

Web Scraping a Football database

- Extracted data from TransferMarkt
- Aim: find bargain deals, based on players goals and assists in relation to MV

The data



- Extract goals and assists top 500 players
- Calculate overall score, assuming assists == goals

Goals scored + assists made = Score

- Ratio: market value per player/score
- Allowing to see which players underhyped, overhyped

Example 1 - Harry Kane

Goals: 5

Assists: 2

Harry Kane's Market Value: 100 million

Score: 5+2 = 7

MV/score = ratio = 100/7 = 14.28 million per goal contribution



• The higher the number, the worse the deal

Findings - overall

Findings - underrated players

Findings - overrated players

Difficulties

- Web page not web scraping friendly (Transfermarkt)

- Multiple links within links that needed to be accessed

Improvements

- Only worked for attackers, goals + assists
- Did not take into account appearances
- bad maths
- Hard code

Questions?