STAT 400 FINAL PROJECT

GLM & BASKETBALL ANALYTICS

Ishan Agrahar, Jake Sturges, Noah Tobias, and Prajwal Bhandari

RESEARCH QUESTION

What set of variables, defensive or offensive, have higher predictive strength for playoff appearance?

We compare three models: defensive stats only, offensive stats only, and combined.

DATA CLEANING & TRANSFORMATION

GROUP & AGGREGATE

Grouped raw game data by season & team.

Computed mean values for all performance metrics (FGA, FGM, FG%, 3PT%, rebounds, assists, turnovers, etc.).

ADD CONFERENCE LABELS

Assigned each team to Eastern or Western Conference based on franchise location.

CREATE WIN PERCENTAGE

Calculated WINPCT as the proportion of games won per season by each team.

MERGING PLAYOFF APPEARANCE

PREPROCESSING STEPS

101010
011011
010101
101010

Converted comma-separated playoff years into binary indicators (1 = playoffs, 0 = no playoffs) for each season from 2010–11 through 2023–24.

Reshaped data from wide to long format using pivot_longer to match the structure of the game statistics dataset.

TEAM NAME STANDARDIZATION

"Philadelphia Sixers" → "Philadelphia 76ers"

"Charlotte Bobcats" → "Charlotte Hornets"

"New Jersey Nets" → "Brooklyn Nets"

"New Orleans Hornets" → "New Orleans Pelicans"

"LA Clippers" → "Los Angeles Clippers"

These corrections ensured successful joins between datasets with zero mismatches.

DATASET SUMMARY & MISSING VALUES

420

TOTAL OBSERVATIONS

Team-season combinations from 2010–2024

27

VARIABLES

Including shooting stats, rebounds, assists, turnovers, and playoff indicator

0

MISSING VALUES

Complete dataset with no NA values across all columns

14

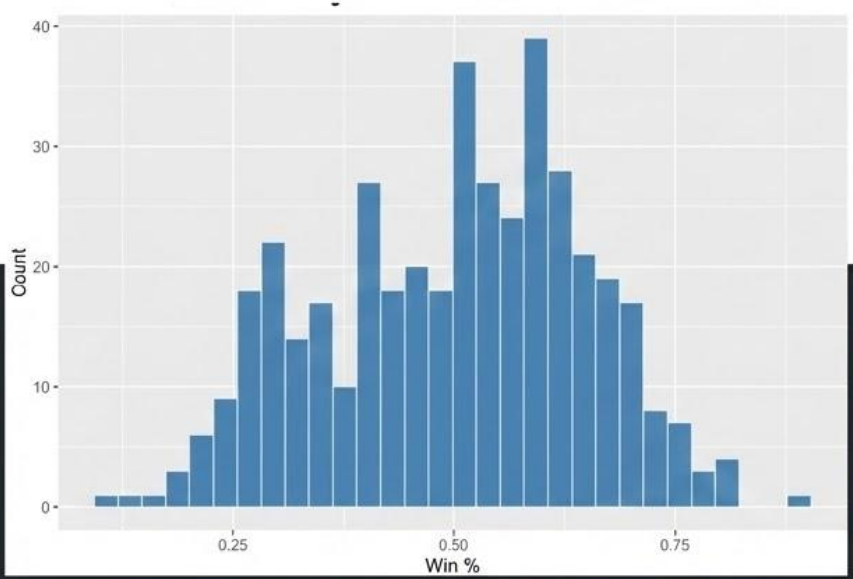
SEASONS COVERED

From 2010–11 through 2023–24 NBA seasons

Key predictors include points per game (PTS), plus-minus (PLUS_MINUS), field goal percentage (FG_PCT), and three-point percentage (FG3_PCT). Win percentage (WINPCT) was intentionally excluded as a direct predictor since the top 8 teams per conference automatically qualify for playoffs based on win record.

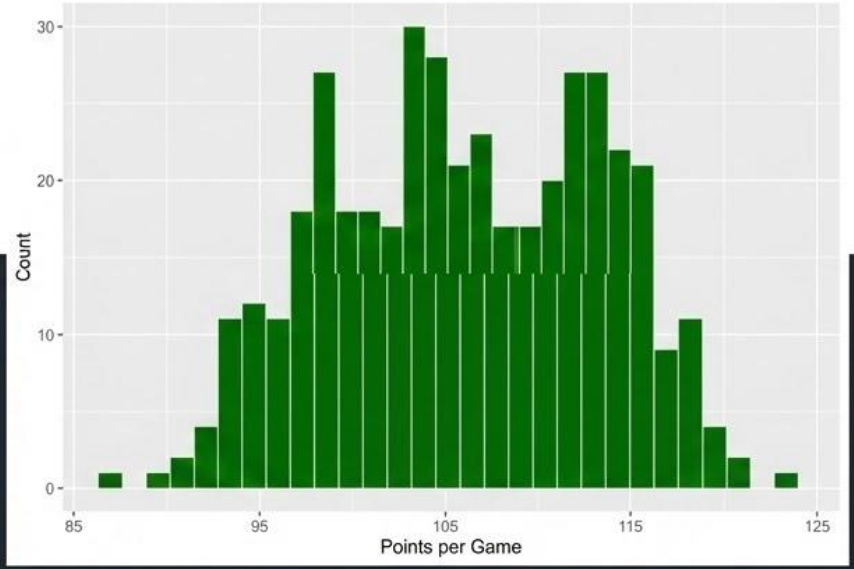
UNIVARIATE EDA: WIN PERCENTAGE & SCORING

WIN PERCENTAGE DISTRIBUTION



Win percentage centers slightly above 50% with a left skew. Extremely low win rates (below 15%) are more common than exceptionally high rates. The legendary 2015–16 Golden State Warriors (73–9) represent the sole team exceeding 85% win rate.

POINTS PER GAME DISTRIBUTION



Points per game averages approximately 105, with a range from 87 to 123 PPG. We hypothesize that higher-scoring teams have greater playoff odds, though defensive performance also plays a critical role.

BIVARIATE EDA: PLAYOFF VS. NON-PLAYOFF

POINTS PER GAME

Playoff teams score nearly 5 points more per game on average compared to non-playoff teams, demonstrating offensive efficiency's importance.

FIELD GOAL PERCENTAGE

Playoff teams maintain approximately 2% higher FG%, indicating more efficient shot selection and execution throughout the season.

TURNOVERS

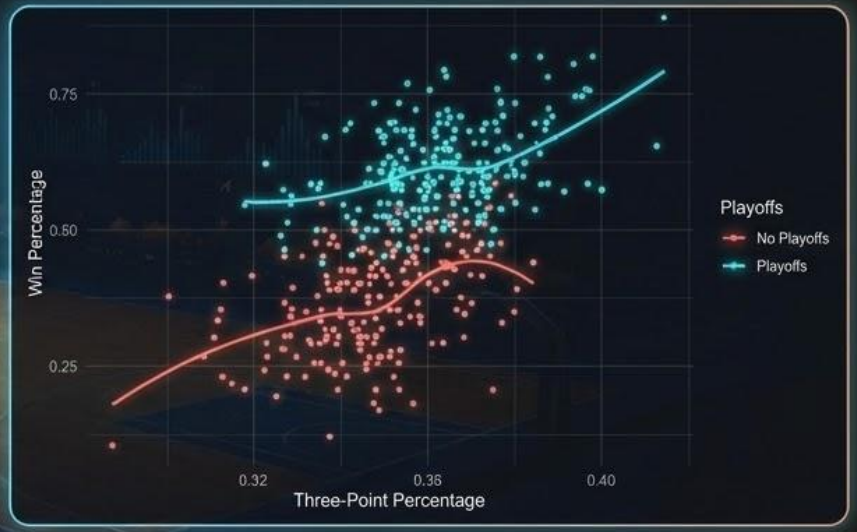
Playoff teams commit fewer turnovers, underscoring that ball security and smart decision-making are key differentiators beyond raw scoring.

SHOOTING EFFICIENCY & DEFENSIVE IMPACT

FIELD GOAL % VS. WIN %

Nearly linear relationship between overall FG% and winning percentage.

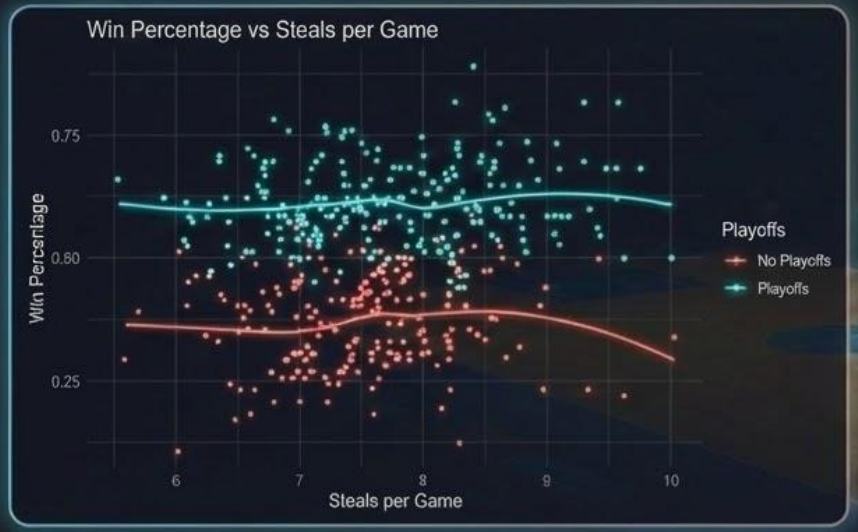
THREE-POINT % VS. WIN %



Elite three-point shooting teams show exponential gains in win percentage, reflecting modern NBA offensive trends.

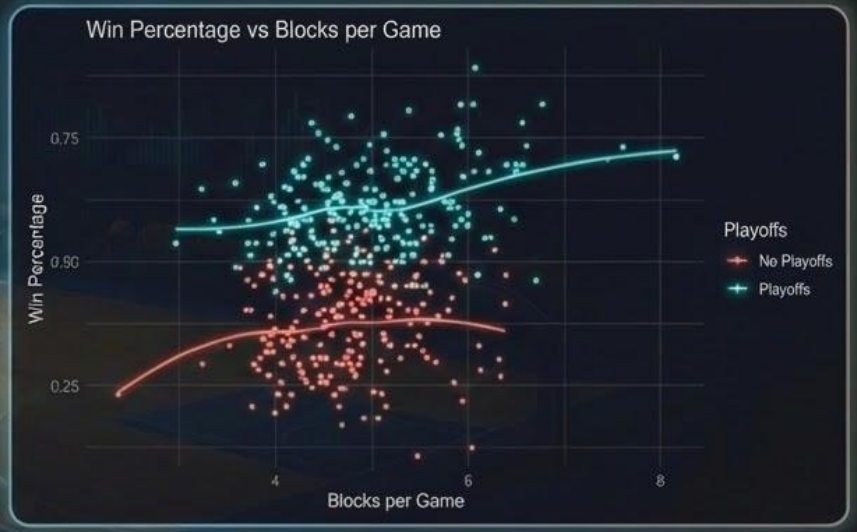
DEFENSIVE IMPACT: STEALS & BLOCKS

STEALS VS. WIN %



More steals correlate with higher win rates, indicating aggressive defense pays off.

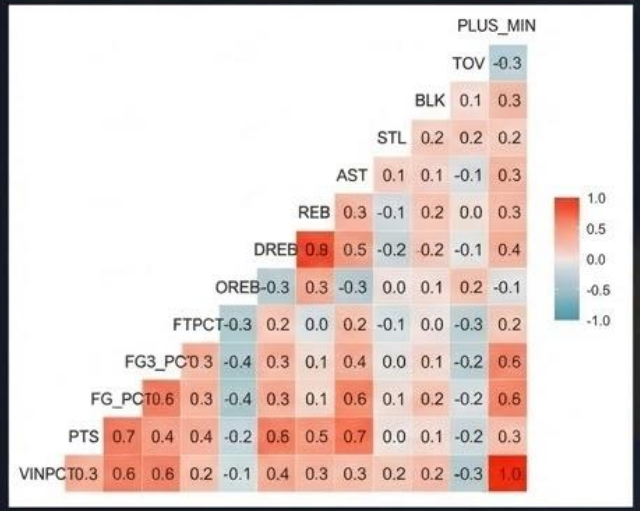
BLOCKS VS. WIN %



Linear relationship shows rim protection is a critical defensive component for winning teams.

CORRELATION ANALYSIS & KEY INSIGHTS

CORRELATION HEATMAP



KEY INSIGHTS

Strongest Correlations with WINPCT

- **PTS:** 0.70 – Scoring output is the top predictor
- **FG_PCT:** 0.60 – Shooting efficiency matters
- **PLUS_MINUS:** 0.60 – Point differential is oritical
- **AST & FG3_PCT:** Moderate positive associations

Multicollinearity Considerations

- REB and DREB are highly correlated (0.80). FG_PCT and PTS show moderate correlation (0.40). Variable selection will require careful consideration to avoid redundancy in the final GLM.

Overall, offensive efficiency, point differential, and scoring emerge as the strongest statistical drivers of playoff success. These insights will guide predictor selection for our logistic regression model.

MODEL BUILDING APPROACH

1. DATA SPLIT

80/20 train-validation split with reproducible seed

2. VARIABLE SETS

Defense: DREB, STL, BLK, PF.
Offense: FGM, FTM, AST, OREB, BLKA, TOV, PTS

3. MODEL CREATION

Build null and full models for each variable set

4. SELECTION

Apply stepwise selection with AIC criterion

KEY ASSUMPTION: Team performance in one year is independent of prior year performance.

DEFENSE MODEL RESULTS

SIGNIFICANT PREDICTORS

DREB: $p < 0.001$, coeff 0.317

STL: $p < 0.001$, coeff 0.582

PF: $p < 0.001$, coeff -0.337

BLK: $p = 0.078$, coeff 0.297

MODEL FIT STATISTICS

AIC:	NULL DEVIANCE:	RESIDUAL DEVIANCE:
425.41	464.83	415.41

MULTICOLLINEARITY CHECK (VIF)

All VIF values near 1 (No Issues)

DEFENSE MODEL INTERPRETATION

DEFENSIVE REBOUNDS

+1 rebound increases playoff odds by factor between **1.20-1.57**

STEALS

+1 steal increases playoff odds by factor between **1.32-2.45**

PERSONAL FOULS

+1 foul reduces playoff odds by factor between **0.68-0.96**

BLOCKS

+1 block changes playoff odds by factor between **0.97-1.88**

All confidence intervals calculated at 95% confidence level using profile likelihood methods.

OFFENSE MODEL RESULTS

SIGNIFICANT PREDICTORS

Blocks Against (BLKA):
coeff -1.335, p < 0.001

Free Throws Made (FTM):
coeff 0.365, p < 0.001

Turnovers (TOV)
p = 0.024, coefficient -0.267.

Field Goals Made (FGM)
p = 0.091, coefficient 0.087.

MODEL FIT STATISTICS

AIC:	NULL DEVIANCE:	RESIDUAL DEVIANCE:
390.33	464.83	380.33

MULTICOLLINEARITY CHECK (VIF)

All VIF values near 1 (No Issues)

AIC: 390.33 | VIF values all near 1 (no multicollinearity)

OFFENSE MODEL INTERPRETATION

BLOCKS AGAINST (BLKA)

+1 block against reduces playoff odds by factor between **0.16-0.39**

FREE THROWS MADE (FTM)

+1 free throw made increases playoff odds by factor between **1.23-1.69**

TURNOVERS (TOV)

+1 turnover reduces playoff odds by factor between **0.60-0.96**

FIELD GOALS MADE (FGM)

+1 field goal made increases playoff odds by factor between **0.98-1.20**

All confidence intervals calculated at 95% confidence level using profile likelihood methods.

COMBINED MODEL: BEST OF BOTH

Finally, we will consider the final model including both sets of variables.

SIGNIFICANT PREDICTORS ($p < 0.05$)

Blocks Against (BLKA)
$p < 0.001$, coefficient -1.426.

Free Throws Made (FTM)
$p < 0.001$, coefficient 0.399.

Steals (STL)
$p < 0.001$, coefficient 0.890.

Defensive Rebounds (DREB)
$p < 0.001$, coefficient 0.362.

Turnovers (TOV)
$p = 0.008$, coefficient -0.365.

MODEL FIT STATISTICS

AIC:	NULL DEVIANCE:	RESIDUAL DEVIANCE:
357.83	464.83	339.83

MULTICOLLINEARITY CHECK (VIF)

All VIF values near 1 (No Issues)

BLKA: 1.21 FTM: 1.20 STL: 1.38 DREB: 1.80 TOV: 1.33

The combined model features 5 significant predictors ($p < 0.05$) out of the original 12 variables. Multicollinearity is not an issue, with all VIF values well below 2 (highest is 1.70 for FGM, see gauges above), indicating a healthy multiple regression model.

COMBINED MODEL INTERPRETATION

SIGNIFICANTLY REDUCE ODDS

BLOCKS AGAINST (BLKA)

+1 block against reduces playoff odds by factor between **0.15-0.38**

TURNOVERS (TOV)

+1 turnover reduces playoff odds by factor between **0.53-0.91**

SIGNIFICANTLY INCREASE ODDS

FREE THROWS MADE (FTM)

+1 free throw made increases playoff odds by factor **1.25-1.79**

STEALS (STL)

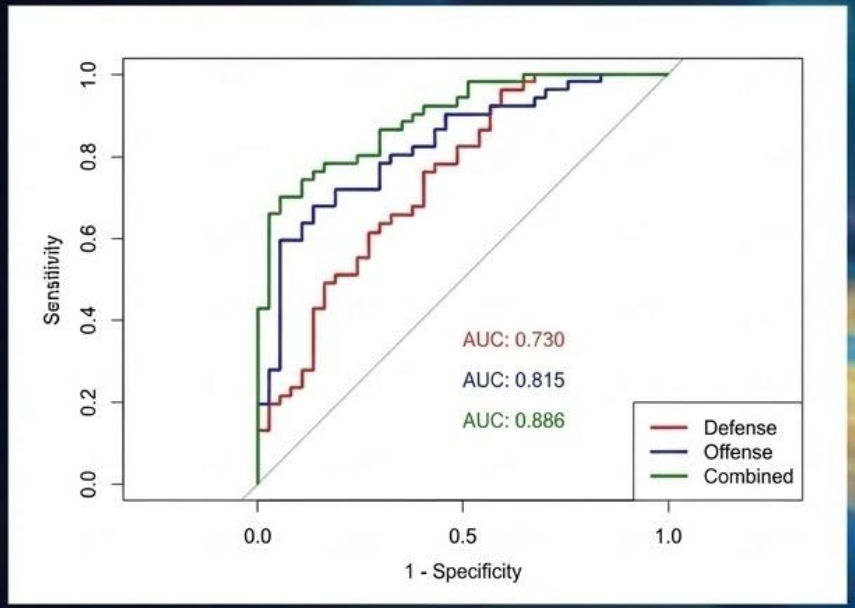
+1 steal increases playoff odds by factor between **1.69-3.60**

DEFENSIVE REBOUNDS (DREB)

+1 rebound increases playoff odds by factor **1.20-1.74**

At a 95% confidence level, Blocks Against and Turnovers significantly reduce playoff odds, while Free Throws Made, Steals, and Defensive Rebounds significantly increase them.

ROC CURVE ANALYSIS



0.730

Defense Model AUC

Lowest predictive power

0.815

Offense Model AUC

0.085 higher than defense

0.886

Combined Model AUC

Highest predictive accuracy

Conclusion: Both defense and offense are necessary, but the combined model (AUC 0.886) outperforms individual offensive (0.815) and defensive (0.730) models. Future analysis should use cross-validation or Monte Carlo simulations for robustness.