

Kansas City Royals Assessment

Questionnaire

- 1. In 100 words or less, which single statistic do you think is most important when evaluating pitchers? Explain.**
 - a. SIERA or Skill-Interactive ERA would be my most important single statistic as it attempts to uncover the underlying skill of a pitcher while controlling for not strikeouts, walks and homeruns like FIP or xFIP, but also adjusting for the balls in play. This statistic eliminates the factors a pitcher cannot control by himself, while also adjusting for the types of ball in play.
- 2. In 100 words or less, which single statistic do you think is most important when evaluating hitters? Explain.**
 - a. xwOBA or Expected Weighted On-base Average would be my most important single statistic for hitters, as it focuses on using exit velocity, launch angle and sprint speed (on certain batted balls) to evaluate a player's overall expected contribution on offense. I prefer xwOBA over wOBA, as it actually offers the expected performance of a batter rather than the on-field result.
- 3. Who is the most undervalued prospect in today's landscape? Explain.**
 - a. For Position players,
 - i. The most undervalued prospect in today's landscape, for me, would be Chase Meidroth. He is a 22 year old 2B (with a fallback position on Left Field), ranked #20 in the Boston Red Sox farm system.
 - ii. He had a nice 2023 season moving from A+ to AA managing to get a .954 and then .761 OPS with a .255 average in AA. He definitely needs another year to prove himself. He has also been good defensively.
 - b. For Pitchers,
 - i. The most undervalued prospect in today's landscape would be Wen Hui Pan, ranked #20 in the Phillies system. His fastball ranges from 96 to 100mph and averages around 96mph and he also has a splitter, slider, changeup and curve in his arsenal.
 - ii. He just had 6 innings in 6 games in A+, so we can't really judge the era from that. However, given the amount of pitches and looking at the xFIP for A+ starts, he seems to be underrated.
- 4. Write a short paragraph explaining who you think the Kansas City Royals should acquire in the offseason – whether through Free Agency or Trade.**
 - a. The Kansas City Royals need a rebuild, and with the Kyle Wright and Nick Anderson trade it makes sense to go for Free Agents whom we can sign for 1-year contracts, like veterans or injury bouncebacks, preferably cheaper contracts. We don't really have a lot of trading pieces.

- b. Lou Severino & Craig Kimbrel could be the cheaper option for rotation and bullpen. Jack Flaherty & Lucas Giolito could be great starting options, but we will have to spill cash and keep the contract longer.
5. **The Royals have a chance to acquire one of the two following relievers. Which player would you prefer? Please provide your reasoning:**
 - Player A: 4.50 ERA, 3.00 FIP, 28 K%, 6 BB%**
 - Player B: 1.75 ERA, 2.80 FIP, 20 K%, 8 BB%**
 - a. I would prefer Player A due to,
 - i. his good K% (around 80th percentile in MLB 2023) and BB% → could lead to him getting more people struck out and would be cheaper to sign (we can take a shot)
 - ii. his FIP being lower than ERA → could be just unlucky
 - b. Also, for Player B even though the FIP and ERA is low, the low K% and relatively higher BB% is a concern. Again, having a low ERA (and higher FIP) could just be due to a good defense.
6. **List 5 things to consider before calling up a prospect to the Majors? Provide a brief explanation.**
 - a. how hard he hits the ball along with barrel %
 - b. how does the prospect handle breaking balls
 - c. how does the prospect handle fastballs up above
 - d. prospect's mentality (ask scouts, whether he has the dawg in him)
 - e. prospect's chase, whiff and k%
7. **Identify a middle inning reliever who you think should be used in high leverage situations more often. Explain.**
 - a. To identify a reliever for this question, I sorted all the relievers with their Win Probability Added (WPA) and then looked at the relievers having gmLI (or average league index entering the game) and inLI close to 1. The stat shows the manager's perspective to indicate the level of fire the manager wanted his reliever to face. After doing so, I found Ryan Brasier and Julian Merryweather as the two options that should be used more often in a high leverage situation.
8. **Prioritize the following traits for catchers: Throwing, Blocking, Framing, Pitch Calling. Explain.**
 - a. Framing
 - b. Throwing
 - c. Blocking
 - d. Pitch Calling

Catchers' skills are ranked with a focus on strategy and statistics. Framing pitches to get more strike calls is now seen as the most important skill, as it can change the game by helping pitchers (roboup may change the outlook here). Throwing is next, important for stopping players from stealing bases, especially since the

bases are bigger now. Blocking is also key to keep wild pitches and passed balls under control. Pitch calling has become a bit less critical due to teams relying more on data and planning for deciding pitches.

9. Player A is an above average defender at all infield and outfield positions with a .700 OPS. Player B is a below average defensive 1st baseman with a .900 OPS.

a. With no other information, which player would you prefer? Briefly explain.

- i. If I'm looking to improve my defense, then I would prefer Player A because .700 OPS is not too bad if he's a contact hitter in the high .250-300. Though, if I'm good on defense and I am just looking to add a bat to my lineup, I would prefer Player B since .900 OPS is nice to add to the offense. For me,

b. You are given the opportunity to ask 3 follow up questions. What 3 questions would you ask to make the most informed decision possible?

- i. What is the current composition of our team in terms of batting and fielding capabilities, and are we in greater need of a versatile defender or a power hitter?
- ii. Does Player A primarily contribute through contact hitting and on-base ability or is he a slugger?/Do we need more contact hitters or power hitters?
- iii. What are the platoon splits for both Player A and Player B?/Do they perform significantly better or worse against right-handed or left-handed pitching?

10. We have identified that our bullpen needs to be better at getting LHHs out, however, there are no LHPs available to acquire. In this scenario, what pitch characteristics would you look for in a RHP that would suggest he is an advantageous matchup versus LHHs?

- a. Horizontal movement of
 - i. RHP with Sinker or 2-Seam in their arsenal with good command can be good options

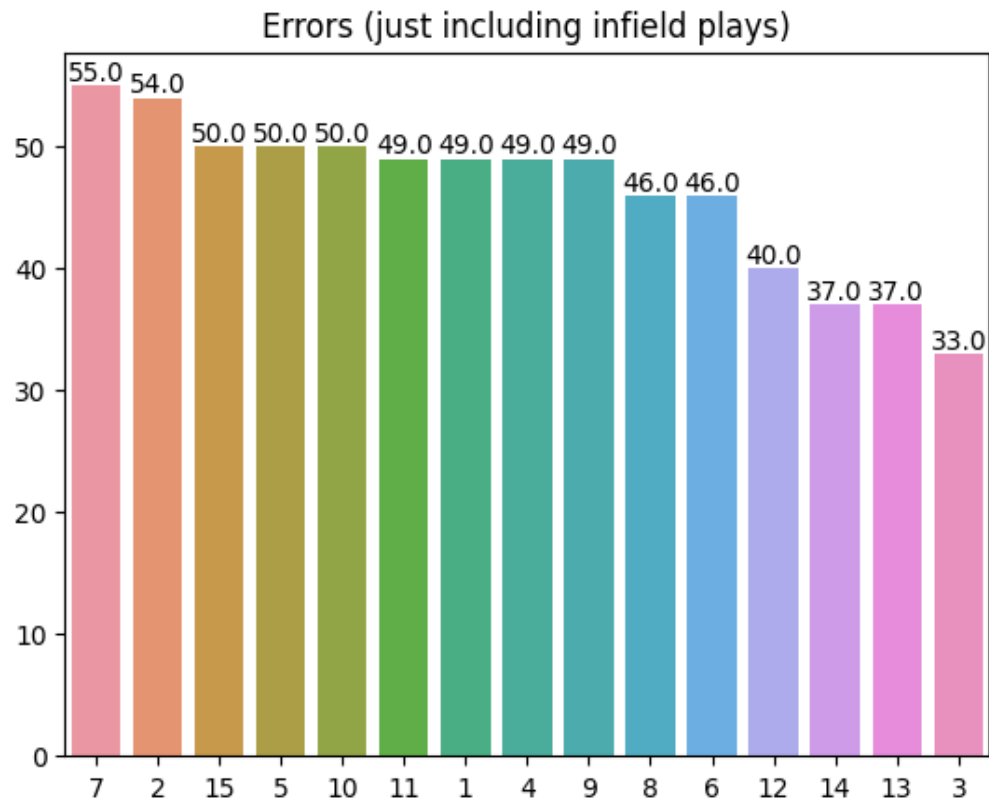
Querying Prompts

Using the attached dataset of throws to first base, please answer the following questions. Please utilize SQL, R or Python to aggregate the data. Attach all code and visualizations that you used throughout your entire process.

1. Which 5 infielders had the quickest exchange times on throws to first base?

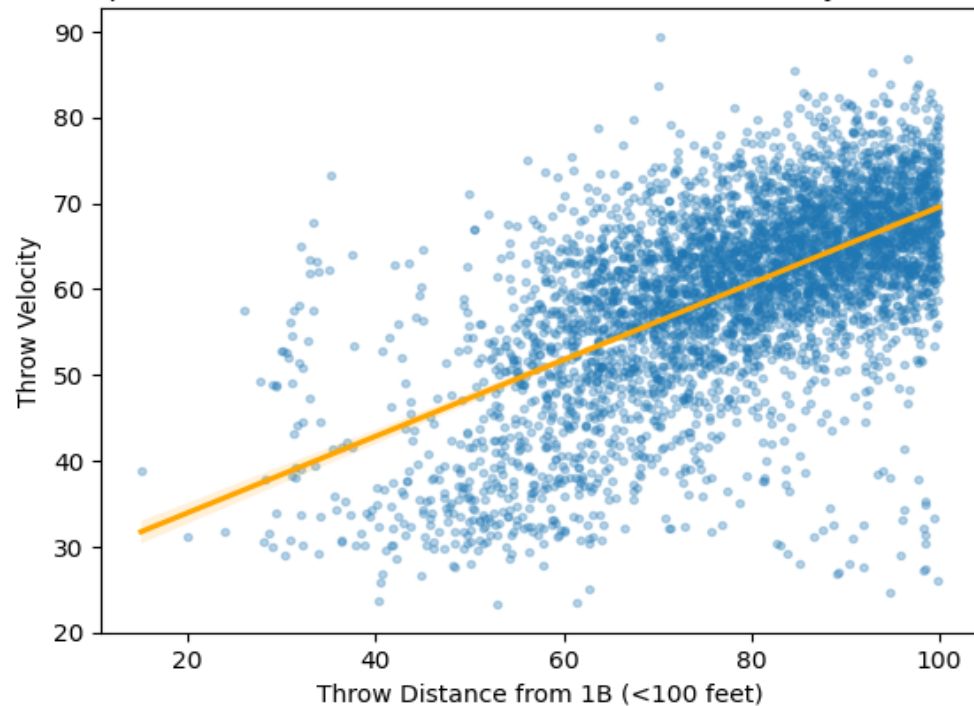
- a. I faced a lot of confusion in this one and had to assume a bunch of stuff before giving the results, and we see that these thrower_id & receiver_id had the quickest exchange time,
 - i. **21-356** with an exchange time of **0.00**
 - ii. **836-451** with an exchange time of **0.00**

- iii. **717-45** with an exchange time of **0.00**
 - iv. **800-339** with an exchange time of **0.4**
 - v. **652-854** with an exchange time of **0.4**
- b. Now the assumptions I took or the questions I had for this prompt,
 - i. We don't know if the exchange time of 0.00 is faulty or not.
 - ii. We don't know what the appropriate number of samples is while considering these exchanges between infielders.
 - iii. We can improve on this by accounting for (thrower_id, receiver_id) pairs of more than just 1. For example, if we keep a **cutoff of 10 exchanges** we get,
 - 1. **159-17** from team_id **17** with an exchange time of **0.749**
 - 2. **166-167** from team_id **9** with an exchange time of **0.783**
 - 3. **655-60** from team_id **11** with an exchange time of **0.820**
 - 4. **783-397** from team_id **4** with an exchange time of **0.836**
 - 5. **300-311** from team_id **1** with an exchange time of **0.841**
- 2. **The infield coach wants to see which teams made the most errant throws to first base. An errant throw is described as a throw that bounced and resulted in the runner being safe. Please create a basic visual that you would present to the infield coach to present your findings.**
 - a. Only considered throws to first base by the infield players for the infield coach, as the outfield throws may be considered as hits if they are not outs, We see that *team_id 7* had the most errors made in the infield.



3. Looking at all infield throws to first base, given that the distance of the throw to first base was in the top 90th percentile,
 - a. What team had the best average exchange time?
 - i. The best average exchange time was by *team_id* 9 with an average exchange time of about 1.121
 - b. Which team had the largest variation in exchange time on these throws?
 - i. The team that had the largest variation in exchange time on these throws was *team_id* 7 with a variance of about 0.205, significantly higher than other teams.
4. Given that a throw was made less than 100 feet from first base, is there a correlation between throw velocity and throw distance? Provide a basic visual alongside a brief explanation.

Regression plot to look at correlation between Throw Velocity and Throw Distance

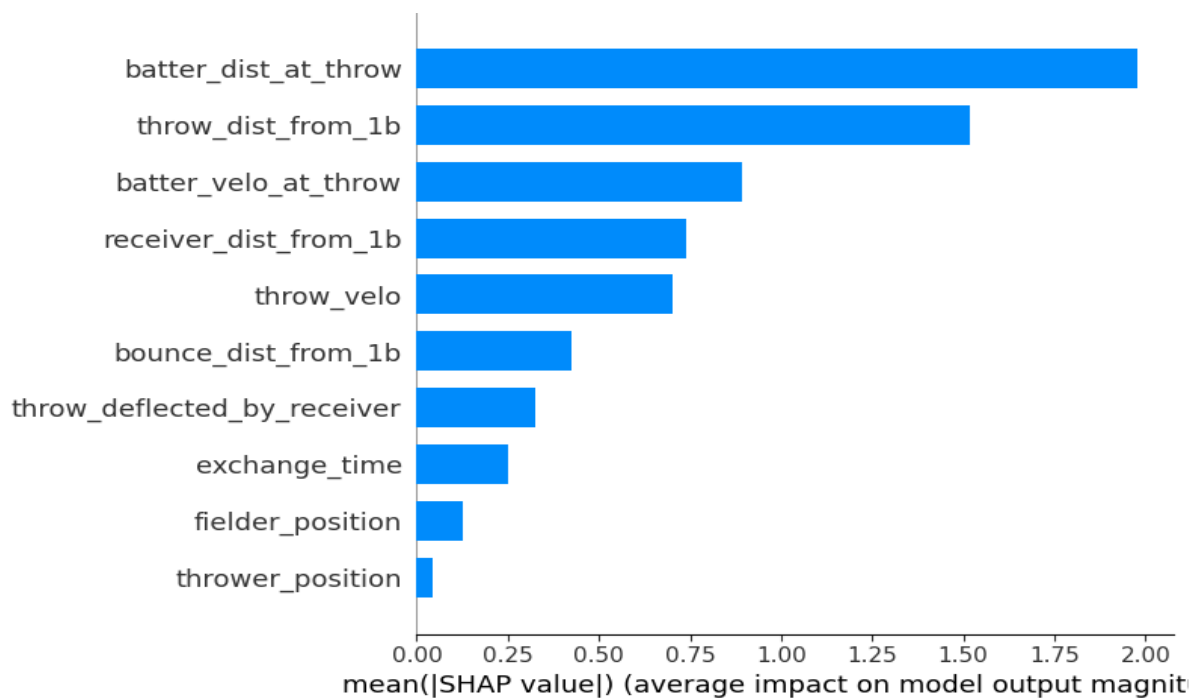


- a. For throws made less than 100 feet from first base, there is a **positive** correlation between the throw velocity and throw distance with a **correlation score of 0.610**.
- b. This means that as the distance of the throw from the first base increases, the velocity of the throw also increases. This makes sense, as you will have to throw harder with an increase in distance.

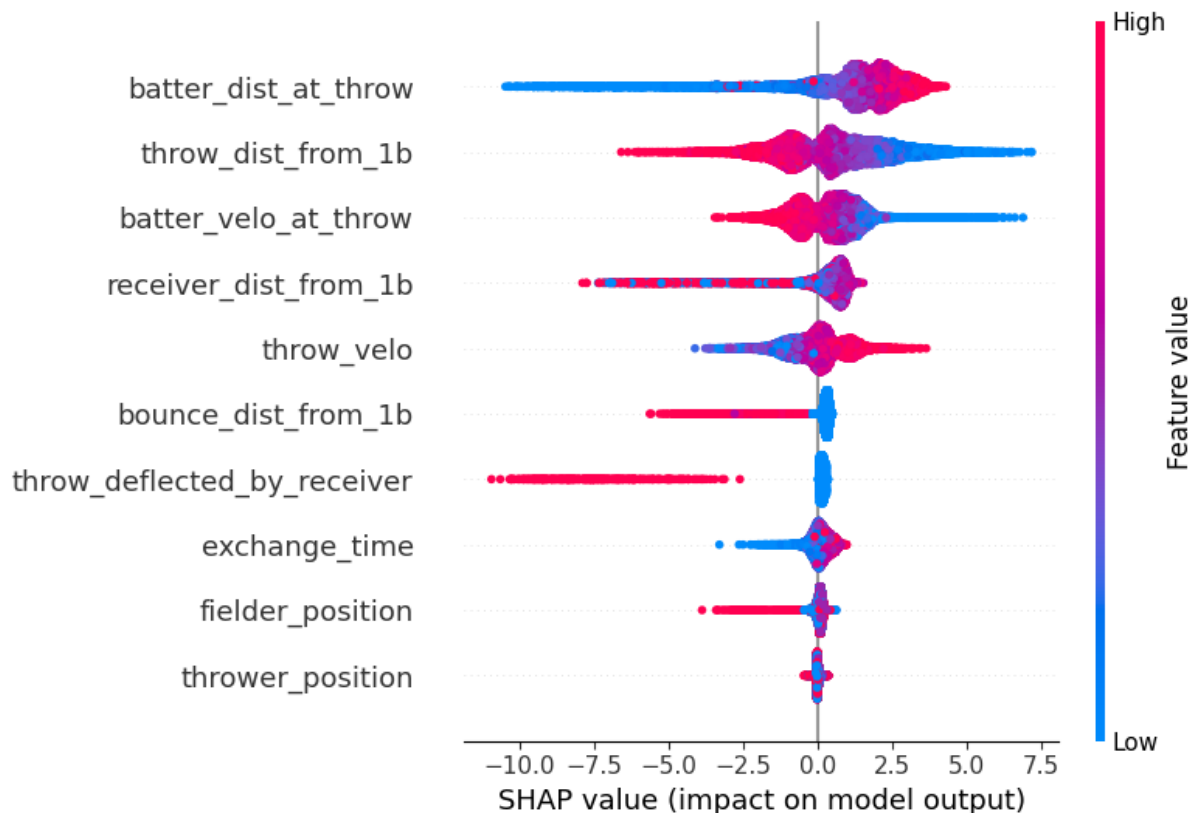
Modeling Project

Being very honest this assessment was by far the most lengthy one I have done in the past month. I can just attach the notebook that I worked on for this project. I built a classification model using XGBoost to predict successful outcomes of getting an out. I used the tree-based method because it allows me to use SHAPley values to understand variable importance and what contributes to determine whether an out can be made on a play.

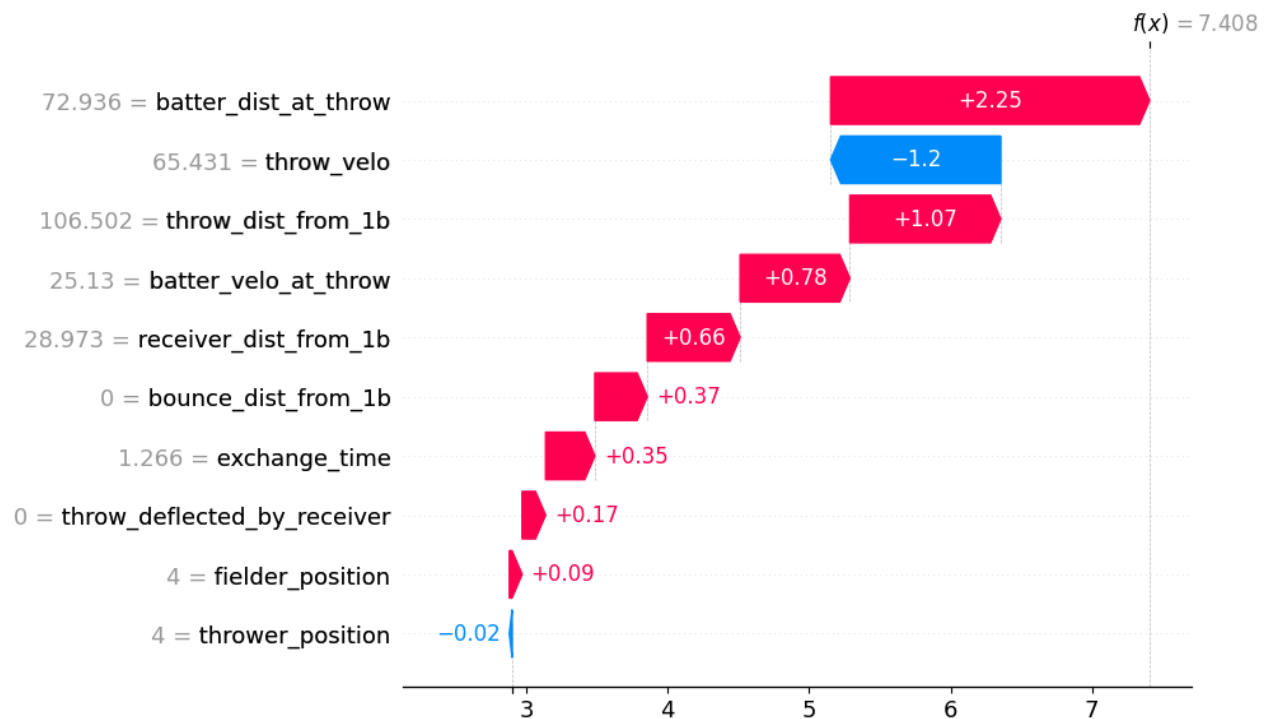
Since, batter_dist_at_throw, throw_dist_from_1b, batter_velo_at_throw and receiver_dist_from_1b are factors we cannot control for, we put our focus on throw_velo, bounce_dist_from_1b, throw_deflected_by_receiver as well as exchange_time. However, these variables that we can control have comparatively lower interaction with the model output.



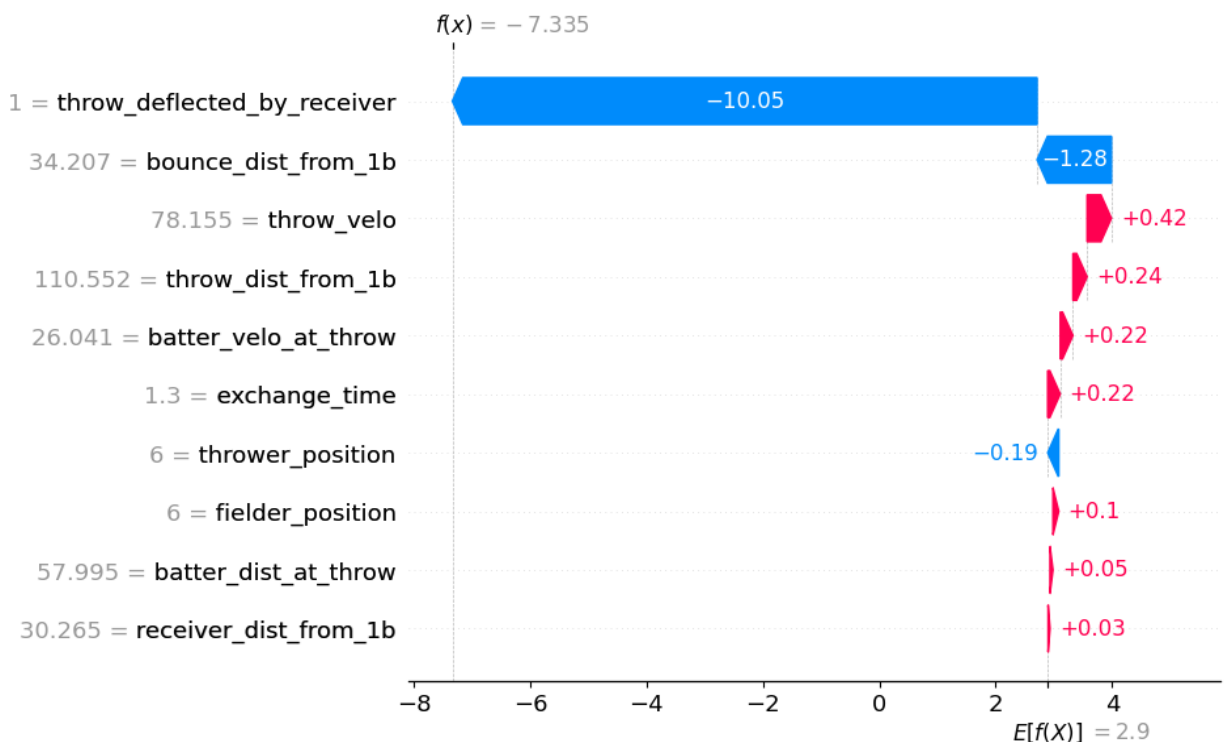
We can also see from the summary plot here that, high values of `batter_dist_at_throw` (in red) have a high and positive SHAP value, meaning when the batter is further from the first base at the time of the throw, it positively impacts the model's output which means that the play will result in an out (1).



Similarly, we can focus on SHAP values of the other predictors, and try to control for what we can.



For example, we see that a throw_velocity of 65.431 pushes the chances of an out by -1.2 SHAP value. Or in the case of a play not resulting in an out, the throw being deflected by the receiver pushes the chances of an out by -10.05 SHAP value.



I understand this is not a complete submission for the modeling project, but if given more time, I would try to explore the cut-offs for predictors that we can control like throw_velo, throw_deflection or even the thrower and receiver positions from each of the throw data.