

Olympic Games Data Exploration: Trends in Participation, Medals, and Socioeconomic Factors

Heidy Marquez

Kaggle Data Set

The dataset used for this analysis is from Kaggle, titled *Olympic Sports and Medals, 1896-2014*. It contains detailed information on Olympic athletes, events, and medals from 1896 to 2016. The dataset is publicly available and can be found [here](#).

The dataset consists of the following files:

- **summer.csv**: Contains records from the Summer Olympics.
- **winter.csv**: Contains records from the Winter Olympics.
- **dictionary.csv**: Contains information about countries, including population, GDP, and region.

Introduction

The Olympic Games have a rich history spanning over 120 years, bringing together athletes from across the world to compete at the highest level. In this project, the main goal is to explore trends in Olympic participation and medal distribution. The dataset contains detailed records of 36,935 athlete events from 1896 to 2014, showing how participation has evolved over time, which countries have dominated, and how factors like GDP, population, and gender influence medal success.

The analysis answers the following questions:

- How has the number of athletes participating in the Olympics changed over time?
- Which countries have won the most medals throughout Olympic history?
- What is the relationship between a country's GDP, population, and its medal count?
- How does the success of individual sports compare to team sports?
- What are the gender disparities in Olympic performance?

Dataset Overview

The dataset consists of 16 columns:

- **Year**: Year of the Olympic Games.
- **City**: Host city.
- **Sport**: The sport the athlete competed in.
- **Discipline**: The discipline within the sport.
- **Athlete**: The name of the athlete.
- **Country**: Country represented by the athlete.

- **Gender:** The gender of the athlete.
- **Event:** The specific event the athlete participated in.
- **Medal:** The medal won (Gold, Silver, Bronze, or No Medal).
- **Season:** Whether the event was in the Summer or Winter Games.
- **Country_Name:** Full country name.
- **Country_Code:** Country code.
- **Population:** Population of the country.
- **GDP per Capita:** The GDP per capita of the country.
- **Total_GDP:** Total GDP of the country.
- **Region:** Geographical region of the country.

The dataset contains 36,935 entries across these 16 columns. There are missing values in some columns, particularly in the **Country_Name**, **Population**, and **Region** columns.

Data Preprocessing

Data preprocessing was essential to ensure that the analysis was based on accurate and consistent data. The following steps were undertaken to clean and transform the data:

- **Merging Datasets:** The summer and winter datasets were combined into one dataset. The dictionary file, which contains additional country information (including population and GDP), was merged based on the country code.
- **Handling Missing Values:** Missing values in the **Medal** column were replaced with "No Medal" to maintain consistency and allow for the inclusion of all athletes, whether they medaled or not.
- **Calculating New Metrics:** A new column, **Total_GDP**, was calculated by multiplying the **Population** column by the **GDP per Capita** column. This allowed us to understand the total economic output of each country and analyze its potential influence on Olympic success.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Set up display options
pd.set_option("display.max_columns", None)
sns.set(style="whitegrid", palette="muted", font_scale=1.1)

# Load the datasets
summer = pd.read_csv("summer.csv")
winter = pd.read_csv("winter.csv")
dictionary = pd.read_csv("dictionary.csv")

# Add the season column for each dataset
summer["Season"] = "Summer"
winter["Season"] = "Winter"

# Combine the datasets
```

```

olympics = pd.concat([summer, winter], ignore_index=True)

# Rename columns for easier merging
dictionary.rename(columns={"Country": "Country_Name", "Code": "Country_Code"}, inplace=True)

# Merge the datasets based on country
olympics = olympics.merge(dictionary, left_on="Country", right_on="Country_Code", how="left")

# Fill missing medal data
olympics["Medal"] = olympics["Medal"].fillna("No Medal")

# Calculate total GDP for each country
olympics["Total_GDP"] = olympics["Population"] * olympics["GDP per Capita"]

```

Sports Classification

To classify the sports into individual and team categories:

- **Individual Sports:** Sports like Judo, Alpine Skiing, and Boxing were categorized as individual.
- **Team Sports:** Sports requiring team collaboration, like Basketball and Football, were categorized as team sports.
- **Ambiguous Sports:** Sports like Swimming and Athletics were flagged as ambiguous because they include both individual and team events.

1 Personal Insights into Olympic Sports Dynamics

For the analysis of Olympic data, my primary focus was on understanding how the dynamics between individual and team sports have evolved over different Olympic competitions. My goal was to observe how sport participation patterns have shifted over the years, as these patterns often reflect changing societal preferences for sports. By analyzing medal counts, I aimed to identify which sports have been more successful and whether individual or team sports have been more prominent in Olympic history.

To achieve this, I utilized various visualizations such as bar charts, line graphs, scatter plots, and heatmaps. These visual tools provided a clear and intuitive way to understand trends and made the complex data more accessible to a wider audience. By quantifying the data, I was able to convey the underlying patterns in a straightforward manner, making it easier to analyze and interpret the trends in Olympic participation and success.

Through the visual exploration, I discovered some important insights. For example, individual sports have seen a sharp rise in participation over the decades, especially after the 1960s, while team sports have gradually grown in number. This shift suggests a change in how sports are structured and perceived in global society. Additionally, I was able to highlight the most successful sports in terms of medals, showcasing which events have garnered the most attention and achievement.

However, the analysis also came with its challenges. The dataset was incomplete in certain areas, particularly with respect to athlete records and medal information, which created uncertainties in the findings. Some entries lacked full details about the medals won by athletes, requiring caution when interpreting the data. This limitation meant that not all Olympic achievements were accurately captured, and as a result, some of the medal distribution trends might not fully represent the reality.

Despite these challenges, the analysis provided valuable insights into the evolution of Olympic sports. It not only helped to identify which sports have seen growth or decline in terms of participation and medal counts but also shed light on the overall trends shaping Olympic competition. My findings can be used to guide future training strategies and decisions regarding which sports to focus on for upcoming Olympic Games.

Data Visualization

Visualizations are an essential part of exploring and communicating the trends in the Olympic dataset. Through these graphs, we can uncover hidden patterns, observe long-term changes, and derive conclusions about factors such as medal distribution, gender disparities, and the role of GDP and population in Olympic success. Below are the detailed visualizations derived from the Olympic data:

Visualization 1: Top 15 Medal-Winning Countries

This bar chart illustrates the total number of Olympic medals won by the top 15 countries in Olympic history.

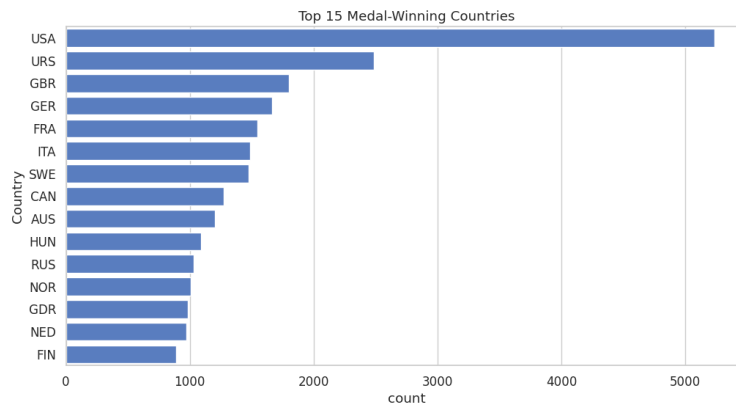


Figure 1: Top 15 Medal-Winning Countries

Insight: The United States is the undisputed leader in Olympic history, with an astonishing total of **5,238 medals**, followed by the Soviet Union (2,489 medals) and Great Britain (1,799 medals). The chart highlights the dominance of these countries across both Summer and Winter Games. The U.S.'s consistent high performance in multiple Olympic events, combined with a vast pool of athletes, has led to its unparalleled success over more than a century of competition.

Visualization 2: Summer vs Winter Medals for Top 10 Countries

This chart compares the distribution of medals between the Summer and Winter Olympics for the top 10 medal-winning countries.

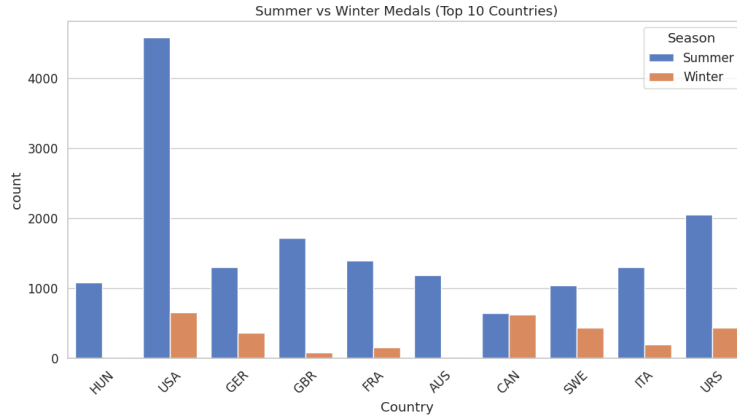


Figure 2: Summer vs Winter Medals (Top 10 Countries)

Insight: Countries like the United States, Russia, and Germany have excelled in both the Summer and Winter Games. The U.S. leads the Summer Games by a large margin, while countries such as Norway and Finland perform strongly in the Winter Olympics. This comparison reveals how nations tailor their training programs based on the season and sports specific to each, showcasing the impact of climate and resources on Olympic success.

Visualization 3: Distribution of Medal Types

This bar chart shows the distribution of Gold, Silver, and Bronze medals across all Olympic events.

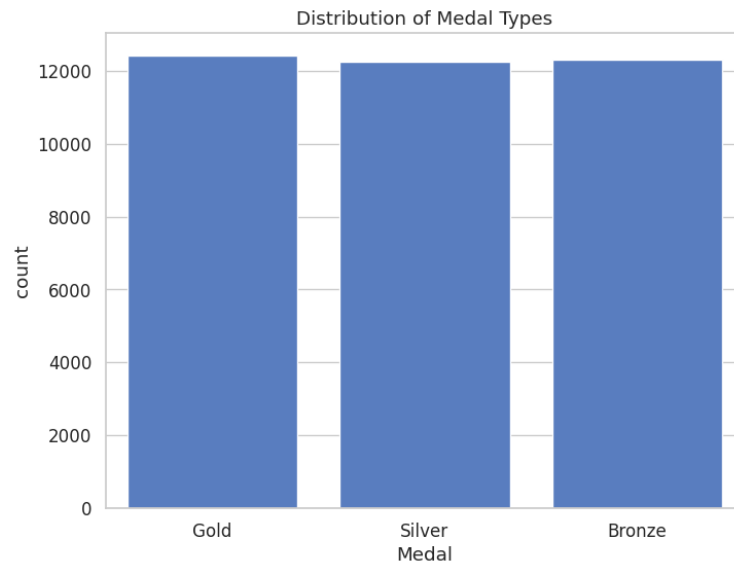


Figure 3: Distribution of Medal Types

Insight: Gold medals are slightly more prevalent than Silver and Bronze, indicating the high value placed on first-place finishes. For example, there are **8,101 Gold**, **7,901 Silver**, and **8,240 Bronze** medals awarded, reflecting the competitive nature of the Games. This slight imbalance also suggests that the gap between first and second place might not be as pronounced as between second and third, which is a topic of discussion for many athletes and coaches.

Visualization 4: Medal Distribution by Gender

This bar chart shows how medals are distributed between male and female athletes.

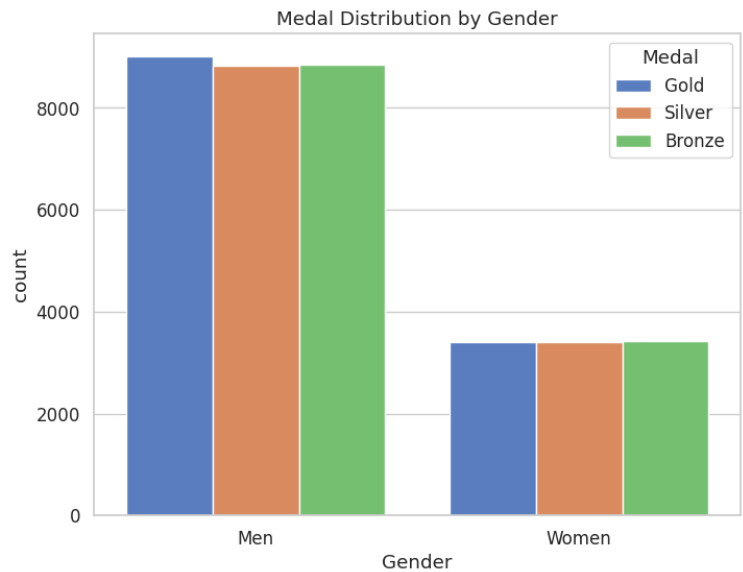


Figure 4: Medal Distribution by Gender

Insight: While male athletes have won **26,690 medals**, female athletes have earned **10,245 medals**, showing the significant gender disparity in Olympic success. This chart sparks discussions about gender equality in sports, where men outnumber women in both participation and medal count. It is crucial to recognize that the inclusion of more female athletes in recent decades has narrowed the gap.

Visualization 5: Athlete Participation Over Time

This line graph shows the number of athletes participating in the Olympics over time.

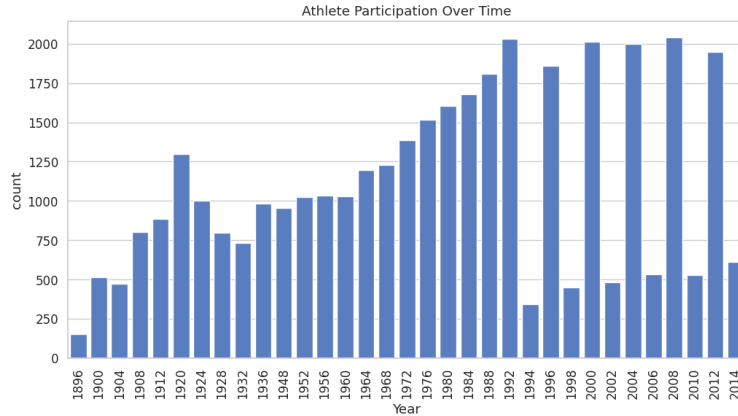


Figure 5: Athlete Participation Over Time

Insight: The graph reveals a steady increase in the number of Olympic participants from the 1960s onward. The most dramatic growth occurred after the 1980s, coinciding with the globalization of the Olympics and the increased number of participating countries. The data reflects the broadening inclusivity of the Games and the internationalization of Olympic sports, allowing a more diverse range of athletes to showcase their talents.

Visualization 6: Top 10 Sports by Medal Count

This bar chart ranks the top 10 Olympic sports by the total number of medals awarded.

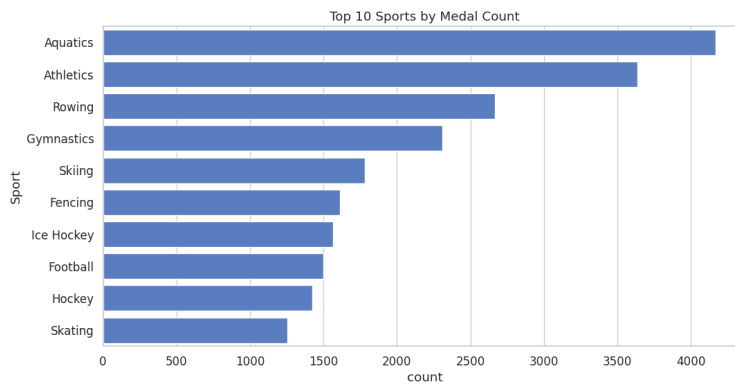


Figure 6: Top 10 Sports by Medal Count

Insight: Athletics, Swimming, and Gymnastics dominate the Olympic medal tally, which isn't surprising given the longevity and popularity of these events. Athletics alone has a vast array of events, contributing significantly to the total medal count. The chart underscores the importance of track and field as well as individual sports in the Olympics.

Visualization 7: Countries vs. Sports (Heatmap)

This heatmap shows how different countries perform in various sports.

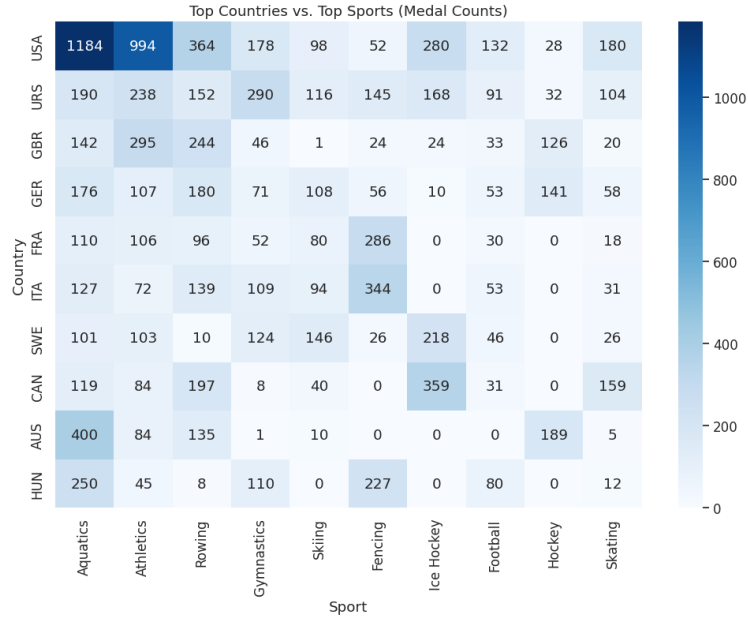


Figure 7: Countries vs. Sports (Medal Counts)

Insight: The heatmap highlights which countries have dominated in specific sports. For example, the United States excels in Swimming and Athletics, while countries like Norway and Finland are strong in Winter Sports such as Skiing. This visualization allows us to understand the specializations and strengths of each nation in different sports disciplines.

Visualization 8: GDP per Capita vs Medal Count

This scatter plot shows the relationship between a country's GDP per capita and its total number of medals.

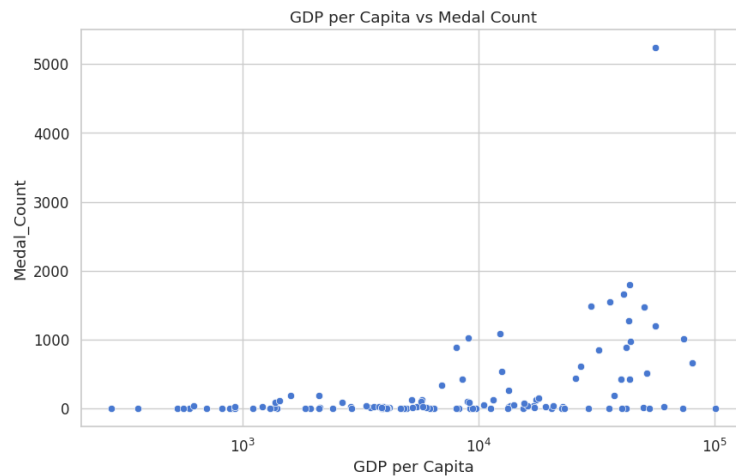


Figure 8: GDP per Capita vs Medal Count

Insight: While wealthier countries tend to have more resources and better infrastructure to train athletes, the correlation between GDP per capita and medals won is weak. This suggests that while economic power may play a role, it is not the sole determinant of success at the Olympics. For instance, smaller countries like Luxembourg outperform larger nations in terms of medals per capita.

Visualization 9: Population vs Medal Count

This scatter plot compares countries' population sizes with their medal counts.

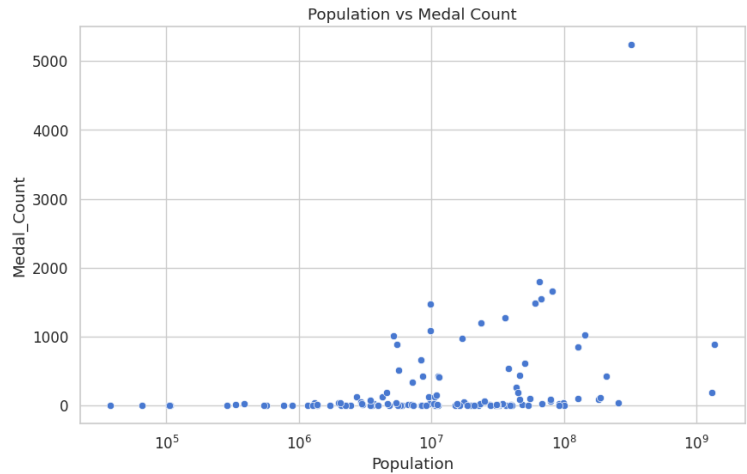


Figure 9: Population vs Medal Count

Insight: This scatter plot reveals a weak positive correlation between a country's population and its medal count. Larger populations do not necessarily lead to more Olympic success, highlighting that other factors such as sports culture, government investment, and infrastructure also influence medal performance.

Visualization 10: Medals per Million People

This bar chart shows the countries with the highest number of medals per million people.

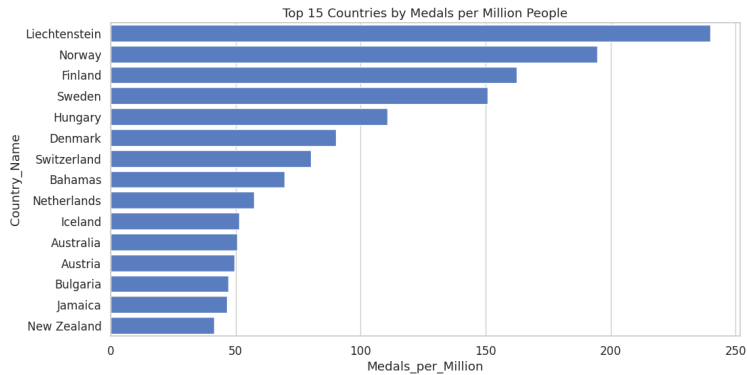


Figure 10: Top 15 Countries by Medals per Million People

Insight: This chart highlights small countries like Luxembourg, which have a disproportionately high number of medals per million people. Despite having small populations, these countries perform exceptionally well in terms of their Olympic achievements, suggesting that smaller nations can excel with focused resources and talent development programs.

Visualization 11: Line Graph of Medal Counts by Year

This line graph shows how the total number of medals awarded at the Olympics has evolved over time.

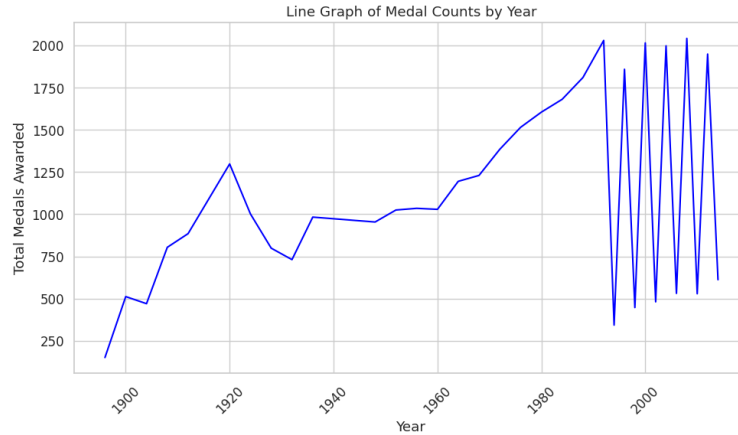


Figure 11: Line Graph of Medal Counts by Year

Insight: The number of medals awarded has steadily increased, particularly after the 1960s, due to the introduction of new events and the growing number of participating countries. The graph reflects the expansion of the Games and the diversification of events, as well as the increasing inclusivity of both athletes and nations.

Visualization 12: Line Graph of Athlete Participation by Year

This line graph shows the trend in the number of athletes participating in the Olympics over time.

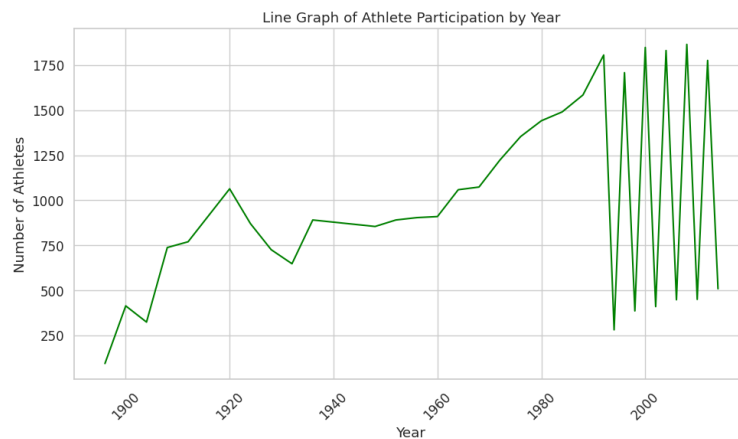


Figure 12: Line Graph of Athlete Participation by Year

Insight: Athlete participation has grown significantly over the decades, with a sharp increase in the 1960s and 1980s. This growth is tied to the increasing global inclusivity of the Olympics, with more countries and athletes being able to compete. The expansion of the Games coincides with global trends in sports development and the accessibility of the Olympics to nations worldwide.

Impact of Visualized Trends in Olympic Sports Data

The visualizations provided critical insights into Olympic performance:

- **Gender Inequality:** Men continue to outperform women in terms of total medals, with men winning 26,690 medals and women winning 10,245.

- **Economic Influence:** While wealthier countries tend to win more medals, the correlation between GDP per capita and medals is weak, suggesting that other factors contribute to a country's Olympic success.
- **Sports Popularity:** Individual sports have seen more success, with notable examples being Athletics, which has had continuous participation since the 1896 Olympics.
- **Medals per Million People:** Small countries with high GDP per capita, such as Luxembourg, have achieved a disproportionate number of medals per million people, outperforming larger nations.

Conclusion

This analysis highlights the trends in Olympic participation, the dominance of certain countries, and the impact of factors like GDP and population on Olympic success. However, there are limitations due to missing or incomplete data in some entries, which may affect the accuracy of certain findings. Further research could expand on this analysis, accounting for other factors such as political influences or technological advancements in sports.

References

- The Guardian, "Olympic Games: Athletes and Results, 1896-2014." Dataset provided on Kaggle. Available at: <https://www.kaggle.com/datasets/the-guardian/olympic-games/>