

# Project Olympian: Unveiling the Impact of a Nations' Stability on Olympic Success



Exploring the Relationship Between Economic, Political, and Health Stability and Olympic Performance

By Dylan Devens, Addie Demhlow, and Alvin Son

Special thanks to Sarah Thong



The Global Unifier: Olympics  
- The Olympics unite nations and showcase the world's best athletes.  
- Winning brings pride to athletes, teams, and entire countries.  
- But how do a country's economic, political, and health conditions influence Olympic success?

## Data Sources

Overview:  
The project utilizes various data sources to predict medal counts for each nation at the 2016 Rio Olympics. These include historical medal counts from the 1956 Melbourne Olympics, the 2012 London Olympics, and the 2016 Rio Olympics. Additionally, the project uses data from the World Bank, the International Monetary Fund (IMF), and the United Nations Development Programme (UNDP) to analyze economic stability, political stability, and health stability.

## Model Selection

Overview:  
The project aims to predict the Olympic medal counts of individual countries using machine learning models. The two models selected are a Random Forest and a Neural Network.

Objectives:  
1. To compare the performance of a Random Forest and a Neural Network in predicting medal counts.  
2. To determine which factors have the greatest influence on a country's Olympic performance.

## Introduction to Tableau Visualizations

Tableau is a powerful tool for visual data analysis, allowing us to uncover patterns and trends that might be hidden in raw data. This section presents key insights from the data visualization, highlighting the factors influencing Olympic performance.

# The Global Unifier: Olympics

- The Olympics unite nations and showcase the world's best athletes.
- Winning brings pride to athletes, teams, and entire countries.
- But how do a country's economic, political, and health conditions influence Olympic success?



# What Drives Olympic Success?

- We explore whether a nation's stability impacts its Olympic performance.
- Does hardship drive athletes to excel or does it hinder their success?
- Our project seeks to uncover the hidden factors behind Olympic outcomes.



# Key Questions Explored

1. How do Economic Conditions Influence Olympic Performance?
2. What Role Does Political Stability Play in Medal Consistency?
3. How Do Disasters and Other Natural Factors Affect Athletes' Training and Performance?



# Data Sources

## Overview:

- Data should describe the athletes and country metrics for each year and country that competed in the olympics.

## Data Sources:

- *Olympic Historical Data:*
  - Source: Kaggle – originally scraped from [www.sports-reference.com](http://www.sports-reference.com) in May 2018
  - Content: Medal counts, athlete participation, country representation over the years, from 1896 – 2016.
- *Socio-Economic Indicators / Political and Environmental Data:*
  - Source: World Bank Group – Large database that aggregates from reputable sources worldwide
  - Content: GDP, GDP per capita, population, education levels, health expenditure, etc.
  - Source: Organisation for Economic Co-operation and Development –
  - Content: Well-being, all-cause mortality, air pollution

# Data Cleaning Process

- group athlete data by country and year
- aggregate individual values for the team
- minimize naming discrepancies

- reshape raw data
- group country statistics by country and year

- match countries and country codes from multiple sources through name changes and wars

(271116, 15)																	
ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal			
0 1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	Nan			
1 2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	Nan			



(5586, 68)																	
Country Name	Country Code	Series Name	Series Code	1960	1961	1962	1963	1964	1965	...	2014	2015	2016	2017	2018	2019	
0 Afghanistan	AFG	Armed forces personnel (% of total labor force)	MS.MIL.TOTL.TF.ZS	Nan	Nan	Nan	Nan	Nan	Nan	...	3.727055	3.700732	3.623539	3.490159	2.952594	3.015079	1.81



NOC	Year	Country Name	Total Athletes	Medal Count	Gold Count	Silver Count	Bronze Count	Average Age (M)	Average Height (M)	...	Total Disaster Aid Contribution (adjusted US\$)	Total Disaster Related Deaths	Total Disaster Related Injured	Total Disaster Damage (adjusted US\$)	Total Dangerous Emissions (tonnes)
1978 AZE 2000 Azerbaijan			31.0	1.0	2.0	0.0	1.0	24.6	173.375	...	341.0	31.0	600.0	17695.0	24.175382

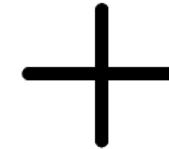
# Data Integration

## Merging Datasets:

- *Process:*
  - Merged multiple datasets using common keys such as country codes and years.
  - Combined data across different sources to create a comprehensive dataset covering all relevant indicators.
- *Challenges:*
  - Different time spans or reporting periods across datasets.
  - Athletes competing without a country
  - Alignment of temporal data (e.g., annual economic indicators with quadrennial Olympic data).
  - Naming conventions after wars
- *Final Dataset:*
  - Structure: 43 features to use for modelling
  - Rows: Each row represents a country-year pair with corresponding socio-economic, political, and environmental indicators.
  - Columns: Features such as GDP, population, political stability, health expenditure, and the target variable (medal count).

(5586, 68)

Country Name	Country Code	Series Name	Series Code	1960	1961	1962	1963	1964	1965	...	2014	2015	2016	2017	2018	2019
0 Afghanistan	AFG	Armed forces personnel (% of total labor force)	MS.MIL.TOTL.TF.ZS	NaN	NaN	NaN	NaN	NaN	NaN	...	3.727055	3.700732	3.623539	3.490159	2.952594	3.015079 1.81



(271116, 15)

ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
0 1	A Dijiang	M	24.0	180.0	80.0	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's Basketball	NaN
1 2	A Lamusi	M	23.0	170.0	60.0	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Lightweight	NaN



```
print(combined.shape)
```

(5191, 46)

# Challenges and Limitations

## Data Challenges:

- *Missing Data:*
  - Some indicators were not consistently reported across all countries or years.
  - Country data began well after the start of the olympics
- *Data Integration:*
  - Merging datasets from different sources and timeframes posed challenges in maintaining consistency.

## Modeling Limitations:

- Some socio-political factors might not be fully captured in the available data.
- Potential for overfitting given the small sample size of countries with high medal counts.

# Model Selection

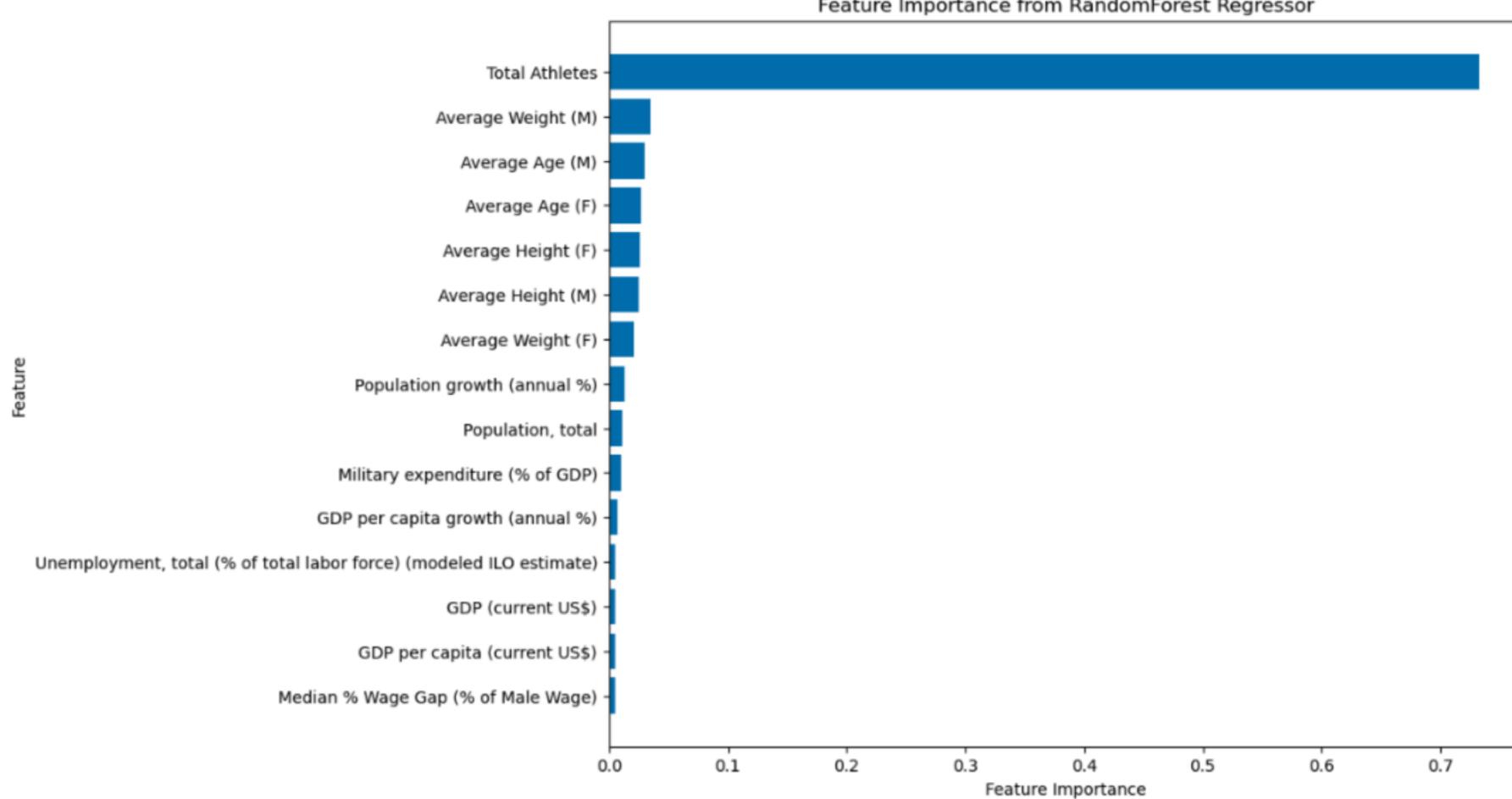
Overview:

- This project aims to predict the Olympic medal counts of various countries using machine learning models.

Objectives:

- To compare the performance of a Random Forest and a Neural Network in predicting medal counts.
- To identify the most important features that influence a country's Olympic performance.

# Feature Selection Using Random Forest



Initial Model:

- Trained a Random Forest model to identify important features.

Feature Importance:

- Identified key features
- Removed less important features to improve model efficiency.

# Model Evaluation: Random Forest Model

## Description:

- Trained the model using the selected features.
- Evaluated using MAE (Mean Absolute Error) and R<sup>2</sup> score.

## Results:

- MAE: 2.94      Selected Features Random Forest – MAE: 2.94, R2: 0.70
- R<sup>2</sup>: 0.70
- Interpretation: The model explains 70% of the variance in medal counts with an average prediction error of about 2.94 medals.

## Flaws:

- The R<sup>2</sup> of 0.70 reflects the complex, multifaceted nature of predicting Olympic medal counts, with some key factors missing from the dataset.
- Data limitations, potential overreliance on key features, and model constraints suggest the need for more data and advanced modeling to improve accuracy.

# Model Evaluation: Neural Network

## Description:

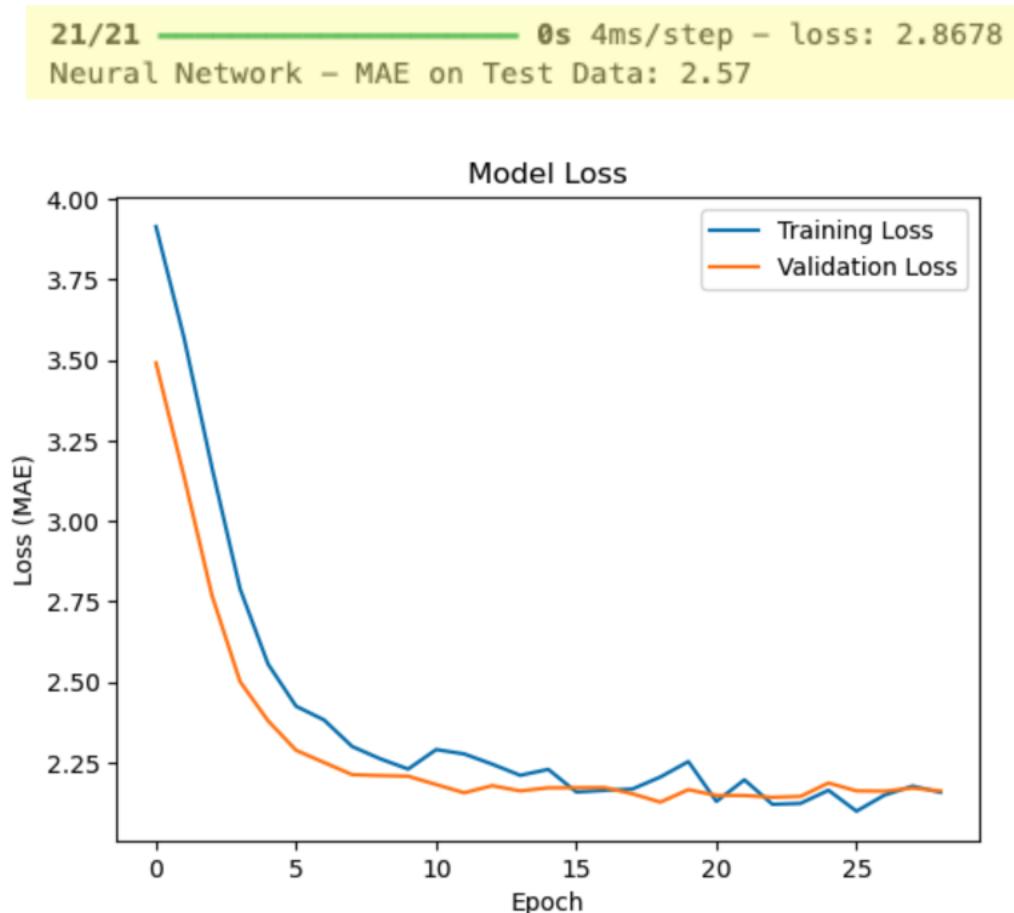
- Used all features to train a Neural Network with two hidden layers and dropout regularization.
- Applied early stopping to prevent overfitting.

## Results:

- MAE: 2.57

## Interpretation:

- The model achieves slightly better prediction accuracy compared to the Random Forest.



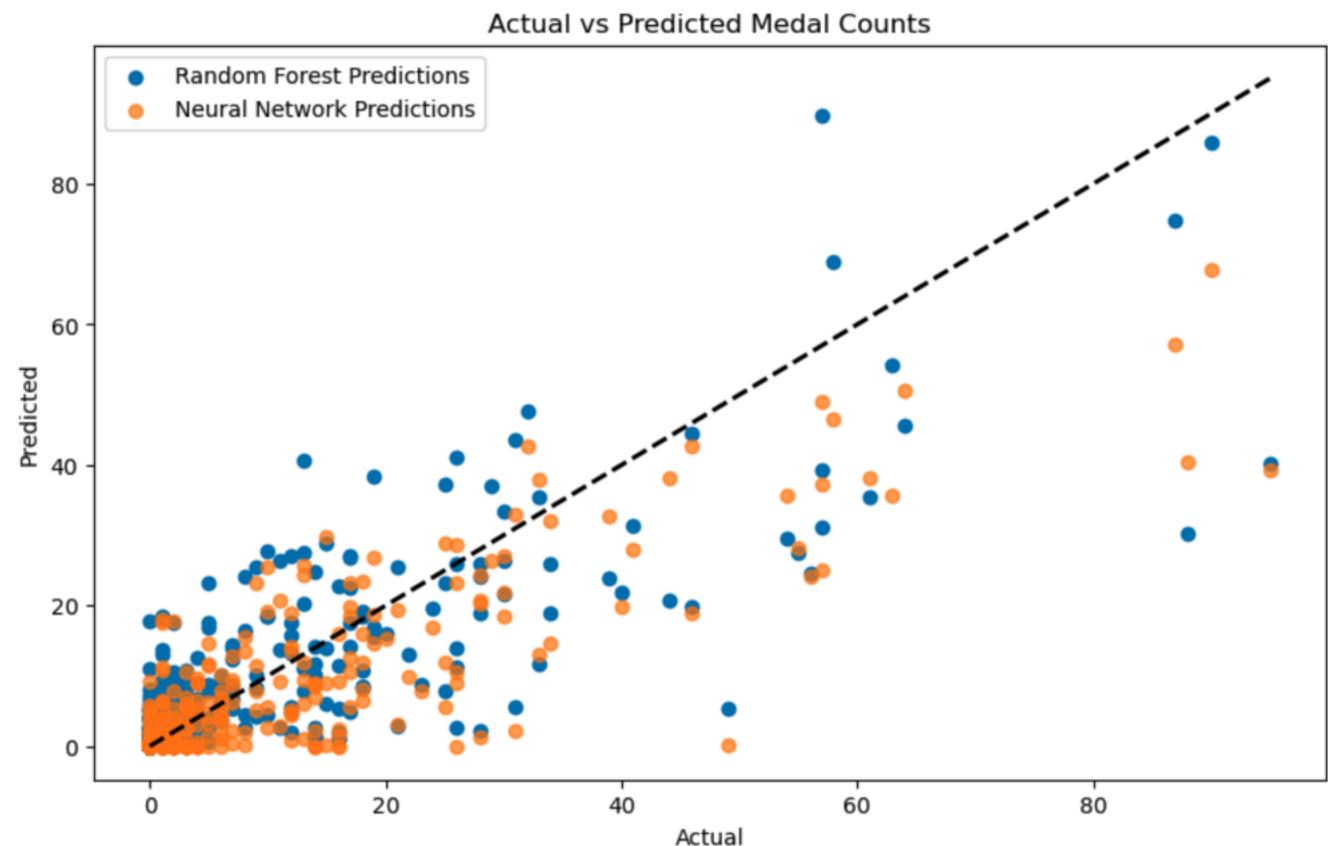
# Comparing the Models

## Scatter Plot:

- Actual vs. Predicted Medal Counts plot comparing both models.

## Key Points:

- Both models perform well on lower medal counts.
- The Neural Network slightly outperforms the Random Forest in prediction accuracy (lower MAE).
- Both models struggle with predicting higher medal counts accurately.



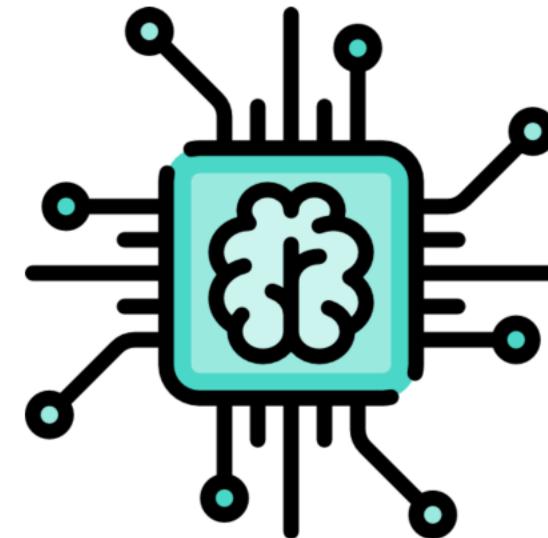
# Conclusion and Next Steps

## Conclusion:

- The models can reasonably predict medal counts with some accuracy, especially for lower counts.
- The feature selection process highlighted important factors related to athlete characteristics.

## Next Steps:

- Further tuning, adding more data, or exploring ensemble methods might improve predictions.
- Potential exploration of other machine learning models like XGBoost for better accuracy.



# Introduction to Tableau Visualizations

- Tableau is a powerful tool for visual data analysis, allowing us to uncover patterns and trends that might be hidden in the raw data.
- This section presents key insights from the data visualized in Tableau, focusing on the factors influencing Olympic performance.



[https://public.tableau.com/app/profile/addie.dehmlow/viz/Project\\_olympian/Story1?publish=yes](https://public.tableau.com/app/profile/addie.dehmlow/viz/Project_olympian/Story1?publish=yes)

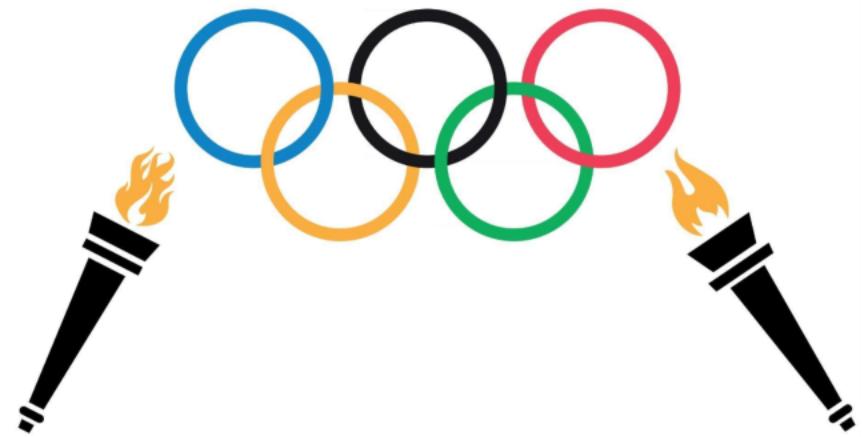
# Tableau Insights

## Summary of Findings:

- Largest factor: Total athletes
- Population played a key role
- Wage gap playing a factor
- Disasters had little to no affect
- Military spending drop

## Implications:

- Seeing is believing
- A change in policies can affect performance



# Project Olympian: Unveiling the Impact of a Nations' Stability on Olympic Success



Exploring the Relationship Between Economic, Political, and Health Stability and Olympic Performance

By Dylan Devens, Addie Demhlow, and Alvin Son

Special thanks to Sarah Thong



The Global Unifier: Olympics  
- The Olympics unite nations and showcase the world's best athletes.  
- Winning brings pride to athletes, teams, and entire countries.  
- But how do a country's economic, political, and health conditions influence Olympic success?

## Data Sources

Overview:  
The project utilizes various data sources to predict medal counts for each nation at the 2020 Tokyo Olympics. These include historical medal counts from the 1956-2016 Olympic games, as well as current data on economic stability, political stability, and health stability.



## Model Selection

Overview:  
The project aims to predict the Olympic medal counts of individual countries using machine learning models.

Objectives:  
1. To compare the performance of a Random Forest and a Neural Network in predicting medal counts.  
2. To determine which factors have the greatest influence on a country's Olympic performance.



## Introduction to Tableau Visualizations

Tableau is a powerful tool for visual data analysis, allowing us to uncover patterns and trends that might be hidden in raw data.

This section presents key insights from the data visualization, highlighting the factors influencing Olympic performance.

