NOVA IMS Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

Business Intelligence approach for *NBA* player & coach performance improvement

Introduction of Smart tools inside *NBA* environment

Marcelo Carrasqueira Sanches

Project Work presented as partial requirement for obtaining the Master's degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

BUSINESS INTELLIGENCE APPROACH FOR *NBA* PLAYER & COACH PERFORMANCE IMPROVEMENT

by

Marcelo Carrasqueira Sanches

Project Work presented as partial requirement for obtaining the Master's degree in Information Management, with a specialization in Business Intelligence and Knowledge Management.

Advisor: Professor Miguel de Castro Neto, Phd

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to everyone that was part of this project.

To my advisor, *Professor Miguel de Castro Neto*, for accepting and encouraging me to develop my own topic and idea. It would not have been possible to finish this project without the follow-up meetings, where his guidance and constructive feedbacks were fundamental to achieve the final goal.

To my sister, mother, and father. Thank you for your support, by understanding the time that I put on this work, and for making this challenge easier with your constant support.

To my friends. Thank you for your honest friendship, feedback and opinions that were very important in the whole process.

ABSTRACT

The National Basket Association (*NBA*) is considered the premier professional basketball league in the world and is seen today as one of the greatest attractions alongside the *NHL* or the *NFL*, due to its great shows and players. This research project will contribute to the theory and practice by establishing the added value that Business Intelligence tools can provide, as well as the positive impact it can have on a player, coach, and consequently on a team's decisions and future results. First, a literature review will be conducted to understand how reputation and status can impact a player’s environment; how statistics are an important asset in the evaluation of player´s performance; and also, in the identification of over-paid and underpaid players. Thereafter, Power BI software will be used to develop reports and dashboards. The main goal is that, both players and coaches, can improve their ability to interpret and analyze their own performances. In the end, they are not only adding value to their teams (through this smart tool), but they are also creating an environment where the teams are one step closer in what is the definition of a smart team.

KEYWORDS

Business Intelligence; Performance; Improvement; Overpaid; Underpaid; Power BI; Dashboard; Smart Team

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

NBA	National Basketball Association
BI	Business Intelligence
DT	Digital Transformation
MVP	Most Valuable Player
DPOY	Defensive Player of the Year
ROY	Rookie of the Year
PER	Player Efficiency Rating
VA	Value Added
EWA	Estimated Wins Added
PROD	Player’s Production
P48	Player’s Production in 48 minutes
WP	Wins Produced
WP48	Wins Produced per 48 minutes
ETL	Extract, Transform & Load
OLTP	Online Transaction Processing
OLAP	Online Analytical Processing

INTRODUCTION

1.1. INTRODUCTION AND PROBLEM IDENTIFICATION

The growing of **Digital Transformation** can be perceived in several areas nowadays, and the use of its benefits may be applied in the fields of health, business, education, and sports. In fact, it is in the sport's field that this project will be conducted. All American professional leagues – *NBA*, *NHL*, *NFL*, *MLB*, etc. – are widely known for their high level of technological, financial, and human resources. Since it would be quite difficult and imprecise to approach all the existing professional sports leagues, this project will focus specifically on the National Basketball Association environment – using teams, coaches, and players. Therefore, the project aims to study – in a more detailed way – the added value that certain Business Intelligence tools can impact the performance and evolution of an *NBA* player, coach, and respective team.

It is possible to argue, the existence of certain *NBA* teams with large investments to achieve better results – such as, reach the playoffs or even win a championship – that most of the times they tend to not achieve the expected results. Although there is a lot of information on statistics; performance metrics; and comparisons to evaluate the added value of the player, there is a small number of academic works that compile – in a single tool – both quantitative and qualitative information as a report or a **Power BI** dashboard. It is increasingly important for a player to be aware of his own performance; his physical form; or his evolution over time, to learn the best way of managing not only his failures but also his performance as a team contributor. As it is equally crucial for a coach – and respective team – to be capable of overcoming problems such as overpaid salaries; reputation issues; overrated and underperformance players. As a result, better decisions can be made on team management, player contracts and results, always taking into consideration the information that gathers a set of facts that does not allow mistakes induced by others.

The utilities of this multifaceted **Business Intelligence** tool can lead – based on a basic interpretation and knowledge created within the outputs – to a higher level of self-evaluation and consequent self-learning. This kind of knowledge, when applied as a whole, will become an improvement for the entire group, allowing better decisions at technical, tactical, physical, and financial levels. This way, it is possible for a franchise to be seen as a "**Smart Team**".

1.2. STUDY OBJECTIVES

This project will contribute to the literature by exploring and highlighting the benefits of a proper use of Business Intelligence, being one of the most well-known and useful concepts nowadays. Moreover, through this literature review, the basketball dimensions that can contribute to better understanding of the *NBA* environment – around players and teams – will be analyzed, to answer the following main research question: *"Is it possible to create a direct connection between players performance, coach decisions, money, and team management, with the aim of creating a highly successful team with better results?"*. To support and prepare a more detailed analysis, the question will be divided into three other research questions that will be addressed individually in the literature review: *"What is the best way to evaluate a player performance?"*; *"Which impact can reputation and status have on players' environment?"*; & *"How can overpaid players affect the team and basketball league?"*.

In a final note, the main goal is to enhance – in a faster and more constant way – the quality of every team, through the players and coach's interpretation of their performances, enabling both coaches and teams for better decisions in the future.

Specific Objectives:

- ✓ Contextualize *NBA* and introduce statistics as a crucial performance measure. *(Literature Review)*
- ✓ Demonstrate that reputation and status have an impact on players environment. *(Literature Review)*
- ✓ Prove, based on statistics and other studies, the existence of overrated and overpaid players, who consequently affect the team and league negatively. *(Literature Review)*

Main Objective:

- ✓ Present a business intelligence solution that can prove the added value of self-learning around the visualization and interpretation of dashboards. (*Project Development)*

1.3. Study Relevance and Importance

Nowadays, data science tools – that have been growing in this era of *DT* – are able to gain influence and affect every field that exists in our daily lives in a way that we are changing our behavior.

It is obvious that sports management is not excluded from this *DT*. 'Outside' of the court, fans increasingly want to have access to statistics in real time on their mobile devices – e.g.: Telephones, Tablets, Smart Watches – most of arenas already have Wi-Fi to provide supporters with the ability to communicate with each other, and also online access to multiple features **(Ströbel et al., 2019)**. When it comes to digital transformation inside of the court, there are elements that are already somehow connected to the player, and to the game quality, **(Bailey, 2016)**, such as:

- ✓ *Sport VU*: camaras that collect 25 times more data per second than normal ones – always following the movement of the ball – providing a more detailed analysis of any illegality that may occur during the game that cannot be seen by the human eye.
- ✓ *Wearable Technology*: Devices that allow you to store data about the player, including heart rate, distance travelled, etc.
- ✓ *Mobile Devices:* Use of tablets during the game as well as training sessions, to replace the usual "whiteboard".

It is important to note that, this last element will have a key connection to Power BI software, since it will be from these mobile devices, already well implemented, that players and coaches will also be able to see and analyze the reports and dashboards corresponding to their performance's evolution.

On one hand, regarding the evolution around the game, and the sport itself, it can be stated that *DT* is already influencing this field in many areas **(Ströbel et al., 2019)**. On the other hand, Vial defends that a holistic understanding of digital transformation in sport needs a broader approach **(Vial, 2019)**. Therefore – from the last statement – we can conclude that it may be interesting if we pay more attention to the player himself when it comes to implementing or developing tools – that could contribute directly to his qualitative evolution. Thus, the potential of this type of tool – in an end-user form – can be individually and collectively explored in a manner of acquiring the most profitable way to use it, and finally achieve its optimal effect and prove its value.

It is worth to mention that, as far as it concerns to projects or dissertations, there is a lack of information on this specific subject, however, it can also become quite challenging to approach this topic. Therefore, this study will certainly come to a highly positive conclusion regarding the way of getting benefits from this type of *BI* tools – facing the lack of decisions and subsequent poor results of some "Super Teams".

1.4. PROJECT RESEARCH DESIGN AND STRUCTURE

1.4.1. Research Methodology

With the initial research, it became clearer that the project development would be based on *Design Science Research*, where the conceptual model is derived from the analysis made in the literature review. According to some authors, this process is known by the combination of qualitative and quantitative research, references and research methods **(Offermann et al., 2009)**. Besides, the process is usually composed by six steps, such as, *problem identification and motivation*; *identify objectives*; *conceptual model design*; *development*; *evaluation;* and *communication* **(Rossi et al., 2015)**. This type of approach can be particularly useful, since it is guided by problems that can be observed in real life; it is also a solution-oriented; and can easily contribute as a positive asset that will develop knowledge and strategies, which are employed by professionals. **(Koneru, 2018)**.

The work that will be developed in this '*Business Intelligence Approach for NBA Player & Coach Performance Improvement'* follows the characteristics mentioned above. The project includes a *qualitative research* – by analyzing both, performance, and results, obtained by the players – as well as *quantitative research* – using statistics. It also follows all steps that Rossi and the other authors describe. Finally, the project has the purpose of reaching the smart team concept, by obtaining a solution that is able to solve the lack of great decisions by coaches and players.

1.4.2. Project Structure

To conduct this study in a more structured and coherent way, the main steps are mentioned below, along with a brief explanation:

The first step consists in the research of the problem found in the best possible way. It will be necessary to understand the relationship between reputation and status within the *NBA* environment; to prove the existence of overpaid players when compared with their performances; and show the importance of statistics, in a way that it is possible to understand the presence of valuable players, and if they can provide added value and success to their teams or not.

The second step involves the Literature Review draft and development. To prepare this literature review, it is part of the process to select the most relevant topics for the problem identification; identify the recognized authors for each topic; contextualize and justify the problem research; include relevant information and citations.

The third step consists in the elaboration of a conceptual model for the project. For this, it will be presented, some statistics; formulas and metrics to analyze the player's evolution; a dimensional model draft; and finally, the players and coach reports design – with dashboard extensions. However, it is relevant to mention that the coach's report will have almost the same player dashboard structure, but with some extra useful information.

The fourth step will focus exclusively on the *ETL* process, which will fulfill all the necessary requirements to create a proper and useful dimensional model. Once the data has been pre-processed, it will be downloaded into the dimensional model, to check if the data warehouse is ready to apply.

The fifth step requires the application of the conceptual model designed to build the first prototypes. The purpose of this prototypes, is to get a real vision of the final output and offer the best information possible to the final user.

The sixth step will explain how Business Intelligence can impact in terms of learning and quality improvement – for player, coach, and associated team. It will have a section to evaluate results, one to draw some conclusions, and another to share some future suggestions.

2. LITERATURE REVIEW

Prior to developing any project with the intention of providing tools that can lead to better decisions and consequently to achieve a smart team, it is crucial to have some contextualization. To clarify the topic, it is important to explain how certain factors can influence aspects such as performance; quality; budgets; and team results. Most of basketball studies – as well as related articles – addresses statistical and factual issues with mathematical approaches, but there are also some ethical and abstract research using logical considerations.

The aim of this literature review is to build a logical path, that can justify the reason for the added value of implementing Business Intelligence tools and dashboards within *NBA* teams. To trace this logical path, the structure of the literature review is formed ***primarily*** by a brief contextualization of the *NBA* and how the player's performance can be measured through statistics. The ***second*** topic is a clarification around the influence that reputation and status can have, in terms of business and results, on a player and a team. ***Subsequently*** – after the conclusion that there may be overrated and overpaid players – an analysis will be made around articles, that can justify the wage disparity in the league. This analysis will be made around two authors who have created two metrics, "Estimated Wins added" and "Wins Produced". This way it is possible to relate the player's performance to the teams' budgets. Followed by this line of thought, will be made an interpretation based on consequences that the evolution of Salary Cap and the player's wages have in team results.

2.1. NBA History and Contextualization

NBA is one of the four major American sports leagues – along with National Hockey League (*NHL*); National Football League (*NFL*); and Major League Baseball (*MLB*) – founded on June 6, 1946. It is considered the most well-known Basketball league in the world due to its competitiveness, its high-quality players, its entertainment offered to the audience, and of course, the millions of dollars it generates. Nowadays, the league is composed by 30 teams (29 American and one Canadian) geographically divided into two conferences (east and west) each one with three separated divisions – making a total of six divisions: Atlantic; Southeast; Central; Northwest; Pacific; and Southwest Division. The competition is structured by 82 regular season games, followed by a playoff between the eight best teams of each conference – each round is decided at the best of seven games – finally, the championship run is settled between each playoff conference winner – also decided at the best of seven games **(NBA, 2019)**.

2.1.1. Measuring performance through statistics

Over the years, the growth of *NBA* players, with a high level of quality, raised the need to analyze their performances in more detail. In the face of this facts, it is normal that the analysis of statistics in modern sports is helpful and quite relevant in these days **(Yarrow & Kranke, 2016)**. Therefore, stats can be used for many purposes, such as, access detailed information, to redirect tactical decisions – before and after competitions – **(Yarrow & Kranke, 2016**); to determine how efficient a player is – while he is on the court – **(Bension, 2019)**; to rating or ranking a player, and give some guidance to his consistency **(Hauser, 2009)**. According to Bension (2009), the classic box score is composed by a set of offensive stats – Points (*PTS*); Offensive Rebounds (*ORB*); Assists (*AST*); Fields Goals Made (*FGM*) and Fields Goals Missed (*FGm*) – as well as defensive stats – Defensive Rebounds (*DRB*); Steals (*STL*) and Blocks (*BLK*). Based on these statistics it is quite simple to create several types of analysis, as an example, we can mention the metrics developed by **(John Hollinger, 2007)** and **(Berri, 1999)** – Player *PER* and *WP* respectively; which are linked to everything that relates the performance of a player and a team. Hollinger's method is more focused on the offensive part of the game by using just offensive stats, while Berri's considers both offensive and defensive stats.

In the development of this project, a more related approach to Berri's will be followed since both the offensive and defensive part of the game will be taken into consideration. Therefore, the dashboards construction for this work, will involve the use of data that are mostly collected from NBA **(NBA & SAP, 2020)** – that provides a much more complete range of statistics of both player and team, when compared to Bension's and Basketball Reference official websites **(Sports Reference, 2020a)**.

2.2. Reputation & Status

One of the purposes of this project is to find potential problems that may affect the quality and performance of a player and the respective team. In April 2013, Gokhan Ertug and Fabrizio Castellucci discussed and compared the impact that Reputation and Status has on a team in terms of performance, revenues and hiring decisions. They state that both intangible assets have effects on the final product quality (game performance) and revenues (tickets sold) **(Ertug & Castellucci, 2013)**.

2.2.1. Difference between Reputation & Status

On one side, *Reputation* is based on past actions or experiences that demonstrate the quality of the actor, assigning a specific value that will distinguish him as a possible quality resource in the future. The organizations use reputation to infer skills, knowledge, and the quality of a resource provider, therefore the actors (players), who have more reputation, are expected to provide high quality resources (players performance) and thus increase the final product quality (team performance). The authors also state that, it is expected to pay a higher price for resources (players) that are obtained from high reputations, since a positive result is expected from that acquisition **(Ertug & Castellucci, 2013)**.

On the other hand, the previous authors argue that, *Status* is based on a hierarchical order, formed on the basis of some factors such as education, awards and prestige. Thus, these indicators are less directly tied to quality than reputation, so, when we are in the presence of a high level of status, we are expecting to have superior ability, and this is what teams prefer to work with. Having said that, status is an intangible asset that is valuable to the organization as it adds more value, more visibility, becoming more attractive in the eyes of their audience (supporters), resulting in a larger effect on revenues (tickets line revenues) than reputation. Finally, the authors also state that, it is expected to pay higher prices to resources that would certify their association (team) with more status resource providers, resulting in the increase of status benefits, such as higher revenues.

In conclusion, the authors argue that, organizations with low quality resources (low quality players) when comparing to their own goals, are likely to acquire more High-Reputation providers than High-Status providers. On the other side, organizations with revenue problems, when compared to their aspiration levels, tend to recruit more High-Status resource providers than High-Reputation providers.

2.2.2. Transition from Reputation to Status

After a deep interpretation of the work carried out by Ertug and Castelluci around reputation and status, it is possible to build a self-logical sequence for the process of this topic. The first *state* of the sequence requires the existence of a certain quality experience in a past behavior, in a way that reputation can exist in a proper and founded way. Since it is possible to say that there is a solid *reputation* of the actor, it is expected that the resource (player) will be able to offer quality in his actions, which in turn will translate into the quality of the Final Product (Success of the team). With these 4 stages completed, it is quite simple for any team – if we are thinking as a whole – to establish a position within a certain qualitative hierarchy, and therefore build a status around this whole process. A team which achieves a high *status* – due to its positive results – may attract more fans and associates, which results in an increase of the number of tickets sold, leading directly to an increase in their revenues.

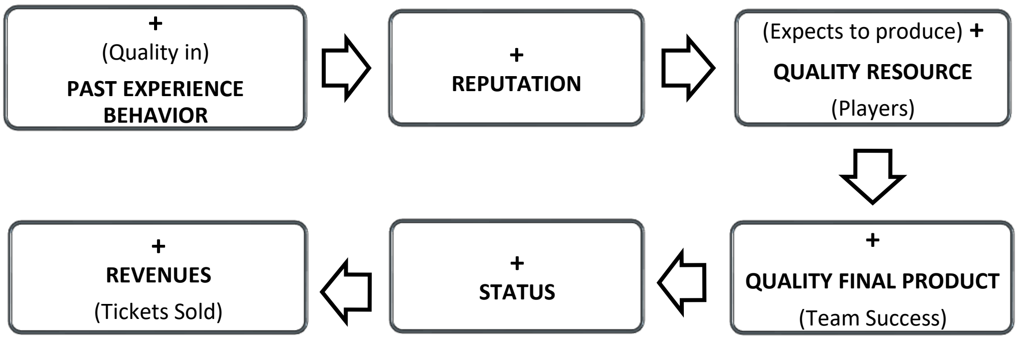


Figure 1: Transition from Reputation to Status

2.2.3. Status Impact in a Long Term

The increase in a team's revenues is always quite significant for financial success, especially if it is directly justified by status. According to **(Washington & Zajac, 2005)**, in a scenario where the status is determined by its past quality, once it is formed, becomes more difficult to decrease in terms of qualitative characteristics of the team, and when it decreases, it tend to happen slowly, even in the face of recent negative changes within the performance quality. This argument is also justified by the fact that Status is less sensitive to changes in performance when compared to reputation **(Ertug & Castellucci, 2013)**. As long as a status ordering is settled, it tends to perpetuate itself and the benefits that are associated with it, independently of the “merit system” **(Bol et al., 2018; Podolny, 2016)**. Based on these arguments, it is possible to conclude that there are players who have already achieved high levels of reputation and have performed below the expectations in some seasons. However, these players still maintain their privileged position in the league due to the status that the team provides to them. With this type of scenarios, it is expected that, in a short-term analysis, ticket line revenues will continue to be positive due to past privileges, independently of the performances presented by players in the present.

2.2.4. Reputation & Status inside NBA

Applying a parallel approach to the NBA, as it was done partially so far, it is possible to deduce that the players are the resource providers and their quality is evaluated through their statistics, such as points (*PTS*), assists (*AST*), rebounds (*RB*), field goal percentage (*%FG*), among others. **(Ertug & Castellucci, 2013)**. According to these stats and the certifications that are given to the players – such as *MVP*, *NBA* All-Star Game[1] selection, All NBA etc. – teams can choose which players they want to keep, which ones they want to dismiss and which ones they want to add to their roster. However, this type of awards and choices are based on votes attributed by coaches, sportswriters, broadcasters, and fans **(Wade et al., 2006)**.

In the end it is quite clear that all sportswriters, broadcasters, and fans are elements with basketball knowledge, but they are not coaches, who are constantly on the pitch dealing with the players every day in every practice. As a result, these three elements probably do not have the internal knowledge – player's behavior in a training environment or even in the "changing room" – to make such decisions. However, they end up with crucial votes in what is the choice of some honors, merely based on player's superficial condition and game performances. The fact that there are fewer professional elements in the voting for individual distinctions – such as *MVP*, *DPOY*, *ROY*, among others – also justify, the existence of overrated players with higher status and reputation, which are not reflected in their performances, and tend to provide them with an excessively remuneration.

To explain this type of situation, a brief *example* can be used about the discrepancy between players' choices in the All-Star game. In the last 2019/2020 season, some players with excellent per game performances – Bradley Beal, second best performer per game of the season with: 30.5 *PTS*, 6.1 *AST*, 4.2 *RB*; Devin Booker with: 26.6 *PTS*, 6.5 *AST*, 10. 8 *RB*; and Zach LaVine with: 25.5 *PTS*, 4.2 *AST*, 4.8 *RB* – were put out of the *All-Star game* and all of them with only one common characteristic, they belong to losing teams that did not reach the playoffs. On the other hand, in the same season, there were players with lower numbers than those mentioned above – the veteran Chris Paul with: 17.6 *PTS*, 6.7 *AST*, 5 *RB*; 2018/2019 champion Kyle Lowry with: 19.4 *PTS*; 7.5 *AST*; 5 *RB*; 2020 runner-up Jimmy Butler: 19.9 *PTS*, 6 *AST*, 6.7 *RB* – that curiously have in common the fact that all of them belong to winning teams that reached the playoffs **(Goldberg, 2020; NBA & SAP, 2020)**. Facing these facts, it is possible to verify a problem with a pattern of decisions that are not directly reflected in the players' performances, but choices that are based on some external factors.

---

[1] The All-Star Game is a basketball exhibition game, normally hosted every February by the National Basketball Association (*NBA*) – as well as it is done by others American professional sports leagues, such as *NFL*, *NHL* and *MLB* – and showcases the best 24 league's players.

2.3. Overrated & Overpaid Players

It is quite common, inside the NBA environment, the existence of opportunity cost problems. Agents, coaches, supporters even the players are very familiar with the fact that in some cases we are dealing with overrated cases who have salaries that seem quite high when compared to their performance. However, there is also the other side of the coin, players who do not have such a great reputation – since they belong to losing teams with lower status – that are the same players who have performances and statistics well above average. Usually these players – e.g., Bradley Beal and Devin Booker – have been affected in terms of wages, sponsorships, advertising agreements or even prestigious brands partnerships. According to the Forbes editor Brett Knight, there is also a challenge in finding the right player at the right price, due to the fact that we often focus on what the player has already done, and not what he is doing at the moment **(Brett Knight, 2019)**.

2.3.1. Overpaid players based on statistics

Statistics are, not only one of the most used tools in data science nowadays, but also the faster way to analyze a player's evolution and check if his performance corresponds, or not, to the expectations. However, there are authors who use these numbers to achieve more qualitative conclusions. Tyler Stanek (2016) has produced a player salary analysis – based on performance and team revenues – and has proven the existence of both, under and overpayments in the league. His line of thought was built on two metrics created by two other authors, that compress the statistics of the players in only one number – *EWA*, developed two years after the creation of PER, by **(John Hollinger, 2007, 2009)**; and *WP* by **(Berri, 1999)**. Stanek was able to compare the contributions inside the field with the salary of certain players and superstars to prove the existence of certain incongruities inside the league **(Stanek, 2016)**.

<u>Estimated Wins Added (EWA) by Hollinger</u>

Hollinger try to approximate the player´s ability to score points for the team, by converting the offensive statistics up to one number. The author started by the creation of *PER,* to understand the player’s per-minute quality of play, however this metric was unable to find the value of the player by the amount of time he was in the game. To solve this issue, he decided to count the number of minutes a player is on the court by using a metric named *VA*. Finally, to understand – in a more accurate way – the importance that each player has in his respective team, he improved his metric based on the creation of *EWA*. This way he was able to calculate the importance of the player according to wins contributed **(John Hollinger, 2009)**. Although it was quite innovative, this metric has a deficiency, because it mainly focuses on offensive statistics **(Stanek, 2016)**.

Wins Produced (WP) by Berri's

On the other hand, David Berri is a sports economist who looks at both offensive and defensive stats and has created the econometric model called *Wins Produced* which links the statistics of the players with the wins of the teams. This model is used to calculate the number of wins each player is able to add to his team, based on the marginal product of the player (Berri, 1999). To obtain the result, Berri had to create 6 important steps in his measure. First, it is necessary to calculate player's production and player's production per 48 minutes. Then, it is necessary to adjust the productivity of teammates defensive rebounds – because some players positions are more prone to get rebounds. The third step requires the adjustment of the productivity for assists – for the same reason as step two. The fourth step consists of integrating the team defense factor and calculate the adjusted value of *P48*. Before reaching the last step it is necessary to adjust the calculation to each player's position and finally calculate *WP48* as well as *WP* **(Stanek, 2016)**. Although *WP* is both offensive and defensive oriented, it is not possible to explain the totality of a player's value in a team, however it is a much more complete than Hollinger's metric.

2.3.2. Salary Cap consequences on team's performance and salaries

Salary Cap was implemented in the NBA in 1940, as previously mentioned in the introductory chapter, each team could not exceed a certain salary limit to spent on players **(Goldrosen, 2018)**. The main reason why there was a need to enforce a salary cap in the NBA, was the creation of a competitive balance sheet and a hinder player mobility within the league. However, with the progress of this restriction, it was possible to see that the competitive balance was not being reached. Since 1984-85, the salary cap has grown from 3.6 million to 109,140 million dollars in 2019-20 season **(Coon, 2020; Stanek, 2016)**.

Vrooman belief that the salary cap and cost sharing collusion of *NBA* lead to a less competitive balance of the league when compared to the other 2 major leagues (*NFL* and *MLB*). The author state that there is a high exploitation of the players and a decrease in the competitive balance within the league. It is fact that some owners are much more interested in maximizing profits than actually thinking about achieving a championship, most of the time they hire players who will bring more money than the money they cost at the time of purchase **(Vrooman, 2009)**. Borghesi has studied *NFL* salary cap and affirmed that teams that opt for a superstar-approach – retaining a small number of high-paid players, complementing the rest of the team with low-paid players – will not only create an environment where there is a "high degree of compensation inequality" but they also will not perform as well as expected, due to a dissatisfaction generated among players who are underpaid when compared to the two or three superstars on the team **(Borghesi,**

2008). In 2004, DeBrock, Hendricks & Koenker also realized that "baseball teams with more dispersion to their salary distribution performed more poorly" **(Debrock et al., 2004)**, thus, the higher the wage dispersion within a team, the more likely the team is to play below expectations. In order to affirm that there is a solution to these wage dispersion problems, Lazear – who studied the "Pay equality and industrial politics" and "psychology of organizations" – stated that, if there is a better balanced distribution of salaries among the workers, it is enough to have an increase in cooperation and effectiveness in the organization **(Lazear, 1989, 1991)**. If we make a parallel comparison with Lazear conclusion, as well as based on the other authors mentioned above, we can argue that, with a smaller dispersion of wages among the team players, the respective cooperation and efficiency on the field will increase.

To fundament this analysis with facts, in the season 1997-1998 the top ten players in the NBA earned 15 percent of the leagues payroll **(Berri & Jewell, 2004)**. Leeds and Kowalewski realized that "superstars and veterans were paid disproportionately at the expense of the rookies and marginal players" **(Leeds & Kowalewski, 2001)**. However, nothing has changed over the years, in addition to the constant increase of the salary cap, Stanek analyzed the wages of players, through his Marginal Product Revenue *(MRP)* metric, always based on the *EWA* and *WP* metrics of Hollinger and Berri, respectively. In the end he also came to the conclusion that, in 2016, superstars are overpaid, veteran players continue to receive on average higher salaries compared to younger players, and coaches tend to pay more for points scored, ignoring the rest of the statistics **(Stanek, 2016)**.

Given these types of facts it is easily understandable that teams were able to spend a higher amount of money on players. Therefore, it is possible to conclude the need of a more efficient salary balance, that should be based on the player's performances and their statistics. Most of managers spent a large percentage of the team´s salary cap on just one or two players. This approach may affect the game, since the rest of the money is spent on low quality players who only demand low salaries.

3. CONCEPTUAL MODEL PROPOSAL

3.1. CONTEXT

As previously referred in the introduction chapter, the conceptual model, based on the literature review it will be presented – under the *Design Science Research* Methodology – as temptation to compute, in a dashboard, all players and team performance issues that were mentioned through the project.

Firstly, it will be clarified which types of ***data*** and ***statistics*** are going to be used inside the dimensional model. *Secondly*, the most relevant ***formulas*** and ***metrics*** will be listed, to increase the quality of the information provided by the future dashboards. *After* this mathematical approach, an attempt to build a ***dimensional model*** – that fulfils all the Business Intelligence concepts and requirements – will be made. To *finalize* this concept*,* both player and coach ***dashboards*** will be designed and described, in a way that the final output will provide the right conclusions to offer a useful performance evaluation.

3.2. DATA AND STATISTICS

The ***data*** that will be used in this project is, in its entirety, taken from the official *NBA* website and therefore considered as a web source. Therefore, the usually extensive and complex *Extract Transform* and *Load* (*ETL*) application process will not be necessary from the outset, as the data is properly processed and available to be used, nevertheless any necessary changes will be made generally inside Excel and Power BI power query editor software.

The following tables shows all the *NBA* ***stats*** that will be used to build the respective dashboards. It should be noted that some of the statistics have been object of minor arithmetic changes – such as Per Game (*Assists Per Game*; *Points Per Game*; *Blocks Per Game; etc.*) and percentage (*% of Field Goal*; *% of Free Throw Made; etc.*) calculations. Therefore, these calculations are already observed as metrics that provide both qualitative and quantitative analysis. As a result, both types can be used to perform any type of analysis.

	PLAYER STATISTICS	FORMATS
PTS	*Points*	Totals **–** Per-Game
FGM	*Field Goals Made*	
FGA	*Field Goals Attempt*	
FGm	*Field Goals Missed*	
FG %	*Field Goals Percentage*	
3PM	*3-Point Made*	
3PA	*3-Point Attempt*	
3P %	*3-Point %*	
2PA	*2-Point Attempt*	
2PM	*2-Point Made*	
2Pm	*2-Point Missed*	
FTA	*Free Throws Attempted*	
FTM	*Free Throw Made*	
FTm	*Free Throw Missed*	
FT %	*Free Throw Percentage*	
REB	*Rebounds*	
OREB	*Offensive Rebounds*	
DREB	*Defensive Rebounds*	
AST	*Assists*	
TO	*Turnovers*	
STL	*Steals*	
BLK	*Blocks*	
PF	*Personal Fouls*	
P. Min	*Player Minutes Played*	

Table 1: Player Standard Statistics

	TEAM STATISTICS	FORMATS
T.Min	*Team Minutes Played*	Totals **–** Per-Game **–** %
T.FGM	*Team Field Goals Made*	
T.AST	*Team Assists*	
T.REB	*Team Rebounds*	
T.DREB	*Team Defensive Rebounds*	
T.TO	*Team Turnovers*	
T.BLK	*Team Blocks*	
T.PC	*Team Pace*	

Table 2: Team Standard Statistics

	OPPONENTS STATISTICS	FORMATS
Opp. PTS	*Opponent Points*	
Opp. FGM	*Opponent Field Goal Made*	
Opp. 3PM	*Opponent 3-Point Made*	
Opp. 2PM	*Opponent 2-Point Made*	Totals **–** Per-Game **–** %
Opp. FTM	*Opponent Free Throws Made*	
Opp. TO	*Opponents Turnovers*	

Table 3: Opponents Standard Statistics

	LEAGUE STATISTICS	FORMATS
L.PTS	*League Points*	
L.FGA	*League Field Goals Attempted*	
L.FGM	*League Field Goals Made*	
L.FTA	*League Free Throws Attempted*	
L.FTM	*League Free Throws Made*	
L.AST	*League Assists*	Totals **–** Per-Game **–** %
L.REB	*League Rebounds*	
L.OREB	*League Offensive Rebound*	
L.TO	*League Turnovers*	
L.PC	*League Pace*	
L.PF	*League Personal Fouls*	

Table 4: League Standard Statistics

From the previous tables, it is notable all the information – abbreviations, respective definition, and respective formats – that will be used in the project. There are stats from players, teams, and opponents' statistics, as well as some fundamental information about the league. It may be important to understand the usefulness of the data, that is available in several formats, however, in the research carried out, for the project development, it was concluded that only three formats will be necessary – ***total***, ***per-game*** and ***percentage*** values. The existence of data in three formats is justified by two reasons, *firstly* it allows a more diversified analysis – providing a deeper overview of the players and teams progress – and *secondly*, is mainly justified by the fact that the, *EWA* and *WP*, formulas and auxiliary calculations will demand several intermediate steps in which are included the use of averages, percentages, and total values. In addition to the web-source, any extra data that will need to be added inside the data warehouse, will be done throw the use of excel by uploading the respective files.

3.3. METRICS AND FORMULAS

3.3.1. John Hollinger Estimated Wins Added (*EWA*) metric

To achieve the expected number of wins added by a player, through the *EWA*, it will be necessary to perform set of primarily calculations. Among these additional calculations, it can be found formulas such as *PER – Unadjusted PER (uPER)*; *Adjusted PER (aPER)* which contains *Pace* and *League Adjustments* –, *VA* – that contains the *Position Replacement Player* (*PRL*) –, and finally *EWA* – which takes *VA* and divides by 30 **(Sports Reference, 2020b)**.

1) $\boldsymbol{uPER} * MP = 3PM + \frac{2}{3}AST + \left(2 - \boldsymbol{f} * \frac{T.AST}{T.FGM}\right) * FGM + \frac{1}{2}FTM\left(2 - \frac{1}{3} * \frac{T.AST}{T.FGM}\right) - \boldsymbol{v}TO - \boldsymbol{vd}(FGA - FGM) - 0.44\boldsymbol{v}(0.44 + 0.56\boldsymbol{d}) * (FTA - FTM) + \boldsymbol{v}(1 - \boldsymbol{d})(REB - OREB) + \boldsymbol{vd}OREB + \boldsymbol{v}STL + \boldsymbol{vd}BLK - PF * \left(\frac{L.FTM}{L.PF} - 0.44 * \frac{L.FTA}{L.PF}\boldsymbol{v}\right)$

$$\boldsymbol{f} = \frac{2}{3} - \left(\frac{1}{4} * \frac{L.AST}{L.FGM} * \frac{L.FTM}{L.FGM}\right)$$

$$\boldsymbol{v} = \frac{L.PTS}{(L.FGA - L.OREB + L.TO + 0.44 * L.FTA)}$$

$$\boldsymbol{d} = \frac{L.REB - L.OREB}{L.REB}$$

As it can be seen above, there is a representation of the Unadjusted *PER* (***uPER***), already with some statistics applied. The calculations of the formulas ***f***, ***v*** and ***d*** are defined in the following box. The *Factor* (***f***) – represents the relation between assists and free throws of the league, with the field goals made; *Value of Possession* (***v***) – divide the league points (numerator) by the calculation that estimates the possessions and represent the average number of points per possession; and finally, *Defensive Rebound Percentage* (***d***) – calculates the number of defensive rebounds and then divides the latter by the number of total rebounds **(Justin Jacobs, 2017)**.

$$\boldsymbol{aPER} = uPER * PaceAdjustment$$

$$\boldsymbol{PaceAdjustment} = \frac{L.PC}{T.PC}$$

To calculate the adjusted *PER*, it is required the use of pace factor – which is a statistic that estimates the number of possessions per 48 minutes by each team. The respective adjustment tends to normalize players in terms of their team pace, where a value of one reflects a team playing at the leave average.

$$\boldsymbol{PER} = aPER * \left(\frac{15}{League(aPER)}\right)$$

$$\boldsymbol{League(aPER)} = \sum_{i=1}^{P} \frac{MPi}{MIN} * aPERi$$

To finalize the first part of *PER* calculation, it is necessary to standardize *aPER* by proceed to a final adjustment – *League(aPER)* – that allows the comparison across the years. This step uses the players minutes played as the weights. Finally, John

Hollinger defines 15, as average value for *PER*, in order to reach the real *PER,* the author divides 15 for the *league(aPER)* **(John Hollinger, 2007; Sports Reference, 2020b**).

<u>Linear Weights ***PER*** by Zach Fein</u>

As mentioned above, *PER* from John Hollinger, represents the omnipresent *NBA* metric which provides a single number that measures the productivity of players per minute **(Justin Willard, 2018)**. However, this formula is known for its level of complexity and detail, and it was based on this assumption that Zach Fein – a writer from Bleacher Report – wrote an article in which he studied the formula for himself. The idea was to develop a faster and easier way to calculate the John Hollinger formula, but with the aim of maintaining its efficiency. To achieve his target, the author uses only linear weights, for that purpose, his "*Player*" were the league player's average stats in 82 games; the "*Team*" stats were the average of all 30 teams in 2008 season, and finally; the "*League*" were the league stats in 2008. To find the linear weights of the stats, the author multiplies the player's total stats by different weights and add them up – but not to the league or team – to study the effect of that addition on *PER* on the next step. Since the goal is to simplify the formula, no pace adjustments are made, due to its complexity in terms of calculations. In addition, the author decides that there was no reason to pass through the value 15 – which corresponds to the league average *PER*. Fortunately, the league average has not changed significantly in the past few years, each league adjustment has been around 54. Having conclude that, it is possible to achieve the final weights by multiply the initially weights – founded for each statistic – by the value 54 **(Zach Fein, 2009)**. Therefore, the simplified *player efficiency rating* will have the following format:

$$\text{L. Weights } \boldsymbol{PER} = \left[\begin{pmatrix} FGM * 85{,}910 \\ +STL * 53{,}897 \\ + * 3PM * 51{,}757 \\ +FTM * 46{,}845 \\ +BLK * 39{,}190 \\ +OREB * 39{,}190 \\ +AST * 34{,}677 \\ +DREB * 14{,}707 \\ -PF * 17{,}174 \\ -FTm * 20{,}091 \\ -FGm * 39{,}190 \\ -TO * 53{,}897 \end{pmatrix} * \left(\frac{1}{P.Min}\right) \right]$$

It should be mentioned that after Zach compiled the player's "linear weights *PER"* of that year, he concluded that the ten players with the highest "linear weights *PER"* of his analysis were practically the same as the top ten players that John Hollinger's has in his regular *PER*, apart from one player who was slightly overtaken by another. The League Average of this "linear" approach is 14.98 – close to the original *PER* value of 15 – and the players with the highest difference in this topic are respectively those whose teams have the lowest or second lowest paces in the league **(Zach Fein, 2009)**.

2) $\boldsymbol{VA} = \frac{Minutes\ Played*(PER-PRL)}{67}$

The second step – after *PER* determination – it is based on the *VA* calculation, which represents the difference between the performance offered by the player and the talent that his position already offers by itself – which is measured through the *PRL*[2] value –, multiplied by the number of minutes played, to find a more accurate value. *VA* determines how many points a player adds to his team at the end of the season. However, it is essential to adjust the calculation according to the player's position, otherwise the result would be flawed. According to John Hollinger, this metric is an important value when it comes to award voting. It also allows the comparison between players with different production and minutes, and finds out which performance was more productive **(John Hollinger, 2009)**.

Position Replacement Level (PRL)

- *Power Forward* = 11.5
- *Point Guard* = 11
- *Center* = 10.6
- *Shooting Guard* = 10.5
- *Small Guard* = 10.5

3) $\boldsymbol{EWA} = \frac{VA}{30}$

Once the *VA* – that represents the number of additional points a player contributes to team's season – is determined, the third and final step is quite simple. The *EWA* value is quite similar with the previous one, the difference is just the fact that *EWA* is specifically oriented for wins. It is known as the estimated number of additional wins a player contributes to its team. Therefore, to obtain the *EWA* value, it is just necessary to divide *VA* by the value of 30 – that corresponds to the estimated number of points, over the course of 82 season games, that enable to add another win.

[2] Player Position Replacement Level (*PRL*) is a popular term – inside the analytics sports industry – for the level of talent available to be picked up off the less valuable players **(John Hollinger, 2009)**. The value was obtained according to a study made with players that had less than 500 minutes played in each season and revealed a pattern that allowed to obtain the respective values above.

3.3.2. David Berri Wins Produced (*WP*) metric

David J. Berri, University professor of Economics at Utah University, dedicated the last two decades of his career to improve his specialization in competitive balance and evaluations of players and coaches in sports **(Utah, 2020)**. To contextualize the topic, David Berri is the author of *WP* metric – as mentioned in the literature review. However, before moving to the formula's clarification, it is pertinent to explain that, to model *NBA* Wins, the author started from a simple model – where it connects all the statistics tracked for individuals to team wins –; followed by a complex regression model – where it builds equations that lead him to the marginal values –; and finally reached the final *WP* formula. All the data used in the regression model, and the respective formula construction, were based on data from the 1991/92 to the 2005/06 season.

In a first stage, David. J. Berri builds the model based on the previous research – that was conducted by John Hollinger and Dean Oliver – and he concluded that these studies contain some limitations. On one hand, it is not specifically clear that wins are only defined by the relative ability of teams to score points through possessions. On the other hand, possession does not include important variables such as defensive rebounds. Facing the complex stage of his work, Berri developed a regression model to determine the marginal values for players and team statistics. In the model construction, the author addresses several aspects, such as, the correlation between points scored and conceded in wins; how field goal attempts derive from possession – which includes all Hollinger's variables and rebounds –; why the number of rebounds is also a consequence of a change in possession; and finally, how wins are associated with points scored per possession **(Berri et al., 2008)**.

The following Table 5, contains the respective marginal values from Berri's regression model and allows to understand the importance that each statistic have on the court. Similar with Hollinger's *EWA*, these values are going to be a part of the final formula construction – which has several steps that need to be followed, to obtain the *WP48*[3].

[3] *WP48* is the abbreviation for wins produced per 48 minutes, the 48 minutes represent the standard time of an NBA game. Since some basketball games could have one or more overtimes, the calculation is done per 48 minutes and not per game.

	PLAYER VARIABLES	MARGINAL VALUES
3PM	*3- Point Field Goal Made*	0,065
2PM	*2- Point Field Goal Made*	0,032
FTM	*Free Throw Made*	0.018
FGm	*Missed Field Goal*	(0,034)
FTm	*Missed Free Throw*	(0,015)
OREB	*Offensive Rebound*	0,034
DREB	*Defensive Rebound*	0,034
TO	*Turnover*	(0,034)
STL	*STL*	0,034
Opp.FTM	*Opponent Free Throw Made*	(0,018)
BLK	*Blocked Shots*	0,019
AST	*Assists*	0,022
	Opponent & Team Variables	
Opp.3PM	*Opponent 3-Point Field Goal Made*	(0,065)
Opp.2PM	*Opponent 2-Point Field Goal Made*	(0,032)
Opp.TO	*Opponent Turnovers*	0,034
T.REB	*Team Rebounds*	0,034

Table 5: *WP* Marginal Values of Statistics

1) $\boldsymbol{PROD} = 3FGM * 0{,}065 + 2FGM * 0{,}032 + FTM * 0{,}018 + FGm * (-0{,}034) + FTm * (-0{,}015) + OREB * 0{,}034 + DREB * 0{,}034 + TO * (-0{,}034) + STL * 0{,}034 + Opp.FTM * (0{,}018) + BLK * 0{,}019 + AST * 0{,}022$

The first step in D. Berri's formula, simply involves the calculation of *PROD* by multiplying the accumulation of each player statistic by the respective marginal value in table 5. However, both *BLK's* and *AST's* are stats that do not have a direct impact on wins, but they do have an impact on teammates' productivity. Therefore, it is necessary to consider the accumulation of these two factors by all teammates and create some adjustments.

$$\boldsymbol{MATE\ 48} = P48\ Team\ AVG\ AST\ \&\ BLK - P48\ League\ AVG\ AST\ \&\ BLK$$

$$P48\ \boldsymbol{Team\ AVG\ AST\ \&\ BLK} = \frac{Team\ AST * 0.022 + Team\ BLK * 0.019}{Accumulation\ of\ Team\ Minutes} * 48$$

$$P48\ \boldsymbol{League\ AVG\ AST\ \&\ BLK} = \frac{\sum[(Team\ AST * 0.022 + Team\ BLK * 0.019)]/30}{\sum Accumulation\ of\ Team\ Minutes\ /30} * 48$$

For this first intermediate calculation, it is necessary to obtain *MATE48* – which represents the teammates' assists and blocks shots adjustment. This step is calculated by the subtraction between *Per 48 Team AVG Accumulation of AST & BLK* – which is done by multiplying the team's accumulated assists and blocks by the respective marginal value and then divide the result by the sum of players minutes accumulation – and *Per 48 League AVG Accumulation of AST & BLK* – which follows the same approach but applied to the league.

2) $\boldsymbol{P48} = \left[\frac{PROD}{Minutes\ Played\ by\ Player}\right] * 48 - MATE48$

From the previous calculation, it becomes quite simple to estimate the production of a player per 48 minutes, by just multiplying the ratio between *PROD* and each player's minutes, by 48, and then subtract the *MATE 48* value.

3) ***POS 48*** (*Position Adjustment)*

Average Productivity at Each Position (POS 48)

- *Center* = 0.296
- *Power Forward* = 0.256
- *Small Forward* = 0.186
- *Shooting Guard* = 0.158
- *Guard* = 0.191

Although the previous calculation could provide an approximate idea of the reality, nevertheless it is important to notice that basketball positions are not the same, and therefore it is also essential to create this adjustment. Over the years, basketball has been changing from an 'inside' game – where Centers and Power Forwards had a lot of influence on the game – to a more 'outside' game – with more individual and three-point plays, where players as Point Guards and Shooting Guards tend to have more relevance. Although D. Berri is one of the "*The Business of Sports*" book contributors – which provides an explanation of how the author developed the formula – the values used were based on a model with data from 1992/1992 to 2005/06 **(Berri et al., 2008)**, which coincides with the time when the game was 'dominated' by inside game players. To create the most updated information possible, the project will use values from the latest regression made by "*The Wages of Wins Journal*"[4] – which result from a model estimated with data from 1987/88 to 2010/2011 **(Berri & Schmidt, 2012)**.

[4] The Wages of Wins Journal, corresponds to an online Website Journal which has as its principal source "The Wages Of Wins – Taking Measure of the Myths in Modern Sport" book, released in 2006 **(Berri et al., 2006)**.

4) $\boldsymbol{TM48} = P48\ Value\ of\ Team\ Factors - P48\ League\ AVG\ Value\ of\ Team\ Factors$

To finally achieve *WP*, first it is important to apply one last adjustment that is specifically focused on team factors – player's team, and opponent's teams. This adjustment uses values from the final part of *table 5* and intends to measure the defensive impact that the team has on the player's productivity. For that, *TM48* is calculated by subtracting the *league average team adjustment* from each *team's adjustment*.

$$\boldsymbol{P48\ Value\ of\ Team\ Factors}$$
$$= \frac{[Opp.3PTM * (-0.065) + Opp.2PTM * (-0.032) + Opp.TO * 0.034 + T.REB * 0.034]}{Accumulation\ of\ Team\ Minutes} * 48$$

$$\boldsymbol{P48\ League\ AVG\ Value\ of\ Team\ Factors} = \frac{\sum P48\ Value\ Of\ Team\ Factors}{30}$$

The adjustment has some intermediate calculations, where it accumulates some team and respective opponent stats. Then, it allocates the numbers across the team members – by dividing the sum by the *accumulation of Team Minutes* according to each player's minutes – following the scale already used in the previous calculations by multiplying the result per *48 minutes*. Finally, it is required to make the same calculation for each team, obtaining the average value of the league factors.

5) $\boldsymbol{WP48} = P48 - POS\ 48 + TM48 + 0.1$

Once the previous adjustments have been calculated, it is possible to determine *WP*. This final calculation is made by simply subtracting the *(3) Position Adjustment (**POS 48**);* from the *(2) player's productivity (**P48**)*; adding the impact that *(4) team factors (**TM 48**)* have on the player; and finally sum the value of 0.1[5]. Again, and since all the previous calculations are made for 48 minutes, this final step will also be analyzed on the same scale – to achieve the closest number of wins produced possible. Having achieved the ***WP48*** result, and, to obtain the number of wins produced for the specific player, we just need to calculate ***WP*** per minute – dividing *WP48* by 48 minutes – and multiply this last value by the player's number of minutes played.

$$Player's\ \boldsymbol{WP} = \frac{\boldsymbol{WP\ 48}}{48} * Player's\ minutes\ played$$

[5] Since each team will win, on average, 0.500 games per 48 minutes. The 0,1 value represents the number of wins that every player will produce, on average, per 48 minutes played **(Berri et al., 2008)**.

3.3.3. Player's Monetary Value metric

This chapter will be focused on a complementary measurement that enable the extraction of some conclusions concerning players' performance. Based on Brett Knight' approach – that make use of *EWA* or *WP* from Hollinger and Berri – the goal is to create a metric that shows which would be the player expected salary. This metric will be based on the Basketball Related Income (*BRI*)[6] amount that is collectively attributed to players – usually ranging between 49-51%, since the 2011 lockout **(Coon, 2020)** – and the amount of wins that players would add to their teams. Assuming this percentage, the *BRI* that is given to the players should be multiplied by 0.5, to finally, obtain the expected salary, by multiplying the added player wins by the monetary value of each win. It is worth to mention that players are paid to contribute to the team's success, by winning games **(Brett Knight, 2019)**, without this assumption, the following formula should not be applied.

1) $$\boldsymbol{Monetary\ Value\ of\ Win\ (MVW)} = \frac{BasketbalL\ Related\ Income\ (BRI) * 0.5}{Regular\ Season\ Games}$$

2) $$\boxed{\begin{array}{c} \boldsymbol{Player\ Expected\ Salary} = Players's\ WP\ * MVW \\ \text{or} \\ \boldsymbol{Player\ Expected\ Salary} = Players's\ EWA\ * MVW \end{array}}$$

To calculate this previous steps, it is required to load additional data, such as, the last ten years of the *NBA* BRI, which will be obtained using **(Statista, 2020)**; players' salaries – from **(ESPN, 2021)**; and the number of games played per season – through **(NBA & SAP, 2020)**. This metric will establish a line that defines whether the player is being *overpaid* or *underpaid*, without considering secondary factors, such as, players' age, reputation, status, or any other variable that would, indirectly, affect their salary. It will provide an analysis, for both, managers, and team owners, who will use this information to judge if the player is being an asset or not. The formula satisfies the literature review, by dealing with the existence of young players – with high performances – who are underpaid, in comparison with veteran; or superstars' players that are paid above to their qualities – who perform below their prime but still has a high market value – due to their status or reputation that was created on a recent past.

[6] Basketball Related Income (BRI), represents the overall operations revenues by NBA league, that will be used to define the players salary cap. This related income, includes several items, such as, broadcast rights, advertising, merchandising, sponsorships, etc. BRI is also used to define player's salary cap **(Trevir Nath, 2020)**.

3.4. Conceptual & Dimensional Model

3.4.1. Conceptual Model Process

Following the project structure and supporting the previous literature review, the conceptual model is presented in this section through a small diagram that briefly represents the process that will lead to the Business Intelligence solution. In a first stage, it is necessary to identify the performance indicators that will be used for players' analysis. After that, it will be verified the data availability – which in this case, all data has public access, and therefore, does not require any user permission. The next stage is composed by a simple *ETL* process, which will lead to the Data Warehouse construction and further dimensional model.

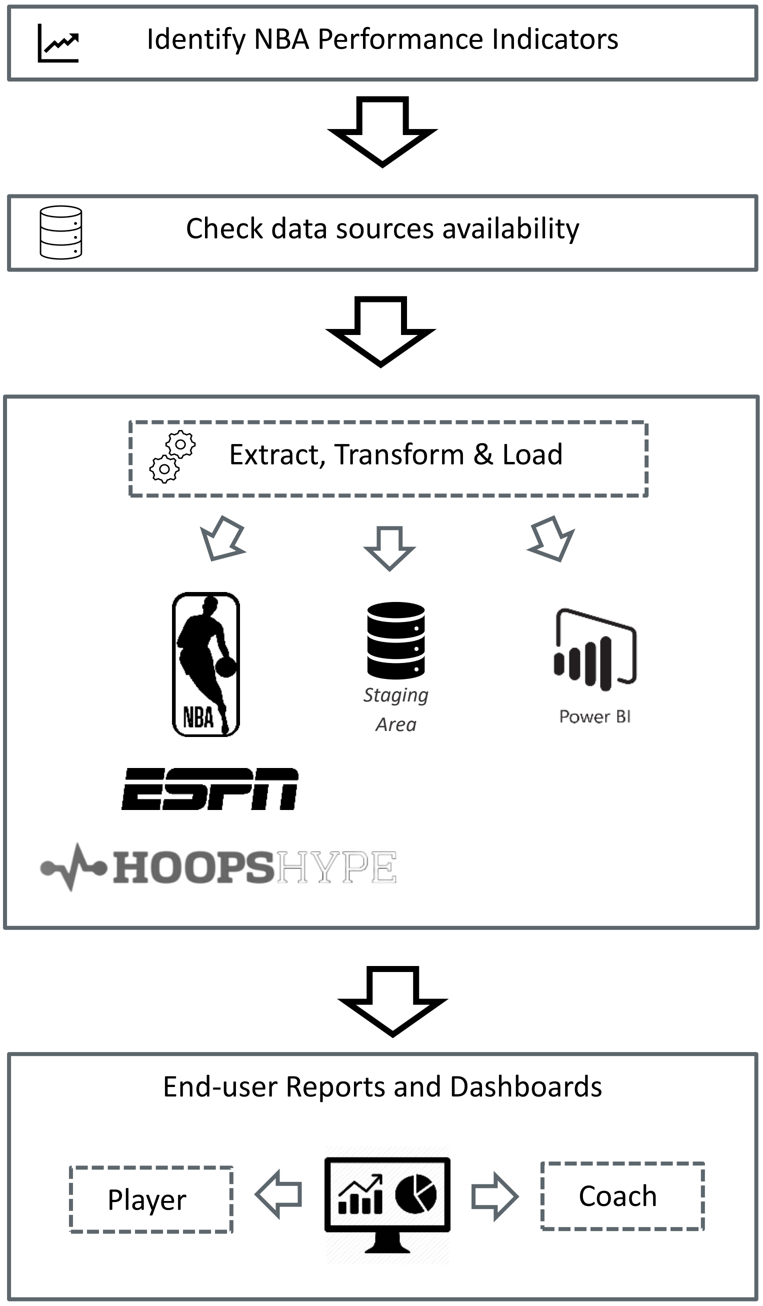


Figure 2: Conceptual Model

All data content were collected from *NBA* official website – for player, team, and league statistics – and from *ESPN* & *HoopsHype* official websites – for salaries. Then, the data will be both, added and transformed inside a staging area, in a way that each dimensional table can be obtained and loaded into the Power BI software. The remaining transformations are going to be applied inside the Power BI query Editor. Once the data is properly prepared – with metrics, measures, and formulas – the final step is the dashboards construction, as represented at the end of *Figure 2*.

3.4.2. Dimensional Model

Contrarily to *OLTP* systems – which are traditionally called as relational models for data capture – this project follows an *OLAP* system – usually known as dimensional model which enable to improve decision making – that is composed by dimensions, measures, hierarchies, cubes, or data warehouses. *Figure 3,* represents a star schema design, that was adopted for the analytical process of this dimensional model. This type of schema is not only defined by its denormalization and redundancy but also by several benefits, such as, less complexity; easy to understand; answer the needs for future changes; and improve the performance by lowering the time of queries. The following data warehouse is composed by three dimensions – *Date_Dimension; Player_Dimension; Team_Dimension* – and one Fact Table – *Fact_Metrics*.

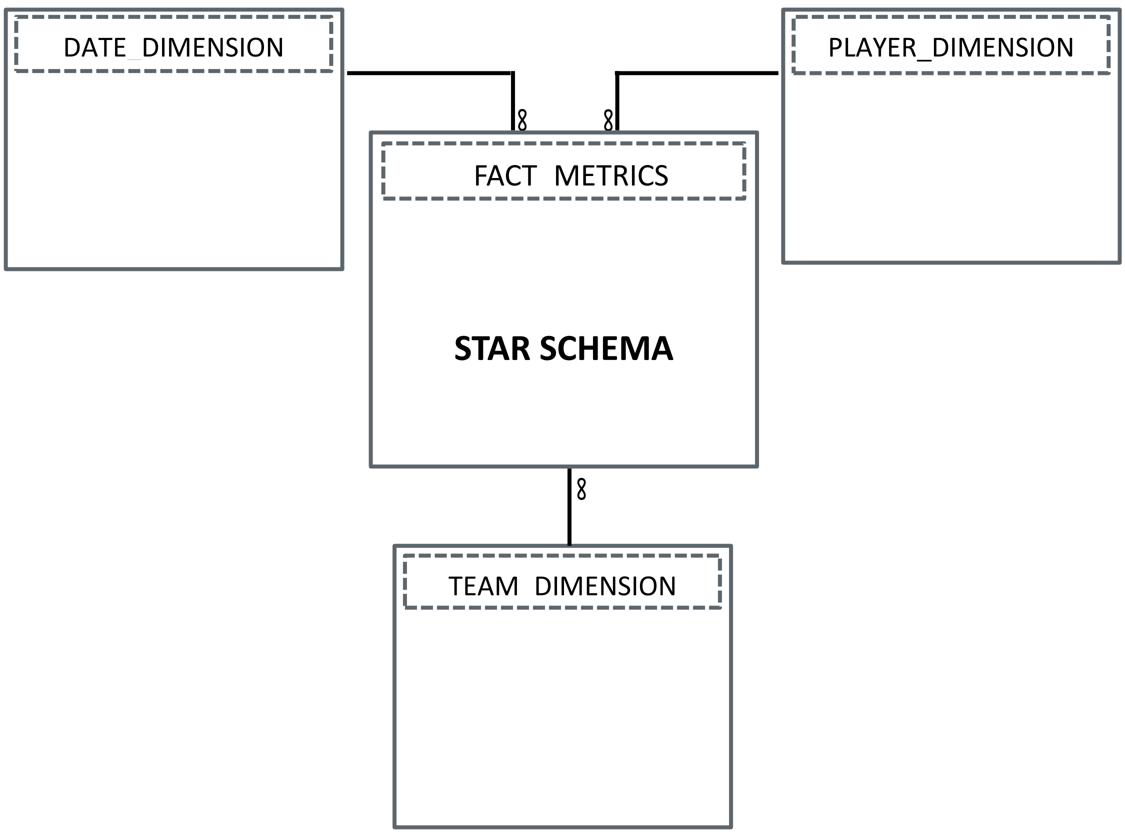


Figure 3: Dimensional Model - Star Schema

3.5. PLAYER & COACH DASHBOARDS DESIGN

3.5.1. Dashboards Interaction

The following figure presents the workflow between dashboards. The diagram starts with a summary overview that offer an overall information, for both player and coach and provides the possibility to navigate between the remaining dashboards through connective buttons. The main functionality of the first prototype page consists in the creation of context. The main goal of this workflow is to provide one type of information for the player, and to create a coach dashboard, named *"Coach-Individual Player analysis"*, that will contain some privileged information for the coach – such as expected player salary and difference between real player salary and expected player salary.

The *"Player – League Overview"* offers an overall analysis based on the five best *NBA* players, both offensive and defensive. In addition to this, the *"Player - Individual Analysis"* dashboard will be created, to analyze each player's performance, how many wins he adds or remove from the team – through Berri and Hollinger metrics – as well as other type of performance visualizations.

On the other perspective, it will be also presented the coach dashboards design, such as, *"Coach - Team Overview"* and *"Coach – Individual Player Analysis"*. The first one, is a *"Player – League Overview"* replica, in terms of design, but focused only on the best five players of each team, instead of focusing on the best league players. Finally, the second coach dashboard, is similar to the previous "*Player's Individual Analysis*", however, it contains some additional measures that allows the evaluation, whether if the player is under or overpaid, depending on his performance.

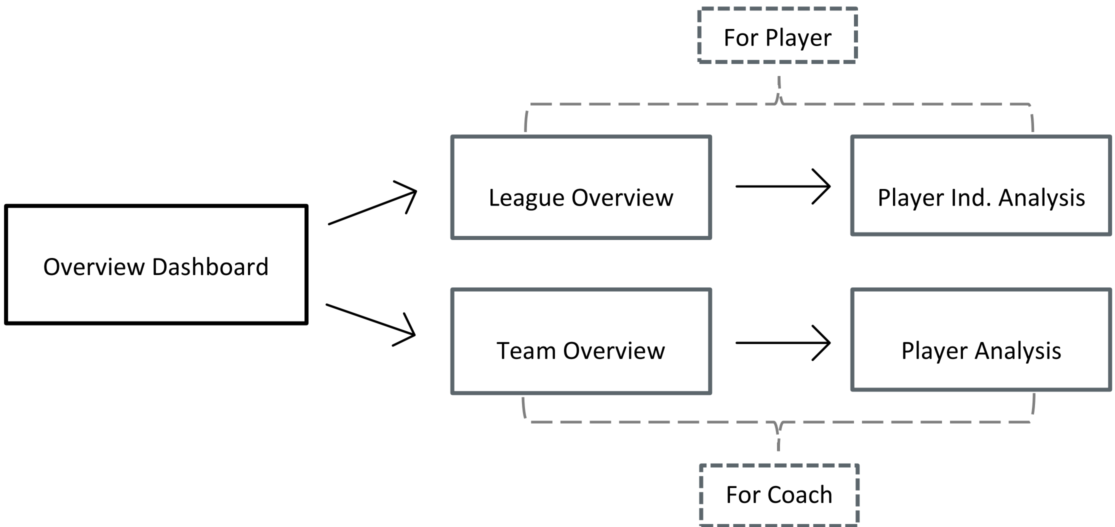


Figure 4: Dashboards Sequence

3.5.2. Dashboard's mockups

3.5.2.1. Dashboard – Initial page overview

As an initial component of the prototype, the project will contain five dashboards, with the first being just an introduction to the analysis. The *Overview Dashboard* function is to display some of the most relevant stats about players and teams in the league, alongside with a small bar chart (7) that shows the league basketball related income evolution across the last ten years – note that this stat work as an important piece to evaluate the presence of overpaid or underpaid players.

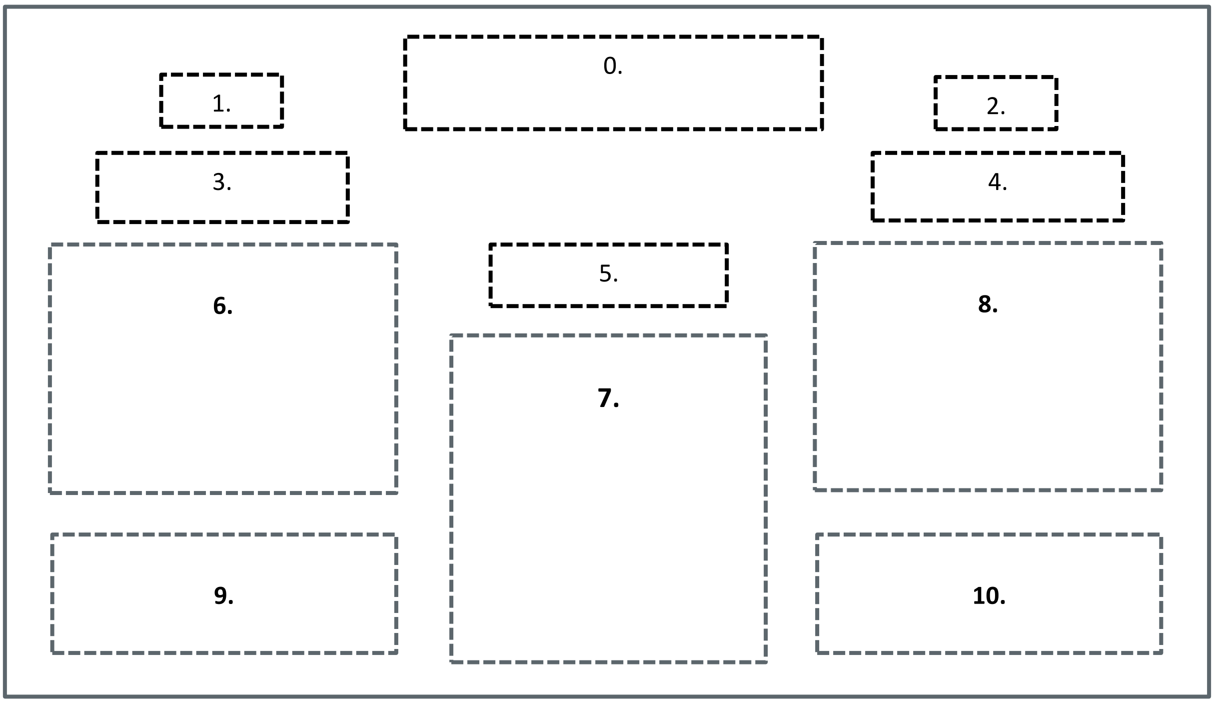


Figure 5: Mockup - Overview Dashboard

0. Slicer: To navigate across the Seasons;

1. Slicer: To select across the League Divisions;

2. Slicer: To select across the League Conferences;

3. Button: To navigate directly to Player's Overview Dashboard;

4. Button: To navigate directly to Coach Overview Dashboard;

5. Text Box: NBA Revenue title;

6. Double Funnel: Showing the TOP 10 player's *EWA* vs *WP*;

7. Stacked Area Chart: NBA Revenue by Season;

8. Scatter Chart: Team's Offensive efficiency vs Defensive Efficiency;

9. Double Funnel: Shows the TOP 3 winning teams with their respective number of losses;

10. Double Funnel: Shows the TOP 3 players with the best Offensive efficiency with their respective Defensive efficiency;

3.5.2.2. Coach Dashboard – Team overview

After clicking on the coach's button from the initial page overview, it will navigate into a dashboard where it is possible to find a summary analysis, based on the five best team players – the number, is justified by the fact that each basketball team plays with five players. This dashboard will illustrate the top five best players in the league by the main statistics, in each conference or in each season[7]. Additionally, it will be also possible to find the top five players with the highest salaries, with the most wins added via *EWA* and finally the most wins produced via *WP* formulas.

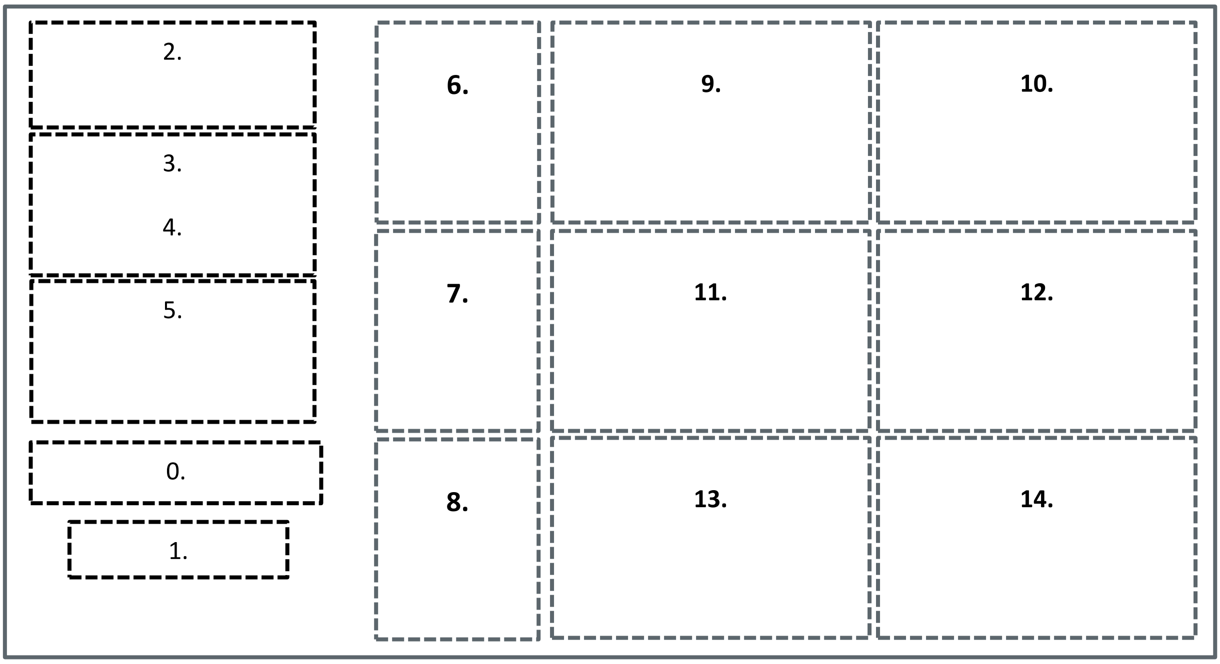


Figure 6: Mockup - Team Overview (Coach Dashboard)

0. Text box: Dashboard Title;

1. Button: To navigate directly to Player's Individual Analysis Dashboard;

2. Slicer: To select team name;

3 & 4. Simple Image: Picture of team and respective conference;

5. Slicer: To select season year;

6. Multi-row card: TOP 5 Player's salaries list;

7 & 8. Multi-row card: TOP 5 Player's *EWA* list and TOP 5 Player's *WP* list, respectively;

9 &10. Funnel: Presents TOP 5 FG and FT player's (by percentage);

11. Clustered column chart: TOP 5 Defensive rebound players with their offensive rebounds;

12. Clustered column chart: Presents TOP 5 three-point players (by percentage);

13. Clustered bar chart: Presents TOP 5 player's double-double (two statistics with double digits);

14. Clustered bar chart: Presents TOP 5 player's triple-double (three statistics with double digits);

[7] To avoid the presentation of similar mockups, the player mockup named "League Overview", instead of the numbers 2,3,4,5 only has number 2 and a 3, both representing two slicers that allows the analysis by season and conference only.

3.5.2.3. Coach Dashboard – Player Individual Analysis

To provide a deeper analysis of each player, the end user (coach), can simply navigate to this "*Player Individual Analysis"* dashboard, by clicking on the button – marked as number one on the previous "*Team Overview*". This dashboard displays more detail of every stat presented previously, but only for one player, and it enables the comparison between them and the league average. It will also show the actual and expected salary[8], to conclude if the player is being overpaid or underpaid.

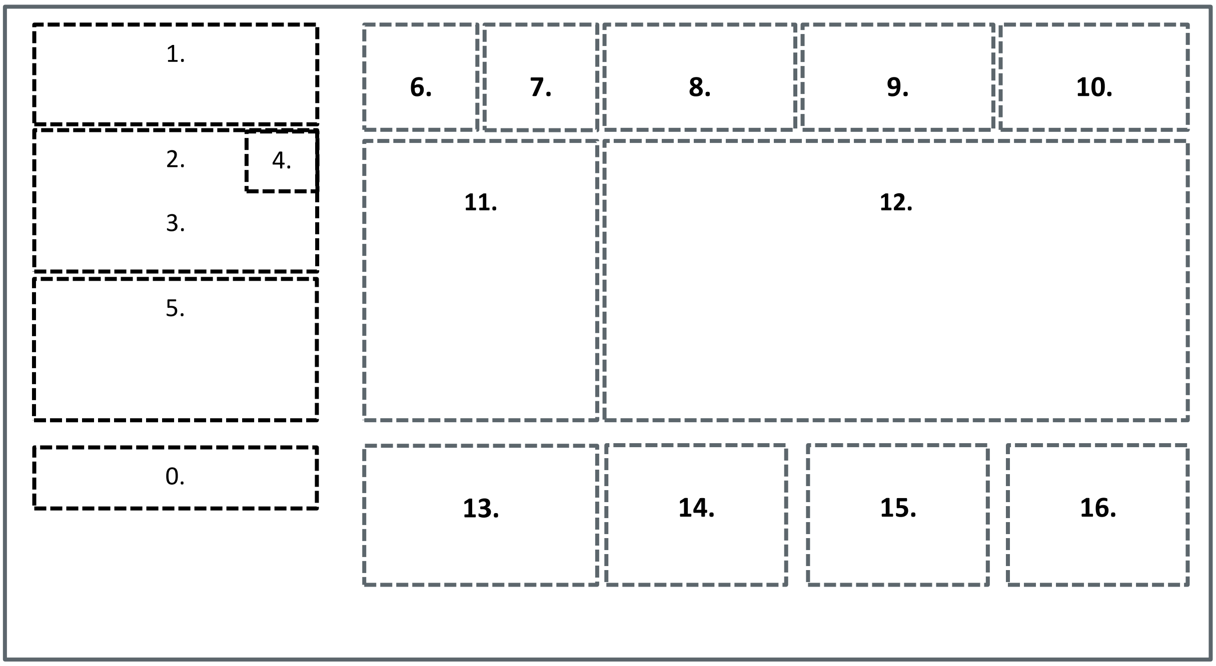


Figure 7: Mockup - Player Individual Analysis (Coach Dashboard)

1. Slicer: To select player name;

2 & 3. Simple Image: Pictures of both team and player (using URL);

4. Card: Player position;

5. Slicer: To select season year;

6 & 7. Card: Player *Estimated Wins Added* (*EWA*) and Player *Wins Produced* (*WP*), respectively;

8 & 9. Card: Player real salary and Player expected salary, respectively;

10. Card: Difference between player real salary and expected salary

11. Area chart: Comparison between points scored vs league AVG;

12. Line and Clustered bar chart: Comparison between player's rebound and assists vs league AVG;

13,14,15 &16. Gauge: Shows if players FG%, 3P%, 2P% and FT% are above or below league AVG;

[8] As previously explained in the note 7, there is also another player mockup, named "Individual Analysis". The latter instead of number 9 and 10, will have a scatter chart representing the player offensive and defensive efficiency, across all seasons.

4. CONCEPTUAL MODEL APPLICATION

In this chapter, the most important steps of the conceptual model development process will be presented. The first section, contains all the necessary DAX codes that were required to develop both measures and calculated columns. Then, it will be introduced the final dimensional model design – ***Figure 15*** – along with its composition in terms of dimensions and fact table. The ***Figure 16***, shows the fact table expanded, in a way that it is possible to see all the columns, calculated columns, and measures. Finally, the last dashboard prototype will be displayed, as well as the explanation of its structure, to conclude if the initial goal was reached successfully.

4.1. DAX Code Adjustments

4.1.1. Additional Calculated Columns

In a way to obtain the right information, as well as the proper *Hollinger's* and *Berri's* formulas, this section shows all the additional data that had to be added to the dimensional model. The following figures provide a list of each calculated columns. These columns are computed only once and automatically generated in the fact table of the *PBI* file, requiring more space memory when compared to measures.

Figure 8, contain a list with the average of some important stats – *Points*, *Assists* and *Rebounds* – which are going to represent a target value for each year, to evaluate if the player or team is below or above the league average. ***Figure 9*** contains several calculations which, in a first stage, were not inside the data warehouse, however they represent an important part of data that will be applied on both *Hollinger* and *Berri's* initial formulas – such as *PER* and *PROD – and* several graphical dashboard visualizations.

```
PTS AVG P/Year = AVERAGEX (FILTER ('Fact Metrics','Fact Metrics'[FK_Date] = EARLIER
('Fact Metrics'[FK_Date])),'Fact Metrics'[PTS])
REB AVG P/Year = AVERAGEX (FILTER ('Fact Metrics','Fact Metrics'[FK_Date] = EARLIER
('Fact Metrics'[FK_Date])),'Fact Metrics'[REB])
AST AVG P/Year = AVERAGEX (FILTER ('Fact Metrics','Fact Metrics'[FK_Date] = EARLIER
('Fact Metrics'[FK_Date])),'Fact Metrics'[AST])
```

Figure 8: League average statistics for *PTS*, *REB*, *AST*

```
2PM = 'Fact Metrics'[FGM] - 'Fact Metrics'[3PM]
2PM = 'Fact Metrics'[FGA] - 'Fact Metrics'[3PA]
FG missed = 'Fact Metrics'[FGA]-'Fact Metrics'[FGM]
FT missed = 'Fact Metrics'[FTA] - 'Fact Metrics'[FTM]
opp.FTM (P/Foul) = ('Fact Metrics'[PF] / 'Fact Metrics'[PF(Team Acu.)]) * 'Fact
Metrics'[Opp.FTM]

TOTAL OREB P/Player = CALCULATE( SUM('Fact Metrics'[OREB]) , FILTER('Fact Metrics','Fact
Metrics'[FK_Player] = EARLIER ('Fact Metrics'[FK_Player])))
TOTAL DREB P/Player = CALCULATE( SUM('Fact Metrics'[DREB]) , FILTER('Fact Metrics','Fact
Metrics'[FK_Player] = EARLIER ('Fact Metrics'[FK_Player])))
TOTAL GP = CALCULATE( SUM('Fact Metrics'[GP]) , FILTER('Fact Metrics','Fact
Metrics'[FK_Player] = EARLIER ('Fact Metrics'[FK_Player])))
```

Figure 9: Additional player statistics

The following two figures are mainly composed by data that was required from *Berri's* formula. ***Figure 10***, is composed by the accumulated values of personal fouls *(PF)*, minutes *(MIN)*, assists *(AST)*, blocks *(BLK)* and rebounds *(REB)* both per team and year, that are directly applied in auxiliary calculations of *Berri's* formula. ***Figure 11,*** also includes auxiliary calculated columns for *Berri's* formula, but only with data that are required for future adjustments, such as *MATE48* and *TM48* – that are explained in points *1)* and *5)* of the *3.3.2*. chapter. It is important to mention the relevance of this additional content – that was not available on the original data sources – and corresponds to a crucial part of the business intelligence approach that increase the technical performance interpretation.

```
PF (Team Acu.) = SUMX (FILTER ('Fact Metrics', AND ('Fact Metrics'[FK_Team] = EARLIER
('Fact Metrics'[FK_Team]),'Fact Metrics'[FK_Date] = EARLIER ('Fact
Metrics'[FK_Date]))),'Fact Metrics'[PF])

MIN (Team Acu.) = SUMX (FILTER ('Fact Metrics', AND ('Fact Metrics'[FK_Team] = EARLIER
('Fact Metrics'[FK_Team]),'Fact Metrics'[FK_Date] = EARLIER ('Fact
Metrics'[FK_Date]))),'Fact Metrics'[MIN])

AST (Team Acu.) = SUMX (FILTER ('Fact Metrics', AND ('Fact Metrics'[FK_Team] = EARLIER
('Fact Metrics'[FK_Team]),'Fact Metrics'[FK_Date] = EARLIER ('Fact
Metrics'[FK_Date]))),'Fact Metrics'[AST])

BLK (Team Acu.) = SUMX (FILTER ('Fact Metrics', AND ('Fact Metrics'[FK_Team] = EARLIER
('Fact Metrics'[FK_Team]),'Fact Metrics'[FK_Date] = EARLIER ('Fact
Metrics'[FK_Date]))),'Fact Metrics'[BLK])

REB (Team Acu.) = SUMX (FILTER ('Fact Metrics', AND ('Fact Metrics'[FK_Team] = EARLIER
('Fact Metrics'[FK_Team]),'Fact Metrics'[FK_Date] = EARLIER ('Fact
Metrics'[FK_Date]))),'Fact Metrics'[REB])

LEAGUE Total MIN P/Year = SUMX (FILTER ('Fact Metrics', 'Fact Metrics'[FK_Date] = EARLIER
('Fact Metrics'[FK_Date])),'Fact Metrics'[MIN])
```

Figure 10: Additional statistics for *WP* formula

```
TEAM AST/BLK P/Year = CALCULATE (SUM ('Fact Metrics'[AST(Team Acu.)])*0.022 + SUM ('Fact
Metrics'[BLK(Team Acu.)])*0.019, FILTER ('Fact Metrics', AND ('Fact Metrics'[FK_Player]
= EARLIER ('Fact Metrics'[FK_Player]),'Fact Metrics'[FK_Date] = EARLIER ('Fact
Metrics'[FK_Date]))))

LEAGUE AST/BLK P/Year = CALCULATE (SUMX (DISTINCT ('Fact Metrics'[FK_Team]),FIRSTNONBLANK
('Fact Metrics'[TEAM AST/BLK P/Year],0)), FILTER ('Fact Metrics','Fact
Metrics'[FK_Date] = EARLIER ('Fact Metrics'[FK_Date])))

TEAM 48 Factor = (('Fact Metrics'[Opp.3PM]*(-0.065) + (('Fact Metrics'[Opp.FGM]-'Fact
Metrics'[Opp.3PM]))*(-0.032) + 'Fact Metrics'[Opp.TO]*0.034 + 'Fact Metrics'[REB(Team
Acu.)]*0.034) / 'Fact Metrics'[MIN(Team Acu.)]) * 48

LEAGUE AVG 48 Factor = CALCULATE (SUMX (DISTINCT ('Fact Metrics'[FK_Team]),FIRSTNONBLANK
('Fact Metrics'[TEAM 48 Factor],0)), FILTER ('Fact Metrics','Fact Metrics'[FK_Date] =
EARLIER('Fact Metrics'[FK_Date])))
```

Figure 11: Intermediate calculations for *WP* formula

4.1.2. Additional Measures

Following the previous approach, this section contains all the metrics created to improve the dashboards analysis. When compared to calculated columns, measures are strongly recommended by the specialists, as they are less memory consuming than calculated columns, and they work as an ongoing tool that can applied whenever the user requires for it.

Figure 14, is composed by measures that represents every step of Hollinger's formula – *PER*, *VA* and *EWA*. It is relevant to note that, in the prototype it is implemented the simplified version of *PER*, since there was not a significant difference between original's *PER* and the one studied by Zach Fein – with linear weights. ***Figure 15***, is composed by measures that represents *Berri's* both adjustments and final formula – *PROD*; *MATE48*; *P48*; *TM48*, *WP48* and *WP*. Since *WP* is defined by a more complex and detailed process, the development of the previous calculated columns – mentioned in grey color – was crucial to enable the creation of final *WP* adjustments such as *TEAM MATE 48*, *LEAGUE MATE 48* and *TM48*. The position adjustment values, was applied directly inside *WP48* measure by the application of the IF condition.

```
PER = ( (SUM('Fact Metrics'[FGM])*85.91)+
        (SUM('Fact Metrics'[STL])*53.897)+
        (SUM('Fact Metrics'[3PM])*51.757)+
        (SUM('Fact Metrics'[FTM])*46.845)+
        (SUM('Fact Metrics'[BLK])*39.190)+
        (SUM('Fact Metrics'[OREB])*39.190)+
        (SUM('Fact Metrics'[DREB])*14.707)+
        (SUM('Fact Metrics'[AST])*34.677)-
        (SUM('Fact Metrics'[PF])*17.174)-
        (SUM('Fact Metrics'[FT missed])*20.091)-
        (SUM('Fact Metrics'[FG missed])*39.190)-
        (SUM('Fact Metrics'[TOV])*53.897))/(SUM('Fact Metrics'[MIN]))

VA = IF(
    VALUES('Fact Metrics'[POSITION])="G",SUM('Fact Metrics'[MIN])*([PER]-11)/67,
     IF(
    VALUES('Fact Metrics'[POSITION])="SG",SUM('Fact Metrics'[MIN])*([PER]-10.5)/67,
     IF(
    VALUES('Fact Metrics'[POSITION])="SF",SUM('Fact Metrics'[MIN])*([PER]-10.5)/67,
     IF(
    VALUES('Fact Metrics'[POSITION])="PF",SUM('Fact Metrics'[MIN])*([PER]-11.5)/67,
     IF(
    VALUES('Fact Metrics'[POSITION])="C",SUM('Fact Metrics'[MIN])*([PER]-10.6)/67)))))

EWA = [VA]/30
```

Figure 12: List of measures for *EWA* calculation

```
PROD = ((SUM('Fact Metrics'[3PM])*0.065)+
        (SUM('Fact Metrics'[2PM])*0.032)+
        (SUM('Fact Metrics'[FTM])*0.018)+
        (SUM('Fact Metrics'[FG missed])*(-0.034))+
        (SUM('Fact Metrics'[FT missed])*(-0.015))+
        (SUM('Fact Metrics'[OREB])*0.034)+
        (SUM('Fact Metrics'[DREB])*0.034)+
        (SUM('Fact Metrics'[TOV])*-0.034)+
        (SUM('Fact Metrics'[STL])*0.034)+
        (SUM('Fact Metrics'[opp.FTM (P/Foul)])*(-0.018))+
        (SUM('Fact Metrics'[BLK])*0.019)+
        (SUM('Fact Metrics'[AST])*0.022))

TEAM MATE 48 = ((SUM('Fact Metrics'[AST(Team Acu.)])*0.022 + SUM('Fact
Metrics'[BLK(Team Acu.)])*0.019) / SUM('Fact Metrics'[MIN(Team Acu.)])) * 48

LEAGUE MATE 48 = ((SUM('Fact Metrics'[LEAGUE AST/BLK P/Year])/30) / (SUM('Fact
Metrics'[LEAGUE Total MIN P/Year])/30)) * 48

TM48 = SUM('Fact Metrics'[TEAM 48 Factor]) - (SUM('Fact Metrics'[LEAGUE AVG 48
Factor])/30)

P48 = ((([PROD]/SUM ('Fact Metrics'[MIN]))*48) - ([TEAM MATE 48]-[LEAGUE MATE 48]))

WP48 = IF(
    VALUES('Fact Metrics'[POSITION])="G",[P48]-0.191+[TM48]+0.1,
       IF(
    VALUES('Fact Metrics'[POSITION])="SG",[P48]-0.158+[TM48]+0.1,
       IF(
    VALUES('Fact Metrics'[POSITION])="SF",[P48]-0.186+[TM48]+0.1,
       IF(
    VALUES('Fact Metrics'[POSITION])="PF",[P48]-0.256+[TM48]+0.1,
       IF(
    VALUES('Fact Metrics'[POSITION])="C",[P48]-0.296+[TM48]+0.1)))))

WP = ([WP48]/48) * SUM('Fact Metrics'[MIN])
```

Figure 13: List of measures for *WP* calculation

To conclude, ***Figure 16*** includes several DAX code percentages – *2P%*; *3P%*; *FG%*; and *FT%* – that were created to deal with data type errors that were inside the original columns, these issues could not be fixed inside power query editor. Subsequently, there are four overall league values representing the last ten years league average – *2P% AVG*; *3P% AVG*; *FG% AVG*; and *FT% AVG.* These percentages will serve as a target value, similarly to the previous calculated columns from 3.5.3, to evaluate if the player or team are above or below the league average. For the evaluation of the most defensive players, two measures were created – *DREB P/Game* and *OREB P/Game* – using the respective calculated columns that were created previously.

Initially, the project does not have many measures that could be applied on overall league visualizations. To establish a solution for this need, two measures were created – *PA (Off.E)* and *PE (Def.E).* These two measures are known as, Offensive Efficiency and Defensive Efficiency, respectively, and both of them are based on *WP* principles from **(Berri & Schmidt, 2012)** approach. Finally, the last measure works as the privileged information – named "Expected Salary" – and it will be available only for coaches, allowing the comparison between the real player salary and the salary, calculated based on his performance. This measure uses the average between both, *WP* and *EWA* formulas, to create a balance for the situations where both measures do not have similar values.

```
2P % = DIVIDE(SUM('Fact Metrics'[2PM]),SUM('Fact Metrics'[2PA]),0)
3P % = DIVIDE(SUM('Fact Metrics'[3PM]),SUM('Fact Metrics'[3PA]),0)
FG % = DIVIDE(SUM('Fact Metrics'[FGM]),SUM('Fact Metrics'[FGA]),0)
FT % = DIVIDE(SUM('Fact Metrics'[FTM]),SUM('Fact Metrics'[FTA]),0)

2P % AVG = AVERAGEX(ALL('Fact Metrics'),[2P %])
3P % AVG = AVERAGEX(ALL('Fact Metrics'),[3P %])
FG % AVG = AVERAGEX(ALL('Fact Metrics'),[FG %])
FT % AVG = AVERAGEX(ALL('Fact Metrics'),[FT %])

DREB P/Game = SUM('Fact Metrics'[TOTAL DREB P/Player]) / SUM('Fact Metrics'[TOTAL GP])
OREB P/Game = SUM('Fact Metrics'[TOTAL OREB P/Player]) / SUM('Fact Metrics'[TOTAL GP])

PA (Off.E) = SUM('Fact Metrics'[Opp.PTS]) / (SUM('Fact Metrics'[Opp.FGM]) + 0.45*SUM('Fact
Metrics'[Opp.FTM]) + SUM('Fact Metrics'[DREB]) + SUM('Fact Metrics'[Opp.TO]) + SUM('Fact
Metrics'[REB(Team Acu.)]))
PE (Def.E) = SUM('Fact Metrics'[PTS]) / (SUM('Fact Metrics'[FGA]) + 0.45*SUM('Fact
Metrics'[FTA]) + SUM('Fact Metrics'[TOV]) - SUM('Fact Metrics'[OREB]))

Exp.SALARY B = ((SUM('Fact Metrics'[NBA Revenue])*0.5)/2460) * [WP]
Exp.SALARY H = (((SUM('Fact Metrics'[NBA Revenue])*0.5)/2460) * [EWA]
Exp. Salary (H&B) = AVERAGEX('Fact Metrics',((SUM('Fact Metrics'[NBA Revenue])*0.5)/2460)
* (([WP]+[EWA])/2))
```

Figure 14: List of additional measures for dashboard visualizations

4.2. FINAL DIMENSIONAL MODEL (DATA SOURCES)

This final *OLAP* dimensional model, was developed inside Power BI software, where most of the data was obtained through WEB sources – specifically from NBA official website. Only the players' pictures, and the respective team and conference logos, were obtained through an excel with a list of *URLs,* that allows the connection between the WEB image and power BI. The dimensional model presented in the figure below, has a Star Schema design, where each dimension tables are already denormalized and holds a relationship with the "Fact Metrics". This fact table is collapsed in this picture, due to its size, but will be presented and explained in the next page.

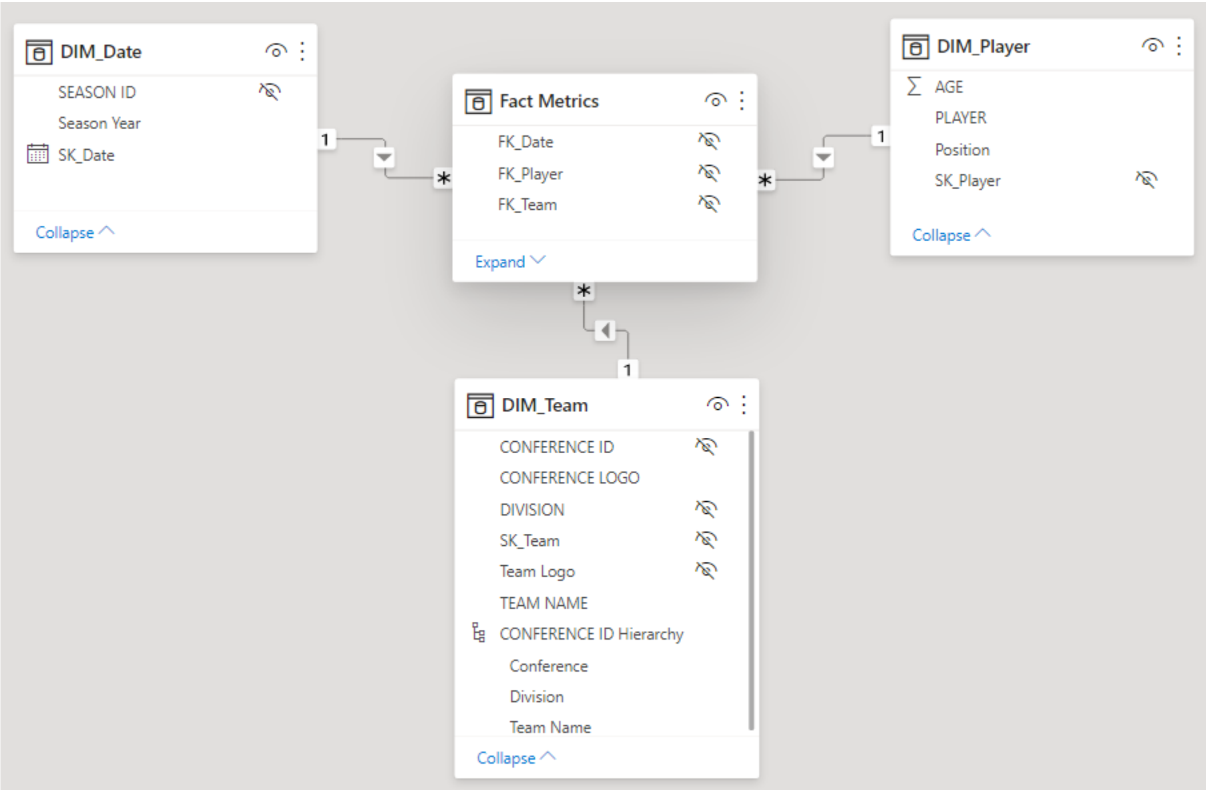


Figure 15*:* Final Dimensional Model (Star Schema)

In the following ***Figure 16***, there is an expanded fact table, named "*Fact Metrics*", that corresponds to the dimensional model presented in the previous ***Figure 15***. This fact table is the central node of the data warehouse, being surrounded by dimensions tables, that are linked by their respective Surrogate Keys (SK) and by Foreign Keys (FK). Additionally, the *"Fact Metrics"* table also includes ***Columns*** – that have been inserted, fixed, transformed and loaded into the model, throughout the 'Power Query Editor' functionality; ***Calculated Columns*** created throughout the project – previously presented with the respective DAX codes explained; and several ***Measures*** – also referred in the previous chapter along with the respective DAX codes explained – that were specifically created to facilitate the dashboards development and improve their visualizations.

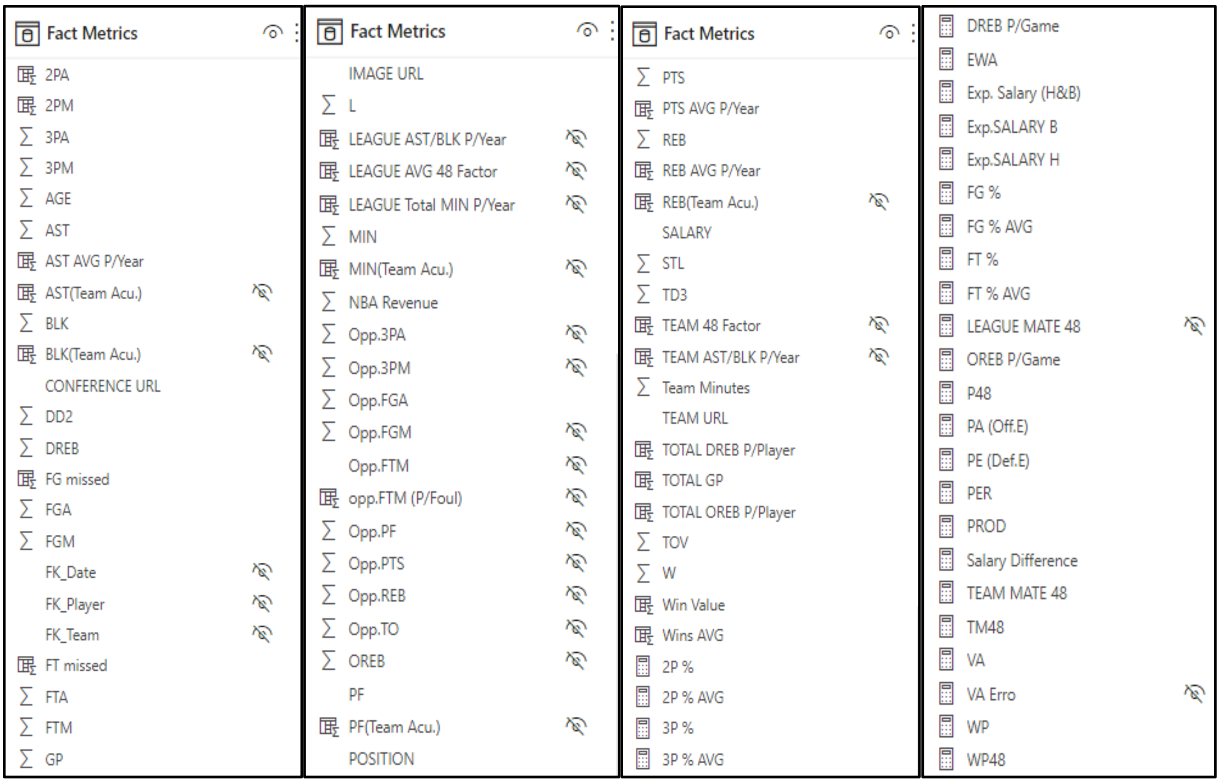


Figure 16: Expanded fact metrics

4.3. Dashboards Prototypes

This chapter will illustrate several prototype examples, that were designed and described previously. In addition to the structure and design explanation, there will be also a brief analytical interpretation, showing the way that the "end-user" should look and interpret these dashboards. The purpose of this section is to verify, if the goals that were established at the beginning of the project are achieved in a successful way.

4.3.1. Prototype – Initial page Overview

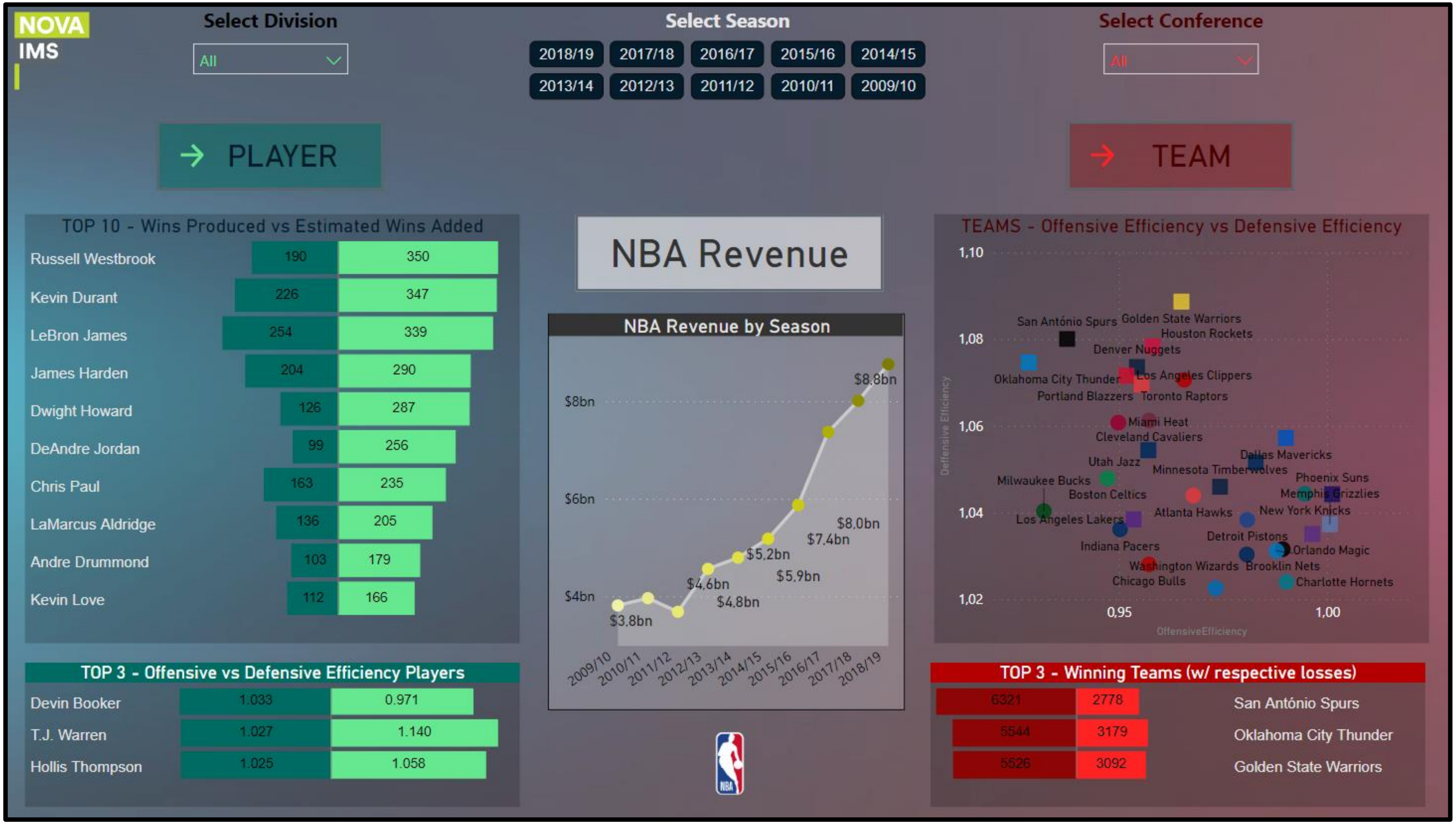


Figure 17: Initial page dashboard

The ***Figure 17*** is the first prototype – named "Overview" – that will display, in a broader way, the most relevant information for this study. As secondary function, it offers the possibility to navigate between both, player, and coach's dashboards, and allows the filtering by division, season, and conference. Its structure is composed by five visualizations, divided into two distinct parts, one for player and another for team. In the center, there is just a small visualization, showing the *NBA* revenue evolution by year. The relevance of this information relies on the fact that it is one of the crucial parts to create the formula that will allow to achieve the player's expected salary.

On **player's** side, in ***Figure 17***, there is some cumulative information, based on PROD formula, regarding players with the most wins added in the league – by *WP* and *EWA* formulas. It is notable that, players who add the most wins, are also the best and most popular ones inside the *NBA* environment. Immediately below, there is a double funnel visualization, which shows the three best players in terms of offensive and defensive efficiency. To avoid some bias and in a way of displaying three players with a high level of consistency when it comes to performance quality, the visualization includes a filter that evaluates players that have more than 500 games.

On **team´s** side, in ***Figure 17***, a scatter chart targeted to the league is presented, with the aim of showing the relationship between offensive and defensive efficiency. The teams that are inside the second quadrant – such as, *Oklahoma City thunder, San Antonio Spurs*, among others – are more defensive than offensive, while those located in the fourth quadrant – such as, *Charlotte Hornets*, *Orlando Magic*, *Washington Wizards*, among others – are more offensive than defensive. Immediately below, there is a double funnel, where it is possible to find the three most winning teams, with their respective losses.

By comparing both 2009/10 (left) and 2018/19 (right) seasons, in ***Figure 18***, it is possible to reflect on the paradigm that is widely discussed about the way of playing inside *NBA* through the years – the transition from an interior and more defensive game to a more exterior and offensive game.

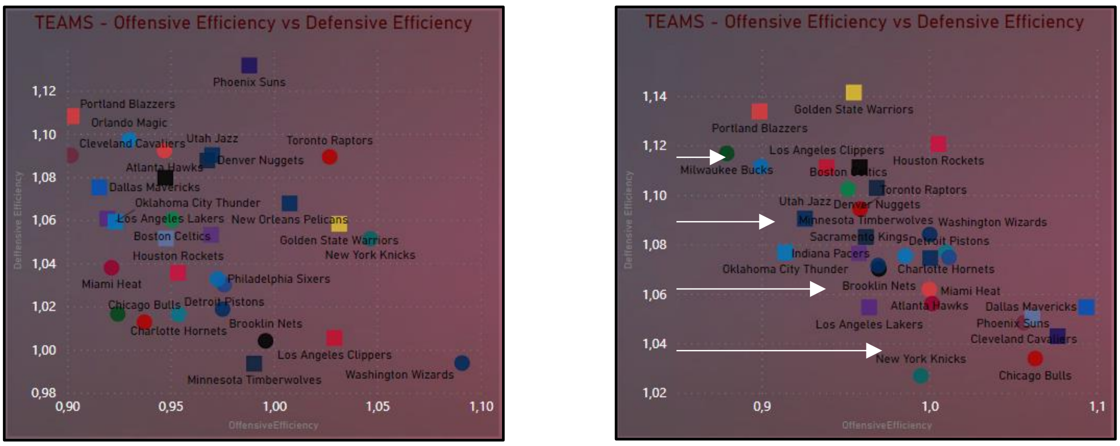


Figure 18: Efficiency transition between seasons 2009/10 & 2018/19

4.3.2. Coach Prototype – Team Overview

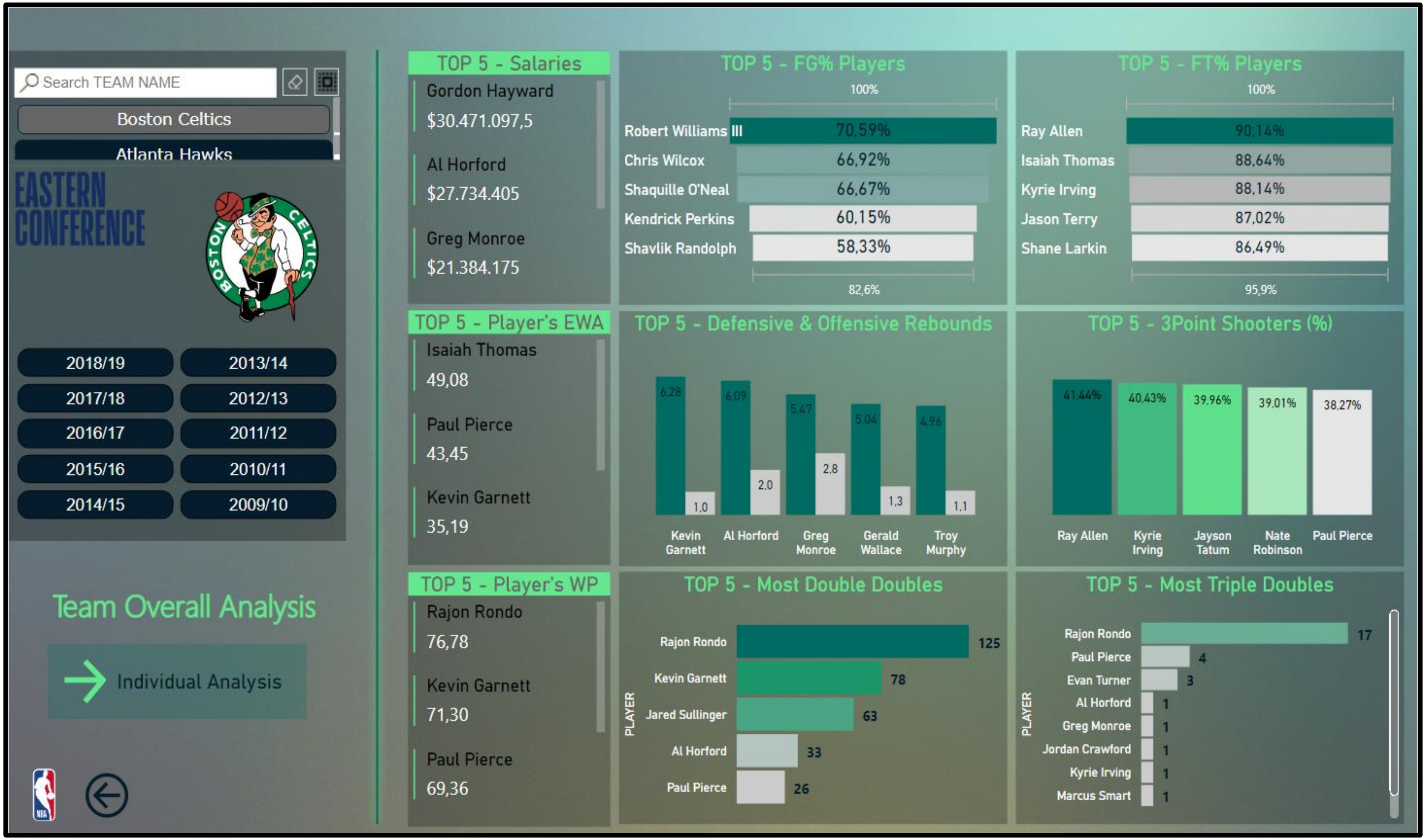


Figure 19: Team overview (Coach dashboard)

To follow the mockups sequence presented in the previous chapter, ***Figure 19***, corresponds to the coach's prototype – named "*Team Overview*". The main functionality of this dashboard, is to provide an evaluation based on the best five players of each team. Every coach has access to his own team and his top five players across several seasons. Since the *figure 18* does not have any season selected, the dashboard displays a global analysis over the last ten seasons.

Figure 20: Team overview dashboard (Reduced)

The structure is divided into 3 distinct parts. Inside the left ***red*** box, the user will find an interactive navigation between teams and seasons. Then, it is presented one list that shows the five most expensive players, and other two lists showing the most contributors, in terms of wins added. Considering that each team is not only composed by offensive factors, there are three visualizations, surrounded by the ***blue*** part, with four types of stats that are predominant on defensive players – usually playing at *Center*; *Power Forward* and *Small Forward* positions. Each visualization will show the top five players from *FG%, both* defensive & offensive rebounds and the most double doubles[9]. The ***yellow*** box, is focused on the opposite. The three remaining visualizations, are representing three types of stats – best *FT* and *3P* percentages as well as the most triple doubles[10]. These last stats are predominant on offensive player – usually playing in positions such as *Guard*; *Shooting Guard* and some *Small Forward*.

Figure 21: League overview (Player dashboard)

This ***Figure 21***, is known as the "*League Overview*" prototype, designed for the player dashboard. In terms of structure, it has practically the same layout as the previous one but evaluates the five best players in the league instead of the team. Finally, the player will be able to gain get useful information by understanding which players are having the best performances within the league, and they will have access to filter by conference and season .

[9] Double doubles is a performance in which a player accumulates a double-digit in two of the five statistical categories – points, rebounds, assists, steals, and blocked shots. More usual in 'inside' game players.

[10] Triple doubles is a performance in which a player accumulates a double-digit in three of the five statistical categories mentioned in note 9. More common in 'outside' and playmaking players.

4.3.3. Coach Prototype – Individual Player Analysis

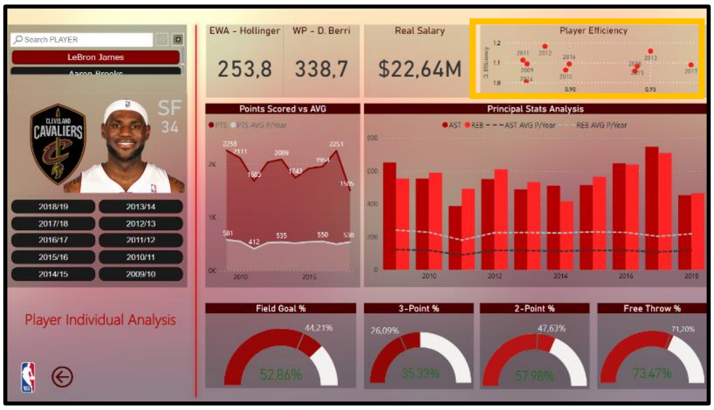
Figure 22: Individual player analysis (Coach dashboard)

Figure 22, corresponds to the coach's prototype – named "*Individual Player Analysis*" – and its main function, is to provide an individual performance analysis, based on statistics, salaries, and the *NBA* revenue. Each coach will have access to every league player, with the possibility to navigate between several seasons. Similar to the *figure 19*, this dashboard does not have a single season selected, and therefore, it displays a global player analysis over the last ten seasons, with the respective extra card information. Since, in this specific case, we are in the presence of an overall evaluation, the cards' values are shown as a sum.

Figure 23: Individual player analysis (reduced)

The structure is also divided into 3 parts. Inside the left ***red*** box, the end-user will find the navigation menu, which enable the player selection – by writing the respective name or by drilling down players names from the list; this dashboard offers the capability to choose a specific season and see the player's position and age. On the top, and inside the ***blue*** surrounded part, it is presented a set of cards showing the player's contributions within his teams in terms of wins – *EWA* and *WP* formulas. It is also displayed some privileged coach information, such as, the real *player salary*, *expected salary* – calculated based on *NBA* revenue and the previous cards – and the difference between both metrics on the right-side card. It is important to note that every time the expected amount is higher than the real amount, the difference will appear in green, as it means the player could receive more, otherwise it will appear in red color. Inside the ***yellow*** box, it is presented the most common statistics – *points*, *rebounds*, *assists*, *FG%, 3P%, 2P%* and *FT%* – and the respective league averages, to enable the comparison between the player performance and the league values. The bottom gauges will also show green if the value surpasses the league average.

Figure 24: Individual analysis (Player dashboard)



This ***Figure 24*** , known as "*Individual Analysis*", was designed for the player dashboard. In terms of structure, it has practically the same layout as the previous one, but instead of "*Expected Salary*" and "*Difference*" cards, it contains a scatter chart showing the "*Player Efficiency*" across his respective seasons.

4.3.4. Dashboard Interpretation

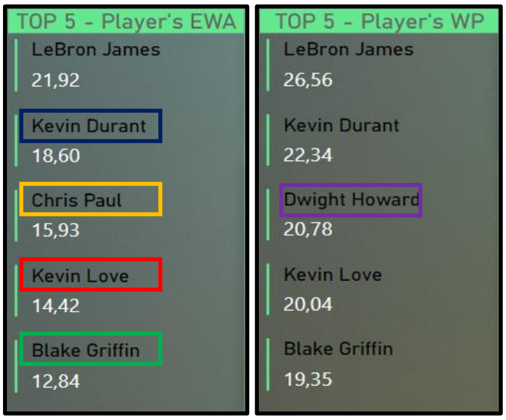
The purpose of this section, is to interpret two of the four prototypes available on the project. The idea is to understand, if the results are giving, or not, some useful information, that would be able to help both players and coaches. To create a more coherent interpretation, it will be presented one example from "*Player - League Overview*" dashboard, followed by one "*Coach - Individual Player Analysis*" dashboard.

<u>*League Overview*</u>

Figure 25: Example - League Overview Dashboard

On this dashboard, only the *2011/12* season is selected, and therefore it shows the respective top five player's analysis covering both conferences – *Western* and *Eastern*. One of the several interpretations that can be made on this dashboard, is to look at *"TOP 5 - Salaries"* list, and notice that the former all-star *Kobe Bryant*, is the highest paid player in the season. This player, has achieved several honors during his career – such as, five *NBA* championships, two *NBA final MVP's*, one season *MVP*, 18 appearances in the *All-Star* game, among many others – and reached a level where his reputation and status are extremely high, when compared to other players. However, after analyzing the remaining dashboard, *Kobe Bryant* – of age 32 – is not amongst the five players who added the most wins to his

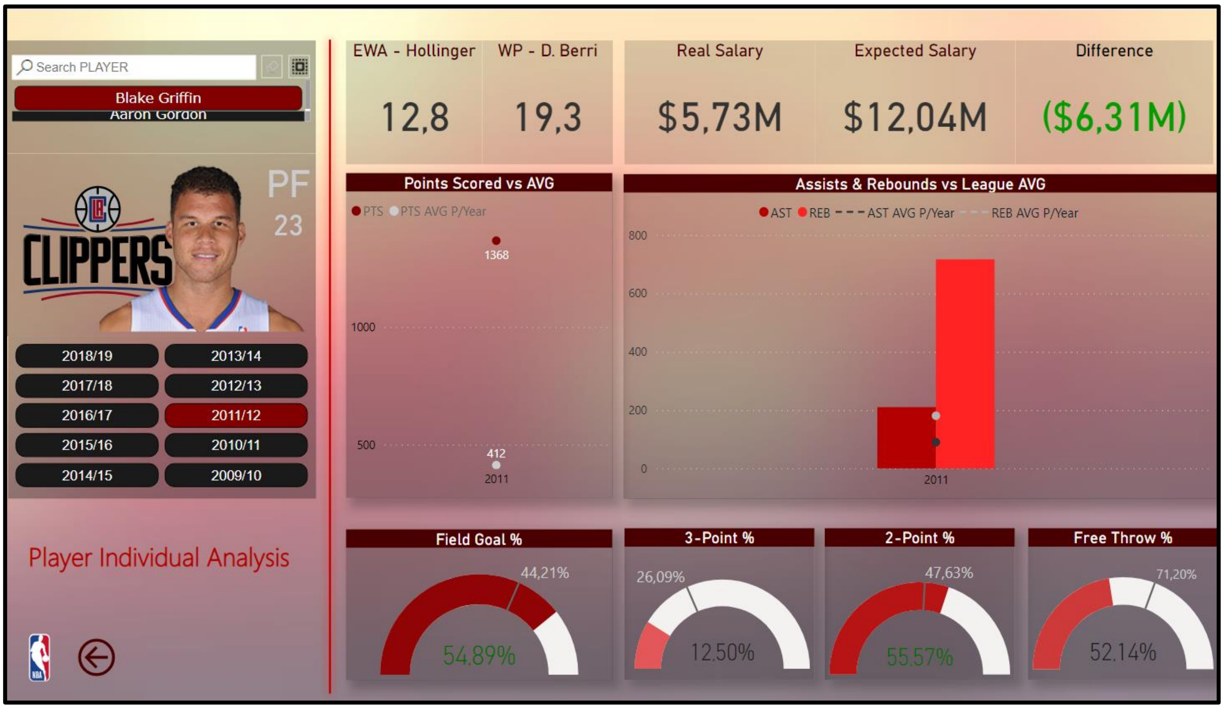
Figure 26: Lists from league overview



teams, neither in the rest of the main statistics. On the other hand, a correlation can be perceived in some young talents – such as, *Kevin Durant*, *Blake Griffin*, *Chris Paul,* and *Dwight Howard* – who are simultaneously between the top five players who have added the most wins and have also great statistical values in some visualizations. Through this small analysis, it seems almost clear that there is a certain probability that *Kobe Bryant* was overpaid during the last years of his career, due to the recognition, the reputation he achieved with the league over the years and the status that he brought to *Los Angeles Lakers*, turning the team more visible and attractive in terms of supporters.

Individual Player Analysis

Figure 27: Example - Blake Griffin Individual Analysis



To maintain the previous line of analysis, the following dashboard presents the values of *Blake Griffin*, who was – in the 2011/12 season – one of the five best players when it comes to wins contributions. His performance was above the league average on almost all stats – apart from his *3P%* and *FT%* conversion – however, his *Real Salary* was considerably less than the *Expected Salary.* Considering the previous and following seasons, it is easily understandable that, *Blake Griffin*, was an over-average player in almost every season, with undeserved salaries. At the time, comparatively to *Kobe Bryant*, *Blake Griffin* was a player with an inferior reputation and his quality was underestimated, probably due to his age and the lack of awards.

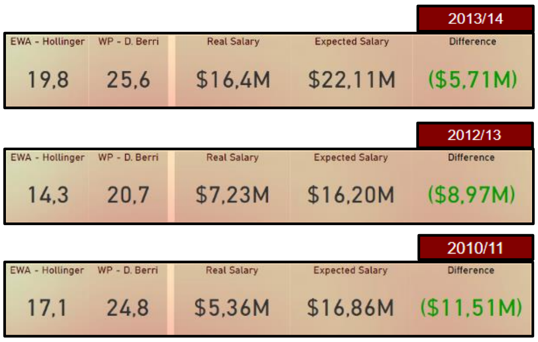


Figure 28: Blake Griffin salary cards

5. PROJECT AND RESULTS EVALUATION

The project stared with an initial literature review, which was meant to connect both, abstract and concrete aspects. After the literature review conclusion, some correlations regarding player's reputation, status, performance, and salaries were made. Then, the goal was to create a realistic solution that would incorporate a variety of data types in a single tool that could be constantly used to improve decision making. To better understand the results, the list below summarizes the major steps that were defined and achieved throughout the project:

- ✓ The conceptual model was built around the foundations of the literature review. The recommended data in the literature review – player and teams' statistics, salaries, league revenues – was properly employed. Additional data was included for illustrations – such as, players' pictures, team logos, and conferences pictures;
- ✓ To ensure a type of information that would allow a consistent analysis, ten years of data were collected;
- ✓ To develop a proper dimensional model, following the star schema design, some changes had to be made around the stats that came from several web sources. Inside the data transformation tool – power query editor – some changes involving the removal of null values; replacement values; addition of columns through examples; construction of surrogate and foreign keys; the merge and expansion of columns were made;
- ✓ Once the data was properly extracted and transformed, it was loaded into the dimensional model, where all the metrics and formulas indicated throughout the conceptual model were inserted successfully, through the Power BI DAX code. Having both different functions and different roles within the dimensional model, a balanced number of calculated columns and measures have been constructed;
- ✓ The data is ready to be used, allowing the end-user to benefit from better decision making;
- ✓ Based on the previous point, the dashboard prototypes are providing a solution that meets the project's requirements. The final output is user friendly, containing a whole range of information to assist the decision-making process and performance management;
- ✓ Being the information presented, simplified, practical and useful, it can be concluded that, the final dashboard itself is pretty accessible, even for those who are not familiar with basketball and the *NBA* league;

Having followed the design science research (*DSR*) methodology, the stated steps were met throughout the project in a positive way.

6. CONCLUSION

The sports industry is deeply ingrained in the culture of some countries, being the *United States* one of those where this rhetoric is strongly present. It is a fact that nations as *US*, are constantly pursuing new types of knowledge management, analytical tools, and technologies inside the sports environment. With the continuous growth of game quality, revenues and requirements, the need of additional solutions, which can bring value to the game, are always in need – the same happens within the *NBA* league as well.

The development of this project intended to create an artifact that could embrace several types of information needs, that nowadays, are a bit scattered in the *NBA* quotidian. The main target was to present "*a business intelligence solution that can prove the added value of self-learning around the visualization and interpretation of power BI dashboards*".

Having tested each one of the prototypes as a final solution to the performance analysis, we can state that the dashboards worked as expected – providing a solution that directly meets the initial goal. To achieve the *WP* and *EWA* formulas, multiple measures and calculated columns were created. However, the dimensional model also includes many additional ones, which have been created with the intention to provide a richer analysis in terms of detail and in terms of facilitating the end-user interpretation.

The connection that was created in the same artifact between, a *performance analysis*, and the presentation of other measures – such as the *number of wins added*, the respective *salaries*, and the *expectable salaries* – is one of the benefits that a business intelligence approach can bring to an organization's strategy.

The indicators proposed in the *player's* perspective, are a tool that follows the league's evolution, by telling the best players and tracing their own evolution. Promoting the player's intention of setting goals, define levels to reach, and promote consistency in their performances by staying aware of their own records. The indicators offered in the *coach's* perspective are a launching pad to improve the management of his players, allowing better decision making that offer more stability and success to the team.

Through the analysis of the process, it is concluded that the project is aligned with the topic addressed in the literature review. The outcome is quite satisfactory, in the sense that we are looking at an artifact that provides a practical and reliable solution that can be applied to the real world.

7. LIMITATIONS AND FUTURE RECOMMENDATONS

This last chapter will contain a list based on the limitations encountered during the project's development, followed by a list of recommendations that can support future projects.

Limitations:

- ✓ It was challenging to structure and frame a literature review that was consistent with the intended goal, due to the lack of data and studies related to the topic;
- ✓ Obtaining all the players' salaries through a web source was a challenging process, as the web HTML was giving problems on the software,
- ✓ The process of introducing auxiliary calculations, intermediate formulas, and final formulas – inside the dimensional model – was quite challenging. In the project, the most diverse DAX code functions were used, requiring a deep understanding of the respective code language;
- ✓ There were some Power query editor difficulties, since there were several errors within the percentage columns, it was important to recreate the columns using DAX code;
- ✓ The team and player's pictures were part of a personal touch within the project, and it took some additional effort. The respective pictures could not be found among the sources mentioned throughout the project;

Future recommendations:

- ✓ Identify reasons and factors for the lack of positive results, in conditions of superiority, where the players performance is supposed to be quite superior;
- ✓ Connect these dashboards directly to the *NBA* database and provide them a real time analysis that can offer real time information;
- ✓ Extend the project to all sports and leagues – such as *NHL*, *NFL*, *MLB* among many others – always being required all the necessary data for its implementation;
- ✓ Test the dashboards with *NBA* experts, so they can express their opinions and suggest future ideas;

8. REFERENCES

Bailey, A. (2016). *5 Technological Advances That Are Changing the NBA*. Bleacherreport.Com. https://bleacherreport.com/articles/2627381-5-technological-advances-that-are-changing-the-nba

Bension, J. B. (2019). *The Importance of NBA Box Score Statistics and the Value of Statistical Outbursts* (Issue November). http://repository.library.csuci.edu/bitstream/handle/10211.3/214528/Bension, Jack MSCS Thesis F 19_done.pdf?sequence=1

Berri, D. J. (1999). Who is 'most valuable'? Measuring the player's production of wins in the National Basketball Association. *Managerial and Decision Economics*, *20*(8), 411–427. https://doi.org/10.1002/1099-1468(199912)20:8<411::AID-MDE957>3.0.CO;2-G

Berri, D. J., Bradbury, J. C., S.Rosentraub, M., Mason, D. S., Zimmerman, D., Almen, P. von, Teixeira, C. P. B. and M., Davis, M. C., Mondello, M., Quinn, K. G., Mason, D. S., Duquette, G. H., & Zillante, A. (2008). *The Business of Sports - Volume 3: Bridging Research and Practice* (B. R. Humphreys & D. R. Howard (eds.); First). Pragear Publishers.

Berri, D. J., & Jewell, R. T. (2004). Wage inequality and firm performance: Professional basketball's natural experiment. *Atlantic Economic Journal*, *32*(2), 130–139. https://doi.org/10.1007/bf02298830

Berri, D. J., & Schmidt, M. (2012). *The Wages of Wins Journal*. Wagesofwins. http://wagesofwins.com/how-to-calculate-wins-produced/

Berri, D. J., Schmidt, M. B., & Brook, S. L. (2006). *The Wages of Wins -Taking Measure of the Many Myths in Modern Sports* (2006th ed.). Stanford University Press.

Bol, T., De Vaan, M., & Van De Rijt, A. (2018). The Matthew effect in science funding. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(19), 4887–4890. https://doi.org/10.1073/pnas.1719557115

Borghesi, R. (2008). Allocation of scarce resources: Insight from the NFL salary cap. *Journal of Economics and Business*, *60*(6), 536–550. https://doi.org/10.1016/j.jeconbus.2007.08.002

Brett Knight. (2019). *NBA's Most Overpaid Players 2019: Wiggins, Jabari And 8 Others Who Underperformed*. Forbes.Com. https://www.forbes.com/sites/brettknight/2019/04/12/nba-overpaid-wiggins-jabari/#2b56de0776e3

Coon, L. (2020). *NBA Salary CAP FAQ*. CBA FAQ. http://www.cbafaq.com/salarycap.htm#Q12

Debrock, L., Hendricks, W., & Koenker, R. (2004). Pay and Performance: The Impact of Salary Distribution on Firm-Level Outcomes in Baseball. *Journal of Sports Economics*, *5*(3), 243–261. https://doi.org/10.1177/1527002503259337

Ertug, G., & Castellucci, F. (2013). Getting what you need: How reputation and status affect team performance, hiring, and salaries in the NBA. *Academy of Management Journal*, *56*(2), 407–431. https://doi.org/10.5465/amj.2010.1084

ESPN. (2021). *NBA Players Salaries*. Espn.Com. http://www.espn.com/nba/salaries

Goldberg, R. (2020). *Wizards' Bradley Beal Slams "Disrespectful" 2020 NBA All-Star Snub*. Bleacherreport.Com. https://bleacherreport.com/articles/2873956-wizards-bradley-beal-slams-disrespectful-2020-nba-all-star-snub

Goldrosen, J. M. (2018). *The Motivations and Effects of the NBA Salary Cap* [Bard College,]. https://digitalcommons.bard.edu/senproj_s2018/119/

Hauser, M. (2009). *Are Statistics Becoming Too Important in Sports?* Bleacherreport.Com. https://bleacherreport.com/articles/130273-are-statistics-becoming-too-important-in-sports´

John Hollinger. (2007). *What is PER?* Espn.Com. https://www.espn.com/nba/columns/story?columnist=hollinger_john&id=2850240

John Hollinger. (2009). *PER Diem: March 25, 2009*. Espn.Com. https://www.espn.com/nba/columns/story?columnist=hollinger_john&page=PERDiem-090325

Justin Jacobs. (2017). *Breaking Down Player Efficiency Rating*. Squared2020.Com. https://squared2020.com/2017/09/01/breaking-down-player-efficiency-rating/

Justin Willard. (2018). *Reinventing PER*. Fansided.Com. https://fansided.com/2017/01/31/nylon-calculus-reinventing-per/

Koneru, L. (2018). Using Design Science Research to Develop a Conceptual Solution for Improving Knowledge Sharing in a Virtual Workspace. In *Andhra University* (Issue December). https://digitalscholarship.unlv.edu/cgi/viewcontent.cgi?article=4505&context=thesesdissertations

Lazear, E. P. (1989). Pay Equality and Industrial Politics. *Journal of Political Economy*, *97*(3), 561–580.

Lazear, E. P. (1991). Labor Economics and the Psychology of Organizations. *Journal of Economic Perspectives*, *5*(2), 89–110. https://doi.org/10.1257/jep.5.2.89

Leeds, M. A., & Kowalewski, S. (2001). Winner Take All in the NFL: The Effect of the Salary Cap and Free Agency on the Compensation of Skill Position Players. *Journal of Sports Economics*, *2*(3), 244–256. https://doi.org/10.1177/152700250100200304

NBA. (2019). *National Basketball Association Constitution and By-Laws*. *September*, 1–91. https://ak-static.cms.nba.com/wp-content/uploads/sites/4/2019/09/NBA-Constitution-By-Laws-September-2019-1.pdf

NBA & SAP. (2020). *NBA Advanced Stats*. Nba.Com. https://www.nba.com/stats/

Offermann, P., Levina, O., Schönherr, M., & Bub, U. (2009). Outline of a Design Science Research Process. *4th International Conference on Design Science Research in Information Systems and Technology*, 11. https://wise.vub.ac.be/sites/default/files/thesis_info/outline_of_design_science_research_process.pdf

Podolny, J. M. (2016). *A Status-Based Model of Market Competition Author (s): Joel M . Podolny Source : American Journal of Sociology , Vol . 98 , No . 4 (Jan ., 1993), pp . 829-872 Published by : The University of Chicago Press Stable URL : http://www.jstor.org/stable/2781*. *98*(4), 829–872.

Rossi, M., Hui, W., Peffers, K., Tuunanen, T., Gengler, C., Virtanen, V., & Bragge, J. (2015). *The Design Science Research Process*. https://arxiv.org/ftp/arxiv/papers/2006/2006.02763.pdf

Sports Reference. (2020a). *Basketball Stats and History*. Basketball-Reference.Com. https://www.basketball-reference.com/

Sports Reference. (2020b). *Calculating PER*. Basketball-Reference.Com. https://www.basketball-reference.com/about/per.htmlhttps://www.basketball-reference.com/about/per.html

Stanek, T. (2016). *Player Performance and Team Revenues: NBA Player Salary Analysis*. http://scholarship.claremont.edu/cmc_theses/1257%0AThis

Statista. (2020). *Total Revenue of the NBA 2001-2019*. Statista.Com.

Ströbel, T., Stieler, M., & Stegmann, P. (2019). Digital transformation in sport : The disruptive potential of digitalization for sport management research. *Sport, Business and Management: An International Journal*, 4. https://bit.ly/2Tt7Rs8

Trevir Nath. (2020). *The NBA's Business Model*. Investopedia.Com. https://www.investopedia.com/articles/investing/070715/nbas-business-model.asp

Utah, U. O. S. (2020). *David Berri - Expert Profile*. Suu.Edu. https://www.suu.edu/experts/profile/?expert=david.berri

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *Journal of Strategic Information Systems*, *28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003

Vrooman, J. (2000). The economics of American sports leagues. *Scottish Journal of Political Economy*, *47*(4), 364–398. https://doi.org/10.1111/1467-9485.00169

Vrooman, J. (2009). Theory of the perfect game: Competitive balance in monopoly sports leagues. *Review of Industrial Organization*, *34*(1), 5–44. https://doi.org/10.1007/s11151-009-9202-7

Wade, J. B., Porac, J. F., Pollock, T. G., & Graffin, S. D. (2006). The burden of celebrity: The impact of CEO certification contests on CEO pay and performance. *Academy of Management Journal*, *49*(4), 643–660. https://doi.org/10.5465/AMJ.2006.22083021

Washington, M., & Zajac, E. J. (2005). Status Evolution and Competition : *Academy of Management Journal*, *48*(2), 282–296. https://doi.org/10.5465/AMJ.2005.16928408

Yarrow, D., & Kranke, M. (2016). The performativity of sports statistics: towards a research agenda. *Journal of Cultural Economy*, *9*(5), 445–457. https://doi.org/10.1080/17530350.2016.1202856

Zach Fein. (2009). *Cracking the Code: How to Calculate Hollinger's PER Without All the Mess*. Bleacherreport.Com. https://bleacherreport.com/articles/113144-cracking-the-code-how-to-calculate-hollingers-per-without-all-the-mess