The Influence of Age on NBA Player Performance

**Introduction**

This study analyzes the NBA 2024 dataset, which contains comprehensive statistics from the 2023-2024 season of 572 NBA players per 36 minutes played. The dataset contains various statistics tracking all aspects of player performance, ranging from basic statistics such as points, rebounds, and assists, to more advanced metrics such as shooting percentages from different ranges.

Basketball players often experience shifts in their playstyle and performance throughout their careers due to factors such as changing physical strength and endurance, agility, experience, and adaptability. Often, age is the underlying cause of many of these changes. This analysis is motivated by an interest in understanding the statistical relationship between a player's age and performance.

Using the NBA 2024 dataset, this research examines:

1. Scoring differences between younger players (<25 years) and older players (>30 years) through a two-sample t-test
2. Factors influencing Fantasy Points (FP), a composite performance metric, using multiple linear regression.
3. The likelihood of being a high-scoring player based on age, minutes played (MP), and field goal percentage (FG%) using logistic regression.

Our analyses particularly focus on scoring ability and efficiency in various age groups. We mainly analyze variables such as points, rebounds, field goal percentages, and minutes played. By analyzing this data, we aim to find patterns and assess whether a player's age influences these metrics.

Additionally, we examine Fantasy Points (FP), a rating system used in fantasy basketball, which is a game/simulation where participants build their own teams using NBA players and compete with friends; the players' FP determines who wins games.

FP is calculated with this formula:

$$PTS + (1.2 * (ORB + DRB)) + 1.5 * AST + 3 * (STL + BLK) - TOV$$

FP provides a way to determine a player's total performance in a straightforward statistic, weighting different aspects of the game according to their relative impact.

**Methodology**

Before conducting the analyses, we performed exploratory data analysis (EDA) to identify general patterns in the data and gain insights into the variables. Various plots were used to visualize distributions and relationships, and five-number summaries were calculated for key variables to better understand their statistical characteristics.

This study employed three different statistical analyses to examine the relationships between age and performance metrics:

1. **Two-Sample T-Test**

   The first analysis used a two-sample t-test to compare scoring metrics (PTS) between players under 25 years old and over 30 years old. This test helped identify any statistically significant differences in scoring ability between these age groups.

2. **Multiple Linear Regression**

   The second analysis utilized multiple linear regression to examine how age, minutes played (MP), and shooting efficiency (FT% and FG%) influence Fantasy Points (FP), a composite measure of player performance.

3. **Logistic Regression**

The third analysis employed logistic regression to predict whether a player is high-scoring, defined as scoring above the league average. Predictor variables included age, minutes played (MP), and field goal percentage (FG%).

To conduct these analyses, we used several R libraries:

- **ggplot2**: For visualizing the data
- **caret**: To generate a confusion matrix for evaluating the logistic regression model
- **pROC**: TO generate a receiver operating characteristic (ROC) curve and calculate the area under the curve (AUC)

Additionally, the na.omit() function was used to remove rows with missing values, ensuring data integrity. By following this structured methodology, we ensured the accuracy of our findings and that patterns and relationships were clearly identified and easy to interpret.

## Exploratory Data Analysis (EDA)

Our dataset is official player statistics for the 2023 – 2024 season, sourced from basketball-reference.com. It contains data for 572 NBA players, standardized per 36 minutes played.

The primary variables analyzed in relation to age were:

- Points (PTS): Scoring ability.
- Minutes Played (MP): Time spent on the court.
- Rebounds (TRB): Total rebounds.
- Field Goal Percentage (FG%): Shooting efficiency.
- Fantasy Points (FP): Score for player performance.

Analysis of Variables

We analyzed the relationships between age and the medians of the listed statistics. We used median instead of mean to account for star players. By plotting scatter plots of age against these statistics, the following trends were observed:

- PTS, TRB, and FG%: These metrics generally decrease as age increases.
- MP: Peaks around the age of 30, likely reflecting the preference for experienced players on the court during critical moments

Outliers

It is also important to consider outliers such as LeBron James, who isn't representative of the dataset, who has exceptionally high statistics for his age and is the only player at the age of 39. It is also important to note that the distribution of players by age is skewed to the left, meaning there are many more younger players than old in the NBA.

Descriptive Statistics

Five-number summaries and box plots showed the following insights:

- **Age:** The median age is 25, with most players in their early to mid-20s. The first quartile (1Q) is 23, and the third quartile (3Q) is 28.
- **PTS:** The median is 14, with a compact distribution (1Q = 11.1, 3Q = 18).
- **MP:** The median is 1038, with data spanning a broad range due to varying player roles.
- **TRB:** The median is 5.8, with most players between 4.2 (1Q) and 8.2 (3Q).
- **FG%:** The median is 0.4525, with most players falling between 0.4098 and 0.5.

**Here is a glimpse at our data:**

| Player | Age | PTS | MP | TRB | FG% |
|---|---|---|---|---|---|
| Precious Achiuwa | 24 | 12.5 | 1624 | 10.8 | 0.501 |
| Bam Adebayo | 26 | 20.4 | 2416 | 11 | 0.521 |
| Ochai Agbaji | 23 | 10 | 1641 | 4.7 | 0.411 |

**Table 1**: *Sample Data Overview*

<u>Important Observations</u>

Any outliers in the data can likely be attributed to either star players or players with low minutes, thus skewed statistics.

**<u>Analysis #1: Comparison of Scoring Performance by Age Group</u>**

This analysis aimed to determine if there was a significant difference in points scored between younger NBA players (<25 years old) and older players (>30 years old). Using a two-sample t-test, we compared the average points per game (PTS) of both of the groups to determine whether older players or younger players were more effective scorers.

<u>Methodology</u>

To begin, the data was divided into two groups:

1. Younger players: Players under 25 years old.
2. Older players: Players over 30 years old.

The average points per game were calculated for each group:

1. **Younger players:** 14.56 points per game (average).
2. **Older players:** 14.53 points per game (average).

The observed difference in means was only 0.03 points. This suggested minimal variation. A Welch's t-test was then performed because we are assuming that the variances aren't equal. This is because younger players may have more variability in scoring due to inexperience or fluctuating roles, while older players may have consistent performance due to a decline in physical performance or reduced playing time.

Results:

The Welch's t-test results were as follows:

- **T-statistic:** 0.0481
- **P-value:** 0.9617

The p-value of 0.9617 indicates that the difference in points scored per game between the groups is statistically insignificant. The 95% confidence interval for the difference in means ranged from -1.37 to 1.44, showing that we are 95% confident that younger players might score up to 1.44 more points per game or up to 1.37 fewer points per game compared to older players. This interval includes zero, reinforcing the idea that any observed differences are likely due to random variation rather than a true disparity.

Visualization:

A visualization using a boxplot of points per game by age group illustrated the similar distributions of scoring performance between the age groups. These results highlight that age alone does not appear to have a significant impact on scoring ability.

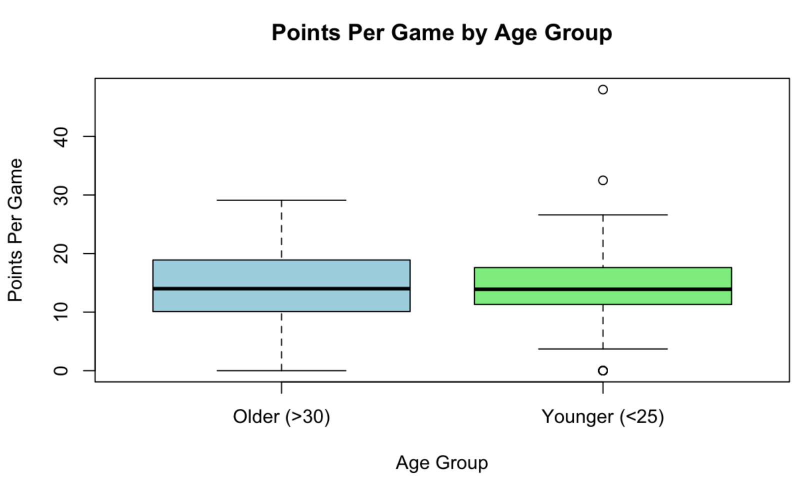


***Figure 1:*** *Points Per Game by Age Group*

Important Observations:

These results indicate that age does not play a significant role in determining scoring performance in the NBA for players under 25 and over 30 years old. However, our analysis excluded players aged 25-30, a group that might show a different scoring pattern and warrants further research.

## Analysis #2: Impact of Key Factors on Fantasy Points

This analysis used multiple linear regression to analyze how different key factors such as age, minutes played, and shooting efficiency impact Fantasy Points (FP). FP are used to measure a player's performance, calculated using points, rebounds, assists, steals, blocks, and turnovers per game. These variables were picked alongside age so we could compare how good of a predictor age is compared to variables that, we assumed, plays a large role in determining FP.

Methodology:

The following variables were selected for the regression analysis:

- **Age:** Examines how a player's experience, matched with potential physical decline, affects overall performance.
- **Minutes Played (MP):** Represents a player's time on the court, indicating their opportunity to contribute to the team.
- **Field Goal Percentage (FG%):** Measures shooting efficiency; higher accuracy generally leads to increased scoring.
- **Free Throw Percentage (FT%):** Indicates shooting efficiency on a smaller scale.

A scatter plot of each variable against one another was used to look for general trends to get a sense of which variables had a strong correlation and which were weak. At a glance, each variable has a correlation with at least moderate strength in terms of relation with Fantasy Points.

Results:

The linear regression model provided the following insights:

- Significant predictors:
  - Minutes Played (MP): $p < 6.28e^{-12}$
  - Field Goal Percentage (FG%): $p < 2e^{-16}$
- Insignificant predictors:
  - Age: $p = 0.260$
  - Free Throw Percentage (FT%): $p > 0.05$

The $R^2$ value for the model was 0.3184, with an adjusted $R^2$ of 0.3131. This means that 31% of the variance in Fantasy Points can be explained by the model.

Variable Importance:

- FG% is the most impactful predictor, with a coefficient of 39.714 (assuming 100% shooting accuracy). Removing FG% as a variable caused the adjusted $R^2$ to drop significantly to 0.1515.
- Age was found to be an insignificant predictor, with a minimal impact on model performance. Removing age reduced the adjusted $R^2$ only slightly, from 0.3131 to 0.3127.

Diagnostic Analysis:

Diagnostic plots were used to evaluate the model's assumptions:

- **Residuals vs. Fitted Values:** Indicated potential issues with heteroscedasticity.
- **Normal Q-Q Plot:** Showed noticeable deviation in the tails, meaning non-normality of residuals.

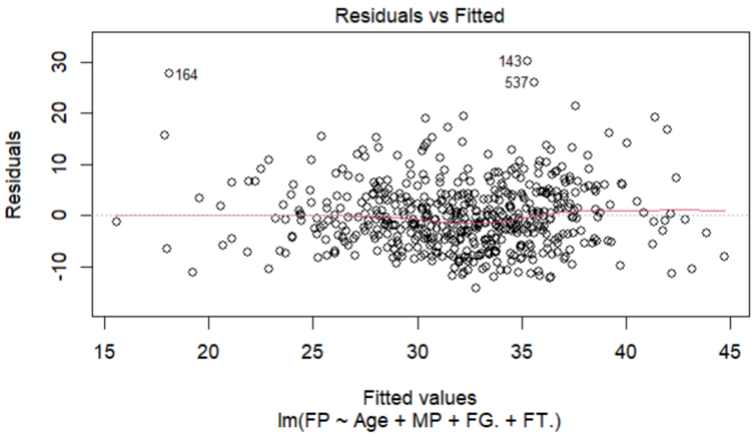


*Figure 2: Residuals vs. Fitted for Fantasy Points*

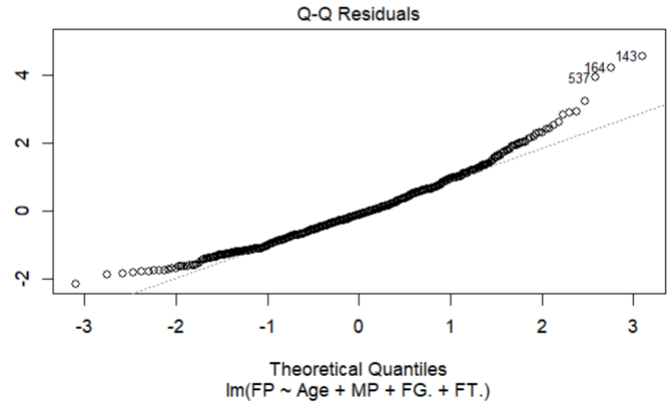


*Figure 3: Q-Q Plot Residuals for Fantasy Points*

These diagnostic results show the model may not fully capture the complexities of age and FT% related to FP. This shows the limitations in the linear regression method for modeling these variables.

**Analysis #3: Predicting High-Scoring Status Using Logistic Regression**

In this analysis, the goal was to determine if a player's age impacts their likelihood of being a high-scoring player in the NBA, with accounting factors such as playtime and efficiency in the shooting. Using a logistic regression model, we analyzed the relationship between age, minutes played (MP), and field goal percentage (FG%) to predict whether a player scores above the league average.

Methodology:

To begin, the explanatory variables selected were hypothesized to influence scoring performance:

- Age: Investigates the trade-off between experience and potential physical decline.
- Minutes Played (MP): Represents playing opportunity and contribution time.
- Field Goal Percentage (FG%): Measures scoring efficiency, which is critical to scoring success.

These variables were used to test whether scoring performance is driven by factors other than age. The logistic regression model predicted the likelihood of being a high-scoring player (above-average points per game).

**Results:**

The logistic regression model outputted following results:

- **Age:** Coefficient = $-0.054$ ($p = 0.0177$). As age increases, the likelihood of being high-scoring decreases slightly, but the effect is minimal.
- **Minutes Played (MP):** Coefficient = $0.00099$ ($p < 0.001$). Players with more playing time are significantly more likely to be high-scoring.
- **Field Goal Percentage (FG%):** Coefficient = $3.459$ ($p < 0.001$). A 1% increase in FG% strongly increases the likelihood of being high-scoring, making it the most impactful predictor.

Model Performance

The model's predictive power was evaluated using an ROC curve and confusion matrix:

- **AUC:** 0.7379, meaning moderate predictive power.
- **Confusion Matrix:**
    - Accuracy: 69.54%
    - Sensitivity: 77.32%
    - Specificity: 60%

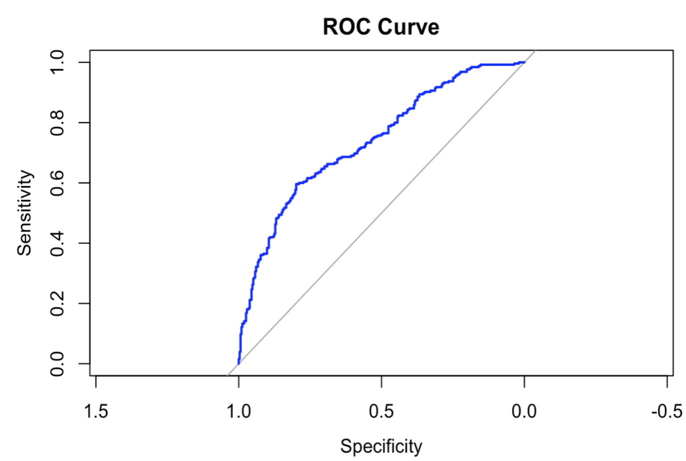


***Figure 4:*** *ROC Curve for High Scoring*

The model is good for random guesses but also indicates that the model can improve with more predictors included.

Predictions:

Predicted probabilities for five new observations showed the following:

- **Observation 5** (highest MP and FG%) had the highest predicted probability of being high-scoring (79%).
- **Observation 4** (lower MP and medium FG%) had the lowest predicted probability (47%).
- Similar probabilities were observed for **Observations 2** and **3**, despite an 8-year age difference, highlighting that MP and FG% are more significant predictors than age.

Visualization

Looking at Figure 5 (High Scoring vs. Age), the points are evenly distributed across ages, with no clear trend, meaning that age has minimal influence on whether a player is high scoring. High-scoring players (1) are slightly more concentrated at younger ages, but there is variability across all age ranges. Next, examining High Scoring vs. MinutesPlayed (MP) in Figure 6, players with more minutes played tend to have a higher likelihood of being high-scoring (concentration near y = 1 increases as minutes played rise). The separation between high-scoring and non-high-scoring players becomes more apparent as minutes played exceed 1500, meaning playing time is a strong predictor of scoring status. Lastly, in Figure

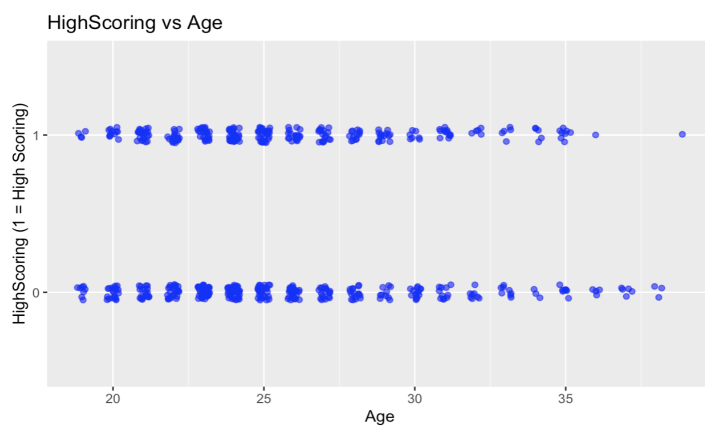


***Figure 5:** High Scoring vs. Age*

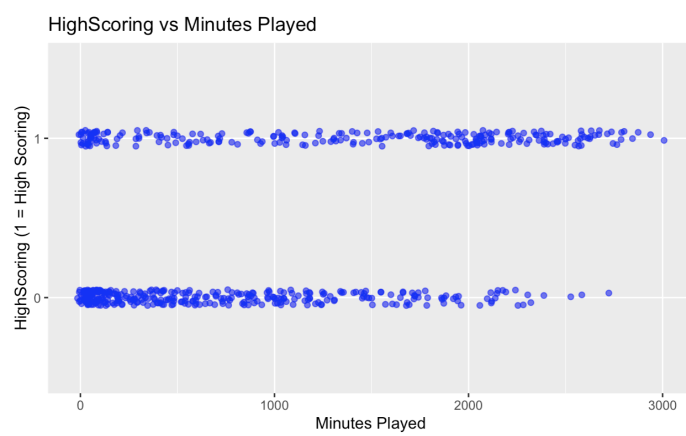


***Figure 6:** High Scoring vs. MP*

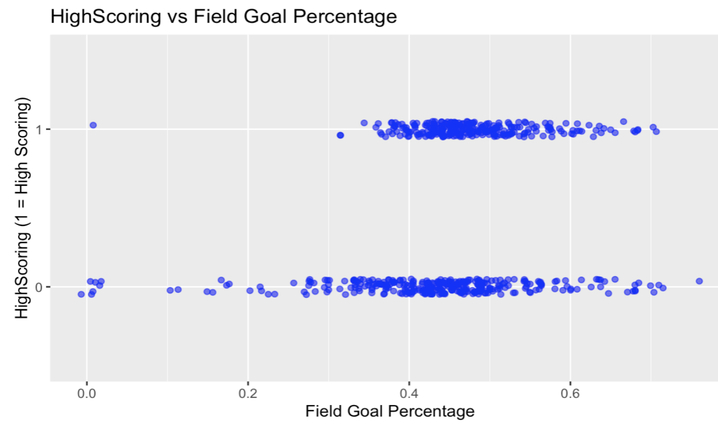


***Figure 7:** High Scoring vs. FG%*

7(High Scoring vs. Field Goal Percentage), players with higher field goal percentages (above 0.5) are more likely to be high-scoring, as seen by the denser concentration near y = 1. Conversely, players with lower field goal percentages (below 0.4) are predominantly non-high-scoring. This means that field goal percentage is a strong predictor of whether a player is high-scoring.

**Conclusion**

The results of our analysis indicate no significant difference in points per game between younger players (<25 years old) and older players (>30 years old), with a p-value of 0.9617 and a 95% confidence interval spanning -1.37 to 1.44. While younger and older players perform similarly in scoring, this does not rule out the influence of other factors such as experience, team role, or skill level. It's worth noting that players aged 25-30 were excluded, and their scoring performance metrics could provide further insights into the relationship between age and scoring in the NBA. Further research could explore more specific ranges of age groups or explore additional variables like points per minute played to better understand the determinants of scoring performance.

The results of our analysis using multiple linear regression to predict Fantasy Points showed that Age was an insignificant predictor of player performance. It yielded a high p-value of 0.260 and had a minimal impact on the adjusted r-squared when removed. On the contrary, we determined that variables like minutes played and field goal percentage play a significant role in predicting FP, as they are better indicators of performance. The diagnostic plots also revealed that there was a flaw in the model suggesting that the linear regression model may not accurately represent the complexities of how age may affect FP. We concluded that age did not play a significant role in predicting FP.

Finally, our logistic regression to predict whether a player is high-scoring (above-average Points) based on their age, playing time (Minutes Played), and shooting efficiency, we can conclude that Minutes Played (MP) and Field Goal Percentage (FG%) are the two significant predictors of high-scoring status. FG% is the most impactful predictor among these three variables. Even though age has a minor negative effect, it is not a strong predictor in comparison to the other two variables. This does not mean that age does not affect a player’s performance, as the statistics show that older players are slightly less likely to be high-scoring. However, the effect of age is not as strong as that of MP and FG%. Thus, we can conclude that while age does matter in an NBA player’s performance, different factors are more significantly important in a player's performance.

Thus, our analysis shows that while age may influence certain aspects of performance, it is not the primary factor in determining scoring ability or overall contribution in the NBA. Instead, variables such as playtime and shooting efficiency are more important factors in a player's performance.

One limitation of this analysis is excluding players aged 25 – 30, a key age group that can show interesting insights into scoring patterns and player performance. Additionally, the regression models assume a linear relationship between variables, which may oversimplify complex interactions. Also, focusing on limited variables, excluding factors like team roles, usage rates, or defensive contributions could provide a more comprehensive understanding. In the future, we will work to include players aged 25 – 30, expand the analysis to larger datasets, and incorporate additional variables to improve model accuracy and predictive power.

**<u>Reference:</u>**

*Basketball Statistics & History of every Team & NBA and WNBA players*. Basketball. (n.d.). https://www.basketball-reference.com/

Brady, B. (2017, November 21). *What happens to NBA players when they age?*. The Harvard Sports Analysis Collective. https://harvardsportsanalysis.org/2017/11/what-happens-to-nba-players-when-they-age/

Salameh, T. (n.d.). *An empirical analysis of prime performing age of NBA players; when do they reach their prime?*. Bryant Digital Repository. https://digitalcommons.bryant.edu/eeb/vol16/iss1/14/