

# NBAll'd - NBA All Data Navigator

## *Project Report*

Frederico Cardoso - 2012138904  
Eduardo Neves - 2018281324

May 17, 2023

## 1 Motivation

The National Basketball Association (NBA) is one of the most lucrative competitions in the world, generating an annual revenue of more than 10 billion dollars[1]. Basketball is, currently, one of the most watched sports of all-time and NBA's indisputable fame and media coverage around the globe, in terms of basketball competitions, is yet to be topped.

As one of the oldest professional leagues of basketball, being around since its foundation, in 1946, there is a vast amount of data collected through this time window. Our proposed work is to research and implement **advanced techniques of visualization** of this information, focusing on the last 30 years.

## 2 Related Work

Other works on behalf of NBA statistics are mainly focused on players. The official NBA site provides data between players and teams, but the focus is mostly on tabular, as similar to other analysis platforms. As our motivation went, our work tried to do something different, to the interest of not only from the specialists but also the fans, by having a different perspective with some advanced visualization as well as some alternative algorithms to enhance the quality of the whole product.

Some related work is worth mentioning:

- The NBADataAnalysis-Visualization[2] repository explained in this Medium article. Based on Python, the author builds a stat scrapper with focus on player statistics. The outcome in terms of visualization remains on the basis of simple bar charts.
- The NBAstatsVis[3] repository gives a visualization resolution for players stats until the 15/16 season. This work tackles specially the scoring of each athlete playing that last season and some statistics around it, like location and frequency.

### 3 Data

For this proposed work, it was used the **NBA Dataset**[4], a daily updated *SQLite* database that comprises information about more than 64,000 games, 4800 players, and all the 30 Teams in the NBA, since its foundation.

Intrinsically, this repository has a very extensive record of different classes of data, ranging from details about players, draft history, games, officials and teams. For the scope of this project, a lot of data preparation steps were needed, as most of these classes will not be needed. We planned to explore the following:

- **draft\_history:** data about the draft system through the years.
- **game:** data about all games.
- **player:** data about all players.
- **player\_game\_logs:** data about all games by player.
- **team:** basic data about all current teams.
- **team\_details:** detailed data about all current teams.
- **team\_info\_common:** data about points, ranking, etc. to use as a reference to construct our own.

### 4 Exploratory Data Analysis (EDA)

The analysis done in this step was aimed to improve the files and even create more useful ones for the project. Based on the .csv file given, some of them were transformed and then saved with another way of organization.

On this step it was given some attention to the years of analysis explained in the scope of this work. In that sense, some inactive players in these years were dropped, as well as some teams and old games in the list. Some tables are used to index players and teams, but have some additional columns that will not be used.

A simple description of the .csv file is done at the beginning of the Jupyter Notebook. We now summarize each analysis and processes done in each file:

#### **common\_player\_info**

This table had some missing values, mostly regarding old players. This will serve as an index table for accessing player information for a simple description of the athlete. The missing values will be ignored and, if needed for the display, will not be showed as they are not relevant for the visualization.

#### **team**

Another index table to show the 30 teams active in the NBA. Just for state, region and name purposes.

### **game**

For this we only care about games from the 81/82 season until last season (21/22). For this table there is a lot of missing values in some older games.

### **line\_score**

There are many missing values on the 'ot' (overtime) score. As these columns are dropped the table seems to have no problems and we are left with the points scored per quarter, a far more important information, and the points home and away.

### **player**

Another index table, this time for players active or not. This will be merged to the common\_player.info

### **play\_by\_play**

A description of every moment in every game from 1996. This table presents a lot of missing values, but this is normal, regarding the level of detail in the file. This happens because of home and away values and many players interactions that are specified in different columns.

Besides the EDA, the *game* and *play\_by\_play* files are used for creating other more useful tables:

### **season\_team\_stats**

This table comes from the game one and the aim is to get season stats for every team. The final table features many stats like 3-pointers, rebounds, etc. recorded in each season of the scope. This will be useful for visualization by season and comparison views between teams. Another component of our work is the player predictor, explained later on.

### **season\_player\_stats**

Similar to the previous one, this will can get even more detail, as the table of origin is the most detailed one in the dataset. We prioritize the season stats as the would be too many variables for a normal use of the visualization.

## **4.1 Details and problems of the dataset**

During this phase of the project some inconveniences that cannot be solved easily and bring some difficulties that cause a lack of completeness in the final delivery.

### **4.1.1 Missing values**

Even with a lot of 'nan' values in the various files of the dataset, most of them occur in because of the build of the set. Some problems occur when the values of the original dataset cannot be transformed and included in some important features. An example is some missing information of some athletes

Figure 1: Caption

in the 'players.csv' leaving some of them with only name and id. This becomes even more harmful when some of those players are famous ones, as happens with some legends like Kobe Bryant and so.

#### 4.1.2 Incongruous entries

Some entries of the files are not in line with other files, regarding the ids. In the clean dataset, some players in the stats per season file do not have a corresponding entry in the 'player.csv'.

#### 4.1.3 Not enough depth

As a mainly stats focused dataset, the files do not provide enough varied data to explore more on the NBA scene. In the original data, there was a lot of unnecessary information and we were left with the basic stats drawn from the games and some additional info. Some more entries like year end position each season and all the teams a player was part of would enrich the final product.

## 5 Design Requirements

We aimed to create an application that runs a simple website where the user can navigate through the different NBA teams. This user can be just a casual fan, and average follower of the league or even an NBA executive.

With the help of our system, the user should be able to:

- **Compare teams in last, any, or all seasons.**
- **Compare players in last, any, or all seasons.**
- **Display teams by classification table in any season.**
- **Rank players by position.**
- **Puzzle pieces:** Rank best fitted players to a team.
- **Player analysis:** relevant information about players' careers.
- **Team analysis:** relevant information about teams in given years.

The most ambitious task of our application was the prediction system, where any user should be able to see a list of players that should, on paper, be a great addiction to a certain team. Unfortunately, this caused some problems during the development phase, that were not resolved, thus not being demonstrated in the final project. In terms of analysis, the application supports a Player vs. Player (PVP) and Team vs. Team (TGT) comparison, complemented with auxiliary visual graphical representations.

## 5.1 Design choices

This simple paragraph serves to indicate the motivation behind the design choices. An NBA fan wants to see everything. He wants to be in control of what stats he sees and loose itself in the immersion of the program. So, it has to appear simple, but intuitive. The displays must be big, informative enough, but not too much. From a executive perspective, it matters more the possible comparisons and some "hidden" stats of a player, team and comparisons.

With that in mind, we created: With the help of our system, the user should be able to:

- **Color Orange:** to resemble a basketball and the combination of light orange and white is known to be easy on the eyes and inspire comfort.
- 2 **Navigate intuitively** in a system in order to search for everything he wants.
- 3 **Visualize**, in an easy way, important stats about players and teams.

# 6 Design steps and implementation

## 6.1 Mockups

The proposed application was expected to be very intuitive. In the first menu, there will be a complete map of the USA, divided by states, and a little of the frontier of Canada, to encompass the city of Toronto, that has an NBA team. At a first glance, the map will have the logos of the 30 teams, at their respective locations. These items will be interactive, to get to the Team menu, that will be explained after. There will also be a button to click, to search by player, to get to the Player menu. In this first and main menu, there will be some classification tables with the selected season, or all-time information about the top players in the league.

The Team menu is the one that comprises the key informations about the teams. There will be a comparison button, to select another team, so it can be viewed how those teams clash against each other, with information about matchups or general stats, by season, or all-time. It is expected that the application, in the current season, can suggest key players that complement the team, using an algorithm that gives an offensive and defensive score to the team and the players, to be implemented in the next weeks.

Similar to the Team menu, the player menu has practically the same information, but this time changing the subject of interest. There will also be graphs about that player, by season, or all-seasons. A offensive-defensive rating is expected, with highs and lows of the player characteristics.

## 6.2 Design implementation

After the development phase, we followed the proposed mockups rigorously, and the suggestion by the professors to include in the application and in this report

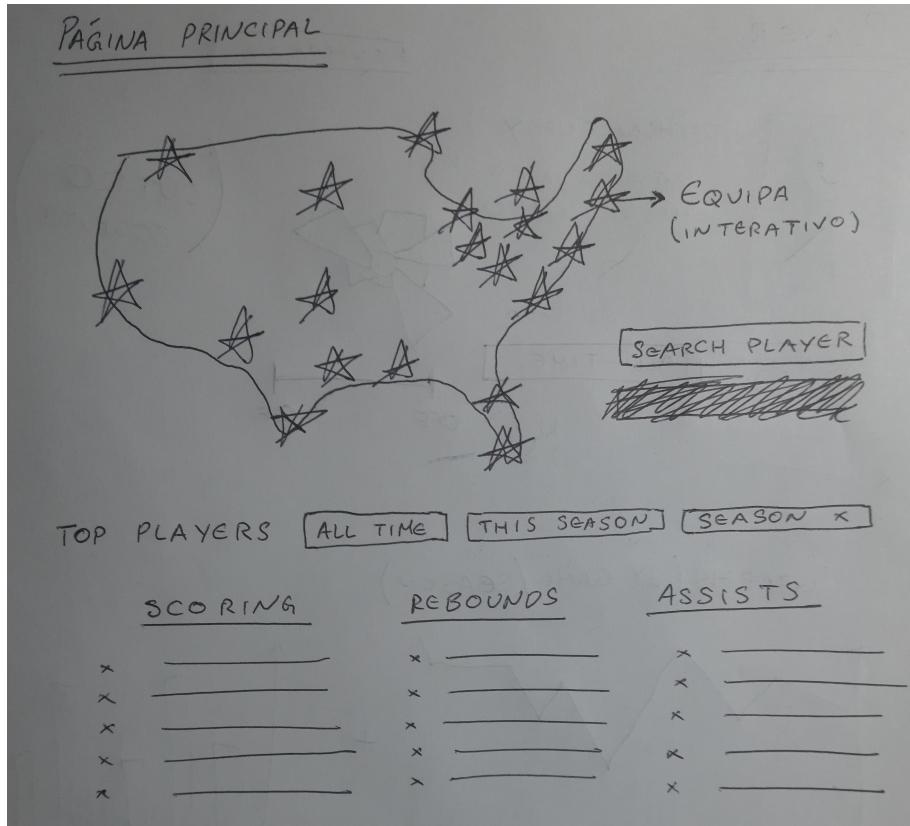


Figure 2: Initial menu mockup.

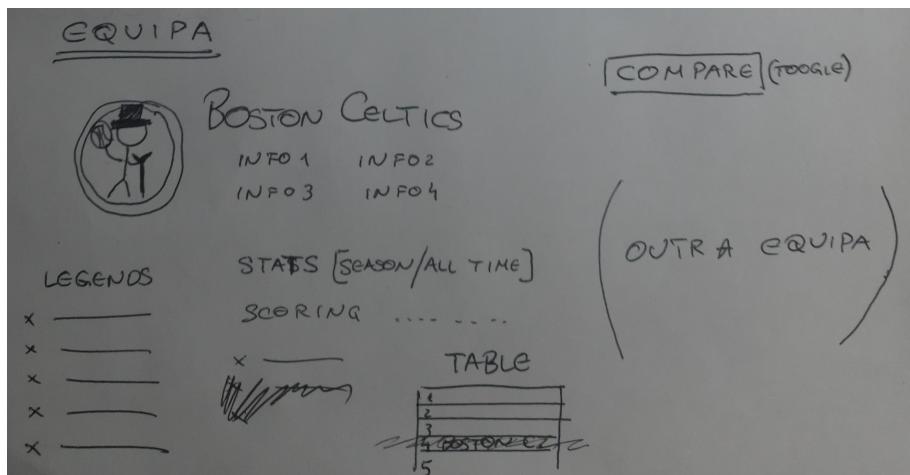


Figure 3: Team menu mockup.

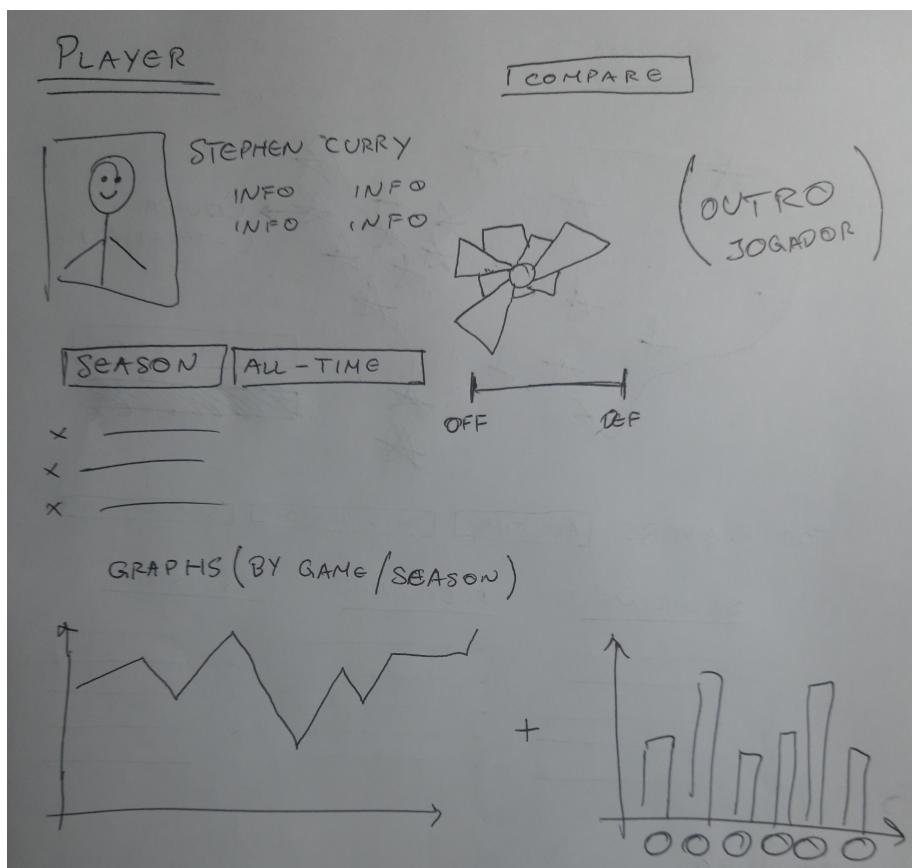


Figure 4: Player menu mockup.

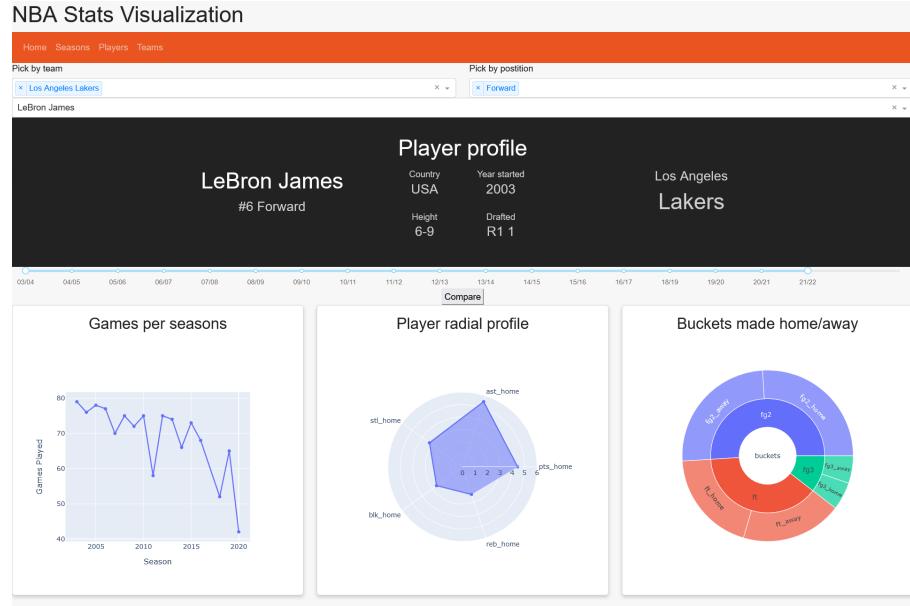


Figure 5: Players menu.

the motivation behind NBA executives and fans to use the application.

That being said, we created the following 3 different menus:

- **Players:** In this section the user can select a player by searching or by selecting the team and position on the team. Then Some stats will pop up and a box to compare is shown. When clicked, another page to select another player is prompted and a simple comparison can be made between players.
- **2 Seasons:** In the seasons page, a simple depiction of the best five teams, both in attack and defense, is shown. The page allows for a season pick range. In the end, the best players of that interval are listed, with the numbers they provided.
- **3 Teams:** This is the simplest one, as there is not that much important information about the teams, both on stats and biography. This only shows stats about all the seasons on the scope of the project.

## 7 User flow diagram

The system can be navigated and traversed in the following way:

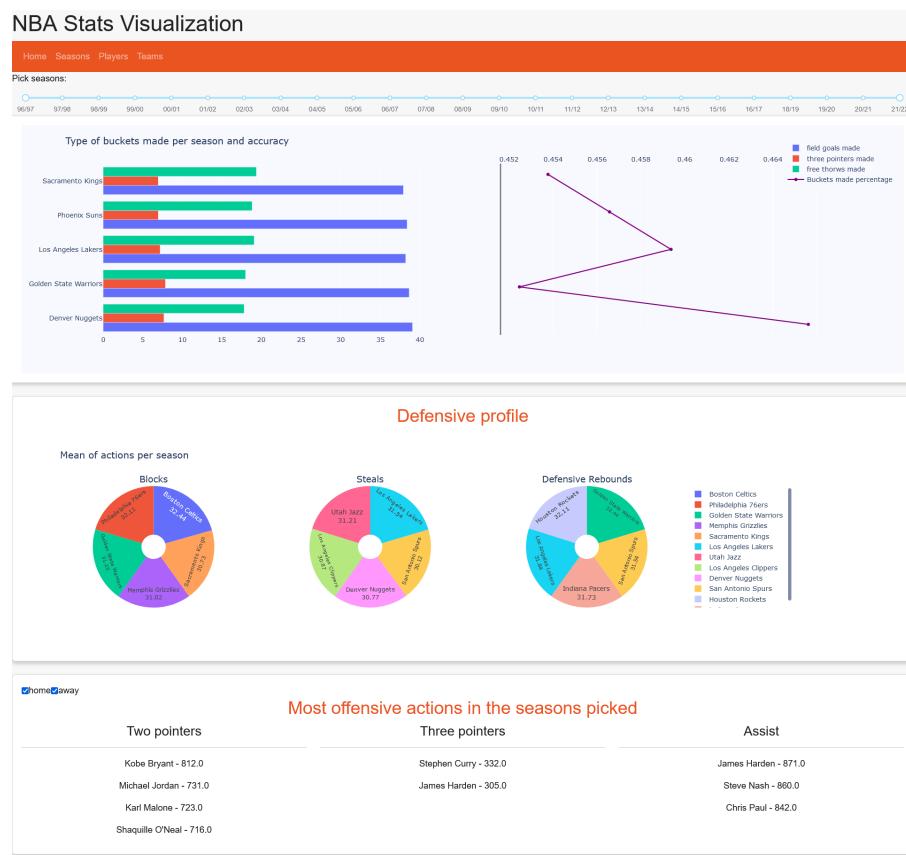


Figure 6: Seasons menu.

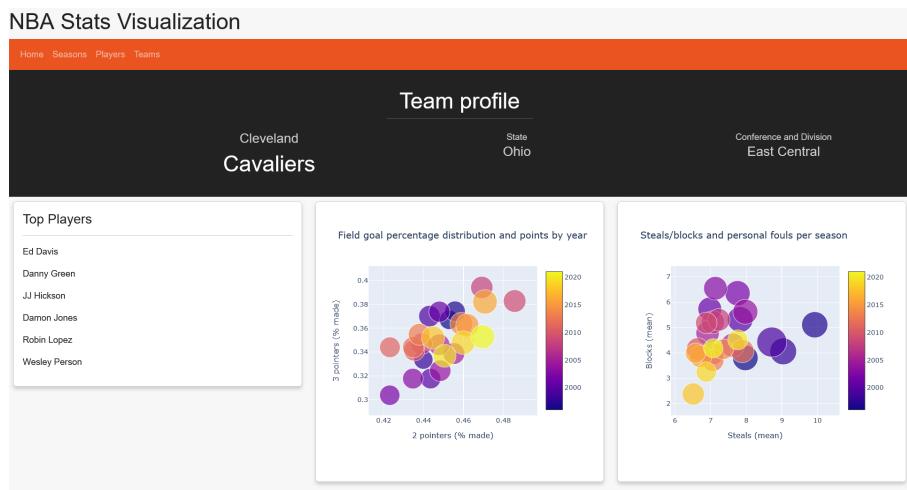


Figure 7: Teams menu.

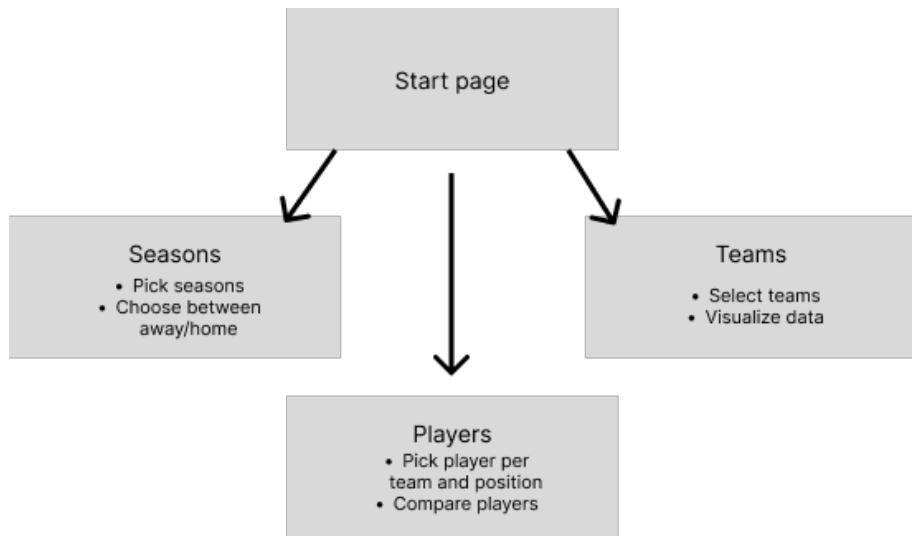


Figure 8: User flow diagram.

## 8 Validation

To validate our system, we asked 5 persons to evaluate it according to different metrics, in the defined terms of an NBA executive and an average fan:

- **A:** Is the Design appealing?
- **B:** Is the system intuitive?
- **C:** Does the system fill the needs of a user?
- **D:** Does the system have sufficient information?
- **E:** Does the system have a good display of information?
- **F:** Would you use the system in the current state?
- **G:** What about a more advanced version?

With the following results in a scale of 0 to 10:

Metric	User 1	User 2	User 3	User 4	User 5	Average
A	4	5	3	5	4	4.20
B	6	7	5	6	7	6.20
C	6	6	5	7	5	5.80
D	7	6	7	6	6	6.40
E	7	5	7	6	6	5.20
F	3	4	2	4	3	3.20
G	8	7	7	7	6	7.40
Total	5.86	5.71	5.14	5.86	5.57	5.63

As we can see in the results above, the overall score is not very good. This means our system still needs a lot of work.

We could clearly see that our strength was in the performance of user navigation and in the amount of information in the system. Besides that, the system design and its display do not provide a fruitful experience to the user. However, giving the early stage of the work, all 5 users considered that it has potential and would use it in a more advanced form.

## 9 Critical reflection

There are a lot of things to be said about the development of this project. First of all, it was needed to learn and understand a lot of new concepts of the Design area, that are a little divergent from the Data Science/Machine Learning ones, which proved a harder task than was imagined.

As for the system itself, we wish that it was more like idealized, but we ran into some development issues that delayed the consistency of the result. Therefore, the design part turned a bit lacking. However, the foundations of

the system are strongly built, which means that, if we want to take over this project in the future, it is easily upgraded.

The user experience tests showed us that this work has potential and it is interesting, beside the underwhelming current state.

## References

- [1] J. Byers, “Nba tops \$10b in revenue for first time ever,” <https://frontofficesports.com/nba-tops-10b-in-revenue-for-first-time-ever/>, 2022.
- [2] J. Leuschen, “Nbadataanalysis\_visualization,” [https://github.com/jleuschen17/NBADATAAnalysis\\_Visualization](https://github.com/jleuschen17/NBADATAAnalysis_Visualization), 2021.
- [3] Q. WU, “Nbastatsvis,” <https://github.com/wilsonCernWq/NBAstatsVIS>, 2017.
- [4] W. Walsh, “Nba database,” <https://www.kaggle.com/datasets/wyattowalsh/basketball>, 2023.