

Excel

- **Introduction to Excel**
 - Advantages of Excel
 - Interface Overview
 - File Formats
 - Navigation and Shortcuts
- **Text Formatting**
 - Font Styles
 - Alignment
 - Other Formattings
 - Text to Columns
- **Data Manipulation**
 - Removing Duplicates
 - Data Filtering
 - Multi-Level Sorting
 - Data Validation
- **Basic Data Visualisation**
 - Charts and Graphs
- **Formula and Functions**
 - Categories of Function
 - Mathematical
 - Logical
 - Text
 - Date & Time
 - Lookup & Reference
 - Statistical
 - Parts of a Function
- **Referencing**
 - Absolute and Relative References
- **Advanced Analysis and Tools**
 - Pivot Tables
 - Power Pivot
 - What If ANalysis

Introduction to Excel

Excel is a powerful spreadsheet program developed by Microsoft. It's widely used across various industries and professions for data analysis, calculation, visualization, reporting, and more.

Other Spreadsheet Programs

1. Google Sheets
2. Apple's Numbers
3. LibreOffice Calc, etc (Open Source)

Advantages of Excel

1. **Versatility**
2. **User-Friendly Interface**
3. **Calculation Capabilities**
4. **Data Visualization**
5. **Integration with Other Tools:** Excel integrates well with other Microsoft Office applications (like Word, PowerPoint, and Outlook) and third-party tools like Power BI, Tableau, Python, enhancing its functionality.
6. **Data Storage and Organization:** Excel allows users to organize large amounts of data in spreadsheets, making it easy to sort, filter, and search information.
7. **Accessibility:** Excel files (.xlsx) are widely supported and can be easily shared and accessed across different platforms and devices.

Interface Overview

The Excel interface consists of the

1. Ribbon (containing tabs and commands)
2. Workbook (the main file where data is stored)
3. Worksheets (individual tabs within a workbook)
4. Cells (individual data entry points)
5. Formulas and Functions for calculations

File Formats

1. **.xlsx**: This is the default file extension for Excel workbooks created in Excel 2007 and later versions.
2. **.xls**: This was the default file format used by Excel 97-2003 versions. It supports basic Excel features but has limitations compared to .xlsx, such as smaller file size limits and fewer formatting options.
3. **.xlsb**: Binary Workbook format, which can result in smaller file sizes and faster opening and saving times, especially for large workbooks.
4. **.csv**: Comma-Separated Values file format. While not strictly an Excel format, .csv files can be opened and edited in Excel. They store data in plain text with each line representing a row of data and commas separating values within each row. Heavily used in Python by Data Scientists and Analysts.
5. **.txt**: Plain text file format. Similar to .csv, .txt files can be opened and edited in Excel. However, they lack structure (like rows and columns) inherent in Excel files.
6. **.xlt**: Excel Template format. It's used for saving a template that can be reused to create new workbooks with predefined settings, formatting, and formulas.

Navigation and Shortcuts

Selection Shortcuts	
Extend the selection by one cell	Shift + Arrow Keys
Extend the selection to the edge of data region	Ctrl + Shift + Arrow Keys
Select All	Ctrl + A
Editing Shortcuts	
Copy	Ctrl + C
Cut	Ctrl + X
Paste	Ctrl + V
Paste Special	Ctrl + Alt+ V
Undo	Ctrl + Z
Redo	Ctrl + Y
Edit current cell	F2

Navigation Shortcuts	
Go to the beginning of the row (A1)	Ctrl + (Fn)+ Home
Go to the end of the used worksheet area	Ctrl + (Fn) + End
Move to the edge of the current data region	Ctrl + Arrow Keys
Move up/down one page	PgUp/PgDn
Move to the next/previous worksheet	Ctrl + Pg Down / Ctrl + Pg Up
Some Useful Shortcuts	
Save	Ctrl + S
Open	Ctrl + O
Find	Ctrl + F
AutoSum	Alt + =

Text Formatting

1. Font Styles: Excel allows you to change various aspects of font styles, such as:
 - a. Font Type, Font Size, Bold, Italic, Underline
 - b. Strikethrough, Superscript, Subscript
2. Alignment
 - a. Horizontal Alignment, Vertical Alignment, Text Orientation
3. Cell Fill and Border
4. Number Formatting
5. Conditional Formatting
6. Format Painter
7. Wrap Text
8. **Text to Columns** : "Text to Columns" is a feature in Excel that allows you to split the content of a single cell or a range of cells into multiple columns based on a delimiter, such as a comma, space, or tab.
Example splitting a name or a csv.

Data Manipulation



- Removing Duplicates
- Data Filtering (Ctrl + Shift + L)
- Multi-Level Sorting
- Data Validation

Basic Data Visualisation



Data visualization in Excel is a powerful feature that allows users to present data graphically for easier interpretation and analysis. Here are some common techniques and tools for data visualization in Excel:

Charts and Graphs:

Excel offers a variety of built-in chart types that can be easily created from your data:

1. **Column Chart:** Shows data changes over a period of time or compares items.
2. **Bar Chart:** Similar to column charts but with horizontal bars.
3. **Line Chart:** Displays trends over time or data series.
4. **Pie Chart:** Represents parts of a whole, useful for showing percentages.
5. **Scatter Plot:** Shows relationships between two sets of data points.
6. **Area Chart:** Similar to line charts but with shaded areas below lines.

Functions and its Categories

Functions in Excel are predefined formulas that perform calculations or manipulate data in various ways. They are designed to simplify complex calculations and make it easier to analyze large amounts of data.

Categories: Functions in Excel are categorized based on their purpose, such as:

- **Mathematical** : Perform mathematical calculations like SUM, AVERAGE, SQRT, etc.
- **Logical**: Evaluate conditions and return true or false values like IF, AND, OR, etc.
- **Text**: Manipulate text strings like CONCATENATE, LEFT, RIGHT, etc.
- **Date & Time**: Perform operations on dates and times like DATE, TODAY, MONTH, etc.
- **Lookup & Reference**: Look up values in tables or ranges like HLOOKUP, VLOOKUP, INDEX, MATCH, etc.
- **Statistical**: Calculate statistical measures like COUNT, MEDIAN, STDEV, etc.

[Formula Sheet](#)

Parts of a Function

Syntax: Functions in Excel follow a specific syntax, typically starting with an equal sign (=) followed by the function name and its arguments enclosed within parentheses.

Arguments: Functions can have one or more arguments (input values) that define how the function operates. These arguments can be constants, cell references, or other functions.

Nested Functions: Functions can be nested within one another to perform more complex calculations. For example, using a function result as an argument for another function.

AutoComplete and Help: Excel provides AutoComplete functionality when typing function names and offers help documentation to assist in understanding the function syntax and usage.

Precedents and Dependents : This help to track the cells used in formula.

[Formula Sheet](#)

Errors in Excel

These common errors in Excel indicate different issues:

#VALUE!: This error occurs when the wrong type of argument or operand is used in a formula, such as trying to perform a mathematical operation on text.

#REF!: This error appears when a formula refers to a cell that is not valid, often because the cell has been deleted or moved.

#NAME?: This error indicates that Excel does not recognize text in a formula, which can happen if a named range is misspelled or a function name is incorrect.

#N/A: This error means "not available." It often occurs with lookup functions (like VLOOKUP or HLOOKUP) when the requested value cannot be found.

Circular dependency : This error in Excel occurs when a formula refers back to its own cell, either directly or indirectly, creating a loop that Excel cannot resolve.

Referencing

Relative Reference:

- A relative reference in a formula is one that adjusts relative to the position where the formula is copied or filled.
- Example: If you have a formula in cell B2 that references cell A1 (`=A1`), when you copy the formula to cell B3, it will automatically adjust to `=A2` (one row down relative to the original reference).

Absolute Reference:

- An absolute reference in a formula refers to a specific cell or range and does not change when the formula is copied or filled to other cells.
- Example: If you want to keep referring to cell A1 in all instances of a copied formula, you would use an absolute reference like `A1`. **Shortcut : F4**

Sheet Reference:

Sheet2!A1

'Sheet 2'!A1

Sheet2!A1:B10

Titanic Dataset Overview



Data Dictionary

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

Practice



IF:

Categorize the passengers in the Titanic dataset into different age groups, such as "Children" and "Adults"?

i. Children ≤ 18

ii. Adult > 18

Nested IF:

- a. Can you categorize the passengers in the Titanic dataset into different age groups, such as "Children," "Adults," and "Elderly"?

i. Children ≤ 18

ii. Adult $> 18 < 60$

iii. Elderly > 60

- a. Categorize fare in four buckets using nested if

i. less than \$50, ii. between \$50 and \$100, iii. between \$100 and \$200, and iv. greater than \$200:

Practice



SUMIF: What is the total fare paid by passengers who survived?

COUNTIF: How many passengers were in the 1st class of the Titanic?

AVERAGEIF: What is the average age of passengers in the 1st class?

Excel Questions <https://excel-practice-online.com/>

<https://www.excelformulabeautifier.com/>

Practice



PIVOT TABLE

1. How many passengers were there for each Pclass ?
2. What was the average survival rate for each Pclass ?
3. What was the average survival rate for Pclass and gender combination?
4. What was the survival rate for different age groups?
5. Total Fare Collected by Embarkation Points
6. Survival Rate by Family Size
7. What was the average fare rate for different age groups for survived people?
8. Age Distribution by Passenger Class: Create a histogram chart to visualize the age distribution of passengers in each passenger class. Calculate the mean and standard deviation of ages in each class.

Disadvantages of Excel

1. **Limited Data Handling:** Excel has limitations in handling large datasets or complex databases compared to dedicated database management systems (DBMS) like SQL Server or Oracle.
2. **Data Integrity:** Human error, such as accidental overwriting of formulas or incorrect data entry, can affect data integrity and lead to errors in analysis.
3. **Version Control:** Managing versions of Excel files and ensuring all collaborators work on the latest version can be challenging, especially in shared environments.
4. **Security Risk**
5. **Scalability**
6. **Learning Curve for Advanced Features:** Mastering advanced features like complex formulas, array functions, and VBA programming may require a steep learning curve for some users.
7. **Lack of Centralization:** Unlike centralized systems, Excel files are typically stored locally or on shared drives, which can lead to data duplication or inconsistent data sources.

Predictive Analytics Titanic Dataset



Titanic Submission On Kaggle

<https://www.kaggle.com/competitions/titanic/data>

Note : We can say both sex and Pclass are playing key roles in determination of survival.