A Work Project, presented as part of the requirements for the Award of a Master’s degree in Business Analytics from the Nova School of Business and Economics.

Exploring Patterns and Influences on the *Hot Hand* Phenomenon in the NBA: A Comprehensive Analysis

The Role of Player Expertise in the Manifestation of the *Hot Hand* in the National Basketball Association

Beatriz dos Santos Leal

Work project carried out under the supervision of:

Michail Batikas

10/01/2024

Abstract

This research assesses the complexity of the rare phenomenon of the *hot hand* in the NBA. A detailed data collection methodology, involving the systematic extraction of information from official NBA and WNBA websites through web scraping, has been employed. The results obtained yield significant insights. The analysis reveals a discernible concave relationship between player experience, age, and occurrences of the hot hand, consistent with existing literature on the delicate balance between accumulated knowledge and the physical decline associated with aging. Furthermore, the study uncovers patterns in shooting behaviour associated with heightened player experience.

Keywords

Sports Analytics, NBA, Hot Hand, Hypothesis Testing, Regression Analysis, Data Extraction, Quarters, Home Advantage, Player Expertise, Shot Difficulty, Defensive Pressure, WNBA, Gender Dynamics, Risk Aversion, Prediction, Classification Algorithms, Machine Learning

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209)

1 Introduction

In recent years, the NBA has continuously leveraged data analytics to enhance team and player performance. Through the course of each basketball game, a vast and detailed cache of data is gathered by a variety of technologies, such as cameras and sensors, which is then delivered to the teams. Analytics are shaping a new era, subjecting players a meticulous scrutiny through metrics that range from a straightforward average field goal percentage to the specific court area in which they achieve higher accuracy, or even the angles at which they shoot. With extensive datasets containing this information, it is possible to identify situations where players are exceeding expectations, surpassing their average performance. On the contrary, areas of improvement can be pinpointed in the case of underperformance.

During basketball games, it is common to have commentators state that a player has a *hot hand* when they achieve a remarkable streak of field goals. Players, coaches, and fans collectively acknowledge that when players are on a streak, often described as being "in the zone", they are almost incapable of missing the next shot. However, humans tend to have incorrect intuitions about the laws of chance, leading them to underestimate the likelihood of the occurrence of streaks. In fact, it is often believed that random events should exhibit more balance that what is indicated by statistics. Upon the controversy around this subject, a primordial question arises – do *hot hand* streaks fit within statistical rules, or, on the contrary, do they defy them?

This question has troubled researchers for decades, motivating numerous studies. However, a consensus has not yet been reached. Despite assertions by a multitude of scholars that this phenomenon is a fallacy, it has also been acknowledged that the ability to detect it lies on the employed definition of heat, along with the foundational assumptions. Nevertheless, even if the *hot hand* is merely an illusion, the belief in it brings changes to NBA games, as players often choose to feed the ball to a teammate thought to be *hot*. In fact, the significance of the *hot hand* lies in the behavioral changes caused by the conviction in its existence.

In this study, a comprehensive examination of the *hot hand* phenomenon is built upon previous findings achieved over the last decades, with the goal of scrutinizing the underlying dynamics. Unraveling the key factors that systematically impact *hot hand* occurrences, along with the prediction of such events, surpasses the confines of the theoretical realm. A deeper understanding of the phenomenon may fundamentally transform the very essence of basketball gameplay, representing an advance in the analytics era of sports.

The foundation of this study is a dataset with NBA shot records from the 2003-04 season to 2022-23. A thorough analysis of the data has been conducted, leading to a profound understanding of the available information. In addition, it is recognized that gathering additional data would be enriching since it could provide further insights, employing a methodology for collecting it. The additional data comprises information that complements the original dataset, broadening the richness of the conducted analysis of the NBA, and WNBA data, to enable gender comparison.

Following an extensive review of previous literature, the *hot hand* exploration in this study initiates with the definition of heat. The broad body of work available on this subject allowed for an assessment of multiple approaches, achieving a definition of *hot hand* that overcomes bias previously experienced by several authors. Subsequently, a comprehensive analysis of the phenomenon is founded on this definition, with the goal of determining the underlying patterns and building a model to predict future occurrences.

The current work employs regression analysis to assess the *hot hand* dynamics from several distinct angles, emphasizing factors that have been proven to be significant in past studies. It commences with a section focused on the quarterly influence and home-court impact on *hot hand* shots, attaining results that elucidate the variation in the *hot hand* specificities through the course of a game and across home and away matches. Subsequently, a new section unfolds, with insights related to the determination of the impact produced by player experience and age, along with an in-depth analysis of the associated shooting patterns, with regard to measures such as the shot distance. Following this section, this study directs its attention to streak patterns and the impact of team strength and defensive pressure on *hot hand* shots. The dissection of patterns within a *hot hand* streak is made with regard to shot difficulty, by employing an adaptation of a shot difficulty measure present on previous literature. Finally, the scrutiny of the *hot hand* is further extended to the WNBA, leveraging the additional data collected. A comparison between the two leagues is performed to compare risk-taking tendencies across genders.

Following the determination of patterns linked with the phenomenon under scrutiny, a predictive model of *hot hand* shots is designed. Several machine learning algorithms are trained in selected features of the data, utilizing techniques that overcome the challenges posed by the rarity of these events. A thorough preprocessing, including preliminary data cleaning, feature engineering, handling dataset imbalance and scaling is performed to ensure the model's efficiency in capturing hidden patterns and delivering accurate predictions. Notably, the chosen algorithms are tailored to the unique characteristics of the dataset, including its imbalance and size, and carefully chosen taking into account previous literature on the area, thereby enhancing the model's predictive performance.

Conducting such a thorough assessment of the *hot hand* phenomenon marks a significant step towards comprehending it, paving the way for leveraging the knowledge in this subject to drive teams and players to success. Further delving into the results achieved in this study, through an incorporation of enhancements and additional methods, holds the potential to unveil findings that can be taken into account for the development of data-driven game strategies with a significant impact in performance.

2 Literature Review

The duality between belief and scepticism regarding the *hot hand* phenomenon in basketball has sparked massive controversy and has led to numerous studies that acknowledge its significance, as the belief in its existence among players changes their behaviour and, consequently, the course of the game. Not long ago, the empirical evidence for the existence of this phenomenon was limited. More recently, the increasing detail and amount of collected in-game data introduces new opportunities for deepening our knowledge on this subject, since it can be leveraged to detect new patterns.

Some researchers have focused on settling the psychological foundations of being *hot*. Iso-Ahola and Dotson (2014) posit that there are performance-enhancing effects of a phycological *momentum*, since players are better performers when their confidence escalates. In this study, it is argued that the *hot hand* is easier to measure in more uniform sports, where the conditions of performance trials do not diverge drastically, such as bowling. Basketball does not provide this uniformity since it is so dynamic and multi-faceted, meaning that it might be more challenging to detect the *hot hand* phenomenon. To overcome this, some authors have settled for more controlled environments where the shot conditions are similar. In fact, Miller and Sanjurjo (2021) argue that the presence of in-game features such as large crowds and overall game environment possibly leads to differences in each *hot hand* shooting and increase the complexity of assessing this phenomenon. In order to address this obstacle, this study used data from the annual NBA Three Point Contest, which is considered to be an ideal setting to study this phenomenon, since it connects several features of a controlled setting to the high stakes and crowd of an NBA game. In this annual event, each shot is roughly taken from the same distance and without most of the strategic features present in a normal game. However, research conducted by Pelechrinis and Winston (2022) provided further insights on this subject, arguing that the conditions of shot attempts within a sequence do not need to be identical when assessing the presence of the *hot hand* effect, in contrast to the approaches followed by Gilovich, Vallone, and Tversky (1985) and, more recently, by Avugos et al. (2013). This is crucial for the analysis of the *hot hand* in basketball, since it is a dynamic sport with many variables involved.

The controversy around the *hot hand* motivated research that gathered in-game data to determine if a player who is outperforming is in fact more likely to make the next shot successfully. Research conducted by Bocskocsky, Ezekowitz, and Stein (2014) focused on identifying the situations where the number of streaks of one player is higher than what would be expected to occur by chance, pinpointing a positive correlation between shots. This study estimates an increase of 1.2% in the probability of a player making a successful next shot for each additional shot he made before. It recognizes that players who believe they are *hot* take more difficult shots, accounting for an increase of 4.5% in the distance from the basket. The findings of this study are in line with the research conducted by Rao (2009), where players were further allocated into one of three groups, depending on their reaction to past success and on the arising consequences in terms of points. While there are players who react to prior success by taking more challenging shots, it is possible that others do not change their behaviour. At the same time, different changes in behaviour mean different consequences on the outcome of the game. Furthermore, Bocskocsky, Ezekowitz, and Stein (2014) alert to the fact that despite *hot* players having a higher shot accuracy overall, their overconfidence might escalate and lead them to take excessively difficult shots, lowering shot success probability. These two effects would then cancel each other out. More recently, Miller and Sanjurjo (2018) also drew attention to the overestimation of the *hot hand* phenomenon. Thus, results of these papers highlight the importance of accounting for the implications of a player’s change in behaviour due to his belief on being *hot*.

In a widely cited paper, Gilovich, Vallone, and Tversky (1985) suggest that there is no evidence of *hot hand* in basketball, contradicting the held belief among fans and arguing that it consists of a mere cognitive illusion. The authors further assert that even if the *hot hand* does, in fact, exist, players are incapable of predicting it accurately. However, subsequent studies questioned the truthfulness of these results after a substantial bias was found. The critical test used by Gilovich, Vallone, and Tversky (1985) is susceptible to bias due to the unreal expectation that a Bernoulli shooter (50% shooter) who takes 100 shots will make half of them after a streak of three. After correcting for that bias, Miller and Sanjurjo (2018) were able to find evidence of the *hot hand* in the data.

The powerful belief in the *hot hand* affects the decision-making during a basketball match, promoting choices that might be decisive to the outcome of the game, especially in a sport where opponents will likely have a close final score. Large evidence suggests that players will most likely pass the ball to their teammate who is believed to have the *hot hand*, as it is stated by Burns (2001), however, that might not be the best strategy in all situations. In fact, results obtained in research conducted by Bocskocsky, Ezekowitz, and Stein (2014) shows that the opponents tighten their defence on the *hot* player, which hinders the success probability of a shot made by that player.

2.1 Definition of *Hot Hand*

The definition of *hot hand* diverges through the extend literature on this subject. Traditionally, the *hot hand* consists in the belief that if a player is on a streak of successful shots, he is more likely to continue to make the next ones, meaning that there is a positive correlation between consecutive successful shots. Contrary to the traditional belief shared among the fans, Pelechrinis and Winston (2022) contend that getting *hot* during a basketball game might be a result of the player’s ability to read the game and explore effective shot opportunities, as a result of his experience. Controversially, Bocskocsky, Ezekowitz, and Stein (2014) argue that a player might have a *hot hand* even if he missed the last shots.

In a vast collection of papers, authors focus on determining the number of situations where a player goes on a streak and further comparing it to the number of expected streaks emerging from identically distributed and independent trials, where a player has a fixed probability of success and is therefore often referred to as a Bernoulli shooter, as explained by Bocskocsky, Ezekowitz, and Stein (2014). Gilovich, Vallone, and Tversky (1985) suggest that the belief in the *hot hand* may be due to the fact that streaks are more memorable, generating a memory bias, and to the underestimated probability of streaks occurring by chance.

It is also relevant to mention that despite it being frequent for a player to shoot a free throw through the course of a basketball game, Koehler and Conley (2003) draw attention to the fact that this category of shots should not be included in a *hot hand* analysis. The logic lying beneath this statement is that the accuracy of these shots is extremely high for all players and that consecutive attempts occur within an extremely short period of time.

Rao (2009) argues that when defining heat, one has to control for difficulty using shot conditions such as distance from basket and location. This control is critical to avoid potential downwards bias in the inference of correlation between consecutive successful shots, particularly because players tend to take more difficult ones. Nevertheless, it is also fundamental to choose the critical number of made shots that constitutes a streak, which is a highly controverse subject among researchers. Miller and Sanjurjo (2018) posit that a compromise should be achieved, since setting a larger streak threshold allows for a stronger detection of a *hot hand*, but it also leads to smaller samples. This study is consistent with Rao (2009)'s choice of streak threshold as they both focus on streaks of three, arguing that a lower threshold will aggravate the bias, and taking into consideration that it is common for people to believe a player is on a streak after three made shots. Regardless, the ability to detect a *hot hand* depends on the assumptions taken and on how heat has been defined, as posited by Csapo and Raab (2014).

2.2 Influencing Factors

2.2.1 Home-Court Effect

Traditionally, an important factor for performance during basketball games is the game location. Basketball is known for having a substantial home advantage, being among the two sports with highest advantage, second only to soccer in Europe, as stated by Harris and Roebber (2019). For instance, when analysing NBA play-by-play data of games spanning thirteen seasons, Ribeiro, Mukherjee, and Zeng (2016) found that teams playing home win 60% of the games, which is more than the 50% percentage that would be expected by chance.

Home-court dynamics have been assessed from a broad range of angles in previous studies. Bustamante-Sánchez, Gómez, and Jiménez-Saiz (2022) investigated game-related statistics for teams playing home, away, or in a neutral location, reporting that home-court teams attempted and scored more points overall. However, it was found that while playing at home increased accuracy in 2-point field goals, neutral-court teams accomplished more 3-point field goals. The findings of this study are in line with the research conducted by Harris and Roebber (2019), that suggests that 2-point shots are the main forecaster of home advantage. It is also argued that when playing home, 2-point shots are crucial for maximizing the home-advantage, whereas away-court players should aim for more 3-point shots, although this study does not address why the type of shot matters to the home advantage effect. Work by Ribeiro, Mukherjee, and Zeng (2016) provides a deeper understanding of this effect, stating that it is mainly accumulated in the first quarter of the game. Jones (2007) adds that it is stronger when the home team is losing at the beginning of the quarter. In addition to the type of shots made, the game location also influences the game pace, more specifically the length of time intervals between shots of both teams, as stated by Ribeiro, Mukherjee, and Zeng (2016). In this study, it was found that home-court teams may exhibit a more rapid scoring rhythm, with thinner intervals between scores.

However, it is further stated that the increasing fatigue among players through the course of the game might continuously increase these time intervals for both teams.

Despite the large number of studies about home-court advantage, opinions about the root of this phenomenon diverge among authors. Some relate the home-court advantage to crowd factors such as noise and density, as it is the case of Jamieson (2010), while others attribute it to referee bias caused by the fan's hostility, as it is argued by Anders and Rotthoff (2014). Travel and familiarity with the court are also ascribed to this phenomenon by Pollard R. and Pollard G. (2005).

Nevertheless, studies by Ribeiro, Mukherjee, and Zeng (2016) and Harris and Roebber (2019) have pinpointed a decrease in the home-court advantage through the course of the last decades. The latter study further adds that, considering that the advantage is mostly drawn from a style of play based on 2-point shots, this decrease might be partially explained by the rise in 3-point shooting, which was introduced in the 1979-80 season. Pollard R. and Pollard G. (2005) argue that the NBA reached the highest levels of home advantage during its early years, and that a drop has taken place since the 80's. This study adds that compared to other leagues, such as the NHL and NFL, the NBA has registered the most substantial variation of home advantage across the years.

2.2.2 Quarterly Influence

Previous work has underlined an influence of quarters on shooting. Using data from the NBA, Grund, Höcker, and Zimmermann (2010) found that both teams take more 3-pointers in the last quarter, revealing a quarterly influence in the taken shots. However, the number of minutes left is also considered to be relevant to the type of shots taken, since the rise in 3-point shots happens especially in the last three minutes within each quarter. Additionally, Csapo and Raab (2014) delve into a more granular level of detail, by relating shot difficulty to the number of seconds left in the shot clock, which gives the team playing an offense a total of 24 seconds to score. It is contended that a considerable number of seconds left generally leads to easier shots, since they often happen during a fast break or after an offensive rebound, and that the opposite is also true.

2.2.3 Player Experience

Further, Pelechrinis and Winston (2022) highlight that another factor connected to a *hot hand* is the changing shooting ability of players throughout their career in the NBA. The power of experience alongside other player features was underlined in several studies, and various authors have adopted a different measurement of experience, using either the number of seasons played or the player's age as their preferred metric. Since not all players enter the NBA at the same time after playing in lower level leagues, age is pointed out by Bakkenbüll (2017) as a more accurate quantifier of expertise. Zhang et al. (2018) argue that the players' proficiency should indeed be considered, since experts hold an advantage over novices in reading the game and anticipating situations in court. Research conducted by Kalén et al. (2021) provides further insights on the impact of experience on an NBA player's performance, focusing on the variable age. Findings of this study showed that there is a disparity in the performance measured both across distinct playing positions and across diverse age groups. The conducted statistical analysis demonstrated that the highest number of points per minute among all players was achieved in the ages from 26 to 29. However, players in different playing positions would reach this peak in distinct ages.

2.2.4 Defensive Pressure

In addition, the impact of the defensive pressure applied by the opposing team on player performance has been assessed by Csapo and Raab (2014). This study draws attention to the fact that opposing teams respond in accordance with the *hot hand* belief, tightening their defence on the player on a *hot* streak, simultaneously highlighting that this might prompt the unobservability of the phenomenon during games. It is also highlighted that more importantly than the number of defenders on each shot, the most meaningful measure of defence pressure is actually the defence intensity. The conducted research demonstrates that players attempt more difficult shots not only due to their overconfidence, but also because a tighter defence might push them into it.

2.2.5 Team Quality

Lastly, previous research has delved into the influence of the team's quality in game-related statistics and dynamics. Research conducted by Sampaio, Drinkwater, and Leite (2010) has identified a team quality effect on shooting, indicating that stronger teams perform better since players have an enhanced training environment. Another result of this research is that superior teams also outperform for other variables, such as passing, which is relevant for creating opportunities to pass the ball to the *hot* player. Therefore, a superior team quality could have a significant impact on the course of basketball games and, by extension, on the occurrence and specificities of the *hot hand* phenomenon.

2.3 *Hot Hand* in the WNBA

While extensive literature has been published for the subject of the *hot hand* in the NBA, there are comparatively fewer studies available on its manifestation in the WNBA.

Moss (2022) took as foundation the four factors considered to be key determinants to a team's success in basketball – Shooting, Turnovers, Rebounding, Free Throws - and further evaluated their weight and evolution throughout the years on the WNBA. One result of this study is that field goals amount for 42% of the overall success in WNBA during a game, as opposed to a lower weight of 40% in NBA. In addition, while rebounds and free throws have a larger weight on the NBA, the opposite happens for turnovers, which are more crucial in the WNBA. The variance in these weights exhibits a different type of play in the two leagues, highlighting the need to account for the distinct type of play when comparing NBA and WNBA. It is therefore crucial to factor in that the *hot hand* might manifest differently in both.

Despite the lacking research on the *hot hand* in the WNBA, there are studies regarding risk-taking. Work by Böheim, Freudenthaler, and Lackner (2016) states that female basketball players are more cautious and risk-adverse than men, particularly in decisive moments, when a triumphant risky shot can change the outcome of the game. In fact, previous literature states that higher risk-taking, traduced in higher shot difficulty, is often associated with the *hot hand* phenomenon, as stated by Bocskocsky, Ezekowitz, and Stein (2014). However, it is crucial to highlight that a higher risk does not always imply that the player is *hot*. Hence, further research on this subject is key for a deeper understanding.

2.4 Methodological Approach

In the realm of empirical investigation into the multifaceted aspects of *hot hand* behaviour within the instances of professional basketball leagues, a necessity for exploration of the direction and magnitude of the relationship of this phenomenon with the proposed influencing factors emerges.

For this purpose, the exhaustive analysis made by Mackenzie and Cushion (2013) underscores crucial considerations for sports analytics investigations, particularly relevant to the exploration of factors influencing performance. Emphasizing the need for precision, the researchers advocate for a thorough operational definition of variables, ensuring clarity in the analytical framework. Simultaneously, their insights highlight the significance of aligning sample sizes with the temporal dynamics of the athletic context, to achieve generalizable patterns. The scholars further underline a different approach to the statistical models built on the realm of performance analysis, which considers athlete and coach practical learning, involving the observation of behavioural changes in consequence of the insights gained from the study.

On the specific context of professional basketball, a recurrent methodological choice among researchers probing phenomena related to player behaviour is the application of regression analysis. Notable studies exemplifying this approach include Grund, Höcker, and Zimmermann (2010) examination of whether heightened risk is more prevalent among low-scoring teams, and Bocskocsky, Ezekowitz, and Stein (2014) research on the influence of heat on shot distances and defender proximity. However, given the versatility of regression analysis, it is crucial to identify the appropriate regression model tailored to the specific intricacies of each problem under consideration. While Ordinary Least Squares (OLS) stands out as particularly efficient for the examination of continuous variables, Logistic Regression serves as a primary tool to deal with binary outcomes, as posited by Ergül, Yavuz A., and Yavuz H. (2014), given its capability to maintain the probabilities of the result within the range of 0 and 1. Nevertheless, the selection of a regression model also hinges on the nature of the relationship between variables. For instance, unlike OLS, which is suitable for linear relationships, Polynomial Regression proves to be effective in the cases where the hypothesized relationship is non-linear.

Within the domain of the predictive analysis concerning the *hot hand* phenomenon, work by Arkes (2013) adds that the ability to detect a *hot hand* utterly depends on the used method and sample size, which might understate the effect and not detecting its significance. Such a complex and rare event entails investigating it through different lenses and considering possible limitations involved. Thus, a paramount consideration arises in the form of highly imbalanced data, which requires appropriate handling.

To deal with this challenge, Chawla et al. (2002) advocates a strategic combination of undersampling and oversampling techniques. This research seeks to show how the combination of traditional re-sampling methods, such as undersampling the majority class or oversampling the minority class, can be an effective manner of improving performance of imbalanced classifiers. As such, the study applies a mixture of SMOTE techniques, which imply oversampling using synthetic instances of the minority class, and the random removal of samples from the majority class.

On the other hand, Domingos (1999) offers a different insightful solution to deal with the challenges posed by imbalanced datasets through a cost-sensitive approach. The essence of his proposal lies in the strategic assignment of different misclassification costs to distinct classes within the classification framework. By recognizing that the consequences of misclassifying instances from one class may differ significantly from those of misclassifying instances of another, MetaCost introduces a mechanism to instil a heightened sensitivity to these varying costs in the learning process, incorporating a cost matrix and explicitly articulating the penalties associated with misclassifying instances.

Nonetheless, opting for ensemble models over traditional classifiers, as exemplified by Zhou et al. (2020)'s comparison of Random Forest and Decision Trees in an imbalanced dataset, can be advantageous in tackling these data challenges, due to their bootstrap characteristics. Beyond Random Forest, prominent examples of ensemble models encompass XGBoost, evidenced by the work of Oughali, Bahloul, and El_Rahman (2019), and AdaBoost, widely elucidated in the work of An and Kim (2010), each wielding unique strengths, and providing robust forecasting.

3 Data

3.1 Initial Dataset

The dataset includes records from 23,993 regular season NBA games spanning from the 2003-04 season to 2022-2023, which have been acquired in csv format from a GitHub repository (https://github.com/DomSamangy/NBA_Shots_04_23/tree/main) containing data gathered from the official NBA website. Pre-season and playoff games are not included. The dataset comprises all shot records for each of the games, accounting for a total of 4,012,561 shots and encompassing a variety of variables.

3.1.1 General Game-Related Data

General game data is available for each shot record, including the IDs for the game and the team taking the shot, the corresponding team name, season indicator variables, the date in which the game took place, and the abbreviations of the home-court and away-court teams.

3.1.2 Player-Related Data

For each shot, player-related data encompasses the ID for the player taking the shot, the player's name, the specific position in court which has been assigned to that player, and the group to which that position belongs to.

3.1.3 Shot-Related Data

The dataset provides details for each taken shot. Shot-related data includes variables denoting the shot outcome, both in string and binary format, the action type, the type of shot (2-point or 3-point), the quarter in which the shot took place, the minutes and seconds remaining within the quarter at the time the shot was taken, and spatial data. Spatial data englobes the name of the court zone and of the side of court in which the shot took place, the corresponding abbreviation for the side of court, the distance from the place where the shot has been taken in relation to the centre of the hoop, the corresponding distance range, and, additionally, the specific coordinates of the shot in the court. Distances are expressed in feet.

3.1.4 Data Cleaning

With the goal of reducing redundancy and retaining solely the pertinent information, the dataset has been subject to filtering. The string variable denoting the shot outcome has been dropped, retaining only the corresponding binary variable, as only one is necessary and a binary variable is more computationally convenient. In addition, the two variables related to the side of the court in which the shot took place have also been excluded. This decision is grounded on the fact that the remaining spatial data already provides all the necessary information. Finally, considering the contemporary shift towards transcending the traditional positions in basketball, reducing their relevance for game dynamics, the columns containing data on the player position have been dropped. With these filtering actions, a more concise dataset has been attained.

Moreover, given the diversity of existing action types in the dataset, a new column *groups_action_ type* has been introduced to categorize these according to their specificities, enhancing the clarity and conciseness of the analysis. Further details about how this column has been built are presented in A1.

3.2 Additional Data Collection

3.2.1 NBA Data

Despite the initial NBA dataset containing game and shot by shot data, gathering additional data is valuable for assessing the *hot hand* phenomenon in basketball. Previous studies have demonstrated how a variety of variables impact the manifestation of a *hot hand* through, drawing attention to the need of assessing these situations from various angles. Thus, the initial dataset has been expanded by incorporating extra variables found to be relevant by prior research, aiming to enhance the quality of the analysis and prediction conducted in this study.

3.2.1.1 Defence Rating of each Team per Season

The works of Bocskocsky, Ezekowitz, and Stein (2014) and Csapo and Raab (2014) have demonstrated that adversaries intensify their defensive efforts when confronted with players considered to be *hot*. This phenomenon assumes particular significance when investigating the *hot hand* effect, as it introduces a notable variable that can influence a player's performance.

Hence, the defence rating metric, sourced from the NBA website, has been included in the dataset. This metric is structured as a season-by-season, team-specific statistic, providing an average per game of how many points a team allows per 100 possessions of the opposing team.

3.2.1.2 Player Experience

The inclusion of a metric able to measure players' expertise aligns with the insights drawn from existing literature, which emphasizes its relevance in understanding player performance, as highlighted by Zhang et al. (2018). Furthermore, as addressed before, there is an enduring debate among researchers regarding the most accurate indicator of expertise, with some advocating for age and others favouring years of experience in NBA.

To compile this information, the birthdate and first season played have been extracted from each player's profile on the NBA website. Subsequently, the birthdate has been refined to reflect the player's age at the time of each game, while experience has been computed by calculating the difference in years between the season of each game and the player's first season in NBA.

3.2.1.3 Final Standing and Conference of each Team per Season

As previously mentioned, work by Sampaio, Drinkwater, and Leite (2010) identified a team quality effect on shooting and passing performance, stating that stronger teams achieve superior results. In addition to shooting, which is directly related to *hot* players, one could argue that proficiency in passing allows for a more strategic ball allocation, potentially feeding the *hot hand*. Thus, team strength warrants consideration. Diverse aspects can define a team's strength, making the task of measuring it quite challenging given the need to consider many metrics.

Hence, to incorporate each team's strength into the dataset, the final standing of each team within its conference has been scrapped for each season, employing it as proxy of team strength.

Depending on their geographic region, NBA teams are divided into Eastern and Western conferences. During the regular season, to which the game data belongs to, teams compete against others belonging to the same conference. The employment of this system during the NBA regular season creates the need to add another column, complementary to the one indicating each team's ranking per season, specifying which conference the team belongs to.

3.2.2 WNBA Data

In addition to exploring the *hot hand* in the NBA, the goal of this study extends to delving into the same phenomenon for the WNBA, deepening the available research on this topic for female basketball players. For enabling a comparison between the two leagues, a similar dataset has been built for the WNBA. General game information, play-by-play data, and spatial data have been extracted from the official WNBA official website, which originated a dataset incorporating all the variables that are simultaneously in the NBA dataset and available on the website, achieving a similar structure. The WNBA and NBA websites follow a close architecture, allowing for a straightforward pinpointing of all elements needed.

However, spatial shot data is only available on the WNBA website for games from 2017 onward. Thus, to ensure comparability between the two datasets and avoid missing values, only data from 2017 onward has been scrapped. Moreover, only regular season data has been included, aligning with the NBA dataset. The additional features mentioned before (defence rating, player characteristics and final standings) have also been scraped for the WNBA.

3.2.2.1 Process Description

To ensure the quality of the information gathered, the data collection process plan has been carefully designed. Firstly, the source code of the website underwent thorough scrutiny to learn the optimal selection of HTML tags and CSS selectors. This meticulous evaluation, while time-consuming, has been undertaken to ensure the utmost efficiency in the web scraping process.

Henceforth, a comprehensive assessment has been conducted to determine the most appropriate tool for data extraction. In this regard, two primary factors have been taken into account: ethical considerations concerning the use of the website and the tool's degree of flexibility. The "Selenium" Python library distinguishes itself by its ability to emulate human browsing behaviour. Notably, in addition to its proficiency in replicating user engagement, often associated with websites' interactions through JavaScript, its slower performance can be advantageous in terms of ethical considerations. By limiting the number of requests that are made within a specified timeframe, it aids in mitigating the risk of overloading websites, aligning with responsible web scraping practices. On top of that, the collected data has been scrutinized through validation checks to identify data points failing to conform to predefined criteria, for subsequent case-by-case adjustment.

Nonetheless, the data collection process has also encountered certain challenges. Frequently, the website returned errors, notably the 502 error. To address this, a retry mechanism with a predefined maximum number of attempts to access the website has been implemented. If the website remained inaccessible despite these efforts, manual data insertion has been resorted to as an alternative course of action. These adaptive strategies underscore the importance of providing the code with significant flexibility in navigating unforeseen obstacles, reinforcing the robustness of the overall data collection methodology.

For an overview of the final set of variables, A1 should be consulted. Additionally, the code used for the complete construction of the WNBA dataset is available in https://github.com/BLeal2/Shot_Data_WNBA/tree/main.

4 Descriptive Statistics

Descriptive statistics play an important role in data analysis by presenting the data in a meaningful and understandable way, assisting the visualization. They are essential for gaining insights from complex datasets and allow to turn data into knowledge. Without the guidance of statistics, the understanding of trends and patterns in data would be limited and decision-making would be hindered in several industries, including sports management and analytics. By leveraging statistics, valuable insights within sports data can be uncovered and decision-making in sports analytics is enhanced.

4.1 NBA

4.1.1 Player Name

In the course of the research, the data was gathered from a diverse array of 2058 NBA players, each with distinctive contributions to the league. Yet, one figure that consistently emerges as a prominent player is LeBron James, as shown in A2. As referred by Beer (2020), since 2003, LeBron has continuously upheld an exceptional performance, establishing himself as one of the league's most dominant players. His excellence is highlighted by his cumulative shot count, amounting to an impressive number of 28,042 shots. Another key aspect is LeBron's extensive playing time, clocking in numerous minutes on the court throughout his career, leading the league in this category. Moreover, LeBron's esteemed reputation as a prolific scorer significantly elevates the dataset's relevance. His consistent presence among the league's top scorers accentuates the inherent correlation between scoring proficiency and the frequency of shot attempts.

4.1.2 Team Name

The dataset gathers information on 36 different NBA teams. Following A3, the Golden State Warriors hold the record for the team with the largest number of shot attempts with 137,601 shots, followed closely by the Phoenix Suns with 137,201 attempts. This accomplishment can be ascribed to the paced and defensive approach of the Warriors to the game, adding to the presence of talents such as Kevin Durant and Stephen Curry, who are among those with the highest number of shot attempts. This style not only encourages shot attempts from the Warriors themselves but also tends to provoke their opponents into taking more shots thereby increasing overall shot frequency in their games.

Additionally, the fact that the New Orleans/ Oklahoma City Hornets have the least appearances can be attributed to their significant relocations and rebranding. In 2005, the team temporarily became the Oklahoma City/ New Orleans Hornets during a move to Oklahoma City after Hurricane Katrina, as highlighted by Conn (2019). Furthermore, in 2007 they officially returned to New Orleans resulting in another change in the team’s name, followed by a franchise rebrand as the New Orleans Pelicans in 2013. These changes explain why we have various names for the same team in our dataset, each accounting for a small number of shots.

4.1.3 Season

According to Gupta (2023) and as seen in A4 and A5, the season of 2011-2012 witnessed an unprecedented alteration in the NBA landscape. This deviation is primarily attributed to the occurrence of the 2011 NBA lockout, a protracted labour dispute between NBA owners and the National Basketball Players Association. Starting at the beginning of July 2011, the lockout lasted for a period of 161 days, marking one of the longest work stoppages in NBA history. As mentioned by Solomom (2011), the core issue of this dispute was the negotiation of a new Collective Bargaining Agreement, dictating the conditions governing NBA’s operations, including player salaries and salary cap regulations. The delayed resolution of the lockout demanded an adapted approach, forcing to reduce regular season games from 82 to 66.

In contrast, 2020 and 2021 faced game reductions due to the global Covid-19 pandemic. The 2019-20 season grappled with the pandemic’s impact, leading to a temporary suspension of play in March 2020, resuming with a modified format and 72 games. The 2020-21 season started late, leading to significant differences in the recorded number of shots.

As shown in A6, the first quarter of the years under analysis, comprising the period from January to March, exhibits a notable difference in the number of shots compared to the rest of the periods, accounting for almost 2,100,000 shots, explained by several factors. One of these factors is the anticipation of roster changes, since February marks the NBA players' trade deadline. This anticipation of potential changes in team dynamics motivates players to make a significant impact before any adjustments, resulting in a higher number of shot attempts.

Furthermore, the NBA All-Star Game traditionally takes place halfway through the season, more specifically in February, which also leads to a higher number of shots in the first three months of the year. As players want to be chosen for that game, as mentioned by Elkins (2020), during this time they aspire to showcase their skills to their fans, resulting in a higher motivation to perform exceptionally well during this period, increasing shot-taking.

In addition to these factors, the drive for late-season excellence also contributes to the elevated number of shots. As stated by Skinner (2013), players strive to attain their peak form when preparing for the playoffs, naturally leading to an increased number of shots. Nonetheless, January and March are central months in the NBA regular season, coinciding with the mid-season and playoff push. In these months, teams compete fiercely for playoff berths and look for favourable seeding, giving a high importance to scoring. This intensified focus on scoring is also a main reason for the observed increase in shot attempts during these periods.

4.1.4 Event Type

The distribution of made and missed shots through the seasons reveals noteworthy observations. As seen in A8, the number of missed shots is higher than made shots, with missed shots accounting for 54% of the total, while made shots represent 46%.

As it is highlighted by A7, over the seasons, the overall shot attempts in the NBA have seen an increase, indicating an uptrend in offensive play styles. According to what was proven by Scaletta (2020), this can be partly attributed to the evolving pace of play in the NBA since 2002. Many teams have adopted a faster game pace, leading to increased possession and scoring opportunities, resulting in a higher volume of shot attempts per game.

According to A7, the average of shots made has grown by approximately 20%, while missed shots have increased by almost 3%. As shown in the study conducted by Wang and Zheng (2022), there has been a shift towards more three-point attempts and improved field goal accuracy, driven by rule modifications, player skill developments and data-driven coaching strategies adaptations, emphasizing efficient scoring and devaluing mid-range shots.

4.1.5 Action Type

The dataset contains a variety of 70 distinct shot actions. This distribution of action types also mirrors the dynamic nature of the game and is in line with the expected, given the level of shot difficulty. On one side, as shown in A9, the “Putback Reverse Dunk Shot” emerges as the rarest action type, constituting only 10 shots of all the shots captured.

On the other side, the basketball classic “Jump Shot” was found to be the most prevalent action type, comprising almost 2 million shots, corresponding to a high percentage of 48.33% of all shots. The “Layup Shot” takes the place of the second most common action type, also covering a significant portion of the shots.

4.1.6 Shot Type

It is clear that 2-point shots are significantly more common than 3-point shots, which only represent 28% of the total attempts, easily attributed to their difficulty and dependency on uncontrollable game factors, such as player ability and specific game dynamics. As observed in recent years, the landscape of the NBA has undergone a seismic shift in offensive strategy, with pronounced changes in shot distribution. As seen on A10 and A11, the rate of 2-point shots has steadily decreased over the last 18 seasons, falling from an average of 126 attempts per game, constituting 79.8% of all field goal tries inside the arc, to 108 attempts, accounting for 61.3%. Conversely, the rate of 3-point shots has seen a dramatic upsurge over the same period, rising from an average of 32 attempts per game, constituting 20.2% of all field goal attempts beyond the arc, to 68 attempts, accounting for 38.7%.

As stated by Pollard R. and Pollard G. (2005), the impact of the 3-point revolution in the NBA extends beyond shot distribution. With the proliferation of three-pointers league-wide, the task of defending NBA offences has become progressively challenging. New records for league-wide efficiency, measured in points scored per 100 possessions, have been consistently established in each of the last three seasons, as reported by Schuhmann (2021), highlighting the formidable offensive capabilities in the current era.

Following Uiti (2022), this shift in offensive efficiency is a direct reflection of the evolution in playing style within the NBA. Prioritizing efficiency and scoring from beyond the arc has become the cornerstone of contemporary basketball. This strategic shift aligns with the principles of the “Moneyball” approach, which underscores the heightened value of the three-point shot due to its potential for higher point yields.

4.1.7 Zone Name and Zone Range

Notably, as seen in A12, 3-point shots in the left and right corners, backcourt and restricted area have shown a distinct steady evolution over the years, with minor deviations over seasons. However, the quantity of shots in the paint has been slightly growing, consistent with the changing landscape of the league, particularly the adopted defensive adjustments in response to the higher weight of 3-point shooting. Since defenders spread their coverage to the perimeter, NBA players had the opportunity to drive to go inside the paint, resulting in incremental growth in paint shots.

In contrast, Uiti (2022) highlighted that the mid-range shots have dropped over the seasons, aligning with the previously mentioned decreasing trend on 2-point shots. Simultaneously, shots above the break have witnessed a sharp increase, directly related to the increasing trend on 3-point attempts analysed above. This shift reflects a fundamental principle where the value of a 3-pointer per attempt exceeds that of mid-range shots.

Turning now to the analysis of the zone range, as highlighted in A13, over the seasons, it is notable that shots attempted in the backcourt and within the 8-16 ft range from the basket have revealed a constant trend. Additionally, shots made less than 8 ft from the basket consistently revealed the highest number of shots, in line with the low difficulty of this type of shot when compared to others. In contrast, 16-24 ft attempts from the basket have shown a consistent fall over the seasons, in conformity with the decreasing trend observed in both 2-point and mid-range shots. However, a substantial increase in shots made from a distance exceeding 24 ft from the basket, beyond the 3-point line, can be witnessed, mirroring the earlier analysed surge in 3-point attempts.

4.1.8 Quarter

The quarters also provide valuable insights into the unique dynamics, game pace and strategies of the NBA league. Although the majority of the games had 4 quarters, 1,442 games have more than 4 quarters, which is explained due to the existence of overtime periods in the league's basketball games in case of a tie, as ruled by NBA Official (2018b) in Section 5 of Rule No. 5.

Moreover, as shown in A14, the first quarter of the games has the highest average number of shots, making it the one with the highest number of successful and missed shots. On average, this quarter records approximately 20 successful shots and 23 missed shots per game. However, as the game progresses, teams often adjust their defensive strategies to limit the other team's ability to score, which can lead to fewer chances for successful shots in the later quarters.

Across the first four quarters, there is a notable trend where the shot attempts are well distributed among all the minutes in the entire quarter, in contrast to overtime periods, where there is a shift towards an increased number of shots attempted in the final two minutes of the periods. This shift can be attributed to heightened pressure and scoring urgency during overtime, as players become more aggressive, aiming to secure victory or close the gap when trailing. Additionally, clock management plays a crucial role, with teams strategically using more of the shot clock early in overtime periods to limit the opposing team's response time.

4.1.9 Seconds Left

Since NBA basketball games are fast paced, the allocation of shot attempts throughout the game can deeply influence its outcome. As notable in A15, it is clear that the last seconds of the quarter are the ones with the highest number of shots, explained by the opportunity to score buzzer-beater shots. Players and teams often seek to score before the quarter ends, as these shots give a psychological boost to teams and have a demoralizing effect on the opposing team. Teams also aim to maximize their scoring opportunities within the limited time available in the quarter, leading to a rush to take advantage of every opportunity to score before the quarter ends.

4.1.10 Home Team and Away Team

The dynamics of home-court advantage also have a great impact on the overall team performance. On the one hand, the average made shots per game consistently prove to be higher when players are at home, as mirrored by A16, on the other hand, the average missed shots per game are greater when players are not on their home court, as seen in A17. As explained in the article by Woo (2019), this fact is verified mainly due to the home-court advantage, mirrored by a winning rate higher than 50% during the regular season, the opposite of what is verified when teams play away from their home court.

4.1.11 Ranking

In this analysis, a distinct pattern emerges when evaluating the number of missed and made shots in the NBA concerning teams' final season rankings. According to A18, as the rankings of teams improve, reflecting enhanced performance, a trend that the number of missed shots diminishes becomes evident, while the number of made shots increases. This recurring pattern underscores the substantial relevance of teams' final season rankings and their direct correlation with shooting performance, a crucial component influencing a team's overall contribution to its season's success.

4.2 WNBA

4.2.1 Player Name

The data gathers information on 318 players from the WNBA since 2017, and, as notable in A19, with almost 2,400 shots, Tina Charles has the highest number of appearances, with around 8% more shots than Courtney Williams, deserving a in-depth analysis.

Tina Charles's remarkable number of made shots in the WNBA can be attributed to several factors. First and foremost, her significant playing time and central role in her team's offensive schemes have provided her with ample shot opportunities. Secondly, Tina Charles is renowned for her exceptional scoring ability, her versatility allows her to score effectively from various areas on the court which makes her a "defender's nightmare", as it was stated by ESPN (2021).

4.2.2 Team Name

As seen in A20, the Atlanta Dream, with almost 12,200 attempts among all the shots presented in the dataset, stands out as the team with the highest number of shot attempts. Their unique approach to outside shooting, as highlighted by Nusbaum (2021), is an inversion of the typical analytics-driven style. Instead of relying on traditional two-point shots, they prioritize three-pointers, leading to an increase in overall shot attempts. However, this strategy exposes a vulnerability in their defence, as opponents often prioritize guarding the paint.

4.2.3 Event Type

Notably, the total number of missed attempts in the WNBA is visibly higher than the number of successful attempts, with unsuccessful shots making up 56% of all attempts, while made shots constitute 44%.

A21 reveals that, over the seasons, the overall number of shots on average per game, encompassing both made and missed attempts, has generally remained stable. Moreover, a distinctive shift occurred during the 2019 season, where a slight decrease in made shots contrasts with a minor increase in missed shots. This divergence is clearly depicted in A21, emphasizing a parallel trend in both types of shots during this particular period. Specifically, the number of shots made decreased by 23.4%, and the number of missed shots also experienced a 23.7% decrease.

4.2.4 Action Type

When analysing game style in the WNBA, in a similar fashion to the NBA, the diversity of game dynamics and techniques is evident, translated in 49 different shot actions. Following A22, “Jump Shot” is clearly the most common, representing 43% of all the recorded shots. As previously analysed, this is also the preferred kind of attempt in the NBA, given its’ versatility to various game situations and easiness of performing. As emphasized by WABC (2020), “the advantage of a jump shot is that it enables an offensive player both to get additional strength in the shot and to shoot over a taller defender”, which mirrors its versatility and again explains the predominance of this type of attempt in the WNBA dataset. In contrast, the Cutting Dunk Shot, the Harrigan Pullup Jump Shot, the Driving Dunk and the Driving Jump Shot together represent less than 1% of the whole dataset. These attempts demand specific game scenarios and player skills, and, therefore, they are more dependent on how opposing teams adjust their defensive strategies and on the skill set of each team player.

4.2.5 Shot Type

When closely looking at the Shot Type variable, it is evident that 2-point shots are much more frequent than 3-point shots, making up a significant 71% of total attempts. Over the past few years, the WNBA's offensive strategy has undergone a change in how shots are distributed. According to the data shown in A23 and A24, the frequency of 2-point shots has consistently dropped in the last six seasons, going from an average of 102 attempts per game (74.3% of all shots inside the arc) to 92 attempts (now 67.1%). Oppositely, the rate of 3-point shots has significantly increased during the same period, going from an average of 35 attempts per game (25.7% of all shots beyond the arc) to 45 attempts (now 32.9%).

As mentioned by Schindler (2023), the increase in three-point shots in the WNBA can be explained by a few key factors. Firstly, there is a noticeable change in how teams approach offence, giving more importance to playing at a faster pace and initiating early offensive moves. This shift is evident in the higher average possessions per game (pace), indicating teams' preference for quicker plays and taking advantage of fast-scoring opportunities.

Looking back at the playing styles of Katie Smith, Becky Hammon, and Diana Taurasi serves as a reference point to understand how offensive strategies have evolved. These players laid the foundation for the significance of three-pointers in the league. With time, teams have adopted modern concepts like flare screens and motion offences, making use of players' shooting abilities to generate open opportunities for three-point shots.

4.2.6 Zone Name

As visible in A25, the backcourt, right corner, and left corner shots are consistently the least used zones in both leagues, suggesting a commonality in shot selection strategies between WNBA and NBA players. Another compelling observation is the shift in mid-range shots over the years. Back in 2016, players averaged around 37 mid-range shots per game. However, this number has significantly decreased over time, aligning with the broader trend seen in the NBA, where mid-range shots are becoming less prevalent as stated by Uiti (2022). In the most recent season, the average dropped to 26 shots per game in that zone of the court.

Conversely, there is a notable uptick in shots taken from the Above the Break zone, increasing from around 29 average shots per game in 2016 to approximately 35 in the latest season. This also aligns with the overarching trend of an increase in three-point shots over the years in both the WNBA and the NBA.

4.2.7 Quarter

Out of all the WNBA games played since 2016, there are 45 matches that went beyond the standard four quarters, about 4.3% of the total. The reason for this extension is typically overtime periods, which is a rule similar to the NBA to avoid ending games in a tie.

Similarly to the NBA, as observed in A26, the first quarter stands out as the period with the highest average number of shots per game. During this quarter, players tend to have more successful shots on average, around 16.19 per game, but also more missed shots, with 19.93 per game. It is notable that the number of shots is well distributed over the entire quarter. However, in quarters 5 and 6 players attempt more shots in the last two minutes of the quarter.

4.2.8 Seconds Left

As evidenced in A27, and similarly to the NBA, the highest number of shots performed take place in the last seconds of the quarter. Teams usually attempt last-second shots to secure possession of the quarter, as it minimizes the possibility for the opposing team to turn the game or respond. These last quarter seconds are usually when the evolving environment boosts players’ and teams’ psychological motivation, discouraging the opponent.

5 Formulation of Target Variable

Several steps have been taken to correctly explore the *hot hand* in the realm of basketball, with the first one being the formulation of a meticulous and accurate definition of *hot hand*, so that it can be employed as the target variable in this study.

The widely accepted definition of *hot hand*, prevalent across numerous research papers, is rooted in the belief that, during a specific time frame, a player's performance notably exceeds the expectations based on his record. This concept, as initially posited by Gilovich, Vallone, and Tversky (1985), discussed in the Literature Review section, essentially suggests that a player's *heat* percentage reflects the proportion of consecutive shots they make within a specified set of *n* shots (*n* represents the count of consecutive shots preceding the current shot under evaluation for its *hot* status). Alternatively, a more mathematical interpretation, as proposed by Bocskocsky, Ezekowitz, and Stein (2014), can be used to describe the heat, where *i* represents each individual shot.

$$\textit{(1)} \qquad Simple\ Heat_i = Actual\ \%_{of}\ success\ over\ past_n\ shots\ = \frac{shots\ hit}{n} \times 100$$

It is crucial to emphasize that this analysis exclusively considers consecutive shots within the same game. Given that this phenomenon is classified as rare, the choice of the time frame for event prediction becomes highly significant. Striking a balance is essential to avoid defining an excessively narrow or overly broad time frame, as suggested by Weiss and Hirsh (2000). Many previous studies, including those by Bocskocsky, Ezekowitz, and Stein (2014) and Arkes (2013), have employed the entire game as the designated time frame, building an analysis upon the assumption that the *hot hand* effect does not extend across different games, but it may manifest within the confines of time-outs and breaks between quarters. Taking that into account, the same approach is followed in this study.

However, the commonly accepted definition of Simple Heat mentioned above faces the limitation of failing to consider how challenging the shots are, as identified by Bocskocsky, Ezekowitz, and Stein (2014). Accounting for the level of challenge players undertake when executing a shot is essential for defining the heat variable. This is illustrated by the following scenario, inspired by the one present in the study conducted by Bocskocsky, Ezekowitz, and Stein (2014), where the 5 past consecutive shots (*n* = 5), taken by two players, are examined, determining the 6th shot heat. A28 provides a clearer perspective on this example.

1. Player A takes 5 shots and successfully hits all of them. Every shot attempted by Player A is considered easy. For that reason, based on this player's record over the past 5 shots, it is expected that none of those shots are missed. According to the traditional definition of *hot hand*, Player A's Simple Heat value of his 6th shot is 100% (five out of five shot attempts succeed).

$$Player\ A_{\ Simple\ Heat_6} = \frac{5}{5} \times 100 = 100\% \tag{2}$$

2. In contrast, Player B also takes five shots but only hits three of them. The key distinction here lies in the fact that all shots made by Player B are exceptionally challenging to execute, resulting in a relatively high probability of missing one or more of them. Therefore, based on this player's record over the past 5 shots, it is expected that some are missed. Player B's calculated Simple Heat of his 6th shot is 60%, which is lower than that of Player A.

$$Player\ B_{\ Simple\ Heat_6} = \frac{3}{5} \times 100 = 60\% \tag{3}$$

This example distinctly illustrates a fundamental flaw in the conventional approach to identifying the *hot hand*. It is illogical why Player A, who had a higher chance of success, has a higher heat score than Player B, who took on more risk by attempting harder shots and, in spite of that, hit three out of five. This discrepancy highlights the need for a more nuanced definition of the *hot hand* phenomenon that addresses these cases.

To address this flaw, an adaptation of the concept of Complex Heat used by Bocskocsky, Ezekowitz, and Stein (2014) in their research has been chosen for this study:

$$(4) \qquad Complex\ Heat_i = Actual\ \%_{of}\ SU_{over}\ \ past_n\ shots - Expected\ \%_{of}\ SU_{over}\ past_n\ shots$$

Following this approach, where "SU" stands for success, the "Expected percentage of success over the past n shots" has been derived as an average of the success rate for each shot, considering the available historical data. Success Rate has been computed by grouping the shot data by player, season, shot type and distance. Subsequently, the ratio of successful shots to the total number of shots within each respective group has been calculated, resulting in the percentage of success expected in each situation.

This definition categorizes players individually, as each player possesses a unique playing style and a distinct approach for every specific shot type they execute. The goal is to evaluate a player's likelihood of making a shot based on the normal success rate for shots of similar challenge. The expected percentage has been calculated on a season-by-season basis, since player performance and game strategies evolve through distinct seasons, creating the need for an individual analysis for each one. In other words, success rate values for each shot have been determined within the context of the entire season in which it took place. Therefore, the existing data from games that occurred after each specific shot, within the same season, has also been used for the calculation. This approach aims to draw meaningful values and conclusions from real occurrences that took place in court during matches, providing a holistic overview.

Mathematically expressed, where *i* represents the shot number and "SR" stands for success rate, the equation of the expected % of successes over past n shots is presented:

$$(5) \qquad Expected\ \%\ of\ success\ over\ past\ n\ shots = \frac{\sum_{i=0}^{n} SR}{n}$$

Using the Complex Heat definition, the heat of the previous example can be calculated more accurately, considering $n = 5$. For a clearer perspective, A28 can be consulted.

1. Revisiting Player A, who initially held 100% of Simple Heat on his 6^{th} shot, has now reduced this value to 20%, according to the Complex Heat equation. Under this approach, the expected percentage over the past five shots amounts to 80%. A higher

expected success rate for each shot implies that Player A faced relatively easy shots, and it is within the realm of expectation to successfully make those shots.

$$(6) \qquad PlayerA_{Complex\ Heat_6} = \left(\frac{5}{5} - \frac{(72\% + 64\% + 84\% + 94\% + 87\%)}{5}\right) \times 100 = 20\%$$

2. Shifting the focus to Player B, who initially reached an 80% value through the Simple Heat equation on his 6th shot, the revised heat now registers at 60. A low expected success rate for each shot suggests that Player B attempted highly challenging shots, being within the expected range for some of these shots to be missed.

$$(7) \qquad PlayersB_{Complex\ Heat_6} = \left(\frac{3}{5} - \frac{(37\% + 8\% + 25\% + 20\% + 10\%)}{5}\right) \times 100 = 60\%$$

Taking this into account, the variable *heat* calculated from the Complex Heat equation has been incorporated into the dataset, encompassing both negative and positive values, including zero. It is essential to note that the calculation of the heat for each shot involves considering the past *n* shots, resulting in heat values consistently equal to 0 for the initial *n* shots for all players in every game. Following the construction of this variable, a sensitive analysis has been executed in order to define both the values of heat in which shots are considered *hot shots*, and the *n* shots chosen to be analysed in retrospect for each shot.

5.1 Sensitivity Analysis

The primary objective of this analysis is twofold, aiming to determine the optimal value of *n* and subsequently define the appropriate heat threshold for categorizing a *hot shot*. In this study, a *hot shot* is indicated by assigning a value of 1 when the heat exceeds a specified threshold, otherwise, it is assigned a value of 0.

Firstly, a computational loop has been employed to test values of *n* ranging from 3 to 7 (inclusive). For each *n*, two criteria have been explored to define the threshold for a *hot shot*:

1. Quantity of *hot shots* if the threshold is above the value of the third quartile (q3) of the heat variable.

2. Quantity of *hot shots* if the threshold is above heat values of 0.

Considering that a *hot shot* should be an infrequent occurrence and given that the average number of shots per player per game is around 8, the results presented on A29 highlight that *n* = 7 and *n* = 6 emerge as the preeminent choices. However, with *n* = 7, the threshold remains notably large, prompting the need to reduce this range. Recognizing the need for a reduced threshold, *n* = 6 has been singled out as the most fitting option.

Nevertheless, at this point, the percentage of *hot shots* continues to be considerably high. Consequently, another sensitivity analysis has been conducted, where *n* is predefined (*n* = 6), and thresholds for the heat variable, above which a shot is considered to be *hot*, are explored within the range from 0.75 to 1.0 with increments of 0.01. To enhance the precision of the analysis and introduce an additional metric, a new column, *hot_hand*, has been generated. In this column, a value of 1 denotes the first *hot shot* in a player's hot streak. Essentially, this column exclusively captures the initial *hot shot* executed by each player in a game following a non-*hot shot*, denoting a *hot hand*, while 0 signifies the opposite. If a player makes a *hot shot* in the same game immediately after another *hot shot*, it is marked as 0.

A30 illustrates the outcomes derived from the second sensitivity analysis. Given the rarity of these events and the necessity for statistically meaningful results to enable a substantial analysis, it is deemed optimal to aim for an average of one *hot hand* and two *hot shots* per game. As depicted in A31, selecting the quantile 0.97 emerges as the most suitable choice to align with the previously mentioned optimal values. Opting for the quantile 0.97 means that shots with values of heat above 0.38 are considered *hot*.

Finally, taking everything into account, the Y variable has been defined, with a value of 1 indicating that a shot is considered *hot* (all heat values exceeding the previously mentioned threshold), while a value of 0 denotes a shot as non-*hot*. With this target variable established, it is possible to conduct an in-depth analysis of several hypotheses and predictions.

6 Hypothesis Testing

6.1 The Role of Player Expertise in the Manifestation of the *Hot Hand* in the National Basketball Association

Building upon the extensive groundwork laid by the literature review, this study embarks on a rigorous empirical investigation to elucidate the intricate relationship between player experience, age, and the occurrences of the *hot hand* phenomenon in the realm of the National Basketball Association. The theoretical underpinnings articulated by Pelechrinis and Winston (2022), Bakkenbüll (2017), Zhang et al. (2018), and Kalén et al. (2021), as previously discussed, have provided a robust foundation, delineating the crucial roles played by cumulative experience and age in shaping players' overall performance. Motivated by these insights, this research seeks to further unveil the nuanced relationships that underlie these dynamics.

6.1.1 Influence of Player Experience and Age on the Occurrences of *Hot hand*

To investigate the relationships among these variables, a comprehensive methodological framework has been formulated. Preliminary exploratory analysis laid the foundation of the study, offering initial insights into the patterns of the data. Subsequently, more sophisticated statistical techniques, including different regression analysis, have been employed to unravel the sophisticated relationships within the data.

6.1.1.1 Exploratory Analysis

Before all, to ensure the subsequent examination is grounded in statistically significant data, one must consider the representativity of distinct years of experience and ages.

The granularity of the dataset, which captures shot by shot data in NBA games, presents a unique set of considerations for the intended analysis. Although it affords a comprehensive understanding of player performance, this level of detail brings forth a potential bias when examining the frequency of occurrences across various experience levels and ages. Specifically, having only a few players with a certain level of experience, even if they shoot frequently, may not yield a representative sample. The inherent variability among these few players may not be reflective of the broader population, introducing a risk of drawing conclusions based on outliers. Hence, to mitigate this risk, only unique combinations of player ID with experience and player ID with age have been considered for this analysis.

The histogram depicted in A32 reveals that, with regard to player experience, values predominantly cluster along the lower edge of the graph, resulting in an average experience calculated at 4.61 years. This observation suggests a tendency towards relatively short-lasting careers for most NBA players. On the other hand, the distribution of age is more dispersed, reflecting an average of 26.29 years.

Following these results and taking into account the rarity of the phenomenon under investigation, the imposition of a minimum threshold of unique players within each specified experience and age levels was required, resulting in the stipulation of a minimum of 100 players at each level. Consequently, the upper bounds for experience and age levels considered in this study were capped at 14 (A33) and 36 years (A34), respectively. The followed strategy ensured that any observed patterns were not unduly influenced by the sporadic occurrence of *hot hand* momentum within limited subsets of the dataset, and guarantees the generalizability and credibility of the findings.

Furthermore, in agreement with the NBA rule implemented since 2005, requiring that players participating in the NBA Draft should be a minimum of 19 years old at the year of the draft and have completed at least one NBA season after high school graduation (Vorkunov and Moore, 2019), a lower bound for age was also established at 19 years old.

In light of these considerations, the groundwork has been laid for a more comprehensive examination of the *hot hand* phenomenon and the hypothesized factors influencing it. The chosen approach for the exploration of initial insights involves an analysis that centres on the percentage of hot shots over overall shots for each experience and age level. These derived probabilities serve as crucial indicators, offering insights into the likelihood of experiencing a *hot hand* at different stages of a player's career.

In Figure 1, an observation of the distribution of the percentage of *hot shots* in relation to overall shots reveals comparable patterns for player experience and age levels. Both exhibit a characteristic U-inverted shape, starting with an initial ascent, reaching a peak, and subsequently declining. Bakkenbüll (2017) elucidates this shape by positing that, until a certain age, the synergy between physical form and the expertise derived from playing practice leads to enhanced performance. However, beyond this point, a marked decline in physical ability ensues, bearing diminished performance irrespective of player experience. This same pattern appears to hold true specifically within the context of the manifestation of the *hot hand* phenomenon.

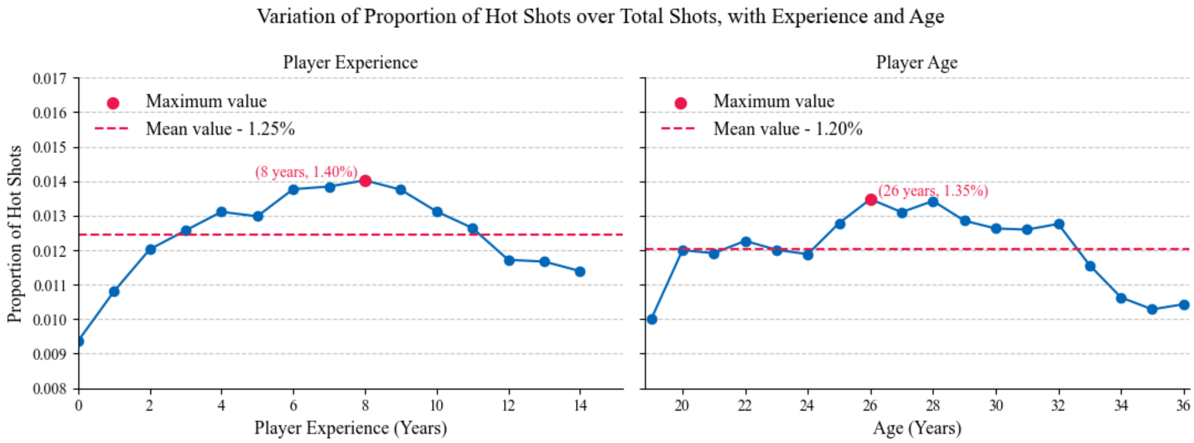


Figure 1 - Average Proportion of Hot Shots over Total Shots by Experience and Age Level.

Additionally, apparent in the graphical representation is a suggestive pattern indicating that players tend to peak in performance at approximately 8 years of experience and between the ages of 26 to 28. This observation aligns with the insights offered by Kalén et al. (2021), who notes a peak in efficiency within the 26–29 age group, with a notable surge from the 23–25 to the 26–29 age brackets. This observed convergence also resonates with the findings outlined by Bakkenbüll (2017), suggesting that player efficiency reaches its zenith around the age of 28.

Following these results, the proposed hypothesis posits that the relationships between both experience and age with the probability of a *hot shot* are non-linear, specifically exhibiting a quadratic pattern. The subsequent phase of the research aims to rigorously examine the validity of these hypotheses, probing the strength and significance of these relationships.

6.1.1.2 Regression Analysis

To disentangle these relationship dynamics, regression analysis serves as the primary statistical tool. This technique allows for a systematic examination of the hypothesized non-linear relationships, facilitating the identification of patterns and trends within the data.

The fundamental objective is to ascertain the substantive impact of player experience and age on the discernment of *hot shot* occurrences, incorporating the entire spectrum of observations. In the pursuit of this objective, characterized by a classification problem essential to determining whether a shot is made during a *hot hand* state, logistic regression emerges as the method of choice. This approach is underpinned by its appropriateness for handling the binary nature of the target variable, capturing the essence of the analysis, and ensuring the maintenance of the probabilities of the event at levels between 0 and 1.

For this purpose, two sets of logistic regression models were crafted, the first exploring the influence of player experience and the other examining the impact of player age (Table 1). Each set comprises two variations, one with linear terms and the other incorporating quadratic elements, to capture the hypothesized non-linear nuances within the data.

Table 1 - Logistic Regression Results with Hot Shot, Experience, Age, and the Squared Terms of Experience and Age.

Logistic Regression				
	Dependent variable: *hot shot*			
	(1)	(2)	(3)	(4)
constant	-4.461***	-4.612***	-4.406***	-6.615***
	(0.008)	(0.011)	(0.031)	(0.198)
experience	0.017***	0.096***		
	(0.001)	(0.004)		
experience squared		-0.007***		
		(0.000)		
age			0.001	0.169***
			(0.001)	(0.015)
age squared				-0.003***

				(0.000)
Observations:	3,938,116	3,938,116	3,966,290	3,966,290
Note:				*p<0.1; **p<0.05; ***p<0.01 *Standard errors in parentheses*

In the context of Model 1 from Table 1, where only linear terms are considered, the positive coefficient associated with experience (0.017, p<.01) implies a proportional increase in the log-odds of a *hot shot* with each unit rise in experience. This positive relationship underscores the notion that heightened player experience corresponds to an increased likelihood of a *hot hand* occurrence.

Model 2 presents a quadratic term for experience. The negative coefficient for the quadratic term (-0.007) introduced curvature to the relationship between experience and log-odds, indicating that, although initially positive, the effect of experience gradually diminishes. Both terms of the model exhibited relatively small standard errors. The p-values (p<.01) associated with each coefficient were consistently statistically significant.

Regarding age, the logistic model exclusively employing linear terms to assess its impact on the log-odds of a *hot shot* (Model 3) reveals a positive coefficient. It is however important to note that the p-value (p>.1) indicates this coefficient is non-significant, which suggests an elevated level of uncertainty in the linear relationship between the two variables. Nevertheless, when the quadratic term is introduced (Model 4*)*, the results unveil a more complex and less uncertain relationship. While the positive coefficient for the linear term persists, with a p-value lower than 0.01, the negative coefficient for the quadratic term (p<.01) advocates for the hypothesized concave relationship between age and the log-odds.

Henceforth, the persistent statistical significance of the quadratic model accentuates the reliability of the identified non-linear relationship between these two variables and the probability of hot shots.

However, although these findings reinforce the belief in the hypothesis, one must consider the delicate balance between model fit and complexity. As outlined by Burnham and Anderson (2004), a comparative analysis of Akaike Information Criterion (AIC) and Bayesian

Information Criterion (BIC) may facilitate the identification of models that not only adequately fit the observed data but also provide generalized insights for future applications of the inferred patterns. The resulting values are listed in Table 2.

Table 2 - AIC & BIC Results of Logistic Regression Models.

AIC and BIC Measures			
	Model	AIC	BIC
Experience	Regression with linear terms	527,823	527,850
	Regression with linear and quadratic terms	527,452	527,491
Age	Regression with linear terms	532,238	532,264
	Regression with linear and quadratic terms	532,109	532,148

In both cases, the models incorporating both linear and quadratic terms demonstrate finer performance compared to their counterparts. This is evident by the lower AIC and BIC values associated with the models, indicating that the inclusion of non-linear components through quadratic terms significantly enhances the models' explanatory power regarding the occurrence of *hot shots*. Notably, this improvement is achieved while also maintaining a balanced level of complexity.

The equations of the best regression models found earlier are the following:

$$logit(p_{hot\ shot}) = -4.612 + 0.096 \times experience + (-0.007) \times experience^2 \tag{8}$$

$$logit(p_{hot\ shot}) = -6.615 + 0.169 \times age + (-0.003) \times age^2 \tag{9}$$

Moreover, the incorporation of time-fixed effects, such as season dummy variables, within the logistic regression framework serves as a refinement in the analysis of the relationship between player experience and age and the occurrence of *hot hand* in professional basketball. By capturing the unique variations associated with individual seasons, this approach may account for external factors that fluctuate across time, including rule modifications, team compositions, and dynamic game conditions. Therefore, the inclusion of season fixed-effects aptly addresses the potential influence of unobserved time-dependent factors and mitigates omitted variable bias.

The results of the models incorporating season-fixed effects can be found in A35 and can be characterized as follows: The coefficient for experience (0.096) and the squared term of experience (-0.006) remains positive and negative, respectively, while statistically significant ($p<.01$), aligning with the findings of the first models. Remarkably, the inclusion of season dummy variables reveals varying impacts across different seasons, as evidenced by the coefficients associated with each season. The positive coefficient for season 2005 (0.052, $p<.1$) implies an increased likelihood of *hot shots* during that season compared to the reference season (2004). Conversely, the negative coefficients for subsequent seasons, including season 2012 (-0.11, $p<.01$) and season 2015 (-0.093, $p<.01$), suggest decreased probabilities of *hot hand*s during those periods, indicative of temporal variations in player performance influenced by factors beyond experience. Similar patterns are found for the model incorporating season-specific effects and age, with age coefficients remaining akin to the ones stated in the previous equations, and season 2005, 2012, and 2015 having identical behaviours to the equivalent model with experience.

Additionally, the model enriched with time-fixed effects demonstrates a lower AIC of 527,399 compared to the previous quadratic model which relates experience and *hot hand* occurrences, indicating an improved fit to the observed data. However, the BIC for the time-fixed effects model is higher at 527,689, as this measure is more stringent in penalizing models for complexity by incorporating a larger penalty for additional parameters. For the models relating to age, the same pattern holds, with AIC (532,059) at a lower value and BIC (532,349) at a lower value. Future research should consider this cautious trade-off between model fit and complexity when considering the inclusion of these variables.

Nonetheless, the stability of coefficients for age and age-squared and experience and experience-squared across these two models suggests that the impact of age and experience on the relationship with *hot hand* occurrences remains relatively consistent. Hence, the model serves as indication that the observed effects are robust and not significantly influenced by the temporal dynamics introduced by season-fixed effects.

Concluding, the comprehensive analysis undertaken in this segment has provided compelling evidence supporting the hypothesized relationships between experience, age, and the occurrences of *hot hand*, shedding light on the interplay between an athlete's accumulated experience, advancing age, and the manifestation of *hot hand* momentum.

6.1.2 Shot Accuracy in *Hot Hand* States at Different Experience and Age Levels

Following an in-depth analysis of the influence of player expertise on the occurrences of the *hot hand*, this smaller section seeks to further understand if the pattern studied before, as highly correlated with performance, holds true for shot accuracy during *hot hand* momentum. Accordingly, to refine the analysis focus, the scope of the sample data for this section has been confined to instances of shots made during *hot hand* states.

6.1.2.1 Exploratory Analysis

To understand the dynamics of the relationships between these variables, a comparable graphical representation, akin to the visual elucidation that bestowed insights in the previous section, has been depicted in Figure 2. The figure shows the proportion of made shots over the total shots attempted during *hot hand* states by experience and age levels.

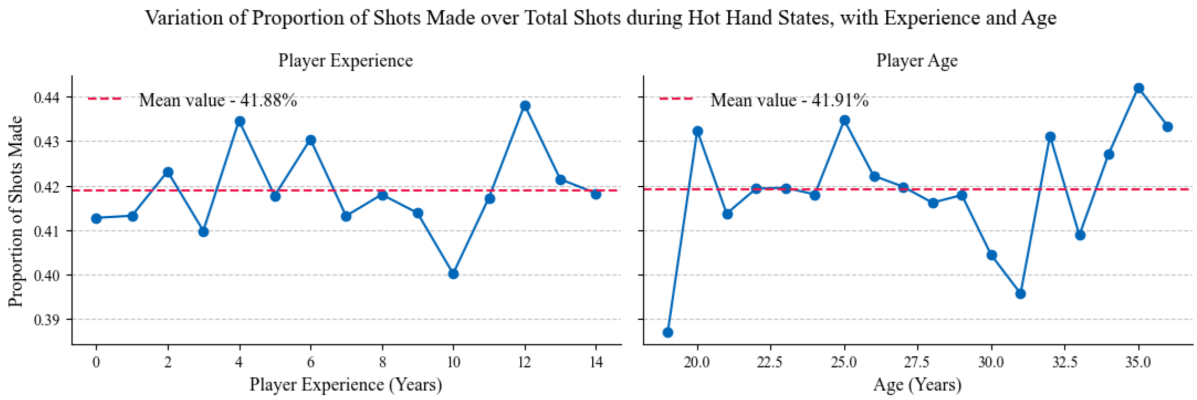


Figure 2 - Average Proportion of Shots Made during Hot Hand States over Total Shots by Experience and Age Level.

Intriguingly, as discerned from both graphical representations, an initial discernible pattern remains elusive. The proposed hypothesis consequently aligns with the belief in the absence of a substantive linear or concave relationship between these variables. This specific analysis proves instrumental primarily in delineating divergent aspects of the influence of age and experience across varied performance metrics, thereby providing caution when inferring patterns of some measures of performance to overall performance.

6.1.2.2 Regression Analysis

To scrutinize this assessment, four logistic regressions, encompassing linear terms, alongside quadratic terms, were conducted. The outcomes are outlined in Table 3.

Table 3 - Logistic Regression Results with Shot Made, Experience, Age, and the Squared Terms of Experience and Age.

Logistic Regression				
	Dependent variable: *shot made*			
	(1)	(2)	(3)	(4)
constant	-0.329***	-0.341***	-0.290***	-0.259
	(0.017)	(0.023)	(0.065)	(0.398)
experience	0.000	0.006		
	(0.003)	(0.009)		
experience squared		-0.000		
		(0.001)		
age			-0.001	-0.004
			(0.002	(0.030)
age squared				0.000
				(0.001)
Observations:	49,084	49,084	49,486	49,486
Note:				*p<0.1; **p<0.05; ***p<0.01
				Standard errors in parentheses

The tabulated results reveal a consistent non-significance in all coefficients (p>.1), excluding the constant, across each model iteration. This collective non-significance rigorously interrogates the reasonableness of a meaningful linear or quadratic association between experience, age, and shot accuracy during *hot hand* occurrences. However, it is advisable upon future research to adopt more complex models, thereby enhancing the depth and complexity this relationship may require.

6.1.3 Exploring Patterns in Shooting Behaviour During *Hot Hand* States Across Different Experience Levels

Subsequent to a comprehensive study discerning the influence of both experience and age on facets of the *hot hand* phenomenon intricately tied to performance, this section seeks to delve deeper into the multifaceted aspects surrounding the shooting behaviour of players during *hot hand* states, discerning how players with varying ages and experiences may respond when in a *hot hand* state. As such, the data under consideration for this section has also been limited to instances of shots made during *hot hand* states.

The methodological approach employed in this analysis matches the preceding study. Opening with a preliminary exploratory analysis, which sets the foundations by providing initial insights into the patterns of the data, and employing similar statistical techniques to prove the strength of the hypothesized behaviours, such as regression analysis.

6.1.3.1 Exploratory Analysis

Building upon the previously elucidated concave relationship between experience and the probability of *hot hand* occurrence, a phenomenon posited by Bakkenbüll (2017) as the trade-off between experience and diminishing physical abilities, it becomes conceivable that players with heightened experience and advanced age may pivot towards adopting an approach that emphasizes strategic finesse over sheer speed. Notably, Allen and Hopkins (2015) research further supports the idea that professional players remain accumulating game intelligence up until the age of 60 years old.

Hence, for the purpose of examining the proposed shooting behaviour, the research concentrates on some key variables, instrumental in characterizing a shot. These variables include shot distance, which provides insights into the spatial dynamics of shooting preferences, and action type, which includes a range of shooting techniques.

Below, in Figure 3, the analysis of average shot distance for shots made during *hot hand* momentum for each player experience level reveals visible trends. As experience increases,

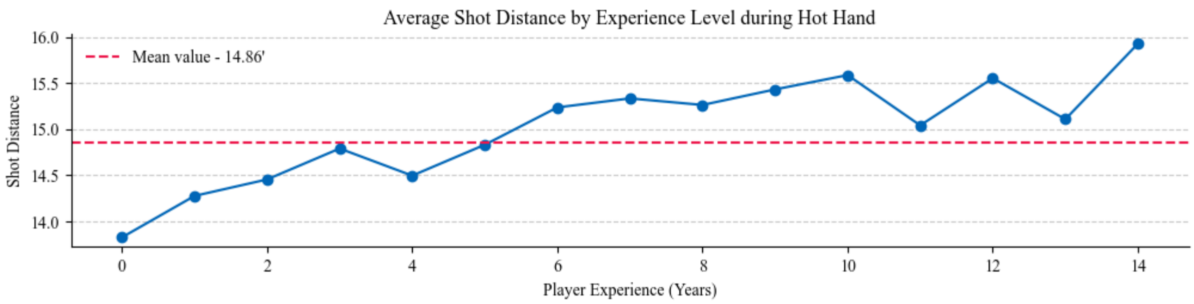


Figure 3 - Average Shot Distance of Shots Attempted during Hot Hand States by Experience Level.

there appears to be a gradual upward trajectory in the average shot distance. This observation suggests a potential correlation between a player's experience and their willingness or ability to attempt shots from greater distances. If proved significant in the subsequent regression analysis, this behavior may indicate a strategic evolution in shooting behavior, with more experienced players possibly demonstrating a higher level of confidence in executing shots from extended ranges.

For the assessment of shooting behaviour concerning action types, it is worthy to mention that the action types under consideration are the ones which were strategically grouped into five distinct categories during the data cleaning process. The categorization comprised "Dunk Shot", "Jump Shot", "Layup Shot", "Bank Shot", and "Hook Shot", facilitating a streamlined analysis for the sake of clarity and interpretability.

The findings of the preliminary analysis, considering the variation of the representativity of each type of shot over the total shots made in *hot hand* with experience, can be fully reviewed in A36. Nonetheless, the results of the preliminary insights drawn from the visualization are explained as follows:

Remarkably, action types such as "Bank Shot" and "Hook Shot" exhibit less discernible trends, which might suggest relative insensitivity to fluctuations in player attributes such as speed, agility, or game intelligence. However, the lower frequency of occurrence of this type of shots requires a cautious approach when drawing conclusions, as minor fluctuations can disproportionately impact the observed patterns.

By opposite, an intriguing divergence emerges in the proportions of “Dunk Shot” and “Layup Shot” relative to total shots, showcasing a diminishing prevalence with increasing player experience. The action type “Jump Shot” demonstrates an opposing trajectory, indicating an augmentation in its prevalence as player experience advances.

These initial findings lay a solid foundation for a more rigorous assessment of the statistical significance of observed patterns, elucidating the quantitative relationships and contributing to a deeper understanding of the dynamics at play.

6.1.3.2 Regression Analysis

The method of choice for scrutinizing the relationship between the distance of attempted shots and the level of player experience is linear regression. This statistical technique seeks to model the linear association between these two variables. The resulting parameters estimation will discern the nature and magnitude of the relationship between these two variables, together with the statistical significance of the proposed association. The findings of the linear regression are detailed in Table 4 and Equation (**10)**.

Table 4 – Linear Regression Results with Shot Distance and Experience.

	Linear Regression
	Dependent variable*: shot distance*
	(1)
constant	14.199***
	(0.081)
experience	0.126***
	(0.013)
Observations:	49,084
R^2	0.002
Adjusted R^2	0.002
Residual Std. Error	9.895 (df=49,082)
F Statistic	95.730*** (df=1; 49,082)
Note:	*p<0.1; **p<0.05; ***p<0.01
	Standard errors in parentheses

$$logit(p_{hot\ shot}) = 14.199 + 0.126 \times shot\ distance \quad (10)$$

The estimated coefficients reveal key insights into the relationship under scrutiny. The intercept, representing the expected average shot distance when player experience is zero, was found to be 14.2. The player experience coefficient is 0.126, suggesting that, on average, there is a positive increase of 0.126 units (feet) in shot distance for each year contributing to player experience. Both coefficients are statistically significant, with p-values lower than 0.01, which strengthens the validity of the previously observed patterns. These findings support the hypothesis drawn earlier – The more experienced the player, the higher the average distance of shots attempted during *hot hand*.

For the assessment of the relationship between player experience and action types during *hot hand* states, a multinomial logistic regression was chosen as the relevant analytical framework. This type of regression is particularly suitable when dealing with a categorical dependent variable with multiple categories, as it allows for the modeling of the probability of each category (in a log-odds scale) relative to a reference category. The reference category under consideration is "Jump Shot", based on two considerations: firstly, "Jump Shot" is the most common category, assumed to be more stable, and secondly, it holds practical significance as a baseline for interpretation. Therefore, by choosing "Jump Shot" as reference, we aim to understand how increasing experience deviates from choosing this category over the others. Results can be found in Table 5.

Table 5 - Multinomial Logistic Regression Results with Action Type and Experience.

Multinomial Logistic Regression				
	Dependent variable*: action type*			
	(1)			
	Dunk Shot	Layup Shot	Hook Shot	Bank Shot
constant	-2.824***	-1.132***	-3.578***	-3.77***
	(0.045)	(0.020)	(0.059)	(0.066)
experience	-0.064***	-0.039***	-0.012	-0.0157
	(0.008)	(0.003)	(0.009)	(0.011)
Observations:	49,083			
Note:				*p<0.1; **p<0.05; ***p<0.01 *Standard errors in parentheses*

Upon examining the results, distinctive insights surface for “Dunk Shots” and “Layup Shots”, as both display particularly significant coefficients ($p<.01$). The constant in the context of each category signifies the log-odds of opting for any of these shots over a “Jump Shot” when experience registers at zero. Furthermore, the associated experience coefficients unravel noteworthy trends, showcasing a decrease of 0.064 in the log-odds of favoring a “Dunk Shot” over a “Jump Shot” for each unit increase in player experience. Similarly, for “Layup Shots”, the coefficient unveils an akin pattern but at a diminishing pace of 0.039. These revelations underscore a decline in the inclination towards both “Dunk Shots” and “Layup Shots” with the increase of experience.

Both categories “Hook Shot” and “Bank Shot” display coefficients for experience with associated p-values exceeding 0.1. Consequently, recognizing a discernible pattern within these latter categories becomes propitious to error.

The revealed trends in the multinomial logistic regression mostly align with the hypothesized relationships discussed previously. “Hook Shot” and “Bank Shot” indeed exhibited a less pronounced trend, as proved by their coefficients and p-values, supporting the notion of relative insensitivity to player’s increased knowledge, and weakened physical ability. Additionally, the diminishing prevalence of “Dunk Shot” and “Layup Shot” with increasing player experience confirms the divergences in shooting behavior.

6.1.4 Results Overview

The extensive analysis regarding the volatility of different aspects of the *hot hand* phenomenon, with changes in experience and age levels, has been proved significant.

This study substantiates the presence of a concave relationship between player experience, age, and occurrences of the *hot hand*. This finding aligns harmoniously with the patterns and conclusions documented in existing literature, particularly in the context of the balance between accumulated knowledge and the decline in physical capabilities as a consequence of ageing.

Additionally, despite a lack of significance in the research of variation of shot accuracy during *hot hand* states with regards to experience, this analysis allows to understand the implications of extrapolating patterns of specific aspects of performance to overall performance, leading to erratic conclusions.

Regarding the examination of shooting behavior patterns, heightened player experience was found to be correlated with both an increased average shot distance and different shooting techniques during *hot hand* states, specifically, the preference for “Jump Shots” in detriment of “Dunk Shots” and “Layup Shots” as experience advances.

Cumulatively, these findings denote nuanced sensitivities of *hot hand* factors to shifts in experience and age, contributing not only to the understanding of the hot hand phenomenon but also to a broader discourse on the intricacies of performance dynamics in the dynamic realm of professional basketball.

7 Managerial Implications

Basketball games are known for their highly dynamic nature and fast-paced rhythm, with a constant interplay between offense and defence. The fluidity of movements in the court, combined with the several number of variables at play, such as the roster of players, fatigue, and inequality of team strength, makes the task of unveiling game patterns quite challenging.

Over the last decades, the incorporation of data analytics into basketball has resulted into a major revolution in how games are played. Analytics departments from NBA and WNBA teams currently stand at the forefront of this shift, being deeply relied on to aid the corresponding teams achieve victory and, ultimately, become the champion. Despite these significant advancements, the *hot hand* phenomenon remains mostly undecipherable. Thus, teams are unable to fully leverage it to excel in games.

This study acknowledges the relevance of the *hot hand* phenomenon, as it can entirely change the course of basketball games. In light of that, an attempt to deconstruct the several layers of complexity related to the *hot hand* is performed, with the goal of determining the factors that impact it the most. Subsequently, an algorithm is built to predict occurrences of *hot hand* shots.

The results acquired in this study can be employed for designing data-driven strategies in the future, empowering teams to tailor their on-court strategies and off-court training to take advantage of the *hot hand*. Each of the sections of this study unveils relevant patterns related to this phenomenon that result in several implications to players, coaches and even to the fans.

The observed quarterly trends indicate that the likelihood of a *hot shot* continuously increase until the last quarter, with the most substantial rise in the transition between the first two quarters, followed by progressively smaller increments in the remaining transitions. It is found that, as the game proceeds, players are able to discern the *hot* player and consequently provide him with more shot opportunities. This behaviour suggests that, despite the fast-paced nature of basketball games, an NBA player is able to focus on executing his own plays while also being aware of the individual performance of his teammates. Nevertheless, it is possible that not all players can balance these two tasks, and that the ones that can often struggle with it.

Holding an external perspective of the game, not being directly involved in court events, coaches can gather insights and aid in this matter by employing the findings of this study regarding quarterly dynamics to ensure a capitalization of *hot hand* moments. Knowing that the largest increase in the odds of a *hot shot* occurs within the first half of the game, coaches should be the most attentive during this time, as it is when the most evident shifts in game dynamics might take place. After evaluating the players' behaviour from an external perspective, coaches should communicate the gathered insights to them during halftime or timeout, directing their focus towards specific players whose performance might signal the manifestation of a *hot hand*. Applying this approach, players can focus on specific teammates rather than the whole team, being more likely to identify more readily the *hot* player, feeding him the ball.

Moreover, the uncovered patterns related to the occurrence of *hot hand* shots in home-court games can also be leveraged for devising an appropriate game strategy, which would place more emphasis on defence or offence depending on the game location.

Mindful of the fact that the home-court advantage in *hot shots* is mostly materialized in terms of accuracy, coaches and players belonging to the away-court team should be attentive to the occurrence of *hot hand* streaks in the home-court team, with the goal of applying strategies to decrease shot accuracy. In those cases, a greater emphasis should be placed on increasing the defensive pressure on the *hot* player. This approach would entail simultaneously hindering the *hot* player from shooting and preventing his teammates to pass him the ball, disrupting the opponent's offense.

Indeed, a comparable strategy can be applied by the home-court team, but with an emphasis on the team itself as opposed to the opponent. In other words, the team playing at home should be more vigilant of players with a *hot hand* within the team itself. In the case of such an event, the remaining teammates should focus on feeding the ball to the *hot* player, in addition to preventing the opposing team from placing more defensive pressure on him.

Regarding the impact of player experience**,** the identification of a concave relationship between player expertise and the frequency of *hot hand* occurrences yields profound insights for coaching strategies and managerial decision-making. By providing an extensive comprehension of the delicate balance between accumulated knowledge and the gradual erosion of physical attributes over time, this research empowers coaches to navigate the landscape of player development with precision, facilitating tailored adjustments in strategic planning.

Crucially, the findings carry substantial ramifications concerning contract lengths, the timing of contract renewals, and salary negotiations, aligning them more closely with the trajectory of player performance. As such, coaches and managers can strategically time contract renewals to coincide with the peaks in performance associated with *hot hand* occurrences, which constitutes a proactive approach that allows teams to capitalize on the heightened effectiveness of players during these states. The same applies for salary negotiations, given that this study's results empower negotiation teams to craft salary packages that reflect a more precise value of a player at different career junctures.

Furthermore, the analysis of shooting behaviour during *hot hand* states among players of diverse experience levels contributes to a distinct factor of strategic decision-making, such as tactical guidance for defensive strategies. The identified patterns enable proactive adjustments to game plans, assuming a palpable manifestation of *hot hand* within the stadium. For instance, recognizing that players with lower experience levels showcase an elevated likelihood in attempting “Dunk Shots” and “Layup Shots” during *hot hand* states prompts strategic adjustments on the defensive tactic, which should be adapted to prioritize defenders adept at challenging such shots. Similarly, understanding the strategic choices regarding shot distance and technique provides a blueprint for defenders to anticipate and counteract offensive moves.

Regarding shot difficulty, the findings of this study indicate a clear elevation during a *hot shot* streak. Coaches can draw valuable insights from this research, intervening when they observe an upswing in player confidence, as increased confidence may result in attempts at nearly impossible shots and end the heat. The likelihood of a shot being categorized as *hot* can be influenced by the increased difficulty of the shots attempted. However, it is essential to strike a balance to avoid setting ambitions beyond what is realistically achievable. In these situations, coaches should guide their players and help them maintain a realistic perspective.

In addition, findings of this study related to the impact produced by the opponent's defence in the *hot hand* are also valuable and can be employed for enhancing game strategies. When analysing defensive pressure in preparation for an upcoming game, teams should categorize the opposing team as either high or low defensive intensity. Following this classification, players should concentrate on areas where it is more likely to generate a *hot shot*. Consequently, players facing teams with strong defensive pressure should avoid the "In the Paint" area and the "Restricted Area" if they want to become *hot*. Contrarily, teams opposing weaker defensive adversaries are advised to concentrate their gameplay in zones closer to the hoop. This strategic focus can prove pivotal in securing the first crucial *hot shot* and ultimately winning the game. Teams should thoroughly study these specific areas and strategize on how to exploit these opportunities to their advantage.

Implications also emerge from the comparison performed between the NBA and the WNBA. In basketball, decision-making and risk-taking shape outcomes and this study notes gender nuances. These findings show that it is important for coaches to adjust their strategies based on how men and women make decisions in basketball. Men tend to take more risks, especially when they feel they have a *hot hand*. Knowing and accepting these differences can help coaches be better at their job, making players perform better and teams more successful. Thus, the study suggests that coaches should employ a coaching style that takes into account these gender differences, making the most out of both male and female players.

Coupled with the findings drawn from hypothesis testing, in the dynamic landscape of professional basketball, the integration of predictive models presents a transformative opportunity for teams to refine their managerial strategies. To optimize both players' and teams' performance, predictive modelling offers valuable insights into patterns and sports dynamics contributing to players' *hot hand.* These findings allow for tailoring game strategies, enhancing teams' and players' capabilities to foster overall team growth and, possibly, to boost sports revenues.

Additionally, during live games, coaches and managers can use the *hot hand* predictive model to make real-time optimized and well-informed decisions. Understanding when a player is likely to be *hot* enables dynamic adjustments in playing strategies, potentially influencing critical moments in a game.

Moreover, predictive modelling in basketball, particularly in the NBA, can also be very useful when performing opponent analysis, providing teams with a competitive edge. By assessing the likelihood of *hot hand* streaks in opposing players, teams can adapt their defensive strategies and substitutions accordingly and craft comprehensive game plans effectively.

Finally, besides the immediate impact on gameplay and game management decisions, the integration of a predictive model for player's *hot hand* in the NBA introduces a strategic shift in long-term planning. Teams can leverage players' contract negotiations, identifying players that better align with their strategic needs, ensuring not only short-term success but also laying the foundation for sustained competitiveness. Incorporating predictive analytics not only enhances on-court performance but also fosters success in the data-driven evolving landscape of professional basketball.

8 Future Research

The present study constitutes a valuable contribution to the existing literature, representing a step further towards unravelling the complexity of the *hot hand* phenomenon. Building upon prior findings through the last decades, the conducted analyses aim to explore the impact of diverse aspects in the manifestation of the *hot hand* to scrutinise the several layers of complexity it entails and predict future occurrences. Nevertheless, this study faces some limitations, and the achieved results are not definitive. Hence, further research is required in order to deepen the existing understanding of this phenomenon.

8.1 Hypothesis Testing

Regarding the performed evaluation of the factors that impact *hot shots*, this study faces the limitation of employing proxies as a tool to handle the restrictions imposed by the available data. Specifically, proxies have been used to overcome the lack of data regarding play-by-play defensive pressure, ball possession, and metrics that quantify a team's strength. In the absence of the required data, the employment of proxies served as a foundation for hypothesis testing, contributing to the exploration of the *hot hand* across different contexts. However, the constraints entailed by this approach must be acknowledged, since these variables do not capture the true patterns of the data they represent. Future work should be built upon a dataset containing the data that is absent in this study, eliminating the need to resort to a proxy.

In addition, a wider variety of methods should be applied in order to test the robustness of the results obtained in this study. Performing regression analysis is key to quantifying the strength of the relationship between variables, however, leveraging additional methods is valuable for approaching hypothesis testing from a different perspective. The results deriving from these additional methods would be useful for validating or disproving the results obtained with the regressions. Some examples of additional methods include discriminant analysis and clustering. By grouping similar observations based on specified features, the application of these methods would aid in identifying additional patterns in the data.

Notwithstanding the fact that the conducted analyses provide new clues about the role of diverse aspects on the structural dynamics of the *hot hand*, a deeper understanding can be obtained by extending each section to further details.

8.1.1 Quarterly Dynamics and Home-Court Influence in *Hot Hand* Shooting in the National Basketball Association

Upon investigating the quarterly dynamics entailed by *hot hand* shooting, overtime periods have been excluded from the analysis on account of the dubiousness of examining them as a whole and of the difficulties imposed by the limited amount of available data. It is further acknowledged that this approach imposes some limitations, as individual overtime periods should be assessed as well, in addition to the four main quarters. This should be considered in future studies concerning the quarterly dynamics related to the *hot hand* phenomenon.

Regarding the home-court influence analysis, several improvements can be introduced by elevating the level of detail involved. Despite providing insights on how playing home affects *hot hand* shooting and the magnitude of this impact, this section does not address 2-point and 3-point shots in detail. Considering that previous research by Harris and Roebber (2019) has highlighted the importance of a style of play focused on 2-point shots for home-teams, future research in this issue is needed.

Additionally, in spite of having identified a significant relation between playing home and *hot shots*, key aspects not accounted for in the performed regressions might contribute for a variation in the prevalence of *hot shots* among teams playing home. For instance, recent changes in the home-court location might negatively impact the home-team's performance in *hot shots*. For this reason, the inclusion of a broader range of features should also be considered in order to gain a deeper understanding of the home-court dynamics.

The presented team-by-team statistics reveal that despite some teams achieving a substantially enhanced performance in *hot shots* in home-court games, other teams exhibit the opposite trend. However, this study does not assess the significance of those differences, thus raising the need to build a model incorporating team fixed effects in future work. Moreover, in the case that the fluctuations in the home-court effect across teams are significant, what factors contribute to such variability? Further questions remained unanswered, requiring some attention in the future.

8.1.2 The Role of Player Expertise in the Manifestation of the *Hot Hand* in the National Basketball Association

In the realm of future research, a meticulous re-examination of the construction of player experience might be useful to refine the research analytical framework. Traditional measures, such as the count of seasons played, might inadvertently overlook the variations in actual on-court exposure. Thus, incorporating additional metrics, particularly the total minutes played, could afford a more refined portrayal of experience, distinguishing between athletes who have accumulated expertise through consistent gameplay and those who have participated in multiple seasons but with limited on-court contributions.

Furthermore, the broadening of the scope of available player statistics, spanning from fitness metrics to indices of physical performance, promises a more holistic understanding of the patterns observed in the data. Specifically, the acknowledgment of the gradual decline in physical capabilities with advancing age is particularly helpful to better refine the conclusion that, as player experience advances, a trade-off between accumulating knowledge and diminishing physical capabilities emerges.

Additionally, the acquisition of characteristics that best define the playing characteristics of an athlete would allow for the aggregation of groups of players. This, in turn, opens the door for the utilization of new methods to study the shooting behaviour of players during *hot hand* states, such as K-Means Clustering. This methodological refinement would not only augment the depth of our insights but also fortify the validity and reliability of future investigations.

8.1.3 Streak Patterns and the Interaction between Team Strength and Defensive Pressure Related to *Hot Hand* in the NBA

Following an examination of shot difficulty and its correlation with the *hot hand* phenomenon, it is advisable to contemplate a more refined equation for this variable. The dataset’s limitations result in several variables included in shot difficulty being proxies, posing challenges to a succinct and accurate definition of the variable. Moreover, if feasible, additional variables should be extracted and incorporated into this equation. As asserted by Bocskocsky, Ezekowitz, and Stein (2014), shot difficulty should encompass four categories of variables, each containing at least two or more variables. However, in spite of these limitations, the analysis conducted in this study does yield significant results.

Regarding additional methods, K-means is particularly promising as employing it would enhance the analysis of shot difficulty and the *hot hand* by grouping players based on risk aversion levels, revealing distinctive characteristics within each group.

Regarding defensive pressure and spatial patterns, a detailed analysis can be conducted in future research using a multinomial logistic regression. The dependent variable would represent different zones, while independent variables include *hot shots*, defensive pressure, and an interaction term between them.

Lastly, the variable of team strength stands out for having considerable potential for extending the research. When considered as an independent variable, its value alone may not provide abundant information. Predicting the ability to secure top positions stands out as a more effective approach. Notably, the application of multinomial logistic regression aligns with this proposed strategy.

8.1.4 Gender Dynamics in Decision-Making during Critical Moments in the National Basketball Association

Comparing the relatively smaller dataset of the WNBA to the more extensive dataset of the NBA, even when sliced to the same time frame, presents both challenges and opportunities. The inherent differences in sample sizes may impact the statistical robustness and generalizability of the findings. While insights drawn from the NBA dataset can offer valuable trends and patterns, the smaller WNBA dataset may limit the depth of analysis and the ability to draw definitive conclusions. Furthermore, collecting and including more extensive specific WNBA data would also enhance the comprehensiveness and reliability of research on gender dynamics in decision-making. Increased data granularity would not only contribute to a more equitable representation of both leagues but also enable a more nuanced understanding of the unique factors influencing decision-making within the WNBA, fostering a more inclusive and accurate analysis of gender dynamics in professional basketball.

Exploring gender dynamics in decision-making across various sports presents an exciting avenue for future research. A comparative analysis could delve into how gender influences strategic choices not only in basketball but also in sports with different dynamics and structures. For instance, examining team sports such as soccer could offer insights into whether gender-related decision-making patterns persist across diverse athletic contexts. This comparative approach could facilitate the identification of overarching trends and shed light on the unique challenges and opportunities that athletes of different genders encounter across various sports.

Furthermore, a public GitHub repository, containing the code to extract the WNBA data, has been established to facilitate ongoing advancements of this research. The repository can be accessed at the following link: https://github.com/BLeal2/Shot_Data_WNBA/tree/main.

8.2 Predictive Modelling of *Hot Hand* Shooting in the National Basketball Association

While constructing the predictive model for *hot shots* in the NBA, several possible future work enhancements have been identified. Firstly, exploring different outlier removal methods and defined thresholds would be a crucial point for refinement. Assessing the impact of different techniques on the model's performance could provide valuable insights on how to improve its accuracy and reliability. Moreover, performing a thorough examination of variable correlations through distinct techniques, like pairwise, on the entire dataset could also offer meaningful insights into relationships between different features, leading to a better understanding of the predictors.

Nonetheless, besides the already employed approaches, applying the K-Nearest Neighbours algorithm to find the best method to scale the data for the final model would also possibly constitute promising future work. Performing this experiment, accommodating its sensitivity to the number applied of neighbours and varying it to identify the most fitting configuration, could lead to further optimized model results.

Furthermore, some aspects of the feature selection process can also be enhanced in future research. Evaluating models for numerical and categorical variables individually, before combining them, might refine the selection criteria, ensuring a more comprehensive consideration of potential predictors. Further, conducting multiple attempts with different variable combinations, with a special focus on the ones identified as important by multiple selection methods, would add granularity to the feature selection process.

Lastly, diversifying the choice of machine learning algorithms and implementing grid-search with different parameters for hyperparameter refinement would also offer a comprehensive strategy for model optimization. Iteratively testing different parameters and values would enhance the model's adaptability and predictive accuracy.

9 Conclusion

The expanding investment in sports analytics has underscored the imperative for sustained research development in the field, facilitating the continual evolution of strategic paradigms predicated upon insights derived from empirical data. Henceforth, in pursuit of further enhancing the existing analytics present in the NBA, this study adeptly elucidates the dynamics involved within the enigmatic phenomenon known as the *hot hand*.

Given the inherently subjective and quasi-emotive nature of this phenomenon, its definition is established upon the work of Bocskocsky, Ezekowitz, and Stein (2014), suitably tailored to the peculiarities of this investigation. The subsequent articulation of this definition is performed through a pragmatic evaluation of the contextual settings of NBA games and a meticulous scrutiny of the repercussions stemming from varied adjustments to this definition, facilitated by means of a sensitivity analysis.

The articulation of this phenomenon's specificities and the insights drawn from existing literature, coupled with the augmentation of the pre-existing dataset through the application of web scraping techniques, laid the bases that underpin the attainment of this study's findings. The subsequent analysis entails not only the examination of relationships between various facets of the *hot hand*, including its frequency, shot accuracy, and the sequencing of shots within a streak, with the hypothesized influential factors, but also the development of a predictive model.

In the exploration of the manifestation of *hot hand* within the purview of temporal dynamics throughout the game, the uncovered findings reveal that the likelihood associated with *hot shots* exhibit a more pronounced escalation from the first to the second quarter, in contradistinction to the transitions between the remaining quarters. Furthermore, although a general positive effect on *hot hand* occurrences is discernible when approaching the end of a quarter, the manifestation of this behaviour undergoes variations from one quarter to another. The identification of temporal asymmetries in this study resonates with those uncovered in the research of Grund, Höcker, and Zimmermann (2010), wherein the phenomenon in question is examined across the broader spectrum of shots, as opposed to *hot shots* in specific. Moreover, it is noteworthy to acknowledge that, as the game unfolds, players exhibit astute decision-making by tactically distributing the ball to a player perceived as being *hot*, leading to an intensified allocation of shot opportunities to the identified individual.

Shifting the focus to the realm of home-court advantage, the results echo the outcomes delineated in Bustamante-Sánchez, Gómez, and Jiménez-Saiz (2022)'s comprehensive study, which substantiates the presence of a substantial advantage for teams playing on their home-court. However, a distinctive aspect surfaces in this study, wherein this advantage manifests as more pronounced in terms of accuracy rather than in the frequency of *hot shots*. In addition, it is paramount to underscore that the outcomes indicate a diminishing strength of home-court advantage across the temporal spectrum analysed in this study (2004-2023).

Regarding the influence of player expertise on the occurrences of the *hot hand*, the analysis unravels a noteworthy inverted U-shaped curve evident in both player experience and age, given their highly correlated nature. This distinctive pattern is juxtaposed with values elucidated in the research made by Bakkenbüll (2017), in which the influence of age in the evolution of performance is assessed. Notwithstanding, this behaviour is intriguingly different from the findings on the variation of shot accuracy during a *hot hand* as regards to experience and age, where no discernible linear or quadratic relationship emerges.

Notably, the opposing dynamics observed in factors such as home advantage, as articulated above, where the most pronounced effect pertains to shot accuracy rather than the occurrence of *hot shots*, accentuates the complex nature of this phenomenon. This underscores the notion that identifying influential factors affecting the *hot hand* may be insufficient if no attention is given to understanding the specific nuances of how these factors truly impact this event.

Delving into the complex issue of understanding shooting behaviour during *hot hand* states, in tandem with experience, the analysis has yielded significant revelations, indicating that higher player experience translates into an augmented average shot distance and a lower tendency towards "Dunk Shots" and "Layup Shots” over "Jump Shots”.

Furthermore, in regard to the difference in shot tendencies within a *hot hand* streak, the study shows a marginal increase in difficulty as the streak’s length increases. However, a counterintuitive growth in accuracy emerges. This departure challenges the work done by Csapo and Raab (2014), which posits that defensive responses during a streak lack a rational foundation, given that accuracy is assumed to fail to show a corresponding increase.

The intriguing revelations unveiled in the earlier analysis propel an interest in the dynamics of defensive interactions during the *hot hand* phenomenon. Consequently, an apparent trend emerged, wherein *hot shots* have been found to display a propensity to occur in closer proximity to the basket when playing against poor defence teams, which could be intricately linked to the heightened ease of reaching these areas.

On the other hand, assuming a superior position in the ranking was found to not unequivocally translate into an abundance of *hot shots*, particularly when confronted with a team boasting challenging defensive pressure. Conversely, when a team occupies a lower rank but plays against a team exhibiting weak defensive capabilities, the log-odds of *hot shots* witness a notable increase. Nevertheless, it is imperative to underscore that, despite suggesting strategic advantages, the fluctuation in log odds remains modest.

Moreover, in the pursuit of unravelling the gender dynamics linked with this event, the conducted research has unveiled significant patterns. Despite the observation that WNBA players exhibit a higher average shot difficulty during *hot hand*, the analysis shows that NBA players display a propensity to venture beyond their usual boundaries, embracing a calculated risk-taking behaviour during high-intensity events. This judgment resonates with the insights posited by Böheim, Freudenthaler, and Lackner (2016), affirming the prevailing tendency for female players to exhibit a more risk-averse disposition.

In the context of the latter research, an effort has been made to maintain a consistent timeframe with the WNBA dataset, by capping the NBA dataset. Intriguing revelations surfaced during this process, particularly when comparing the outcomes of the earlier study on the impact of the *hot hand* on shot difficulty for all seasons with the current research focused on later years (2017-2022). Particularly, the average difficulty of shots and the increase in shot difficulty during a streak are found to be lower, which might signify an evolving trend in the game over seasons, with a heightened sense of caution during shooting moments in high-pressure events.

Finally, the culmination of this study involves the construction of a classification model adept at comprehensively capturing the delineated relationships and generating predictive insights. On this context, emphasis must be placed on the profound impact of the extensive and imbalanced dataset, affecting model preparation and results.

The model that has yielded the most favourable outcomes is the Random Forest, trained with a focus on the F0.5-score to address class imbalance and implemented without any sampling methods. For model evaluation, precision emerged as the chosen metric, considering the contextual significance of false positives as the more consequential error.

The produced results, when compared to previous literature, considering the complexity and rarity of the *hot hand* phenomenon, are favourable. The precision of the final model stands at 0.53, a notable departure from the precision of 0.22 reported by Zhou et al. (2020) in a study that centres on crash prediction using Random Forest models. Similarly, Oughali, Bahloul, and El_Rahman (2019)'s work, which encompasses shooting prediction using Random Forest, has demonstrated a lower TPR and higher FPR compared to this study´s model. This examination accentuates the effectiveness of the final model in the context of rare-event prediction in basketball.

10 References

Allen, Sian, and Will Hopkins. 2015. 'Age of Peak Competitive Performance of Elite Athletes: A Systematic Review'. *Sports Medicine*. Springer International Publishing. https://doi.org/10.1007/s40279-014-0239-x.

An, Tae, and Moon Kim. 2010. 'A New Diverse AdaBoost Classifier'. In *Proceedings - International Conference on Artificial Intelligence and Computational Intelligence, AICI 2010*, 1:359–63. https://doi.org/10.1109/AICI.2010.82.

Anders, Anne, and Kurt Rotthoff. 2014. 'Is Home-Field Advantage Driven by the Fans? Evidence from Across the Ocean'. *Applied Economics Letters* 21 (16).

Arkes, Jeremy. 2013. 'Misses in "Hot Hand" Research'. *Journal of Sports Economics* 14 (4): 401–10. https://doi.org/10.1177/1527002513496013.

Avugos, Simcha, Michael Bar-Eli, Ilana Ritov, and Eran Sher. 2013. 'The Elusive Reality of Efficacy-Performance Cycles in Basketball Shooting: An Analysis of Players' Performance under Invariant Conditions'. *International Journal of Sport and Exercise Psychology* 11 (2): 184–202. https://doi.org/10.1080/1612197X.2013.773661.

Bakkenbüll, Linn-Brit. 2017. 'Physical Constitution Matters for Athletic Performance and Salary of NBA Players'. *1*. www.wiwi.uni-muenster.de/io.

Beer, Tommy. 2020. 'These Stats Suggest LeBron James Is The Greatest Basketball Player Ever'. *Forbes*, 12 October 2020. https://www.forbes.com/sites/tommybeer/2020/10/12/these-stats-suggest-lebron-james-is-the-greatest-basketball-player-ever/?sh=2711f85d1cab.

Bocskocsky, Andrew, John Ezekowitz, and Carolyn Stein. 2014. 'Heat Check: New Evidence on the Hot Hand in Basketball', July. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2481494.

Böheim, René, Christoph Freudenthaler, and Mario Lackner. 2016. ‘Gender Differences in Risk-Taking: Evidence from Professional Basketball’.

Burns, Bruce. 2001. ‘The Hot Hand in Basketball: Fallacy or Adaptive Thinking?’ *Proceedings of the Annual Meeting of the Cognitive Science Society* 23 (23). https://escholarship.org/uc/item/056552kc.

Bustamante-Sánchez, Álvaro, Miguel Gómez, and Sergio Jiménez-Saiz. 2022. ‘Game Location Effect in the NBA: A Comparative Analysis of Playing at Home, Away and in a Neutral Court during the COVID-19 Season’. *International Journal of Performance Analysis in Sport* 22 (3): 370–81. https://doi.org/10.1080/24748668.2022.2062178.

Chawla, Nitesh, Kevin Bowyer, Lawrence Hall, and Philip Kegelmeyer. 2002. ‘SMOTE: Synthetic Minority Over-Sampling Technique’. *Journal of Artificial Intelligence Research* 16: 321–57.

Conn, Jordan. 2019. ‘How the Hornets and Hurricane Katrina Paved the Way for the OKC Thunder’. *The Ringer,* 24 October 2019. https://www.theringer.com/2019/10/24/20929974/sonic-boom-episode-4-new-orleans-hornets-oklahoma-city-katrina.

Csapo, P, and M Raab. 2014. "“‘Hand down, Man down.’” Analysis of Defensive Adjustments in Response to the Hot Hand in Basketball Using Novel Defense Metrics’. *PLoS ONE* 9 (12). https://doi.org/10.1371/journal.pone.0114184.

Domingos, Pedro. 1999. ‘MetaCost: A General Method for Making Classifiers Cost-Sensitive’. *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.,* August, 155–64. http://www.gia.ist.utl.pt/-pedrod.

Elkins, Kathleen. 2020. ‘How Much Money the Winners of the NBA All-Star Game Will Earn’. *CNBC,* February. https://www.cnbc.com/2020/02/15/nba-all-star-game-2020-how-much-money-the-winning-team-earns.html.

Ergül, Barış, Arzu Yavuz, and Hasan Yavuz. 2014. 'Classification of NBA League Teams Using Discriminant and Logistic Regression Analyses'. *Pamukkale Journal of Sport Sciences* 5 (1): 48–60. http://pjss.pau.edu.tr.

ESPN. 2021. 'WNBA's Greatest Players of All Time: Ranking the 25 Best in League History'. ESPN. 21 September 2021. https://www.espn.com/wnba/story/_/id/32210623/wnba-greatest-players-all-ranking-25-best-league-history.

Gilovich, Thomas, Robert Vallone, and Amos Tversky. 1985. 'The Hot Hand in Basketball: On the Misperception of Random Sequences'. *Cognitive Psychology* 17 (3): 295–314. https://doi.org/10.1016/0010-0285(85)90010-6.

Grund, Christian, Jan Höcker, and Stefan Zimmermann. 2010. 'Risk Taking Behavior in Tournaments-Evidence from the NBA'. http://hdl.handle.net/10419/37372.

Gupta, Siddhant. 2023. 'Why Was There an NBA Lockout in 2011? Taking a Closer Look at the Moment the League Stopped'. *Sports Keeda*, 28 June 2023. https://www.sportskeeda.com/basketball/news-why-nba-lockout-2011-taking-closer-look-moment-league-stopped.

Harris, Austin, and Paul Roebber. 2019. 'NBA Team Home Advantage: Identifying Key Factors Using an Artificial Neural Network'. *PLoS ONE* 14 (7). https://doi.org/10.1371/journal.pone.0220630.

Iso-Ahola, Seppo, and Charles Dotson. 2014. 'Psychological Momentum: Why Success Breeds Success'. *American Psychological Association* 18 (1): 19–33. https://doi.org/10.1016/J.EUROECOREV.2021.103771.

Jamieson, Jeremy. 2010. 'The Home Field Advantage in Athletics: A Meta-Analysis'. *Journal of Applied Social Psychology* 40 (7): 1819–48.

Jones, Marshall. 2007. 'Home Advantage in the NBA as a Game-Long Process'. *Journal of Quantitative Analysis in Sports* 3 (4). https://doi.org/10.2202/1559-0410.1081.

Kalén, Anton, Alexandra Pérez-Ferreirós, Pablo B. Costa, and Ezequiel Rey. 2021. 'Effects of Age on Physical and Technical Performance in National Basketball Association (NBA) Players'. *Research in Sports Medicine* 29 (3): 277–88. https://doi.org/10.1080/15438627.2020.1809411.

Koehler, Jonathan, and Caryin Conley. 2003. 'The "Hot Hand" Myth in Professional Basketball'. *Journal of Sports & Exercise Psychology* 25: 253–59.

Mackenzie, Rob, and Chris Cushion. 2013. 'Performance Analysis in Football: A Critical Review and Implications for Future Research'. *Journal of Sports Sciences*. https://doi.org/10.1080/02640414.2012.746720.

Miller, Joshua, and Adam Sanjurjo. 2018. 'Surprised by the Hot Hand Fallacy?A Truth in the Law of Small Numbers'. *Econometrica, Vol. 86, No. 6, 2019-2047* 86 (6). https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA14943.

———. 2021. 'Is It a Fallacy to Believe in the Hot Hand in the NBA Three-Point Contest?' *European Economic Review* 138 (September). https://doi.org/10.1016/j.euroecorev.2021.103771.

Moss, Hollye. 2022. 'FOUR FACTORS IN THE WNBA'. https://www.basketball-reference.com/wnba/.

NBA Official. 2018. 'RULE NO. 5: Scoring and Timing'. NBA. 2018. https://official.nba.com/rule-no-5-scoring-and-timing/.

Nusbaum, Spencer. 2021. 'How the Atlanta Dream Can Bump up Those Shooting Percentages'. The Next. 25 May 2021. https://www.thenexthoops.com/features/how-the-atlanta-dream-can-bump-up/.

Oughali, Maram Shikh, Mariah Bahloul, and Sahar El_Rahman. 2019. *Analysis of NBA Players and Shot Prediction Using Random Forest and XGBoost Models*. IEEE.

Pelechrinis, Konstantinos, and Wayne Winston. 2022. ‘The Hot Hand in the Wild’. *PLoS ONE* 17 (1). https://doi.org/10.1371/journal.pone.0261890.

Pollard, R., and G. Pollard. 2005. ‘Long-Term Trends in Home Advantage in Professional Team Sports in North America and England (1876-2003)’. *Journal of Sports Sciences* 23 (4): 337–50. https://doi.org/10.1080/02640410400021559.

Rao, Justin. 2009. ‘Experts’ Perceptions of Autocorrelation: The Hot Hand Fallacy Among Professional Basketball Players’.

Ribeiro, Haroldo, Satyam Mukherjee, and Xiao Zeng. 2016. ‘The Advantage of Playing Home in NBA: Microscopic, Team-Specific and Evolving Features’. *PLoS ONE* 11 (3). https://doi.org/10.1371/journal.pone.0152440.

Sampaio, Jaime, Eric J. Drinkwater, and Nuno M. Leite. 2010. ‘Effects of Season Period, Team Quality, and Playing Time on Basketball Players’ Game-Related Statistics’. *European Journal of Sport Science* 10 (2): 141–49. https://doi.org/10.1080/17461390903311935.

Scaletta, Kelly. 2020. ‘Why NBA Game Pace Is Historicically High’. Lineups. 2 March 2020. https://www.lineups.com/articles/why-nba-game-pace-is-at-historic-high/.

Schindler, Mark. 2023. ‘The WNBA Shooting Evolution, and Why This Is the Best Year Yet’. Compare. 10 August 2023. https://www.compare.bet/en-us/news/the-wnba-shooting-evolution-and-why-this-is-the-best-year-yet.

Schuhmann, John. 2021. ‘NBA’s 3-Point Revolution: How 1 Shot Is Changing the Game’. NBA. 14 October 2021. https://www.nba.com/news/3-point-era-nba-75.

Skinner, Benjiman. 2013. ‘The Relationship Between Confidence and Performance Throughout a Competitive Season’. *All Graduate Plan B and Other Reports, Spring 1920 to Spring 2023* 285 (May). https://doi.org/https://doi.org/10.26076/3fdf-fab6.

Solomom, Michael. 2011. 'NBA Lockout: Why Was the Pay Deal Refused and What Happens Next?' *The Guardian,* 15 November 2011. https://www.theguardian.com/sport/blog/2011/nov/15/nba-lockout-pay-deal-explained.

Uiti, Jacob. 2022. 'Loved, Loathed and Everywhere: How the Three-Pointer Came to Dominate the NBA | NBA | The Guardian'. The Guardian. 13 October 2022. https://www.theguardian.com/sport/2022/oct/13/loved-loathed-and-everywhere-how-the-three-pointer-came-to-dominate-the-nba.

Vorkunov, Mike, and C.J. Moore. 2019. 'NBA, NBPA Talks Explained: New Draft Age Limit Rules and What Else to Expect from Proposed CBA Changes'. The Athletic. 19 September 2019. https://theathletic.com/3608468/2022/09/19/nba-cba-draft-age-limit-eligibility/.

WABC. 2020. 'Basic Shooting - Jump Shot - A Player Should Not Be Taught to Jump Shoot Unless They Have a Consistent Shooting Action with a High Release.' WABC. 2020. https://wabc.fiba.com/manual/level-1/l1-player/l1-2-offensive-basketball-skills/2-7-shooting/2-7-7-basic-shooting-jump-shot/.

Wang, Feng, and Guohua Zheng. 2022. 'What Are the Changes in Basketball Shooting Pattern and Accuracy in National Basketball Association in the Past Decade?' *Frontiers in Psychology* 13 (September). https://doi.org/10.3389/fpsyg.2022.917980.

Weiss, Gary, and Haym Hirsh. 2000. 'Learning to Predict Extremely Rare Events'. *AAAI Workshop on Learning from Imbalanced Data Sets*. www.aaai.org.

Woo, Marcus. 2019. 'The Subtle Biases That Influence Home-Court Advantage'. Inside Science. 22 October 2019. https://www.insidescience.org/news/subtle-biases-influence-home-court-advantage.

Zhang, Shaoliang, Alberto Lorenzo, Miguel Angel Gómez, Nuno Mateus, Bruno Gonçalves, and Jaime Sampaio. 2018. 'Clustering Performances in the NBA According to Players'

Anthropometric Attributes and Playing Experience'. *Journal of Sports Sciences* 36 (22): 2511–20. https://doi.org/10.1080/02640414.2018.1466493.

Zhou, Xiaoyi, Pan Lu, Zijian Zheng, Denver Tolliver, and Amin Keramati. 2020. 'Accident Prediction Accuracy Assessment for Highway-Rail Grade Crossings Using Random Forest Algorithm Compared with Decision Tree'. *Reliability Engineering & System Safety* 200 (August). https://doi.org/10.1016/J.RESS.2020.106931.

11 Appendix

11.1 A1 - Data Dictionary

The following tables present descriptions and data types for each variable in the NBA and WNBA datasets. Further details for specific variables are provided after the tables, to offer more comprehensive information beyond the brief descriptions.

Table 1 - Data Dictionary NBA Dataset.

Data Dictionary – NBA Dataset		
Variable	**Data Type**	**Description**
season_1	Date (Y)	Year when the shot has taken place
season_2	Date (Y-Y)	Season when the shot has taken place
team_id	Categorical	Unique number identifier of the team
team_name	Categorical	Team name of player taking the shot
player_id	Categorical	Unique number identifier of the team
player_name	Categorical	Player name
game_date	Date (D-M-Y)	Timestamp of the game
game_id	Categorical	Unique number identifier of the game
home_team	Categorical	Three letter abbreviation of the home team's name
away_team	Categorical	Three letter abbreviation of the away team's name
event_type_binary	Binary	Says if the player made the shot (1) or not (0)
action_type	Categorical	Description of shot type[*1]
group_action_type	Categorical	Grouped data for the column action type[*1]
shot_type	Categorical	The type of shot with regard to points the shot corresponds to[*2]
zone_name	Categorical	Name of the court zone the shot took place in[*3]
zone_range	Categorical	Distance range of shot by zones[*3]
loc_x	Spatial	X coordinate of the shot[*3]
loc_y	Spatial	Y coordinate of the shot[*3]
shot_distance	Numerical	Distance of the shot from the centre of the hoop, in feet
quarter	Categorical	Quarter of the game[*4]
mins_left	Numerical	Minutes remaining in the quarter
secs_left	Numerical	Seconds remaining in minute of the quarter
drtg_opposing_team	Numerical	Defence rating of the opposing team[*5]
ranking	Categorical	Final standing of the team within its conference for each season
team_conference	Categorical	Conference of the team[*6]
player_experience	Numerical	Number of years that the player spent playing in the NBA
player_birthdate	Date (D-M-Y)	Birthdate of the player
age	Numerical	Age of the player at the time of the shot

Table 2 - Data Dictionary WNBA Dataset.

Data Dictionary – WNBA Dataset		
Variable	**Data Type**	**Description**
season_1	Date (Y)	Year when the shot has taken place
season_2	Date (Y-Y)	Season when the shot has taken place
team_id	Categorical	Unique number identifier of the team
team_name	Categorical	Team name of player taking the shot
player_id	Categorical	Unique number identifier of the team
player_name	Categorical	Player name
game_date	Date (D-M-Y)	Timestamp of the game
game_id	Categorical	Unique number identifier of the game
home_team	Categorical	Three letter abbreviation of the home team's name
away_team	Categorical	Three letter abbreviation of the away team's name
event_type_binary	Binary	Says if the player made the shot (1) or not (0)
action_type	Categorical	Description of shot type[*1]
group_action_type	Categorical	Grouped data for the column action type[*1]
shot_type	Categorical	The type of shot with regard to the points the shot corresponds to[*2]
zone_name	Categorical	Name of the court zone the shot took place in[*3]
loc_x	Spatial	X coordinate of the shot[*3]
loc_y	Spatial	Y coordinate of the shot[*3]
shot_distance	Numerical	Distance of the shot with respect to the centre of the hoop, in feet
quarter	Categorical	Quarter of the game[*4]
mins_left	Numerical	Minutes remaining in the quarter
secs_left	Numerical	Seconds remaining in minute of the quarter
drtg_opposing_team	Numerical	Defence rating of the opposing team[*5]
ranking	Categorical	Final standing of the team within its conference for each season
team_conference	Categorical	Conference of the team[*6]
player_experience	Numerical	Number of years that the player spent playing in the NBA
player_birthdate	Date (D-M-Y)	Birthdate of the player
age	Numerical	Age of the player at the time of the shot

[*1] Action Type and Group Actions Type

In the NBA and WNBA, there are various types of shots that take place during a game. These shots are classified based on the location from which they are taken and the circumstances surrounding the shot. Given the 70 distinct values in the *action_type* column, it becomes imperative to introduce a new column for the classification of major groups. Below is the table of the correspondence between the new column and action type.

Table 3 - Correspondence Between the Column Group Actions Type and Actions Type.

Assignment of group action types to various values in the "action type" column	
Group Action Type	**Action Type**
Jump Shot	Jump Shot, Step Back Jump Shot, Pullup Jump Shot, Turnaround Jump Shot, Running Pull-Up Jump Shot, Turnaround Fadeaway shot, Fadeaway Jump Shot, Driving Jump shot, Running Jump Shot, Floating Jump shot, Driving Floating Jump Shot, Driving Jump shot, Pullup Jump shot, Step Back Jump shot
Bank shot	Turnaround Bank shot, Hook Bank Shot, Jump Bank Shot, Driving Bank Hook Shot, Fadeaway Bank shot, Turnaround Bank Hook Shot, Step Back Bank Jump Shot, Driving Bank shot, Pullup Bank shot, Running Bank shot, Turnaround Fadeaway Bank Jump Shot, Driving Floating Bank Jump Shot, Running Bank Hook Shot, Jump Bank Hook Shot
Layup shot	Running Finger Roll Layup Shot, Driving Layup Shot, Driving Finger Roll Layup Shot, Tip Layup Shot, Driving Reverse Layup Shot, Running Reverse Layup Shot, Layup Shot, Reverse Layup Shot, Cutting Layup Shot, Putback Layup Shot, Running Layup Shot, Cutting Finger Roll Layup Shot, Running Alley Oop Layup Shot, Alley Oop Layup shot, Finger Roll Layup Shot, Tip Shot, Running Tip Shot, Driving Finger Roll Shot, Finger Roll Shot, Turnaround Finger Roll Shot, Running Finger Roll Shot
Dunk shot	Driving Slam Dunk Shot, Reverse Slam Dunk Shot, Running Slam Dunk Shot, Putback Reverse Dunk Shot, Follow Up Dunk Shot, Slam Dunk Shot, Putback Slam Dunk Shot, Reverse Dunk Shot, Driving Reverse Dunk Shot, Running Reverse Dunk Shot, Driving Dunk Shot, Running Alley Oop Dunk Shot, Alley Oop Dunk Shot, Putback Dunk Shot, Cutting Dunk Shot, Dunk Shot, Running Dunk Shot, Tip Dunk Shot
Hook shot	Running Hook Shot, Jump Hook Shot, Turnaround Hook Shot, Driving Hook Shot, Hook Shot
Not specified	Not specified

The main categories descriptions which correspond to the group actions type column are the following:

1. **Jump Shot:** This is a standard shot in which a player jumps off the ground and releases the ball while in mid-air. Jump shots can be taken from various distances, including mid-range jumpers and three-pointers.
2. **Bank Shot:** A Bank Shot refers to a method of shooting the ball by intentionally rebounding it off the backboard. This technique is often used when a player is positioned at an angle to the basket, allowing them to use the backboard as a target for a more controlled and strategic shot.

3. **Layup Shot:** A layup is a close-range shot where a player drives towards the basket and releases the ball with one hand while laying it off the backboard. It's often used when a player is close to the hoop.
4. **Dunk Shot:** A dunk involves a player jumping and forcefully putting the ball through the hoop with one or both hands. Dunks are high-percentage shots and often result in spectacular plays.
5. **Hook Shot:** A hook shot is a basketball shooting technique in which a player, typically a post player or a player closer to the basket, releases the ball with a sweeping, hooking motion. The shooter uses one hand to control the ball and extends the arm away from the defender, creating a curved or hook-like trajectory.
6. **Not Specified**: There is no specification of the action type.

[*2] Shot Type

In basketball, there are three different types of points that a shot can make, each with its own point value:

1. **2-Point Field Goal (2 points):** This is the most common type of shot in basketball. When a player makes a field goal inside the 3-point line, it is worth 2 points.
2. **3-Point Field Goal (3 points):** When a player makes a shot from beyond the 3-point line, it is worth 3 points. In the NBA, the 3-point line has an arc of 23 feet and 9 inches (approximately 7.24 meters), while in the WNBA it has an arc of approximately 22 feet and 1.75 inches (approximately 6.75 meters).
3. **Free Throw (1 point each):** Free throws are awarded in various situations when a player is fouled during a basketball game.

However, it is relevant to mention that the dataset used in this study does not have records regarding Free Throws.

*3 Zone Names and Coordinates

Information about the different zones of the court, the name lines and the axis of the coordinates:

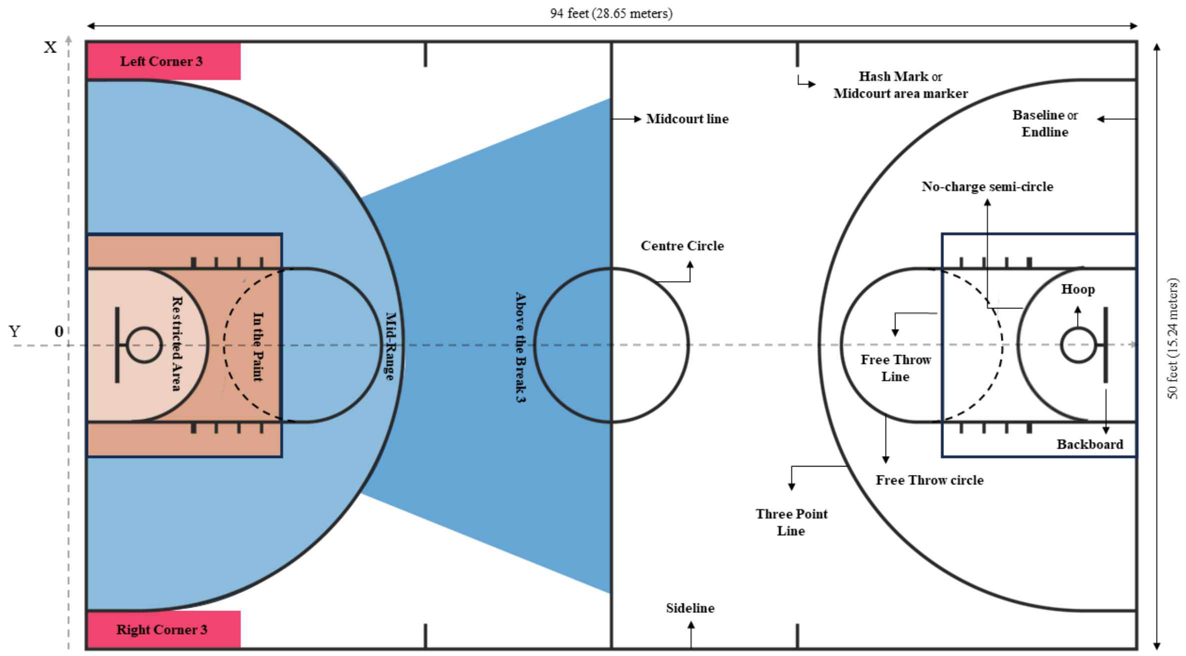


Figure 1 - Court Dimensions, Zones Placements, lines names and axis placements.

It should be highlighted that despite the differences in the 3-point line between NBA and WNBA, the court size is the same for both leagues.

*4 Quarter

Every basketball game is divided into four quarters. Each quarter usually lasts 12 minutes. If the score is tied at the end of regulation time (four quarters or two halves), an overtime period is played to determine the winner. In the NBA and WNBA, overtime periods are 5 minutes long.

There is a break between the second and third quarters, known as halftime. This break typically lasts for 15 minutes, during which teams and players can regroup and adjust. Teams are allowed a certain number of timeouts during the game. These timeouts provide an opportunity for coaches to strategize, make substitutions, and give players a brief rest. In the

NBA, each team is allowed seven timeouts during regulation play, with additional timeouts in overtime.

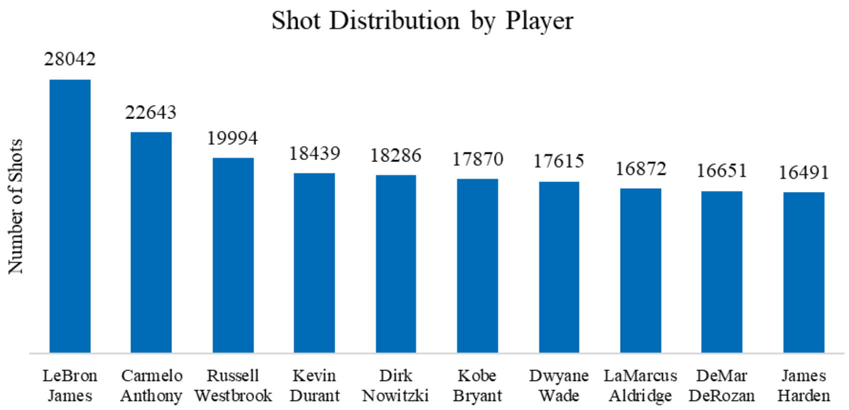
[*5] Defensive Rating

The defensive rating indicates how many points a team allows per 100 possessions by the opposing team. This basketball stat differs from the rest, since lower numbers are more favourable.

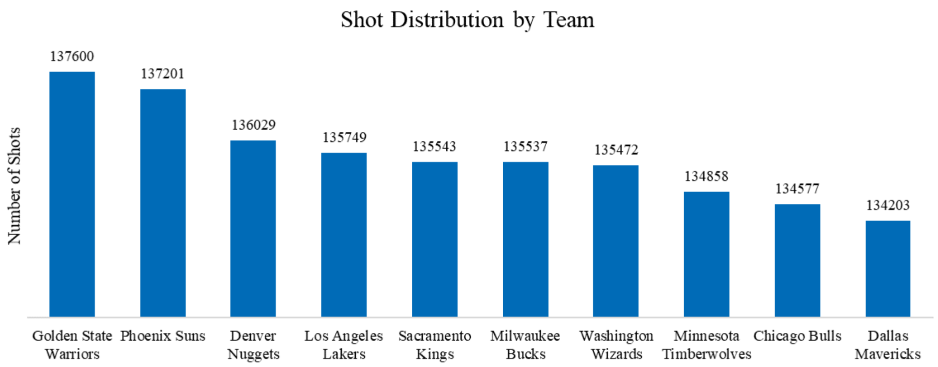
[*6] Team Conference

NBA teams are divided into Eastern and Western conferences, depending on their geographic region. Among each conference, teams are further allocated into one of three divisions. During the regular season, to which our game data belongs to, teams compete against others belonging to the same conference. In the end of each regular season, only the top seed of each division is selected for the playoffs.

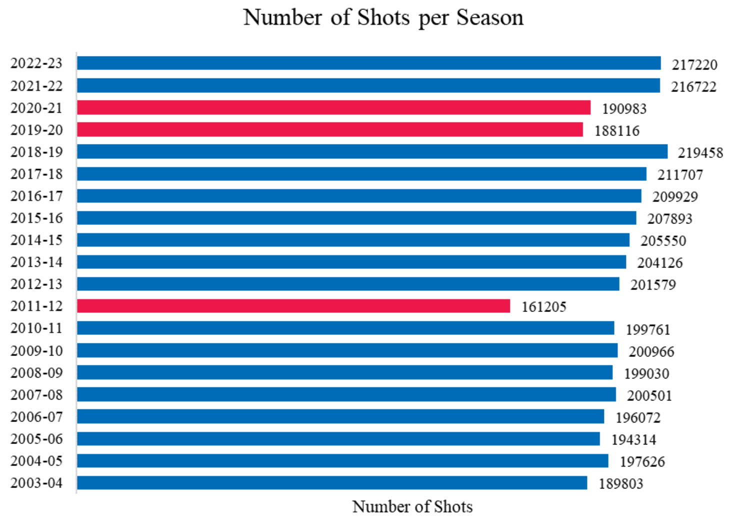
11.2 Descriptive Statistics



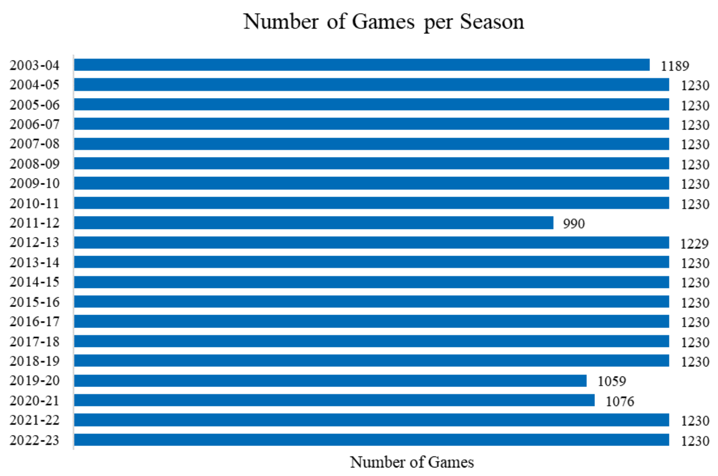
A2 - Distribution by top 10 NBA Players.



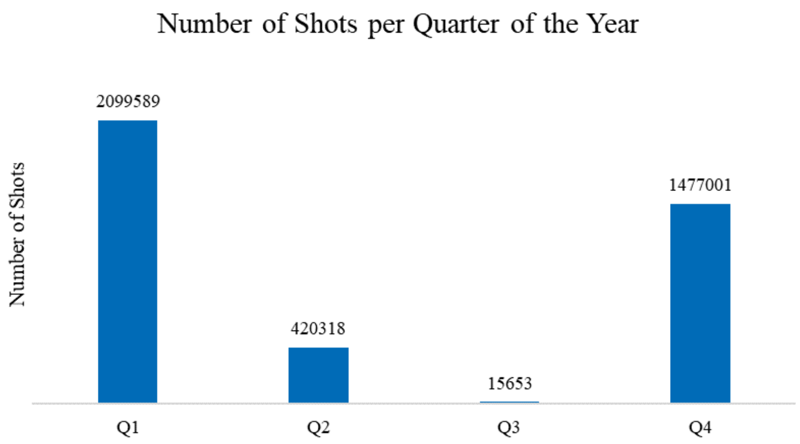
A3 - Shot Distribution by top 10 NBA Teams.



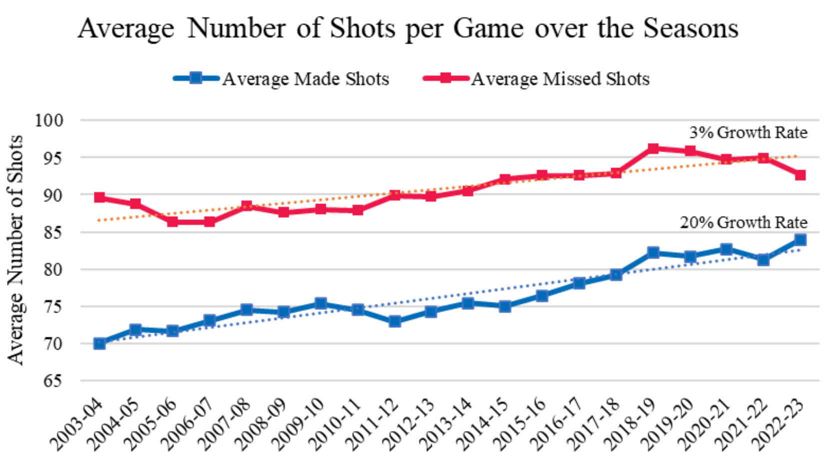
A4 - Number of Shots per Season.



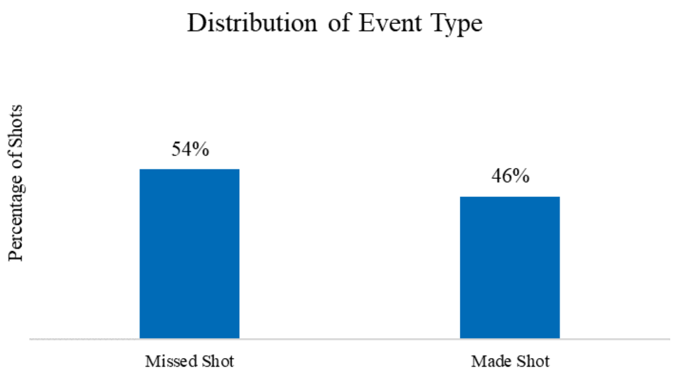
A5 - Number of Games per Season.



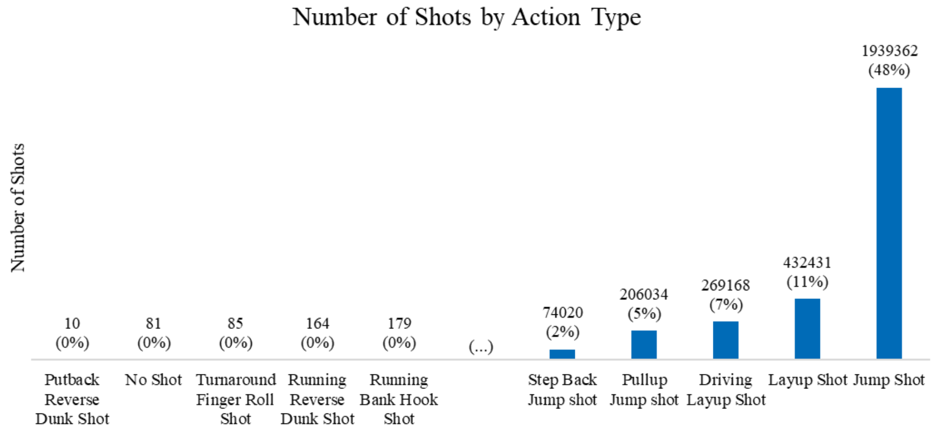
A6 - Number of Shots per Quarter of the Year.



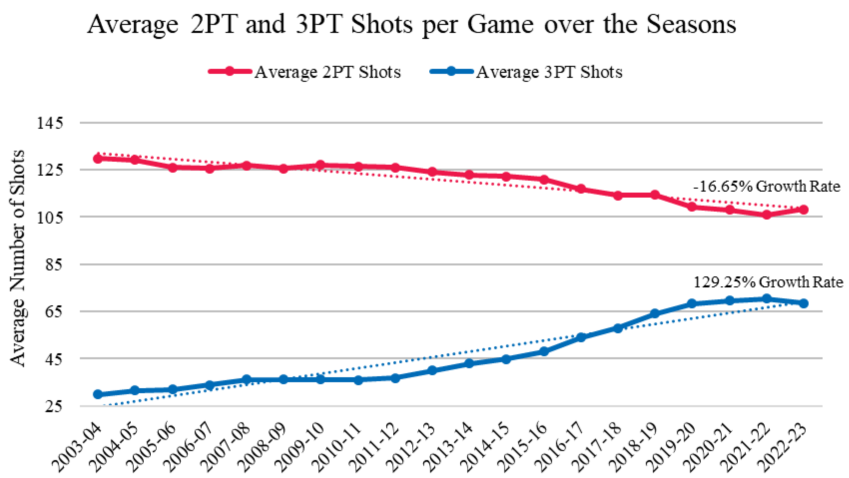
A7 – Average Number of Shots per Game over the Seasons.



A8 - Distribution of Event Type.



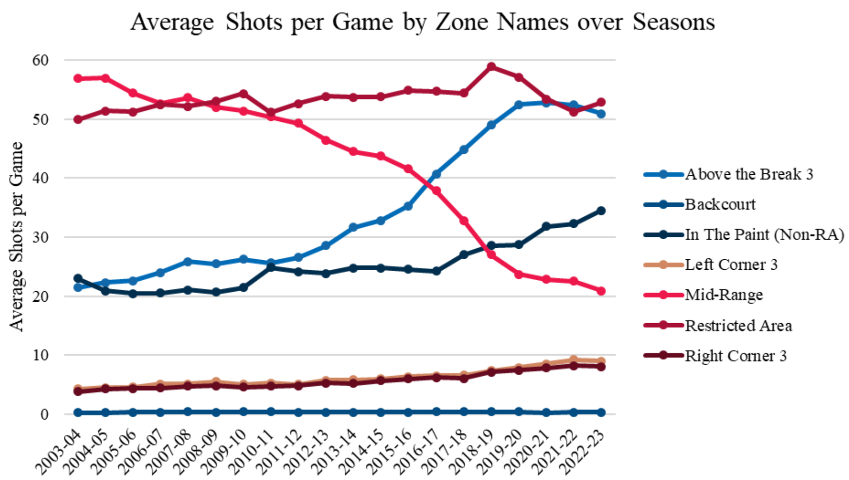
A9 – Number of Shots by Action Type.



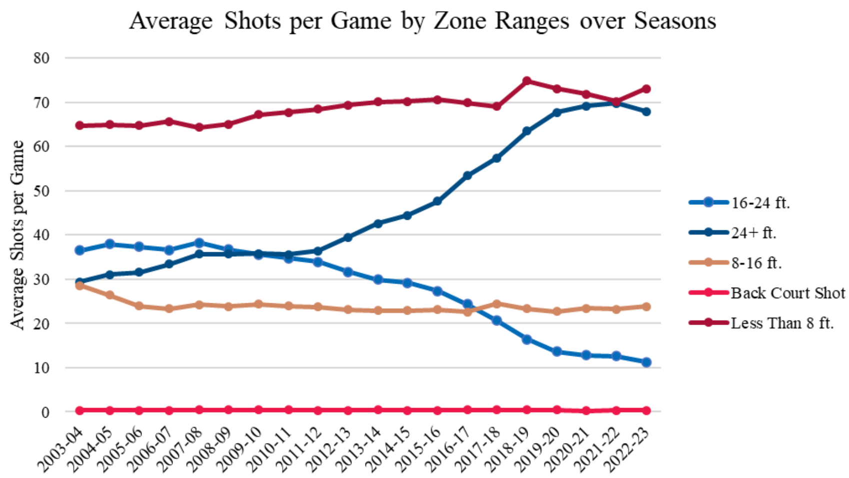
A10 – Average 2PT and 3PT Shots per Game over the Season.

A11 - Number of 2PT and 3PT Shots per Game and respective Percentages.

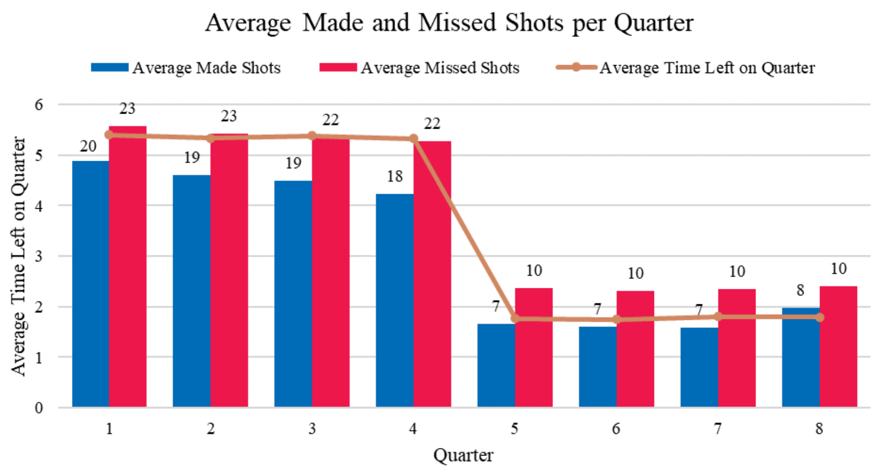
Number of 2PT and 3PT Shots per Game and respective Percentages				
Season	**Average 2PT Shots per Game**	**Average 3PT Shots per Game**	**Percentage 2PT**	**Percentage 3PT**
2003-04	130	30	81%	19%
2004-05	129	32	80%	20%
2005-06	126	32	80%	20%
2006-07	126	34	79%	21%
2007-08	127	36	78%	22%
2008-09	126	36	78%	22%
2009-10	127	36	78%	22%
2010-11	126	36	78%	22%
2011-12	126	37	77%	23%
2012-13	124	40	76%	24%
2013-14	123	43	74%	26%
2014-15	122	45	73%	27%
2015-16	121	48	72%	28%
2016-17	117	54	68%	32%
2017-18	114	58	66%	34%
2018-19	114	64	64%	36%
2019-20	109	68	62%	38%
2020-21	108	70	61%	39%
2021-22	106	70	60%	40%
2022-23	108	68	61%	39%



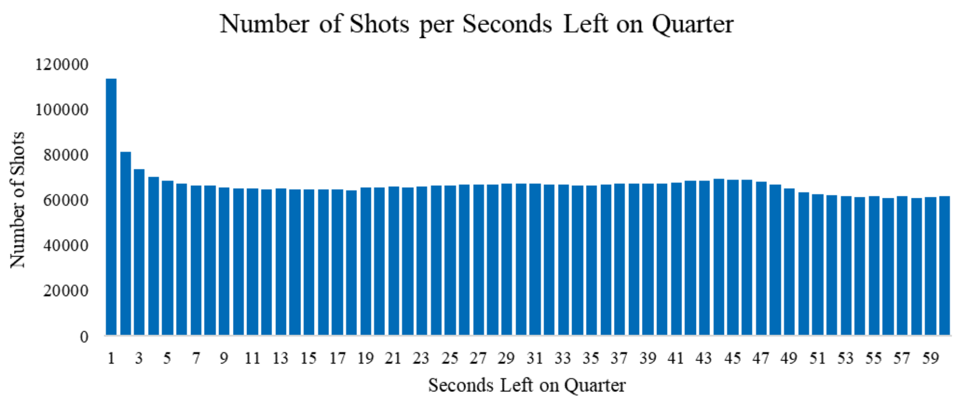
A12 - Average Shots per Game by Zone Names over Seasons.



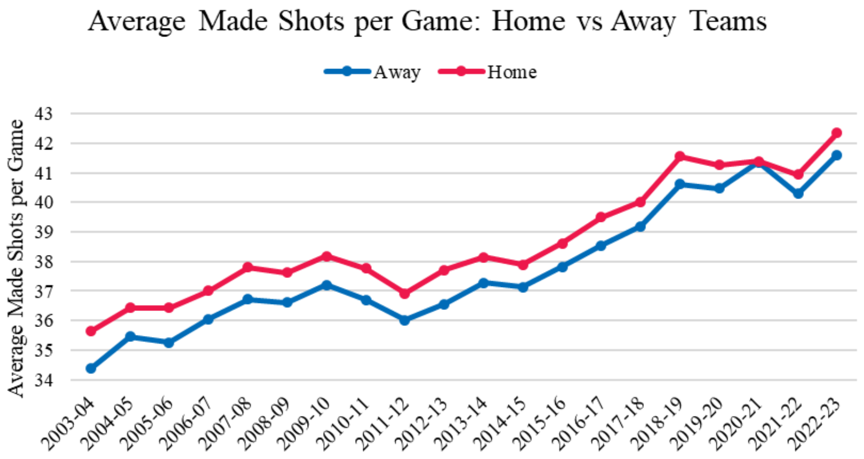
A13 - Average Shots per Game per Zone Range over the Seasons.



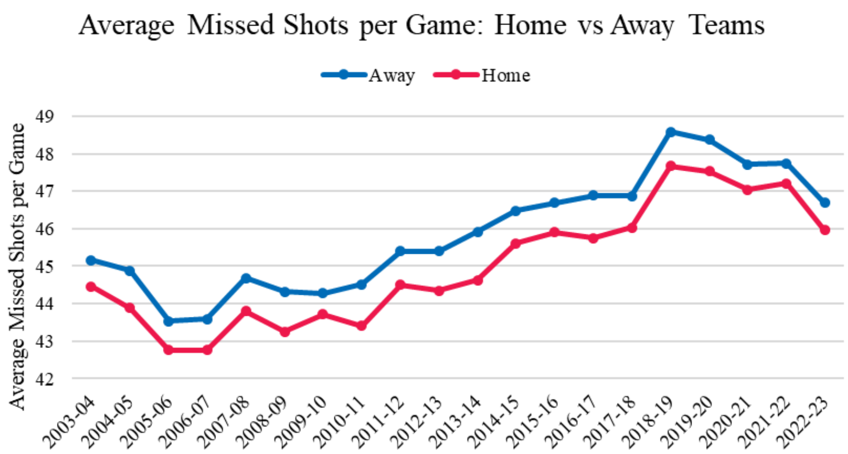
A14 – Average Made and Missed Shots per Quarter and by Minutes Left in Quarter.



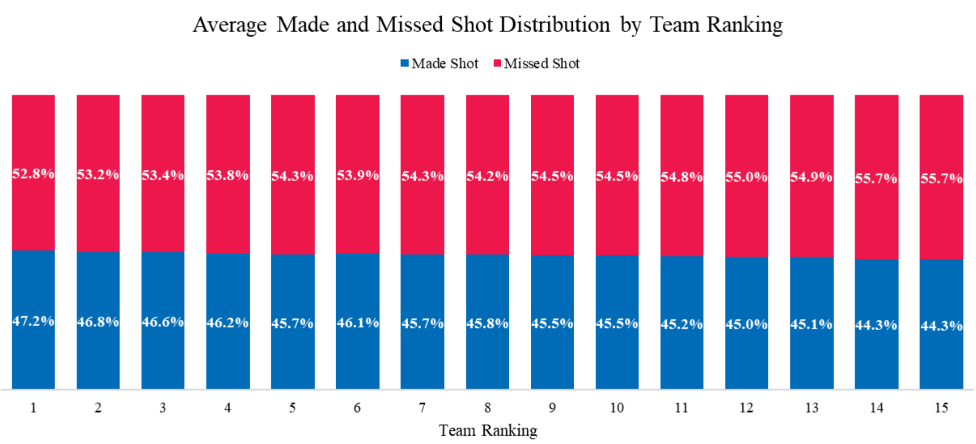
A15 - Number of Shots per Seconds Left on Quarters.



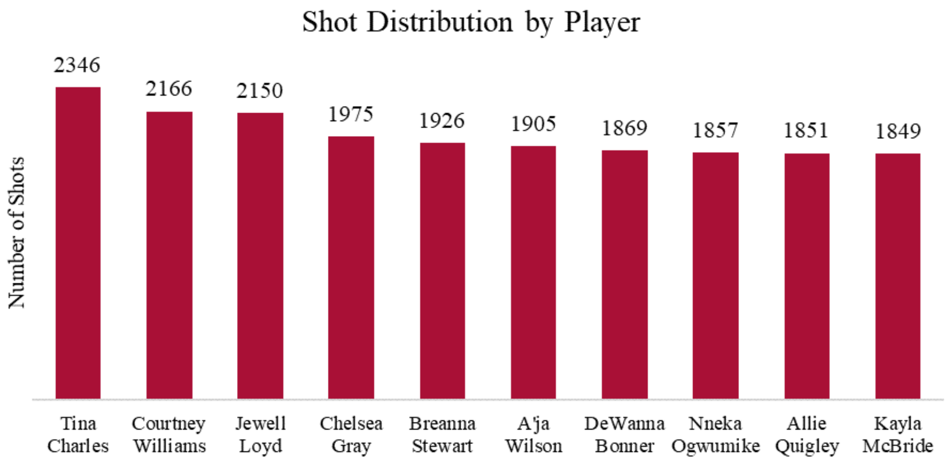
A16 - Average Made Shots per Game over Seasons: Home vs Away Teams.



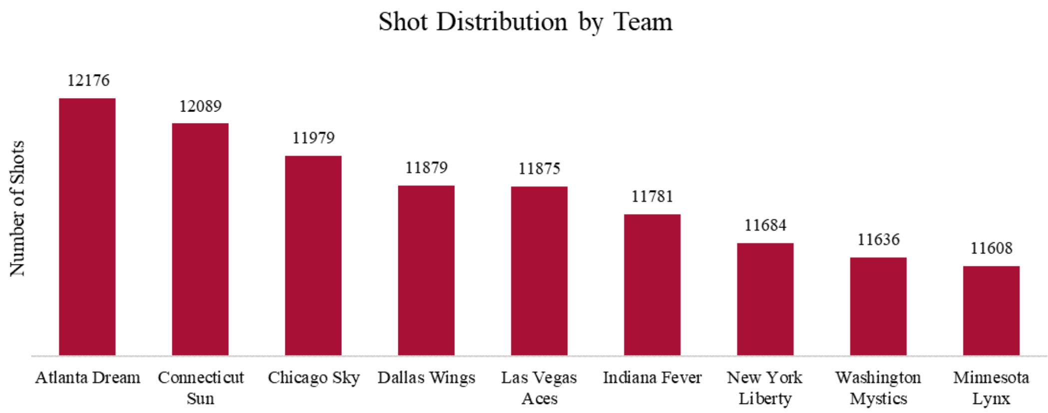
A17 - Average Missed Shots per Game over Seasons: Home vs Away Teams.



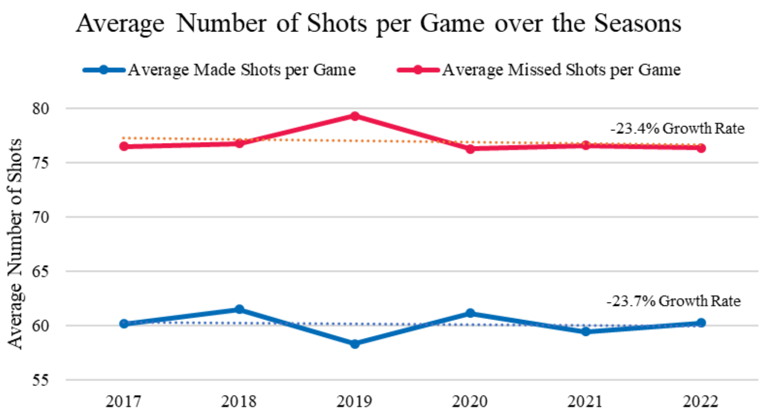
A18 -Average Made and Missed Shot Distribution by Team Ranking.



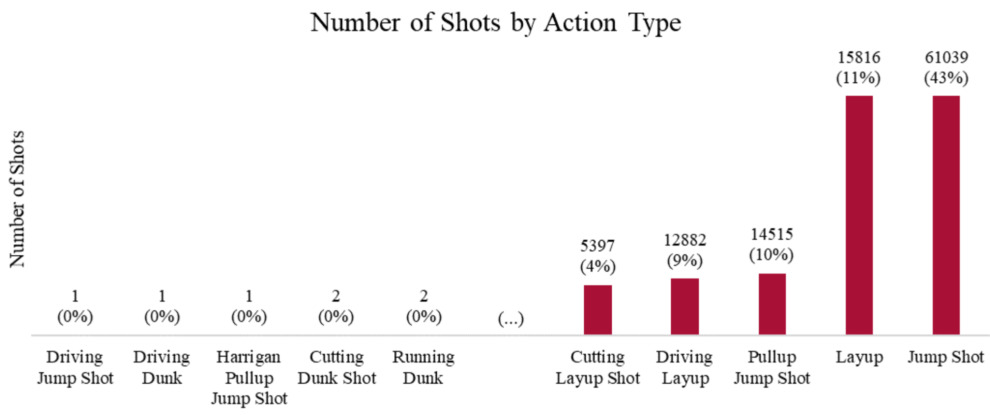
A19 - Shot Distribution by top 10 WNBA Players.



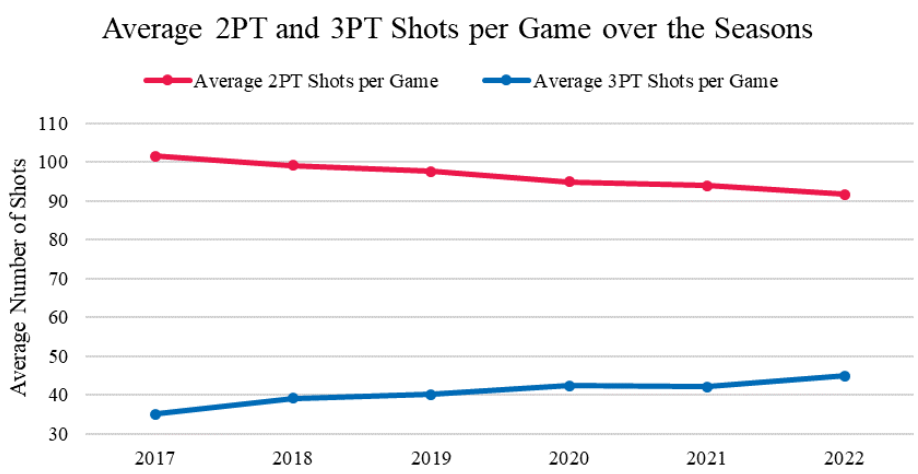
A20 - Shot Distribution by top 10 WNBA Teams.



A21 - Average Shots per Game over the Seasons.



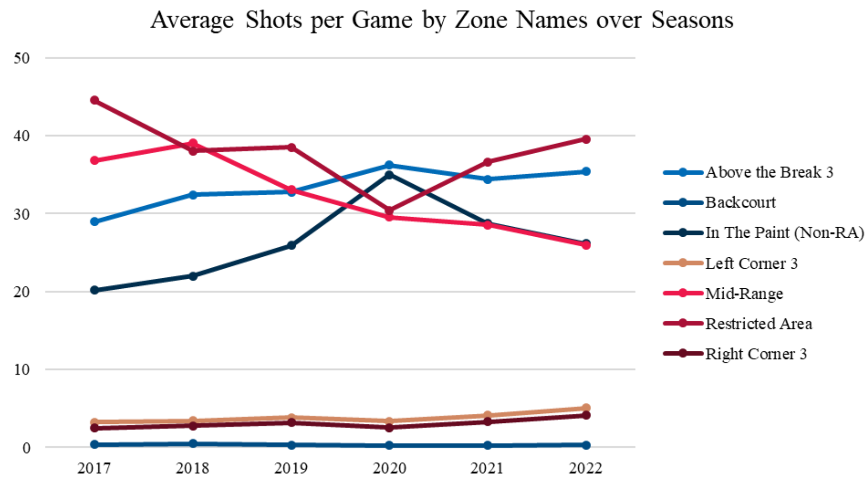
A22 - Number of Shots by Action Type.



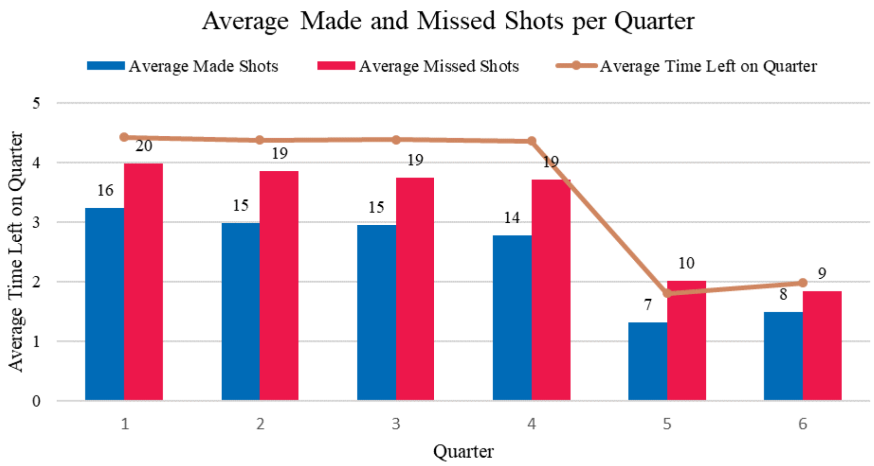
A23 - Average 2PT and 3PT Shots per Game over the Seasons.

A24- Number of 2PT and 3PT Shots per Game and respective Percentages.

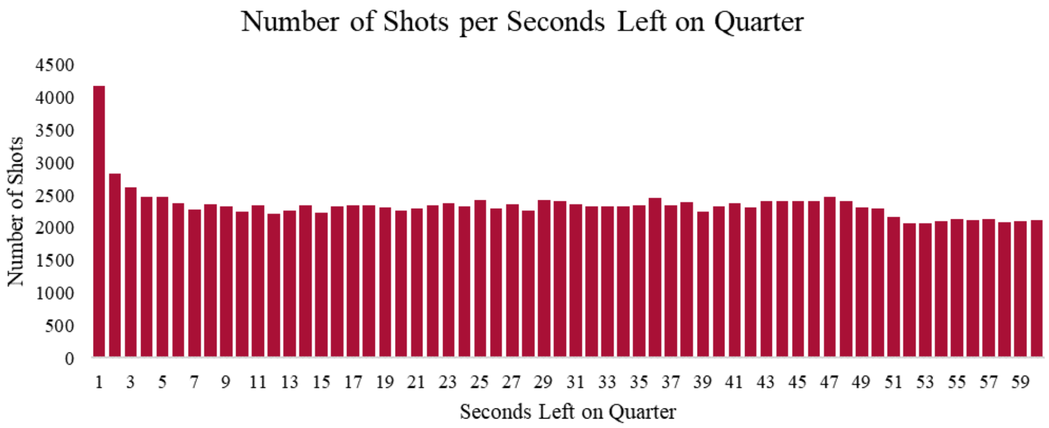
Number of Shot Types per Game and respective Percentages				
Season	**Average 2PT Shots per Game**	**Average 3PT Shots per Game**	**Percentage 2PT**	**Percentage 3PT**
2017	102	35	74%	26%
2018	99	39	72%	28%
2019	98	40	71%	29%
2020	95	42	69%	31%
2021	94	42	69%	31%
2022	92	45	67%	33%



A25 - Average Shots per Game by Zone Names over Seasons.

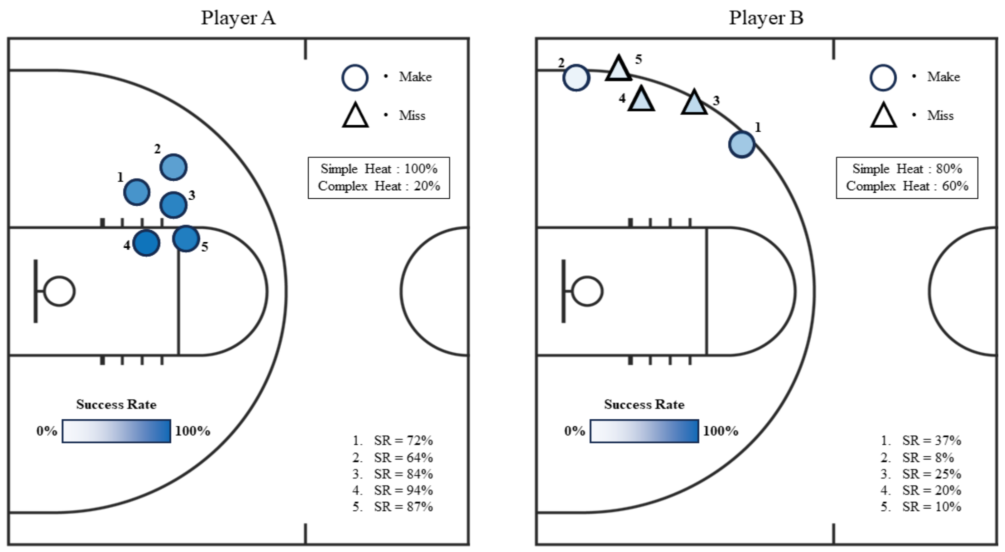


A26 - Average Made and Missed Shots per Quarter and by Minutes Left on Quarter.



A27 - Number of Shots per Seconds Left on Quarter.

11.3 Formulation of Target Variable



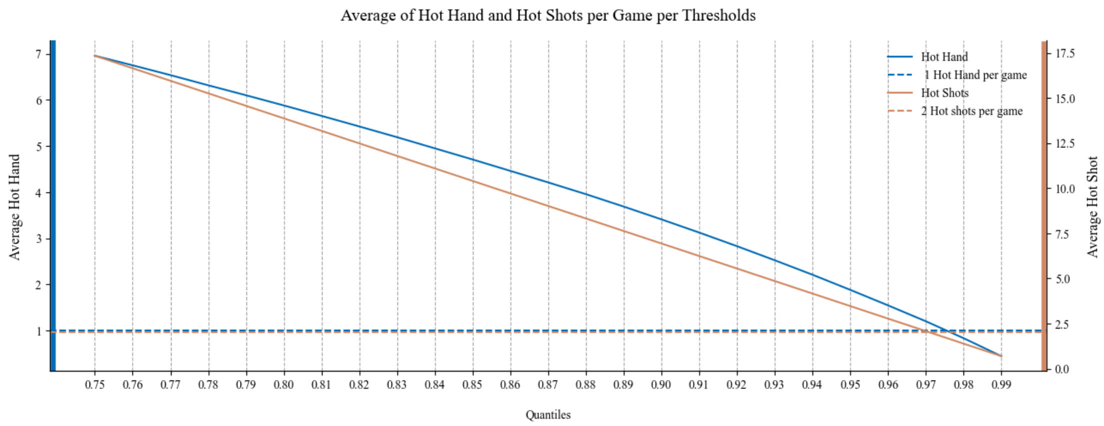
A28 - Simple and complex heat calculation, Player's shot location and their Success Rate.

A29 - Quantity and Percentage of Hot Shots based on the Threshold used to define if a Shot is Hot.

Sensitive Analysis 1 – Quantity and Percentage of *Hot* Shots based on the Threshold				
n	**Third Quantile**		**Above values of 0 of *Heat***	
	Quantity	**Percentage**	**Quantity**	**Percentage**
3	671,459	16.730%	1,246,692	31.070%
4	577,427	14.390%	1,239,250	30.880%
5	492,225	12.270%	993,214	24.750%
6	415,977	10.370%	872,670	21.750%
7	348,542	8.690%	731,383	18.230%

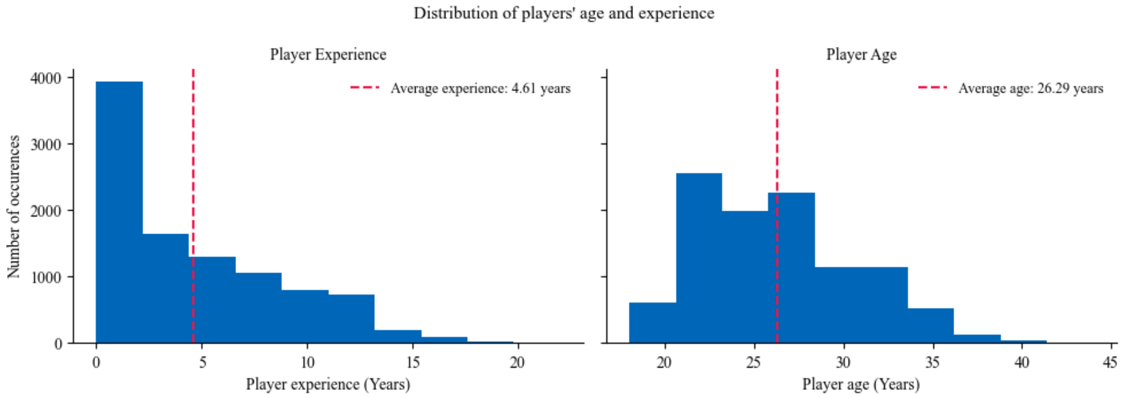
A30 - Results from the Second Sensitive Analysis to Define the Threshold to Create the Target Variable.

Sensitive analysis 2 - Average of Shots by Groups by Hot Hand column and Hot Shot column								
Quantiles	**Average by game and player**		**Average by game**		**Average by game by team**		**Average Overall**	
	Hot Hand	**Hot Shot**	**Hot Hand**	**Hot Shot**	**Hot Hand**	**Hot Shot**	**Hot Hand**	**Hot Shot**
0.75	0.3526	0.8788	6.9561	17.3375	3.4775	8.6673	0.0416	0.1037
0.76	0.3421	0.8437	6.7489	16.6440	3.3739	8.3206	0.0404	0.0995
0.77	0.3313	0.8085	6.5362	15.9505	3.2676	7.9739	0.0391	0.0954
0.78	0.3202	0.7734	6.3159	15.2570	3.1574	7.6272	0.0378	0.0912
0.79	0.3092	0.7382	6.1001	14.5635	3.0495	7.2805	0.0365	0.0871
0.80	0.2980	0.7031	5.8794	13.8701	2.9392	6.9339	0.0352	0.0829
0.81	0.2866	0.6679	5.6531	13.1766	2.8261	6.5872	0.0338	0.0788
0.82	0.2749	0.6328	5.4236	12.4830	2.7114	6.2404	0.0324	0.0746
0.83	0.2631	0.5976	5.1905	11.7896	2.5948	5.8938	0.0310	0.0705
0.84	0.2509	0.5625	4.9489	11.0960	2.4740	5.5471	0.0296	0.0663
0.85	0.2386	0.5273	4.7077	10.4025	2.3535	5.2004	0.0281	0.0622
0.86	0.2261	0.4921	4.4602	9.7090	2.2297	4.8537	0.0267	0.0581
0.87	0.2134	0.4570	4.2107	9.0155	2.1050	4.5070	0.0252	0.0539
0.88	0.2005	0.4218	3.9556	8.3220	1.9775	4.1603	0.0237	0.0498
0.89	0.1869	0.3867	3.6868	7.6285	1.8431	3.8136	0.0220	0.0456
0.90	0.1728	0.3515	3.4092	6.9351	1.7043	3.4670	0.0204	0.0415
0.91	0.1583	0.3164	3.1235	6.2416	1.5615	3.1203	0.0187	0.0373
0.92	0.1434	0.2812	2.8284	5.5480	1.4140	2.7735	0.0169	0.0332
0.93	0.1279	0.2461	2.5231	4.8547	1.2613	2.4269	0.0151	0.0290
0.94	0.1120	0.2109	2.2095	4.1610	1.1046	2.0802	0.0132	0.0249
0.95	0.0954	0.1758	1.8814	3.4675	0.9405	1.7335	0.0112	0.0207
0.96	0.0783	0.1406	1.5451	2.7740	0.7724	1.3868	0.0092	0.0166
0.97	0.0609	0.1055	1.2009	2.0805	0.6003	1.0401	0.0072	0.0124
0.98	0.0423	0.0703	0.8342	1.3871	0.4170	0.6934	0.0050	0.0083
0.99	0.0225	0.0352	0.4444	0.6935	0.2222	0.3467	0.0027	0.0041



A31 - Averages Hot Hand Shots and Hot Shots per game per Quantiles. The left y axis showcases the average hot hand shots (blue lines), and the right y axis showcases the average hot shots (brown lines).

11.4 The Role of Player Expertise in the Manifestation of the *Hot Hand* in the National Basketball Association



A32 – Histogram with Distribution of Different Levels of Experience and Age.

A33 - Distribution of Players by Level of Experience.

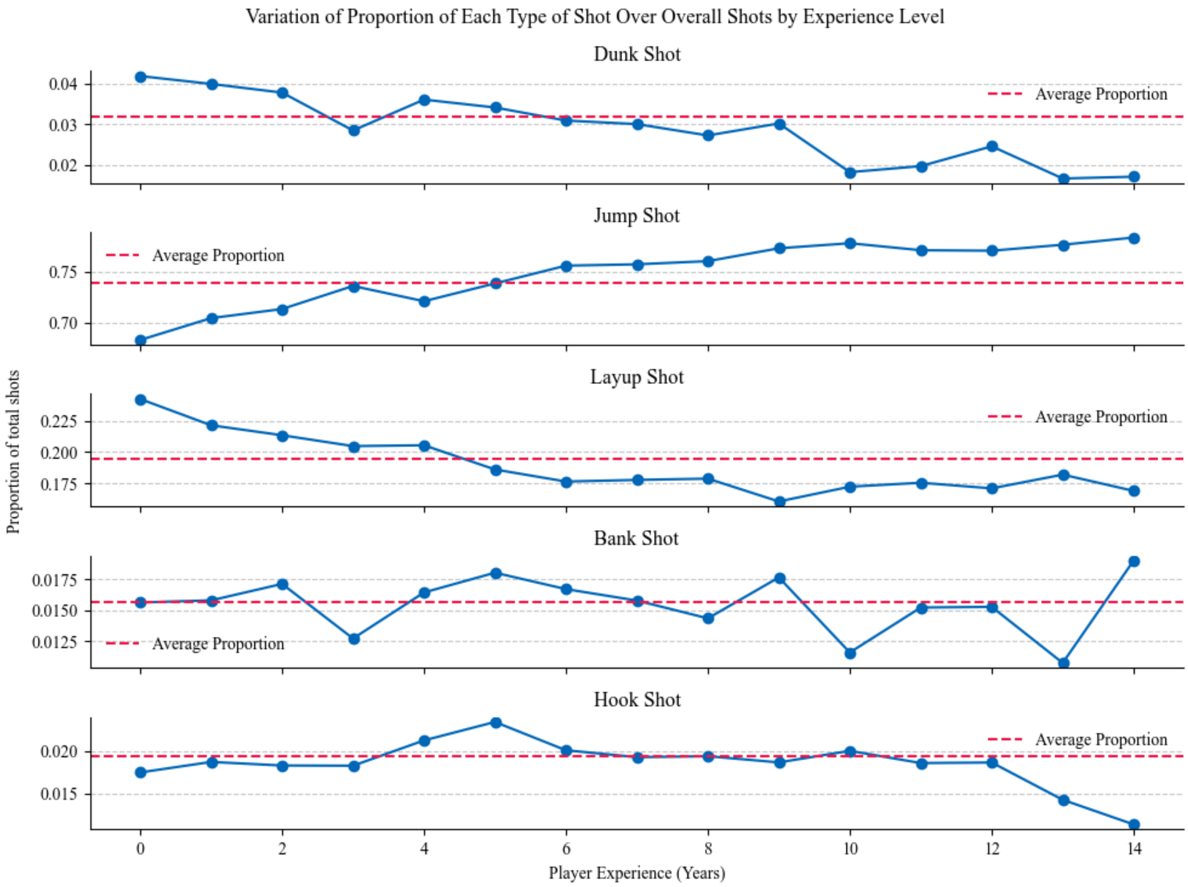
Distribution of Players by Experience	
Level of Experience	**Count**
0	1,657
1	1,251
2	1,014
3	881
4	753
5	675
6	614
7	559
8	500
9	428
10	365
11	313
12	232
13	181
14	118
15	75
16	54
17	34
18	19
19	8
20	5
21	1
22	1

A34 - Distribution of Players by Age Level.

Distribution of Players by Age	
Age Level	**Count**
18	17
19	185
20	402
21	587
22	893
23	1,066
24	1,059
25	923
26	818
27	762
28	671
29	607
30	537
31	451
32	386
33	312
34	234
35	169
36	126
37	76
38	46
39	27
40	17
41	4
42	4
43	0
44	1

A35 - Regression Analysis Results with Incorporation of Season-Specific Effects.

Logistic Regression		
	Dependent variable*: hot shot*	
	(1)	(2)
constant	-4.594***	-6.586***
	(0.023)	(0.200)
experience	0.096***	
	(0.004)	
experience squared	-0.006***	
	(0.000)	
age		0.168***
		(0.015)
age squared		-0.003***
		(0.000)
season 2005	0.052*	0.051*
	(0.028)	(0.028)
season 2006	-0.014	-0.016
	(0.029)	(0.029)
season 2007	0.049*	0.042
	(0.028)	(0.029)
season 2008	-0.009	-0.010
	(0.029)	(0.029)
season 2009	-0.017	-0.025
	(0.029)	(0.029)
season 2010	-0.026	-0.037
	(0.029)	(0.029)
season 2011	-0.028	-0.039
	(0.029)	(0.029)
season 2012	-0.110***	-0.117***
	(0.032)	(0.031)
season 2013	-0.024	-0.030
	(0.029)	(0.029)
season 2014	-0.008	-0.017
	(0.029)	(0.029)
season 2015	-0.093***	-0.097***
	(0.029)	(0.029)
season 2016	-0.069**	-0.076***
	(0.029)	(0.029)
season 2017	-0.060**	-0.053*
	(0.029)	(0.029)
season 2018	-0.058**	-0.072**
	(0.029)	(0.029)
season 2019	0.022	0.007
	(0.028)	(0.028)
season 2020	0.056*	0.042
	(0.029)	(0.029)
season 2021	-0.012	-0.022
	(0.029)	(0.029)
season 2022	0.013	-0.003
	(0.028)	(0.028)
season 2023	-0.046	-0.050*
	(0.029)	(0.029)
Observations:	3,938,116	3,966,290
Note:		*p<0.1; **p<0.05; ***p<0.01
		Standard errors in parentheses



A36 – Proportion of each Category of Shots over Total Shots by Experience Level.