

### Question/need:

- ❑ The goal of this project is to build a linear regression model that can predict market value of a soccer player.
- ❑ Being able to predict the market value of players will be essential for soccer clubs. Clubs pay an enormous amount of money for players and they will have to accurately know whether each player is worth the money so they can take careful consideration when buying/selling players for their club.

### Data Description:

- ❑ The dataset that I plan to use will come from 2 different sites:  
<https://www.premierleague.com/players> &  
<https://www.transfermarkt.com/wettbewerb/startseite/wettbewerb/GB1>  
. Using web scraping, all of the players from English Premier League will be collected from 2 seasons which sums up to roughly 1000+ observations. Similarly, market value of players will be collected from another site from 2 corresponding seasons.
- ❑ Some of the features that will be used to predict the market value of players will be:

Goals, Goals per match, Headed goals, Goals with right foot, Goals with left foot, Penalties scored, Freekicks scored, Shots, Shots on target, Shooting accuracy %, Hit woodwork, Big chances missed, Assists, Passes, Passes per match...etc

- ❑ The target for this model will be to predict the market value of each player.

### Tools:

- ❑ Use BeautifulSoup + Selenium to web scrape information needed from the site. The EPL site definitely needs Selenium for the scrolling part.

### MVP Goal:

- ❑ My goal is to be able to scrape all the information that I need, merge them and draw out the pair plot to see the overall correlation.