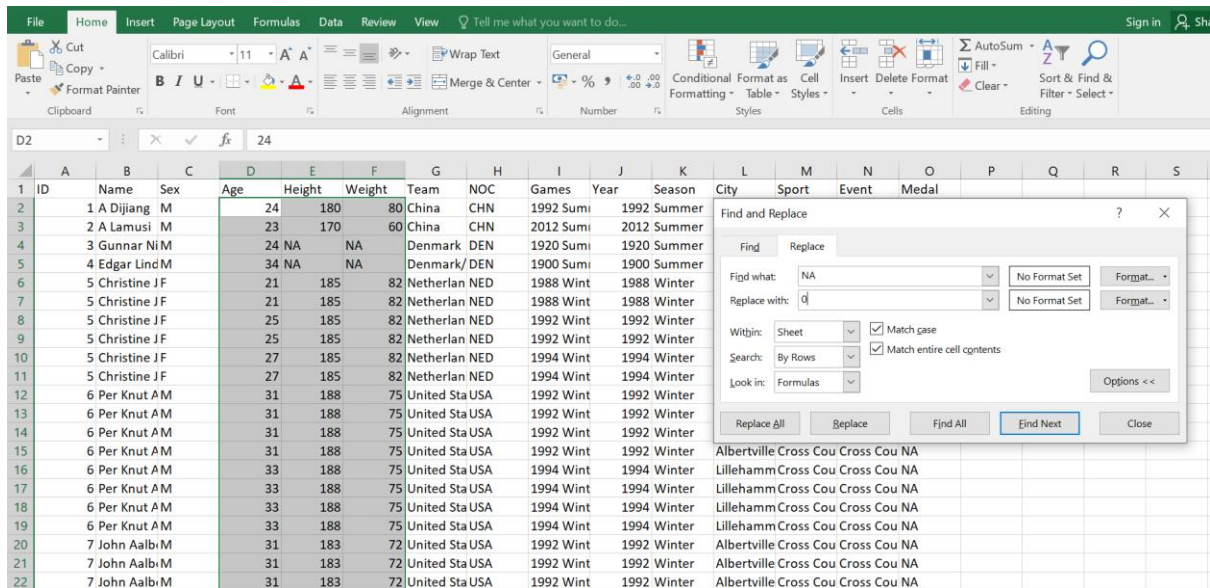


Data Cleaning can be done in Excel or Python

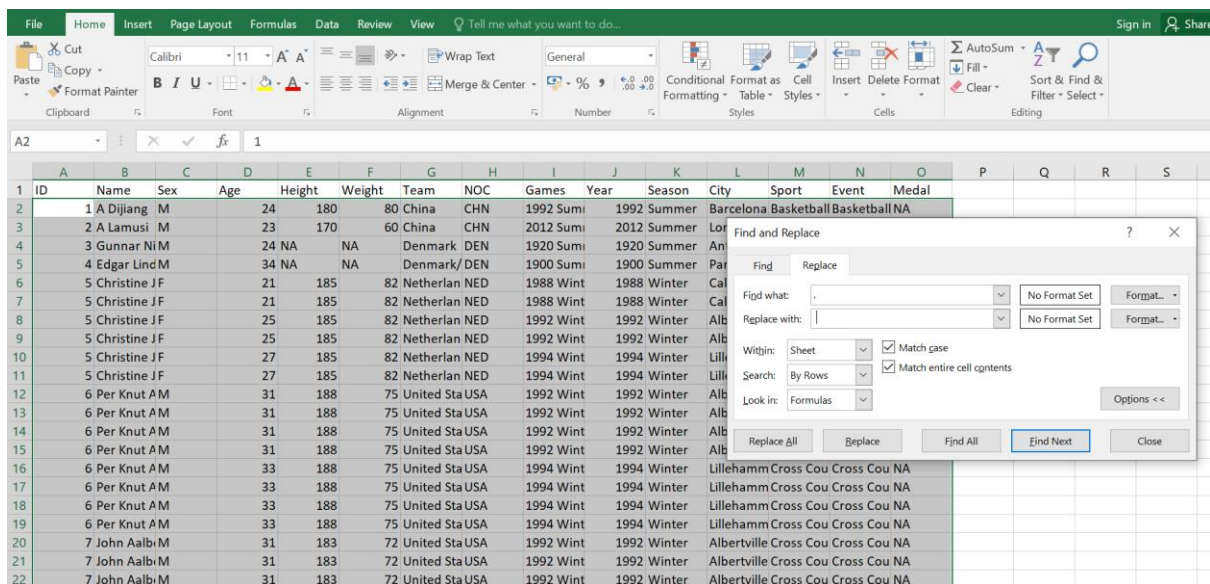
Replace NA from numeric columns to 0.



ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
1	A Djiang	M	24	180	80	China	CHN	1992 Sumi	1992 Summer					
2	A Lamusi	M	23	170	60	China	CHN	2012 Sumi	2012 Summer					
3	Gunnar NiM	M	24	NA	NA	Denmark	DEN	1920 Sumi	1920 Summer					
4	Edgar LindM	M	34	NA	NA	Denmark/ DEN		1900 Sumi	1900 Summer					
5	Christine JF	F	21	185	82	Netherlan	NED	1988 Wint	1988 Winter					
6	Christine JF	F	21	185	82	Netherlan	NED	1988 Wint	1988 Winter					
7	Christine JF	F	25	185	82	Netherlan	NED	1992 Wint	1992 Winter					
8	Christine JF	F	25	185	82	Netherlan	NED	1992 Wint	1992 Winter					
9	Christine JF	F	27	185	82	Netherlan	NED	1994 Wint	1994 Winter					
10	Christine JF	F	27	185	82	Netherlan	NED	1994 Wint	1994 Winter					
11	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter					
12	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter					
13	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter					
14	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter					
15	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter					
16	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter					
17	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter					
18	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter					
19	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter					
20	John AalbiM	M	31	183	72	United Sta	USA	1992 Wint	1992 Winter					
21	John AalbiM	M	31	183	72	United Sta	USA	1992 Wint	1992 Winter					
22	John AalbiM	M	31	183	72	United Sta	USA	1992 Wint	1992 Winter					

As we will be importing the dataset in csv format, commas inside entities can cause problems.

Therefore, replace commas with space.



ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
1	A Djiang	M	24	180	80	China	CHN	1992 Sumi	1992 Summer		Barcelona	Basketball	Basketball	NA
2	A Lamusi	M	23	170	60	China	CHN	2012 Sumi	2012 Summer					
3	Gunnar NiM	M	24	NA	NA	Denmark	DEN	1920 Sumi	1920 Summer					
4	Edgar LindM	M	34	NA	NA	Denmark/ DEN		1900 Sumi	1900 Summer					
5	Christine JF	F	21	185	82	Netherlan	NED	1988 Wint	1988 Winter					
6	Christine JF	F	21	185	82	Netherlan	NED	1988 Wint	1988 Winter					
7	Christine JF	F	25	185	82	Netherlan	NED	1992 Wint	1992 Winter					
8	Christine JF	F	25	185	82	Netherlan	NED	1992 Wint	1992 Winter					
9	Christine JF	F	27	185	82	Netherlan	NED	1994 Wint	1994 Winter					
10	Christine JF	F	27	185	82	Netherlan	NED	1994 Wint	1994 Winter					
11	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter					
12	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter					
13	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter					
14	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter					
15	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter					
16	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter					
17	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter					
18	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter					
19	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter					
20	John AalbiM	M	31	183	72	United Sta	USA	1992 Wint	1992 Winter					
21	John AalbiM	M	31	183	72	United Sta	USA	1992 Wint	1992 Winter					
22	John AalbiM	M	31	183	72	United Sta	USA	1992 Wint	1992 Winter					

Eliminate all the duplicate values.

File Home Insert Page Layout Formulas Data Review View Tell me what you want to do... Sign in Share

From Access From Web From Other Sources Existing Connections New Query Recent Sources Show Queries From Table From Table Refresh All Edit Links Connections Properties Sort Filter Clear Reapply Advanced Text to Columns Data Validation Data Tools What-If Analysis Forecast Sheet Subtotal

A2 1

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
1	A Djiang	M	24	180	80	China	CHN	1992 Sumi	1992 Summer	Barcelona	Basketball	Basketball	NA	
2	A Lamusi	M	23	170	60	China	CHN	2012 Sumi	2012 Summer	London	Judo	Judo	NA	
3	Gunnar Ni	M	24	NA	NA	Denmark	DEN	1920 Sumi	1920 Summer	Antwerpe	Football	Football	NA	
4	Edgar Lind	M	34	NA	NA	Denmark/	DEN	1900 Sumi	1900 Summer	Paris	Tug-Of-W	Tug-Of-W	Gold	
5	Christine JF	F	21	185	82	Netherlan	NED	1988 Wint	1988 Winter	Calgary	Speed	Speed	NA	
5	Christine JF	F	21	185	82	Netherlan	NED	1992 Wint	1992 Winter	Albertville	Speed	Speed	NA	
5	Christine JF	F	25	185	82	Netherlan	NED	1992 Wint	1992 Winter	Albertville	Speed	Speed	NA	
5	Christine JF	F	27	185	82	Netherlan	NED	1994 Wint	1994 Winter	Lillehamm	Speed	Speed	NA	
5	Christine JF	F	27	185	82	Netherlan	NED	1994 Wint	1994 Winter	Lillehamm	Speed	Speed	NA	
6	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA	
6	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA	
6	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA	
6	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA	
6	Per Knut AM	M	31	188	75	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA	
6	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter	Lillehamm	Cross	Cross	NA	
6	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter	Lillehamm	Cross	Cross	NA	
6	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter	Lillehamm	Cross	Cross	NA	
6	Per Knut AM	M	33	188	75	United Sta	USA	1994 Wint	1994 Winter	Lillehamm	Cross	Cross	NA	
7	John Aalb	M	31	183	72	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA	
7	John Aalb	M	31	183	72	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA	
7	John Aalb	M	31	183	72	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA	
7	John Aalb	M	31	183	72	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA	
7	John Aalb	M	31	183	72	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA	

Remove Duplicates

To delete duplicate values, select one or more columns that contain duplicates.

Select All Unselect All My data has headers

Columns

- ☒ ID
- ☒ Name
- ☒ Sex
- ☒ Age
- ☒ Height

OK Cancel

athlete_events

Ready Average: 15615.05597 Count: 4066740 Sum: 19098150354 100%

60	China	CHN	2012 Sumi	2012 Summer	London	Judo	Judo Men	NA
A	Denmark	DEN	1920 Sumi	1920 Summer	Antwerpe	Football	Football M	NA
A	Denmark/	DEN	1900 Sumi	1900 Summer	Paris	Tug-Of-W	Tug-Of-W	Gold
82	Neth							ka Speed Ska
82	Neth							ka Speed Ska
82	Neth							ka Speed Ska
82	Neth							ka Speed Ska
82	Neth							ka Speed Ska
82	Neth							ka Speed Ska
75	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA
75	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA
75	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA
75	United Sta	USA	1992 Wint	1992 Winter	Albertville	Cross	Cross	NA

Microsoft Excel

1385 duplicate values found and removed; 269731 unique values remain.

OK