



Olympics Dataset Project

Yasmin Soltani

Inspired by techTFQ

INTRODUCTION

The goal of this project is to become acquainted with SQL concepts by solving real-world data questions; we will accomplish this by downloading 120 years of Olympics History dataset from Kaggle and answering each question provided by techTFQ accordingly.

Table 1: athlete_events

column	type
ID	FLOAT
Name	VARCHAR(255)
Sex	VARCHAR(255)
Age	FLOAT
Height	FLOAT
Weight	FLOAT
Team	VARCHAR(255)
NOC	VARCHAR(255)
Games	VARCHAR(255)
Year	FLOAT
Season	VARCHAR(255)
City	VARCHAR(255)
Sport	VARCHAR(255)
Event	VARCHAR(255)
Medal	VARCHAR(255)

Table 2: noc_regions

column	type
NOC	VARCHAR(255)
region	VARCHAR(255)
notes	VARCHAR(255)

Case Study Questions

1. How many olympics games have been held?

Problem statement: Write a SQL query to find the total no of Olympic Games held as per the dataset.

```
SELECT COUNT(DISTINCT(Games)) as NumberOfGames
FROM athlete_events$
```

Answer

NumberOfGames
51

2. List down all Olympics games held so far.

Problem statement: Write a SQL query to list down all the Olympic Games held so far

```
SELECT DISTINCT Year, Season, City
FROM athlete_events$
ORDER BY Year
```

Answer

Year	Season	City
1896	Summer	Athina
1900	Summer	Paris
1904	Summer	St. Louis
1906	Summer	Athina
1908	Summer	London
1912	Summer	Stockholm
1920	Summer	Antwerpen
1924	Winter	Chamonix
1924	Summer	Paris
1928	Summer	Amsterdam
1928	Winter	Sankt Moritz
1932	Summer	Los Angeles
1932	Winter	Lake Placid
1936	Summer	Berlin
1936	Winter	Garmisch-Partenkirchen
1948	Summer	London
1948	Winter	Sankt Moritz

Showing few records only

3. Mention the total no of nations who participated in each olympics game?

Problem statement: SQL query to fetch total no of countries participated in each olympic games

```
with all_countries as
(
  SELECT Distinct Games, reg.noc
  FROM athlete_events$ ev
  JOIN noc_regions$ reg
  ON ev.noc = reg.noc
  GROUP BY Games, reg.noc)

SELECT Games, count(1) as total_countires
FROM all_countries
GROUP BY Games
ORDER BY Games
```

Answer

Games	total_countries
1896 Summer	12
1900 Summer	31
1904 Summer	15
1906 Summer	21
1908 Summer	22
1912 Summer	29
1920 Summer	29
1924 Summer	45
1924 Winter	19
1928 Summer	46
1928 Winter	25
1932 Summer	47

Showing few records only

4. Which year saw the highest and lowest no of countries participating in olympics

Problem statement: write a SQL query to return the Olympic Games which had the highest participating countries and the lowest participating countries

```
with all_countries as

(SELECT Distinct Games, reg.noc
FROM athlete_events$ ev
JOIN noc_regions$ reg
ON ev.noc = reg.noc
GROUP BY Games, reg.noc),

Tot_Countries as
(
SELECT Games, count(1) as total_countries
FROM all_countries
GROUP BY Games
)

SELECT DISTINCT
CONCAT(first_value(Games) over (order by total_countries), ' - '
, first_value(total_countries) over(order by total_countries)) as Lowest_Countries,
concat(first_value(Games) over(order by total_countries desc), ' - '
, first_value(total_countries) over(order by total_countries desc)) as
Highest_Countries
from tot_countries
order by 1;
```

Answer

Lowest_Countries	Highest_Countries
1896 Summer - 12	2016 Summer - 206

5. Which nation has participated in all of the olympic games

Problem statement: SQL query to return the list of countries who have been part of every Olympics games

```
WITH countries as
(SELECT Games, reg.noc as country
FROM athlete_events$ ev
JOIN noc_regions$ reg --joining the other table
ON ev.noc = reg.noc --setting noc number of columns to be the same
GROUP BY Games, reg.noc),

games as
(SELECT country, COUNT(DISTINCT Games) as total_number_games
FROM countries
GROUP BY country)

SELECT country, total_number_games
FROM games
WHERE total_number_games >= (SELECT COUNT(DISTINCT Games) FROM countries)
ORDER BY total_number_games desc
```

Answer

country	total_number_games
FRA	51
GBR	51
ITA	51
SUI	51

6. Identify the sport which was played in all summer olympics

Problem statement: SQL query to fetch the list of all sports which have been part of every olympics

```
With t1 as
(Select count(distinct Games) as total_summer_games
FROM athlete_events$
WHERE Season = 'Summer'
),
t2 as
(SELECT DISTINCT Sport, Games
FROM athlete_events$
WHERE Season = 'Summer'
),
t3 as
(SELECT Sport, Count(Games) as no_of_games
from t2
group by Sport)

SELECT *
FROM t3
JOIN t1
on t1.total_summer_games = t3.no_of_games
```

Answer

Sport	no_of_games	total_summer_games
Gymnastics	29	29
Fencing	29	29
Swimming	29	29
Cycling	29	29
Athletics	29	29

7. Which Sports were just played only once in the olympics

Problem statement: using SQL query, Identify the sport which were just played once in all of olympics

```
WITH t1 as
(SELECT distinct games, Sport
FROM athlete_events$
),
t2 as
(SELECT sport, Count(games) as no_of_games
from t1
group by sport)

select t2.*, t1.games
from t2
join t1 on t1.sport = t2.sport
where t2.no_of_games = 1
order by sport;
```

Answer

Sport	no_of_games	games
Aeronautics	1	1936 Summer
Basque Pelota	1	1900 Summer
Cricket	1	1900 Summer
Croquet	1	1900 Summer
Jeu De Paume	1	1908 Summer
Military Ski Patrol	1	1924 Winter
Motorboating	1	1908 Summer
Racquets	1	1908 Summer
Roque	1	1904 Summer
Rugby Sevens	1	2016 Summer

8. Fetch the total no of sports played in each olympic games

Problem statement: write SQL query to fetch the total no of sports played in each olympics

```
SELECT Games, count(DISTINCT (Sport)) as no_of_sports
from athlete_events$
group by games
order by no_of_sports desc
```

Answer

Games	no_of_sports
2016 Summer	34
2008 Summer	34
2004 Summer	34
2000 Summer	34
2012 Summer	32
1996 Summer	31
1992 Summer	29
1988 Summer	27
1984 Summer	25
1920 Summer	25
1908 Summer	24
1936 Summer	24
1976 Summer	23
1980 Summer	23
1972 Summer	23

Showing few records only

9. Fetch oldest athletes to win a gold medal

Problem statement: SQL Query to fetch the details of the oldest athletes to win a gold medal at the olympics

```
SELECT TOP 2 name, sex, age, team, games, city, sport, event, medal
FROM athlete_events$
WHERE Medal = 'Gold'
ORDER BY AGE desc
```

Answer

name	sex	age	team	games	city	sport	event	medal
Oscar Gomer Swahn	M	64	Sweden	1912 Summer	Stockholm	Shooting	Shooting Men's Running Target, Single Shot, Team	Gold
Charles Jacobus	M	64	United States	1904 Summer	St. Louis	Roque	Roque Men's Singles	Gold

10. Find the Ratio of male and female athletes participated in all olympic games

Problem statement: write a SQL query to get the ratio of male and female participants

```
with m as
(SELECT COUNT(Sex) as tot_males
FROM athlete_events$
WHERE Sex = 'M'
GROUP BY Sex
),
f as
(SELECT COUNT(Sex) as tot_females
FROM athlete_events$
WHERE Sex = 'F'
GROUP BY Sex
)
SELECT concat('1: ', str(CAST(tot_males/tot_females AS decimal), 4, 3)) as ratio
FROM m, f
```

Answer

ratio

1: 2.00

11. Fetch the top 5 athletes who have won the most gold medals

Problem statement: SQL query to fetch the top 5 athletes who have won the most gold medals

```
WITH t1 as
(SELECT Name, Team, COUNT(Medal) as total_gold_medals
FROM athlete_events$
WHERE Medal ='Gold'
GROUP BY Name, Team),

t2 as
(SELECT *, dense_rank() over (order by total_gold_medals desc) as rank
FROM t1)

SELECT *
FROM t2
WHERE rank <=5;
```

Answer

name	Team	total_gold_medals	rank
Michael Fred Phelps, II	United States	23	1
Raymond Clarence "Ray" Ewry	United States	10	2
Mark Andrew Spitz	United States	9	3
Frederick Carlton "Carl" Lewis	United States	9	3
Larysa Semenivna Latynina (Diriy-)	Soviet Union	9	3
Paavo Johannes Nurmi	Finland	9	3
Usain St. Leo Bolt	Jamaica	8	4
Sawao Kato	Japan	8	4
Ole Einar Bjrndalen	Norway	8	4
Jennifer Elisabeth "Jenny" Thompson (-Cumpelik)	United States	8	4
Matthew Nicholas "Matt" Biondi	United States	8	4
Donald Arthur "Don" Schollander	United States	7	5

Showing few records only

12. Fetch the top 5 athletes who have won the most medals (gold/silver/bronze)

Problem statement: SQL Query to fetch the top 5 athletes who have won the most medals (Medals include gold, silver and bronze)

```
SELECT TOP 5 Name, Team, COUNT(Medal) as tot_number_medals
FROM athlete_events$
WHERE Medal IN ('Gold', 'Silver', 'Bronze')
GROUP BY Name, Team
ORDER BY tot_number_medals desc
```

Answer

Name	Team	tot_number_medals
Michael Fred Phelps, II	United States	28
Larysa Semenivna Latynina (Diriy-)	Soviet Union	18
Nikolay Yefimovich Andrianov	Soviet Union	15
Edoardo Mangiarotti	Italy	13
Takashi Ono	Japan	13

13. Fetch the top 5 most successful countries in olympics. Success is defined by no of medals won

Problem statement: write a SQL query to fetch the top 5 most successful countries in olympics. (Success is defined by no of medals won)

```
WITH countries AS
( SELECT n.region, count(medal) AS total_medals
  FROM athlete_events$ a
    JOIN noc_regions$ n
    ON n.noc = a.noc
 WHERE medal <> 'NA'
 GROUP BY n.region),

rank AS
( SELECT *,
  DENSE_RANK() OVER (ORDER BY total_medals DESC) AS rnk
  FROM countries)

SELECT *
FROM rank
WHERE rnk <=5
ORDER BY total_medals DESC
```

Answer

Team	tot_number_medals	rnk
USA	5637	1
Russia	3947	2
Germany	3756	3
UK	2068	4
France	1777	5

14. List down total gold, silver and bronze medals won by each country

Problem statement: write a SQL query to list down the total gold, silver and bronze medals won by each country

```
SELECT country,
coalesce([Gold], 0) as Gold, --coalesce is to remove NULLS
coalesce([Silver], 0) as Silver,
coalesce([Bronze], 0) as Bronze

FROM
(
SELECT n.region as country, Medal, COUNT(1) as tot_medals
FROM athlete_events$ a
INNER JOIN noc_regions$ n
ON a.noc = n.noc
WHERE Medal <> 'NA' -- excluding NA from our result
GROUP BY n.region, Medal
--ORDER BY n.region, Medal
) AS Source_table

PIVOT --converts rows into columns
(
SUM(tot_medals)
FOR[Medal] IN ([Gold], [Silver], [Bronze])
) AS Pivot_table

ORDER BY gold desc, silver desc, bronze desc
```

Answer

country	gold	silver	bronze
USA	2638	1641	1358
Russia	1599	1170	1178
Germany	1301	1195	1260
UK	678	739	651
Italy	575	531	531
France	501	610	666
Sweden	479	522	535
Canada	463	438	451
Hungary	432	332	371
Norway	378	361	294
Australia	368	459	522
China	351	349	293
Netherlands	287	340	413

Showing few records only

15. List down total gold, silver and bronze medals won by each country corresponding to each olympic games

Problem statement: write a SQL query to list down the total gold, silver and bronze medals won by each country corresponding to each olympic games

```
SELECT n.region, Games,
SUM(CASE WHEN Medal = 'Gold' THEN 1 ELSE 0 END) as gold,
SUM(CASE WHEN Medal = 'Silver' THEN 1 ELSE 0 END) as silver,
SUM(CASE WHEN Medal = 'Bronze' THEN 1 ELSE 0 END) as bronze
FROM noc_regions$ n
INNER JOIN athlete_events$ a
ON n.noc = a.noc
GROUP BY region, games
ORDER BY Games
```

Answer

region	Games	gold	silver	bronze
Australia	1896 Summer	2	0	1
Austria	1896 Summer	2	1	2
Denmark	1896 Summer	1	2	3
France	1896 Summer	5	4	2
Germany	1896 Summer	25	5	2
Greece	1896 Summer	10	18	20
Hungary	1896 Summer	2	1	3
Italy	1896 Summer	0	0	0
Sweden	1896 Summer	0	0	0
Switzerland	1896 Summer	1	2	0
UK	1896 Summer	3	3	3
USA	1896 Summer	11	7	2

Showing few records only

16. Identify which country won the most gold, most silver and most bronze medals in each olympic games

Problem statement: write SQL query to display for each Olympic Games, which country won the highest gold, silver and bronze medals

```
with temp as
(
SELECT substring(games_country, 1, 11 ) as games,
substring(games_country, 13, 15 ) as country,
coalesce([Gold], 0) as Gold, --coalesce is to remove NULLS
coalesce([Silver], 0) as Silver,
coalesce([Bronze], 0) as Bronze

FROM
(
SELECT concat(games, '-', n.region) as games_country, Medal, COUNT(1) as tot_medals
FROM athlete_events$ a
INNER JOIN noc_regions$ n
ON a.noc = n.noc
WHERE Medal <> 'NA' -- excluding NA from our result
GROUP BY n.region, games, Medal
--ORDER BY n.region, Medal
) AS Source_table

PIVOT --converts rows into columns
(
SUM(tot_medals)
FOR[Medal] IN ([Gold], [Silver], [Bronze])
) AS Pivot_table
)

SELECT DISTINCT games , concat(first_value(country) over(partition by games order
by gold desc), ' - ',
first_value(gold) over(partition by games order by gold desc)) as Max_Gold,
concat(first_value(country) over(partition by games order by silver desc), ' - ',
first_value(silver) over(partition by games order by silver desc)) as Max_Silver,
concat(first_value(country) over(partition by games order by bronze desc), ' - ',
first_value(bronze) over(partition by games order by bronze desc)) as Max_Bronze

FROM temp
ORDER BY games;
```

Answer

games	Max_Gold	Max_Silver	Max_Bronze
1896 Summer	Germany - 25	Greece - 18	Greece - 20
1900 Summer	UK - 59	France - 101	France - 82
1904 Summer	USA - 128	USA - 141	USA - 125
1906 Summer	Greece - 24	Greece - 48	Greece - 30
1908 Summer	UK - 147	UK - 131	UK - 90
1912 Summer	Sweden - 103	UK - 64	UK - 59
1920 Summer	USA - 111	France - 71	Belgium - 66
1924 Summer	USA - 97	France - 51	USA - 49
1924 Winter	UK - 16	USA - 10	UK - 11
1928 Summer	USA - 47	Netherlands - 29	Germany - 41
1928 Winter	Canada - 12	Sweden - 13	Switzerland - 12
1932 Summer	USA - 81	USA - 47	USA - 61
1932 Winter	Canada - 14	USA - 21	Germany - 14

Showing few records only

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

Problem statement: similar to the previous query, identify during each Olympic Games, which country won the highest gold, silver and bronze medals. Along with this, identify also the country with the most medals in each olympic games

```

with temp as
(
SELECT substring(games_country, 1, 11 ) as games,
substring(games_country, 13, 15 ) as country,
coalesce([Gold], 0) as Gold, --coalesce is to remove NULLS
coalesce([Silver], 0) as Silver,
coalesce([Bronze], 0) as Bronze

FROM
(
SELECT concat(games, '-', n.region) as games_country, Medal, COUNT(1) as tot_medals
FROM athlete_events$ a
INNER JOIN noc_regions$ n
ON a.noc = n.noc
WHERE Medal <> 'NA' -- excluding NA from our result
GROUP BY n.region, games, Medal
--ORDER BY n.region, Medal
) AS Source_table

PIVOT --converts rows into columns
(
    SUM(tot_medals)
    FOR[Medal] IN ([Gold], [Silver], [Bronze])
) AS Pivot_table
),

tot_medals as
(
SELECT games, nr.region as country, count(1) as total_medals
FROM athlete_events$ oh
JOIN noc_regions$ nr
    ON nr.noc = oh.noc
where medal <> 'NA'
GROUP BY games,nr.region )

SELECT DISTINCT t.games , concat(first_value(t.country) over(partition by t.games
order by gold desc), ' - ',
first_value(gold) over(partition by t.games order by gold desc)) as Max_Gold,
concat(first_value(t.country) over(partition by t.games order by silver desc), ' - ',
first_value(silver) over(partition by t.games order by silver desc)) as Max_Silver,
concat(first_value(t.country) over(partition by t.games order by bronze desc), ' - ',
first_value(bronze) over(partition by t.games order by bronze desc)) as Max_Bronze,
concat(first_value(tm.country) over (partition by tm.games order by total_medals
desc) , ' - '
, first_value(tm.total_medals) over(partition by tm.games order by
total_medals desc)) as Max_Medals
FROM temp t
join tot_medals tm on tm.games = t.games and tm.country = t.country
ORDER BY games;

```

Answer

games	Max_Gold	Max_Silver	Max_Bronze	Max_Medals
1896 Summer	Germany - 25	Greece - 18	Greece - 20	Greece - 48
1900 Summer	UK - 59	France - 101	France - 82	France - 235
1904 Summer	USA - 128	USA - 141	USA - 125	USA - 394
1906 Summer	Greece - 24	Greece - 48	Greece - 30	Greece - 102
1908 Summer	UK - 147	UK - 131	UK - 90	UK - 368
1912 Summer	Sweden - 103	UK - 64	UK - 59	Sweden - 190
1920 Summer	USA - 111	France - 71	Belgium - 66	USA - 194
1924 Summer	USA - 97	France - 51	USA - 49	USA - 182
1924 Winter	UK - 16	USA - 10	UK - 11	UK - 31

Showing few records only

18. Which countries have never won gold medal but have won silver/bronze medals?

Problem statement: write a SQL Query to fetch details of countries which have won silver or bronze medal but never won a gold medal

```

with number_of_medals as
(
  SELECT n.region,
  SUM(CASE WHEN Medal = 'Gold' THEN 1 ELSE 0 END) as gold,
  SUM(CASE WHEN Medal = 'Silver' THEN 1 ELSE 0 END) as silver,
  SUM(CASE WHEN Medal = 'Bronze' THEN 1 ELSE 0 END) as bronze

  from noc_regions$ n
  inner join athlete_events$ a
  on n.NOC = a.NOC
  group by region
)
SELECT region, gold, silver, bronze
from number_of_medals
where gold = 0
and (silver > 0 or bronze > 0 )
order by silver, bronze desc

```

Answer

region	gold	silver	bronze
Afghanistan	0	0	2
Kuwait	0	0	2
Guyana	0	0	1
Macedonia	0	0	1
Mauritius	0	0	1
Barbados	0	0	1
Bermuda	0	0	1
Djibouti	0	0	1
Eritrea	0	0	1
Monaco	0	0	1
Iraq	0	0	1
Togo	0	0	1
Ghana	0	1	22

19. In which Sport/event, India has won highest medals

Problem statement: write SQL Query to return the sport which has won India the highest no of medals

```
SELECT top 1 sport, count(medal) as tot_medals
from athlete_events$
where team = 'India'
and medal != 'NA'
group by sport
order by tot_medals desc
```

Answer

sport	tot_medals
Hockey	173

20. Break down all olympic games where India won medal for Hockey and how many medals in each olympic games

Problem statement: write an SQL Query to fetch details of all Olympic Games where India won medal(s) in hockey

```
SELECT team, sport, games, count(medal) as total_medals
FROM athlete_events$
where sport = 'Hockey'
and team = 'India'
and medal != 'NA'
group by Team, Games, sport
order by total_medals desc
```

Answer

team	sport	games	total_medals
India	Hockey	1948 Summer	20
India	Hockey	1936 Summer	19
India	Hockey	1956 Summer	17
India	Hockey	1968 Summer	16
India	Hockey	1980 Summer	16
India	Hockey	1964 Summer	15
India	Hockey	1932 Summer	15
India	Hockey	1928 Summer	14
India	Hockey	1952 Summer	14
India	Hockey	1972 Summer	14
India	Hockey	1960 Summer	13

THE END :)