

Analysing 120 Years Olympics Dataset

Olympic games are an event where athletes from all over the world participate to compete with each other. In this paper, I have tried to study the data of Olympic games from the year 1896 – 2016.

Empty Records

```
-- Checking count of empty records in important columns
SELECT
COUNT(*) AS total_participants_records,
SUM(CASE WHEN sex IS NULL THEN 1 ELSE 0 END) AS unspecified_gender,
SUM(CASE WHEN age IS NULL THEN 1 ELSE 0 END) AS unspecified_age,
SUM(
CASE WHEN height IS NULL THEN 1 ELSE 0 END
) AS unspecified_height,
SUM(
CASE WHEN weight IS NULL THEN 1 ELSE 0 END
) AS unspecified_weight
FROM
gondal.athlete_events
```

Records with Missing values

total_participants_records	unspecified_gender	unspecified_age	unspecified_height	unspecified_weight
271116	0	9474	60171	62875

« < Page 1 of 1 > » Showing rows 1 of 1

Exploring Event Records

```
SELECT
id,
name,
sex,
age,
height,
weight,
team,
games,
year,
season,
city,
sport,
event,
region,
medal
FROM
```

Combined Both Tables

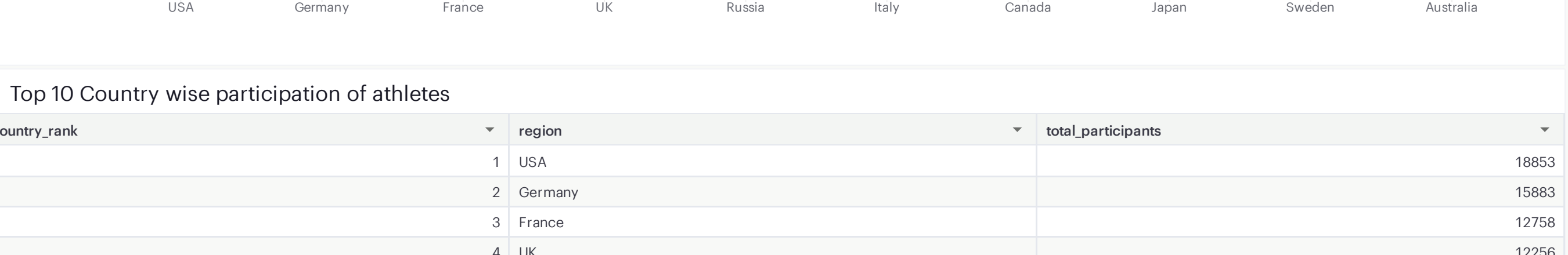
id	name	sex	age	height	weight	team	games	year	season
1	A Djang	M	24	180	80	China	1992 Summer	1992	Summer
2	A Lamusi	M	23	170	60	China	2012 Summer	2012	Summer
3	Gunnar Nielsen Aaby	M	24			Denmark	1920 Summer	1920	Summer
4	Edgar Lindensau Aab...	M	34			Denmark/Sweden	1900 Summer	1900	Summer
5	Christine Jacobsa Aaf...	F	21	185	82	Netherlands	1968 Winter	1988	Winter
5	Christine Jacobsa Aaf...	F	21	185	82	Netherlands	1968 Winter	1988	Winter
5	Christine Jacobsa Aaf...	F	25	185	82	Netherlands	1992 Winter	1992	Winter
5	Christine Jacobsa Aaf...	F	25	185	82	Netherlands	1992 Winter	1992	Winter
5	Christine Jacobsa Aaf...	F	27	185	82	Netherlands	1994 Winter	1994	Winter
5	Christine Jacobsa Aaf...	F	27	185	82	Netherlands	1994 Winter	1994	Winter

« < Page 1 of 1 > » Showing rows 1 of 10

Country wise participation

```
SELECT
RANK() OVER (
CASE BY
COUNT(ID) DESC
) AS country_rank,
region,
COUNT(ID) AS total_participants
FROM
```

Top 10 Country wise participation of athletes



Top 10 Country wise participation of athletes

country_rank	region	total_participants
1	USA	18853
2	Germany	15883
3	France	12758
4	UK	12256
5	Russia	11692
6	Italy	10715
7	Canada	9734
8	Japan	8444
9	Sweden	8339
10	Australia	7724

« < Page 1 of 1 > » Showing rows 1 of 10

Gender Distribution of Male and Female athletes

```
SELECT
sex,
COUNT(ID) AS participants
FROM
```

Gender Distribution of Male and Female athletes



Gender Distribution of Male and Female athletes

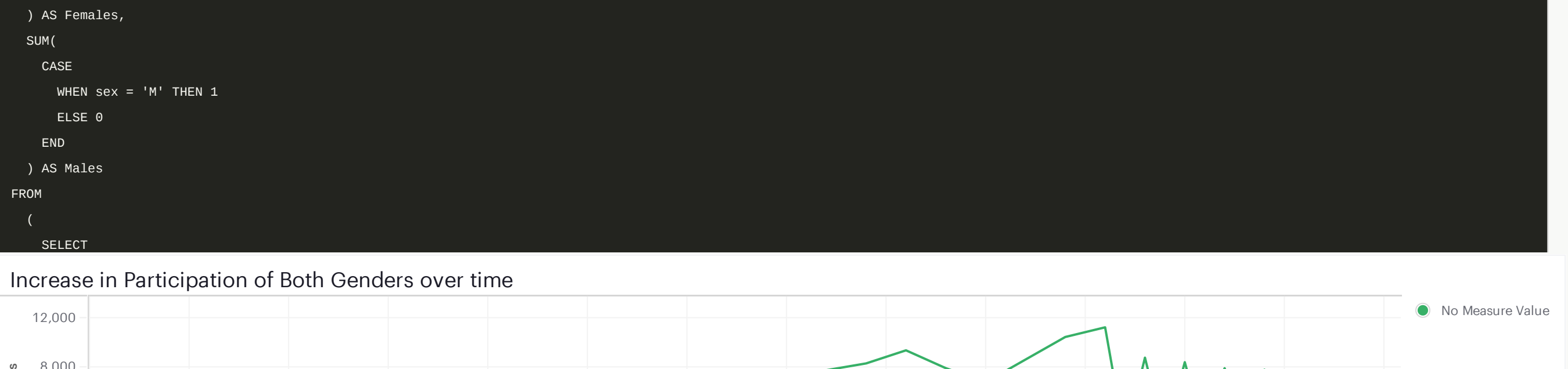
sex	participants
F	74393
M	196374

« < Page 1 of 1 > » Showing rows 1 of 2

Increase in participation of both genders over the time

```
SELECT
year,
SUM(
CASE
WHEN sex = 'F' THEN 1
ELSE 0
END
) AS Females,
SUM(
CASE
WHEN sex = 'M' THEN 1
ELSE 0
END
) AS Males
FROM
```

Increase in Participation of Both Genders over time



Increase in Participation of Both Genders over time

year	females	males
1896	0	380
1900	0	1903
1904	0	1285
1908	0	1722
1912	0	3054
1916	0	3953
1920	0	4158
1924	0	5432
1928	0	5137
1932	0	2952
1936	0	6852

« < Page 1 of 1 > » Showing rows 1 of 35

Female participants in both Seasons

```
SELECT
year,
season,
SUM(CASE WHEN sex = 'F' THEN 1 ELSE 0 END) AS participants
FROM
```

Overall Female Participation in Both Seasons

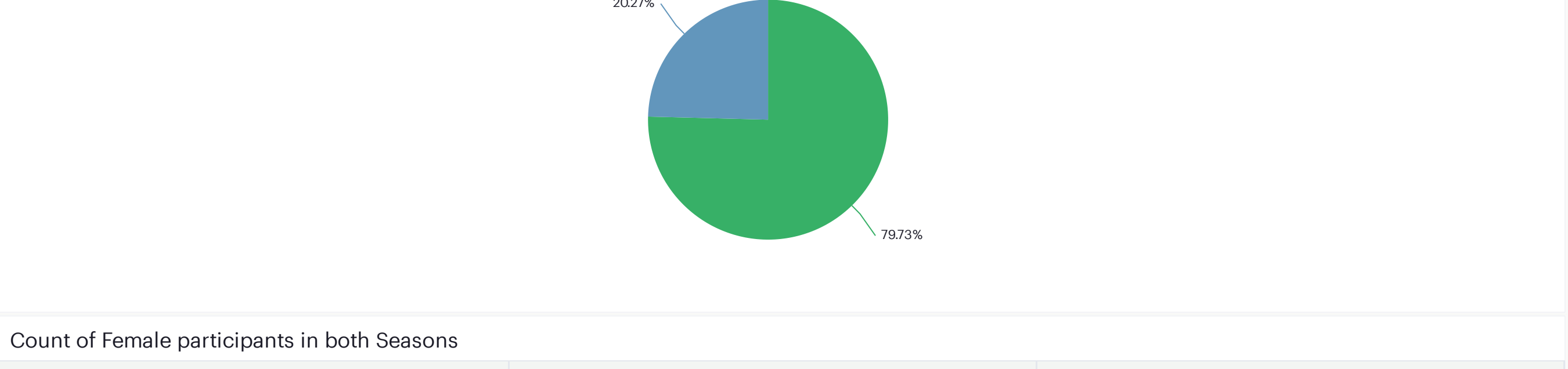


Count of Female participants in both Seasons

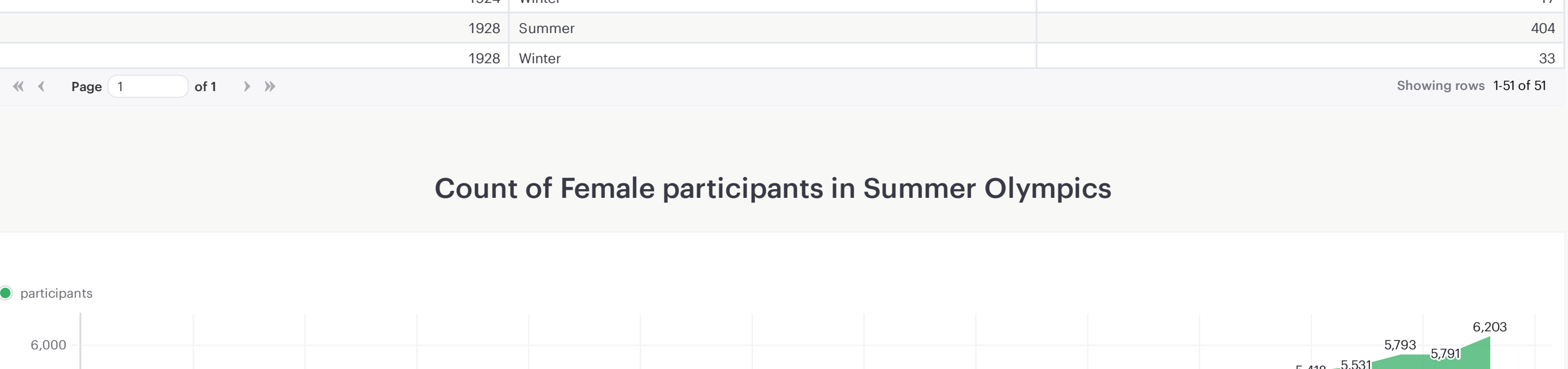
year	season	participants
1896	Summer	0
1900	Summer	33
1904	Summer	16
1908	Summer	11
1912	Summer	47
1916	Summer	87
1920	Summer	134
1924	Summer	244
1928	Summer	404
1932	Summer	369
1936	Summer	549
1940	Summer	626
1944	Summer	1003
1948	Summer	1435
1952	Summer	1360
1956	Summer	2290
1960	Summer	2170
1964	Summer	2445
1968	Summer	3543
1972	Summer	4305
1976	Summer	4990
1980	Summer	5418
1984	Summer	5531
1988	Summer	5793
1992	Summer	5757
1996	Summer	6203

« < Page 1 of 1 > » Showing rows 1 of 31

Count of Female participants in Summer Olympics



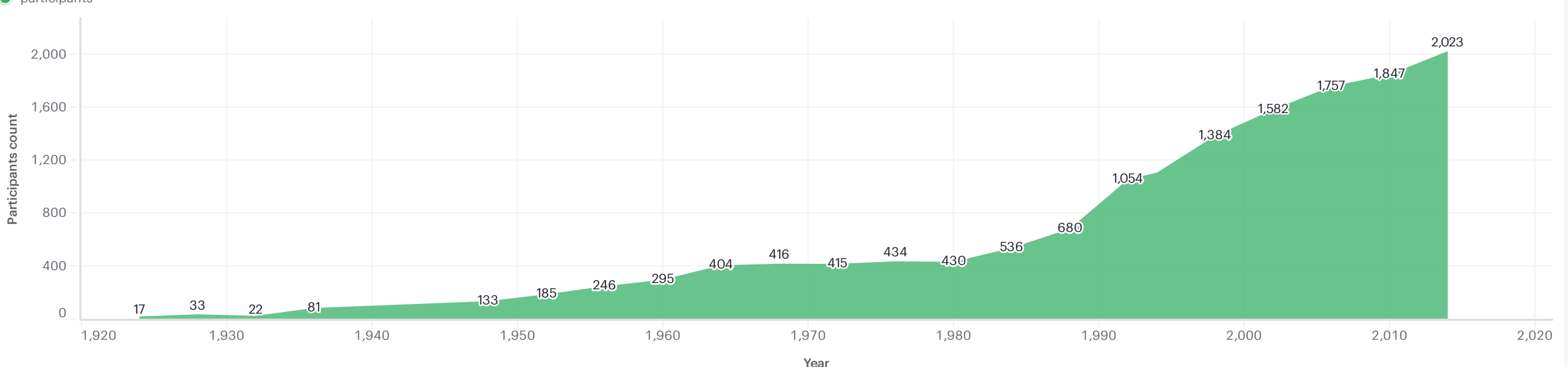
Count of Female participants in Winter Olympics



Top 10 Countries to win Medals

```
WITH medals_country AS (
SELECT
id,
name,
sex,
age,
height,
weight,
team,
games,
year,
season,
city,
sport,
event,
region,
medal
FROM
```

Top 10 Countries to win Medals

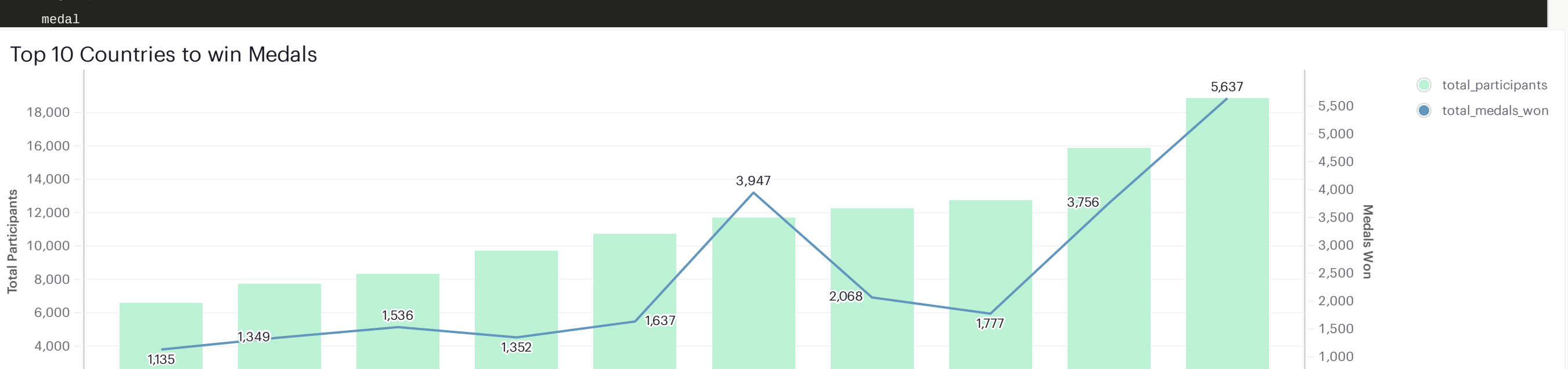


« < Page 1 of 1 > » Showing rows 1 of 10

Winner count - both Genders

```
WITH medals_gender AS (
SELECT
id,
name,
sex,
age,
height,
weight,
team,
games,
year,
season,
city,
sport,
event,
region,
medal
FROM
```

Participants vs Winners - Both Genders



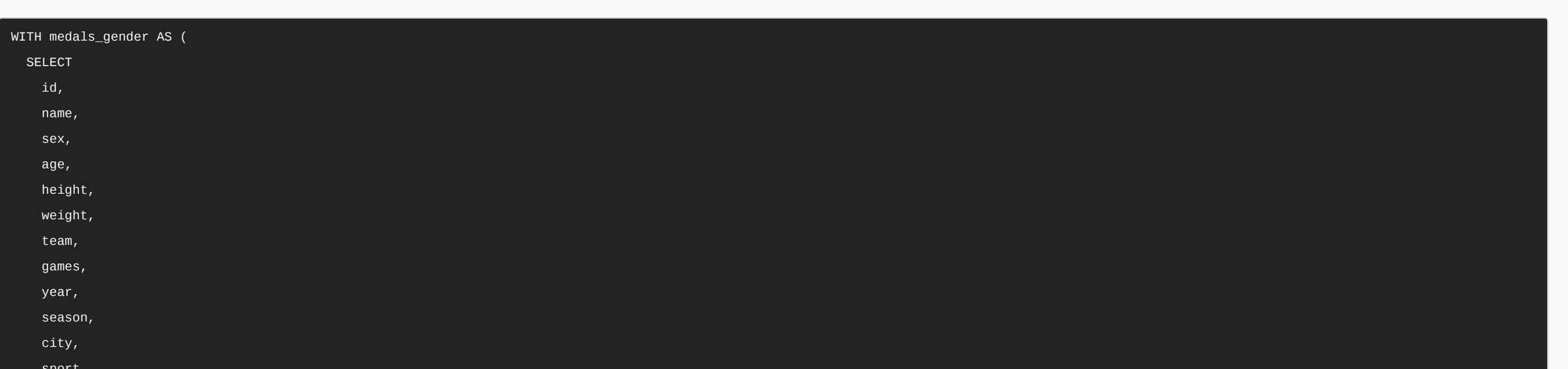
Participants vs Winners - Both Genders

sex	total_participants	total_medals_won	gold_medals	silver_medals	bronze_medals
F	74393	11246	171	3747	3732
M	196374	28528	624	9624	9380

« < Page 1 of 1 > » Showing rows 1 of 2

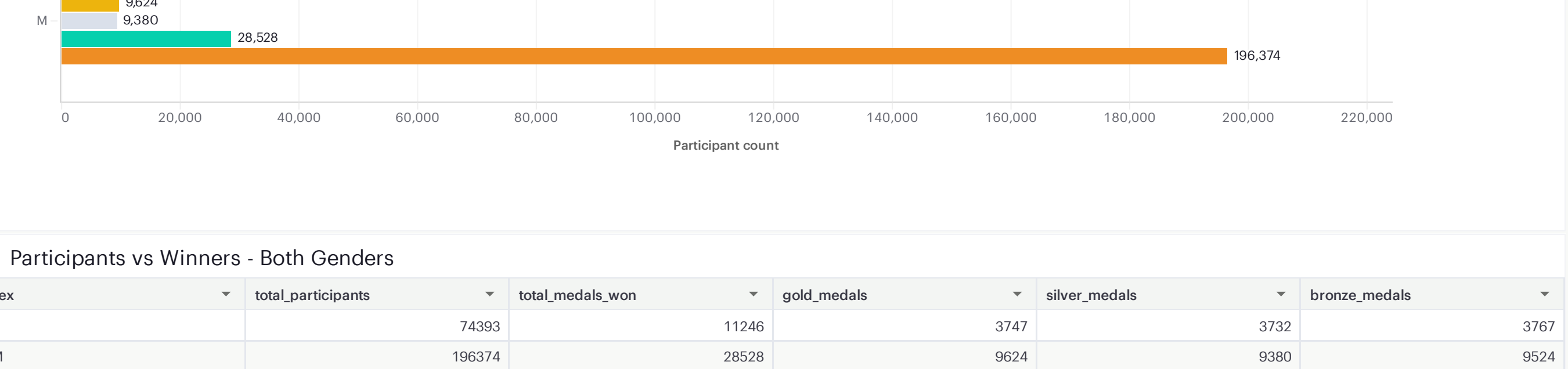
Height vs Weight of winners

Height vs Weight of Winners



Average Physical stats of both genders (kg and cm)

Average physical stats of both genders (winners)



« < Page 1 of 1 > » Showing rows 1 of 2

Female participants - winners vs non-winners stats

Females: Winners vs Non-winners stats

medal	medal_count	average_height	average_weight	average_age
Gold	3747	171	63	24
Silver	3732	171	63	24
Bronze	3767	170	63	25
NA	0	167	59	24

« < Page 1 of 1 > » Showing rows 1 of 4

Male participants - winners vs non-winners stats

Males: Winners vs Non-winners stats

medal	medal_count	average_height	average_weight	average_age
Gold	9624	181	80	27
Silver	9380	181	79	27
Bronze	9524	181	79	26
NA	0	178	75	26

« < Page 1 of 1 > » Showing rows 1 of 4