

# Olympics Competition

1896 - 2016

# Olympics Background

The Olympics is a tradition that has been there since the 18<sup>th</sup> century, the competition brings together best athletes in various events to compete for the top prize money, the same is not limited to just specific countries rather ahead of the competition there are trials that are organized in the every country and the athletes that meet specified threshold are admitted for participation:

- a. Free of banned substances
- b. Country must be a member on NOC (Olympics organizing body)
- c. Athletes must meet the set timing to qualify for the competition

<b>Id</b>	<b>Name</b>	<b>Gender</b>	<b>Age</b>	<b>Height</b>	<b>Weight</b>	<b>Team</b>	<b>Noc</b>	<b>Games</b>	<b>Year</b>	<b>Season</b>	<b>City</b>	<b>Sport</b>	<b>Event</b>	<b>Medal</b>
1	A Dijiang	M	24	180	80	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NA
2	A Lamusi	M	23	170	60	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NA
3	Gunnar Nielsen Aaby	M	24	NA	NA	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	NA
4	Edgar Lindenau Aabye	M	34	NA	NA	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
5	Christine Jacoba Aaftink	F	21	185	82	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NA

# Data Summary

There are various columns in the data and the same has been constrained in a specific format to ease the analysis process, 3 columns have numerical values but have instead been imported as text, this is to take care of the NA values in them

## Summary

- columns – 15
- Rows – 271,116
- Missing values Age column – 9474 – 3.5% of total rows
- Missing values Height column – 60171 – 22% of total rows
- Missing values weight column – 62575 – 23% of total rows

Three columns have missing values, there are several ways of looking into the same and having them sorted out.

- The first option is to work with the data without interfering with NAs but the same will not allow us to look at interesting analysis such as mean and median
- The second option is to clean the data by age these values based on present information will either be added or subtracted to fill the NA values, for the case of Weight and Height the means of the columns can be taken into account

# Data Summary

Data Output

Messages

Notifications

	age text	height text	weight text
1	NA	NA	NA
2	NA	NA	NA
3	NA	NA	NA
4	NA	156	48
5	NA	NA	NA

Sample of the missing values in the three categories










- ✓ Since much interest is on understanding how the countries and individuals fared in terms of the medals we shall at this time disregard the missing values in Age, Height and Weight columns

# Analysis










Data Output Messages Notifications					
<div><div><div>≡+</div><div></div><div><div>▼</div><div></div><div><div>▼</div><div></div><div></div><div><div>↓</div><div></div></div></div></div></div></div>					
	total_number_of_athletes bigint	athletes_participations bigint	years_of_competition bigint	host_cities bigint	unique_sports bigint
1	271116	134732	35	42	66

- ✓ The competition has been in existence for the last 35 years, based on the data the final year was 2016 meaning the data is missing out on the most current competition, the event has been hosted in at least **42** cities in the world (we shall dive into these cities to know the ones that have hosted the most)
- ✓ There are at least **66** different competitions that have been held in the Olympics competition, with **134,732** athletes taking part in the competition.
- ✓ Overall the data has **271,116** rows but of these only **134,732** rows are unique meaning there is repetition of names, but the same can be explained by the fact that athletes participated in several competitions and also that they may as well have made appearance in more than one Olympic, it is hosted at least after every four years

# Analysis

Data Output		Messages	Notifications
			
			
			
	gender text	count bigint	
1	F	33808	
2	M	100979	

Gender distribution, there are more males as compared to the females that have featured in the competition, the competition is always subdivided into male and female categories, it would be interesting to dive deeper into the data and understand why the discrepancies, there was expectation that there would an equal representation of the both genders.

Data Output		Messages	Notifications
			
			
			
	season text	participants bigint	
1	Summer	116122	
2	Winter	18923	

Summer season tend to attract more athletes as compared to winter seasons, this requires further research, the next table will explain the distribution of athletes in both summer and winter

# Analysis

	year text	summer bigint	winter bigint
1	1896	380	[null]
2	1900	1936	[null]
3	1904	1301	[null]
4	1906	1733	[null]
5	1908	3101	[null]
6	1912	4040	[null]
7	1920	4292	[null]
8	1924	5233	460
9	1928	4992	582
10	1932	2969	352
11	1936	6506	895
12	1948	6405	1075
13	1952	8270	1088
14	1956	5127	1307
15	1960	8119	1116
16	1964	7702	1778
17	1968	8588	1891

	year text	summer bigint	winter bigint
18	1972	10304	1655
19	1976	8641	1861
20	1980	7191	1746
21	1984	9454	2134
22	1988	12037	2639
23	1992	12977	3436
24	1994	[null]	3160
25	1996	13780	[null]
26	1998	[null]	3605
27	2000	13821	[null]
28	2002	[null]	4109
29	2004	13443	[null]
30	2006	[null]	4382
31	2008	13602	[null]
32	2010	[null]	4402
33	2012	12920	[null]
34	2014	[null]	4891
35	2016	13688	[null]

Judging by the pivot table most of the events take place in the summer as compared to winter, and event that take place in the winter tend to record low number of participants

## Winter and Summer Olympics

Both are organized by the International Olympics committee, The summer Olympics is much bigger with 206 participating countries as per count in the 2021 competition held in Tokyo Japan, the winter on the other hand is smaller scale with 91 participating countries as per count for the competition held in China.

Essence of the 2 events though Olympics in general is held every four years is to ensure that Olympics happen atleast every 2 years

# Analysis

Data Output

Messages

Notifications

	city text	hosting bigint
1	Athina	3
2	London	3
3	Innsbruck	2
4	Lake Placid	2
5	Los Angeles	2
6	Paris	2
7	Sankt Moritz	2
8	Stockholm	2

Host cities, especially those that have hosted the competition more than once, its only Athina(Athens Greece) and London that have had the privilege to host the competition more than twice.

The cities mentioned below hosted only twice, the country is recorded in the brackets

- ✓ Innsbruck(Austria)
- ✓ Lake placid(Canada)
- ✓ Los Angeles(United states)
- ✓ Paris(France)
- ✓ Sankt Moritz(Switzerland)
- ✓ Stockholm(Sweden)

**There have been 42 different host city since the inception of the competition.**



# Analysis

	sport text	count bigint
1	Athletics	22053
2	Swimming	8761
3	Rowing	7684
4	Football	6161
5	Cycling	5819
6	Boxing	5254
7	Wrestling	4987
8	Shooting	4879
9	Sailing	4480
10	Gymnastics	4132
11	Fencing	4118
12	Ice Hockey	3882
13	Hockey	3813
14	Basketball	3412
15	Canoeing	3204
16	Weightlifting	2880
17	Alpine Skiing	2734

	sport text	count bigint
18	Judo	2721
19	Handball	2699
20	Water Polo	2599
21	Volleyball	2502
22	Cross Country Skiing	2400
23	Equestrianism	2345
24	Art Competitions	1814
25	Bobsleigh	1694
26	Speed Skating	1582
27	Figure Skating	1572
28	Diving	1466
29	Tennis	1246
30	Biathlon	1135
31	Archery	1113
32	Ski Jumping	874
33	Modern Pentathlon	864
34	Badminton	811

	sport text	count bigint
35	Luge	771
36	Baseball	761
37	Table Tennis	749
38	Freestyle Skiing	626
39	Nordic Combined	604
40	Rhythmic Gymnastics	567
41	Snowboarding	567
42	Synchronized Swimming	549
43	Taekwondo	470
44	Short Track Speed Skating	443
45	Beach Volleyball	383
46	Softball	367
47	Triathlon	355
48	Curling	346
49	Rugby Sevens	299
50	Golf	218
51	Tug-Of-War	160

	sport text	count bigint
52	Rugby	155
53	Skeleton	146
54	Trampolining	93
55	Polo	87
56	Lacrosse	60
57	Alpinism	25
58	Military Ski Patrol	24
59	Cricket	24
60	Motorboating	14
61	Jeu De Paume	11
62	Croquet	10
63	Racquets	7
64	Roque	4
65	Basque Pelota	2
66	Aeronautics	1

- ✓ Racquets, Roque, Basque, Aeronautics have less than 10 participants for the period of time
- ✓ Top 11 sports with more than 4000 participants are Athletics, Swimming, Rowing, Football, Cycling, Boxing, Wrestling, Shooting, Sailing, Gymnastics, Fencing.

# Analysis

Data Output			Messages	Notifications
	region	count		
	text	bigint		
196	Micronesia	5		
197	Brunei	5		
198	Individual Olympic Athletes	5		
199	Eritrea	5		
200	Montenegro	5		
201	Palau	5		
202	Kiribati	4		
203	Marshall Islands	3		
204	South Sudan	1		
205	Kosovo	1		

Countries that have the least participation since inception of the campaign, interesting to note there is a group of athletes that are registered as individuals.

We have countries that joined NOC late i.e. for the case of south Sudan who gained independence later after separating from Sudan in 2011.

Data Output			Messages	Notifications
	number_of_participation	count_of_countries		
	text	bigint		
1	between_10_14	62		
2	between_5_9	48		
3	between_15_19	34		
4	between_20_24	18		
5	between_25_29	18		
6	between_30_34	14		
7	participation_35	7		
8	between_0_4	4		

Number of times for participation.

- ✓ 62 countries have featured between 10 - 14 times
- ✓ 48 countries have made appearance between 5 – 9 times
- ✓ 7 Countries have fully featured since the competition making 35 appearances

Data Output			Messages	Notifications
	region	count		
	text	bigint		
1	France	35		
2	USA	35		
3	UK	35		
4	Australia	35		
5	Italy	35		
6	Greece	35		
7	Switzerland	35		

The countries that have featured every year since inception of the competition, same have cities that have hosted the competition more than once apart from Italy.

# Analysis

	year text	female bigint	male bigint
1	1896	[null]	380
2	1900	33	1903
3	1904	16	1285
4	1906	11	1722
5	1908	47	3054
6	1912	87	3953
7	1920	134	4158
8	1924	261	5432
9	1928	437	5137
10	1932	369	2952
11	1936	549	6852
12	1948	761	6719
13	1952	1682	7676
14	1956	1139	5295
15	1960	1730	7505
16	1964	1752	7728
17	1968	2193	8286

	year text	female bigint	male bigint
18	1972	2608	9351
19	1976	2606	7896
20	1980	2186	6751
21	1984	2983	8605
22	1988	4223	10453
23	1992	5178	11235
24	1994	1105	2055
25	1996	5008	8772
26	1998	1384	2221
27	2000	5431	8390
28	2002	1582	2527
29	2004	5546	7897
30	2006	1757	2625
31	2008	5816	7786
32	2010	1847	2555
33	2012	5815	7105
34	2014	2023	2868
35	2016	6223	7465

The discrepancy on gender is quite clear, up until 1936 the competition record less than 1000 women, this was a huge gap but the same has closed as years have gone by.

# Analysis

	sport text	female bigint	male bigint
1	Aeronautics	[null]	1
2	Alpine Skiing	996	1738
3	Alpinism	1	24
4	Archery	500	613
5	Art Competitions	204	1610
6	Athletics	6527	15526
7	Badminton	412	399
8	Baseball	[null]	761
9	Basketball	932	2480
10	Basque Pelota	[null]	2
11	Beach Volleyball	189	194
12	Biathlon	371	764
13	Bobsleigh	109	1585
14	Boxing	65	5189
15	Canoeing	702	2502
16	Cricket	[null]	24
17	Croquet	3	7

	sport text	female bigint	male bigint
18	Cross Country Skiing	717	1683
19	Curling	160	186
20	Cycling	714	5105
21	Diving	635	831
22	Equestrianism	459	1886
23	Fencing	880	3238
24	Figure Skating	824	748
25	Football	734	5427
26	Freestyle Skiing	267	359
27	Golf	70	148
28	Gymnastics	1499	2633
29	Handball	1027	1672
30	Hockey	995	2819
31	Ice Hockey	498	3384
32	Jeu De Paume	[null]	11
33	Judo	756	1965
34	Lacrosse	[null]	60
35	Luge	228	543

	sport text	female bigint	male bigint
37	Modern Pentathlon	114	750
38	Motorboating	1	13
39	Nordic Combined	[null]	604
40	Polo	[null]	87
41	Racquets	[null]	7
42	Rhythmic Gymnastics	567	[null]
43	Roque	[null]	4
44	Rowing	1483	6201
45	Rugby	[null]	155
46	Rugby Sevens	148	151
47	Sailing	629	3851
48	Shooting	737	4142
49	Short Track Speed Skating	208	235
50	Skeleton	45	101
51	Ski Jumping	30	844
52	Snowboarding	239	328

	sport text	female bigint	male bigint
53	Softball	367	[null]
54	Speed Skating	528	1054
55	Swimming	3618	5143
56	Synchronized Swimming	549	[null]
57	Table Tennis	377	372
58	Taekwondo	229	241
59	Tennis	486	760
60	Trampoline	44	49
61	Triathlon	175	180
62	Tug-Of-War	[null]	160
63	Volleyball	1128	1374
64	Water Polo	337	2262
65	Weightlifting	356	2524
66	Wrestling	222	4765

- ✓ Notable competition such as Baseball, Nordic combined, Rugby had not registered any women hence the low number of women in the competition, other sports such as athletics have huge difference between the genders.
- ✓ Good to note that sports such as Softball, Synchronized swimming, Rhythmic gymnastics have on the other hand recorded more females and than males suggesting the mentioned competitions are female dominated.
- ✓ The number are a pointer incase a sports brand wanted to sponsor the any of the genders equally, this could easily plus other insights could be used in making a decision

# Analysis

Data Output

Messages

Notifications

	<div>sport</div> <div>text</div> <div></div>	<div>medal_count</div> <div>bigint</div> <div></div>
1	Athletics	38624
2	Gymnastics	26707
3	Swimming	23195
4	Shooting	11448
5	Cycling	10859
6	Fencing	10735
7	Rowing	10595
8	Cross Country Skiing	9133
9	Alpine Skiing	8829
10	Wrestling	7154

Medals table for the various disciplines, the same are for those that have over 7,000 medals awarded.

The top 5 include, athletics, Gymnastics, Swimming, Shooting and cycling, this is the same as the order of participation by discipline.

Data Output

Messages

Notifications

region

text

medal\_count

bigint

1

USA

18853

2

Germany

15883

3

France

12758

4

UK

12256

5

Russia

11692

6

Italy

10715

The USA tops the list of medals.

Apart from Germany and Russia the balance of the countries had made appearance since its inception that is 35 times hence explains the reason as to why there have more medals.

# Conclusions

- ✓ Data analysis gives one an opportunity to review through the data and even develop sound insights that can be presented to the management for actioning.
- ✓ Looking at the data on fly there is little that can be made out of it but after running analysis below are some of the questions that cropped up and the data in combination with further research can conclusively help in ensuring balance, sponsorship deals and marginal gains.
  - ✓ From the simple analysis the first thing would be how do we bridge the gap between the genders? This is to ensure equally representation for both genders.
  - ✓ Also, is it important to run the winter campaign? The numbers from the winter campaign are pretty low as compared to the summer campaign, this decision should be made after checking on revenue generated during the campaign against the cost for running.
  - ✓ Are there sports that need to be scraped off? – there are sports that had very low number of participants, is it important for them to continue featuring
  - ✓ Hosting rights, what makes other cities favorable to host? Are there ideas that prospective cities can borrow to help them in the bidding process
  - ✓ Medal standing is critical as well, can we use successfully countries as case studies for those who aspire to win more medals?

**The beauty about analysis is that the it gives insights that can further be researched on before making a final decision, these decisions are meant to help the business pivot in the right direction more so in ensuring there are minimal issues along the way and also maximum profits are generated.**