NBA PLAYERS PERFORMANCE INTERACTIVE VISUAL ANALYSIS

VISUAL ANALYTICS PROJECT

Carmella Sta Ana Pana
Sapienza University of Rome
Student ID: 1794019
pana.1794019@studenti.uniroma1.it

ABSTRACT

Basketball analytics, or sports analytics in general, has been a great instrument to evaluate the athletes' characteristics and performances over the time, to understand and study strategies and to optimize gameplays. The purpose of this project is to provide a user-friendly interface from which other NBA enthusiasts, but also team managers, coaches etc., can analyze and easily visualize the players' performances during the Regular Season of 2024-25.

1. INTRODUCTION

The National Basketball Association (NBA) is one of the most popular professional basketball leagues in the world and has become a global phenomenon over the decades, accumulating millions of fans and playing a significant part in sports culture.
In the world of professional basketball, data and data analysis play a fundamental role in studying and providing insights into a player's performance, team dynamics and game strategies, either based on simple or traditional statistics, or on more complex and advanced ones.
The principal goal of this project was to build a comprehensible system that shows NBA players metrics, on which users can compare multiple players and analyze their performance with simple interactive visualizations.

2. DATASET

Each year, the NBA season is divided into two main phases: the Regular Season and the Playoffs. During the Regular season, that consists typically of 82 games per team, each team competes against others across the league. In the Playoffs, the top teams from the Regular Season compete in a series of elimination rounds, leading to the coronation of the champion in the NBA finals.

For this project, I made the choice of focusing solely on the dataset from the ongoing 2024-25 Regular Season. In particular, I worked on the dataset found in *Kaggle* which collected the players' traditional statistics in the said season up to December 30th 2024. It provides a consistent and comprehensive record of players' performance across all teams. By analyzing this specific data, consistency in playing styles and delivery and in team compositions is ensured. The system therefore can focus on providing timely and applicable insights.

The original data set used contains 512 tuples with 30 columns. So, the AS index is:

$$512 \, x \, 29 \; = \; 14848$$

which fully satisfies the project requirement. Then, the data has also been integrated with other data from different sources to collect more information about the players.

The main categories of the different features present in the final datasets are:

- Game performance: Games played, Wins, Losses, Minutes played;
- Player's general attributes: Age, Height, Weight, Team, Position, Number;
- Scoring: Points, Field Goal, 3-Points, Free-Throws percentages;
- Playmaking and ball handling: Assists, Turnovers;
- Offensive and defensive metrics: Rebounds, Steals, Blocks, Personal Fouls;
- Other advanced metrics: Fantasy points, Double-doubles, Triple-doubles, Plus-minus statistic.

The dataset was mostly used as it is, with the need of only some preprocessing using Python scripts, with libraries like *panda* and *scikit-learn* to manage the data files. Some of the operations done on them were finding intersections and joining data (for example, on the *Player* column to collect different attributes of each player), data conversion (Height from inches to cm and Weight from lbs to kg) or normalizing values for the wanted analysis, and dimensionality reduction.

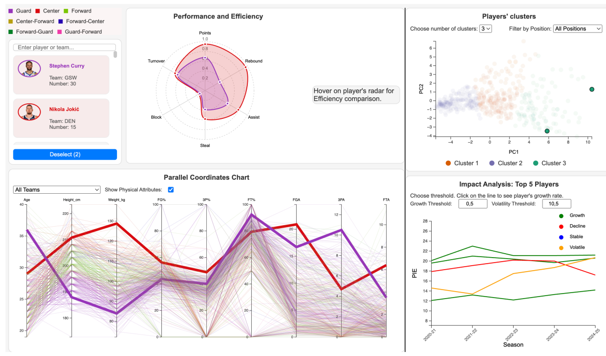
2. SYSTEM OVERVIEW

The main technologies used to develop the system are: Python (Flask), HTML, CSS and JavaScript (d3.js).

To run the program, on terminal we just need to execute the command:

```
python app.py
```

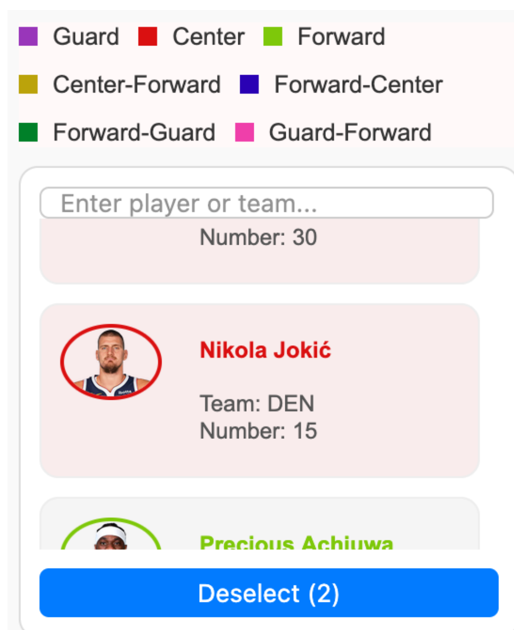
and then go to *127.0.0.1:5000* to view the webpage.
The user is greeted with a simple welcome page that invites the user to access the interface containing the visualization of the 2024-25 Regular Season NBA players performance analytics. In fact, by clicking the button at the center, the page shown after will be the official dashboard of the system:



As we can see, it is composed of different elements, each contributing into showing different aspects of the information available.

3. VISUALIZATION COMPONENTS

3.1 Players List



On the top left of the page, there is a section that allows the user to visualize a simple scrollable view that contains the list of all the

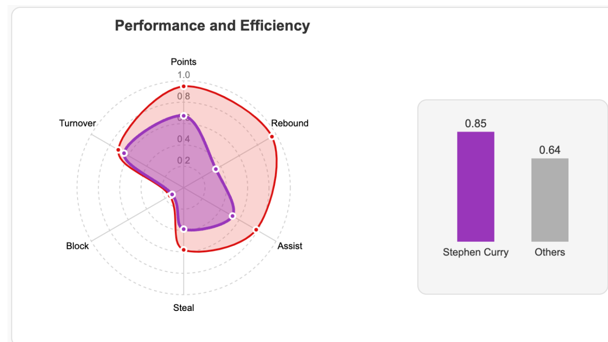
official players of the current season. The players are ordered alphabetically based on their surnames, and each entry has the image of the player on the left and general information on the right, which includes: full name, team (abbreviated with 3 letters) and their player number.
Above the scrollable list, there is a color legend that assigns different colors to each player based on their position in the game. The 3 main positions are Guard, Center, and Forward, but there are also hybrid-positions (Center-Forward, Forward-Center, Guard-Forward, Forward-Guard). I personally chose the color for each of them given the fact that there are 7 items and I wanted to create as much contrast as possible to avoid confusion. These colors are used in the name of the players and also in the images' borders, to quickly visualize which Position is assigned to that player.

3.1.1 Interaction

The user is allowed to search for a player by entering its name on the search bar, and the system will look for a player that matches the input, ignoring case. Additionally, it is also possible to filter the rendered players on the list by team: if a user inserts the name of the team, either parts of it or abbreviated, the system will look for the players with a matching value in their *Team* column and will display all the found players.

Users are also able to select players from the list. This is coordinated with other sections of the system (radar chart, scatterplot and parallel coordinates chart) in order for the user to easily visualize the representation of selected players in those charts, and so, to also compare them between each other. However, the maximum number of players selected is only 3, for more focused and meaningful comparisons and to avoid creating too many overlapping views.
To select a player, user can simply click on it on the list and the selected player will be highlighted and put at the top of the list, to avoid losing sight of it (given that there are more than 500 entries). While selecting, there is also a button below the list that lets the user cancel previous selections (deselect button). If user tries to select more than 3 players, an alert message is shown instead.

3.2 Radar Chart



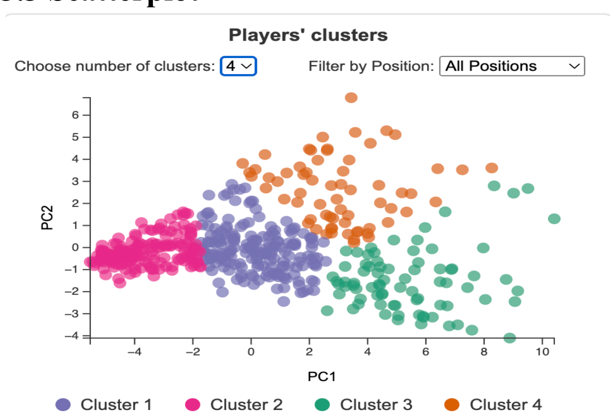
The next component shown besides the player's list section is the radar chart section. It is designed to visualize and compare multiple players' overall performance during the games of the season across a subset of the metrics simultaneously.
Each player's performance is visualized as a closed polygon, with rounded edges for softer visibility, and the polygon's highlighted points correspond to the values (normalized) along each axis. The axes represent a subset of the dataset's metrics, which describes in general the player's gameplay: PTS, REB, AST, TOV, STL, BLK.
Therefore, a larger polygon area typically means better overall performance, and different shapes highlights specific strengths (or weaknesses).

3.2.1 Interaction

The radar chart of a player is only shown in this section if the player belongs to the selected ones from the players' list or scatterplot sections (see 3.1). Since there can be selected up to 3 players, the chart can display up to 3

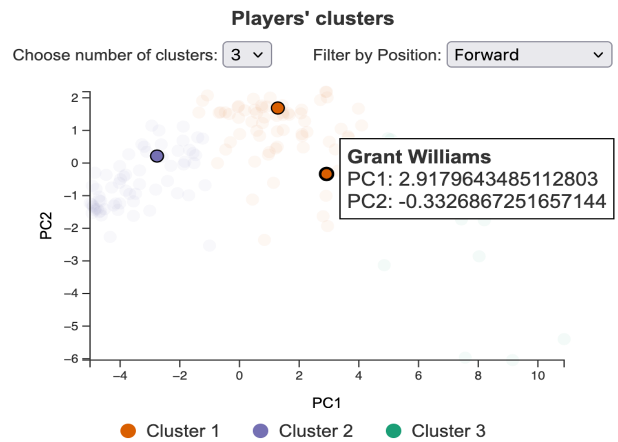
different overlapping polygons, with regulated opacity in order to have clear visibility of all of them and better comparison.
When hovering over the dots of the polygons, which are basically their intersection with the axes, the exact value for that axis is displayed. Furthermore, users can also move the mouse over the polygon's area and this will trigger the highlighting of that figure but also the display of a simple bar diagram, which visually represents a player's efficiency (calculated on the metrics considered for the radar and using a defined formula) in comparison to the average efficiency of all the other players. This will help visualize a general positioning of the player with respect to the average of all players of the season.
Lastly, the chart also supports zooming, for better visibility specially if the chart is too small.

3.3 Scatterplot



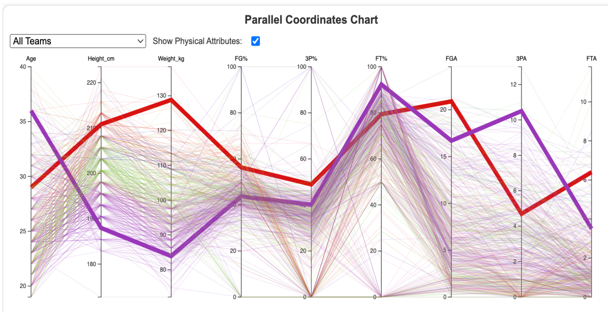
On the top right of the page, we have a scatterplot section that displays player performance clusters based on Principal Component Analysis and K-Means clustering. Two principal components are identified and they represent the x- and y- axes of the plot. Each cluster is represented by a unique color, generated using *d3.scaleOrdinal,* while excluding the colors already used for the Positions. This will help users explore and identify groups of players with similar performance characteristics and statistical profiles through PCA projection. Under the plot there is a legend that displays the mapping of colors to corresponding clusters.

3.3.1 Interaction



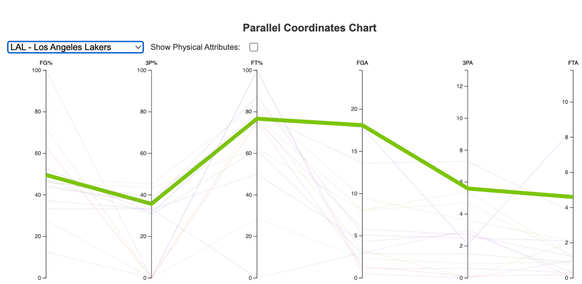
Similarly to the players list, users can select specific players by selecting the dots from the scatterplot. Again, only up to three dots can be selected. If user attempts to click more, the alert message is shown.
When selection happens, the points selected are highlighted by lowering the opacity of the other points. Furthermore, this selection with the scatterplot is coordinated with other sections of the page, so it works like the selection of the players list.
Additionally, by hovering over the points, a tooltip is displayed to show the name of the player and values of that particular dot for the two principal components.
Users are also allowed to dynamically compute the clusters by choosing the amount of clusters wanted. There is a dropdown menu that enables to select 2, 3 or 4 clusters.
Another dropdown menu enables users to choose to perform the PCA and clustering to all players, or to only a subset of them based on their position.
Zooming and dragging is also possible, to get a closer view of the points for accurate visibility and selection.

3.3 Parallel Coordinates



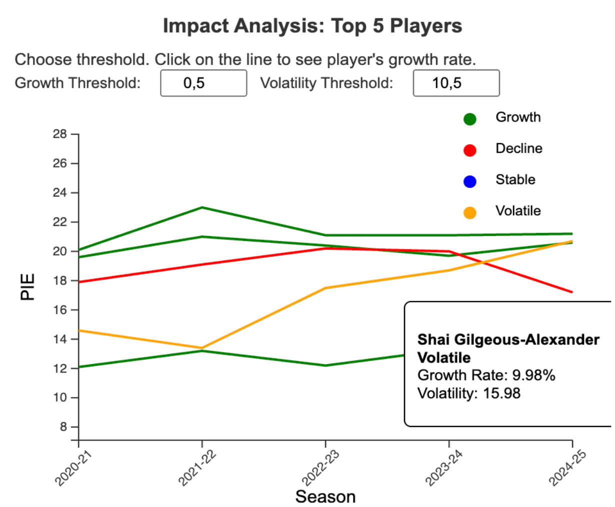
In the lower part of the page, we have the parallel coordinates chart. It helps compare all the players, but differently from the other charts, it focuses more on the players' physical attributes (Age, Height, Weight) and shooting metrics values (FG%, FGA, FT%, FTA, 3P%, 3PA).
It can help highlight relationships between these metrics by showing lines that connect each player's values across all parallel axes. Each line is color coded by the player's position, providing the possibility to differentiate and analyze the players also within their specific roles, while remaining visually consistent with the other sections of the web page. Some lines can be highlighted and have a thicker structure to be easily seen by the user.

3.3.1 Interaction



Here it is also possible to hover over the lines and a tooltip will show the specific values for that player for all the considered metrics. Since it is coordinated with the player selection logic either from the players list and from the scatterplot, if a player is chosen by the user then the opacity of the lines will change and the line corresponding to the selected player(s) will be thicker than the rest. Also, since there are about 500 lines for all the players of the league, too much overlapping is present. Hence, to give a better visualization, the lines in this chart can be filtered by team, which can be done by selecting a team from the drop down menu. All 30 teams are present and once selected, the parallel coordinates chart will only show the lines of the players of that team. Furthermore, it is also possible to focus solely on the shooting attributes of the players by excluding the physical ones: this is implemented with the help of the checkbox. User can zoom in on the chart too, for closer visuals.
Last but not the least, it is also possible to interchange the position of the axes, by clicking and dragging one of them wherever the user pleases, and the chart will update itself accordingly. This was implemented in order to allow the user to see different possible correlations between various metrics.

3.3 Line Chart



The line chart is the last visualization component added to the system. Its aim was to simply highlight the top 5 players of the 2024-25 regular sesaon (at least up until the data was downloaded). These are Shai Gilgeous-Alexander, Giannis Antetokounmpo, Nikola

Jokic, LaMelo Ball and Jayson Tatum. In particular, it shows the PIE value of these players over the seasons starting from 2020-21 until 2024-25. The PIE (Player's Impact Estimate) value is the estimate of player's overall impact and contribution to the game relative to their team's performance during a season league. The chart plots the PIEs of these players across the seasons, which helps reveal trends and determine growth, stability or decline over time, or eventual fluctuations in their impact.
The lines corresponding to the top 5 players are represented simultaneously and they are color-coded based on the trend of their growth rate. Color-coding is displayed with a simple legend.

3.3.1 Interaction

Another tooltip is shown to the users when they click one of the lines: in this case, it will show the name of the player, their average growth rate percentage, eventual volatility value, and a label that determines the trend of their performance.
The trend can change based on the thresholds for growth and/or volatility, which can be adjusted by the users, dynamically updating the colors of the lines.

4. ANALYTICS

Now, we delve more into the analytics part of the project, in other words, the process of preparing, transforming and interpreting the data that are then translated into the visualization components of the system.

For the radar chart, only a subset of the metrics were considered to help focus more on the key aspects of the general performance of each player. As mentioned on section 3.2, the metrics chosen for this analysis were:

- PTS: Points, describes scoring ability;
- REB: Rebounds, player's ability to secure ball possession after a missed shot;
- AST: Assists, passes that lead directly to a made basket, highlighting player's playmaking ability;
- TOV: Turnovers, the errors and mistakes in ball-handling, which is a negative contribution but offers a balanced perspective on the overall player's performance;
- STL: Steals, defensive skill of taking the ball from a player on offense;
- BLK: Blocks, also a defensive ability that prevents the opponent to score.

Since some of these metrics can vary widely in scale (for example PTS with respect to BLK), to make all of them comparable, their values have been normalized using a python script that executes a Min-Max scaling. After calculating the normalized values, they are joined to the original dataset as new columns through the Player column, and saved as a new *.csv* file, from which the backend script then renders the radar chart.
Another analytics used for this section was for the bar chart. The player's *Efficiency* is calculated, by using the general formula (REFERENCE):

$$EFF = (PTS + REB + AST + STL + BLK - ((FGA - FGM) + (FTA - FTM) + TO))/GP$$

The metrics used in this formula are all present in the dataset, so the system, triggered by user interaction (hover on the polygon), is able to process on demand through a Flask endpoint *(/get_eff_comparison*) and dynamically compute the Efficiency of the player in question and the mean of the Efficiency values of all the other players, displaying finally a simple and intuitive bar chart that compares them. Users can therefore see if a player's efficiency is above, below or in the league average.

Moving on to the scatterplot, the Principal Components Analysis dimensionality reduction, crucial for reducing high-dimensional data in 2D space, and computation is done on the original dataset and is triggered at document load. On the Flask server, a */pca_clusters* endpoint is defined: this function takes care of, first, standardizing the data using `StandardScaler()` to make sure that all metrics contribute equally to the analysis, then, computing both the PCA results and the clusterization of the data. In fact, when user wants to change the number of clusters, after selecting from the dropdown menu, the endpoint is called and the clusters are re-computed using K-Means algorithm. This is useful for more accurate categorization of the players, to find outliers, and data very similar or very far from each other. Lastly, the PCA is also re-computed on-demand when users filter the scatterplot by positions.

The line chart is useful to get more information on the progress of the top 5 players of the current season. As mentioned earlier, it is based on the PIE values recorded for the players in the past 5 seasons. The data for each player was downloaded from the official NBA Stats website and then preprocessed to be combined into one file (*top_pies.csv*).
The evolution of the players is measured as:

$$PIE_{\text{Growth Rate}} = \frac{\text{PIE}_{\text{current season}} - \text{PIE}_{\text{previous season}}}{\text{PIE}_{\text{previous season}}} \times 100$$

which calculates the percentage changes in season-over-season PIE growth rates for each player. This is done with the help of the Python pandas `pct_change()` method. The volatility is calculated as the standard deviation of the PIE growth rate and it is intended to measure the fluctuations of the growth rates of the player.
Two thresholds, the growth and the volatility, which can be user-defined, can categorize the players into 4 trend types:

- Growth: consistent PIE increase and low volatility
- Decline: consistent PIE decrease and low volatility
- Stable: minimal PIE changes and low volatility
- Volatile: inconsistent impact with significant fluctuations in growth rates.

5. INSIGHTS

The developed system makes it easier and more intuitive to understand players' current and most recent game statistics by combining analytics, visualization and interactivity, and there can be derived various insights and potential use cases.

NBA enthusiasts can make use of the system to have a clear view of their favorite player(s)' performance during the season, instead of reading large datasets and huge tables of metrics.

A coach can evaluate a players' strengths or weaknesses by looking at the radar chart: for instance, a player can be great at scoring but struggles with playmaking. Another possible use is comparing two players of the same position, and the chart could reveal that one has high defensive skills while the other can be more efficient in the offense or scoring. Therefore, new strategies for the future games or leagues can be designed, with eventual adjustment in player roles.

The efficiency comparison can highlight players that are significantly above the league average, indicating elite performers and overachievers, or significantly below for players currently underperforming. Team managers can use this information and take notes on the contribution of these players before considering contract extensions or trades.

The parallel coordinates chart can help evaluate if a player has a good set of shooting or physical attributes and can be analyzed to find correlation between them.

The scatterplot and clusters can identify players that have a statistical profile distant from the majority of the league, making them a potential outlier with unique contributions to the game. Opposing teams can find this useful to prepare strategies to counter these players.
For instance, we can see who are the players that are particularly excelling in their roles: Nikola Jokić, who is a Center, is notably distant from everyone both in the PCA analysis with all positions considered and in the one filtered with only Center roles, making him a player to look forward to for the rest of the season. He is in fact currently among the Top 5 players of the season, and fans can have a look at his performance in the past five years by analyzing the line chart and seeing his team impact growth rate.

All of these components can be used individually but also combined, and they can become a versatile tool that explores and transforms NBA data into actionable insights.

6. RELATED WORKS

Basketball analytics have grown and evolved remarkably over the years. Coaches, team managers, analysts have relied on statistics to make informed and valuable decisions both inside and outside the court. There are numerous platforms and systems that provide access to extensive game, team and player data and their visualizations. Furthermore, basketball analytics have also been explored widely in academic studies. These systems and researches demonstrate the consistent increasing importance of data-driven decision-making in basketball, especially in a hugely and worldwide known association like the NBA. In fact, an article by Mahima Phalkey focused on explaining the "*7 Ways NBA Teams Use Analytics for Performance Improvement"* and one of them is for 'Player Performance Analysis', which is exactly the purpose of the project presented.

The article *"Correlation matrix and Descriptive analysis - NBA",* published by Bianca Isaac on *medium.com* worked on the 2019-20 season data and performed a deeper analysis on how data-driven strategies improve team performance. The difference with my work is that the author used other visualizations like a heatmap and used team statistics instead of player ones, but both are focusing on a particular season to understand performance and metrics importance.

This other medium article, *"Basketball & Data: Understanding changes in style of play within the NBA through Data"* by Kushagra Luthra shares similarities with my project in its use of PCA and K-Means Clustering to uncover patterns in performance data. However, while this analysis focuses on dynamics of teams and historical trends, my project emphasizes player-level analysis during the ongoing 2024-25 Regular season. Additionally, my system offers dynamic interactivity, allowing users to explore player clusters, efficiency metrics, and role-based performance in real time.

An interesting research is also *"Exploring Game Performance in the National Basketball Association using Player Tracking Data*", Sampaio J, McGarry T, Calleja-González J, Jiménez Sáiz S, Schelling I Del Alcázar X, Balciunas M. This study explores player performance using data from the 2013-2014 NBA regular season. The similarities with my project is, again, its use of clustering algorithms: the research identified performance profiles like scorers, passers, defenders, and all-round players, while my system groups players into clusters based on their overall statistical profiles and filtering positions. While

this study focuses on player tracking technology and specific in-game actions from 2013-2014, my project utilizes more recent data (2024-2025 season) and evaluates broader statistical metrics.
Both works emphasize on applying analytics to support coaching decisions. However, again my project also focuses on integrating interactive visualizations to enable dynamic exploration of player performance and trends.

The paper *" Visual Analysis of NBA Player Data",* by Jiazhi Di, Ben Wang , Hua Hu, Qihang Zhao and Yinggui Wang, makes use of similar visual charts as my work, but again it is more traditional and research-oriented, focusing on past data to understand player and team dynamics, with no real-time interactivity.

In regards to data visual analysis about NBA data, my work is just one of the already present platforms that enables interactive visualization: *ViziBall*, for example, also makes use of radar chart and other visualization components to display individual player profile. It allows more interactivity and analyzes different set of metrics related to the player. Another example, is *NBAVis* which uses dynamic scatterplot visualization and a bar chart showing summary of the team's games. These are just two of the many examples of NBA interactive visual analysis.

In conclusion, the field of basketball analytics has seen meaningful advancements, with a considerable amount of studies and systems leveraging data to uncover trends, evaluate performance, and inform decision-making. My system offers an innovative, intuitive and quick approach for player-level analysis to visualize compare, and derive actionable insights in real time.

6. FUTURE IMPROVEMENTS

While the system achieves its primary goals, there are areas where it could be further improved and enhanced. For instances, by:
- incorporating more past data, organized by seasons, to have easy access and visualizations of historical performances
- incorporating live game data, which would enable users to analyze ongoing player performances and make in-game decisions;
- including more advanced metrics that would provide a more comprehensive evaluation of players' contributions;
- extending the system to include team-level dynamics to broaden its applicability;
- improving user experience by incorporating more interesting and intuitive UI elements.

7. CONCLUSION

The combination of interactivity, advanced statistical techniques, and dynamic visual exploration in the development of the system helps eliminate the gap between raw data and actionable insights. In this way, it gives the possibility for casual and professional users to uncover trends and make informed decisions, while also representing the growing role of data in shaping the future of basketball, and in general, sports analytics.

REFERENCES

1. NBA Traditional Stats 2024-25 dataset by Tilii, https://www.kaggle.com/datasets/tilii7/nba-traditional-stats

2. NBA Official Website player statistics:
- Players Index, https://www.nba.com/players
- Shai Gilgeous-Alexander, https://www.nba.com/stats/player/1628983
- Giannis Antetokounmpo, https://www.nba.com/stats/player/203507,
- Nikola Jokic, https://www.nba.com/stats/player/203999
- LaMelo Ball, https://www.nba.com/stats/player/1630163
- Jayson Tatum, https://www.nba.com/stats/player/1628369

3. ViziBall, https://viziball.app/nba/en

4. NBAVis, https://sites.cc.gatech.edu/gvu/ii/sportvis/nbaVis/vis-new/index.html

5. Efficiency formula, https://en.wikipedia.org/wiki/Efficiency_(basketball)

6. "Correlation matrix and Descriptive analysis - NBA", by Bianca Isaac, https://biancarisaac.medium.com/correlation-matrix-and-descriptive-analysis-nba-c479c8613dcc

7. "Basketball & Data: Understanding changes in style of play within the NBA through Data", https://medium.com/@kushluthra15/basketball-data-part-1-understanding-changes-in-the-style-of-play-within-the-nba-through-data-a8a0b6a9100e

8. "7 Ways NBA Teams Use Analytics for Performance Improvement", Mahima Phalkey, https://www.almabetter.com/bytes/articles/how-nba-teams-use-data-analytics-to-improve-their-performance

9. "Exploring Game Performance in the National Basketball Association using Player Tracking Data", Sampaio J, McGarry T, Calleja-González J, Jiménez Sáiz S, Schelling I Del Alcázar X, Balciunas M, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0132894

10. " Visual Analysis of NBA Player Data", Jiazhi Di, Ben Wang, Hua Hu, Qihang Zhao and Yinggui Wang, https://www.scitepress.org/Papers/2022/117688/117688.pdf

11. D3.js, https://d3js.org/