Rounding Third:

Modeling When a Third Base Coach Should Send a Runner Home

Team 66

Abstract:

Our project attempts to grade and optimize decisions made by third base coaches. We do this by first modeling the probability of a player being safe or out if they are sent home. This is calculated from information that should be available to a third base coach during a play, such as player distances to home, baserunner speed, outfielder momentum, and game state. We then use this information to see how good third base coaches are at making the correct send or stay decision, and what they could be doing to make better decisions.

Introduction:

In baseball, every run matters, and because of that a third base coach could be the deciding factor in a game. A runner rounding third can't be looking back trying to assess the play, they're relying on their third base coach to do that. And sometimes they're going to make the wrong choice. Take this play for example.

May 31st 2024, the Cubs are down 5-3 to the Reds in the bottom of the 9th. They have runners on first and third with 1 out with Seiya Suzuki at the plate. Suzuki hits one down the left field line that lands fair in the corner, obviously scoring Pete Crow-Armstrong from third and making it a one run game. Nick Madrigal was on first and is now rounding third with the throw on its way to cutoff man Elly De La Cruz. Cubs third base coach Willie Harris has a game-deciding decision on his hands. Either send Madrigal to try and tie the game, or hold him at third and trust the Cubs heart of the lineup coming up. Harris chooses to send him, wrong choice. De La Cruz guns down Madrigal at the plate, which is followed up by a Cody Bellinger fly out to end a 5-4 Reds win.

Figure 1: Screenshot of the play from the MLB X account

This is just one example, but we've all seen our fair share of bad third base coach decisions (especially if you're a Yankees fan). And it got us thinking, how good are third base coaches? And what could they do to be better? In our project, we attempt to tackle these questions about possibly the most underestimated person on the baseball field.

Main Factors:

We decided on three major factors a third base coach would have at their disposal to decide on sending a runner home or not. These factors being player distances to home plate, the speed of the runner, and the momentum of the outfielder (see Appendix A on page 10 if you want to see other factors we considered).

Factor 1: Distances to Home

Our first factor involves how far the runner and outfielder are from home, both taken at the time the ball is fielded. We decided on this time as it should be around when a third base coach will start making their decision. The distance between the runner and home (denoted as **run_dist**) is a straight line distance between their location and the next base, plus 90 if that next base is third, and 180 if it's second. The distance between the outfielder and home (denoted as **OF_dist**) is just a straight line from the outfielder's location to home plate.

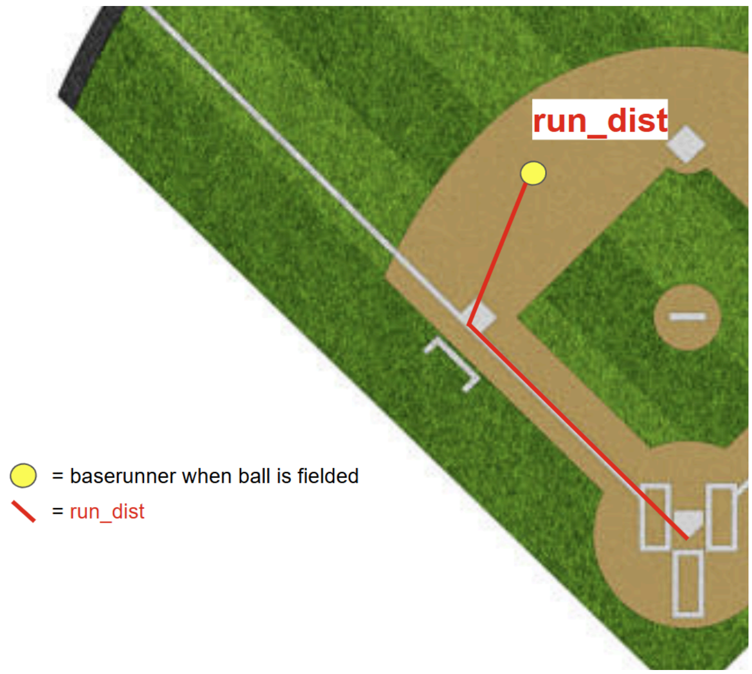

Figure 2: Visualization of run_dist

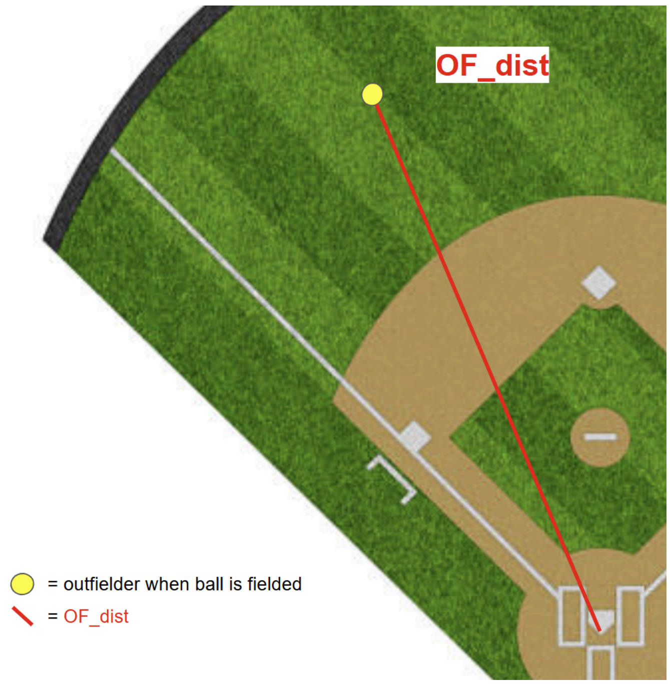

Figure 3: Visualization of OF_dist

Factor 2: Base Running Speed

The second factor was the speed of the baserunner. We decided to use two different speed variables. Firstly, we took the speed of the runner as the ball was fielded (denoted as **run_speed**). We calculated this by recording where the baserunner was ~0.5 seconds before the ball was fielded and when the ball was fielded, and getting speed over that time in ft/sec. Then we calculated the top speed of the runner (denoted as **top_speed**) by finding the most distance the runner covered during the play over ~0.5 second period, and converting that into ft/sec.

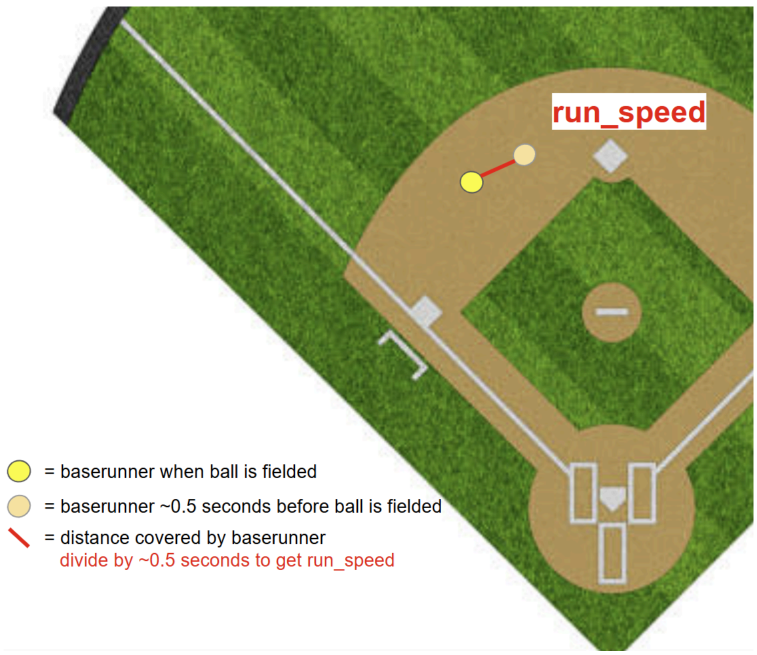

Figure 4: Visualization of run_speed

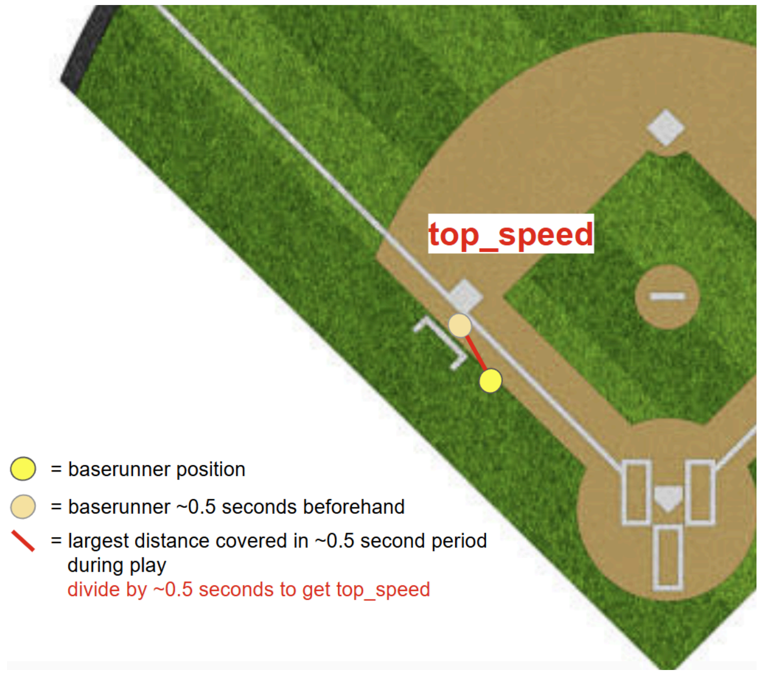

Figure 5: Visualization of top_speed

Factor 3: Outfielder Momentum

Our final main factor was the momentum of the outfielder. We feel like this is something that isn't discussed nearly enough with outfielder throwing. An outfielder can throw a ball much harder if they're getting a good run into it compared to being flat-footed. They'll also lose accuracy if they have to make a throw on the run to the side, or will lose time having to stop before throwing the ball. To calculate the momentum stats, we take where the outfielder was when they fielded the ball and ~0.5 seconds beforehand, then make a vector of the outfielder's speed over that time period. And to calculate the momentum towards home (denoted as **OF_momentum_home**) we project that speed vector onto a vector between where the outfielder was when they fielded the ball and home plate (essentially OF_dist) and that shows how much of their speed was directed towards home. Then for side momentum (denoted as **OF_momentum_side**) we project the speed vector onto a vector perpendicular to OF_dist. Home momentum is positive when they're going towards home, negative if they're going away from home. Side momentum is always positive, no matter if they're going left or right.

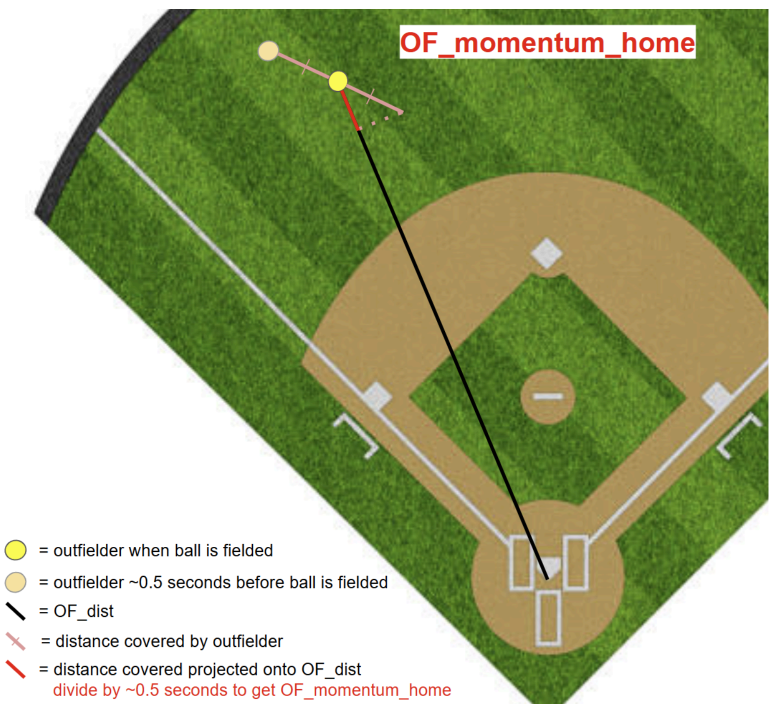

Figure 6: Visualization of OF_momentum_home

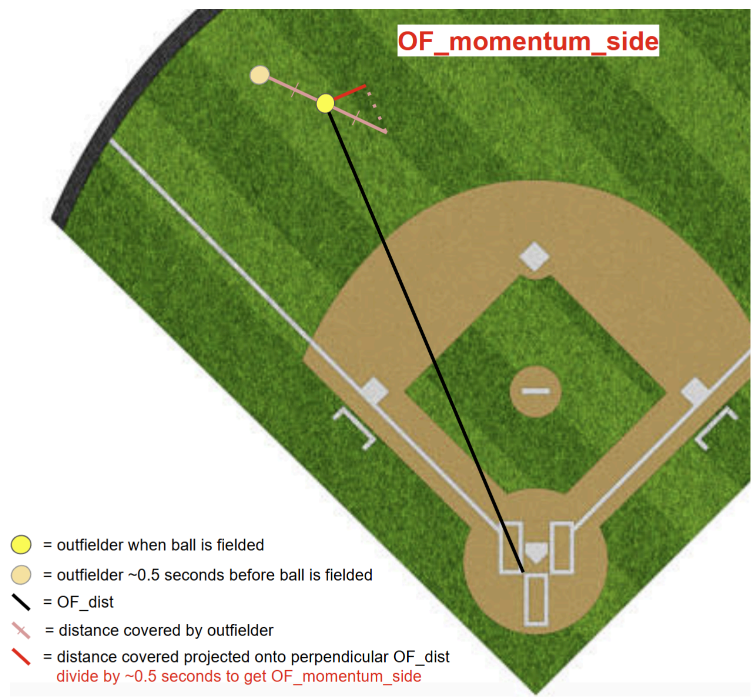

Figure 7: Visualization of OF_momentum_side

Safe, Out, or Didn't Go?

Now we needed to determine the actual outcome of all the plays. Since the outcomes of plays were not given, the best option was to watch animations of each play to make our best guess as to whether runners were safe or out. To help with this, for each play we found what was the closest the runner got to home plate and at what time, and if/when the catcher acquired the ball. This narrowed down the amount of plays we had to watch dramatically. From there the plays were watched and we predicted whether the runner was safe or out. We did this on a scale of 0 to 1, with 0 being definitely out and 1 being definitely safe.

Modeling

Then we needed to find probabilities of whether a player would be safe or out based on our factors. Ultimately, a logistic regression model was decided on after numerous tests of various model types (see Appendix B on pages 10 & 11 for more on the model testing). Here were the results of how important each variable was.

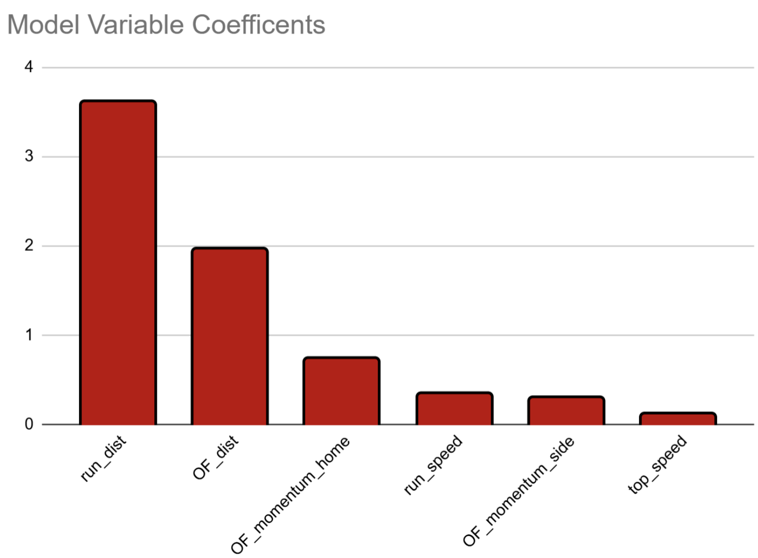

Figure 8: Coefficient values for each variable in the model predicting the probability of a player being safe or out

The coefficients show how important each variable is for predicting if a runner will be safe or out. (see Appendix C on pages 11 & 12 for some more interpretable results of how important each variable is). As expected, how far the runner and outfielder are from home were most important. We were a bit surprised to see how low the runner speed variables were, we thought they would be a bit more important.

Then since the model was trained only on plays where the runner went home, we use the model on the plays where the runner did not go home. This now gave us a safe probability for every play.

Run Expectancy

One more key factor is needed for determining if a runner should be sent, which is run expectancy. A run expectancy table shows the average number of runs a team would be expected to score for the remainder of an inning based on the baserunners and the number of outs. The intersection of a row and column tells you the run expectancy for that given baserunner situation and number of outs.

2021 MiLB Run Expectancy		OUTS		
		0	1	2
BASERUNNERS	___	0.304	0.118	0.005
	1__	0.720	0.365	0.113
	_2_	0.809	0.425	0.176
	__3	0.922	0.463	0.184
	12_	1.329	0.765	0.454
	1_3	1.490	0.874	0.542
	_23	1.650	0.924	0.634
	123	2.184	1.363	0.858

Figure 9: Run Expectancy table from all MiLB levels during the 2021 season

The MiLB level of the data wasn't specified, and it was said the data came before the big rule changes like shift restrictions and larger bases. So taking it from all MiLB levels from a year like 2021 seemed like the best approach.

The run expectancy for each potential outcome of a play is crucial for making a send or go decision, as we can calculate the safe probability needed for it to be a good decision to go. This probability was calculated for every play and was denoted as **prob_to_go**.

prob_to_go

p = safe probability
RE_S = run expectancy if the runner goes and is **safe**
RE_O = run expectancy if the runner goes and is **out**
RE_T = run expectancy if the runner **stays at third (doesn't go)**

Run expectancy of going = Run expectancy of staying

$$RE_S(p) + RE_O(1 - p) = RE_T$$

$$\ldots$$

$$p = (RE_T - RE_O) / (RE_S - RE_O)$$

Figure 10: How prob_to_go (p) is calculated

So essentially, we're calculating what safe probability is needed for the run expectancy if the runner goes to exceed the run expectancy if the runner stays at third. If the safe probability is greater than prob_to_go, then the run expectancy of going will exceed the run expectancy of staying and the runner should be sent (see Appendix D on page 12 for an example play for prob_to_go).

Obtaining the baserunner situations for all the plays was fairly simple. Outs however, were not given with the data, and trying to make an accurate prediction of the outs during a play would have taken too much time. So instead, we assumed that if a play occurred in the first third of a half inning, there were no outs, middle third 1 out, and final third 2 outs.

Results

So using the safe probability and prob_to_go of all the plays, we can determine what would be the correct send or stay decision according to the model. We can then compare these to the actual decisions made by third base coaches in a confusion matrix.

Stay or Go Confusion Matrix		ACTUAL	
		Stayed	Went Home
MODEL	Should Stay	449	36
	Should Go	56	528

Figure 11: Confusion matrix of decisions the model would have made versus the actual decisions. To help with interpretation, the bottom left corner is saying there were 56 plays where the model says the runner should have gone home, but the third base coach had them stay at third.

So according to the model, third base coaches made the correct decision 91.4% of the time. It also appears that third base coaches weren't quite as aggressive sending runners as the model would be. Third base coaches didn't send 9.5% of runners the model would have sent, while sending 7.4% of runners the model would have had stay at third.

We also measured how good the third base coaches were with run expectancy. By taking the difference between the run expectancy if the runner goes and run expectancy if the runner stays (both seen in Figure 10) for all plays, we can measure the expected runs lost by incorrect decisions. We found third base coaches lost 1.65 runs per 100 decisions made. They lost 1.8 runs per 100 decisions where the model would have sent the runner, and 1.48 runs per 100 decisions where the model would have had the runner stay at third. This again shows the third base coaches were too conservative, losing more runs from plays where the model would have sent the runner, but they had them stay, than the reverse.

RE_go = Run expectancy if the runner goes home
RE_stay = Run expectancy if the runner stays at third
RE_diff = RE_go - RE_stay
should_go =
1: RE_go > RE_stay, model says should go home
0: RE_go < RE_stay, model says should stay at third
went_home =
1: Runner went home
0: Runner stayed at third

RE_go	RE_stay	RE_diff	should_go	went_home
1.0	0.9	0.1	1	1
1.1	0.8	0.3	1	1
0.9	0.8	0.1	1	1
1.2	1.0	0.2	1	1
1.0	**0.8**	**0.2**	**1**	**0**
1.0	0.9	0.1	1	1
0.8	0.7	0.1	1	1
0.9	0.7	0.2	1	1
1.1	0.8	0.3	1	1
1.0	0.9	0.1	1	1

10 plays where the runner should've been sent home
1 incorrect decision to not send the runner, that cost the team 0.2 expected runs (RE_diff)
So they lost 2 runs per 100 decisions where the model would've sent the runner home

Figure 12: Example of how expected runs lost stat works

What Could They Do Better?

So we've already established the third base coaches as a whole were too conservative with sending runners. We then wanted to see what variables third base coaches might be over or under valuing. So two new logistic regression models were created, both using the six variables used for the safe or out model, plus prob_to_go. One was trained on the decisions the model would make, and the other on the actual decisions made. We then compared the coefficient values of both models. Here were the results (note a positive coefficient means a larger value = more likely to be safe, negative means a smaller value = more likely to be safe, zero = no effect).

COEFFICIENTS	run_dist	OF_dist	run_speed	top_speed	OF_momentum_home	OF_momentum_side	prob_to_go
ACTUAL	-2.40	1.01	0.15	0.42	-0.20	0.10	0.05
MODEL	-2.44	0.92	0.25	0.06	-0.38	0.14	-0.26

Figure 13: A comparison between variable coefficient values of models based on models decisions and actual decisions

What we noticed first is third base coaches barely seemed to take prob_to_go into account. In fact, the coefficient is going the wrong direction (larger prob_to_go = larger safe probability needed to send the runner = less likely the runner should be sent, so the coefficient should be negative like with the model). Runner speed as a whole seems to be valued more by third base coaches, especially top_speed. Meanwhile the model values run_speed more. The third base coaches also seem to be undervaluing OF_momentum_home.

So all in all, our findings are that these third base coaches were not aggressive enough. While also overvaluing a runners top speed, and undervaluing their speed at the time the ball is fielded, the outfielders momentum towards home, and the game state (see Appendix E on pages 13 & 14 for this analysis for individual team third base coaches).

Conclusions

The obvious use for analysis like this to help improve the decision making of third base coaching. Whether that is telling them to be a bit more aggressive/conservative, or to start valuing certain variables more or less. Perhaps even having them practice on simulated situations to hone their decision making. This would only get better with more plays and variables that could not be obtained with the data and time we had for this project.

It would also be very interesting to potentially work more with the runs lost stat we used to create a WAR stat for third base coaches. It could maybe be used to grade, improve, or even scout third base coaches for potential hiring by an organization.

References

Fryer, B. (2022, September 8). *Should I Stay or Should I Go.* https://billyfryer.com/projects/should-i-stay-or-should-i-go/

Appendix A: Other Potential Factors

Obviously these are not all the factors that come into play with deciding whether to send a runner home or not. Having the actual outs in an inning, the score of the game, what inning the play occurred in, where in the batting order the team was, and how strong and accurate the fielder's arm is, etc. would be very useful. But, we thought that the variables we did choose were the best to use with the time and data we had. Getting the actual outs in an inning would obviously be better than the assumption we made for this project. The score of the game and inning would obviously be useful for calculating win probability, which could be used in place of run expectancy for deciding whether to send a runner or not. Where in the order the team is could also factor in with the run expectancy / win probability as you'd be less likely to send a runner with your cleanup hitter coming up with 0 outs than your nine hitter with 2 outs. Finally, knowing the arm strength and the accuracy of the outfielder and/or cutoff men would be a game changer as knowing if they had the arm of Ronald Acuna Jr compared to a player like Ben Revere could make a significant difference. We could have done better with the run_dist by trying to calculate how far they'd actually run with the curvature of their path. But, that would have taken a lot of work and likely would have made a minimal difference. We also thought of possibly taking into account cut off men, but could not think of a good way to do so.

Appendix B: Modeling

For modeling the probability of whether a runner would be safe or out, six different methods were tested. Those being logistic regression, XGBoost, random forest, naive bayes, LDA, and QDA. To test the accuracy of the models, 100 simulations of k-fold cross validation with log loss as the accuracy measure was used. K-fold cross validation is a model testing method where a dataset is split into k evenly sized folds (subsets of the original dataset), a model is trained on k - 1 of those folds, then tested on the one fold not used in training. Then this is repeated k times until every fold is tested on. For this case, the folds were stratified so they each contained roughly the same distribution of safe and out plays. A k of 10 was used for this testing, as it gave a good balance of training and testing data. Simple right or wrong accuracy wasn't used since a model could predict every play to be safe since there are so few out plays and still have a good accuracy. So log loss was used since it will punish confident wrong guesses, which would punish a model that just predicts everyone will be safe. Then for models like XGBoost, some testing was needed to find what values of certain parameters would give the best results. Same with logistic regression seeing if interactions between certain variables or curved values for variables made the models more accurate. The best of each model type was chosen and these were the results (a lower log loss is better).

SAFE OR OUT MODEL ACCURACIES	
MODEL	LOG LOSS
Logistic	0.156
XGBoost	0.168
LDA	0.168
Random Forest	0.177
Naive Bayes	0.206
QDA	0.262

So ultimately, a logistic regression model came out on top, just edging out XGBoost and LDA. The logistic regression model used was fully linear for all variables and no variable interactions were used. There actually were a couple logistic regression models that had slightly better log losses that used some curves and interaction variables. But, upon using them and checking some of the probabilities, there were a few plays where the probability was very clearly wrong and we decided against using them. The simple logistic regression model had the best combination of accuracy in the tests and having seemingly accurate results for the plays where runners didn't go.

I would also like to note that for the modeling we originally tried just using plays where we were fairly confident the runner was out as out plays. However, after some testing it became clear there were not enough out datapoints for the model to be as accurate as we wanted. Most notably, we noticed with a lot of the models that the runners speed wasn't being taken into account. For example, there were two plays where run_dist, OF_dist, and the OF_momentum variables were nearly identical, the only difference being in one play the runner was significantly faster than the runner in the other play. But, the safe probabilities were coming out nearly identical, which was clearly not correct. So we used every play we found to be a "bang-bang" play at the plate as an out, and that seemed to give us enough data points to fix the speed issue. We also thought this could help with an issue with this modeling; Extrapolating the model to plays where runners did not go, where the imbalance of safe and out plays will be flipped compared to plays where the runners did go that we trained with.

Appendix C: Interpretation of Model Coefficients

For every **1 foot further the runner is from home**, the odds of scoring **decrease by 11.9%**

For every **1 foot further the outfielder is from home**, the odds of scoring **increase by 3.9%**

For every **1 ft/sec faster the runner is going when the ball is fielded**, the odds of scoring **increase by 13.6%**

For every **1 ft/sec faster the runners top speed is**, the odds of scoring **increase by 5.5%**

For every **1 ft/sec more of momentum towards home the outfielder has**, the odds of scoring **decrease by 6.5%**

For every **1 ft/sec more of momentum to the side the outfielder has**, the odds of scoring **increase by 5.5%**

Probability = Odds / (1 + Odds)

Ex.
Odds = 1,
Probability = 1 / (1 + 1) = 1 / 2 = 0.5

Runner one foot further from home, odds decrease by 11.9%
New Odds = 1 * (1 - 0.119) = 1 * 0.881 = 0.881
New Probability = 0.881 / (1 + 0.881) = 0.468

So on a 50/50 safe or out play…
If the **runner was 1 foot further from home**, the probability of scoring would **decrease to 46.8%**
If the **outfielder was 1 foot further from home**, the probability of scoring would **increase to 51%**
If the **runner was 1 ft/sec faster when the ball is fielded**, the probability of scoring would **increase to 53.2%**
If the **runners top speed was 1 ft/sec faster**, the probability of scoring would **increase to 51.3%**
If the **outfielder had 1 ft/sec more of momentum towards home**, the probability of scoring would **decrease to 48.3%**
If the **outfielder had 1 ft/sec more of momentum to the side**, the probability of scoring would **increase to 51.3%**

Appendix D: Probability To Go

An example of how prob_to_go should help here if there's any confusion. Let's take our introduction example, and say there are runners on first and third with 1 out, and then a double is hit. So if the runner from first goes and scores, we have a run scored and a runner on third (batter advanced on the throw home) with 1 out, we have a run expectancy of 1.463 (the 1 run already scored plus the 0.463 expected runs from a runner on third with 1 out, we don't need to take the runner who scored from third into account at all because it happens in all situations and will just cancel out in the equation). If the runner goes and is out at the plate, we have a runner on third with 2 outs, for a run expectancy of 0.184. And if the runner stays at third, we have runners on second and third with 1 out, and a run expectancy of 0.924. Then we do the math from Figure 10:

(0.924 - 0.184) / (1.463 - 0.176) = 0.740 / 1.287 = **0.575 = prob_to_go**

So this tells us that the runner would need a 57.5% chance of being safe for the run expectancy if they are sent home to equal the run expectancy if they stay at third. Anything more than 57.5%, the run expectancy of going exceeds the run expectancy of staying at third, so they should be sent.

Appendix E: Individual Team Third Base Coaches

The main teams in the dataset were denoted as QEA, RZQ, and YJD. So we went through the same analysis done for third base coaches as a whole on these individual teams.

QEA Confusion Matrix		ACTUAL: Stayed	ACTUAL: Went Home
MODEL	Should Stay	20	5
	Should Go	6	40

Made correct decision 84.5% of the time (-6.9%)
Didn't send 13% of players the model would have sent (+4.5%)
Sent 20% of players the model would not have sent (+12.6%)

Lost 1.54 runs per 100 decisions (-0.11)
Lost 1.57 runs per 100 runner should have went decisions (-0.23)
Lost 1.5 runs per 100 runner should have stayed decisions (+0.02)

COEFFICIENTS	run_dist	OF_dist	run_speed	top_speed	OF_momentum_home	OF_momentum_side	prob_to_go
QEA	-2.27	0.94	0.18	0.33	-0.60	-0.27	0.20
MODEL	-2.44	0.92	0.25	0.06	-0.38	0.14	-0.26

The numbers next to the stats show comparison to total stats from the Result section. So for example, the QEA third base coach's correct decision percentage of 84.5% was 6.9% worse than the 91.4% for all third base coaches. Just looking at percentages you'd think this is not the best third base coach, but the expected runs lost actually has him as above average. After looking at their incorrect decisions, this third base coach just had a lot of really close calls. They had a lot of plays where they made the incorrect choice, but it would only be by 0.02 expected runs.

So from this analysis, we can see the QEA third base coach was properly aggressive, losing a similar amount of runs per 100 decisions on should have gone and should have stayed decisions. From the coefficients, we can see they overvalued the runners top speed and outfielders momentum towards home, while wrongly valuing the outfielders momentum to the side and game state.

RZQ Confusion Matrix		ACTUAL: Stayed	ACTUAL: Went Home
MODEL	Should Stay	95	6
	Should Go	9	107

Made correct decision 93.1% of the time (+1.7%)
Didn't send 7.8% of players the model would have sent (-1.7%)
Sent 5.9% of players the model would not have sent (-1.5%)

Lost 1.48 runs per 100 decisions (-0.17)
Lost 1.81 runs per 100 runner should have went decisions (+0.01)
Lost 1.09 runs per 100 runner should have stayed decisions (-0.39)

COEFFICIENTS	run_dist	OF_dist	run_speed	top_speed	OF_momentum_home	OF_momentum_side	prob_to_go
RZQ	-2.41	1.01	0.12	0.38	0.04	0.07	0.14
MODEL	-2.44	0.92	0.25	0.06	-0.38	0.14	-0.26

The RZQ third base coach was pretty good, he had the best correct decision percentage and lost runs per 100 decisions of all three third base coaches looked at here. But, they were way too conservative looking at that difference in lost runs between should have gone and should have stayed decisions. The only spot they were below average is with the lost runs on should have gone decisions (so holding runners at third that should have been sent home). From the coefficients, we see they overvalue a runners top speed, while undervaluing the runners speed at the time the ball is fielded and the outfielders momentum towards home. They also wrongly value the game state like QEA did.

YJD Confusion Matrix		ACTUAL	
		Stayed	Went Home
MODEL	Should Stay	77	6
	Should Go	12	97

Made correct decision 90.6% of the time (-0.8%)
Didn't send 11% of players the model would have sent (+1.5%)
Sent 7.2% of players the model would not have sent (-0.2%)

Lost 1.51 runs per 100 decisions (-0.14)
Lost 1.92 runs per 100 runner should have went decisions (+0.12)
Lost 0.96 runs per 100 runner should have stayed decisions (-0.52)

COEFFICIENTS	run_dist	OF_dist	run_speed	top_speed	OF_momentum_home	OF_momentum_side	prob_to_go
YJD	-2.28	1.20	0.51	0.14	0.03	0.02	-0.02
MODEL	-2.44	0.92	0.25	0.06	-0.38	0.14	-0.26

Like QEA, the YJD third base coach was below average by correct decision percentage but above average by runs lost. And like RZQ, YJD was way too conservative, even more so in fact. They lost double the expected runs per 100 decisions on plays where the model said the runner should have gone than when they should have stayed. From the coefficients, we can see they are the first third base coach to value the runners speed when the ball is fielded over their top speed as the model suggests they should. But, they still overvalue speed as a whole and how far the outfielder is from home, while undervaluing the outfielder's momentum and the game state.

Appendix F: Limitations

Obviously there are plenty of limitations to what was presented in this project. There are so many factors that determine whether a runner will be safe or out at home, and we could not account for all of them. The limited amount of plays where a runner was probably out at home for the model to train on, then having to use it on plays where the runner stayed at third could've led to some inaccuracies in calculating safe probability. We had to make assumptions with the outs in the game state, and while it should be accurate most of the time, it is not going to be correct for all of the plays. But, overall we think this project is comprehensive and shows the potential of work like this in improving third base coaches decision making.