

Excel Data Cleaning & Preparation

- Remove Duplicates
- Remove blank spaces
- Remove blank rows
- Change cases
- Fix negative values
- Split data

Remove Duplicates

2 Scenarios:

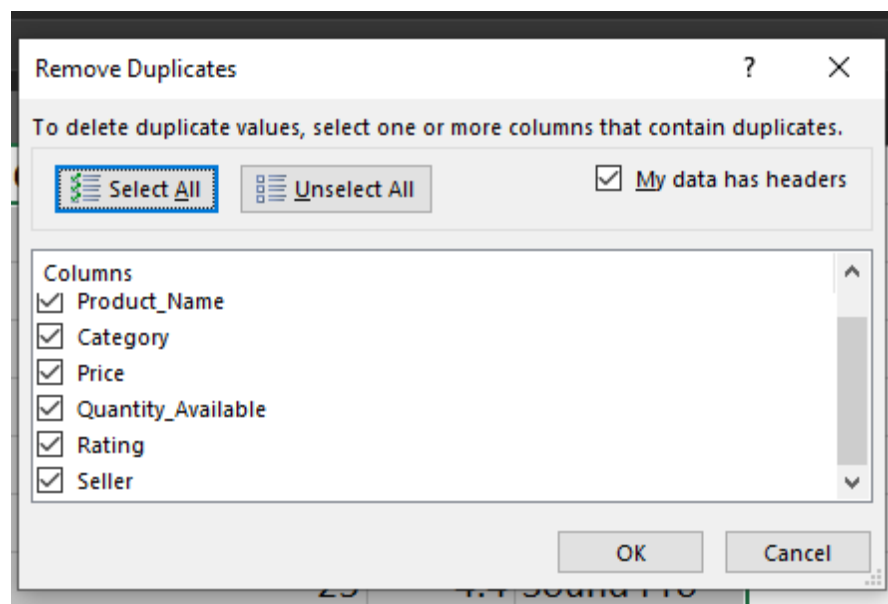
- Entire row duplicate
- Duplicate cell → eg. Duplicate Mob No.

Steps:

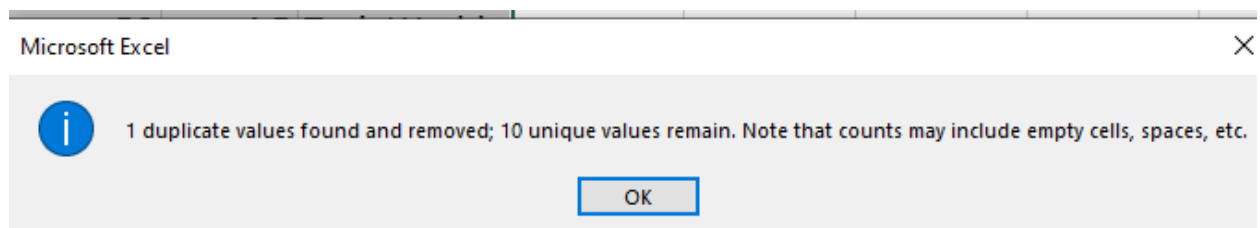
- Select the cells
- Go to Data → Remove Duplicates

File Home Insert Page Layout Formulas Data Review View Help Tell me what you want to do								
Get data			From Text/CSV			From Table/Range		
Recent Sources			Existing Connections			Refresh All		
Get & Transform Data			Queries & Connections			Sort & Filter		
Sort			Filter			Advanced		
Text to Columns			Flash Fill			Remove Duplicates		
Data Validation			Data Tools					

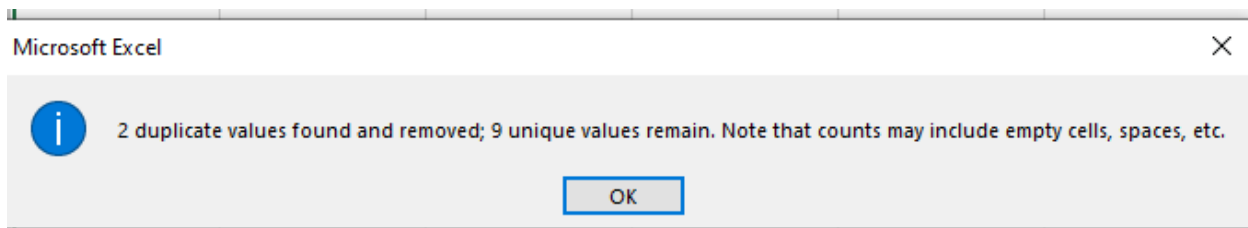
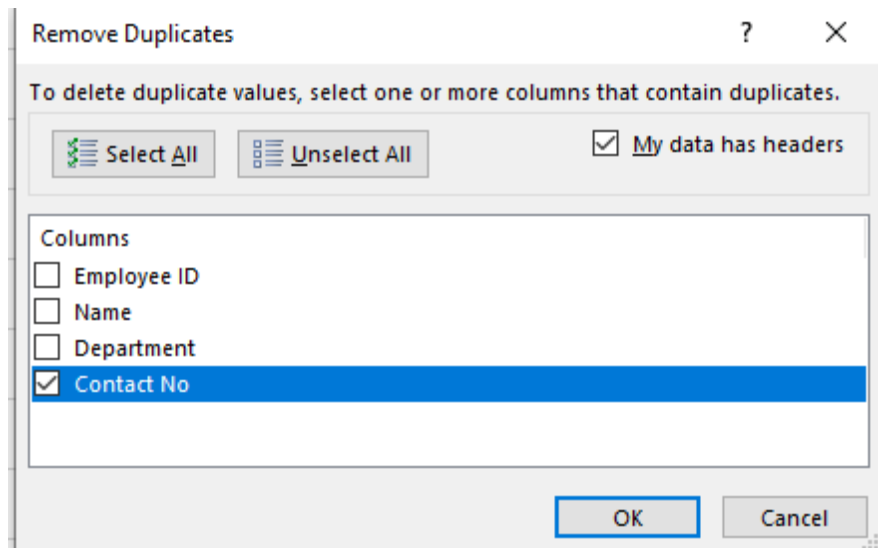
Product_ID	Product_Name	Category	Price	Quantity_Available	Rating	Seller
101	Apple iPhone 14	Mobile	799	50	4.5	Tech World
102	Samsung Galaxy S23	Mobile	999	30	4.7	Gadget Hub
103	Sony WH-1000XM5	Headphones	399	15	4.6	Sound Pro
104	Dell XPS 13	Laptop	1249	10	4.8	Laptop Store
101	Apple iPhone 14	Mobile	799	50	4.5	Tech World
105	HP Spectre x360	Laptop	1349	8	4.7	HP Store
106	Bose QuietComfort 45	Headphones	329	25	4.4	Sound Pro
107	Sony WH-1000XM5	Headphones	399	15	4.6	Sound Pro
108	Lenovo ThinkPad X1	Laptop	1299	6	4.8	Laptop Store
103	Sony WH-1000XM5	Headphones	399	15	4.6	Sound Pro



- If headers are present, keep the *My data has headers* box checked

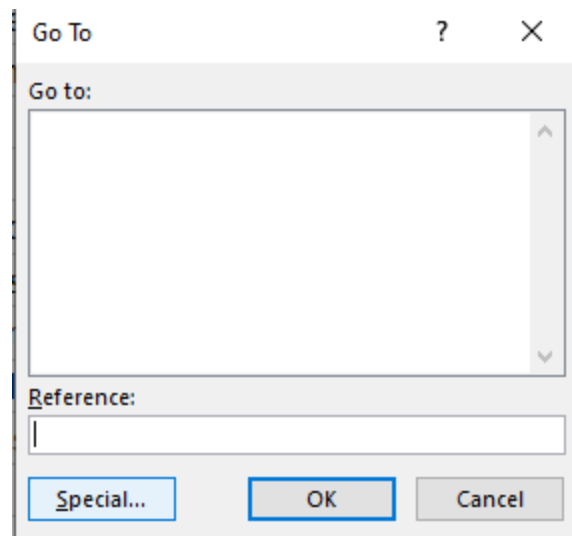


Remove Duplicate contact no.:

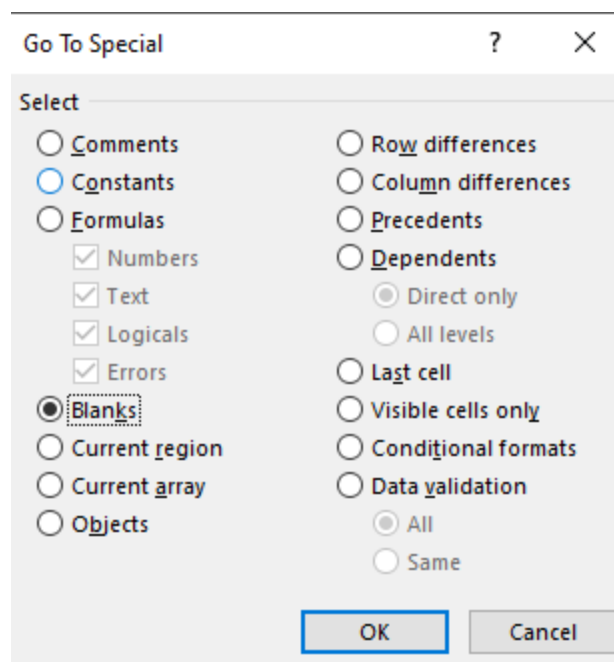


Remove Blank Rows

- Select all data
 - Ctrl + A A (Press A twice)
- Ctrl + G → Special



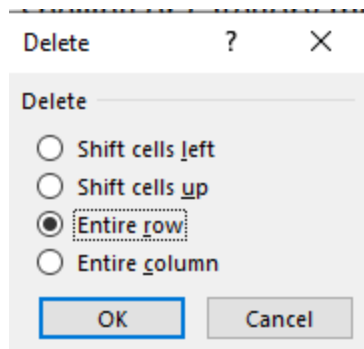
- click **Blanks**



- It will select the blank rows

5	P0004	Product_4	Clothing	903.5	69	Let me know if you'd like further guidance on implementing these steps!
6						
7						
8	P0005	Product_5	Furniture	352.9	87	Use conditional formatting to highlight outliers or filter using IF formulas or slicers.
9	P0006	Product_6	Electronics	549.18	31	Correct spelling or category mismatches (e.g., "Laptops" vs. "Laptop").
10	P0007	Product_7	FOOD	222.56	59	Ensure consistent case for Product_Name (e.g., capitalize each word).
11	P0008	Product_8	Electronics	969.7	78	Fill missing values with appropriate replacements (e.g., average price or a placeholder).
12	P0009	Product_9	Stationery	321.88	92	In Excel, go to Data > Remove Duplicates.
13						
14						
15	P0010	Product_10	FOOD	230.28	-5	Identify duplicate rows based on Product_ID or Product_Name.
16	P0011	Product_11	FOOD	763.66	22	Use Data Validation under the Data tab to restrict input types.
17	P0012	Product_12	Electronics	368.76	45	Ensure Price and Quantity_Available are numeric, and Rating is a decimal.
18	P0013	Product_13	Stationery	669.92	87	Highlight unnecessary columns and delete them or hide them if required for later use.
19	P0014	Product_14	FOOD	914.74	-1	Let me know if you'd like further guidance on implementing these steps!
20	P0015	Product_15	Stationery	137.79	78	Use conditional formatting to highlight outliers or filter using IF formulas or slicers.
21	P0016	Product_16	Furniture	840.84	39	Correct spelling or category mismatches (e.g., "Laptops" vs. "Laptop").
22						
23						
24						
25	P0017	Product_17	Stationery	784.71	96	Ensure consistent case for Product_Name (e.g., capitalize each word).
26	P0018	Product_18	Electronics	588.75	51	Fill missing values with appropriate replacements (e.g., average price or a placeholder).

- Then press → Ctrl + -
 - Entire row



Remove blank spaces

Headphones

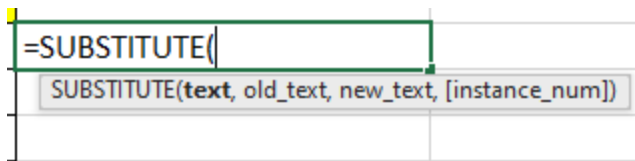
Mobile

=TRIM()

Unbreakable Space

- `=TRIM()` does not work for unbreakable space

Use `=SUBSTITUTE(`



Unbreakable space → `CHAR(160)`

`= SUBSTITUTE (A2, CHAR(160) , "")`

- 🙅 single quote will not work

Fix Letter Case

Category
ElecTronics
Electronics
FOOD
Clothing
Furniture
ElecTronics
FOOD
Electronics
Stationery
FOOD
FOOD
Electronics
StationERY
FOOD
Stationerv

USE → = **PROPER()**

Category	Proper
Electronics	=PROPER(C2)
Electronics	PROPER(text)

Category	Proper
Electronics	Electronics
Electronics	Electronics
FOOD	Food
Clothing	Clothing
Furniture	Furniture
Electronics	Electronics
FOOD	Food
Electronics	Electronics
Stationery	Stationery
FOOD	Food
FOOD	Food
Electronics	Electronics
Stationery	Stationery
FOOD	Food
Stationery	Stationery

ajit Das	Ajit Das
vijay singh	Vijay Singh
ANIL VERMA	Anil Verma

All CAPS

UPPER()

all small

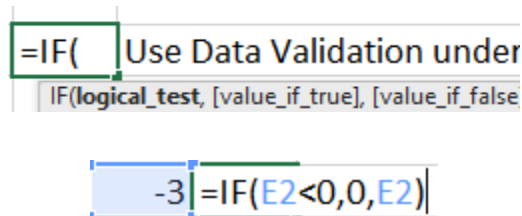
LOWER()

Remove Formula after we adjust the case:

- Select the row
- Copy
- Paste **values**
 - Alt + H +V +V

Fix negative values

=IF()



- 🖐 If the value is less than 0 (**Value<0**) , return 0
 - Else: Return the **value**

-3	0
40	40
26	26
69	69
87	87
31	31
59	59
78	78
92	92
-5	0
22	22
45	45
87	87
-1	0
78	78

Formula to convert **Negative** → **Positive**

-3 =IF(E2<0,-E2,E2)

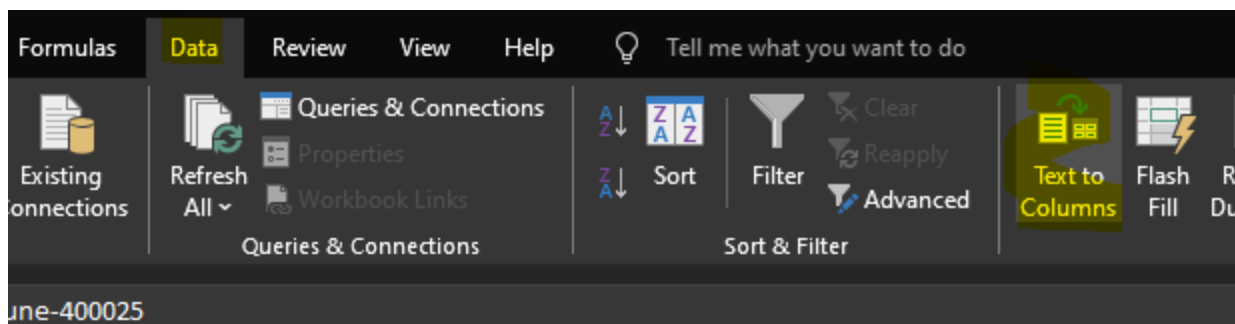
-E2 → Double negative becomes positive

-3	3
40	40
26	26
69	69
87	87
31	31
59	59
78	78
92	92
-5	5

Split Data/Flash fill

Method 1:

- Use text to columns
- Give delimiter



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

Pune	400025
Mumbai	5146351
Nashik	5416
Kolkata	524
Uri	63514


Cancel < Back **Next >** Finish

Pune	400025
Mumbai	514635
Nashik	5416
Kolkata	524
Uri	63514
Puri	6534
Bhaji	6534
Yuvi	6534
Rays	6534

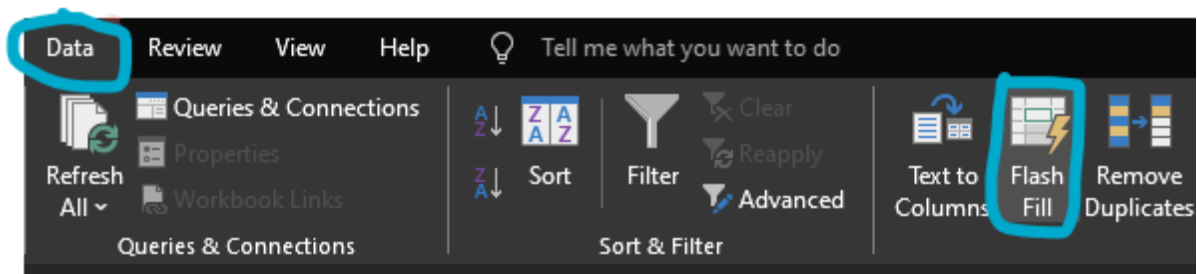
Method 2:

Pune-400025
Mumbai-5146351
Nashik-5416
Kolkata-524
Uri-63514
Puri-6534
Bhaji-65341
Yuvi-654
Rays-654

- Copy & Paste **Pune**
- **Ctrl + E**
- 🖱️ Flash fill
- It'll auto fill the column with respected cities

Pune-400025	Pune	
Mumbai-5146351	Mumbai	
Nashik-5416	Nashik	
Kolkata-524	Kolkata	
Uri-63514	Uri	
Puri-6534	Puri	
Bhaji-65341	Bhaji	
Yuvi-654	Yuvi	
Rays-654	Rays	

OR



Things we can do with Ctrl + E

- CAPS → small
 - Or any case
 - Extract number/text
 - Combining Data
 - If you have columns for **First Name** and **Last Name** (e.g., "John" and "Smith") and want to combine them into **Full Name** (e.g., "John Smith"), type the first combined name in the first row, and then press **Ctrl + E**.
 - Excel will automatically combine the other rows in the same way.
 - Extracting Data (Substrings)
 - **Formatting Numbers**
 - Flash Fill can be used to add specific formatting to numbers (e.g., adding dashes to phone numbers or currency symbols).
- Example:**
- If you have a column with **phone numbers** (e.g., "5551234567") and want to format them as "(555) 123-4567", type the formatted version in the next column, and then press **Ctrl + E** to apply the pattern to the rest of the numbers.

FIND()

=FIND(

FIND(find_text, within_text, [start_num])

=FIND("-",F2)

- 👉 Find -

Pune-400025	5
Mumbai-5146351	7
Nashik-5416	7
Kolkata-524	8
Uri-63514	4
Puri-6534	5
Bhaji-65341	6
Yuvi-654	5
Rays-654	5

LEFT()

- Returns the **leftmost characters** from a text string, based on the number of characters you specify.

LEFT(text, [num_chars])

- **text**: The text string or cell reference containing the text you want to extract from.
- **num_chars** (optional): The number of characters you want to extract from the left. If omitted, it defaults to 1.

Example:

- **LEFT("Hello", 2)** will return "He".
- **LEFT(A1, 3)** will extract the first 3 characters from the value in cell A1.



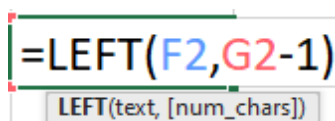
Pune-400025 5 =LEFT(F2,G2)

- Here, 5 is the position of -



Pune-400025	5 Pune-
-------------	---------

- To remove the -, add -1

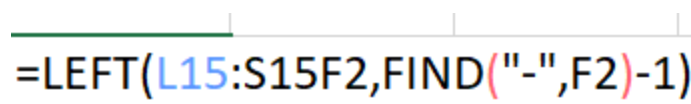


=LEFT(F2,G2-1)

LEFT(text, [num_chars])

OR

- **Without adding an additional column:**



=LEFT(L15:S15F2,FIND("-",F2)-1)

Pune-400025	5	Pune
Mumbai-5146351	7	Mumbai
Nashik-5416	7	Nashik
Kolkata-524	8	Kolkata
Uri-63514	4	Uri
Puri-6534	5	Puri
Bhaji-65341	6	Bhaji
Yuvi-654	5	Yuvi
Rays-654	5	Rays

RIGHT()

- We'll need `LEN()`

	Hyphen Position	Total LENGTH
Pune-400025	5	=LEN(F2

G	H	I
Hyphen Position	Total LENGTH	Len - Hyphen
5	11	=H2-G2

F	G	H	I
	Hyphen Position	Total LENGTH	Len - Hyphen
Pune-400025	5	11	6
Mumbai-5146351	7	14	7
Nashik-5416	7	11	4
Kolkata-524	8	11	3
Uri-63514	4	9	5
Puri-6534	5	9	4
Bhaji-65341	6	11	5
Yuvi-654	5	8	3
Rays-654	5	8	3

- Highlighted column is the characters on right side of hyphen

Now apply the **RIGHT()** Formula

F	G	H	I	J
	Hyphen Position	Total LENGTH	Len - Hyphen	
Pune-400025	5	11	6	=RIGHT(F2,I2)
Mumbai-5146351	7	14	7	RIGHT(text, [num_chars])

OR

- You can directly apply the formula without adding 2 columns

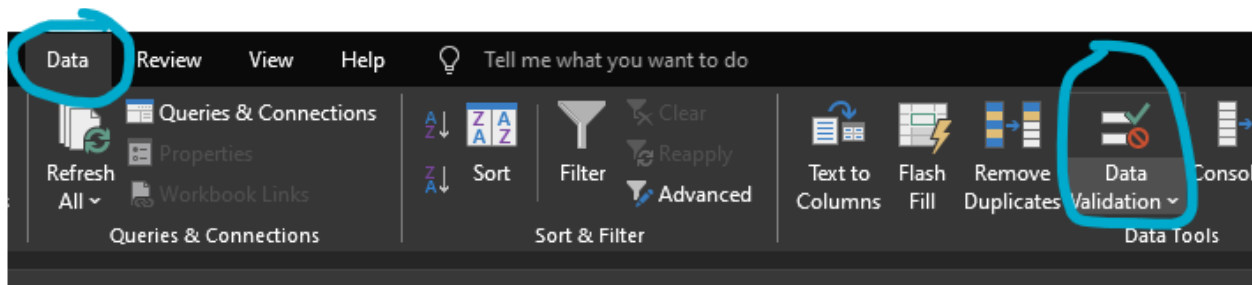
F	G	H	I	J	K	L
	Hyphen Position	Total LENGTH	Len - Hyphen			
Pune-400025	5	11	6	=RIGHT(F2,LEN(F2)-FIND("-",F2))		

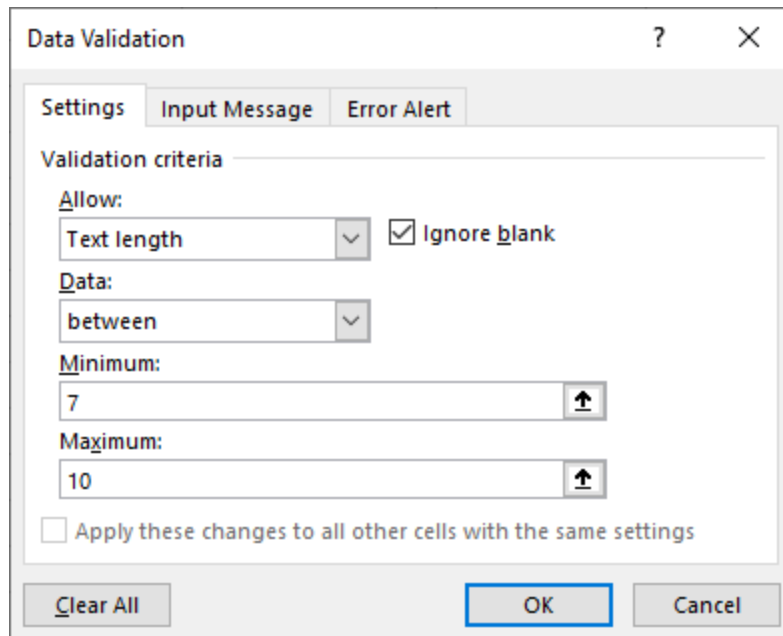
F	G	H	I	J
	Hyphen Position	Total LENGTH	Len - Hyphen	
Pune-400025	5	11	6	400025
Mumbai-5146351	7	14	7	5146351
Nashik-5416	7	11	4	5416
Kolkata-524	8	11	3	524
Uri-63514	4	9	5	63514
Puri-6534	5	9	4	6534
Bhaji-65341	6	11	5	65341
Yuvi-654	5	8	3	654
Rays-654	5	8	3	654

Data Validation

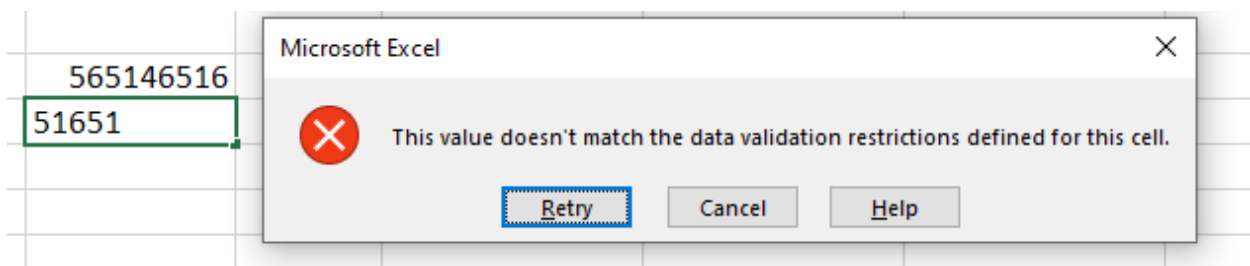
Task: Find out mob no. outside the range of 7 to 10 digits

- Data tab → Data validation





- If you try to enter a number outside this range, it'll throw an error



- 🙌 Here, it's now showing what's the condition.
- You can enter an **error alert**

Data Validation ? X

Settings Input Message Error Alert

☒ Show error alert after invalid data is entered

When user enters invalid data, show this error alert:

Style: Stop Title: Wrong Input

Error message: Enter 7 to 10 digit number

Clear All OK Cancel

565146516
asdd

Wrong Input

Enter 7 to 10 digit number

Retry Cancel Help