

|  |  |
| --- | --- |
| National basketball association  Business Intelligence | **Master’s Degree in Data Science and Advanced Analytics – Major in Business Analytics**  NOVA Information Management School  Instituto Superior de Estatística e Gestão de Informação  Universidade Nova de Lisboa  AUTHORS:  Carolina Neves // 20200049  Diogo Gonçalves // 20200632  Inês Pires // 20200757  João César // 20200669  Nuno Pais // 20200576  **March 2021**  **Professors**: Miguel de Castro Neto, Bruno Jardim |

Index

[1. Introduction 2](#_Toc73741804)

[2. Presentation of Business – NBA 3](#_Toc73741805)

[3. Business Needs 5](#_Toc73741806)

[4. Data Source and Discovery Process 6](#_Toc73741807)

[5. Perspectives of Analysis 7](#_Toc73741808)

[5.1. Operational Analysis 7](#_Toc73741809)

[5.2. Competitive Analysis 8](#_Toc73741810)

[6. Dimensional Model 9](#_Toc73741811)

[7. Data Integration, Transformation and Modelling 11](#_Toc73741812)

[8. Description of PowerBI Report 12](#_Toc73741813)

[9. Analysis and Discussion 17](#_Toc73741814)

[10. Critical Assessment 20](#_Toc73741815)

[11. Conclusion 20](#_Toc73741816)

[12. Extra – Work 21](#_Toc73741817)

[References 21](#_Toc73741818)

# Introduction

This project comes with the purpose of studying the National Basketball Association, the professional sports league with the highest player wages worldwide. This report seeks to reinforce the conceptual knowledge acquired throughout the Business Intelligence course and is meant to apply core concepts of Business Intelligence, as well as to develop an end-to-end Self-Service BI solution.

The information addressed in this project aims to be usable either by external stakeholders (sponsors) who want to get a deeper understanding on certain players’ or teams’ performance and evolution throughout the seasons, so as to better plan and allocate their sponsorship budgets, by NBA itself as an organization, by an NBA teams who wants to evaluate its players’ performance, or even by players themselves, as further explained on chapter 3.

Thus, the final goal of this project is to deliver a complete dashboard with useful visualizations and KPIs that allow for the user to interpret several aspects of the NBA, through several techniques which aim to evaluate both players and teams.

Throughout the report, two main perspectives will be adopted – the operational and competitive perspectives. As such, several teams and players can be analysed in the dashboard, as it will be detailed throughout this report, with different examples being explained in order to provide actual analyses that can be obtained through the dashboard, both in an Operational and a Competitive perspective.

Finally, in the “Extra – Work” chapter of this report, the additional work performed for the dashboard, such as PowerBI techniques not covered during class, and other functions and visuals from the marketplace will be detailed.

# Presentation of Business – NBA

The National Basketball Association, mostly known as NBA, is a professional basketball league founded in North America in 1946. It has a total of 29 teams in the United States and 1 in Canada, adding up to a total of 30 teams. During the offseason, NBA may have as many as 600 players, 20 per team; however, during the regular season NBA rosters are limited to 15 players, plus two addition spots for players under “two-way-contracts”, making a total of 450 players. The NBA’s regular season runs from October to April, with each team playing a total of 82 games, but the league's playoff tournament can extend into June.

According to *Statista*, NBA players earned an average annual salary of 8.32 million U.S dollars in 2019/20 season, being the professional sports league with the highest player wages worldwide. Although the wages are quite high, these can vary quite a lot from team to team, varying from an average of 7.07 million U.S. dollars a year in New York to 10.04 in Portland Trail Blazers in 2019/20.[[1]](#footnote-2) NBA wages have shown a continuous positive growth, comparatively to a value of 4.6 million U.S. dollars in 2015/16 season, making a growth ratio of around 81%.

In order to pay these enormous values, NBA has been quite innovative throughout the years, turning its all-star game into a three-day weekend loaded with events, becoming the first sports league to draw a significant portion of its revenues through merchandising. NBA become not only attractive due to its games but become a brand itself. These allowed to create other different streams of revenue rather than just ticket games, such as, the merchandising, television, and sponsorships. Since the live events include the ticket cost, parking and food, television has been the main option for many fans, increasing NBA’s advertising revenue and television contracts. International revenue is another stream of income that has been growing throughout the years, becoming a key revenue source. For example, it is estimated that, every year, China generates 500 million dollars in revenue. The international revenue stream is tied to international investors, such as Alibaba which purchased 49% stake in the Brooklyn Nets last year for an alleged 1.15 billion dollars.[[2]](#footnote-3) Although NBA has many captivating different streams of revenue, it is not possible to analyse them since it is not a public company and, thus, there is no access to their financial reports.

Regarding the players there are some interesting key features about them. In the 2020-21 season, 107 out of 450 players are from outside of the United States, from 41 different countries. The most represented country, after the US, is Canada with 17 players, followed by France with 9 players, Australia with 8 players and, Serbia and Germany with 6 players each. All 30 teams have at least one international player. [[3]](#footnote-4) The average NBA player, in this season, is 199cm tall, weighting 99kgs, 26 years old and has 4.56 seasons of NBA experience. Concerning the Season Leaders, Bradley Beal stands out for having the highest number of points per game, 32.1, however, Damian Lillard has the highest total points, 1244. Stephen Curry is the player that has made the most three pointers so far, a value of 182. All the mentioned players are from different teams, respectively, Washington Wizards, Portland Trail Blazers and Golden State Warriors. Interesting enough, Utah Jazz is the team with highest winning percentage, a value of 73.8, followed by Philadelphia 76ers, with a value of 69.8. Minnesota Timberwolves and Detroit Pistons are the teams with the lowest percentage values of winning with, respectively, 23.3 and 28.6.[[4]](#footnote-5)

NBA is also an organization with a high presence in the communities, having 8 different programs incorporated in the NBA Cares. This segment is the league’s global social responsibility program that has the mission of addressing important social issues in the U.S. and around the world. It has provided more than 5.8 million hours of hands-on service, created more than 1,650 places where kids and families can live, learn, or play and engaged more than 60 million kids in basketball programs in communities around the world. At an international level, NBA Cares has created more than 332 places where kids and families can live, learn, or play in 40 different countries.[[5]](#footnote-6) Another important civic engagement that NBA does is the celebration of the Black History and support the ongoing pursuit of racial justice by taking leaguewide action that spotlights the voices, experiences and perspectives of Black players, coaches, employees, and fans.[[6]](#footnote-7)

# Business Needs

As stated in the previous chapter, NBA is a competition that generates large amounts of revenue and being on the front lead of the league gives teams the possibility of making millionaire contracts.

Thus, a performance evaluation is crucial for a team to succeed. It is very important to get to know their players’ skills, strengths, and weaknesses. Being aware of a team’s performance and players’ momentum, alongside knowing past performances of players in certain counties and against specific teams, is a key success factor for any sports nowadays.

Besides, there is a huge need to know your opponent. This report aims to generate tools to access opponent players’ momentum, performance, and past statistics. With this information, a team can then create and adapt different strategies based on an outstanding scouting process. Some of these needs could be answered by this analysis, as such:

* Who is the player on my team (or opponent’s team) with the most scored points, the most assists and/or the most rebounds?
* Who is the player on my team (or opponent’s team) with the most minutes played and the most points/assists/rebounds per minute?
* Who is the player on my team (or opponent’s team) with the most played games?
* Which of my players are most likely to get injured during the season?
* What type of different players do I have on my team and the opponent’s team?
* How are my players in comparison with their past performances?

All this information will not only allow the team to prepare themselves for the current games, but to prepare for the upcoming seasons, while observing and discovering who their next acquisition could be.

Finally, sponsorships are a considerable part of a team’s revenue. This analysis can also be used by external stakeholders, such as sponsors, to create more detailed and accurate partnerships both with teams and players according to their performances and drive. An example of business needs that this project might solve are, for instance:

* Which teams have been consistently scoring more points over the last years?
* Which teams have the biggest arenas and therefore the biggest quantity of supporters in their stadiums?
* Which teams belong to the same county?

# Data Source and Discovery Process

* https://www.kaggle.com/nathanlauga/nba-games?select=teams.csv

The data source that enabled this report and analysis is composed by 6 different datasets. The games.csv includes home and way teams of all games from 2004 season until May 2020. The dataset players.csv include information regarding player’s average injuries per season, considering the total number of injuries throughout his career and the total seasons played. Moreover, the dataset games\_dates.xlsx complements the first dataset with information regarding the date when the match was played, and the location of the games are then given by the dataset games\_location.csv. The dataset games\_details.csv, which includes the most important information for the present report and dashboard, includes information regarding the games and the players that were in the match, alongside statistics as points, assists and rebounds, as well the total minutes played. Finally, the dataset teams.xlsx contains information from all the teams, as its city, its foundation year, its general manager, etc, which played an important role on the visualizations delivered on the dashboard.

NBA official website provides free access to all this information in their stats website, however, this Kaggle repository is continuously collecting and slightly cleaning the data to make it more feasible and easier to use.

All data sources follow the requests stated in the project guidelines. They represent transactional records, with files of csv and xlsx type, with quantitative variables and a regular pattern across time, making them easy to understand and to operate on. After a thorough analysis, it was found that it is possible to create a dimensional model with five dimensions, which can be found in chapter 6.

# Perspectives of Analysis

In order to perform a Business Intelligence analysis on the NBA teams, it was decided that a focused analysis would be more effective. As such, two business perspectives were chosen to be analysed throughout this project. The chosen perspectives were the Operational and Competitive viewpoints of the NBA teams.

## 5.1. Operational Analysis

According to Bridgman, an Operations Analysis is the study of operational systems with the purpose of recognizing opportunities for improvement[[7]](#footnote-8). As such, regarding NBA teams, an Operational Analysis would involve the analysis of the variables through which the team’s performance is measured. Thus, through an analysis of these variables, the teams would be capable to identify potential strengths and weaknesses, and therefore find prospective threats and opportunities.

As basketball teams are the companies under analysis, the variables that were chosen through which the teams can be analysed were:

* Points – which can be accumulated by either making field goals (two points if it is made within the three-point line, or three points if it is made beyond the three-point line) or free throws (one point).
* Rebounds - it is the statistic granted to the player who retrieves the ball after a missed field goal or free throw.
* Assists – statistic awarded to a player who passes the ball to a teammate, leading to the teammate scoring by field goal.
* Injuries - Professional NBA athletes NBA experience a high rate of game-related injuries. This measure has the intent to see if certain players consistently have higher injury rates.
* Games played – This measure has the objective to measure if the team is playing above or below average amount of games.
* Minutes played – This measure has the goal of ascertaining if the team is playing effectively, that is, if more minutes played lead to more points per team.

## 5.2. Competitive Analysis

On the other hand, a competitive analysis is a critical part of any company’s strategy, by assessing the strengths and weaknesses of the competitors, thus providing an offensive and defensive strategy for the company at hand.[[8]](#footnote-9) Given that the company using this analysis is any NBA team, the corresponding competitors will be the remaining NBA teams. As such, the remaining companies will be analysed in variables through which they can be assessed.

The variables chosen to analyse the competitor teams’ strengths and weaknesses were:

* Number of games played – this variable will be used in order to find out whether the competitor teams are playing on average more or less games than the team using the analysis.
* Win% - Verifies which teams have higher win percentages and those who have lower, in order to access the threats for future games with competitors.
* Injury incidence - This measure has the intent to gauge whether certain teams have practices that constantly lead to higher injury rates.

# Dimensional Model

The methodology presented for dimensional modelling follows the approaches that the group assumed were best for the quick, relevant, and powerful retrieval of data, through querying. Given the necessities of the stakeholders, the dimensions that surround the facts should provide the most complete contextualization for the analysis desired. Because the objective of analysis is the deeper comprehension of a team and player’s performance (in terms of point scoring), the dimensions created should help explain and provide a narrative to those metrics.

In the common dimensional modelling for a company, usually, the Fact are Sales. Here, the unitary subjects of the decision analysis are the Points. This represents a Snapshot Fact as it will contain a combination of all dimensions in the model at specific time intervals.

The objective of the dashboard will be to explain and analyse performance through the points scored, according to 5 dimensions: Game, Team, Player, Location and Date. The ending output will be an analysis of the points scored by a specific player, on a specific game, when he was playing for a certain team, by date and also by location. Game answers the question of What, Player answers the question of Who, Location the question of Where, Date the question When and Team contextualizes the Point scoring on the account of the team the player belongs to.

Other measures that can be aggregated are the minutes played by the player in question, the number of rebounds, a total of ball interactions, and point related ratios.

This way, a particular team, according to its own operational and competitive objectives of analysis, can understand a player’s performance per game, per team, per location where the game is held and per date. An understanding of the Points scored can drive teams, the NBA, and external stakeholders to make informed business decisions regarding performance, salaries, sponsorships, among others.

**Dimension Game:** This dimension has no hierarchy but provides relevant information for any deeper analysis. The objective is to describe the fact table according to the kind of game that occurred. For this reason, Dim Game contains, firstly, the Game\_id as its primary key and the two teams that participated (home and away). Initially, this dimension included Date and City, but those attributes are already in the Fact Table. One suggestion given was the removal of Team Home and Team Away, making them descriptive attributes. The group believed these two attributes are relevant in the context of dimension game, given that, at any given game, the fact table can’t inform the user whether the player’s team was playing Home or Away, simply stating the team he was playing for.

**Dimension Team:** Team contains all the teams playing in the NBA, uniquely identified, with a specific 4-level hierarchy: The Team Name, City from where it’s from, Division and Conference[[9]](#footnote-10) . In addition, the dimension also provides information on the arena capacity, general manager, head coach (current), owner and year founded.

**Dimension Player:** Being one of the central dimensions that support the fact table, Player describes the name of the player at hand that scored the points in analysis, along with a measure of his injuries and cluster label he belongs. In this case, injuries could be defined as a slowly changing dimension, meaning that the number of injuries a player suffers changes over time, requiring update. There are 7 types of ways of dealing with these slowly changing dimensions and the method initially used was Type 2. Before, Type 4 had been used, which adds a history table associated with the dimension in question. Dimension Player would always have updated information of its players (current total injuries), but Player\_History would contain all past injury information, along with a new variable ‘Create\_Date’ indicating when the injury value was changed. Type 2 was considered better, as it kept the Star Schema formation, adding a ‘Version’ variable to the Player table. Type 1 and Type 3 could not be applied as the storing of historical data is important and also because Type 3 can only update its values one time. The downside is that this ends up creating several primary keys for the same player, according to the number of injuries he has over time. To simplify the data, Injuries was turned into a global Injury per season ratio, which took the total value of injuries a player had suffered throughout his career, divided by the number of seasons he played. This way, Injuries becomes a fixed numeric measure over time, disregarding the slowly changing dimension specifications.

**Dimension Location:** Location is meant to describe where the game took place, where a certain number of points were scored. This dimension contains another 4-level hierarchy, starting on: The Location of the Stadium, the City, and State of the U.S.A.

**Dimension Date:** The source of this dimension is the csv file ‘games\_dates’. It allows the user of the dashboard to pinpoint temporally when the point scoring took place, representing the final dimensional hierarchy.

The type of data warehousing schema applied was a Star Schema, which is easier to understand (less complex), scalable and easily changeable.

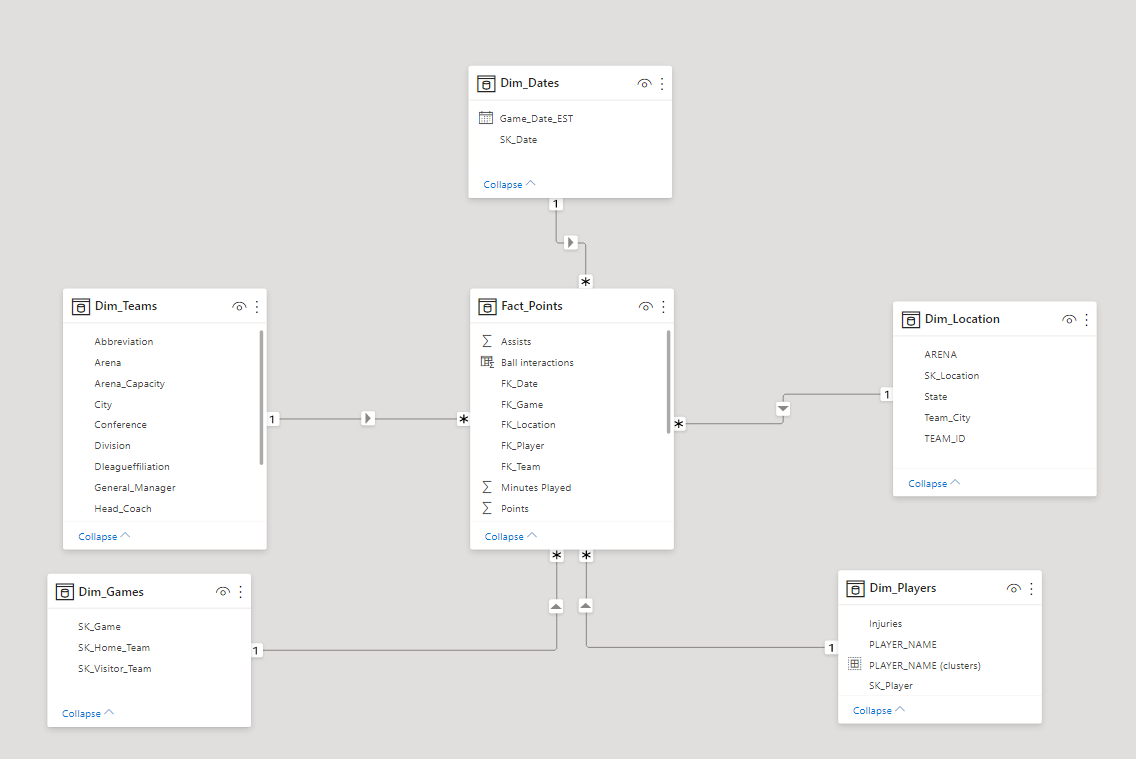


Fig.1- Star Schema Dimensional Model

# Data Integration, Transformation and Modelling

The two existent data sources, NBA Games Data (5 datasets) and NBA Injuries, required some processing for the ease of their integration into Power BI. For example, the data within the 6 datasets did not completely match, and even inside NBA Games Data some inconsistencies were found. To accomplish the objective of transforming the data into the dimensional model defined above, the datasets had to be separated into the 5 dimensions.

The first processing step into constructing the Fact\_Table started by the realization that it would be easier to build upon the dataset ‘games\_details.csv’, as it contained already most of the information that would be present in the Fact\_Table.

The initial transformation occurred when transforming Dim Game, as it had more unique Games than the ones present within the Fact Table. To reassure this consistency, a merge was performed by inner joining the tables on Game\_ID. Regarding Dim Team, each Team’s Division and State was manually introduced to the dataset, as it previously did not contain this information, but no merging had to be done. Regarding Dim Location, manually, an ID was assigned to each location and later joined back to the Fact\_Table by performing an inner join on ‘Team\_ID’. This ‘Team\_ID’ was later dropped, as its only purpose was to connect Dim Location to the Fact Table (thus allowing the Fact Table to contain a Location\_ID).

For the Dimension Date, a new manual index was created to define each Date\_ID, which was added back to the Fact Table through Game\_ID.

As was previously mentioned, instead of having Dim Player be a slowly changing dimension of type 2, with several ‘Versions’ for each Player depending on his number of injuries throughout the seasons, the group decided it was best to create an Injury per Season ratio which is fixed over time, by taking the total number of injuries a Player had and dividing it per total number of seasons played. This value can then be relevant and evaluated independently of the time selection the user defines, inside the dashboard.

Consequently, each variable on each dimension had to be assigned the proper type, and its names had to change. In particular, every primary key was labelled *“SK\_*” (surrogate key) and foreign keys in the Fact Table as “*FK\_*”. Rows referring to null values were removed and the variable referring to ‘Minutes Played’, in the Fact Table, had to be transformed into an integer, after being split by a colon delimiter.

# Description of PowerBI Report

The final PowerBI report starts with a Main Menu, where the three main parts that it contains can be accessed via page navigation, by clicking on each respective image. The three main components are: an **Overview**- corresponding to the **Operational Perspective of Analysis**, through a global outlook on each Team’s performance attributes and characteristics, followed by a Player Cluster Analysis, which seeks to group all players into four categories, allowing for an easier classification of each athlete’s performance; **Competitive Analysis- Players**- the first of the two Competitive Analysis reports, used to compare metrics between a player in analysis and any competitor; **Competitive Analysis- Teams**- following the logic of the previous report page, permitting comparisons between a Team’s main KPIs and how it measures up to the competition.

**Overview**

In the upper portion of the Overview, three cards are presented to the viewer, which inform the user on the global characteristics of the report. These include the total number of teams- 30, 1745 different players and 24.05 thousand games. Each of these is done by simply presenting in a card the Count of Team Nickname, Count Distinct of Player Name and Games foreign key.

Below, the Operational Perspective on a Team’s global performance is presented, starting with a dropdown, allowing the user to define which Team he seeks to explore. This filter will consequently affect all visualizations surrounding it, until the Player Cluster Analysis. On the right, the team location is displayed on the map, with the size of the bubbles referring to the total points that team has scored over the course of all available seasons (2003 to 2020). The Lakers showcase a significant larger difference in points scored over all other teams, judging by their size comparison.

Below the dropdown, a bar chart presents the user with the total players that have played for the chosen Team, in the available time range, sorting them according to their ball interactions. Ball Interactions are made up of Points, Assists and Rebounds, the three most important player metrics. This is a new calculated measure, being the result of a sum of those three metrics, which were in the Fact Table. Further explanation on this can be observed by hovering the cursor over the graph, where a tooltip defines the proportions of singular ball interactions that make up the global metric. [[10]](#footnote-11)

On the right, these ball interaction metrics can be compared to a previous year, for each team or even for a specific player. On top, a slicer filters through the available years in the Dimension Date, with a KPI below indicating the variation between the most recent year chosen and its previous one.[[11]](#footnote-12)   
The ball interactions defined as the target were a calculated measure named “LY Interactions Amount”, which utilizes the *SAMEPERIODLASTYEAR* function on the Interactions Amount measure. Interactions Amount is the summation of all ball interactions.

Following this narrative line, the user might be interested in going beyond analysing past performance, and as such a forecast is presented, below and on the left. Like all previous visualizations, this forecast can be applied to both teams and players, attempting to predict how many points they might score in the next two seasons. For this reason, the length of the forecast is two, with a confidence interval of 75%, represented by the upper and lower bounds of the forecast, in yellow. Higher confidence intervals result in wider bounds, which can be a bit less informative, even though more accurate. To achieve this, the visualization uses Points Amount, a measure that sums all Points throughout the Fact\_Table, which can be useful when filtered through Teams or Players. On its x-axis, the Year (present within the Date Hierarchy) is used.

To the right of the forecast, the total minutes played (sum of Minutes Played), segmented by player, is shown as a donut chart.[[12]](#footnote-13)

Lastly a clustering analysis of the players was developed through the creation of a new measure termed PLAYER\_NAME (clusters). This measure was developed through the variables Points, Assists and Rebounds which can be found in the Fact Table. This generated four different clusters which were renamed according to the type of player each cluster was composed of. The interpretation of the clusters was performed through a table where the number of players, as well as the averages for Points scored, Assists and Rebounds performed were provided for each cluster. Thus, the final clusters obtained were “Greatest of All Time”, “Very Good Players”, “Average Players”, and “Weak Players”.

Furthermore, in the clustering analysis section, two types of filters were employed only affecting this area. The first filter allows for the selection of a team from a drop-down list of all of the teams’ names, while the second filter permits the choice of analysing only one chosen cluster at a time.

Finally, a dynamic scatter plot visualization of the clusters’ behaviour regarding Points scored in the x-axis and Assists performed in the y-axis, with the players’ markers coloured according to their respective clusters. This visualization used a play axis in order to show how each player’s performance changed over the years. Through the usage of the filters described above, it was possible to observe the performance of each individual team and/or each individual cluster.

**Competitive Analysis- Players**

The first of the two competitive analysis pages deals with Players, in an attempt to inform the user on how a particular player is performing while allowing for a quick direct comparison against another. This transcends the operational perspective of analysis, by rather placing emphasis on the relationship between an entity and its competition.

To start, two players can be selected through the dropdown of Player 1 and Player 2, and this filter is applied on all visualizations.

Below, a slicer can filter the available years, as one might be interested in comparing one player in a specific season versus another in a same or different time range. In the four cards below, a new measure was calculated and shown describing the average points per game. This measure Points per Game uses *AVERAGEX* to calculate Points Amount, specifically over the total number of distinct games.[[13]](#footnote-14)

Additionally, the total minutes played, number of teams and average injuries per season is displayed. The number of teams refers to the number of times the specific player switched and played for a different team, displayed solely using the count distinct of a team’s foreign key in the Fact Table. The Injuries per Season ratio, as has been mentioned, is fixed, thus not changing depending on the chosen time range.

Below, a KPI with a background scatterplot shows the evolution of points scored by the player, with the respective percent change from the most recent year to the year before. The measure in use is Points per Game, described above, including a new measure called LY Points per Game (which serves as the target variable for the KPI), a metric computed using SAMEPERIODLASTYEAR.

Furthermore, the end-user might be interested in understanding how changing teams affected player performance. In the scatter charts below, each player’s team is defined by a specific colour and nickname. The size of each bubble corresponds to the number of assists performed and the average points scored is on the y-axis.[[14]](#footnote-15)

Below, the violin plot solely focus on rebound numbers. When selecting all players (i.e., not selecting any specific one), a distribution of the number of rebounds of the entire NBA across the years is shown, in which is possible to see how these values vary from season to season. Moreover, it is also possible to check the mean value, median value, as well the maximum and minimum for the rebounds amount per season. This way, one can easily compare one player to the rest of the league and conclude whether or not its statistics have been outperforming the average player.

**Competitive Analysis- Teams**

The second of the two competitive analysis pages deals with Teams, so as to provide the end user with information how a particular team or set of teams are performing while allowing for a direct comparison against another. Once again, this also transcends the operational perspective of analysis by placing emphasis on the relationship between an entity and its competitors.

At the top of the page, a filter displayed in horizontal buttons is present where one or multiple teams can be selected by their team names, and this filter is applied on all visualizations and information displayed on this page.

Below the team’s name filter, there is a table which displays information regarding the selected team or teams. The information displayed on the table is the Team’s name (Nickname), the Arena of the team (Arena), the team’s city (Team\_City) and the corresponding state (State), the last recorded Coach, obtained by selecting the Last option on the Head\_Coach variable, the team’s corresponding Conference (Conference) and finally the Arena capacity (Arena\_Capacity).

Beside the table dispaying the teams’ information there is a map visualization which shows us the cities where there are teams, it the size of the markers representing the number of points of the team (or teams) in the city. When a specific team is selected, the map visualization will amplify in the selected team’s city. This visualization uses the measures Team\_City and Points.

Below the previously described table there are four visualization. The first visualization is a line plot of the points of each team, throughout the years, where the years are the x-axis, using the measures Date hierarchy of Year, Points, and Abbreviation (for the team abbreviation). The following visualization is a scatter plot correlating the Arena Size of each team with the Points scored per team, so as to ascertain if there is any connection between the two measures, using the measures Arena\_Capacity in the x-axis, the Points in the y-axis and the team Nickname as the legend. This page also includes a KPI which compares the points of the current year with the ones of the previous year for the selected team, with the goal of surpassing the points scored in the previous year. In order t achieve the KPI visualization a new DAX time measure was created, LY Points Amount, for the amount of points scored the previous year. Finally, an horizontal bar plot was created in order to access which players have the most Ball Interaction for the selected team, using the measures Ball Interactions and PLAYER\_NAME.

# Analysis and Discussion

The final PowerBI dashboard allows the user to meet and solve several needs that one might face during an NBA season. In fact, all the answers to the Business Needs stated on chapter 3 can be answered. Assuming one’s team is Golden State Warriors (GSW), and assuming also that it will face Los Angeles Lakers in the upcoming match, these answers might provide valuable insights for the staff.

The dashboard allows to check that the most influent players on GSW during the 2020 season are Stephen Curry, Andrew Wiggins and Damion Lee, accounting together for 30% of total minutes played. Stephen Curry is the most influent player with an average of 37 points per game in the 2020 season. Nevertheless, he has an average of around 5 injuries per season, which might bring some instability to team’s momentum in case he gets injured.

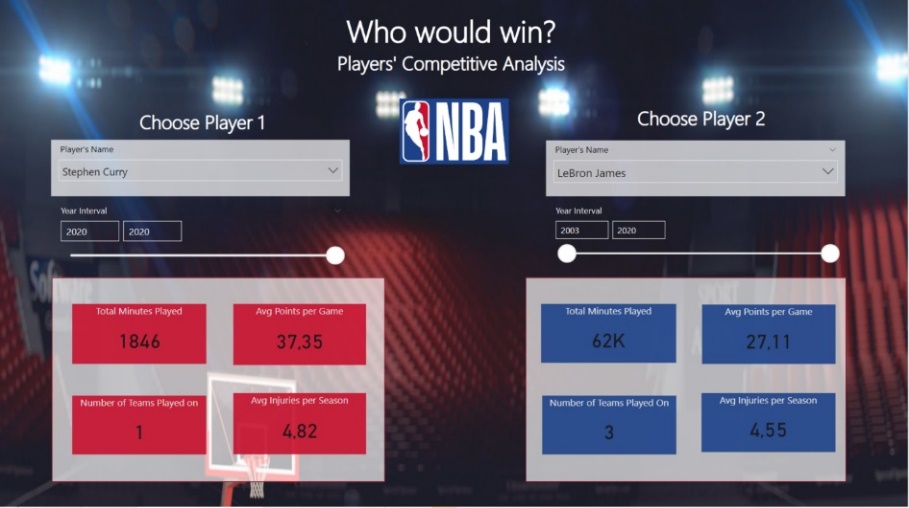


Figure 2 – Player Competitive analysis on NBA Dashboard

When facing Los Angeles Lakers, special attention must be given to players like LeBron James, who has consistently scored on average above 25 point per game in the last 18 years. Also, Anthony Davis had above 700 Rebounds per season in 6 of the last 7 seasons. These statistics should influence the team’s strategy to mitigate its own weaknesses when facing other teams.

Uma imagem com texto, captura de ecrã, armário

Descrição gerada automaticamente

Figure 3 – Cluster distribution on Player Competitive analysis on NBA Dashboard in Overview (Operational Perspective)

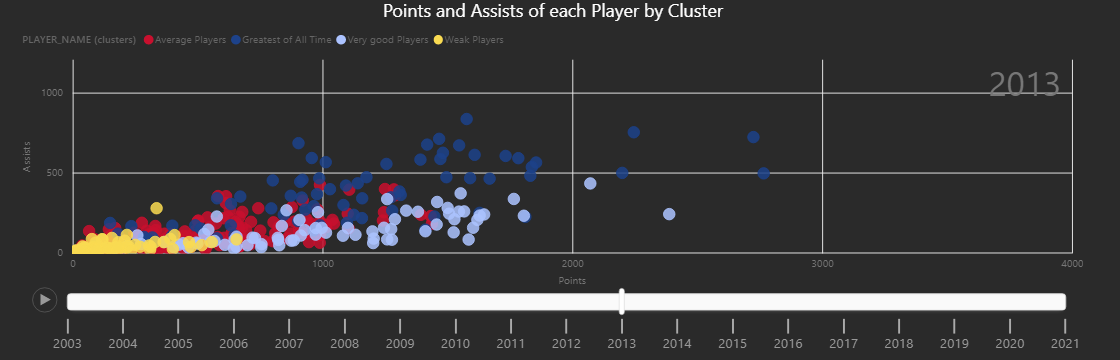


Figure 4 – Playable Cluster Visualization of Points and Assists per player throughout the years, in Overview (Operational Perspective)

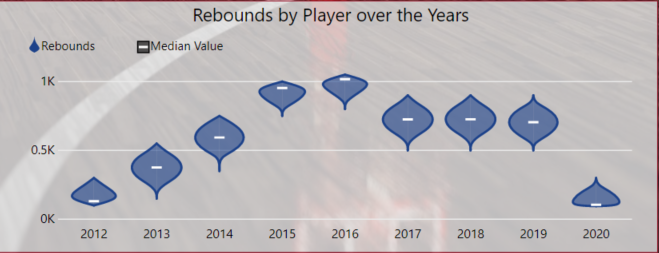
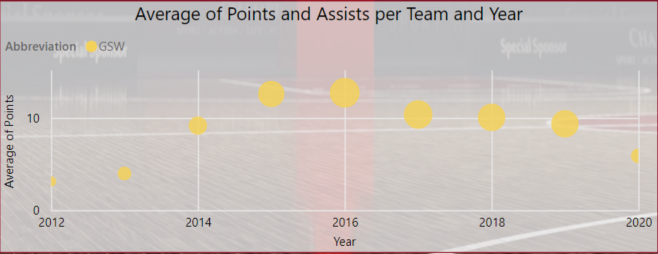
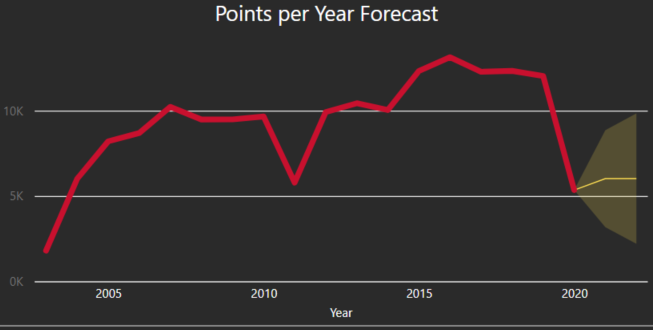
At the end of each season, there is always the need to prepare the upcoming season. Therefore, it is crucial for a staff to know which type of players the team is mainly composed of. Other than that, when looking for a specific type of player to trade, the staff should look for similar players on other teams. Thus, a cluster analysis allows the user to see which type of player a team has and how many per each type. This eases the trading and the draft process and allows also to study the opponent for future games, once the staff adapts its strategy according to a team having a lot of defenders or a lot of all-stars (here stated as very good players and greatest of all time).

Figure 5 – Visualization in Player Competitive Analysis of Average Points and Assists throughout the years.

Figure 6 - Visualization in Player Competitive Analysis Rebounds throughout the years.

An individual and detailed training program can be achieved when considering each player’s stage of career. Considering the database available, the average number of seasons per player playing at NBA is around 4 seasons. Therefore, players like Draymond Green (GSW), should be doing detailed training plans. As can be seen in figures X and Y, the player appears to be decreasing its numbers, seeming that his top seasons were 2015 and 2016.

Uma imagem com texto

Descrição gerada automaticamente

Figure 7 – Current year points compared to previous year KPI in Team Competitive Analysis

Figure 8 – Points per year forecast per team in Team Competitive Analysis

Moreover, an analysis can be performed looking to the Team as a whole. GSW have been outperforming its competitors for the past 7 years, but in 2020 there was a considerable decrease on the total of points scored by the team. In fact, the total of points was less than half of the points from 2019, which can also be explained by the fact that, in 2019, GSW won the championship and, therefore, played a considerable higher number of games. It is then expected that the team gets back to better results, forecasting a total of six thousand points for 2020 season.

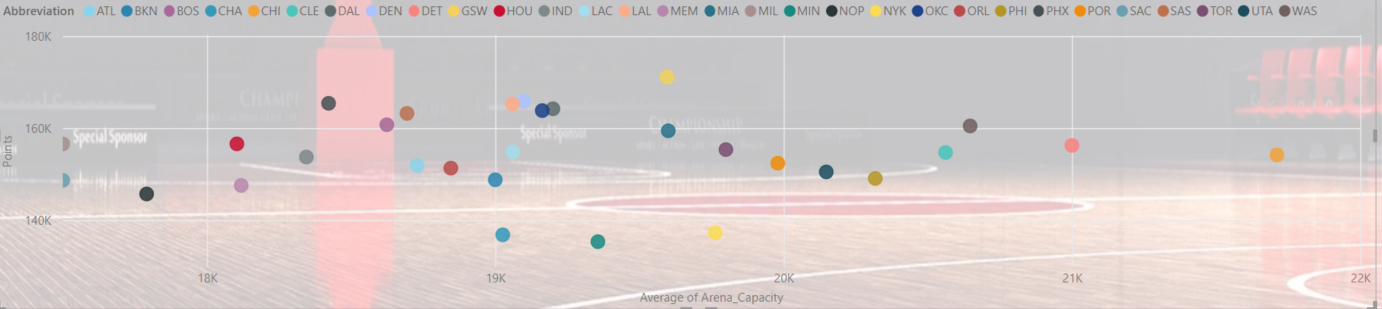


Figure 9 – Correlation Teams’ Points and Arena sizes in Team Competitive Analysis

In the case you are an investor or a sponsor looking for a team to invest and want to consider teams that have more supporters at their arenas, you should definitely be given attention to teams like Chicago Bulls (Capacity of around 22 thousand supporters), Detroit Pistons (21 thousand) and Washington Wizards (20,6 thousand). If you want to combine the arena capacity with adrenaline, i.e., more points scored, then the outlier you are looking for is Golden State Warriors, with an arena capacity of 19,6 thousand supporters and having a total of 170 thousand points scored in the last 18 years, overcoming its competitors for the past 7 years.

# Critical Assessment

Although the group considered the achieved results of this project quite good, a better dataset could have led to perform even better and deeper analysis. A dataset that would include more statistics per game would make possible to see how players perform during the season and not only on the whole season. NBA is a competition divided between regular season and playoffs, so it would have been interesting to see the difference in the player’s statistics when having more important games.

Also, it would have been interesting to see which players perform better against certain teams or against teams that they have played in for several years - for instance, see LeBron James’ performance against Cleveland Cavaliers and spot whether his performances tend to decrease/increase when comparing to games against any other team.

Some other analysis could have been made, if there was some information regarding the teams playing home and away. This way, stronger teams playing home with their supporters could have been differently highlighted in the dashboard. This would make even more accurate visualizations like the correlation between the arena capacity and the total of points scored, which aimed for the points from the whole season, instead of the home points.

Finally, more information regarding the players itself (Age, Height, etc) and more indicators on their performances could led to deeper analysis. For instance, having access to statistics like Post-Ups, Isolation, Hand-off, among others, would allow to characterize the players and create a more powerful clustering analysis, with clusters that would distinguish players based on their quality and position on the field, for instance.

# Conclusion

This report is primarily a preparation for a more complete and detailed report which will focus on data integration and creation of a Power Bi dashboard, formerly, it will culminate on a powerful analysis and discussion alongside a critical assessment. An initial assessment of the data to be used was performed, to make sure this one is adequate to do a business intelligence analysis.

Thus, the current report focused on critical aspects such as the data sources, which will allow a further analysis on two perspectives: Operational Analysis, based on the study of the variables that impact a teams’ performance; Competitive analysis, which, in the mainstream business language, would be the market competitive analysis but, instead of companies, the focus relies on NBA teams.

Finally, as previously stated on chapter 6, a Star Schema was used for data warehousing, which is typically easier to understand and perform this type of analysis.

# Extra – Work

This chapter, as stated in the Introduction, aims to simplify the process of identifying which of the data processes is considered to be extra work. As such, these were some extra steps:

* Dynamic scatter plot visualization of the clusters’ behaviour regarding points and assists, through a play axis in order to show how each player’s and cluster’s performance changed over the years.
* Violin Plot – used for describing total Rebounds per player in each season. Uploaded from the marketplace.
* Horizontal Bar Chart – used for differentiating players of each team by total of ball interactions. Uploaded from marketplace.
* KPI – It was defined 1 KPI for each perspective, being that competitive analysis (Players and Teams) had a total of 3 KPIs (2 being extra).
* Dax Measures Time Intelligence – there are a total of 4-time intelligence measures, 2 extra.

# References

Bridgman, P. (1938). Operational Analysis. Philosophy of Science, 5(2), 114-131. Retrieved March 26, 2021, from <http://www.jstor.org/stable/184626>

1. Cares.nba.com. 2021. NBA Cares. [online] Available at: <https://cares.nba.com/> [Accessed 23 March 2021].

Entrepreneur. 2021. Competitive Analysis Definition - Entrepreneur Small Business Encyclopedia. [online] Available at: <https://www.entrepreneur.com/encyclopedia/competitive-analysis> [Accessed 23 March 2021].

1. Gough, C., 2021. Average sports salaries by league 2019/20 | Statista. [online] Statista. Available at: <https://www.statista.com/statistics/675120/average-sports-salaries-by-league/> [Accessed 21 March 2021].

Nba.com. 2020. NBA rosters feature 107 international players from 41 countries. [online] Available at: <https://www.nba.com/news/nba-rosters-feature-107-international-players-from-41-countries> [Accessed 22 March 2021].

1. NBA Stats. 2021. NBA Stats. [online] Available at: <https://www.nba.com/stats/> [Accessed 23 March 2021].

Reiff, N., 2020. How The NBA Makes Money: The Second-Largest Sport in the Country. [online] Investopedia. Available at: <https://www.investopedia.com/articles/personal-finance/071415/how-nba-makes-money.asp> [Accessed 22 March 2021].

Voices.nba.com. 2021. NBA Voices – NBA Voices. [online] Available at: <https://voices.nba.com/> [Accessed 23 March 2021].

1. Gough, C., 2021. *Average sports salaries by league 2019/20 | Statista*. [online] Statista. Available at: <https://www.statista.com/statistics/675120/average-sports-salaries-by-league/> [Accessed 21 March 2021]. [↑](#footnote-ref-2)
2. Reiff, N., 2020. *How The NBA Makes Money: The Second-Largest Sport in the Country*. [online] Investopedia. Available at: <https://www.investopedia.com/articles/personal-finance/071415/how-nba-makes-money.asp> [Accessed 22 March 2021]. [↑](#footnote-ref-3)
3. Nba.com. 2020. *NBA rosters feature 107 international players from 41 countries*. [online] Available at: <https://www.nba.com/news/nba-rosters-feature-107-international-players-from-41-countries> [Accessed 22 March 2021]. [↑](#footnote-ref-4)
4. NBA Stats. 2021. *NBA Stats*. [online] Available at: <https://www.nba.com/stats/> [Accessed 23 March 2021]. [↑](#footnote-ref-5)
5. Cares.nba.com. 2021. *NBA Cares*. [online] Available at: <https://cares.nba.com/> [Accessed 23 March 2021]. [↑](#footnote-ref-6)
6. Voices.nba.com. 2021. *NBA Voices – NBA Voices*. [online] Available at: <https://voices.nba.com/> [Accessed 23 March 2021]. [↑](#footnote-ref-7)
7. Bridgman, P. (1938). Operational Analysis. *Philosophy of Science,* *5*(2), 114-131. Retrieved March 26, 2021, from http://www.jstor.org/stable/184626 [↑](#footnote-ref-8)
8. Entrepreneur. 2021. *Competitive Analysis Definition - Entrepreneur Small Business Encyclopedia*. [online] Available at: <https://www.entrepreneur.com/encyclopedia/competitive-analysis> [Accessed 23 March 2021]. [↑](#footnote-ref-9)
9. There are two Conferences in the NBA: Western Conference and Eastern Conference, both of which can be divided in 3 divisions: Atlantic, Southeast, Central for East and Southwest, Northwest and Pacific for West. [↑](#footnote-ref-10)
10. For example, **choosing the Philadelphia 76ers** and hovering over Andre Iguodala, one can quickly see that his total ball interactions are made up of 58,8% Points he scored, 18,65% his number of assists and 22,55% Rebounds. These values can be contextualized through the indication of the total minutes he played, given the reasonable assumption that, the more minutes a player has been on the pitch, the higher his total ball interactions will be. [↑](#footnote-ref-11)
11. Example: Having scored 18.087 points in 2019, the 76ers decreased sharply their total points scored in 2020, through a reduction of 36,36% to 11.510 points [↑](#footnote-ref-12)
12. For the 76ers, the player that is more distinguished in seniority is Andre Iguodala, despite not having played for the team since 2012. One interesting conclusion to be immediately retained is that the player named Joel Embiid, third on the ball interactions chart, has played considerably less minutes than the top 3 most senior players by total minutes played, an indication of talent and high-quality performance, by having scored more with less minutes on the pitch. [↑](#footnote-ref-13)
13. For example, if James Harden is sliced for the time interval 2019-2020, we can see he scored an average of 36.99 points per game, an extremely positive result. By comparison, Carmelo scored an average of 15.90 in that same season. This disparity could only be seen by filtering through the slicer, as initially, considering the entire time interval, both players’ averages of points per game seemed relatively similar. [↑](#footnote-ref-14)
14. Example: James Harden seems to be on a growing part of his career, possibly reaching his prime, very evidently improving after moving to the Houston Rockets from the Oklahoma City Thunder, in 2012. On the opposite side, Carmelo has been on a downward path of both points and assists, having performed at the highest level in the earlier stages of his career, under the Denver Nuggets, slowly decreasing until 2020, where he now plays for the Portland Trailblazers. [↑](#footnote-ref-15)