Sports Analytics Using Data Mining: NBA Player Quarter-by-Quarter Performance Exploration

Dimitrios Iatropoulos

SID: 3308220015

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Data Science

JANUARY 2025

THESSALONIKI – GREECE

Sports Analytics Using Data Mining: NBA Player Quarter-by-Quarter Performance Exploration

Dimitrios Iatropoulos

SID: 3308220015

Supervisor:	Prof. Christos Tjortjis
Supervising Committee Members:	Dr. Christos Berberidis Dr. Paraskevas Koukaras

SCHOOL OF SCIENCE & TECHNOLOGY

A thesis submitted for the degree of

Master of Science (MSc) in Data Science

JANUARY 2025

THESSALONIKI – GREECE

Abstract

This dissertation is submitted for the MSc in Data Science at the International Hellenic University. It analyzes quarter-by-quarter variations in NBA player performance metrics and their impact on game outcomes during the regular season and playoffs from the 2004-05 to 2023-24 seasons.

Using Association Rule Mining (ARM) and FP-Growth, the study examines advanced statistics across five key categories: Offense, Defense, Ball Handling, Overall Impact, and Tempo. Data from the NBA API was cleaned to address missing values and outliers. The findings show that defensive metrics play a bigger role in the fourth quarter and playoffs, while offensive metrics like Effective Field Goal Percentage (eFG%) and True Shooting Percentage (TS%) are especially important in playoff games.

Descriptive analysis highlights standout performances from players like Nikola Jokić, Stephen Curry, Tim Duncan, and LeBron James, emphasizing the value of position-specific skills and overall impact. This research offers actionable insights for strategy optimization and player development, helping teams succeed in high-pressure situations.

KEY WORDS:

Association Rule Mining, Basketball Analytics, Data Mining, Data Science, Defensive Metrics, Machine Learning, Offensive Efficiency, Performance Metrics, Quarter-by-Quarter Analysis, Sports Analytics

Acknowledgments

I sincerely thank my supervisor, Professor Christos Tjortjis, for his invaluable guidance, constructive feedback, and continuous support throughout this dissertation. His expertise and insights were essential to the success of this work.

I am also grateful to Postdoctoral Researcher Dr. Vangelis Sarlis for his expert advice and willingness to assist whenever needed, as well as to PhD candidate Georgios Papageorgiou for his helpful suggestions and efficient feedback.

Finally, I am deeply thankful to my family for their constant encouragement and understanding during this demanding period. Their support was my strength throughout this effort.

Dimitrios Iatropoulos

15-01-2025

Contents

List of Figures

List of Tables

1 Introduction

Advanced analytics has transformed the way we understand professional sports. In basketball, especially in the NBA, data science techniques like machine learning (ML) have revealed new ways to analyze player performance, team strategy and game outcomes. While performance metrics like offensive and defensive ratings, and assist ratios are common, few studies explore how these stats shift from quarter to quarter and how they change between the regular season and the playoffs.

This thesis fills that gap by using Association Rule Mining (ARM) to analyze per-quarter player performance over two decades of NBA games (2004-05 to 2023-24 seasons). Unlike past research that focuses on entire games or just their final moments, this work examines all quarters to uncover patterns that impact game outcomes.

Data from the NBA API, covering regular season and playoff games, was cleaned and processed to handle missing values and outliers. The FP-Growth algorithm was applied to discretized data to identify patterns and generate association rules. The analysis focused on five key categories: Offense, Defense, Ball Handling, Overall Impact, and Tempo, allowing for a detailed examination of different game aspects. Composite scores were calculated to estimate each player's contributions by combining metrics into a single measure.

The results highlight how performance metrics influence game outcomes and how these relationships change between the regular season and playoffs. Defensive metrics became more crucial in the fourth quarter, while offensive metrics like Effective Field Goal Percentage (eFG%) and True Shooting Percentage (TS%) were more impactful in playoff games.

Nikola Jokic exceled in Overall Impact and Offense during both the regular season and playoffs. Stephen Curry's control of Tempo was key to his team's playoff success. Tim Duncan consistently stood out defensively, ranking high in defensive rating. LeBron James demonstrated unmatched versatility, ranking among the best in both offensive and defensive metrics. His ability to elevate his performance in high-pressure moments, such as the fourth quarter and overtime, reinforces his status as one of the most impactful players in NBA history.

Position-specific insights showed that Guards with strong ball-handling, Centers with solid defense and rebounding, and Forwards with high assist rates were crucial to their teams' success.

This thesis fills a gap in NBA analytics by offering a quarter-by-quarter analysis of player performance. The findings provide actionable insights for coaches and teams, emphasizing the importance of defense and overall impact in playoff games. By tracking how performance metrics evolve throughout the game, this research introduces a new perspective to sports analytics.

2 Background

Key concepts related to the research are explained to provide the necessary background for understanding the methods and analyses discussed later in this dissertation. The topics covered include:

- Data Mining (DM)
- Machine Learning (ML)
- Association Rule Mining (ARM)
- Sports Analytics

2.1 Data Mining

Data Mining (DM) is the process of finding patterns, relationships, and unusual trends in large datasets using techniques from statistics, machine learning, and database systems. It helps uncover useful insights in areas like business, healthcare, and sports analytics.

As Figure 1 shows, it includes gathering data from different sources, cleaning and adjusting it, analyzing it to extract patterns, and finally reporting the findings. Each step is crucial for ensuring accurate and valuable insights.

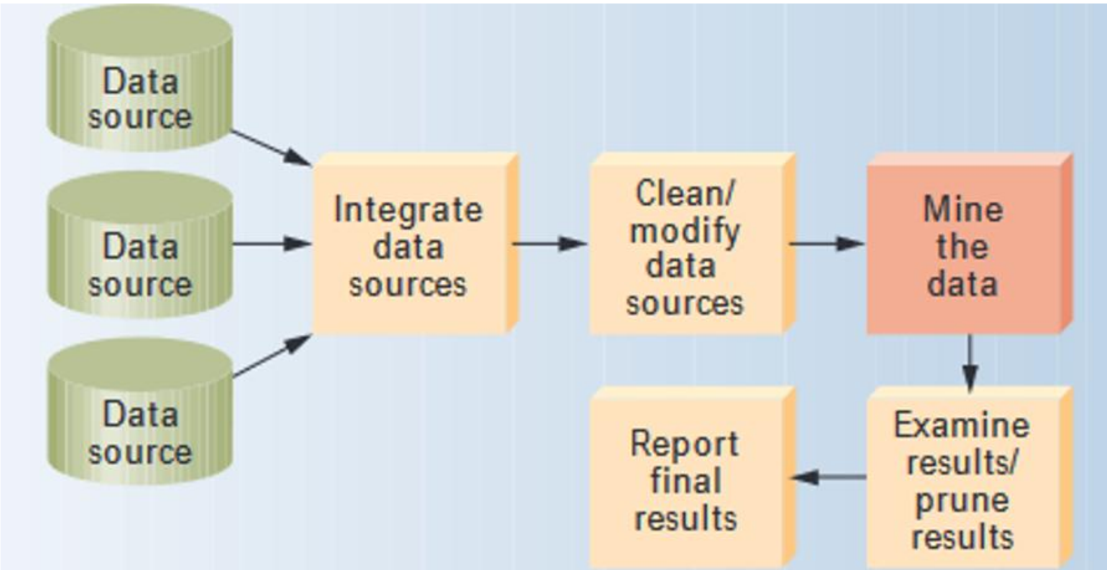

Figure 1: Data Mining Process Flow [1]

- **Data Selection**: The first step is choosing and extracting relevant data from various sources. This involves defining the analysis scope and making sure the data matches the research goals [1].

- **Data Preprocessing**: Raw data is often messy, incomplete, or inconsistent. This step cleans the data by filling in missing values, reducing noise, and removing outliers. In sports analytics, preprocessing is especially important since factors like injuries, fouls, or player substitutions can affect the data [2], [3].
- **Data Transformation**: The data is then formatted for analysis. One common method is discretization, which converts continuous variables (such as player efficiency scores) into categories. This is crucial for algorithms like Association Rule Mining (ARM), which work better with categorical data [2], [3].

Data Mining Tasks

Once the data is ready for analysis, DM techniques are applied to extract insights. These tasks help us understand patterns and relationships within the data that are otherwise difficult to detect. Some key tasks include:

- **Clustering** is a DM technique used to group similar data points into clusters. Several clustering methods exist, including hierarchical, partitioning, density-based, and model-based clustering. Clustering is useful for unsupervised learning tasks where the goal is to discover hidden patterns or relationships in the data. K-means algorithm is one of the most popular partitioning methods, which minimizes the squared distance between data points and cluster centroids. Hierarchical clustering, on the other hand, forms a dendrogram and can either merge or divide clusters based on a similarity threshold [4].

- **Anomaly detection** involves identifying data points that deviate significantly from the norm. These outliers may represent rare events or errors in data collection. In sports analytics, anomalies often result from player injuries or unexpectedly poor or exceptional performances. Various methods detect these anomalies, including distance-based methods, statistical approaches, and ML techniques [5].

- **Association Rule Mining** (ARM) identifies relationships between variables in large datasets. Two widely used algorithms in this field are **Apriori** and **FP-Growth**. Apriori Builds larger sets of frequently co-occurring items step by step. However, it requires multiple database scans, making it computationally expensive. FP-Growth improves efficiency by using a prefix-tree (FP-Tree) structure, reducing database scans while discovering frequent itemsets. Key parameters in ARM include [6]:

i) Support measures how frequently an itemset appears in the dataset. It is a parameter for identifying frequently occurring patterns.

ii) Confidence reflects the likelihood that the consequence occurs when the antecedent is present, indicating the rule's reliability.

iii) Lift compares the confidence of the rule to the expected confidence if the antecedent and consequent were independent. A lift greater than 1 suggests a strong association between them.

2.2 Machine Learning (ML)

ML is a branch of artificial intelligence (AI) that enables computers to learn from data and improve their performance without being explicitly programmed. It involves creating algorithms that allow machines to learn from patterns in data and make predictions or decisions based on that data.

As a result, ML is widely used in applications such as fraud detection, image recognition, recommendation systems, and more.

The goal of ML is to create models that generalize from known examples to new, unseen data. These models use statistical techniques to identify patterns and relationships in large datasets, improving their performance as they process more data.

There are three main types of ML: Supervised Learning, Unsupervised Learning, and Reinforcement Learning [7], [8], [9], [10].

2.2.1 Supervised Learning

Supervised learning is the most common type of machine learning, where models are trained on labeled data. Each input has a corresponding correct output, allowing the model to learn by minimizing the difference between its predictions and actual labels.

Examples of **Supervised Learning**:

- Predicting player injuries based on past injury data.
- Classifying emails as spam or not spam.

Algorithms:

- **Decision Trees** split data into subsets based on feature values, creating an interpretable tree-like structure for decision-making. Useful in medical diagnosis but prone to overfitting with complex datasets.

- **Support Vector Machines (SVM)** find the best hyperplane to separate classes, making them effective in high-dimensional spaces like image recognition. They handle non-linear relationships but can be computationally expensive.

- **Linear Regression** predicts continuous outcomes using a linear relationship between input features and the target variable. Simple and interpretable but struggles with non-linear data and outliers.

- **Neural Networks** are inspired by the human brain, they capture complex patterns and power deep learning applications like speech recognition and NLP. They require large datasets, high computational power, and are difficult to interpret.

The performance of these models is evaluated using metrics such as accuracy, precision, recall, and the F1 score [7], [8], [9], [10].

2.2.2 Unsupervised Learning

In unsupervised learning, the dataset lacks labeled outputs, so the model identifies hidden patterns, correlations, or structures without explicit guidance. It is used for tasks like clustering to group similar data points.

Examples of **Unsupervised Learning**:

- Grouping customers into segments based on purchasing behavior.
- Detecting anomalies in player performance during games.

Algorithms:

- **DBSCAN (Density-Based Spatial Clustering of Applications with Noise)** is a density-based clustering algorithm that identifies anomalies as points far from dense clusters. It classifies core points as those with many neighbors within a set radius (epsilon) and labels points with few neighbors as outliers. Unlike Z-Score and IQR, DBSCAN does not rely on a specific data distribution, making it useful for datasets with uneven distributions. Its strength lies in distinguishing between dense clusters and noise [11].

- **K-Means** is a centroid-based clustering method that groups data points into clusters based on their natural distribution patterns. By calculating the midpoints between cluster centers, this method allows for the creation of bins that reflect the underlying structure of the data. It is useful for complex datasets where natural groupings are not immediately visible [12].

- **Principal Component Analysis (PCA)** is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional space while retaining most of the original variance. It is generally used for data visualization, noise reduction, and improving the efficiency of other ML algorithms. PCA reduces the complexity of datasets by finding the most important features, but it assumes linear relationships, which can limit its ability to capture complex patterns in the data.

- **Autoencoders** are neural networks that learn efficient data representations by compressing inputs into a latent space and then reconstructing them. They are useful for tasks like anomaly detection and feature learning, uncovering hidden patterns without needing labeled data. However, they require considerable computational power and careful tuning to avoid overfitting.

Since unsupervised learning models don't use labels, evaluating their success is challenging. Common techniques include visualizing clusters to identify patterns and using metrics like the silhouette score [7], [8], [9], [10].

2.2.3 Reinforcement Learning

Reinforcement learning is a goal-driven approach where an agent learns by interacting with its environment. It makes decisions, gets feedback through rewards or penalties, and adjusts its actions to maximize long-term rewards. Unlike supervised learning, there's no fixed correct answer and the agent improves through trial and error.

Examples of **Reinforcement Learning**:

- Training robots to perform tasks such as walking or picking up objects.
- Developing AI systems that can play complex games like chess.

Algorithms:

- **Q-Learning** is a model-free reinforcement learning algorithm that helps an agent determine the best action to take in a given state. It updates Q-values based on rewards received, gradually learning an optimal policy to maximize total rewards. While effective in discrete environments like simple games and navigation tasks, it struggles with large or continuous state spaces due to the growing size of the Q-table.

- **Deep Q-Networks (DQN)** improve Q-Learning by using deep neural networks to approximate Q-values, allowing them to handle complex, high-dimensional environments like video games and robotic control. DQNs generalize better across states and actions but require significant computational power and careful tuning to ensure stability.

- **The Monte Carlo Method** estimates values by averaging returns from multiple episodes without needing a model of the environment. It works well in scenarios with clearly defined episodes, such as board games like chess, where repeated sampling provides reliable learning. While Monte Carlo methods are simple and straightforward, they can be inefficient for environments with long or infinite episodes and may have high variance in their estimates.

Reinforcement learning is frequently used in fields such as robotics, gaming, and autonomous systems. The agent's performance is measured by how well it maximizes its reward over time [7], [8], [9], [10].

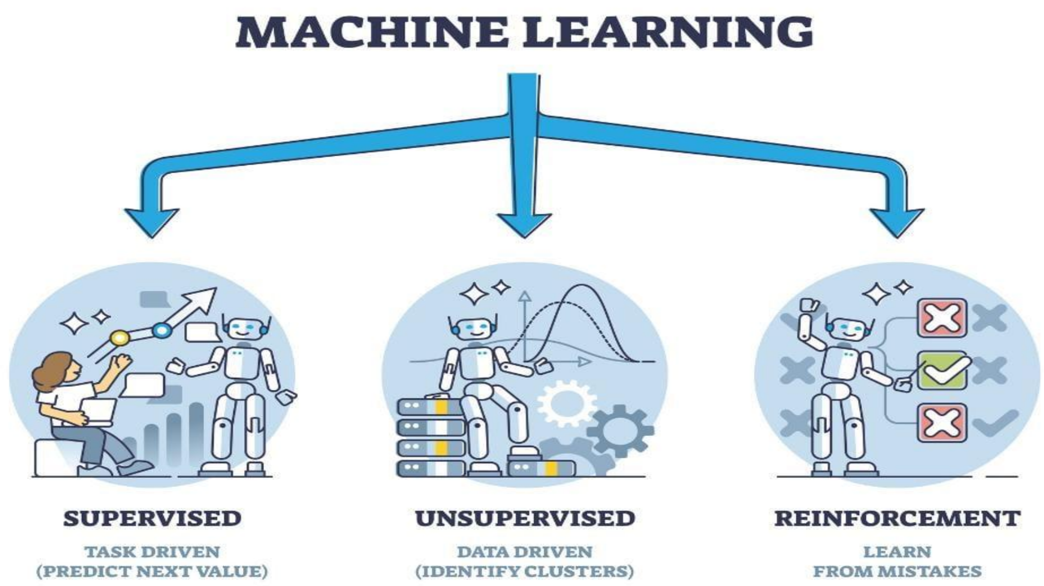

Figure 2: Machine Learning Categories [13]

Figure 2 uses robot-themed illustrations to show the unique characteristics of each category.

2.3 Sports Analytics

Sports analytics refers to the systematic use of data to analyze player performance, team strategies, and operational efficiency within sports organizations. By leveraging advanced statistics, ML, and DM techniques, sports teams and organizations gain insights that inform better decision-making both on and off the field [14]. First gaining attention by the Moneyball approach in Major League Baseball, sports analytics has grown in popularity across many sports, applying to areas like player selection, injury prevention, and even predicting game outcomes [15]. As technology and data availability have advanced, sports analytics has become vital to fan engagement and business strategies, influencing decisions like ticket pricing, sponsorship returns, and personalized marketing campaigns [16].

In the following sections, we explore the applications of sports analytics, around both on-field and off-field strategies that enhance competitive advantage and organizational success. We examine how data optimizes player and team performance through real-time metrics, and we investigate advanced basketball analytics, focusing on metrics that provide insights into player efficiency.

2.3.1 Sports Analytics Applications

Sports analytics can be divided into two categories: on-field analytics and off-field analytics. On-field analytics helps teams maximize performance by analyzing stats, strategies, and player health. Wearable devices and tracking systems collect real-time data on metrics like heart rate, speed, and positioning. Coaches use this data to refine tactics, personalize training, and even predict injuries. Some teams have reduced injuries by integrating predictive models into their routines.[14].

Off-field analytics focuses on the business side of sports, analyzing fan engagement, ticket sales, and sponsorships to boost revenue. Teams like the Tampa Bay Lightning have used data-driven pricing strategies to adjust ticket prices based on demand, maximizing sales. Fan interactions on mobile apps, social media, and stadium Wi-Fi are also analyzed to create targeted marketing campaigns.[16], [17].

Sports analytics are not only limited to team performance and business operations. It is also gaining traction in areas such as player recruitment, where predictive models are used to assess the future performance of athletes. This helps teams make better decisions on signing and developing players, thereby maintaining a competitive edge in the market [17].

Figure 3 presents the benefits of sports analytics, including performance optimization, injury prevention, training excellence, strategic decision-making, fan engagement, and talent scouting.

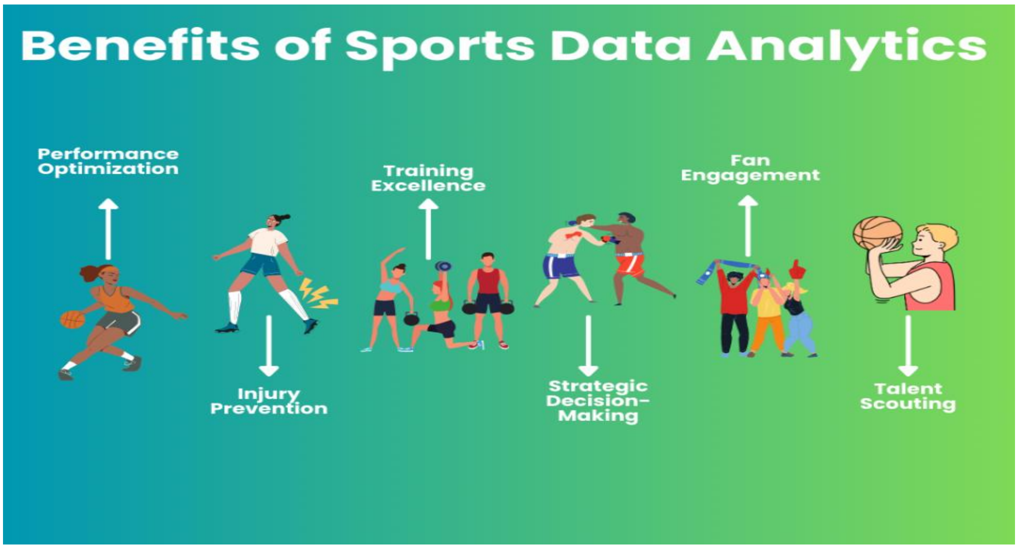

Figure 3: Benefits of Sports Analytics [18]

2.3.2 Advanced Basketball Analytics

To analyze player performance in basketball, especially in the context of on-field analytics, a range of advanced metrics have been developed. These metrics go beyond traditional box score statistics to provide more detailed insights into player efficiency, impact, and overall performance. Below, we outline the advanced metrics used in this work. For detailed definitions of each variable in the formulas, please refer to Appendix A, Table 5.

- **Offensive Rating (OffRtg)**

Definition: The team's points scored per 100 possessions when a specific player is on the court.

$$OffRtg = 100\frac{PTS}{POSS}$$

Explanation: This metric measures how effectively a player contributes to their team's offense. Higher values indicate better offensive performance by the player and their team [19].

- **Defensive Rating (DefRtg)**

Definition: The number of points per 100 possessions that the opposing team scores while a specific player is on the court.

$$DefRtg = 100\frac{OppPTS}{OppPOSS}$$

Explanation: This metric measures how effectively a player contributes to their team's defense. Lower values indicate better defensive performance by the player and their team during that time [19].

- **Net Rating (NetRtg)**

Definition: The team's point differential per 100 possessions when a specific player is on the court.

$$NetRtg = OffRtg - DefRtg$$

Explanation: This metric measures the overall impact a player has on the game by combining their offensive and defensive performances. Positive values indicate that the team outperforms opponents while the player is on the court [19].

- **True Shooting Percentage (TS%)**

Definition: A measure of a player's shooting efficiency, accounting for field goals, three-point field goals, and free throws.

$$TS\% = \frac{PTS}{2 \times (FGA + 0.44 \times FTA)}$$

Explanation: This metric gives a comprehensive measure of a player's shooting efficiency, containing all scoring attempts. Higher values indicate greater scoring efficiency [19].

- **Usage Percentage (USG%)**

Definition: The percentage of team plays used by a player.

$$USG\% = \frac{(FGA + PossEndFTA + TO)}{POSS}$$

Explanation: This metric measures how much of a team’s offense a player is responsible for when on the floor. Higher values indicate the player is heavily involved in the team’s offensive plays [19].

- **Possessions**

Definition: A measure of the total number of offensive opportunities a player has during a game.

$$POSS = 0.5 \times ((TmFGA + 0.4 \times TmFTA - 1.07 \times (\frac{TmORB}{TmORB + OppDRB}) \times (TmFGA - TmFG) + TmTOV) + (OppFGA + 0.4 \times OppFTA - 1.07 \times (\frac{OppORB}{OppORB + TmDRB}) \times (OppFGA - OppFG) + OppTOV))$$

Explanation: This metric estimates a player's involvement in offensive opportunities by considering their contributions while excluding possessions that continue after an offensive rebound. Higher values indicate more offensive involvement [20].

- **Pace**

Definition: The number of possessions a player involved per 48 minutes of play.

$$PACE = 48 \frac{(TmPoss + OppPoss)}{2 \times (\frac{TmMp}{5})}$$

Explanation: This metric indicates the tempo at which a player is involved in the game. Higher values indicate a faster game involvement with more possessions [20].

- **Player Impact Estimate (PIE)**

Definition: A measure of a player's overall statistical contribution to a game.

$$PIE = \frac{PTS + FGM + FTM - FGA - FTA + DREB + (0.5 \times OREB) + AST + STL + (0,5 \times BLK) - PF - TO}{GmPTS + GmFGM + GmFTM - GmFGA - GmFTA + GmDREB + (0.5 \times GmOREB) + GmAST + GmSTL + (0,5 \times GmBLK) - GmPF - GmTO}$$

Explanation: This metric evaluates a player's total impact on their team's performance, with higher values indicating a greater contribution to overall success [19].

- **Effective Field Goal Percentage (eFG%)**

Definition: A measure of Field Goal Percentage adjusted for the additional value of three-point field goals.

$$eFG\% = \frac{FGM + (0.5 \times 3PM)}{FGA}$$

Explanation: This metric provides a more accurate measure of a player's shooting efficiency by giving extra weight to three-point field goals. Higher values indicate better shooting performance [19].

- **Assist Percentage (AST%)**

Definition: The percentage of teammate field goals a player assisted.

$$AST\% = \frac{AST}{TmFGM - FGM}$$

Explanation: This metric measures a player's contribution to their team's scoring by assisting on field goals. Higher values indicate that the player is more involved in creating scoring opportunities for their teammates [19].

- **Assist to Turnover Ratio (AST/TO)**

Definition: The number of assists for a player compared to the number of turnovers they have committed.

$$AST/TO = \frac{AST}{TO}$$

Explanation: This metric evaluates a player's efficiency in handling the ball. Higher ratios indicate better ball control and playmaking efficiency [19]

- **Assist Ratio (AST Ratio)**

Definition: The number of assists a player averages per 100 possessions.

$$AST\ Ratio = 100\frac{AST}{POSS}$$

Explanation: This metric measures how frequently a player assists with scoring plays. Higher values indicate that a player is more effective in creating scoring opportunities for teammates [19].

- **Turnover Ratio (TOV%)**

Definition: The number of turnovers a player averages per 100 possessions.

$$TOV\% = 100\frac{TO}{POSS}$$

Explanation: This metric estimates ball security, with lower values indicating greater control and fewer turnovers [19].

- **Offensive Rebound Percentage (ORB%)**

Definition: The percentage of available offensive rebounds a player grabs.

$$ORB\% = 100\frac{ORB \times \frac{TmMP}{5}}{MP \times (TmORB + OppDRB)}$$

Explanation: This metric estimates a player's effectiveness in securing offensive rebounds, with higher values indicating better performance [20].

- **Defensive Rebound Percentage (DRB%)**

Definition: The percentage of available defensive rebounds a player grabs.

$$DRB\% = 100\frac{DRB \times \frac{TmMP}{5}}{MP \times (TmDRB + OppDRB)}$$

Explanation: This metric estimates a player's effectiveness in securing defensive rebounds, with higher values indicating better performance [46].

- **Rebound Percentage (TRB%)**

Definition: The percentage of total rebounds (offensive and defensive) a player secures.

$$TRB\% = 100\frac{TRB \times \frac{TmMP}{5}}{MP \times (TmTRB + OppTRB)}$$

Explanation: This metric estimates a player's effectiveness in securing total rebounds, with higher values indicating better performance [20].

3 Literature Review

Sports analytics have reshaped the game, using data to improve performance and decision-making. It all started with early experiments and key breakthroughs that paved the way for today's advanced methods. Different sports, like baseball, tennis, football, and basketball, apply analytics in their own unique ways. From evaluating players to crafting game strategies and boosting team performance, data has become a game-changer in modern sports.

3.1 Historical Background

Sports analytics has become a key part of modern sports, helping teams make smarter decisions, develop strategies, and improve player performance. With technology advancing, collecting and analyzing massive amounts of data is easier than ever, leading to new tactics, playstyles, and deeper insights. Back in the day, the process of data collection was entirely manual, and therefore, it was a slow and time-consuming operation. The current version of the sport data tools has so much altered the whole picture that it is already hard to imagine sports management without data [14], [16].

3.1.1 Baseball

Baseball is seen as the pioneer of sports analytics, with its roots in statistical tracking going back to the 19th century. The first recorded baseball stats appeared in the 1830s when teams started using scorebooks to monitor player performance, mainly focusing on runs scored. In the mid-19th century, Henry Chadwick, often credited as the father of baseball, introduced the box score, which changed how games were recorded and analyzed [21].

The development of advanced metrics such as batting average, on-base percentage, and earned run average (ERA) in the early 20th century provided a way to measure a player's contribution to the game beyond simple observations, making performance assessments more data driven [22].

Baseball's use of data grew stronger with the rise of sabermetrics. Bill James, through his influential *Baseball Abstract* series, challenged traditional views of player evaluation and introduced new statistics that better predicted player and team success [23]. This analytical approach gained recognition with the publication of *Moneyball* in 2003, which discussed how Oakland Athletics used data to build a competitive team on a limited budget, essentially changing the sport's strategic approach [24].

The integration of advanced analytics into baseball has extended beyond player performance metrics. Studies have shown significant correlations between anthropometric characteristics, such as height, weight, and body fat percentage, and performance outcomes like batting average and slugging percentage [25]. Fitness metrics like grip and squat strength have been identified as predictors of success in both offensive and defensive play, emphasizing the role of physical conditioning in the sport [26].

Today, baseball remains the sport most closely tied to sports analytics, with technologies like tracking systems and ML models being applied to optimize player performance, predict injuries and refine game strategies [22]. The evolution of baseball analytics has not only changed how the game is played but has also influenced the way fans, analysts, and teams engage with the sport [23].

3.1.2 Tennis

In recent years, tennis has embraced data analytics to enhance player performance and strategy, helping in a new era of data-driven coaching. Technologies like the Hawk-Eye system, which tracks balls and players in detail, has changed both training and competition by generating vast amounts of spatiotemporal data [27], [28]. To a great extent, this data has been instrumental in the player shots' analysis, and even the player's best strategies to acquire maximum shots' efficiency. Serve power, in fact, is a strong predictor of match results, as shown in random forest models that the players who often win a high percentage of first serves are the ones with the most significant advantage [29].

Visual analytics tools like CourtTime take data a step further by adding match context, helping players and coaches spot patterns beyond traditional stats. They can analyze key moments in rallies and see how different shot combinations impact winning or losing points [30]. As analytics play a bigger role, tennis has become more data-driven, giving players and coaches sharper insights for match preparation and in-game decisions.

3.1.3 Football

Analytics have transformed soccer, shifting the game from coach intuition to data-driven decision-making. Advanced statistical tools and big data now influence everything from player performance analysis to strategic planning [31], [32].

Machine learning and big data help teams predict outcomes, optimize player positioning, and reduce injury risks.[33], [34]. Predictive models, like those used in the Greek League, assess team strength, player performance, and match probabilities, helping coaches make informed tactical decisions [35].

Metrics like Expected Goals (xG) go beyond traditional stats by analyzing shot quality and possession control, offering a clearer view of performance [36].

As soccer analytics evolve, advancements in AI and wearable technology will further enhance data collection and analysis, pushing the sport into an even more data-driven future.. The use of AI, in decision-making and tactical execution, offers exciting possibilities for teams looking to gain a competitive edge [37].

3.1.4 Basketball

The integration of analytics into basketball took off with Dean Oliver's work, "Basketball on Paper," which is frequently credited as the start of modern basketball analytics. Published in 2004, the book introduced metrics like Offensive Rating (ORtg) and Defensive Rating (DRtg), which evaluate team and player efficiency per possession. Oliver's work provided a foundation for evaluating players and teams through a focus on efficiency in performance analysis. "Basketball on Paper" set the stage for the broader application of advanced statistics in basketball [38], [39].

In 2010-2011, SportVU cameras took things further by tracking player and ball movement 25 times per second. By 2013-2014, every NBA arena had SportVU. Analysts used this real-time data to study positioning, spacing, and speed [40].

3.2 Basketball Analytics Evolution

Many researchers have explored how injuries affect NBA players and teams. Sarlis et al. [41] in 2021 used data science methods to study injury patterns and found weak, but statistically significant correlations between injuries and player/team performance over ten seasons, explaining the complex impact of injuries on the game.

Applying text mining techniques to injury reports, Sarlis et al. [42] in 2023 analyzed the recovery dynamics from injuries, contributing valuable insights into player rehabilitation and performance in professional basketball.

In 2024, Papageorgiou et al. [43] examined the relationships between injury types, recovery durations, and economic impacts, revealing that players with lower salaries experienced shorter recovery periods, suggesting a financial influence on recovery decisions, which could affect both player performance and team dynamics.

In 2024, Sarlis et al. [44] explored the economic consequences of injuries, identifying links between player recovery times and team financial outcomes, emphasizing the financial impact of high-salary players on team losses. Lastly, studies by Sarlis and Tjortjis [45] in 2024 applied DM techniques to analyze the influence of player positioning, age, and injuries, revealing significant associations between these factors and both performance outcomes and economic impacts, such as salary and team success.

ML models have increasingly been applied to predict NBA game outcomes and player performance. Using regression analysis in 2015, Yang [46] predicted regular season results, and found that player efficiency ratings are strongly correlated with team performance, identifying key performance indicators for predictive modeling.

Chen et al. [47] in 2021 developed hybrid models combining multiple DM techniques to predict final game scores, providing more detailed insights into game dynamics beyond simple win/loss outcomes. J. K. Nakul et al., [48] in 2022 utilized Artificial Neural Networks (ANN) integrated with advanced statistics, achieving notable predictive accuracy. More recently, Papageorgiou et al. [49] in 2024 conducted a comparative study evaluating the effectiveness of 14 ML in forecasting player performance, concluding that tree-based models showed strong predictive power when applied to basketball statistics.

The development and application of advanced metrics have transformed how player performance is analyzed in the NBA. Ruano et al. [50] in 2016 analyzed the dynamics of NBA games during critical moments, such as the fourth quarter of close games, revealing how performance metrics can influence outcomes in high-pressure situations.

Research by Dehesa et al. [51] identified key performance profiles of NBA players using metrics such as effective field-goal percentage, offensive/defensive ratings. Their findings revealed variations in performance clusters between the regular season and playoffs, with playoff profiles being more stable. More recently, Sarlis et al. [52] in 2024 introduced the Estimation of Clutch Competency (EoCC), a metric that evaluates a player’s impact in the final minutes of close games by combining offensive, defensive, and playmaking contributions.

Sarlis [53] in 2024 also discussed how data science in sports analytics has refined sports analytics, improving the accuracy of player evaluations. Insights from Ben Taylor’s "Thinking Basketball" [54] in 2024 emphasize advanced metrics like Player Efficiency Rating (PER) and True Shooting Percentage (TS%), which offer a clearer view of a player’s impact beyond traditional stats.

AI and computational intelligence are further transforming sports analytics. Sha et al. [55] in 2018 developed intelligent interfaces for real-time player trajectory analysis, using tracking data to visualize and predict player movements. These tools allow users to modify player positions and see how different scenarios affect performance. Ghosh et al. [56] in 2021 discussed the growing role of AI in player performance analysis, strategic decision-making and injury management.

Despite extensive research on NBA player performance, gaps remain in understanding how advanced metrics vary on a quarter-by-quarter basis and across different season types. Most existing studies focus on overall game dynamics or solely on critical moments like the fourth quarter. There is a lack of analysis that examines all the quarters in a game to understand performance trends.

Advanced data mining techniques like ARM have been rarely utilized in NBA analytics. This gap prevents the discovery of complex patterns and relationships within quarter-by-quarter performance data. This thesis aims to fill these gaps by conducting an analysis of advanced metrics across all game quarters and comparing these metrics between regular season and playoff games. By employing ARM, the research will uncover and analyze the patterns within these metrics, providing an understanding of player performance dynamics and their impact on game outcomes in different competitive contexts.

4 Methodology

This chapter details the methodology used to collect, process, and prepare the data for this work. Data was collected from both regular season and playoff games over the past 20 years, focusing on per-quarter player performance. The NBA API [57] was utilized for data retrieval, with Python scripts managing all aspects of the process. Data processing was performed on Google Collaboratory [58], utilizing its TPU and 334 GB of RAM. The mlxtend library [59] was used to run the FP-Growth algorithm for ARM. For a complete explanation of the API endpoints and parameters used, refer to Appendix B. The code and files used in this study are hosted in a private repository, accessible via https://github.com/Dimitrios-Iatropoulos/Thesis only if access is granted. To request access, please contact me at diatropoulos@ihu.edu.gr with your GitHub username.

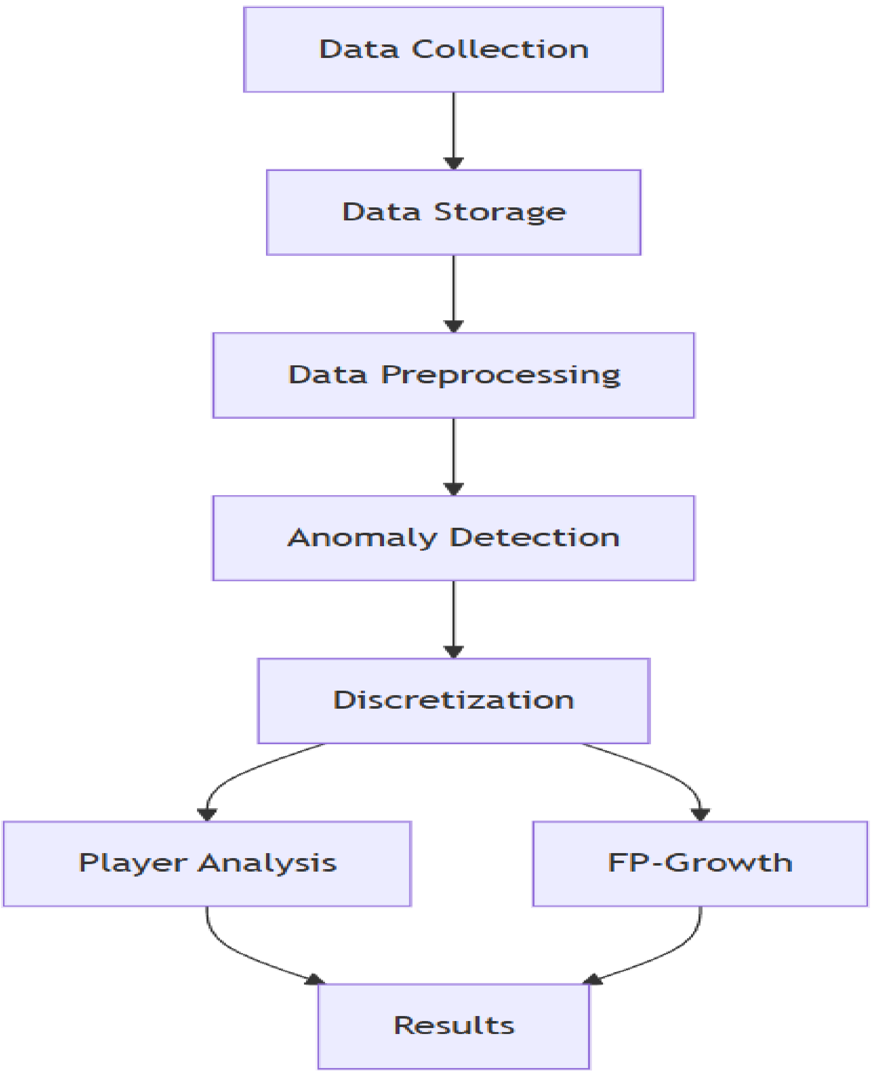

Figure 4: Process Flow Diagram of the Methodology

4.1 Data Collection

This section explains how data was collected to analyze NBA player performance across multiple seasons. The process had two steps: identifying active players for each season and gathering advanced player statistics for each quarter of every game. By focusing only on active players, the analysis stays accurate and relevant. The following sections describe the tools and methods used to ensure the data is reliable and ready for analysis.

4.1.1 Player Identification

For each regular and playoff season from 2004-05 to 2023-24, active players were identified using the commonallplayers endpoint [60]. To ensure that only players who participated in at least one game were included, the leaguegamelog [61] was utilized. This excluded players who were on the roster but never played a game.

```
all_players = commonallplayers.CommonAllPlayers(is_only_current_season=0, league_id='00', season=season)
players_active_in_season = all_players.get_data_frames()[0][players['ROSTERSTATUS'] == 1]

game_logs = leaguegamelog.LeagueGameLog(season=season, season_type_all_star='Regular Season',
player_or_team_abbreviation='P').get_data_frames()[0]
player_ids_with_games = game_logs['PLAYER_ID'].unique()

players_who_played = players_active_in_season[players_active_in_season['PERSON_ID'].isin(player_ids_with_games)]
players_who_played.to_csv(f'nba_players_{season}_names_ids.csv', index=False)
```

Figure 5: Code for Identifying Active Players in a Season

- For each season, game logs were retrieved, and player participation in each game was confirmed.
- Player names were formatted as "Initial. LastName" for consistency, and their IDs were stored for future analysis.

All data was saved as a CSV file for later use.

4.1.2 Per-Quarter Advanced Stats

Advanced player statistics were collected per quarter and overtime using the NBA API's boxscoreadvancedv3 endpoint [62] across regular seasons and playoffs from 2004-05 to 2023-24.

4.1.2.1 Retrieving Player Game Logs and Handling API Errors

The process started by retrieving each player's game log for every season using the playergamelog endpoint [63]. These logs provided game IDs, which were needed to query per-quarter stats via the boxscoreadvancedv3. A retry mechanism was implemented to handle network issues and API timeouts.

```
def get_player_game_ids(player_id, season):
  for _ in range(5):  # retry logic with up to 5 attempts
    try:
      gamelog = playergamelog.PlayerGameLog(player_id=player_id, season=season,
      season_type_all_star='Regular Season')
      return gamelog.get_data_frames()[0]['Game_ID'].tolist()
    except Timeout:
      time.sleep(5)  # wait and retry in case of timeout
      continue
  return []
```

Figure 6: Code for Retrieving Player Game Logs

This prevented temporary API failures from disrupting data collection. It applied up to five retry attempts with pauses between them, ensuring reliable data retrieval.

4.1.2.2 Retrieving Per-Quarter Advanced Stats

Advanced statistics were collected for every quarter of each game using the boxscoreadvancedv3 endpoint. The key to retrieving per-quarter stats was in the use of the range_type=1 parameter, which specifies that the data should be segmented by periods, rather than cumulative game totals.

```
response = boxscoreadvancedv3.BoxScoreAdvancedV3(
  game_id=game_id,
  start_period=period,  # the quarter to get stats for
  end_period=period,   # same value to get stats for a single quarter
  start_range=0,     # fetches entire quarter
  end_range=0,       # ensures the entire quarter is included
  range_type=1       # ensures stats are segmented by period
)
```

Figure 7: Code for Retrieving Per-Quarter Advanced Stats

To ensure that the statistics were associated with the correct player, the response data was filtered by the player's unique id using the personId field.

4.1.2.3 Adding a Column to Track Periods

To track the quarter for which the stats were retrieved, a column named PERIOD was added to the dataset. This column indicated whether the stats corresponded to the 1st, 2nd, 3rd, 4th quarter, or an overtime period.

```python
if not player_stats.empty:
    player_stats['PERIOD'] = period
    all_quarters_stats.append(player_stats)
```

Figure 8: Code for Adding a Column to Track Game Periods

4.1.2.4 Handling Overtime Periods

In cases where a game extended into overtime, the code was designed to handle up to four additional periods. Periods 5 through 8 were treated as overtime periods, and stats were only retrieved if data existed for those periods. If no data was found for a given overtime period, the loop would break, avoiding unnecessary API calls.

```python
for period in range(5, 9):
    response = boxscoreadvancedv3.BoxScoreAdvancedV3(
        game_id=game_id,
        start_period=period,
        end_period=period
    )
    if response is None or response.get_data_frames()[0].empty:
        break
```

Figure 9: Code for Handling Overtime Periods

4.2 Data Storage

Given the large volume of data being processed, intermediate results were saved after each player's data was processed. This ensured that progress would not be lost in the event of an interruption and allowed the process to be resumed at any point. Each player's data was saved as CSV file to Google Drive immediately after processing.

After all players in a season were processed, a final file was created, containing advanced stats for every quarter of every game.

This process was repeated for both regular seasons and playoffs from 2004-05 to 2023-24.

4.3 Data Preprocessing

The collected data was processed to ensure consistency, handle missing values, and prepare it for analysis. This included removing duplicates, filling missing values, adding season information, and merging the datasets into a single file.

4.3.1 Removing Duplicates and Handling Missing Values

First, duplicate rows were identified and removed from each CSV file. After removing duplicates, the cleaned data was saved back to the original CSV files for further processing.

A check for missing values was performed across all datasets. Missing values were primarily found in the 'position' column. To address this, the NBA API's commonplayerinfo endpoint [64] was used to retrieve the player's position based on their unique ID.

```
def get_player_position(person_id, position_cache):
    player_info = commonplayerinfo.CommonPlayerInfo(player_id=person_id)
    position = player_info.common_player_info.get_dict()['data'][0][15]
    return position[0].upper() if position else None
```

Figure 10: Code for Finding Missing Player Positions

The retrieved positions were then used to update the missing values in the CSV files. A cache was implemented to reduce the number of API calls and improve efficiency during this process. Inconsistent positions were identified across players. To resolve this, the most frequent position for each player was reassigned based on their unique ID.

The updated data was saved back to the CSV files. This ensured that both missing and inconsistent values were properly addressed.

4.3.3 Adding Season Information and Combining CSV Files

A new column, 'year,' was added to each dataset to ensure clear identification of the season for which the data was collected. The year information was extracted from the file names.

After processing each individual file, the CSVs were combined into a single file for both regular seasons and playoffs. This step combined all player statistics across multiple seasons into one dataset, making it easier for analysis.

By combining the datasets, the final output provided a comprehensive view of player performance across 20 years of NBA regular seasons and playoffs.

Table 1: Dataset Size and Unique Players

Dataset	Rows	Unique Players
Playoffs	115386	1213
Regular	1670947	1905

4.3.4 Adding Game Results

To include game outcomes in the dataset, a new column, Result, was added. This column displays whether the player's team won ('W') or lost ('L') each game. Game logs were retrieved using the leaguegamelog endpoint, and the results were mapped based on the gameId and teamTricode.

```
def check_game_result(row):
    game_id = str(row['gameId'])
    team_tricode = row['teamTricode']
    return game_results.get(game_id, {}).get(team_tricode, 'no data')

nba_stats['Result'] = nba_stats.apply(check_game_result, axis=1)
```

Figure 11: Code for Adding Game Results

This step ensured that each game was tagged with its corresponding result, allowing for an analysis of performance relative to game outcomes.

4.3.5 Removing Unnecessary Columns

To simplify the dataset and focus only on relevant performance metrics, we removed unnecessary metadata and estimated statistics.

```
columns_to_remove = [
    'teamId', 'teamCity', 'teamName', 'teamSlug',
    'playerSlug', 'comment', 'jerseyNum', 'teamTricode',
    'firstName', 'familyName', 'estimatedOffensiveRating',
    'estimatedDefensiveRating', 'estimatedNetRating', 'estimatedUsagePercentage',
    'estimatedPace', 'pacePer40'
]

df.drop(columns=columns_to_remove)
```

Figure 12: Code for Removing Unnecessary Columns

4.3.6 Filtering Players

To ensure the dataset included only players with a meaningful impact during the season, two filters were applied:

- Players had to participate in at least 20% of their team's games, based on league averages from each season. For example, in 2023-2024, this meant at least 20% of 82 games.

- Players needed to play at least 20% of the total game time (48 minutes per game) in each game they participated in. The minutes column, originally contained time in 'MM:SS' format, was converted into total minutes for correct filtering.

```
# 'MM:SS' format to total minutes
def convert_time_to_minutes(time_str):
    minutes, seconds = map(int, time_str.split(':'))
    return minutes + seconds / 60

# conversion and filtering
min_minutes_required = 48 * 0.2
```

Figure 13: Code for Filtering Players by Minutes Played

This ensured that only players with sufficient playing time and participation were included in the final dataset, with the same approach applied to the playoffs.

Table 2: Filtered Dataset Size and Unique Players

Dataset	Rows	Unique Players
Playoffs	104118	932
Regular	1560372	1624

4.3.7 Anomaly Detection

To refine the dataset and remove outliers, four anomaly detection techniques were applied after standardizing the data: Z-score, Interquartile Range (IQR), Isolation Forest, and DBSCAN. These methods identified statistical outliers that could distort our analysis if left unaddressed.

The following features were selected for anomaly detection:

- minutes
- offensiveRating
- defensiveRating
- netRating
- assistPercentage
- assistToTurnover
- assistRatio
- turnoverRatio
- offensiveReboundPercentage
- defensiveReboundPercentage
- reboundPercentage
- effectiveFieldGoalPercentage
- trueShootingPercentage
- usagePercentage
- pace
- possessions

- PIE

The selected features represented our chosen advanced performance metrics, with minutes included to account for players with limited playing time. This was necessary to prevent skewed metric values caused by small sample sizes. For example, a player who played only a few seconds and made a single shot would have an abnormally high shooting percentage, misrepresenting typical performance.

4.3.7.1 Data Standardization

Before applying anomaly detection, we standardized the data using StandardScaler [65] to ensure all metrics were on the same scale, which is important for distance-based algorithms like DBSCAN and Isolation Forest [66].

4.3.7.2 Z-Score Method

The Z-Score method measured how far each data point was from the mean in terms of standard deviations. Any value with a Z-Score beyond 3 was considered an outlier. This threshold follows the empirical rule, which states that 99.7% of data in a normal distribution falls within three standard deviations. Removing only the most extreme values minimized the risk of discarding valid data points with natural variability. However, its effectiveness decreased when dealing with noisy data or distributions that were not Gaussian [5].

```
def z_score_outliers(df, threshold=3):
    outliers_indices = set()
    for column in metrics_columns:
        z_scores = (df[column] - df[column].mean()) / df[column].std()
        outliers_indices.update(df[np.abs(z_scores) > threshold].index)
    return outliers_indices
```

Figure 14: Code for Applying Z-Score Anomaly Detection

4.3.7.3 Interquartile Range (IQR)

The IQR method detects outliers by measuring the range between the first and third quartiles. Any data point beyond 1.5 times the IQR from the lower or upper quartile is flagged as an outlier. This approach is effective for skewed data, making it useful for detecting anomalies in datasets that do not follow a normal distribution.[5].

```
def iqr_outliers(df):
    outliers_indices = set()
    for column in metrics_columns:
        Q1 = df[column].quantile(0.25)
        Q3 = df[column].quantile(0.75)
        IQR = Q3 - Q1
        outliers_indices.update(df[(df[column] < (Q1 - 1.5 * IQR)) | (df[column] > (Q3 + 1.5 * IQR))].index)
    return outliers_indices
```

Figure 15: Code for Applying IQR Anomaly Detection

4.3.7.4 Isolation Forest

Isolation Forest (IF) is a tree-based anomaly detection technique designed for high-dimensional data. It isolates anomalies by randomly selecting features and partitioning the data into subsets. Anomalies are those points that require fewer partitions (splits) to be isolated. Points requiring fewer partitions to be isolated are considered anomalies [67]. We set a contamination level of **5%**, meaning the model expected 5% of the data to be outliers.

```
def isolation_forest_outliers(df, contamination=0.05):
    iso_forest = IsolationForest(contamination=contamination, random_state=42)
    is_outlier = iso_forest.fit_predict(df[metrics_columns])
    return set(df[is_outlier == -1].index)
```

Figure 16: Code for Applying Isolation Forest Anomaly Detection

4.3.7.5 DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN was applied with an epsilon (eps) of 0.5, which defines the radius for neighbors, and a minimum sample size of 5, which is the minimum number of points required to form a cluster. This setup highlighted anomalies as points lying outside dense clusters.

```
def dbscan_outliers(df, eps=0.5, min_samples=5):
    dbscan = DBSCAN(eps=eps, min_samples=min_samples)
    labels = dbscan.fit_predict(df[metrics_columns])
    return set(df[labels == -1].index)
```

Figure 17: Code for Applying DBSCAN Anomaly Detection

4.3.7.6 Identifying Consensus Outliers

To improve the robustness of the outlier detection process, only data points that were flagged as outliers by at least three of the four methods were considered consensus outliers. These consensus outliers were then removed from the dataset.

After removing the consensus outliers, the dataset was saved for further analysis.

Table 3: Final Dataset Size and Unique Players

Dataset	Rows	Unique Players
Playoffs	89133	932
Regular	1362697	1624

4.3.7.7 Excluding Redundant Columns

After standardization and outlier removal, the dataset was further cleaned by removing columns that were no longer necessary for the analysis. The following columns were deleted:

- total_games
- min_games_played
- game_count
- minutes_total
- minutes
- year

These columns were primarily used during data processing and filtering. Once the dataset was filtered accordingly, these columns became redundant for following analyses, such as discretization and ARM.

4.3.8 Discretization

After anomaly detection, the selected features were discretized into categorical bins to facilitate ARM. Four discretization methods were employed, utilizing Python's cut [68], qcut [69], and KMeans [70] functions.

4.3.8.1 Equal Width Binning

This method divides the continuous values of a feature into intervals of equal size. It is simple to implement but can lead to poor results when the data distribution is skewed or when outliers are present [12], [71].

```
def equal_width_binning(df, column_name):
    df[f'binned_{column_name}'], bin_intervals = pd.cut(df[column_name], bins=3,
    labels=["low", "medium", "high"], retbins=True)
```

Figure 18: Code for Equal Width Binning of Features

4.3.8.2 Equal Frequency Binning

This approach divides the data so that each interval contains an equal number of observations. It helps handle skewed data distributions better than equal width binning but can still be influenced by outliers [12], [71].

```
def equal_frequency_binning(df, column_name):
    df[f'binned_{column_name}'], bin_intervals = pd.qcut(df[column_name], q=3,
    labels=["low", "medium", "high"], retbins=True)
```

Figure 19: Code for Equal Frequency Binning of Features

4.3.8.3 Manual Binning

For certain performance metrics, manual binning was applied. This technique involves using custom thresholds to define bins based on domain knowledge. This was useful when certain metrics (e.g., assistPercentage, assistToTurnover) was highly skewed towards zero [71].

Standard discretization methods, like equal width and equal frequency binning, were not suitable because they could not provide three meaningful bins due to the skewness of the data. For example, a large proportion of the data points might fall into a single bin, making the other bins insignificant.

To tackle this, the distribution and counts of the data were examined. Appropriate thresholds were created to meaningfully differentiate between low, medium, and high values, ensuring each bin contained a sufficient number of observations to be statistically relevant.

The following code shows how bin edges and labels were defined and applied to these features:

```
new_bins = {
    'assistPercentage': ([df['assistPercentage'].min() - 0.01, 0.10, 0.30, df['assistPercentage'].max() + 0.01], ['low',
    'medium', 'high']),
    'assistToTurnover': ([df['assistToTurnover'].min() - 0.01, 0.50, 1.20, df['assistToTurnover'].max() + 0.01], ['low',
    'medium', 'high'])
}

for feature, (edges, labels) in new_bins.items():
    df[f'binned_{feature}'] = pd.cut(df[feature], bins=edges, labels=labels)
```

Figure 20: Code for Manual Binning of Features

- For **assistPercentage**, the bins were set with edges at 0.10 and 0.30, representing meaningful thresholds based on the data's distribution. Values below 10% were considered 'low', between 10% and 30% as 'medium', and above 30% as 'high'.
- For **assistToTurnover**, thresholds at 0.50 and 1.20 were used to create the bins. These values were chosen to effectively separate the data into three categories.

4.3.8.4 K-Means Clustering

K-Means was applied with 3 clusters, 10 separate runs and a random state of 42. These settings ensured consistent results and uncovered natural groupings in the data.

```
# K-Means clustering
def kmeans_clustering(df, column_name):
    kmeans = KMeans(n_clusters=3, n_init=10, random_state=42)
    df[f'{column_name}_kmeans'] = kmeans.fit_predict(df[[column_name]])

  # Define thresholds based on cluster centers
    centers = sorted(kmeans.cluster_centers_.flatten())
    low_threshold = (centers[0] + centers[1]) / 2
    high_threshold = (centers[1] + centers[2]) / 2

  # Assign categories based on thresholds
    conditions = [
      (df[column_name] <= low_threshold),
      (df[column_name] > low_threshold) & (df[column_name] <= high_threshold),
      (df[column_name] > high_threshold)
    ]
    choices = ['low', 'medium', 'high']

    df[f'{column_name}_category'] = np.select(conditions, choices)
```

Figure 21: Code for K-Means Clustering-Based Binning

4.3.8.5 Selecting the Best Discretization Method

Once the four discretization methods were applied, the final decision on the best method for each feature was made by manually reviewing the results in Excel. The method that resulted in the most **meaningful** distribution for each feature was chosen.

The criterion for selection was the range and distribution of data within each bin.

The goal was to choose the method that provided the most balanced and meaningful distribution across the 'low', 'medium', and 'high' categories. This often meant selecting the method with the smallest range within bins.

	Equal Width	Equal Frequency	Manual	K-means
				effectiveFieldGoalPercentage
Low	892016	685576	892016	559408
Medium	411884		124511	528432
High	58797	677121	346170	274857

Figure 22: Example of choosing discretization method for a metric

In this example:

- The Equal Width method resulted in a highly skewed distribution, with the majority of data points in the 'Low' category.
- The Equal Frequency method failed to provide three bins, as indicated by the missing value in 'Medium'.
- The Manual Binning method improved the distribution but still had an imbalance.
- K-Means produced the most balanced distribution across the three categories, with a reasonable number of data points in each bin.

Finally, after discretizing the features, a new CSV file was generated that included all the columns except for the original continuous features that were updated to their binned versions.

4.4 FP-Growth

Due to the memory constraints encountered when using Colab’s TPU, the FP-Growth could not be run on all features simultaneously. As a result, the features were grouped into five distinct categories, based on their function in player performance. This division allowed for a more efficient analysis, while still maintaining meaningful associations within each group. FP-Growth ran separately for the regular season and playoffs datasets, with association rules extracted to detect patterns of performance metrics that are strongly associated with game outcomes.

4.4.1 Offense

- offensiveReboundPercentage
- effectiveFieldGoalPercentage
- trueShootingPercentage
- usagePercentage
- offensiveRating

These metrics measure a player's contributions on offense. Offensive rebound percentage captures second-chance opportunities, while effective and true shooting percentages reflect shooting efficiency. Usage percentage shows how often a player is involved in offensive plays and offensive rating counts their overall offensive impact.

4.4.2 Defense

- defensiveReboundPercentage
- defensiveRating

These metrics measure a player's defensive contributions. Defensive rebound percentage indicates how well a player helps secure possession after an opponent’s missed shot, and defensive rating evaluates the player's overall impact in preventing opponent scoring.

4.4.3 Ball Handling

- assistPercentage
- assistToTurnover
- assistRatio
- turnoverRatio

This category focuses on metrics that measure a player's ability to facilitate plays and handle the ball effectively. Assist-related metrics (assist percentage, assist-to-turnover ratio, assist ratio) represent a player's role in creating scoring opportunities for teammates, while turnover ratio indicates how well they maintain possession.

4.4.4 Overall Impact

- PIE
- netRating
- reboundPercentage

These metrics summarize a player's overall impact on the game. PIE (Player Impact Estimate) gives an overall measure of a player's contributions, net rating balances both offensive and defensive contributions, and rebound percentage captures their effectiveness in securing the ball.

4.4.5 Tempo

- pace
- possessions

These metrics reflect the speed at which the team plays when the player is on the court. Pace measures the number of possessions per game, and total possessions indicate the volume of opportunities a player is involved in.

4.4.6 Running FP-Growth

The following steps were taken to prepare the data and apply the FP-Growth:

- Transforming Overtime Periods:

Before running the FP-Growth, periods 5, 6, 7, and 8 were grouped together and labeled as "5+" due to their low frequency.

- Feature Selection

For each category, only the **relevant** performance metrics were selected. For example, in the Defense, 'binned_defensiveRating' and 'binned_defensiveReboundPercentage' were included.

- Grouping Player Metrics by Game

Each player's performance metrics were grouped by gameId and personId. The performance metrics were aggregated for each player across game periods and combined into a single list.

- Combining Metrics with Period, Result, and Position

Each player's performance metrics were combined with their game results and position. The metrics were linked to game periods to ensure that the algorithm detects performance differences across the game.

- Transaction Encoding

The combined metrics were converted into a transactional format using the TransactionEncoder [72], which transformed the data into a format suitable for the FP-Growth.

- Running FP-Growth

The FP-Growth [73] was executed with a minimum support threshold of **0.005**. Association rules were generated with a minimum confidence of **0.6** and a lift of **1.1** to ensure only meaningful and strong patterns were retained.

- Initial parameters were set with a minimum support of 1%, minimum confidence of 70%, and lift of 1.2 [44] but when attempting to run the FP-Growth, computational limitations were encountered, due to RAM overload.
- To resolve this, we gradually adjusted the parameters downward through trial and error, reducing the minimum support threshold to 0.005 and minimum confidence to 0.6. We set the lift threshold at 1.1 to focus on meaningful associations.

4.4.6.1 Filtering and Sorting Rules

- Filtering Rules

Only the rules where the consequent was W_Result (win) or L_Result (loss) were kept, focusing on patterns that influenced game outcomes.

- Sorting by Period

To improve interpretability across different game periods, we sorted the antecedents were by game period.

The final filtered rules were saved as CSV files, with the process repeated for each performance category in both regular season and playoffs datasets.

4.5 Descriptive Statistics

The following explains how descriptive statistics were generated and how player performance was analyzed. The goal was to identify top-performing players across different performance areas and assess their impact in both regular seasons and playoffs.

4.5.1 Data Loading and Preparation

Regular season and playoff data were loaded from CSV files and combined into a single dataset, with a **"Season"** label added to distinguish between them. All binned metrics were then mapped to numerical values using a predefined mapping of assigning 'low' to 1, 'medium' to 2, and 'high' to 3.

4.5.2 Filtering and Aggregation

To focus on players with meaningful game time, the following filters were applied:

- Minimum Games Played: At least **100** regular-season games or **40** playoff games.
- Quarterly Filtering: For overtime (Period 5+), players needed at least 10 quarters of participation. For other periods, the minimum game criteria applied based on season type.

After filtering, performance metrics were aggregated in two ways. First, by player and season, average values for each binned metric across all five categories were calculated for each player per season. Second, composite scores were computed for each category using a weighted average, where more influential metrics within their categories were given greater weight. For metrics where lower values indicate better performance (defensiveRating, turnoverRatio) values were inverted so that higher scores indicate better performance. This inversion was applied only during the composite score calculation to align with the scoring system.

Category Weights

- **Offense**
  - Effective Field Goal Percentage: 20%
  - True Shooting Percentage: 20%
  - Usage Percentage: 30%
  - Offensive Rebound Percentage: 10%
  - Offensive Rating: 20%
- **Defense**
  - Defensive Rebound Percentage: 30%
  - Defensive Rating: 70%
- **Ball Handling**
  - Assist Percentage: 40%
  - Assist to Turnover Ratio: 30%
  - Assist Ratio: 20%
  - Turnover Ratio: 10%
- **Overall Impact**
  - PIE: 45%
  - Net Rating: 45%
  - Rebound Percentage: 10%

- **Tempo**
  - Pace: 60%
  - Possessions: 40%

These weights represent the rank of each metric within its category based on their impact on team success and player contribution. For offense, usage percentage is given higher weight (30%) as it captures a player's involvement in scoring opportunities while shooting efficiency metrics are prioritized equally (20%) to emphasize scoring precision. For defense, defensive rating is heavily weighted (70%) due to its comprehensive measure of defensive impact. In ball handling, assist percentage carries the most weight (40%) as it indicates playmaking ability, supported by metrics related turnovers and passing efficiency. Overall impact balances holistic metrics like PIE and net rating equally (45%), while tempo prioritizes pace (60%) to reflect the strategic emphasis on game speed, with possessions having a supplementary measure.

4.5.3 Identifying Top Players and Visualization

Top players were identified using composite scores, selecting the top 5 players in each performance category for both regular seasons and playoffs.
Complete data can be found in Appendix D, Table 9 (Ball Handling), Table 10 (Defense), Table 11 (Offense), Table 12 (Overall Impact), and Table 13 (Tempo).

To present the findings clearly, charts and tables were used:

- Bar Plots: Comparing composite scores of top players between regular seasons and playoffs to visualize performance differences.

- Summary Tables: Displaying individual metric scores, composite scores, games played, and performance across game periods for each top player to provide detailed rankings.

5 Results

This section summarizes the findings from the ARM analysis of player performance data from the 2004-05 to 2023-24 seasons, covering both the regular season and playoffs. The results are grouped into 5 categories: Ball Handling, Defense, Offense, Overall Impact, and Tempo, making it easy to compare performance across different game periods and player positions. The analysis includes tables and visuals to show patterns and trends. Examples of standout players are also included to show how their performances match the identified trends. Appendix C, Table 7 and Table 8, list the top association rules for regular season and playoffs and Appendix D includes detailed player scores by category and quarter.

5.1 Rule Counts

To gain an initial understanding of the association frequencies, we summarized the total number of rules generated for each performance category during the regular season and playoffs.

Table 4: Total Number of Rules per Category

Category	**Regular Season**	**Playoffs**
Ball Handling	7	2829
Defense	2904	2903
Offense	173435	220715
Overall Impact	17154	18480
Tempo	2	8

Support Thresholds and Dataset Size:

- Since the same minimum support threshold (0.005) was applied for both datasets, differences in rule counts are largely due to the smaller playoff dataset. For instance, fewer rules in Ball Handling during the regular season (7 rules) compared to the playoffs (2,829 rules) suggest that support threshold and dataset size significantly impact rule generation.

Trends Across Categories:

- The **Offense** category, with 173,435 rules in the regular season and 220,715 in the playoffs, shows how a large dataset with a low support threshold can generate a high volume of associations. Offense metrics show strong associations with game outcomes across both seasons, indicating their significant role in determining results.
- **Defense** has nearly identical rule counts between regular season and playoffs, suggesting that defensive associations with game outcomes are stable and less impacted by the dataset size.
- **Ball Handling**: Higher rule counts in playoffs imply changes in gameplay strategies that emphasize possession control under increased defensive pressure.
- **Overall Impact**: More rules in the playoffs indicate that cumulative performance factors become crucial in postseason games.

5.2 Top Rules by Category

This section analyzes the top association rules across Ball Handling, Defense, Offense, Overall Impact, and Tempo. These rules help show which metrics guide winning outcomes and reveal standout players.

Ball Handling

In the regular season, the top association rules for Ball Handling are characterized by lift values exceeding **1.20**, indicating a significant association with game outcomes (Table 7). **Rule 1**, with a lift of **1.3146**, shows that a combination of a low assist percentage and medium assist ratio in **Period 3** is associated with winning results. This rule has a confidence of **65.73%** and a support of **0.0057**.

This suggests that during the regular season, teams benefit when players balance ball distribution and shot creation at key points in the game.

In the playoffs, the lift values for Ball Handling rules increase, with top rules demonstrating lifts around **1.3570** (Table 8). For example, **Rule 1** indicates that a combination of a medium assist percentage in Period 1, low assist-to-turnover ratio in Period 3, high assist ratio in Period 3, and medium assist percentage in Period 3 leads to wins. This rule has confidence of **67.76%** and a support of **0.0053**.

This pattern suggests a heavier focus on ball security and distribution under playoff pressure, especially in the third quarter.

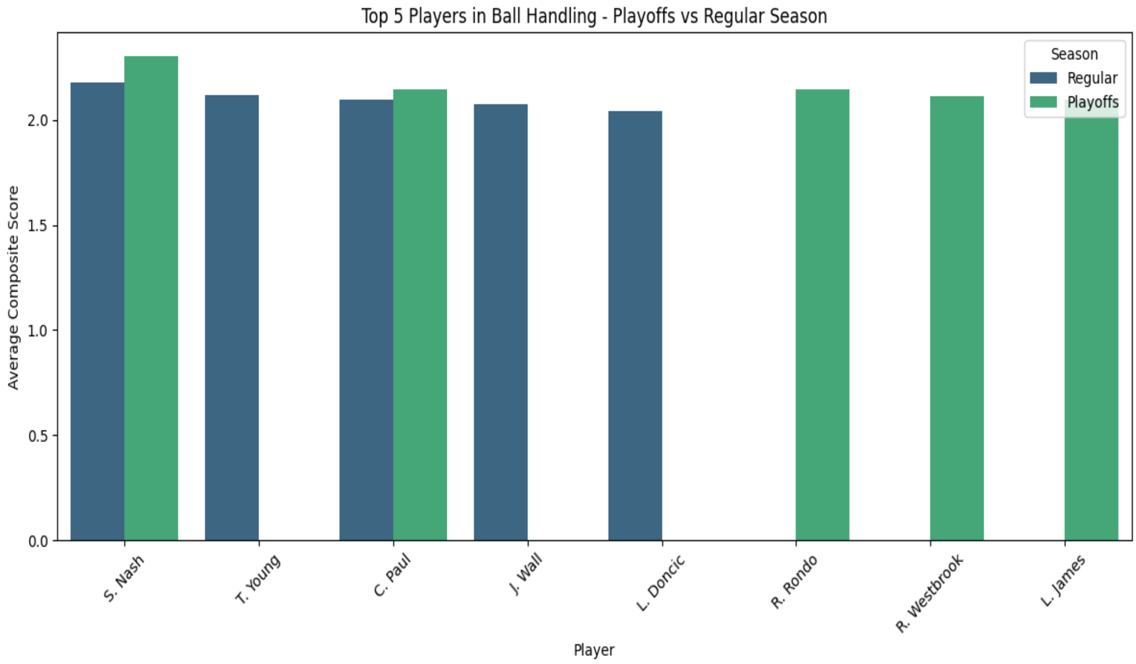

Figure 23: Top Ball Handling Players

Player Highlights:

- **Steve Nash** excels in the playoffs with a Composite score of **2.302**. Over 66 playoff games, Nash maintains a high assist percentage of **2.676** and a low turnover ratio of **1.743**, displaying excellent playmaking under pressure.
- **Trae Young** appears as an important ball handler in the regular season, achieving a Composite Ball Handling score of **2.122** over **404 games**. His assist percentage of **2.472** and assist-to-turnover ratio of **1.696** indicate his influence on the offensive flow.
- **Chris Paul** performs strongly in both the regular season and playoffs. In the regular season, he achieves a Composite Ball Handling score of **2.095** across **1,257 games**, with an assist percentage of **2.445** and an assist ratio of **2.156**. In the playoffs, his score is **2.145** over **149 games**, maintaining high assist metrics and steady ball control.
- **R. Rondo**, known for his playoff skills, achieves a high assist ratio (**2.444**) and balanced turnover ratio (1**.616**) across 129 playoff games.

  Please see Appendix D, Table 9

Defense

Regular-season rules show lift values up to **1.6555** (Table 7). **Rule 1** indicates that players with high defensive ratings in Periods 1 to 3 combined with medium defensive rebound percentages in Period 1 are associated with losses. This rule has a support of **82.77%** and a confidence of **0.0050**. This suggests that higher defensive inefficiencies correlate strongly with losing outcomes.

Lower defensive ratings with medium defensive rebound percentages are linked to winning outcomes. For instance, **Rule 2** has a lift of **1.6426**, indicating that players with low defensive ratings in Period 1, medium defensive rebound percentages in Period 1, and low defensive ratings in Periods 3 and 4 are associated with winning, with a support of **82.13%**.

In the playoffs, the lift values for Defense rules increase further, with the top rule having a lift of **1.7516** (Table 8). **Rule 1** shows that forwards with high defensive ratings in Periods 1, 3, and 4 are associated with losses, indicating that defensive inefficiencies negatively impact playoff success.

These findings show that low defensive ratings (i.e., better defensive performance) and rebounding are key to success in both the regular season and playoffs.

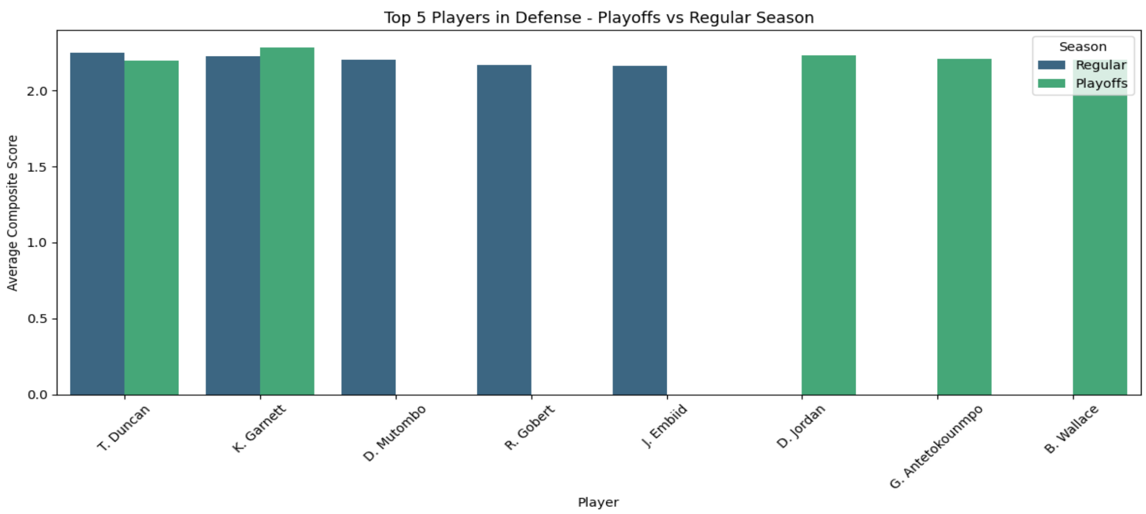

Figure 24: Top Defense Players

Player Highlights:

- **Kevin Garnett** excels defensively in both the regular season and playoffs. In the regular season, he has a defensive rating of **1.819** and a defensive rebound percentage of **2.319** over 756 games. In the playoffs, his defensive rating improves to **1.791**, and his defensive rebound percentage to **2.452** over **96 games**, supporting his consistent defensive ability.

- **Tim Duncan** achieved the highest Composite Defense score of **2.248** over **868 games**. He maintains a defensive rebound percentage of **2.323** and the best defensive rating of **1.784**, highlighting his long-term defensive impact.

- **Giannis Antetokounmpo** shines in the playoffs with a defensive rebound percentage of **2.521** and a defensive rating of **1.927** over 79 games. His ability to dominate the defensive side and the boards is remarkable.

- **Ben Wallace** and **Dikembe Mutombo**, widely regarded as two of the best defenders in NBA history, showcase their defensive power. Wallace achieves a strong playoff defensive rating of 1.811 and a defensive rebound percentage of 2.235 over 76 games, emphasizing his tenacity on defense. Mutombo, known for his dominant shot-blocking and rebounding, has an impressive defensive rating of 1.735 and a defensive rebound percentage of 2.048 in the regular season over 186 games.

  Please see Appendix D, Table 10

Offense

In the regular season, the top rules for Offense exhibit high lift values, with the highest being **1.7077** (Table 7). **Rule 1** indicates that players with high offensive ratings across all periods (Periods 1 to 4) are strongly associated with winning outcomes, with a confidence of **85.39%** and a support of **0.0051**.

Rules involving high effective field goal percentage and true shooting percentage also have high lifts. For example, **Rule 2** has a lift of **1.6517**, showing that high offensive ratings in Periods 1 and 3, combined with high effective field goal percentage in Period 4, are associated with wins.

In the playoffs, the lift values for Offense rules increase, with the top rule having a lift of **1.8014** (Table 8). **Rule 1** indicates that medium effective field goal percentage and true shooting percentage in Period 1, combined with low offensive ratings in Periods 1 to 3, are associated with losses. This shows how important high offensive ratings are to avoid losses in playoff games.

Rules involving high offensive ratings and true shooting percentages continue to be associated with winning outcomes. For example, **Rule 3**, with a lift of **1.7732**, shows that high offensive ratings in Periods 1, 3, and 4, along with high effective field goal percentage in Period 3 and low offensive rebound percentages in Periods 2 and 4, are linked to wins.

These findings reveal that scoring efficiency and offensive consistency are key to wins, especially in the playoffs.

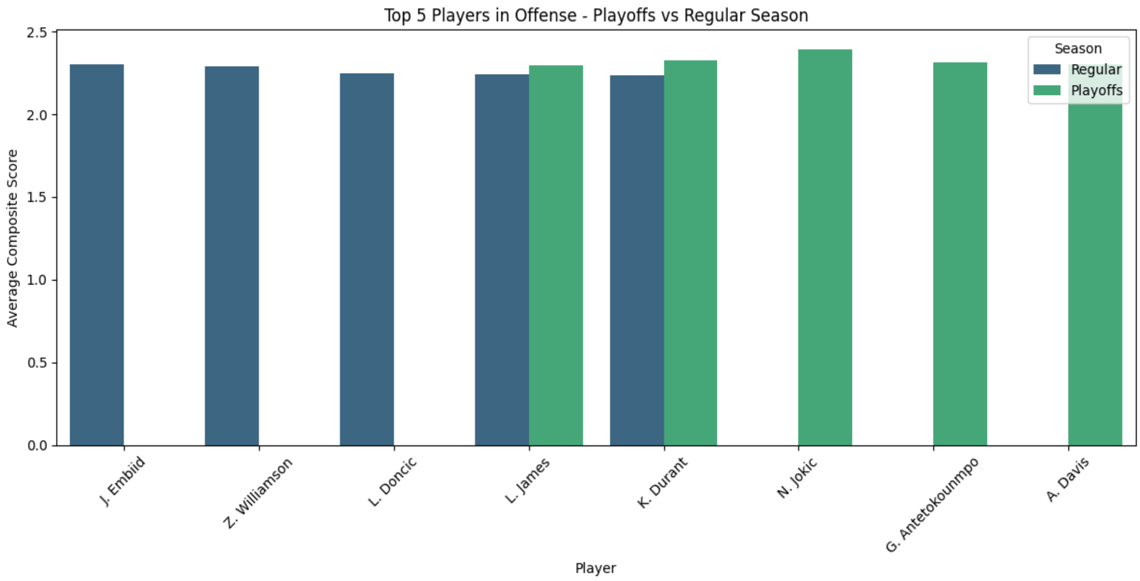

Figure 25: Top Offense Players

Player Highlights:

- **LeBron James** has incredible consistency across both the regular season and playoffs. During the regular season, he maintains a composite offense score of 2.241 over 1,408 games, with an effective field goal percentage of 2.009 and a true shooting percentage of 2.313. In the playoffs, he slightly improves to a composite offense score of 2.395 over 286 games, with an effective field goal percentage of 2.317 and a true shooting percentage of 2.277, highlighting his ability to adapt in the postseason.

- **Kevin Durant** excels offensively in both regular season and playoff contexts. Over 1,058 regular-season games, he achieves a composite offense score of 2.238, with an effective field goal percentage of 2.027 and a true shooting percentage of 2.379. In the playoffs, Durant’s offensive ability peaks with a composite offense score of 2.326 over 170 games, featuring an effective field goal percentage of 2.354 and a true shooting percentage of 2.359, supporting his elite scoring skill in high-stakes situations.

- **Nikola Jokic** stands out in the playoffs, achieving the highest composite offense score among the listed players at 2.391 over 79 games. His effective field goal percentage of 2.457 and true shooting percentage of 2.386 illustrate his offensive efficiency, while a usage percentage of 2.646 and an offensive rating of 2.278 confirm his role as a central playmaker and scorer for his team.

  Please see Appendix D, Table 11

Overall Impact

In the regular season, the top rules for Overall Impact have the highest lift values among all categories, with lifts approaching **1.9497** (Table 7). **Rule 1** indicates that high net ratings and high Player Impact Estimate (PIE) in Periods 5+ are strongly associated with winning outcomes, with a confidence of **97.49%** and a support of **0.0064**.

Low net rating values, when combined with specific positions or medium rebound percentages, are associated with losses. For example, **Rule 4** shows that guards with low net ratings in Periods 5+ are linked to losses, with a lift of **1.9262**.

In the playoffs, we observe similar patterns, with lift values remaining high. **Rule 1** has a lift of **1.9892** (Table 8)., indicating that high PIE and high net ratings across Periods 1 to 3 are strongly associated with wins, with a confidence of **99.32%** and a support of **0.0053**.

These rules emphasize that achieving a constantly high impact on the court, as measured by net rating and PIE, often results in wins.

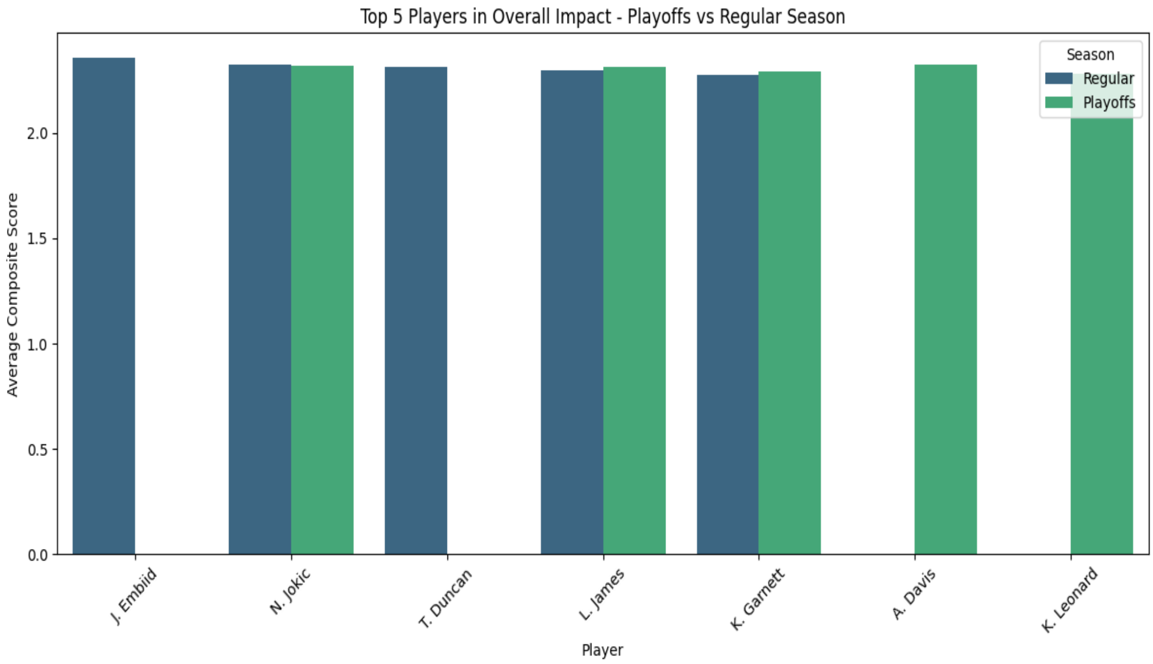

Figure 26: Top Overall Impact Players

Player Highlights:

- **Joel Embiid** leads with a composite overall impact score of **2.355** in the regular season, with a PIE of **2.525**, net rating of **2.149** and rebound percentage of **2.516** over **428 games**.

- **Nikola Jokic** follows closely with a composite overall impact score of **2.323** in the regular season and maintains this level in the playoffs. In the playoffs, Jokic's PIE is **2.561**, the highest among all players, and his net rating is **2.040** over **79 games**.

- LeBron James has exceptional versatility and longevity once again, ranking among the top players with a composite overall impact score of 2.293 in the regular season across 1408 games. In the playoffs, he raises his performance with a score of 2.313.

- Anthony Davis excels in the playoffs with a composite overall impact score of 2.321, supported by a PIE of 2.495, a strong net rating of 2.099, and the highest rebound percentage of 2.533 among playoff players, exhibiting his dominance over 59 games.

  Please see Appendix D, Table 12

Tempo

In the regular season, the top rules for Tempo have lift values around **1.2350** (Table 7). **Rule 1** indicates that guards with high possessions in Period 1 and low possessions in Period 3 are associated with losses, with a confidence of **61.75%** and a support of **0.0053**. This suggests that pacing fluctuations may hurt game outcomes.

In the playoffs, lift values for Tempo rules are similar, with slightly higher values. For example, **Rule 1** has a lift of **1.2618** (Table 8)., indicating that high possessions in Period 2, medium possessions in Period 3, and low possessions in Period 4 are associated with losses, with a confidence of **63.18%**.

These patterns imply that sustaining a consistent possession strategy matter, with playoff success depending more on stable pacing across periods. The data indicates that teams may benefit from sustaining a high-paced approach throughout the game rather than changing their possession rates.

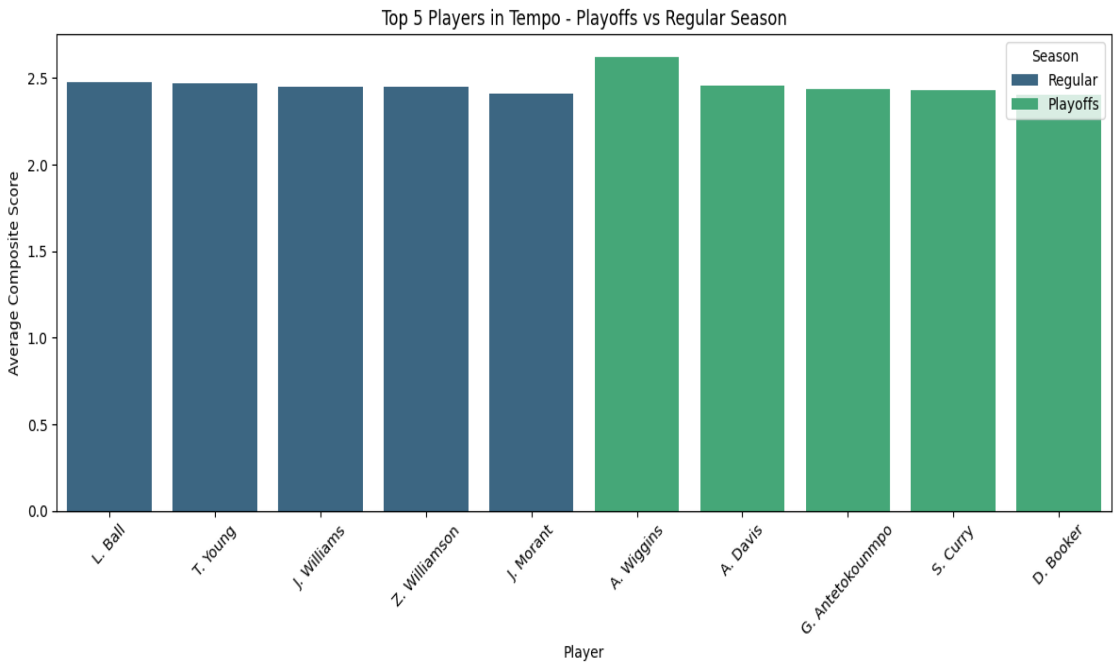

Figure 27: Top Tempo Players

Player Highlights:

- **LaMelo Ball** leads the regular season with a Composite Tempo score of **2.475**, achieving a pace of **2.513** and possessions of **2.418** over **183 games**. His control of the game's tempo over a high number of possessions showcases his impact on the offensive flow.

- **Trae Young** follows closely in the regular season with a Composite Tempo score of **2.471**, a pace of **2.479**, and possessions of **2.460** over **404 games**.
- **Stephen Curry** stands out in the **playoffs** with a **Composite Tempo score of 2.428** across **147 games**. His **pace of 2.3675** and **possessions of 2.506** confirm that he keeps an aggressive tempo.

  Please see Appendix D, Table 13

5.3 Quarter-by-Quarter Comparison

The following analysis explores how the average lift of performance metrics within each category changes across quarters during the regular season and playoffs.

Ball Handling

As shown in Figure 28, the average lift for Ball Handling peaks during the second quarter of the regular season, reaching approximately **1.31**, while the playoffs display a more stable pattern across quarters with lifts around **1.27**. This indicates that specific ball-handling metrics have a greater impact on game outcomes during the second quarter of regular-season games.

In the **second quarter of the regular season**, Steve Nash showed his elite ball-handling skills, leading with a Composite Ball Handling score of 2.162, an assist percentage of 2.495, and an assist ratio of 2.267 over 457 quarters (Table 9).

In the **playoffs' second quarter**, Chris Paul stands out with a Composite Ball Handling score of 2.194 over 123 quarters, with an assist percentage of 2.553 and an assist ratio of 2.512.

In the **regular season's fourth quarter**, Deron Williams shows his composure and playmaking ability, achieving a Composite Ball Handling score of 1.999 over 571 quarters. His assist percentage of 2.289 and assist ratio of 1.949.

In the **playoffs' fourth quarter and overtime periods**, Rajon Rondo excels as a key playmaker, recording a Composite Ball Handling score of 2.074 over 86 quarters in the fourth and 1.670 over 10 overtime periods. With an assist percentage of 2.360 in the fourth quarter and 1.800 in overtime, along with assist ratios of 2.395 and 1.800, Rondo leads under pressure, adapts to extended game situations and supports his team’s offensive success.

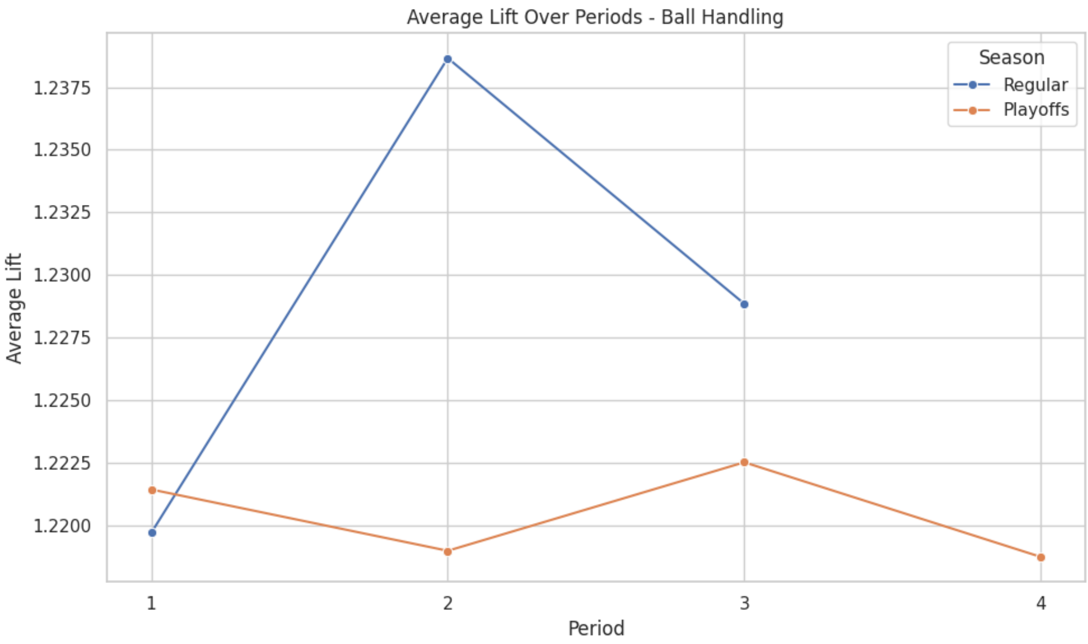

Figure 28: Ball Handling Lift by Quarter and Season

Defense

Figure 29 shows that the average lift values for Defense increase sharply during the fourth quarter in both the regular season and playoffs, peaking at approximately **1.65** in the regular season and **1.75** in the playoffs. This trend suggests that defensive performance becomes more influential in the later stages of games.

In the regular season, in the fourth quarter, **Tim Duncan** shows strong defensive impact with a Composite Defense score of **2.167** over **524 quarters** (Table 10).

Dwight Howard exhibits increased defensive metrics during the fourth quarter of playoff games, with a Composite Defense score of **2.265**, a defensive rebound percentage of **2.591**, and a defensive rating of **1.875** over **88 quarters**. His effectiveness in protecting the rim and securing defensive rebounds was incredible.

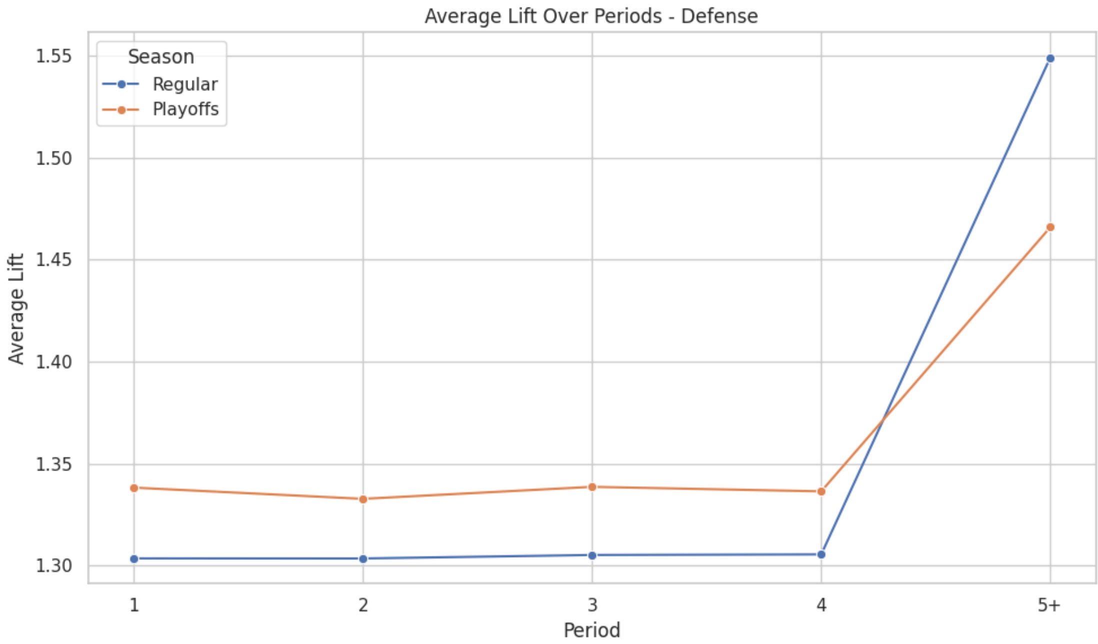

Figure 29: Defense Lift by Quarter and Season

Offense

Figure 30 shows that the average lift values for Offense increase in the fourth quarter, reaching around **1.77** at the playoffs. This suggests that offensive performance becomes impactful toward the end of games.

In the playoffs, **Nikola Jokic** consistently shows a strong offensive impact in late-game situations. In the fourth quarter, his Composite Offense score peaks at **2.410**, with an effective field goal percentage of **2.375** and true shooting percentage of **2.354** over **48 quarters** (Table 11).

Stephen Curry stands out offensively in the playoffs, with a Composite Offense score of **2.325** in the fourth quarter and an effective field goal percentage of **2.357** over **115 fourth quarters**. His consistent performance supports his reputation as an elite late-game scorer.

Kevin Durant has a remarkable offensive impact on the playoffs. During the fourth quarter, his Composite Offense score reaches **2.338**, highlighted by an offensive rating of **2.195** and a true shooting percentage of **2.376** over **133 quarters** (Table 11).

In playoff overtime, both Damian Lillard and LeBron James exhibit high offensive scores. Lillard's Composite Offense score of 2.264, combined with an effective field goal percentage of 2.273 and a true shooting percentage of 2.273 across 11 overtime periods. LeBron James has a Composite Offense score of 2.480 in overtime, complemented by an effective field goal percentage of 2.500 and a usage percentage of 3 over 10 periods.

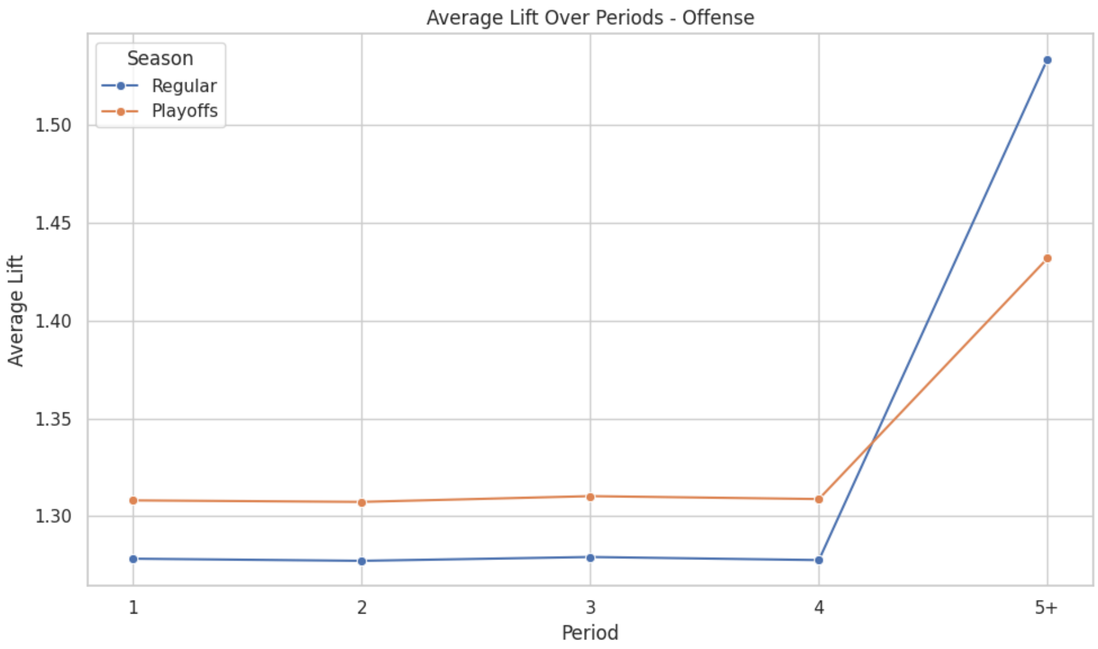

Figure 30: Offense Lift by Quarter and Season

Overall Impact

Figure 31 illustrates that the Overall Impact category exhibits a strong increase in lift in the fourth quarter, peaking in period 5+ for regular season games at nearly **1.95** and remaining high in the playoffs. This aligns with the idea that a player's comprehensive performance becomes increasingly decisive as games approach the final stages.

In the playoffs, **LeBron James** and **Kawhi Leonard** emerge as closers in the fourth quarter. LeBron's Composite Overall Impact score is **2.335** over **204 quarters**, paired with a PIE of **2.520** and a net rating of **2.142**. **Kawhi Leonard** records a Composite Overall Impact score of **2.323**, with the highest net rating (**2.250**) among the players analyzed and a PIE of **2.420** over **100 quarters** (Table 12).

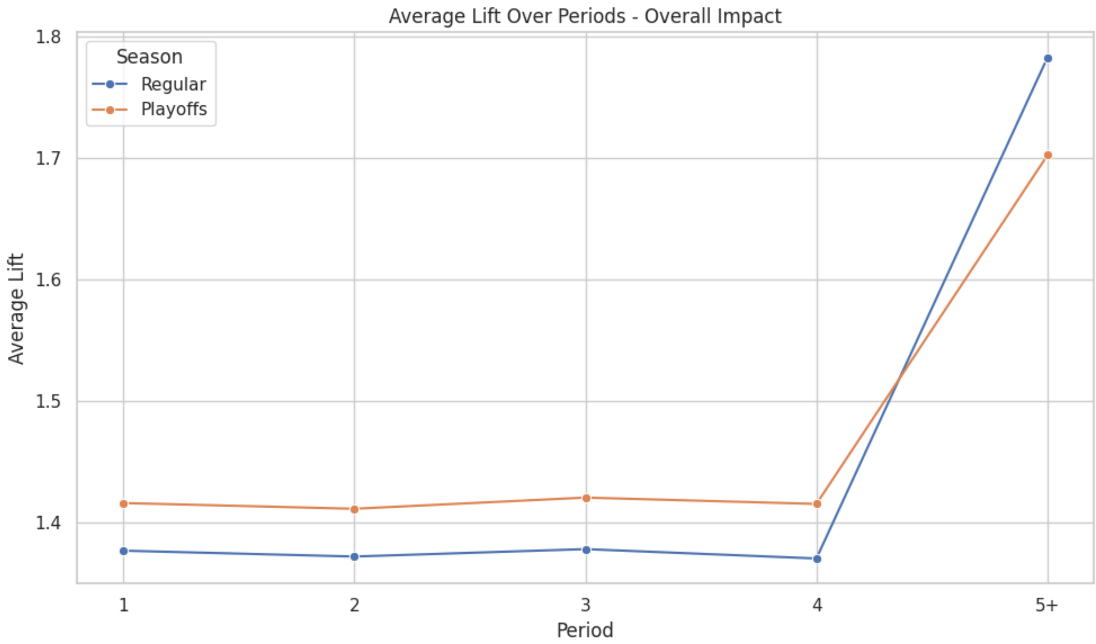

Figure 31: Overall Impact Lift by Quarter and Season

Tempo

Figure 32 shows that for Tempo, the regular season has variability across quarters, with a dip in lift during the second quarter to around **1.23**, whereas playoff lift values steadily increase up to the third quarter before decreasing in the fourth. This trend hints that possession and pacing strategies are slightly more influential in the middle quarters of playoff games.

During the first quarter of the regular season, **Trae Young** and **Stephen Curry** lead with a Composite Tempo score of **2.630**. Trae Young achieves this over **294 quarters**, setting an aggressive pace for his team with a pace of **2.578** and possessions of **2.707**. Stephen Curry matches this score over **793 quarters** (Table 13).

In the playoffs, Stephen Curry and Klay Thompson rank among the best in composite score during the first two quarters, showing how the Splash Brothers set the tone early in games.

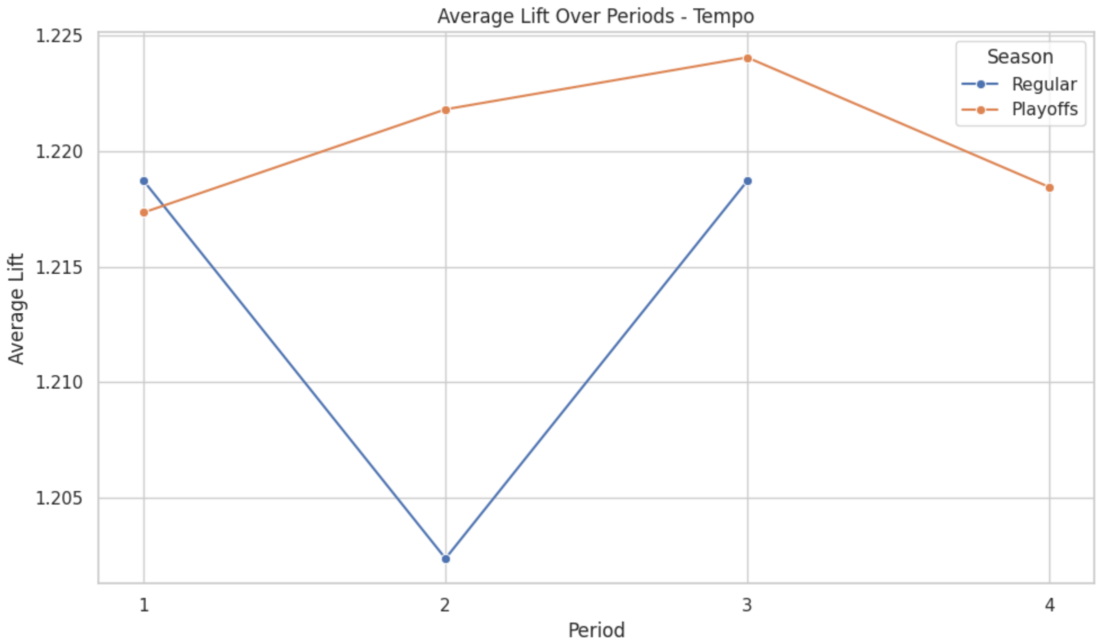

Figure 32: Tempo Lift by Quarter and Season

5.4 Position-Based Breakdown

This section analyzes how different player positions Guards, Centers, and Forwards affect game outcomes across performance categories during the regular season and playoffs.

Guards

In the regular season, offensive ratings play a dominant role for Guards. Rules involving high offensive ratings in early periods are associated with winning outcomes, with lifts around **1.6096** (Table 7). For instance, **Rule 1** for Offense/Guards indicates that Guards with high offensive ratings in Periods 1 and 2, low offensive rebound percentages in Periods 2 and 3, and high offensive ratings in Period 3 are associated with wins.

Defensively, Guards with high defensive ratings in Periods 1 to 3 are associated with losses, as seen in **Rule 1** for Defense/Guards with a lift of **1.6002**.

In the playoffs, Guards with medium assist percentages in Period 1 and high assist ratios in Period 3 are associated with winning outcomes, as seen in **Rule 1** for Ball Handling/Guards with a lift of **1.2713** (Table 8). Guards with medium turnover ratios in Period 1 and high turnover ratios in Period 2 are associated with losses, revealing how crucial is to control turnovers.

Centers

In the regular season, Centers with high offensive ratings in Periods 1 and 3 are associated with wins, as shown in Rule 2 for Offense/Centers with a lift of **1.4195**. Centers with low offensive ratings in Periods 1 and 3 are associated with losses.

Defensively, Centers with high defensive ratings in early periods are associated with losses, as seen in **Rule 1** for Defense/Centers with a lift of **1.4487** (Table 7), emphasizing the need for defensive consistency for Centers.

In the playoffs, Centers with high rebound percentages and high net ratings in early periods are associated with wins, as seen in **Rule 1** for Overall Impact/Centers with a lift of **1.7565**. showing how valuable strong rebounding and overall impact is (Table 8).

Forwards

During the regular season, Forwards with high offensive ratings in early periods are associated with wins, with lifts around 1.6053 (Table 7). Conversely, high defensive ratings in early periods are associated with losses, as shown in Rule 1 for Defense/Forwards with a lift of 1.6182.

In the playoffs, Forwards with low turnover ratios and high assist ratios in early periods are associated with winning outcomes, as seen in **Rule 1** for Ball Handling/Forwards with a lift of **1.3112** (Table 8). High net ratings in initial periods also contribute to victories.

For Forwards, offensive efficiency and ball control correlates with wins, especially in the playoffs.

5.5 Average Lift Values Across Categories

This section analyzes the average lift across the five performance categories during both the regular season and playoffs. By evaluating these lift values, we aim to identify which categories influence game outcomes the most and how their impact shifts between different stages of the competition.

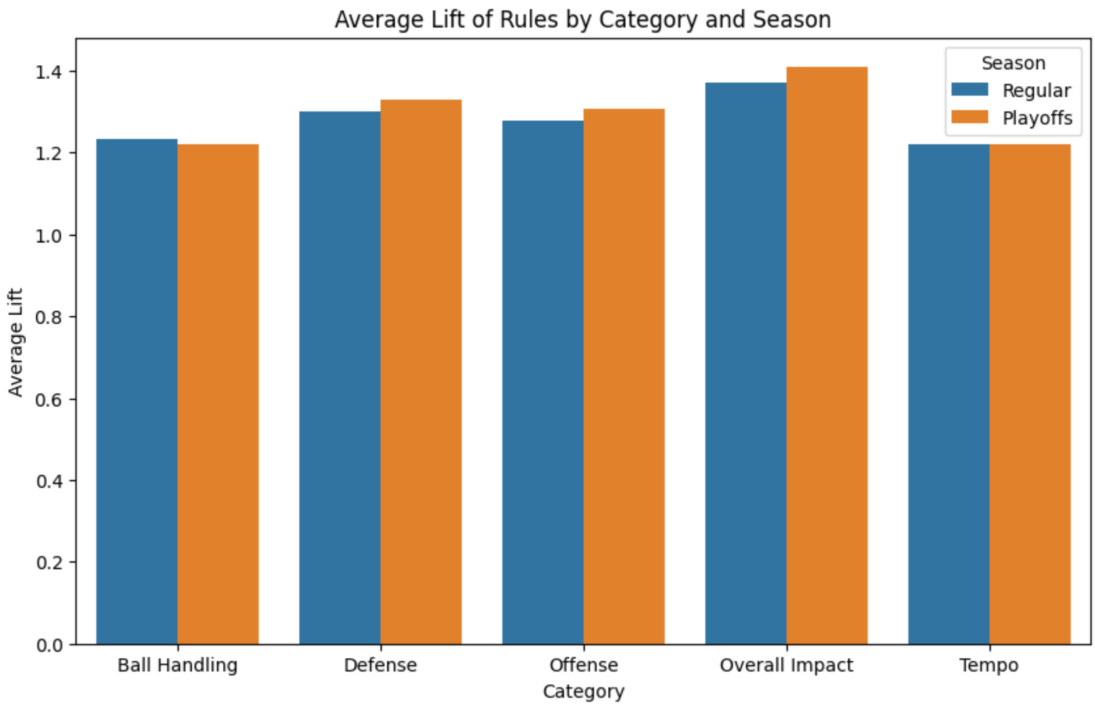

Figure 33: Lift by Category and Season

As we can see from Figure 33:

- **Overall Impact** has the highest average lift in both the regular season (around **1.35**) and playoffs (around **1.40**), highlighting the value of cumulative metrics in capturing a player's overall contribution to game outcomes.

- **Defense** shows a slightly higher average lift than **Offense** across both seasons, with lifts around **1.29** in the regular season and **1.32** in the playoffs for Defense, compared to **1.27** and **1.30** for Offense. This suggests that defensive performance matters more for game outcomes than offensive metrics.

- **Ball Handling** and **Tempo** have the lowest average lift values, both hovering around **1.22**, indicating their lower direct influence on game outcomes.

We can clearly observe that while all performance categories influence game outcomes, **Overall Impact** and **Defense** are more significant.

5.6 Top Overall Rule for Each Season

Playoffs:

- **Top Rule**: "high_PIE_Period1, high_netRating_Period1, high_netRating_Period2, high_PIE_Period2, high_netRating_Period3 → W_Result"
- **Support**: 0.0053
- **Confidence**: 0.9932
- **Lift**: 1.9892

Players with high PIE and net ratings in the first three periods form the strongest playoff association, indicating that strong early-game performance increases the likelihood of winning in the playoffs.

Regular Season:

- **Top Rule**: "high_netRating_Period5+, high_PIE_Period5+ → W_Result"
- **Support**: 0.0064
- **Confidence**: 0.9749
- **Lift**: 1.9497

In the regular season, the top rule reveals that high net ratings and PIE in late-game or overtime stretches (Period 5+) strongly predict regular-season wins.

6 Discussion

This section explains the study's main findings and how they apply to professional basketball. It starts by looking at performance metrics and how they affect game outcomes, then it explores the strategic differences between regular season and playoff performances. Individual player performances are also analyzed, with a focus on standout players and how coaching strategies contribute to team success. The data science perspective is incorporated to emphasize the analytical methods used in the study. The section also discusses the study's limitations and potential threats to its validity. By linking the data to real-life basketball situations and existing research, it provides a clear understanding of what drives success in both regular and playoff games.

6.1 Interpretation

The analysis of NBA player performance points to which metrics carry the most weight in winning games during the regular season and the playoffs. By using composite scores, association rules, and position-specific data across different quarters, we can explain why some trends emerge and relate them to established studies. This discussion connects the numerical findings to actual basketball scenarios, tactics and dynamics, while comparing them with previous studies.

6.1.1 Strategic Context and Game Dynamics

Our findings show that defensive performance becomes more critical in the playoffs, as evidenced by higher lift values for defensive rules. This aligns with Dehesa et al. [51], who identified differences between regular season and playoff performances, in metrics like defensive ratings and effective field-goal percentage. Their analysis revealed how performance changes across game quarters, influenced by situational variables and game dynamics. Extending their work, we analyzed all quarters and found that strong defense matters throughout the playoffs, not limited to specific periods. The postseason's slower pace, higher stakes, and defensive intensity play a key role in shaping game outcomes.

In these games, every possession counts, making metrics like offensive rating and effective field goal percentage play key roles in determining success. This aligns with Ruano et al. [50], who found that performance metrics shift during the 4th quarter under different game conditions.

While their study used statistical methods, like discriminant analysis to uncover performance trends, our ARM application offers complementary insights into how certain metrics influence game outcomes each quarter. By examining composite scores and association rules across performance categories, the findings reveal that successful playoff teams rely more on defense and efficiency than in the regular season.

6.1.2 Player Performance

Star players like Tim Duncan, Nikola Jokić, and Stephen Curry stand out for their success, shaped by coaching, team culture, and the quality of teammates. Tim Duncan's excellent defensive ratings and rebound percentages confirm his consistent impact on the game, aided by the Spurs' system emphasizing team defense and unselfish play. While Ruano et al. [50], analyzed how situational variables and performance metrics influence outcomes during the 4th quarter of close games, our analysis broadens this approach by examining how players' metrics contribute to game outcomes across all quarters.

Giannis Antetokounmpo’s playoff defense reinforces the value of versatile players who can adjust to different in-game demands. His ability to dominate the boards and defend multiple positions under playoff pressure underscores the value of adaptable defenders in securing postseason wins. Top performers like Kawhi Leonard and LeBron James thrive under playoff pressure by adapting their game to tougher defenses and high-stakes situations. This reflects Ben Taylor's [54] emphasis that adaptability, whether on offense or defense, is necessary for succeeding against tough competition in the playoffs.

Stephen Curry maintains his reputation as an elite scorer, especially in high-pressure situations. His ability to sustain an aggressive tempo and deliver scoring in the postseason is pivotal for his team's success. This aligns with Sarlis et al. [52], who identified Curry as a consistent top performer in clutch situations.

Players like Zion Williamson and Joel Embiid struggle with playoff efficiency due to injuries, tougher defenses, and the extra physical demands of postseason play. As evidenced by Sarlis et. al [41], injuries not only disrupt individual contributions but also hinder team outcomes in high-stakes scenarios like the playoffs.

6.1.3 Coaching Influence and Team Dynamics

Coaching styles and systems play a major role in the observed trends. Coaches who prioritize defense and efficient offense usually see their teams perform better in the playoffs. The Spurs' emphasis on defense and ball movement under Coach Gregg Popovich, aligns with the high lift values associated with defensive and offensive metrics in our study. Ben Taylor [54] notes Popovich's strategic adjustments and ability to build team chemistry as central to his success, which mirrors the patterns we observe.

Similarly, Steve Kerr's high-paced, high-scoring style enables the Golden State Warriors to maintain high offensive ratings even under playoff pressure. His system focuses ball movement, spacing, and off-ball actions, principles Taylor [54] identifies as critical for countering the tighter defenses encountered in the playoffs. This adaptability allows Kerr's teams to continue generating quality shots despite tighter postseason defense.

LeBron James thrives in systems designed to leverage his versatility and leadership. Coaches like Erik Spoelstra and Frank Vogel have developed strategies that amplify his strengths, including playmaking, multi-positional defense, and clutch decision-making. Ben Taylor [54] points out that building around players unique styles is important, which helps explain LeBron's consistent playoff success.

Teamwork and synergy also make a difference. Effective support from teammates allows players to succeed in their roles. The high Overall Impact lift values in our study suggest that players contributing across multiple aspects of the game have a great impact on outcomes. This resonates Dehesa et al. [51] emphasis on team dynamics influencing performance during critical game moments.

In the playoffs, when defenses focus on stopping star players, role players who step up become invaluable. Teams like the Warriors, with a cohesive unit understanding their roles, can maintain high offensive ratings because of collective effort. The data shows that high net ratings and Player Impact Estimates (PIE) are associated with winning outcomes, supporting the value of overall team effectiveness.

6.1.4 Interpreting Results on Real-Life Scenarios

The increased lift values for defense in the playoffs reflect real-life strategies where teams rely on shutting down opponents through detailed defensive schemes. For example, the 2004 Detroit Pistons, anchored by Ben Wallace's defensive dominance, won the championship by holding opponents to low scoring. This aligns with our data showing that low defensive ratings drive playoff success.

Performances of Tim Duncan and Kevin Garnett exemplify how individual defensive excellence can anchor a team's success. Their consistent defensive impact in both regular-season and playoff games reveals the value of strong defensive players, supporting prior studies on the importance of defensive ratings.

On offense, players like Nikola Jokic, Stephen Curry, Kevin Durant, and Damian Lillard thrive by maintaining or improving efficiency under playoff pressure.

Stephen Curry's high tempo play and scoring in the playoffs reflect the Warriors' strategy of overwhelming opponents with pace and shooting. His ability to sustain this style in the postseason shows the effectiveness of their system, as discussed in "Thinking Basketball" [54].

Damian Lillard's clutch performances, often referred to as "Dame Time”, mirror scenarios where his ability to score and make decisive plays in the fourth quarter leads his team to victory. LeBron James's versatility and clutch performances display his ability to dominate multiple facets of the game, leading his teams to success in high-pressure situations. Both players' contributions in critical moments highlight their value in closing out games, aligning with findings of Sarlis et. al on clutch ability [41].

6.2 Data Science Perspective

This research highlights the critical role of data science in solving real-world challenges. The large-scale NBA dataset, sourced from the NBA API, underwent extensive preprocessing to handle missing values, outliers, and inconsistencies. These steps represent the approach required for high-quality analysis.

By employing tools such as FP-Growth, we uncovered meaningful patterns in quarter-by-quarter performance metrics. The feature engineering process, including discretization and computation of composite scores, demonstrates how raw data can be transformed into actionable insights, reflecting core best practices in data science.

An integrated anomaly detection and discretization framework ensured data reliability, showcasing how multiple methodologies can enhance the robustness of analysis. While our focus was sports, these same techniques and processes can be used wherever large datasets need to be turned into practical solutions and are applicable across industries.

6.3 Limitations and Threats to Validity

Our work has certain limitations and potential threats to validity, including issues related to quarter-specific analysis, sample size imbalances, player participation thresholds, feature selection and categorization, discretization choices, computational limits of FP-Growth, and the calculation of composite scores. Each subsection details how these limitations may affect the reliability and generalizability of the findings.

6.3.1 Influence of Quarter on Composite Scores

Our analysis heavily relies on per-quarter statistics to compute composite scores for players, but specific game situations may not be fully captured. Coaching adjustments, substitutions, or uneven playing time can cause variations in per-quarter performance data.

6.3.2 Sample Size

The playoff dataset is considerably smaller compared to the regular season dataset. This disparity can affect the robustness of the association rules generated, especially in categories like Ball Handling, where only a limited number of rules were found during the regular season. The small sample size may lead to less reliable results.

6.3.3 Player Participation

Players below specific thresholds in games and minutes were excluded from the analysis, potentially excluding impactful performances from players with fewer games or minutes, such as emerging talents or those returning from injuries.

The filtering process may also lead to situations where a player meets the games played threshold but contributes meaningfully to only a small subset of those games. For example, a player who participated in 22 games but met the minimum playing time threshold in only 5 games would only have those 5 games included in the analysis.

Players who only play a few games or quarters can end up with inflated stats that make them look better than they really are. For example, the stats of players who put up great numbers in a small sample appear better than players who perform consistently over time.

6.3.4 Feature Selection and Categorization

The categorization of performance metrics into Offensive, Defensive, Ball Handling, Overall Impact, and Tempo was based on subjective judgments of how individual metrics might correlate. An alternative grouping might produce different conclusions and reflect the complexity of player performance more accurately.

6.3.5 Discretization Methods

Discretization shapes both the calculation of composite scores and the ARM process. The choice of discretization methods impacts the results and the final selection, which was based on manual reviews, may introduce subjectivity. Since the binned datasets were directly used to calculate composite scores, the accuracy of these scores depends heavily on the quality and distribution of the bins. Different discretization strategies could yield alternative results, affecting both the association rules and the composite scores.

6.3.6 FP-Growth Limitations

Due to RAM limitations, FP-Growth could not run on all features simultaneously, so we grouped them into five categories. This splitting may have hidden association rules that connect multiple categories. Different parameter settings for the FP-Growth, such as support and confidence thresholds, would likely uncover different patterns.

Alternative algorithms like ARMICA-Improved, may offer solutions to these limitations. ARMICA-Improved automates parameter selection and scans the database only once, enabling simultaneous processing of all features while reducing dependency on user-defined thresholds [74]. Swarm-based approaches such as Ant-ARM and BSO-ARM leverage parallelization to handle larger datasets more efficiently, preserving full feature relationships without requiring segmentation. These methods are not only faster but also focus on generating fewer, higher-quality rules by emphasizing interestingness and relevance, minimizing noise introduced by over-segmentation. GPU-based algorithms like PMES and HSBO-TS excel in scalability, addressing memory limitations and enabling the exploration of more complex rule sets. [74], [75].

6.3.7 Score Calculation

Composite scores relied on weighted averages within each category to reflect importance. Although the weights were carefully chosen, they may not fully capture the real impact of every metric, creating a degree of subjectivity in the final scores, introducing potential subjectivity into the scoring process.

7 Conclusions

This chapter wraps up the analysis of NBA player performance metrics drawn from two decades of data. It also covers practical takeaways for teams and analysts, along with ideas for future studies.

7.1 Conclusion

This thesis addressed a gap in NBA analytics by analyzing quarter-by-quarter player performance metrics from the 2004-05 to 2023-24 seasons. By using FP-Growth Association Rule Mining (ARM), it revealed how advanced performance metrics influence game outcomes in five categories: Offense, Defense, Ball Handling, Overall Impact, and Tempo.

The findings reveal several insights:

- **Effectiveness of FP-Growth:** FP-Growth efficiently extracted association rules from high-dimensional data while handling large datasets. By setting threshold values for support and confidence, it produced clear and meaningful rules for game analysis.
- **Discretization Challenges**: The choice of discretization method impacted our results. Equal frequency binning and K-means clustering had a balance between interpretability and preserving meaningful data distributions, though some subjective decisions were inevitable.
- **Anomaly Detection Contribution:** Our combining of anomaly detection techniques (Z-score, IQR, Isolation Forest, and DBSCAN) improved data cleaning and prevented outliers from skewing results.
- **Defensive Performance in Crucial Moments**: Defensive metrics during the fourth quarter and overtime periods show a stronger association with winning. This was even more notable in the playoffs where strong defensive ratings often led to success.
- **Offensive Efficiency in the Playoffs**: Offensive metrics, like effective field goal percentage and offensive rating become more influential in the playoffs. Players who scored efficiently under tough defensive pressure boosted their teams' chances of winning.
- **Ball Handling and Turnovers**: Association rules indicate that a high assist ratio

combined with a low turnover ratio correlate with victory, reflecting strong possession control.

- **Overall Impact is Vital**: Metrics that capture a player's comprehensive contribution, like Player Impact Estimate (PIE) and net rating, had the highest lift values. Players with strong all-around stats often have the biggest influence on the game.
- **Positional Insights**:
  - **Guards**: Offensive efficiency and ball control are crucial, with high offensive ratings and assist ratios being key to winning.
  - **Centers**: Both offensive contributions and defensive consistency are important. High rebound percentages and net ratings are strongly associated with success.
  - **Forwards**: Offensive efficiency and ball security are critical, especially in the playoffs.
- **Quarter-by-Quarter Variations**: Defensive and offensive mattered most in the fourth quarter and overtime, aligning with the increased pressure of those moments.

Implications

The study provides valuable insights for coaches, analysts, and teams:

- **Strategic Focus on Defense**: Focusing on defense late in games can enhance winning probabilities, especially in the playoffs.
- **Player Development**: Training programs can be tailored to improve players' performance in metrics identified as influential, such as offensive efficiency and ball handling.
- **Game Planning**: Knowing which metrics matter most in each quarter can guide real-time adjustments and strategic planning.

While this study provides valuable insights, certain limitations discussed earlier, such as the smaller playoff sample size, subjective discretization methods, and computational constraints, may influence the findings.

7.2 Future Work

There are many ways to build on this research. One area for further exploration is the **integration of additional metrics**, such as shot location and defensive matchups. These might reveal new details about what affects performance. Employing **advanced analytical techniques**, such as ML models like deep learning or ensemble methods, could better capture the nonlinear relationships and interactions in the data. Studies such as that by Papageorgiou et al. [49] showed the strong predictive power of Random Forest models for player performance, while J. K. Nakul et al. [48] found Artificial Neural Networks (ANN) to be very effective at forecasting outcomes based on advanced stats, highlighting the potential of advanced models to uncover complex patterns in sports analytics.

To improve ARM's ability to analyze every feature at once, we need to address computational limitations. More advanced methods, like ARMICA-Improved or swarm-based approaches, could help find connections across different categories and reveal deeper patterns. A **dynamic weighting** system for composite scores would also be useful. By adjusting weights based on factors like game stakes, opponent strength, or specific situations, we could make performance metrics more accurate.

Considering **external factors** like home-court advantage, back-to-back games, or travel fatigue could add valuable insights into how these elements impact performance and outcomes. Including **qualitative data**, like player interviews or coaching reports, would bring in a human perspective to better interpret performance metrics.

It would also help to expand the analysis beyond the NBA. Comparing international basketball leagues could highlight which performance metrics are universal and which ones are unique to the NBA.

The development of **real-time analytics applications** could help coaches make quick strategic adjustments during games based on live player performance metrics.

Lastly, it is important to analyze the **impact of rule changes** within the NBA over the studied period. Understanding how these changes have affected player performance metrics can provide insights on how the sport keeps evolving and inform future rule-making decisions.

References

[1] B. Thuraisingham, “Primer for understanding and applying data mining,” *IT Prof*, vol. 2, no. 1, pp. 28–31, Jan. 2000, doi: 10.1109/6294.819936.

[2] J. Jackson, “Data Mining; A Conceptual Overview,” *Communications of the Association for Information Systems*, vol. 8, no. 1, p. 19, Mar. 2002, doi: 10.17705/1CAIS.00819.

[3] D. T. Larose, “Discovering Knowledge in Data: An Introduction to Data Mining,” *Discovering Knowledge in Data: An Introduction to Data Mining*, pp. 1–222, Jan. 2005, doi: 10.1002/0471687545.

[4] L. Rokach and O. Maimon, “Clustering Methods,” *Data Mining and Knowledge Discovery Handbook*, pp. 321–352, May 2005, doi: 10.1007/0-387-25465-X_15.

[5] V. Chandola, A. Banerjee, and V. Kumar, “Anomaly Detection: A Survey,” *ACM Comput. Surv.*, vol. 41, Jul. 2009, doi: 10.1145/1541880.1541882.

[6] S. B. Kotsiantis and D. N. Kanellopoulos, “Association Rules Mining: A Recent Overview,” 2006. [Online]. Available: https://api.semanticscholar.org/CorpusID:671244

[7] M. L. Refined, J. Watt, R. Borhani, and A. K. Katsaggelos, “Machine learning foundations, Algorithms, and Applications,” pp. 1–593, 2020.

[8] J. Alzubi, A. Nayyar, and A. Kumar, “Machine Learning from Theory to Algorithms: An Overview,” *J Phys Conf Ser*, vol. 1142, no. 1, p. 012012, Nov. 2018, doi: 10.1088/1742-6596/1142/1/012012.

[9] B. Mahesh, “Machine Learning Algorithms-A Review,” *International Journal of Science and Research*, 2018, doi: 10.21275/ART20203995.

[10] X. Wang, “Machine Learning Basics,” 2019.

[11] M. Çelik, F. Dadaşer-Çelik, and A. Ş. Dokuz, “Anomaly detection in temperature data using DBSCAN algorithm,” in *2011 International Symposium on Innovations in Intelligent Systems and Applications*, 2011, pp. 91–95. doi: 10.1109/INISTA.2011.5946052.

[12] H. Liu, F. Hussain, C. L. Tan, and M. Dash, “Discretization: An Enabling Technique,” *Data Min Knowl Discov*, vol. 6, no. 4, pp. 393–423, 2002, doi: 10.1023/A:1016304305535.

[13] “What is Machine Learning (ML)? Types, Models, Examples | Enterprise Tech News

EM360.” Accessed: Nov. 02, 2024. [Online]. Available: https://em360tech.com/tech-article/what-is-machine-learning-ml

[14] W. Tichy, “Changing the Game: ‘Dr. Dave’ Schrader on Sports Analytics,” 2016, Accessed: Nov. 02, 2024. [Online]. Available: http://ubiquity.acm.org1

[15] “Moneyball: The Power of Sports Analytics 15.071-The Analytics Edge”.

[16] M. Mondello and C. Kamke, “The Introduction and Application of Sports Analytics in Professional Sport Organizations,” *Journal of Applied Sport Management*, vol. 6, no. 2.

[17] “Analytics in Sports: The New Science of Winning - February 2014 Thomas H. Davenport.” Accessed: Nov. 02, 2024. [Online]. Available: https://www.readkong.com/page/analytics-in-sports-the-new-science-of-winning-6739419

[18] “Data Analytics in Sports: Enhancing Performance and Fan Engagement | LinkedIn.” Accessed: Nov. 02, 2024. [Online]. Available: https://www.linkedin.com/pulse/data-analytics-sports-enhancing-performance-fan-durgesh-kekare-rvtlf/

[19] “Stat Glossary | Stats | NBA.com.” Accessed: Nov. 03, 2024. [Online]. Available: https://www.nba.com/stats/help/glossary

[20] “Glossary | Basketball-Reference.com.” Accessed: Nov. 03, 2024. [Online]. Available: https://www.basketball-reference.com/about/glossary.html

[21] “History of baseball | Events, Dates, People, & Facts | Britannica.” Accessed: Nov. 02, 2024. [Online]. Available: https://www.britannica.com/sports/history-of-baseball

[22] J. Mizels, B. Erickson, and P. Chalmers, “Current State of Data and Analytics Research in Baseball,” *Curr Rev Musculoskelet Med*, vol. 15, no. 4, p. 283, Aug. 2022, doi: 10.1007/S12178-022-09763-6.

[23] B. Burroughs, “Statistics and Baseball Fandom: Sabermetric Infrastructure of Expertise,” *Games Cult*, vol. 15, no. 3, pp. 248–265, May 2020, doi: 10.1177/1555412018783319.

[24] M. S. Triady and A. F. Utami, “Analysis of Decision Making Process in Moneyball: The Art of Winning an Unfair Game,” *The Winners*, vol. 16, no. 1, p. 57, Mar. 2015, doi: 10.21512/TW.V16I1.1555.

[25] J. R. Hoffman, J. Vazquez, N. Pichardo, and G. Tenenbaum, “Anthropometric and performance comparisons in professional baseball players,” *J Strength Cond Res*, vol. 23, no. 8, pp. 2173–2178, Nov. 2009, doi: 10.1519/JSC.0B013E3181BCD5FE.

[26] Z. Papadakis, R. N. Padgett, A. Stamatis, and R. A. Karasch, “Baseball performance via the lens of anthropometric testing, fitness metrics, and statistics: a longitudinal cross-sectional study,” *Curr Orthop Pract*, vol. 32, no. 2, pp. 151–160, Mar. 2021, doi:

10.1097/BCO.0000000000000962.

[27] S. Mecheri, F. Rioult, B. Mantel, F. Kauffmann, and N. Benguigui, “The serve impact in tennis: First large-scale study of big Hawk-Eye data,” *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 9, no. 5, pp. 310–325, Oct. 2016, doi: 10.1002/SAM.11316.

[28] X. Wei, P. Lucey, S. Morgan, and S. Sridharan, “Forecasting the Next Shot Location in Tennis Using Fine-Grained Spatiotemporal Tracking Data,” *IEEE Trans Knowl Data Eng*, vol. 28, no. 11, pp. 2988–2997, 2016, doi: 10.1109/TKDE.2016.2594787.

[29] Z. Gao and A. Kowalczyk, “Random forest model identifies serve strength as a key predictor of tennis match outcome,” Oct. 2019, Accessed: Nov. 02, 2024. [Online]. Available: https://arxiv.org/abs/1910.03203v1

[30] T. Polk, D. Jackle, J. Haussler, and J. Yang, “Courttime: Generating actionable insights into tennis matches using visual analytics,” *IEEE Trans Vis Comput Graph*, vol. 26, no. 1, pp. 397–406, Jan. 2020, doi: 10.1109/TVCG.2019.2934243.

[31] R. Rein and D. Memmert, “Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science,” *SpringerPlus 2016 5:1*, vol. 5, no. 1, pp. 1–13, Aug. 2016, doi: 10.1186/S40064-016-3108-2.

[32] L. Bornn, D. Cervone, and J. Fernandez, “Soccer analytics: Unravelling the complexity of ‘the beautiful game,’” *Significance*, vol. 15, no. 3, pp. 26–29, Jun. 2018, doi: 10.1111/J.1740-9713.2018.01146.X.

[33] Markopoulou and Christina, “Sport Analytics Algorithms for Football Performance Prediction,” Jun. 2024, Accessed: Nov. 03, 2024. [Online]. Available: https://repository.ihu.edu.gr//xmlui/handle/11544/30418

[34] K. Apostolou and C. Tjortjis, “Sports Analytics algorithms for performance prediction,” *10th International Conference on Information, Intelligence, Systems and Applications, IISA 2019*, Jul. 2019, doi: 10.1109/IISA.2019.8900754.

[35] M. C. Malamatinos, E. Vrochidou, and G. A. Papakostas, “On Predicting Soccer Outcomes in the Greek League Using Machine Learning,” *Computers 2022, Vol. 11, Page 133*, vol. 11, no. 9, p. 133, Aug. 2022, doi: 10.3390/COMPUTERS11090133.

[36] V. C. Pantzalis and C. Tjortjis, “Sports Analytics for Football League Table and Player Performance Prediction,” *11th International Conference on Information, Intelligence, Systems and Applications, IISA 2020*, Jul. 2020, doi: 10.1109/IISA50023.2020.9284352.

[37] Z. Pu *et al.*, “Orientation and Decision-Making for Soccer Based on Sports Analytics and AI: A Systematic Review,” *IEEE/CAA Journal of Automatica Sinica, 2024, Vol.*

11, Issue 1, Pages: 37-57, vol. 11, no. 1, pp. 37–57, Jan. 2024, doi: 10.1109/JAS.2023.123807.

[38] J. Kubatko, D. Oliver, K. Pelton, and D. T. Rosenbaum, "A Starting Point for Analyzing Basketball Statistics," *J Quant Anal Sports*, vol. 3, no. 3, Jul. 2007, doi: 10.2202/1559-0410.1070.

[39] "How the 'Idiots Who Believe' in the Analytics Movement Have Forever Changed Basketball | Opta Analyst." Accessed: Nov. 03, 2024. [Online]. Available: https://theanalyst.com/eu/2021/04/how-advanced-analytics-have-changed-basketball

[40] "SportVU - Wikipedia." Accessed: Nov. 03, 2024. [Online]. Available: https://en.wikipedia.org/wiki/SportVU

[41] V. Sarlis, V. Chatziilias, C. Tjortjis, and D. Mandalidis, "A Data Science approach analysing the Impact of Injuries on Basketball Player and Team Performance," *Inf Syst*, vol. 99, p. 101750, 2021, doi: 10.1016/j.is.2021.101750.

[42] V. Sarlis, G. Papageorgiou, and C. Tjortjis, "Sports Analytics and Text Mining NBA Data to Assess Recovery from Injuries and Their Economic Impact," *Computers*, vol. 12, no. 12, Dec. 2023, doi: 10.3390/COMPUTERS12120261.

[43] G. Papageorgiou, V. Sarlis, and C. Tjortjis, "Unsupervised Learning in NBA Injury Recovery: Advanced Data Mining to Decode Recovery Durations and Economic Impacts," *Information 2024, Vol. 15, Page 61*, vol. 15, no. 1, p. 61, Jan. 2024, doi: 10.3390/INFO15010061.

[44] V. Sarlis, G. Papageorgiou, and C. Tjortjis, "Leveraging Sports Analytics and Association Rule Mining to Uncover Recovery and Economic Impacts in NBA Basketball," *Data 2024, Vol. 9, Page 83*, vol. 9, no. 7, p. 83, Jun. 2024, doi: 10.3390/DATA9070083.

[45] V. Sarlis and C. Tjortjis, "Sports Analytics: Data Mining to Uncover NBA Player Position, Age, and Injury Impact on Performance and Economics," *Information 2024, Vol. 15, Page 242*, vol. 15, no. 4, p. 242, Apr. 2024, doi: 10.3390/INFO15040242.

[46] "Stanley Yang Thesis - Predicting Regular Season Results of NBA Teams Based on Regression Analysis of Common Basketball Statistics by Yuanhao Stanley | Course Hero." Accessed: Nov. 07, 2024. [Online]. Available: https://www.coursehero.com/file/22705629/Stanley-Yang-Thesis/

[47] W. J. Chen, M. J. Jhou, T. S. Lee, and C. J. Lu, "Hybrid Basketball Game Outcome Prediction Model by Integrating Data Mining Methods for the National Basketball Association," *Entropy 2021, Vol. 23, Page 477*, vol. 23, no. 4, p. 477, Apr. 2021, doi: 10.3390/E23040477.

[48] J. K. Nakul *et al.*, “Integrating Advanced Statistics to Predict the Outcome of National Basketball Association Games with Machine Learning,” *ECSTr*, vol. 107, no. 1, pp. 15263–15272, Apr. 2022, doi: 10.1149/10701.15263ECST.

[49] G. Papageorgiou, V. Sarlis, and C. Tjortjis, “Evaluating the effectiveness of machine learning models for performance forecasting in basketball: a comparative study,” *Knowl Inf Syst*, vol. 66, no. 7, pp. 4333–4375, Jul. 2024, doi: 10.1007/S10115-024-02092-9/FIGURES/1.

[50] M. Á. Gómez Ruano, L. Gasperi, and C. Lupo, “Performance analysis of game dynamics during the 4th game quarter of NBA close games,” *Int J Perform Anal Sport*, vol. 16, no. 1, pp. 249–263, Jan. 2016, doi: 10.1080/24748668.2016.11868884.

[51] R. Dehesa, A. Vaquera, M. A. Gomez-Ruano, B. Gonçalves, N. Mateus, and J. Sampaio, “KEY PERFORMANCE INDICATORS IN NBA PLAYERS’ PERFORMANCE PROFILES,” *Kinesiology*, vol. 51, no. 1, pp. 92–101, Mar. 2019, doi: 10.26582/k.51.1.9.

[52] V. Sarlis, D. Gerakas, and C. Tjortjis, “A Data Science and Sports Analytics Approach to Decode Clutch Dynamics in the Last Minutes of NBA Games,” *Machine Learning and Knowledge Extraction 2024, Vol. 6, Pages 2074-2095*, vol. 6, no. 3, pp. 2074–2095, Sep. 2024, doi: 10.3390/MAKE6030102.

[53] Sarlis and Vangelis, “Data Science for Sports Analytics,” May 2024, Accessed: Nov. 03, 2024. [Online]. Available: https://repository.ihu.edu.gr//xmlui/handle/11544/30399

[54] “Thinking Basketball Summary PDF | Ben Taylor.” Accessed: Nov. 03, 2024. [Online]. Available: https://www.bookey.app/book/thinking-basketball

[55] L. Sha *et al.*, “Interactive Sports Analytics,” *ACM Transactions on Computer-Human Interaction (TOCHI)*, vol. 25, no. 2, Apr. 2018, doi: 10.1145/3185596.

[56] I. Ghosh, S. Ramasamy Ramamurthy, A. Chakma, and N. Roy, “Sports analytics review: Artificial intelligence applications, emerging technologies, and algorithmic perspective,” *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 13, no. 5, p. e1496, Sep. 2023, doi: 10.1002/WIDM.1496.

[57] “GitHub - swar/nba_api: An API Client package to access the APIs for NBA.com.” Accessed: Oct. 13, 2024. [Online]. Available: https://github.com/swar/nba_api?tab=readme-ov-file

[58] “colab.google.” Accessed: Oct. 13, 2024. [Online]. Available: https://colab.google/

[59] “mlxtend.” Accessed: Oct. 13, 2024. [Online]. Available: https://rasbt.github.io/mlxtend/

[60] “nba_api/docs/nba_api/stats/endpoints/commonallplayers.md at master · swar/nba_api

· GitHub.” Accessed: Oct. 13, 2024. [Online]. Available: https://github.com/swar/nba_api/blob/master/docs/nba_api/stats/endpoints/commonallplayers.md

[61] “nba_api/docs/nba_api/stats/endpoints/leaguegamelog.md at master · swar/nba_api · GitHub.” Accessed: Oct. 13, 2024. [Online]. Available: https://github.com/swar/nba_api/blob/master/docs/nba_api/stats/endpoints/leaguegamelog.md

[62] “nba_api/docs/nba_api/stats/endpoints/boxscoreadvancedv3.md at master · swar/nba_api · GitHub.” Accessed: Oct. 13, 2024. [Online]. Available: https://github.com/swar/nba_api/blob/master/docs/nba_api/stats/endpoints/boxscoreadvancedv3.md

[63] “nba_api/docs/nba_api/stats/endpoints/playergamelog.md at master · swar/nba_api · GitHub.” Accessed: Oct. 13, 2024. [Online]. Available: https://github.com/swar/nba_api/blob/master/docs/nba_api/stats/endpoints/playergamelog.md

[64] “nba_api/docs/nba_api/stats/endpoints/commonplayerinfo.md at master · swar/nba_api · GitHub.” Accessed: Oct. 13, 2024. [Online]. Available: https://github.com/swar/nba_api/blob/master/docs/nba_api/stats/endpoints/commonplayerinfo.md

[65] “StandardScaler — scikit-learn 1.6.dev0 documentation.” Accessed: Oct. 13, 2024. [Online]. Available: https://scikit-learn.org/dev/modules/generated/sklearn.preprocessing.StandardScaler.html

[66] “When and Why to Standardize Your Data | Built In.” Accessed: Oct. 13, 2024. [Online]. Available: https://builtin.com/data-science/when-and-why-standardize-your-data

[67] W. S. Al Farizi, I. Hidayah, and M. N. Rizal, “Isolation Forest Based Anomaly Detection: A Systematic Literature Review,” in *2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, 2021, pp. 118–122. doi: 10.1109/ICITACEE53184.2021.9617498.

[68] “pandas.cut — pandas 2.2.3 documentation.” Accessed: Oct. 13, 2024. [Online]. Available: https://pandas.pydata.org/docs/reference/api/pandas.cut.html

[69] “pandas.qcut — pandas 2.2.3 documentation.” Accessed: Oct. 13, 2024. [Online]. Available: https://pandas.pydata.org/docs/reference/api/pandas.qcut.html

[70] “KMeans — scikit-learn 1.5.2 documentation.” Accessed: Oct. 13, 2024. [Online]. Available: https://scikit-learn.org/1.5/modules/generated/sklearn.cluster.KMeans.html

[71] S. Kotsiantis and D. Kanellopoulos, “Discretization techniques: A recent survey,” *GESTS International Transactions on Computer Science and Engineering*, vol. 32, no. 1, pp. 47–58, 2006.

[72] “TransactionEncoder - mlxtend.” Accessed: Oct. 13, 2024. [Online]. Available: https://rasbt.github.io/mlxtend/user_guide/preprocessing/TransactionEncoder/

[73] “Fpgrowth - mlxtend.” Accessed: Oct. 13, 2024. [Online]. Available: https://rasbt.github.io/mlxtend/user_guide/frequent_patterns/fpgrowth/#overview

[74] S. Yakhchi, S. M. Ghafari, C. Tjortjis, and M. Fazeli, “ARMICA-Improved: A New Approach for Association Rule Mining,” 2017, doi: 10.1007/978-3-319-63558-3_25.

[75] S. M. Ghafari and C. Tjortjis, “A survey on association rules mining using heuristics,” *Wiley Interdiscip Rev Data Min Knowl Discov*, vol. 9, no. 4, p. e1307, Jul. 2019, doi: 10.1002/WIDM.1307.

Appendices

Appendix A: Variables Definitions

This appendix provides definitions for all variables and metrics used in the analysis. Table 5 lists these variables, which are used to calculate the basketball performance metrics for this study.

Table 5: Variables and Definitions

Variables	Defenition
PTS	Points scored
POSS	Total number of possessions
OppPTS	Points scored by the opposing team
OppPOSS	Total number of possessions by the opposing team
FGA	Field Goal Attempts
FGM	Field Goals Made
3PM	Three-Point Field Goals Made
FTA	Free Throw Attempts
TmFGA	Team Field Goal Attempts
TmFTA	Team Free Throw Attempts
TmFG	Team Field Goals Made
TmTOV	Team Turnovers
OppFGA	Opponent Field Goal Attempts
OppFTA	Opponent Free Throw Attempts
OppFG	Opponent Field Goals Made
OppTOV	Opponent Turnovers
TmORB	Team Offensive Rebounds
OppDRB	Opponent Defensive Rebounds
TmDRB	Team Defensive Rebounds
OppORB	Opponent Offensive Rebounds
TmMP	Team Minutes Played
MP	Player Minutes Played
OREB	Offensive Rebounds
DREB	Defensive Rebounds
TRB	Total Rebounds
TmTRB	Team Total Rebounds
OppTRB	Opponent Total Rebounds
AST	Assists by the player
TO	Turnovers by the player

GmPTS	Total Game Points Scored
GmFGM	Total Game Field Goals Made
GmFTM	Total Game Free Throws Made
GmFGA	Total Game Field Goal Attempts
GmFTA	Total Game Free Throw Attempts
GmDREB	Total Game Defensive Rebounds
GmOREB	Total Game Offensive Rebounds
GmAST	Total Game Assists
GmSTL	Total Game Steals
GmBLK	Total Game Blocks
GmPF	Total Game Personal Fouls
GmTO	Total Game Turnovers

Appendix B: Endpoints and Parameters

This appendix outlines the NBA API endpoints and parameters used to collect and organize player and game data. Table 6 describes endpoints and parameters, which were used for retrieving and structuring the dataset by game quarters.

Table 6: NBA API Endpoints & Key Parameters

Endpoints and Parameters	Explanation
commonallplayers	NBA API endpoint used to retrieve all active players for a specific season.
leaguegamelog	NBA API endpoint to retrieve game logs for a given season, including the playoffs.
boxscoreadvancedv3	NBA API endpoint that provides advanced stats for individual games, including per-quarter and overtime performance.
playergamelog	NBA API endpoint that returns game logs for individual players in a season.
commonplayerinfo	NBA API endpoint used to retrieve detailed player information, such as their position.
PLAYER_ID	A unique identifier for each player in the NBA.
PERSON_ID	Another identifier for players, used similarly to PLAYER_ID.
ROSTERSTATUS	A flag indicating whether a player was active in the given season (1 = active).
gameId	A unique identifier for each game.
teamTricode	The three-letter abbreviation represents each NBA team.
Result	Indicates whether the player's team won ('W') or lost ('L') the game.
PERIOD	Indicates the specific quarter (1st, 2nd, 3rd, 4th) or overtime period (5th-8th) for which the stats are recorded.
minutes	Total minutes a player spent on the court.
start_period	Defines the starting period (quarter or overtime) for which stats are retrieved.
end_period	Defines the ending period, ensuring the data retrieval covers only the specified period.
start_range	Indicates the start of the time range within the specified period.
end_range	Indicates the end of the time range within the specified period.
range_type	Specifies how the statistics are segmented.

Appendix C: Association Rules (Regular Season and Playoffs)

This appendix summarizes the association rules generated from the FP-Growth for the regular season and playoffs. Tables 7 and 8 present these rules, highlighting key performance metrics and their associations with game outcomes (wins/losses) for different player categories and positions.

Table 7: Association Rules for All Categories (Regular Season)

	REGULAR				
Category/Position	**Antecedents**	**Consequents**	**Confidence**	**Support**	**Lift ↓**
Ball Handling/All					
Rule 1	low_assistPercentage_Period3, medium_assistRatio_Period3	W_Result	0.0057	0.6573	1.3146
Rule 2	low_assistPercentage_Period1, medium_assistRatio_Period1	W_Result	0.0077	0.6250	1.2499
Rule 3	low_assistPercentage_Period1, medium_assistRatio_Period1, low_assistToTurnover_Period2	W_Result	0.0058	0.6194	1.2387
Rule 4	low_assistPercentage_Period1, medium_assistRatio_Period1, low_assistToTurnover_Period3	W_Result	0.0056	0.6149	1.2298
Rule 5	high_assistRatio_Period1, low_assistToTurnover_Period1, low_assistToTurnover_Period3, low_turnoverRatio_Period3, medium_assistPercentage_Period3	W_Result	0.0071	0.6000	1.2000
Ball Handling/Guards					
	No rules found for this category and position.				
Ball Handling/Centers					
	No rules found for this category and position.				
Ball					

Handling/Forwards					
	No rules found for this category and position.				
Defence/All					
Rule 1	medium_defensiveReboundPercentage_Period1, high_defensiveRating_Period1, high_defensiveRating_Period2, high_defensiveRating_Period3	L_Result	0.0050	0.8277	1.6555
Rule 2	low_defensiveRating_Period1, medium_defensiveReboundPercentage_Period1, low_defensiveRating_Period3, low_defensiveRating_Period4	W_Result	0.0056	0.8213	1.6426
Rule 3	low_defensiveRating_Period1, medium_defensiveReboundPercentage_Period1, low_defensiveRating_Period2, low_defensiveRating_Period3	W_Result	0.0069	0.8164	1.6327
Rule 4	low_defensiveReboundPercentage_Period5+, high_defensiveRating_Period5+	L_Result	0.0069	0.8156	1.6313
Rule 5	high_defensiveRating_Period5+	L_Result	0.0093	0.8126	1.6253
Defence/Guards					
Rule 1	G_position, high_defensiveRating_Period1, high_defensiveRating_Period2, high_defensiveRating_Period3	L_Result	0.0073	0.8001	1.6002
Rule 2	G_position, low_defensiveRating_Period1, low_defensiveRating_Period3, low_defensiveRating_Period4	W_Result	0.0066	0.7915	1.5829
Defence/Centers					
Rule 1	C_position, high_defensiveRating_Period1, high_defensiveRating_Period2	L_Result	0.0075	0.7243	1.4487
Rule 2	C_position, high_defensiveRating_Period1, high_defensiveRating_Period3	L_Result	0.0080	0.7237	1.4475
Defence/Forwards					
Rule 1	F_position, high_defensiveRating_Period1,	L_Result	0.0072	0.8091	1.6182

	high_defensiveRating_Period2, high_defensiveRating_Period3				
Rule 2	F_position, high_defensiveRating_Period1, high_defensiveRating_Period3, high_defensiveRating_Period4	L_Result	0.0058	0.7988	1.5977
Offense/All					
Rule 1	high_offensiveRating_Period1, high_offensiveRating_Period2, high_offensiveRating_Period3, high_offensiveRating_Period4	W_Result	0.0051	0.8539	1.7077
Rule 2	high_offensiveRating_Period1, high_offensiveRating_Period3, high_effectiveFieldGoalPercentage_Period4, high_offensiveRating_Period4	W_Result	0.0052	0.8259	1.6517
Rule 3	high_offensiveRating_Period1, high_offensiveRating_Period3, high_trueShootingPercentage_Period4, high_effectiveFieldGoalPercentage_Period4, high_offensiveRating_Period4	W_Result	0.0052	0.8256	1.6512
Rule 4	low_offensiveRating_Period1, low_offensiveRating_Period2, low_offensiveRating_Period3, medium_effectiveFieldGoalPercentage_Period3	L_Result	0.0055	0.8254	1.6509
Rule 5	high_offensiveRating_Period1, high_offensiveRating_Period2, low_offensiveReboundPercentage_Period3, high_effectiveFieldGoalPercentage_Period3, high_offensiveRating_Period3	W_Result	0.0052	0.8219	1.6437
Offense/Guards					
Rule 1	G_position, high_offensiveRating_Period1, high_offensiveRating_Period2, low_offensiveReboundPercentage_Period2, low_offensiveReboundPercentage_Period3, high_offensiveRating_Period3	W_Result	0.0063	0.8048	1.6096

Rule 2	G_position, high_offensiveRating_Period1, high_offensiveRating_Period2, low_offensiveReboundPercentage_Period2, high_offensiveRating_Period3	W_Result	0.0073	0.8031	1.6061
Offense/Centers					
Rule 1	C_position, low_offensiveRating_Period1, low_offensiveRating_Period3	L_Result	0.0090	0.7156	1.4312
Rule 2	C_position, high_offensiveRating_Period1, high_offensiveRating_Period3	W_Result	0.0082	0.7098	1.4195
Offense/Forwards					
Rule 1	F_position, high_offensiveRating_Period1, high_offensiveRating_Period3, high_offensiveRating_Period4, low_offensiveReboundPercentage_Period4	W_Result	0.0050	0.8027	1.6053
Rule 2	F_position, high_offensiveRating_Period1, high_offensiveRating_Period2, low_offensiveReboundPercentage_Period2, high_offensiveRating_Period3	W_Result	0.0061	0.8007	1.6013
Overall Impact/All					
Rule 1	high_netRating_Period5+, high_PIE_Period5+	W_Result	0.0064	0.9749	1.9497
Rule 2	medium_reboundPercentage_Period4, high_netRating_Period5+	W_Result	0.0052	0.9712	1.9423
Rule 3	medium_reboundPercentage_Period4, low_netRating_Period5+	L_Result	0.0055	0.9706	1.9412
Rule 4	G_position, low_netRating_Period5+	L_Result	0.0057	0.9631	1.9262
Rule 5	F_position, low_netRating_Period5+	L_Result	0.0051	0.9587	1.9175
Overall Impact/Guards					
Rule 1	G_position, low_netRating_Period5+	L_Result	0.0057	0.9631	1.9262
Rule 2	G_position, high_netRating_Period5+	W_Result	0.0059	0.9564	1.9128
Overall Impact/Centers					

Rule 1	C_position, high_netRating_Period1, high_netRating_Period3	W_Result	0.0092	0.7972	1.5943
Rule 2	C_position, low_netRating_Period1, low_netRating_Period2	L_Result	0.0089	0.7946	1.5893
Overall Impact/Forwards					
Rule 1	F_position, low_netRating_Period5+	L_Result	0.0051	0.9587	1.9175
Rule 2	F_position, high_netRating_Period1, high_netRating_Period3, high_netRating_Period4	W_Result	0.0063	0.9032	1.8064
Tempo/All					
Rule 1	G_position, high_possessions_Period1, low_possessions_Period3	L_Result	0.0053	0.6175	1.2350
Rule 2	high_possessions_Period1, low_possessions_Period2, low_possessions_Period3	L_Result	0.0057	0.6012	1.2024
Tempo/Guards					
Rule 1	G_position, high_possessions_Period1, low_possessions_Period3	L_Result	0.0053	0.6175	1.2350
Tempo/Centers					
	No rules found for this category and position.				
Tempo/Forwards					
	No rules found for this category and position.				

Table 8: Association Rules for All Categories (Playoffs)

	PLAYOFFS				
Category/Position	**Antecedents**	**Consequents**	**Confidence**	**Support**	**Lift ↓**
Ball Handling/All					
Rule 1	medium_assistPercentage_Period1, low_assistToTurnover_Period3, high_assistRatio_Period3, medium_assistPercentage_Period3	W_Result	0.0053	0.6776	1.3570
Rule 2	medium_assistPercentage_Period1, low_turnoverRatio_Period3, high_assistRatio_Period3, medium_assistPercentage_Period3	W_Result	0.0053	0.6776	1.3570
Rule 3	medium_assistPercentage_Period1, low_assistToTurnover_Period3, high_assistRatio_Period3, medium_assistPercentage_Period3, low_turnoverRatio_Period3	W_Result	0.0053	0.6776	1.3570
Rule 4	medium_assistPercentage_Period1, high_assistRatio_Period3, medium_assistPercentage_Period3	W_Result	0.0067	0.6583	1.3184
Rule 5	F_position, low_turnoverRatio_Period1, high_assistRatio_Period1, low_assistToTurnover_Period1, low_assistToTurnover_Period3, high_assistRatio_Period3, low_turnoverRatio_Period3	W_Result	0.0053	0.6547	1.3112

Ball Handling/Guards					
Rule 1	G_position, medium_assistPercentage_Period1, low_assistToTurnover_Period1, low_assistToTurnover_Period2, high_assistRatio_Period3	W_Result	0.0053	0.6348	1.2713
Rule 2	G_position, medium_turnoverRatio_Period1, high_turnoverRatio_Period2	L_Result	0.0053	0.6293	1.2569
Ball Handling/Centers					
	No rules found for this category and position.				
Ball Handling/Forwards					
Rule 1	F_position, low_turnoverRatio_Period1, high_assistRatio_Period1, low_assistToTurnover_Period3, high_assistRatio_Period3	W_Result	0.0053	0.6547	1.3112
Rule 2	F_position, low_turnoverRatio_Period1, high_assistRatio_Period1, low_assistToTurnover_Period1, low_assistToTurnover_Period3, high_assistRatio_Period3, low_turnoverRatio_Period3	W_Result	0.0053	0.6547	1.3112
Defence/All					
Rule 1	F_position, high_defensiveRating_Period1, high_defensiveRating_Period3, high_defensiveRating_Period4	L_Result	0.0060	0.8770	1.7516
Rule 2	high_defensiveRating_Period1, high_defensiveRating_Period2, high_defensiveRating_Period3	L_Result	0.0156	0.8442	1.6860

Rule 3	high_defensiveRating_Period1, high_defensiveRating_Period3, high_defensiveRating_Period4	L_Result	0.0146	0.8436	1.6848
Rule 4	low_defensiveRating_Period1, high_defensiveReboundPercentage_Period1, low_defensiveRating_Period3, low_defensiveRating_Period4	W_Result	0.0057	0.8396	1.6815
Rule 5	F_position, low_defensiveRating_Period1, low_defensiveRating_Period3, low_defensiveRating_Period4	W_Result	0.0068	0.8386	1.6795
Defence/Guards					
Rule 1	G_position, high_defensiveRating_Period1, high_defensiveRating_Period2, high_defensiveRating_Period3	L_Result	0.0071	0.8405	1.6787
Rule 2	G_position, medium_defensiveReboundPercentage_Period1, low_defensiveRating_Period2, low_defensiveRating_Period4	W_Result	0.0058	0.8144	1.6311
Defence/Centers					
Rule 1	C_position, high_defensiveRating_Period1, high_defensiveRating_Period3	L_Result	0.0078	0.7852	1.5682
Rule 2	C_position, high_defensiveRating_Period1, high_defensiveRating_Period2	L_Result	0.0069	0.7692	1.5363
Defence/Forwards					
Rule 1	F_position, high_defensiveRating_Period1, high_defensiveRating_Period3, high_defensiveRating_Period4	L_Result	0.0060	0.8770	1.7516

Rule 2	F_position, low_defensiveRating_Period1, low_defensiveRating_Period3, low_defensiveRating_Period4	W_Result	0.0068	0.8386	1.6795
Offense/All					
Rule 1	medium_effectiveFieldGoalPercentage_Period1, medium_trueShootingPercentage_Period1, low_offensiveRating_Period1, low_offensiveRating_Period2, low_offensiveRating_Period3	L_Result	0.0050	0.9020	1.8014
Rule 2	medium_effectiveFieldGoalPercentage_Period1, low_offensiveRating_Period1, low_offensiveRating_Period2, low_offensiveRating_Period3	L_Result	0.0061	0.8978	1.7932
Rule 3	high_offensiveRating_Period1, low_offensiveReboundPercentage_Period2, high_effectiveFieldGoalPercentage_Period3, high_offensiveRating_Period3, high_offensiveRating_Period4, low_offensiveReboundPercentage_Period4	W_Result	0.0051	0.8854	1.7732
Rule 4	high_offensiveRating_Period2, high_trueShootingPercentage_Period2, high_offensiveRating_Period3, high_offensiveRating_Period4, high_trueShootingPercentage_Period4	W_Result	0.0051	0.8854	1.7732
Rule 5	high_offensiveRating_Period1, high_trueShootingPercentage_Period1, high_offensiveRating_Period2, high_effectiveFieldGoalPercentage_Period3, high_offensiveRating_Period3	W_Result	0.0050	0.8839	1.7702

Offense/Guards					
Rule 1	G_position, high_offensiveRating_Period1, high_trueShootingPercentage_Period3, high_effectiveFieldGoalPercentage_Period3, high_offensiveRating_Period4, low_offensiveReboundPercentage_Period4	W_Result	0.0064	0.8406	1.6835
Rule 2	G_position, low_offensiveReboundPercentage_Period1, low_offensiveRating_Period1, low_offensiveRating_Period2, low_offensiveReboundPercentage_Period3, low_offensiveRating_Period4	L_Result	0.0051	0.8424	1.6825
Offense/Centers					
Rule 1	C_position, high_offensiveRating_Period2, high_offensiveRating_Period4	W_Result	0.0064	0.7521	1.5064
Rule 2	C_position, low_offensiveRating_Period2, low_offensiveRating_Period3	L_Result	0.0080	0.7526	1.5031
Offense/Forwards					
Rule 1	F_position, high_offensiveRating_Period1, high_offensiveRating_Period2, low_offensiveReboundPercentage_Period3, high_offensiveRating_Period3	W_Result	0.0058	0.8743	1.7511
Rule 2	F_position, high_offensiveRating_Period1, high_offensiveRating_Period2, high_offensiveRating_Period3	W_Result	0.0077	0.8678	1.7380
Overall Impact/All					
Rule 1	high_PIE_Period1, high_netRating_Period1, high_netRating_Period2, high_PIE_Period2, high_netRating_Period3	W_Result	0.0053	0.9932	1.9892

Rule 2	high_reboundPercentage_Period1, high_netRating_Period1, high_netRating_Period2, high_netRating_Period3	W_Result	0.0071	0.9797	1.9621
Rule 3	high_PIE_Period1, high_netRating_Period1, high_netRating_Period2, high_PIE_Period3, high_netRating_Period3	W_Result	0.0053	0.9732	1.9490
Rule 4	high_PIE_Period1, high_netRating_Period1, high_netRating_Period2, high_netRating_Period3	W_Result	0.0104	0.9727	1.9481
Rule 5	medium_reboundPercentage_Period4, low_netRating_Period5+	L_Result	0.0050	0.9716	1.9406
Overall Impact/Guards					
Rule 1	G_position, low_netRating_Period5+	L_Result	0.0066	0.9476	1.8927
Rule 2	G_position, high_netRating_Period1, high_netRating_Period3, high_netRating_Period4	W_Result	0.0074	0.9186	1.8397
Overall Impact/Centers					
Rule 1	C_position, high_reboundPercentage_Period1, high_netRating_Period1, high_netRating_Period3	W_Result	0.0060	0.8770	1.7565
Rule 2	C_position, high_netRating_Period1, high_reboundPercentage_Period3, high_netRating_Period3	W_Result	0.0053	0.8580	1.7184
Overall Impact/Forwards					
Rule 1	F_position, high_netRating_Period1, high_netRating_Period2, high_netRating_Period3	W_Result	0.0086	0.9551	1.9129
Rule 2	F_position, low_netRating_Period5+	L_Result	0.0051	0.9329	1.8632
Tempo/All					
Rule 1	high_possessions_Period2, medium_possessions_Period3, low_possessions_Period4	L_Result	0.0055	0.6318	1.2618

Rule 2	G_position, low_possessions_Period1, high_pace_Period1, medium_possessions_Period3	L_Result	0.0050	0.6199	1.2381
Rule 3	low_possessions_Period1, low_pace_Period2, high_pace_Period4	L_Result	0.0067	0.6141	1.2265
Rule 4	F_position, medium_possessions_Period1, low_pace_Period4, low_possessions_Period4	L_Result	0.0056	0.6111	1.2205
Rule 5	high_possessions_Period2, medium_pace_Period3, low_possessions_Period4	L_Result	0.0052	0.6104	1.2191
Tempo/Guards					
Rule 1	G_position, low_possessions_Period1, high_pace_Period1, medium_possessions_Period3	L_Result	0.0050	0.6199	1.2381
Tempo/Centers					
	No rules found for this category and position.				
Tempo/Forwards					
Rule 1	F_position, medium_possessions_Period1, low_pace_Period4, low_possessions_Period4	L_Result	0.0056	0.6111	1.2205
Rule 2	F_position, high_possessions_Period1, high_pace_Period2, high_possessions_Period3, high_possessions_Period4	W_Result	0.0059	0.6000	1.2017

Appendix D: Detailed Player Scores by Category and Quarter

This section details the performance scores of players across the five categories, broken down by game quarter. Tables 9 - 13 showcase composite scores and quarter-specific metrics for top-performing players, emphasizing how performance patterns differ across quarters and game contexts. Red formatting highlights the top 20% values in each column.

Table 9: Ball Handling Summary & by Quarter

		Ball Handling					
	Summary						
nameI	**Season** ↓	**Composite_Ball Handling** ↓	**assistPercentage**	**assistToTurnover**	**assistRatio**	**turnoverRatio**	**games/quarters_played**
S. Nash	Regular	2.179	2.508	1.639	2.258	1.672	631
Trae Young	Regular	2.122	2.472	1.696	1.937	1.631	404
C. Paul	Regular	2.095	2.445	1.428	2.156	1.424	1257
J. Wall	Regular	2.077	2.403	1.616	1.974	1.640	610
L. Doncic	Regular	2.044	2.383	1.641	1.801	1.618	394
S. Nash	Playoffs	2.302	2.676	1.620	2.598	1.743	66
R. Rondo	Playoffs	2.146	2.460	1.447	2.444	1.616	129
C. Paul	Playoffs	2.145	2.493	1.428	2.390	1.585	149
R. Westbrook	Playoffs	2.116	2.463	1.583	2.150	1.741	117
L. James	Playoffs	2.097	2.413	1.571	2.155	1.703	286

	Quarter 1						
S. Nash	Regular	2.240	2.554	1.718	2.355	1.681	408
C. Paul	Regular	2.176	2.524	1.447	2.378	1.433	924
Trae Young	Regular	2.176	2.507	1.776	2.031	1.660	294
J. Wall	Regular	2.156	2.491	1.677	2.104	1.642	424
L. Doncic	Regular	2.116	2.509	1.693	1.823	1.597	293
C. Paul	Playoffs	2.234	2.560	1.596	2.541	1.771	109

R. Westbrook	Playoffs	2.224	2.640	1.610	2.300	1.750	100
M. Conley	Playoffs	2.204	2.481	1.469	2.556	1.407	81
S. Nash	Playoffs	2.198	2.545	1.545	2.432	1.705	44
R. Rondo	Playoffs	2.171	2.478	1.500	2.478	1.663	92
	Quarter 2						
S. Nash	Regular	2.162	2.495	1.611	2.267	1.720	457
Trae Young	Regular	2.152	2.547	1.685	1.966	1.654	298
C. Paul	Regular	2.098	2.477	1.384	2.161	1.395	988
B. Simmons	Regular	2.078	2.349	1.580	2.151	1.660	238
J. Wall	Regular	2.058	2.395	1.601	1.942	1.683	479
S. Nash	Playoffs	2.296	2.708	1.479	2.646	1.604	48
C. Paul	Playoffs	2.194	2.553	1.415	2.512	1.537	123
R. Rondo	Playoffs	2.141	2.464	1.423	2.443	1.598	97
J. Harden	Playoffs	2.089	2.406	1.486	2.239	1.667	138
L. James	Playoffs	2.079	2.384	1.563	2.122	1.681	229
	Quarter 3						
S. Nash	Regular	2.238	2.585	1.709	2.305	1.695	436
Trae Young	Regular	2.165	2.510	1.764	1.969	1.620	292
J. Wall	Regular	2.155	2.481	1.735	2.040	1.656	453
C. Paul	Regular	2.137	2.497	1.514	2.155	1.471	940
L. Doncic	Regular	2.130	2.446	1.854	1.837	1.718	294
S. Nash	Playoffs	2.393	2.733	1.889	2.644	1.956	45
R. Rondo	Playoffs	2.242	2.602	1.538	2.527	1.656	93
L. James	Playoffs	2.169	2.494	1.636	2.264	1.715	239
R. Westbrook	Playoffs	2.152	2.500	1.663	2.130	1.728	92
C. Paul	Playoffs	2.139	2.508	1.383	2.375	1.542	120
	Quarter 4						

S. Nash	Regular	2.085	2.403	1.530	2.119	1.587	385
Lam. Ball	Regular	2.047	2.333	1.593	1.926	1.496	135
Trae Young	Regular	2.002	2.334	1.573	1.782	1.601	293
D. Williams	Regular	1.999	2.289	1.485	1.949	1.520	571
B. Davis	Regular	1.987	2.298	1.436	1.911	1.454	326
S. Nash	Playoffs	2.319	2.714	1.571	2.667	1.714	42
R. Rondo	Playoffs	2.074	2.360	1.360	2.395	1.570	86
D. Williams	Playoffs	2.062	2.379	1.379	2.258	1.545	66
L. James	Playoffs	2.058	2.387	1.549	2.029	1.676	204
D. Lillard	Playoffs	2.017	2.302	1.528	2.019	1.660	53

	Quarter 5+						
D. Fox	Regular	2.005	2.421	1.211	2.053	1.368	19
B. Davis	Regular	2.000	2.400	1.133	2.067	1.133	15
Trae Young	Regular	1.989	2.263	1.474	1.947	1.474	19
S. Nash	Regular	1.976	2.324	1.412	1.912	1.588	34
J. Kidd	Regular	1.966	2.276	1.138	2.138	1.138	29
M. Conley	Playoffs	1.869	2.154	1.154	2.000	1.385	13
L. James	Playoffs	1.860	2.200	1.100	2.000	1.500	10
M. Ginobili	Playoffs	1.780	1.900	1.200	2.000	1.400	10
R. Rondo	Playoffs	1.670	1.800	1.100	1.800	1.400	10
T. Parker	Playoffs	1.600	1.643	1.143	1.643	1.286	14

Table 10: Defense Summary & by Quarter

		Defense			
	Summary				
nameI	**Season ↓**	**Composite_Defense ↓**	**defensiveReboundPercentage**	**defensiveRating**	**games/quarters_played**
T. Duncan	Regular	2.248	2.323	1.784	868
K. Garnett	Regular	2.222	2.319	1.819	756
D. Mutombo	Regular	2.200	2.048	1.735	186
R. Gobert	Regular	2.170	2.380	1.920	724
J. Embiid	Regular	2.163	2.433	1.953	428
K. Garnett	Playoffs	2.282	2.452	1.791	96
D. Jordan	Playoffs	2.231	2.653	1.950	59
G. Antetokounmpo	Playoffs	2.208	2.521	1.927	79
B. Wallace	Playoffs	2.203	2.235	1.811	76
T. Duncan	Playoffs	2.197	2.422	1.899	169
	Quarter 1				
T. Duncan	Regular	2.330	2.419	1.708	785
K. Garnett	Regular	2.314	2.415	1.730	677
D. Mutombo	Regular	2.290	2.110	1.632	136
R. Gobert	Regular	2.239	2.538	1.889	610
T. Splitter	Regular	2.239	2.024	1.669	251
D. Jordan	Playoffs	2.347	2.745	1.824	51
T. Duncan	Playoffs	2.336	2.459	1.717	159
Z. Randolph	Playoffs	2.305	2.317	1.700	60
K. Garnett	Playoffs	2.304	2.438	1.753	89

Z. Ilgauskas	Playoffs	2.287	2.333	1.733	60

	Quarter 2				
T. Duncan	Regular	2.254	2.325	1.776	747
K. Garnett	Regular	2.220	2.248	1.792	638
D. Mutombo	Regular	2.216	2.065	1.720	107
L. Wright	Regular	2.205	2.075	1.739	161
J. Sullinger	Regular	2.195	2.114	1.771	201
K. Garnett	Playoffs	2.345	2.446	1.699	83
M. Gortat	Playoffs	2.264	2.396	1.792	53
G. Antetokounmpo	Playoffs	2.263	2.493	1.836	67
S. O'Neal	Playoffs	2.253	2.378	1.800	45
L. Odom	Playoffs	2.251	2.307	1.773	88

	Quarter 3				
T. Duncan	Regular	2.220	2.308	1.818	792
A. Bogut	Regular	2.183	2.252	1.846	532
D. Howard	Regular	2.182	2.436	1.927	1003
K. Garnett	Regular	2.181	2.307	1.873	670
E. Okafor	Regular	2.175	2.225	1.847	542
K. Garnett	Playoffs	2.315	2.473	1.753	93
J. Noah	Playoffs	2.277	2.528	1.830	53
C. Boozer	Playoffs	2.252	2.616	1.904	73
L. Deng	Playoffs	2.247	2.083	1.683	60
B. Wallace	Playoffs	2.246	2.305	1.780	59

	Quarter 4				
T. Duncan	Regular	2.167	2.233	1.861	524
C. Aldrich	Regular	2.162	2.113	1.817	115
K. Garnett	Regular	2.156	2.306	1.908	530

J. Embiid	Regular	2.154	2.379	1.943	280
D. Howard	Regular	2.153	2.309	1.914	871
A. Varejao	Playoffs	2.294	2.038	1.596	52
D. Howard	Playoffs	2.265	2.591	1.875	88
Thaddeus Young	Playoffs	2.195	1.909	1.682	44
G. Antetokounmpo	Playoffs	2.186	2.333	1.877	57
T. Chandler	Playoffs	2.165	1.980	1.755	49

	Quarter 5+				
J. Pargo	Regular	2.500	1.800	1.200	10
L. Nance Jr.	Regular	2.446	2.231	1.462	13
J. Magloire	Regular	2.440	1.600	1.200	10
C. Boozer	Regular	2.385	2.350	1.600	20
K. Perkins	Regular	2.333	2.000	1.524	21
L. James	Playoffs	2.280	2.000	1.600	10
J. Noah	Playoffs	2.270	2.200	1.700	10
D. Green	Playoffs	2.192	1.667	1.583	12
M. Ginobili	Playoffs	2.160	1.600	1.600	10
M. Gasol	Playoffs	2.157	1.857	1.714	14

Table 11: Offense Summary & by Quarter

		Offense						
	Summary							
nameI	**Season ↓**	**Composite_Offense ↓**	**offensiveReboundPercentage**	**effectiveFieldGoalPercentage**	**trueShootingPercentage**	**usagePercentage**	**offensiveRating**	**games/quarters_played**
J. Embiid	Regular	2.304	1.565	1.968	2.353	2.835	2.165	428
Z. Williamson	Regular	2.291	1.583	2.113	2.417	2.614	2.213	184
L. Doncic	Regular	2.248	1.233	1.965	2.243	2.851	2.137	394
L. James	Regular	2.241	1.295	2.009	2.313	2.740	2.123	1408
K. Durant	Regular	2.238	1.184	2.027	2.379	2.702	2.139	1058
N. Jokic	Playoffs	2.391	1.731	2.457	2.386	2.646	2.278	79
K. Durant	Playoffs	2.326	1.208	2.354	2.359	2.750	2.186	170
G. Antetokounmpo	Playoffs	2.313	1.556	2.363	2.251	2.772	2.015	79
A. Davis	Playoffs	2.300	1.623	2.358	2.354	2.575	2.113	59
L. James	Playoffs	2.295	1.382	2.317	2.277	2.733	2.092	286

	Quarter 1							
J. Embiid	Regular	2.333	1.603	1.976	2.365	2.861	2.231	373
Z. Williamson	Regular	2.312	1.610	2.171	2.463	2.549	2.299	164
L. Doncic	Regular	2.270	1.266	1.973	2.229	2.911	2.150	293
K. Towns	Regular	2.253	1.694	2.091	2.343	2.525	2.196	516
K. Durant	Regular	2.247	1.196	2.051	2.384	2.723	2.115	963
N. Jokic	Playoffs	2.420	1.759	2.556	2.463	2.593	2.315	54
A. Davis	Playoffs	2.350	1.768	2.375	2.339	2.661	2.161	56
S. O'Neal	Playoffs	2.338	1.750	2.417	2.250	2.833	1.896	48
J. Embiid	Playoffs	2.313	1.625	2.161	2.250	2.750	2.214	56
D. Booker	Playoffs	2.307	1.357	2.357	2.357	2.762	2.000	42

	Quarter 2							
J. Embiid	Regular	2.309	1.540	1.974	2.374	2.839	2.167	348
Z. Williamson	Regular	2.301	1.646	2.061	2.378	2.677	2.226	164
L. James	Regular	2.270	1.308	2.045	2.359	2.758	2.155	1107
L. Doncic	Regular	2.252	1.205	1.970	2.262	2.832	2.178	298
K. Durant	Regular	2.241	1.192	2.040	2.412	2.655	2.177	984
N. Jokic	Playoffs	2.357	1.733	2.433	2.350	2.600	2.233	60
K. Durant	Playoffs	2.354	1.244	2.350	2.406	2.700	2.344	160
L. James	Playoffs	2.328	1.362	2.349	2.336	2.742	2.157	229
G. Antetokounmpo	Playoffs	2.300	1.493	2.373	2.254	2.672	2.119	67
A. Davis	Playoffs	2.300	1.418	2.364	2.455	2.600	2.073	55

	Quarter 3							
Z. Williamson	Regular	2.297	1.572	2.127	2.428	2.602	2.241	166
J. Embiid	Regular	2.296	1.520	2.011	2.360	2.813	2.131	358
L. Doncic	Regular	2.265	1.248	1.963	2.296	2.861	2.150	294
K. Durant	Regular	2.255	1.197	2.065	2.406	2.714	2.135	948
D. Mitchell	Regular	2.254	1.238	2.014	2.269	2.757	2.231	424
G. Antetokounmpo	Playoffs	2.426	1.529	2.500	2.412	2.853	2.176	68
N. Jokic	Playoffs	2.384	1.689	2.459	2.377	2.672	2.230	61
J. Embiid	Playoffs	2.380	1.451	2.333	2.392	2.824	2.216	51
K. Durant	Playoffs	2.353	1.182	2.396	2.384	2.792	2.208	159
D. Booker	Playoffs	2.331	1.286	2.381	2.262	2.690	2.333	42

	Quarter 4							
J. Embiid	Regular	2.279	1.604	1.900	2.293	2.846	2.132	280
L. James	Regular	2.258	1.308	1.955	2.281	2.816	2.178	948
Z. Williamson	Regular	2.250	1.493	2.090	2.396	2.632	2.069	144

K. Durant	Regular	2.210	1.144	1.960	2.315	2.711	2.137	818
D. Mitchell	Regular	2.208	1.245	1.871	2.145	2.774	2.242	380
N. Jokic	Playoffs	2.410	1.750	2.375	2.354	2.729	2.354	48
K. Durant	Playoffs	2.338	1.203	2.353	2.376	2.774	2.195	133
A. Stoudemire	Playoffs	2.334	1.723	2.277	2.447	2.596	2.191	47
S. Curry	Playoffs	2.325	1.148	2.357	2.322	2.765	2.226	115
L. James	Playoffs	2.289	1.426	2.186	2.196	2.828	2.108	204

	Quarter 5+							
J. Clarkson	Regular	2.306	1.000	2.188	2.313	2.688	2.500	16
K. Irving	Regular	2.293	1.200	1.833	2.400	2.800	2.433	30
G. Arenas	Regular	2.218	1.235	2.000	2.353	2.588	2.235	17
S. Curry	Regular	2.216	1.158	1.974	2.368	2.737	2.053	38
D. Wade	Regular	2.206	1.354	2.000	2.229	2.583	2.250	48
L. James	Playoffs	2.480	1.800	2.500	2.400	3.000	2.100	10
D. Lillard	Playoffs	2.264	1.182	2.273	2.273	2.727	2.091	11
P. Pierce	Playoffs	2.091	1.273	2.000	2.091	2.545	1.909	11
M. Conley	Playoffs	2.038	1.231	1.923	1.923	2.385	2.154	13
T. Parker	Playoffs	2.029	1.143	2.143	2.143	2.143	2.071	14

Table 12: Overall Impact Summary & by Quarter

		Overall Impact				
nameI	Summary Season ↓	Composite_Overall Impact ↓	PIE	netRating	reboundPercentage	games/quarters_played
J. Embiid	Regular	2.355	2.525	2.149	2.516	428
N. Jokic	Regular	2.323	2.495	2.123	2.451	658
T. Duncan	Regular	2.310	2.416	2.175	2.446	868
L. James	Regular	2.293	2.521	2.119	2.057	1408
K. Garnett	Regular	2.276	2.420	2.106	2.391	756
A. Davis	Playoffs	2.321	2.495	2.099	2.533	59
N. Jokic	Playoffs	2.319	2.561	2.040	2.484	79
L. James	Playoffs	2.313	2.538	2.099	2.258	286
K. Garnett	Playoffs	2.288	2.400	2.130	2.496	96
K. Leonard	Playoffs	2.279	2.416	2.147	2.260	139

	Quarter 1					
J. Embiid	Regular	2.429	2.598	2.225	2.587	373
T. Duncan	Regular	2.371	2.504	2.208	2.501	785
N. Jokic	Regular	2.356	2.537	2.142	2.506	445
K. Garnett	Regular	2.339	2.504	2.148	2.462	677
R. Gobert	Regular	2.316	2.430	2.134	2.623	610
A. Davis	Playoffs	2.430	2.589	2.232	2.607	56
J. Embiid	Playoffs	2.351	2.411	2.250	2.536	56
K. Leonard	Playoffs	2.331	2.477	2.192	2.300	130
L. James	Playoffs	2.316	2.593	2.068	2.182	236
N. Jokic	Playoffs	2.313	2.593	2.000	2.463	54

	Quarter 2					
N. Jokic	Regular	2.337	2.484	2.180	2.384	471
J. Embiid	Regular	2.333	2.480	2.155	2.468	348
L. James	Regular	2.312	2.511	2.170	2.051	1107
T. Duncan	Regular	2.285	2.365	2.171	2.438	747
C. Capela	Regular	2.264	2.351	2.115	2.544	485
K. Garnett	Playoffs	2.364	2.446	2.241	2.554	83
G. Antetokounmpo	Playoffs	2.331	2.522	2.104	2.493	67
J. Embiid	Playoffs	2.329	2.482	2.125	2.554	56
L. James	Playoffs	2.320	2.528	2.127	2.249	229
T. Duncan	Playoffs	2.316	2.500	2.090	2.507	144

	Quarter 3					
J. Embiid	Regular	2.339	2.542	2.089	2.547	358
T. Duncan	Regular	2.298	2.434	2.128	2.455	792
N. Jokic	Regular	2.297	2.501	2.045	2.517	489
R. Gobert	Regular	2.290	2.412	2.105	2.574	592
K. Durant	Regular	2.278	2.495	2.114	2.037	948
G. Antetokounmpo	Playoffs	2.374	2.662	2.044	2.559	68
A. Davis	Playoffs	2.373	2.642	2.057	2.585	53
N. Jokic	Playoffs	2.342	2.607	2.033	2.541	61
K. Garnett	Playoffs	2.326	2.398	2.226	2.452	93
J. Embiid	Playoffs	2.314	2.412	2.176	2.490	51

	Quarter 4					
L. James	Regular	2.323	2.549	2.138	2.138	948
N. Jokic	Regular	2.306	2.471	2.120	2.397	408
J. Embiid	Regular	2.305	2.457	2.121	2.450	280
T. Duncan	Regular	2.284	2.342	2.204	2.384	524

A. Davis	Regular	2.283	2.485	2.047	2.434	555
N. Jokic	Playoffs	2.354	2.521	2.188	2.354	48
L. James	Playoffs	2.335	2.520	2.142	2.373	204
K. Leonard	Playoffs	2.323	2.420	2.250	2.210	100
K. Durant	Playoffs	2.292	2.444	2.195	2.045	133
S. Nash	Playoffs	2.274	2.548	2.167	1.524	42

	Quarter 5					
L. Nance Jr.	Regular	2.465	2.538	2.462	2.154	13
J. Hill	Regular	2.436	2.636	2.273	2.273	11
J. Pargo	Regular	2.430	2.300	2.700	1.800	10
E. Payton	Regular	2.400	2.588	2.353	1.765	17
J. Embiid	Regular	2.339	2.611	2.056	2.389	18
L. James	Playoffs	2.460	2.600	2.400	2.100	10
M. Ginobili	Playoffs	2.305	2.200	2.500	1.900	10
J. Noah	Playoffs	2.155	2.300	2.000	2.200	10
D. Green	Playoffs	2.154	2.083	2.333	1.667	12
K. Leonard	Playoffs	2.145	2.200	2.100	2.100	10

Table 13: Tempo Summary & by Quarter

		Tempo			
	Summary				
nameI	**Season ↓**	**Composite_Tempo ↓**	**pace**	**possessions**	**games/quarters_played**
Lam. Ball	Regular	2.475	2.513	2.418	183
Trae Young	Regular	2.471	2.479	2.460	404
J. Williams	Regular	2.449	2.528	2.330	145
Z. Williamson	Regular	2.448	2.453	2.440	184
J. Morant	Regular	2.407	2.451	2.342	245
A. Wiggins	Playoffs	2.620	2.600	2.650	40
A. Davis	Playoffs	2.458	2.335	2.642	59
G. Antetokounmpo	Playoffs	2.439	2.332	2.598	79
S. Curry	Playoffs	2.428	2.375	2.508	147
D. Booker	Playoffs	2.403	2.233	2.659	47

	Quarter 1				
Trae Young	Regular	2.630	2.578	2.707	294
S. Curry	Regular	2.630	2.484	2.847	793
Lam. Ball	Regular	2.628	2.691	2.533	152
A. Edwards	Regular	2.614	2.497	2.790	286
J. Morant	Regular	2.600	2.584	2.624	202
A. Davis	Playoffs	2.686	2.536	2.911	56
S. Curry	Playoffs	2.668	2.558	2.833	120
A. Wiggins	Playoffs	2.620	2.600	2.650	40
G. Antetokounmpo	Playoffs	2.618	2.522	2.761	67
D. Booker	Playoffs	2.590	2.333	2.976	42

	Quarter 2				
Z. Williamson	Regular	2.589	2.652	2.494	164
J. Williams	Regular	2.553	2.679	2.364	140
Trae Young	Regular	2.552	2.674	2.369	298
B. Simmons	Regular	2.520	2.534	2.500	238
Lam. Ball	Regular	2.505	2.664	2.267	146
K. Thompson	Playoffs	2.544	2.458	2.673	153
C. McCollum	Playoffs	2.511	2.339	2.768	56
G. Antetokounmpo	Playoffs	2.484	2.478	2.493	67
E. Gordon	Playoffs	2.477	2.419	2.565	62
D. Green	Playoffs	2.455	2.565	2.290	124

	Quarter 3				
J. Morant	Regular	2.499	2.396	2.655	197
D. Booker	Regular	2.462	2.278	2.738	526
Trae Young	Regular	2.461	2.366	2.603	292
S. Curry	Regular	2.459	2.259	2.760	792
J. Green	Regular	2.434	2.235	2.732	213
A. Davis	Playoffs	2.547	2.283	2.943	53
D. Booker	Playoffs	2.471	2.167	2.929	42
G. Antetokounmpo	Playoffs	2.435	2.294	2.647	68
S. Curry	Playoffs	2.427	2.222	2.735	117
D. Lillard	Playoffs	2.427	2.100	2.917	60

	Quarter 4				
Lam. Ball	Regular	2.316	2.274	2.378	135
B. Mathurin	Regular	2.312	2.336	2.276	116
J. Williams	Regular	2.302	2.197	2.459	122
Trae Young	Regular	2.287	2.311	2.253	293

F. Jackson	Regular	2.277	2.397	2.096	146
C. McCollum	Playoffs	2.404	2.208	2.698	53
D. Lillard	Playoffs	2.377	2.302	2.491	53
R. Westbrook	Playoffs	2.231	2.156	2.344	90
E. Gordon	Playoffs	2.226	2.129	2.371	62
K. Middleton	Playoffs	2.209	2.119	2.343	67

	Quarter 5+				
N. Mirotic	Regular	1.917	2.417	1.167	12
E. Payton	Regular	1.871	2.294	1.235	17
J. Meeks	Regular	1.857	2.286	1.214	14
G. Vasquez	Regular	1.855	2.364	1.091	11
K. Bates-Diop	Regular	1.850	2.417	1.000	12
T. Allen	Playoffs	1.692	2.154	1.000	13
R. Rondo	Playoffs	1.660	2.100	1.000	10
R. Allen	Playoffs	1.646	2.077	1.000	13
Z. Randolph	Playoffs	1.600	2.000	1.000	12
D. Green	Playoffs	1.600	2.000	1.000	12