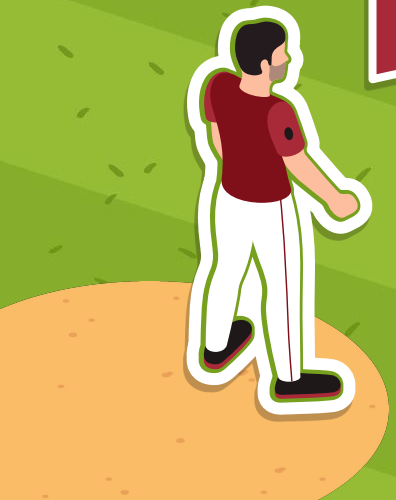
MAJOR LEAGUE BASEBALL 2024

TEAM MANAGER STATISTICAL ANALYSIS

Dani DiTomasso, Data Analytics 2024

INTRODUCTION

Background Information, The Data, and the Hypothesis.



BACKGROUND INFORMATION

Within the sport of baseball at the professional level, 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.

It is for this reason that the data focuses on and studies team statistics based on stragecial plays called by the manager.

THE DATA

DATASET 1

Managerial Records 2024

- Manager Name
- Team Abbreviation
- Wins
- Losses
- Wins-Loss Percentage
- Ties
- Games Played or Pitched
- Team Finish (average of all years weighted by the number games played or managed)
- Challenges
- Overturns & Overturn Percentage
- Ejections

DATASET 2

Managerial Tendencies 2024

- Stealing 2nd Base / 3rd Base, Sacrificial Bunts
  - Chances
  - Attempts
  - Rate
  - League Rate
- Intentional Walks
  - Plate Appearances
  - Intentional Walks
  - Rate
  - League Rate
- Substitutions
  - Pinch Hitters
  - Pinch Hitter Rate
  - Pinch Runner
  - Pinch Runner Rate
  - Pitchers
  - Picters Rate

HYPOTHESIS

Given this information, the hypothesis for this study is the **more aggressive offensive plays a manager chooses to take, the more success the team will have overall.**

Note that “team success” is measured by any team having a Win/Loss percentage that is **higher than 53%**.

WHY 53 PERCENT?

Although 2024 postseason statistics were not utilized within this study, they were still used to determine what teams were “successful” compared to “unsuccessful.” In 2024...

POSTSEASON TEAMS

Had above a 53 Win-Loss percentage
(12/12 Teams)

ALL TEAMS

Had above a 53 Win-Loss Percentage
(13/33 Teams)

EXPLORATORY ANALYSIS

Data Cleaning, Merging, and EDA's

DATA CLEANING & MERGING

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 listed by manager number, allowing for the merging of data by the 'Rk' field, or the 'Mgr' name.

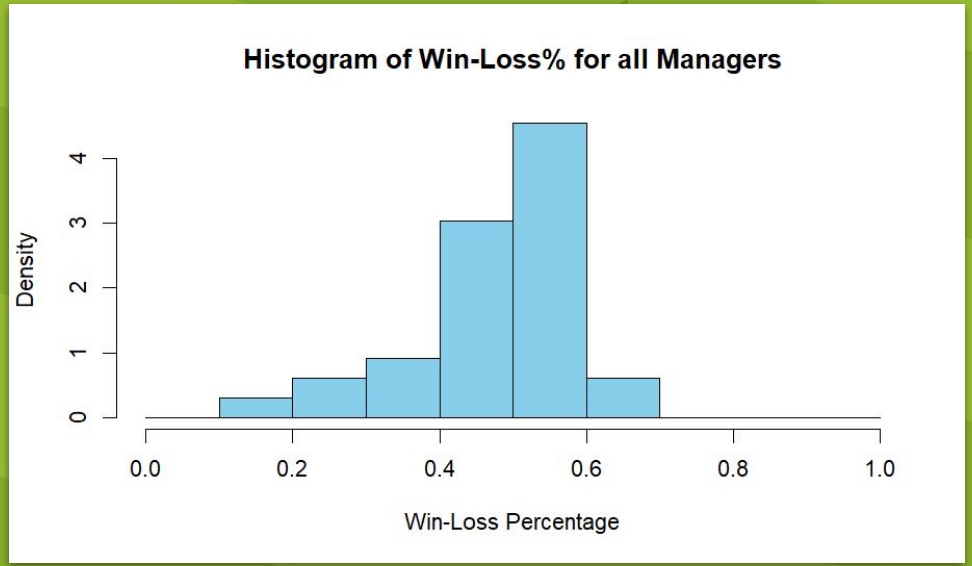
Dataset 2 - 2024 Managerial Tendencies

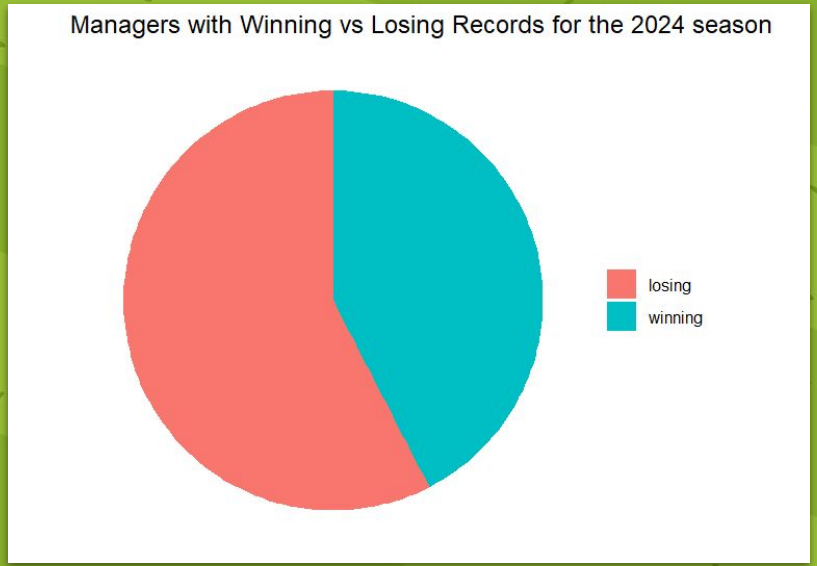
					Stealing 2nd				Stealing 3rd				Sac Bunts				Intentional Walks				Substitutions					
Rk	Mgr	Tm	Age	G	Ch	Att	Rate	Rate+	Ch	Att	Rate	Rate+	Ch	Att	Rate	Rate+	PA	IBB	Rate	Rate+	PH/G	PH/G+	PR/G	PR/G+	P/G	P/G+
1	Rocco Baldelli	MIN	42	162	1386	79	5.70%	68	1028	5	0.50%	26	1456	10	0.70%	79	6078	14	0.20%	76	1.18	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

*Repetitive columns and blank columns were removed when datasets merged

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

EXPLORATORY ANALYSIS

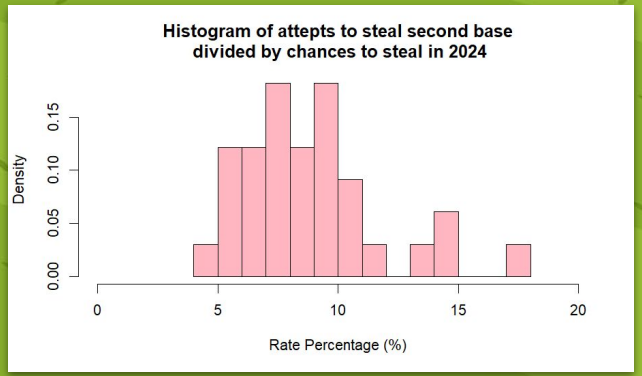


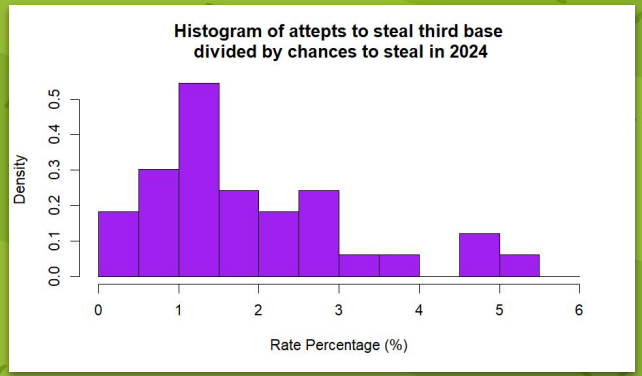


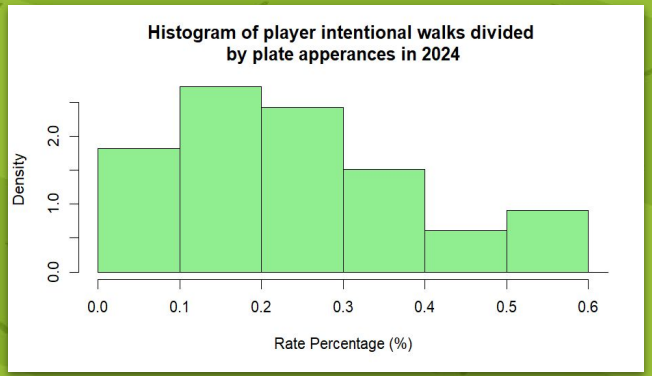
Most Teams ranged from a 40% - 60% win / loss record for the 2024 season

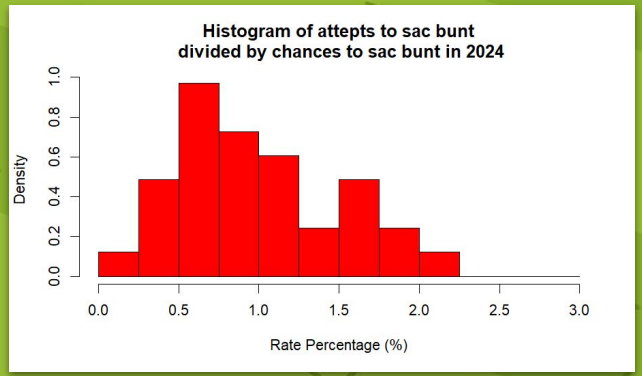
61% of teams had a "losing" status (W.L. < 0.53)

EXPLORATORY ANALYSIS





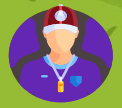




DATA MODELS

Linear Models, Decision Trees, Random Forest, and Naive Bayes

LINEAR MODELS



Manager Records – "Finish" (p-value = 1.42e-06)

Stealing Second – "Ch" (p-value = 0.0276)

Stealing Third – "Ch.1" (p-value = 1.23e-05)

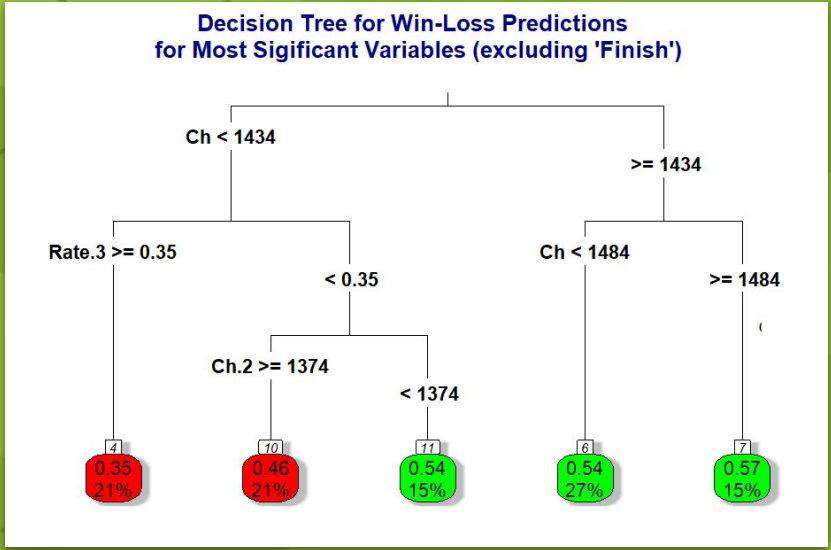
Sac Bunts – "Ch.2" (p-value = 9.66e-05)

Intentional Walks – "IBB" (p-value = 0.00188) & "Rate.3" (p-value = 0.00301)

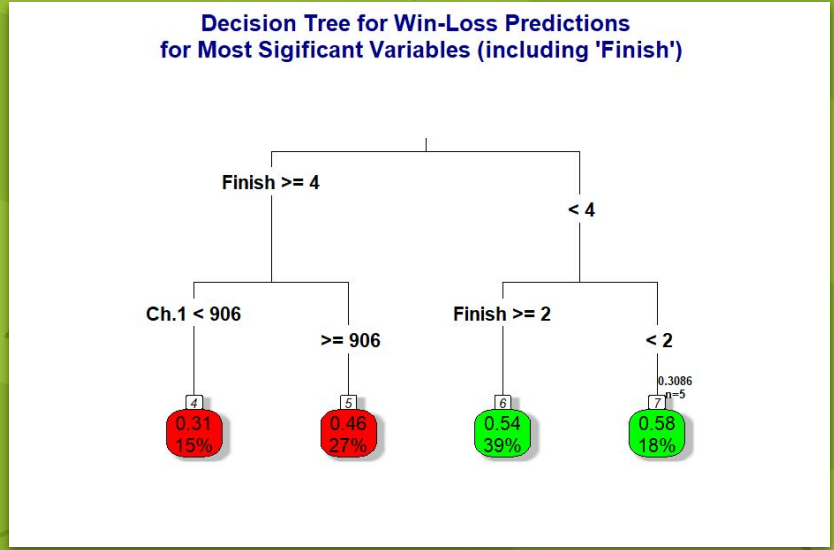
Substitutions – *None* "PH/G+" (p-value = 0.0811)

Each **category** of data was compared to the W.L. variable.

DECISION TREES

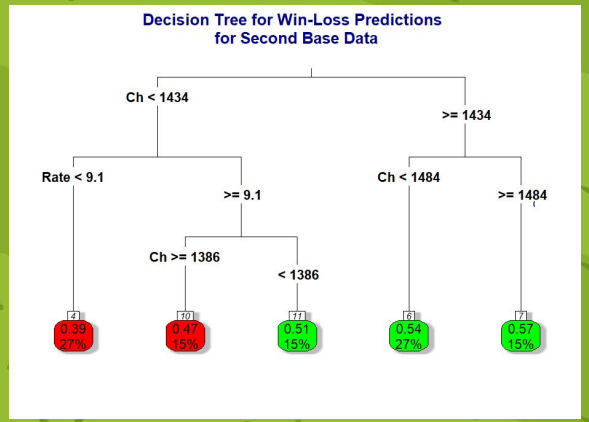


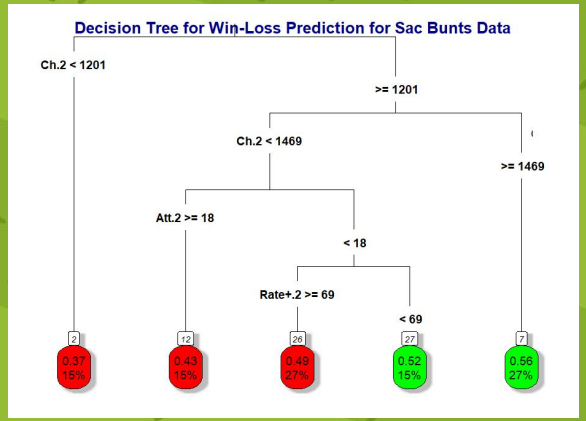
Mean Squared Error: 0.00222707845117845
R-squared: 0.771327433622756

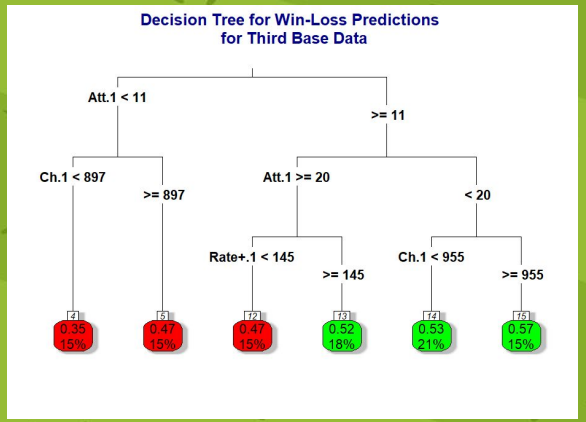


Mean Squared Error: 0.00360962116402116
R-squared: 0.629370337273226

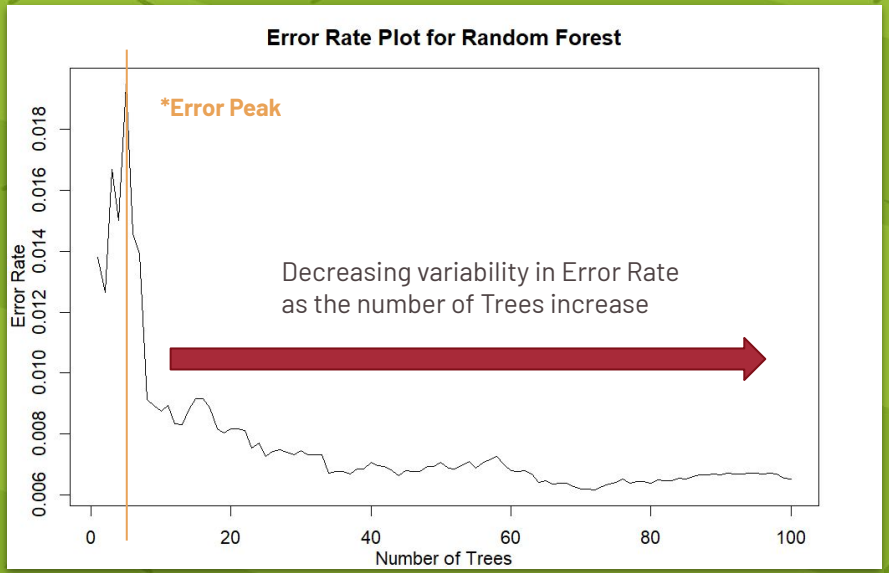
DECISION TREES



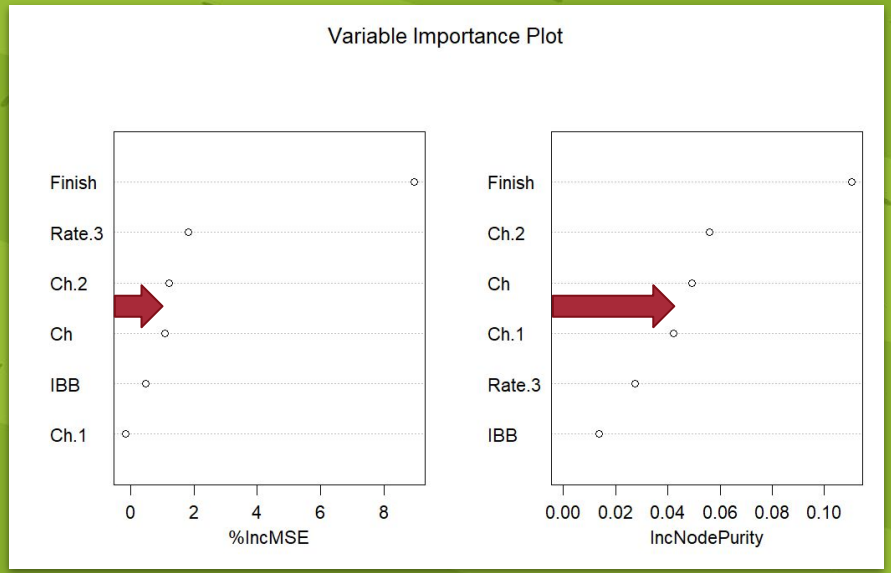




RANDOM FOREST



Error Rate is the smallest at around 70-100 decision trees.

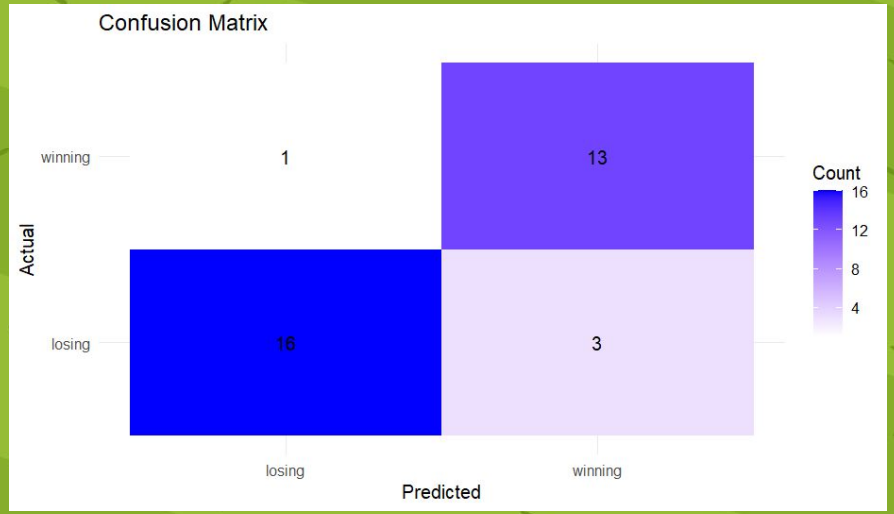


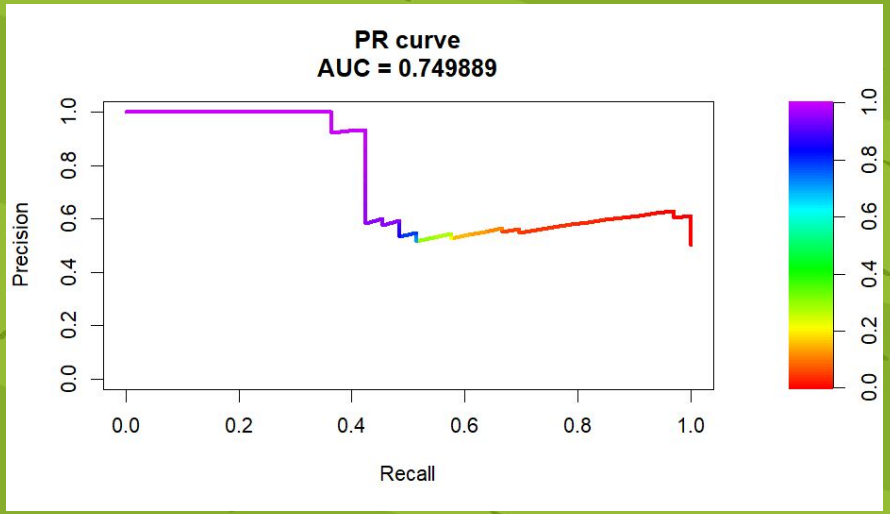
Greater significance in Node Purity for variables compared to % increase of MSE.

Prediction Model:

Mean Squared Error: 0.00126226367072559
R-squared: 0.870392947820565

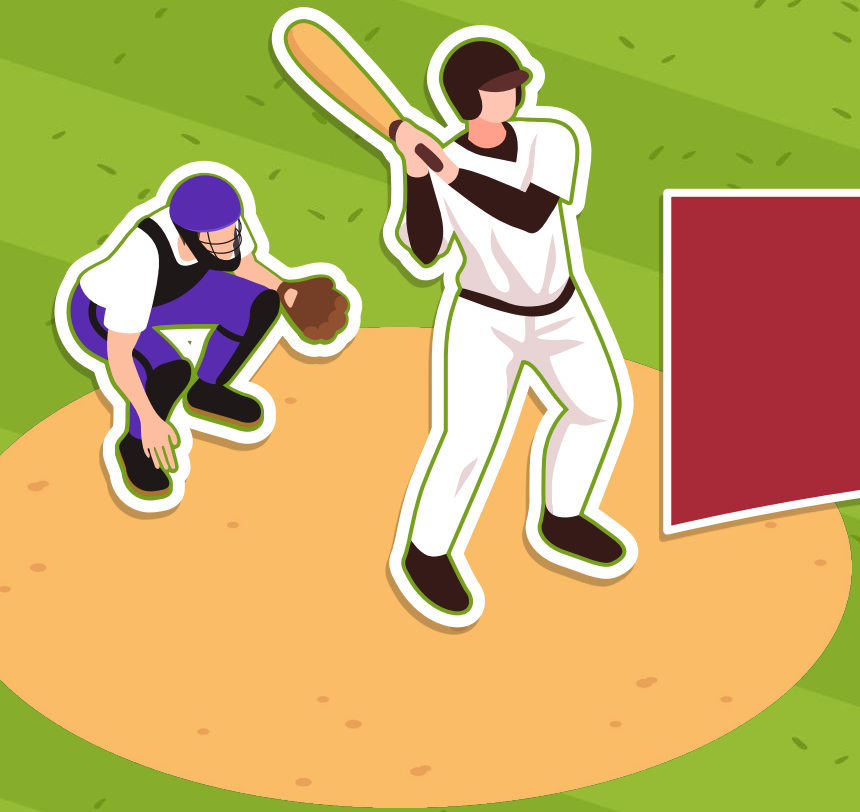
NAIVE BAYES





Confusion Matrix shows a highly accurate prediction model.

Precision Recall area under the curve shows about a 75% accuracy classifier performance.



CONCLUSION

Hypothesis Results & Reflections

HYPOTHESIS RESULTS

Failed to Reject the Null Hypothesis

- Although an offensive playing style can be beneficial for some plays, an overall more aggressive decision making style as a manager does not lead to more team wins.
- However, it is certain that the decisions managers make are crucial to the outcome of a game, as multiple variables have been proven to have statistical significance for a teams win-loss percentage.

KEY TAKEAWAYS

1. **Linear Models** are more effective and accurate when run with a **smaller** amount of variables compared to a large amount of variables.
2. **Decision Trees** were **crucial** in assessing the influence of aggressive plays on win-loss ratios.
3. **Data cleaning and merging** is critical for an effective analysis.

THANKS!

Do you have any questions?