

Project Report for CS661: Big Data Visual Analytics

Olympus Insight

Group 10

Even Semester 2024–2025
Indian Institute of Technology Kanpur

Team Member	Email Address
Abhishek Choudhary	cabhishek21@iitk.ac.in
Arnesh Dadhich	arneshd21@iitk.ac.in
Nirmal Prajapati	nirmal21@iitk.ac.in
Rajat Gattani	rajatg21@iitk.ac.in
Utkarsh Agrawal	utkarsha21@iitk.ac.in
Vipul Chanchlani	vipulch21@iitk.ac.in
Vishal Himmatsinghka	vishalh21@iitk.ac.in
Yash Verma	yashv21@iitk.ac.in

April 27, 2025

Abstract

Olympus Insight is an interactive analytics platform that transforms over a century of Olympic Games data into actionable insights. Leveraging Dash, Plotly, and Pandas, the dashboard integrates athlete demographics, socioeconomic indicators, and environmental factors to deliver comprehensive visualizations and predictive analytics. This report details the objectives, data pipeline, visual analytics modules, and key findings, highlighting how data-driven storytelling can inform athletes, analysts, and policymakers.

1. Introduction

The Olympic Games, a global sporting phenomenon since 1896, provide a rich dataset reflecting athletic achievement, national development, and sociopolitical trends. However, extracting meaningful insights from this wealth of information requires advanced analytical and visualization tools. Our project, **Olympus Insight**, addresses this challenge by transforming historical Olympic data into an interactive analytics platform. Built with Dash, Plotly, and Pandas, the dashboard empowers users to explore participation, performance, and predictive insights through intuitive and dynamic visualizations.

2. Tasks

To realize the vision of Olympus Insight, the following key tasks were accomplished:

- Aggregation and cleaning of Olympic historical records, socioeconomic indicators, and environmental data.
- Development of interactive, web-based visualizations for deep data exploration.
- Implementation of predictive models for medal forecasts.
- Design of a modular, multi-page navigation system for seamless user experience.

3. Data Sources and Composition

To enable a multidimensional and robust analysis of the Olympic Games, we integrated several authoritative datasets spanning athlete, event, national, and temporal perspectives:

- **Olympic Historical Dataset (1896–2016):** Over 74,000 athlete profiles with demographics (age, gender, nationality, career duration) and complete medal records across 53 sports and hundreds of events.
- **World Bank Development Indicators (1960–2024):** Socioeconomic metrics such as GDP, Human Development Index (HDI), and national population, enabling correlation of sporting success with economic and social factors.

Data Composition:

- **Athlete-Level Data:** Attributes include age, gender, nationality, sport discipline, and career span, allowing for demographic and performance analytics.
- **Event-Specific Data:** Details on sport types, venue specifications, participation rates, and world records, supporting event-level and temporal analyses.
- **National Metrics:** Aggregated medal tallies, GDP correlations, and host city investments, enabling country-wise benchmarking and socioeconomic studies.
- **Temporal Patterns:** Decadal trends, seasonal variations (Summer/Winter), and legacy effects, facilitating longitudinal and comparative studies.

This integrated data foundation empowers Olympus Insight to deliver comprehensive, context-rich analytics and visualizations, supporting both macro and micro-level exploration of Olympic history.

4. Solution Flow

Olympus Insight is architected as a modular, scalable analytics platform:

- **Data Pipeline:** An ETL process using Pandas for cleaning, normalization, and merging of diverse datasets.
- **Visualization Framework:** A Dash application with dedicated pages for global, country, and comparative analytics, each tailored to answer specific analytical questions.
- **Predictive Modeling:** Decision Tree regressors and LSTM model made available to the user for future prediction of a country's performance.
- **Deployment:** Deployed via AWS at Olympus Insight.

5. Results and Visual Analytics

5.1 Globe View (Global Olympic Analysis)

Purpose:

To provide a global, interactive visualization of Olympic medal efficiency across all participating countries, enabling users to identify regional patterns and outliers at a glance.

Key Features:

- 3D interactive globe with color mapping for medal efficiency (total points per athlete).
- Region filter and country drilldown for focused analysis.
- Clickable navigation to country profile pages.
- Clear metric definitions and user guidance for interpretation.

Insights:

- Medal efficiency varies widely across continents, with smaller nations sometimes outperforming larger ones on a per-athlete basis.
- Developed countries and traditional sporting powerhouses dominate both total and efficient medal counts.
- The visualization quickly reveals underdog nations with high efficiency, highlighting the impact of focused investment and specialization.
- Regional filters help users identify continental trends and anomalies in performance.

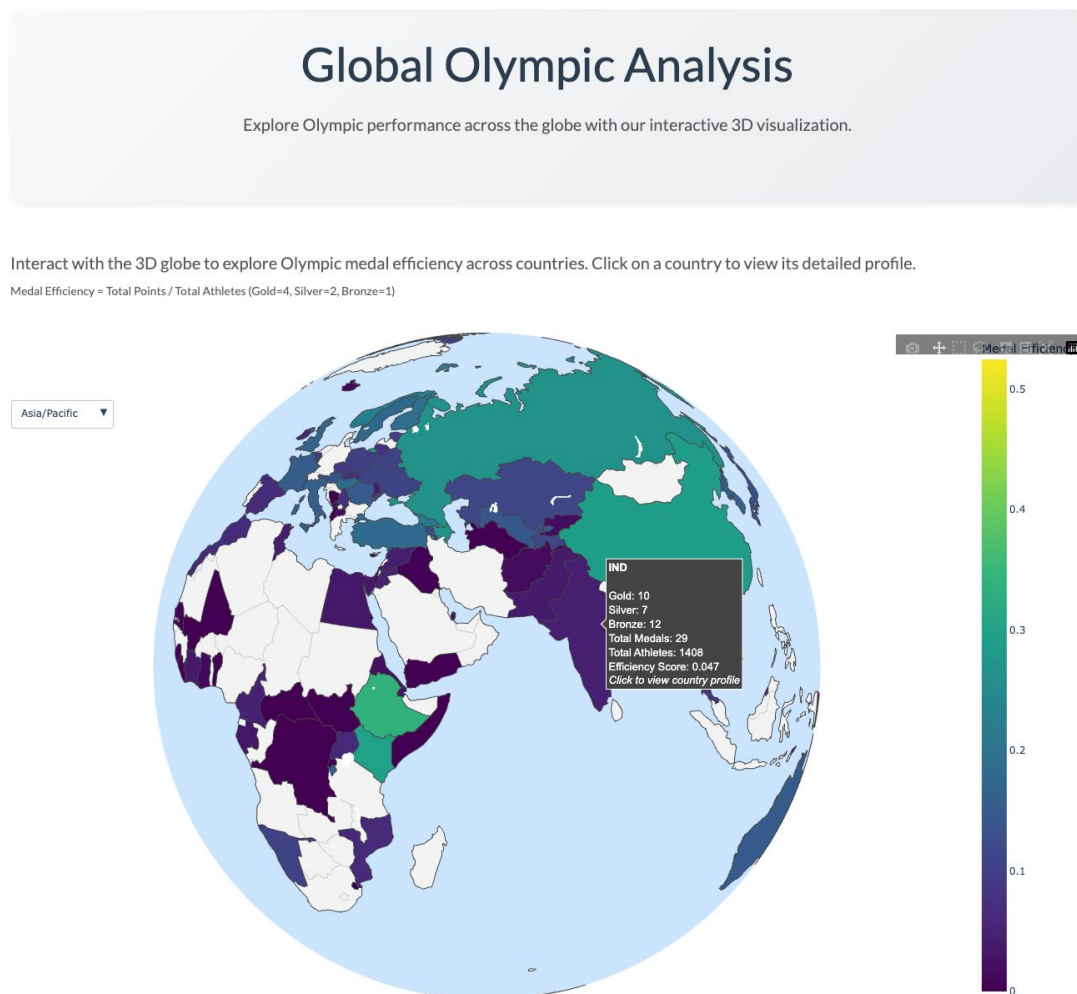


Figure 1: Global Olympic Analysis: Interactive 3D globe visualizing medal efficiency by country.

5.2 Country Comparison

Purpose:

Enables side-by-side comparison of Olympic performance metrics and statistics between any two nations, facilitating benchmarking and trend analysis.

Key Features:

- Medal breakdown by type and year for both countries.
- Temporal trends in medals and athlete participation.
- Sport-wise medal distribution for each country.
- Efficiency metrics, normalized radar and bubble charts.
- Animated medal accumulation over time to visualize historical momentum.

Insights:

- Users can identify eras of dominance, rapid improvement, or decline for any nation.
- Efficiency and normalized metrics reveal not just raw totals, but performance relative to resources and participation.
- Animated views make it easy to spot historical turning points and the effect of major events (e.g., hosting, political changes).
- The tool highlights how some countries excel in particular sports, while others show broad-based strength.

Select two countries to compare their Olympic achievements, participation trends, and performance metrics.

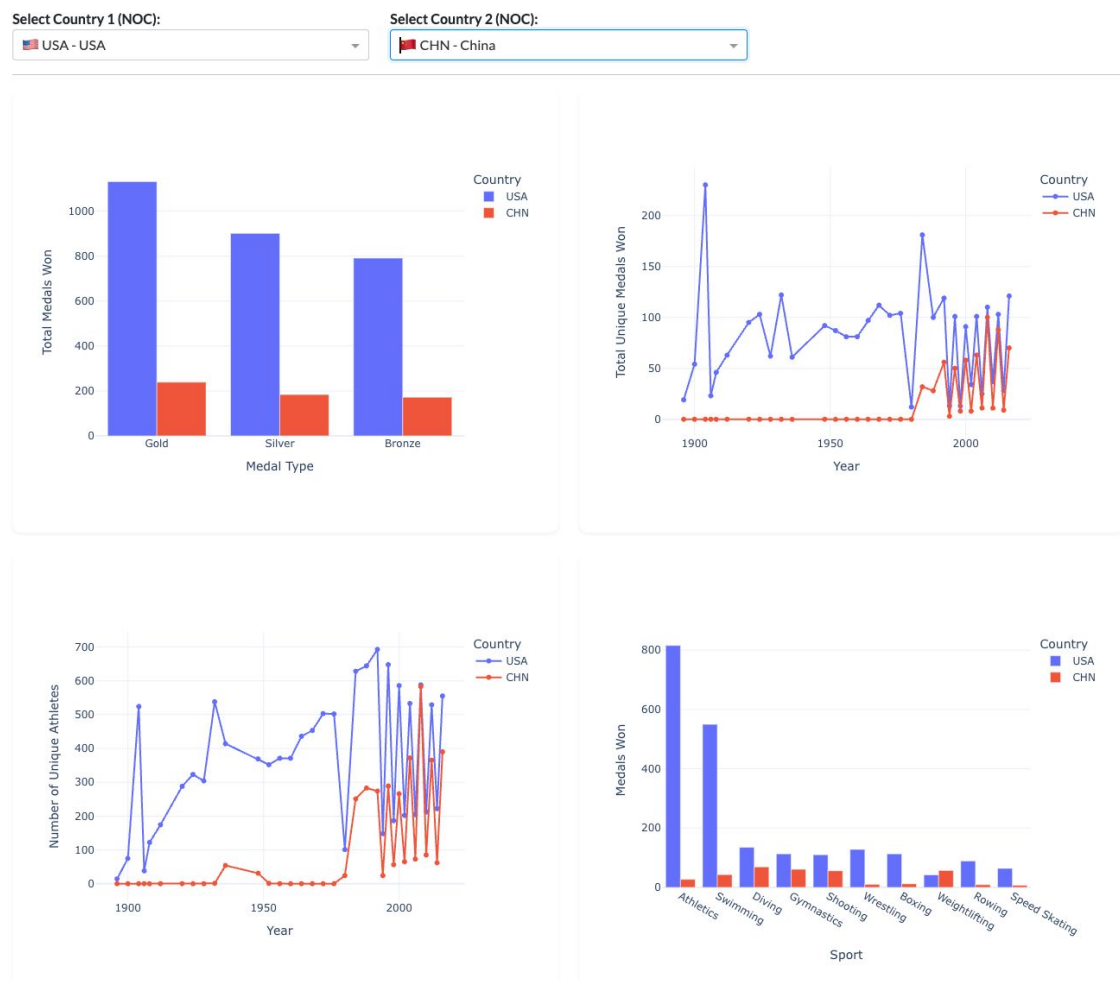


Figure 2: Animated Medal Accumulation Graph, Radar Chart of Normalized Performance Metrics, and Bubble Chart for Sports Performance Comparison.

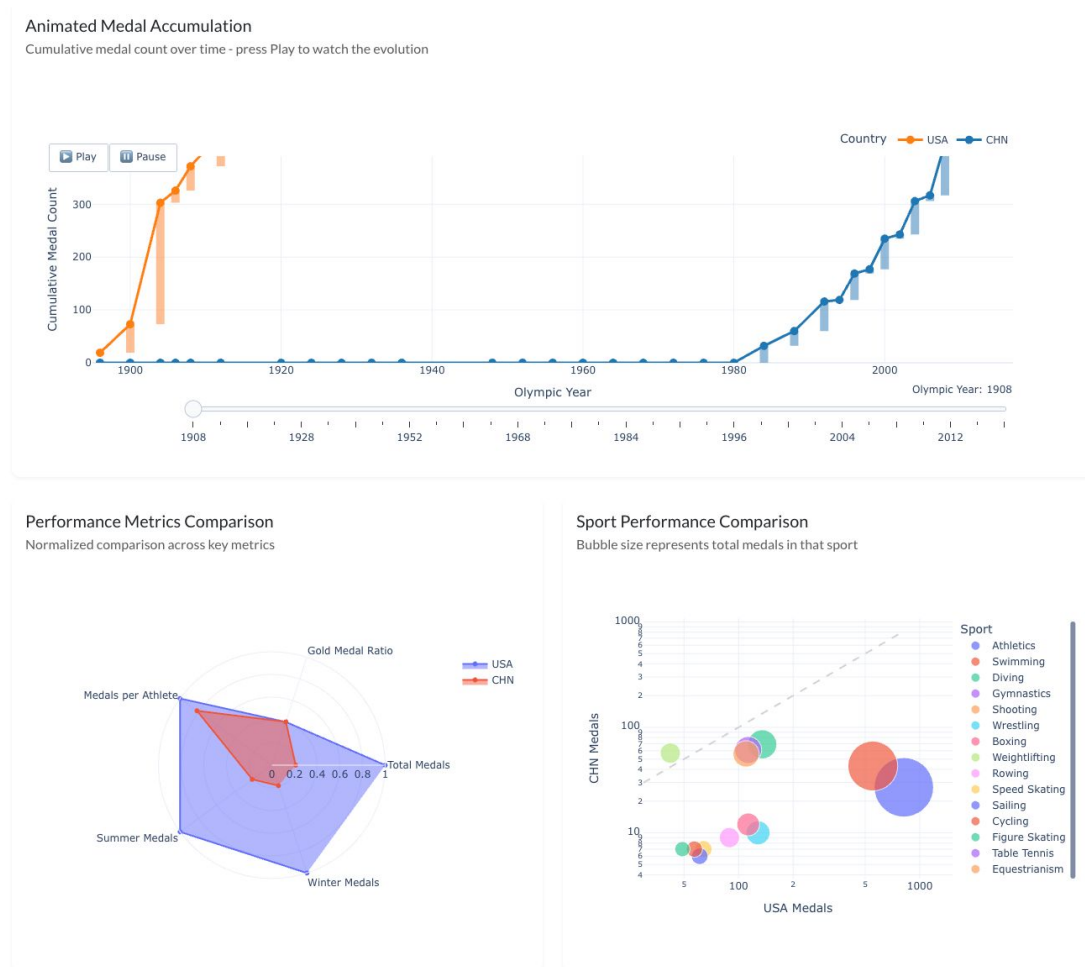


Figure 3: Medal Type Distribution, Year-wise Medal Trends, Year-wise Athlete Trend, and Sports Comparison between Two Countries

5.3 Country Profile

Purpose:

Provides an in-depth overview of a single country's Olympic journey, including historical trends, demographic makeup, and performance drivers.

Key Features:

- Historical medal timeline (cumulative and per-edition).
- Age and gender participation demographics.
- Sport specialty breakdown (sunburst chart).
- Socioeconomic correlation (GDP vs. medals per capita).
- Age-performance curve and top athletes spotlight.

Insights:

- Medal trends reveal periods of growth, stagnation, or resurgence for each country.
- Demographic charts show progress in gender equity and shifting age profiles of elite

athletes.

- Sport breakdowns highlight national strengths and evolving priorities.
- Socioeconomic plots confirm that higher GDP often correlates with greater medal success, but outliers exist.
- Age-performance curves illustrate optimal age ranges for medal-winning athletes in different sports.

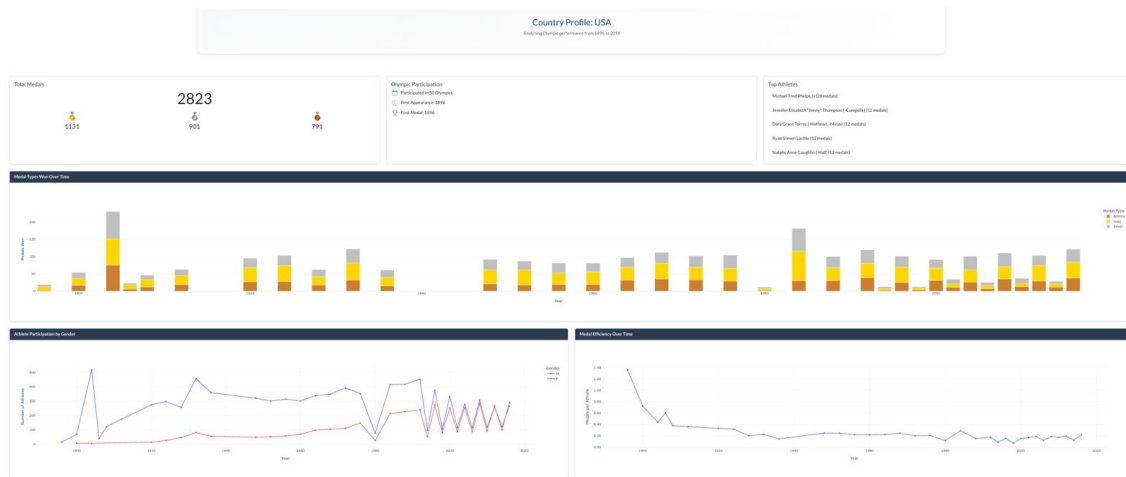


Figure 4: Historical Medals Timeline for the selected country, showing progression across different Olympic editions.

5.4 Host Country Analysis

Purpose:

Analyzes the 'Home Field Advantage' by comparing host nations' performance during hosting years versus previous and subsequent Games.

Key Features:

- Interactive host selection and year navigation.
- Medal and athlete participation comparison for previous, host, and next Games.
- Summary statistics for medals, athletes, and efficiency.
- Percentage change indicators for all metrics.

Insights:

- Host nations typically experience a significant boost in both medal count and athlete participation during their hosting year.
- For example, China at the 2008 Beijing Olympics saw a 40.9% increase in total medals and a 56.7% increase in athlete participation compared to the previous Games.
- Medal efficiency (medals per athlete) also tends to peak in the host year, reflecting improved performance.

- The "home advantage" is consistent across most hosts, attributed to factors like increased quotas, familiar environments, and greater investment.



Figure 5: Host Country Analysis: Medal and athlete comparisons for the host year, previous, and next Games, with summary statistics.

5.5 More Analysis (Advanced Olympic Analysis)

Purpose:

Enables users to explore deeper insights and patterns in Olympic data through a suite of advanced, interactive visualizations and analysis tools.

Key Features:

- **Flexible Filtering:** Users can select year range, season (Summer/Winter), and specific sports to tailor the analysis to their interests.
- **Tabbed Analytics:** Multiple analysis tabs, including Medal Trends, Physical Characteristics, Age Analysis, Gender Analysis, and Sport Deep Dive, provide a comprehensive view of the data.
- **Medal Trends Visualization:** Dynamic line charts display medal count trends for the top countries over time, with clear legends and color-coding for easy comparison.
- **Interactive Controls:** Real-time updates as users adjust filters, allowing for immediate exploration of new questions and hypotheses.
- **Sport-Specific Analysis:** Users can drill down into individual sports (e.g., Badminton) to observe country-wise dominance and evolving trends within that discipline.
- **User Guidance:** Clear instructions and contextual information help users interpret the visualizations effectively.

Insights:

- The advanced analysis module uncovers nuanced trends in medal counts, participation, and demographic patterns that may not be visible in broader views.
- Filtering by sport and season reveals how different countries have risen or declined in specific disciplines, highlighting shifts in global sporting power.
- Age and gender analysis tabs provide a window into the evolving demographics of Olympic athletes, supporting research into factors influencing performance.
- The ability to interactively explore and compare across multiple dimensions empowers users to generate custom insights relevant to their interests, from sports historians to policy makers and fans.
- This module exemplifies the value of interactive analytics in making complex, high-dimensional Olympic data accessible and actionable.

5.6 Olympic Year Analysis

Purpose:

This module enables users to explore detailed statistics and performance metrics for each Olympic year, providing a comprehensive snapshot of a selected edition of the Games.

Key Features:

- **Year and Season Selection:** Users can choose any Olympic year and filter by Summer or Winter Games for focused analysis.
- **Edition Overview:** At a glance, view the host city and logo, number of participating

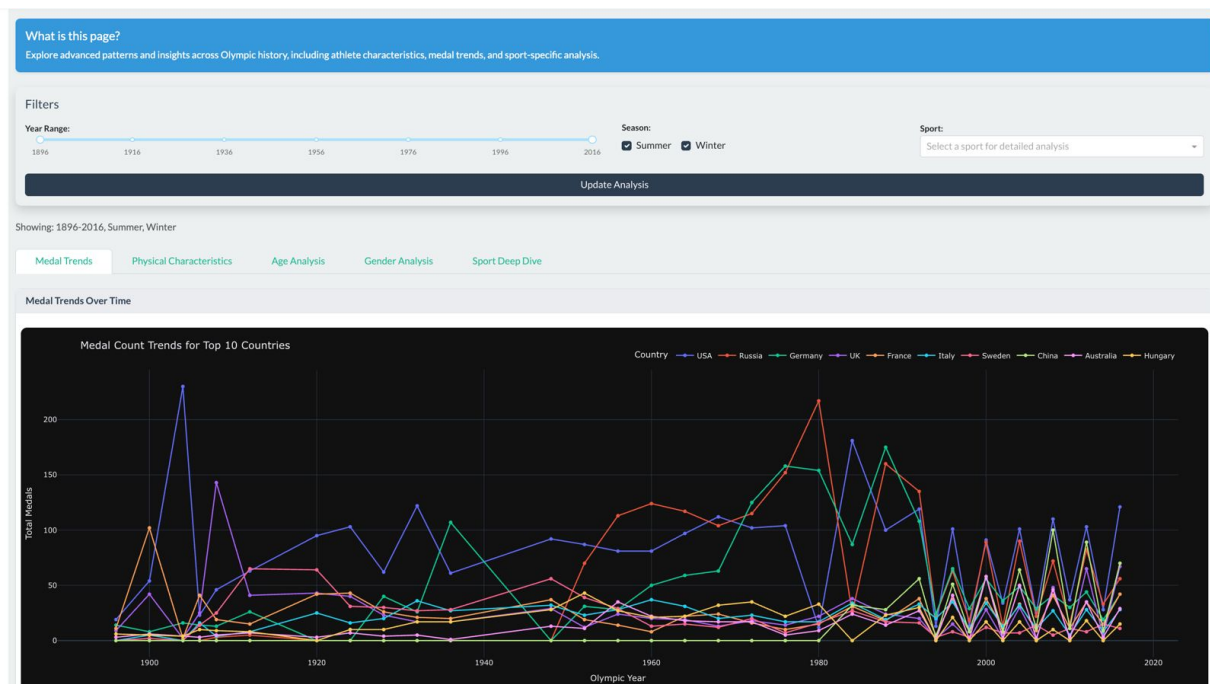


Figure 6: Advanced Olympic Analysis: Interactive dashboard for deep exploration of medal trends, demographics, and sport-specific patterns. Shown: Medal count trends for top 10 countries in Badminton.

nations, total athletes, sports, and events for the selected year and season.

- **Dynamic Medal Table:** An interactive, sortable table presents the full medal tally by country, including gold, silver, bronze, and total medals, with live search and pagination for usability.
- **Participation Metrics:** See how many athletes and nations competed, and how the scale of the Games has evolved over time.
- **Event Breakdown:** Quickly identify the breadth of sports and events contested in each edition.
- **User-Friendly Interface:** Modern design with Olympic branding and mascots enhances engagement.

Insights:

- The Olympic Year Analysis module provides a clear, at-a-glance summary of each Games, making it easy to compare editions and spot trends in participation and performance.
- Users can identify dominant countries in any year, observe changes in the number of sports and events, and track the growing scale of the Olympics.
- The interactive medal table allows for quick discovery of medal standings and facilitates deeper exploration, such as tracking emerging nations or shifts in the global sporting landscape.
- This module is invaluable for researchers, journalists, and enthusiasts seeking detailed, edition-specific Olympic data in an accessible format.



Figure 7: Olympic Year Analysis: Select a year and season to view comprehensive statistics, edition overview, and an interactive medal table. Shown: Rio de Janeiro 2016 Summer Olympics.

5.7 Sport Performance Profile

Purpose:

This module enables users to analyze detailed statistics, trends, and notable achievements for any Olympic sport, offering a comprehensive view of sport-specific performance across the Games.

Key Features:

- **Sport Selection:** Users can choose any Olympic sport (e.g., Badminton) to focus the analysis.
- **Flexible Filtering:** Filter data by season (Summer/Winter), gender, event, and year range to tailor the insights.
- **Sport Overview:** Brief description and rules overview for the selected sport provide context.
- **Current Selection Summary:** Key metrics such as number of athletes, events, and physical characteristics (height/weight stats) for the chosen filters.
- **Top Athletes Visualization:** Bar chart highlighting the most decorated athletes in the selected sport.
- **Medal Distribution:** Interactive charts showing medal counts by country, gender distribution, and age distribution for athletes in the sport.
- **Country Dominance and Trends:** Bar chart of top countries by total medals and line chart of medal trends over time.
- **User-Friendly Interface:** Clean layout, intuitive filters, and informative tooltips en-

hance user experience.

Insights:

- The Sport Performance Profile reveals which countries and athletes have historically dominated each sport, and how this has evolved over time.
- Physical and demographic analyses (age, gender, height, weight) provide insight into the typical profiles of successful athletes in each sport.
- Medal trends and country dominance charts highlight shifts in global sporting power and the emergence of new contenders.
- The module supports both broad exploration and deep dives into specific events, empowering analysts, coaches, and fans to uncover nuanced patterns in Olympic sports.

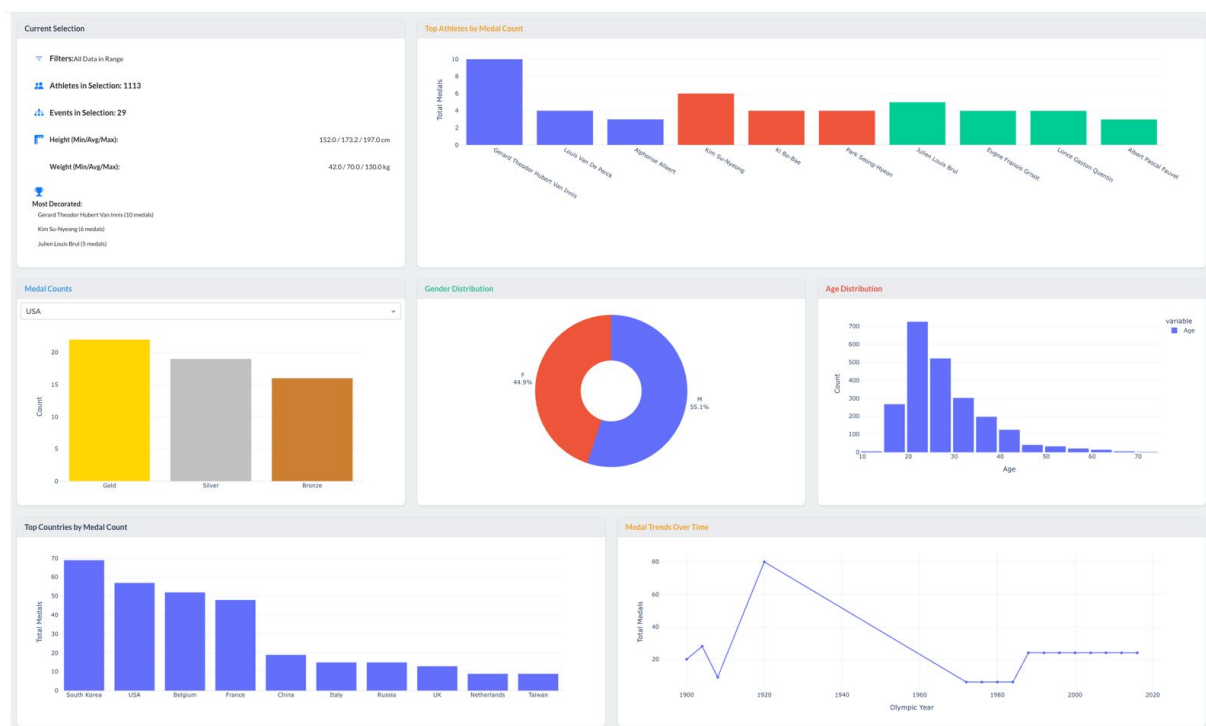


Figure 8: Sport Performance Profile: Key metrics, top athletes, country medal counts, gender and age distribution for the selected sport.

5.8 Economic Factor Analysis

Purpose:

This module explores how national economic and human development factors-especially GDP and Human Development Index (HDI)-correlate with Olympic success, using both correlation analysis and categorical breakdowns.

Key Features:

- **HDI Impact Insights:** Presents the overall correlation between a country's latest HDI and its total Olympic medal count, with interpretation of the correlation coefficient.
- **Medal Distribution by HDI Category:** Bar charts and summary statistics show how Olympic medals are distributed among nations grouped by HDI level (Low, Medium, High, Very High).
- **HDI vs. Medal Count Scatterplot:** Visualizes the relationship between HDI and total medal count, with bubble size indicating medal volume and color denoting HDI category.
- **Contextual Guidance:** Explanations clarify the meaning of correlation and the socio-economic mechanisms that may underlie observed patterns.

Insights:

- There is a moderate positive correlation (**0.41**) between a country's HDI and its total Olympic medal count, suggesting that nations with higher human development tend to achieve greater Olympic success.
- Medal distribution is highly skewed: **84.7%** of all Olympic medals are won by countries with Very High HDI, while nations with Low or Medium HDI collectively account for less than 2.5% of medals.
- The scatterplot reveals that while a higher HDI generally aligns with more medals, there are notable outliers-some high-HDI countries with modest medal counts (often due to small population) and a few lower-HDI countries that excel due to specialization in specific sports.

5.9 Prediction Module

Purpose:

This module predicts future Olympic medal counts for countries using machine learning models trained on historical data (1896–2016), offering insights into potential upcoming performances.

Key Features:

- **Model Choice:** The interface allows user to choose between the decision tree and LSTM model for appropriate prediction.

Decision Tree Regression:

Given a dataset $D = \{(x_i, y_i)\}_{i=1}^n$, a decision tree splits the input space recursively to minimize the Mean Squared Error (MSE) at each node:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

where \hat{y}_i is the predicted value in a region (leaf) and y_i is the true target value.

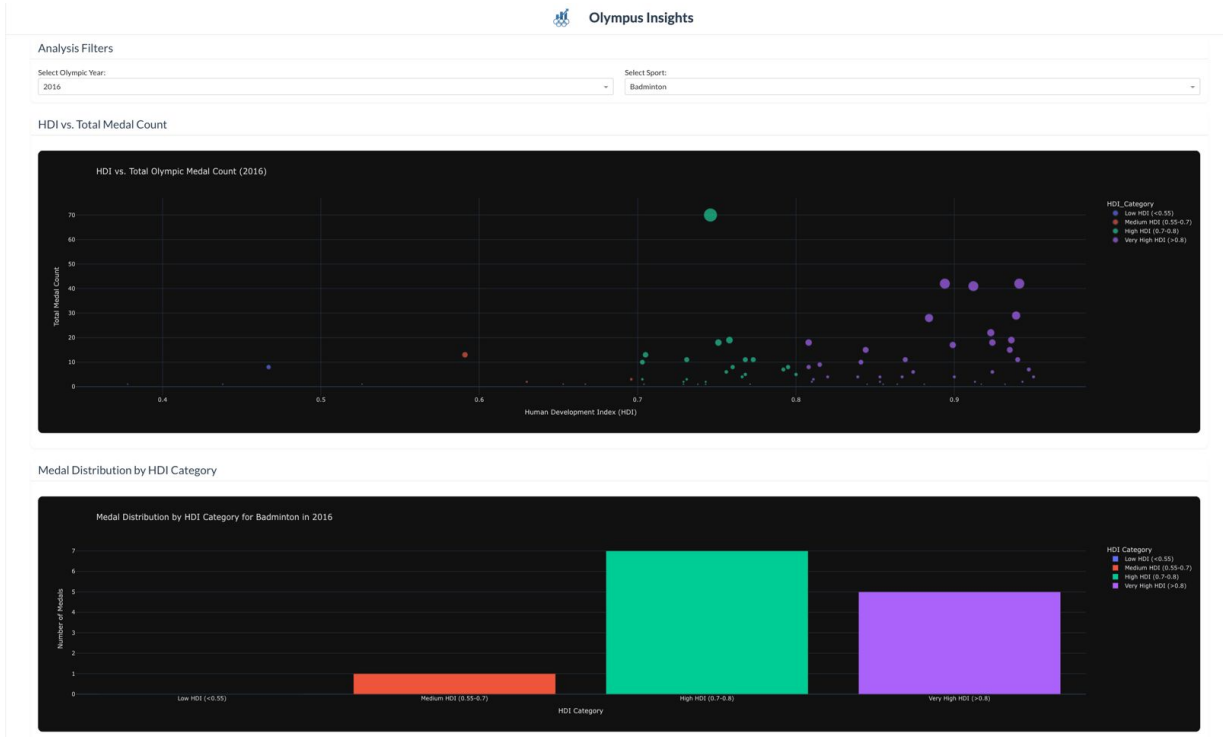


Figure 9: Key Insights: Correlation between HDI and total medal count, and medal distribution by HDI category.

Each split aims to partition the data such that the overall variance within each leaf node is minimized.

LSTM Model:

An LSTM unit maintains two states: the cell state (c_t) and the hidden state (h_t). The core LSTM operations are:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$

$$c_t = f_t * c_{t-1} + i_t * \tilde{c}_t$$

$$h_t = o_t * \tanh(c_t)$$

where:

- f_t : forget gate activation
- i_t : input gate activation
- o_t : output gate activation
- \tilde{c}_t : candidate cell state

- σ : sigmoid activation function
- $*$: element-wise multiplication
- **Historical vs Predicted Medal Trends:** Displays line plots comparing historical medal counts with future predictions for selected countries.
- **Model Summary Panel:** Provides key metrics such as RMSE (Root Mean Squared Error), R^2 Score, and last known athlete/medal counts to assess model reliability.
- **Feature Importance Analysis:** Bar chart ranks which factors (e.g., Year, Athlete Count, Previous Medals, Host Status) most influence medal predictions.
- **Breakout Sports Identification:** Highlights top 5 sports showing the highest medal growth for a country, helping pinpoint potential future strengths.

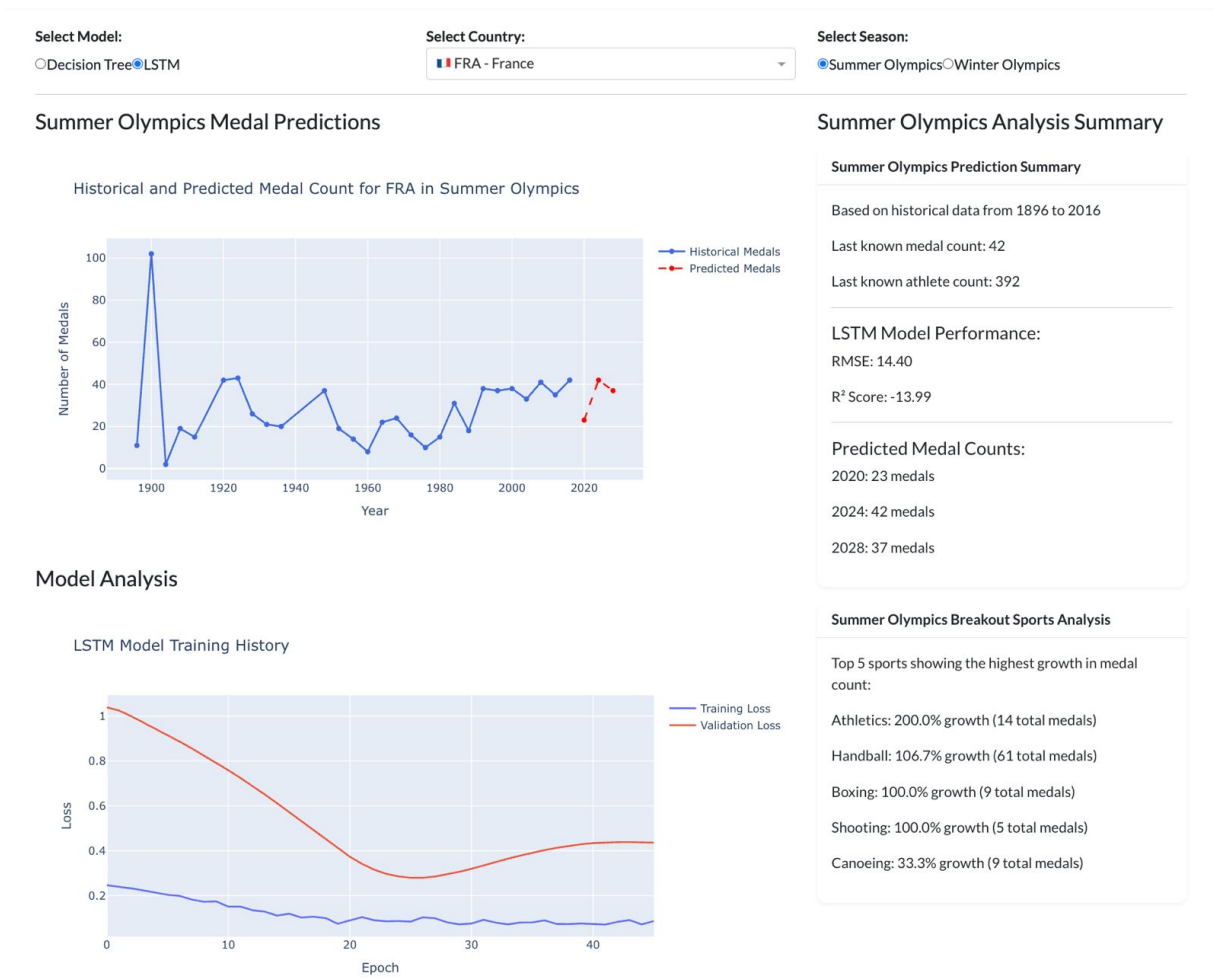


Figure 10: Prediction Model

Insights:

- Medal prediction models show that countries with consistently high athlete counts and historical momentum tend to maintain or grow their medal hauls.
- Feature importance indicates that which feature for say, Athlete Count is the most influential predictors of future medal counts, while which factors have lower impact.

- Breakout sports analysis reveals specific disciplines (e.g., Athletics, Handball) where medal growth is fastest, offering countries strategic focus areas for investment.
- These findings support economic research that links Olympic success to national wealth, infrastructure, health, and education [0].
- However, correlation does not imply causation; other factors such as population size, sports culture, government support, and historical legacy also play critical roles.

6. Design Choices Justification

In designing the Olympus Insight dashboard, we selected visualization types and interaction patterns to maximize clarity, comparability, and user engagement. Below we justify each major choice and highlight how it supports the analytical goals.

6.1 Chart Types

- **Bar Charts** • Used for medal breakdowns (total and by type), top-athlete rankings, and country comparisons. • Facilitate side-by-side category comparisons with precise value perception, outperforming pie charts when comparing more than five categories.
- **Line Charts** • Employed for time-series trends (e.g., medal counts, participation over editions). • Ideal for uneven intervals (Olympic years) with markers clarifying data points and emphasizing growth or decline.
- **Donut/Pie Charts** • Applied sparingly for simple part-to-whole views such as gender distribution. • Limited to two–three slices to avoid clutter and preserve interpretability of proportions.
- **Histograms** • Used to display age, height, and weight distributions of athletes. • Reveal modality, skewness, and outliers in demographic data.
- **Sunburst Charts** • Show hierarchical breakdown of medals by sport and discipline in the Country Profile. • Convey nested proportions in a compact, intuitive layout.
- **Radar Charts** • Compare normalized metrics (e.g., total medals, gold ratio, medals per athlete, seasonal tallies). • Provide a holistic, multivariate snapshot for two countries simultaneously.
- **Bubble Charts** • Encode three dimensions (sport, country, medal count) in the Sport Performance Profile. • Use bubble size and color to highlight relative strengths across disciplines.
- **3D Globe Visualization** • Offers an immersive view of global medal efficiency, leveraging SciVis principles. • Enables intuitive geographic exploration, with rotation and zoom for spatial context.
- **Animated Accumulation** • Depicts cumulative medal counts over time in the Country Comparison module. • Animation emphasizes historical momentum and turning points.

- **Interactive Tables** • Present Olympic Year medal standings in a sortable, paginated format.
- Support search and filtering for detailed, record-level inspection.

6.2 Interactivity and Layout

- **Dynamic Filters** • Season, year range, country, gender, and sport filters enable task-oriented exploration.
- Real-time updates ensure immediate feedback and hypothesis testing.
- **Drilldowns and Navigation** • Clicking on map countries or chart elements navigates to detailed pages (Country Profile, Sport Profile), preserving context and filters.
- Aligns with dashboard best practices for guided, multi-level analysis.
- **Tabbed Interface** • Groups related visualizations (e.g., Medal Trends, Demographics, Sport Deep Dive) to reduce cognitive load.
- Allows users to switch analytical perspectives without losing filter settings.
- **Responsive Design and Aesthetics** • Consistent color palettes (colorblind-friendly), typography, and spacing improve readability.
- Modular layout adapts to various screen sizes, ensuring usability on desktops and tablets.

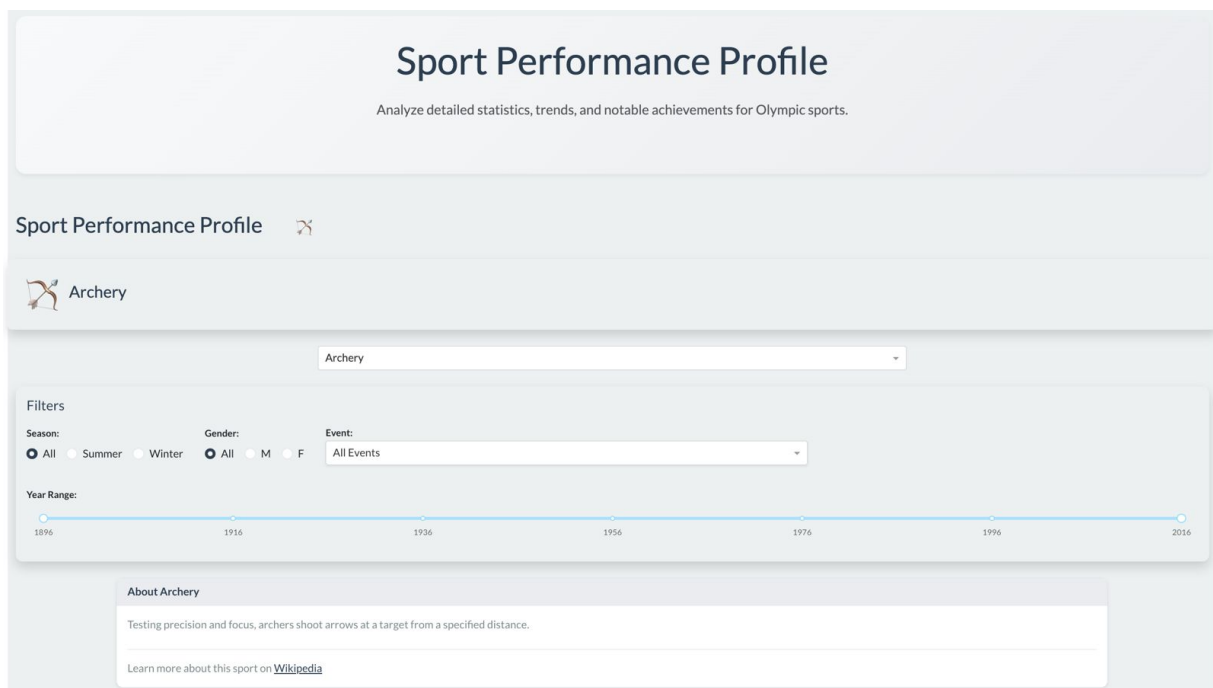


Figure 11: Dynamic Filtering: Select a sport and apply filters to explore athlete demographics, medal counts, and top performers. Shown: Archery overview and filters.

These design choices collectively ensure that Olympus Insight communicates complex, multidimensional Olympic data in a clear, accurate, and engaging manner, empowering users to uncover insights with confidence.

7. Conclusion

The visual analytics modules of Olympus Insight reveal a range of trends across Olympic history. While established sporting nations consistently dominate total medal counts, efficiency metrics—such as medals per athlete—highlight the remarkable achievements of smaller countries that strategically invest in specific sports.

The Globe View, Country Comparison, Country Profile, and Host Country Analysis modules empower users to explore the Olympics at both macro and micro levels. The country comparison and profile modules enable detailed, granular investigations into national performances, while the host country analysis quantitatively validates the often-cited "home advantage" hypothesis.

Despite the depth of insights, certain limitations persist, such as potential inaccuracies in early Olympic records, lack of recent data in the analysis and the absence of detailed socioeconomic and geopolitical factors. Future work could focus on integrating real-time data from ongoing games, refining the predictive models, and expanding the scope to include deeper socioeconomic, environmental, and geopolitical variables.

8. Link to Source Code

The complete codebase and deployment instructions are available on GitHub: <https://github.com/isuniverseok/olympus>

9. How to Run the Project

To set up and run the Olympus Insight dashboard locally, follow these steps:

1. Clone the GitHub repository:

```
git clone https://github.com/isuniverseok/olympus.git
cd olympus
```

2. Create and activate a virtual environment (optional but recommended):

```
python -m venv venv
source venv/bin/activate # On Windows: venv\Scripts\activate
```

3. Install the required packages:

```
pip install -r requirements.txt
```

4. Run the application:

```
cd olympus-dash
python app.py
```

5. Access the dashboard:

Open your browser and navigate to: <http://127.0.0.1:8050/>

10. Team Responsibilities and Individual Contributions

- **Abhishek Choudhary (210037)**

Worked on landing page and acknowledgement page, age and gender-based analysis, and landing page

- **Arnesh Dadhich (210189)**

Designed Country profile page, host analysis, economic comparison data and contributed to frontend development

- **Nirmal Prajapati (210735)**

Focused on extracting insights from age and gender participation, and worked on predictive analytics

- **Rajat Gattani (210813)**

Worked on Globe view, dataset integration, dashboard design, frontend enhancement and website integration

- **Utkarsh Agrawal (211127)**

Worked on dashboard UI/UX, frontend animations, and user interaction optimizations. Contributed to report structuring

- **Vipul Chanchlani (211173)**

Helped manage the dataset integration and deployment. Contributed to Olympic Season wise analysis and Country Comparison.

- **Vishal Himmatsinghka (211175)**

Contributed to dashboard design, layout structuring, responsiveness and user interactivity. Contributed in Sports Profiling and Misc Analysis

- **Yash Verma (211197)**

Prediction model design along with LSTM and Decision Tree model. Contributed to report structuring.

References

- Kaggle, *Olympic Games Medals Dataset (1896-2018)*, Available at: <https://www.kaggle.com/datasets/piterfm/olympic-games-medals-19862018>
- World Bank, *World Development Indicators*, Available at: <https://databank.worldbank.org/source/world-development-indicators>
- LLM Models, *OpenAI, Claud, Perplexity*,
- Stack Overflow, *At Multiple instances for troubleshooting* , Available at: <https://databank.worldbank.org/source/world-development-indicators>
- World Bank, *World Development Indicators*, Available at: <https://databank.worldbank.org/source/world-development-indicators>
- Michael Klein, *What Determines Countries' Olympic Success?*, Available at: <https://econofact.org/what-determines-countries-olympic-success>