

# pyHoops: A Python package for advanced basketball data analytics

Alessandro Bombelli<sup>1</sup>

<sup>1</sup> Air Transport & Operations, Faculty of Aerospace Engineering, Delft University of Technology

DOI: [10.21105/joss.01845](https://doi.org/10.21105/joss.01845)

## Software

- [Review](#) ↗
- [Repository](#) ↗
- [Archive](#) ↗

**Submitted:** 04 October 2019

**Published:** 30 October 2019

## License

Authors of papers retain copyright and release the work under a Creative Commons Attribution 4.0 International License ([CC-BY](#)).

## Summary

Advanced data analytics has received contrasting opinions within the sport community. One of the greatest American Football coaches of all times, Bill Belichick, stated that he uses analytics “less than zero” when preparing games (Bleacher Report, 2019). On the other end, the movie “Money Ball” stressed the necessity of advanced data analytics in baseball, i.e., sabermetrics, to improve team’s performances. Restricting the field of view to basketball, there exist, especially in the United States, a strong interest in using data science to enhance players’ performances. Europe, on the other hand, is still a step behind in this sense. While statistics on individual performances of players are generally available after games in the form of box scores, such statistics are generally miopic. In fact, being individual statistics, they cannot fully grasp the impact of a player on his/her team while on the court and, equally importantly, the impact on the opposing team.

Basketball games in most European leagues and in Euroleague (the most important European basketball competition) are paired with a play-by-play report that maps every action (field goal made or attempted, foul, assist, substitution) in the game. The goal of pyHoops is to process such play-by-play reports to extrapolate performance indices that map the impact of players or lineups on the both teams. As example, a player might be a prolific scorer, hence positively affecting his/her own offense, but a below-average defender, and thus negatively affect the team.

pyHoops is based on two major blocks: (1) web-parsing to translate play-by-play html tables into pandas databases, and (2) computation of statistics and performance indices. The web-parsing block easily stores play-by-play reports by automatically accessing the associated webpage. Since webpages (and html tables) vary according to the specific league, the web-parsing block is league-specific and should be modified according to the underlying html structure of the webpages of interest. As a last resort, play-by-play reports and boxscores can be directly copied from the webpage, saved as spreadsheets and loaded as databases to skip the web-parsing block. The second block uses play-by-play information to aggregate statistics for every distinct player and lineup, both for their own team and the opposing team.

Having now team-aggregated and not player-specific information, pyHoops can easily compute and compare, for example, the field goal percentage of the home team when a particular player was on the court, versus the field goal percentage of the away team when the same player was on the court. While each call of pyHoops focuses on a single game, season-wide statistics can be computed by running pyHoops every time a new game occurs and storing new information in a pandas database, with the overarching goal of creating season-specific databases.

We hope pyHoops can foster a better understanding and interest in advanced team-oriented basketball statistics within the European basketball movement. Although some works already exist in the academic literature addressing advanced basketball analytics for European basketball (Metulini, 2018; Travassos, Davids, Araújo, & Esteves, 2013), we believe this work to

be the first (i) to focus on aggregated team-oriented statistics rather than individual-oriented statistics, and (ii) to provide an open-source package to be used by team data analysts of basketball enthusiasts to perform such analyses.

## Acknowledgements

The contribution of Luca Cappelletti in the development phase of the package is warmly acknowledged.

## References

- Bleacher Report. (2019). Bleacher report website. <https://bleacherreport.com/articles/2855622-patriots-bill-belichick-uses-analytics-less-than-zero-in-decision-making>.
- Metulini, R. (2018). Players movements and team shooting performance: A data mining approach for basketball. *arXiv preprint arXiv:1805.02501*.
- Travassos, B., Davids, K., Araújo, D., & Esteves, T. P. (2013). Performance analysis in team sports: Advances from an ecological dynamics approach. *International Journal of Performance Analysis in Sport*, 13(1), 83–95. doi:[10.1080/24748668.2013.11868633](https://doi.org/10.1080/24748668.2013.11868633)