

# A NEW PERSPECTIVE ON ANALYSIS AND EVALUATION OF BASKETBALL TEAM PERFORMANCE AND PLAYERS USING CLUSTER ANALYSIS

Ajinaja, M.O.<sup>1</sup>, Abiona, A.A.<sup>1</sup> and Adewuyi, J.A.<sup>2</sup>

<sup>1</sup>Department of Computer Science, Federal Polytechnic Ile-Oluji, Ondo State, Nigeria

<sup>2</sup>Department of Computer Science, Federal University of Technology, Akure, Ondo State, Nigeria

Corresponding email: [ajinajalekan@gmail.com](mailto:ajinajalekan@gmail.com)

## ABSTRACT

*Basketball is a game that demands a complete list of criteria to be enumerated in order to fully comprehend the game and assess strategy and decisions while limiting unpredictability. Its success heavily depends on identifying the players' playing styles both offensive and defensive, preferred and assembling the ideal team. This paper reviews a new perspective on analysis and evaluation of basketball team performance and players using both h-cluster and k-cluster techniques. Though there are known techniques for analysis and player performance evaluation which have been developed overtime, several other techniques have been proposed. However, cluster analysis hasn't been much explored in player performance. This study used both the k-cluster and the h-cluster to assess and analyze the basketball team's and players' performance. There are 23 factors in the NBA player statistics from the 2017–2018 regular season that describe a team's overall offensive, defensive, offensive rebounding, and general capabilities. After standardizing the variables, two clustering techniques—hierarchical clustering and K-means clustering—were used to conduct the cluster analysis. K-means clustering produced superior outcomes than h-cluster. According to the study, participants in cluster 3 spend a lot more time than those in the other 3 clusters. The cluster 2 players spend the least amount of time. While players in cluster 2 scored very little throughout the 2017–2018 season, players in cluster 3 end up with a lot more points. For offensive rebound (OREB), players in cluster 4 had higher values compared to other clusters. While cluster 4 had the highest blocks meaning players in cluster 4 are mostly defensive players. An overall analysis showed in details which player performed better generally and where to improve for next season.*

**Keywords:** evaluation, cluster, k-cluster, h-cluster

## 1 INTRODUCTION

**B**asketball has been around for more than a century, and as the game changed and added new laws, so did the players. The NBA has advanced to the point where Point Guards pull down 10 or more rebounds and Centers make 3-pointers with accuracy. Players 7 feet tall or above are adept at handling the ball as the primary ball handler, while players under 6 feet 6 receive playing time as Centers. With 82 games in a regular season that last roughly 24 weeks, the National Basketball Association (NBA) of the United States of America is the most competitive basketball league in the world. Throughout the entire competition time, the coaching staff must plan and supervise the training loads placed on players, a challenging process that puts a lot of physiological strain on the athletes and provide detailed and concise evaluation based on data gathered from the players every single match. For every match, there is an intense amount of preparation and process that had gone under. All players are required to manage their work load, study game plans and eat right whether they are starting or not. It is necessary to make adjustments throughout the season to a

number of changing unpredictable constraints like player injuries. Thus, it is essential for optimizing the choices made by the coaching staff regarding the individual training loads that practice sessions and game performance are continuously planned and monitored [1]. Using game-related statistics to assess technical and tactical conduct, as well as the effectiveness of players and teams throughout the season, is one of the most popular ways to keep track of sports performance. In studies reporting these characteristics, data from European league games is routinely included, but not data from NBA games. This research, which makes use of NBA data, attempts to remedy this discrepancy. In order to comprehend how players, interact with one another by spotting affordances, performance variables describe the dualities of the performer and the environment [2]. For instance, the assists most likely came about because open teammates provided the ball carrier with affordances. In fact, perception-action coupling suggests that movement and information are interdependent [3]. Information drives movement, and movement drives information available to players. In this way, game-related statistics might shed light on players' perceptions and behavior. Additionally, they might offer a measure of co-adaptation in the way that individuals work together as a team to adapt to little yet significant changes in one another's structure and function [4]. Following the turn of the millennium and in the early years of the new century, data and mathematical approaches to basketball emerged. Economists and statisticians tended to do this type of analysis for academic or recreational reasons. The National Basketball Association (NBA) began to pay closer attention as a sizable fan-based movement of data junkies grew daily. Numerous websites, including 82games, fivethirtyeight, Basketball-Reference, and the NBA website, now provide a very wide and comprehensive range of data, statistics, and in-depth analysis about the NBA [5]. With the focus of this research being basketball, data-driven analysis paired with machine learning is increasingly used in most sporting events today. Sports provide a lot of information about the participants, the effectiveness of the teams, and the spectators. The vast volume of data at hand makes it possible to examine and make use of a lot of this data for player evaluation and team performance generally. It is elevating sports to a new level. Different individuals have worked on using data from team performance to evaluate or determine other aspects in sport. According on their on-court performances and statistical data over the course of a season, basketball players' roles and positions will be redefined as part of the initiative. Clustering techniques will be employed to evaluate and analyze performance of teams and players.

## **2 LITERATURE REVIEW**

Using machine learning approaches (random forest, linear and non-linear regression models, multiple regression models, deep learning, and artificial neural networks (ANN)), [6] investigated the effects of injuries on basketball player and team performance. They determined which general health conditions and injuries were most prevalent in the NBA league over the past ten years (RQ1), which teams and players sustained the most injuries (RQ2), and how injuries affected NBA player and team performance in general from 2010 to 2020. (RQ3). No standardized form-based injury factor existed, which would have facilitated data entry for injury diagnosis and classification. [7] investigated how to increase youth interest in math and science by using basketball analytics. It looked at how sports analytics can give students who aren't typically interested in STEM (Science, Technology, Engineering, and Mathematics) a rigorous, yet practical application of math and statistics in data collection and analysis, which in turn allows them to be intrinsically motivated to use STEM concepts as a tool for basketball training. [8] sought to deliver important insights and advanced analytics for teams, technical personnel, and players in order to satisfy the growing need for innovative methodologies. Predicting a player's performance depends on a number of factors, including psychology, injury risk [9], poor shots in the first minutes of a game, and opponent match-ups, all of which have a significant impact on future performance but are challenging to describe and measure [6]. However, technical personnel and data analysts must assess these indicators, monitor and track performance, and ultimately make critical judgments for the team's upcoming hirings and selections [10]. The plus-minus statistic was enhanced in [11], and a new visual platform for investigating derived values was also included. [12] attempted to forecast NBA talent using a vast amount of tracking data from college basketball. [13] introduced a framework for machine-learning modelling for games' outcomes and compared nine most commonly used algorithms for classification. The AdaBoost algorithm outperformed the others in this classification task. [14] developed a data-driven basketball web application for support in making decisions. The Basketball Coach Assistant (BCA)

gave decision-makers the crucial data they needed to train and coach basketball teams. The outcomes of the BCA information system's utilization were found to be beneficial for planning training exercises and for making strategic, tactical, and operational choices. [15] worked on heart rate analysis of high-level basketball players during training sessions. Leb Oro's team of eleven basketball players wore HR monitors for 12 training sessions, yielding 238 HR files. According to the positions they played during each exercise, the number of players, style of exercise, type of defense, field size, and duration were all taken into consideration (Guard, Forward and Centers). Based on the number of players or field size, no discernible variations were discovered. [16] applied data analytics as an alternative to subjective rankings of players in fantasy basketball. [17] explored factor analysis of measures of basketball players' physical and functional readiness. The study's findings showed that adding functional loop training (also known as suspension training) using TRX® to a basketball player's training regimen significantly improved both their overall physical performance and aerobic capacity. [18] explained how big data along with the SAPHANA platform has benefitted different sports using football and basketball as case study. [19] created an algorithm (Court Detection, Pedestrian Detection by Histogram of Oriented Gradients, and Color-Based Detection and Classification) that follows the motions of ten different players in a video of a basketball game. The goal of player monitoring was to give basketball coaches and organizations as much data as possible so they could better create defensive and offensive strategies. The model had a high degree of identification and tracking of the players in the court.

### 3 METHODOLOGY

Statistics for every player for 2017/2018 season were gathered by scraping data from Basketball-Reference. Using custom R [20] scripts to gather box score data from basketball reference. Precisely there are 30 teams and had initially 23 variables related to their performance.

#### 3.1 Clustering

For cluster analysis, quantitatively variables are needed hence 2 columns were removed. This was done to remove any data that might not case the algorithm to have an error as it requires only quantitative variables to run. The first column in the dataset specifies the name of the team. The fundamental box-score count statistic totals are abbreviated as in Table 1.

Table 1. Acronyms for the basic box-score count statistic

aCRONYMS FOR THE BASIC BOX-SCORE	
1	FTM free-throws made
2	P3A three-point shots attempted
3	ORB offensive rebounds
4	DRB defensive rebounds
5	TRB all rebounds
6	P2A two-point shots attempted
7	P3M three-point shots made
8	TOV turnovers
9	FCM fouls committed
10	FRV fouls received
11	FTA free-throws attempted
12	P2M two-point shots made
13	AST assists
14	BLK blocks
15	STL steals

Analysis was done using R [20] and packages ggplot2 [21]. One of the most straightforward and well-liked unsupervised machine learning algorithms is K-means clustering. Unsupervised algorithms typically draw conclusions from datasets using only input vectors without taking into account predetermined or labelled results. The x-axis captured the number of clusters while the y-axis captured BD/TD. BD is between deviant which captures the variability between the clusters and TD is the total variability.

### 3.2 K-means clustering

K-means clustering [22] aims to partition the  $n$  observations into  $N (\leq n)$  sets  $S = 1, 2, \dots, N_k$  so as to reduce the within group sum of squares ( $SS_w$ ). To be specific, its objective is to find:

$$\arg \min \sum_{i=1}^N \sum_{x \in S_i} \|x - m_i\|^2 \quad (1)$$

where  $m_i$  is the cluster center of  $S_i$ . In K-means clustering, the following presumptions are necessary:

1. The number of clusters is  $N = 6$ , according to the result of hierarchical clustering.
2. The distance measure between observations is Euclidean distance,  $dist(x_i, x_j) = \|x_i - x_j\|$ .

Introducing the algorithm:

1. Randomly choose  $N$  "cluster centers"  $m_1, m_2, \dots, m_N$ .
2. Iterate until convergence ( $k =$  iteration number):

- a. Assign each  $x_i$  to the closest mean

$$S_t^{(k)} = \{x_i : \|x_i - m_i^{(k)}\|^2 \leq \|x_i - m_j^{(k)}\|^2, \forall j, 1 \leq j \leq N\} \quad (2)$$

- b. Calculate the new means to be the centroids of the observation in the new cluster [25].

$$m_i^{(k+1)} = \frac{1}{|S_i^{(k)}|} \sum_{x_j \in S_i^{(k)}} x_j \quad (3)$$

### 3.3 Hierarchical clustering

Hierarchical clustering seeks to build a hierarchy of clusters and the method needs a measure of similarity between clusters of observations as a requirement [23]. The Ward's approach, which tries to reduce the error sum of squares, will be used in this study [24]. Introducing the algorithm; 1. Start with each point in a cluster of its own, 2. Until there is only one cluster with size  $n$ : (a) Find the closest pair of clusters; (b) Merge them; (c) Compute distances between the new cluster and each of the old clusters, and 3. Return the tree of cluster-mergers.

## 4 RESULTS AND DISCUSSION

### 4.1 k-means clustering Result Analysis (k = 6)

Looking at the chart in fig 1, the BD/TD was plotted against the number of clusters in the cluster graph to determine the optimal number of clusters to generate. The difference between each of the point in the graph was studied and number cluster 6 has the highest improvement (greater than 9.5%).

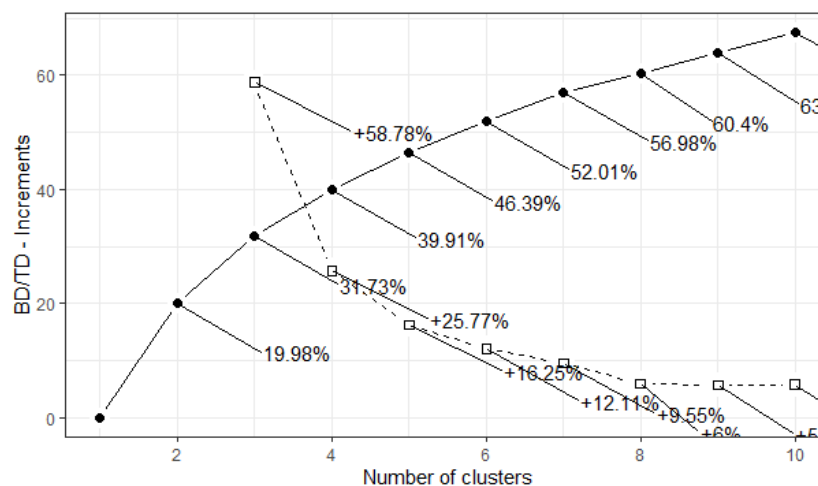


Figure 1. K - cluster graph to determine optimal number of clusters

Six spider plot was generated in fig 2 to show how the teams performed compared to average performance during the season. For example, teams in the first cluster, the dotted blue line is average performance during that season from all the teams, it was noticed that teams performed even lower than average performance in many cases and very slightly in some other cases. Also, on cluster 1, it was noticed the defensive rebound (RREB) was lower than the average performance. For 2 points attempted (PSA), teams under the cluster performed slightly above average. Also, there was an incredible high number of losses for teams in that cluster 1. For cluster 2, teams had high 3 points attempt as compared to the average performance.

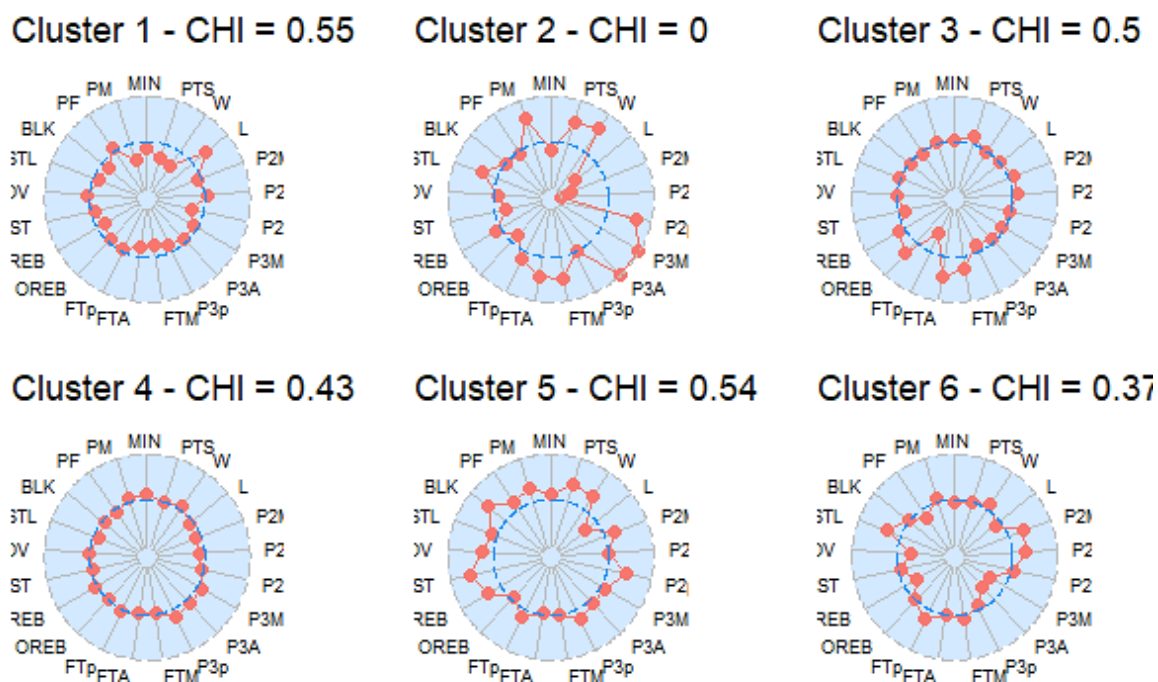


Figure 2. Six (6) k - cluster group

Table 2 shows that cluster each team teams belong to. Milwaukee Bucks and Minnesota Timberwolves both belong to cluster 6 while Cleveland Cavaliers and Boston Celtics belong to cluster 4.

Table 2. Team and corresponding cluster they belong to

Label/Teams	Cluster	Label/Teams	Cluster
1 Atlanta Hawks	1	16 Miami Heat	4
2 Boston Celtics	4	17 Milwaukee Bucks	6
3 Brooklyn Nets	4	18 Minnesota Timberwolves	6
4 Charlotte Hornets	3	19 New Orleans Pelicans	5
5 Chicago Bulls	1	20 New York Knicks	1
6 Cleveland Cavaliers	4	21 Oklahoma City Thunder	3
7 Dallas Mavericks	1	22 Orlando Magic	1
8 Denver Nuggets	4	23 Philadelphia 76ers	5
9 Detroit Pistons	4	24 Phoenix Suns	1
10 Golden State Warriors	5	25 Portland Trail Blazers	4
11 Houston Rockets	2	26 Sacramento Kings	1
12 Indiana Pacers	6	27 San Antonio Spurs	6
13 LA Clippers	3	28 Toronto Raptors	5
14 Los Angeles Lakers	3	29 Utah Jazz	4
15 Memphis Grizzlies	1	30 Washington Wizards	4

The plot generated in figure 2 is based on figure 3. Representation of each team and the cluster they belong to is showed in figure 3. Only 1 team Houston Rockets is in cluster 2 while cluster 6 has 4 teams.

[1]	"Atlanta Hawks"	"Chicago Bulls"	"Dallas Mavericks"	"Memphis Grizzlies"
[5]	"New York Knicks"	"Orlando Magic"	"Phoenix Suns"	"Sacramento Kings"
[1]	"Houston Rockets"			
[1]	"Charlotte Hornets"	"LA Clippers"	"Los Angeles Lakers"	
[4]	"Oklahoma City Thunder"			
[1]	"Boston Celtics"	"Brooklyn Nets"	"Cleveland Cavaliers"	
[4]	"Denver Nuggets"	"Detroit Pistons"		"Miami Heat"
[7]	"Portland Trail Blazers"	"Utah Jazz"	"Washington Wizards"	
[1]	"Golden State Warriors"	"New Orleans Pelicans"	"Philadelphia 76ers"	
[4]	"Toronto Raptors"			
[1]	"Indiana Pacers"	"Milwaukee Bucks"	"Minnesota Timberwolves"	
[4]	"San Antonio Spurs"			

Figure 3. Cluster list

The plots in figure 2 seen for each cluster is based on values in figure 4 where the ridges are formed.

ID	MIN	PTS	W	L	P2M	P2A	
1	-0.4910350	-1.06012689	-1.31933322	1.31933322	-0.2654125	0.22269237	
2	-0.5819675	1.58972072	1.96365875	-1.96365875	-2.6392319	-3.50230942	
3	0.1454919	0.51496861	-0.06136434	0.06136434	0.3563188	0.39386775	
4	0.3475639	-0.05760636	0.30000342	-0.30000342	-0.4810866	-0.39751511	
5	0.4182891	1.40697261	1.08410327	-1.08410327	0.7848568	-0.04146718	
6	-0.2182378	-0.06950332	0.45000513	-0.45000513	1.1319023	0.97220105	
	P2p	P3M	P3A	P3p	FTM	FTA	FTP
1	-0.85769732	-0.3207205	-0.1968527	-0.59670025	-0.81463291	-0.76953822	-0.1434091
2	2.14178890	3.2147867	3.3153025	0.02641482	1.71996866	1.50199025	0.5523260
3	-0.09939465	-0.2780751	-0.1577655	-0.56093592	0.95127085	1.45165901	-1.5590751
4	-0.14795513	0.4457686	0.3276409	0.59901159	-0.07930175	-0.13097897	0.1498614
5	1.40352872	0.3880055	0.2422141	0.72223117	0.10027508	-0.09006643	0.6638469
6	0.20871238	-1.2751653	-1.2567609	-0.32227452	0.32615667	0.09668897	0.7067769
	OREB	DREB	AST	TOV	STL	BLK	PF
1	-0.31432429	-0.6095530	-0.3166840	0.16797807	-0.5075120	-0.54527514	0.3688250
2	-0.74768776	0.4680141	-0.8649711	-0.41074497	1.2147308	-0.04388738	-0.2691072
3	1.21711055	0.3928607	-0.5010084	0.02552317	0.1474899	-0.12808990	-0.1645996
4	-0.03084592	0.2558163	-0.1096227	0.01530071	-0.3975816	-0.21229243	-0.1733086
5	-0.29151568	1.1399739	1.8608437	0.76450797	0.4085746	1.67460962	0.5647769
6	-0.04062104	-1.0063187	-0.2635735	-1.05772764	1.0498352	0.03266037	-0.6806062

Figure 4. Value representing the ridges of the cluster

#### 4.2 k-means clustering Result Analysis (k = 6)

The hierarchical cluster is divided into 6 clusters as seen in figure 5 with the representation of where a particular team belongs to.

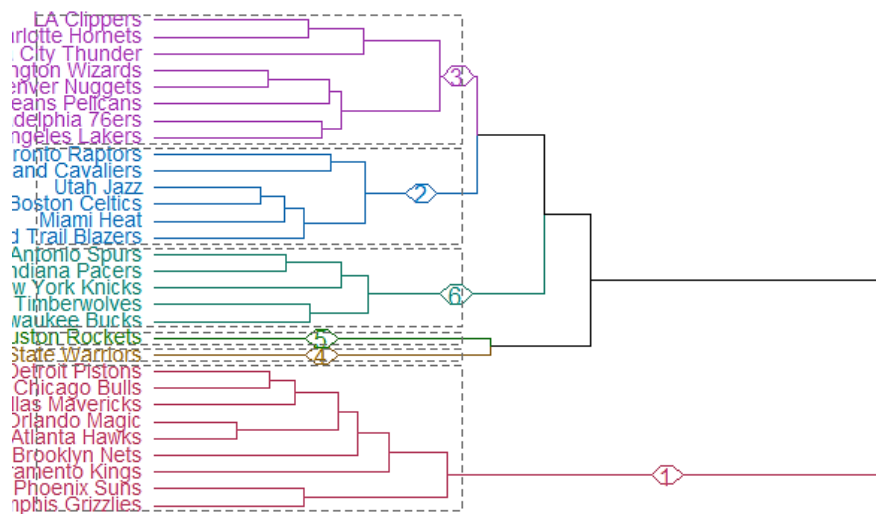


Figure 5. dendrogram h-cluster of teams

Similar groups can be done for players as well. From the earlier observations, there are 605 players and this might be difficult to put all the players in the plot. To reduce the number of players without adversely uttering the result of the evaluation, players who have played more than 2000 minutes of basketball will be used which then gives 99 players. 99 players from the 2017/2018 season played more than 2000 minutes.

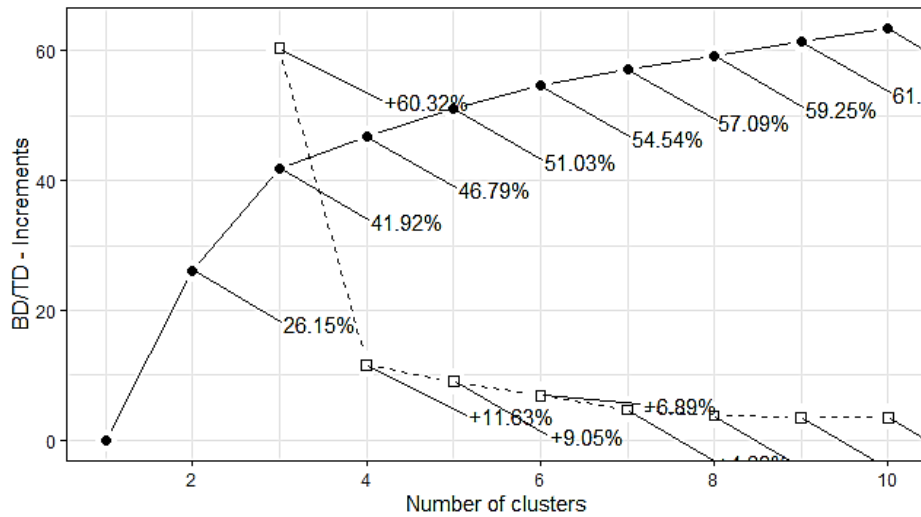
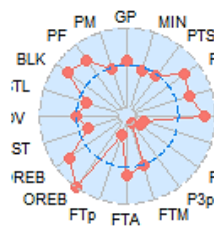


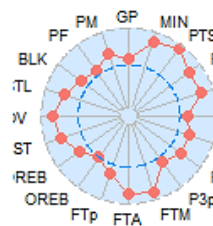
Figure 6. k-Cluster graph for k = 4

From figure 6, moving from cluster 2 to 4 has the highest improvement and also provides the optimal number of clusters to use. Moving from cluster 3 to 4 is 11% and moving from cluster 5 to 6 is 9%, hence cluster 4 is picked as the optimal number. Figure 7 shows the division of k-cluster into 4 cluster. It is observed that cluster 1 has some outstanding performance with both defensive rebound (REB) and offensive rebound (OREB) which means players in cluster 1 have an impressive rebound both offensively and defensively. Players in cluster 1 has have high number of blocks compared to average. Players in cluster 2 also have very high overall performance generally with high assist, high 2 points attempted and number of points made. Also, free throws attempted and made were quite high. Players in cluster 3 are very close to the average (blue dotted line) while players in cluster 4 had below average season.

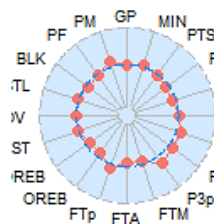
Cluster 1 - CHI = 0.72



Cluster 2 - CHI = 0.76



Cluster 3 - CHI = 0.5



Cluster 4 - CHI = 0.34

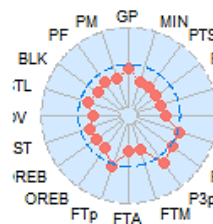


Figure 7. k-cluster group

Figure 8 shows the players who belong to a particular cluster indicating where they belonged to. Players who belonged to cluster 2 had outstanding performances overall. The 3<sup>rd</sup> group of players who were like very close to average and 4<sup>th</sup> group had a poor performance during the 2017/2018 basketball season.

[1]	"Dwight Howard"	"Andre Drummond"	"Clint Capela"
[4]	"DeAndre Jordan"	"Julius Randle"	"Taj Gibson"
[7]	"Steven Adams"	"Ben Simmons"	"TJ Warren"
[10]	"Jusuf Nurkic"	"Willie CauleyStein"	"LaMarcus Aldridge"
[13]	"Derrick Favors"	"Marcin Gortat"	
[1]	"Kemba Walker"	"LeBron James"	"Nikola Jokic"
[4]	"Kevin Durant"	"James Harden"	"Victor Oladipo"
[7]	"Lou Williams"	"Khris Middleton"	"Giannis Antetokounmpo"
[10]	"KarlAnthony Towns"	"Jrue Holiday"	"Anthony Davis"
[13]	"Russell Westbrook"	"Paul George"	"CJ McCollum"
[16]	"Damian Lillard"	"DeMar DeRozan"	"Donovan Mitchell"
[19]	"Bradley Beal"		
[1]	"Taurean Prince"	"Dennis Schroder"	"Jayson Tatum"
[4]	"Al Horford"	"Jaylen Brown"	"Spencer Dinwiddie"
[7]	"Harrison Barnes"	"Will Barton"	"Jamal Murray"
[10]	"Gary Harris"	"Klay Thompson"	"Draymond Green"
[13]	"Eric Gordon"	"Thaddeus Young"	"Bojan Bogdanovic"
[16]	"Kyle Kuzma"	"Marc Gasol"	"Josh Richardson"
[19]	"Goran Dragic"	"Eric Bledsoe"	"Andrew Wiggins"
[22]	"Jeff Teague"	"Jimmy Butler"	"E'Twaun Moore"
[25]	"Carmelo Anthony"	"Robert Covington"	"Dario Saric"
[28]	"Kyle Lowry"	"Serge Ibaka"	"Joe Ingles"
[31]	"Ricky Rubio"	"Otto Porter Jr"	

Figure 8. Players belonging to k-cluster group

### 4.3 The Hierarchical cluster analysis

Figure 9 showed that 3 to 4 cluster would provide the optimal number of clusters as moving from 3 to 4 gave the highest percentage difference.

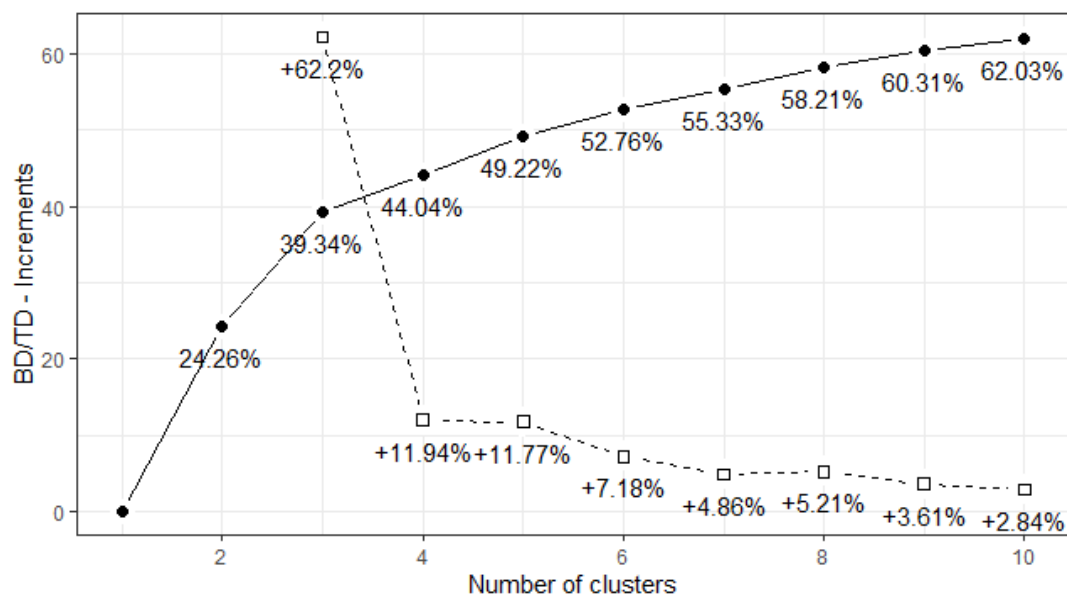


Figure 9. Optimal number of h-cluster

Figures 10 and 11 shows the cluster list for h-cluster and where each player belongs to. The 1<sup>st</sup> cluster contained players like Dennis Schroder in h-cluster. Figure 12 gave the standardized averages in each cluster from 1 to 4.



[1]	"Taurean Prince"	"Dennis Schroder"	"Jayson Tatum"
[4]	"Al Horford"	"Jaylen Brown"	"Spencer Dinwiddie"
[7]	"Harrison Barnes"	"Wesley Matthews"	"Dennis Smith Jr"
[10]	"Will Barton"	"Jamal Murray"	"Nikola Jokic"
[13]	"Gary Harris"	"Klay Thompson"	"Kevin Durant"
[16]	"Draymond Green"	"Eric Gordon"	"Thaddeus Young"
[19]	"Bojan Bogdanovic"	"Darren Collison"	"Austin Rivers"
[22]	"Kyle Kuzma"	"Marc Gasol"	"Josh Richardson"
[25]	"Goran Dragic"	"Jeff Teague"	"Jimmy Butler"
[28]	"Etwan Moore"	"Carmelo Anthony"	"Jonathon Simmons"
[31]	"Robert Covington"	"Dario Saric"	"JJ Redick"
[34]	"TJ Warren"	"Willie CauleyStein"	"DeAaron Fox"
[37]	"Kyle Lowry"	"Serge Ibaka"	"Joe Ingles"
[40]	"Ricky Rubio"	"Otto Porter Jr"	
-----			
[1]	"Terry Rozier"	"Allen Crabbe"	
[3]	"DeMarre Carroll"	"Marvin Williams"	
[5]	"Justin Holiday"	"Denzel Valentine"	
[7]	"Lauri Markkanen"	"JR Smith"	
[9]	"Yogi Ferrell"	"Wilson Chandler"	
[11]	"Ish Smith"	"PJ Tucker"	
[13]	"Trevor Ariza"	"Cory Joseph"	
[15]	"Kentavious CaldwellPope"	"Dillon Brooks"	
[17]	"Tyler Johnson"	"Wayne Ellington"	
[19]	"Tony Snell"	"Courtney Lee"	
[21]	"Dragan Bender"	"AlFarouq Aminu"	
[23]	"Evan Turner"	"Bogdan Bogdanovic"	
[25]	"Buddy Hield"	"Patty Mills"	
[27]	"Kelly Oubre Jr"		

Figure 10. h-cluster list 1 and 2

[1]	"Kemba Walker"	"LeBron James"	"James Harden"
[4]	"Victor Oladipo"	"Lou Williams"	"Khris Middleton"
[7]	"Giannis Antetokounmpo"	"Eric Bledsoe"	"Andrew Wiggins"
[10]	"KarlAnthony Towns"	"Jrue Holiday"	"Anthony Davis"
[13]	"Russell Westbrook"	"Paul George"	"Ben Simmons"
[16]	"CJ McCollum"	"Damian Lillard"	"LaMarcus Aldridge"
[19]	"DeMar DeRozan"	"Donovan Mitchell"	"Bradley Beal"
-----			
[1]	"Dwight Howard"	"Andre Drummond"	"Clint Capela"
[5]	"Julius Randle"	"Taj Gibson"	"Steven Adams"
[9]	"Derrick Favors"	"Marcin Gortat"	

Figure 11. h-cluster list 3 and 4

ID	GP	MIN	PTS	P2M	P2A	P2p
1	1	-0.4310537	-0.1666134	-0.1144805	-0.1743349	-0.1423651
2	2	0.1540230	-0.7386786	-0.9035612	-0.9754435	-0.9560555
3	3	0.4108174	1.3570114	1.5190171	1.2305966	1.2770321
4	4	0.4887413	-0.1721768	-0.2809507	0.7642177	0.4832793
-----						
	P3M	P3A	P3p	FTM	FTA	FTP
1	0.1411888	0.1132364	0.33552360	-0.16858857	-0.2202364	0.1844981
2	0.1720096	0.1595567	0.36775703	-0.79573039	-0.8463349	0.1979324
3	0.3409477	0.4330423	0.01772944	1.34177258	1.3029610	0.2007594
4	-1.7592901	-1.8044612	-2.40582255	0.02196277	0.4518554	-1.7124543
-----						
	OREB	DREB	AST	TOV	STL	BLK
1	-0.26013330	-0.2528221	0.07838189	-0.04347149	-0.07716757	-0.1617369
2	-0.49008448	-0.4152487	-0.57815358	-0.86561874	-0.43141221	-0.5343134
3	0.09714432	0.3638793	0.95086577	1.17812077	0.98016830	0.3639329
4	2.18577154	1.3935956	-0.75716917	0.04135008	-0.57715343	1.3415083
-----						
	PF	PM	CHI			
1	-0.1489011	-0.004159499	0.57			
2	-0.4230779	-0.552737408	0.30			
3	0.2763333	0.678787651	0.86			
4	1.1725049	0.083990880	0.54			

Figure 12. Standardized averages for each attribute

For example, players in cluster 3 spend much more time compared to other 3 groups. Players in cluster 2 spend least amount of time. Players in cluster 3 end up with much more point while players in cluster 2 made very few points in 2017/2018 season. For OREB, players in cluster 4 had higher values compared to other clusters. While cluster 4 had the highest blocks meaning players in cluster 4 are mostly defensive players.

## 5 CONCLUSIONS

Based on the clustering analysis result, each player who had played at least 2000 mins of basketball were evaluated and analyzed using attributes such as FTA free-throws attempted, FTM

free-throws made, P2A two-point shots attempted and P3M three-point shots made. The study was able to determine which players or team had better offensive or defensive attributes overall. According to the study, participants in cluster 3 spent a lot more time than those in the other 3 clusters. The cluster 2 players spent the least amount of time. While players in cluster 2 scored very little throughout the 2017–2018 season, players in cluster 3 ended up with a lot more points. For offensive rebound (OREB), players in cluster 4 had higher values compared to other clusters. While cluster 4 had the highest blocks meaning players in cluster 4 are mostly defensive players. An overall analysis showed in details which player performed better generally and where to improve. It is recommended that other ML techniques like Network analysis can be used for evaluation and analysis and compared with this study to know which had better evaluation metrics.

## REFERENCES

1. Jaime, S., Tim, M., Julio, C.G., Sergio, J.S., Xavi, S. and Mindaugas, B. (2014). Exploring Game Performance in the National Basketball Association Using Player Tracking Data, *PLOS ONE*, DOI:10.1371/journal.pone.0132894, pp. 1-14.
2. Gibson, J. (2014). The ecological approach to visual perception. *Boston. Houghton Mifflin*, pp. 308-309.
3. Savelsbergh, G, Davids, K., Van der Kamp, J. and Bennett, S.J. (2013). Development of Movement Coordination in Children. *Applications in the Field of Ergonomics, Health Sciences and Sport: Taylor & Francis*, vol. 45, no. 12, pp. 847-848.
4. Kauffman, S.A. (1993). The Origins of Order: Self Organization and Selection in Evolution. *Oxford University Press*, pp. 2698-2699.
5. Vadrucchio, L.S. (2017). Ranking methods for data analytics on players' performance in basketball games. *Università degli studi di Padova Dipartimento di Scienze Statistiche Corso di Laurea Magistrale in Scienze statistiche*, pp. 1-126.
6. Sarlis, V., Chatziilias, V., Tjortjis, C. and Mandalidis, D. (2021). A Data Science approach analysing the Impact of Injuries on Basketball Player and Team Performance. *Information Systems*, vol. 99, pp. 1-16.
7. John, F.D., Amy, K.L., Benjamin, D.H. and Ron, E. (2017). MIT Sloan Sport Analytics Conference from Sports to Science. *Using Basketball Analytics to Broaden the Appeal of Math and Science among Youth MIT Sloan, Sport Analytics Conference*, pp. 1-14.
8. Sarlis V. and Tjortjis, C. (2020). Sports analytics-Evaluation of basketball players and team performance. *Information Systems*, vol. 93, pp. 1-30.
9. Pelechrinis, K. and Papalexakis, E. (2018). Athlytics. Proc. Elev. ACM Int. Conf. Web Search Data Min. – WSDM, vol. 18, pp. 787-788.
10. Zhou, C. and Wang, R. (2016). A new perspective on positions in basketball players using cluster analysis. *Academia.edu.*, vol. 29, pp. 1-8.
11. Robert, S. and Mark, V.M. (2022). Expanding Plus-Minus for Visual and Statistical Analysis of NBA Box-Score Data. *IEEE Technical Community on Visualization and Graphics*, pp. 1-7.
12. Patton, A.N. Scott, M., Walker, N., Ottenwess, A., Power, P., Cherukumudi, A. and Lucey, P. (2020). Predicting NBA Talent from Enormous Amounts of College Basketball Tracking Data, , pp. 1-14.
13. Georgios, G. (2020). Descriptive and Predictive Analysis of Euro league Basketball Games and the Wisdom of Basketball Crowds. arXiv:2002.08465v1, pp. 1-24.
14. Tomislav, H., Ladislav, H., Dunja, S. and Vladimir, M. (2019). Data-driven Basketball Web Application for Support in Making Decisions. *Proceedings of the 7th International Conference on Sport Sciences Research and Technology Support*, pp. 239-244.

15. Rubén, D., Alejandro, V., José Vicente, G.T. and Paula, B. (2015). Heart rate analysis of high-level basketball players during training sessions. *Scientific Information System*, vol. 24, no. 1, pp. 17-19.
16. Collins, C. (2022). Applying Data Analytics as an Alternative to Subjective Rankings of Players in Fantasy Basketball. *Mathematica Militaris*, vol. 25, no. 1, pp. 1-14.
17. Valeria, T., Yaroslav, H., Viktoria, P., Oleksij-Oleg, B. and Zoryana, S. (2018). Factor analysis of indicators of physical and functional preparation of basketball players. *Journal of Physical Education and Sport*, vol. 18, no. 4, pp. 1839-1844.
18. Ruth, D. and Asher, D. (2017). Big Data in Sports - Leverage Big Data in Sports: An Insight using SAP HANA. *International Journal of Engineering Research & Technology*, vol. 6, no. 1, pp. 380-386.
19. Evan, C., Cibele, H. and Jose, K.P. (2015). Player Tracking and Analysis of Basketball Plays, pp. 1-6.
20. Core Team, R. (2018). A Language and Environment for Statistical Computing, 3833p.
21. Wickham, H. (2016). Ggplot2: Elegant Graphics for Data Analysis. *Springer-Verlag; New York*, 186p.
22. Hastie, T., Robert, T., Jerome, F. and James, F. (2005). The elements of statistical learning: data mining, inference and prediction, vol. 27, pp. 1-200.
23. What is Hierarchical Clustering (2022). Retrieved from <https://www.displayr.com/what-is-hierarchical-clustering/>. pp. 1-7.
24. Johnson, R.A. and Dean, W.W. (2014). Pearson Education, Applied multivariate statistical analysis. *Pearson*, pp. 106-393.