Olympic Athletes Through the Ages: A Data Analysis

Keon Augustine•04-24-2025

Overview

Introduction

This dataset contains information on over 270,000 Olympic athletes from 1896 to 2016. Key features include athlete age, gender, height, weight, team, sport, and medal status. The analysis explores demographic and performance trends across this period.

Objective & Goal

The goal is to understand physical and demographic patterns among Olympic athletes and identify correlations between attributes like height, weight, and medal performance.

```
df.isnull().sum()
```

```
ID              0
Name            0
Sex             0
Age          9474
Height      60171
Weight      62875
Team            0
NOC             0
Games           0
Year            0
Season          0
City            0
Sport           0
Event           0
Medal      231333
dtype: int64
```

Cleaning the Data

Significant missing data was found in Age, Height, and Weight. These records were dropped to ensure analytical accuracy, resulting in a cleaned file: athlete_events_cleaned.csv

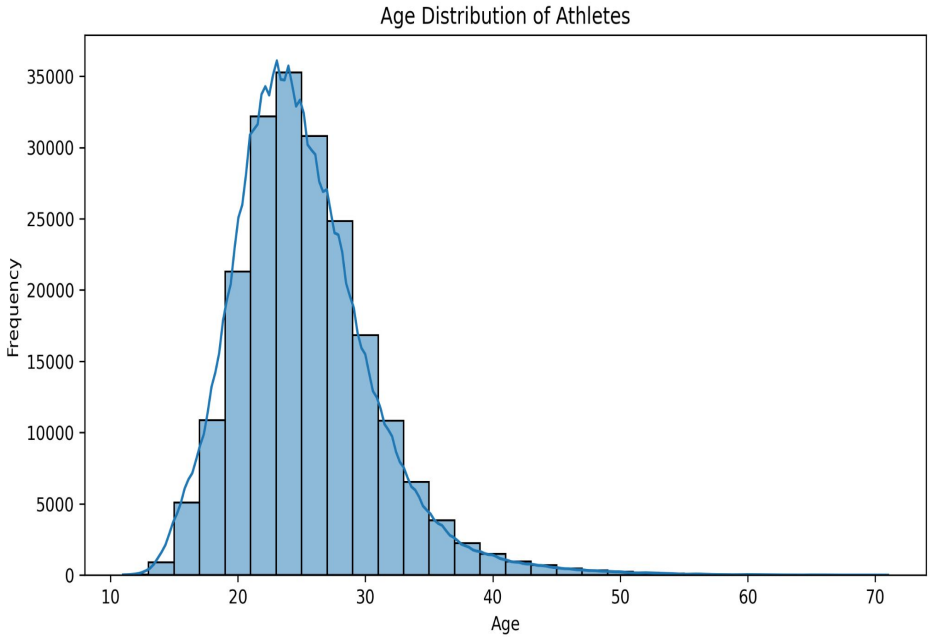
Mean, Mode, Medal count

```
Mean Age:  25.055508937016466
Mode Age:  23.0
Mean Height:  175.3719496519778
Mean Weight:  70.68833701161691

Medal
Gold      10167
Bronze    10148
Silver     9866
Name: count, dtype: int64
```

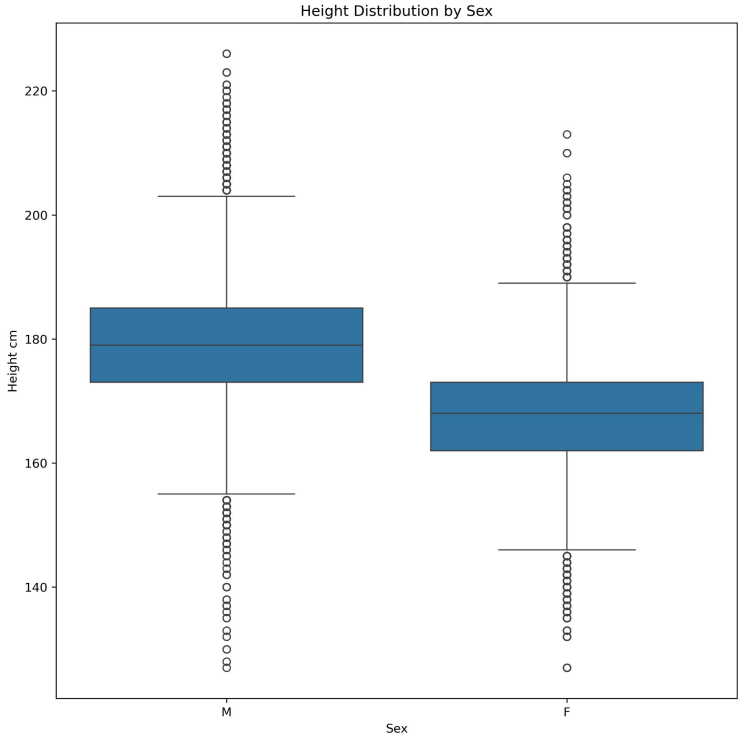
To further understand the data and create a starting point, I took a look at the Mean and Mode of the ages of athletes, as well as the Mean of both height and weight. Moreover, I wanted to see the amount of Medals won in each ranking.

Age Distribution



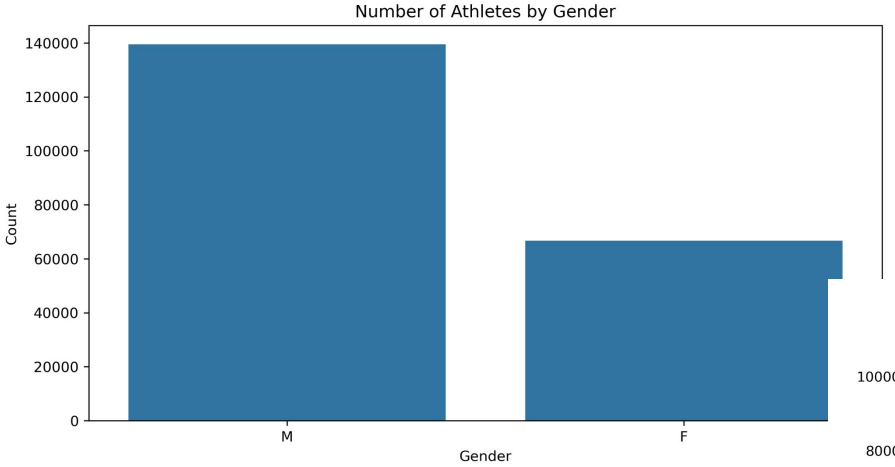
Most Olympic athletes are between 20 and 30 years old, with a sharp decline after age 35, indicating peak athletic performance years.

Height Distribution by Gender

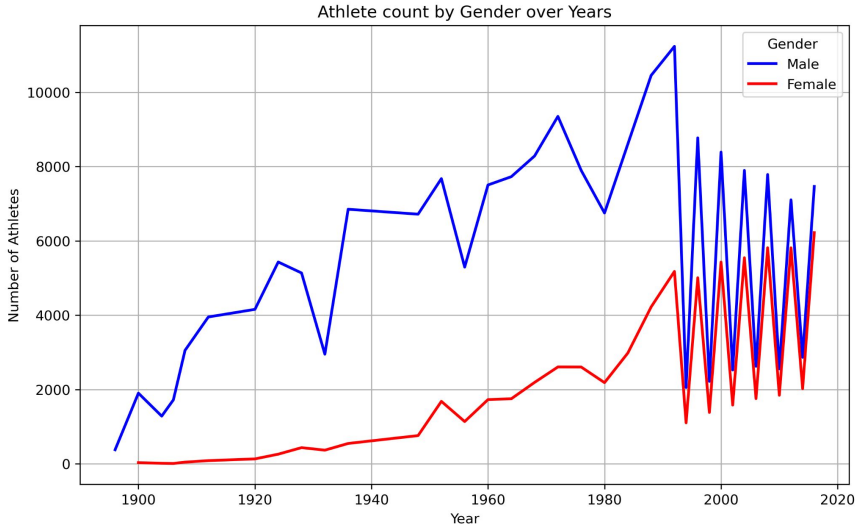


Males are taller on average, though both genders show a broad range of heights, depending on sport type.

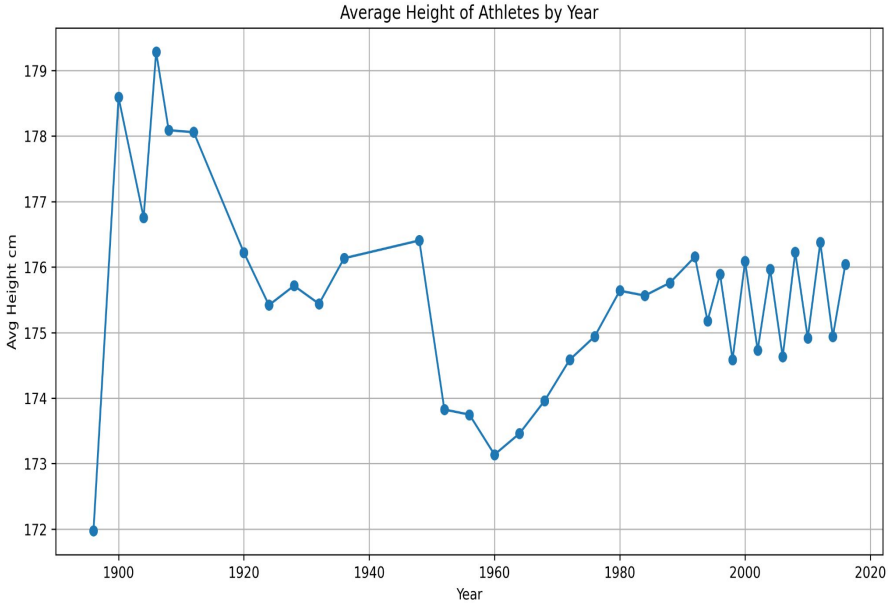
Gender Representation



While male athletes have historically dominated, recent decades show a near-equal gender split, reflecting improvements in representation.

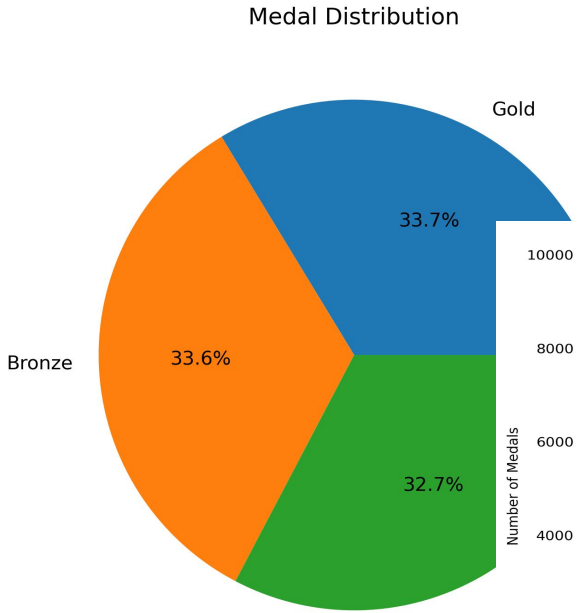


Trends in Height Over Time

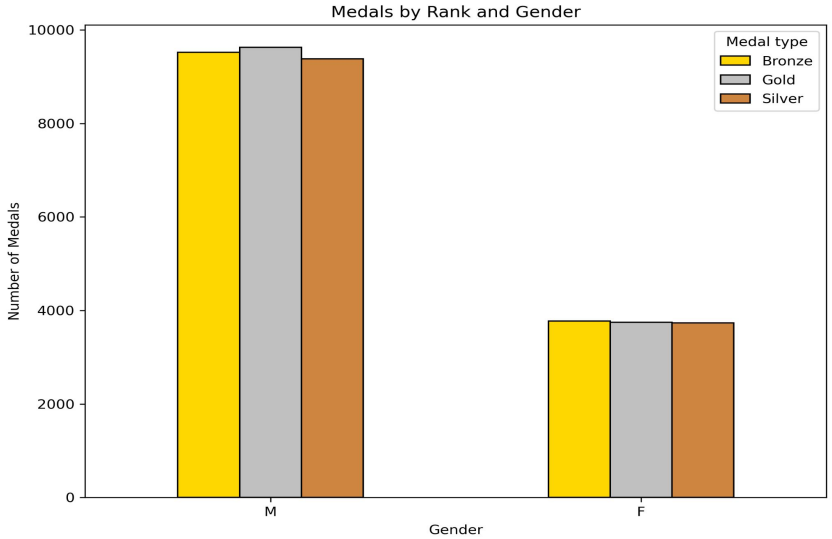


A steady increase in average athlete height is seen over time, possibly due to training improvements and changing selection standards.

Medal Distribution

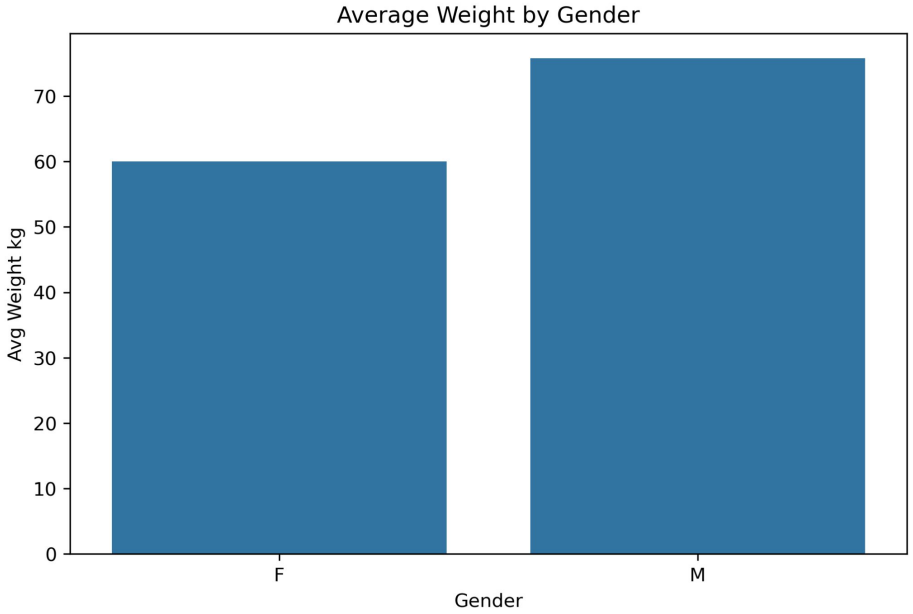


Gold medals are the most common. This could be due to more athletes receiving Gold in team events and a wider medal spread.



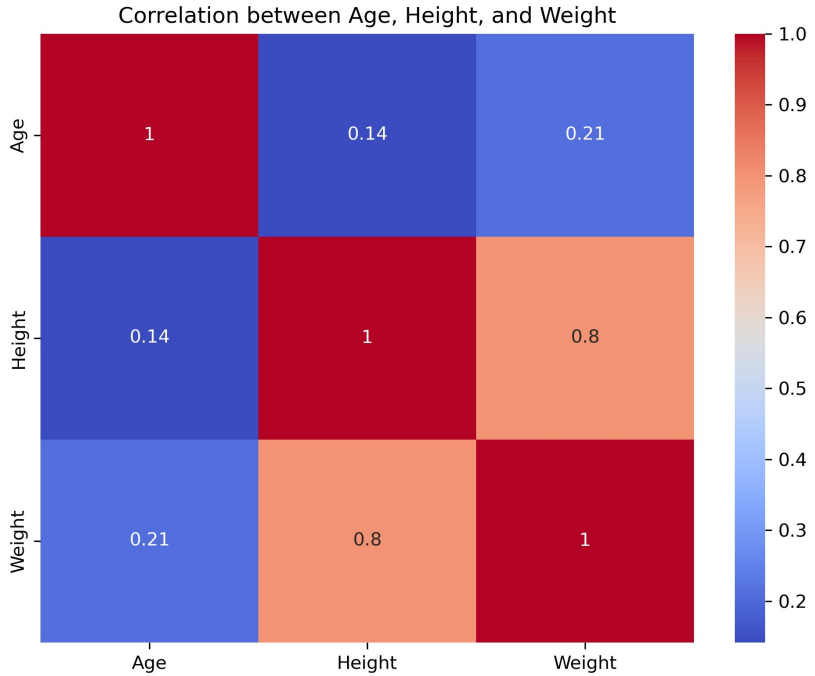
However stepping down into the data, we see that Males have won more medals overall the Females, this could also be due to more female participants in recent years

Average Weight by Gender



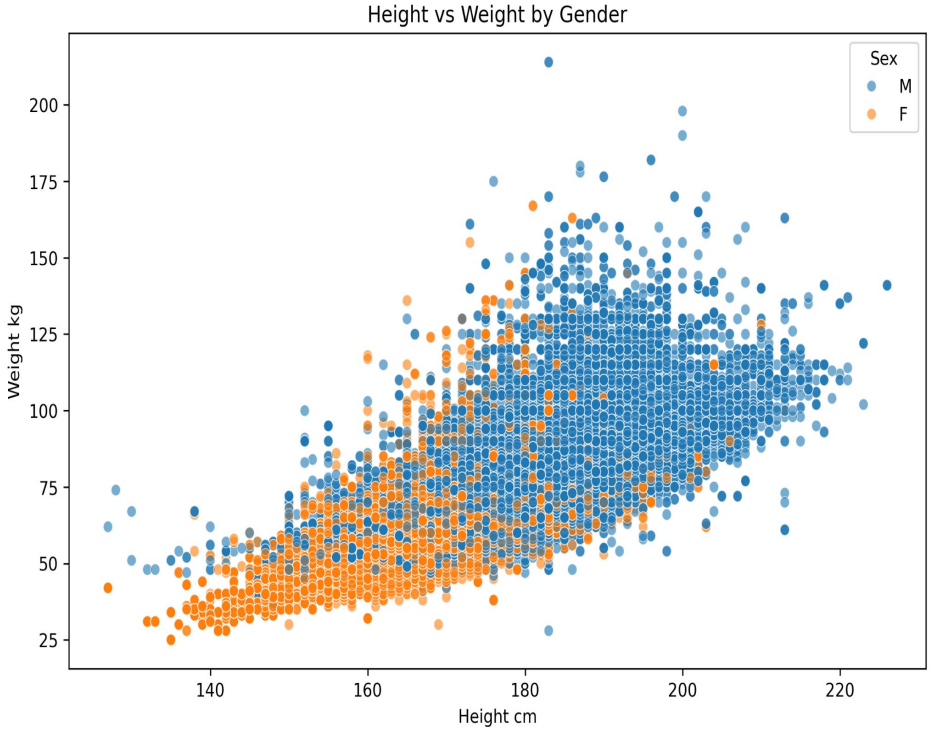
Male athletes weigh more on average, aligned with natural physiological differences and differing sports disciplines.

Correlation Heatmap



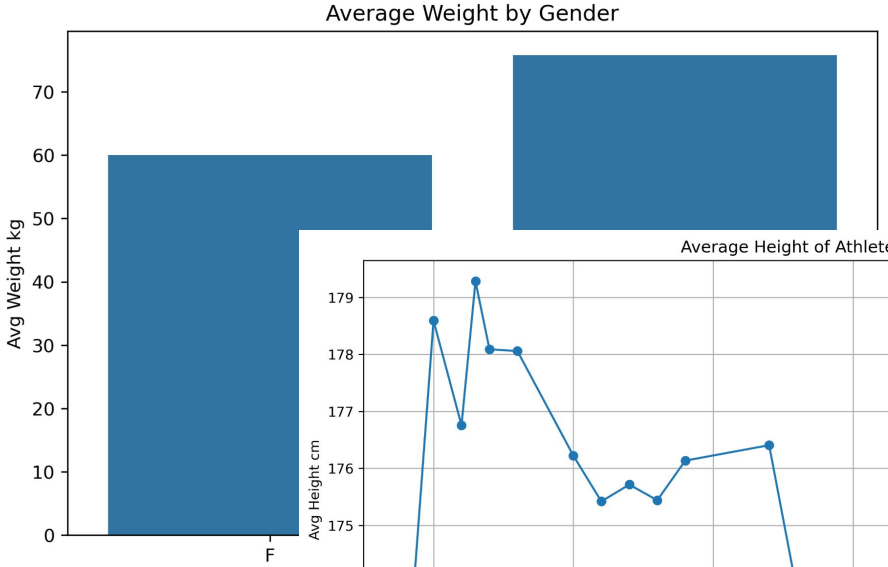
Height and weight show a strong positive correlation. Age is moderately correlated with weight, less so with height.

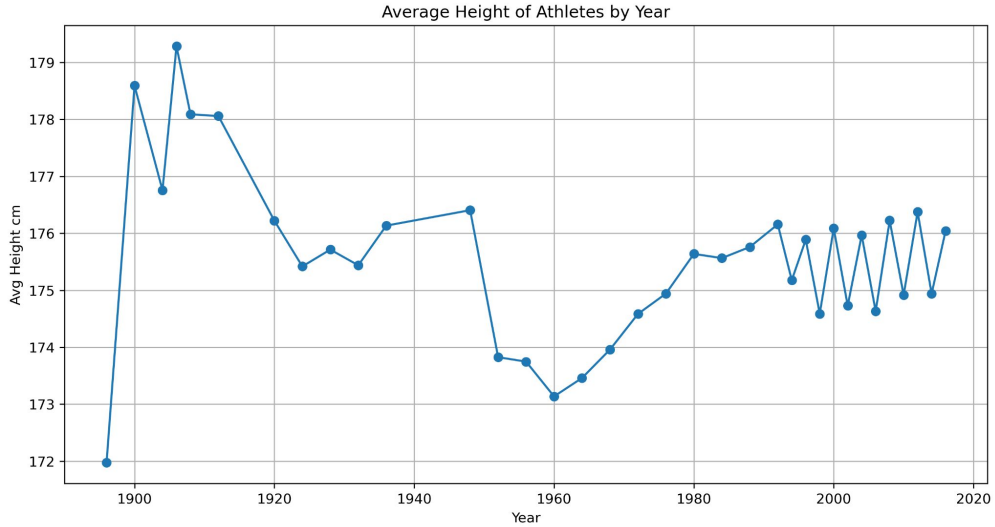
Height vs Weight Scatter Plot



A near-linear relationship exists between height and weight. Gender clusters are distinct, reflecting different body compositions and sports.

Grouped Analysis with .groupby()





Using groupby(), we compared average height and weight by gender, revealing consistent differences across groups.

As well as height by year which allows us better understanding.

Challenges and Interpretation

Missing values in key numeric columns posed a challenge. Interpretation required assumptions about training regimes, biological norms, and event-specific needs.

Conclusion

This analysis highlighted trends in age, height, gender, and performance. Future work could include longitudinal tracking of individual athletes or comparing different nations' training trends.

Acknowledgements

- Data sourced from a provided olympic dataset.
- Visualizations created using Python, pandas, seaborn, and matplotlib.