

Baseball Batting Orders

Optimization Using Simulated Gameplay and Neural Networks

Claude Fried

2021

Problem Statement

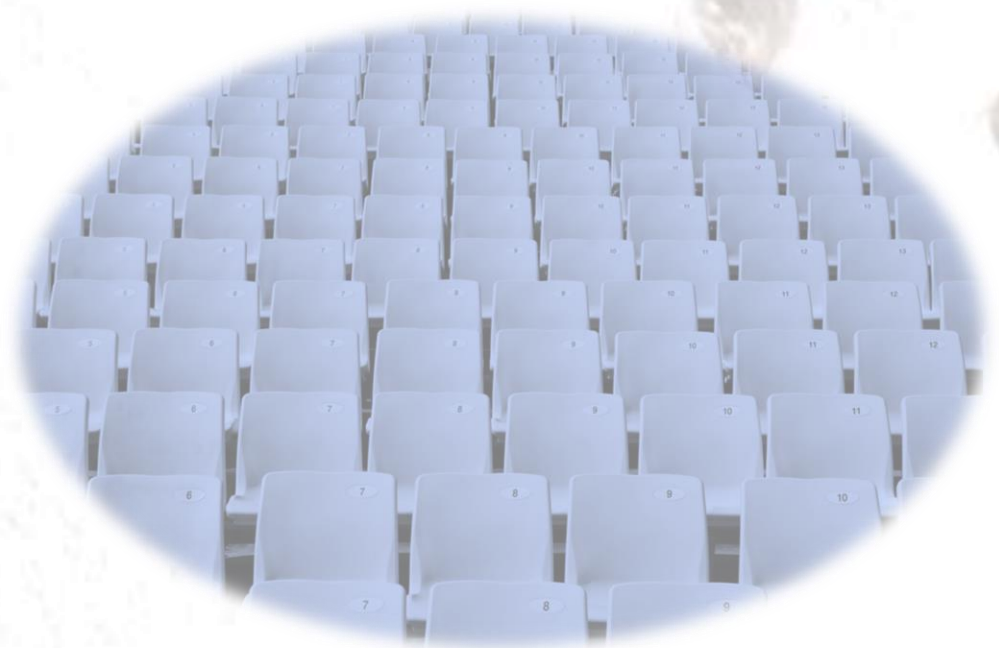
- Can the **outcome of an at-bat** be modeled and predicted?

How can these predictions be used to improve performance?

- Does the **batting order** impact the expected runs scored?
- Can team performance improve by setting an **optimal batting order**?

Business Value

- Finding the best batting order will improve the team's performance.
- By scoring more runs, the team is more likely to win more games.
- Increased team performance will lead to:
 1. Increased ticket sales.
 2. Increased merchandise sales.
 3. Increased publicity.



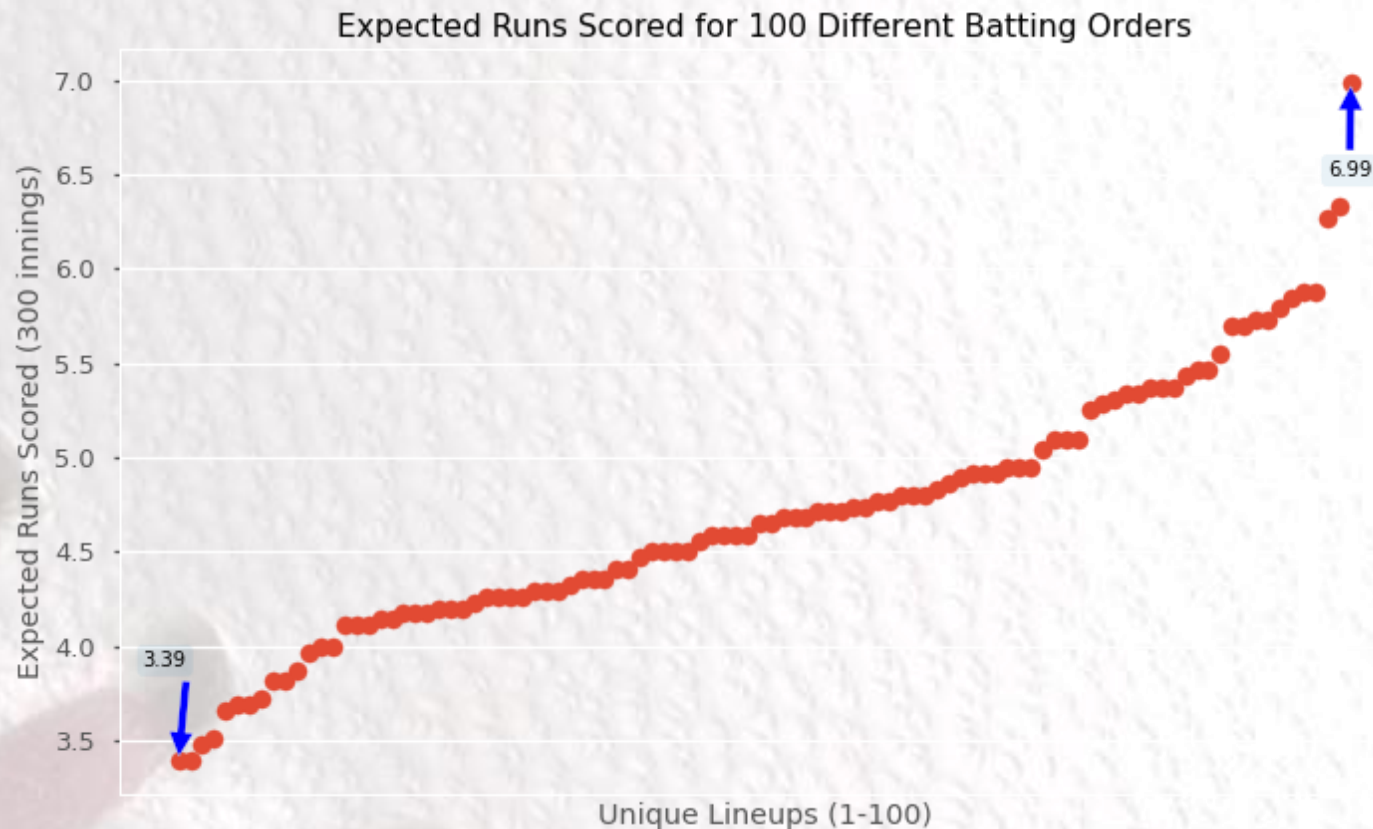
Methodology

1. **Gather** regular-season data for every at-bat since 1950.
2. **Engineer** *Players* to track career stats and Pitcher/Hitter interactions.
3. **Model** to predict the outcome of a given at-bat (probabilities).
4. Create and deploy ***Simulator*** to simulate games and optimize batting orders.

Baseball Recommendations

- Batting Order Matters!

- When running simulations of different lineup-combinations against a given pitcher, the performance of these lineups were drastically different.

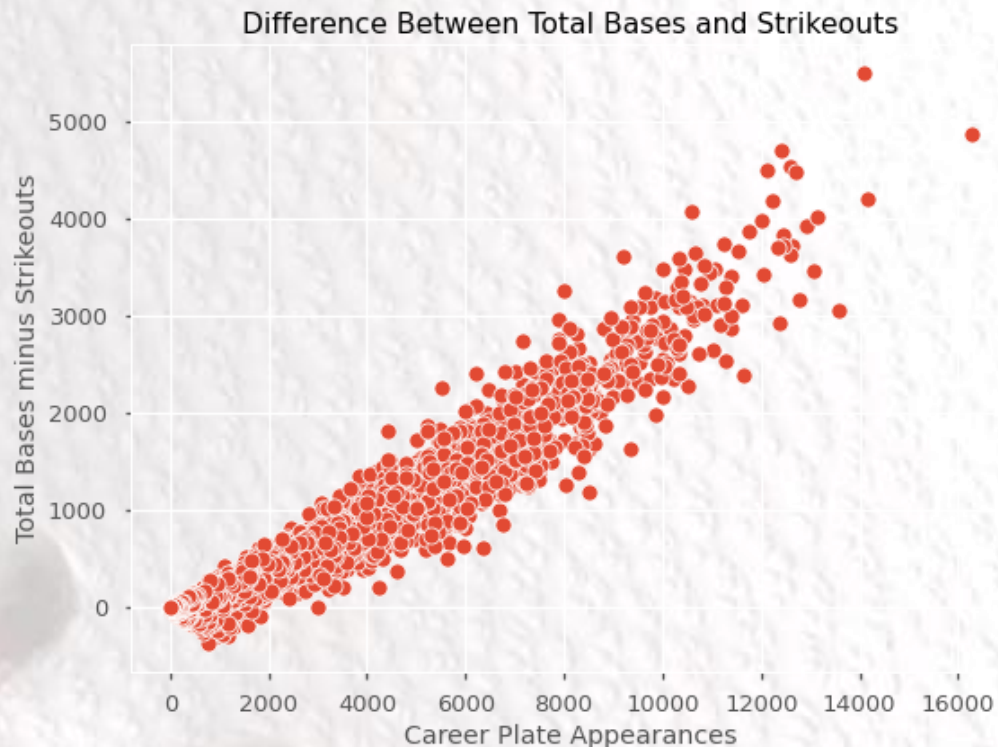


It's clear how impactful the batting order is on expected runs scored.

After 300 innings, some batting orders were outperforming others by over 100%.

Baseball Recommendations - Players

- When evaluating young players, watch *total-bases* and *strikeouts*.
 - There is a strong correlation between number of career-at-bats and difference between total-bases and strikeouts.



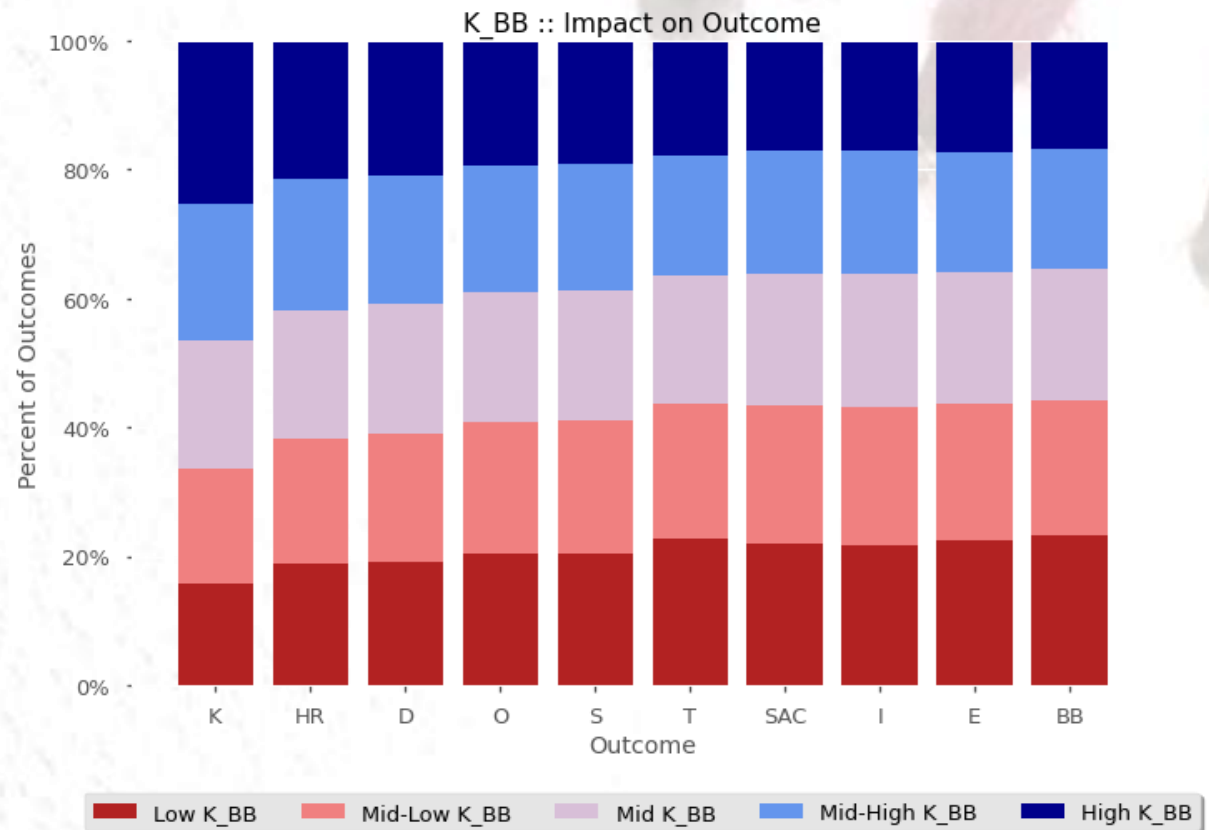
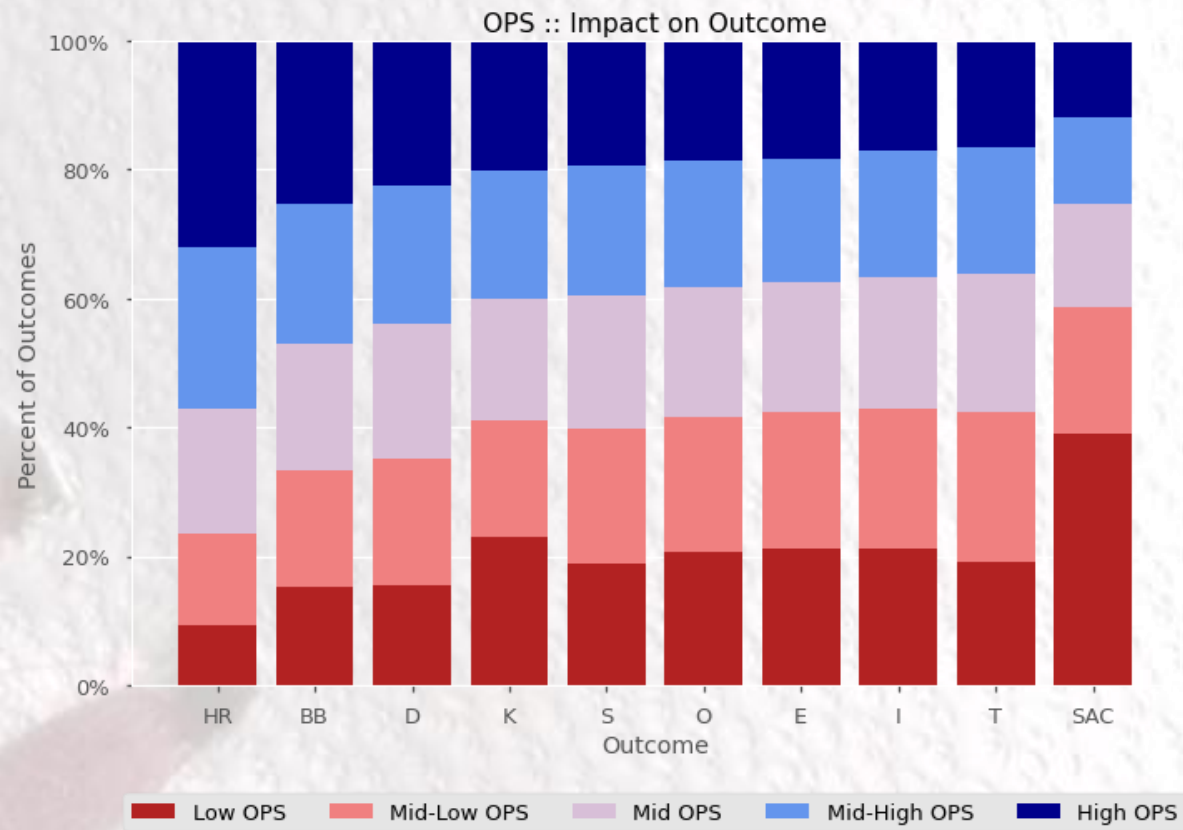
A hitter's ability to hit for a high total-bases and low strikeout total shows longevity in the league.

Player contracts can use this information to offer long-term deals for team-friendly money.

Baseball Recommendations - Outcomes

- **Homeruns** more often hit by players with a high *on-base-plus-slugging (OPS)*.
- **Sacrifices** are most common with players with a lower OPS.

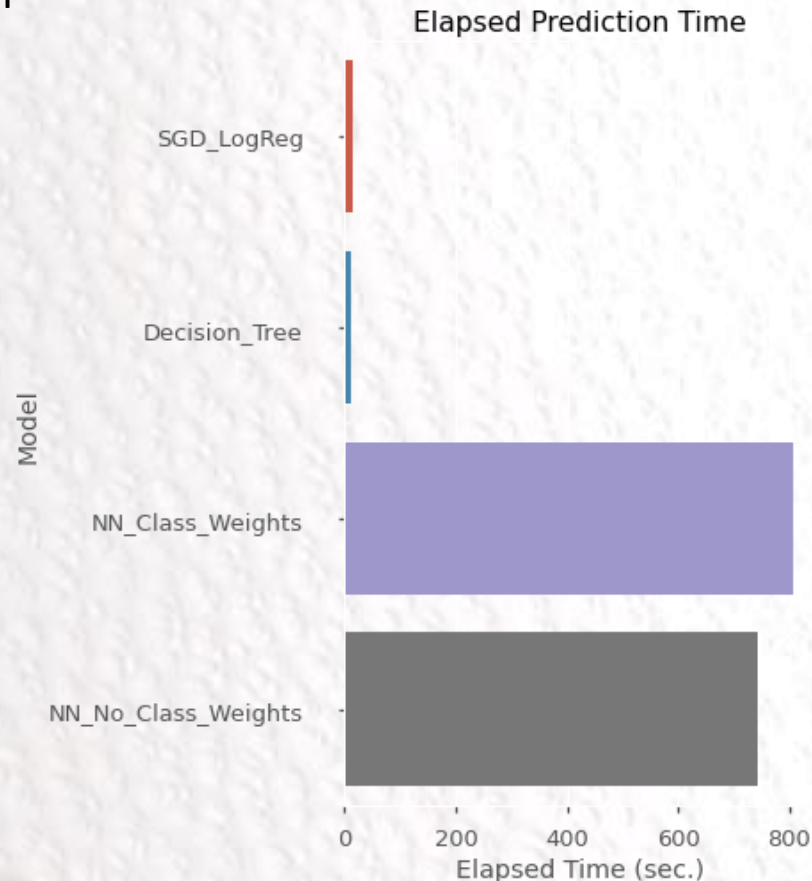
- While **strikeouts** are most common outcome when pitchers have a high *strikeout-to-walk-ratio* (which is good for the pitcher)...
- **Homeruns** and **Doubles** are the second and third most common (which are very good for the hitters)!



Modeling Recommendations

1. Model size and performance should be considered.

- A stronger model with better predictions will be much bigger in size and slower to compute predictions and optimizations.



There is a massive difference in model performance. The fastest models were making predictions several hundred times faster than the slowest models.

Modeling Validation

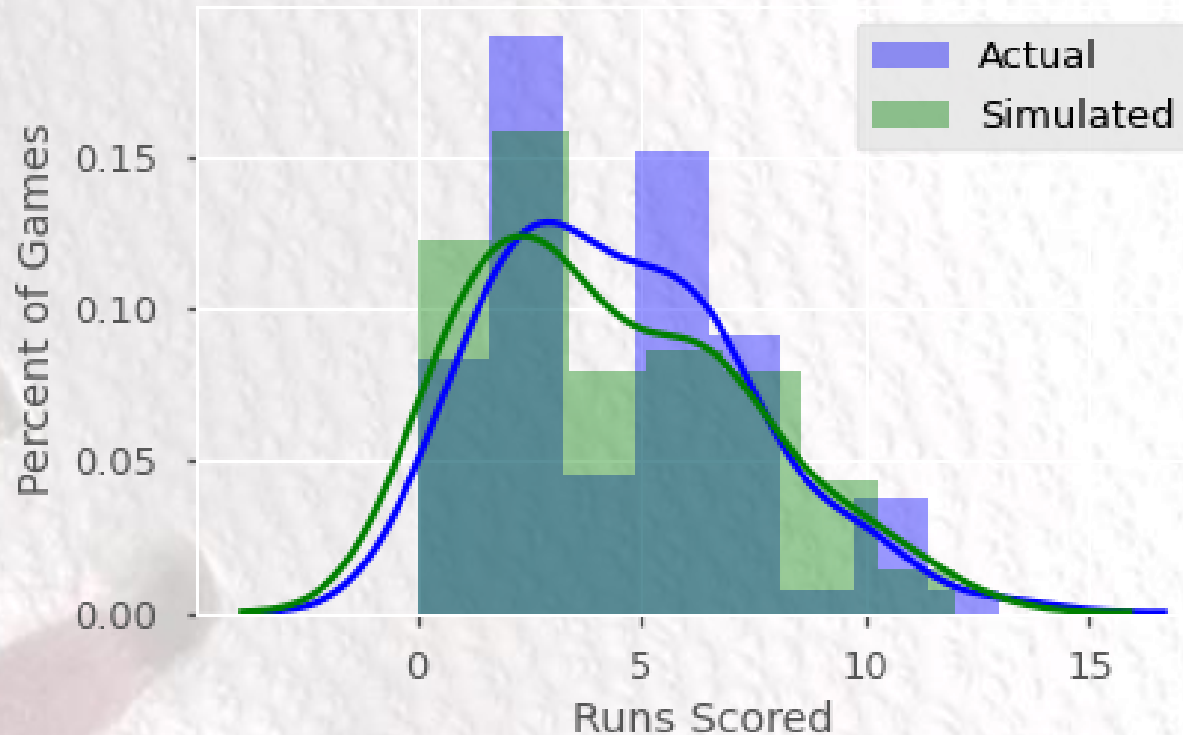
Without having any prior knowledge about “runs scored”, the models’ predictions (compared to real past data) are quite good.

Yankees: Average Runs per Game

Actual:-----4.519

Simulated:-----4.296

NN - 2010 Yankees Runs per Home Game

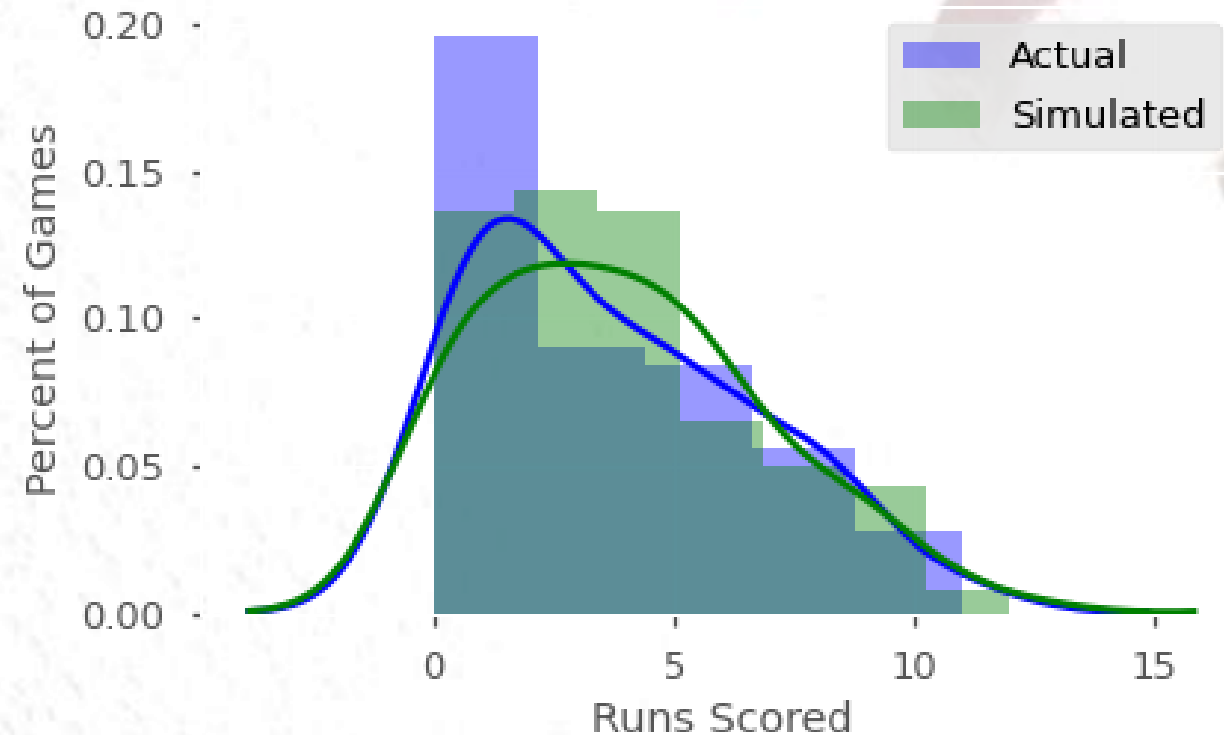


Mariners: Average Runs per Game

Actual:-----3.753

Simulated:-----3.914

NN - 2010 Mariners Runs per Home Game



Future Work / Next Steps

- Experiment with different modeling architectures to try to improve performance.
- Engineer more features:
 - More stats for players.
 - *Hitler-vs.-Pitcher* interaction stats.
- Collect data and do research on the minor league system to acquire high-quality prospects to fit the team's roster.

Thank you!

Questions?

- Data:
 - Retrosheet.org
- Flatiron School – Data Science Bootcamp
 - Instructor: James Irving

The information used here was obtained free of charge from and is copyrighted by Retrosheet. Interested parties may contact Retrosheet at “www.retrosheet.org”