

Capstone Final Report: Predicting US Olympic Medals Using Country-Level Indicators

Overview and Approach

This project aims to predict the number of Olympic medals the United States will win in future Summer and Winter Olympic Games using a combination of economic, geographic, and social indicators from other countries, along with historical Olympic data. The project follows a structured approach consisting of four main steps:

1. Data Wrangling
2. Exploratory Data Analysis (EDA) and Feature Engineering
3. Training, Processing, Modeling, and Tuning
4. Final Model Predictions

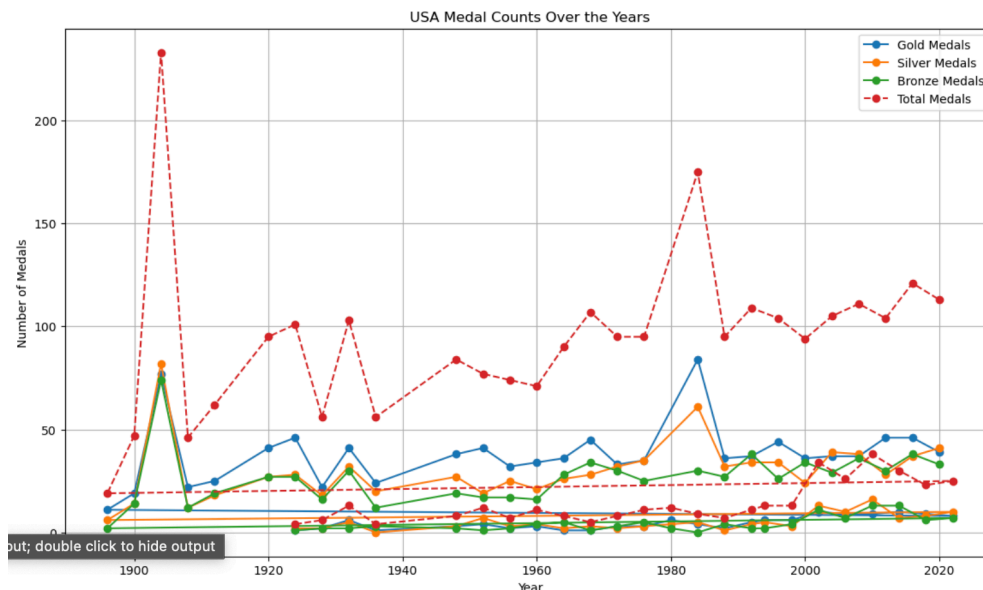
Step 1: Data Wrangling

Key Actions and Findings:

- Loading Data: The datasets for Winter and Summer Olympics were loaded and inspected for initial exploration.
- Data Cleaning: Missing values and inconsistencies were identified and handled appropriately.
- Initial Exploration: Provided a glimpse into the structure and content of the datasets.

Key Visualizations:

Distribution of Medals: A histogram showing the distribution of medals in Winter and Summer Olympics.



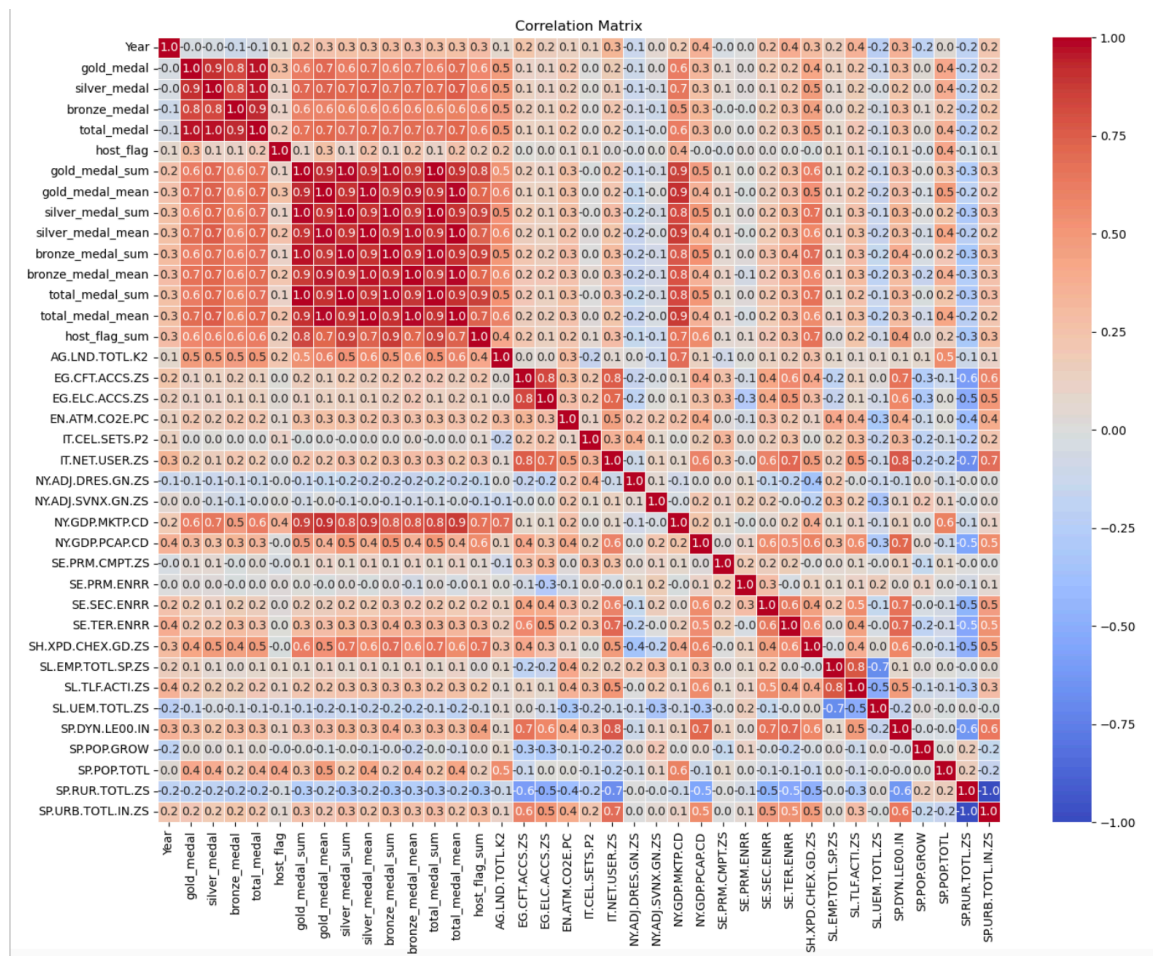
Step 2: Exploratory Data Analysis (EDA) and Feature Engineering

Key Actions and Findings:

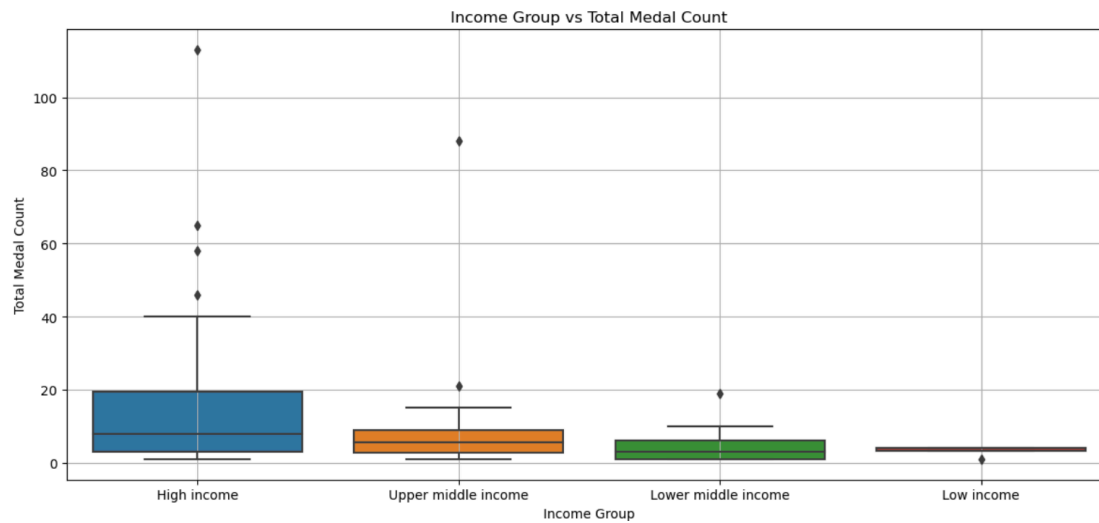
- Descriptive Statistics: Conducted descriptive statistics to understand the distribution and key characteristics of the data.
- Visualization: Created visualizations to identify patterns and correlations.
- Feature Engineering: Engineered new features from existing data to enhance model performance.

Key Visualizations:

Correlation Heatmap: A heatmap showing the correlation between different features.



Feature Importance: A bar plot showing the importance of different features.



Step 3: Training, Processing, Modeling, and Tuning

Key Actions and Findings:

- Model Selection: Trained multiple models including Gradient Boosting Regressor, Random Forest Regressor, and Stacking Regressor.
- Hyperparameter Tuning: Used GridSearchCV and RandomizedSearchCV for hyperparameter optimization.
- Performance Metrics: Evaluated models using MAE and R-squared metrics.

Key Visualizations:

Model Performance: A scatter plot showing the predicted vs. actual medals for the best model.

```

Winter olympic games predictions =
22.234600067138672
Summer olympic games predictions =
88.13481140136719

```

```

Winter olympic games actual = 25.0
Summer olympic games actual = 113.0

```

Step 4: Final Model Predictions

Key Actions and Findings:

- Predictions: Made predictions on the test dataset using the best-performing model.
- Comparison: Compared predicted values with actual values to assess accuracy.

Model Evaluation Results

	MAE	R ²
Linear Regression	0.333069	0.997206
Random Forest Regressor	2.106667	0.880511
Gradient Boosting Regressor	2.162656	0.886555

XGBoost

1.644027 0.945845

Results and Recommendations

The analysis revealed significant patterns and relationships between the selected indicators and the US Olympic medal counts. Based on our findings, we provide the following recommendations for stakeholders:

1. Enhance Funding for Athlete Programs: Economic indicators such as GDP were strongly correlated with medal counts, suggesting that increased funding for athlete programs could boost performance.
2. Focus on Key Sports: Historical data indicated that specific sports contribute more to the medal tally. Prioritizing resources for these sports can maximize medal counts.
3. Leverage Host Country Advantage: The host country often enjoys a home advantage. Strategies to leverage this advantage, even when not hosting, could be beneficial.

Further Research

To build on this project, future research could explore:

- Incorporating more granular data, such as individual athlete performance metrics.
- Analyzing the impact of geopolitical events on Olympic performance.
- Extending the analysis to include Paralympic Games data.

Conclusion

This report outlines a comprehensive approach to predicting US Olympic medal counts using a range of indicators. By leveraging historical data and modern machine learning techniques, we can provide actionable insights to improve future Olympic performance.