NBAll'd - NBA All Data Navigator

Project Proposal

Frederico Cardoso - *2012138904*
Eduardo Neves - *2018281324*

March 26, 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 40 years.

2 Related Work

Other works on behalf of NBA statistics is 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 goes, our work will be from the specialists perspective, and will focus on advanced visualization as well as some alternative algorithms to enhance the quality of the whole product.

Some related work 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 will be 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 are needed, as most of these classes will not be needed. We plan 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.

As this is only a project proposal, these selected classes ans methods may differ from the final ones, since we may change what we want to do or extend the application, to include other ideas that may arise in the development phase.

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 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 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 priorize the season stats as the would be too many variables for a normal use of the visualization.

5 Design Requirements

We are aiming 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 will be 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. In terms of analysis, the application should support a Player vs. Player (PVP) and Team vs. Team (TVT) comparison, complemented with auxiliary visual graphical representations.

6 Design steps and Mockups

The proposed application is 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.

7 Disclosure

At this moment, this is still, as expected, a proposal of a prototype. There may be some relevant changes to the scope of the project, to be documented in its own section.

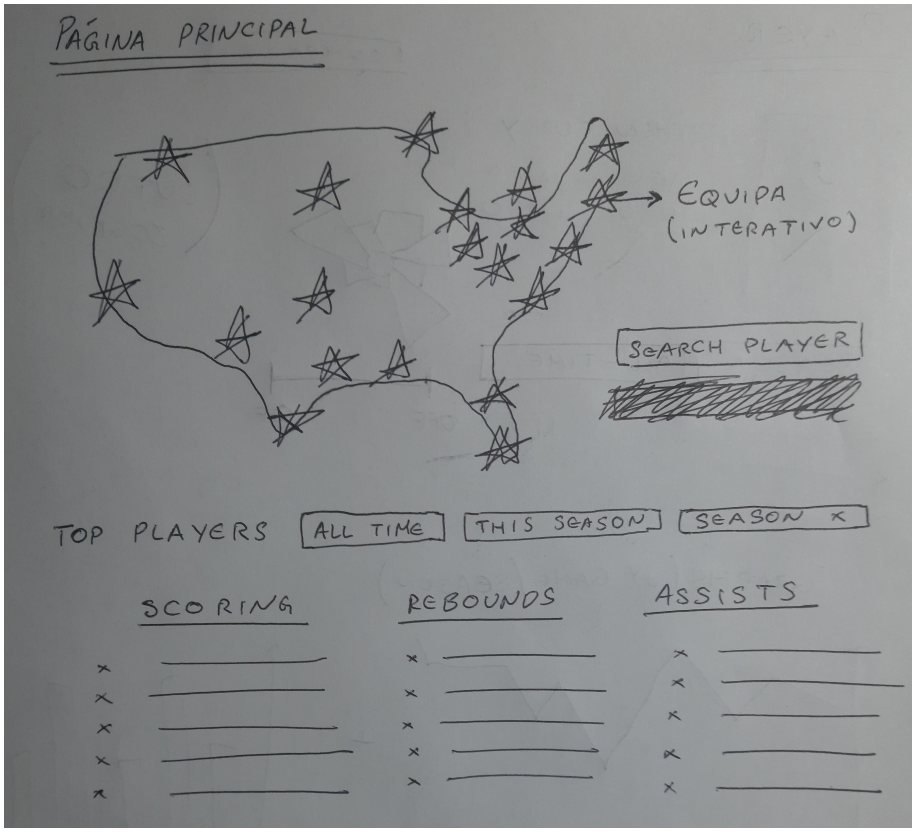

Figure 1: Initial menu mockup.

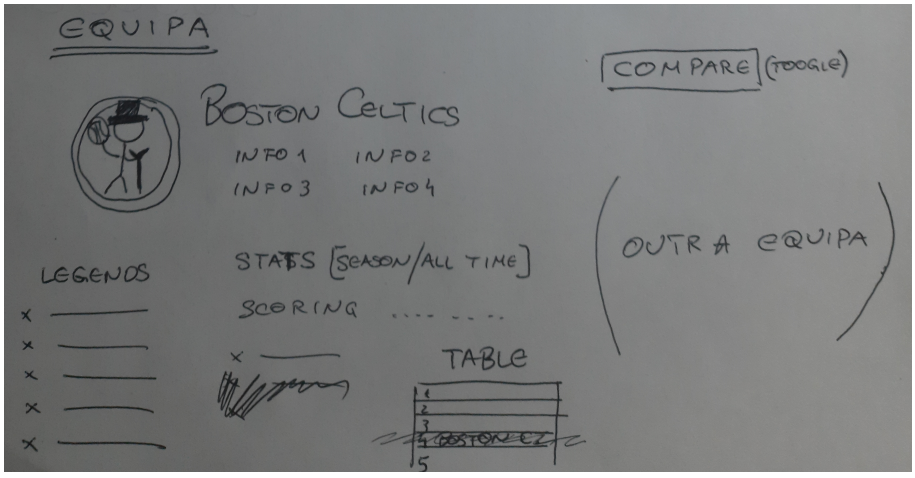

Figure 2: Team menu mockup.

Figure 3: Player menu mockup.

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, 2021.

[3] Q. WU, "Nbastatsvis," https://github.com/wilsonCernWq/NBAstatsVIS, 2017.

[4] W. Walsh, "Nba database," https://www.kaggle.com/datasets/wyattowalsh/basketball, 2023.