1\_cleaning\_data.r

abhia

2019-11-17

# IMPORT BY USING THE SUITABLE ENCODING  
fifa = read.csv("0\_raw\_downloaded\_data.csv")  
  
# DELETING THE UNECCESSARY ROWS  
fifa <- fifa[, -c(2,5,7,11,14,16,17,18,19,21,24,25,26,29,30,31,32,33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,51,52,53,54)]  
  
#CHECK FOR THE CLASS  
sapply(colnames(fifa), function(x) class(fifa[[x]]))

## Sl.No Name Age Nationality   
## "integer" "factor" "integer" "factor"   
## Overall Potential Club Value   
## "integer" "integer" "factor" "integer"   
## Wage Preferred.Foot Body.Type Position   
## "integer" "factor" "factor" "factor"   
## Jersey.Number Height Weight Crossing   
## "integer" "numeric" "numeric" "integer"   
## Finishing HeadingAccuracy ShortPassing Volleys   
## "integer" "integer" "integer" "integer"   
## Dribbling Curve FKAccuracy LongPassing   
## "integer" "integer" "integer" "integer"   
## BallControl Acceleration SprintSpeed Agility   
## "integer" "integer" "integer" "integer"   
## Reactions Balance ShotPower Jumping   
## "integer" "integer" "integer" "integer"   
## Stamina Strength LongShots Aggression   
## "integer" "integer" "integer" "integer"   
## Interceptions Positioning Vision Penalties   
## "integer" "integer" "integer" "integer"   
## Composure Marking StandingTackle SlidingTackle   
## "integer" "integer" "integer" "integer"   
## GKDiving GKHandling GKKicking GKPositioning   
## "integer" "integer" "integer" "integer"   
## GKReflexes Release.Clause   
## "integer" "integer"

#REPLACING CATEGORICAL MISSING VALUE  
for (j in which(fifa$Name == '')) {  
 fifa$Name[j] <-fifa$Name[j-1]  
}  
for (j in which(fifa$Nationality == '')) {  
 fifa$Nationality[j] <-fifa$Nationality[j-1]  
}  
for (j in which(fifa$Club == '')) {  
 fifa$Club[j] <-fifa$Club[j-1]  
}  
for (j in which(fifa$Preferred.Foot == '')) {  
 fifa$Preferred.Foot[j] <-fifa$Preferred.Foot[j-1]  
}  
for (j in which(fifa$Body.Type == '')) {  
 fifa$Body.Type[j] <-fifa$Body.Type[j-1]  
}  
for (j in which(fifa$Position == '')) {  
 fifa$Position[j] <-fifa$Position[j-1]  
}  
  
#REPLACING NUMERICAL MISSING VALUES  
for(i in c(8,9, 14:50))  
{  
 m <- mean(fifa[,i], na.rm = TRUE)  
 fifa[is.na(fifa[,i]), i] <-m  
 fifa[[i]] = lapply(fifa[[i]], function(x){replace(x,x==0,m)})  
 fifa[[i]] = unlist(lapply(fifa[[i]], as.numeric))  
}  
for(i in c(3,5,6,13))  
{  
 if(i!=12){  
 m=as.integer(mean(fifa[,i], na.rm = TRUE))  
 fifa[is.na(fifa[,i]), i] <-m  
 fifa[[i]] = lapply(fifa[[i]], function(x){replace(x,x==0,m)})  
 fifa[[i]] = unlist(lapply(fifa[[i]], as.numeric))  
 }  
}  
#CHECK IF THERE IS ANY MISSING VALUES  
print(any(is.na(fifa)))

## [1] FALSE

#VIEW THE CLEANED DATA  
View(fifa)  
  
#WRITE THE CLEANED DATA INTO A NEW FILE  
write.csv(fifa, "1\_cleaned\_data.csv", row.names = FALSE)