

# Data Engineer Coding Challenge

## Basketball player rankings

### Overview

In this challenge you will scrape data from a sports website, store the data and perform a set of operations on them. The website is <https://espn.com/nba>, where one can find information for all the NBA teams and their players.

### Tasks

1. Find all the teams who participated in the 2020-2021 playoffs. For these teams, create a scraper to collect and store each player in the roster of these teams.
2. For each player from step 1, collect and store their yearly salary.
3. Additionally, scrape the points scored, assists, rebounds, and time played in the post season for each player from step 1.
4. How would you use the data collected in step 2 to rank the players from best to worst? Decide on a metric and make a list of the top 10 players in the post season.
5. Create a Command Line tool that runs the above steps and prints out the top 10 players in a tab- or comma-separated manner.

### Notes

- You can use any programming language and libraries you are comfortable with
- You can store the scraped data any way you see fit (SQL db, non-relational db, json file)
- Do not spend more time than 3 hours on the challenge. If you haven't finished in 3 hours, we can discuss and give feedback on the completed/attempted steps
- For any questions/clarifications regarding the challenge, feel free to contact [christos.tsiapalis@diva-e.com](mailto:christos.tsiapalis@diva-e.com) (cc [maren.dubowy@diva-e.com](mailto:maren.dubowy@diva-e.com))