Momentum Makers: Value Beyond Salary and Stats

Evanna Yang Dexter Wei Chloe Zhou

April 28, 2025

1 Introduction

Basketball games are often decided by momentum swings and decisive scoring runs. A sudden 12-0 run can shift a game's outcome, and these surges are not always driven by household superstars. Our project asks: *Who are the undervalued NBA players that contribute most to scoring runs, especially when their team is trailing?* This question is important because many role players or young talents provide critical sparks that go beyond what traditional box score statistics capture. Identifying these players can help teams recognize hidden strengths and market inefficiencies.

We tackle this problem by analyzing detailed play-by-play data from the 2024–2025 NBA season. We introduce novel metrics to quantify a player's impact during scoring runs and specifically during comeback situations. Using these metrics, we develop predictive models to evaluate player contributions in context, controlling for factors like team strength, playing time, and salary. By examining which players consistently outperform expectations in momentum-shifting moments, we aim to highlight under-the-radar contributors who may be undervalued relative to their impact. In this report, we describe our data collection and feature engineering, present exploratory data analysis, outline our modeling approaches, and discuss results including model performance and identified standout players. In summary, our team-by-team regression models achieved roughly a 15–20% reduction in RMSE compared to the baseline, and residual analyses highlighted under-the-radar contributors—such as Shaedon Sharpe and Jalen Green—who consistently exceed expectations in scoring runs and comebacks.

2 Data

We constructed a comprehensive dataset from multiple sources covering the entire 2024–2025 NBA regular season. The primary data came from detailed play-by-play logs (sourced via the *cdnnba* repository), which provided information on every scoring event and context. We supplemented this with team statistics from NBA Advanced Stats (team advanced stats) and player information from Basketball Reference (including advanced metrics and salary data). After cleaning and merging these sources, our dataset included over 600,000 play-by-play entries distilled into player and team performance indicators during scoring runs.

From the play-by-play data, we identified segments corresponding to scoring runs (e.g., a 9-0 run or a 13-4 run). We detected scoring runs by scanning play-by-play data to identify uninterrupted periods where one team accumulated a significant point differential, grouping plays by a unique $runId$, and recording team performances during these stretches. For each run, we computed each player's contributions: points scored, assists leading to run points, steals, blocks, and fouls drawn. These contributions were then aggregated by player across the season to form our key response variables. We engineered two novel metrics to quantify impact:

- **RunImpactScore**: A player's total contribution during all scoring runs, combining points and high-impact plays. We computed this as

$$\begin{aligned}\text{RunImpactScore} = \text{RunPoints} &+ 2.3 \times \text{RunAssists} + 1.2 \times \text{RunSteals} \\ &+ 1 \times \text{RunBlocks} + 1.5 \times \text{RunFoulsDrawn}.\end{aligned}$$

  essentially a weighted sum of the player's offensive contributions during runs.

- **TrailingImpactScore**: A player's total contribution during scoring runs *while their team was trailing*, calculated with the same formula but considering only runs where the player's team was behind. This metric highlights a player's impact in comeback situations.

In addition to these run-based metrics, we gathered a variety of features for each player to use in modeling. These included advanced individual metrics like Player Efficiency Rating (PER), Win Shares (WS), Box Plus/Minus (BPM), Usage Rate (USG), etc. We also incorporated team-level stats such as team net rating and pace, since team context can influence opportunities for runs. Each player's annual salary was included as a proxy for their market value. The final dataset has one entry per player for the season, containing their RunImpactScore, TrailingImpactScore, and all aforementioned features. We standardized all continuous features to have mean 0 and unit variance to ensure fair comparison and to aid in model training.

Exploratory Data Analysis: Before modeling, we performed exploratory analysis to understand patterns in scoring runs at both the team and player levels. We visualized three aspects of the data (Figures 1–3) to glean key insights:

Team runs created vs. runs allowed: Figure 1 plots each team's average runs created minus runs allowed per game, showing that every team with a positive net run made the playoffs.

Player run points vs. usage rate: Figure 2 shows a positive relationship between player run points and usage rate, with efficient outliers like Mikal Bridges punching above their usage.

Player performance when trailing: Figure 3 compares run points when trailing, highlighting role players who elevate performance under pressure—often on lower-ranked teams.

These insights confirm that non-superstars drive momentum swings and motivated our inclusion of usage rate and team-level features in the modeling stage.

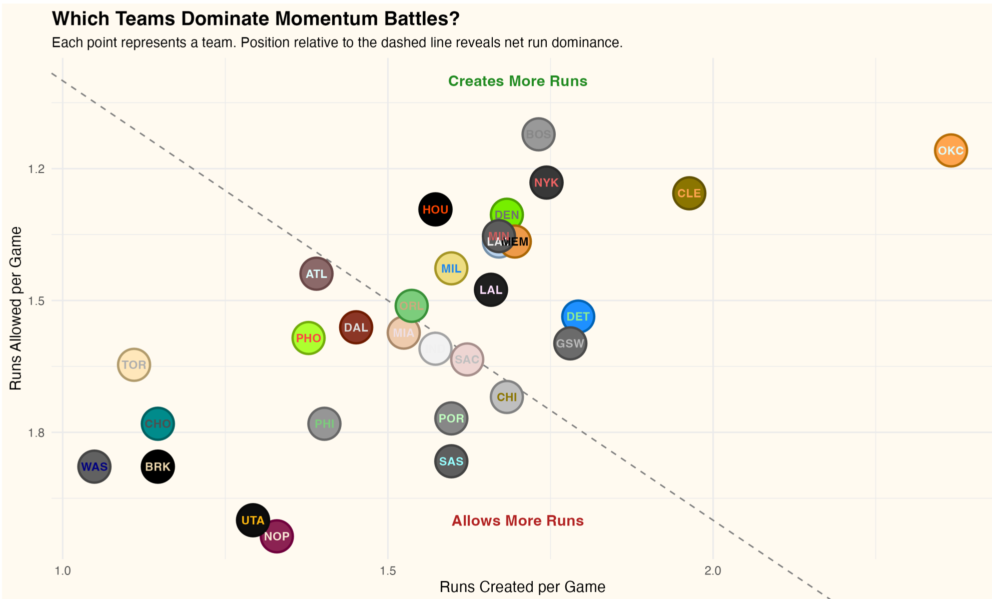

Figure 1: Team Runs created vs. runs allowed

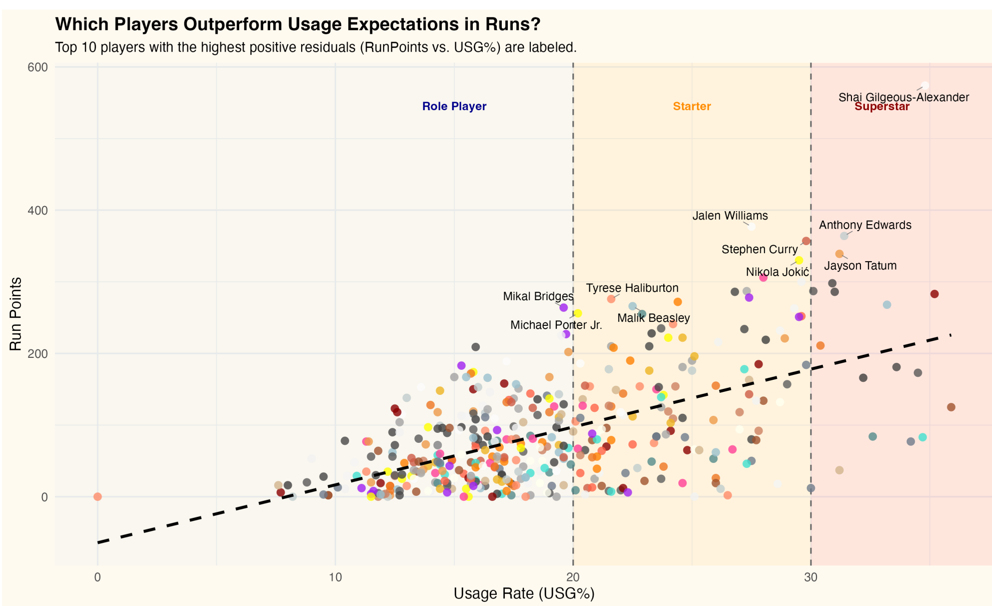

Figure 2: Player usage rate vs. run points

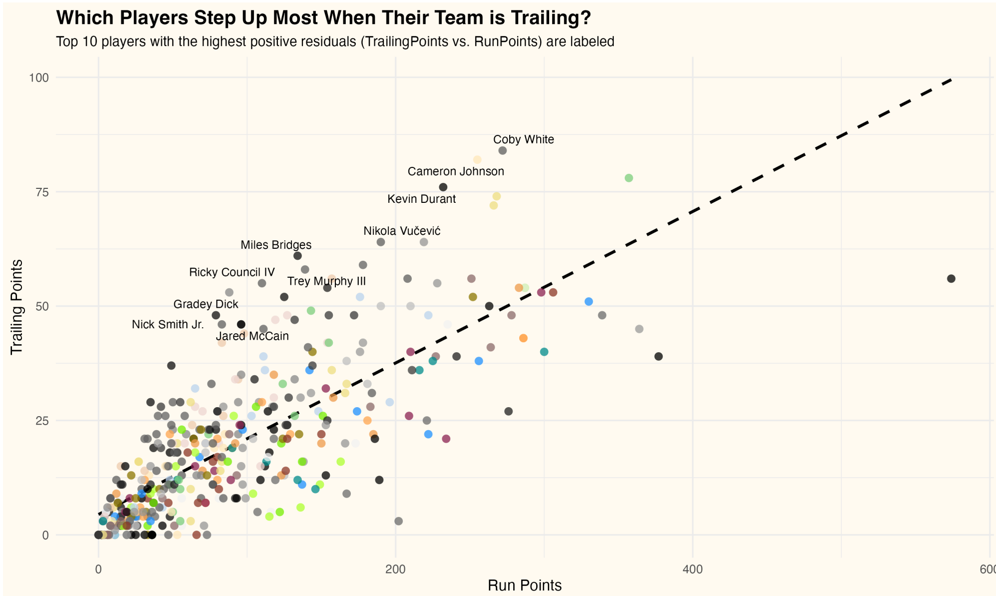

Figure 3: Player run points vs. trailing run points

3 Methods

We employed a series of models to predict player impact scores and identify players who outperform expectations. Our approach progresses from simple linear regression to more complex methods accounting for feature selection, nonlinearity, and team context. Residual analysis is the main tool to spot players who contribute more during runs than expected. Below we outline each model:

1. **Baseline Linear Regression.** We first fit a multiple linear regression to predict a player's RunImpactScore using established metrics: Salary, PER, USG, TS%, TOV%, WS, BPM, and VORP. This model serves as an interpretable benchmark for expected contributions based on a player's overall stats and salary. We applied the same predictors to TrailingImpactScore. After fitting, we computed residuals (Actual − Predicted), where large positive residuals identify players who exceed conventional expectations during scoring runs. We assume a linear relationship between the predictors and response with independent, homoscedastic, normally distributed errors.

2. **Lasso Regression.** Next, to allow a wider set of features (including team-level metrics) and perform automatic variable selection, we employed Lasso regularized regression. The Lasso optimizes

$$\hat{\beta} = \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (y_i - X_i\beta)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\},$$

which adds an L_1 penalty to shrink less informative coefficients to zero. We standardized all predictors before fitting and used cross-validation to choose λ. The Lasso model retained a subset of the most predictive features for RunImpactScore, such as usage rate, win shares, BPM, and certain team efficiency metrics. We again examined model residuals to find players with unusually high impact given the features selected. We assume the underlying linear model assumptions—linearity, independent and homoscedastic Gaussian errors—hold, with the L_1 penalty enforcing sparsity.

3. **Generalized Additive Model (GAM).** To capture potential nonlinear relationships, we fit a GAM with smooth spline terms for key continuous predictors (e.g., Salary, PER, TS%) and included a fixed effect for each player's team. For example:

$$\text{RunImpactScore} \sim s(\text{Salary}) + s(\text{PER}) + s(\text{USG}) + s(\text{WS}) + s(\text{BPM}) + \cdots + \text{Team}\,,$$

   where $s(\cdot)$ denotes a spline. The smooth terms allow for effects like diminishing returns (e.g., the marginal impact of an extra point of PER might decrease at higher PER values). Including Team as a predictor gives each team its own baseline adjustment, helping control for team-level differences in playing style and pacing. We fit analogous GAMs for TrailingImpactScore as well. We assume the response is an additive combination of smooth functions plus independent, homoscedastic, normally distributed errors.

4. **Team-by-Team Linear Models.** As an alternative way to account for team context, we fit separate linear regressions for each team's players. These models use the same form as the baseline (Salary and advanced stats) but are trained on one team at a time, yielding 30 distinct models. This approach lets the intercept and coefficients vary freely by team. It highlights which players on a given team exceed that team's expectations. However, because each team model is independent, residuals from different teams are not directly comparable across the league. The team-by-team approach tends to fit within-team performance very well. We assume that within each team's data the usual linear regression conditions hold: linearity, independent and homoscedastic Gaussian residuals.

5. **Multilevel (Hierarchical) Model.** Finally, we implemented a multilevel linear model with a random intercept for each team. For RunImpactScore:

$$\text{RunImpactScore}_{ij} = \beta_0 + \beta_1 \text{Salary}_{ij} + \beta_2 \text{PER}_{ij} + \cdots + u_{0j} + \epsilon_{ij}\,,$$

   where i indexes players and j indexes teams. The random term $u_{0j} \sim N(0, \sigma_u^2)$ lets each team have its own baseline after adjusting for player-level predictors. This approach "borrows strength" across teams, informing the overall regression while allowing team-specific deviations. We fitted the model using `lme4`, aiming for more robust estimates, especially for teams with fewer players, and more comparable residuals across teams. (A Bayesian variant was also attempted but failed to converge reliably.) We assume player-level residuals are independent and normally distributed with constant variance, and that team-specific random intercepts follow a normal distribution.

Across all models, our primary interest lies in players with large positive residuals, as these are the individuals performing better in scoring runs than expected.

After fitting each model, we evaluated the predictive performance using RMSE and MAE metrics. To assess the stability of these estimates, we applied cross-validation across multiple data folds. Additionally, we used bootstrapping to generate 95% confidence intervals, providing a measure of uncertainty around the performance metrics.

Most importantly, we analyzed residuals to compile rankings of players who most exceeded model expectations, emphasizing multilevel model residuals as the most context-aware measure. We also examined team-level random effects to identify teams systematically exceeding or underperforming expectations.

4 Results

We evaluated five models—Baseline, Lasso regression, GAM with team features, Team-by-Team models, and a Multilevel Team model—on predicting **RunImpactScore** and **TrailingImpactScore**. Performance is summarized using RMSE and MAE (mean $\pm$ 95% CI). Models incorporating team-specific effects and nonlinear terms generally outperformed simpler baselines, particularly for trailing-run contexts. We now detail model performance for each target and use residuals to identify players who exceeded or underperformed relative to expectations.

Table 1: Model performance metrics (RMSE and MAE) with 95% CIs for predicting RunImpactScore

Model	**RMSE**	**MAE**
Baseline	59.4 ± 7.8	42.9 ± 3.6
Lasso	68.7 ± 7.5	50.8 ± 4.5
GAM + Team	64.7 ± 7.2	46.6 ± 4.0
Team-by-Team	50.0 ± 4.7	37.5 ± 3.0
Multilevel Team	67.1 ± 7.2	49.1 ± 4.1

RunImpactScore Model Performance: For predicting RunImpactScore, the **Team-by-Team** approach achieved the best accuracy. It obtained an RMSE of 50.0 ± 4.7 and an MAE of 37.5 ± 3.0, substantially outperforming the Baseline model (RMSE 59.4 ± 7.8, MAE 42.9 ± 3.6). This indicates that accounting for team-specific dynamics reduces error by roughly 15–20% relative to a generic approach. By contrast, the Lasso model performed worse than the baseline (RMSE 68.7 ± 7.5, MAE 50.8 ± 4.5), suggesting that a strictly linear fit with heavy regularization failed to capture important features of run performance. The **GAM + Team** model (which allows nonlinear effects and includes team as a predictor) showed moderate improvement over Lasso, with RMSE 64.7 ± 7.2 and MAE 46.6 ± 4.0. However, even this GAM did not match the baseline's accuracy for RunImpactScore. The **Multilevel Team** model also underperformed (RMSE 67.1 ± 7.2, MAE 49.1 ± 4.1), possibly due to oversmoothing or limited data per team. In summary, for RunImpactScore the only method that clearly improved on the baseline was the Team-by-Team model, implying that individual team contexts have a large impact on a player's run contributions. The 95% confidence intervals for Team-by-Team's error metrics did not overlap with those of the

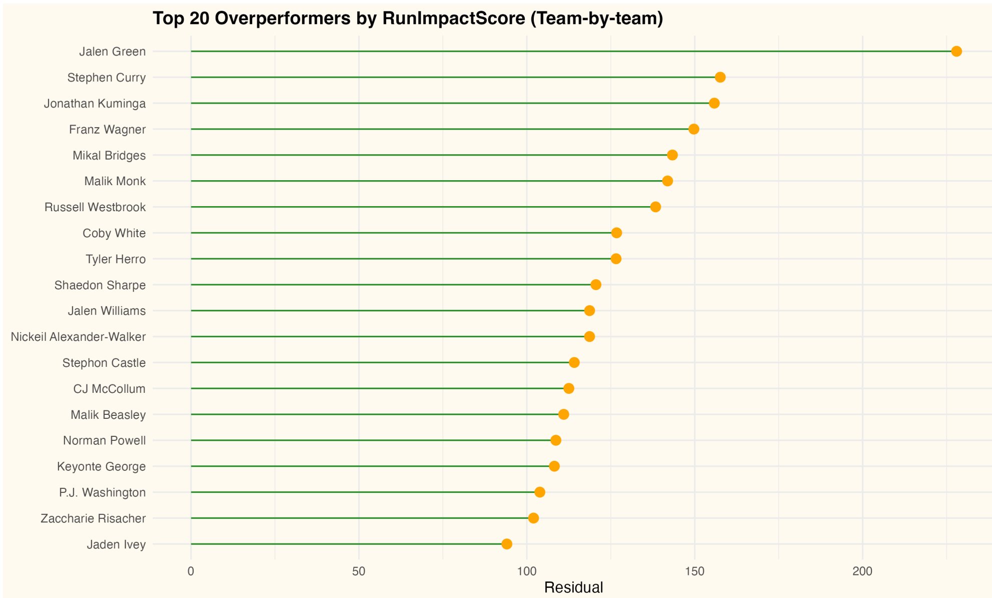

Figure 4: Top 20 overperformers by RunImpactScore

weaker models, indicating a statistically significant gain in predictive power. For example, Team-by-Team's MAE (37.5 ± 3.0) is noticeably lower than the baseline's (42.9 ± 3.6), while other models' intervals overlap with or even exceed the baseline error, reinforcing the benefit of the team-specific approach.

Table 2: Model performance metrics (RMSE and MAE) with 95% CIs for predicting TrailingImpactScore

Model	**RMSE**	**MAE**
Baseline	21.5 ± 1.8	16.3 ± 1.3
Lasso	21.1 ± 1.6	16.4 ± 1.2
GAM + Team	19.6 ± 1.7	15.1 ± 1.2
Team-by-Team	14.9 ± 1.3	11.3 ± 0.9
Multilevel Team	20.5 ± 1.7	15.8 ± 1.2

TrailingImpactScore Model Performance: A similar pattern was observed for the TrailingImpactScore, with even more pronounced differences. The Baseline model had an RMSE of 21.5 ± 1.8 and MAE of 16.3 ± 1.3. The Team-by-Team model again yielded by far the lowest error (RMSE 14.9 ± 1.3, MAE 11.3 ± 0.9), representing roughly a 30% reduction in RMSE compared to baseline. The advantage of incorporating team-specific models is evident: the Team-by-Team 95% CI ranges (e.g., RMSE 13.6–16.2) do not overlap with those of the Baseline (19.7–23.3), confirming a significant improvement. The **GAM + Team** model also improved on the baseline, achieving RMSE 19.6 ± 1.7 and MAE 15.1 ± 1.2. In contrast to the RunImpactScore case, even the GAM's performance here is meaningfully better than Baseline (CIs barely overlap). The Lasso model for TrailingImpactScore was on par with baseline (RMSE 21.1 ± 1.6, MAE 16.4 ± 1.2), underscoring that without allowing nonlinearity

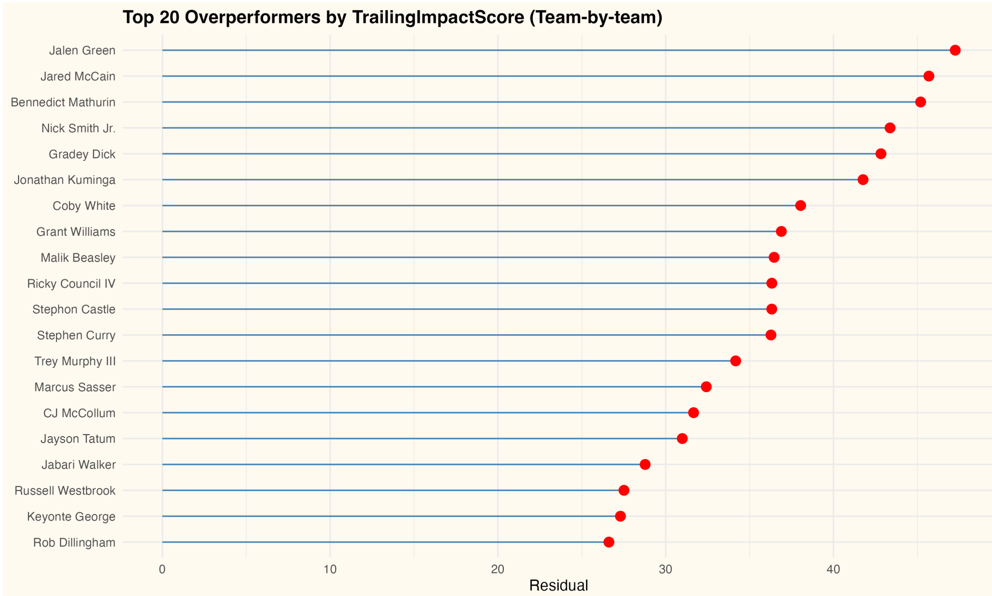

Figure 5: Top 20 overperformers by TrailingImpactScore

or team factors, the model cannot gain predictive accuracy in this context. The Multilevel Team model (RMSE 20.5 ± 1.7, MAE 15.8 ± 1.2) showed a slight improvement over baseline but was still far worse than the Team-by-Team approach. These results demonstrate that players' performance when their team is trailing is highly dependent on team context and possibly nonlinear interactions – a one-size-fits-all model struggles to capture these patterns. Including team-specific effects (either via separate team models or as features in GAM) substantially tightened the error bounds and increased accuracy for the trailing scenario.

Residual Analysis of Player Impact: To understand which players the models struggled with, we analyzed the residuals. Positive residuals indicate players who **overperformed** relative to model expectations. The Team-by-Team model's residuals were examined since it was the best predictor – any remaining large residuals highlight exceptional cases.

In Figure 4, the top residuals from the multilevel model based on RunImpactScore – how players perform during scoring runs. These players, ranked by decreasing residuals, exceeded the model's expectations the most. While you'll definitely spot superstars like **Stephen Curry**, we also see rising young talents like **Shaedon Sharpe**, **Stephon Castle**, and **Keyonte George**. Veterans like **Russell Westbrook** still offer steady momentum-driving contributions, as he always does.

In Figure 5, the lollipop plot for TrailingImpactScore highlights players who shine when their team is behind and under high pressure. Some names reappear, like **Jalen Green**, **Jonathan Kuminga**, and **Malik Beasley**, showing that they're reliable in both momentum and adversity. Expanding this list beyond the top 20, even more under-the-radar talents emerge as consistent momentum shifters for their teams. Again, by focusing on residuals from a multilevel model, we reveal underappreciated contributors – often overlooked by standard evaluations – that teams might be undervaluing.

5 Discussion

This study introduced a new way to evaluate NBA players by focusing on their impact during scoring runs and comebacks. By moving beyond traditional box score stats, we identified undervalued players who consistently spark momentum shifts. Our results showed that accounting for context—both non-linear effects and team adjustments—is crucial. The multilevel model, which separated player effects from team style, emerged as the best method, providing both strong predictive performance and interpretable residuals.

One major takeaway is that many players add hidden value not reflected in season aggregates. Players highlighted in our residual analysis often act as "igniters," delivering bursts of energy or critical defensive plays that swing momentum. These situational contributions, captured by our RunImpactScore and TrailingImpactScore metrics, often go unnoticed but can be game-changing.

Practically, NBA teams could use this approach to identify undervalued talent. Players with consistently high run impact residuals might be ideal trade targets or candidates for expanded roles. Our metrics also offer a fresh lens for recognizing "clutch" contributions—not just final-shot heroes, but players who fuel critical mid-game runs.

However, there are several limitations. Our analysis used only one season of data; expanding to multiple seasons and playoffs would improve robustness. Although we included many advanced stats, features like lineup context, coaching strategies, fatigue, and richer defensive measures (like deflections or contested shots) could further enhance our models.

For future work, developing a Bayesian multilevel model could be valuable. Bayesian methods would provide full posterior distributions for player effects, improving uncertainty quantification and handling small sample sizes more rigorously. Our initial attempt at a Bayesian approach was inconclusive, but further refinement is promising.

Additionally, integrating modern player tracking data could deepen insights. Tracking metrics—like speed, off-ball movement, screen-setting, or hockey assists—would capture critical run-making behaviors not visible in basic stats. This could refine RunImpactScore or inspire new momentum-based metrics.

In conclusion, our project shows that focusing on momentum swings reveals hidden layers of player value. Multilevel modeling combined with residual analysis provides a powerful framework to spotlight players who consistently exceed expectations. By expanding the data and refining the models, we can continue uncovering the unsung engines behind a team's success in the game of runs.