How effectively do baserunners judge when to be aggressive?

Introduction:

Baserunning decisions on contact are one of the few moments in baseball where a player must make an immediate, irreversible choice with meaningful run-value consequences. Upon contact, a runner must instantly decide whether to be aggressive - attempting to take an extra base - or to hold. This project aims to evaluate how well runners make these on-contact baserunning decisions and to quantify the run value gained or lost relative to typical baserunning behavior.

Rather than focusing solely on outcomes (safe versus out), the analysis explicitly treats each baserunning opportunity as a decision problem with probabilistic consequences, allowing the quality of the decision itself to be evaluated independently of its realized result.

Background:

Publicly available baserunning metrics, most notably Statcast's Extra Bases Taken Run Value, model the probability that an aggressive attempt will succeed based on factors including runner speed, outfielder arm strength, and ball location. The runner's actual outcome is then compared to this expected success probability, and cumulative run value is assigned accordingly.

While this framework provides valuable insight into baserunning performance, it is fundamentally outcome-focused. In addition, Extra Bases Taken Run Value only evaluates plays in which the runner attempts to advance aggressively. As a result, it does not account for:

1. baserunning opportunities where the runner chooses not to be aggressive, or
2. whether the decision to be aggressive or conservative was itself optimal or atypical in that context.

Data and Methodology:

Using MLB Statcast pitch-level data from the 2022–2025 seasons, this study focuses on six common on-contact baserunning situations. These are grouped into three core situations, each of which can result in either an aggressive or non-aggressive decision:

<u>Aggressive events:</u>

- First to third on a single
- First to home on a single or double
- Second to home on a single

<u>Non-aggressive events, mapped to the same situations for modeling purposes:</u>

- First to second on a single → treated as first to third
- First to third on a double → treated as first to home
- Second to third on a single → treated as second to home

This mapping ensures that aggressive and non-aggressive decisions are evaluated within a consistent situational framework.

For each baserunning opportunity, we define an Average Decision Value, representing the expected run value of typical baserunning behavior in that context. We then compare the runner's actual decision to this baseline to measure how much run value was gained or lost due to the decision itself.

To construct this framework, each baserunning opportunity is decomposed into three components:

1. the probability that a runner chooses to be aggressive - P(aggressive),
2. the probability that a runner would be safe if aggressive - P(safe | aggressive),
3. the run value associated with each possible outcome:

a. aggressive & safe - $RV_{agg, safe}$
b. aggressive & out - $RV_{agg, out}$
c. Hold - RV_{hold}

Decision Value above Average:
To evaluate baserunning decision quality, each on-contact baserunning opportunity is framed as a decision. The runner must choose whether to be aggressive or hold, with each choice carrying probabilistic outcomes and associated run values.

Average Decision Value:
The Average Decision Value represents the expected run value of a baserunning opportunity before the runner makes a choice. It quantifies the run-value context of the situation itself, based on how runners typically behave in similar plays.

Specifically, it combines two components of average behavior:
1. How often runners are aggressive in this situation
2. The expected run value of each possible outcome

$$P(aggressive)*[P(safe \mid aggressive)*RV_{agg, safe} + (1-P(safe \mid aggressive))*RV_{agg, out}]$$
$$+$$
$$(1-P(aggressive))*RV_{hold}$$

Actual Decision Value:
The Actual Decision Value reflects the run value of the choice the runner actually made:
- If the runner was aggressive, the value is the expected run value of an aggressive attempt:
$$P(safe \mid aggressive)*RV_{agg, safe} + (1-P(safe \mid aggressive))*RV_{agg, out}$$
- If the runner was not aggressive, the value is simply:
$$RV_{hold}$$

Decision Value above Average:
Finally, Decision Value Above Average is defined as:
$$\text{Decision Value Above Average} = \text{Actual Decision Value} - \text{Average Decision Value}$$

This metric measures how much run value a runner's decision added or lost relative to typical baserunning behavior in that same situation.
- A positive value indicates the runner made a better-than-average decision.
- A negative value indicates the runner deviated from typical behavior in a way that reduced expected run value.

Modeling:
To estimate the components of the decision framework, we model (i) the probability that a runner chooses to be aggressive and (ii) the probability that a runner would be safe if aggressive.

We use XGBoost binary classifiers for both tasks. XGBoost is well suited to this setting because it captures nonlinear effects and interactions among batted-ball geometry, runner speed, and defensive positioning.

Run values for each possible outcome are derived from RE24 base-out values sourced from FanGraphs.

Aggression Model:
The aggression model predicts the probability that a runner chooses to attempt the extra base on a given batted-ball event, P(aggressive).

A single global model is used across all situations, with *situation* included as a categorical feature. This allows the model to learn both shared decision logic (e.g., faster runners are more aggressive) and situation-specific thresholds (e.g., second-to-home vs. first-to-third).

Two visualizations illustrate how the model behaves:

1. Batted-ball location heatmap (Figure 1):
   Focusing on first-to-third opportunities, aggression probability is highest on balls hit to right field, where the longer throw to third base reduces defensive leverage. Shallow balls and balls hit toward left field correspond to lower predicted aggression due to shorter, more direct throws.

2. Partial dependence on sprint speed (Figure 2):
   The partial dependence plot demonstrates a strong monotonic relationship between sprint speed and aggression probability. Holding all other features constant, faster runners are substantially more likely to attempt the extra base, confirming that the model captures intuitive baserunning behavior rather than spurious correlations.

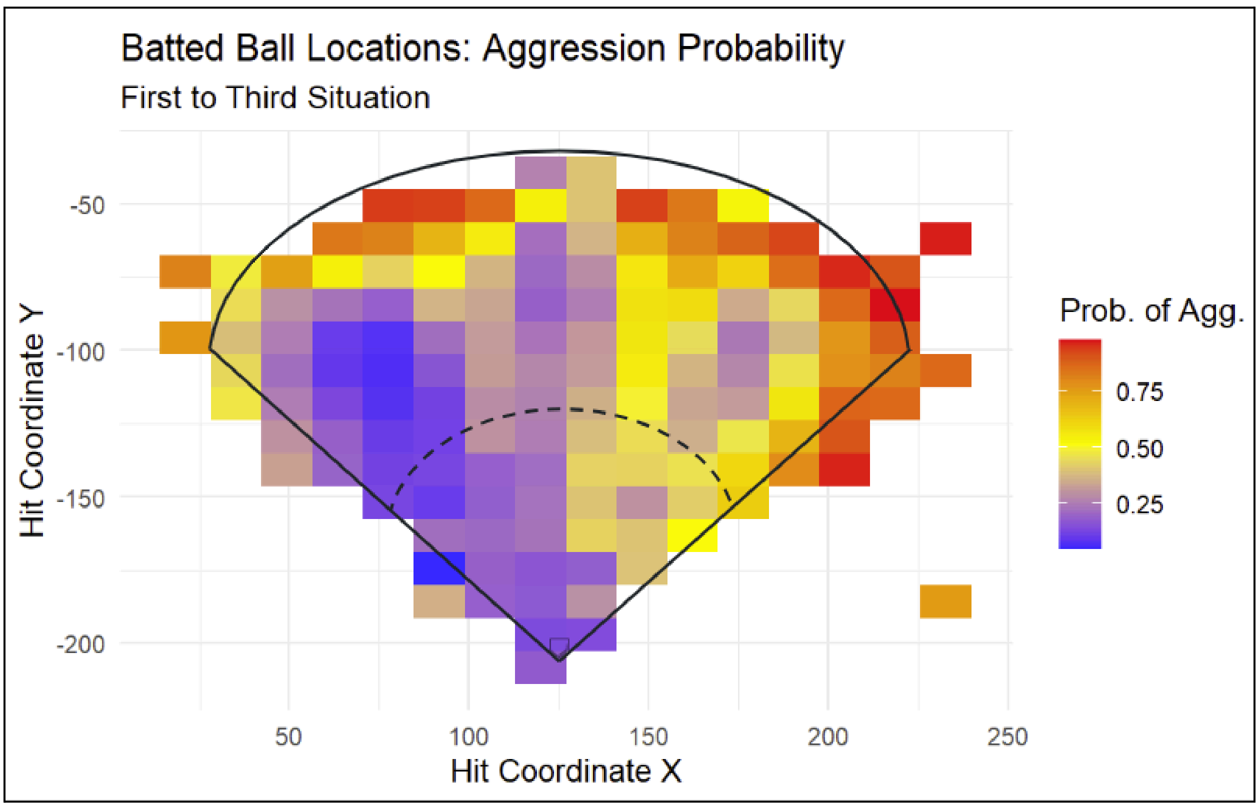

Figure 1: Predicted aggression probability in first-to-third situations is highest for balls hit deep to right field.

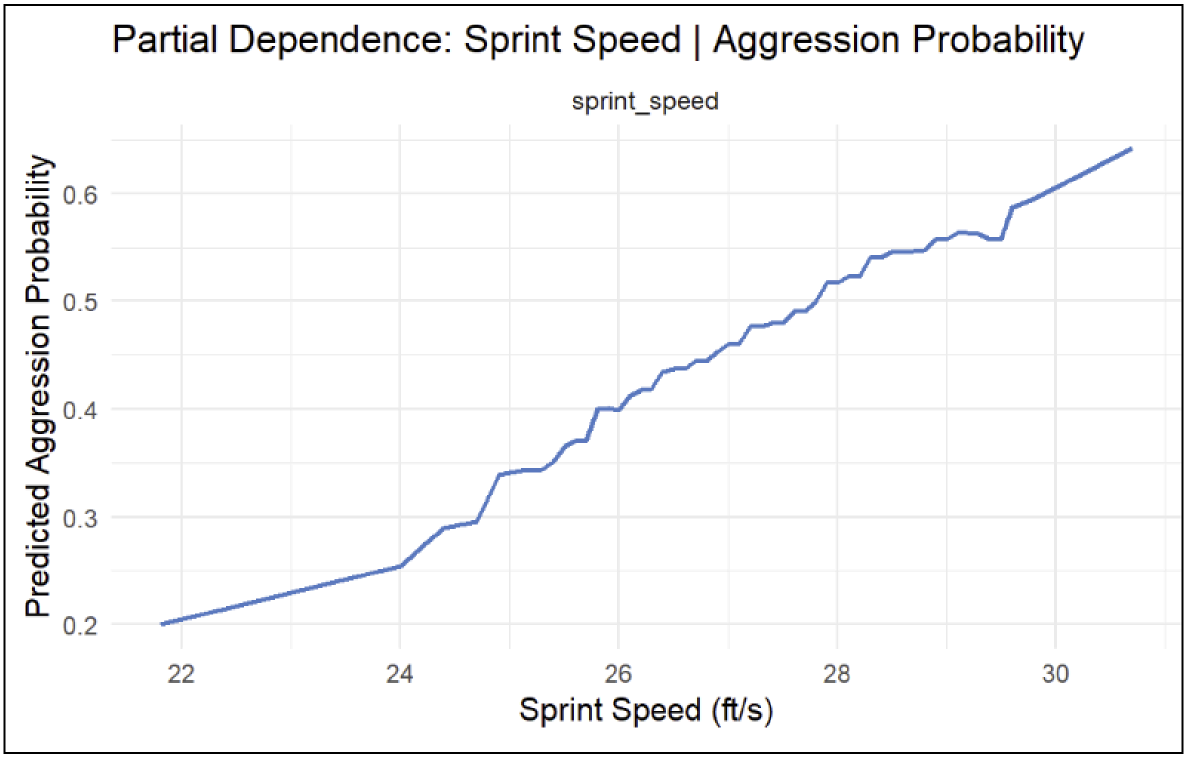

Figure 2: Aggression probability increases monotonically with runner sprint speed.

<u>Safety Model:</u>
The safety model estimates the probability that a runner would be safe conditional on choosing to be aggressive, P(safe | aggressive).

Separate safety models are trained for each baserunning situation to account for the fundamentally different geometry and throw dynamics across plays (e.g., first-to-third vs. second-to-home). Each model is trained only on aggressive events, since we want to compute the probability a runner is safe **given** he attempts to take the extra base.

A key challenge in this framework is that non-aggressive events often lie in regions of the feature space where aggressive attempts are uncommon. When a safety model trained only on aggressive plays is applied to these situations, it can generate unrealistically high success probabilities because it is extrapolating beyond the data it was trained on.

To address this, the predicted aggression probability from the first model is included as a covariate in the safety model. Runners are more likely to attempt aggressive advances when the chance of success is high and tend to hold when the risk of being thrown out is greater. Conditioning on aggression probability helps the safety model account for this selection effect and produces more realistic success estimates for non-aggressive opportunities (Figure 3).

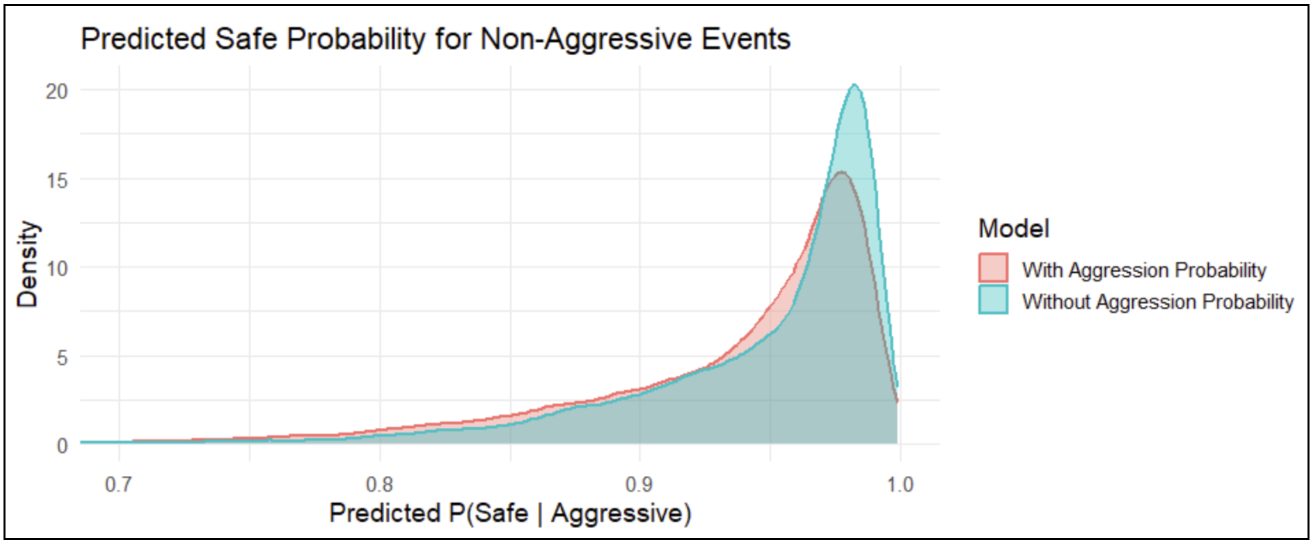

Figure 3: Including aggression probability as a covariate shifts predicted safety probabilities for non-aggressive *events away from unrealistically near-certain success and toward a more realistic, better-calibrated distribution.*

Results:

	runner_name	opportunities	agg_events	sum_dec_val_above_avg
1	Profar, Jurickson	131	76	4.2083840
2	Andrus, Elvis	54	42	4.1261383
3	McMahon, Ryan	125	61	3.5892531
4	Santana, Carlos	88	48	3.5271592
5	Betts, Mookie	166	85	3.4398231
6	Devers, Rafael	123	60	3.3720698
7	Albies, Ozzie	90	60	3.3182273
8	Arenado, Nolan	108	43	3.2028693
9	Perdomo, Geraldo	100	63	2.9371133
10	Henderson, Gunnar	115	70	2.9336024
11	Olson, Matt	122	56	2.7705617
12	Margot, Manuel	53	39	2.6956266
13	Altuve, Jose	137	61	2.6049096
14	Goldschmidt, Paul	146	71	2.5376677
15	Tovar, Ezequiel	78	51	2.5205390
16	Kemp, Tony	56	31	2.4764140
17	Winn, Masyn	74	48	2.4678214
18	Walker, Christian	106	55	2.4463647
19	Frazier, Adam	103	54	2.4337359
20	Marte, Starling	66	37	2.3747007
21	Ramírez, José	102	70	2.3429568
22	Hernández, Enrique	77	44	2.3382070
23	Pederson, Joc	65	31	2.3104206
24	Tatis Jr., Fernando	77	45	2.2848151
25	McKinstry, Zach	63	43	2.2471826
26	Chisholm Jr., Jazz	75	48	2.2128420
27	Yelich, Christian	111	68	2.2050095
28	Torres, Gleyber	112	55	2.2012184
29	Isbel, Kyle	83	58	2.1810504
30	Melendez, MJ	62	41	2.1609582

Table 1: Decision Value Above Average Leaderboard

Table 1 presents a leaderboard of runners ranked by cumulative Decision Value Above Average, aggregated across all qualifying on-contact baserunning opportunities.

Interpreting and validating decision-based baserunning metrics is inherently challenging. Unlike outcome-based metrics, which can be evaluated against observable results (e.g., safe vs. out), this framework assigns value to the decision itself relative to typical behavior in the same context. As a result, there is no direct "ground truth" against which individual decisions can be conclusively verified.

Nevertheless, several names near the top of the leaderboard align well with intuition. Jurickson Profar ranks first, combining a high number of opportunities with strong cumulative decision value. Profar has long been regarded as an instinctive baserunner, and his placement here reflects frequent, well-timed aggression rather than simply elite speed.

Elvis Andrus provides an especially illustrative case. Andrus recorded 42 aggressive attempts across just 54 opportunities, one of the highest aggression rates in the sample. Despite this high frequency of aggression, his cumulative decision value remains strongly positive, suggesting that his aggressiveness was generally well-aligned with favorable contexts rather than reckless risk-taking.

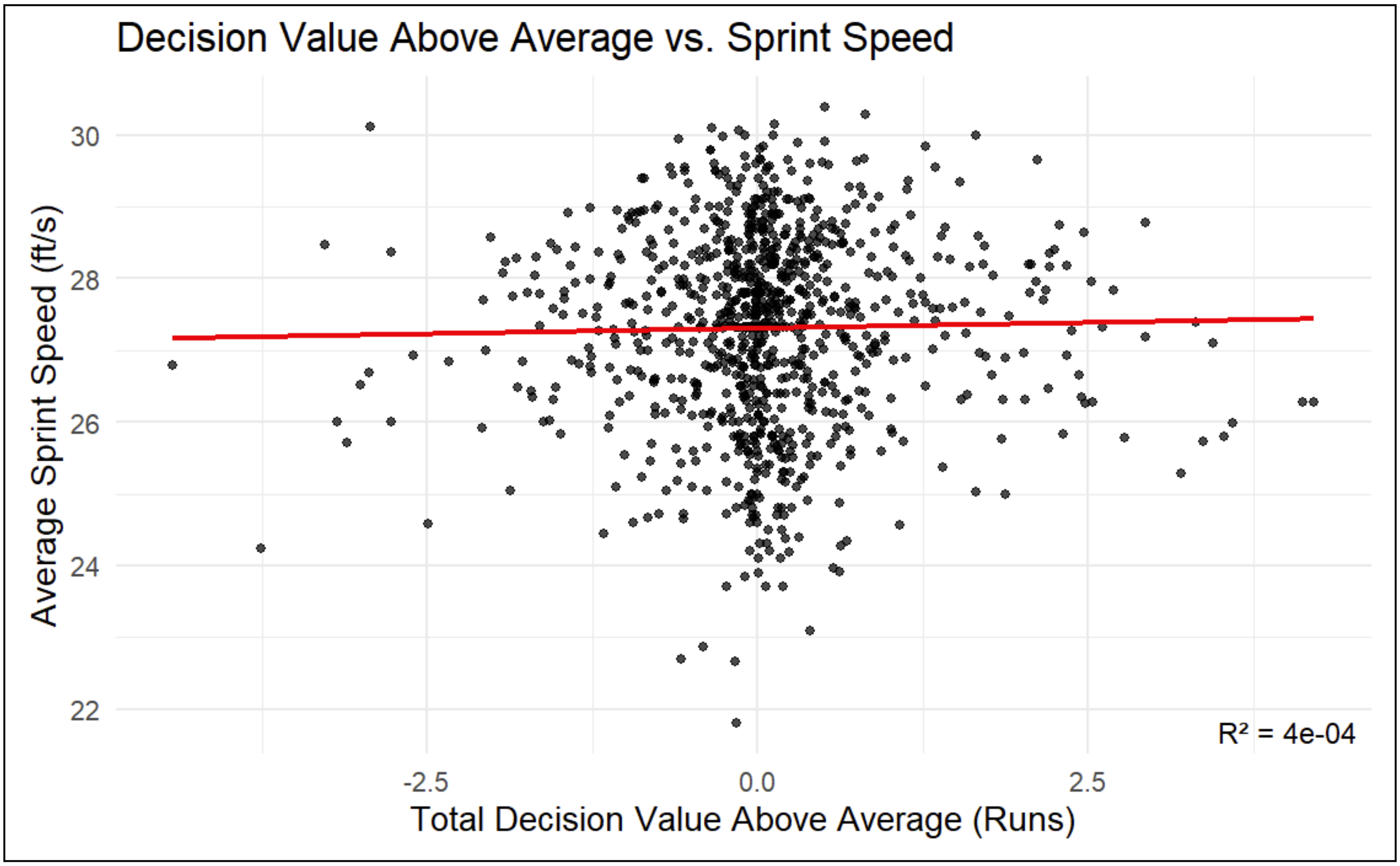

Figure 4: There is no correlation between Sprint Speed and Decision Value Above Average.

Other high-ranking players such as Mookie Betts, Rafael Devers, and Ozzie Albies similarly fit an intuitive profile - runners who combine athletic ability with strong situational awareness. Importantly, the leaderboard is not dominated exclusively by the fastest players, reinforcing that the metric is capturing decision quality rather than raw speed alone (Figure 4).

<u>The Eye Test:</u>
While player-level validation is inherently difficult for a decision-based metric, play-level evaluation provides a more intuitive lens. Here are two contrasting examples:

High Decision-Value Run (+0.38):
https://baseballsavant.mlb.com/sporty-videos?playId=b18ba8c2-41e0-42ab-b60a-bdd583eea660
On this play, the shortstop fails to make the initial stop, sending the ball into right field and forcing the defense into a rushed recovery. The right fielder hesitates and relays the ball casually toward the infield rather than throwing home, allowing the runner to score from first, a case where aggressive baserunning clearly paid off.

Low Decision-Value Run (-0.43):
https://baseballsavant.mlb.com/sporty-videos?playId=fede7f56-5f3b-4d7c-b76b-8d935343d430
In this low-value example, the batted ball is fielded closer to the interior of left field than the runner likely perceived, providing the fielder with a favorable throwing opportunity. The runner attempts to advance to third despite the shallow angle and is thrown out, illustrating an overly aggressive decision relative to the context.

Conclusion and Future Work:

This project introduces a novel framework for evaluating on-contact baserunning by explicitly treating each opportunity as a decision under uncertainty. By combining the probability of aggression, the probability of success conditional on aggression, and run values derived from base-out states, the Decision Value Above Average metric isolates the value of the decision itself.

The results suggest that this approach captures meaningful variation in baserunning behavior that is not explained by speed alone. At the same time, the analysis underscores the inherent difficulty of validating decision-based metrics at an aggregate level, reinforcing the importance of play-level interpretation.

There are several clear avenues for improvement. Incorporating stadium-specific effects would help account for differences in outfield dimensions and throwing environments. Including exact fielder starting positions and movement would better capture defensive leverage at the moment of contact. Explicitly modeling fielding mistakes and incorporating runner lead distance at contact would further refine both the aggression and safety models.

Overall, this work demonstrates that decision-based evaluation offers a complementary perspective to existing outcome-based baserunning metrics and provides a foundation for more granular analysis of baserunning strategy and value.