

# 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?

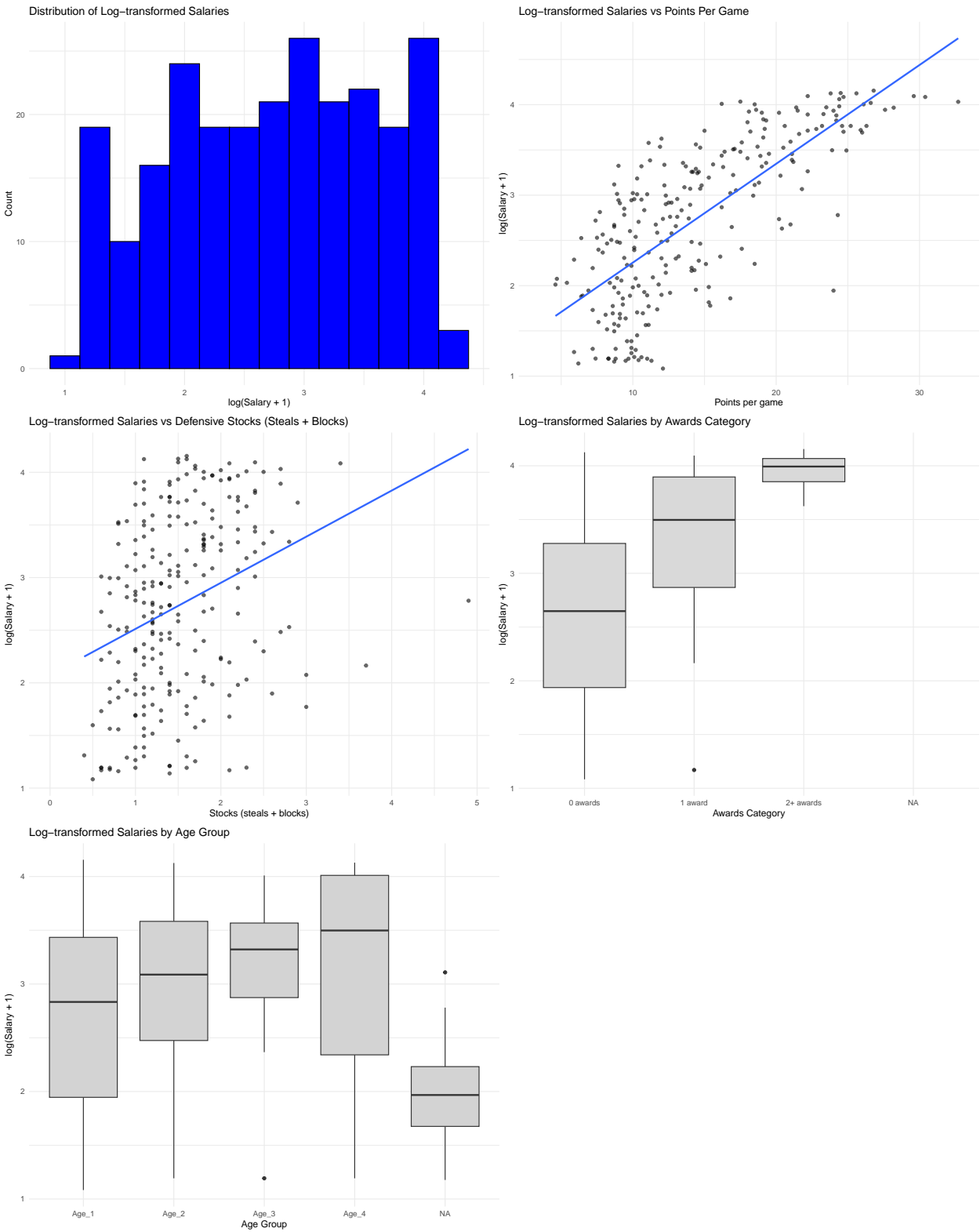
These questions allow us to assess how performance and recognition jointly relate to compensation and how our findings compare with established sports-economics patterns.

## 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 ( $\sqrt{\text{Salary} + 1}$ ). 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. After standardizing player names, we merged all sources so each player occupied one row. Awards were simplified into a three-level categorical variable, and age was grouped into brackets.

# Exploratory Data Analysis



## EDA Summary

Log-transformed salary shows a strong positive linear relationship with scoring: players above ~20 PPG consistently fall in the upper salary range on the log scale, while low-scoring players cluster near the lower end. Defensive performance, measured by stocks (steals + blocks), also has a positive association with salary, though the relationship is noticeably weaker and contains more variance, as shown by the scattered cloud in the plot.

Awards and experience further separate players into clear salary tiers. The boxplots show that players with even one award earn substantially more on the log scale, and those with multiple awards form the highest and most compressed salary group. Age groups display a similar trend: older, more experienced players (particularly those in the 27–31 and 32–34 brackets) show higher median log salaries, reflecting the market value of veteran players. Younger players (Age\_1) display a much wider spread, consistent with rookie deals and early-career variability.

## Methods and Analysis

### Assessing Multicollinearity

We began our analysis by assessing multicollinearity among predictors. From our EDA correlation heatmap, we observed potential overlap between PTS and other offensive statistics such as FTA and MP. We quantified these relationships using pairwise correlations and VIFs. If VIFs fell below standard thresholds (individual < 10, average < 3), we proceeded; otherwise, we used variable screening and stepwise regression to identify a reduced subset of predictors. After selecting a more stable set of variables, we reassessed multicollinearity before model building.

### Building the Model

We will begin 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:

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

We estimated this model with `lm()` and evaluated overall utility using a Global F-Test. Significant predictors—including the interaction—were identified through individual t-tests.

Non-significant predictors were removed, and the updated model was rechecked with another Global F-Test.

Next, we incorporated qualitative predictors and one interaction (Awards  $\times$  Age). We again performed a Global F-Test, followed by t-tests for the interaction and a Nested F-Test for the Awards factor. Only significant categorical effects were retained. No quantitative  $\times$  qualitative interactions were considered. Once the final set of predictors was selected, we confirmed model adequacy with a concluding Global F-Test.

### **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 metrics such as  $R^2$ , Adjusted  $R^2$ , and  $\sqrt{MSE}$ , 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  $R^2$ . This process helped determine whether the model generalized well to new data or required further refinement.

### **Checking Model Assumptions**

We evaluated model assumptions using residual diagnostics. Residual vs. fitted plots, QQ plots, histograms, and lack-of-fit assessments enabled us to check constant variance, normality, independence, and overall fit. We then examined outliers and influential points through Cook's Distance, leverage, Studentized Residuals, and Deleted Studentized Residuals. If any observations exerted undue influence, we considered removing them to improve model integrity.

## Results

### Final Model

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

$$\begin{aligned}\text{Salary} = & \beta_0 + \beta_1 MP + \beta_2 PTS + \beta_3 TRB + \beta_4 AST \\ & + \beta_5 \text{awards}_1 + \beta_6 \text{awards}_{2plus} + \beta_7 \text{Age } 22-26 \\ & + \beta_8 \text{Age } 27-31 + \beta_9 \text{Age } 32-34 + \beta_{10} (Pos_{PF}) \\ & + \beta_{11} (Pos_{PG}) + \beta_{12} (Pos_{SF}) + \beta_{13} (Pos_{SG})\end{aligned}$$

### Model Development

Initial VIF analysis showed severe multicollinearity in the interaction term, so it was removed. Stepwise selection identified six predictors, but t-tests reduced the quantitative model to MP, PTS, TRB, and AST. ANOVA F-tests indicated that adding qualitative predictors (Awards, Age, Position) significantly improved model fit (all  $p < 0.01$ ).

### Model Performance

The final model achieved an adjusted  $R^2 = 0.7681$ , explaining 76.8% of salary variation, with a highly significant overall F-statistic ( $F = 75.39$ ,  $p < 2.2e-16$ ). The residual standard error was \$7.832M. Five-fold cross-validation produced  $RMSE = \$8.052M$  and  $R^2 = 0.761$ , confirming strong out-of-sample performance.

### Predictor Significance

Points per game was the strongest predictor ( $\beta = 1.453$ ,  $p < 2e-16$ ), and assists were also significant ( $\beta = 1.545$ ,  $p = 0.00089$ ). Minutes played and rebounds were not significant in the full model. Awards showed large premiums: one award added \$4.36M ( $p = 0.0544$ ), and two or more added \$11.83M ( $p < 2.8e-06$ ). Players aged 27–31 earned about \$4.19M more than the oldest group ( $p = 0.0537$ ). Position also mattered: point guards earned \$4.67M less ( $p = 0.0472$ ), shooting guards \$5.24M less ( $p = 0.00865$ ), and small forwards trended lower ( $\beta = -3.38$ ,  $p = 0.0667$ ) relative to power forwards.

### Diagnostics

Residual analysis identified observations 86 and 154 as influential points with high leverage and large studentized residuals (exceeding  $\pm 2$ ). These observations were flagged for removal to ensure robust parameter estimates.

## Conclusions

Our final multiple linear regression model explained approximately 76.8% of the variation in NBA player salary, indicating strong predictive usefulness. The most influential quantitative predictors were points per game and assists, both of which showed large and statistically significant positive effects on salary. Award recognition and age group also contributed meaningfully, with players earning one award receiving an estimated \$4.36M salary increase and players with two or more awards receiving nearly \$12M more than otherwise similar players. Mid-career players (ages 27–31) earned significantly higher salaries than the oldest age group, aligning with typical NBA performance peaks.

### Prediction Equation (in millions)

$$\begin{aligned}\widehat{\text{Salary}} = & \beta_0 + \beta_1 MP + \beta_2 PTS + \beta_3 TRB + \beta_4 AST \\ & + \beta_5(\text{awards}_1) + \beta_6(\text{awards}_{2+}) \\ & + \beta_7(\text{Age}_{22-26}) + \beta_8(\text{Age}_{27-31}) \\ & + \beta_9(\text{Age}_{32-34}) \\ & + \beta_{10}(\text{Pos}_{PF}) + \beta_{11}(\text{Pos}_{PG}) \\ & + \beta_{12}(\text{Pos}_{SF}) + \beta_{13}(\text{Pos}_{SG})\end{aligned}$$

### Example Prediction

For a player who averages **24 PTS**, **6 AST**, **7 TRB**, plays **33 minutes**, has **one award**, and is in the **27–31 age group**, the model predicts a salary in the approximate range of **\$30–\$40 million**, depending on position category. This demonstrates that the model produces intuitive and practically meaningful estimates based on real performance indicators.

### Future Improvements

Although the model performed well, several limitations remained. Two influential outliers (observations 86 and 154) may have affected parameter estimates, and salary is influenced by factors not captured in our dataset—such as contract timing, injuries, popularity, or CBA constraints. Including these variables or using mixed-effects approaches could improve realism. Because NBA salaries are highly right-skewed, future work could also explore log-salary models, nonlinear effects, or regularization methods like ridge regression to address remaining variance and multicollinearity. Overall, the model offered a solid foundation for understanding salary drivers while leaving room for deeper economic and behavioral 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
4	Luka Dončić	2TM	50	35.4	0.450	7.9	8.2	7.7	2.2	28.2
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
	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				
4	NA	0	0	51.8379	1	0				
5	3.60	0	1	50.6117	1	0				
6	6.70	0	1	62.7867	1	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	
4	0	0	0	1	0	0	Age_1	0 awards	3.967229	
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	

## 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
-29.7128	-6.6206	0.2972	5.6668	24.8545

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-14.9563	5.0312	-2.973	0.00326	**
MP	0.3855	0.2238	1.722	0.08636	.
PTS	1.7350	0.1872	9.269	< 2e-16	***
TRB	-0.2298	0.4545	-0.505	0.61369	
AST	1.6084	0.5493	2.928	0.00375	**
Pos_PF	-0.4317	2.4306	-0.178	0.85919	
Pos_PG	-7.1756	3.6079	-1.989	0.04787	*
Pos_SF	-4.4441	2.7486	-1.617	0.10723	
Pos_SG	-8.3569	3.0208	-2.766	0.00611	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.464 on 237 degrees of freedom

(24 observations deleted due to missingness)

Multiple R-squared: 0.6913, Adjusted R-squared: 0.6809

F-statistic: 66.34 on 8 and 237 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](https://www.basketball-reference.com/leagues/NBA_2025_per_game.html)

### Additional Help