

Investigation of FIBA World Cup 2019: Evidence Using Advanced Statistical Analysis and Quantitative Tools [†]

Christos Katris

Department of Mathematics, University of Patras, 26504 Patras, Greece; chriskatris@upatras.gr

[†] Presented at the 9th International Conference on Time Series and Forecasting, Gran Canaria, Spain, 12–14 July 2023.

Abstract: The purpose of this study is the quantitative investigation of the basketball tournament of the FIBA World Cup 2019. Firstly, it identified the performance of a team by using Principal Components Analysis (PCA). Then, the contributions of shooting, rebounding, turnover, and free-throw factors are identified and compared with Offense vs. Defense in terms of their contribution to the team's performance. Moreover, other factors are identified that affected the performance, the teams which performed better than expected are detected, and finally, machine learning models which enhance the 'Power Rankings' for the prediction of the final position of the teams in the tournament are suggested.

Keywords: basketball analytics; team performance; multiple regression; k-means clustering; machine learning models

1. Introduction

It is widely known and accepted that traditional statistics cannot accurately describe some aspects of basketball, and for this reason, an advanced statistics revolution has taken place in research on basketball in order to produce statistics that are more meaningful and useful for the analysis of the game. The advanced statistics for basketball can be found in the works [1,2]. However, these analyses are valid only in a league format, where all teams play with all other teams, and seasons last for a long time. When the situation is a tournament which is a fast-track competition where teams do not face all other teams, these statistics could be misleading. Moreover, the view of this work is "macroscopic", i.e., the aim is to specify factors that can lead to overall (performance-based) success in the tournament and not to winning in a single game. The aim of this paper is to offer a quantitative method of answering questions in a tournament situation, such as the FIBA World Cup, and to be a starting point for analyzing tournaments in other sports. The focus of most previous papers regarding research on basketball has mostly been on league situations and comparisons or factors of discrimination between winning and losing teams. The focus here is on overall tournament performance and not only on single-game winning factors.

Some previously published related works include the work [3] that explored the factors that influenced the performance of the Chinese team in the 15th Men's World Basketball Championship; they found that the team's ability was imbalanced, that a flexible attack strategy was needed in order to increase attacking ability, and that players' mental regulation needed improving greatly. Furthermore, in work [4], the authors for the matches of the Chinese basketball team in the 14th Men's World Basketball Championship analyzed all kinds of causes of the losses and gains in the match and indicated that speediness, agility, precision, and antagonism are the everlasting trends in the world basketball, while in work [5], the authors used regression analysis to examine the influence of certain basketball elements (FIBA standard indicators of performance) on the final result of a basketball game (they considered games from the 13th, 14th, and 15th Men's World Basketball Championships). Additionally, in the paper [6], the authors compared the Chinese team with the



Citation: Katris, C. Investigation of FIBA World Cup 2019: Evidence Using Advanced Statistical Analysis and Quantitative Tools. *Eng. Proc.* **2023**, *39*, 85. <https://doi.org/10.3390/engproc2023039085>

Academic Editors: Ignacio Rojas, Hector Pomares, Luis Javier Herrera, Fernando Rojas and Olga Valenzuela

Published: 14 July 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

six other top teams of the 2006 Men's World Basketball Championship in terms of statistics. They analyzed the gaps and detected the weaknesses of the Chinese team. In work [7], the author determined which basketball performance indicators can discriminate winners from losers using a dataset of 76 matches from the world championships in Spain in 2014, of which the official statistical parameters were downloaded from FIBA. Finally, in work [8], they compared and analyzed differences between the technical styles of the Chinese and American men's basketball teams in the 15th FIBA World Cup.

The explanation of the aims of this paper follows. A crucial term is team performance. The consideration of only the final ranking of the team in the tournament is obviously misleading. Maybe a team is playing very well in all games but has a blackout in a knock-out match, and then the rating is unfair for this team. On the other hand, if performance is the extent of victory, then a strong team might be lucky in the draw of the groups and easily win against their first opponents but, when facing another strong team, not be able to cope with the situation. The previous examples have led to the consideration of performance as a multivariate measure, with the target being to extract a single value for the performance of each team. To achieve this, we used principal component analysis (PCA). The next goal of this work was to determine which factors contributed to the performance of a team. The basis for this analysis is the concept of the four factors of Dean Oliver as a standard for determining the winner of a game. Another big debate is whether offense or defense is more important for success in such tournaments. Both questions are answered with the use of multivariate regression. Additionally, we studied the effects of other factors, such as (i) the height of the team, (ii) the age of the team, (iii) the coach's experience with the team, (iv) the players' percentage (pcg.) usage of the ball (or from the first five players), (v) the distance shooting in a team, (vi) the balance in team scoring, and (vii) the efficiency of small players. Multiple regression and the correlation of variables with team performance were the tools for measuring these effects. Another very popular debate is whether a team performed as well as expected in the tournament. In this manuscript, we make an attempt to determine whether team performance is compatible with pre-tournament expectations, which are specified with the help of hierarchical k-means clustering based on variables that were found to affect the performance of teams. Groups of teams were formed according to their pre-tournament characteristics, and post-tournament actual performance was compared with the expected performance of the teams. A final question that was studied is whether we can have better pre-tournament predictions than power rankings. We employed machine learning models (Random Forests and neural networks) for the prediction of the final positions (based on performance) of the teams in the tournament. Power rankings incorporate information and knowledge from experts that should not be wasted, and this is the reason why they were considered among the inputs in our models. Moreover, they were the benchmark for our models, i.e., we were interested in whether a model could enhance power rankings, and if so, then the model was considered useful. The models were compared in terms of correlation (through the pseudo-R-square measure) with performance-based final positions.

A brief overview of the problems that are studied in this work is the following: the measurement of the performance of a team in a tournament, the detection of which factors played an important role in the performance of a team, whether a team fulfilled expectations in the tournament, and, finally, the suggestion of an improvement in 'Power Rankings'.

The rest of the paper is as follows: Section 2 presents the definitions and meanings of the statistical measures and the statistical methods that were used to tackle the problems in this work. Section 3 is an overview of the questions and problems we considered for the tournament and a detailed description of the procedures we used to deal with them. Section 4 presents the data analysis, and Section 5 contains the summary and the conclusions of the paper.

2. Statistical Definitions, Measures, and Tools

In this section are briefly presented the elements which are used throughout this work. The Principal Component Analysis (PCA) method is a statistical method that was introduced by Pearson and later independently developed and named by Hotelling, and the aim is to express multivariate data with fewer dimensions. A detailed analysis of this method can be found in the book [9].

The correlation of 2 variables can be measured using a coefficient that quantifies this correlation (the value of the coefficient is between -1 and 1 , the magnitude displays the strength of the correlation while the sign displays the direction of the correlation). In this work, we use two such coefficients for completeness: Pearson correlation (r) (details about this method can be found in [10]) and the Spearman rank correlation coefficient (ρ) (details about this method can be found in [11]).

In statistics, linear regression is a linear approach of the form $y = Xb + \varepsilon$, which is used to model the relationship between a (dependent) variable and one or more explanatory (independent) variables. Details about linear regression can be found in many statistics books, such as [12]. It is known that factors that affect the outcome of a game are the shooting factor, turnover factor, rebounding factor, and free throw factor, and they are introduced and described in the works [1,2]. Their formulas are mentioned briefly: The shooting factor (*Sh.F.*) formula for both offense and defense is $(FG + 0.5 \times 3P)/FGA$. The turnover factor (*To.F.*) formula for both offense and defense is $TOV/(FGA + 0.44 \times FTA + TOV)$. The rebounding factor (*Reb.F.*) formula for offense is $ORB/(ORB + Opp\ DRB)$, while the formula for defense is $DRB/(Opp\ ORB + DRB)$. The free throw factor (*FT.F.*) formula for both offense and defense is FT/FGA . Possessions of a team are computed through the formula: $FGA + 0.475 \times FTA - ORB + TO$. The possessions are calculated for both the offensive and defensive teams, and the average is considered to decide a game's overall possessions.

Random Forests (are described in [13], were introduced in [14], and each node is split using the best split among a subset of predictors randomly chosen at that node. The output is the mean of all trees for regression. This strategy performs very well against other classifiers and is robust against overfitting. Neural networks are computing systems that are inspired by biological neural networks that constitute animal brains. An overview of neural networks can be found in reference [15]. K-means clustering ([16]) is a popular method for cluster analysis in data mining. In this work, we use the method of Hierarchical k-means clustering ([17]), and the method is implemented in the R package 'factoextra' ([18]).

3. FIBA World Cup 2019: Problems and Procedures to Solve Them

To be decided the success of a team in tournament competition, they are used some metrics. Because the most important definition is the performance of a team, the Winning percentage is naturally the first used metric. However, in the case of a tournament is not a suitable metric because teams do not face all other teams (only a subset of them after a draw). Another measure of the performance of a team is the point difference (PD) between the team and its opponents (this metric displays the dominance of the team). Another metric of success of a team could be the final ranking of the team in the tournament. This metric is also inappropriate. In order to achieve a complete metric of the success of a team, we consider all the above metrics, and we derive an overall metric of success (team score) with the use of the concept of Principal Component Analysis (which explains a large portion of variance).

Furthermore, in this work, it is specified whether the four factors (*Sh.F.*, *TO.F.*, *Reb.F.*, and *FT.F.*) affect the overall performance of the team. To achieve this, we use a multiple regression model (Model 1) with these factors as independent variables and performance as dependent variable. The factors for each team are calculated based on team statistics per game (were extracted from the site of basketball reference—<https://www.basketball-reference.com/international/fiba-world-cup/2019.html> accessed on 1 March 2023).

Additionally, this work replies to another very interesting question, which is whether offence or defense played the most important role in the performance of a team in the

tournament. To answer this question, are formulated, applied, and compared two multiple regression models (Models 2 and 3).

$$Y = a + b_1(Sh.F_{offence}) + b_2(Sh.F_{defence}) + b_3(To.F_{offence}) + b_4(To.F_{defence}) + b_5(Reb.F_{offence}) + b_6(Reb.F_{defence}) + b_7(FT.F_{offence}) + b_8(FT.F_{defence}) + \varepsilon \quad (\text{Model 1})$$

$$Y = a + b_1(Sh.F_{offence}) + b_2(To.F_{offence}) + b_3(Reb.F_{offence}) + b_4(FT.F_{offence}) + \varepsilon \quad (\text{Model 2})$$

$$Y = a + b_1(Sh.F_{defence}) + b_2(To.F_{defence}) + b_3(Reb.F_{defence}) + b_4(FT.F_{defence}) + \varepsilon \quad (\text{Model 3})$$

Furthermore, many effects are tested for their effect on the performance of a team in the tournament. Firstly, are tested the effects which are related to player usage percentage (usg%). The formula of the concept of usage percentage (usg%) is the following: $usg\% = 100 \times ((FGA + 0.44 \times FTA + TOV) \times (Tm MP/5)) / (MP \times (Tm FGA + 0.44 \times Tm FTA + Tm TOV))$. The usage percentage (usg%) is an estimate of the percentage (pcg.) of the team's offensive attempts (plays), which are used by a player while he is on the floor.

Except for the usg%, we consider the position of the player with the greatest usg% in the team (or the avg. position of the five players with the greatest usg%), the played minutes of the player with the greatest usg% in the team (or the avg. played minutes of the five players with the greatest usg%) and the percentage of plays of the player with the greatest usg% in the team (or the avg. percentage of plays of the five players with the greatest usg%). The effect of the player with the greatest usg% (or of five players with the greatest usg%) is tested with multiple regression models (Models 4 and 5, respectively).

$$Y = a + b_1(usg\% \text{ of first Player}) + b_2(\text{Position of first Player}) + b_3(\text{Minutes of first Player}) + b_4(\% \text{ of Plays of first Player}) + \varepsilon \quad (\text{Model 4})$$

$$Y = a + b_1(\text{Avg. usg\% of first 5 Players}) + b_2(\text{Avg Position of first 5 Players}) + b_3(\text{Avg. Minutes of first 5 Players}) + b_4(\% \text{ of Plays of first 5 players}) \quad (\text{Model 5})$$

Next, there is tested if the players who are competing in a specific league (League Effect) can affect the performance of a team in the competition. The most important leagues (and their weights for building an overall League Effect score) is an ad-hoc decision. There are considered players who play in the NBA, the Euroleague, the Eurocup, the Basketball Champions League (BCL), and the NCAA. In this work, the scores for the leagues are respectively 1, 1, 0.5, 0.5 and 0.5. Other effects which are tested include whether they affect the performance of a team, the heights of players of the team (this is measured by the average height of the players of the team and by the number of players in a team with a height over 200 cm.), the ages of the players of the team (this effect is measured by the average age of the players of the team and by the number of players in the team with age over 30 years old), the coach experience to the bench of the team (in Years), and the importance of shooting (this is measured by the percentage of 3 point attempts over the overall attempts and by the points scored from players who plays in the positions 1, 2, and 3 (small players) versus the points scored from players who plays in the positions 4 and 5 (high players)). These effects are tested with regression models (Models 6–10).

$$Y = a + b_1(\text{League Effects} + \varepsilon) \quad (\text{Model 6})$$

$$Y = a + b_1(\text{Avg. Height}) + b_2(\text{number of players with height over 200 cm}) + \varepsilon \quad (\text{Model 7})$$

$$Y = a + b_1(\text{Avg. Age}) + b_2(\text{Number of players with age over 30 yearsold}) + \varepsilon \quad (\text{Model 8})$$

$$Y = a + b_1(\text{Coach experience in the team (in Years)}) + \varepsilon \quad (\text{Model 9})$$

$$Y = a + b_1(\% \text{ of 3pt Attempts}) + b_2(\text{Pts of small.vs. Pts of high players}) + \varepsilon \quad (\text{Model 10})$$

Moreover, two formulas are defined:

$$(i) \quad \text{Efficiency of Small Players} = \frac{\text{Pts from Small Players (PG,SG,SF)}}{\text{Pts from Tall Players (PF,C)}} - \frac{\% \text{usg. from Small Players}}{\% \text{usg. from Tall Players}}.$$

$$(ii) \quad \text{Balance} = \frac{\text{Pts from Small Players (PG,SG,SF)}}{\text{Pts from Tall Players (PF,C)}} + \frac{\% \text{usg. from Small Players}}{\% \text{usg. from Tall Players}} - 2.$$

Additionally, the intention is to check whether the team pace (tempo) affects the performance, i.e., faster or slower teams found to perform well (this is measured by the number of possessions of a team per game of the competition). These additional effects are tested with regression models (Models 11–13) and with Spearman and Pearson correlations.

$$Y = a + b_1(\text{Efficiency of Small Players}) + \varepsilon \quad (\text{Model 11})$$

$$Y = a + b_1(\text{Balance}) + \varepsilon \quad (\text{Model 12})$$

$$Y = a + b_1(\text{Possesions per Game}) + \varepsilon \quad (\text{Model 13})$$

4. Data Analysis

This section contains the data analysis of the tournament and the conclusions from this analysis. The fundamental concept here is the calculation of the total score for each team which represents its performance. The tables with the data are in the link https://docs.google.com/document/d/1QMERMfeckZNY9LZT1BCHdbI7UK6tctNc/edit?usp=share_link&ouid=118393132040933122489&rtpof=true&sd=true, (accessed on 14 July 2023). The conclusions from the analyses which were implemented are presented here. All methods in Sections 4.1–4.3 were implemented in the statistical software R (Version 3.6.0).

4.1. Calculation of Team Performance and Descriptive Analysis

At first, we calculate the team scores using PCA. Table S1 displays both the score of each team and the ranking of the team according to this score. We consider the success of each team, the percentage of wins in its games, the point difference on average in its games, and the final position of the team in the ranking of the FIBA World Cup 2019 (we consider the value 4/Final Position). The goal is to achieve an overall score for each team. This score reflects the overall performance and takes into account all three aforementioned dimensions of the performance. The first component, after the application of the PCA method, explains almost 80% of the variance of the three initial variables, so we consider it as a measure of performance. The variables were rescaled before the analysis in order to have unit variance.

4.2. Analysis of the Four Factors in the Performance of a Team

In this subsection, at first, we calculate the offensive and the defensive factors, and then we run a multiple regression with these factors as independent variables and Team Score as the dependent variable. The regression shows that these factors explain approximately 80% of team performance (R^2 values). The signs of the factors are expected for all factors. According to the p-value, we found the following: the shooting factor (offensive and defensive) affected the score of each team, the free-throw factor (in defense) and the turnover factor (in defense) were found to affect the score of each team at the 5% level. The

other factors were found to be statistically insignificant. Moreover, we attempt to compare offense and defense. The procedure is following:

- *Offense vs. Defense:* We consider only the offensive factors and their contribution to the team's success, then only the defensive factors and their contribution to the team's success, and then we compare the contribution of the factors to the explanation of success through multiple regression models. From the above models, we conclude that the defensive functioning was found to be more influential than the offensive functioning according to R^2 values (while both were found to be statistically significant).

4.3. Exploration of Effects

In this subsection, many effects were tested through regression analysis and correlation measurement using the team performance (team score) as the dependent variable.

4.3.1. Effect of the Player (of Five Players) with the Greatest Usage

There are considered the five players (who play more than 10 min.) with the greatest usage in the game. They are taken into account and tested the next variables for these players: the average usage, the average position (i.e., PG = 1, SG = 2, SF = 3, PF = 4, C = 5), the average minutes, and the percentage of plays. There are considered the five players (who play more than 10 min.) with the greatest usage in the game. They are taken into account and tested the next variables for these players: the average usage, the average position (i.e., PG = 1, SG = 2, SF = 3, PF = 4, C = 5), the average minutes, and the percentage of plays. Table S7 displays the correlation results, and Table S8 displays the results of the regression of each factor with team performance (team score). There is no significant correlation between any of the factors with team performance, and the regression is not significant, so there is no significant effect of the usage, position, or minutes on the team score. The same applies to the effect of the leader on the team's performance. Tables S9 and S10 display the results.

4.3.2. League Effects

The goal of this subsection is to study the effect of the players who compete in more competitive leagues on the performance of the team in the tournament. We consider the number of players who compete in the NBA, the Euroleague, the Eurocup, and on BCL, and the NCAA. Table S11 displays the correlation, and Table S12 the regression of these factors with the team score. There was found to be a significant correlation between team performance and the number of NBA players and between team performance and the number of Euroleague players in a team. The regression was found to be significant, too (F-value). Specifically, the number of NBA players and the number of Euroleague players affect the performance significantly, while the number of Eurocup players, the number of BCL players, and the number of NCAA players do not seem to affect the team's performance. Furthermore, we derive a value for each team using the following formula: *Top-League Effect* = No. of NBA players + No. of Euroleague players + $0.5 \times$ No. Eurocup players + $0.5 \times$ No. of BCL players + $0.5 \times$ No. of NCAA players. The Regression is significant, and this effect explains over 45% of team performance (Table S13).

4.3.3. Height Impact, Age Impact, and Coach Experience in the Team

The first goal of this subsection is the study of the effect of the height of available players of a team. A common question is if increased height leads to increased chances of winning. We consider as variables: (i) the average (avg.) height of players and (ii) the number (no.) of players with height over 200 cm. Table S14 displays the correlation between the variables, and Tables S15 and S16 display the results of the regression of each factor with team performance (team score). It was found significant correlation and regression between average height of the team and team performance.

Furthermore, it studied the effect of the age of available players of a team. Another common belief is that increased age leads to decreased performance. This is maybe rational

for leagues, but is this true for a tournament? We consider (i) the average (avg.) age of the players of the team and (ii) the number (no.) of the players of the team with ages over 30 years old. There were not found significant correlations and regressions between the considered variables and the performance of the team (see Tables S17–S19).

Finally, it studied the effect of coach experience on the performance of a team. A common belief is that the long-term incorporation of a coach into a team leads to increased chemistry, thus, performance. We consider, as a variable, the years (yrs.) of the coach in the team. There is a significant correlation and regression between “Coach experience” and “team performance” (see Tables S20 and S21).

4.3.4. Existence of Shooters and Other Effects

In this subsection, the first goal is to study the effect of the existence of shooters inside the roster of a team on its performance. A common belief is that the existence of many shooters inside the roster of a team leads to an increased offensive threat, thus to an improved performance of the team. In order to test this fact, we consider as variables: (i) the percentage (%) of 3 pt attempts and (ii) the points scored from the players who play in the positions of PG, SG, and SF (‘small’ players) divided with the points scored from the players who play in the positions of PF and C (‘high’ players).

Furthermore, for the effects of (i) the team balance, (ii) the efficiency of small players’ and (iii) the pace of the team in its performance, it was found that there were no significant correlations and regressions.

4.4. Detect Surprises and Upsets—Clustering (with Hierarchical k-Means)

The goal of this subsection is to suggest a procedure to detect the positive and negative surprises of the tournament. At the core of this procedure is the generation of clusters of teams. From the previous analysis (in Section 4.3) there are considering the variables which were found to significantly affect the performance of a team (these are the Coach’s Experience in the team, the Average Height of the team, and the number of players who compete within Top Leagues) and the Power-Rankings before the tournament. Figure S3 presents the elbow method (a graph of the total within the sum of squares (WSS) which is explained by the increase of clusters). The decision about the number of clusters is a number for which the addition of an additional cluster is not improving much the total WSS, which is explained. In the case of the FIBA World Cup 2019 tournament, the generation of 3 clusters of teams is the decision. The first cluster represents the strongest teams (S), the second cluster represents the 2nd tier teams (M), and the third cluster represents the weakest teams (W) of the tournament. From Table S28 and Figure S4, are observed some notable facts: (i) stronger teams display higher offensive and defensive efficiencies, and (ii) we observe that Higher Pace is not a characteristic of stronger teams necessarily and rather displays the style of play of each team (see Figure S2). However, the weakest teams play, on average, on a lower tempo. Finally, we consider ranking according to clustering as expected from the team characteristics beforehand; this ranking is compared with the actual ranking of the tournament, and the procedure for the detection of surprises is implemented (see Table S29 for surprise detection).

4.5. Improve Power Rankings—Predict Team Position

The goal of this subsection is the improvement of the virtual ranking of the teams before the tournament. The accuracy of such ranking is very important for betting reasons and for coaches and fans who can adjust their expectations about the performance of the team they support. The main tool for such a ranking is the so-called ‘Power Rankings’, which are released before the tournament and take into account all the relevant information. The improvement of ‘Power Rankings’ is the aim of this subsection. The main idea is the use of ‘Power Rankings’ as input to other models. To achieve such an improvement, we use the machine learning methods of Random Forests and Feed-Forward Neural Networks (ANN). The analysis is performed with the use of the package Rattle [19] of the R statistical

software. The parameters of the models are for Random Forests 500 trees and two variables and for ANN 1 hidden layer with 10 nodes. Input variables are: (i) the ‘Power Rankings’ (which are released officially by FIBA) and (ii) the significant variables from Section 4.3 (the Coach Experience in the Team, the Average Height of the team, and the number of players of the team who compete in Top-Leagues), whilst the output of the models is the team ranking. The crucial part is the evaluation of the models. At first, we build the models, i.e., we consider the entire dataset in order to train the models, and the training performance of them is measured using as metric the pseudo-R-square. Secondly is evaluated the predictive accuracy of the models with the following procedure: there are considered 10 teams randomly, and then their ranking is predicted based on the model. The experiment is repeated 10,000 times. The evaluation is performed using the metric of pseudo-R-square (the mean, the standard deviation and a 90% confidence interval are calculated). We observe that the training accuracy of the models is greater than that of Power Rankings, and Random Forest is the preferred approach (see Table S31).

5. Summary and Conclusions

In this work was made a quantitative analysis of the FIBA World Cup 2019. Firstly, it was determined as a metric of the performance of a team in a tournament. Furthermore, was studied the importance of the four factors in the performance of the team and was answered the question of offense vs. defense. Moreover, the coach’s experience in the team, the average height of the team, and the number of players who compete in top leagues are found to affect the performance of a team. Next, a procedure was presented, which was based on clustering in order to detect ‘surprises’ in the tournament. Finally, a procedure with the aim of improving Power Rankings through machine learning methods was suggested. This work can serve as a source of thought for tournament analysis in basketball and other sports.

The Data Analysis of this paper can be found in the following link: https://docs.google.com/document/d/1QMERMfeckZNy9LZT1BCHdbI7UK6tctNc/edit?usp=share_link&ouid=118393132040933122489&rtpof=true&sd=true, (accessed on 14 July 2023).

Supplementary Materials: The data used for this paper and all the relevant analysis (tables and figures) are provided with the paper in a link as supplementary material of the paper (https://docs.google.com/document/d/1QMERMfeckZNy9LZT1BCHdbI7UK6tctNc/edit?usp=share_link&ouid=118393132040933122489&rtpof=true&sd=true).

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used for this paper and all the relevant analysis (tables and figures) are provided with the paper in a link as an additional file as Supplementary Material of the paper.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Kubatko, J.; Oliver, D.; Pelton, K.; Rosenbaum, D.T. A starting point for analyzing basketball statistics. *J. Quant. Anal. Sport.* **2007**, *3*, 1. [CrossRef]
2. Oliver, D. *Basketball on Paper: Rules and Tools for Performance Analysis*; Potomac Books, Inc.: Sterling, VA, USA, 2004.
3. Wu, M.; Yan, J. Statistic Analysis on Chinese Men’s Basketball Team in the 15th World Basketball Championship. *J. Chengdu Sport Univ.* **2007**, *3*, 63–66.
4. Zhang, H.; Pan, S. China the Development Strategies for China’s Man Basketball Studied with the Case of the 14th World Man’s Basketball Championship. *Sport. Sci.* **2002**, *6*, 42–44.
5. Simović, S.; Komić, J. Analysis of influence of certain elements of basketball game on final result based on differetiant at the XIII, XIV and XV World Championship. *Acta Kinesiol.* **2008**, *2*, 57–65.
6. Wang, D.; Zhou, Y. Analysis of the Gaps in Main Techniques between Chinese and European Teams in 2006 FIBA World Championship. *J. Mianyang Norm. Univ.* **2008**, *5*, 125–129.

7. Čaušević, D. Game-related statistics that discriminate winning and losing teams from the world championships in Spain in 2014. *Homo Sport*. **2015**, *17*, 16–19.
8. Zhang, Z.D.; Sun, Z.J. Analysis of different technical styles between China and American Men's Basketball in the 15th FIBA World Championship. *J. Nanyang Norm. Univ.* **2006**, *12*, 79–82.
9. Jolliffe, I. *Principal Component Analysis*; Springer: Berlin Heidelberg, Germany, 2011; pp. 1094–1096.
10. Sharma, A.K. *Text Book of Correlations and Regression*; Discovery Publishing House: Delhi, India, 2005.
11. Myers, L.; Sirois, M.J. Spearman correlation coefficients, differences between. In *Encyclopedia of Statistical Sciences*; John Wiley & Sons: Hoboken, NJ, USA, 2004; Volume 12.
12. Seber, G.A.; Lee, A.J. *Linear Regression Analysis (Vol. 329)*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
13. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
14. Breiman, L. Random forests. *Mach. Learning*. **2001**, *45*, 5–32. [[CrossRef](#)]
15. Macukow, B. Neural networks—state of art, brief history, basic models and architecture. In Proceedings of the IFIP International Conference on Computer Information Systems and Industrial Management, Vilnius, Lithuania, 14–16 September 2016; Springer: Cham, Switzerland; pp. 3–14.
16. Celebi, M.E. (Ed.) *Partitional Clustering Algorithms*; Springer: Berlin/Heidelberg, Germany, 2014.
17. Xu, T.S.; Chiang, H.D.; Liu, G.Y.; Tan, C.W. Hierarchical K-means method for clustering large-scale advanced metering infrastructure data. *IEEE Trans. Power Deliv.* **2015**, *32*, 609–616. [[CrossRef](#)]
18. Kassambara, A.; Mundt, F. Package 'factoextra'. Extract and visualize the results of multivariate data analyses. Available online: <https://www.rdocumentation.org/packages/factoextra/versions/1.0.7> (accessed on 1 March 2023).
19. Williams, G.J. Rattle: A data mining GUI for R. *R J.* **2009**, *1*, 45–55. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.