NBA's Triple-Play Evolution: Data-Driven Insights

Unveiling the evolution of basketball through statistical metamorphosis, this project dissects the three-point revolution in the NBA. Explore data-driven insights, from historical shifts to the impact on iconic players, reshaping the game's dynamics.

Python programming project report

by

Giulia Beccaria

Alessandro Degani

Roberta Di Santo

Lorenzo Lobosco

JUNIA, Graduate School of Science and Engineering

Junia - ISEN

February 2024

Contents

Abstract **1**

1 Introduction **2**
1.1 Aim of the project 2

2 Web Scraping **3**
2.1 Teams Table 3
2.2 Players Tables 3

3 Pre-processing **5**
3.1 Team table Pre-processing 5
3.2 Players table Pre-processing 6

4 Analysis **7**
4.1 3 points impact 7
4.2 3 points time series 8
4.3 Differents position analysis 10
4.4 LeBron James and Stephen Curry evolution 12

5 Predictive model **17**

6 Conclusions **20**
6.1 Results 20
6.2 Further developments 20

Bibliography **22**

Abstract

This project aims to analyze the evolution of basketball gameplay, focusing on the significant increase in three-point shooting using official NBA statistics. The analysis explores shifts in player roles, team dynamics, and the influential role of key players such as LeBron James and Stephen Curry. The findings highlight a widespread preference for three-point shots and emphasize Curry's impact on the strategic approach of the Golden State Warriors.

The project also involves developing a predictive model for the NBA 2023-2024 season, with a specific emphasis on three-point statistics. The Linear Regression model is identified as the most effective, predicting the Dallas Mavericks as potential frontrunners. However, it is crucial to note that these initial predictions require continuous monitoring and refinement.

In conclusion, the project provides insights into the dynamic changes in basketball, emphasizing the growing significance of three-point shooting. Future developments may include more nuanced player-specific analyses and real-time updates to enhance the understanding of evolving trends.

1 | Introduction

In the world of basketball, as in many other sports, change is the only constant. If we consider the game played as recently as 10 or 20 years ago, it is evident how much basketball has undergone a remarkable metamorphosis. In a not-so-distant era, the idea of focusing on three-point shooting was viewed with skepticism, even disdained by many. It was felt that anyone who dared to shoot from behind the three-point arc would be doing their team a disservice. And it was not just perception, but also statistics: teams that relied on three-point shooting seemed to be assigning their hopes to a failing strategy.
However, the evolution of basketball has shown that the beliefs of the past are not immutable. Today, almost all basketball teams not only take advantage of the three-point shot, they cannot do without it. It has become a fundamental pillar of the style of play, and success on the court is largely determined by the accuracy of three-point shooters. In other words, basketball is not simply changing; it has already been significantly transformed. This change is an exemplary demonstration of how innovation and statistics have radically affected a game that has a centuries-old history.

1.1 Aim of the project

The goal of this project is to highlight in a convincing and data-driven way the significant evolution in the style of basketball play over the past decades. To achieve this, reliance will be placed on the official statistics site of the NBA [1], the National Basketball Association. It will be critical to quantify the degree and magnitude of this change by identifying in what aspects of the game, such as players, teams, and roles, this evolution has manifested itself. This will be done through the use of descriptive statistics and data visualizations.
Next, we will focus on two emblematic basketball players of the past few decades: Stephen Curry [4], known as the pioneer of a strategy focused on three-point shooting and widely recognized as the best three-point shooter in basketball history, and LeBron James [5], an extraordinary athlete often considered the strongest and longest-tenured player in NBA history. This will allow us to analyze how the change in playing style has affected their individual performances and their approach to basketball over time.
In pursuit of a comprehensive understanding of a team's performance in the upcoming 2023-2024 regular season, we have endeavored to construct a predictive machine learning model. Our methodology involved training and testing models utilizing statistics acquired through web scraping. Subsequently, we employed the statistics from the initial games of this season as input for our model, aiming to forecast the team's win percentage at the conclusion of the season.
This predictive model goes beyond conventional analyses by offering insights into the nuanced impacts of variables such as three-point shooting on a team's overall performance. By exploring these facets, we aim to provide a deeper perspective on a team's chances of success in the dynamic landscape of the modern NBA.
In summary, this project aims to comprehensively and thoroughly explore the evolution of basketball as a game, demonstrating how the three-point shot has become crucial, analyzing the influence on key players, and finally experimenting with a predictive model to understand how these dynamics have a tangible effect on teams' performance in the regular season.

2 | Web Scraping

The initial phase of our project naturally involved data extraction and acquisition. To accomplish this, we conducted web scraping from the official NBA website [1]. Specifically, we utilized the Python programming language, particularly the Selenium library for automated web browsing, and Beautiful Soup for extracting data tables once we reached the desired web pages. For a more comprehensive understanding of the process, the code and detailed commands used can be found in the attached .py file.

From the NBA website, we scraped two main tables. The first table contained general team statistics for each available year and for each player position. The second table focused on player statistics for Stephen Curry and LeBron James. During the extraction phase, we downloaded all available variables, and a comprehensive feature selection process was subsequently conducted during the pre-processing phase to refine the dataset and focus on the most relevant variables.

2.1 Teams Table

In the context of scraping data for the Teams table, our approach involved navigating the website using the site's HTML code to access the traditional statistics for each team. Subsequently, we extracted the data table for each team's performance in a given year. We then proceeded to iterate through the data extraction process for each player role within the selected year, ensuring that we only collected information for players occupying that specific role on each team.

Furthermore, we repeated this extraction process for every available year, spanning from the 2023-2024 season all the way back to the 1996-1997 season. This comprehensive approach allowed us to capture all the available data within the target table, resulting in a dataset with 28 columns. We also appended columns indicating the year and player role for contextual information.

Ultimately, the output of this data scraping endeavor was a dataset comprising 3327 observations and 30 columns, including 'Year,' 'Position,' 'id,' 'TEAM,' 'GP,' 'W,' 'L,' 'WIN%,' 'MIN,' 'PTS,' 'FGM,' 'FGA,' 'FG%,' '3PM,' '3PA,' '3P%,' 'FTM,' 'FTA,' 'FT%,' 'OREB,' 'DREB,' 'REB,' 'AST,' 'TOV,' 'STL,' 'BLK,' 'BLKA,' 'PF,' 'PFD,' and '+/-' which were subsequently saved in a CSV format for further analysis.

2.2 Players Tables

In terms of scraping the tables for players, we chose to analyse the data of two specific players: LeBron James and Stephen Curry. By examining the HTML code, we were able to reach LeBron James' statistics table and, using the BeautifulSoup library, we extracted the table we were interested in and exported it to a dataframe, which was then transformed into CSV. Since there were 5 tables of interest to us ("Traditional Splits", "Advanced Splits", "Misc Splits", "Scoring Splits", "Usage Splits"), we repeated this process, selecting a different table each time. We did the same thing with Stephen Curry's statistics.

We then took all 5 tables of LeBron James and merged them into a single dataset of 87 columns; we performed the same procedure for Curry's data, again obtaining a dataset of 87 columns.

The tables we have analysed are divided into five categories:

- Traditional Splits: these statistics include basic data such as points, rebounds, assists, stolen balls and stoppages.
- Advanced Splits: these statistics provide a more in-depth view of a player's performance, taking into account factors such as shooting rate, shooting efficiency and scoring.
- Misc Splits: these statistics include data on specific aspects of the game, such as three-point shots, free throws and defensive shots.
- Scoring Splits: These statistics provide detailed information on a player's shooting, such as two-point shooting percentage, three-point shooting percentage and free throw percentage.
- Usage Splits: these statistics measure a player's ball possession percentage.

3 | Pre-processing

During the preprocessing phase, we utilized the raw tables obtained from the web scraping stage and meticulously cleaned them to derive refined datasets. This meticulous cleansing process was undertaken to ensure the acquisition of pristine data, thereby facilitating a more streamlined and immediate analysis.

3.1 Team table Pre-processing

As an initial step, we scrutinized the first five rows of the dataset to gain an initial overview. During this analysis, we noted the presence of missing values in the "Position" column. Upon careful evaluation, it became apparent that all the missing values corresponded to unspecified positions. Consequently, we employed a method for handling missing values, assigning them the value 'All' based on an analysis involving the unique values of the column and the management of missing values.
For the second step, a decision was made to exclude data pertaining to the current season (2023-2024), preserving it for subsequent use in the prediction phase. This choice is underpinned by the fact that, given the early stage of the season, such data might be considered outliers or have a distorted impact on analyses.
Given the dataset's complexity with 30 variables, it was imperative to perform feature selection, opting for a meaningful subset. In this context, the criterion employed was based on correlation with the 'WIN%' variable. To visually depict this correlation comprehensively, we utilized the heatmap function from the seaborn library, enabling a graphical representation of the intricate interplay among variables.

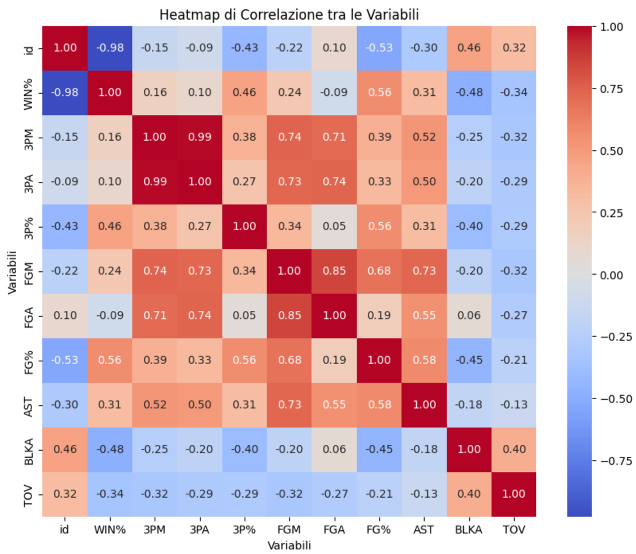


Figure 3.1: The heatmap illustrates the correlation among all variables in the dataset

In our pursuit of a more streamlined and insightful analysis, we made deliberate decisions during the preprocessing phase. By excluding variables like W, WIN%, plus/minus, L, and id that exhibited a conspicuous correlation with the number of wins, we aimed to eliminate potential sources of

redundancy and noise in our subsequent analyses. This strategic reduction not only simplifies the dataset but also focuses our attention on the variables that bear the most significant relevance to our analytical objectives.
The emphasis on variables such as FG and 3P, driven by their high correlations, was not merely an exploratory exercise but a conscious effort to delve deeper into key performance indicators. By closely examining associated metrics like FGA, FGM, 3PA, and 3PM, we sought to unravel nuanced patterns and nuances in the dataset that might contribute valuable insights to our overarching analysis.

The transformation of the 'id' column into a composite key, comprising the team's name, the regular season year, and the position, was motivated by a desire for clarity and specificity in our dataset. This composite identifier not only facilitates unique data retrieval but also aligns with best practices for database management, paving the way for a more intuitive and efficient analysis.
Furthermore, the adaptation of variable types, such as "teams," "position," and "anno," without invoking explicit functions, was a meticulous choice aimed at maintaining the integrity and interpretability of our data. The conversion to categorical data types and the nuanced handling of the "anno" variable underscore our commitment to ensuring that the dataset is not only amenable to statistical analyses but also comprehensible and coherent in its representation of team performance over the specified years and roles. These deliberate steps lay the foundation for a more effective and nuanced analysis, enabling us to derive meaningful insights from the dataset at hand.

3.2 Players table Pre-processing

In the initial stages of preprocessing the player tables, we embarked on the task by importing various scraped dataframes, each corresponding to a distinct table on the player's comprehensive stats page. This step was crucial in laying the foundation for a thorough analysis of the player's performance over time. Early in the process, it became evident that the initial two rows and a specific column required removal to streamline the data for further examination.

Following this data cleansing, a meticulous examination for duplicates or missing data was conducted to ensure the integrity of the dataset. This step was essential to guarantee the accuracy of subsequent analyses. Once the quality of the individual tables was confirmed, the next step involved merging them into a cohesive dataset. This consolidation was a pivotal move, providing a consolidated view of the player's statistics across different aspects of their game.

With the merged dataset at hand, we turned our attention to enhancing its usability for future analytical tasks. Specifically, we focused on the 'By Year' column, recognizing its significance in tracking the player's performance across seasons. An adjustment was made to this column to optimize its utility for graph creation, involving the conversion of the date format from, for example, 2022-23 to 2023. This transformation facilitated a more seamless integration of the data into graphical representations, aiding in visualizing trends and patterns over time.

Moreover, recognizing the ongoing nature of the data updates, especially for the current season (2023-24), we implemented a strategic decision to exclude the corresponding row from our dataset. This exclusion was deemed necessary as the ongoing season data is inherently incomplete and subject to weekly updates. By doing so, we ensured the dataset's accuracy and relevance, laying the groundwork for meaningful and insightful analyses.

Expanding beyond data cleansing and formatting, our preprocessing efforts extended to feature selection. This involved a thoughtful consideration of features that would be most pertinent to analyzing the selected player's game evolution over the years. The objective was to choose features that could provide valuable insights into the player's strategy, playing style, and overall performance trajectory.

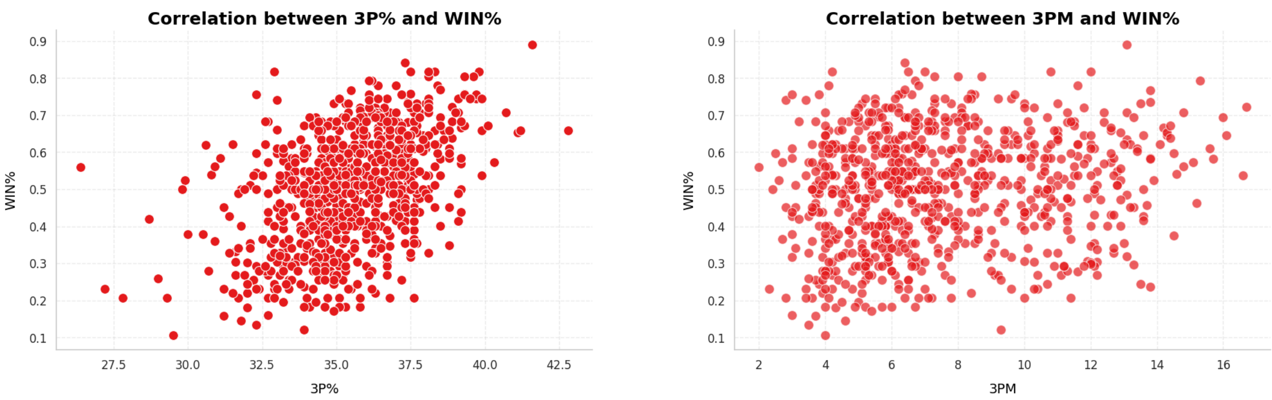
4 | Analysis

We have directed our attention towards the three-point shooting aspect, and our analysis will be segmented into the following components:

In the first part (4.1), we will evaluate the impact of three-point shooting. Moving on to the second part (4.2), we will conduct a time series analysis of three-point shooting trends. The third part (4.3) will involve an exploration of three-point shooting performance across different player positions. Subsequently, in the fourth part (4.4), we will delve into the evolution of three-point shooting for specific players, namely LeBron James and Stephen Curry. Lastly, section 4.5 ?????.

4.1 3 points impact

We commenced our analysis with a comprehensive examination of teams, filtering the dataset to include only the "All" values, representing season-by-season data for all teams. Specifically, our focus was on exploring correlations between 3-point shooting and victories, emphasizing both shooting percentages (3P%) and the total number of 3-pointers made (3PM).



(a) Scatter plot tra 3P% e WIN%

(b) Scatter plot tra 3PM e WIN%

Figure 4.1: Scatter plot relationship with Win

From these plots, it is evident that shooting percentage holds more significance than the sheer number of 3-pointers made. In the first scatterplot (Figure 4.1), the correlation is much clearer, whereas in Figure 4.2, the data appears more scattered, still suggesting a positive correlation between the two variables. It is evident, as previously inferred from correlations, that precision matters more than sheer scoring. The focus lies not only on making three-pointers but on minimizing errors.
Furthermore, we find it intriguing to focus on the performance of teams with the highest 3-point shooting percentage. Utilizing the groupby() and idmax functions, we can select, for each year, the team with the highest value in the variable "3P%."

Notably, excluding the 2002 season, all teams with the highest 3P% secured a spot in the playoffs, ranking within the top 16.
In this case, the variable "Ranking" represents the ranking only in 2004, and in 2001, the team with the highest average of 3-pointers made did not make the playoffs (classified outside the top 16). The

Table 4.1: Best 3P score and best 3P percentage between the 2001 and 2004

Anno	3PM			3P%		
	ID	TEAM	3PM	ID	TEAM	3P%
2001	20	Boston Celtics	7.2	1	San Antonio Spurs	40.7
2002	8	Boston Celtics	8.5	19	Washington Wizards	38.8
2003	13	Boston Celtics	8.8	16	Milwaukee Bucks	38.3
2004	18	Seattle SuperSonics	8.8	5	Sacramento Kings	40.1

significance of this data appears to grow over time, with the number of made 3-pointers also showing a noticeable increase over the years.

In conclusion, our analysis of the impact of 3-point shooting on basketball teams has revealed several noteworthy trends. The exploration of shooting percentages (3P%) and the total number of 3-pointers made (3PM) provided valuable insights into the dynamics between these variables and team victories. The visual representation through scatter plots highlighted the dominance of 3P% over the sheer volume of 3-pointers made in influencing team success. Teams with higher shooting percentages tended to exhibit a more pronounced positive correlation with win percentages, emphasizing the importance of accuracy over quantity in 3-point shooting. Focusing on the top-performing teams with the highest 3P%, our analysis showcased consistent playoff appearances for these teams, reinforcing the significance of effective 3-point shooting in achieving success on the court. Additionally, the temporal evolution of 3-point shooting statistics indicated a growing emphasis on this aspect of the game over the years.

Overall, our findings underscore the strategic relevance of 3-point shooting in contemporary basketball, emphasizing the impact of precision in shooting percentages on team outcomes. As the sport evolves, understanding and optimizing 3-point performance emerges as a crucial factor for sustained success in the competitive landscape of basketball.

4.2 3 points time series

Now, we shift our focus to the evolution of 3-point shooting over time. Specifically, we initiate this exploration by comparing the variables 3PA and FGA, representing attempted 3-pointers and total field goal attempts (including 3-pointers).

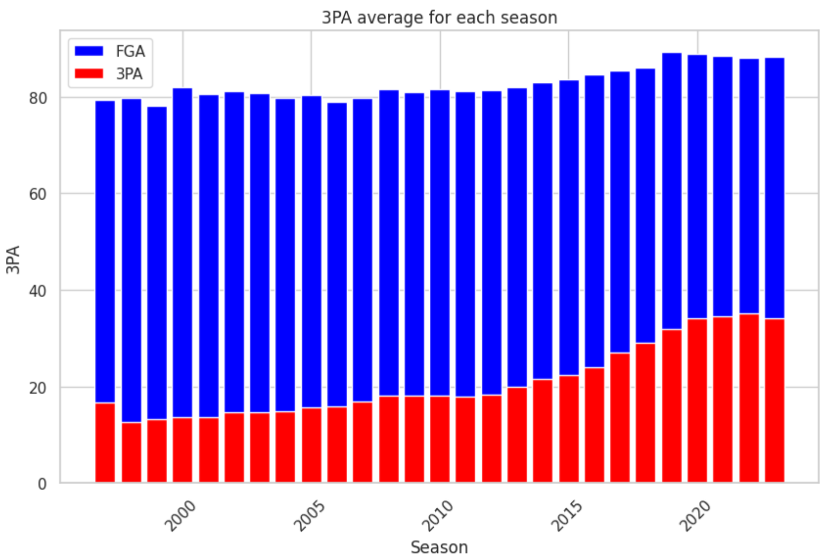


Figure 4.2: Histogram FGA and 3PA mean for each season

Through this histogram, it becomes evident that the prevalence of 3-point attempts within the total field goal attempts has been consistently rising.
In the 1996-1997 NBA season, there is a noticeable outlier. Further investigation revealed that, leading up to the 1994-95 season, the league faced a scoring challenge, with the previous season seeing a 35-year low in average scores. Teams shifted from the high-tempo offenses of the '70s and '80s, adopting a slower, more deliberate style. During the 1990s, the success of the Chicago Bulls, averaging less than 95 possessions per game, influenced other teams to prioritize efficient shot attempts. In an effort to address scoring issues, the NBA briefly shortened the three-point line, resulting in increased attempts during the 1994-95 season. However, this didn't resolve the overall scoring problem as teams continued to play at a slower pace. The NBA restored the three-point line to its original position before the 1997-98 season. It took another 20 years for scoring to return to levels seen in the high-scoring eras of the '70s and '80s.
To assess the impact of 3-point shooting compared to field goals, we examined three seasons (an initial, a mid-term, and a concluding one). We utilized pie charts to visually represent the distribution of shot types, incorporating corresponding percentages for each season.

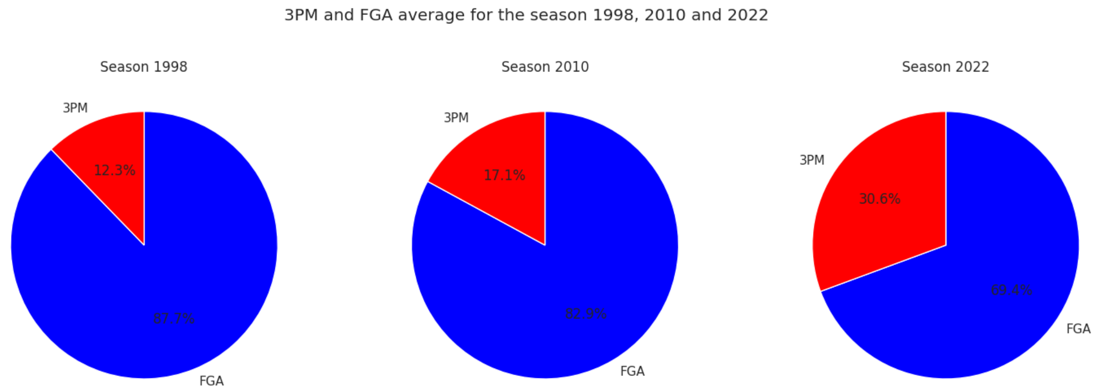


Figure 4.3: Pieplot for 3 season

From this graph, it is evident that 3-point attempts have increased over the years. In 1998, the three-point shooting percentage accounted for 12.3% of total field goal attempts. Over the subsequent 12 years, a gradual increase of nearly 5% was observed, reaching 17.1%. Another 12 years later, in 2022, this percentage nearly doubled, surpassing the 30% threshold.
However, how have the shooting percentages evolved? Through a line plot, we observe a consistent upward trend in percentages, especially those for field goals.

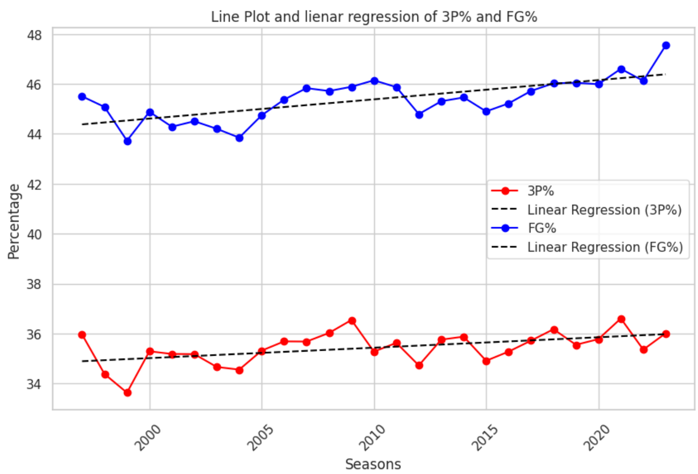


Figure 4.4: Line Plot for 3P% and FG%

The line plot clearly illustrates the temporal trends of three-point shooting percentages and overall field goal percentages for each season. The presence of two upward-trending lines indicates a gradual increase over the years for both three-point shots and total field goals. However, the notable observation of a greater growth in field goal percentages suggests that, despite the increasing popularity of three-pointers, the success rate in overall field goals has experienced a more significant improvement. This trend may reflect a more balanced gameplay strategy or an overall enhancement in the effectiveness of shooting, encompassing both three-pointers and two-pointers. The divergence between the two lines could be interpreted as an indication of how teams continue to evolve and optimize their offensive strategies over the years. It is apparent that players tend to opt for easy 2-point shots (close to the basket), and when shooting from a distance, they prefer attempting 3-pointers.

4.3 Differents position analysis

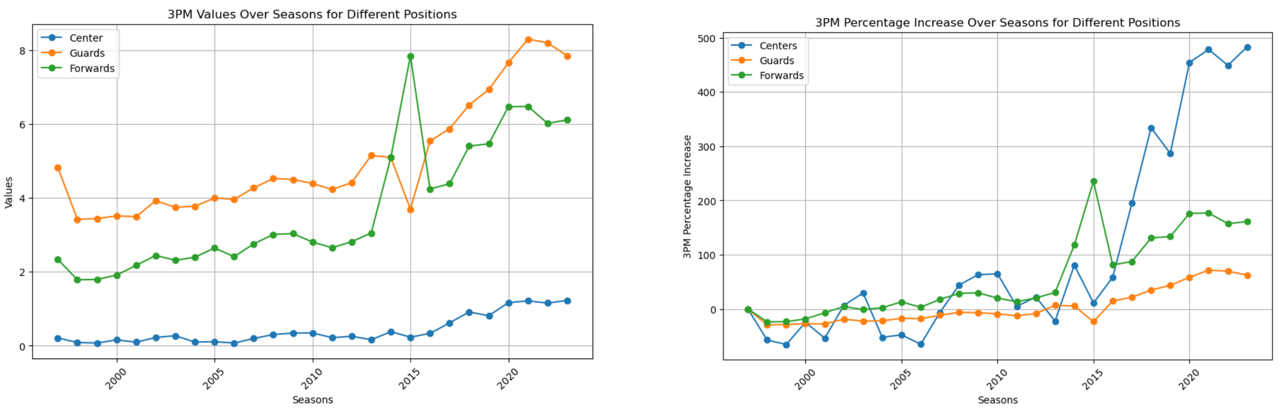
To examine the impact of the three-point shot in the NBA over the years, we took a detailed approach to analyzing which role has undergone a significant change in its playing style. Initially, we divided players into three main categories based on their playing positions: Centre (C), Guard (G), and Forward (F). This subdivision allowed us to explore trends and variations in role-specific playing statistics over the course of different seasons.

The DataFrame is divided into three subsets based on the playing position of the players: Centre, Guard, and Wing. The 'Position' column is removed to simplify the analysis.

After removing the 'Position' column to simplify the analysis, we examined the descriptive statistics in detail, highlighting changes in player performance over time. Next, we applied a more in-depth temporal analysis, grouping the data by year and team, and calculated averages to obtain a clear view of the evolutionary dynamics.

To explore the impact of the three-point shot (3PM) in the NBA over the years, we employed the matplotlib library to conduct a detailed analysis of the performance of each player position. Initially, we categorized the players according to their main positions: Centre (C), Guard (G), and Forward (F). This categorization provided us with a specific picture of how the different roles in basketball coped with the evolution of the game over time. Next, we took advantage of matplotlib's graph plotting capabilities to create line visualizations of 3PM values for each position. These graphs allowed us to identify trends and fluctuations in the number of three-point shots made during different seasons.

To further enrich the analysis, we calculated and graphically represented the percentage value indicating the increase in 3PM over the years. This approach allowed us to clearly and visually highlight how the frequency of three-point shots increased or decreased compared to the first year analyzed.



(a) Line Graph between 3PM

(b) Line Graph between 3PM%

Figure 4.5: Line Graph for 3PM and 3PM% for every position

After careful analysis of the graphs, it became clear that the position that experienced the most significant increase in the number of three-point shots (3PM) was the Centre. Faced with this evidence, we focused on a more in-depth analysis of this specific position.

To obtain a more complete view of the Centres' playing style, we compared 3PM with three-point shot attempts (3PA), shots made from the field (FGM) with total attempts (FGA), and finally, we examined the percentage variation between three-point shot attempts (3PA) and total attempts (FGA). This approach allowed us to identify significant details in the game dynamics of the Centres, understanding how their involvement in three-point shots and total shots evolved over time. The detailed analysis of these metrics provided us with an in-depth insight into the Centre's position-specific game strategies in addressing the rise of the three-point shot in the NBA.

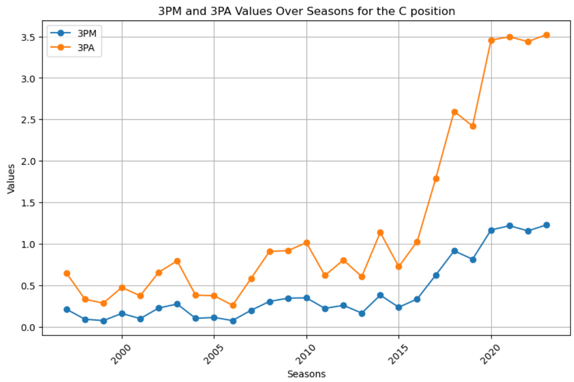


Figure 4.6: Line graph for 3PM and 3PA for the centre position.

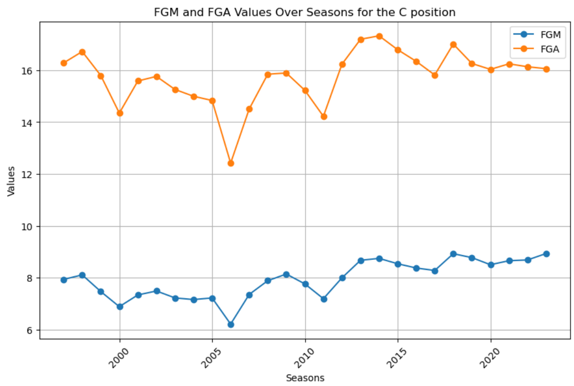


Figure 4.7: Line graph for FGM and FGA for the centre position

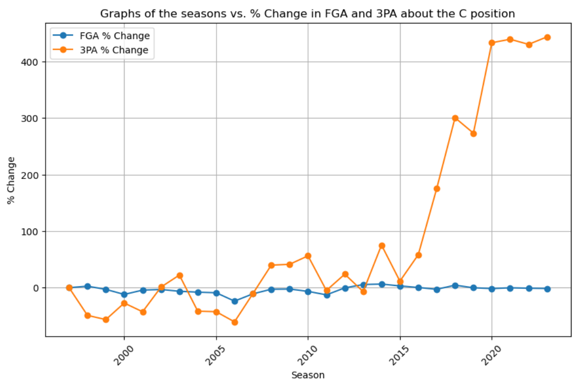


Figure 4.8: Line graph for FGA% and 3PA % for the centre position

In-depth analysis of the data revealed a notable trend in the Centres' position, evidenced in particular by comparing three-point shots (3PA) to total shot attempts (FGA). Interestingly, the

Centres significantly increased the number of three-point shots, while total shots attempted remained relatively constant over the years. However, the most intriguing finding emerges from the third graph (Figure 4.8), which illustrates the percentage change between three-point shot attempts (3PA) and total shot attempts (FGA). This graph reveals a crucial aspect: while maintaining a constant number of total attempts, the Centres have radically changed their style of play over time. Instead of concentrating mainly on shooting under the basket, they started to move towards the perimeter, showing a clear shift towards the three-point shot. This strategic shift indicates a significant evolution in the Centres' style of play, as they have moved away from the area closest to the basket to embrace a more prominent presence on the perimeter. This transition is clearly evidenced by the increasing number of three-point shots, underlining a relevant trend in the way Centres contribute to their teams' attack in the NBA.

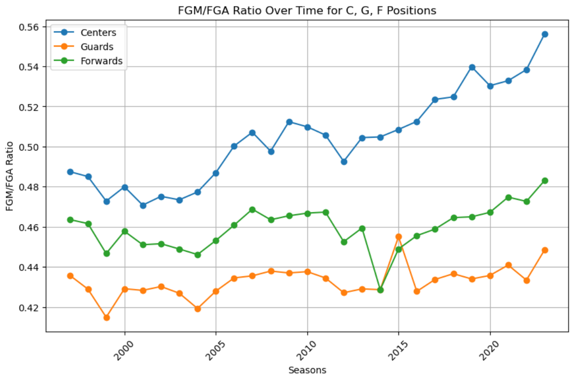


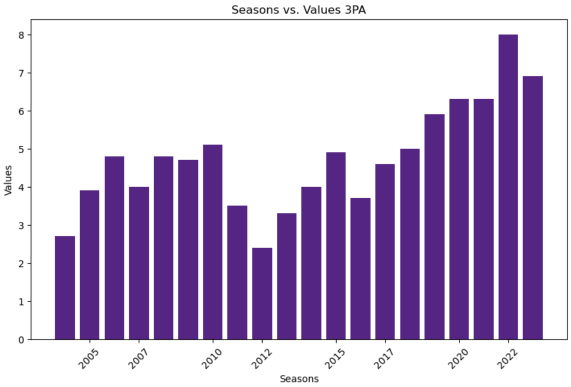
Figure 4.9: Line graph for FGM and FGA ratio for every position

When analyzing the ratio of shots made from the field (FGM) to total shot attempts (FGA) for all positions, we clearly observed that it is the role of the Centre that has seen a significant increase in shooting efficiency over the years. The most striking fact emerges when comparing this ratio over time, revealing a steady growth in efficiency in the shots made by the Centres. This indicates a significant improvement in the accuracy of their shots, suggesting that, despite the increase in three-point shots, the Centres are also becoming more efficient in their more traditional actions under the basket. This trend highlights the increasing versatility of Centres in the NBA, who are demonstrating improvement in both their ability to score from three-point range and in shooting closer to the basket over the years.

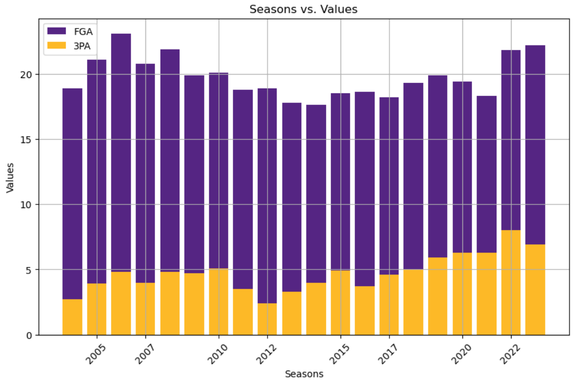
4.4 LeBron James and Stephen Curry evolution

LeBron James

We chose to analyze this to player for what they represent in the basketball world, one being the all-time leading three-point scorer who is still playing, and the other being the all-time leading scorer in general and the oldest player in the league who has experienced the evolution of three-point shooting and adapted to it. LeBron James stands as the most prolific scorer in the history of basketball. Therefore, envisioning a true revolution in the basketball world involves capturing this transformative shift through the lens of LeBron James' evolutionary journey. Our analysis commenced with the creation of a bar chart illustrating the progression of attempted and successful three-point shots. The charts depicts the evolution in LeBron's gameplay from the early 2000s to the present, also in comparison to the whole of field goals attempted. Looking at these graphs, we can definitely say the Lebron James has changed is game in the last years to adapt to a developing environment.



(a) 3 Points Attempted



(b) 3 Points Attempted

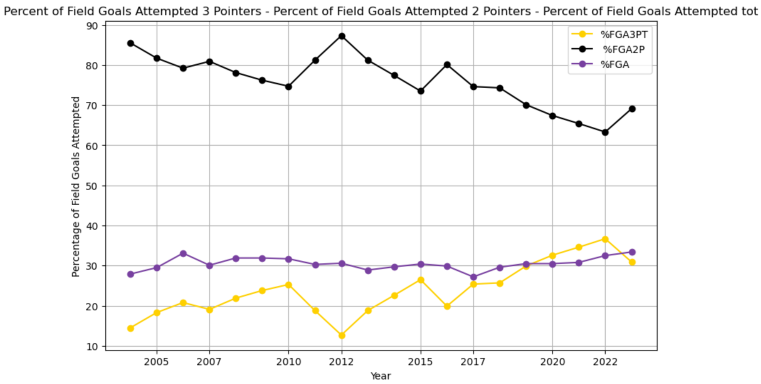


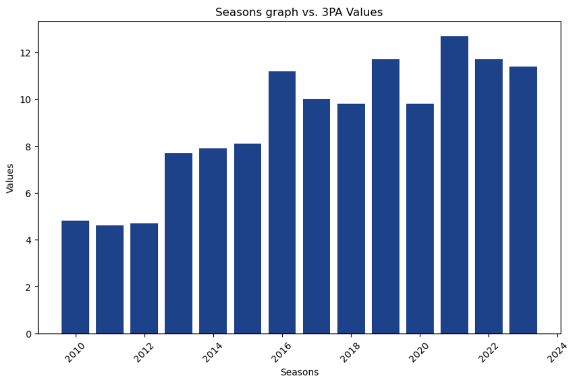
Figure 4.10: Percent of Field Goals Attempted 3 Pointers - Percent of Field Goals Attempted 2 Pointers - Percent of Field Goals Attempted tot

In the above graph, we observe a complementary relationship between 3-point shooting and 2-point shooting, signifying a preference for and selection of 3-point shots over their 2-point counterparts.

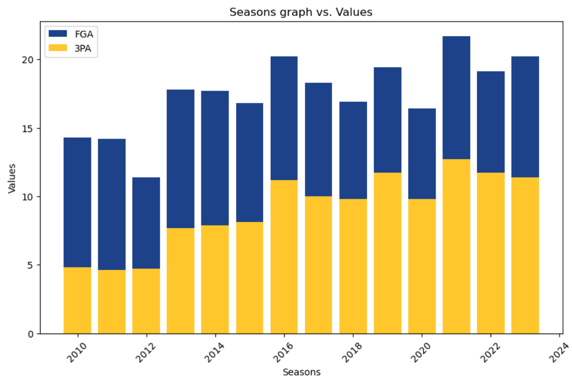
Stephen Curry

Stephen Curry, he is widely regarded as one of the greatest three-point shooters in basketball history due to his exceptional shooting accuracy, high volume of successful three-point shots, ability to create shots off the dribble, influence in revolutionizing the game's style towards more emphasis on three-pointers, and his consistent, long-term success in maintaining a high level of performance. Our analysis started with the creation of a bar chart illustrating the progression of attempted and successful three-point shots. The charts depicts Curry's game play from the 2010s to the present, also in comparison to the whole of field goals attempted.

We can observe a drop in 3-points shoots made in 2020, with some research we realized that it is



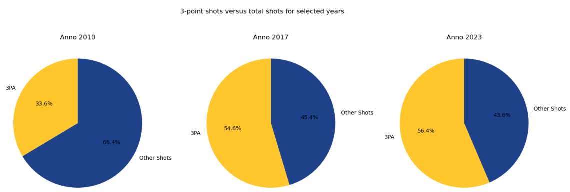
(a) 3 Points Attempted



(b) 3 Points Made

Figure 4.11

due to a hand injury. Overall, it is remarkable the on going growth in 3 points shootings.



(a) 3PA vs FGA

The graph above shows how in almost every season the shots Stephan Curry attempts are more than 50% of the time 3-Pointers.
It's noticeable the increase in 3 Points attempts in LeBron's game, but it is evident how this move is dominated by Stephen Curry.

Shifting the focus on the contribution that Curry brought to the Golden State Warriors, the team he played in for all his career, we want to analyze the impact he had on the strategy of the team. For this reason we analyzed a line graph to show the evolution of the 3 Points shooting attempts over time.

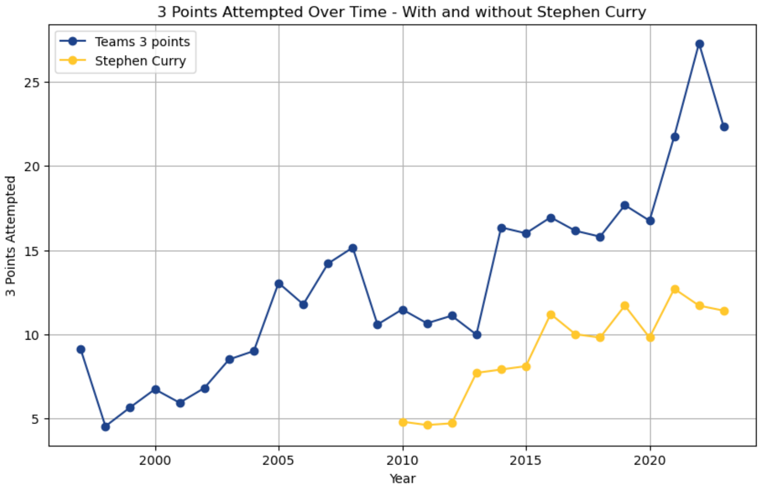


Figure 4.12: GSW with and Without Stephen Curry

We can notice how the strategy of the team developed throughout the years, we can also see how the 3 points shoots attempted follow the lead of Curry, with a significant drop in 2020 when he was injured.

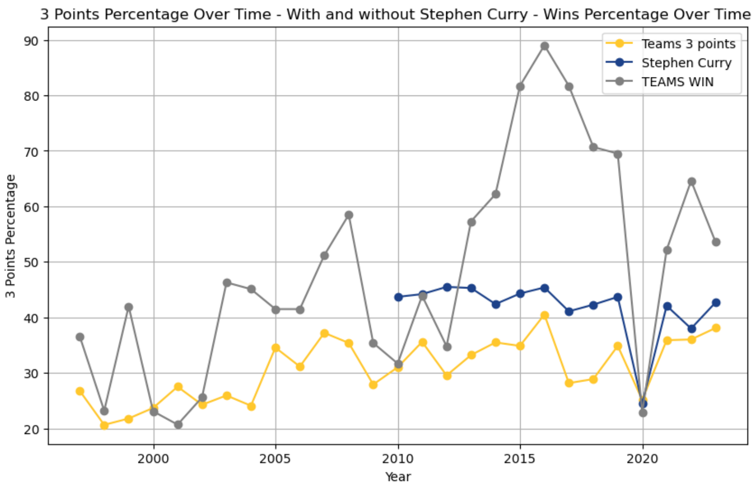


Figure 4.13: How Curry influenced GSW wins

Analyzing the data, it's clear that Stephen Curry's role in the Golden State Warriors' wins is visually apparent. The graph highlights a notable connection between Curry's performance and the team's success. His scoring prowess, particularly from three-point range, corresponds with high points in the team's overall performance, signaling that Curry's efforts play a substantial role in securing wins. The data indicates a straightforward influence, demonstrating how Curry's offensive abilities, strategic plays, and leadership directly impact on-court outcomes, contributing to the team's overall success.

LeBron vs Curry

Over the past two decades, LeBron James and Stephen Curry have emerged as two of the most influential and dominant players in the NBA. While their playing styles differ significantly, both players have left an indelible mark on the league. We compared the three-point shooting performance of LeBron James and Stephen Curry in their respective games. Specifically, our analysis will focus on the frequency of attempted three-point shots to provide insights into their playing styles and strategic choices in this aspect of the game.

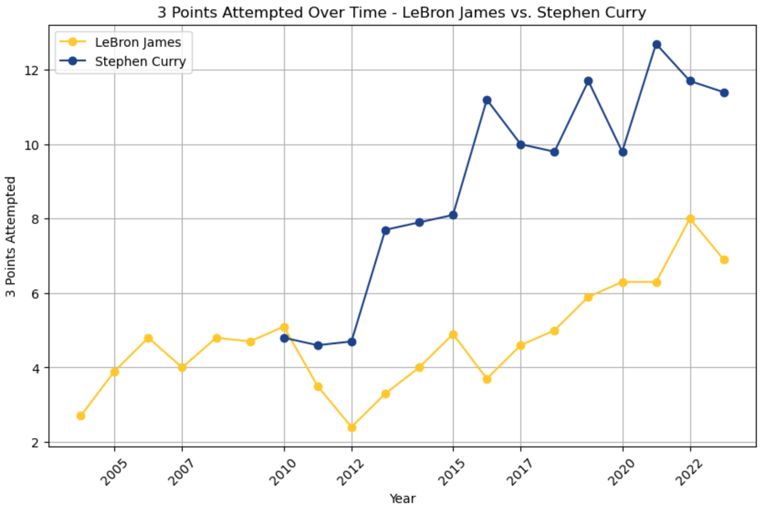


Figure 4.14: LeBron vs Curry 3 Points Attempted

5 | Predictive model

In our ongoing effort to enhance our understanding of team performance in the NBA, we've embarked on the development and evaluation of predictive models. The focus is on forecasting the win percentage of each team for the current season (2023-2024). Our objective is to identify the most effective model by engaging in a rigorous process of training and testing on two distinct datasets.

The first dataset is constructed solely with variables related to 3-point statistics, emphasizing the significance of accurate three-point predictions. The second dataset, in contrast, includes all the variables gathered and employed in our previous analyses. This comprehensive approach recognizes the potential limitations of predicting win percentages solely based on 3-point variables, thus ensuring a more holistic understanding of team dynamics.

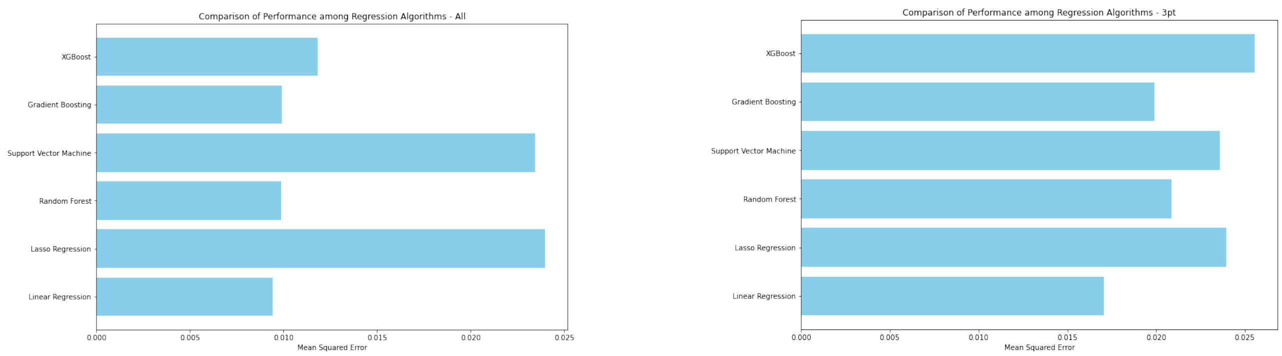
Data Preprocessing and Model Training

To initiate the predictive modeling process, we utilized essential Python libraries and imported the required datasets. The first dataset ('df_All') encompasses all players without distinguishing positions, while the second dataset ('X_3pt') specifically includes variables related to three-point shots. We employed the train-test split method to divide our datasets into training and test sets, crucial for training and evaluating our predictive models.

Various regression algorithms, including Linear Regression, Lasso Regression, Random Forest, Support Vector Machine, Gradient Boosting, and XGBoost, were employed for the analysis. Mean Squared Error (MSE) was chosen as the evaluation metric to assess the performance of each algorithm.

Algorithm Performance Comparison

After training and testing each model on both the complete dataset and the three-point variables dataset, we visualized the results using a bar chart. The chart illustrated the Mean Squared Error for each algorithm, providing a clear comparison of their performance.



(a) df All **(b)** df 3pt

Figure 5.1: Comparison of Models Performance

The results were then presented in tabular form, offering a concise overview of each algorithm's MSE.

Model	Mean Squared Error
Linear Regression	0.009
Lasso Regression	0.024
Random Forest	0.01
Support Vector Machine	0.023
Gradient Boosting	0.01
XGBoost	0.012

Table 5.1: MSE for Algorithms using df All

Model	Mean Squared Error
Linear Regression	0.017
Lasso Regression	0.024
Random Forest	0.021
Support Vector Machine	0.024
Gradient Boosting	0.02
XGBoost	0.026

Table 5.2: MSE for Algorithms using df 3pt

Selection of Linear Regression Model

From the graphical representations and tables above, it is evident that the Linear Regression model outperforms others in terms of performance for both dataframes, although there is not a significant difference observed among some other algorithms. The evaluation of performance was based on Mean Squared Error (MSE) as the metric. After careful consideration of the consistent accuracy demonstrated by Linear Regression across various datasets, we have opted to employ this model for predicting the win percentage for the 2024 season.

2024 Win Percentage Predictions

To forecast win percentage values for the 2024 season, we incorporated data extracted from the NBA stats website up until November 3rd. Linear Regression models were trained separately on the entire dataset and the three-point variables dataset. Predictions for the 2024 season were then generated using these models.

Results and Standings Prediction

The predicted win percentage values were rounded for clarity, and the results were organized in a combined dataframe for comparison. Linear Regression predictions based on all variables and those specific to three-point statistics were presented side by side. Additionally, we generated a table showcasing the predicted NBA standings at the end of the season for both models.

We will illustrate below the top 5 predictions made by each of the two models, along with their corresponding win percentages. From the predictions of our models, the following results were obtained:

Team	Predicted Value
Dallas Mavericks	0.775
Boston Celtics	0.713
Brooklyn Nets	0.708
LA Clippers	0.685
Denver Nuggets	0.635

Table 5.3: Predicted Values for 2024 using all variables

Team	Predicted Value
Dallas Mavericks	0.907
LA Clippers	0.863
Brooklyn Nets	0.838
Boston Celtics	0.833
Philadelphia 76ers	0.674

Table 5.4: Top 5 Predicted Values for 2024 using 3pt variables

The table predictions indicate that based on early game data, the Dallas Mavericks are expected to emerge as the winner of the NBA 2023-2024 regular season in both scenarios. However, the positions of the other teams vary between the two models.

Discussion on Predictions

It is crucial to note that these predictions are preliminary and are based on limited data, specifically encompassing statistics from the initial 5-6 games of the regular season. We have attempted to fit a model focusing solely on variables related to three-pointers, both for consistency with the project's theme and to explore feasibility. However, it is important to recognize that basketball is a sport with numerous variables, and it is highly unlikely that predictions from a model considering only three-point shooting would be reliable. Continuous monitoring and re-evaluation of the models will be imperative as the season progresses to refine predictions, given the dynamic nature of the sport and the potential impact of other contributing factors.

6 | Conclusions

In conclusion, this project constitutes a thorough exploration of the evolution of basketball over the past decades, with a particular focus on the transformative impact of the three-point shot. Leveraging official NBA statistics, we employed descriptive statistics and data visualizations to quantify the magnitude of change in various aspects of the game, including player roles and team dynamics. We obtained several intriguing results on the evolution of basketball in recent decades, applying various tasks such as web scraping, data exploration, and machine learning to gain a more comprehensive view of this increasingly fundamental aspect of the sport.

6.1 Results

Through our analyses, we have observed a significant transformation in the playing style of NBA teams over the past few years, as anticipated. Notably, there has been a radical shift towards an increased frequency of attempted three-pointers. Furthermore, our findings underscore the growing importance of three-point shooting accuracy for achieving strong performances in the NBA regular season. It is a necessary but not solely sufficient factor; thus, while having proficient three-point shooters is crucial for most teams, it is not the sole determinant of success.

The metamorphosis of three-point shooting over the last decade has seen teams opt for more three-point attempts rather than relying on orchestrated plays for two-pointers. Another noteworthy insight pertains to the evolution of player roles in basketball, particularly the transformation in the playing style of centers. The remarkable surge in three-point attempts by centers in recent years is nothing short of astonishing.

A glaring example of this shift in basketball dynamics is the evolution of arguably the greatest player of all time, LeBron James. Over the course of his career, he has increasingly transitioned towards three-point shooting, despite initially being more inclined towards penetrations and orchestrated plays. Examining the impact of the greatest three-point shooter in NBA history, Stephen Curry, on his team's performance speaks volumes about the evolving nature of the game and the indispensability of three-point shooters in a winning team.

Finally, our predictive model offers a compelling forecast for the 2023-2024 NBA regular season standings, suggesting that the Dallas Mavericks could clinch the top spot based on their proficiency in three-point shooting.

6.2 Further developments

To enhance the project and analyses, a more nuanced approach could involve employing specific and in-depth analyses of three-point shooting to provide advanced insights into the basketball transformation. A more targeted examination of the players under analysis could be conducted, delving into the mechanics of their actions and three-point shooting performances. This would help ascertain whether their proficiency in this aspect is intentional or if it aligns with a strategy dictated by the coach and the team's playing style.

For the predictive model, incorporating additional variables related to three-pointers would be valuable. Expanding the scope of data acquisition to automate the process could enable real-time updates to predictions after each NBA game day. This would not only ensure the model remains current but also allow for a more dynamic and responsive analysis of the evolving trends in three-point shooting and its impact on team performance.

Bibliography

[1] NBA. 2023, NBA Statistics. https://www.nba.com/stats

[2] Schuhmann, J. 2023, NBA's 3-point revolution: How 1 shot is changing the game. https://www.nba.com/news/3-point-era-nba-75

[3] Shea, S. 2019, THE 3-POINT REVOLUTION. https://shottracker.com/articles/the-3-point-revolution#:~:text=The%20change%20was%20gradual%20at,3%20FGA%20is%20a%203PA

[4] Wikipedia. 2023, Stephen Curry. https://it.wikipedia.org/wiki/Stephen_Curry

[5] —. 2023, LeBron James. https://it.wikipedia.org/wiki/LeBron_James