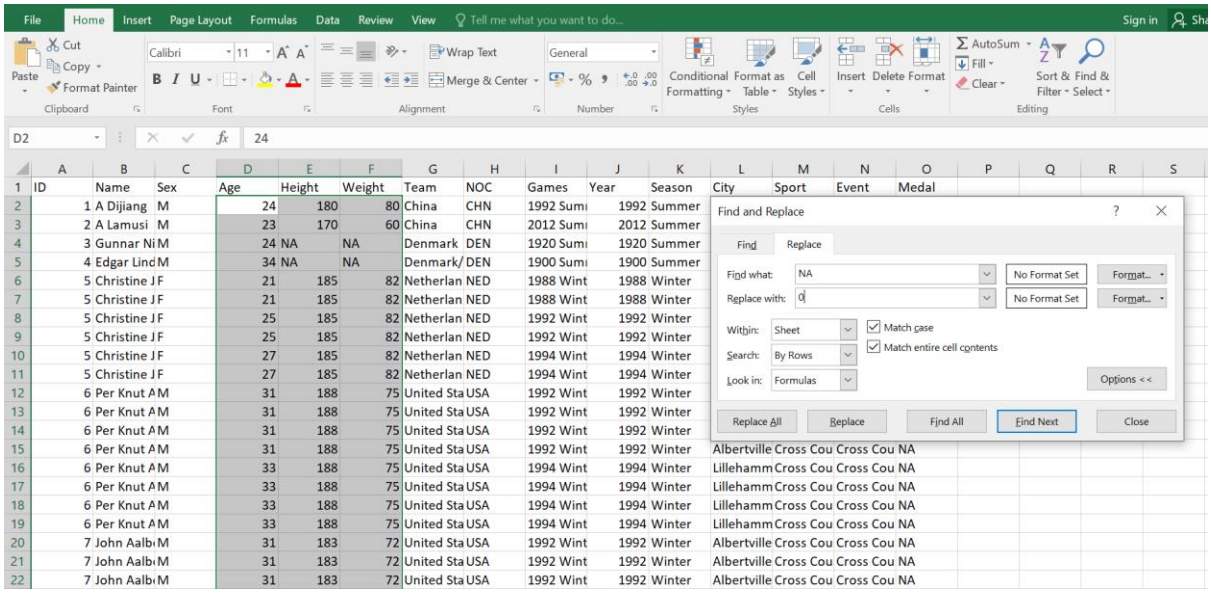
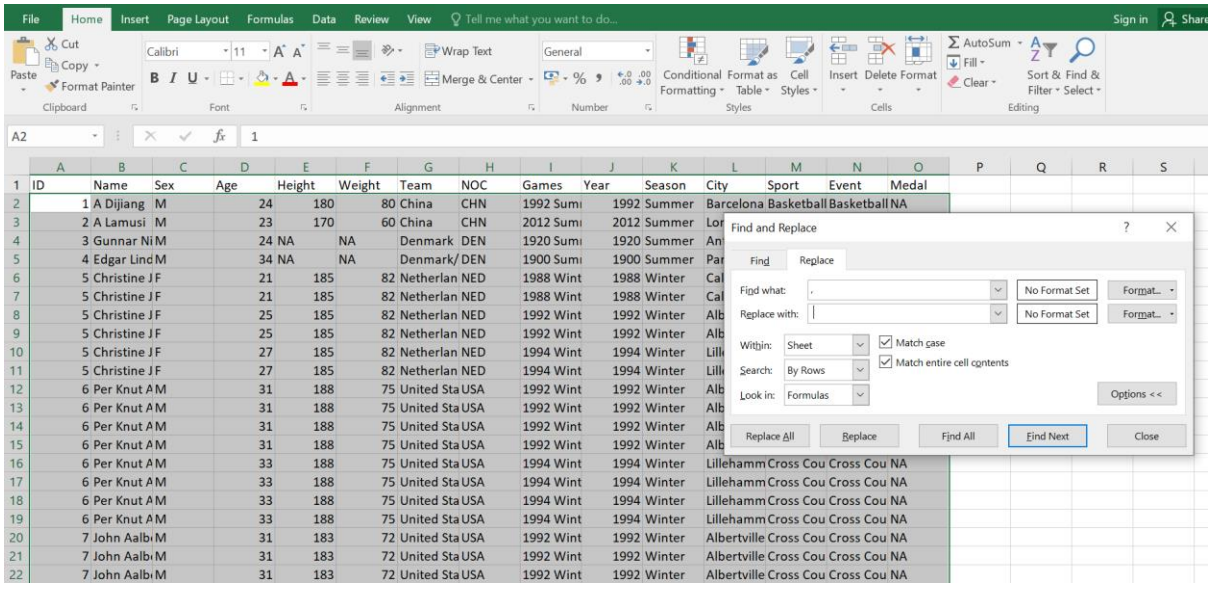
Data Cleaning can be done in Excel or Python

Replace NA from numeric columns to 0.

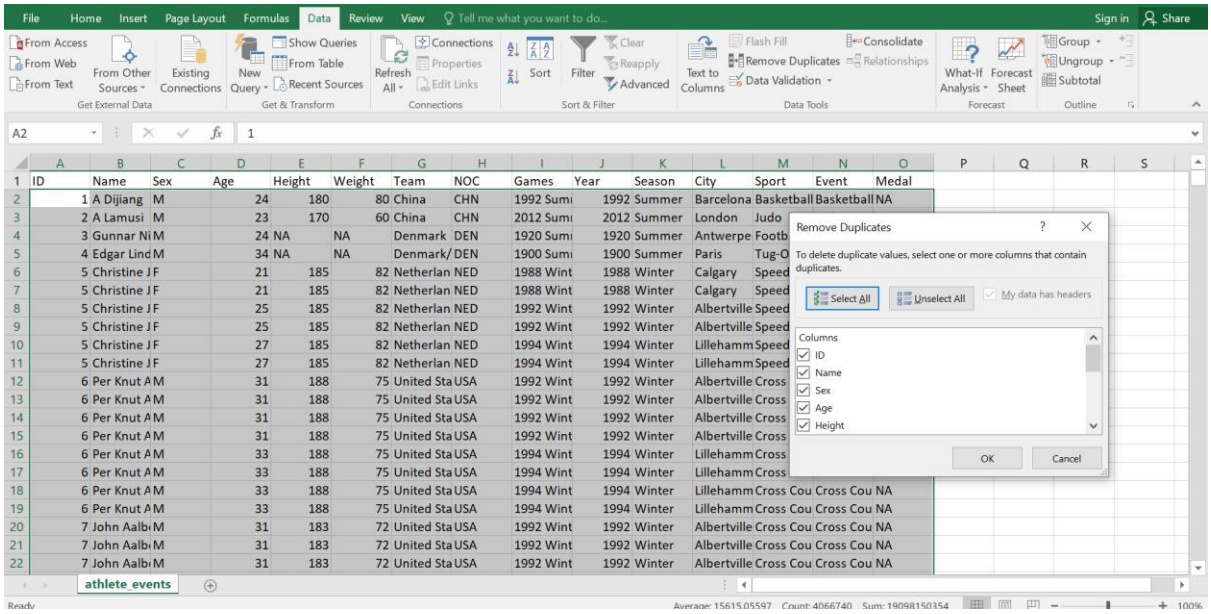


As we will be importing the dataset in csv format, commas inside entities can cause problems.

Therefore, replace commas with space.



Eliminate all the duplicate values.



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

Microsoft Excel

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

OK