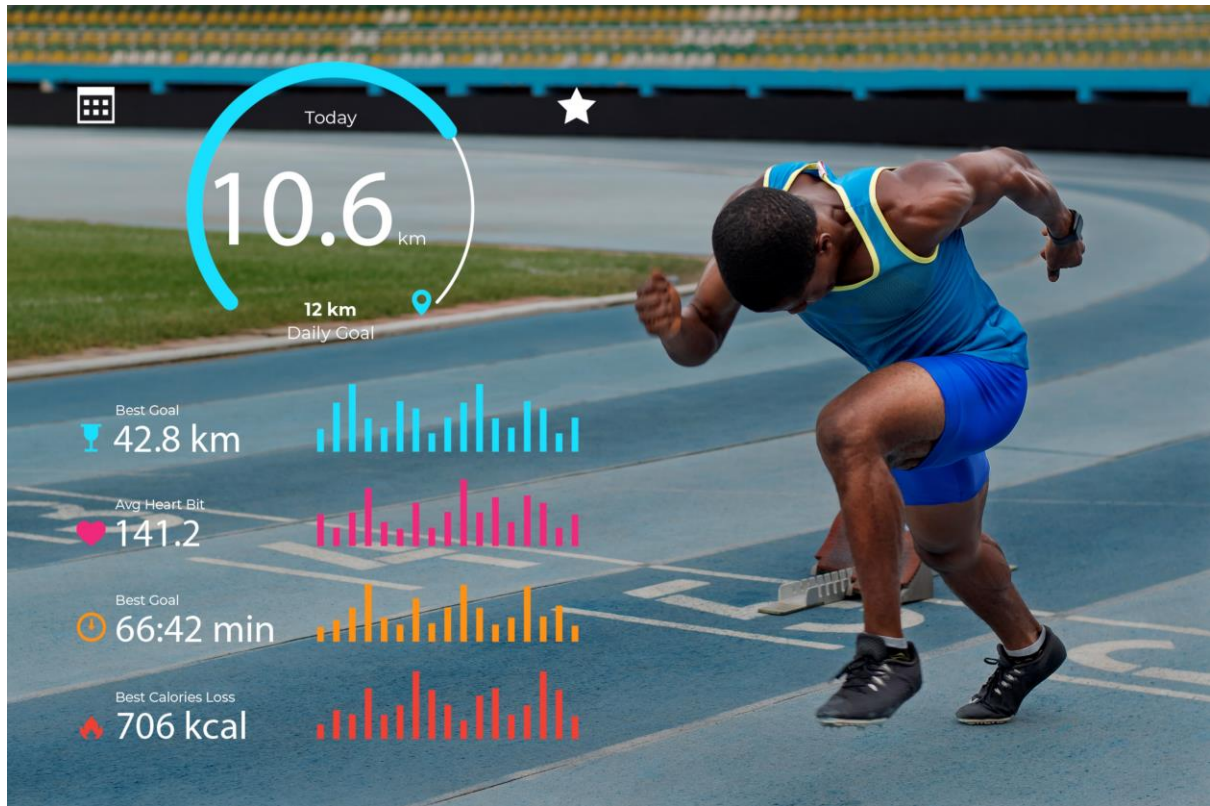


Project Title : Olympic Data Analysis Project



Name – Ankit Kumar

Date – 02/07/2025

Project Overview

Objective:

The objective of this project is to perform a comprehensive analysis of the Summer Olympic medals from 1976 to 2008. This involves:

- Cleaning the dataset
- Analysing medal distributions by countries, years, gender, and sports
- Visualizing key trends and patterns
- Extracting meaningful insights about country performances, gender participation, and sport popularity

DATASET DESCRIPTION

- **Source:** Summer Olympic Medals Dataset (1976–2008)
- **Total Records:** 15,433 rows (before cleaning)
- **Features:**
 - City: Host city of the Olympic Games
 - Year: Year of the Games
 - Sport: Sport category
 - Discipline: Specific discipline within the sport
 - Event: Event name
 - Athlete: Athlete's name
 - Gender: Gender of the athlete
 - Country_Code: Abbreviation of country
 - Country: Full name of the country
 - Event_gender: Gender category of the event (M, W, or X)
 - Medal: Medal type (Gold, Silver, Bronze)

DATA CLEANING PROCESS

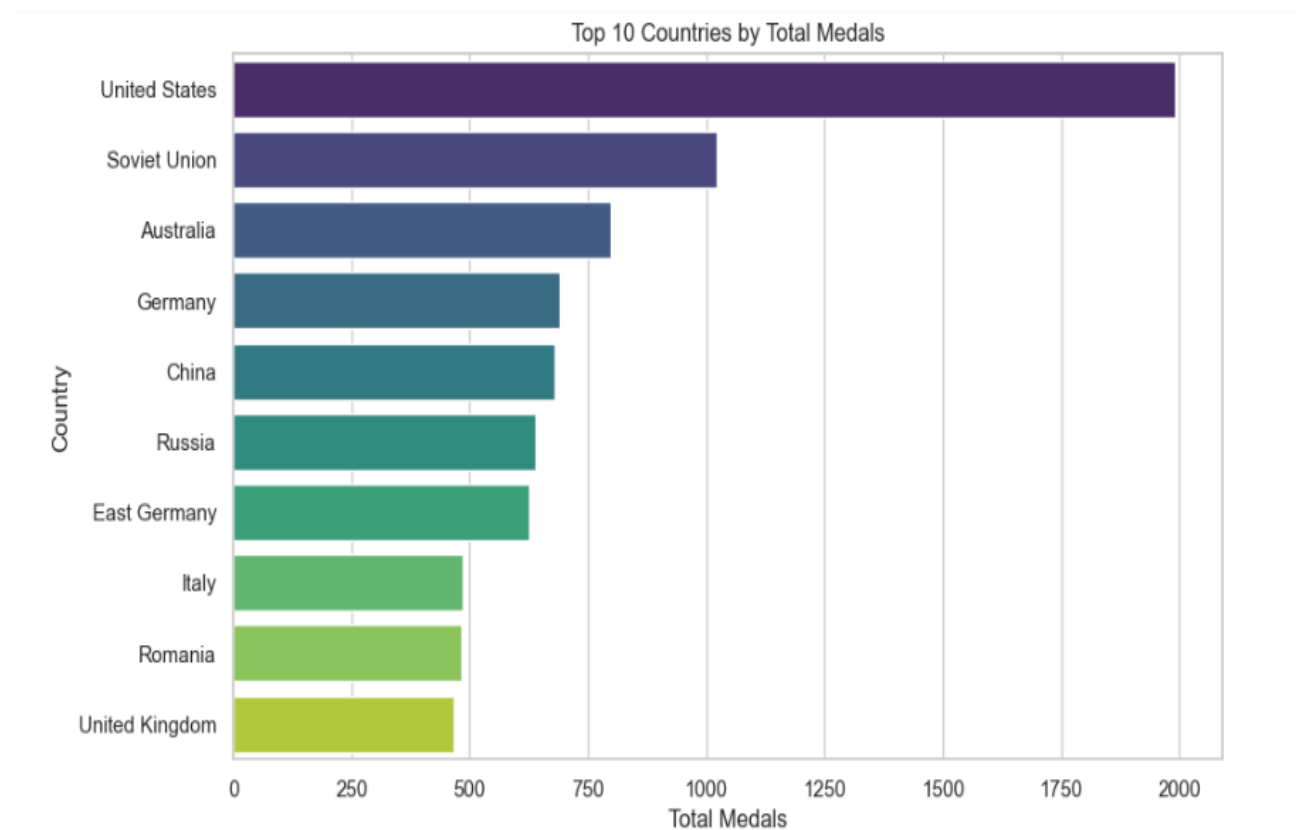
Steps Taken:

- **Encoding Issue Fixed:** Loaded using encoding='ISO-8859-1'.
- **Removed Duplicates:** 117 duplicate rows dropped.
- **Handled Missing Values:** Dropped 117 rows with nulls (~0.75%).
- **Data Types Corrected:**
 - Converted Year from float to integer.
 - Converted Gender, Event_gender, and Medal to categorical types.
- **Standardized Text Columns:**
 - Removed leading/trailing spaces.
 - Standardized capitalization (e.g., 'united states' → 'United States').
 - Standardized country codes to uppercase.
- **Feature Engineering:**
 - Added Medal_Value column (Gold=3, Silver=2, Bronze=1).
 - Extracted First_Name and Last_Name from Athlete.
 - Added Decade column.

Data Analysis Phase

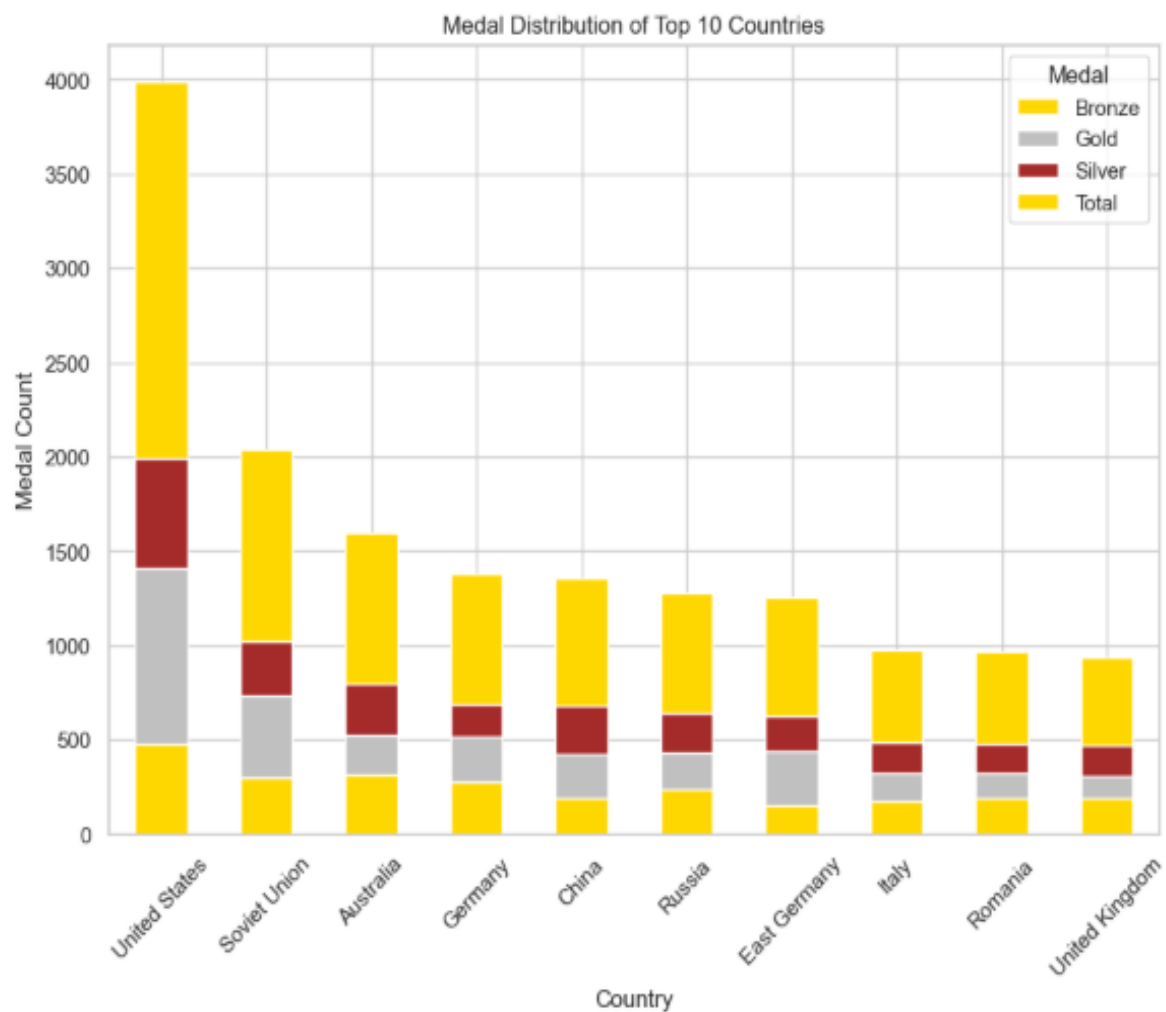
Ques-1) Top Countries by Total Medals (1976–2008).

- **United States** dominated overall, consistently winning the most medals.
- Other top countries include:
 - Soviet Union (until 1992)
 - Germany (including East & West Germany in earlier years)
 - China (rapid growth post-1988)
 - Russia (post-1996)
- USA remained the most dominant Olympic country throughout the timeline.



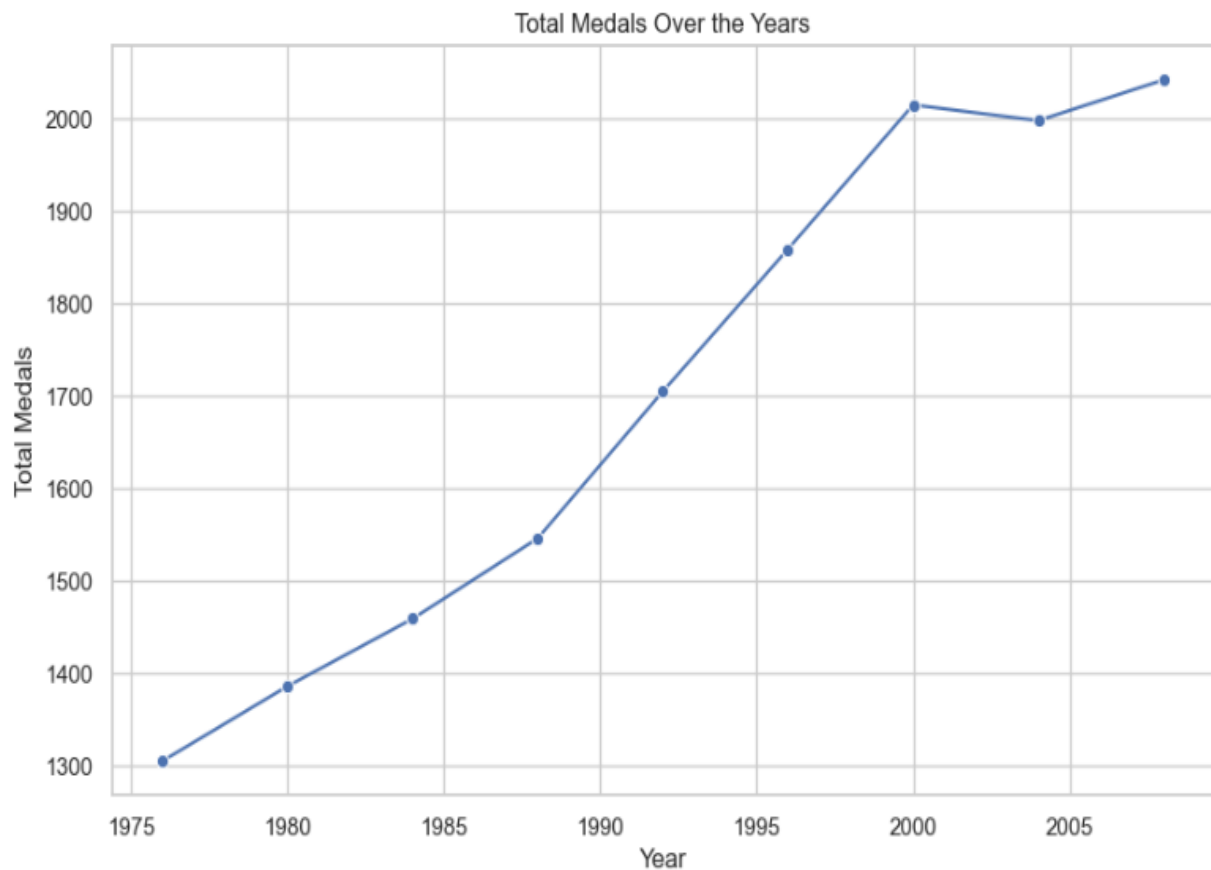
Ques-2) Medal Distribution by Country.

- Medal distribution heavily skewed towards a few top nations.
- **USA**, **Soviet Union**, and **China** accumulated high counts across all medal types.
- Some countries like **East Germany** performed extremely well before reunification in 1990.



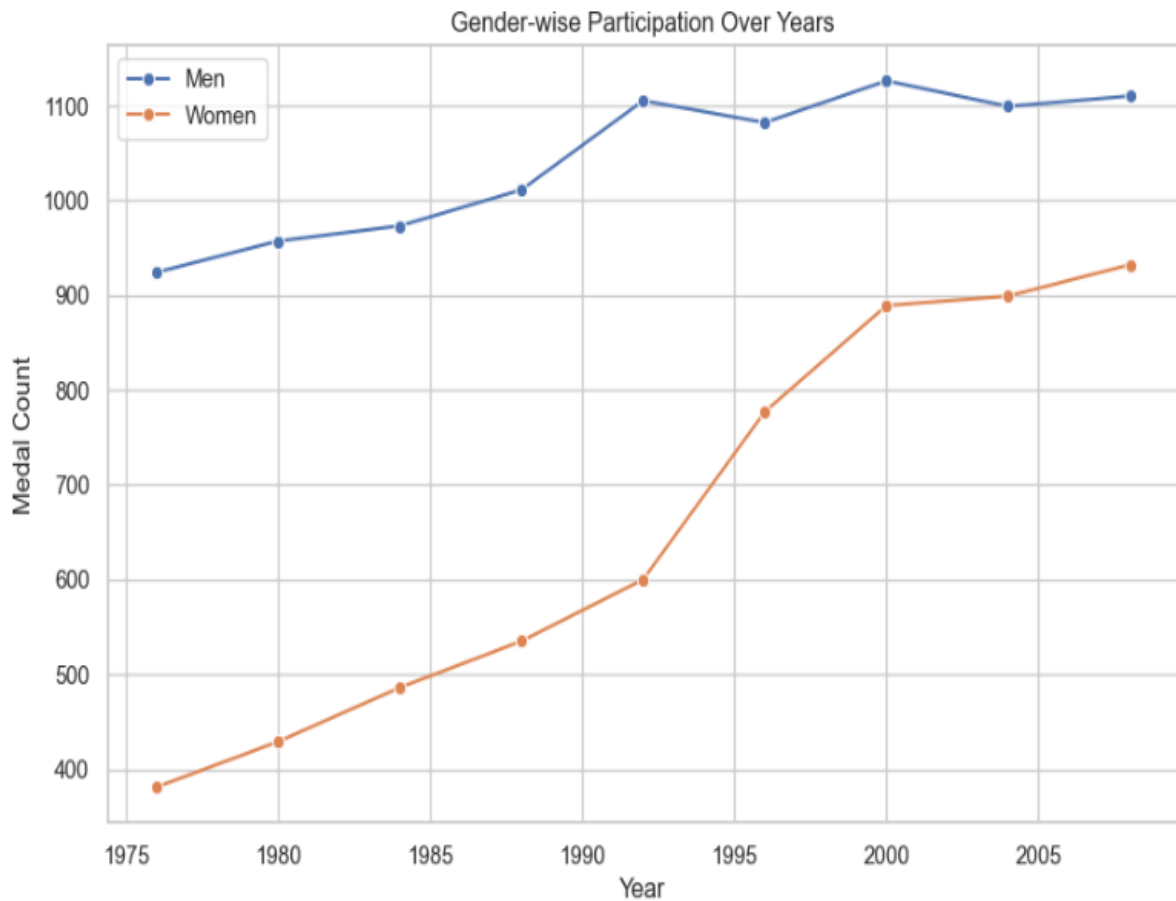
Ques-3) Year-wise Medal Trends.

- Fluctuations were observed due to political events (e.g., boycotts in 1980 and 1984).
- General trend shows an increase in the total number of medals over the years, indicating:
 - Growth in event categories
 - Increased participation worldwide



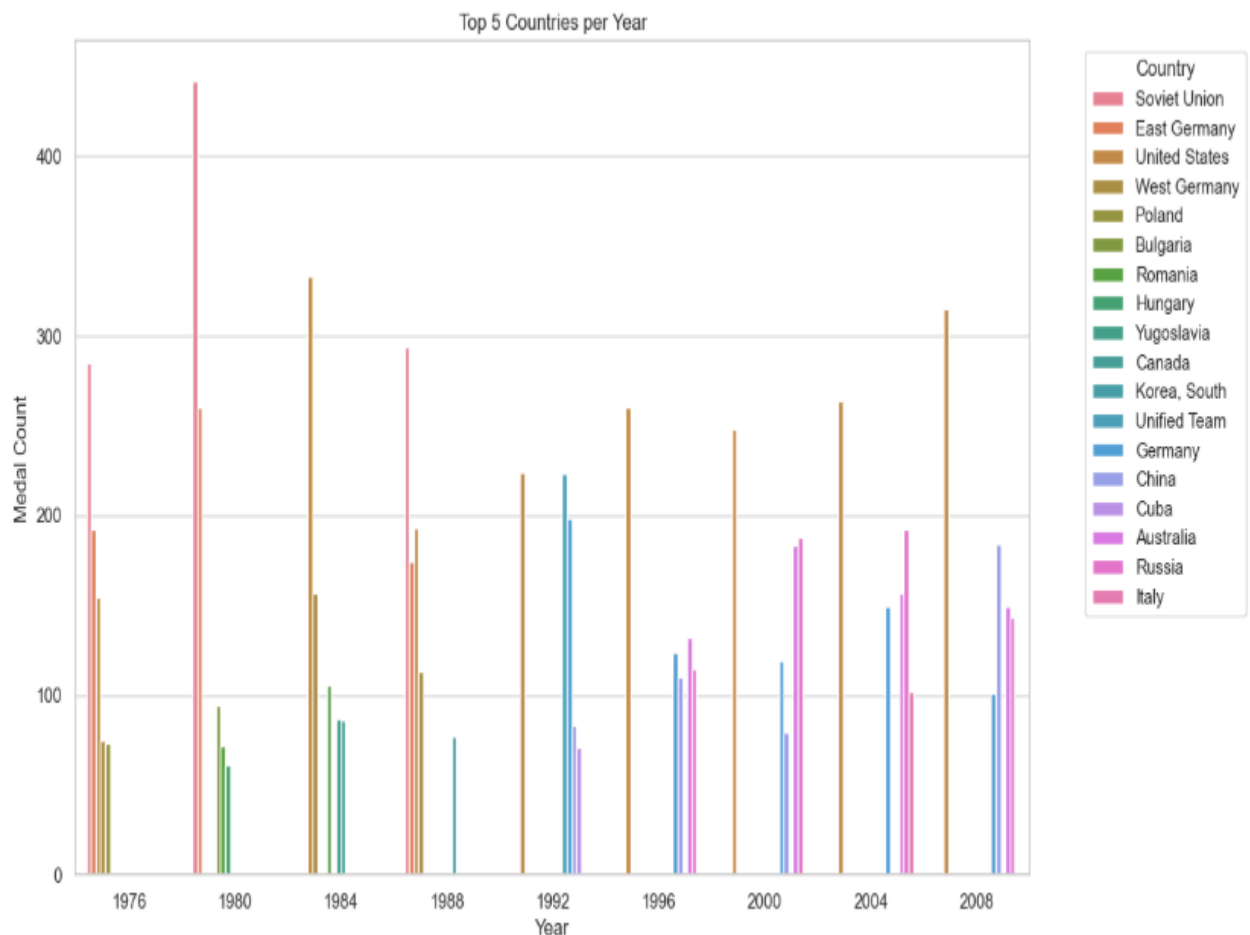
Ques-4) Gender-wise Participation over the years.

- **Male athletes dominated participation in early years (1976–1988).**
- **Women's participation increased steadily from 1992 onward.**
- By 2008, the gender gap narrowed significantly.



Ques-5) Top 5 Countries Each Year.

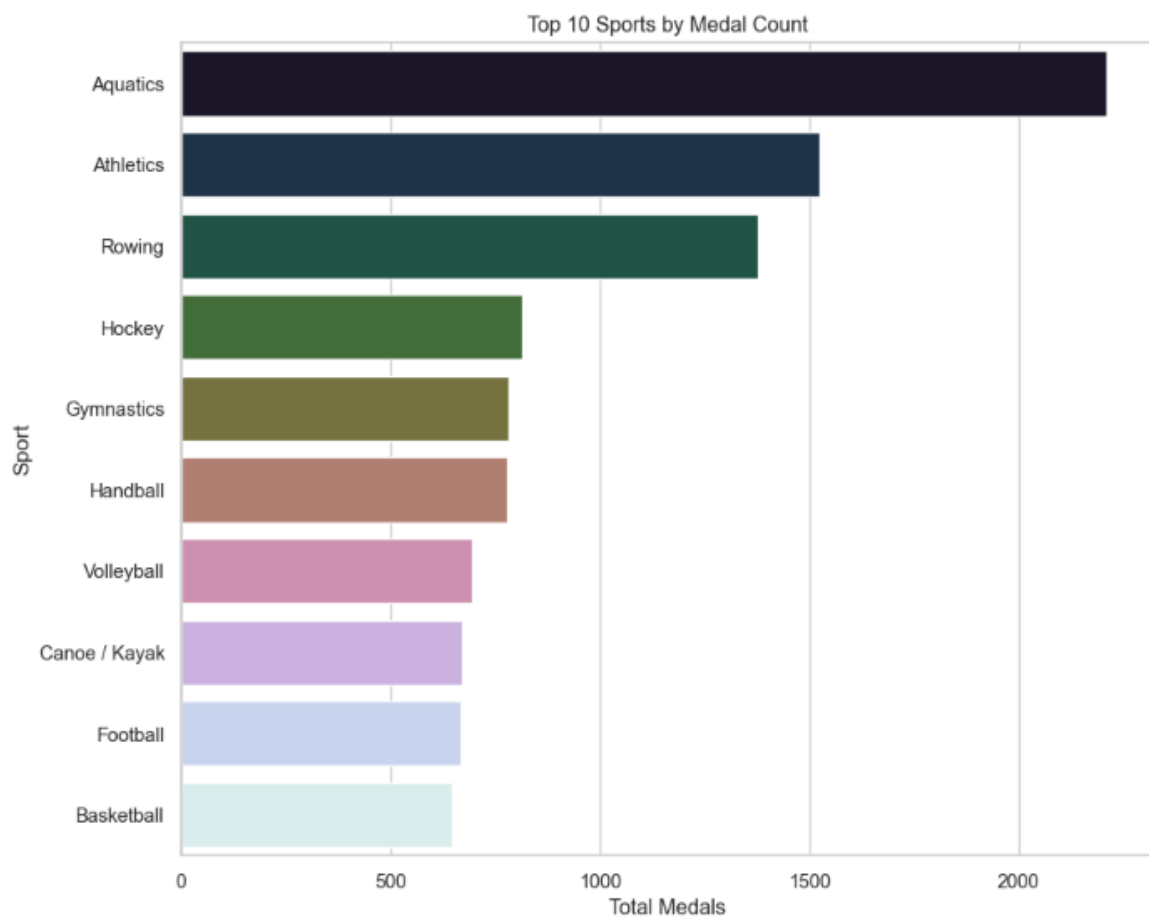
- The USA consistently ranked in the top 5 every year.
- The **Soviet Union/Unified Team/Russia** was also consistently a top performer.
- **China's rise** was particularly noticeable post-1992, culminating in strong performances in 2004 and 2008.



Most popular sports by medal count

Ques-6) Most Popular Sports by Medal Count.

- **Aquatics (Swimming, Diving):** Highest number of medals.
- **Athletics (Track & Field):** Second-highest medals.
- Other popular sports:
 - **Gymnastics**
 - **Wrestling**
 - **Rowing**



CHALLENGES FACED

- **Encoding Issues:** CSV had non-UTF-8 characters; resolved using ISO-8859-1.
- **Country Name Standardization:** Managed inconsistencies in country names due to geopolitical changes (e.g., Soviet Union, Unified Team, Russia).
- **Event Gender Categories:** Managed 'X' representing mixed events.

CONCLUSION

- The analysis provides a comprehensive view of Olympic medal distributions from 1976 to 2008.
- It highlights:
 - Dominance of certain nations (especially USA and Soviet Union/Russia).
 - The increasing role of female athletes over time.
 - The changing geopolitical landscape reflected in medal counts (e.g., the breakup of the Soviet Union).
 - Sports like Aquatics and Athletics consistently contributed the most medals.