FPL Solver – Model Doc



Contents

[What we are trying to model 3](#_Toc106208822)

[Expected Points 3](#_Toc106208823)

[Player Value 3](#_Toc106208824)

[Goalkeeper Model 3](#_Toc106208825)

[Defender Model 3](#_Toc106208826)

[Multiple Model Considerations 3](#_Toc106208827)

[Data Dictionary 4](#_Toc106208828)

[Exploratory Analysis 8](#_Toc106208829)

[xValue Model 8](#_Toc106208830)

# What we are trying to model

Pull the superset joined data together using datautils. get\_gw\_pl\_superset()

## Expected Points

The dependent variable is ‘total\_points\_gw’, also a separate model will be required to predict total value <field required here> - so we can try and figure out who is under and over valued and trade accordingly.

There are two datasets. The “player aggregate data” (which is currently already pulled as part of the superset, and the “understat summary data” which is not.

There are a lot of gameweek-level datapoints which could be used to model this – there is a table below in the below section which looks at the inputs.

Understats are done on an aggregate level but there is a week by week breakdown when you access the csv file named after the specific player (which goes back very far, as long as each player is in the PL).

## Player Value

Player value will be run on end of season aggregate stats so we can get a feel for how certain positions and certain metrics affect this, and we can attempt to get a feel for who is under-overpriced thus.

## Goalkeeper Model

It may be necessary to strip out goalkeepers from the dataset and create a bespoke model with specific fields just for them as there are a lot of metrics that work **specifically** for goalkeepers. There may need to be a separate model for Defenders. Not sure if this is to be lumped in with the GK model?

## Defender Model

There is very little correlation between the datapoints selected for the xPoints model which applies to defenders in the same way they do for midfielders and forwards. Some thought is going to have to be put into this.

## Multiple Model Considerations

What are the performance issues w.r.t using multiple models in an optimizer scenario?

Is there any way to collapse these into one model (would you even want to do that?)

## FDR Modelling

This is going to have to replace the massive amount of columns “team” and “opposition team” does. Big time dimensionality reduction necessary here.

# Data Dictionary

DataHelper has a function called get\_gw\_pl\_us\_superset() which has a gameweek-level breakdown. This will be used to model the expected points output on the back of xG, xA, whatever else. Rows in blue would be good for reports and visualisations.

| Field Name | Definition | XPoints Model | Value Model | GK Model | Viz |
| --- | --- | --- | --- | --- | --- |
| player\_id | Unique key | No | No | No | No |
| player\_name\_us | Player name in understats | No | No | No | No |
| goals | Goals scored | No | No | No | Yes |
| shots | Shots taken | No | No | No | Yes |
| xG | Expected goals | Yes | Yes | No | No |
| time | Minutes played | No | No | No | Yes |
| position | Position (understats) | No | No | No | No |
| h\_team | Home team | No | No | No | No |
| a\_team | Away Team | No | No | No | No |
| h\_goals | Home goals | No | No | No | No |
| a\_goals | Away goals | No | No | No | No |
| date | Date played | No | No | No | No |
| id | Some key | No | No | No | No |
| season | Season year | No | No | No | No |
| roster\_id | Some Key | No | No | No | No |
| xA | Expected Assists | Yes | Yes | No | Yes |
| assists\_us | Assists, understats | No | No | No | No |
| key\_passes | Key passes | Yes | Yes | No | Yes |
| npg | Non penalty goals | No | No | No | Yes |
| npxG | Non penalty goals xG | No | No | No | Yes |
| xGChain | When a player has played a role in a shot being taken. Might imply thread / importance in team. Such as fullbacks. Must plot too. | Yes | Yes | No | Yes |
| xGBuildup | More focus on the buildup. Will be good to plot reports on this but not necessarily model. Must plot. | No | Yes | No | Yes |
| player\_name\_pl | Player name, per FPL | No | No | No | Yes |
| assists\_pl | Assists | No | No | No | Yes |
| bonus | Bonus points | No | No | No | Yes |
| bps | [Bonus points system defined here. Will be worth plotting https://www.premierleague.com/news/106533#:~:text=Bonus%20points%20are%20a%20key,performance%20score%20for%20every%20player.](https://www.premierleague.com/news/106533#:~:text=Bonus%20points%20are%20a%20key,performance%20score%20for%20every%20player) | No | No | No | Yes |
| clean\_sheets | Clean sheets | No | No | Yes | Yes |
| creativity | Creativity, per FPL | Yes | No | No | No |
| element | ? | No | No | No | No |
| fixture | Fixture foreign key? | No | No | No | No |
| goals\_conceded | Goals conceded | No | No | No | Yes |
| goals\_scored | Goals conceded | No | No | No | No |
| ict\_index | Influence, creativity and threat index. Would prefer these metrics separately | No | No | No | Yes |
| influence | Influence per ICT | Yes | Yes | No | Yes |
| kickoff\_time | Kickoff time | No | No | No | No |
| minutes | Minutes played | No | No | No | Yes |
| opponent\_team | Foreign key | No | No | No | No |
| own\_goals | Own goals | No | No | No | Yes |
| penalties\_missed | Pens missed | No | No | No | Yes |
| penalties\_saved | Pens saved | No | No | Yes | Yes |
| red\_cards | Red cards | No | No | No | No |
| round | GW Round | No | No | No | No |
| saves | Saves | No | No | Yes | Yes |
| selected | Number selected. Good to plot. | No | No | No | No |
| team\_a\_score | Away goals | No | No | No | No |
| team\_h\_score | Home goals | No | No | No | No |
| threat | Threat, per FPL. | Yes | Yes | No | Yes |
| total\_points | Total points – dependent variable! | Yes | No | Yes | Yes |
| transfers\_balance | Net transfers | No | No | No | Yes |
| transfers\_in | Transfers In. | No | No | No | Yes |
| transfers\_out | Transfers Out | No | No | No | Yes |
| value | Value in gameweek. | No | Yes | No | Yes |
| was\_home | Bool, was home or not. | Yes | No | No | No |
| yellow\_cards | Yellow cards | No | No | No | Yes |
| kickoff\_date | Kick off date | No | No | No | No |
| team | Player’s team | Yes | Yes | Yes | Yes |
| opposition\_team | Player’s opposition team | Yes | Yes | Yes | Yes |
| xP | Expected Passes | Yes | Yes | No | Yes |

# Exploratory Analysis

## xValue Model

