# Basic Filtering

## Learner Guide

# CONTENTS

# How to Use This Workbook

### Activity
Alongside this icon you'll find details of the group/individual activity or a point for everyone to discuss.

### Useful Tool
This icon indicates a technique that will help you put what you learn into practice.

### Important Idea or Concept
Generally, this icon is used to draw your attention to ideas that you need to understand by this point in the course. Let your trainer know if you do not understand or see the relevance of this idea or concept.

### Helpful Hint
This icon guides you to tips or hints that will help you avoid the standard pitfalls that await the unwary practitioner or to show you how you might increase your effectiveness or efficiency in practising what you have learnt.

### Key Point
This icon is used to indicate something that practitioners in this field should know. It's likely to be one of the major things to remember from the course, so check you do understand these key points.

### Reference Material
When we have only touched briefly on a topic this icon highlights where to look for additional information on the subject. It may also be used to draw your attention to International or National Standards or Web addresses that have interesting collections of information.

### Definition
Where a word with a very specific definition (or one that could be described as jargon) is introduced this will highlight that a definition is provided. (These words will also be found in the Glossary at the back of the workbook.)

### Warning
This icon is used to point out important information that may affect you and your use of the product or service in question.

# Introduction

Filtering data is a fundamental process in data management and analysis used to extract specific information from a larger dataset based on defined criteria.

Data filtering serves the purpose of refining, extracting, and manipulating data to extract valuable insights, improve data quality, and aid decision-making processes. It helps users work with manageable subsets of data that are relevant to their specific analysis or task at hand.

Your task is to apply basic filtering in Microsoft Excel.

## Objectives

The objectives of this activity are to:

- Introduce the filter feature of Excel tables.

- Apply and remove different types of filters to a small example dataset. Filters include:

    - Checkbox filters

    - Numeric filters

    - Text filters

- Using multiple filters to achieve a result.

# Basic Filtering

Filtering in Excel allows you to display only the data that meets certain criteria, hiding the rows that don't match those criteria.

In this section, you will use the filter feature of Excel tables. These steps will help you perform basic filtering in Excel, allowing you to display specific data that meets your defined conditions.

## Table filtering

**Guided activity:**
**Filtering in Excel Tables**

You will be using the Palmer Penguins dataset. This dataset provides information on three species of penguins, including Adélie, Gentoo, and Chinstrap, and covers essential biological metrics such as bill dimensions and body mass.

The dataset consists of the following columns:

1. **Species**: Species of the penguin (Adelie, Chinstrap, Gentoo)
2. **Island**: Island where the penguin was found (Biscoe, Dream, Torgersen)
3. **Bill Length (mm)**: Bill length in millimeters
4. **Bill Depth (mm)**: Bill depth in millimeters
5. **Flipper Length (mm)**: Flipper length in millimeters
6. **Body Mass (g)**: Body mass in grams
7. **Gender**: Sex of the penguin (Male, Female)
8. **Year**: Year the data was collected (2007-2009)

Here are the basic steps to filter data in Excel:

1. Open the **Palmer Penguins.XLSX** workbook.

2. Click on any cell within the range of the Penguins data you want to filter.

3. In the Excel ribbon, navigate to the **Data** tab.

4. Click on the **Filter** button found on the **Sort and Filter** group. This button typically looks like a funnel or filter icon in the toolbar. Alternatively, you can use the keyboard shortcut **Ctrl + Shift + L** to apply or remove filters.

**Helpful hint**

If your data is already in a table, the filters are automatically applied to your columns.

To tabulate your data, select **Ctrl + A** on your keyboard, navigate to the **Insert** tab, and select **Table**. A **Create table** dialog box will appear and click **OK**.

5. Once the filter is applied, small drop-down arrows will appear next to the column headers. Click on the arrow in the column you want to filter. For example, let's filter **bill_length_mm**.

6. In the drop-down menu, you'll see a list of unique values from the **bill_length_mm** column. You can either select specific values to display, use number filters or the search option.

7. Uncheck the **Select All** box and check all boxes with values greater than or equal to 50.

8. After setting your criteria, click **OK** to filter the data. Excel will show only the rows that meet your specified criteria, hiding the rest.

9. To remove the filter and display all the data again, click on the **filter** button for the **bill_length_mm** column and select **Clear Filter from bill_length_mm**.

## Number filters

In this section, you will apply numeric filters to columns containing numerical data in the penguin dataset.

The number filter process can be repeated for different columns or criteria as needed to filter the dataset based on specific numeric conditions. The following steps will allow you to apply numeric filters.

**Guided activity:**
**Apply number filters on numerical columns.**

1. Enable filters by going to the **Data** tab in the Excel ribbon and selecting the **filter icon**.

2. Click on the drop-down arrow in the header of the **bill_length_mm** column containing numeric data that you want to filter.

3. Choose **Number Filters** from the drop-down menu.

4. You will see various options such as Equals, Greater Than, Less Than, Between, etc.

5. Choose the specific numeric filter criterion you want to apply. For example: select **between** and enter the range 50 and 50.9 This means we want numbers that are greater than or equal to 50 **And** is less than or equal to 50. Ensure that **And** is selected.

6. After choosing the filter criteria and entering the values, click **OK**. Excel will filter the dataset based on your specified numeric conditions. The filter will present values within 50 and 50.9 for the **bill_length_mm** column.

7. To remove the filter and display the entire dataset again, click on the filter icon in the column header and select **Clear Filter from bill_length_mm**.

## Text filters

In this section, you will apply text filters to columns containing text data in the penguin dataset.

**Guided activity:**
**Apply text filters based on specific text conditions.**

1. Ensure **Filters** are enabled before you proceed.

2. Click on the drop-down arrow in the columns containing text data that you want to filter. For example, "species", "island", and "gender".

3. You'll see options such as "Sort A to Z," "Sort Z to A," and "Text Filters."

4. Apply "Sort A to Z," "Sort Z to A," on one of the text columns (e.g., species). This allows you to sort the data in alphabetical order or in reverse alphabetical order.

5. Click on **Text Filters**. You'll see options like "Equals," "Does Not Equal," "Begins With," "Contains," etc.

6. **To filter species that start with a particular letter**: Select **Begins With** and enter the starting letter or text. For example, "ade" for "Adelia". It is not case sensitive.

7. Click **OK** after selecting the text filter criterion and entering the text. Excel will now filter the "Species" column to display only those rows where the species name contains the text you specified.

8. **To filter species that contain a specific word**: Select **Contains** and enter the word or part of the word. For example: you want to show species that contain the letter "e". This shows species that contains the letter "e", and these species are, "Adelia" and "Gentoo". You can also enter, parts of a word you want to filter by in the text box. For example, "lie". This will present all species containing "lie", such as Adelie. Click **OK** after entering the letter and text.

9. **To filter species that match exactly**: Select **Equals** and enter the exact species name. For example, "Chinstrap". This will present the "Chinstrap" penguin species. Click **OK** after entering the text.

## Multiple filtered columns

In this section, you will apply filters to multiple columns simultaneously to narrow down the dataset based on combined criteria, using the "species," "island," and "gender" columns.

This process allows you to filter the dataset based on combined criteria across multiple columns, focusing on subsets of data that meet all the specified conditions.

Excel's advanced filtering command also allows multiple criteria to be specified in a variety of ways. However, this will be covered in the next topic.

**Guided activity:**
**Apply filters to multiple columns simultaneously**

Let's apply filter to multiple columns to determine "Female" "Adelie" penguins found in the "Dream" island.

1. Ensure that the **Palmer Penguins.XLSX** workbook is still open from the previous activity.

2. Ensure **Filters** are enabled before you proceed. Remember a short cut is **Ctrl + Shift + L**.

3. Click on the drop-down arrow in the header of the **Species** column. Uncheck the **Select All** box and check **Adelie** or use the **Text Filters** to filter species containing a particular word.

4. Click on the drop-down arrows in the header of the **Gender** column. Uncheck the **Select All** box and check **Female**.

5. Click on the drop-down arrows in the header of the **Island** column. Uncheck the **Select All** box and check **Dream**.

6. Excel will apply the combined filters, displaying only the rows that meet all the specified criteria across the selected columns.

7. You can further adjust or refine the filters in any column to narrow down the dataset based on your desired criteria.

8. To clear the filters and view the entire dataset again: Click on any filtered column's filter icon. Select **Clear Filter from [Column Name]**.

# Activity on Basic Filtering

## Task - Superstore Dataset.

You have been tasked to explore and filter the Superstore dataset using different criteria in Excel to understand the distribution of data in various categories.

**Objective:** Familiarise yourself with practical filtering functionalities in Excel using real-world data from the Superstore dataset.

The dataset consists of 9994 rows and 21 columns.

|    | Column headers | Definition |
|----|----------------|------------|
| 1  | Row ID | Unique ID for each row. |
| 2  | Order ID | Unique Order ID for each Customer. |
| 3  | Order Date | Order Date of the product |
| 4  | Ship Date | Shipping Date of the Product. |
| 5  | Ship Mode | Shipping Mode specified by the Customer. |
| 6  | Customer ID | Unique ID to identify each Customer. |
| 7  | Customer Name | Name of the Customer. |
| 8  | Segment | The segment where the Customer belongs. |
| 9  | Country | Country of residence of the Customer. |
| 10 | City | City of residence of the Customer. |
| 11 | State | State of residence of the Customer. |
| 12 | Postal Code | Postal Code of every Customer. |
| 13 | Region | Region where the Customer belong. |
| 14 | Product ID | Unique ID of the Product. |
| 15 | Category | Category of the product ordered. |
| 16 | Sub-Category | Sub-Category of the product ordered. |
| 17 | Product Name | Name of the Product |
| 18 | Sales | Sales of the Product |
| 19 | Quantity | Quantity of the Product |
| 20 | Discount | Discount provided |
| 21 | Profit | Profit/Loss incurred |

**Independent activity:**
**Apply basic filtering methods to multiple columns on the superstore dataset.**

1. Open the **Superstore.XLSX** workbook.
2. Enable **filters**.
3. **Filter by Category:**
   a. Filter the **Category** column to show only **Office Supplies** items.

4. **Filter by Sub-category:**
   a. Within the **Office Supplies** category, filter the **Sub-Category** column to display only **Binders** and **Paper**.
5. **Filter by Region and Sales amount:**
   a. Filter the **Region** column to show data only from the **East** region.
   b. In the **Sales** column, filter to display sales greater than $500.
6. **Multiple column Filtering:**
   a. Simultaneously filter the **Category** to **Furniture** and the **Sub-Category** to **Chairs**.
7. Make observations about the filtered data in each task on a MS word doc:
   a. Note the number of records displayed after each filtering.
   b. Analyse trends or patterns in the filtered data.
   c. Make note of any interesting insights or observations from the subsets obtained.
8. Discuss the importance of filtering data to analyse specific subsets and draw meaningful insights.
9. Clear all filters, save, and close the workbook.
10. Save your MS word doc.