Breaking the Olympic Mold: A Century of Athletic Evolution through Data

Ruhama Baruda

Department of Mathematics, Howard University

MATH014: Introduction to Data Science

April 24, 2025

Breaking the Olympic Mold: A Century of Athletic Evolution through Data

This project explores historical trends in Olympic athlete participation, medal distribution, and physical characteristics using a dataset from Kaggle that spans from Athens 1896 to Rio 2016. The objective is to identify how gender, nationality, sport, and physical metrics like height, weight, and BMI have influenced outcomes and evolved over time.

Introduction

The Olympic Games have long served as a global stage for athletic excellence and international unity. Since their modern inception in 1896, the Games have grown not only in size and scope but also in their reflection of social and political change. Using a comprehensive dataset of athlete-event records, this project examines how patterns of inclusion, performance, and representation have shifted across gender, nationality, and sport. The analysis uses a data set that includes athlete-level records from the modern Olympic Games spanning 1896 to 2016, encompassing both Summer and Winter Games. Each row in the dataset corresponds to a single athlete's participation in a specific Olympic event.

Methodology

The original dataset included over 270,000 rows, each representing an athlete's participation in a specific event. Initial steps involved:

- Removing 1,385 duplicate records.
- Imputing missing values in Age, Height, and Weight using median values grouped by Sport and Sex.
- Filling all missing Medal values with "No Medal."
- Renaming the "Team" column to "Country" for clarity.
- Standardizing text across categorical columns.
- Creating new features:
    - BMI (Body Mass Index) = Weight / (Height in meters)^2
    - Is_Medalist: binary flag for winning any medal
    - Medal_Value: ordinal scale (Gold = 3, Silver = 2, Bronze = 1, No Medal = 0)
    - Decade: decade extracted from the Year column

Data Visualization and Interpretation

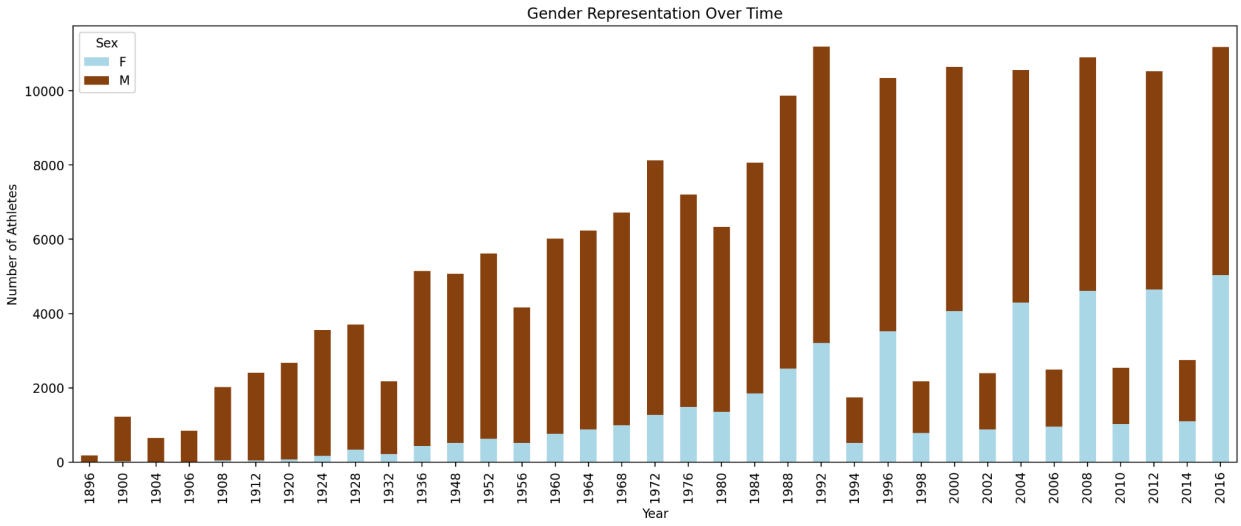

Figure 1: Gender Representation Over Time

This stacked bar chart by year shows increasing gender parity over time from 1896 to 2016. In the earliest Games, participation was entirely male, with the first women competing in 1900. The most dramatic increase in female athlete representation occurred post-1980, following widespread reforms by the International Olympic Committee (IOC) and global policy shifts toward gender equity.

This visualization captures the impact of systemic changes—such as the expansion of women's events in athletics, swimming, and weightlifting—and marks progress toward gender parity in global sport.

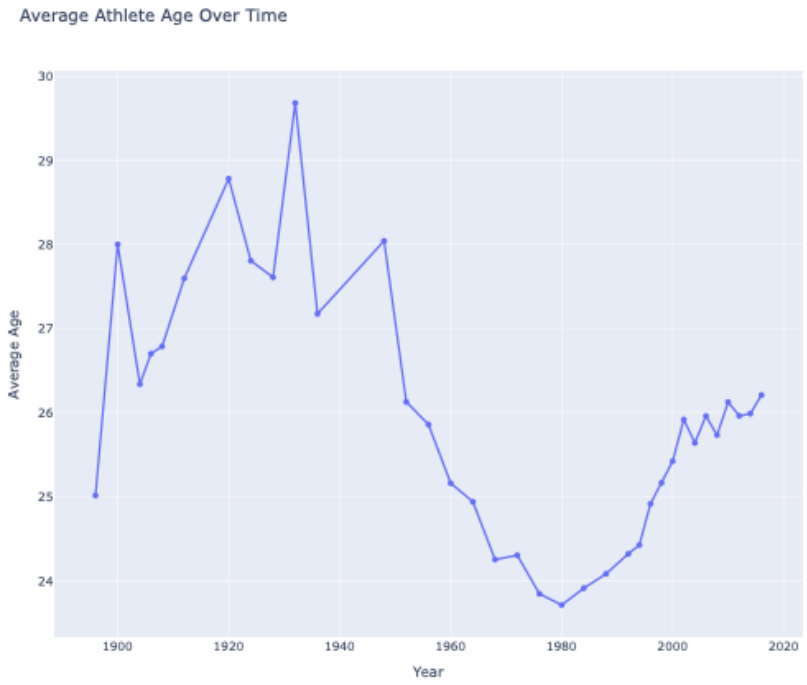

Figure 2: Average Athlete Age Over Time

A line graph of average athlete age reveals that despite significant changes in Olympic scale and format, athlete age has remained fairly consistent (24-26 years old). Slight dips and peaks align with historical events (e.g., post-WWII recovery) or the introduction of youth-skewed sports like skateboarding and BMX. This steadiness suggests that while sport composition evolves, the optimal age window for Olympic competition has remained biologically and strategically consistent.

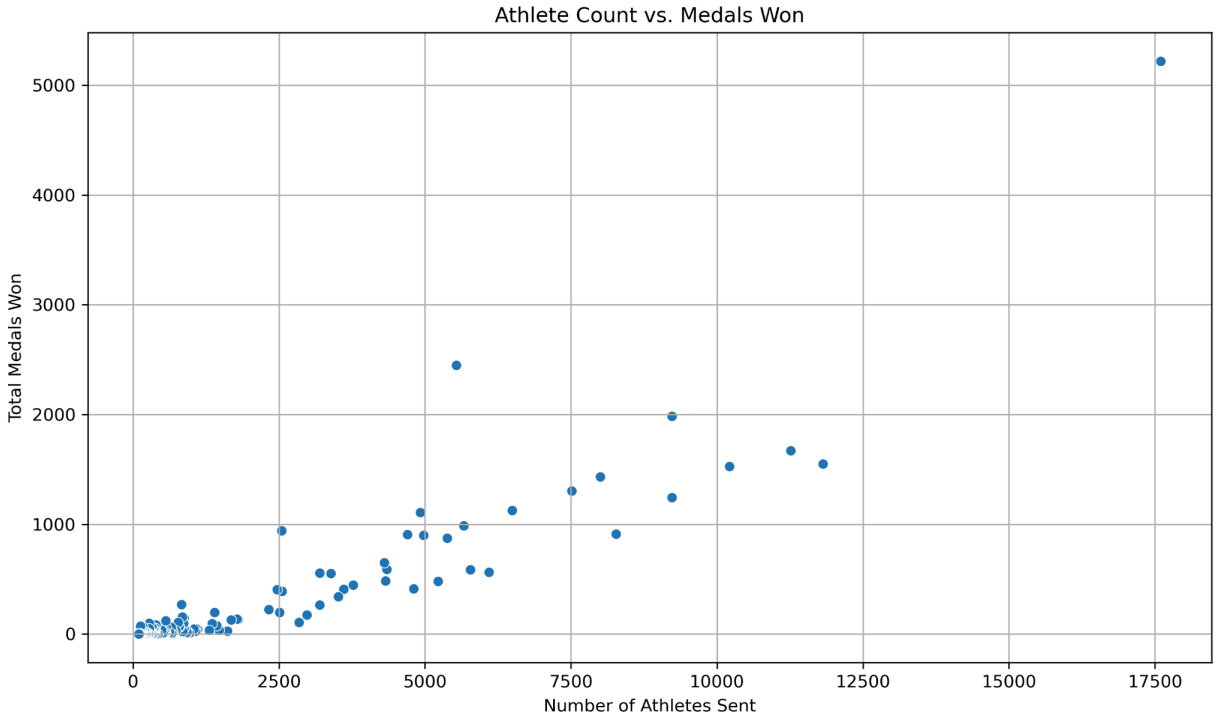

Figure 3: Athlete Count vs Medals Won

This scatter plot compares countries' total athlete counts to total medals revealing a clear positive correlation (r = 0.9). Countries that send more athletes tend to win more medals, but some nations like Kenya and Jamaica punch above their weight, achieving high medal counts with relatively small delegations.

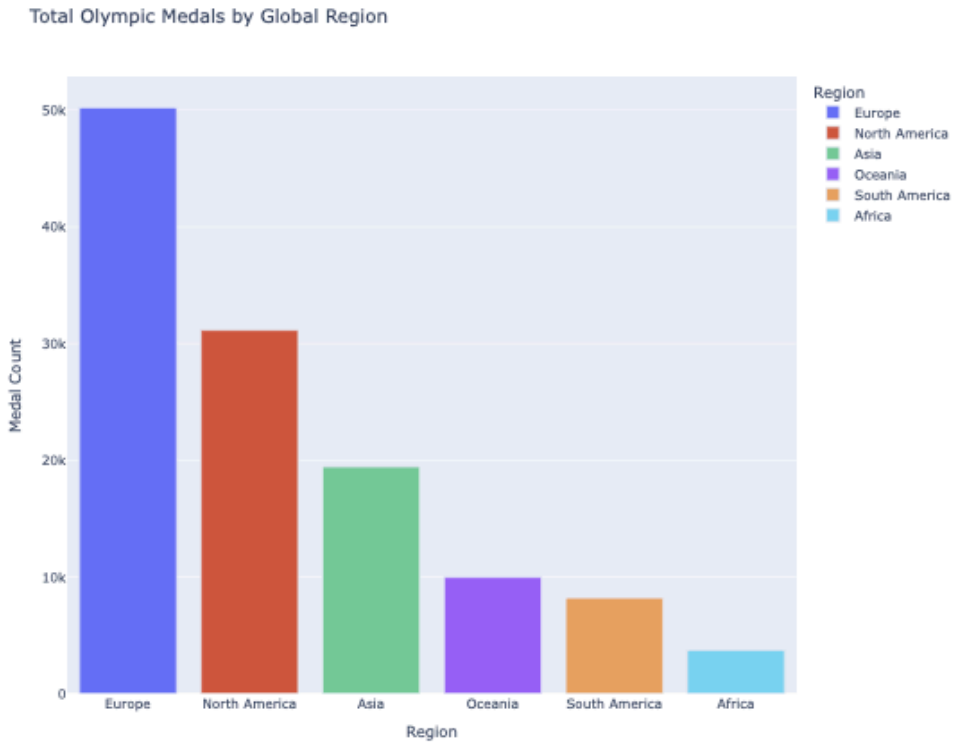

Figure 4: Total Olympic Medals by Global Region

To explore global disparities in Olympic performance, a regional analysis was conducted by mapping NOC codes to global regions. A bar chart shows that Europe has consistently dominated in total medals, followed by North America and Asia. Africa and Oceania had substantially fewer total medals, reflecting regional inequalities in resources, access to training infrastructure, and Olympic representation.

This visualization highlights the geopolitical and economic factors that influence international athletic success.

Figure 5: Medal Count by Gender Across Decades

Using the Decade column, a grouped bar chart displayed decade-over-decade growth in women's medal counts, underscoring the long-term impact of gender-focused policy reforms.

While male totals remain higher, the gap has narrowed significantly due to these policies by the IOC and member nations

Conclusion

This analysis of over a century of Olympics history reveals that success at the Games is shaped by far more than athleticism alone; it is deeply influenced by structural, social, and geopolitical forces. A key finding was the strong correlation ($r = 0.90$) between the number of athletes a country sends and the number of medals it wins, illustrating how access, investment, and inclusion directly drive performance. The Summer Olympics, in particular, continue to dominate in both athlete representation and medal counts, reflecting the broader accessibility and diversity of summer sports compared to their winter counterparts.

While physical traits such as BMI play a role in certain events, the weak correlation with medal outcomes ($r = 0.01$) confirms that no single body type guarantees success; performance is context-specific, varying widely by sport. Social change is also evident in the data—female athlete participation and medal achievements have grown significantly since the 1980s, especially in traditionally male-dominated sports like athletics and swimming.

Finally, the data reveals how the Olympics mirror historical and cultural shifts, from the disappearance of early events like Art Competitions to the rise of sports like Judo and Triathlon, and the visible impact of political events such as boycotts. In this way, the Olympics emerge not only as a showcase of global athleticism but also as a living record of progress, inequality, and transformation across nations and generations.

References

120 years of Olympic history: athletes and results. (2018, June 15). Kaggle. https://www.kaggle.com/datasets/heesoo37/120-years-of-olympic-history-athletes-and-results

Sports Reference | Sports Stats, fast, easy, and up-to-date | Sports-Reference.com. (n.d.). Sports-Reference.com. https://sports-reference.com/

Acknowledgments

Thanks to our course instructor, Kaggle contributors, and the Olympic historians who compiled and preserved the data used in this analysis.