BAIS:3200
Project Report

Davis Hoffstatter

Major League Baseball 2012-2018 Season Analysis G12

Introduction

Baseball has always been a game of numbers, and teams are constantly looking for the key stats that lead to wins and losses. In this project, we will perform a descriptive analysis to identify which baseball metrics have the biggest impact on a team's success over a five-season period.

This data set and analysis may be especially useful for people who are familiar with baseball metrics and want to dig deeper into what drives team performance. However, we will do our best to break things down so that anyone can follow along and understand the key takeaways. We were motivated to do this analysis because we are all passionate about the game of baseball and wanted to gain further insight into what plays into success amongst Major League Baseball teams using basis Baseball statistics.

Data

This project will use data from a 2019 Kaggle dataset[1] examining how specific baseball metrics contribute to a team's win-loss record during the 2012–2018 Major League Baseball seasons. The original dataset includes data from all 30 MLB teams across seven seasons, totaling 350 unique values. For this analysis, we will focus on the 2014–2018 seasons, reducing the dataset to 210 unique values. We will analyze all relevant columns that contribute to our research questions. Table 1 displays a description of the data.

Table 1 Data Dictionary

FIELD	TYPE	DESCRIPTION
TeamName	Text	Unique ID for each instance
RA/G	Numeric	Totals runs allowed by team, per season divided by 162
DefEff	Numeric	1 - ((H + ROE - HR) / (PA - BB - SO - HBP - HR))
E	Numeric	E is the number of Errors the team committed throughout the season

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

DP	Numeric	The number of doubles plays the turned during the season
W	Numeric	The Number of Wins the team had on the year
L	Numeric	The Number of Losses the team had on the yea
W-L%	Numeric	W / 162
ERA	Numeric	Team ERA during season
H	Numeric	Number of hits team allowed during season
ER	Numeric	Number of Earned runs allowed on season
HR	Numeric	Homeruns allowed by pitchers in season
BB	Numeric	Walks allowed by pitchers in season
SO	Numeric	Strikeouts by teams' pitchers in season
R/G	Numeric	Runs per game scored
R	Numeric	Runs scored in total by team on season
H1	Numeric	Hits by team on season
RBI	Numeric	Runs batted in by team per season
SB	Numeric	Stolen bases by team per season
SO1	Numeric	Number of times the team struck out on the year
BA	Numeric	Team batting average over season
OBP	Numeric	Team on base percentage over season
SLG	Numeric	Team slugging percentage over season
GDP	Numeric	Number of times team grounded into a double play on season
LOB	Numeric	Number of baserunners left on base by team per season

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

Salary	Numeric	Total dollar amount spent on players by team per season
WAR	Numeric	How much better the team was than a "average" team

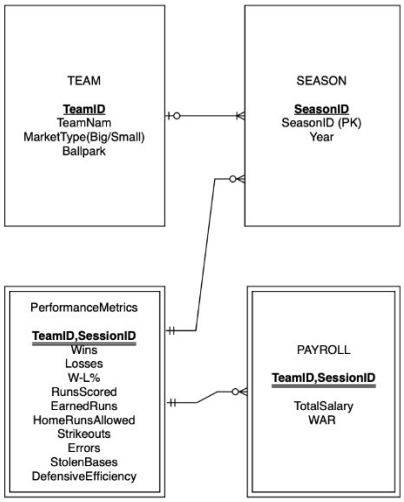
This data is essentially one entity (TeamName) because it is unique, and one Team and the Year is identified by a surrogate TeamName. All attributes are required because they contribute to the analysis. Essentially, all attributes are statistically measured (some derived) and are put in place based on gameplay in the past. We believe we can normalize the data with the transitive entities in the data.

Based on the tables created, we normalized the data and created a relational schema with 4 tables.

We normalized the data and created a relational schema based on the baseball performance data by normalizing it into four tables. The main table, **Team Performance**, stores overall team information like wins, losses, win percentage, WAR, and salary. To keep the data organized and avoid redundancy, we separated the stats into three child tables: **Batting**, **Fielding**, and **Pitching**. Each of these tables includes specific metrics for its category and uses Team Name as a foreign key to link back to the main table. This structure helps us manage the data more efficiently and keeps everything connected to the correct team.

ERD Diagram

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics



We ultimately altered things from this ERD, but we are including it to show our original thought process and foundation of our work. We

Graphical Relation Schema

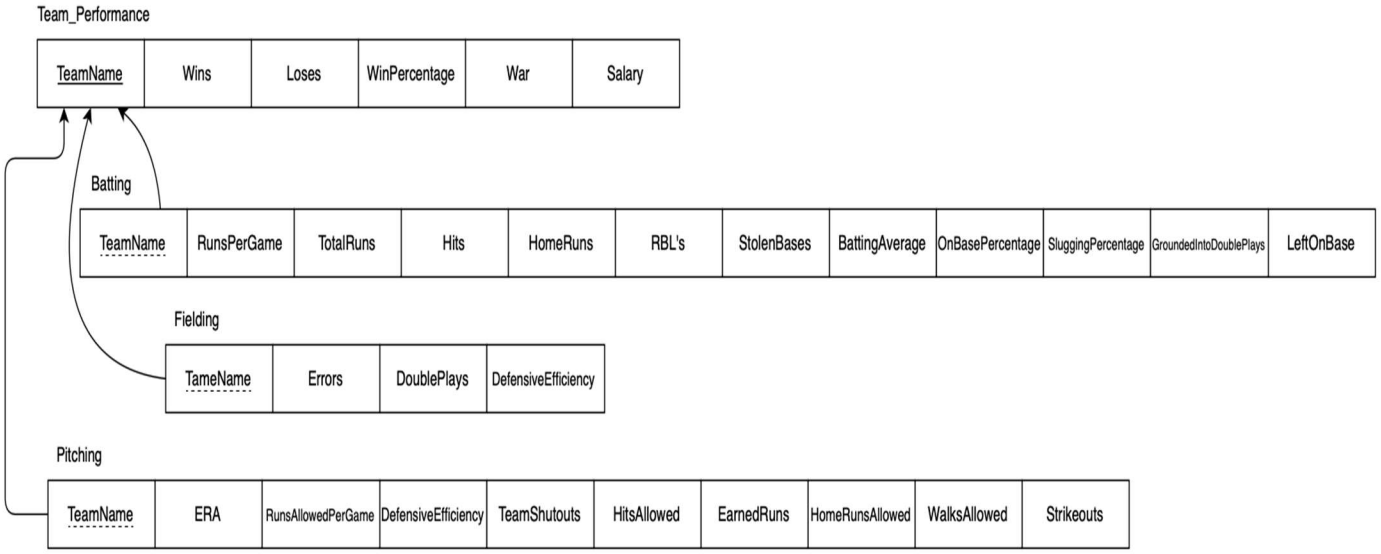


Figure 2 Graphical Relational Schema

Database Implementation

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

To implement the database in APEX, we wrote CREATE TABLE commands for each table in the relational schema. As the parent table, TEAM was created first.

TEAM

```
CREATE TABLE Team_Performance (
   TeamName VARCHAR(25) PRIMARY KEY,
   Wins INT,
   Losses INT,
   WinPercentage FLOAT,
   WAR FLOAT,
   Salary DECIMAL(12, 2)
);
```

BATTING

```
CREATE TABLE Batting_Stats (
   TeamName VARCHAR(25) PRIMARY KEY,
   RunsPerGame FLOAT,
   TotalRuns INT,
   Hits INT,
   HomeRuns INT,
   RBIs INT,
   StolenBases INT,
   Strikeouts INT,
   BattingAverage FLOAT,
   OnBasePercentage FLOAT,
   SluggingPercentage FLOAT,
   GroundedIntoDoublePlays INT,
   LeftOnBase INT
);
```

FIELDING

```
CREATE TABLE Fielding_Stats (
   TeamName VARCHAR(25) PRIMARY KEY,
   Errors INT,
   DoublePlays INT,
   DefensiveEfficiency FLOAT
);
```

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

PITCHING

```
CREATE TABLE Pitching_Stats (
   TeamName VARCHAR(25) PRIMARY KEY,
   ERA FLOAT,
   RunsAllowedPerGame FLOAT,
   DefensiveEfficiency FLOAT,
   TeamShutouts INT,
   HitsAllowed INT,
   EarnedRuns INT,
   HomeRunsAllowed INT,
   WalksAllowed INT,
   Strikeouts INT
);
```

The cleaned and normalized data was then imported into the database tables using the APEX Data Upload tool. The final dataset included 350 rows in all tables due to 350 represented teams. We manually went into the data upload feature and mapped out what columns were represented for our queries. You will find our analysis below which consists of 5 Business Questions with Baseball concentrated queries. The Queries range from Simple to advanced CASE queries and all attempt to answer a practical question. We also believe that our questions fit best with this group of data, over any other dataset we checked, and we finally manually input data into the CSV file along with cleaning it to even better help us prepare to take on this challenge.

INSERT COMMANDS

INSERT INTO TEAM (TeamName, W, L, W-L%, salary, WAR)

VALUES ('2018 ARI', 82, 80, 0.506, 143324597, 34.1);

INSERT INTO BATTING (TeamName, R/G, R, H 1, RBI, SB, SO 1, BA, OBP, SLG, GDP, LOB)

VALUES ('2018 ARI', 4.28, 693, 1283, 658, 79, 1460, 0.235, 0.31, 0.397, 110, 1086);

INSERT INTO FIELDING (TeamName, E, DP, DefEff)

VALUES ('2018 ARI', 75, 152, 0.698);

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

INSERT INTO PITCHING (TeamName, RA/G, ERA, H, ER, HR, BB, SO)

VALUES ('2018 ARI', 3.98, 3.72, 1313, 605, 174, 522, 1448);

Analysis

This analysis is for anyone interested in gaining a deeper understanding of advanced baseball metrics. It breaks down key stats beyond traditional numbers, offering clear insights for fans, students, or aspiring analysts looking to explore the data-driven side of the game.

Question 1: How did high-spending teams perform relative to their expected win percentage based on WAR?

Context: A team's General Manager or front office may want to evaluate whether their financial investment in talent is translating into actual wins. By comparing each team's actual win percentage to an expected value based on WAR, we can identify which teams are efficiently converting player performance into results—and which are falling short despite high payrolls. This deeper analysis provides a more accurate picture of team efficiency than WAR per dollar alone.

This query provides a detailed performance analysis of high-spending Major League Baseball teams by comparing their actual win percentage to an expected win percentage derived from their team WAR. Using the assumption that one WAR equals roughly one additional win over a replacement-level team, we calculate expected wins as `52 + WAR`, then convert that to a win percentage over a 162-game season. By subtracting this expected win percentage from the actual win percentage, we quantify whether teams are overperforming or underperforming. Teams are then categorized as "Overperforming Significantly," "Underperforming Significantly," or "Performing as Expected," based on whether their actual performance deviates by more than 5%. This approach allows analysts and team executives to assess the return on investment of their high payrolls in a more precise, data-driven way.

```
SELECT
  TeamName,
  Salary,
  WAR,
  WinPercentage,
```

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

```
  ROUND((52 + WAR) / 162, 3) AS Expected_WinPct,
  ROUND(WinPercentage - ((52 + WAR) / 162), 3) AS WinPct_Difference,
  CASE
    WHEN ((52 + WAR) / 162) - WinPercentage > 0.05 THEN 'Underperforming Significantly'
    WHEN WinPercentage - ((52 + WAR) / 162) > 0.05 THEN 'Overperforming Significantly'
    ELSE 'Performing as Expected'
  END AS Performance_Accuracy
FROM
  Team_Performance
WHERE
  Salary > 190000000
  AND WAR IS NOT NULL
  AND WinPercentage IS NOT NULL
ORDER BY
  WinPct_Difference ASC;
```

The results of this query revealed that several high-spending teams with strong WAR totals still underperformed relative to their expected win percentage. In particular, a few teams missed their projected win rate by more than 5%, signaling inefficiency despite substantial roster value. On the other hand, a handful of teams exceeded expectations, outperforming their WAR-based predictions and suggesting stronger team chemistry, managerial influence, or clutch performance. Most teams, however, fell within the "Performing as Expected" range, reinforcing that WAR is generally a reliable predictor when tied to financial investment.

TEAMNAME	SALARY	WAR	WINPERCENTAGE	EXPECTED_WINPCT	WINPCT_DIFFERENCE	PERFORMANCE_ACCURACY
2017 NYY	209690952	53.1	.562	.649	-.087	Underperforming Significantly
2016 BOS	207758836	53	.574	.648	-.074	Underperforming Significantly
2018 LAD	199582045	49.9	.564	.629	-.065	Underperforming Significantly
2012 NYY	228495456	51	.586	.636	-.05	Performing as Expected
2015 NYY	222528373	42.1	.537	.581	-.044	Performing as Expected
2013 LAD	239894375	46	.568	.605	-.037	Performing as Expected
2014 LAD	246367142	47.1	.58	.612	-.032	Performing as Expected
2016 DET	200179285	38.6	.534	.559	-.025	Performing as Expected
2015 LAD	301735080	43.7	.568	.591	-.023	Performing as Expected
2018 CHC	194259933	45	.583	.599	-.016	Performing as Expected

Figure 3 Actual vs expected performance

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

***Question 2:** How does pitching performance correlate with team success?*

Context: Think maybe a pitching coach or analyst wants to find what pitching stats are most relevant and associated with win percentage.

We joined team performance data with pitching statistics to examine whether strong pitching correlates with higher win percentages. This query lists each team's ERA and total strikeouts alongside their win percentage, allowing us to observe patterns between pitching effectiveness and team success. Teams are ordered from highest to lowest win percentage to spotlight top-performing clubs.

```
SELECT
  tp.TeamName,
  ps.ERA,
  ps.Strikeouts,
  tp.WinPercentage
FROM
  Team_Performance tp
JOIN
  Pitching_Stats ps ON tp.TeamName = ps.TeamName
ORDER BY
  tp.WinPercentage DESC;
```

We identified the top 10 team seasons based on winning percentage and examined their corresponding pitching statistics to look for patterns in elite performance.
Interestingly, most of these high-performing teams shared a common trait: an **ERA in the 3.0 to 4.0 range**, indicating consistent, above-average pitching. One notable outlier was the **2017 Houston Astros**, who maintained a strong win record despite a slightly higher ERA.
Another surprising finding was the **2016 Chicago Cubs**, who dominated the league while recording relatively fewer strikeouts—a reminder that there are multiple paths to success.
Overall, this analysis reveals that while pitching performance is a key contributor to winning, elite teams may excel in different ways, adding depth to how we evaluate success in baseball. The most obvious insight to these queries is the fact that in the later years of the study, strikeout numbers were significantly higher. This is not a coincidence, as the game is seeing much higher strikeout numbers and it is reflected in this study.

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

TEAMNAME	ERA	STRIKEOUTS	WINPERCENTAGE
2018 BOS	3.75	1558	.667
2017 LAD	3.38	1549	.642
2016 CHC	3.15	1441	.64
2018 HOU	3.11	1687	.636
2017 CLE	3.3	1614	.63
2017 HOU	4.12	1593	.623
2015 STL	2.94	1329	.617
2018 NYY	3.78	1634	.617
2012 WSN	3.33	1325	.605
2014 LAA	3.58	1342	.605
More than 10 rows available. Increase rows selector to view more rows.			

Figure 4 Pitching metrics and win percentage

<u>Question 3:</u> Are teams with better fielding metrics (low number of errors, high defensive efficiency) winning more games?

<u>Context:</u> A manager wants to know how much team defense is affecting performance

We categorized MLB teams into two tiers—**High Defense** and **Low Defense**—based on their defensive efficiency.
Using a CASE statement, we grouped teams by this classification and calculated the average win percentage for each group.
This allowed us to explore whether teams with stronger fielding performance tend to win more games

```
SELECT
  CASE
    WHEN fs.DefensiveEfficiency >= 0.700 THEN 'High Defense'
    ELSE 'Low Defense'
  END AS Defense_Tier,
  ROUND(AVG(tp.WinPercentage), 3) AS Avg_WinPercentage
FROM
  Team_Performance tp
JOIN
  Fielding_Stats fs ON tp.TeamName = fs.TeamName
```

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

```
GROUP BY
  CASE
    WHEN fs.DefensiveEfficiency >= 0.700 THEN 'High Defense'
    ELSE 'Low Defense'
  END;
```

DEFENSE_TIER	AVG_WINPERCENTAGE
Low Defense	.487
High Defense	.547

2 rows returned in 0.02 seconds Download

Figure 5 Defense and win percentage

Using a CASE statement for this analysis allowed us to clearly separate teams into defensive performance tiers. By applying a key metric like Defensive Efficiency and comparing it against win percentage, we were able to highlight a strong correlation between defensive ability and overall team success.

If a coach or analyst were reviewing this data, they would easily see that top-performing teams tend to excel defensively, while teams with lower win percentages often struggle in the field. This reinforces the importance of defense in building winning ball clubs.

***Question 4:* What teams were the most well rounded in all of Baseball from 2012 – 2018?**

Context: In modern baseball, success is no longer judged solely by win-loss records. Front offices, analysts, and coaching staff use advanced metrics like WAR (Wins Above Replacement) and strikeouts to evaluate team strength and sustainability over time. WAR provides a holistic view of a team's overall contribution to winning, while strikeouts are a key pitching metric reflecting dominance and control on the mound.

By using **INTERSECT**, the query returns only those teams that appear in the **top 10 of all three categories**, allowing us to isolate the most complete and well-rounded teams. These are teams that not only win games but also demonstrate strong underlying performance in both overall value and pitching power.

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

This analysis could be particularly valuable to front offices and coaching staff looking to model team-building strategies after consistently successful teams.

```
SELECT
  tp.TeamName,
  tp.WAR,
  tp.WinPercentage,
  ps.Strikeouts
FROM
  Team_Performance tp
JOIN
  Pitching_Stats ps ON tp.TeamName = ps.TeamName
WHERE
  tp.TeamName IN (
    -- Teams in Top 10 WAR
    SELECT TeamName FROM (
      SELECT TeamName FROM Team_Performance
      ORDER BY WAR DESC
    ) WHERE ROWNUM <= 10
    INTERSECT
    -- Teams in Top 10 Win Percentage
    SELECT TeamName FROM (
      SELECT TeamName FROM Team_Performance
      ORDER BY WinPercentage DESC
    ) WHERE ROWNUM <= 10
    INTERSECT
    -- Teams in Top 10 Strikeouts
    SELECT TeamName FROM (
      SELECT TeamName FROM Pitching_Stats
      ORDER BY Strikeouts DESC
    ) WHERE ROWNUM <= 10
  )
ORDER BY
  tp.WinPercentage DESC;
```

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

Our Group wanted to create an analysis on the most well-rounded teams in all of Baseball during this timeframe. We used metrics such as WAR, which is more offensive invested, Strikeouts, and Win Percentage to evaluate the best teams over this time. It is important to note that all these team are American League teams, and all had deep playoff runs, with the 2018 Red Sox and 2017 Astros winning the World Series. What can be concluded is that a team must be well rounded on offense and defense for success.

TEAMNAME	WAR	WINPERCENTAGE	STRIKEOUTS
2018 BOS	56.5	.667	1558
2018 HOU	57.1	.636	1687
2017 CLE	61.2	.63	1614
2017 HOU	52.9	.623	1593
2018 NYY	51.3	.617	1634

5 rows returned in 0.03 seconds Download

Figure 6 WAR, strikeouts, and win percentage

Question 5: Do teams with higher total salaries tend to win more games?

Context: A fan that is curious if spending is associated with wins, and if the high dollar teams are guaranteed success. This question is interesting because if the game is paid to win, shouldn't every team spend as much as possible? The season is 162 games, so anything around a .500-win percentage represents an 81-81 season.

This question provides deep insight and ability for analysis and is a basic simple query.

```
SELECT
  tp.TeamName,
  tp.Salary,
  tp.WinPercentage
FROM
  Team_Performance tp
```

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

```
ORDER BY
   tp.Salary DESC;
```

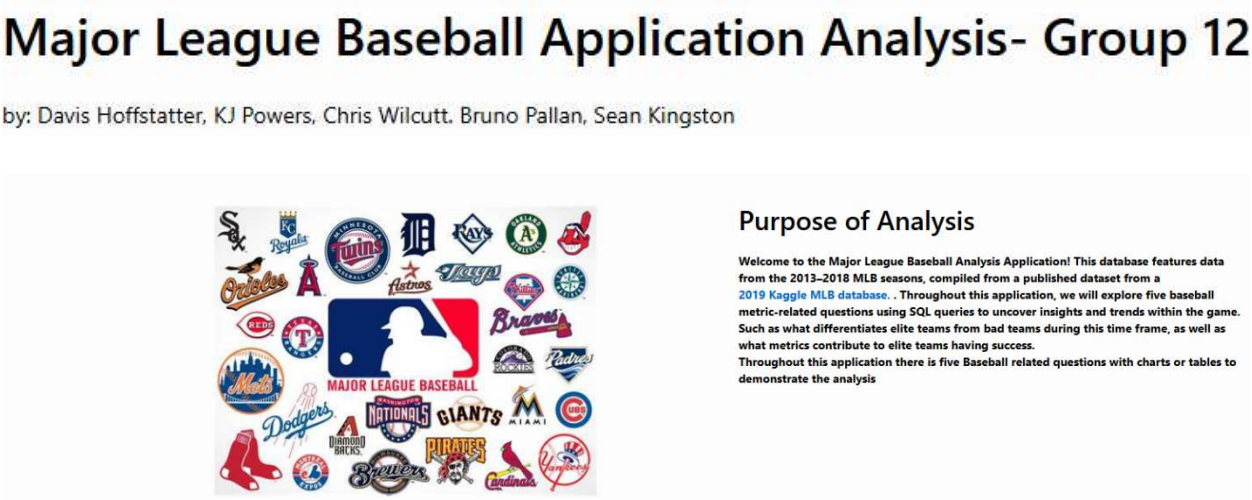
TEAMNAME	SALARY	WINPERCENTAGE
2015 LAD	301735080	.568
2016 LAD	268745494	.562
2017 LAD	259119625	.642
2014 LAD	246367142	.58
2013 LAD	239894375	.568
2013 NYY	239288940	.525
2012 NYY	228495456	.586
2018 BOS	227398860	.667
2015 NYY	222528373	.537
2016 NYY	220743376	.519

Figure 7 Sarlary and win percentage

As you can see from this analysis, all teams listed have records above .500. Being above .500 is going to give you a shot at the postseason, so we can conclude the teams that spend the most money probably will make it to the postseason although there is not a direct correlation to this.

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

Web Design



The web application's home page features a concise overview of the project, including a hyperlink to the original data source. It also includes a main navigation menu with nested items and custom icons for easy exploration. An image of the thirty major league baseball team's logos is presented for visual assistance to the reader to better understand. We used a bright blue accent color on our bread crumb bar to help tie into the colors of the logos and make it pleasing to view. Each navigation icon is thoughtfully designed to represent the type of visualization on its respective page, and a dropdown menu conveniently organizes all table views under a single heading. Also included is a short description regarding our purpose of analysis and what the viewer will find within.

Tables

The table below we joined team performance data with pitching statistics to examine whether strong pitching correlates with higher win percentages. This query lists each team's ERA and total strikeouts alongside their win percentage, allowing us to observe patterns between pitching effectiveness and team success. Teams are ordered from highest to lowest win percentage to spotlight top-performing clubs.

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

Teamname	Era	Strikeouts	Winpercentage
2018 BOS	3.75	1558	.667
2017 LAD	3.38	1549	.642
2016 CHC	3.15	1441	.64
2018 HOU	3.11	1687	.636
2017 CLE	3.3	1614	.63
2017 HOU	4.12	1593	.623
2018 NYY	3.78	1634	.617
2015 STL	2.94	1329	.617
2012 WSN	3.33	1325	.605
2015 PIT	3.21	1338	.605
2014 LAA	3.58	1342	.605
2012 CIN	3.34	1248	.599
2013 STL	3.42	1254	.599
2017 WSN	3.88	1457	.599
2015 CHC	3.36	1431	.599
2018 OAK	3.81	1237	.599
2013 BOS	3.79	1294	.599
2014 WSN	3.03	1288	.593
2013 OAK	3.56	1183	.593
2014 BAL	3.43	1174	.593
2013 ATL	3.18	1232	.593
2018 MIL	3.73	1428	.588
2016 WSN	3.51	1476	.586
2016 TEX	4.37	1154	.586
2012 NYY	3.84	1318	.586
2015 KCR	3.73	1160	.586
2016 CLE	3.84	1398	.584
2018 CHC	3.65	1333	.583
2013 PIT	3.26	1261	.58
2012 SFG	3.68	1237	.58
2012 OAK	3.48	1136	.58
2012 ATL	3.42	1232	.58
2014 LAD	3.4	1373	.58
2012 BAL	3.9	1177	.574
2012 TEX	3.99	1286	.574
2013 DET	3.61	1428	.574
2017 BOS	3.7	1580	.574
2016 BOS	4	1362	.574
2017 ARI	3.66	1482	.574
2015 TOR	3.8	1117	.574
2013 LAD	3.25	1292	.568
2013 CLE	3.82	1379	.568
2015 LAD	3.44	1396	.568
2017 CHC	3.95	1439	.568
2013 TBR	3.74	1310	.564
2018 LAD	3.38	1565	.564
2018 CLE	3.77	1544	.562
2017 NYY	3.72	1560	.562
2016 LAD	3.7	1510	.562
2018 COL	4.33	1409	.558

Teamname	War	Winpercentage	Strikeouts
2018 BOS	56.5	.667	1558
2018 HOU	57.1	.636	1687
2017 CLE	61.2	.63	1614
2017 HOU	52.9	.623	1593
2018 NYY	51.3	.617	1634

This table below shows the relationship between salary amount compared to win percentage.

[1]

https://www.kaggle.com/datasets/omipelcastre/

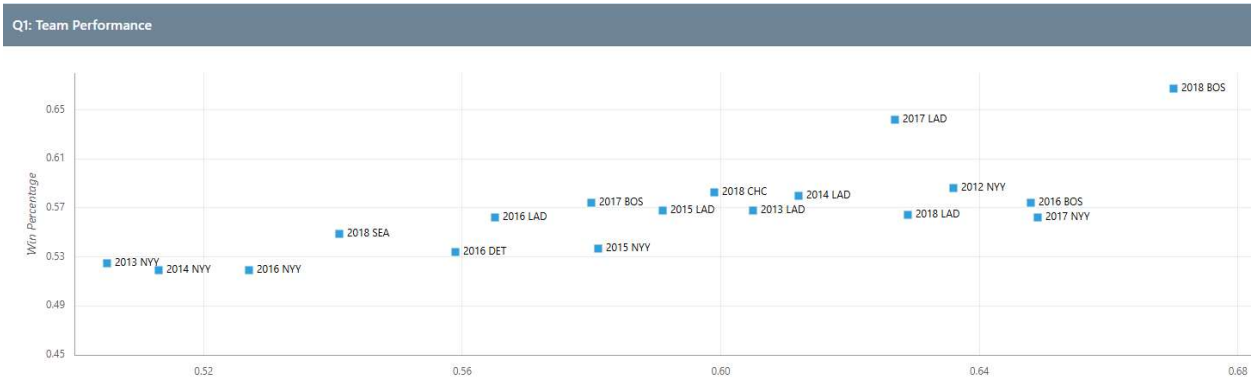
mlb-team-statistics

Inverse Relationship Between Salary Rank and Win Percentage Rank is Imperfect: This highlights that simply spending more money on player salaries does not guarantee more wins. But you can see instances where teams with significantly higher salaries higher win percentages as the left hand of the graph shows this.
Identification of Potentially Efficient and Inefficient Spending: Teams with relatively low spending but high winning percentages could be considered to be getting a good return on investment.
The Range of Salaries and Win Percentages: The range of well over $200 million to under $50 million. This highlights the complexities of building a winning baseball team and the different pathways to success.

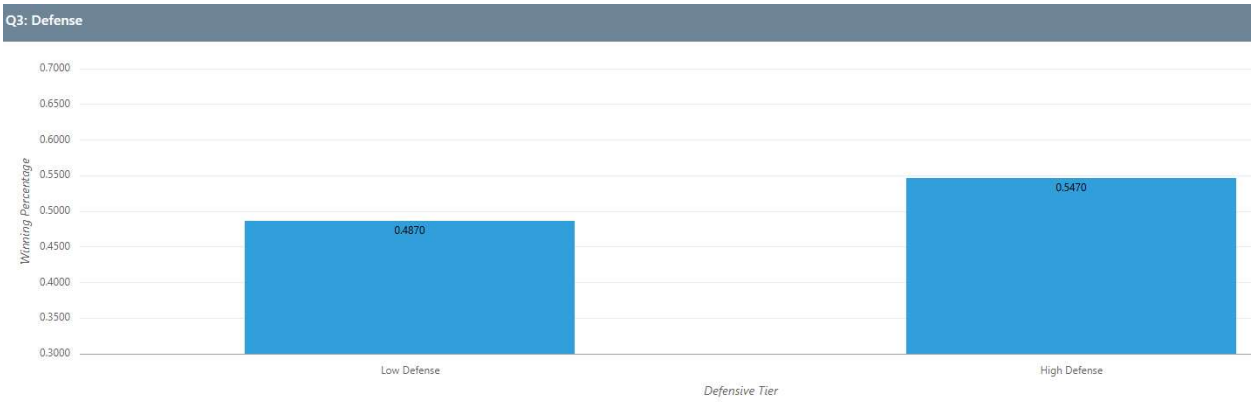
Teamname	Salary	Winpercentage
2015 LAD	301735080	.568
2016 LAD	268745494	.562
2017 LAD	259119625	.642
2014 LAD	246367142	.58
2013 LAD	239894375	.568
2013 NYY	239288940	.525
2012 NYY	228495456	.586
2018 BOS	227398860	.667
2015 NYY	222528373	.537
2016 NYY	220743376	.519
2014 NYY	214898031	.519
2017 BOS	209872508	.574
2017 NYY	209690952	.562
2016 BOS	207758836	.574
2018 SEA	205665348	.549
2016 DET	200179285	.534
2018 LAD	199582045	.564
2018 CHC	194259933	.583
2017 SEA	189806809	.481
2017 DET	188498884	.395
2017 TOR	187067487	.469
2015 SEA	186413630	.469
2016 CHC	184352494	.64
2015 BOS	183481478	.481
2017 WSN	182238854	.599
2018 WSN	181382609	.506
2012 PHI	179900523	.5
2018 NYY	179598151	.617
2016 SEA	178326844	.531
2017 BAL	176349127	.463
2012 BOS	175279051	.426
2014 PHI	175061605	.451
2016 LAA	174363407	.457
2018 LAA	173784989	.494
2017 CHC	172092536	.568
2014 DET	171996220	.556
2014 SEA	171336782	.537
2013 BOS	170734264	.599
2014 BOS	170095758	.438
2017 LAA	166161209	.494
2015 WSN	165655095	.512
2014 LAA	165403251	.605
2018 STL	163784311	.543
2013 PHI	163536898	.451
2018 HOU	163524216	.636
2017 TEX	163350840	.481
2016 TEX	163021690	.586
2015 DET	162160921	.46
2017 SFG	161397330	.395
2018 SFG	160993827	.451

[1] https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics

Interactive reports were developed for each database table, enabling users to search, filter, and group the data with ease To improve clarity, column headers and number formats were customized for readability. Each report page also includes a descriptive text box that provides context about the data presented in the table.

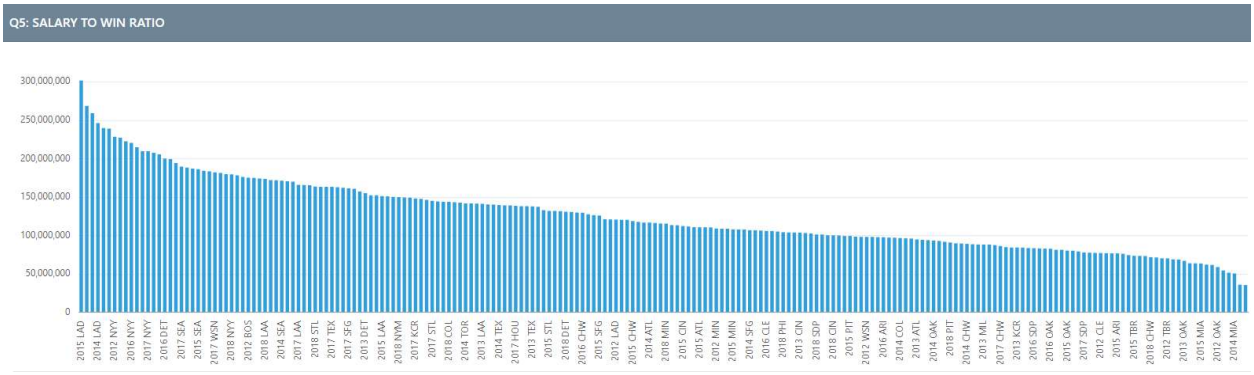


This scatter plot displays each team's expected win percentage (based on WAR) on the x-axis and their actual win percentage on the y-axis. The purpose of the graph is to compare how closely teams performed relative to expectations. Teams that appear above the diagonal line performed better than their WAR would suggest, while those below the line underperformed. Points near the line indicate teams that performed as expected. This visual provides a clear way to see which high-spending teams exceeded, met, or fell short of their projected success based on overall player contributions.



This bar chart compares the average win percentage of teams with high defensive efficiency (≥ 0.700) versus those with lower efficiency. The chart clearly shows that teams classified as "High Defense" tend to have a higher average win percentage than "Low Defense" teams. This visual reinforces the idea that strong defensive performance contributes meaningfully to overall team success, providing a simple yet powerful insight for coaches, analysts, and front offices focused on improving win outcomes through defense. We chose to keep this chart very simple and easy to understand for readers.

1 https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics



This graph shows the relationship between salary amount compared to win percentage.

Inverse Relationship Between Salary Rank and Win Percentage Rank is Imperfect: This highlights that simply spending more money on player salaries does not guarantee more wins. But you can see instances where teams with significantly higher salaries higher win percentages as the left hand of the graph shows this.

Identification of Potentially Efficient and Inefficient Spending: Teams with relatively low spending but high winning percentages could be considered to be getting a good return on investment.

The Range of Salaries and Win Percentages: The range of well over $200 million to under $50 million. This highlights the complexities of building a winning baseball team and the different pathways to success.

1 https://www.kaggle.com/datasets/omipelcastre/mlb-team-statistics