New York Mets Edwin Diaz (39) walking towards the mound. (Getty Images/Marc Levine) Source: Getty Images/Marc Levine

Decision-Making Analytics for MLB

Connor Train and Rafael Gonzalez

MSDS 456 – Sports Performance Analytics

Northwestern University

August 22, 2024

Introduction

Statistics are plentiful in Major League Baseball (MLB) today and teams must find ways to use it to gain a competitive advantage. Lately data analysis has been commonly used for decision making during different in game scenarios. One of those being the challenging task of deciding when to use a starting pitcher versus a relief pitcher during a game. Pitchers are extremely valuable assets for an organization so it's imperative to attempt to maximize performance while also minimizing the risk of injury. This project aims to analyze data from the 2022 and 2023 MLB regular seasons to find key metrics that help with these difficult decisions. Pitching metrics historically used to evaluate pitching performance include Earned Run Average (ERA), Walks plus Hits per Inning Pitched (WHIP), and Strikeout Ratio (SO). These traditional statistics have been a cornerstone in baseball due to their straightforward nature. A pitcher's success can be measured by their ability to get outs, limit baserunners, and prevent runs. In recent history advanced statistics have become very popular due to the extent of what they measure. Even though new metrics have become readily available, the standard stats remain significant in their ability to provide insights into pitcher effectiveness and assisting in making strategic decisions.

The project will analyze these key statistics using a comprehensive dataset from Baseball Reference and Baseball Savant. The use of advanced statistical techniques will help identify patterns and correlations used to create guidelines for optimal utilization of starting and relief pitchers. By providing actionable insights we can improve team success and prevent player injuries. The following report examines the findings from this project by first providing a detailed review of existing literature on pitcher performance and injury prevention. The methods section focuses on the statistical techniques used to filter and analyze the data. Results will be presented to highlight the identified key relationships and recommendations on utilizing them for decision-making strategies on pitcher management.

Professional sports have an increasing dependence on data analytics underlining the importance of this project. This organization can stay ahead of the curve by leveraging these actionable insights to gain a competitive advantage. Simplifying decisions for optimal pitcher management will improve team outcomes and ensure the health and longevity of an important piece of our team.

Literature Review

An important aspect of a manager's job is to strategically utilize the pitching staff. It is difficult to know when to pull a starter and replace them with a relief pitcher since there are many factors that can influence the decision. There has been a large amount of research conducted attempting to evaluate the effectiveness of pitchers in different scenarios. This literature review will explore previous research on performance, injury prevention, and strategic optimization.

Evaluation of a pitcher's effectiveness has been thoroughly researched. In Chapter 5 of the book *Mathletics*, it explains how standard metrics like ERA are critical when evaluating a pitcher's ability to limit runs scored. A pitcher's allowed runs are adjusted for the number of innings pitched providing our project with valuable information for their effectiveness and fatigue. ERA does not always tell the entire story of a pitcher's performance since there are other factors beyond the pitcher's control that could impact their ERA (Schwarz, 2020).

WHIP is another traditional statistic that measures a pitcher's ability to control baserunners.

Controlling a game from the mound involves limiting base traffic which is important for preventing runs. Research shows that evaluations based on WHIP and ERA remain extremely reliable, especially when comparing different pitchers over multiple seasons (Lewis, 2020). The number of hits allowed by a pitcher directly impacts both their ERA and WHIP. Preventing runs leads to winning games and that starts with limiting hits from the opponents. Batting Average (BA) of balls in play is much higher when compared to the standard BA at the start of an at bat. When a ball is hit in play it increases the number of variables on the outcome of the play. Limiting hits decreases the pitcher's reliance on the defense and the inherent randomness of a ball in play (Lewis, 2020).

The inevitability of allowing baserunners during a game creates a more difficult situation for a pitcher's approach. Strikeout rate (SO) is the statistic used to measure a pitcher's ability to retire batters without a ball being put into play. Outs become more valuable with runners on base especially when they come without a ball being put in play. A pitcher's dominance can be associated with a high strikeout rate since those outs reduce the number of variables, like a defensive error, that are out of a pitcher's control (Dunbar, 2020). Strikeouts help pitchers accrue the needed outs during the span of a game, but they can have a significant impact on the outcome of the game when timely executed. This metric is important when analyzing relief pitchers' effectiveness due to the high-leverage situations they frequently face. This helps managers make decisions on pitcher deployment as a key strikeout late in a game can be the difference between winning and losing a game.

The pitching staff is a valuable part of the team's overall success and as such their workload needs to be monitored to prevent injuries. The number of innings pitched shows how durable and effective a pitcher can be, which is key when comparing starting pitchers. When starters pitch late into games it generally indicates that they are executing their pitches well. This provides value to a team by reducing the need to rely on the bullpen and creating stability for a pitching staff (Schwarz, 2020). To prevent injuries, teams should be mindful of each pitcher's IP due to the significance of durability. As they always say, a player's best ability is availability. The use of these standard statistics by organizations across MLB remains consistent. They are straightforward, which allows for the analysis of data to be clearly communicated throughout the team. Using these reliable and simplistic metrics for our project will be key since they directly correlate with the outcome of a game. Analyzing these metrics will provide a solid foundation when evaluating pitcher effectiveness and supporting strategic decisions when utilizing pitchers during the regular season.

This project aims to guide the team's strategies on pitcher deployment. Chapter 14 of Mathletics explains different methods used and how these traditional metrics like IP and SO can assist in making crucial pitching decisions. It discusses the concept of a pitcher facing a lineup for the third time and how it impacts the performance of the starting pitcher (Schwarz, 2020). When facing a lineup for the third time a starter becomes more fatigued, and the batters are more familiar with their pitch arsenal. Teams often rely on IP and SO when deciding whether to pull the starting pitcher. If the pitcher's SO rate declines with the increasing number of pitches, managers need to utilize their bullpen to sustain a competitive advantage.

Monitoring pitch count and the number of hits allowed can provide valuable insights into a pitcher's performance and injury risk. Protecting a pitcher when they start fading is imperative to prevent overexertion of a pitcher's arm throughout the season. Utilizing these common metrics has helped inform team's decision-making strategies for optimal pitcher use. Using pitchers in the correct roles at the right time leads to winning more games. A balanced approach is required to ensure the long-term health and effectiveness of a pitching staff.

Methods

Data Acquisition: For the project scope, we have obtained data from both Baseball Reference and Baseball Savant that cover the 2022 and 2023 MLB seasons. The datasets include detailed pitching statistics for all the starting pitchers and relievers across the entire MLB. By acquiring all this data, we would be able to notice patterns in our data that ensure the accuracy without outliers or anomalies that occur in one season. By pulling two years across the league, we can identify consistent patterns and trends in pitcher performances to provide a more reliable projection. We have captured key performance indicators across these two seasons that will work as our variables to help analyze our findings.

Earned Run Average (ERA): A measure of a pitcher's effectiveness in preventing earned runs amongst the number of innings pitched.

Walks plus Hits per Inning Pitched (WHIP): A crucial metric for evaluating how often a pitcher allows a base runner per inning.

Strikeout Ratio (SO): A measure of a pitcher's ability to retire batters via strikeouts, which will demonstrate a key indicator of dominance on the pitching mound.

Innings Pitched (IP): A measure of the number of innings that a pitcher has covered that provides us with a pitcher's workload and durability.

Hits (H): A measure of the number of hits that a pitcher allows throughout the season.

The purpose of pulling all these indicators stems from the project's overall objective in determining the optimal use of relief pitchers versus starting pitchers to maximize team performance. Once we begin analyzing stats of pitchers against these measures, we will be able to provide more insight that can inform us on better in-game decisions to improve the teams ERA, WHIP, and strikeout ratio.

Data Filtering and Cleaning: After acquiring the data, we then began filtering and cleaning out the initial data pull to ensure the accuracy and relevance of the dataset. Since there are occasions where MLB players are designated for assignment (sent to Minor Leagues) or rotated from the bullpen to the starting rotation, we wanted to ensure that we were pulling only the necessary data to avoid outliers in our analysis. We first started with the number of games (innings) that a pitcher was able to contribute during the season. Since relief pitchers at times only accounted for 1 inning per game, we have decided to set the limit that a pitcher needs to attribute more than 20.0 innings per game to ensure we are capturing Closers that get traded or come in mid-season. This threshold allows us to remove pitchers with limited appearances who had injuries or moved up/down from the Minors. By focusing on players with 20.0 innings or more it allows us to attribute to a more reliable and consistent analysis amongst pitchers in the league. Additionally, we have removed any incomplete records, outliers, null data, or inconsistencies based on data type that would skew our results. It regarded moving unnecessary attributes, incorrect data entry, and meeting our innings criteria. This enables our project to maintain a high level of data integrity and relevance for our results.

Appropriateness of Methods: The methods that we have used to filter and clean our dataset ensure that the pitchers that we are focusing on provide a more significant impact on their teams and overall game. These thresholds help us eliminate any outliers and remove any inconsistencies. This step is essential in maintaining the integrity of our dataset and ensuring the results are accurate and complete. Aligning these filters and criteria without variables are all directly tied to our project's objective in optimizing our pitching performance and decision making. This focus ensures that our analysis is relevant and provides insightful decisions to help improve team performance. We plan to implement different methods such as regression analysis and decision trees to allow for a more robust examination of relationships that occur between our metrics and pitching roles. The methods utilized will be essential for delivering a thorough and comprehensive analysis that can guide our strategic decision making and optimize our pitching rotation.

Results

Initial Trends: As we began looking at the data sets that we have pulled for the 2022 and 2023 MLB pitchers, we began analyzing at a closer look each season separately to begin understanding the current differences amongst starting pitchers and relief pitchers. Within the 2022 MLB season, we saw an increase in ERA, WHIP, and a decrease in Strikeouts amongst relief pitchers to starting pitchers. It was interesting to see this trend as we had expected starting pitchers to have a higher ERA or WHIP since these pitchers are enduring a longer season and

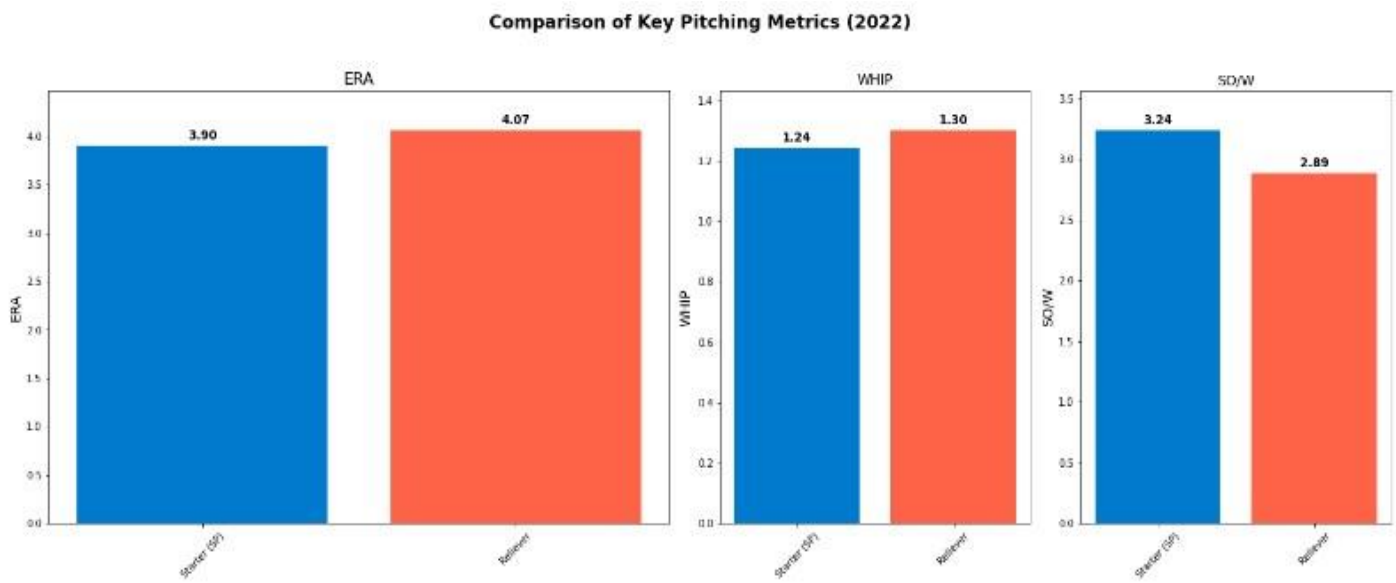


Figure 1 Pitcher Comparison 2022 Regular Season

would have expected them to see a decrease in strikeout and an increase in ERA. Usually when a pitcher is experiencing an over usage of pitches or have higher pitch counts, they begin to reduce in velocity and trajectory of their pitches which would result in errors occurring but that didn't seem to occur here. We realized the same trend occurred in 2023 which can be seen below.

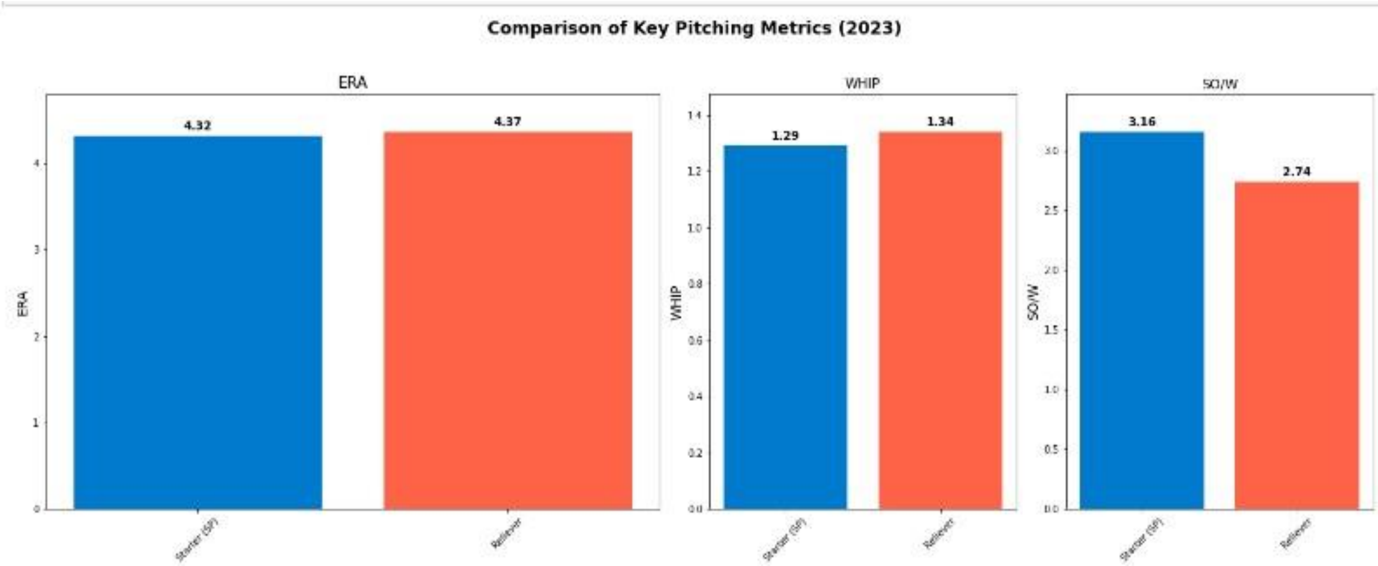


Figure 2 Pitcher Comparison 2023 Regular Season

Overall, we saw an increase in every variable but specifically saw the relievers increase in ERA, WHIP, and decrease in Strikeouts the most. We believe that individuals from minor leagues coming to the Majors could be a result of these poor ERA's, WHIP score, and strikeouts. It even occurred that most of the time relievers come into a game with base runners on which results in higher ERAs against their performance. As we looked at the initial trends we began to take a deeper dive of efficiency per inning for each type of pitcher.

Strikeout Efficiency: We began to take a closer look at each pitcher's efficiency per inning and exactly how well each pitcher contributes to a game. In the diagram below, we looked at Starting Pitchers and Relief Pitchers strikeouts per inning to give us an example of how well each pitcher is performing. We notice immediately that most of the time Relief pitchers are only pitching one to two innings resulting in one to two strikeouts per inning. This demonstrates to us that relief pitchers are more efficient if they are used for only an inning or two to usually close out the games or pressured scenarios. The more time they stayed on the field it resulted in a lower strikeout to inning ratio that made them less efficient in the long run. For starting pitchers on the other hand they become more efficient the longer they stay on the field, usually around 5 to 7 innings before they are rotated out during the game. This could be since Starting Pitchers face more batters and have more opportunities to increase their strikeout efficiency whereas relief pitchers are meant to perform at a peak level within one to two innings before resulting in a rotation. From this analysis we then wanted to investigate the ERA efficiency of the two roles pitchers to determine how well they performed against the increased innings.

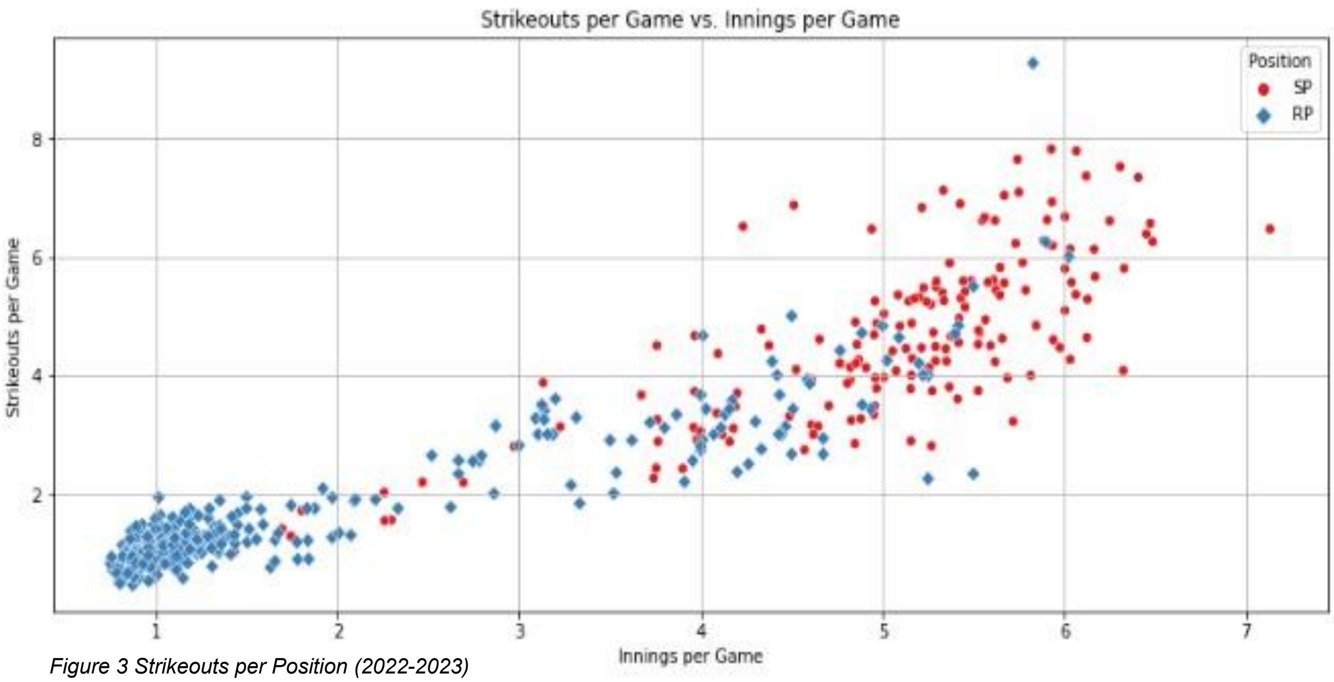


Figure 3 Strikeouts per Position (2022-2023)

ERA Efficiency: As we moved on to player ERAs based on the number of innings each pitching role pitches, we noticed a trend that occurs amongst the pitchers. In the diagram below, we can

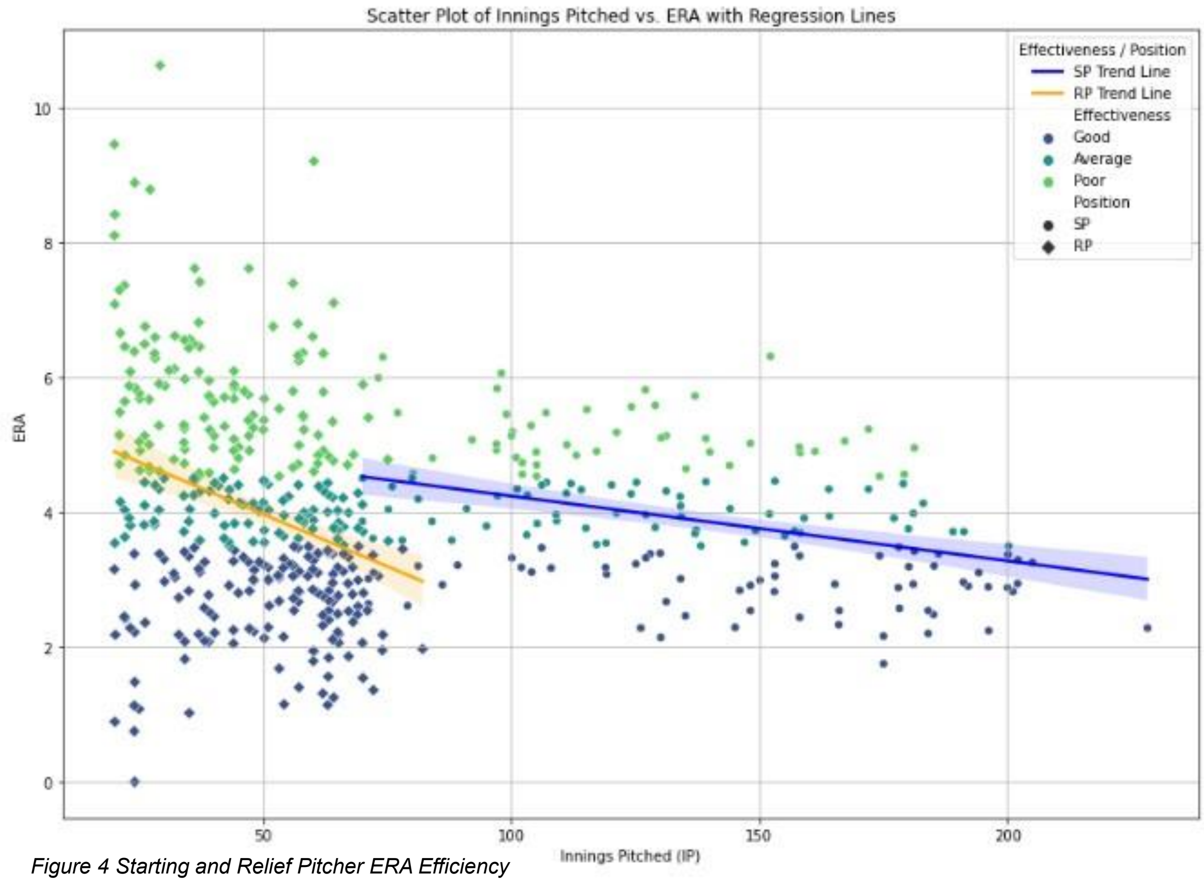


Figure 4 Starting and Relief Pitcher ERA Efficiency

see that relief pitchers and starting pitchers see a regression in ERA the longer they are asked to pitch during the length of a season. We noticed that the main cause for relief pitchers having such a high ERA amongst starting pitchers are usually due to minor league players that are usually assigned and getting their first take glance within the major league. The length of innings demonstrates that pitchers used briefly in a season cause the entire average of ERAs to increase for relief pitchers, but closers and actual relievers in the majors see a more drastic decrease in ERA compared to starting pitchers. In the same chart we incorporated starting pitchers' length of innings to ERA and see the same regression except at a lengthier time frame. Starting pitchers don’t drop as drastically as relief pitchers which could give the implication that relievers perform better in a shorter period. This informs us that relievers can perform in a more specific high pressure scenario at the end of the game compared to starting pitchers. We could probably get a better look at future results by including more specific situations and runners on base to determine exactly how well relievers perform compared to starting pitchers. The drastic regression line for relief pitchers compared to starters really dictates how well each pitcher performs over the entire season. Another aspect we would consider in the future is to increase our innings pitched to remove any individuals that are brought up from the minors that could be altering the overall ERA and Strikeout efficiency of relievers or even getting a dataset that would include relief pitchers that are set for the majors all season versus pitchers being moved up from minors. Overall, this would be interesting to see more accurate and specific models present data that could back up our case of when to decrease pitching ERAs and increase pitcher performance.

Decision Tree Model: In our previous two analyses we looked at current trends that have occurred in the past two years amongst starting pitchers and relief pitchers. We then looked at how well each pitcher was performing based on their ERA and Strikeout ratio within games throughout the regular season. Now, we have taken a closer look at a decision tree model based on the datasets that we have collected to determine exact scenarios of when it would be more efficient to use relief pitchers to starting pitchers. The model considers both positions per game averages of innings pitched (IP), runs allowed (R), and home runs allowed (HR). The model is then split into two training and test sets that achieve a strong accuracy score against our model.

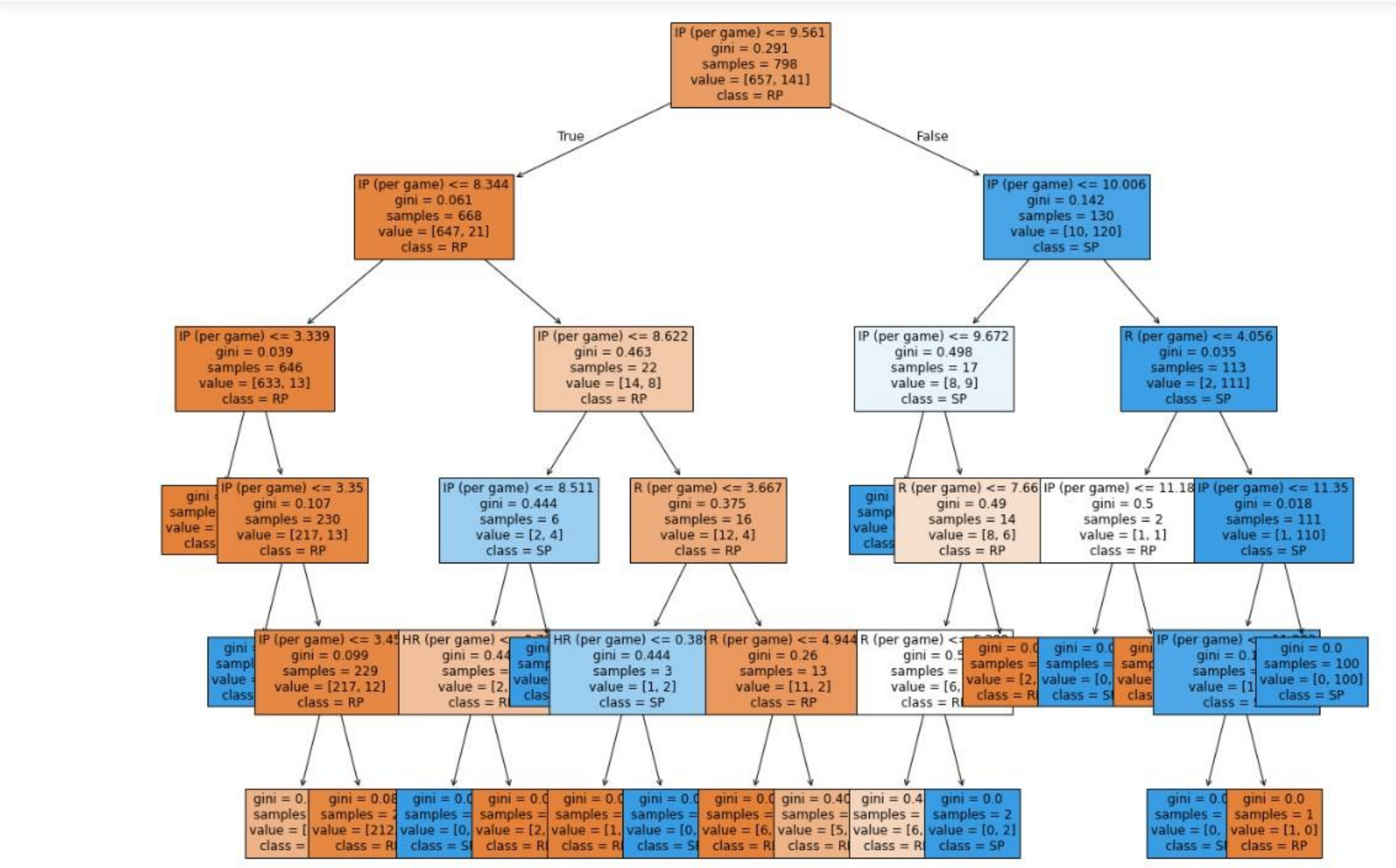


Figure 5 Decision Model on Pitcher Scenarios

Key Results:

1. The model accurately distinguishes which pitchers to use in specific scenarios between SPs and RPs based on the amount of runs or innings pitched during a game
2. The Decision tree illustrates the critical decision points such as the number of innings to runs already allowed or the number of home runs given up by specific pitchers to determine if a rotation is needed even amongst relievers.

Improvements:

1. To improve our current state, we believe that adding more pitching metrics (strikeouts, WHIP, Walks, etc.) could provide more contextual factors so that the decision tree can provide more specific scenarios and accurate decisions.
2. Experiment with different tree depths and consider using more advanced models like Random Forests. This was a preliminary model for data we were able to pull from the past two seasons. Incorporating more seasons or even playoffs would help provide more outcomes that could be beneficial for a team's overall performance.
3. Incorporate more seasons of data and balance the dataset to avoid bias. This model is a solid foundation, and with these improvements, it can provide even more precise insights into pitcher roles in the future.

Conclusion/Recommendations

This project has provided significant insights into optimal pitching management from analyzing traditional metrics including ERA, WHIP, SO, IP, and Hits Allowed. We can use this analysis to guide our decision-making strategies on utilizing pitchers to improve effectiveness and prevention of long-term injuries throughout the regular season. We found that using relief pitchers are most effective when used for short (one to two innings), high-leverage situations. By doing this we can maximize their strikeout efficiency without causing fatigue due to extended use. Strategies for optimal use of our starting pitchers require careful monitoring as they pitch deeper into the game. Our data shows that there is a significant decline in effectiveness when pitching to a lineup for the third time. Knowing when to go to the bullpen will be critical for pitching performance and health.

Preventing injuries was an important focus of this project to have sustained success throughout a long regular season. As we closely examined metrics such as IP, Hits Allowed, and

SO we found its crucial for early identification of signs that a pitcher is fading. Pitcher fatigue and decline in performance are critical indicators for injury risk that we need to identify. We can manage a pitcher's workload proactively by creating thresholds based on the data. This will help reduce the likelihood of injuries and maintain pitcher effectiveness throughout the entire season. Keeping our pitching staff healthy can be accomplished by managing pitch counts and ensuring ample rest time between outings.

Our analysis found an importance on strategically utilizing relief pitchers in situation specific, high-leverage scenarios. Bringing in a reliver in late-game critical situations, utilizes their strengths to get outs without allowing baserunners. To guide decision making we developed visual representations like the graph showing strikeout efficiency over innings pitched and the decision tree model for substitutions. These provide clear, actionable insights on how this analysis can be utilized to assist management strategies during a game.

The methods used for this project were chosen for their ability to provide insights into managing pitchers. To ensure that we provided clear and usable findings, we focused on traditional metrics that apply to our team. To better understand how extended innings effect pitcher performance we used regression analysis and efficiency metrics. For in-game pitching decisions we used the decision tree model to translate complex data into practical insights that can be used in management strategies.

Based on our findings we recommended that coaching staff use strategies driven by data to enhance team success. Using the decision tree model and monitoring key metrics will aid optimal pitching management leading to improved performance and stable health of our pitching staff. Continuing to monitor data throughout the season is important to support the reliability of our findings and maintain a competitive advantage. Further analysis can be conducted to expand on our solid foundation of findings by refining the models and exploring more statistical measures for analysis. Finally, to maintain the long-term health of our pitching staff we need to monitor workload to keep them available throughout the season. Following these recommendations will maximize pitcher effectiveness and utilization for sustained success of our team.

References:

"2022 Major League Baseball Standard Pitching." Baseball. Accessed August 19, 2024. https://www.baseball-reference.com/leagues/majors/2022-standard-pitching.shtml.

"2023 Major League Baseball Standard Pitching." Baseball. Accessed August 19, 2024. https://www.baseball-reference.com/leagues/majors/2023-standard-pitching.shtml.

Winston, W. L., Nestler, S., & Pelechrinis, K. (2022). *Mathletics: How gamblers, managers, and fans use mathematics in sports*. Princeton University Press.

Loinaz, X. (2018, January 17). *A brief analysis of predictive pitching metrics*. Community Blog. https://community.fangraphs.com/a-brief-analysis-of-predictive-pitching-metrics/

Dunbar, P. by J. (2020, May 31). *Study: What are the best metrics for evaluating a pitcher?*. MVP Sports Talk. https://mvpsportstalk.com/2020/05/20/study-what-are-the-bestmetrics-for-evaluating-a-pitcher/

Appendix:

Figure 1: 2022 Pitching Metrics

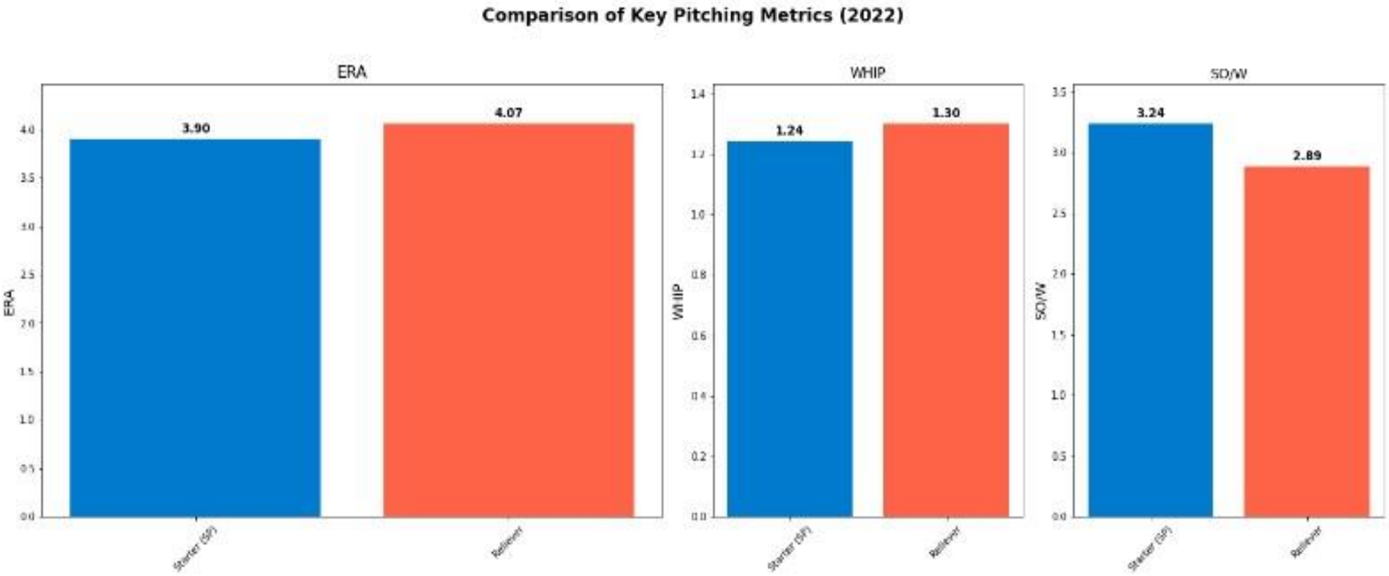


Figure 2: 2023 Pitching Metrics

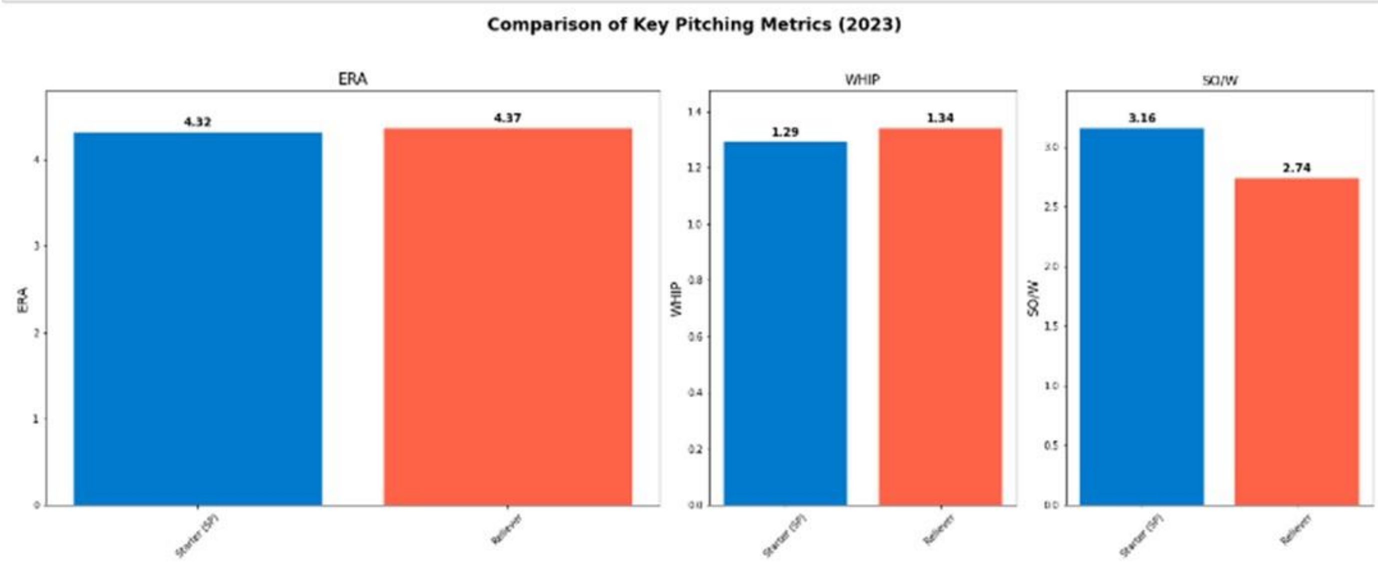


Figure 3: Strikeouts per Position

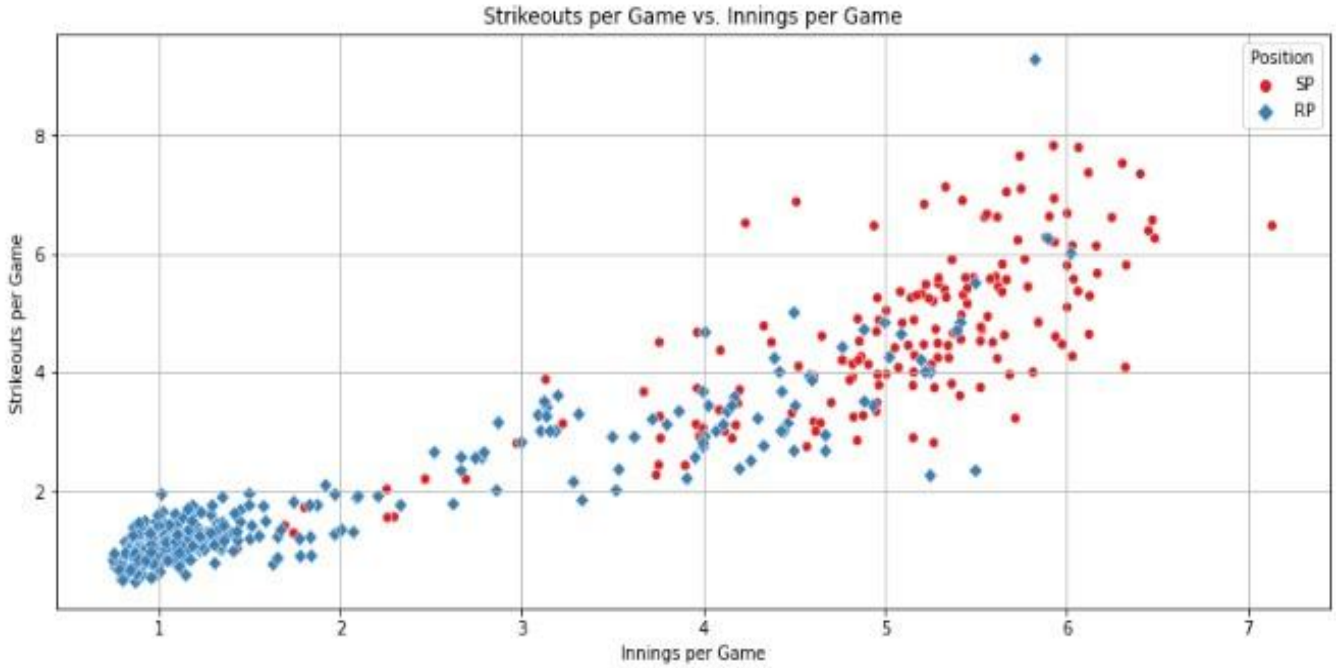


Figure 4: Starting and Relief Pitcher ERA Efficiency

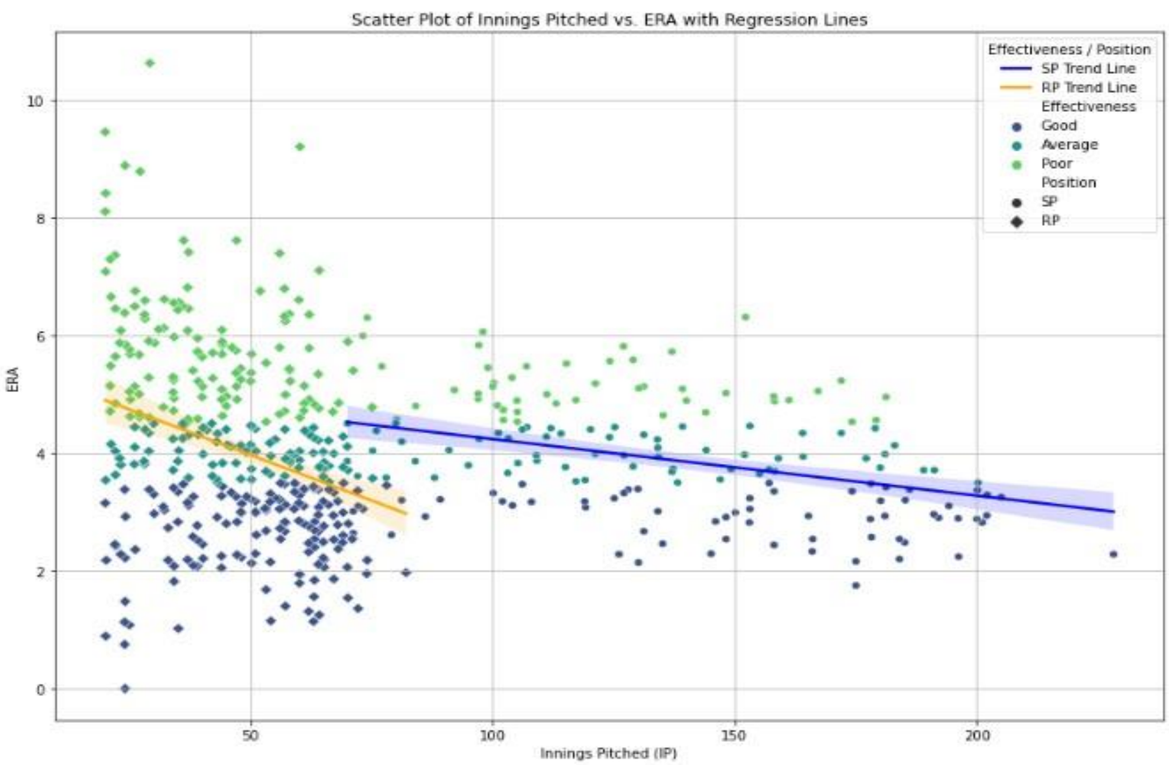
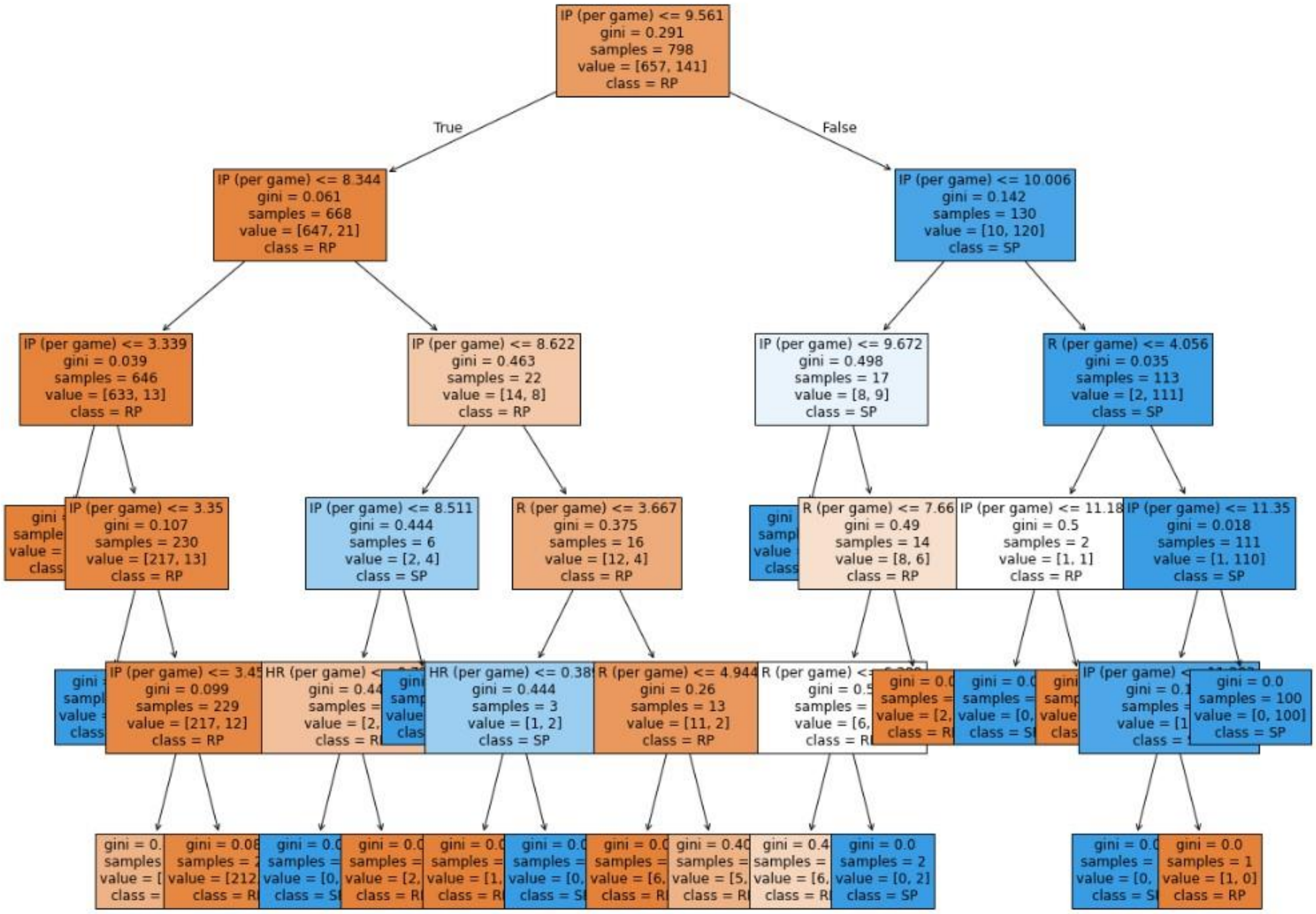


Figure 5: Decision Model on Pitcher Scenarios


IP (per game) <= 9.561
gini = 0.291
samples = 798
value = [657, 141]
class = RP
True
False
IP (per game) <= 8.344
gini = 0.061
samples = 668
value = [647, 21]
class = RP
IP (per game) <= 10.006
gini = 0.142
samples = 130
value = [10, 120]
class = SP
IP (per game) <= 3.339
gini = 0.039
samples = 646
value = [633, 13]
class = RP
IP (per game) <= 8.622
gini = 0.463
samples = 22
value = [14, 8]
class = RP
IP (per game) <= 9.672
gini = 0.498
samples = 17
value = [8, 9]
class = SP
R (per game) <= 4.056
gini = 0.035
samples = 113
value = [2, 111]
class = SP
IP (per game) <= 3.35
gini = 0.107
samples = 230
value = [217, 13]
class = RP
IP (per game) <= 8.511
gini = 0.444
samples = 6
value = [2, 4]
class = SP
R (per game) <= 3.667
gini = 0.375
samples = 16
value = [12, 4]
class = RP
gini = 0.49
samples = 14
value = [8, 6]
class = RP
gini = 0.5
samples = 2
value = [1, 1]
class = RP
IP (per game) <= 11.35
gini = 0.018
samples = 111
value = [1, 110]
class = SP
gini = 0.099
samples = 229
value = [217, 12]
class = RP
gini = 0.444
samples = 3
value = [1, 2]
class = SP
R (per game) <= 4.944
gini = 0.26
samples = 13
value = [11, 2]
class = RP
gini = 0.0
samples = 100
value = [0, 100]
class = SP
gini = 0.0
samples = 2
value = [0, 2]
class = SP
gini = 0.0
samples = 1
value = [1, 0]
class = RP