Tableau to Bokeh: A Case Study of Basketball Analytics Reporting

CSC-475: Seminar in Computer Science
Fall 2023

Jonathan Lawrence
jonny.lawrence@furman.edu
Department of Computer Science
Furman University

Abstract

This project is the journey and creation of an eventual business planned to launch in the upcoming year. Currently in the product development phase, the mission is to make the most efficient and effective system possible for basketball analytics reporting, specifically at the collegiate level. The process involves grading out practices statistically, organizing cumulative data efficiently, writing code to display the data meaningfully in a dashboard format, then distributing this dashboard to the client, whom in this case are members of the coaching staff and players. The mission is being accomplished by writing Python code that automates these processes, and creating a website that is updated and managed for each team brought on as a client. Furman University Men's Basketball is currently the only client in which results were produced for. The initial system, via Tableau static report generation, was effective, but not the most efficient, and not scalable. The system that is being transitioned into, via Bokeh interactive report generation and personalized website creation, is both efficient and scalable, but not yet as effective. This meaning that the dashboards aren't currently as effective telling a complete story of the data as the prior system. This is all very achievable however with more time and effort, which there will be plenty of both, as the goals and vision of this project are not satisfied, and not bound by the time constraints of this semester.

1 Introduction

Data analytics is a rapidly growing industry in terms of both demand and prevalance. It is such a versatile job that just about any field, business, or organization can benefit from regardless of what they are trying to accomplish. Sports teams, specifically basketball in this case, also fall under this umbrella where there is a lot to be gained by teams through sports analytics. The current market for basketball analytics is primarily focused on analyzing in season game data. This is natural as the big business of basketball, ESPN, and the NCAA is centered around game revenue. The potential for the business of this project however is taking a more team-development driven approach and offering teams in-depth reports similar to what teams commonly hire for an in-season analysis, but for off-season practice data instead. So, this will be a much easier market to enter and is a unique space that the business can thrive in. The off-season is many months in the May-November period where teams can identify problems and figure out roles and minute rotations of players before the season comes, and constant clarity of where players stand in relation to others and the team is essential for decision making by coaches. This is a field the author personally has a lot of knowledge and experience in being a college basketball player himself as well as a computer science major/data analytics minor, and is confident and able to build a unique product/service, start a small business, pitch knowledgeably to a head coach, and aid teams closer to success.

Within the time parameters of this semester project deadline, the objective was not to accomplish all of this and start my own business already, but rather to lay the groundwork for launching a successful college basketball data analytics business to work further towards beyond the semester. "Lay the groundwork" meaning establishing a system that takes in all practice data and outputs detailed reports on the team's and individual player's stats, progress, and standing in relation to the team and other players both effectively and efficiently. Effectively, referring to the actual quality and usefulness of the report, and efficiently referring to establishing a system that can take in multiple clients and service them all in a reasonable, timely manner every time they need a new report, which will be every new practice or two. The groundwork was laid by treating Furman University Men's Basketball as my first client and creating a growing portfolio on them as reports accumulated. College basketball is a huge market and there is definitely a demand for this work. This project was a great way to initially attempt to monetize a personal strength and passion.

The data being used is raw practice data of the clients/teams (Currently only Furman University Men's Basketball) sent by the coaching staff via email after every practice. The data is expanded and manipulated using the Pandas library and displayed using the Bokeh library in the Python programming language. The old system that is being transitioned away from in this project was using Tableau to create the reports. However, this system was not as efficient, distributable, accessible, interactive, or scalable as a system that should be in place for multiple clients. The basic premise behind how the new system works is with every new practice, the data will be updated and accumulated, written from scratch custom Python code will run updating all visuals, and then post the latest version to a custom website that has been built for the client to access whenever they please. This system is currently being expanded, making plenty of more visuals and interactivity options, eventually moving all graphics over from the Tableau system, but the basic structure and necessary visuals are already established and have already been given to the Furman coaching staff access to their website.

2 Related Work

A wide range of diverse sources have been gathered related to the topic of interest that will aid in accomplishing the goals that have been set out to achieve. There are a couple of different fields that are important to find expert knowledge in that when put all together, will provide the best frame of mind to contextualize this project. Those general fields are graphic design/visualization, data analytics in athletics, and implications/impact on the basketball world. Some of these sources are broader in scope that helped gain necessary knowledge that was applied to the project. And others are seminal papers more closely related to what is being set out to accomplish.

One of the broader sources that helped is "What do we Talk About When we Talk About Dashboards?" [1]. This source is all about how to make better visualization tools to better reach your audience. This information was used to try and make the reports as simple, effective, and visually pleasing as possible. "Designing Data Visualizations" [2] is another tool that was used to potentially increase the effectiveness of the reports similar to that of the prior source mentioned. This book contains a good balance of ways to improve your graphics, but also does a good job at explaining the importance and motivations behind data visualization. These two sources were go to guides for improving the optics of the reports.

There were a couple different sources referring to data analytics, and the relation, impact, and opportunity to be gained when applying it to the realm of athletics. One source that is extremely interesting, "A.V.R. Hybrid Design for Sports Data Visualization Using AI and Big Data Analytics" [3], explains the importance of big data and analytics with a sports lens and explores the impact that AI can have in this field and potential for greatly improving model accuracy. This is a great paper to reference as it emphasizes the importance of the sports data science field, as well as offering other thoughts as to what is possible in this industry. This could possibly be used to expand upon the project even further in the future. Another source, [4], puts an implication lens on this topic and explains the lucrative and vast publicly accessible sports analytics industry that is rapidly growing through technology and is utilized greatly by organizations, businesses, fans, and media. The last source in this area that isn't explicitly focusing on basketball is "Developing Athlete Monitoring Systems in Team Sports: Data Analysis and Visualization" [5]. This article is interesting because it talks about tracking the things that aren't just about the sports statistics, but rather things like the likelihood of injury for an athlete. This was used to relate to the Kinexon piece of my project, where not just the basketball statistics are tracked, but rather the athletes movements, moreso metadata if you will. The Furman Men's

Basketball strength coach also currently tracks the players hydration levels before every practice, which is used as a likelihood of injury measure.

The remainder of the sources are directly related to data analytics and basketball data analytics in particular. Three of these sources "Basketball Analytics. Data Mining for Acquiring Performances" [6], "Sports Analytics—Evaluation of Basketball Players and Team Performance" [7], and "Modeling Player and Team Performance in Basketball" [8] all have very similar messages that most directly relate to what was set out to accomplish. However, the two main differences in these papers and what is done in this project is that they are working with the NBA, not college basketball, and they are only working with game data, where the emphasis of this project is on practice data and then possibly relating it to game data. These sources talk about using data analytics and reporting to increase understanding within a program to help player development, identify team strengths and weaknesses, help coaches' decision making process on who to play and value easier, and thus increase the overall success of the organization. Individual and team performance will be constantly evaluated to assess and suggest change to the current state of the team, while also laying out a predictable future path for the team. Another source that was very helpful is "Basketball Analytics: Objective and Efficient Strategies for Understanding how Teams Win" [9]. This source talks more specifically about how teams win and what is most important that contributes to winning. Obviously increasing winning is the tangible bottom line and end goal of the project's value added that makes it appealing, so this source was used to back up and possibly get suggestions for what else should be tracked in the analysis. Lastly, "Basketball Analytics Using Spatial Tracking Data" [10], is pretty much exactly the motive behind the measurement of the Kinexon movement data, so this was treated as a seminal paper in that area of the report.

3 Data

The data will be pulled from two different sources, making a custom dataset that combines the two with some other analytical additions, and will get larger and larger with each new practice that the Furman basketball team has. The coaching staff watches the practice film after every practice and puts all raw stats into an Excel file in which they send usually the next day. From there meaningful basketball statistics are calculated that gauge multiple different areas of play to get a complete picture of the team's strengths and weaknesses. That is the main source, but the Kinexon database for Furman Basketball will also be drawn from daily. Kinexon is a company that sells electronic chips that players put in their practice shorts that tracks their movements, jumps, accelerations, etc. to essentially track how hard the player is working during a given practice. This is used to try and find correlations between player stats and how hard the player is playing. Some meaningful features of the dataset include Three Point Percent, Effective Field Goal Percent, Offensive and Defensive Rebounds per Possession, Assist to Turnover Ratio, alongside many many others. All of these metrics are recorded for each individual player as well as the collective team for every practice since the beginning of the summer. The rows are equal to the number of players + 1, each player being one row plus the team row. The number of rows varies a little bit over time due to transfers, graduates, and freshmen/walk-ons joining the team. The columns are multiple different statistics, over 60 which do get filtered down to the most important ones. These vary too as the coach requests to track different things as the team progresses. Every sheet in the Excel file is a different practice, so essentially the dataset is three dimensional. The focus of this project is this year's data so it includes data from June-November, however last year's data is also included. There are 63 sheets just in that June-November period, so the dataset grows quickly. Figure 1 is provided as a glimpse of the dataset to provide an idea.

Player Name	Off. Poss.	Def. Poss.	A 2FGM	A 2FGm	B 2FGM	B FGm	3FG	3FGm	FT	FTm	AST	TO	DF	ST	Foul	Foul Drawn	OREB
#0 Smith, Pjay	34	28	1		1	1	5	2					6	2	1		3
#1 Pegues, JP	34	30	1		1	1		3			6	1	7	2			
#2 Bronson, Eddrin	25	26		1			3	2				2	2				1
#4 VanderWal, Ben	26	29	1				2	1	2		2	1	3	2		1	
#5 Foster, Marcus	33	29	1		1		1	2			3	2	3			1	1
#10 Humphries, Tommy	22	18					1	1			3	0	1		2		
#11 Whitt, Carter	31	36	1	1	3			2	1	1	3	5	2			1	
#12 Molnar, Davis	31	35	1	1	1				2			3	3	3		1	1
#13 Hien, Garrett	34	32		1	1	1	1	1	1		4	4	5	1	2	1	1
#15 Hughey, Tyrese	6	7	1								1	0	1	1	1		1
#21 Bowser, Cooper	31	36		1	1							1	4	1			1
#22 Collins, Cason																1	
#24 Williams, Alex	18	19	1			1		1					4	3			
Total	325	325	8	5	9	4	13	15	6	1	22	19	41	15	6	6	9

... | 10-7-23 | 10-9-23 | 10-10-23 | 10-11-23 | 10-12-23 | 10-16-23 | 10-17-23 | 10-19-23 | 10-20-23 | 10-21-23 | 10-23-23 | 10-24-23 | 10-26-23 | 10-27-; ...

Figure 1: Excel File Data Format

4 Methods

The process includes pooling all data from the two different sources (receiving it from coach and Kinexon), making edits, cleansing it, and making calculations for certain basketball metrics on that data.

4.1 Old System

From there, the old system was to import this data into Tableau in which informative visual reports in the form of collections of dashboards were created that would be exported to a PDF and texted and emailed to all the players and coaching staff, as well as printed out hard copies delivered to each coach. The early part of the research was just trying to expedite this process as much as possible, which happened very efficiently. However, about halfway through the semester, the focus shifted to transitioning to a new system that would make this process more efficient, more interactive, more accessible, and most importantly more scalable.

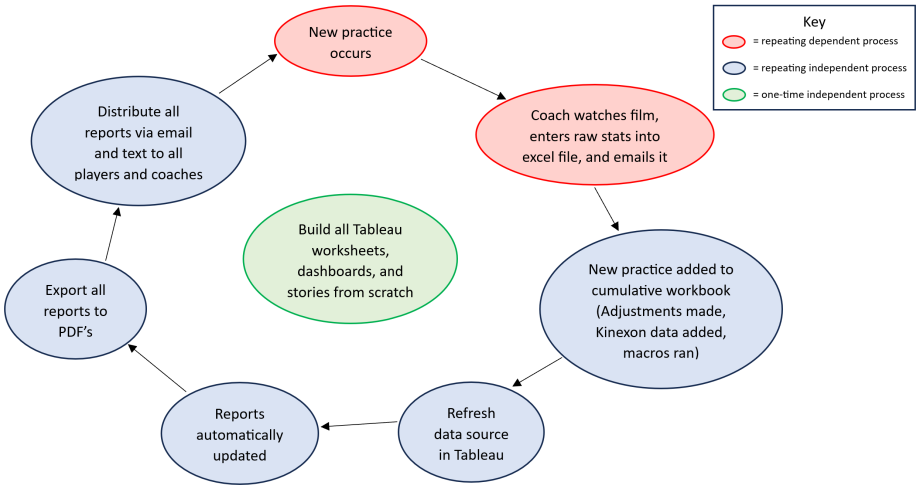

Figure 2: Old System- Tableau Pipeline Diagram

4.2 New System

The new system involves writing code using the Pandas and Bokeh libraries in Python. The Pandas library is very useful in having efficient vectorized operations on the dataset. And the Bokeh library is very useful for the visual and interactive display of the dashboard of visuals to tell the story of my data. Finally, instead of exporting these visuals into reports that need to be distributed after every

practice, a custom website was made for the client via a GitHub Code Repository that is constantly updated. Each client in the future will have their own statistics website that is uploaded to with every new practice. This way all that needs to be sent to the coaches and players is one persistent link to their website in which they can go access whenever they want. This will need to happen after every practice or every two practices. There are usually 3- 5 practices a week depending on the time of year. See Figure 2 for the Tableau pipeline. Here is what the website pipeline looks like: new practice takes place –> coaches watch practice film and email over raw stats for that day –> new day added into the existing Excel file making small tweaks and adding Kinexon data –> refresh the Excel file and run Python code –> update website with new practice data using git commands in the Terminal.

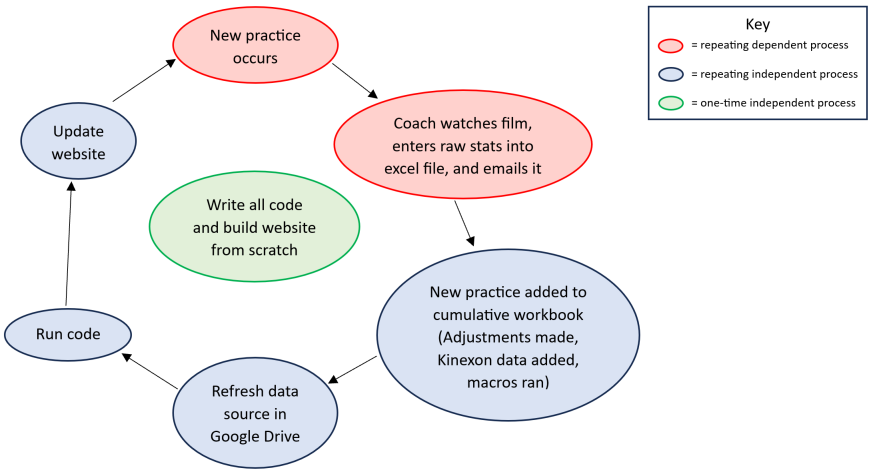

Figure 3: New System- Interactive Website Pipeline Diagram

5 Results

In the course of a single semester, there are plenty of tangible goals that were able to be achieved. First off, the Furman Basketball players and staff have found it very useful which is the entire goal in the first place. It is brought up in coach to player conversations, caused changes in practice plans when a focus shift is realized, and regularly discussed in staff meetings and presented in team meetings with players. Also, until realizing the need and value of switching systems, great improvements were made on the old reporting system in both content and efficiency. In terms of content, new visuals and metrics were created to tell a more complete picture of the status of the team including the creation of a net ranking formula that essentially took all important metrics and spit out one overall grade per player. The efficiency of the system was improved roughly by a factor of 4, as it used to take around 2 hours to generate and distribute each new report to now around 30 minutes. This involved cleaning and cutting up the datasets in a more efficient to use manner, programming excel macros to automate certain tasks, and pooling all 3 of my regular reports into one Tableau workbook with a collective Story page each that included all dashboards per report. The first two points will help tremendously towards the new system as well. Also note that "Old System" is said to indicate the transition, but until all of its contents are successfully transitioned over to the website the client has been and will continue to receive the Tableau reports as well because it is a client that expects results. So this has been managed daily throughout the semester as well.

After getting pretty much the most out of the old system that was possible, the benefits of a complete system change were realized, and the process of writing a program from scratch using the Pandas and Bokeh libraries in Python began. In a fairly short amount of time, the majority of the structural code has been written, and Furman basketball already has a new statistics website with a homepage and three hyperlinks that takes the user to current game stats, this year's practice stats, or last year's practice stats. The website is fully interactive allowing the user to navigate to whatever page they

want, look at what ever metric they want for the entire team or any individual player in the form of a line graph of cumulative averages, and hover over the points in the line graph for more specific statistics details. With this new system the client can visit the website whenever they want instead of having to wait on the next email or physical copy. It is all in one persistent location for them now in which they can access the data whenever they desire to do so.

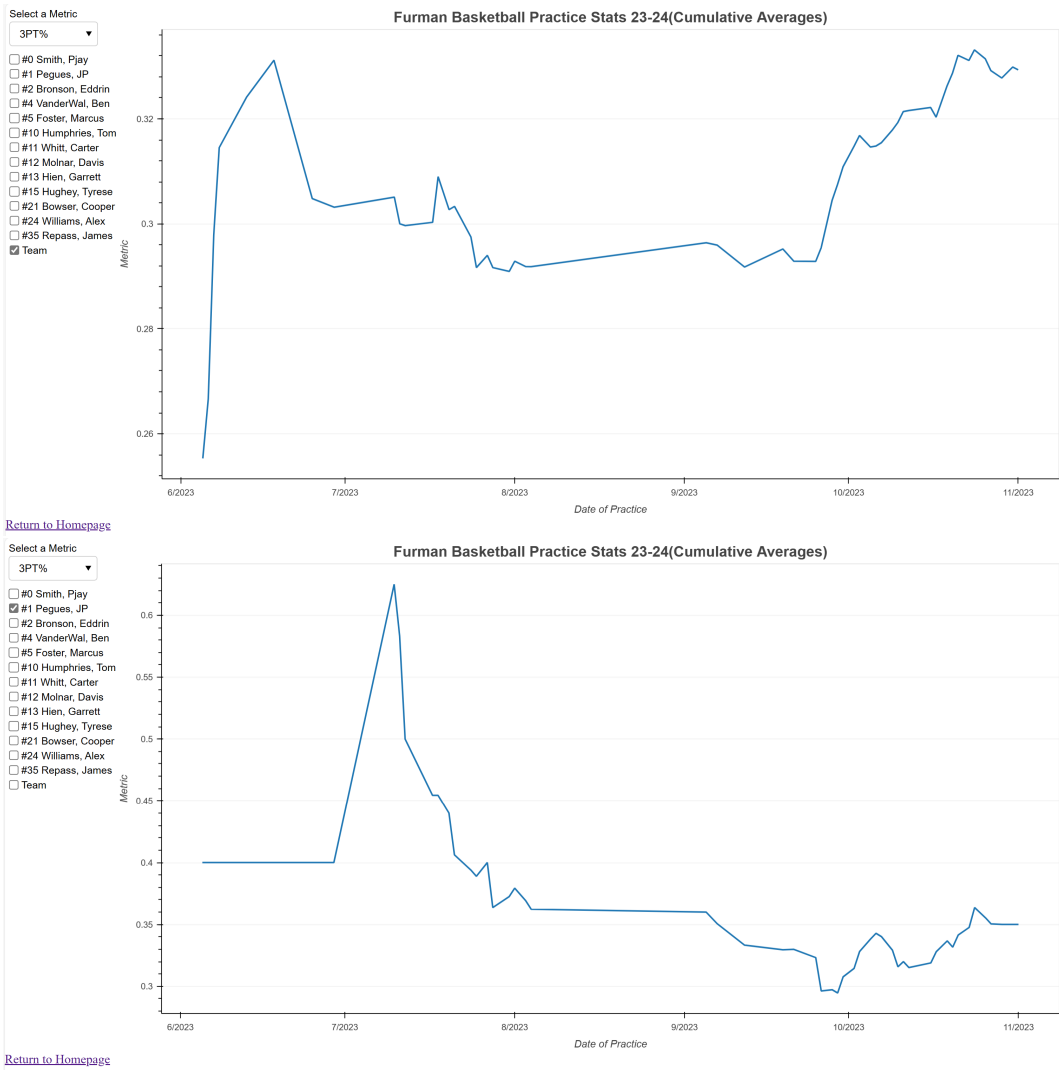

Figure 4: One Website interactivity option- Cumulative Average Three Point Percentage for Individual Player vs Team Comparison (All Practices June-November)

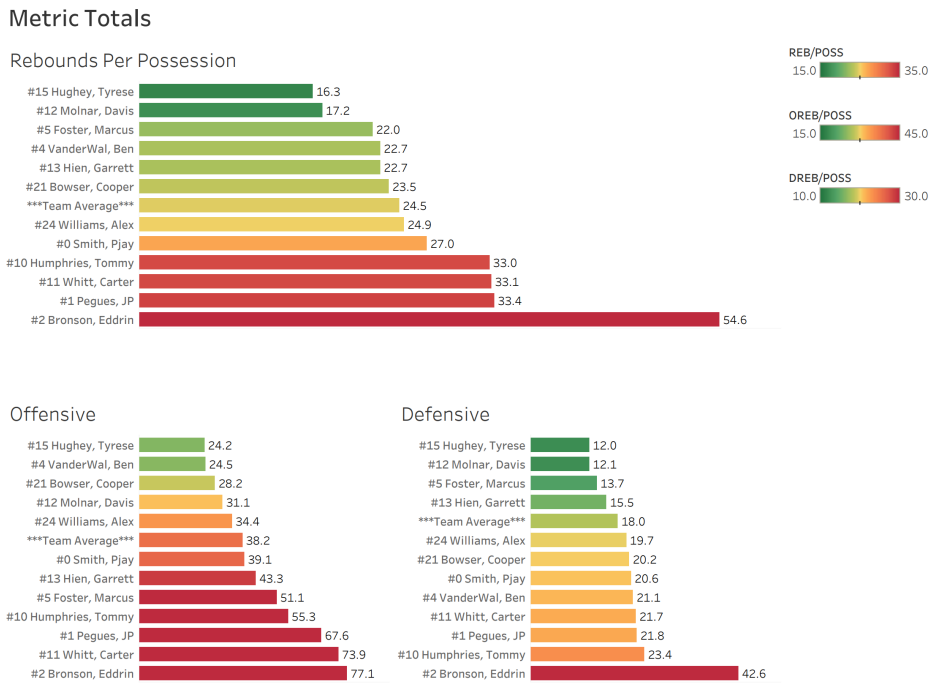

Figure 5: One Page from Tableau Report- Cumulative Aggregation Rebounds per Possession(For all In-Season Practices)

5.1 Dissemination

The benefits of this project will continue to be shared initially to the Furman Men's Basketball program, and then to any other college basketball programs that can be taken on as clients. After this year, a full profile on Furman will be tangible and will be used in meetings with other colleges and market what can be done for them through experience and example. In the future a website of the business will most likely be used as marketing with sample reports included. There will be a demo and possibly a physical copy of the latest report to show evolution of the system at Furman Engaged.

6 Future Work

Before describing future work, a few limitations need to be discussed. For one, a main limitation is the issues and constant communication needing to be had regarding the underlying dataset (Excel file) that gets sent to me. Before work can be started each time, the coach has to grade the practice film for that day and send it which can vary in completion time. the Excel file rows and columns change somewhat frequently, columns moreso with every time the head coach wants to stat something new, and there were two instances this year needing to adjust for row changes with a transferred player and with walk-ons joining the team. Also, every client worked with will have a different formatted dataset that will need to be customly adjusted for in the code with each team. This is just the nature of not watching the practice film and recording the stats firsthand.

As far as what the websites will look like in the future, here are some tangibles that will be added to the dashboard: Date Range Slider so the coach/player can look solely at a certain set of practices, a "Show Previous Year" checkbox so the coach/player can look at this year's team compared to last year's team, being able to select two players at once displaying two comparative lines, more visuals besides just the line graphs such as cumulative bar charts and scatter plots, and lastly more statistical analyses such as correlation coefficients between variables, regression, box and whisker plots, etc. This project is not confined to the time restraints of this semester, and the goal for the end completion date is this May. That way client outreach can begin in the dead period between this season and off-season workouts. Hopefully, a small client base can be achieved in this window for the upcoming year.

References

[1] L. Bartram M. Tory A. Sarikaya, M. Correll and D. Fisher. What do we talk about when we talk about dashboards? *IEEE Transactions on Visualization and Computer Graphics*, 25(1):682–692, 2019.

[2] Noah Iliinsky and Julie Steele. Designing data visualizations: Representing informational relationships. 2011.

[3] Mahapatra R.P. Liu, A. and Mayuri. A.v.r. hybrid design for sports data visualization using ai and big data analytics. *Complex Intell*, Syst. 9:2969–2980, 2023.

[4] R. C. Basole and D. Saupe. Sports data visualization [guest editors' introduction]. *IEEE Computer Graphics and Applications*, 36(5):24–26, 2016.

[5] Delaney J. A. Duthie G. M. Thornton, H. R. and B. J. Dascombe. Developing athlete monitoring systems in team sports: Data analysis and visualization. *International Journal of Sports Physiology and Performance*, 14(6):698–705, 2023.

[6] et al. Hamdad, Leila. Basketball analytics. data mining for acquiring performances. *Computational Intelligence and Its Applications*, Proceedings 6, 2018.

[7] Vangelis Sarlis and Christos Tjortjis. Sports analytics—evaluation of basketball players and team performance. *Information Systems*, 93, 2020.

[8] Zachary Terner and Alexander Franks. Modeling player and team performance in basketball. *Annual Review of Statistics and Its Application*, 8:1–23, 2021.

[9] Stephen M. Shea and Christopher E. Baker. Basketball analytics: Objective and efficient strategies for understanding how teams win. *Advanced Metrics*, 2013.

[10] Rodolfo Metulini Manisera, Marica and Paola Zuccolotto. Basketball analytics using spatial tracking data. *New Statistical Developments in Data Science*, pages 305–318, 2019.