Manager Statistics

MLB 2024 Season

Data Analytics Term Project Proposal

Dani DiTomasso

Interest in MLB

This summer, I had the opportunity to participate in a full time internship at MLB headquarters as a IT Project Manager.

While there, I networked with people within many different departments, including those working in data analytics for the Baseball Operations department.

I had a very positive experience, and found a passion for working in the sports industry, which is my goal post-graduation.

Hypothesis

Managers are a huge determinant on how a major league team will perform for any given season. They make imperative decisions including player subbing, types of plays, overall game strategy, and so much more.

Therefore, it is my hypothesis that the **more aggressive offensive plays a manager chooses to take, the more success the team will have overall.**

Success in this context is defined by having more than or equal to a 52 win-to-loss percentage within the 2024 season (excluding postseason statistics)

The Data

Dataset 1

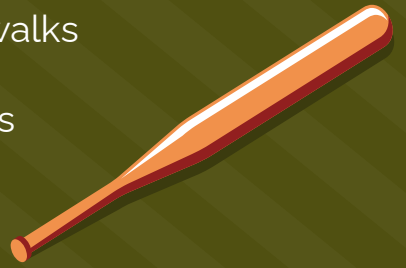
2024 Manager Records

- Rank, manager name, team
- Wins, Losses, Wins-Loss percentage, ties
- Challenges, overturns, overturn percentages
- Ejections

Dataset 2

2024 Managerial Tendencies

- Stealing 2nd base
- Stealing 3rd base
- Sac bunts
- Intentional walks
- Substitutions



Merging the Data

Dataset 1 - 2024 Manager Records

Rk	Mgr	Tm	W	L	W-L%	Ties	G	Finish	Wpost	Lpost	W-L%post	Challenges	Overturned	Overturn%	Ejections
1	Rocco Baldelli	MIN	82	80	0.506	0	162	4				32	17	53.10%	0
2	David Bell	CIN	76	81	0.484	0	157	4				40	18	45.00%	5
3	Freddie Benavides	CIN	1	4	0.2	0	5	4				1	1	100.00%	0
4	Bud Black	COL	61	101	0.377	0	162	5				32	20	62.50%	3
5	Bruce Bochy	TEX	78	84	0.481	0	162	3				35	19	54.30%	5
6	Aaron Boone	NYY	94	68	0.58	0	162	1				31	22	71.00%	6
7	Kevin Cash	TBR	80	82	0.494	0	162								
8	Alex Cora	BOS	81	81	0.5	0	162								
9	Craig Counsell	CHC	83	79	0.512	0	162								
10	Joe Espada	HOU	88	73	0.547	0	161								
11	Pedro Grifol	CHW	28	89	0.239	0	117								
12	A.J. Hinch	DET	86	76	0.531	0	162								
13	Brandon Hyde	BAL	91	71	0.562	0	162								
14	Mark Kotsay	OAK	69	93	0.426	0	162								
15	Torey Lovullo	ARI	89	73	0.549	0	162								
16	Oliver Marmol	STL	83	79	0.512	0	162								
17	Dave Martinez	WSN	71	91	0.438	0	162								
18	Bob Melvin	SFG	80	82	0.494	0	162								
19	Carlos Mendoza	NYM	89	73	0.549	0	162								

Each dataset is sorted by manager rank, allowing for the merging of data by the 'Rk' field, or the 'Mgr' name.

Dataset 2 - 2024 Managerial Tendencies

					Stealing 2nd				Stealing 3rd				Sac Bunts				Intentional Walks				Substitutions					
Rk	Mgr	Tm	Age	G	Ch	Att	Rate	Rate+	Ch	Att	Rate	Rate+	Ch	Att	Rate	Rate+	PA	IBB	Rate	Rate+	PH/G	PH/G+	PR/G	PR/G+	P/G	P/G+
1	Rocco Baldelli	MIN	42	162	1386	79	5.70%	68	1028	5	0.50%	26	1456	10	0.70%	79	6078	14	0.20%	76	1.18	138	0.2	103	4.3	101
2	David Bell	CIN	51	157	1315	186	14.10%	152	917	47	5.10%	245	1305	16	1.20%	103	5893	13	0.20%	86	0.81	104	0.2	115	4.3	101
3	Freddie Benavides	CIN	58	5	34	3	8.80%	95	24	0	0.00%	0	37	0	0.00%	0	171	1	0.60%	227	0.2	26	0.4	227	4.2	99
4	Bud Black	COL	67	162	1350	102	7.60%	81	963	8	0.80%	40	1381	19	1.40%	115	6322	12	0.20%	74	0.44	56	0.15	87	4.1	95
5	Bruce Bochy	TEX	69	162	1420	112	7.90%	94	892	7	0.80%	43	1343	6	0.40%	51	6108	26	0.40%	141	0.62	73	0.12	59	4	93
6	Aaron Boone	NYY	51	162	1521	78	5.10%	61	934	25	2.70%	145	1480	14	0.90%	108	6131	8	0.10%	43	0.37	43	0.24	122	4.2	99
7	Kevin Cash	TBR	46	162	1421	188	13.20%	158	996	32	3.20%	174	1442	10	0.70%	79	6022	15	0.20%	83	0.84	98	0.23	115	4.3	102
8	Alex Cora	BOS	48	162	1475	153	10.40%	124	1019	27	2.60%	144	1485	7	0.50%	54	6154	26	0.40%	140	1.27	149	0.12	59	4.3	101
9	Craig Counsell	CHC	53	162	1407	136	9.70%	104	975	29	3.00%	142	1398	16	1.10%	96	6032	7	0.10%	45	0.86	110	0.09	49	4.1	95
10	Joe Espada	HOU	48	161	1483	98	6.60%	79	965	13	1.30%	73	1487	12	0.80%	92	6018	2	0.00%	11	0.7	82	0.2	100	4.2	98
11	Pedro Grifol	CHW	54	117	933	83	8.90%	106	642	10	1.60%	85	926	15	1.60%	185	4493	17	0.40%	125	0.88	103	0.24	121	4.2	99
12	A.J. Hinch	DET	50	162	1434	92	6.40%	77	877	8	0.90%	50	1332	4	0.30%	34	5969	11	0.20%	61	1.19	140	0.1	50	4.1	96
13	Brandon Hyde	BAL	50	162	1496	113	7.60%	90	907	10	1.10%	60	1410	6	0.40%	49	6071	9	0.10%	49	0.87	102	0.15	78	4.4	102
14	Mark Kotsay	OAK	48	162	1418	112	7.90%	95	902	8	0.90%	48	1364	19	1.40%	159	6161	34	0.60%	183	0.85	100	0.22	112	4.3	100
15	Torey Lovullo	ARI	58	162	1512	124	8.20%	88	1071	15	1.40%	67	1586	34	2.10%	179	6195	32	0.50%	201	1.03	133	0.17	98	4.3	101
16	Oliver Marmol	STL	37	162	1472	98	6.70%	72	917	14	1.50%	73	1384	16	1.20%	97	6044	16	0.30%	103	0.73	95	0.18	101	4.1	97
17	Dave Martinez	WSN	59	162	1378	235	17.10%	183	999	47	4.70%	225	1454	26	1.80%	150	6114	11	0.20%	70	0.57	73	0.17	98	4.5	106
18	Bob Melvin	SFG	62	162	1426	73	5.10%	55	910	13	1.40%	68	1441	10	0.70%	58	6095	11	0.20%	70	1.01	130	0.15	87	4.5	105

Data Source: https://www.baseball-reference.com/leagues/majors/2024-managers.shtml

Data is updated on a daily basis to reflect current stats in the 2024 season

Conducting the Data Analysis

- Is there a correlation between a managers' rank and their win-loss percentage? → Linear Model
- Do higher ranking managers take more risks during gameplay?
  - Increased attempts to steal bases, bunts, and intentional walks compared to the average. → K-Means Cluster
- Do number of managerial ejections correlate to team performance? → Histogram
- Do higher play challenges and overturn percentages equate to better game outcomes? → Linear Model

Subject to change upon new knowledge gained further in the semester

Data Application & Uncertainties

Model Application

Each model will be stored, either as a png or jpg image if it is a visual plot, or within a text document if it is written content that is printed in the R terminal.

These models will then be compared and contrasted against each other, and determine whether or not the hypothesis is either accepted or rejected.

Dataset Uncertainty

Given the continuance of the MLB Postseason games until the end of october, this could potentially impact and change the data as analysis is being conducted. The nature of the MLB season structure must be taken into consideration in order to cultivate an accurate data analysis.

Analysis Predictions

For this project, a “good result” is **not** the hypothesis being supported, nor it being rejected.

The goal is to analyze data that has statistically significant results, proving some type of correlation between managers and their teams’ performance, regardless whether that correlation is positive or negative,

The analysis should provide insight to the different type of calls that managers make, and how those decisions ultimately affects the performance of the team, if there is any relation to be found.

Thanks!

Do you have any questions?

Resources:

"MLB Team Win Trends - All Games." Www.teamrankings.com, www.teamrankings.com/mlb/trends/win_trends/.

Baseball Reference. "2024 Major League Baseball Managers | Baseball-Reference.com." Baseball-Reference.com, 2024, www.baseball-reference.com/leagues/majors/2024-managers.shtml. Accessed 10 Oct. 2024.