

Postwork Sesión 5 BEDU Santander Programación en R

José David Vázquez Rojas

A partir del conjunto de datos de soccer de la liga española de las temporadas 2017/2018, 2018/2019 y 2019/2020, crea el data frame SmallData, que contenga las columnas date, home.team, home.score, away.team y away.score; esto lo puede hacer con ayuda de la función select del paquete dplyr. Luego establece un directorio de trabajo y con ayuda de la función write.csv guarda el data frame como un archivo csv con nombre soccer.csv. Puedes colocar como argumento row.names = FALSE en write.csv.

```
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

df1<-read.csv("https://www.football-data.co.uk/mmz4281/1718/SP1.csv")
df2<-read.csv("https://www.football-data.co.uk/mmz4281/1819/SP1.csv")
df3<-read.csv("https://www.football-data.co.uk/mmz4281/1920/SP1.csv")
df1 <- rename(df1, date = Date, home.team = HomeTeam, home.score=FTHG, away.team= AwayTeam, away.score=AWAYFTHG)
df2 <- rename(df2, date = Date, home.team = HomeTeam, home.score=FTHG, away.team= AwayTeam, away.score=AWAYFTHG)
df3 <- rename(df3, date = Date, home.team = HomeTeam, home.score=FTHG, away.team= AwayTeam, away.score=AWAYFTHG)
lista<-list(df1,df2,df3)

df1<- df1 %>% select(date,home.team,home.score,away.team,away.score)
df2 <- df2 %>% select(date,home.team,home.score,away.team,away.score)
df3 <- df3 %>% select(date,home.team,home.score,away.team,away.score)
SmallData <- rbind(df1, df2, df3)
SmallData <- mutate(SmallData, date = as.Date(date, "%d/%m/%Y"))
SmallData$date <- gsub('0017', '2017', SmallData$date)
SmallData$date <- gsub('0018', '2018', SmallData$date)
setwd("~/Desktop/Santander/BEDU/Módulos/Programación con R/postworks/sesion 5")
write.csv(SmallData, "soccer.csv", row.names = FALSE)
```

Con la función create.fbRanks.dataframes del paquete fbRanks importe el archivo soccer.csv a R y al mismo tiempo asignelo a una variable llamada listasoccer. Se creará una lista con los elementos scores y teams que son data frames listos para la función rank.teams. Asigna estos data frames a variables llamadas anotaciones y equipos.

```
library(fbRanks)
listasoccer <- create.fbRanks.dataframes(scores.file = "soccer.csv")

## Alert: teams info file was not passed in.
## Will construct one from the scores data frame but teams in the scores file must use a unique name.
## Alert: teams resolver was not passed in.
```

```
## Will construct one from the team info data frame.
```

```
anotaciones <- listasoccer$scores
equipos <- listasoccer$teams
```

Con ayuda de la función `unique` crea un vector de fechas (fecha) que no se repitan y que correspondan a las fechas en las que se jugaron partidos. Crea una variable llamada `n` que contenga el número de fechas diferentes. Posteriormente, con la función `rank.teams` y usando como argumentos los data frames `anotaciones` y `equipos`, crea un ranking de equipos usando únicamente datos desde la fecha inicial y hasta la penúltima fecha en la que se jugaron partidos, estas fechas las deberá especificar en `max.date` y `min.date`. Guarda los resultados con el nombre `ranking`.

```
fechas <- unique(anotaciones$date)
fechas <- sort(fechas)
n <- length(fechas)
ranking <- rank.teams(anotaciones, teams = equipos, max.date = fechas[n-1], min.date = fechas[1])
```

```
##
## Team Rankings based on matches 2017-08-18 to 2020-07-16
##   team      total attack defense n.games.Var1 n.games.Freq
## 1  Barcelona    1.51 2.23    1.28    Barcelona    113
## 2  Ath Madrid    1.24 1.33    1.78    Ath Madrid    113
## 3  Real Madrid    1.15 1.86    1.19    Real Madrid    113
## 4  Valencia      0.56 1.34    1.10    Valencia      113
## 5  Getafe         0.55 1.10    1.33    Getafe         113
## 6  Sevilla        0.43 1.37    0.98    Sevilla        113
## 7  Granada        0.37 1.26    1.03    Granada         37
## 8  Villarreal     0.33 1.38    0.91    Villarreal     113
## 9  Sociedad       0.32 1.39    0.90    Sociedad       113
## 10 Ath Bilbao     0.15 1.02    1.09    Ath Bilbao     113
## 11 Osasuna        0.07 1.18    0.89    Osasuna         37
## 12 Betis          0.05 1.28    0.81    Betis          113
## 13 Celta          0.02 1.26    0.81    Celta          113
## 14 Eibar         -0.02 1.08    0.91    Eibar          113
## 15 Levante       -0.03 1.26    0.78    Levante        113
## 16 Girona        -0.18 1.07    0.83    Girona         76
## 17 Espanol       -0.21 0.93    0.93    Espanol        113
## 18 Alaves        -0.23 0.95    0.90    Alaves         113
## 19 Leganes       -0.31 0.82    0.98    Leganes        113
## 20 Valladolid    -0.33 0.79    1.00    Valladolid     75
## 21 Huesca        -0.35 1.09    0.72    Huesca         38
## 22 Mallorca     -0.41 1.02    0.74    Mallorca       37
## 23 Vallecana     -0.51 1.04    0.67    Vallecana      38
## 24 La Coruna     -0.82 0.94    0.60    La Coruna      38
## 25 Malaga        -1.17 0.58    0.76    Malaga         38
## 26 Las Palmas    -1.43 0.59    0.63    Las Palmas     38
```

Finalmente estima las probabilidades de los eventos, el equipo de casa gana, el equipo visitante gana o el resultado es un empate para los partidos que se jugaron en la última fecha del vector de fechas `fecha`. Esto lo puedes hacer con ayuda de la función `predict` y usando como argumentos `ranking` y `fecha[n]` que deberá especificar en `date`.

```
predict(ranking, date = fechas[n])
```

```
## Predicted Match Results for 1900-05-01 to 2100-06-01
## Model based on data from 2017-08-18 to 2020-07-16
## -----
```

2020-07-19 Alaves vs Barcelona, HW 9%, AW 75%, T 16%, pred score 0.7-2.5 actual: AW (0-5)
2020-07-19 Valladolid vs Betis, HW 28%, AW 43%, T 28%, pred score 1-1.3 actual: HW (2-0)
2020-07-19 Villarreal vs Eibar, HW 45%, AW 30%, T 25%, pred score 1.5-1.2 actual: HW (4-0)
2020-07-19 Ath Madrid vs Sociedad, HW 54%, AW 20%, T 26%, pred score 1.5-0.8 actual: T (1-1)
2020-07-19 Espanol vs Celta, HW 32%, AW 41%, T 27%, pred score 1.2-1.4 actual: T (0-0)
2020-07-19 Granada vs Ath Bilbao, HW 39%, AW 31%, T 29%, pred score 1.2-1 actual: HW (4-0)
2020-07-19 Leganes vs Real Madrid, HW 13%, AW 66%, T 21%, pred score 0.7-1.9 actual: T (2-2)
2020-07-19 Levante vs Getafe, HW 25%, AW 48%, T 27%, pred score 0.9-1.4 actual: HW (1-0)
2020-07-19 Osasuna vs Mallorca, HW 48%, AW 27%, T 25%, pred score 1.6-1.1 actual: T (2-2)
2020-07-19 Sevilla vs Valencia, HW 34%, AW 40%, T 26%, pred score 1.2-1.4 actual: HW (1-0)