

Toward Los Angeles 2028: A Prognosticative Study on Olympic Medal Performance

Summary

This comprehensive study leverages the upcoming 2028 Los Angeles Summer Olympics as an opportunity to develop predictive models based on Olympic medal data since 2000, aiming to forecast the medal standings and assess the impact of various influencing factors, such as coaching and event changes. The study is designed to gain an in-depth understanding of the influence of these factors on Olympic medal winners and to provide strategic guidance for National Olympic Committees (NOCs) worldwide.

The research employs advanced analytical techniques, including XGBoost, Cox survival analysis, and the Difference-in-Differences (DID) method, to evaluate the potential of countries to secure medals. The study addresses several key issues: predicting the 2028 medal standings, estimating the potential for new countries to win their first medals, analyzing the significance of specific sports events, assessing the "coach effect," and revealing unique insights into the Olympic medal table.

For countries that have already secured medals, we predict the 2028 Los Angeles Summer Olympics medal standings through data preprocessing, feature engineering, and the construction of an XGBoost model.

Regarding countries that have not yet won medals, we employ the Cox Proportional Hazards model to forecast the likelihood of these nations earning medals. We conducted data preprocessing, feature engineering, model training, and evaluation. Through analysis, the model identified key factors affecting the number of medals, such as the number of athletes and event diversity, and used survival curves to predict the probability of countries winning medals in the 2028 Olympics.

For the "coach effect," we assessed the impact of distinguished coaches like Lang Ping and Sandro Damilano on the Olympic achievements of various nations. We established a Difference-in-Differences (DID) model, supplemented by cross-period DID and dynamic DID models, to evaluate the true impact of coaches by comparing performance metrics before and after their involvement, as well as across countries with different coaching scenarios. The diminishing influence of time on the rate of winning medals suggests that the coach's impact on performance increment gradually decreases.

This paper provides a clear overview of the assumptions and justifications for the study, including consideration of countries participating in the Olympics after 2000, and a stratified forecasting approach based on the level of sport development. The study also acknowledges the limitations and potential weaknesses of the model and makes recommendations for Olympic committees.

Keywords: Medal prediction, XGBoost, Cox survival analysis, DID method, the great coach effect

Contents

1	Introduction	2
1.1	Background	2
1.2	Restatement of the Problem	2
1.3	Our work	2
2	Assumptions and Justification	3
3	Problem 1: Predicting the Medal Table for the 2028 Los Angeles Summer Olympics	4
3.1	Data Preparation	4
3.1.1	Countries Classification	4
3.2	Model building	5
3.3	Model Prediction Result	8
3.3.1	Leading Sports Countries	8
3.3.2	Countries with Medium Sports Strength	11
4	Problem 2: Predictions for Non-Medal Countries in the 2028 Olympics	13
4.1	Model Establishment	13
4.1.1	Model Development	13
4.1.2	Data Preprocessing	13
4.1.3	Feature Engineering	13
4.1.4	Model Training	14
4.1.5	Model Evaluation	14
4.2	Result	15
5	Problem 3: The impact of a head coach on team performance	16
5.1	Model Establishment	16
5.2	Conclusion	19
6	Problem 4: Other Influencing Factors–The Impact of Changes in the Number of Sports Events on Competition Outcomes	19
7	Strength and Weakness	21
7.1	Strength	21
7.2	Weakness and Further Discussion	21
	Reference	21
8	Memorandum for The Coaches	23
	Appendices	24
	Appendix A Feature Importance Plot of different countries	24
	Appendix B Features extracted for XGBoost Model	24

1 Introduction

During the Olympic Games, there is a prevalent focus on both the gold and overall medal standings of participating nations. These rankings are subject to a multitude of influencing factors, encompassing athletes, sports events, coaching, and numerous other elements. This report aims to analyze the impact of these factors on the fluctuation of the standings and to construct predictive models to forecast the outcomes of the 2028 Olympic Games. The analysis will delve into the intricate interplay of these variables to discern their relative contributions to the medal tally, thereby enhancing the predictive accuracy of the models.

1.1 Background

The 2024 Paris Olympics illuminated the profound impact of various factors on a nation's medal count. This event not only captivated sports enthusiasts worldwide but also highlighted a complex, yet underexplored, aspect of international sports competition: the ability of nations to shift their performance dynamics through factors such as historical trends, Olympic program changes, and the influence of elite coaches. The significance of these factors in determining medal outcomes has become a key area of interest. For example, the "coach effect" has been observed in sports like gymnastics and volleyball, where renowned coaches can significantly impact team performance. This underscores the need for a sophisticated model that captures the essence of these dynamics and explains the underlying rationale.

1.2 Restatement of the Problem

The challenge at hand involves the analysis of data from the 2024 Paris Olympics with the objective of gaining insights into the dynamics of Olympic medal performance. The primary tasks to be addressed include:

- **Medal Forecasting.** Predict the medal standings for the 2028 Los Angeles Summer Olympics, providing a range of outcomes. Identify which countries are likely to improve or regress compared to 2024.
- **Inclusion of Newcomers.** Estimate how many countries without prior medals will earn their first in the next Olympics and assess the confidence in this prediction.
- **Event Analysis.** Consider the specific events of the Olympics, exploring the relationship between event types and medal counts. Determine which sports are most significant for various countries and how the host's chosen events influence results.
- **Coach Effect Evaluation.** Analyze data for evidence of changes due to the "coach effect," estimating its contribution to medal counts. Select three countries and suggest sports where hiring renowned coaches could be impactful.
- **Model Insights.** Uncover unique insights about the Olympic medal table and explain how these can inform national Olympic committees.

1.3 Our work

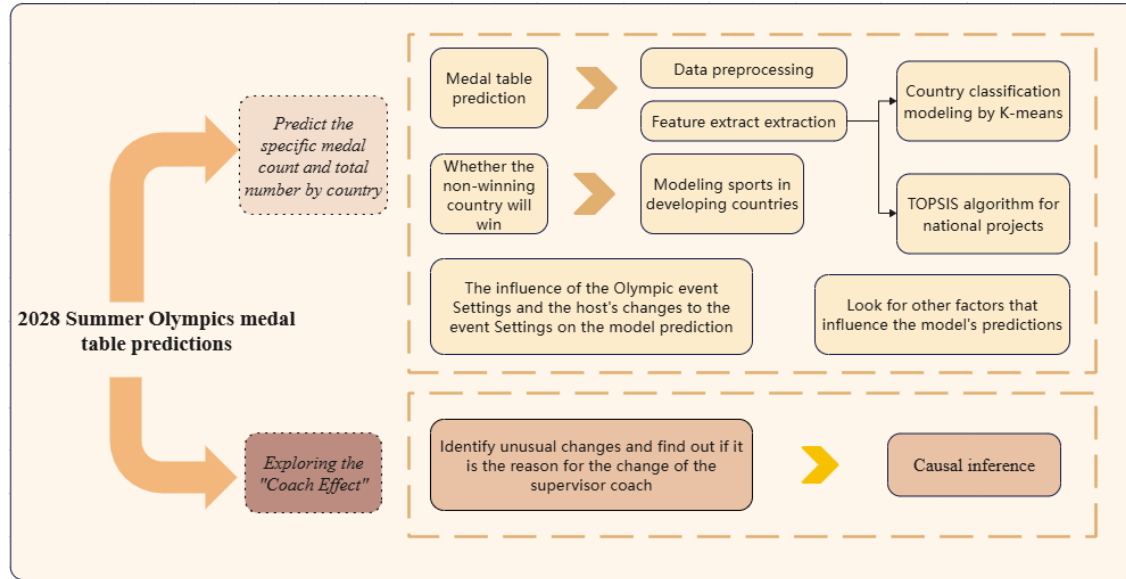


Figure 1: Our Work

2 Assumptions and Justification

To simplify the problem and make it convenient for us to simulate real conditions, we make the following basic assumptions, each of which is properly justified.

- **We only consider the countries that have participated in the Olympic Games after 2000.** The dataset provided in the question has a wide time span. Some countries have undergone changes such as disintegration or consolidation (for instance, the Soviet Union, West Germany and East Germany), and these countries will no longer be within the scope of our model.
- **The Russian Olympic Committee will not be taken into account within the scope of the model consideration.** Due to political factors, Russian athletes are only allowed to participate in competitions as individuals in a neutral capacity. As a result, relevant data has been missing since 2024, and it is still undetermined whether the Russian Olympic delegation can participate normally in the 2028 Los Angeles Olympics. Therefore, this model does not consider making predictions for them..
- **Considering that the factors affecting the medals of countries with different sports levels are quite different, we adopt stratified prediction.** Through K-means clustering and analysis of the historical medal table, we assume that leading sports nations will win 50% of the total medals, mid-level countries will share the remaining 45%, and 5% will be won by sports developing countries for the first time.
- **We suppose the number of total medals and gold medals for 2028 Los Angeles Olympics has been determined.** According to Wikipedia[2], the 2028 Summer Olympics in Los Angeles will have a total of 35 sports and 50 disciplines with 325 events, which means that 325 gold medals and 975 medals will be awarded.

where we define the main parameters while specific value of those parameters will be given later.

3 Problem 1: Predicting the Medal Table for the 2028 Los Angeles Summer Olympics

3.1 Data Preparation

Firstly, we import the csv files named *summerOly_athletes*, *summerOly_hosts*, *summerOly_medal_counts* and *summerOly_programs*, respectively, and name them *df_athlete*, *df_host*, *df_medal* and *df_programs*. Then we add the corresponding NOC for each country in *df_host* and *df_programs*.

Then we break down the medal table by using *df_athlete* to analyze the composition of the medals, which means counting the number of medals each country has won in each corresponding event.

3.1.1 Countries Classification

Taking into account the varying levels of sport development between different countries, the factors determining the number of Olympic medals won by each country and the magnitudes of their respective influences are also different. Therefore, we have decided to adopt the method of establishing models by stratification.

We first identify and mark those countries that sent athletes to compete after 2000 but never won any medals as developing countries in sports.

Using K-means clustering, we divide the remaining countries into two groups: leading sports countries and median sports countries, then analyze the average number of gold per Olympic Games, the average total number of medal per Olympic Games, the average number of participation, the average total number of gold medals, the average total number of medal and the proportion of gold medals to total medals among countries that have won medals. From Figure 2 we can observe significant difference between the two countries group in these statistical indicators.

Category	Leading Sports Nations	Median Sports Nations
Number of Countries	8	61
Average Gold Medals per Edition	29.59	2.85
Average Total Medals per Edition	78.78	9.68
Average Participation Times	6.8	6.5
Average Total Gold Medals	201.2	19.6
Average Total Medals	533.5	66.5
Gold Medal Ratio (%)	37.7	29.5

Table 1: Statistical Indicators for Leading and Median Sports Nations

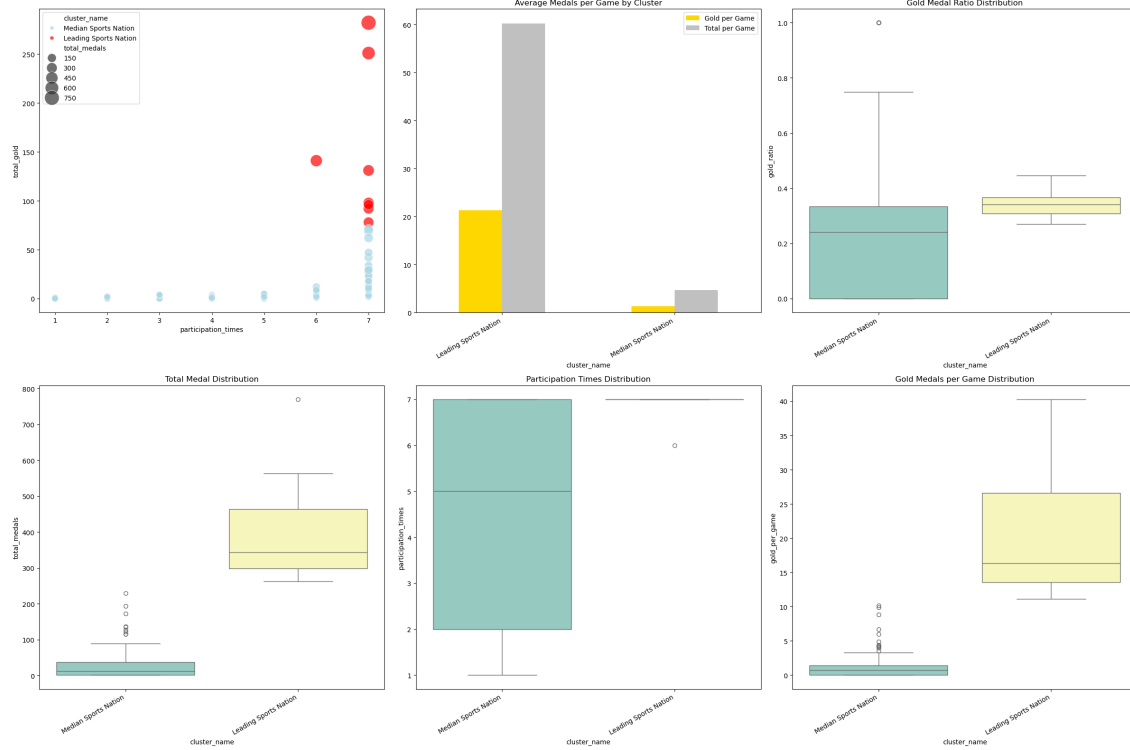


Figure 2: Clustering Result

3.2 Model building

In the initial phase of our research, we undertook a meticulous process of data preprocessing to ensure the integrity and utility of the dataset for subsequent model training. This process encompassed three primary steps: data cleaning, feature engineering, and data normalization. During the data cleaning stage, we eliminated redundant records, addressed missing values, and standardized the nomenclature of countries to achieve consistency. In the feature engineering phase, we extracted a suite of features, including the number of participating athletes, the diversity of sports events, and the historical medal tally. Subsequently, we employed the MinMaxScaler technique to normalize the feature set, thereby mitigating the potential for scale-induced biases in the model.

Given the complexity of the factors that affect a country's medal count, we consider using the XGBoost model to aggregate the features to make accurate predictions about the number of medals won by each country.

In total, we extracted the following features in Table 6. What we need to emphasize here is the construction of *Tech_value*, *Power_value*, *Arti_value*, *Water_value*, *Team_value*, *Endu_value* and *Mixed_value*. In order to take the quantity and type of specific Olympic Games' events into consideration, we apply the TOPSIS method to calculate the scores based on the number of medal, gold, silver and bronze each country has won in various sports during the past three Olympic Games. Then we classify the sports into the following categories based on their features: technical sports, power sports, artistic sports, water sports, team sports, endurance sports and other

mixed sports.

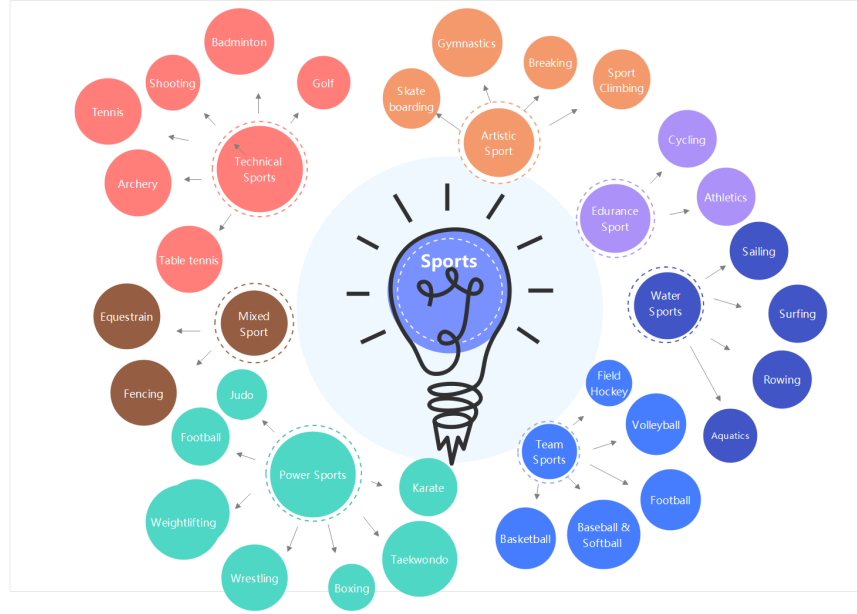


Figure 3: Sports classification

We calculated the proportion of events included in these sports classification in each Olympic Games, then weighted the scores obtained by the previous topsis method according to these proportion.

Then we use the XGBoost algorithm to predict the number of Olympic medals and gold medals for each country in 2028 based on the extracted features. XGBoost is an optimized distributed gradient boosting library, aiming to achieve efficiency, flexibility and portability. It implements machine learning algorithms under the Gradient Boosting framework. XGBoost provides parallel tree boosting (also known as GBDT, GBM), which can quickly and accurately solve many data science problems.

The objective function of XGBoost consists of two parts: Loss function L (1): Measures the gap between the predicted values and the true values of the model. Regularization term Ω (2): Controls the complexity of the model and prevents overfitting. T is the number of leaves in the tree, ω is the weight of the leaf nodes, and γ and λ are regularization hyperparameters.

$$\mathcal{L}(y, \hat{y}) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k), \quad (1)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2, \quad (2)$$

For sample i , its predicted value is the weighted sum of all decision trees:

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad (3)$$

where $f_k \in \mathcal{F}$ indicates the k th tree, and \mathcal{F} is the function space that encompasses all trees. Using Taylor expansion to perform second-order approximation on the objective function:

$$\mathcal{L}^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t), \quad (4)$$

where

$$g_i = \frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i}, \quad h_i = \frac{\partial^2 l(y_i, \hat{y}_i)}{\partial \hat{y}_i^2}, \quad (5)$$

g_i, h_i are respectively the first-order and second-order derivatives of the loss function. Splitting leaf nodes to minimize a loss function:

$$\mathcal{L}_{\text{split}} = \frac{1}{2} \left[\frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I} g_i)^2}{\sum_{i \in I} h_i + \lambda} \right] - \gamma, \quad (6)$$

where I_L and I_R are respectively the left and right child nodes after the split, and γ is the regularization coefficient of the leaf node.

The final model is continuously optimized through the gradient boosting method. At each step, a new tree $f_t(x)$ is added to minimize the objective function.

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i). \quad (7)$$

During model training, we leveraged grid search to meticulously tune the model parameters, and adopted time series cross-validation to evaluate the performance of our model. It is applicable to scenarios where the data exhibits time dependence and does not meet the assumption of random splitting. Unlike the ordinary k-fold cross-validation, time series cross-validation retains the sequence of time and avoids using future data to predict the past. The data of the validation set is always the future data in time. It is also worth noting that we trained the models separately for different category of countries to capture their unique characteristics and trends more accurately. Finally, in the model evaluation phase, we utilized a suite of metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2), to assess the model's performance. The model demonstrated satisfactory performance on both the training and validation datasets, thereby validating the robustness of our modeling approach.

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (8)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

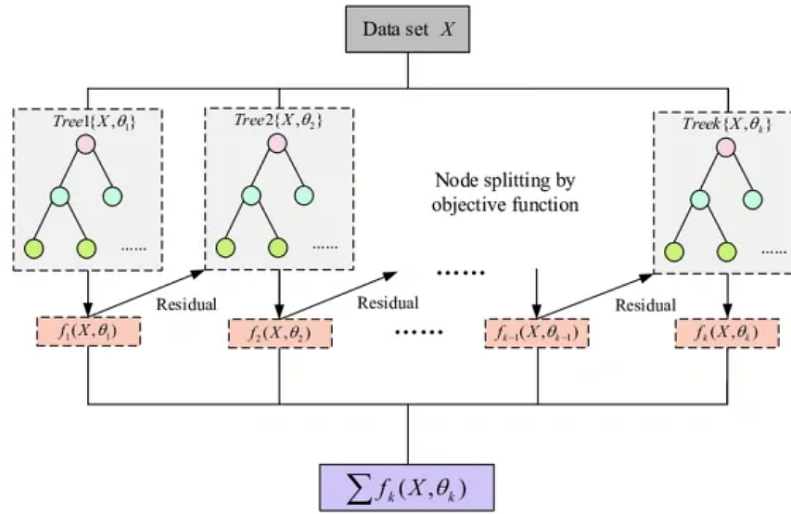


Figure 4: XGboost model

3.3 Model Prediction Result

3.3.1 Leading Sports Countries

Due to the assumption of limitations on the total number of medals won by leading sports nation, we add a nonlinear weighting adjustment to the XGBoost forecast based on the forecast results, taking into account medal trends and the relative strength of countries. The result of cross-validation shows that the average absolute value of the error between the predicted number of medals and the actual number of medals is approximately 8.12 and the square root of the average of the squared errors between the predicted values and the actual values is approximately 9.99. The model explains approximately 86.1% of the variance of the target variable. We also obtained the corresponding 95% confidence intervals by adding random noise to the features through the Bootstrap method.

Table 2: Time series cross validation result for medal prediction model

Dataset	R^2	MAE	RMSE
train data: 2004,2008,2012 test data: 2016	0.804	7.90	11.96
train data: 2004, 2008, 2012, 2016 test data: 2020	0.857	9.01	10.12
train data: 2004, 2008, 2012, 2016, 2020 test data: 2024	0.923	7.46	7.88
Average result of cross-validation (standard deviation)	0.861 (± 0.097)	8.12 (± 1.31)	9.99 (± 3.34)
Test data (2024)	0.925	7.04	7.04

The prediction results for the top eight sports nations in the 2028 Los Angeles Olympic Games highlight general characteristics where the United States (USA) is likely to maintain its leading position with a broad base of athletic talent and strong performances across multiple sports including athletics, swimming, and gymnastics. China (CHN) is expected to remain a dominant force with a focus on excelling in sports such as gymnastics, diving, and table tennis where they have historically

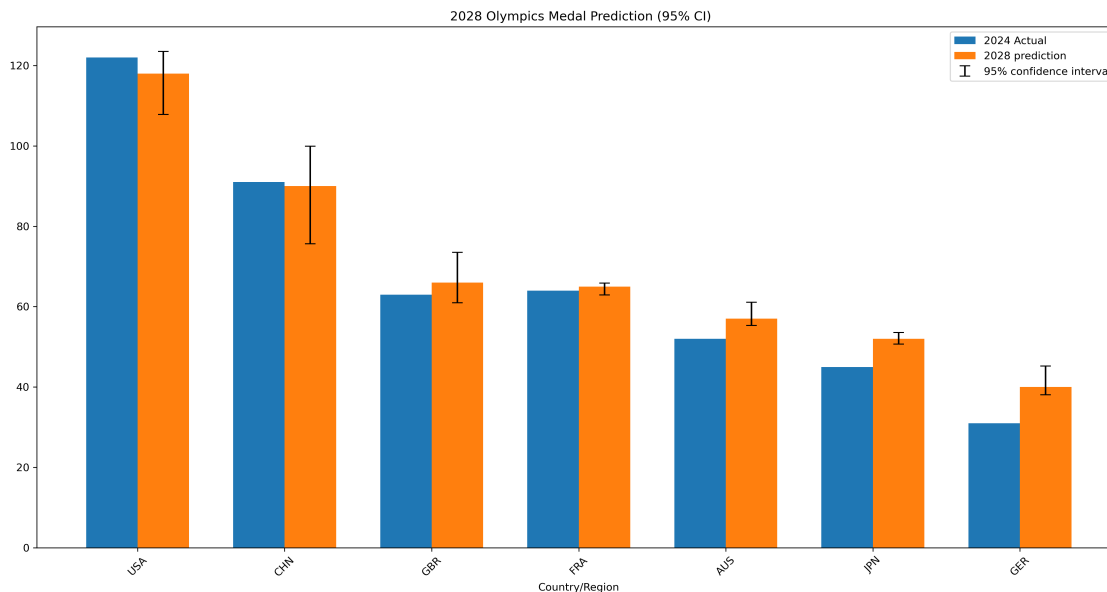


Figure 5: 2028 Olympics Medal Prediction(95% CI)

demonstrated exceptional skill and consistency. Japan (JPN) is projected to see an increase in medal count driven by continued strength in judo, wrestling, and swimming. Other nations like Great Britain (GBR), France (FRA), Germany (GER), Australia (AUS), and Canada (CAN) are also expected to perform well, each with their own strengths in specific sports contributing to their overall medal tally.

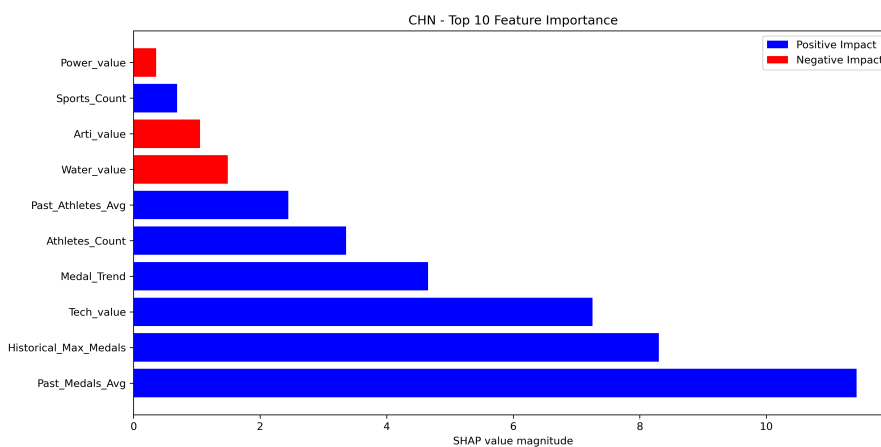


Figure 6: Feature importance plot for medal prediction model of CHN

The predicted slight decrease in the total number of medals won by the USA and CHN in the 2028 Olympics compared to 2024 may be attributed to increased competitiveness from other countries, changes in Olympic events, and potential shifts in performance or strategic focus. Emerging sports programs and diversification of medal-winning nations could reduce the dominance of leading countries. Additionally, factors like new sports inclusion and generational transitions may influence outcomes. Despite this, the 95% confidence interval indicates that medal counts for the

USA and CHN could still increase, highlighting the uncertainty and complexity of such predictions, influenced by both statistical variability and unforeseen developments.

In order to explore the relationship between the events held and the medals won by each country, we drew the feature importance graph corresponding to the predicted results of the XGBoost model and the features of each country, and selected the most representative CHN as the representative for analysis. According to the Figure 7, we can observe that the feature *Power_value* has a negative effect on the predicted value. As a country with a monopoly advantage in weightlifting, China is faced with the decrease of events in weightlifting, which is bound to have a negative impact on the number of medals, and also confirms the conclusion in the figure.

Similarly, we trained a model for the 2028 gold medal predictions of leading sports countries. The prediction and cross-validation results are showing in table 3.3.1 and Figure 7. Both China (CHN) and the United States (USA) are projected to experience slight decreases in gold medals, with predictions of 38 and 37 respectively, down from 40 and 39 in 2024, representing declines of 5.0% and 5.1%. However, their 95% CIs suggest the possibility of maintaining or slightly increasing their medal counts. On the other hand, countries like Japan (JPN), Australia (AUS), France (FRA), Great Britain (GBR), and Germany (GER) are expected to see increases. Overall, while traditional powerhouses like CHN and USA may face slight declines, European countries like FRA, GBR, and GER are expected to perform better, potentially leading to a more balanced distribution of gold medals.

Table 3: Time series cross validation result for gold prediction model

Dataset	R ²	MAE	RMSE
train data: 2004,2008,2012 test data: 2016	0.676	4.69	6.60
train data: 2004, 2008, 2012, 2016 test data: 2020	0.815	3.85	4.79
train data: 2004, 2008, 2012, 2016, 2020 test data: 2024	0.899	2.99	3.54
Average result of cross-validation (standard deviation)	0.797 (± 0.184)	3.85 (± 1.39)	4.98 (± 2.51)
Test data (2024)	0.899	2.99	3.54

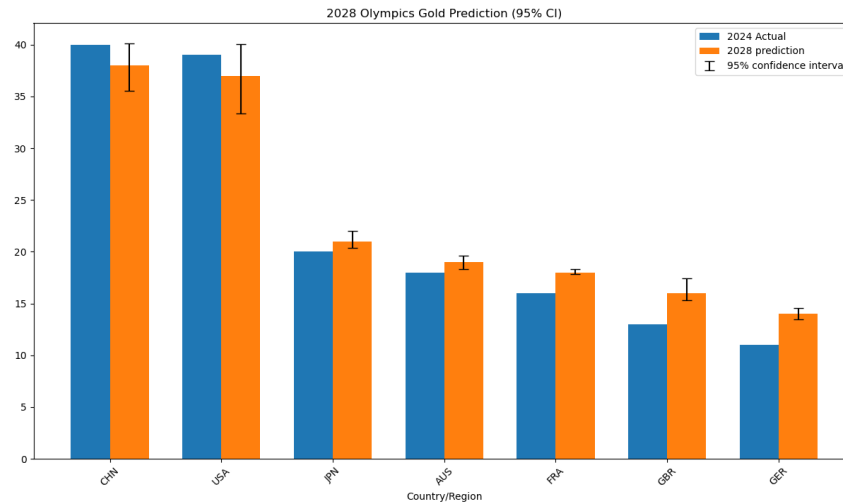


Figure 7: 2028 Olympics Gold Prediction(95% CI)

3.3.2 Countries with Medium Sports Strength

The Figure 9 provides medal predictions for the 2028 Olympics for Countries with Medium Sports Strength, along with their 95% confidence intervals (CIs), changes compared to 2024, and confidence scores. Countries like Italy (ITA), Netherlands (NED), and South Korea (KOR) are expected to see slight increases in medal counts, with ITA showing the largest growth at 5.6%. Canada (CAN) is predicted to maintain its medal count at 27. Conversely, several countries are expected to experience declines, such as New Zealand (NZL) and Brazil (BRA) with a 5.0% decrease, and Spain (ESP) and Hungary (HUN) showing significant drops of 11.1%. The steepest decreases are observed in countries like Norway (NOR), Greece (GRE), and Switzerland (SUI), all projected to lose 25% of their medals compared to 2024. Confidence scores indicate high reliability for most predictions (e.g., ITA, NED, and KOR at 100%), but lower scores for countries like Cuba (CUB) and Switzerland (SUI), suggesting greater uncertainty. Overall, while a few nations are poised for slight gains, many are likely to face challenges in maintaining their medal count.

Table 4: Time series cross-validation result for medal prediction model

Dataset	R ²	MAE	RMSE
train data: 2004,2008,2012 test data: 2016	0.954	0.25	0.46
train data: 2004, 2008, 2012, 2016 test data: 2020	0.892	0.36	0.72
train data: 2004, 2008, 2012, 2016, 2020 test data: 2024	0.925	0.33	0.77
Average result of cross-validation (standard deviation)	0.924 (± 0.05)	0.32 (± 0.09)	0.65 (± 0.27)
Test data (2024)	0.925	0.33	0.77

Notably in the Figure 9, the Netherlands (NED), South Korea (KOR), and Italy (ITA) are expected to see increases in their medal counts, with NED projected to win 17 golds (+13.3%) and KOR 14 golds (+7.7%), supported by high confidence scores and narrow prediction intervals.

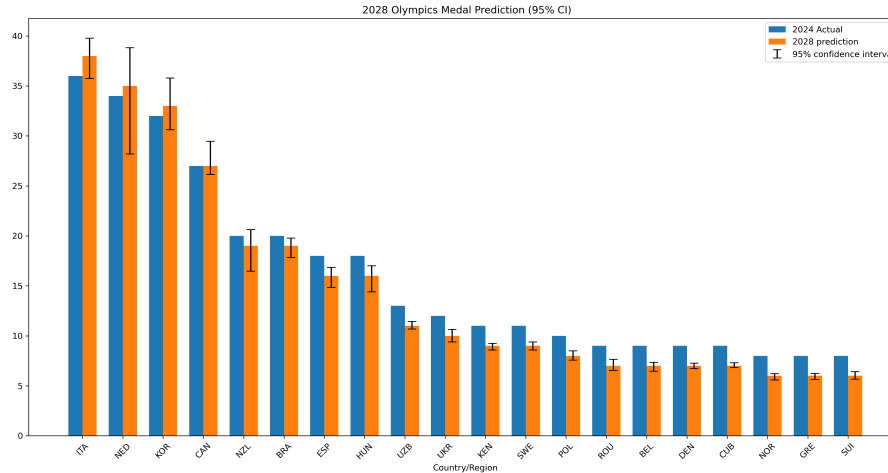


Figure 8: 2028 Olympics Medal Prediction(95% CI)

Table 5: Time series cross validation result for gold prediction model

Dataset	R ²	MAE	RMSE
train data: 2004,2008,2012 test data: 2016	0.902	1.11	1.94
train data: 2004, 2008, 2012, 2016 test data: 2020	0.898	1.14	2.33
train data: 2004, 2008, 2012, 2016, 2020 test data: 2024	0.964	0.82	1.35
Average result of cross-validation (standard deviation)	0.921 (± 0.06)	1.02 (± 0.29)	1.87 (± 0.80)
Test data (2024)	0.977	0.60	1.08

Conversely, countries like Ireland (IRL), Ukraine (UKR), and Romania (ROU) face predicted decreases of 33.3%, with relatively lower confidence scores indicating higher uncertainty. Some countries, including Canada (CAN), Uzbekistan (UZB), and Hungary (HUN), are predicted to maintain their 2024 gold medal counts

Sports developing countries generally show the same or lower golds in the 2028 Olympic medal count projections, reflecting their relatively weak competitiveness in global competitive sports. The forecast data reveal the scope for these countries to grow in medal count, although their growth may be smaller compared to leading sports countries. The adjusted projections and confidence intervals provide a reasonable expectation of the future performance of these countries, with the percentage change and confidence scores further explaining the reliability of the projections and the potential volatility of medal count changes. The medal projections for these mid-tier nations highlight their potential for development in specific sports, as well as the progress that may be achieved in the following Olympic cycle.

Regarding the predictions for countries that have not yet secured any medals, we will elaborate in detail in the subsequent section.

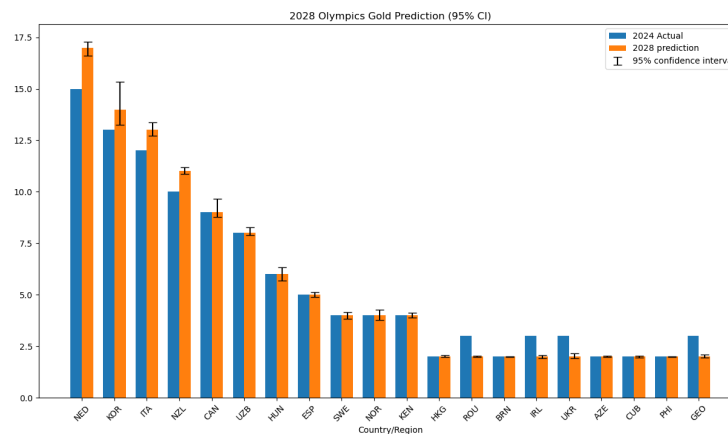


Figure 9: 2028 Olympics Medal Prediction(95% CI)

4 Problem 2: Predictions for Non-Medal Countries in the 2028 Olympics

This section presents an analytical examination of Olympic medal data, with a particular focus on nations that have not yet secured any medals. The analysis employs a Cox Proportional Hazards Model to predict the likelihood of these countries earning medals in the 2028 Olympic Games.

4.1 Model Establishment

In the "Model Establishment" section of the academic paper, we detail the process of model development, encompassing data preprocessing, feature engineering, model selection, training, and evaluation. The following provides an elaborate description of the model-building process, integrating Figure 10 (the heatmap analysis):

4.1.1 Model Development

Our study aims to predict the number of medals won by countries that have not yet secured any Olympic medals in the 2028 Olympic Games. To achieve this, we established a survival analysis model based on the Cox Proportional Hazards model, which considers several features related to a country's Olympic performance.

4.1.2 Data Preprocessing

Initially, we collected historical Olympic medal data and performed data cleaning to ensure accuracy and consistency. The dataset included the number of athletes (Athletes_Count), the number of sports events (Events_Count), the diversity of events (Events_Diversity), and the frequency of participation (Participation_Frequency) for each country.

4.1.3 Feature Engineering

During the feature engineering phase, we normalized the original features to eliminate scale differences among various metrics. Additionally, we constructed a correlation matrix to assess

the relationships between features. The heatmap (Figure ??)revealed a strong correlation between the number of sports events (Events_Count) and event diversity (Events_Diversity), with a correlation coefficient of 0.88, indicating potential redundancy in the model. Conversely, the correlation between the number of athletes (Athletes_Count) and participation frequency (Participation_Frequency) was low (correlation coefficient of 0.042), suggesting these features provide relatively independent information. Based on the correlation analysis, we selected the most informative features, including the number of athletes, the number of events, event diversity, and participation frequency for model training.

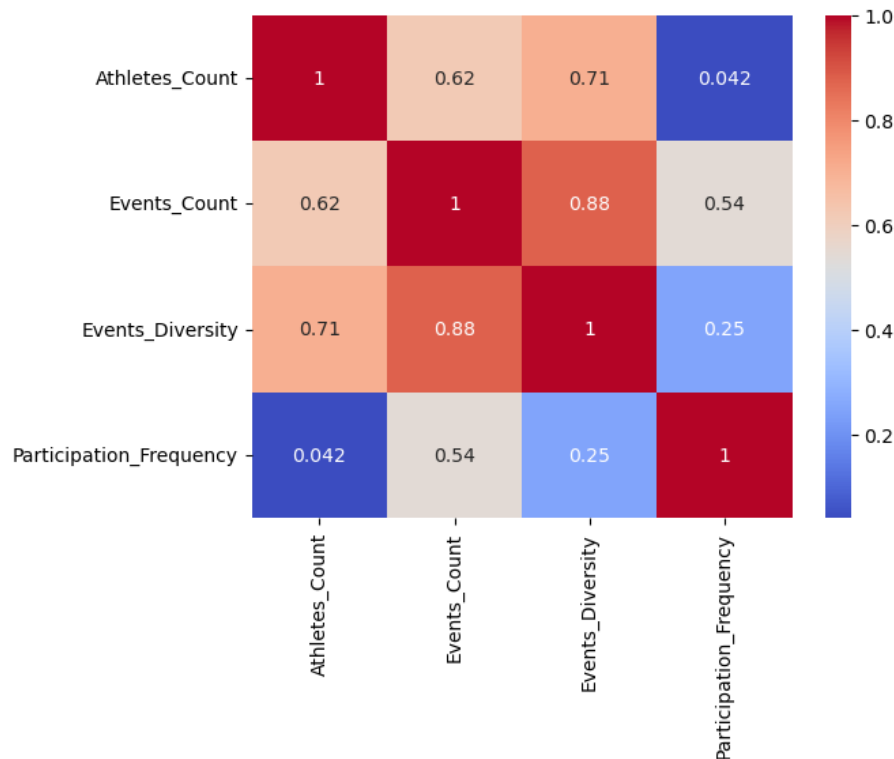


Figure 10: Correlation Heatmap of Olympic Participation Metrics

4.1.4 Model Training

We employed the Cox Proportional Hazards model (CoxPHFitter), a survival analysis model suitable for time-to-event data. The model's duration was defined as the time elapsed since a country's first participation in the Olympics until its first medal win, with the event indicator set to 1, representing at least one medal won. During model training, we used cross-validation to evaluate the model's performance and adjusted model parameters to optimize prediction accuracy. Model evaluation metrics included the concordance index (C-index) and cross-validation scores.

4.1.5 Model Evaluation

In the model evaluation phase, we calculated the importance of each feature and analyzed the model's prediction outcomes. Feature importance analysis indicated that the number of athletes and event diversity were key factors in predicting the number of medals. Furthermore, we visualized

survival curves for different countries to predict the probability of winning medals in the 2028 Olympics. The survival curves displayed the decreasing trend of winning medals over time since the first participation.

4.2 Result

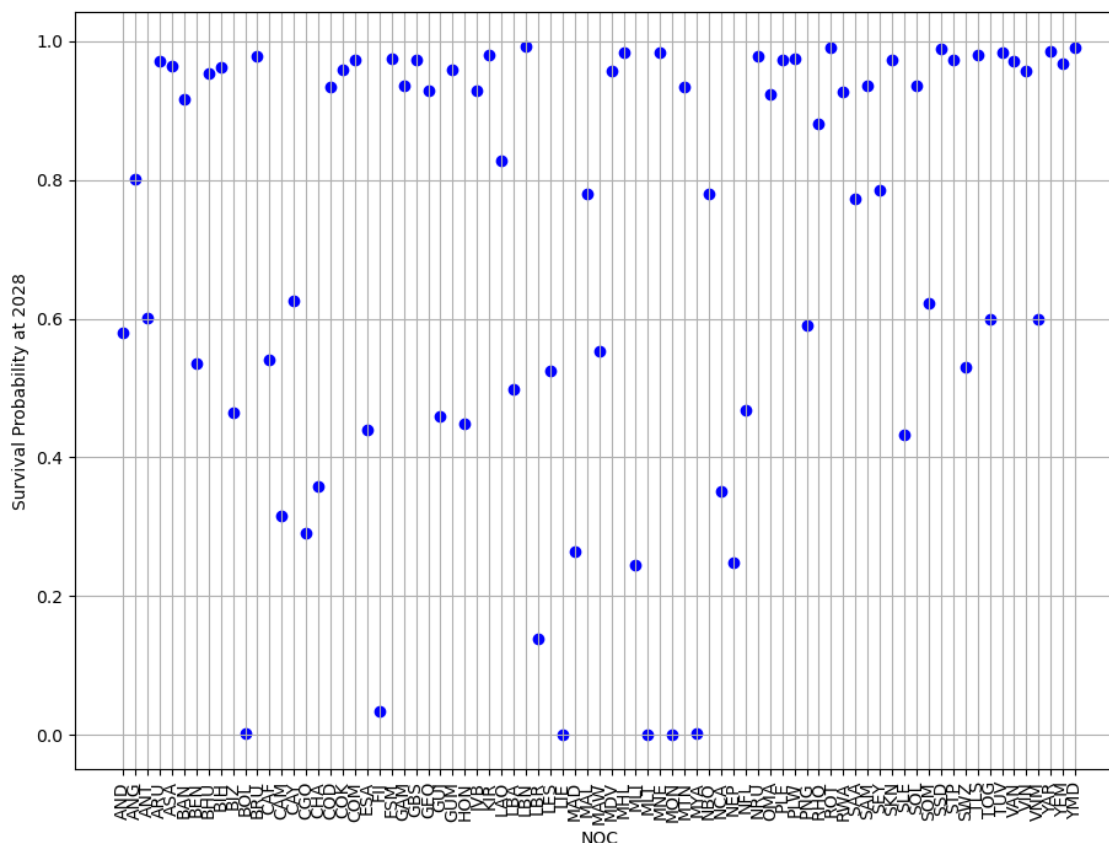


Figure 11: The probability of each country not winning a medal by 2028

The chart (Figure 11) illustrates the probability of not winning a medal, for each country (NOC) by the year 2028. It is observed that the majority of countries have a high survival probability, indicating a lower likelihood of winning a medal before 2028. Conversely, a minority of countries exhibit a lower survival probability, suggesting a relatively higher chance of securing a medal before the year 2028.

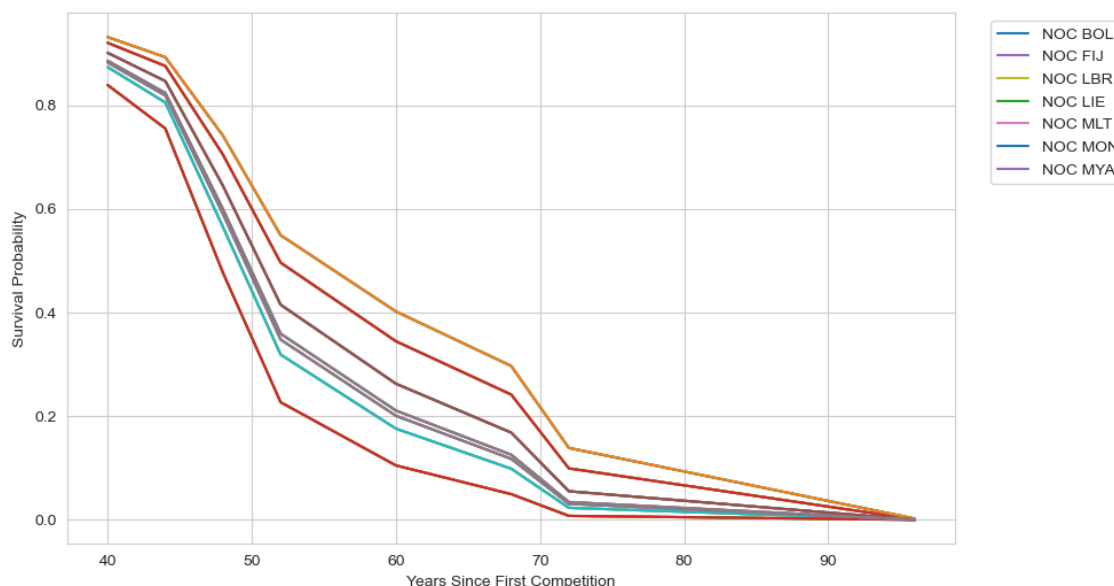


Figure 12: Survival Curves for Countries with Low Award Risk (<0.2) in 2028

The figure (Figure 12) presents survival curves for a selection of countries identified by their National Olympic Committee (NOC) codes, which depict the probability of these nations not securing a medal by 2028. Specifically, the curves are for countries with a low risk of remaining without a medal (survival probability < 0.2). The x-axis measures the duration in years since each country's initial competition, while the y-axis quantifies the survival probability. Notably, countries such as NOC BOL (Bolivia), NOC FIJ (Fiji), and NOC LBR (Liberia) exhibit distinct trajectories in their survival probabilities, with NOC LBR showing a particularly steep decline, suggesting a higher likelihood of winning a medal sooner than others. Conversely, NOC MON (Monaco) and NOC MYA (Myanmar) display a more gradual decrease, indicating a lower immediate risk of medal acquisition. These curves collectively offer a comparative analysis of the medal-winning potential for these low-risk countries over time.

5 Problem 3: The impact of a head coach on team performance

5.1 Model Establishment

In this study, we delve into the impact of eminent coaches on a nation's Olympic achievements, with a particular focus on the enhancement in medal tally brought about by globally recognized coaches such as Lang Ping and Sandro Damilano during their tenure in various countries. To quantify the "great coach effect," we employ the Difference-in-Differences (DID) model, complemented by cross-period DID and dynamic DID models, which allow us to evaluate the true impact by contrasting performance metrics before and after the coaches' involvement, as well as across countries with different coaching scenarios.

What requires special emphasis is the construction of various value metrics that reflect different aspects of Olympic performance. These include Technical_value, Power_value, Artistic_value, Water_value, Team_value, Endurance_value, and Mixed_value. To account for the quantity and

type of events in the Olympic Games, we apply the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) method to calculate scores based on the number of medals, including gold, silver, and bronze, that each country has won in various sports during the past three Olympic Games.

We then categorize the sports into technical sports, power sports, artistic sports, water sports, team sports, endurance sports, and other mixed sports based on their characteristics. This classification enables us to analyze the specific areas where the influence of great coaches is most pronounced.

By leveraging the DID models, we can control for time-varying and unobservable factors, thus providing a more accurate assessment of the coaches' impact. This approach not only captures the immediate effects of coaching but also the long-term influence on a country's Olympic performance. Our findings are expected to shed light on the strategic importance of coaching in enhancing a nation's competitive edge in the global sporting arena.

Our analysis is grounded in a dataset encompassing two emblematic Olympic sports: volleyball, with a focus on the U.S. and China, and racewalking, centered on China and Italy. These sports epitomize the dual of Olympic events, where one is predicated on team dynamics and the other intricately blends individual and collective efforts.(Figure 16)

We hone in on two key dependent variables: the scoring rate, which is a weighted metric based on the count of medals accrued, and the score change rate, which monitors the trajectory of improvements over successive Olympic cycles. Building upon the features extracted from our prior models, we integrate additional external variables to bolster the reliability and robustness of our analytical framework.

The DID model outcomes(Figure 13d Figure 13i) demonstrate that the appointment of a renowned coach is correlated with a marked escalation in the scoring rate, translating to an approximate increment of 0.1 and ensuring at least one elevated tier in the medal hierarchy during the most recent Olympic cycle. As we observe in the subsequent cycles, the initial surge in performance tends to taper off, leading to a plateau in the scoring rate. This trend suggests that while the immediate influence of illustrious coaches is palpable, their legacy endures, fostering sustained enhancements in team performance over an extended period. Our study thus underscores the pivotal role of coaching excellence in catalyzing both short-term gains and long-term advancements in a nation's Olympic standing.

China, India, and Brazil could greatly benefit from hiring "star coaches" to enhance their performance in key sports. For China, focusing on football could help improve the national team's tactical and technical levels, inspire youth participation, and boost the country's international reputation in the sport. India, with its vast population, could unlock its potential in athletics by hiring experts in events such as sprinting, long jump, and throwing, thus enhancing competitiveness, improving coaching standards, and increasing its medal tally in global competitions. Brazil, already a football powerhouse, could invest in basketball by bringing in top NBA or European league coaches to increase the national team's performance, strengthen youth training systems, and popularize the sport through increased sponsorship and visibility. By strategically investing in these sports, these countries can achieve short-term success while fostering sustainable growth in their athletic programs.

Figure 13: Comparison of scoring rates and changes between different countries and models.



5.2 Conclusion

In conclusion, this study employs a Difference-in-Differences (DID) model, complemented by cross-period and dynamic DID models, to evaluate the impact of renowned coaches such as Lang Ping and Sandro Damilano on the Olympic achievements of various nations. The findings suggest that the appointment of these distinguished coaches correlates with a significant increase in scoring rates, indicating their substantial influence on enhancing the medal tally. The analysis reveals both immediate and long-term effects, with an initial surge in performance that tends to stabilize over time, highlighting the enduring benefits of their coaching legacies. The results underscore the strategic importance of coaching excellence in catalyzing both short-term gains and long-term advancements in a nation's Olympic standing. While the model provides valuable insights, it also acknowledges the need for further research to address potential limitations such as sample size and the generalizability of the model's assumptions. Overall, this study offers a robust quantitative assessment of the "great coach effect," demonstrating the pivotal role of coaching in elevating a country's competitive edge in the global sporting arena.

6 Problem 4: Other Influencing Factors–The Impact of Changes in the Number of Sports Events on Competition Outcomes

In our study, we meticulously extracted data pertaining to the quantity of various sports events in the Olympic Games and conducted a trend analysis to discern patterns and fluctuations.(Figure 14) By extracting the number of medals garnered by nations in each sport, we incorporated this metric as a predictive factor in our forecasting model. (Figure 15)This inclusion aimed to enhance the model's explanatory power and predictive accuracy, providing a more nuanced understanding of the multifaceted determinants of Olympic success.

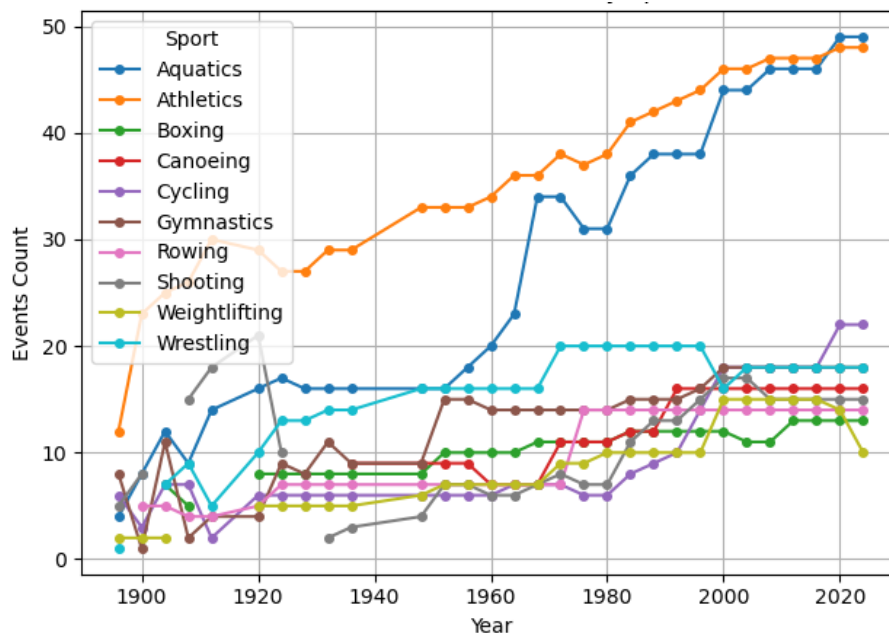


Figure 14: Events Count Over Time by Sport

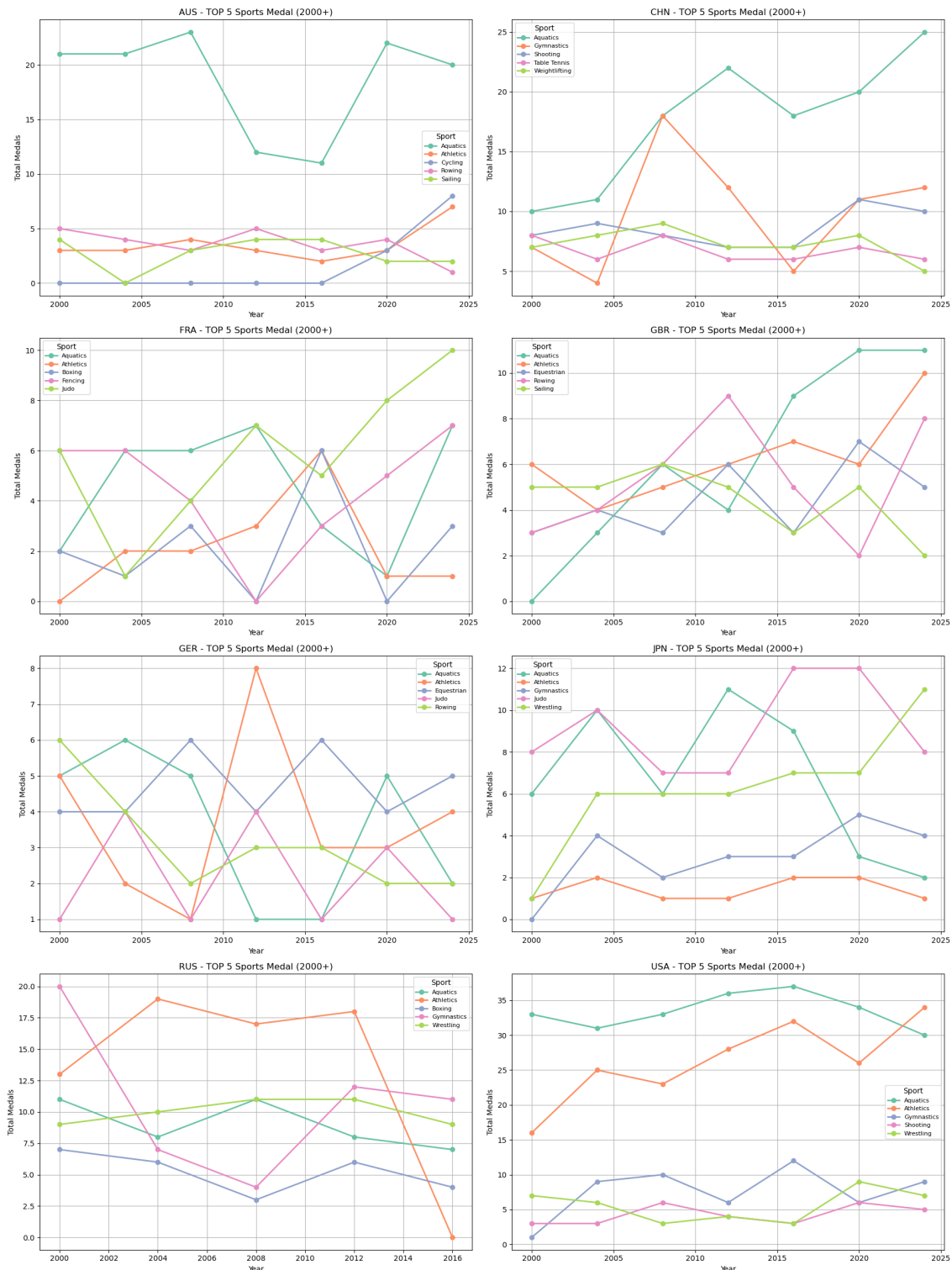


Figure 15: Events Count Over Time by Sport

By incorporating data on the number and type of sports events, the model can account for variations in competition intensity and diversity across different Olympic Games. This allows for

a more nuanced understanding of how changes in the sports program or the introduction of new sports affects a nation's medal haul.

The model can identify which sports are the most influential in driving a nation's overall medal count. By analyzing trends in medal distribution across sports, the model can pinpoint sports where coaching interventions are most effective, thus guiding resource allocation and training focus.

Understanding the impact of different sports on medal counts enables sports federations and Olympic committees to strategically allocate resources. For instance, if certain sports consistently yield a high number of medals, prioritizing investment in these areas could be more beneficial.

The analysis can also reveal sports that have shown a growing trend in medal count over successive Olympic Games. For sports where a nation has been historically strong, maintaining or enhancing coaching expertise in these areas could be crucial for sustaining or improving medal performance.

7 Strength and Weakness

7.1 Strength

- **Comprehensive analysis:** The model takes into account multiple factors that affect the number of MEDALS, such as the number of athletes, the diversity of sports, the number of historical MEDALS, etc., to provide a comprehensive analysis framework.
- **Adaptability:** Through the hierarchical prediction method, the model can forecast according to the characteristics of countries at different sports levels, improving the accuracy and adaptability of the prediction.
- **Data-driven:** The model is trained and validated based on historical data to ensure the reliability and scientificity of the predicted results.

7.2 Weakness and Further Discussion

- The model posits that leading sports nations will retain their competitive advantage, capturing 50% of the total medals. This assumption is overly deterministic and may underestimate the potential of developing countries to rise in Olympic competition.
- The model's reliance on data from Olympic Games held after 2000 may fail to capture long-term trends in the time series, which could affect the accuracy of the predictions.

References

- [1] CHEN T. Q., GUESTRIN C. XGBoost: A Scalable Tree Boosting System[C]//Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM: 785–794.
- [2] https://en.wikipedia.org/wiki/2028_Summer_Olympics

- [3] <https://la28.org/en/games-plan/olympics.html>
- [4] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2), 187–202.
- [5] *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

8 Memorandum for The Coaches

To: Olympic Committee

From: Team #2509743

Date: January 28, 2025

Our team has conducted an extensive analysis of the factors influencing Olympic medal counts and has developed predictive models to forecast the outcomes of the 2028 Los Angeles Summer Olympics. We believe that our findings can offer valuable insights and strategic guidance for National Olympic Committees (NOCs) around the world.

Our predictive models forecast the medal standings for the 2028 Olympics, identifying countries likely to improve or regress compared to 2024. This analysis can help NOCs set realistic goals and allocate resources effectively. For instance, our models suggest that leading sports nations such as the United States and China may experience slight declines in their medal counts, while countries like France, Great Britain, and Germany are expected to see increases. This information can guide NOCs in focusing on areas where they can maximize their medal potential.

We have analyzed the impact of renowned coaches on team performance using the Difference-in-Differences (DID) model. Our findings indicate that hiring distinguished coaches can significantly enhance a country's medal tally. For instance, the appointment of coaches like Lang Ping and Sandro Damilano has been correlated with a marked increase in scoring rates and medal wins. NOCs can use this information to strategically invest in coaching talent to improve their Olympic performance.

In conclusion, our study offers a comprehensive analysis of the factors influencing Olympic medal performance and provides actionable recommendations for NOCs. By leveraging our predictive models and insights, NOCs can make informed decisions to enhance their Olympic performance and achieve better results in the 2028 Los Angeles Summer Olympics. We hope that our analysis will be of great value to your organization and look forward to any further discussion or collaboration.

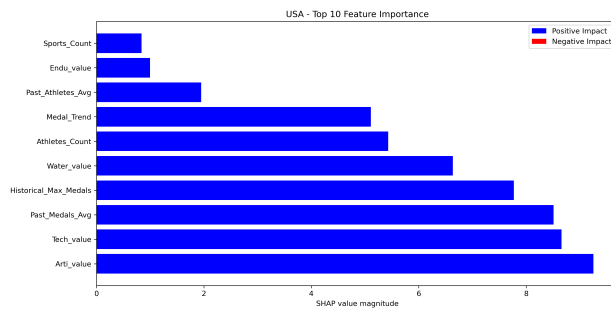
Sincerely yours,

Team #2509743

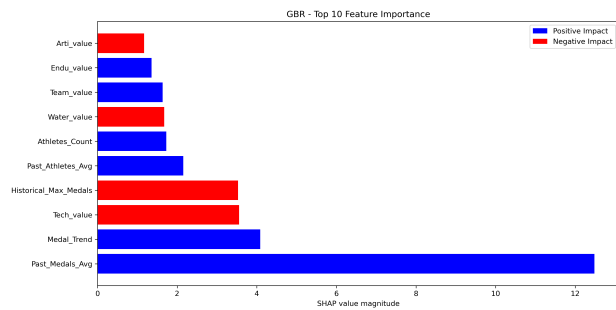
Appendices

Appendix A Feature Importance Plot of different countries

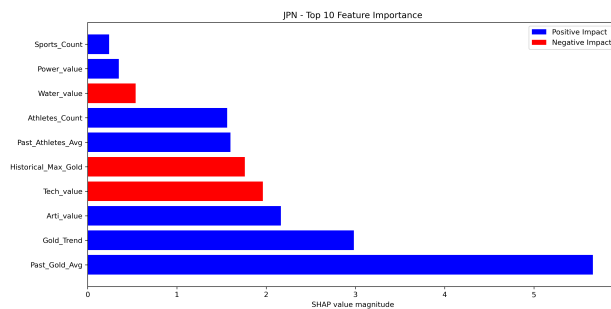
Figure 16: Feature Importance Plot of different countries



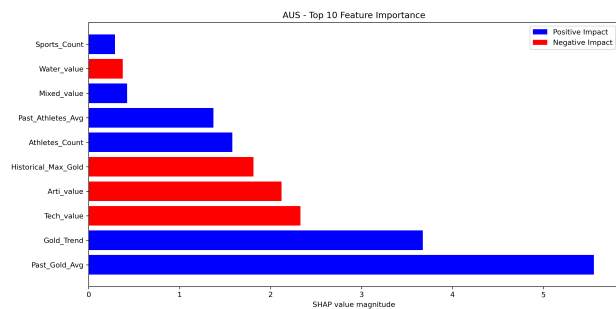
(a) USA Medal model feature importance plot



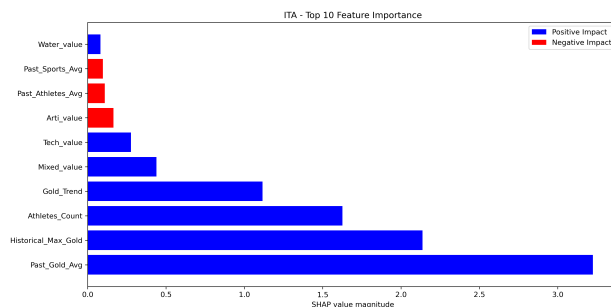
(b) GBR Medal model feature importance plot



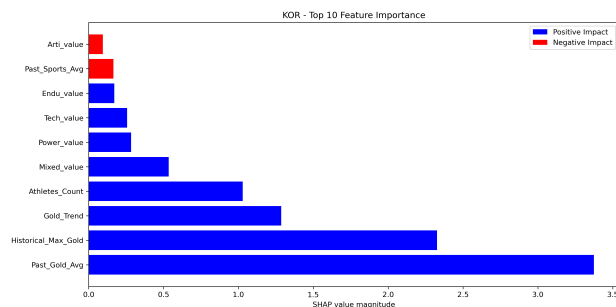
(c) JPN Medal model feature importance plot



(d) AUS Medal model feature importance plot



(e) ITA Medal model feature importance plot



(f) KOR Medal model feature importance plot

Appendix B Features extracted for XGBoost Model

Feature	Feature Meaning
NOC	National Olympic Committee
Year	Olympic year
Athletes_Count	Number of athletes participating in that Games
Sports_Count	Number of sports participated in that Olympics
Past_Athletes_Avg	The average number of athletes participating in the last two Olympic Games
Past_Sports_Avg	The average number of sports participated in the last two Olympic Games
Past_Medals_Avg	The average number of medals won in the last two Olympic Games
Historical_Max_Medals	The most medals won by the country in the Olympics
Medal_Trend	The trend of medal number: First difference of medal count
Is_host	Category variable: Identifies whether the country is the host
Tech_value	This is an indicator that represents the overall ability of the athletes in the technical sports of the country, combined with the influence of the weight of the sports in the Olympic Games
Power_value	This is an indicator that represents the overall ability of the athletes in the power sports of the country, combined with the influence of the weight of the sports in the Olympic Games
Arti_value	This is an indicator that represents the overall ability of the athletes in the artistic sports of the country, combined with the influence of the weight of the sports in the Olympic Games
Water_value	This is an indicator that represents the overall ability of the athletes in the water sports of the country, combined with the influence of the weight of the sports in the Olympic Games
Team_value	This is an indicator that represents the overall ability of the athletes in the team sports of the country, combined with the influence of the weight of the sports in the Olympic Games
Endu_value	This is an indicator that represents the overall ability of the athletes in the endurance sports of the country, combined with the influence of the weight of the sports in the Olympic Games
Mixed_value	This is an indicator that represents the overall ability of the athletes in other mixed sports of the country, combined with the influence of the weight of the sports in the Olympic Games

Table 6: Features for XGBoost Model

Report on Use of AI

The team used AI in the writing of the paper to polish the inauthentic expressions, and all AI-generated polished language was strictly reviewed by the team before being applied to the corresponding part of the paper. At the same time, all AI-generated texts in the paper were generated under strict prompt words and sufficient content, and the team did not abuse AI.

Question: Please provide some cases where renowned coaches have significantly enhanced the Olympic performance of a certain country after taking up coaching positions abroad.

Answer(chatGPT-4o): Cases of coaches who have coached internationally being successful at the Olympic Games are not common. However, some renowned coaches have indeed significantly improved the Olympic performance of certain countries through international coaching. Here are some famous cases:

1. Michael Schuller

National Team: China (Men's Basketball Team)

Achievement: Michael Schuller is an American basketball coach. In the mid-1990s, he served as the head coach of the Chinese men's basketball team. His arrival played a crucial role in enhancing the overall level of the Chinese men's basketball team. Particularly, under his guidance, the technical level and tactical system of the Chinese team improved, and they achieved better international results. Although the Chinese team did not directly win Olympic medals during his coaching tenure, his coaching helped the Chinese basketball team gain stronger competitiveness in the Asian region.

2. Mark Hodgkinson

National Team: Singapore (Men's Ice Hockey Team)

Achievement: British coach Mark Hodgkinson once coached the men's ice hockey team of Singapore. He introduced the European ice hockey training system to Singapore and improved the training methods for players, significantly enhancing the ice hockey level of the Singapore team. Hodgkinson's arrival made the performance of the Singapore team more outstanding in Asian competitions, helping them establish a stronger foothold in international events. Although the Singapore team did not directly win Olympic medals, they made significant breakthroughs in the Asian ice hockey field.

3. Shabbir Hossain

National Team: Bangladesh (Men's Hockey Team)

Achievement: Shabbir Hossain is a hockey coach who moved from India to Bangladesh. By introducing advanced training concepts and systems from India, he helped Bangladesh's hockey team make significant progress. Under his coaching, Bangladesh's hockey team not only performed well in domestic competitions but also gained more recognition on the international stage, especially in Asian Cup events, where their performance improved significantly.

4. Gonzalo Garcia

National Team: Bahrain (Track and Field)

Achievement: Gonzalo Garcia is an Argentine coach who was hired as the head coach of the Bahrain track and field team in the early 2000s. By introducing more modern training methods and scientific sports physiology, he successfully improved Bahrain's performance in sprint and long-distance events. Garcia's training methods helped Bahrain achieve better results in international track and field events, including improving performance in the World Track and Field Championships.

5. **Floyd Pera**

National Team: Qatar (Track and Field)

Achievement: Floyd Pera is an American coach who became the head coach of Qatar's national track and field team in the early 2010s. By introducing some new training concepts and methods, he achieved outstanding results in sprint events for Qatar. Under Pera's leadership, Qatar's track and field performance has been continuously improving, especially in sprint and hurdle events, where Qatar gradually achieved better results in international competitions.

6. **Mihai Bilau**

National Team: United Arab Emirates (Weightlifting)

Achievement: Mihai Bilau was the head coach of the United Arab Emirates weightlifting team, helping the country improve its weightlifting level. Bilau's training methods and concepts focused on enhancing strength and explosive power. His training system enabled the weightlifting athletes of the United Arab Emirates to make significant progress in Asian and world-class competitions. Although the United Arab Emirates did not win Olympic medals, their performance in Asian weightlifting competitions improved significantly.

7. **Joseph Mutua**

National Team: Brazil (Basketball)

Achievement: Joseph Mutua is a basketball coach from Kenya who served as the assistant coach of the Brazilian national basketball team in the early 2010s. By introducing more tactical analysis and systematic training methods, he helped Brazil improve its overall performance. With the assistance of Mutu, Brazil's performance in international competitions became more stable. Although it did not directly affect Brazil's Olympic results, Brazil's performance in the World Basketball Championship was improved.

Summary: These cases of cross-national coaching demonstrate how foreign coaches can bring their knowledge and training systems into other countries, thereby enhancing the performance of these countries in international events such as the Olympics. Although some cases may not have directly led to Olympic medals, they did indeed enhance the sports level at the national level and made some countries stronger competitors in specific sports.