

WHAT PAYS IN THE NBA?

YOUYU FU, DAVID LEE, RYAN LI

STA302



1. INTRODUCTION

This report investigates the research question: **Which factors, including points per game, assists, total rebounds, age, average views, and position, have the strongest positive or negative impact on NBA player salaries?** Using confidence intervals on linear regression, we analyze these relationships to help NBA teams make informed financial decisions.

2. DATA COLLECTION

The data was procured from kaggle, created for a University project out in Milan. It is appropriate for answering our research question as the dataset contains the relevant aforementioned variables of players from 2016 through to 2019. This dataset is trustworthy, the data is verifiable through the NBA's official statistics website.

5. FINAL MODEL

$$\hat{\text{Salary_transformed}} = 125.278572 + 0.102745 \cdot \text{Age_sq} + 1.505158 \cdot \text{PTS} + 9.694878 \cdot \text{AST} + 7.700196 \cdot \text{TRB}$$

$$+ 5.842177 \cdot \log(\text{mean_views}) + 0.008243 \cdot (\text{Age_sq} \cdot \text{PTS}) + \text{Pos1Effect}$$

Where:

$$\text{Pos1Effect} = \begin{cases} -12.698336 & \text{if Pos1 = PF (Power Forward)} \\ -40.944888 & \text{if Pos1 = PG (Point Guard)} \\ -9.080352 & \text{if Pos1 = SF (Small Forward)} \\ -13.662659 & \text{if Pos1 = SG (Shooting Guard)} \\ 0 & \text{if Pos1 = C (Center; baseline category)} \end{cases}$$

3. PRELIMINARY MODEL

We cleaned and performed EDA on the dataset, and fitted a preliminary model with the aforementioned variables. We also created two interaction terms, age:points and position:mean_views, to give us better insight on a team's decision making process (exp. would younger players valued higher for the same PTS because of growth potential?).

4. DATA TRANSFORMATION

Our preliminary model had issues with patterns in the residual graphs (figure 1), we applied a BOX-COX transformation on salaries at lambda = 0.3, As well as quadratic and log transformations on age and mean views, respectively, which fixed many issues (figure 2) but to make the model more accurate we applied a partial F-test to drop insignificant terms, which allowed us to arrive at our final model.

6. PERFORMING CONFIDENCE INTERVAL

95% confidence interval created with final model

Coefficients	2.5%	97.5%
AST (assists)	5.32682973	14.0629263
TRB (total rebounds)	4.67371144	10.7266814
log_mean_views (wiki clicks)	2.71126309	8.97309121

7. CONCLUSION

We conclude that log_mean_views, AST, and TRB have the strongest positive influence on Salary_transformed, since these intervals have the largest values and don't include zero. We are 95% confident that in every unit increase the predictors above, transformed salary will increase by an amount within that interval.

8. LIMITATIONS

The reliance on transformations, such as Box-Cox and logarithmic adjustments, while necessary to meet assumptions of normality and linearity, complicates interpretability by altering the scale of variables. We also found moderate levels of multicollinearity. For future studies perhaps some transformations can be forgone for the sake of interpretability, so that more confident conclusions can be drawn from the research.

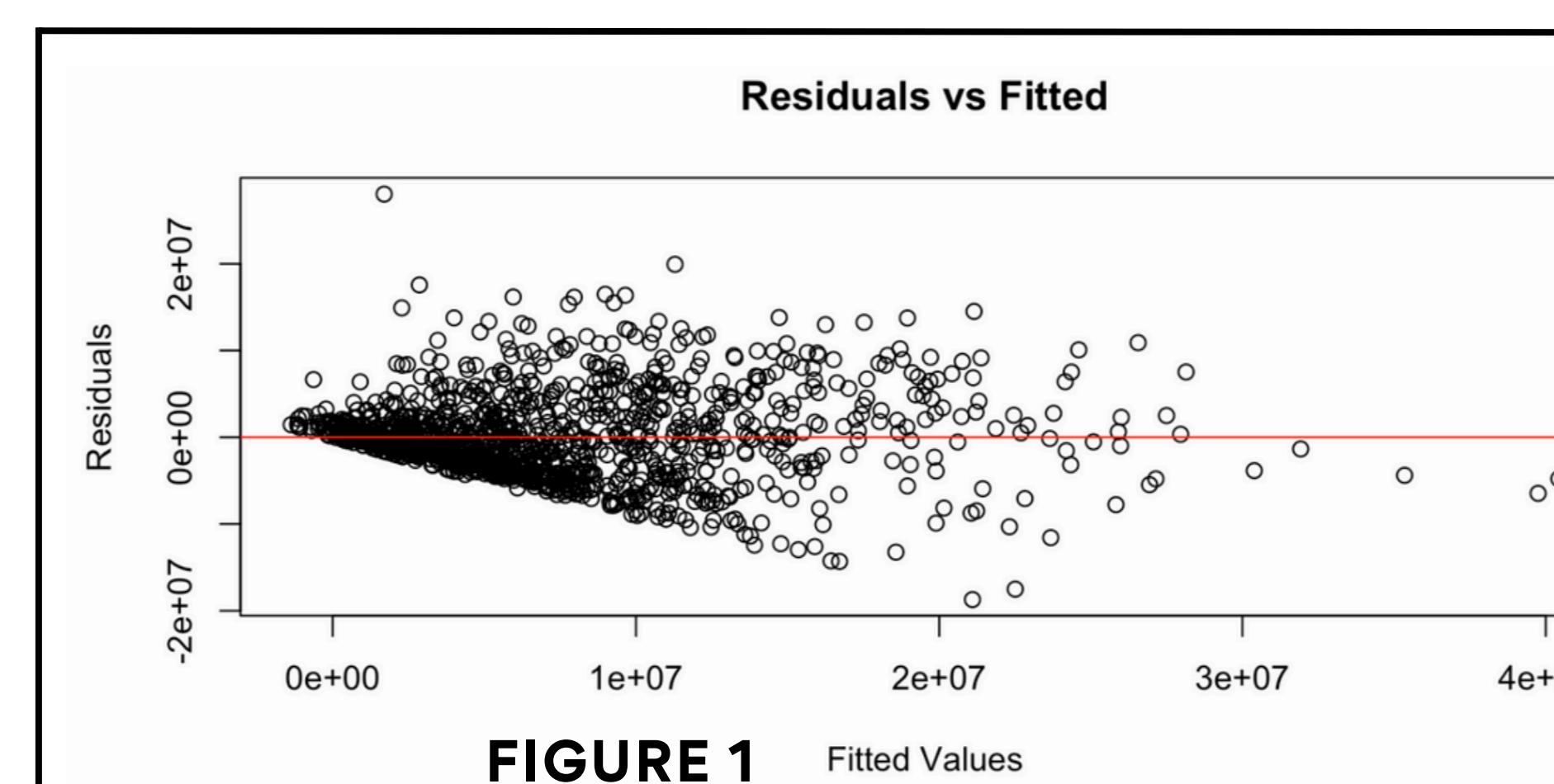


FIGURE 1

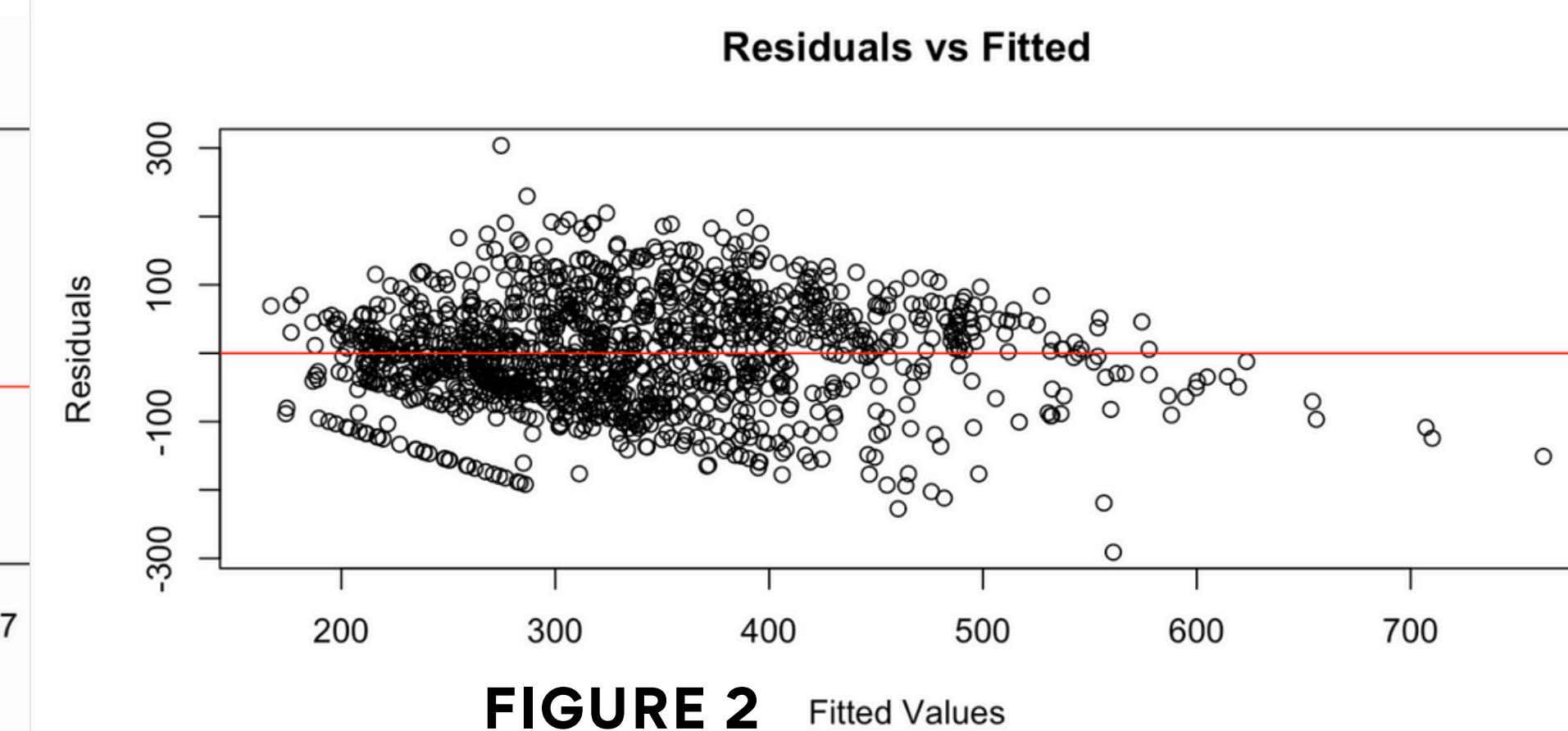


FIGURE 2