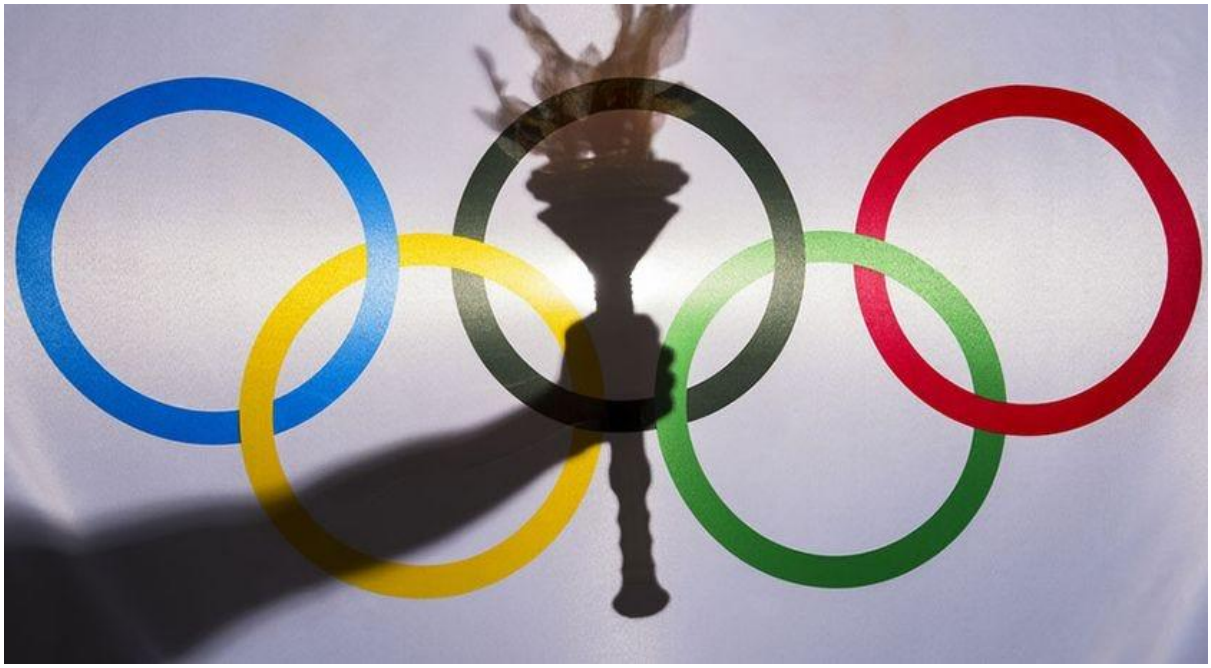
120 years of Olympic Dataset :- SQL Case Study

Context

This is a historical dataset on the modern Olympic Games, including all the Games from Athens 1896 to Rio 2016. The reference to this dataset is Kaggle.

Note that the Winter and Summer Games were held in the same year up until 1992. After that, they staggered them such that Winter Games occur on a four year cycle starting with 1994, then Summer in 1996, then Winter in 1998, and so on. A common mistake people make when analyzing this data is to assume that the Summer and Winter Games have always been staggered.



Content

The file athlete_events.csv contains 271116 rows and 15 columns. Each row corresponds to an individual athlete competing in an individual Olympic event (athlete-events). The columns are:

1. **ID** — Unique number for each athlete
2. **Name** — Athlete's name
3. **Sex** — M or F
4. **Age** — Integer
5. **Height** — In centimeters

6. **Weight** — In kilograms
7. **Team** — Team name
8. **NOC** — National Olympic Committee 3-letter code
9. **Games** — Year and season
10. **Year** — Integer
11. **Season** — Summer or Winter
12. **City** — Host city
13. **Sport** — Sport
14. **Event** — Event
15. **Medal** — Gold, Silver, Bronze, or NA

Inspiration

This dataset provides an opportunity to ask questions about how the Olympics have evolved over time, including questions about the participation and performance of women, different nations, and different sports and events.

Problem Statement

Through the data, the analysis for how many games were held in the past and who were the teams which won the highest number of medals, what exactly has been the ratio male and female participating in the different games.

Additionally, the data would also provide the insights over successful countries in olympics,medals won by each country corresponding to each olympic games,the oldest athletes to win a gold medal.

Q1. How many olympics games have been held ?

```
SELECT
  COUNT(DISTINCT Games) AS total_games
FROM
  `olympics_project.athlete_events`;
```

Based on the information provided in the original table, we can calculate the total distinct games, played in olympic.

JOB INFORMATION | RESULTS

Row	total_games
1	51

Q2. List down all Olympics games held so far ?

```
SELECT
  Year,
  Season,
  City
FROM
  `olympics_project.athlete_events`
ORDER BY
 1
```

JOB INFORMATION | RESULTS | CHART | JSON

Row	Year	Season	City
1	1896	Summer	Athina
2	1896	Summer	Athina
3	1896	Summer	Athina
4	1896	Summer	Athina
5	1896	Summer	Athina
6	1896	Summer	Athina
7	1896	Summer	Athina
8	1896	Summer	Athina
9	1896	Summer	Athina
10	1896	Summer	Athina

Q3. Mention the total no of nations who participated in each olympics game?

```
SELECT
  A.Games AS TOTAL_GAMES,
  COUNT(DISTINCT R.string_field_1) AS COUNTRIES
FROM
  `olympics_project.athlete_events` AS A
JOIN
  `olympics_project.noc_regions` AS R
ON
  A.NOC = R.string_field_0
GROUP BY
  1
ORDER BY
  1
```

JOB INFORMATION | RESULTS | CHART

Row	TOTAL_GAMES	COUNTRIES
1	1896 Summer	12
2	1900 Summer	31
3	1904 Summer	14
4	1906 Summer	20
5	1908 Summer	22
6	1912 Summer	29
7	1920 Summer	29
8	1924 Summer	45
9	1924 Winter	19
10	1928 Summer	46

Insights:- The data shows the total number of countries participated in olympics, we can observe that, the participation in the summer games are more than that of winters, also there has been a growth into the participation, over the years.

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

```
(SELECT
  A.Games AS TOTAL_GAMES,
  COUNT(DISTINCT R.string_field_0) AS COUNTRIES
FROM
  `olympics_project.athlete_events` AS A
JOIN
  `olympics_project.noc_regions` AS R
ON
  A.NOC = R.string_field_0
GROUP BY
  1
ORDER BY
  1
Limit 1)
union all
(SELECT
  A.Games AS TOTAL_GAMES,
  COUNT(DISTINCT R.string_field_1) AS COUNTRIES
FROM
  `olympics_project.athlete_events` AS A
JOIN
  `olympics_project.noc_regions` AS R
ON
  A.NOC = R.string_field_0
GROUP BY
  1
ORDER BY
  1 DESC
Limit 1)
```

JOB INFORMATION | RESULTS | CHART

Row	TOTAL_GAMES	COUNTRIES
1	2016 Summer	204
2	1896 Summer	12

Insights :- The lowest country count has been in the summer of 1896, and the highest has been in 2016. The data shows the growth in the participation, over the years.
The year with the lowest number of participating countries may suggest factors such as geopolitical tensions, economic challenges, or logistical issues that deterred countries from participating. Understanding the reasons behind low participation can provide insights into potential barriers to global cooperation or challenges facing the Olympic movement. Businesses may need to consider these factors when planning international marketing campaigns or partnerships.

Q5. Which nation has participated in all of the olympic games

```
SELECT
  R.string_field_1 AS COUNTRY_NAME,
  COUNT(DISTINCT A.Games) AS TOTAL_GAMES
FROM
  `olympics_project.athlete_events` AS A
JOIN
  `olympics_project.noc_regions` AS R
ON
  A.NOC = R.string_field_0
GROUP BY
  1
HAVING
  (TOTAL_GAMES) = 51
ORDER BY
  1
```

Row	COUNTRY_NAME	TOTAL_GAMES
1	France	51
2	Italy	51
3	Switzerland	51
4	UK	51

Insights: The countries that have been a part of every game are : France, Italy, Switzerland and UK.

Q6. Identify the sport which was played in all summer olympics ?

```
with CTE as

(SELECT
  COUNT(DISTINCT Games) AS total_games
FROM
  `olympics_project.athlete_events`
WHERE Season = 'Summer'),

CTE2 AS

(SELECT
  Sport,
  COUNT(DISTINCT Games) AS sport_count
FROM
  `olympics_project.athlete_events`
WHERE Season = 'Summer'
GROUP BY 1)

SELECT
*
FROM
   CTE AS C1
JOIN
  CTE2 AS C2
ON
  C1.total_games = C2.sport_count
```

Row	total_games	Sport	sport_count
1	29	Athletics	29
2	29	Fencing	29
3	29	Gymnastics	29
4	29	Swimming	29
5	29	Cycling	29

Insights: The games which were part of every olympics are given, these 5 games were played in all 29 olympics being held.

Q7. Which Sports were just played only once in the olympics ?

```
WITH t1 AS
(SELECT
  DISTINCT games,
  sport
FROM
  `olympics_project.athlete_events`),
t2 AS
(SELECT
  sport,
  COUNT(1) 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 1
```

Row	sport	no_of_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

Insights: These 10 games were played only once in the olympic, this reflects that, sports that are played infrequently in the Olympics presents an opportunity for businesses to explore niche markets, differentiate their offerings, and potentially tap into new revenue streams through strategic partnerships and investments.

Q8. Fetch the total no of sports played in each olympic games?

```
WITH t1 AS
  (SELECT DISTINCT
  games,
  sport
FROM
  `olympics_project.athlete_events`),
t2 AS
  (SELECT
  games,
  COUNT(1) AS no_of_sports
FROM
  t1
GROUP BY
  games)
SELECT
*
FROM t2
ORDER BY
  no_of_sports DESC;
```

Row	games	no_of_sports
1	2000 Summer	34
2	2004 Summer	34
3	2008 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

Insights: Through this data, we can fetch the trends in sports participation, the seasonal variation, as in summers the games are held more, this can be useful for businesses involved in seasonal markets or industries that cater to specific sports, understanding historical trends in sports participation, can help us to analyze, the market for future opportunities and challenges.

Q9. **Fetch oldest athletes to win a gold medal?**

```
WITH TEMP AS
  (SELECT
   name,
   sex,
   cast(case when age = 'NA' then '0' else age end as int) as age,
   team,
   games,
   city,
   sport,
   event,
   medal
from
  `olympics_project.athlete_events`),
ranking as
```

```
 (select *, rank() over(order by age desc) as rnk
from
 temp
where
 medal='Gold')
select
*
from
 ranking
where
 rnk = 1;
```

Row	name	sex	age	team	games	city	sport	event	medal	rnk
1	Oscar Gomer S...	M	64	Sweden	1912 Summer	Stockholm	Shooting	Shooting Men's Running Target,...	Gold	1
2	Charles Jacobus	M	64	United States	1904 Summer	St. Louis	Roque	Roque Men's Singles	Gold	1

Insights: The data reflects the oldest athletes to win a gold medal was of age 64, this can boost in market opportunities, recognizing the achievements of older athletes can highlight the market opportunities in industries related to sports, wellness, and leisure for older demographics. Businesses can develop products, services, and experiences that cater to the unique needs and interests of older athletes and sports enthusiasts.

Q10. Find the Ratio of male and female athletes participated in all olympic games?

```
with t1 as
 (select sex, count(1) as total_count
 from
 `olympics_project.athlete_events`
 group by 1),
t2 as
 (select sum(case when sex ="M" then total_count else 0 end )as male_count,
 sum(case when sex ="F" then total_count else 0 end)as female_count
 from t1
 )
 select
 male_count, female_count,
 round(male_count/female_count) as male_female_ratio
 from t2
```

Row	male_count	female_count	male_female_ratio
1	196594	74522	3.0

Insights: The data reflects that for every 1 female athlete, there are approximately 3 male athletes in the dataset, the insight from the male-to-female athlete ratio of approximately 3.0 suggests a significant gender disparity, particularly in the realm of sports.

This insight underscores the importance of initiatives and policies aimed at promoting gender inclusivity and equal opportunities in athletics.

Market Opportunities: The imbalance in male-to-female athlete representation presents opportunities for businesses and brands to target specific demographics with tailored products, services, and marketing campaigns. Recognizing and catering to the interests and needs of female athletes can help businesses tap into a growing market segment and promote gender diversity and inclusivity.

Q11. Fetch the top 5 athletes who have won the most gold medals ?

```
Select
  Name,
  Team,
  COUNT(Medal) AS cnt
FROM
  `olympics_project.athlete_events`
WHERE
  Medal = "Gold"
GROUP BY
  1,2
ORDER BY
  3 DESC
```

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS

Row	Name	Team	cnt
1	Michael Fred Phelps, II	United States	23
2	Raymond Clarence "Ray" Ewry	United States	10
3	Paavo Johannes Nurmi	Finland	9
4	Larysa Semenivna Latynina (Di...	Soviet Union	9
5	Mark Andrew Spitz	United States	9
6	Frederick Carlton "Carl" Lewis	United States	9
7	Sawao Kato	Japan	8
8	Ole Einar Bjrndalen	Norway	8
9	Usain St. Leo Bolt	Jamaica	8
10	Matthew Nicholas "Matt" Biondi	United States	8

Insights: The data reflects the top 5 athlete who won the most gold medals, Understanding the achievements of these athletes can provide insights into the factors contributing to their success, such as talent, training programs, and national support systems.

Businesses can leverage the popularity and recognition of top Olympic athletes to enhance brand visibility and credibility. Partnering with these athletes for endorsements, sponsorships, and marketing campaigns can help businesses reach a broader audience and enhance their brand image.

Q12. Fetch the top 5 athletes who have won the most medals (gold/silver/bronze) ?

```
with t1 as
 (select
   name,
   team,
   count(1) as total_medals
 from
   `olympics_project.athlete_events`
 where
   medal in ('Gold', 'Silver', 'Bronze')
 group by
   name, team
 order by
   total_medals desc),
t2 as
```

```
 (select *, dense_rank() over (order by total_medals desc) as rnk
 from t1)
select
name,
team,
total_medals
from t2
where rnk <= 5
```

Row	name	team	total_medals
1	Edoardo Mangiarotti	Italy	13
2	Ryan Steven Lochte	United States	12
3	Larysa Semenivna Latynina (Di...	Soviet Union	18
4	Paavo Johannes Nurmi	Finland	12
5	Jennifer Elisabeth "Jenny" Tho...	United States	12
6	Natalie Anne Coughlin (-Hall)	United States	12
7	Ole Einar Bjrndalen	Norway	13
8	Michael Fred Phelps, II	United States	28

Insights: The data reflects the top 5 athlete who won the most medals. Understanding the achievements of these athletes can provide insights into the factors contributing to their success, such as talent, training programs, and national support systems, as per the data, most of the athlete are from US.

Q13. Fetch the top 5 most successful countries in olympics. Success is defined by no of medals won?

```
select
 R.string_field_1 as COUNTRY_NAME,
 count(A.Medal) as TOTAL_COUNTS
from
 `olympics_project.athlete_events` as A
join
 `olympics_project.noc_regions` as R
on
 A.NOC = R.string_field_0
where
 A.Medal != "NA"
group by
```

```
 1
order by
  2 desc
limit 5;
```

Row	COUNTRY_NAME	TOTAL_COUNTS
1	USA	5637
2	Russia	3947
3	Germany	3756
4	UK	2068
5	France	1777

Insights: The data reflects medal counts which include gold, silver, and bronze medals won by each country across various Olympic games over the years. It's clear that the USA leads with the highest total count of Olympic medals, followed by Russia, Germany, the UK, and France respectively. This suggests that these countries have been consistently successful in Olympic competitions over the years.

Q14. List down total gold, silver and bronze medals won by each country corresponding to each olympic games ?

```
select
  R.string_field_1 as COUNTRY_NAME,
  A.Games,
  sum(case when A.Medal = "Gold" then 1 else 0 end ) as total_gold_medal,
  sum(case when A.Medal = "Silver" then 1 else 0 end ) as total_silver_medal,
  sum(case when A.Medal = "Bronze" then 1 else 0 end ) as total_bronze_medal
from
  `olympics_project.athlete_events` as A
join
  `olympics_project.noc_regions` as R
on
  A.NOC = R.string_field_0
where
  A.Medal != "NA"
group by
  1,2
order by
  3 desc, 4 desc,5 desc
```

Row	COUNTRY_NAME	Games	total_gold_medal	total_silver_medal	total_bronze_medal
1	Russia	1980 Summer	187	129	126
2	USA	1984 Summer	186	116	50
3	USA	1996 Summer	159	48	52
4	UK	1908 Summer	147	131	90
5	USA	2012 Summer	145	57	46
6	USA	2016 Summer	139	54	71
7	Russia	1988 Summer	134	67	99
8	USA	2000 Summer	130	61	51
9	USA	1904 Summer	128	141	125
10	USA	2008 Summer	127	110	80

Insights: The data reflects the country name along with the different medals, won in various games.

Q15. List down total gold, silver and bronze medals won by each country?

```
select
  R.string_field_1 as COUNTRY_NAME,
  sum(case when A.Medal = "Gold" then 1 else 0 end ) as total_gold_medal,
  sum(case when A.Medal = "Silver" then 1 else 0 end ) as total_silver_medal,
  sum(case when A.Medal = "Bronze" then 1 else 0 end ) as total_bronze_medal
from
  `olympics_project.athlete_events` as A
join
  `olympics_project.noc_regions` as R
on
  A.NOC = R.string_field_0
where
  A.Medal != "NA"
group by
  1
```

Row	COUNTRY_NAME	total_gold_medal	total_silver_medal	total_bronze_medal
1	Norway	378	361	294
2	Bahamas	14	11	15
3	UK	678	739	651
4	Sweden	479	522	535
5	Denmark	179	241	177
6	Belgium	98	197	173
7	Cuba	164	129	116
8	Fiji	13	0	0
9	Iran	18	21	29
10	Iraq	0	0	1

Insights: The data reflects query to list down the total gold, silver and bronze medals won by each country, where USA has the most number of medals, this suggests that they have been consistently successful in Olympic competitions over the years.

Q16. Which countries have never won gold medal but have won silver/bronze medals ?

```
with cte as(
select
  R.string_field_1 as country_name,
  sum(case when A.Medal = "Gold" then 1 else 0 end ) as total_gold_medal,
  sum(case when A.Medal = "Silver" then 1 else 0 end ) as total_silver_medal,
  sum(case when A.Medal = "Bronze" then 1 else 0 end ) as total_bronze_medal
from
  `olympics_project.athlete_events` as A
join
  `olympics_project.noc_regions` as R
on
  A.NOC = R.string_field_0
where
  A.Medal != "NA"
group by
  1
),
Gold_medal as(
  select
  country_name,
  total_gold_medal,
```

```
 total_silver_medal,
 total_bronze_medal
from
 cte
where
total_gold_medal = 0
),
silver_bronze_medal as(
 select
 country_name,
 total_gold_medal,
 total_silver_medal,
 total_bronze_medal
from
 Gold_medal
where
 total_silver_medal  >=1 or total_bronze_medal >=1
)
select
 country_name,
 total_gold_medal,
 total_silver_medal,
 total_bronze_medal
from
silver_bronze_medal
order by 3 desc, 4 desc
```

Row	country_name	total_gold_medal	total_silver_medal	total_bronze_medal
1	Paraguay	0	17	0
2	Iceland	0	15	2
3	Montenegro	0	14	0
4	Malaysia	0	11	5
5	Namibia	0	4	0
6	Philippines	0	3	7
7	Moldova	0	3	5
8	Lebanon	0	2	2
9	Tanzania	0	2	0
10	Sri Lanka	0	2	0

Insights: That data has aimed at analyzing Olympic data to find countries that have won either silver or bronze medals but have never won a gold medal.

Q17. In which Sport/event, India has won highest medals ?

```
with t1 as
(select
  sport,
  count(1) as total_medals
from
  `olympics_project.athlete_events`
where
  medal != 'NA' and team = 'India'
group by
  1
order by
  total_medals desc),
t2 as
(select *, rank() over(order by total_medals desc) as rnk
from t1)
select
  sport, total_medals
from
  t2
where
  rnk = 1
```

Row	sport	total_medals
1	Hockey	173

Insights: The provided data indicates that India has achieved the highest number of medals in the sport of hockey, with a total of 173 medals. This signifies India's remarkable success and dominance in field hockey at the Olympic Games.

Q18. Break down all olympic games where India won medal for Hockey and how many medals in each olympic games?

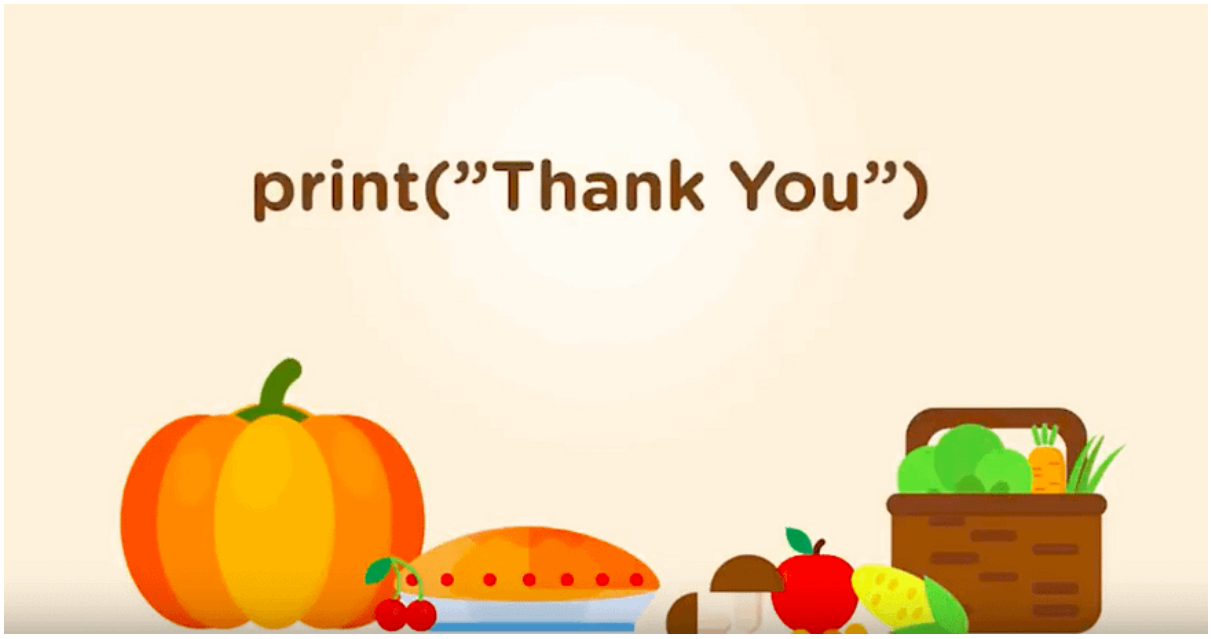
```
select
  team,
  sport,
  games,
  count(1) as total_medals
```

```
from
  `olympics_project.athlete_events`
where
  medal !='NA' and team = 'India' and sport = 'Hockey'
group by
  1,2,3
order by
  4 desc
```

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	team	sport	games	total_medals
1	India	Hockey	1948 Summer	20
2	India	Hockey	1936 Summer	19
3	India	Hockey	1956 Summer	17
4	India	Hockey	1968 Summer	16
5	India	Hockey	1980 Summer	16
6	India	Hockey	1932 Summer	15
7	India	Hockey	1964 Summer	15
8	India	Hockey	1928 Summer	14
9	India	Hockey	1952 Summer	14
10	India	Hockey	1972 Summer	14

Insights: The data reflects to represent India's performance in the sport of hockey across different Summer Olympic Games, including the number of medals won in each respective year, the insights that can be derived from this data would be consistent presence in Hockey and dominance in the 20th Century, while India's performance in hockey has seen fluctuations in more recent years, with fewer gold medals won compared to the earlier decades, the country has remained competitive and continues to be a strong contender in international hockey tournaments


print(”Thank You”)