

# Unit 1. Fundamentals of data Analysis

## 1.1. Meaning of Data

### Definition:

The term data refers to raw facts and figures that are collected from various sources and have not yet been processed, analyzed, or interpreted. By themselves, data do not convey any meaningful information until they are organized and analyzed.

### Key Points:

- Data is the basic input and starting point of all statistical analysis.
- Data may be **qualitative** (descriptive in nature) or **quantitative** (numerical in nature).
- Data can be collected through observations, surveys, experiments, interviews, or existing records.
- Proper classification and analysis of data are necessary to convert it into meaningful information.

### Examples:

- **Marks obtained by students in an examination:** 65, 70, 55, 90, 85
- **Daily sales of a shop (in ₹):** 1,000; 1,200; 950; 1,400; 1,100

## 1.2. Meaning of Information

### Definition:

Information refers to data that has been processed, organized, analyzed, or summarized in a meaningful form so that it provides knowledge or insight useful for decision-making.

### Key Points:

- Information is derived after proper processing and analysis of raw data.
- It helps in identifying patterns, relationships, and trends in data.
- Information is meaningful and relevant, unlike raw data.
- Managers and decision-makers use information for planning, control, evaluation, and decision-making.

### Examples:

- The average marks of a class is 73, which helps teachers assess the overall academic performance of students.
- The average daily sales of a shop are ₹1,130, which assists the owner in planning inventory and sales strategies.

### 1.3. Difference between Data and Information

| Basis              | Data   | Information   |
|--------------------|--|---|
| Meaning            | Data refers to raw facts and figures without any processing. | Information is processed and organized data that has meaning. |
| Nature             | Unorganized and unprocessed.                                 | Organized, summarized, and analyzed.                          |
| Significance       | By itself, data has little or no meaning.                    | Information is meaningful and useful.                         |
| Stage              | Data is the input for analysis.                              | Information is the output of analysis.                        |
| Usefulness         | Not directly useful for decision-making.                     | Highly useful for planning and decision-making.               |
| Understanding      | Difficult to understand without processing.                  | Easy to understand and interpret.                             |
| Example            | Marks of students: 65, 70, 55, 90                            | Average marks of class: 73                                    |
| Example (Business) | Daily sales figures of a shop                                | Average daily sales or sales trend                            |

### 1.4. Concept of Data Analysis

#### ➤ Meaning:

Data analysis is the systematic and scientific process of collecting, organizing, classifying, summarizing, and interpreting data in order to extract useful information. It involves applying statistical tools and techniques to transform raw data into meaningful insights that support planning, decision-making, and control in business and research activities.

In simple words, data analysis helps convert large volumes of raw data into understandable information that can be used to draw conclusions and make informed decisions.

#### ➤ Purpose of Data Analysis:

##### 1. To Convert Raw Data into Meaningful Information

Data collected in its raw form has little value. Data analysis organizes and processes this data so that it becomes useful and understandable.

##### 2. To Support Business Decisions

Managers rely on analyzed data to make decisions related to pricing, production, marketing, finance, and human resources. Data analysis reduces uncertainty and risk in decision-making.

### 3. To Identify Problems and Opportunities

By analyzing trends, patterns, and variations in data, organizations can identify operational problems as well as new growth opportunities.

### 4. To Make Predictions and Forecasts

Past data, when properly analyzed, helps in forecasting future outcomes such as sales, demand, profits, and market trends.

### 5. To Test Hypotheses and Draw Conclusions

Data analysis helps in testing assumptions or hypotheses using statistical tools, thereby enabling researchers to accept or reject hypotheses scientifically.

#### ➤ Example:

A company collects monthly sales data for a period of 12 months. After analyzing the data, it is observed that sales increase during festive seasons and decline during summer months. This information helps the company plan higher production and promotional activities during festivals and reduce costs during low-demand periods.

## 1.5. Types of Data Analysis

### 1.5.1. Descriptive Statistics

#### ➤ Meaning:

Descriptive statistics is a branch of statistics that deals with **collecting, organizing, summarizing, and presenting data** in a meaningful manner. Its main objective is to describe the main characteristics of a given set of data so that patterns, trends, and variations can be easily understood.

Descriptive statistics does not draw conclusions beyond the data collected; it simply explains what the data shows.

#### ➤ Common Tools of Descriptive Statistics:

- Measures of Central Tendency: **Mean, Median, Mode**
- Measures of Dispersion: **Range, Variance, Standard Deviation**
- Tabulation
- Graphs and Charts (Bar chart, Pie chart, Line graph)
- Percentages and Ratios

#### ➤ Examples:

##### • Example 1:

Average sales of 5 shops = ₹12,000

(This summarizes large sales figures into a single representative value.)

- **Example 2:**

A pie chart showing the market share of four mobile brands

(This helps in visual comparison and understanding market position.)

- **Features of Descriptive Statistics**

1. **Summarizes Large Data:**

Descriptive statistics helps in reducing large volumes of raw data into simple summaries such as averages, percentages, or totals. This makes complex data easy to handle and understand.

2. **Easier Interpretation:**

Data is presented using tables, graphs, and charts, which makes interpretation simple and clear even for non-technical users.

3. **No Generalization:**

The results of descriptive statistics are limited only to the data under study. It does not make predictions or generalizations about the entire population.

4. **Foundation for Further Analysis:**

Descriptive statistics serves as the base for inferential statistics. It helps identify patterns and trends that can later be tested statistically.

5. **Quantitative Presentation:**

It focuses mainly on numerical representation of data through figures, percentages, averages, and graphical forms.

6. **Helps in Decision Making:**

Although limited in scope, descriptive statistics provides useful insights for preliminary business decisions such as sales performance evaluation and trend analysis.

### 1.5.2. Inferential Statistics

- **Meaning:**

Inferential statistics is a branch of statistics that deals with drawing conclusions, making predictions, or taking decisions about a large group (population) based on data collected from a small representative group (sample).

Since studying the entire population is often costly, time-consuming, or impractical, inferential statistics provides scientific methods to make reliable estimates and decisions using sample data.

- **Common Tools of Inferential Statistics:**

- **Hypothesis Testing**
- **Correlation Analysis**
- **Regression Analysis**
- **Confidence Intervals**
- **t-test, F-test, Chi-square test**

➤ **Examples:**

- **Example 1:**

Predicting election results based on a survey conducted among 2,000 voters instead of the entire voting population.

- **Example 2:**

Testing whether a new teaching method improves student performance by comparing test scores of a sample of students.

➤ **Features of Inferential Statistics**

**1. Based on Sampling:**

Inferential statistics relies on sample data to draw conclusions about the entire population. The accuracy of inference depends on how well the sample represents the population.

**2. Helps in Decision Making:**

It supports managerial and policy decisions by predicting future trends and outcomes, such as demand forecasting or quality control.

**3. Involves Probability Theory:**

Inferential statistics deals with uncertainty and uses probability theory to estimate the likelihood of events and measure risk.

**4. Hypothesis Testing:**

It involves formulating and testing hypotheses to determine whether assumptions or claims are supported by data. Decisions are made to accept or reject the null hypothesis.

**5. Generalization:**

Unlike descriptive statistics, inferential statistics allows results obtained from sample data to be generalized to the entire population.

**6. Measures Relationship and Impact:**

Tools like correlation and regression help in understanding relationships between variables and estimating the impact of one variable on another.

## 1.6. Meaning of Hypothesis

### ➤ Definition:

A **hypothesis** is a clear and testable statement or assumption made about a population parameter. It represents a tentative explanation or proposed answer to a research question, which can be verified or rejected using statistical data and methods.

In research and data analysis, a hypothesis provides direction to the study and forms the basis for statistical testing and decision-making.

A hypothesis is formulated before conducting the analysis. It helps researchers focus on specific relationships or differences between variables. Since it is based on assumptions, it must be tested using sample data and appropriate statistical tools such as t-tests, F-tests, or correlation analysis.

### ➤ Examples:

- “Advertising has a positive effect on sales.”
- “There is no difference in performance between male and female students.”

### ➤ Types of Hypotheses

#### 1. Null Hypothesis ( $H_0$ ):

The **null hypothesis** states that there is **no difference, no relationship, or no effect** between variables. It represents the default assumption that any observed difference is due to chance or random variation.

### ➤ Purpose:

It is tested statistically and either accepted or rejected based on evidence from sample data.

### ➤ Example:

- “Advertising has no effect on sales.”
- “There is no difference in academic performance between male and female students.”

#### 2. Alternative Hypothesis ( $H_1$ ):

The **alternative hypothesis** states that there is a **significant difference, relationship, or effect** between variables. It contradicts the null hypothesis and represents what the researcher aims to prove.

### ➤ Purpose:

It is accepted when sufficient statistical evidence leads to the rejection of the null hypothesis.

### ➤ Example:

- “Advertising increases sales.”

- “There is a significant difference in performance between male and female students.”

## 1.7. Process of Hypothesis Testing

Hypothesis testing is a systematic procedure used in statistics to determine whether a given assumption about a population is supported by sample data or not. The process involves the following steps:

### Step 1: Formulation of Hypotheses

The first step is to clearly state the hypotheses to be tested:

- **Null Hypothesis ( $H_0$ ):** Assumes no difference or no effect.
- **Alternative Hypothesis ( $H_1$ ):** Assumes the presence of a difference or effect.

*Example:*

$H_0$ : Advertising has no effect on sales.

$H_1$ : Advertising has a significant effect on sales.

### Step 2: Selection of Significance Level ( $\alpha$ )

The significance level represents the probability of rejecting a true null hypothesis. Commonly used levels are **5% (0.05)** or **1% (0.01)**.

This level determines how strong the evidence must be to reject the null hypothesis.

### Step 3: Selection of Appropriate Test Statistic

Based on the nature of data, sample size, and hypothesis, a suitable statistical test is selected, such as:

- t-test
- F-test
- Z-test
- Chi-square test

### Step 4: Collection of Sample Data

Relevant data is collected from the population using appropriate sampling methods. The data must be reliable and representative.

### Step 5: Calculation of Test Statistic

Using the collected sample data, the value of the test statistic is calculated with the help of statistical formulas or software such as Excel or SPSS.

### Step 6: Determination of Critical Value / p-value

The calculated test statistic is compared with:

- **Critical value**, or
- **p-value** corresponding to the chosen significance level.

### Step 7: Decision Rule

- If the calculated value falls in the rejection region or  $p\text{-value} \leq \alpha \rightarrow$  **Reject  $H_0$**
- If the calculated value does not fall in the rejection region or  $p\text{-value} > \alpha \rightarrow$  **Accept  $H_0$**

### Step 8: Conclusion and Interpretation

A conclusion is drawn in the context of the problem, stating whether the sample data provides sufficient evidence to support the alternative hypothesis.

*Example:*

“There is sufficient statistical evidence to conclude that advertising significantly affects sales.”

Dr. Rudri & Dr. Gautam

# Unit 2: Data Visualisation and Descriptive Statistics using Excel

## 2.1. Data visualisation in Excel

### 1.1 What a Good Business Chart Must Have — Detailed Theoretical Notes

A business chart is not decoration; it is a decision-making tool. Its primary purpose is to communicate information clearly, accurately, and quickly to managers, investors, students, or policymakers. A good chart reduces cognitive effort and highlights insights that would otherwise be hidden in raw data.

#### 1. Clear Objective: The Chart Must Answer a Question

Every chart must start with a single, well-defined question.

If the question is unclear, the chart will confuse rather than inform.

#### Examples of clear questions:

- *Which region has the highest sales?*
- *How are profits changing over time?*
- *What share does each product contribute to total revenue?*
- *Is there a relationship between advertising spend and sales?*

#### A chart should answer one question only.

Trying to answer multiple questions in one chart leads to clutter and misinterpretation.

#### 2. Appropriate Chart Type Selection

Choosing the correct chart type is critical. Each chart type is designed to convey a specific kind of information.

##### a) Comparison (Differences Across Categories)

Purpose: Compare values across groups

Best charts: Bar chart, Column chart

Example: Sales by region, expenses by department

Avoid using pie charts for comparisons when values are close, as human eyes compare lengths better than angles.

##### b) Trend (Change Over Time)

Purpose: Show movement, growth, or decline

Best charts: Line chart, Area chart

Example: Monthly revenue, yearly profit growth

Time should always be on the horizontal (X) axis, and data values on the vertical (Y) axis.

##### c) Composition (Parts of a Whole)

Purpose: Show how a total is divided

Best charts: Pie chart (limited categories), Stacked bar chart

Example: Market share, cost structure

Pie charts should ideally have no more than 5–6 segments to remain readable.

#### **d) Relationship (Correlation Between Variables)**

Purpose: Show how two variables relate

Best charts: Scatter plot, Bubble chart

Example: Advertising spend vs. sales

These charts help identify patterns such as positive, negative, or no correlation.

### **3. Essential Elements of a Good Chart**

#### **a) Descriptive Title**

A good title:

- States what is being shown
- Includes time period where applicable

Example:

- “*Quarterly Sales Performance (FY 2024–25)*”
- Avoid vague titles like “*Sales Chart*”.

#### **b) Clearly Labelled Axes**

Both axes must be labelled clearly:

- What is measured
- Unit of measurement

Examples:

- ₹ (Rupees)
- Units sold
- Percentage (%)

Unlabelled axes force viewers to guess and reduce credibility.

#### **c) Readable and Honest Scale**

- Use consistent intervals.
- Start axes at zero where possible.
- If a non-zero baseline is used, it must be clearly stated, as it can exaggerate differences.

Misleading scales can unintentionally (or intentionally) distort conclusions.

#### **d) Limited and Meaningful Use of Colors**

- Use 2–4 colors ideally.
- Do not exceed 6–8 categories.
- Use contrasting colors for clarity.
- Maintain consistency (e.g., the same region always has the same color across charts).

Avoid decorative or neon colors that distract from data.

### **4. Readability and Simplicity**

A good chart should be readable from a distance (important for presentations).

Key guidelines:

- Adequate font size for titles, labels, and legends

- Minimal gridlines (light grey if used)
- Avoid unnecessary data labels—label only where needed

Simplicity increases comprehension speed.

## 5. Common Mistakes in Business Charts and How to Avoid Them

### 1: Too Many Data Series

Problem: Overcrowding and confusion

Solution:

- Break into two or more charts
- Use filters or small multiples

### 2: Use of 3-D Charts

Problem: Distorts perception of size and angle

Solution:

- Use clean 2-D charts
- Flat charts improve accuracy and professionalism
- 

### 3: Tiny Fonts and Overloaded Labels

Problem: Poor readability

Solution:

- Increase font size
- Rotate labels only if necessary
- Remove redundant text

### 4: Chart Without Context

Problem: Viewer cannot interpret meaning

Solution:

- Add time period, units, or short annotation if needed
- Provide a brief caption for complex charts

### 6. Key Principle: “Less Ink, More Insight”

A good chart:

- Minimizes non-data elements
- Maximizes insight
- Directs attention to what matters most

If a chart does not clearly support a business decision, it should be redesigned or replaced.

## 1.2 Column/Bar charts (category comparison)

**Purpose:** Compare size of categories (products, regions, departments).

**Types:**

- Clustered Column: several series side by side (e.g. Region vs Year).

- Stacked Column: composition within one bar (e.g. total sales split by product).
- 100% Stacked Column: all bars scaled to 100% to compare proportions.

### Detailed steps – simple clustered column

1. Arrange data:

| Category | Sales |
|----------|-------|
| North    | 150   |
| South    | 120   |
| East     | 100   |
| West     | 180   |

2. Select A1:B5.

3. Insert → Column → Clustered Column.

4. Click chart title → type “Sales by Region (₹ ‘000)”.

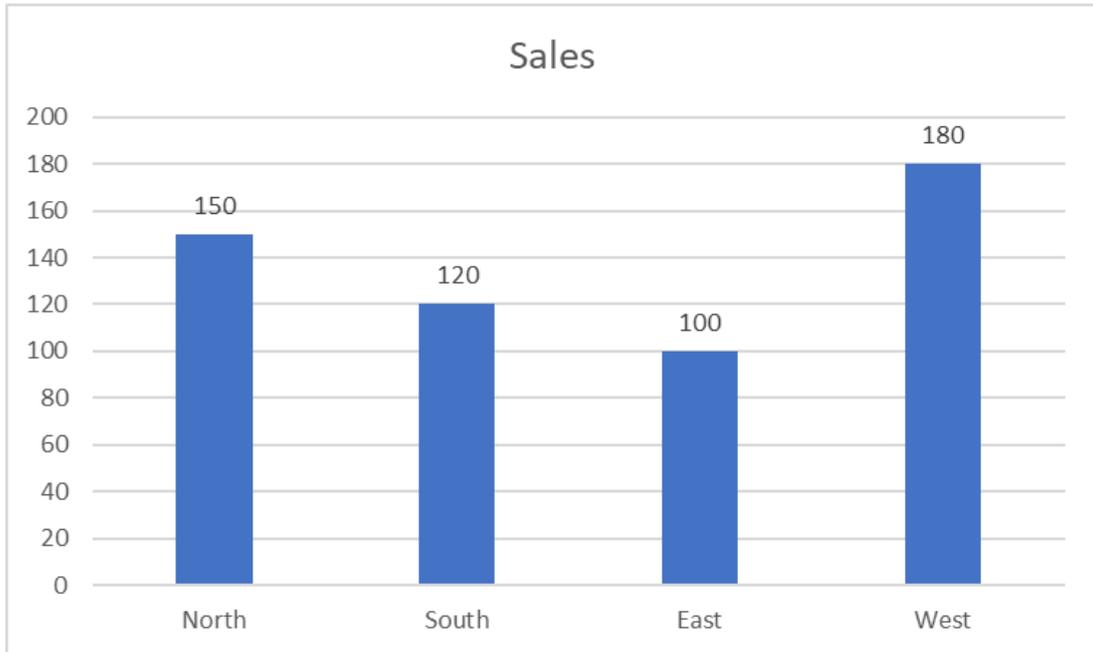
5. Layout:

- Chart Design → Add Chart Element → Axis Titles → Primary Horizontal/Vertical.
- Enter “Region” on X-axis and “Sales (₹ ‘000)” on Y-axis.

6. Format Y-axis:

- Right-click → Format Axis → set Minimum = 0.
- Set Major unit (e.g. 50) so gridlines are clear.

7. Add Data Labels (optional): right-click bars → Add Data Labels.



### Common variations

- **Multiple years:** add another column (Sales 2025). Excel will automatically show two bars per region.
- **Sorted bar chart:** sort data (largest to smallest) before inserting chart so bars descend visually.

### 1.3 Line charts (trend over time)

**Purpose:** Show movement over continuous time (day/week/month/quarter/year).

#### Rules:

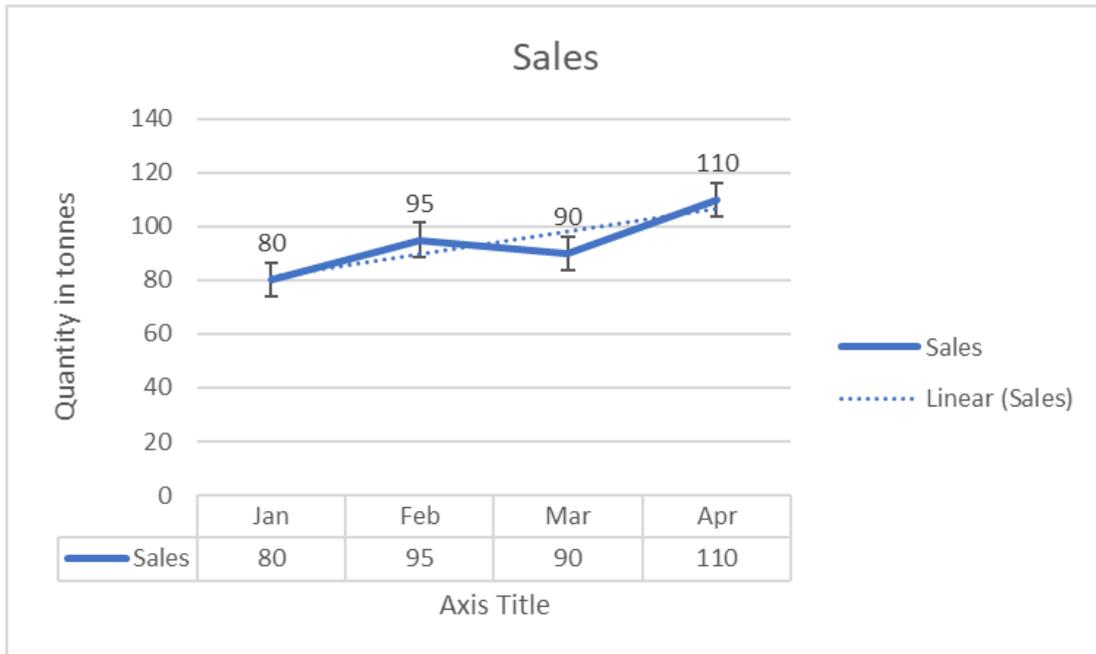
- X-axis must be time and equally spaced.
- Use lines for 1–3 series; with many series, chart becomes confusing.

#### Steps

1. Data:

| Month | Sales |
|-------|-------|
| Jan   | 80    |
| Feb   | 95    |
| Mar   | 90    |
| Apr   | 110   |

2. Select range → Insert → Line → “Line with Markers”.
3. Use Chart Design → Add Chart Element:
  - Axis titles (“Month”, “Sales (units)”).
4. Right-click line → Add Trendline → choose “Linear” if you want overall growth direction; tick “Display Equation” and “Display R-squared”.



Reading line charts:

- Upward slope: increasing trend.
- Downward slope: decreasing trend.
- Flat: stable performance.
- Sudden jump/fall: may indicate a special event or problem.

#### 1.4 Pie charts (composition of total)

**Purpose:** Show percentage contribution of each category to a total.

**When to use / when not to use**

- Use when there is **one** series that sums to a meaningful whole (100%).
- Limit to 5–6 slices; otherwise labels and colors clutter.

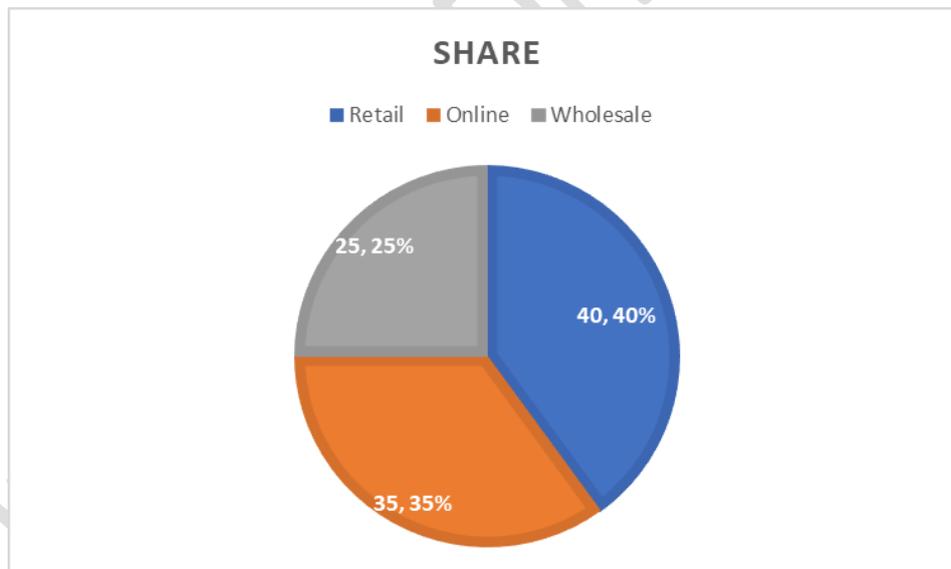
- Avoid for precise comparison; use bar chart if exact differences matter.

## Steps

1. Data:

| Channel   | Share |
|-----------|-------|
| Retail    | 40    |
| Online    | 35    |
| Wholesale | 25    |

2. Select table → Insert → Pie → 2-D Pie.
3. Right-click → Add Data Labels.
4. Format Data Labels → check “Category Name” and “Percentage”.
5. Move labels outside for clarity; avoid 3-D, explosion except for highlighting one slice.



## 1.5 Scatter plots (relationship between two variables)

**Purpose:** Show correlation between **two numerical variables**.

Examples for management:

- Advertising expenditure vs. sales.
- Years of experience vs. salary.

## Steps

1. X variable in column A, Y variable in column B.
2. Select both columns → Insert → Scatter → first type (only markers).
3. Add axis titles: “Advertising Spend (₹)” and “Sales (₹)”
4. Right-click any point → Add Trendline → Linear.
5. Tick “Display Equation on chart” and “Display R-squared value”.

## Interpretation

- Positive slope: as X increases, Y increases.
- Negative slope: as X increases, Y decreases.
- $R^2$  close to 1 → strong linear relation; near 0 → weak.

## 1.6 Sparklines (mini charts in cells)

**Types:** Line, Column, Win/Loss.

- Line: general trend.
- Column: up/down bar pattern.
- Win/Loss: only direction (positive or negative).

## Detailed steps

| Month | Sales (Units) |
|-------|---------------|
| Jan   | 150           |
| Feb   | 180           |
| Mar   | 165           |
| Apr   | 200           |
| May   | 220           |
| Jun   | 210           |
| Jul   | 195           |
| Aug   | 240           |
| Sep   | 260           |
| Oct   | 255           |
| Nov   | 290           |
| Dec   | 310           |

1. Suppose B2:M2 contains monthly sales for Product A.
2. Select cell N2 → Insert → Sparklines → Line.
3. Data Range: B2:M2; Location Range: N2 → OK.
4. Use Sparkline Design tab:
  - Check “Markers” and “High Point/Low Point”.
  - Change line color to dark; high point marker to green, low to red.
5. Drag fill handle from N2 downward to copy sparklines for other products

You can also adjust vertical axis to be same for all sparklines: Sparkline Design → Axis → Same for all Sparklines.



## 1.7 Introduction to Pivot Table

A Pivot Table is a data summarization tool found in spreadsheet programs like Excel and Google Sheets. It is called a "pivot" table because you can rotate (pivot) its structure to look at the same data from different perspectives without changing the original dataset.

### 1. What is a Pivot Table?

At its core, a pivot table is a dynamic engine that processes raw, unorganized data into a meaningful summary. If a standard table is a "list," a pivot table is a "report."

- Non-Destructive: It does not change your original data; it simply creates a "view" of it.

- **Formula-Free:** You can perform complex calculations (like finding the average sales per region) by dragging and dropping fields rather than writing long formulas like `=SUMIFS` or `=VLOOKUP`.
- **Multidimensional:** It allows you to cross-reference multiple categories at once (e.g., viewing "Sales" by both "Product Category" and "Quarter").

## 2. The Four Pillars of a Pivot Table

When you create a pivot table, you organize your data into four distinct areas:

1. **Rows:** The labels that appear on the left side (e.g., Employee names or Product categories).
2. **Columns:** The labels that appear across the top (e.g., Months or Years).
3. **Values:** The actual data you want to measure (e.g., Total Salary, Total Sales, or Count of Students).
4. **Filters:** A way to isolate specific data (e.g., showing only the "HR Department" data).

## 3. What are its Main Uses?

Pivot tables are used whenever you have a large dataset (hundreds or thousands of rows) and need quick answers.

- **Summarizing Large Data:** Instantly turn 5,000 rows of transactions into a small table showing total revenue per store.
- **Data Analysis & Trends:** Identify which products are growing in popularity over time or which regions are underperforming.
- **Grouping & Categorizing:** Automatically group dates into months or quarters, or group ages into brackets (e.g., 20–30, 31–40).
- **Comparing Datasets:** Side-by-side comparisons of "Actual vs. Budget" or "This Year vs. Last Year."
- **Cleaning & Auditing:** Quickly spot duplicates or missing values in a dataset by looking at the "Count" of unique entries.

*Using PivotTables to answer questions*

**Consider the example below. Let's say we wanted to answer the question**

What is the amount sold by each salesperson? Answering it could be time consuming and difficult; each salesperson appears on multiple rows, and we would need to total all of their different orders individually. We could use the Subtotal command to help find the total for each salesperson, but we would still have a lot of data to work with.

|    | A                | B      | C       | D            | E        |
|----|------------------|--------|---------|--------------|----------|
| 1  | Salesperson      | Region | Account | Order Amount | Month    |
| 2  | Albertson, Kathy | East   | 29386   | \$925.00     | January  |
| 3  | Albertson, Kathy | East   | 74830   | \$875.00     | February |
| 4  | Albertson, Kathy | East   | 90099   | \$500.00     | February |
| 5  | Albertson, Kathy | East   | 74830   | \$350.00     | March    |
| 6  | Brennan, Michael | West   | 82853   | \$400.00     | January  |
| 7  | Brennan, Michael | West   | 72949   | \$850.00     | January  |
| 8  | Brennan, Michael | West   | 90044   | \$1,500.00   | January  |
| 9  | Brennan, Michael | West   | 82853   | \$550.00     | February |
| 10 | Brennan, Michael | West   | 72949   | \$400.00     | March    |
| 11 | Davis, William   | South  | 55223   | \$235.00     | February |
| 12 | Davis, William   | South  | 10354   | \$850.00     | January  |
| 13 | Davis, William   | South  | 50192   | \$600.00     | March    |
| 14 | Davis, William   | South  | 27589   | \$250.00     | January  |
| 15 | Dumlao, Richard  | West   | 67275   | \$400.00     | January  |
| 16 | Dumlao, Richard  | West   | 41828   | \$965.00     | February |

Fortunately, a PivotTable can instantly calculate and summarize the data in a way that will make it much easier to read. When we're done, the PivotTable will look something like this:

| Row Labels         | Sum of Order Amount |
|--------------------|---------------------|
| Albertson, Kathy   | \$2,650.00          |
| Brennan, Michael   | \$3,700.00          |
| Davis, William     | \$1,935.00          |
| Dumlao, Richard    | \$1,490.00          |
| Flores, Tia        | \$4,565.00          |
| Post, Melissa      | \$1,690.00          |
| Thompson, Shannon  | \$3,160.00          |
| Walters, Chris     | \$4,375.00          |
| <b>Grand Total</b> | <b>\$23,565.00</b>  |

Once you've created a PivotTable, you can use it to answer different questions by rearranging—or pivoting—the data. For example, let's say we wanted to answer What

is the total amount sold in each month? We could modify our PivotTable to look like this:

| Row Labels         | Sum of Order Amount |
|--------------------|---------------------|
| January            | \$9,090.00          |
| February           | \$9,160.00          |
| March              | \$5,315.00          |
| <b>Grand Total</b> | <b>\$23,565.00</b>  |

*To create a PivotTable:* Select the table or cells (including column headers) you want to include in your PivotTable.

|    | A                | B      | C       | D            | E        |
|----|------------------|--------|---------|--------------|----------|
| 1  | Salesperson      | Region | Account | Order Amount | Month    |
| 2  | Albertson, Kathy | East   | 29386   | \$925.00     | January  |
| 3  | Albertson, Kathy | East   | 74830   | \$875.00     | February |
| 4  | Albertson, Kathy | East   | 90099   | \$500.00     | February |
| 5  | Albertson, Kathy | East   | 74830   | \$350.00     | March    |
| 6  | Brennan, Michael | West   | 82853   | \$400.00     | January  |
| 7  | Brennan, Michael | West   | 72949   | \$850.00     | January  |
| 8  | Brennan, Michael | West   | 90044   | \$1,500.00   | January  |
| 9  | Brennan, Michael | West   | 82853   | \$550.00     | February |
| 10 | Brennan, Michael | West   | 72949   | \$400.00     | March    |
| 11 | Davis, William   | South  | 55223   | \$235.00     | February |
| 12 | Davis, William   | South  | 10354   | \$850.00     | January  |
| 13 | Davis, William   | South  | 50192   | \$600.00     | March    |
| 14 | Davis, William   | South  | 27589   | \$250.00     | January  |
| 15 | Dumlao, Richard  | West   | 67275   | \$400.00     | January  |
| 16 | Dumlao, Richard  | West   | 41828   | \$965.00     | February |

From the Insert tab, click the PivotTable command.

The screenshot shows the Microsoft Excel ribbon with the 'Insert' tab selected. The 'PivotTable' command is highlighted with a mouse cursor. A tooltip for the 'PivotTable' command is displayed, providing instructions: 'Easily arrange and summarize complex data in a PivotTable.' and a note: 'FYI: You can double-click a value to see which detailed values make up the summarized total.'

Below the ribbon, a small data table is visible, showing the following data:

|  | B      | C       | D            | E        |
|--|--------|---------|--------------|----------|
|  | Region | Account | Order Amount | Month    |
|  | East   | 29386   | \$925.00     | January  |
|  | East   | 74830   | \$875.00     | February |

source data and place the PivotTable in a new worksheet.

Create PivotTable

Choose the data that you want to analyze

Select a table or range

Table/Range: Table1

Use an external data source

Choose Connection...

Connection name:

Use this workbook's Data Model

Choose where you want the PivotTable report to be placed

New Worksheet

Existing Worksheet

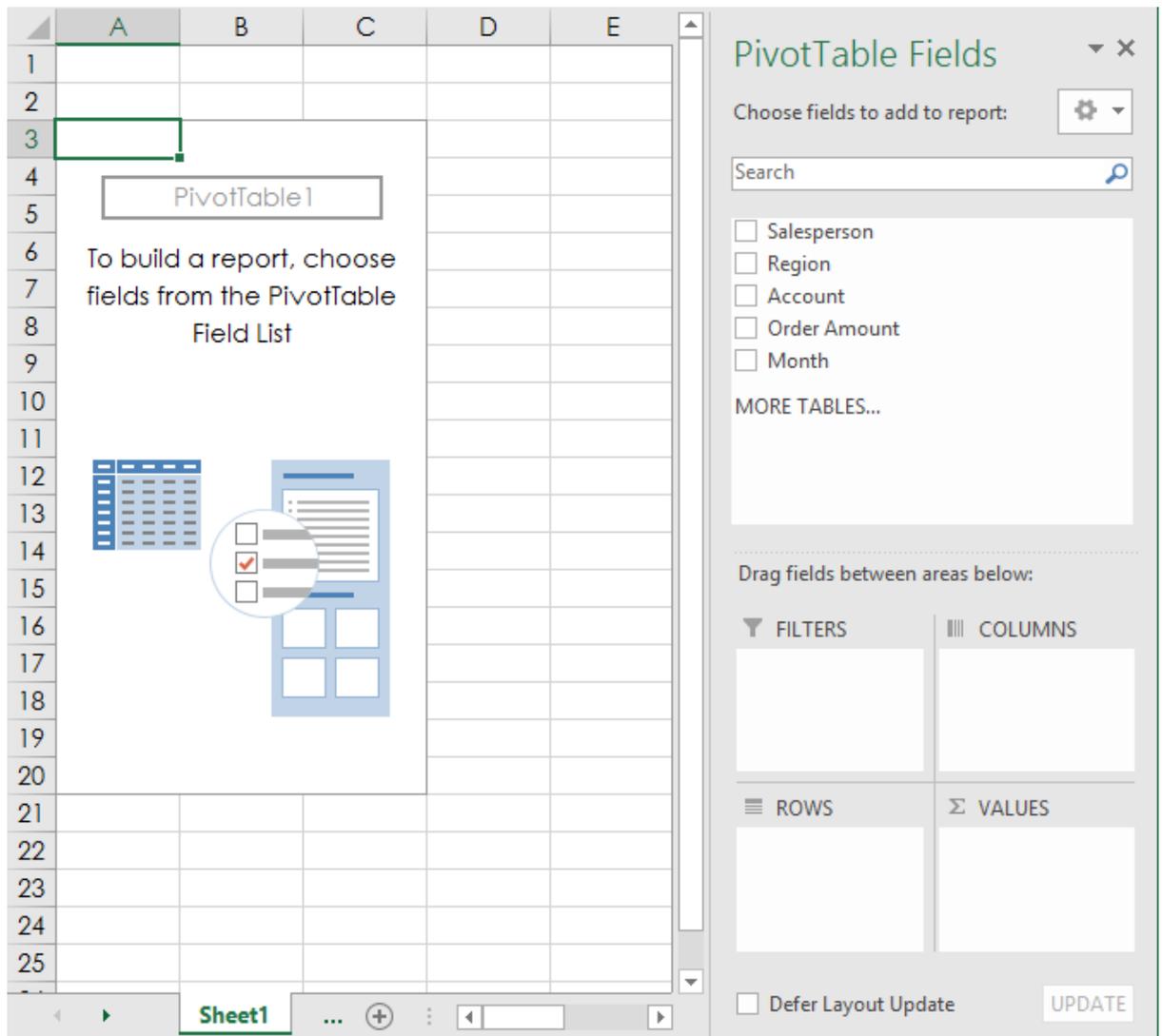
Location:

Choose whether you want to analyze multiple tables

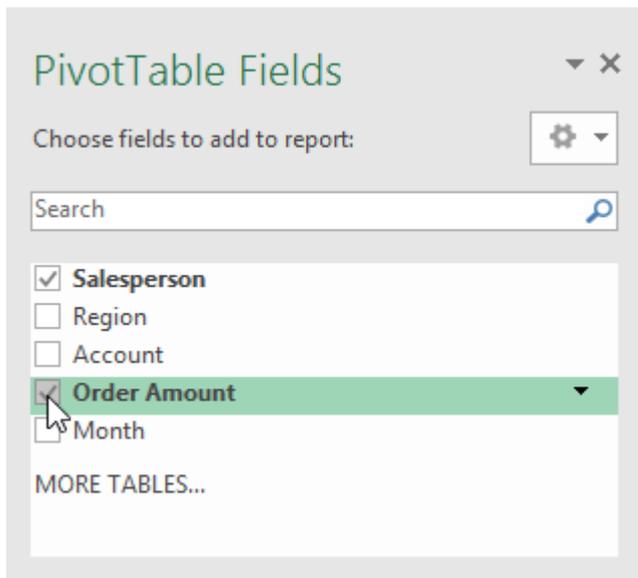
Add this data to the Data Model

OK Cancel

**A blank PivotTable and Field List will appear in a new worksheet.**

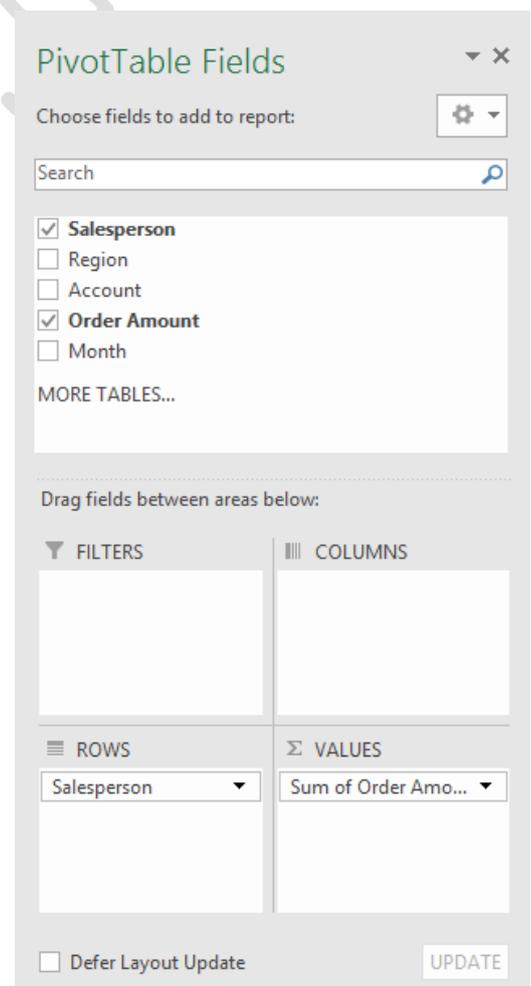


**Once you create a PivotTable, you'll need to decide which fields to add. Each field is simply a column header from the source data. In the PivotTable Fields list, check the box for each field you want to add. In our example, we want to know the total amount sold by each salesperson, so we'll check the Salesperson and Order Amount fields.**



The selected fields will be added to one of the four areas below. In our example, the Salesperson field has been added to the Rows area, while Order Amount has been added to Values. You can also drag and drop fields directly into the desired area.

The PivotTable will calculate and summarize the selected fields. In our example, the PivotTable shows the amount sold by each salesperson.



The screenshot shows a PivotTable in Excel with the following data:

| Row Labels         | Sum of Order Amount |
|--------------------|---------------------|
| Albertson, Kathy   | 2650                |
| Brennan, Michael   | 3700                |
| Davis, William     | 1935                |
| Dumlao, Richard    | 1490                |
| Flores, Tia        | 4565                |
| Post, Melissa      | 1690                |
| Thompson, Shannon  | 3160                |
| Walters, Chris     | 4375                |
| <b>Grand Total</b> | <b>23565</b>        |

The PivotTable Fields task pane on the right shows the following configuration:

- Choose fields to add to report:
  - Salesperson
  - Region
  - Account
  - Order Amount
  - Month
- Drag fields between areas below:
  - FILTERS:** (Empty)
  - COLUMNS:** (Empty)
  - ROWS:** Salesperson
  - VALUES:** Sum of Order Amo...

Just like with normal spreadsheets, you can sort the data in a PivotTable using the Sort & Filter command on the Home tab. You can also apply any type of number formatting you want. For example, you may want to change the number format to Currency. However, be aware that some types of formatting may disappear when you modify the PivotTable.

| Row Labels         | Sum of Order Amount |
|--------------------|---------------------|
| Flores, Tia        | \$4,565.00          |
| Walters, Chris     | \$4,375.00          |
| Brennan, Michael   | \$3,700.00          |
| Thompson, Shannon  | \$3,160.00          |
| Albertson, Kathy   | \$2,650.00          |
| Davis, William     | \$1,935.00          |
| Post, Melissa      | \$1,690.00          |
| Dumlao, Richard    | \$1,490.00          |
| <b>Grand Total</b> | <b>\$23,565.00</b>  |

**If you change any of the data in your source worksheet, the PivotTable will not update automatically. To manually update it, select the PivotTable and then go to Analyze > Refresh.**

### **Pivoting data**

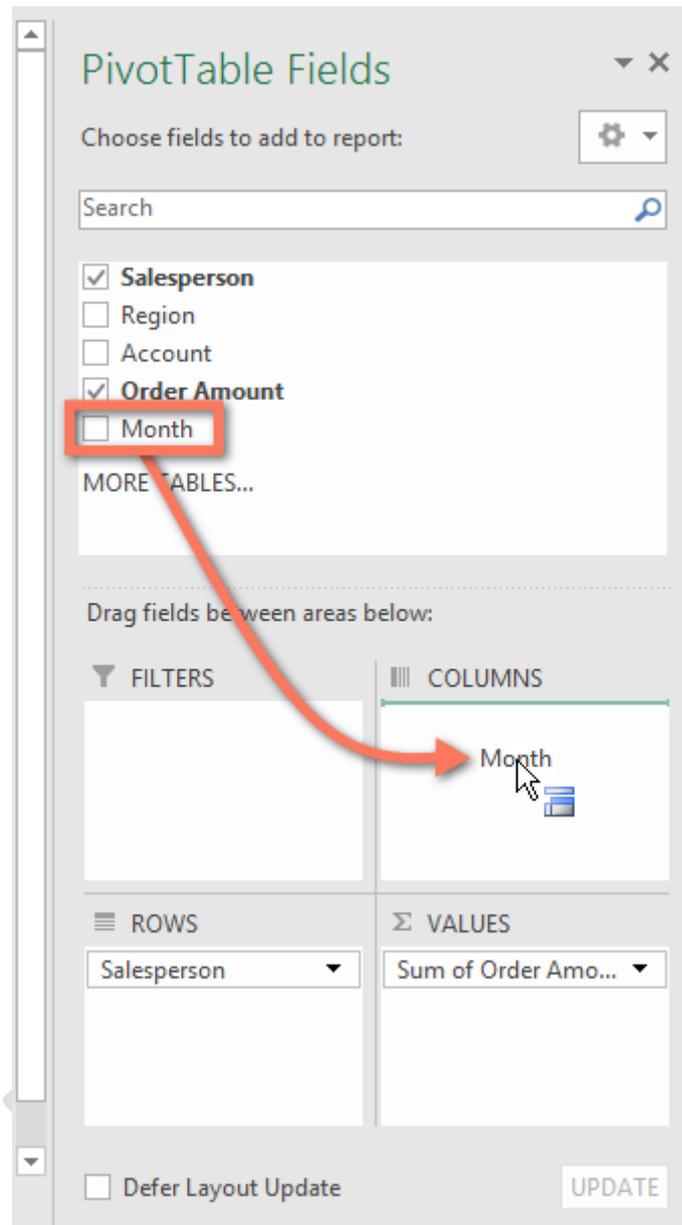
**One of the best things about PivotTables is that they can quickly pivot—or reorganize—your data, allowing you to examine your worksheet in several ways. Pivoting data can help you answer different questions and even experiment with your data to discover new trends and patterns.**

#### *To add columns:*

**So far, our PivotTable has only shown one column of data at a time. To show multiple columns, you'll need to add a field to the Columns area.**

Dr. Rudri & Di.

**Drag a field from the Field List into the Columns area. In our example, we'll use the Month field.**



**The PivotTable will include multiple columns. In our example, there is now a column for each person's monthly sales, in addition to the grand total.**

|    | A                          | B              | C               | D            | E                  |
|----|----------------------------|----------------|-----------------|--------------|--------------------|
| 1  |                            |                |                 |              |                    |
| 2  |                            |                |                 |              |                    |
| 3  | <b>Sum of Order Amount</b> | <b>Column</b>  |                 |              |                    |
| 4  | <b>Row Labels</b>          | <b>January</b> | <b>February</b> | <b>March</b> | <b>Grand Total</b> |
| 5  | Flores, Tia                | 1655           | 985             | 1925         | 4565               |
| 6  | Walters, Chris             | 355            | 2755            | 1265         | 4375               |
| 7  | Brennan, Michael           | 2750           | 550             | 400          | 3700               |
| 8  | Thompson, Shannon          | 1140           | 1720            | 300          | 3160               |
| 9  | Albertson, Kathy           | 925            | 1375            | 350          | 2650               |
| 10 | Davis, William             | 1100           | 235             | 600          | 1935               |
| 11 | Post, Melissa              | 765            | 575             | 350          | 1690               |
| 12 | Dumlao, Richard            | 400            | 965             | 125          | 1490               |
| 13 | <b>Grand Total</b>         | <b>9090</b>    | <b>9160</b>     | <b>5315</b>  | <b>23565</b>       |
| 14 |                            |                |                 |              |                    |
| 15 |                            |                |                 |              |                    |
| 16 |                            |                |                 |              |                    |
| 17 |                            |                |                 |              |                    |
| 18 |                            |                |                 |              |                    |
| 19 |                            |                |                 |              |                    |
| 20 |                            |                |                 |              |                    |
| 21 |                            |                |                 |              |                    |
| 22 |                            |                |                 |              |                    |
| 23 |                            |                |                 |              |                    |
| 24 |                            |                |                 |              |                    |

### PivotTable Fields

Choose fields to add to report: ⚙️

Search 🔍

- Salesperson
- Region
- Account
- Order Amount
- Month

MORE TABLES...

---

Drag fields between areas below:

|   |  |
|---|--|
| <p style="margin: 0; font-size: small;">🔽 FILTERS</p> <div style="border: 1px solid #ccc; height: 40px; margin-top: 5px;"></div>  | <p style="margin: 0; font-size: small;">📊 COLUMNS</p> <div style="border: 1px solid #ccc; padding: 2px; margin-top: 5px;">             Month <span style="float: right;">▼</span> </div>           |
| <p style="margin: 0; font-size: small;">📋 ROWS</p> <div style="border: 1px solid #ccc; padding: 2px; margin-top: 5px;">             Salesperson <span style="float: right;">▼</span> </div> | <p style="margin: 0; font-size: small;">Σ VALUES</p> <div style="border: 1px solid #ccc; padding: 2px; margin-top: 5px;">             Sum of Order ... <span style="float: right;">▼</span> </div> |

*To change a row or column:*

**Changing a row or column can give you a completely different perspective on your data. All you have to do is remove the field in question, then replace it with another.**

**Drag the field you want to remove out of its current area. You can also uncheck the appropriate box in the Field List. In this example, we've removed the Month and Salesperson fields.**

**PivotTable Fields**

Choose fields to add to report: [Settings]

Search [Search]

- Salesperson
- Region
- Account
- Order Amount
- Month

MORE TABLES...

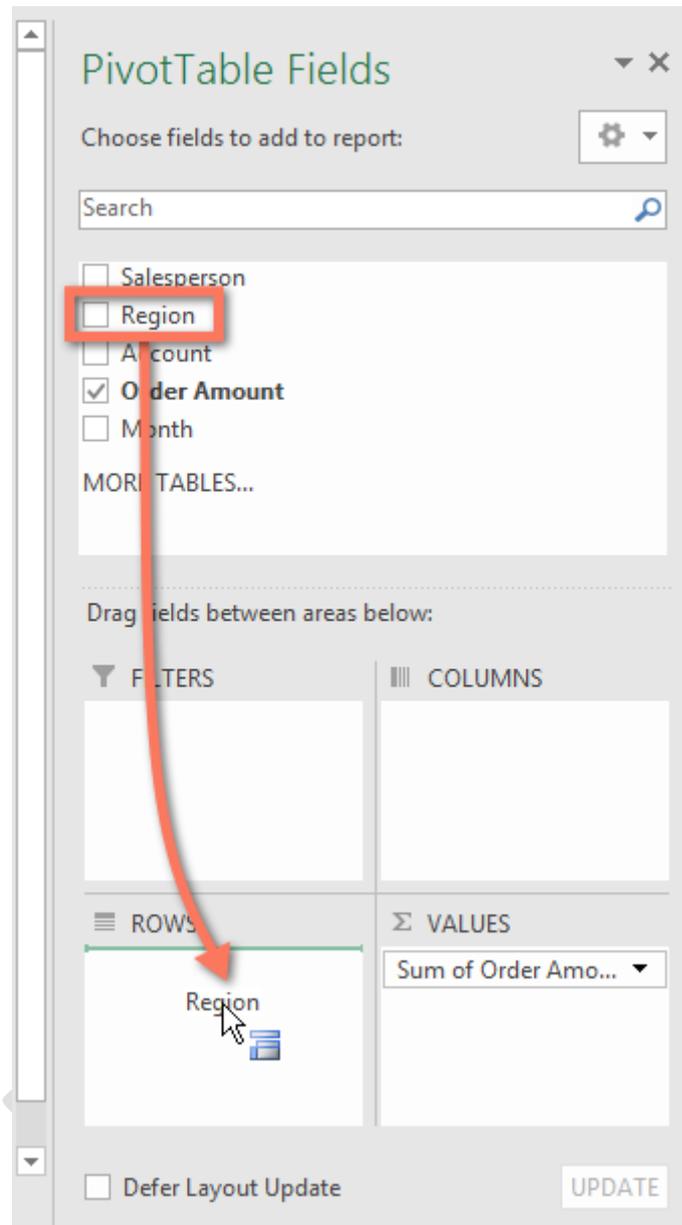
Drag fields between areas below:

| FILTERS | COLUMNS |
|---------|---------|
|         |         |

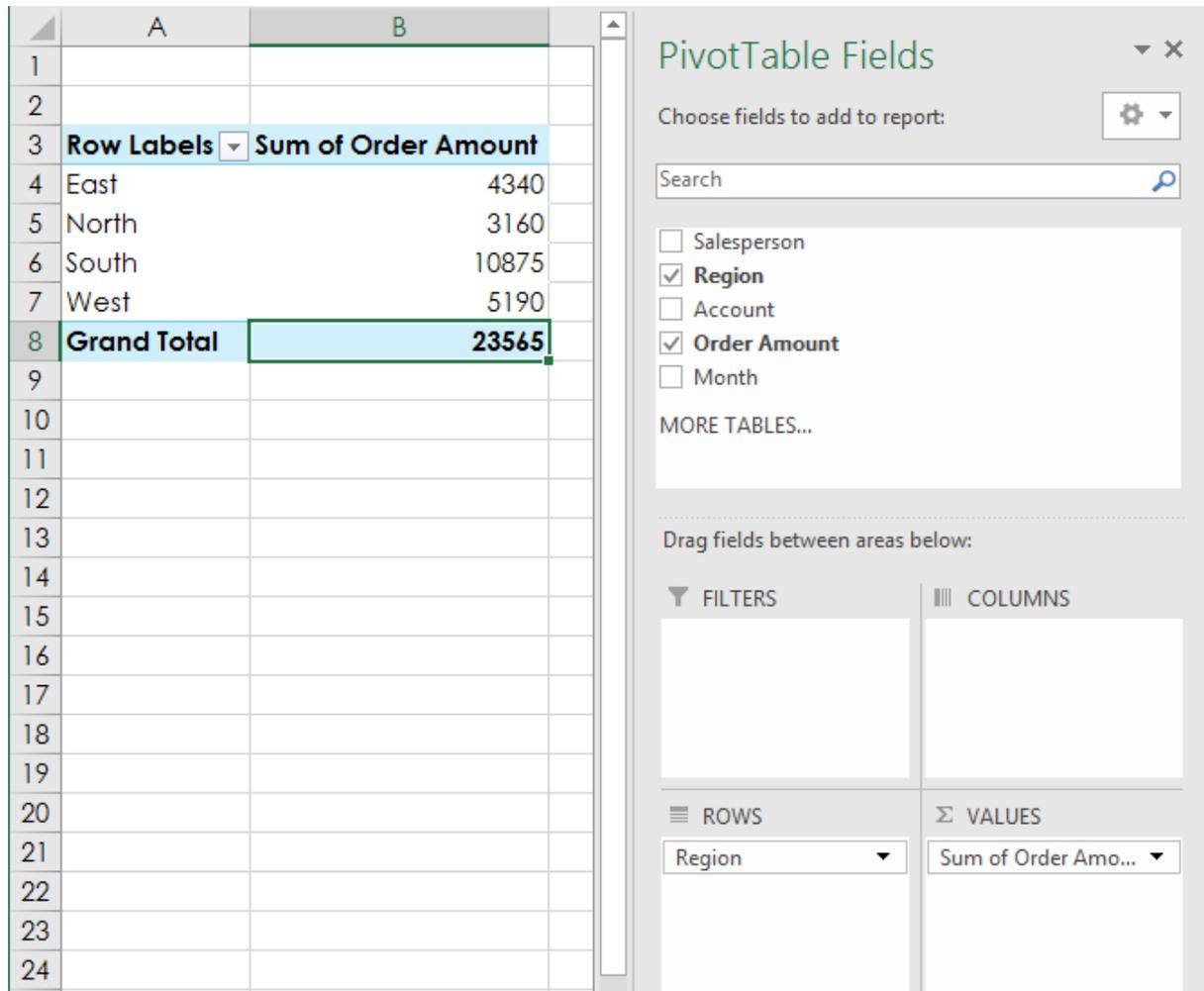
| ROWS        | VALUES              |
|-------------|---------------------|
| Salesperson | Sum of Order Amo... |

Defer Layout Update [UPDATE]

**Drag a new field into the desired area. In our example, we'll place the Region field under Rows.**



The PivotTable will adjust—or pivot—to show the new data. In our example, it now shows the amount sold by each region.



|    | A           | B                   |
|----|-------------|---------------------|
| 1  |             |                     |
| 2  |             |                     |
| 3  | Row Labels  | Sum of Order Amount |
| 4  | East        | 4340                |
| 5  | North       | 3160                |
| 6  | South       | 10875               |
| 7  | West        | 5190                |
| 8  | Grand Total | 23565               |
| 9  |             |                     |
| 10 |             |                     |
| 11 |             |                     |
| 12 |             |                     |
| 13 |             |                     |
| 14 |             |                     |
| 15 |             |                     |
| 16 |             |                     |
| 17 |             |                     |
| 18 |             |                     |
| 19 |             |                     |
| 20 |             |                     |
| 21 |             |                     |
| 22 |             |                     |
| 23 |             |                     |
| 24 |             |                     |

**PivotTable Fields**

Choose fields to add to report:

Search

- Salesperson
- Region
- Account
- Order Amount
- Month

MORE TABLES...

Drag fields between areas below:

**FILTERS**

**COLUMNS**

**ROWS**: Region

**VALUES**: Sum of Order Amo...

*To learn more:*

Once you're comfortable with PivotTables, review our [Doing More with PivotTables](#) lesson for additional ways to customize and manipulate data.

### Challenge!

Open our [practice workbook](#).

Create a PivotTable in a separate sheet.

We want to answer the question **What is the total amount sold in each region?** To do this, select **Region** and **Order Amount**. When

you're finished, your workbook should look like this:

| Row Labels         | Sum of Order Amount |
|--------------------|---------------------|
| East               | 4340                |
| North              | 3160                |
| South              | 10875               |
| West               | 5190                |
| <b>Grand Total</b> | <b>23565</b>        |

In the Rows area, remove Region and replace it with Salesperson.

Add Month to the Columns area.

Change the number format of cells B5:E13 to Currency. Note:

You might have to make columns C and D wider to see the values.

When you're finished, your workbook should look like this:

The screenshot shows an Excel PivotTable with the following data:

| Row Labels         | January           | February          | March             | Grand Total        |
|--------------------|-------------------|-------------------|-------------------|--------------------|
| Albertson, Kathy   | \$925.00          | \$1,375.00        | \$350.00          | \$2,650.00         |
| Brennan, Michael   | \$2,750.00        | \$550.00          | \$400.00          | \$3,700.00         |
| Davis, William     | \$1,100.00        | \$235.00          | \$600.00          | \$1,935.00         |
| Dumlao, Richard    | \$400.00          | \$965.00          | \$125.00          | \$1,490.00         |
| Flores, Tia        | \$1,655.00        | \$985.00          | \$1,925.00        | \$4,565.00         |
| Post, Melissa      | \$765.00          | \$575.00          | \$350.00          | \$1,690.00         |
| Thompson, Shannon  | \$1,140.00        | \$1,720.00        | \$300.00          | \$3,160.00         |
| Walters, Chris     | \$355.00          | \$2,755.00        | \$1,265.00        | \$4,375.00         |
| <b>Grand Total</b> | <b>\$9,090.00</b> | <b>\$9,160.00</b> | <b>\$5,315.00</b> | <b>\$23,565.00</b> |

The PivotTable Fields task pane on the right shows the following configuration:

- Choose fields to add to report:** Salesperson, Order Amount, Month (checked).
- Drag fields between areas below:**
  - FILTERS:** (Empty)
  - COLUMNS:** Month
  - ROWS:** Salespers...
  - VALUES:** Sum of Or...

## Lesson 28: Doing More with PivotTables

### Introduction

As you learned in our previous lesson, PivotTables can be used to summarize and analyze almost any type of data. To manipulate your PivotTable—and gain even more insight into your data—Excel offers three additional tools: filters, slicers, and PivotCharts.

**Optional:** Download our practice workbook.

Watch the video below to learn more about enhancing PivotTables.

### *Filters*

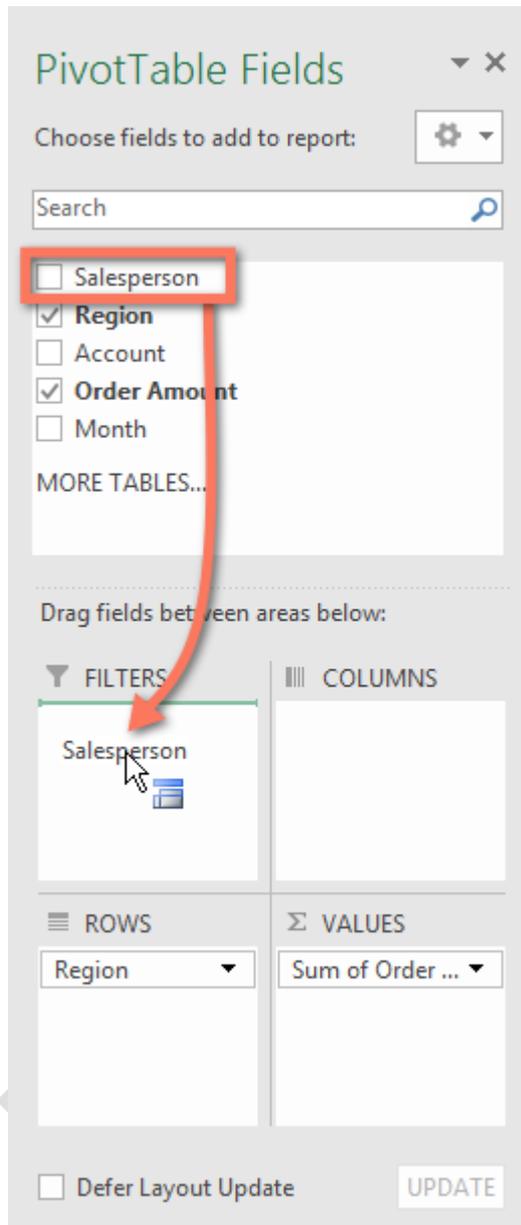
Sometimes you may want focus on a certain section of your data. Filters can be used to narrow down the data in your PivotTable so you can view only the information you need.

*To add a filter:*

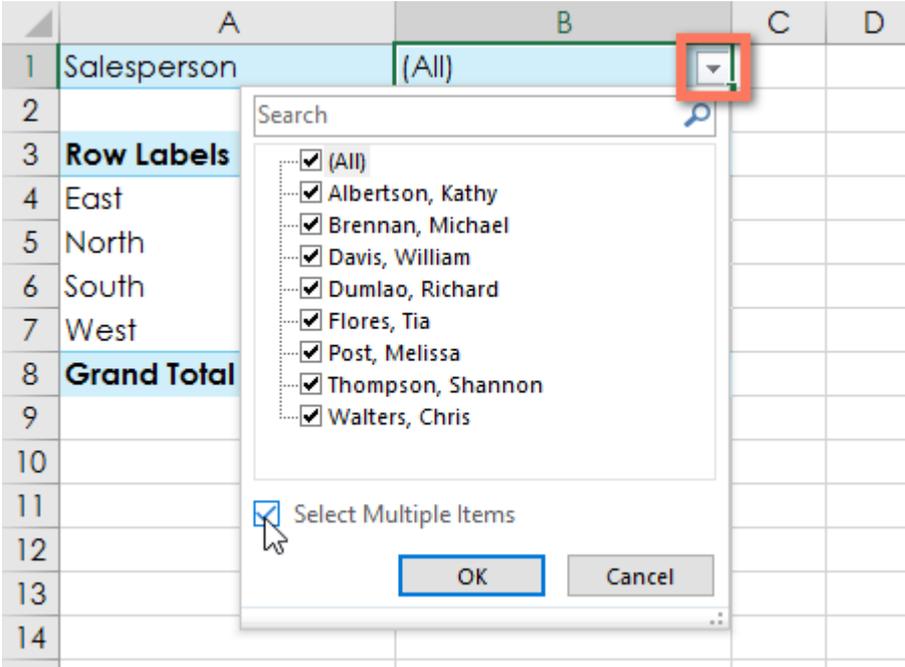
In the example below, we'll filter out certain salespeople to determine how their individual sales are impacting each region.

Dr. Rudri

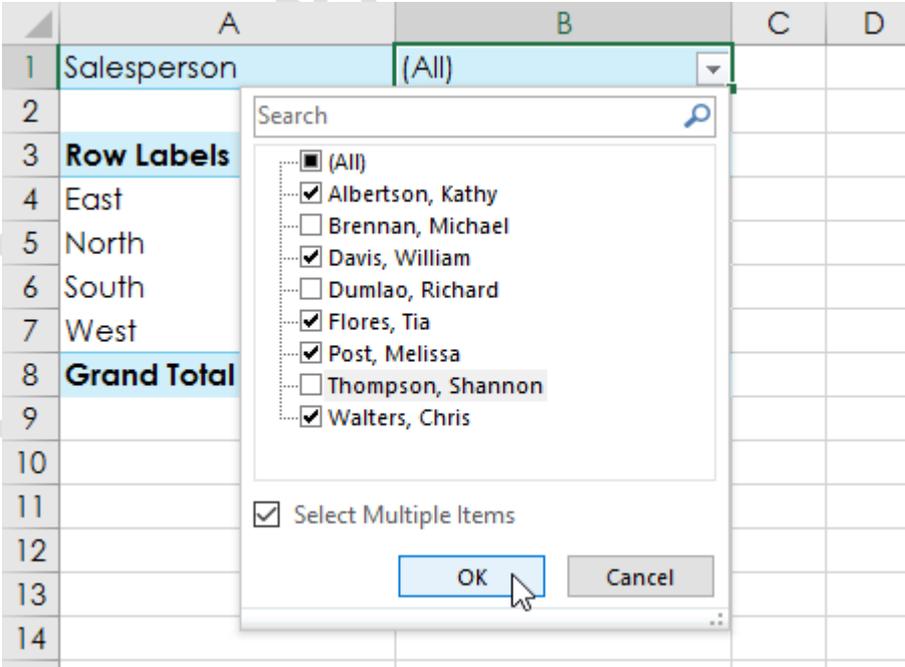
**Drag a field from the Field List to the Filters area. In this example, we'll use the Salesperson field.**



The filter will appear above the PivotTable. Click the drop-down arrow, then check the box next to Select Multiple Items.



Uncheck the box next to any item you don't want to include in the PivotTable. In our example, we'll uncheck the boxes for a few salespeople, then click OK.



The PivotTable will adjust to reflect the changes.

|   | A                  | B                          | C | D |
|---|--------------------|----------------------------|---|---|
| 1 | Salesperson        | (Multiple Items)           |   |   |
| 2 |                    |                            |   |   |
| 3 | <b>Row Labels</b>  | <b>Sum of Order Amount</b> |   |   |
| 4 | East               | 4340                       |   |   |
| 5 | South              | 10875                      |   |   |
| 6 | <b>Grand Total</b> | <b>15215</b>               |   |   |
| 7 |                    |                            |   |   |
| 8 |                    |                            |   |   |

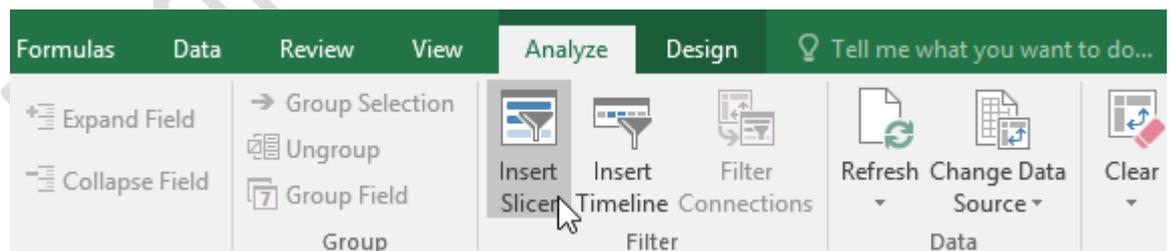
## Slicers

Slicers make filtering data in PivotTables even easier. Slicers are basically just filters but are easier and faster to use, allowing you to instantly pivot your data. If you frequently filter your PivotTables, you may want to consider using slicers instead of filters.

*To add a slicer:*

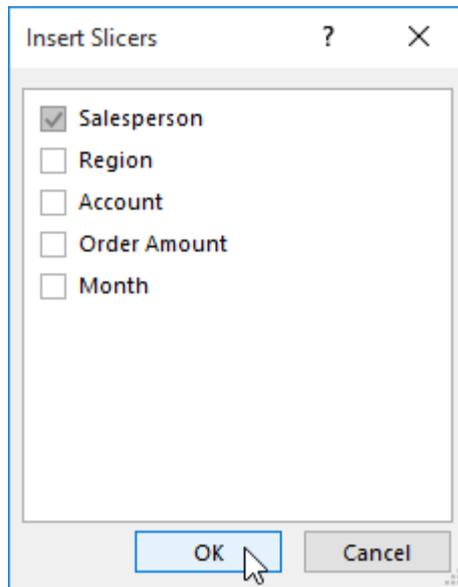
Select any cell in the PivotTable.

From the Analyze tab, click the Insert Slicer command.



A dialog box will appear. Check the box next to the desired field.

In our example, we'll select Salesperson, then click OK.



The slicer will appear next to the PivotTable. Each selected item will be highlighted in blue. In the example below, the slicer contains all eight salespeople, but only five of them are currently selected.

|    | A                  | B                   | C | D | E | F | G | H |
|----|--------------------|---------------------|---|---|---|---|---|---|
| 1  | Salesperson        | (Multiple Items)    |   |   |   |   |   |   |
| 2  |                    |                     |   |   |   |   |   |   |
| 3  | Row Labels         | Sum of Order Amount |   |   |   |   |   |   |
| 4  | East               | 4340                |   |   |   |   |   |   |
| 5  | South              | 10875               |   |   |   |   |   |   |
| 6  | <b>Grand Total</b> | <b>15215</b>        |   |   |   |   |   |   |
| 7  |                    |                     |   |   |   |   |   |   |
| 8  |                    |                     |   |   |   |   |   |   |
| 9  |                    |                     |   |   |   |   |   |   |
| 10 |                    |                     |   |   |   |   |   |   |
| 11 |                    |                     |   |   |   |   |   |   |
| 12 |                    |                     |   |   |   |   |   |   |
| 13 |                    |                     |   |   |   |   |   |   |
| 14 |                    |                     |   |   |   |   |   |   |
| 15 |                    |                     |   |   |   |   |   |   |

| Salesperson       |  |
|-------------------|--|
| Albertson, Kathy  |  |
| Brennan, Michael  |  |
| Davis, William    |  |
| Dumlao, Richard   |  |
| Flores, Tia       |  |
| Post, Melissa     |  |
| Thompson, Shannon |  |
| Walters, Chris    |  |

Just like filters, only selected items are used in the PivotTable.

When you select or deselect an item, the PivotTable will instantly reflect the change. Try selecting different items to see how they affect the PivotTable. Press and hold the Ctrl key on your

keyboard to select multiple items at once.

|    | A                  | B                   | C | D | E | F | G | H |
|----|--------------------|---------------------|---|---|---|---|---|---|
| 1  | Salesperson        | (Multiple Items)    |   |   |   |   |   |   |
| 2  |                    |                     |   |   |   |   |   |   |
| 3  | Row Labels         | Sum of Order Amount |   |   |   |   |   |   |
| 4  | North              | 3160                |   |   |   |   |   |   |
| 5  | South              | 10875               |   |   |   |   |   |   |
| 6  | West               | 5190                |   |   |   |   |   |   |
| 7  | <b>Grand Total</b> | <b>19225</b>        |   |   |   |   |   |   |
| 8  |                    |                     |   |   |   |   |   |   |
| 9  |                    |                     |   |   |   |   |   |   |
| 10 |                    |                     |   |   |   |   |   |   |
| 11 |                    |                     |   |   |   |   |   |   |
| 12 |                    |                     |   |   |   |   |   |   |
| 13 |                    |                     |   |   |   |   |   |   |
| 14 |                    |                     |   |   |   |   |   |   |
| 15 |                    |                     |   |   |   |   |   |   |

The Slicer for 'Salesperson' is open, showing a list of names: Albertson, Kathy; Brennan, Michael; Davis, William; Dumlao, Richard; Flores, Tia; Post, Melissa; Thompson, Shannon; and Walters, Chris. The name 'Walters, Chris' is highlighted in yellow, and a mouse cursor is pointing at it.

You can also click the Filter icon in the top-right corner of the slicer to select all items at once.

## PivotCharts

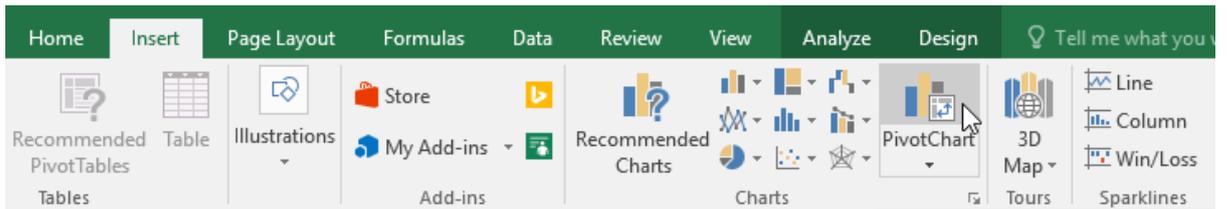
PivotCharts are like regular charts, except they display data from a PivotTable. Just like regular charts, you'll be able to select a chart type, layout, and style that will best represent the data.

*To create a PivotChart:*

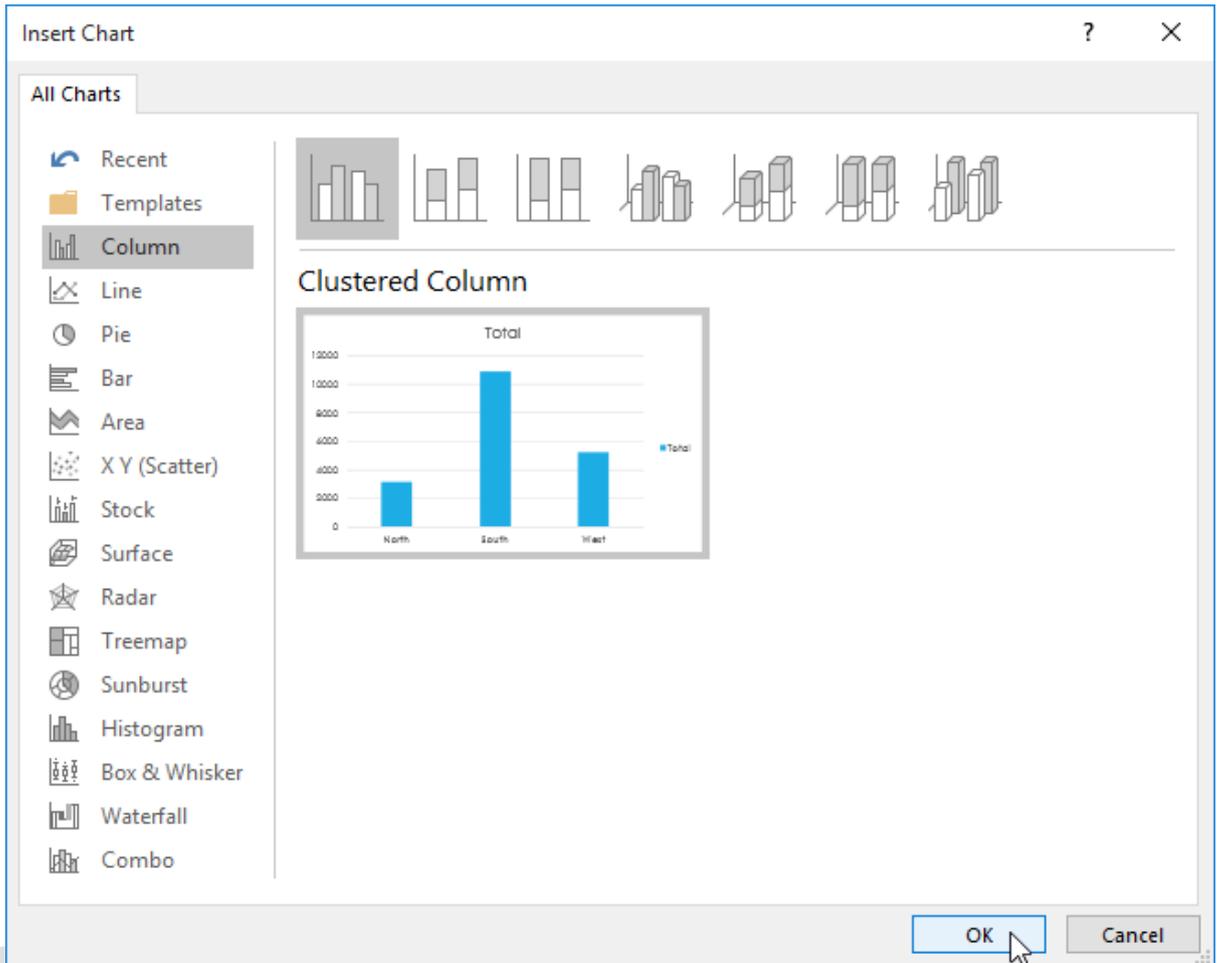
In the example below, our PivotTable is showing a portion of each region's sales figures. We'll use a PivotChart so we can see the information more clearly.

Select any cell in your PivotTable.

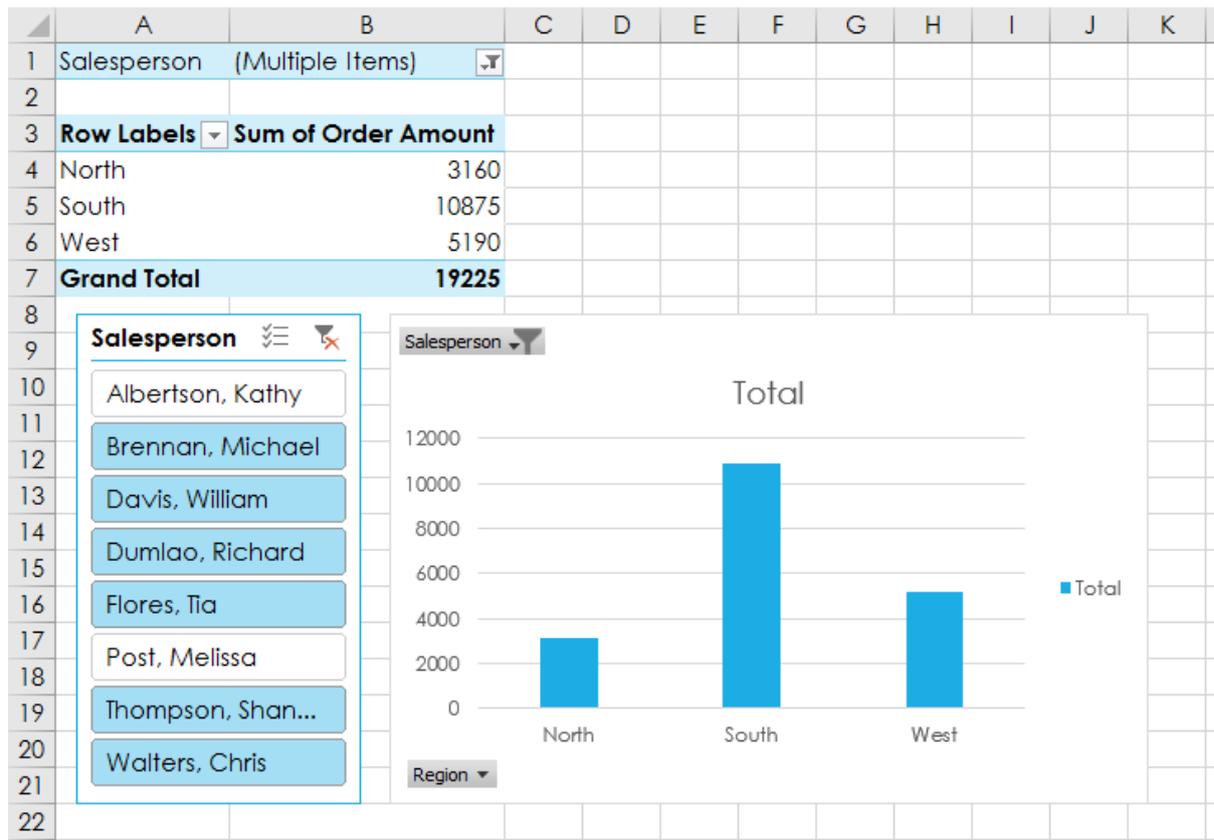
**From the Insert tab, click the PivotChart command.**



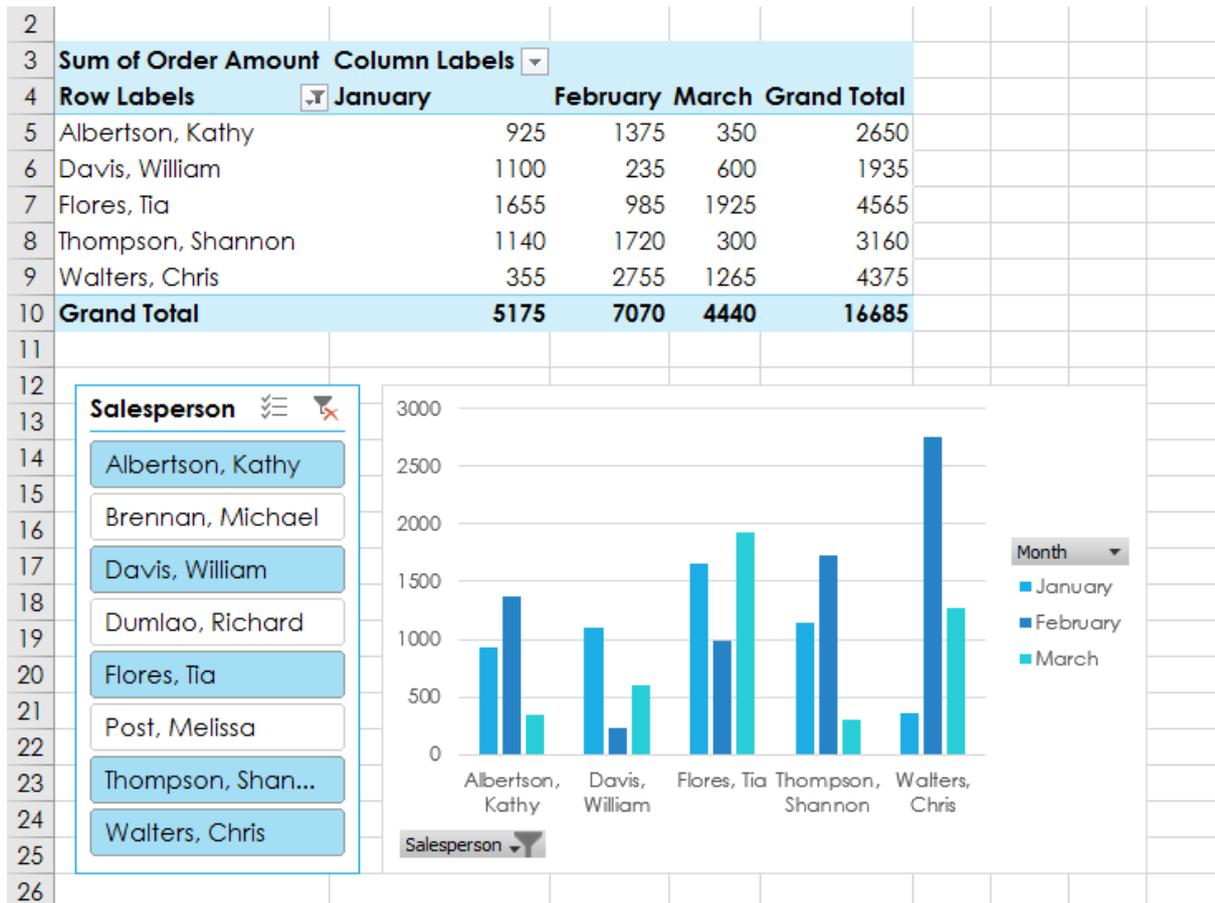
**The Insert Chart dialog box will appear. Select the desired chart type and layout, then click OK.**



**The PivotChart will appear.**



**Try using filters or slicers to narrow down the data in your PivotChart. To view different subsets of information, change the columns or rows in your PivotTable. In the example below, we've changed the PivotTable to view the monthly sales for each salesperson.**



### Challenge!

Open our [practice workbook](#).

In the Rows area, remove Region and replace it with Salesperson.

Insert a PivotChart and choose the type Line with Markers.

Insert a slicer for Regions.

Use the slicer to only show the South and East regions.

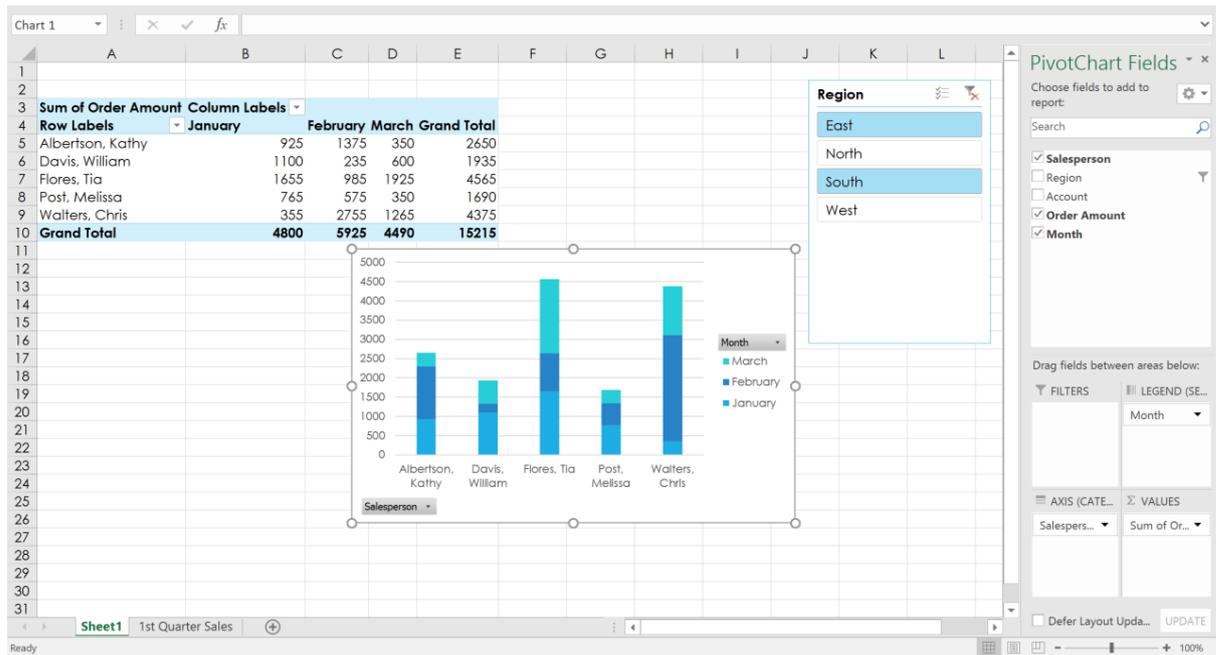
Change the PivotChart type to Stacked Column.

In the PivotChart Fields pane to the right, add Month to the

Legend (Series) area. Note: You can also click the PivotTable and

add Month to the Columns area for the same results.

When you're finished, your workbook should look something like this:



## 2.2. Descriptive Statistics

Descriptive Statistics refers to the branch of statistics concerned with organizing, summarizing, and presenting data in a meaningful way so that its main characteristics can be easily understood.

Descriptive statistics does not involve making predictions or generalizations about a population; rather, it focuses on describing the features of a given dataset.

### ➤ Key Objectives of Descriptive Statistics

- To simplify large volumes of data into a clear and understandable form
- To describe the central tendency, variability, and distribution of data
- To present data systematically using numerical and graphical methods

### ➤ Main Components of Descriptive Statistics

#### 1. Mean

The mean is the average value of a dataset. It is obtained by dividing the sum of all observations by the total number of observations. The mean represents the central value

of the data and is widely used in business, economics, and research. However, it is affected by extreme values (outliers).

- **Formula**

$$\text{Mean} = \frac{\sum X}{N}$$

Where:

- $\sum X$  = Sum of all values
- $N$  = Number of observations

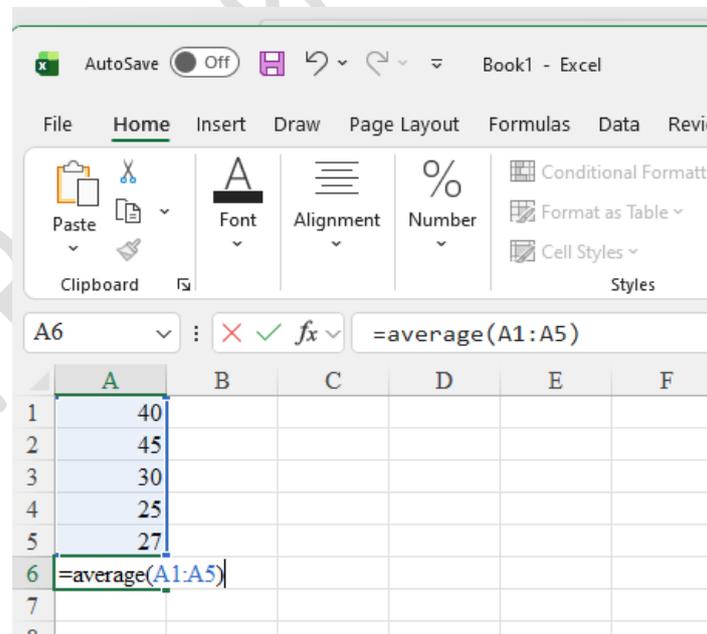
- **Example**

Dataset: **10, 20, 30, 40, 50**

$$\text{Mean} = \frac{10 + 20 + 30 + 40 + 50}{5} = \frac{150}{5} = 30$$

- **Excel Formula**

**=AVERAGE(A1:A5)**



- **Business Use**

- Average sales
- Average cost
- Average marks of students

## 2. Median

The median is the middle value of a dataset when the data is arranged in ascending or descending order. The median is **not affected by extreme values**, making it suitable for income, salary, and skewed data analysis.

- **Example 1 (Odd Number)**

Dataset: **10, 20, 30, 40, 50**

Median = **30**

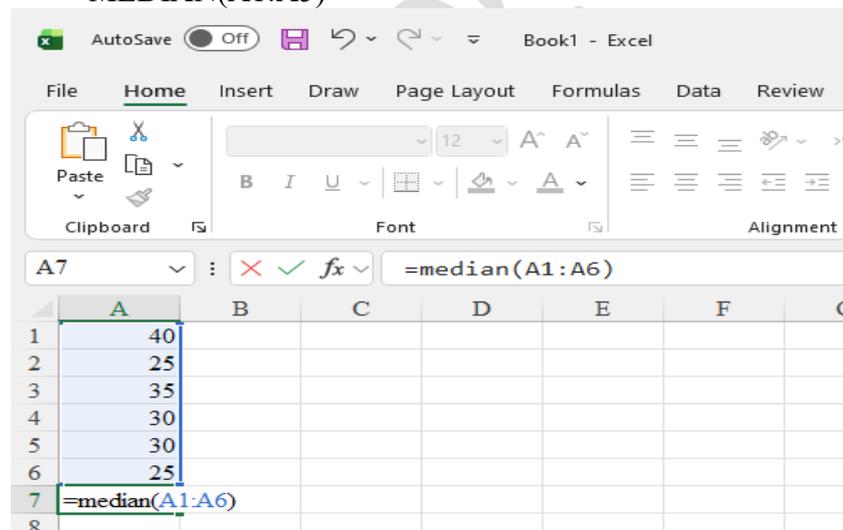
- **Example 2 (Even Number)**

Dataset: **10, 20, 30, 40**

$$\text{Median} = \frac{20 + 30}{2} = 25$$

- **Excel Formula**

=MEDIAN(A1:A5)



- **Business Use**

- Income analysis
- Salary distribution
- Real estate price analysis

## 3. Mode

The mode is the value that occurs most frequently in a dataset.

- **Example**

Data: 2, 4, 6, 6, 6, 8

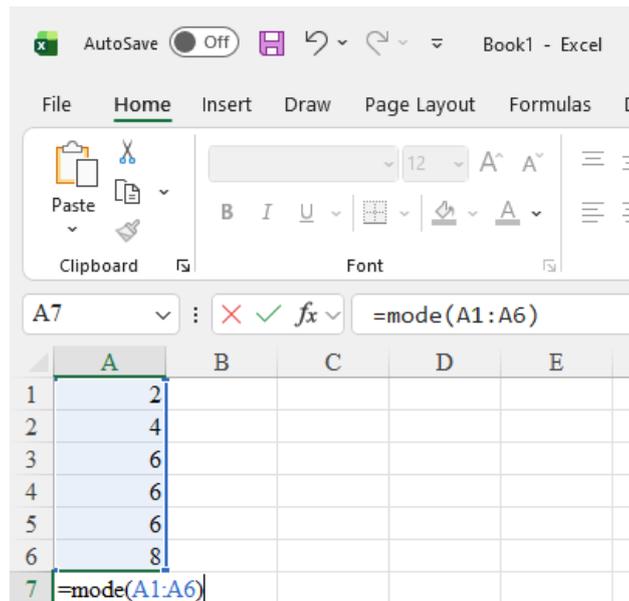
Mode = 6

- **Types of Mode**

- **Unimodal** – one mode
- **Bimodal** – two modes
- **Multimodal** – more than two modes

- **Excel Formula**

=MODE.SNGL(A1:A6)



- **Business Use**

- Most popular product size
- Most common customer age group
- Demand forecasting

#### 4. Range

The **range** measures the **spread of data** by finding the difference between the **maximum and minimum values**.

- **Formula**

Range = Maximum Value – Minimum Value

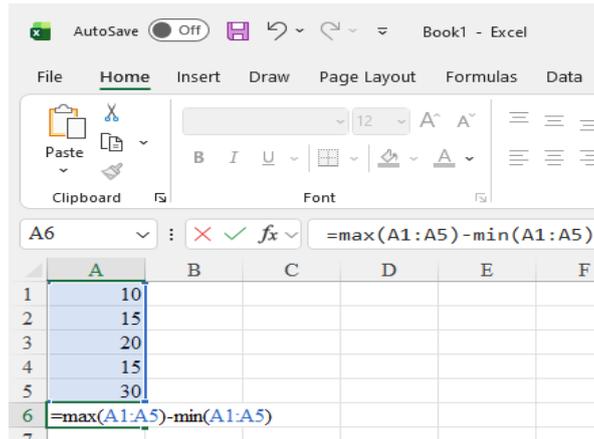
- **Example**

Data: 10, 15, 20, 25, 30

Range = 30 – 10 = 20

- **Excel Formula**

=MAX (A1:A5)-MIN (A1:A5)



- **Limitation**

- Highly affected by extreme values
- Does not show full variability

- **Business Use**

- Price fluctuation analysis
- Stock market volatility (basic level)

## 5. Variance

Variance measures how far each value in the dataset deviates from the mean. It shows the degree of dispersion.

- **Formula (Population Variance)**

$$\sigma^2 = \frac{\sum(X - \bar{X})^2}{n}$$

- **Example**

- Data: 2, 4, 6
- Mean = 4

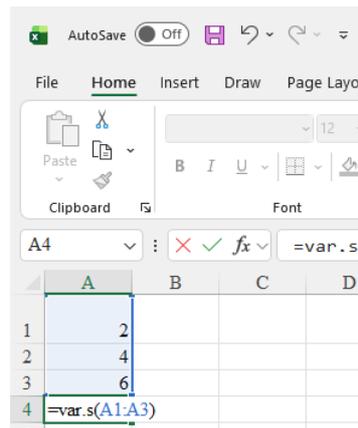
| X | X - Mean | (X - Mean) <sup>2</sup> |
|---|----------|-------------------------|
| 2 | -2       | 4                       |
| 4 | 0        | 0                       |
| 6 | 2        | 4                       |

$$\sigma^2 = \frac{8}{3} = 2.67$$

- **Excel Formula**

- Sample variance:

$$=VAR.S(A1:A3)$$



- **Business Use**

- Risk measurement
- Quality control
- Financial analysis

## 6. Standard Deviation

Standard deviation is the square root of variance. It shows how much data deviates from the mean on average.

- **Formula**

$$\sigma = \sqrt{\sigma^2}$$

- **Example**

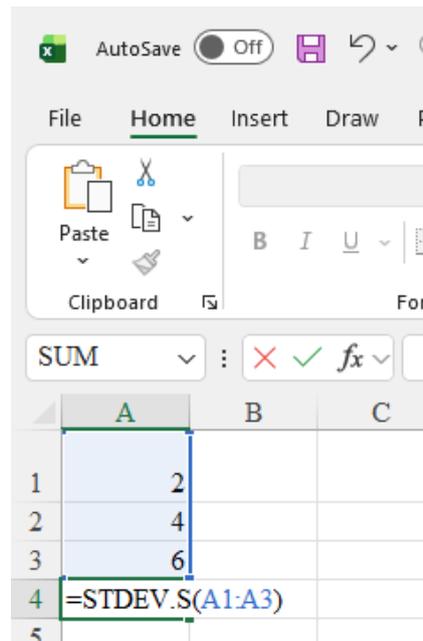
Variance = 2.67

$$\sigma = \sqrt{2.67} = 1.63$$

- **Excel Formula**

- Sample:

=STDEV.S(A1:A3)



- **Interpretation**
  - **Low SD** → Data is closely clustered
  - **High SD** → Data is widely spread
- **Business Use**
  - Market risk analysis
  - Performance consistency
  - Investment decision-making

## 2.3. Other tools of Statistics using MS-Excel:

### A. t-test (Assuming Equal Variance)

A Two-Sample t-Test Assuming Equal Variances is a parametric test used to determine whether there is a significant difference between the means of two independent samples, if both populations have equal variances.

#### ➤ When to Use t-Test (Assuming Equal Variances)

This test is used when:

1. You want to **compare the means of two independent groups**
2. The data is **quantitative (numerical)**
3. Sample sizes are **small ( $n < 30$ )** or population standard deviation is unknown
4. The **variances of the two populations are approximately equal**
5. Data is **approximately normally distributed**

➤ **Examples:**

- Comparing **average marks of two different classes**
- Comparing **average salaries of employees in two departments**
- Comparing **sales of two different regions**

➤ **Basic Assumptions of t-Test (Equal Variance)**

1. **Normality**

Data in both groups should follow a normal distribution.

2. **Independence**

Observations in each sample must be independent.

3. **Equality of Variance (Homoscedasticity)**

The population variances of the two groups are equal.

4. **Scale of Measurement**

Data should be measured on **interval or ratio scale**.

➤ **Null and Alternative Hypothesis**

Let:

$\mu_1$  = Mean of population 1

$\mu_2$  = Mean of population 2

○ **Null Hypothesis ( $H_0$ )**

There is **no significant difference** between the two population means.

$$H_0: \mu_1 = \mu_2$$

○ **Alternative Hypothesis ( $H_1$ )**

There is **a significant difference** between the two population means.

$$H_1: \mu_1 \neq \mu_2$$

(For one-tailed tests,  $H_1$  may be  $\mu_1 > \mu_2$  or  $\mu_1 < \mu_2$ )

➤ **5. How to Run t-Test (Assuming Equal Variance) in Excel**

○ **Excel Formula Used**

$$=T.TEST(A2:A6, B2:B6, 2, 2)$$

### ○ Meaning of Each Argument

- A2:A6 → Sample data for **Group A**
- B2:B6 → Sample data for **Group B**
- 2 (tails) → **Two-tailed test**, used when checking for any difference between means (If a one-tailed test is required, the value **1** should be used.)
- 2 (type) → **Two-sample t-test assuming equal variances**

|    | Group A | Group B |  |  |  |  |  |  |  |  |  |  |  |  |
|----|---------|---------|--|--|--|--|--|--|--|--|--|--|--|--|
| 1  | Group A | Group B |  |  |  |  |  |  |  |  |  |  |  |  |
| 2  | 45      | 48      |  |  |  |  |  |  |  |  |  |  |  |  |
| 3  | 50      | 52      |  |  |  |  |  |  |  |  |  |  |  |  |
| 4  | 47      | 49      |  |  |  |  |  |  |  |  |  |  |  |  |
| 5  | 53      | 55      |  |  |  |  |  |  |  |  |  |  |  |  |
| 6  | 49      | 51      |  |  |  |  |  |  |  |  |  |  |  |  |
| 7  |         |         |  |  |  |  |  |  |  |  |  |  |  |  |
| 8  |         |         |  |  |  |  |  |  |  |  |  |  |  |  |
| 9  |         |         |  |  |  |  |  |  |  |  |  |  |  |  |
| 10 |         |         |  |  |  |  |  |  |  |  |  |  |  |  |
| 11 |         |         |  |  |  |  |  |  |  |  |  |  |  |  |
| 12 |         |         |  |  |  |  |  |  |  |  |  |  |  |  |
| 13 |         |         |  |  |  |  |  |  |  |  |  |  |  |  |
| 14 |         |         |  |  |  |  |  |  |  |  |  |  |  |  |

### ○ Hypotheses Being Tested

- **Null Hypothesis (H<sub>0</sub>):** There is no significant difference between the mean of Group A and Group B
- **Alternative Hypothesis (H<sub>1</sub>):** There is a significant difference between the two means

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

### ○ Interpretation of the Result

- The output of the formula is a **p-value**
- If **p-value** ≤ **0.05**, reject H<sub>0</sub> → significant difference exists
- If **p-value** > **0.05**, accept H<sub>0</sub> → no significant difference

In our example the p-value (0.26308) is greater than the level of significance (0.05), the null hypothesis is accepted. Hence, there is no significant difference between the means of the two groups.

## B. t-test (Assuming No Equal Variance)

A Two-Sample t-Test Assuming Unequal Variances is a parametric statistical test used to determine whether there is a significant difference between the means of two independent samples when the population variances are not equal.

### ➤ When to Use t-Test (Assuming Unequal Variances)

1. You want to **compare the means of two independent groups**
2. The data is **quantitative (numerical)**
3. Population variances of the two groups are **unequal**
4. Sample sizes may be **different**
5. Data is **approximately normally distributed**
6. Population standard deviations are **unknown**

### ➤ Examples:

- Comparing **salaries of permanent and contractual employees**
- Comparing **sales of two regions with different variability**
- Comparing **performance of two departments with unequal dispersion**

### ➤ Basic Assumptions of t-Test (Unequal Variance)

1. **Normality**  
Each sample should be approximately normally distributed.
2. **Independence**  
Observations in one group are independent of the other group.
3. **Unequal Variances Allowed**  
The two populations do **not** need to have equal variances.
4. **Scale of Measurement**  
Data must be measured on an **interval or ratio scale**.

### ➤ Null and Alternative Hypothesis

Let:

$\mu_1$  = Mean of population 1

$\mu_2$  = Mean of population 2

#### ○ Null Hypothesis ( $H_0$ )

There is **no significant difference** between the two population means.

$$H_0: \mu_1 = \mu_2$$

- **Alternative Hypothesis ( $H_1$ )**

There is a **significant difference** between the two population means.

$$H_1: \mu_1 \neq \mu_2$$

(For one-tailed tests,  $H_1$  may be  $\mu_1 > \mu_2$  or  $\mu_1 < \mu_2$ )

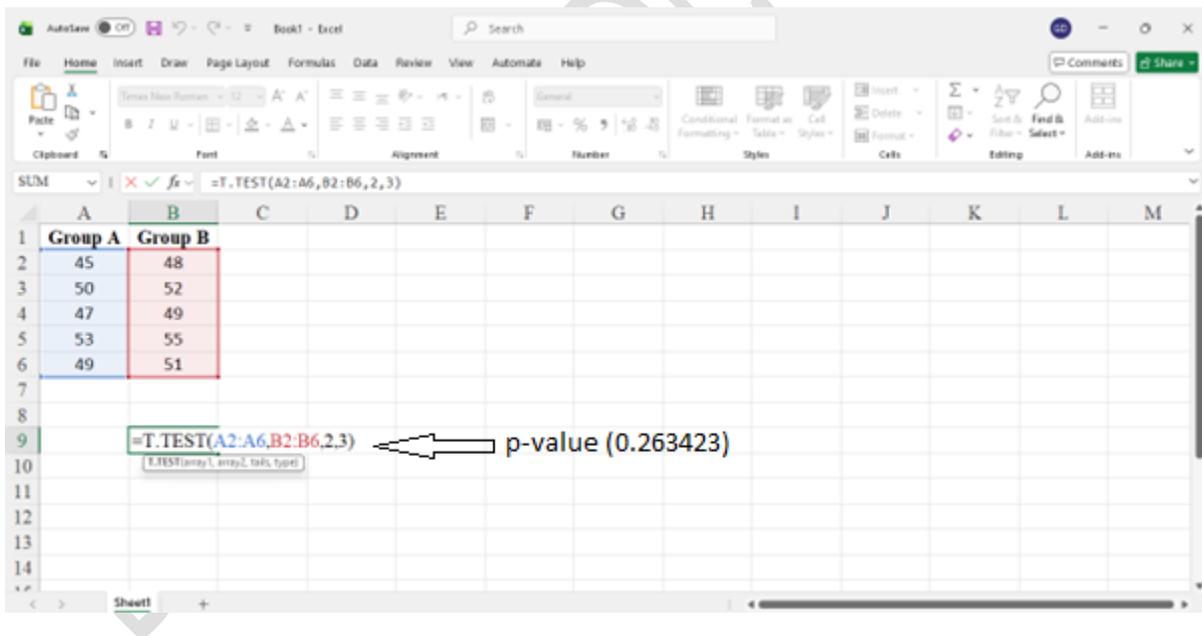
- **How to Run t-Test (Assuming Unequal Variances) in Excel**

- **Excel Formula Used**

=T.TEST(A2:A6, B2:B6, 2, 3)

- **Meaning of Each Argument**

- A2:A6 → Sample data for **Group A**
- B2:B6 → Sample data for **Group B**
- 2 (tails) → **Two-tailed test**, used when checking for any difference between means (If a one-tailed test is required, the value **1** should be used.)
- 3 (type) → **Two-sample t-test assuming unequal variances**



- **Interpretation of the Result**

- The output of the formula is a **p-value**
- If **p-value** ≤ **0.05**, reject  $H_0$  → significant difference exists
- If **p-value** > **0.05**, accept  $H_0$  → no significant difference

In our example the p-value (0.263423) is greater than the level of significance (0.05), the null hypothesis is accepted. Hence, there is no significant difference between the means of the two groups.

### C. F-Test

An F-Test (Two-Sample for Variances) is a parametric statistical test used to determine whether the variances of two independent populations are significantly different. It is mainly used to check the equality of variances before applying a two-sample t-test.

#### ➤ When to Use an F-Test

An F-test is used when:

1. You want to **compare the variances of two independent samples**
2. The data is **quantitative (numerical)**
3. Samples are **independent**
4. The populations are **normally distributed**

#### ➤ Examples:

- Comparing **variability in salaries** of two departments
- Comparing **risk (variance) of returns** of two investment portfolios
- Comparing **production variability** of two machines

#### ➤ Basic Assumptions of F-Test

1. **Normality**  
Both populations should be normally distributed.
2. **Independence**  
Samples must be independent of each other.
3. **Scale of Measurement**  
Data should be measured on an **interval or ratio scale**.
4. **Random Sampling**  
Samples should be randomly selected.

#### ➤ Null and Alternative Hypothesis

Let:

$$\sigma_1^2 = \text{Variance of population 1}$$

$$\sigma_2^2 = \text{Variance of population 2}$$

- **Null Hypothesis ( $H_0$ )**

There is **no significant difference** between the variances of the two populations.

$$H_0: \sigma_1^2 = \sigma_2^2$$

- **Alternative Hypothesis ( $H_1$ )**

There is a **significant difference** between the variances of the two populations.

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

*(One-tailed alternatives may also be used when appropriate.)*

- **How to Run F-Test in Excel**

- **Use the Excel function:**

$$=F.TEST(A2:A6, B2:B6)$$

- **Meaning of Each Argument**

- **A2:A6** → Sample data for **Group A**
  - **B2:B6** → Sample data for **Group B**
- The F.TEST function automatically performs a **two-tailed test** for equality of variances.

- **Interpretation of Result (Using p-value)**

- If **p-value  $\leq 0.05$**  → Reject  $H_0$
- If **p-value  $> 0.05$**  → Accept  $H_0$

|    | Group A | Group B               |
|----|---------|-----------------------|
| 1  |         |                       |
| 2  | 45      | 48                    |
| 3  | 50      | 52                    |
| 4  | 47      | 49                    |
| 5  | 53      | 55                    |
| 6  | 49      | 51                    |
| 7  |         |                       |
| 8  |         |                       |
| 9  |         | =F.TEST(A2:A6, B2:B6) |
| 10 |         |                       |
| 11 |         |                       |
| 12 |         |                       |
| 13 |         |                       |
| 14 |         |                       |

Since the **p-value (0.847833)** is **greater than the level of significance (0.05)**, the **null hypothesis is accepted**. Hence, there is **no statistically significant difference** between the variances of the two populations.