NBA ‘Bigs’ vs European Big Men

# Capstone Project 1 Proposal

## What is the problem you are trying to solve?

The NBA is widely considered to be the best basketball league in the world, attracting the best players, the biggest sponsorships and boasting the highest viewership. In the past the league had very few players from outside of US borders, however this has started to change in recent times as more and more players are coming over from European countries. In this project I will compare NBA centers originating from the US to the ‘European big men’ in the NBA. The goal of the project will be to quantifiably determine the strengths and weaknesses of each group and predict key performance metrics for each player.

## Who is your client and why do they care about this problem?

There are many possible clients. NBA scouts want to make sure the players they provide scouting reports on for their NBA teams are likely to be successful, this study would help them make a more informed selection. Another group of clients are NBA coaches, they are interested in the strengths and weaknesses of the players as that allows them to best utilise the players on the court. NBA teams are another obvious client, as the prediction of key performance metrics into the future allows the front office to make informed financial decisions regarding the contracts they are willing to offer their players. Players themselves are likely clients, as it allows them to figure out areas that they need to focus on and lets them work on their game in an informed way.

## What data are you going to use for this?

Basketball-referance.com provides large amounts of player, team and draft data. There is no API, so the information will need to be scraped. I will thus need to build a scraper in Python. If basketball-reference lacks required data, then I will make use of stats.nba.com. Given that we are comparing European players to non-European players in the NBA, in the case that a European player had begun his career elsewhere before being drafted into the NBA, that data will not be used as is would skew the results (given the lesser opposition and so on). Thus, NBA data suffices for this project.

## In brief, outline your approach to solving this problem

I’ve briefly outlined the data collection process above, however after importing the data, I will need to clean up the dataset (make sure units and data types are correct, deal with missing values, add ‘European’ and ‘non-European’ column to classify each player etc. etc.). Having cleaned up the data and removed unnecessary columns of data I will perform EDA and then check if ‘there is a difference’ between ‘European’ and ‘Non-European’ players, most likely using hypothesis testing. To predict future performance for each NBA center by season (PPG, APG, STL, REB), we will perform multiple linear regression algorithms (one for each stat).

## What are your deliverables?

- Python files

- Jupyter Notebook including detailed documentation and programming choice justification

- Possibly a video describing the project in detail.