# Tools For Data Science: Optimizing Team USA's Gymnastics Roster for the Paris 2024 Olympics

## Executive Summary

Rami Pellumbi*

December 20, 2023

*M.S., Statistics & Data Science

**Introduction** This report outlines a sophisticated system designed to optimize Team USA's roster selection for the Paris 2024 Olympics in Men's and Women's Artistic Gymnastics. Utilizing a blend of statistical modeling and simulation techniques, the system provides a user interface that empowers users to have a data-driven approach to formulating competitive team compositions.

**Methodology** At the heart of the system are linear regression models for each gender-apparatus pair, predicting athletes' execution scores based on difficulty scores and athlete names. These models are grounded in rigorous data analysis, ensuring robust and relevant insights for team selection.

**Data Integrity** A meticulous data exploration and cleansing process was undertaken to ensure the accuracy of the models. This included addressing inconsistencies and duplicate athlete names, resulting in a highly reliable dataset.

**Model Performance** The linear regression models, chosen for their simplicity and interpretability, demonstrate satisfactory performance. This is evidenced by reasonable $R^2$ values across different apparatuses and genders, indicating a good fit with the data.

**Application of the Model** Monte Carlo simulations, based on the model's predictions, are used to forecast the outcomes of various Olympic events. These simulations offer a view of simulated medal counts for allocated teams. The client application is the real strength of the system, offering a user-friendly interface for exploring the results of the simulations. Figures 1 and 2 provide examples of the application's output.

**Analysis of Results** The results from these simulations provide strategic insights into team composition, highlighting which athletes have the greatest potential impact on Team USA's success.

**Implications and Recommendations** The system offers a data-driven approach to team selection, allowing coaches and decision-makers to make informed choices. Beyond gymnastics, this methodology has potential applications in other sports and fields where team composition is crucial.

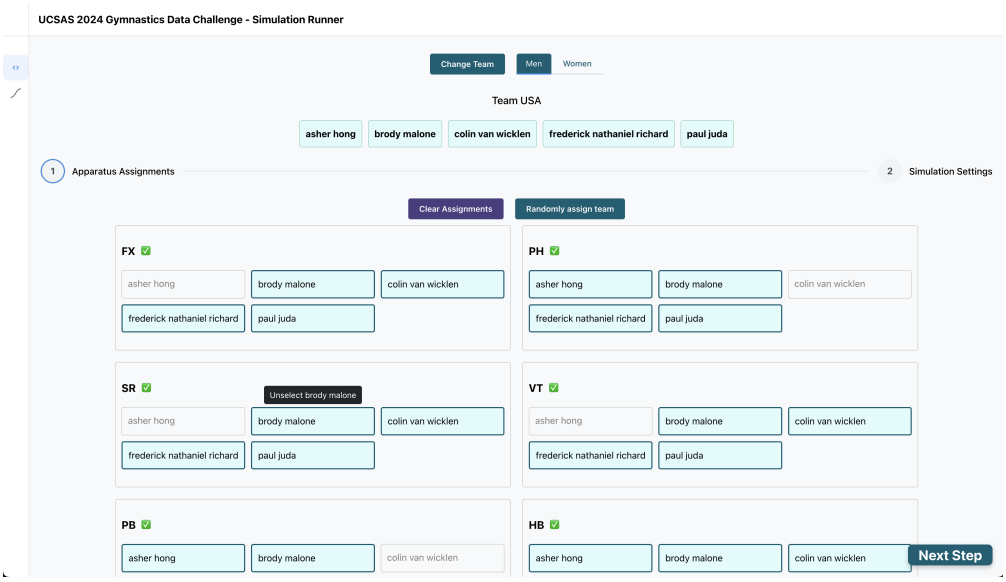


**Figure 1:** Simulation Runner Apparatus Allocations

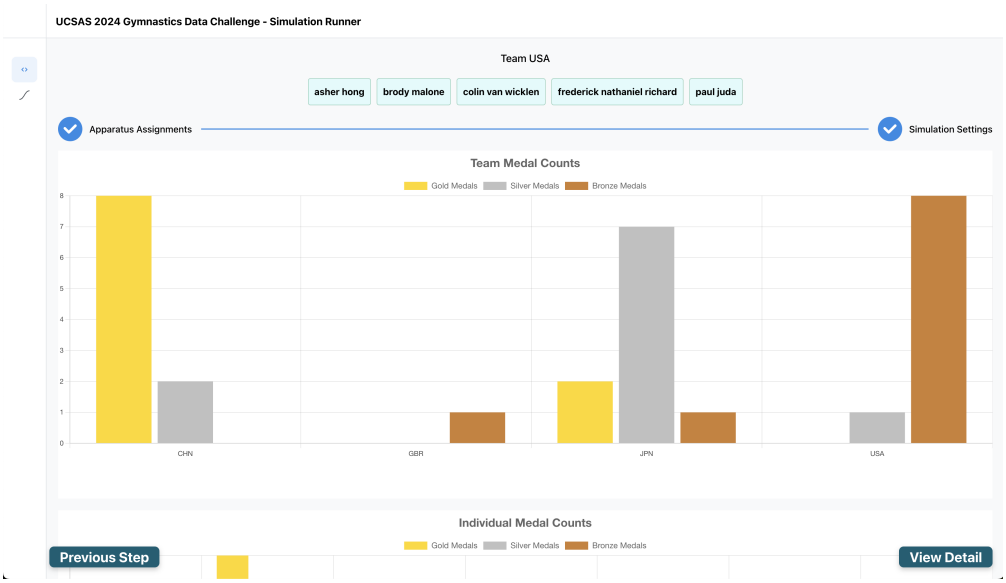


**Figure 2:** Simulation Runner Results