Final Project Report

NBA Advanced/Traditional and Player Salaries

1. Introduction

In the NBA, teams invest significant amounts of money in players based on their perceived contributions to success on the court. However, are these investments truly aligned with player performance? How effectively do traditional and advanced performance measurements predict player salaries? numbers such as points per game, assists per game, and true shooting % provide information about individual contributions, but do these numbers hold equal value between player positions? For example, are assists per game more useful to guards than forwards or centers? Another relevant topic is whether players who are more efficient—as assessed by advanced metrics such as true shooting percentage and win shares—are better compensated. Is there a pattern in which players with greater efficiency metrics earn much more than their peers and how does this differ by position? Finally, do advanced statistics such as offensive and defensive win shares have a higher predictive potential for salary than traditional data like points, rebounds, and assists? Understanding these relationships may offer light on whether NBA teams are using the appropriate parameters when establishing player pay. In this project, I will analyze NBA player salary data scraped from Hoopshype and performance metrics data sourced from Kaggle to investigate the relationship between player performance and salary.

2. Data

This project uses two primary sources of data: NBA player performance metrics obtained from a Kaggle dataset and player salary data scraped from Hoopshype.

2.1 Player Performance Metrics

The Player Performance Metrics dataset contains both traditional and advanced performance metrics for NBA players from the 2019 to 2023 seasons. This dataset was obtained from Kaggle and includes key metrics usually used to evaluate individual performance. Before using the data, I performed several cleaning and transformation steps:

1. Filtered the data to include only the 2019–2023 NBA seasons.
2. Standardized column names to match across datasets (e.g., player_name to player).
3. Aggregated data for players who played for multiple teams in a single season. Specifically, I summed numeric stats such as games played (games) and minutes played (minutes).
4. Calculated weighted averages for stats like points per game, true shooting percentage, and win shares based on the number of games played.

The cleaned Player Performance Metrics dataset is saved in the file *Player_Stats.*csv in the project folder. The code for cleaning and aggregating this data is contained in the script *CombinedStats.ipynb.*

2.2 Player Salary Data

The Player Salary Data was collected by scraping salary information for NBA players from the Hoopshype website for the 2019–2023 NBA seasons. The dataset includes: Player name (player_name), Season year (season), Player salary (salary). To collect the salary data, I wrote a web scraping script using libraries requests and BeautifulSoup. The script iterates over salary pages for each season, extracting player names and their corresponding salaries. The scraped data was stored in a CSV file named Player_Salaries.csv in the project folder. To ensure consistency with the Player Stats dataset, I performed the following steps:

1. Transformed the season column to include only the ending year (e.g., "2019-2020" becomes "2020").
2. Cleaned the player name column by trimming whitespace and standardizing formatting.
3. Removed any duplicate records or rows with missing salary data.

The web scraping script is stored in the file *ProjectScraping.ipynb*

2.3 Combining Player Performance and Salary Data

Both datasets share the columns player (player name) and season, which were used as the keys for merging. I used an inner join to combine the two datasets, ensuring that only players with both performance metrics and salary data were included in the final dataset. This final dataset includes key variables such as games played, minutes, points per game, assists, rebounds, win shares, true shooting percentage, and salary. Minimal cleaning was needed for the performance metrics dataset as it was already well-structured, while the scraping process ensured clean salary values. The resulting merged dataset, saved as *merged_player_stats_salaries.csv,* is clean, consistent, and ready for analysis. A description of each variable is contained in Table 1.

Table 1: Data Dictionary

Field	Type	Description
player	Text	Name of the player
season	Numeric	Season year of the data (e.g., 2020 for the 2019-2020 season)
position	Text	Player's primary position (e.g., Guard, Forward, Center)
games	Numeric	Total games played during the season
minutes	Numeric	Total minutes played during the season
points_per_game	Numeric	Weighted average of points scored per game
assists_per_game	Numeric	Weighted average of assists per game
rebounds_per_game	Numeric	Weighted average of rebounds per game
offensive_win_shares	Numeric	Total offensive win shares for the player
defensive_win_shares	Numeric	Total defensive win shares for the player
true_shooting_percentage	Numeric	Weighted average of true shooting percentage
salary	Numeric	Player's salary in USD for the corresponding season

3. Analysis

3.1 Correlations Between Salary and Performance Metrics by Position

To understand how player performance metrics influence salaries across different positions, I calculated correlation coefficients for key metrics such as points per game, rebounds per game, assists per game, offensive win shares (OWS), defensive win shares (DWS), and true shooting percentage (TS%), and visualized the results using heatmaps for positions like Point Guards (PG) and Centers (C) as seen in Figure 1 and 2. For Point Guards, assists per game (0.68) and points per game (0.71) showed the strongest correlations with salary, reflecting the critical role of scoring and playmaking for guards. Additionally, rebounds per game (0.66) exhibited a notable correlation, suggesting that point guards who contribute beyond traditional scoring and assists are rewarded. Interestingly, advanced metrics like OWS (0.46) and DWS (0.42) had weaker correlations with salary, which indicates that traditional stats like points and assists remain the primary determinants of compensation for guards. For Centers, points per game (0.73) and rebounds per game (0.67) were the strongest predictors of salary, highlighting the importance of scoring and rebounding in evaluating centers' contributions. However, advanced stats such as OWS (0.21) and DWS (0.29) demonstrated much weaker correlations for this position, suggesting that teams rely less on advanced efficiency metrics when determining salaries for centers. Notably, true shooting percentage (TS%) had a weak correlation for both positions (0.31

for PG and 0.14 for C), which implies that while scoring efficiency is valuable, it is not a primary factor in salary determination. The results show that points per game and assists per game are the strongest predictors of salary for most positions, particularly for PGs and SFs. For positions like C and PF, rebounds per game and scoring remain critical. Advanced metrics such as OWS and DWS appear to have weaker correlations, suggesting that basic stats still play a dominant role in salary determination. Hybrid positions show very high correlations, likely due to limited data.

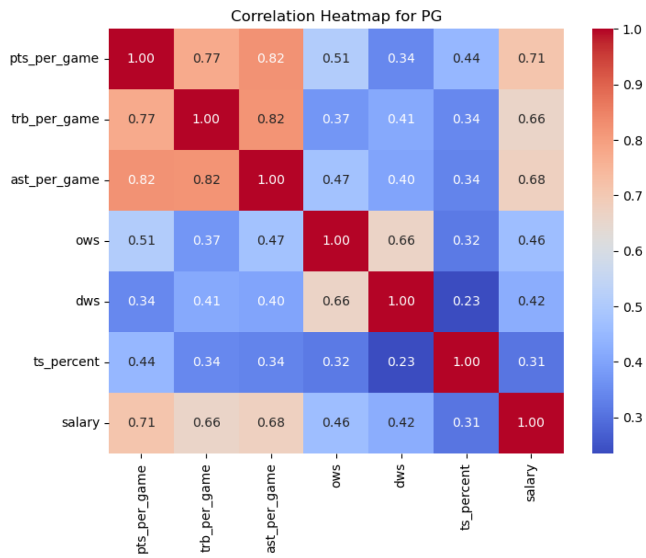

Figure 1

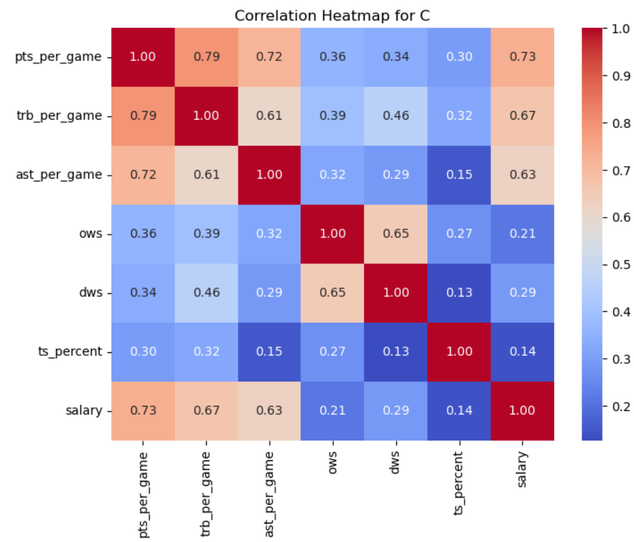

Figure 2

3.2 Efficiency Metrics and Player Compensation

To determine whether players with higher efficiency metrics are compensated more effectively and whether this trend varies by position, I calculated an efficiency score combining offensive win shares (OWS), defensive win shares (DWS), and true shooting percentage (TS%). Players were then grouped into quartiles representing efficiency percentiles: Low, Medium, High, and Very High. The results, visualized in Figure 3, clearly demonstrate a positive relationship between efficiency and salary, with average salaries increasing significantly across percentiles. Players in the Very High Efficiency group earned an average salary of $14.8 million, nearly seven times the average salary of players in the Low Efficiency group, who earned only $2.2 million.

This highlights that NBA teams reward players with higher efficiency metrics substantially more than their less efficient counterparts.

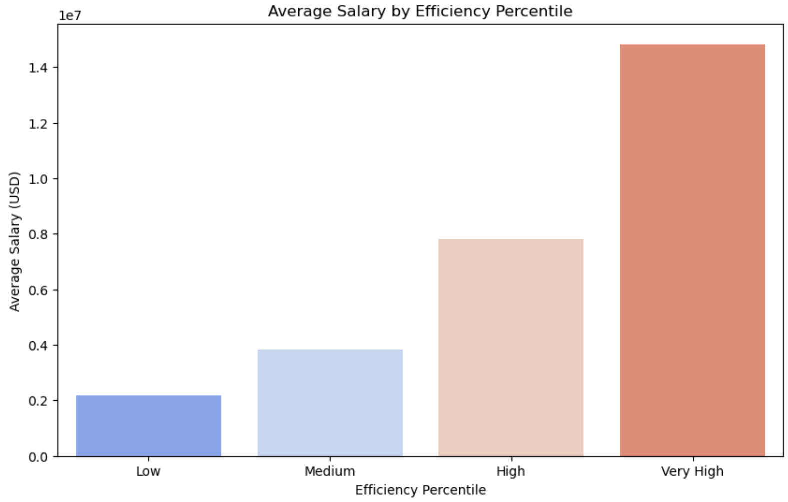

Figure 3

When analyzing these trends by position, further insights emerge. As shown in Figure 4, players in the Very High Efficiency group, particularly positions like Centers (C), Point Guards (PG), and hybrid positions such as C-PF and PG-SG, had the highest salaries, with some reaching upwards of $40 million. This suggests that for top-performing players, positions that emphasize versatility (e.g., hybrid roles) and traditional high-impact roles (e.g., Center and Point Guard) are valued most highly. On the other hand, positions in the Low Efficiency group showed relatively low average salaries across the board, reinforcing the idea that teams are unwilling to invest heavily in players with lower efficiency contributions, regardless of their position.

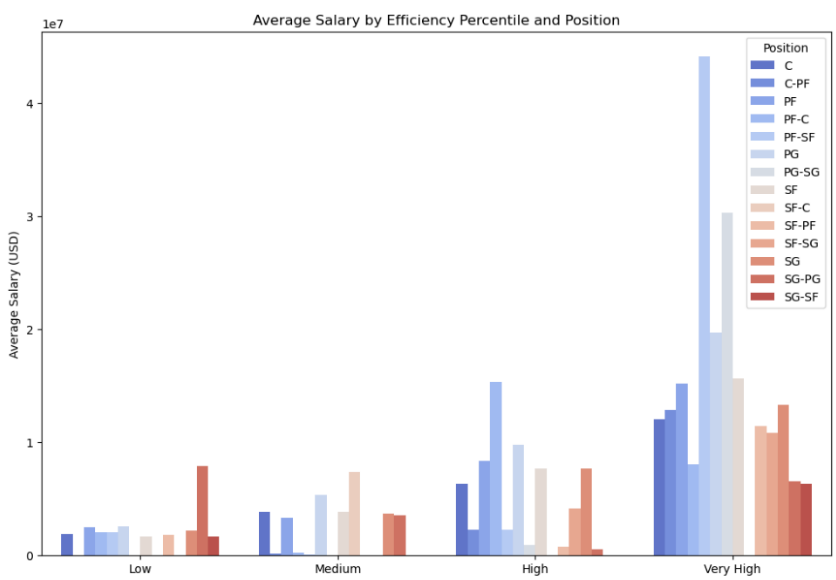

Figure 4

Interestingly, positions such as Shooting Guards (SG) and Small Forwards (SF) showed relatively steady increases in salary across efficiency percentiles, but the salary difference between the High Efficiency and Very High Efficiency groups was less pronounced compared to positions like Center. This could indicate that while efficiency is important, certain positions—especially those where scoring or rebounding is central—benefit more from higher efficiency metrics in terms of salary compensation.

In conclusion, this analysis highlights that players with higher efficiency scores are indeed compensated significantly more, with the effect being strongest for top positions like Centers and Point Guards. Additionally, the data suggests that versatility, as reflected in hybrid positions, further amplifies the relationship between efficiency and compensation. NBA teams appear to prioritize efficiency heavily when determining salaries, particularly for roles where advanced metrics align with traditional performance expectations.

3.3 Predictive Power of Advanced Vs. Traditional Metrics

To determine whether advanced metrics (offensive win shares, defensive win shares, and true shooting percentage) provide stronger predictive power for player salaries compared to traditional per-game statistics (points, rebounds, and assists), three regression models were constructed: one using traditional metrics, one using advanced metrics, and a combined model incorporating both sets of metrics. The results, measured using R-squared values, clearly show that traditional metrics are much better predictors of salary. The Traditional Metrics Model achieved an R-squared value of 0.542, explaining over half of the variation in salaries. By comparison, the Advanced Metrics Model performed poorly, with an R-squared value of only 0.162, indicating that advanced efficiency measures alone do not strongly influence player compensation. When combining both traditional and advanced metrics into a single model, the Combined Metrics Model slightly improved the R-squared value to 0.551, suggesting that advanced metrics add very little predictive value beyond traditional metrics.

The scatterplot in Figure 5, which visualizes actual vs. predicted salaries for the Traditional Metrics Model, further reinforces these findings. While the predictions generally align with the actual salaries, particularly for lower salary ranges, significant deviations appear for higher salaries. This could indicate that factors outside the scope of statistical metrics, such as market demand, player reputation, or team needs, influence the highest salaries in the NBA. Additionally, the clustering of points near the lower range suggests that traditional metrics are more reliable predictors for players with mid-tier salaries but become less accurate for extreme values.

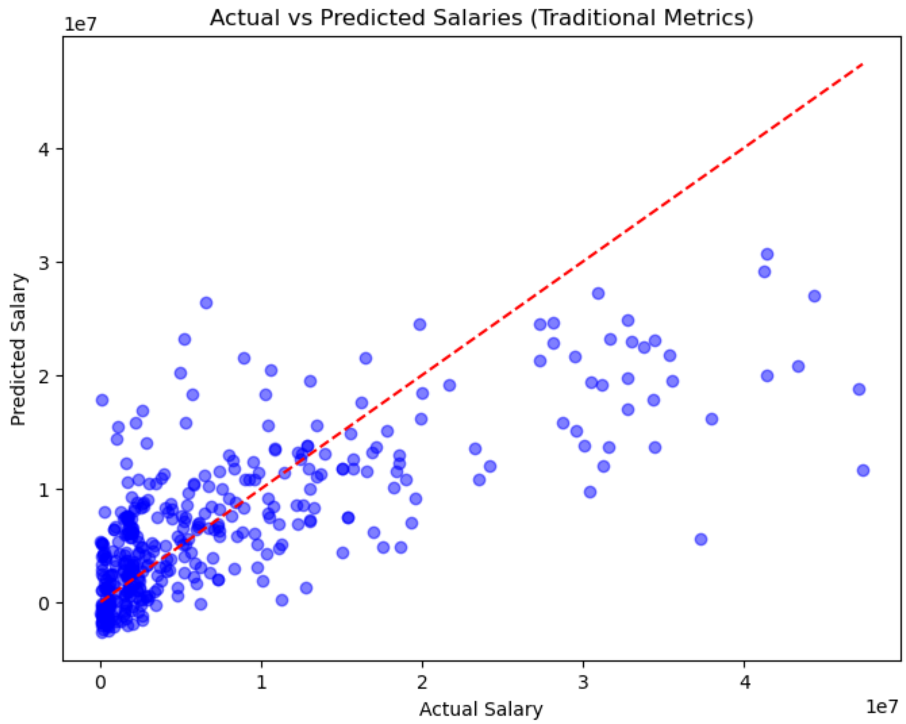

Figure 5

The results suggest that traditional stats, such as points, rebounds, and assists, continue to dominate NBA salary negotiations. Scoring, in particular, remains a primary factor, as shown by the strong performance of the Traditional Metrics Model. Advanced metrics like offensive win shares and defensive win shares, while insightful for measuring player efficiency, appear to play a secondary role in determining player salaries. Teams may rely on these metrics for internal evaluation but seem to default to traditional, highly visible stats when allocating compensation. The combined model's marginal improvement over the traditional model highlights a gap between the use of advanced metrics and their actual impact on salaries. Future studies could explore whether the growing popularity of advanced stats will influence salary determination in the coming years.

4. Conclusion

In this project, I explored the relationship between NBA player salaries and key performance metrics, focusing on three research questions. The analysis utilized traditional per-game statistics (points, rebounds, assists) and advanced metrics (offensive win shares, defensive win shares, and true shooting percentage), combining player performance data with salary data from 2019 to 2023. Below is a summary of the findings from each question:

1. How do correlations between salary and specific performance metrics vary across positions?
   Traditional metrics such as points per game, rebounds per game, and assists per game

showed the strongest correlations with salaries across all positions. For Point Guards (PG), assists per game and points per game were the most significant predictors of salary, highlighting the importance of playmaking and scoring for guards. For Centers (C), points per game and rebounds per game stood out, reflecting the traditional emphasis on scoring and rebounding for big men. Advanced metrics like offensive win shares (OWS) and defensive win shares (DWS) displayed weaker correlations across all positions, indicating that salary decisions are still primarily driven by traditional performance measures.

2. Are players with higher efficiency metrics compensated more effectively, and does this trend differ by position?
   Players with higher combined efficiency scores (based on OWS, DWS, and TS%) were compensated significantly more, as shown by the sharp increase in average salaries across efficiency percentiles. Players in the Very High Efficiency group earned an average of $14.8 million, nearly seven times more than those in the Low Efficiency group. This trend was particularly pronounced for positions like Centers (C), Point Guards (PG), and hybrid roles such as C-PF and PG-SG, where top-performing players commanded salaries exceeding $40 million. However, certain positions, such as Shooting Guards (SG) and Small Forwards (SF), showed more gradual salary increases across efficiency percentiles.

3. Do advanced metrics provide stronger predictive power for salary than traditional per-game stats?
   Regression analysis revealed that traditional metrics such as points, rebounds, and assists are significantly better predictors of salary than advanced metrics. The Traditional Metrics Model achieved an R-squared value of 0.542, explaining over half of the variance in salaries. In contrast, the Advanced Metrics Model performed poorly with an R-squared of 0.162, suggesting that advanced efficiency measures play a limited role in salary determination. The Combined Metrics Model showed a marginal improvement (R-squared = 0.551), indicating that advanced metrics add little predictive power beyond what traditional metrics already provide. The scatterplot in Figure 5 further supports this finding, showing that while predictions align well for mid-tier salaries, they are less accurate for higher salary ranges.

This project highlights that NBA salaries remain heavily influenced by traditional performance statistics, with scoring, assists, and rebounding being the most valued metrics across positions. Advanced metrics like offensive win shares and true shooting percentage, despite their growing popularity, have yet to meaningfully impact salary determination.

There are a few limitations to this analysis. First, the dataset does not account for external factors influencing salaries, such as team-specific needs, player reputation, or marketability.

Additionally, the inclusion of hybrid positions (e.g., PG-SG) may complicate comparisons and could benefit from further categorization. Future work could involve incorporating qualitative factors such as player awards, team success, or endorsements to build a more comprehensive model. It would also be valuable to explore salary trends over longer timeframes to assess how the influence of advanced metrics evolves in the NBA. Finally, analyzing contract structures (e.g., rookie contracts vs. max contracts) could provide deeper insights into salary allocation strategies.