

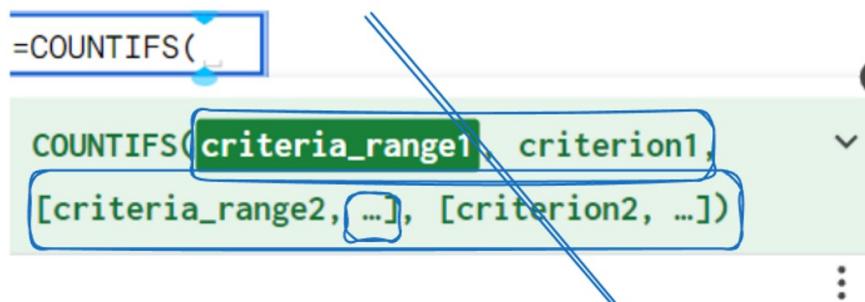
Data Cleaning & Formatting

22 July 2024 21:20

Data Cleaning & Formatting

IF - Condition is being used for 1 criteria to met.

IFS - conditions are being used for multiple criteria's to be met.



1 or more represent
IFS

Challenges :

1. Missing Values.
2. Data Types Issue.
3. Duplicates Entry.
4. Splitting Text To Columns
5. Replacing Values.

Missing Values:

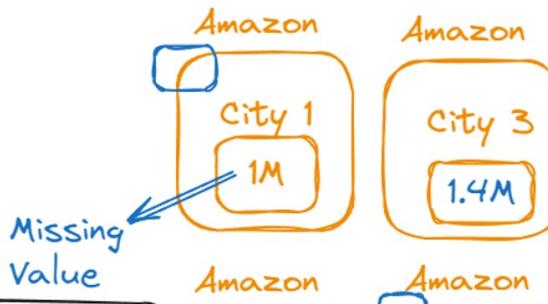
→ NA / Delete

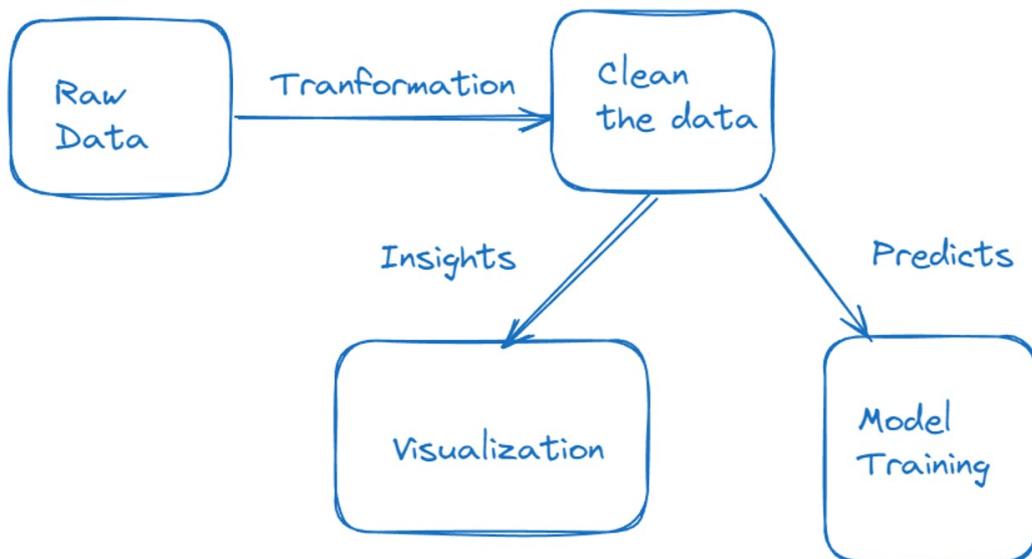
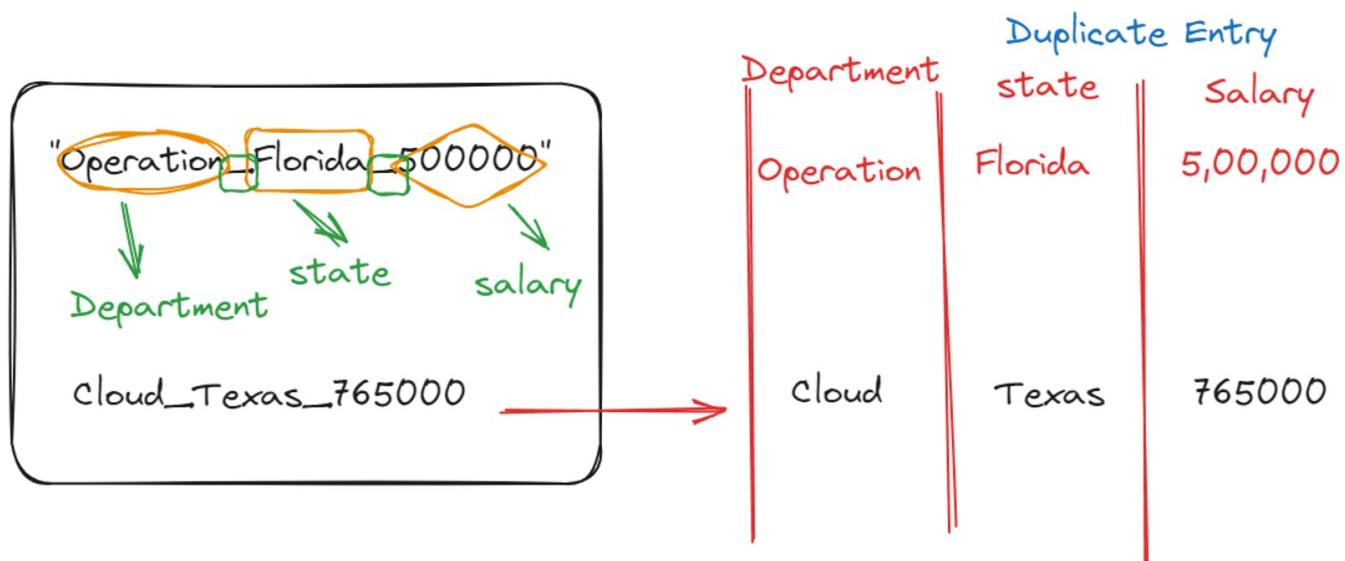
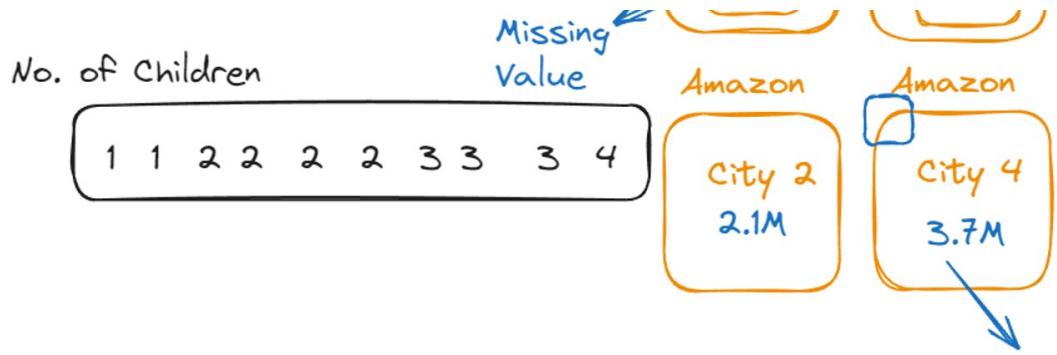
→ Mean/Mode/Median

No. of Children

Krishna	- 100
Devendra	- 290
Vinay	- 230
Puneet	-
Aman	- 400
Krishna	- 100
Devendra	- 290
Vinay	- 230
Puneet	-
Aman	- 400

Krishna - 200





Removing Duplicate Entry

1. UNIQUE()
2. Remove Duplicate as a setting.

2. Remove Duplicate as a setting.

19849	198691 AA Batteries (4-pack)	1	3.84	582 Chestnut St, San Francisco, CA 94016
-------	------------------------------	---	------	--

19849 17318 = 2531 [duplicate]

B1041 ▾ | fx =UNIQUE(Data!A1:F19849)

	Unique!A1			
19833	=UNIQUE(Data	UNIQUE([range], [by_column], [exactly_once])	590	
19834	206871 AAA		689	
19835	nan		nan	
19836	204428 Lightning Charging Cable	1	14.95	347
19837	206148 AA Batteries (4-pack)	1	3.84	38
19838	nan		nan	
19839	205741 Macbook Pro Laptop	1	1700	863
19840	204785 27in FHD Monitor	1	149.99	516
19841	201404 AAA Batteries (4-pack)	2	2.99	75
19842	201186 ThinkPad Laptop	2	999.99	112
19843	198821 AAA Batteries (4-pack)	2	2.99	59
19844	207863 27in FHD Monitor	1	149.99	377
19845	208239 AAA Batteries (4-pack)	2	2.99	321
19846	195045 Wired Headphones	1	11.99	95
19847	202021 Bose SoundSport Headphones	1	99.99	811
19848	202681 27in 4K Gaming Monitor	1	389.99	844
19849	198691 AA Batteries (4-pack)	1	3.84	582
19850				
19851				

Add 1000 more rows at the bottom

+ Index Data Unique cell reference

17318	198691 AA Batteries (4-pack)	1	3.84	582 Chestnut St, San Francisco, CA 94016
-------	------------------------------	---	------	--

17318

A1 ▾ | fx =UNIQUE(Data!A:F)

	A	B
1	Order ID	Product
2	194095	Wired Headphones

Data Tools Extensions Help

- Sort sheet
- Sort range
- Create a filter
- Create group by view
- Create filter view
- Add a slicer

E	F
ark City, NY 10001	1556.2
as, TX 75001	10743.31
v York City, NY 10001	15828.97
ork City, NY 10001	13791.10

as, TX 75001 10743.31
v York City, NY 10001 15828.97
ork City, NY 10001 13791.10
e, WA 98101 15496.48
Francisco, CA 94016 11341.98
on, MA 02215 33509.81
TX 75001 32703.00
, WA 98101 13130.91
/A 98101 32703.91

n Francisco, CA 94016 39736.37
York City, NY 10001 17938.84
Austin, TX 73301 7626.042
le, WA 98101 26968.08

Remove duplicates

2531 duplicate rows found and removed.

17318 unique rows remain.

OK

Krishna - 29
Aman - 22
Karan - 23
Manav - 24
Krishna - 33
Krishna - 32
Krishna - 44
Krishna - 33

B. Handling Missing Data

1. Explanation:

- Missing data can lead to biased results.

2. Example:

- Use `=IF(ISBLANK(A2), "Unknown", A2)` to replace blank cells with "Unknown".

2. Example:

- Use `=IF(ISBLANK(A2), "Unknown", A2)` to replace blank cells with "Unknown".

Formatting Data with correct Data Type

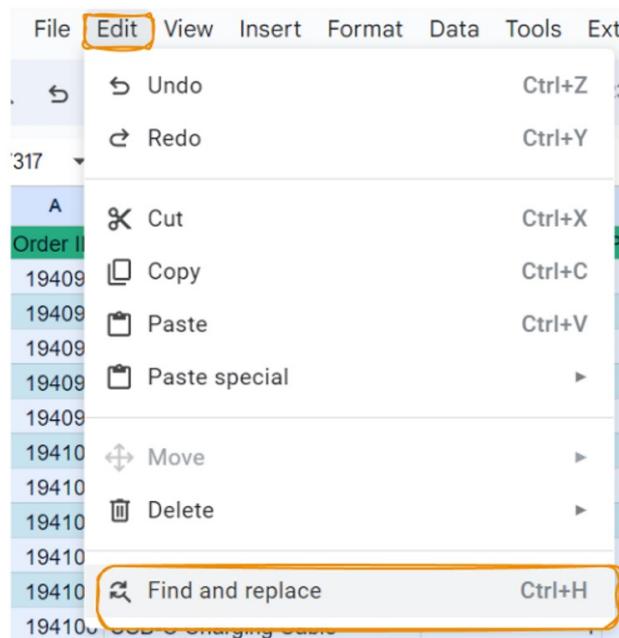
The screenshot shows the Microsoft Excel ribbon with the 'Format' tab selected. In the main area, there's a table with columns for Date, Price, and Quantity. The 'Date' column contains dates like '04/04/1904', '30/05/1929', etc. The 'Format' menu is open, and the 'Number' option is selected. A dropdown menu lists various number formats: Number, Percent, Scientific, Accounting, Financial, Currency, and Currency rounded. One of the options, 'Date', is highlighted with a green box and shows the value '26/09/2008'.

`ISBLANK()` - Finding the empty cell - `True` / `False`

The screenshot shows a table with columns F, G, and H. The 'Date' column (F) contains dates like '04/04/1904', '30/05/1929', etc. The 'G' column contains the formula `=IF(ISBLANK(F2), "Unkonwn", F2)`. The formula is highlighted with an orange box. An orange arrow points from the formula in cell G2 to the text 'True/False' located below the table.

Regex - Find & Replace

Regex - Find & Replace



regex - Empty cell

regex for empty cell in google sheet

All Videos Images Shopping News Books Web More Tools

Did you mean: regex for empty cell in google *sheets*

In the "Find" field, enter the regular expression `^\s*$` to search for blank cells. This expression matches cells that are empty or contain only whitespace.

Find and replace

Find

Replace with

Search Data!A1:F17317

Match case
 Match entire cell contents
 Search using regular expressions [Help](#)
 Also search within formulae
 Also search within links

Replaced 17,256 instances of `^\s*$` with Unknown



TRIM() : Whenever you want to remove some leading , trailing whitespaces.

The screenshot shows a Microsoft Excel spreadsheet titled "Copy of Data Cleaning and formatting". The Data tab is selected in the ribbon. A context menu is open over a table, with the "Data clean-up" option highlighted. Under "Data clean-up", the "Trim whitespace" option is also highlighted with a green box.

The screenshot shows the Microsoft Excel formula bar with the formula =TRIM(G5) entered. Below the formula bar, the cell G5 contains the text "Krishna Madan".

Price Formatting : Custom Currency for Indian Rupees.

The screenshot shows the "Custom currencies" dialog box from Google Sheets. It lists various currencies with their symbols and exchange rates. The Indian Rupee is selected, indicated by a green box around its row.

Currency	Symbol	Rate
US Dollar	\$	\$1,000.12
Indian Rupee	₹	₹1,000.12
British Pound Sterling	£	£1,000.12
Afghan Afghani	Af.	1,000 Af.
Albanian Lek	Lek	Lek1,000
Algerian Dinar	din	din1,000.12

The screenshot shows a context menu for a cell in a spreadsheet. The 'Format' tab is selected. The 'Number' section is highlighted with a green border. Other options include 'Text', 'Alignment', 'Wrapping', 'Rotation', 'Font size', 'Merge cells', 'Convert to table' (New), 'Conditional formatting', 'Alternating colours', 'Clear formatting' (Ctrl+V), and a preview of the current row's data.

	Albanian Lek	Lek1,000
	Algerian Dinar	din1,000.12
	Angolan Kwanza	Kz1,000.12
	Argentine Peso	\$1,000.12

Freezing the row is important for referencing.

The screenshot shows the 'View' menu open, with the 'Freeze' option selected. A dropdown menu for freezing rows and columns is displayed, with '1 row' and '1 column' selected and highlighted with orange boxes.

	Lightning Charging Cable	1	₹11.99	330 1st St, Seattle, WA 98101
7in FHD Monitor	1	₹14.99		
Vired Headphones	1	₹11.99	330 1st St, Seattle, WA 98101	
JSB-C Charging Cable	1	₹11.95	261 Jackson St, San Francisco, CA 94016	