0\_clean.r

abhia

2019-11-18

# IMPORT BY USING THE SUITABLE ENCODING  
fifa = read.csv("raw\_downloaded\_data.csv",fileEncoding="UTF-8-BOM")  
  
#LIBRARY TO PLAY WITH STRINGS  
library("stringr")  
  
#TO PREVENT SCIENTIFIC NOTATION  
options(scipen = 999)  
  
#REMOVE "€" FROM THE CURRENCY COLUMNS  
#RUN IN TERMINAL   
fifa$Release.Clause = lapply(fifa$Release.Clause, function(x){gsub("€", "", x)})  
fifa$Value = lapply(fifa$Value, function(x){gsub("€", "", x)})  
fifa$Wage = lapply(fifa$Wage, function(x){gsub("€", "", x)})  
  
#CONVERT HEIGHTS FROM INCHES TO CENTIMETERS  
fifa$Height = lapply(fifa$Height, function(i) {  
 a = strsplit(as.character(i), "[[:punct:]]")[[1]][1]  
 b = strsplit(as.character(i), "[[:punct:]]")[[1]][2]  
 return (as.integer(a)\*30.48 + as.integer(b)\*2.54)})  
  
#CONVERT CURRENCY(FROM 'K' AND 'M') TO NUMBERS  
  
km <- function(i){  
 if(isTRUE(str\_detect(i, "K"))){  
 a= strsplit(as.character(i[[1]][1]), "[[:upper:]]")  
 b = as.double(a[[1]][1])  
 c = b\*1000  
 return(c)  
 } else if (isTRUE(str\_detect(i, "M"))){  
 a= strsplit(as.character(i[[1]][1]), "[[:upper:]]")  
 b = as.double(a[[1]][1])  
 c = b\*1000000  
 return(c)  
 } else{  
 return(i)}}  
  
fifa$Value = lapply(fifa$Value, km)  
fifa$Wage = lapply(fifa$Wage, km)  
fifa$Release.Clause = lapply(fifa$Release.Clause, km)  
  
#CONVERT WEIGHTS FROM LBS TO KG  
fifa$Weight = lapply(fifa$Weight, function(i) {  
 a= strsplit(as.character(i[[1]][1]), "[[:lower:]]")  
 b = as.double(a[[1]][1])  
 c = round(b\*0.45359237, digits=2)  
 return(c)})  
  
  
#VIEW DATA FRAME  
View(fifa)  
  
  
  
#CONVERT LIST(EVERY CELL) TO NUMERIC  
fifa$Value = unlist(lapply(fifa$Value, as.numeric))  
fifa$Wage = unlist(lapply(fifa$Wage, as.numeric))  
fifa$Height = unlist(lapply(fifa$Height, as.numeric))  
fifa$Weight = unlist(lapply(fifa$Weight, as.numeric))  
fifa$Release.Clause = unlist(lapply(fifa$Release.Clause, as.numeric))  
  
  
  
  
#WRITE THE CLEANED DATA INTO A NEW FILE  
write.csv(fifa, "0\_raw\_downloaded\_data.csv", row.names = FALSE)