Olympic dataset

In [1]:

```
import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings('ignore')
```

Display all columns

In [2]:

```
pd.set_option('display.max_columns', None)
```

In [3]:

```
data = pd.read_csv('./athlete_events.csv')
data
```

Out[3]:

	ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year
0	1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992
1	2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012
2	3	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920
3	4	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900
4	5	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988
...	...	...	...	...	...	...	...	...	...	...
271111	135569	Andrzej ya	M	29.0	179.0	89.0	Poland-1	POL	1976 Winter	1976
271112	135570	Piotr ya	M	27.0	176.0	59.0	Poland	POL	2014 Winter	2014
271113	135570	Piotr ya	M	27.0	176.0	59.0	Poland	POL	2014 Winter	2014
271114	135571	Tomasz Ireneusz ya	M	30.0	185.0	96.0	Poland	POL	1998 Winter	1998
271115	135571	Tomasz Ireneusz ya	M	34.0	185.0	96.0	Poland	POL	2002 Winter	2002

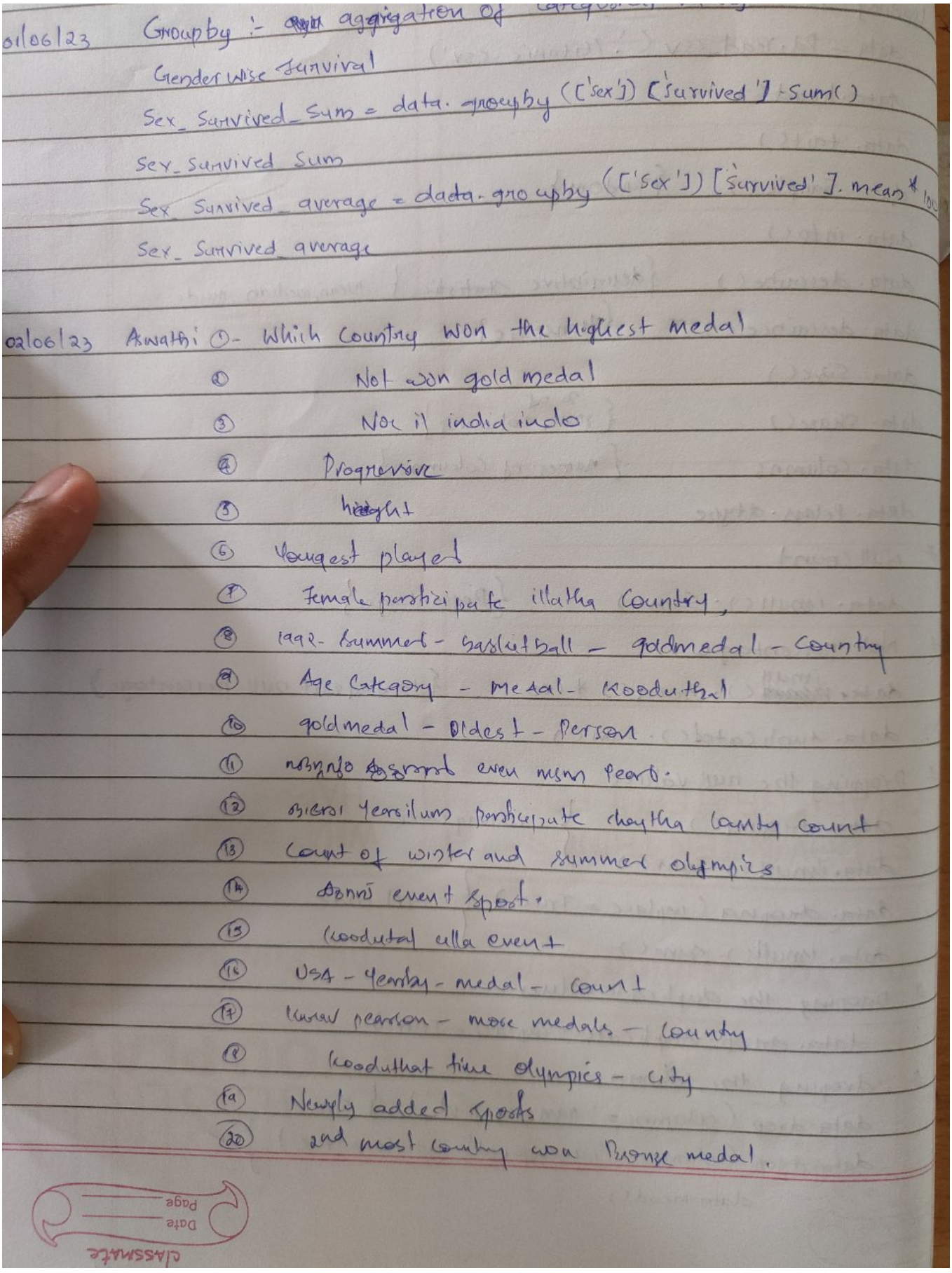
271116 rows × 15 columns

Droping the unwanted columns

In [4]:

```
data.drop(columns=['ID','Games'], inplace=True)
```

Questions and answers



1. Which country has won the most number of medals

2. Countries with no gold medal at all

3. Is India included in the above?

4. Progressive report of medals by countries

5. Athlete with most height

6. Youngest player

7. Countries without any female players

8. Gold medalist in basketball event, 1992 summer olympics

9. Age category wise medal count

10. Oldest person who won gold medal

11. Year in which most number of events

12. Year wise count of participant countries

13. count of winter and summer olympics

14. Sport in which there is less events

15. Most occuring event

16. USA's - Yearly medal count

17. Country with more medals and less number of players

18. City with most number of olympics held

19. Which are the newly added sports

20. Second most country won bronze medal

Answers

Feature engineering the Medal column

In [5]:

```
data.Medal.unique()
```

Out[5]:

```
array([nan, 'Gold', 'Bronze', 'Silver'], dtype=object)
```

Lets create Seperate columns using if else statements for number of medals

In [6]:

```
data['Gold'] = [1 if medal == 'Gold' else 0 for medal in data['Medal']]
data['Silver'] = [1 if medal == 'Silver' else 0 for medal in data['Medal']]
data['Bronze'] = [1 if medal =='Bronze' else 0 for medal in data['Medal']]
```

In [7]:

```
data.head()
```

Out[7]:

	Name	Sex	Age	Height	Weight	Team	NOC	Year	Season	City	
0	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992	Summer	Barcelona	Ba
1	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012	Summer	London	
2	Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920	Summer	Antwerpen	
3	Edgar Lindenau Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900	Summer	Paris	
4	Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988	Winter	Calgary	

1. Which country won most medals

In [8]:

```
country_wise_medals = data.groupby(['Team'])['Gold', 'Silver', 'Bronze'].sum()
country_wise_medals
```

Out[8]:

Team	Gold	Silver	Bronze
30. Februar	0	0	0
A North American Team	0	0	4
Acipactli	0	0	0
Acturus	0	0	0
Afghanistan	0	0	2
...	...	...	...
Zambia	0	1	1
Zefyros	0	0	0
Zimbabwe	17	4	1
Zut	0	3	0
rn-2	0	0	0

1184 rows × 3 columns

Lets add a new column to the groupby object

In [9]:

```
country_wise_medals['Total'] = country_wise_medals['Gold']+country_wise_medals['Silver']+
```

In [10]:

```
country_wise_medals
```

Out[10]:

Team	Gold	Silver	Bronze	Total
30. Februar	0	0	0	0
A North American Team	0	0	4	4
Acipactli	0	0	0	0
Acturus	0	0	0	0
Afghanistan	0	0	2	2
...	...	...	...	...
Zambia	0	1	1	2
Zefyros	0	0	0	0
Zimbabwe	17	4	1	22
Zut	0	3	0	3
rn-2	0	0	0	0

1184 rows × 4 columns

In [11]:

```
country_wise_medals.sort_values(by="Total", ascending=False)
```

Out[11]:

Team	Gold	Silver	Bronze	Total
United States	2474	1512	1233	5219
Soviet Union	1058	716	677	2451
Germany	679	627	678	1984
Great Britain	519	582	572	1673
France	455	518	577	1550
...	...	...	...	...
Ireland-1	0	0	0	0
Israel-1	0	0	0	0
Israel-2	0	0	0	0
Italy-3	0	0	0	0
rn-2	0	0	0	0

1184 rows × 4 columns

In [12]:

```
country_wise_medals.nlargest(1, 'Total')
```

Out[12]:

Team	Gold	Silver	Bronze	Total
United States	2474	1512	1233	5219

2. Country with no gold medal at all

In [13]:

```
No_gold_medal_at_all = country_wise_medals[country_wise_medals['Gold']==0]
```

In [14]:

```
No_gold_medal_at_all
```

Team	Gold	Silver	Bronze	Total
30. Februar	0	0	0	0
A North American Team	0	0	4	4
Acipactli	0	0	0	0
Acturus	0	0	0	0
Afghanistan	0	0	2	2
...	...	...	...	...
Yugoslavia-2	0	0	0	0
Zambia	0	1	1	2
Zefyros	0	0	0	0
Zut	0	3	0	3
rn-2	0	0	0	0

In [15]:

```
List_No_gold_medal_at_all=list(No_gold_medal_at_all.index)
```

In [16]:

```
print("Number of countries with no Gold medals at all :", len(List_No_gold_medal_at_all))
```

```
Number of countries with no Gold medals at all : 942
```

3. Is India included in the above list?

In [17]:

```
Zero_medals_countries_with_India = data.query("Gold == 0 & Silver == 0 & Bronze == 0 & NO
```

In [18]:

```
Zero_medals_countries_with_India
```

Out[18]:

	Name	Sex	Age	Height	Weight	Team	NOC	Year	Season	City	
505	S. Abdul Hamid	M	NaN	NaN	NaN	India	IND	1928	Summer	Amsterdam	At
506	S. Abdul Hamid	M	NaN	NaN	NaN	India	IND	1928	Summer	Amsterdam	At
895	Shiny Kurisingal Abraham-Wilson	F	19.0	167.0	53.0	India	IND	1984	Summer	Los Angeles	At
896	Shiny Kurisingal Abraham-Wilson	F	19.0	167.0	53.0	India	IND	1984	Summer	Los Angeles	At
897	Shiny Kurisingal Abraham-Wilson	F	23.0	167.0	53.0	India	IND	1988	Summer	Seoul	At
...	...	...	...	...	...	...	...	...	...	...	
264138	Mohammad Anas Yahiya	M	21.0	177.0	69.0	India	IND	2016	Summer	Rio de Janeiro	At
264139	Mohammad Anas Yahiya	M	21.0	177.0	69.0	India	IND	2016	Summer	Rio de Janeiro	At
265876	Thyadathuvilla Chandrapillai "T. C." Yohannan	M	29.0	174.0	62.0	India	IND	1976	Summer	Montreal	At
270912	Geeta Zutshi	F	23.0	167.0	51.0	India	IND	1980	Summer	Moskva	At
270913	Geeta Zutshi	F	27.0	167.0	51.0	India	IND	1984	Summer	Los Angeles	At

1211 rows × 16 columns

Sadly India is also in the list.

4. Progressive reoprt of medals by country

In [19]:

```
Cross_tab01= pd.crosstab(index=data.Team, columns=data.Year, values=data.Gold, aggfunc='s
```

In [20]:

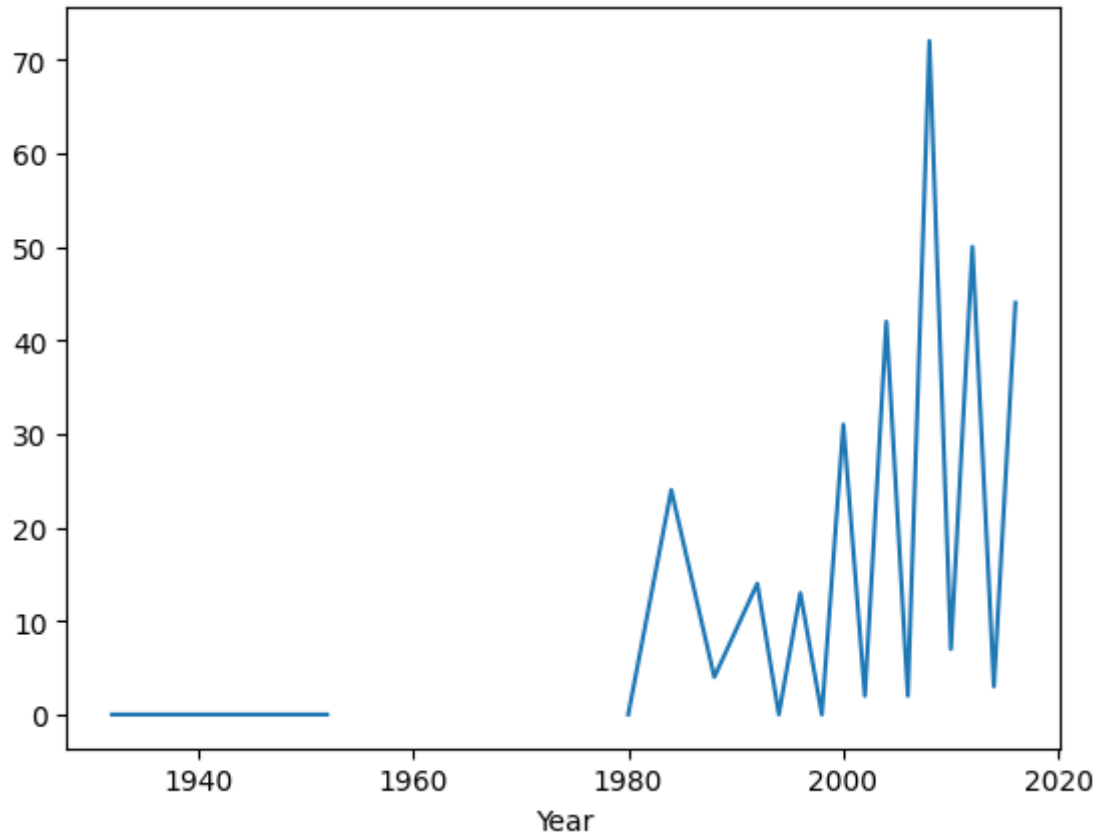
```
Cross_tab01.loc['China']
```

Out[20]:

```
Year
1896     NaN
1900     NaN
1904     NaN
1906     NaN
1908     NaN
1912     NaN
1920     NaN
1924     NaN
1928     NaN
1932     0.0
1936     0.0
1948     0.0
1952     0.0
1956     NaN
1960     NaN
1964     NaN
1968     NaN
1972     NaN
1976     NaN
1980     0.0
1984    24.0
1988     4.0
1992    14.0
1994     0.0
1996    13.0
1998     0.0
2000    31.0
2002     2.0
2004    42.0
2006     2.0
2008    72.0
2010     7.0
2012    50.0
2014     3.0
2016    44.0
Name: China, dtype: float64
```

In [21]:

```
Cross_tab01.loc['China'].plot();
```



5. Athlete with most height

In [22]:

```
data.nlargest(1, 'Height')
```

Out[22]:

	Name	Sex	Age	Height	Weight	Team	NOC	Year	Season	City	Sport	
265040	Yao Ming	M	20.0	226.0	141.0	China	CHN	2000	Summer	Sydney	Basketball	Ba Ba

Yao Ming from China is the tallest player in the dataset

6. Youngest player

In [23]:

```
data.nsmallest(1, 'Age')
```

Out[23]:

	Name	Sex	Age	Height	Weight	Team	NOC	Year	Season	City	
142882	Dimitrios Loundras	M	10.0	NaN	NaN	Ethnikos Gymnastikos Syllogos	GRE	1896	Summer	Athina	Gymr

Dimitrios Loundras of age 10 is the youngest player in the dataset.

7. Countries without female representatives

In [24]:

```
Gender_count = pd.crosstab(data.Sex, data.NOC)
```

In [25]:

```
Gender_count
```

Out[25]:

NOC	AFG	AHO	ALB	ALG	AND	ANG	ANT	ANZ	ARG	ARM	ARU	ASA	AUS	AUT	A
Sex															
F	5	12	27	94	42	129	37	2	643	38	16	7	2749	1100	
M	121	67	43	457	127	138	96	84	2654	183	26	30	4889	4041	

In [26]:

```
Gender_count_01 = Gender_count.unstack().unstack()
```

In [27]:

```
Gender_count_01
```

Out[27]:

Sex	F	M
NOC		
AFG	5	121
AHO	12	67
ALB	27	43
ALG	94	457
AND	42	127
...	...	...
YEM	4	28
YMD	0	5
YUG	378	2205
ZAM	19	164
ZIM	113	198

230 rows × 2 columns

In [28]:

```
Countries_without_female_players = Gender_count_01[Gender_count_01['F'] == 0]
```

In [29]:

```
Countries_without_female_players
```

Out[29]:

Sex	F	M
NOC		
CRT	0	11
NBO	0	2
NFL	0	1
UAR	0	123
UNK	0	2
WIF	0	20
YAR	0	11
YMD	0	5

8. Gold medalist in basketball event, 1992 summer olympics

In [30]:

```
data.query("Year == 1992 & Sport == 'Basketball' & Medal == 'Gold' & Sex == 'M'")
```

Out[30]:

	Name	Sex	Age	Height	Weight	Team	NOC	Year	Season	City	S
14656	Charles Wade Barkley	M	29.0	198.0	114.0	United States	USA	1992	Summer	Barcelona	Baske
22150	Larry Joe Bird	M	35.0	205.0	100.0	United States	USA	1992	Summer	Barcelona	Baske
58646	Clyde Austin Drexler	M	30.0	200.0	101.0	United States	USA	1992	Summer	Barcelona	Baske
65566	Patrick Aloysius Ewing	M	29.0	213.0	109.0	United States	USA	1992	Summer	Barcelona	Baske
109625	Earvin "Magic" Johnson, Jr.	M	32.0	205.0	100.0	United States	USA	1992	Summer	Barcelona	Baske
110517	Michael Jeffrey Jordan	M	29.0	198.0	90.0	United States	USA	1992	Summer	Barcelona	Baske
130713	Christian Donald Laettner	M	22.0	211.0	107.0	United States	USA	1992	Summer	Barcelona	Baske
147819	Karl Malone	M	29.0	205.0	116.0	United States	USA	1992	Summer	Barcelona	Baske
165616	Christopher Paul "Chris" Mullin	M	28.0	200.0	98.0	United States	USA	1992	Summer	Barcelona	Baske
189367	Scottie Maurice Pippen	M	26.0	200.0	102.0	United States	USA	1992	Summer	Barcelona	Baske
201956	David Maurice Robinson	M	26.0	216.0	107.0	United States	USA	1992	Summer	Barcelona	Baske
229678	John Houston Stockton	M	30.0	185.0	79.0	United States	USA	1992	Summer	Barcelona	Baske

USA is the winner

9. Age category wise medal count

In [31]:

```
Age_category_wise_medal_count = data.groupby(['Medal'])['Age'].value_counts()
```

In [32]:

```
Age_category_wise_medal_count.unstack().unstack().unstack()
```

Out[32]:

Medal	Bronze	Gold	Silver
Age			
10.0	1.0	NaN	NaN
11.0	NaN	NaN	1.0
12.0	3.0	NaN	3.0
13.0	2.0	7.0	7.0
14.0	18.0	27.0	30.0
...	...	...	...
68.0	NaN	NaN	2.0
69.0	1.0	NaN	1.0
71.0	1.0	NaN	1.0
72.0	1.0	NaN	1.0
73.0	NaN	NaN	1.0

61 rows × 3 columns

10. Oldest person who won gold medal

In [33]:

```
Gold = data.query("Medal == 'Gold'")
```

In [34]:

```
Gold_oldest = Gold[Gold['Age'] == Gold.Age.max()]
```

In [35]:

```
Gold_oldest
```

Out[35]:

	Name	Sex	Age	Height	Weight	Team	NOC	Year	Season	City	Spo
105199	Charles Jacobus	M	64.0	NaN	NaN	United States	USA	1904	Summer	St. Louis	Roqu
233390	Oscar Gomer Swahn	M	64.0	NaN	NaN	Sweden	SWE	1912	Summer	Stockholm	Shootir

11. Year in which most number of events

In [36]:

```
Most_events_year = pd.crosstab(data.Year, data.Event ,margins=True)
```

In [37]:

```
Most_events_year.sort_values(by='All', ascending=False)
```

Out[37]:

Event	Aeronautics Mixed Aeronautics	Alpine Skiing Men's Combined	Alpine Skiing Men's Downhill	Alpine Skiing Men's Giant Slalom	Alpine Skiing Men's Slalom	Alpine Skiing Men's Super G	Alpine Skiing Women's Combined	Alpine Skiing Women's Downhill	Al Sk Wom G Sla
Year									
All	1	569	1164	1535	1593	570	361	708	
1992	0	66	55	131	119	118	40	30	
1988	0	56	51	117	109	94	39	35	
2000	0	0	0	0	0	0	0	0	
1996	0	0	0	0	0	0	0	0	
2016	0	0	0	0	0	0	0	0	
2008	0	0	0	0	0	0	0	0	
2004	0	0	0	0	0	0	0	0	
2012	0	0	0	0	0	0	0	0	
1972	0	0	55	73	72	0	0	41	
1984	0	0	61	108	101	0	0	32	
1976	0	0	74	97	94	0	0	38	
1968	0	0	86	101	100	0	0	39	
1964	0	0	84	96	96	0	0	43	
1952	0	0	82	83	86	0	0	42	
1960	0	0	63	65	63	0	0	42	
1980	0	0	47	78	79	0	0	28	
1948	0	79	111	0	77	0	28	37	
1936	1	66	0	0	0	0	37	0	
1956	0	0	75	95	89	0	0	47	
1924	0	0	0	0	0	0	0	0	
1928	0	0	0	0	0	0	0	0	
2014	0	50	49	107	115	63	39	41	
2010	0	52	64	101	101	63	34	44	
2006	0	59	55	82	93	63	43	44	
1920	0	0	0	0	0	0	0	0	
2002	0	47	54	78	77	55	31	38	
1912	0	0	0	0	0	0	0	0	
1998	0	38	43	62	65	45	29	39	

Event	Aeronautics Mixed Aeronautics	Alpine Skiing Men's Combined	Alpine Skiing Men's Downhill	Alpine Skiing Men's Giant Slalom	Alpine Skiing Men's Slalom	Alpine Skiing Men's Super G	Alpine Skiing Women's Combined	Alpine Skiing Women's Downhill	Al Sk Wom G Sla
Year									
1932	0	0	0	0	0	0	0	0	
1994	0	56	55	61	57	69	41	48	
1908	0	0	0	0	0	0	0	0	
1900	0	0	0	0	0	0	0	0	
1906	0	0	0	0	0	0	0	0	
1904	0	0	0	0	0	0	0	0	
1896	0	0	0	0	0	0	0	0	

~~Year 1992 is having the most number of events~~

12. Year wise count of participant countries

In [38]:

```
Yearly_participating_country_count =pd.crosstab( data.Year,data.NOC, margins=True)
Yearly_participating_country_count
```

Out[38]:

NOC	AFG	AHO	ALB	ALG	AND	ANG	ANT	ANZ	ARG	ARM	ARU	ASA	AUS	AUT	A
Year															
1896	0	0	0	0	0	0	0	0	0	0	0	0	5	8	
1900	0	0	0	0	0	0	0	0	1	0	0	0	6	26	
1904	0	0	0	0	0	0	0	0	0	0	0	0	6	7	
1906	0	0	0	0	0	0	0	0	0	0	0	0	10	100	
1908	0	0	0	0	0	0	0	48	1	0	0	0	0	13	
1912	0	0	0	0	0	0	0	38	0	0	0	0	0	132	
1920	0	0	0	0	0	0	0	0	1	0	0	0	29	0	
1924	0	0	0	0	0	0	0	0	109	0	0	0	61	60	
1928	0	0	0	0	0	0	0	0	108	0	0	0	29	162	
1932	0	0	0	0	0	0	0	0	47	0	0	0	20	45	
1936	16	0	0	0	0	0	0	0	68	0	0	0	42	435	
1948	25	0	0	0	0	0	0	0	315	0	0	0	103	323	
1952	0	11	0	0	0	0	0	0	195	0	0	0	159	312	
1956	12	0	0	0	0	0	0	0	44	0	0	0	429	144	
1960	16	5	0	0	0	0	0	0	130	0	0	0	318	240	
1964	8	4	0	7	0	0	0	0	170	0	0	0	406	198	
1968	5	5	0	9	0	0	0	0	144	0	0	0	231	171	
1972	8	2	5	7	0	0	0	0	129	0	0	0	278	211	
1976	0	4	0	0	14	0	17	0	122	0	0	0	293	189	
1980	11	0	0	63	11	17	0	0	26	0	0	0	209	158	
1984	0	13	0	33	8	0	27	0	152	0	0	0	357	235	
1988	5	7	0	45	13	33	25	0	179	0	12	6	375	249	
1992	0	6	9	44	22	33	18	0	157	0	5	3	431	248	
1994	0	0	0	0	13	0	0	0	20	2	0	2	47	130	
1996	2	7	9	47	8	30	19	0	221	38	3	7	550	92	
1998	0	0	0	0	7	0	0	0	6	7	0	0	34	154	
2000	0	8	5	51	5	30	4	0	165	26	5	4	788	114	
2002	0	0	0	0	5	0	0	0	26	17	0	0	41	148	
2004	5	3	7	71	6	30	5	0	179	19	4	3	601	92	
2006	0	0	3	3	10	0	0	0	17	9	0	0	45	132	
2008	4	4	12	57	5	33	5	0	145	26	2	4	570	93	
2010	0	0	2	1	20	0	0	0	20	6	0	0	49	132	
2012	6	0	10	39	6	35	4	0	148	31	4	4	514	90	
2014	0	0	2	0	12	0	0	0	20	6	0	0	84	213	
2016	3	0	6	74	4	26	9	0	232	34	7	4	518	85	

NOC	AFG	AHO	ALB	ALG	AND	ANG	ANT	ANZ	ARG	ARM	ARU	ASA	AUS	AUT	A
Year															
All	128	79	70	351	169	267	133	86	5297	221	42	37	7638	5141	

13. count of winter and summer olympics

In [39]:

```
summer = data[data['Season']=='Summer']
winter = data[data['Season']!='Summer']
```

In [40]:

```
len(summer), len(winter)
```

Out[40]:

```
(222552, 48564)
```

14. Sport in which there is less events

In [41]:

```
Sport_with_less_events = pd.crosstab(data.Sport, data.Event, margins=True)
```

In [42]:

```
Sport_with_less_events.sort_values(by='All')
```

Out[42]:

Event	Aeronautics Mixed Aeronautics	Alpine Skiing Men's Combined	Alpine Skiing Men's Downhill	Alpine Skiing Men's Giant Slalom	Alpine Skiing Men's Slalom	Alpine Skiing Men's Super G	Alpine Skiing Women's Combined	Alpine Skiing Women's Downhill
Sport								
Aeronautics	1	0	0	0	0	0	0	0
Basque Pelota	0	0	0	0	0	0	0	0
Roque	0	0	0	0	0	0	0	0
Jeu De Paume	0	0	0	0	0	0	0	0
Racquets	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...
Shooting	0	0	0	0	0	0	0	0
Swimming	0	0	0	0	0	0	0	0
Gymnastics	0	0	0	0	0	0	0	0
Athletics	0	0	0	0	0	0	0	0
All	1	569	1164	1535	1593	570	361	708

67 rows × 766 columns

15. Most occuring event

In [43]:

```
data.Event.value_counts()
```

Out[43]:

```
Football Men's Football                                5733
Ice Hockey Men's Ice Hockey                            4762
Hockey Men's Hockey                                    3958
Water Polo Men's Water Polo                            3358
Basketball Men's Basketball                            3280
                                                       ...
Croquet Mixed Doubles                                     2
Archery Men's Target Archery, 50 metres, Individual       2
Archery Men's Target Archery, 33 metres, Individual       2
Archery Men's Target Archery, 28 metres, Individual       2
Aeronautics Mixed Aeronautics                             1
Name: Event, Length: 765, dtype: int64
```

16. USA's - Yearly medal count

In [44]:

```
USA = data[data['Team'] =='United States']
```

In [45]:

```
pd.crosstab(USA.Year, USA.Medal)
```

Out[45]:

Medal	Bronze	Gold	Silver
Year			
1896	2	11	7
1900	13	18	14
1904	66	65	68
1906	6	12	6
1908	15	34	16
1912	36	46	25
1920	38	111	45
1924	50	92	44
1928	19	48	25
1932	62	71	57
1936	28	51	29
1948	33	82	34
1952	29	77	38
1956	21	61	61
1960	21	97	27
1964	32	96	37
1968	36	100	37
1972	46	72	77
1976	39	73	61
1980	2	24	4
1984	50	190	119
1988	56	89	67
1992	87	92	57
1994	5	6	8
1996	52	157	46
1998	3	25	2
2000	51	128	61
2002	9	9	52
2004	69	115	75
2006	32	9	7
2008	78	121	110
2010	20	8	61
2012	44	139	55
2014	16	8	28
2016	67	137	52

17. Country with more medals and less number of players

In [46]:

```
team_and_medal_count = pd.crosstab(data.Team, data.Medal)
```

In [47]:

```
team_and_medal_count['Total'] = team_and_medal_count['Bronze']+team_and_medal_count['Silv
```

In [48]:

```
team_and_medal_count
```

Out[48]:

Medal	Bronze	Gold	Silver	Total
Team				
A North American Team	4	0	0	4
Afghanistan	2	0	0	2
Algeria	8	5	4	17
Ali-Baba II	5	0	0	5
Amateur Athletic Association	0	5	0	5
...	...	...	...	...
Winnipeg Shamrocks-1	0	12	0	12
Yugoslavia	93	130	167	390
Zambia	1	0	1	2
Zimbabwe	1	17	4	22
Zut	0	0	3	3

498 rows × 4 columns

In [49]:

```
len(team_and_medal_count)
```

Out[49]:

498

Number of players in a Team

In [50]:

```
top_countries = data.Team.value_counts()
top_countries
```

Out[50]:

```
United States    17847
France           11988
Great Britain    11404
Italy            10260
Germany           9326
                 ...
Briar                1
Hannover             1
Nan-2                1
Brentina             1
Digby                1
Name: Team, Length: 1184, dtype: int64
```

In [51]:

```
team_and_medal_count
```

Out[51]:

Medal Team	Bronze	Gold	Silver	Total
A North American Team	4	0	0	4
Afghanistan	2	0	0	2
Algeria	8	5	4	17
Ali-Baba II	5	0	0	5
Amateur Athletic Association	0	5	0	5
...	...	...	...	...
Winnipeg Shamrocks-1	0	12	0	12
Yugoslavia	93	130	167	390
Zambia	1	0	1	2
Zimbabwe	1	17	4	22
Zut	0	0	3	3

498 rows × 4 columns

In [52]:

```
team_and_medal_count['Number_of_Players'] = top_countries
```

In [53]:

```
team_and_medal_count
```

Out[53]:

Medal Team	Bronze	Gold	Silver	Total	Number_of_Players
A North American Team	4	0	0	4	4
Afghanistan	2	0	0	2	126
Algeria	8	5	4	17	551
Ali-Baba II	5	0	0	5	7
Amateur Athletic Association	0	5	0	5	5
...	...	...	...	...	...
Winnipeg Shamrocks-1	0	12	0	12	12
Yugoslavia	93	130	167	390	2558
Zambia	1	0	1	2	183

18. City with most number of olymipcs held

In [54]:

```
data.groupby(['City'])['Year'].value_counts()
```

Out[54]:

```
City                    Year
Albertville             1992     3436
Amsterdam               1928     4992
Antwerpen               1920     4292
Athina                  2004    13443
                        1906     1733
                        1896      380
Atlanta                 1996    13780
Barcelona               1992    12977
Beijing                 2008    13602
Berlin                  1936     6506
Calgary                 1988     2639
Chamonix                1924      460
Cortina d'Ampezzo       1956     1307
Garmisch-Partenkirchen  1936      895
Grenoble                1968     1891
Helsinki                1952     8270
Innsbruck               1976     1861
                        1964     1778
Lake Placid             1980     1746
                        1932      352
Lillehammer             1994     3160
London                  2012    12920
                        1948     6405
                        1908     3101
Los Angeles             1984     9454
                        1932     2969
Melbourne               1956     4829
Mexico City             1968     8588
Montreal                1976     8641
Moskva                  1980     7191
Munich                  1972    10304
Nagano                  1998     3605
Oslo                    1952     1088
Paris                   1924     5233
                        1900     1936
Rio de Janeiro          2016    13688
Roma                    1960     8119
Salt Lake City          2002     4109
Sankt Moritz            1948     1075
                        1928      582
Sapporo                 1972     1655
Sarajevo                1984     2134
Seoul                   1988    12037
Sochi                   2014     4891
Squaw Valley            1960     1116
St. Louis               1904     1301
Stockholm               1912     4040
                        1956      298
Sydney                  2000    13821
Tokyo                   1964     7702
Torino                  2006     4382
Vancouver               2010     4402
Name: Year, dtype: int64
```

19. Which are the newly added sports

Lets make 2 data frames : before 2016 and 2016. Then extract the events from it in lists, then compare both the lists.

In [55]:

```
data2016 = data[data['Year'] == 2016]
data_not_2016 = data[data['Year'] != 2016]
```

In [56]:

```
list1 = list(data2016.Event)
list2 = list(data_not_2016.Event)
```

In [57]:

```
New_events = list(set(list1).difference(list2))
print(New_events)
```

```
["Wrestling Women's Light-Heavyweight, Freestyle", "Wrestling Women's Feat
herweight, Freestyle", "Rugby Sevens Women's Rugby Sevens", "Sailing Wome
n's Skiff", "Rugby Sevens Men's Rugby Sevens"]
```

20. Second most country won bronze medal

In [58]:

```
Bronze = data.groupby(['Team'])['Bronze'].sum()
```

In [59]:

```
Bronze.sort_values(ascending=False)
```

Out[59]:

```
Team
United States    1233
Germany           678
Soviet Union      677
France            577
Great Britain     572
                 ...
Hb-20               0
Hb-24               0
Heidelberg          0
Heira II            0
rn-2                0
Name: Bronze, Length: 1184, dtype: int64
```

Germany is the contry having second place in the case of Bronze medals.