STRUCTURING YOUR DATA THE RIGHT WAY THE FIRST TIME📈

Sandhya Kambhampati @sandhya__k


Yeah, I have a lot of questions.

PDFS

ARE THE WORST

@sandhya__k

GOVERNMENT FILE STRUCTURE

C	D	E	F	G
Sonderauswertung Personal am 15.12.2013 in Pflegeheimen auf Kreisebene in Bayern				
			Stationäre Pflege	
Personal Insgesamt	Auszubildende/-r, (Um-)Schüler/-in	Überwiegender Tätigkeitsbereich im Pflegeheim		
		Pflege und Betreuung	soziale Betreuung	zusätzliche Betreuung (§ 87b SGB XI)
917	59	599	38	32
5 210	443	3 675	171	227

@sandhya__k

FOIA DOCUMENTS

	A	B	C	D	E	F	G	H	I	J	K
1		Ticket Sales.	$103,410.68	Include revenue received for sales of admissions to athletics events. Include ticket sales to the public, faculty and students, and money received for shipping and handling of tickets. Do not include ticket sales for conference and national tournaments that are pass-through transactions. Report amounts in excess of a ticket's face value paid by ticket purchasers (for example, to obtain preferential seating) in Category 4 (Contributions).							

Revenues by Source	Men's Teams Only Ticket Sales. 1	Women's Teams Only Ticket Sales. 1	Not Allocated by Gender Ticket Sales. 1
Archery			
Badmiton			
Baseball			

Summary		
	Men's	$97,432.97
	Women's	$5,977.71
	No Gender	$0.00
	Grand Total	$103,410.68

@sandhya__k

THE BASICS

- Rows
- Columns
- Column Headers
- Cells

COLUMN

COLUMN HEADER

	A	B	C	D
1	first_name	last_name	phone_number	email_address
2				
3				
4				
5				
6				
7				
8				
9				
10				

ROW

CELL

@sandhya__k

WHY DOES STRUCTURE MATTER?

ARE THESE THE SAME?

A	B	C
Student	**test_score_1**	**test_score_2**
Bob Smith	88	67
Anita Doe	100	100
John Phillips	76	87

Student	Bob Smith	Anita Doe	John Phillips
test_score_1	88	100	76
test_score_2	67	100	87

THINK ABOUT FUNCTIONS

SORT

AGGREGATE

TRANSFORM

FILTER

@sandhya__k

STOP THIS! 🗑️ 🔥

	A	B	C	D	E	F
1	unique id	total3	total2	namef	namel	other1
2	1	64328	392902	Sam	Smith	738393
3	2	42910	2048349	Bob	Miller	28102
4	3	27293	2302938	Sandhya	Kambhampa	1281912
5	4	91213	2392302	Sam	Jones	12182913
6	5	42323	23826	Nikil	Patel	4289328943
7	6	492303	3422	Samantha	Smith	32849
8	7	23900	3423422	John	Doe	29401

@sandhya__k

EXAMPLE FILE

http://bit.ly/2lw2JAV

@sandhya__k

INSTEAD OF THIS

A	B	C	D	E	F	G
ID	Player	Date_of_birth	Day	Month	Year	Age
1	Sergio Romero	2/22/87	22	February	1987	29
2	Nahuel Guzmán	2/10/86	10	February	1986	30
3	Mariano Andújar	7/30/83	30	July	1983	33
4	Pablo Zabaleta	1/16/85	16	January	1985	31
5	Martín Demichelis	12/20/80	20	December	1980	35
6	Marcos Rojo	3/20/90	20	March	1990	26
7	Nicolás Otamendi	2/12/88	12	February	1988	28
8	Ramiro Funes Mori	3/5/91	5	March	1991	25
9	Facundo Roncaglia	2/10/87	10	February	1987	29
10	Gabriel Mercado	3/18/87	18	March	1987	29
11	Emmanuel Más	1/15/89	15	January	1989	27
12	Mateo Musacchio	8/26/90	26	August	1990	25
13	Javier Mascherano	6/8/84	8	June	1984	32
14	Ángel Di María	2/14/88	14	February	1988	28

@sandhya__k

USE FLAGS

A	B	C	D	E	F	G	H
ID	**Player**	**Date_of_birth**	**Month**	**Day**	**Year**	**Age**	**Checked**
1	Sergio Romero	2/22/87	February	22	1987	30	1
2	Nahuel Guzmán	2/10/86	February	10	1986	31	1
3	Mariano Andújar	7/30/83	July	30	1983	33	0
4	Pablo Zabaleta	1/16/85	January	16	1985	32	1
5	Martín Demichelis	12/20/80	December	20	1980	36	1
6	Marcos Rojo	3/20/90	March	20	1990	26	1
7	Nicolás Otamendi	2/12/88	February	12	1988	29	0
8	Ramiro Funes Mori	3/5/91	March	5	1991	25	0
9	Facundo Roncaglia	2/10/87	February	10	1987	30	0
10	Gabriel Mercado	3/18/87	March	18	1987	29	0
11	Emmanuel Más	1/15/89	January	15	1989	28	0
12	Mateo Musacchio	8/26/90	August	26	1990	26	0

@sandhya__k

File Edit View Insert Format Data Tools Add-ons Help All changes saved in Drive

$ %

Number ▸

Font size ▸

B Bold ⌘B

I Italic ⌘I

U Underline ⌘U

S Strikethrough Option+Shift+5

Align ▸

Merge cells ▸

Text wrapping ▸

Text rotation ▸

Conditional formatting...

Alternating colors...

Clear formatting ⌘\

B			F	G	H	
layer	Date_		Year	Age	Weekday	Caps
ergio Romero		22	1987	30	1	
ahuel Guzmán		10	1986	31	2	
Mariano Andújar		30	1983	33	7	
ablo Zabaleta		16	1985	32	4	
Martín Demichelis		20	1980	36	7	
Marcos Rojo		20	1990	26	3	
icolás Otamendi		12	1988	29	6	
amiro Funes Mori		5	1991	25	3	
acundo Roncaglia		10	1987	30	3	
abriel Mercado		18	1987	29	4	
mmanuel Más		15	1989	28	1	
Mateo Musacchio		26	1990	26	1	
avier Mascherano		8	1984	32	6	
ngel Di María	2/14/88	February	14	1988	29	1

@sandhya__k

Conditional format rules

Single color | Color scale

Apply to range

L1:L1000

Format cells if...

Text contains

1

Formatting style

Custom

B I U S A

Done | Cancel

Add another rule

ranks_pflege_datadictionary.md

```
## Data Dictionary for `ranks_pflege_data.xlsx`
____
**Source**: Data comes from Statisches Landesamt for December 2013.
**Note**: This file contains the rankings for each county for the major data
questions we asked.
__________
### Contents

*All files contain the same two beginning A and B columns (`Name` and `State`
), where `name` means the name of the district or county and `state` is the
state the county is found in.*

#### Sheets
* **More than 1 bed**
    - `Rank`: ranking of the total # beds that are in rooms with more than 1
    beds in the county or district in descending order
    - `Total of >1bed`: total number of beds that are in rooms with more than
    1 bed
    - **About this data**: the rankings of the data show that `Berlin` has the
    highest number of beds overall at `11,896`. This shows that `Berlin` has a
    lot more older nursing homes than other places as they have a lot more
    rooms with more than 1 bed.
* **Ratio of beds**
    - `Total of >1 beds`: total number of beds that are in rooms with more
    than 1 bed
    -  `All beds`: total number of beds in the county or district
    -  `Ratio of >1bed to allbeds`: `total of >1 beds`/ `all beds`; in other
    words, what percentage of beds are in rooms with more than 1 bed
    -  `Rank`: rank of the `ratio of >1bed to allbeds`
```

DATA DICTIONARY

GOOD HEADERS & CLEAN DATA

	A	B	C	D	E	F
1	unique_id	total_debt	total_revenue	first_name	last_name	taxes
2	1	64328	392902	Sam	Smith	738393
3	2	42910	2048349	Bob	Miller	28102
4	3	27293	2302938	Sandhya	Kambhampa	1281912
5	4	91213	2392302	Sam	Jones	12182913
6	5	42323	23826	Nikil	Patel	4289328943
7	6	492303	3422	Samantha	Smith	32849
8	7	23900	3423422	John	Doe	29401

QUESTIONS?
THANK YOU!

@sandhya__k