



OLYMPICS DATASET ANALYSIS

Tamil Selvan R

PROJECT DESCRIPTION

OBJECTIVE:

- The goal of this project is to analyze the Olympics dataset spanning 120 years of history. The dataset comprises **two CSV files** - 'athletes' and 'athlete_events,' providing comprehensive information about the nations participating in the Olympics and the events that occurred over the years.
- The project involves importing these datasets into an SQL Server and addressing specific analytical questions related to Olympic performance.

DATASETS

1. ATHLETES DATASET:

1. Contains detailed information about all players who participated in the Olympics.

2. ATHLETE EVENTS DATASET:

1. Provides information about all events that occurred over the years, with the 'athlete_id' linking back to the 'id' column in the 'athletes' table.



ANALYSIS TASKS

TEAM WITH MAXIMUM GOLD MEDALS

- **OBJECTIVE:** To Identify the team that has won the maximum gold medals over the years.
- From the results we got, it is evident that **United States/France** has got the maximum number of gold medals

```
1 | -- Team with maximum gold medals over the years.
2 |
3 | SELECT a.team, COUNT(a.team) as no_of_gold
4 | FROM
5 |   athletes a JOIN athlete_events b ON a.id = b.athlete_id
6 |   WHERE b.medal = 'Gold' and a.team = 'united states/france'
7 |   GROUP BY a.team
8 |   ORDER BY no_of_gold desc
9 |
10 |
11 |
12 |
13 |
14 |
15 |
16 |
```

	team	no_of_gold
1	United States/France	4

TOTAL SILVER MEDALS FOR EACH TEAM

- **OBJECTIVE:** Total silver medals for each team and the year in which they won the maximum silver medals.
- From the analysis, the **United States** has the highest number of silver(**1546**) followed by the **Soviet Union**(**766**).

```
12 --total silver medals of each team and year in which they won maximum silver medal
13 with cte as(SELECT a.team,COUNT(a.team) as no_of_silver
14 from
15     athletes a JOIN athlete_events b ON a.id = b.athlete_id
16     where b.medal = 'Silver'
17     group by a.team
18 ),
19 cte1 as
20 (select *
21 from (select team,no_of_silver1,year,ROW_NUMBER() over (partition by team
22         order by no_of_silver1 desc) match
23 from (select team,COUNT(c.team) as no_of_silver1,year
24 from
25     athletes c JOIN athlete_events d ON c.id = d.athlete_id
26     where d.medal = 'Silver'
27     group by c.team,year
28 ) a) b
29 where match = 1)
30 select a.team,a.no_of_silver,b.year
31 from cte a join cte1 b on a.team = b.team
32 order by a.no_of_silver desc
33
34
35
```

Results Messages

Query executed successfully.

DESKTOP-30HQGF2\SQLEXPRESS ... DESKTOP-30HQGF2\hp (60) | olympic

PLAYER WITH MAXIMUM GOLD MEDALS (NO SILVER OR BRONZE):

- **OBJECTIVE:** To identify the player who has won the maximum gold medals among those who have never won silver or bronze.
- From the analysis, it is found **Raymond Clarence** has the highest Gold Medal

```
46 --player who won maximum gold medals amongst the players
47 --which have won only gold medal (never won silver or bronze) over the years
48
49 select top 2 a.name, COUNT(b.medal) no_of_gold
50 from athletes a join athlete_events b on a.id = b.athlete_id
51 where b.medal = 'Gold'
52 group by a.name
53 having a.name not in (select a.name from athletes a join athlete_events b on a.id = b.athlete_id
54 order by no_of_gold desc
55
56
57
```

	name	no_of_gold
1	Raymond Clarence "Ray" Ewry	10
2	Usain St. Leo Bolt	8

MAXIMUM GOLD MEDALS PER YEAR

- **OBJECTIVE:** For each year, find the player(s) who won the maximum gold medals.

```
56 --4 in each year which player has won maximum gold medal.
57 with cte as
58 (select year,name as player_name,count(b.medal) no_of_gold
59 from athletes a join athlete_events b on a.id = b.athlete_id
60 where b.medal = 'Gold'
61 group by year,name)
62
63 select year,STRING_AGG(player_name,',') as player_name,no_of_gold
64 from(select year,player_name,no_of_gold
65 from (select *,rank() over (partition by year order by no_of_gold desc) rn
66 from cte) a
67 where rn = 1) b
68 group by year,no_of_gold
69 order by year
70
```

	year	player_name	no_of_gold
1	1896	Carl Schuhmann	4
2	1900	Alvin Christian "Al" Kraenzlein	4
3	1904	Anton Heida	5
4	1906	Giorgio Cesana,Francesco Verri,Emilio Fontanella...	3
5	1908	Melvin Winfield "Mel" Sheppard,Henry Taylor	3

Query executed successfully. DESKTOP-30HQGF2\SQLEXPRESS ... DESKTOP-30HQGF2\hp (60)

INDIA'S FIRST GOLD, SILVER, AND BRONZE MEDAL

- **OBJECTIVE:** Identify the event and year in which India won its first gold, silver, and bronze medals.

```
71 | --India has won its first gold medal,first silver medal and first bronze medal
72 |
73 | with medal as(select team,medal,min(year) year,sport,ROW_NUMBER() over
74 | (partition by medal order by min(year)) rn
75 | from athletes a join athlete_events b on a.id = b.athlete_id
76 | where team = 'india'
77 | group by medal,sport,team)
78 |
79 | select YEAR,medal,Sport
80 | from medal
81 | where rn = 1 and medal not in ('na')
82 | order by sport
83 |
84 |
85 |
```

	YEAR	medal	Sport
1	1924	Gold	Alpinism
2	1900	Silver	Athletics
3	1952	Bronze	Wrestling

PLAYERS WINNING GOLD IN BOTH SUMMER AND WINTER OLYMPICS

- **Objective:** To find players who won gold medals in both the Summer and winter Olympics.
- We found 4 players winning Gold in both the Summer and Winter Olympics.

```
86
87 --players who won gold medal in summer and winter olympics both.
88
89 select name
90 from athletes a join athlete_events b on a.id = b.athlete_id
91 where medal = 'gold' and season in ('summer' , 'winter')
92 group by name
93 HAVING COUNT(DISTINCT season) = 2;
94
95
96
97
98
99
100
101
```

75 %

Results Messages

	name
1	Aleksandr Vladimirovich Popov
2	Edward Patrick Francis "Eddie" Eagan
3	Gillis Emanuel Grafstrm
4	Kim So-Hui

PLAYERS WINNING GOLD, SILVER, AND BRONZE IN A SINGLE OLYMPICS

- **OBJECTIVE:** To identify players who won gold, silver, and bronze medals in a single Olympics.

```
101 --players who won gold, silver and bronze medal in a single olympics. print player name
102 --along with year.
103
104 select b.year,a.name
105 from athletes a join athlete_events b on a.id = b.athlete_id
106 where medal not in ('NA') and b.medal in ('Gold','Silver','Bronze') and b.medal IS NOT NULL
107 group by year,name
108 having COUNT(distinct medal) = 3
109 order by year
110
111
112
113
114
```

	year	name
1	1896	Alexander Viggo Jensen
2	1896	Hermann Otto Ludwig Weingrtnr
3	1896	Ioannis Frangoudis
4	1896	James Brendan Bennet Connolly
5	1896	Marie Lon Flameng

PLAYERS WITH CONSECUTIVE GOLD MEDALS IN THREE SUMMER OLYMPICS

- **OBJECTIVE:** To find players who have won gold medals in the same event in three consecutive Summer Olympics starting from 2000.

```
116 --8 find players who have won gold medals in consecutive 3 summer olympics in the same event
117 with cte as
118 (SELECT name,
119         event, lag(b.year,1) OVER (PARTITION BY a.name, b.event ORDER BY b.year)
120         as previous_year,
121         year as present_year,
122         lead(b.year,1) OVER (PARTITION BY a.name, b.event ORDER BY b.year) as following_year
123 from athletes a join athlete_events b on a.id = b.athlete_id
124 where medal = 'Gold' and
125         season = 'Summer' and
126         b.year >= '2000'
127 )
128 select *
129 from cte
130 where previous_year = present_year - 4 and following_year = present_year + 4
131
```

	name	event	previous_year	pres
1	Allyson Michelle Felix	Athletics Women's 4 x 400 metres Relay	2008	201
2	Anastasiya Semyonovna Davydova	Synchronized Swimming Women's Team	2004	200
3	Artur Borisovich Taymazov	Wrestling Men's Super-Heavyweight, Freestyle	2004	200
4	Brendan Joseph Hansen	Swimming Men's 4 x 100 metres Medley Relay	2004	200
5	Carmelo Kvan Anthony	Baskethall Men's Baskethall	2008	201

Query executed successfully. | DESKTOP-30HQGF2\SQLEXPRESS ... | DESKTOP-30HQGF2\hp (60) | olympics | 00:00:00 | 62 rows



SQL TECHNIQUES USED FOR ANALYSIS

KEY SQL FUNCTIONS/CONCEPTS UTILIZED:

- WINDOW FUNCTIONS

- In the course of analyzing the 120-year-old Olympics dataset, I employed a range of SQL techniques. Functions like `LEAD()`, and `LAG()` allowed me to navigate sequential data effectively.
- Applied `ROW_NUMBER()` to assign unique row numbers to result sets
- `RANK()` was used for assigning ranks to rows to handle ties.

- AGGREGATION

- Use of `GROUP BY`, `PARTITION BY`, and `HAVING` clauses helped to perform aggregations which helped to gain insights on team performance over the years.
- `COUNT()`, `MIN/MAX` functions helped to perform statistical analysis

- STRING AGGREGATION

- Use of `STRING_AGG()` for concatenating string values

- CTE

- played a key role in streamlining the complex queries.

- SUBQUERIES:

- filtering data based on conditions specified



THANK YOU...