Universitat Politècnica de Catalunya
Facultat d'Informàtica de Barcelona
Escola Tècnica Superior d'Enginyeria de Telecomunicació de Barcelona
Facultat de Matemàtiques i Estadística

Degree in Data Science and Engineering

Bachelor's Degree Thesis

Development of value metrics for specific basketball contexts: a positional approach to the defensive rebound value

Arnau Turch Ferreres

Supervised by

Sergi Oliva	Portland Trail Blazers, NBA
Jordi Cortadella	Dept. of Computer Science, UPC
Ferran Marqués	Dept. of Signal Theory and Communications, UPC

June, 2022

First and foremost, I would like to thank Sergi Oliva for his selfless commitment and involvement. Without him this project would not have been posible. Thanks for giving me the opportunity to learn from a profesional who I deeply admire and look up to.

I also want to deeply thank the UPC, especially Jordi Cortadella, Ferran Marqués and David Anglada. Thank you for your advice and guidance during all this time.

Thanks also to NBA Properties, Inc, for their cordiality and generosity in providing the data necessary to carry out this work.

Finally, I would like to thank all the people I have met in Gran Canaria, as they have been a fundamental part of the experience lived during these last months. I will always keep them with me, especially Raquel, who I want to thank for her infinite patience and support.

Abstract

Over the past decade, NBA teams are making an increasing investment in the development and growth of data analytics to help each team's ultimate goal of winning the championship. This growth has been accompanied by improvements in data collection from the league's games, moving from superficial, manually annotated statistics - box scores - to much richer data that can provide much more information - the optical tracking data.

This study aims to analyze a very specific facet of the game: defensive rebounding. This section of the game is one of the least information that can be collected manually. Using optical game tracking data from the 2020/2021 season, we will try to capture the contribution that players have in achieving the rebounding for the team.

In this work, multiple metrics have been developed to explain different dimensions of the defensive rebounding process, in addition to different analyses of these metrics. This has allowed us to discover different player profiles in this phase of the game and team behaviors that were invisible to the eyes of traditional statistics.

Keywords

basketball, defensive rebound, optical tracking data, Voronoi diagram, principal component analysis, data visualizations, probability density estimation

Contents

List of Figures

List of Tables

1. Introduction

Robin and Russell are two friends who love to play basketball, every afternoon they meet at the park where they have 2-on-2 games with other neighborhood kids. I'm not going to lie, they are pretty good, they make a duo where they complement each other very well: Robin is tall and strong for his age and Russell is shorter but fast and explosive.

This duo has found a tactic that gives them a lot of advantage over the rest and this makes them almost unstoppable:

When an opponent is going to shoot, Robin, who is tall and strong, ignores the ball and looks for the other opponent. Simply with his physical attributes, Robin prevents him from fighting for the rebound. Then Russell, quick, agile, and explosive, is the one who, with almost no competition, secures the defensive rebound and quickly initiates the play to attack with plenty of advantage. In short, no one has yet managed to grab an offensive rebound from this duo.

All the other kids, bothered by Robin and Russell's dominance of the playground, started a discussion about how to beat them. Rebounding quickly became the hottest topic of conversation. They all felt they needed to stop and nullify the duo's best rebounder to stop their quick attacks.

But, wait, who was actually the best rebounder of the two? Russell clearly finishes the games with a ton more rebounds than anybody else, so the answer seemed simple.

But was it?

1.1 Definition of the problem

To this day, when asking who is the best rebounder, the most common answer would be the player who has secured more rebounds. This logic works when applied to other fields, such as points. Normally, a scorer is a player who can score a high number of points for his team and therefore, this characteristic is attributed as a quality of that player. But with rebounds, and in particular defensive rebounds, it doesn't work exactly like that.

The rebounding process does not simply take place when the ball drops to a certain height where players can start fighting for the ball, this process begins much earlier.

At the moment the shot is taken, the real fight starts on the court to get the rebound for the team: players trying to get into the paint, defenders trying to push opponents out of the zone, people jumping... a physical war.

That is why it is not so fair and correct to evaluate the contribution that a player has had in obtaining the defensive rebound by observing the simple, but also important, fact of grabbing it. This is because behind this achievement of the rebound, there are multiple layers of "dirty work" that are not reflected in any of the typical statistics present in the box scores.

This fact hides and detracts in one way or another the work of certain players, whose mission is not to grab the rebound, but to facilitate it for the team.

For example, look at the following play:

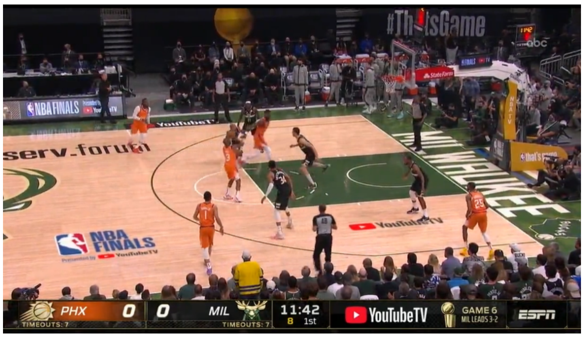


Figure 1: Chris Paul's shot.

Figure 2: Ball in the air: Brook Lopez goes to the opposing team's pivot, Deandre Ayton.

Figure 3: Ball in the air: the pivot is expelled from the paint.

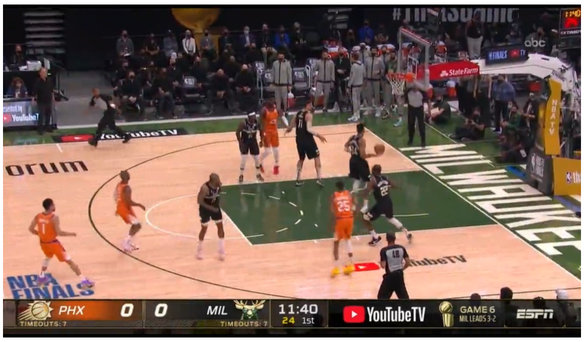


Figure 4: The rebound is secured by Giannis Antetokounmpo.

This is a capture from a game between the Phoenix Suns (in orange) and the Milwaukee Bucks (in green) from the 2020-21 NBA Finals, and a clear example of what was discussed above.

When Chris Paul (number 3 for the Phoenix Suns) shoots, number 11 for the Milwaukee Bucks, Brook Lopez, does not even look at the ball and goes directly to the pivot of the opposing team, Deandre Ayton, to expel him from the paint as shown in figures 2 and 3. The result of the action is the achievement of the rebound by number 34, Giannis Antetokounmpo, with no competition.

The statistics will reflect that Antetokounmpo caught the rebound. However, the "merit" is not really his. It was Brook Lopez's protection of the rebounding area that allowed Antetokounmpo to grab the rebound. This is not reflected in any of the traditional statistics, as it does not represent the description of an outcome, but rather of its process.

In the past, it was not possible to evaluate the rebounding or calculate any metric of contribution other than counting who and how many rebounds were achieved for the team, but now the scene is different.

The NBA has made a qualitative leap in recent years in relation to data collection and analysis. The 2013-14 season was the first in which optical-tracking cameras were installed in all the arenas, which collect the positioning of the players (X, Y) and of the ball (X, Y, Z) at a resolution of 25 frames per second. At first, the main information provider was STATS Sport VU [8] until 2017 when it was replaced by Second Spectrum [9].

This information captured by the cameras is processed by different computer vision algorithms that generate the positional data of the players and the ball. Then, with this information, many other statistics and information are generated covering a wide range of aspects of the game, including shooting, rebounding, defense, or dribbling.

This advance provides new game-related information which goes far beyond that which is got from the traditional box score.

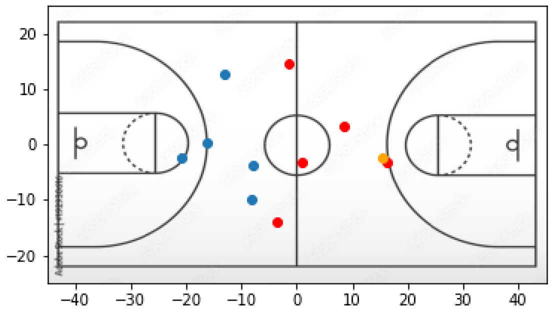


Figure 5: Example of players and ball position information

This type of data will be used during this project to improve on the problem of assigning credit and responsibility in the defensive rebound phase of the game. This data position of the players will bring a new dimension of information which will help to quantify the contribution and importance of the players in the positive outcome for the team in this important task of the game.

1.2 Importance of the study for the NBA

The field of sports data analytics is a market that is growing and has more value every year. According to [13], in 2021 its market value was USD 889.4 million, and they expected to grow by 21.3% between 2021 and 2028. In other words, they predicted that the value of this market for that year will be about USD 3.4 billion.

Reports such as [22] show that in 2021 its market value was 2.5 billion USD and a growth of 27.3% is forecasted until 2026, reaching a value of 8.4 billion USD.

What is clear is that it is a market that is on the rise and that there is an investment behind it aiming for its improvement and development.

This rise in the market value of game data analysis is partially because of the large investment the NBA and its teams are making. Fifteen years ago, the position of data analyst did not exist in the NBA. In contrast to nowadays, every team in the league has a dedicated data analysis department, some with over 10 people, showing this growth and investment in the industry.

Hence, any advance that involves the quantification of game actions and that generates metrics to evaluate the behavior of players in various aspects of the game is beneficial to the league and the teams [19] in their pursuit to better identify the players and strategies that help them win.

Here, as mentioned above, defensive rebounding is a very important aspect of the game and very little information is extracted from it. If the shared notion of rebounding is changed to understand it as a process that begins long before it is known whether the shot goes in or out, the work of teams and players can be understood so that in the end the team gets the expected result of grabbing the rebound.

This is information that few people have worked with, and it would help to find players that are vital for the achievement of the team's rebound even if their mission is not to grab it or team and player behaviors, among many other concepts.

2. Objectives

The main objectives of the present work are:

- To investigate and understand previous work related to extracting information about the contribution and performance of players and teams in defensive rebounding using positioning data.
- To reproduce the state-of-the art in this area and to introduce and develop new concepts and metrics that add more value and provide more information on the contribution of players to defensive rebounding.
- To perform different analyses of player and team behaviors and support them with real examples of game scenarios.
- Provide positioning data for each player for each play in which a defensive rebound occurs in order to build a model that considers this information and also combines it with play-by-play information.

  This aim is the continuation of the project Armand Alarcón will carry out [2] in his final bachelor's thesis.

3. State of the art

Since the term "Moneyball" [12] appeared, the popularity and use of data analysis in sports has been increasing, as well as the research and scientific work related to it. A growing number of studies are using tracking data to find new information and develop new tactics for teams.

In studies such as [7, 6, 21], it can be shown that using positional data of the players on the court can provide information on tactical and technical questions of individuals or estimations of the value of the play that is being developed. These studies focus on the offensive facet of the game and how players and teams perform in this phase.

Going a step further, in the article *Using Voronoi diagrams to describe tactical behavior in invasive team sports: an application in basketball* (António Lopes et al.) [16], the authors mention that "Team sports are recognized as dynamic systems of interaction, where individual and collective patterns of behavior emerge from a confluence of multiple constraints on the players". In order to carry out the analysis respecting these constraints, the concept of Voronoi diagram is introduced as it allows, unlike other approaches, to define the dominant region of both players and teams.

In the aforementioned study, the tactical offensive behavior in basketball is presented. The results show that the percentage of area occupied by each team, at each instant, can describe the spatial organization patterns of the teams during a positional attack, allowing to classify and distinguish between the transition phase and the organized phase of attack.

Focusing more on rebounding, in *Analysis of factors predicting who gets a ball in basketball rebounding situations* (Hojo et al.) [15], the authors deal with the problem of trying to predict who will get the rebound. They make use of positioning data and variables developed from this type of data divided into three categories: individual position, individual movement, and interpersonal relationship. The main purpose of this study, besides the prediction, is to find out which factors are the most important and the most determining in achieving the rebound. The authors highlight in the results that one of the best techniques to secure the defensive rebound is to protect the area near the hoop (box out).

In *To Crash or Not To Crash: A quantitative look at the relationship between offensive rebounding and transition defense in the NBA* (Winsen et al.) [24] the authors discuss the trade-off between crashing the glass or retreating to defense, showing that going back to defense contributes to a successful defense, although the probability of getting an offensive rebound is reduced by more than half.

Finally, in the article *The Three Dimensions of Rebounding* (Maheswaran et al.) [17], the authors mix the notions of the use of tracking data and defensive rebound. This paper shows, for the first time, the idea that all rebounds are not equal and that they should be understood as a process instead of just the exact moment when someone catches the ball. Following this reasoning, the course of the rebound is divided into three phases or dimensions: positioning, hustle and conversion. These dimensions correlate to three distinct skills that reflect what is observed in the game during the process of the rebound.

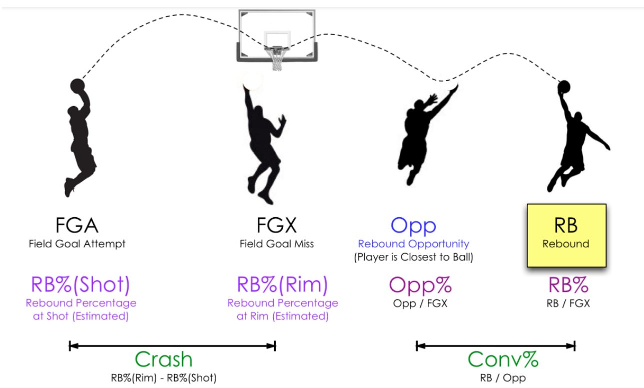


Figure 6: The three dimensions of rebounding

In *The Three Dimensions of Rebounding* (Rajiv Maheswaran et al.) [17] **positioning** is understood as the ability of players to be in a zone where the rebound is likely to fall. It is more likely for a player to get the rebound while being close to the basket than being outside the three-point line. Players who tend to have a high positional value, whether because of talent, luck or strategy, are known as positional rebounders. Between the moment the shot is taken and the instant the ball finally touches the hoop, players fight for space on the court where attackers try to penetrate the areas where the rebound is most likely to fall while defenders try to block those attackers. In short, there is a variation of positioning between these two moments that is referred to as **crash**.

Once the ball has touched the hoop and bounces off the rim, players fight for the rebound and try to convert that positioning into an **opportunity** to get the rebound. It is understood that the player who has an **opportunity** to get the ball at a particular moment is the player who is closest to the ball at that moment. This is the second dimension of rebounding, the **hustle**: the ability of players to generate opportunities for themselves.

Finally, the **conversion** tries to see if, once the player has generated such an opportunity, he has also managed to catch the rebound.

It is based on this previous work that the present study will be developed, reproducing some of its methods and introducing several new aspects.

4. Methodology

In order to achieve the main aim of this study, to quantify the contribution in the defensive rebounding phase of players using tracking data, the following methodological techniques are used: data acquisition, data processing, development of metrics, and a subsequent analysis of these metrics.

Because of it, this is a quantitative study, i.e., a study that "uses data collection to test hypotheses, based on numerical measurement and statistical analysis to establish patterns of behavior" (Hernandez et al., 2006) [14], since it is based primarily on the provision of information from objective data. Therefore, the reproducibility of the study is guaranteed.

4.1 Data acquisition

The National Basketball Association (NBA) provided the data used in this project for academic purposes.

Player positioning data is information describing the position (X, Y) of the players and the position (X, Y, Z) of the ball. Besides who the players are, the two teams present in the game and information about the period and time remaining in the game.

All the data was generated by Second Spectrum and corresponds to the complete data for the 2020/2021 season.

4.2 Data processing

For this study, data related to shooting and rebounding is also used. In addition, not all the tracking data is used, but rather only those frames compressed between the moment of the shot and the securing of the rebound. The rest of the frames are not useful for the purpose of this study as the tracking data has high resolution (25 frames per second) and for this study, it is unnecessary to work at such resolution. For example, to calculate the positioning values, only one frame is needed, since it is the analysis of an image of a certain instant of the game (it will be explained in more detail in section 5).

All this cleaning work and the subsequent analysis of the data is developed solely with Python with the help of libraries like Pandas, Numpy or Scipy among others.

4.3 Metrics development

As it will be explained in 5, once the data is obtained and processed, the metrics calculation begins.

This work will study the definition and implementation of different existing metrics (such as position, hustle and conversion). In addition, some of these metrics have been innovated in order to provide more value.

Finally, an original metric called box-out influence is presented.

4.4 Metrics analysis

Once the computation of the metrics is completed, different analyses are carried out based on these metrics, as it can be seen in 6. Different analyses are developed: metrics breakdown according to the shooting position, team analysis, metrics adjustment, and studies of similar players, among many others.

5. Implementation

This section defines and formally explains how the different metrics analyzed in this study have been calculated, as well as intermediate terms and concepts that are needed to understand the computation of these metrics.

First, a discussion will be carried out about the three metrics mentioned in section 3 specific of *The Three Dimensions of Rebounding* (Maheswaran et al.) [17]: positioning, hustle and conversion: An explanation of how they have been interpreted in this study and how they have been calculated. To conclude, the original metric of this study, called box out influence, will be introduced.

For all the above, it is necessary to take into account the rebound process considered in this study: it begins when the opposing player shoots the ball and ends when the rebound is secured.

5.1 Positioning

This metric aims to evaluate how players position themselves to secure the rebound. To achieve this evaluation, two factors must be considered:

First, given a missed shot, find the positional distribution of where the ball tends to fall after hitting the rim, therefore establishing a distribution of court value regarding the shooting position.

Second, given the positions of the players on the court, determine which player had the easiest access to each particular point of the court, or in other words, which parts of the court real estate he "controlled".

5.1.1 Spatial probability distribution of the rebound

To solve the problem of assigning a value to the playing court, a *spatial probability of the rebound conditional on the area the shot is taken from* has been developed. This assigns a higher value to the areas where the rebound is more likely to fall.

In order to achieve these probability distributions, a 2-D kernel density estimator (KDE) was used [18]. This is a non-parametric technique to estimate the probability density function of a random variable, which its final goal is to make inferences about the density function in all space regions using a finite dataset.

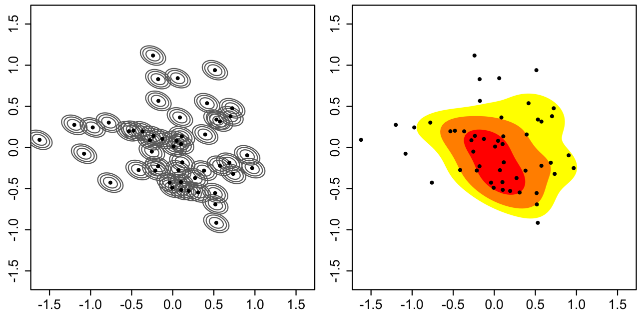


Figure 7: Construction of a KDE.

Formally defined:

Let $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_i$ be d-dimensional points drawn of the same unknown random distribution f, the KDE is defined as:

$$\hat{f}_{\mathbf{H}}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}} (\mathbf{x} - \mathbf{x}_i) \tag{1}$$

where $K_{\mathbf{H}}$ is the kernel function used, it has to be a symmetric multivariate density. $\mathbf{H}$ is the bandwidth/smoothing $d \times d$ matrix, symmetric and positive defined.

In this case, a Gaussian Kernel was used. It is defined as:

$$K_{\mathbf{H}}(\mathbf{x}) = (2\pi)^{-d/2} |\mathbf{H}|^{-1/2} e^{-\frac{1}{2}\mathbf{x}^{\mathbf{T}}\mathbf{H}^{-1}\mathbf{x}}$$

where $\mathbf{H}$ acts as the covariance matrix and controls the orientation and the amount of smoothing applied. The type of bandwidth selected influences the estimation of the KDE, more than the Kernel shape. In this study the $\mathbf{H}$ will be selected by a "rule of thumb", more precisely using Scott's Rule [20].

This computation of the rebound KDE will be calculated for each shooting zone, thus having several distinct functions which will be called, using equation 1:

$$\hat{f}_{reb}(\mathbf{X}|Zone = z) \tag{2}$$

where z is each region of the court.

5.1.2 Voronoi diagram

Once defined the value of the floor of the court, it is necessary to face the problem of defining to whom each part of the land "belongs". For this, the technique used in [16] and in [17] will be employed: the Voronoi tessellations.

The Voronoi diagram [5] is a division of the plane determined by the configuration of the data. For each reference point, its corresponding region (or cell) is defined as all the points of the plane closer to this reference point than to any other data points. The reason why this type of distribution is used is that it defines physical barriers in space, just as basketball players use their bodies in the process of rebounding to prevent opponents from advancing.

Formally defined as:

$$R_{k,t} = \{x \in X \mid d_t(x, p_k) \leq d_t(x, p_j) \text{ for all } j \neq k\} \quad (3)$$

defining X as a metric space, K as the set of indices, p_k a point in X and d_t as a distance measure at instant t, in this case, the euclidean distance. p_j is another point in X where j is other index from K different from k.

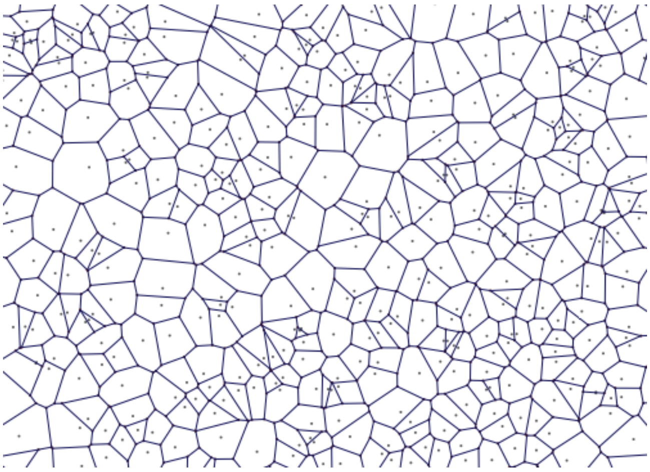


Figure 8: Example of Voronoi tessellation

We can interpret this in basketball terms more informally: the area of "control" of each player is all those points of the court that are closer to him than to any other player.

This definition assumes that all players are equally capable of controlling their respective cells, although, due to physical or skill conditions, it may be thought that this is not the reality, and therefore it would be necessary to define another partition of the space taking into account some factor that reflects this difference between players.

5.1.2.1 Power diagram

A power diagram [3] is another partition of the Euclidean plane into cells based on the data points. The difference between the aforementioned Voronoi diagram is that here, the points are defined as a set of circles with defined radius, representing the weight of each observation.

The definition of a cell for a circle C are all the points for which the **power distance** to C is closer than to the other circles.

The definition of **power distance** $\Pi(P)$ of a point P, to a circle c, being O the center and r the radius:

$$\Pi(P) = |PO|^2 - r^2$$

Power diagrams are a generalization of Voronoi diagrams, these two divisions of the space are the same when the radii of the circles are equal.

This technique is a way to represent the inequalities that may exist between players on the court and reflect the fact that a person who controls more ground of the floor thanks to a particular aspect of his skill or physique, has a larger cell.

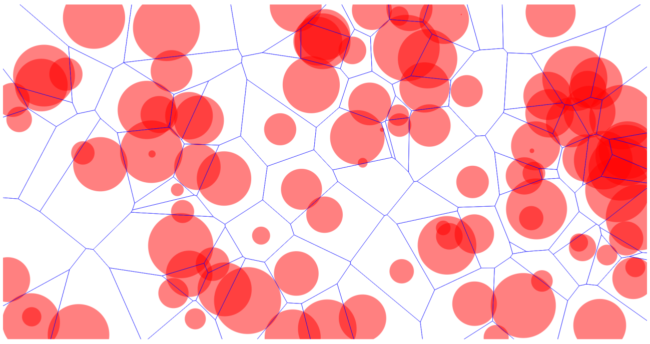


Figure 9: Example of a power diagram

5.1.3 Mixing it all

Once solved the following problems:

1. The value that each point of the court has for rebounding. See 5.1.1

2. Assign, to each player, the area of the court he controls. See 5.1.2

We can now assign to each player the worth of his control area.

Defining j as a player on the floor, z the spot from which the shot was taken in a specific time t, the *Position Value* (PV) is obtained for j as the integral over the area defined in 3 of the equation 2.

$$PV(J = j|Z = z, T = t) = \int_{R_{j,t}} \hat{f}_{reb}(\mathbf{x}|Zone = z)\, dx \tag{4}$$

This metric can be interpreted as the prior probability that each player has to get the rebound at a given instant if nobody moves. The key is not who controls more space, but who controls the most valuable area.

5.1.4 The crash

Using 4, we can perform the calculation for every play at two different instants. Firstly at the moment the shot is taken, the *release*, and secondly at the moment the ball touches the *rim*. The intention of this analysis is to see how the players move and fight for a better position while the ball is in the air.

Formally defined as:

$$\begin{aligned} Crash &= \Delta Position\ Value = \\ &= Position\ Value(J = j|Z = z, T = t_{rim}) - Position\ Value(J = j|Z = z, T = t_{release}) \end{aligned} \tag{5}$$

That is the variation of the position value, players that enter to compete for the rebound and therefore that present a better positioning at the rim moment, will have a positive position value.

5.2 Hustle

Before defining what *hustle* is, the notion of *Opportunity* is introduced.

A player is able to generate an *Opportunity* if he is the closest player to the ball, at any point between the ball touches the rim and someone grabs the rebound. The player who is closest to the ball for the first time may not catch the ball because of many reasons; for example, he may not be able to grab it, the ball might keep moving, and then another player may become the closest player to the ball. In other words, there may be more than one opportunity per rebound.

Let t_{rim} be the instant when the ball hits the rim, t_{grab} the instant where the rebound is captured and d_t the distance (euclidean) at the instant t between player p and the *ball*. The opportunity can be defined as:

$$Opportunity(p_i) = \begin{cases} 1, & \text{if } \exists\ t_{rim} \leq t \leq t_{grab} \mid \text{argmin}_{p \in players}\, d_t(p, \text{ball}) = p_i \\ 0, & \text{otherwise} \end{cases}$$

If the number of *Opportunities* that a player has generated is divided by all the missed field goals that have occurred when the player was on the floor, the result is what in [17] is defined as *hustle*.

$$Hustle(p_i) = \#Opp_{p_i} / \#Missed\ field\ goals_{p_i}$$

If the *Position Value* (5.1.3) represents the probability that the players have of getting the rebound if they do not move, that is, prior to really knowing where the rebound would fall for that play, *Hustle* reflects the moment after the ball touches the hoop, and it is known where it goes. It looks at the ability of the players to actually be in the correct position a posteriori, i.e., closest to the ball.

5.3 Conversion

As stated before, one missed shot can generate more than one opportunity, as the closest player to the ball in a certain moment could be unable to secure the rebound. Hence, this metric reflects this ability of grabbing the rebound when a player is the most likely to grab it.

Conversion is calculated as the number of defensive rebounds secured divided by the total number of opportunities generated:

$$Conversion(p_i) = \#DREB_{p_i} / \#Opp_{p_i}$$

5.4 The box out influence

Note that all of the above metrics relate to the ability of a particular player to rebound the ball. While that is relevant, it is not necessarily what the goal of the rebounding phase is: what we care about, as a team, is that our team gets that rebound, independently of which player does it. While the individual approach we have introduced up to this point likely translates to a team approach when referring to offensive rebounding, as it is complex to orchestrate a predetermined team strategy, that is not the case in defensive rebounding.

Most often, in order to maximize the ability of a team to secure defensive rebounds, the team strategy is not to go after them, but to "box out" the opponents - this is, to impede their progress towards high-value rebounding areas - thus making the rebound easily obtainable by the defense.

(a) Original image

(b) Deleted player

(c) Deleted player 33

Figure 10: The box-out influence.

Note that none of the metrics above accounts for this concept, which is key in order to evaluate the true contribution of a player to his team, obtaining a defensive rebound. To address that, we will introduce the following metric.

Before formally defining this metric, the following case is presented to understand the intuition behind it. Imagine the real game situation presented in figure 10a.

It shows a typical game situation in which a shot is taken. Look at the defense, i.e., the Sacramento Kings in purple. What would happen if a defender is eliminated from the court as in 10b? Surely his disappearance would not affect the team much for their chances of getting the rebound, although this player would have a very high positional value since he is in a very valuable rebounding area of the court.

On the other hand, what would happen if number 33 is eliminated, as in 10c? His disappearance leaves the field open to an attacker, who would surely have no difficulty in grabbing the offensive rebound in case the shot does not go in. Number 44, being out of the paint, would probably not have a higher positioning value than the one in 10b but his presence in that position of the court is key for the defending team to secure the rebound.

This example intends to illustrate that some players block and impede the passage of opponents to high-value areas with their presence, even if their positioning value is not high. They contribute to make the defensive rebound easier to capture, since they are "nullifying" a potential threat.

Letting be P the set of players on the floor, we will define P_{-j} as all the set P excluding player j. Moreover, D will be defined as all the players in P that are on defense. Then the box out influence (BOI) of a player $j \in D \subset P$, for an instant t, with a shot taken at zone z will be defined as:

$$BOI(j,t,z) = \sum_{p_i \in D \subset P} PV(J = p_i | Z = z, T = t) - \sum_{p_i \in D \subset P_{-j}} PV(J = p_i | Z = z, T = t) \tag{6}$$

This definition is derived from the notion of positioning value. In a first step, the positioning value of the whole defending team is calculated, i.e., the sum of the value of each cell of each player in the same team. Then, to see the influence of a specific player, the same team-wide positioning value is recalculated without taking into account the presence of player j in the play. This creates new configurations of the different Voronoi cells due to the absence of the player under study. Then, to see how much difference there has been in the value of the floor controlled by the team, the difference is performed.

A result close to zero would indicate that the terrain controlled by the player either does not have much value, or would be controlled by other players of the same team. On the other hand, a very high value would indicate that the player controls a high-value area, and that high-value area would have been under the opponent's control had he not been in that position.

6. Results

This section will show the results obtained using the methods described in section 5. This information will be shown and accompanied by an analysis of these results if necessary, besides trying to relate it to some examples of real cases.

6.1 Positioning

6.1.1 The rebound spatial probability distribution

The following figure, 11, is a collection of plots of the estimated probability density functions conditioned on the shooting zone. These were calculated using the rebound position information.

It can be seen that all the functions are distinct and have a certain semantic meaning:

Look at the first column. The first four pictures show shots taken from a centered area, but the lower the picture is in the column, the farther away the shot from the rim. In the KDEs you can see that they all have the highest probability zone in the same place, centered near the hoop. The main difference observed is that the further the distance of the shot, the more dispersion is seen in the functions. This is due to the fact that the greater the distance, the harder the ball rebounds and, therefore, the farther the rebound goes. The last two rows are shots from very far zones, and this dispersion is even more noticeable.

Continuing along the second and third columns, we can observe lateral shots, both two and three pointers. We can see that there is a clear difference between the distribution of shooting from the right or left zone: shots taken from the right zone tend to have the highest probability zone on the right of the hoop, and vice versa for shots taken from the left zone.

This is basically due to the momentum of such shots and, as stated in [10], "The ball tends to go to the other side of the basket for rebounds because of Newton's First Law, the relevant part of which is 'an object in motion continues in motion unless acted on by a force.'"

With this, it can be verified that the shooting position affects the position of the rebound and therefore justifies using a density function conditioned by the shooting zone.

These results are consistent with the ones that *Kirk Goldsberry* obtained in [11].

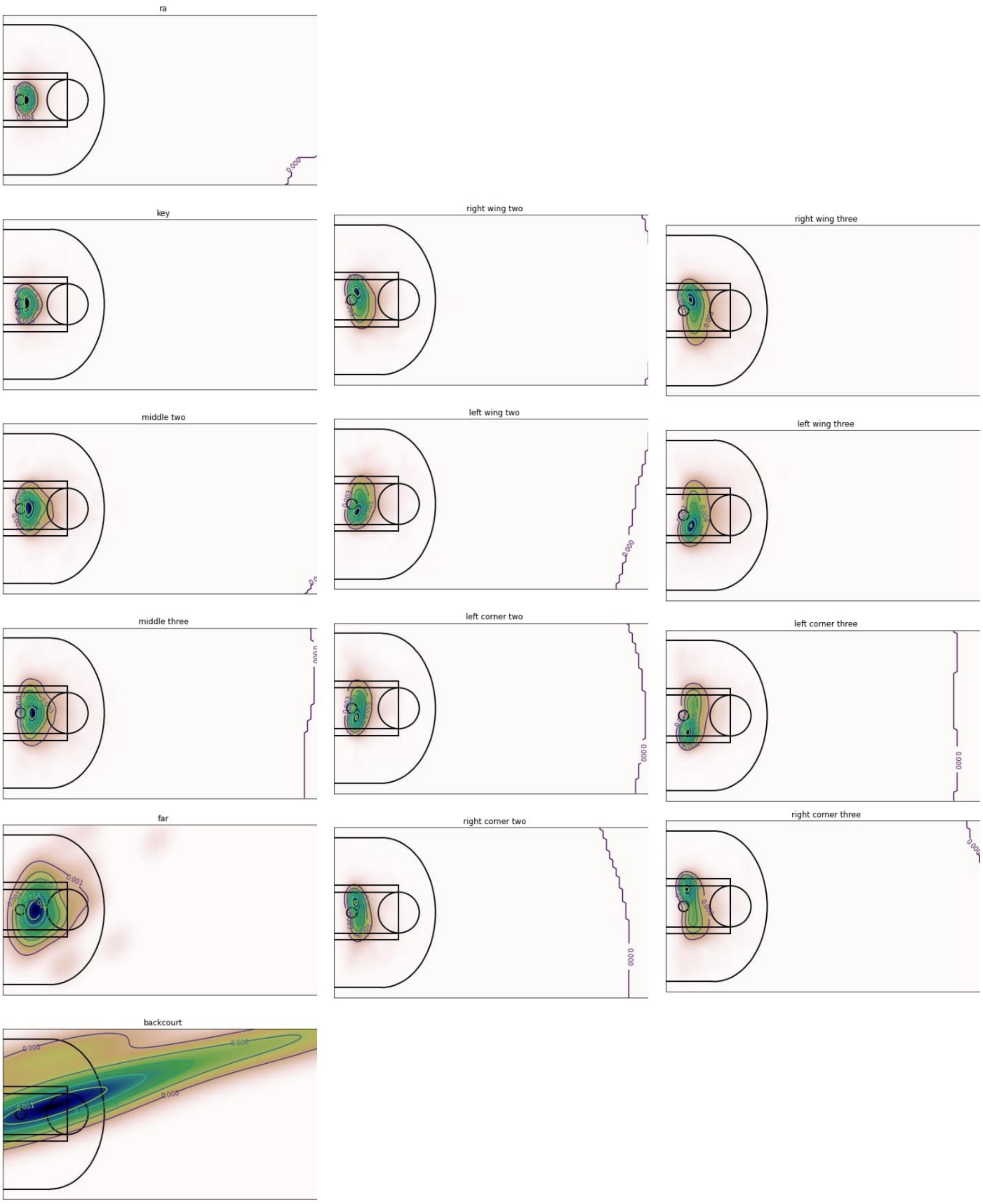


Figure 11: Plots of the KDEs functions obtained per each zone shot

6.1.2 Voronoi diagrams

In this section we will briefly show an example of a match frame, its mapping to positions and finally the result of having applied Voronoi.

Figure 12: Original frames of the match

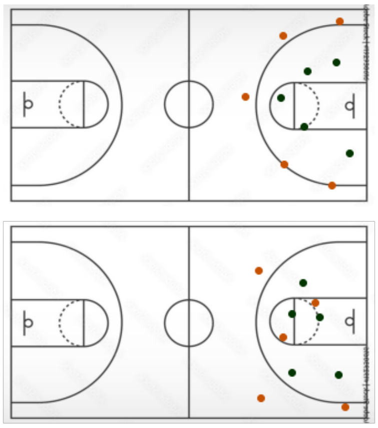


Figure 13: Players' position

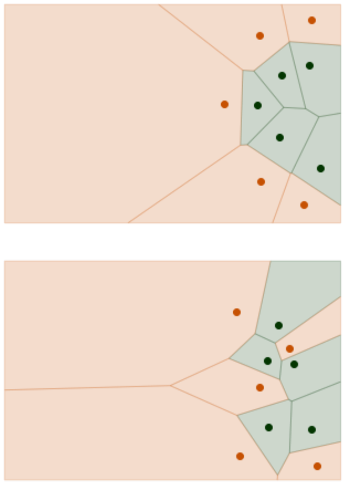


Figure 14: Players' position and Voronoi cell

Here it is possible to check how is the result of applying Voronoi in a concrete frame of the match: the field is distributed in cells that belong to each of the players. Depending on whether there are many players nearby or not, the cell of a particular player will be larger or smaller.

6.1.3 Power diagrams

As mentioned in 5.1.2.1, the Voronoi algorithm could be modified changing the definition of distance and, instead of assuming that all players are equal and have the same characteristics, add some differential factor between these players. This means a modification on the cell definition being more favorable to those players who excel in that quality. The factors used in this study are the height of the players, their weight and finally their body mass index (BMI).

Below, in 15, we show an example of a frame of a play, the results of both the Voronoi and the power diagrams using the aforementioned qualities as a differential factor.

Take some time to observe the differences that can be seen. The most significant would be the green player inside the paint at the baseline: he loses area when height is used as a factor, but gains territory if we use weight.

6.1.4 Position value

Here is where the results obtained in 6.1.1 and 6.1.2 are put together in a similar way as shown in Figure 16. For each player on the court, the estimated density function for the region obtained in the court division is integrated to obtain the positioning value.

This exercise is done for each shot taken at two specific moments: when the shot is taken (release)

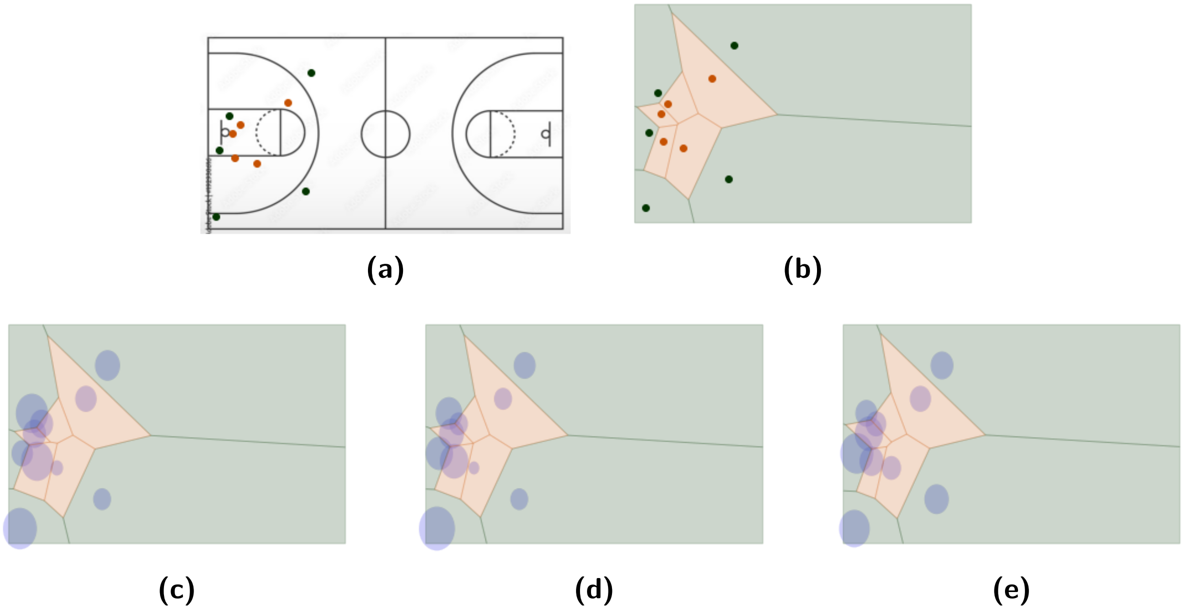


Figure 15: **(a)** Player's position **(b)** Voronoi **(c)** Power diagram using height information **(d)** Power diagram using weight information **(e)** Power diagram using BMI information

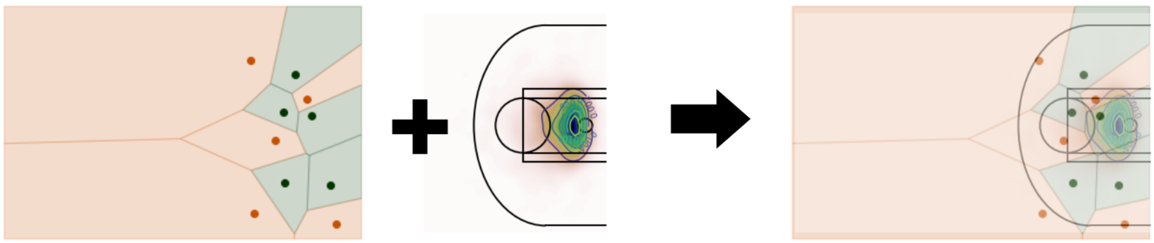


Figure 16: Computation of the positioning value

and when the ball touches the hoop (rim), this is done to observe how the positions of the players have changed while the ball was in the air.

The following are the 10 players with the highest average positioning value for both release and rim instants, the ones in bold font mean that are in the top 10 players with more DRB per game. Note that a filtering of players by minutes has been performed, all those whose number of minutes played in the whole season is less than the 0.6 quantile: 1792.4 minutes, are ignored.

As we could expect, the players present at the table are all interior players who usually move around areas where the rebounding is very likely to fall and, in the end, are responsible for fighting for the defensive rebounding.

It can be observed that the current Utah Jazz player and the one with the most defensive rebounds per game during the 2020-2021 season, Rudy Gobert, is present at the top of both tables, which indicates that thanks to his positioning ability he is able to grab more rebounds than anyone else.

However, the complete opposite case is found: the second highest defensive rebounder in that season, Russell Westbrook, is not present in any position of any table.

	Player Name	Positioning Value
1	**Rudy Gobert**	0.325
2	**Jonas Valanciunas**	0.311
3	Dwight Howard	0.307
4	**Joel Embiid**	0.303
5	Brook Lopez	0.300
6	Ivica Zubac	0.298
7	Enes Kanter	0.291
8	Wendell Carter-Jr.	0.286
9	**Nikola Vucevic**	0.285
10	Robin Lopez	0.282

(a) Top 10 players (Positioning value rim)

	Player Name	Positioning Value
1	Dwight Howard	0.284
2	**Rudy Gobert**	0.272
3	Enes Kanter	0.269
4	**Jonas Valanciunas**	0.266
5	**Joel Embiid**	0.266
6	Andre Drummond	0.254
7	Mason Plumlee	0.251
8	**Nikola Vucevic**	0.244
9	**Clint Capela**	0.239
10	**Nikola Jokic**	0.239

(b) Top 10 players (Positioning value release)

Table 1: Top 10 players with the highest average positioning value

There are also cases of players who “disappear” from one table to the other, like Brook Lopez. He is a player who, at the time of release, is almost in the 3rd place. However, when the ball touches the rim, he is not present in the top 10, in fact, he becomes the 54th player in the entire league in this value.

With this, we can already guess that there are different player profiles during the rebounding process present in the league, but this will be discussed later.

6.1.4.1 Power diagram study

Before continuing with the study, we wanted to check which was the best way to divide the playing field using the power diagram definition shown above.

For this, we have computed the positioning value according to the three variables exposed in point 6.1.3, together with the Voronoi value, with the result that, for each play, we have a different positioning value for each player (since the court partitions are not equal).

To distinguish and evaluate which partition is better, it has been defined that the best one will be the one that maximizes the positioning value of the player who finally gets the rebound. If we understand the definition of position value as the a priori probability that the players on the floor have of getting the rebound, the one that maximizes this probability will express the reality more accurately and therefore its partitions will be considered to be better.

In 17, it is shown the results of this study. “Unit” refers to the results obtained using the conventional Voronoi algorithm.

Although semantically it makes sense not to treat all the players on the court equally, assuming that there are certain physical inequalities between them, with the results shown in 17, not a very significant difference can be distinguished.The average rebounding position value is very similar in all cases, hovering around the value of 0.26.

Since there is no significant difference when performing the field division, the original Voronoi division will be used throughout this study, treating all players equally.

This does not detract from the fact that this technique can be useful in other subjects of study, although they do not include the main subjects of study of this work.

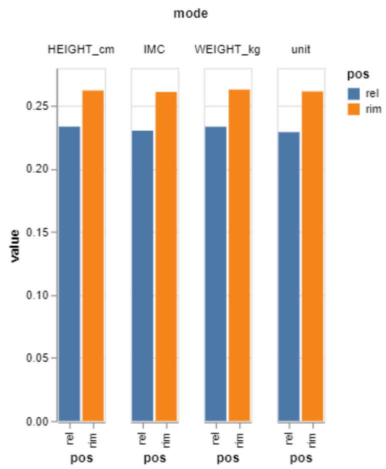


Figure 17: Results of the experiment

6.1.5 Crash value

In 2 it is shown the 10 players with the highest and lowest values of this metric explained in 5.1.4.

	Player Name	Crash value
1	Dejounte Murray	0.0686
2	**Russell Westbrook**	0.0421
3	Royce O'Neale	0.0391
4	**Giannis Antetokounmpo**	0.0390
5	Josh Hart	0.0384
6	Donte DiVincenzo	0.0380
7	Deni Avdija	0.0359
8	Luka Doncic	0.0329
9	Bruce Brown	0.0315
10	Cole Anthony	0.0312

(a) Top 10 players

	Player Name	Crash value
215	Brook Lopez*	-0.13143
214	Robin Lopez*	-0.10859
213	Ivica Zubac*	-0.06743
212	Jakob Poeltl	-0.06685
211	**Rudy-Gobert**	-0.05356
210	DeAndre Jordan	-0.04949
209	Naz Reid	-0.04802
208	Wendell Carter-Jr.*	-0.04768
207	Isaiah Stewart	-0.0469
206	Myles Turner	-0.04663

(b) Last 10 players

Table 2: Top 10 players with the highest and lowest average crash value

Again, the top 10 players with more defensive rebounds per game are marked in bold. Also, those players present in any of the 1 tables are marked with the symbol "*".

At a glance, two facts can be observed. On the one hand, players with a higher value are mostly exterior players. On the other hand, players with a lower value are interior players, even though some of them have some of the highest positional values in the league.

This can be easily checked with the help of 18 and 19, a plot that relates the positioning value (release and rim) to the crash value. As expected, a negative correlation is seen between these variables, which means that the higher the positioning value, the lower the crash value.

This makes sense basketball-wise, since outside players, just by the fact of advancing a little towards more interior areas, without necessarily entering the rebounding battle, already benefit them in terms of possessing higher value areas. On the other hand, for interior players it is just the opposite: as the ball is

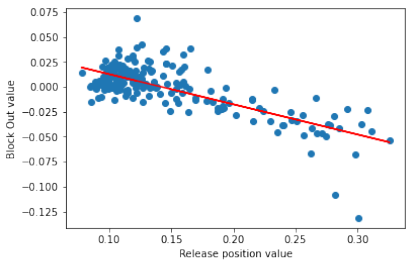


Figure 18: Position value release Vs crash value

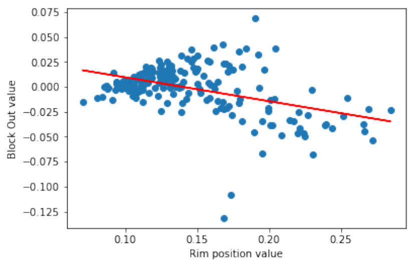


Figure 19: Position value rim Vs crash value

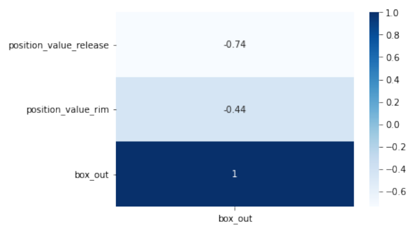


Figure 20: Correlation values

in the air, the players are concentrating and adding to the fight for the rebound, which reduces the space they possess and therefore lose positioning value with respect to the instant of release.

As a result, while this metric clearly describes a reality, it is not necessarily an appropriate means of comparison across all players - as their initial positioning conditions it. Players like Russell Westbrook or Giannis Antetokounmpo are some of the players with more defensive rebounds, and although they do not have a very high positioning value, it is reflected in the data and results obtained that going to more favorable areas helps them to get rebounds.

But what is the difference between these two, and the other 8 players present in 2? Why only these two are present among the top 10 defensive rebounders in the league? This is a topic that will be discussed in a more open and detailed way in point 6.4, but it has a lot to do precisely with the two individuals who have the lowest values (by far) of this metric: Brook and Robin Lopez.

6.2 Hustle & Conversion

Here it is show again the players with the highest values in both metrics: hustle and conversion.

	Player Name	Hustle
1	Dwight Howard*	0.70917
2	**Jonas Valanciunas***	0.65181
3	**Rudy Gobert***	0.64784
4	Enes Kanter*	0.6447
5	**Joel Embiid***	0.6314
6	Andre Drummond*	0.62374
7	Mason Plumlee*	0.61369
8	**Clint Capela***	0.60206
9	**Nikola Vucevic***	0.58072
10	Myles Turner	0.5616

(a) Top 10 players hustle per shot

	Player Name	Conversion
1	Ben Simmons	0.61466
2	John Wall	0.59524
3	**Russell Westbrook**	0.59509
4	Seth Curry	0.58036
5	LaMelo Ball	0.57925
6	Facundo Campazzo	0.57576
7	Jayson Tatum	0.5729
8	Jerami Grant	0.57018
9	Alec Burks	0.5698
10	**Giannis Antetokounmpo**	0.56683

(b) Last 10 players conversion

Table 3: Top 10 players with the highest hustle and conversion values

It can be seen that many players in the hustle table are those with a high positioning value since, as shown in 21, 22, 23, they are highly correlated variables. This makes sense because if they are in favorable rebounding areas, many of these times they will be the closest player to the ball once it has touched the hoop.

On the other hand, the conversion values do not highlight great rebounders, except for Giannis and Russell (again). This is because we are in front of a case of survivorship bias [4], a type of selection bias. As the definition of conversion is based on counting how many rebounds you have grabbed among all the opportunities you got, players who have few but "easy" opportunities will clearly be benefited. On the other hand, players who generate more opportunities but disputed and more difficult to convert into rebounds, will be disadvantaged in this metric because, although they will have more rebounds, they will also have generated more opportunities, thus decreasing the value of this metric.

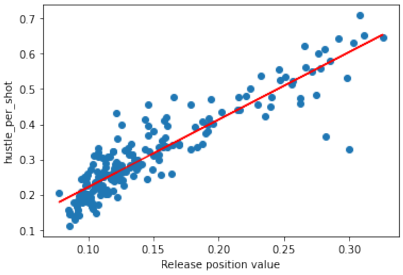


Figure 21: Position value release Vs hustle

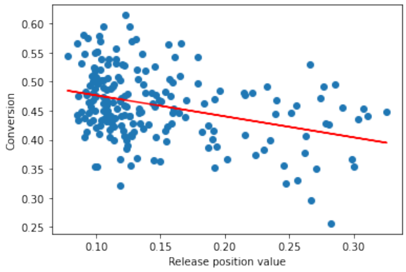


Figure 22: Position value rim Vs conversion

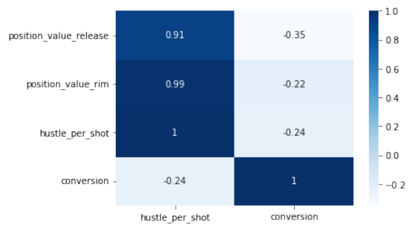


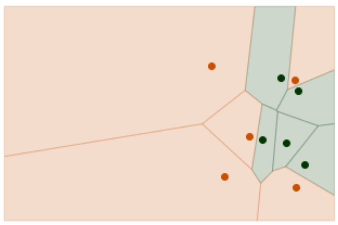
Figure 23: Correlation values

6.3 The box out influence

Finally, the values of the original metric of this study will be presented. But first, a small example will be shown to help to understand this concept and how it really indicates the contribution that players have in the rebounding process.

Let's take for example a play that was seen in 12, a shot by Chris Paul. The Voronoi diagram at the moment when the ball touches the rim obtained in 6.1.2 is analyzed.

(a) Original frame



(b) Original Voronoi

Figure 24: Original example

Now, we reapply multiple divisions of the court but eliminating a different defensive player each time, as seen in 25.

Here is the opportunity to see how the Voronoi distribution changes depending on which player is eliminated. Focus on images a) and b). If they are compared with the original image, it is seen that in a) the removal of a defender gives way to an attacker to get more control area of the floor, on the other

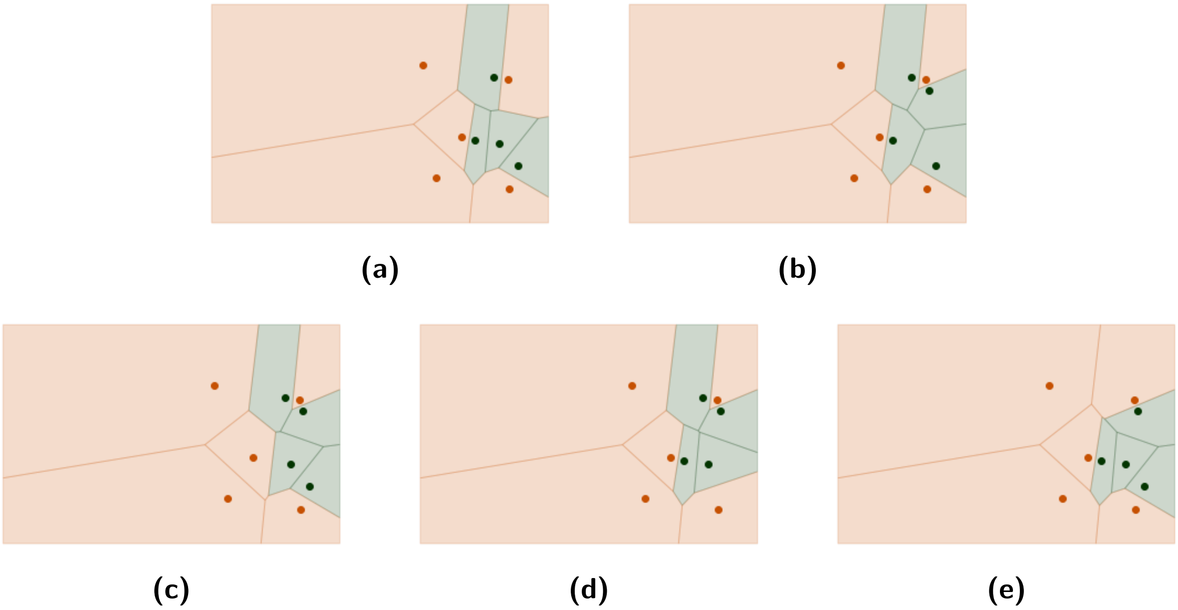


Figure 25: **(a)** Voronoi without Brook Lopez **(b)** Voronoi without Giannis Antetokounmpo **(c)** Voronoi without P.J. Tucker **(d)** Voronoi without Khris Middleton **(e)** Voronoi without Jrue Holiday

hand, if player b) is removed, it is noticed that the space that he controlled, is now controlled by players of the same team so that his absence should not be very noticeable for the team.

Now, if the position values for all defending players are computed using the original Voronoi partition and sum those values, a team metric is obtained that relates to the probability for the team of obtaining the rebound, at the end, the same as before, but instead of focusing on players, now focusing on the whole team.

Then, for each Voronoi shown in 25, the same calculations are performed to obtain the team positioning metric. After obtaining these results, each value is subtracted to the whole team positioning number to obtain the value of the zone of control that was lost with the removal of the player.

The results for this example, accompanied by the release and rim positioning values, are the following:

Player	Box-out Influence	Release positioning value	Rim positioning value
Brook Lopez	0.381	0.674	0.522
Jrue Holiday	0.037	0.028	0.053
P.J. Tucker	0.037	0.027	0.027
Khris Middleton	0.010	0.099	0.076
Giannis Antetokounmpo	0.002	0.014	0.238

Table 4: Example box-out influence results

These results show how the disappearance of Brook Lopez would cause an opposing player to gain valuable floor area, while the disappearance of Giannis would not affect the floor area controlled by the team at all. On the other hand, if you follow the play, you see that Brook Lopez is going to block the attacker (as the results show) being Giannis free to grab the rebound placidly.

To the traditional stats, it would count as a rebound for Giannis, ignoring all the work by the other team members to finally get the rebound. Instead, thanks to this new metric, it can be said that who has had a greater influence for the achievement of the rebound for the team has been Brook Lopez doing this hidden task of blocking the way to the attacker.

Once the use of this metric is understood and justified, the players with the highest value in this metric are presented, both at the time of release and at the moment of the ball hitting the rim.

	Player Name	Box Out Influence
1	**Joel Embiid***	0.158
2	**Rudy Gobert***	0.156
3	Dwight Howard*	0.149
4	Mason Plumlee*	0.146
5	**Jonas Valanciunas***	0.146
6	Myles Turner	0.145
7	Isaiah Stewart	0.145
8	**Clint Capela***	0.145
9	Jakob Poeltl	0.142
10	**Nikola Jokic***	0.141

(a) Top 10 players box out influence release

	Player Name	Box Out Influence
1	**Joel Embiid***	0.141
2	Dwight Howard*	0.133
3	Andre Drummond*	0.131
4	Mason Plumlee*	0.128
5	**Rudy Gobert***	0.124
6	Jarrett Allen	0.121
7	**Jonas Valanciunas***	0.121
8	**Nikola Jokic***	0.119
9	**Clint Capela***	0.119
10	Enes Kanter*	0.119

(b) Top 10 players box out influence rim

Table 5: Top 10 players with the highest box out influence values

6.4 We need to talk about the Lopezes

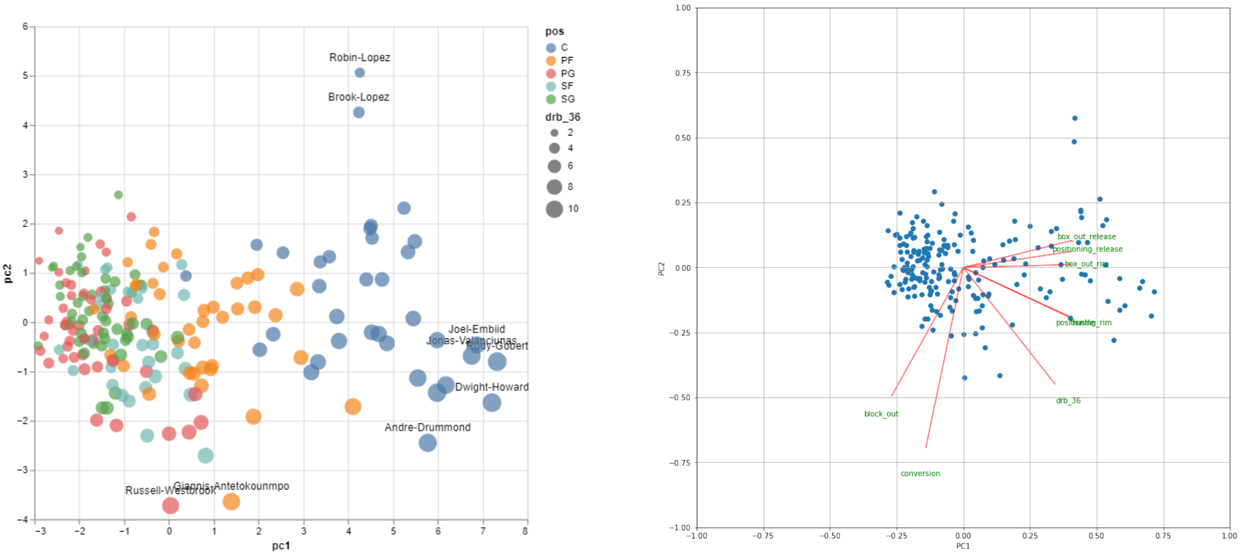
Once all the metrics have been developed, explained and shown, it is time to perform an analysis from these metrics to extract information about the rebounding process: similar players, different player profiles involved in this process, synergies between players, etc.

One way to visually analyze different observations, in this case players, is to perform a Principal Component Analysis, PCA [1]. The following are the results of this analysis, in which the next variables have been used:

- Positioning value (release) (6.1.4)
- Positioning value (rim) (6.1.4)
- Crash value (6.1.5)
- Hustle (6.2)
- Conversion (6.2)
- Box out influence (release) (6.3)
- Box out influence (rim) (6.3)
- Defensive rebounds per 36 minutes ([23])

The results are shown below in 26:

The figure on the left shows a two-dimensional plot to visualize the players. The color differentiates their position, and the size of the circle differentiates the number of DRB per 36 minutes. On the right, the same plot is shown, but accompanied by vectors that project the influence that each variable has on the two principal components. In addition, the angles between these vectors represent how the characteristics correlate with each other.



(a) PCA

(b) Biplot

Figure 26: PCA Results

As mentioned in 6.2, the high correlation between the hustle variable and the positioning value at the instant the ball touches the rim can be observed since the vectors are almost indistinguishable. Moreover, it is seen that the two principal components have a similar weight in these variables (hustle and positioning value (rim)) since they form almost a 45-degree angle with the X and Y-axis. This fact is also visible in the variable DRB per 36 minutes. Variables such as box out and positioning (release) have strong weight in the principal component of the X-axis, while crash and conversion gain importance in the Y-axis, the second principal component.

Knowing this and looking at the plot a) of Figure 26, it can be observed that, as we would expect, the players who are more to the right - meaning that they have a large value in the box out, positioning and hustle variables - are interior players. This is natural since those players are the main responsible for fighting for the rebounding for the team and for that they usually move in high value areas. Regarding the Y-axis we can observe that notably, those players who are lower, are players who are characterized by moving into the paint during the phase and with a high conversion. To conclude this analysis, the players who are located in the right-bottom zone of the plot project a high value in DRB per 36 minutes, as can be seen with the size of the circles they represent.

With this information, it is possible to identify different types of players during the rebounding process:

The first group would be the one located in the rightmost part of the plot. These players are characterized by having a high projection in the values of positioning and box out influence, and also are able to maintain their positioning value between the moments of release and rim.

They are players who from the beginning are located in the area of high rebounding value and also remain to fight for it because, as can be seen, they are players with a high DRB value. Players of this type would be Rudy Gobert or Joel Embiid for example, players who are distinguished by being very good defenders of the paint, responsible for blocking the pass to the zone and fight to collect rebounds for the team.

Continuing, two observations should be noted. These are Brook and Robin Lopez, who also happen to be twins. They are players that present very large box out influence and release positioning values, that is, they are players that dominate the inside zone at the moment the shot is taken and in both moments, they are key to block the passage of other opponents to high value zones. As they are usually located in a rebounding zone at the moment of the shot, we could think that they will also have high hustle and conversion values, but what happens is the opposite. Their crash values are the lowest in the league, and their positioning value decreases considerably, even though they are variables with a high positive correlation. Furthermore, even though they have a high projection on the hustle variable, their conversion value is very low. This would represent that although they are often close to the ball, they do not usually catch it.

In summary, they are two players that at the beginning of the rebounding process are in good positions, but end up moving away from them, ignoring the ball, and always with high box out values. This difference with respect to the other players would have two possible explanations: one would be that they are not very good players in this process of rebounding, the other, that their main objective is not to rebound. It is preferred to think that the correct answer is the second option, but then, what do their teams do to get the rebound? Someone should grab it, shouldn't it?

Note that at the lower part of the plot, two players separated from the rest are found, not as much as the Lopezes, but there is a difference between these two and the other players. What characterizes these players? They are players with high conversion and crash values but low positioning and box out values. In addition, they are players with low positioning at the time of the shot but high value at the time they touch the rim, as indicates the crash variable. In short, they are players who are expected to grab a high number of rebounds without initially being in areas of high probability, entering to look for these rebounds while the ball is in the air and without much competition since they project low box out values.

Those players are Giannis and Russell. In this dataset (2020-21 season), Robin and Russell are they are teammates of the Lopezes. Robin and Russell are teammates on the Washington Wizards, and Giannis and Brook are teammates on the Milwaukee Bucks.

It is clear that there is a great synergy between these two pairs of players in the rebounding process. Investigating how these two pairs actually play, it can be seen that they behave as described above. The Lopezes are very big players, both are 2,13 m tall and are characterized by defending very deep in the paint. They, when the shot is taken, go straight for the bigger attacker who can be a threat for the rebound and block him and try to move him out of the rebounding area. At the same time, Russell and Giannis get close to the rebounding area to try to grab the rebound and initiate the offensive play.

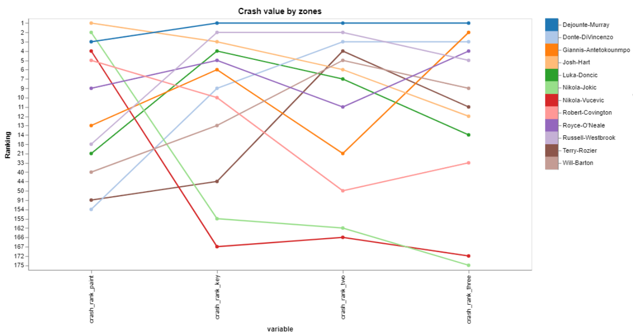
Using this new information, different profiles of players at the time of rebounding can be identified, which were invisible to the eyes of classical statistics.

6.5 Breaking down the metrics

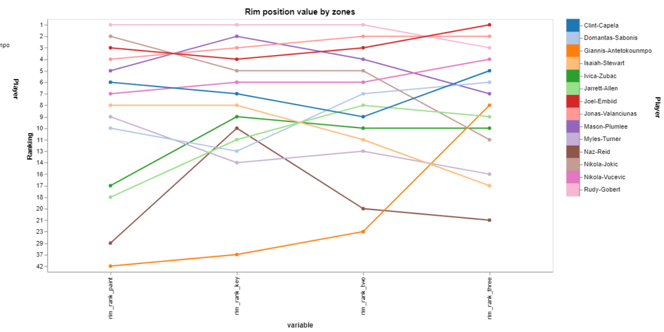
Having reached this point, questions may arise as to whether players may have a different role or performance during the process of rebounding depending on the area in which the shot is taken, so a breakdown of metrics based on shooting regions will be performed to see if this fact may exist.

The study will not focus on all metrics, but will show those that are of high interest and provide significant information. The rest of the information regarding this topic that is not present in this point can be found in A for the reader's curiosity.

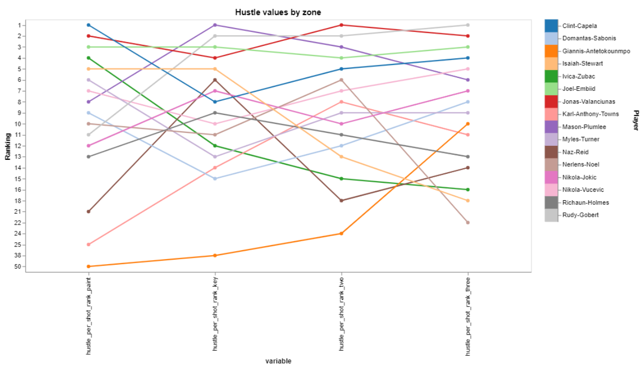
The most significant plots obtained in this breakdown are shown in 27, all following the same structure: On the X-axis is the metric breakdown between: paint (which is the restricted area), key, two point shots and three point shots. Then on the Y-axis is the ranking, i.e., which players have a higher value in that metric. Finally, the colors mark the different players. This observation is useful for noticing the irruption or descent of players by a certain shooting area.



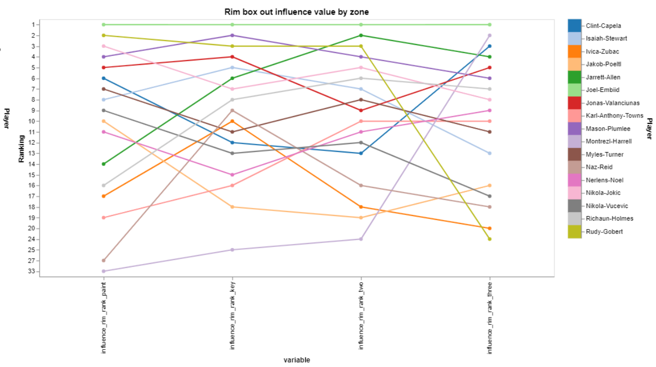
(a) Crash values per shooting zone



(b) Rim positioning values per shooting zone



(c) Hustle values per shooting zone



(d) Rim box out influence values per shooting zone

Figure 27: Metric breakdown per shooting zone

In plots **b)**, **c)** and **d)** a similar pattern is repeated: players who in the shots made in areas closer to the rim does not have much importance, but in three-point shots significantly increases his metric value. This is clear in plot **b)** with Giannis. In addition, in some occasions, it can also be observed that players with a high value in the areas close to the rim lose importance as the shot gets farther away. This is evident in **d)** with Rudy Gobert.

All these plots are metrics that are captured at the time of rim, it could be thought that the crash variable could have some kind of influence on these results obtained since it is the variable that reflects whether there has been a position gain between the time of release and the time of rim. To check if there is a relationship between the crash variable and the variables shown in 27, their correlation values can be easily stated, as follows:

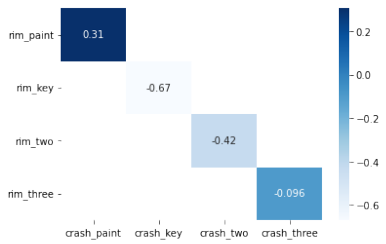


Figure 28: Rim positioning values

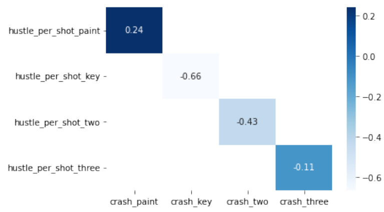


Figure 29: Hustle values

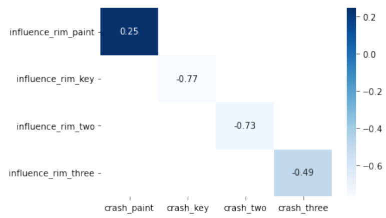


Figure 30: Rim box out influence values

The same pattern is observed in the three figures.

If one focuses on the correlations between the key, two and three pointer metrics, the correlation becomes less and less, going from being a negative correlation (that is, the more positioning, hustle or box out influence value, the less crash value and vice versa) to becoming less and less negative. This indicates that if shots are taken from a greater distance, the higher crash values obtained could mean higher values of the three metrics due to the low negative correlation, as observed in the case of Giannis Antetokounmpo in 27. This happens because the ball travels more distance, players have more time to move, and that translates into more time to win high value areas. In the opposite case, a player who is in zones close to the ball from the beginning, is negatively affected by this fact: more people will enter zones that he controlled at the beginning, meaning more competition, as observed in the case of Rudy Gobert in 27.

The interesting case that was not expected is a high positive correlation between the metrics of the shots taken in the restricted area (paint) and their crash value. This is clearly due to the short time that elapses between the moment of release and rim. In these cases, the players who are closer to the rim, and do enough to gain a minimum of positioning value, are by far the most likely to increase the value of their metrics. It is precisely when more interior players have a high value in the crash metric, as for example Nikola Jokić, Nikola Vucevic or Jonas Valanciunas. When the shot is made from farther away, the exterior players are gaining more value in this metric as seen before.

This demonstrates, once again, what we have been trying to convey throughout this study: defensive rebounding is a very complex aspect of the game that goes far beyond which team or which player grabs the rebound. This process includes many cause-effect relationships that are difficult to detect and isolate using the traditional statistics.

For example: the presence of a player on the court can cause the opposing team to be forced to shoot from further away, therefore making it easier for the team to rebound on defense on average, although this would not be reflected in the individual statistics of the player who is causing this effect. That is one of the reasons the box out influence has been developed. However, this factor could also be isolated taking into account which types of shots are the most predominant in the league, the different values of the metrics of a player by zones could be weighted by the frequency in which they occur in a way such as this:

$$Adjusted\_metric(P = p) = \sum f\_zone_z \dot{m}etric_{z,p}$$

Being p the player and z the different zones present on the court. *f_zone* would be the following in 6:

Zone name	Frequency
Backcourt	0.0016
Far	0.0018
Paint	0.2402
Corner 3 - left	0.0485
Corner 2 - left	0.0241
Wing 3 - left	0.0898
Wing 2 - left	0.0303
Middle 3	0.0984
Middle 2	0.0365
Restricted area	0.2480
Corner 3 - right	0.0448
Corner 2 - right	0.0234
Wing 3 - right	0.0809
Wing 2 - right	0.0312

Table 6: Shot zone frequency

This would try to isolate the effect on the distribution of the shooting zones caused by the presence of a player on the court and try to adjust the value of the metrics to the standard league shot distribution. Some significant results of this adjustment are shown below:

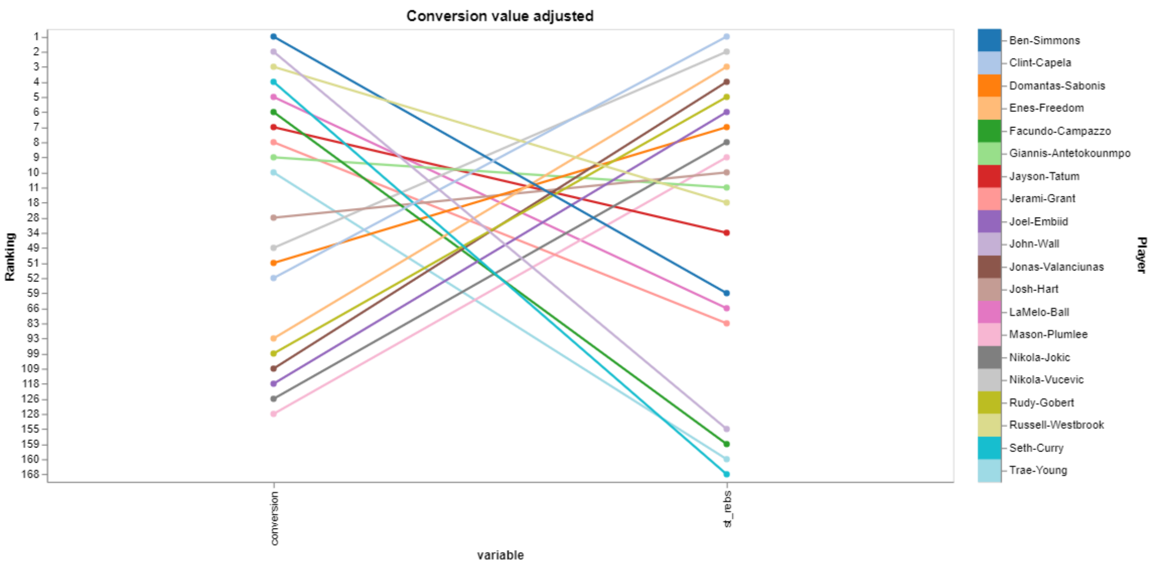


Figure 31: Conversion value vs conversion value adjusted

On the one hand, in 31, the effect of this adjustment can be clearly seen. As discussed in 6.2, the conversion value was greatly affected by those exterior players who had few and easy opportunities, which logically, benefited this metric. By adjusting and giving much more weight to the most frequent shooting areas (restricted area and key) it is seen that the interior players become much more valuable in the metric and the players who had the best values of these metrics in the beginning, become much less valuable.

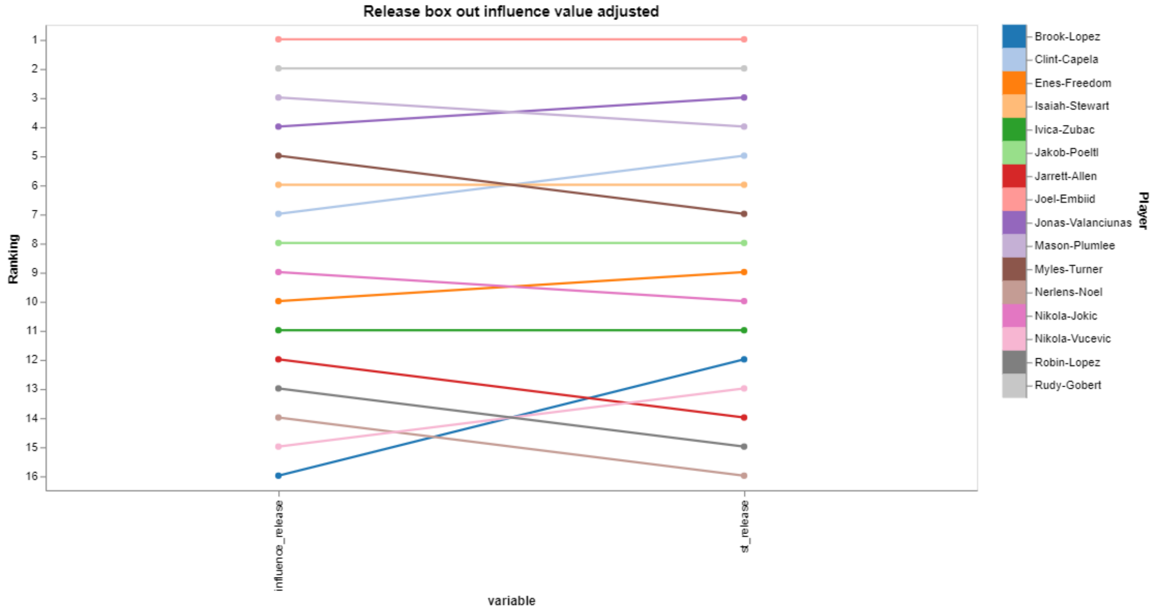


Figure 32: Release box out influence vs Release box out influence adjusted

However, if you look closely, you can see that two players are affected, but not in a very noticeable way compared to the others, those two are precisely Russell Westbrook and Giannis Antetokounmpo. This clearly supports what was discussed in 6.4: they are players who are very involved in the task of getting the rebound, and who are also very effective at it. On the other plot, 32, we can see how Brook Lopez benefits from this adjustment, as he has much more presence and influence on interior shots than on exterior shots. Notice that the shots that he contributes at avoiding are precisely those that he's more able to rebound himself, once again showing the counterintuitive nature of defensive rebounding.

Other plots regarding this adjustment can be seen in B for reader's curiosity.

6.6 No Brook, no party

To conclude the analysis of this study, it could be observed how teams and players behave when a key player for the rebounding process is not on the court, to understand even more the importance of this player in this process. This final section will focus on how the different metrics of the Milwaukee Bucks players are affected by the absence on the court of a player that has been studied in depth in this work: Brook Lopez.

The following plot, 33, shows how the metrics of the players who play more minutes of the team vary, depending on whether Brook Lopez is on the court or not

The results obtained show that the Hustle and conversion metrics do not vary greatly. However, it is true that there is a slight general increase in these metrics when Brook Lopez is on the court, which means that his presence is beneficial as it allows his teammates to be closer to the ball and to catch it. The first difference in the behavior of the players is in the positioning value metric at the time of release. Bobby Portis presents a considerable increase in this metric, which leads to think that he is the player who performs the role defending the interior zone during Brook's absence. There is also a big difference in the Crash metric. In general, players during Brook's presence tend to enter more in high value zones, since

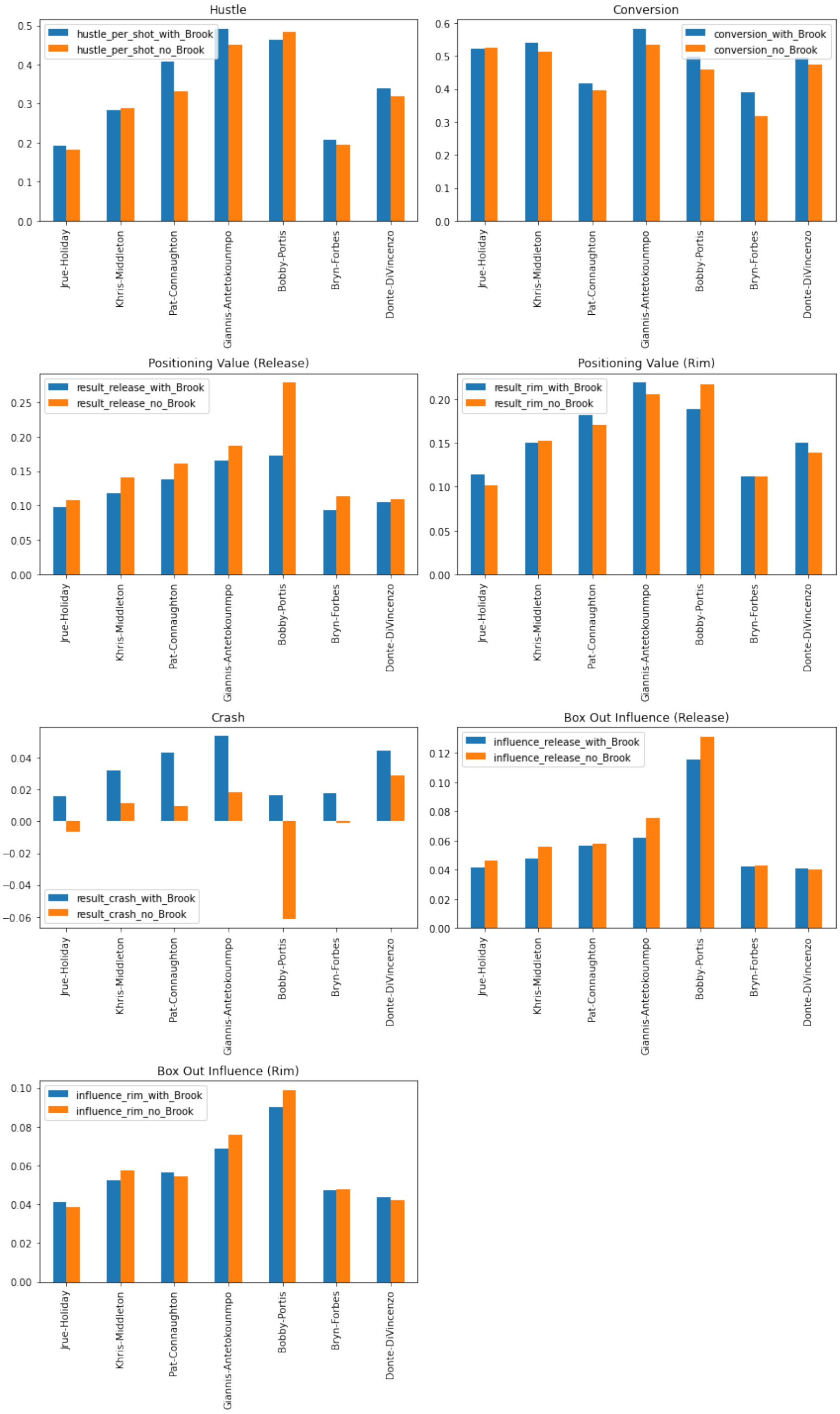


Figure 33: Metrics difference between when Brook Lopez is or is not on the court

the value of this metric is very high when he is on the court compared to when he is absent. When he is not on the court, this metric decreases considerably for all players, especially in Bobby Portis' case. Portis' metric acquires a negative value, probably due to the fact that he occupies the areas of higher value and, as discussed above, this type of players tend to have a negative crash value.

The above fact suggests that thanks to the work of Brook's box out, players have more opportunities to enter high value areas. This happens because when this player is not playing, no other player increases the box out value significantly.

7. Conclusions

Once the work is finished, it is possible to go back to point 2 to check if the aims established at the beginning have been fulfilled or not.

The first objective, to investigate and understand previous work done in this area, has been fulfilled. This learning and understanding of previous concepts is reflected in 3, as these have been used as a basis for the construction of this study. The second objective, one of the most important, has also been achieved. It was possible to reproduce the metrics used in previous studies along with the development of the *box out influence*, the original metric of the present work. Finally, a complete analysis was also presented using the aforementioned developed metrics as a basis, extrapolating player and team behaviors that go beyond traditional statistics and taking advantage of player positioning data. In addition, the development of some metrics has been useful to fulfill one of the objectives of the [2] study, thus mixing play-by-play data with metrics from this study to provide more value and contextual information.

To conclude this section, the importance of understanding rebounding not as an instant, but as a process that starts at the moment a player shoots (if not before), must be emphasized. As seen throughout this work, Brook Lopez may be the poster child of this type of analysis: a player whose very defensive function is avoiding the shots he is most able to rebound (rim/paint shots), and who creates great advantages for his team's rebounding on 3s while posting very mediocre traditional numbers.

In a nutshell, this project has been about appreciating his selfless approach from a quantitative standpoint. By building out new metrics, complex, but much more intimately related to the nature of the game, we can identify winning behaviors that, with traditional analysis, would have been ignored or penalized.

On a more personal note, this work has allowed me to work in a field that I am passionate about, and arouses my curiosity and desire to research those hidden aspects of the game that can only be found through the lens of data analysis. I have enjoyed every moment of the process of working on this project. This process and the meetings I have had the chance to hold with such experienced professionals in this field have greatly deepened my knowledge and passion for this game.

8. Future work

The following points discuss what the next steps should be to improve the existing results:

- Test new factors that weight the space partitioning using power diagrams. There may be a factor not mentioned in this study that is significant and makes the values obtained more directly related to reality. In addition, it could also be reconsidered to redo the analysis using the partitions weighted by the variables presented in this study and do a comparison work.

- Research on techniques or ways to avoid the correlation between the metrics and the initial positioning of the players. Introducing further analysis and conditioning to make all those metrics comparable across players, establishing what are realistic thresholds given initial positioning.
- Using data from other seasons would make it possible to analyze the evolution in players' behavior or new interactions between players who do not share a team right now but who did in the past. The type of analysis that we did at the end, understanding those interactions when a player was on or off the court, could be expanded as more seasons are added and the graph of interactions between players gets denser.
- Understand the value of particular players rebounding the ball, and how this influences the start and outcome of the following offensive play.

References

[1] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

[2] A Alarcón Román. Development of value metrics for specific basketball contexts: Evaluating player contribution by means of regression. Technical Report TFG, 2022.

[3] F. Aurenhammer. Power Diagrams: Properties, Algorithms and Applications. *SIAM Journal on Computing*, 16(1):78–96, 1987.

[4] Rasmus Bro and Age K Smilde. Principal component analysis. *Analytical methods*, 6(9):2812–2831, 2014.

[5] Peter A Burrough, Rachael A McDonnell, and Christopher D Lloyd. *Principles of geographical information systems*. Oxford university press, 2015.

[6] Daniel Cervone, Luke Bornn, and Kirk Goldsberry. Pointwise: Predicting points and valuing decisions in real time with nba optical tracking data. 2014.

[7] Ahmed Cheema. Analyzing NBA spacing with player tracking data, 11 2019.

[8] NBA Communications. NBA expands partnership with STATS LLC to unveil player tracking technology for all 30 NBA teams, 09 2013.

[9] NBA Communications. NBA announces multiyear partnership with Sportradar and Second Spectrum, 09 2016.

[10] John Fontanella. *The Physics of Basketball*. Johns Hopkins University Press, 2006.

[11] Kirk Goldsberry. How Rebounds Work, 2014.

[12] Phil Goldstein. Splunk .conf2017: Oakland a's billy beane says data analytics has transformed baseball, 05 2020.

[13] Grand View Research. Sports analytics market size, share & trends analysis report by component (software, service), by analysis type (on-field, off-field), by sports (football, cricket, basketball, baseball), and segment forecasts, 2021 - 2028. Technical report, 04 2021.

[14] Roberto Hernandez Hernández. *Metodologia de la Investigacion*. McGraw-Hill Education, New York, United States, 2006.

[15] Motokazu Hojo, Keisuke Fujii, and Yoshinobu Kawahara. Analysis of factors predicting who obtains a ball in basketball rebounding situations. *International Journal of Performance Analysis in Sport*, 2019.

[16] António Lopes, Sonia Fonseca, Roland Lese, and Arnold Baca. Using voronoi diagrams to describe tactical behaviour in invasive team sports: an application in basketball. *Cuadernos de Psicologia del Deporte*, 15(1):123–130, 2015.

[17] R Masheswaran, Y-H Chang, Jeff Su, Sheldon Kwok, Tal Levy, Adam Wexler, and Noel Hollingsworth. The three dimensions of rebounding. *MIT SSAC*, 2014.

[18] Emanuel Parzen. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3):1065 – 1076, 1962.

[19] Abhas Ricky. How Data Analysis In Sports Is Changing The Game, 12 2021.

[20] David W Scott. *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons, 2015.

[21] Brian Skinner and Stephen J. Guy. A Method for Using Player Tracking Data in Basketball to Learn Player Skills and Predict Team Performance. *PLOS ONE*, 10(9):e0136393, 2015.

[22] Sports Analytics Market. Sports analytics market by component (solutions, services), application, deployment mode, organization size, industry vertical, and region (APAC, north america, MEA, europe, latin america) (2022 - 2026). Technical report, 03 2022.

[23] Sports Reference. 2020-21 NBA Player Stats: Per 36 Minutes, 2021.

[24] Jenna Wiens, Joel Brooks Guha Balakrishnan, and John Guttag. To crash or not to crash: A quantitative look at the relationship between offensive rebounding and transition defense in the nba.

A. Appendix A: Metric Breakdown by zones

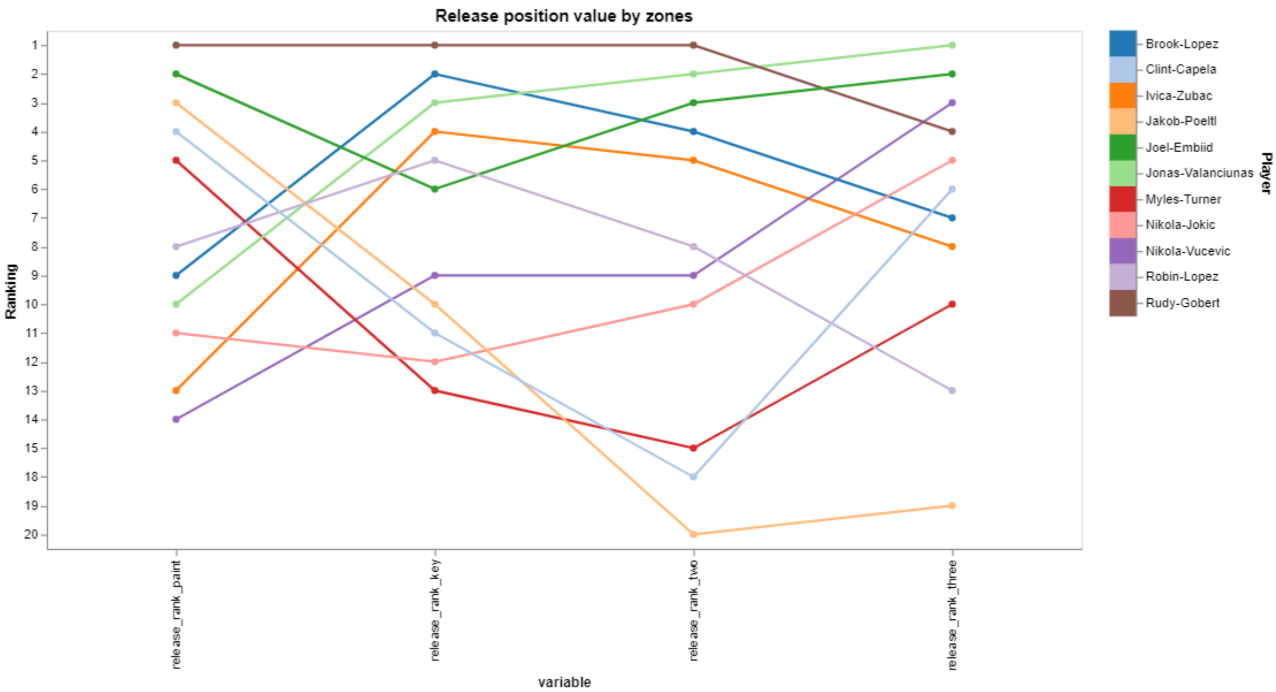


Figure 34: Release positioning values per shooting zone

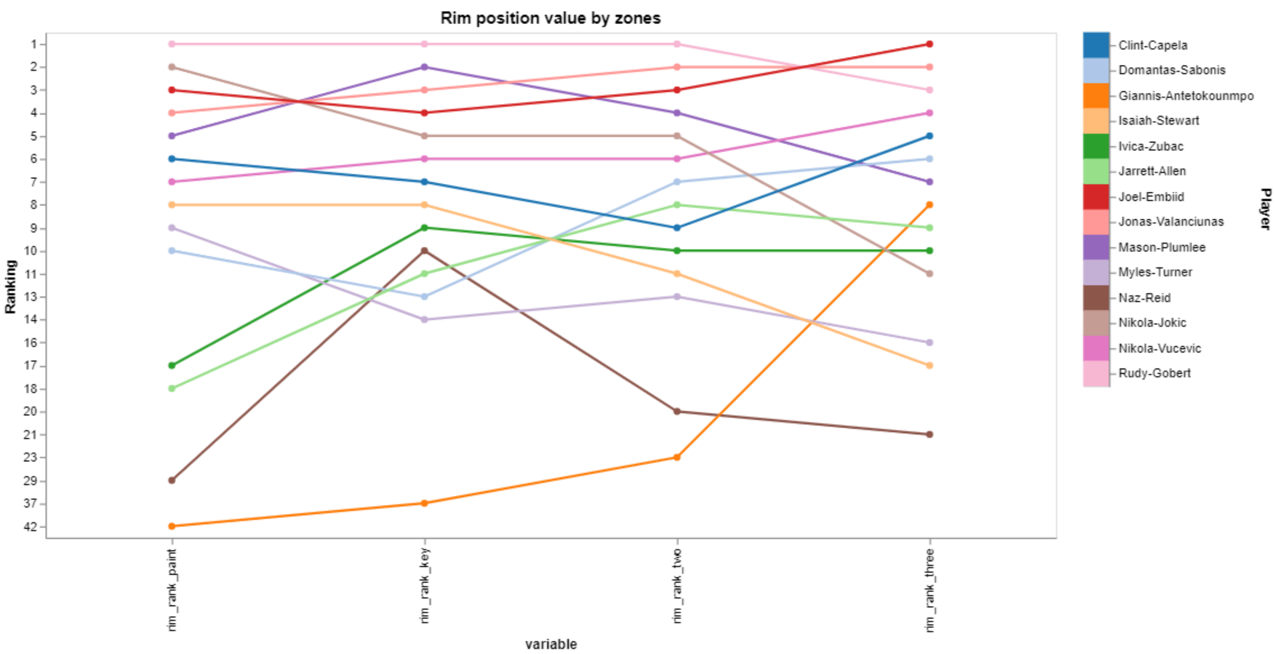


Figure 35: Rim positioning values per shooting zone

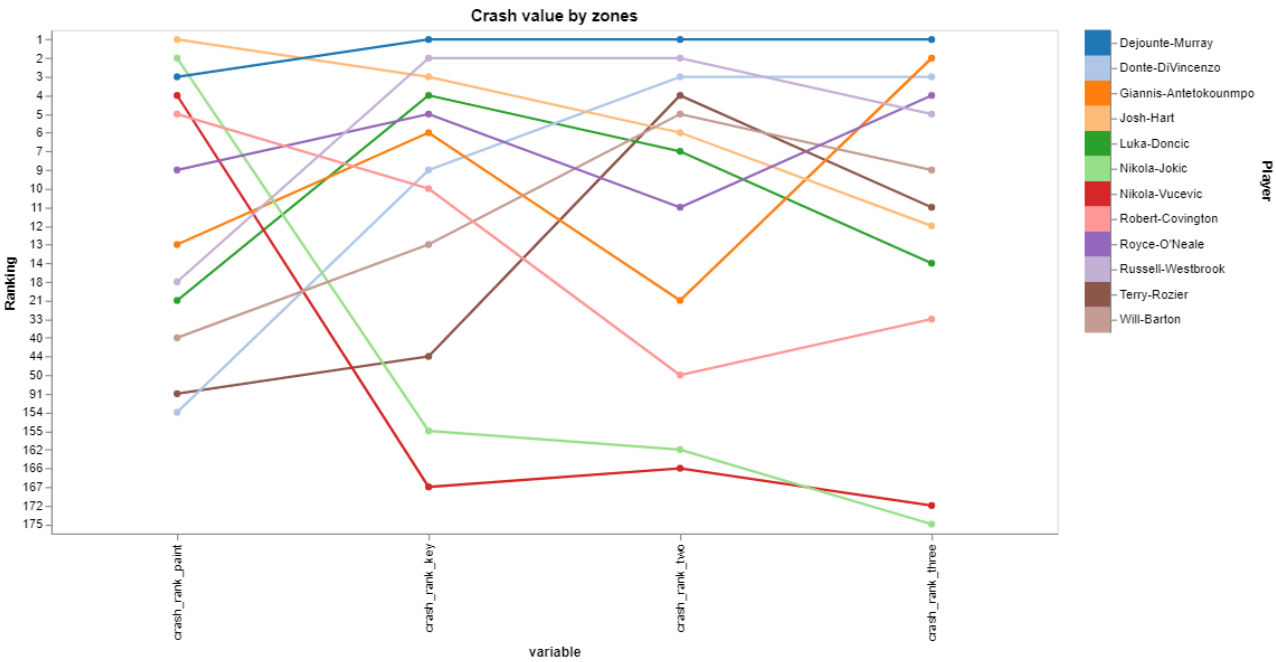


Figure 36: Crash values per shooting zone

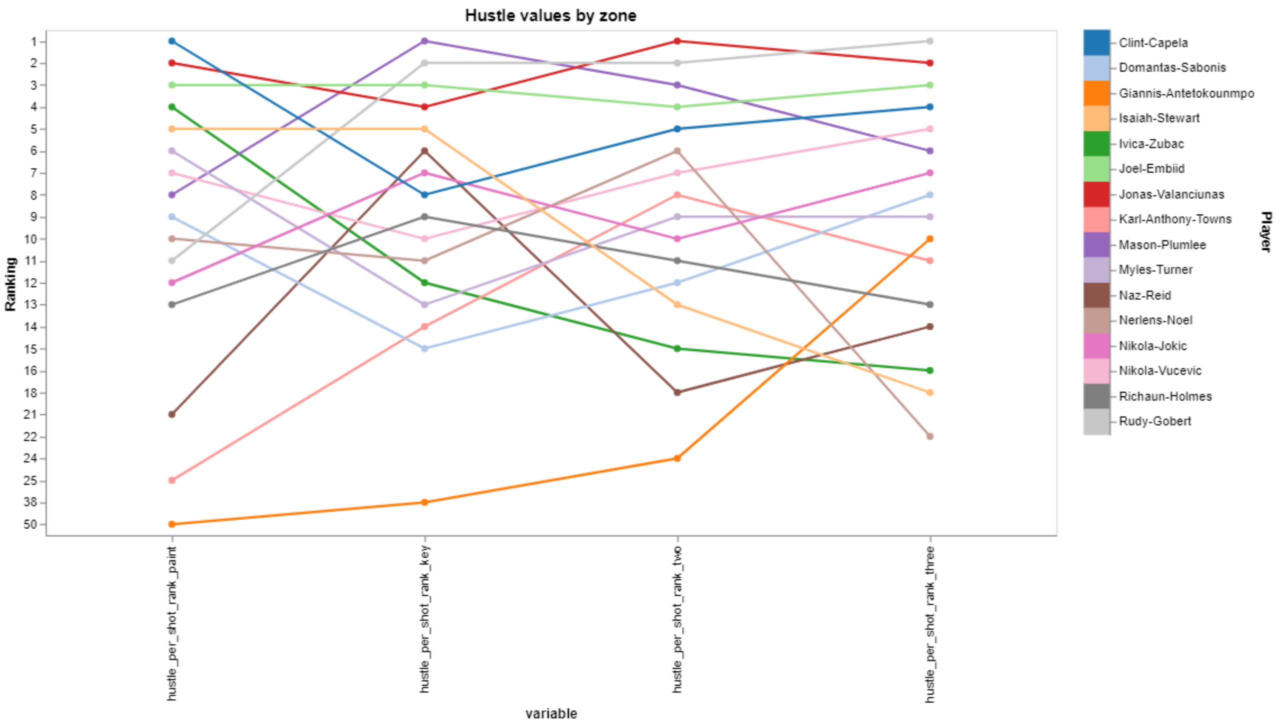


Figure 37: Hustle values per shooting zone

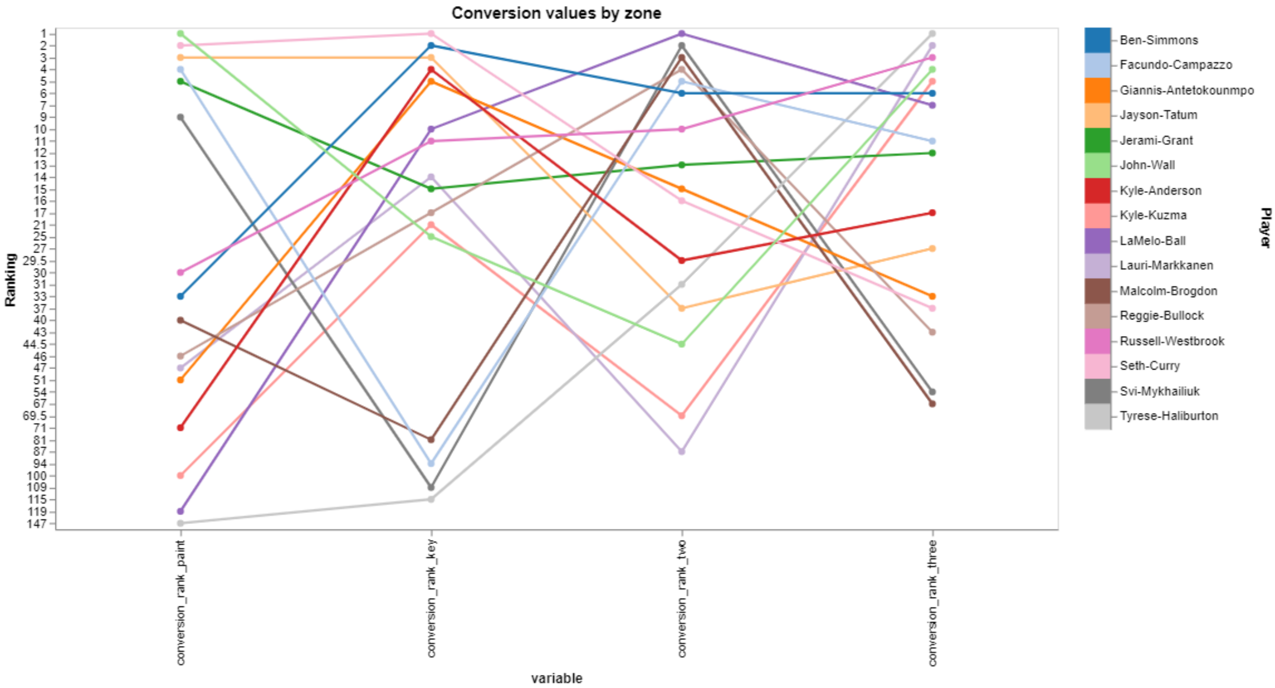


Figure 38: Conversion values per shooting zone

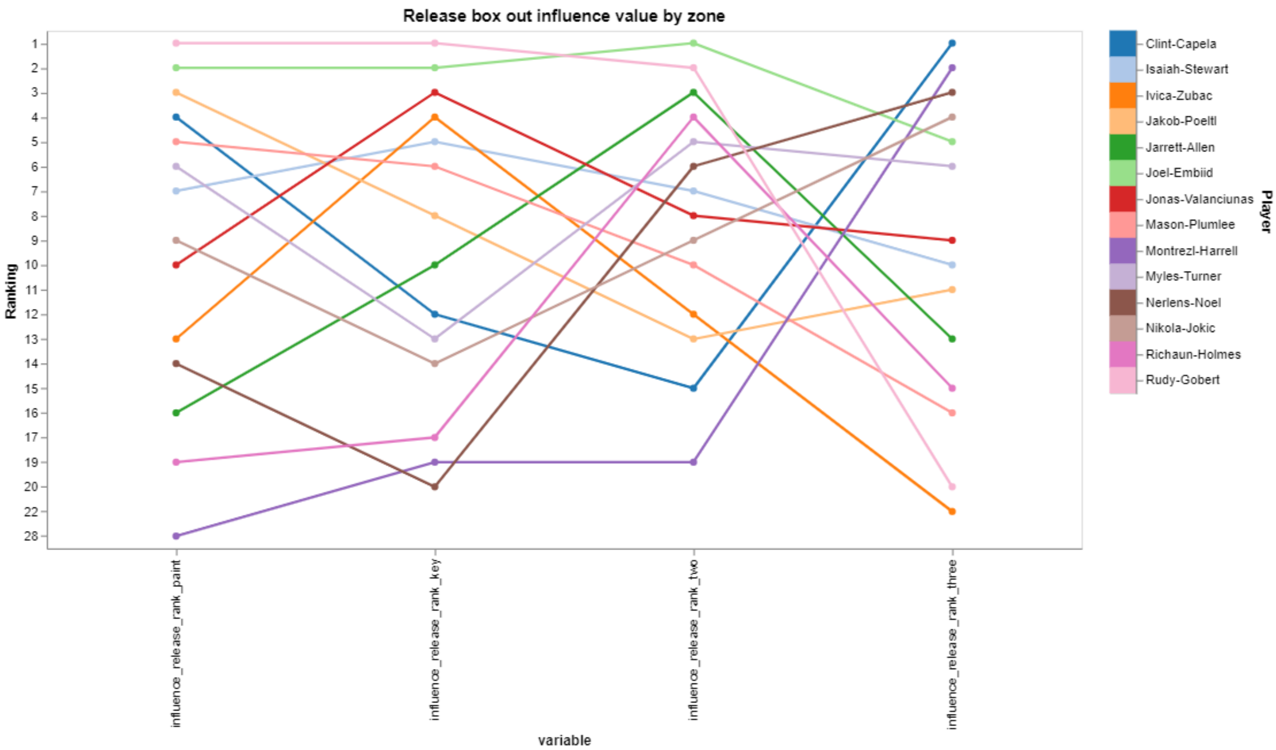


Figure 39: Release box out influence values per shooting zone

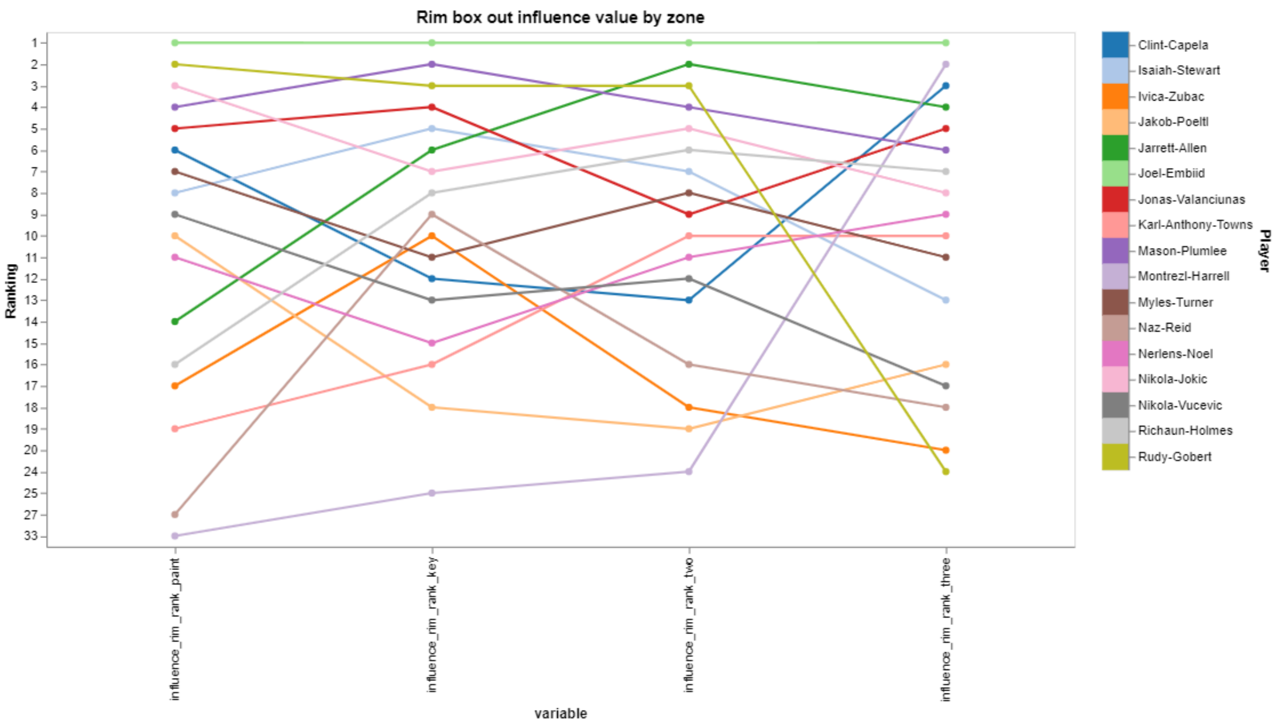


Figure 40: Rim box out influence values per shooting zone

B. Appendix B: Metric Adjustment

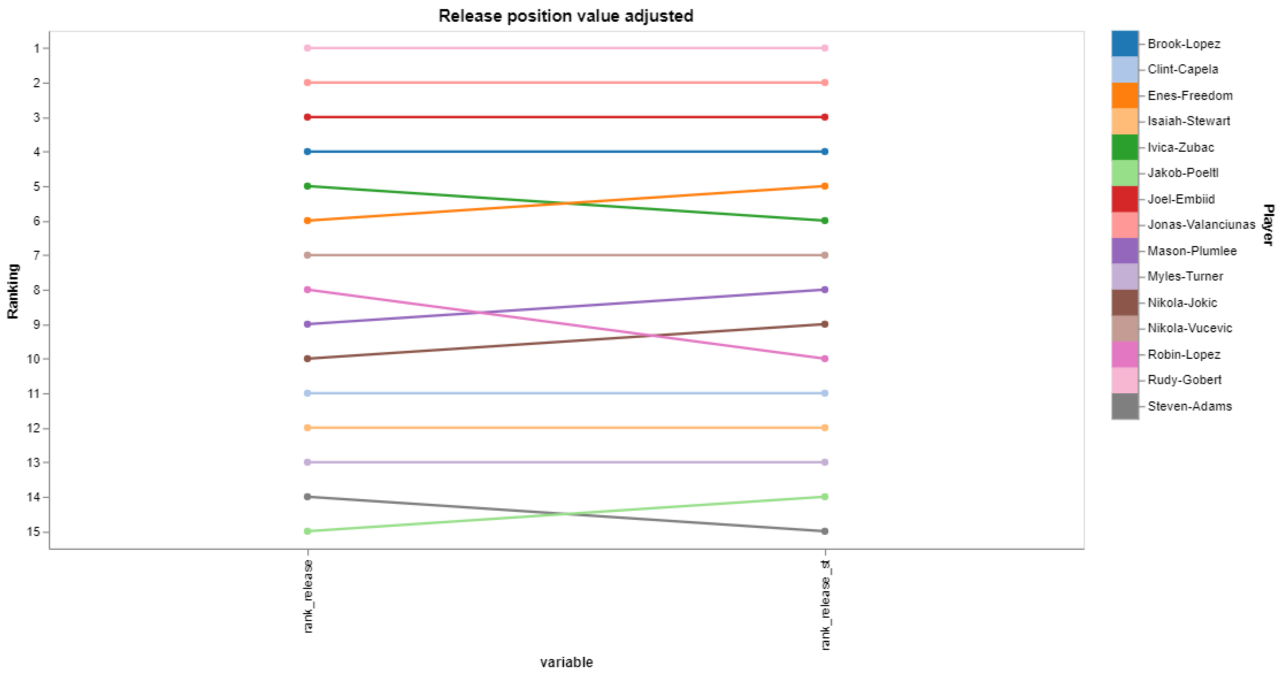


Figure 41: Release positioning adjusted

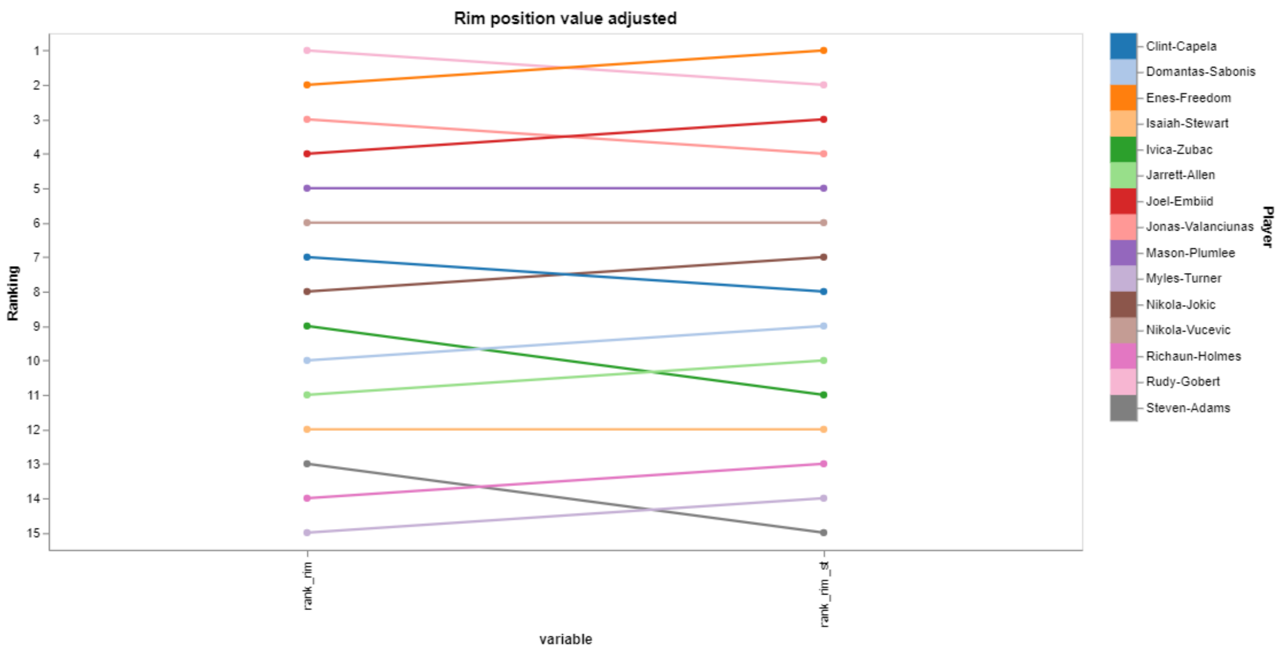


Figure 42: Rim positioning adjusted

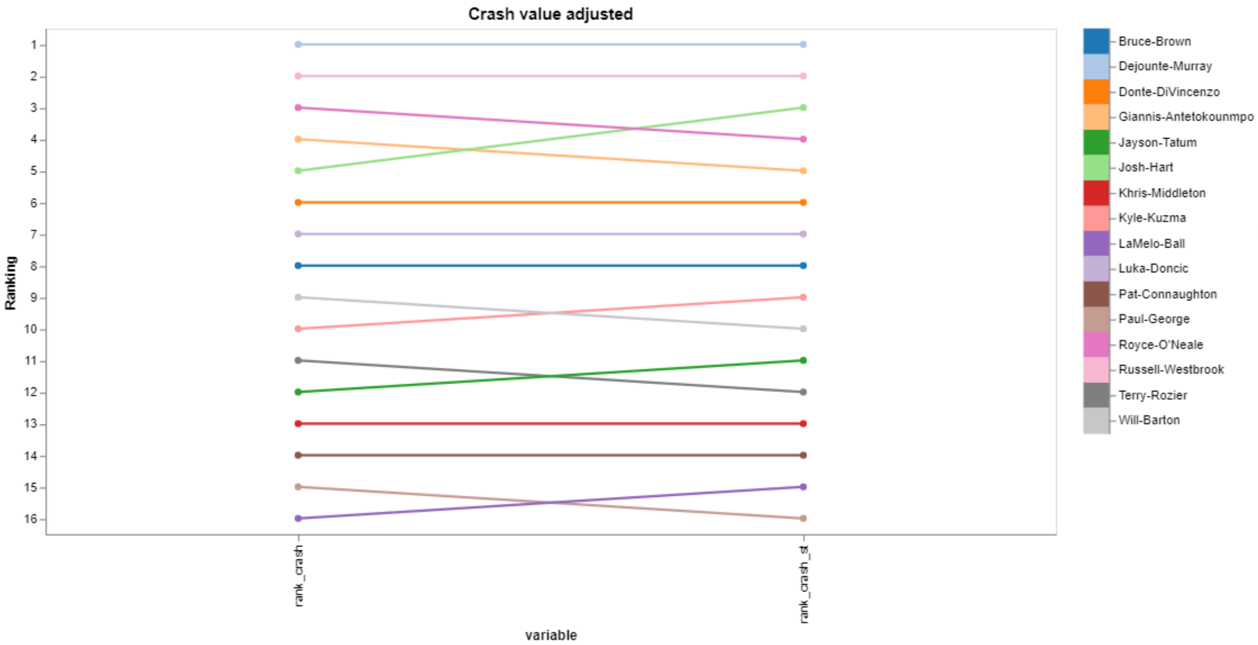


Figure 43: Crash values adjusted

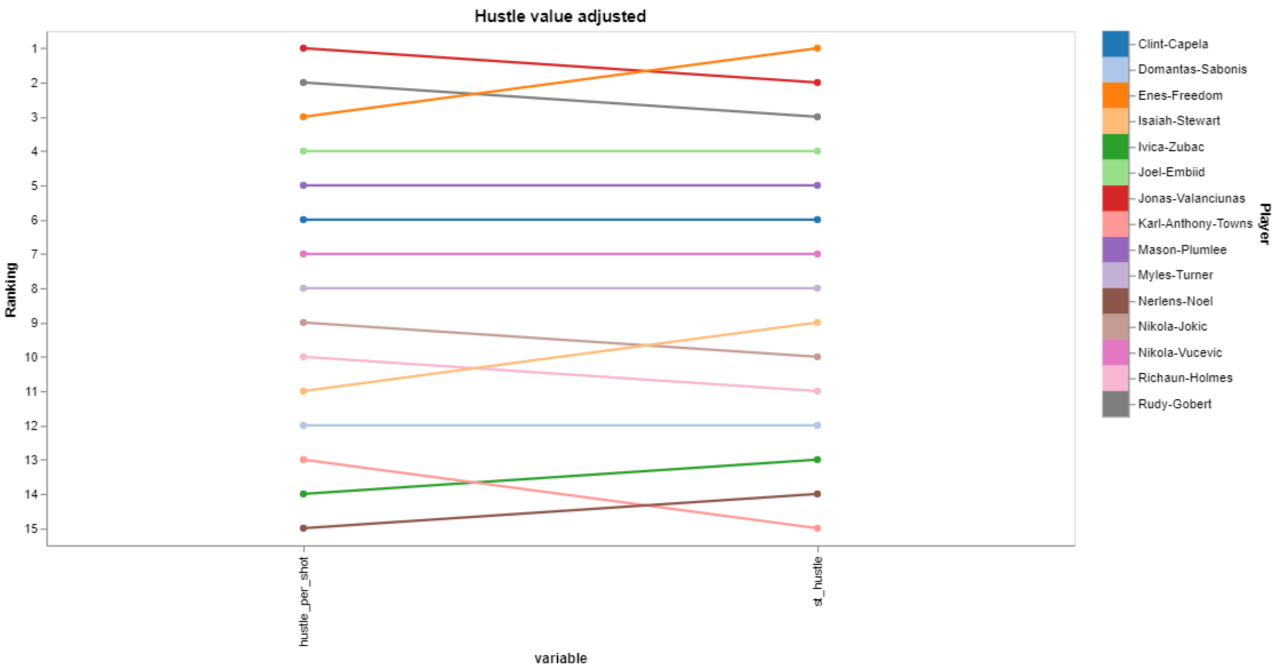


Figure 44: Hustle values adjusted

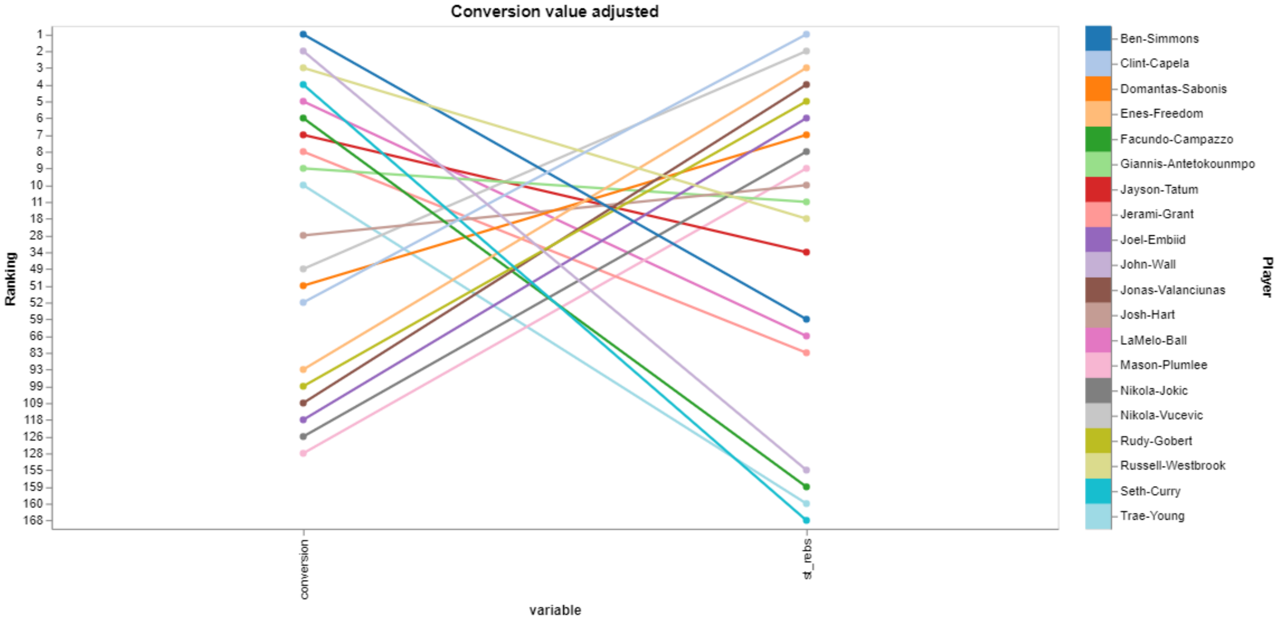


Figure 45: Conversion values adjusted

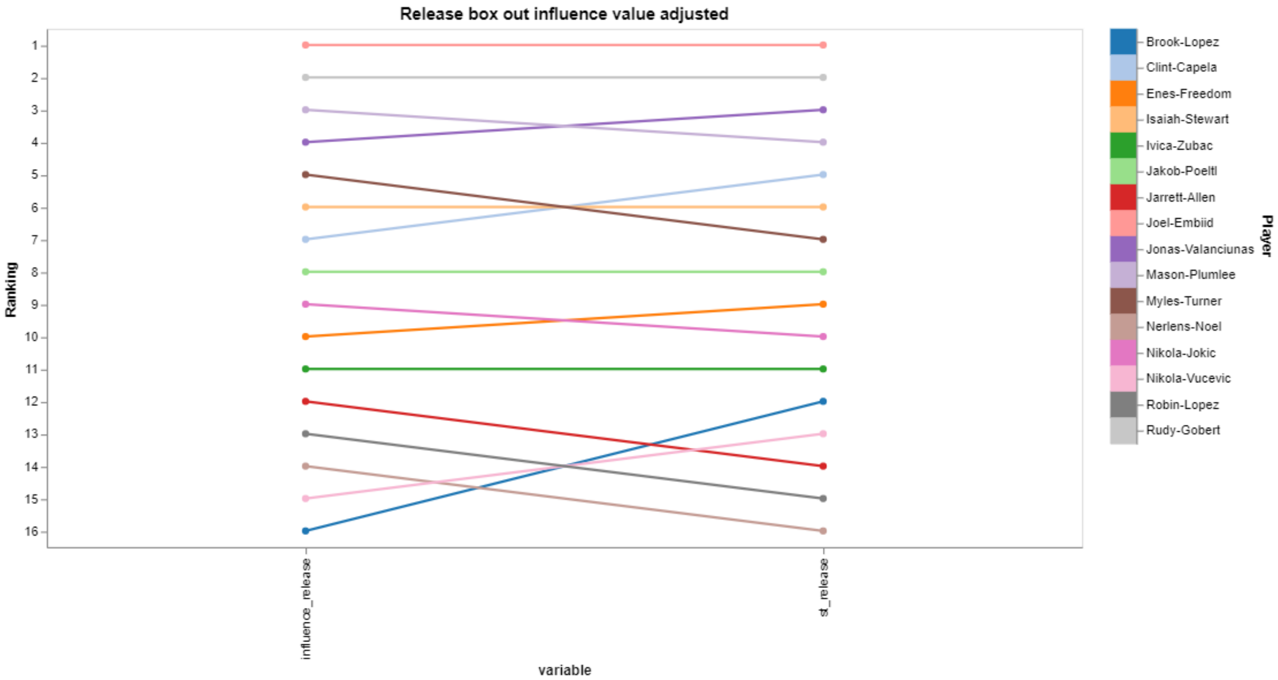


Figure 46: Release box out influence values adjusted

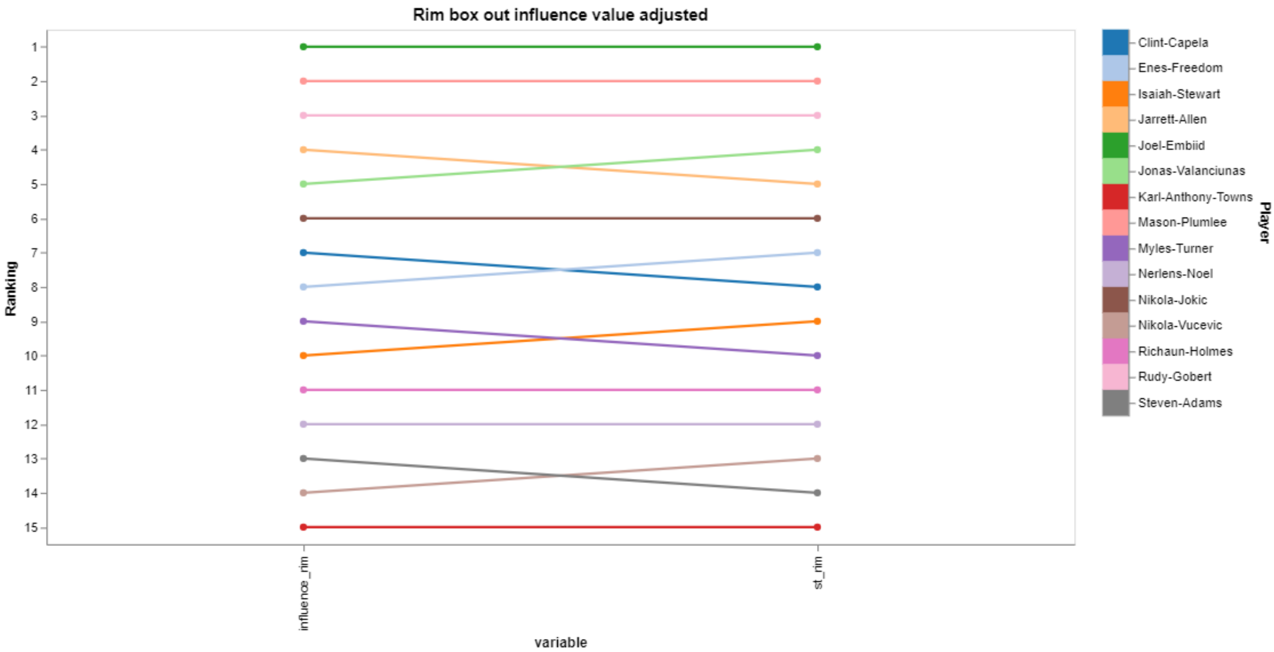


Figure 47: Rim box out influence values adjusted