Data Organisation & Analysis

Data Organisation

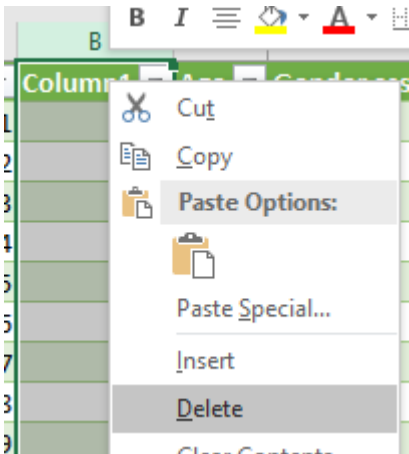| Date:	18-01-22
Section worked on:	Importing Data
Work Completed:	

We've been asked to import data from a **text** file into an excel spreadsheet. This file **will** have empty data, which we have been tasked to sort out. We will be:

- Making sure there are no duplicated records, unnecessary spaces or blank cells
- Making sure the mean (average) is the num. of hours spent taking part in outdoor activities **rounded to a whole number**
- The **PREFERENCES** column is identified in uppercase
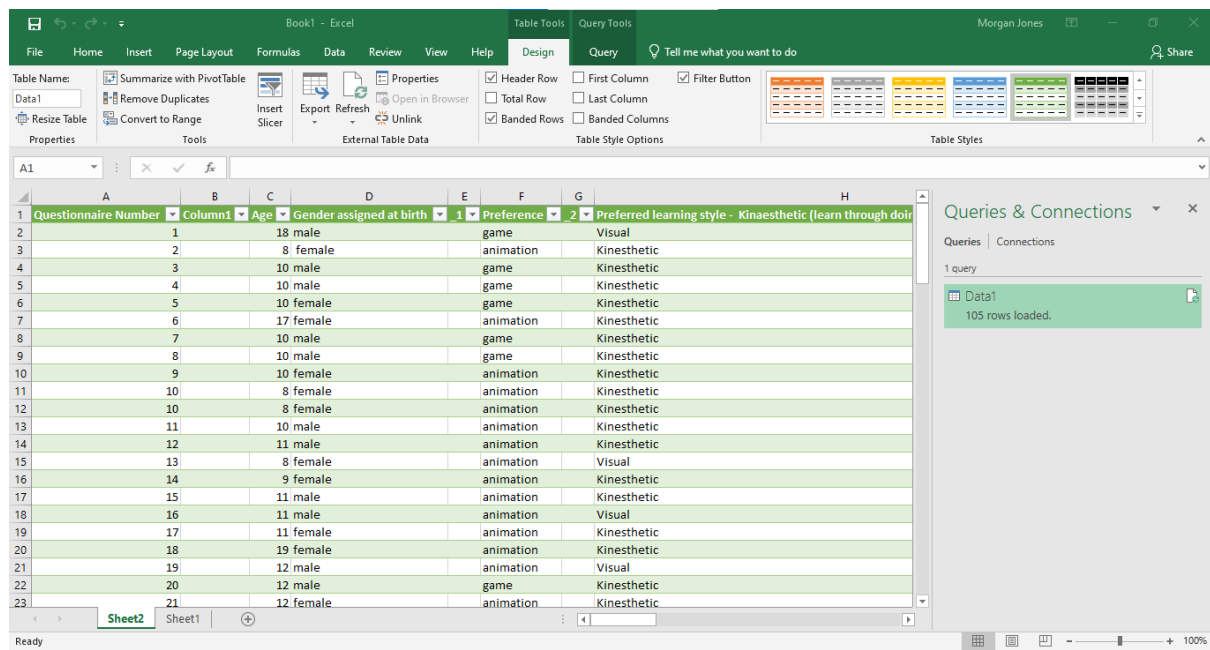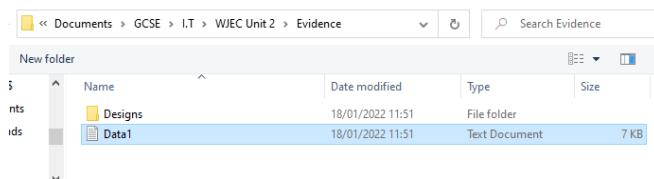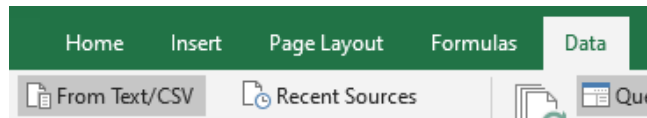
Removing Blank Columns

To remove a column, right click at the top of it (Where it says A, B, C, etc.) and press delete.

Before:

	A Questionnaire Number	B Column1	C Age	D Gender assigned at birth	E _1	F Preference	G _2	H Preferred learning style – Kinaesthetic (learn through doing and moving) Visual (learn by	Average number
2	1		18	male		game		Visual	
3	2		8	female		animation		Kinesthetic	
4	3		10	male		game		Kinesthetic	
5	4		10	male		game		Kinesthetic	
6	5		10	female		game		Kinesthetic	
7	6		17	female		animation		Kinesthetic	
8	7		10	male		game		Kinesthetic	
9	8		10	male		game		Kinesthetic	
10	9		10	female		animation		Kinesthetic	
11	10		8	female		animation		Kinesthetic	
12	10		8	female		animation		Kinesthetic	
13	11		10	male		animation		Kinesthetic	
14	12		11	male		animation		Kinesthetic	
15	13		8	female		animation		Visual	
16	14		9	female		animation		Kinesthetic	
17	15		11	male		animation		Kinesthetic	
18	16		11	male		animation		Visual	
19	17		11	female		animation		Kinesthetic	
20	18		19	female		animation		Kinesthetic	
21	19		12	male		animation		Visual	
22	20		12	male		game		Kinesthetic	
23	21		12	female		animation		Kinesthetic	

After:

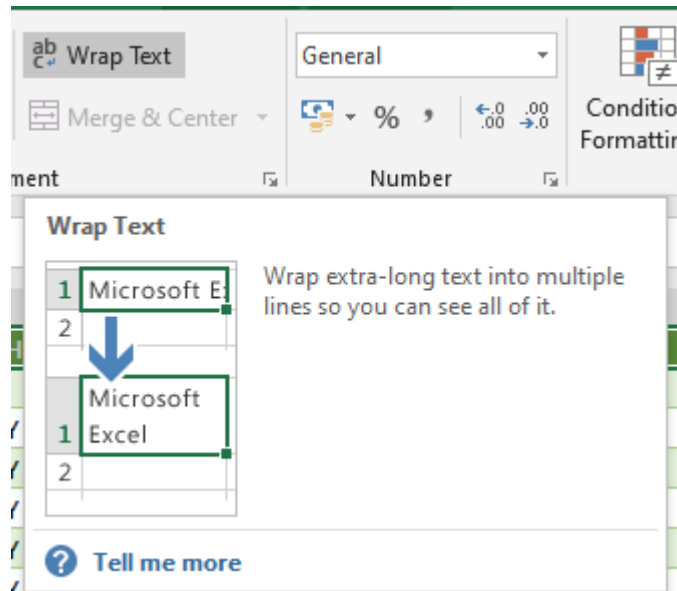|	A Questionnaire Number	B Age	C Gen	D Pre	E Preferred le	F Av	G Have you ever been camping?	H Activity Choice
2	1	18	male	game	Visual	4		Climbing
3	2	8	female	anima	Kinesthetic	3 Y		Paddle boarding
4	3	10	male	game	Kinesthetic	6 Y		Paddle boarding
5	4	10	male	game	Kinesthetic	3 Y		Paddle boarding
6	5	10	female	game	Kinesthetic	7 Y		Paddle boarding
7	6	17	female	anima	Kinesthetic	8 Y		Paddle boarding
8	7	10	male	game	Kinesthetic	1 N		Paddle boarding
9	8	10	male	game	Kinesthetic	6 Y		Paddle boarding
10	9	10	female	anima	Kinesthetic	6 Y		Paddle boarding
11	10	8	female	anima	Kinesthetic	2 N		Kite Surfing
12	10	8	female	anima	Kinesthetic	2 N		Kite Surfing
13	11	10	male	anima	Kinesthetic	3 N		Kite Surfing
14	12	11	male	anima	Kinesthetic	4 N		Kite Surfing
15	13	8	female	anima	Visual	3 N		Kite Surfing
16	14	9	female	anima	Kinesthetic	3 N		Kite Surfing
17	15	11	male	anima	Kinesthetic	0 N		Paddle boarding
18	16	11	male	anima	Visual	1 N		Paddle boarding
19	17	11	female	anima	Kinesthetic	8 Y		Paddle boarding
20	18	19	female	anima	Kinesthetic	8 Y		Paddle boarding
21	19	12	male	anima	Visual	2 N		Paddle boarding
22	20	12	male	game	Kinesthetic	8 Y		Paddle boarding
23	21	12	female	anima	Kinesthetic	8 Y		Paddle boarding

Sheet2   Sheet1

Text Wrapping Titles

By text wrapping our titles, it allows us to increase the height of our titles and make them thinner so we don't have to scroll horizontally across our spreadsheet.

First, we will select all of our titles in the first row

And press **Home > Wrap Text**:

After:

Obviously, this top row is too tall. The main culprit is column F. We will now change the width of all of these columns to make the spreadsheet titles more appropriate to the titles and the columns stored data.

This is much better because now we can see all the columns in the spreadsheet without having to scroll.

Removing Duplicates

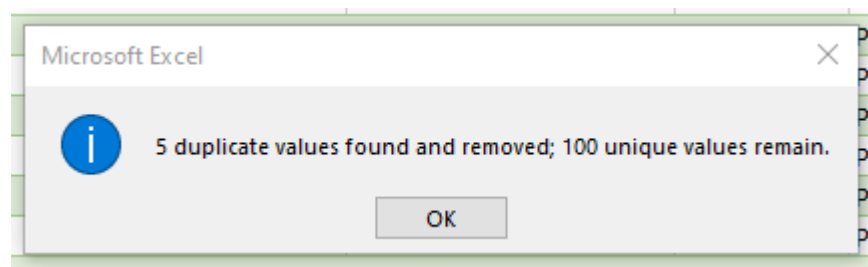Firstly, we will highlight every cell with **Ctrl + A**

And then we go to **Data > Remove Duplicates** and press it

Press **OK** on the next two windows
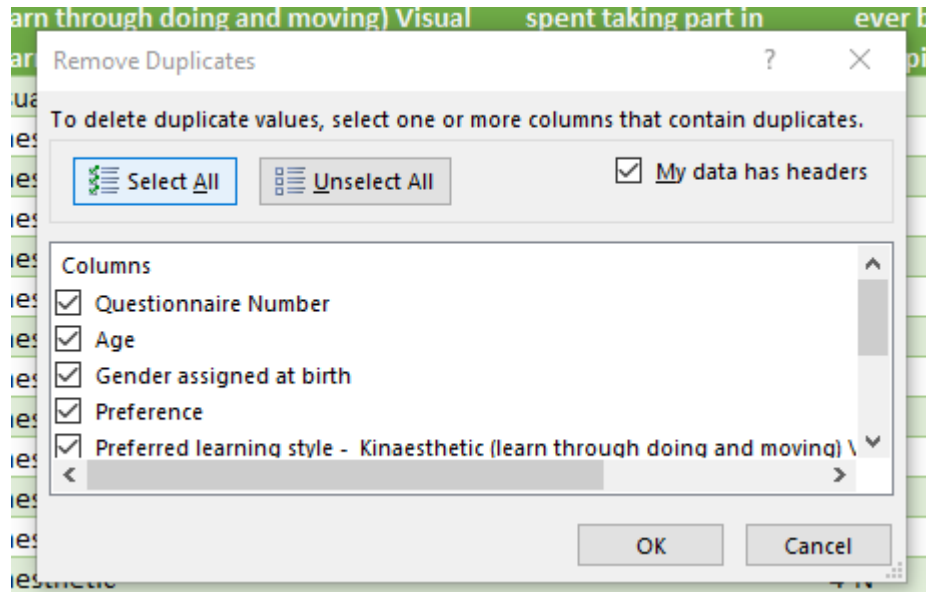
Removing Blank Cells

Once again, highlight all the data with **Ctrl + A**

Then, we will click on **Home > Find & Select > Go To Special…**
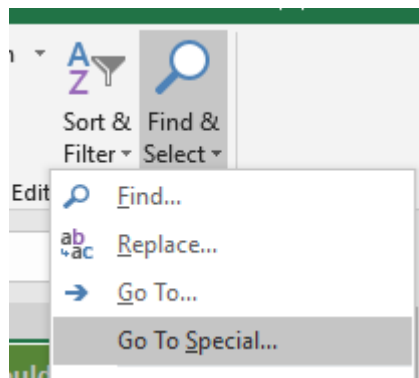
Select **Blanks** and click **OK**
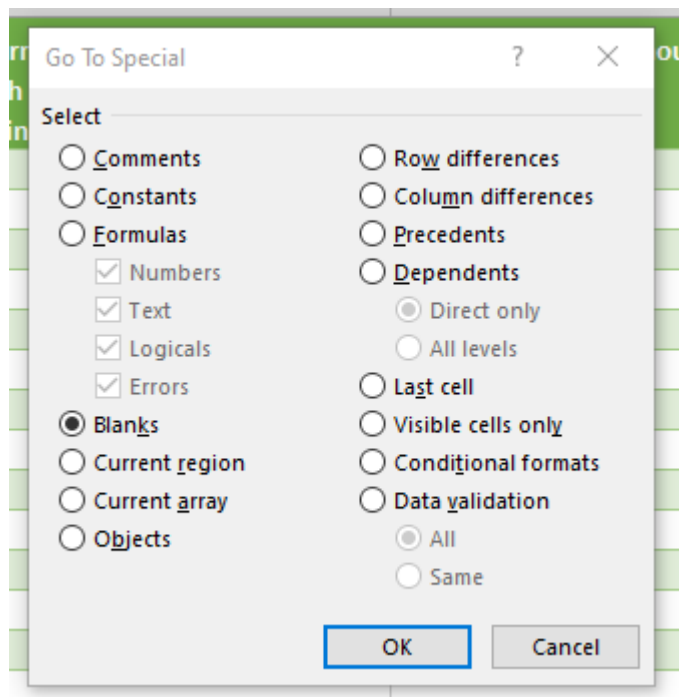
Date:	20-01-22
Section worked on:	Roundup Mean Number & Changing Preferences to Upper Case
Work Completed:	

Reducing decimals on the Average column

First, we will click on the column **letter** (F) to select the whole column

Then, under the Number section, press the expand button
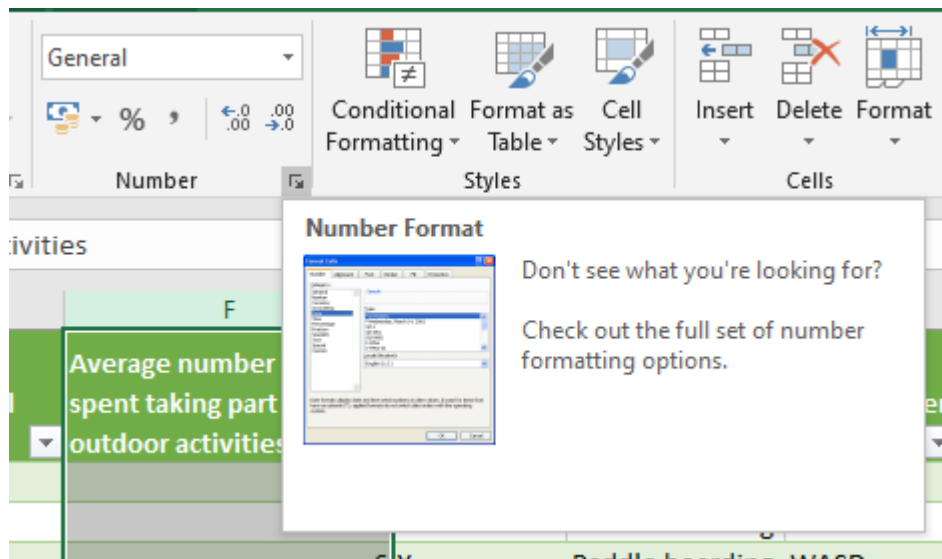
In the window that pops up, click on **Number** and set **Decimal places** to 0
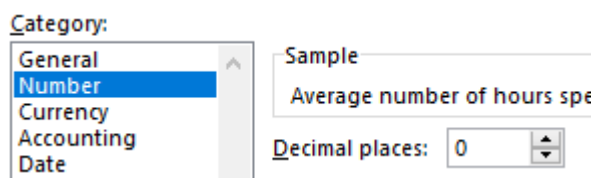
Change the **Preferences** column to upper case

We select the column **K** and type **=upper(E2:E101)** in the formula field and press **Enter**
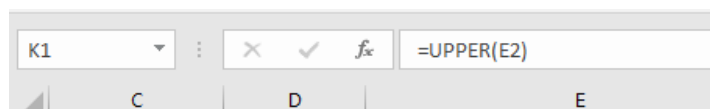
Our spreadsheet should now look something like this

Gender assigned at birth	Preference	Preferred learning style - Kinaesthetic (learn through doing and moving) Visual (learn by seeing)	Average number of hours spent taking part in outdoor activities	Have you ever been camping?	Activity Choice	Game Player Movement	Where should the game start and end? 1= Start Inside and End Outside. 2 = Start Outside and End Inside.	0
male	game	Visual	4		Climbing	WASD		2 VISUAL
female	animation	Kinesthetic	3	Y	Paddle boarding	WASD		2 KINESTHETIC
male	game	Kinesthetic	6	Y	Paddle boarding	WASD		1 KINESTHETIC
male	game	Kinesthetic	3	Y	Paddle boarding	WASD		2 KINESTHETIC
female	game	Kinesthetic	7	Y	Paddle boarding	ARROW KEYS		2 KINESTHETIC
female	animation	Kinesthetic	8	Y	Paddle boarding	WASD		2 KINESTHETIC
male	game	Kinesthetic	1	N	Paddle boarding	WASD		1 KINESTHETIC
male	game	Kinesthetic	6	Y	Paddle boarding	ARROW KEYS		1 KINESTHETIC
female	animation	Kinesthetic	6	Y	Paddle boarding	ARROW KEYS		1 KINESTHETIC
female	animation	Kinesthetic	2	N	Kite Surfing	WASD		2 KINESTHETIC
male	animation	Kinesthetic	3	N	Kite Surfing	WASD		2 KINESTHETIC
male	animation	Kinesthetic	4	N	Kite Surfing	WASD		1 KINESTHETIC
female	animation	Visual	3	N	Kite Surfing	WASD		2 VISUAL
female	animation	Kinesthetic	3	N	Kite Surfing	WASD		2 KINESTHETIC
male	animation	Kinesthetic	0	N	Paddle boarding	ARROW KEYS		2 KINESTHETIC
male	animation	Visual	1	N	Paddle boarding	ARROW KEYS		2 VISUAL
female	animation	Kinesthetic	8	Y	Paddle boarding	ARROW KEYS		1 KINESTHETIC
female	animation	Kinesthetic	8	Y	Paddle boarding	WASD		1 KINESTHETIC
male	animation	Visual	2	N	Paddle boarding	ARROW KEYS		1 VISUAL
male	game	Kinesthetic	0	Y	Paddle boarding	ARROW KEYS		1 KINESTHETIC

Now, we select from **K2** to **K101** and press **Ctrl+C**
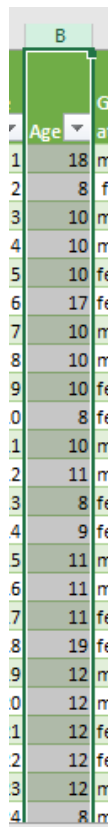
K
l=
0
2 VISUAL
2 KINESTHETIC
1 KINESTHETIC
2 KINESTHETIC
2 KINESTHETIC
2 KINESTHETIC
1 KINESTHETIC
1 KINESTHETIC
1 KINESTHETIC
2 KINESTHETIC
2 KINESTHETIC
1 KINESTHETIC
2 VISUAL
2 KINESTHETIC
2 KINESTHETIC
2 VISUAL
1 KINESTHETIC
1 KINESTHETIC
1 VISUAL
1 KINESTHETIC
1 KINESTHETIC
1 KINESTHETIC
1 KINESTHETIC
1 KINESTHETIC

Then right click on **E2** and press **Paste Values**

Preferred learning style - Kinaesthetic (learn through doing and moving) Visual (learn by seeing)	Avera spent outdo
VISUAL	
KINESTHETIC	
KINESTHETIC	
KINESTHETIC	
KINESTHETIC	
KINESTHETIC	

Paste Options:

Values (V)

	E
	Preferred learning style - Kinaesthetic (learn through doing and moving) Visual (learn by seeing)
	VISUAL
	KINESTHETIC
	KINESTHETIC
	KINESTHETIC
	KINESTHETIC
	KINESTHETIC
	KINESTHETIC
	KINESTHETIC
	KINESTHETIC
	KINESTHETIC
	KINESTHETIC
	KINESTHETIC
	VISUAL

We can now right click on **column K** and press **Delete**

The spreadsheet now looks like this

Data Analysis

Date:	27-01-22
Section worked on:	Sorting by age
Work Completed:	

First, select the **Age** column

Then go to **Data > Sort**

Press **OK**

	A	B	C	D	E	F	G	H	I	J
1	Questionnaire Number	Age	Gender assigned at birth	Preference	Preferred learning style - Kinaesthetic (learn through doing and moving) Visual (learn by seeing)	Average number of hours spent taking part in outdoor activities	Have you ever been camping?	Activity Choice	Game Player Movement	Where should the game start and end? 1= Start Inside and End Outside. 2 = Start Outside and End Inside.
2	25	7	female	game	KINESTHETIC		1	Kite Surfing	ARROW KEYS	2
3	100	7	female	game	VISUAL		7 Y	Climbing	WASD	2
4	2	8	female	animation	KINESTHETIC		3 Y	Paddle boarding	WASD	2
5	10	8	female	animation	KINESTHETIC		2 N	Kite Surfing	WASD	2
6	13	8	female	animation	VISUAL		3 N	Kite Surfing	WASD	2
7	24	8	male	game	KINESTHETIC		1 N	Kite Surfing	ARROW KEYS	1
8	14	9	female	animation	KINESTHETIC		3 N	Kite Surfing	WASD	2
9	3	10	male	game	KINESTHETIC		6 Y	Paddle boarding	WASD	1
10	4	10	male	game	KINESTHETIC		3 Y	Paddle boarding	WASD	2
11	5	10	female	game	KINESTHETIC		7 Y	Paddle boarding	ARROW KEYS	2
12	7	10	male	game	KINESTHETIC		1 N	Paddle boarding	WASD	1
13	8	10	male	game	KINESTHETIC		6 Y	Paddle boarding	ARROW KEYS	1
14	9	10	female	animation	KINESTHETIC		6 Y	Paddle boarding	ARROW KEYS	1
15	11	10	male	animation	KINESTHETIC		3 N	Kite Surfing	WASD	2
16	26	10	female	game	KINESTHETIC		3	Kite Surfing	WASD	2
17	27	10	male	game	KINESTHETIC		2 N	Kite Surfing	ARROW KEYS	2
18	28	10	male	animation	KINESTHETIC		3 Y	Kite Surfing	ARROW KEYS	2
19	29	10	male	game	KINESTHETIC		4	Kite Surfing	ARROW KEYS	2
20	30	10	female	game	KINESTHETIC		5 Y	Kite Surfing	WASD	2
21	34	10	male	game	KINESTHETIC		7 Y	Kayaking	ARROW KEYS	2
22	35	10	male	game	KINESTHETIC		2 N	Climbing	ARROW KEYS	2
23	36	10	male	game	VISUAL		2 N	Kite Surfing	ARROW KEYS	2
24	40	10	female	game	KINESTHETIC		2	Kite Surfing	ARROW KEYS	2
25	41	10	female	game	VISUAL		3 Y	Kite Surfing	WASD	2

Analysing ages

Starting from L2, we will create this table

Age Analysis	
Average	
Youngest	
Oldest	
Common Age	
Middle of age group	

We then add in these formulae into the fields

Age Analysis	
Average	=ROUND(AVERAGE(B:B),0)
Youngest	=MIN(B:B)
Oldest	=MAX(B:B)
Common Age	=MODE(B:B)
Middle of age group	=MEDIAN(B:B)

Age Analysis	
Average	13
Youngest	7
Oldest	19
Common Age	10
Middle of age group	12

From this information I can tell that the average age of people that go to Adventure Cymru is 13, the youngest is 7, the oldest is 19, the most common age is 10 and the middle age is 12.

=ROUND(x,y) rounds the first argument to the decimal place of the second argument

=MIN(B:B) gets the smallest value from the cells or arguments

=MAX(B:B) get the biggest value from the cells or arguments

=MODE(B:B) gets the most used value from the cells or arguments

=MEDIAN(B:B) gets the middle-most value from the cells or arguments

Analysing sex

To do this, we will create a table under the age table

Sex	
Male	
Female	

We will now use the **COUNTIF** function.

=COUNTIF(xx:yx, z) will look between the columns **xx** and **yx** and if the value of **z** is found in a cell, it will increment the count.

Sex	
Male	=COUNTIF(C:C, "male")
Female	=COUNTIF(C:C, "female")

Sex	
Male	46
Female	43

This information tells us that 46 males and 43 females are participating in Adventure Cymru.

Showing how many people are doing each activity with a pivot table

To use this, we will use a **pivot table**

Click on **Insert > Pivot Table**

This window will pop up:
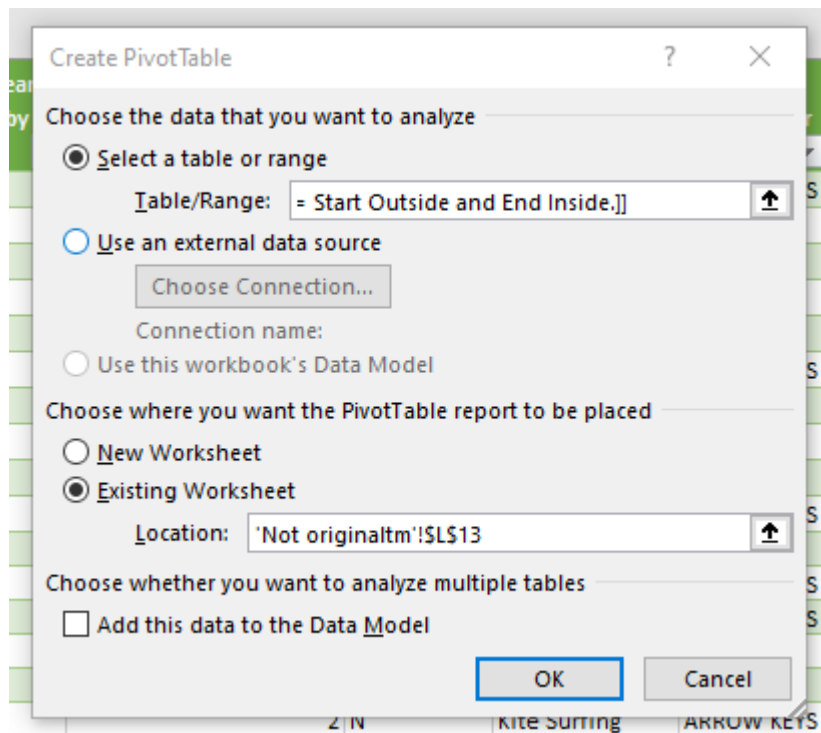
Click on the up arrow under **Select a table or range** and select the whole table including headings

Click on the up arrow under **Existing worksheet** and select a cell under the **Sex** table

Press **OK**

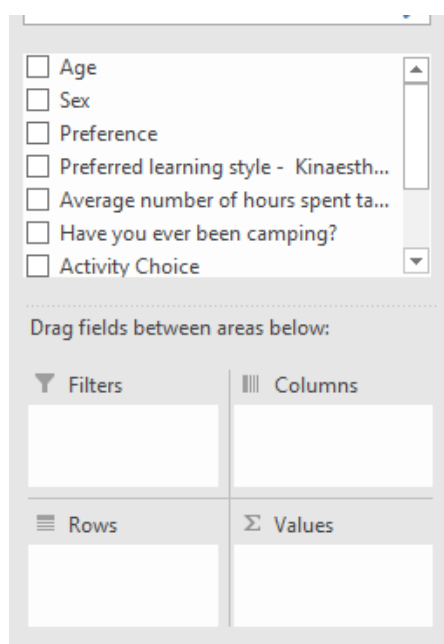We can now drag these values down to create the pivot table

Example:

Age Analysis	
Average	13
Youngest	7
Oldest	19
Common Age	10
Middle of age	12

Sex	
Male	46
Female	43

Row Labels	Count of Game Player Movement
ARROW KEYS	52
WASD	48
Grand Total	100

This tells us that 52 people prefer using the arrow keys and the other 48 prefer WASD

Youngest	7
Oldest	19
Common Age	10
Middle of age	12

Sex	
Male	46
Female	43

Row Labels	Count of Game Player Movement
ARROW KEYS	52
WASD	48
Grand Total	100

Row Labels	Count of Activity Choice
Climbing	32
Kayaking	10
Kite Surfing	40
Paddle boardi	18
Grand Total	100

This shows us that 32 people enjoy climbing, 10 enjoy kayaking, 40 enjoy kite surfing and 18 enjoy paddle boarding

Youngest	7
Oldest	19
Common Age	10
Middle of age	12

Sex	
Male	46
Female	43

Row Labels	Count of Game Player Movement
ARROW KEYS	52
WASD	48
Grand Total	100

Row Labels	Count of Activity Choice
Climbing	32
Kayaking	10
Kite Surfing	40
Paddle boardi	18
Grand Total	100

Row Labels	Count of Preference
animation	51
game	49
Grand Total	100

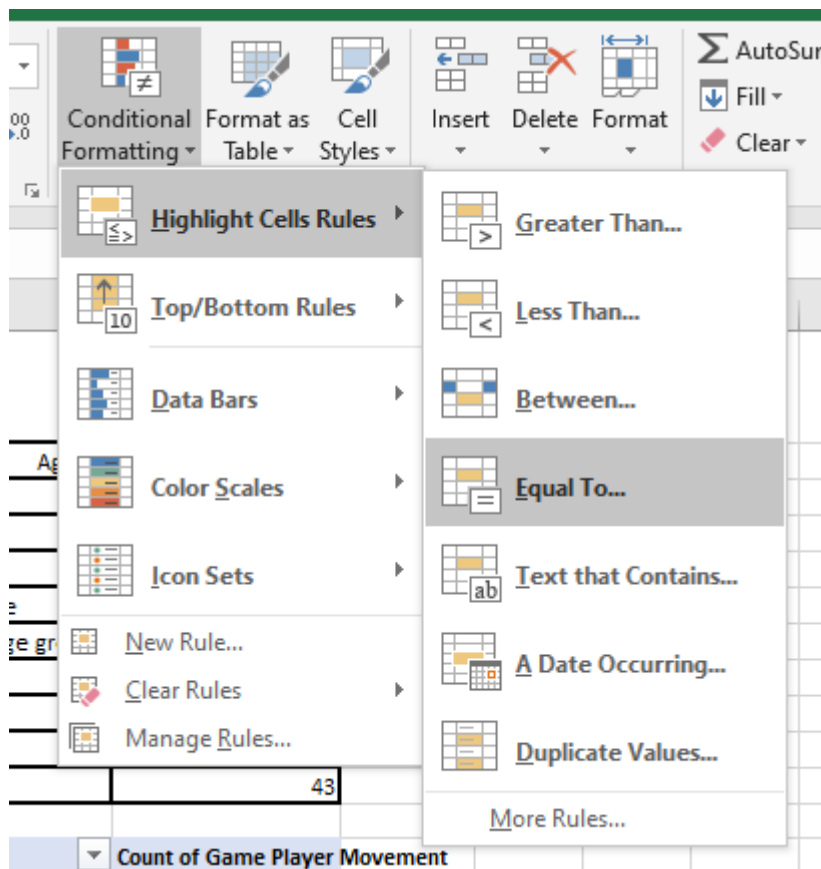This shows us that 51 people prefer animation over the other 49 that prefer games.

Conditional formatting

Conditional formatting is a non-numerical way of visualising information.

First, we select the whole of the last column of data (excluding the header)
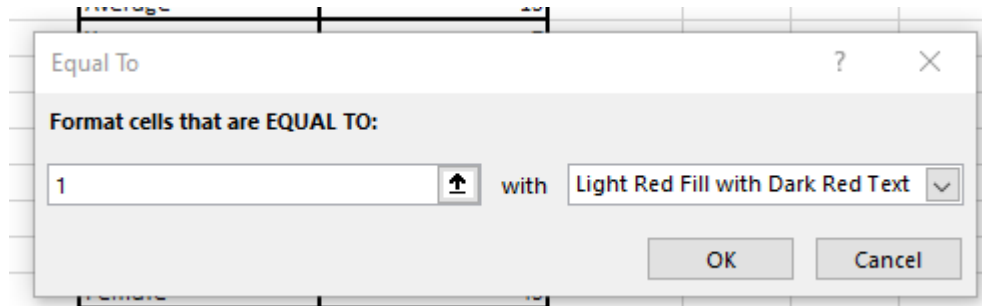

We then go to **Home > Conditional Formatting > Highlight Cell Rules > Equal To...**

We will then enter 1 **OR** 2 into the left input field then choose our styling on the right.

We will make 1 = Orange and 2 = Yellow

Press **OK**

With the column still selected, press **Equal To…**, type in 2 and make the colour yellow

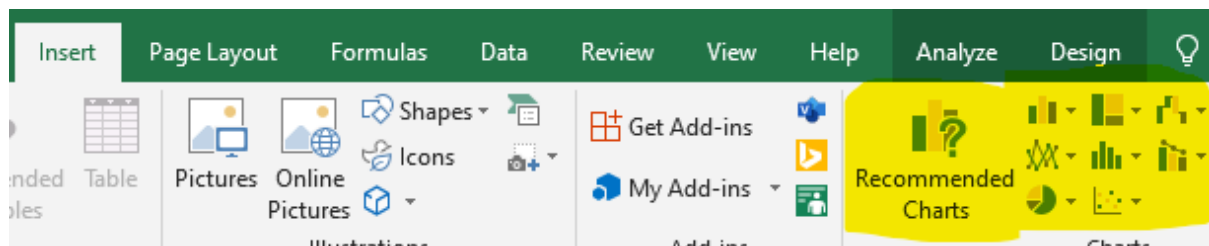The column should now look like this (Red has been changed to orange)

J
Where should the game start and end? 1= Start Inside and End Outside. 2 = Start Outside and End Inside.
2
2
2
2
2
1
2
1
2
2
1
1
1
2
2
2
2
2
2
2
2
2
2

Creating a Table of Activities

First, we will select **inside** the activity pivot table

Row Labels	Count of Activity Choice
Climbing	32
Kayaking	10
Kite Surfing	40
Paddle boarding	18
Grand Total	**100**

Then go to **Insert** and select a type of table

And voila! You have a table!