

Project Proposal — Olympic Sports Data Analytics

Executive Summary

This project aims to uncover key trends in Olympic history by analyzing over a century of athlete participation and medal performance data. The analysis will focus on **medal distribution**, **gender participation**, and **athlete demographics**, providing **evidence-based insights** for sports organizations, broadcasters, and sponsors. By the end of the project, we will deliver actionable intelligence on **which sports and nations dominate**, **how participation has evolved**, and **what demographic patterns drive performance**.

1. Project Title & Basic Info

Title: “Exploring Olympic Trends: Medal Distribution, Gender Participation, and Athlete Profiles (1896–2016)”

Client: A global sports analytics firm advising **sports federations**, **Olympic committees**, **broadcasters**, and **sponsors** who need **data-driven insights** to guide Olympic strategy and investment.

Dataset: Historical Olympic dataset (`athlete_events.csv`) — covering over **271,000 athlete entries** from **1896 to 2016**, including athlete demographics, sports, events, medal outcomes, and countries (NOC).

Why this dataset:

- Rich longitudinal data covering 120 years of Olympic history.
 - Enables performance, demographic, and participation analysis.
 - Highly relevant to real-world strategic decisions in sports.
-

2. Key Questions (What We Want to Find Out)

1. How has **medal distribution** evolved over time across different sports, countries, and Olympic editions?
 2. Which **sports and nations** have historically dominated the Olympics in terms of total medals?
 3. How has **gender participation** changed over time, and what were the major inflection points?
 4. What are the **typical age patterns** of Olympic athletes and medalists, and how do they vary across sports?
-

3. Initial Hypotheses (What We Believe to Be True)

1. Sports with more events (e.g., **Athletics**, **Swimming**) dominate medal counts.
 2. Female participation increased significantly from the late 20th century onward.
 3. Most Olympic athletes and medalists fall between **20–30 years old**.
 4. Countries with a **long Olympic history** (e.g., USA) maintain consistent medal dominance.
-

4. Analytical Approach (How We'll Prove or Disprove)

- **Step 1: Data Preparation**

- Import and clean dataset using Pandas.
- Impute missing values for age; leave height/weight untouched initially.
- Create helper datasets (`medal_data` , `age_series`) to support focused analysis.

- **Step 2: Exploratory Data Analysis**

- Aggregate medal counts by **Year**, **Sport**, and **NOC**.
- Compare **Summer vs Winter Games** trends over time.

- **Step 3: Gender & Demographics**

- Analyze gender participation trends using `Year × Sex`.
- Explore age distribution of athletes vs. medalists.

- **Step 4: Visualization & Insight Building**

- Use Matplotlib and Seaborn for visual storytelling (line charts, bar plots, histograms, KDE, boxplots).
- Validate or revise hypotheses based on the results.

- **Step 5: Final Insights & Delivery**

- Summarize trends into key findings for decision-makers.
- Highlight dominant sports, participation evolution, and demographic profiles.
- Prepare **presentation visuals** and final **management report**.

This proposal provides a clear path from exploration to insight, aligning with industry-standard data analytics workflows and enabling impactful, actionable recommendations for stakeholders.

Prepared by: Bibhudendu Behera

Date: 18 October 2025