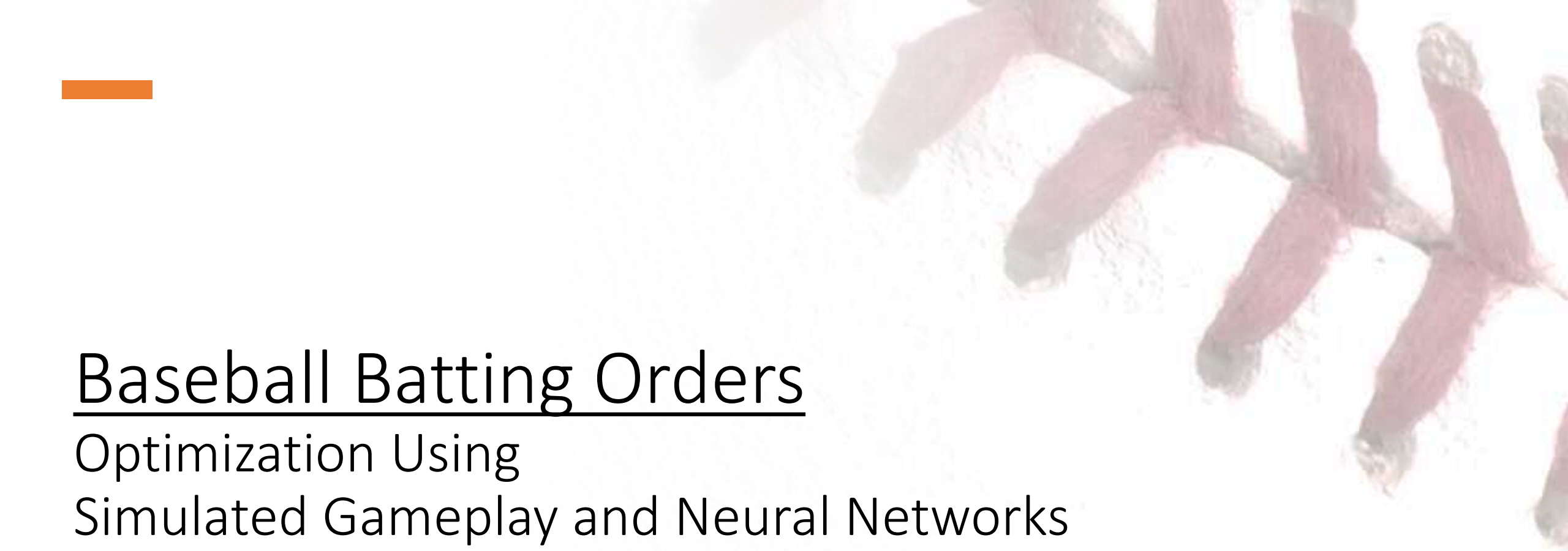


Baseball Batting Orders

Optimization Using Simulated Gameplay and Neural Networks

Claude Fried

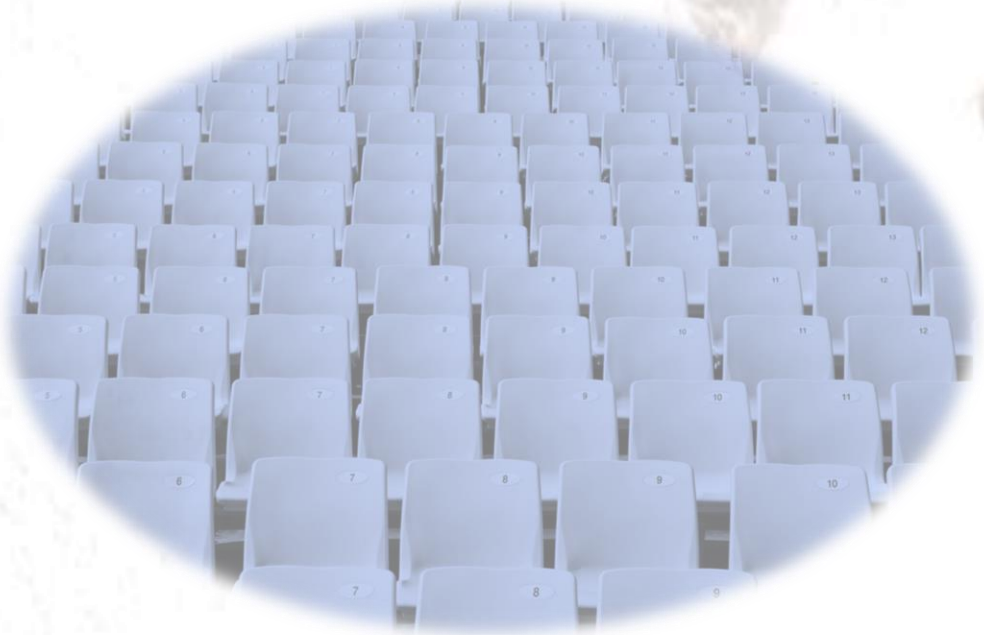
2021

Problem Statement

- Does the batting order impact the expected runs scored?
- Are baseball lineups optimal?
- Can we improve on tradition?

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. **Scrape** regular-season data for every at-bat since 1950.

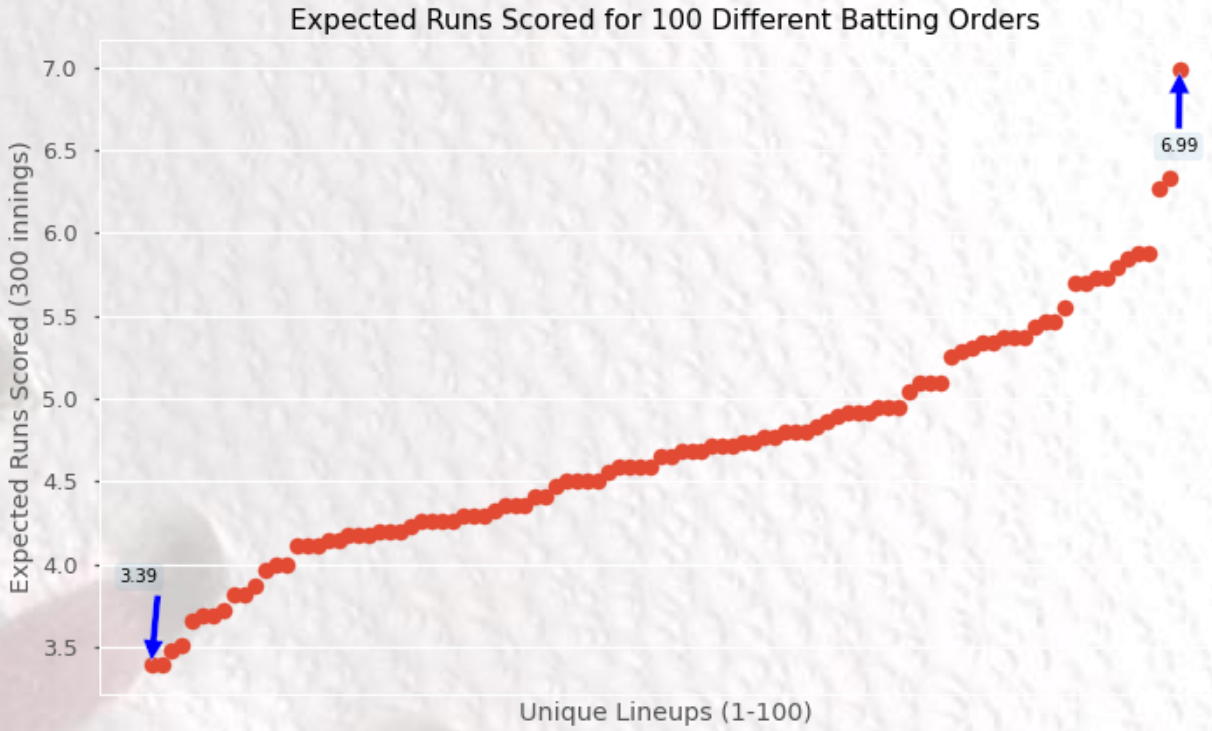
2. **Engineer** *Players* to track career stats and Pitcher/Hitter interactions.

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

- There is no *one-size-fits-all* method for setting your lineup.
  - Opposing pitchers' tendencies, game conditions, and player interactions should be at the forefront of the decision-making process.

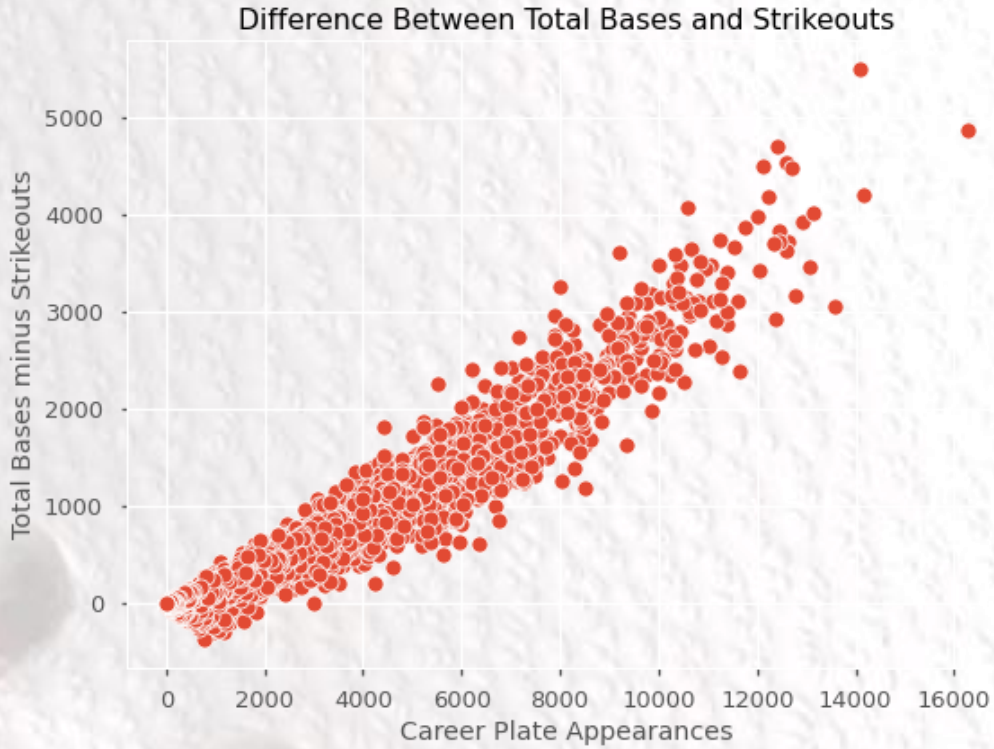


After simulating 100 randomly-shuffled lineup permutations, it is clear how impactful the batting order is on expected runs scored.

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

Baseball Recommendations

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

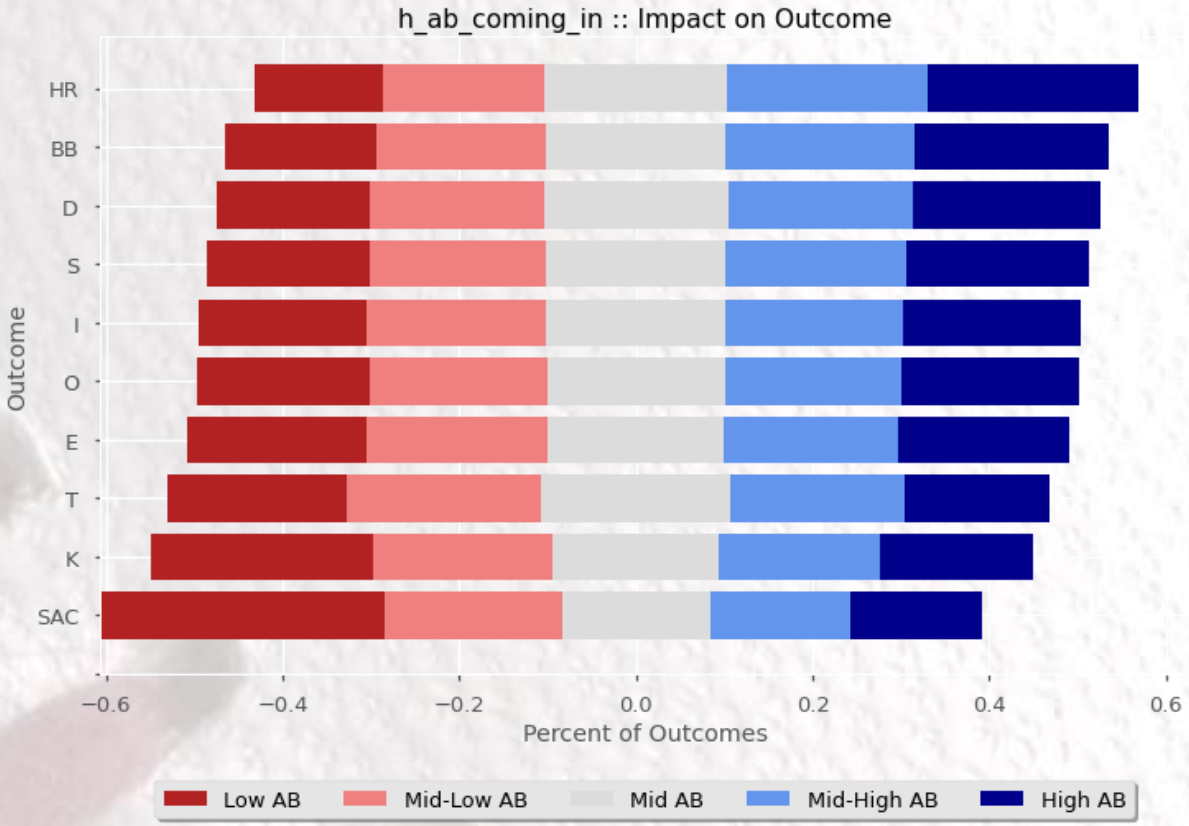


A hitter's ability to hit for a high total-bases and low strikeout total shows longevity in the league.

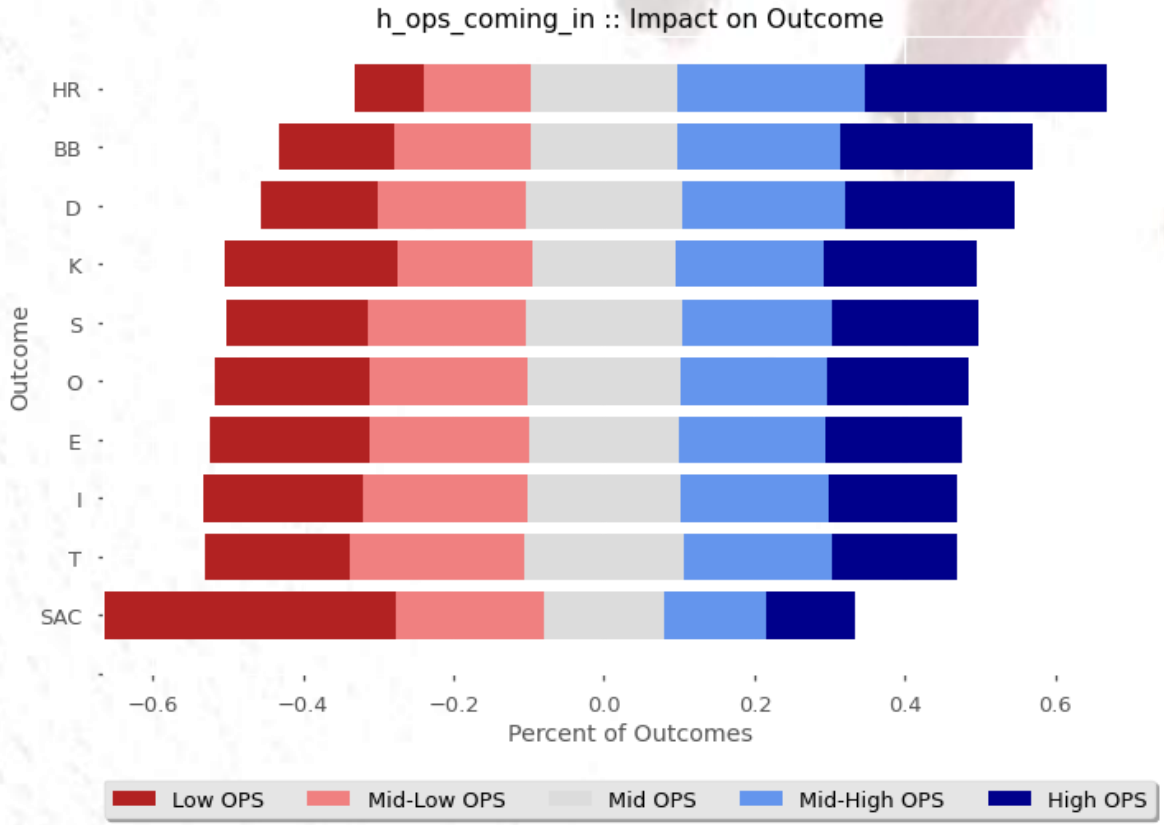
Player contracts can use this information to offer long-term deals for team-friendly money.

Baseball Recommendations

- Young hitters are prone to *sacrifice* and *strikeout*.
- More experienced hitters are prone to hit *homeruns* and draw *walks.*



- Hitters with a high *on-base-plus-slugging (OPS)* tend to sacrifice very rarely while hitting many *homeruns* and drawing *walks*.



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.

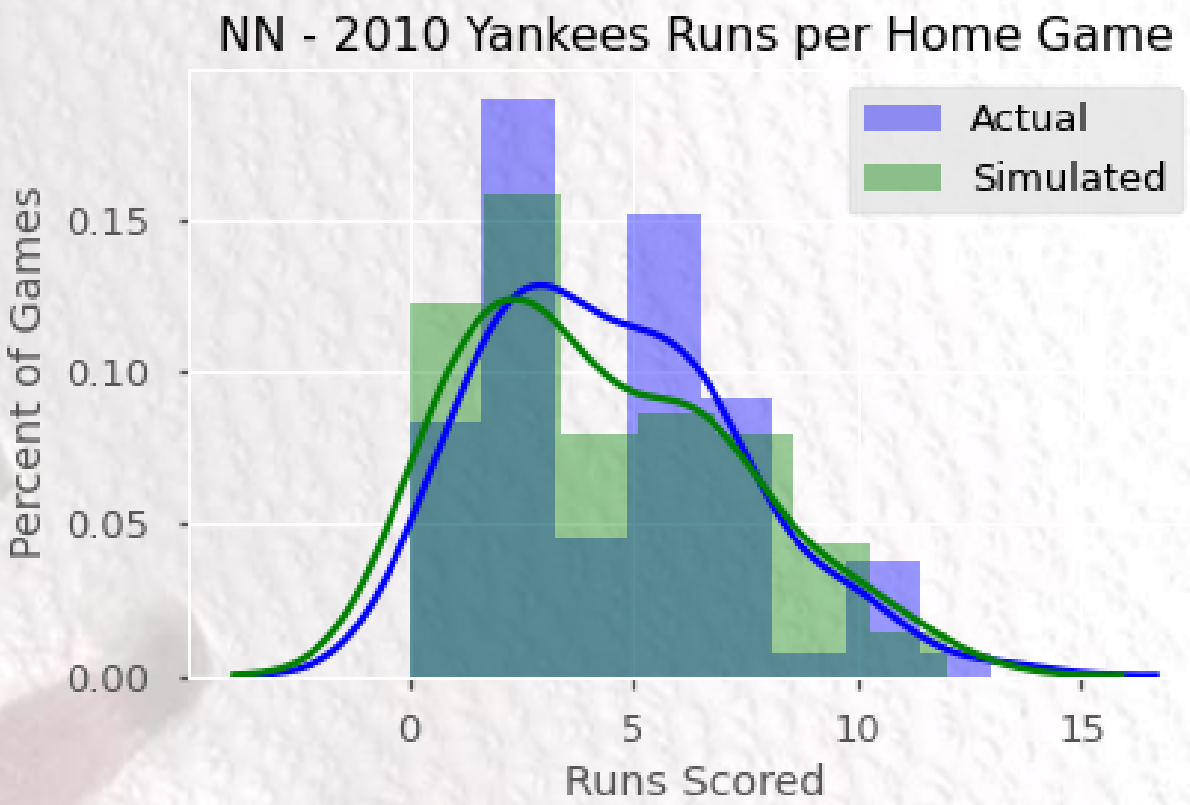
Modeling Validation

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

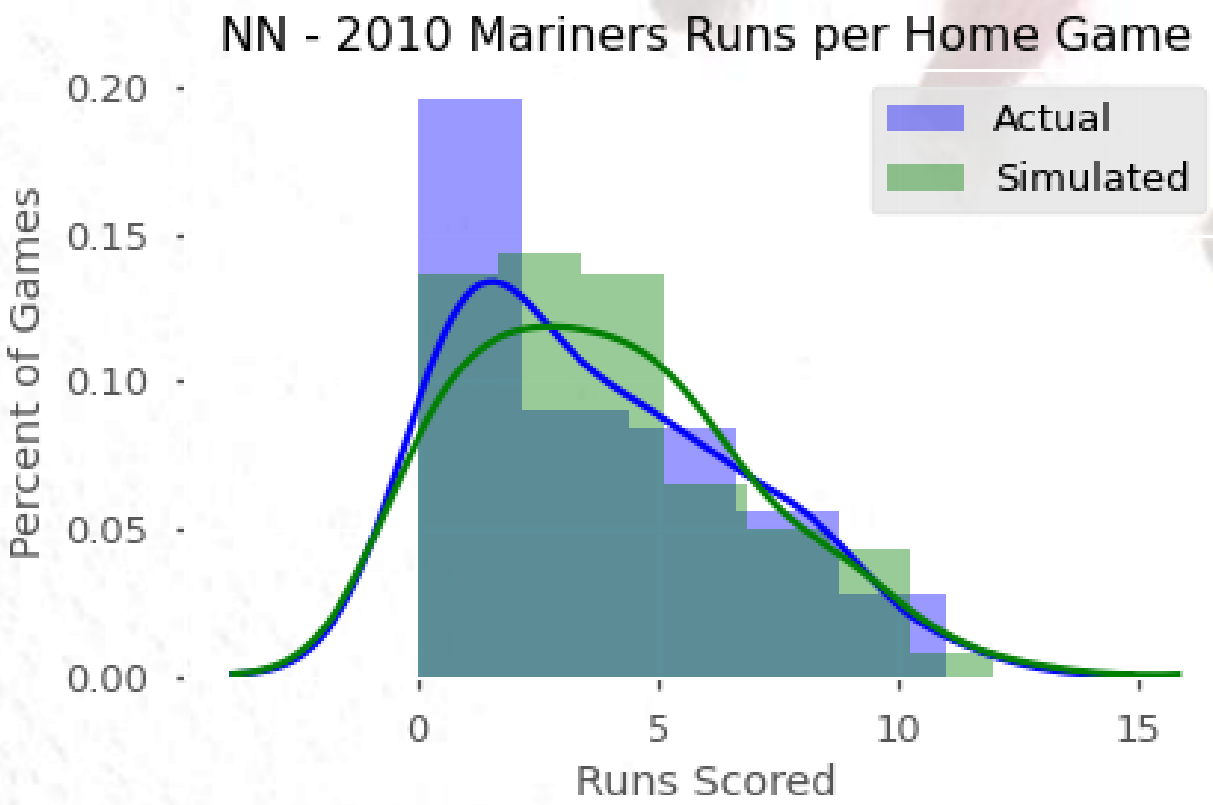
Simulated:------4.296



Mariners: Average Runs per Game

Actual:----------3.753

Simulated:------3.914



Future Work / Next Steps

- Try out different modeling architectures to try to improve performance.

- Engineer more features:
  - More stats for players.
  - *Hitter-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 source: Retrosheet.org
- Flatiron School – Data Science Bootcamp