Title

The 2021 San Francisco Giants: A Deep Dive Into One of Major League Baseball's Most Overperforming Teams

Author

Carey, Philip Robert

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

The 2021 San Francisco Giants: A Deep Dive Into One of Major League Baseball's Most Overperforming Teams

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Applied Statistics and Data Science

by

Philip Carey

2024

ABSTRACT OF THE THESIS

The 2021 San Francisco Giants:
A Deep Dive Into One of Major League Baseball's
Most Overperforming Teams

by

Philip Carey
Master of Applied Statistics and Data Science
University of California, Los Angeles, 2024
Professor Frederic R. Paik Schoenberg, Chair

In the world of Major League Baseball, the 2021 San Francisco Giants season stands as a remarkable instance of overperformance, defying the expectations of analysts and fans alike. This study seeks to unravel the factors contributing to the Giants' unexpected success by addressing the research question: what made the 2021 San Francisco Giants so much better than expected?

The primary objective of this research is to analyze the discrepancies between forecasted and observed data to pinpoint areas of overperformance. To achieve this, I conducted a thorough examination of both team-level and individual-level statistics. The analysis extends beyond basic metrics to include an in-depth look at the resurgence of veteran players and the critical role of pitching.

Methodologically, this research involved gathering extensive data on player performances and team outcomes. By comparing preseason projections with end-of-season statistics, I identified key variances and their potential causes. The study encompassed various performance indicators, including win-loss percentages, ERA, FIP, and WAR, both at the team and individual levels.

The findings reveal that the Giants' overperformance in 2021 can be attributed to a combination of factors. At the team level, strategic decisions and effective management

played a significant role. On an individual level, several veteran players experienced career resurgences, significantly surpassing their projected statistics. Additionally, the importance of pitching cannot be understated, as key pitchers outperformed expectations, contributing to the team's overall success.

In conclusion, the 2021 season for the San Francisco Giants illustrates a complex interplay of factors leading to their overperformance. While it remains challenging to definitively classify their success as purely luck or a result of precise strategic execution, the contributions of standout players were undeniably crucial. This research underscores the multifaceted nature of baseball success and highlights the potential for unexpected outcomes even in a data-driven sport.

The thesis of Philip Carey is approved.

David Anthony Zes

Michael Tsiang

Yingnian Wu

Frederic R. Paik Schoenberg, Committee Chair

University of California, Los Angeles

2024

To my dad . . .

who recorded nearly every statistic of my childhood sports,

enjoyed with me many an afternoon at the ballpark,

and taught me to love the game of baseball.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

My primary advisor for this project was Dave Zes, who I consistently met with and talked over the potential avenues of researching this problem. His guidance greatly influenced my work and I have a lot to thank him for.

And I would like to acknowledge the rest of the statistics department at UCLA and to anyone here with whom I worked with, providing me with the foundational knowledge to complete a project such as this one, and always encouraging me to think outside the box.

VITA

2018-2022	B.S. (Statistics and Data Science), UCSB, Santa Barbara, California.
2022-2024	Master of Applied Statistics and Data Science, UCLA, Los Angeles, California. Enrolled at time of writing.

CHAPTER 1

Introduction

Every year, sports fans pour their heart and soul into rooting for their teams to win a game, a series, a division, a championship—whatever it may be. For most of these fans, it is simply about the love of the game. Like many others, I find myself in that same position. However, with a background in statistics, I appreciate viewing the game through a more analytical lens.

Sports inherently contain a degree of randomness. Players are human beings, and with that humanity comes unpredictability. A sudden illness or injury, or events outside of the game that affect physical well-being, can influence a player's performance and, consequently, the game itself. Considering that every player has a degree of randomness associated with them adds to the complexity. Even before addressing the dynamics of the game, we already face a quite complex problem.

Now, let us consider baseball, the focus of this paper. Baseball is a game built on pure, man-made, measurable statistics, but that does not mean the game is free from randomness. Take pitching, for example: a pitcher may intend to throw a fastball low, but miss up in the zone instead. Variations in spin rate, arm angle, and other metrics can differ slightly from pitch to pitch. The hitter, in turn, must react to this irregularity. Even with thorough preparation, they're still playing the odds, making decisions in real-time depending on how the ball is thrown.

In essence, sports are events filled with randomness, and baseball is no exception. However, human nature drives us to control, forecast, and problem-solve in any scenario, especially in high-stakes sports. Teams strive to gain even the slightest advantage to optimize player performance. This leads us to the core of this thesis aims to uncover: understanding

how to account for the efficacy of our forecasts.

Having grown up in the Bay Area as a San Francisco Giants fan, I have experienced this particular organization's highs and lows. Winning three World Series titles in five years between 2010 and 2014 was a remarkable achievement. However, the team has also endured multiple losing seasons over the past two decades, with 2016-2020 being particularly mediocre. Heading into 2021, the Dan Szymborski (ZiPS) projection system predicted another losing season for the Giants. Yet, they defied expectations, achieving a franchise-high 107 wins out of a 162 game season (meaning nearly 2/3 of their games were wins). This stunning performance surprised fans who expected dominance from other teams in the division, surely not the Giants.

Even with the impressive 107 win season under the Giants' belts, a fierce Dodger competition was right on their tail, winning 106 games and clinching the wild card spot to have a chance at knocking them out of the playoffs. It came down to the wire, with both both teams winning 2 out of the first 4 games and sending it to a game 5. That final game was decided by a singular run to the Dodgers, ending the Giants' upset season and their chance at proving to everyone that their success was no fluke. At the end of the day, when two teams are so neck and neck in terms of talent, it all comes down to those small random chance events that add up slightly in favor of one team or the other. But random chance or not, the Giants proved that while being undervalued at the beginning of the season, they could hang with the best of the best by the end.

This brings us to the central question of my thesis: what made the 2021 San Francisco Giants so much better than expected?

CHAPTER 2

Data Acquisition Process

Before diving into the analysis, I think it would be useful, for the reader's sake, to provide context on the data collection process and the challenges involved. The data collected is comprehensive and insightful, but I should mention I am not part of an MLB organization's R&D team. I am a student with a passion for baseball and a desire to use my web-scraping knowledge to create my own dataset.

The first challenge was determining which source to use as the ground truth. With numerous resources available, I initially considered several, including Baseball Reference [bbr] and Baseball Savant [sav]. These large data sources are rich in information and often overlap. However, for this project, I chose to use data from Fangraphs [fan]. This choice was driven by the focus on comparing forecasted and observed statistics, and Fangraphs' consistent publication of the ZiPS projection system by Dan Szymborski was particularly useful. Additionally, James LeDoux's Python module 'pybaseball' [LeD] was very helpful in identifying the best data sources for various statistics and simplifying the retrieval of basic statistics, such as team standings.

For observed statistics, I wrote straightforward functions to gather data from the web. However, collecting forecasted data was a more complex challenge. This paper analyzes both team-level and player-level statistics, neither of which are hosted on easily accessible APIs.

For team-level forecasts using the ZiPS system, I found that Szymborski publishes his predictions annually on Fangraphs, typically around the same time each year since 2019. I manually copied and pasted the relevant tables into a large Excel spreadsheet for the years 2019-2023. Additionally, I found three more years of data from Szymborski's ESPN blog posts, and added that to the spreadsheet. His blog posts on Fangraphs and ESPN follow a

consistent format, such as his 2021 blog post [Szy21b] and his 2017 ESPN post [Szy16a].

Collecting individual player statistics was somewhat more straightforward but still challenging. Szymborski includes these forecasts in his team-specific blog posts each season. From 2016 to 2023, with 30 teams per season, there were 240 blog posts to review for individual player statistics. To have a chance at performing any sort of meaningful analysis, I consolidated all this information into my own database. The blog posts followed a consistent linking style, as seen in the 2021 Giants post [Szy20a]. After addressing some edge cases (e.g., team name changes and alterations in the style of the link), I created a function to extract data from these tables. Some statistics were deprecated over time, and new ones emerged, so certain values were used sparingly due to inconsistency over the eight years of data collection.

After creating these scraper functions, I stored all observed and forecasted statistics in a database of my own, including tables for forecasted batter/pitcher statistics with corresponding standard and advanced stats, and tables with observed values from the Fangraphs API. This solution enabled the conclusions drawn in this paper, the creation of any tables and charts you may see, and also provided a framework for future analysis. As a result, you may not notice an abundance of citations throughout the paper, because the conclusions have been derived from the dataset I have created. However, it should be noted that the rich source of data provided by Dan Szymborski and Fangraphs is the only reason that a project like this, and the rest of my findings, could be made possible.

CHAPTER 3

Team-Level Overview

To assess the San Francisco Giants' performance relative to preseason projections, we turn to the preseason forecasts provided by Dan Szymborski of Fangraphs. Szymborski's projections, a mainstay in baseball analytics, have been available on Fangraphs for several years, offering a valuable dataset starting from 2016. While our analysis is limited to the past eight seasons due to data availability constraints, this timeframe encompasses a comprehensive sample of 30 teams per season. This dataset allows us to gauge the likelihood of a team achieving a performance akin to that of the 2021 Giants. Here, we take a look at the difference in Win/Loss percentage across teams actual vs. predicted rates.

Tm	Year	Win Actual%	Win Proj.%	WL Percent Diff
San Francisco Giants	2021	0.660	0.463	0.197
Boston Red Sox	2018	0.667	0.537	0.130
Baltimore Orioles	2023	0.623	0.494	0.129
Oakland Athletics	2018	0.599	0.481	0.118
Baltimore Orioles	2022	0.512	0.395	0.117
...	...	...	...	...
Detroit Tigers	2019	0.292	0.420	-0.128
Washington Nationals	2022	0.340	0.469	-0.129
Washington Nationals	2020	0.433	0.567	-0.134
San Francisco Giants	2017	0.395	0.531	-0.136
Baltimore Orioles	2018	0.290	0.512	-0.222

Table 3.1: Win-Loss Differentials for Teams between 2016-2023

Table 2.1 doesn't necessarily pinpoint the 'best' performing team each year, but it highlights the top overperformers, or rather, the teams that exceeded their preseason projections the most. In particular, the San Francisco Giants not only claim the top spot, but do so by a considerable margin. Their actual win rate surpasses the projected rate by nearly 20 points, with almost a seven-point difference between them and the next team in terms of differential.

To drive this point home on just how much of an anomaly season this was for the Giants, I'll use a simple betting analogy. Of course the world of sports betting is far more intricate that what I'm describing, but let's assume that the Giants' presumed win rate of 0.463 determined their odds for the rest of the season. Using the standard betting line formula, a 0.463 probability of winning will give a money line of +216, or in other words, placing a $100 bet will yield $216 in profit. If we were to assume this money line for the Giants every game, and we placed $100 for them to win with these odds every game, we would've made $17,612 over the course of one season. And again, no sports betting practice is foolish enough to keep these odds as they recognize the Giants' success, but in some world where we only go off of how the team was projected to do before seeing them in action, there's a lot of money to be made.

Following this line of curiosity, I decided to visualize what the rest of the Giants Win/Loss percentages looked like across the scope of our timeframe.

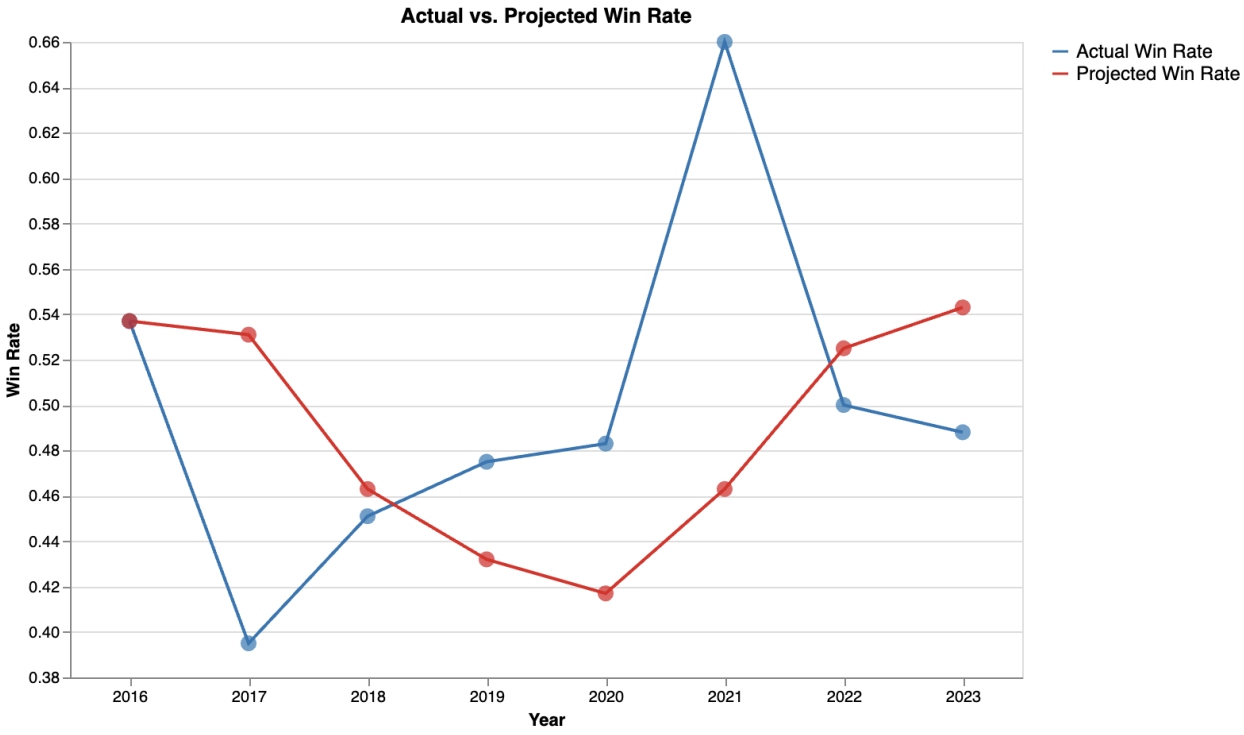

Figure 3.1: Win Rate of SF Giants From 2016-2023

It's clear the organization has been a middling team as of late, hovering around that .500 line and, in fact, dipping a ways below for many other years. We also notice that they do both over and underperform at different paths along their trajectory, with 3 overperformances, 4 under, and 1 that just about perfectly predicts it. There also appears to be a bit of a lag in correlation between the two values, where the actual trend clearly points in one direction until the predicted rate begins to catch up to it, which would seem reasonable in the difficult task of forecasting a team's success.

Going back to reviewing the rankings in Table 2.1, it too becomes evident that the 2018 Baltimore Orioles fall short, performing a staggering 22 points below their preseason projection. This observation prompts an intriguing question: is it more common for a team to surpass expectations, as exemplified by the Giants, or to fall short, akin to the Orioles' performance?

To look into this inquiry, I opted to visualize the frequencies of the Win/Loss percentage difference across every team from the seasons of 2016 to 2023.

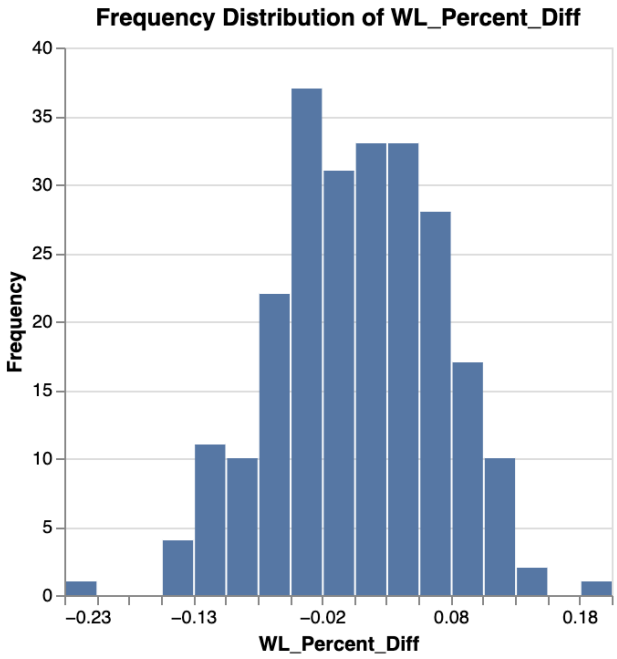

Figure 3.2: Empirical Distribution of Win-Loss Differential Across All Teams From 2016-2023

The visualization depicts a distribution that aligns somewhat predictably with a normal distribution—a reassuring outcome given the reliability of Dan Szymborski's projection system. The mean of the difference between wins and losses hovers remarkably close to 0 (approximately 0.0000416), exhibiting the characteristic bell curve shape indicative of a normal distribution. Assuming normality, we can compute a variance statistic, which yields a standard deviation of approximately **0.0636**, a seemingly reasonable outcome.

It's noteworthy that both the Giants and the Orioles deviate significantly from this norm, each exceeding 3 standard deviations from the mean. In a sample encompassing 240 teams (30 teams across 8 seasons), such extreme deviations are exceptionally rare, with only one team overperforming expectations to such a degree and another underperforming similarly.

Looking further into this distribution of the difference between win/loss percentages, I conducted a simulation based on the available 240 observations. The approach involved assuming a normal distribution for win/loss percentages, utilizing mean and standard deviation values derived from the dataset (mean = 0, standard deviation = 0.0636). Subsequently, 10,000 win/loss samples were generated for each team's preseason projection. This process

yielded 10,000 new 'actual' win rates per team/season pairing. The simulation was repeated using five different random seeds.

The objective was to assess how frequently the generated win/loss percentages equaled or exceeded the 2021 Giants' value of 0.197. The results indicated that, on average, **2,416.8** out of **300,000** samples per trial (30 teams * 10,000 samples each) surpassed the threshold of 0.197. Dividing these values, it can be expected that approximately **0.806%** of teams in future seasons will outperform at the same level as the Giants.

Contrasting this with the current reality of only 240 observations, it becomes evident that only one team (the 2021 Giants) has achieved this level of performance, representing approximately **0.416%** of teams in the dataset. This result underscores the discrepancy between simulated outcomes and actual observations, indicating that the performance of many teams falls short of expectations, and perhaps we may have even expected another team to outperform expectations at the level of the 2021 Giants.

Our analysis begins to reveal some insight into the nature of team performance relative to preseason projections. The rarity of extreme overperformance, as evidenced by the 2021 San Francisco Giants, underscores the exceptional nature of such outliers in a largely predictable distribution. The simulated data suggest that while outlier performances are statistically possible, they remain infrequent in actual observed seasons. This discrepancy between simulated probabilities and real-world outcomes emphasizes the influence of unquantifiable factors, such as team dynamics, management decisions, and individual player contributions, which may not be fully captured in preseason projections.

Transitioning from the team-level analysis, it is crucial to go further into the contributions at the individual player level. Understanding how key players' performances diverge from expectations can provide further clarity on why some teams manage to significantly outperform their projections. In the following section, we will explore the impact of individual player statistics, focusing on how standout performances or unexpected contributions can drive overall team success.

CHAPTER 4

The Veteran Offense

There are various avenues we can explore when it comes to analyzing the individual players' performances as a contribution to the team's ability to win. At the end of the day, there's so many variables that it can be difficult to discern one being more meaningful than others from year to year. However, one trend that caught my eye when digging deeper into the player stats was the rejuvenation of a few veterans of that 2021 Giants team.

Firstly, the Giants had the highest average player age in the MLB in 2021, with an average age of 30.32 years. Only one other team had an average age above 30, and two more teams had average ages above 29, while the majority of teams had average player ages between 27 and 29. Here is the age breakdown:

Season	TeamName	Mean Age
2021	SFG	30.32
2021	OAK	30.28
2021	CHC	29.79
2021	LAA	29.40
2021	ATL	28.90
...	...	...
2021	TEX	26.89
2021	CLE	26.81
2021	BAL	26.67
2021	SEA	26.21

Table 4.1: Mean Age of Teams in 2021

Typically, having some veterans on a team is beneficial for leadership and mentoring younger players. However, the 2021 Giants were predominantly composed of veterans, especially in the starting hitting lineup. While this could imply a wealth of experience and wisdom, it also means a higher risk of injury and declining physical abilities. Despite these risks, we can examine one of the most comprehensive statistics for evaluating a player's hitting performance: wRC+ (weighted Runs Created Plus). This statistic adjusts for park and league factors, allowing for better comparisons of players across different teams and seasons. In wRC+, a score of 100 represents the league average. Below is the breakdown of veteran hitters' wRC+ over the past few seasons.

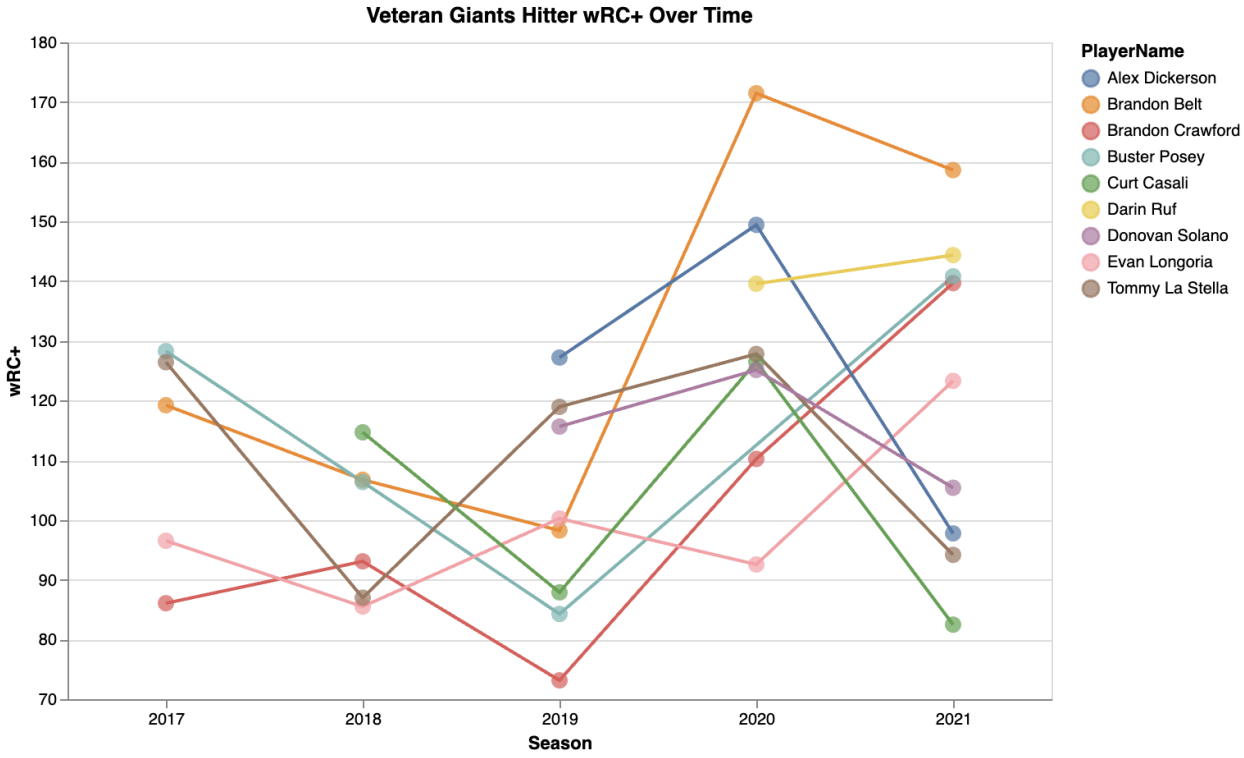

Figure 4.1: Veteran Giants' Hitting

Highlights include Buster Posey's remarkable comeback from 2019, Darin Ruf joining the Giants in 2020 after spending several years playing in Korea, Brandon Belt's resurgence in 2020 and 2021 following a slow decline, Brandon Crawford's breakout season, and Evan Longoria's bounce-back performance after a period of steady decline. Among other players, Donovan Solano had a decent season, while two other players were utilized more as bench

players. Here is the breakdown of plate appearances for these players:

Player Name	Age	PA	wRC+	WAR
Brandon Crawford	34	549	139.68	6.22
Buster Posey	34	454	140.82	5.33
Brandon Belt	33	381	158.61	3.59
Donovan Solano	33	344	105.40	1.39
Alex Dickerson	31	312	97.76	0.22
Darin Ruf	34	312	144.36	2.97
Evan Longoria	35	291	123.30	1.97
Tommy La Stella	32	242	94.17	0.64
Curt Casali	32	231	82.48	0.34

Table 4.2: Player Offensive Statistics with WAR

What stands out from this table is the significant number of strong plate appearances from these veteran players. Crawford, for instance, had a substantial 549 plate appearances while maintaining respectable wRC+ numbers. The other high-performing veterans also had solid plate appearances. Interestingly, the top five performers (Crawford, Posey, Belt, Ruf, Longoria) were all older than the team's average age of 30.32 years, with one player at 33, three at 34, and one at 35. Having one veteran perform well at this age and stage of their career is something to take note of, but having five perform at such a high level after less impressive previous seasons is an anomaly.

In the previous section, we compared actual versus projected stats extensively. A similar approach may be useful here to evaluate the team's performance. Let us a take a look at the disparity in some of these veteran's observed statistics with respect to their projections:

Player	Age	Projected OPS+	Observed wRC+	Projected WAR	Observed WAR	WAR Disparity
Brandon Crawford	34	83	139.68	1.2	6.22	5.02
Buster Posey	34	91	140.82	2.0	5.33	3.33
Darin Ruf	34	87	144.36	0.2	2.97	2.77
Steven Duggar	27	77	107.32	0.2	1.78	1.58
Brandon Belt	33	110	158.61	2.3	3.59	1.29
Evan Longoria	35	87	123.30	1.2	1.97	0.77
Donovan Solano	33	87	105.40	1.1	1.39	0.29
Wilmer Flores	29	105	113.66	1.7	1.87	0.17
Austin Slater	28	107	102.93	1.9	1.49	-0.41
Alex Dickerson	31	105	97.76	0.9	0.22	-0.68
Mike Yastrzemski	30	120	106.34	3.2	2.26	-0.94
Jason Vosler	27	80	64.13	0.9	-0.45	-1.35
Mauricio Dubón	26	84	75.09	1.7	0.14	-1.56

Table 4.3: Comparisons between Observed and Projected Statistics, Sorted by WAR Disparity

It is worth noting that Brandon Crawford alone posted a WAR value 5 points higher than expected, indicating that the team won 5 more games than projected this season largely due to his performance. Combined with the other veterans mentioned earlier, the team achieved a production value exceeding 12 additional wins beyond expectations. Although some younger players underperformed relative to their projections, the total WAR disparity among the players listed in this chart still results in over 10 more wins than anticipated.

CHAPTER 5

The Pitchers

Moving on to the pitching side, the Giants demonstrated strong performance in this area as well. Interestingly, the players they acquired through free agency or other means were not considered superstars before joining the team. Whether due to excellent coaching or some other factor, many of these players, both in the starting rotation and the bullpen, posted impressive numbers that seemed unlikely based on their prior seasons. Let us take a look at some of these pivotal players.

The most impressive performance came from starter Kevin Gausman. Before 2021, Gausman was regarded as a decent pitcher, but not close to a Cy Young candidate. Let us start by examining Gausman's stats from the seasons leading up to 2021:

Season	PlayerName	TeamName	Age	ERA	IP	ER	FIP	WAR
2021	Kevin Gausman	SFG	30	2.81	192.0	60.0	3.00	4.80
2020	Kevin Gausman	SFG	29	3.62	59.2	24.0	3.09	1.58
2019	Kevin Gausman	ATL	28	6.19	80.0	55.0	4.20	1.21
2019	Kevin Gausman	CIN	28	4.03	22.1	10.0	3.17	0.40
2018	Kevin Gausman	BAL	27	4.43	124.0	61.0	4.58	1.48
2018	Kevin Gausman	ATL	27	2.87	59.2	19.0	3.78	0.86
2017	Kevin Gausman	BAL	26	4.68	186.2	97.0	4.48	2.56

Table 5.1: Kevin Gausman's Performance Summary

Aside from his brief 22-inning stint in Cincinnati, it is evident that his FIP (Fielding Independent Pitching, a more comprehensive statistic than ERA) improved from the high 3's and low 4's to the low 3's over consecutive seasons. Additionally, his ERA, which was 4.30

in previous seasons, dropped to 3.00 during his time with the Giants. This 1.3 run difference over 9 innings could easily influence the outcome of several games, which is reflected in his rising WAR (Wins Above Replacement) with the Giants.

Although his WAR value for 2020 was 1.577, if we adjust for a full 162-game season instead of the shortened 60-game season, it jumps to 4.25, approaching his remarkable level of production in 2021.

Another pitcher to take notice of is Anthony DeSclafani, who was acquired from the Reds and debuted for the Giants in 2021. Here is his stat line up until that point:

Season	PlayerName	TeamName	Age	ERA	IP	ER	FIP	WAR
2021	Anthony DeSclafani	SFG	31	3.17	167.2	59.0	3.62	3.03
2020	Anthony DeSclafani	CIN	30	7.22	33.2	27.0	6.10	-0.06
2019	Anthony DeSclafani	CIN	29	3.89	166.2	72.0	4.43	2.56
2018	Anthony DeSclafani	CIN	28	4.93	115.0	63.0	4.83	0.81

Table 5.2: Anthony DeSclafani's Performance Summary

DeSclafani posted decent numbers in 2019, but his 2018 and 2020 seasons were less effective. However, his performance improved significantly with the Giants, with his ERA dropping to the low 3's and his WAR surpassing 3 for the first time in his career.

In addition to Gausman's and DeSclafani's strong performances, third-year player Logan Webb found his stride, reducing his ERA to 3.00, and increasing his WAR to a whopping 4.08. Previously, Webb had seen some struggles on the mound, posting two 5+ ERA seasons and never surpassing a WAR of more than 0.65. He nearly quadrupled his career WAR in just one standout season here. Here is the breakdown of his performance:

Season	PlayerName	TeamName	Age	ERA	IP	ER	FIP	WAR
2021	Logan Webb	SFG	24	3.03	148.1	50.0	2.72	4.09
2020	Logan Webb	SFG	23	5.47	54.1	33.0	4.17	0.65
2019	Logan Webb	SFG	22	5.22	39.2	23.0	4.12	0.50

Table 5.3: Logan Webb's Performance Summary

Other players to take account of in the starting rotation include Alex Wood, a former standout pitcher for the Dodgers, who contributed 138.2 innings with a respectable 3.82 ERA. Even a past-his-prime Johnny Cueto provided 114.2 innings with an ERA slightly above 4.00. The starting rotation undoubtedly played a crucial role in the team's success. Here is the breakdown amongst all other starters:

Season	PlayerName	TeamName	Age	ERA	IP	ER	FIP	WAR
2021	Alex Wood	SFG	30	3.83	138.2	59.0	3.48	2.53
2019	Alex Wood	CIN	28	5.80	35.2	23.0	6.38	-0.14
2018	Alex Wood	LAD	27	3.68	151.2	62.0	3.53	2.41
2017	Alex Wood	LAD	26	2.72	152.1	46.0	3.32	3.15
2021	Johnny Cueto	SFG	35	4.08	114.2	52.0	4.05	1.52
2020	Johnny Cueto	SFG	34	5.40	63.1	38.0	4.64	0.51
2018	Johnny Cueto	SFG	32	3.23	53.0	19.0	4.71	0.26
2017	Johnny Cueto	SFG	31	4.52	147.1	74.0	4.49	1.17

Table 5.4: Performance summary for Alex Wood and Johnny Cueto

I also wanted to analyze the total innings pitched by the Giants' starters and relievers compared to other teams. The Giants' starting pitchers logged 831.1 innings, ranking them 11th in the league for starters' innings. However, their starters produced the 5th highest WAR value among rotations that year, totaling 16.53. Here is the breakdown for starting pitching:

Season	Team	IP	ERA	FIP	WAR
2021	LAD	843.1	2.93	3.33	20.77
2021	MIL	847.2	3.13	3.29	20.32
2021	CHW	855.1	3.57	3.73	19.83
2021	PHI	843.0	4.25	3.84	16.64
2021	SFG	831.1	3.44	3.43	16.53
2021	BOS	812.0	4.49	3.87	15.21
2021	CIN	864.0	4.03	4.19	14.41
2021	NYY	829.1	3.91	4.01	14.38
2021	OAK	894.0	3.91	3.93	13.61
2021	TOR	836.1	3.79	4.05	12.94

Table 5.5: Team-Wide Starting Pitcher Statistics for the 2021 season

On the other hand, the Giants' bullpen threw a total of 623.2 innings, placing them 10th overall in bullpen innings. Despite this, they generated a WAR value of 5.42, ranking them 6th in that category. Here is the breakdown for the relief pitching statistics from that year:

Season	Team	IP	ERA	FIP	WAR
2021	CHW	548.0	3.97	3.75	8.09
2021	TBR	703.0	3.24	3.59	7.95
2021	NYY	606.0	3.56	3.76	7.23
2021	SEA	618.2	3.88	3.72	6.83
2021	LAD	608.2	3.16	3.83	6.75
2021	SFG	623.2	2.99	3.71	5.42
2021	MIA	635.2	3.81	3.80	5.36
2021	BOS	607.0	3.99	4.06	5.11
2021	LAA	645.1	4.59	4.19	4.71
2021	NYM	595.2	3.90	4.05	4.69

Table 5.6: Pitching statistics for selected teams in the 2021 season

While neither the bullpen nor the starting rotation ranked at the top among teams individually, their combined value was a remarkable achievement. It underscores the importance of both sides of the pitching staff working in tandem for overall effectiveness. I thought it might be interesting to check out the sum of the ranks for each teams starters and relievers. For example, the lowest score would be 2 if both the starting and relief pitching of a team ranked 1st, while the max score is 60 if they rank 30th in both categories. Here are the results I generated from looking at that:

Team	Sum of Rankings
CHW	4
LAD	6
SFG	11
NYY	11
BOS	14
TBR	17
MIL	18

Table 5.7: Sum of Rankings for the WAR Statistic Across Teams

In terms of their combined ranking of starters and relievers, the Giants tied with the Yankees that year for the 3rd best ranking total in the MLB that year. Paired with their offensive prowess, the cohesiveness of this pitching unit proved to be much better than expected. Rather than looking at the distinct rankings, I thought it may also be useful to take a look at the total WAR produced between both units. Here's what I got from that.

Team	Total WAR
CHW	27.92
LAD	27.52
MIL	23.55
SFG	21.94
NYY	21.61
BOS	20.32
TBR	18.91

Table 5.8: Total Pitching WAR for 2021 teams

Once again, the Giants prove their standing in that they posted the 4th best overall WAR as a team of pitchers in the league this year.

One key difference in this section compared to earlier ones is the focus on the exceptional

performance rather than simply the level of 'over'-performance. While the Giants exhibited elite pitching throughout the season, it is the trend of overperforming that stands out as particularly impressive.

Let us revisit some players previously discussed and introduce a few new names into the mix. Going back to the starting rotation, here's a breakdown of their projected versus actual statistics, ranked by the disparity in WAR (Wins Above Replacement) between observed and projected values:

Player	IP_proj	IP_actual	ERA_proj	ERA_actual	FIP_proj	FIP_actual
Kevin Gausman	151.3	192.0	3.93	2.812500	3.61	3.003367
Logan Webb	117.0	148.1	4.00	3.033708	3.99	2.718349
Anthony DeSclafani	118.0	167.2	5.03	3.166998	5.07	3.617350
Johnny Cueto	100.3	114.2	4.66	4.081396	4.64	4.050848
Alex Wood	98.3	138.2	3.75	3.829327	3.88	3.480130

Table 5.9: Pitching statistics comparison (Part 1)

Player	WAR_proj	WAR_actual	WAR_disparity
Kevin Gausman	2.5	4.800728	2.300728
Logan Webb	1.8	4.085827	2.285827
Anthony DeSclafani	1.0	3.031672	2.031672
Johnny Cueto	0.8	1.515719	0.715719
Alex Wood	1.9	2.525144	0.625144

Table 5.10: Pitching statistics comparison (Part 2)

Three players stand out with a WAR disparity of over 2.00, indicating they each contributed significantly more to the team's success than anticipated. Collectively, these players (Gausman, Webb, DeSclafani) were projected to pitch 387 innings but delivered 508 innings, surpassing expectations by 121 innings. This surplus not only saved bullpen arms for critical situations but also featured pitchers with a combined ERA around a very respectable 3.00.

To put it in perspective, this extra 121 innings equates to nearly 13 and a half full games of pitching over a 162-game season—truly remarkable.

Additionally, while Wood and Cueto didn't match the extraordinary levels of their peers, they still made substantial contributions. At 35 years old, Cueto posted a WAR value of 1.52, his highest since 2016, and pitched over 114. Wood also achieved the second-highest WAR of his career, surpassed only by his 2017 performance with the Dodgers, boasting a value of 3.15 in that year.

The success of the pitching staff compared to their projections extends beyond the starting rotation. There are many more names to consider as we look into the heart of the bullpen. Before examining the observed stats, let's first look at the preseason projections for the players who comprised the 2021 Giants bullpen:

Year	Team	Player	IP	ERA	FIP	WAR
2021	San Francisco Giants	Camilo Doval	48.3	5.21	5.350	-0.5
2021	San Francisco Giants	Dominic Leone	56.7	4.45	4.345	0.0
2021	Los Angeles Dodgers	Jake McGee	45.3	4.17	4.220	0.2
2021	San Francisco Giants	Jarlin García	68.3	4.74	4.950	0.1
2021	Cincinnati Reds	Jay Jackson	58.0	4.34	4.270	0.3
2021	Philadelphia Phillies	Jose Alvarez	47.7	4.15	4.190	0.4
2021	San Francisco Giants	Tyler Rogers	70.3	3.71	3.830	0.8
2021	Minnesota Twins	Zack Littell	68.0	4.37	4.420	0.2

Table 5.11: Relief Pitchers' 2021 Preseason Projections

A couple of interesting points stand out. First, only about half of the bullpen contributors were even predicted to be on the Giants' roster for the season. Second, there were no standout stars in the projections. Tyler Rogers had the highest projected WAR at just 0.8, while most others hovered around 0.2. Particularly, rookie Camilo Doval was predicted to have a negative contribution to the team. Now, let's take a look at what actually happened.

Season	TeamName	PlayerName	IP	ERA	FIP	WAR
2021	SFG	Camilo Doval	27.0	3.000000	3.466330	0.272361
2021	SFG	Dominic Leone	53.2	1.509317	3.076866	0.898540
2021	SFG	Jake McGee	59.2	2.715084	3.354392	1.108436
2021	SFG	Jarlín García	68.2	2.621360	3.767122	0.413589
2021	SFG	Jay Jackson	21.2	3.738462	4.046957	0.060009
2021	SFG	Jose Alvarez	64.2	2.365980	3.154570	1.024039
2021	SFG	Tyler Rogers	81.0	2.222222	3.281145	1.266276
2021	SFG	Zack Littell	61.2	2.918920	3.867332	0.276496

Table 5.12: Relief Pitchers' Observed Statistics

To highlight the differences between the projections and actual performance, I calculated a comparison statistic for each column. The bullpen was projected to throw 520 innings but only ended up throwing 438 innings, which underscores the effectiveness of the starting rotation this season. The projected average ERA for the bullpen was a substantial 4.40, while the actual average ERA was an impressive 2.64. For FIP, the projected average was 4.44, but the observed value was 3.50. This significant improvement suggests a combination of good fortune, elite defense, and the pitcher-friendly environment of their home ballpark. Despite these factors, lowering the FIP by nearly a full point is remarkable.

Finally, we turn to the tell-all statistic: WAR. The projections estimated a total bullpen WAR of 1.5, while the actual value was 5.31, showcasing another impressive improvement. Almost every player in the bullpen outperformed their projected WAR, with some experiencing breakout seasons. To further explore this, let's revisit some specific names on the roster.

First of all, let's take a look at rookie Camilo Doval. As much as I like to rely on statistics and numbers to generate these results, sometimes they just simply don't tell the whole story. In the above table 4.12, we'll notice that Doval threw just 27 innings across the season and posted a WAR of 0.27, the 2nd lowest WAR contribution of the bunch. While this is true, it

doesn't capture the fact that Doval's "ERA at the end of May after nearly 11 innings pitched over 13 games was 7.59 with a WHIP of 1.59." To add onto that, he pitched only two more innings in one outing in August before returning to the majors on September 5th, the last month of the season. Doval was seriously struggling. And as talented as any rookie may be, the majors is a different beast, so it doesn't come as much of a surprise. But the way that he turned his year around in a singular month proved to everyone that he had what it took to be an elite pitcher in the league. "In 13 games in September, his ERA was 0.00000000000. His WHIP was 0.65. He didn't let in a single run for the rest of the season. He was named the National League Reliever of the Month in early October." [Ken21b] Now, of course Doval does not tell the whole story for the Giants' success. After all, he was only effective for about a month of the season. However, he turned what could've been a disastrous season into becoming the most feared reliever in baseball.

Second off, let's take a look at Tyler Rogers. Rogers was not a young guy when he debuted with the Giants in 2020, already 29 years old, with an uncommon "submarine" delivery that boasted an average fastball speed of just 82.8 miles per hour. [rog] In a world where the average fastball in the MLB is 93.7 miles per hour [mlb22], it begs the question: how on Earth can somebody throw so slowly, yet so effectively? Rogers was the epitome of that answer in this season. He posted the highest WAR total amongst all relievers on the team at a value of 1.26, and threw the most innings out of any reliever on the team with a remarkable 81 innings from the bullpen (the 3rd most in the majors). Across the entire year, Rogers was consistent in his performance, appearing in 80 games (the 2nd most of any pitcher in the majors that year) and posting an ERA of 2.22 in that time. Simply put, Rogers was the workhorse that the team needed and could count on time and again to hold a close game.

Moving deeper into the bullpen, we'll touch on the numbers of Jake McGee. McGee is an interesting case here. One might notice when looking at the above table that his stats in innings pitched are pretty similar to that of Jarlin Garcia, Jose Alvarez, or Zack Littell, but what distinguished him from the crowd and allowed him to rack up that 2nd highest total WAR of 1.11 was the fact that McGee was a lights-out closer for the majority of the season.

He pitched in 62 games across the season [Ken21a], providing solid contribution to the team and giving them those saves, especially towards the beginning of the season, that provided just a little extra oomph into getting that 107 win season.

It is no secret that the Giants' pitching staff, both starters and relievers, played an integral role in the team's success in 2021. The remarkable performances of the starting rotation, including significant contributions from Gausman, Webb, and DeSclafani, set the stage for a consistently competitive team. Meanwhile, the bullpen, anchored by standout performances from Rogers, Doval, and McGee, demonstrated resilience and effectiveness far beyond preseason projections. The collective effort of these pitchers, who often surpassed expectations, underscores the importance of depth and versatility in a successful pitching staff. This cohesive and unexpectedly stellar performance from the pitching unit was a crucial factor in the Giants' ability to defy the odds and achieve their record-setting season.

CHAPTER 6

Comparable Teams

I had hoped this section would be longer, but the reality is that baseball's inherent randomness, coupled with the limited data points within my collection timeframe, made it challenging to discern meaningful patterns. And in the end, it was difficult to draw comparisons between teams that significantly overperformed their projections. To illustrate this, the San Francisco Giants had a differential of 0.197 between their actual win rate and their projected win rate. The table below shows other teams that we could potentially compare:

Tm	Year	W-L% (Actual)	W-L% (Expected)	WL Percent Diff
San Francisco Giants	2021	0.660	0.463	0.197
Boston Red Sox	2018	0.667	0.537	0.130
Baltimore Orioles	2023	0.623	0.494	0.129
Oakland Athletics	2018	0.599	0.481	0.118
Baltimore Orioles	2022	0.512	0.395	0.117
Miami Marlins	2020	0.517	0.400	0.117
Los Angeles Dodgers	2022	0.685	0.574	0.111
Los Angeles Dodgers	2017	0.642	0.531	0.111
Minnesota Twins	2019	0.623	0.512	0.111
Cleveland Indians	2017	0.630	0.519	0.111

Table 6.1: Win-Loss Percentage Comparison for Top 10 Overperformers

As shown in the table, the Giants' overperformance is significantly higher than that of the next team. The win-loss differential decreases to 0.111 by the time we reach the 10th team on the list. While I examined the characteristics of the next three most overperforming teams (2018 Red Sox, 2023 Orioles, 2018 Athletics), no particular traits stood out. These teams not only performed better than expected but also excelled overall, each finishing first in their divisions or winning the Wild Card. They had multiple standout players, though none with as significant a disparity as Brandon Crawford's 5+ WAR.

An important observation from this distribution is that teams expected to perform poorly have a greater potential for overperformance. Interestingly, half of the teams listed were projected to have a win rate above 0.500, while the other half, including the Giants, were projected below 0.500. Teams already expected to be average or good face a larger challenge to exceed expectations significantly compared to those initially counted out.

For future breakout overperformers like the 2021 Giants, we might see a team like the 2018 Red Sox, which posted a 108-win season, already considered better than average. It's the teams like the Giants, underestimated from the start, that have the potential to evolve into unforeseen powerhouses.

CHAPTER 7

Conclusion

The 2021 San Francisco Giants' remarkable performance serves to show the unpredictable nature of sports and the limitations of preseason projections. Despite the sophisticated statistical models and extensive historical data used to forecast team performance, the Giants defied expectations by achieving a franchise-record 107 wins. This exceptional outcome, while rare, highlights several key insights regarding the nature of sports performance and the factors that can lead to such extraordinary results.

The analysis revealed that extreme overperformance, like that of the Giants, is an outlier within a largely predictable distribution. The simulated data indicated a low probability of such exceptional performances, reflecting the rarity of achieving such a high differential between actual and projected win rates. This reinforces the notion that while statistical models provide valuable insights, they cannot account for every variable influencing team success.

At the individual level, the rejuvenation of veteran players on the 2021 Giants was a pivotal factor in the team's success. The standout performances of veterans such as Brandon Crawford, Buster Posey, and Kevin Gausman were instrumental. Crawford's resurgence, alongside significant contributions from other experienced players, demonstrated that age and past performance are not always reliable indicators of future success. The extraordinary WAR values and wRC+ statistics among these players were key drivers in the Giants' overperformance.

On the pitching side, the Giants benefitted from remarkable improvements by both veteran and emerging players. Kevin Gausman's transformation into a Cy Young candidate, Anthony DeSclafani's significant improvement, and Logan Webb's breakout season were crit-

ical to the team's success. The collective performance of the starting rotation, marked by significant reductions in ERA and increases in WAR, provided a solid foundation for the team's winning season. The analysis of the total innings pitched by the Giants' starters and relievers compared to other teams highlighted the combined value of their pitching staff. Despite not having the highest total innings, the Giants' pitchers produced one of the highest WAR values, underscoring the importance of both the starting rotation and the bullpen working effectively in tandem.

The success of the bullpen, particularly the performance of relievers like Camilo Doval, Tyler Rogers, and Jake McGee, further accentuated the team's overperformance. Doval's late-season turnaround, Rogers' consistent effectiveness with an unconventional pitching style, and McGee's reliability as a closer contributed significantly to the Giants' success. The bullpen's collective performance, with a much lower ERA and FIP than projected, highlighted the impact of elite defense, a pitcher-friendly environment, and perhaps a touch of good fortune.

The 2021 San Francisco Giants exemplify the unpredictability and excitement of sports. Their success story highlights the limitations of statistical projections and the importance of intangible factors that contribute to team dynamics and individual performances. While advanced analytics will continue to play a crucial role in understanding and forecasting sports outcomes, the human element remains an important and unpredictable aspect of the game. The Giants' remarkable season serves as a reminder that in sports, anything is possible, and that the love of the game is often fueled by these extraordinary and unexpected achievements.

REFERENCES

[bbr] https://www.baseball-reference.com/.

[Bra22] Tim Keown Buster Olney Bradford Doolittle, Alden Gonzalez. "Why San Francisco Giants have gone from 107 wins to sub-.500." https://www.espn.com/mlb/story/_/id/34549809/why-san-francisco-giants-gone-107-wins-sub-500, September 2022.

[Bri21] Grant Brisbee. "The 2021 Giants were a treasure trove of fun facts, records and historical anomalies." https://www.nytimes.com/athletic/2905388/2021/10/21/the-2021-giants-were-a-treasure-trove-of-fun-facts-records-and-historical-anomalies/, October 2021.

[fan] https://www.fangraphs.com/.

[gau] https://baseballsavant.mlb.com/leaderboard/pitch-arsenal-stats?type=pitcher&pitchType=&year=2021&team=&min=10&_gl=1*5tuklx*_ga*MTc5NTgyODE5Mi4xNzA5NDE4NDkx*_ga_N8YFCZLYSZ*MTcxOTI2MTIxNC4yNS4wLjE3MTkyNjEyMTQuMC4wLjk3Nzg3NDYwMQ.

[Ken21a] Steven Kennedy. "Country Hard Ball." https://www.mccoveychronicles.com/2021/12/9/22826504/san-francisco-giants-jake-mcgee-2021, December 2021.

[Ken21b] Steven Kennedy. "The Iceman Cometh." https://www.mccoveychronicles.com/2021/11/18/22790202/san-francisco-giants-camilo-doval-2021, November 2021.

[LeD] James LeDoux. "Pybaseball." https://github.com/jldbc/pybaseball.

[mlb22] "MLB velocity, shifts set records; average lowest since 1968." https://spectrumlocalnews.com/nys/buffalo/ap-online/2022/10/06/mlb-velocity-shifts-set-records-average-lowest-since-1968, October 2022.

[Pet21] Mike Petriello. "How 2021's most shocking conteder did it." https://www.mlb.com/news/how-giants-became-2021-most-shocking-contender Accessed: 2024-07-12, July 2021.

[rog] https://pitcherlist.com/player/tyler-rogers/ Accessed 2024-07-12.

[sav] https://baseballsavant.mlb.com/.

[Szy16a] Dan Szymborski. "ZiPS 2017 standings projections: Cubs' odds to repeat, surprise contenders and more." https://www.espn.com/mlb/insider/story/_/id/17964821/zips-2017-standings-projections-cubs-odds-repeat-surprise-contenders-more-mlb, November 2016.

[Szy16b] Dan Szymborski. "ZiPS projected standings/World Series chances: Where do Royals, Mets check in?" https://www.espn.com/mlb/insider/story/_/id/15123627/zips-projected-standings-world-series-chances-where-do-royals-mets-check-mlb, April 2016.

[Szy17] Dan Szymborski. "Early projections for 2018 standings: ZiPS loves L.A." https://www.espn.com/mlb/insider/story/_/id/21265001/early-projections-2018-standings-zips-loves-la, November 2017.

[Szy19] Dan Szymborski. "The Super-Duper Official Final Preseason-2019 ZiPS Projected Standings." https://blogs.fangraphs.com/the-super-duper-official-final-preseason-2019-zips-projected-standings/, March 2019.

[Szy20a] Dan Szymborski. "2021 ZiPS Projections: San Francisco Giants." https://blogs.fangraphs.com/2021-zips-projections-san-francisco-giants/, December 2020.

[Szy20b] Dan Szymborski. "The Obscenely Late, Obscenely Early ZiPS Projected Standings." https://blogs.fangraphs.com/the-obscenely-late-obscenely-early-zips-projected-standings/, June 2020.

[Szy21a] Dan Szymborski. "The Hopefully-Not-Too-Regretful 2021 ZiPS Projections, American League." https://blogs.fangraphs.com/the-hopefully-not-too-regretful-2021-zips-projections-american-league/, March 2021.

[Szy21b] Dan Szymborski. "The Hopefully-Not-Too-Regretful 2021 ZiPS Projections, National League." https://blogs.fangraphs.com/the-hopefully-not-too-regretful-2021-zips-projections-national-league/, March 2021.

[Szy22a] Dan Szymborski. "The Hopefully-Not-Horrifyingly-Inaccurate 2022 ZiPS Projections: American League." https://blogs.fangraphs.com/the-hopefully-not-horrifyingly-inaccurate-2022-zips-projections-american-league/, April 2022.

[Szy22b] Dan Szymborski. "The Hopefully-Not-Horrifyingly-Inaccurate 2022 ZiPS Projections: National League." https://blogs.fangraphs.com/the-hopefully-not-horrifyingly-inaccurate-2022-zips-projections-national-league/, April 2022.

[Szy23] Dan Szymborski. "The Official Hopefully-Not-Too-Erroneous 2023 ZiPS Projections." https://blogs.fangraphs.com/the-official-hopefully-not-too-erroneous-2023-zips-projections/, March 2023.

[Szy24] Dan Szymborski. "The Official (And Hopefully Not Too Cringe) 2024 ZiPS Projections." https://blogs.fangraphs.com/the-official-and-hopefully-not-too-cringe-2024-zips-projections/, March 2024.