

DATA AND ARTIFICIAL INTELLIGENCE



Business Analytics with Excel



Data Cleaning and Preparation

Learning Objectives

By the end of this lesson, you will be able to:

- 🕒 Implement sort and filter functionalities to order or filter data
- 🕒 Organize the data using group by and ungroup functions
- 🕒 Execute Remove duplicates function to rid the data of duplicates
- 🕒 Implement data validation function to a given data



A Day in the Life of Business Analyst

As a business analyst of an organization:

You are required to sort and filter data. Also, improper data needs to be eliminated and data must be cleaned and be meaningful which serves the business purpose of the organization

To achieve these tasks, you will be learning a few concepts, such as sort, filter, group by, subtotal and removing duplicates.



Sort and Filter

Sort and Filter

The sort and filter functionalities are available in Excel to order or filter the data for further analysis.



Sort and Filter

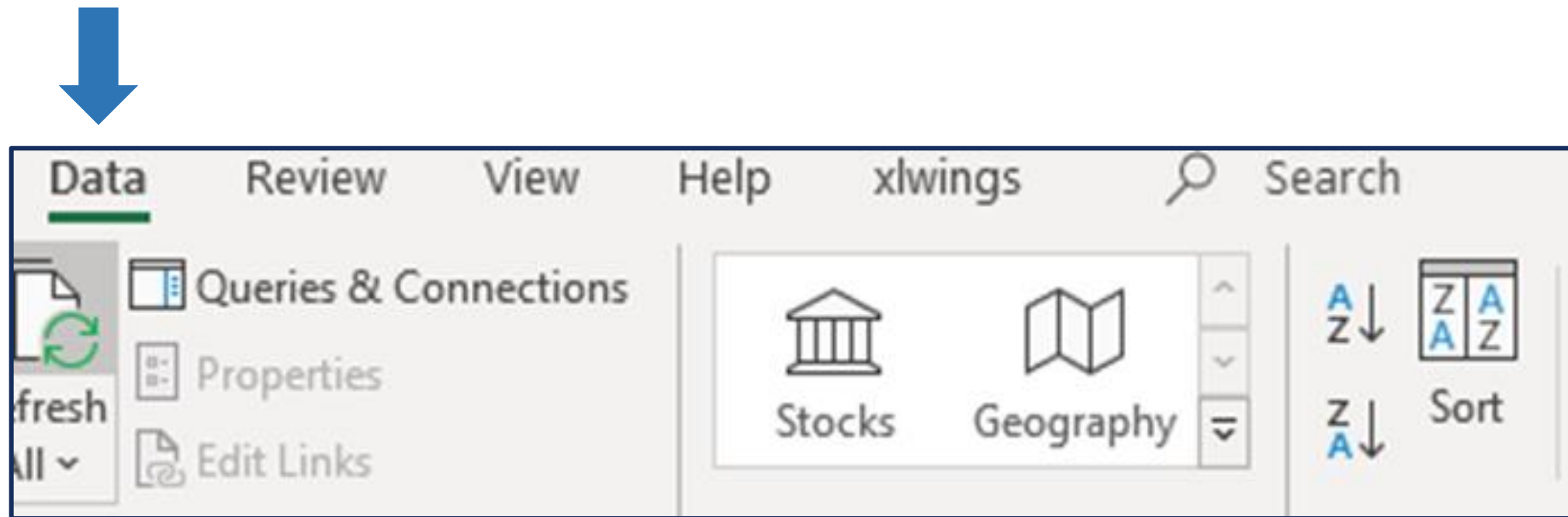
Example: The results of nine students for the Maths subject.

Name	Subject	CGPA
Alison Cox	Maths	4.3
Anise Jeff	Maths	4.2
Beatrice Cane	Maths	3.4
Danish Xavier	Maths	3.9
Rada Hofman	Maths	4.1
James Alan	Maths	4.8
Xavier Alex	Maths	4.6
Albert Dane	Maths	4.8
Hassan Alburi	Maths	4.6



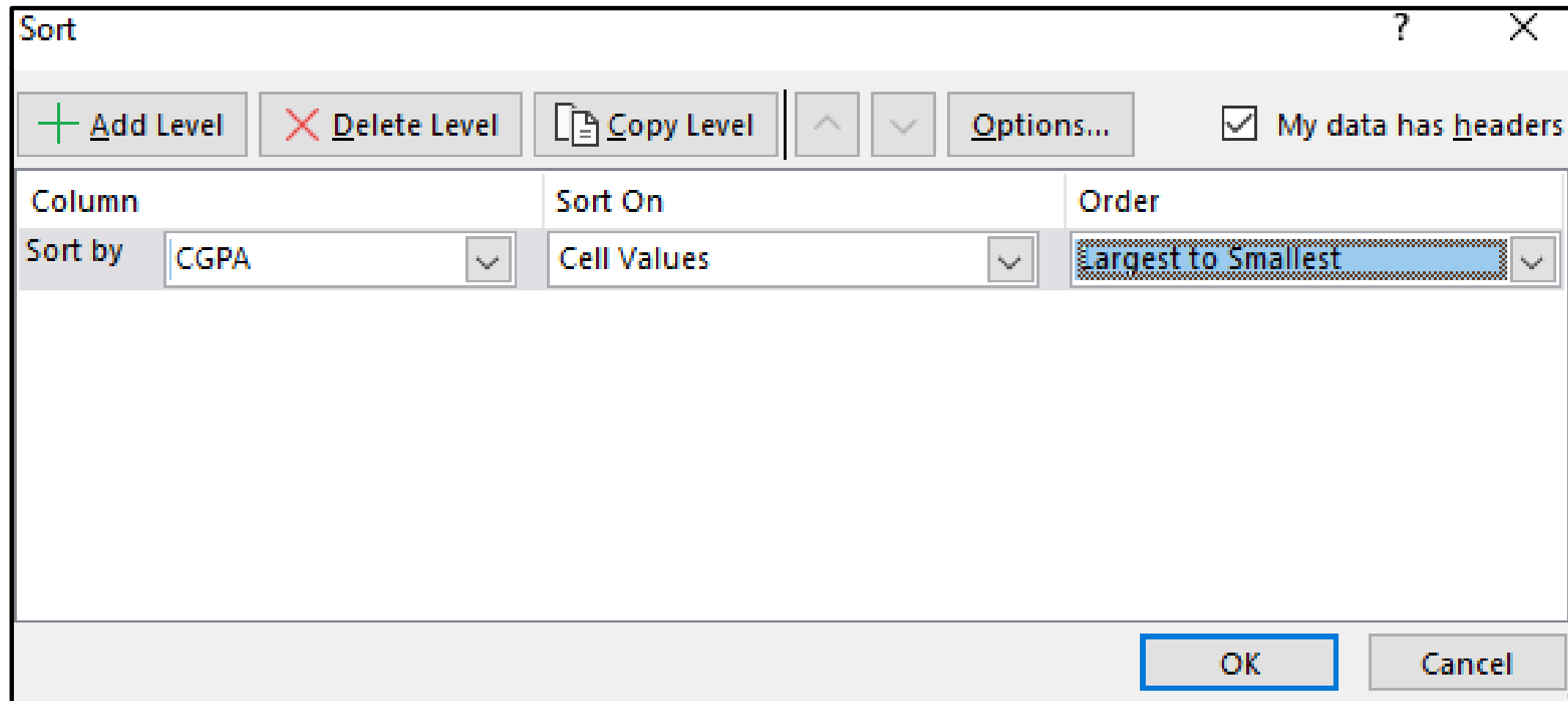
Steps to Sort Value

To sort the data based on the CGPA in descending order, choose the CGPA column name and then click on Sort under the Data tab



Steps to Sort Value

Under sort columns, choose CGPA and then select 'Largest to Smallest'



The screenshot shows the 'Sort' dialog box in Microsoft Excel. The dialog has a title bar with a question mark and a close button. Below the title bar are buttons for '+ Add Level', 'X Delete Level', 'Copy Level', and 'Options...'. There is also a checkbox labeled 'My data has headers' which is checked. The main area of the dialog is divided into three columns: 'Column', 'Sort On', and 'Order'. Under 'Column', the 'Sort by' dropdown is set to 'CGPA'. Under 'Sort On', the dropdown is set to 'Cell Values'. Under 'Order', the dropdown is set to 'Largest to Smallest'. At the bottom right are 'OK' and 'Cancel' buttons. A blue arrow points to the 'Sort by' dropdown, and another blue arrow points to the 'Order' dropdown. A faint illustration of a robot is visible in the background on the right side.

Column	Sort On	Order
Sort by: CGPA	Cell Values	Largest to Smallest

Steps to Sort Value

These are the sorted values:

Name	Subject	CGPA
James Alan	Maths	4.8
Albert Dane	Maths	4.8
Xavier Alex	Maths	4.6
Hassan Alhuri	Maths	4.6
Alison Cox	Maths	4.3
Anise Jeff	Maths	4.2
Rada Hofman	Maths	4.1
Danish Xavier	Maths	3.9
Beatrice Cane	Maths	3.4



Steps to Sort Value

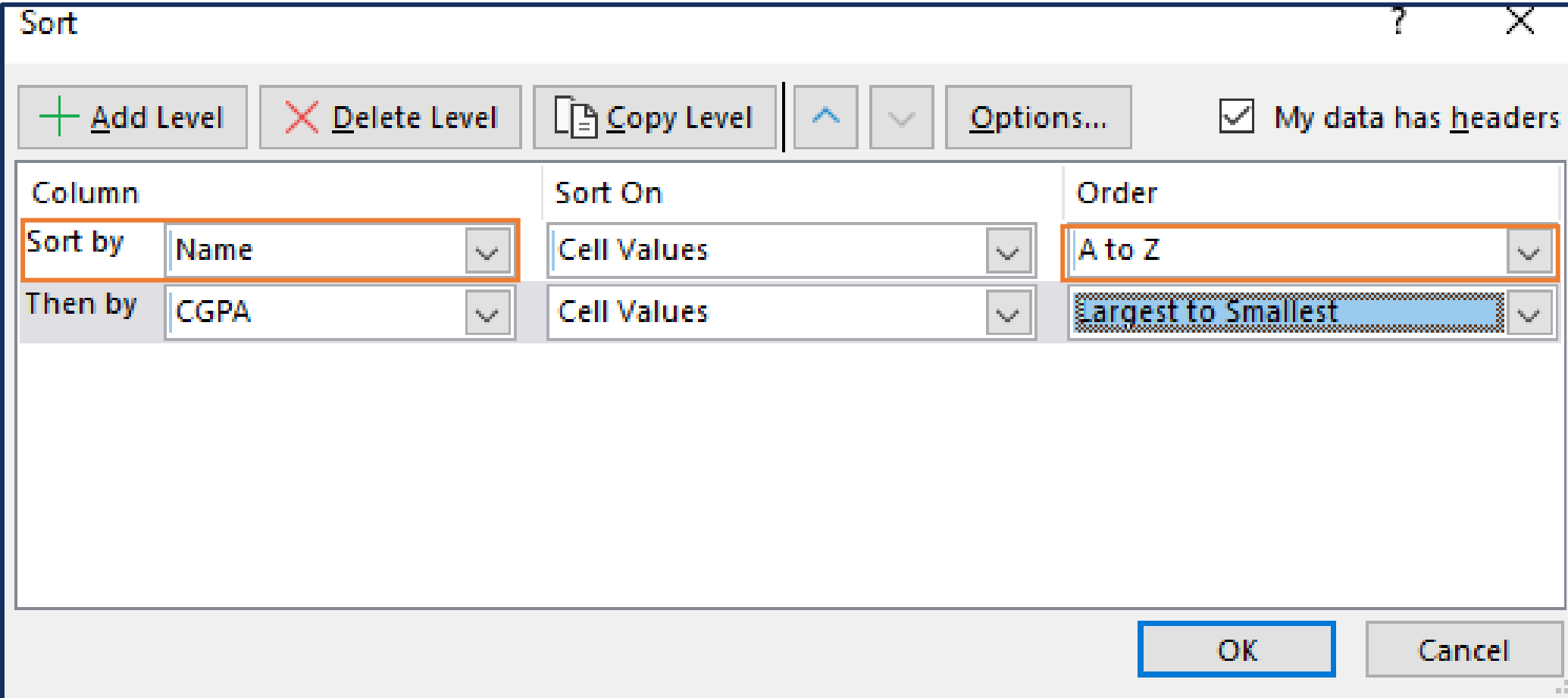
Sorting can be done based on character values from either A-Z or Z-A.

Name	Subject	CGPA
Alison Cox	Maths	4.3
Anise Jeff	Maths	4.2
Beatrice Cane	Maths	3.4
Danish Xavier	Maths	3.9
Rada Hofman	Maths	4.1
James Alan	Maths	4.8
Xavier Alex	Maths	4.6
Albert Dane	Maths	4.8
Hassan Alhuri	Maths	4.6

Any type of data
can be sorted
based on multiple
columns

Steps to Sort Value

Under the Sort by tab, select Name and choose order as A to Z and click OK



The screenshot shows the 'Sort' dialog box in Microsoft Excel. The dialog has a title bar with a question mark and a close button. Below the title bar are buttons for '+ Add Level', '- Delete Level', and 'Copy Level', along with up and down arrow buttons and an 'Options...' button. A checkbox labeled 'My data has headers' is checked. The main area is divided into two sections: 'Sort by' and 'Then by'. The 'Sort by' section has three dropdowns: 'Column' (set to 'Name'), 'Sort On' (set to 'Cell Values'), and 'Order' (set to 'A to Z'). The 'Then by' section has three dropdowns: 'Column' (set to 'CGPA'), 'Sort On' (set to 'Cell Values'), and 'Order' (set to 'Largest to Smallest'). At the bottom are 'OK' and 'Cancel' buttons. A blue arrow points to the 'Sort by' section, and another blue arrow points to the 'Order' dropdown in the 'Sort by' section.

Column	Sort On	Order
Sort by: Name	Cell Values	A to Z
Then by: CGPA	Cell Values	Largest to Smallest

Steps to Sort Value

The results will be in the following order.

Name	Subject	CGPA
Albert Dane	Maths	4.8
Alison Cox	Maths	4.3
Anise Jeff	Maths	4.2
Beatrice Cane	Maths	3.4
Danish Xavier	Maths	3.9
Hassan Alhuri	Maths	4.6
James Alan	Maths	4.8
Rada Hofman	Maths	4.1
Xavier Alex	Maths	4.6



Filter

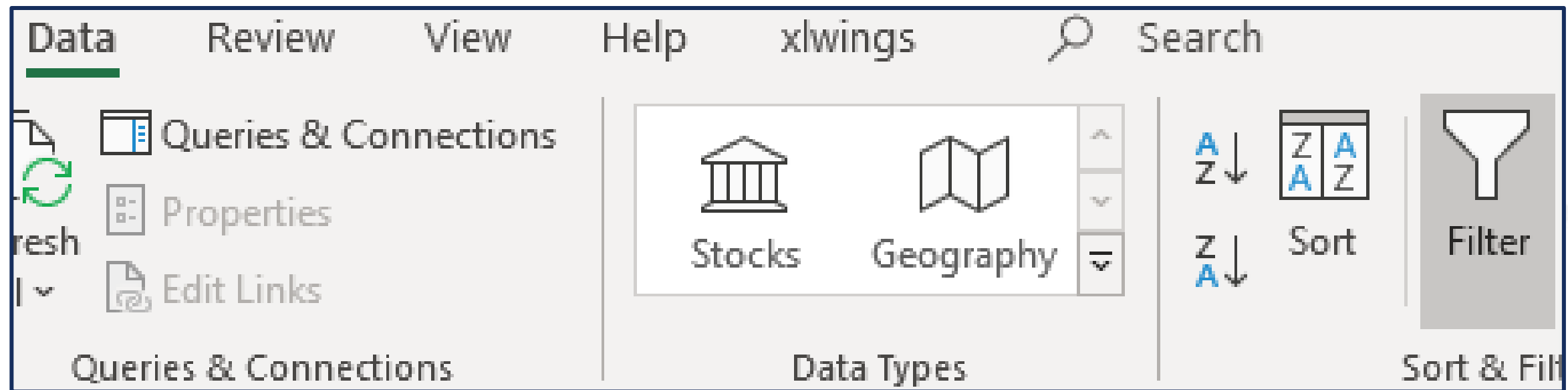
Filter option allows us to choose any column we would like to filter the data on.

Name	Subject	CGPA
Alison Cox	Maths	4.3
Anise Jeff	Maths	4.2
Beatrice Cane	Maths	3.4
Danish Xavier	Maths	3.9
Rada Hofman	Maths	4.1
James Alan	Maths	4.8
Xavier Alex	Maths	4.6
Albert Dane	Maths	4.8
Hassan Alhuri	Maths	4.6



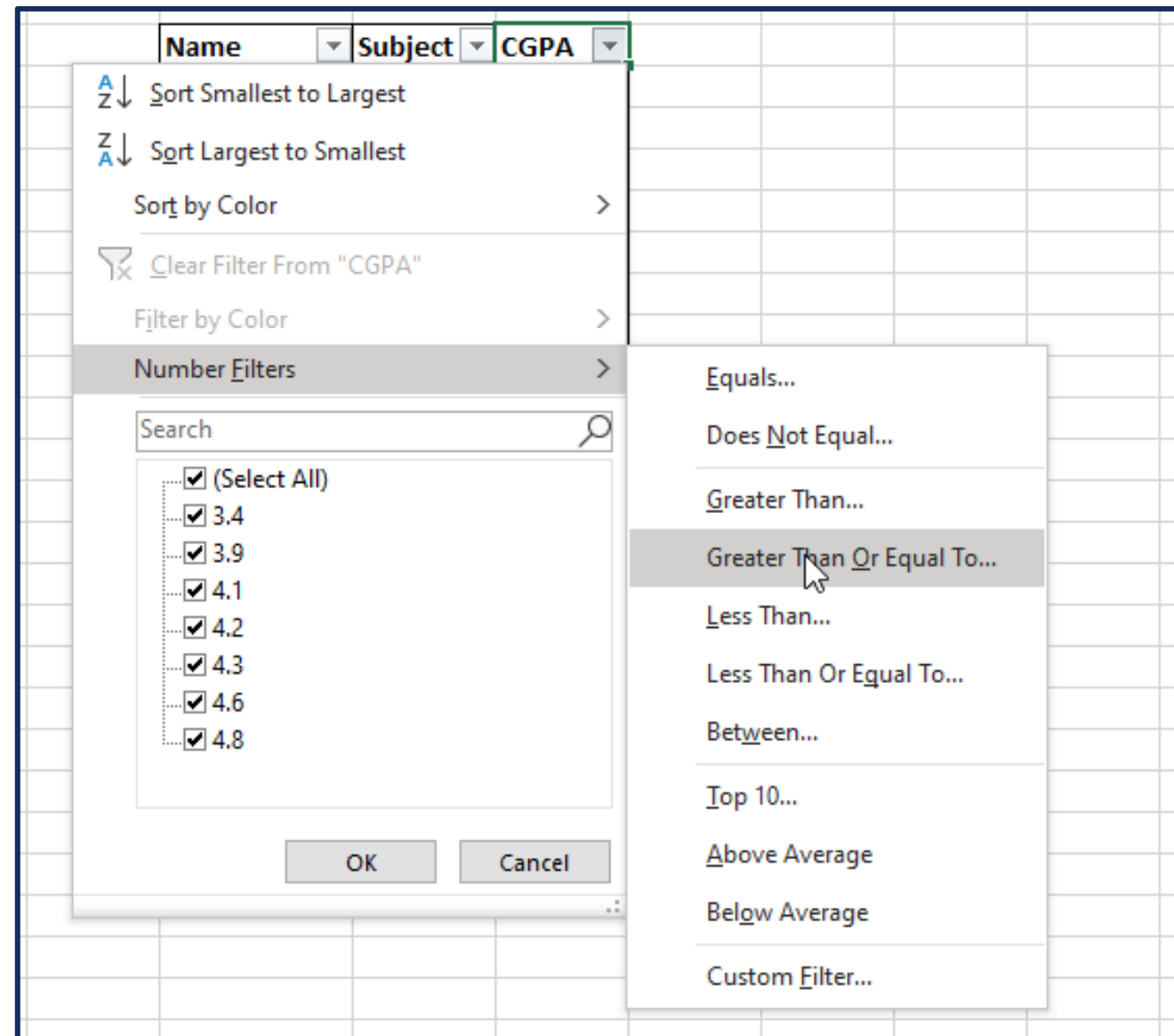
Steps to Filter Data

Choose a column to filter and in the Data tab, click on Filter



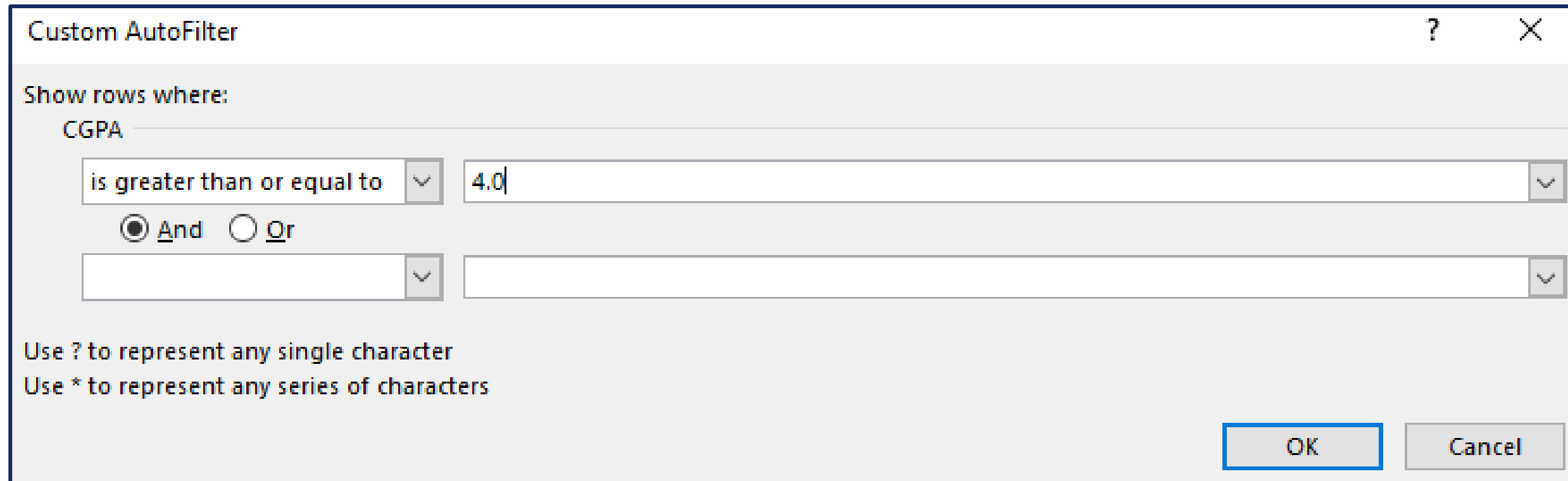
Steps to Filter Data

Example: To view CGPA's that are greater than or equal to four.



Steps to Filter Data

Choose the greater than or equal to option from the dropdown, and then mention the number



Custom AutoFilter

Show rows where:

CGPA

is greater than or equal to 4.0

☒ And ☐ Or

Use ? to represent any single character
Use * to represent any series of characters

OK Cancel

In this case, it is 4.0.

Steps to Filter Data

The result will be in the following order.

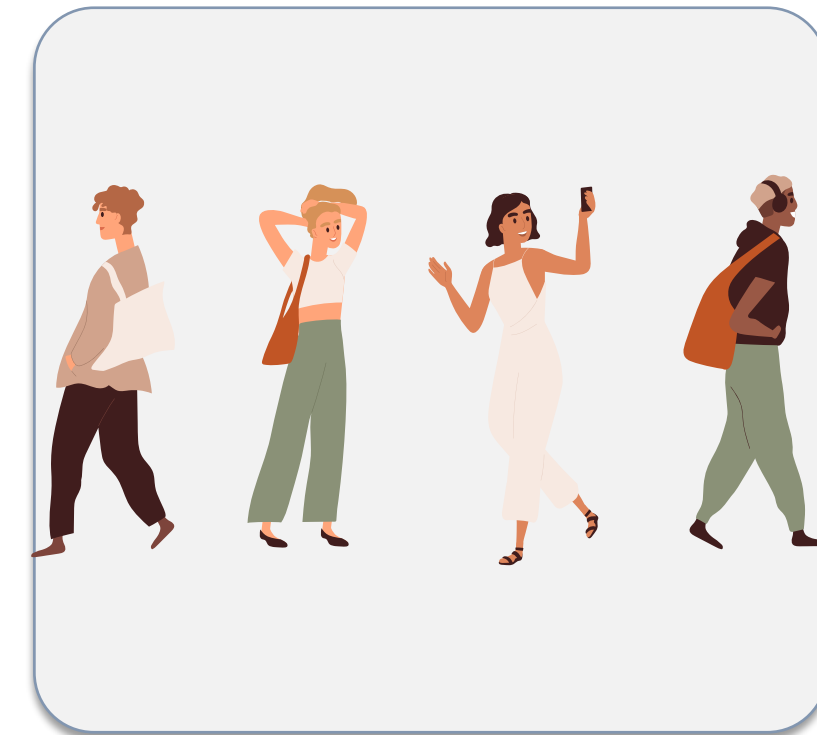
Name	Subject	CGPA
Albert Dane	Maths	4.8
Alison Cox	Maths	4.3
Anise Jeff	Maths	4.2
Hassan Alburi	Maths	4.6
James Alan	Maths	4.8
Rada Hofman	Maths	4.1
Xavier Alex	Maths	4.6



Group by and Subtotal

Group by and Ungroup

Group by and ungroup allow data to group data by collapse and expanding rows with similar content to create more compact and understandable views.



Group by and ungroup by are available under the Data tab within the outline section.

Group By

The group by functionality in Excel allows us to show necessary data for easy viewing and analysis.



It is possible to create subtotals and outline for a given set of data.

Group By

Group by can be done for rows or columns.

1	2		A	B	C	D
+	8		James Alan	4.8	3.9	3.9
	9		Rada Hofman	4.1	3.5	3.7
	10		Xavier Alex	4.6	4.6	3.9

Grouping for Rows

1		+
2		
	A	E
1	Date	
2	Albert Dane	
3	Alison Cox	
4	Anise Jeff	
5	Beatrice Cane	
6	Danish Xavier	
7	Hassan Alhuri	
8	James Alan	
9	Rada Hofman	
10	Xavier Alex	

Grouping for Columns

Steps for Grouping

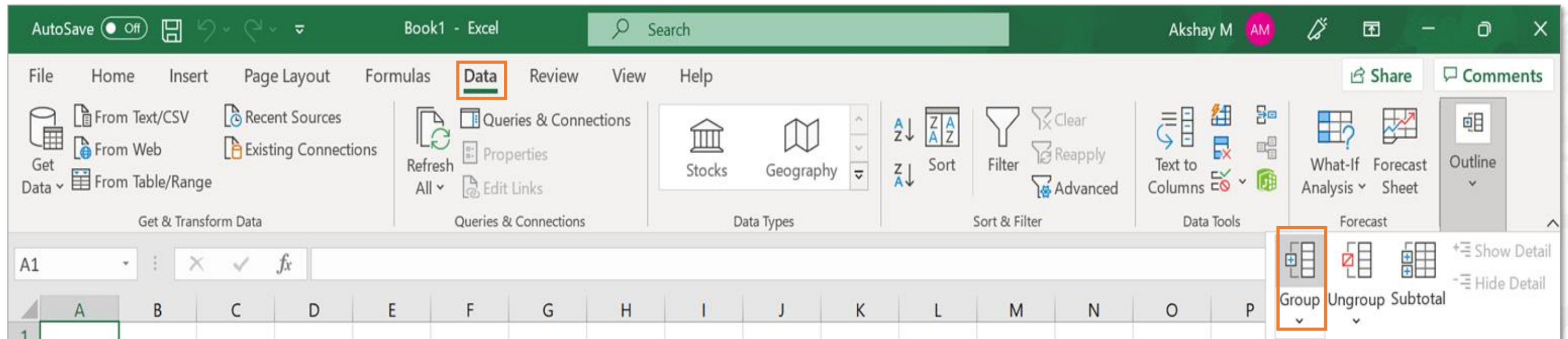
Let us discuss the steps for grouping data.

	A	B	C	D
1	Date	Maths	English	Science
2	Albert Dane	4.8	4.8	4.3
3	Alison Cox	4.3	3.4	4.2
4	Anise Jeff	4.2	3.5	4.1
5	Beatrice Cane	3.4	3.6	4.6
6	Danish Xavier	3.9	3.5	4.5
7	Hassan Alburi	4.6	3.6	3.8
8	James Alan	4.8	3.9	3.9
9	Rada Hofman	4.1	3.5	3.7
10	Xavier Alex	4.6	4.6	3.9

Step 1: To group data, select the rows and columns you want to group

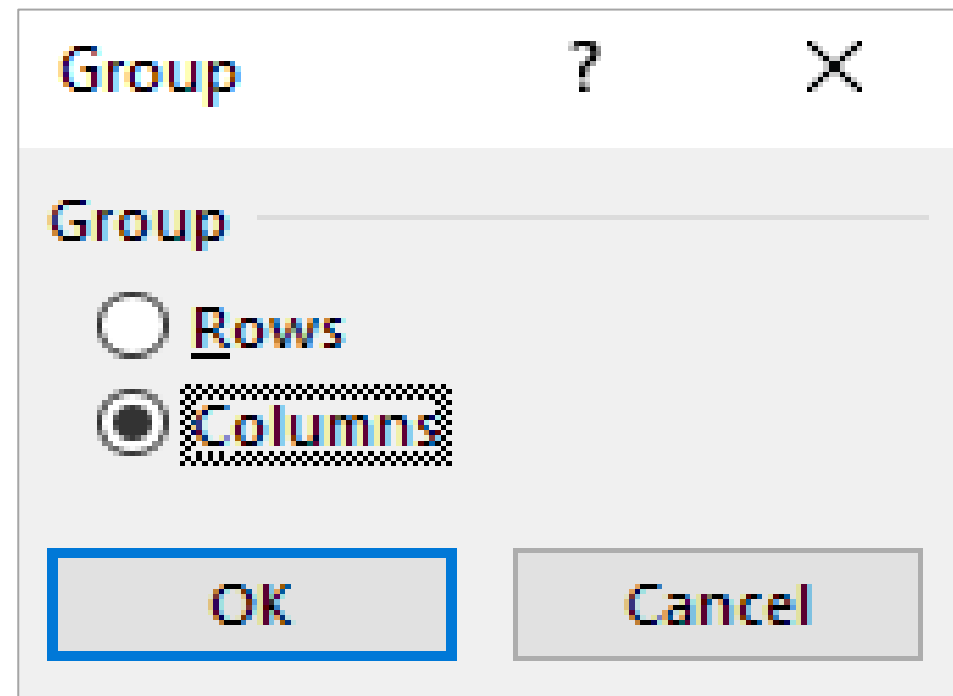
Steps for Grouping

Step 2: Click on Group under Data tab



Grouping for Columns

Step 3: Choose Columns and click on OK



Grouping for Columns

Step 4: This groups the three columns chosen, and applies a control to show or hide the grouped content

	A	B	C	D	E
1	Date	Maths	English	Science	
2	Albert Dane	4.8	4.8	4.3	
3	Alison Cox	4.3	3.4	4.2	
4	Anise Jeff	4.2	3.5	4.1	
5	Beatrice Cane	3.4	3.6	4.6	
6	Danish Xavier	3.9	3.5	4.5	
7	Hassan Alhuri	4.6	3.6	3.8	
8	James Alan	4.8	3.9	3.9	
9	Rada Hofman	4.1	3.5	3.7	
10	Xavier Alex	4.6	4.6	3.9	



Grouping for Columns

Clicking on – hides the content, while clicking on + shows the grouped content.

1					–
2					
	A	B	C	D	E
1	Date	Maths	English	Science	
2	Albert Dane	4.8	4.8	4.3	
3	Alison Cox	4.3	3.4	4.2	
4	Anise Jeff	4.2	3.5	4.1	
5	Beatrice Cane	3.4	3.6	4.6	
6	Danish Xavier	3.9	3.5	4.5	
7	Hassan Alhuri	4.6	3.6	3.8	
8	James Alan	4.8	3.9	3.9	
9	Rada Hofman	4.1	3.5	3.7	
10	Xavier Alex	4.6	4.6	3.9	

1			+
2			
	A	E	
1	Date		
2	Albert Dane		
3	Alison Cox		
4	Anise Jeff		
5	Beatrice Cane		
6	Danish Xavier		
7	Hassan Alhuri		
8	James Alan		
9	Rada Hofman		
10	Xavier Alex		

Grouping for Rows

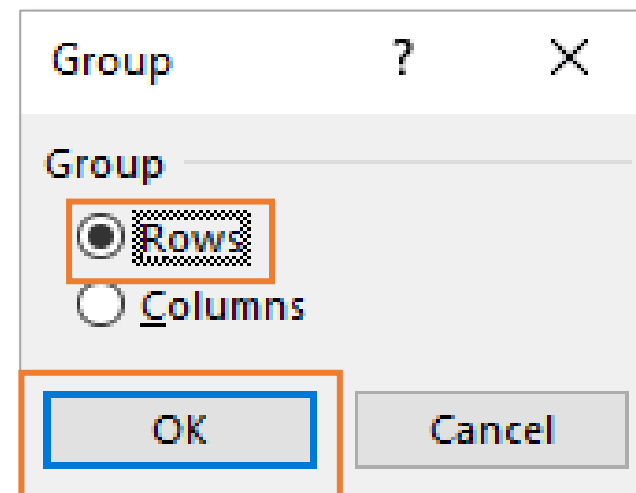
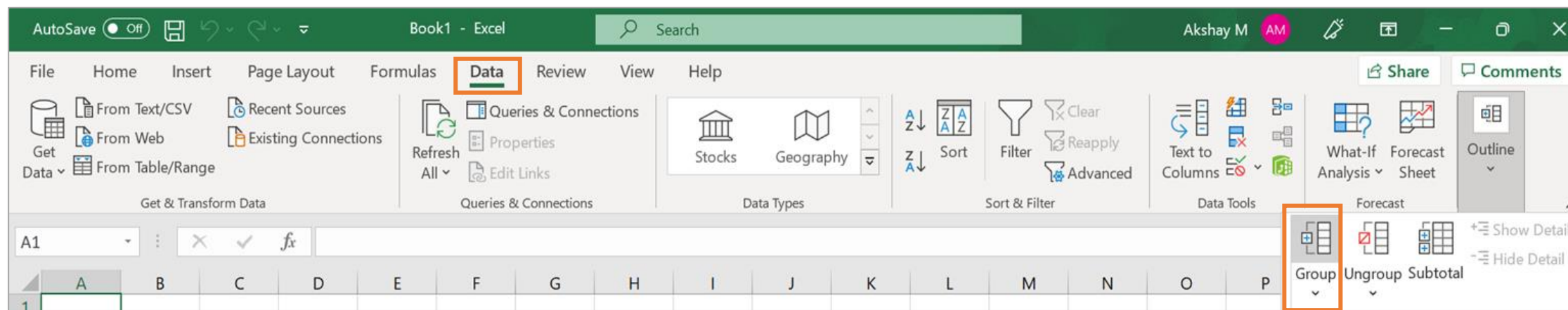
Similarly, for row-wise grouping, select the rows you want to group.

	A	B	C	D
1	Date	Maths	English	Science
2	Albert Dane	4.8	4.8	4.3
3	Alison Cox	4.3	3.4	4.2
4	Anise Jeff	4.2	3.5	4.1
5	Beatrice Cane	3.4	3.6	4.6
6	Danish Xavier	3.9	3.5	4.5
7	Hassan Alhuri	4.6	3.6	3.8
8	James Alan	4.8	3.9	3.9
9	Rada Hofman	4.1	3.5	3.7
10	Xavier Alex	4.6	4.6	3.9



Grouping for Rows

Step 1: Click on Group under the Data tab, and then select rows option from the dialog box



Step 2: Click on OK

Grouping for Rows

Clicking on – hides the content and clicking on + shows the grouped content.

1	2		A	B	C	D
	1		Date	Maths	English	Science
	2		Albert Dane	4.8	4.8	4.3
	3		Alison Cox	4.3	3.4	4.2
	4		Anise Jeff	4.2	3.5	4.1
	5		Beatrice Cane	3.4	3.6	4.6
	6		Danish Xavier	3.9	3.5	4.5
	7		Hassan Alhuri	4.6	3.6	3.8
	8		James Alan	4.8	3.9	3.9
	9		Rada Hofman	4.1	3.5	3.7
	10		Xavier Alex	4.6	4.6	3.9

1	2		A	B	C	D
	8		James Alan	4.8	3.9	3.9
	9		Rada Hofman	4.1	3.5	3.7
	10		Xavier Alex	4.6	4.6	3.9

Grouping for Rows

We can create a group within a group by choosing rows or columns within the grouped data.

Create a row or column group again.



Grouping for Rows

This will be the result.

1	2	3		A	B	C	D
	.		1	Date	Maths	English	Science
	.		2	Albert Dane	4.8	4.8	4.3
	.		3	Alison Cox	4.3	3.4	4.2
	.		4	Anise Jeff	4.2	3.5	4.1
	.		5	Beatrice Cane	3.4	3.6	4.6
	.		6	Danish Xavier	3.9	3.5	4.5
	.		7	Hassan Alhuri	4.6	3.6	3.8
	.		8	James Alan	4.8	3.9	3.9
		-	9	Rada Hofman	4.1	3.5	3.7
		-	10	Xavier Alex	4.6	4.6	3.9



Ungrouping

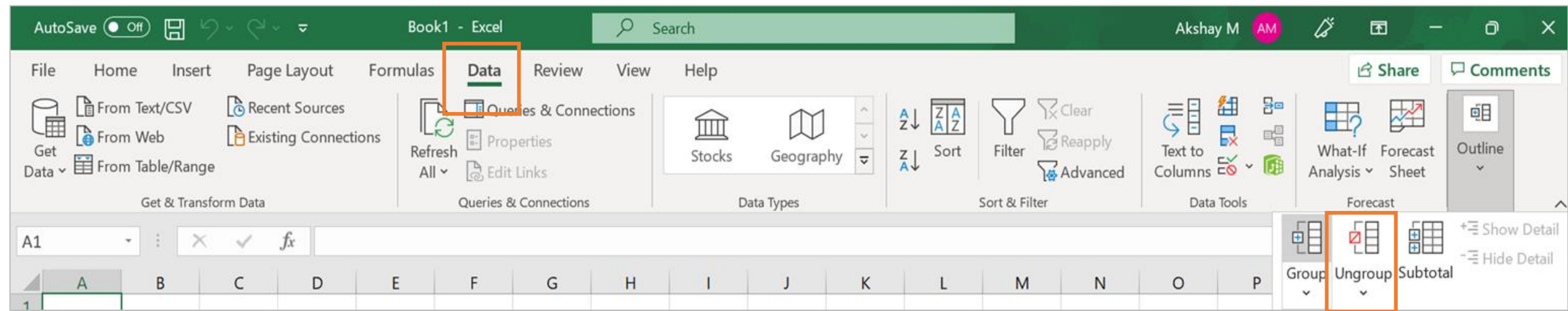
The ungroup option allows us to remove the groups created by group.

Step 1:
Choose the data already chosen for grouping
(row/column)



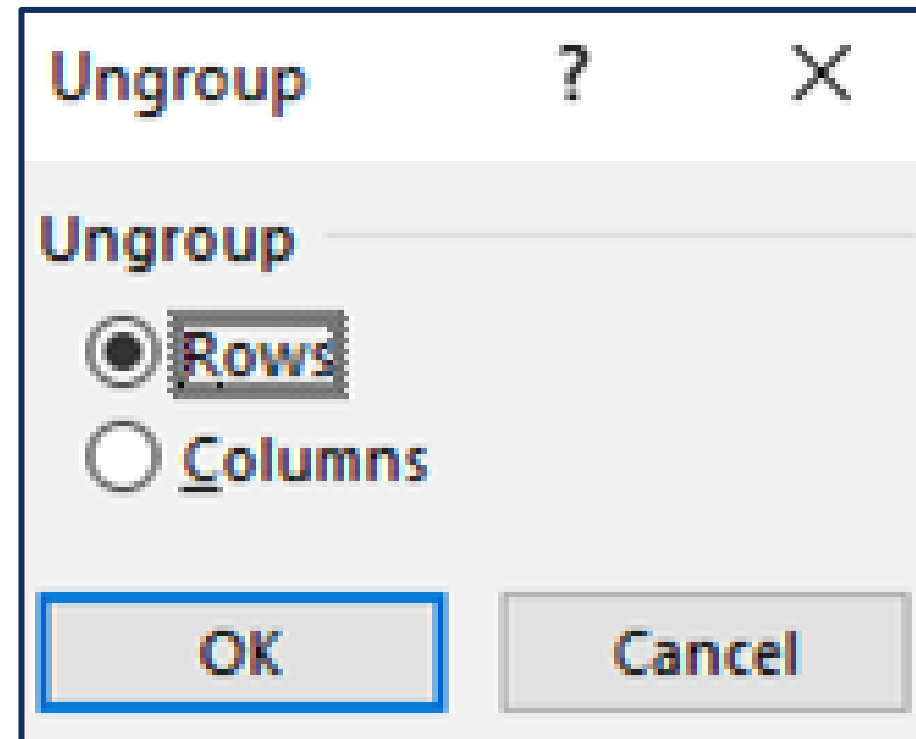
Ungrouping

Step 2: Click on Ungroup under the Data tab



Ungrouping

Step 3: Choose Rows to remove row-level grouping



Ungrouping

Step 4: The group chosen will be removed

1	2		A	B	C	D
	•	1	Date	Maths	English	Science
	•	2	Albert Dane	4.8	4.8	4.3
	•	3	Alison Cox	4.3	3.4	4.2
	•	4	Anise Jeff	4.2	3.5	4.1
	•	5	Beatrice Cane	3.4	3.6	4.6
	•	6	Danish Xavier	3.9	3.5	4.5
	•	7	Hassan Alhuri	4.6	3.6	3.8
	•	8	James Alan	4.8	3.9	3.9
-		9	Rada Hofman	4.1	3.5	3.7
		10	Xavier Alex	4.6	4.6	3.9



Subtotal

Subtotal allows us to create groups and have a subtotal for each group.



Subtotal: Example

Let us understand this by taking an example.

Name	Subject	Marks
Albert Dane	Maths	4.8
Albert Dane	English	4.8
Albert Dane	Science	4.3
Alison Cox	Maths	4.3
Alison Cox	English	3.4
Alison Cox	Science	4.2
Anise Jeff	Maths	4.2
Anise Jeff	English	3.5
Anise Jeff	Science	4.1
Beatrice Cane	Maths	3.4
Beatrice Cane	English	3.6
Beatrice Cane	Science	4.6
Danish Xavier	Maths	3.9
Danish Xavier	English	3.5
Danish Xavier	Science	4.5
Hassan Alhuri	Maths	4.6
Hassan Alhuri	English	3.6
Hassan Alhuri	Science	3.8
James Alan	Maths	4.8
James Alan	English	3.9
James Alan	Science	3.9
Rada Hofman	Maths	4.1
Rada Hofman	English	3.5
Rada Hofman	Science	3.7
Xavier Alex	Maths	4.6
Xavier Alex	English	4.6
Xavier Alex	Science	3.9

For the following data set,
find the total per student by
grouping students and adding their
marks.

Subtotal: Example

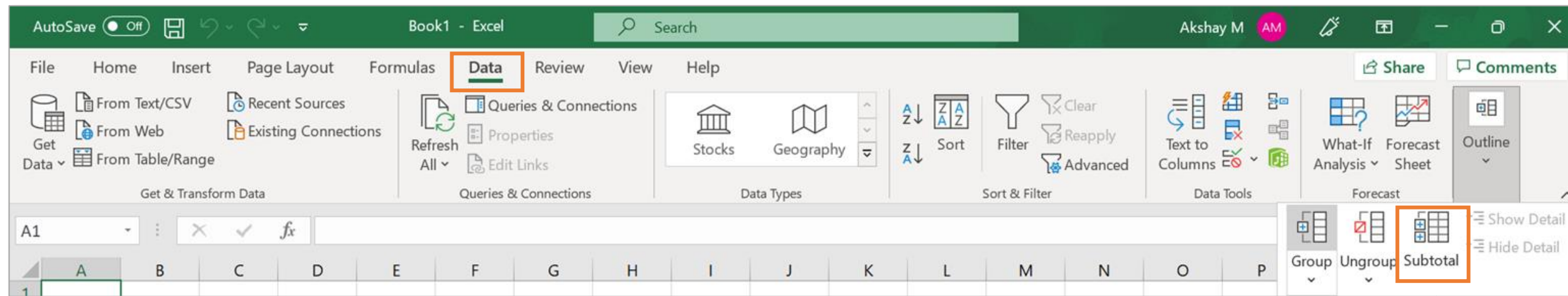
Step 1: Select the data we need to group by and subtotal

Name	Subject	Marks
Albert Dane	Maths	4.8
Albert Dane	English	4.8
Albert Dane	Science	4.3
Alison Cox	Maths	4.3
Alison Cox	English	3.4
Alison Cox	Science	4.2
Anise Jeff	Maths	4.2
Anise Jeff	English	3.5
Anise Jeff	Science	4.1
Beatrice Cane	Maths	3.4
Beatrice Cane	English	3.6
Beatrice Cane	Science	4.6
Danish Xavier	Maths	3.9
Danish Xavier	English	3.5
Danish Xavier	Science	4.5
Hassan Alhuri	Maths	4.6
Hassan Alhuri	English	3.6
Hassan Alhuri	Science	3.8
James Alan	Maths	4.8
James Alan	English	3.9
James Alan	Science	3.9
Rada Hofman	Maths	4.1
Rada Hofman	English	3.5
Rada Hofman	Science	3.7
Xavier Alex	Maths	4.6
Xavier Alex	English	4.6
Xavier Alex	Science	3.9



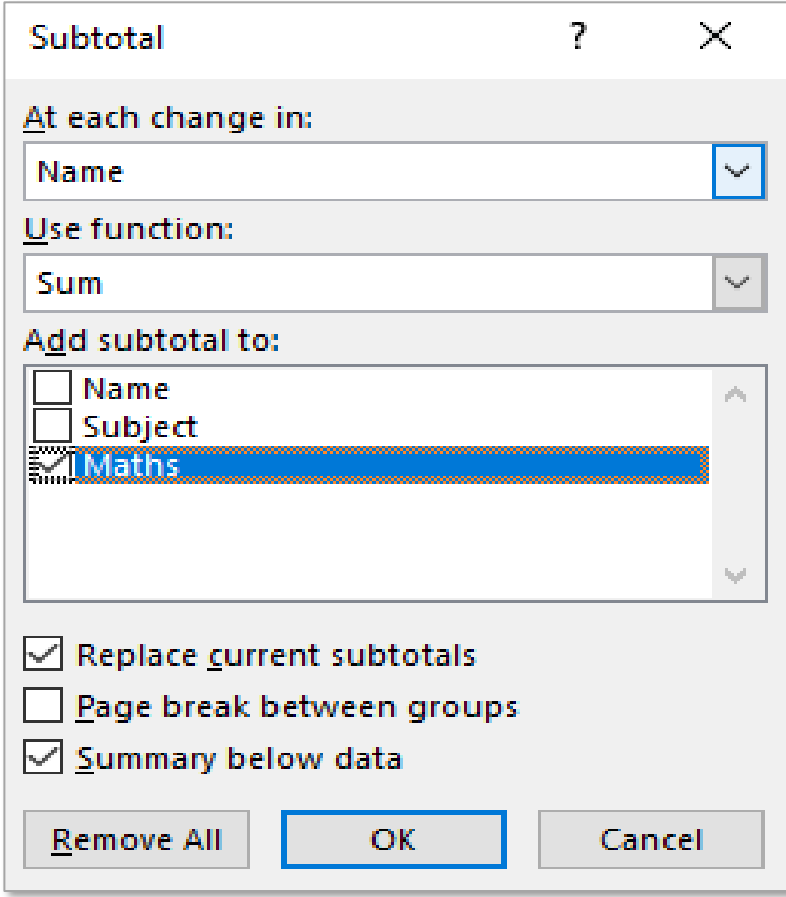
Subtotal: Example

Step 2: Click on Subtotal under Data tab



Subtotal: Example

Step 3: Click on the column to which the sum function has to be applied



Subtotal

At each change in:
Name

Use function:
Sum

Add subtotal to:
☐ Name
☐ Subject
☒ Maths

☒ Replace current subtotals
☐ Page break between groups
☒ Summary below data

Remove All OK Cancel



Subtotal: Example

The subtotalling provides control to the group and shows subtotals per student.

	A	B	C
1	Name	Subject	Marks
2	Albert Dane	Maths	4.8
3	Albert Dane	English	4.8
4	Albert Dane	Science	4.3
5	Albert Dane Total		13.9
6	Alison Cox	Maths	4.3
7	Alison Cox	English	3.4
8	Alison Cox	Science	4.2
9	Alison Cox Total		11.9
10	Anise Jeff	Maths	4.2
11	Anise Jeff	English	3.5
12	Anise Jeff	Science	4.1
13	Anise Jeff Total		11.8
14	Beatrice Cane	Maths	3.4
15	Beatrice Cane	English	3.6
16	Beatrice Cane	Science	4.6
17	Beatrice Cane Total		11.6
18	Danish Xavier	Maths	3.9
19	Danish Xavier	English	3.5
20	Danish Xavier	Science	4.5
21	Danish Xavier Total		11.9
22	Hassan Alhuri	Maths	4.6
23	Hassan Alhuri	English	3.6
24	Hassan Alhuri	Science	3.8
25	Hassan Alhuri Total		12
26	James Alan	Maths	4.8
27	James Alan	English	3.9
28	James Alan	Science	3.9
29	James Alan Total		12.6
30	Rada Hofman	Maths	4.1
31	Rada Hofman	English	3.5
32	Rada Hofman	Science	3.7
33	Rada Hofman Total		11.3
34	Xavier Alex	Maths	4.6
35	Xavier Alex	English	4.6
36	Xavier Alex	Science	3.9
37	Xavier Alex Total		13.1
38	Grand Total		110.1




Text to Column

Text to Column

It converts raw text into columns in excel, which can save a user the time of manually separating the text in a cell into several columns.

Name, age, address, phone number, university
Tom Smith, 22,4th street, 8998798901, St Gallen University

Raw text



	A	B	C	D	E
1	Name	age	address	phone number	university
2	Tom Smith	22	4th street	8998798901	St Gallen University

Text put in excel columns

Text to Column: Example

Step 1: Open Excel and paste the content into a sheet

	A	B	C	D	E	F
1	Name, age, address, phone number, university					
2	Tom Smith, 22,4 th street, 8998798901, St Gallen University					



Text to Column: Example

Step 2: Choose Column A

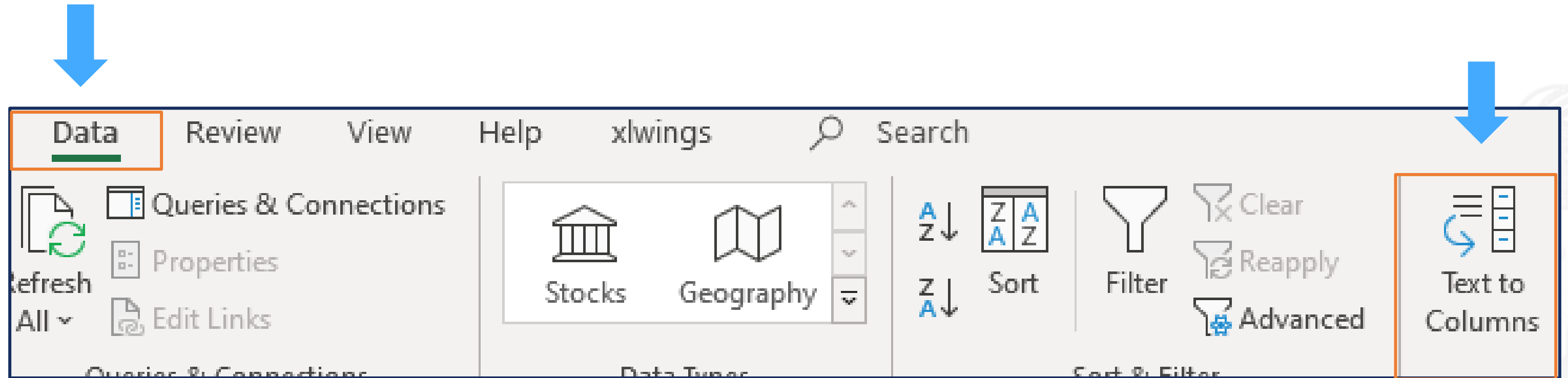


	A	B	C	D	E	F
1	Name, age, address, phone number, university					
2	Tom Smith, 22,4 th street, 8998798901, St Gallen University					



Text to Column: Example

Step 3: Go to the **Data** tab and click on **Text to Columns**



Text to Column: Example

Convert Text to Columns Wizard - Step 1 of 3

The Text Wizard has determined that your data is Fixed Width.
If this is correct, choose Next, or choose the data type that best describes your data.

Original data type

Choose the file type that best describes your data:

☒ Delimited - Characters such as commas or tabs separate each field.

☐ Fixed width - Fields are aligned in columns with spaces between each field.

Preview of selected data:

1	Name, age, address, phone number, university
2	Tom Smith, 22,4th street, 8998798901, St Gallen University
3	
4	
5	

Cancel < Back **Next >** Finish

Step 4: Select the **Delimited** option in the dialog box and click **Next**

Text to Column: Example

Convert Text to Columns Wizard - Step 2 of 3

This screen lets you set the delimiters your data contains. You can see how your text is affected in the preview below.

Delimiters

☐ Tab

☐ Semicolon

☒ Comma

☐ Space

☐ Other:

☐ Treat consecutive delimiters as one

Text qualifier:

Data preview

Name	age	address	phone number	university
Tom Smith	22	4th street	8998798901	St Gallen University

Cancel

< Back

Next >

Finish

Step 5: Choose **Comma** and click **Next**, since the delimiter is a comma here.

©Simplilearn. All rights reserved.

simplilearn

Text to Column: Example

	A	B	C	D	E
1	Name	age	address	phone number	university
2	Tom Smith	22	4th street	8998798901	St Gallen University

The text to column function puts each element separated by comma in an individual box

Removing Duplicates

Duplicate

Duplicate refers to a copy of the original.



Removing Duplicates in Excel

In any data analytics work, there will always be cases where we get duplicates in different columns.



Excel is very handy in removing duplicates in the data.

Causes of Duplicates

Duplicates can occur in data and cause errors in analytics.



Duplicates occur when there is an incorrect submission of user data.

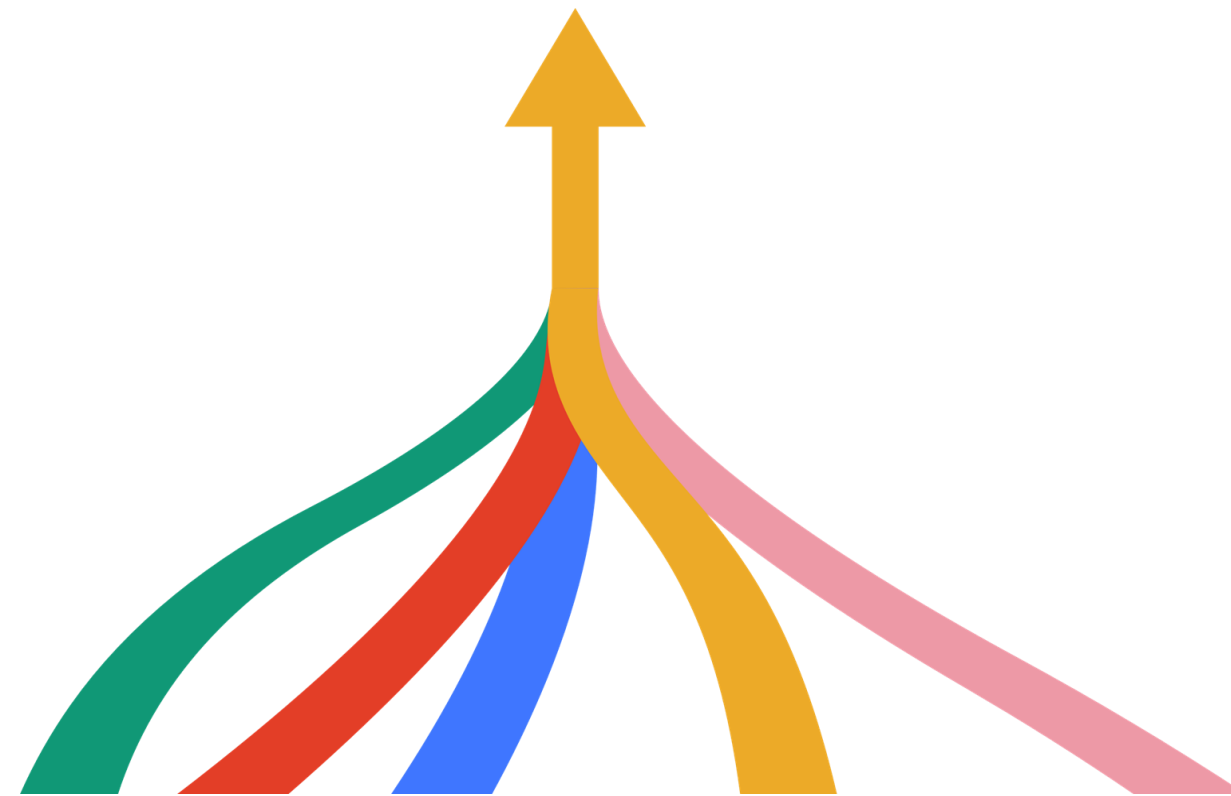
Causes of Duplicates

When there is a missing validation in the data set.



Causes of Duplicates

Duplicates occur when we merge multiple data sources using Joins.



Causes of Duplicates

When data is copy pasted multiple times.



When duplicates are removed using Excel, we can choose a single column or multiple columns to check the data.

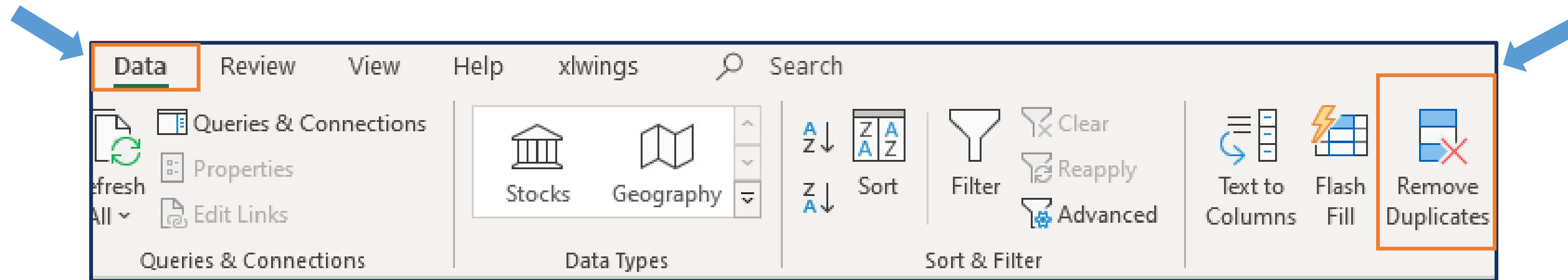
Removing Duplicates Using Single Column: Example

Step 1: Choose the column with a set of rows to remove duplicates

	A
1	Count
2	1
3	2
4	3
5	4
6	32
7	2
8	2
9	2
10	2
11	1
12	2
13	2

There are many duplicates in this column.

Removing Duplicates Using Single Column: Example



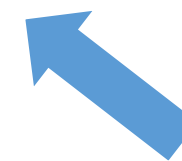
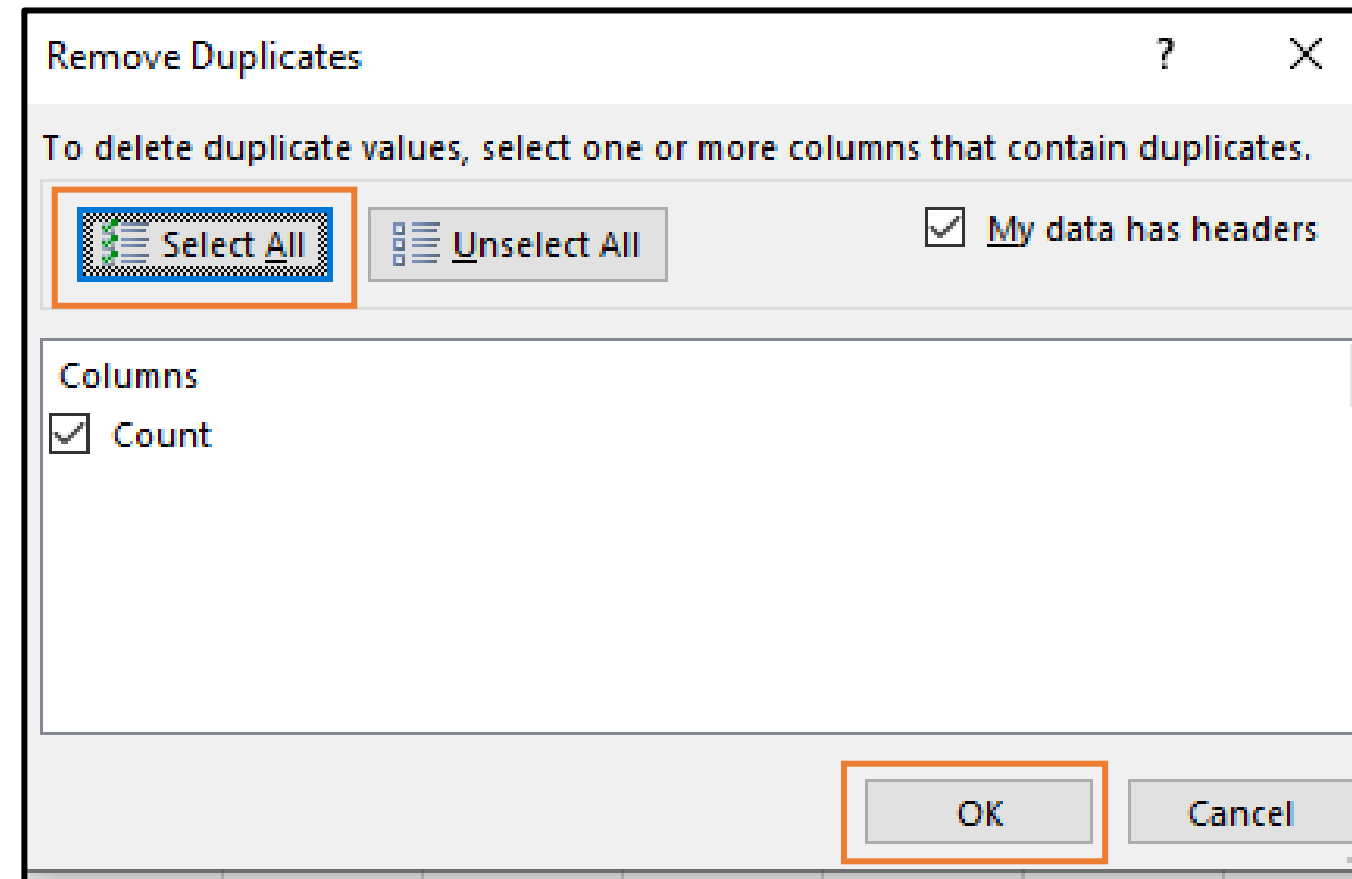
	A
1	Count
2	1
3	2
4	3
5	4
6	32
7	2
8	2
9	2
10	2
11	1
12	2
13	2

Step 2:

- Select the entire column
- Click on Data
- Click on Remove Duplicates


Removing Duplicates Using Single Column: Example

Step 3: After clicking the option a pop up will appear



Click OK

Removing Duplicates Using Single Column: Example



	A
1	Count
2	1
3	2
4	3
5	4
6	32

Step 4: It is clearly visible that all the duplicates are removed

Removing Duplicates Using Multiple Columns: Example

Let us consider the following data set as an example for removing duplicates with multiple columns:

Name	Subject	CGPA
Albert Dane	Maths	4.8
Albert Dane	Maths	3.4
Alison Cox	Maths	4.3
Anise Jeff	Maths	4.2

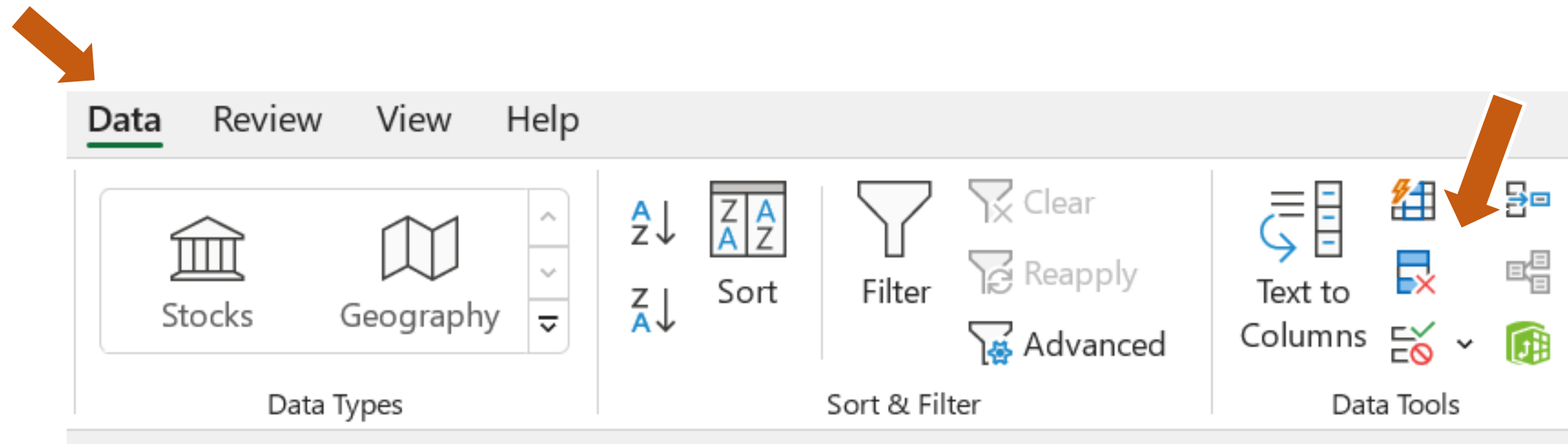
- Here, there are two entries for Maths subject under the same name, Albert Dane.
- When removing duplicates for this, only the first row is retained.

Removing Duplicates Using Multiple Columns: Example

Name	Subject	CGPA
Albert Dane	Maths	4.8
Albert Dane	Maths	3.4
Alison Cox	Maths	4.3
Anise Jeff	Maths	4.2

Step 1: Let us choose the data to remove duplicates

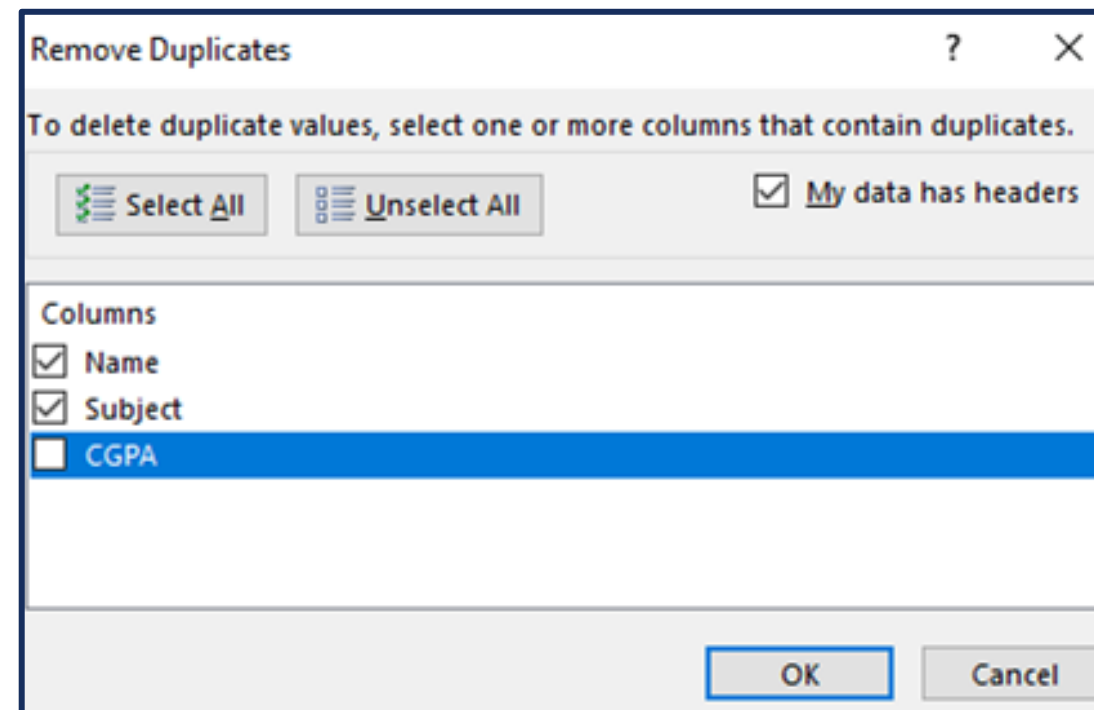
Removing Duplicates Using Multiple Columns: Example



Step 2: Now click on Remove Duplicates from Data tab

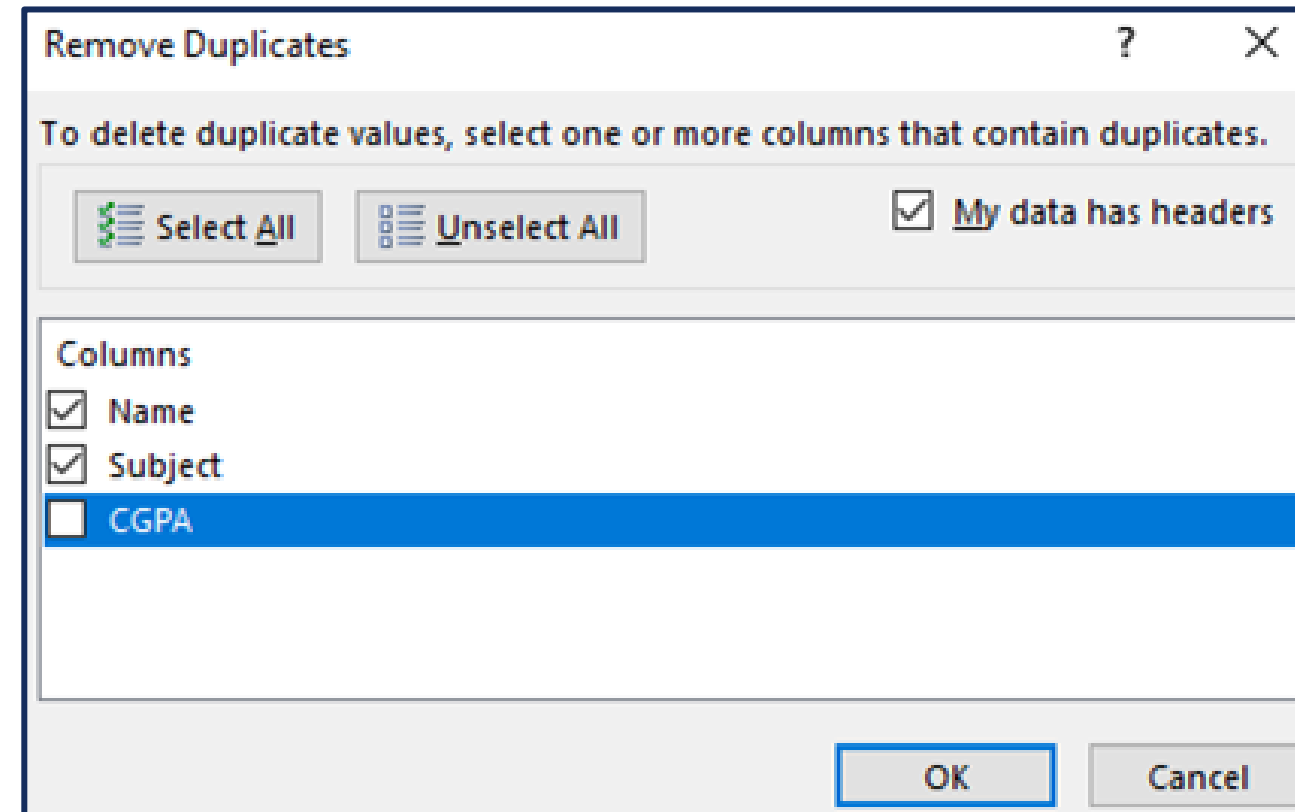
Removing Duplicates Using Multiple Columns: Example

A pop up will occur to remove duplicates.



Step 3: Choose the columns where duplicates need to be checked

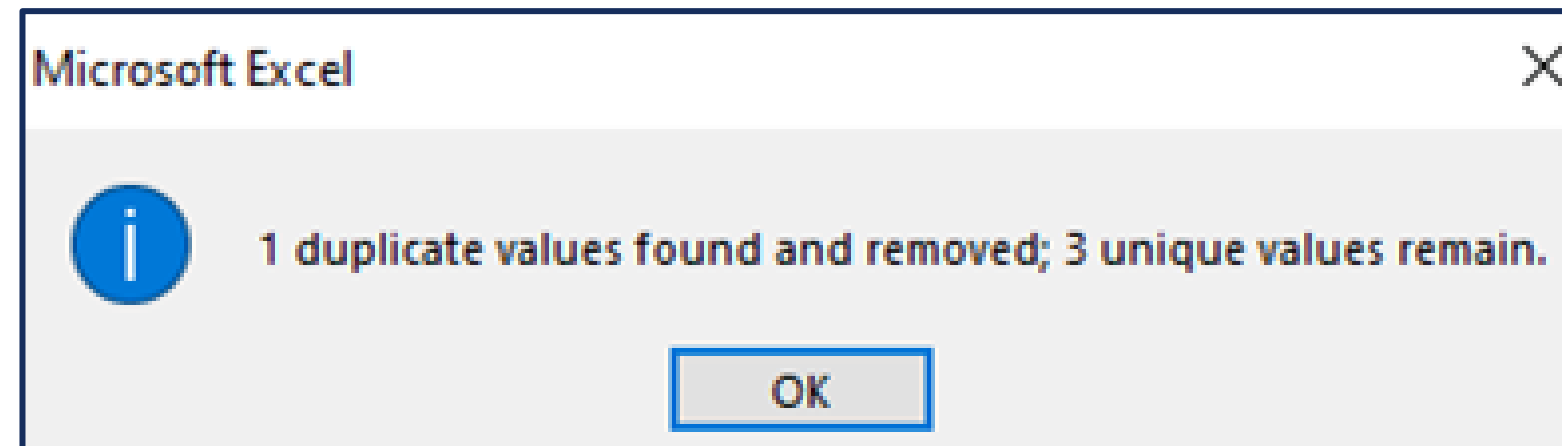
Removing Duplicates Using Multiple Columns: Example



Step 4: Once it is checked, click OK

Removing Duplicates Using Multiple Columns: Example

Another pop up will appear which notifies that, 1 duplicate value was found and removed, also 3 unique values remain.



Removing Duplicates Using Multiple Columns: Example

Name	Subject	CGPA
Albert Dane	Maths	4.8
Alison Cox	Maths	4.3
Anise Jeff	Maths	4.2

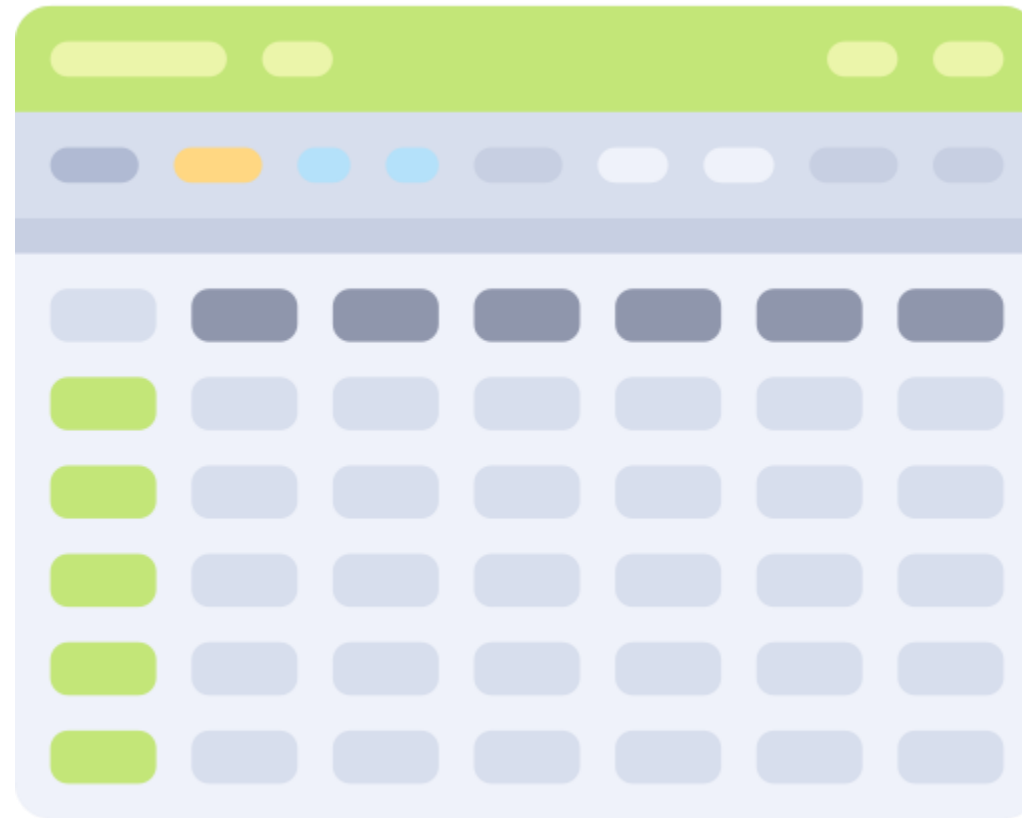


This is the final data set

Data Validation

Data Validation

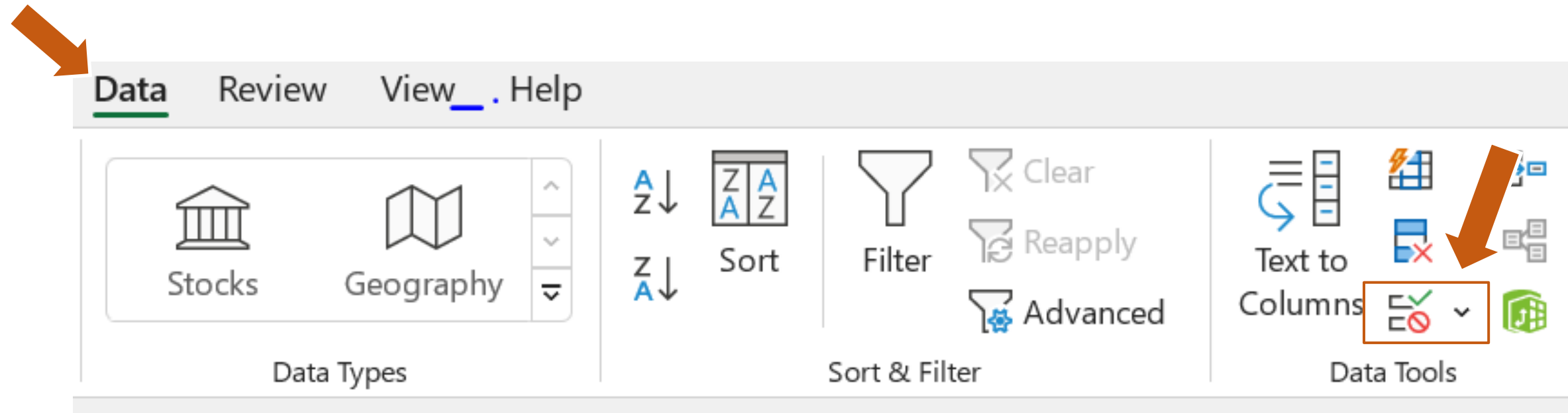
Data in Excel can be validated using some rules set in data validation dialog.



This helps in reducing the amount of unstandardized data, errors, or irrelevant information in the worksheet.

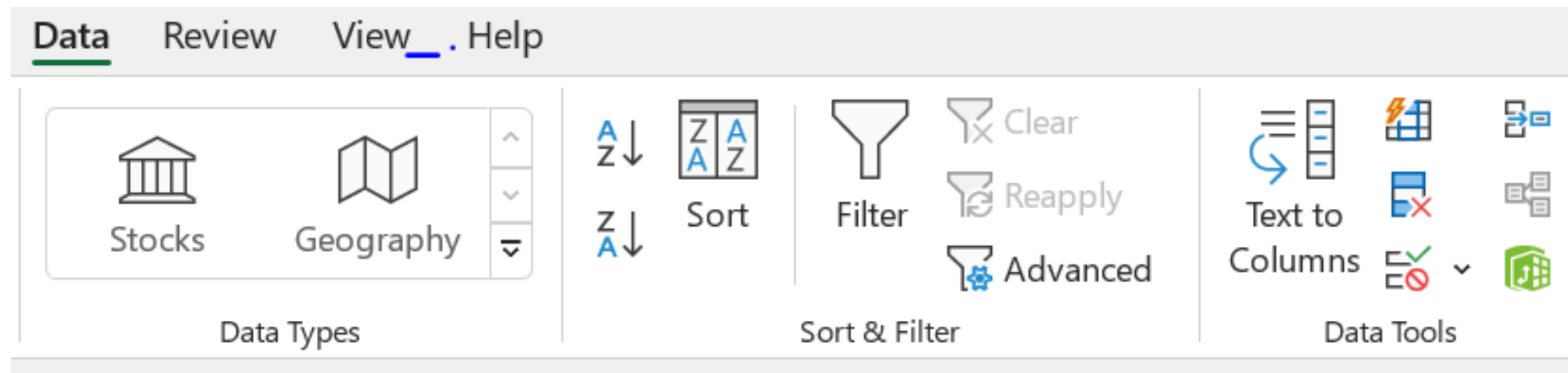
Data Validation: Example

Let us understand data validation through an example.



- Choose a cell or a group of cells to validate
- Click on Data Validation under Data tab

Data Validation: Example

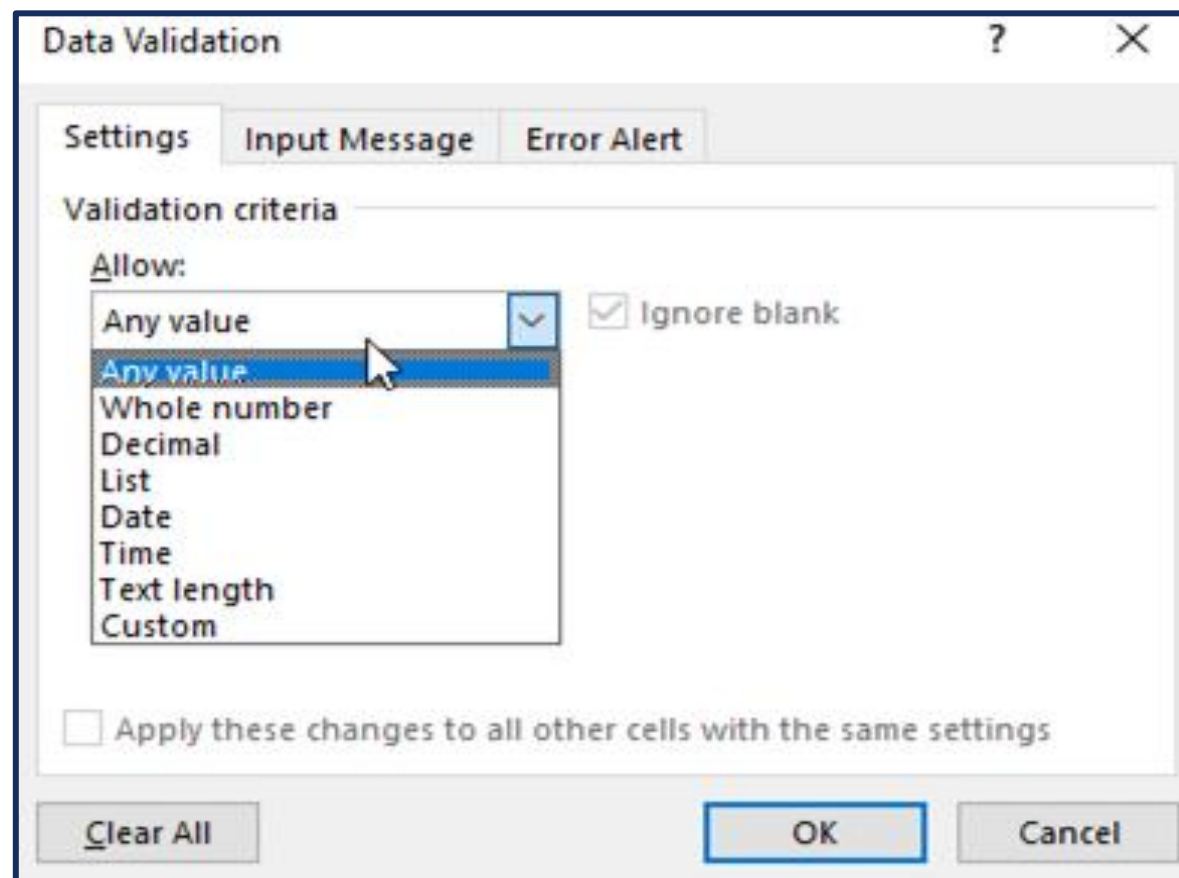


It is important to remember that:

- Validation applies to new data entered in the cells where rules are placed.
- Existing data is not validated.

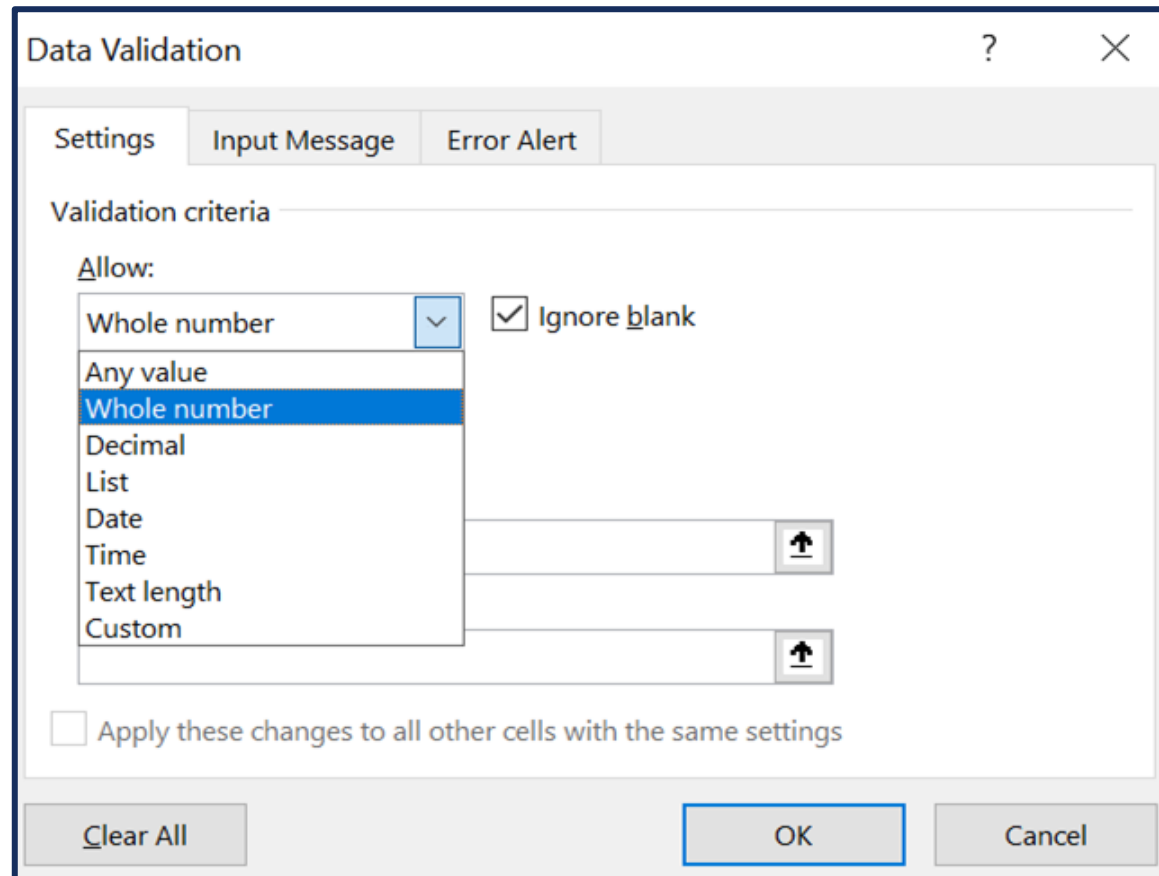
Data Validation: Example

After clicking on the data validation, a pop-up appears regarding the validation criteria and the following validations are possible.



'Any value' allows any alphanumeric value in the cells.

Data Validation: Example



Data Validation

Settings Input Message Error Alert

Validation criteria

Allow:

Whole number (selected) ☒ Ignore blank

Any value

Decimal

List

Date

Time

Text length

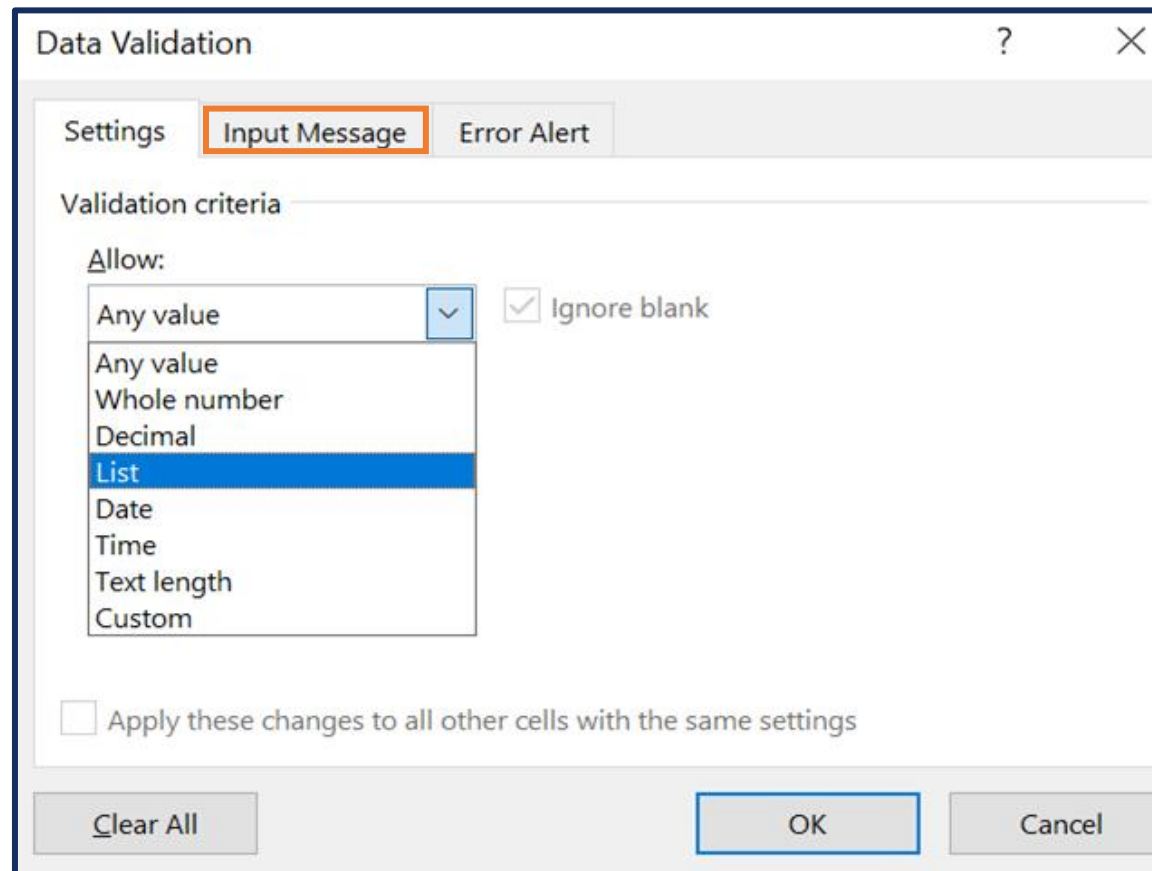
Custom

☐ Apply these changes to all other cells with the same settings

Clear All OK Cancel

'Whole number' allows whole numbers and a set of rules including a range of minimum and maximum to be set.

Data Validation: Example



Data Validation

Settings **Input Message** Error Alert

Validation criteria

Allow:

Any value ☒ Ignore blank

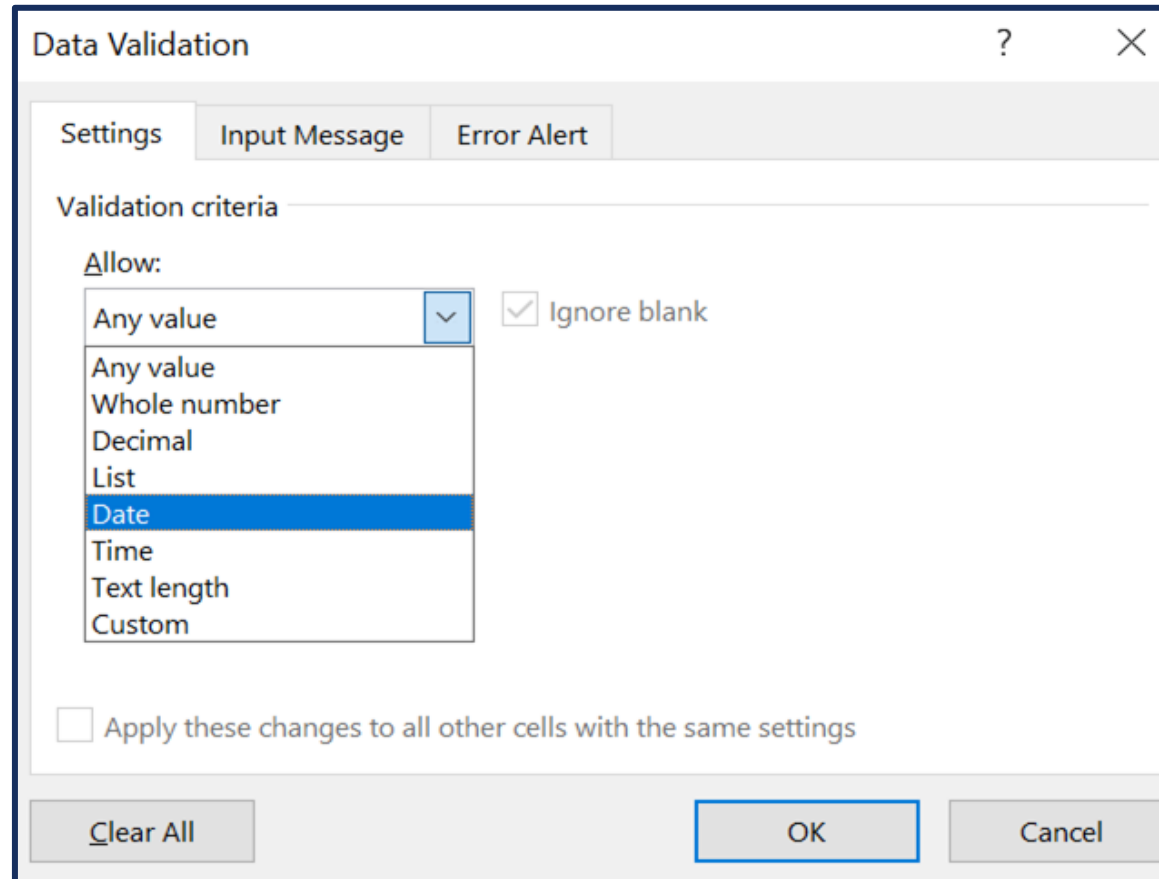
Any value
Whole number
Decimal
List
Date
Time
Text length
Custom

☐ Apply these changes to all other cells with the same settings

Clear All OK Cancel

'List' allows only a list of values specified in a range of cells or written manually in the 'source' input box.

Data Validation: Example



Data Validation

Settings Input Message Error Alert

Validation criteria

Allow:

Any value (selected)
Any value
Whole number
Decimal
List
Date
Time
Text length
Custom

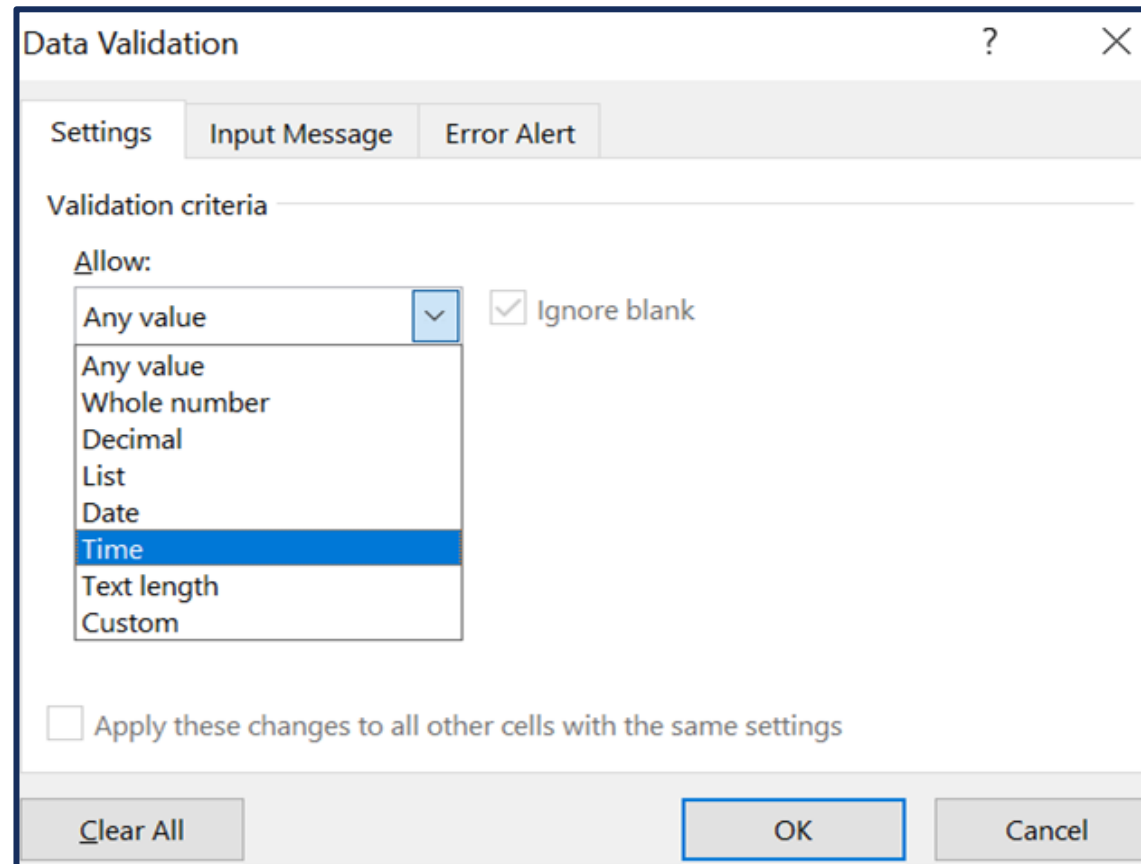
☒ Ignore blank

☐ Apply these changes to all other cells with the same settings

Clear All OK Cancel

'Date' allows only dates and a set of rules including a range of minimum and maximum to be set.

Data Validation: Example



Data Validation

Settings Input Message Error Alert

Validation criteria

Allow:

Any value ☒ Ignore blank

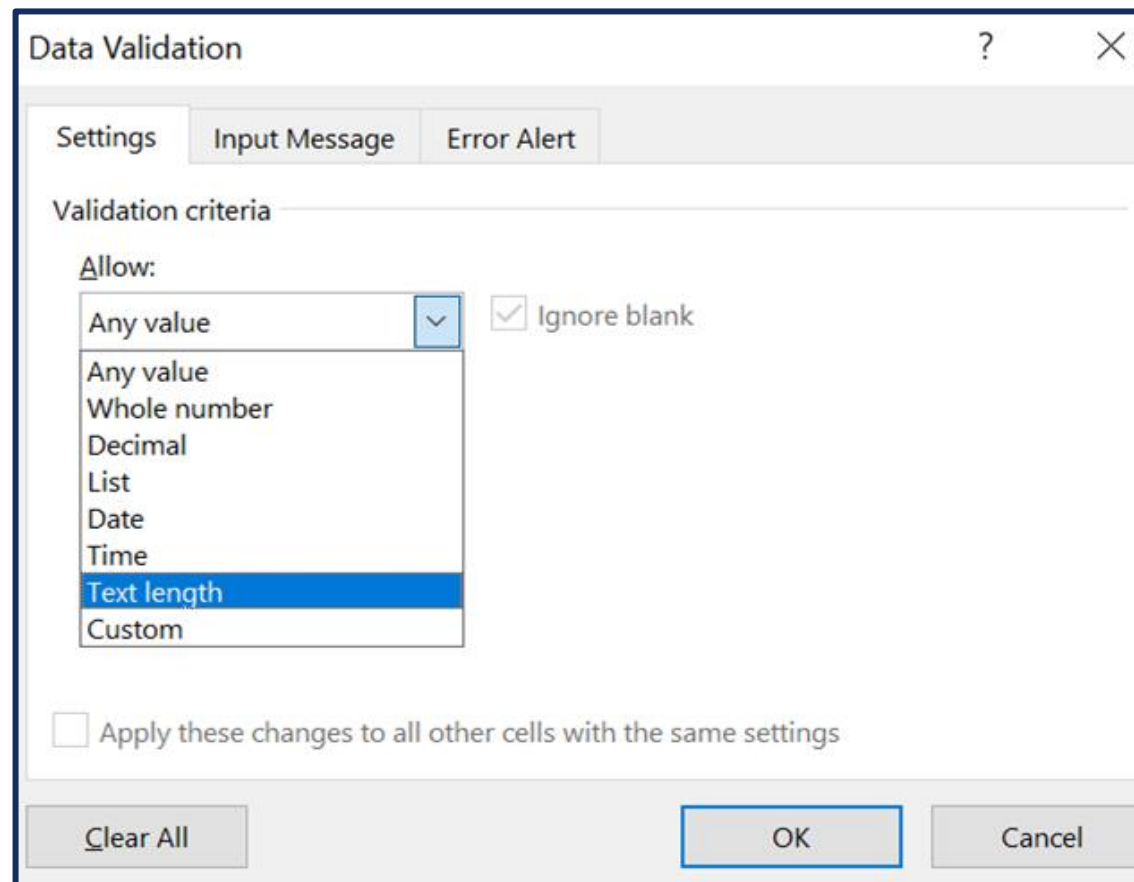
Any value
Whole number
Decimal
List
Date
Time
Text length
Custom

☐ Apply these changes to all other cells with the same settings

Clear All OK Cancel

'Time' allows only time values and a set of rules including a range of minimum and maximum to be set.

Data Validation: Example



Data Validation

Settings Input Message Error Alert

Validation criteria

Allow:

- Any value
- Any value
- Whole number
- Decimal
- List
- Date
- Time
- Text length**
- Custom

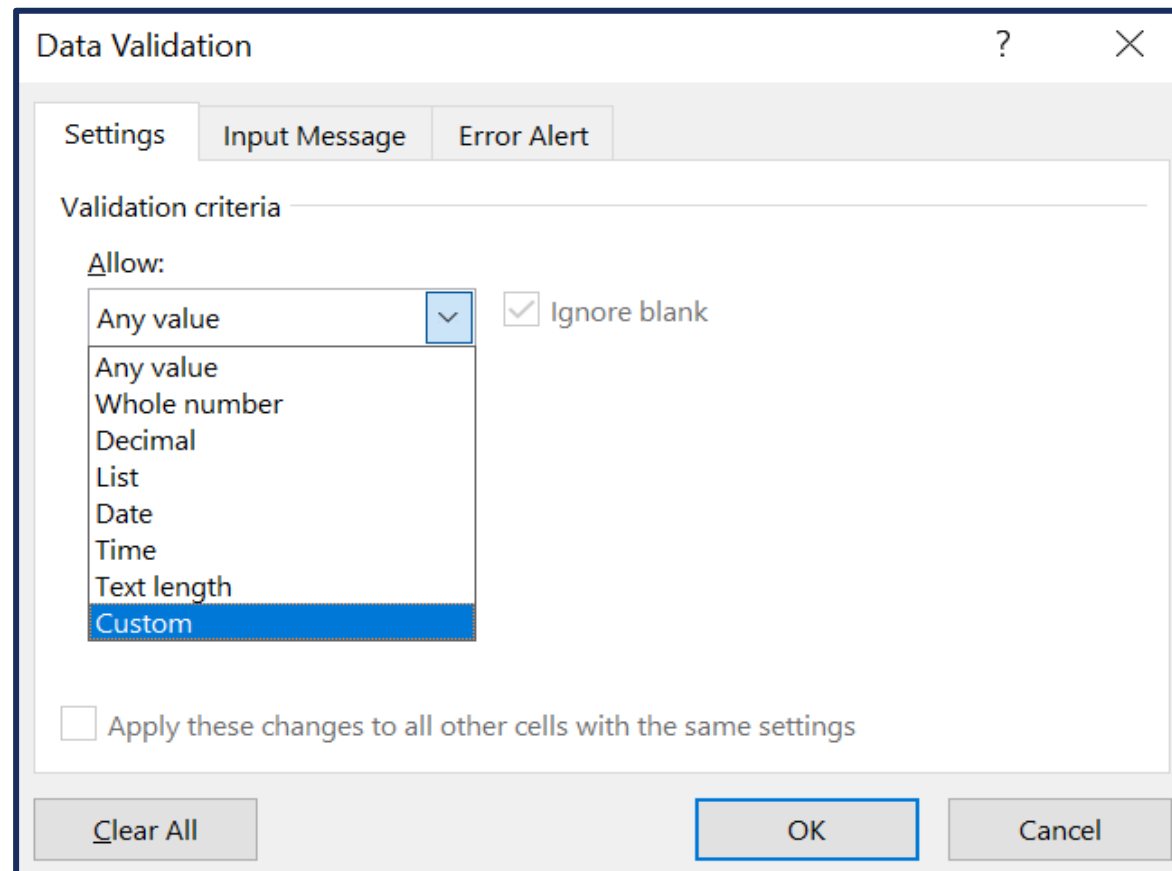
☒ Ignore blank

☐ Apply these changes to all other cells with the same settings

Clear All OK Cancel

'Text length' allows only text within the specified length and a set of rules on the length to be set.

Data Validation: Example



Data Validation

Settings Input Message Error Alert

Validation criteria

Allow:

Any value ☒ Ignore blank

Any value
Whole number
Decimal
List
Date
Time
Text length
Custom

☐ Apply these changes to all other cells with the same settings

Clear All OK Cancel

'Custom' allows custom rules on data to be set.

Key Takeaways

- The sort and filter functionalities are available to order or filter the data for further analysis.
- Group by functionality in Excel allows us to show necessary data for easy viewing and analysis.
- The ungroup option allows us to remove the groups created by group.
- While removing duplicates, we can choose a single column or multiple columns to check the data.
- Data validation applies only to new data entered in the cells where rules are placed.





Knowledge Check

Knowledge Check

1

In which of the following sections can we find Group By and Subtotal under the data tab?

- a. Sort & Filter
- b. Data Tools
- c. Outline
- d. Analyze



Knowledge Check

1

In which of the following sections can we find Group By and Subtotal under the data tab?

- a. Sort & Filter
- b. Data Tools
- c. Outline
- d. Analyze



The correct answer is **c**

Outline section under Data tab allows group by and subtotal.

Knowledge Check

2

Group By within a Group By is possible. True or False.

- a. True
- b. False



Knowledge
Check

2

Group By within a Group By is possible. True or False.

- a. True
- b. False



The correct answer is **a**

True. Group By within a Group By is possible.

Knowledge Check

3

Which of the following options can be used for sorting on multiple columns?

- a. Options
- b. Add Level
- c. Sort On
- d. Order



Knowledge
Check

3

Which of the following options can be used for sorting on multiple columns?

- a. Options
- b. Add Level
- c. Sort On
- d. Order



The correct answer is **b**

Add Level helps to add multiple columns for sorting.

Knowledge Check

4

Pattern matching is possible in filters. True or False.

- a. True
- b. False



Knowledge
Check

4

Pattern matching is possible in filters. True or False.

- a. True
- b. False



The correct answer is **a**

True. Pattern matching is done using regular expressions such as ? and *.

**Knowledge
Check**
5

Which of the following options is used to convert text to columns when there is no delimiters?

- a. Delimiter
- b. Fixed Width
- c. Comma
- d. Space



**Knowledge
Check**
5

Which of the following options is used to convert text to columns when there is no delimiters?

- a. Delimiter
- b. Fixed Width
- c. Comma
- d. Space



The correct answer is **b**

Fixed width allows us to convert data into columns based on the length of each column.

Knowledge Check

6

How to convert a CSV format data into excel?

- a. Use text to columns
- b. Use remove duplicates
- c. Use copy paste to take out each CSV value



Knowledge Check

6

How to convert a CSV format data into excel?

- a. Use text to columns
- b. Use remove duplicates
- c. Use copy paste to take out each CSV value



The correct answer is **a**

Text to columns is the easiest way to convert data to columns

**Knowledge
Check**

7

Is it possible to separate data with multiple delimiters into columns?(Example 1,2,3;4,5 | 6)? True or False.

- a. True
- b. False



**Knowledge
Check**

7

Is it possible to separate data with multiple delimiters into columns?(Example 1,2,3;4,5 | 6)? True or False.

- a. True
- b. False



The correct answer is **a**

True. Multiple delimiters can be specified in Text to Columns

Knowledge Check

8

Why do duplicates occur in a dataset?

- a. Missing validation
- b. Duplicates cannot occur in a dataset
- c. Excel has a feature to create duplicates



Knowledge Check

8

Why do duplicates occur in a dataset?

- a. Missing validation
- b. Duplicates cannot occur in a dataset
- c. Excel has a feature to create duplicates



The correct answer is **a**

Duplicates occur if the input feed has not validated the data and allowed duplicates.

Knowledge Check

9

How do you specify that data has header while removing duplicates?

- a. Click on "My data has headers" Checkbox
- b. Remove headers manually
- c. Cannot be specified



Knowledge Check

9

How do you specify that data has header while removing duplicates?

- a. Click on "My data has headers" Checkbox
- b. Remove headers manually
- c. Cannot be specified



The correct answer is **a**

The "My data has headers" checkbox specifies that the data has headers

**Knowledge
Check**
10

Is it possible to remove rows in a dataset where only one row has duplicates? True or False.

- a. True
- b. False



**Knowledge
Check**
10

Is it possible to remove rows in a dataset where only one row has duplicates? True or False.

- a. True
- b. False



The correct answer is **a**

True. It is possible to remove all rows in a dataset where one column only has duplicates.

**Knowledge
Check**

11

Which of the following options in data validation allows us to validate a list of values?

- a. Any Value
- b. Data
- c. List
- d. Custom



Knowledge
Check

11

Which of the following options in data validation allows us to validate a list of values?

- a. Any Value
- b. Data
- c. List
- d. Custom



The correct answer is **b**

Outline section under Data tab allows group by and subtotal.

**Knowledge
Check**

12

Which of the following range of values can be provided in data validation?

- a. not between
- b. equal to
- c. greater than
- d. between



Knowledge
Check

12

Which of the following range of values can be provided in data validation?

- a. not between
- b. equal to
- c. greater than
- d. between



The correct answer is **d**

Between allows us to set range of values.