

EXPLORATORY DATA ANALYSIS ON OLYMPICS HISTORY [1896 – 2016]



By Ben Joan C

INTRODUCTION

The modern Olympic games is a grand international sporting events featuring winter and summer sports competitions in which thousands of athletes and countries participate in variety of sports events. The event is held ever four years period and this exploratory analysis using SQL (structured Query Language),is to derive information about the Olympics games over the years.

The dataset contains information on the games from 1896 – 2016 and is comprised of two files: (a)athlete_events (b)noc_regions.

Athlete_events being the major file used for the analysis is made up of 271,116 rows and 15 columns while the Noc_regions contains three columns of data.

Tool Used for the Analysis: **SQLite Studio**

TABLES



EXPLORING THE DATASET

To analyze and derive information from the dataset, a few questions (source: techtfq.com & some from my own curiosity) will be asked to help know what queries to write. The questions and queries include the following:

1.How many Olympics games have been held between the given date

```
11
12 SELECT COUNT( Distinct games) AS Games
13 FROM athlete_events;
14
15
```

Grid viewForm view

1

Total rows loaded: 1

	Games
1	51

2.List down all Olympics games held so far

```
19
20 SELECT DISTINCT games
21 FROM athlete_events ORDER BY games ;
22
23
```

Grid viewForm view

1

Total rows loaded: 51

	games
1	1896 Summer
2	1900 Summer
3	1904 Summer
4	1906 Summer
5	1908 Summer
6	1912 Summer
7	1920 Summer
8	1924 Summer
9	1924 Winter
10	1928 Summer
11	1928 Winter
12	1932 Summer
13	1932 Winter
.....	

3.How many teams participated in the Olympics

```
25
26 SELECT COUNT( DISTINCT team) AS Team
27 FROM athlete_events;
28
29
30
```

Grid view		Form view							
					1				Total
Team									
1	1184								

5.Which year saw the highest & lowest no of countries participating in Olympics

```
WITH least_country AS (
    SELECT
        games || '-' || COUNT(distinct noc) AS least_country
    FROM athlete_events
    GROUP BY games
    ORDER BY COUNT(distinct noc) LIMIT 1 ),
most_country AS (
    SELECT
        games|| '-'|| COUNT(distinct noc) AS most_country
    FROM athlete_events
    GROUP BY games
    ORDER BY COUNT(distinct noc) DESC LIMIT 1
)
SELECT least_country, most_country
FROM least_country, most_country;
```

Grid view

Form view

✓

✕

⏪

⏩

1

⏴

⏵

🖨

Total rows loaded: 1

least country	most country
1896 Summer-12	2016 Summer-207

4.The total no of nations who participated in each Olympics games

```
35 SELECT games,
36        COUNT(DISTINCT noc) AS countries
37 FROM athlete_events
38 GROUP BY games;
39
40
```









Grid viewForm view

Total rows loaded: 51

	games	countries
1	1896 Summer	12
2	1900 Summer	31
3	1904 Summer	15
4	1906 Summer	21
5	1908 Summer	22
6	1912 Summer	29
7	1920 Summer	29
8	1924 Summer	45
9	1924 Winter	19
10	1928 Summer	46
11	1928 Winter	25
12	1932 Summer	47
13	1932 Winter	17
14	1936 Summer	49
15	1936 Winter	28
16	1948 Summer	50









6. Which Nation has participated in all the Olympics

```
62
63 WITH total_participants AS (
64     SELECT region,
65            COUNT(DISTINCT games) AS total_games,
66            DENSE_RANK() OVER(ORDER BY COUNT(distinct games)DESC) AS ranking
67     FROM athlete_events ae
68     LEFT JOIN noc_regions nr
69           USING(noc)
70     GROUP BY region
71 )
72 SELECT region, total_games
73 FROM total_participants
74 WHERE ranking = 1 ;
75
```

Grid view		Form view	
			
			
Total rows loaded: 4			
region	total games		
1 UK	51		
2 Switzerland	51		
3 Italy	51		
4 France	51		

7. Identify the sports which was played in all summer Olympics

```
80
81 WITH all_summer_game AS (
82     SELECT sport,
83            (SELECT COUNT(distinct games)
84             FROM athlete_events WHERE season = 'Summer') AS summer_games,
85            COUNT(distinct games) AS total_games
86     FROM athlete_events
87     GROUP BY sport
88 )
89 SELECT sport,
90        summer_games,
91        total_games
92 FROM all_summer_game
93 WHERE summer_games = total_games ;
94
```

Grid view		Form view	
			
			
Total rows loaded: 5			
sport	summer games	total games	
1 Athletics	29	29	
2 Cycling	29	29	
3 Fencing	29	29	
4 Gymnastics	29	29	
5 Swimming	29	29	

8.Which sports were played just once in the Olympics

```
99
100 WITH one_olympic_game AS (
101     SELECT sport,
102            COUNT(DISTINCT games) AS total_games,
103            games,
104            DENSE_RANK() OVER(ORDER BY count(distinct games) ) AS ranking
105     FROM athlete_events
106     GROUP BY sport
107 )
108 SELECT sport, total_games, games
109 FROM one_olympic_game
110 WHERE ranking = 1
111 ORDER BY sport;
```

Grid viewForm view

✓

✗

⏪

⏩

1

⏴

⏵

🖨

Total rows loaded: 10

	sport	total games	games
1	Aeronautics	1	1936 Summer
2	Basque Pelota	1	1900 Summer
3	Cricket	1	1900 Summer
4	Croquet	1	1900 Summer
5	Jeu De Paume	1	1908 Summer
6	Military Ski Patrol	1	1924 Winter
7	Motorboating	1	1908 Summer
8	Racquets	1	1908 Summer
9	Roque	1	1904 Summer
10	Rugby Sevens	1	2016 Summer

9.Fetch the total no of sports played in each Olympic games

```
117
118 SELECT games,
119        COUNT(DISTINCT sport) AS no_of_sports
120 FROM athlete_events
121 GROUP BY games ;
122
123
```

Grid viewForm view

✓

✗

⏪

⏩

1

⏴

⏵

🖨

Total rows loaded: 51

	games	no of sports
1	1896 Summer	9
2	1900 Summer	20
3	1904 Summer	18
4	1906 Summer	13
5	1908 Summer	24
6	1912 Summer	17
7	1920 Summer	25
8	1924 Summer	20
9	1924 Winter	10
10	1928 Summer	17
11	1928 Winter	8
12	1932 Summer	18
..	1936 Summer	7

10.How many teams participated in the Olympics

```
127
128 WITH oldest_athlete AS
129     ( SELECT name, sex, CAST(age as int) as age,
130         games, city, sport, medal
131       FROM athlete_events ),
132  ranking AS( SELECT *,
133              DENSE_RANK() OVER( ORDER BY age desc) AS ranking
134              FROM oldest_athlete
135              WHERE medal = 'Gold'
136            )
137 SELECT * FROM ranking
138 WHERE ranking = 1 ;
139
140
```

Grid view Form view

↺

✓

✕

⏪

⏩

1

⏴

⏵

🖨

Total rows loaded: 2

	name	sex	age	games	city	sport	medal	ranking
1	Charles Jacobus	M	64	1904 Summer	St. Louis	Roque	Gold	1
2	Oscar Gomer Swahn	M	64	1912 Summer	Stockholm	Shooting	Gold	1

Oldest athlete to participate in the Olympics

```
142
143 SELECT name, age, sport, games, medal
144 FROM athlete_events
145 WHERE age <> 'NA'
146 ORDER BY age DESC LIMIT 1;
147
148
149
```

Grid view Form view

↺

✓

✕

⏪

⏩

1

⏴

⏵

🖨

Total rows loaded: 1

	name	age	sport	games	medal
1	John Quincy Adams Ward	97	Art Competitions	1928 Summer	NA

11.The total no of nations who participated in each Olympics games

```
153 WITH Females AS (
154     SELECT COUNT(sex) AS Females FROM athlete_events
155     WHERE Sex = 'F'),
156  Males AS (SELECT COUNT(Sex) AS Males FROM athlete_events
157            WHERE sex = 'M'
158          )
159 --concatination of values
160 SELECT 1|| ' : ' ||ROUND(CAST(Males as Float)/CAST(Females as Float),2) AS gender_ratio
161 FROM Females, Males;
162
163
```

Grid view Form view

↺

✓

✕

⏪

⏩

1

⏴

⏵

🖨

Total rows loaded: 1

	gender ratio
1	1 : 2.64

12.Top 5 athlete who have won the most gold medal

```
168 WITH top_5_gold_winners AS (
169     SELECT name ,
170           COUNT(medal) AS total_medal,
171           DENSE_RANK() OVER(ORDER BY COUNT(medal) DESC) AS ranking
172     FROM athlete_events
173     WHERE medal = 'Gold'
174     GROUP BY name
175   )
176 SELECT name, total_medal
177 FROM top_5_gold_winners
178 WHERE ranking between 1 and 5;
179
```

Grid view Form view

↺

✓

✕

⏪

⏩

1

⏴

⏵

🖨

Total rows loaded: 18

	name	total medal
1	Michael Fred Phelps, II	23
2	Raymond Clarence "Ray" Ewry	10
3	Paavo Johannes Nurmi	9
4	Mark Andrew Spitz	9

13.Top 5 athletes who have won the most medals [gold/silver/bronze]

```
185 WITH Top_5_athletes AS (  
186     SELECT name, region, team, sport,  
187           COUNT(medal) AS total_medal,  
188           DENSE_RANK() OVER(ORDER BY count(medal) DESC) AS ranking  
189     FROM athlete_events AE  
190     JOIN noc_regions R  
191       USING(noc)  
192     WHERE medal <> 'NA'  
193     GROUP BY name, region  
194   )  
195   SELECT name, region, sport, total_medal  
196   FROM Top_5_athletes WHERE ranking <= 5;  
197  
198
```

Grid viewForm view

1

Total rows loaded: 15

	name	region	sport	total medal
1	Michael Fred Phelps, II	USA	Swimming	28
2	Larysa Semenivna Latynina (Diriy-)	Russia	Gymnastics	18
3	Nikolay Yefimovich Andrianov	Russia	Gymnastics	15
4	Borys Anfiyanovych Shakhlin	Russia	Gymnastics	13
5	Edoardo Mangiarotti	Italy	Fencing	13
6	Ole Einar Bjrndalen	Norway	Biathlon	13
7	Takashi Ono	Japan	Gymnastics	13
8	Aleksey Yuryevich Nemov	Russia	Gymnastics	12
9	Boris Filippovich Shcherbak	Soviet Union	Gymnastics	12

14.Fetch the top 5 most successful country in Olympics. Success is defined by the no of medals won in total.

```
202  
203 WITH most_successful AS (  
204     SELECT R.region AS country,  
205           COUNT(medal) AS no_of_medals,  
206           DENSE_RANK() OVER(ORDER BY count(medal) DESC) AS ranking  
207     FROM athlete_events E  
208     JOIN noc_regions R  
209       USING(noc)  
210     WHERE medal <> 'NA'  
211     GROUP BY R.region  
212   )  
213   SELECT country, no_of_medals  
214   FROM most_successful  
215   WHERE ranking <= 5;  
216
```

Grid viewForm view

1

Total rows loaded: 5

	country	no of medals
1	USA	5637
2	Russia	3947
3	Germany	3756
4	UK	2068
5	France	1777

15. List the Total gold, silver & bronze won by each country.

```
223
224 SELECT region,
225         COUNT(Case When medal = 'Gold' Then 1 END) AS Gold,
226         COUNT(Case When medal = 'Silver' Then 1 END) AS Silver,
227         COUNT(Case When medal = 'Bronze' Then 1 END) AS Bronze
228 FROM athlete_events AE
229 JOIN noc_regions R
230     USING(noc)
231 WHERE medal <> 'NA'
232 GROUP BY region
233 ORDER BY Gold desc, Silver desc, Bronze desc;
234
```

Grid viewForm view

✓

✗

⏪

⏩

1

⏪

⏩

Total rows loaded: 136

	region	Gold	Silver	Bronze
1	USA	2638	1641	1358
2	Russia	1599	1170	1178
3	Germany	1301	1195	1260
4	UK	678	739	651
5	Italy	575	531	531
6	France	501	610	666
-	Sweden	470	522	525

16. List the total gold, silver & bronze medals won by each country corresponding to each Olympic game.

```
239
240 SELECT games, region,
241         COUNT(Case When AE.medal = 'Gold' Then 1 END) AS Gold,
242         COUNT(Case When AE.medal = 'Silver' Then 1 END) AS Silver,
243         COUNT(Case When AE.medal = 'Bronze' Then 1 END) AS Bronze
244 FROM athlete_events AE
245 JOIN noc_regions R
246     USING(noc)
247 WHERE medal <> 'NA'
248 GROUP BY games, region;
249
```

Grid viewForm view

✓

✗

⏪

⏩

1

⏪

⏩

Total rows loaded: 1640

	games	region	Gold	Silver	Bronze
1	1896 Summer	Australia	2	0	1
2	1896 Summer	Austria	2	1	2
3	1896 Summer	Denmark	1	2	3
4	1896 Summer	France	5	4	2
5	1896 Summer	Germany	25	5	2
6	1896 Summer	Greece	10	18	20
7	1896 Summer	Hungary	2	1	3
8	1896 Summer	Switzerland	1	2	0
9	1896 Summer	UK	3	3	3
10	1896 Summer	USA	11	7	2

17. Identify which country won the most gold, most silver, & most bronze medals in each Olympic games.

```
254
255 WITH Temp as (
256     SELECT games, region,
257           COUNT(Case When medal = 'Gold' Then 1 END) AS Gold,
258           SUM(Case When medal = 'Silver' Then 1 else 0 END) AS Silver,
259           SUM(Case When medal = 'Bronze' Then 1 else 0 END) AS Bronze
260     FROM athlete_events AE
261     JOIN noc_regions R
262       USING(noc)
263     WHERE medal <> 'NA'
264     GROUP BY games, region
265 )
266 SELECT distinct games,
267       (FIRST_VALUE(region) over(partition by games order by Gold desc) || ' - ' || FIRST_VALUE(Gold) over(partition by games order by Gold Desc)) AS Max_Gold,
268       (FIRST_VALUE(region) over(partition by games order by Silver desc) || ' - ' || FIRST_VALUE(Silver) over(partition by games order by Silver Desc)) AS Max_Silver,
269       (FIRST_VALUE(region) over(partition by games order by Bronze desc) || ' - ' || FIRST_VALUE(Bronze) over(partition by games order by Bronze Desc)) AS Max_Bronze
270 FROM Temp
271 ORDER BY games;
272
```

Grid viewForm view

✓

✕

⏪

⏩

1

⏴

⏵

Total rows loaded: 51

	games	Max Gold	Max Silver	Max Bronze
1	1896 Summer	Germany - 25	Greece - 18	Greece - 20
2	1900 Summer	UK - 59	France - 101	France - 82
3	1904 Summer	USA - 128	USA - 141	USA - 125
4	1906 Summer	Greece - 24	Greece - 48	Greece - 30
5	1908 Summer	UK - 147	UK - 131	UK - 90
6	1912 Summer	Sweden - 103	UK - 64	UK - 59
7	1920 Summer	USA - 111	France - 71	Belgium - 66
8	1924 Summer	USA - 97	France - 51	USA - 49

18.Oldest and youngest Nigerian athletes.

```
366
367 WITH Oldest_athlete AS (
368     SELECT name|| ' - ' || age|| ' - ' ||sport AS oldest
369     FROM athlete_events
370     WHERE age <> 'NA' AND noc = 'NGR'
371     ORDER BY age desc LIMIT 1
372 ),
373 Youngest_athlete AS (
374     SELECT name|| ' - ' || age|| ' - ' ||sport AS Youngest
375     FROM athlete_events
376     WHERE age <> 'NA' AND noc = 'NGR'
377     ORDER BY age LIMIT 1
378 )
379 SELECT oldest, youngest
380 FROM Oldest_athlete, Youngest_athlete;
381
```

Grid viewForm view

1

Total rows loaded: 1

	oldest	youngest
1	Segun Moses Toriola - 41 - Table Tennis	Barbara Diribe - 15 - Handball

21.Which Nigerian athlete has won the most medal

```
395
396 WITH athlete_with_most_medal AS (
397     SELECT name, count( medal) as medals,
398           DENSE_RANK() OVER(order by count(medal) desc ) rank
399     FROM athlete_events
400     WHERE noc = 'NGR' AND medal <> 'NA'
401     GROUP BY name
402 )
403 SELECT name, medals
404 FROM athlete_with_most_medal
405 WHERE rank = 1;
406
407
```

Grid viewForm view

1

Total rows loaded: 3

	name	medals
1	Nkemdilim "Mary" Onyali-Omagbemi	2
2	Falilat Ogunkoya (-Osheku)	2
3	Enee Enefiok Udo-Obong	2

20.Which sport/event has Nigeria won the most medals.

```
321 WITH Nigeria_most_sport_win AS (
322     SELECT sport,
323           count(medal) AS total_medal,
324           DENSE_RANK() OVER(ORDER BY count(medal) DESC) AS ranking
325     FROM athlete_events AE
326     JOIN noc_regions R
327     USING(noc)
328     WHERE region = 'Nigeria' AND medal <> 'NA'
329     GROUP BY sport
330 )
331 SELECT sport, total_medal
332 FROM Nigeria_most_sport_win
333 WHERE ranking = 1;
334
```

Grid viewForm view

1

Total rows loaded: 1

	sport	total medal
1	Football	50

22.How many medals has Nigeria won

```
410
411 SELECT count(medal) AS Total_medal
412 FROM athlete_events
413 WHERE noc = 'NGR' AND medal <> 'NA';
414
415
416
```

Grid viewForm view

1

Total rows loaded: 1

	Total medal
1	99

SOME INSIGHTS FROM THE ANALYSING OF THE DATA INCLUDE ;

- A total of 51 games was played between 1896 – 2016 comprising of 1184 different teams.
- 1896 summer had the least participants of 12 countries while 2016 summer had the most of 207 countries.
- UK, Switzerland, Italy, and France took part in all 51 Olympic games.
- The oldest athlete to take part in the Olympics is John Quincy Adams Ward, 97 years old from Amsterdam.
- Michael Fred Phelps of USA have won the most medal of 28 in total.
- USA is the most successful country, 5637 medals won i.e defining success by the total medal won.
- Nigeria participated in 16 of the Olympic games and won the most medal in the football sport (50 medals). In total they have won 99 medals between 1896 – 2016.
- Out of the 565 Nigerian athletes, only 96 won medals.

By doing this project, I was able to properly put into use lots of SQL syntax and functions.

I'm officially a big fan of CTEs and GROUP BY 😎

THANK YOU FOR READING