

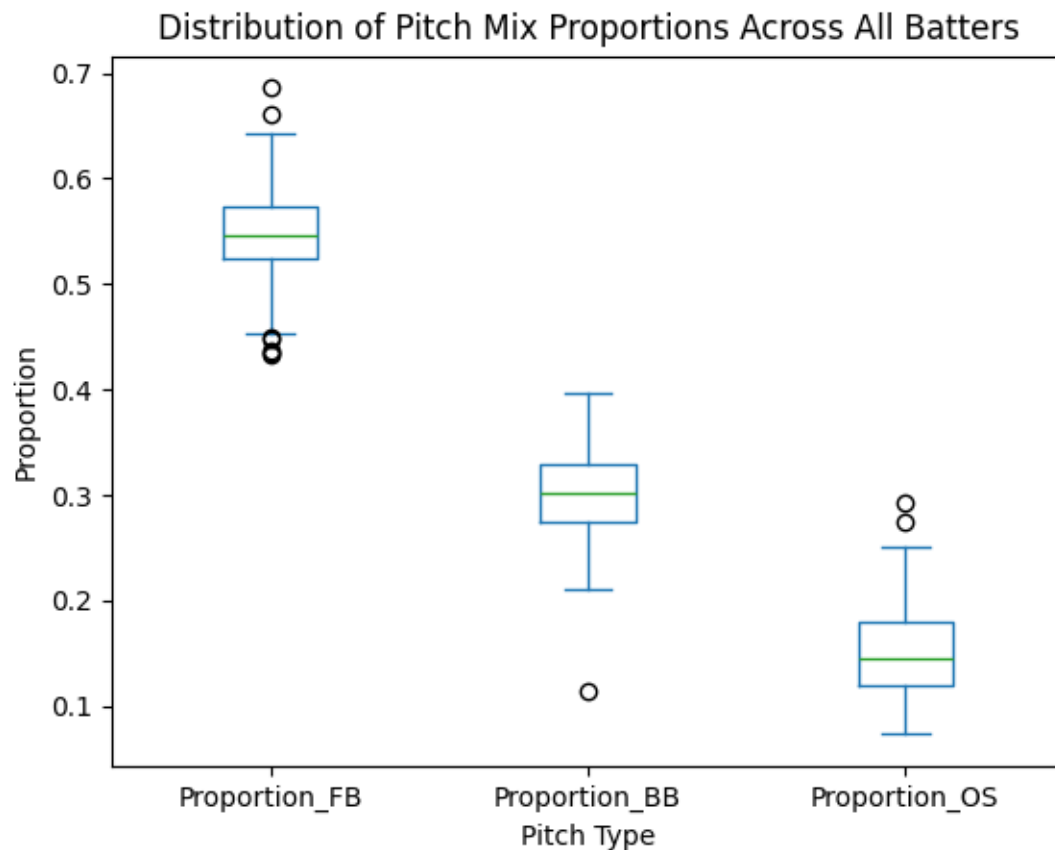
## Technical Report:

The data provided was a two-year, 2021 to 2023, set of pitch-by-pitch record, focusing on batters who faced a minimum of 1,000 MLB pitches during the respective years. Within the data set, each pitch was categorized by type, allowing me to group these pitches into three main categories: fastballs (four-seam, two-seam, and sinker types), breaking balls (slider, curveball, and knucklecurve), and off-speed pitches (changeups and splitters). Any pitch that does not fit into this category was filtered out. After this categorization, the data was grouped by individual batters given their unique identifier, BATTER\_ID, given in the csv file. Each batter's total number of fastballs, breaking balls, and off-speed pitches faced within the sample were tallied. With the total number of pitches tallied, proportions of pitch type were calculated. For example, if a batter faced 1,200 FBs, 1,500 BBs, and 300 OS pitches their respective proportions would be: 40% FBs, 50% BBs, and 10% OS. These proportions for each batter were then used as the projected pitch mix for 2024. I implemented a hybrid approach combining a deep learning model and Bayesian update for a more accurate predictive model. The LSTM was trained on sequences of 10 pitches per batter, where each sequence is used to predict the next likely pitch type based on patterns observed in the previous pitches. This training is grounded on the need to capture meaningful patterns as opposed to a historical proportion-based projection. This will account for any changes within the batter's approach within the time frame given in the csv providing a more accurate future projection. Additionally, I used Bayesian updating to dynamically refine pitch mic predictions. Bayesian updating enabled adjustment of these initial proportions as new pitches are observed in 2024, allowing the model to adapt to in-season changes in pitch strategy. During training, the model's performance is monitored through epoch and loss outputs. Each epoch represents a complete pass through the entire training data set and provides the model with

another opportunity to learn from the data. Loss is a measure of prediction accuracy calculated by Cross Entropy Loss. It quantifies how closely the model's predictions align with the actual labels in the training data. My approach, while flexible and adaptive, does come with some limitations. This approach is solely based on past data, which may not fully account for strategic adjustments that occur in real time. Bayesian updating partially accounts for this issue but cannot possibly resolve it. My sequence length of 10 captures relevant patterns, but it also limits the ability to see long-term trends. This approach is also very computationally intensive. It has a long runtime and requires hardware specifications to be used consistently. Even with these limitations, I still believe that this intensive approach is the best approach for the problem presented.

#### Coaching Staff Report:

Given the pitch-by-pitch report from the 2021-2023 seasons, our analysis aims to understand the expected pitch mix that certain batters will face in the 2024 season. By examining historical trends and trends within a 10-pitch sequence, we capture both the historical tendencies as well as any adjustments made within smaller timeframes. This dual approach provides a more nuanced set of data that is better applicable to a broader range of scenarios, helping to tailor training strategies to specific game situations. I will provide two visuals to better explain what I mean.



The first of the visuals, as seen above, represents the distribution of pitch mix proportions across all batters who have seen at least 1,000 pitches from 2021-2023.

Given this information we can see that Fastballs are the most common pitch type, as most batters are expected to see between 45%-65% fastballs. Mastering fastball recognition and timing is essential, as these pitches are the foundation of most pitchers' arsenals. Knowing this the coaches can now prioritize velocity training, reaction time drills, tracking and barrel awareness, and count based fastball situations.

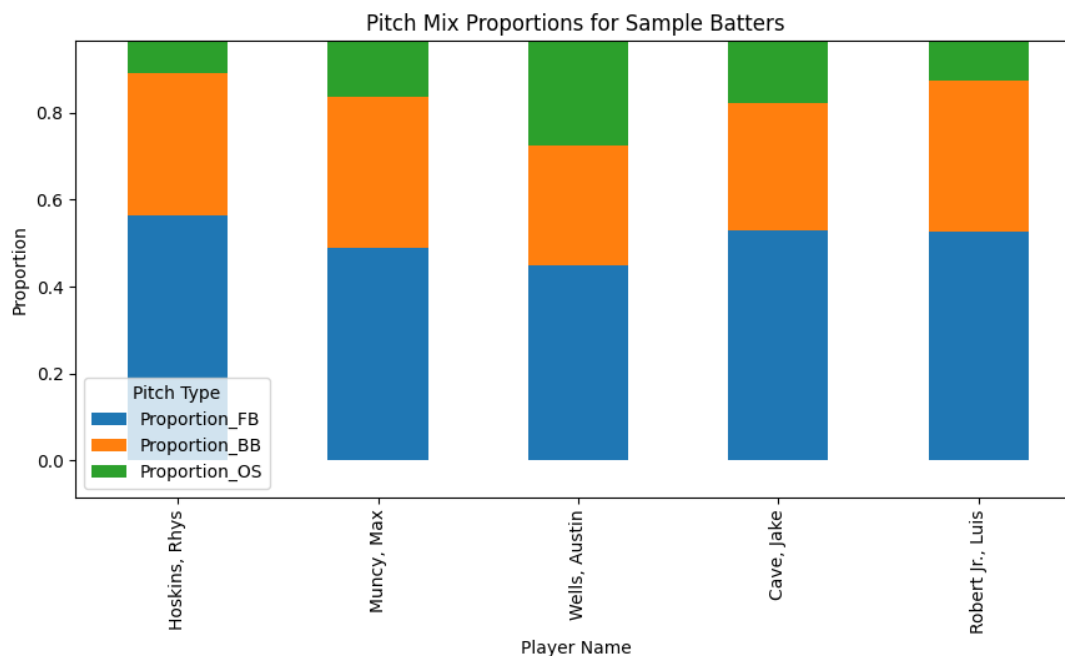
Breaking balls are the second most common pitch type, with a broader range of frequency, generally 30%, across batters. These pitches require different timing and adjustments.

Drills recommended are pitch recognition drills, tracking and layoff drills, two-strike approach, and breaking ball balance.

Off-speed pitches, while the least common, only shown at a rate of 10%-20%, need to be coached as they disrupt timing and balance. Drills recommended are timing adjustment drills, recognition drills, and reaction drills.

While this information is broad and not applicable to individual players, it is important to know these general trends. Through these trends coaches can better allocate their resources and allow for a more efficient practice session. This data-driven approach creates customized training regimes that directly correlate to the trends in which we see at the MLB level.

Now let's look at this information from an individual player's perspective.



The second visual, as shown above, represents the pitch mix proportions for individual batters: Rhys Hoskins, Max Muncy, Austin Wells, Jack Cave, and Luis Rober Jr.

I will now begin to explore what this graph shows and the certain coaching strategies that can help each individual batter.

### **Rhys Hoskins:**

Hoskins' expected pitch mix is roughly 50% fastballs, 35% breaking balls, and 15% off-speed pitches. Given this data, we can assume that Rhys is likely to see a significant number of fastballs, so the coaching staff can prioritize fastball-heavy counts to better prepare him.

### **Austin Wells:**

Wells' expected pitch mix is roughly 45% fastballs, 35% breaking balls, and 20% off-speed pitches. Wells' pitch mix includes a slightly higher off-speed proportion. Training sessions should include a mix of off-speed pitches to improve timing and adjustment to slower pitches.

### **Jack Cave:**

Caves' expected pitch mix is roughly 50% fastballs, 35% breaking balls, and 15% off-speed pitches. Focusing on timing drills with fastballs and keeping him prepared for breaking balls in late counts will be beneficial.

### **Luis Robert Jr:**

Robert's expected pitch mix is roughly 45% fastballs, 35% breaking balls, and 20% off-speed pitches. With an expected increase in off-speed pitches, Luis may benefit from practicing off-speed pitch recognition and improving his approach to slower pitches.

These individualized predictions allow the coaching staff to tailor training sessions for each batter based on the specific pitch mix they are likely to face. By focusing on the most relevant pitch types and using targeted drills, the team can improve each player's timing, pitch recognition, and overall adaptability. This data-driven, customized approach not only prepares each player for the unique challenges they will encounter in-game but also enhances their readiness across a broader range of scenarios. As a result, the coaching staff can make the most of practice time, ensuring that each batter is equipped to handle the types of pitches they are most likely to see in the 2024 season