

Project Name: Olympic Medal Predictions

Objective: This project was undertaken to demonstrate how historical data can be leveraged to predict the number of medals a country will win at the Olympics.

Method:

Python served as the primary tool for this analysis, involving machine learning techniques. The methodology encompassed the following key steps:

1. **Data Collection:** Data was originally sourced from Kaggle and subsequently consolidated by Dataquest, ensuring data quality and consistency.
2. **Data Cleaning:** Data cleaning followed rigorous data science standards, involving the removal of missing, NaN, and infinite (inf) values. Relevant columns were selected for the study, and decimal values were rounded to enhance accuracy.
3. **Data Analysis:** In-depth data analysis adhering to data science standards was conducted, providing a solid foundation for the subsequent modeling.
4. **Statistical Techniques:** A linear regression algorithm was utilized for machine learning. Multiple training models were developed to predict the number of medals each country would earn. Mean absolute error (MAE) was used as the error metric, with a focus on ensuring that MAE remained well below the standard deviation.

Key Findings:

1. **Age and Olympic Success:** An examination of whether age had an effect on Olympic team success.
2. **Athlete and Event Impact:** Insights into how the number of athletes and events influenced medal count.
3. **Past Performance Influence:** The study found that past performance had a significant impact on the creation of strong predictive models.
4. **Data Balance:** Despite data imbalance, the project demonstrated that it was possible to create effective predictive models.
5. **Predictor Correlations:** Identification of predictors with the highest correlation to the target variable, medals.
6. **Model Refinement:** Continuous model refinement led to a decrease in error frequency and discrepancies, enhancing prediction accuracy.
7. **Reliable Models:** Identification of the most reliable model for overall predictions, as well as models tailored for low-scoring and high-scoring countries.
8. **Refined Prediction:** Refinement of predictive models, while increasing MAE overall, did not necessarily lead to less accurate predictions for the majority of countries.
9. **Upcoming Olympic Predictions:** The project successfully generated predictions for the upcoming Olympic Games.

Actionable Insights:

Although not extensively detailed in the analysis, the ability to predict medal counts using the selected predictors can guide teams and organizations in understanding what they need to do to increase their medal count.

What's Missing:

To further enhance the analysis, potential future iterations could consider the inclusion of:

- Most recent Olympic data
- Specific data on which Olympic medals were won in
- Information on the coaches who contributed to medal achievements
- Information on the categories of sports that were analyzed in the dataset.