**“An Athlete’s Dream”**

A data analysis of ‘120 years of Olympic history: athletes and results’ (Retrieved from Kaggle)

Quentin Bone (@03054253)

Math 014-01

Dr. Nerolu Meenshaki

April 29th, 2025

**Introduction**

The dataset used in this project, "120 years of Olympic history: athletes and results," was retrieved from Kaggle. It contains 271,116 rows across 15 columns describing athlete statistics, performances, and country-level results from the inception of the modern Olympic Games in 1896 in Athens through 2016 in Rio.

**Objective**

This project aims to explore historical patterns in Olympic participation, performance, and athlete demographics. Through data cleaning, exploratory analysis, visual storytelling, and interactive visualizations, the study uncovers how the Games have transformed over time into a global phenomenon reflecting broader shifts in inclusion, geopolitics, and athletic achievement.

**Method**

The dataset was first verified for structure and completeness. Columns such as Age, Height, Weight, and Medal were verified for missing values. Data wrangling steps taken were:

Making a working copy of the dataset

Removing duplicate records

Imputing missing Age, Height, and Weight with gender-specific medians

Converting respective columns into numeric or categorical type

Engineering new features such as BMI and Age Group

These steps rendered the dataset clean, uniform, and prepared for meaningful analysis.

**Storytelling**

Initial exploratory analysis showed that the number of athletes, countries, and sports has been growing consistently over the years. There was an explosion in participation following World War II and even more so after the 1980s. Men held sway in the Games in the initial decades, but the expansion of female participation—more so from the 1970s—mirrored broader developments in gender integration. Countries like the United States, Soviet Union, and Germany always found themselves at the top of medal counts, with new countries being added along the way.

The exploratory data analysis phase was used to establish questions about physical attributes of competitors, medal counts, and participation gaps, opening up further deeper visual scrutiny.

**Data Visualization and Interpretation**

To clarify and enrich exploratory findings, visualizations were created:

A graph of a bar graph

AI-generated content may be incorrect.

**Figure 4.1:** Top 10 Countries by Total Medals

This chart shows the ten countries that have won the most total Olympic medals, gold, silver, and bronze. The United States is far ahead of all other countries by a very large margin. Countries like the Soviet Union and Germany also have very high medal counts, but they are still much behind the United States. This shows that some countries have had a good tradition of winning at the Olympics for many decades.

A graph with orange and blue lines

AI-generated content may be incorrect.

**Figure 4.2:** Olympic Participation Over Time by Gender

This line graph shows how the number of female and male athletes has grown over the years at the Olympics. In the beginning, almost all the athletes were males, but since about the 1970s, the number of female athletes began growing extremely fast. In the latest years, female participation came close to male participation remarkably, showing that the Olympics have become much more inclusive for females. It also shows that profound changes in society did indeed have a real impact on sport over time.

A diagram of a graph

AI-generated content may be incorrect.

**Figure 4.3:** Age Distribution by Medal Type

This boxplot is utilized to compare the age of gold, silver, and bronze medalists. Most of the medalists are twenties or early thirties, but there are also some younger or older medalists. Gold medalists are slightly younger in average than silver and bronze winners. There are also a few outliers on the graph — some very young and very old medalists — but the great majority of athletes are in a pretty similar age range.

A diagram of a graph

AI-generated content may be incorrect.

**Figure 4.4:** Age Distribution by Medal Type (Under 50 years old)

This graph is similar to the previous graph but removes some of the outliers to take a closer look at the boxplots themselves.

A screenshot of a graph

AI-generated content may be incorrect.

**Figure 4.5:** Pairplot of Physical Attributes

This pairplot shows how different physical attributes like age, height, weight, and BMI are related to each other. As is clear from the graph, there is a clear positive relationship between weight and height, i.e., taller athletes weigh more. Weight and BMI are also closely related. Relations like age and height or age and BMI are not as strong, i.e., age is not a good predictor of how big an athlete is.

A graph of a number of medals

AI-generated content may be incorrect.

**Figure 4.6:** Top 10 Sports by Medal Type

This bar chart shows which sport has awarded the most medals in total, segregated by gold, silver, and bronze. There are most medals in athletics and swimming because these sports have so many various events and races. Other sports, like wrestling, gymnastics, and rowing, also account for a big number of medals. The chart is used to show that certain sports offer quite a bit more medal opportunities compared to others, giving nations more chances to succeed in some divisions.

A pie chart with numbers and text

AI-generated content may be incorrect.

**Figure 4.7:** Distribution of Medals (Gold, Silver, Bronze)

This pie chart represents the overall percentage of gold, silver, and bronze medals that have been presented throughout Olympic history. The medals are distributed relatively evenly between gold, silver, and bronze and each represents approximately a third of the total. This makes sense because for the majority of events, three medals are awarded. The pie chart is used to show that there is not some large imbalance in the distribution of medals across all events.

A screenshot of a graph

AI-generated content may be incorrect.

**Figure 4.8:** Correlation Heatmap of Physical Attributes

This heatmap illustrates to what extent various physical attributes are correlated with one another. Height and weight are positively correlated and very strongly, meaning taller players are more likely to be heavier. BMI and weight are also well correlated. But age is not correlated with height, weight, or BMI very strongly, meaning an athlete's age does not predict his/her body size well. In general, the heatmap quickly identifies which traits are associated and which are not.

Analysis also included animated visualizations:

**Question 1**: How has the number of athletes increased by decade?

**Figure 5.1:** Number of Athletes Over Decades

A graph with a line going up

AI-generated content may be incorrect.

This animated line plot shows how the number of athletes has increased per decade. The Olympics have seen steady growth throughout the years, especially after World War II and again after the 1980s, reflecting how the Games expanded globally.

**Question 2:** Which sports have had the most female participation over time?

**Figure 5.2:** Top 10 Sports with Most Female Participation

A graph showing a number of athletes

AI-generated content may be incorrect.

This interactive bar plot shows the top 10 sports with the highest number of female athletes. Athletics, Swimming, and Gymnastics are among the leading sports where women have been highly active, reflecting changes in opportunities and popularity over time.

**Question 3:** How has the average age of medalists changed over time?

**Figure 5.3:** Average Age of Olympic Medalists Over Time

A graph showing the growth of the olympic games

AI-generated content may be incorrect.

This trend line shows how the average age of the Olympic medalists has changed over the years. While there are fluctuations, the general trend is that athletes have continued to be in their early-to-mid twenties, with minor distinctions depending on the era and sport specialization.

**ScatterGeo Map:**

**Figure 5.4:** Total Olympic Medals by Country

A map of the world

AI-generated content may be incorrect.

This map uses dots placed on each country to represent the total count of Olympic medals won. The size of each bubble corresponds to the amount of medals that country has won, with larger bubbles representing higher numbers. A custom range of sizes was applied so that even countries with only a few medals still appear on the map with a visible bubble, which helps to highlight their presence without being completely overshadowed by the frontrunner nations. Bubble color also indicates medal count, allowing for a nice visual differentiation between countries with fewer and more awards. Country borders, land and sea colors were added to increase the map's readability and visual balance.

These dynamic visuals offered a fresh look at how global engagement and athletes have evolved.

**Conclusion**

This project provided a rich and informative analysis of the evolution of the Olympic Games. The data showed that the Olympics have become more inclusive and global over time. Female athlete participation has increased steadily, and more countries are earning medals across a wider range of sports. A few powerful nations continue to dominate the medal tables, reflecting both historical investment in sports infrastructure and geopolitical influence.

By combining traditional EDA, visual storytelling, and interactive mapping, this analysis demonstrates how sports can reflect broader societal trends. The Olympic Games are not just a celebration of physical ability—they are a mirror of social progress, international development, and global unity.

**References**

* Kaggle: 120 years of Olympic history: athletes and results

**Acknowledgments**

Special thanks to Doctor Nerolu Meenashki for the guidance in making this project happen, and to the Kaggle contributors who made this project possible.