Byron Burnett

# The Impact of Data Analytics in Basketball. A Case Study from the National Basketball Association

Metropolia University of Applied Sciences

Degree: IBL

Degree Program: International Business and Logistics

Thesis

Date: 10.9.2023

# Abstract

| | |
|---|---|
| Author(s): | Byron Burnett |
| Title: | The Impact of Data Analytics in Basketball. A Case Study from the National Basketball Association |
| Number of Pages: | 44 |
| Date: | 22 October 2023 |
| | |
| Degree: | Bachelor of Business Administration |
| Degree Programme: | International Business and Logistics |
| Specialisation option: | Marketing |
| Instructor(s): | Michael Keaney, Senior Lecturer<br>Kevin McIntire, Senior Lecturer |

The trends towards Big Data have influenced many business sectors of the economy. Data Analytics were introduced in sports in the early 2000's to improve efficiency, track players performance and give organizations, general managers, and coaches raw statistical data to analyse for improved decision making. Most professional sports clubs use data analytics to make informed decisions in some capacity. The NBA (National Basketball Association) also use analytics to answer questions about its players, team, and organizational needs. There is an ongoing debate about how data analytics has changed the game of basketball. The purpose of this research is to investigate the overall impact of data analytics and its current use. Optical tracking technology like SportsVU cameras have opened new possibilities to gain insight in a competitive sporting landscape. This paper will highlight how analytics have altered the shot selection of players in the National Basketball Association using play-by-play data along with visualization tools. This research provides a fundamental understanding of data analytics in basketball for spectator, organization, and athletes. Furthermore, it will highlight the practical uses, its impact, and limitations by reviewing literature of past studies. This study provides a statistical analysis of 30 NBA seasons from 1993 using team and player statistics. The results illustrate the trends of 3P and 2P shot attempt before and after the inception of analytics in the sport. To emphasize the modern-day impact of analytics in the NBA the paper examines over 427,737 shot attempts from the 2021-2023 NBA seasons. Using Pearson's correlation coefficient, the paper investigates which statistical categories have a strong relationship with wins for NBA teams.

Keywords: data analytics, basketball, efficiency, shooting, statistics

**Contents**

## List of Tables and Figures

# 1 Introduction

Sports are rooted in numerical data, whether it be through statistics, physical measurements of athletes, or the final score to decide the game, numbers are everywhere. All this numerical data can be captured, cleaned, analysed, and articulated to key stakeholders within sport organizations. The terms data analytics and sport analytics can be used interchangeably both describes processing data. According to Pykes (2022: 1) "sport analytics is the study of athletic performance and business health that provides value and insight to a sports organization."

No other sport has invested more effort in data analytics than basketball. In a 2017 Wharton School article the NBA Commissioner Adam Silver states that "Analytics are part and parcel of virtually everything we do now" (Wharton School 2017). Currently there is a competition to extract valuable information through various methods including wearable technologies, statistics, and sophisticated video technology installed in arenas to track play-by-play data during live game action. The NBA is the number one premier professional basketball league in the world with global recognition and respect. The game of basketball, especially in the NBA has gone through many rule changes and trends since the inception of the league in 1946.

The most significant change over the last 20 years, involves the introduction of data analytics into virtual all decision-making and day-to-day responsibilities within an NBA franchise. Focus areas include team strategy, player evaluation, injury prevention and business management.  The visual impact of data analytics is seen in the way the game is played, the pace of the game, the shot selection players attempt during the game, along with the spacing of teammates on the basketball court. Historically the game of basketball was a slow pace, half-court game where players attempted to score around the "painted" area near the basketball rim, known as the mid-range position on the court.

## 1.1 A Brief Description About Basketball

To assist readers who have no concept of basketball I will outline court dimensions to help follow this report more sufficiently. Figure 1 gives the dimension and court size of the standard basketball court.

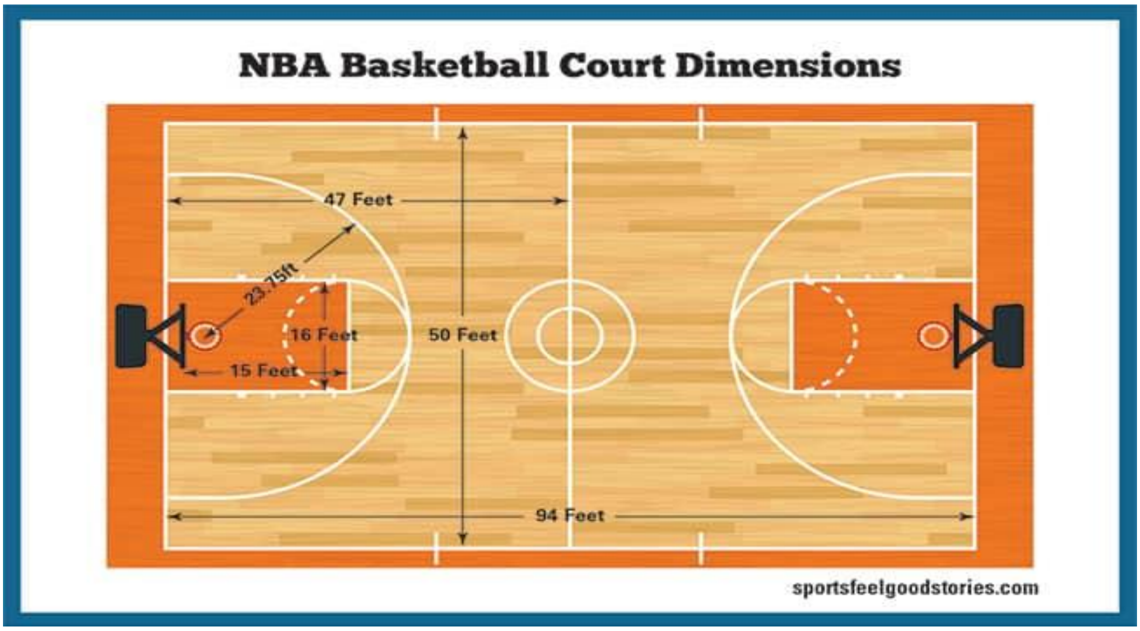


Figure 1: Standard Basketball Court (www.sportsfeelgoodstories.com).

The game of basketball is played with 10 players on the court and the objective is to score more points than your opponent. The scoring tiers are 1 point for *free throws*, 2 points for *mid-range* shots and 3-points for *long distance* shots. Basketball courts are 94 feet long and 50 feet wide. The painted area (*mid-range*) and free throw lane are 16 feet across. The 3-point line is 23.75 feet away from the rim, and the shot clock is 24 seconds in the beginning of each possession. If the team who has possession, rebounds (*offensive rebound*) the basketball during that possession the shot clock resets to 14 seconds in the offensive team's front court. In this research we will focus on the painted area (*mid-range*) and the 3-point area (*3-point range*) to describe how analytics have impacted NBA gameplay.

## 1.2 The Rise and Use of Analytics in the NBA

Figure 2 illustrates the gradual increase of points per game beginning in the year 2000 with another rise in points per-game starting in the year 2014. The increase in points is strongly correlated with 2PA (2-point attempts) and 3PA (3-point attempts) throughout an 82-game NBA season. The aim of this research paper is to (1) investigate the impact and uses of data analytics in the NBA, (2) illustrate how the shot selection has changed using NBA play-by-play data, (3) which statistical variables strongly correlates with the wins of an NBA team?

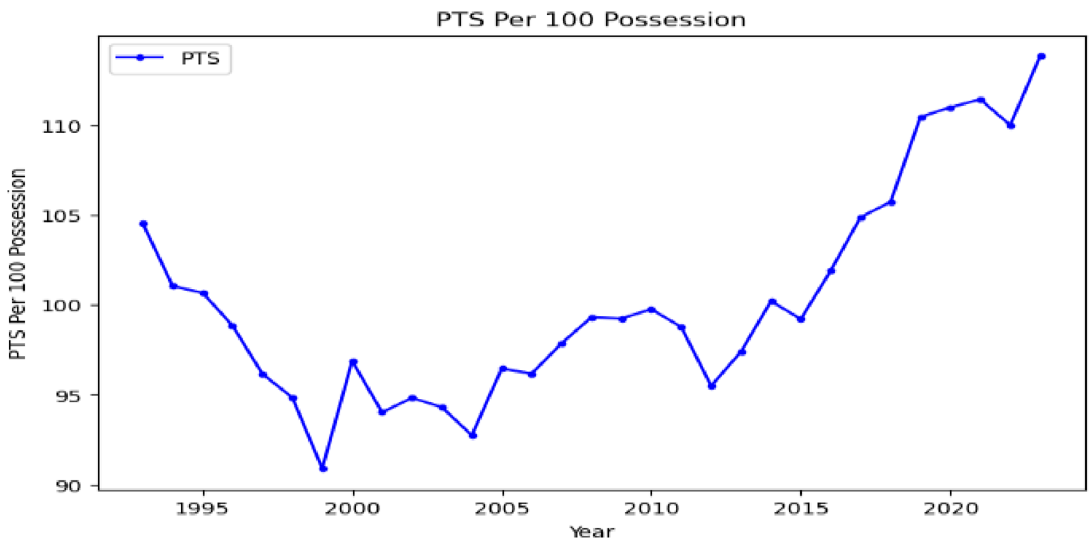


Figure 2 Points Per 100 Possessions

To visualize the shot attempt distribution over the last 30 years Figure 3 shows a sharp increase in 3PT (3-point) attempts and a gradual decrease in 2PT (2-point) attempts. The decline in 2PA and the increase in 3PA is a growing trend that has the attention of fans, coaches, athletes, media personnel and academic researchers. In the early 1990's the most frequent shot attempts were inside the 3-point line which is the mid-range area, usually anywhere 22-2 feet from the basketball rim.

Currently the most common shot type is the 3-pointer, although the 3PT shot was introduced in 1979, NBA players still did not attempt a lot, even though they had ample opportunities. The 3PT shot are considered to be difficult to make, successful attempts count as 3 points which is 50% more than 2-point conversions. This trend has produced an outcome were the incentive to shoot more 3-point shots, have awarded players with more notoriety and higher salaries, while others have been punished for not adapting fast enough. Figure 3 show the gradual increase in 3PA and the decline in 2PA.

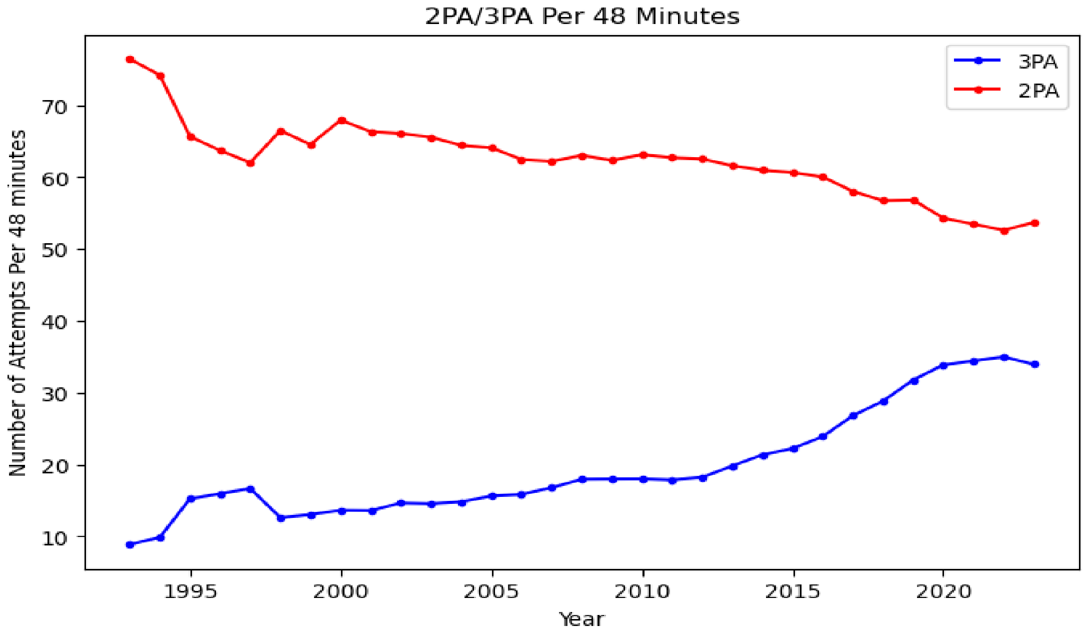


Figure 3 Number of Attempts Per 48 Minutes

## 1.3 A Shift Towards Efficiency

In a 2003 book *"Moneyball"* Michael Lewis, who detailed the work done in statistical modelling by Bill James who is known for his work on sabermetrics and finding undervalued talent was an early researcher in statistical analysis. According to the Society for American Baseball research, sabermetrics is "the search for objective knowledge about baseball" (Sabr.org, 2011). Using the work of James, researcher from other sports began investigating how those principles of analysis could contribute to their respective sport.

In basketball early researchers in the field include Dean Oliver who wrote a book in 2004 entitled *"Basketball on Paper"* rules and tools for performance analysis and John Hollinger who created the evaluation metric *player efficiency rating* (PER). Oliver and Hollinger thought it was best to evaluate NBA players on a *per-minute* and *possession* basis rather than the outcome of statistical totals. The use of analytics in basketball is rooted in finding the most efficient way to play and focusing teams' efforts on specific statistical categories that could influence the outcome of the game.

Another early adapter of data analytics was Darly Morey who became the general manager of the Houston Rockets in 2007. Morey led the first analytical movement in an NBA front office. Through his tenure as the Rockets general manager, which lasted from 2007-2019, Morey led the most data-driven franchise in NBA history. Thereafter teams began to invest heavily in staff and other technologies to enhance its understanding of analytics within a framework of winning and business insight. Figure 4 shows how many staff members are employed in the "analytics departments" of NBA franchises.

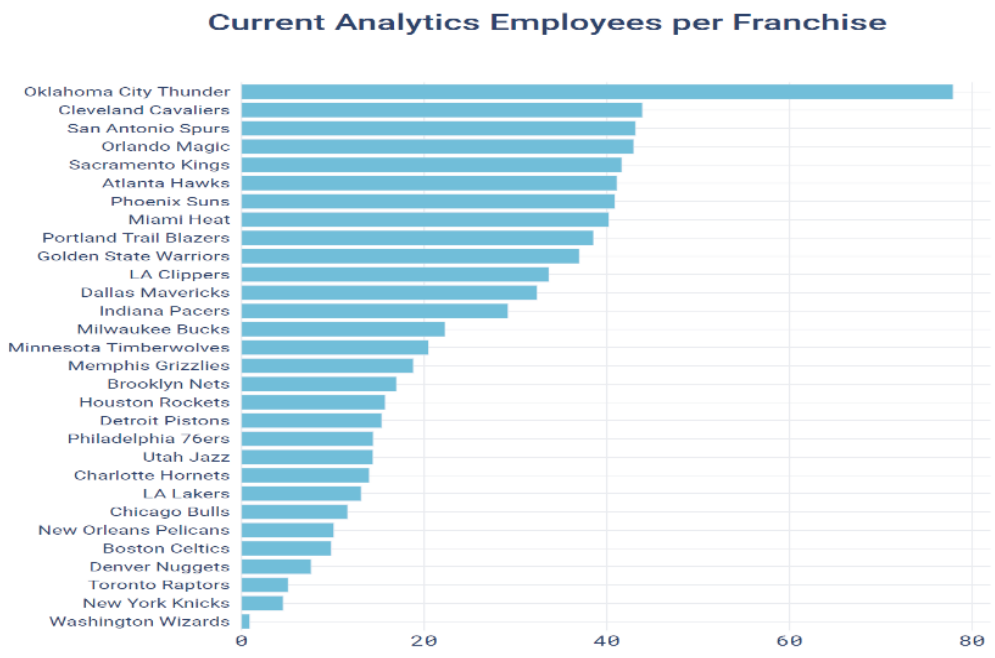


Figure 4 NBA Analytics Staff (Reveliolabs.com, 2021)

It is clear the trend of analytics is here to stay and will influence future developments in the NBA as the league moves forward. Team's strategies, business insight, player evaluation and the integration of technical staff clearly shows the interest the NBA has for the data analytics.

# 2 Literature Review

The introduction of statistical analysis with the intent to become more efficient started with the game of baseball, thereafter other sports took a keen interest in learning about the theory. Basketball a sport rich in statistical information became a reservoir for scholars interested in data science and analytics. Several researchers began to investigate the use of data analytics in basketball. These areas include the evaluation of team performance, player efficiency statistics, spatial analysis, shot selection theory, injury prevention and offensive/defensive strategy. This section highlights some of the past studies that introduced the idea and practical uses of data analytics in the NBA.

## 2.1 Statistical Analysis

A key theme in modern statistical analysis is the principle of efficiency*,* and how it can be an indicator in evaluating a team or individual performance throughout a game or even through an entire season. A common field of study in sports analytics is predictive analysis, the ability to model and predict outcomes to better help coaches, scouts, players, and general managers make informed decisions in the future. A 2006 study *Efficiency in the National Basketball Association: A Stochastic Frontier Approach with Panel Data,* investigates if a team is playing to its full potential and does the quality of players and coaches influence if a team is playing to its full potential. Their study is based on using the Stochastic Frontier Model developed by Aigner et al (1977) and modified by Battese and Coelli (1995).

The variables used in the study are *field goal, free throws, offensive rebounds, defence rebounds, assist, steals, turnovers, and blocks.* The finding of this research concluded that both the players quality and the quality of coaching influence the overall win total during the NBA season. They found that scoring and rebounding increased the numbers of wins and turnovers decreased the total number of victories, while defensive statistics such as steals and blocks also provided positive impact on win totals (Hofler and Payne 2006: 283).

Although the study was based on team's efficiency and how players could affect winning totals it did not isolate individual statistics that played a role in overall win totals. The research highlights how the quality of players along with coaching quality determines the full potential and efficiency of a team. It contributes to the study of efficiency in terms of teams but does not give sufficient information about individuals players.

Kubatko et al (2007) journal "*A Starting Point for Analysing Basketball Statistics"* provides a list of basics variables and methods used in modern statistical analysis. Their combined research laid the foundation that contributes to the development of modern statistical analysis. The authors state that the possession and per-minute formulas are an acceptable starting point for analysing variables in box-score statistics. Some of the concepts include *possessions*, *offensive and defensive ratings, per-minute statistics*, *pace rate*, *four factors for winning* and many more.

$$\boxed{POSSt = FGAt + 0.44 \times FTAt - OREBt + TOt.}$$

The above formula calculates the number of possessions a team has during a 48-minute game in the NBA. From these methods other statistical models can be formulated. One example is offensive/defensive rating, which uses *the possession concept* to find out the per-possession efficiency as a "rating" (Kubatko et al 2007). The research points out that since possession are equal, meaning that each team possesses the basketball an equal number of times, measuring efficiency through offensive/defensive rating recognizes the quality of each possession respectively. Below is the formula used for offensive/defensive rating:

| |
|---|
| Offensive Rating (ORtgt) = PTSt/POSSt × 100 |
| Defensive Rating (DRtgt) = PTSo/POSSo × 100 |

Kubatko et al, points out the importance of per-minute statistics, this allows for a fairer comparison of starters and reserves. Players who do not play a lot of minutes (*rookies, older veterans)* do not qualify if their total playing time is between 500-1000 minutes in an NBA season (2007: 9). The formula for *Per-Minutes Statistics* is the following:

$$\boxed{PTS40p = PTSp / MINp \times 40}$$

Two other important statistical analysis the authors identify is *true shooting percentage*, and *effective field goal percentage*. As Kubatko et al (2007: 10) writes "*field goal percentage (FG%)* does not add made three pointers, or free throws to those statistics". Effective field goal percentage considers made 3PT shots, and true shooting accounts for made 3PT and FT(free throws). Below are the formulas use to calculate both percentages:

| |
|---|
| $FG\% = FGM/FGA.$ |
| $eFG\% = (FGM + 0.5 \times 3PM)/FGA.$ |
| $TS\% = (PTS/2)/(FGA + 0.44 \times FTA).16$ |

Other key findings in this study are the four factors, this concept seeks to breakdown the offensive and defensive ratings in statistical terms. The four factors concept of (Oliver 2004) examines the key factor which a team needs to capture wins in a basketball game, in other words the more factors a team accumulates the better their chances in winning the game. The four factors are the following:

| |
|---|
| Effective field goal percentage (eFG%t). |
| Turnovers per possession (TOt/POSSt). |
| Offensive rebounding percentage (OREB%t). |
| Free throw rate (FTMt/FGAt). |

This study provides researchers with agreed upon concepts and methods to analyse individual and team performance on a per-minute, per-possession basis. Many concepts and methods from this study remain useful for statistical analysis, as mentioned before this journal was intended to give researchers a "starting point" for analysis in the future. Modern NBA stat sheets record the eFG% (effective field goal), pace, TS% (true-shooting) and +/- ratings.

Modern statistical analysis tends to focus on predictive analysis to assess and grade an individual or team performance. Sarlis and Tjortjis, (2020: 1) points out in their research paper *Sports analytics – Evaluation of basketball players and team performance,* "that player performance predictions using current and past data has gained attention especially in basketball". Their research is intended to review past studies done in advanced basketball metrics used in the NBA and Euroleague matches. The overall purpose of the paper is to benchmark current performance analytics used in studies that evaluate teams and players.

Sarlis and Tjortjis (2020) explain the increased use in performance metrics that could add value and insight about a basketball player's behaviour. The authors believe new findings in technology could increase data collection which could spark new methods of analysis about players and teams for increase opportunities in predicting outcomes.

This increase would provide more information to staff member and coaches when making decisions about offensive, defensive and preventive measure such as injuries management. Sarlis and Tjortjis identifies, that there is a trend to transform sports analytics from a descriptive model to a predictive model to gain even more insight and understanding about the game in general. Below in Figure 5 is a chart that describes the correlation between power of sports analytics, and it's required skills.

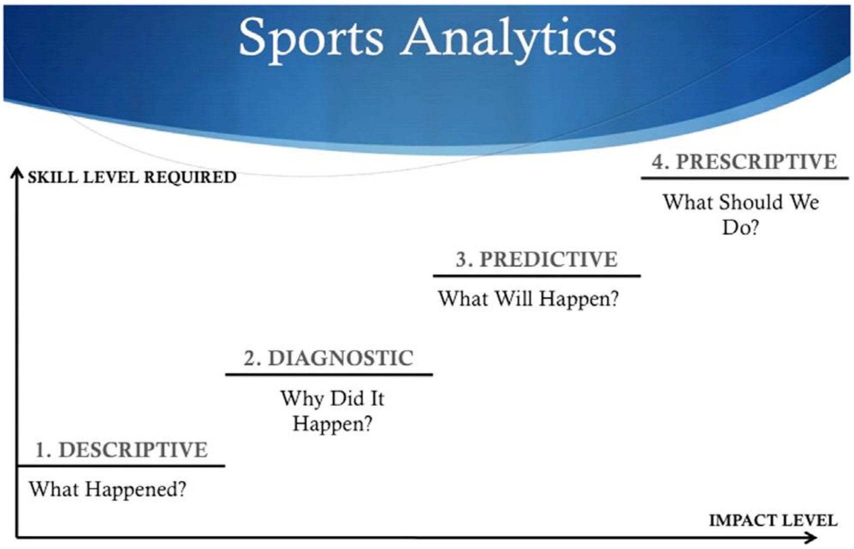


*Figure 5: Sports Analytcis vs Skills (Sarlis and Tjortis 2015)*

This research paper provides the definition and explanation of the useful basketball metrics for the offensive and defensive side of the game. The authors highlight the importance of data mining in sports and give a working definition for the process. According to Sarlis and Tjortjis (2020: 2) "data mining is the discovery of patterns or rules from large amounts of data, that leads to the search of valuable information within that data". Sports franchises use DM techniques to find advantages to gain an upper hand against their competitors. There exist a few sophisticated techniques used to predict outcomes in data analysis, those include Bayesian Networks, Linear/Logistics Regression and Neural Networks.

Most elite athletes tend to score high in achievement in performance metrics while reserve players tend to rank below average in the same areas. This shows proof that the metrics have credibility in assessing a player's true ability. Below in Table 1 is a sample of metrics that are used worldwide in the NBA and Euro leagues:

*Table 1: Examples Anlytics Metrics (Sarlis and Tjorjtis 2016)*

| **Metric** | **Explanation** |
|---|---|
| Plus/Minus (+/−or PM) | measures the impact of a player in a game |
| Adjusted Plus Minus (Adj +/−or APM) | is the player statistics for rating |
| Real Plus Minus (Real +/− or RPM or RAPM) | Inclucded the Real Plus Minus wins (RPM Wins) |
| PIPM (Player Impact Plus Minus) | is another version of plus-minus that adjust the box-score |
| Player Impact Estimate (PIE) | calculates the overall player's contribution |
| CARMELO | focus on win forecasting topics |
| Expected Possession Value (EPV) | evalute and quantity values that makes a player to during the game |
| Wins Above Replacement (WAR) | refelcts a combination of player's projected playing time and his projected productivity |

The study is limited because basketball remains unpredictable in nature, the events that make it impossible to predict are factors of luck, the physical fitness of players, behavioural tendencies, injuries, and quick shots taken in offensive sets. The study provides value because it outlines the predictive basketball analytics that are accepted by a wide community of analyst. Each year new researchers improve old metrics by adjusting weights and developing new concepts. Behavioural and psychological studies could add insight into a player's thought process, links could be drawn from the psychological side into the physical performance side for better indicators of success.

## 2.2 Spatial Analysis in Basketball

As data became more prevalent in the NBA coaches began to use shot charts to details where player likes to shoot the basketball. A basic shot chart could summarize a player's shooting tendencies with make or miss indicators using an image of a basketball court. Spatial analysis started to gain traction in the NBA at the early stages of the analytics movement. Reich et al (2006: 1) study, *Spatial Analysis of Basketball Shot Chart Data* "develops a hierarchical spatial model for shot chart data that allows for spatially varying effects using different covariates".

The authors developed a model using covariates that potentially could affect a player's shot behaviour or habit. Each covariate carried a frequency value that could be weight against his decision to take shots during the game. Example of covariates include home game, team rest for 2 days, opponents average blocks, opponent allowed field goal percentage and whether it was the second half or overtime. The goal of the study was to develop a statistical model for analysing shot chart data using covariates and spatial correlation in data, which transformed shot charts from descriptive, to one that is wholly inferential. (Reich et al 2006: 11)

A 2012 research called *CourtVision: New Visual and Spatial Analytics for the NBA* introduces "new analytical techniques designed to quantify, visualize and communicate spatial aspects of NBA performance with precision and clarity" (Goldsberry 2012: 1). The study introduces new ways to quantify the shooting range, with charts that revel tendencies in NBA player's shooting abilities. The main goal of the study is to inspect shot site performance using spatial awareness and determine which NBA player has the best spatial shooting behaviour. As Goldberry (2012: 2) states "spatial analysis and visual analytics greatly enhance shooting evaluation".

The data analysed in the study is compiled from the 2006-2012 NBA seasons, Goldsberry charted every field goal attempt using (x, y) coordinates including the shot location, player name and shot outcome, Figure 6 and Figure 7 illustrate the final output of the visuals.

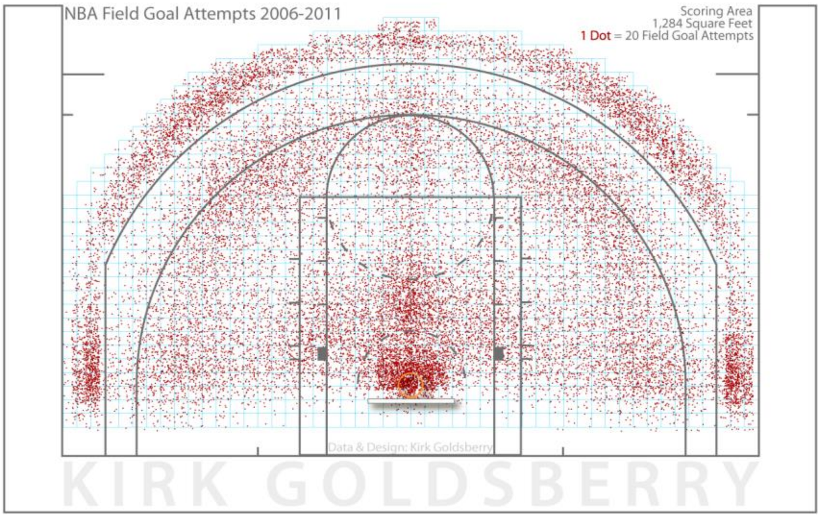


*Figure 6: Example of Spatial Analysis Chart (Goldsberry 2012)*

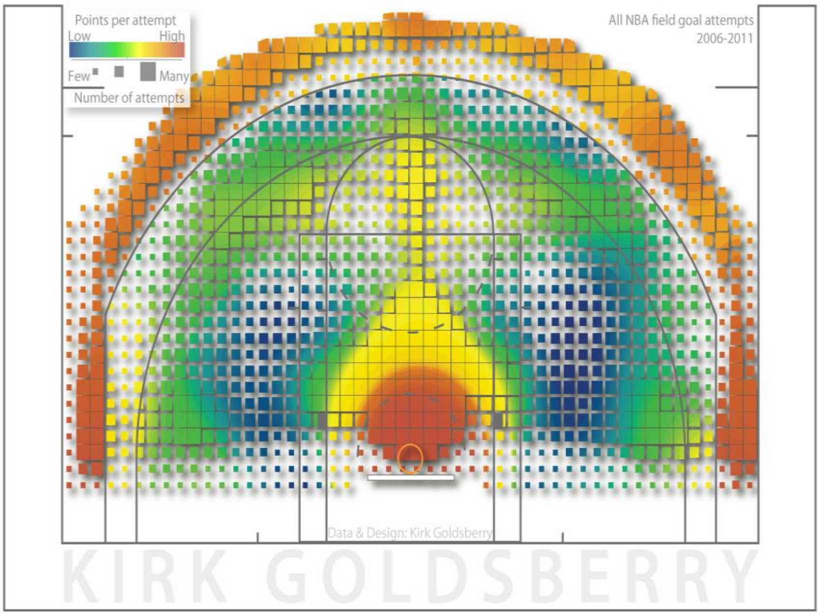


*Figure 7 Example of Spatial Analysis Chart (Goldsberry 2012)*

After collecting the data Goldsberry used 2 metrics to describe a player's shooting ability, *spread* and *range*. Spread was the count of at least 1 field goal attempt in a cell. The range identifies if a player scores at least 1 point per attempt in a giving area. The players who have to most point per range could be considered as a diverse shooter in the NBA with an ability to shoot from various positions. The finding of the study revealed that Kobe Bryant had the largest spread percentage with 83.4% and Steve Nash had the highest range percentages with 31.6%.

To summarize, the study contributes to the growing field of spatial analysis and visual analytics using maps and charts to make inferences about a player's ability to shoot and which location the player prefers. The study is limited because it does not give information about the position of defenders on the court which could help coaches have better knowledge about predicting defensive strategy from opposing teams.

As advancement and research in spatial analysis continued the NBA took note. In a 2013 article "*How the NBA's SportsVU ball and player tracking tech changes the face of sports"* Plafke writes "the NBA now has access to SportsVU, a camera system so technologically advanced that it has opened the door for Big Data to invade and shape the NBA" (2013: 1). The article details how the NBA installed the groundbreaking technology in all arenas as it provides teams and spectator with an immense amount of information than previously possible.

SportsVU created by STATS LLC, is a sophisticated camera system, with the ability to track player movement 25 times per second. Plafke (2013:3) illustrates how the SportsVU system will have six computer vision camera and software, that identifies 11 data points per game visualizing 10 players and the basketball. Other findings in the articles highlights that the camera will record X, Y coordinates of the players and X, Y, Z coordinate of the basketball 72,000 times per game. Mr. Plafke writes that the system will identify spatial patterns on the court between players, also which player performs in certain geometrical positions.

Sampaio et al (2015) research on playing tracking data, demonstrates how much new information is available with the use of player tracking data. Their study *Exploring Game Performance in the National Basketball Association Using Player Tracking Data* used data from the 2013-2014 NBA season to compare two different profiles of players: (1) NBA starters who play large minute and (2) NBA reserves who play a smaller number of minutes. The authors explain how the data tracking technology opens more possibilities in understanding game performance by highlighting key features of the technology.

According to Sampaio et al (2015: 3) "movement patterns (kinematics) complement variables the physiological side, technical (game action) and tactical (individual/team behavioural patterns) leads to a better description and understanding of player's behaviour". Some of the game action that SportsVU provides during their recordings are pull-up shots, catch and shoot, close shots, drives, passing-variables, and touch-variables.

The research in this section contributes to the advancement in spatial analysis studies. Early research used shot chart recording methods with covariates to determine shooting habits. Goldsberry followed that research by creating *CourtVision* an advance way in analysing spatial information from statistics with range and spread metrics. Lastly advancement in analytics lead the NBA to invest in SportVU technology and install sophisticated camera networks to capture data 25 times per second. This research is key to the advancement in future concept of capturing spatial information that provides coaches and technical staff with useful information.

## 2.3 Analytics Impact on Shot Selection

Shot selection is based on which type of shot (3PT or 2PT) should a player attempt within an offensive possession. Early researchers in shot selection theory found links to behaviour theory. In a 2009 study *Generality of the Matching Law as Descriptor of Shot Selection in Basketball* identifies that the generalized matching law corresponds with shot selection (two vs three-point field goals). Alferink et al (2009: 1) details a study from Vollmer and Bourret (2000) "that suggest behaviour allocation (a measurement of a specified shot proportion two-points and three-points field goals) had a significant correlation with relative reinforcement frequency (number of shots made of each type)". Below is the formula for GML (*Generalized Matching Law*) according to the research:

| GML = General Matching Law |
|---|
| log(Bx=By) =a log(rx=ry) + log b |

$B_x$ and $B_y$ are frequencies of occurrence in the behaviours while $r_x$ and $r_y$ represents the frequencies of reinforcements.

> According to the study (1) most of the variance in behaviour allocation which are the ratio of shot taken varied systemically with the ratio of shots made. (2) Sensitivity (a parameter) usually less than 1 (undermatching) in covariation between behaviour ratio (shots taken) and reinforcement ratio (shots made). (3) There was heavy bias for taking three-point shots, across all shot making ratios, more 3-point shots were taken than were predicted (Alferink et al 2009: 295).

The authors found that shot selection matching varied with a system for players who were on successful and unsuccessful teams, which were highly dependent on the level of play and if the player was a starter rather than a reserve.

The research paper concludes that their experimental study involved players from hundreds of teams, while all past analyses represented a pool of players. Study one and study two included 320 NCAA teams for two experiments and a pool of individual NBA players for the third study. Alferink et al (2009: 606) summarizes by "stating that GML accounted for most of the variance, demonstrating the reliable describing basketball shot selection process". This research is key in understanding shot selection behaviour including the reinforcement and bias in taken certain shots.

In 2011 Brian Skinner wrote a journal entitled *The Problem of Shot Selection in Basketball,* the author explores "when a team should shoot and when they should pass up the shot by considering a simple theoretical model of shot selection process" (Skinner, 2012: 1) This study improves the following research of (Alferink et al 2009) in using a model that fall randomly in a uniform distribution. Shot selection theory is grounded in the probability of shot making. In the beginning of each possession, the ball moves throughout the court between teammates. Eventually one player will have to decide when to shoot the best quality of shot in the most efficient way possible before the 24 second shot clock ends.

Skinner's study is theoretical in nature, he highlights this throughout the paper as he writes "much of disagreement between observed and theoretical optimum shooting rates can be attributed to an inaccuracy in the theory's assumptions" (Skinner 2012: 7). The research done in shot selection present many pitfalls because the teams and players make and take shots from a reaction of circumstances on the court not from a theoretical decision. Further studies could use a probability model to predict when a shot can be taken rather if it can be decided.

## 2.4 Other Areas of Impact

Data Analytics have provided coaches and staff members with valuable insight and information to make better decisions about vital areas of concern within a sport organization. Another area in basketball where analytics play a role is injury management. Spectators and media personnel have come to know this process as load management. According to (Heindl, 2022) load management "is the practice of monitoring and restricting a player's on-court activity to reduce their injury risk".

Research on specific use in data analytics that pertains to injury prevention is limited in scope. According to Bishop (2023) the NBA is interested more than ever in the use of technology that prevent injuries and keep players active during the season. Mr. Bishop found that wearable technology helps to track heart rate, how a player is moving, and exertion levels during physical activity. Other devices include motion capture that deploy cameras that track players movement patterns, which provides enough information to reduce workloads. Although, this technology is accessible to NBA players currently it is not permitted for use in official NBA games. Data analytics play a key role in injury preventing, Bishop (2023) points out that analysing data on specific players and collecting injury history, clubs can identify trends that could predict future injuries.

There is a significant gap in analytical research that evaluates defensive play. Although offense (*scoring*) is a key component in being successful in basketball the defence (*prevent from scoring*) is also equally important. A study by Goldsberry and Weiss (2017) *Dwight Effect: A New Ensemble of Interior Defence Analytics for the NBA,* attempts to introduce new spatial and visual analytics that assess and describes how effective interior defence is in the NBA.

The paper highlights the opportunities and challenges in measuring defensive ability using optical tracking data. The objective of the study is to understand key aspect of interior defence, and identify the challenges associated with measuring defence as new forms of data emerge in the NBA. Several limitations are associated with statistics in basketball, one key limitation is their relatively low explanatory ability (Goldsberry and Weiss 2017).

The authors introduce "spatial splits this concept was borrowed from baseball's "triple-slash" lines. Spatial splits identify shortcomings and explains how players and teams perform within court space, that gives analyst a better idea of their true value. The study was conducted with player tracking data provided by STATS LL (SportsVU) that provided location data, shooting habits, and shot outcomes for 75,000 NBA shots from the 2011-2013 season respectively. According to the authors they introduced two variation of shooting splits: *frequency splits* and *efficiency splits.* The two splits represented a shooting percentage value from 3 locations: close-range, mid-range, 3-point range. Frequencies splits recorded details on shot distribution while efficiency splits focused on how efficient a team or player shot from a specified zone.

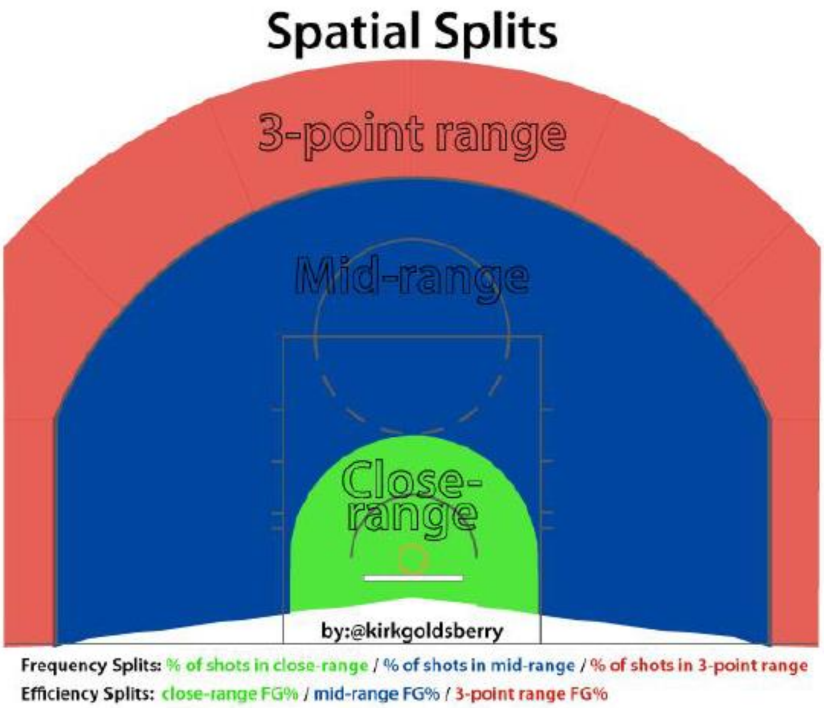


*Figure 8: Spatial Splits (Goldsberry 2013)*

Although the splits were insightful in describing offensive shooting ability in this study it was used to evaluate the defence. The findings conclude that when measured in terms of the *"a defender's location"* that's approximate to the offensive player attempting a shot.  Two key insights were discovered first it identifies reduced shooting efficiencies close to the basket. Secondly it exposed a reduced shooting frequency closed to the basket, with an increased number of attempts in the mid-range and 3-point areas. Combined this translates into fewer offensive rebounds, shot attempts and extra possessions.

In *Counterpoints: Advanced Defensive Metrics for NBA Basketball,* the authors try to bridge the gap between offensive advanced statistics and defensive statistics. For example, due to the widespread use of analytics searching for offensive metrics such as

PER (player efficiency rating), plus/minus stats, and other possession base statistics would be easily accessible with a simple google search. Looking up advanced defensive metrics hardly exist. Franks et al (2015) bridges the gap between, by formulating new defensive metrics that quantifies defensive play.

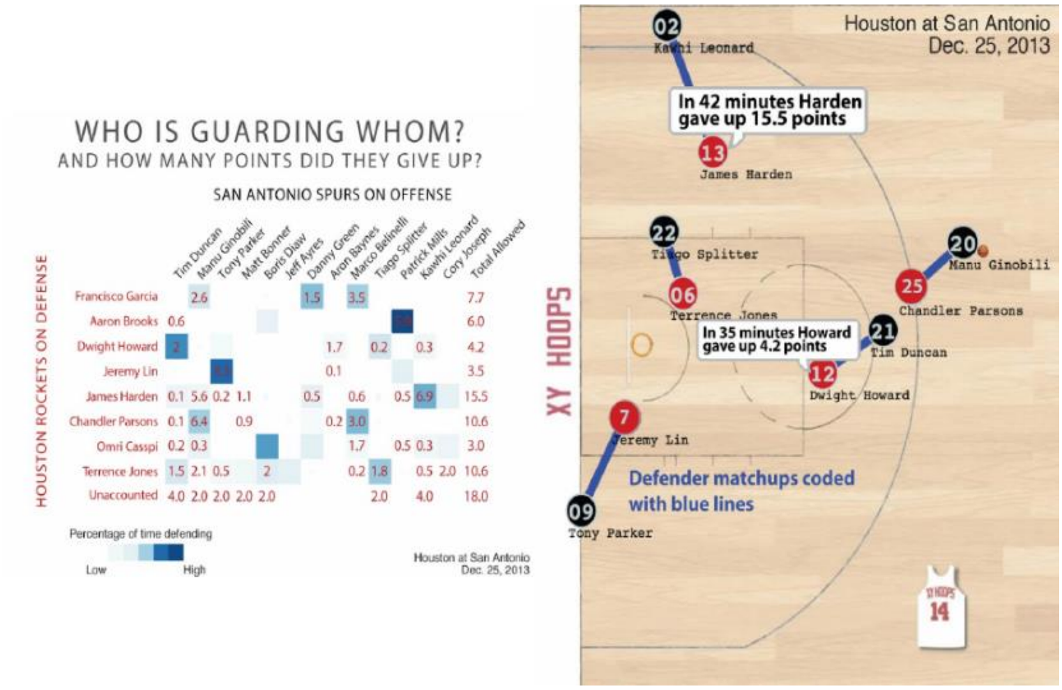


*Figure 9: Defensive Metrics (Franks et al 2015)*

Figure 9 describes the "matchup box score" which visuals the amount of time a player spends guarding an offensive player during the game. Counterpoints are assigned to certain fractions. The "matchup estimation creates the ability to sum up the defender's responsibility at multiple stages of offensive possessions" (Franks et al 2015: 2).

Thereafter the authors can track defensive matchup's better using matchup information alongside a spatial regression model. Five new findings emerge (1) *volume score,* which determined how many shot a defender faces, (2) *disruption score* reducing effectiveness, (3) *defensive shot charts* and inverse of offensive shot charts, (4) *shot against* weighed average shots per 100 possessions and (5) *counterpoints* weighted average of points scored against that defender.

The purpose of this paper is to quantify traditional defensive statistics such as blocks, steals and into a more inferential statistical category. This study improves and contributes to study of defensive analytics, it provides a starting point for quantitative analysis on a defensive players full ability.

To conclude the authors, concede that evaluating defence is a difficult task. The use of optical tracking data is important and groundbreaking but reducing players to x, y coordinates do not capture their full abilities as humans. They are running, jumping, and moving laterally at unusual speed not captured by data tracking hardware. This research is key for understanding the limits and value of player tracking in the NBA, it also exposes the need for more research into the uses and true benefits of this technology.

Coaches and players need to be in sync to execute strategies and tactics to be successful on the court. The player's attention is focused on basketball needs such as scoring, rebounding, defending and being cohesive as a unit. The coaching staff primary function is to create practice plan, choose the starting lineup and formulate a successful game plan to win the game. In a 2020 journal *Identifying group contribution in NBA lineups with spectral analysis* explores ways to quantify the contributions of groups and individual players within a team. Their approach involves the use of spectral analysis, which borrows from algebraic signal processing. Spectral analysis is based on algebraic signal processing, a method that the machine learning community is currently embracing (Devlin and Uminsky, 2020).

The study involves the use of spectral analysis using NBA play-by-play data from the 2015-16 season. Devlin and Uminsky (2020: 2) state that "the premise of spectral analysis in a basketball context is: team success can be viewed as a function on lineups and such functions have rich structure which can be analysed for data analytic insights". The findings reveal that spectral analysis proposes new way in understanding the group effects in basketball.  This research presents ways that spectral analysis help in identifying lineups and answering questions like, are these group of players compatible and which players do not perform well together.

## 2.5 Potential Limitation in Data Analytics Use

Research on the limitations of data analytics does not exist currently. The implementation of data analytics within an organization could be considered a philosophy of an organization instead of a principle. Certain organizations were quick to embrace analytics, and some were slow implementing its usability within their team structures. Internet articles, blogs and podcasts present most of the resistance in opinion when talking about data analytics. NBA players seems to be first in line when it comes to the push back against data analytics.

In a 2019 article *Kobe Bryant explains why he hates analytics in NBA* the author reveals Kobe Bryant's opinion:

> "I hate it," Kobe said. "It's ridiculous. What numbers don't tell you is they don't tell you the emotion. I don't like analytics. "You see the numbers, but the numbers don't tell you *how* or *why* they are the way they are. You must be able to feel that, to sense that. Tendencies." (Brown 2019).

Kobe Bryant a five-time NBA champion who opinion holds weight in the NBA must be taking into consideration. Analytics provide valuable insight on what happens, but it does not describe how it occurs. Several factors go into winning a basketball game including the practice regiment, team chemistry, home/away games, rest, psychological factors, and luck. Data analytics do not account for these factors. Calculating efficiency, is based on statistical data that can be hard to access, analyse and able to draw conclusions from.

Another area where analytics has limitations in basketball is game play. Aesthetically the game of basketball has no diversity in playing style. Most players are attempting high percentage shots. These shots include a high number of free throws, shot within 2ft-4ft of the basket and 3-points shots. The NBA has incentivized players to shoot more with rule changes and practice regiment that encourage players to shoot these types of shots.

As pointed out in the literature review data analytics is not balanced in its approach to analysing basketball. First it focuses a lot of attention on offensive metrics and ignores the defensive side of the game. Several correlating factors are involved and some uncorrelated which have no impact on wins. One such correlation is win/loss percentage and eFG%, meaning that being more efficient isn't necessarily correlated with wins.

## 2.6 Conclusion

To conclude data analytics is a growing field of study in the context of analysing basketball. This literature review documents some of the important sources in research regarding statistical analysis, spatial analysis, shot selection theory, player, and team performance evaluation. Studies on statistical analysis, seem to be the area where the most consistency is prevalent. Concepts such possessions and per-minute estimates are use throughout research to quantify the ability of teams and individuals. Spatial analysis research is limited because it doesn't give the full account of on court activity, identifying space and the position of all 10 players require more technological development to record player location information. Advanced defensive analytics research has the biggest gap in terms of available material.

Although basketball is primarily an offensive game with a focus on scoring points, defence plays a major role in strategy and execution of an overall gameplan. Future research in this area could investigate how to evaluate a defensive player more accurately. Injury prevention research is also limited in scope, one reason is the reluctance of the NBA to allow players to play with wearable technologies during live matches. The reason the NBA does not allow the use of wearable technology during game is related to privacy issues dealing with disclosure of health outcomes.

Data analytics remains in the infant stage of development as new research comes forward new findings will appear and stand to be evaluated. Therefore, this literature review provides a broad set of knowledge that pertains to the current use of analytics in the NBA. Fresh ideas and concepts will increase the knowledge and understanding of this complex subject. As data mining techniques increase more information about player's abilities will be discovered. Research in data analytics will only continue to progress, in the future researchers should find ways to articulate this information and make it accessible not only to coaches, players and general managers but to spectators and fans as well.

## 2.7 Theoretical Framework

A correlational research design was implemented for this project, which investigates the relationship between two or more variables. Data analytics was introduced to encourage teams and players to be more efficient in scoring opportunities. As found in the literature review to measure efficiency, researchers defined the possessions concept which provides a basis for an accurate estimation for evaluating teams and individuals. A correlation design is relevant in quantitative studies because its help to find relatedness factors between two or more variables. A common mistake in correlation research is two variables that seem to have a positive relationship, but other variables not accounted into analysis influence their outcome.

This study uses correlations to explore NBA box score variables that influence factors of efficiency as defined in the literature review. One such measure is Pearson's correlation coefficient. According to Abbot (2017: 332) "Pearson's correlation coefficient, symbolized by "r" is used to measure the relationship between two interval-level variables. The Pearson's correlation coefficient (r) represents the relationship between two numeric values and expresses the relationship through values between -1 and 1. If the correlation coefficient outcome is near 1, both variables express a linear relationship (positive). When the coefficient outcome is -1, it may also express a correlation but if one variable increase and the other decreases this is known as an inverse or negative correlation relationship. Two variables with a correlation near or equal to zero entails that there is no relationship and variables near or equal to 1 or -1, expresses strength in relationship.

A correlative relationship is not the cause for one variable to increase or decrease when comparing with other variables. One of the challenges in analysing statistics is to determine which outside factors influence the nature of a relationship between variables.

# 3 Research Methodology

## 3.1 NBA Statistical Data

The purpose of this study is to gain a better understanding of the impact data analytics have in the NBA. To address the purpose of research quantitative data was collected from two secondary sources at www.nba.com/stats and www.basketball-reference.com. The first dataset is from the 1993-2023 NBA seasons which represents 30 seasons of player statistics. This dataset provides exploratory data analysis of statistical variables that include 3PA, 2PA, PTS, POSS, 3PA%. Detailing historical trends in shot distribution will explain how the league has moved from a primary 2PA league to one that is embracing 3PA.

The first dataset was scraped from www.basketball-reference.com using the python programming language. The columns in the dataset are box score statistics, the total numbers of columns variables are 30. The significance of using this data shows the shot distribution between 2PA and 3PA along with FG% and 3P% over a 30-year period. Other key variables include PTS (Points), FTA (Free Throw Attempts), FGA (Field Gola Attempts), POSS (Possession), eFG%, TOV (Turnovers) and TRB (Total Rebounds). Below is a sample set of variables used in this study. Table 2 outline the variables used:

Table 2: Statisitcs Variables

| Variables | Definition |
|---|---|
| MP | Minutes Played |
| FG | Field Goals |
| FGA | Field Goals Attempts |
| FG% | Field Goals Percentage |
| 3P | 3 Point |
| 3PA | 3 Point Attempts |
| 3P% | 3 Point Percentage |
| 2P | 2 Points |
| 2PA | 2 Point Attempts |
| 2P% | 2 Point Percentage |
| eFG% | Effective Field Goal Percentage |
| FT | Free Throw |
| FTA | Free Throw Attempts |
| FT% | Free Three Percentage |
| ORB | Offensive Rebound Percentage |
| DRB | Defensive Rebound Percentage |
| TRB | Total Rebounds |
| AST | Assist |
| STL | Steals |
| BLK | Blocks |
| TOV | Turnovers |
| PF | Personal Fouls |
| PTS | Points |
| POSS | Possession Estimates |

The second set of data is from https://www.nba.com/stats/shotcharts the data was again scraped with the python programming language. This dataset is from the 2021-2023 seasons, which represents 2 years of data in the modern NBA and 427,737 shots. SportsVU cameras were installed in NBA arenas in 2013-14 so this data gives us valuable insight into play-by-play data. The usefulness of these variables as mentioned in the literature review include insightful players data that is not recorded in a normal box score stat sheets. Accessing play-by-play data in shot chart detail is important because we get column variables that include shot range, shot distance, action type, event type, shot type among other parameters. Table 3 is a sample set of variables used in this study:

Table 3: Play-by-Play Variables

| **Variables** | **Definition** |
|---|---|
| PLAYER_NAME | Player |
| TEAM_NAME | Team |
| EVENT_TYPE | Made/Miss |
| ACTION_TYPE | Jump/Layup etc |
| SHOT_TYPE | 2PT FG 3PT FG |
| SHOT_ZONE_BASIC | Mid-Range Restricted Area etc |
| SHOT_ZONE_AREA | Left, Right, Center, Corner |
| Feet (8-16 ft) | Feet (8-16 ft) |
| SHOT_DISTANCE | 1-27 FT |

The third dataset was scraped from www.basketball-reference.com. This dataset is teams' total statistics from 1993-2023, other variables merged into this dataset is the win and loss records over the same period. Wins and losses reflect the success of a team during a given season which can be used compare against to other statistics. One of the main talking points around data analytics is the focus on team performance and taking efficient shots on offense, those include 3PT, FT, and layups.

Furthermore, this paper seeks to answer the question which statistical variables show a strong correlative relationship with the wins of an NBA team? We chose basics statistics recorded in an NBA stat sheet, ignoring complicated analytical metrics used by analyst. Table 4 illustrates the variables tested:

Table 4: Correlation Variables

| **Variables** | **Definition** |
|---|---|
| W | Wins |
| AST | Assist |
| DRB | Defensive Rebounds |
| PTS | Points |
| FT | Free Throws |
| BLK | Blocks |
| STL | Steals |
| 2P | 2-Points |
| 3P | 3-Points |
| FT% | Free Throw Percentage |
| TOV | Turnovers |
| PF | Personal Fouls |

## 3.2 Methodology

There are three datasets in this project analysed by two methods. The first dataset will be analysed using the python programming language. Python is an open-source object-oriented programming language that works well with complex data sets. Basic line graphs are used to show the trends in different statistical categories, plotting data with line graph show information that changes over time using the x, y axis.

The second dataset will be analysed with Tableau. Tableau is a data visualization tool used for data analysis and business intelligence. Using the shot distance filter (1ft-27ft), FG% parameter, along with total FGA will illustrate what type of shots are taken in each NBA game and which shots are efficient. Bar charts visualize data using rectangular bars either horizontally or vertically to compare the heights and lengths of the values they represent. This method is important because a total sum of FGA (3PA, 2PA) is used in this study. Other variables measured using bar charts are shot zone basic, shot zone range, etc.

The third dataset was analysed by python, to figure out the correlative values we used a basic correlative function to process the output. To create the heatmap we used YData Profiling, this tool helps data scientist extract value information and provides insight for EDA (exploratory data analysis)

# 4 Results

The overall impact of data analytics can be visualized from the 1993 season until the 2023 season. Analytical theory was applied first by Darly Morey's Houston Rockets in 2007 and a shift in strategy by most NBA teams begin in the 2013-2014 season. To give a historical outlook we use provided a data from 1993-2023 which illustrate a 30-year period. It is important to note that the NBA had a lockout in the 1995-96 season for three months, a lockout in 1998-99 for five months with 52 total games out of 82 and a lockout in 2012-13 season with 64 total games out of 82.  Basic statistical categories increased or decreased on a yearly basis revealing the shift in philosophy in NBA. The basic categories are 3PA, 2PA, PTS, POSS and 3PA%.

The number of 3PA in 1995 were 16 and increased 18 per game in the 1997 season. This slight increase doesn't represent any indication the league was shifting towards 3PA the league was still heavily influenced by mid-range or close range shot attempts. As the years progressed 3PA declined then increased around the year 2009. In 2009 3PA were at 19 per game, in the year 2023 the number of attempts rose to 33 this represents a 74% increase in the number of attempts.

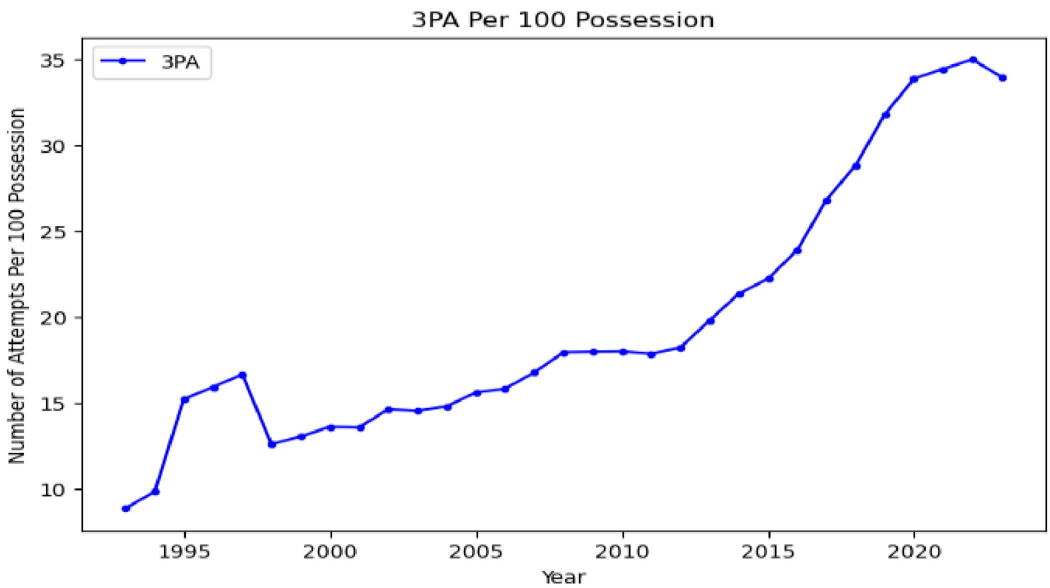


Figure 10: 3PA Per 100 Possessions

The number of 2PA decreased yearly, which illustrates a concerted effort to stop taking close-range shots and attempt long distance shots in the NBA. In 1995 NBA players attempted 77 2PA per game, in 2023 players attempted 53 2PA per game. The percentage change in 2PA from 1995-2023 is -31.1% overall which shows a negative trend in 2PA. These finding are key in explaining the drastic impact of data analytics in the NBA.

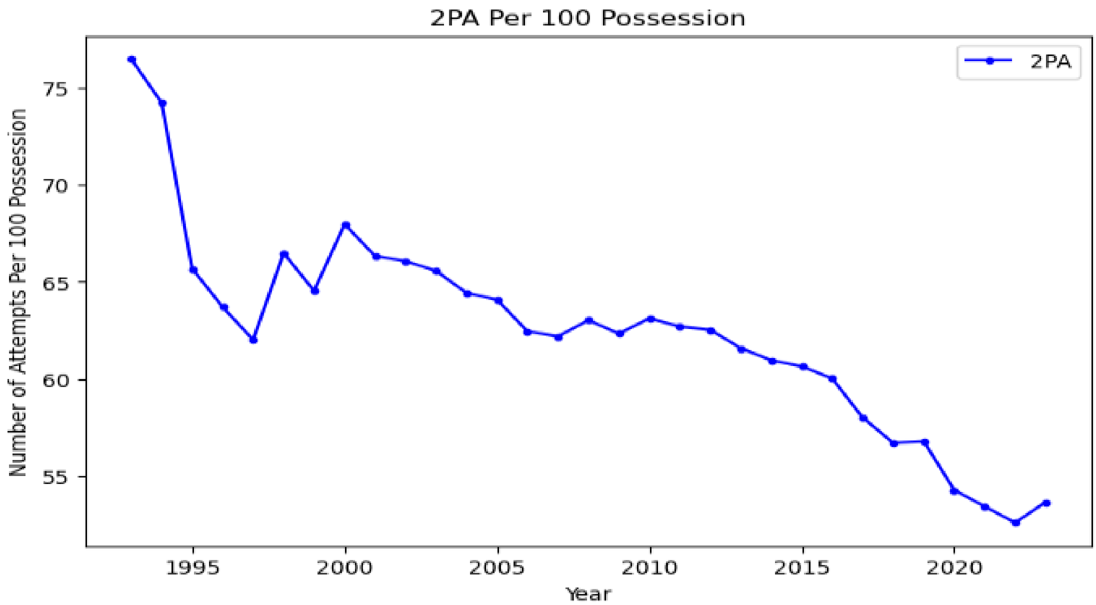


Figure 11: 2PA Per 100 Possession

The impact of PTS scored per 100 possessions clearly illustrates the NBA shift from a slow pace to a fast-paced game. An overall increase in PTS not only indicates a faster pace but more 3PA per game. The object of basketball is to score more points than your opponent, data analytics finds out which shots are most efficient during each possession to increase that point total. In 1993 scoring was 104 points per game, decreasing significantly in the next 6 seasons, with a total output of 90 points per game in 1998-99 season (lockout year). In the year 2000 points began to increase but fluctuated in totals until the year 2014 which could mark the period where all teams were somewhat analytically driven. From 2000-2023 points increased by 17.8% illustrating the impact of analytics in the PTS category.

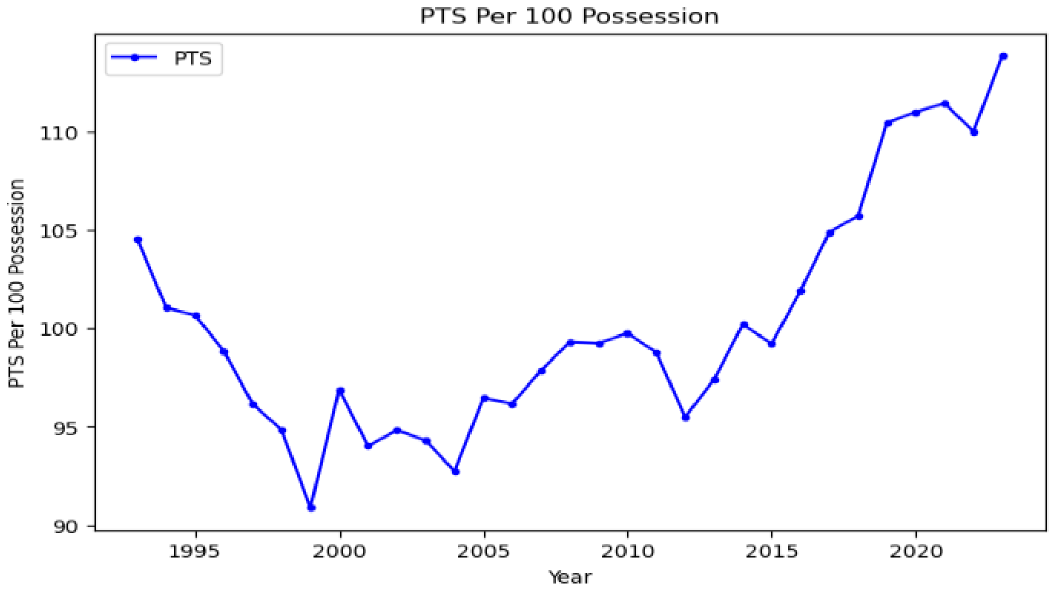


Figure 12: PTS Per 100 Possession

The rate or percentage of possessions is key in understanding data analytics in the NBA. Possessions are the number of times a team has the ball to score points. A low number of possessions indicate a slow game with passing, screens and play calling. A high number of possessions indicate a fast-paced game with quick shots and no play calling. The points and possessions graph mirror each other in yearly trends because they are connected, the more PTS scored per possession will increase the number of possession available during the game.

In 1993 possessions were 99 a game for both teams, decreasing yearly until 1999 at its lowest at 90 POSS a game. In the year 2000 possessions began to increase but saw a sharp decrease until 2014 again mirroring point totals graph. From 2000-2023 possessions changed by 5.2%, this illustrates the impact in pace and being more efficient lending its shift to the data analytics revolution.

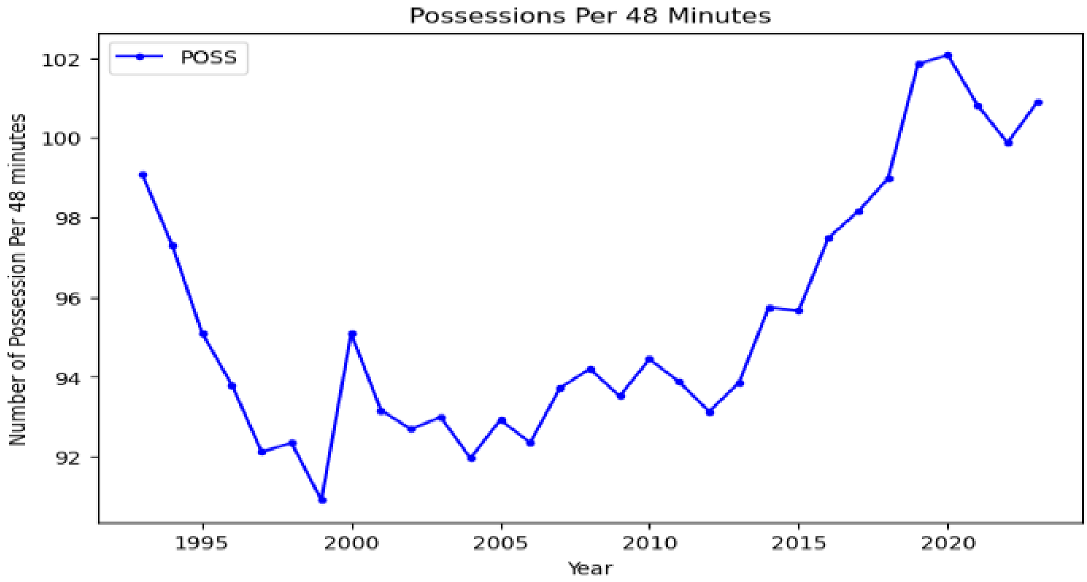


Figure 13: Points Per 48 Minutes

The 3PA% (3-point attempt percentage) shows the percentage of 3-pointers attempted during a game. In an NBA game there is only 3 types of shots being attempted 2PA, 3PA, and FTA. In 1993 there was only 10.4% of 3PA, displaying a low attempt percentage rate. In 2023 the number of 3PA in an NBA game was 38.7% this represents a 28% increase in 3PA's. Data analytics encourage teams to shoot more 3 pointers because they are 50% more than 2 pointers so this shows a clear trend in attempting more long-distance shots in the NBA.

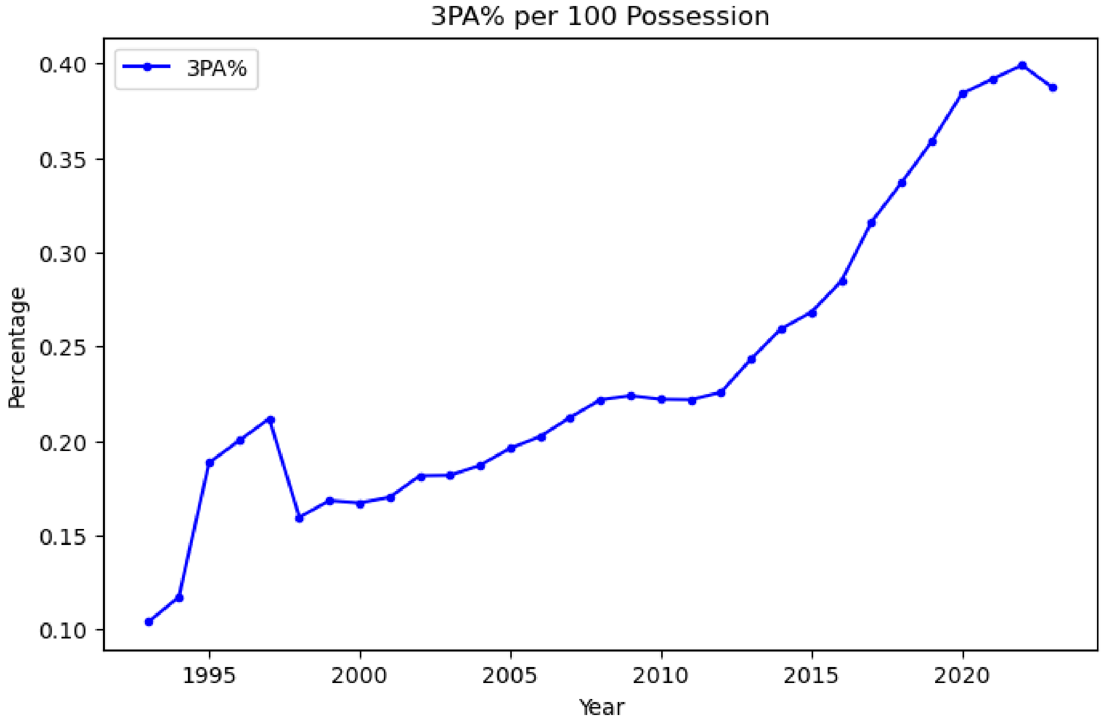


Figure 14: 3PA% Per 100 Possession

To illustrate the shot selection of NBA players due to the implementation of analytics are the following. Using NBA shot chart data from 427,737 shots through two full NBA seasons. 28,806 shots from 0 feet from the basket, these represent dunk shots which are the most efficient shots and the closes. Shots at 0 feet from the basket convert 76% of the time. Shots 1 feet from the basket totalling 34,639 and convert on 73% of the attempts. These are layups which are also highly efficient shots. The next highest total are shots from 2 feet away from the basket, 31,316 shots from that distance went in at 60%. As shown in the findings after 3 feet, shots decline significantly until 22 feet.

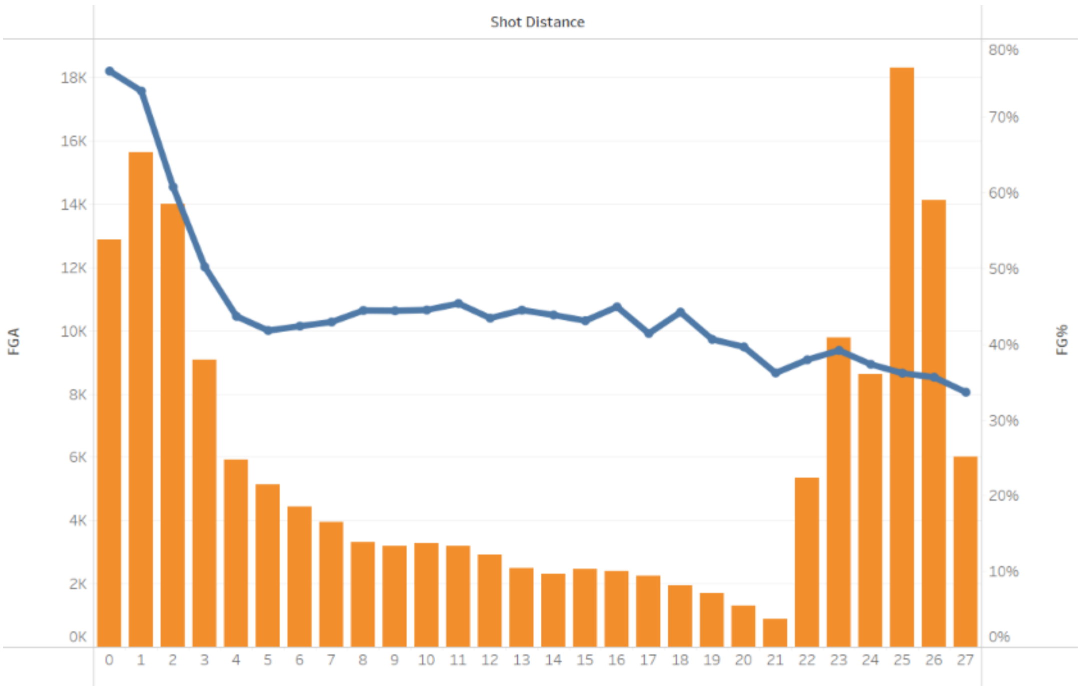


Figure 15: Measure of Shot Distance

Corner 3-point shots are 22 feet away and standard 2-point shots are 23 feet and beyond. With the advent of analytics players can shoot from distances as far as 27 feet with no problem. There were 12,250 shots from the corner location with 38% success rate. Standard 3-point shots at 23 feet were attempted 22,463 times with a 39% FG%. Long distance shots from 25 feet away are the most frequent shots followed by 26 footers. Over 2-year period 42,282 shots were taken from 25 at 36% FG% and 31,898 from 26 feet with a 35% FG% respectively. The results of the shot distance graph show the impact data analytics have on the shot selection in the NBA. From 1-3 feet shots are attempted frequently, from 4-21 feet a decrease in attempts occur, finally from 22-27 feet shots began to increase in numbers.

In terms of shot zone, play-by-play NBA data records the location where a player has taken his last shot. There seems to be specific location that players intend to shoot from. The most frequent shots are restricted area shots, those are plays from 1-3ft from the basketball rim. From the 2021-2023 NBA season players took 51,588 restricted area shots, 44,680 above the break threes, 33,303 shots in the painted area, 20,574 mid-range, 8,722 left corner threes and 7,856 right corner threes.

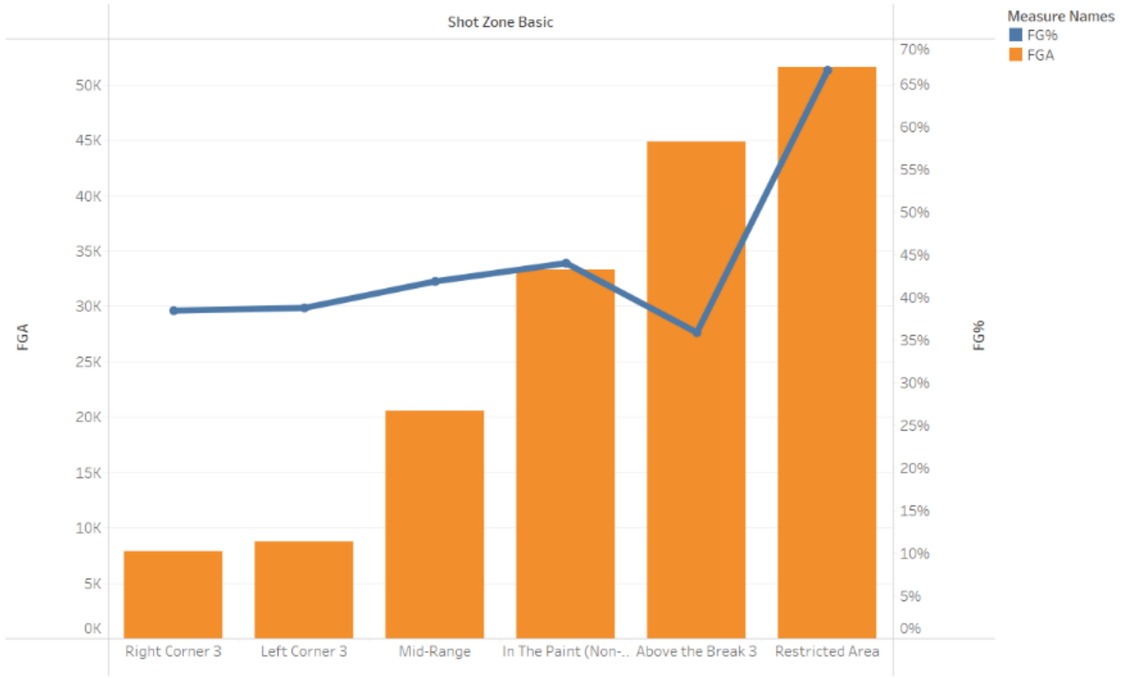


Figure 16: Shot Zone Basic Chart

In terms of efficiency (field goal percentage) the restricted area converts 67% of the time, followed by "in the paint" points which are successful at 44%. Above the break three points are effective 36% of the time, which is still considered to be efficient in the context of data analytics in basketball.

The shot zone range is the distance measured in feet from the basketball rim to the location of the shot attempt. There are 4 distance area in NBA play-by-play data, (1) less than 8 feet, (2) 8-16 feet, (3) 16-24 feet and (4) 24+ feet. This information provides the coaching staff with valuable information for scouting reports, also it helps with predicting the location of jump shots throughout the game. Over 157,409 shots were taken from less than 8 feet away from basket, followed by 141,239 taken from 24 plus feet away. As seen the previous findings there is a trend of taken shots close to rim, then there is a clear indication that players prefer to take long distance shots thereafter. The mid-range shot (8-16) was attempted 51,730 times, before the analytics revolution this was the common shot taken in the NBA. The decrease in attempts show the impact of data analytics in the overall offensive strategy and philosophy of coaching staffs in the league.

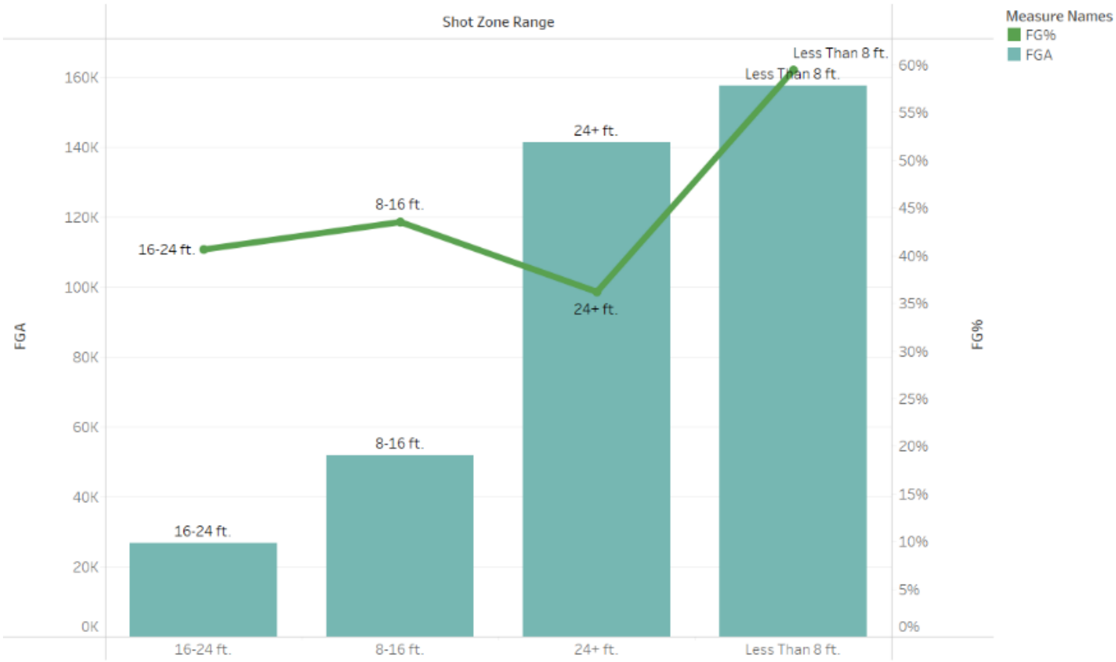


Figure 17: Shot Zone Range

## 4.1 Correlation Findings

The question of which statistical variables correlate with team wins provide some interesting findings. The results of the Pearson's correlation coefficient (r) are the following:

| Variable | Value |
|---|---|
| DRB | 0.732 |
| AST | 0.666 |
| PTS | 0.604 |
| FT% | 0.569 |
| 3P | 0.543 |
| BLK | 0.191 |
| FT | -0.02 |
| STL | -0.042 |
| 2P | -0.099 |
| PF | -0.275 |
| ORB | -0.319 |
| TOV | -0.319 |

*Figure 18: Correlation Results*

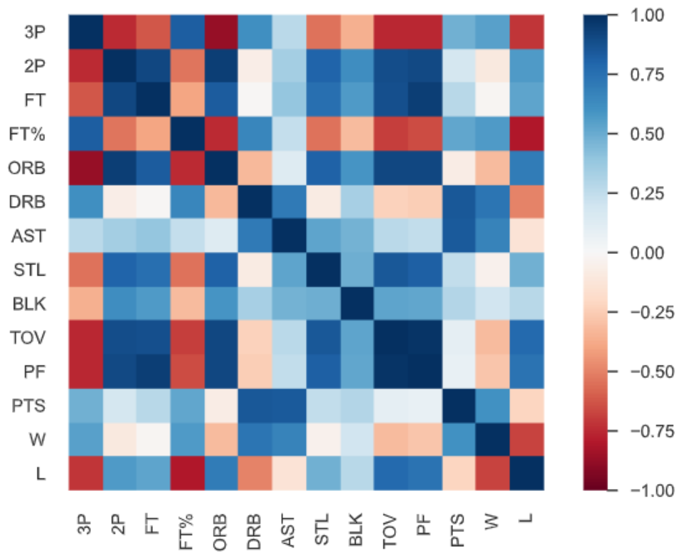


Figure 19: Correlation Heatmap

Based on the study 6 variables highly correlated with wins. Those were DRB, AST, PTS, FT%, 3P and BLK. Defensive rebound had the highest value at 0.732, when a team grabs a rebound a new possession starts for that team, which equates to a failed shot attempt, a turnover or miss free throw by the opponent. Assist showed a positive correlative relationship with wins 0.666, an assist is a pass from a teammate that results in a successful FGA. High assist players are valued and contribute to a team overall success, they usually find open players on the court for easy conversion.

PTS were found to have a strong correlation at 0.604, followed by FT% with a 0.569 correlation. 3P suggest a strong relationship with wins at 0.543 and BLK have 0.191 correlation. As noted in the study 3PT are 50% more than 2PT which leads to more points, these are considered valuable and efficient shot. Negative correlated values are FT at -0.020, STL with -0.042, 2P negatively correlate with -0.099. Personal fouls (PF) have -0.275 with wins followed by ORB and TOV with -0.319 respectively. ORD (offensive rebounds) surprisingly have a negatively correlate with wins, the number of offensive rebounds in a game varies because defenders are always in positions to obtain defensive rebounds.

# 5 Conclusion and Recommendations

The outcome of the line graphs highlights the major points of the thesis that data analytics has impacted the NBA game in several ways. 3PA from 1993-2008 never exceeded over 17 a game, thereafter 3PA attempts reached 20 per game for the first time in 2013. Over 34 3PA were shot in the 2022-2023 season this is proof that analytics has impacted strategy on the offensive side of the game. The decrease in 2PA, those are mid-range and attempts just outside the painted area have declined significantly. In 1993 75 2PA were the league average, a negative trend over the last 30 years have resulted in only 53 2PA per game in 2023. This also show a clear indication that teams have become analytically driven and encourage their players not to attempt those shots.

The increase in the number of possessions relate to the pace of the game.  Pace is the speed at which a team plays it determines the total possession a team may possess in a 48-minute game. One reason for the dramatic shift is to attract a bigger audience, faster play and quick shot with higher percentage effect would make the game more entertaining for fans and spectators. Recommendation for future studies in analytics could talk about rule changes and how it impacts the overall NBA business model. Total PTS scored were at 105 per game in 1993 and declined steadily until the year 2000 after 1998-99 lockout season. Shifting to a high scoring, fast paced game indicates the NBA willingness to integrate analytics in its overall philosophy hence its strong impact on all levels in a franchise.

In terms of identifying how the shot selection in the NBA has been impacted by data analytics, play-by-play data was used to demonstrate the shot distance and FGA from NBA players in a season. The outcome of the bar charts confirms that most the preferred shots in an NBA game are between 1-3ft thereafter, shots between 23-26 feet are selected. An interesting insight from the bar chart shows the steep decline in attempts from 4-21 feet (mid-range or long 2PT jumpers) this indicates a huge impact on gameplay and shot selection.

The basic theory in data analytics look for the most efficient shots in an offensive possession as discussed earlier, shots close to the basket, or 3-point shots are the main category of shot types currently in NBA. The shot zone basic bar chart confirms that over

51,588 shots were taken from restricted area at 65% success rate. Above the break 3 were taken 44,680 times during the season. The specific location of these events confirms again the overall impact of data analytics in NBA. One criticism of analytics is how its implementation has reduced other actions that might occur during a match. A critical analysis of NBA rule changes that occurred since the inception of analytics could provide useful information of its effect. Defensive 3 seconds, that is no player defender is allowed to stand in the painted area longer than three seconds opens up the possibility for more space and easy attempts at the rim.

The shot zone range is measurement from where the shot is taken and the rim. This suggests that players are either taking very long shots or close shots at the rim. As stated in the literature review, the shot range could be altered by defensive play, currently not enough information is retained regarding defensive analytics. Using shot chart data coaches can draw specific defensive strategies to keep players from taking certain shots. 157,409 shots were taken from 1-8 feet away and 141,239 from 24 feet plus. This is an overwhelming number of shots from specified locations which suggest that the impact of analytics is involved in the phenomenon.

The Pearson's correlation coefficient method to measure which variables where correlated with wins suggests that 6 statistical categories have a positive correlation with wins. Those were DRB, AST, PTS, FT% 3P and BLK. This suggests these factors influence winning by a wide margin. Although 3PT are heavily correlated it does not mean it is the primary factor in winning. It is recommended in the future that advance metrics such as TS% and eFG% could be used because they combined various basic statistics that could give a better overview of what influence wins.

This purpose of this paper was to investigate the impact and uses of analytics in the NBA. Using historical data, the paper showed how the trend in several statistical categories changed in the last 30 NBA seasons. Using play-by-play data, it was possible to highlight the volume of shots taken from certain locations. As mentioned throughout the paper data analytics is merely a tool to gain insight into a team or player's ability. Future studies could find new ways to connect sociological and psychological factors in the decision making of players. Basketball remains an unpredictable sport that requires preparation, a competitive drive, and luck to be successful.

## References

Abbott, M. (2017). *Using statistics in the social and health sciences with SPSS and Excel*. Hoboken, New Jersey: Wiley.

Alferink, L.A., Critchfield, T.S., Hitt, J.L. and Higgins, W.J. (2009). GENERALITY OF THE MATCHING LAW AS A DESCRIPTOR OF SHOT SELECTION IN BASKETBALL. *Journal of Applied Behavior Analysis*, 42(3), pp.595–608. DOI: https://doi.org/10.1901/jaba.2009.42-595

Bishop, E. (2023). *The Science of Injury Prevention in the NBA: How Teams are Using Technology to Keep Players Healthy*. [online] Available at: <https://www.sportskeeda.com/basketball/the-science-injury-prevention-nba-how-teams-using-technology-keep-players-healthy> [Accessed 5 October 2023].

Brown, L. (2019). *Kobe Bryant explains why he hates analytics in NBA*. [online] Larry Brown Sports. Available at: https://larrybrownsports.com/basketball/kobe-bryant-hates-analytics-stats-nba/511218 [Accessed 9 Oct. 2023].

Devlin, S. and Uminsky, D. (2020). Identifying group contributions in NBA lineups with spectral analysis. *Journal of Sports Analytics*, 6(3), pp.215–234. DOI: https://doi.org/10.3233/jsa-200407

Kubatko, Justin; Oliver, Dean; Pelton, Kevin; and Rosenbaum, Dan T. (2007) "A Starting Point for Analyzing Basketball Statistics," *Journal of Quantitative Analysis in Sports*: Vol. 3: Issue: 3, Article 1.

Knowledge@Wharton. (2017). *The NBA's Adam Silver: How Analytics Is Transforming Basketball Knowledge@Wharton*. [online] Available at: <https://knowledge.wharton.upenn.edu/article/nbas-adam-silver-analytics-transforming-basketball/> [Accessed 7 September 2023].

Goldsberry, Kirk & Eric Weiss. "The Dwight Effect: A New Ensemble of Interior Defense Analytics for the NBA." *Harvard University*, MIT Sloan Sports Analytics Conference, 1 March 2013, Boston, MA, USA.

Goldsberry, Kirk. "CourtVision: New Visual and Spatial Analytics for the NBA." *Harvard University,* MIT Sloan Sports Analytics Conference, 2 March 2012, Boston, MA, USA.

Heindl, K. (2022). *The Conflicting Catch-All of NBA Load Management*. [online] Global Sport Matters. Available at: <https://globalsportmatters.com/business/2022/02/22/conflicting-catch-all-nba-load-management/> [Accessed 5 October 2023].

Hofler, R.A. and Payne, J.E. (2006). Efficiency in the National Basketball Association: a stochastic frontier approach with panel data. *Managerial and Decision Economics*, [online] 27(4), pp.279–285. https://doi.org/10.1002/mde.1252

Laver, L., Kocaoglu, B., Cole, B., Arundale, A.J.H., Bytomski, J.R. and Amendola, A. (2020). *Basketball sports medicine and science*. Berlin, Germany: Springer.

Plafke, J. (2013). *How the NBA's SportVU ball and player tracking tech changes the face of sports - ExtremeTech*. [online] Available at: <https://www.extremetech.com/extreme/169775-how-the-nbas-sportvu-ball-and-player-tracking-tech-changes-the-face-of-sports> [Accessed: 27 September 2023].

Pykes, K. (2022) *Sports analytics: How different sports use data analytics*, *DataCamp*. [online] Available at: <https://www.datacamp.com/blog/sports-analytics-how-different-sports-use-data-analysis> [Accessed 26 September 2023].

Reveliolabs.com. (2021). *Arrow point top right*. [online] Available at: <https://www.reveliolabs.com/news/business/which-nba-team-is-the-most-data-driven/> [Accessed 28 Sep. 2023].

Reich, B.J., Hodges, J.S., Carlin, B.P. and Reich, A.M. (2006). A Spatial Analysis of Basketball Shot Chart Data. *The American Statistician*, 60(1), pp.3–12. DOI: https://doi.org/10.1198/000313006x90305

Sabr.org. (2011). *A Guide to Sabermetric Research | Society for American Baseball Research*. [online] Available at: <https://sabr.org/sabermetrics> [ Accessed 27 Sep. 2023]

Sampaio, J., McGarry, T., Calleja-González, J., Jiménez Sáiz, S., Schelling i del Alcázar, X. and Balciunas, M. (2015). Exploring Game Performance in the National Basketball Association Using Player Tracking Data. *PLOS ONE*, 10(7), p.e0132894. DOI: https://doi.org/10.1371/journal.pone.0132894

Sarlis, V. and Tjortjis, C. (2020). Sports analytics — Evaluation of basketball players and team performance. *Information Systems*, 93, p.101562. DOI: https://doi.org/10.1016/j.is.2020.101562

Skinner, B. (2012). The Problem of Shot Selection in Basketball. *PLoS ONE*, 7(1), p.e30776. DOI: https://doi.org/10.1371/journal.pone.0030776

Skinner, B. (2010). The Price of Anarchy in Basketball. *Journal of Quantitative Analysis in Sports*, 6(1). DOI: https://doi.org/10.2202/1559-0410.1217