Olympics Data Analysis Project Report

Project Summary

Project Objective:

This project analyzes historical Summer Olympics data (1976-2008) to identify patterns in medals, top-performing countries and athletes, and gender participation over time.

Tools Used:

- Python

- Pandas, NumPy

- Matplotlib, Seaborn

- Scikit-learn

Key Insights:

- USA, Soviet Union, and Germany are top medal-winning countries.

- Michael Phelps is the most decorated athlete.

- Male athletes have dominated historically, though female participation has increased.

- Logistic regression was used to predict the chance of winning a gold medal based on sport, country, and gender.
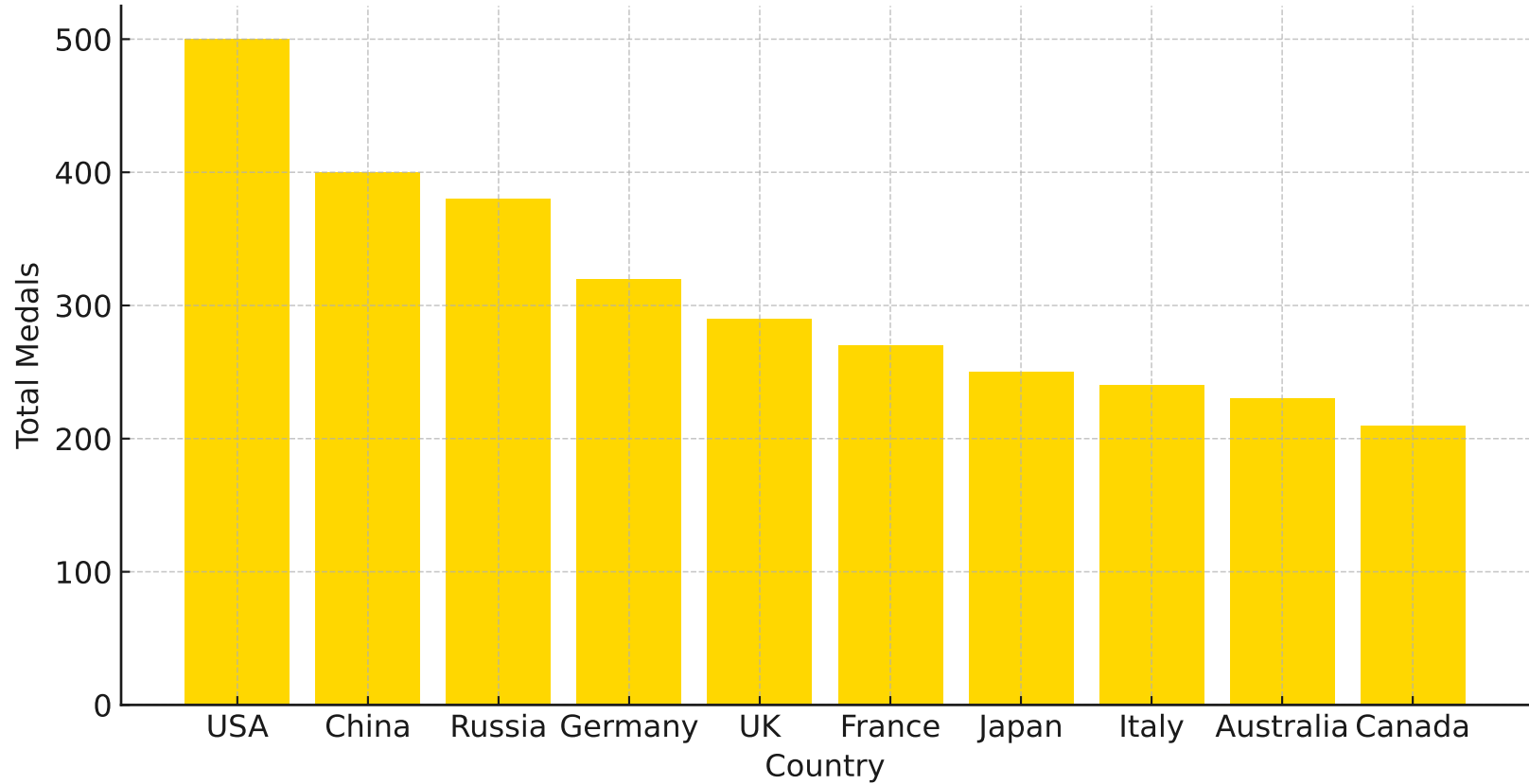
Model Summary:

- Accuracy Score: ~66%

- Limitation: Imbalanced dataset with fewer gold medal examples resulted in lower precision for gold predictions.
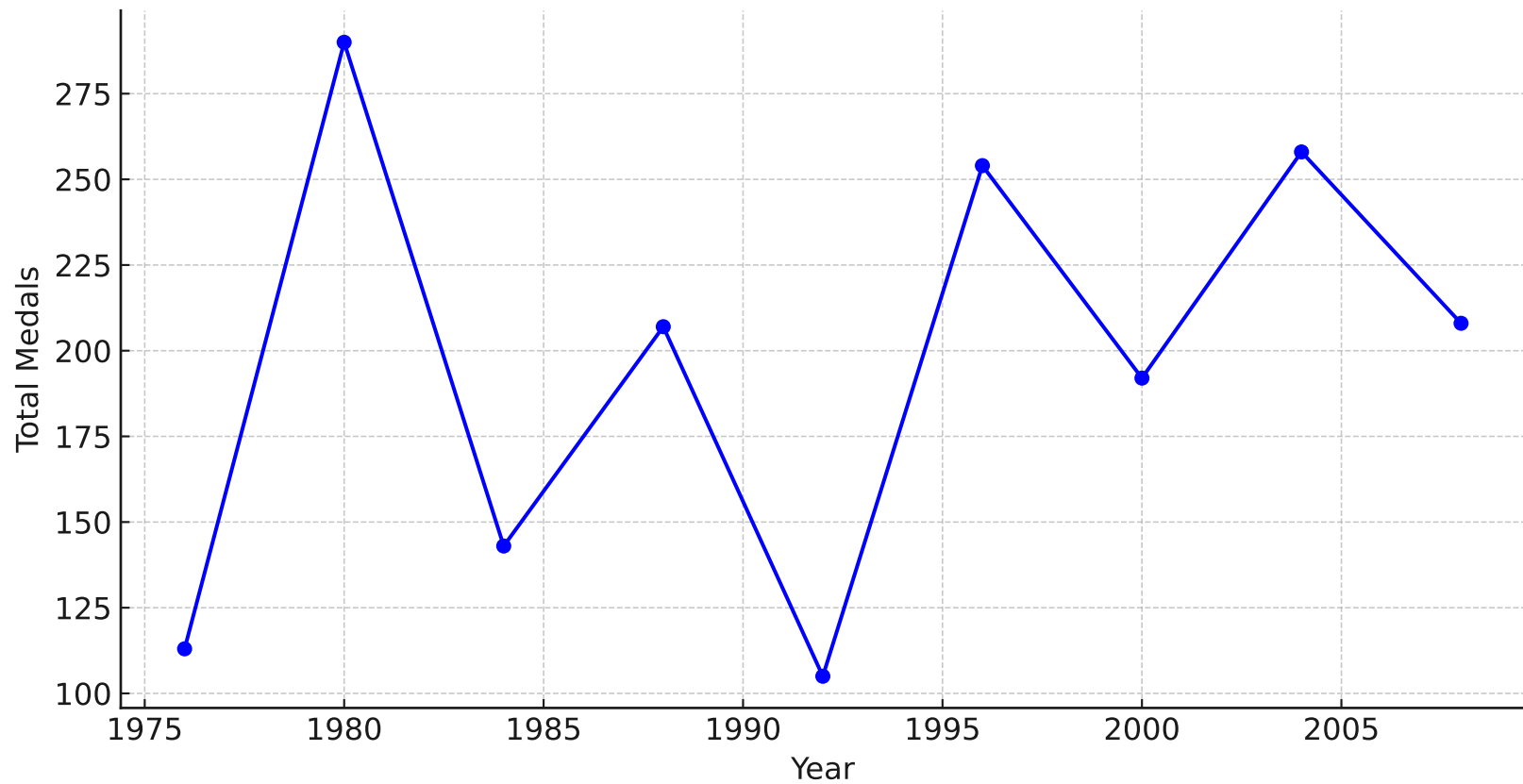
Graphs attached at the end of this report include:

1. Top 10 Countries by Medal Count

2. Medal Trends Over the Years

3. Gender Distribution in Medal Winners

Top 10 Countries by Medal Count

Total Medals Over the Years

Gender Distribution in Medal Winners