What Truly AffectsWinning

Seth Wilkins *CSCI department*  
AUM

swilkin4@aum.edu

# Abstract

This project looks at how the most important factors for winning in Major League Baseball have changed over the course of MLB history. Using data from the Lahman Baseball Database put through SQL queries, we analyzed team performance based on key stats like home runs, ERA, walks and strikeouts allowed, runs scored, stolen base value, and payroll.

Our findings show a clear shift in offensive strategy starting in the late 1960s. Since then, teams that rank towards the top of their league in homeruns have generally been much more successful. The value of stolen bases has also increased, with teams generally having better records the higher their stolen base value is.

Pitching and defense have remained consistently important. Teams with top-five rankings in ERA, walks allowed, and strikeouts allowed have had a much higher chance on average of making the postseason across all of MLB history. When it comes to payroll, teams that spend within the top fifth of the league tend to win more championships, however about half of teams in this bracket have missed the playoffs. In all, the project highlights how the game has evolved and which stats teams should still focus on if they want to win.

# Introduction

This project will examine how winning has changed throughout MLB history. It will view a few different statistics: win percentage, home runs, stolen base value, strikeouts, earned run average, strikeouts allowed, runs allowed, and team salary. This project will help to determine how these stats have changed over the 100+ year history of the MLB. Unfortunately for the early years of MLB history, 1876 until the 1920s, not all information is available.

## Abbreviations and Acronyms

* teamID: The ID for the team.
* yearID: The year the following statistics come from
* franchise: the franchise each teamID belongs to. Links ID together for teams that had name changes but still belong to the same franchise. For example: the Brooklyn Dodgers changed their name to the Los Angeles Dodgers but still belong to the same franchise
* winPCT: The percentage of games that the team won
* DivWin: Team won their division
* LgWin: Team won their league
* WsWin: Team won the World Series
* runs: amount of runs the offense scored
* hr: amount of home runs the offense hit
* bb: amount of times the offense walked
* so: amount of times the offense struckout
* sbvalue: the overall value provided by stolen bases. Formula: (successful stolen bases x 0.3) / (caught stealing x 0.6)
* ra: runs allowed by the pitching staff
* era: average earned runs allowed by the pitching staff
* hra: home runs allowed by the pitching staff
* bba: walks allowed by the pitching staff
* soa: strikeouts by the pitching staff
* e: errors by the defense
* salary: team salary

##### Methodology

##### The project utilized the Lahman Baseball Database, developed by Sean Lahman of the Society for American Baseball Research. The analysis specifically focused sections on team history, team offense, team pitching, and team salary data.

Key queries included:

Identifying teams with the highest/lowest salary and comparing it to win percentage and postseason success.

Ranking teams by performance metrics like home runs, ERA, and runs scored.

Calculating averages and differences from league norms to evaluate what statistical advantages championship teams had over their competition.

Conclusion

As baseball history has evolved, the ways teams try to win have changed a lot, especially on offense. It found that in today’s game, teams tend to be more successful when they hit more home runs and get more value out of stolen bases. That shift really started in the late 1960s and has only gotten stronger since then.

Before the 1950s, only a few teams each year had a positive stolen base value. But in the last 30 years, about 25 out of 30 teams each season have had a positive value. Around 90% of the time post 1960, the teams that have been in the top 5 home runs hit per year have made the playoffs. Prior to the 60s, it was closer to 50%. Another finding made was that despite the number of total strikeouts increasing in the last 25 years, teams that made the playoffs had lower strikeout rates on average than teams before.

On the defensive side, statistics have remained steady. Generally, teams that were top 5 in ERA, walks allowed, or strikeouts allowed have made the playoffs far more often than they have not, and this has not changed much over the decades. The correlation of salary to winning has largely remained unchanged as well. Teams with a bottom 5 salary have only managed to make the playoff 3 times since 1980. However, teams with a top 5 salary have an increased number of World Series winners but a decreased rate of teams making the playoffs. Another finding made was that despite the number of total strikeouts increasing in the last 25 years, teams that made the playoffs had lower strikeout rates on average than teams before then.

Working on this project has strengthened my understanding of database design and data management. It also has improved my problem-solving skills by teaching me how to structure information efficiently.

##### References

1. Samford University Sports Analytics, “MLB Winning Percentage Breakdown: Which Statistics Help Teams Win More Games,” *Samford University Sports Analytics Fans*, 2022. [Online]. Available: <https://www.samford.edu/sports-analytics/fans/2022/MLB-Winning-Percentage-Breakdown-Which-Statistics-Help-Teams-Win-More-Games>.
2. J. R. Chabot and S. K. Normand, “The Contributing Factors to Winning in Baseball,” *ResearchGate*, 2015. [Online]. Available: <https://www.researchgate.net/publication/276026339_The_Contributing_Factors_to_Winning_in_Baseball>.
3. Society for American Baseball Research, “Lahman Baseball Database,” *SABR*, 2024. [Online]. Available: <https://sabr.org/lahman-database/>.
4. GitHub repository <https://github.com/Seth0370/What-Truly-Affects-Winning.git>

Full Code:

-- salary queries

-- highest budget

select teamID, yearID, salary, wins, winPCT, WsWin

from salaries natural join team\_record

where (yearID, salary)

in (select yearID, max(salary)

from salaries

group by yearID);

-- highest budget albb highest win

select teamID, yearID, salary, wins, winPCT, WsWin

from salaries natural join team\_record

where (yearID, wins)

in (select yearID, max(wins)

from team\_record

group by yearID)

and (yearID, salary)

in (select yearID, max(salary)

from salaries

group by yearID);

-- lowest budget

select teamID, yearID, salary, wins, winPCT, DivWin, LgWin, WsWin

from salaries natural join team\_record

where (yearID, salary)

in (select yearID, min(salary)

from salaries

group by yearID);

-- lowest 5 budgets

SELECT teamID, yearID, salary, wins, winPCT, WsWin

FROM (

SELECT s.teamID, s.yearID, s.salary, t.wins, t.winPCT, t.WsWin,

RANK() OVER (PARTITION BY s.yearID ORDER BY s.salary asc) AS salary\_rank

FROM salaries s

NATURAL JOIN team\_record t

) ranked\_salaries

WHERE salary\_rank <= 5

ORDER BY yearID, salary asc;

-- era rank for ws winner

with era\_stats as

(select yearID, teamID, era, rank() over (Partition by yearID order by era asc) as era\_rank,

AVG(era) over (Partition by yearID) as league\_avg\_era

from team\_record natural join team\_pitching)

select yearID, teamID, era, era\_rank, round(league\_avg\_era,2), round(league\_avg\_era-era, 2) as era\_better\_than\_avg

from era\_stats natural join team\_record

where WsWin = 'Y'

order by yearID;

-- how runs rank change

with run\_stats as

(select yearID, teamID, runs, rank() over (Partition by yearID order by runs desc) as run\_rank,

AVG(runs) over (Partition by yearID) as league\_avg\_runs

from team\_record natural join team\_offense)

select yearID, teamID, runs, run\_rank, round(league\_avg\_runs,2), round(runs-league\_avg\_runs, 2) as runs\_better\_than\_avg

from run\_stats natural join team\_record

where WsWin = 'Y'

order by yearID;

-- how hr rank changed

with hr\_stats as

(select yearID, teamID, hr, rank() over (Partition by yearID order by hr desc) as hr\_rank,

AVG(hr) over (Partition by yearID) as league\_avg\_hr

from team\_record natural join team\_offense)

select yearID, teamID, hr, hr\_rank, round(league\_avg\_hr,2), round(hr-league\_avg\_hr, 2) as hr\_better\_than\_avg

from hr\_stats natural join team\_record

where WsWin = 'Y'

order by yearID;

-- salary rank of WS winner

with budget\_stats as

(select yearID, teamID, salary, rank() over (Partition by yearID order by salary desc) as salary\_rank,

AVG(salary) over (Partition by yearID) as league\_avg\_salary

from team\_record natural join salaries)

select yearID, teamID, salary, salary\_rank, round(league\_avg\_salary,2), round(salary-league\_avg\_salary, 2) as salary\_better\_than\_avg

from budget\_stats natural join team\_record

where WsWin = 'Y'

order by yearID;

-- runs scored vs era

with win\_stats as

(select yearID, teamID, salary, rank() over (Partition by yearID order by salary desc) as salary\_rank, runs, rank() over (Partition by yearID order by runs desc) as run\_rank,

hr, rank() over (Partition by yearID order by hr desc) as hr\_rank, era, rank() over (Partition by yearID order by era asc) as era\_rank,

AVG(era) over (Partition by yearID) as league\_avg\_era, AVG(wins) over (Partition by yearID), rank() over (Partition by yearID order by wins desc) as win\_rank

from team\_record natural join salaries natural join team\_offense natural join team\_pitching)

select yearID, teamID, salary, salary\_rank, runs, run\_rank, hr, hr\_rank, era, era\_rank, wins, win\_rank

from team\_record natural join win\_stats

where win\_rank <= 5

order by yearID;

-- how so have changed

with so\_stats as

(select yearID, teamID, so, rank() over (Partition by yearID order by so desc) as so\_rank, hr, rank() over (Partition by yearID order by hr desc) as hr\_rank

from team\_record natural join team\_offense)

select yearID, teamID, hr, hr\_rank, so, so\_rank

from team\_record natural join so\_stats

where so\_rank <= 10

order by yearID;