

Python-based Visualization Platform Implementation of CBA Players' Regular Season 2022-2023 Data

Hailong Tan, Chaoling Qin*, Zhengban Ran, Kai Wang and Zhirong Zheng

School of Physical Education, Southwest University, Chongqing, 400799, China

* Corresponding Author

Abstract

Based on the current CBA league's insufficient use of data visualization means and the lack of data analysis awareness of relevant practitioners, the platform obtains the CBA 2022-2023 regular season players' average base data and total base data through Python, then uses the calculation formula to calculate the commonly used players' PER value in the NBA league as an analysis index for higher-order data, and finally combines with the Echarts, Flask and other tools to complete the construction of the data visualization platform, designed and implemented the platform's three kinds of visualization and analysis functions: first, the analysis of the players' average data ranking. The second is the comparative analysis of players' ability. The third is the analysis between the higher-order data PER value and the players' average basic data. Aiming at the functional deficiencies of the platform and the defects of the CBA league's data statistics and analysis, this study puts forward two feasibility suggestions: first, the CBA league should fully learn from the NBA league's statistics and management methods of the data, and further improve the statistics of the game data. The second is to further enrich the functions of the visualization platform for the existing statistics of the CBA league to improve the applicability of the analysis.

Keywords

Python; data visualization; basketball game data; CBA.

1. Introduction

As the highest level of Chinese basketball league, CBA league has many excellent players, and through the wonderful performance in the game, it brings excellent visual enjoyment to the fans. A player's performance in the game can be largely reflected by the game data, and the statistics and analysis of the game data play an important role in the decision-making process of coaches and self-improvement of players. It can be said that data analysis is the decisive factor in the decision-making process of the team, and it is also the key link to play the value of data in the era of big data. From this aspect, the NBA league as the world's best basketball league, it is because of its storage of extremely detailed statistical records, historical data is extremely perfect, and through the integration of advanced technological means, from the basic statistical data to "scientific, comprehensive, real-time, detailed" data analysis, so as to provide the NBA league and other professional leagues around the world to compete. This has brought stronger competitiveness to the NBA league in the competition with other professional leagues around the world.

At present, China's CBA league has not yet reached the appropriate level in the strategic height of the use of data analysis, and the data management and analysis awareness of the main participants in the CBA has not yet been formed, and there is a lack of effective data analysis platform, which to a large extent constitutes a constraint on the development of the CBA league.[2] This is to a large extent a constraint to the development of the CBA league. Therefore,

the establishment of a visualization and analysis platform for players' game data and the improvement of coaches' grasp of players' competitive ability and targeted training for players through data visualization and analysis, as well as the improvement of game excitement and the overall level, play an important role in promoting the CBA league to improve the overall level.

Python, as one of the hottest programming languages nowadays, is widely used in the fields of artificial intelligence, web crawling, machine learning, etc., and using python as a research tool for data visualization research has become more and more common lately.[3] It can not only realize the player's game data. It can not only achieve the crawling of players' game data, but also combine with other computer technologies for the development of players' game data visualization platform. Therefore, this study used python to crawl the average base data and total base data of players in the regular season of CBA 2022-2023, and obtained the higher-order data PER value through calculation, and used the two as data cases for the development of the visualization platform, which aimed to enhance the richness and convenience of the current CBA league in the analysis of game data, and at the same time provide ideas for the CBA players' game data visualization analysis of CBA players' game data.

2. Research Objects and Research Methods

2.1. Objects of study

Taking the CBA player game data visualization and analysis platform as the research object, using python language, Echarts visualization chart library and Flask framework and other tools to complete the design and implementation of the relevant functions of the visualization platform, and using the average basic data and total basic data of 82 games of the regular season of the CBA 2022-2023 season as the data cases to carry out the various functions of the platform. Display. Among them, the average data refers to the average data of each game in which the player participated, and the total data refers to the sum of all the data obtained by the player participating in the game.

2.2. Research methodology

In this study, all the basic data of the CBA regular season players for the 2022-2023 season were collected in two different directions, regular season average and regular season total, respectively, using python, and the cleaning and storing of the data was accomplished in accordance with the format of the data presentation presented in the original web page.

There are a total of nine identical data items in the two categories such as player name, team played, court position, number of appearances, number of starts, playing time, regular season shooting percentage, regular season three-point shooting percentage and regular season free throw shooting percentage, and the remaining different data items are shown in [Table 1](#).

Table 1:CBA 2022-2023 Regular Season Statistical Items

Regular season averages			Regular season totals		
scoring average	Rebounds per game	Assists per game	Total number of points scored	Total Rebounds	Total assists
Steals per game	Caps per game	Turnovers per game	Total steals	Total Caps	Total number of errors

Fouls per game	Field goal attempts	Field goal attempts	Total Fouls	Total shots taken	Total number of shots
Three-point attempts per game	Three-point field goal percentage	Free throw attempts per game	Total three-point attempts	Total three-point hits	Total free throw attempts
Free throws made per game	Average frontcourt rebounds per game	Backcourt rebounds per game	Total free throws made	Total frontcourt rebounds	Total Backcourt Rebounds

2.3. Structured software development methodology

Structured software development method is one of the commonly used software development methods, whose basic idea is to solve complex problems in phases, using top-down, layer-by-layer decomposition, so that the problems handled in each phase are controlled in a way that is easy for people to understand and deal with.[4] The basic idea is to solve complex problems in stages, using a top-down, layer-by-layer decomposition approach, so that the problems dealt with in each stage are controlled within the scope that people can easily understand and handle. Its advantage is that the development process is clear, easy to manage and control, and can clarify the tasks and requirements of each stage and step.

For the development of the visualization platform, the first is the data acquisition and the selection of visualization indexes for the CBA players' regular season 2022-2023 in this platform, and the second is the specific development of the three parts of the platform: the backend data, the front-end display and the platform integration. The development of the background data part includes the acquisition and cleaning of CBA players' regular season data and the storage of CBA players' regular season data; the development of the front-end display part includes the selection interface of CBA players' regular season data visualization function and the design of CBA players' regular season data visualization page; the integration of the visualization platform completes the data interaction between the front-end display and the background data, and finally realizes the integration of the platform. The integration of the visualization platform completes the data interaction process between the front-end display and the back-end data, and finally realizes the integration of the platform, whose complete process is shown in [Figure 1](#).

The development tools used in this study include Python programming language, Navicat Premium database management tool, MySQL database, Echarts visual charting library and Flask framework.

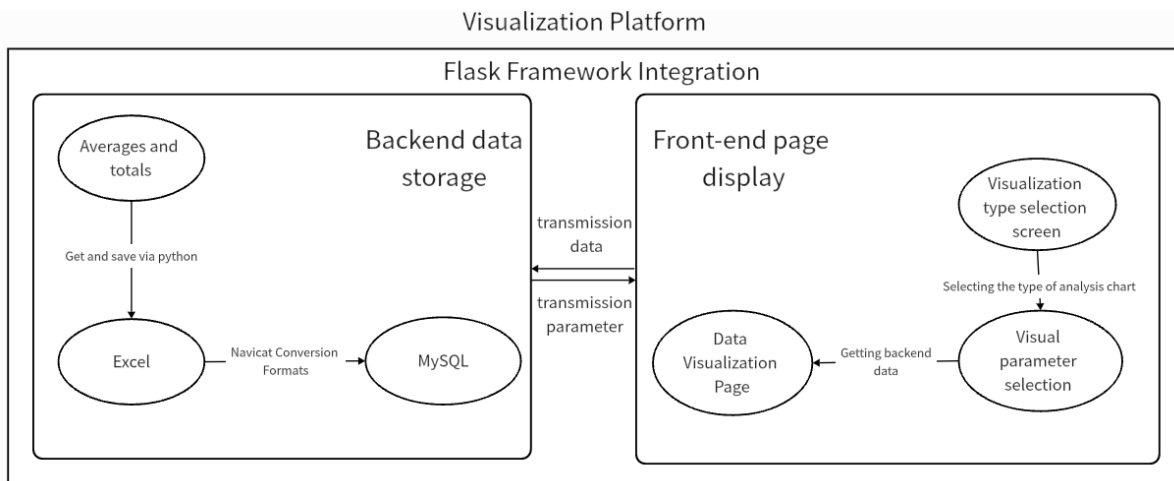


Figure 1:Flowchart of visualization platform construction

3. Visualization platform implementation process

3.1. Data Sources and Indicator Selection for CBA Player Visualization Analysis

3.1.1. CBA Player Match Data Sources

The data used in this study are all from the official CBA statistics website (<https://www.cbaleague.com/data/#/leaders>). By selecting the average base data and total base data of the regular season of 2022-2023 in the options of the webpage, the relevant data will be displayed to the webpage, as shown in Figure 2, and finally the source code of the webpage is saved to the computer, and the Python programming language is utilized to obtain, clean, and store the player's game data in the source code.

The official website of the Chinese Men's Basketball Professional League

Data homepage Players Team race Record ranking Playoff matchups Enter the player's name or the num. Q

Home/Basic data

season 2022-2023 Leg Regular season Statistical dimensions Averages per g... Player range All players Advanced filtering

Historical Data Missing Details >>>

ranking	Players	Playing for the team	location	Number of fields	Start	Time	score	Front plate	Rebound	Assist	Steals	Cap	blunde	foul	Shot hit	Shoot the shot	Shooting %	Three-point hit	Three-point shot	Three points %	Free throw hit	Free throws	
1	Blakeney	Jiangsu Kendia	Scoring guard	37	36	38.1	32.7	1.0	6.4	7.4	4.8	1.1	0.8	3.3	1.1	11.3	24.8	45.5%	3.3	8.3	39.4%	6.9	8.1
2	Darius-Adams	Fujian	Scoring guard	25	25	36.4	31.8	0.8	6.3	7.1	6.6	2.1	0.3	4.7	2.2	10.8	25.1	42.8%	3.8	11.8	32.5%	6.4	7.5
3	Peterson	Nanjing Tongxi	Scoring guard	38	36	32.9	28.7	1.3	5.8	7.1	5.9	2.0	0.2	3.7	1.7	9.5	21.8	43.4%	4.7	12.4	37.9%	5.1	5.7
4	Dominic Jones	Jiutai	Small forward	38	37	42.0	27.9	1.8	9.4	11.2	10.9	2.6	0.8	4.0	3.4	10.3	22.4	46.0%	1.5	6.8	22.3%	5.8	7.8
5	Germain	Qingdao	Point guard	41	8	22.6	24.2	0.5	4.8	5.3	3.7	1.1	0.3	2.4	1.2	8.8	19.7	44.5%	2.8	9.5	29.2%	3.9	5.2
6	Marchand-Brooks	Guangdong Dongguan	Scoring guard	41	22	27.3	23.4	1.2	6.3	7.5	5.9	1.4	0.4	2.5	1.2	8.5	17.3	49.1%	3.2	8.2	39.3%	3.2	3.9
7	Gillenwater	Shandong	Power forward	36	24	23.7	22.2	2.2	5.6	7.8	1.1	0.8	0.9	1.9	1.7	7.4	13.9	53.4%	1.6	4.1	39.2%	5.7	6.3

Figure 2:Data of CBA official website

3.1.2. CBA player data visualization metrics selection

For the various statistics appearing in basketball games, they can be roughly categorized into two types, one is the basic data of the players, such as the scores, rebounds, assists, and other data commonly found in the news reports of the games. The other category is the various higher-order data obtained by utilizing the various basic data of the players combined with the higher-order data calculation formulas proposed by the data statistics experts. In this study, the average basic data of the players in the regular season and the PER value of the higher-order data obtained by totaling the basic data are selected as the visualization indexes of this study. Basic data of CBA players in the regular season

In this study, the visual analysis of the base data is mainly carried out by using the average base data collected by Python for all the players in the regular season of CBA 2022-2023, including the average score, average rebound, average assist, average steal, average three-point attempts, average free throw attempts, etc. The total base data is mainly used to calculate the PER value of CBA players, which is not used as an index for visual analysis. The total base data, on the other hand, is mainly used to calculate the PER value of CBA players and is not used as an indicator for visual analysis.

On one hand, the average basic data can be ranked and analyzed to quickly find out the top players and the overall situation of other players in the current CBA league, and on the other hand, it can also be compared between different players at , which realizes the comparison and analysis of many abilities of different players to understand the advantages and disadvantages of the two players in the same position, and helps coaches to make more accurate decisions in different situations.

Table 2:Regular season base data visualization metrics

BASIC DATA ITEMS		
scoring average	Rebounds per game	Assists per game
Steals per game	Caps per game	Turnovers per game
Fouls per game	Field goal attempts	Field goal attempts
Three-point attempts per game	Three-point field goal percentage	Free throw attempts per game
Free throws made per game	Average frontcourt rebounds per game	Backcourt rebounds per game

Higher-order data PER values for CBA players in the regular season

Higher-order data is the use of a variety of basic data to perform a variety of operations to produce new data that represent certain meanings.[5] Higher-order data is the use of basic data to generate new data that represent certain meanings. Nowadays, it is difficult to analyze the players' performance by only relying on the traditional basic stats. The emergence of advanced stats, such as PER, win percentage, and true shooting percentage, provides a brand new perspective to analyze the players' performance on the field.

In this study, the Player Efficiency Rating (PER) efficiency value, proposed by ESPN expert John Hollinger, was chosen as the metric for analyzing higher-order data[6] It is one of the hottest

statistical indicators of higher-order data at present, and has not been used for the time being among the statistical indicators of higher-order data in recent years in CBA. This indicator is calculated by synthesizing the total base data of players in the current season, so that players of different positions and different generations can be evaluated and compared. [7] The metric is a composite of a player's total base stats for the current season. Using the total base stats of the players in the 2022-2023 regular season and the PER calculation formula, the PER value of each player in the 2022-2023 regular season can be calculated.

The formula for PER:

Athlete PER = [(Points + Assists + Total Rebounds + Steals + Caps) - (Shot attempts - Shot attempts) - (Free throws attempts - Free throws attempts) - Turnovers] / Athlete's games played [7]

3.2. Data Acquisition and Storage for Visual Analytics of CBA Players

3.2.1. Acquisition and Cleaning of CBA Players' Basic Data and Higher Order Data PER Values

This study looked up the official CBA stats website in the 2022-2023 After the average base data and total base data of players in the regular season, the webpage is saved to the computer. Python is utilized to parse the saved webpage to obtain the players' game data in it, and the key code is shown in [Figure 3](#):

```
url = 'https://data-server.cbaleague.com/api/player-base-list?pageNumber=1&pageSize=20'

headers = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) '
    'Chrome/118.0.0.0 Safari/537.36 Edg/118.0.2088.61'
}

tree = etree.parse( source: './CBA_HTML/CBA2022-2023.html', etree.HTMLParser())

# Getting data about players
playerData = tree.xpath('//div[@class="data-table-column"]/a/div/text()')

# Get Header Row
head = tree.xpath('//div[@class="data-table-column-title"]//text()')
```

Figure 3:Key codes for data acquisition

Due to the first parsing of the original data obtained from the web page there are some such as empty strings, garbled and other wrong format data, and can not be directly for data storage, but also need to use Python to clean the acquired data, remove the wrong format data, in order to carry out the operation of data storage, the key code shown in [Figure 4](#):

```
# Get Header Row
head = tree.xpath('//div[@class="data-table-column-title"]//text()')

# Clear the % in the list
head.pop(19)
head.pop(22)
head.pop(25)

# Clear the list of gibberish and receive it in data.
data = [x.strip() for x in playerData if x.strip() != '']
```

Figure 4:Cleaning of acquired data

Since the higher-order data PER values were not counted in the website, this study completed the storage of the total base data, and when it was necessary to visualize it after the platform was launched, it was calculated in real time in the platform, and the calculated PER values of

the players were passed to the page for the generation of the visualization graphs without any subsequent storage operation of the PER values.

3.2.2. Storage of CBA player base data

MySQL is a relational database management system, one of the most popular relational database management systems and the best RDBMS application for WEB applications.[8] . In this study, MySQL database is used to complete the storage operation of the average base data and total base data of players' games in the regular season of CBA 2022-2023.

Firstly, Python is utilized to write the code to save each average base data and total base data of 385 CBA players in the regular season of 2022-2023 obtained after data cleaning into two Excel data tables, and its key code is shown in [Figure 5](#).

```
# Save data
col = 1
row = 1

# Create an Excel sheet for storage
book = openpyxl.Workbook()
sheet = book.create_sheet(title='gameData', index=0)

for j in range(len(data)):
    if (j + 1) % 385 == 0:
        sheet.cell(row, col, data[j])
        col = col + 1
        row = row + 1
    else:
        sheet.cell(row, col, data[j])
        row = row + 1

# An additional row of cells is used to write the title
sheet.insert_rows(1)
col = 1
row = 1
for i in range(len(head)):
    sheet.cell(row, col, head[i])
    col = col + 1

book.save('CBA 2022-2023 Season Player Averages.xlsx')
print('File saved successfully!!!')
```

Figure 5:Data storage to Excel

Secondly, in order to avoid the complicated process of writing MySQL storage statements, this study uses Navicat Premium to complete the operation of storing two Excel tables of player data into a MySQL database. Navicat Premium, as a set of heterogeneous database management and development tools, can be used as a back-end management tool for this visualization platform, which allows creating editing databases, import and export data, create and run SQL queries, with an intuitive graphical user interface and tools, which is conducive to improving work efficiency[9] It has an intuitive graphical user interface and tools, which is helpful to improve the working efficiency.

After the final conversion to the MySQL database, the field average base data table is named cba22-23pergame, and the total base data table is named cba22-23alldata, and the two data tables are stored in the MySQL database in part as shown in [Figure 6](#).

1	Blakeney	37	36	38.1	32.7	1	6.4	7.4	4.8	1.1	0.8	3.3	1.1	11.3
2	Darius Adams	25	25	36.4	31.8	0.8	6.3	7.1	6.6	2.1	0.3	4.7	2.2	10.8
3	Peterson	38	36	32.9	28.7	1.3	5.8	7.1	5.9	2	0.2	3.7	1.7	9.5
4	Dominic Jones	38	37	42	27.9	1.8	9.4	11.2	10.9	2.6	0.8	4	3.4	10.3
5	Gaiman	41	8	22.6	24.2	0.5	4.8	5.3	3.7	1.1	0.3	2.4	1.2	8.8
6	Marshawn Brooks	41	22	27.3	23.4	1.2	6.3	7.5	5.9	1.4	0.4	2.5	1.2	8.5
7	Troy Gillenwater	36	24	23.7	22.2	2.2	5.6	7.8	1.1	0.8	0.9	1.9	1.7	7.4
8	Salinger	42	9	30.3	22	2.9	10.7	13.6	4.7	1.5	0.6	1.9	2.3	7.4
9	Wang Zhelin	41	41	33.7	21.8	4.1	7.6	11.7	2.6	1.1	0.9	2.8	2.4	8.9
10	Felder	41	35	29.9	21.4	0.8	4.3	5.1	8.7	1.8	0.3	2.2	1.7	7.9
11	Fogg	40	37	28.2	21.1	1	4.1	5.1	3.7	2.1	0.6	2	2.8	6.6
12	Hu Jinqiu	25	25	37.4	20.9	4.5	6	10.4	1.1	1	1.2	1.7	1.1	8.2
13	Gordon	38	26	30.7	20.8	0.9	3.1	4	5.5	0.9	0	2.9	1.4	7.3
14	Malan	37	24	27.2	18.8	1.1	4.2	5.3	5.9	1.6	0.4	3	2.1	5.9
15	Brandon Jefferson	34	14	27.1	18.6	0.9	3.3	4.2	3.7	1.7	0	2.1	1.4	6.2
16	Brandon Taylor	42	34	29.9	18.5	0.4	3.6	4	5.9	1.3	0	2.7	1.9	6.6
17	Achour	38	11	28	18.3	4.2	6.4	10.6	1.9	0.9	1.7	1.9	1.9	7.1
18	Zhang Fan	37	37	38.8	18.2	0.6	3.7	4.3	3.6	1	0.2	2	1.3	6.9
20	Trimble	40	22	23.9	18.2	0.4	3.7	4.1	5	1.5	0.1	2.7	2	5.4

Figure 6:MySQL partial storage results

3.3. Visualization page design for basic and higher order data of CBA players' regular season games

Echarts, a JavaScript-based data visualization charting library, provides intuitive, vivid, interactive, and personalizable data visualization charts and supports mixing and matching between charts and graphs[10]. This study uses the bar chart, radar chart and scatter chart provided by Echarts as the templates for designing the three visualization functions in this platform, reads the data that needs to be visualized and analyzed in MySQL through Python, and passes it to Echarts in Flask framework for the generation of visualization charts.

3.3.1. CBA players' average stats ranking analysis chart

Averages is a visual analysis of a player's averages, including points per game, rebounds per game, steals per game, and so on, using a bar chart as the basic format. By comparing the size of each data, they are arranged in the graph in the order of big left and small right, where the horizontal coordinate indicates the player's name and the vertical coordinate presents the selected player's data. In terms of the customization settings of the graphs, since the number of players is up to 385, this study designed the function of selecting different average data for visualization in the platform according to individual needs, and only one kind of average data is displayed in the same average data ranking graph, avoiding the phenomenon of blurring caused by displaying all the average data in the same graph. In order to further improve the clarity of the image, a segment display function with every 20 rankings as an interval segment is also designed to make the comparison of before and after data more reasonable and obvious. Taking the average score as an example, the corresponding graph can be generated after selecting it as the analyzing index and specifying the ranking interval, as shown in [Figure 7](#).

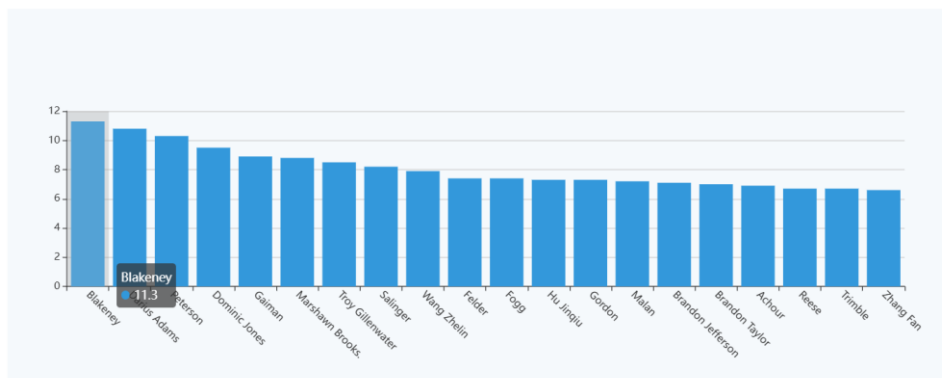


Figure 7: Analyzing the ranking of players' average data

3.3.2. Comparative Analysis Chart of CBA Players' Abilities

Radar charts show the independent values of each data item by means of radial axes, which extend from the center outward and connect the points on the axes with polylines to form polygons, which are used for comparing the overall situation of different data series on multiple indicators, and can be used as a modeling tool to effectively reflect qualitative problems with quantitative indicators.[11]. In order to carry out the analysis of individual players' abilities and the comparison between different players' abilities, this study utilizes the radar chart and the basic data indicators in Table 2 to complete the construction of the comparative analysis chart of players' abilities. The six indicators in the graph can be customized in the platform, and the maximum value of each indicator in the graph adopts the highest value of the indicator in the current season, so as to avoid the data from exceeding the boundary range of the graph, and at the same time, the size of the shaded area occupied by the data in the graph is more reasonable, and its basic style is shown in [Figure 8](#).

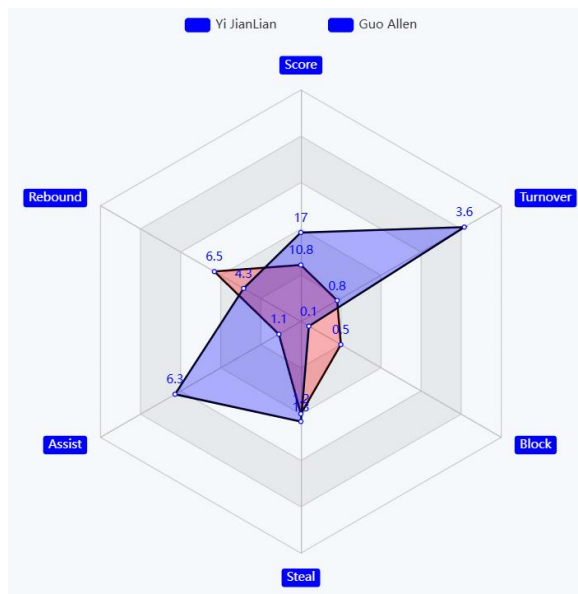


Figure 8:Comparative analysis of players' abilities

After selecting two players and the six indicators to be compared, the corresponding shaded area will be generated to be displayed in the graph and differentiated by different colors, and the corresponding graphs of the six indicators of a single player can be displayed after the graph is generated. The design not only compares the six data of a single player with the maximum value of the current season to analyze the overall performance of the player in the season, but also compares the two players to explore their respective strengths and weaknesses.

3.3.3. Analytical Plot of CBA Players' Base Data in Relation to PER Value

PER is an indicator that combines a variety of statistics in order to comprehensively evaluate a player's performance in a game. When comparing a player's basic stats, the introduction of PER can provide a more comprehensive explanation of a player's individual basic stats. For example, when comparing a player's scoring average, the PER value can be used to understand the player's efficiency in scoring. If the scoring is high and the PER value is also high, it means that the player not only scores a lot of points, but also makes fewer shots on the court, which is more efficient; if the PER value is high but the scoring is not, it means that the player may have an excellent performance in other aspects other than scoring.

Therefore, in this study, the design of the graph of the PER values of the players' higher-order data versus the base data analysis has been carried out by utilizing scatterplots, the basic style of which is shown in Figure 9. In the graph, the data on the X-axis is presented using only the PER values, while the data on the Y-axis can be presented by selecting individual metrics from Table 2 in the web page. In addition, the maximum scale values of the X-axis and Y-axis are automatically adjusted by Echarts according to the data to be presented in order to ensure proper display of the graph. Meanwhile, in order to avoid the blurring of the graph caused by the overlapping of data points when all the players' data are presented in the same graph, this study also designed a function to visualize the data according to the number of players' appearances, which not only avoids the overlapping phenomenon of the data points when the data volume is too large, but also improves the flexibility of the analysis.

CBA player stats analysis

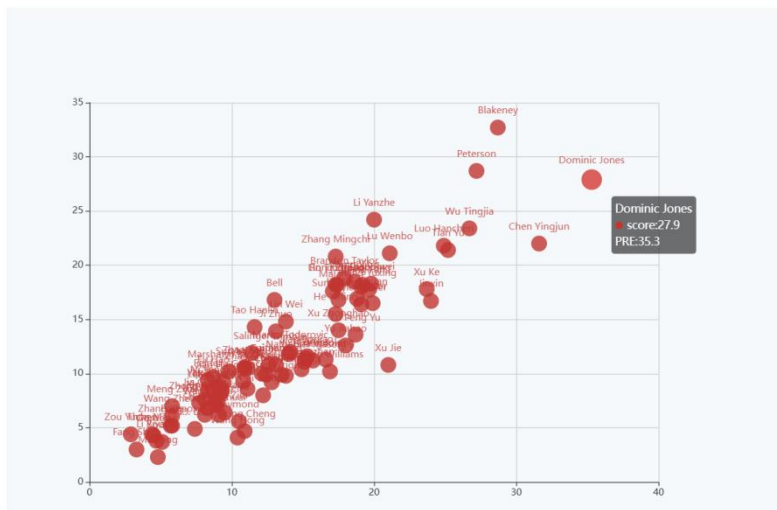


Figure 9:Player PER value vs. base data analysis chart

3.4. Integration of visualization platforms

Flask framework was born in 2010, is the use of python language based on the Werkzeug toolkit to write a lightweight web development framework , using it combined with the most popular and powerful python library can easily carry out the development of small and medium-sized websites .[12] . In this study, it can be used for front-end display and back-end data interaction, the front-end web page in the request to the Flask framework, Flask framework will request parameters passed to the MySQL database to get the corresponding data, and finally the data will be returned to the front-end web page Echarts for visualization and rendered in the web page.

4. Results and analysis

4.1. Visualization and analysis of CBA players' data

4.1.1. A Case for Analyzing Player Average Base Stat Rankings

Through the player average data ranking analysis chart on all the players' data visualization can help team coaches and players quickly understand the current CBA league all the players' game situation, for coaches to grasp the overall situation to provide some help.

Taking the scoring data as an example, you can view the players ranked 1-20 in the scoring data after selecting the corresponding parameter and submitting it in the field data ranking analysis page, as shown in [Figure 10](#).

CBA player stats analysis



Figure 10:Parameter selection interface for player data ranking analysis graphs

After Echarts processing, in the graph shown in [Figure 11](#), you can see that the player with the highest scoring data is Blakeney from Jiangsu Kentia. As a point guard, his average scoring data is as high as 32.7 points, for the coaches when the team will play against Blackie's team, inform the players in advance of the player's scoring performance on the court, and discuss with the players and set up the corresponding defense tactics and strategies to limit the player's scoring will be of great significance for the team to get the initiative in the game.

On the other hand, only three of the players in the top 20 in terms of scoring are local Chinese players, namely Wang Zhelin, Hu Jinqiu and Zhang Fan. This is a reflection of a problem that has always existed in the CBA league, i.e., the team's main offensive power is basically on the team's foreign aid, while most of the domestic players in the offensive rounds are more to cooperate with the foreign aid to complete the offense, and the independent offense plays less. Although it is true that the scoring ability of foreign players is stronger than that of domestic players, it is undeniable that taking measures to cultivate the independent offense ability of domestic players is of great significance to improve the performance of our national team in international competitions and gradually reduce the dependence of CBA teams on foreign players as well as to change the role played by domestic players in the offense.

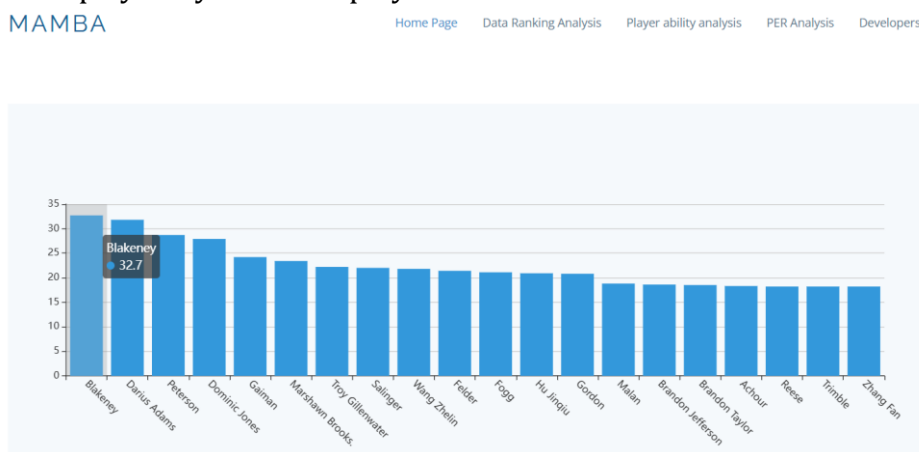


Figure 11: Ranking chart of players' scoring average data

4.1.2. A case of comparative analysis of players' abilities based on fielding base data

In this case, Guo Allen from Liaoning Bensteel and Li YiYang from Fujian Xunxing were selected for a comprehensive comparative analysis, and the data selected included six items of data, including average points, average rebounds, average assists, average steals, average caps and average turnovers, as shown in [Figure 12](#).



Figure 12: Parameter selection interface for comparative analysis chart of players' abilities

After analyzing the player ability comparison between the two players as shown in [Figure 13](#), it can be seen that in terms of scoring data, Guo Allen is ahead of Li YiYang 's 10.8 points with an average of 17 points per game. In the CBA 2022-2023 season, Guo Allen should have played as a point guard, due to the team's Zhao Jiwei as a point guard position, in most of the season as a point guard to focus on the team's scoring tasks, in the organization of the team is not fully engaged in the team compared to Li YiYang, resulting in the number of assists lagged behind

the results of RiYiYang. But even so, Li YiYang also with lower number of turnovers and steals ahead of Guo Ailun, showing good team organization and defensive ability.

Accordingly, for former CUBA player Li YiYang, who gained entry to the CBA through the draft, enhancing his scoring proficiency and sustaining his organizational and defensive capabilities at this juncture of the game, while additionally amassing experience in the game, will facilitate a more optimal performance in the game. For Guo Allen, if he continues to play the point guard position, and maintains his current scoring ability to further reduce his own mistakes and improve his defensive ability, he can make a greater contribution to the team's offense and defense.

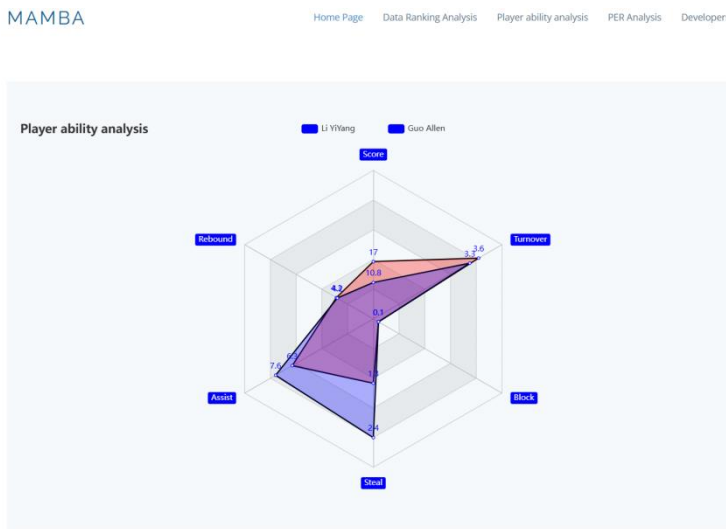


Figure 13:Comparative analysis of ability between players

4.1.3. Higher-order data PERs and fielding base data analysis cases

In order to avoid the chance factor on the field and the influence of too large amount of data on the graphic display, the visualization parameters of this case were selected to analyze a total of 77 players whose seasonal appearances ranged from 37 to 42 games, and the X-axis data of the graphic was fixed to the players' PER value, while the Y-axis was selected for the analysis of the average score of the game, as shown in Figure 14.



Figure 14:Parameter selection interface for higher order data and basic data analysis

In the graph shown in Figure 15, this study analyzes Dominic Jones from Jiutai Agricultural and Commercial Bank and Blackney from Jiangsu Kentia. It can be seen in the graph that Blackney's average score of 32.7 points per game is higher than Jones' 27.9 points per game, and from the point of average score alone, Blackney's performance in this season is really better than Jones. But the chart of Blackie's PER value of 28.7, much lower than Jones's 35.3, which indicates that Jones in addition to scoring in other areas, such as rebounds, assists, steals, etc. has more outstanding performance than Blackie. After checking the other data items of the two players

in the Excel table, it is found that Jones is ahead of Blakeney in a number of data in addition to the average score, the overall performance is more excellent, and its leading data items are shown in [Table 3](#).

Table 3: Comparison of data items

	Rebounds per game	Assists per game	Steals per game	shooting percentage
Blakeney	7.4	4.8	1.1	45.5%
Dominic Jones	11.2	10.9	2.6	46.0%

In the CBA league, the selection of the regular season best international player awards will often be a comprehensive examination of the player's performance and comprehensive contribution to the team, and at the end of the group voting points selected, rather than a separate examination of a particular indicator. Therefore, a total of 12 foreign aid candidates, including Jones, participated in the selection of the 2022-2023 CBA regular season best international player awards, Jones' group voting with 349 points ranked first, Shenzhen's Salinger received a total of 307 points ranked second, and the third place was Guangdong's foreign aid Marshon Brooks, who received a total of 246 points.[13] The third place was Guangdong's foreign aid Marshon Brooks, who scored 246 points. In the end, Jones was unquestionably selected as the best international player of the regular season.

In conclusion, the PER value plays an important role in the selection of comprehensive awards for CBA players. It provides an indicator for comprehensively evaluating players' performance, which provides objective data support for the selection process and makes the results more fair and accurate. At the same time, the PER value also helps us to understand the contribution and value of players more comprehensively, so that the fans and media have a better perception of the performance of players in the field.

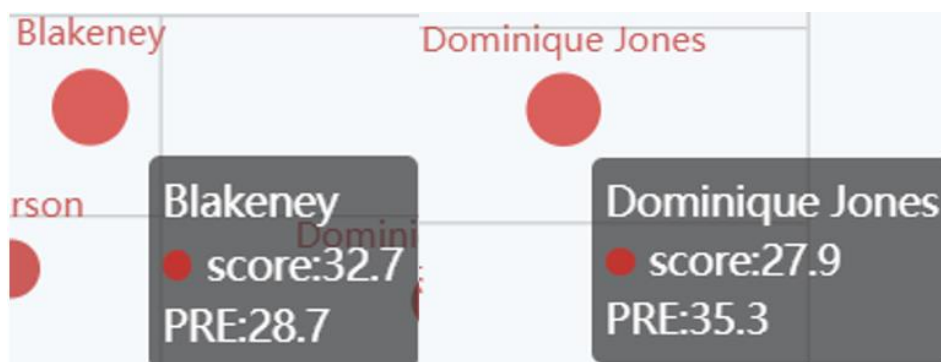


Figure 15: Higher order data vs. base data between players

5. Conclusion and outlook

5.1. Conclusion

Aiming at the shortcomings of the current CBA league in the use of data visualization means and the incomplete use of high-level data analysis indicators, this study firstly obtained and stored the average basic data and total data of the CBA 2022-2023 regular season through Python, and then calculated the PER value, which has not been counted by the CBA league for the time being, as a high-level data analysis indicator with the total data obtained. Finally, we combined the Flask framework, Echarts visualization chart library and other tools to complete the construction of the data visualization platform, and realized the three major functions of

the platform: ranking analysis of the average data, comparative analysis of the players' data, and analysis between the higher-order data and the dependent variable. Through the construction of the visualization platform, it provides convenience for the visualization and analysis of CBA players' game data, and the introduction of the PER value, a high-order data analysis index used in the NBA league, not only helps to enrich the comprehensiveness of the data statistics of the CBA league, but also provides a reference basis for the selection of comprehensive awards in the CBA league.

5.2. Outlook

5.2.1. Enriching platform functionality for existing CBA statistics

Although this study completed the development of the platform function by using the CBA players' game data, it only obtained the average and total basic data of the players in the regular season of the CBA 2022-2023 season, and did not obtain all the statistics of the official website of the CBA, and the developed platform has the shortcoming of not being able to carry out the visual analysis of the players' game data in terms of the year, so it not only helps to further develop the corresponding data visualization and analysis function of the platform, but also improves the accuracy and comprehensiveness of the data visualization platform's analysis of the players' game data. Therefore, obtaining all the statistics of the CBA league at this stage is not only conducive to the further development of the corresponding data visualization and analysis functions in the platform, but also improves the accuracy and comprehensiveness of the data visualization platform in the analysis of the players' game data.

5.2.2. CBA league should enhance the comprehensiveness of statistics

Compared with the NBA league's perfect data management system, the CBA league's statistics on players' game data started late, and the players' running distance, shot tendency, and data obtained from a single game in the season were not counted in its official data website, which not only restricts the development of the corresponding functions of the data visualization platform, but also greatly restricts the progress of the CBA league to the world's first-class league. This not only restricts the development of the corresponding functions of the data visualization platform, but also greatly restricts the pace of CBA League to advance to the world-class league. Therefore, it is of great significance to improve the data management system of the CBA league, learn the method of data statistics from the NBA league, and strengthen the data awareness of the relevant practitioners, in order to promote the improvement of the overall level of the CBA league and the development of a more comprehensive data visualization platform.

References

- [1] Yang Zhenxing, Yang Jun, Bai Jie et al. Research on the American professional basketball league based on big data technology[J]. *China Sports Science and Technology*, 2016, 52(01): 96-104. DOI: 10.16470/j.csst.201601014.
- [2] Jia Baojian, Yang Zhenxing, Yao Jian. Application of data analysis and revelation of the U.S. professional basketball league[J]. *China Sports Science and Technology*, 2018, 54(06): 118-126.
- [3] Zhu YZ, Jing J. Research on Chinese word segmentation technology based on Python language[J]. *Communication Technology*, 2019, 52(07):1612-1619.
- [4] Post Kaihua. Research and implementation of the twenty-first CUBA players' data visualization based on Python [D] Xi'an:Xi'an Institute of Physical Education and Sports, 2022.
- [5] Yang Zhenxing, Bai Jie, Yao Jian. Research on big data statistics of NBA league[J]. *Sports Culture Guide*, 2015,(07):103-107.
- [6] Zhao Shuqiang. Comparison of positional efficiency of CBA foreign aid and analysis of influencing factors[J]. *Sports Research and Education*, 2014, 29(06): 91-94.

- [7] Liu Boyang, B. Analysis and evaluation of NBA player ranking system[J]. Sporting Goods and Technology, 2016,(22):45-46.
- [8] Bell C, Kindahl M, Thalmann L. MySQL High Availability: tools for Building Robust Data Centers [M].2010.
- [9] Wang YX, Liu F. Practice of data visualization application in university library based on Navicat+Tableau[J].Electronic World, 2020, (21): 94-95+99.
- [10] Li JL, Yuan X, Yang B. Design and implementation of infectious disease data visualization platform based on Web technology[J]. Computer Application and Software, 2023, 40(10):101-106+173.
- [11] Fu, Eddie, and Fang, De-Ying. Research on the application of radar chart method in comprehensive evaluation[J]. Statistics and Decision Making, 2007,(24):176-178.
- [12] Niu Zuodong, Li Handong. Building a practical MVC framework that can be efficiently developed based on Python and flask tools[J]. Computer Application and Software, 2019, 36(07):21-25
- [13] Sina Sports. Jones elected the best international player in the CBA [EB\OL] (2023-04-08) [2024-06-01].<https://sports.sina.com.cn/basketball/cba/2023-04-08/doc-impstnw6450950.shtml>
- [14] Jin Yingyan, Jia Juncheng, Hong Minjie et al. Data analysis and visualization for NBA players[J]. Computer Applications and Software, 2021, 38(08):84-91+174.