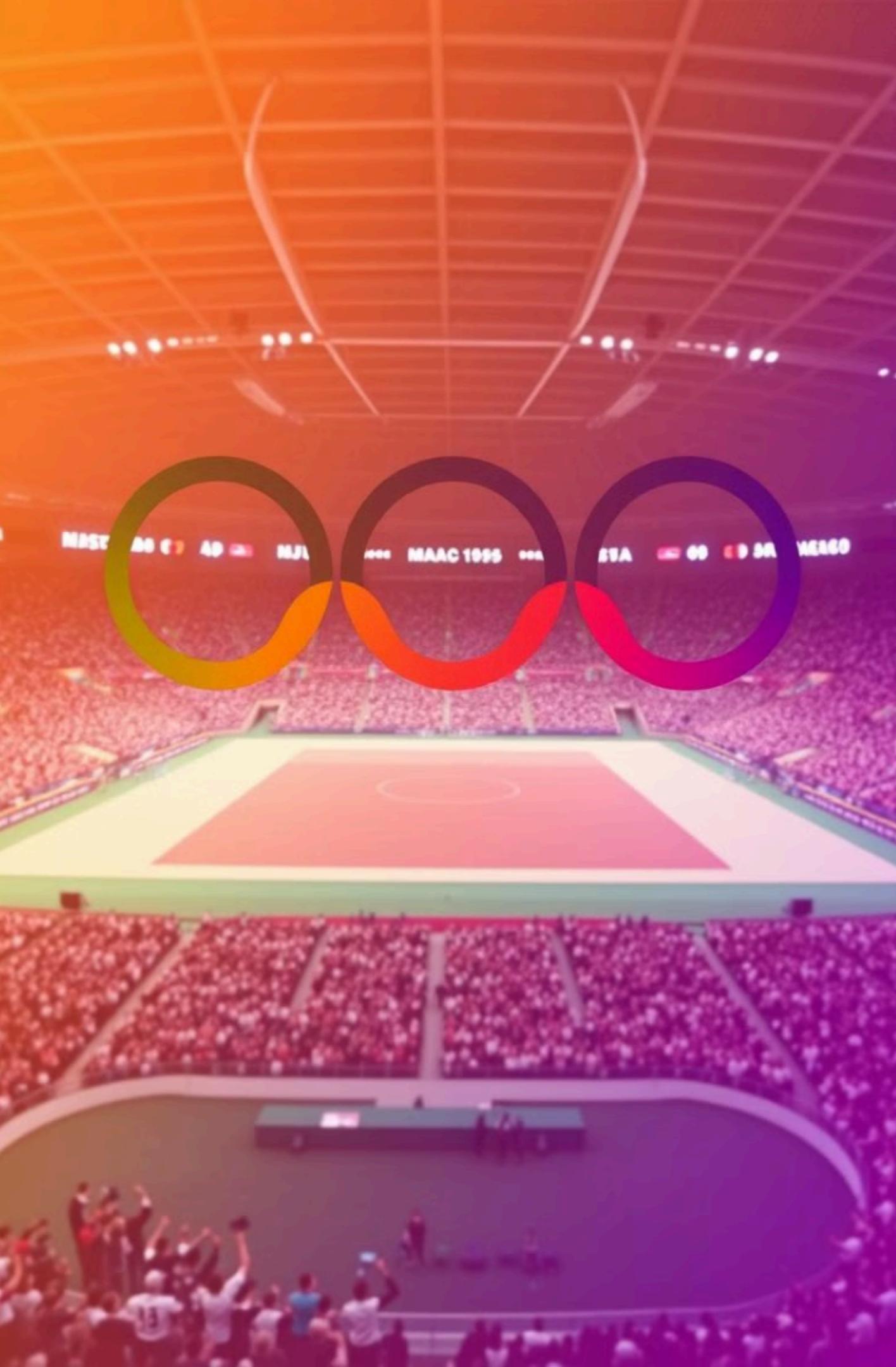


Olympics Data Analysis

Data-Driven Insights into Olympic Performance (1976 - 2008)

 by Dhruv Singh





Introduction to the



Data Exploration

Olympic data from 1976 to 2008.



Key Focus

Trends, top performers, athlete performance.



Analysis Areas

Medal distribution, gender participation, country trends.



Predictive Analysis

Estimate medal chances based on factors.

Technologies & Tools

Programming

Python

Libraries

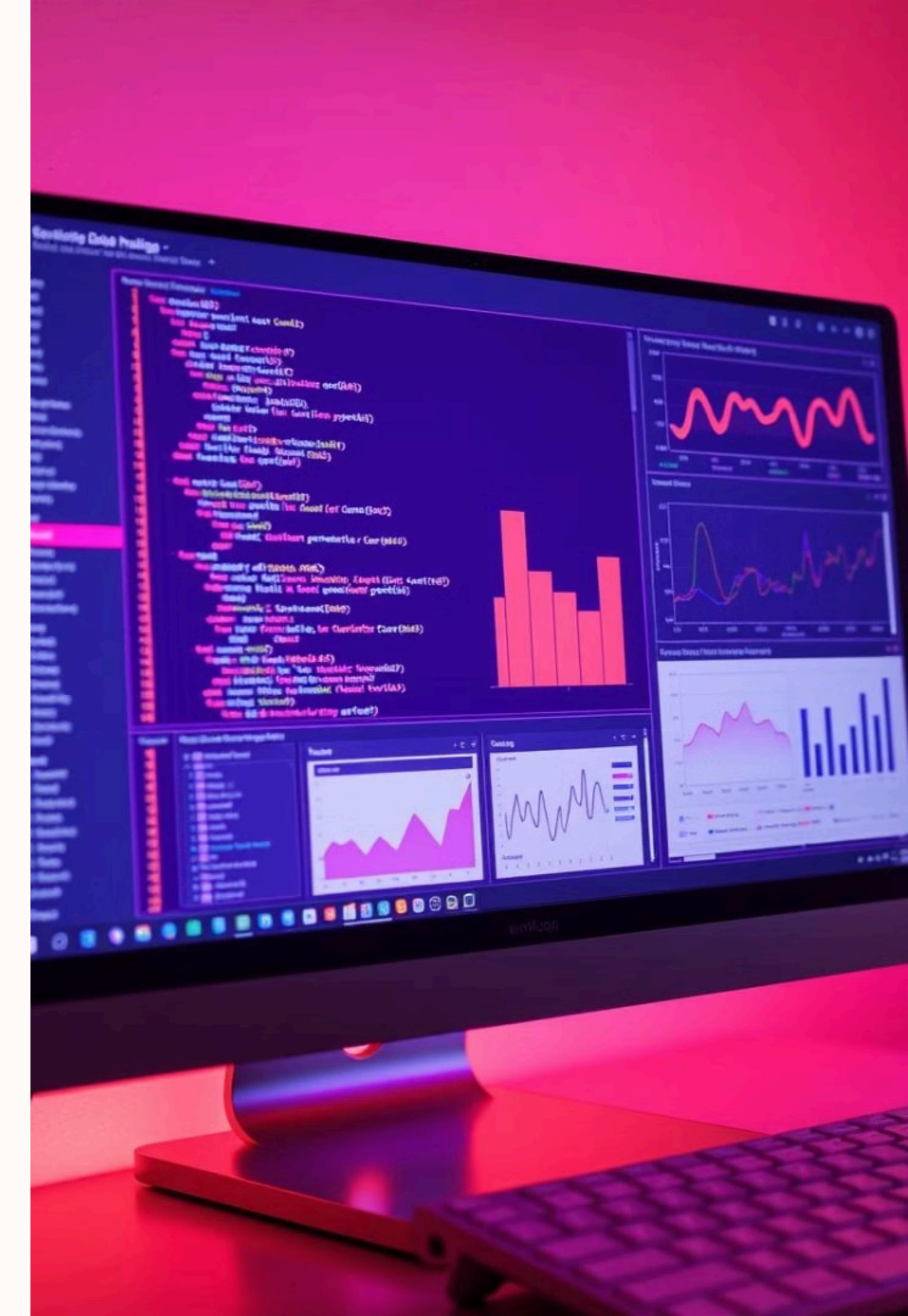
Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn

Data Analysis

SQL, Excel

Machine Learning

Logistic Regression



Dataset Overview

Data Scope

Olympic medal winners from 1976 to 2008.

Key Columns

- City
- Year
- Sport & Discipline
- Athlete Name
- Country
- Medal
- Gender

Data Cleaning & Preprocessing

Missing Values

Removed null values from key columns.

Data Standardization

Converted data types, removed duplicates, formatted names.

Feature Engineering

Added features like total medals, gender ratio, athlete performance.

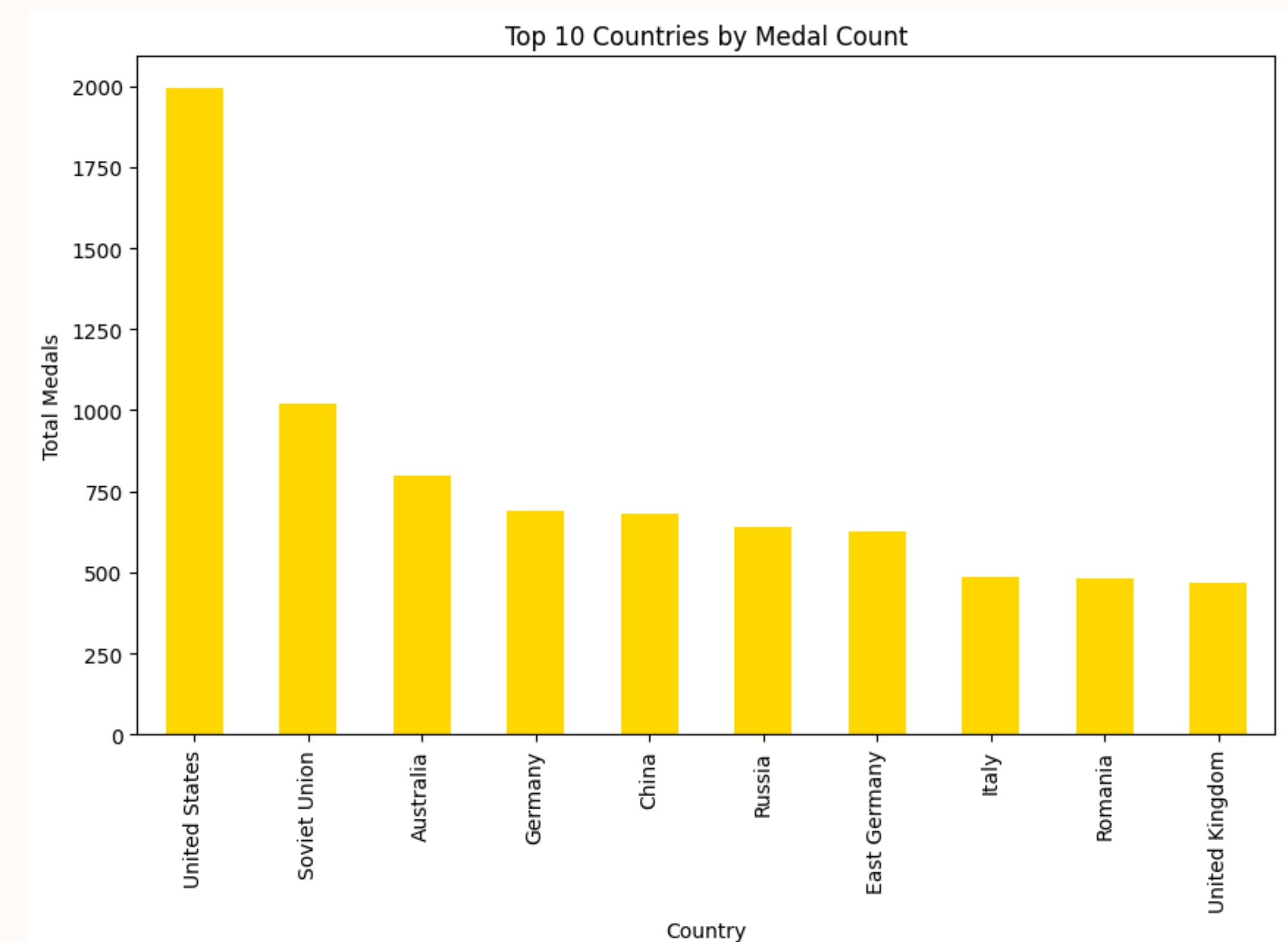
DATA CLEANING

Mationas oning tr remet and magi cleang.



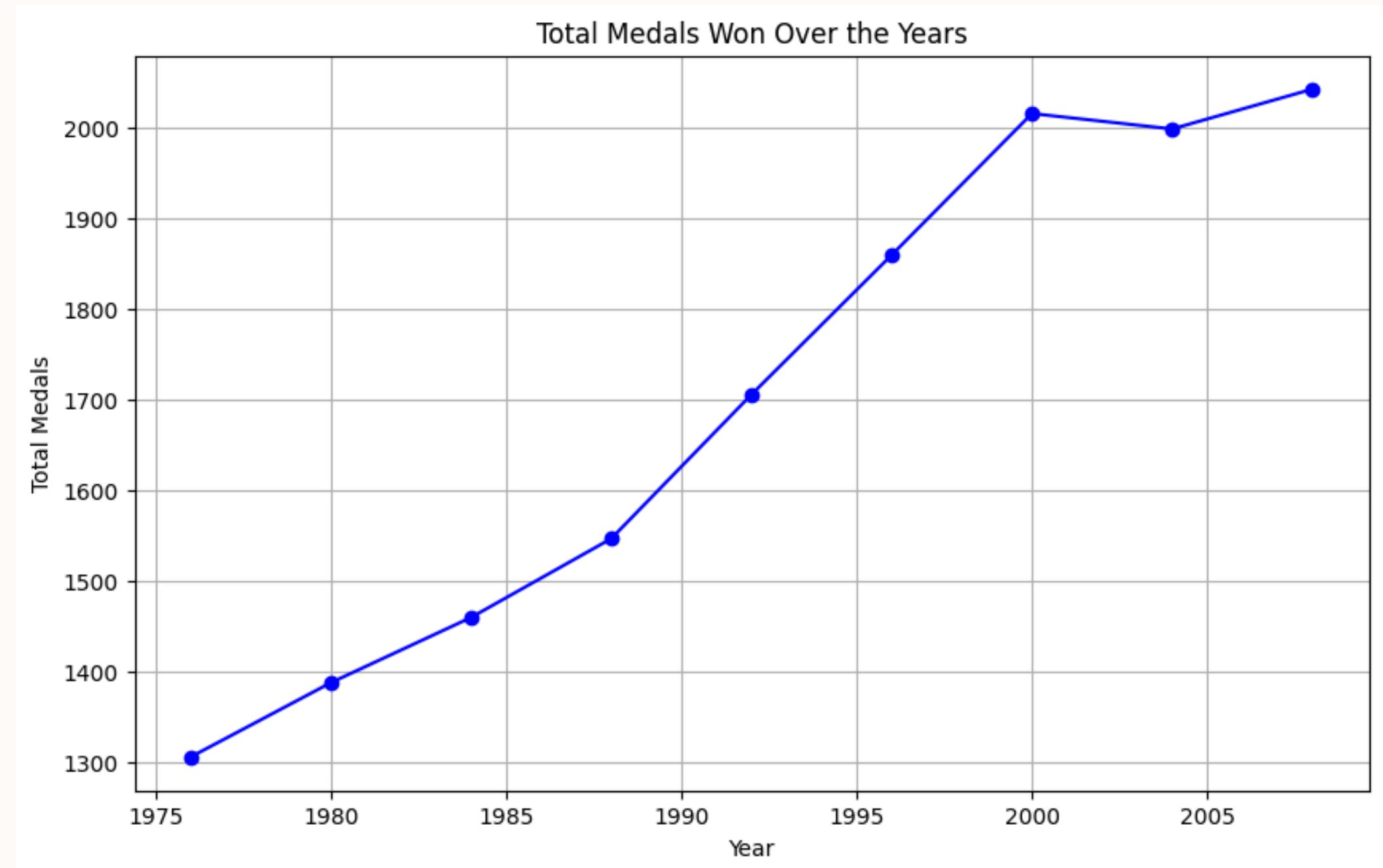
Medal Distribution by Country

Key Insights : The United States, Soviet Union, and China dominate the medal tally.



Medals Won Over the Years

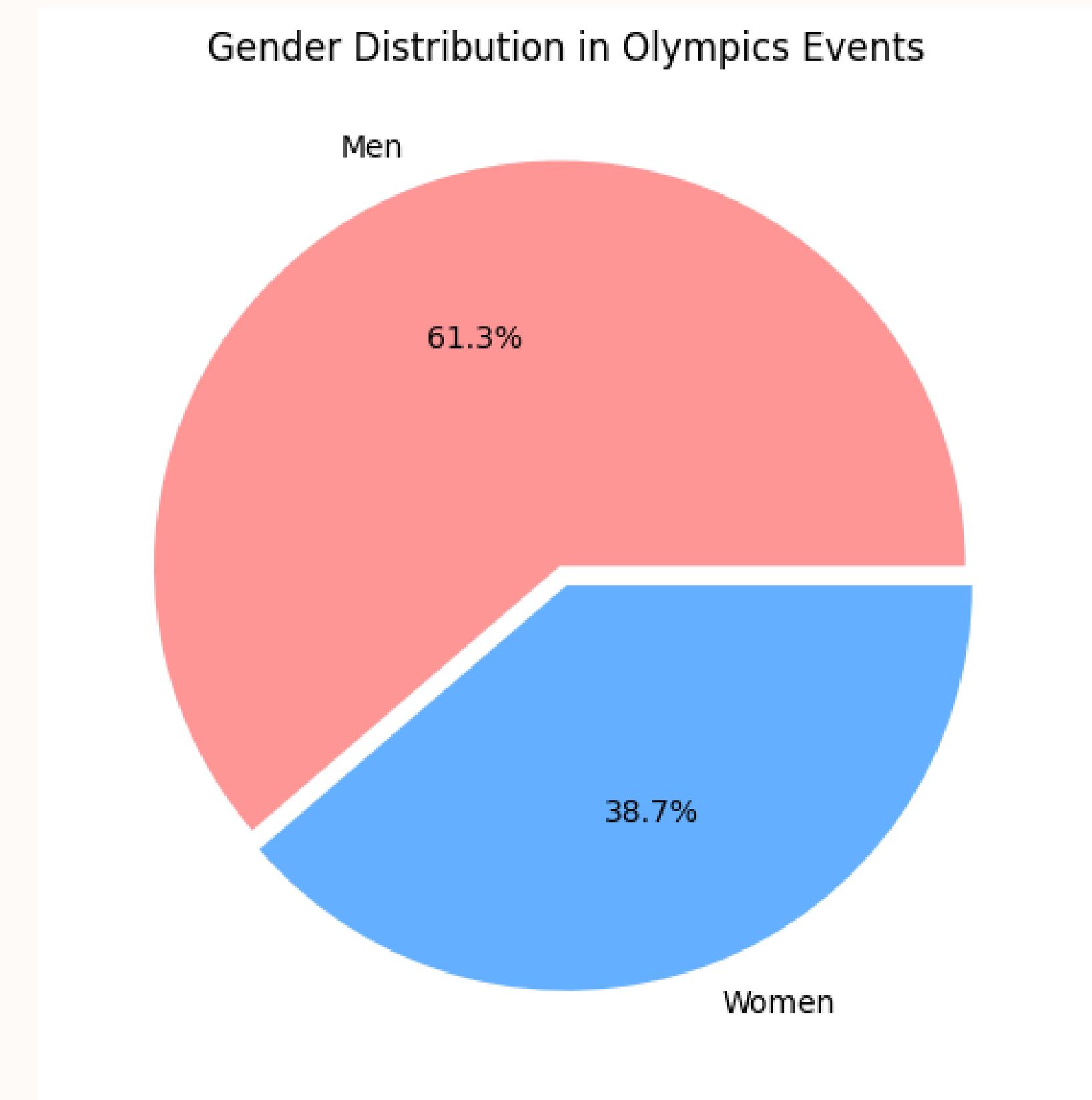
Key Insights : A spike in medals in some years due to increased participation. Some countries saw a decline post-1990 due to political factors.



Gender Distribution over Events

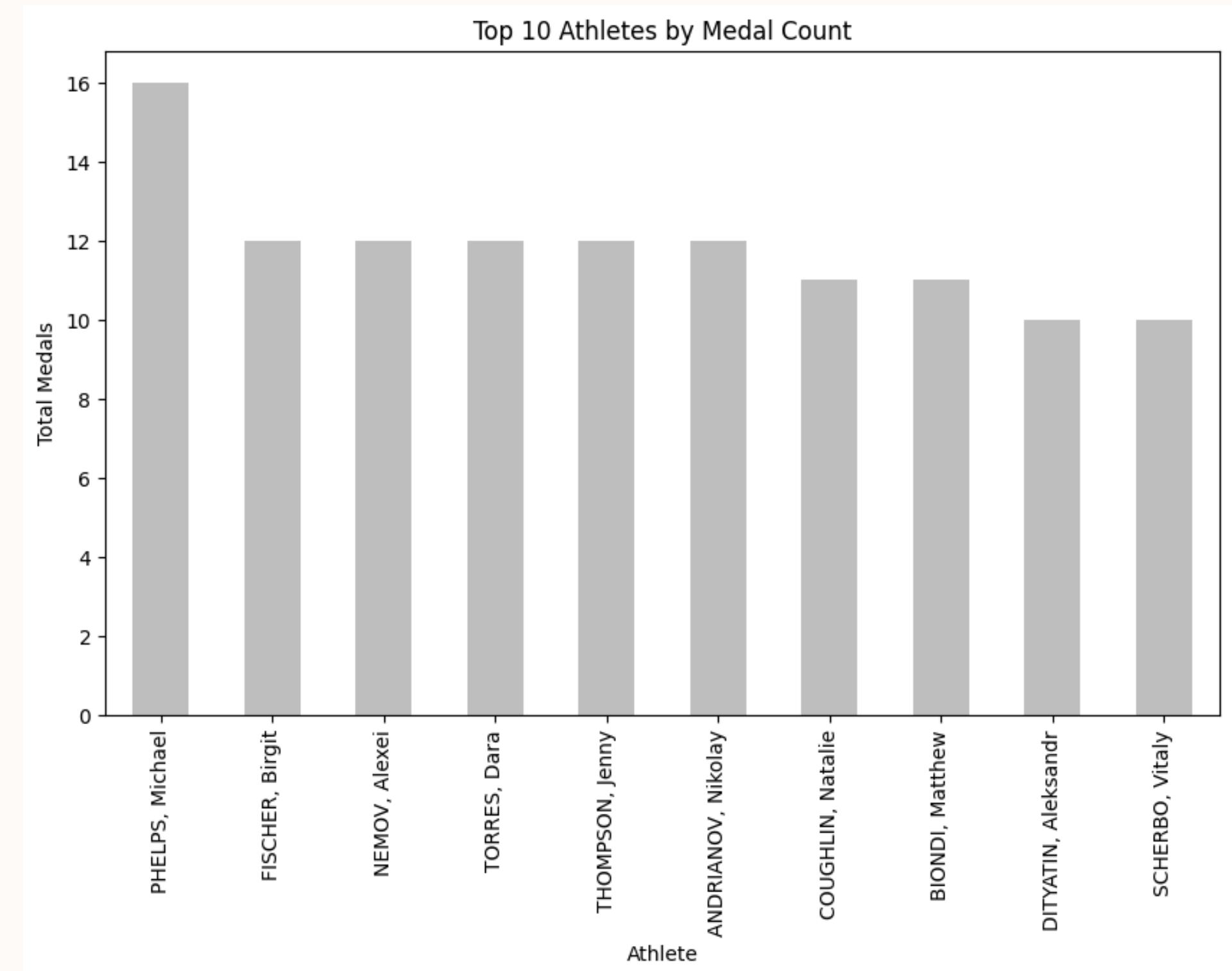
Key Insights :

- The number of female participants increased significantly after 1980.
- Some events are still male-dominated, while others have balanced participation.



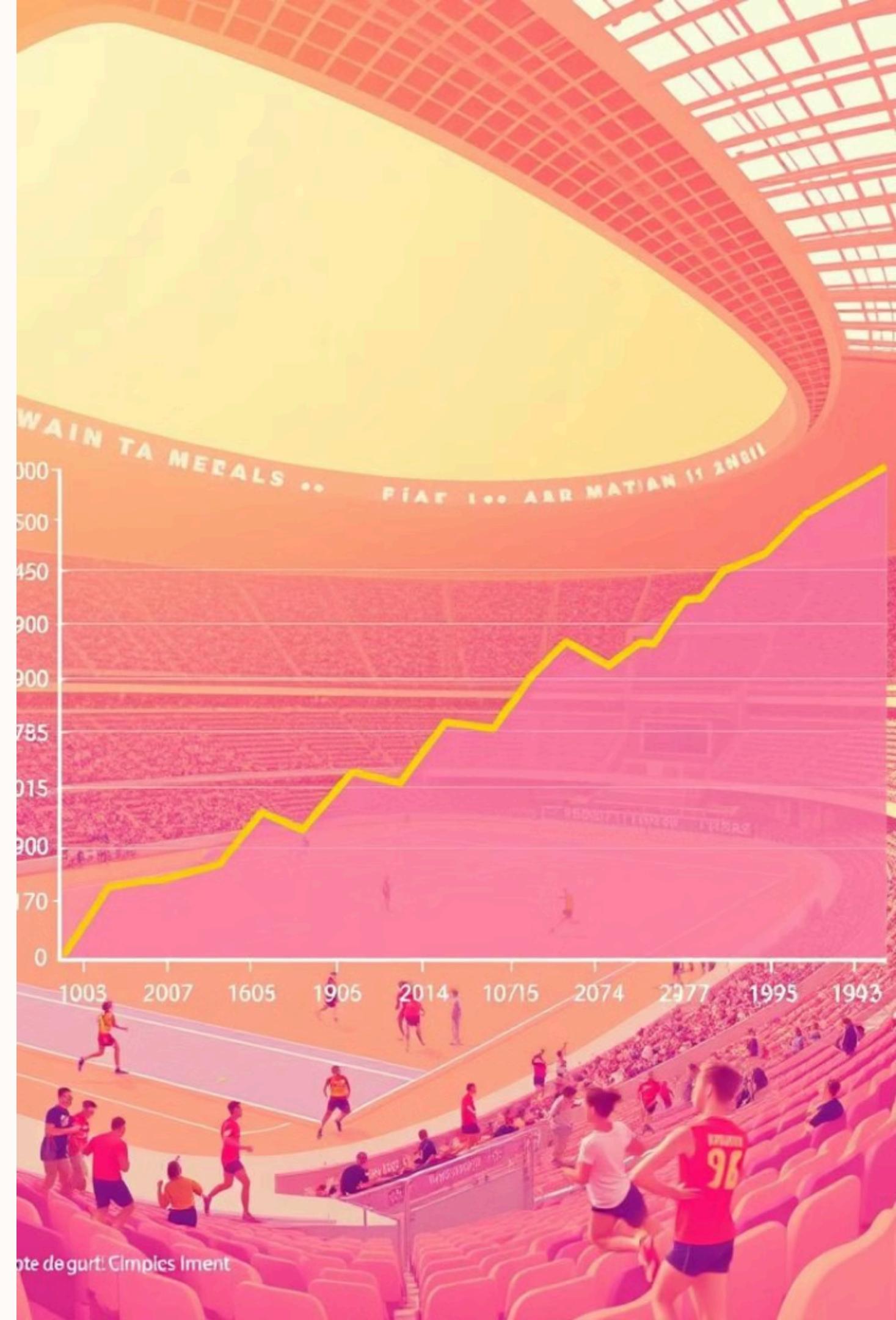
Top Athletes in Olympic History

Key Insights :
Michael Phelps has dominated swimming. Gymnastics and Athletics contribute significantly to medal wins.



Trends in Medal Wins Over the Years

- 1 **Medal Spike**
Increased participation in some years.
- 2 **Post-1990 Decline**
Political factors influenced some countries.



Can We Predict Medal Winners?



Logistic Regression

Predict medal wins based on factors.



Features Used

Country, sport, gender, event type.



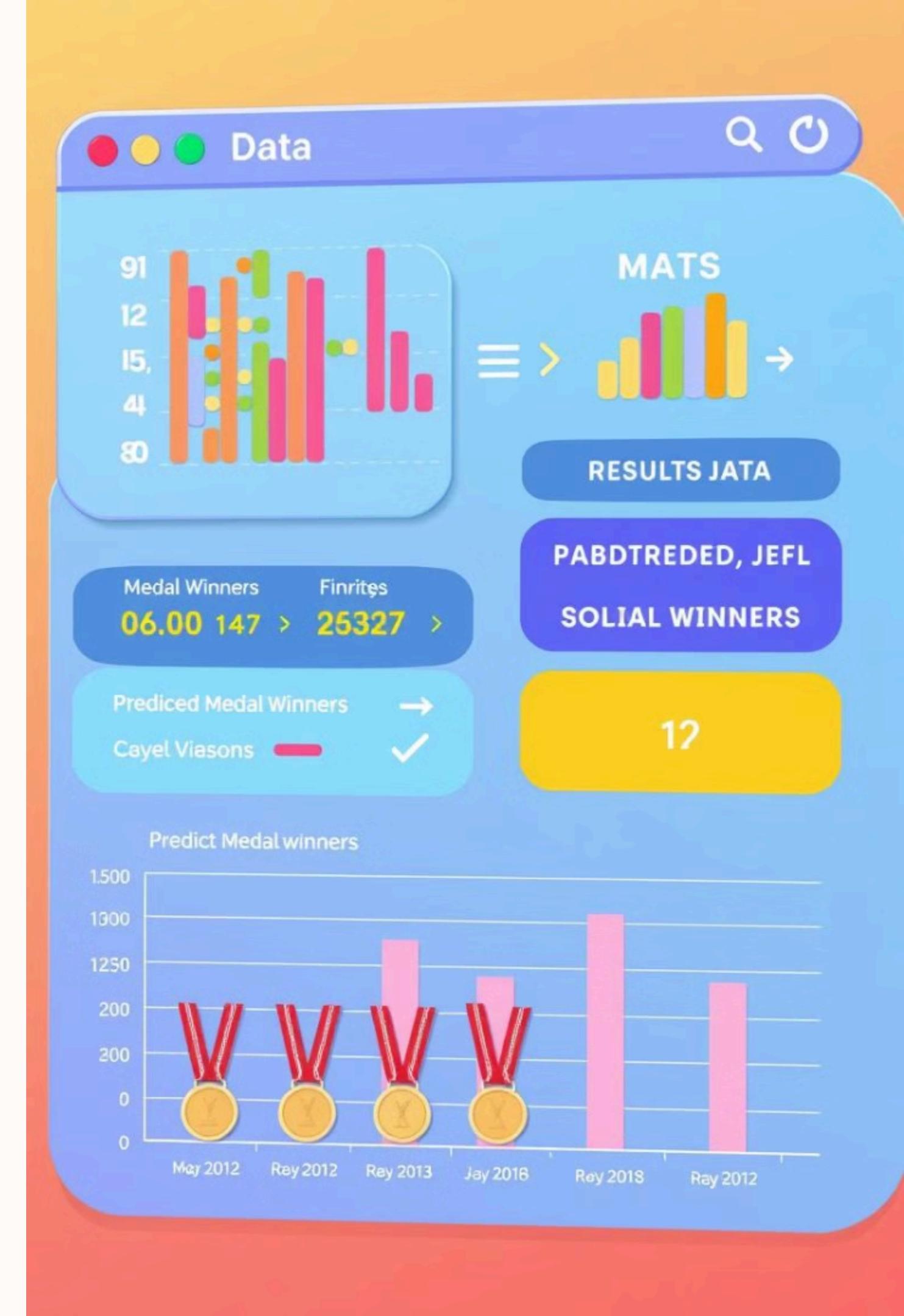
Accuracy

75% accuracy achieved.



Key Finding

Country and sport type have the highest impact.



Summary of Insights

Top Performers

USA, USSR, China.

Popular Sports

Athletics, Swimming, Gymnastics.

Gender Trends

Steady increase in female participation.

Medal Trends

Political and economic changes influenced performance.

Machine Learning

Predicting medal chances with reasonable accuracy.



Next Steps & Further

- 1**
- 2**
- 3**

Model Improvement

Add more features, use advanced models.

Dataset Expansion

Include more recent Olympics.

Real-Time Integration

Connect with live Olympic stats.

Thank
you





Contact Me

- Email: ds075156@gmail.com
- LinkedIn:
[https://www.linkedin.com/in/dhruv
singh-data-analyst/](https://www.linkedin.com/in/dhruv-singh-data-analyst/)
- GitHub: <http://github.com/dhru07>