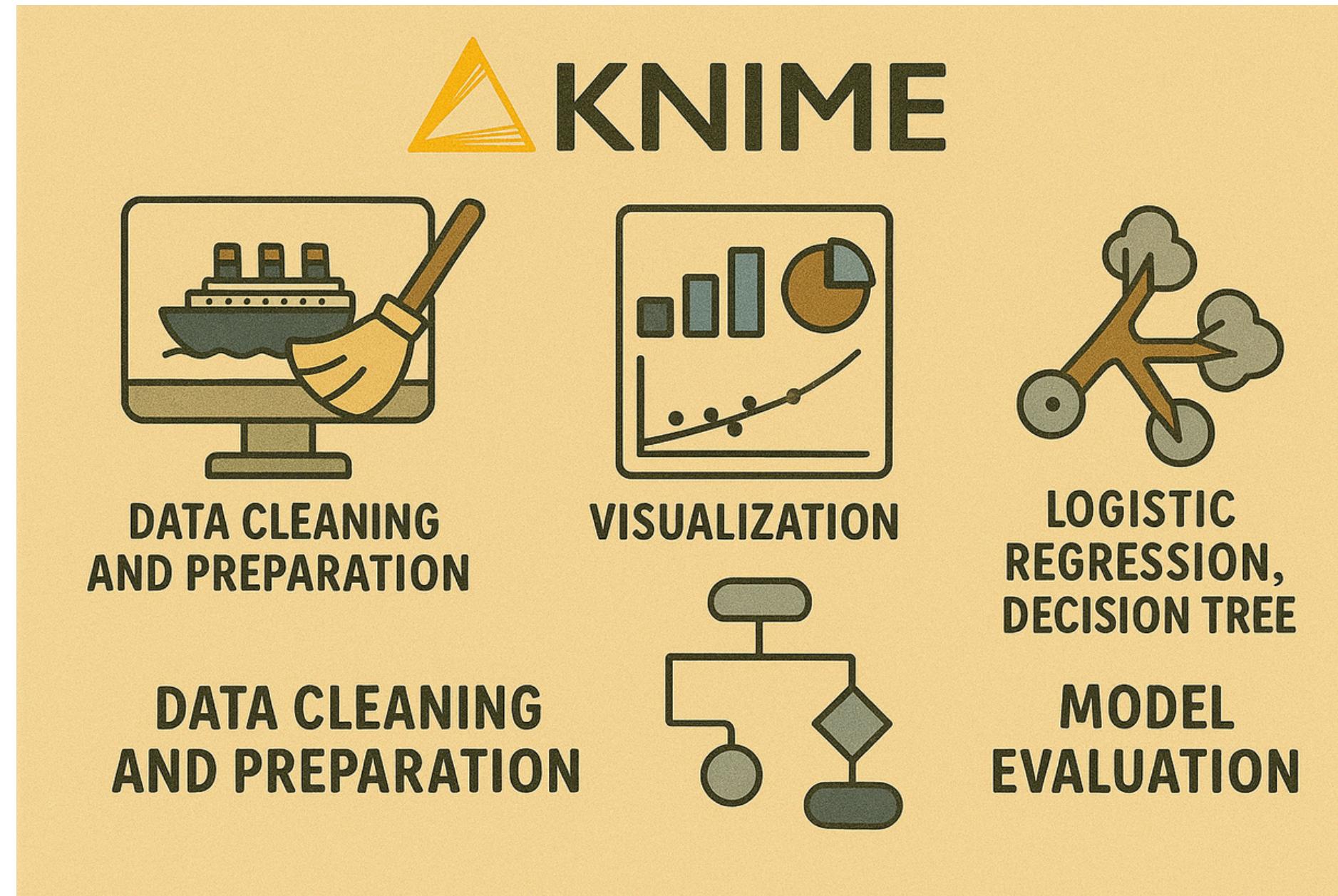


# Athlete and Medal Analysis with Olympic Data

The objective of this project is to predict whether an Olympic athlete is likely to win a medal based on historical data. The study includes visual and statistical analysis by country, sport, and gender, followed by classification using Logistic Regression and Decision Tree models.



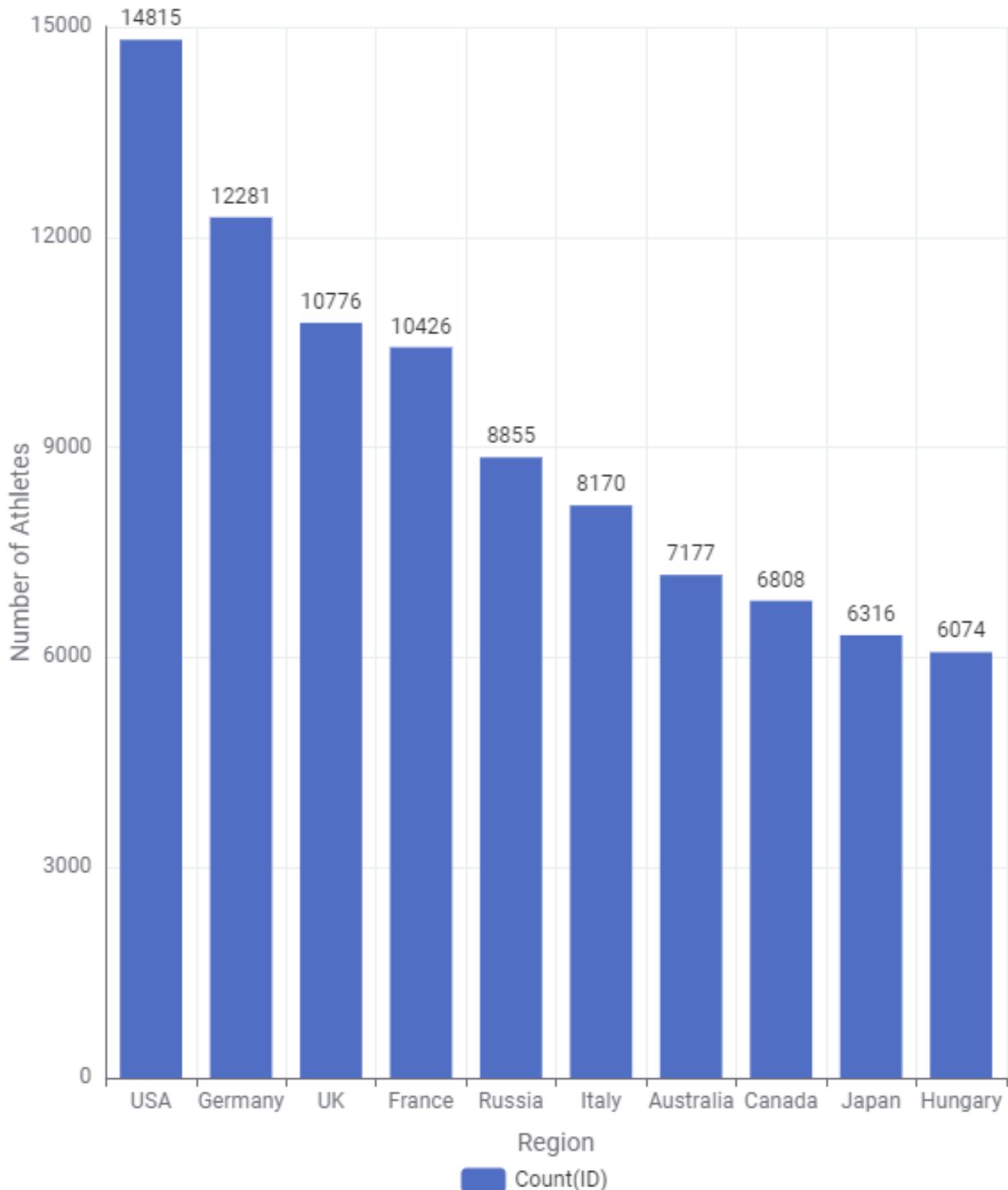
Most athletes represented developed countries such as

the USA, Germany, France, Russia, Italy, Japan, Canada,

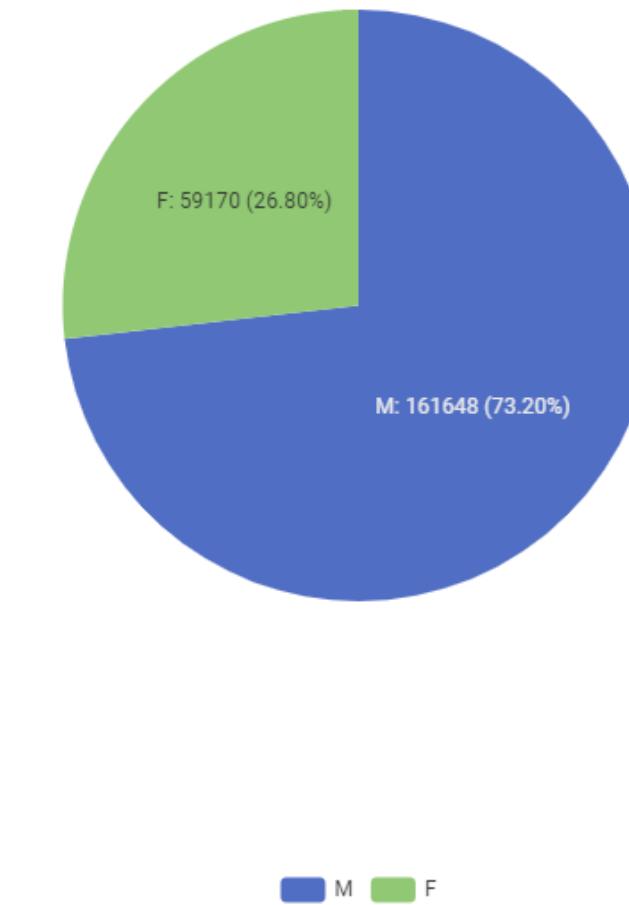
and the UK. These nations are dominant in Olympic

participation.

#### Top 10 Countries with most Athletes

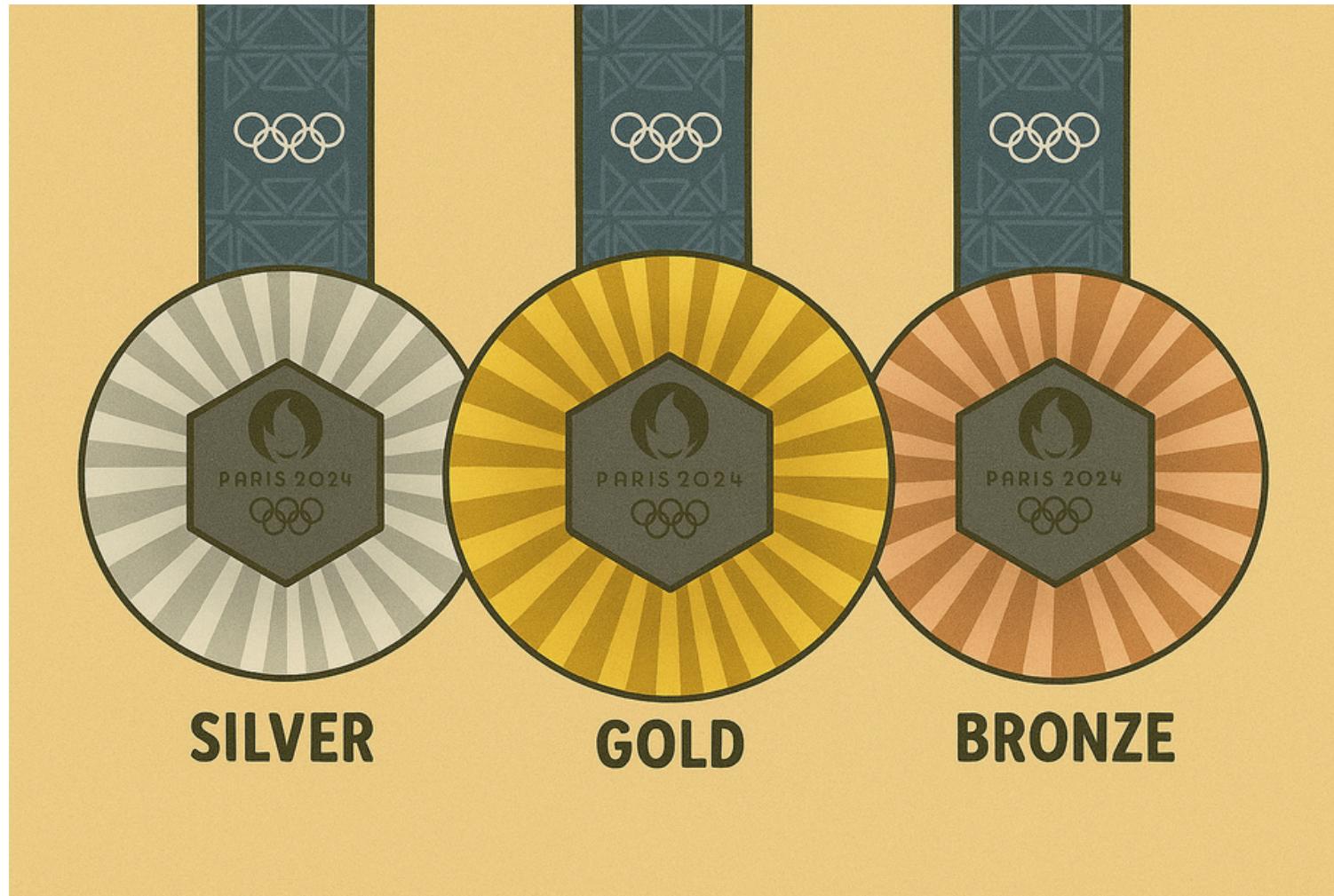


Gender Distribution Visualization



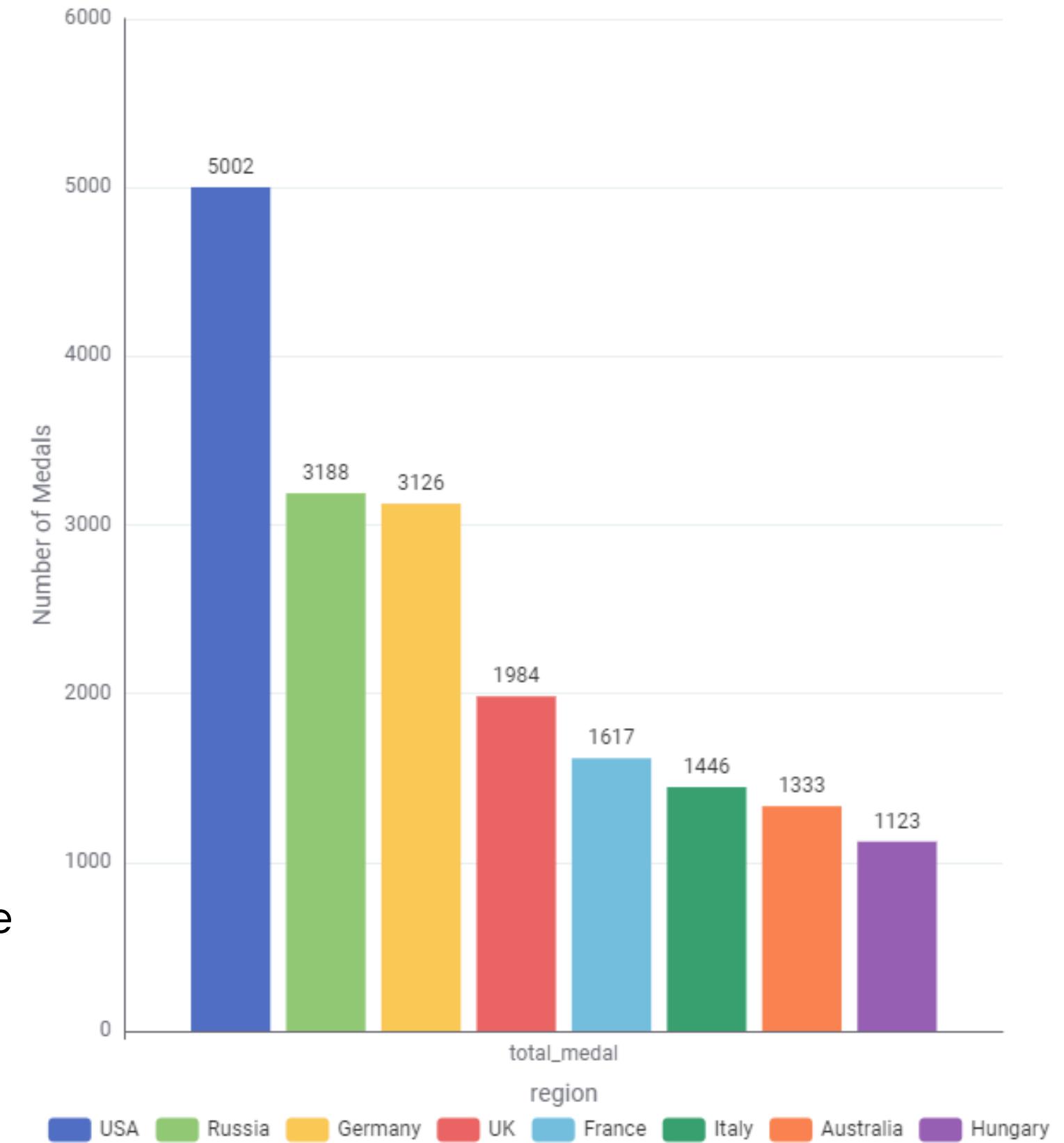
Approximately 75% of the athletes are male, while 25% are female. There is a significant gender gap in Olympic participation.

## Countries Won the Most Medals in the Olympics

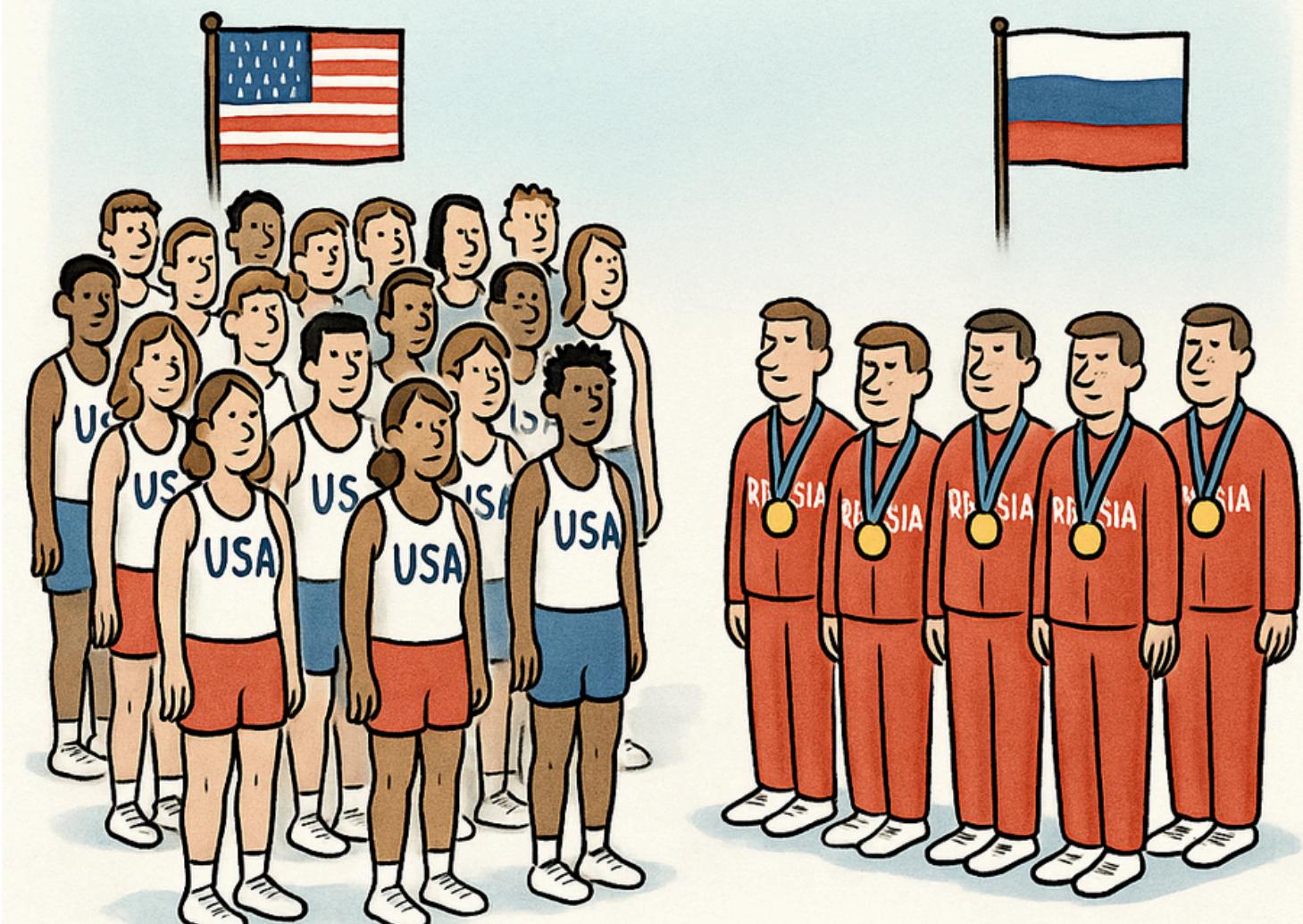


SILVER      GOLD      BRONZE

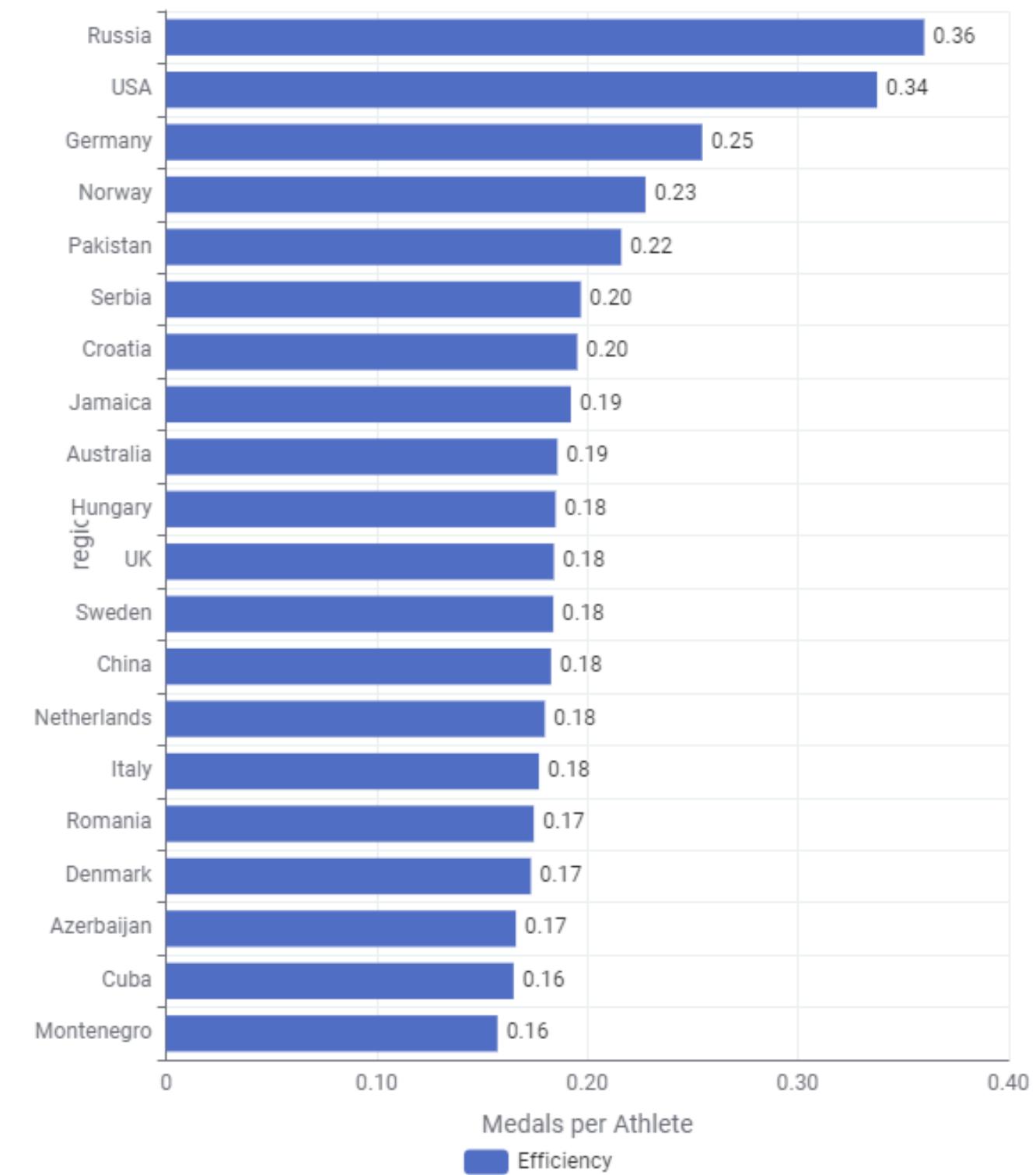
The USA, Russia, and Germany have won the most medals, followed by the UK, Italy, and Australia. These countries show strong performance across multiple games.



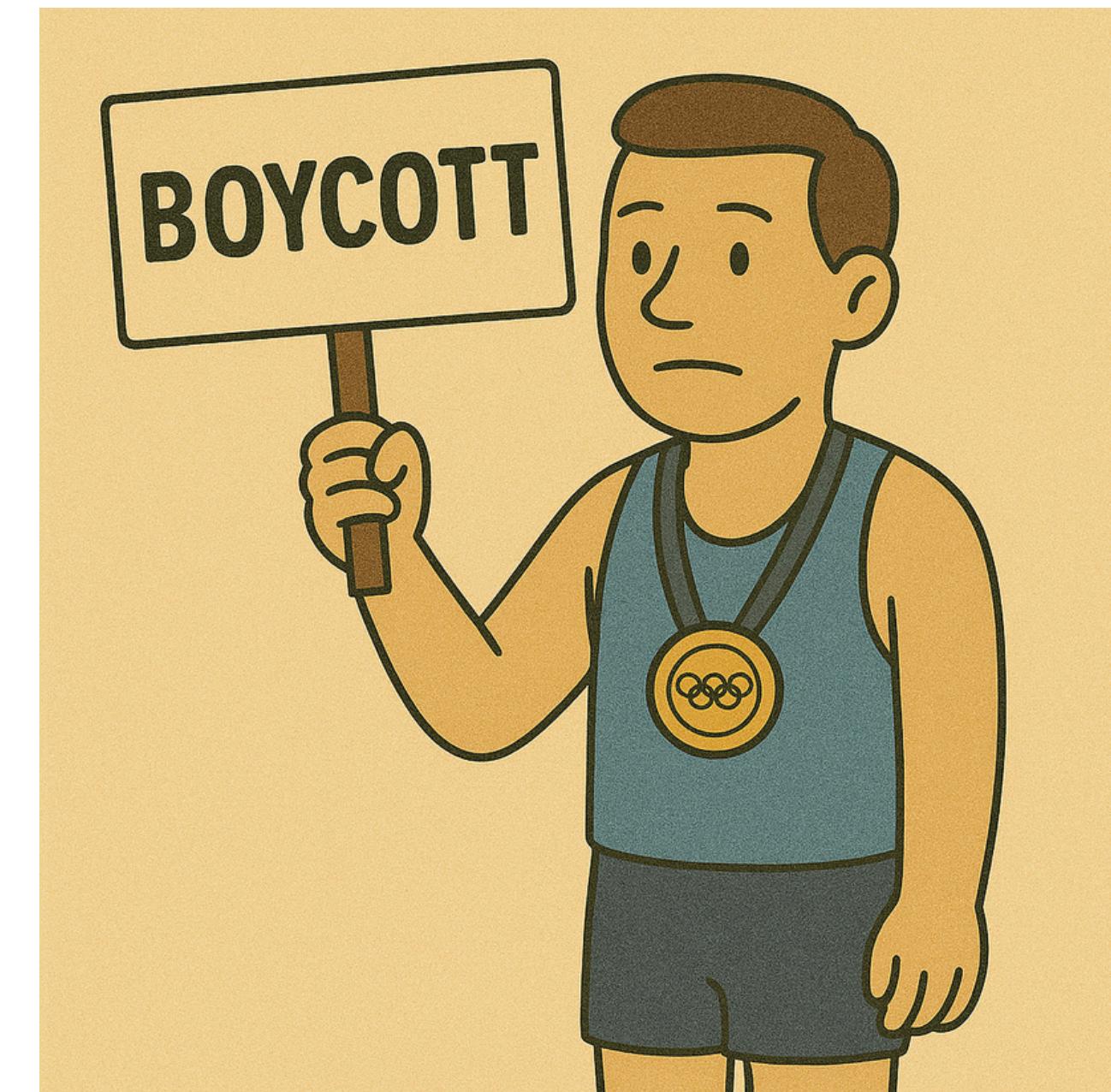
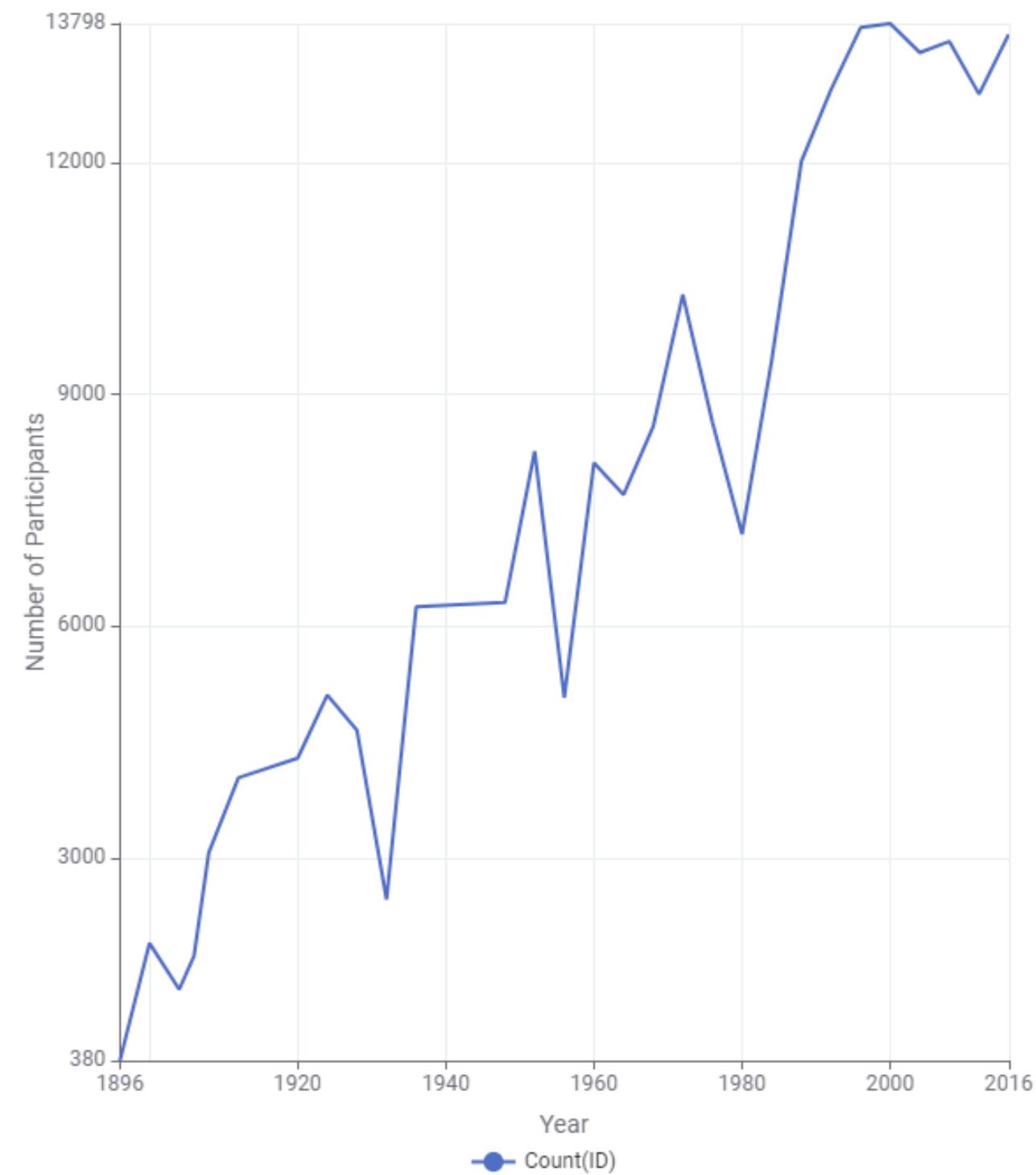
→ More athletes does not always mean more medals



Medal Efficiency by Country



Change in the Number of the Athletes Over the Years

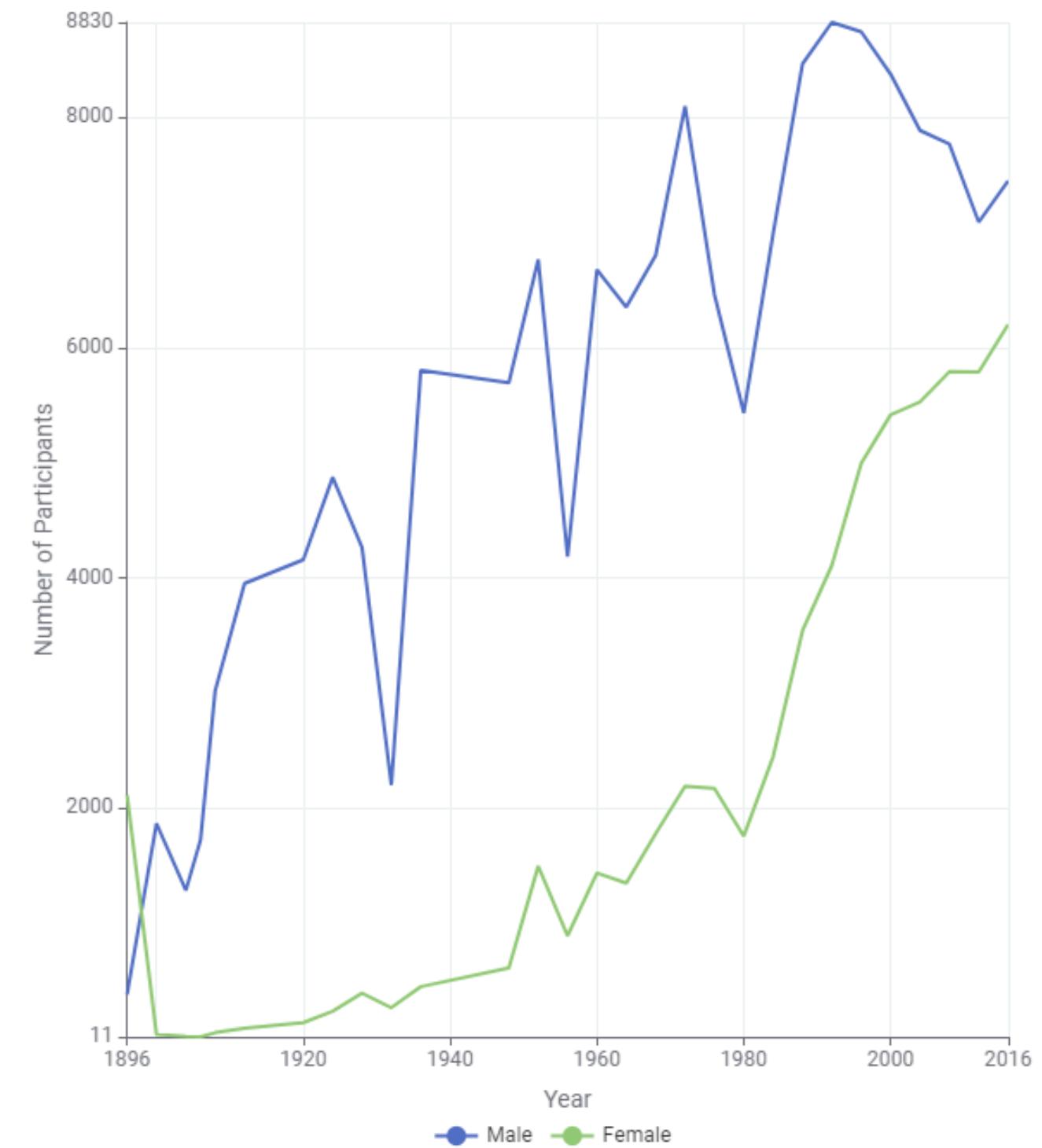


The total number of Olympic participants has generally increased over time. However, there were significant drops in 1932, 1956, and 1980—these anomalies may be linked to historical or political events.

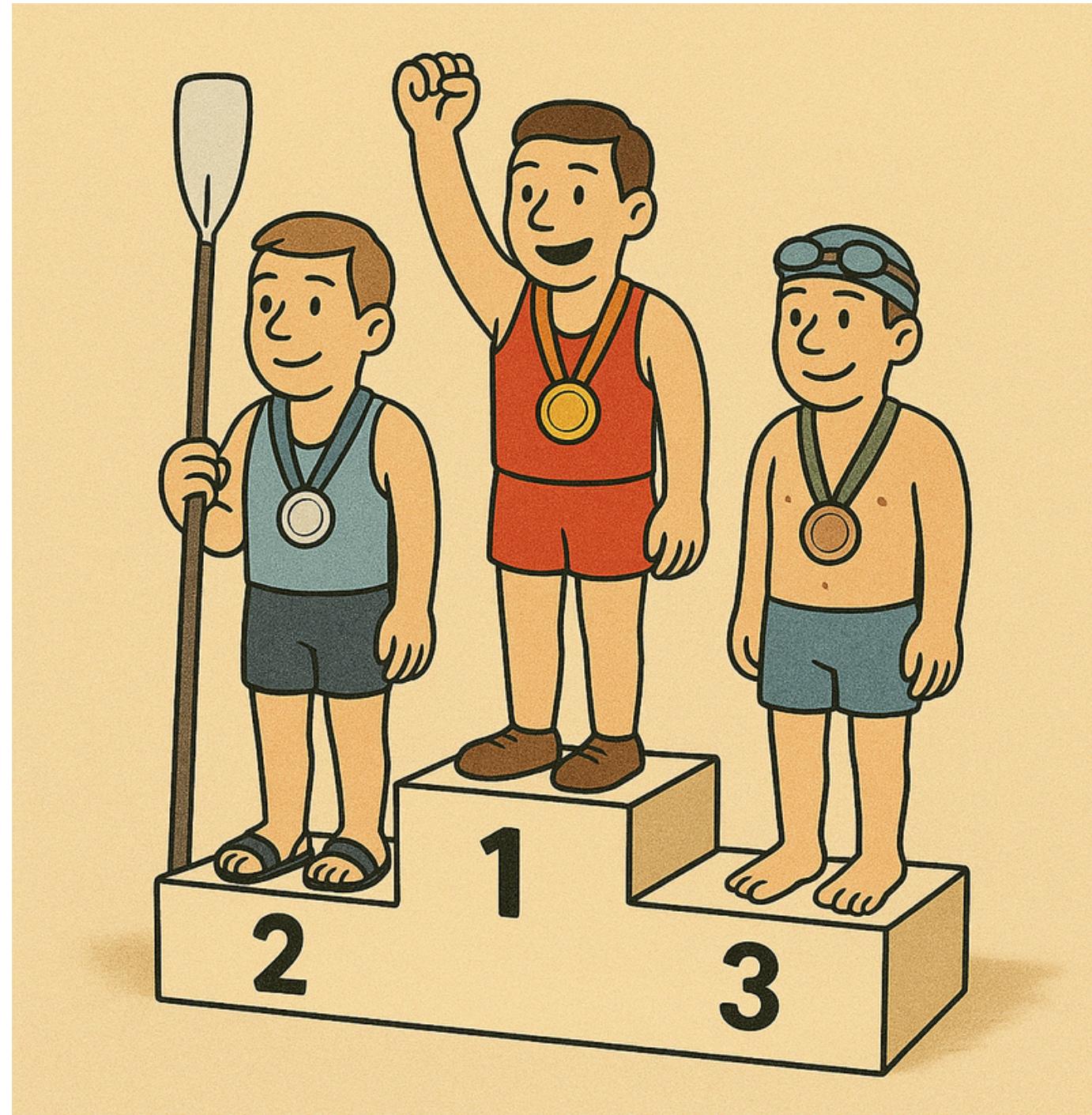




Changes in Female and Male Participants Over the Years



Male participation has always been higher, but female athlete participation has grown steadily and consistently over the years.



This heatmap shows in detail how many medals were won in each sport and at each medal level (Gold, Silver, Bronze).

It helps identify the most competitive sports and where the most medals are awarded.

How many medals were won in which sport and at which level?

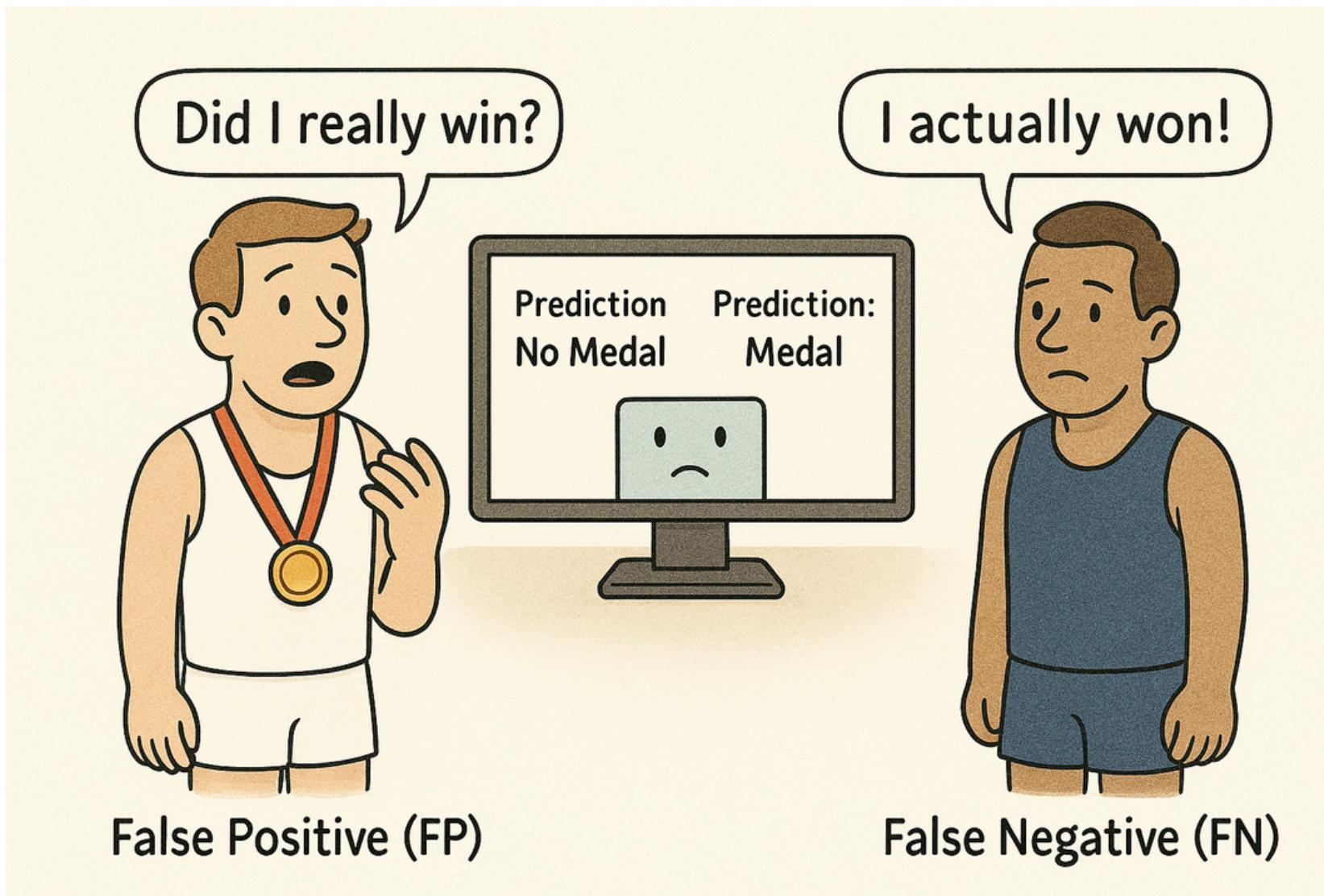
Sport	Gold	Bronze	Silver
Aeronautics	1	4	
Archery	91	89	96
Athletics	1031	1183	1196
Baseball	44	59	49
Basque Pelota	100	112	108
Boxing	299	322	318
Cricket	2	24	22
Cycling	230	428	244
Equestrianism	274	334	316
Figure Skating	12	2	12
Golf	3	2	2
Handball	341	366	384
Ice Hockey	100	123	118
Judo	234	271	258
Modern Pentathlon	406	459	467
Polo	9	9	9
Rhythmic Gymnastics	477	518	490
Rowing	15	16	15
Rugby Sevens	531	584	575
Shooting	302	324	325
Swimming	450	478	475
Table Tennis	8	8	11
Tennis	1	1	1
Triathlon	123	252	129
Volleyball	24	12	24
Weightlifting	50	56	54

Gold      Bronze      Silver

Medal



Actual Medal_Won	Predicted	Meaning	Count
1 → 1	True Positive (TP)	Correctly predicted as winning a medal	3867
1 → 0	False Negative (FN)	Actually won but predicted as not winning	2942
0 → 1	False Positive (FP)	Didn't win but predicted as won	2316
0 → 0	True Negative (TN)	Correctly predicted as not winning	4503



**Overall Accuracy: 61.4%**  
**This means the model correctly predicted the outcome in 61.4% of the cases.**

Metric	Meaning
Recall (1)	Ability to detect actual medal winners 0.568 ( $\approx 57\%$ )
Precision (1)	How many predicted winners actually won 0.625 ( $\approx 63\%$ )
Recall (0)	Correct detection of those who didn't win 0.660 ( $\approx 66\%$ )
Precision (0)	How accurate we are when predicting no medal 0.605 ( $\approx 61\%$ )
F1-Score (1)	Harmonic mean of precision and recall (1) 0.595 ( $\approx 60\%$ )
F1-Score (0)	Harmonic mean of precision and recall (0) 0.631 ( $\approx 64\%$ )

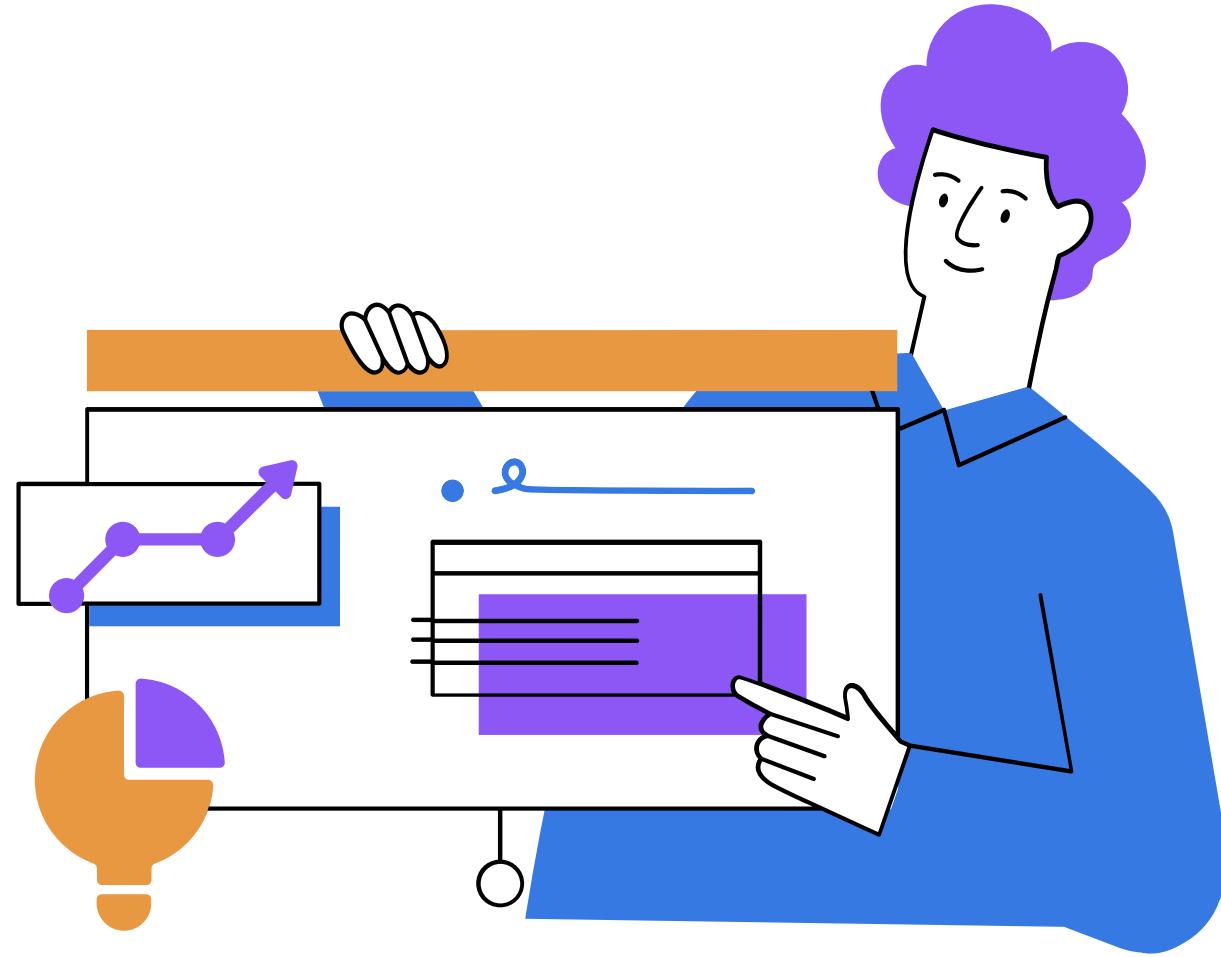
Medal Won	Prediction	Meaning
1 → 1	True Positive (TP)	<b>4059</b> athletes won a medal and were correctly predicted.
1 → 0	False Negative (FN)	<b>2630</b> athletes won a medal but were incorrectly predicted as not having won.
0 → 1	False Positive (FP)	<b>2842</b> athletes did not win a medal but were incorrectly predicted as having won.
0 → 0	True Negative (TN)	<b>4097</b> athletes did not win a medal and were correctly predicted.

Metric	Meaning	Value
Recall (1)	Proportion of actual medal winners correctly predicted	0.607 ( $\approx 61\%$ )
Precision (1)	Proportion of predicted winners who actually won	0.588 ( $\approx 59\%$ )
F1-Score (1)	Harmonic mean of precision and recall	0.597 ( $\approx 60\%$ )





- Extend the dataset to include more recent Olympic events and athletes.
- Analyze the correlation between a country's GDP, population, and Olympic performance.
- Use machine learning models to predict medal counts based on athlete or country data.
- Create interactive dashboards for real-time exploration of the data.
- Include Paralympic and Winter Olympics data for broader comparison.
- Improve gender-based analysis with more detailed demographic variables.



# Thanks

- Deniz Drahyalı - Batuhan Unal
- hello@dbanalytics.com
- www.dbanalytics.com