FPL points predictor Data Import and Wrangling

This is a file for wrangling the collected datasets

FIFA 17, FIFA 18 and FIFA 19 files which contained players' attributes were first imported.

```r
FIFA.17 <- read.csv("~/DSI-SRP FILES/FIFA 17.csv")
FIFA.18 <- read.csv("~/DSI-SRP FILES/FIFA 18.csv")
FIFA.19 <- read.csv("~/DSI-SRP FILES/FIFA 19.csv")
```

Fantasy Premier League files from the season of 2016-17 to 2018-19 which contained players' attributes and performance were also imported.

```r
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")
```

The imported datasets were wrangled in order to include important rows and columns only

```r
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()
```

```r
create_new <- function(df) {
  df %>%
    mutate(Full_Name = as.character(Full.Name)) %>%
    select(Full_Name, age, overall, potential, player_positions, club)
```

```
}

FIFA.17_new <- create_new(FIFA.17)
FIFA.18_new <- create_new(FIFA.18)
FIFA.19_new <- create_new(FIFA.19)
```

The newly formed FIFA datasets were merged with the FPL datasets

```
season19 <- inner_join(FIFA.19_new, FPL_201819, by = c(Full_Name = "Full Name"))
season18 <- inner_join(FIFA.18_new, FPL_201718, by = c(Full_Name = "Full Name"))
season17 <- inner_join(FIFA.17_new, FPL_201617, by = c(Full_Name = "Full Name"))
```

Further wrangling was used to separate the player positions' columns into multiple columns. This made players with multiple positions have one position.

```
season17 <- season17 %>%
  separate(player_positions, into = c("Position1", "Position2", "Position3"), sep = ",")
season18 <- season18 %>%
  separate(player_positions, into = c("Position1", "Position2", "Position3"), sep = ",")
season19 <- season19 %>%
  separate(player_positions, into = c("Position1", "Position2", "Position3"), sep = ",")
```

The two additional columns were deleted in order to get rid of missing values

```
remove_2p <- function(df) {
  df %>%
    select(-Position2,-Position3)
}
season17 <- remove_2p(season17)
season18 <- remove_2p(season18)
season19 <- remove_2p(season19)
```

Converting the players' positions' data type to factor and obtaining the unique levels of player positions

```
season17$Position1 <- as_factor(season17$Position1)
season18$Position1 <- as_factor(season18$Position1)
season19$Position1 <- as_factor(season19$Position1)
levels(season17$Position1)
```

```
##  [1] "GK"  "CM"  "CAM" "LM"  "CB"  "RM"  "ST"  "LB"  "CDM" "RB"  "LW"  "RW"
## [13] "LWB"
```

```
levels(season18$Position1)
```

```
##  [1] "LW"  "GK"  "CAM" "ST"  "CM"  "CB"  "RM"  "CDM" "RB"  "LM"  "RW"  "LB"
## [13] "RWB" "LWB"
```

```
levels(season19$Position1)
```

```
##  [1] "CAM" "LW"  "GK"  "ST"  "CM"  "CB"  "CDM" "RW"  "RB"  "CF"  "LB"  "RM"
## [13] "LM"  "LWB" "RWB"
```

Creating a function that indexes players' positions

```
position_indexer <- function(df) {
  df %>%
    mutate(position_index = if_else(
      Position1 == "GK", '1',if_else(Position1 %in% c("CB", "LWB", "RWB", "LB", "RB"),
'2', if_else(Position1 %in% c("CM", "CDM", "RM", "CAM", "LM"),
'3', if_else(Position1 %in% c("RW", "LW", "ST", "CF"),
'4', '0')))))
}
```

Applying the function above to the datasets

```
season17 <- position_indexer(season17)
season18 <- position_indexer(season18)
season19 <- position_indexer(season19)
```

Converting position index data type to double

```
season17$position_index <- as.double(season17$position_index)
season18$position_index <- as.double(season18$position_index)
season19$position_index <- as.double(season19$position_index)
```

Saving files as csv files

```
write_csv(season17, "players17.csv")
write_csv(season18, "players18.csv")
write_csv(season19, "players19.csv")
```