SQL Project



Querying the Olympics Dataset

Introduction

The Olympic Games with its rich history and global significance, have captivated the world for over a century. As an integral part of our global sports culture, the Olympics have witnessed extraordinary achievements, inspiring stories, and moments of triumph. The wealth of data accumulated over the years provides a treasure trove of insights waiting to be discovered. In this project I would be delving into the vast Olympic dataset, aiming to unveil patterns, trends, and fascinating details about the Games' history.

Project Brief

The objective of this project is to analyze and extract meaningful insights from the comprehensive Olympic dataset. We will explore various aspects of the games, including participation trends, medal distributions, athlete demographics, successful nations, and more. By employing SQL query technique, we aim to highlight exceptional performances and gain a deeper understanding of the Olympic Games' evolution over time.

Data Source

The dataset used for the analysis was gotten from Kaggle. The dataset is a historical dataset on the modern Olympic Games, including all the games from Athens 1896 to Rio 2016.

Below is the link to the dataset:

https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results

Overview of Table Structure

Olympics_History Table:

This table consists of 271116 rows and 15 columns.

Each row corresponds to an individual athlete competing in an individual Olympic event.

The columns are as follows:

- **ID** Unique number for each Athlete
- Name Athlete's name
- Sex M or F
- Age Integer
- **Height** In centimeters
- Weight In kilograms
- **Team** Team name
- **NOC** National Olympic Committee 3-letter code
- Games Year and season
- Year Integer
- Season Summer or Winter
- **City** Host city
- Sport Sport
- Event Event
- Medal Gold, Silver, Bronze, or NA

Olympics_History

Name Name	Sex	Age I	Height	Weight	Team	NOC	Games	Year :	Season	City	Sport	Event	Medal
1 A Dijiang	M	24	180	8	China	CHN	1992 Summer	1992 9	lummer	Barcelona	Basketball	Basketball Men's Basketball	NA
2 A Lamusi	M	23	170	6) China	CHN	2012 Summer	2012 9	Summer	London	Judo	Judo Men's Extra-Lightweight	NA
3 Gunnar Nielsen Aaby	M	24 1	NA	NA	Denmark	DEN	1920 Summer	1920 9	ummer	Antwerpen	Football	Football Men's Football	NA
4 Edgar Lindenau Aabye	M	34 N	NA	NA	Denmark/Sweden	DEN	1900 Summer	1900 9	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
5 Christine Jacoba Aaftink	F	21	185	8	Netherlands	NED	1988 Winter	1988 \	Vinter	Calgary	Speed Skating	Speed Skating Women's 500 metres	NA
5 Christine Jacoba Aaftink	F	21	185	8:	Netherlands	NED	1988 Winter	1988 \	Vinter	Calgary	Speed Skating	Speed Skating Women's 1,000 metres	NA
5 Christine Jacoba Aaftink	F	25	185	8:	Netherlands	NED	1992 Winter	1992 \	Vinter	Albertville	Speed Skating	Speed Skating Women's 500 metres	NA
5 Christine Jacoba Aaftink	F	25	185	8:	Netherlands	NED	1992 Winter	1992 \	Vinter	Albertville	Speed Skating	Speed Skating Women's 1,000 metres	NA
5 Christine Jacoba Aaftink	F	27	185	8	Netherlands	NED	1994 Winter	1994 \	Winter	Lillehammer	Speed Skating	Speed Skating Women's 500 metres	NA
5 Christine Jacoba Aaftink	F	27	185	8	Netherlands	NED	1994 Winter	1994 \	Winter	Lillehammer	Speed Skating	Speed Skating Women's 1,000 metres	NA
6 Per Knut Aaland	M	31	188	7	United States	USA	1992 Winter	1992 \	Winter	Albertville	Cross Country Skiing	Cross Country Skiing Men's 10 kilometres	NA
6 Per Knut Aaland	M	31	188	7	United States	USA	1992 Winter	1992 \	Vinter	Albertville	Cross Country Skiing	Cross Country Skiing Men's 50 kilometres	NA
6 Per Knut Aaland	M	31	188	7	United States	USA	1992 Winter	1992 \	Vinter	Albertville	Cross Country Skiing	Cross Country Skiing Men's 10/15 kilometres	NA
6 Per Knut Aaland	M	31	188	7	United States	USA	1992 Winter	1992 \	Winter	Albertville	Cross Country Skiing	Cross Country Skiing Men's 4 x 10 kilometres	NA NA
6 Per Knut Aaland	M	33	188	7	United States	USA	1994 Winter	1994 \	Winter	Lillehammer	Cross Country Skiing	Cross Country Skiing Men's 10 kilometres	NA
6 Per Knut Aaland	M	33	188	7	United States	USA	1994 Winter	1994 \	Vinter	Lillehammer	Cross Country Skiing	Cross Country Skiing Men's 30 kilometres	NA
6 Per Knut Aaland	M	33	188	7	United States	USA	1994 Winter	1994 \	Winter	Lillehammer	Cross Country Skiing	Cross Country Skiing Men's 10/15 kilometres	NA
6 Per Knut Aaland	M	33	188	7	United States	USA	1994 Winter	1994 \	Winter	Lillehammer	Cross Country Skiing	Cross Country Skiing Men's 4 x 10 kilometres	NA NA
7 John Aalberg	M	31	183	7	2 United States	USA	1992 Winter	1992 \	Winter	Albertville	Cross Country Skiing	Cross Country Skiing Men's 10 kilometres	NA
7 John Aalberg	M	31	183	7.	2 United States	USA	1992 Winter	1992 \	Vinter	Albertville	Cross Country Skiing	Cross Country Skiing Men's 50 kilometres	NA
7 John Aalberg	M	31	183	7	United States	USA	1992 Winter	1992 \	Vinter	Albertville	Cross Country Skiing	Cross Country Skiing Men's 10/15 kilometres	NA

Table is continuous as it has 271116 rows and 15 columns.

Olympics_History_noc_regions Table:

This table consists of 230 rows and 3 columns.

The columns are as follows:

- NOC- National Olympic Committee 3-letter code
- Region- Country
- Note- Comment

Olympics_History_noc_regions

NOC	region	notes
AFG	Afghanistan	
AHO	Curacao	Netherlands Antilles
ALB	Albania	
ALG	Algeria	
AND	Andorra	
ANG	Angola	
ANT	Antigua	Antigua and Barbuda
ANZ	Australia	Australasia
ARG	Argentina	
ARM	Armenia	
ARU	Aruba	
ASA	American Samoa	
AUS	Australia	
AUT	Austria	
AZE	Azerbaijan	
BAH	Bahamas	
BAN	Bangladesh	
BAR	Barbados	
	5325	

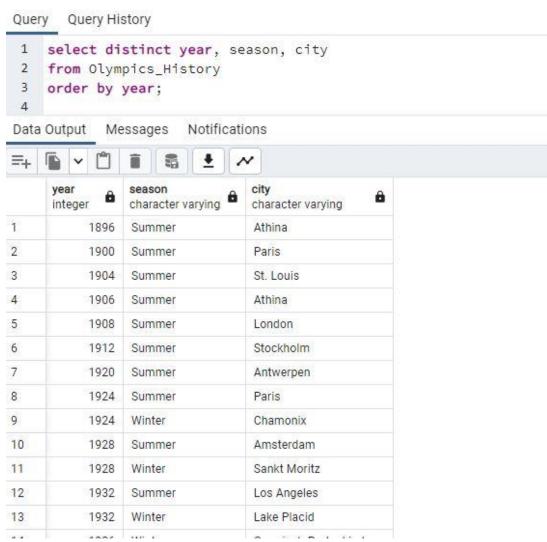
Table is continuous as it has 230 rows and 3 columns.

Case Study Questions

1. How many Olympics games have been held?



2. List down all Olympics games held so far with respect to the year, season, and city.?



3. What is the total number of nations who participated in each Olympics game?

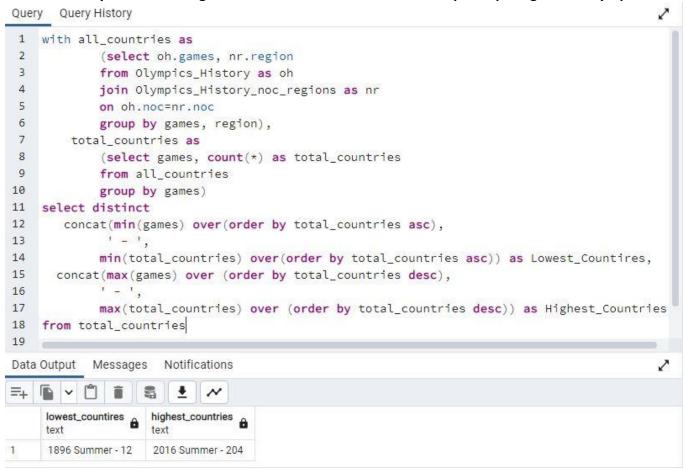


Output record is continuous.

Question 3 Alternative Solution

```
Query
      Query History
    with all_countries as
1
2
                (select oh.games, nr.region
3
                from Olympics_History as oh
4
                inner join Olympics_history_noc_regions as nr
5
                on oh.noc=nr.noc
                group by games, nr.region)
6
7
    select games, count(games) as total_countries
8
    from all_countries
9
    group by games
    order by games
10
11
```

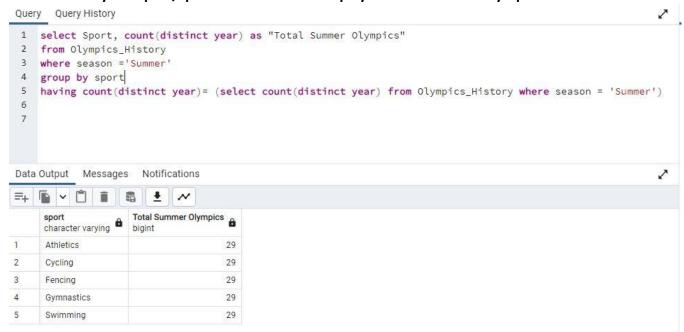
4. Which year saw the highest and lowest number of countries participating in the Olympics?



5. Which nation has participated in all Olympics Games?



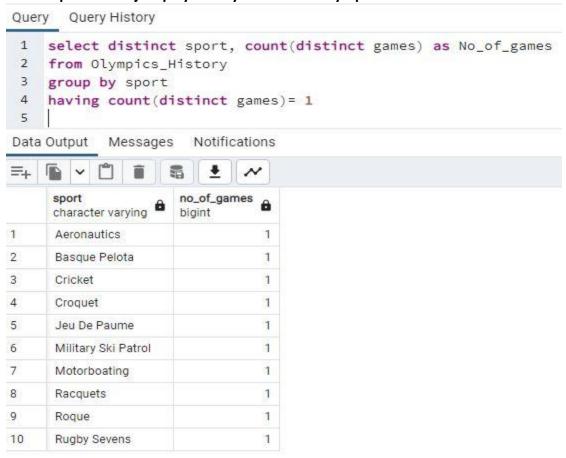
6. Identify the sport/sports which have been played in all summer Olympics?



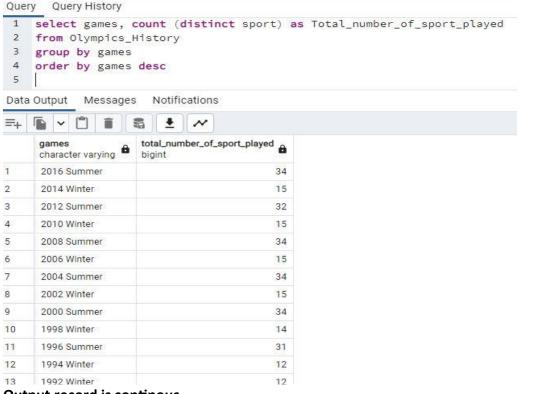
Question 6. Alternative Solution

```
Query
     Query History
   with t1 as
 2
        (select count( distinct games) as total_summer_games
 3
         from Olympics_history
 4
         where season = 'Summer' ),
 5
   t2 as
 6
        (select distinct sport, games
 7
           from Olympics_history
 8
            where season = 'Summer' order by games),
9
    t3 as
10
            (select sport, count(games) as no_of_games
11
            from t2
12
            group by sport)
13
    select *
   from t3
14
15
    join t1
16
    on t1.total_summer_games = t3.no_of_games;
17
```

7. Which Sports were just played only once in the Olympics.?



8. Fetch the total no of sports played in each Olympic game.?



Output record is continous

Question 8. Alternative Solution

Data Output Messages Notifications

M

Oscar Gomer Swahn

Query Query History

```
1
  with t1 as
2
           (select distinct games, sport
3
           from olympics_history),
4
5
           (select games, count (distinct sport) as no_of_sports
6
7
           group by games)
8
  select * from t2
9
   order by games desc;
```

9. Fetch oldest Athlete/Athletes to win a gold medal?

```
Query Query History
1 select name, sex, age, team, sport, medal
2 from Olympics_History
3 where medal ='Gold' and age <> 'NA' and age <> ' '
   order by age desc
   limit 2;
6
```

=+ medal age sport character varying character varying character varying character varying character varying character varying 1 Gold Charles Jacobus 64 United States Roque 2 Gold

Sweden

Shooting

64

Question 9. Alternative Solution

```
Query Query History
 1 with tl as
 2
            (select name, sex, cast(case when age = 'NA' then '0' else age end as integer) as age,team,
 3
           sport, event, medal
 4
           from Olympics_History),
 5
      ranking as
 6
           (select *,
 7
            rank() over (order by age desc ) as rnk
 8
            from t1
 9
           where medal = 'Gold')
10 select *
11 from ranking
12 where rnk =1;
13
```

10. What is the ratio of male and female athletes who has participated in all Olympic games?

```
Query Query History
   with t1 as
 1
 2
             (select sex, count(sex) as cnt
 3
             from Olympics_History
 4
             group by sex),
 5
        t2 as
 6
             (select *, row_number() over(order by cnt asc) as ranking
 7
              from t1),
 8
        min_cnt as
 9
             (select cnt from t2 where ranking =1),
10
        max cnt as
11
             (select cnt from t2 where ranking =2)
12
    select concat(' 1 : ', round(max_cnt.cnt::decimal/min_cnt.cnt, 2 )) as ratio
13
     from min_cnt, max_cnt;
14
Data Output Messages Notifications
=+
     ratio
            0
     text
1
     1:2.64
```

Question 10. Alternative Solution

Query Query History

```
1 select concat('1 :', round(max_cnt.cnt::decimal/min_cnt.cnt, 2)) as Ratio
2 from
 3
        (select cnt
         from
4
 5
             (select sex, count(sex) as cnt
 6
              from Olympics_History
7
              group by sex
8
              order by cnt asc) as t1
9
              limit 1) as min_cnt,
10
        (select cnt
11
        from
12
             (select sex, count(sex) as cnt
13
             from Olympics_History
14
             group by sex
15
             order by cnt desc) as t2
16
             limit 1) as max_cnt;
17
```

11. Fetch the top 5 Athletes who have won the most Gold Medal

Query Query History select name, gold_medals, rank 1 2 3 (select name, count(medal) as gold_medals, 4 dense_rank() over(order by count(medal) desc) as rank 5 from Olympics_History where medal= 'Gold' 6 7 group by name) as Ranked_Athletics where rank <=5 8 9 Data Output Notifications Messages gold_medals rank name â bigint bigint character varying 1 Michael Fred Phelps, II 23 1 2 Raymond Clarence "Ray" Ewry 10 2 3 Frederick Carlton "Carl" Lewis 9 3 4 Mark Andrew Spitz 9 3 5 Paavo Johannes Nurmi 9 3 6 Larysa Semenivna Latynina (Diriy-) 9 3 7 Matthew Nicholas "Matt" Biondi 8 4 8 Jennifer Elisabeth "Jenny" Thompson (-Cumpelik) 8 4 9 Usain St. Leo Bolt 8 4 Rirait Fischer-Schmidt 0 1

Question 11. Alternative Solution

```
Query Query History
1 with t1 as
2
       (select name, count(medal) as total_medals
3
      from Olympics_History
4
      where medal= 'Gold'
5
       group by name
6
       order by count(*) desc),
7 t2 as
8
       (select *, dense_rank() over(order by total_medals desc) as ranking
9
         from t1)
10 select *
11 from t2
12 where ranking <=5;
```

12. Fetch the top 5 athletes who have won the most medals (Gold/Silver/Bronze)

```
Query Query History
    select name, total_medals, ranking
2
 3
        (select name, count(medal) as total_medals,
 4
        dense_rank () over (order by count(medal) desc ) as ranking
 5
        from Olympics_history
        where medal in ('Gold', 'Silver', 'Bronze')
 6
7
         group by name
8
        ) as Ranked_Athletics
9 where ranking <=5;</pre>
10
Data Output Messages Notifications
```

	name character varying	total_medals bigint	ranking bigint	
1	Michael Fred Phelps, II	28	1	
2	Larysa Semenivna Latynina (Diriy-)	18	2	
3	Nikolay Yefimovich Andrianov	15	3	
4	Edoardo Mangiarotti	13	4	
5	Ole Einar Bjrndalen	13	4	
6	Borys Anfiyanovych Shakhlin	13	4	
7	Takashi Ono	13	4	
8	Natalie Anne Coughlin (-Hall)	12	5	
9	Rvan Steven Lochte	12	5	

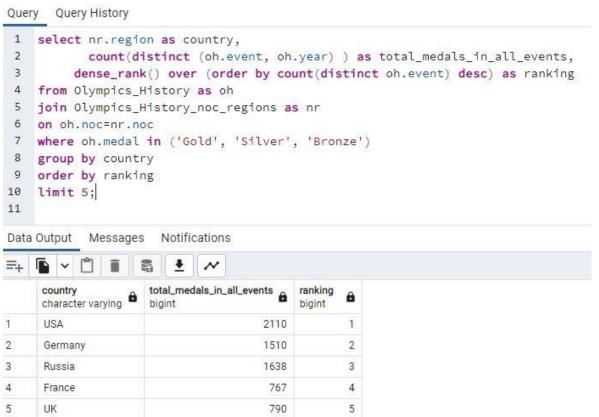
Output record is continuous.

= [6] (1 [6] = [8] 4

Question 12. Alternative Solution

```
Query
      Query History
1
    with t1 as
 2
              (select name, count(medal) as total_medals
 3
               from olympics_history
               where medal in ('Gold', 'Silver', 'Bronze')
 4
 5
               group by name, team
 6
               order by total_medals desc),
 7
          t2 as
 8
               (select *, dense_rank() over (order by total_medals desc) as ranking
 9
               from t1)
10
    select name, total_medals, ranking
11
    from t2
12
    where ranking <= 5;
13
```

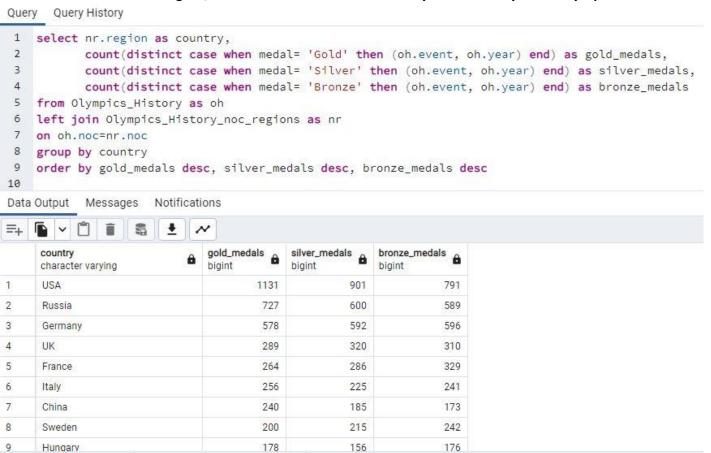
13. Fetch the top 5 most successful countries in the Olympics. Success is defined by number of medals won.



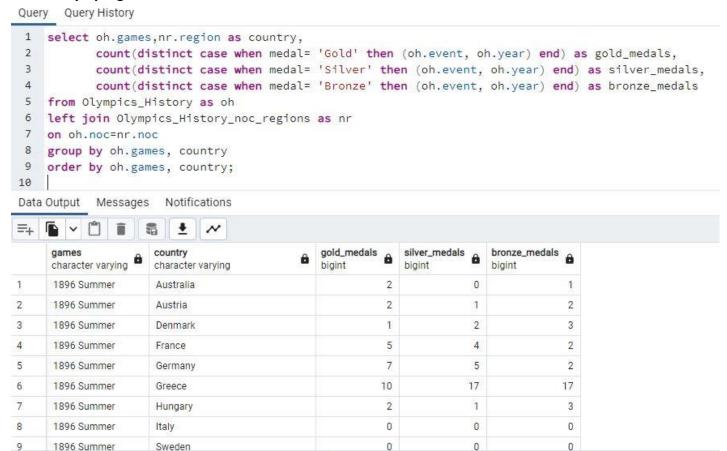
Question 13. Alternative Solution

```
Query History
    with t1 as
2
                (select nr.region, count(distinct (oh.event, oh.year) ) as total_medals_in_all_events
3
                from olympics_history oh
4
                join olympics_history_noc_regions nr on nr.noc = oh.noc
 5
                where medal <> 'NA'
 6
                group by nr.region
 7
                order by total_medals_in_all_events desc),
8
9
                (select *, dense_rank() over(order by total_medals_in_all_events desc) as rnk
10
                from t1)
11
   select *
12
   from t2
13
    where rnk <= 5;
14
```

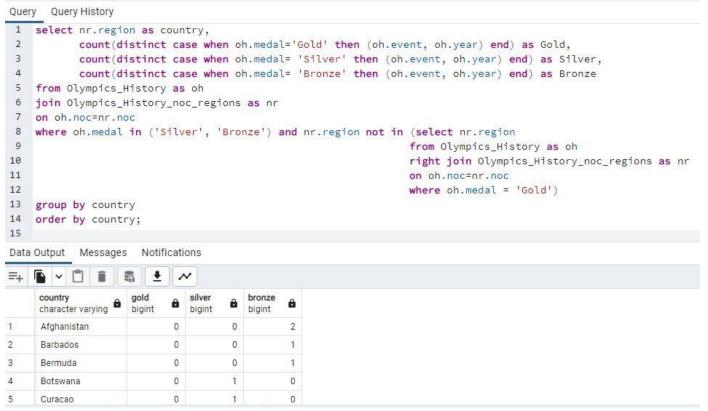
14. List down the total gold, silver and bronze medals won by each country in all Olympics.



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

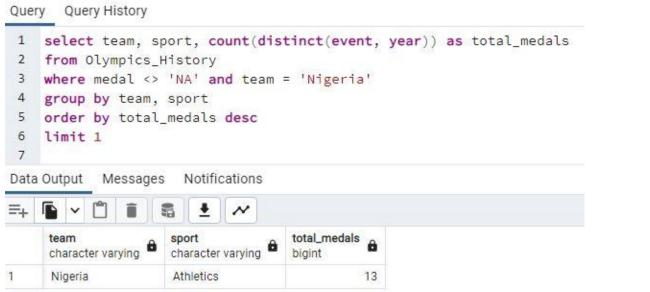


16. Which countries have never won gold medal but have won silver/bronze medals?



Output record is continuous.

17. In which Sport has team Nigeria won the highest medals?



Question 17. Alternative Solution

Query Query History

```
1 with t1 as
2
           (select team, sport, count(distinct(event, year)) as total_medals
3
           from olympics_history
          where medal <> 'NA'
4
           and team = 'Nigeria'
 5
6
           group by team, sport
7
           order by total_medals desc),
8
       t2 as
9
           (select *, rank() over(order by total_medals desc) as rnk
10
            from t1)
11 select team, sport, total_medals
12 from t2
13 where rnk = 1;
14
```

A few Insights and Findings

One interesting insight that was uncovered from analyzing the Olympic dataset is the remarkable dominance of the United States in the Olympic Games. Over the years, the United States has consistently emerged as a powerhouse in terms of medal count across various Olympic events.

By aggregating the medal data for all Olympic Games, it was found that the United States has secured the highest number of gold medals, silver medals, and bronze medals. This sustained success showcases the country's exceptional athletic prowess and commitment to sporting excellence.

This insight not only highlights the sporting achievements of the United States but also underscores the significance of investment in sports infrastructure, talent development programs, and overall national commitment to athletic excellence. It serves as a testament to the country's rich sporting culture and its ability to produce world-class athletes who consistently excel on the Olympic stage.

Furthermore, this insight encourages a deeper exploration of the factors contributing to the United States' success in the Olympics, including training methodologies, sports policies, and the overall sporting ecosystem. It serves as a source of inspiration for aspiring athletes and nations striving to emulate the United States' success in the pursuit of Olympic glory.

Over the years, there has been a noticeable increase in the number of nations participating in the Olympics. This demonstrates the growing global interest and inclusivity of the games.

The Olympics have featured a variety of sports throughout its history. Some sports, such as Aeronautics, Motorboating, Military Ski Patrol, Roque, Ruby Sevens etc. were only played once in the Olympics. This highlights the diverse range of sports that have been included in the games over time.

Nigeria has achieved notable success in the sport of Athletics in the Olympics. However, it does not necessarily mean that Nigeria has the highest overall medal count in the Olympics. The success in Athletics indicates a strong performance by Nigerian athletes in this particular sport compared to other sports. To improve overall success and increase the medal count in the Olympics, Nigeria should diversify its sports focus, invest in athlete development, leverage sports science, foster collaborations, and improve sports governance. These strategies will help identify talent across various sports, provide necessary resources and support, optimize training methodologies, and ensure effective coordination and accountability. By implementing these measures, Nigeria can aim for broader success in the Olympics and enhance its overall standing in international sporting events.

Athletics, Cycling, Fencing, Gymnastics, and Swimming have been played in every edition of the Olympic Games. These sports have maintained their presence and popularity, highlighting their enduring significance in the Olympics. These insights highlight the historical significance, global appeal, and enduring popularity of these sports in the Olympic Games, emphasizing their role in shaping the spirit and success of the Olympics.

Appendix

SQL is the technique that has been used to perform analysis on this Olympics dataset. SQL is used in many relational database management systems such as PostgreSQL, MySQL, Microsoft SQL Server etc. In this analysis I used PostgreSQL

The dataset used for the analysis was gotten from Kaggle. The dataset is a historical dataset on the modern Olympic Games, including all the games from Athens 1896 to Rio 2016.

Below is the link to the dataset:

https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results

Dataset was imported into PostgreSQL by the creating the two tables for the two dataset csv files, the two tables were called Olympics History and Olympics History noc regions

Below is the syntax for creating columns for the tables

Olympics_History Table

```
Query Query History
   Create table Olympics_History
1
2
    (
3
        id int,
4
        name varchar,
5
        sex varchar,
6
        age varchar,
7
        height varchar,
8
        weight varchar,
9
        team varchar.
10
        noc varchar,
11
        games varchar,
12
        year int,
13
        season varchar,
14
        city varchar,
15
        sport varchar,
16
        event varchar,
17
        medal varchar
18
    );
19
```

Olympics_History_noc_regions Table