OLYMPICS

1896 - 2016

DATA ANALYSIS

SQL CASE STUDY

PROJECT BY: KOUSTAV SANTRA

OVERVIEW: EXPLORING OLYMPIC DATA USING SQL

The Olympics symbolize global unity and athletic excellence, with a wealth of data on athletes, events, and trends. Leveraging this data through SQL (Structured Query Language) provides a powerful tool for analysis, insight generation, and decision-making.

The SQL case study on the Olympics aims to uncover insights using SQL queries. It explores various aspects, including the total number of Olympic games, nations' participation, trends over time, and notable achievements like top athletes and oldest medallists. Additionally, it examines sports played maximum time, and India's hockey performance. Through SQL analysis, the project aims to provide concise insights into the rich Olympic dataset.

The dataset consists of one table with the following columns:

public	
olympics_history	
id integer	noc character varying
name character varying	games character varying
sex character varying	year integer
age character varying	season character varying
height character varying	city character varying
weight character varying	sport character varying
team character varying	event character varying
	medal character varying

1) HOW MANY OLYMPICS GAMES HAVE BEEN HELD?

```sql
select count(distinct(games)) from OLYMPICS_HISTORY
```

	count bigint 🔒
1	51

2) LIST DOWN ALL OLYMPICS GAMES HELD SO FAR.

```
select distinct(year),season,city from OLYMPICS_HISTORY order by year
```

	year integer	season character varying	city character varying
1	1896	Summer	Athina
2	1900	Summer	Paris
3	1904	Summer	St. Louis
4	1906	Summer	Athina
5	1908	Summer	London
6	1912	Summer	Stockholm
7	1920	Summer	Antwerpen
8	1924	Summer	Paris
9	1924	Winter	Chamonix
10	1928	Summer	Amsterdam
11	1928	Winter	Sankt Moritz
12	1932	Summer	Los Angeles
13	1932	Winter	Lake Placid
14	1936	Winter	Garmisch-Partenkirchen
15	1936	Summer	Berlin
16	1948	Winter	Sankt Moritz
17	1948	Summer	London
18	1952	Winter	Oslo

	year integer	season character varying	city character varying
35	1984	Winter	Sarajevo
36	1984	Summer	Los Angeles
37	1988	Summer	Seoul
38	1988	Winter	Calgary
39	1992	Winter	Albertville
40	1992	Summer	Barcelona
41	1994	Winter	Lillehammer
42	1996	Summer	Atlanta
43	1998	Winter	Nagano
44	2000	Summer	Sydney
45	2002	Winter	Salt Lake City
46	2004	Summer	Athina
47	2006	Winter	Torino
48	2008	Summer	Beijing
49	2010	Winter	Vancouver
50	2012	Summer	London
51	2014	Winter	Sochi
52	2016	Summer	Rio de Janeiro

18	1952	Winter	Oslo
19	1952	Summer	Helsinki
20	1956	Summer	Melbourne
21	1956	Summer	Stockholm
22	1956	Winter	Cortina d'Ampezzo
23	1960	Winter	Squaw Valley
24	1960	Summer	Roma
25	1964	Summer	Tokyo
26	1964	Winter	Innsbruck
27	1968	Winter	Grenoble
28	1968	Summer	Mexico City
29	1972	Summer	Munich
30	1972	Winter	Sapporo
31	1976	Winter	Innsbruck
32	1976	Summer	Montreal
33	1980	Summer	Moskva
34	1980	Winter	Lake Placid

3) Which year saw the highest and lowest no of countries participating in Olympics?

```sql
with cte as
(
select games,noc
from OLYMPICS_HISTORY
group by games,noc
order by games,noc
),
cte1 as
(
select games,count(1) as no_of_country, rank() over(order by count(1))as rnk
from cte
group by  games
order by 2
)
select *
from cte1
where rnk=1 or rnk=(select max(rnk) from cte1)
```

	games character varying	no_of_country bigint	rnk bigint
1	1896 Summer	12	1
2	2016 Summer	207	51

4) WHICH NATION HAS PARTICIPATED IN ALL OF THE OLYMPIC GAMES?

```sql
with number_of_olympic as
(
select count(distinct(games))
from OLYMPICS_HISTORY
),
cte as
(
select noc,games
from OLYMPICS_HISTORY group by noc,games order by noc
),
cte1 as
(
select noc,count(1) as country_times_participated from cte group by noc
)

select ohn.region as country ,c1.country_times_participated
from cte1 c1
inner join OLYMPICS_HISTORY_NOC_REGIONS ohn on c1.noc=ohn.noc
where country_times_participated =(select * from number_of_olympic)
```

	country character varying 🔒	country_times_participated bigint 🔒
1	France	51
2	UK	51
3	Italy	51
4	Switzerland	51

5) IDENTIFY THE SPORT WHICH WAS PLAYED IN ALL SUMMER OLYMPICS.

```sql
with cte as
(
select count(distinct(games)) from OLYMPICS_HISTORY where season='Summer'
)
select sport,count(distinct games) as no_times_played
from OLYMPICS_HISTORY
where season='Summer' group by sport
having count (distinct(games))=(select * from cte)
```

	sport character varying 🔒	no_times_played 🔒 bigint
1	Athletics	29
2	Cycling	29
3	Fencing	29
4	Gymnastics	29
5	Swimming	29

6) WHICH SPORTS WERE JUST PLAYED ONLY ONCE IN THE OLYMPICS?

```sql
with cte as
(
select count(distinct(games)) from OLYMPICS_HISTORY
)
select sport,count(distinct games) as no_times_played
from OLYMPICS_HISTORY
group by sport
having count (distinct(games))=1
```

	sport character varying 🔒	no_times_played 🔒 bigint
1	Aeronautics	1
2	Basque Pelota	1
3	Cricket	1
4	Croquet	1
5	Jeu De Paume	1
6	Military Ski Patrol	1
7	Motorboating	1
8	Racquets	1
9	Roque	1
10	Rugby Sevens	1

7) FETCH THE TOTAL NO OF SPORTS PLAYED IN EACH OLYMPIC GAME.

```
with cte as
(
select games,sport as no_times_played
from OLYMPICS_HISTORY
group by games, sport
order by games
)

select games,count(1) as number_of_sports from cte group by games order by 2 desc
```

	games character varying	number_of_sports bigint
1	2008 Summer	34
2	2000 Summer	34
3	2004 Summer	34
4	2016 Summer	34
5	2012 Summer	32
6	1996 Summer	31
7	1992 Summer	29
8	1988 Summer	27
9	1920 Summer	25
10	1984 Summer	25
11	1908 Summer	24
12	1936 Summer	24
13	1972 Summer	23
14	1980 Summer	23
15	1976 Summer	23
16	1964 Summer	21
17	1968 Summer	20
18	1900 Summer	20

	games character varying	number_of_sports bigint
19	1924 Summer	20
20	1948 Summer	20
21	1960 Summer	19
22	1956 Summer	19
23	1952 Summer	19
24	1904 Summer	18
25	1932 Summer	18
26	1928 Summer	17
27	1912 Summer	17
28	2006 Winter	15
29	2010 Winter	15
30	2002 Winter	15
31	2014 Winter	15
32	1998 Winter	14
33	1906 Summer	13
34	1992 Winter	12
35	1994 Winter	12
36	1924 Winter	10

	games character varying	number_of_sports bigint
34	1992 Winter	12
35	1994 Winter	12
36	1924 Winter	10
37	1968 Winter	10
38	1972 Winter	10
39	1976 Winter	10
40	1980 Winter	10
41	1984 Winter	10
42	1988 Winter	10
43	1964 Winter	10
44	1948 Winter	9
45	1896 Summer	9
46	1956 Winter	8
47	1960 Winter	8
48	1952 Winter	8
49	1936 Winter	8
50	1928 Winter	8
51	1932 Winter	7

8) Fetch details of the oldest athletes to win a gold medal.

```sql
-- i have done rnk 2 ,because data contain null values and sql have ranked null values as 1
with cte as
(
select * ,dense_rank()over(order by age desc) as rnk
from OLYMPICS_HISTORY
where medal='Gold'
)
select * from cte where rnk=2
```

	name character varying	sex character varying	age character varying	team character varying	games character varying	city character varying	sport character varying	medal character varying	rnk bigint
1	Charles Jacobus	M	64	United States	1904 Summer	St. Louis	Roque	Gold	2
2	Oscar Gomer Swahn	M	64	Sweden	1912 Summer	Stockholm	Shooting	Gold	2

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

```sql
with cte1 as
(
select id,count(1) from OLYMPICS_HISTORY where medal='Gold' group by id
),
cte2 as
(
select oh.id,oh.name,c1.count,dense_rank()over(order by count desc) as rnk
from cte1 c1 inner join (select distinct id,name from OLYMPICS_HISTORY )oh
on oh.id=c1.id
)
select * from cte2 where rnk < 6
```

	id integer	name character varying	count bigint	rnk bigint
1	94406	Michael Fred Phelps, II	23	1
2	33557	Raymond Clarence "Ray" Ewry	10	2
3	67046	Larysa Semenivna Latynina (Diriy-)	9	3
4	69210	Frederick Carlton "Carl" Lewis	9	3
5	113912	Mark Andrew Spitz	9	3
6	87390	Paavo Johannes Nurmi	9	3
7	35550	Birgit Fischer-Schmidt	8	4
8	119922	Jennifer Elisabeth "Jenny" Thompson (-Cumpelik)	8	4
9	13029	Usain St. Leo Bolt	8	4
10	11642	Matthew Nicholas "Matt" Biondi	8	4
11	11951	Ole Einar Bjrndalen	8	4
12	57998	Sawao Kato	8	4
13	21402	Viktor Ivanovych Chukarin	7	5
14	109161	Borys Anfiyanovych Shakhlin	7	5
15	4198	Nikolay Yefimovich Andrianov	7	5
16	107383	Donald Arthur "Don" Schollander	7	5
17	18826	Vra slavsk (-Odloilov)	7	5
18	39726	Aladr Gerevich (-Gerei)	7	5

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

```sql
with cte1 as
(
select id,count(1) from OLYMPICS_HISTORY where medal in('Gold','Silver','Bronze') group by id
),
cte2 as
(
select oh.id,oh.name,c1.count,dense_rank()over(order by count desc) as rnk
from cte1 c1 inner join (select distinct id,name from OLYMPICS_HISTORY )oh
on oh.id=c1.id
)
select * from cte2 where rnk < 6
```

	id integer	name character varying	count bigint	rnk bigint
1	94406	Michael Fred Phelps, II	28	1
2	67046	Larysa Semenivna Latynina (Diriy-)	18	2
3	4198	Nikolay Yefimovich Andrianov	15	3
4	11951	Ole Einar Bjrndalen	13	4
5	89187	Takashi Ono	13	4
6	109161	Borys Anfiyanovych Shakhlin	13	4
7	74420	Edoardo Mangiarotti	13	4
8	57998	Sawao Kato	12	5
9	23426	Natalie Anne Coughlin (-Hall)	12	5
10	119922	Jennifer Elisabeth "Jenny" Thompson (-Cumpelik)	12	5
11	87390	Paavo Johannes Nurmi	12	5
12	35550	Birgit Fischer-Schmidt	12	5
13	70965	Ryan Steven Lochte	12	5
14	121258	Dara Grace Torres (-Hoffman, -Minas)	12	5
15	85286	Aleksey Yuryevich Nemov	12	5

11) FETCH THE TOP 5 MOST SUCCESSFUL COUNTRIES IN OLYMPICS. SUCCESS IS DEFINED BY NO OF MEDALS WON.

```sql
with cte1 as
(
select noc,count(1),dense_rank()over(order by count(1)desc) as rnk
from OLYMPICS_HISTORY where medal in('Gold','Silver','Bronze')
group by noc
)
select * from cte1 where rnk < 6
```

	noc character varying	count bigint	rnk bigint
1	USA	5637	1
2	URS	2503	2
3	GER	2165	3
4	GBR	2068	4
5	FRA	1777	5

12) IN WHICH SPORT/EVENT, INDIA HAS WON HIGHEST MEDALS.

```
with cte as
(
select sport ,count(1),dense_rank()over(order by count(1)desc)
from OLYMPICS_HISTORY
where team='India' and medal in ('Gold','Silver','Bronze')
group by sport
)

select * from cte where dense_rank=1
```

	sport character varying	count bigint	dense_rank bigint
1	Hockey	173	1

13) Break down all Olympic games where India won medal for Hockey and how many medals in each Olympic game.

```sql
select games,count(1)
from OLYMPICS_HISTORY
where team='India' and sport='Hockey' and medal in ('Gold','Silver','Bronze')
group by games
order by 2 desc
```

	games character varying 🔒	count bigint 🔒
1	1948 Summer	20
2	1936 Summer	19
3	1956 Summer	17
4	1980 Summer	16
5	1968 Summer	16
6	1932 Summer	15
7	1964 Summer	15
8	1928 Summer	14
9	1972 Summer	14
10	1952 Summer	14
11	1960 Summer	13