MLB Teams and the Probability of Playoffs

Grace Coccagna and Ryan Thompson

Introduction

Objective

- Identify key statistical thresholds that differentiate playoff and non-playoff teams in the MLB.

- Use historical MLB data from Baseball Savant to determine which metrics are most influential in predicting playoff success

- Built a model that returns a percentage chance a team will make the playoffs based on number of wins at a certain point in the year, and certain statistics at that point in the season.

Why this matters:

- Baseball analytics is a crucial tool for Front Offices

- Understanding performance benchmarks can guide roster decisions and player evaluations

Methodology

Data Collection & Analysis

- Historical MLB data covering seasons 2018-2024 (excluding 2020)

- Team performance statistics (batting, pitching, fielding)

- Win-loss records by month and season

- Playoff qualification outcomes

Analytical Approach

- Identify a statistical correlation analysis between performance metrics and playoff qualification

- Built machine learning models to identify predictive patterns (Radom Forrest, Gradient Boosting, Logistic Regression)

- Testing and validation against historical outcomes

Three Initial Models

Logistic Regression, Gradient Boosting, Random Forest

- We built three models to help identify the key statistics and their thresholds for predicting playoff clinching seasons.

- The Random Forest model proved to be the most accurate, with a .9000 ROC-AUC Score

```
Model: Logistic Regression
              precision    recall  f1-score   support

           0       0.80      0.80      0.80         5
           1       0.67      0.67      0.67         3

    accuracy                           0.75         8
   macro avg       0.73      0.73      0.73         8
weighted avg       0.75      0.75      0.75         8

ROC-AUC Score: 0.8667
```

```
Model: Random Forest
              precision    recall  f1-score   support

           0       0.80      0.80      0.80         5
           1       0.67      0.67      0.67         3

    accuracy                           0.75         8
   macro avg       0.73      0.73      0.73         8
weighted avg       0.75      0.75      0.75         8

ROC-AUC Score: 0.9000
```

```
Model: Gradient Boosting
              precision    recall  f1-score   support

           0       0.71      1.00      0.83         5
           1       1.00      0.33      0.50         3

    accuracy                           0.75         8
   macro avg       0.86      0.67      0.67         8
weighted avg       0.82      0.75      0.71         8

ROC-AUC Score: 0.6667
```
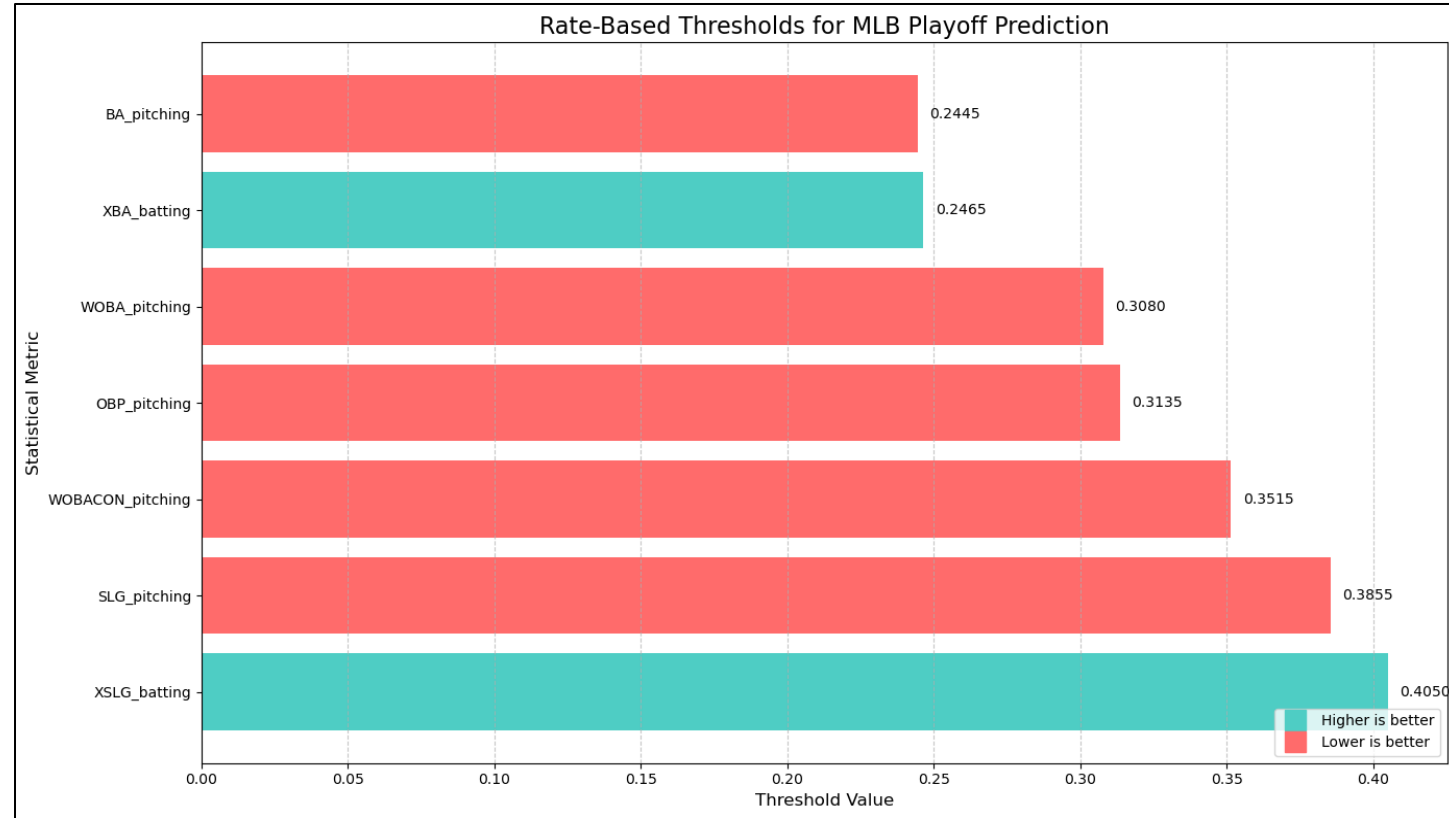
Key Statistical Thresholds

- From our Random Forest Model, we identified the feature importance for the KPI's

- Number of Hits, Slugging Percentage, Batting Average against pitchers, and Expected Batting Average for batters were the top five most influential values

Feature Importances (Random Forest)
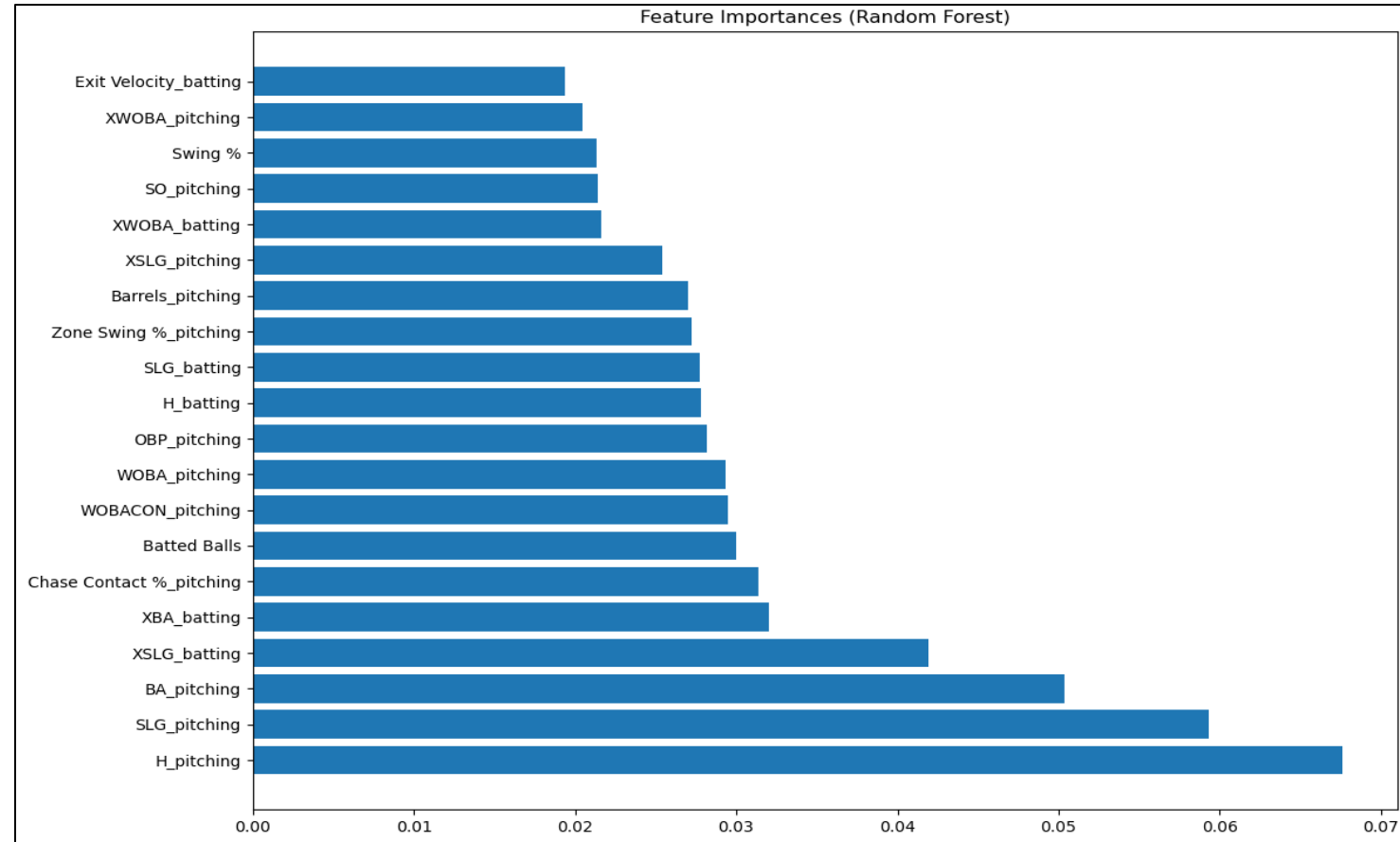
Key Statistics

Shown here is the most important statistics that predict playoff clinching seasons.

It also includes the thresholds marks of these statistics that teams should aim to achieve.

Rate-Based Thresholds for MLB Playoff Prediction

MLB Playoff Prediction: Key Statistical Thresholds

Statistical Metric	Threshold Value	Interpretation	Category
XSLG_batting	0.4050	Higher is better	XBatting
XBA_batting	0.2465	Higher is better	XBatting
Batted Balls	4135	Higher is better	General
H_pitching	1354	Higher is better	Pitching
Chase Contact %_pitching	54.9500	Higher is better	Pitching
SLG_pitching	0.3855	Lower is better	Pitching
WOBACON_pitching	0.3515	Lower is better	Pitching
OBP_pitching	0.3135	Lower is better	OPitching
WOBA_pitching	0.3080	Lower is better	Pitching
BA_pitching	0.2445	Lower is better	Pitching

Example Data Probability Output

```
Playoff Prediction: 56.0%

Additional Examples:
High-performing team: 58.0%
Low-performing team: 53.0%

Interactive Playoff Predictor
Enter your team's stats to get a playoff prediction!
Enter team wins (e.g., 85): 90
Enter team ERA (e.g., 3.75): 4.00
Enter team OPS (e.g., 0.750): .900
Enter team home runs (e.g., 200): 300

Playoff Prediction: 57.0%
This team has a good chance of making the playoffs.
```

- This graphic represents our Random Forest prediction model, ran on the 2023 season

- Our model correctly predicted all but one team to make the playoffs, and only incorrectly predicted one team.

- Based on our projections, the Cubs should've made the playoffs, and the Phillies should've missed.

- Important to note, the best teams in baseball had the highest chance to make the playoffs, while the worst teams in baseball had the lowest chance. There is very little fraudulent success in baseball, especially in such a long season.

- This model can be beneficial at the beginning of playoffs to predict which team may go the furthest in the postseason
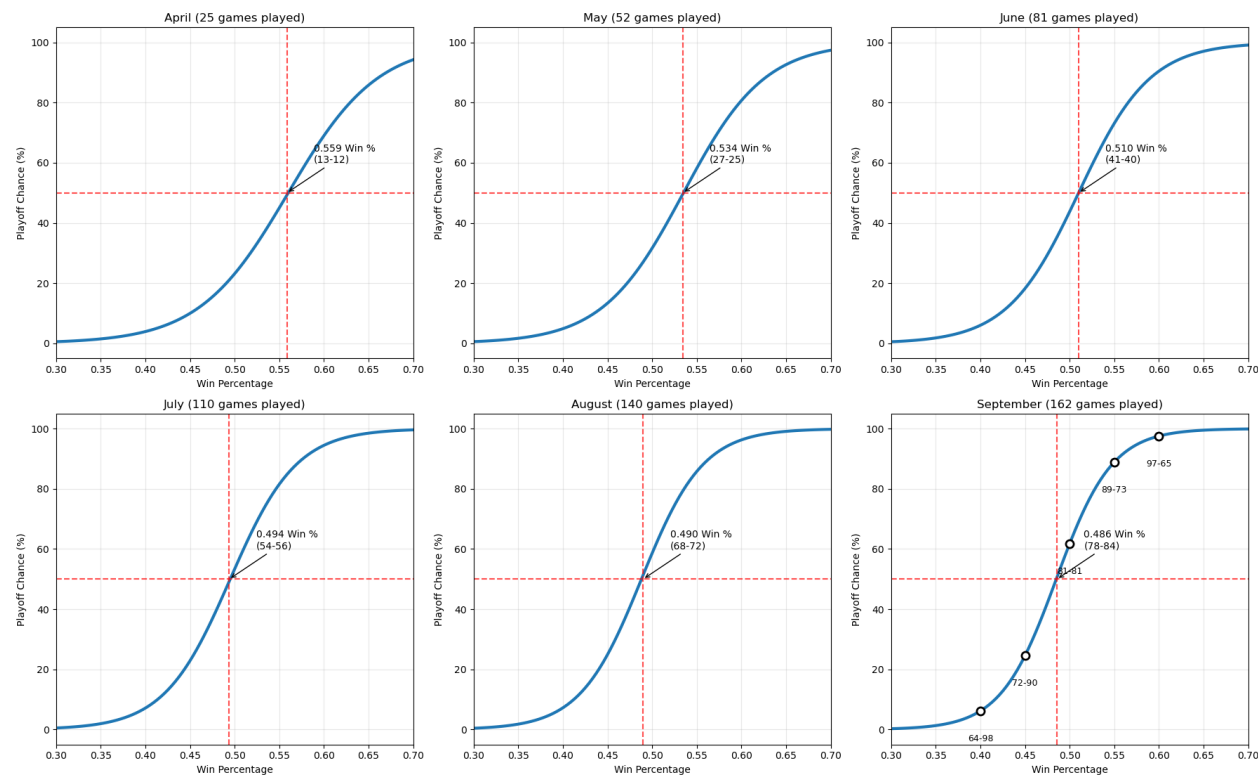
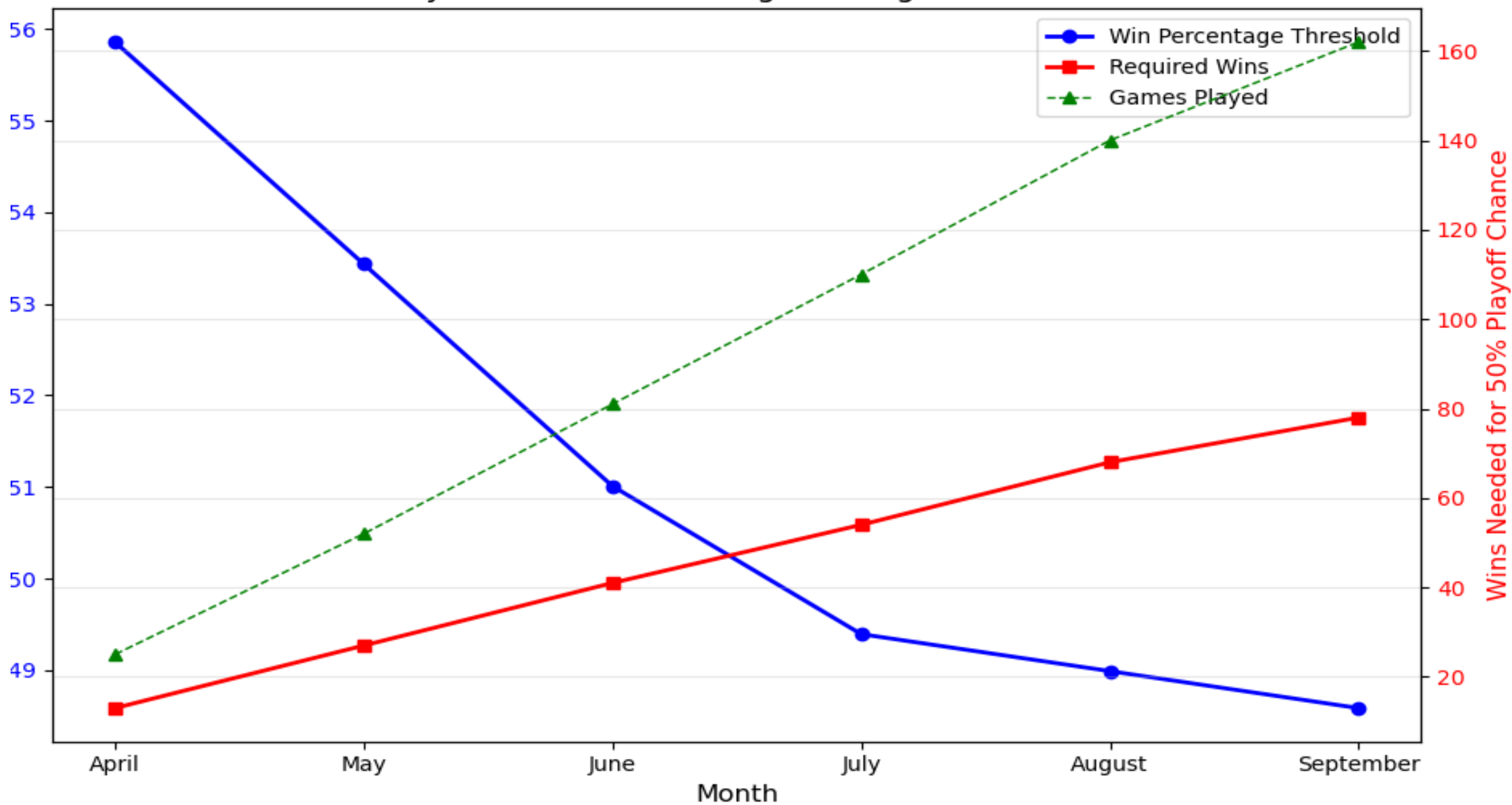MLB 2024 Season: Predicted Playoff Probability by Team
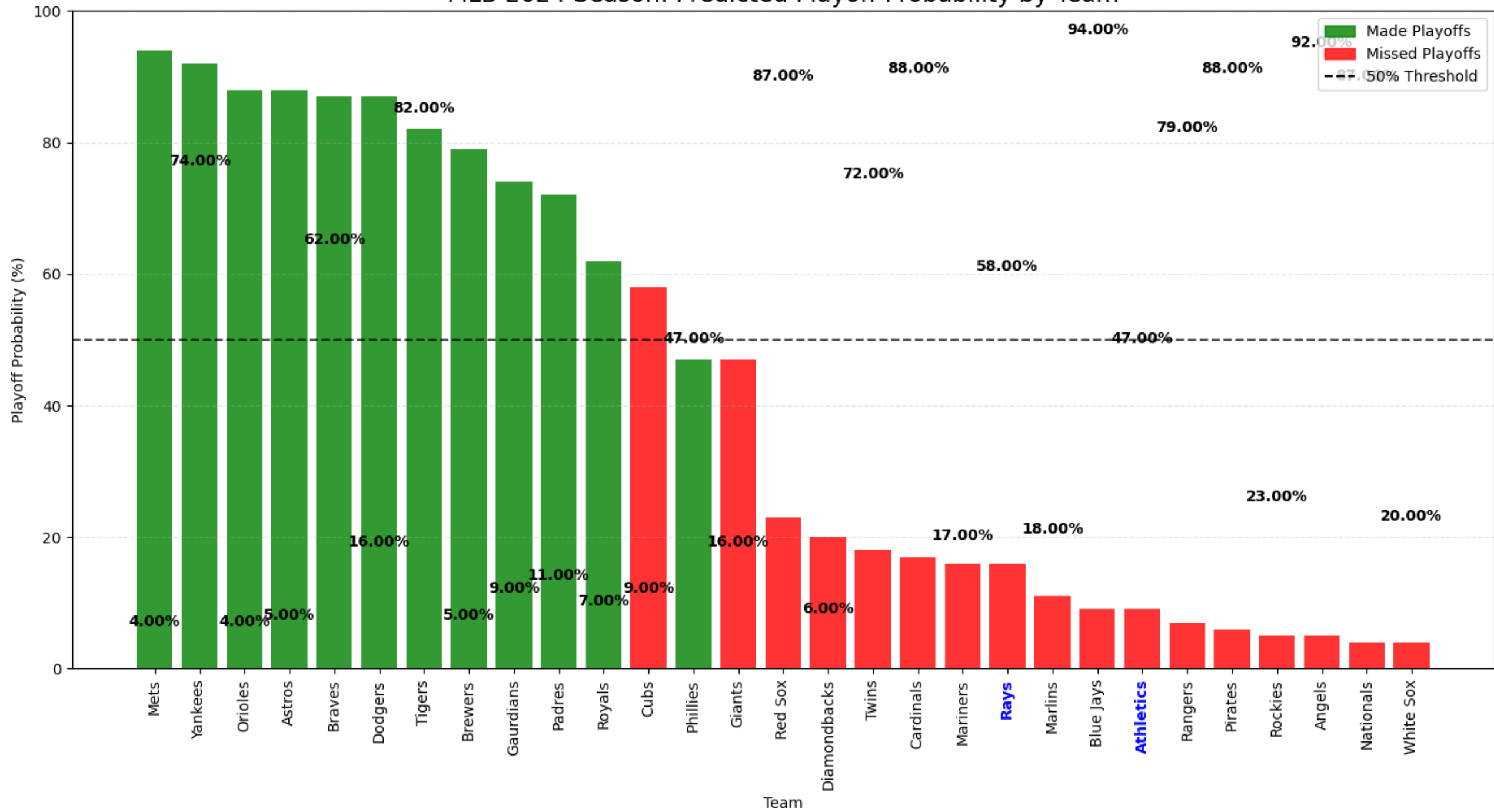
Record Benchmarks

MLB Playoff Chances by Win Percentage and Month

50% playoff chance thresholds by month:

Month	Win %	Record	Games Played
April	0.559	13-12	25
May	0.534	27-25	52
June	0.510	41-40	81
July	0.494	54-56	110
August	0.490	68-72	140
September	0.486	78-84	162

How Playoff Thresholds Change Throughout the Season

Model Use Case Example

New York Yankees 17-12 (4/29)

Baltimore Orioles 11-17 (4/29)

```
▶| get_playoff_prediction()


=== MLB Playoff Predictor ===
Enter month (3-9, where 3=March, 9=September): 4
Enter current number of wins: 17
Enter current number of losses: 12

Based on a record of 17-12 in month 4:
Playoff chance: 67.8%
Outlook: Better than even playoff chances
```

```
▶| get_playoff_prediction()


=== MLB Playoff Predictor ===
Enter month (3-9, where 3=March, 9=September): 4
Enter current number of wins: 11
Enter current number of losses: 17

Based on a record of 11-17 in month 4:
Playoff chance: 25.0%
Outlook: Long shot for playoffs
```

Accuracy and Limitations

```
Cross-Validation Results (5-fold):

Month 4 model:
Accuracy: 0.7667 (±0.1333)
F1 Score: 0.6933 (±0.1902)
ROC-AUC: 0.8500 (±0.1333)

Month 5 model:
Accuracy: 0.8000 (±0.1247)
F1 Score: 0.7200 (±0.1939)
ROC-AUC: 0.7333 (±0.1839)

Month 6 model:
Accuracy: 0.7000 (±0.0667)
F1 Score: 0.5933 (±0.1218)
ROC-AUC: 0.8500 (±0.0952)
```

```
Month 7 model:
Accuracy: 0.7000 (±0.1247)
F1 Score: 0.6676 (±0.1203)
ROC-AUC: 0.7722 (±0.1827)

Month 8 model:
Accuracy: 0.8000 (±0.1944)
F1 Score: 0.7743 (±0.2093)
ROC-AUC: 0.7639 (±0.2334)

Month 9 model:
Accuracy: 0.9000 (±0.0816)
F1 Score: 0.9029 (±0.0820)
ROC-AUC: 0.9222 (±0.0969)

Quick accuracy check: 0.8750
```

- This model does not take into account other teams in the league record
  - If a lot of teams in the league have a great record, or a handful of teams within a division have a great record, it can give a percentage that is too high.
- Model only processes monthly records not day of month. This can cause some limitations in the playoff percentage in the beginning of the month vs the end

MLB Playoff Chances by Win Percentage (Full Season)
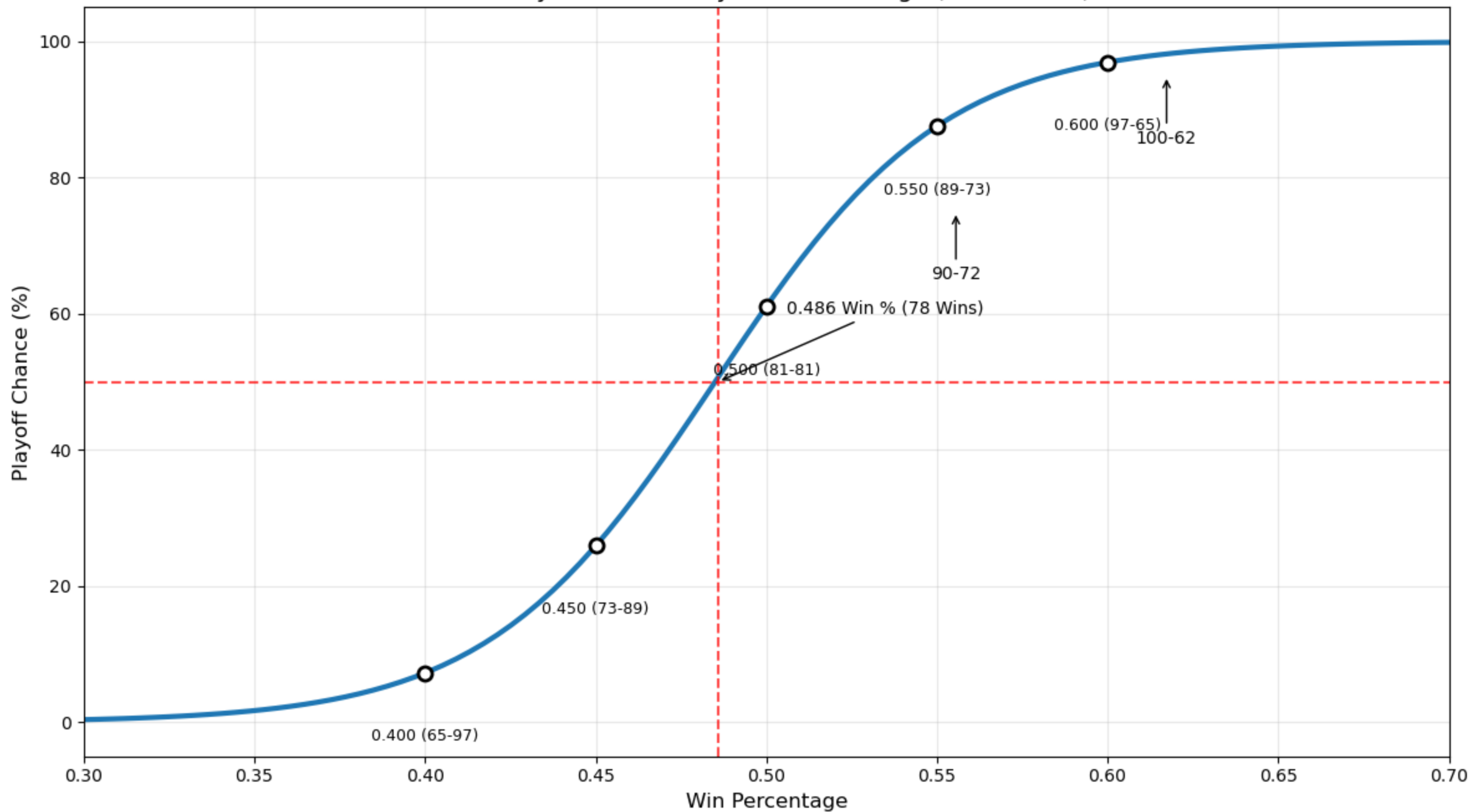
Ethical Considerations

Bias in Data & Decision-Making:

•Over-reliance on historical data may undervalue **intangibles like leadership, team chemistry, and clutch performance**.

•Risk of reinforcing **inequalities in player evaluation**, potentially overlooking players with unconventional success paths, such as guys who are more valuable to a team for things that do not show up on a stat sheet.

Transparency & Fairness:

•Teams must balance analytics with **fair player evaluations** to avoid unfair roster decisions based solely on data.

•Ethical use of analytics ensures **players are not unfairly penalized for small sample sizes or systemic biases**.

Data Privacy:

•Advanced analytics often require **biometric and performance tracking data**—it is important to make sure player data is protected and kept safe, especially regarding injury history, etc.

Summary:

Overall, we were confident in our results and being able to predict what it takes to make the playoffs, and how a team could achieve getting there.

We would feel confident bringing this to a General Manager/Front Office of a team. It can help making decisions to construct a roster that would successfully make the playoffs year in and year out.

Our Playoff Model can be used for sports betting decisions on which team could win the World Series

Questions?