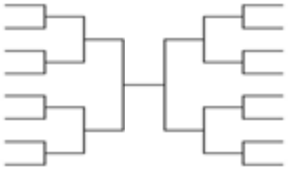
Box Score to Bracket



Evaluating the Predictive Power of NBA Regular Season Performance on Playoff Outcomes

"Can watching a few NBA games tell you who's making a deep playoff run?"

Andrew Scheiner & Sid Lamsal - DATA400 - Spring 2025

Research Question

- Can we use NBA regular season box score data to predict how far a team will go in the playoffs.
- Use sample sizes of 5-10 games at a time to make predictions
- Use neural networks to train a classifier model
    - Playoffs?, *n*th round exit?

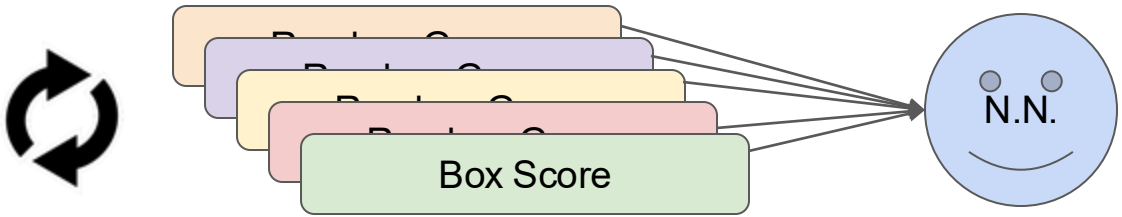
Data Collection

- Scraped standard and advanced box scores from NBA.com
- Used requests and pandas libraries in Python
- Used nba_api to get playoff wins
- Total: 26418 rows × 91 columns
- “Playoff Wins” is changed into a categorical variable

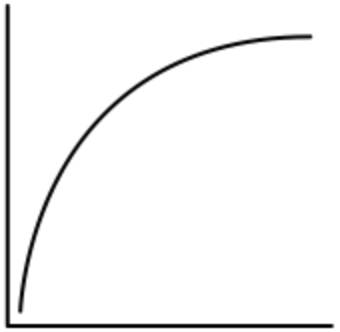
Season	Team	GameId	*[Stats]*	Playoff Wins
2023-24	Bos	0042300405	…	16
⋮	⋮	⋮	⋮	⋮

```
2014-15: dict
  x: ndarray
  y: Series
2015-16: dict
  x: ndarray
  y: Series
2016-17: dict
  x: ndarray
  y: Series
2017-18: dict
  x: ndarray
  y: Series
2018-19: dict
  x: ndarray
  y: Series
2019-20: dict
  x: ndarray
  y: Series
2020-21: dict
  x: ndarray
  y: Series
2021-22: dict
  x: ndarray
  y: Series
2022-23: dict
```

Model:



Accuracy v Games



Big Goal: Predict the Champs

See how “valuable” watching regular season games is for estimating Playoff performance

Things to consider:

- Different parts of the season
- Rolling averages
- Different models
- Cumulative data

Two Input Types:

Raw Boxscores

Bulls 1
Celtics 1
NYK 1
Bulls 2

Input consists of box score stats from all games of season *x*

Training:

These performances from season *x* led to these playoff outcomes

Testing:

Given these performances from season x+1, what are the outcomes?

Rolling Averages

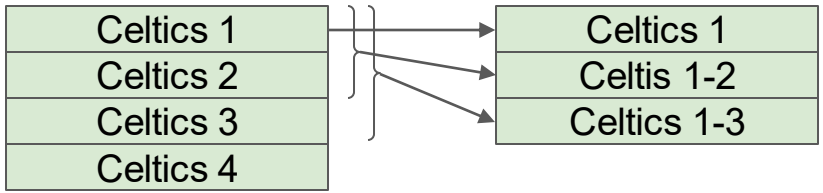
Input is the average performance of the teams at game *n* of season *x*

Training:

At game *n* of season x these are the averages or the teams and here are the playoff outcomes.

Testing:

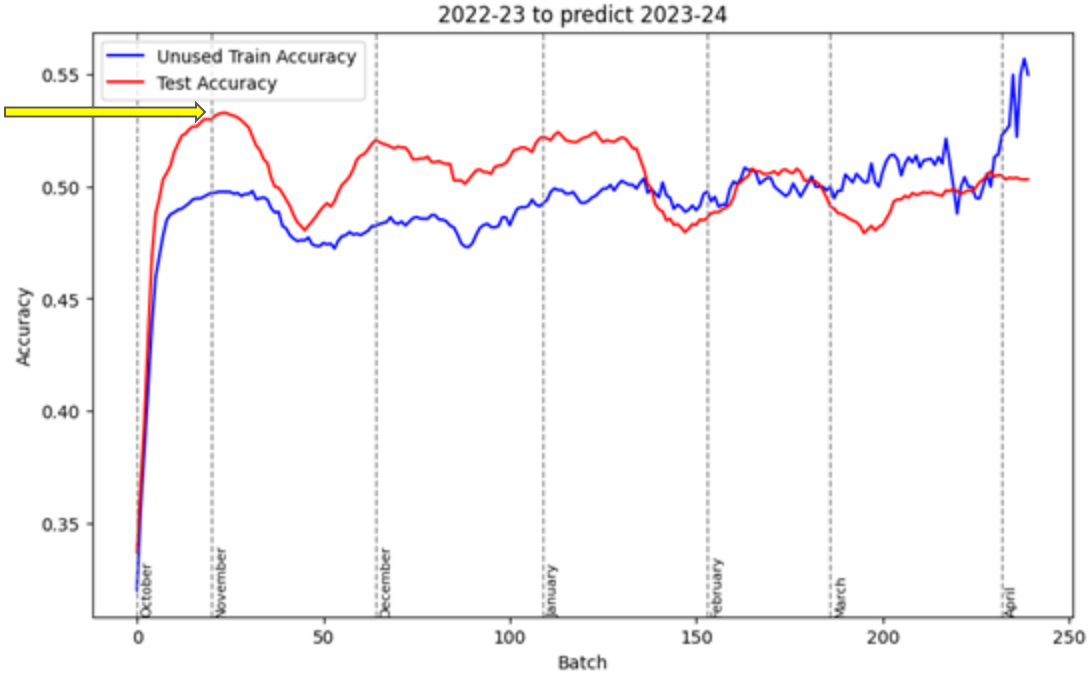
Given the averages at game n, season x+1, what are the playoff outcomes



Model Results: Box score predictions

1 batch = 5 games,

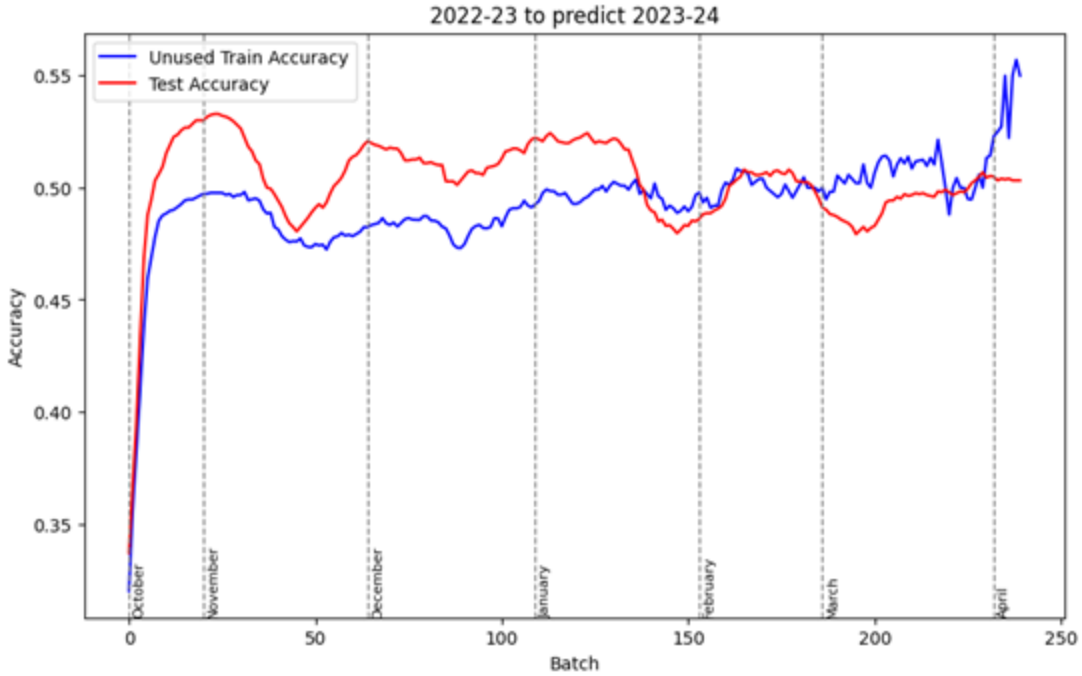
At about batch 20 (game 100), given a boxscore of a game, the model can predict where the team will finish in the playoffs at about 53% accuracy



Model Results: Box score predictions

Accuracy peaks around game 100 (batch 20) meaning:

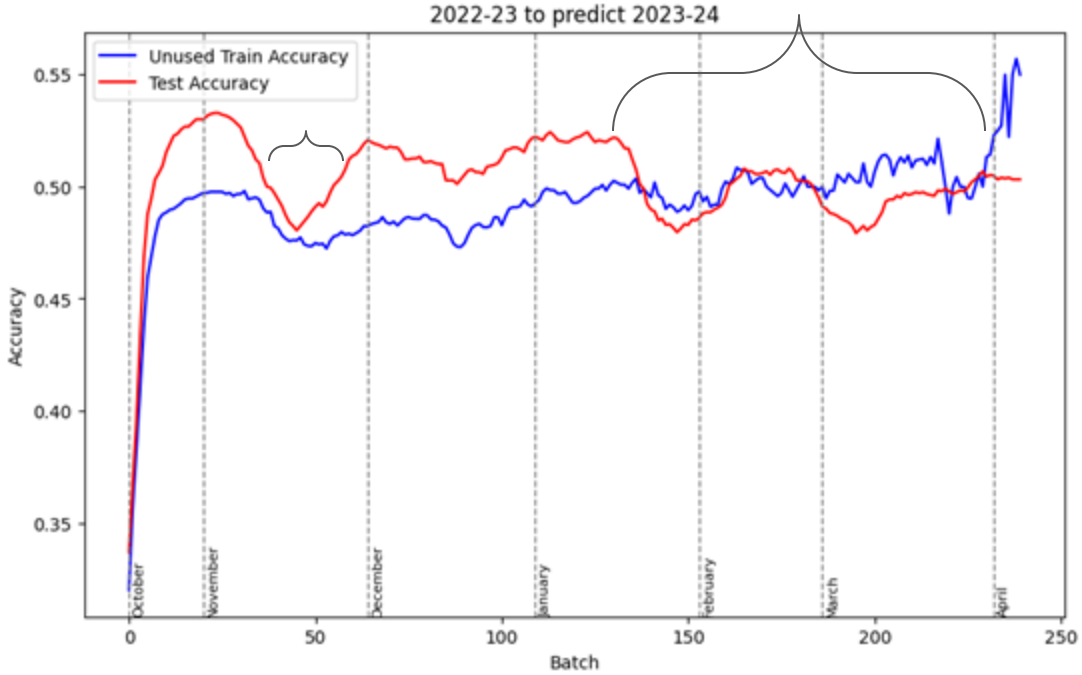
- At game 100 the model is as good as it gets at predicting playoff performance based on a single game statline



Model Results: Box score predictions

As season goes on, accuracy decreases meaning:

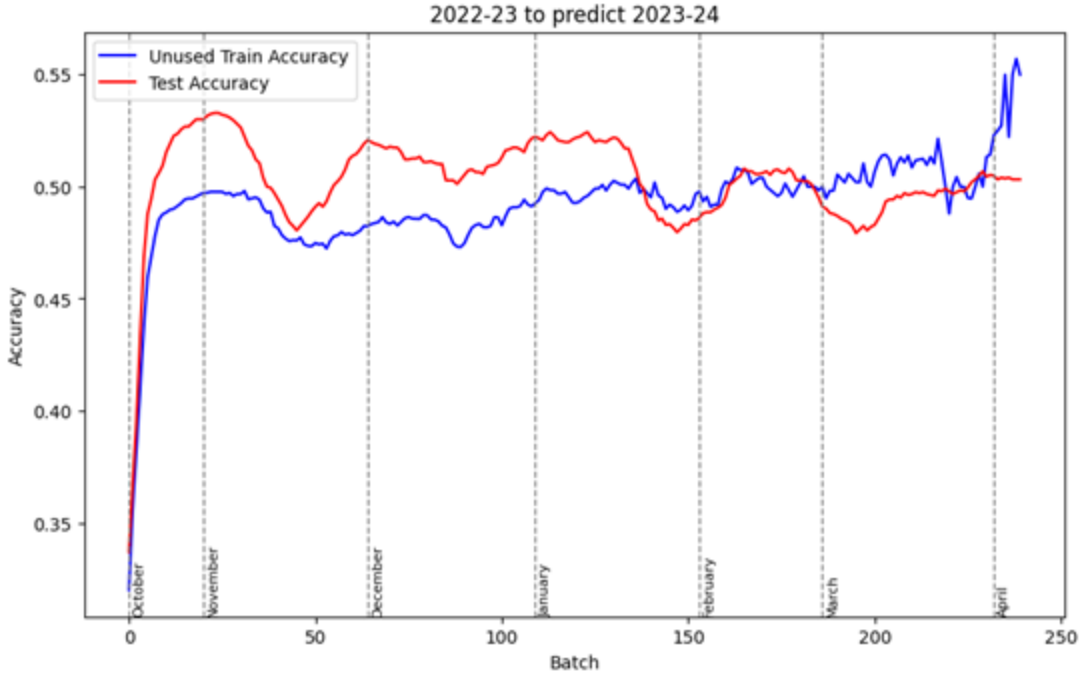
- Maybe higher variance of play
  - Bad teams have good nights
  - Good teams have off nights
- Later games are less indicative of playoff performance/real capability of teams



Model Results: Box score predictions

As season goes on, games may feel “boring” to some as they might not reflect what teams are really capable of as opposed to start of the season.

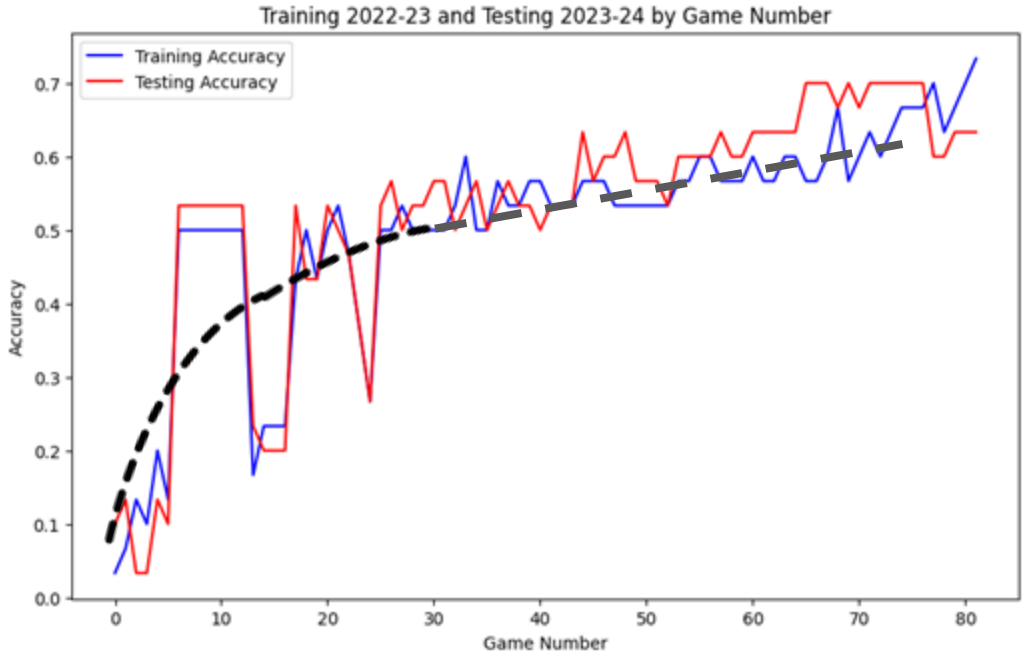
If you want to know who will win it all, later games feel less meaningful



Model Results: Rolling Averages

Looking at game by game averages, accuracy increases drastically around game 15 (mid November). Meaning:

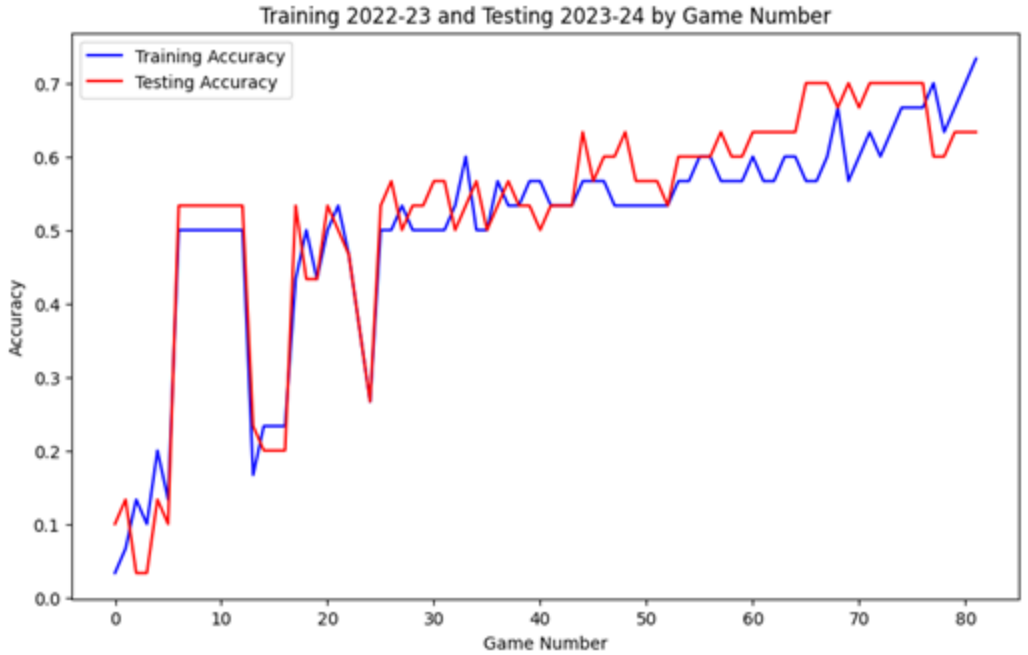
- Around November is a good time to gauge how good teams are.
- You will have close, but not the best predictions



Model Results: Rolling Averages

Accuracy continues to increase as season progresses. Meaning:

- As teams play more games their averages are increasingly better predictors for playoff performance



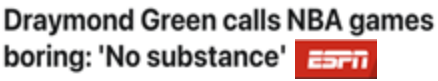
Verdict: Regular Season Does Matter

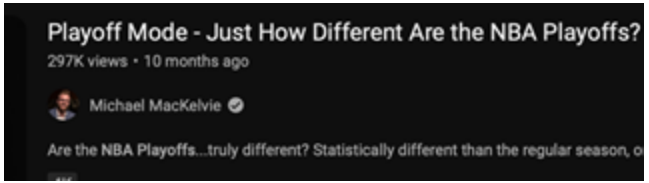
The results suggest that regular season is indicative of Playoff performance, however:

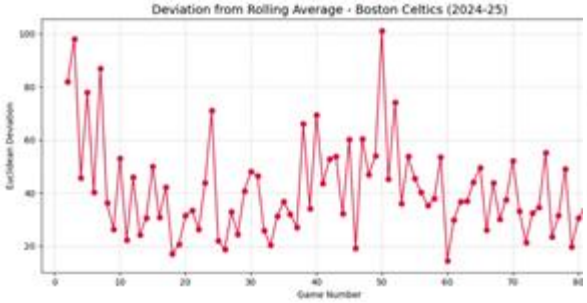
- Per unit predictive power of watching games is lower after mid November (that only a month into the season)
- Watching games beyond this point decreases accuracy game by game
- However, holistically, accuracy goes up.

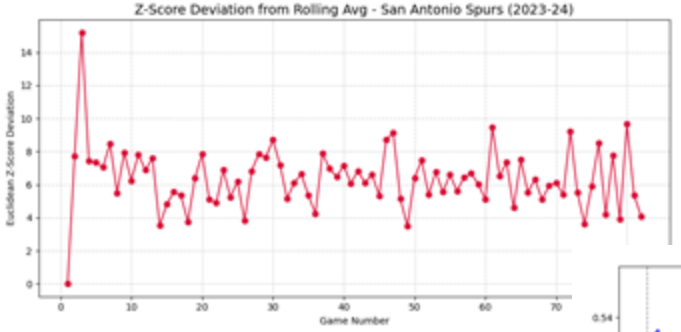
Verdict: Regular Season Does Matter

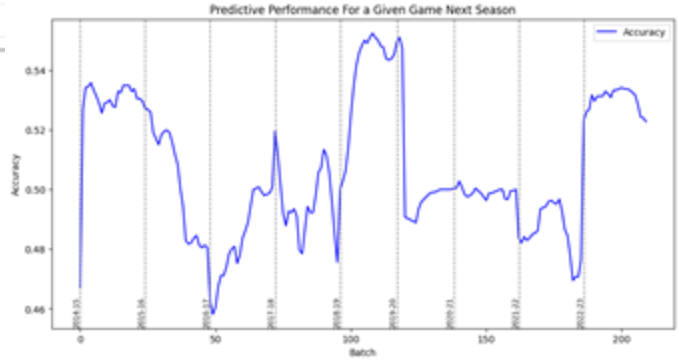
NBA Games later into the season may seem boring to avid fans (the people complaining) as team performance game by game may not be indicative of how they will perform when it matters: the playoffs.



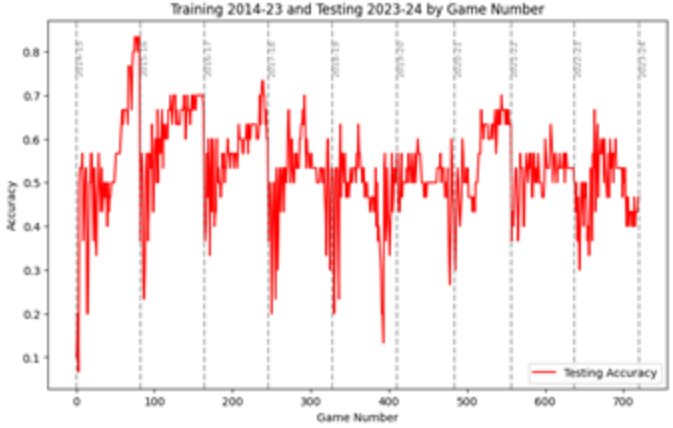








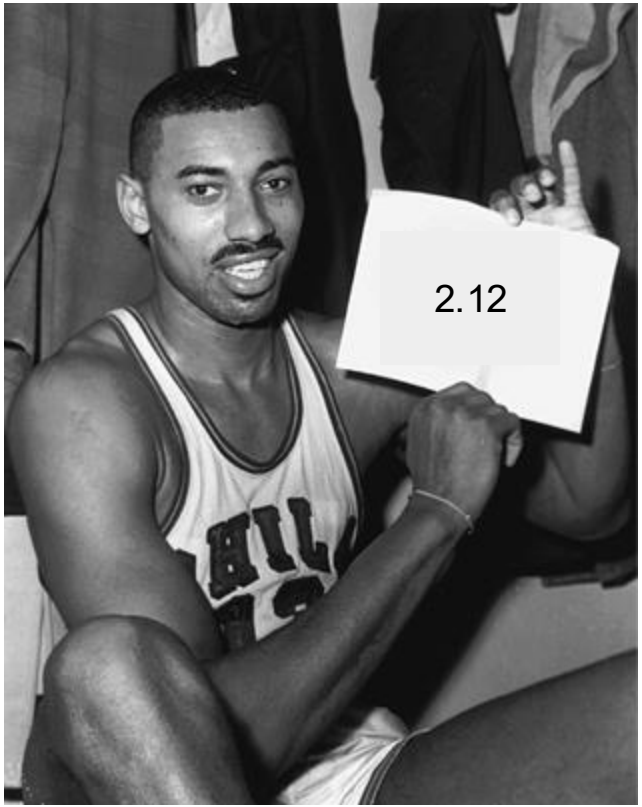
Other Things We Looked At



Quick Summary

- Training from 2014-2024 and predicting the next season showed:
  - Certain seasons were more predictive than others
  - If we carry over weights, accuracy drops start of each season
  - Long term training wasn't much better

- Numerical response variable showed:
  - Can predict how many playoff wins plus/minus 2-3 games
  - Best was plus/minus 2.12 games



2025 Playoffs Currently …

2025 Playoffs Prediction …

Predictions

7	Orlando Magic - se	41	41	.500
8	Atlanta Hawks - pi	40	42	.488
9	Chicago Bulls - pi	39	43	.476
10	Miami Heat - pi	37	45	.451
11	Toronto Raptors - o	30	52	.366
12	Brooklyn Nets - o	26	56	.317

Predictions are fairly accurate so far

Incorrect:

- Nuggets vs Clippers: Very close series, had a game winning dunk
- Bulls - Heat, Magic - Nets: These are bad teams, the model predicted none of them to win a single playoff game

Predictions

Predictions seem nuanced

- The Lakers (3rd seed) and Rockets (2nd seed) were upset by the Timberwolves (6th seed) and Warriors (7th seed) respectively. The model picked up on this meaning it was able to look past the gap in regular season wins and pick out the better team.

Predictions

Early Predictions are solid

- As previously mentioned, the best MAE was plus/minus 2 games at game 55 of the season.
- At game 55 the bracket looks almost identical to the game 82 bracket (1 difference). Meaning, the graphs were not wrong, the model can get pretty close to its best performance early on.

XGBoost Tree

- Input: Weighted average at game 82 of each season from 2014-2023
- Test: 2023-24
- Used for variable importance

Classification accuracy on 2023-24: 0.667

Classification report:

	precision	recall	f1-score	support
0	1.00	0.75	0.86	16
1	0.57	0.67	0.62	6
2	0.33	0.75	0.46	4
3	1.00	0.50	0.67	2
4	0.00	0.00	0.00	1
5	0.00	0.00	0.00	1
accuracy			0.67	30
macro avg	0.48	0.44	0.43	30
weighted avg	0.76	0.67	0.69	30

Top 10 important features:

	Feature	Importance
0	WL	0.121258
26	E_NET_RATING	0.077294
30	AST_RATIO	0.048791
13	AST	0.044858
3	FG_PCT	0.042614
14	TOV	0.032086
35	EFG_PCT	0.030244
18	PF	0.028444
16	BLK	0.027607
36	TS_PCT	0.025817

Stakeholders

- NBA admins
  - Should the commissioner shorten the season?
  - Can use our findings to make schedule more exciting for fans
    - In season tournament during “boring” periods to get teams to play their best
- Teams (owner, GM, coaches)
  - Strategy updates like using certain stats to predict team performance
  - Teams might start valuing certain stretches of the season more

Stakeholders

- NBA players
  - More emphasis on consistency and specific metrics that correlate with better playoff chances
  - Urgency to get off to a hot start

- Fans (casual, bettors)
  - Fans might be able to optimize the amount of games that they have to watch.
  - Bettors and analytical fans can benefit from our findings to predict playoffs.

Implications

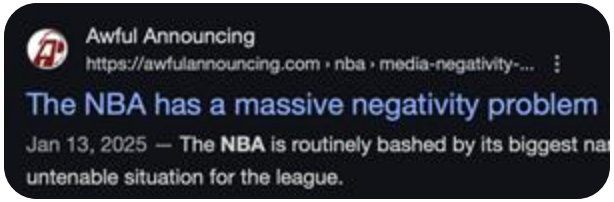
Societal: Sports betting, social media takes, fan interest in NBA

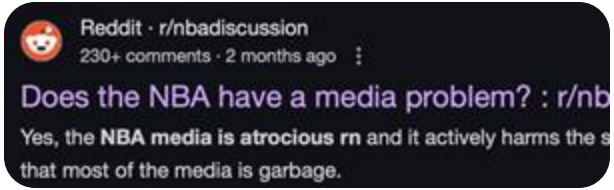
Ethical: Is forcing player to play 82 games justified? Is the play-in/in-season tournament justified?

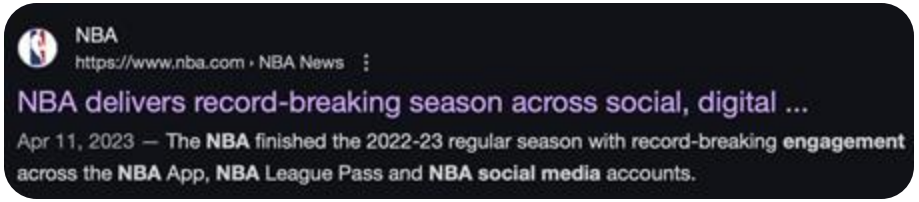
Legal: If season is shortened, how would the NBA would go about restructuring player contracts, TV deals, and deals with owners to avoid a lockout?

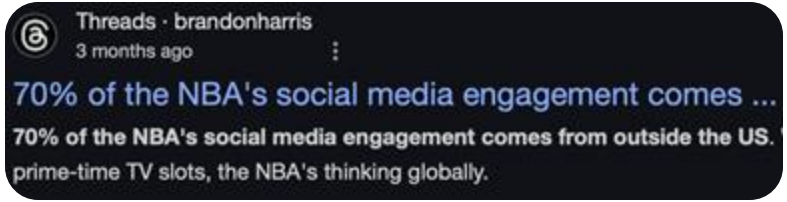
Implications: What's really the problem?

- Increasing issues with negative NBA media coverage
  - TNT/ESPN trashing the “quality” of the game
  - Allegations of hot take hunting and rage bait
  - Divide between fans and mainstream media
- Media seems unable to appreciate new product.









Thanks! Questions?