Automated Scouting Report Generation for the Adidas Next Generation Tournament (ANGT)

Aleix Vindel

June 2025

1 Introduction

“ *”Information is a source of learning. But unless it is organized, processed and available to the right people in a format for decision making, it is a burden, not a benefit.”* — William Pollard ”

In the current landscape of data science applied to sports, this statement resonates strongly. The availability of data does not guarantee competitive advantage of the game. In contrast, the lack of automated structures to organize and analyze data often makes information a burden for coaches, analysts, and technical staff. This challenge is particularly evident in development tournaments such as the **Adidas Next Generation Tournament (ANGT)**, where data access is limited, fragmented, and lacks standardization.

In these contexts, manual scraping of statistics — though widely used by scouts and agents — is a repetitive, non-reusable, and time-consuming task. This presents a significant obstacle for those who want to produce scouting reports based on easily shareable data.

This project proposes the development of a **tool aimed at automating the extraction and structuring of statistical data from matches and players in the ANGT competition.** Through a technical approach, the objective is to make the data available so that teams can turn them into useful information for decision making and performance analysis.

2 Context

This project is not designed for the exclusive benefit of a single club, but rather expects to provide a useful tool for all organizations participating in the Adidas Next Generation Tournament (ANGT). The ANGT is an international U18 basketball competition organized by Euroleague Basketball, in which only selected teams are invited to compete. These are typically elite-level clubs or academies with highly regarded prospects. In recent years, teams such as Žalgiris Kaunas, Real Madrid, Mega Basket, and FC Barcelona have been among the winners.

The structure of the competition consists of four qualifying tournaments held in different cities throughout the season. Each tournament features eight invited teams, divided into two groups of four. Teams compete within their groups, and the top teams from each group face off in a final to determine the winner.

The winners of the four qualifiers, along with additional invited teams, advance to a final tournament with the same group-stage format. The winner of this final event is crowned the ANGT champion for the season.

3 Justification and State of the Art

One of the persistent challenges in the field of sports analytics is how to store and structure data so that it becomes usable and meaningful. While access to performance data has increased over recent years, most organizations still lack the technical knowledge, infrastructure, or financial resources required to build and maintain data systems that allow for scalable and actionable use.

Furthermore, many professionals in the sector — including statisticians and mathematicians — often lack experience in areas such as software architecture, web scraping, and automated data pipelines. This technical gap creates a significant barrier when trying to develop tools that go beyond isolated analyses and move toward fully integrated, long-term solutions.

With these limitations in mind, this project proposes the creation of a customizable and user-friendly infrastructure for extracting, storing, and structuring ANGT data. The goal is to offer a flexible and accessible system that empowers analysts and coaches to focus on what truly matters: interpreting data and making informed decisions.

State of the Art

There is a wide range of existing tools and studies related to data scraping and analysis in other basketball contexts. In Spain, one of the most prominent initiatives was developed by Adrià Arbués: *Bue Stats*, a platform that provided users with a graphical interface to download and explore basketball statistics [Arbués, 2020]. Building upon that foundation, Nil Crespo later launched *Bue Stats 2.0*, expanding the functionality through a web-based interface that improved accessibility and performance [Crespo, 2023].

In one of my previous works, I also explored this concept by developing a scouting platform for FEB leagues (Spanish Basketball Federation), with a focus on structuring player data for easier evaluation and comparison [Vindel, 2022].

At the global level, numerous examples exist, including open-source repositories containing play-by-play data for professional leagues like the ACB [Díaz, 2021] or NBA [Shufinskiy, 2020]. In addition, various academic and technical works have analyzed how to structure sports data using spreadsheets or lightweight databases, emphasizing reproducibility and accessibility [Broman and Woo, 2005].

However, despite this body of work, there appear to be no publicly available tools or studies focused on the Adidas Next Generation Tournament (ANGT).

4 Objectives and Methodology

The project is structured around two main components: (1) the creation of scripts to extract and store play-by-play and box score data; and (2) the use of this data to generate detailed scouting reports.

This division is essential to ensure that the project is both customizable and reusable. Once the data collection is complete and the information is stored in structured files, these datasets become adaptable resources that teams can use according to their specific needs and priorities.

The first phase is primarily technical and will be most useful for professionals who may not have expertise in web scraping, but who do have a background in statistics or performance analysis. This part of the system will provide pre-processed and summarized data files, allowing users to bypass the complexity of raw data extraction.

The second phase is more flexible and oriented toward customization. It focuses on applying the previously collected data to produce tailored scouting reports. In this case, a demonstration will be developed to illustrate the analytical potential of combining play-by-play and box score data. This demo report will showcase how the data can be used to analyze a specific team's performance in depth, offering a tangible example of the project's practical utility.

General Objective

To develop an automated system for extracting, structuring, and analyzing data from the ANGT tournament to support scouting and performance analysis.

Specific Objectives

- To implement scraping scripts capable of retrieving box score and play-by-play data from ANGT sources.
- To store the extracted data in structured and reusable formats.
- To generate scouting reports using pre-processed datasets.
- To demonstrate the utility of the system through a case study on a specific team.

Methodology

The technical development of the project is divided into two main stages: data extraction and data analysis.

For the data extraction phase, the project will use **Python** as the primary programming language, alongside libraries such as **BeautifulSoup** and **Selenium**. These tools will be employed to scrape both box score and play-by-play statistics from publicly available sources related to the ANGT tournament. Selenium will allow interaction with dynamically loaded web elements, while BeautifulSoup will be used for parsing static HTML content.

The extracted data will be stored in structured **CSV files**, chosen for their simplicity, portability, and ease of use in a wide range of data analysis environments. Each game's box score and event log will be saved in separate files, with standardized formats to enable reuse across different analysis pipelines.

In the second phase, the project will use **Python** libraries such as **pandas**, **matplotlib**, and potentially **plotly** or **seaborn** to analyze and visualize the data. This will include both aggregated metrics and game-specific event flows. The goal is to translate raw statistical data into actionable insights through clear and interpretable visualizations and summaries.

The report generation will be based on this processed data, although the final visualization interface (e.g., web app, dashboard, or notebook) is still under consideration.

Timeline and Resource Planning

Timeline. The project is expected to be completed over a period of approximately 4 weeks, as the duration of the course, divided into three main phases:

- **Week 1:** Identification of relevant data sources, and definition of scraping targets and file structures.
- **Week 2:** Development and testing of the scraping scripts using Python, Selenium, and BeautifulSoup. The goal is to extract both box score and play-by-play data and store it in standardized CSV files.
- **Weeks 3–4:** Data cleaning, analysis, and visualization. During this phase, processed datasets will be used to create scouting summaries and performance reports. A case study based on a specific ANGT team will be developed as a demonstration of the system's utility.

Technical Resources. The project will be developed using a personal computer equipped with:

- Python 3.10+
- Libraries: Selenium, BeautifulSoup, pandas, matplotlib, seaborn, plotly
- Development tools: VSCode

Data will be stored locally in CSV format. Git may be used for version control.

Human Resources. The project will be developed individually by the author.

5 References

References

[Arbués, 2020] Arbués, A. (2020). Bue stats: A gui-based tool for advanced basketball statistics retrieval. https://www.upf.edu/web/adria-arbues/buestats.

[Broman and Woo, 2005] Broman, K. W. and Woo, K. M. (2005). Data organization in spreadsheets. *The American Statistician*, 59(1):1–6.

[Crespo, 2023] Crespo, N. (2023). Bue stats 2.0: A web application for basketball data analysis. https://buestats2.com.

[Díaz, 2021] Díaz, P. (2021). Acb database project. https://github.com/PabloDMC/ACB_DB/tree/main. GitHub repository.

[Shufinskiy, 2020] Shufinskiy, D. (2020). Nba data scraping and analytics tools. https://github.com/shufinskiy/nba_data. GitHub repository.

[Vindel, 2022] Vindel, A. (2022). Scouting platform for spanish feb basketball (feb). https://upcommons.upc.edu/handle/2117/379944?show=full).