Analyzing Factors that Lead to NBA Regular Season Success

Mohamad El-Hajj, Jackson Steed, Victor Gore, Craeg Jethro Infante, Raniel Flores, Danindu Wakista and Mohammed Elmorsy

Computer Science Department, MacEwan University, Edmonton, Canada

elhajjm@macewan.ca, {steedj3, gorev2, infantej2, floresr24,wakistad}@mymacewan.ca, elmorsym@macewan.ca

Keywords: Clustering, K-Means, Decision Trees, Sports Analytics, National Basketball Association.

Abstract: The National Basketball Association (NBA) values regular-season success and acknowledges the crucial role of a team's roster composition in determining overall performance. This study uses machine learning techniques, specifically unsupervised learning clustering and decision tree models, to predict the composition of a winning roster. Our research identified three distinct clusters based on win percentage and the distribution of players across different skill levels. Successful teams typically have more top-tier players and a significant representation of players in the lowest skill level. In contrast, teams that spread their talent across the entire roster are less successful. We have noticed that players with average to above-average skills are notably affected by excessive playing time in the previous game, which leads to decreased performance and potential losses for the team in the next game. Considering the time of year and the gap between games, we recommend prioritizing the rest and recovery of top players, especially in the latter half of the season. It's crucial to ensure that players who are not as skilled as the top players but still make significant contributions to the team maintain consistent performance, especially during the first half of the season. Analyzing height's impact on basketball player performance has revealed practical insights that can empower coaches and management. We found that the shortest and tallest players often perform less than those of average height. Most top performers in the NBA tend to have heights closer to the average. However, for players who frequently operate near the net and encounter numerous rebound opportunities, it is generally preferable to have an average or taller player for slightly enhanced overall performance compared to below-average height players. Teams can use these insights to improve their roster construction and maximize player utilization by coaches from one game to the next. This research provides practical strategies that can be immediately implemented to enhance team performance.

1 INTRODUCTION

The NBA is considered the top professional basketball league globally, with 30 teams, each with 15 talented players. Teams range from smaller market ones like the New Orleans Pelicans to globally celebrated franchises like the Los Angeles Lakers (NBA, 2023), (Burns, 2023).

NBA teams increasingly use advanced analytics to improve their operations and overall performance. This includes identifying players with long-term potential, reducing injury risks, and maintaining consistent performance. Analytics have led to increased revenue and a stronger winning record (Bishop, 2023). Teams analyze player longevity, injury susceptibility, performance at different stages of the season, play styles, and more to evaluate a player's suitability and alignment with the team's vision. This work suggests going beyond standard statistics to using data mining to ensure the selection of effective players and the formation of successful team combinations.

NBA teams must balance building competitive rosters with financial stability. The league enforces a salary cap to create fairness and equal opportunities for all teams. Teams aim to optimize player combinations while managing expenses. Identifying undervalued players through data analysis can help teams secure talented players at lower costs.

Teams closely guard their proprietary advanced analytics systems and methods to gain a competitive edge. Despite this secrecy, certain statistics are widely used by sports media, basketball aficionados, and team management. One such advanced statistic is **win-shares**, which aims to estimate the number of wins each player contributes to their team over a season. It serves as a valuable metric that essentially functions as a scorecard, measuring a player's overall impact on their team's success. Another crucial statis-

El-Hajj, M., Steed, J., Gore, V., Infante, C., Flores, R., Wakista, D. and Elmorsy, M.
Analyzing Factors that Lead to NBA Regular Season Success.
DOI: 10.5220/0013041500003828

In *Proceedings of the 12th International Conference on Sport Sciences Research and Technology Support (icSPORTS 2024)*, pages 83-94
ISBN: 978-989-758-719-1; ISSN: 2184-3201

tical measure is a player's **offensive rating**, which assesses their scoring effectiveness by evaluating their scoring efficiency while considering the number of possessions they utilize. Additionally, (Basketball Reference, 2023a), the **plus/minus** statistic helps in gauging whether a team outscores their opponent or is outscored when a specific player is on the floor. In this work, we use data mining methods to answer four research questions were as follows:

1. What is the **optimal strategy of a roster**: Should a roster concentrate on the best players and supplement with weaker ones, or distribute talent evenly across the team?
2. Can the **timing of the year** or the duration between games impact the performance?
3. Could a **player's performance** in the current game be influenced by their duration of participation in the previous game? What are the different ways in which a player's height may affect their performance?
4. Can we explore the elements that led to the **team's exceptional performance** in a particular season?

Our primary objective is to have a positive impact on coaches and management by offering them pioneering and effective strategies for the development and management of their rosters. We are dedicated to providing valuable insights and implementing practical solutions that will significantly improve the efficiency and effectiveness of their roster management processes.

2 RELATED WORKS

Franks et al. (Franks et al., 2015). evaluated the defensive metrics that influence the outcome of an NBA game. They used a matchup matrix and a spatial regression model to create a new metric. Their analysis involved closely evaluating the number of points a defensive player prevents an opponent from scoring. They then identified the specific location on the court and developed a disruption score to determine where a defender is most likely to stop a shot.

McIntyre et al. (McIntyre et al., 2016) conducted a study on defensive strategies used in response to offensive screens in basketball. They used data mainly from SportVU and carefully analyzed how the defensive team reacts when the offensive team sets a screen. The researchers categorized the screens based on their location on the court and developed four distinct classifications for the defensive tactics used in response to these screens. This detailed analysis better explains how offensive and defensive strategies interact in NBA games.

Gonzalez et al. (Gonzalez et al., 2013) studied how a player's performance changes throughout a season. They examined two main factors: the number of minutes a player spent in a game and their Vertical Jump Power (VJP). The researchers compared the VJP of players in the starting lineup with those who were nonstarters. Their analysis found that starters who played an average of 27.8 ± 6.9 minutes per game tended to increase their VJP compared to nonstarters, who played an average of 11.3 ± 7.0 minutes per game. Specifically, starters increased their VJP by 77.3 ± 78.1 W, while nonstarters increased by 2160.0 ± 151.0 W.

In their study, Drakos et al. (Drakos et al., 2010) thoroughly analyzed NBA injuries over 17 years. They examined the total number of injuries related to the number of games played and calculated the injury rate per thousand athletes. The researchers also looked into the specific body areas affected by these injuries. They attempted to identify potential correlations between injuries and demographic factors such as weight, height, player age, and NBA experience. However, they did not find any significant correlations between these variables.

Berri et al. (Berri et al., 2011) critically examined the reverse-order draft system for amateur players. This system is designed to give weaker teams an advantage by allowing them to secure the first draft picks. The researchers evaluated various factors such as the players' college performance, draft age, years of college basketball experience, player height, and position played. The study focused on college basketball players' performance metrics and their influence on draft day. It found that the number of points scored in college was a significant factor in the draft, but it had minimal correlation with a player's scoring potential in the NBA. This suggests that the current draft system may overlook crucial performance metrics when selecting future star players.

Fearnhead and Taylor (Fearnhead and Taylor, 2011) critically examine the prevalent rating systems used to evaluate an NBA player's performance. They start by looking at the conventional regression model that correlates a player's performance with the number of wins their team achieves. They argue that this model, while helpful, falls short in capturing the player's complete individual performance as it tends to diminish the player's contribution to the team's success. Fearnhead and Taylor have developed a new model that provides a more accurate assessment of a player's abilities by separating their performance into offensive and defensive ratings. This approach allows for a more comprehensive evaluation of a player's skill set, taking into account their contribution to the

team beyond just the number of wins. The method leverages data from multiple seasons to estimate a player's ability in a specific season and measures defensive and offensive ratings separately, combining them to give an overall rating.

Most literature used statistical models to obtain their results. In our work, we introduced machine learning models, such as classifications and clustering, to predict the answers to our questions.

3 DATA

For our analysis, we utilized two main datasets gathered from Kaggle (Kaggle, 2022) and the NBA open data (ESPN, 2023). The first dataset contains 26,652 rows and 21 columns, offering a comprehensive overview of overall NBA game statistics. The second dataset consists of 668,629 rows and 29 columns, providing detailed individual NBA player statistics per game. These available datasets cover the period from 2003 to 2020 and were merged to create our primary dataset. To enhance our analysis, we incorporated win-share, offensive win-share, defensive win-share, season team wins, season team losses from basketball (Basketball Reference, 2023b), and NBA player height data from ESPN (ESPN, 2023).

We analyzed individual player game performances from 2003 to 2020, focusing on the more recent style of play. The dataset contained over 600,000 rows. After data cleaning, around 550,000 rows were left. We used feature selection and creation to retain relevant columns such as season, plus-minus, height, points, assists, rebounds, steals, turnovers, and more. This helped prevent the curse of dimensionality. We excluded irrelevant features like players' nicknames and team abbreviations from our analysis.

As per Basketball Reference, the win-share metric is a player statistic designed to apportion credit for team success among team members (NBA Stuffer, 2023a). Win-shares estimate the number of wins a player contributes to their team through offensive and defensive performances. Offensive win-shares center on a player's offensive contributions, such as scoring points, creating team opportunities, and efficient shooting (Sporting Charts, 2023). Defensive win-shares isolate a player's defensive impact, including blocking shots, stealing the ball, and overall defensive prowess (Sports Lingo, 2023). These metrics assess an individual's collective offensive and defensive performance in a season.

We obtained the team's season wins and losses data from Basketball (Basketball Reference, 2023b), which included the number of wins, losses, team name, and season. Combining the wins and losses provided the team's total games for that season. We then calculated the win percentage by dividing the number of games won by the total games played that season.

$$\text{Win Percentage} = \frac{\text{Number of Games Won}}{\text{Total Games}}$$

We introduced game dates and minutes played as key components to create multiple new dimensions. The game date was crucial for identifying the date of the previous game and calculating the gap between each game. Additionally, we leveraged the game date to categorize the season into early, mid, and late stages, with each stage representing a three-month period. By accessing the previous game date, we were able to extract the minutes played in the preceding game. These additional dimensions enable us to examine how the stage of the season, the duration between games, and the minutes played in the previous game influence an individual's current game performance.

In the world of the NBA, plus-minus (+/-) is a statistical tool used to gauge the point differential when a player is on the court (NBA Stuffer, 2023b). It provides valuable insights into a team's performance with a specific player on the floor. A positive plus-minus value indicates that the player's team outscored the opponents while they were on the court. Conversely, a negative plus-minus value suggests that the opposing team outscored the player's team during their time on the court. Win shares are crucial metrics for a player's season performance, making plus-minus an important measure of a player's game performance. Additionally, points, field goals attempted, free throws attempted, and turnovers are factored in to create an offensive rating metric, offering a comprehensive analysis of a player's offensive game performance. The offensive rating is designed to quantify a player's offensive efficiency and contribution to their team's scoring, often expressed as the number of points a player produces per 100 possessions (Fromal, 2023).

$$\text{Player's Possessions} = \text{Field Goals Attempted} + 0.44 \times \text{Free Throws Attempted} + \text{Turnovers}$$

$$\text{Offensive Rating} = \frac{\text{Points}}{\text{Player's Possessions}} \times 100$$

In our data analysis, we observed a wide range of values within each category. To address this, we opted to use a straightforward discretization method called equal frequency binning for the mentioned values. Equal frequency binning involves dividing a dimension into bins to ensure that each bin contains a

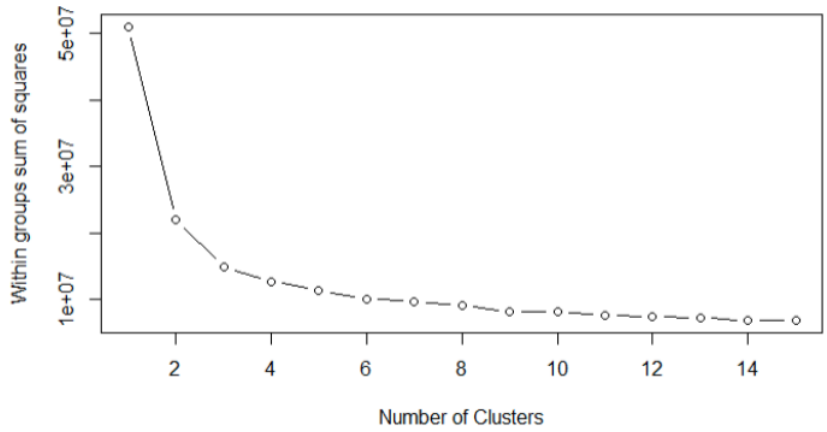

Figure 1: Optimal number of clusters.

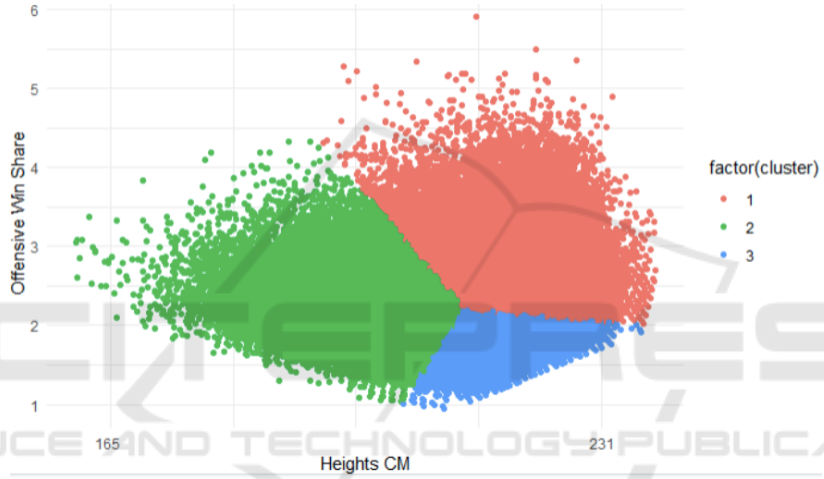

Figure 2: K-means clustering depicting heights and win-shares.

similar frequency of values. In this case, we created six bins for each dimension. This approach ensures that each category or bin will have an equally distributed representation when we run algorithms on the data, thereby enhancing the resilience and accuracy of our analysis.

4 ANALYSIS AND METHODOLOGY

In the next section of our study, we will carefully analyze and discuss the particular techniques utilized to address each of the four research questions we have identified. Subsequently, we will provide an in-depth explanation of the results and valuable insights obtained from our extensive analysis of each of these research inquiries.

4.1 Question One: Optimal Strategy

When deciding whether teams should focus on having a few standout players or distributing talent across various positions, we utilized the k-means clustering technique. This method helps identify data points that are more similar to each other than to others. It involves randomly placing centroids, which represent the center of a cluster, and then assigning data points to clusters. The algorithm then calculates the distance between each cluster's centroid and the specific data point and assigns the point to the nearest centroid. New centroids are computed based on the points belonging to each cluster, and this process is repeated until the centroids stop changing significantly.

We also used the elbow method to determine that three clusters, as in Figure 1, are optimal for analyzing the distribution of player types from each team and their respective win percentages. This strategic insight allows us to understand the balance between

superstar players and talent spread across many positions. Before this analysis, all player features had been categorized into bins, with bin 1 denoting the lowest rating and bin 6 representing the highest. Subsequently, we tabulated the number of players at each level for every team and year, using these figures as the basis for the clustering features.

After running the k-means algorithm on the dataset, we observed the formation of three well-defined clusters, as illustrated in Figure 2. These clusters exhibit noticeable similarities in their features, providing us with crucial insights into the characteristics required for constructing a successful roster.

4.2 Question Two: Timing of the Year

Decision trees are used to evaluate how specific factors impact an NBA player's performance. This involves categorizing input data into different classes using a classification method. Classification decision trees make decisions based on the features of the data and create rules to assign each instance to a specific class (Raj, 2023). The key factors being examined include parts of the season, the number of days between games, and the player's previous and current game minutes. Other features, such as the winning team, were included to provide additional insight. The decision trees use plus-minus and offensive ratings as target variables to determine if the specified features influence the player's performance. Since this is a classification problem, each feature was divided into six equal-frequency bins to ensure the readability of the decision tree outcomes. Before running the algorithm, the dataset is split into a training set for building the tree and a testing set for evaluating performance, with a random seeding and a 0.7 ratio. The classification decision tree is applied to different aspects of the data. First, the data is examined as a whole, and then each of the six bins in win-share, offensive win-share, and defensive win-share are analyzed to compare differences between different player levels. This results in 38 decision trees: 19 targeting plus-minus and 19 targeting offensive ratings.

The study aimed to analyze how a gap influences a player's performance across different points in a season and to determine if this influence varies based on the player's skill level. Players in consecutive win share brackets were combined to form three distinct categories to simplify the findings.

The analysis utilized the f-regression function (Pedregosa et al., 2011), incorporating the target plus-minus to refine the assessment. This function evaluates the correlation between each regressor and the target variable, converting these correlations into F-scores. These scores measure the degree of linear dependency between each regressor and the target, thus aiding in identifying the most predictive features of the outcome.

4.3 Question Three: Player's Performance

We conducted an in-depth analysis of various performance metrics to investigate the connection between a player's height and their performance on the basketball court. One of the metrics we found to be particularly useful was offensive win-shares, which take into account a variety of offensive statistics. As we sought to develop a comprehensive performance statistic, we initially considered multiple factors but eventually honed in on the relationship between player height and win-shares as the key components.

In narrowing our focus to player heights and offensive win-shares, we employed the elbow method as a crucial step in our analysis. This meticulous approach allowed us to determine that the optimal number of clusters was 3, laying the groundwork for implementing a k-means clustering algorithm on the data. This reaffirmed the precision and rigour of our analysis.

Upon applying the k-means clustering algorithm, we uncovered three distinct clusters that encapsulated player heights ranging from 165cm to 231cm. By dividing win-shares into six bins, we could visually depict player performance across different height categories described later

This led us to delve deeper into the relationship between a player's height and their ability to secure rebounds. To analyze this, we sorted the players into six performance tiers and then further categorized them into three height groups. H1 represents the tallest players, H2 consists of players of average NBA height, and H3 encompasses the shortest players.

4.4 Question Four: Team's Exceptional Performance

During our analysis, we delved into the factors contributing to exceptional athletic performances, focusing on the remarkable success of the 2015-2016 Warriors team. Notably, the Warriors concluded the regular season with a historic 73–9 record, surpassing the previous record of 72–10 established by the 1995–1996 Chicago Bulls. This achievement solidified their position with the best regular-season record in the history of the NBA.

To conduct a comprehensive analysis of the Golden State Warriors' performance during the 2015-

2016 NBA season, we implemented a thorough methodology. Our approach commenced by meticulously filtering the dataset to exclusively focus on the 2015-2016 season. Subsequently, we meticulously gathered and organized detailed statistical information, including but not limited to points per game and three-point shooting percentages, for each individual player on the team. Given that each entry in the dataset corresponded to a player's statistical contribution to a specific game on a specific date, we conscientiously compiled multiple entries for each player to ensure an accurate representation of their performance throughout the season.

We gathered detailed data for each player, including their points per game and three-point percentages. This allowed us to analyze their performance throughout the season thoroughly. We expanded our analysis to include the Cleveland Cavaliers, also known as the Cavs, a professional basketball team based in Cleveland. Between 2015 and 2018, the Cavaliers faced the Golden State Warriors in four consecutive NBA Finals, igniting a fierce rivalry and creating one of the most memorable matchups in modern NBA history. For the Cavaliers, we collected and examined their corresponding performance metrics. By using a similar methodology, we calculated the Cavaliers' mean points per game and three-point percentages, enabling a comprehensive comparison between the two teams' performance. This comprehensive approach provided valuable insights into the factors contributing to each team's success.

5 RESULTS AND DISCUSSION

In this section, we will thoroughly examine the results for each research question.

5.1 Question 1: Optimal Strategy

In our upcoming discussion, we will thoroughly analyze the three primary clusters identified during the initial cluster analysis. Our focus will be primarily on examining the mean values for each feature. We will also provide some insights based on the analysis of median values, albeit to a lesser extent.

In Cluster 1 in Figure 3, teams experienced the least success, with an average win percentage of 1.9, equivalent to roughly 27-55 in the regular season. These teams had the lowest number of top-end players, averaging 1.02, and the highest number of low-end players, averaging 5. Their struggle to win is attributed to a need for more high-end talent and an abundance of low-end players.

The most compelling analysis arises from the comparison of clusters 2 and 3. Cluster 2 has a diverse distribution of talent, with teams possessing between 2.06 and 3.50 players in bins 1-5 and 1.53 players in bin 6. Their win percentage is approximately 0.379, translating to a 31-51 record. This group displays marginal improvement over Cluster 1. Conversely, Cluster 3 showcases a top-heavy lineup, with teams having 3.21 players in bin 6, 2.45 in bin 5, fewer than two players in bins 2-4, and 2.71 in bin 1. Their regular season record is roughly 0.622, equivalent to 51-31. Although teams in Cluster 3 have more players in bins 5 and 6, they also have more in Bin 1 compared to the Cluster 2 teams. An analysis of win percentage makes it clear that the more successful teams are those in Cluster 3.

A particularly interesting statistic reveals that when there are three players in bin 6 and 2 players in bin 5, successful teams in cluster 3 can assemble a lineup of 5 players who are well above average, ensuring a cohesive team with no weak links on the floor. On the other hand, teams in cluster 2 cannot achieve this and are more likely to field a lineup where at least one of the five players on the floor is only slightly above average or even below average. Additionally, teams with lower win percentages need more superstar players to rely on during crucial game moments. At times, all a team needs is a brief period where their best players completely take over a game, and the more superstar players a team has, the more likely they are to accomplish this.

Based on our research, it is recommended that teams prioritize acquiring top-tier talent rather than focusing on the depth of their roster or evenly distributing talent. This is particularly relevant in the NBA, which is characterized as a superstar-driven league. Our findings indicate that teams with a few dominant superstars tend to achieve greater success and have a higher regular-season win percentage. In contrast, those lacking a superstar player often need help to keep up with the competition.

5.2 Question 2: Timing of the Year

In this study, we analyzed three statistics: Defensive Rebounds (DREB), Rebounds (REB), and Offensive Rebounds (OREB). We categorized their values into 6 bins, as shown in Figure 4. In our analysis, regardless of the win-share category, each bin displayed a consistent pattern in the decision trees we later discuss, with offensive rating as the target classifier.

We found that players who were on the court for bin 1 or bin 2 minutes (equivalent to 20 minutes or less, considered a low amount of time) tended to have

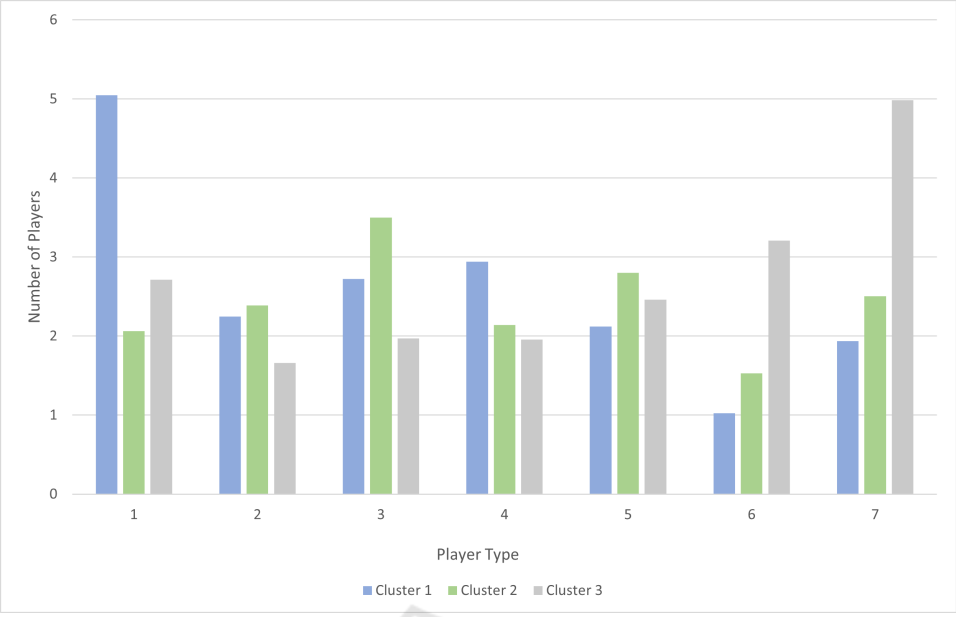

Figure 3: Number of players of each type and win percentage for teams in each cluster.

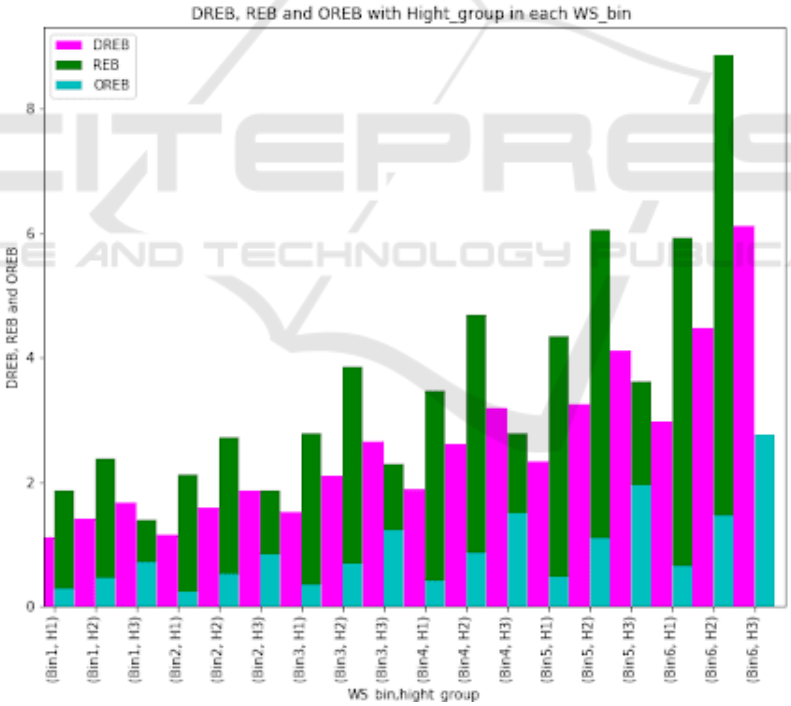

Figure 4: Defensive Rebounds, Rebounds, and Offensive rebounds used with Win share bins.

a bin 1 or bin 2 offensive rating, ranging from 0 to 70. This outcome was expected, as offensive rating is heavily influenced by the points scored by an NBA player. Limited time on the court leads to fewer opportunities to contribute offensively.

When we change the target classifier to plus-minus, we observe different outcomes. The results are affected by whether the team wins or loses. Generally, when a player's team loses, their plus-minus is between bins 1 to 3, regardless of their win-share variation. On the other hand, if the team wins, the plus-minus is either bin 5 or bin 6, indicating a high overall performance if the player played 20 minutes or more. These patterns align with our expectations,

suggesting that a player's overall performance is positively influenced by the team's victory and significant playing time, showcasing their consistent contribution to the team's success.

These trends are generally expected. When we look at plus-minus, we focus on players who fall into the fourth category for each win-share category. These players are average or above average and have significantly contributed to their team's success during the season. They can be described as "role players" or good players but are not considered "all-star" caliber players. These players either support high-level talent in a roster by assisting them when they are on the court together or by holding down the team when their all-star players are not on the court.

When analyzing their decision trees, the number of minutes played in the previous game becomes an important feature in the player's plus-minus. If the team lost and the player's previous game minutes were over 25, the player's current game performance had a plus-minus in the first category, which is extremely low. However, if their minutes were less than 25, their current game performance had a plus-minus in the second category, which is still low but not as bad. The difference in plus-minus is small; however, we can see that the excessive use of role players in the previous game not only negatively affects their overall performance in the current game but may also cause their team's loss. This suggests that role players should be used cautiously if their previous game was strenuous.

In the early part of the season, players in groups 1 and 2 are most affected by the gap between games, while the impact on the rest of the players is minimal. However, as we move into the later part of the season, players in groups 1 and 2 still experience an impact on their +/-. This impact is reduced compared to the early part of the season. Once again, players in the middle of the pack do not experience a major impact, while top players see a significant impact on their performance due to the gaps between games.

The results are shown in Figure, 5, it is worth noting that the accuracy of these decision trees ranges from 0.3 to 0.5, representing extremely low accuracy. This could be attributed to the extensive use of equal-frequency binning. Increasing the number of bins from 6 might lead to more accurate results.

As the season begins, players in the lower skill level groups, specifically bins 1 and 2, are most affected by the extended breaks between games. This can have a significant impact on their performance and readiness. However, as the season advances, the influence on players in these skill-level groups gradually decreases. Meanwhile, players in the intermediate skill level range continue to encounter minimal impact from the gaps between games, while the top-tier players are notably affected by the extended breaks, potentially impacting their momentum and form.

Upon analyzing the impact of the timing of the season and the duration between games on player performance, the findings are as follows: Figure 6. In the early stages of the season, underperforming players are notably more affected by longer breaks between games. As the season progresses, top-performing players are increasingly impacted by the duration of breaks between games, while mid-range players tend to maintain consistent performance regardless of the gap. From a strategic perspective, top players should prioritize rest at the season's commencement and reduce rest as the season advances. Moreover, due to their versatility, role players can be utilized more frequently. Lastly, players ranked at the bottom in terms of performance would benefit from consistent game time, particularly at the season's onset.

5.3 Question 3: Player's Performance

After analyzing the k-means clustering plot, we discovered some fascinating results. Previous studies, such as the one by Berri et al. (Berri et al., 2011), have indicated that height plays a significant role in the selection of amateur players in drafts. However, our plot unveiled an intriguing pattern. The players were effectively categorized into three clusters based on height, with distinct groups for shorter, medium-height, and taller players.

The analysis of the players' heights in relation to their performance yielded intriguing findings. Upon closer examination, it was noted that the shortest players tended to exhibit lower offensive win-shares, suggesting a diminished level of performance. However, this trend was not confined to shorter players, as taller players also displayed a similar pattern. Interestingly, a performance peak was identified within the medium-height range, followed by a decline among the taller players. These observations point to the possibility that optimal performance may not necessarily correlate with extreme heights, but rather lie somewhere within the middle range of heights.

Upon analyzing the height-rebounds graph, a clear pattern emerges indicating that players with higher overall performance also excel in rebounds, in line with expectations. Notably, players in the average height category (H2) exhibit the most impressive rebound performance within each group. Consistently, a trend is evident wherein shorter players tend to underperform compared to their average or tall counter-

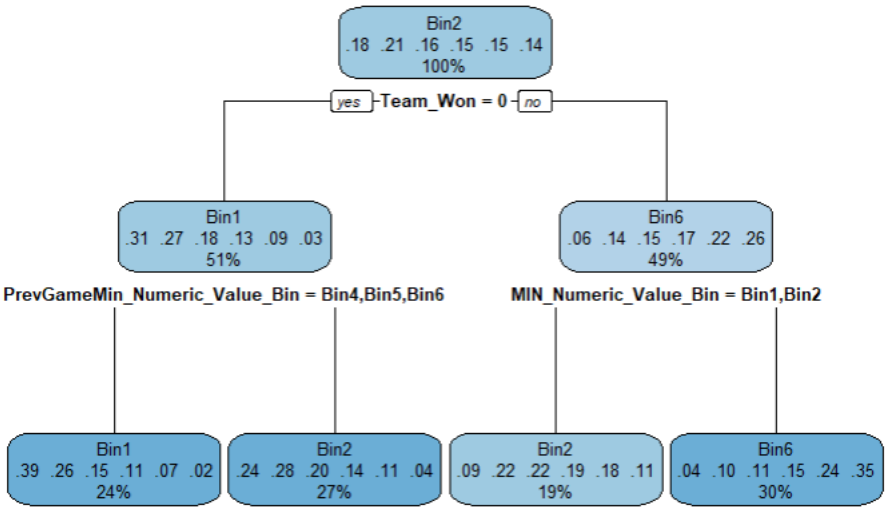

Figure 5: Decision tree with plus-minus as target classifiers in bin 4 of the win-share categories.

parts across offensive and defensive rebounds within the same or lower performance groups. In summary, it can be inferred that an average-height player is well-suited for positions such as Center or Power Forward, although locating such players may not always be feasible. In such instances, sacrificing some performance in favour of added height can lead to favourable outcomes. However, it is imperative to stress that the primary focus should not solely revolve around acquiring the tallest players; prioritizing the best performers remains crucial. In scenarios where a choice must be made between a shorter top performer and a slightly less skilled player of average or above-average height, the latter should take precedence.

5.4 Question 4: Team's Exceptional Performance

In the 2015-2016 season, the Golden State Warriors finished first in the Western Conference with an unprecedented 73-9 record, surpassing the previous record set by the 72-10 Chicago Bulls led by Michael Jordan. The Cleveland Cavaliers finished first in their Eastern Conference with a 57-25 record. In our analysis of the most correlated factors contributing to exceptional achievements, we found that two statistics played significant roles. These factors are the ability to score high points per game and the ability to score many 3-pointers. These two aspects emerged as strong indicators of exceptional performance and were closely linked to achieving outstanding results.

When examining the statistical data of the Cleveland Cavaliers and the Golden State Warriors, it is evident from Figure 7 that the points per game graph illustrates the significant scoring advantage of the Golden State stars, Stephen Curry and Klay Thompson, over the Cavaliers' stars, LeBron James and Kyrie Irving. While the remaining players on both teams make valuable contributions to their respective performances, it is noteworthy that Stephen Curry's exceptional average of 30.1 points per game stands out as a rare achievement in the NBA, playing a pivotal role in the Warriors' historic season. Our analysis indicates a strong correlation between the presence of high-scoring players and the attainment of this remarkable achievement.

The second highly influential factor contributing to exceptional achievements is the 3-point score percentage. In Figure 7.B, a significant disparity between the two teams is evident, emphasizing one of the primary reasons for Golden State's formidable performance. Steph Curry and Klay Thompson's shooting percentages far surpass those of the Cavaliers' starting lineup (excluding centers due to limited data near the basket). Furthermore, Harrison Barnes and Draymond Green demonstrate superior 3-point shooting percentages compared to Lebron James, Kyrie Irving, and Kevin Love. It's noteworthy that although the Cavaliers' 3-point percentages were considered good, the Warriors' dominance completely overshadowed them.

6 LIMITATIONS AND FUTURE WORK

As we move forward, we will focus on a comprehensive analysis of NBA postseason games. It's important to note that our dataset is primarily focused on the 82-game regular season, which means our insights

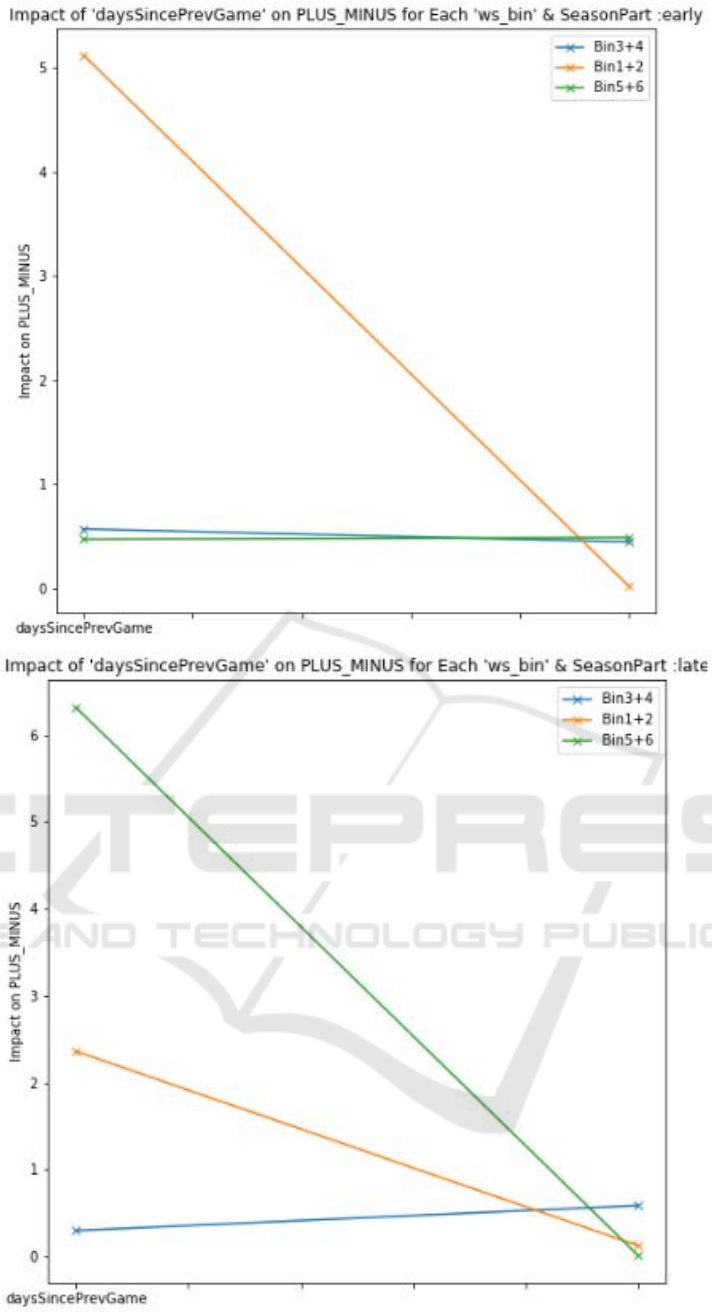

Figure 6: Fregression function visualization with specific bins correlated together.

into team performance during the playoffs will be somewhat limited. However, the best-of-seven playoff series format provides a unique opportunity for in-depth analysis, offering a more streamlined approach compared to the extensive data entries from the regular season. Moreover, we can delve into a thorough examination of individual player performances, aiming to gain insights into their influence on their respective teams, especially those who made significant contributions to their teams' advancement in the playoffs or those who were eliminated early.

7 CONCLUSION

Our analysis of the factors influencing NBA regular season performance shows that a team's roster composition significantly affects its success. We've iden-

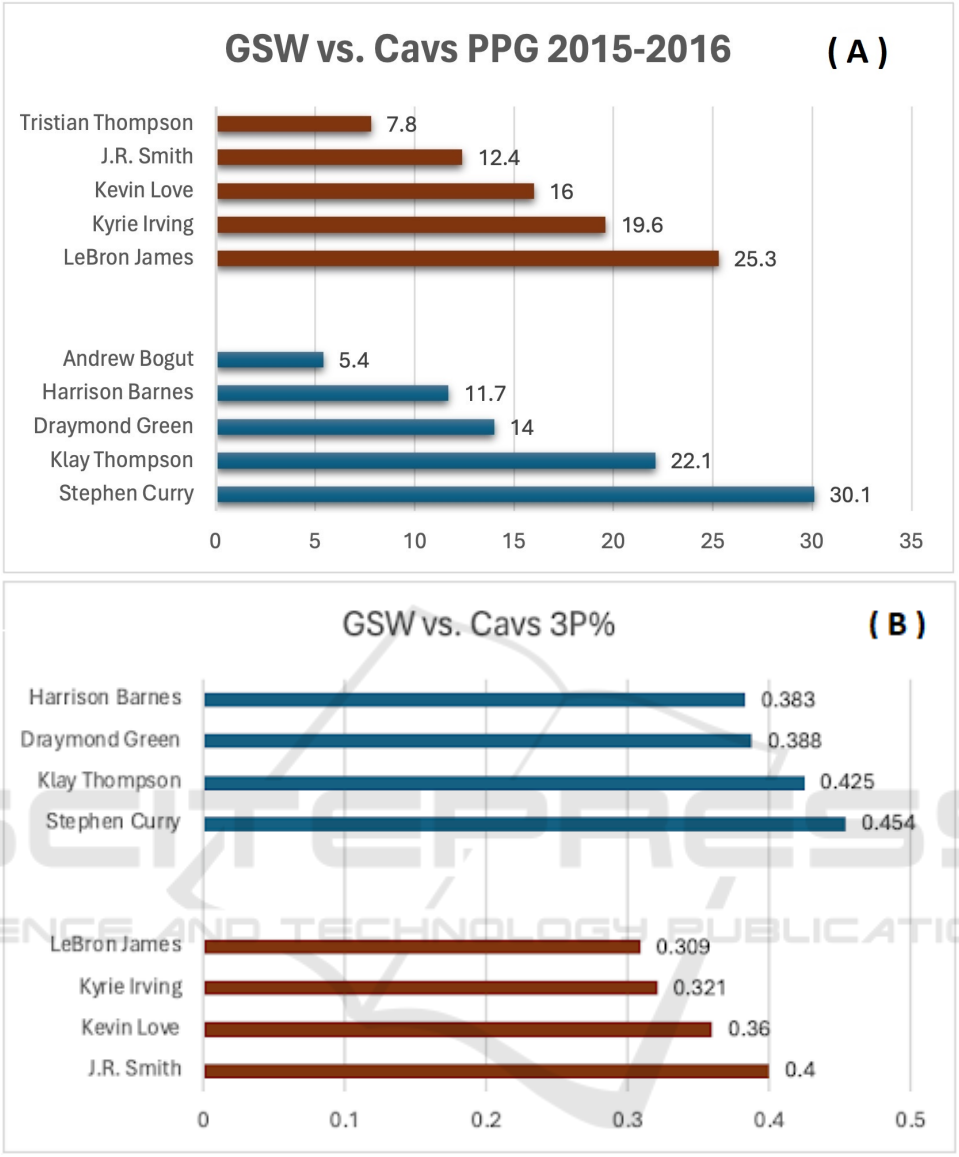

Figure 7: A. Points per game and B. three-point percentage for Cleveland. Cavaliers (brown) and the Golden State Warriors (blue).

tified three clusters based on win percentage and the number of players at different skill levels. The most successful teams tend to have a higher number of top-end players and a significant number of players in the lowest skill level. On the other hand, teams that evenly distribute their talent across the roster tend to be less successful. We've also discovered that players of average to above-average skill levels are most affected by excessive playing time in the previous game. If these players have logged significant minutes in the previous game, they are more likely to show a decline in performance and potentially lead the team to a loss in the next game. Taking into account the time of year and the gap between games, we recommend giving priority to the rest and recovery of the top players, particularly in the latter half of the season. It is crucial to ensure that below-average players maintain consistent performance, especially during the first half of the season.

When we consider the impact of height on player performance, we find that the shortest and tallest players tend to underperform compared to those closer to average height. The majority of top performers in the NBA have an average height compared to other players. However, for players who operate near the net and encounter many rebound opportunities, an average or taller player is preferable to a below-average height player for slightly better overall performance.

Coaches and management could use this information to construct and deploy teams more effectively, leading to an increased win percentage in regular season games. Additionally, coaches could analyze successful seasons, such as the Golden State Warriors in 2015-2016, to identify important factors leading to these achievements, such as having players who can effectively score three-pointers.

REFERENCES

Basketball Reference (2023a). Glossary. https://www.basketball-reference.com/about/glossary.html. Accessed: 10-15-2023.

Basketball Reference (2023b). Nba player stats: Advanced. https://www.basketball-reference.com/leagues/NBA_2022_advanced.html. Accessed: 12-10-2023.

Berri, D. J., Brook, S. L., and Fenn, A. J. (2011). From college to the pros: Predicting the nba amateur player draft. *Journal of Productivity Analysis*, 35:25–35.

Bishop, E. (2023). 5 ways nba teams use analytics to gain a competitive edge. https://www.sportskeeda.com/basketball/5-ways-nba-teams-use-analytics-gain-competitive-edge. Accessed: 12-10-2023.

Burns, M. (2023). Nba team market size rankings. https://hoop-social.com/nba-team-market-size-rankings/. Accessed: 12-10-2023.

Drakos, M. C., Domb, B., Starkey, C., Callahan, L., and Allen, A. A. (2010). Injury in the national basketball association: a 17-year overview. *Sports health*, 2(4):284–290.

ESPN (2023). Nba player stats. https://www.espn.com/nba/stats/player/_/season/2020/seasontype/2. Accessed: 12-10-2023.

Fearnhead, P. and Taylor, B. M. (2011). On estimating the ability of nba players. *Journal of Quantitative analysis in sports*, 7(3).

Franks, A., Miller, A., Bornn, L., and Goldsberry, K. (2015). Counterpoints: Advanced defensive metrics for nba basketball. In *MIT Sloan Sports Analytics Conference*, Boston, MA. Presented at the MIT Sloan Sports Analytics Conference.

Fromal, A. (2023). Understanding the nba: Explaining advanced offensive stats and metrics. https://bleacherreport.com/articles/1039116-understanding-the-nba-. Accessed: 12-14-2023.

Gonzalez, A. M., Hoffman, J. R., Rogowski, J. P., Burgos, W., Manalo, E., Weise, K., Fragala, M. S., and Stout, J. R. (2013). Performance changes in nba basketball players vary in starters vs. nonstarters over a competitive season. *The Journal of Strength & Conditioning Research*, 27(3):611–615.

Kaggle (2022). Nba games data. https://www.kaggle.com/datasets/nathanlauga/nba-games?select=games_details.csv. Accessed: 12-10-2023.

McIntyre, A., Brooks, J., Guttag, J., and Wiens, J. (2016). Recognizing and analyzing ball screen defense in the nba. In *Proceedings of the MIT sloan sports analytics conference, Boston, MA, USA*, pages 11–12.

NBA (2023). About the nba. https://www.nba.com/news/about. Accessed: 12-10-2023.

NBA Stuffer (2023a). Analytics movement in the nba. https://www.nbastuffer.com/analytics101/nba-analytics-movement/. Accessed: 12-10-2023.

NBA Stuffer (2023b). Nba plus-minus and impact metrics in basketball explained. https://www.nbastuffer.com/analytics101/plus-minus/. Accessed: 12-14-2023.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É. (2011). Scikit-learn: Machine learning in python.

Raj, A. (2023). An exhaustive guide to decision tree classification in python 3.x. https://towardsdatascience.com/an-exhaustive-guide-to-\newlineclassification-using-decision-trees-8d472e77223f. Accessed: 12-14-2023.

Sporting Charts (2023). What is offensive win shares. https://www.sportingcharts.com/dictionary/nba/offensive-win-shares-ows.aspx. Accessed: 12-14-2023.

Sports Lingo (2023). Defensive win shares (dws). https://www.sportslingo.com/sports-glossary/d/defensive-win-shares-dws/. Accessed: 12-14-2023.