

How Do Different Aspects of MLB Gameplay Impact Team Success?



Nolan Lowe

DATA6150 – Data Science Foundations

Professor Pang

November 30th, 2025

Table of Contents

Introduction.....	3
Batting and Baserunning.....	3
Fielding.....	4
Pitching.....	4
Dataset	5
Methodology.....	6
Results and Analysis	7
Discussion and Conclusions	11
References	12

Introduction

Baseball has often been referred to as “America’s Pastime”, largely earning this connotation through its combination of historical significance and its slow, segmented play style. Its significance within American culture and the sporting world gives rise to a multitude questions that can be answered with the help of data science practices. Among these questions is the focus of this report, namely, which aspects of MLB gameplay are most closely related with team success?

This question can be broken down into four secondary questions, which are as follows:

- 1) What is the impact of batting and baserunning on winning in the regular season?**
- 2) What is the impact of fielding on winning in the regular season?**
- 3) What is the impact of pitching on winning in the regular season?**
- 4) How do these relationships change in the playoffs?**

The first three inherently represent three different areas of play, with batting and baserunning being determined by offensive statistics and fielding and pitching being determined by defensive statistics. The fourth evidently references these same areas, but in the context of the postseason. In understanding how these facets of the game impact winning both in the regular season and the playoffs, fans of the sport can gain greater appreciation for the statistics within baseball and team officials can make educated decisions on how to structure their rosters for the most success.

Batting and Baserunning

Investigating batting and baserunning is done with the intention of determining how a team’s offensive output correlates to winning games. The specific statistics that will be used to represent this component of play are runs scored per game, total bases stolen, batting average, on base percentage, slugging percentage, and home runs. Breaking this category down into individual statistics not only provides quantitative values to perform analysis on, but also allows differing components of offense to be examined and compared. For example, there has long been a debate on whether contact hitting (hitting for average) or slugging (hitting for power) is a better strategy for team success. Such a debate will be settled within the contents of this report.

Fielding

Fielding represents the core component of a team's defense. Being able to turn plays to get outs, particularly those that may be considered more difficult, is often the difference between winning and losing teams. The statistics to be examined here are defensive efficiency, defensive runs saved, and errors.

Pitching

Pitching represents a secondary, yet arguably more important aspect of team defense. For the purpose of this investigation, the pitching category will capture both starting pitcher and bullpen statistics. These statistics are inclusive of quality start percentage, average game score, innings pitched per game started (on average), and save percentage.

Dataset

The dataset that will be used to perform all quantitative analysis is linked below.

Baseball Reference:

https://www.baseball-reference.com/leagues/MLB-standings.shtml#all_expanded_standings_overall

This website features all data necessary for the analysis, including team standings, playoff results, and all batting, pitching and fielding stats by team for the 2025 season. Other years can also be accessed if needed.

Methodology

In order to answer the question posed in the report's introduction, a data science-centered approach will be undertaken. Using Python, and the associated libraries numpy, pandas, matplotlib and seaborn, a linear regression will be performed between each statistic identified in the introduction and wins on a team-by-team basis. The resulting correlation coefficients will then be compared to determine which of these stats are most telling of a winning team. Further, with the correlation coefficients known, a general impression can be made for each area of the game and their relative importance to overall success. Once this is done, the process is then repeated for the playoff data.

Results and Analysis

The following tables provide a summary of the correlation coefficients between wins and the considered statistics.

Table 1: Correlation coefficients for batting and baserunning statistics.

Batting and Baserunning	
Statistic	Correlation Coefficient
Runs per Game	0.712
Total Bases Stolen	0.172
Batting Average	0.557
On Base Percentage	0.724
Slugging Percentage	0.500
Total Home Runs	0.342
Average	0.501

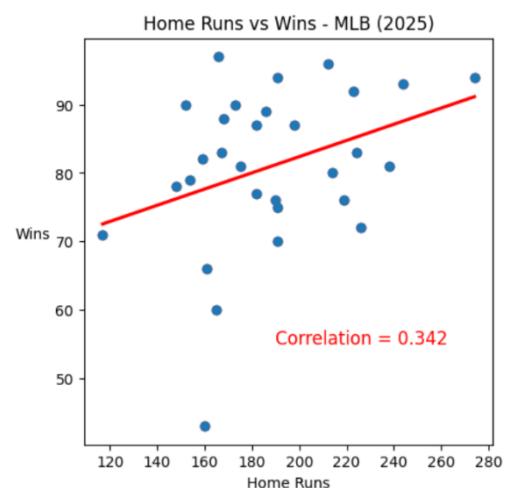
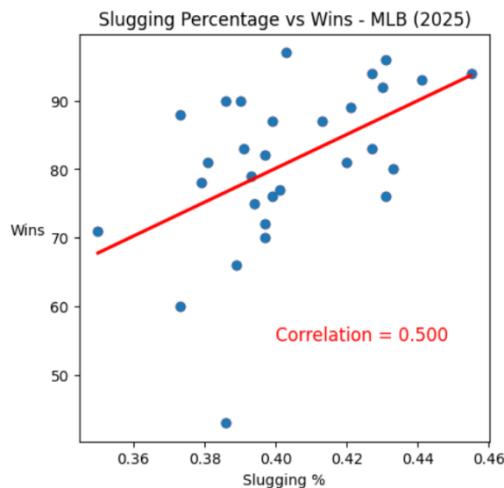
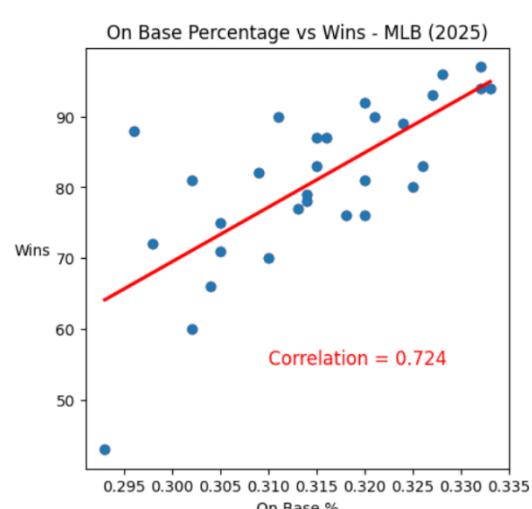
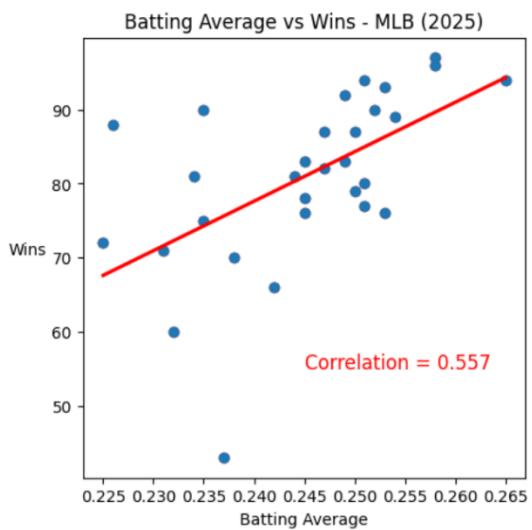
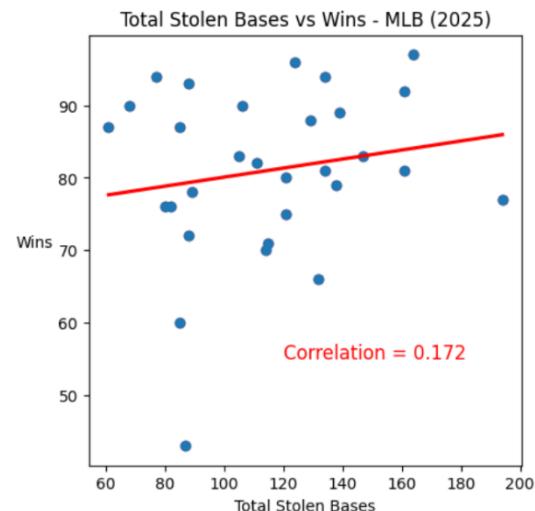
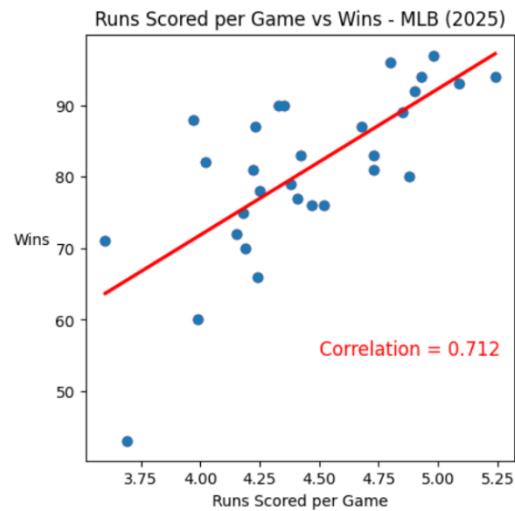
Table 2: Correlation coefficients for fielding statistics.

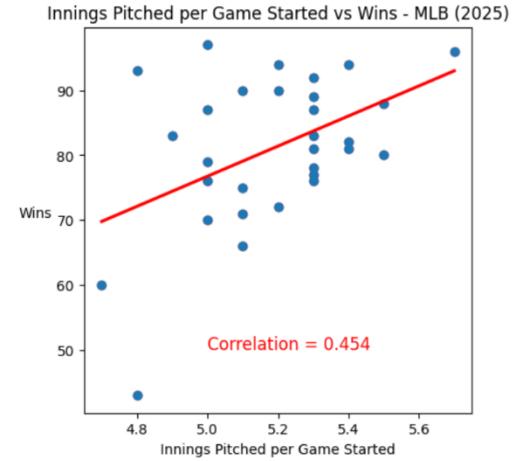
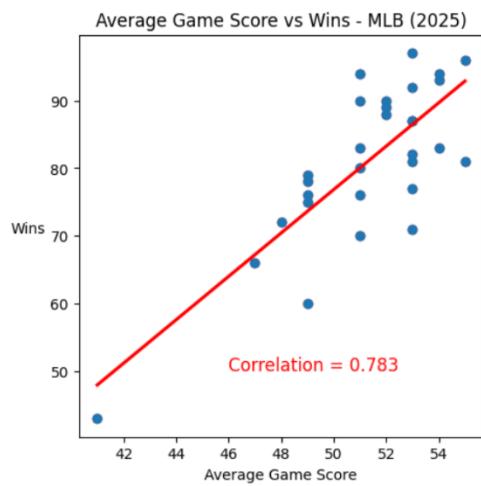
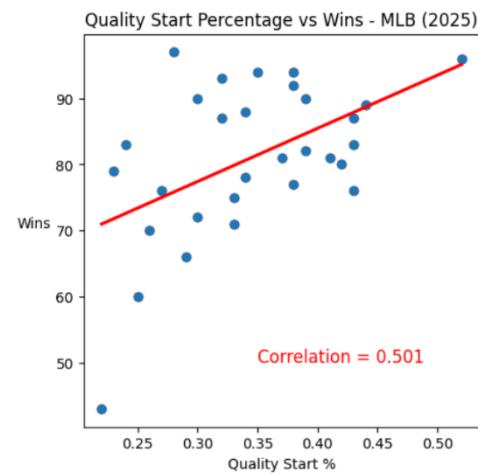
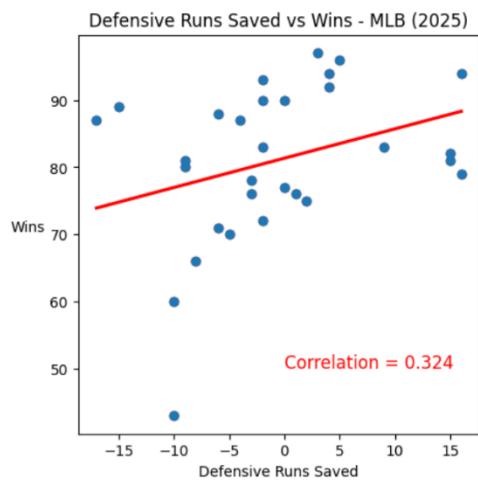
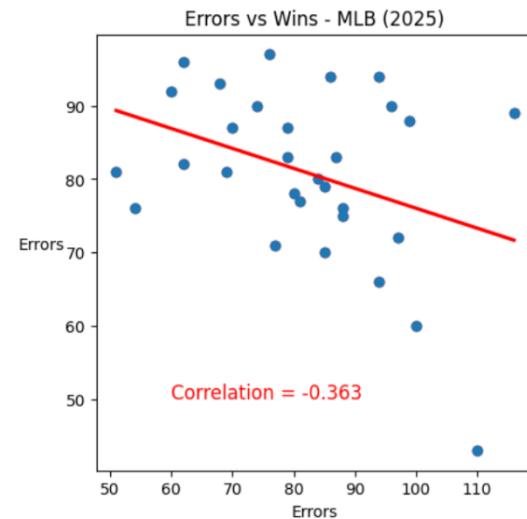
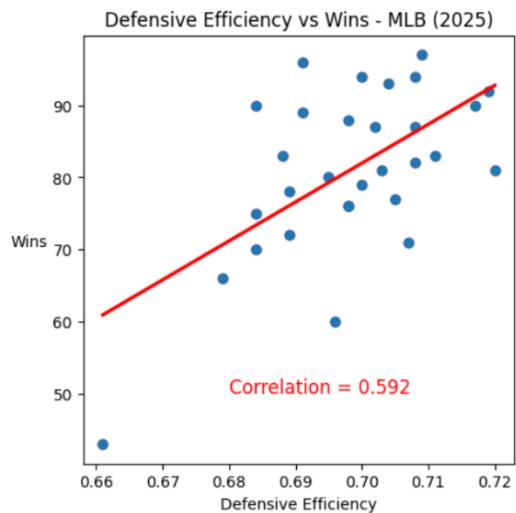
Fielding	
Statistic	Correlation Coefficient
Defensive Efficiency	0.592
Defensive Runs Saved	0.324
Errors	-0.363
Average	0.426

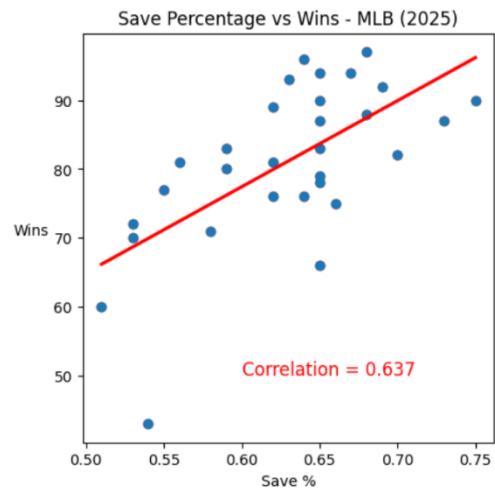
Table 3: Correlation coefficients for pitching statistics.

Pitching	
Statistic	Correlation Coefficient
Quality Start Percentage	0.501
Average Game Score	0.783
Innings Pitched per Start	0.454
Save Percentage	0.637
Average	0.594

The following figures detail the correlations between the identified variables and wins.







Discussion and Conclusions

References