NBA Salaries: How Much is that Buzzer Beater Worth?

Eddie and Co.

Introduction

NBA salaries reflect the combined effects of player performance, recognition, and team strategy. Prior research shows that compensation is strongly tied to measurable statistics and awards-based reputation (Berri & Schmidt, 2006; Stiroh, 2007; Simmons & Berri, 2011). Our project examines which factors best explain salary variation in the 2024–2025 season, focusing on three dimensions of player value: scoring (PTS), defensive activity (stocks), and award recognition. These variables represent offensive production, defensive impact, and league-wide acknowledgment—each commonly cited in contract evaluations

Our general research question is:

What factors from the 2024-2025 season best explain variation in future salaries (2025-2026 onwards)?

To address this, we developed three specific research questions:

- Do NBA players who score more points (PTS) earn higher salaries on average?

- Is there a meaningful positive relationship between a player's defensive activity ("stocks") and their salary?

- Do players who have earned at least one major NBA award receive higher salaries compared to players with no awards?
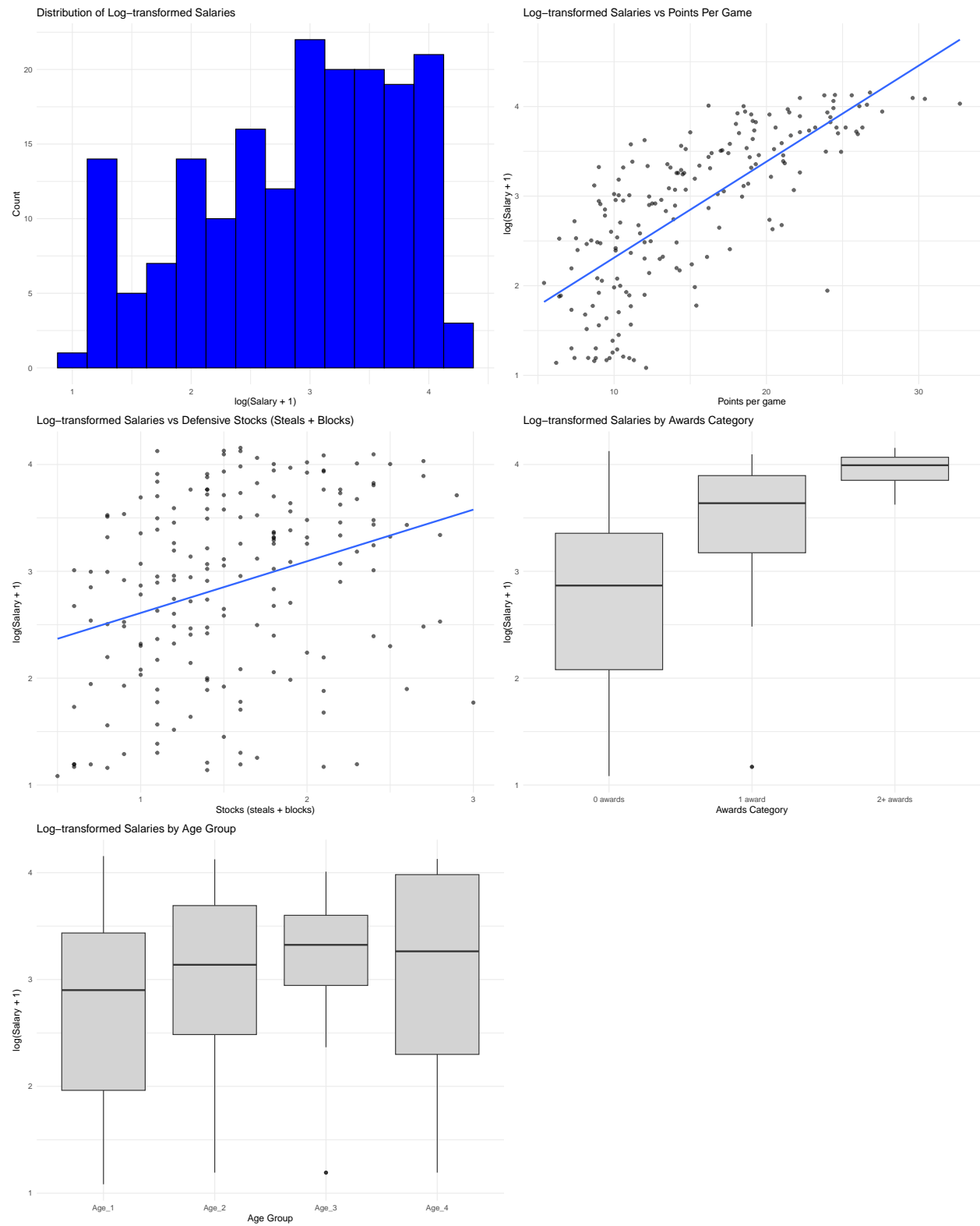
These questions allow us to assess how performance and recognition jointly relate to compensation and how our findings compare with established sports-economics patterns.

Data Summary

Data Sources

Our original dataset included all NBA players who appeared during the 2024–2025 season, combining salary, performance, demographic, and award information. However, to stabilize variance and reduce right skew in the response variable, we restricted the dataset to players averaging more than 20 minutes per game and applied a square-root transformation to salary (sqrt(Salary + 1)). The transformed distribution is noticeably more symmetric and appropriate for linear modeling. Player statistics and awards were sourced from Basketball Reference, while Statista provided team franchise values to add economic context. Player names were standardized, awards was simplified into a three-level categorical variable, and age was grouped into brackets.

Exploratory Data Analysis

Distribution of Log-transformed Salaries

Log-transformed Salaries vs Points Per Game

Log-transformed Salaries vs Defensive Stocks (Steals + Blocks)

Log-transformed Salaries by Awards Category

Log-transformed Salaries by Age Group

EDA Summary

The square-root–transformed salary variable is suitable for regression: its distribution is more symmetric and shows clearer linear relationships with key predictors. Higher scorers ($\approx$ 20+ PPG) consistently fall in the upper salary range, while low scorers cluster near the bottom. Defensive impact (stocks) has a weaker but positive trend. Awards and age also matter—players with awards and those in the 27–34 age range show higher median salaries, while younger players have wider variation due to rookie contracts.

Assessing Multicollinearity

For multicollinearity, we used correlations and VIFs to evaluate overlap among predictors, especially between PTS, FTA, and MP. When VIFs exceeded standard thresholds, we applied variable screening and stepwise selection to reduce the model and reassessed VIFs before final model building.

Methods and Analysis

Building the Model

We began by building our model by estimating our model parameters, starting with only our quantitative predictors and a quantitative interaction term that we believe may contribute to the model. Our initial model using quantitative predictors is as follows:

$$\log(\text{Salary}) = \beta_0 + \beta_1 PTS + \beta_2 MP + \beta_3 PFG\_\%TS + \beta_4 FTA + \beta_5 TRB$$
$$+ \beta_6 AST + \beta_7 \text{Stocks} + \beta_8 \text{Value} + \beta_9 (PTS \times FTA)$$

A Global F-test confirmed the model had predictive value, but VIF analysis revealed severe multicollinearity, especially in the interaction term (VIF $\approx$ 23.4 for PTS×FTA). Because of this, we removed the interaction and refit a model with only main effects.

To determine the most important quantitative variables, we conducted stepwise regression. Stepwise selection chose: PTS, MP, stocks. This produced a simpler quantitative model:

$$\log(\text{Salary}) = \beta_0 + \beta_1 PTS + \beta_2 MP + \beta_3 stocks$$

All three predictors were highly significant and the model achieved Adjusted $R^2 = 0.5945$

Next, we incorporated our qualitative predictors: awards (awards_1, awards_2plus), age groups (22-26, 27-31, 32-34, 35+), positions (PF, PG, SF, SG), and one interaction (Awards

$\times$ Age). Using nested F-tests, we found that age group significantly improved the model, while awards and position did not. Removing non-contributing categorical variables yielded the final model:

$$\log(\text{Salary}) = \beta_0 + \beta_1 PTS + \beta_2 MP + \beta_3 stocks + \beta_4 \text{Age 22-26}$$
$$+ \beta_5 \text{Age 27-31} + \beta_6 \text{Age 32-34}$$

A final Global F-test confirmed the adequacy and significance of this model.

Assessing the Model (Including Cross Validation)

After finalizing our model through variable testing, we conducted a holistic assessment to understand its strengths and limitations. We evaluated our metrics which had values of Adjusted $R^2 = 0.6221$ and RSE $= 0.530$, and we interpreted confidence and prediction intervals to assess practical performance.

We also incorporated cross-validation to evaluate model reliability. Cross-validation allowed us to check for overfitting or underfitting by training the model on most of the data and testing it on the remainder. Using the Validation Set Approach in R, we randomly split the dataset into 80% training and 20% testing, then evaluated the resulting models using $\sqrt{MSE}$ and R^2 which came out to be 0.540 and 0.638 respectively. This process helped determine whether the model generalized well to new data or required further refinement.

Checking Model Assumptions

We examined key regression assumptions using residual diagnostics:

Linearity & Homoscedasticity: Residual–fitted and residual–predictor plots showed no major violations, with only mild non-constant variance.

Normality: The Q–Q plot displayed slight deviation in the left tail due to a few outliers, but overall residuals were reasonably normal.

Influential Observations: Cook's distance, leverage, and studentized residual analyses identified several influential cases (notably observations 154, 161, and 235). These points may warrant further investigation or sensitivity analysis.

Overall, the diagnostics suggest that the assumptions are reasonably met, with only minor departures that do not meaningfully undermine inference.

Results

Final Model

The final model for predicting NBA player average salary (in millions) is:

$$\text{Salary} = \beta_0 + \beta_1 PTS + \beta_2 MP + \beta_3 stocks + \beta_4 \text{Age 22-26}$$
$$+ \beta_5 \text{Age 27-31} + \beta_6 \text{Age 32-34}$$

The final model achieved: Multiple $R^2 = 0.6345$, Adjusted $R^2 = 0.6221$, Residual Standard Error $= 0.530$

A Global F-test indicated the model was highly significant: ($F = 51.21$, $p < 2.2 \times 10^{-16}$).

We evaluated generalization using **5-fold cross-validation**, obtaining: RMSE $= 0.540$, Cross-validated $R^2 = 0.638$, MAE $= 0.440$

These results closely matched the in-sample performance, indicating **no evidence of overfitting**.

Predictor Significance

From the final model output:

- **PTS** was the strongest predictor ($\beta = 0.08228$, $p = 1.03 \times 10^{-12}$)

- **MP** had a significant but smaller positive effect ($\beta = 0.03258$, $p = 0.0234$)

- **stocks** was also significant ($\beta = 0.23531$, $p = 0.0018$)

For age groups (relative to the baseline category):

- **Age 27–31**: significant positive effect ($\beta = 0.39501$, $p = 0.0404$)

- **Age 32–34**: marginally significant ($\beta = 0.43997$, $p = 0.0512$)

- **Age 22–26**: not significant ($\beta = 0.08700$, $p = 0.6420$)

Awards and position effects were discarded because they were not statistically useful in improving model fit.

Diagnostics

Influence metrics (Cook's distance, leverage, and studentized residuals) identified several influential observations—particularly **154**, **161**, and **235**.
Observations **86** and **154** were marked for potential removal in follow-up sensitivity analysis.

Conclusions

Our final regression model explains approximately **62%** of the variation in log-salary, showing strong predictive usefulness while maintaining parsimony.

The most influential predictors were: **Points per game (PTS)** - by far the strongest driver of salary, **Minutes played (MP)** — indicating playing time remains an important valuation metric, **Defensive activity ("stocks")** — also contributed meaningfully

Age group effects were also important.
Players aged **27–31** and **32–34** earned higher salaries than the youngest group, consistent with peak-career performance in the NBA.

Overall, the final model provides a statistically sound and interpretable framework for understanding NBA salary determination.

Example Prediction

Because the final model is fitted on **log-transformed salary**, predictions must be exponentiated back to the original salary scale.

For illustration, consider a player with: **PTS = 24**, **MP = 33**, **stocks = 2.0**, in the **27–31 age group** (indicator = 1), all other age indicators = 0

Using the final model and evaluating the expression:

$$\log(\hat{S}) \approx 4.06 \quad \implies \quad \hat{S} = e^{4.06} \approx 58.1$$

Thus, the predicted **average salary is approximately 58 million dollars**.
This demonstrates that the model produces intuitive, performance-driven estimates that scale naturally with scoring, playing time, defensive contributions, and peak-career age ranges.

Future Improvements

Although the model performed well overall, several limitations remain: While the model performed well, several limitations remain. A few observations (154, 161, 235) showed high influence and may affect coefficient stability. Important salary factors not in the dataset—such as contract timing, injuries, reputation, and CBA rules—could improve realism if added. Finally, future modeling could test nonlinear effects, interaction terms, or regularization methods (ridge/lasso) to better capture skewed salaries and correlated predictors. Overall, the model is strong but leaves room for richer extensions.

Appendix A: Data Dictionary

Variable Name	Abbreviated Name	Description
Games	G	This represents the games played by each respective player.
Minutes Played	MP	This represents the minutes played by each respective player in the overall season.
Points Scored	PTS	This represents the total points scored by each respective player.
Field Goal Percentage	FG_PCT	This represents the percentage of field goals scored by each respective player.
Free Throw Attempts	FTA	This represents the free throw attempts made by each respective player.
Total Rebounds	TRB	This represents the total rebounds made by each respective player.
Total Assists	AST	This represents the total assits by each respective player.
Steals and Blocks Combined	Stocks	This represents the steals and blocks achieved by each respective player.
Franchise Value	Value	This represents the franchise value of the team of each respective player.

Variable Name	Abbreviated Name	Description
Position	Pos	This represents the position of the player. The position can either be center (reference), power forward (pf), point guard (pg), small forward (sf), or shooting guard (sg).
Awards	awards	This signifies the number of awards a player has won, sorted into three groups: 0 awards (reference), 1 award, or at least 2 awards.
Age	Age	This represents the player's age, sorted into 4 groups: 22-26, 27-31, 32-24, and 35+ (reference).

Appendix B: Data Rows

	Player	Team	G	MP	FG_pct	FTA	TRB	AST	stocks	PTS
1	Shai Gilgeous-Alexander	OKC	76	34.2	0.519	8.8	5.0	6.4	2.7	32.7
2	Giannis Antetokounmpo	MIL	67	34.2	0.601	10.6	11.9	6.5	2.1	30.4
3	Nikola Jokić	DEN	70	36.7	0.576	6.4	12.7	10.2	2.4	29.6
5	Anthony Edwards	MIN	79	36.3	0.447	6.3	5.7	4.5	1.8	27.6
6	Jayson Tatum	BOS	72	36.4	0.452	6.1	8.7	6.0	1.6	26.8
7	Kevin Durant	PHO	62	36.5	0.527	5.8	6.0	4.2	2.0	26.6

	Value_Billions	awards_1	awards_2plus	avg_salary_millions	Age_22_26	Age_27_31
1	4.35	0	1	55.3591	1	0
2	4.30	0	1	58.4566	0	1
3	4.60	0	1	59.0331	0	1
5	3.60	0	1	50.6117	1	0
6	6.70	0	1	62.7867	1	0
7	5.43	1	0	54.7086	0	0

	Age_32_34	Age_35_plus	Pos_PF	Pos_PG	Pos_SF	Pos_SG	Age	Awards	log_salary
1	0	0	0	1	0	0	Age_1	2+ awards	4.031744
2	0	0	1	0	0	0	Age_2	2+ awards	4.085247
3	0	0	0	0	0	0	Age_2	2+ awards	4.094896
5	0	0	0	0	0	1	Age_1	2+ awards	3.943748
6	0	0	1	0	0	0	Age_1	2+ awards	4.155545
7	0	1	1	0	0	0	Age_4	1 award	4.020135

Appendix C: Final Model Output and Plots

```
Call:
lm(formula = avg_salary_millions ~ MP + PTS + TRB + AST + Pos_PF +
    Pos_PG + Pos_SF + Pos_SG, data = df)


Residuals:
     Min       1Q   Median       3Q      Max
-29.6270  -6.1698  -0.0801   5.9922  22.7640


Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -14.7630     6.0544  -2.438  0.01575 *
MP            0.4598     0.2511   1.831  0.06876 .
PTS           1.6955     0.2085   8.132 7.47e-14 ***
TRB          -0.1666     0.4924  -0.338  0.73547
AST           1.3765     0.6064   2.270  0.02443 *
Pos_PF       -1.3210     2.7440  -0.481  0.63082
Pos_PG       -6.4342     4.1176  -1.563  0.11995
Pos_SF       -6.6225     3.0456  -2.174  0.03102 *
Pos_SG       -9.3352     3.4545  -2.702  0.00756 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 9.446 on 175 degrees of freedom
Multiple R-squared:  0.7031,    Adjusted R-squared:  0.6895
F-statistic: 51.81 on 8 and 175 DF,  p-value: < 2.2e-16
```

Appendix D: References

Background

Berri, D. J., & Schmidt, M. B. (2006). The wages of wins. Stanford University Press.

Kubatko, J., Oliver, D., Pelton, K., & Rosenbaum, D. T. (2007). A starting point for analyzing basketball statistics. Journal of Quantitative Analysis in Sports, 3(3). https://doi.org/10.2202/1559-0410.1070

NBA. (2023). Collective Bargaining Agreement: Summary of principal terms. https://nbpa.com/cba

Simmons, R., & Berri, D. J. (2011). Mixing the princes and the paupers: Pay and performance in the NBA. Labour Economics, 18(3), 381–388. https://doi.org/10.1016/j.labeco.2010.12.003

Staw, B. M., & Hoang, H. (1995). Sunk costs in the NBA: Why draft order affects playing time and survival in professional basketball. Administrative Science Quarterly, 40(3), 474–494. https://doi.org/10.2307/2393794

Stiroh, K. (2007). Playing for keeps: Pay and performance in the NBA. Economic Inquiry, 45(1), 145–161. https://doi.org/10.1111/j.1465-7295.2006.00002.x

Data Sources

Statista (NBA franchise values): Statista. (2010). Franchise value of National Basketball Association teams in 2010. https://www.statista.com/statistics/193696/franchise-value-of-national-basketball-association-teams-in-2010/

Basketball-Reference (Player contracts): Basketball-Reference.com. (n.d.). NBA player contracts. https://www.basketball-reference.com/contracts/players.html

Basketball-Reference (2025 Per-game stats): Basketball-Reference.com. (2025). NBA per game statistics: 2024–25 season. https://www.basketball-reference.com/leagues/NBA_2025_per_game.html

Additional Help