**MLB Front Office Consultants**

Riley Sansone, Eli Young, Isaac Glasser, Nick Adgar

DSO 578: Fundamentals of Sports Analytics

Instructor: Lorena Martin

October 15, 2024

# 

# 

# 

# 

# Abstract

This paper investigates the roster construction and performance metrics of the past five World Series winners (not including 2020 due to COVID-19 shortened season) to identify key trends and factors contributing to their success. By examining team strategies regarding homegrown talent, draft approaches, international signings, trades, and free agency, we aim to establish performance indicators (KPIs) that correlate with World Series victories. Using statistical methods, including ANOVA, we analyze player performance metrics across various categories. Our findings will inform a new initiative aimed at improving league-wide quality of play.

# Introduction

The landscape of Major League Baseball (MLB) is constantly evolving, influenced by the strategies employed by its most successful teams. This report focuses on the past five World Series champions to uncover the elements contributing most to their success. By analyzing various aspects of roster construction and performance metrics, we seek to provide actionable insights for all MLB teams. Our primary hypothesis is that a roster that is made up of more homegrown players and that accumulates more wins-above-replacement (WAR) is more likely to win the World Series.

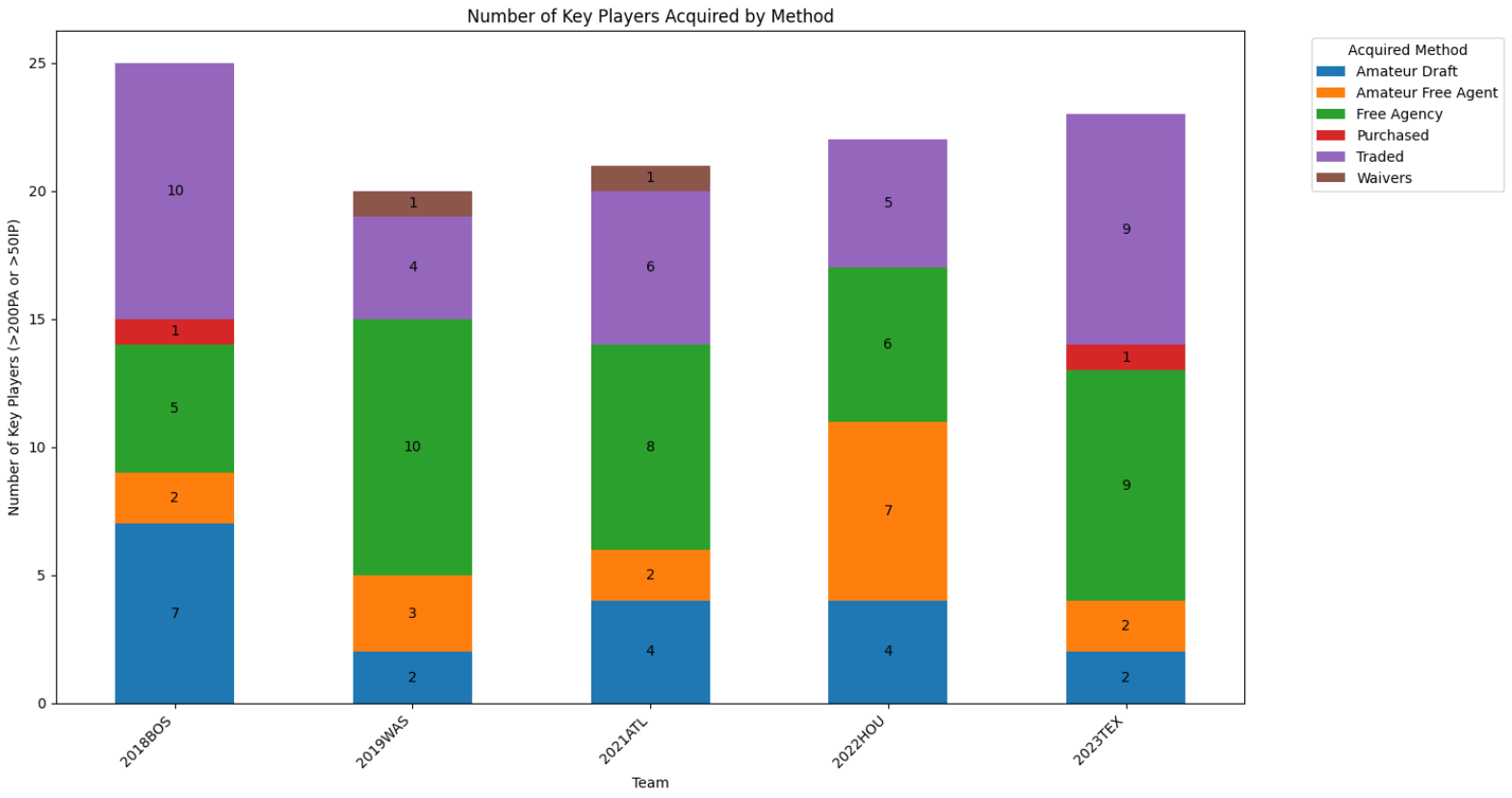
While each of us draw from our own personal interests and experiences as baseball fans as a catalyst for this research, we drew inspiration from existing literature to guide us. Ahead of the 2023 World Series, MLB.com writer Jim Callis published an article titled “How the World Series teams were built” wherein he examines the different methods by which the eventual champion Texas Rangers and runner-up Arizona Diamondbacks had assembled their rosters. Callis concludes that “recent World Series championships [sic] have been built in a variety of ways” but makes note of the stark contrast between the free-spending Rangers and the stingy Diamondbacks4. Further back, a 2013 article on popular sports news site BleacherReport.com titled “Blueprint for Building a World Series Winner” claims to offer just that by pointing out eight key statistical similarities between the previous ten champions5. While author Zachary D. Rymer writes tongue-in-cheek and acknowledges the absurdity of trying to distill a simple formula for winning a title, he does draw some interesting conclusions. For example, Rymer recommends that teams “Load Up Corners with 100 Home Runs and .470 Slugging Percentage”, and that the “First Two Hitters Should Combine for a .340 OBP and 200 Runs”5. Trying to tease out predictive insights from small-sample outcome data is flawed for a number of reasons that won't be discussed here, but some of the underlying information in Rymer’s analysis could prove useful in predicting future champions. For example, knowing that every World Series champion has two hitters at the top of their order who average – at minimum – a .340 OBP, Managers might prioritize this metric when selecting batting lineups, and General Managers might consider a high OBP top-of-the-order bat a priority when seeking team upgrades ahead of the trade deadline.

A 2005 article written by Kent von Scheliha for the Society for American Baseball Research (SABR) looks specifically at the differences between World Series winners and losers across a century's worth of data. von Scheliha concludes that “the World Series is simply too short for the stronger team to win consistently”, an important point to consider when analyzing World Series winners in hopes of informing future operational decisions6. The article does, however, suggest that World Series winners tend to be slightly stronger in pitching and defense than in hitting, arguing that “when the game is on the line, success is more likely to come from a timely strikeout or double play than it is from a clutch hit.”6 While it’s worth noting that this article is now nearly two decades old, and professional baseball has evolved greatly in the years since, it may be worth further examining whether this particular trend has continued. In our analysis, we break down an array of individual hitting and pitching KPIs to see which have been most predictive of championship success in the past five years.

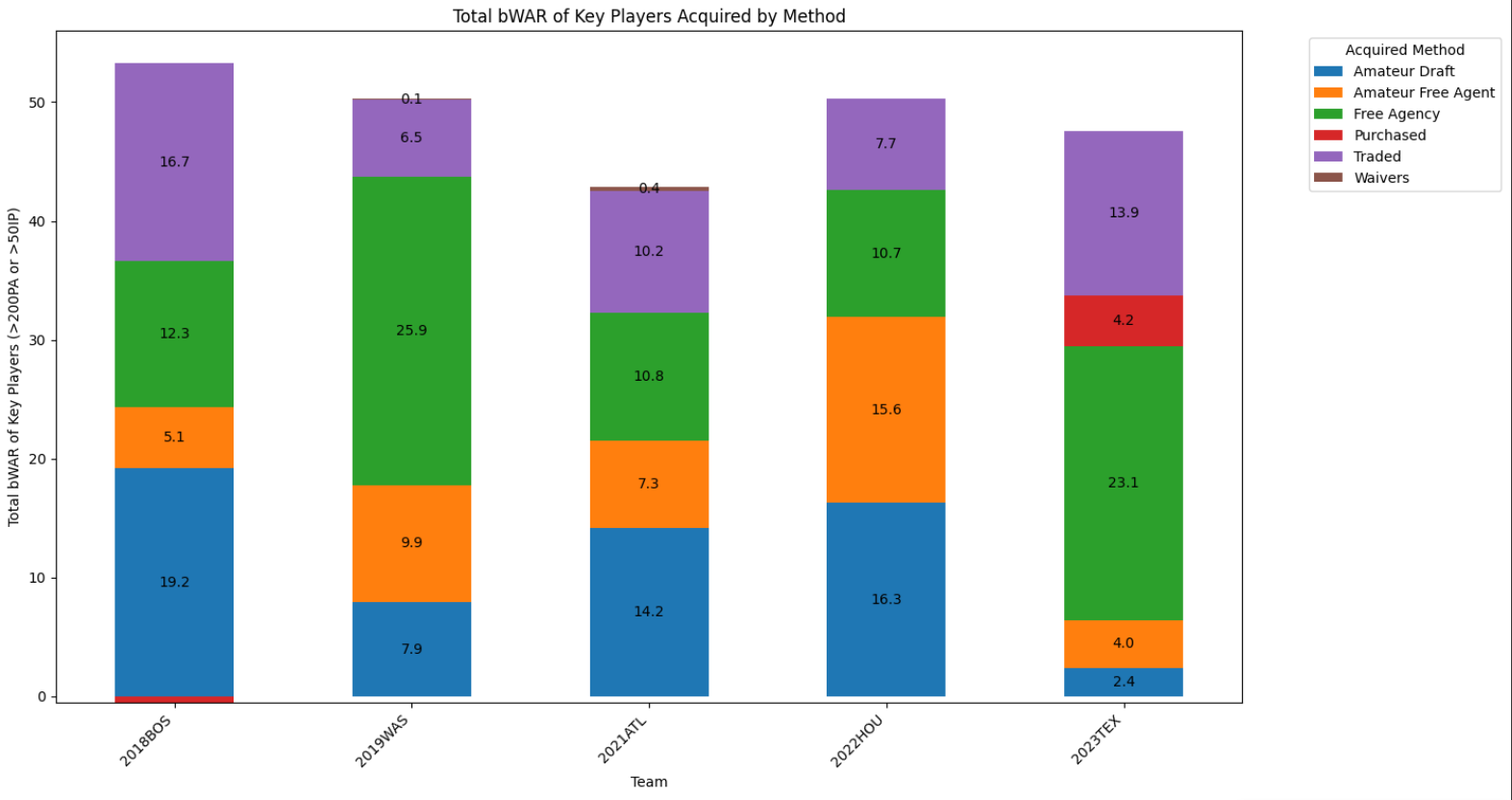
# Methodology

One of the primary sources we used for data collection was Baseball Reference1. From this source, we collected data on the five World Series teams we studied, and filtered to only include hitters from these teams that had at least 200 plate appearances and pitchers from these teams that had at least 50 innings pitched during the season they won a championship. In our analysis, we will refer to this subset as “key players”, as they were the primary contributors to their teams’ success. In our dataset, we included the KPIs of age, games, plate appearances, innings pitched, bWAR, oWAR, dWAR, AVG, OBP, SLG, Stolen Bases, ERA, ERA+, K/9, FIP, WHIP, Salary, and how they were acquired. We acquired data for wRC+ and OPS+ through FanGraphs, and data on contract information through Spotrac (see Appendix for more information).

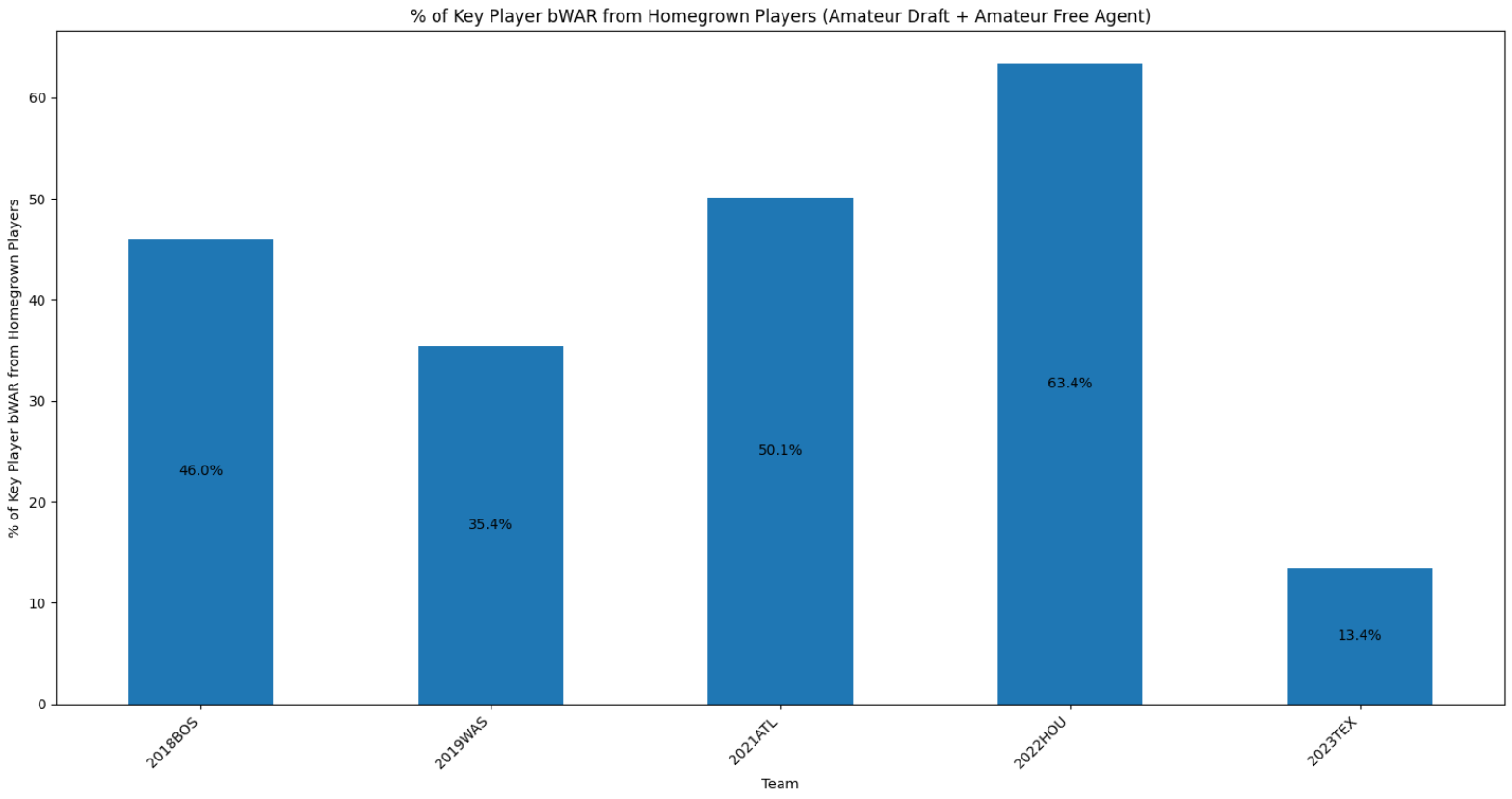
Before doing any high-level analysis, we opted to do some simple data visualization using Matplotlib and Pandas packages in Python in order to tease out any obvious trends. We looked at roster construction, i.e. how each of the past five World Series winners had acquired the key players on their rosters. We first looked purely at roster makeup, and noticed a) that every team used some mix of all available roster construction methods to assemble a team and b) that Boston and Houston leaned far more heavily on homegrown players than Washington, Atlanta or Texas.



We looked deeper at how much value each team was extracting during their World Series winning season from each player acquisition method. This analysis revealed a deeper understanding of how each winner was constructed. For this, we used Baseball Reference Wins-above-replacement (abbreviated as bWAR or simply WAR), a volume-based metric that tries to measure the sum total of a players value contribution (offense, defense and baserunning) over some period of time, in this case the entire relevant season7. In the figure below, we see that the proportion of WAR the 2022 Houston Astros gained from the amateur draft and amateur free agency far exceeds any other, especially the 2023 Texas Rangers and 2019 Washington Nationals. Both of these teams attained around half of their total team WAR from players they signed during a prior offseason free agency period.



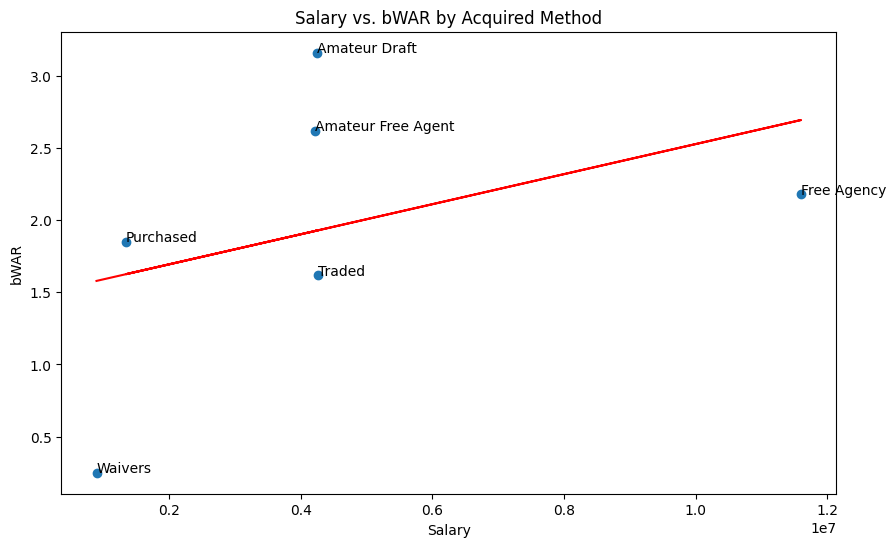
We designate “homegrown” players as players acquired via either the Amateur Draft (as high school or US college graduates) or Amateur Free Agency (as international free agents, typically from Latin America and the Caribbean. Drafted players typically take anywhere from 2-4 years to reach the major league level, while international free agents typically take even longer8. For the sake of analysis, we want to distinguish between these homegrown players and players acquired by teams after their minor league development has occurred. The caveat here is that the “Traded” method includes players who were traded as minor league players, meaning that the trade recipient may still have had a major role to play in that player’s development.We see in the following chart that there is a wide discrepancy between how much value World Series winners gained from these homegrown players, particularly between the 2022 Astros and the 2023 Rangers.



During the offseason before the 2022 season, the Texas Rangers – having spent half a decade as one of the worst teams in the league – opened up their coffers to the tune of $500 million, split between two of the most coveted free agent hitters on the market: Corey Seager and Marcus Semien9. That same winter, 3-time Cy Young Award winner Max Scherzer signed a contract worth $130 million with the New York Mets10. Scherzer would be traded to Texas during the 2023 season for a prospect and cash. During the offseason prior to the 2023 season, 2-time Cy Young Award winner Jacob DeGrom inked a huge contract of his own with Texas to the tune of $185 million over 5 years11. While it’s rarely as simple as spend money; win games, the Rangers’ spending spree, combined with a timely wave of young contributors, provided just enough juice for Texas to sneak into a wildcard playoff spot before catching fire en route to a World Series title in 2023.

During that 2023 regular season, Seager and Semien contributed a combined 14.3 WAR, accounting for more than 40% of the total WAR contributed by all 26 hitters that appeared in at least one game for that Rangers team12. Had those Rangers lost just two fewer games on their 162-game slate, they would have missed the playoffs entirely. The approximate 14 wins above the replacement provided by Semien and Seager were – by far – the biggest reason that Rangers team had the opportunity to win the World Series at all13.

By contrast, the 2022 Houston Astros won a title many years in the making, with a formidable core of homegrown talent drafted, signed and developed by the Astros elite organizational apparatus. While those Astros still ran the league's 8th highest payroll (compared to Texas’ 3rd), including large allocations to homegrown stars like Jose Altuve and Alex Bregman, they got massive value from cheaply-signed homegrown superstars Yordan Alvarez, Kyle Tucker, and rookie phenom Jeremy Pena14.



In consideration of the data on how each World Series winner acquired their roster, we quickly compared the different acquisition methods on a cost-to-value ratio. In other words, the above figure evaluates how many win shares each team gained per dollar spent from each acquisition method. In the aggregate, the methods above the trendline (homegrown players) are unsurprisingly more cost-efficient. This aligns with our findings when comparing the 2022 Houston Astros and the 2023 Texas Rangers.

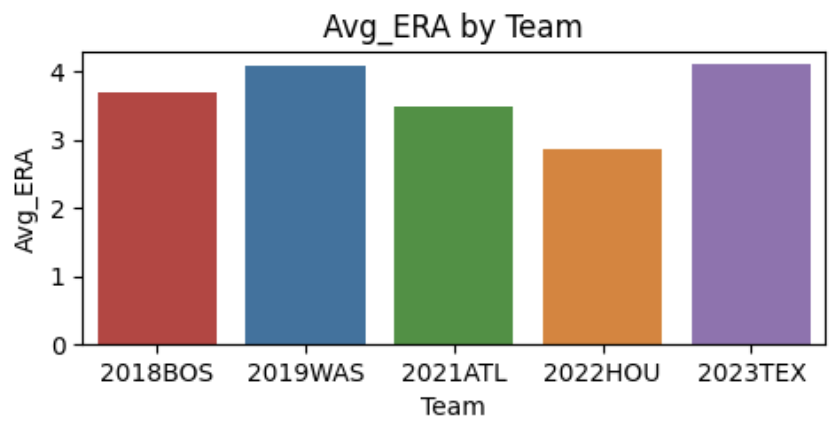
## KPI Analysis Overview

When examining the traditional baseball KPIs, our team hypothesized that WAR would be the leading predictor of team success and World Series victories. In order to test this hypothesis we conducted an ANOVA test among the World Series winners that would alert us of any statistical differences in KPIs, calculated z-scores for each World Series winner to compare their KPIs to the league in that year, and ran a logistic regression model to identify the KPIs that had the most significance in predicting a World Series win.

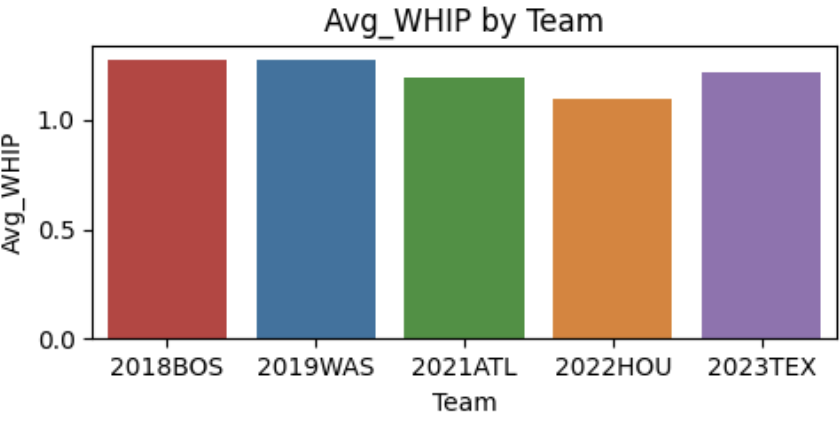
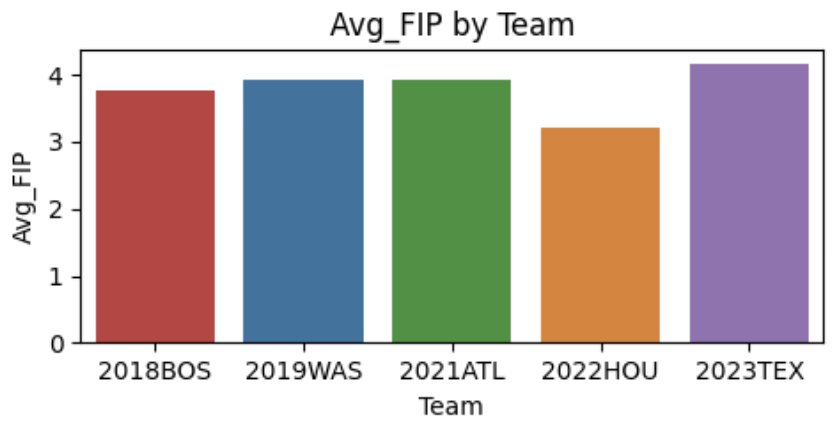
## ANOVA Testing of Measured KPIs

What we found from the ANOVA test of World Series winners was that there was no statistically significant difference among any of the batting KPIs and only one difference in the pitching KPIs at the 95 percent confidence level: team ERA.

The ERA difference was found between the 2019 Washington Nationals, 2022 Houston Astros, and 2023 Texas Rangers where the Astros were significantly lower than both the Nationals and Rangers. This difference can be attributed to the highly regarded performance of the Astros’ pitching staff that year which helped attain the fourth-lowest BAA (Batting Average Against) of any team since integration in 194715.



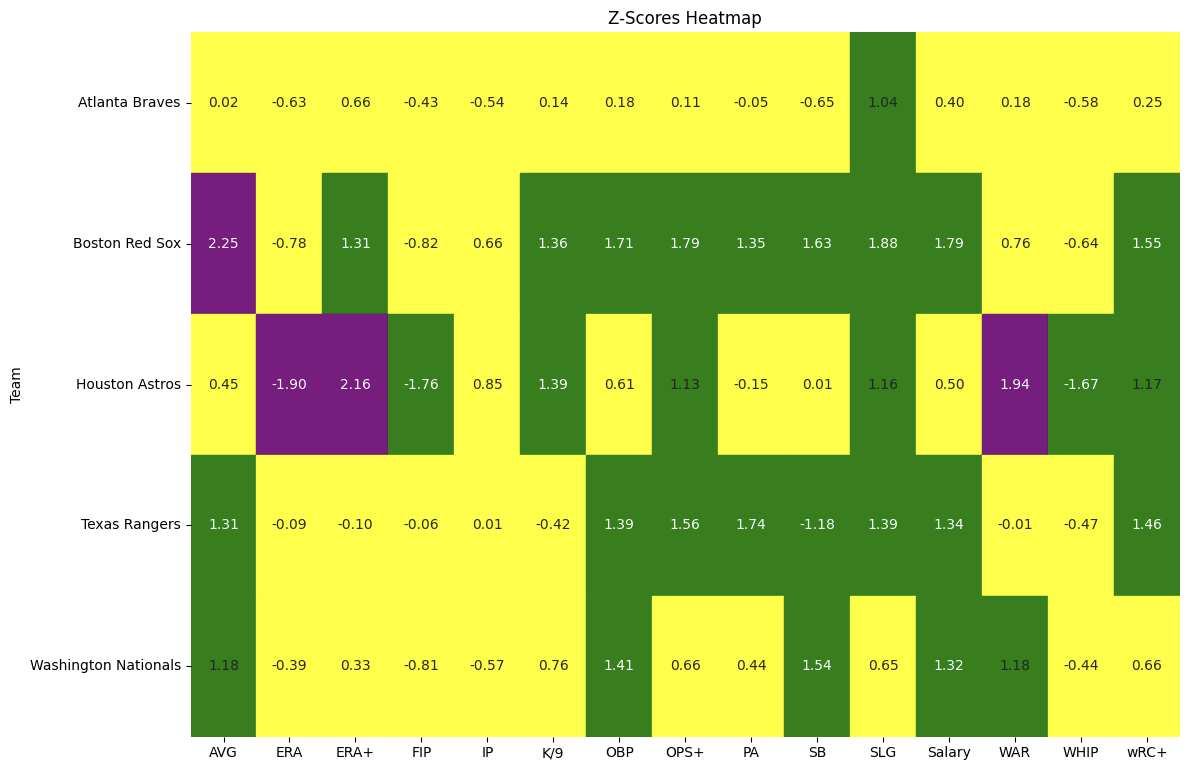
Although ERA was the only KPI that achieved significance, we believed that it was worth mentioning FIP and WHIP which came close to significance with p-values of 0.116 and 0.079 respectively. The teams that were involved in the significance tests were the FIP of the 2022 Houston Astros and 2023 Texas Rangers and the WHIP of the 2018 Boston Red Sox and 2022 Houston Astros. This once again highlights the dominance of the Astros pitching that year.



Since batting KPIs were found to not be significantly different among the World Series winners and the only significant difference in pitching KPIs was due to the dominance of the 2022 Houston Astros, we concluded that World Series winners likely have KPIs in common with one another. We attempted to find these similar KPIs by comparing each winning team to the league for their respective year through a z-score analysis.

## Z-Score Comparison of World Series Winners and League Average

From the z-score analysis and visualization we found that World Series winners commonly stood out from the entire league in each year through AVG (batting average), OBP (on base percentage), OPS+ (on base percentage plus slugging), SLG (slugging), SB (stolen bases), wRC+ (weighted runs created plus), and salary.



In this mapping a yellow color corresponds to a z-score which is close to league average and can be considered common or expected, taking values from -1 to 1.

The green color corresponds to values that are slightly above league average and take values from -1 to -1.9 and 1 to 1.9.

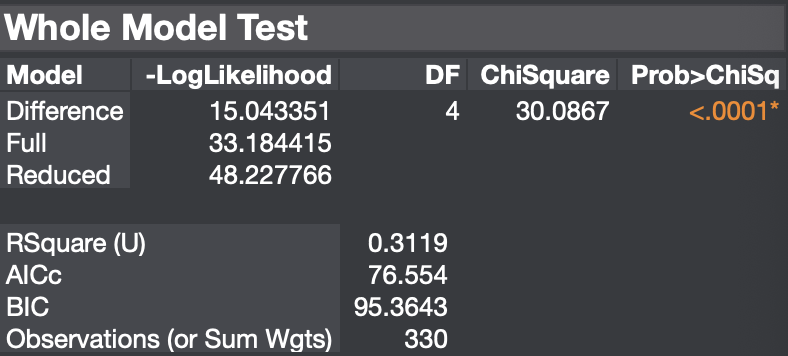
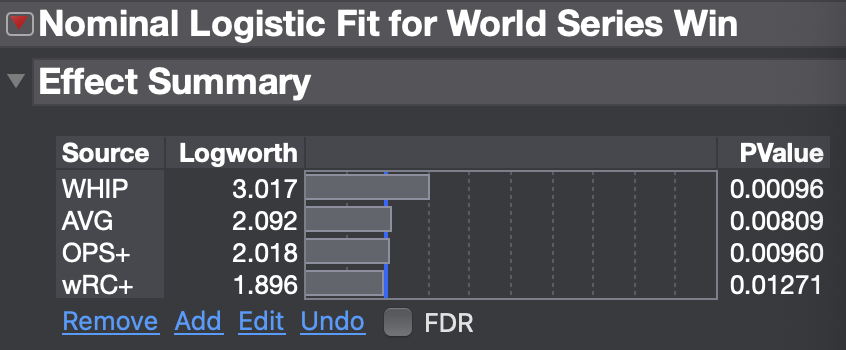
The purple color represents values that are uncommon and can be considered significantly different from league average, suggesting that a team was performing exceptionally well for that KPI, and are considered when a team has values over 1.9 or under -1.9. Traditionally, the threshold for significant difference is around 1.95 but we decided to lower it as the performance of the 2022 Houston Astros was not over this threshold even though they had what was considered an exceptional season.

Interestingly, all of the KPIs are batting related with the exception of salary which leads us to believe that World Series winning teams are more focused on a dominant offense rather than an overwhelming pitching game and defense. The 2022 Houston Astros stick out as the exception to this belief, but as noted earlier, they had a stellar pitching season that was highly unusual.

To confirm our findings and validate that these KPIs could be translated into predicting World Series victories, we created a logistic regression model using all of the KPIs measured and expanded our data to include information from 2012 until 2023.

## Nominal Logistic Regression Prediction Model

Using JMP to run a logistic regression model we found that WHIP, AVG , OPS+, and wRC+ were the only statistically significant predictors of World Series victories.



This refutes our initial hypothesis that WAR would be the most useful metric for predicting World Series victories and confirms some of our findings from the z-score tests. AVG, OPS+, and wRC+ were all metrics that we found were similar among World Series winners in the z-score test and they made it into the model, but WHIP was not one that we expected to make it into the model given the z-score results. This model had an explanatory power of around 32 percent which is an adequate result given the amount of explanatory variables present in the sport.

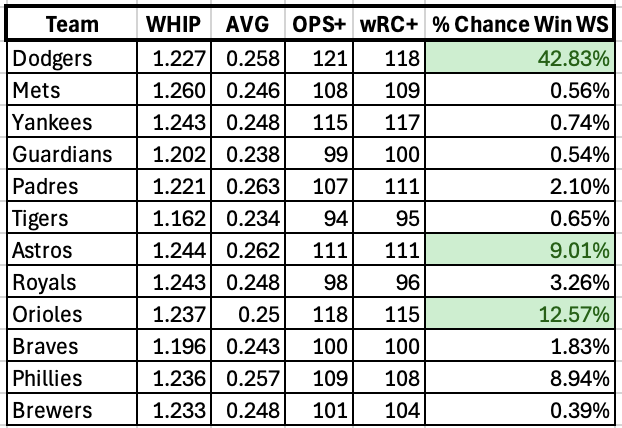
However, thinking about these KPIs from what they measure and not just from their statistical significance in World Series winners, they are some of the most useful statistics that can be picked for prediction. OPS+ and wRC+ are both designed to try and standardize the offensive capabilities of a team, AVG directly measures a team’s consistency in offensive production, and WHIP is an all around metric for a pitcher’s defensive ability.

An interesting point to notice is that the coefficients for the statistics seem to follow the normal assumption of the statistic except for wRC+. A higher AVG and OPS+ is rewarded in the model and so is a lower the WHIP, which is how the statistics work, but wRC+ is penalized for being higher even though a larger wRC+ value is better. This could be a limitation of the model as it is incorrectly penalizing a statistic that should be rewarded for a higher value. A possible explanation for this occurrence could be the limitation of our data as we are not including playoff statistics for each team and team performance during the playoff season can vary dramatically from their regular season. This opens up a possibility for future inclusion of playoff data to try and identify what causes this interaction.

# 

# Prediction Results

Using the nominal logistic regression model that we created, we were able to make predictions on how likely each team in the playoffs is to win the World Series this year. We found that the Dodgers have the highest chance at around 43 percent with the Orioles and Astros, who have already been knocked out at the time of writing, come in second and third with a 13 and 9 percent chance respectively.



# Discussion

While we are proud of our study, there are some limitations to our findings. For example, some of the key players that helped these teams win the World Series (Joc Pederson in 2021, Trey Mancini in 2022, and Evan Carter in 2023) were either traded to the winning team in the middle of the year, or made their MLB debut in the middle of the year, and thus didn’t collect the minimum 200 plate appearances required to be a part of our study. In addition, it would have been beneficial to have broken down spending attribution to different position groups, to see how these teams allocate money based on position. In addition, we use the players’ regular season statistics, since the playoffs are a relatively small sample size, and are not representative of the key players' traits. This is a problem because winning a World Series requires the clutch factor that is inherent in postseason performance. Some players who are great in the regular season struggle in the postseason, or vice versa, and this is why some regression predictions may be incorrect. It also would have been interesting to include two-way interaction in our regression with PA/IP qualifiers and performance.16

# Conclusion

MLB GM’s should build rosters with a higher balance of homegrown talent. This should mean that General Managers should try and convince the team owner to spend more money on minor league development. Acquiring proven stars is great for a team’s future success and financial valuation, and could lead a team to the mountaintop, as we saw with the 2023 Rangers, but spending money on minor league development allows the team to get the best out of their young prospects, and increases the chance for sustained success year after year, as we saw with the 2022 Astros.

Our initial hypothesis was proven incorrect, as WHIP was the most useful metric for predicting a World Series win according to our analysis. This means that GM’s should look for pitchers who limit baserunners to play for their team. Based on some of the concepts we learned in class, this makes sense. When a pitcher pitches with no baserunners, there is reduced stress and a reduced psychological load on the pitcher’s mind, which allows him to work more freely. This in turn leads to less runs per game, which more likely means more wins for the team. Additionally, teams should make sure to track and actively improve AVG, OPS+, and wRC+ as much as possible if they are trying to maximize their chances of winning the World Series. This will require extensive training and emphasis on batter efficiency and can comprise the bulk of talent recognition metrics for potential homegrown players.

# References

**Statistics**

1. "Baseball Reference." *Baseball-Reference.com*, <https://www.baseball-reference.com/>.
2. "FanGraphs." *FanGraphs.com*, <https://www.fangraphs.com/>.

**Contract/Financial Data**

1. "Spotrac." *Spotrac.com*, <https://www.spotrac.com/>.

**Articles**

1. "How the World Series Teams Were Built: 2023." *MLB.com*, October 27, 2023 <https://www.mlb.com/news/how-the-world-series-teams-were-built-2023>.
2. O'Connor, John. "Blueprint for Building a World Series Winner." *Bleacher Report*, April 1, 2013 <https://bleacherreport.com/articles/1492758-blueprint-for-building-a-world-series-winner>.
3. "World Series Winners and Losers: What's the Difference?" *SABR Journal*, <https://sabr.org/journal/article/world-series-winners-and-losers-whats-the-difference/>.
4. "WAR Explained." *Baseball-Reference.com*, [https://www.baseball-reference.com](https://www.baseball-reference.com/).
5. "Time from Draft Until Majors." *FanGraphs Community*, February 21, 2015 <https://community.fangraphs.com/time-from-draft-until-majors/>.
6. "Texas Rangers Finalize Mega Deals for SS Seager, 2B Semien." *ESPN*, December 1, 2021<https://www.espn.com/mlb/story/_/id/32766464/texas-rangers-finalize-mega-deals-ss-seager-2b-semien>.
7. "Max Scherzer's Deal with the Mets." *MLB.com*, December 1, 2021 <https://www.mlb.com/mets/news/max-scherzer-mets-deal>.
8. "Jacob deGrom Was a Mistake, but Rangers Can Still Win." *Texas Monthly*, June 7, 2023 <https://www.texasmonthly.com/arts-entertainment/jacob-degrom-was-mistake-but-rangers-can-still-win/>.
9. "Texas Rangers 2023 Team Page." *Baseball-Reference.com*, <https://www.baseball-reference.com/teams/TEX/2023.shtml>.
10. "How the Texas Rangers Built Their 2023 World Series Roster." *Sports Illustrated*, October 25, 2023 <https://www.si.com/mlb/rangers/news/how-texas-rangers-built-2023-world-series-roster-arizona-diamondbacks>.
11. "Houston Astros 2022 Team Page." *Baseball-Reference.com*, <https://www.baseball-reference.com/teams/HOU/2022.shtml>.
12. "Is This the Best Astros Pitching Staff We've Ever Seen?" *Fox Sports*, October 19. 2022 <https://www.foxsports.com/stories/mlb/2022-mlb-playoffs-is-this-the-best-astros-pitching-staff-weve-ever-seen>.
13. "Evan Carter Won't Be Overlooked Anymore." *Sports Illustrated*, October 18, 2023 <https://www.si.com/mlb/2023/10/18/texas-rangers-evan-carter-wont-be-overlooked-anymore>.

# Appendix

1. **G:** Games Played
2. **PA**: Plate Appearances
3. **bWAR**: Wins Above Replacement (Total WAR from Baseball-Reference)
4. **oWAR**: Offensive Wins Above Replacement
5. **dWAR**: Defensive Wins Above Replacement
6. **AVG**: Batting Average
7. **OBP**: On-Base Percentage
8. **SLG**: Slugging Percentage
9. **OPS+**: On-Base Plus Slugging Plus (Adjusted for park factors and league averages)
10. **wRC+**: Weighted Runs Created Plus (Adjusted for park factors and league averages)
11. **SB**: Stolen Bases
12. **Salary**: Player Annual Salary
13. **GS**: Games Started
14. **IP**: Innings Pitched
15. **WAR**: Wins Above Replacement
16. **ERA**: Earned Run Average
17. **ERA+**: Adjusted ERA (Adjusted for park factors and league averages)
18. **K/9**: Strikeouts per Nine Innings
19. **FIP**: Fielding Independent Pitching
20. **WHIP**: Walks Plus Hits per Inning Pitched