# Capstone Project Proposal

Name: Philippe Tchokokam

## Objective

In recent times, the market for professional footballers has been extremely inflated. Ever since famous soccer player, Neymar Jr., transferred to PSG from F.C Barcelona in 2017 for 200 million euros, transfers for the most highly sought after players have reached fees that are uncomparable to those of the past. As a result, many teams overvalue their players and use these higher transfer deals as precedents in order to justify the fees set for their players. As of recently, there have been a number of high transfer fees for players whose performance don’t match the weight of their price tags. This project aims to use predictive modeling in order to determine what the correct transfer market values of professional football players should be

## Impact

This project will be able to provide for coaches, scouts, sporting directors, and other soccer club personnel the proper benchmarking of prices that should be associated with players. This benchmarking will allow these personnel to properly assess the market price of team members or potential interests thereby minimizing the loss in the case of poor judgement of the transfer market. If successful, this project could create more transparency by limiting the overestimation of players created by media speculation.

## Dataset(s)

## Getting performance metrics of players with <https://fbref.com/en/>

Scraping data for previous transfer market price history from <https://www.transfermarkt.com/>

## Approach

1. Collecting data from fbref and webscraping with transfermarkt
2. Feature scaling/engineering in order to derive the best variables for analysis
3. Perform EDA on a number of different variables such as age, past performance, position, league
4. Perform machine learning algorithms such as linear regression, random forests, and/or gradient boosting
5. Compare different models to determine the best model for the dataset.
6. Use RMSE to evaluate the performance of the model

## Timeline

This is a rough timeline for this project:

- (3 Weeks) Data Collection

- (3 Weeks) Data pre-processing and EDA

- (4 Weeks) Modeling

- (2 Weeks) Interpreting results, creating visualizations, error analysis

- (1 Weeks) Writing up report

- (2 Week) Poser and Final Presentation ]

## 

## Possible Issues

[List some of the prospective challenges and issues and discuss how you envision overcome them]

1. Feature Selection

Problem: A potential challenge can be the possibility of not choosing the best features to include, or rather not being able to quantify realistic factors such as player hype or team interest for certain players.

Solution: Focusing on the most quantifiable data and then using proxy variables like social media followers as “popularity”.

2. Domain Expertise

Problem: Although not a problem, it’s important to have a good understanding of soccer and the most important statistics

Solution: Conducting research on statistics by reading forums, blogs, and other sources might help reinforce knowledge on the most important

3. Project Scope

Problem: Taking a global approach might lead to potential problems during the data collection and modeling stages.

Solution: Starting small, with one league or few more might alleviate the strain on these steps. If proven to be successful, it would then be possible to open up the project to more leagues and countries.