

# End-to-End Data Cleaning and Standardization Workflow in

## Excel

Prepared a raw client dataset for analysis and reporting using Excel.



Project Type: Practical Excel Workflow Demonstration

Presented by: Seema Kumari  
Tool Used: Microsoft Excel

# Business Problem

The client dataset could not be directly used for reporting because:

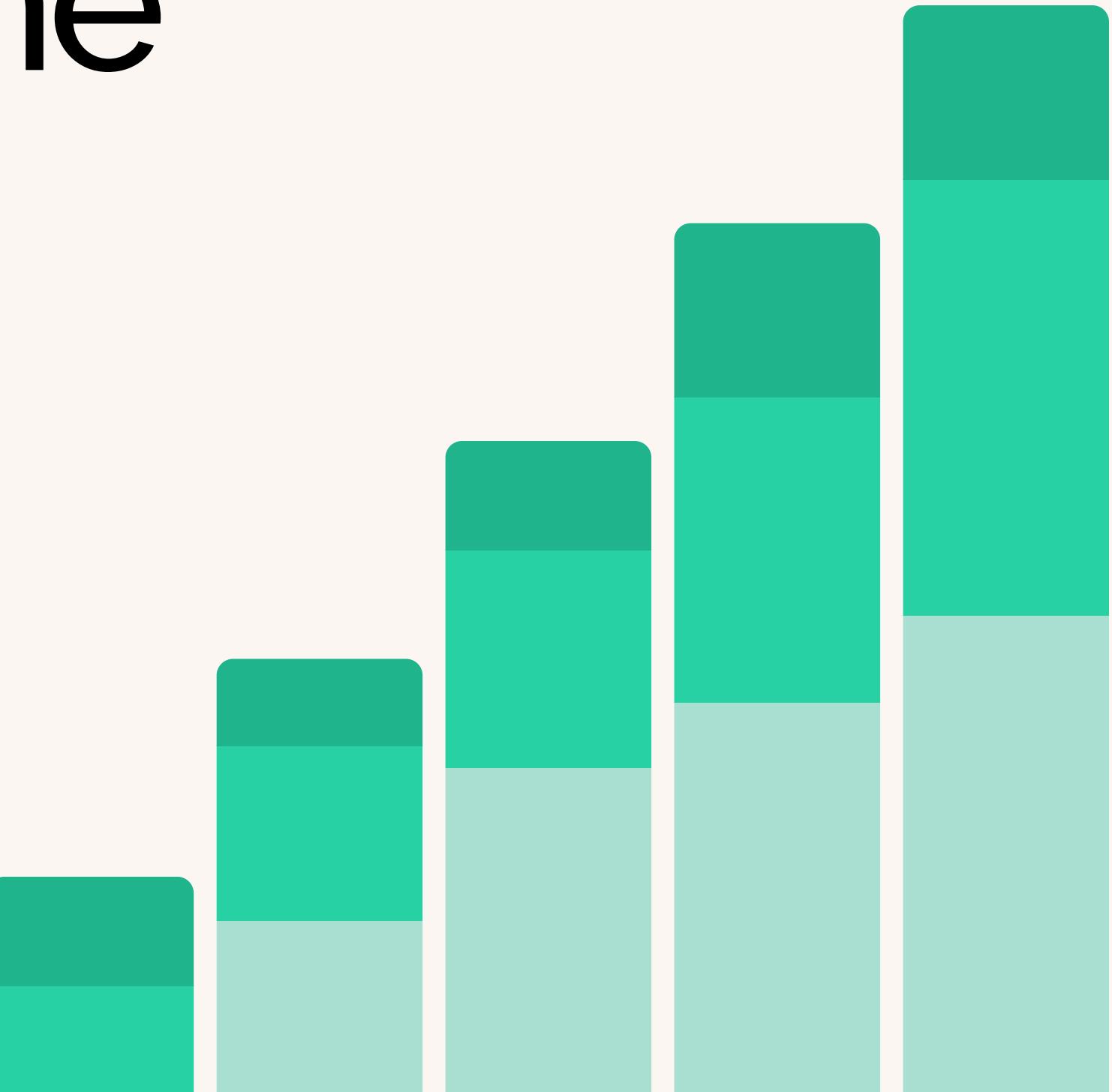
- Inconsistent naming conventions
- Missing and duplicate values
- Combined fields reducing clarity
- Formula errors affecting accuracy
- Poor formatting impacting readability

Impact: Reports based on this data would be unreliable.



# Objective of the Project

- Improve data quality and consistency
- Ensure accuracy in reporting metrics
- Transform raw data into an analysis-ready format
- Maintain data integrity using Excel best practices





# Problem Statement

The original dataset contained multiple issues:

- Inconsistent name formats
  - Extra spaces and unwanted text
  - Combined columns (Department + Region)
  - Duplicate records
  - Blank cells and formula errors
  - Poor visual presentation
  - 📌 This data was not ready for analysis or reporting.
- 

# Solution Approach

- Followed a structured, step-by-step Excel data cleaning workflow
- Used:
  1. Excel formulas
  2. Built-in Excel tools
  3. Keyboard shortcuts
- Focused on:
  4. Accuracy
  5. Standardization
  6. Professional presentation



# Step 0 – Data Backup

- Created a duplicate copy of the original sheet
- Ensured:
  - Data safety
  - No loss of original information

**Shortcut Used:**

**Ctrl + Drag sheet tab**



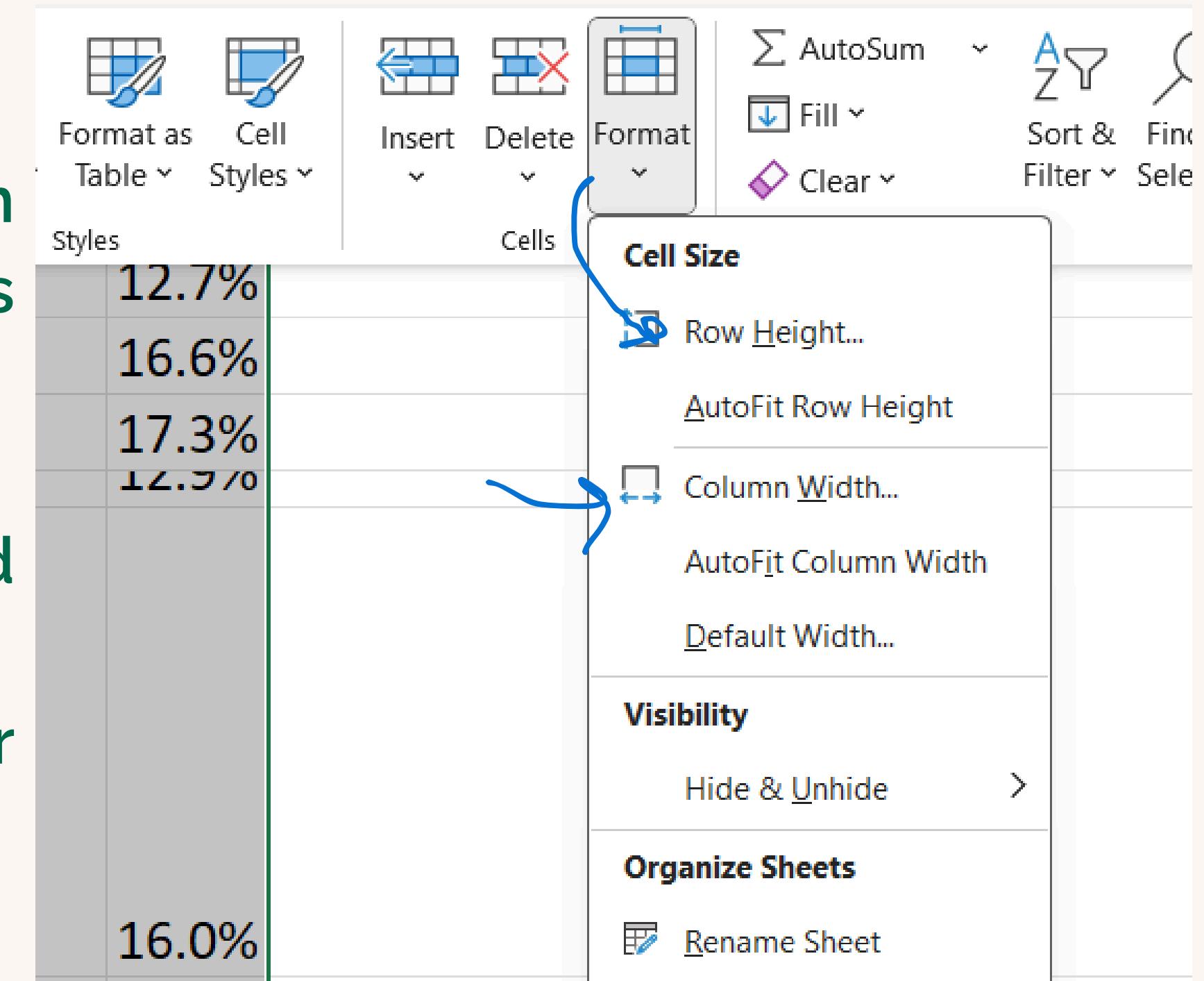
# Step 1: Data Visibility Check

## Action Taken:

Adjusted row and column dimensions to ensure all data points were fully visible.

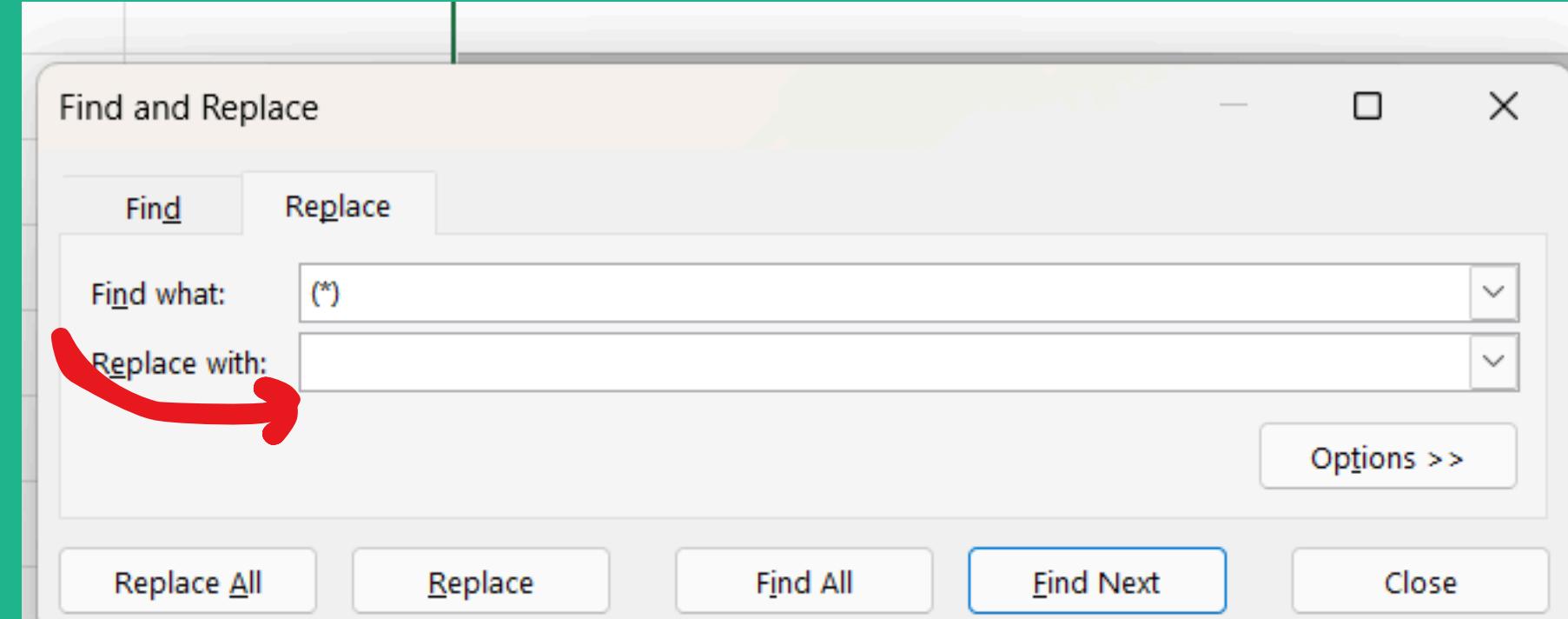
## Analytical Reasoning

- Hidden data can lead to missed errors
- Full visibility is required for accurate assessment



# Step 2: Noise Removal in Text Fields

- Client names had extra information in brackets
- Removed unwanted text using Find and Replace



The screenshot shows a 'Find and Replace' dialog box overlaid on a spreadsheet. The dialog has tabs for 'Find' and 'Replace', with 'Replace' selected. The 'Find what:' field contains the wildcard '(\*)'. The 'Replace with:' field is empty, indicated by a red arrow. Below the dialog are several rows of data in the spreadsheet:

31-05-2023	MORGAN STANLEY (XNYS:MS)
31-05-2023	CITIGROUP INC. (XNYS:C)
31-05-2023	BANK OF AMERICA CORPORATION (XNYS:BAC)
31-05-2023	WALMART INC. (XNYS:WMT)
31-05-2023	TARGET CORPORATION (XNYS:TGT)
31-05-2023	COSTCO WHOLESALE CORPORATION (XNAS:COST)

# Step 3: Text Case Standardization

- Converted text into lowercase for consistency
  - Formula:LOWER(cell)
- 
- Pasted values to lock the cleaned data
  - Deleted the old column

Client	Client
AMAZON.COM, INC.	=LOWER(C3)
TESLA, INC.	
NETFLIX, INC.	
THE PROCTER & GAMBLE COMPANY	
THE GOLDMAN SACHS GROUP, INC.	

# Step 4: Spacing & Naming Consistency

- Removed extra spaces and fixed capitalization
- Used nested formula:  
`=PROPER(TRIM(cell))`

Result:

- Clean and readable text
- Standard naming format

D	E
Contact	
=TRIM(E3)	
	Bill SmITH
	KEN Singh
	Harley Fritz
	Nyla Novak

D	E
Contact	
=PROPER(TRIM(E3))	
	PROPER(text)
	Bill SMITH
	KEN Singh
	Harley Fritz
	Nyla Novak
	David Rasmussen
	IVAN HINEY
	JONha Ma

# Step 5: Data Structure Optimization

Split combined columns into separate analytical fields.

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

Cloud Tech	Texas
Strategy	New York
Strategy	New York
Operations	Florida
Operations	Florida
Cloud Tech	Texas

Cancel < Back Next > Finish

Big Data	California	Transfer	\$3,713
Big Data	California	Transfer	\$4,950
Big Data	California	Transfer	\$4,750
Operations	Florida	Transfer	\$810
Operations	Florida	Transfer	\$7,320

iuu\_tech\_texas

Convert Text to Columns Wizard - Step 1 of 3

The Text Wizard has determined that your data is Delimited.

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:

3 Cloud Tech_Texas
4 Strategy_New York
5 Strategy_New York
6 Operations_Florida
7 Operations_Florida
8 Cloud Tech_Texas

Cancel < Back Next > Finish

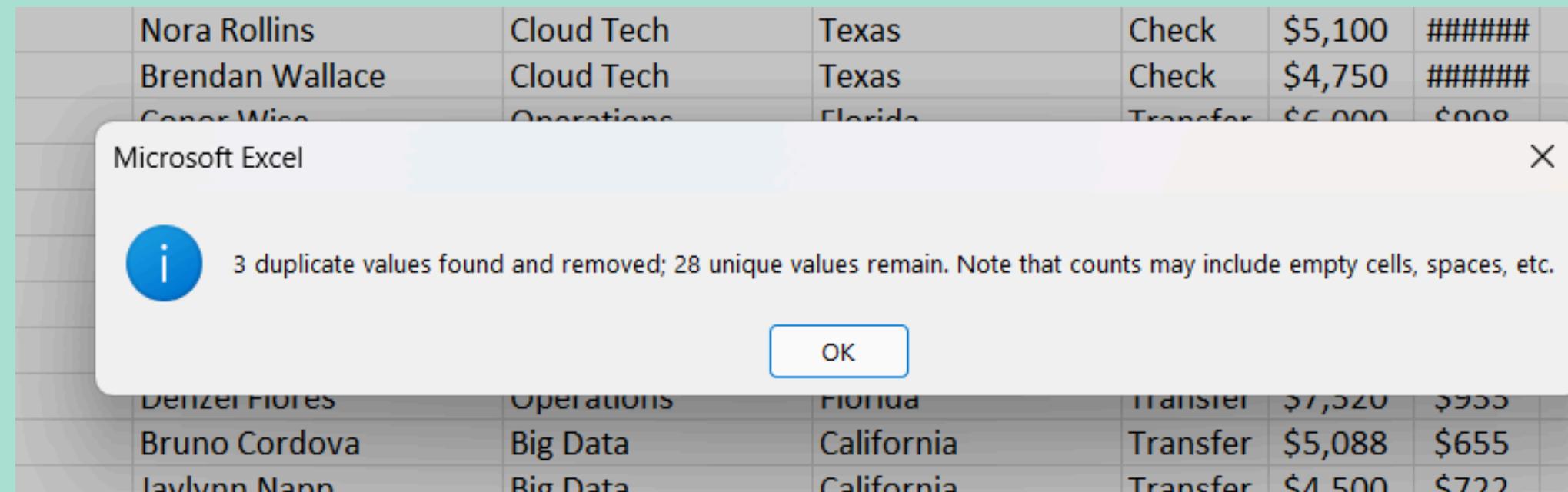
Big Data_California	Transfer	\$3,713
Big Data_California	Transfer	\$4,950
Big Data_California	Transfer	\$4,750
Operations_Florida	Transfer	\$810
Operations_Florida	Transfer	\$7,320

# Step 6: Duplicate Record Resolution

- Duplicate data can cause wrong analysis
- Removed duplicates using Excel's built-in feature

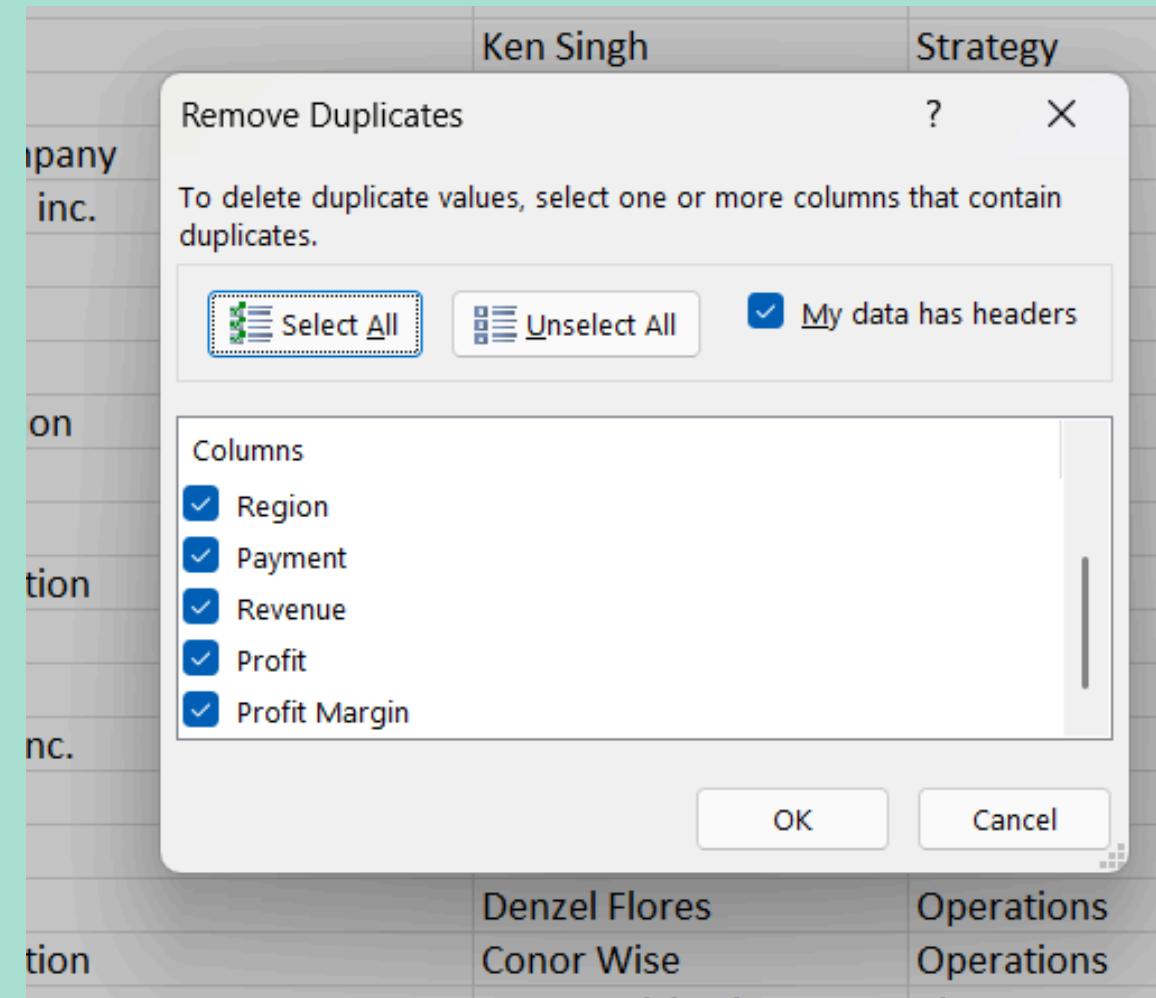
## Method:

- Data Tab → Remove Duplicates



A screenshot of a Microsoft Excel spreadsheet titled "Microsoft Excel". The spreadsheet contains data in columns A through F. A message box is overlaid on the screen, indicating that 3 duplicate values were found and removed, leaving 28 unique values. The message box has an "OK" button at the bottom.

Nora Rollins	Cloud Tech	Texas	Check	\$5,100	#####
Brendan Wallace	Cloud Tech	Texas	Check	\$4,750	#####
Conor Wise	Operations	Florida	Transfer	\$6,000	\$600
Microsoft Excel					
Denzel Flores	Operations	Florida	Transfer	\$7,520	\$955
Bruno Cordova	Big Data	California	Transfer	\$5,088	\$655
Iavlyn Nann	Big Data	California	Transfer	\$4,500	\$722



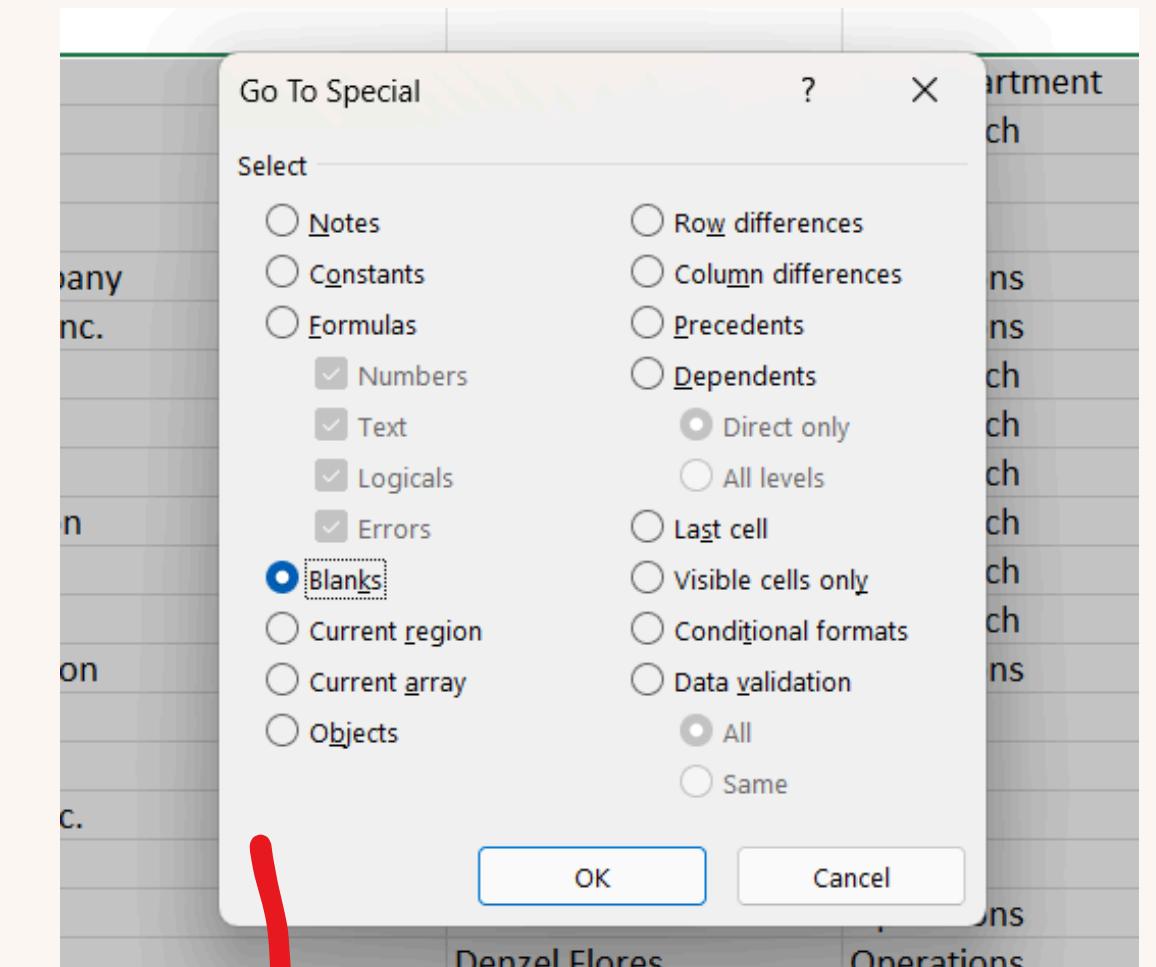
# Step 7 - Handle Missing Values

- Blank cells were present in the dataset
- Filled them with “N/A”

## Steps:

- Go To Special → Blanks
- Type N/A → Press Ctrl + Enter

F	G	H	I	J
Region	Payment	Revenue	Profit	Profit
Texas	Transfer	\$4,500	\$598	
New York	PayPal	\$3,800	#####	
New York	NA	\$3,713	#####	
Florida	NA	NA	\$779	#V
Florida	Check	\$5,000	\$684	
Texas	Transfer	\$6,100	\$544	
Texas	Transfer	\$4,625	\$670	
Texas	Transfer	\$3,800	#####	
Texas	Card	\$3,600	#####	
Texas	Check	\$5,100	#####	
Texas	Check	\$4,750	#####	
Florida	Transfer	\$6,000	\$998	
California	Check	\$4,500	\$780	
California	Card	NA	#####	#V
California	Transfer	\$3,713	#####	



# Step 8 – Handle Formula Errors

- Implemented error-handling logic in formulas
- Prevents broken reports
- Ensures outputs remain usable for stakeholders

Region	Payment	Revenue	Profit	Profit Margin	
North America	Transfer	\$4,500	\$598	13.3%	
Europe	PayPal	\$3,800	\$1,045	27.5%	
Asia	NA	\$3,713	\$1,009	27.2%	
Australia	NA	NA	\$779	NA	
Latin America	Check	\$5,000	\$684	13.7%	
Middle East	Transfer	\$6,100	\$544	8.9%	
Africa	Transfer	\$4,625	\$670	14.5%	

H	I	J	K
venue	Profit	Profit Margin	
,500	\$598	13.3%	
,800	\$1,045	27.5%	
,713	\$1,009	27.2%	
NA	\$779	#VALUE!	
,000	\$684	13.7%	
,100	\$544	8.9%	
,625	\$670	14.5%	
NA	NA	NA	

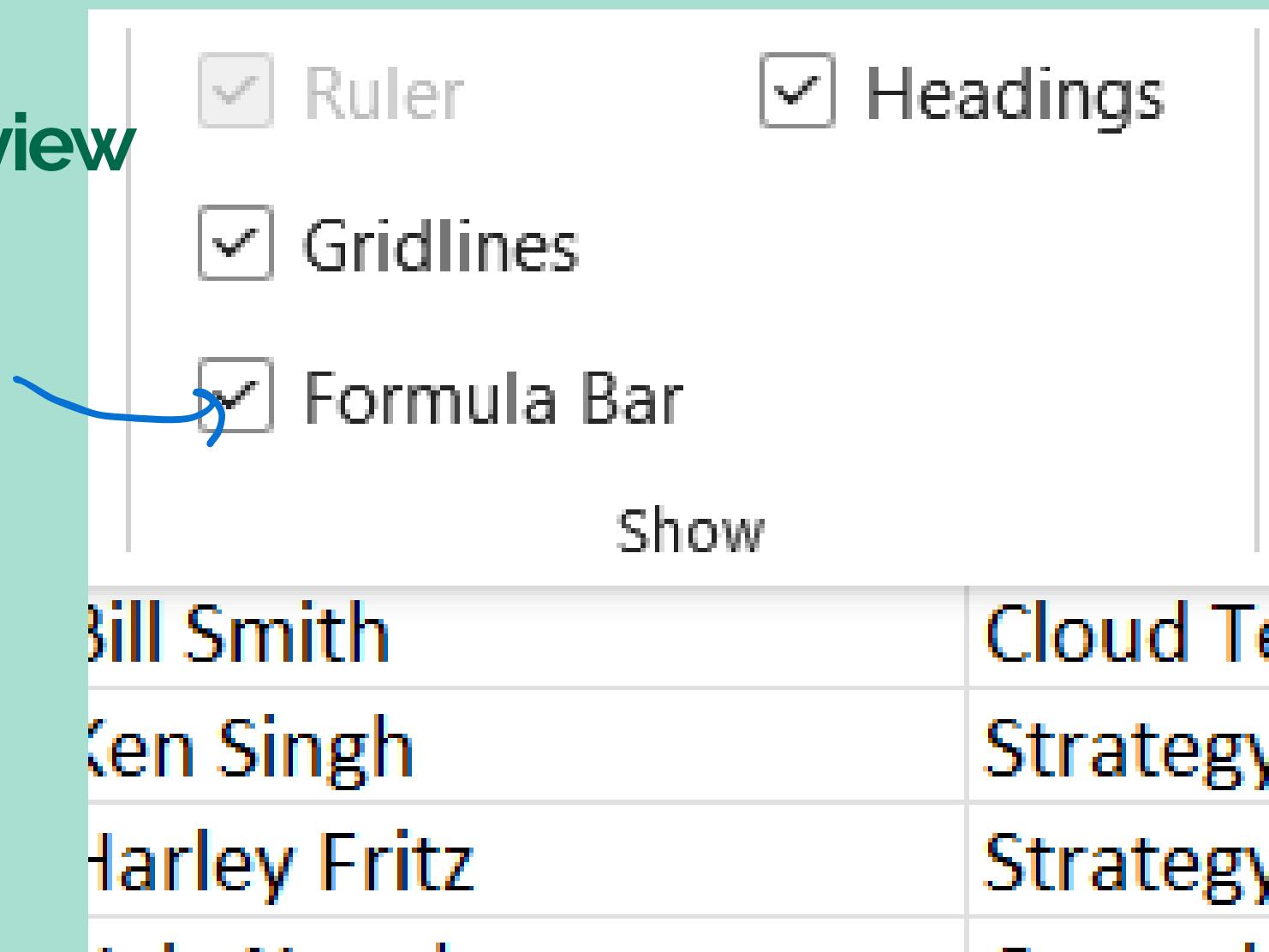
# Step 9 - Reporting Format Enhancement

- Applied professional header formatting
- Improves readability for business users
- Aligns with reporting standards

Date	Client
30-05-2023	amazon.com, inc.
30-05-2023	tesla, inc.
30-05-2023	netflix, inc.
30-05-2023	the procter & gamble company
30-05-2023	the goldman sachs group, inc.
21 05 2022	inmigrant choco & co

# Step 10: Presentation Readiness

- Removed gridlines to deliver a clean report view
- Reduces visual noise
- Makes the dataset presentation-ready



# Outcome & Impact

## Result

- Clean, standardized, error-free dataset
- Ready for analysis and reporting

## Business Impact

- Increased confidence in data
- Reduced manual correction effort
- Enabled faster and more accurate insights

Date	Client	Contact	Department	Region	Payment	Revenue	Profit	Profit Margin
30-05-2023	amazon.com, inc.	Bill Smith	Cloud Tech	Texas	Transfer	\$4,500	\$598	13.3%
30-05-2023	tesla, inc.	Ken Singh	Strategy	New York	PayPal	\$3,800	\$1,045	27.5%
30-05-2023	netflix, inc.	Harley Fritz	Strategy	New York	NA	\$3,713	\$1,009	27.2%
30-05-2023	the procter & gamble company	Nyla Novak	Operations	Florida	NA	NA	\$779	NA
30-05-2023	the goldman sachs group, inc.	David Rasmussen	Operations	Florida	Check	\$5,000	\$684	13.7%
31-05-2023	jpmorgan chase & co.	Ivan Hiney	Cloud Tech	Texas	Transfer	\$6,100	\$544	8.9%
31-05-2023	morgan stanley	Jonha Ma	Cloud Tech	Texas	Transfer	\$4,625	\$670	14.5%
31-05-2023	citigroup inc.	Jordan Boone	Cloud Tech	Texas	Transfer	\$3,800	\$2,045	53.8%
31-05-2023	bank of america corporation	Kylee Townsend	Cloud Tech	Texas	Card	\$3,600	\$1,564	43.4%
31-05-2023	walmart inc.	Nora Rollins	Cloud Tech	Texas	Check	\$5,100	\$1,220	23.9%
31-05-2023	target corporation	Brendan Wallace	Cloud Tech	Texas	Check	\$4,750	\$1,435	30.2%
31-05-2023	costco wholesale corporation	Conor Wise	Operations	Florida	Transfer	\$6,000	\$998	16.6%
31-05-2023	mcdonald's corporation	Steven Michael	Big Data	California	Check	\$4,500	\$780	17.3%
01-06-2023	exxon mobil corporation	Lucia Mckay	Big Data	California	Card	NA	\$1,044	NA
01-06-2023	verizon communications inc.	Jose Roach	Big Data	California	Transfer	\$3,713	\$1,222	32.9%
01-06-2023	the home depot, inc.	Franklin Wright	Big Data	California	Transfer	\$4,950	\$1,065	21.5%
01-06-2023	cisco systems, inc.	Alia Thornton	Operations	Florida	Transfer	\$4,750	\$810	17.1%
01-06-2023	chevron corporation	Denzel Flores	Operations	Florida	Transfer	\$7,320	\$933	12.7%
01-06-2023	at&t inc.	Bruno Cordova	Big Data	California	Transfer	\$5,088	\$655	12.9%
01-06-2023	intel corporation	Jaylynn Napp	Big Data	California	Transfer	\$4,500	\$722	16.0%
01-06-2023	general motors company	Bruce Rich	Big Data	California	Card	\$4,250	\$901	21.2%
02-06-2023	microsoft corporation	Arturo Moore	Big Data	California	PayPal	\$5,250	\$1,349	25.7%
02-06-2023	comcast corporation	Bryce Carpenter	Strategy	New York	PayPal	\$6,500	\$1,288	19.8%
02-06-2023	dell technologies inc.	Jaidyn Andersen	Strategy	New York	PayPal	\$7,500	\$1,664	22.2%
02-06-2023	johnson & johnson	Mark Walm	Strategy	New York	Transfer	\$5,500	\$1,320	24.0%
02-06-2023	fedex corporation	Harry Lee	Strategy	New York	Transfer	\$4,625	\$1,001	21.6%
02-06-2023	general electric company	Josh Johnson	Strategy	New York	Transfer	\$4,500	\$960	21.3%
02-06-2023	lockheed martin corporation	Mik Naam	Strategy	New York	Card	\$5,400	\$540	10.0%

# THANK YOU!

