Premier League Players Data Cleaning

Byron Caviedes

2023-05-29

Table of Contents

[Installing Packages we are be using 1](#_Toc136276072)

[Procede to upload the xslx file using read\_xlsx 2](#_Toc136276073)

[Dataset first sight 2](#_Toc136276074)

[Cleaning names 3](#_Toc136276075)

[Separate abbreviations in the country column and age of the players 3](#_Toc136276076)

[Changing data type of the age from char into numeric 3](#_Toc136276077)

[Graph of the goals converted by position in each team using ggplot and labs 5](#_Toc136276078)

[Goals converted by the first 6 teams in the ranking table using pipelines to filter 5](#_Toc136276079)

In this document I will apply some functions and packages to upload, clean, modify and download the data of the Premier League Players

# Installing Packages we are be using

install.packages("tidyverse", repos = "http://cran.us.r-project.org")

## package 'tidyverse' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\USUARIO\AppData\Local\Temp\RtmpclOPMp\downloaded\_packages

library(tidyverse)  
  
install.packages("readxl", repos = "http://cran.us.r-project.org")

## package 'readxl' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\USUARIO\AppData\Local\Temp\RtmpclOPMp\downloaded\_packages

library(readxl)  
  
install.packages("skimr", repos = "http://cran.us.r-project.org")

## package 'skimr' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\USUARIO\AppData\Local\Temp\RtmpclOPMp\downloaded\_packages

library(skimr)  
  
install.packages("janitor", repos = "http://cran.us.r-project.org")

## package 'janitor' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\USUARIO\AppData\Local\Temp\RtmpclOPMp\downloaded\_packages

library(janitor)  
  
install.packages("ggplot2", repos = "http://cran.us.r-project.org")  
library(ggplot2)  
  
install.packages("xlsx", repos = "http://cran.us.r-project.org")

## package 'xlsx' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\USUARIO\AppData\Local\Temp\RtmpclOPMp\downloaded\_packages

library("xlsx")

## Procede to upload the xslx file using read\_xlsx

players\_data <- read\_xlsx("C:\\Users\\USUARIO\\Desktop\\DATA ANALYTICS\\PORTAFOLIO\\PREMIER LEAGUE DATA\\Estadística\_jugadores.xlsx")

## Dataset first sight

head(players\_data)

## # A tibble: 6 × 38  
## Clasificación Jugador País Posc Equipo Edad Nacimiento PJ Titular Mín  
## <dbl> <chr> <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 1 Brenden… us U… CCDL Leeds… 22-2… 2000 35 28 2340  
## 2 2 Che Ada… sct … DL South… 26-3… 1996 28 23 1992  
## 3 3 Tyler A… us U… CC Leeds… 24-0… 1999 24 24 2156  
## 4 4 Tosin A… eng … DF Fulham 25-2… 1997 24 22 1995  
## 5 5 Nayef A… ma M… DF West … 27-0… 1996 17 16 1505  
## 6 6 Naouiro… fr F… CCDL Cryst… 21-0… 2002 7 0 93  
## # ℹ 28 more variables: `Minutos\_jugados/90` <dbl>, Goles <dbl>,  
## # Asistencias <dbl>, Goles\_mas\_asistencias <dbl>,  
## # Goles\_Sin\_Penalización <dbl>, Tiros\_Penales <dbl>,  
## # Tiros\_Penales\_Intentados <dbl>, Tarjetas\_Amarillas <dbl>,  
## # Tarjetas\_Rojas <dbl>, Goles\_Esperados <dbl>,  
## # `Goles\_Esperados\_-Penaltis` <dbl>, `Goles\_esperados\_+Asistencia` <dbl>,  
## # `Goles\_esperados\_-Penales\_+Asistencia` <dbl>, Acarreos\_progresivos <dbl>, …

## Cleaning names

players\_data\_clean <- clean\_names(players\_data)

## Separate abbreviations in the country column and age of the players

players\_data\_clean <- players\_data\_clean %>%   
 separate(pais, c("siglas", "pais\_siglas"), " ")  
  
players\_data\_clean <- players\_data\_clean %>%  
 separate(edad, c("edad", NA), sep = 2)

## Changing data type of the age from char into numeric

str(players\_data\_clean, 6)

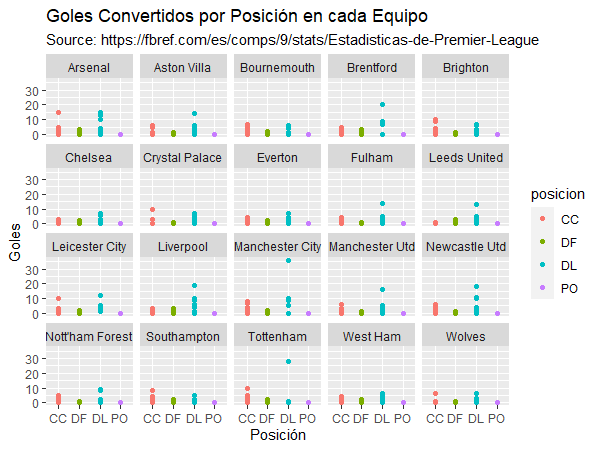
## tibble [562 × 39] (S3: tbl\_df/tbl/data.frame)  
## $ clasificacion : num [1:562] 1 2 3 4 5 6 7 8 9 10 ...  
## $ jugador : chr [1:562] "Brenden Aaronson" "Che Adams" "Tyler Adams" "Tosin Adarabioyo" ...  
## $ siglas : chr [1:562] "us" "sct" "us" "eng" ...  
## $ pais\_siglas : chr [1:562] "USA" "SCO" "USA" "ENG" ...  
## $ posc : chr [1:562] "CCDL" "DL" "CC" "DF" ...  
## $ equipo : chr [1:562] "Leeds United" "Southampton" "Leeds United" "Fulham" ...  
## $ edad : chr [1:562] "22" "26" "24" "25" ...  
## $ nacimiento : num [1:562] 2000 1996 1999 1997 1996 ...  
## $ pj : num [1:562] 35 28 24 24 17 7 20 9 29 25 ...  
## $ titular : num [1:562] 28 23 24 22 16 0 9 9 24 21 ...  
## $ min : num [1:562] 2340 1992 2156 1995 1505 ...  
## $ minutos\_jugados\_90 : num [1:562] 26 22.1 24 22.2 16.7 1 11.5 8.5 25.4 20.1 ...  
## $ goles : num [1:562] 1 5 0 1 2 0 1 0 0 1 ...  
## $ asistencias : num [1:562] 3 3 0 0 0 0 0 0 1 0 ...  
## $ goles\_mas\_asistencias : num [1:562] 4 8 0 1 2 0 1 0 1 1 ...  
## $ goles\_sin\_penalizacion : num [1:562] 1 5 0 1 2 0 1 0 0 1 ...  
## $ tiros\_penales : num [1:562] 0 0 0 0 0 0 0 0 0 0 ...  
## $ tiros\_penales\_intentados : num [1:562] 0 0 0 0 0 0 0 0 0 0 ...  
## $ tarjetas\_amarillas : num [1:562] 2 1 6 4 4 1 4 0 2 2 ...  
## $ tarjetas\_rojas : num [1:562] 0 0 1 0 0 0 0 0 0 0 ...  
## $ goles\_esperados : num [1:562] 3.9 6.6 0.1 0.7 2.2 0.2 0.7 0 1.3 1.6 ...  
## $ goles\_esperados\_penaltis : num [1:562] 3.9 6.6 0.1 0.7 2.2 0.2 0.7 0 1.3 1.6 ...  
## $ goles\_esperados\_asistencia : num [1:562] 4.2 2.8 1.4 0.4 0.5 0.1 1 0 0.4 0.6 ...  
## $ goles\_esperados\_penales\_asistencia : num [1:562] 8.1 9.4 1.5 1.1 2.7 0.2 1.7 0.1 1.7 2.2 ...  
## $ acarreos\_progresivos : num [1:562] 43 17 19 16 2 1 30 10 24 25 ...  
## $ pases\_progresivo : num [1:562] 86 34 136 92 24 2 43 31 116 139 ...  
## $ pases\_progresivo\_rec : num [1:562] 149 115 20 3 4 11 76 12 17 30 ...  
## $ goles\_90 : num [1:562] 0.04 0.23 0 0.05 0.12 0 0.09 0 0 0.05 ...  
## $ asistencias\_90 : num [1:562] 0.12 0.14 0 0 0 0 0 0 0.04 0 ...  
## $ goles\_asistencias\_90 : num [1:562] 0.15 0.36 0 0.05 0.12 0 0.09 0 0.04 0.05 ...  
## $ goles\_penales\_90 : num [1:562] 0.04 0.23 0 0.05 0.12 0 0.09 0 0 0.05 ...  
## $ goles\_penalizacion\_asistencias\_90 : num [1:562] 0.15 0.36 0 0.05 0.12 0 0.09 0 0.04 0.05 ...  
## $ goles\_esperados\_90 : num [1:562] 0.15 0.3 0 0.03 0.13 0.15 0.06 0 0.05 0.08 ...  
## $ asistencias\_esperadas\_90 : num [1:562] 0.16 0.13 0.06 0.02 0.03 0.07 0.09 0 0.02 0.03 ...  
## $ goles\_esperados\_asistencias\_esperadas\_90: num [1:562] 0.31 0.42 0.06 0.05 0.16 0.22 0.15 0.01 0.07 0.11 ...  
## $ goles\_esperados\_penales\_90 : num [1:562] 0.15 0.3 0 0.03 0.13 0.15 0.06 0 0.05 0.08 ...  
## $ goles\_esperados\_penales\_asistencias\_90 : num [1:562] 0.31 0.42 0.06 0.05 0.16 0.22 0.15 0.01 0.07 0.11 ...  
## $ partidos : chr [1:562] "Partidos" "Partidos" "Partidos" "Partidos" ...  
## $ id\_unica : chr [1:562] "5bc43860" "f2bf1b0f" "2b09d998" "c81d773d" ...

players\_data\_clean <- players\_data\_clean %>%   
 transform(edad = as.numeric(edad))

## Graph of the goals converted by position in each team using ggplot and labs

CODE: ggplot(players\_data\_clean, aes(x = posicion, y = goles, color = posicion)) + geom\_point() + facet\_wrap(~equipo) + labs(x = “Posición”, y = “Goles”, title = “Goles Convertidos por Posición en cada Equipo”, subtitle = “Source: <https://fbref.com/es/comps/9/stats/Estadisticas-de-Premier-League>”)

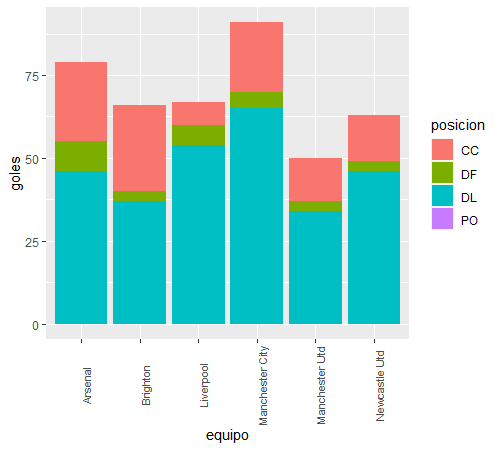
knitr::include\_graphics("C:\\Users\\USUARIO\\Desktop\\DATA ANALYTICS\\PORTAFOLIO\\PREMIER LEAGUE DATA\\images/goalsTeam.png")



## Goals converted by the first 6 teams in the ranking table using pipelines to filter

CODE: players\_data\_clean %>% filter(equipo == “Arsenal” | equipo == “Manchester City” | equipo == “Manchester Utd” | equipo == “Liverpool” | equipo == “Newcastle Utd” | equipo == “Brighton”) %>% ggplot() + geom\_col(aes(x = equipo, y = goles, fill = posicion))

knitr::include\_graphics("C:\\Users\\USUARIO\\Desktop\\DATA ANALYTICS\\PORTAFOLIO\\PREMIER LEAGUE DATA\\images/bigsixg.png")

 ## Download the clean table as xlsx format for further use

write.xlsx(players\_data\_clean, "C:/Users/USUARIO/Desktop/pl\_data.xlsx", sheetName="Sheet1")