

Cincinnati Reds Hackathon 2024

Collaborators: Ethan Mann and Brett Laderman

A. Introduction

Pitching roles in baseball are ever-changing. The Tampa Bay Rays introduced a strategy billed as an “Opener”, typically a reliever to start the game and pitch through the order at least once, followed by a “Bulk guy”, typically a starter/swingman/long reliever to stretch out through multiple innings and pitch through the order multiple times. This limits the number of times a hitter gets to see a pitcher in a game which eases the damage produced by a lineup. This innovative thought process has never been more popular as all organizations have gone to the Opener or some extension of it.

In this writing, we attempt to build a predictive model that showcases pitchers that would succeed in a different role. For instance, a starter that may succeed as a reliever, and vice versa. The idea behind our approach is using several key statistics that define pitcher success and tailoring them to their new predicted role. This not only includes statistical analysis, but baseball reasoning as well to “fact check” and certify how accurate our model is.

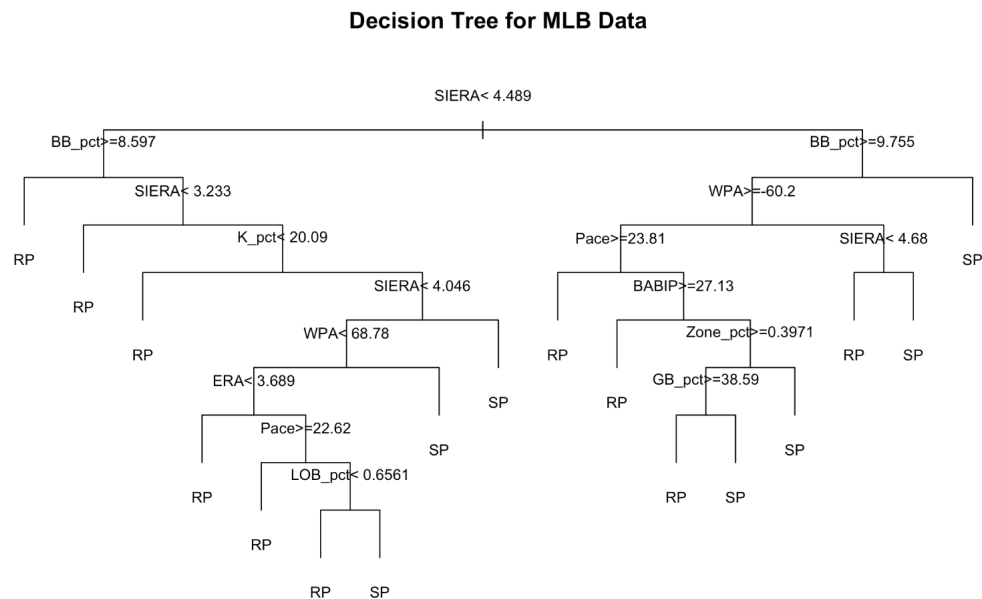
B. Analytical Application

We believe we made the correct choice in building a decision tree, hoping to predict our pitcher’s new roles. Decision trees are used to handle regression and classification tasks, using our data to classify and predict what is to follow. The tree was built using a three-year sample size, dating back to 2021. Our tree incorporates several variables, listed as an “if-then clause”. The variables that we used in this dataframe comprise:

- Zone_pct (Pitches Thrown in the zone)
- CSW_pct (Called Strikes and Whiffs)
- SIERA (Skill-Interactive-Earned-Run-Average)
- K_pct (Strikeout Percentage)
- BB_pct (Walk Percentage)
- LOB_pct (Percentage of Baserunners left on base)
- BABIP (Batting Average on Balls In Play)
- Pitch_Arsenal (# of above or above average graded pitches)
 - To explain further, from 2021-2023 the average Pitching+ grade for a Curveball was 104. If a pitcher threw that pitch at least once and their average grade was 104 or above they would receive a 1. We used 6 pitches: 4-Seam Fastball, Changeup, Slider, Curveball, Sinker, Splitter, Cutter. For example, Joe Musgrove had 6 pitches that were graded at least league average, thus giving him a 6 to Pitch_Arsenal.
- Pace (Average time taken to throw a pitch in between each pitch)
- WPA (Winning Probability Added)

If a pitcher both started a game, or came out of the bullpen, they were separated from one another instead of grouped together. This was enforced to see the pitcher's statistics and skill sets in the two roles to understand if their skill set was different. This was most noticeable in pitchers like Aaron Ashby, Kris Bubich, and Yonny Chirinos.

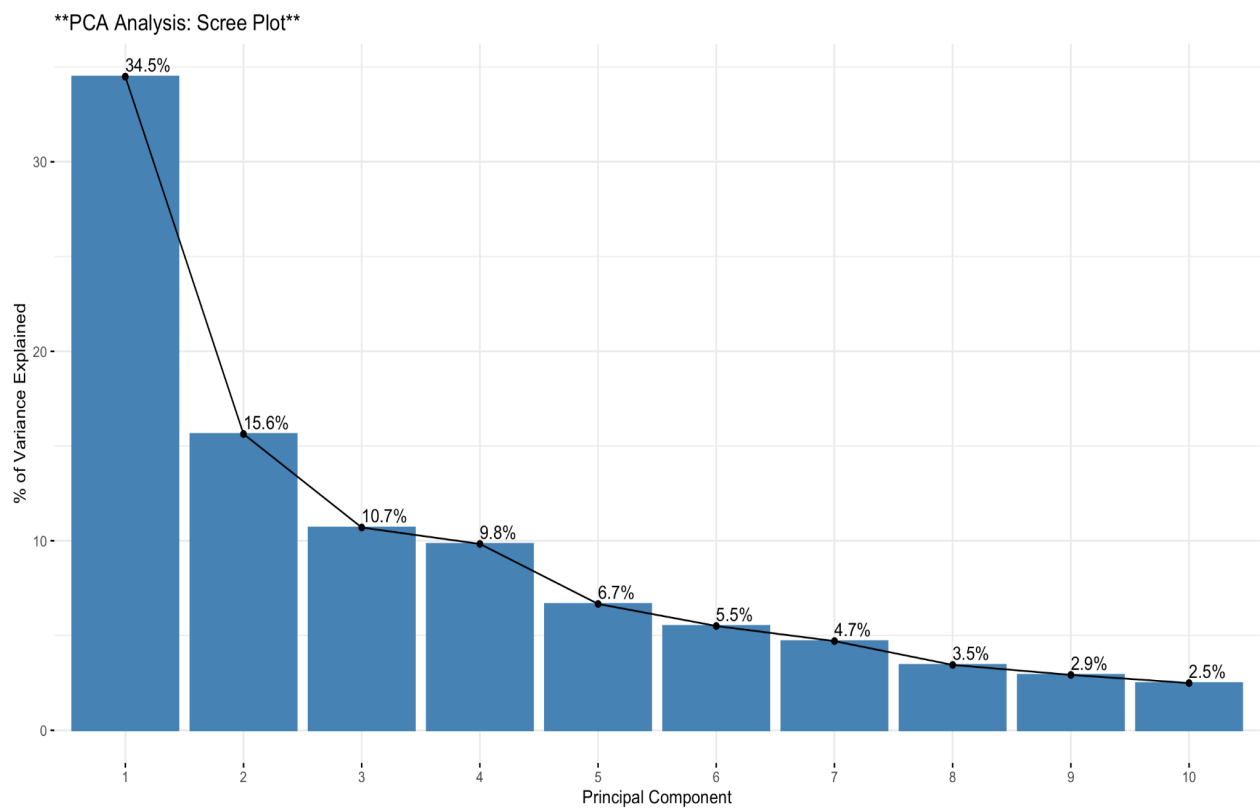
This is a detailed overview of our decision tree:

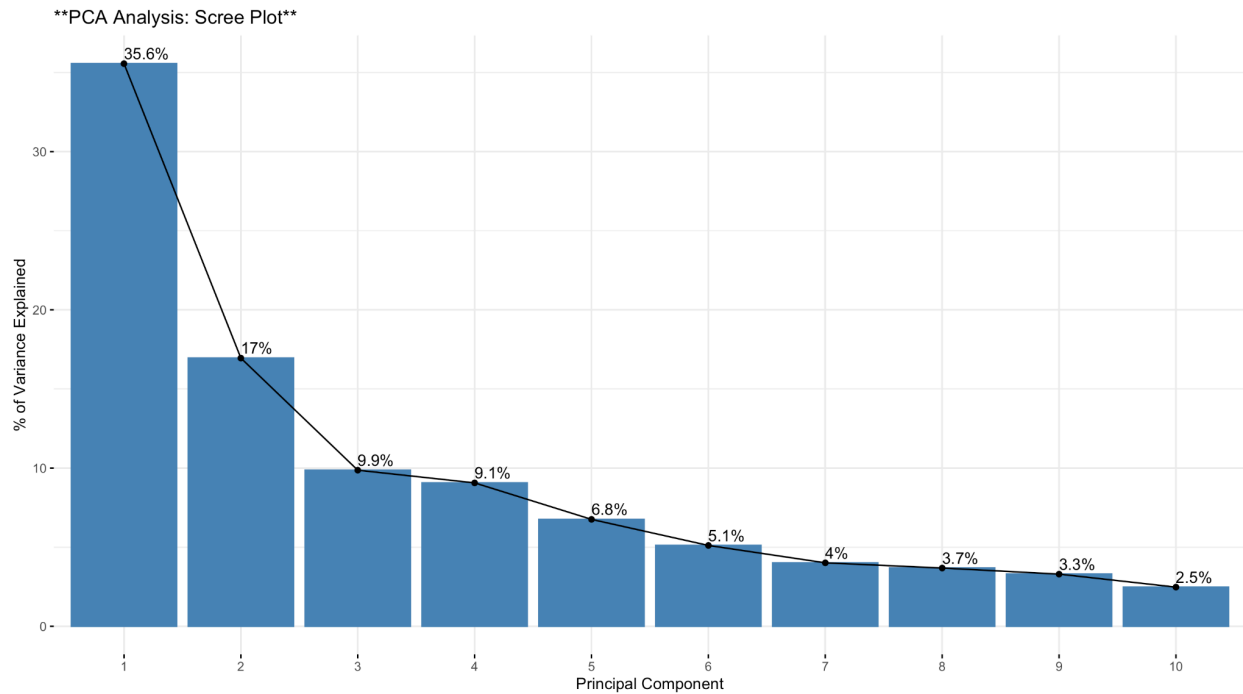


To follow the decision tree, if your decision is true, you would move to the left. To start off the tree, if you have a SIERA of 4.489 or less, the pitcher moves to the left and is decided by a BB_pct of 8.597. If you have a SIERA of 4.489 or more, move to the right and deal with the BB_pct of 9.755. The process remains consistent until the classification is determined by LOB_pct on the left, or GB_pct on the right, leading to categorization as a reliever or starter. Our finished tree had a “miscalculation” rate of 13%. That is around the range we would like for it to be, between 10% to 15%. We want some miscalculation because we want some starting pitchers to be predicted as relievers in response, and vice versa. However, we don't want a high percentage like 20% to 25% because then it shows that our decision tree is flawed with significant errors.

| Role | | |
|-------|-----|-----|
| group | RP | SP |
| RP | 539 | 74 |
| SP | 54 | 308 |

In total, there were 54 relief pitchers that became starters and 74 starters that became relief pitchers, creating an even mix. After constructing the tree, we moved to clustering, and clustered just the new starters by their skill sets, and the same for the new relievers. With slimmer observations, it was easier to dissect the respective pitchers' arsenal of skills for their new roles.





We then performed a cluster analysis of the grouped pitchers. In the process of clustering the starters, a “Scree Plot” was created. The top scree plot is for the new starting pitchers and the bottom is registered for relief pitchers. The purpose of a scree plot is to illustrate the amount of clusters that are needed to be created. To figure that out, you follow the line trending down until it moves horizontally rather than vertically. Follow until the eigenvalue levels out like we see in both graphs. That is considered the “elbow” of the plot. In both instances, only three clusters were made for each dataset.

C. Pitchers

Traditional Starter/Bulk Guy

Drey Jameson (RHP) - Arizona Diamondbacks

Our first relief pitcher turned starter is Drey Jameson. Jameson underwent Tommy John surgery this past September, so this new role will have to be in effect for the 2025 season. Amongst the new starters, Jameson was clustered in with pitchers who have a more well-rounded pitch arsenal. A factor that we heavily consider being an important factor in becoming a starting pitcher. How often have we heard that a pitcher can’t stick in the rotation because they don’t have a second pitch, let alone a third pitch? Jameson fits the bill with a Pitch_Arsenal score of 3. Here is his pitch arsenal and peripherals:

Drey Jameson Arsenal

| PITCH | PERCENT THROWN | PITCHES THROWN | INDUCED VERTICAL BREAK | HORIZONTAL BREAK | STUFF+ | LOCATION+ | PITCH+ | RUN VALUE | RUN/VALUE OVER 100 PITCHES |
|-----------|----------------|----------------|------------------------|------------------|--------|-----------|--------|-----------|----------------------------|
| Sinker | 29.8 | 201 | 5.7 | 16.0 | 92 | 95 | 89 | 0 | 0.1 |
| 4-Seam FB | 27.2 | 183 | 15.6 | 6.7 | 102 | 104 | 103 | 1 | 0.3 |
| Slider | 26.1 | 176 | 0.0 | -8.8 | 115 | 99 | 106 | 3 | 1.4 |
| Changeup | 13.2 | 89 | 1.4 | 15.0 | 95 | 94 | 93 | 2 | 2.4 |
| Curveball | 3.7 | 25 | -4.3 | -15.1 | 119 | 101 | 107 | 0 | 1.9 |

Table: Ethan Mann | Data: Baseball Savant

A 5-pitch mix with his 4SFB/SL/CB grading out to be his best three pitches, with his changeup and sinker not far behind. With the 26-year-old having a history of both starting games and coming out of the bullpen, him permanently becoming a starter in 2025 might be a best case scenario for Jameson and the Diamondbacks.

Opener

Nick Anderson (RHP) - Kansas City Royals

Our next relief pitcher converted into a starter is recently dealt pitcher, Nick Anderson. Anderson was one of the rare relief pitchers that the Rays never looked to start a game due to him being one of the game's more dominant relief pitchers in a season and a half stretch from the second half of 2019 through the 2020 season. Now a few seasons removed from a UCL surgery, Anderson has come back down to earth and him being a versatile pitcher might be in the best interest of the Kansas City Royals. In deciding an opener, they will always have to face the opposing team's best hitters. At the very most, an opener might pitch through the order once, so a vast pitch arsenal isn't principal for openers. We looked into pitchers that threw strikes and threw strikes often, and weren't giving up free bases. In the company of 54 turned starters, Anderson ranked first in Called Strikes and Whiffs %, first in first pitch strike %, 4th highest strikeout %, and had the 6th lowest walk percentage.

Full time Reliever

Nick Pivetta (RHP) - Boston Red Sox

A starting pitcher that would benefit from becoming a full-time reliever is Nick Pivetta. The tall right-handed pitcher has been extremely durable having back-to-back 30+ starts in 2021 and 2022, while still throwing 142.2 innings last year splitting time in and out of the rotation. In our model Pivetta was predicted to be a relief pitcher based on his starting stats, which was a huge success. Pivetta flourished when he was moved to the bullpen emphasizing his strike throwing prowess. That is the skill that we encourage out of relievers, combined with a high octane arm. He was in the top half of statistics like CSW%, FStrike%, and Zone%. One drawback that he suffers from is a relatively low ground ball% making him more prone to the long ball, an issue that we see in relievers, preferably making him not an opener candidate.

D. Conclusion

In today's game, pitching roles are no longer as strict as they once were. Pitchers are shuffling back and forth between the rotation and the bullpen, and getting the first out in the game and getting the last out in a different game. We chose three pitchers that would best suit in different pitching roles based on a series of statistics devoted to these present-day positions. In the meantime, there are more pitchers out there that would be better off in a new role, and some of them are listed below as our finished prediction piece with our new relievers grouped on the top; and our new starters grouped at the bottom.

```
# A tibble: 49 × 11
```

| | pitcher | player_name | Role | group | ERA | K_pct | BB_pct | GB_pct | CSW_pct | Zone_pct | FStrike_pct |
|----|---------|--------------------|-------|-------|-------|-------|--------|--------|---------|----------|-------------|
| | <dbl> | <chr> | <chr> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| 1 | 500779 | Quintana, José | SP | RP | 4.91 | 22.6 | 9.71 | 43.6 | 26.0 | 0.354 | 0.610 |
| 2 | 502179 | Espino, Paolo | SP | RP | 5.15 | 19.1 | 5.50 | 38.7 | 27.9 | 0.427 | 0.578 |
| 3 | 503449 | Peralta, Wily | SP | RP | 3.12 | 14.4 | 9.32 | 50.8 | 23.6 | 0.384 | 0.552 |
| 4 | 527048 | Pérez, Martín | SP | RP | 4.21 | 18.0 | 8.13 | 44.8 | 26.3 | 0.385 | 0.628 |
| 5 | 547001 | Poteet, Cody | SP | RP | 4.99 | 24.2 | 12.1 | 32.9 | 27.3 | 0.450 | 0.621 |
| 6 | 547179 | Lorenzen, Micha... | SP | RP | 4.15 | 19.2 | 8.93 | 45.5 | 25.7 | 0.422 | 0.645 |
| 7 | 592761 | Smith, Caleb | SP | RP | 6.95 | 24.6 | 15.2 | 29.2 | 26.9 | 0.429 | 0.572 |
| 8 | 592767 | Smyly, Drew | SP | RP | 4.61 | 20.7 | 7.29 | 37.7 | 27.2 | 0.422 | 0.646 |
| 9 | 592826 | Velasquez, Vince | SP | RP | 5.13 | 21.9 | 9.27 | 32.0 | 27.8 | 0.426 | 0.616 |
| 10 | 592866 | Williams, Trevor | SP | RP | 4.69 | 19.0 | 7.09 | 40.8 | 24.1 | 0.400 | 0.658 |

```
# A tibble: 12 × 12
```

| | pitcher | player_name | Role | group | ERA | Pitch_Arsenal | K_pct | BB_pct | GB_pct | Pace | Zone_pct |
|----|---------|--------------------|-------|-------|-------|---------------|-------|--------|--------|-------|----------|
| | <dbl> | <chr> | <chr> | <fct> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| 1 | 572193 | Tepera, Ryan | RP | SP | 3.20 | 0 | 25.5 | 8.27 | 42.5 | 26.8 | 0.366 |
| 2 | 608716 | Steckenrider, D... | RP | SP | 2.00 | 2 | 21.7 | 6.37 | 37.4 | 23.4 | 0.461 |
| 3 | 623433 | Anderson, Nick | RP | SP | 3.06 | 1 | 25.5 | 6.38 | 41.7 | 18.0 | 0.453 |
| 4 | 623437 | Topa, Justin | RP | SP | 2.61 | 4 | 21.9 | 6.45 | 56.7 | 19.0 | 0.463 |
| 5 | 641755 | Kinley, Tyler | RP | SP | 2.74 | 1 | 25.0 | 7.40 | 39.1 | 28.5 | 0.453 |
| 6 | 642528 | Loáisiga, Jonat... | RP | SP | 3.15 | 4 | 21.3 | 7.51 | 60.3 | 27.7 | 0.393 |
| 7 | 656557 | Houck, Tanner | RP | SP | 2.70 | 1 | 24.2 | 8.43 | 53.0 | 23.9 | 0.356 |
| 8 | 656578 | Jackson, Andre | RP | SP | 4.72 | 3 | 22.5 | 3.52 | 39.0 | 19.5 | 0.464 |
| 9 | 663947 | Holton, Tyler | RP | SP | 2.13 | 3 | 22.7 | 5.3 | 45.4 | 17.9 | 0.442 |
| 10 | 666745 | Brito, Jhony | RP | SP | 1.43 | 3 | 24.3 | 5 | 46.9 | 18.6 | 0.453 |

Link

The link to our code and writing pdf can be found on our Github: [Here](#).