

# Identifying Underpaid and Overpaid MLB Hitters

Terence Moriarty, DS 4002 CS2, May 2nd, 2023



**Scenario:** You are in your second year at UVA, and your Data Science professor has been inspired by Michael Lewis' book *Moneyball: The Art of Winning an Unfair Game*, and is assigning your class a related project. If you are not familiar, the book is about the Oakland Athletics using an analytical approach to find undervalued baseball players to build a great team despite their small budget. You are being tasked with creating a model to evaluate the performance of baseball hitters, then comparing their performance to their salaries to determine the most underpaid and overpaid players.

**Deliverable:** Develop a model that analyzes the statistics of every hitter from the 2022 MLB season, and qualifies their overall level of contribution. The model should also include a comparison of level of contribution with salary. From here, identify the ten most underpaid and ten most overpaid hitters in the MLB for the season. Two datasets are provided, one with statistics for every hitter in the 2022 season, and one with every MLB player's 2022 salary. You will be graded based upon the consideration you put into your model, your choice of model, and your written explanation of the model and its results. You will not be assessed on the statistics and salary of the players chosen.