**🏅 Olympic Games Data Analysis (1896–2016)**

***Historical Analysis of Athlete Performance***

**📌 1. Executive Summary**

Analysis of **271,116 athlete records** spanning over a century reveals key patterns:

* 🇺🇸 **United States** dominates in both participation (**17,847 athletes**) and gold medals (**2,638**).
* 👨 **Male athletes** constitute **72.5%**, but 👩 **female participation** has steadily risen.
* 🧓 Most athletes are **aged 20–30**, with rare medalists over 60 (all in 🎯 shooting/archery).

**📂 2. Dataset Overview**

**Datasets Used:**

* athlete\_events.csv – Individual athlete performances
* noc\_regions.csv – Country codes and regions

**Key Variables:**

python

CopyEdit

['ID', 'Name', 'Sex', 'Age', 'Height', 'Weight', 'Team', 'NOC',

'Games', 'Year', 'Season', 'City', 'Sport', 'Event', 'Medal',

'region', 'Notes']

**Data Notes:**

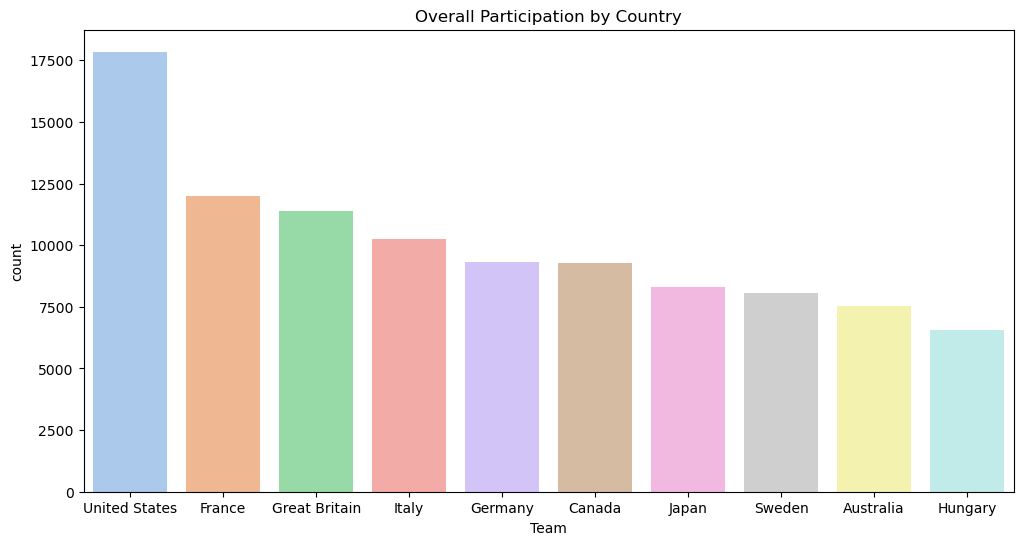
* 📉 Missing values: ~22% athletes lack Height or Weight.
* ❌ **231,333** entries have no medal (non-medal participants).

**📊 3. Key Insights**

**📈 3.1 Participation Trends**

**Top 10 Countries by Athlete Count:**

| **Country** | **Athletes** |
| --- | --- |
| 🇺🇸 United States | 17,847 |
| 🇫🇷 France | 11,988 |
| 🇬🇧 Great Britain | 11,404 |

📊 *Visualization:*****

**Gender Breakdown:**

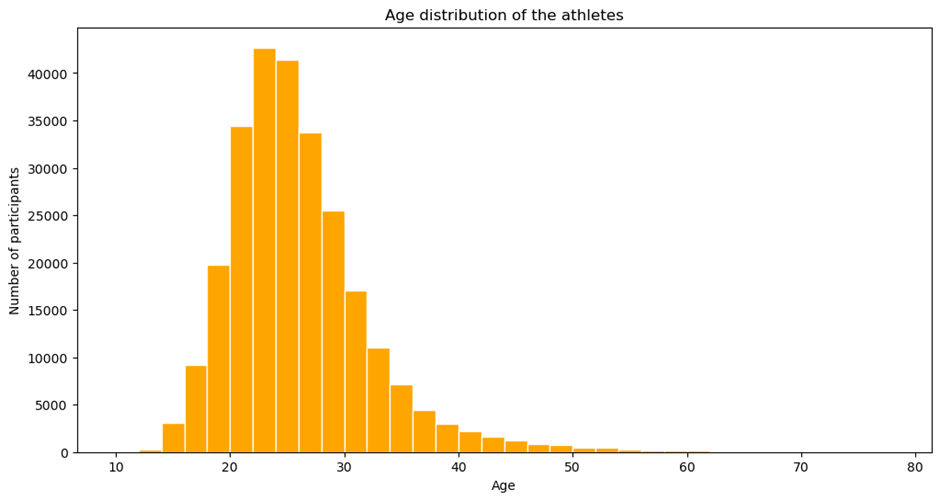
python

CopyEdit

Male: 196,594 (72.5%)

Female: 74,522 (27.5%)

📊 *Visualization:*

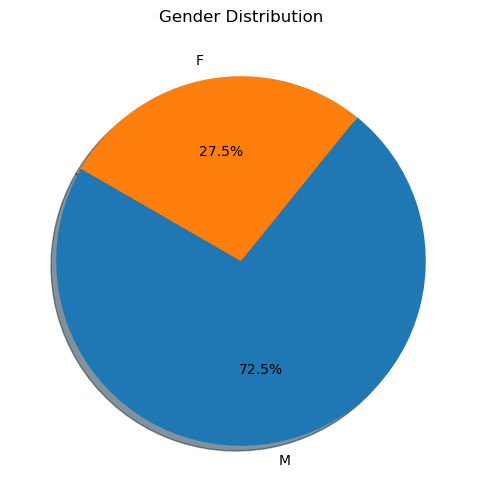
**

**🥇 3.2 Medal Analysis**

**All-Time Gold Medal Leaders:**

| **Country** | **Gold Medals** |
| --- | --- |
| 🇺🇸 USA | 2,638 |
| 🇷🇺 Russia | 1,599 |
| 🇩🇪 Germany | 1,301 |

📊 Visualization:



**2016 Rio Olympics Top Performers:**

python

CopyEdit

United States: 137 gold

Great Britain: 64 gold

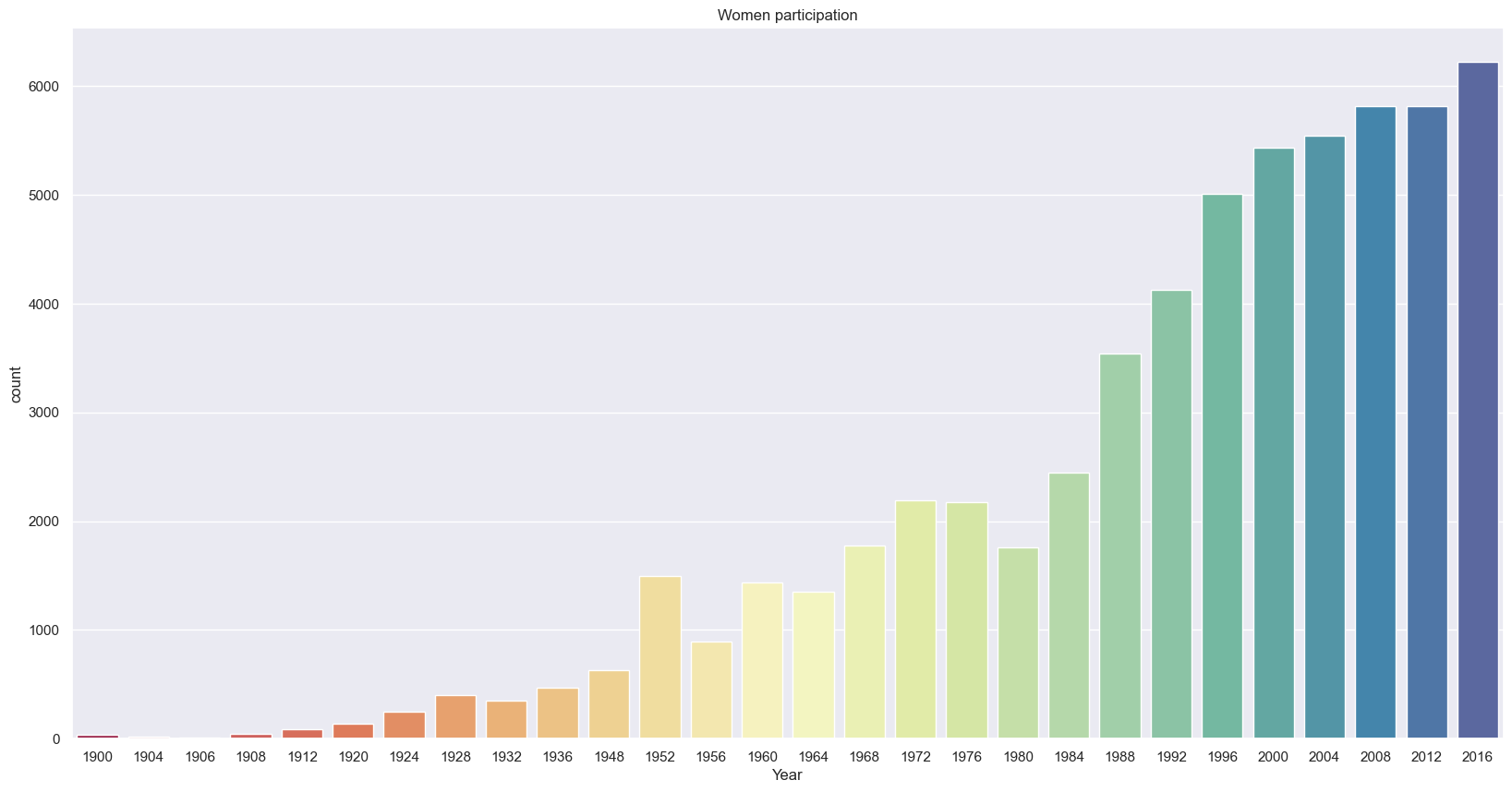
Russia: 50 gold

**🧬 3.3 Athlete Demographics**

**Age Distribution:**

* 📊 Average Age: **25.6 years**
* 👴 Oldest competitor: **97 years**
* 🥇 Medalists over 60: **6**, all in shooting/archery

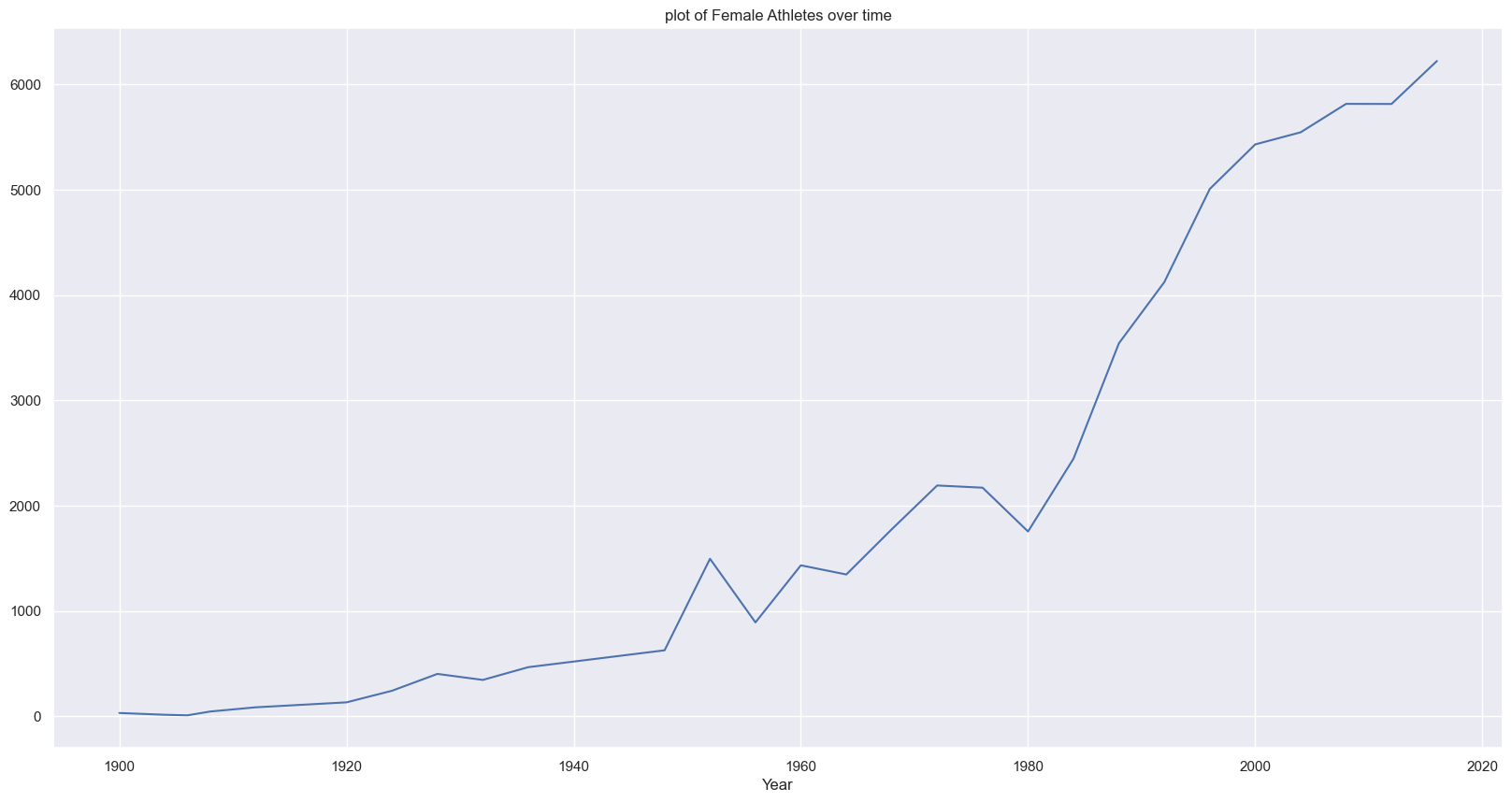
📊 Visualization:



**Height vs Weight Patterns:**

* Strong **gender-based clustering** visible

📊 Visualization:

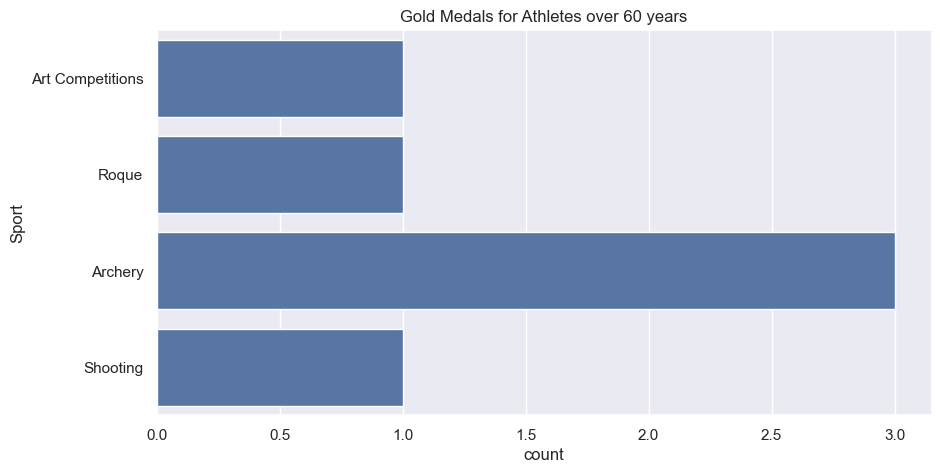


**🚺 3.4 Female Participation Growth**

**Female Athletes Over Time:**

| **Year** | **Female Athletes** |
| --- | --- |
| 2000 | 5,431 |
| 2016 | 6,223 (+14.6%) |

📊 *Visualization:*



**🧪 4. Technical Appendix**

**Data Preparation:**

python

CopyEdit

# Merge athlete and region data

athletes\_df = athletes.merge(regions, how='left', on='NOC')

# Handle missing values

null\_columns = ['Age', 'Height', 'Weight', 'Medal', 'region', 'Notes']

athletes\_df[null\_columns] = athletes\_df[null\_columns].fillna('Not Recorded')

**Gold Medal Analysis:**

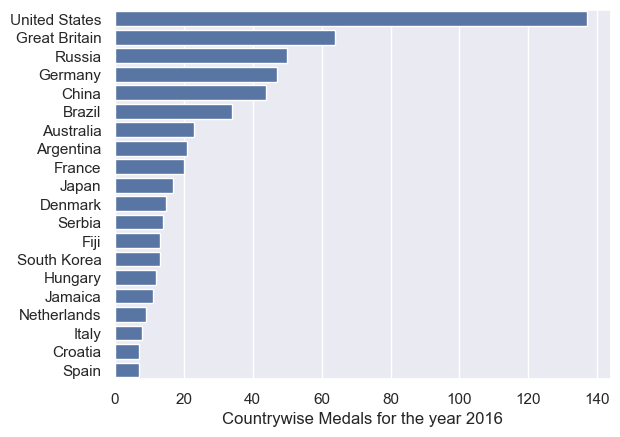
python

CopyEdit

goldMedals = athletes\_df[(athletes\_df.Medal == 'Gold') &

(athletes\_df.Age.notna())]

📊 *Visualization:*



**Visualization Example:**

python

CopyEdit

plt.figure(figsize=(12,6))

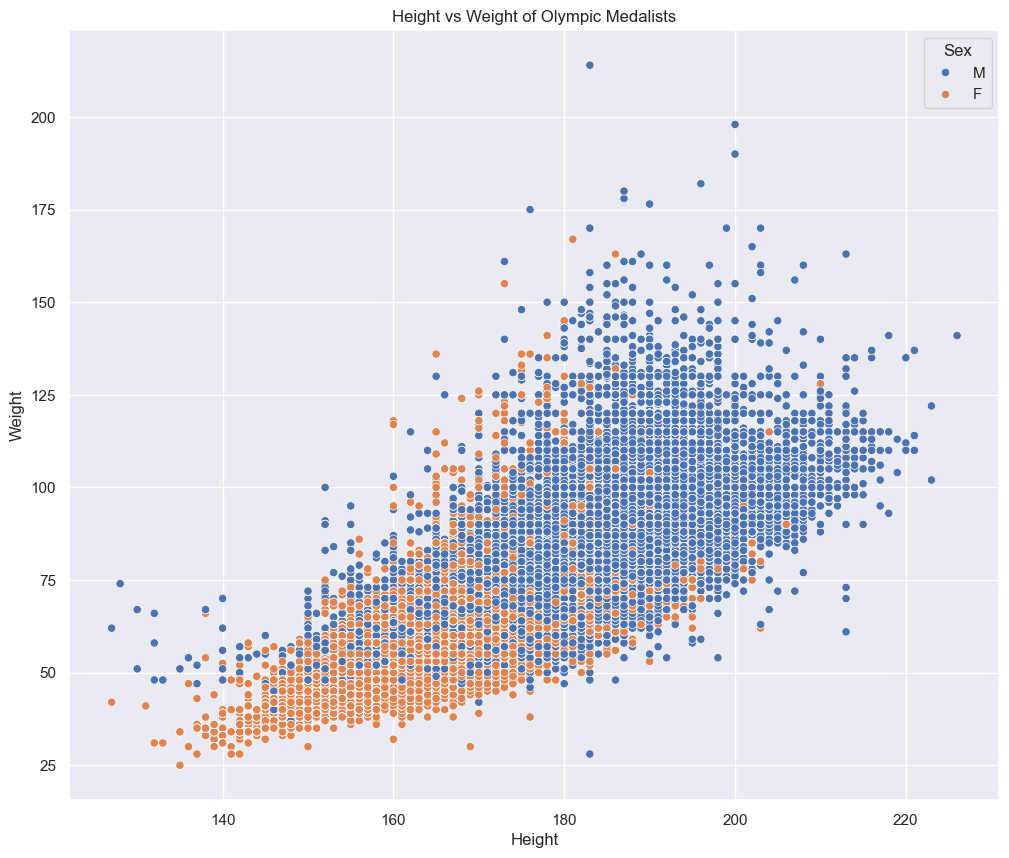
sns.barplot(x=top\_10\_countries.index,

y=top\_10\_countries.values,

palette='viridis')

plt.title('Top 10 Participating Countries', fontsize=16)

📊 *Visualization:*



**✅ 5. Recommendations**

1. 🔬 **Analyze sport-specific age limits** – especially for older sports like shooting/archery.
2. 👩‍🎓 **Assess the impact of gender equality initiatives** post-2000.
3. 📏 **Compare height/weight by sport** for athletic profile optimization.