

FPL points predictor Data Import and Wrangling 2

This is a file for wrangling the second set of collected datasets

Importing the datasets on the premier league players

```
library(readxl)
FPL_201617 <- read_excel("~/DSI-SRP FILES/FPL 2016-17.xlsx")
FPL_201718 <- read_excel("~/DSI-SRP FILES/FPL 2017-18.xlsx")
FPL_201819 <- read_excel("~/DSI-SRP FILES/FPL 2018-19.xlsx")
Prem_201617 <- read_excel("~/DSI-SRP FILES/Premier League 2016-17.xlsx")
Prem_201718 <- read_excel("~/DSI-SRP FILES/Premier League 2017-18.xlsx")
Prem_201819 <- read_excel("~/DSI-SRP FILES/Premier League 2018-19.xlsx")
```

Loading tidyverse

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --

## v ggplot2 3.3.0      v purrr  0.3.4
## v tibble  3.0.1      v dplyr  0.8.5
## v tidyr   1.0.3      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Merging the datasets with the use of inner_join

```
Prem_201617 <- as_tibble(Prem_201617)
play17 <- inner_join(Prem_201617, FPL_201617, by = c(player_name = "Full Name"))
play18 <- inner_join(Prem_201718, FPL_201718, by = c(player_name = "Full Name"))
play19 <- inner_join(Prem_201819, FPL_201819, by = c(player_name = "Full Name"))
```

Transforming elements in the position variable into factors

```
play17$position <- as_factor(play17$position)
levels(play17$position)
```

```
## [1] "Goalkeeper"      "Centre-Back"      "Left-Back"
## [4] "Right-Back"      "Defensive Midfield" "Central Midfield"
## [7] "Attacking Midfield" "Left Winger"      "Right Winger"
## [10] "Centre-Forward"   "Left Midfield"    "Right Midfield"
## [13] "Midfielder"      "Second Striker"   "Forward"
```

```
play18$position <- as_factor(play18$position)
levels(play18$position)
```

```
## [1] "Goalkeeper"      "Centre-Back"      "Left-Back"
## [4] "Right-Back"      "Defensive Midfield" "Central Midfield"
## [7] "Left Midfield"    "Left Winger"      "Right Winger"
## [10] "Centre-Forward"   "Attacking Midfield" "Right Midfield"
## [13] "Midfielder"      "Forward"          "Second Striker"
```

```
play19$position <- as_factor(play19$position)
levels(play19$position)
```

```
## [1] "Goalkeeper"      "Centre-Back"      "Left-Back"
## [4] "Right-Back"      "Defensive Midfield" "Central Midfield"
## [7] "Attacking Midfield" "Left Winger"      "Right Winger"
## [10] "Centre-Forward"   "Left Midfield"    "Right Midfield"
## [13] "Second Striker"
```

Creating function called `position_indexer` that builds a new column called `position_index` which indexes the position of the players.

```
position_indexer <- function(df) {
  df %>%
    mutate(position_index = if_else(position %in% c("Goalkeeper"),
1,if_else(position %in% c("Centre-Back", "Left-Back", "Right-Back"),
2,if_else(position %in% c("Defensive Midfield", "Central Midfield", "Left Midfield",
"Right Midfield", "Midfielder"),
3,if_else(position %in% c("Left Winger","Right Winger","Attacking Midfield"),
4,if_else(position %in% c("Forward", "Second Striker", "Centre-Forward"),5,0))))))
}
```

Applying position indexer to the three datasets

```
play17 <- position_indexer(play17)
play18 <- position_indexer(play18)
play19 <- position_indexer(play19)
```

Saving the datasets as csv files

```
write_csv(play17, "season17.csv")  
write_csv(play18, "season18.csv")  
write_csv(play19, "season19.csv")
```