Laken Rivet

Sports Performance Analytics

March 12, 2023

Assessing MLB Team Performance

Part One: Updated Pythagorean Wins Formula

The first step in determining an updated Pythagorean wins formula was acquiring data from Sean Lahman's Baseball Database. This was done via his website, seanlahman.com, where a CSV file containing yearly team data for all major league teams from 1871 to 2022 was downloaded. Once the file was read into the software, R, data filtering took place. First, the data was filtered to include team information for the years 2012 to 2022. The decision to include only the past ten years of data was made to ensure that the updated formula coincides with recent playing styles. Next, the data was filtered to include only the variables necessary for formula determination which were team, year, league, number of games played, wins, losses, runs, and runs allowed. The data was imported in a format that allowed for computation as is, so data cleaning was not required.

The next step in determining an updated formula was exploring the relationship between team performance and wins. To measure overall team performance, a run differential variable was created by subtracting runs allowed from runs scored for each team. To remain consistent with typical baseball terminology, rather than look at wins, winning percentages were calculated. This was done by dividing wins by the sum of wins and losses, or total games played. The relationship between winning percentage and run differential is visualized is Figure 1.

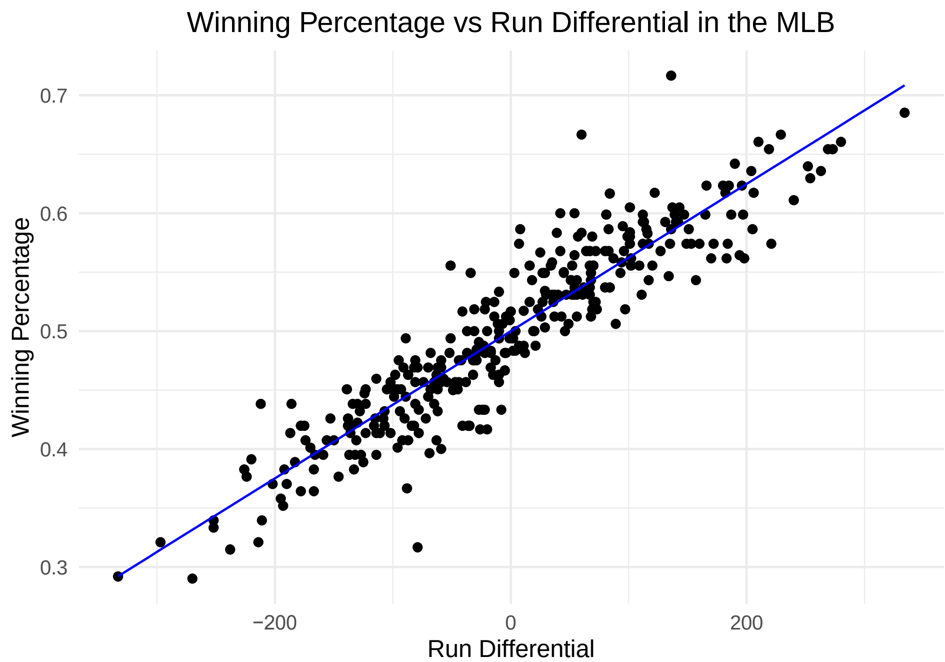

Figure 1. Winning percentage plotted against run differential for all teams in the MLB from 2012 to 2022.

Visually, it's evident there is a strong, positive linear relationship between winning percentage and run differential. To further confirm this observation, the correlation coefficient was calculated and resulted in a value of 0.921. Because such a strong linear relationship was observed, a linear regression line was calculated as well. The equation for the line can be seen below:

$$Wpct = 0.499979 + 0.000624 \times RD$$

Where 'Wpct' is winning percentage and 'RD' is run differential. The line is plotted in blue in Figure 1. The regression formula acts logically as a team with a run differential of zero is expected to win around half of their games. However, there is a serious limitation in the fact that a win percentage can be above one, which doesn't happen in the real world. After gaining a deeper understanding of the relationship between winning percentage and run differential across the entire MLB, comparisons of correlation coefficients and regression formulas were done between the two leagues, the National League (NL) and the American League (AL). The results can be found in Table 1.

League	Correlation Coefficient	Regression Line
American	0.918	$Wpct = 0.5005211 + 0.0006106 \times RD$
National	0.925	$Wpct = 0.499539 + 0.000639 \times RD$

Table 1. Comparison of correlation coefficients and regression lines between teams in the National League and American League from 2012 to 2022.

From the results in Table 1, it can be observed that winning percentage and run differential are slightly more correlated for teams in the NL compared to teams in the AL. The regression lines further support this observation as a larger coefficient value is assigned to the run differential in the NL. These results are logical as historically, the teams in the AL consistently produce more home runs per game than the teams in the NL (Rymer 2013). Thus, in the NL where there are less home runs overall, they are more valuable to winning. While comparing the leagues provides insight into the differences between them, their correlation coefficients and linear regression lines are similar enough that the updated Pythagorean wins formula should be determined using teams from both leagues. The standard Pythagorean wins equation can be seen below:

$$Wpct = \frac{R^2}{R^2 + RA^2}$$

Where 'Wpct' is winning percentage, 'R' is runs scored, and 'RA' is runs allowed. Prior to determining the exponent that results in the lowest prediction error, the above equation was used to calculate winning percentages for the data set. After comparison of the estimated winning percentages and actual winning percentages, the root mean square error (RMSE) was calculated to be 0.0274. To identify an alternative exponent to use the equation, a linear regression with a response variable of the log of the ratio of wins to losses and a response variable of the log of the ratio of runs scored to runs allowed was performed. The regression identified the ideal exponent as 1.783. This result was confirmed as the RMSE was calculated to be 0.0260. Thus, the equation for updated Pythagorean wins formula is:

$$Wpct = \frac{R^{1.783}}{R^{1.783} + RA^{1.783}}$$

According to this equation, in the 2022 season, the San Francisco Giants were predicted to have a winning percentage of 0.51. Our actual winning percentage for the season was 0.50, so we slightly underperformed compared to what the updated Pythagorean wins formula predicted.

Part Two: Ranking Players Compared To The League

To identify the three players on the San Francisco Giants with the most at-bats for the 2022 season, Sean Lahman's Baseball Database was utilized. Again, a CSV file was downloaded from his website, seanlahman.com, that contained batting data for all players in the MLB from 1871 to 2022. The data was filtered to include only the 2022 season and players from the San Francisco Giants. This led to identifying the players with the most at-bats to be Wilmer Flores, Thairo Estrada, and Mike Yastrzemski with 525, 488, and 485 at-bats, respectively. Another data source, fangraphs.com, was used to obtain advanced hitting data. A total of 150 of the highest performing hitters in the MLB for the 2022 season, based on Wins Above Replacement (WAR), were included in the advanced hitting dataset, which was downloaded in CSV format. The decision to include only the leading hitters in the league was made for two reasons; first, it's more valuable for our organization to see how our star hitters compare to the star hitters on other teams. That is, for our organization to have a competitive edge against others, we need to be able to understand the strengths and weaknesses of their best players. The second reason is certain metrics require a very large sample size, which in this case would be the number of plate appearances or at-bats, to be accurate (Slowinski 2010c). Because the leaders of the league hit the most, their statistics will be the most reliable.

The four metrics selected to evaluate hitters come from three different types of data - seasonal, play-by-play, and pitch-by-pitch. The statistic based on seasonal data is weighted on-base average (wOBA). wOBA is like on-base plus slugging (OPS) in that it considers many different aspects of hitting, except wOBA attempts to credit the hitter based on the value of each outcome with the value being how well the player contributes to run scoring (Slowinski 2010a). Essentially, wOBA helps us to evaluate our hitter's offensive value to the team in terms of run scoring. The wOBA statistics for Flores, Estrada, and Yastrzemski can be found in Table 2, as well as their league-wide ranking for wOBA in the 2022 season.

Name	AB	wOBA	wOBA Rank
Wilmer Flores	525	0.312	94
Thairo Estrada	488	0.317	82
Mike Yastrzemski	485	0.307	108

Table 2. wOBA statistics and rankings for the three players with the most At-Bats (AB) on the San Francisco Giants in the 2022 season.

Our hitters rank poorly in wOBA compared to the rest of the data set, which is confirmed by the visualization in Figure 2.

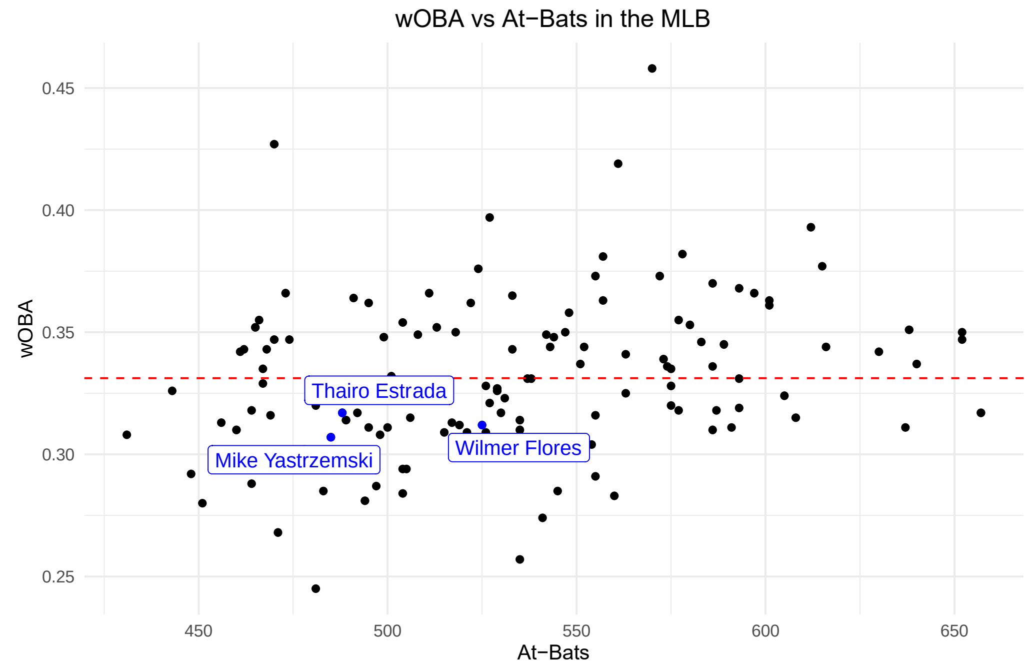

Figure 2. wOBA plotted against At-Bats (AB) for the top 150 hitters in the 2022 season.

The red dashed line in Figure 2 is the mean wOBA for hitters in the data set. Thus, all our most frequent hitters are performing below-average. The play-by-play metric to be evaluated is run expectancy based on 24 base-out states (RE24). As the name suggests, this metric is based on a run expectancy matrix for the 24 base-out states of a baseball game. It measures the difference in run expectancy before and after a plate appearance and then credits the hitter with that difference. RE24 is a great hitting metric because it is context-dependent, meaning more credit is assigned to players who produce a hit with players on base rather than empty (Weinberg 2014). Ultimately, RE24 helps us gauge how well a player capitalizes on situations that can lead to scoring. The 2022 RE24 statistics and rankings for our most frequent hitters can be seen in Table 3.

Name	AB	RE24	RE24 Rank
Wilmer Flores	525	15.21	52
Thairo Estrada	488	10.72	62
Mike Yastrzemski	485	-5.18	106

Table 3. RE24 statistics and rankings for the three players with the most At-Bats (AB) on the San Francisco Giants in the 2022 season.

Overall, our rankings in RE24 are much better than wOBA. However, there is a glaring issue with the fact that Yastrzemski has a negative value for the season. This tells us that he is producing less runs than would be expected in game situations. The visualization in Figure 3 further highlights this disparity.

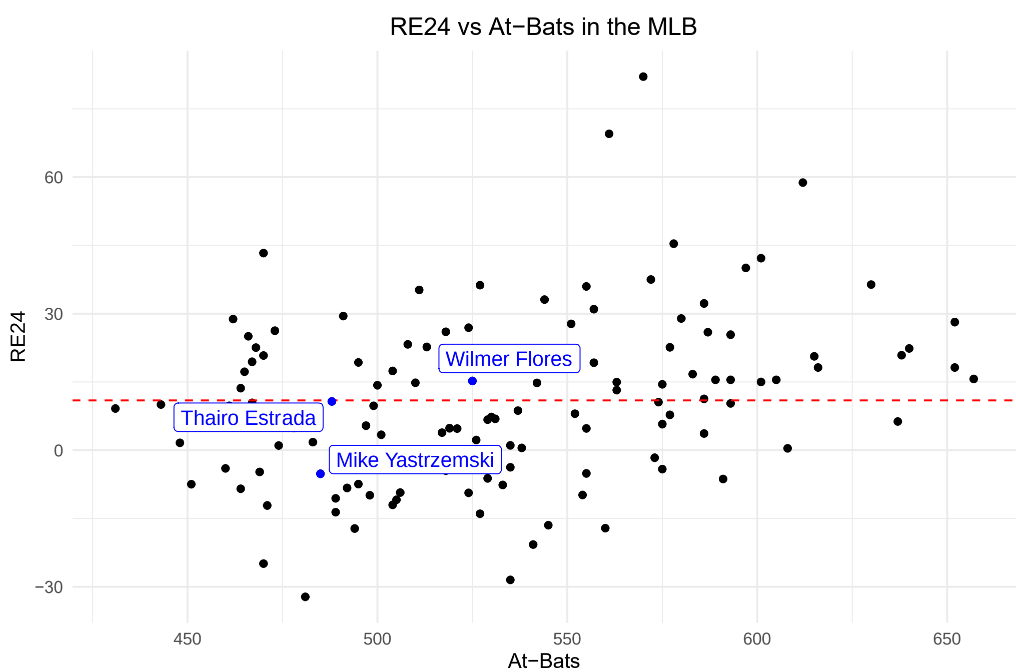

Figure 3. RE24 plotted against At-Bats (AB) for the top 150 hitters in the 2022 season.

As with Figure 2, the red dashed line represents the mean RE24 for the data set. As such, Flores is performing above average while Estrada is sitting nearly exactly on the mean. Yastrzemski is far below his teammates as he is falling into the negative ratings. The final two metrics, O-Swing percentage (O-Swing%) and Z-Swing percentage (Z-Swing%) are retrieved from pitch-by-pitch data. O-Swing% is the number of swings at pitches outside the zone over the number of pitches outside the zone. Z-Swing% is the number of swings at pitches inside the zone over the number of pitches inside the zone. Both metrics are categorized under “plate discipline”, essentially measuring how well a hitter can differentiate between pitches in or out of the zone and how aggressive they are in swinging at them (Slowinski 2010b). Note that a low O-Swing% indicates a very disciplined hitter, for this reason throughout this report a low O-Swing% is considered ideal. Because of this, the ranking system is reversed for O-Swing% where lower values are ranked higher. Z-Swing% is ranked like all other metrics where a higher value is ranked higher. The O-Swing% and Z-Swing% statistics for the top Giants hitters can be found in Table 4.

Name	AB	O-Swing %	O-Swing % Rank	Z-Swing %	Z-Swing % Rank
Wilmer Flores	525	30.7	48	68.1	76
Thairo Estrada	488	35.5	97	68.7	67
Mike Yastrzemski	485	25.4	16	64	107

Table 4. O-Swing Percentage statistics and rankings for the three players with the most At-Bats (AB) on the San Francisco Giants in the 2022 season.

It’s in O-Swing% that Yastrzemski obtains the Giants’ highest rating for all four metrics. It can be observed how far above Yastrzemski is below the mean value of O-Swing% in Figure 4. Again, note that the y-axis is reversed in Figure 4 to account for the fact that lower O-Swing% is ideal.

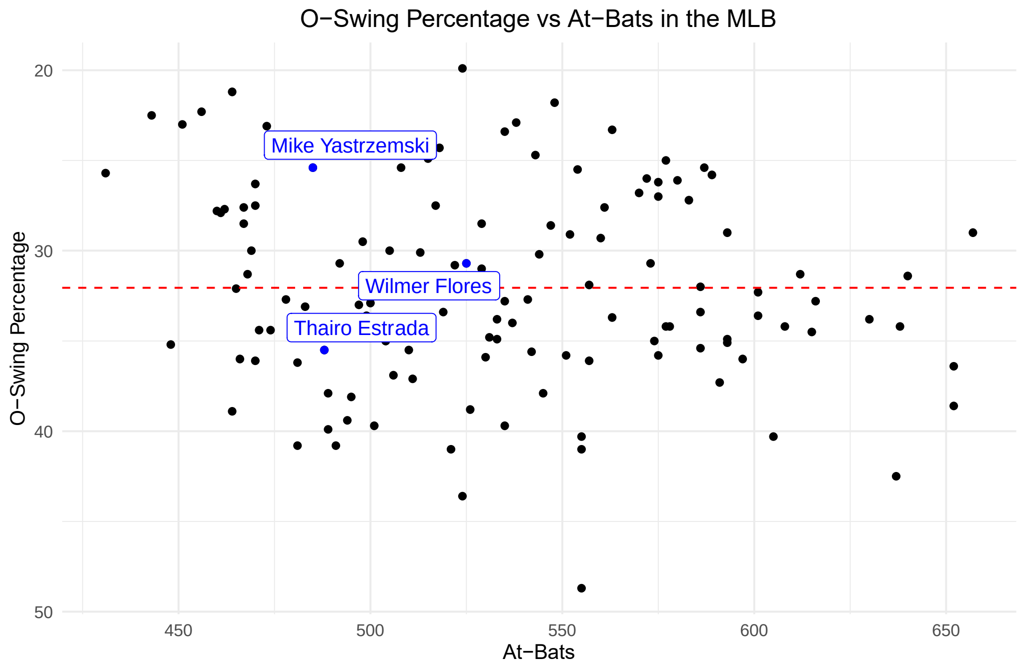

Figure 4. O-Swing Percentage plotted against At-Bats (AB) for the top 150 hitters in the 2022 season. Note the reversed y-axis.

For Z-Swing%, all identified Giants players are performing below-average which can be seen in Figure 5.

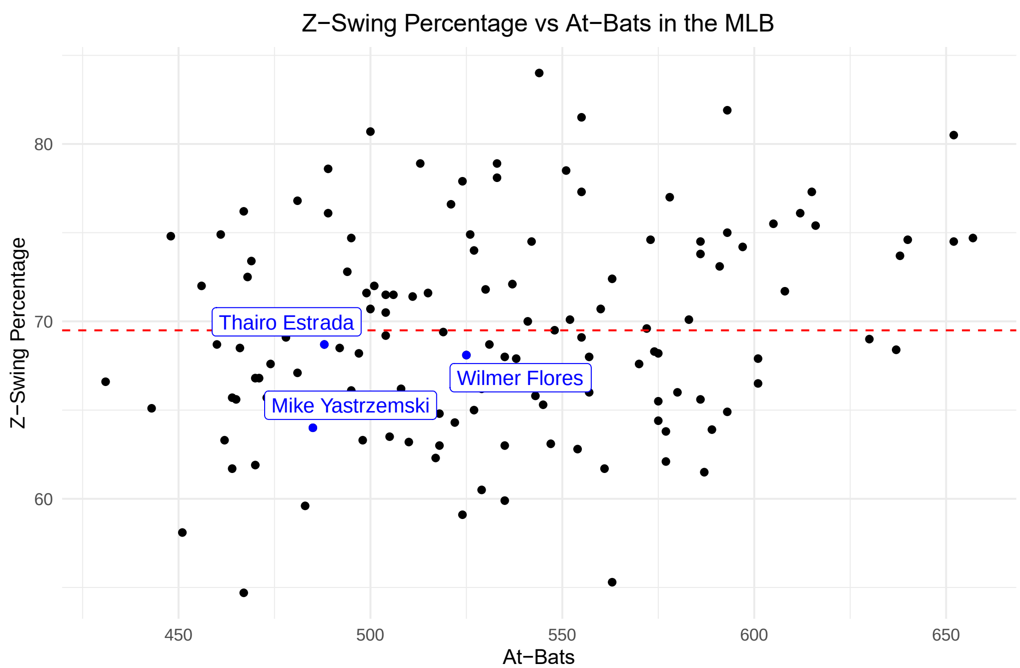

Figure 5. Z-Swing Percentage plotted against At-Bats (AB) for the top 150 hitters in the 2022 season.

When considering what these metrics suggest about our top hitters, we must consider the combination of statistics for each player. Beginning with Wilmer Flores, the San Francisco player with the most at-bats in the 2022 season, his wOBA value of 0.312 suggests he is not contributing as much to scoring as we'd hope. However, his impressive RE24 of 15.21 says that when he does contribute, it's at key moments that have a large impact in a game. According to his O-Swing% of 30.7, he's a disciplined hitter that can identify if a ball is out of the zone fairly well. His Z-Swing% of 68.1 says the same is true for balls in the zone. However, in comparison to other players in the league, he falls short on Z-Swing%. Overall, Flores is a solid hitter who needs to focus on improving his Z-Swing% for next season. If he's able to swing and make meaningful contact with pitches in the zone more often, that will improve his wOBA as well.

Next, Thairo Estrada had the second most at-bats for the 2022 season. While he received the highest wOBA of the three players at 0.317, it's still lower than we'd hope of our best hitters. Estrada's RE24 of 10.72 makes for an interesting comparison between him and Flores. Estrada's higher wOBA and lower RE24 suggests that he contributes to scoring more than Flores, but when he does it's less impactful. With an O-Swing% of 35.5 and a Z-Swing% of 68.7, Estrada is better at identifying pitches in the zone rather than out. Having the highest percentages in both categories suggests he's the most aggressive hitter of the three. With that, he should focus on improving his O-Swing% for next season. Improving his plate discipline will force pitchers to either walk him or throw in the zone, which is something that would play in his favor.

Finally, Mike Yastrzemski had the third most at-bats for the 2022 season. His wOBA is comparable to Flores and Estrada but is notably the lowest at 0.307. Yastrzemski's greatest weakness is his RE24 at -5.18. Essentially, what we learn from this is that he is not creating as many runs as we'd expect for a player in a given situation. Yastrzemski appears to make up for this weakness by having a phenomenal O-Swing% of 25.4. From his O-Swing%, we'd assume he's a very well-disciplined hitter, yet his Z-Swing% is the worst of the three at 64. The trend of having lower percentages for O-Swing% and Z-Swing% appears to show that Yastrzemski is a less aggressive hitter in general. For next season, he should focus on swinging more often and confidently. While his O-Swing% may suffer slightly, he will be able to improve his overall offensive statistics such as wOBA and RE24.

Part Three: Runs Added by Replacing a Player

Prior to evaluating which of the aforementioned players should be replaced and by whom, additional data was collected. A CSV file from Sean Lahman's Baseball Database containing demographic information for all players was downloaded and read into R. From this dataset the player ID, name, birth year, eight, height, handedness of batting, and handedness of throwing were extracted and joined to the batting dataset, previously mentioned in Part Two, by player ID. Next this dataset was filtered to only include player batting and demographic information for the 2022 MLB season. A new dataset was then created from the annual team dataset mentioned in Part One. The new dataset was filtered to only include team data from the 2012 to 2022 seasons, as well as the variables year, league, team, at bats, games, wins, losses, runs, hits, doubles, triples, homeruns, walks, batters hit by pitch, stolen base, caught stealing, outs pitched, and errors. Two additional variables were created, singles and walks. Singles was created by subtracting the sum of doubles, triples, and home runs from total hits. Walks was created by adding walks and batters hit by pitch.

To determine which player to replace on our roster, runs created above average was calculated following the method outlined in chapter 3 of *Mathletics* (Winston, Nestler, and Pelechrinis 2022). The first step in calculating runs created above average, which essentially measures how many runs a player adds to an average MLB team, is creating a linear regression model to predict runs based on seasonal team hitting data. The model includes singles (X1B), doubles (X2B), triples (X3B), home runs (HR), walks (walks + batters hit by pitch, WLKS), stolen bases (SB), and caught stealing (CS) as predictor variables. Note that in *Mathletics* the model did not include caught stealing due an insignificant p-value. However, caught stealing was found to have a significant p-value in this analysis and thus was kept in the model. The model was trained on the newly created dataset containing team hitting data for the most recent ten seasons, this decision was made to ensure the model would be accurate to playing trends in recent years. The resulting regression equation can be seen below:

$$R = -5.65 + 0.18(X1B) + 0.78(X2B) + 0.83(X3B) + 1.27(HR) + 0.19(WLKS) + 0.26(SB) - 1.2(CS)$$

Next, the data for an average MLB team was generated. This was done by averaging the statistics in the 2012-2022 team hitting dataset. In addition to determining the average predictor variables for the model, we also needed to determine the average number of outs for a season. This was done by averaging the outs pitched variable, which is equivalent to innings pitched multiplied by 3. The statistics of the average MLB team data were then run through the model to predict the number of runs. Once we knew how many runs an average team would be predicted to score, we then had to determine how many runs they would score by adding a specific player to the team. To do this, the individual player statistics were retrieved from the individual batting dataset. Outs "created" by an individual player also needed to be calculated. The method outlined in chapter 2 of *Mathletics* was used to do so. This involved determining the average error rate at bat by dividing the sum of errors by the sum of at bats in the team hitting data set. The sum of the player's hits, grounding into double plays, sacrifice flys, sacrifice hits, and caught stealing were then subtracted from the product of their at bats and the aforementioned error rate. Finally, the resulting number was then divided by the average number of outs in a game. The average number of outs in a game was calculated by dividing outs pitched by number of games.

To add an individual player's statistics to a team, an individual multiplier had to be calculated. This was done by subtracting the player's predicted outs from the total average team's outs, then dividing by the total average team's outs. The new team statistics were then calculated by adding the player's statistics to the product of the team statistics and the calculated multiplier. For example, to calculate the new team's singles the player's seasonal singles would be added to the product of the multiplier and the average team's singles. This was repeated for all predictor variables necessary to the model. Once all the statistics for the team with a particular player added were calculated, they were run through the model to predict runs scored. Finally, the number of predicted runs of an average team was subtracted from the number of predicted runs of an average team with a particular player to estimate how many runs that player added to the team's total.

This procedure was performed for Flores, Estrada, and Yastrzemski. The results indicated that Flores and Yastrzemski would add 5.09 and 6.18 runs to an average MLB team, respectively. Estrada, on the other hand, was found to reduce the average MLB team's predicted runs by 2.4. As such, it was determined that Estrada should be replaced on the roster. Estrada plays both the second basemen and shortstop positions and has a current AAV of $2.3M. In looking for potential players to add to the roster, I was cognizant of the fact that acquiring a player who will contribute to more wins will likely cost the organization more money. With that in mind, I also attempted to find players who had AAVs of a relatively similar size. Additionally, rather than look exclusively for a player who plays both second base and shortstop like Estrada, I also extended the search to players who solely play second base. This decision was made as Brandon Crawford is our primary shortstop and the new addition would rarely be played over him. The search resulted in three options to be evaluated: Kolten Wong, Gleyber Torres, and Jake Cronenworth.

Before delving into the details of each player, I first wanted to outline how I determined the impact a new player would have on the team. First, as with our own players, I calculated the runs created above average with the potential new player's 2022 seasonal hitting data. To take this analysis a step further and translate a roster change to winning, I then created three new versions of the 2022 seasonal data. In all versions I removed Estrada's statistics and added Wong's, Torres', and Cronenworth's statistics to the first, second, and third versions, respectively. All three versions as well as the original team data were then input into the predicted runs model mentioned previously. The predicted run amounts were then used in the updated Pythagorean runs formula, holding the runs allowed the same, to calculate predicted winning percentages and games won given the new roster changes. These results can be found in Table 5. Finally, I compared the four metrics outlined in Part Two, as well as the predicted runs created above average (Pred. RCAA), for the new players with Estrada's metrics (Table 6).

Team	Pred. Runs	Pred. Win %	Pred. Games Won
Current	712.38	50.97	82.58
+Kolten Wong, - Thairo Estrada	714.05	51.08	82.74
+Gleyber Torres, - Thairo Estrada	725.08	51.76	83.85
+Jake Cronenworth, - Thairo Estrada	734.2	52.32	84.75

Table 5. Results from imputing current team hitting data, team hitting data with Wong and without Estrada, team hitting data with Torres and without Estrada, and team hitting data with Cronenworth and without Estrada into predicted runs model and updated Pythagorean Wins formula from Part One.

Name	AB	Pred. RCAA	wOBA	RE24	O_Swing%	Z_Swing%
Kolten Wong	430	6.26	0.336	7.02	24.9	65.8
Gleyber Torres	526	5.58	0.328	2.24	31.5	74.9
Jake Cronenworth	587	7.38	0.318	25.21	25.4	61.5

Table 6. Statistics from Part Two and predicted runs created above average for potential new players. Red cells indicate the player underperformed in a specific statistic compared to Thairo Estrada while green cells indicate the player overperformed.

The first potential replacement player evaluated was Kolten Wong from the Seattle Mariners. Wong is a second basemen with a current AAV of $10.0M. Wong would be predicted to improve the team's winning percentage by 0.11% and increase wins by approximately 0.16 (Table 5). As can be seen in Table 6, Wong had 58 less at bats than Estrada last season. Wong outperformed Estrada in wOBA and O-Swing%, but slightly underperformed in RE24 and Z-Swing %. With less batting experience, it may take Wong some time to adjust to batting as often as Estrada did. From Wong's wOBA it's clear he'd be of offensive value to the team. However, his RE24 and Z-Swing% suggest he may not capitalize on crucial opportunities as often as we'd hope. Wong also comes with a hefty price tag, costing nearly five times as much as Estrada annually.

The next potential player evaluated was Gleyber Torres from the New York Yankees. Torres is a second basemen and shortstop with a current AAV of $10.0M. Torres is predicted to increase the team winning percentage by 0.79% and increase wins by approximately 1.27 (Table 5). In Table 6 it can be observed that Torres had 38 more at bats than Estrada last season. Torres outperformed Estrada in wOBA, O-Swing% and Z-Swing%, only underperforming in RE24. Torres' O-Swing% and Z-Swing% suggest he's the most disciplined option of the three, yet his

RE24 is significantly lower than the others. His number of at bats is encouraging as he's acclimated to batting more often. Again, a major limiting factor is Torres' AAV, especially considering him and Wong are the same price, yet Wong is predicted to contribute more to winning.

The final potential replacement player considered was Jake Cronenworth of the San Diego Padres. Cronenworth is a second basemen with a current AAV of $4.2M. He is predicted to increase the team winning percentage by 1.35% and increase the number of wins by approximately 2.17 (Table 5). Table 6 demonstrates that Cronenworth had 99 more at bats than Estrada last season. Cronenworth outperformed Estrada in wOBA, RE24, and O-Swing%, but underperformed in Z-Swing%. Of the three candidates, Cronenworth outperforms Estrada in the most categories. His astonishingly high RE24 presents him as a key player in game-deciding moments. While it's beneficial he is at bat so often, it must also be kept in mind that his stats could be inflated because he has more opportunities to hit. What's most appealing about Cronenworth is his significantly lower AAV. While it's still a jump compared to Estrada's $2.25M AAV, it's much more attainable for the organization. His affordability in combination with him being predicted to increase winning percentage the most make him the ideal candidate to add to the roster. Not to mention he's already located in California and will require less time to adjust to the area and climate than Torres or Wong.

Part Four: Recommendations, Limitations and Future Directions

As mentioned in Part One, we underperformed in comparison to what the updated Pythagorean wins formula predicted last season. As this prediction is based solely on runs allowed and runs scored, we need to shift our focus during spring training to address these issues. When breaking down the hitting metrics in Part Two, I'd argue that O-Swing% and Z-Swing%

are foundational to hitting. While power is important after contact is made with the ball, a hitter needs to be able to make contact first. We've seen that our top hitters struggle the most with Z-Swing%, so efforts should be focused on improving this metric for next season. Performance analytics can be utilized to do just that. I recommend that coaching staff and players consult with performance analysts on staff to use pitch tracking technologies to evaluate a player's swinging profile. With the 2020 addition of Hawk-Eye to the Statcast system, we have access to the most reliable pitching data to date (Jedlovec 2020). As such, this will allow performance analytics staff to answer questions for individual players such as: what types of pitches do you swing at most frequently? Are you better against right- or lefthanded pitchers? At what velocity are the pitches that you make the most contact with? Insights like these can help players understand their strengths and weaknesses as well as aid coaching staff in designing individual training exercises.

The other major recommendation stemming from this analysis is for management consider releasing Estrada and acquiring Cronenworth. I recognize that there are many factors to consider when making roster changes, and I suggest that management take the time evaluate if they agree with this recommendation as soon as possible. Myself and other members of the analytics team can compile additional information on both players should management require them. If they are in agreeance with this recommendation, the process should be initiated as soon as possible so both players have as much time as possible in the off season to get acclimated to their new homes and teammates.

Equally as important as presenting the recommendations from this analysis is presenting the limitations of this analysis. The linear regression run model that predicts run based on seasonal data referenced in Part Three was calculated to have an R-squared value of 0.966, suggesting the predictor variables used explain 96.6% of variation in the resulting runs. Overall, this is a well-performing model, but it's based on many assumptions. Linear regression models assume that the predictors are independent of one another, yet some of our predictors like doubles and triples are indeed correlated. Linear regression models also assume that residuals, or the difference in predicted runs and actual runs in this case, are normally distributed. This has not been verified for the run model. To address the multicollinearity amongst the variables in future work, additional regression model types such a ridge regression will be utilized as they help adjust for the fact that predictors may not be entirely independent (Severini 2020). Residuals will be calculated, plotted, and verified for normalcy as well.

While the linear regression run model performs well, it is by no means perfect. For example, using the 2022 Giants' seasonal data, the model predicted they would score 712.38 runs. In actuality, the Giants scored 716 runs last season. I point out these discrepancies to remind those receiving and interpreting the results of this analysis that these are predictions. They are well-informed with plentiful past data, but still are predictions, and should not be taken as the end all be all. Human consideration should always be used in combination with analyses.

In this analysis, we accounted for the hitting aspect of the Pythagorean wins formula by predicting runs scored. However, another limitation is the fact that runs allowed was assumed to remain the same when altering the roster and calculating estimated win percentages. Logically, adding a new player and removing a current player will result in differences in all statistics, not just hitting. As such, to develop this analysis further in the future, I'd like to consider how pitching and fielding data impact runs allowed. My goal would be to create another linear regression model that predicts a team's runs allowed for a season. That way both aspects of the Pythagorean wins formula can be predicted and more accurately reflect how a team would be impacted in changing a roster.

The next step required to improve the analysis would be data wrangling, which should not require additional resources as publicly available data can be easily accessed and utilized. Following the acquisition of data, a linear model would need to be constructed that predicts runs allowed. Such a model would need to find an ideal combination of predictor variables from team pitching and fielding data. This could be approached many different ways using feature selection methods such as SelectKBest or human trial and error using the variables that would make logical sense. Variable multicollinearity would have to be examined as well as residual distribution as mentioned earlier. All in all, this is an attainable improvement that will simply require additional time and will provide even more insight into both team performance and player acquisition analyses.

References

Jedlovec, Ben. 2020. "Introducing Statcast 2020: Hawk-Eye and Google Cloud." Medium, July 20, 2020. https://technology.mlblogs.com/introducing-statcast-2020-hawk-eye-and-google-cloud-a5f5c20321b8.

Rymer, Zachary, D. 2013. "What Is Baseball's More Powerful League, the AL or NL?". Bleacher Report, May 16, 2013. https://bleacherreport.com/articles/1641924-what-is-baseballs-more-powerful-league-the-al-or-nl.

Severini, Thomas A. 2020. *Analytic Methods in Sports: Using Mathematics and Statistics to Understand Data from Baseball, Football, Basketball, and Other Sports*. Boca Raton, FL: CRC Press.

Slowinski, Piper. 2010a. "wOBA". Fan Graphs: Sabermetrics Library, February 15, 2010. https://library.fangraphs.com/offense/woba/.

Slowinski, Piper. 2010b. "Plate Discipline (O-Swing%, Z-Swing%, etc.)". Fan Graphs: Sabermetrics Library, February 18, 2010. https://library.fangraphs.com/offense/plate-discipline/.

Slowinski, Piper. 2010c. "Sample Size". Fan Graphs: Sabermetrics Library, February 18, 2010. https://library.fangraphs.com/principles/sample-size/.

Weinberg, Neil. 2014. "RE24". Fan Graphs: Sabermetrics Library, July 30, 2014. https://library.fangraphs.com/misc/re24/.

Winston, Wayne L., Scott Nestler, and Konstantinos Pelechrinis. 2022. *Mathletics: How Gamblers, Managers, and Sports Enthusiasts Use Mathematics in Baseball, Basketball, and Football.* Princeton, NJ: Princeton University Press.