David Alvarado

11 November 2024

Bellevue University

**Summary of Exploratory Data Analysis on NBA Player Performance**

The exploratory data analysis (EDA) provided insights into the relationship between NBA player performance metrics, with a focus on Points Per Game (PPG) and Usage Percentage (USG%). The scatter plots highlighted a strong positive correlation, showing that players with higher usage rates often achieve higher scoring outputs. Similarly, Player Efficiency Rating (PER) demonstrated a positive relationship with PPG, confirming that more efficient players generally contribute more to their teams' scoring. However, the Cumulative Distribution Function (CDF) and histogram analyses revealed a right-skewed distribution of PPG, where most players score in lower ranges, with only a small subset of players achieving high PPG values.

One notable oversight in the analysis was the exclusion of variables like True Shooting Percentage (TS%), minutes played, or team context. Including these factors could provide a more well-rounded understanding of scoring efficiency. Additionally, the regression analysis assumed a linear relationship between USG% and PPG. While the model confirmed a statistically significant positive association, the residual diagnostics – such as the Omnibus and Jarque-Bera test – indicated that residuals deviate from normality. This suggests that the linearity assumption may oversimplify the relationship, and other influential factors were likely excluded.

Key challenges included interpreting the regression diagnostics and addressing non-normal residuals. Outliers, such as players with high usage rates but low scoring, raised questions about specific roles or inefficiencies within their teams. Expanding the dataset to include more comprehensive metrics could provide deeper insights and enhance model performance.

In conclusion, the analysis successfully identified significant predictors of scoring, emphasizing the importance of usage and efficiency. However, incorporating additional variables and refining modeling techniques could yield a more nuanced and comprehensive understanding of player performance in the NBA.