**FPL Points Prediction Tool**

Developed by: Nick Fisher

*Introduction*

This documentation covers how to use the FPL Predicted Points tool. This tool is designed to integrate statistical information on past seasons to predict points for the current season. This tool is designed to also integrate for new players entering the league, applying penalties based on league strength using the FiveThirtyEight Global Club Rankings.[[1]](#footnote-0) The repository for this tool can be found at the following link:

Running this tool requires the following software:

* R: This tool was developed using R version 4.1.2. R can be installed here: <https://cran.r-project.org/bin/windows/base/>
* RStudio: A visual engine for running R. Can be installed here: <https://posit.co/download/rstudio-desktop/>
* The following R packages are required for running the markdown file:
  + worldfootballR
  + rstudioapi
  + tidyverse
  + kableExtra
  + ggplot2
  + corrplot
  + RColorBrewer
  + openxlsx
  + broom
    - To install these packages, use the following command for each package before running the Rmarkdown file:

install.packages(“[insert package name]”)

Before running any scripts, open “FPL.Rproj”, then choose your subsequent script or markdown file from the provided directories. It is not necessary to edit the directory structure.

*Step 1: Run the scripts for extracting statistics (ALREADY RUN)*

In the submodules folder are sub-folders for each statistic type. NOTE: These scripts do not to be run, they have already been run and data has been extracted for seasons 2013/14 through 2022/23. These should only serve as templates for extracting further data in the future for subsequent models.

These scripts utilize the worldfootballR package to extract data for standard, shooting, defense, passing, possession, and goalkeeping statistics stored in the free data site FBref.[[2]](#footnote-1) Clean sheet and goals against data were borrowed from previous analyses.[[3]](#footnote-2) There are currently pre-extracted CSV files for seasons dating back to 2013/14. As noted above, new seasons will need to be extracted for future models. It is recommended to not perform this extraction until the end of the current season.

Note only 2017/18 - 2022/23 data is used in the model as passing and possession data are only available for those seasons.

*Step 2: Run combine data scripts (ALREADY RUN)*

The script “combine\_data.R” will combine the various statistics into one CSV file for outfield players and keepers. This has already been run and produced a CSV output for use in training the model.

*Step 3: Get current rosters (ALREADY RUN)*

The script “get\_current\_rosters.R” can be used to create a CSV file of FBref for data for current rosters. This file integrates the FiveThirtyEight SPI to determine projected points for incoming players.

This is a lengthy file that contains code from worldfootballR that must be uncommented to run. It is highly recommended that you save a CSV copy of the data for each statistic once you have run the worldfootballR section (ex. Once the code for extracting shooting data for each player URL is complete, save the file as a CSV). This provides an avenue for much easier access the data, as the worldfootballR commands will take a long time to run. The current code includes both:

* Commented sections for worldfootballR commands;
* And commands to save the output as CSV files, and then load the subsequent file from the directory structure.

This script has already been run and an up-to-date CSV file produced for use in the model. It is recommended that this script is periodically during transfer windows, and at the end of each transfer window to maintain up-to-date predictions. Future work on this script includes adjustments for winter transfer windows, so predictions will only include the half-season a player will feature for a team.

*Step 4: Run the predicted points R markdown file (ALREADY RUN)*

This file evaluates a series of models and their validity and effectiveness at predicting FPL points for the current season using a variety of regression techniques. This model can only be run once the CSV and Excel outputs from the above files have been produced. It is recommended that this process be edited only after each FPL season, as that is when it will be re-evaluated.

This will produce the file “final\_predictions.xlsx”, which contains predicted points for the current FPL season based on previous historic player performance. This process is still being refined to more accurately model player results, as there a still a small number of unexpected or nonsensical predictions.

1. <https://projects.fivethirtyeight.com/soccer-predictions/global-club-rankings/> [↑](#footnote-ref-0)
2. <https://fbref.com/en/> [↑](#footnote-ref-1)
3. <https://github.com/vaastav/Fantasy-Premier-League/tree/master> [↑](#footnote-ref-2)