

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green color. They are positioned diagonally, with the blue one in front of the green one.

NESCAC Baseball Lineup Optimization

Sam Berman, Clay Kynor, Alex Maleno



Overview - English Explanation

- Web application that runs baseball games using a Monte Carlo simulation to predict average number of runs given a specified batting order
- Given a batting order, simulation runs multiple 9-inning games to predict runs scored (runs as a performance/win indicator)
- Originally designed for Tufts Baseball, but can be used by any NESCAC team!



Motivation

Found way to target website to much broader demographic due to the generality of the data. We've found this could be utilized by any baseball/softball team given the user has a csv file tracking the following stats for their specific players:

- At bats (AB), Home Runs (HR)
- Hits (H), Strikeouts (K)
- Doubles (2B), Triple (3B),
- Walks (BB), Hit By Pitch (HBP)

Batting order is incredibly important and hard to optimize - lots of unwritten rules (are they accurate?)

- Sluggers are often placed 3rd or 4th with consistent base-hitters before in an attempt to optimize runs
- 7th position is often given to a good bunter or defensive player who suffers at bat




The Data












- Obtained batting data for each NESAC team for the entire 2021 season
- Stats used: AB, H, 2B, 3B, HR, K, BB, HBP
- Stats calculated: 1B, outs

Issue: Inconsistent stats on players who don't play






- Coaches can input personalized data



New England Small College
Athletic Conference

SPORTS
CHAMPIONSHIPS
SCOREBOARD
INSIDE THE NESCAC
ADMINISTRATION
NEWS & AWARDS

Baseball

Schedule / Standings / Statistics / Championship / Coaches / More +

2022 Tufts Baseball Overall Team Stats

Conference Only Stats

Select a Season...

HITTING

PITCHING

FIELDING

Hitting

	AVG	GP-GS	AB	R	H	2B	3B	HR	RBI	TB	SLG%	BB	HBP	SO	GDP	OB%	SF	SH	SB-ATT
Ozzie Fleischer	.394	27-24	99	31	39	8	2	0	16	51	.515	15	6	4	1	0.492	2	3	6-9
Jimmy Evans	.348	27-26	115	29	40	9	0	4	23	61	.530	10	2	8	2	0.397	4	0	9-11
Miles Reid	.444	27-26	108	27	48	10	0	0	23	58	.537	17	4	6	1	0.527	2	0	7-8
Peter Demaria	.385	28-27	104	23	40	10	0	5	24	65	.625	14	5	9	2	0.472	2	0	1-1
Connor Flavin	.211	18-2	19	4	4	0	0	1	4	7	.368	1	1	3	0	0.286	0	0	0-0
Kyle Cortese	.293	26-25	92	25	27	9	1	5	21	53	.576	10	11	26	0	0.417	2	1	4-4
Ryan Noone	.347	23-19	75	13	26	3	1	2	20	37	.493	4	5	12	1	0.407	2	0	1-1
Clay Sowell	.196	18-13	46	13	9	0	0	3	7	18	.391	8	8	22	1	0.403	0	0	2-2
Jackson Duffy	.186	22-17	59	7	11	2	0	0	9	13	.220	7	4	18	2	0.306	2	1	1-2
Connor Bowman	.260	22-19	73	13	19	2	0	4	19	33	.452	8	3	13	0	0.357	0	1	3-3
Patrick Solomon	.288	24-23	80	22	23	3	0	1	17	29	.363	12	1	17	0	0.371	4	1	6-7
Michael Volgende	.000	7-4	0	0	0	0	0	0	0	0	.000	0	0	0	0	0.000	0	0	0-0

Figure 1: NESCAC Website Data

<https://nescac.com/teamstats.aspx?path=baseball&year=2022&school=tufts>



Obtaining Data and Processing

- Pulled data from [NESCAC website](#)
 - Used 2021 Season data
- Converted data to CSV format, made it easier to process as pandas DataFrame

- Calculate player thresholds for results to simulate each at bat
- Use probabilities to simulate games tracking runs scored as indicator of performance


Figure 2: Example Player Thresholds

Plate Appearances	6225		
Errors	102	.016	.016
Outs (in play)	3125	.502	.518
Strikeouts	922	.148	.666
BB	531	.085	.751
HBP	61	.010	.761
Singles	981	.157	.918
2B	292	.047	.965
3B	27	.004	.969
HR	184	.031	1.000



Our Monte Carlo Simulation

- Dictionary of game state (i.e. player on first/second/third, outs, inning, runs)
- Generate a random number for each “at bat”
- Game state dictionary is updated accordingly
- Depending which threshold range number is in that event occurs in the game
- Iterate through batting lineup next at bat for different player threshold
- Repeat until inning ends and tally score



Features That We (Wanted To) Implement

- Coach inputs a list of at least 9 players
- We run 440 Monte Carlo simulations of games and compute the average runs with any given lineup (brute-force combination)
- Return the best lineup out of that list of players

BUT!

1 Monte Carlo Simulation = **~0.1s**

9 players in a lineup = 9! permutations = **362,880**

440 simulations per perm = 9! * 440 = **159,667,200**



0.1s * 159,667,200 = **15,966,720 s**

= **266,114 min**

= **4435 hrs** = **184 days**



Features That We (Actually) Implemented

- Coach inputs a list of at least 9 players and we predict the number of runs scored with that lineup
- User inputs a lineup of 9 players – we do some optimization (less than $9! * 440$ permutations) and report to them if any minor switches would improve the chance of winning
- Compare two lineups (# of runs predicted) - and random lineup



Streamlit Web Application

Build web applications, but using Python instead of HTML!

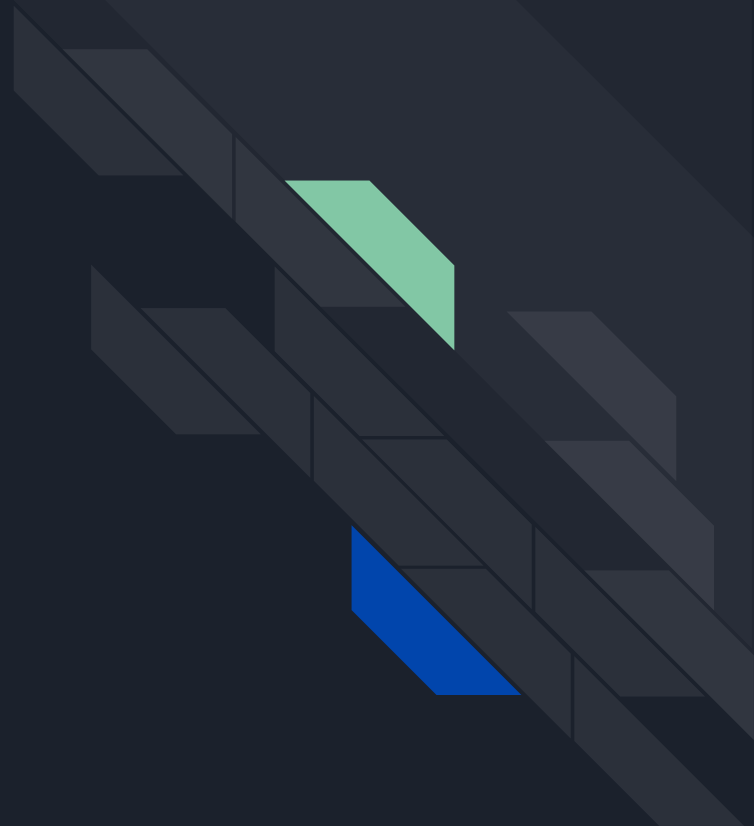
Allowed us to run Monte Carlo sim (written in Python) directly on the web app

Incredibly easy to use and produce good-looking web applications

<https://streamlit.io/>

DEMO!

<https://bit.ly/lineup-optimization>





Actionability

- Better batting lineup = More runs
- More runs = Winning games!
- Runs is the best predictor of winning baseball games
 - Ex: The first three batters in the lineup get on base consistently, the fourth is a slugger than the drive them home (lots of runs)
- Important to note that batting order can be optimized in theory by previous stats, but everything comes down to day-of conditions, opponent, etc.



Future Work

(1) Batting lineup with all “average” players and one slugger or horrible player

Find batting position of that player to maximize indicator of performance (runs)

(2) Weighting opponent’s pitching staff

(3) Softball

Questions?



Bibliography:

- Batting order importance:

<https://www.justbats.com/blog/post/what-is-the-best-spot-to-bat-in-the-batting-order/>

- Starter NESCAC dataset:

<https://nescac.com/teamstats.aspx?path=baseball&year=2022&school=tufts>

