Manipulating dataset rows in Excel

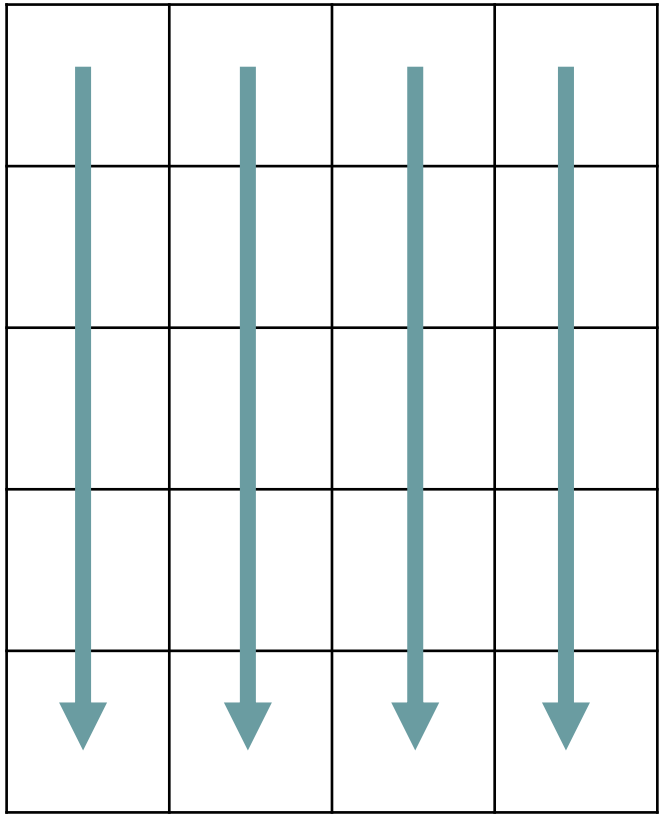
Learning intentions

We will be learning how to manipulate data in Excel, specifically,

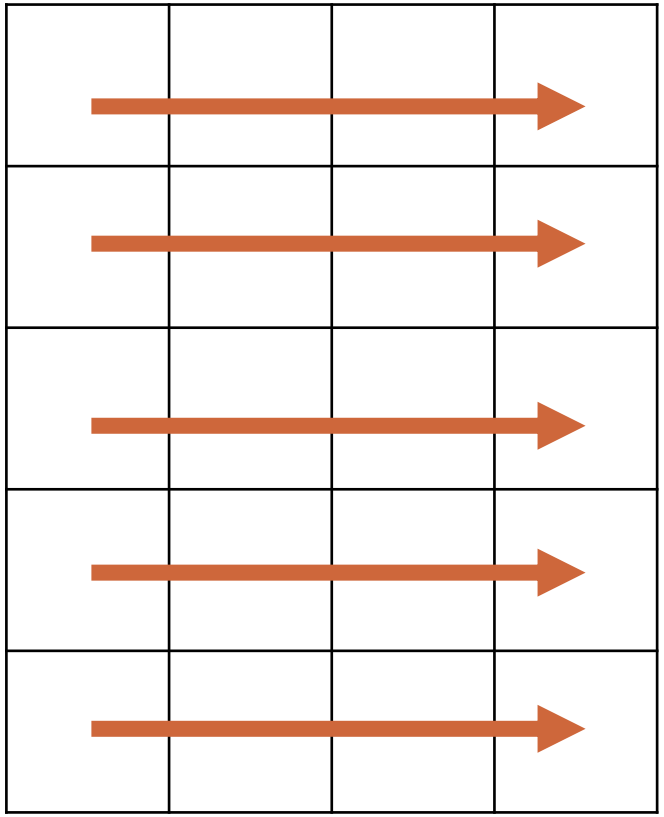
- to be able to **filter and sort** rows
- to be able to **subset** data to select the parts of the data you are interested in
- to be able to **remove duplicates** from the data

Background

In previous lessons we have looked at manipulating columns…



…now we are going to look at manipulating rows.



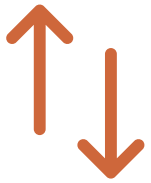
Why this is important?

Some benefits of manipulating the rows in a dataset are,

Helps you spot **missing** or **unusual data** (outliers)

Hides non-relevant information

To **group data** into useful chunks



To **sort it** into a useful order

Show me…

Here is a dataset which has a row with missing data.

ID	first_name	last_name	age
DF 12 24	Jack	Brown	20
JC 56 54	Isla	Smith	25
XX XX XX	Jamie		

The last row has missing information.

By manipulating the rows it makes it easier to spot issues like this. Then you can decide how you will handle the data.

Definition

Sorting

To re-order the rows of a dataset

Show me…

Datasets can be sorted by number

Position	Name	Start time
1	Fred	10:45
2	Amy	10:30
3	Vicky	10:15

or alphabetically

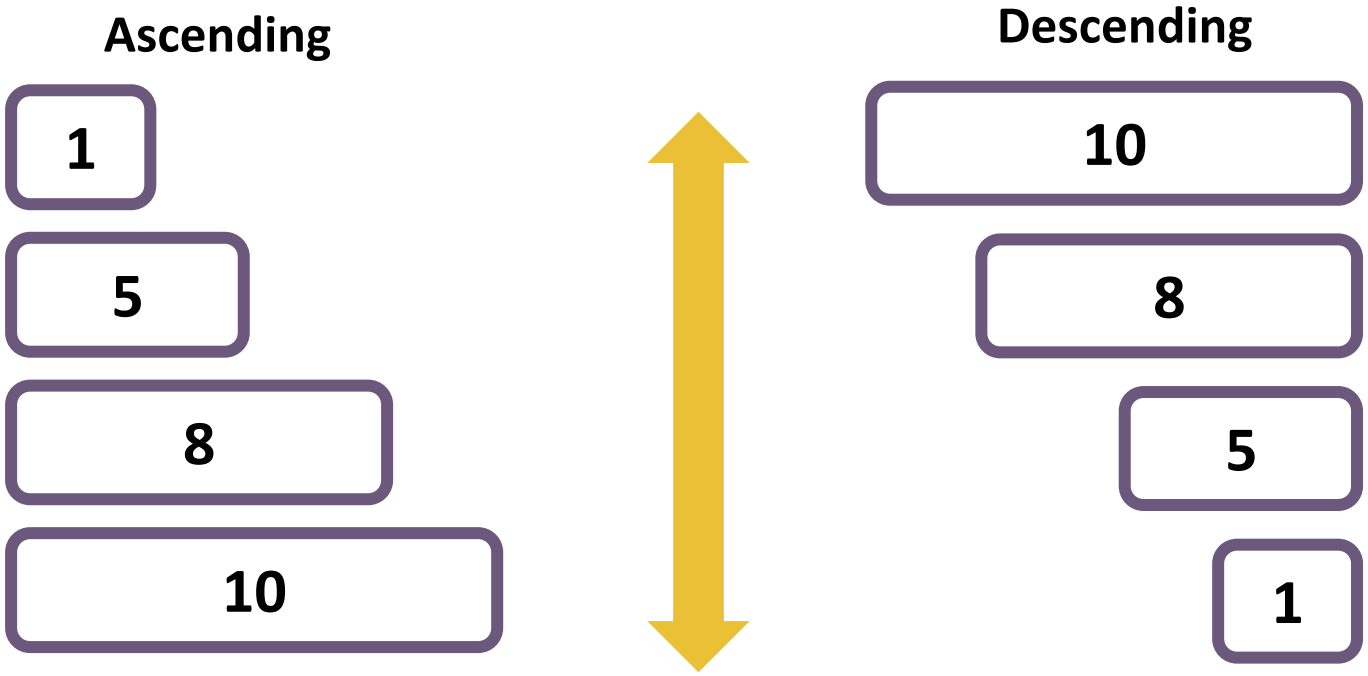
Position	Name	Start time
2	Amy	10:30
1	Fred	10:45
3	Vicky	10:15

or by date/time

Position	Name	Start time
3	Vicky	10:15
2	Amy	10:30
1	Fred	10:45

Sorting data

When sorting data you can decide if the order is **ascending** (smallest to largest) or **descending** (largest to smallest)



Your turn…

Which mountain do you think would be at the bottom of the dataset if these these mountains were sorted by ascending height (smallest to largest)?

mountain_name	height_m
Ben More	1,174
Ben Lomond	974
Ben Nevis	1,344
Schiehallion	1,083

Your turn…

Sort these mountains by their height (smallest to largest)

mountain_name	height_m
Ben More	1,174
Ben Lomond	974
Ben Nevis	1,344
Schiehallion	1,083

Sort by number

mountain_name	height_m
Ben Lomond	974
Schiehallion	1,083
Ben More	1,174
Ben Nevis	1,344

Sorting best practice

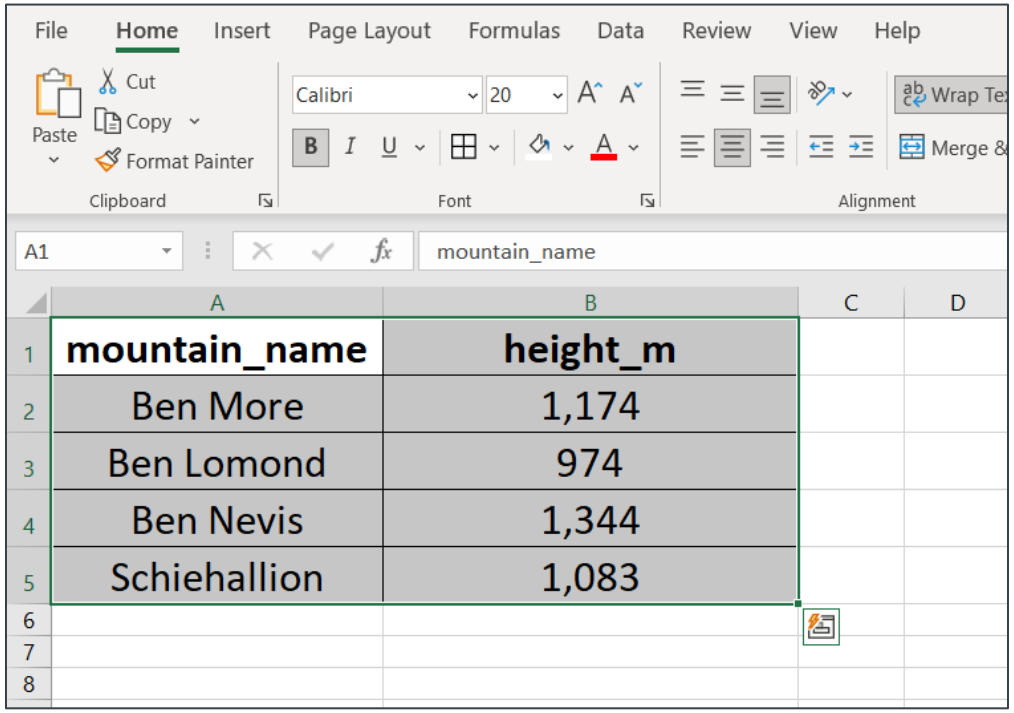
It can be **very difficult to un-sort** data back to its original order.

Think about saving an original copy before sorting.

How to Sort in Excel

Step 1.

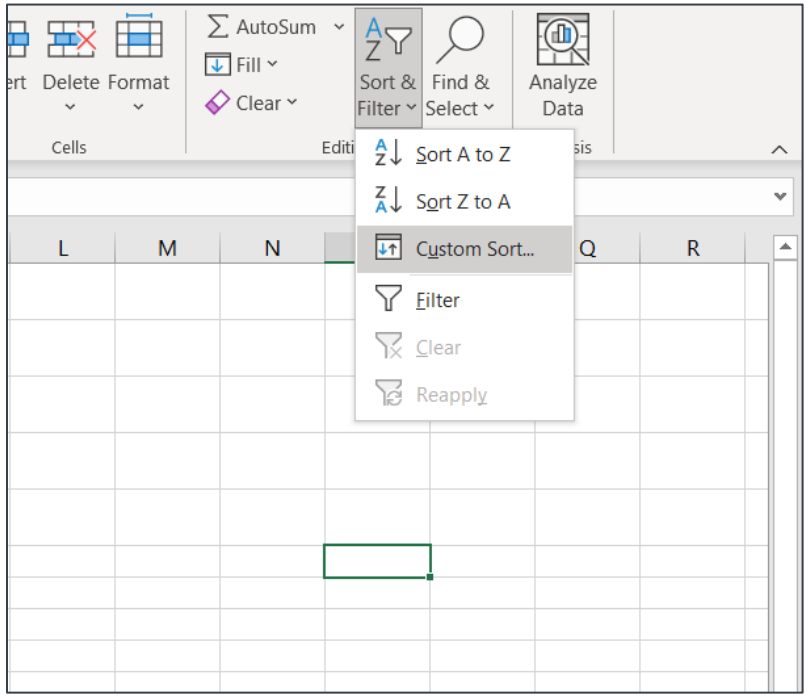
Select all the data including the headings that you need to sort.



Sorting in Excel

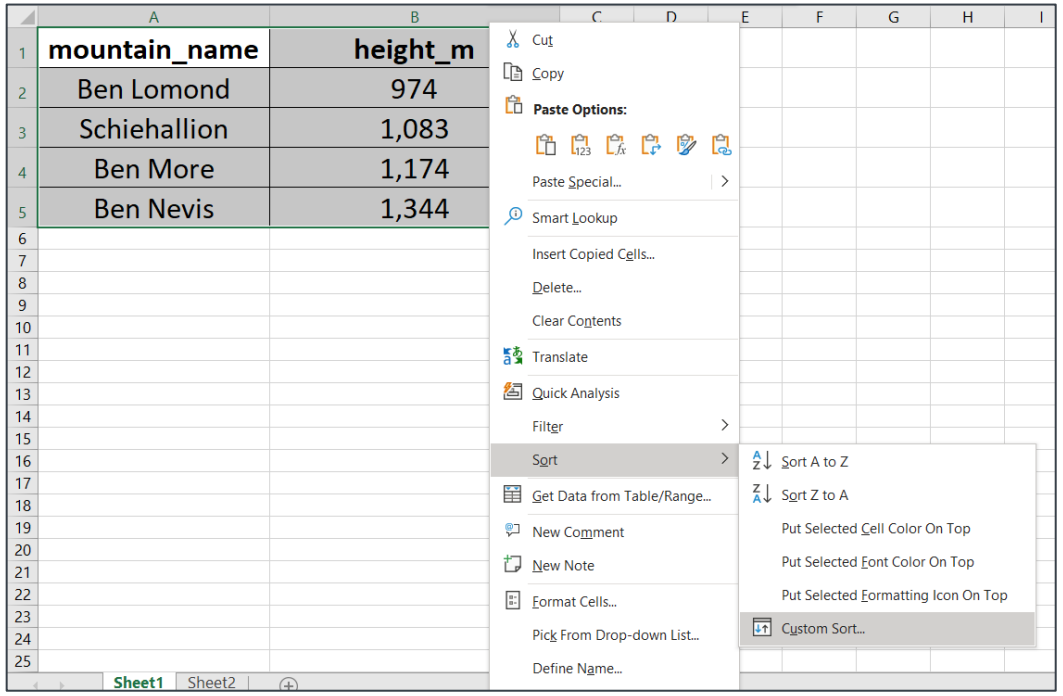
Step 2.

In the Home ribbon, click on ‘Sort & Filter’ then on ‘**Custom Sort…**’



or

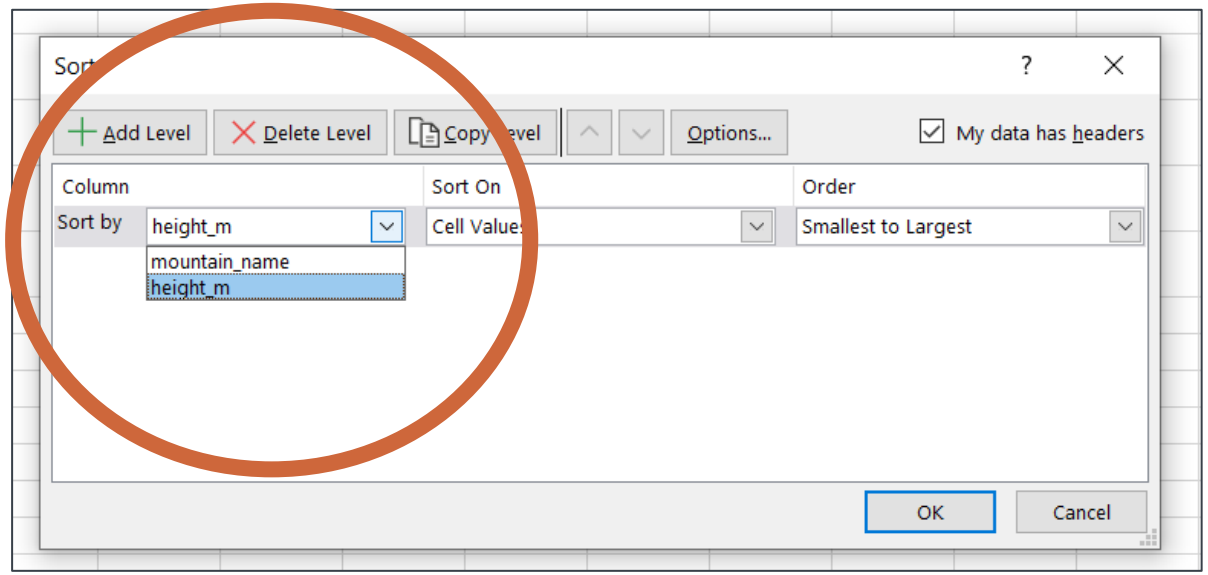
Right click and find the **‘Custom Sort…** option.



Sorting in Excel

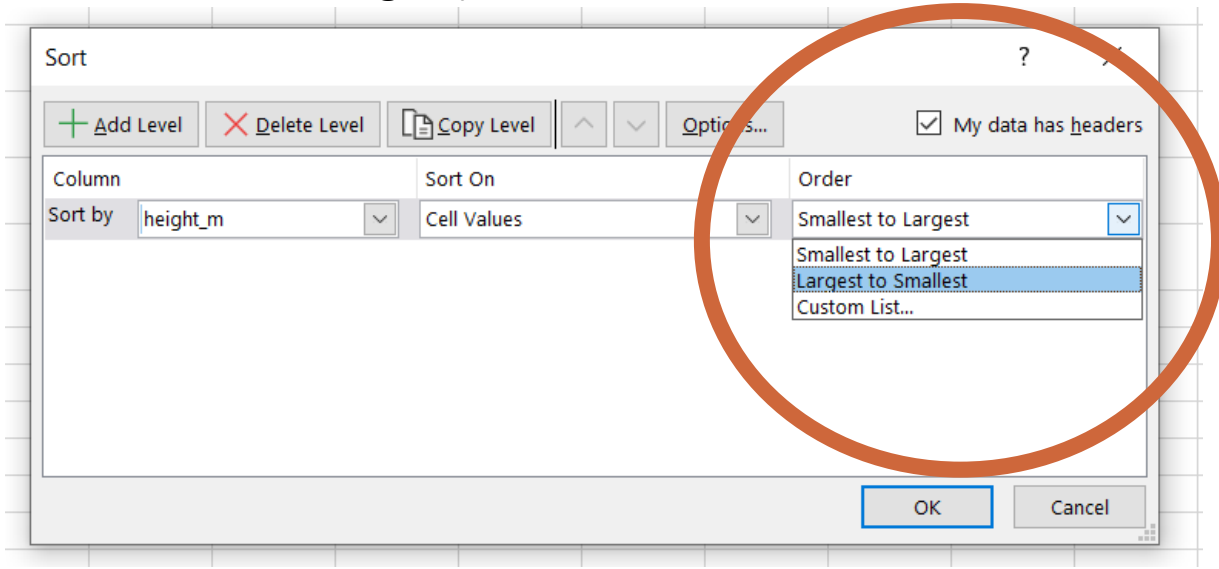
Step 3.

Choose the column header you would like to **sort by** (e.g. height_m)



Step 4.

Select the **order** you want to sort by (e.g. smallest to largest)



Step 5. Press OK, your data will now be sorted.

Next steps

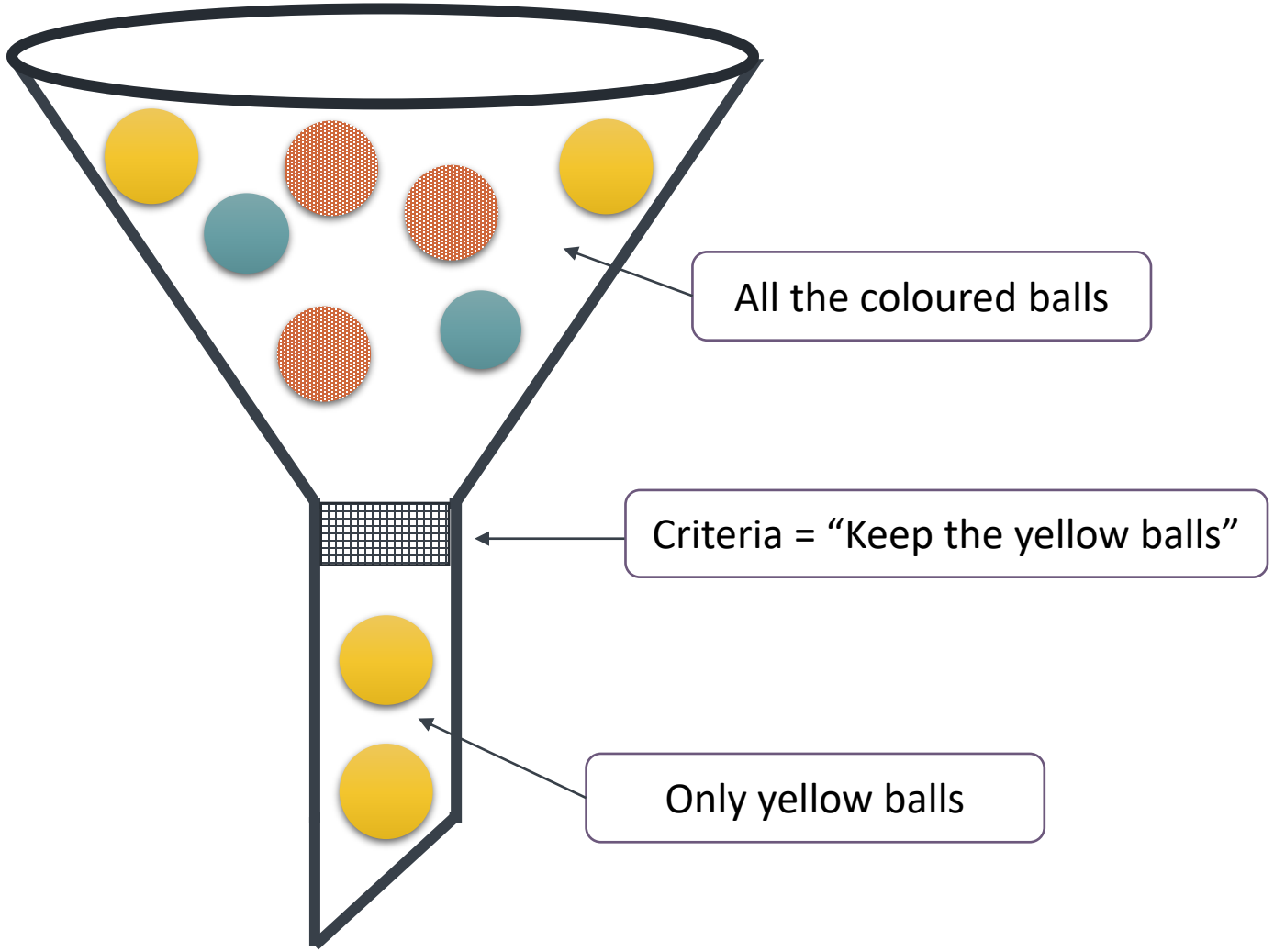
Complete **questions 1 to 8**
in **section 1** of the
‘Manipulating datasets rows in Excel’ workbook.

Definition

Filter

To choose some of the rows in a dataset based on some criteria

Filtering data



Imagine you had **a group of coloured balls.**

These balls are trying to pass through **a filter.**

This filter **only allows yellow balls** through (the **criteria**).

After the filter you would only see the yellow balls.

Filtering data

When filtering data it can help to think about the following questions,

What **data** do I have?

What do you **need** from the data?

What **criteria** do I need to filter my data by?

Filtering data

What **data** do I have?

What do you **need** from the data?

What **criteria** do I need to filter my data by?

Details of people attending a match including,

Name,

Email address,

Age

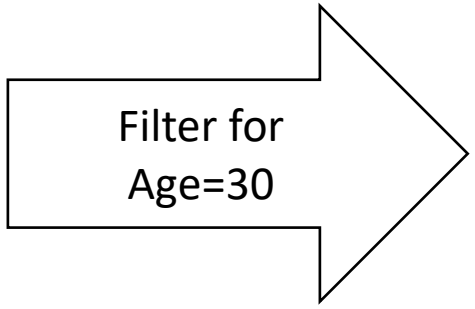
"I need to **email** all attendees that are 30 years old."

Age = 30

Filtering data

"I need to **email** all attendees that are 30 years old."

name	email_address	age
Jack	jack@gmail.com	20
Jane	jane@outlook.com	30
Isla	Isla_g@hotmail.com	30
John	john@hotmail.com	18
Lily	lily@gmail.com	32
James	james@gmail.com	30



name	email_address	age
Jane	jane@outlook.com	30
Isla	isla@hotmail.com	30
James	james@gmail.com	30

Show me…

Here is an example of filtering rows. You can filter based on **numbers**, **strings** or **dates**.

Type	Size	Colour
Hat	M	Red
Coat	L	Blue
Hat	S	Black
Top	S	White

Filter for
Type = Hat

Type	Size	Colour
Hat	M	Red
Hat	S	Black

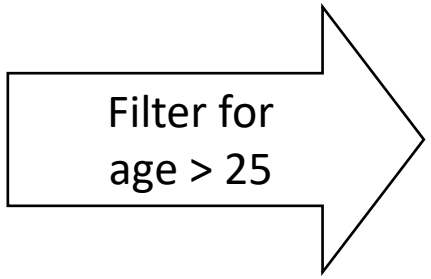
Example

Filter

To choose some of the rows in a dataset based on some criteria

Filter this dataset to choose the rows where the age is greater than 25 years old.

name	age
Charlie	18
Max	30
Rowan	32
Jess	21



name	age
Max	30
Rowan	32

Your turn…

Which locations would be in your dataset if you filtered it to choose the rows where **dawn is before 6am?**

location	dawn
Edinburgh	03:30
Paris	05:09
Sydney	06:25
New York	04:54

Your turn…

Filter this dataset to choose the rows where **dawn is before 6am**.

location	dawn
Edinburgh	03:30
Paris	05:09
Sydney	06:25
New York	04:54

Filter for dawn is before 6am

location	dawn
Edinburgh	03:30
Paris	05:09
New York	04:54

How to Filter in Excel

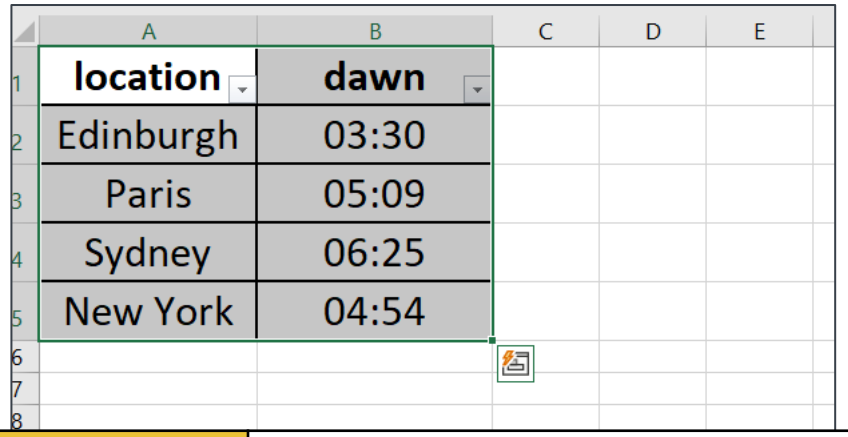
Step 1.

To turn on the filter options,

Highlight all the data you want to filter

Press **'Ctrl' and 'Shift' and 'L'** if you are using Windows or **'Command' and 'Shift' and 'L'** for Mac at the same time.

This tells Excel to make your selected cells a table. Excel now shows filter arrows next to each heading. We can use these to filter our data.



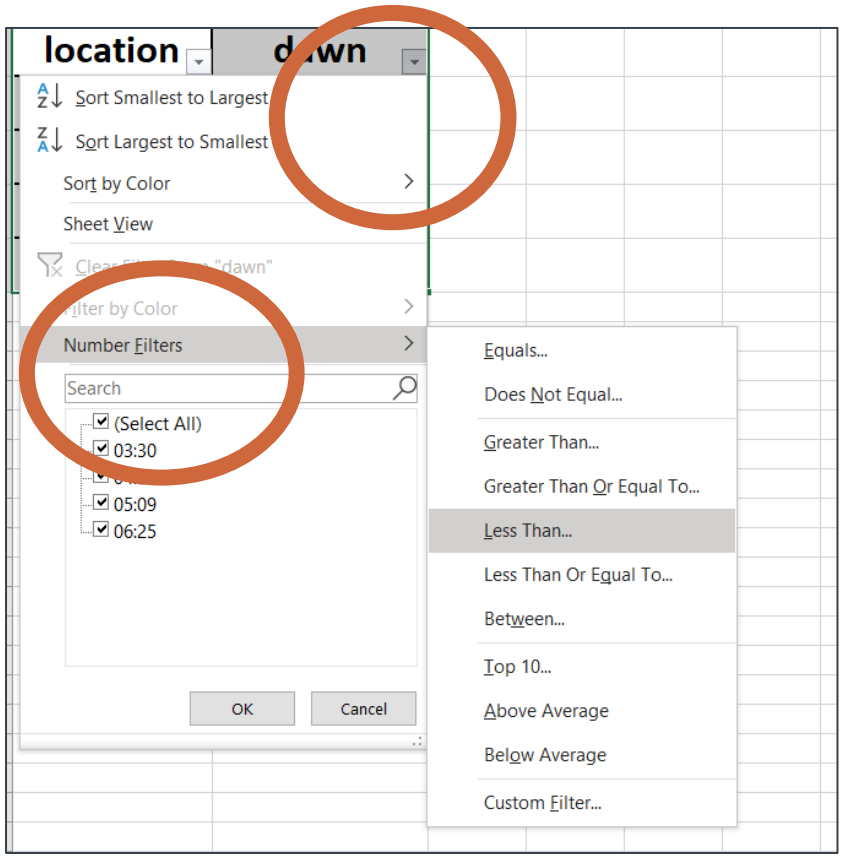
location	dawn
Edinburgh	03:30
Paris	05:09
Sydney	06:25
New York	04:54

Windows	Ctrl + Shift + L
Mac	Command + Shift +L

Filter in Excel

Step 2.

Click on the small arrow that has now appeared next to the column you want to filter, and select ‘Number Filters’.



Step 3.

Fill in the ‘Custom AutoFilter’ box and press OK

Next steps

Complete **questions 1 to 9**
in **section 2** of the
‘Manipulating datasets rows in Excel’ workbook.

Definition

Subset

To choose some rows and columns from a dataset based on some criteria

Show me…

Here is an example of subsetting a dataset.

ID	flower	colour	size_cm
1	Rose	White	0.96
2	Heather	White	0.12
3	Sunflower	Yellow	2.34

Select columns
flower and **colour**
and
filter where
colour = "White"

flower	colour
Rose	White
Heather	White

Subsetting involves **selecting columns** and **filtering rows** at the same time.

Your turn…

Which parts of this dataset to you think you would have if you subset this dataset so you kept,

- columns **day**, **location** and **weather_forecast**
- rows where **day = Tuesday**

day	location	weather_forecast	wind_speed_mph
Monday	Ettrick	sunny	3
Monday	Rockcliffe	cloudy	15
Tuesday	Creetown	rain	10
Tuesday	Musselburgh	sunny	5

Your turn…

Subset this dataset so you keep the **day**, **location** and **weather_forecast** where **day = Tuesday.**

day	location	weather_ forecast	wind_speed _mph
Monday	Ettrick	sunny	3
Monday	Rockcliffe	cloudy	15
Tuesday	Creetown	rain	10
Tuesday	Musselburgh	sunny	5

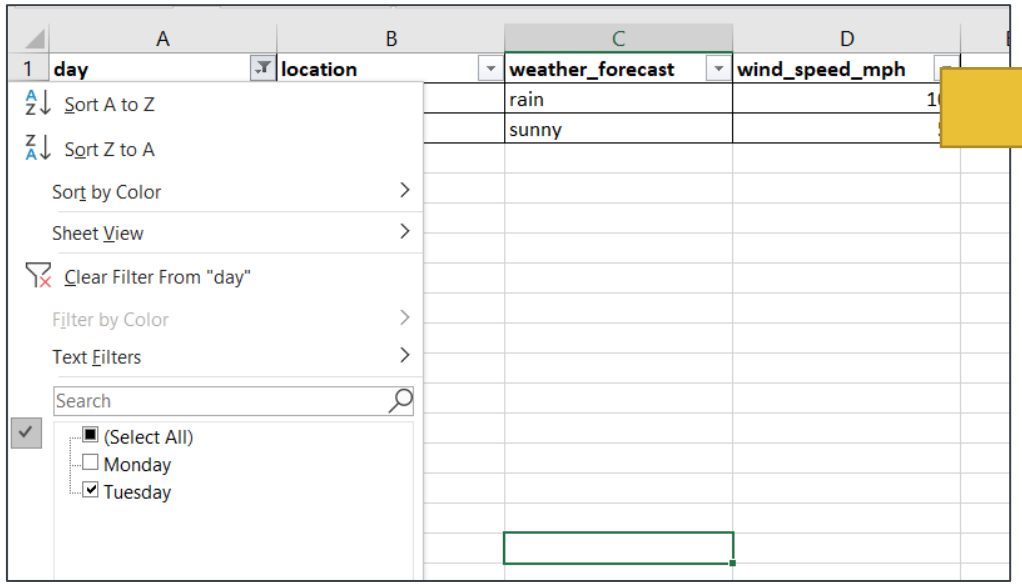
Subset

day	location	weather_ forecast
Tuesday	Creetown	rain
Tuesday	Musselburgh	sunny

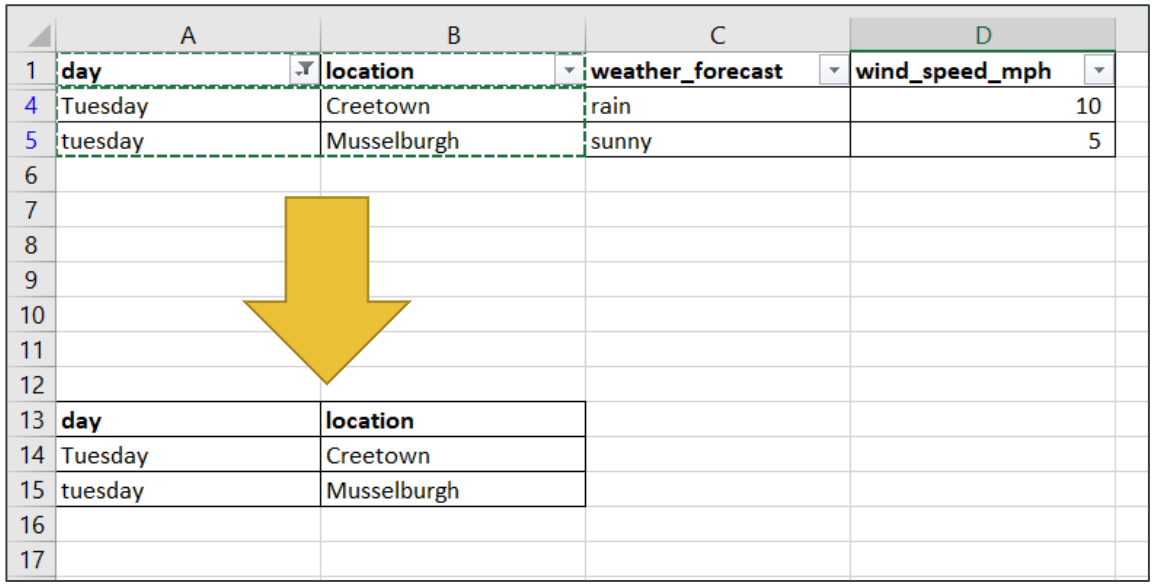
How to Subset in Excel

To subset in Excel you need to follow the steps to **'filter' the rows** and **'select' columns.**

Filter the rows….



… then select the columns



Next steps

Complete **questions 1 to 8**
in **section 3** of the
‘Manipulating datasets rows in Excel’ workbook.

Definition

Remove duplicates (deduping)

To remove un-needed repeated rows of data

Show me…

Fruit	Colour	Price
Apple	Red	£1
Apple	Red	£1
Apple	Red	50p
Grapes	Green	£2

This dataset has 2 rows that contain the same information (‘Apple’ and ‘Red’ and ‘£1’).

Fruit	Colour	Price
Apple	Red	£1
Apple	Red	50p
Grapes	Green	£2

We can remove one of these rows without losing any information from the dataset.

Show me…

In this table you **would not** be able to remove any duplicate rows.

day	town	temperature
Monday	Stranraer	10
Monday	Stranraer	11
Tuesday	Wigtown	14
Tuesday	Stranraer	15
Tuesday	Portpatrick	12
Tuesday	Kirkcudbright	13

Although there are **2 rows for Stranraer on Monday,** the temperature column has different values.

You would **miss some information** if you removed one of the rows.

Your turn…

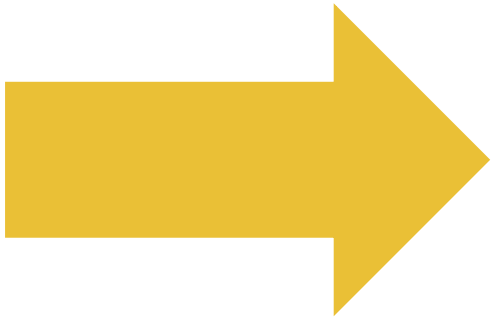
Which rows would you have if you **remove the duplicates** from this dataset?

loch	max_depth
Loch Ness	227
Loch Ness	227
Loch Lomond	190
Loch Ness	227

Your turn…

Remove the duplicates from this dataset

loch	max_depth
Loch Ness	227
Loch Ness	227
Loch Lomond	190
Loch Ness	227



loch	max_depth
Loch Ness	227
Loch Lomond	190

Now there is only 1 row per loch

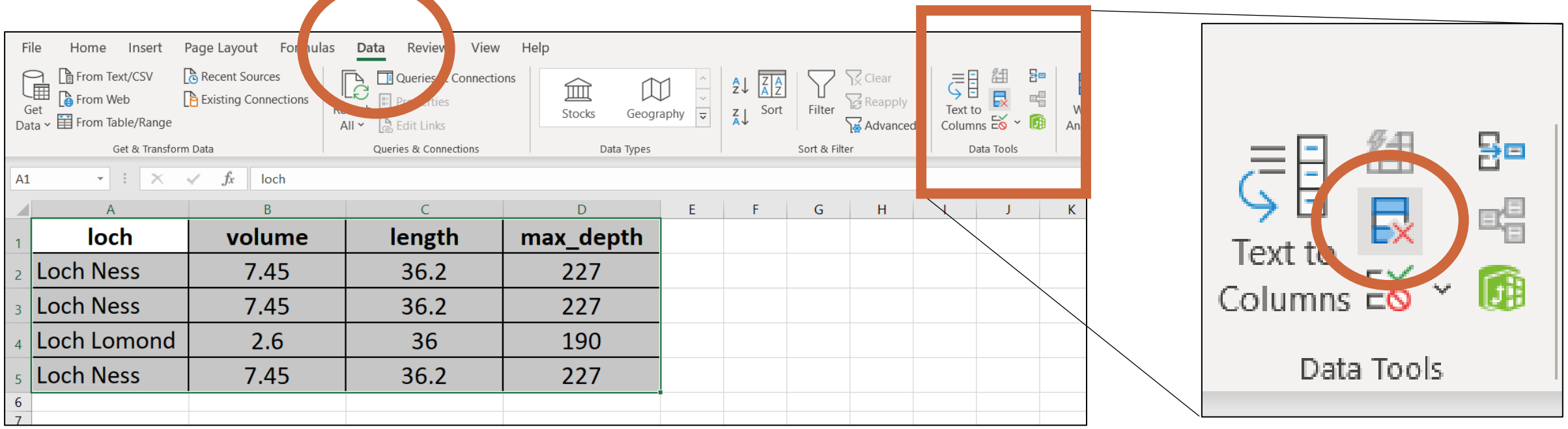
Remember, you can only remove un-needed repeated rows of data if **everything in the rows is identical**.

Remove duplicates in Excel

Step 1.

Highlight the data you would like to remove duplicates from.

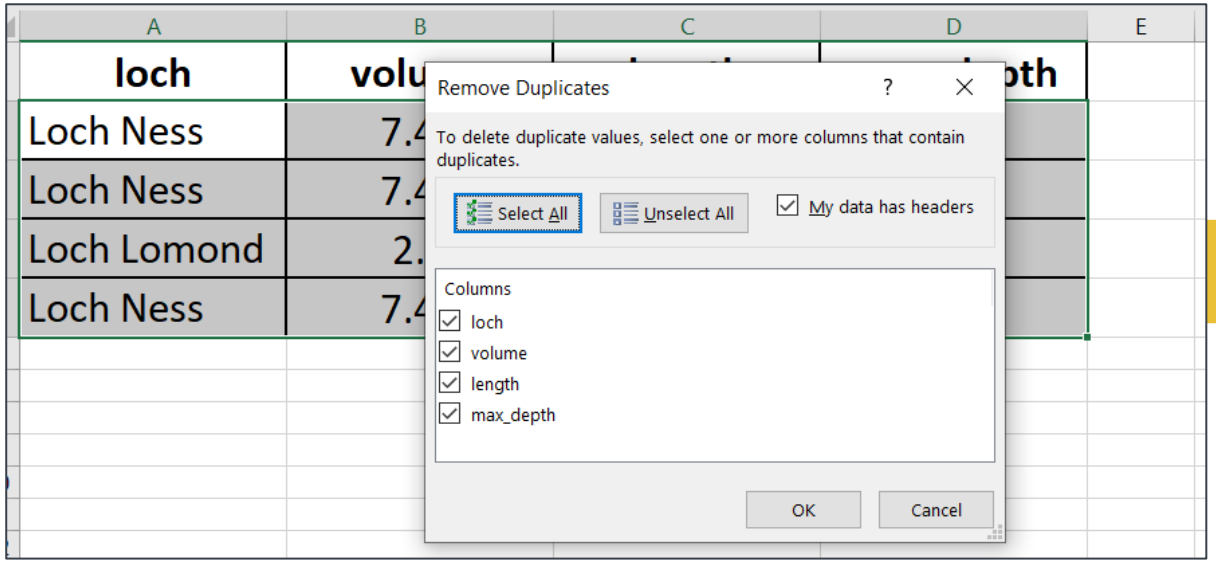
On the ‘Data’ ribbon find the option in the ‘Data Tools’ section called ‘**Remove Duplicates**’. It looks like some boxes with a cross next to them.

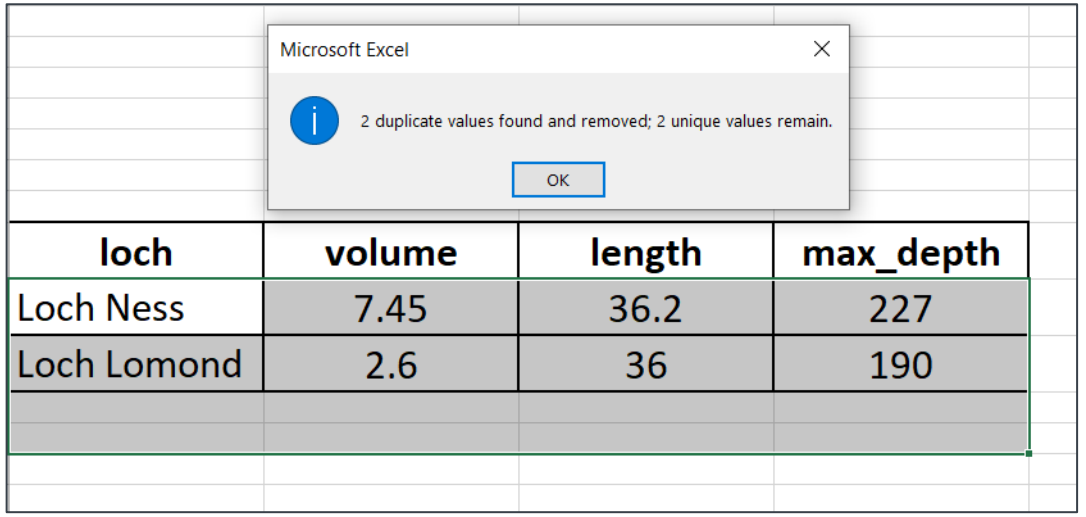


Remove duplicates in Excel

Step 2.

The ‘remove duplicates’ box will open. Press OK.





Next steps

Complete **questions 1 to 4**
in **section 4** of the
‘Manipulating datasets rows in Excel’ workbook.

Additional information

In this lesson we have covered some of the procedures for manipulating data in Microsoft Excel.

However, there are different options (including keyboard shortcuts) available.

For more information, please see

https://support.microsoft.com/en-us/excel

Learning checklist

I can *describe* what is means to sort, filter, subset and remove duplicates from a dataset.

I can *manipulate* data by sorting, filtering, subsetting and removing duplicates in Excel.

How you can use this lesson

You are free to:

- **Share** – copy and redistribute the material in any medium or format
- **Adapt** – remix, transform and build upon the material

Under the following terms:

- **Attribution** — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- **NonCommercial** — You may not use the material for commercial purposes.
- **ShareAlike** — If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

Created by Effini in partnership with Data Education in Schools, The Data Lab and Data Skills for Work, with funding from the Scottish Government.

Effini

THE DATA LAB
value from data

Data Skills FOR WORK

Data Education in Schools