Training Data - Description

Each row represents one player's performance in a single fixture, and will be unique across the player's name, their team, and kickoff time fields:

* player (player name)
* team (the player's team)
* kickoff\_time (kickoff time for the fixture)

The fixtures are further defined with the following fields:

* opponent\_team (the opposition team)
* was\_home (was it a home game for the player)
* season (e.g., '1920' for the 2019/20 season)
* gw (the FPL gameweek in which the fixture occured)
* crowds (were there crowds present at the match)

There is evidence that the lack of crowds during the Covid-19 pandemic reduced the home advantage, so I added a field for this. All games played between the 15th of March 2020 and 17th June 2021 are marked as having no crowds (in practice some fans were allowed in certain stadiums in November 2020, but it's relatively few matches with very few fans so I just left them all as false).

Note that there can be multiple fixtures (i.e. rows for a given player) in a single gameweek - so called double or triple gameweeks.

The position that a player plays is also given, this will be consistent for each player within seasons, but may change between seasons:

* position (1 - goalkeeper, 2 - defender, 3 - midfielder, 4 - forward)

There are then fields with the player (or team's) FPL metrics for fixture e.g., the number of minutes played, points scored, assists, goals, goals conceded while on the field, etc. Anything that is used in the game. This should be 100% complete for all rows.

There are also further stats from the start of the 2017/18 season, taken from sources outside of the FPL game, such as expected goals, expected assists, passes, dribbles, interceptions, etc. These will only be present if the player played (i.e., had at least 1 in the minutes field) in the fixture, otherwise the fields are null.

Other incomplete fields for FPL data are:

* transfer and selected values (transfers\_in, transfers\_out, transfers\_balance, selected) - these were only collected from the start of the 2019/20 season, and require further investigation as to what they actually represent (in other words, treat with caution when modelling); values prior to the 2019/20 are set to 0
* play\_proba - again only collected from the start of the 2019/20 season, this is the probability that the player would actually be available for the fixture according to the FPL website (note that the time that this is captured each week varies); values prior to the 2019/20 are null, and they are also null for any new players in a given gameweek (i.e., players that FPL has added to the game during that gameweek)

Finally, team transfer market value is taken from transfermarkt each week (from the 2019/20 season onwards) or a single value has been taken for the whole season:

* relative\_market\_value\_team - the market value for the team taken during that gameweek (non-null from start of 2019/20 season)
* relative\_market\_value\_opponent\_team - the market value for the opposition team taken during that gameweek (non-null from start of 2019/20 season)
* relative\_market\_value\_team\_season - a single value for the team's value from the start of each season
* relative\_market\_value\_opponent\_team\_season - a single value for the opposition team's value from the start of each season