

Practise combining datasets in Excel

Version: 1.1



Learning intentions

We will be learning about **how to combine datasets in Excel**, specifically

- how to use Power Query Editor to **append rows** to a dataset
- how to use Power Query Editor to **join columns** to a dataset

Background

In the “Combining datasets” lesson we looked at what we mean by appending rows and joining columns.

In this lesson we will use Power Query to **practise appending** and **joining datasets in Excel**.

Appending rows

Joining columns

What is Power Query?



“With Power Query you can import or **connect to external** data, and then **shape that data** [...] in ways that meet your needs.

Then, you can **load your query into Excel** to create charts and reports.

Periodically, you can **refresh the data** to make it up-to-date”

Using Power Query in Excel



In this lesson we are going to look at how to use the **combine** part of Power Query to allow us to append and join datasets.

If you would like to know more about Power Query in Excel, please see

<https://powerquery.microsoft.com/en-us/>

Reminder of Power Query Editor

As with other Microsoft packages, there is a **ribbon** at the top with the buttons you will need.

Any steps you apply to your data are recorded here.

The screenshot displays the Power Query Editor interface. At the top, the ribbon is visible with tabs for File, Home, Transform, Add Column, and View. The Home tab is active, showing various data manipulation options. Below the ribbon, a data preview table is shown with columns for race, start, and end times. On the right side, the 'Query Settings' pane is open, showing the 'APPLIED STEPS' section with a list of transformations applied to the data.

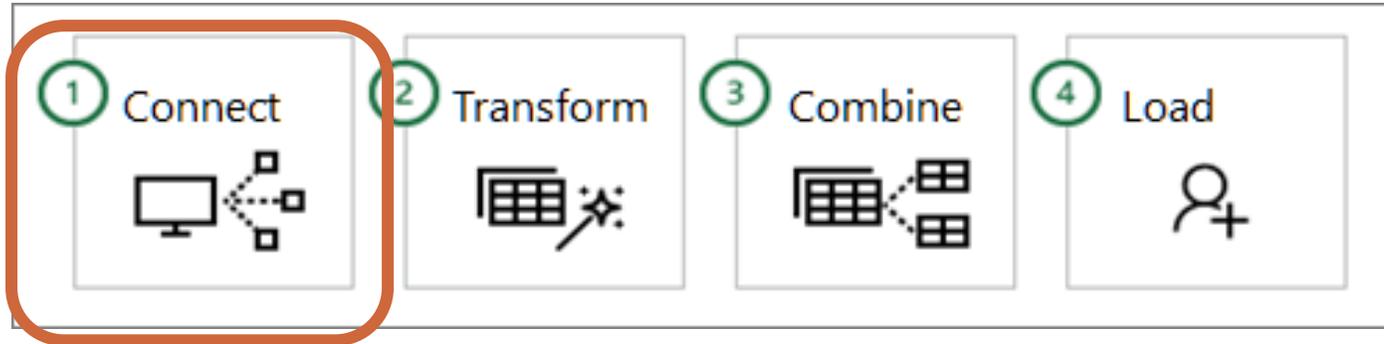
	race	1.2 start	1.2 end
1	Dumfries	0.427083333	1.501460291
2	Glasgow	0.4375	1.687573161
3	Leven	0.416666667	0.544444444
4	Stirling	0.395833333	0.839493237

APPLIED STEPS

- Source
- Changed Type
- Pivoted Column
- Reordered Columns**

Here you can see a **preview of your data** you are transforming

Creating tables in Excel



Before combining datasets in Excel you need to **import the data into a table**.

This is part of the 'connect' stage of using Power Query.

In this lesson we will use datasets that have **already been loaded into Power Query**.

For more information on importing data please see

<https://support.microsoft.com/en-us/office/import-data-from-data-sources-power-query>

Appending data in Power Query Editor

We are going to use Power Query to append the UK CO₂ emissions from 2011 to 2019 to the dataset on the left hand side.

	A	B	C	D	E	F
1	year	co2_emissions		year	co2_emissions	
2	2000	530,890		2011	445,590	
3	2001	545,260		2012	467,780	
4	2002	530,790		2013	453,760	
5	2003	543,040		2014	415,600	
6	2004	543,080		2015	401,080	
7	2005	540,920		2016	380,810	
8	2006	542,060		2017	367,000	
9	2007	530,500		2018	360,730	
10	2008	515,340		2019	348,920	
11	2009	466,490				
12	2010	482,440				
13						



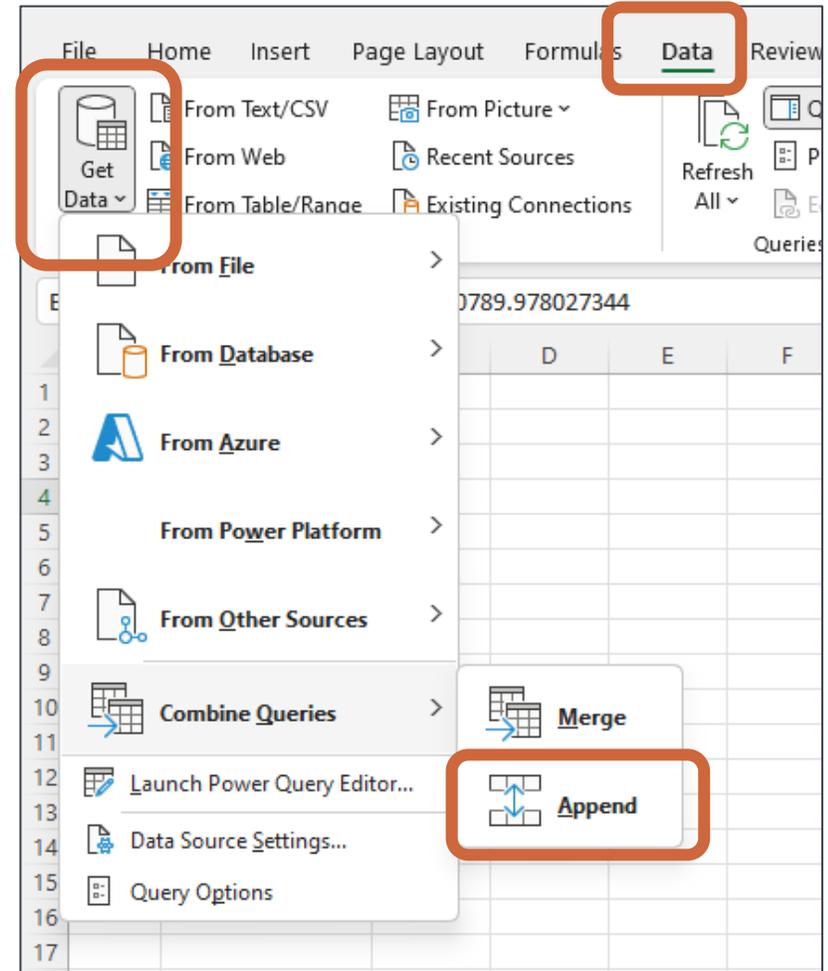
Open the append window

Reminder: the datasets have already been loaded into Power Query.

Step 1.

Click on

- **Data** ribbon then ...
- **Get Data** then ...
- **Combine Queries** then ...
- **Append**



Using the append window

Step 2.

Underneath **First table**, select the name of the first dataset from the drop down list.

Then select the dataset you wish to append from the drop down list underneath the **Second table** drop list.

Then press **OK**.

Append

Concatenate rows from two tables into a single table.

Two tables Three or more tables

First table

CO2_2000_2010

Second table

CO2_2011_2019

CO2_2000_2010

CO2_2011_2019

OK Cancel

Checking the data in Power Query

Step 3.

Check that the dataset looks how you expect. (e.g. you have the number of rows and columns you are expecting)

Under the properties section of the Query Settings, change the name of the dataset from the default name.

Click on **Close & Load**.

The screenshot displays the Power Query Editor window. The 'File' menu is highlighted with a red box, and the 'Close & Load' option is selected. The main data view shows a table with 20 rows and 2 columns. The first column is labeled 'year' and the second is 'co2_emissions'. The data is as follows:

year	co2_emissions
2000	530890
2001	545260
2002	530789.978
2003	543039.978
2004	543080.0171
2005	540919.9829
2006	542059.9976
2007	530500
2008	515340.0269
2009	466489.9902
2010	482440.0024
2011	445589.9963
2012	467779.9988
2013	453760.0098
2014	415600.0061
2015	401079.9866
2016	380809.9976
2017	367000
2018	360730.011
2019	348920.0134

The 'Query Settings' pane on the right is also highlighted with a red box, showing the 'PROPERTIES' section where the 'Name' field is set to 'Append1'. The 'APPLIED STEPS' section shows a single step named 'Source'.

Dataset is loaded into the worksheet

Step 4.

The dataset is now loaded into a worksheet.

The name we gave the dataset in the last step is now the name of the tab.



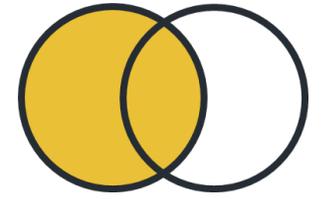
	A	B	C
1	year	co2_emissions	
2	2000	530890	
3	2001	545260	
4	2002	530789.978	
5	2003	543039.978	
6	2004	543080.0171	
7	2005	540919.9829	
8	2006	542059.9976	
9	2007	530500	
10	2008	515340.0269	
11	2009	466489.9902	
12	2010	482440.0024	
13	2011	445589.9963	
14	2012	467779.9988	
15	2013	453760.0098	
16	2014	415600.0061	
17	2015	401079.9866	
18	2016	380809.9976	
19	2017	367000	
20	2018	360730.011	
21	2019	348920.0134	
22			
23			

The screenshot shows a spreadsheet with a worksheet named "CO2_2000_2019" selected, which is highlighted with a red box. The data is organized into columns A and B, with rows 1 through 21 containing the years and corresponding CO2 emissions values.

Next steps

Complete **questions 1 to 3**
in **section 1** of the
'Practise combining datasets' workbook.

Left join datasets in Power Query Editor



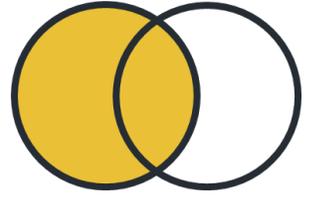
We are now going to LEFT join these datasets. Again, the datasets have already been loaded into Power Query Editor.

	A	B	C	D	E	F
1	Name	max_height		Name	good_with_children	good_with_other_dogs
2	Border Collie	22		Border Collie	3	3
3	Boxer	25		Yorkshire Terrier	5	3
4	Cocker Spaniel	16		Boxer	5	3
5	Dalmatian	24		West Highland White Terrier	5	3
6	Great Dane	32		Dalmatian	3	3
7	Greyhound	30		Rottweiler	3	3
8	Irish Setter	27		Greyhound	3	4
9	Newfoundland	28		Irish Setter	5	5
10	Poodle (Miniature)	15		Cocker Spaniel	5	5
11	Rottweiler	27		Newfoundland	5	5
12	West Highland White Terrier	11				
13	Yorkshire Terrier	8				
14						

Left dataset

Right dataset

Open the merge window

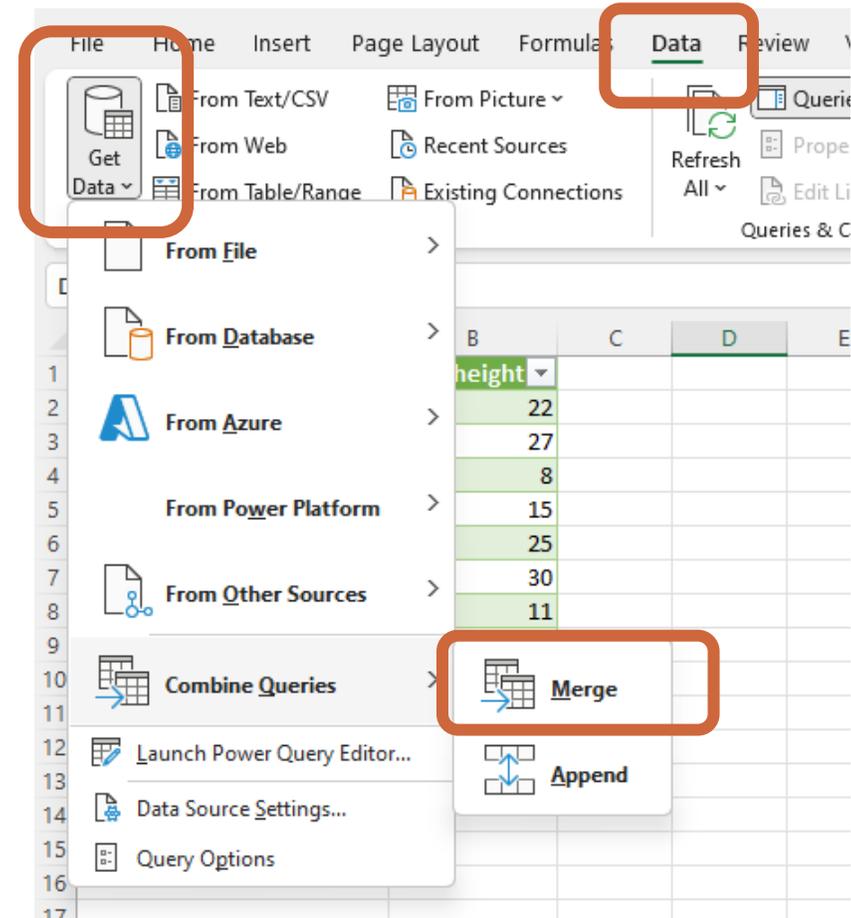


Reminder: the datasets have already been loaded into Power Query.

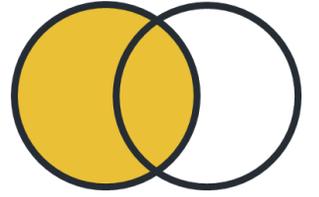
Step 1.

Click on

- **Data** ribbon then ...
- **Get Data** then ...
- **Combine Queries** then ...
- **Merge**



Select the datasets



Step 2.

Select the datasets you wish to join from the **drop down lists**.

You need to choose,

- the **left** dataset in the **top drop down list**
- the **right** dataset from the **bottom drop down list**.

Merge

Select tables and matching columns to create a merged table.

dog_height

Left dataset

Name	max_height
Border Collie	22
Irish Setter	27
Yorkshire Terrier	8
Poodle (Miniature)	15
Boxer	25

dog_good
dog_height

Right dataset

No preview is available

Join Kind

Left Outer (all from first, matching from second)

Use fuzzy matching to perform the merge

› Fuzzy matching options

OK Cancel

Identifying the key

Key

one or more columns that datasets have common

Step 3.

In each of the datasets **left click on the key** that you will use to join the datasets.

Merge

Select tables and matching columns to create a merged table.

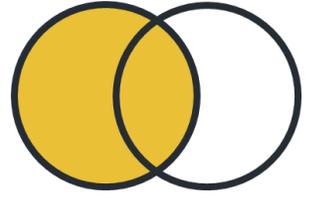
dog_height

Name	max_height
Border Collie	22
Irish Setter	27
Yorkshire Terrier	8
Poodle (Miniature)	15
Boxer	25

dog_good

Name	good_with_children	good_with_other_dogs
Border Collie	3	3
Irish Setter	5	5
Yorkshire Terrier	5	3
Boxer	5	3
Greyhound	3	4

Choose the join kind



Step 4.

Under the heading **Join Kind**, select **Left Outer** join option from the drop down list.

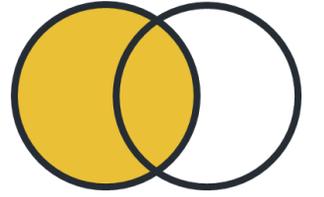
Then press **OK**.

Irish Setter	5	5
Yorkshire Terrier	5	3
Boxer	5	3
Greyhound	3	4

Join Kind

- Left Outer (all from first, matching from second)
- Left Outer (all from first, matching from second)
- Right Outer (all from second, matching from first)
- Full Outer (all rows from both)
- Inner (only matching rows)
- Left Anti (rows only in first)
- Right Anti (rows only in second)

Expand the column list



Step 5.

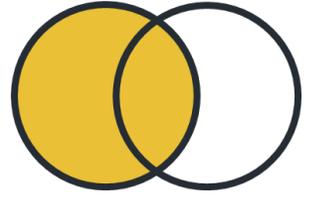
The Power Query Editor will have opened. The data from the left dataset will be visible.

You now need to decide which of the right hand dataset columns you wish to see.

To expand the list of right hand dataset columns **click on the double arrow icon** at the top right corner of the table.

	A ^B _C Name	1 ² ₃ max_height	dog_good
1	Border Collie	22	Table
2	Irish Setter	27	Table
3	Yorkshire Terrier	8	Table
4	Poodle (Miniature)	15	Table
5	Boxer	25	Table
6	Greyhound	30	Table
7	West Highland White Terrier	11	Table
8	Cocker Spaniel	16	Table
9	Great Dane	32	Table
10	Dalmatian	24	Table
11	Rottweiler	27	Table
12	Newfoundland	28	Table

Selecting the columns



Step 6.

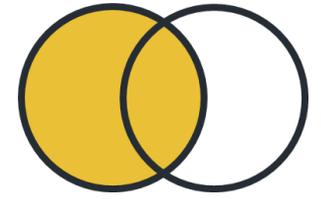
Once you have clicked on the double arrow icon you can choose which of the right hand dataset columns you would like to include.

In this case we are keeping **good_with_children** and **good_with_other_dogs**, by ticking the box next to the column names.

As a default, **'Use original column name as prefix'** is selected. This is not normally needed so you can un-tick this box.

	A ^B C Name	1 ² 3 max_height	dog_good
1	Border Collie		
2	Irish Setter		
3	Yorkshire Terrier		
4	Poodle (Miniature)		
5	Boxer		
6	Greyhound		
7	West Highland White Terrier		
8	Cocker Spaniel		
9	Great Dane		
10	Dalmatian		
11	Rottweiler		
12	Newfoundland		

Naming and loading the dataset



Step 7. The dataset has now been left joined. The final stage is to **name the dataset** under Properties of Query Settings and then **Close & Load** the dataset into an Excel workbook.

The screenshot shows the Power Query Editor interface. The 'File' tab is active, and the 'Close & Load' button is highlighted with a red box. The main area displays a table with 12 rows of dog data. The 'Query Settings' pane on the right is open, and the 'Name' field under 'PROPERTIES' is highlighted with a red box, containing the text 'dog_left_join'. The 'APPLIED STEPS' pane shows the 'Expanded dog_good' step.

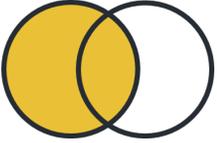
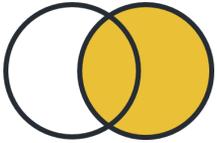
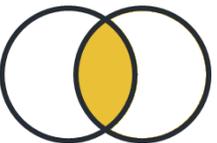
	Name	max_height	good_with_children	good_with_other_dogs
1	Border Collie	22	3	3
2	Irish Setter	27	5	5
3	Yorkshire Terrier	8	5	3
4	Poodle (Miniature)	15	null	null
5	Boxer	25	5	3
6	Greyhound	30	3	4
7	West Highland White Terrier	11	5	3
8	Cocker Spaniel	16	5	5
9	Dalmatian	24	3	3
10	Great Dane	32	null	null
11	Rottweiler	27	3	3
12	Newfoundland	28	5	5

Next steps

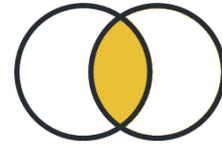
Complete **questions 1 to 4**
in **section 2** of the
'Practise combining datasets' workbook.

More joining options

In the Combining Dataset lesson we looked at the different join types. This table shows you the names that Power Query uses for these join types.

Join Type	Power Query Name
Left Join 	Left Outer (all from first, matching from second)
Right Join 	Right Outer (all from second, matching from first)
Inner join 	Inner (only matching rows)
Outer join 	Full Outer (all rows from both)

Show me... inner join



To complete an inner join, you follow the same steps as the left join except you select **Inner** as the **Join Kind**.

Merge

Select tables and matching columns to create a merged table.

dog_height

Name	max_height
Border Collie	22
Irish Setter	27
Yorkshire Terrier	8
Poodle (Miniature)	15
Boxer	25

dog_good

Name	good_with_children	good_with_other_dogs
Border Collie	3	3
Irish Setter	5	5
Yorkshire Terrier	5	3
Boxer	5	3
Greyhound	3	4

Join Kind

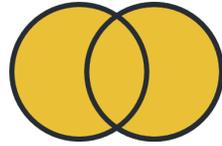
Inner (only matching rows)

Use fuzzy matching to perform the merge



	A ^B C Name	1 ² 3 max_height	1 ² 3 good_with_children	1 ² 3 good_with_other_dogs
1	Border Collie	22	3	3
2	Irish Setter	27	5	5
3	Yorkshire Terrier	8	5	3
4	Boxer	25	5	3
5	Greyhound	30	3	4
6	West Highland White Terrier	11	5	3
7	Cocker Spaniel	16	5	5
8	Dalmatian	24	3	3
9	Rottweiler	27	3	3
10	Newfoundland	28	5	5

Show me... outer join



To complete an outer join, you follow the same steps as the left join except you select **Full Outer** as the **Join Kind**.

Merge

Select tables and matching columns to create a merged table.

dog_height

Name	max_height
Border Collie	22
Irish Setter	27
Yorkshire Terrier	8
Poodle (Miniature)	15
Boxer	25

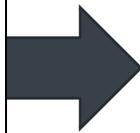
dog_good

Name	good_with_children	good_with_other_dogs
Border Collie	3	3
Irish Setter	5	5
Yorkshire Terrier	5	3
Boxer	5	3
Greyhound	3	4

Join Kind

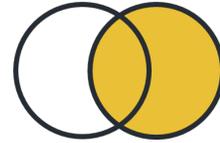
Full Outer (all rows from both)

Use fuzzy matching to perform the merge



	A ^B _C Name	1 ² ₃ max_height	A ^B _C Name.1	1 ² ₃ good_with_children
1	Border Collie	22	Border Collie	3
2	Irish Setter	27	Irish Setter	5
3	Yorkshire Terrier	8	Yorkshire Terrier	5
4	Poodle (Miniature)	15		null
5	Boxer	25	Boxer	5
6	Greyhound	30	Greyhound	3
7	West Highland White Terrier	11	West Highland White Terrier	5
8	Cocker Spaniel	16	Cocker Spaniel	5
9	Dalmatian	24	Dalmatian	3
10	Great Dane	32		null
11	Rottweiler	27	Rottweiler	3
12	Newfoundland	28	Newfoundland	5

Show me... right join



To complete a right join, you follow the same steps as the left join except you select **Right Outer** as the **Join Kind**.

Merge

Select tables and matching columns to create a merged table.

dog_height

Name	max_height
Border Collie	22
Irish Setter	27
Yorkshire Terrier	8
Poodle (Miniature)	15
Boxer	25

dog_good

Name	good_with_children	good_with_other_dogs
Border Collie	3	3
Irish Setter	5	5
Yorkshire Terrier	5	3
Boxer	5	3
Greyhound	3	4

Join Kind

Right Outer (all from second, matching from first)



	A ^B Name	1 ² 3 max_height	1 ² 3 good_with_children	1 ² 3 good_with_other_dogs
1	Border Collie	22	3	3
2	Irish Setter	27	5	5
3	Yorkshire Terrier	8	5	3
4	Boxer	25	5	3
5	Greyhound	30	3	4
6	West Highland White Terrier	11	5	3
7	Cocker Spaniel	16	5	5
8	Dalmatian	24	3	3
9	Rottweiler	27	3	3
10	Newfoundland	28	5	5

Next steps

Complete **questions 1 to 7**
in **section 3** of the
'Practise combining datasets' workbook.

Learning checklist

I can *use* Power Query Editor in Excel to append rows to a simple dataset.

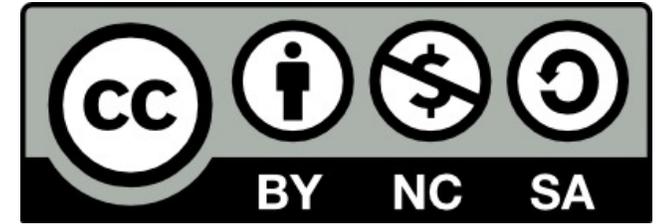
I can *use* Power Query Editor in Excel to left join simple datasets.

I can *use* Power Query Editor in Excel to right join simple datasets.

I can *use* Power Query Editor in Excel to inner join simple datasets.

I can *use* Power Query Editor in Excel to outer join simple datasets.

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.

© 2023. This work is licensed under a [CC BY-NC-SA 4.0 license](#).

Created by effini in partnership with The Data Lab.



Alternative format

If you require this document in an alternative format, such as large print or a coloured background, please contact

hello@effini.com

or

4th Floor, The Bayes Centre
47 Potterrow
Edinburgh
EH8 9BT

