From Marketing to the Court: Applying MMM and Fatigue Analysis for Optimal Basketball Lineups

Nachi Lieder

April 14, 2025

Abstract

Lineup optimization in basketball is crucial for maximizing team performance. Traditional methods often overlook the effects of player fatigue and the absence of historical data for certain lineups. This study introduces an advanced model inspired by marketing mix models (MMM), incorporating player fatigue to optimize lineups and maximize the plus-minus metric. By examining two distinct lineups: One high-tempo, fast-breaking, and the other slow-paced and durable, we highlight the varying impacts on productivity and fatigue. The fast-breaking lineup may show a higher immediate productivity but suffers from quicker fatigue, while the durable lineup maintains consistent performance over a longer period.

I Introduction

Lineup optimization in basketball is a critical component of maximizing team performance.Traditional approaches often assume linear and uniform productivity across different lineups,failing to account for the dynamic effects of fatigue that accumulate within the course of a single game. As players spend prolonged minutes on the court, fatigue sets in, reducing both individual performance and the overall effectiveness of the lineup. Even the best-performing lineups, while initially dominant, can see diminishing returns as their ability to sustain high levels of play deteriorates over time.

For instance, during the 2016 NBA Finals, the Golden State Warriors' demonstrated the "death lineup", a high-tempo and versatile group, which began to falter as fatigue from extended court time took its toll. Despite being one of the league's most effective lineups, their ability to execute plays and maintain defensive intensity declined noticeably in the later stages of the game. Similarly, in the 2014 NBA Finals, the Miami Heat leaned heavily on their star players, LeBron James and Dwyane Wade, who saw their effectiveness drop significantly in the second halves of games as extended minutes without rest left them unable to counter the fresh rotations of the San Antonio Spurs. These examples illustrate the critical need for models that dynamically account for within-game fatigue and its impact on lineup performance.

To address these limitations, this study introduces a novel model inspired by Marketing Mix Models (MMM), a statistical framework traditionally used to measure the impact of marketing channels on outcomes like sales and ROI. In this adaptation, lineups are treated as analogous to marketing channels, with their contributions to team performance analyzed dynamically over time. The proposed model incorporates player fatigue as a decay function, enabling a more realistic and actionable understanding of lineup performance as it fluctuates during a game.

By examining two distinct lineup types: one high-tempo and fast-breaking, the other slow-paced and durable, we explore the trade-offs between immediate productivity and sustained performance. The fast-breaking lineup may achieve higher initial returns but suffers from quicker fatigue and sharp diminishing returns, while the durable lineup maintains steadier performance over longer stretches, though at a lower peak productivity. Instead of favoring one type over the other, our approach emphasizes dynamically balancing both lineup styles throughout the game to optimize overall team performance.

This study leverages advanced data analytics and machine learning techniques to enhance traditional regression models, integrating a fatigue transformation kernel to account for diminishing returns within a game. By clustering players based on performance characteristics, analyzed on Israeli Basketball League data, we demonstrate the model's robustness and its superiority over baseline methods. The findings reveal significant scoring differentials among lineup clusters and fatigue function setups, underscoring the importance of strategic player utilization and providing actionable insights for coaches and analysts.

II Methods

1 Data Collection

The dataset includes detailed metrics from the Israeli Basketball League, covering player statistics, play-by-play data, and game outcomes. This comprehensive data set allows robust analysis and model development.

The data pertain to information from the 2023-24 season with the intention that due to the variance in teams rosters, we can address the core issue of dealing with small sampled data.

2 Terminology

In this section, we will explain common terminology used in the analysis of basketball through data science.

- **Plus-Minus**: Plus-Minus, often abbreviated as +/- or PM, is a statistic that measures the point differential when a player/set of players are on the court. It is calculated by subtracting the points allowed from the points scored by the player's team while they are in the game. This metric helps in understanding the overall impact of a set of players on the team's performance, accounting for both offensive and defensive contributions. [3]

- **Lineups**: In basketball, a lineup refers to the combination of five players on the court for a team at any given time. Analyzing different lineups helps coaches and analysts determine which groupings of players work best together, providing insights into the most effective combinations for various situations during a game. Data on lineups can reveal synergies between players and the overall balance of the team.

- **Possessions**: A possession in basketball is a period during which a team controls the ball and attempts to score. It starts when a team gains control of the ball and ends when they either score, turn the ball over, or the other team gains possession. Possession-based metrics, such as points per possession (PPP) and turnover rate, are crucial for understanding a team's efficiency and effectiveness on both offense and defense.

3 Initial Linear Regression Model

We start with a linear regression model as a baseline model similarly used within the literature [2], $Y \sim X_i$, where Y is the plus-minus metric, and X_i represents the number possessions per lineup i. This model serves as the baseline for further enhancements. The regression model was chosen for its interpretability, based on the assumption that the coefficients can represent the solution, which, when normalized, can describe the optimal timeshare proportion for each lineup. While this approach draws conceptual inspiration from Marketing Mix Models (MMM), which have been widely used in marketing to allocate resources across channels, it diverges by focusing on dynamic lineup performance in basketball. Unlike past work that primarily examines static efficiency metrics or linear contributions, this method incorporates player fatigue as a key variable, providing a distinct perspective on lineup optimization. [5]

Each lineup combination is represented as a column, and each game as a row, such that if there are M games and N different combinations of lineups played that year , we will represent our problem as MxN. It is important to note that not all possible lineups play in each game. For that case we set the number of possessions to 0 per that game per that lineup.

The target in this case is the difference in scores between the current team and the opposing team. Match scores are considered an ideal measure point [8]. For example, if the final score was 95:94 to the team being evaluated, the target will be 1. If the score was a loss of the same score, it would be -1.

The initial thought of modeling was to optimize per team, but due to the variance within the teams' roster over the year and dimension problem due to small data, we will look at a more global look at the problem. We will bundle all teams together into one dataset, and instead of representing lineups by actual players, we will represent them by combination of "player styles". (which will be elaborated below). Once we optimize the process we can decode the solution into a relevant set of lineups per a single team.

4 Enhancing the Model with Fatigue Decay

While the initial linear regression model provides a basic understanding of how different lineups affect the plus-minus metric to a certain level, it does not account for the impact of player fatigue. This limitation arises because the coefficients in the linear model are assumed to be stable over time and are not time-variant, meaning they do not decay as a function of volume. This stability leads to a regression-to-the-mean effect, where the influence of each lineup's minutes on the plus-minus metric remains constant regardless of the duration of play.

To address these limitations, we introduce a decay function, $K_i(X_i)$, that dynamically adjusts the contribution of each lineup's minutes to the plus-minus metric, accounting for the effects of fatigue over time. This approach allows the model to reflect the diminishing returns on performance as players become fatigued, providing a more accurate representation of the impact of lineups over extended periods of play.

5 Decay Function Definition

The decay function $K_i(X_i)$ is designed to model the diminishing returns of player performance as fatigue sets in. We define $K_i(X_i)$ as an exponential decay function [6] described below:

$$\mathbf{K}(X_i, \tau_i) = \frac{1}{1 + \frac{X_i}{\tau_i}}$$

where X_i is the possessions played by lineup i , and τ is a parameter that controls the rate of decay for lineup i. A higher value of τ_i indicates a faster rate of fatigue, whereas a lower value suggests more endurance.

6 Graphical Representation

To visualize the effect of the decay function, consider the figure below, which plots $K_i(X_i, \tau_i)$ for various values of τ_i . The graph demonstrates how the contribution of lineup minutes decreases as playing time increases, reflecting the impact of fatigue.

This transformation is applied to the dataset, where each lineup's minutes are adjusted based on their respective decay rates. The transformed dataset is then used to re-fit the regression model.

7 Optimizing the Decay Function Parameters

Determining the optimal values of τ_i for each lineup is crucial for accurately modeling the effects of fatigue. We employ hyperparameter optimization to

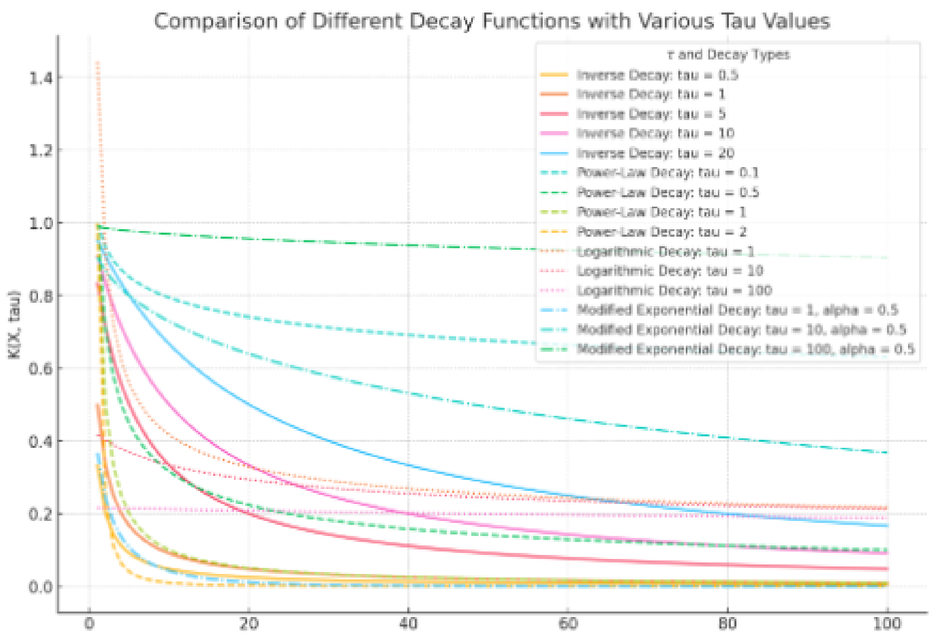

Figure 1: Graph of the decay function $K_i(X_i)$ for different values of τ_i.

find these values, focusing on minimizing the regression error.

7.a Hyperparameter Optimization

As shown in Algorithm 1, the optimal τ_i is selected based on regression error minimization.

Algorithm 1: Hyperparameter Optimization for τ_i

Input: Range of candidate τ_i values for each lineup
Output: Optimal τ_i minimizing regression error

foreach *lineup* **do**
  foreach *candidate* τ_i **do**
    Transform the lineup minutes using τ_i;
    Fit a regression model to the transformed data;
    Compute the regression error per τ_i ;
  Select the τ_i with lowest regression error;

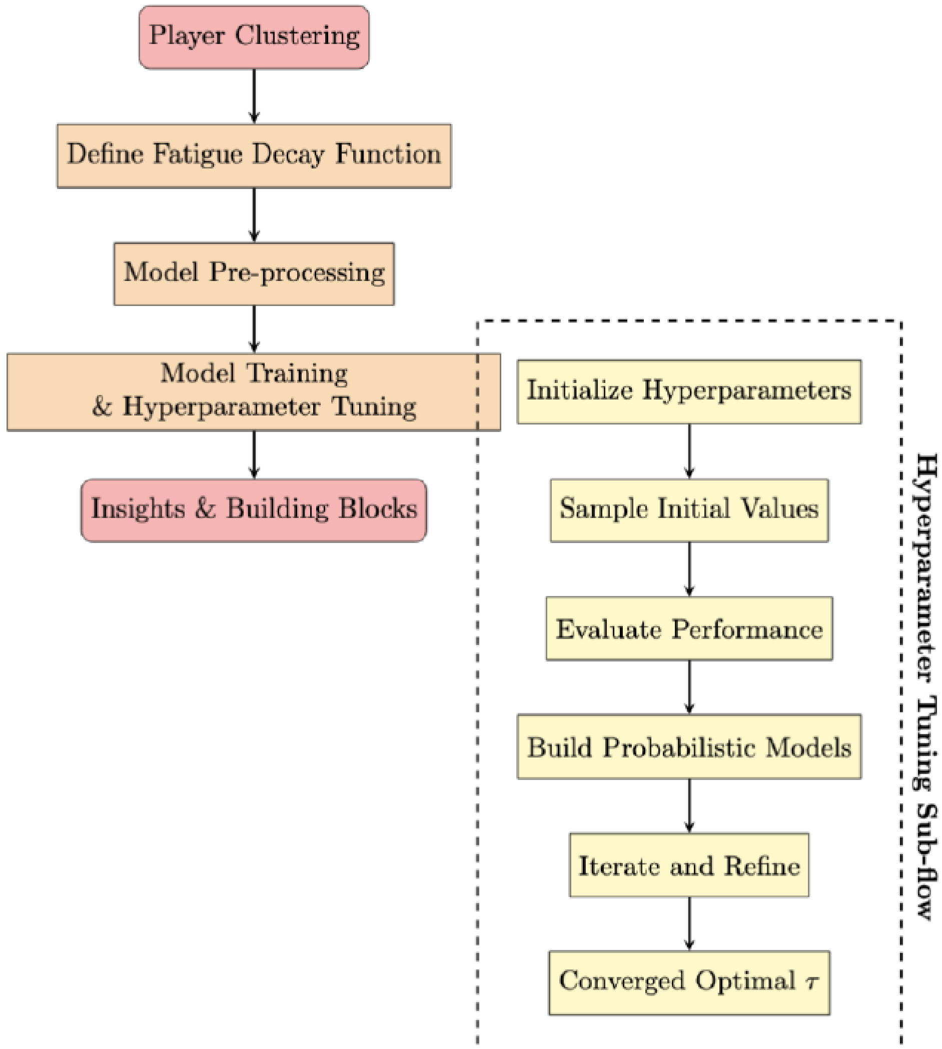

Figure 2: Flowchart for Lineup Optimization Process with Hyperparemter Tuning Sub-flow, optimizing the *tau* values to incorporate within the decay function

We utilize Tree-structured Parzen Estimations (TPE) for efficient convergence during hyperparameter optimization. TPE is particularly suited for high-dimensional and complex search spaces, making it an ideal choice for our problem.

7.b Implementation

The implementation involves iterating over possible values of τ_i and evaluating the performance of the transformed model. The optimal τ_i values are those that result in the lowest regression error, ensuring that the model accurately reflects the effects of fatigue.

8 Player Clustering

To handle high dimensionality and limited data, we cluster players based on performance characteristics. These clusters form distinct lineup types, enabling generalization and robustness in our models. The method of clustering is open to interpretation in a way that represents the clusters ideally. For the sake of this evaluation, the clustering process was performed based on players' performances from the previous year isolated by player. The method in this process of clustering was K-means , though other methods may apply as well.

This paper does not focus on the method of clustering, as significant work has already been done in this area. Xu and Martens [10] used k-means to group players into roles like "primary scorers," while Terner and Franks [9] applied hierarchical clustering to identify traditional and hybrid positions. Luo and Wu [1] further extended this by using mixed-data clustering for a nuanced view of player efficiency. Building on these established methods, this paper focuses on the application of these clusters in lineup optimization.

Clustering players by "types" and generalizing the lineups as combinations of five labels helps to mitigate the problem of limited data and high dimensionality. By grouping players with similar performance characteristics, we can create more generalized lineup types that can be analyzed across multiple games and teams. This approach allows us to aggregate data from various sources, increasing the robustness and reliability of our models. Additionally, clustering makes the model more adaptable to dynamic team compositions, ensuring that it remains relevant even as player rosters change.

III Results

1 Analysis of Optimized τ_i Values

The enhanced model, utilizing the optimized τ_i values obtained through the hyperparameter tuning process, consistently outperforms the baseline model where no decay function is applied. The optimization of τ across different data sets has shown to significantly improve predictive accuracy, particularly by capturing the nuances of player performance over time. This demonstrates that incorporating tailored decay rates provides a more realistic and effective representation of fatigue effects within the model.

2 Distribution of τ Values Across Lineups

The analysis across various lineup clusters reveals that the optimized τ values are distributed unevenly, with certain lineups exhibiting notably larger τ values than others. This non-uniform distribution suggests that different lineups require varying degrees of decay to accurately model performance.

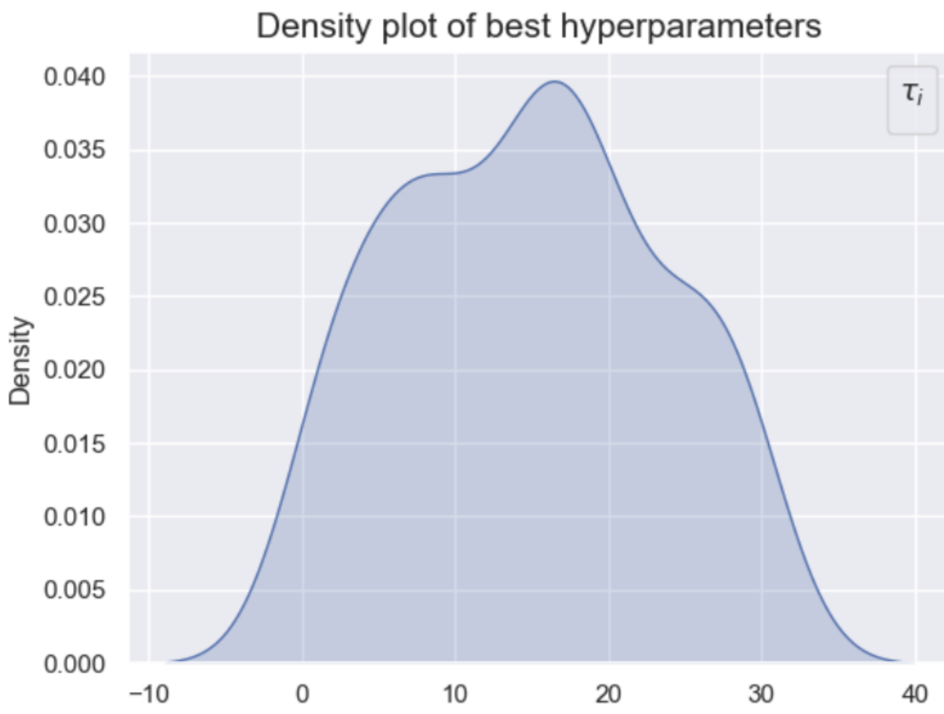

Figure 3: Density plot of τ values across different lineups, showing the distribution and variation in optimized decay rates.

3 presents the density plot of τ values, highlighting the variation in decay rates across different lineups. The plot shows that some lineups benefit from more aggressive decay (higher τ), leading to quicker declines in performance, while others maintain more consistent performance over extended periods with lower τ values.

3 Improvement Over Baseline

The improvements gained by using the optimized τ values are further illustrated in the table below, which presents a comparison of RMSE values across different decay functions relative to the baseline model. The table highlights the mean and median RMSE values, along with their percentage improvements. Notably, the majority of the decay functions outperform the baseline across nearly all trials, with the mean and median values of the exponential, inverse, and Gaussian decay functions showing significant reductions in error. This strongly indicates that in most cases, applying the optimized decay functions is preferable, as they consistently lead to better model performance compared to the baseline.

The Mean RMSE represents the average RMSE over 20 different folds of the data in a cross validation comparison.

Table 1: Mean RMSE and % improvement over the baseline across cross-validation folds using decay functions. A visual summary is shown below.

Decay Function	Mean RMSE	% Improvement
Exponential	4.454	2.79%
Inverse	4.488	2.05%
Gaussian	4.548	0.74%
Baseline	4.582	–

4 Optimal Lineups and Fatigue Management

The results underscore the necessity of managing player fatigue to optimize overall team performance. Lineups designed for short bursts of high productivity can be highly effective in specific situations but require careful rotation to prevent rapid declines in performance. Conversely, durable lineups provide steady performance, making them valuable for maintaining consistency throughout the game. A team cannot sustain optimal high paced lineups throughout the entire game without "running out of minutes". Therefore one would need to balance between high valued τ lineups and low valued ones.

IV Conclusion

This research introduces a comprehensive framework for optimizing basketball lineups by incorporating the effects of player fatigue. The integration of decay functions and sophisticated statistical models offers coaches data-driven strategies to enhance team performance. By balancing productivity and fatigue, teams can dynamically adjust to in-game conditions and maintain peak performance throughout the game. The insights gained have broad implications for the sports industry, providing a nuanced approach to lineup management that can give teams a competitive edge.

The analysis reveals significant scoring differentials among lineup clusters, demonstrating the model's robustness and superiority over baseline approaches. Optimal lineups exhibit diminishing returns with prolonged play due to fatigue, underscoring the need for balanced player utilization.

The insights gained offer coaches data-driven strategies to enhance team performance. By leveraging sophisticated statistical models and clustering techniques, teams can maintain peak performance and gain a competitive edge which is well known across the industry [7]. The integration of MMM principles and fatigue modeling provides a nuanced approach to lineup management, ensuring optimal performance throughout games.[4]

V Discussion

Future work could involve improving the trained model and decay function. In this paper we can see the emphasis of the methodology as a breakthrough for future work. In addition, incorporating opposing team data to enhance the robustness of the models. Additionally, exploring more advanced clustering methods could yield even more precise player groupings. Another potential avenue is to apply the models on a global scale, leveraging data from the entire league or across multiple leagues, and then tailoring the insights to optimize individual team strategies.

The improved models developed in this study provide valuable insights into lineup optimization and game planning. By converging on more accurate coefficients, these models facilitate better decision-making. For instance, while a high-tempo lineup may show strong initial performance, its effectiveness can quickly diminish, whereas a lineup designed for endurance tends to maintain steady performance throughout the game.

References

[1] Jian Luo and Bin Wu. "Player Types and Efficiency in the NBA: A Clustering Analysis Using Mixed Data". In: *Journal of Sports Science and Medicine* 17.2 (2018), pp. 161–169.

[2] Brian Macdonald. "A regression-based adjusted plus-minus statistic for NHL players". In: *Journal of Quantitative Analysis in Sports* 8.3 (2012).

[3] NBA. *NBA Plus-Minus: Glossary*. Accessed: 2024-08-24. 2024. URL: `https://www.nba.com/stats/help/glossary/#plusminus`.

[4] Thomas Reilly and A Mark Williams. *Science and Soccer*. Routledge, 2003.

[5] Dan T Rosenbaum. "Measuring how NBA players help their teams win". In: *MIT Sloan Sports Analytics Conference*. 2004.

[6] Jaime Sampaio and Manuel Janeira. "Statistical analyses of basketball team performance: Understanding teams' wins and losses according to a different index of ball possessions". In: *International Journal of Performance Analysis in Sport* 3.1 (2003), pp. 40–49.

[7] Hal S Stern. "On the probability of winning a football game". In: *The American Statistician* 45.3 (1991), pp. 179–183.

[8] Clinton Sullivan et al. "Match score affects activity profile and skill performance in professional Australian Football players". In: *Journal of Science and Medicine in Sport* 17.3 (2014), pp. 326–331.

[9] Ziv Terner and Andy Franks. "Position and Style Differences Among NBA Players: A Clustering Approach". In: *Journal of Quantitative Analysis in Sports* 12.3 (2016), pp. 145–160.

[10] Xiaoyan Xu and David Martens. "Clustering NBA Players: Examining the Influence of Advanced Statistics on Player Categorization". In: *Journal of Sports Analytics* 5.4 (2019), pp. 297–312.

VI Appendix

The improvements gained by using the optimized τ values are further illustrated in Table 2, which presents a comparison of RMSE values across different decay functions relative to the baseline model. The table highlights the mean and median RMSE values, along with their percentage improvements. Notably, the majority of the decay functions outperform the baseline across nearly all trials, with the mean and median values of the exponential, inverse, and Gaussian decay functions showing significant reductions in error. This strongly indicates that in most cases, applying the optimized decay functions is preferable, as they consistently lead to better model performance compared to the baseline.

Table 2: Comparison of RMSE Values for Different Decay Functions with Mean, Median, and Improvement

Trial	Exponential Decay	Inverse Decay	Power-Law Decay	Gaussian Decay	Baseline
0	4.416	4.486	4.890	4.452	4.529
1	4.504	4.556	4.745	4.554	4.593
2	4.509	4.551	4.768	4.663	4.678
3	4.486	4.531	4.808	4.614	4.618
4	4.125	4.180	4.524	4.329	4.319
5	4.582	4.619	4.899	4.615	4.696
6	4.554	4.598	4.848	4.628	4.663
7	4.305	4.302	4.407	4.395	4.438
8	4.488	4.519	4.765	4.498	4.605
9	4.530	4.562	4.822	4.691	4.706
10	3.745	3.778	4.077	3.824	3.826
11	4.502	4.540	4.597	4.675	4.724
12	4.531	4.564	4.780	4.589	4.632
13	4.572	4.605	4.892	4.688	4.687
14	4.574	4.611	4.900	4.750	4.725
15	4.278	4.326	4.651	4.516	4.463
16	4.509	4.585	4.858	4.609	4.660
17	4.275	4.314	4.605	4.399	4.427
18	4.449	4.486	4.712	4.613	4.618
19	4.392	4.446	4.688	4.484	4.498
Mean	4.454	4.488	4.728	4.548	4.582
Median	4.486	4.519	4.768	4.589	4.618
Pct Improvement (Mean)	2.79%	2.05%	-3.21%	0.74%	–
Pct Improvement (Median)	2.85%	2.14%	-3.25%	0.63%	–