Abstract

Basketball is a very popular sport that has been developing itself constantly year after year with changes in rules, equipments and technology that directly impact game dynamics. Data Science both in general and focused in sports analytics, in turn, has been directly utilized to help decision making and performance enhancing in athletes, teams and organizations.

This study aims to analyze the National Basketball League's Regular Seasons from 2013-14 to 2022-23, through time series analysis and forecasting. Firstly, by creating a latent variable correspondent to the players' performance, via machine learning algorithms such as Principal Component Analysis. Which, consequently, allowed us to better analyze and understand the data, and made possible to group the PCA's results, creating a hierarchical time series structure encompassing data from each player, team, conference and season.

By combining an exploratory and descriptive analysis of data with time series forecasting models, the obtained results provide a complete overview of the historical characteristics of the NBA's Regular Seasons between 2013 and 2023. In this study, we were able to analyze teams's historical successes and failures, the impact of a mid or preseason trade, the league's players usage efficiency, specially at the COVID-19 period, and the impact of highly forbidden practices by the NBA, such as tanking, in teams' decision making during and between regular seasons.

In addition, with forecasting models, it was possible to obtain point predictions based on past regular seasons, or early regular season performances, to project expected results for the trade deadline or the end of the season, estimating the eight best franchises in each conference, that aim for a playoff appearance, and, on the other side, teams that are projected to be ranked as the worst of each conference, looking for good odds in the draft lottery.

Overall, this study contributes to data science and sports analytics' advances, especially in the NBA and other basketball leagues, providing a deeper comprehension of the historical performances of franchises, as well a forecasting tool that has the main goal to assist strategic decision making at the level of teams, divisions, franchises or the whole association. Consequently, the present study's results can be useful not only for a major sport league like the NBA, but for other organizations sportive or not, that have the main goal of best enable its managers and decision makers aiming in maximizing their teams' potentials and results in the area of action.

Introduction

Sports, in all their modalities, forms and purposes, arouse the interest of a large part of the population. They encompass not only leisure, but also several other areas such as teaching and education, health and prevention, economy and business, and, consequently, the formation and development of citizens, who learn through sports ethical and moral values that guide them on their journey through life chave35, chave36.

The interest in improving athletes' performance is something inherent to sports, where the clear goal is to win in the best possible way. Sports analysis with emphasis in increasing the performance of athletes, teams, or organizations can be done in several ways and multiple areas. However, the presence of statistics and data science in sports has been growing exponentially in the information era. Increasingly, decision-making is assisted, and often guided, by statistical analyzes of the massive amount of data that is constantly produced in the various sports leagues at all levels.

Along with the way we watch, appreciate, study, and analyze sports, the way we practice it also evolves over time. The change or addition of rules, equipment, and technologies that make the dynamics of a sport more complete and fascinating, can be cited as some of the ways in which the practice of a sport has changed over the years. And in basketball, consequently, it would not be different.

Although the essence of basketball, created by James Naismith in 1891, remains the same - to score points by shooting a ball through a hoop - the way the sport is played, watched, taught, learned, and analyzed has undergone significant transformations over time chave5. Each new rule implemented changes the understanding and gameplay of the sport considerably such as the introduction of the three-point line chave11, the creation of the shot clock chave12, and the prohibition of goaltending, among others, have defined different eras of basketball, each with its own unique characteristics and features chave1, chave13. These changes directly influence the strategies, roles, and responsibilities that players assume on the court chave4.

The North American basketball league, the National Basketball League (NBA), is a notable example

of the constant evolution of the sport. Since its founding in 1947, the NBA has implemented various rule changes that have shaped the different eras of basketball chave6. From the banning of zone defense in 1947 aimed to increase the league's popularity by allowing more dominant players to showcase their offensive skills more freely chave25, chave26, chave9. Navegating through changes in court dimensions were made to prevent the excessive dominance of classic centers. And, more recently, changes to defensive rules, such as the prohibition of unnecessary contact between the defender and the ball-handler, and mainly, aiming to prevent offensive players from taking advantage of these changes, demonstrates how the NBA constantly seeks to improve the dynamics and competitiveness of the sport chave37.

Simultaneously, several changes in the way the NBA is analyzed have also occurred. With the emergence of new methods, robust statistical models, and new technologies, the amount of data produced is overwhelming chave7. Currently, it is possible to find and study virtually all information about any game, from the positioning of players on the court, their movements and actions chave23, to the moments when they are not even on the court. And particularly, this is what makes studying the National Basketball League so interesting.

Another point of great relevance is the exponential growth of the sports industry, which brings with it the negotiation of gigantic contracts that put even more pressure on the importance of winning and being relevant chave8, chave24, in any modality. And this fact has directed organizations to an approach based on sports analysis to better make decisions regarding the planning, coordination, performance and allocation of athletes.

Within data science and statistical applications, we can highlight five relevant varieties from the range of possibilities of sports analysis: physical analysis, technical analysis, psychological analysis, tactical analysis and historical analysis. In the physical context, previous studies have investigated the absence of impact of athlete height on the ability to win games in the NBA chave3 and the difference between basketball positions, their on-court functions and contributions to victories chave2, chave10.

From a tactical and technical point of view, four factors have been generally determined as ideal to control in order to win a game: shooting efficiency, offensive rebounds, turnovers, and free throws chave14, chave15. Other studies have also highlighted the importance of controlling turnovers and defensive rebounds chave16, and steals, which help create offensive opportunities through defensive actions chave17.

The differences between winning and losing teams were explored by the studies of chave18, chave45, and chave20. Tactical and technical factors in high-stakes games, such as the NBA Finals series, were studied by works such as chave32. Factors associated with winning over an entire season, by controlling the team's defensive skills and teamwork through the creation of assists, were explored by chave22.

In the area of sports psychology, works such as chave27 and chave28 draw attention to the mental preparation of athletes, as an addition to the traditional and exclusive physical training and performance preparation, with the aim of improving the way high-performance athletes deal with the pressure of training demands, travel, injuries chave29, and high-level competitions.

Finally, in historical analyses, chave34 explores the history of the league's best players, suggesting an ordering based on statistical models. Studies such as chave1 and chave13 detail the NBA Eras and the respective determining factors for success in the regular seasons of each era. From an aggregated historical and tactical point of view, chave30 and chave31 show that the classic basketball positions and their definitions no longer represent the complexity and modernity of the sport as it is played today, evidencing, once more, the evolution of the sport.

However, to the best of our knowledge, only one previous study has aimed to historically study the NBA through time series. The work of chave38 studied the results (win or loss) of all teams in twenty-two NBA seasons, looking for memory in the time series analysis of the franchise results, focusing on helping bettors make better decisions about where to allocate their resources.

This difficulty in analyzing the NBA through time series may be due to the large number of variables that can be studied for each player, franchise, conference, or even season.

chave31 explored this abundance of variables by studying the versatility of modern athletes through multivariate analysis. By creating a latent variable to measure overall player performance and ranking, encompassing the main offensive and defensive statistics of players, regardless of their position labels, they made it possible to evaluate individual regular seasons.

The historical study of a sport (in this case, basketball, and in particular the NBA) helps us to better

understand the dynamics of the game. Investigating trends, events, and historical events provides a better understanding of how to analyze, study, practice, and appreciate it.

This study aimed to analyze the regular seasons from 2013-14 to 2022-23, specifically chosen because compose the Modern Era of the NBA, as defined in chave1. Multivariate time series analysis and machine learning to define a latent variable for player performance, allowed us to replicate the process in chave31 and create historical series for the ranking (and consequently performance) of franchises, divisions, conferences, and the league in general.

By aggregating team's time series, we obtain division's time series, which, in turn, can be aggregated to obtain conference's time series, and finally the overall aggregate, which represents the time series of the league's performance in a season. Consequently, we have a hierarchical time series structure, which allows us to explore the relationship between levels and observe the data in its isolated or aggregated forms.

The hierarchical structure of the series also provides us with different approaches for analysis and forecasting, observing and forecasting them from the lowest to the highest levels, and vice versa, taking advantage of the advantages that each level can provide for a better analysis.

In particular, this analysis aims to descriptively analyze the performance of teams, divisions, conferences, and the league as a whole over time, providing a comprehensive understanding of the dynamics of the NBA, through a latent variable. In addition, we seek to investigate the impact of player trades during the season, the influence of talent recruitment between regular seasons and team trading windows, and to conduct a historical study of the quality and efficiency of the league.

Through the creation of time series of rankings, it was possible to develop time series analysis and forecasting models that offer a valuable tool to assist decision-makers, such as general managers, coaches, and athletic trainers. These models allow for a more accurate intervention on teams during the regular season and between seasons.

Using time series analysis and forecasting models, we were able to determine which teams were projected to qualify for the postseason not only at the beginning of the regular season, but also at crucial moments, such as the trade deadline. This ability to provide forecasts at strategic moments allows for more informed decision-making by sports managers.

By combining a descriptive analysis of past performance with predictive models, this study offers a comprehensive approach to understanding and predicting team success in the NBA. These findings have the potential to positively impact the field of sports analytics, allowing decision-makers to have a clearer and more analytical view of the prospects of their teams, athletes, and goals. Additionally, they can implement more effective strategies to achieve success in the long or short term.

Therefore, we believe that this work contributes to the advancement of sports analytics, an area of growing importance in statistics, and in data science in general by providing valuable information on team performance over time, as well as predictive tools that can improve strategic decision-making. The results of this present work can be useful not only for the NBA, but also for other sports leagues, or even other organizations in other areas, such as retail and wholesale, empowering managers, directors and organizations to maximize the potential of their teams and achieve better results.

referencias