

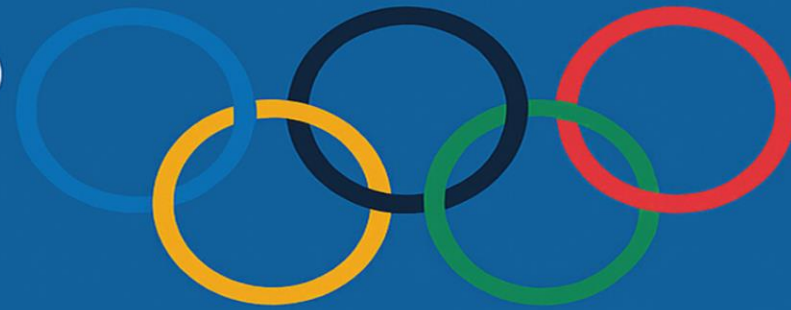
Medals Over Time: An Olympic Data Analysis

SQL For Data Science Capstone Project (Milestone 1)

Name: Kalyani Khond

Client: SportsStats

(Olympics Dataset – 120 years of data)



SportsStats is a sports analysis firm partnering with local news and elite personal trainers to provide “interesting” insights to help their partners.

➤ Project Proposal

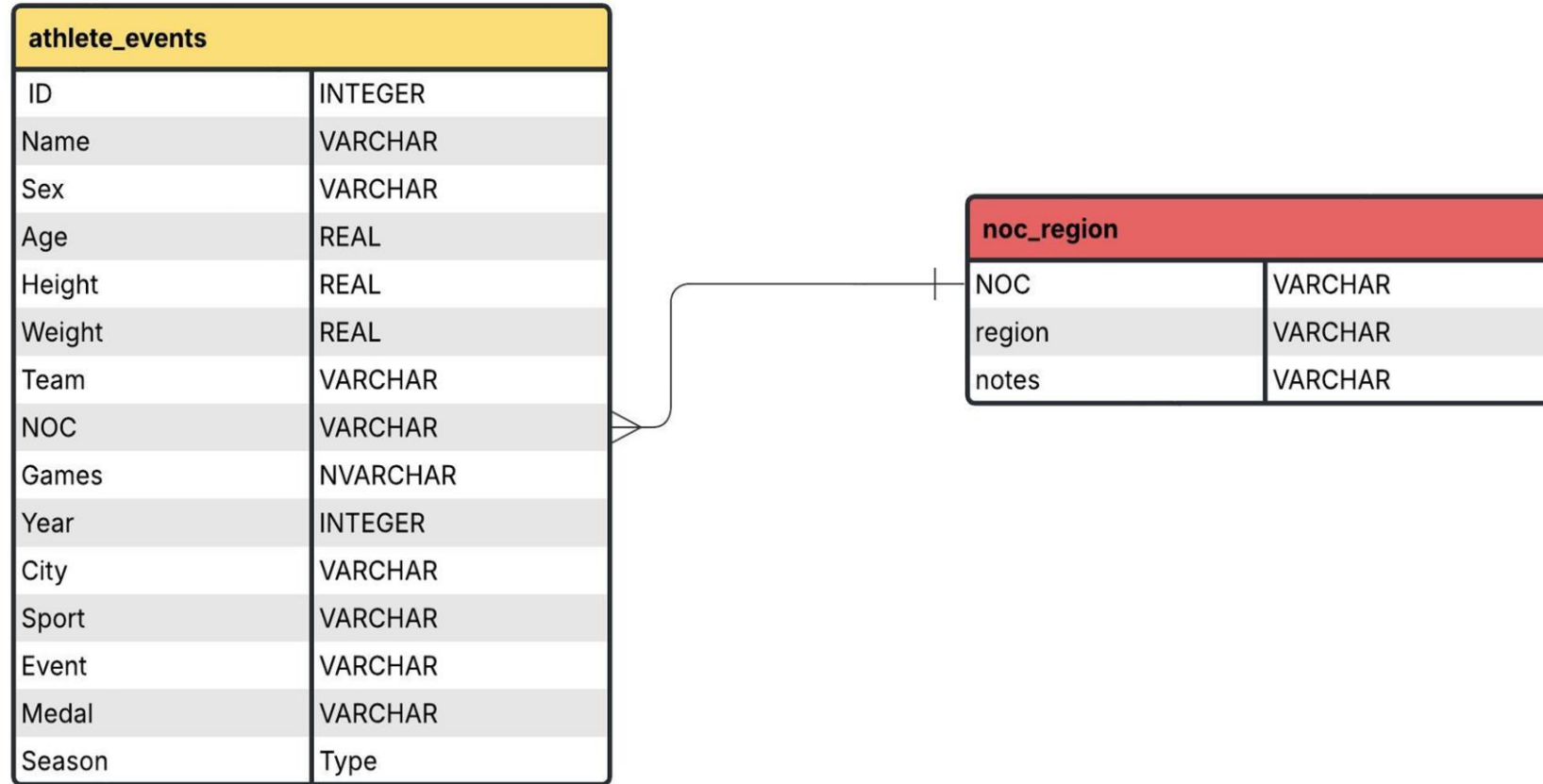
The Olympic Games feature a wide range of sports, from traditional events like athletics and swimming to newer additions such as skateboarding and surfing.

Over time, new sports have been added, and more countries have taken part. This means the number of athletes, events, and medals has grown a lot. The dataset shows interest in the Olympics over 120 years and highlights changes in which countries perform best at the Olympics. It helps us explore how athlete demographics, like age and gender, have shifted over the years.

This Olympic dataset can be useful for sports analysts, researchers and Sports Federations. It helps study how countries have performed over time and how participation has changed. National sports bodies might use it to plan training or understand medal trends. Students and fans can also explore it to learn more about the history of the Games.

In this project, the dataset was used as-is without any cleaning. The analysis was based on the raw dataset to reflect the original Olympic records and preserve the integrity of the data.

➤ Entity Relationship Diagram

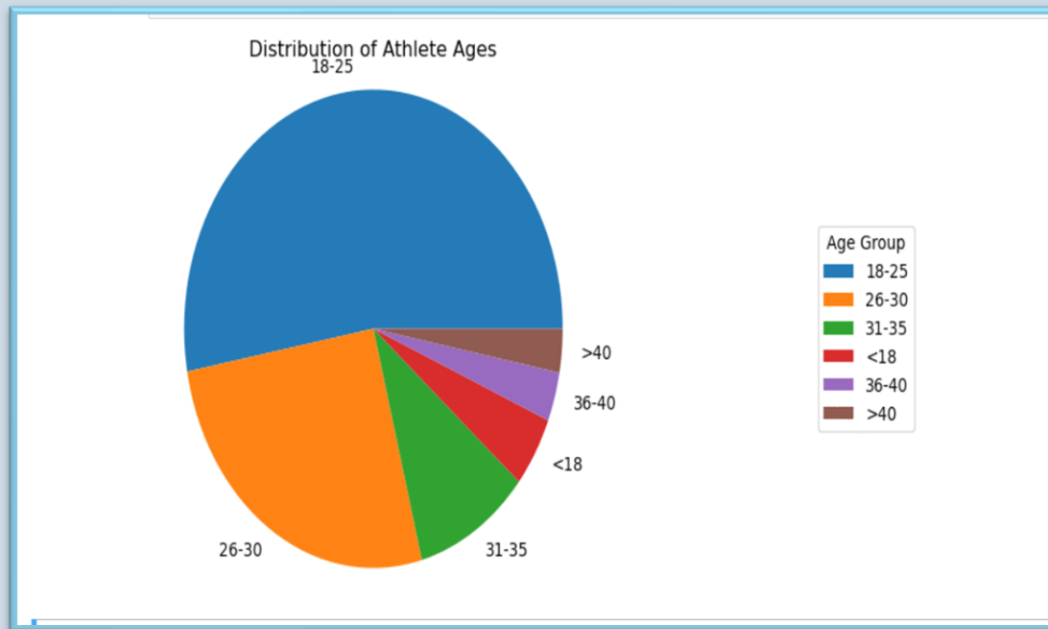


➤ Initial Data Exploration

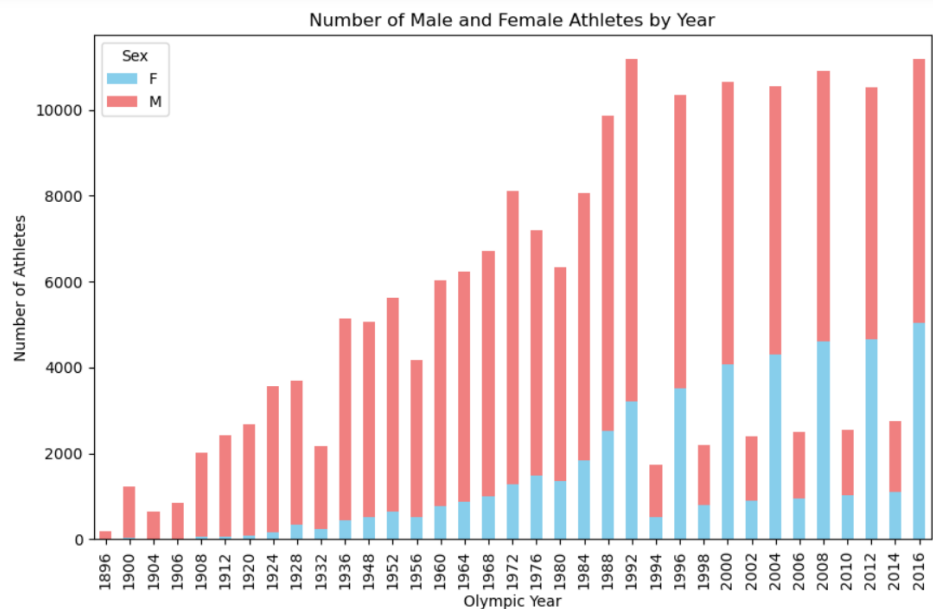
The dataset was loaded using pandas and queried with pandasql. Basic functions like `.info()` and `.describe()` were used to understand the structure.

A histogram showed the number of male and female athletes over the years, while a pie chart displayed the age distribution of participants.

This gave a simple overview of the data.



Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	None
A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	None
Gunnar Nielsen Aaby	M	24.0	NaN	NaN	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Football	None
Edgar Lindenu Aabye	M	34.0	NaN	NaN	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's Tug-Of-War	Gold
Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 500 metres	None
Christine Jacoba Aaftink	F	21.0	185.0	82.0	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Women's 1,000 metres	None
Christine Jacoba Aaftink	F	25.0	185.0	82.0	Netherlands	NED	1992 Winter	1992	Winter	Albertville	Speed Skating	Speed Skating Women's 500 metres	None
Christine Jacoba Aaftink	F	25.0	185.0	82.0	Netherlands	NED	1992 Winter	1992	Winter	Albertville	Speed Skating	Speed Skating Women's 1,000 metres	None
Christine Jacoba Aaftink	F	27.0	185.0	82.0	Netherlands	NED	1994 Winter	1994	Winter	Lillehammer	Speed Skating	Speed Skating Women's 500 metres	None



Columns: ['ID', 'Name', 'Sex', 'Age', 'Height', 'Weight', 'City', 'Sport', 'Event', 'Medal']

6]:

	ID	Age	Height	Weight
count	271116.000000	261642.000000	210945.000000	208241.000000
mean	68248.954396	25.556898	175.338970	70.702393
std	39022.286345	6.393561	10.518462	14.348020
min	1.000000	10.000000	127.000000	25.000000
25%	34643.000000	21.000000	168.000000	60.000000
50%	68205.000000	24.000000	175.000000	70.000000
75%	102097.250000	28.000000	183.000000	79.000000
max	135571.000000	97.000000	226.000000	214.000000

In [19]: `pysql("SELECT Team, COUNT(Medal) AS Total_Medals_in_120_Y FROM df_athletes")`

Out[19]:

	Team	Total_Medals_in_120_Y
0	United States	5219
1	Soviet Union	2451
2	Germany	1984
3	Great Britain	1673
4	France	1550
...
493	Brynild-2	1
494	Botswana	1
495	Bonaparte	1
496	Bermuda	1
497	Barbados	1

➤ Questions

1. How has the ratio of male and female participants changed over the years?
2. How does the age of an athlete affect his/her performance in the competitions?
3. What is the relationship between height of participants and the sport they play?
4. Is the performance of a nation in the Summer Olympics correlated to that in the Winter Olympics?
5. Do certain sports have a higher proportion of young or older medalists?

➤ Hypothesis

1. The ratio of female to male participants in the Olympics has steadily increased over time.
2. Athletes performance declines with age. Athletes physical strength and agility generally decreases as they grow older.
3. There is a relationship between an athlete's height and the sport they participate in, suggesting that certain sports tend to attract athletes with specific physical characteristics
4. A nation's performance in the Summer Olympics is correlated with its performance in the Winter Olympics.
5. There is a variation in the age of medalists across different sports, indicating that some sports are dominated by younger athletes while others see greater success among older participants.

➤ Approach

- Depending on the question, the data can then be grouped or aggregated using columns such as Year, Sport, or NOC.
- Specific filters can be applied, such as including only medal winners or focusing on a particular Olympic season.
- Calculations like average age, participant count, or total medals are then performed to uncover trends.
- Finally, the results are visualized using charts such as bar plots, pie charts, or line graphs to better interpret the findings.
- The key columns to look for the analysis include Year, Sex, Age, Height, Sport, Medal, Season, and NOC.