MY SQL Project

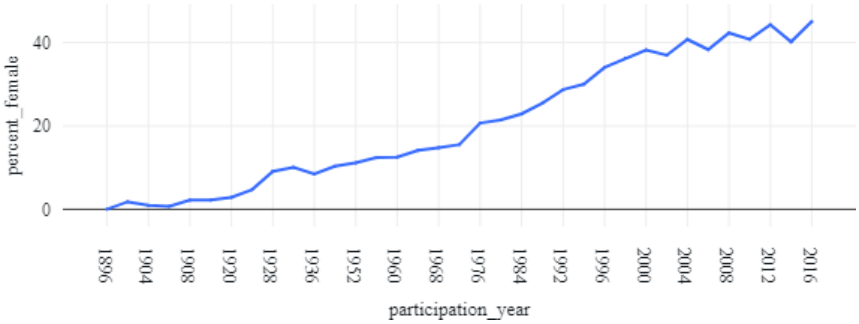
Introduction:

I embarked on an exploration of the Olympic participation history dataset from 1896 to 2016 to uncover evolving trends in Olympic games, with the aim of addressing trainers and coaches. My analysis focused on two features from the dataset:

1. Influence of Gender in Olympics
2. Influence of participants age in Olympics

I began with the hypothesis that sports are predominantly male- dominated, leading me to believe that trainers and coaches should prioritize the gender with the highest participation rate. Additionally, I initially subscribed to the notion that 'the early bird catches the worm,' assuming trainers would focus on young athletes for training and coaching purposes. However, deep diving into the dataset challenged these assumptions, leading me to realize their inaccuracy.

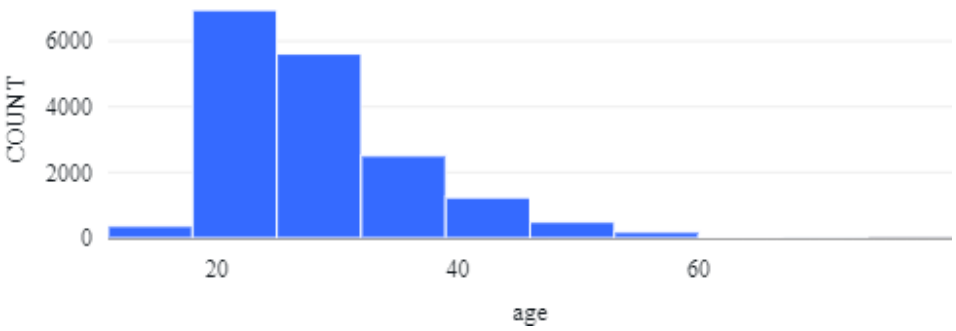
After sanitizing the data, my initial endeavour was to understand how female participation has evolved over the years, and I made an astounding discovery-



45% of participants are Females.

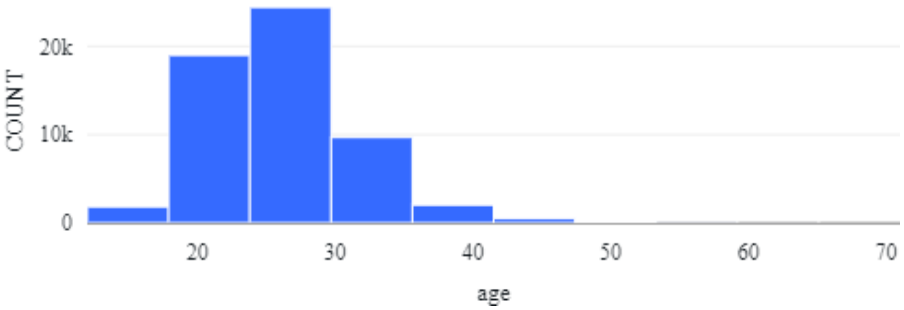
The analysis revealed a significant increase in female participation, with females comprising 45% of total participants in the 2016 Olympics. This realization underscored the potential female participants hold for trainers and coaches as valuable clients.

In my second attempt, I sought to understand the role age plays in Olympic participation. I observed substantial changes over time, particularly noting that during the early days of modern Olympics, participants were predominantly young, with an average age between 18 and 20. This trend is evident from the age distribution histogram plotted below.



Avg. age of participants in the early 20th century.

Over the next 100 years, there has been a drastic change in this pattern, with the current average age of participants ranging between 24 and 26. This shift is clearly depicted in the age distribution histogram of Olympic participants in the 21st Century.



Avg. age of participants in the early 21st Century.

Diving Deep

This prompted me to delve deeper into the data to investigate if there is any correlation between participants' age and the likelihood of winning an Olympic medal. For this analysis, I focused on the data from the Summer Olympics in 2000, 2004, 2008, 2012, and 2016. To conduct this study, I devised two new metrics:

1. average_participation_age
2. average_medalwinning_age

I created this metric for all 5 Olympic events in the 21st century and then made a Union of all these 5 tables to make a master table which is given below.

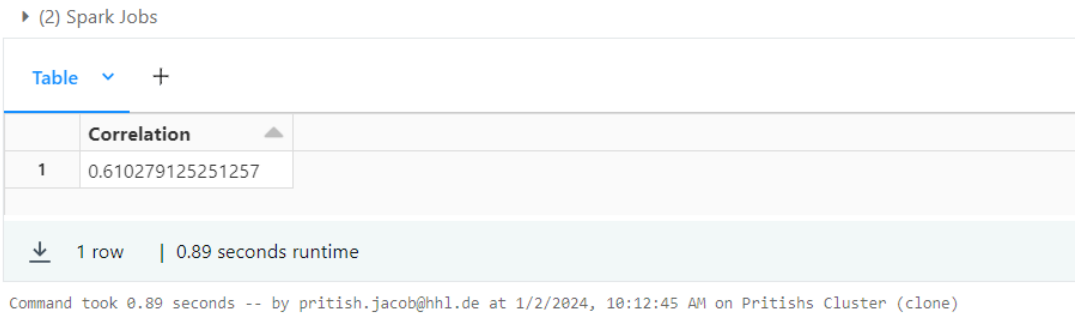
	year	avg_participation_age	avg_medal_age
1	2016	26.21	26.33
2	2012	25.96	26.07
3	2008	25.73	26.05
4	2004	25.64	26.15
5	2000	25.42	26.11

These findings unveiled two intriguing insights:

1. The average participation age shows a consistent upward trend, increasing from 25.4 in 2000 to 26.2 in 2016.
2. The average age of winning a medal consistently surpasses the average age of participation.

Further exploration into the correlation between these variables revealed a correlation coefficient of 0.61, indicating a mild correlation between them.

```
select
corr(rounded_decimal.avg_participation_age,rounded_decimal.avg_medal_age ) as Correlation
from
(select
year,
round(average_participation_age,2) as avg_participation_age,
round(average_medal_age,2) as avg_medal_age
from default.statistics_21st_century) rounded_decimal
```



Conclusion

Data analysis has been a humbling experience, revealing the limitations of the human mind in comprehending vast amounts of data. I began with the assumption that sports were predominantly a 'Men's Game,' only to have that notion challenged by the data. Additionally, the data illustrated an increase in the average age of participation over time.

However, it's crucial to delve deeper into the distribution of age in both participation and medal-winning for individual events to formulate event-based training and coaching strategies effectively. Despite this limitation, I remain hopeful that a century from now, individuals in their 30s, regardless of gender, will aspire to become Olympic participants.

My code is published in databricks and will be available till 5th Nov 2024.

https://databricks-prod-cloudfront.cloud.databricks.com/public/4027ec902e239c93eaaa8714f173bcfc/7775459193810626/545963420449008/1812343117218619/latest.html