ANA 515 Assignment 2

Sree Veereshwar Kumar Vallem

6/17/2022

Description

This dataset consists of NBA data. This dataset has 72044 observations for 27 variables. It was collected from https://projects.fivethirtyeight.com/2022-nba-predictions/ (https://projects.fivethirtyeight.com/2022-nba-predictions/). The dataset contains links to the data behind The Complete History Of The NBA and our NBA Predictions. nba_elo.csv contains game-by-game Elo ratings and forecasts back to 1946. The dataset is a Microsoft Excel Comma Separated Values File (.csv)

Using this Data, will obtain the summary stats of Overal Golden State Warriors Home and Away game totals

Loading Packages

```
library(tidyverse)
library(dplyr)
library(knitr)
library(bslib)
library(readr)
library(stringr)
library(DT)
```

Reading the data with read.csv function from the package readr

```
nba_r <- read.csv("C:\\Users\\Veere\\Downloads\\nba_elo.csv")
```

Cleaning/Pre-Processing Data

```
#nba_r_df <- nba_r %>%

is.na(nba_r) #Finding the missing values with NA in dataset and assign True False
colSums(is.na(nba_r)) #Sums of NA values based on SUm of True NA
which(colSums(is.na(nba_r))>67093) #Filtering the Columns which has more than 67093 NA's
names(which(colSums(is.na(nba_r))>67093)) # Names of the columns where NA's more than 67093 rows

  #rename(away = team1) #Renaming the column team1 to away
#rename(home = team2) #Renaming the column team2 to home
nba_2022 <-filter(nba_r, season =="2022") #Filtering NBA season 2022
head (nba_2022, n=10) #look up top 10 rows
nba_2022_homeaway <- rename (nba_2022, c(away = team1, home=team2, awayscore = score1, homescore = score2))

head (nba_2022_homeaway, n=10) #look up top 10 rows

nbarating <- filter(nba_2022_homeaway, total_rating > 95)
gswnbarating <- filter(nbarating, away=="GSW" | home=="GSW")

gswnbarating %>%
  select(c(date, season, away, home, awayscore, homescore, quality,importance, total_rating))

keeps <- c("date", "season", "away", "home", "awayscore", "homescore", "quality","importance", "total_rating")

gswnbarating = nbarating[keeps]
## Hidden the output for this code chunk as it produces a large amount of data
```

```
gswnbarating['GSWTotalScore'] =  gswnbarating['awayscore'] + gswnbarating ['homescore'] #This is overall total score of Gold
en State Warrior NBA games in 2022 season with greater than 95 NBA Rating
```

Characteristics of Data

```
observations <- nrow(gswnbarating)
variables <- ncol(gswnbarating)
```

This data frame has 45 rows and 10 columns. The names of the columns and a brief description of each are in the table below:

Table

```
kable(str(gswnbarating))
```

```
## 'data.frame':    45 obs. of  10 variables:
##  $ date         : chr  "2022-04-16" "2022-04-17" "2022-04-19" "2022-04-23" ...
##  $ season       : int  2022 2022 2022 2022 2022 2022 2022 2022 2022 2022 ...
##  $ away         : chr  "PHI" "MIA" "MIA" "UTA" ...
##  $ home         : chr  "TOR" "ATL" "ATL" "DAL" ...
##  $ awayscore    : int  131 115 115 100 102 96 89 116 121 109 ...
##  $ homescore    : int  111 91 105 99 77 98 101 117 114 86 ...
##  $ quality      : int  91 92 91 92 93 93 99 97 98 99 ...
##  $ importance   : int  100 100 100 100 100 100 100 100 99 100 ...
##  $ total_rating : int  96 96 96 96 97 97 100 99 99 100 ...
##  $ GSWTotalScore: int  242 206 220 199 179 194 190 233 235 195 ...
```

|| || || ||

Calculating Golden State Warriors Home and Away Totals and their Summary

```
GSWHomeMin <- min(gswnbarating[,9])
GSWHomeMax <- max(gswnbarating[,9])
GSWHomeMean <- mean(gswnbarating[,9])
GSWAwayMin <- min(gswnbarating[,8])
GSWAwayMax <- max(gswnbarating[,8])
GSWAwayMean <- mean(gswnbarating[,8])

summary(gswnbarating) #Using summary function
```

```
##      date               season         away               home
##  Length:45          Min.   :2022   Length:45          Length:45
##  Class :character   1st Qu.:2022   Class :character   Class :character
##  Mode  :character   Median :2022   Mode  :character   Mode  :character
##                     Mean   :2022
##                     3rd Qu.:2022
##                     Max.   :2022
##
##    awayscore       homescore         quality         importance
##  Min.   : 80.0   Min.   : 77.00   Min.   : 91.00   Min.   : 96.00
##  1st Qu.:100.2   1st Qu.: 93.75   1st Qu.: 94.00   1st Qu.:100.00
##  Median :108.0   Median :101.00   Median : 98.00   Median :100.00
##  Mean   :108.0   Mean   :101.69   Mean   : 96.67   Mean   : 99.89
##  3rd Qu.:115.8   3rd Qu.:110.00   3rd Qu.: 99.00   3rd Qu.:100.00
##  Max.   :142.0   Max.   :127.00   Max.   :100.00   Max.   :100.00
##  NA's   :3       NA's   :3
##   total_rating    GSWTotalScore
##  Min.   : 96.00   Min.   :173.0
##  1st Qu.: 97.00   1st Qu.:196.8
##  Median : 99.00   Median :206.5
##  Mean   : 98.58   Mean   :209.7
##  3rd Qu.:100.00   3rd Qu.:224.0
##  Max.   :100.00   Max.   :254.0
##                   NA's   :3
```

##Missing Values

```
    which(is.na(gswnbarating$awayscore)) #Missing Values of Away Score
```

```
## [1] 43 44 45
```

```
    which(is.na(gswnbarating$homescore)) #Missing Values of Home Scores
```

```
## [1] 43 44 45
```

Saving summary stats

```
NBASummary <- data.frame(GSWHomeMax, GSWHomeMin, GSWHomeMean, GSWAwayMax, GSWAwayMin, GSWAwayMean)
print(NBASummary)
```

```
##   GSWHomeMax GSWHomeMin GSWHomeMean GSWAwayMax GSWAwayMin GSWAwayMean
## 1        100         96    98.57778        100         96    99.88889
```