

**NBA Player Stats and Salaries  
REPORT**

GROUP 4

George Cole  
Xavier Watts  
Vibha Naiknavare

University of Notre Dame - Mendoza School of Business

MSBA - Sports Analytics

**Data Analysis with Python- 60230-01**

**2023 Fall**

**Introduction:**

In the dynamic and high-stakes world of professional basketball, the intersection of player performance and compensation stands as a virtual and often debated arena. As the NBA continues to evolve, we need to keep up with what's fair and justified when it comes to player salaries. This report delves into the relationship between NBA player statistics and their corresponding salaries, seeking to unravel the complexities that shape contract negotiations and financial decisions within the league.

**Dataset:**

Our dataset originates from Gigasheet, a comprehensive platform containing thousands of diverse datasets. The dataset under analysis pertains to the NBA Player Salaries for the 2022-2023 season, including all 466 players in the NBA. This dataset comprises an extensive array of approximately 50 variables, including crucial player statistics such as age, player efficiency rating, points per game, and win share percentage, among others. The richness of this dataset provides a valuable opportunity for in-depth analysis, allowing us to explore a multitude of factors influencing playing time and player compensation in professional basketball.

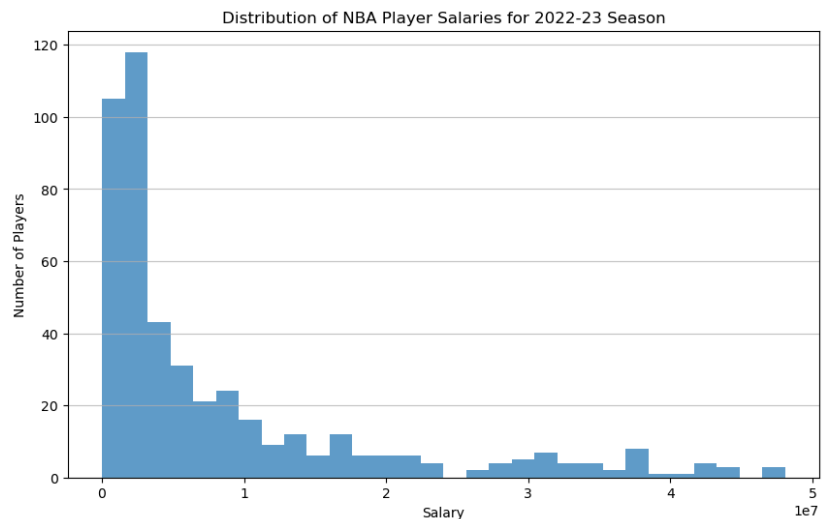
**Data Preparation and Cleaning:**

The dataset was meticulously cleaned before export, eliminating the need for any additional cleaning on our part. To gain a preliminary understanding of the data, we opted to compare various statistics against each player's salary. This approach enabled us to conduct analyses and see the significance of different metrics in influencing how teams determine player compensation. Metrics such as points per game, player efficiency rating, box score plus/minus, and win shares emerged as impactful factors influencing the salaries allocated to players.

**Initial Visualization:** In exploring NBA player statistics and salaries for the 2022-23 season, various data analysis methods were employed to derive meaningful insights. The primary tool for visualizing trends and patterns in the dataset was the Matplotlib library in Python. Methods used were `plt.hist`, `plt.bar()`, and `plt.scatter()` for all visuals below

### 1. Distribution of NBA Salaries:

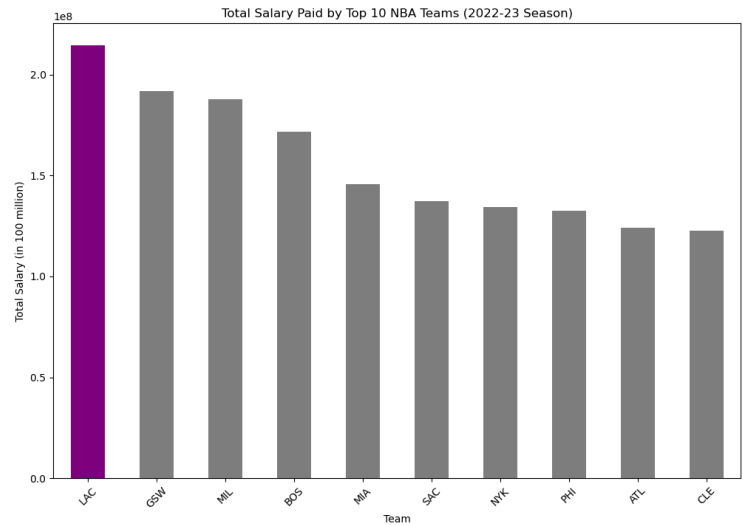
The histogram presented here illustrates the distribution of NBA player salaries for the 2022-23 season. By capturing the frequency of players at various salary ranges the visualization offers a portrait of the income landscape within the league. Notably, the distribution skews to the right, signaling that a substantial proportion of players earn salaries on the lower end, while few selected players significantly earn higher compensation. This clear representation provides an insightful overview of the salary dynamic among NBA players for the season, offering a glimpse into the diversity of the earnings within the professional league.



### 2. Total salary paid by each NBA team for the 2022-23 season - TOP 10 Teams:

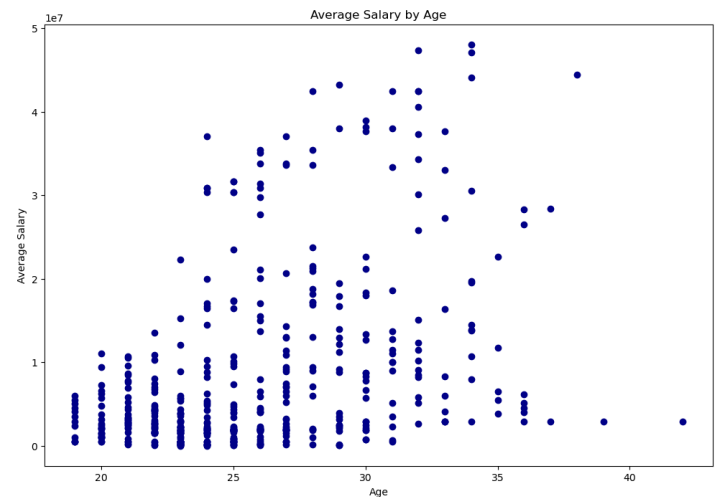
The graph illustrates the total salaries allocated by the top 10 NBA teams for the season. The financial commitments presented for each team showcase the substantial investments made in player salaries. Leading in the list is the Los Angeles Clippers (LAC) with a total salary expenditure of \$214 million, demonstrating their significant financial decision in building a competitive roster. Following closely are the Golden State Warriors (GSW) and the Milwaukee

Bucks (MIL) with total salaries of \$190 and \$187 million respectively. These figures provide valuable insights into the financial landscape of NBA teams, revealing the diverse strategies employed to secure talented players. The data additionally highlights the competitive nature of the league, where teams prioritize substantial financial commitments to enhance their chance to be successful on the court and in season overall.



### 3. Average Salary by Age:

The scatter plot serves as a visual representation of the correlation between the age of NBA players and their respective salaries. Each data point on the plot corresponds to an individual player, with their age plotted along the x-axis and their salary along the y-axis. This visualization offers a



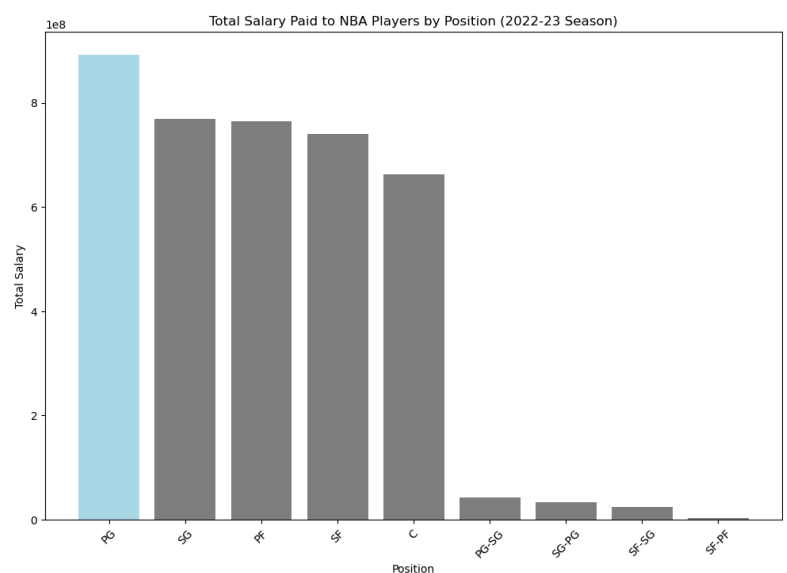
comprehensive overview allowing for observations and analyses of potential trends in the relationship between age and salary. The plot helps us see if there's a common trend in NBA players' salaries as they get older, indicating whether experience is valued. In the plot, the prime age is from 25 to 35 years for players earning a high salary. Notably, there are exceptions like LeBron James at 38, who stand out from the usual patterns making them the outliers on the scatter plot.

### Recommendation 1:

Given that the Clippers and GSW are among the older teams in the league, prioritizing the acquisition of talented, young players with lower salaries can strategically position these teams for the future. Building a strong foundation beyond the current star players, such as Steph Curry and PG/Kawhi Leonard, will be instrumental in sustaining competitiveness. Both teams might want to explore trade packages that include younger, cheaper players for one of the older team's star players. For example, the Warriors might want to explore trading Klay Thompson who is getting to the end of his career, to save some money for a bundle of young players that are less expensive. The Knicks, with promising young talents but potential playoff limitations, should consider adding a seasoned veteran to bolster playoff performance. Across the league, teams can leverage the pool of highly skilled players with lower salaries, represented by talents like Ja Morant or Austin Reeves, to enhance rosters without compromising financial flexibility, ensuring both short-term competitiveness and long-term success.

#### 4. Total Salaries by Position:

The bar chart illustrates how NBA player's total salaries are distributed across different positions during the season. Point Guards (PG) command the highest total salaries, emphasizing the value placed on skilled playmakers and floor leaders. Shooting Guards (SG) closely follow, underscoring the significance of scoring guards in team strategies. Power Forwards (PF), Small Forwards (SF), and Centers (C) hold the third to fifth positions, each indicative of teams allocating resources to players with



diverse skill sets tailored to both offensive and defensive contributions. Dual-position players, such as PG-SG, SG-PG, SF-SG, and SF-PF, showcase the importance teams attribute to versatility in their lineups, offering flexibility and strategic advantages. This breakdown provides a comprehensive overview of positional investments, shedding light on the perceived value and strategic roles assigned to each position within the league.

**Models:**

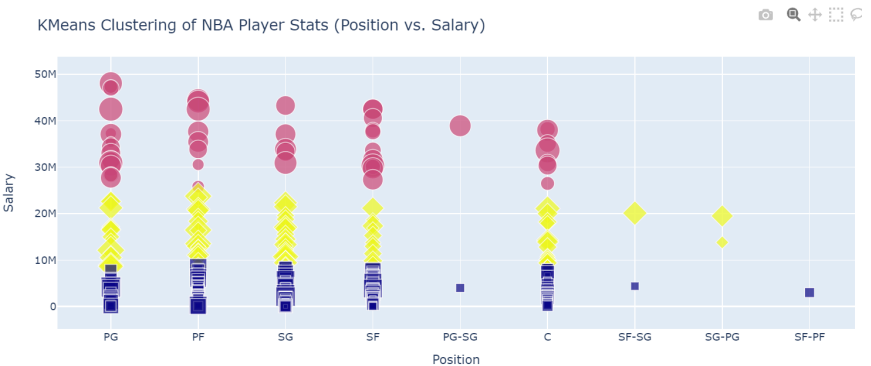
**5. Clustering:**

In the analysis of NBA, Kmeans clustering, implemented through Scikit-Learn, was employed to unveil underlying patterns and groupings within the dataset. Non-numeric columns were excluded to endure a focused numeric dataset, and subsequent clustering resulted in three distinct player groups. These clusters were integrated back into the original dataset and visualized using Plotly Express, producing an insightful scatter plot that color-coded players based on their positions and salaries.

In this visualization, each point represents an NBA player, as its size reflects the player’s impact on the game, combining their

Value Over Replacement Player (VORP) and Point (PTS). Larger points indicate players with higher points and VORPSSs. The color of each point signifies the cluster assigned by the KMeans

algorithm, grouping players with similar statistical profiles. Additionally, symbols corresponding to the cluster labels aid in visually distinguishing these groups, ensuring clarity. This holistic approach visually communicates both individual player impact and groupings based on



statistical similarities, providing a comprehensive view of the players' contributions and relationships within the data.

## Recommendation 2:

Upon analyzing salary distribution by position and our clustering model, a notable trend emerges: centers appear undervalued, drawing the lowest total salaries. The clustering model reinforces this by highlighting that while point guards receive high salaries, many may not outperform replacement players. The success of NBA champions, the Denver Nuggets, with a center as a key player, challenges the prevailing focus on guards. The trend suggests a shift in NBA strategy, emphasizing the importance of skilled centers for a competitive advantage against teams overly invested in guards. This evolution is exemplified by recent successes like the Denver Nuggets.

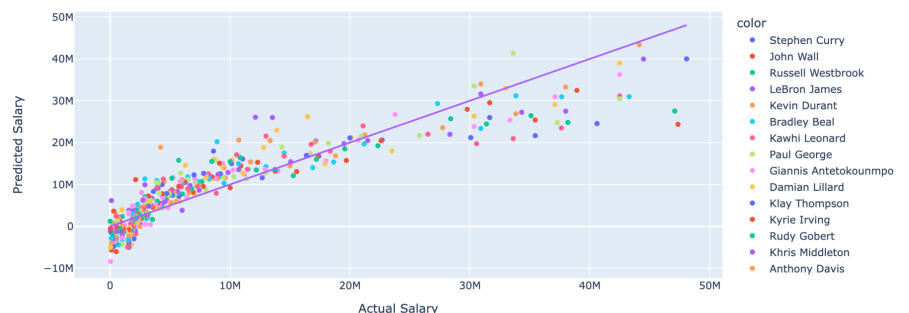
## 6. Salary Regression:

In our pursuit of understanding the fairness of player salaries, we developed a regression model using `Sklearn.linear_model` Python library. Impressively, the model yielded an R-squared value of 0.829, signifying its good

predictive capability in the challenging task of determining a player's deserving salary. Players positioned above the regression line were predicted to be

underpaid. A notable example includes Joel Embiid, the reigning MVP, whom the model suggested deserved a contract worth approximately eight million more than his current one.

This regression analysis effectively captures a distribution of players, distinguishing those who have exceeded expectations from those who may be underperforming in relation to their salaries.



**Recommendation 3:**

When determining player salaries, it is crucial to consider a comprehensive set of variables that holistically reflect a player's performance and contribution to the team. Key metrics such as win share %, plus/minus, and true shooting provide valuable insights into a player's efficiency, impact on team success, and scoring effectiveness. While these metrics offer valuable quantitative measures, it is equally essential to factor in age as one of the most significant variables. Age plays a pivotal role as it can be challenging to predict if a player can consistently maintain a high level of play over the long term. Additionally, unforeseen factors like fluke seasons and injuries can significantly impact a player's performance, emphasizing the need for a holistic approach to player valuation. By considering both on-court metrics and age-related considerations, teams can make more informed decisions when determining player salaries, ensuring a comprehensive evaluation of a player's value to the team.

**Conclusion:**

In summary, our team suggests a strategic team-building approach: prioritize young players on affordable contracts and invest in skilled centers or power forwards. The data reveals that many players earning less than 15 million dollars outperform their salaries, offering cost-effective improvements without straining the salary cap. The undervaluation of centers is evident, emphasizing the potential for a strong return on investment in securing a talented center. We recommend prioritizing the acquisition of solid centers, possibly through drafting, and taking advantage of their talent during rookie contracts. Furthermore, the analysis underscores the impact of unpredictable factors like injuries and age on player progression. This highlights the need for thorough research before committing to max contracts. Teams should exercise diligence in assessing a player's overall performance and potential impact on team success.