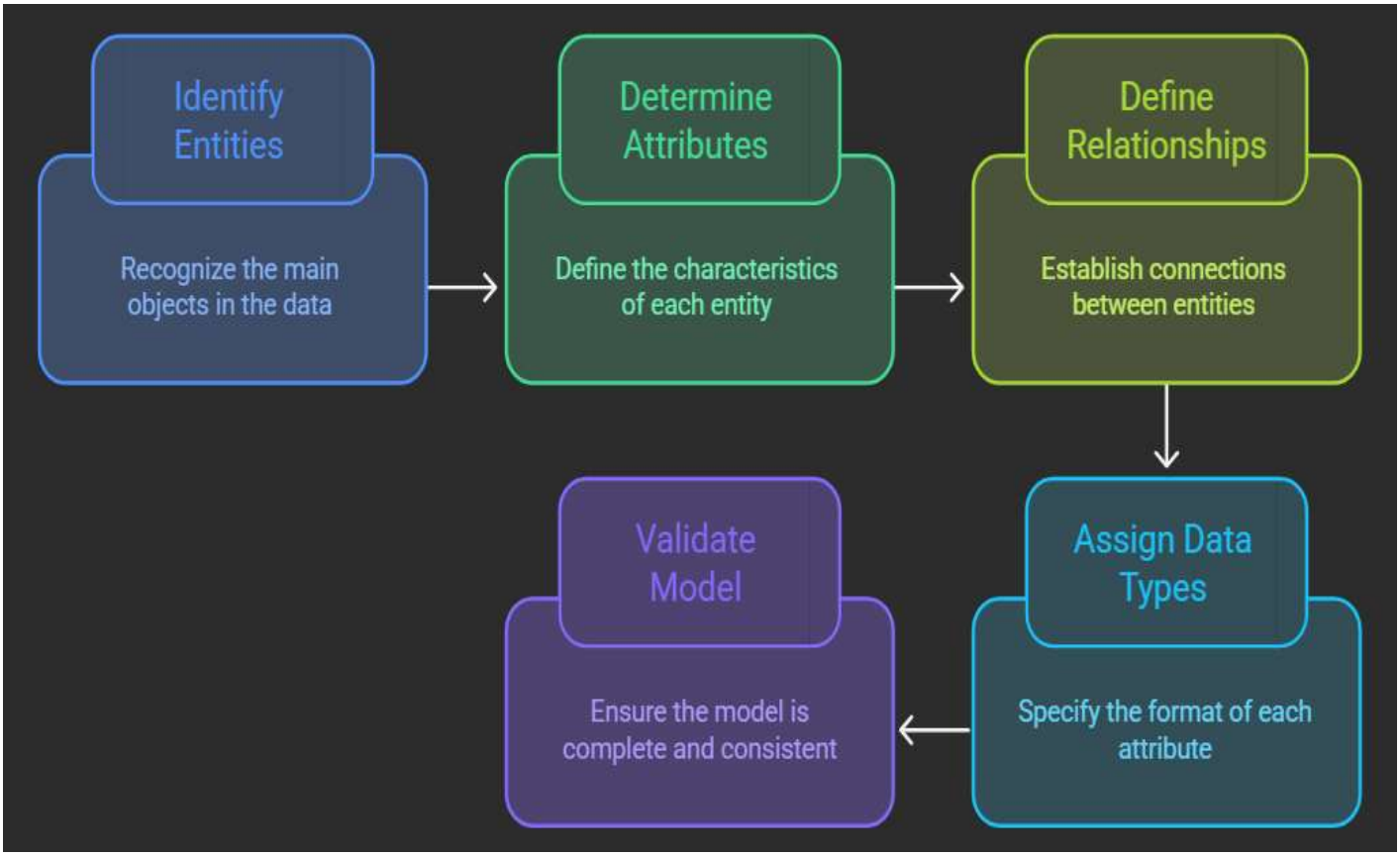
#2.3: DATA MODELLING

Data **modelling** is the process of creating a structured representation of real-world information to organize, analyze, and communicate **data** effectively. It helps in predicting **trends**, making **decisions**, and ensuring consistency in **datasets**. Core concepts include identifying **entities**, **attributes**, **relationships**, **constraints**, and selecting appropriate **data types**. **Data models** can be conceptual, logical, or physical depending on abstraction level. The **algorithm**/**process** involves identifying the main **entities** (e.g., students, courses), determining the **attributes** for each entity (e.g., student name, age, course ID), defining **relationships** between entities (e.g., students enroll in courses), assigning appropriate **data types** to each attribute (e.g., text, number, date), and validating the **model** for completeness and consistency..



#2.3.1: Data Vs Information

Data consists of raw **facts**, **numbers**, or **text** without context, whereas **information** is **data** that has been processed and organized to provide **meaning**. This distinction is crucial because only **information** can guide **decisions** and **problem-solving**. Core concepts include **collection**, **processing**, **summarization**, and **visualization** of **data**. Converting **data** into **information** may involve **sorting**, **filtering**, **calculating**, or creating **charts**. Understanding this **concept** ensures that students can distinguish between unprocessed **data** and actionable **insights**.

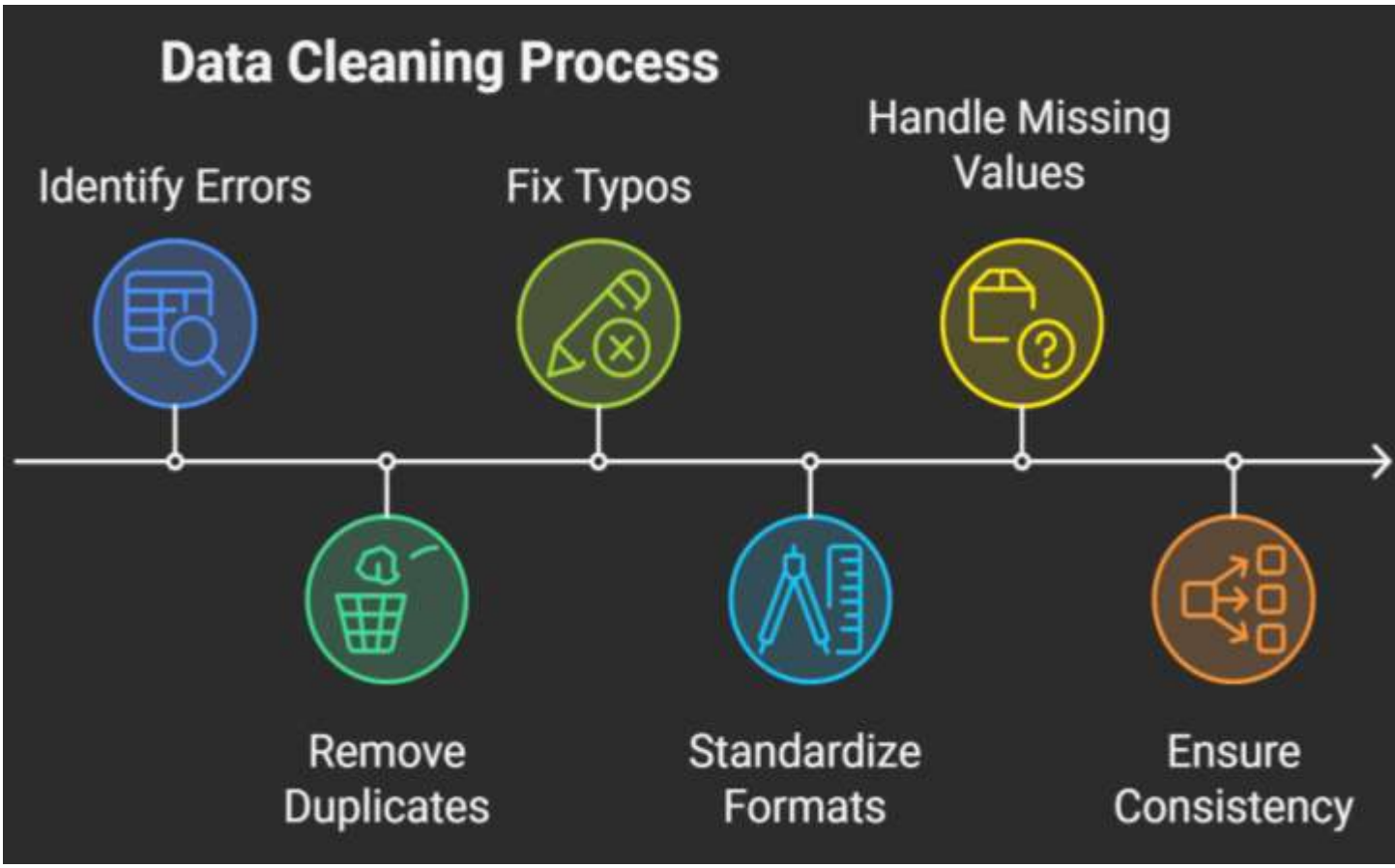
#2.3.3: Data Collection: MS Excel

Data collection is the process of gathering raw **facts** and **figures** from various **sources** for **analysis**. It is important because accurate and complete **data** is the foundation for reliable **conclusions**. Core concepts include identifying **sources**, selecting **relevant data**, **recording systematically**, and ensuring **consistency**. Proper **collection in Excel** involves creating **structured tables** and using **data validation** to prevent **errors**. For example, your school wants to record students' **favorite sports**. The sheet **Sports Collection** already has headers **Name**, **Class**, **Favorite Sport** and some blank rows. Complete the **dataset** by entering **data** for all students by opening **StudentData.xlsx**, checking **headers**, entering **data for 10 students** under each column, using **Data → Data Validation → Allow: Text** for **Favorite Sport** column, making **headers bold**, and saving the **workbook**.

#2.3.4: Data Cleaning

Data cleaning is the process of identifying and correcting errors,

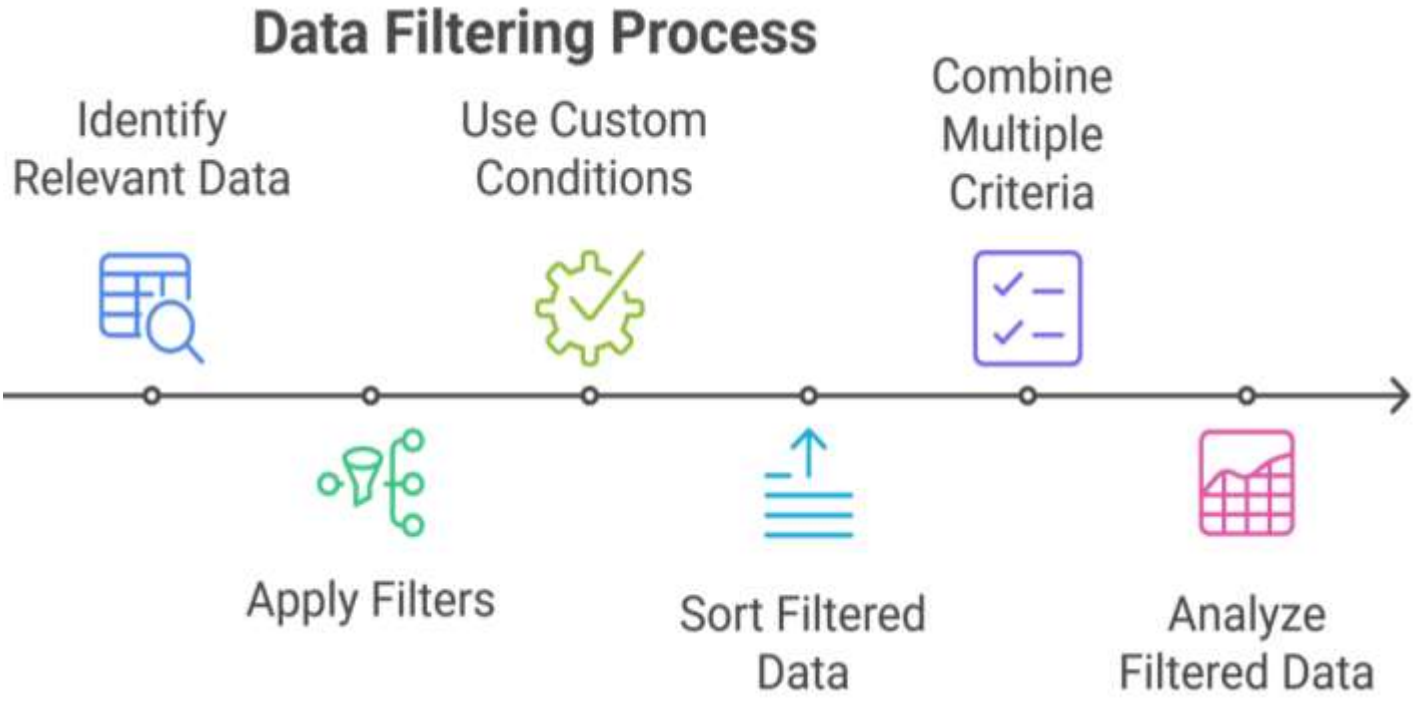
inconsistencies, or missing values in a dataset. It is important because dirty data can lead to misleading results and poor decisions. Core concepts include removing duplicates, fixing typos, standardizing formats, handling missing values, and ensuring consistency. For example, in the sheet ScoresCleaning, some names are inconsistent and a few scores are missing. Clean the dataset by using TRIM() to remove extra spaces, standardizing names with UPPER() or LOWER(), highlighting missing scores using Conditional Formatting → Blank Cells, filling in missing scores or replacing with 0, and saving the workbook.



#2.3.5: Data Filtering: MS Excel

Data filtering is the process of displaying only the rows in a dataset that meet certain criteria. It is important because it

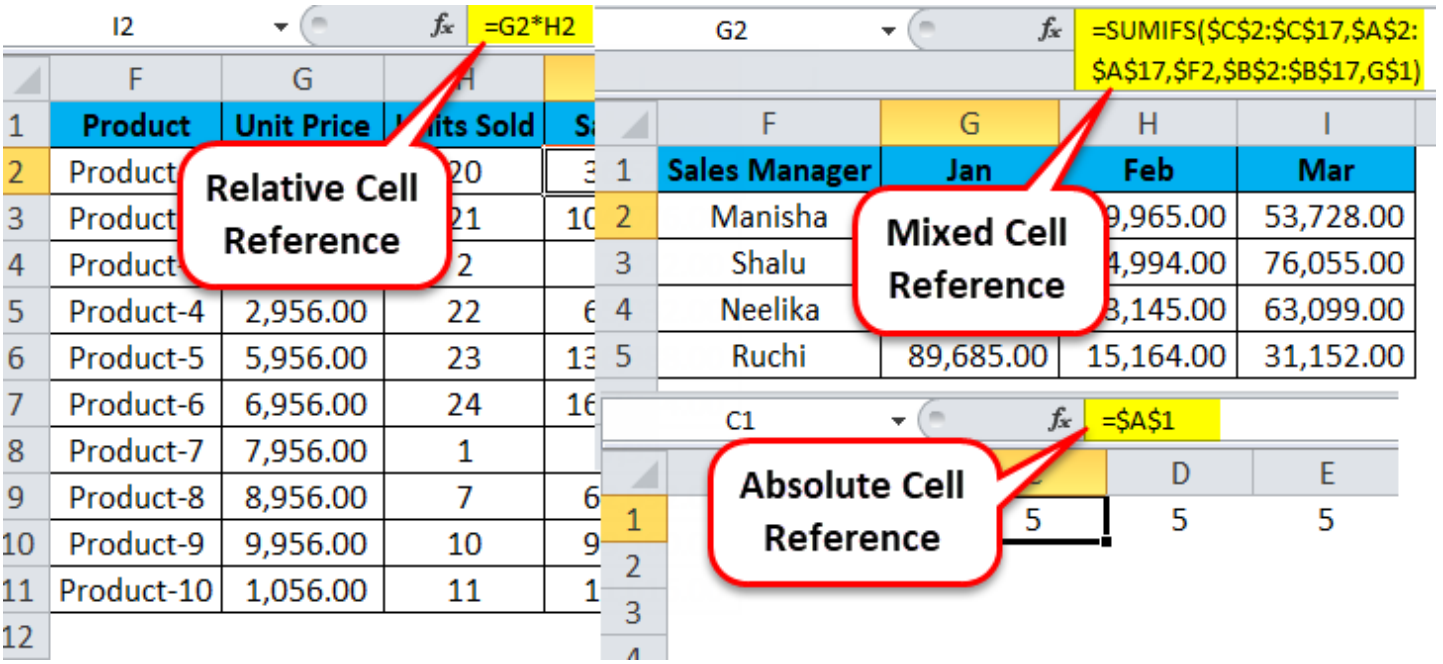
allows users to focus on relevant information and quickly analyze subsets of data. Core concepts include applying filters, using custom conditions, sorting within filtered data, and combining multiple criteria. For example, in the sheet ScoresFiltering, filter students who scored more than 80 in Test 1 and are in Class 10 by applying Data → Filter, using Number Filters → Greater Than → 80 on Test 1, selecting Class 10, reviewing filtered results, clearing filters to view all, and saving the workbook.



#2.3.6: Cell Referencing

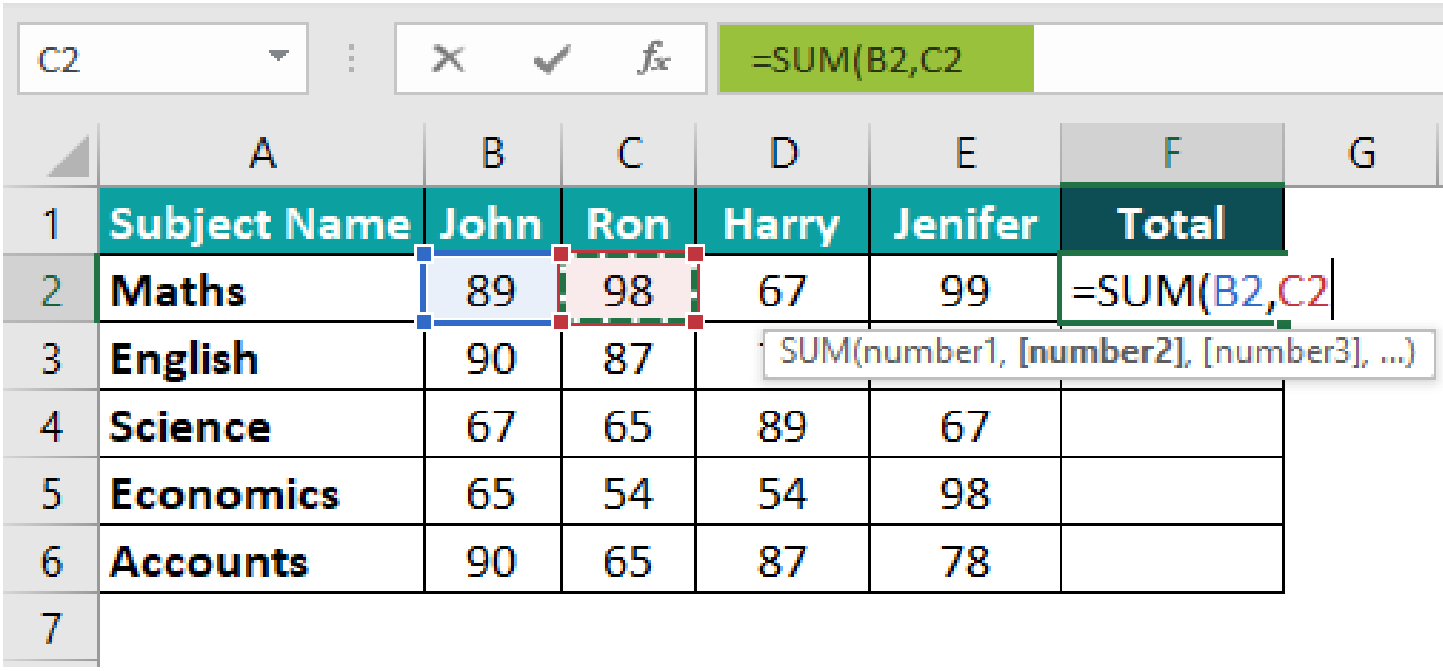
Cell referencing is the method of referring to specific cells in a worksheet to perform calculations or analyses. It is important because it allows formulas to update automatically when copied,

saving time and reducing errors. Relative cell references change automatically when a formula is copied (e.g., B2). Absolute cell references remain fixed using the $ symbol (e.g., $B$2). Mixed cell references fix either the row or the column while allowing the other to change (e.g., $B2 or B$2). For example, in the sheet CellReferencing, calculate final marks using relative, absolute, and mixed references by entering formulas =B2+C2, =B2+C2+$F$1, =B2+$C2+$F$1, copying down, and saving the workbook.

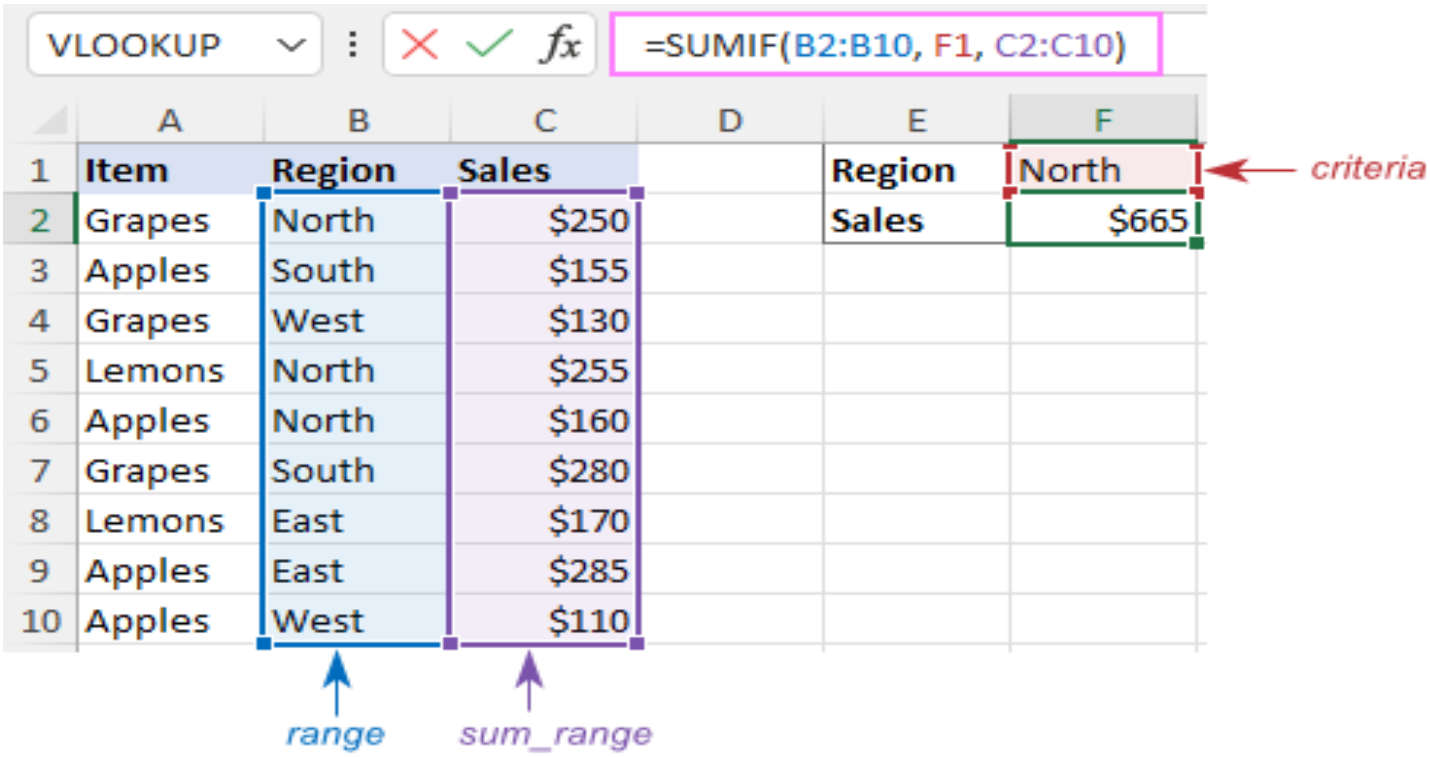


#2.3.7: Data Calculations: SUM

The SUM function is used to add values in a range of cells efficiently. It is important because it simplifies calculations for large datasets, ensuring accuracy and saving time. Core concepts include summing rows, columns, or specific ranges. For example, in the sheet SumCalculation, calculate total sales for each product using =SUM(B2:D2), copy the formula down, and save the workbook.



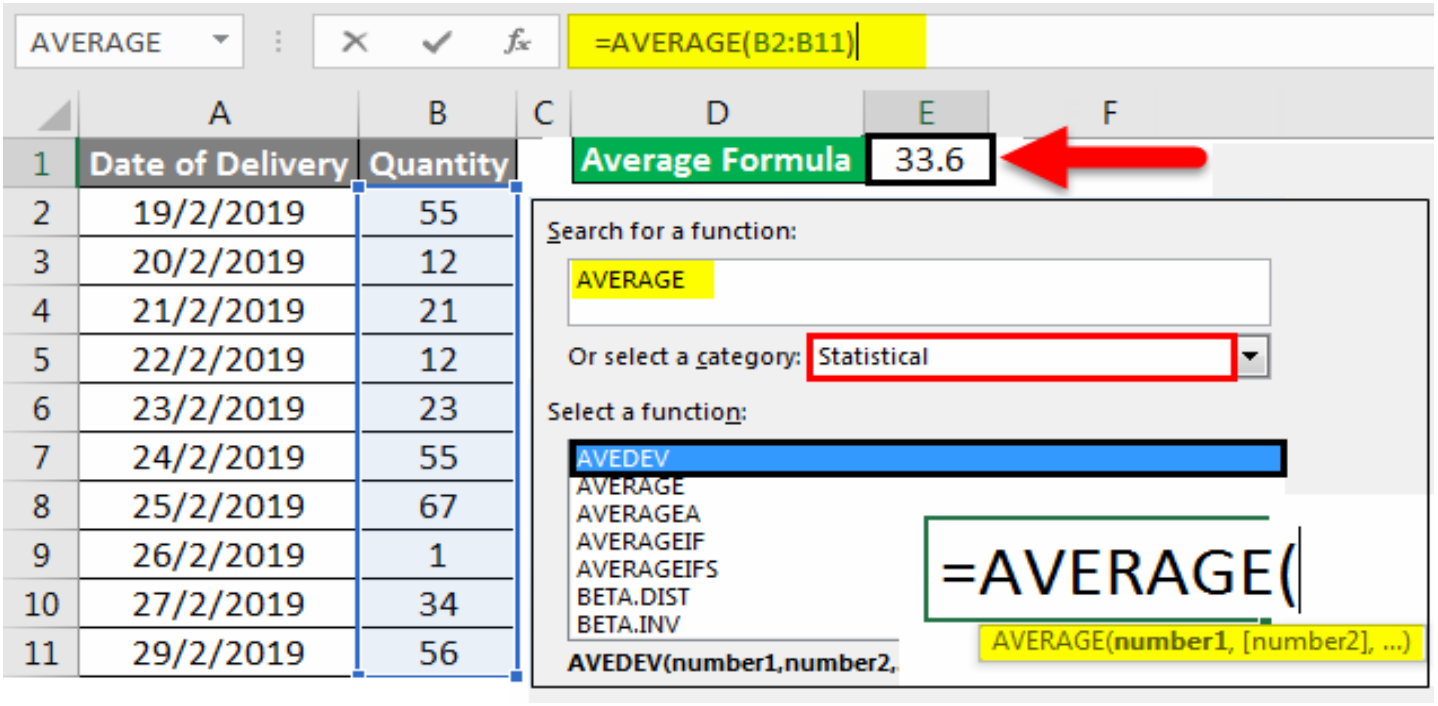
2.3.8: Conditional Summarization with SUMIF
SUMIF allows summing of values that meet a specific condition. It is important because it enables targeted calculations, such as totals for a category or group, without manual filtering. Core concepts include defining the criteria range, condition, and sum range. For example, in the sheet SumIfExample, calculate total marks for students in Class 10 using =SUMIF(B2:B11,"10",C2:C11), repeat for other subjects, and save the workbook.



	A	B	C	D	E	F
1	**Item**	**Region**	**Sales**		**Region**	North
2	Grapes	North	$250		**Sales**	$665
3	Apples	South	$155			
4	Grapes	West	$130			
5	Lemons	North	$255			
6	Apples	North	$160			
7	Grapes	South	$280			
8	Lemons	East	$170			
9	Apples	East	$285			
10	Apples	West	$110			

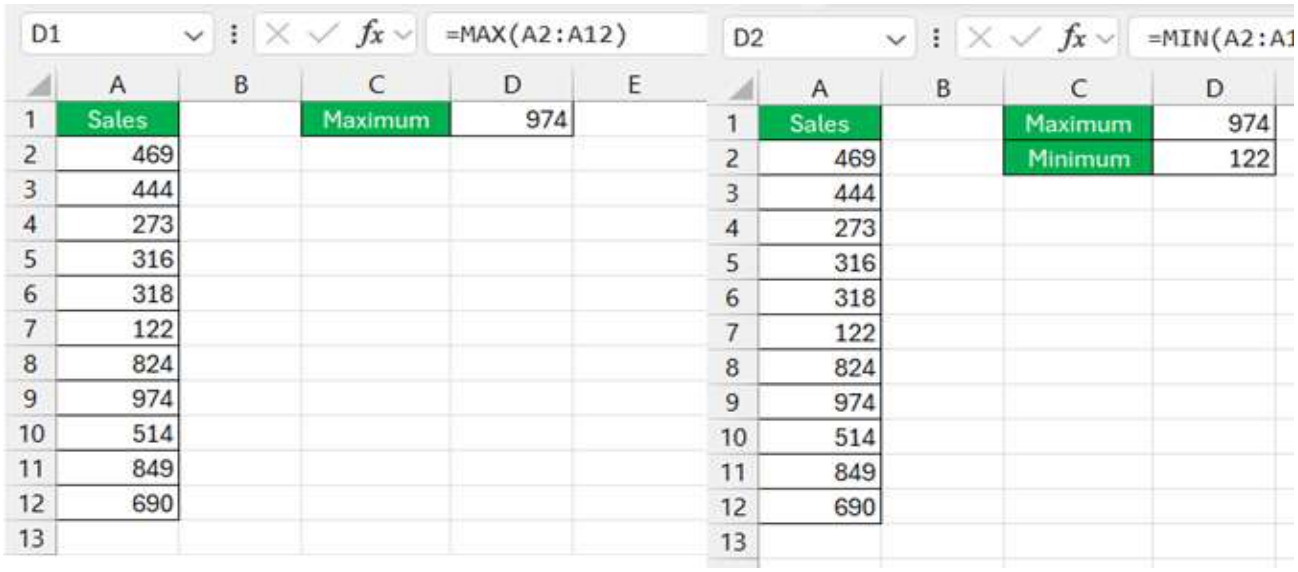
#2.3.9: Calculating Mean Values with AVERAGE

The AVERAGE function calculates the arithmetic mean of a set of numbers. It is important because it helps summarize data and identify overall trends or typical values. Core concepts include averaging ranges, ignoring blank cells, and combining multiple ranges. For example, in the sheet AverageCalculation, calculate the average score for each student using =AVERAGE(B2:D2), copy down for all students, and save the workbook.



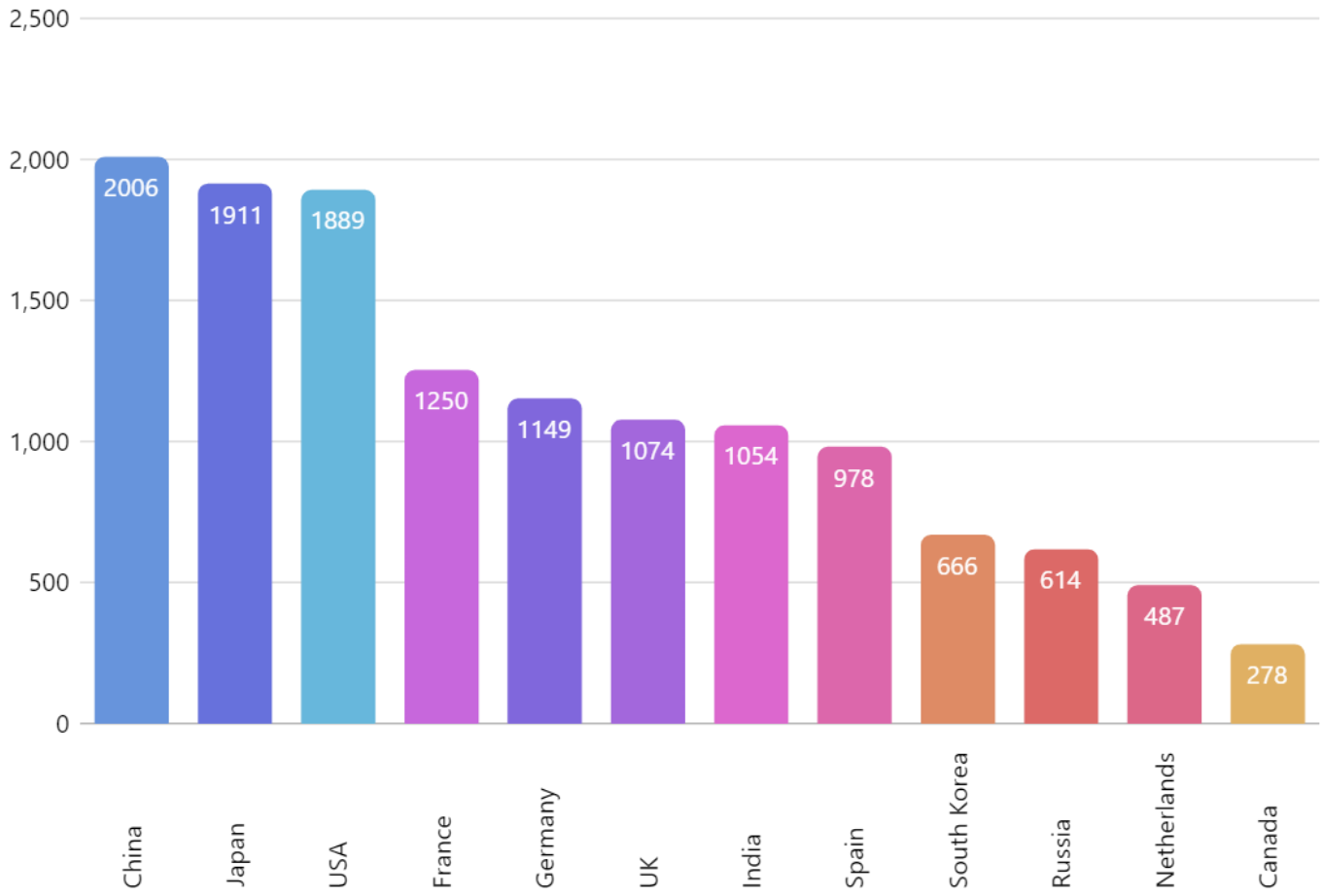
#2.3.10: MAX / MIN

MAX returns the largest value in a range, while MIN returns the smallest value. These functions are important because they quickly identify extremes in datasets, such as top scores or lowest sales. Core concepts include applying MAX/MIN to rows, columns, or ranges and combining with other functions for analysis. For example, in the sheet MaxMinExample, find highest and lowest scores using =MAX(B2:B11) and =MIN(B2:B11), repeat for other tests, and save the workbook.



#2.3.11: Data Sorting

Data sorting is arranging data in a specific order, such as ascending, descending, or by custom criteria. It is important because it allows quick identification of trends, rankings, or patterns in datasets. Core concepts include single-column sorting, multi-level sorting, and sorting numerical or text data. For example, in the sheet DataSorting, sort students by Class ascending and then by Total Marks descending using Data → Sort, apply, review results, and save the workbook.



#Lesson Summary

Cell Referencing enables formulas to dynamically link worksheet data using Relative, Absolute ($), and Mixed References, ensuring accuracy and reusability across datasets. SUM simplifies adding values across rows or ranges (e.g.,

=SUM(B2:D2)), saving time and reducing errors. SUMIF supports conditional calculations for grouped analysis. AVERAGE calculates mean values to identify typical performance trends while ignoring blanks. MAX and MIN detect highest and lowest values, helping identify extremes and outliers. Data Sorting organizes information in ascending, descending, or multi-level order to reveal patterns and rankings. Together, these tools support efficient Data Modelling, accurate analysis, and structured reporting in spreadsheets.