Swing+ Basics

Swing+ is a bat-tracking metric that measures how efficient and effective a hitter's swing mechanics are, essentially how well a swing produces optimal contact independent of raw strength or exit velocity. It captures the quality of the swing itself rather than the result of the contact. The metric evaluates the mechanical traits that allow a player to consistently generate ideal bat-ball matchups and contact quality, serving as a benchmark for swing efficiency.

What Swing+ Measures

At its core, Swing+ evaluates the biomechanical and kinematic characteristics of a hitter's swing. The model uses inputs such as average bat speed, attack angle, swing tilt, swing length, and attack direction to estimate how well a swing aligns with the ideal movement pattern for solid contact. High Swing+ values indicate swings that efficiently translate bat speed and motion into squared-up contact, while lower values suggest inefficiencies in movement sequencing or barrel control that may limit a hitter's ability to optimize their contact point.

Core Concept

Swing+ operates on the principle that not all swings with the same bat speed produce the same results. Some hitters extract more effective contact from their motion due to better path alignment, plane control, and direction. The model isolates these underlying mechanical traits to quantify swing efficiency, or how well a player's movement pattern supports consistent, high-quality contact across pitch locations and types. In other words, Swing+ distinguishes between players who swing hard and those who swing well.

Model Logic

The Swing+ model is trained on a dataset linking swing parameters to expected offensive outcomes, typically expected weighted on-base average (xwOBA). A gradient boosting or LightGBM regression learns which mechanical traits are most predictive of quality contact. After prediction, the model's output is scaled to a familiar index where **100 represents league-average swing quality, and each 10 points represents roughly one standard deviation from that mean.** For example, a Swing+ of 110 indicates a swing pattern one standard deviation more efficient than average, while 90 indicates one standard deviation less efficient. This scaling makes Swing+ directly comparable across players.

Swing+ Nitty Gritty & Model Creation

Model Framework and Objective

The Swing+ model is a supervised learning system built to predict a hitter's expected offensive performance (proxied by expected wOBA) using swing-level bat-tracking variables. The central goal was to isolate *swing quality* from *batted-ball results* and raw power, creating a standardized measure of swing efficiency that can be compared across players. The model was trained on a large dataset of swing characteristics, each labeled with its corresponding expected wOBA outcome based on Statcast contact quality models.

The resulting regression outputs a continuous prediction representing the "expected contribution" of swing mechanics to offensive success. After training, the predictions are normalized to an index scale where 100 equals the league average, mirroring the structure of metrics like wRC+ or Stuff+.

Feature Engineering

The model relies heavily on **bat-tracking inputs** measured during the swing phase, representing the geometry and efficiency of the hitter's movement path. Key features include:

- **avg_bat_speed:** overall swing velocity at contact, which captures a player's ability to generate energy through the kinetic chain.
- **attack_direction and attack_angle:** describe the orientation of the swing path relative to the incoming pitch, showing whether the bat enters the hitting zone on an optimal plane.
- **swing_tilt and swing_length:** measure the vertical and horizontal arc of the swing, balancing power generation with barrel control.
- **batter_x_position and batter_y_position:** capture stance and setup differences that influence reach and plate coverage.
- **intercept_y_vs_batter / intercept_y_vs_plate:** quantify where in 3D space the bat crosses the hitting zone, useful for identifying whether the player's contact point consistently matches ideal timing windows.

These features were normalized and standardized to remove player-level bias (e.g., different swing sensor calibrations or tracking systems). Correlated variables were reduced using variance inflation and feature importance analysis before model fitting.

Model Training Process

The model was trained using LightGBM, a gradient-boosted decision tree algorithm that handles nonlinear relationships and feature interactions efficiently. The target variable, expected wOBA, served as the measure of swing outcome quality. The LightGBM approach allows the model to capture complex interactions, for instance, how bat speed and attack angle interact to create ideal launch conditions — without manually defining these relationships.

Hyperparameters such as tree depth, learning rate, and number of boosting rounds were optimized via cross-validation. Feature importance was later verified using SHAP (SHapley Additive exPlanations) values to ensure interpretability.

Model Validation and Interpretation

Figure 1: Swing+ vs Expected wOBA

The first plot, *Swing+ vs Expected wOBA*, shows the model's predictive strength and validation fit. Each point represents a player's mean Swing+ value versus their expected wOBA, with a clear positive correlation. The red regression line demonstrates that as Swing+ increases, expected wOBA rises accordingly. High-end hitters such as Aaron Judge, Juan Soto, and Ronald Acuña Jr. sit far above the mean, illustrating the model's ability to identify efficient, high-quality swing mechanics. Conversely, hitters like César Prieto and Garrett Stubbs appear at the lower end, reflecting below-average swing quality despite potentially adequate contact skills.

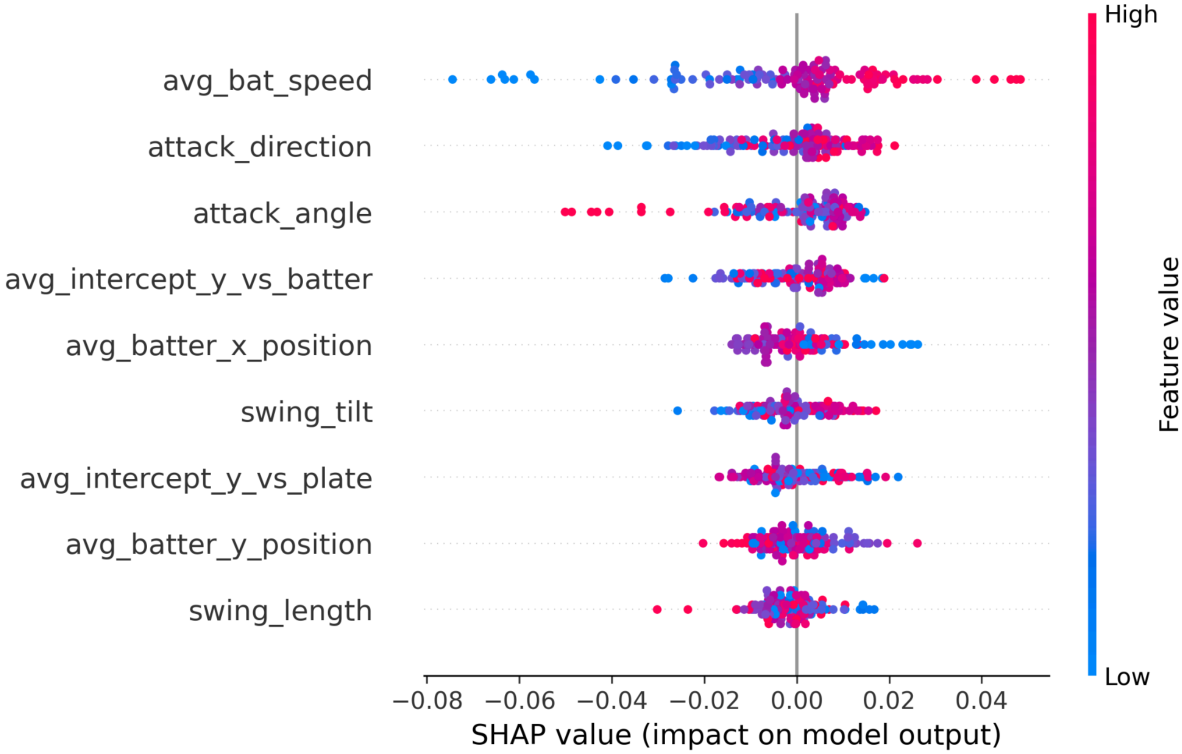

Figure 2: Swing+ SHAP Feature Impact Summary

The second plot, *SHAP Feature Impact Summary*, provides insight into which variables most strongly influence the model's output. avg_bat_speed is the single most impactful predictor, confirming that swing velocity is a fundamental driver of swing efficiency when combined with optimal angles. attack_direction and attack_angle also rank highly, underscoring that swing plane orientation determines whether a hitter's mechanics consistently produce high-quality contact trajectories. Meanwhile, variables such as swing_tilt and swing_length contribute moderate but meaningful adjustments, shaping whether a player's bat path trades contact for power or vice versa.

The color gradient in the SHAP plot (blue = low, red = high) visualizes the directionality of influence: higher bat speeds and more efficient directional planes (red points) push Swing+ values upward, while inefficient movement profiles (blue points) pull them downward.

Scaling and Interpretation

After training, model predictions were transformed into the Swing+ index, with 100 set as league average and each 10-point interval representing roughly one standard deviation. This scaling allows for intuitive comparisons:

- A player with Swing+ = 110 swings one standard deviation better than the average hitter in terms of mechanical efficiency.

- A Swing+ = 90 indicates a swing one standard deviation less efficient than average.

This standardization enables talent evaluators to compare swing mechanics on a common scale across levels, ages, and player types, providing a foundation for deeper models like **ProjSwing+**, which projects future swing-driven performance using additional contextual features (e.g., bat speed growth, power index, and aging curves)

Swing+ Nuance: Cole Young vs. James Wood vs. Roman Anthony

The process of refining the Swing+ model began to evolve once we started testing it across distinctly different swing archetypes. Three young hitters; Cole Young, James Wood, and Roman Anthony—provided a perfect case study. Each possessed elite underlying skill in bat speed and plane control, but the model revealed that they generated their results in fundamentally different mechanical ways.

Bat Speed vs. Swing Efficiency

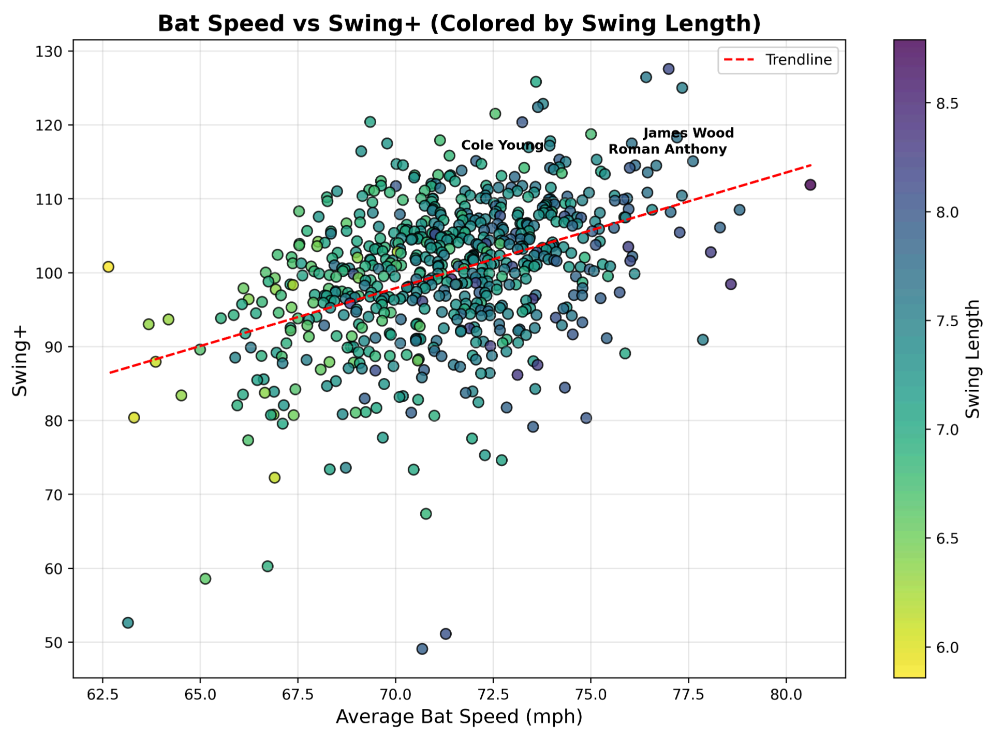

Figure 3: Bat Speed vs. Swing+

In the scatter plot *Bat Speed vs Swing+ (Colored by Swing Length)*, the positive trendline shows the expected relationship: as average bat speed increases, Swing+ tends to rise. However, outliers such as

Cole Young demonstrated that bat speed alone does not guarantee an efficient swing. Despite possessing a more compact stroke (shorter swing length, indicated by the yellow-green hue), Young's Swing+ remained near league-average compared to the longer, more explosive swings of Wood and Anthony, who sat comfortably above 110 on the Swing+ scale.

The distinction highlighted one of Swing+'s core design goals; isolating *how* a player creates bat speed, not just how much they generate. Young's compact, rotational path produces efficient contact at moderate speeds, while Wood and Anthony leverage greater leverage length and attack direction to translate their higher swing intensity into superior contact quality.

Mechanical Fingerprint Comparison

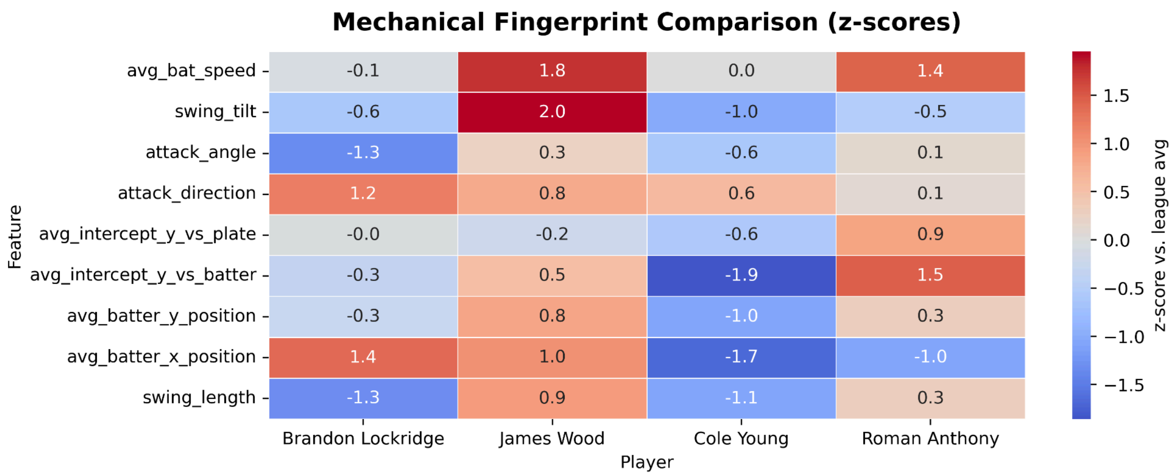

Figure 4: Mechanical Fingerprint Comparison (z-scores)

The *Mechanical Fingerprint Comparison (z-scores)* heatmap provided a deeper layer of evidence. James Wood's z-score profile (bright red in **avg_bat_speed** and **swing_tilt**) reflects a high-power, high-plane swing built for damage, while Roman Anthony displayed a more balanced pattern—slightly elevated in bat speed but steadier across posture and attack metrics. Cole Young, conversely, showed cooler shades of blue, underscoring his flatter attack angle, shorter swing length, and slightly below-average bat speed.

This heatmap was crucial in demonstrating how Swing+ captures a multi-dimensional interaction between speed, direction, and geometry. While Wood's mechanical fingerprint mirrors a high-variance, slugging profile, Young's compact pattern aligns with a contact-oriented hitter. Both can produce value, but the model differentiates them by how efficiently those movements translate into optimized contact—an insight that purely bat-speed-based scouting metrics often miss.

SHAP Decomposition: Cole Young's Swing Drivers

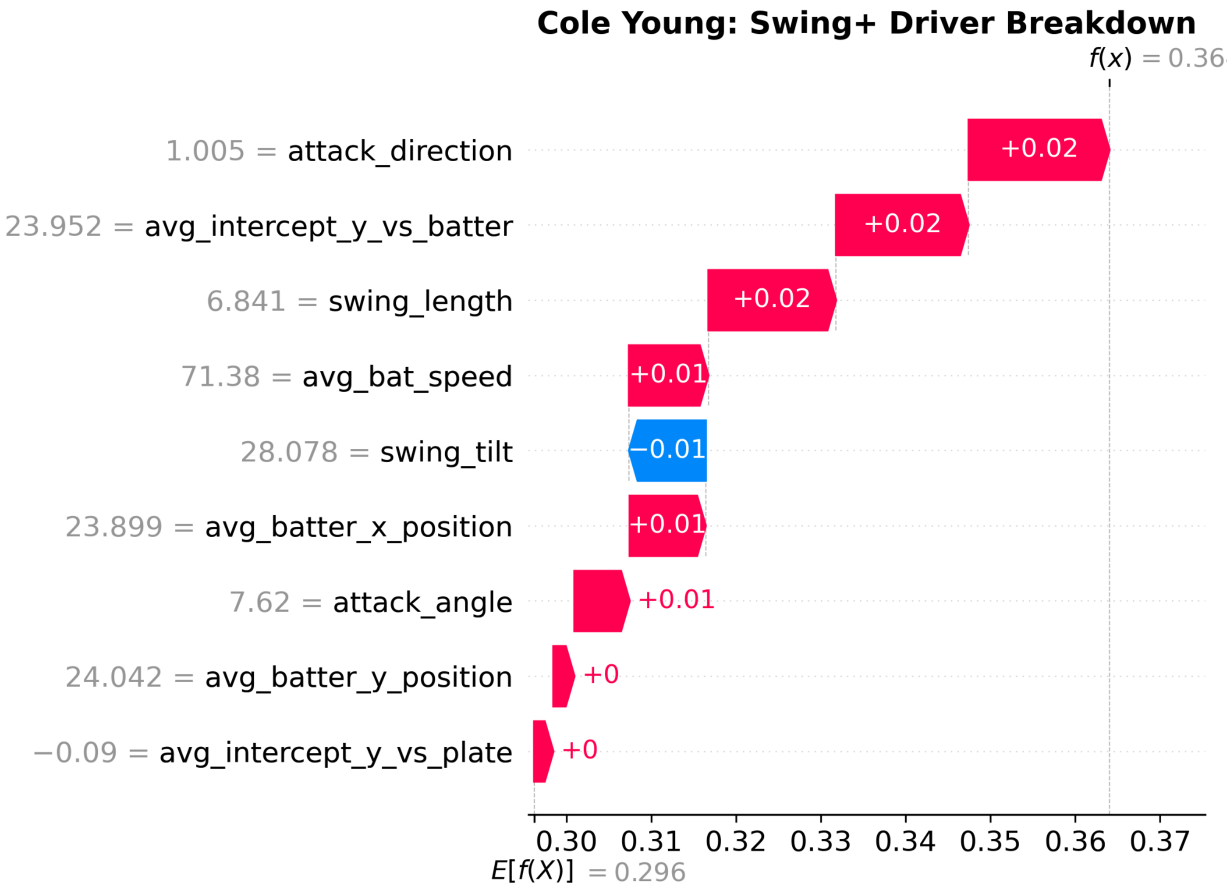

Figure 5: Cole Young SHAP

To interpret these mechanical differences in a more quantitative way, we applied SHAP (Shapley Additive Explanations) analysis to Cole Young's individual prediction. The *Swing+ Driver Breakdown* plot shows how each swing-level variable contributes to his overall score relative to league average.

For Young, attack_direction (+0.02) and avg_intercept_y_vs_batter (+0.02) were his top positive contributors, suggesting precise timing and an efficient bat path through the hitting zone. Meanwhile, swing_tilt (−0.01) and swing_length (−0.01) modestly detracted from his efficiency score—consistent with a flatter, contact-first swing designed to minimize miss but cap maximum energy transfer. By contrast, the same decomposition for James Wood would show positive contributions from avg_bat_speed and swing_tilt, emphasizing a steeper, power-loaded swing plane.

Insights and Transition to Projection Metrics

Through this three-player lens, Swing+ evolved from a descriptive swing-efficiency score into a diagnostic framework capable of capturing both *how* and *why* hitters succeed. The Cole Young–James Wood–Roman Anthony comparison proved that Swing+ could distinguish between mechanically efficient contact hitters and power-driven sluggers, even when both produce strong offensive outcomes.

These insights directly informed the next generation of our modeling system:

- **ProjSwing+** expands Swing+ into a forward-looking projection by integrating bat-speed growth, age curves, and mechanical stability, forecasting how a player's swing efficiency will evolve over time.

- **PowerIndex+** complements Swing+ by quantifying how effectively a hitter converts swing efficiency into *batted-ball impact*, bridging mechanics with raw output metrics such as exit velocity and slugging potential.

Development of ProjSwing+ and PowerIndex+

To extend the interpretability and predictive utility of Swing+, two derivative models; PowerIndex+ and ProjSwing+, were developed using integrated bat-tracking and swing-level mechanical data. These models quantify different aspects of swing performance: PowerIndex+ captures mechanical power potential, while ProjSwing+ projects swing efficiency and scalability across time.

The PowerIndex+ model was constructed by combining key biomechanical inputs known to influence batted-ball quality: average bat speed, swing length, attack angle, swing tilt, and attack direction. Each variable was standardized using *z-score normalization* via StandardScaler to control for differences in scale and variance. Weighted coefficients were then assigned based on their theoretical and empirical contributions to power generation—bat speed (0.50), swing length (0.20), attack angle (0.15), swing tilt (0.10), and attack direction (0.05). The resulting linear composite, termed PowerIndex, was normalized around 100 to create PowerIndex+, where 100 represents league-average mechanical power potential and each 10-point increment corresponds approximately to one standard deviation. This transformation produces a context-independent measure of how efficiently a player's swing mechanics generate force.

Building on that foundation, ProjSwing+ integrates Swing+ (a measure of current swing efficiency) and PowerIndex+ (a measure of physical projection) into a single forward-looking indicator of sustainable swing quality. The model applies a 70/30 weighting between Swing+ and PowerIndex+, emphasizing that mechanical efficiency is the dominant predictor of long-term success, while embedded power traits contribute secondary but meaningful upside.

The resulting metric contextualizes Swing+ in a developmental framework, allowing analysts to distinguish between hitters whose efficient swing patterns are likely to scale with strength gains versus those whose efficiency may be near its ceiling.