The Complete History Of The NBA Predictions

Ted Kim

2022-09-02

Overview

The complete history of the NBA https://projects.fivethirtyeight.com/complete-history-of-the-nba/#warriors article is about Elo ratings for every NBA franchise - over 60,000 ratings in total. The article describes a simple formula for Elo rating calculation and the option to explore each and every NBA team's history. But they get more credit for upset victories and for winning by larger margins. Elo ratings are zero-sum, however.

The article briefly talks about how dominant the 1990s Bulls were, the unbelievable 33 games winning streak by the 1971s Lakers, a strong but unacknowledged 60s Celtics, and the most brilliant season of the New York Knicks.

```
  library(data.table)
  library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:data.table':
##
##     between, first, last
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
  nba_carmelo <- read.csv('https://raw.githubusercontent.com/blacksmilez/DATA607/0b1e776fb7bc21f3ad1e9b1
    mutate(
      team1 = as.factor(team1),
      team2 = as.factor(team2),
      playoff = ifelse(playoff == "f", 'FINAL', playoff),
      playoff = ifelse(playoff == "c", 'CONFERENCE FINAL', playoff),
      playoff = ifelse(playoff == "p", 'PLAY_IN ROUND', playoff),
      playoff = ifelse(playoff == "q", 'FIRST ROUND', playoff),
      playoff = ifelse(playoff == "s", 'PLAYOFFS', playoff),
      neutral = ifelse(neutral == 1, TRUE, FALSE)
    )
  nba_carmelo_final <- subset(nba_carmelo, season == 2018 & playoff == 'FINAL',
                              select=c('date','playoff','season','team1','team2',
                                       'score1','score2','elo1_pre','elo2_pre'))

  nba_carmelo_final
```

```
##             date playoff season team1 team2 score1 score2 elo1_pre elo2_pre
## 67091 2018-05-31   FINAL   2018   GSW   CLE    124    114 1709.991 1611.343
## 67092 2018-06-03   FINAL   2018   GSW   CLE    122    103 1714.320 1607.015
## 67093 2018-06-06   FINAL   2018   CLE   GSW    102    110 1600.706 1720.629
## 67094 2018-06-08   FINAL   2018   CLE   GSW     85    108 1592.281 1729.054
```

Data frame for NBA_ELO

The data frame for NBA_ELO contains 20 variables. I have yet to think about which of these variables to take and discard. (*to be revised later*)

Column Name	Description
date	Date
season	Season year, 1947-2022
neutral	TRUE if the game was played on neutral territory, FALSE if not
playoff	TRUE if the game was a playoff game, FALSE if not
team1	The name of one participating team
team2	The name of the other participating team
elo1_pre	Team 1's Elo rating before the game
elo2_pre	Team 2's Elo rating before the game
elo_prob1	Team 1's probability of winning based on Elo rating
elo_prob2	Team 2's probability of winning based on Elo rating
elo1_post	Team 1's Elo rating after the game
elo2_post	Team 2's Elo rating after the game
carmelo1_pre	Team 1's CARMELO rating before the game
carmelo2_pre	Team 2's CARMELO rating before the game
carmelo1_post	Team 1's CARMELO rating after the game
carmelo2_post	Team 2's CARMELO rating after the game
carmelo_prob1	Team 1's probability winning based on CARMELO rating
carmelo_prob2	Team 2's probability of winning based on CARMELO rating
score1	Points scored by Team 1
score2	Points scored by Team 2

Conclusion

to be revised later

- GitHub - https://github.com/blacksmilez/DATA607/tree/main/Assignment01
- RPubs - https://rpubs.com/blacksmilez/938241