

COURSE 2: EXCEL BASICS FOR DATA ANALYSIS

MODULE 1: INTRODUCTION TO DATA ANALYSIS USING SPREADSHEETS

1.1 COURSE INTRODUCTION

This course by **IBM** is designed to:

- Introduce **Excel as a data analysis tool**
- Build a strong foundation in:
 - Data cleaning
 - Data wrangling
 - Data analysis using spreadsheets
- Provide **hands-on labs** to practice real-world scenarios

What You Will Learn Across Modules

- **Module 1:** Spreadsheet basics, terminology, interface, worksheets & workbooks
- **Module 2:** Selecting, entering, editing data, formatting, formulas & functions
- **Module 3:** Data cleaning & wrangling (duplicates, missing data, inconsistencies)
- **Module 4:** Data analysis (filtering, sorting, pivot tables, slicers)
- **Module 5:** End-to-end hands-on project (business scenario → analysis → insights)

Course Outcomes

After completing the course, you will be able to:

- Use spreadsheets for data analysis
 - Understand **when** to use spreadsheets and their **limitations**
 - Perform data cleaning and wrangling in Excel
 - Analyze data using **filters, sorting, and pivot tables**
 - Create a **data analyst deliverable**
-

1.2 INTRODUCTION TO SPREADSHEETS

1. Common Spreadsheet Applications

Popular Spreadsheet Tools

- **Microsoft Excel**
 - Most widely used and feature-rich
 - Available as:
 - Desktop (paid – Office / Microsoft 365)
 - Web-based (Excel Online – free, limited features)
- **Google Sheets**
 - Free with Google account
 - Web-based
 - Integrates with Google Forms, Analytics, Data Studio
- **LibreOffice Calc**
 - Free and open-source
 - Desktop-based
 - Supports charts, conditional formatting, pivot tables

Other Spreadsheet Applications

- Zoho Sheet
- OpenOffice Calc
- Quip (Salesforce)
- Smartsheet (project management focused)
- Apple Numbers

📌 Choice depends on:

- Features required
- Budget
- Cloud vs desktop preference

2. Why Spreadsheets Are Useful for Data Analysts

Advantages

- Automatic and accurate calculations
- Easy data organization
- Quick editing & error correction
- Built-in filtering and sorting
- Data visualization (charts, graphs, reports)

Evolution of Spreadsheets

- First spreadsheet: **VisiCalc (1970s)** on Apple II
- Progressed from simple tables → advanced data analysis tools

3. Common Uses of Spreadsheets

Business Uses

- Data entry & storage
- Comparing large datasets
- Modeling & planning
- Trend identification
- Financial forecasting
- Budgeting & accounting
- Statistical analysis
- Payroll, invoicing, scheduling
- Fraud & forensic auditing

Personal Uses

- Personal expenses & budgeting
- Fitness & calorie tracking
- Recipe libraries
- Sports leagues (fantasy football)
- Contact & shopping lists

4. How Data Analysts Use Spreadsheets

- Data collection from multiple sources
 - Data cleaning (duplicates, errors, missing values)
 - Data analysis (filtering, sorting, interpretation)
 - Data visualization for storytelling & stakeholder communication
-

1.3 SPREADSHEET BASICS – PART 1 (TERMINOLOGY)

1. Workbook

- Highest-level Excel file
- File extension: **.XLSX**
- Contains all data, formulas, and calculations

2. Worksheet

- Individual tabs inside a workbook
- Default names: Sheet1, Sheet2, ...
- Should be renamed to reflect purpose (e.g., *January Sales*)
- Active worksheet = currently selected tab

3. Cells

- Rectangular boxes containing data
- Can store:
 - Text
 - Numbers
 - Formulas
 - Calculated results

Rows & Columns

- **Columns:** Vertical (A, B, C...)
- **Rows:** Horizontal (1, 2, 3...)

Cell Reference

- Combination of column + row
- Example: **M20**
- Always written as **Column first, Row second**

Active Cell

- Currently selected cell
- Highlighted border
- Shown in the Name Box (top-left)

4. Cell Range

- A group of selected cells
- Written using a colon (:)
 - D9:D19 → same column
 - D9:H9 → same row
 - D9:H19 → rows + columns

📌 Used extensively in:

- Formulas
- Calculations
- Data analysis

3D References

- Cell ranges across multiple worksheets

1.4 SPREADSHEET BASICS – PART 2 (NAVIGATION & SELECTION)

1. Opening Files

- File → Open
- Access via:
 - Recent files

- Browse
- Excel Options available via **Backstage View**

2. Ribbon & Menus

- Ribbon tabs:
 - Home, Insert, View
 - Formulas, Data, Power Pivot
- Ribbon can be hidden/unhidden:
 - Double-click tab
 - Shortcut: **CTRL + F1**

Ribbon Groups

- Organized tool sections (Font, Alignment, Number, Styles)
- Extra options accessed via small arrow (dialog launcher)

3. Quick Access Toolbar

- Frequently used tools:
 - Save, Undo, Redo, New, Open
- Customizable (e.g., Sort Ascending/Descending)

4. Navigating a Worksheet

Keyboard Navigation

- Arrow keys → move one cell
- Page Up / Page Down → faster movement
- Scroll bars → navigate large datasets

Useful Shortcuts

- **CTRL + Home** → go to cell A1
- **CTRL + End** → last used cell
- **CTRL + ↓ / ↑** → jump to end/top of column

5. Selecting Data

Selection Types

- Single cell → mouse or arrow key
- Multiple cells → drag or SHIFT + arrows
- Entire column → click column letter
- Entire row → click row number
- Multiple columns/rows → drag or SHIFT + arrows
- Non-contiguous selections → CTRL + click
- Entire worksheet → top-left corner
- Data-only selection → **CTRL + A**

6. Cursor Symbols (Important!)

- **Large white cross:** Select cells
- **Four-arrow cross:** Move data
- **Small black cross:** Fill Handle (copy/fill data)

⚠ Misusing these can unintentionally move or overwrite data

HANDS-ON LAB: GETTING STARTED WITH EXCEL ONLINE

HANDS-ON LAB 2: SPREADSHEET BASICS

Task	Shortcut
Close a workbook	Ctrl+W
Open a workbook	Ctrl+O
Save a workbook	Ctrl+S
Copy	Ctrl+C
Cut	Ctrl+X
Paste	Ctrl+V
Undo	Ctrl+Z
Remove cell contents	Delete
Bold	Ctrl+B
Open context menu	Shift+F10
Expand or collapse the ribbon	Ctrl+F1
Move up one cell in the worksheet	Up arrow key
Move down one cell in the worksheet	Down arrow key
Move one cell left in the worksheet	Left arrow key
Move one cell right in the worksheet	Right arrow key
Move to the edge of the current data region in the worksheet (e.g. end of column)	Ctrl+Arrow key (e.g. Ctrl+Down arrow)
Move to the last cell on a worksheet	Ctrl+End
Move to the beginning of a worksheet	Ctrl+Home
Extend the selection of cells to the last used cell on a worksheet (lower right corner)	Ctrl+Shift+End
Move to the cell in the upper-left corner of the window (when Scroll Lock is On)	Home+Scroll Lock
Move one screen down in a worksheet	Page Down
Move one screen up in a worksheet	Page Up
Move one screen to the right in a worksheet	Alt+Page Down
Move one screen to the left in a worksheet	Alt+Page Up
Move to the next sheet in a workbook	Ctrl+Page Down
Move to the previous sheet in a workbook	Ctrl+Page Up
Edit the active cell and put the cursor at the end of the cell's contents	F2
Enter the current time	Ctrl+Shift+colon (:)
Enter the current date	Ctrl+semi-colon (;)

MODULE 2: GETTING STARTED WITH USING EXCEL SPREADSHEETS

2.1 VIEWING, ENTERING, AND EDITING DATA

A. Viewing Data in Excel

Zoom Options

- **Zoom Slider** (bottom-right corner)
 - Click  /  or drag slider
- **View Tab → Zoom**
 - Predefined or custom zoom
 - **100%** → original size
 - **Zoom to Selection** → zooms only selected area

Split View

- **Split button**
 - Divides worksheet into multiple scrollable sections
- Remove split:
 - Double-click horizontal or vertical split line

Freeze Panes

Used to keep headers visible while scrolling

Options:

- Freeze Top Row
- Freeze Columns
- Freeze Rows & Columns together

Rule:

- Select the cell **one row below** and **one column right** of what you want to freeze
- Then click **Freeze Panes**

Switching Between Workbooks

- View → Switch Windows
- Shortcut: **CTRL + F6**

B. Entering Data

Header Row

- Column headings are called **header row**
- Use **TAB** to move right while entering headers
- Use **ENTER** to move down rows

Adjusting Column Width

- Manual: Drag column divider
- Automatic: Double-click column divider
- Auto-fit multiple columns:
 - Select columns → double-click divider

Inserting Columns

- Select column → Right-click → Insert
- New column appears to the **left**

C. Editing Data in Cells

Ways to edit cell content:

1. Select cell → type over it
 2. Press **F2**
 3. Double-click cell
 4. Edit in **Formula Bar**
-

2.2 COPYING, FILLING, AND FORMATTING DATA

A. Moving & Copying Data

Move Data

- Select cells → hover edge → drag

Copy Data

- Drag while holding **CTRL**
- Or use:
 - **CTRL + C** (Copy)
 - **CTRL + V** (Paste)

Copy Between Worksheets

- Default paste changes column widths
- To retain widths:
 - Paste Options → **Keep Source Column Widths**

B. AutoFill Feature

Used to fill **patterns or sequences**

Examples:

- Months: Jan → drag → fills Feb, Mar...
- Days: Mon → drag → Tue, Wed...
- Numeric pattern:
 - Single value (5) → copies same value
 - Two values (5,10) → detects pattern (+5)

 Always select enough cells to define the pattern

C. Formatting Cells vs Data

1 Cell Formatting

- Fill color

- Borders
- Bold / Italic
- Font size & style
- Cell styles (Home → Styles)

2 Data Formatting

- Text
- Number
- Decimal places
- Currency
- Accounting

📌 Example issue:

- Imported CSV misreads **9-5** as a date
- Solution: Format cell as **Text**, then re-enter value

Currency Formatting

Steps:

1. Select column
 2. Number Format → More Number Formats
 3. Choose Currency
 4. Select symbol (e.g., \$)
-

HANDS-ON LAB 3: ENTERING AND FORMATTING DATA



2.3 BASICS OF FORMULAS

A. Formula Structure

A formula always starts with =

Components:

- **Function** → performs calculation (e.g., SUM)
- **References** → cells or ranges
- **Operators** → + - * /
- **Constants** → fixed values (5, 10%, date)

Example:

=SUM(B5*20)

B. Selecting Ranges in Formulas

Ways:

- Manual typing: E2,E3,E4
- Better way: **E2:E13**
- Best way: select range using mouse or **SHIFT + arrows**

📌 Excel auto-adds closing parenthesis

C. Copying Formulas (AutoFill)

- Drag fill handle
 - Formula adjusts cell references automatically
 - This uses **relative references**
- 📌 Double-click fill handle to copy formula down long columns

D. AutoSum

Located in **Home → Editing**

Functions:

- SUM
- AVERAGE
- COUNT

- MAX
- MIN

Shortcut:

ALT + =

2.4 INTRODUCTION TO FUNCTIONS

Common Statistical Functions

- **AVERAGE** → mean
 - **MIN** → smallest value
 - **MAX** → largest value
 - **COUNT** → number of values
 - **MEDIAN** → middle value
- 📌 Even count → average of two middle values
📌 Odd count → exact middle value

Function Categories (Formulas Tab)

- Financial
- Logical → IF, AND, OR
- Text → CONCAT, FIND, SEARCH
- Date & Time → WEEKDAY, NETWORKDAYS
- Lookup & Reference → VLOOKUP, HLOOKUP
- Math & Trig → SUMIF, POWER
- Statistical
- Engineering & Information

Insert Function Tool

- Search by name
- Browse by category
- Shows description & syntax

🔗 2.5 REFERENCING DATA IN FORMULAS

A. Relative References (Default)

- Example: =A1+A3
- Change when copied
- Adjust automatically based on position

B. Absolute References

- Uses \$
- Example: =\$A\$1+\$A\$3
- Does **not change** when copied

C. Mixed References

- Part fixed, part relative

Examples:

- A\$1 → row fixed
- \$A3 → column fixed

📌 Used in tables, tax rates, constants

D. Reference Summary Table

Type	Example	Changes on Copy
Relative	A1	Yes
Absolute	\$A\$1	No
Mixed	A\$1 / \$A1	Partial

HANDS-ON LAB 4: SIMPLE USE OF FUNCTIONS

⚠ 2.6 FORMULA ERRORS IN EXCEL

Common Error Indicators

- #####
 - Column too narrow
 - Negative date/time
 - Not a real formula error

#NAME? Error

Cause:

- Wrong operator (X instead of *)
- Misspelled function
- Invalid reference

Error Handling Tools

- Green triangle indicator
- Error icon (!)
- Options:
 - Help on error
 - Show calculation steps
 - Edit in Formula Bar
 - Ignore error
 - Error checking options

MODULE 3: CLEANING & WRANGLING DATA USING SPREADSHEETS

3.1 INTRODUCTION TO DATA QUALITY

◆ Why Data Quality Matters

- Data analysis supports **business decisions**
 - Poor-quality data leads to **incorrect insights & risky decisions**
 - Cleaning and qualifying data is a **core responsibility of a Data Analyst**
-

3.2 FIVE TRAITS OF HIGH-QUALITY DATA

1 Accuracy

- Data must be **correct and error-free**
- Common cleaning tasks:
 - Remove duplicates
 - Fix formatting errors
 - Remove blank rows

📌 *Most critical trait of data quality*

2 Completeness

- All required information must be present
- Example:
 - Task: Calculate revenue per region
 - Issue: Region data missing → dataset is incomplete

📌 Incomplete data may require **additional data sources**

3 Reliability

- Data should come from **trustworthy and consistent sources**
- Example:
 - Agents maintain personal records but don't update company DB

- Shared database becomes **unreliable**
 - 📌 May require **process changes** to ensure accuracy
- 4 Relevance**
- Data should be **useful for the specific project**
 - Example:
 - Sales analysis includes customer birthdays → unnecessary

- 📌 Removing irrelevant data:
- Saves time
 - Improves focus
 - Reduces privacy risks

- 5 Timeliness**
- Data must be **up-to-date and accessible**
 - Example:
 - Weekly employee reviews using monthly-updated data
- 📌 Outdated data can cause **serious business consequences**
-

3.4 IMPORTING FILE DATA INTO EXCEL

- ◆ **Supported File Types**
 - .xlsx, .xls (default Excel)
 - .txt (plain text)
 - .csv (Comma-Separated Values)
- ◆ **Text Import Wizard (CSV / TXT Files)**

Step 1: Identify File Type

- Choose **Delimited**
- Enable “**My data has headers**”

Step 2: Select Delimiter

- Common delimiters:
 - Comma
 - Tab
- Preview data layout before proceeding

Step 3: Set Data Formats

- Choose column data types:

- General
- Text
- Date

- Finish import

◆ Adjusting Columns & Rows

- **Auto-fit columns:** Select all → Double-click divider
- **Resize rows:** Drag or double-click row divider

◆ Editing Structure

- **Delete columns:** Right-click → Delete
- **Insert columns:** Right-click → Insert
- **Delete rows:** Right-click → Delete
- **Insert rows:** Right-click → Insert

◆ Saving Imported Files

- Save as **Excel Workbook (.xlsx)**
- Use **Save As** or yellow import notification

3.5 BASICS OF DATA PRIVACY

◆ Why Data Privacy Is Important

- Prevents **legal penalties**

- Maintains **customer trust**
 - Ensures **ethical data handling**
-

3.6 THREE FUNDAMENTALS OF DATA PRIVACY

1 Confidentiality

- Personal data belongs to the individual
- Analysts must protect:
 - Employee records
 - Customer data
 - Medical information

2 Collection and Use

- Collect **only necessary data**
- Use data **only for its intended purpose**

3 Compliance

- Follow **regional & industry regulations**
-

3.7 TYPES OF PERSONAL DATA

Personal Information (PI)

- Any data linked to a person
- Examples:
 - Emails
 - Photos

Personally Identifiable Information (PII)

- Can directly identify an individual
- Examples:
 - Social Security Number

- Driver's License Number

Sensitive Personal Information (SPI)

- Highly private data
- Examples:
 - Race
 - Sexual orientation
 - Biometric & genetic data

📌 Must be **strictly protected**

3.8 MAJOR DATA PRIVACY REGULATIONS

- **General Data Protection Regulation (GDPR)** – European Union
- **Lei Geral de Proteção de Dados (LGPD)** – Brazil
- **California Consumer Privacy Act (CCPA)** – USA (California)
- **Health Insurance Portability and Accountability Act (HIPAA)** – Healthcare
- **Payment Card Industry Data Security Standard (PCI DSS)** – Retail & credit card data

3.9 DATA BREACHES & RESPONSIBILITY

- Companies are responsible for **data security**
- Example:
 - Analyst takes laptop home
 - Laptop stolen → **data breach**

📌 Consequences:

- Heavy fines
- Loss of consumer trust
- Revenue impact

3.10 ANONYMOUS DATA EXCEPTION

- Privacy laws **do not apply** if data is:
 - Fully anonymized
 - Cannot be traced back to individuals
-  Not always practical, but reduces restrictions
-

3.11 WHY DATA CLEANING IS NECESSARY

When collecting or importing data (manual or automated):

- Errors and inconsistencies are common
- Examples:
 - Spelling mistakes
 - Extra white spaces
 - Incorrect text case
 - Empty rows
 - Missing values
 - Duplicate records

! Problems Caused by Dirty Data

- Formulas may not work correctly
- Sorting and filtering fail
- Incorrect analysis and visualizations
- Poor business decisions

 **Data cleaning improves data quality, accuracy, and usability**

3.12 SPELL CHECKING INACCURATE DATA

◆ Purpose

- Identify and fix spelling errors in text data

◆ How Spell Check Works in Excel

1. Select the column to check
2. Go to **Review tab → Spelling**
3. Excel suggests corrections
4. Options:
 - **Change** (accept suggestion)
 - **Ignore** (if data is correct)
 - Choose another suggestion

◆ Example Columns Checked

- Product line
- Country names
- Deal size (e.g., *small, medium*)

📌 Similar to spell check in Microsoft Word

3.13 REMOVING EMPTY ROWS

◆ Why Empty Rows Are a Problem

- Break continuous datasets
- Cause issues with:
 - CTRL + ↓ navigation
 - Formulas
 - Sorting & filtering

◆ Method 1: Manual Deletion (Not Recommended for Large Data)

- Scroll and delete empty rows one by one

- Suitable only for **very small datasets**
- ◆ **Method 2: Filter-Based Deletion (Recommended)**

Steps

1. Select entire dataset
 - Mouse OR CTRL + SHIFT + END
2. Go to **Data tab → Filter**
3. Choose a column (e.g., Customer Name)
4. Uncheck **Select All**
5. Check **Blanks**
6. Click **OK**
7. Empty rows appear at top
8. Select those rows
9. Delete rows
10. Clear and turn off filter

- Dataset becomes continuous
 - CTRL + ↓ now works correctly
-

3.15 REMOVING DUPLICATE DATA

- ◆ **Why Duplicates Occur**
 - Human data entry errors
 - Import process errors

Method 1: Identify Duplicates Before Deleting (Recommended)

- ◆ **Why This Method Is Safer**
 - Allows review before deletion
 - Reduces risk of removing valid data
- ◆ **Steps**

1. Select a column that **should NOT have duplicates**

- Price Each (bad choice)
 - Sales column (good choice)
2. Go to:
 - **Home → Conditional Formatting**
 - **Highlight Cells Rules → Duplicate Values**
 3. Excel highlights duplicate values
 4. Review rows visually
 5. Delete confirmed duplicate rows manually

 Best practice for **data security and accuracy**

Method 2: Remove Duplicates Automatically (Less Secure)

◆ **Steps**

1. Select entire dataset
2. Go to **Data tab → Remove Duplicates**
3. Unselect all columns
4. Select only the **Sales** column
5. Click **OK**

 Duplicates are deleted **without preview**

3.16 FIXING INACCURATE TEXT USING FIND & REPLACE

◆ **Purpose**

- Correct repeated text errors efficiently

◆ **Steps**

1. Go to **Home → Find & Select**
2. Choose **Find or Replace**
3. Enter incorrect text in **Find what**
4. Use:

- **Find Next**
 - **Find All**
5. Enter corrected text in **Replace with**
 6. Click **Replace All**

◆ **Example**

- Incorrect surname: *Larson*
- Correct surname: *Larsson*

❖ Ensures consistency across the dataset

3.17 DEALING WITH INCONSISTENCIES IN DATA

When data comes from multiple sources, it often contains:

- Mixed text cases
- Inconsistent date formats
- Extra spaces (whitespace)

Excel provides **functions and tools** to fix these issues.

3.18 CHANGING TEXT CASE IN EXCEL

Excel does not have a direct *Change Case* button (like Word), so we use **functions**.

◆ **Text Case Functions**

Function	Purpose	Example
UPPER()	Converts text to uppercase	india → INDIA
LOWER()	Converts text to lowercase	India → india
PROPER()	Capitalizes first letter of each word	india sales → India Sales

◆ Using PROPER Function (Uppercase → Proper Case)

Use Case: Cleaning column headers or names

Steps:

1. Insert a **helper row** below the header.
2. In A2, type:
3. =PROPER(A1)
4. Fill across the row (drag OR SHIFT + RIGHT ARROW → F2 → CTRL + ENTER)
5. Copy helper row
6. Paste into original row using **Paste Values**
7. Delete helper row

⚠ Important:

Do **not** delete the original row before pasting values → it causes #REF! error.

◆ Using UPPER Function (Proper Case → Uppercase)

Steps:

1. Insert a **helper column**
2. In first data cell:
3. =UPPER(T2)
4. Double-click Fill Handle
5. Copy helper column
6. Paste into original column → **Paste Values**
7. Delete helper column

◆ Using LOWER Function (Proper Case → Lowercase)

Steps:

1. Insert helper column
2. Formula:
3. =LOWER(K2)
4. Autofill

5. Paste values back
 6. Delete helper column
-

3.19 FIXING DATE FORMATTING ERRORS

Dates may:

- Use wrong locale (UK vs US)
 - Appear inconsistent
- ◆ **Changing Date Locale**

Steps:

1. Select date column
2. Open **Format Cells**
3. Choose **Date**
4. Change **Locale**:
 - English (United States)
 - English (United Kingdom)

◆ **Custom Date Format**

Example format:

dd-mmm-yyyy

Steps:

1. Format Cells → Custom
 2. Modify an existing format
 3. Apply to column using:
 - Format Painter OR
 - Select column → choose format
-

3.20 TRIMMING WHITESPACE FROM DATA

Whitespace issues include:

- Leading spaces
- Trailing spaces
- Double spaces between words

◆ **Using Find & Replace (Double Spaces)**

Steps:

1. Select all data
2. Home → Find & Select → Replace
3. Find:
4.
5. Replace with:
6.
7. Use **Replace or Replace All**

⚠ Do NOT replace single spaces globally.

◆ **Using TRIM Function (Leading & Trailing Spaces)**

Purpose: Removes extra spaces except single spaces between words

Steps:

1. Insert helper column
 2. Formula:
 3. `=TRIM(M2)`
 4. Autofill
 5. Copy → Paste Values to original column
 6. Delete helper column
-

3.21 FLASH FILL – SMART DATA CLEANING TOOL

Flash Fill detects patterns and auto-fills data.

- ◆ **Use Cases of Flash Fill**

- ✓ Combine columns
- ✓ Split names
- ✓ Change name format

- ◆ **Flash Fill – Excel for the Web**

Steps:

1. Type desired output in first cell
2. Go to **Data → Flash Fill**
3. Excel fills remaining rows automatically

- ◆ **Flash Fill – Desktop Excel**

Steps:

1. Type first value
2. Start typing second value
3. Preview appears
4. Press **Enter**

3.22 TEXT TO COLUMNS FEATURE

Purpose: Split one column into multiple columns

Example:

John Smith → John | Smith

- ◆ **Using Text to Columns (Desktop Only)**

Steps:

1. Select column
2. Data → Text to Columns
3. Choose **Delimited**

4. Select delimiter (Space)
5. Set destination cell
6. Finish

 **Excel for Web does NOT support Text to Columns**

3.23 SPLITTING NAMES USING FUNCTIONS (EXCEL FOR WEB)

◆ **Extract First Name**

=LEFT(C2, SEARCH(" ", C2, 1))

◆ **Extract Last Name**

=RIGHT(C2, LEN(C2) - SEARCH(" ", C2, 1))

Functions Used:

- LEFT()
- RIGHT()
- LEN()
- SEARCH()

3.24 HANDS-ON LAB 5 – SUMMARY

◆ **Skills Practiced**

- Removing duplicates
- Removing empty rows
- Fixing spelling
- Case conversion
- Date formatting
- Trimming whitespace
- Flash Fill
- Text cleaning using functions

◆ **Dataset**

- IBM fictitious customer & sales data
 - Used in **Excel for the Web**
-

HANDS-ON LAB 5: CLEANING DATA

MODULE 4: ANALYZING DATA USING SPREADSHEETS

4.1 INTRODUCTION TO DATA ANALYSIS USING SPREADSHEETS

Once data is **collected and cleaned**, the next step is **analysis** to extract meaningful insights using tools available in **Microsoft Excel**.

◆ Key Questions Before Analysis

Before manipulating data, always visualize the final output and ask:

- How large is the dataset?
- What filters are required?
- How should the data be sorted?
- What calculations are needed?

◆ Core Analysis Techniques

- **Filtering** → Display only required records
 - **Sorting** → Organize data logically (A–Z, smallest–largest, oldest–newest)
 - **Functions** → Perform fast, error-free calculations
 - **Shaping Data** → Tables & Pivot Tables
-

4.2 FILTERING AND SORTING DATA

◆ Filtering Data

Filtering shows **only rows that meet specific criteria**.

Steps to Apply Filter:

1. Select any cell in the dataset
2. Go to **Data** → **Filter**
3. Use dropdown arrows in column headers

Examples:

- Filter by **Year (2004)**
- Filter **Product Line = Classic Cars**
- Filter **Customer = Mini Gifts Distributors Ltd**

❖ **Multiple filters can be applied together**

❖ **Filtered rows are hidden, not deleted**

◆ **Custom Filters**

- **Number Filters** ($>$, $<$, Top 10)
- **Text Filters** (contains, equals, begins with)

Example:

- Sales $>$ \$2000
- Sales $<$ \$2000

◆ **Sorting Data**

Sorting organizes data to improve readability and analysis.

Types of Sorting:

- Text → A to Z / Z to A
- Numbers → Smallest to Largest / Largest to Smallest
- Dates → Oldest to Newest / Newest to Oldest

◆ **Multi-Level Sorting**

Steps:

1. Data → Sort
2. Sort by: Order Date (Oldest → Newest)
3. Add Level → Then by: Sales (Largest → Smallest)
4. Check **My data has headers**

4.3 TABLES & PIVOT TABLES

◆ **Why Convert Data to a Table?**

- Automatic filtering
- Column calculations update dynamically
- Headers remain visible

- Auto-expands when adding rows
- Banded rows improve readability

 Example:

Filtering Country = Japan updates **Sum of MSRP automatically**

◆ **Pivot Tables & Pivot Charts**

Used for **summarizing large datasets quickly.**

Advantages:

- No formulas required
- Fast data summarization
- Interactive filtering
- Visual insights using Pivot Charts

 Example:

- Orders placed in **October**
- Apply month filter → Table & Chart update instantly

4.4 USEFUL EXCEL FUNCTIONS FOR DATA ANALYSIS

◆ **IF Function (Logical)**

Syntax:

=IF(condition, value_if_true, value_if_false)

Example:

=IF(G2="shipped","Yes","No")

◆ **Nested IF (Multiple Conditions)**

=IF(A2>300,"Large",IF(A2>100,"Medium","Small"))

 Hard to maintain if conditions increase

◆ **IFS Function (Better Alternative)**

=IFS(A2>300,"Large", A2>100,"Medium", A2>0,"Small")

- ✓ Cleaner
 - ✓ Easier to manage
 - ✓ Excel 2019+, Web, 365
-

4.5 CONDITIONAL FORMATTING WITH IF

Used to **visually highlight results**.

Example:

- Retention % > 69% → **GOOD (Green)**
- Else → **POOR (Red)**

Steps:

1. Conditional Formatting → New Rule
 2. Format based on text value
-

4.6 COUNTIF & COUNTIFS

◆ **COUNTIF (Single Condition)**

=COUNTIF(range, criteria)

Example:

=COUNTIF(N2:N195,"VISA")

- ✓ Case-insensitive

◆ **COUNTIFS (Multiple Conditions)**

=COUNTIFS(range1,criteria1, range2,criteria2)

Example:

=COUNTIFS(B2:B6,"West",C2:C6,>500")

4.7 SUMIF & SUMIFS

- ◆ **SUMIF (Single Condition)**

=SUMIF(range, criteria, sum_range)

Example:

=SUMIF(AE2:AE195,"Large",AD2:AD195)

- ◆ **SUMIFS (Multiple Conditions)**

=SUMIFS(sum_range, range1, criteria1, range2, criteria2)

Example:

=SUMIFS(AD2:AD195, AE2:AE195,"Large", AL2:AL195,"BABY_BOOMERS")

4.8 LOOKUP FUNCTIONS

- ◆ **VLOOKUP (Vertical Lookup)**

Searches **left → right**.

=VLOOKUP(value, table_array, col_index, FALSE)

Example:

=VLOOKUP("Cherry", B2:C5, 2, FALSE)

 Requires lookup column to be the **leftmost**

- ◆ **HLOOKUP (Horizontal Lookup)**

Searches **top → bottom**.

=HLOOKUP(value, table_array, row_index, FALSE)

◆ **XLOOKUP (Modern Replacement)**

- ✓ Works vertically & horizontally
- ✓ No column index needed
- ✓ Better error handling

=XLOOKUP(lookup_value, lookup_array, return_array)

HANDS-ON LAB 6: FILTERING AND SORTING DATA

4.9 INTRODUCTION TO CREATING PIVOT TABLES IN MICROSOFT EXCEL

◆ **Why Pivot Tables?**

A worksheet full of data is not enough to gain insights. **Pivot Tables** allow:

- Fast data summarization
- Trend & pattern analysis
- Easy comparison of large datasets
- Dynamic updates when source data changes

👉 Pivot Tables go beyond **filters, formulas, VLOOKUP & HLOOKUP**.

◆ **What is a Pivot Table?**

A **Pivot Table** is a dynamic summary table that:

- Analyzes large datasets quickly
 - Automatically updates when data changes
 - Allows drag-and-drop analysis
 - Helps Data Analysts generate insights for stakeholders
-

4.10 PREPARING DATA FOR PIVOT TABLES (CHECKLIST)

Before creating a Pivot Table, ensure:

- ✓ Data is **formatted as a Table**
 - ✓ Only **one header row** exists
 - ✓ Column headings are clear & correct
 - ✓ No blank rows or blank columns
 - ✓ Minimal blank cells
 - ✓ Numeric values are formatted as **Numbers**
 - ✓ Date fields are formatted as **Dates**
-

4.11 FORMATTING DATA AS A TABLE

◆ Why format as a table?

- Improves structure & readability
- Auto-adds filters
- Automatically includes new rows
- Best practice for Pivot Tables

◆ Steps:

1. Select any cell in the dataset
2. **Home → Format as Table**
3. Choose a style
4. Ensure “**My table has headers**” is checked
5. Click **OK**

 New rows added automatically become part of the table

4.12 CREATING A BASIC PIVOT TABLE

◆ Steps:

1. Select any cell in the table
2. **Insert → PivotTable**
3. Excel auto-detects the table name
4. Choose **New Worksheet**

5. Click **OK**

A new sheet opens with:

- Pivot Table placeholder
 - **PivotTable Fields pane**
-

4.13 UNDERSTANDING PIVOT TABLE AREAS

◆ **PivotTable Fields Pane**

Fields can be dragged into:

Area	Purpose
Rows	Categories (e.g., Manufacturer)
Columns	Compare values across columns
Values	Calculations (Sum, Count, Avg)
Filters	High-level filtering

4.14 BUILDING A PIVOT TABLE (EXAMPLE)

⌚ Goal: Total sales per car model

Steps:

- Drag **Manufacturer** → Rows
- Drag **Model** → Rows (below Manufacturer)
- Drag **Price** → Values
- Drag **Unit Sales** → Values

📌 Fields can be rearranged by dragging

📌 Fields can be removed by dragging them out or using dropdown

4.15 FORMATTING VALUE FIELDS

Pivot Table values default to **General format**.

- ◆ **Change formatting:**

1. Click dropdown on value field
 2. **Value Field Settings**
 3. **Number Format**
 4. Choose **Currency (USD)**
 5. Set decimal places if needed
-

4.16 CALCULATIONS USING PIVOT TABLES

- ◆ **Calculated Fields**

Used to create new metrics from existing fields.

Example: Total Model Sales

Total Model Sales = Price × Unit Sales

- ◆ **Steps:**

1. PivotTable Analyze tab
2. **Fields, Items & Sets → Calculated Field**
3. Enter formula
4. Format as currency

❖ New calculated field appears in Pivot Table automatically

4.17 RECOMMENDED PIVOT TABLES

- ◆ **What are they?**

Excel-suggested Pivot Table layouts based on selected data.

- ◆ **How they work:**

- Select a column

- **Insert → Recommended PivotTables**
- Suggestions change depending on selected column
- Choose one → Excel builds it instantly

- ✓ Ideal for beginners
 - ✓ Fully customizable after creation
-

4.18 PIVOT TABLE FILTERS

- ◆ **Built-in Filters:**

- Row Labels filter
- Column Labels filter

- ✓ Easy to apply
- ✓ Easy to clear

- ◆ **Filters Area (Field Filters)**

- Drag a field into **Filters**
- Acts like a standard filter
- Shows **combined totals**, not split columns

- ❖ Columns show **separate data**
 - ❖ Filters show **aggregated totals**
-

4.19 SLICERS

- ◆ **What are Slicers?**

- Visual, button-based filters
- Show current filter state clearly
- User-friendly & interactive

- ◆ **Steps to Add Slicer:**

1. Select Pivot Table
2. **PivotTable Analyze → Insert Slicer**

3. Choose field (e.g., Territory)

- ✓ Supports multi-select
- ✓ Can clear filters instantly
- ✓ Can be freely moved & resized

 Multi-select filters **exclude** selected items

4.20 TIMELINES

◆ What is a Timeline?

- Visual filter specifically for **date fields**
- Faster than manual date filters

◆ Steps:

1. Select Pivot Table
2. **Insert → Timeline**
3. Choose date field

◆ Timeline Options:

- Filter by **Days**
- **Months**
- **Quarters**
- **Years**
- Select date ranges easily

 Can be combined with slicers

4.21 COMBINING FILTERS, SLICERS & TIMELINES

You can filter by:

- Product type (Slicer)
- Territory (Slicer)
- Year / Quarter / Month (Timeline)

- ✓ Extremely powerful for analysis
 - ✓ Shows absence of data clearly (e.g., no sales)
-

4.22 CUSTOMIZING & REMOVING SLICERS/TIMELINES

- Selecting them opens special ribbon tabs
 - Styles & colors can be changed
 - Remove by:
 - Pressing **Delete**
 - Right-click → **Cut**
-

HANDS-ON LAB 7: USING PIVOT TABLES

FINAL ASSIGNMENT

[Final Assignment - Part 1: Clean and Prepare the Data](#)

[Final Assignment - Part 2: Analyze the Data](#)