independence or association between two categorical variables

4. **Chi-Square Test:** The Chi-Square test is a statistical test used to assess the independence or association between two categorical variables. It is used to determine if there is a significant relationship or difference between the observed and expected frequencies of the variables. The test calculates the Chi-Square statistic, which follows the Chi-Square distribution. The test is suitable for analysing categorical data and is often used in contingency tables or to assess goodness-of-fit to a specific distribution.

- ANOVA compares the means of two or more groups simultaneously

5. **Analysis of Variance (ANOVA):** ANOVA is a statistical test used to compare the means of two or more groups simultaneously. It assesses if there are significant differences among the means and determines if the group differences are statistically significant. ANOVA partitions the total variation in the data into variation between groups and within groups. It calculates the

F-statistic, comparing the variation between groups to the variation within groups. ANOVA is commonly used when there are categorical independent variables and a continuous dependent variable.

6. **Wilcoxon Signed-Rank Test:** The Wilcoxon signed-rank test is a non-parametric test used to compare the medians of two related or paired samples. It is commonly used when the data do not meet the assumptions of parametric tests, such as normality or equal variances. The test assesses if there is a significant difference between the paired observations by comparing the ranks of the differences. It is robust to outliers and can handle skewed data. The test provides a p-value indicating the statistical significance of the observed differences.

- Wilcoxon Signed-Rank **Test** compares the medians of two related samples

7. **Runs Test:** The runs test is a non-parametric test used to analyse the randomness or lack of pattern in a sequence of observations. It assesses if there are significant departures from randomness in the data. The test counts the number of runs (sequences of consecutive increasing or decreasing values) in the data and compares it to the expected number of runs under the assumption of randomness. The test is often used in quality control or to check for serial dependence in time series data.

- Runs Test examines the randomness or patterns in a sequence of observations

8. **Kolmogorov-Smirnov Test:** The Kolmogorov-Smirnov test (KS test) is a non-parametric test used to assess whether a sample follows a specified distribution. It compares the empirical distribution function (EDF) of the sample to the cumulative distribution function (CDF) of the specified distribution. The test calculates the maximum difference (D-statistic) between the EDF and the CDF and determines if it exceeds the critical value. The KS test is used to test the goodness-of-fit of a sample to a particular distribution or compare two samples' distributions.

- Kolmogorov-Smirnov Test evaluates whether a sample follows a given distribution

9. **Mann-Whitney U Test:** The Mann-Whitney U test, also known as the Wilcoxon rank-sum test, is a non-parametric test used to compare the medians of two independent groups. It assesses if there is a significant difference between the distributions of the two groups. The test ranks

- Mann-Whitney U Test compares the medians of two independent groups

- Spearman Rank Correlation assess the strength and direction of the monotonic relationship between two variables

- Pearson's Correlation assess the strength and direction of the linear relationship between two continuous variables

all the observations from both groups, calculates the U-statistic (the sum of ranks in one group), and compares it to the critical values. The Mann-Whitney U test is appropriate when the assumptions of parametric tests, such as normality and equal variances, are not met.

10. **Spearman Rank Correlation:** Spearman's rank correlation coefficient is a non-parametric measure of the strength and direction of the monotonic relationship between two variables. It assesses the correlation between variables based on the ranks of the observations. The Spearman correlation coefficient ranges from -1 to +1, where -1 indicates a perfect negative monotonic relationship, +1 indicates a perfect positive monotonic relationship, and 0 indicates no monotonic relationship. The test provides a p-value to determine if the observed correlation is statistically significant.
11. **Pearson's Correlation:** Pearson's correlation coefficient is a parametric measure of the linear relationship between two continuous variables. It assesses the strength and direction of the linear association. The Pearson correlation coefficient ranges from -1 to +1, where -1 indicates a perfect negative linear relationship, +1 indicates a perfect positive linear relationship, and 0 indicates no linear relationship. The test provides a p-value to determine if the observed correlation is statistically significant.

2.2.4 Important Terms in Hypothesis Testing

Understanding the following concepts is essential for conducting hypothesis testing and interpreting the results accurately.

2.2.4.1 P-Value

The p-value, short for "probability value," is a fundamental concept in hypothesis testing. It quantifies the strength of evidence against the null hypothesis (H_0). Specifically, the p-value represents the probability of obtaining the observed results, or more extreme results, assuming that the null hypothesis is true. In other words, it measures the likelihood of observing the data under the assumption that there is no

effect or difference (no statistical significance).

If the p-value is small (typically less than the chosen significance level, alpha), it suggests strong evidence against the null hypothesis. This leads to the rejection of H_0 in favor of the alternative hypothesis (H_a), indicating that the observed effect or difference is likely not due to chance alone. Here the result is statistically significant. If the p-value is large (typically greater than alpha), it suggests weak evidence against H_0. In this case, there is insufficient evidence to reject the null hypothesis, and the results are considered not statistically significant.

- The likelihood of obtaining the observed results

2.2.4.2 Level of Significance (Alpha)

The level of significance, denoted by the symbol alpha (α), is the predetermined threshold used to determine statistical significance in hypothesis testing. It represents the maximum allowable probability of making a Type I error (incorrectly rejecting the null hypothesis when it is true). Commonly used levels of significance are 0.05 (5%) and 0.01 (1%). The choice of alpha depends on the desired balance between Type I and Type II errors and the nature of the research question.

- Determines the probability threshold

If the p-value is less than or equal to alpha, the result is statistically significant, and the null hypothesis is rejected. If the p-value is greater than alpha, the result is not statistically significant, and the null hypothesis is not rejected.

2.2.4.3 One-Tailed (Directional) and Two-Tailed (Non-Directional) Tests:

In hypothesis testing, the choice between a one-tailed and a two-tailed test depends on the research question and the expected direction of the effect.

One-Tailed Test: Also known as a directional test, a one-tailed test is used when the research hypothesis specifies a particular direction of the effect. For example, "The new drug will increase test scores." In this case, the hypothesis is tested against the null hypothesis in only one direction, either greater than or less than. A one-tailed test has more power (greater sensitivity) to detect an effect in the specified direction.

- One-tailed tests are used when the research hypothesis predicts a specific direction of the effect

Two-Tailed Test: Also known as a non-directional test, a two-tailed test is used when the research hypothesis does not specify a particular direction of the effect. For example, "The

new drug will have a different effect on test scores." In this case, the hypothesis is tested against the null hypothesis in both directions, greater than and less than. A two-tailed test is more conservative and allows for detecting effects in either direction.

2.2.4.4 Degrees of Freedom (df):

Degrees of freedom is a concept that arises in statistical calculations, particularly in t-tests and chi-square tests. In hypothesis testing, it represents the number of independent pieces of information available to estimate a parameter or perform a test.

In a one-sample t-test, the degrees of freedom are (n - 1), where "n" is the sample size.

- The number of independent pieces of information available for estimating parameters

In an independent samples t-test, the degrees of freedom depend on the sample sizes of the two groups and are calculated using a formula that considers both sample sizes and assumes equal variances (pooled variance t-test) or unequal variances (Welch's t-test). In a chi-square test, the degrees of freedom depend on the number of categories and constraints in the contingency table. Degrees of freedom play a critical role in determining critical values from the t-distribution and chi-square distribution tables for hypothesis testing.

2.2.4.5 Types of Errors

There are two types of errors that can occur: Type I error and Type II error. Let us understand each of these errors in detail:

Type I Error: A Type I error occurs when the null hypothesis is rejected, even though it is true. In other words, it is a false positive result. This error implies that you have concluded there is a significant effect or difference when, in reality, there is no true effect or difference in the population. The probability of committing a Type I error is denoted by the significance level (α). It represents the maximum tolerable level of chance you are willing to accept to reject the null hypothesis. Commonly used significance levels are 0.05 (5%) and 0.01 (1%). If the p-value (the probability of observing the data given that the null hypothesis is true) is smaller than the chosen significance level, the null hypothesis is rejected, leading to a Type I error.

- Type I error occurs when the null hypothesis is wrongly rejected

Type II Error: A Type II error occurs when the null hypothesis is not rejected, even though it is false. In other words,

it is a false negative result. This error implies that you have failed to detect a significant effect or difference when there is a true effect or difference in the population. The probability of committing a Type II error is denoted by β (beta). It depends on various factors such as sample size, effect size, and the chosen significance level. The power of a statistical test is equal to 1 - β and represents the probability of correctly rejecting the null hypothesis when it is false. Power analysis is often conducted to estimate the sample size required to achieve a desired level of power. It is important to note that Type I and Type II errors are inversely related. Reducing the probability of one type of error increases the probability of the other. Therefore, finding an appropriate balance between Type I and Type II errors is crucial in hypothesis testing. By considering the significance level, effect size, power, and sample size, researchers aim to minimize the likelihood of both Type I and Type II errors to draw accurate conclusions from their statistical analyses.

- Type II error occurs when the null hypothesis is wrongly accepted.

The choice of the appropriate test and the interpretation of the results depend on a clear understanding of p-values, level of significance, the directionality of the test, and degrees of freedom.

2.2.5 T-tests For Means Comparison

T-tests are a group of statistical tests used to compare the means of two groups and assess if there is a significant difference between them. T-tests are commonly employed when working with continuous data and are based on the t-distribution. There are three main types of t-tests for means comparison:

1. One-Sample T-Test:

- One-sample t-test compares the mean of a single sample to a known or hypothesised population mean.

The one-sample t-test compares the mean of a single sample to a known or hypothesised population mean. It assesses whether there is a significant difference between the sample mean and the population mean. This test is appropriate when you have one group and want to determine if its mean is significantly different from a specified value.

Assumptions:

- Data in the sample follow a normal distribution (approximately).

2. Independent Samples T-Test:

The independent samples t-test compares the means of two

- Independent samples t-test compares the means of two independent groups

independent groups to determine if there is a significant difference between them. It is used when the two groups are unrelated and their observations are independent. For example, you might use an independent samples t-test to compare the average test scores of male and female students or to examine if there is a significant difference in sales between two different marketing strategies.

Assumptions:

- Data in each group follow a normal distribution (approximately).
- The two groups have equal variances (homoscedasticity) or similar variances if the sample sizes are similar.

3. Paired Samples T-Test:

- Paired samples t-test, compares the means of two related or paired samples.

The paired samples t-test, also known as the dependent samples or matched pairs t-test, compares the means of two related or paired samples. It is used when the observations in the two groups are dependent or paired. This test is often employed when you have before-and-after measurements, repeated measures on the same subjects, or matching subjects based on specific criteria.

Assumptions:

- The differences between paired observations are approximately normally distributed.
- The differences are independent of each other (no carry-over effects).

It is important to validate the assumptions for each t-test before applying them to the data. If the assumptions are not met, consider using non-parametric alternatives like the Mann-Whitney U test for independent samples or the Wilcoxon signed-rank test for paired samples. Non-parametric tests do not rely on the assumption of normality and are more robust to violations of assumptions.

- Statistical tests used to compare means of two groups

In Excel, there are two versions of the two-sample t-test available: one that assumes equal variances (pooled variance t-test) and another that assumes unequal variances (Welch's t-test). Using t-tests with larger sample sizes provides more reliable results and reduces the impact of non-normality in the data. However, it's essential to remember that the assumptions of normality and equal variances should be verified whenever possible, especially for smaller sample sizes (i.e., sample size essentially below 20), and if they are significantly violated, non-parametric tests or data transforma-

tions may be considered as alternatives.

2.2.5.1 One Sample t-test

The step-by-step instructions for running a one-sample t-test in Excel using the Data Analysis ToolPak is as follows;

Example: A school administrator wants to determine if the average exam score of a group of students differs significantly from the known population mean of 75.

1. Enter the data into Excel

	A
1	**Exam Scores**
2	80
3	85
4	78
5	72
6	88
7	90
8	82
9	79
10	85
11	81
12	

Fig 2.2.2 Entering Data into Excel

2. Set Null and Alternative Hypothesis

   - Null Hypothesis (H_0): The average exam score of the group is equal to the population mean ($\mu = 75$).
   - Alternative Hypothesis (H_a): The average exam score of the group is not equal to the population mean ($\mu \neq 75$).

3. Select the data and go to the "Data" tab in the Excel ribbon. Select the "Data Analysis" option in the "Analysis" group as shown in the Fig 2.2.3 Data Analysis

4. Select Descriptive statistics from the data analysis options as shown in the Fig 2.2.3 Data Analysis

5. Then a new window will open up and click on the input range option to select the input range. Select the data including the title

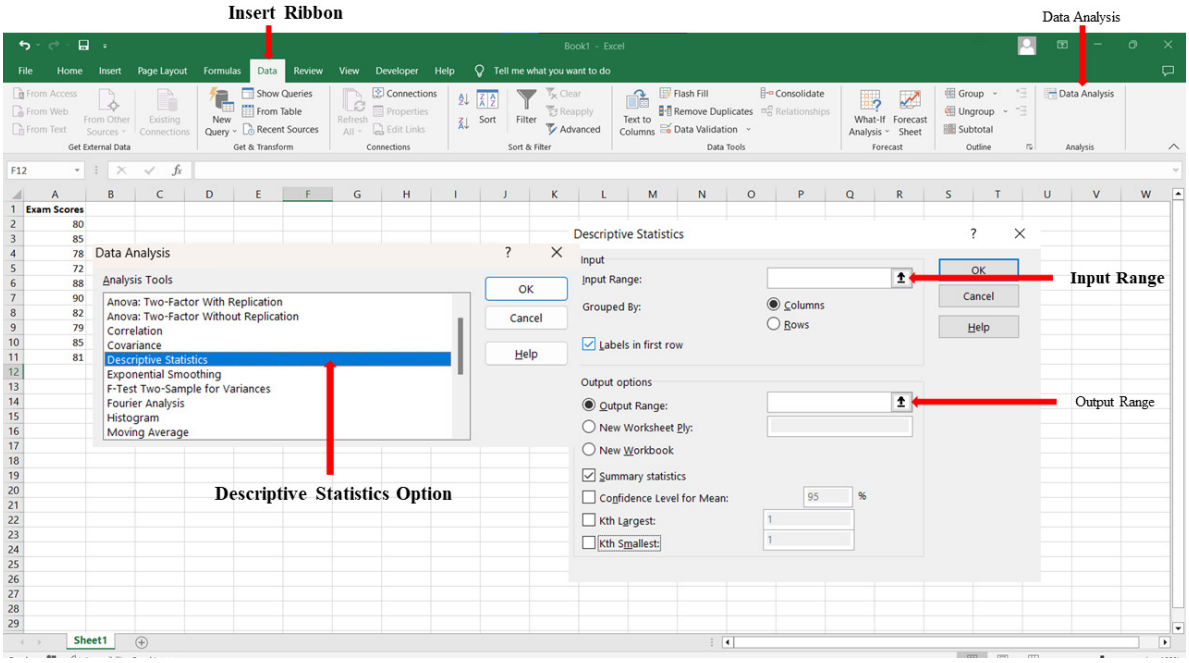


Fig 2.2.3 Data Analysis

6. Click on the output range to display the output. The output can be shown in a new worksheet or workbook. Here we have selected a cell from the current worksheet itself i.e., cell A14
7. The result of the descriptive statistics will be as shown in the Fig 2.2.4

Exam Scores	
Mean	82
Standard Error	1.659986613
Median	81.5
Mode	85
Standard Deviation	5.249338583
Sample Variance	27.55555556
Kurtosis	0.233274026
Skewness	-0.305338007
Range	18
Minimum	72
Maximum	90
Sum	820
Count	10

Fig 2.2.4 Descriptive Statistics Results

8. In order to find the t test statistics, we can use the formula

$$t = \frac{Sample\ mean - Hypothesised\ mean}{Standord\ Error}$$

9. Use the above given formula and calculate the t test statistics. It is shown below.

10. Type out t- statistic in cell C9 and Type = to insert formula in cell D9 Since we already know that the hypothesised mean is 75 and sample mean as per descriptive statistics calculation is 82.2 and standard error is 1.84. Type the formula as shown in Fig 2.2.5 and click Enter

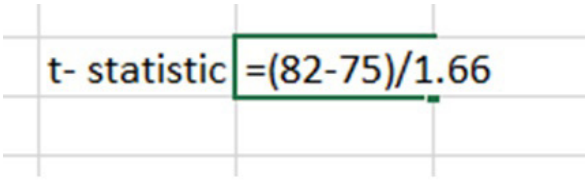


Fig 2.2.5 Inserting Formula

11. The result will be as shown in Fig 2.2.6

Fig 2.2.6 T-Statistic

12. Now we have to calculate p value

13. To calculate click on cell C10 and type p-value

14. In D10 type in the following formula as shown in Fig 2.2.7

Fig 2.2.7 Formulae For P-Value

15. Here the TDIST is the formula, D9 is the value of t-statistics, 9 is the degrees of freedom which is 1 less than the total count and 2 is denoting the 2 tailed test.

The formula for calculating p-value is " =TDIST(t-statistics, degrees of freedom, number of tails in the distribution)

16. The result will be as shown in the Fig 2.2.8

p-value	0.00225

Fig 2.2.8 P-Value

Interpretation of Results:

- If the p-value is less than the chosen significance level (alpha), you can reject the null hypothesis (H_0) and conclude that there is a significant difference between the sample mean and the hypothesised mean.
- If the p-value is greater than the chosen significance level, you fail to reject the null hypothesis, indicating that there is insufficient evidence to conclude a significant difference.

In our example, since the p-value is less than 0.05 (assuming a 5% significance level), we can reject the null hypothesis and conclude that the average exam score of the group is significantly different from the population mean of 75.

- One-sample t-test compares a sample mean to a known population mean

2.2.5.2 Two-Independent Sample t-test

The steps of running a two-independent sample t-test in Excel with a simple example is as follows:

Example:Suppose we want to compare the test scores of two groups of students (Group A and Group B) to see if there is a significant difference in their average scores. Each group has 10 students, and their scores are as follows:

Group A: 78, 85, 92, 88, 79, 90, 86, 80, 84, 89

Group B: 72, 81, 76, 85, 82, 78, 77, 80, 84, 79

1. Enter the data for Group A and Group B in Excel as shown below.

	A	B
1	**Group A**	**Group B**
2	78	72
3	85	81
4	92	76
5	88	85
6	79	82
7	90	78
8	86	77
9	80	80
10	84	84
11	89	79

Fig 2.2. 9 Two-Independent Sample Dataset

2. Set Null and Alternative Hypothesis

- Null Hypothesis (H_0): There is no significant difference in the average test scores between Group A and Group B.
- Alternative Hypothesis (H_a): There is a significant difference in the average test scores between Group A and Group B.

3. Calculate the descriptive statistics for each sample using Data Analysis ToolPak. The results will be as shown below.

	A	B	C	D	E	F	G	H
1	**Group A**	**Group B**						
2	78	72		*Group A*			*Group B*	
3	85	81						
4	92	76		Mean	85.1		Mean	79.4
5	88	85		Standard Error	1.530795		Standard Error	1.231079021
6	79	82		Median	85.5		Median	79.5
7	90	78		Mode	#N/A		Mode	#N/A
8	86	77		Standard Deviation	4.840798832		Standard Deviation	3.893013686
9	80	80		Sample Variance	23.43333333		Sample Variance	15.15555556
10	84	84		Kurtosis	-1.25910293		Kurtosis	0.08194589
11	89	79		Skewness	-0.230526357		Skewness	-0.391520213
12				Range	14		Range	13
13				Minimum	78		Minimum	72
14				Maximum	92		Maximum	85
15				Sum	851		Sum	794
16				Count	10		Count	10

Fig 2.2.10 Descriptive Statistics of Two-Independent Sample

Calculating the descriptive statistics before conducting the two-independent sample t-test served to summarise the data, check assumptions, determine effect size, and facilitate the t-test calculations. These steps are fundamental in conducting a robust and meaningful statistical analysis to compare the test scores of the two groups.

4. Go to the "Data" tab in Excel and Click on "Data Analysis" in the Analysis group
5. Select "t-Test: Two-Sample Assuming Unequal Variances" from the list as shown below and click "OK."
6 In the "t-Test: Two-Sample Assuming Unequal Variances" dialog box, for the "Variable 1 Range," select cells A2:A11 (Group A data) and for the"Variable 2 Range," select cells B2:B11 (Group B data). Leave the "Hypothesised Mean Difference"

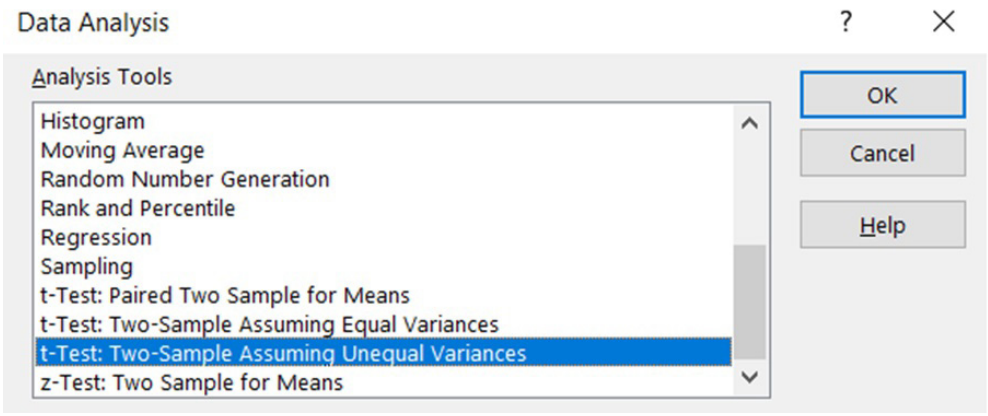


Fig 2.2.11 Selecting Two-Independent Sample T Test

blank (we want to test if the means are significantly different). as shown in Fig 2.2.12.

7. Set "Alpha" to 0.05 (5% significance level). Make sure "Unequal variances assumed" is selected.
8. Click on the output range to display the output. The output can be shown in a new worksheet or workbook. Here we have selected a cell from the current worksheet itself i.e., cell D21. It is as shown below, and click OK.

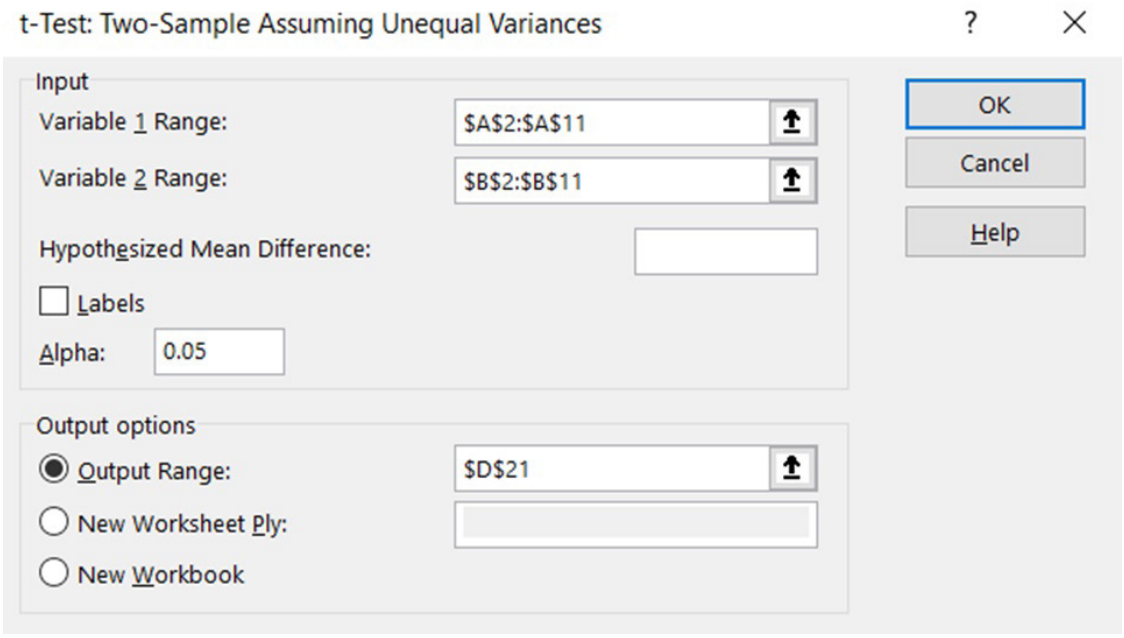


Fig 2.2.12 Inserting Variable Range

9. Excel will generate the result as shown below

t-Test: Two-Sample Assuming Unequal Variances		
	Variable 1	*Variable 2*
Mean	85.1	79.4
Variance	23.43333333	15.15555556
Observations	10	10
Hypothesized Mean Difference	0	
df	17	
t Stat	2.901641264	
P(T<=t) one-tail	0.004963989	
t Critical one-tail	1.739606726	
P(T<=t) two-tail	0.009927978	
t Critical two-tail	2.109815578	

Fig 2.2.13 Results of Two-Independent Sample T Test

Interpretation of Results:

- The independent samples t-test compares means of unrelated groups

Since the p-value (0.0099) is less than the chosen significance level (0.05), we can reject the null hypothesis. Therefore, there is sufficient evidence to conclude that there is a significant difference in the average test scores between Group A and Group B.

2.2.5.3 Paired Sample t-Test

The step-by-step instructions for running a paired sample t-test in Excel using the Data Analysis ToolPak with a simple example is as follows:

Example: Let us compare the effectiveness of a new training program on a group of individuals by measuring their performance before and after the training. We have the scores of 10 individuals before and after the training program. The dataset is as follows:

Table 2.2.1 Dataset of Pre& Post Training

Individual	Before Training	After Training
1	78	85
2	92	88
3	79	90
4	86	80
5	84	89
6	72	81
7	76	85
8	82	78
9	77	80
10	84	79

1. Create an Excel worksheet and enter the data as shown above. Label the columns “Individual,” Before Training,” and “After Training" as shown below

	A	B	C
1	**Individual**	**Before Training**	**After Training**
2	1	78	85
3	2	92	88
4	3	79	90
5	4	86	80
6	5	84	89
7	6	72	81
8	7	76	85
9	8	82	78
10	9	77	80
11	10	84	79

Fig 2.2.14 Dataset of Pre& Post Training

2. Set Null and Alternative Hypothesis

   - Null Hypothesis (H_0): The mean difference in performance before and after the training program is zero (i.e., there is no significant change in performance).
   - Alternative Hypothesis (H_a): The mean difference in performance before and after the training program is not zero (i.e., there is a significant change in performance).

3. Go to the "Data" tab in Excel and click on "Data Analysis" in the Analysis group.

4. Select "t-Test: Paired Two-Sample " from the list and click "OK as shown below

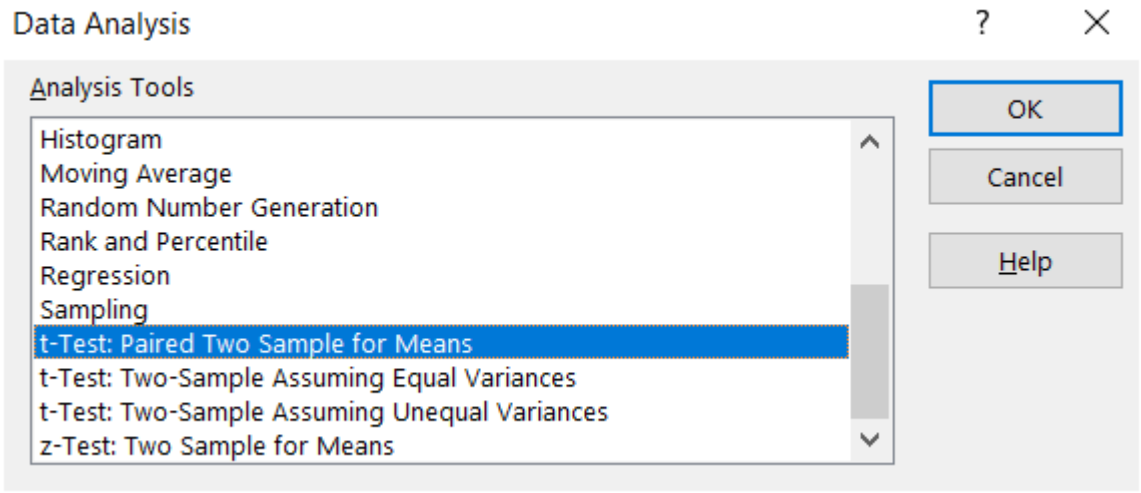


Fig 2.2.15 Selecting Paired Two Sample T-Test

5. In the "t-Test: Two-Sample Assuming Paired" dialog box, for the "Variable 1 Range," select cells B2:B11 (Before Training data) and for the "Variable 2 Range," select cells C2:C11 (After Training data). Leave the "Hypothesised Mean Difference" blank (we want to test if the means are significantly different). Set "Alpha" to 0.05 (5% significance level), and click OK.

6. Click on the output range to display the output. The output can be shown in a new worksheet or workbook. Here we have selected a cell from the current worksheet itself i.e., cell B15. It is as shown below, and click OK. It is shown in the figure below.

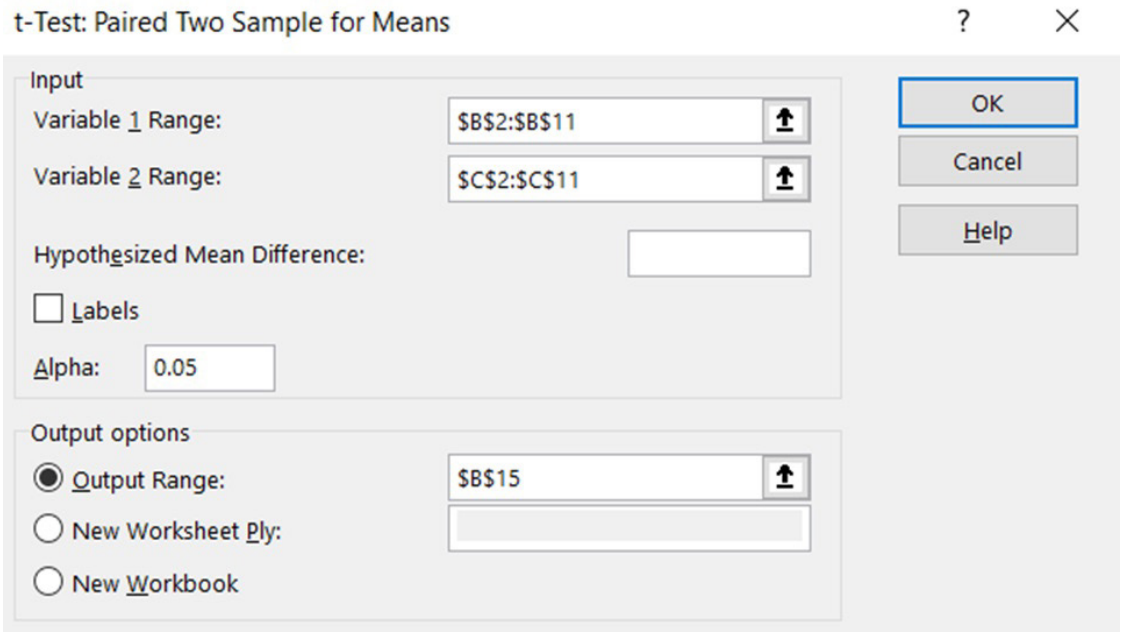


Fig 2.2.16 Selecting Data For Paired Two Sample T-Test

7. Excel will generate the result as shown below

t-Test: Paired Two Sample for Means		
	Variable 1	*Variable 2*
Mean	81	83.5
Variance	33.33333333	19.83333333
Observations	10	10
Pearson Correlation	0.177175684	
Hypothesized Mean Difference	0	
df	9	
t Stat	-1.191076532	
P(T<=t) one-tail	0.132043664	
t Critical one-tail	1.833112933	
P(T<=t) two-tail	0.264087328	
t Critical two-tail	2.262157163	

Fig 2.2.17 Result Of Paired Two Sample T-Test

Interpretation of the result:

- The paired samples t-test compares means of related samples

The p-value of 0.264087328 suggests that there is a 26.41% chance of observing the observed difference in performance (or a more extreme difference) between before and after the training program, assuming that there is no significant difference in performance. Since the p-value is greater than the chosen significance level (alpha = 0.05), we do not have enough statistical evidence to conclude that the training program led to a significant change in performance. In other words, we fail to reject the null hypothesis, indicating that we do not have significant evidence to support that the training program had a substantial impact on the participants' performance. As a result, based on this analysis, we cannot claim with a 5% level of significance that the training program resulted in a significant improvement in performance.

In Excel, you can conduct a t-test using built-in formulas. Specifically, to perform a two-sample independent t-test, you can use the T.TEST function. The T.TEST function calculates the probability (p-value) associated with the t-test. Follow steps 1-3 and then

Use the T.TEST function to perform the t-test as shown below.

Syntax: T.TEST(array1, array2, tails, type)

- array1: The data range for Group A.
- array2: The data range for Group B.
- tails: Specifies the number of distribution tails for which to calculate the probability. For a two-tailed test (most common), use 2. For a one-tailed test (specifying whether the means are greater or less than each other), use 1.
- type: Specifies the type of t-test to be performed: 1 for paired, 2 for two-sample equal variance, and 3 for two-sample unequal variance (which we usually assume unless we have reason to believe otherwise).

- Can conduct a two-sample independent t-test using the T.TEST function

Both the Data Analysis ToolPak and the T.TEST function are valuable tools for conducting t-tests in Excel. The Data Analysis ToolPak provides a more comprehensive and organised output, making it easier for users who prefer a step-by-step approach or require more detailed information. On the other hand, the T.TEST function is a convenient choice when you need to perform quick t-tests directly within your Excel sheet

without additional setup or output management. The choice between these two options depends on your preferences, the level of detail needed, and the complexity of the analysis you wish to perform.

2.2.6 Performing Chi-Square Tests for Categorical Data

Performing chi-square tests for categorical data is a statistical method used to determine if there is a significant association or relationship between two or more categorical variables. The test compares the observed frequencies of different categories in a contingency table to the frequencies we would expect if the variables were independent. It is commonly used in various fields, such as social sciences, biology, marketing, and more.

Let us go through the steps of performing a chi-square test using an example:

Example: Let us conduct a survey to investigate whether there is a significant relationship between gender (male or female) and favourite colour (red, blue, green, or yellow) among a group of 100 people. The data are as follows:

Table 2.2.2 Dataset For Chi Square Test

	Red	Blue	Green	Yellow	Total
Male	15	20	10	5	50
Female	20	15	10	5	50
Total	35	35	20	10	100

1. Enter the data in cells as shown below.

	A	B	C	D	E	F
1		**Red**	**Blue**	**Green**	**Yellow**	**Total**
2	**Male**	15	20	10	5	50
3	**Female**	20	15	10	5	50
4	**Total**	35	35	20	10	100

Fig 2.2.18 Dataset For Chi Square Test

2. Set up the Hypothesis

- Null Hypothesis (H_0): There is no association between gender and favourite colour (i.e., the variables are independent).
- Alternative Hypothesis (H_a): There is an association between gender and favourite colour (i.e., the variables are dependent).

3. Create a contingency table with observed frequencies alone. It will be as shown below

6		Red	Blue	Green	Yellow
7	Male	15	20	10	5
8	Female	20	15	10	5

Fig 2.2.19 Contingency Table For Chi Square Test

4. Calculate the Expected Frequencies and Create a contingency table with expected frequencies

Next, we calculate the expected frequencies for each cell in the contingency table under the assumption that the variables are independent. The formula for calculating expected frequencies is:

Expected Frequency = (row total * column total) / overall total

Using this formula, we can calculate the expected frequencies for each cell in the table. Create an expected frequency table below the original or observed frequency table. It will be as shown below.

10					
11		Red	Blue	Green	Yellow
12	Male	17.5	17.5	10	5
13	Female	17.5	17.5	10	5
14					

Fig 2.2.20 Expected Frequency Table for Chi Square Test

5. Calculate the Chi-Square Test Statistic

The chi-square test statistic is calculated as the sum of the squared differences between the observed and expected frequencies, divided by the expected frequencies. The formula for calculating the chi-square test statistic is:

χ^2 = Σ ((Observed Frequency - Expected Frequency)² / Expected Frequency)

To calculate the chi-square test statistic in Excel, we can use the CHISQ.TEST function.

The syntax for using **CHISQ.TEST(observed range, expected range)**

- observed range: The range of cells containing the observed frequencies.
- expected range: The range of cells containing the expected frequencies of the contingency table (if you have calculated them separately).

6. We entered the below formula

```
=CHISQ.TEST(B2:E3,B12:E13)
```

7. Excel generated the chi square result and it is as shown below.

Fig 2.2.21 Result Of Chi Square Test

Interpretation of the result:

The p-value of 0.69885 suggests that there is a 69.88% chance of observing the observed association between gender and favourite colour (or even a more extreme association) if the null hypothesis is true (i.e., if there is no significant relationship between the two variables). Since the p-value is greater than the chosen significance level of 0.05, we fail to reject the null hypothesis. Therefore, we do not have enough evidence to claim a significant association between gender and favourite colour among the surveyed group at the 5% significance level.

In simple terms, the chi-square test did not find enough statistical evidence to support the presence of a significant relationship between gender and favourite colour in the given dataset. It means that we cannot conclude that gender and

- Involves comparing observed and expected frequencies in a contingency table

favourite colour are related based on the data we have analysed.

2.2.7 Regression Analysis

Regression analysis is a statistical method used to examine the relationship between a dependent variable (also known as the outcome or response variable) and one or more independent variables (also known as predictor or explanatory variables). It aims to understand how the changes in the independent variables are associated with changes in the dependent variable.

The primary goal of regression analysis is to build a mathematical model that best describes the relationship between the variables. This model can then be used to make predictions, understand the effect of different factors on the outcome, and identify patterns and trends in the data.

There are several types of regression analysis, but the two most common ones are:

- **Simple Linear Regression:** In this type of regression, there is only one independent variable used to predict the dependent variable. The relationship between the two variables is assumed to be linear, which means it can be represented by a straight line.
- **Multiple Linear Regression:** Here, there are two or more independent variables used to predict the dependent variable. The relationship is still assumed to be linear, but the model becomes a multi-dimensional plane instead of a straight line.

- Examine the relationship between a dependent variable and one or more independent variables

Regression analysis is widely used in various fields, including economics, finance, social sciences, healthcare, and engineering, to gain insights, make predictions, and understand the underlying relationships between variables. It is a powerful tool for data analysis and decision-making in both research and practical applications.

2.2.7.1 Building Regression Models in Excel

In Excel, you can build regression models using the built-in Data Analysis Toolpak, which provides various statistical analysis tools, including regression.

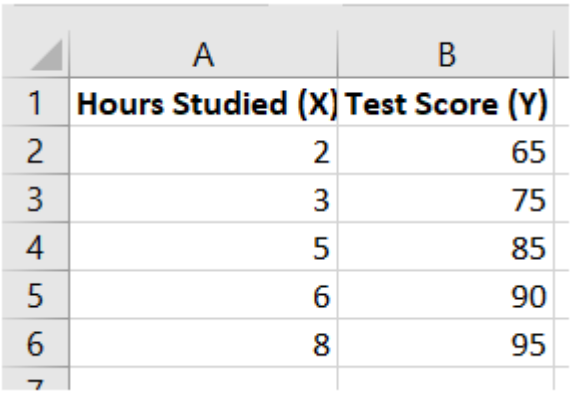
Let us go through a step-by-step process of creating a simple linear regression model in Excel using a sample dataset. For this example, we will use a hypothetical dataset of students'

test scores and the number of hours they studied.

Table 2.2.3 Dataset for Regression Analysis

Hours Studied (X)	Test Score (Y)
2	65
3	75
5	85
6	90
8	95

1. Create an Excel worksheet and enter the data as shown above.



	A	B
1	Hours Studied (X)	Test Score (Y)
2	2	65
3	3	75
4	5	85
5	6	90
6	8	95

Fig 2.2.22 Dataset for Regression Analysis

2. Go to the "Data" tab in Excel and Click on "Data Analysis" in the Analysis group

3. Select "Regression" from the list as shown below and click "OK."

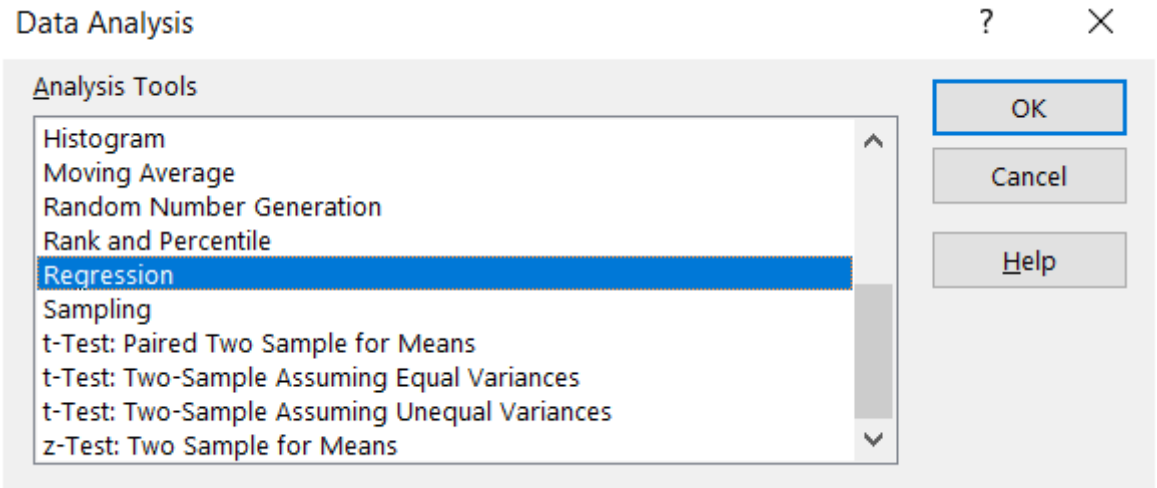


Fig 2.2.23 Selecting Regression Analysis

4. In the Regression dialog box, you will need to fill in the following information:

Input Y Range: Click the button at the right end of the text box and select the range containing your dependent variable (Test Score). Here we have selected B1:B6

Input X Range: Click the button at the right end of the text box and select the range containing your independent variable (Hours Studied). Here we have selected A1:A6

Check "Labels" if your data includes headers, and Excel will use them as variable names in the output.

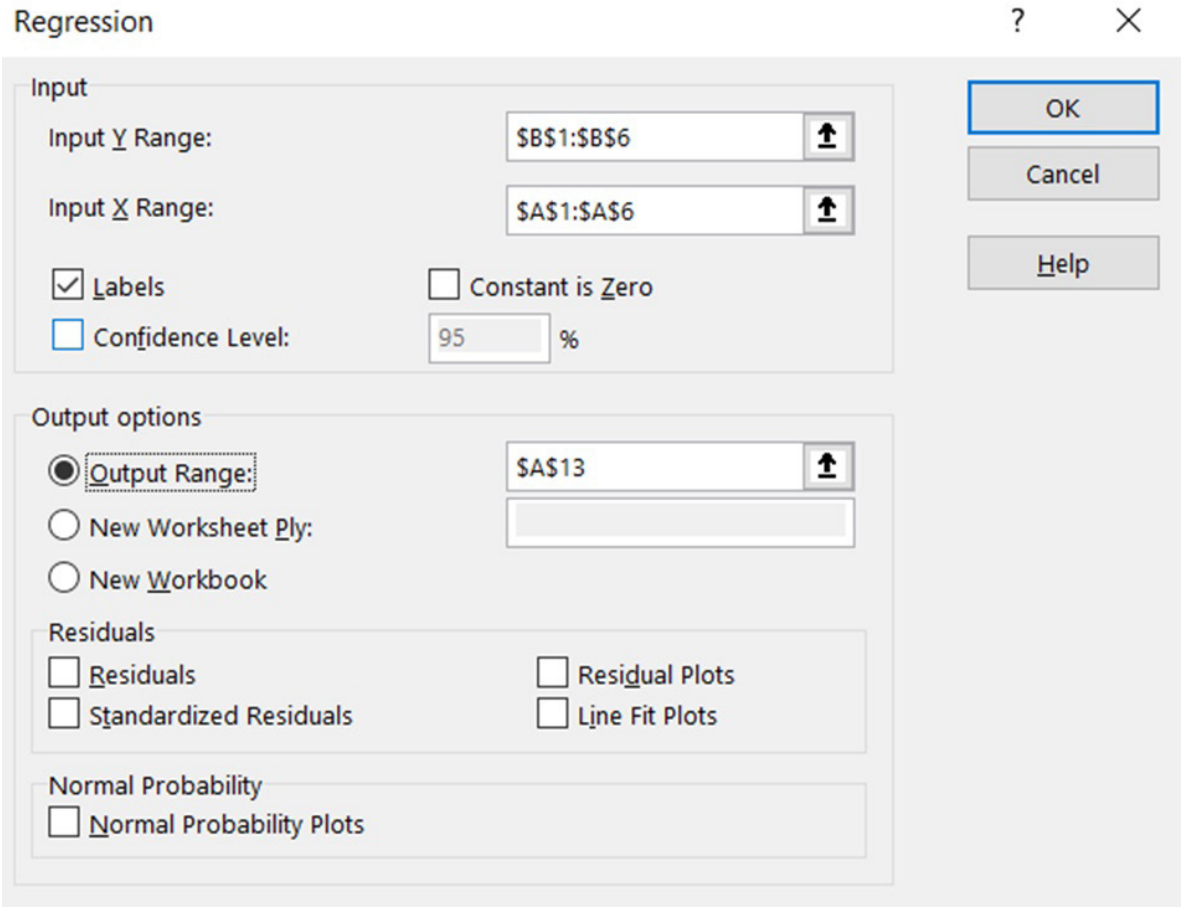


Fig 2.2.24 Selecting Data Range for Regression Analysis

5. Choose where you want the output to be displayed. You can select "Output Range" and specify a cell where the regression results will be placed, or choose "New Worksheet Ply" to create a new sheet for the output. Here we have selected A13 as output range

6. Click"OK" to perform the regression analysis.

7. Excel will generate the results as show below

13	SUMMARY OUTPUT								
14									
15	*Regression Statistics*								
16	Multiple R	0.973949866							
17	R Square	0.948578342							
18	Adjusted R Square	0.93143779							
19	Standard Error	3.153017676							
20	Observations	5							
21									
22	ANOVA								
23		*df*	*SS*	*MS*	*F*	*Significance F*			
24	Regression	1	550.1754386	550.1754386	55.34117647	0.005027412			
25	Residual	3	29.8245614	9.941520468					
26	Total	4	580						
27									
28		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
29	Intercept	58.42105263	3.469073673	16.84053386	0.00045595	47.38091194	69.46119332	47.38091194	69.46119332
30	Hours Studied (X)	4.912280702	0.660326888	7.439165039	0.005027412	2.810825836	7.013735568	2.810825836	7.013735568
31									

Fig 2.2.25 Result of Regression Analysis

- Enables to analyse the relationship between variables and make predictions

The R-squared value of 0.9486 indicates that about 94.86% of the variation in the Test Score can be explained by the Hours Studied.

The p-value for the Hours Studied coefficient is 0.005, which is less than the significance level of 0.05. It means that the number of Hours Studied is statistically significant in predicting Test Score.

2.2.7.2 Analysing Regression Results and Interpreting Coefficients

The regression statistics you provided are the results of a multiple linear regression analysis. Let us interpret each statistic accordingly:

The Regression statistics is as given below

15	*Regression Statistics*	
16	Multiple R	0.973949866
17	R Square	0.948578342
18	Adjusted R Square	0.93143779
19	Standard Error	3.153017676
20	Observations	5

Fig 2.2.26 Regression Statistics

- **Multiple R**: The multiple correlation coefficient (Multiple R) measures the strength and direction of the linear relationship between the dependent variable and the combination of independent variables. In this case, the value of 0.9739 indicates a strong positive correlation between the dependent variable and the independent variables.
- **R Square**: The coefficient of determination (R-squared) represents the proportion of the variance in the dependent variable that is explained by the independent variables. It ranges from 0 to 1, with 1 indicating a perfect fit. In this case, an R-squared value of 0.9486 means that approximately 94.86% of the variability in the dependent variable can be explained by the independent variables included in the model. This suggests that the model is a good fit for the data.

- Regression model shows a strong positive relationship

- **Adjusted R Square**: The adjusted R-squared adjusts the R-squared value for the number of independent variables in the model and the sample size. It penalises the R-squared for including unnecessary variables that might not significantly improve the model. An adjusted R-squared value of 0.9314 indicates that around 93.14% of the variability in the dependent variable is explained by the independent variables, considering the number of predictors and sample size.
- **Standard Error**: The standard error of the regression (standard error) measures the average deviation of the observed values from the predicted values by the regression model. It represents the accuracy of the model's predictions. In this case, a standard error of 3.1530 suggests that, on average, the actual dependent variable values deviate from the predicted values by approximately 3.1530 units.
- **Observations**: The number of observations refers to the sample size used to build the regression model. In this case, there are 5 observations in the dataset.

		Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
29	Intercept	58.42105263	3.469073673	16.84053386	0.00045595	47.38091194	69.46119332	47.38091194	69.46119332
30	Hours Studied (X)	4.912280702	0.660326888	7.439165039	0.005027412	2.810825836	7.013735568	2.810825836	7.013735568

Fig 2.2.27 Coefficients Of Regression Analysis

Let us interpret the coefficients

- **Intercept**: The intercept is the value of the dependent variable (Test Score) when all independent variables (Hours Studied and any other included predictors) are zero. In this case, the intercept is 58.4211, which means that when the Hours Studied is zero (which doesn't make sense in this context), the predicted Test Score would be approximately 58.42.
- **Hours Studied (X) Coefficient**: The coefficient of the Hours Studied represents the change in the Test Score for each one-unit increase in Hours Studied, while holding all other predictors constant. In this case, the coefficient is 4.9123, which means that, on average, for every additional hour a student studies, their Test Score is predicted to increase by approximately 4.9123 units.
- The **p-value** associated with the Hours Studied coefficient (0.0050) is less than the significance level of 0.05. This indicates that the number of Hours Studied is statistically significant in predicting Test Score, suggesting that there is a significant relationship between Hours Studied and Test Score in this dataset.

The regression statistics indicate that the multiple linear regression model has a strong positive relationship with a high degree of explanation (R-squared) between the dependent variable and the independent variables. The adjusted R-squared suggests that the model is not overly complex and still provides a good fit to the data, considering the number of predictors and the sample size. The standard error indicates the accuracy of the model's predictions, with lower values indicating better prediction performance.

Overall, based on the provided regression statistics, the model appears to be a good fit for the data, and the Hours Studied variable has a statistically significant impact on predicting Test Score. However, always keep in mind that proper interpretation and validation of the results depend on the context of the data, the research question, and the underlying assumptions of the regression model.

Summarised Overview

Hypothesis testing is a statistical method used to draw conclusions about a population based on sample data. It involves comparing observed data with a hypothesis to determine if there is enough evidence to support or reject that hypothesis. The p-value quantifies the strength of evidence against the null hypothesis, and the significance level (alpha) sets the threshold for rejecting the null hypothesis. There are one-tailed tests, used when a specific direction is predicted, and two-tailed tests for non-directional predictions. Type I error occurs when the null hypothesis is falsely rejected, and Type II error occurs when it is falsely accepted. The appropriate test choice and interpretation depend on these factors to draw accurate conclusions from statistical analyses.

T-tests are statistical tests used to compare means between two groups. There are three main types of t-tests: One-Sample T-Test compares a sample mean to a known population mean, Independent Samples T-Test compares means of two unrelated groups, and Paired Samples T-Test compares means of related samples. In Excel, you can perform these tests using the Data Analysis ToolPak or the T.TEST function. Performing a chi-square test for categorical data involves comparing observed frequencies in a contingency table to expected frequencies to determine if there is a significant association between two or more categorical variables.

Regression analysis is a statistical method used to examine the relationship between a dependent variable and one or more independent variables. It helps build a mathematical model to understand how changes in the independent variables are associated with changes in the dependent variable. In Excel, you can perform regression analysis using the Data Analysis Toolpak, which provides detailed statistics like R-squared, p-values, and coefficients to interpret the relationship and significance of variables. R-squared indicates the proportion of variability explained by the model, while p-values assess the statistical significance of predictors. Coefficients show how much the dependent variable changes for each unit increase in the independent variable. Proper interpretation and context are essential to draw meaningful conclusions from regression results.

Self-Assessment

1. Explain the concept of hypothesis testing. What are the steps involved in hypothesis testing?
2. What are Type I and Type II errors in hypothesis testing? Provide examples of each.
3. What is a paired sample t-test? Give an example where a paired sample t-test would be appropriate.
4. Explain the chi-square test. When is it used, and what type of data does it analyse?
5. What is regression analysis? Discuss the difference between simple linear regression and multiple linear regression.

Assigments

1. Given below is a dataset of weights of 10 individuals before and after a weight loss programme. Conduct the 2-sample paired t-test with a significance level of 0.05 to determine if there is a significant difference in weight before and after the weight loss program.

Individual	Weight Before (kg)	Weight After (kg)
1	81.65	79.37
2	90.72	87.09
3	72.57	69.85
4	83.91	81.19
5	99.79	95.25
6	79.37	76.2
7	88.45	84.82
8	68.04	66.68
9	92.99	90.72
10	77.11	74.84

2. You have survey data on the preferred mode of transportation (car, bus, bike) for two age groups (18-25 and 26-35). Perform a chi-square test to determine if there is a significant association between age group and mode of transportation. The dataset is given below

3. A researcher wants to investigate the weekly consumption expenditure of households in a village. They randomly selected a sample of 30 households and recorded their weekly consumption expenditures in dollars. The researcher knows that the national average weekly consumption expenditure for households is Rs.1000. Based on the dataset provided, perform a one-sample t-test to determine if there is a significant difference between the average weekly consumption expenditure of households in the village and the national average of Rs.1000. Use a significance level of 0.05.

Household ID	Consumption Expenditure (in INR)
1	1200
2	800
3	950
4	1100
5	750
6	1300
7	900
8	1050
9	980
10	850
11	950
12	800
13	1150
14	900
15	700
16	850
17	950
18	1100

19	850
20	950
21	1000
22	1200
23	850
24	900
25	800
26	1100
27	750
28	950
29	1050
30	900
31	950
32	1200
33	750
34	800
35	950
36	1100
37	850
38	950
39	900
40	800
41	1150
42	900
43	700
44	850
45	950
46	1300
47	900
48	1050
49	980
50	850

4. A researcher wanted to investigate whether there is a significant difference in the average height between Group A and Group B. They collected height data in centimetres for 20 individuals from each group. The data is summarised in the table given below. Using this data, perform an independent two-sample t-test to determine whether there is a statistically significant difference in the average height between Group A and Group B.

Group A (cm)	Group B (cm)
165	161
172	164
160	159
180	165
170	152
178	163
168	157
175	162
162	158
167	160
155	150
162	153
149	148
157	151
152	149
160	155
147	146
158	151
149	150
155	149

5 . Given below is a dataset of a small retail store that records its monthly advertising expenses and corresponding sales over a period of 12 months. Create a regression model based on advertising expenses and sales. Evaluate the model's performance on the testing set using appropriate metrics (e.g., Mean Squared Error, R-squared). Interpret the results.

Month	Advertising Expenses (in INR)	Sales (in INR)
Jan	5000	25000
Feb	6000	27000
Mar	8000	32000
Apr	7000	29000
May	9000	35000
Jun	10000	38000
Jul	11000	39000
Aug	12000	40000
Sep	13000	42000
Oct	14000	43000
Nov	15000	46000
Dec	16000	47000

Reference

1. Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., & Cochran, J. J. (2016). *Statistics for business & economics*. Cengage Learning.

2. Carlberg, C. (2016). *Regression analysis Microsoft excel*. Que Publishing.

3. Nigam, M. (2019). *Advanced Analytics with Excel 2019 (FIRST EDITION).* BPB Publications.

4. Alexander, M., Kusleika, R., & Walkenbach, J. (2018). *Excel 2019 bible*. John Wiley & Sons.

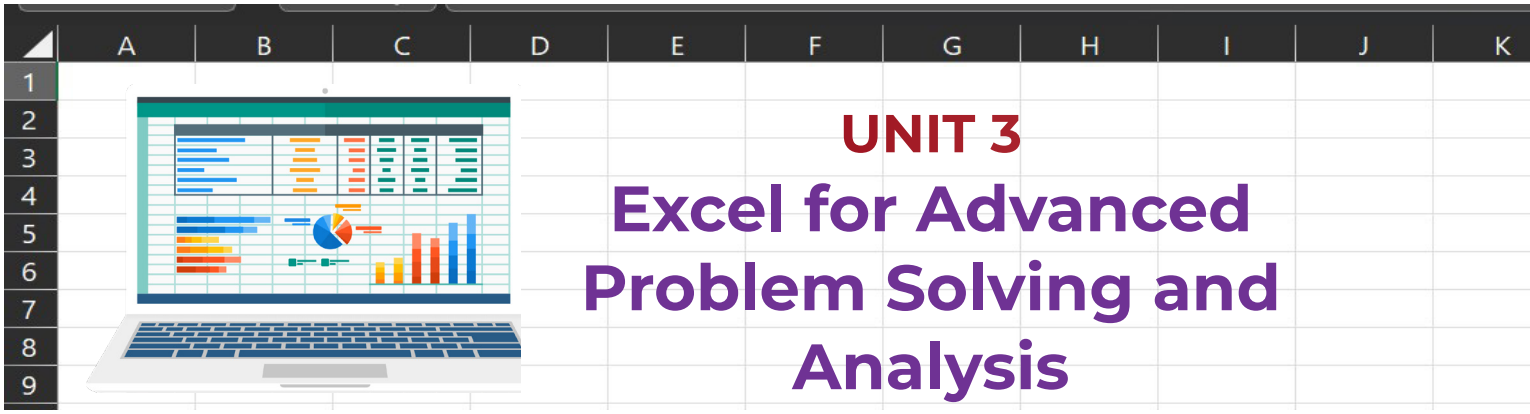
5. Rudraksh, N., & McDonald, A. (2020). *Data Analysis Using Microsoft Excel: Updated for Office 365 (2nd ed.)*. Cambridge University Press

Suggested Reading

1. Winston, W. L. (2019). *Microsoft Excel 2019 Data Analysis and Business Modelling (6th ed.)*. Microsoft Press.
2. Lind, D. A., Marchal, W. G., & Wathen, S. A. (2018). *Statistical Techniques in Business and Economics with Microsoft Excel (17th ed.*). McGraw-Hill Education.
3. Field, A. (2018). *Discovering Statistics Using Microsoft Excel*. SAGE Publications.
4. Albright, S. C., Winston, W. L., & Zappe, C. (2019). *Data Analysis and Decision Making with Microsoft Excel (6th ed.)*. Cengage Learning.
5. Peck, R., Olsen, C., & Devore, J. (2018). *Introduction to Statistics and Data Analysis (6th ed.)*. Cengage Learning.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.



UNIT 3
Excel for Advanced Problem Solving and Analysis

Learning Outcomes

After the completion of this unit, the learner will be able to

- understand the concept of optimisation
- apply Solver to find optimal solutions
- perform analysis using advanced functions
- incorporate logical functions
- use array formulas to perform multi-cell calculations

Background

Advanced problem-solving and analysis is a comprehensive approach to address complex issues by leveraging critical thinking, creativity, and a systematic process for effective solutions. It goes beyond traditional methods, emphasising a deeper understanding of the problem and considering diverse perspectives. Commonly utilised in business, engineering, research, and other fields, this methodology is crucial for tackling intricate problems and achieving progress.

The process commences with a clear and comprehensive definition of the problem, identifying underlying issues, understanding its scope, and setting specific solution goals. Gathering relevant data, conducting thorough research, and engaging stakeholders are critical in gaining insights into root causes and implications. Subsequently, both qualitative and quantitative analysis techniques are employed to scrutinise the collected information. After a thorough analysis, the next step involves generating potential solutions through brainstorming and creative thinking. The chosen solutions must align with the defined problem and be supported by evidence and feasibility. Finally, the selected solution is evaluated, considering advantages, disadvantages, and potential risks. An action plan is devised for implementation, and monitoring and

feedback mechanisms are established to track its effectiveness and make necessary adjustments.

Advanced problem-solving and analysis empower individuals and teams to make informed decisions and drive meaningful outcomes in diverse domains. This multi-dimensional approach fosters critical thinking, analytical skills, creativity, and collaboration, enabling successful resolution of complex challenges in today's rapidly changing world.

Keywords

Optimisation, Logical Functions, Array, Dynamic Arrays, Solver

Discussion

2.3.1 Introduction to Optimisation Problems

Optimisation problems involve finding the best possible solution to a given problem from a set of feasible solutions. These problems arise in various real-world scenarios where you need to maximise or minimise an objective function subject to certain constraints. Excel provides built-in tools and functions that allow you to solve optimisation problems efficiently.

Types of Optimisation Problems in Excel

The types of optimisation problems in Excel are as follows:

a. **Linear Programming (LP):** Linear programming deals with linear objective functions and linear constraints. The goal is to maximise or minimise a linear objective function while adhering to linear constraints.

b. **Nonlinear Programming (NLP)**: Nonlinear programming involves optimisation problems where the objective function or constraints are nonlinear. Solving NLP problems may require more advanced techniques compared to LP problems.

c. **Integer Programming (IP):** Integer programming is an extension of linear programming, where some or all decision variables are restricted to integer values. This type

of problem is particularly useful when the variables represent discrete quantities.

d. **Mixed-Integer Linear Programming (MILP):** MILP combines linear objective functions and constraints with some integer decision variables.

- Finding the best possible solution from a set of feasible options

An optimisation problem in Excel typically consists of the following components:

a. **Objective Function:** The objective function is a mathematical expression that defines the quantity to be optimised. It can be a formula involving decision variables, and the goal is to either maximise or minimise this function.

b. **Decision Variables:** Decision variables are the unknowns in the problem that you can control and adjust to achieve the optimal solution. These variables represent the quantities you want to find to achieve the best outcome.

c. **Constraints:** Constraints are conditions or limitations that the decision variables must satisfy. They can be inequalities (e.g., less than or equal to, greater than or equal to) or equalities (e.g., exact values). Constraints define the feasible region within which the optimal solution must lie.

- Comprises an objective function to optimise, decision variables to control, and constraints

d. **Feasible Region:** The feasible region is the set of all possible solutions that satisfy all the constraints. It is the intersection of the constraints that define the boundaries of the optimisation problem.

e. **Objective**: The objective is to find the optimal values for the decision variables that either maximise or minimise the objective function while still satisfying all the constraints within the feasible region.

2.3.2 Solver Tool

The Solver is an add-in tool that allows users to find the optimal solution for complex problems involving constraints and multiple variables. It is a powerful optimisation tool that can be used to maximise or minimise a specific objective by changing certain input values while adhering to certain limitations or restrictions.

The Solver can be useful in various scenarios where you need to:

a. **Optimise Financial Decisions:** It can be used for financial modelling and investment planning, such as portfo-

lio optimisation or finding the best mix of investments to achieve a specific return.

b. **Resource Allocation:** When dealing with limited resources, like manpower or production capacity, the Solver can help you determine the most efficient way to allocate these resources to maximise output.

c. **Scheduling and Planning:** It can be used to optimise schedules, project timelines, or production plans to meet deadlines and minimise costs.

d. **Regression Analysis:** Solver can be applied in regression analysis to find the best-fit curve for a set of data points.

e. **Supply Chain Optimisation**: For businesses with complex supply chains, the Solver can help optimise inventory levels, transportation routes, and distribution strategies.

f. **Marketing and Sales Strategies:** It can assist in determining the best pricing strategy, product mix, or advertising allocation to maximise revenue or profit.

- Powerful add-in tool in Excel that enables users to find optimal solutions

Solver is best suited for relatively small to medium-sized problems. Very complex optimisation tasks may require more specialised tools or programming solutions.

2.3.2.1 Using Solver Tool for Goal-Seeking and Optimisation

Solver is an add-in tool in Microsoft Excel that allows you to perform various optimisation tasks, such as goal-seeking, linear programming, nonlinear programming, integer programming, and more. It is a powerful tool that can be used for data analysis, decision-making, and problem-solving in a wide range of fields, including business, finance, engineering, operations research, and science.

The steps to enable Solver Add-In are as follows:

1. Click on the “File” menu in Excel.

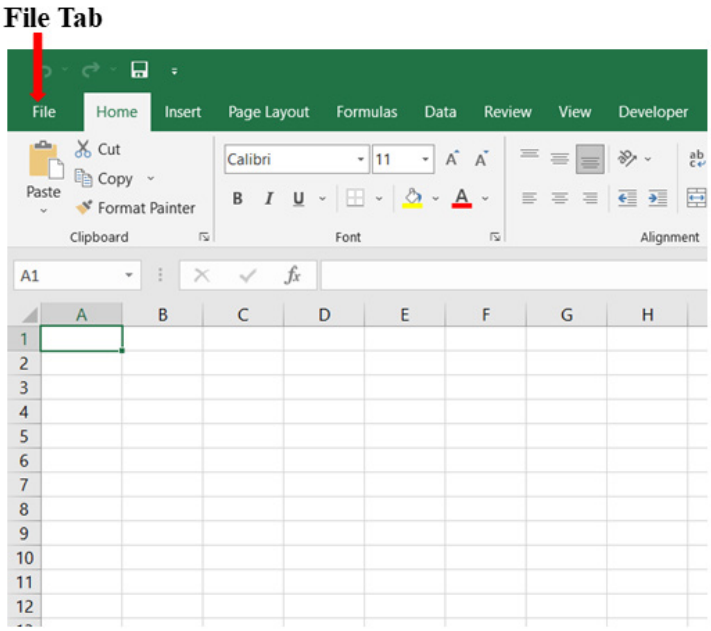


Fig 2.3.1 Selecting File Tab

2. Select “Options,” and in the Excel Options dialog box, click on “Add-Ins.”

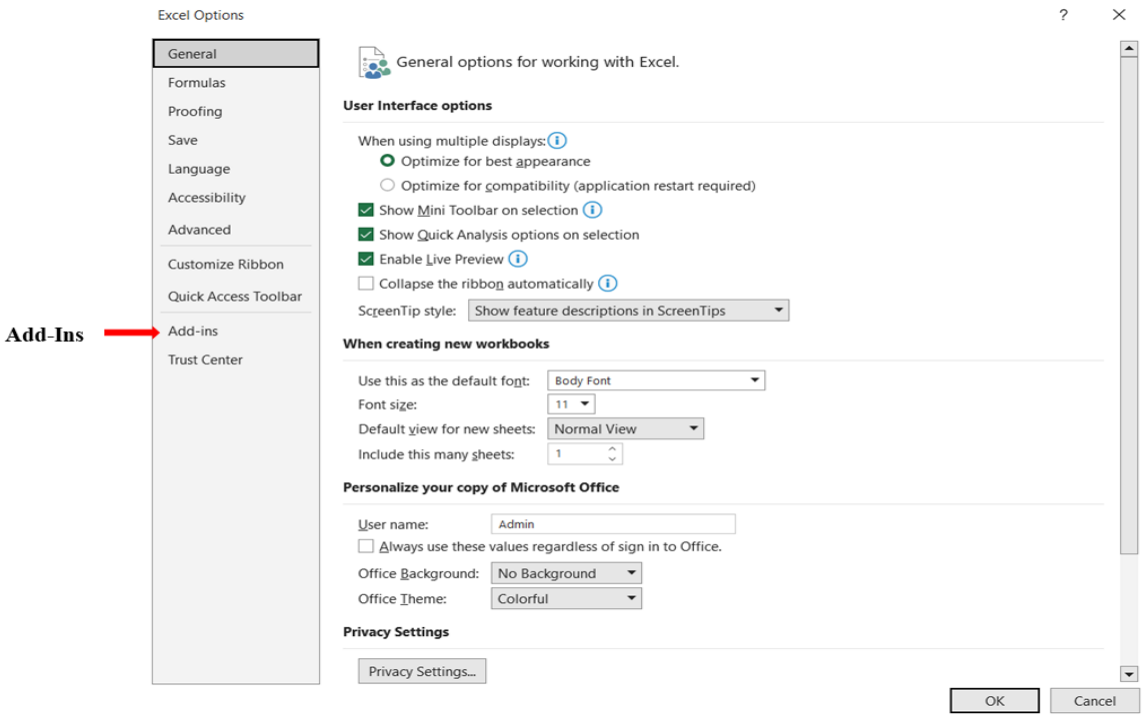


Fig 2.3.2 Selecting Add-ins

3. In the Add-Ins window, select “Solver Add-In” and click “Go.”

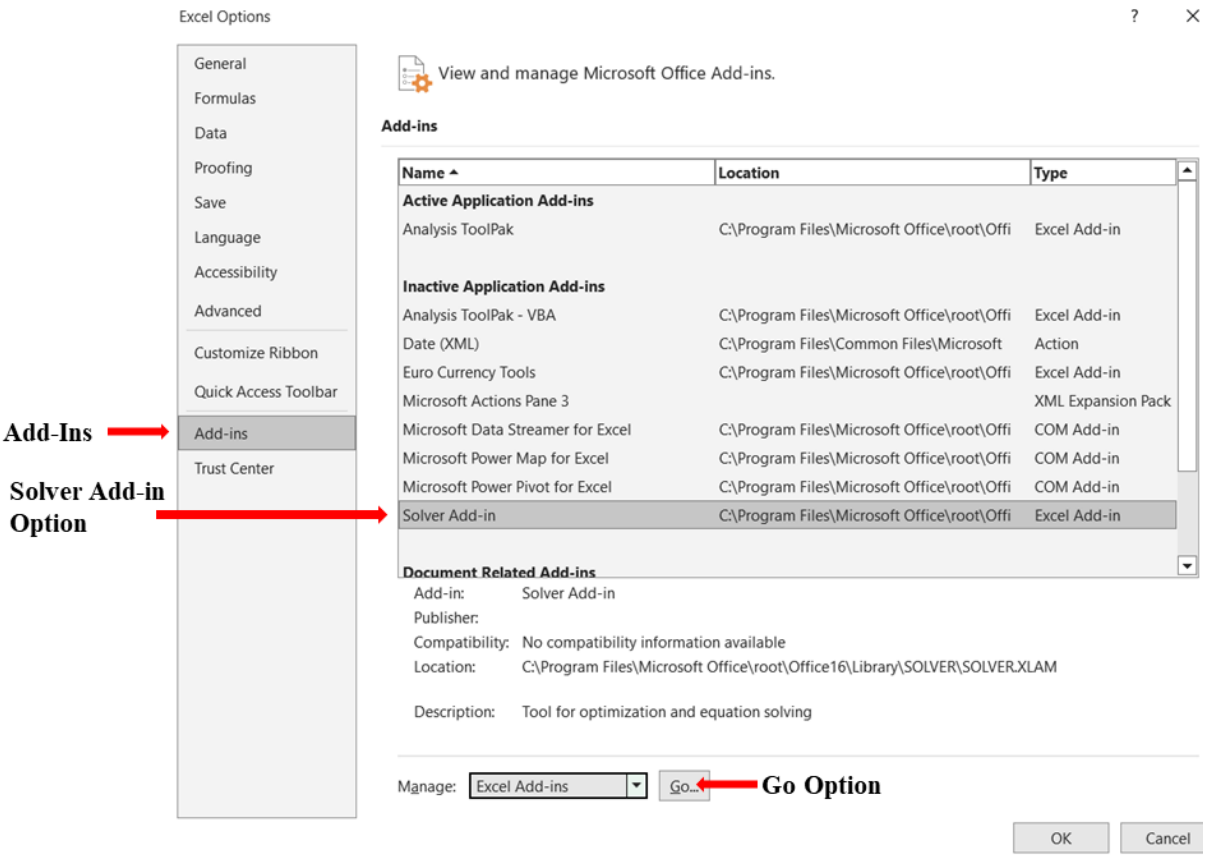


Fig 2.3.3 Selecting Add-Inn Option

4. Check the box next to “Solver Add-In,” then click “OK.”

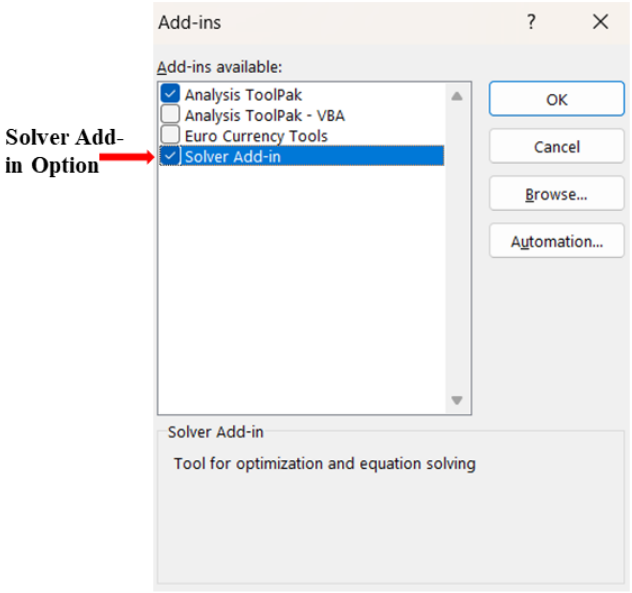


Fig 2.3.4 Selecting Solver Add-In

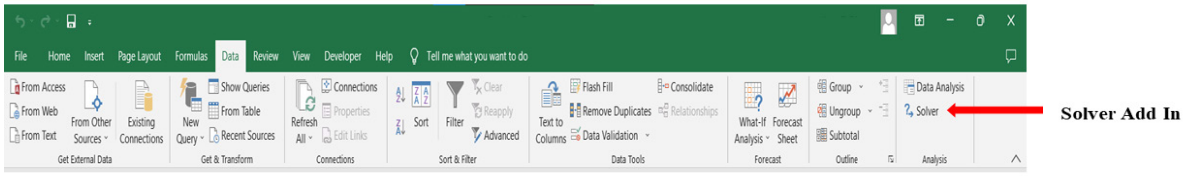


Fig 2.3.5 Solver Icon In Data Tab

- The Solver icon will appear in the Data ribbon

Let us create a simple data set to demonstrate how to use Solver for optimisation in Excel.

The accountant of a small firm made deposits into the bank for some of the days but forgot to record them in his journal. Now, he wants to find out the days for which he made the payments. The total deposited amount is Rs. 17,989. So, let us use excel and solver to solve the problem. The dataset of daily journal entries is given below.

	A	B
1	**Day**	**Sales (INR)**
2	Jul-01	2534
3	Jul-02	1875
4	Jul-03	3150
5	Jul-04	4026
6	Jul-05	5261
7	Jul-06	3702
8	Jul-07	4789
9	Jul-08	2750
10	Jul-09	3248
11	Jul-10	4389
12	Jul-11	5207
13	Jul-12	2136
14	Jul-13	3998
15	Jul-14	5967
16	Jul-15	3425

Fig 2.3.6 Journal Entry

1. In the cell after all the journal entries add another enter titled Sum Total, and in the cell next to this sum total cell we are going to type in the formula =SUMPRODUCT.

2.The formula SUMPRODUCT will generate the total sum of products of all arrays we need. It is as given below.

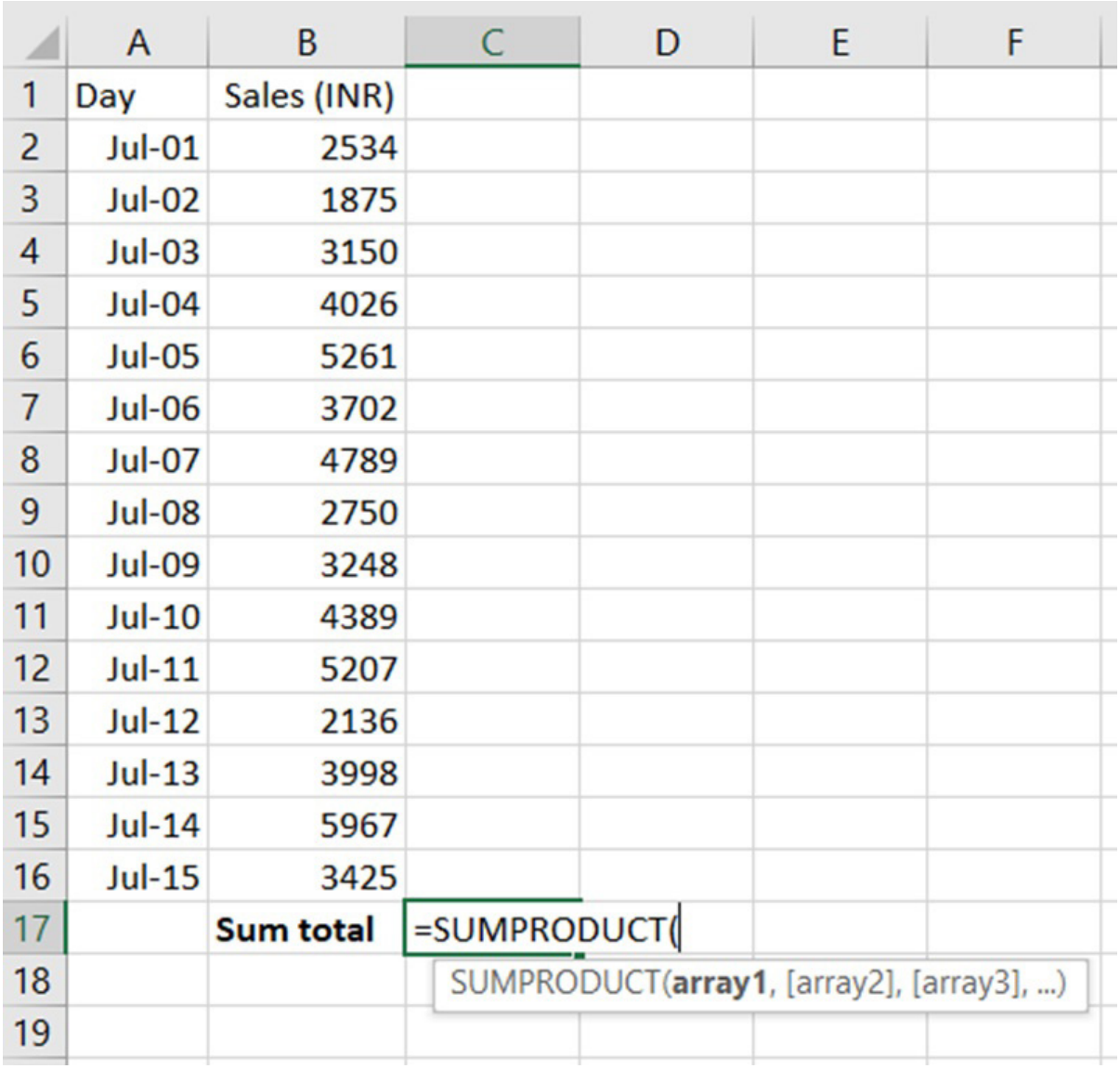


Fig 2.3.7 Using Sumproduct Function

3. Select the Sales entries as array1 and the cell adjacent to the sales as array2, as shown below.

	A	B	C	D	E
1	Day	Sales (INR)			
2	Jul-01	2534			
3	Jul-02	1875			
4	Jul-03	3150			
5	Jul-04	4026			
6	Jul-05	5261			
7	Jul-06	3702			
8	Jul-07	4789			
9	Jul-08	2750			
10	Jul-09	3248			
11	Jul-10	4389			
12	Jul-11	5207			
13	Jul-12	2136			
14	Jul-13	3998			

Fig 2.3.8 Selecting Data

4. This means that the sum total of Products of the arrays will be displayed below.

Now we will get a result of 0 as we have not made any entry in column C. It is as shown below.

	A	B	C
1	Day	Sales (INR)	
2	Jul-01	2534	
3	Jul-02	1875	
4	Jul-03	3150	
5	Jul-04	4026	
6	Jul-05	5261	
7	Jul-06	3702	
8	Jul-07	4789	
9	Jul-08	2750	
10	Jul-09	3248	
11	Jul-10	4389	
12	Jul-11	5207	
13	Jul-12	2136	
14	Jul-13	3998	
15	Jul-14	5967	
16	Jul-15	3425	
17		**Sum total**	0
18			

Fig 2.3.9 Finding Sum Total

Now let us make use of Solver tool in excel to solve for this. Solver will generate the results and will help us with the problem.

5. Click on Solver from Data Tab. It will open up a new menu as shown below.

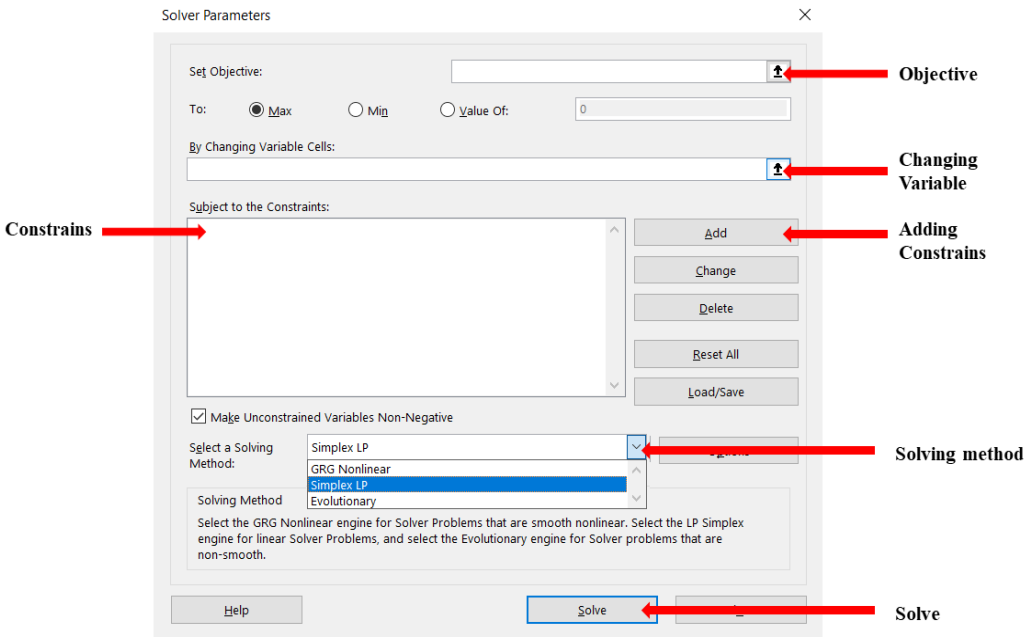


Fig 2.3.10 Options in Solver Menu

1. **Objective:** The "Objective" refers to the cell that you want to optimise or achieve a specific value for. In other words, it represents the goal of your optimisation problem. You select the cell containing the objective function or formula, and Solver will adjust the values of the changing variables to achieve the desired objective.
2. **Changing Variable Cells:** The "Changing Variable Cells" are the cells that will be adjusted by Solver to find the optimal solution for the objective. These cells are the decision variables that you want Solver to change within specified constraints to achieve the desired outcome.
3. **Constraints:** Constraints are restrictions or limitations that you want to impose on the changing variable cells to reflect real-world limitations or business requirements. These constraints define the feasible region within which the changing variables can vary. You can add multiple constraints to limit the potential solutions. Constraints can be of different types, such as equalities (=), inequalities (<=, >=), or integer constraints.

4. **Solving Method:** The "Solving Method" allows you to choose how Solver will search for the optimal solution. Excel provides three solving methods:

   - **GRG Nonlinear**: This method is suitable for smooth, nonlinear problems with continuous variables. It is the default method for most optimisation problems.
   - **Simplex LP**: This method is specifically designed for linear programming problems, where both the objective function and constraints are linear.
   - **Evolutionary**: This method is used for more complex problems that may have non-smooth functions, integer constraints, or discrete decision variables.

The solving method you choose will impact the time taken to find the solution and the accuracy of the result. In most cases, the default GRG Nonlinear method should work well for non-linear optimisation problems.

Example 1: We will be setting the objective, defining the constraining, and adding the changing variables as shown below.

- Our objective cell is the cell C17. That is where we want the result of Rs.17989 and objective is set to value of Rs.17989. We have to type in and give it.
- Our changing variable cells will be C2:C16.
- Constrains is that the values in cells C2:C16 should be 0 or 1. Then only the sum total of product with sales will generate our result.

The constraint will pop a new menu as shown below

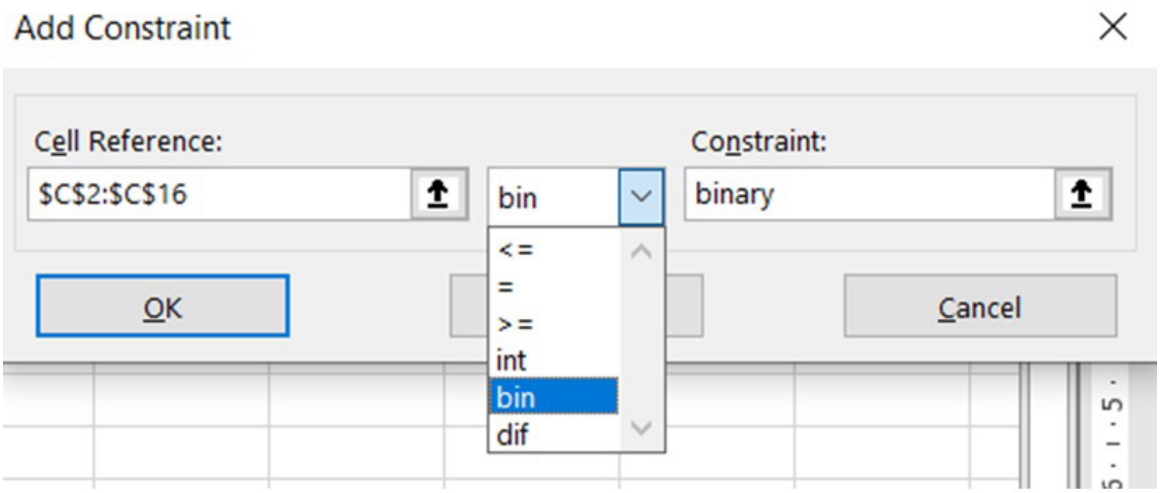


Fig 2.3.11 Inserting Constraints

Select the reference cell that should be restricted. In the next option we can give less than or equal to, equal to, greater than or equal to, integer, binary and all different. Here we have chosen binary since we want either 0 or 1. Then click OK. We can give multiple constraints. Here we have only one constraint so we have given that.

After entering and giving the conditions the menu will be as shown below.

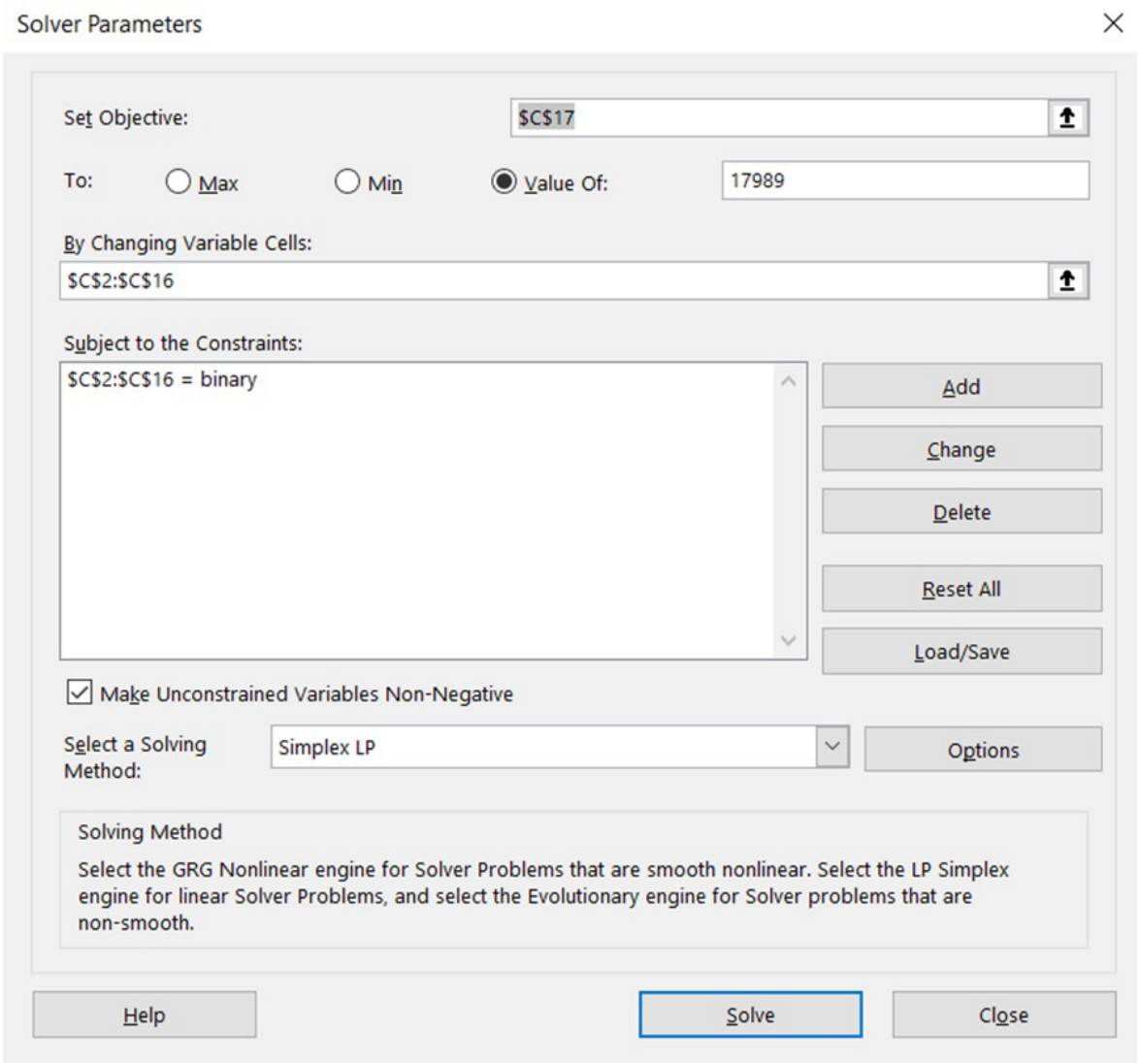


Fig 2.3.12 Setting Parameters

Click on Solve button and the excel will generate the result. Along with the result the excel will also generate the full summary of the operations. The resultant menu will be as shown below

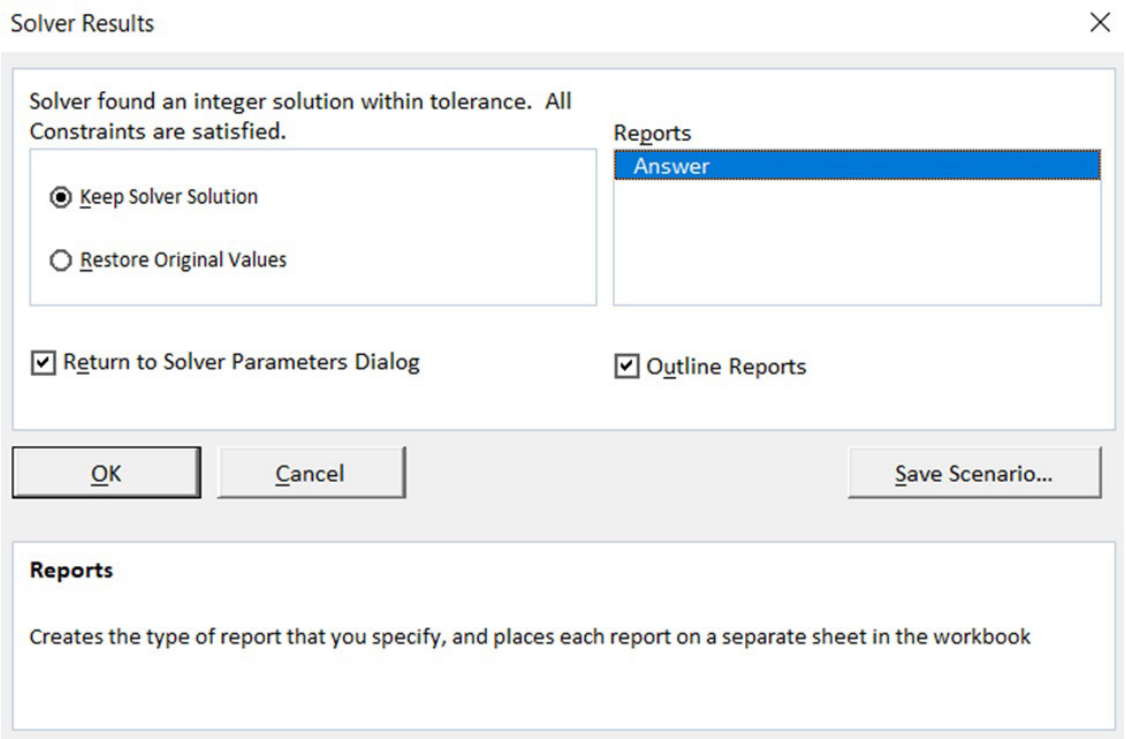


Fig 2.3.13 Solver Result Menu

Click on OK and Excel will show you the result of the solver and a detailed summary in another sheet. We can make use of answer report if we need it, or else we can just use the solver result. The result of solver will be as shown below

	A	B	C
1	Day	Sales (INR)	
2	Jul-01	2534	0
3	Jul-02	1875	1
4	Jul-03	3150	0
5	Jul-04	4026	0
6	Jul-05	5261	1
7	Jul-06	3702	0
8	Jul-07	4789	0
9	Jul-08	2750	1
10	Jul-09	3248	0
11	Jul-10	4389	0
12	Jul-11	5207	0
13	Jul-12	2136	1
14	Jul-13	3998	0
15	Jul-14	5967	1
16	Jul-15	3425	0
17		**Sum total**	17989
18			

Fig 2.3.14 Results using Solver

So now we can see that excel generated a column with 0 and 1. Adding up the "Sales" values with their row value as "Number 1" will give Rs.17989. The corresponding date values will give the dates on which the deposits were made.

This is one of the simplest and easiest method for solving an issue with solver. Solver is mainly used for solving optimisation problems. Let us work on a linear programming problem and see how we can use Solver to solve the problem.

Example 2: Let us say we have a manufacturing company that produces two products: Product A and Product B. The objective is to maximise the profit while considering resource limitations.

Decision Variables:

Let: x = Number of units of Product A to produce, y = Number of units of Product B to produce

Objective Function: Maximise Profit: Z = 8x + 5y

Constraints:

- Labor hours: The total labor hours used to produce both products should not exceed 400 hours.

  2x + 3y <= 400

- Raw materials: The total raw materials used to produce both products should not exceed 600 units.

  4x + 2y <= 600

- Storage space: The company has limited storage space, and the total units of both products should not exceed 200 units.

  x + y <= 200

- Non-negativity: The number of units produced cannot be negative.

  x >= 0

  y >= 0

According to a standard Linear Programming Problem, we need to have decision variables, we are trying to find the optimum amount of decision variables that will maximise or minimise according to an objective function subjected to constraints.

Here in our example, we are trying to find the quantities of Product A and Product B that will maximise the profit of the manufacturing company, subject to the given constraints. So, since this is in a text format, we need to convert it into excel format then only we can solve it using Solver.

Let us see how we can solve the above problem in Excel using Solver. Follow the steps as shown below.

1. Enter the data into Excel in the format given below so that it will make our procedure easier and convenient.
2. The variables are given as 0 so that it will make our calculations easier.
3. Listing the objective function and constraints in such a manner will make our calculations easier. There should be a column left empty between inequality sign and constraint value. Here in this example we have left the whole E column empty as shown below.

	A	B	C	D	E	F
1	**Variables**					
2	x	0				
3	y	0				
4						
5		**x**	**y**	**inequality**		**Value**
6	**Objective Function**					
7	Z=8x+3y	8	5	=		
8						
9	**Labour Hours**					
10	2x+3y<=400	2	3	<=		400
11						
12	**Raw Materials**					
13	4x+2y <=600	4	2	<=		600
14						
15	**Storage Space**					
16	x+y<=200	1	1	<=		200
17						
18	**Non Negativity**					
19	x>=0	1		>=		0
20	y>=0		1	>=		0
21						

Fig 2.3.15 Data into Excel Format

4. Now we calculate the values in that E column corresponding to each function. It will be as shown below

	A	B	C	D	E	F
1	**Variables**					
2	x	0				
3	y	0				
4						
5		**x**	**y**	**inequality**		**Value**
6	**Objective Function**					
7	Z=8x+3y	8	5	=	=B7*B2+C7*B3	
8						
9	**Labour Hours**					
10	2x+3y<=400	2	3	<=		400
11						
12	**Raw Materials**					
13	4x+2y <=600	4	2	<=		600
14						
15	**Storage Space**					
16	x+y<=200	1	1	<=		200
17						
18	**Non Negativity**					
19	x>=0	1		>=		0
20	y>=0		1	>=		0
21						

Fig 2.3.16 Calculating Objective Function

5. Similarly, we will find the values for each function. It will be as shown below.

	A	B	C	D	E	F
1	**Variables**					
2	x	0				
3	y	0				
4						
5		**x**	**y**	**inequality**		**Value**
6	**Objective Function**					
7	Z=8x+3y	8	5	=	0	
8						
9	**Labour Hours**					
10	2x+3y<=400	2	3	<=	0	400
11						
12	**Raw Materials**					
13	4x+2y <=600	4	2	<=	0	600
14						
15	**Storage Space**					
16	x+y<=200	1	1	<=	0	200
17						
18	**Non Negativity**					
19	x>=0	1		>=	0	0
20	y>=0		1	>=	0	0
21						

Fig 2.3.17 Calculating Constraints

6. All the resultant values will be equal to 0 as we have given the values of x and y as 0. Here we have tried to convert the mathematical equations with x and y variables into a constant value. Here we are trying to create a link between functions and variables. You can give any value to x and y.
7. Now we can try to solve the linear programming problem with the help of solver.

8. Click on Solver from the data tab as we have discussed earlier. A window will popup as shown below

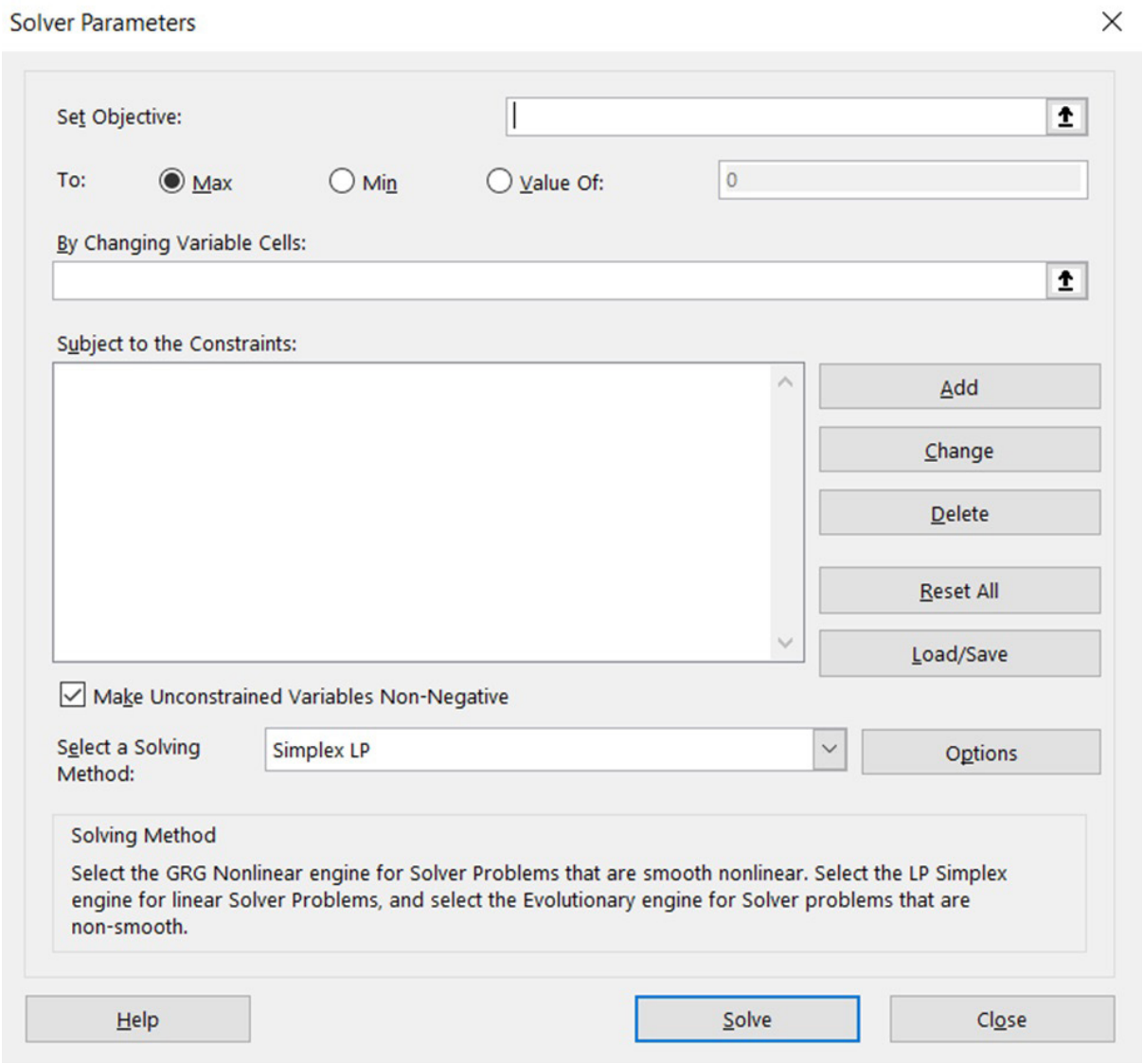


Fig 2.3.18 Setting Parameters

9. We have discussed earlier that we have to give an objective. And then changing variables and then finally the constrains to solve the linear programming problem.

10. The objective is the cell where we have found the value of the objective function. Here it is E7. And changing variables are B2 and B3 cells. It is as shown below. When you select the particular cell, the Excel will change it into the given format as shown below.

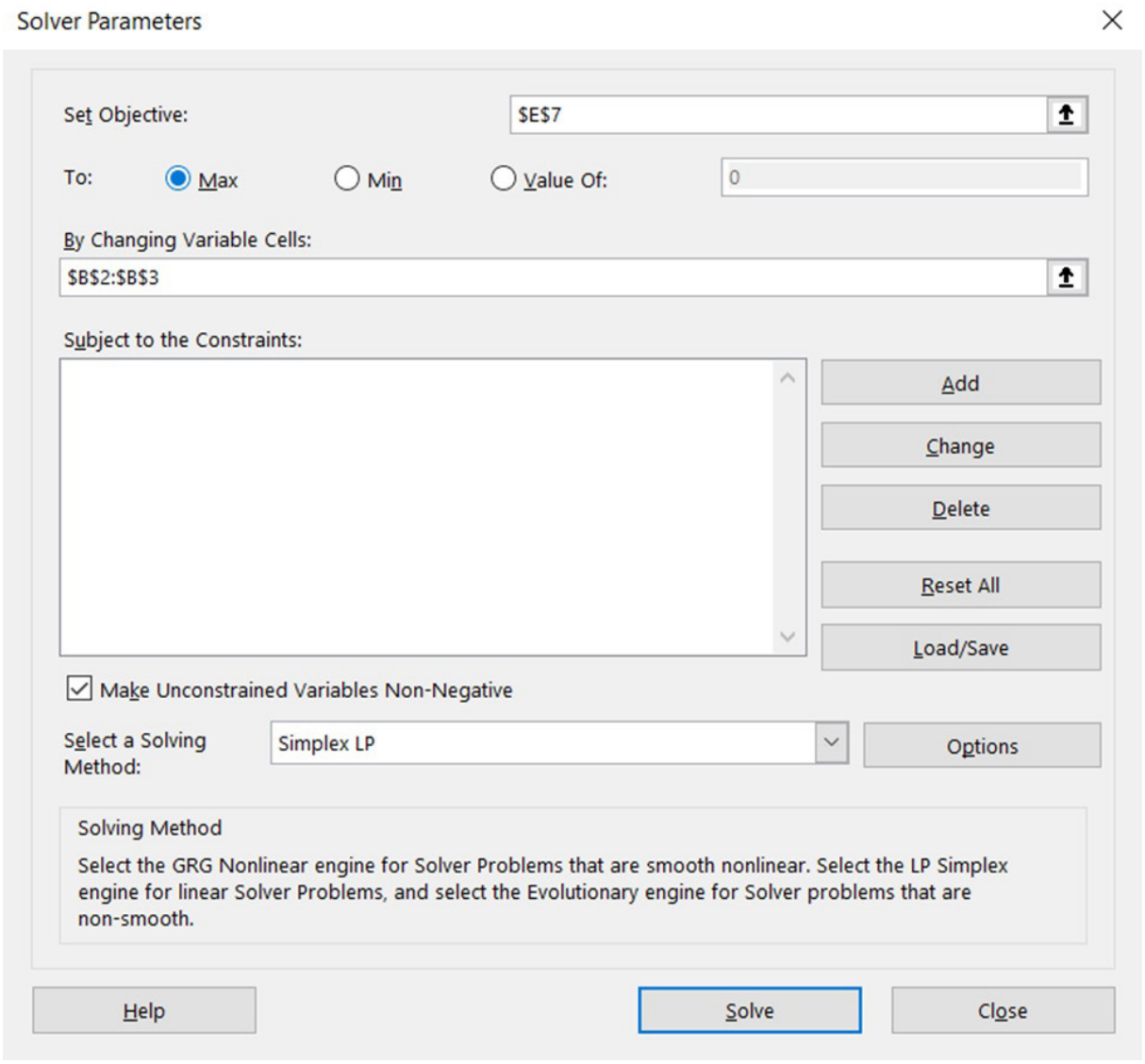


Fig 2.3.19 Setting Objective

11. Now we can add constraints one by one using the add option. Then a new window will pop up as shown below. Select the constraints the same way we have done in the previous example. Here the inequality symbols are noted in Excel itself so it will be easier for us to enter it without making errors.

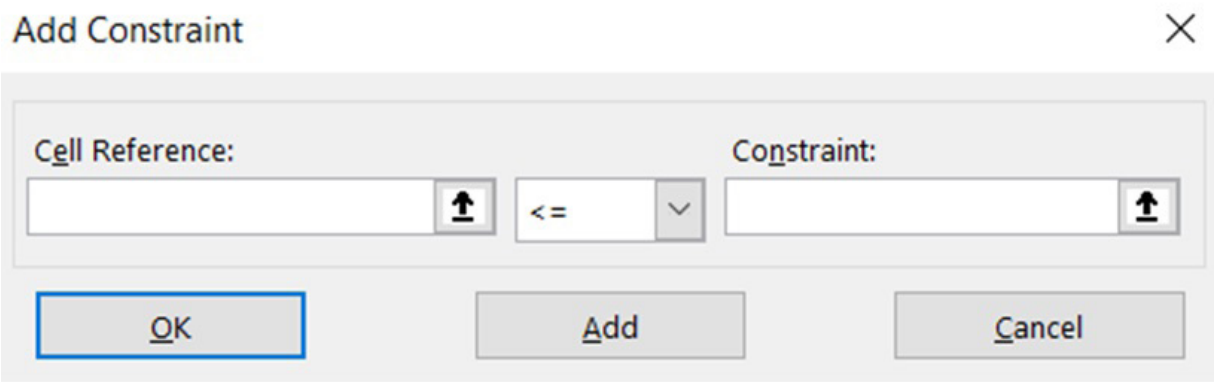
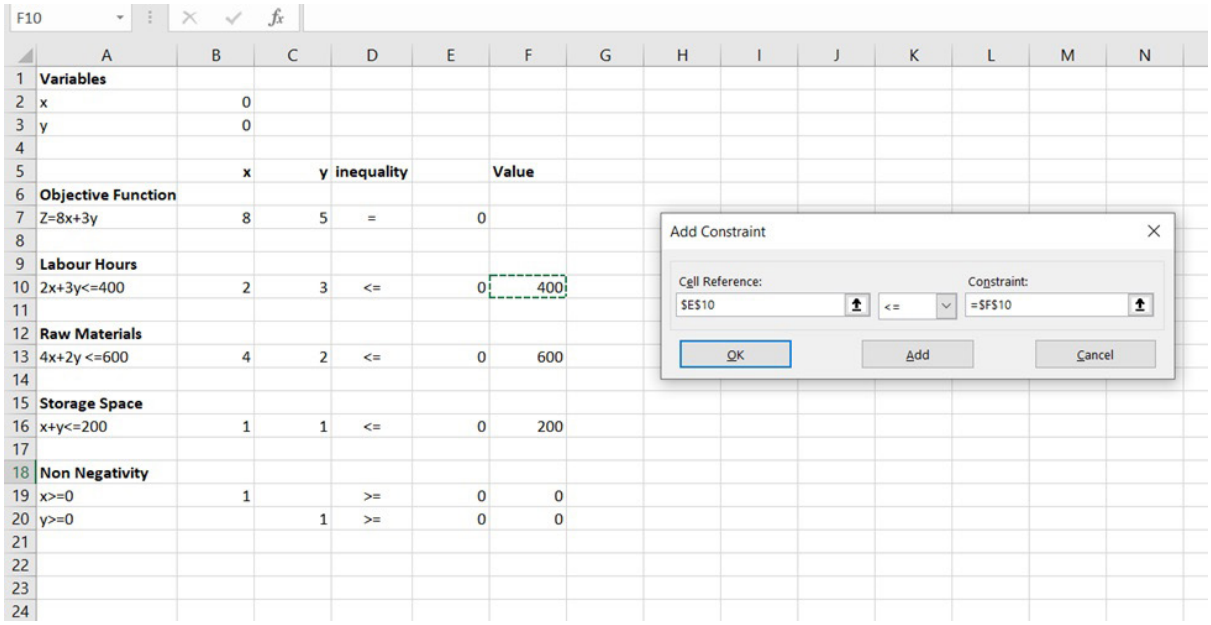


Fig 2.3.20 Setting Constraints

12. The cell reference is where we have to find the reultant value of the constraint function and constraint is the value of the constraint function. The symbol sign in the middle defines what kind of inequality we are defining. It will be as shown as below.



	A	B	C	D	E	F
1	**Variables**					
2	x	0				
3	y	0				
4						
5		**x**	**y**	**inequality**		**Value**
6	**Objective Function**					
7	Z=8x+3y	8	5	=	0	
8						
9	**Labour Hours**					
10	2x+3y<=400	2	3	<=	0	400
11						
12	**Raw Materials**					
13	4x+2y <=600	4	2	<=	0	600
14						
15	**Storage Space**					
16	x+y<=200	1	1	<=	0	200
17						
18	**Non Negativity**					
19	x>=0	1		>=	0	0
20	y>=0		1	>=	0	0
21						
22						
23						
24						

Fig 2.3.21 Selecting Values

13. After giving the input, click on "Add" to add new constraints. The first three are less than or equal to constraints and the last two ones are greater than or equal to constraint. So be careful when you give the inputs. Once completed give OK and the original solver window will pop back with all the inputs given as shown below.

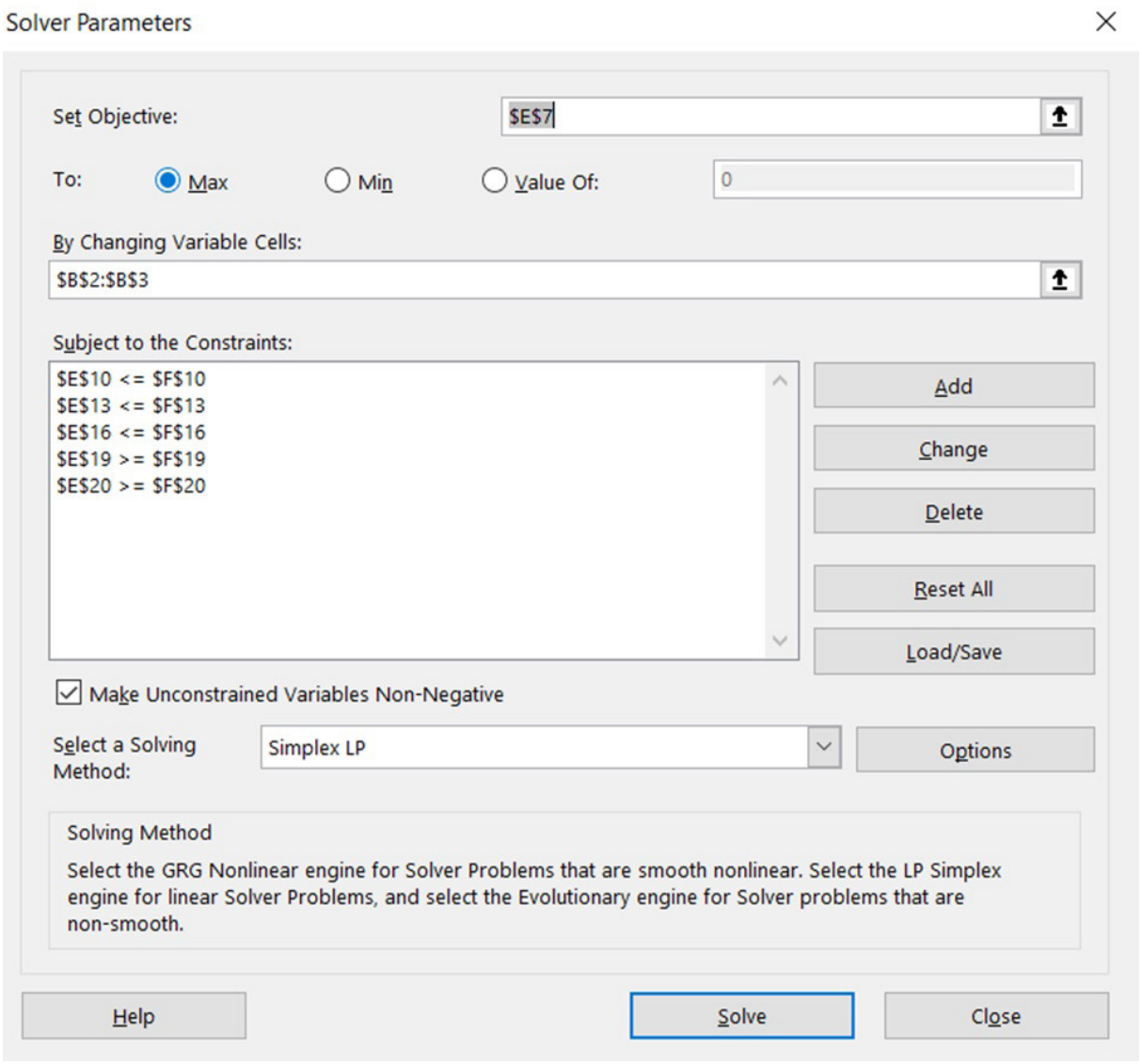


Fig 2.3.22 Solver Parameters

14. Make sure that the solving method selected is Simplex LP and click on Solve.
15. The new window will pop up as shown below

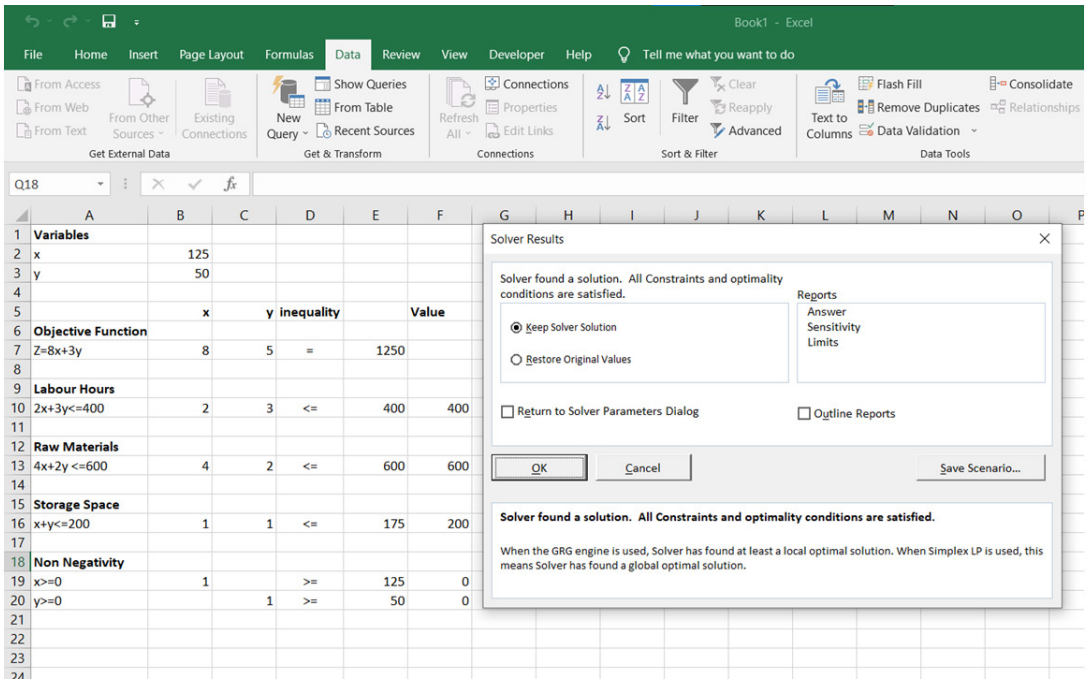


Fig 2.3.23 Solver Result Menu

16. Click on OK and the results will be as shown below.

	A	B	C	D	E	F
1	**Variables**					
2	x	125				
3	y	50				
4						
5		**x**	**y**	**inequality**		**Value**
6	**Objective Function**					
7	Z=8x+3y	8	5	=	1250	
8						
9	**Labour Hours**					
10	2x+3y<=400	2	3	<=	400	400
11						
12	**Raw Materials**					
13	4x+2y <=600	4	2	<=	600	600
14						
15	**Storage Space**					
16	x+y<=200	1	1	<=	175	200
17						
18	**Non Negativity**					
19	x>=0	1		>=	125	0
20	y>=0		1	>=	50	0

Fig 2.3.24 Results

We can see that in the B2 cell and B3 cells there are new values instead of 0. That is the result we were trying to find. The result can be interpreted as follows

- The company should produce 125 units of Product A and 50 units of Product B.
- The amount of maximum Profit will be 1250.
- All the constraints are satisfied.

Solver is a powerful tool, but it may not always find the global optimal solution, especially for complex non-linear problems or problems with multiple local optima.

2.3.2.2 Applying Solver to Real-World Scenarios

Solver is a powerful tool in Excel that helps in real-world scenarios by solving complex optimisation problems and providing optimal solutions to achieve specific goals. It can be applied to a wide range of fields and challenges, enabling data-driven decision-making, resource allocation, and process optimisation. Here is a detailed explanation of how Solver helps in real-world scenarios:

1. **Production Planning and Inventory Management:** Companies can use Solver to optimise production quantities of different products to maximise profit while considering resource constraints, demand, and inventory levels. It helps in determining the optimal inventory levels to minimise holding costs and stockouts while meeting customer demand.

2. **Resource Allocation and Workforce Management:** Solver assists in allocating resources effectively to different projects or tasks to maximise productivity and minimise costs. It helps in creating optimal work schedules for employees to meet demand while minimising labor costs and ensuring compliance with labor laws and work regulations.

3. **Financial and Investment Decision-Making:** Solver can be used to optimise investment allocations across different assets or projects to maximise returns while managing risk. It helps in determining the optimal capital budgeting decisions to select projects that yield the highest return on investment.

4. **Supply Chain Optimisation:** Solver helps in optimising the flow of goods and materials through the supply chain to minimise costs, reduce lead times, and improve efficiency. It aids in strategic decisions regarding sourcing, transportation, and inventory management to meet customer demands effectively.

5. **Project Management:** Solver assists in optimising project schedules and resource allocations to minimise project duration and costs. It helps in identifying critical paths and resource constraints to ensure projects are completed on time and within budget.

6. **Transportation and Logistics:** Solver can be used to optimise transportation routes and schedules to minimise transportation costs while meeting delivery demands. It helps in route optimisation, vehicle routing, and scheduling of shipments to improve efficiency and reduce operational costs.

7. **Portfolio Optimisation:** Solver aids in optimising investment portfolios by determining the optimal allocation of assets to achieve the desired risk-return trade-off. It helps investors to make informed decisions about asset allocation based on their risk tolerance and return objectives.

- Enables data-driven decision-making, resource allocation, process optimisation, and sensitivity analysis

8. **Marketing and Pricing Strategies:** Solver assists in finding the optimal pricing strategy to maximise revenue or profit based on market demand and cost considerations. It helps in determining the best mix of products or services to offer to customers to maximise overall profitability.

9. **Process Optimisation:** Solver can be used to optimise manufacturing processes, service operations, and supply chain workflows to improve efficiency and reduce costs. It helps in identifying the best sequence of operations and resource allocations to minimise bottlenecks and maximise throughput.

10. **Sensitivity Analysis and What-If Scenarios:** Solver allows for sensitivity analysis to understand how changes

in input parameters affect the optimal solution. It enables decision-makers to explore different what-if scenarios and assess the impact of various constraints and assumptions.

By leveraging the Solver, businesses and organisations can make more informed and data-driven decisions, optimise their operations, and find the best possible solutions to complex real-world challenges. It empowers users to solve sophisticated optimisation problems without the need for specialised software or complex mathematical models, making it a valuable tool for a wide range of industries and applications.

2.3.4 Advanced Functions in Excel

Excel offers a wide range of advanced functions that go beyond basic arithmetic and statistical operations. These advanced functions are designed to perform complex calculations, manipulate data, and solve specific problems efficiently. Here are some of the advanced functions in Excel:

1. **VLOOKUP**: The VLOOKUP function is used for vertical lookup. It searches for a specified value (the lookup value) in the leftmost column of a table or range. Once the match is found, it returns a value from a specified column in the same row. The syntax is: =VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])

2. **HLOOKUP**: The HLOOKUP function is used for horizontal lookup. It works similarly to VLOOKUP but searches for the lookup value in the top row of a table or range and returns a value from a specified row in the same column. The syntax is: =HLOOKUP(lookup_value, table_array, row_index_num, [range_lookup])

3. **INDEX**: The INDEX function returns the value of a cell in a specified row and column of a given range. It is often used in combination with other functions to retrieve values from a table. The syntax is: =INDEX(array, row_num, [column_num])

4. **MATCH**: The MATCH function searches for a specified value in a single row or column of a range and returns its relative position. It's commonly used to find the po-

sition of a lookup value in a lookup array. The syntax is: =MATCH(lookup_value, lookup_array, [match_type])

5. **IFERROR**: The IFERROR function allows you to handle errors in formulas. It returns a specified value if the formula evaluates to an error, and it returns the result of the formula if it does not produce an error. The syntax is: =IFERROR(value, value_if_error)
6. **ISERROR**: The ISERROR function checks if a cell contains an error value and returns TRUE if it does or FALSE if it does not. It is often used with IF functions to handle errors in calculations. The syntax is: =ISERROR(value).
7. **CONCATENATE**: The CONCATENATE function is used to join multiple text strings into one. You can specify the strings as separate arguments or as a range of cells to be concatenated. The syntax is: =CONCATENATE(-text1, [text2], ...)
8. **TEXTJOIN**: The TEXTJOIN function is similar to CONCATENATE but allows you to specify a delimiter to separate the text strings. It also provides options for ignoring empty cells or including non-text values. The syntax is: =TEXTJOIN(delimiter, ignore_empty, text1, [text2], ...)

- Advanced functions perform conditional calculations based on criteria

9. **SUMIFS**: The SUMIFS function is used for conditional summing. It adds up values in a range based on multiple criteria specified in different ranges. The syntax is: =SUMIFS(sum_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)
10. **COUNTIFS**: The COUNTIFS function is used for conditional counting. It counts the number of cells in a range that meet multiple criteria specified in different ranges. The syntax is: =COUNTIFS(criteria_range1, criteria1, [criteria_range2, criteria2], ...)
11. **AVERAGEIFS**: The AVERAGEIFS function calculates the average of a range based on multiple criteria. It considers only the cells that meet all the specified criteria. The syntax is: =AVERAGEIFS(average_range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)
12. **MAXIFS**: The MAXIFS function finds the maximum

value in a range based on specified conditions. It considers only the cells that meet all the specified criteria. The syntax is: =MAXIFS(range, criteria_range1, criteria1, [criteria_range2, criteria2], ...)

13. **INDIRECT**: The INDIRECT function allows you to create a reference to a cell or range indirectly using a text string as the reference. This is useful when working with dynamic cell references. The syntax is: =INDIRECT(ref_text, [a1]) . ref_text is the cell reference as a text string, and a1 is an optional argument to specify the type of reference (TRUE for A1-style, FALSE for R1C1-style).

By mastering these advanced functions, Excel users can perform complex calculations, analyse data, and solve specific problems more efficiently. These functions provide valuable tools for data manipulation, decision-making, and scenario analysis in various real-world applications.

2.3.4.1 VLOOKUP Function

Let us create a dataset for explaining VLOOKUP function in Excel. Suppose we have a dataset of employees in a company with the following columns:

Table 2.3.1 Employee Dataset

Employee ID	Name	Department	Salary (₹)	Experience (Years)
101	Aarav	Marketing	45000	5
102	Diya	Sales	55000	3
103	Aditi	HR	60000	8
104	Arjun	Operations	50000	6
105	Avni	Finance	65000	10
106	Ishaan	Marketing	48000	4
107	Aanya	Sales	52000	2
108	Aryan	HR	55000	7
109	Anaya	Operations	52000	5
110	Advait	Finance	68000	12
111	Sai	Marketing	47000	3
112	Virat	Sales	54000	5
113	Aadhya	HR	59000	9
114	Tanvi	Operations	51000	7
115	Aadi	Finance	67000	11

1. Enter the data into Excel and let us try using advanced functions in excel. It will be as shown below

	A	B	C	D	E
1	**Employee ID**	**Name**	**Department**	**Salary (₹)**	**Experience (Years)**
2	101	Aarav	Marketing	45000	5
3	102	Diya	Sales	55000	3
4	103	Aditi	HR	60000	8
5	104	Arjun	Operations	50000	6
6	105	Avni	Finance	65000	10
7	106	Ishaan	Marketing	48000	4
8	107	Aanya	Sales	52000	2
9	108	Aryan	HR	55000	7
10	109	Anaya	Operations	52000	5
11	110	Advait	Finance	68000	12
12	111	Sai	Marketing	47000	3
13	112	Virat	Sales	54000	5
14	113	Aadhya	HR	59000	9
15	114	Tanvi	Operations	51000	7
16	115	Aadi	Finance	67000	11

Fig 2.3.25 Employee Dataset In Excel

2. Now we are trying to find out the salary of employees against their name. Since the dataset is small it can be easily visually calculated but if the dataset has thousands of entries, then it will be difficult to find the necessary data. That is when we use the VLOOKUP function.
3. For that create another table adjacent to the existing table with column titles, Name and Salary, as shown below. Type in the name of employees, whose salary you want to identity, and on the cell below the salary column title type in the VLOOKUP syntax as shown below

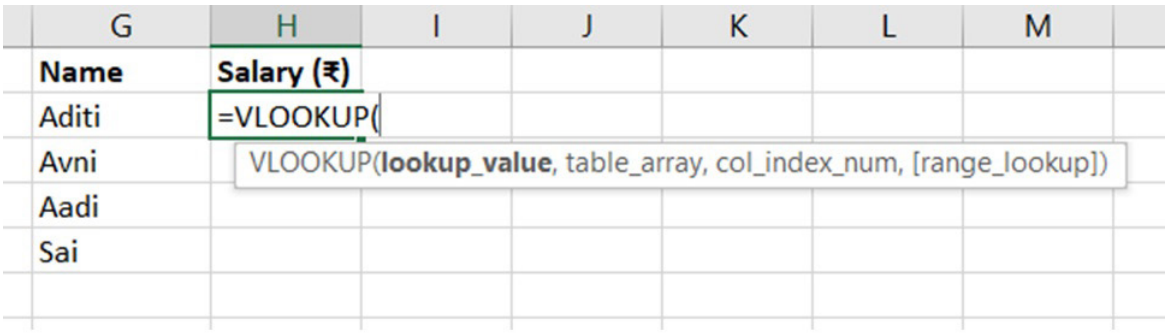


Fig 2.3.26 VLOOKUP Function

- **lookup_value** – is the Names column in the new table we created. So select the first cell in that i.e., cell G2
- **table_array** – is the original table values that the function should look up, so here we will select from Name column to Salary columns. Here it is the B:D. which means columns B to D .
- **col_index_num -** is the column position of the query in the table array. Here our query is salary and it is the 3rd column in the table array, so we should type in number 3.
- **range_lookup-** TRUE if you want an approximate match or FALSE if you want an exact match of the return value. Here we want an exact match so select-False.

4. The syntax will look as given below after entering the data,

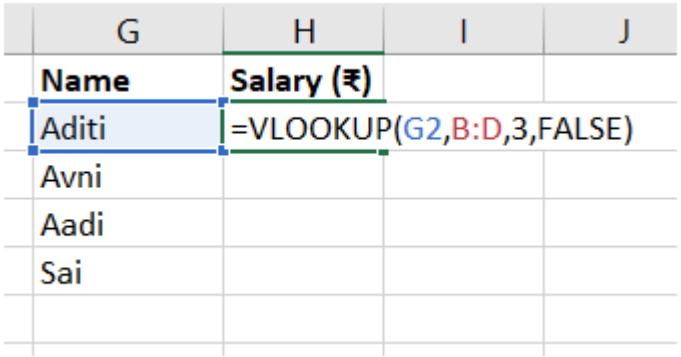


Fig 2.3.27 VLOOKUP Syntax

5. Press Enter and excel will generate the result as shown below

	G	H
	Name	**Salary (₹)**
	Aditi	60000
	Avni	
	Aadi	
	Sai	

Fig 2.3.28 Results of VLOOKUP Function

6. Repeat the steps for each query or use auto fill function. The result will be as shown below.

- Allows users to retrieve information from large datasets and organise data

	G	H
	Name	**Salary (₹)**
	Aditi	60000
	Avni	65000
	Aadi	67000
	Sai	47000

Fig 2.3.29 Results of VLOOKUP Function for the Dataset

By following these steps, you have used VLOOKUP to find the salary for each employee based on their name. VLOOKUP is a powerful function in Excel for performing lookups and retrieving information from other tables, making it easier to analyse and organise data from different sources.

2.3.4.2 INDEX Function

The INDEX function in Microsoft Excel is a powerful and versatile tool used to retrieve data from a specified range or array. It allows you to extract a value from a table based on its row and column positions.

Let us create a dataset and use it to explain step-by-step how to use the INDEX function in Excel. let us create a dataset of mobile phone sales in India for Q1, Q2, Q3, and Q4 for the first 10 mobile brands.

Table 2.3. 2 Mobile Brand - Quarterly Sales Data

Brand	Q1 Sales	Q2 Sales	Q3 Sales	Q4 Sales
Samsung	25000	28000	32000	30000
Xiaomi	20000	22000	26000	27000
Apple	15000	18000	19000	23000
Vivo	18000	20000	21000	24000
Oppo	22000	24000	27000	28000
Realme	16000	19000	22000	25000
OnePlus	10000	12000	14000	18000
Lenovo	8000	9000	10000	12000
Nokia	12000	14000	16000	20000
Micromax	10000	11000	13000	16000

1. Enter the dataset into Excel and it will be as follows

	A	B	C	D	E
1	**Brand**	**Q1 Sales**	**Q2 Sales**	**Q3 Sales**	**Q4 Sales**
2	Samsung	25000	28000	32000	30000
3	Xiaomi	20000	22000	26000	27000
4	Apple	15000	18000	19000	23000
5	Vivo	18000	20000	21000	24000
6	Oppo	22000	24000	27000	28000
7	Realme	16000	19000	22000	25000
8	OnePlus	10000	12000	14000	18000
9	Lenovo	8000	9000	10000	12000
10	Nokia	12000	14000	16000	20000
11	Micromax	10000	11000	13000	16000

Fig 2.3.30 Mobile Brand - Quarterly Sales Data in Excel

1. The INDEX function in Excel has the following

syntax:

=INDEX(array, row_num, [column_num])

- **array:** This is the range or array from which you want to retrieve data. It can be a single column, single row, or a range containing multiple rows and columns.
- **row_num:** This is the row number of the value you want to retrieve from the array. The first row is numbered as 1.
- **column_num (optional):** This is the column number of the value you want to retrieve from the array. If omitted, the function will return the entire row specified by row_num.

2. Suppose we want to retrieve the Q2 sales figure for the brand "Apple". To do this, we'll use the INDEX function as follows:

- Click on an empty cell where you want the result and type the formula
- The array will be just the data from A2 to E11
- Since we want Q2 sales of Apple the row_num will be 3
- And column_num will also be 3.

- Retrieve data from a specified range or array based on row and column positions

3. The excel will generate the result in that particular output cell

The INDEX function in Excel allows you to extract specific values from a given range or array based on row and column positions. By understanding its syntax and using it effectively, you can efficiently retrieve data from your datasets and perform various calculations and analyses in your spreadsheets.

2.3.4.3 MATCH Function

We can use the dataset from table for calculating MATCH function.

1. Enter the dataset into excel as follows,

	A	B	C	D	E	F
1	**Position**	**Brand**	**Q1 Sales**	**Q2 Sales**	**Q3 Sales**	**Q4 Sales**
2	1	Samsung	25000	28000	32000	30000
3	2	Xiaomi	20000	22000	26000	27000
4	3	Apple	15000	18000	19000	23000
5	4	Vivo	18000	20000	21000	24000
6	5	Oppo	22000	24000	27000	28000
7	6	Realme	16000	19000	22000	25000
8	7	OnePlus	10000	12000	14000	18000
9	8	Lenovo	8000	9000	10000	12000
10	9	Nokia	12000	14000	16000	20000
11	10	Micromax	10000	11000	13000	16000

Fig 2.3.31 Mobile Brand - Quarterly Sales Data for MATCH Function

2. The MATCH function in Excel has the following syntax:

=MATCH(lookup_value, lookup_array, [match_type])

- **lookup_value:** This is the value you want to find within the lookup_array.
- **lookup_array:** This is the range or array where you want to search for the lookup_value.
- **match_type (optional)**: This is an optional argument that specifies how the match should be performed. It can be one of the following values:
    - 0 or omitted: Exact match (default).
    - 1: Find the largest value less than or equal to the lookup_value (lookup_array must be in ascending order).
    - -1: Find the smallest value greater than or equal to the lookup_value (lookup_array must be in descending order).

Let's use the MATCH function to find the position of a specified value within the dataset.

3. Suppose we want to find the Position of the brands within the dataset. To do this, we will use the MATCH function as follows:

- Create another table next to the original table
- Give the column titles as given below

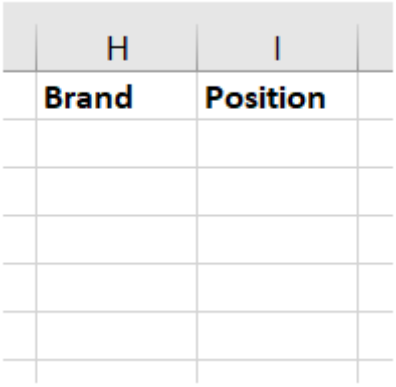
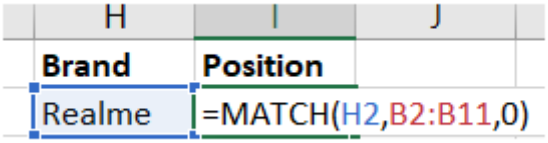


Fig 2.3.32 Creating New Columns

- Type in any phone brand in the cell below the "Brand" column and type the following formula in the cell below "Position" column as shown below



-

Fig 2.3.33 Typing MATCH Function

- Find the position of a specified value within a given range or array

H2 is the cell where we give the input, to match. B2 to B11 is the reference point, and we have given 0 since we wanted an exact match.

4. Excel will generate the result and give us the position of Realme. Similarly, we can give any input and find out the position. It will be as shown below

H	I
Brand	**Position**
Realme	6
Vivo	4
Apple	3

Fig 2.3.34 Results of MATCH Function

The MATCH function searches for a specified item in a range of cells, and then returns the relative position of that item in the range.

2.3.4.4 INDEX-MATCH Function

The INDEX-MATCH combination is a powerful technique in Excel that allows you to retrieve data from a specific cell in a table based on the values in the row and column headers. It is often used when VLOOKUP or HLOOKUP is not suitable for the task, or when you need to perform a two-dimensional lookup.

Let us go through the steps of using the INDEX-MATCH function in the above dataset:

1. The INDEX-MATCH function is a combination of two functions:
   - INDEX: This function retrieves a value from a specified range or array based on its row and column numbers.
   - MATCH: This function finds the position of a specified value within a range or array.

   The basic syntax of INDEX-MATCH is as follows:

   =INDEX(range, MATCH(lookup_value, lookup_array, [match_type]), [column_index])

   - **range:** This is the range from which you want to retrieve the value.
   - **lookup_value:** This is the value you want to find within the lookup_array.
   - **lookup_array:** This is the range or array where you want to search for the lookup_value.
   - **match_type:** This is an optional argument that specifies how the match should be performed. It can

be 0 (exact match) or omitted, -1 (find the smallest value greater than or equal to the lookup_value), or 1 (find the largest value less than or equal to the lookup_value).

- **column_index**: This is the optional argument that specifies the column number from which the value should be retrieved from the range. If omitted, the INDEX function returns the entire row.

Let us look at an example to use INDEX-MATCH combination.

1. Let us use the same dataset we have used in the previous example. It is as shown below.

	A	B	C	D	E
1	**Brand**	**Q1 Sales**	**Q2 Sales**	**Q3 Sales**	**Q4 Sales**
2	Samsung	25000	28000	32000	30000
3	Xiaomi	20000	22000	26000	27000
4	Apple	15000	18000	19000	23000
5	Vivo	18000	20000	21000	24000
6	Oppo	22000	24000	27000	28000
7	Realme	16000	19000	22000	25000
8	OnePlus	10000	12000	14000	18000
9	Lenovo	8000	9000	10000	12000
10	Nokia	12000	14000	16000	20000
11	Micromax	10000	11000	13000	16000

Fig 2.3.35 Mobile Brand - Quarterly Sales Data for INDEX-MATCH Function

2. To make the process easier let us make a drop-down list of Brands and Quarter using Data Validation tool.
   - Create two new columns titled "Brand" and "Quarter"
   - Below Brand title, select a cell and click on Data Ribbon and select data Validation option from that. Select data validation from the list as shown below.

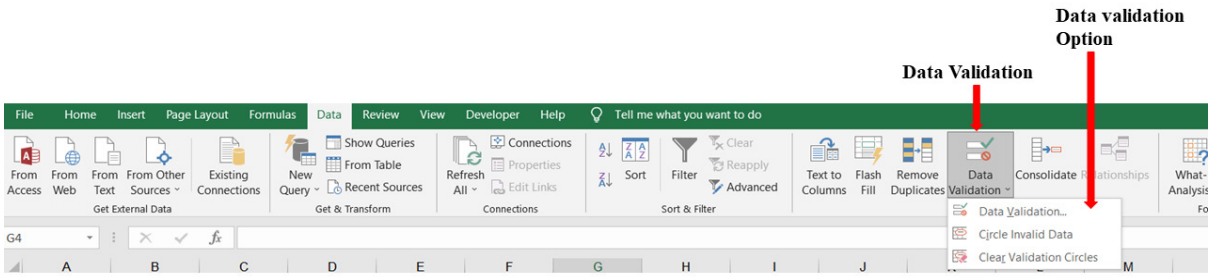


Fig 2.3.36 Data Validation Tool

- Select List option from the menu and select the source from the data set as shown below

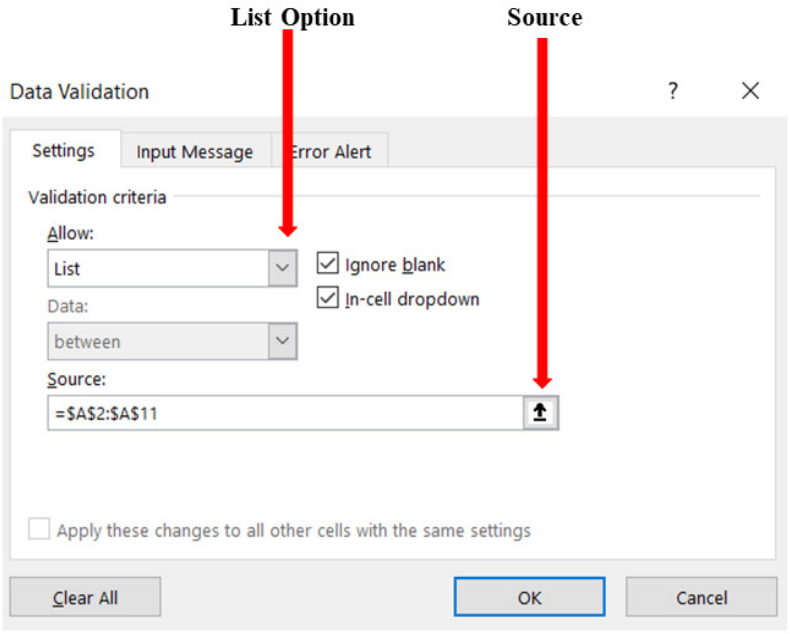


Fig 2.3.37 Making A List Using Data Validation

- Click OK and Excel will generate a drop-down list of all the brands from the list we have given. It will be as shown below.

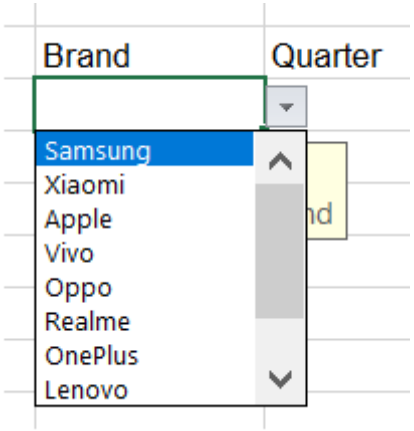


Fig 2.3.38 List of Mobiles

- Similarly make a list of all the 4 quarters and it will be as shown below

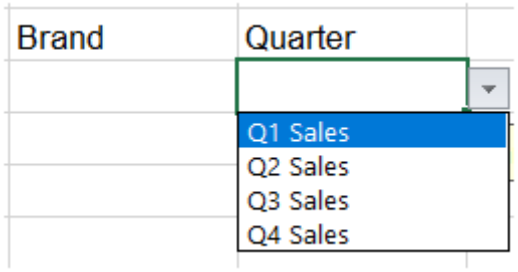


Fig 2.3.39 List Of Quarters

3. Now we can give the formula for INDEX-MATCH combination to find out the Q3 sales of Apple.
   - First choose apple from the brands list in the drop-down menu and Q3 from the quarter list as shown below.
   - On the cell adjacent to the Q3 sales or where ever you want the result to be published type in the INDEX-MATCH combination formula as given below

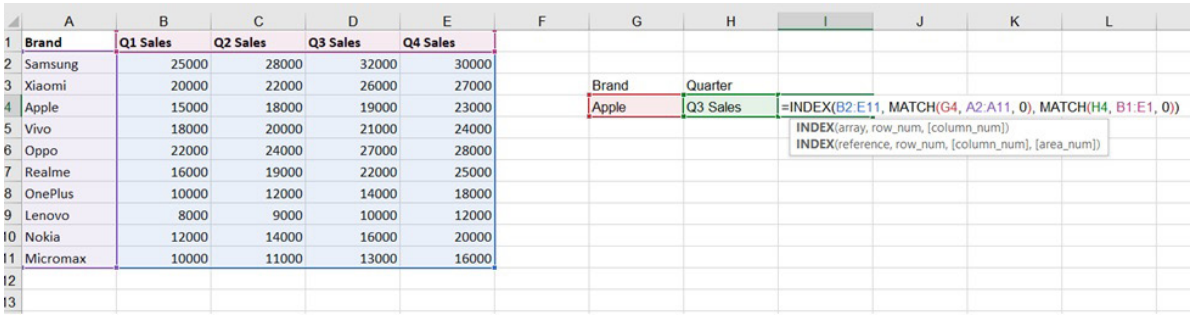


Fig 2.3.40 Selecting Data For INDEX-MATCH Function

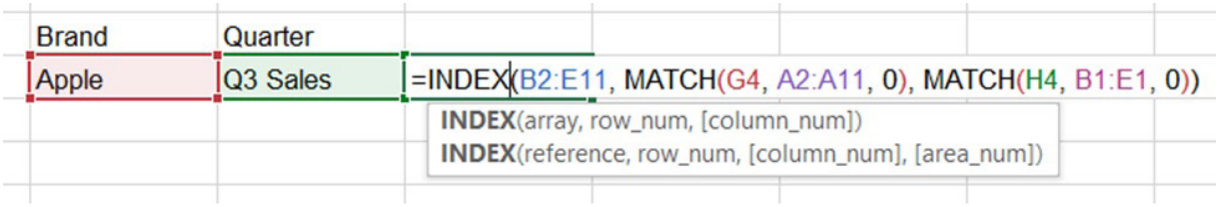


Fig 2.3.41 INDEX-MATCH Function

Here we have used INDEX-MATCH combination

- First, we typed in the = sign and then we entered the INDEX formula.
- Then we have selected the data set from B2 to E11 as array
- In order to find the row_num we have used the MATCH formula
- Inside the MATCH formula, we have selected the cell with the brands list as it is the lookup value
- The lookup array is the list of all the brands that are from A2 to A11.
- Since we wanted the exact match, we typed 0
- Repeated the same for column_num also.
- We have used the MATCH function again and given the quarter list as lookup value and selected the Quarter title cells as lookup arrays. Those cells are B1 to E1. Since we wanted the exact match, we typed 0
- Close the brackets carefully and click enter

4. Excel will generate the result as shown below

Brand	Quarter	
Apple	Q3 Sales	19000

Fig 2.3.42 Results of INDEX-MATCH Function

Here we can change the Brand and Quarter from the list, and Excel will generate the correct result each time.

The INDEX-MATCH function in Excel provides a flexible and powerful way to perform two-dimensional lookups, making it useful for various data analysis and retrieval tasks. By understanding its syntax and using it effectively, you can efficiently retrieve data from your datasets based on specific criteria and conditions. The INDEX-MATCH combination is a valuable addition to your Excel toolkit, especially when VLOOKUP or HLOOKUP does not meet your requirements

2.3.5 Logical Functions in Excel

Logical functions in Excel are tools that allow you to perform logical operations and make decisions based on certain conditions. They return either TRUE or FALSE values, which can

be used in various ways to analyse and manipulate data. Excel provides several important logical functions, including IF, AND, OR, NOT, and more.

2.3.5.1 IF Function:

The IF function is one of the most commonly used logical functions in Excel. It allows you to perform a test and return different values based on whether the test is true or false.

The syntax of the IF function is as follows: **= IF(logical_test, value_if_true, value_if_false)**

- **logical_test:** This is the condition you want to evaluate. It can be a comparison, text, number, or any expression that results in either TRUE or FALSE.
- **value_if_true:** This is the value that the function returns if the logical_test is TRUE.
- **value_if_false:** This is the value that the function returns if the logical_test is FALSE.

Let us create a dataset and demonstrate how to use the IF functions in Excel. For this example, we will create a dataset of employees and their performance ratings for that particular year.

Table 2.3.3 Employee Rating

Employee Name	Yearly rating
John	60
Jane	55
Mike	63
Sarah	59
Alex	70
Emily	71

1. Enter the data into Excel as shown below

	A	B
1	**Employee Name**	**Yearly rating**
2	John	60
3	Jane	55
4	Mike	63
5	Sarah	59
6	Alex	70
7	Emily	71

Fig 2.3.43 Employee Rating in Excel Format

2. Here we are trying to convert the numerical rating into a rating system where it will inform whether they are Good or whether they Need Improvement.

3. So, for that the management has come up with a rating system where any employee with a rating more than 60 should be given Good rating and anyone with rating below 60 should be given a Need Improvement rating.

4. Here we are using IF condition for generating the results. Type the formula as shown below.

	A	B	C	D	E
1	**Employee Name**	**Yearly rating**	**Rating**		
2	John	60	=IF(B2>=60, "Good", "Needs Improvement")		
3	Jane	55			
4	Mike	63			
5	Sarah	59			
6	Alex	70			
7	Emily	71			

IF(**logical_test**, [value_if_true], [value_if_false])

Fig 2.3.44 Using IF Function

- B2>= 60 is the logical condition according to which the IF condition will work.
- If the condition is true then the test will return the value “Good” or else it will return the value “Needs Improvement”.

5. The result will be as shown below

	A	B	C
1	**Employee Name**	**Yearly rating**	**Rating**
2	John	60	Good
3	Jane	55	
4	Mike	63	
5	Sarah	59	
6	Alex	70	
7	Emily	71	

Fig 2.3.45 Result of IF Function

6. The result can be formulated for other employees too by simple auto filling. The result will be as shown below.

	A	B	C
1	**Employee Name**	**Yearly rating**	**Rating**
2	John	60	Good
3	Jane	55	Needs Improvement
4	Mike	63	Good
5	Sarah	59	Needs Improvement
6	Alex	70	Good
7	Emily	71	Good

Fig 2.3.46 Result of IF Function for the Entire Dataset

7. We can give more conditions and use multiple IF conditions inside the IF condition and it is called Nesting.

8. Let us give more conditions than given before. Let the conditions be,

Table 2.3.4 Conditions

Value	Rating
<= 60	Needs Improvement
<60 & >70	Good
>=70	Excellent

9. So here we have added one more condition so we will need to add one more IF condition to get the specified results

10. The formula for this will be as shown below

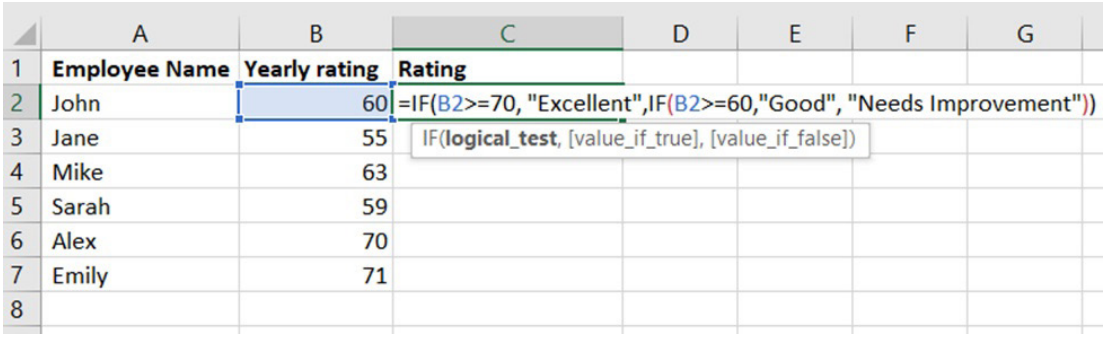


Fig 2.3.47 IF Function for Nesting

The first condition is for the logical test. i.e. if the value of B2 is >= 70 then it will return the result Excellent. The next IF is added in the false value area, so it means that Excel will check it only if the first condition is false, so inside the second IF condition we have given our true and false conditions. It will return the results as shown below

	A	B	C
1	**Employee Name**	**Yearly rating**	**Rating**
2	John	60	Good
3	Jane	55	
4	Mike	63	
5	Sarah	59	
6	Alex	70	
7	Emily	71	
8			

Fig 2.3.48 Result of Nesting

11. The results for the rest of the employees will be as shown below.

	A	B	C
1	**Employee Name**	**Yearly rating**	**Rating**
2	John	60	Good
3	Jane	55	Needs Improvement
4	Mike	63	Good
5	Sarah	59	Needs Improvement
6	Alex	70	Excellent
7	Emily	71	Excellent

Fig 2.3.49 Result of Nesting for the Entire Dataset

We can see the new rating is displayed. This is how we use multiple IF conditions or nesting for improving the quality of our results.

2.3.5.2 AND Function:

The AND function checks multiple conditions simultaneously and returns TRUE only if all the conditions are TRUE.

The syntax of the AND function is as follows: **=AND(logical1, logical2, ...)**

- logical1, logical2, ...: These are the conditions you want to check. You can have multiple conditions sepa-

rated by commas.

Here we are using a company's employee ratings for 4 quarters.

Table 2.3.5 Employee Quarter Wise Rating

Employee Name	Q1 Rating	Q2 Rating	Q3 Rating	Q4 Rating
John	Excellent	Good	Excellent	Good
Jane	Good	Good	Excellent	Excellent
Mike	Excellent	Excellent	Excellent	Excellent
Sarah	Good	Good	Good	Good
Alex	Good	Excellent	Good	Good
Emily	Excellent	Good	Good	Excellent

1. Enter the data set into Excel as shown below.

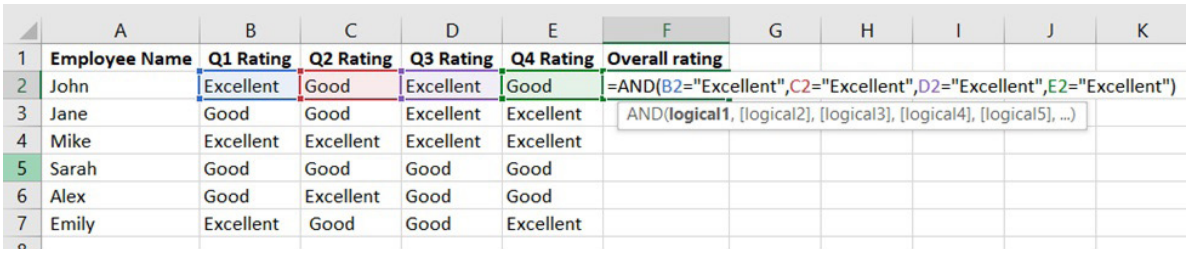
	A	B	C	D	E
1	**Employee Name**	**Q1 Rating**	**Q2 Rating**	**Q3 Rating**	**Q4 Rating**
2	John	Excellent	Good	Excellent	Good
3	Jane	Good	Good	Excellent	Excellent
4	Mike	Excellent	Excellent	Excellent	Excellent
5	Sarah	Good	Good	Good	Good
6	Alex	Good	Excellent	Good	Good
7	Emily	Excellent	Good	Good	Excellent

Fig 2.3.50 Employee Quarter Wise Rating in Excel

2. Now we are trying to find who all are the Excellent employees in our company. So for that we are using AND function. It will generate TRUE or FALSE responses. If all the logical conditions are true then it will generate a TRUE response or else it will be a FALSE. Here we are trying to find the excellent employee based on rating for 4 quarters. If all the ratings are Excellent then it should generate TRUE

and if not then FALSE.

3. So, enter the conditions as given below.



	A	B	C	D	E	F	G	H	I	J	K
1	Employee Name	Q1 Rating	Q2 Rating	Q3 Rating	Q4 Rating	Overall rating					
2	John	Excellent	Good	Excellent	Good	=AND(B2="Excellent",C2="Excellent",D2="Excellent",E2="Excellent")					
3	Jane	Good	Good	Excellent	Excellent						
4	Mike	Excellent	Excellent	Excellent	Excellent						
5	Sarah	Good	Good	Good	Good						
6	Alex	Good	Excellent	Good	Good						
7	Emily	Excellent	Good	Good	Excellent						

Fig 2.3.51 Using AND Function

4. Excel will generate the result as shown below.

	A	B	C	D	E	F
1	Employee Name	Q1 Rating	Q2 Rating	Q3 Rating	Q4 Rating	Overall rating
2	John	Excellent	Good	Excellent	Good	FALSE
3	Jane	Good	Good	Excellent	Excellent	
4	Mike	Excellent	Excellent	Excellent	Excellent	
5	Sarah	Good	Good	Good	Good	
6	Alex	Good	Excellent	Good	Good	
7	Emily	Excellent	Good	Good	Excellent	

Fig 2.3.52 Result of AND Formula

Here we can see that John did not get a TRUE response because he as got both excellent and good. Let us check the results of other employees too.

5. The results of other employees are as shown below.

	A	B	C	D	E	F
1	Employee Name	Q1 Rating	Q2 Rating	Q3 Rating	Q4 Rating	Overall rating
2	John	Excellent	Good	Excellent	Good	FALSE
3	Jane	Good	Good	Excellent	Excellent	FALSE
4	Mike	Excellent	Excellent	Excellent	Excellent	TRUE
5	Sarah	Good	Good	Good	Good	FALSE
6	Alex	Good	Excellent	Good	Good	FALSE
7	Emily	Excellent	Good	Good	Excellent	FALSE

Fig 2.3.53 Result of AND Function for the Entire Dataset

Here we can clearly identify that only MIKE got a TRUE response as he is the only one with consistent results in all the quarters.

OR Function:

The OR function checks multiple conditions simultaneously and returns TRUE if any of the conditions are TRUE.

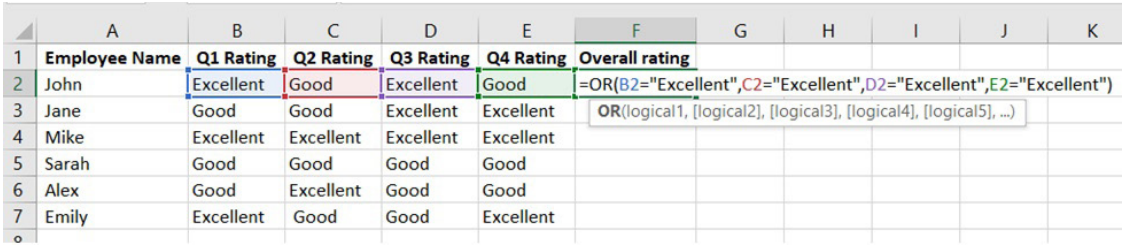
The syntax of the OR function is as follows: **=OR(logical1, logical2, ...)**

- logical1, logical2, ...: These are the conditions you want to check. You can have multiple conditions separated by commas.

We can use the same example we used in the previous section. There we were trying to find who was the true excellent employee with all excellent ratings. Here we can try and see who is not able to perform that well and who could not score even one excellent rating in any one quarter.

So inorder to find that we can use the formula OR. The steps are as follows:

1. Enter the data and use the formula given below to find out which all employees are not performing well and who all could not score even one excellent rating in atleast one quarter.



	A	B	C	D	E	F	G	H	I	J	K
1	Employee Name	Q1 Rating	Q2 Rating	Q3 Rating	Q4 Rating	Overall rating					
2	John	Excellent	Good	Excellent	Good	=OR(B2="Excellent",C2="Excellent",D2="Excellent",E2="Excellent")					
3	Jane	Good	Good	Excellent	Excellent	OR(logical1, [logical2], [logical3], [logical4], [logical5], ...)					
4	Mike	Excellent	Excellent	Excellent	Excellent						
5	Sarah	Good	Good	Good	Good						
6	Alex	Good	Excellent	Good	Good						
7	Emily	Excellent	Good	Good	Excellent						

Fig 2.3.54 Using OR Function

2. So the OR function will generate a TRUE is any one of the logical arguments satisfies or generate a FALSE if all the logical conditions fail. Here we have given the logical condition for 4 cells as Excellent. So if any one of the 4 conditions satisfies then it will show the result as TRUE.

3. The result for the first row will be as shown below.

	A	B	C	D	E	F
1	**Employee Name**	**Q1 Rating**	**Q2 Rating**	**Q3 Rating**	**Q4 Rating**	**Overall rating**
2	John	Excellent	Good	Excellent	Good	TRUE
3	Jane	Good	Good	Excellent	Excellent	
4	Mike	Excellent	Excellent	Excellent	Excellent	
5	Sarah	Good	Good	Good	Good	
6	Alex	Good	Excellent	Good	Good	
7	Emily	Excellent	Good	Good	Excellent	
8						

Fig 2.3.55 Result of OR Function

So here we can see that unlike the previous section we actually got a TRUE here, it is because John was rated Excellent in Q1 and Q3. So the condition was satisfied and Excel generated the result as TRUE.

4. Let us find the result for rest of the employees. It will be as shown below.

	A	B	C	D	E	F
1	**Employee Name**	**Q1 Rating**	**Q2 Rating**	**Q3 Rating**	**Q4 Rating**	**Overall rating**
2	John	Excellent	Good	Excellent	Good	TRUE
3	Jane	Good	Good	Excellent	Excellent	TRUE
4	Mike	Excellent	Excellent	Excellent	Excellent	TRUE
5	Sarah	Good	Good	Good	Good	FALSE
6	Alex	Good	Excellent	Good	Good	TRUE
7	Emily	Excellent	Good	Good	Excellent	TRUE
8						

Fig 2.3.56 Result of OR Function for the Entire Dataset

Here we can see that everyone except Sarah was able to get a TRUE. Sarah is the only employee who could not score an Excellent rating in any of the quarters.

Logical functions in Excel are used to perform logical operations and evaluate conditions within a spreadsheet. These functions return either TRUE or FALSE based on the specified criteria.

2.3.6 Array Formulas and Dynamic Arrays

In Excel, an array is a collection of values or data elements arranged in a rectangular grid of rows and columns. Excel allows you to work with arrays to perform various calculations and operations efficiently. Unlike traditional cell references that refer to a single cell or range, arrays in Excel can refer to multiple cells or ranges simultaneously.

An array constant is a group of values entered directly into a formula or a range of cells without using any functions. When you enclose the values with curly braces {}, Excel automatically recognises them as an array constant. For example, you can create a simple array constant by entering {1, 2, 3, 4} into a cell, and it will be treated as an array with four elements.

1. Array Formulas:

In Excel, an array formula is a special type of formula that operates on arrays of data, rather than single cell references. It allows you to perform calculations on multiple cells at once and return multiple results. When using an array formula, you must enter it by pressing CTRL + SHIFT + ENTER, which adds curly braces {} around the formula to indicate that it's an array formula.

Example of a regular formula:

- Enable performing calculations on arrays of data

Let us say you have a range of numbers from A1 to A5, and you want to sum these numbers. You could use a regular formula like =SUM(A1:A5), which will give you the sum of those five cells.

Example of an array formula:

Using an array formula, you can achieve the same result in a single cell, without using the SUM function. To do this, you would enter the following formula: {=A1:A5}, followed by pressing CTRL + SHIFT + ENTER. The formula will automatically return the sum of the numbers in the array.

Array formulas can be used for various complex calculations, including mathematical operations, conditional statements, and many more. They are particularly useful when dealing with large datasets and performing calculations that would otherwise require multiple cells and formulas.

Example : We have the sales data of a company for 12 months

Table 2.3.6 Monthly - Quantity - Price Dataset

Month	Quantity Sold	Price per Unit (INR)
January	150	100
February	200	110
March	180	120
April	250	105
May	220	115
June	190	125
July	170	130
August	210	120
September	240	115
October	200	110
November	180	125
December	230	105

1. Enter the data into Excel as shown below

	A	B	C
1	**Month**	**Quantity Sold**	**Price per Unit (INR)**
2	January	150	100
3	February	200	110
4	March	180	120
5	April	250	105
6	May	220	115
7	June	190	125
8	July	170	130
9	August	210	120
10	September	240	115
11	October	200	110
12	November	180	125
13	December	230	105

Fig 2.3.57 Monthly - Quantity - Price Dataset in Excel

Let us try and find out the total sales in each month and total sales in the whole year using array formula

2. Create another column titled Total Sales after the Price per unit column, as shown below

	A	B	C	D
1	**Month**	**Quantity Sold**	**Price per Unit (INR)**	**Total Sales**
2	January	150	100	
3	February	200	110	
4	March	180	120	
5	April	250	105	
6	May	220	115	
7	June	190	125	
8	July	170	130	
9	August	210	120	
10	September	240	115	
11	October	200	110	
12	November	180	125	
13	December	230	105	
14				

Fig 2.3.58 Creating New Column

3. Starting from cell D2, select all the cells where you want the Total Sales to appear, as shown below.

	A	B	C	D
1	**Month**	**Quantity Sold**	**Price per Unit (INR)**	**Total Sales**
2	January	150	100	
3	February	200	110	
4	March	180	120	
5	April	250	105	
6	May	220	115	
7	June	190	125	
8	July	170	130	
9	August	210	120	
10	September	240	115	
11	October	200	110	
12	November	180	125	
13	December	230	105	
14				

Fig 2.3.59 Selecting Cells

4. Now when you start typing the formula it will come in the first cell itself. Now let us enter the following formula into that cell. It is as shown below.

	A	B	C	D
1	**Month**	**Quantity Sold**	**Price per Unit (INR)**	**Total Sales**
2	January	150	100	=B2:B13*C2:C13
3	February	200	110	
4	March	180	120	
5	April	250	105	
6	May	220	115	
7	June	190	125	
8	July	170	130	
9	August	210	120	
10	September	240	115	
11	October	200	110	
12	November	180	125	
13	December	230	105	

Fig 2.3.60 Entering Formula

We started the formula with = sign then we selected the quantity sold column values and then we entered the * sign as we wanted to multiply the values. Then we selected the price per unit cell as total sales = quantity sold * price per unit.

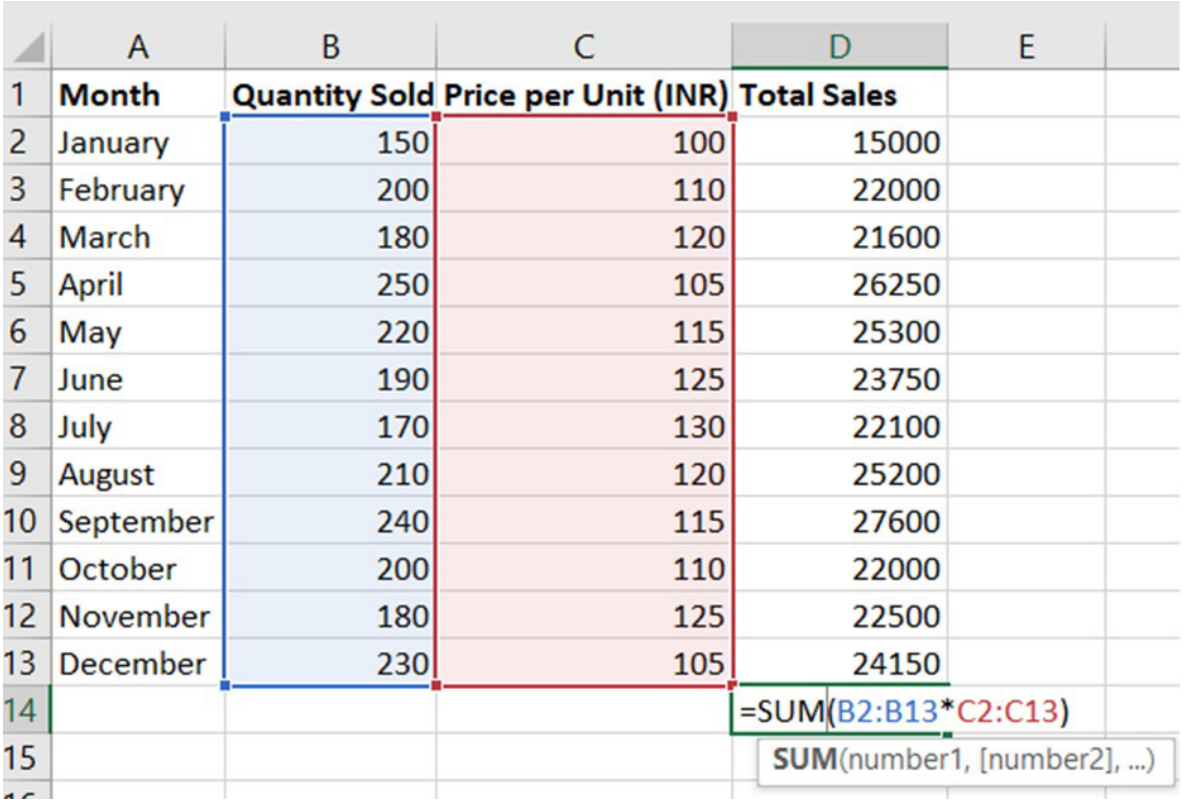
5. Now instead of pressing enter we have to press CTRL + SHIFT + ENTER simultaneously. Now the result will come as shown below.

	A	B	C	D
1	**Month**	**Quantity Sold**	**Price per Unit (INR)**	**Total Sales**
2	January	150	100	15000
3	February	200	110	22000
4	March	180	120	21600
5	April	250	105	26250
6	May	220	115	25300
7	June	190	125	23750
8	July	170	130	22100
9	August	210	120	25200
10	September	240	115	27600
11	October	200	110	22000
12	November	180	125	22500
13	December	230	105	24150
14				

Fig 2.3.61 Result of Array Formula

We can see that total sales for all the months are generated in one click. In traditional formulas we have to enter for one row and then auto fill it to rest of the cells. Here we can generate the result for the whole dataset in few clicks.

6. Let us find out the grant total of all the sales in a year in one cell from quantity sold and price per unit. It is as given below.



	A	B	C	D	E
1	**Month**	**Quantity Sold**	**Price per Unit (INR)**	**Total Sales**	
2	January	150	100	15000	
3	February	200	110	22000	
4	March	180	120	21600	
5	April	250	105	26250	
6	May	220	115	25300	
7	June	190	125	23750	
8	July	170	130	22100	
9	August	210	120	25200	
10	September	240	115	27600	
11	October	200	110	22000	
12	November	180	125	22500	
13	December	230	105	24150	
14				=SUM(B2:B13*C2:C13)	
15					

SUM(number1, [number2], ...)

Fig 2.3.62 Formula for Finding Total Sales Using Array Formula

Here we have used the SUM function and array formula. We have entered the =sum formula and select quantity sold column fist and then entered * sign and then selected the price per unit. Then we have to press CTRL + SHIFT + ENTER simultaneously. Then the excel will generate the result as shown below.

	A	B	C	D
1	**Month**	**Quantity Sold**	**Price per Unit (INR)**	**Total Sales**
2	January	150	100	15000
3	February	200	110	22000
4	March	180	120	21600
5	April	250	105	26250
6	May	220	115	25300
7	June	190	125	23750
8	July	170	130	22100
9	August	210	120	25200
10	September	240	115	27600
11	October	200	110	22000
12	November	180	125	22500
13	December	230	105	24150
14				277450
15				

Fig 2.3.63 Result of Total Sales Using Array Formula

The result will be as shown above. It is same as using SUMPRODUCT function. But we have used the basic SUM formula and array formula and generated the same result.

2. Dynamic Arrays:

Dynamic arrays were introduced in Excel 365 (version 2020) and represent a significant improvement in array formulas. Unlike traditional array formulas, dynamic arrays automatically "spill" their results across multiple cells without the need for pressing CTRL + SHIFT + ENTER. This behavior is known as "spilling."

Example of dynamic arrays:

For instance, suppose you have a list of numbers in cells A1:A5, and you want to calculate the square of each number and return the results in the adjacent cells. In traditional Excel versions, you would have to enter an array formula like {=A1:A5^2} by pressing CTRL + SHIFT + ENTER. But with dynamic arrays in Excel 365, you can enter the same formula simply as =A1:A5^2, and Excel will automatically spill the squared values into cells B1:B5.

- Automatically spill their results across multiple cells

Dynamic arrays make complex calculations more accessible and improve the readability and usability of Excel formulas, as you do not need to remember the array formula syntax or use special key combinations. It is essential to note that dynamic arrays are only available in the newer versions of Excel (Excel 365 or later). In older versions, you will still need to use traditional array formulas for the same operations.

Summarised Overview

Solver is a powerful add-in tool in Microsoft Excel that allows users to find optimal solutions to complex optimisation problems. It can be used for tasks such as production planning, resource allocation, financial decision-making, supply chain optimisation, and more. By defining an objective function, decision variables, and constraints, Solver efficiently searches for the best possible solution within the specified limitations. It is a valuable tool for businesses and individuals seeking data-driven decision-making, process optimisation, and cost reduction. Solver enables users to solve sophisticated optimisation problems without the need for specialised software, making it applicable to a wide range of real-world scenarios.

In Excel, advanced functions offer powerful capabilities for complex calculations and data manipulation. Some of these functions include VLOOKUP and HLOOKUP for vertical and horizontal lookups, INDEX to retrieve values from a specified range, MATCH to find the position of a value in a range, IFERROR to handle errors in formulas, CONCATENATE to join text strings, and more. Mastering these functions enables users to efficiently analyse data, perform conditional calculations, and solve real-world problems. Furthermore, logical functions like IF, AND, and OR allow users to make decisions based on conditions and evaluate logical statements. Additionally, Excel's array formulas and dynamic arrays simplify calculations on multiple cells or datasets, saving time and enhancing formula readability. Dynamic arrays, available in Excel 365 or later versions, automatically spill results across cells, eliminating the need for special key combinations.

Self-Assessment

1. Define optimisation problems in the context of decision-making and resource allocation
2. Describe the purpose of the Solver tool in Microsoft Excel. How does it help in finding optimal solutions for problems involving constraints and objective functions?
3. Compare and contrast the VLOOKUP and INDEX-MATCH functions in Excel.
4. Describe the role of logical functions (IF, AND, OR) in data analysis.
5. Define array formulas in Excel and explain their advantages over regular formulas

Assignments

1. Suppose you run a bakery that produces two types of products: cupcakes and cookies. Each cupcake requires 2 units of flour, 1 unit of sugar, and 3 units of butter to make, while each cookie requires 1 unit of flour, 2 units of sugar, and 1 unit of butter. You have 100 units of flour, 80 units of sugar, and 120 units of butter available for production. The selling price for each cupcake is Rs.2, and for each cookie is Rs.1. Maximise the total revenue (profit) from the sale of cupcakes and cookies. (Hint - Objective function: Maximise $Z = 2x + y$ (Total revenue from cupcakes and cookies)
2. Suppose you have a dataset containing information about employees in a company. The dataset is as given below

Employee ID	Employee Name	Department	Salary
101	John Doe	Sales	₹ 45,000
102	Jane Smith	Marketing	₹ 52,000
103	Robert Johnson	Finance	₹ 60,000
104	Emily Brown	HR	₹ 48,000
105	Michael Lee	Sales	₹ 50,000
106	Lisa Anderson	Finance	₹ 55,000
107	William Clark	HR	₹ 47,000
108	Sarah Walker	Marketing	₹ 53,000
109	James Wilson	Sales	₹ 49,000
110	Linda Garcia	HR	₹ 46,000

Answer the following questions:

a. What is the salary of the employee named "Michael Lee" in the Sales department?

b. Which department does Employee ID 107 belong to?

c. What is the job title of Employee Jane Smith?

3. Suppose we have information about students and their exam scores. As shown below

Student Name	Mathematics Score	Science Score	English Score
John Smith	85	90	78
Mary Johnson	92	88	85
David Lee	78	80	75
Sarah Brown	90	85	88
James Wilson	82	78	90

Answer the following questions:

a. What is the exam score of John Smith in Mathematics?
b. What is the exam score of Sarah Brown in Science.?
c. What is the exam score of James Wilson in English?

4. Suppose we have dataset representing a list of employees and their corresponding performance ratings in different categories as shown below.

Employee Name	Communication Skills	Technical Proficiency	Problem-Solving	Teamwork
David	4	5	4	3
Manya	5	4	3	5
Arjun	3	3	3	4
Basheer	4	4	5	5
Anjali	5	3	4	4

Answer the following questions:

a. Identify employees who have a Communication Skills rating of 5 and Technical Proficiency rating of 5.
b. Identify employees who have received "5" ratings in both Problem Solving and Creativity.
c. Identify employees who have a "4" rating or above in all the categories.

5. Given below is the sales data of a firm

Product	Sales (Week 1)	Sales (Week 2)	Sales (Week 3)	Sales (Week 4)
Product A	120	150	180	200
Product B	80	90	100	110
Product C	50	70	60	80
Product D	200	220	240	260
Product E	30	40	50	60

Answer the following questions:

a. Find the total sales for each product over the four weeks.
b. Calculate the total sales across all products for each week.

Reference

1. Nag, B., Hillier, F. S., Lieberman, G. J., & Basu, P. (2017). *Introduction to Operations Research.*
2. McFedries, P. (2019). *Microsoft Excel 2019 formulas and functions.* Microsoft Press.
3. Jelen, B., & Alexander, M. (2018). *Microsoft Excel 2019 Pivot Table Data Crunching*. Microsoft Press.
4. Jelen, B. (2009). *Excel Gurus Gone Wild: Do the Impossible with Microsoft Exce*l. Tickling Keys, Inc..
5. Winston, W. L. (2022). *Operations research: applications and algorithms*. Cengage Learning.

Suggested Reading

1. Albright, S. C., Winston, W. L., Zappe, C. J., & Broadie, M. N. (2011). *Data analysis and decision making* (Vol. 577). South-Western/Cengage Learning.
2. McFedries, P., & Harvey, G. (2021). *Excel All-in-one for Dummies*. John Wiley & Sons.
3. Alexander, M., & Kusleika, D. (2019). *Excel 2019 power programming with VBA*. John Wiley & Sons.
4. Ragsdale, C. (2021). *Spreadsheet modelling and decision analysis: a practical introduction to business analytics*. Cengage Learning.

Space for Learner Engagement for Objective Questions

Learners are encouraged to develop objective questions based on the content in the paragraph as a sign of their comprehension of the content. The Learners may reflect on the recap bullets and relate their understanding with the narrative in order to frame objective questions from the given text. The University expects that 1 - 2 questions are developed for each paragraph. The space given below can be used for listing the questions.

Appendix I

Keyboard Shortcuts

Cells & Ranges: Group & Hide

Group rows or columns (with rows/columns selected)	Shift+Alt+→
Ungroup rows or columns (with rows/columns selected)	Shift+Alt+←
Open Group window (with cell or range selected)	Shift+Alt+→
Open Ungroup window (with cell or range selected)	Shift+Alt+←
Hide or show outline symbols	Ctrl+8
Hide rows	Ctrl+9
Hide columns	Ctrl+0
Unhide rows	Ctrl+Shift+9
Unhide columns	Ctrl+Shift+0
Unhide columns (Alternate: Ctrl+Shift+0 does not work on some versions of Windows)	Alt,H,O,U,L

Cells & Ranges: Active Cell

Activate next cell in selected range (based on direction set in Excel Options)	Enter
Activate next cell to right/below in selected range	Tab
Activate each corner cell in selected range in clockwise direction	Ctrl+.
Activate previous cell in selection (based on direction set in Excel Options)	Shift+Enter
Activate previous cell to left/above in selected range	Shift+Tab
Scroll active cell into view	Ctrl+Backspace
Select single (active) cell when range is selected	Shift+Backspace

Cells & Ranges: Extend Selection

Extend selection by one cell up	Shift+↑
Extend selection by one cell right	Shift+→
Extend selection by one cell down	Shift+↓
Extend selection by one cell left	Shift+←
Extend the selection to the last cell up	Ctrl+Shift+↑
Extend the selection to the last cell right	Ctrl+Shift+→
Extend the selection to the last cell down	Ctrl+Shift+↓
Extend the selection to the last cell left	Ctrl+Shift+←
Extend selection to first cell in worksheet	Ctrl+Shift+Home
Extend selection to last cell in worksheet	Ctrl+Shift+End

Extend selection right one screen	Shift+Alt+PgDn
Extend selection left one screen	Shift+Alt+PgUp
Extend selection to start of row	Shift+Home
Extend selection down one screen	Shift+PgDn
Extend selection up one screen	Shift+PgUp
Toggle extend selection mode	F8
Select Spill Range	Ctrl+A

Cells & Ranges: Select Special

Open the Go to window	F5 OR Ctrl+G
Open the Go to Special window	Alt,H,F,D,S OR F5,Alt,S
Select visible cells only (when range is selected)	Alt+;
Select current region	Ctrl+A
Select current region around active cell	Ctrl+Shift+8
Select current array	Ctrl+/
Select row differences	Ctrl+\
Select column differences	Ctrl+Shift+\
Select direct dependents	Ctrl+]
Select direct precedents	Ctrl+[
Select all precedents	Ctrl+Shift+{
Select all dependents	Ctrl+Shift+}
Select cells with comments	Ctrl+Shift+O

Cells & Ranges: View

Zoom in	Ctrl+Alt++
Zoom out	Ctrl+Alt+-
Zoom in (with mouse)	Ctrl+Scroll Up
Zoom out (with mouse)	Ctrl+Scroll Down
Zoom to Selection	Alt,W,G
Zoom to Selection	Alt,W,J
Toggle Scroll Lock on/off	ScrLk
Toggle Gridlines on/off	Alt,W,V,G
Toggle Freeze Panes on/off	Alt,W,F,F

Cells & Ranges: Borders

Add Outside Borders to selected cell/range	Ctrl+Shift+&
Remove all borders (apply No Borders) to selection	Ctrl+Shift+_

Open Border dropdown menu on Home Tab	Alt,H,B
Add all borders to cells in selected range	Alt,H,B,A
Add or remove Bottom Border	Alt,H,B,O
Add or remove Top Border	Alt,H,B,P
Add or remove Left Border	Alt,H,B,L
Add or remove Right Border	Alt,H,B,R
Open Borders tab in Format Cells window (More Borders)	Alt,H,B,M

Worksheet Tabs

Insert new worksheet	Shift+F11
Delete selected worksheets	Alt,E,L
Rename current worksheet	Alt,O,H,R
Go to next worksheet	Ctrl+PgDn
Go to previous worksheet	Ctrl+PgUp
Select adjacent worksheets	Ctrl+Shift+PgUp/ PgDn
Select adjacent worksheets (mouse)	Shift+Click
Select non-adjacent worksheets (mouse)	Ctrl+Click
Open Move or Copy window for selected sheets	Alt,E,M
Duplicate worksheet (mouse)	Ctrl+drag
Protect or Unprotect Sheet	Alt,R,P,S
Clear print area	Alt,P,R,C
Set print area	Alt,P,R,S

Workbook: File

Create new workbook	Ctrl+N
Open workbook	Ctrl+O
Save workbook	Ctrl+S
Save as	F12
Close current workbook	Ctrl+F4 OR Ctrl+W
Close Excel	Alt+F4

Workbook: Print

Undo last action	Ctrl+Z
Redo last action	Ctrl+Y

Repeat last action	F4
Open Excel Options window	Alt,F,T
Open help	F1
Open the Control Menu (minimize, maximize,close window)	Alt+Space
Open Spelling window (Spell Check)	F7
Open Thesaurus dialog box	Shift+F7
Open Right-click menu	Shift+F10 OR
Move to next pane in application window	F6
Move to previous pane	Shift+F6
Protect and Unprotect Workbook	Alt,R,P,W

Workbook: View

Minimize current workbook window	Ctrl+F9
Maximize or restore current workbook window (toggle)	Ctrl+F10
Toggle full screen	Ctrl+Shift+F1
Go to previous workbook	Ctrl+Shift+Tab
Go to next workbook	Ctrl+Tab

Tables & Filters

Insert Table	Ctrl+T
Insert Table (select style first)	Alt,H,T
Insert a row above the current row	Alt,H, I, R
Insert a column to the left of the current column	Alt,H, I, C
Select Table Body (Press again to include Headers)	Ctrl+A
Select Table row	Shift+Space
Select Table column	Ctrl+Space
Toggle Autofilter on/off	Ctrl+Shift+L
Open Filter Dropdown Menu (with header cell selected)	Alt+↓
Open Filter Dropdown Menu (from any cell in a Table)	Shift+Alt+↓
Toggle Total Row on/off	Ctrl+Shift+T
Clear slicer filter (with slicer selected)	Alt+C

Charts & Shapes

Create chart or PivotChart on same worksheet	Alt+F1
Create chart in new ChartSheet	F11
Open Format Chart Area Task Pane	Ctrl+1
Show or Hide All Field Buttons on PivotChart	Alt,J,T,D,A

Snap to grid	Alt
Hide or show objects (shapes, charts, slicers)	Ctrl+6

Copy & Paste

Copy selected object (cells, range, shape)	Ctrl+C
Cut selected object	Ctrl+X
Paste content from clipboard	Ctrl+V
Open the Paste Special window	Ctrl+Alt+V
Duplicate selected object (shapes, charts,slicers)	Ctrl+D
Copy selected objects (mouse)	Ctrl+drag
Copy selected ojbects and retain vertical orhorizontal alignment (mouse)	Ctrl+Shift+drag
Copy selected objects (mouse right-click)	Right-click+Drag
Copy selected ojbects and retain vertical or horizontal alignment (right-click mouse)	Right-click+Shift+Drag

Pivot Tables

Insert Pivot Table	Alt,N,V
Refresh Pivot Table or Query	Alt+F5
Refresh All	Ctrl+Alt+F5
Open pivot table wizard	Alt,D,P
Select pivot table body	Ctrl+A
Group pivot table items	Shift+Alt+→
Ungroup pivot table items	Shift+Alt+←
Unhide (clear filter on) pivot table item	Alt,H,S,C
Filter out (exclude/hide) pivot item	Ctrl+-
Open Filter Dropdown Menu in Pivot Table cells	Alt+↓
Toggle pivot table field checkbox	Space
Insert pivot chart	Alt+F1
Create pivot chart on new Chart Sheet	F11

Ribbon

Activate Ribbon shortcut keys	Alt OR Shift+F6
Move to next ribbon control	Tab
Move through Ribbon tabs and groups	→ ← ↑ ↓
Open dropdown menu of selected button	Alt+↓
Open help window for selected control	F1

Expand or collapse ribbon	Ctrl+F1

Macros & VBA

Open the VB Editor window (VBA)	Alt+F11
Open the Macro window (run or edit macro)	Alt+F8
Record Macro	Alt,L,R
Open Excel Add-ins Window	Alt,L,R
Open COM Add-ins Window	Alt,L,J

Application: Dialog boxes

Move to next tab	Ctrl+Tab
Move to previous tab	Ctrl+Shift+Tab
Move to next control or button	Tab
Move to previous control or button	Shift+Tab
Press selected control or button	Enter
Check and uncheck checkboxes	Space
Close the window or dialog box	Esc

Power Query

Open the Power Query Editor	Alt,A,P,L,N
Press any button in the Power Query Ribbon	Alt, Letter Sequence
Quick Access Toolbar in Power Query	Alt+Number
Rename Column	F2
Remove Column(s)	Delete
Undo last action	Tab (6 times), Delete
Select All Columns	Ctrl+A
Select First Column	Home
Select Last Column	End
Select Multiple Adjacent Columns	Shift+→ or ←
Select Multiple Non-Adjacent Columns	Ctrl+→ or ←, Space
Open Filter Drop-down Menu	Alt+↓
Open Right-click Menus	Menu Key
Open Table Options Menu	Select First Column,←,Space
Select Task Panes & Ribbon	Tab & Shift+Tab
Close Power Query Window	Alt+F4
Close & Load	Alt,F,Enter

സർവ്വകലാശാലാഗീതം

വിദ്യയാൽ സ്വതന്ത്രരാകണം
വിശ്വപൗരരായി മാറണം
ഗ്രഹപ്രസാദമായ് വിളങ്ങണം
ഗുരുപ്രകാശമേ നയിക്കണേ

കൂരിരുട്ടിൽ നിന്നു ഞങ്ങളെ
സൂര്യവീഥിയിൽ തെളിക്കണം
സ്നേഹദീപ്തിയായ് വിളങ്ങണം
നീതിവൈജയന്തി പാറണം

ശാസ്ത്രവ്യാപ്തിയെന്നുമേകണം
ജാതിഭേദമാകെ മാറണം
ബോധരശ്മിയിൽ തിളങ്ങുവാൻ
ജ്ഞാനകേന്ദ്രമേ ജ്വലിക്കണേ

കുരീപ്പുഴ ശ്രീകുമാർ

SREENARAYANAGURU OPEN UNIVERSITY

Regional Centres

Kozhikode
Govt. Arts and Science College
Meenchantha, Kozhikode,
Kerala, Pin: 673002
Ph: 04952920228
email: rckdirector@sgou.ac.in

Thalassery
Govt. Brennen College
Dharmadam, Thalassery,
Kannur, Pin: 670106
Ph: 04902990494
email: rctdirector@sgou.ac.in

Tripunithura
Govt. College
Tripunithura, Ernakulam,
Kerala, Pin: 682301
Ph: 04842927436
email: rcedirector@sgou.ac.in

Pattambi
Sree Neelakanta Govt. Sanskrit College
Pattambi, Palakkad,
Kerala, Pin: 679303
Ph: 04662912009
email: rcpdirector@sgou.ac.in

Software Packages for Economic Analysis (Spreadsheet)

Course Code: M21EC01AC

SREENARAYANAGURU
OPEN UNIVERSITY

ISBN 978-81-967121-4-3

Sreenarayanaguru Open University
Kollam, Kerala Pin- 691601, email: info@sgou.ac.in, www.sgou.ac.in Ph: +91 474 2966841