# ESPN_VAR_Complete_Table

Oliver Hagger

2023-08-27

### R VAR Table Data

```
library(rvest)
library(dplyr)
library(stringr)
library(tidyr)
```

### Set Working Directory

```
setwd("C:~~/VAR_scraping")
```

### Define the URL

```
VAR_PAGE_2020_2021 <- "https://www.espn.com/soccer/english-premier-league/story/3929823/how-var-decision
```

## Read the webpage

```
page <- read_html(VAR_PAGE_2020_2021)
```

### Get team list data

```
team_list <- page %>%
  html_nodes("div.article-body h2") %>%
  html_text() %>%
  gsub("\\s+$", "", .) %>%  # Remove extra spaces at the end
  gsub("[^a-zA-Z]+$", "", .) %>%
  gsub("^Editor's Picks|How to fix VAR.*|The ultimate guide.*|ESPN's .*|Marcotti: .*", "", .) %>%
  unique()

team_list <- team_list[team_list != ""]
```

## Get net score list

```r
net_score_list <- page %>%
  html_nodes("div.article-body h2") %>%
  html_text() %>%
  gsub("^.* ", "", .) %>%
  head(20)
```

## Extract general statistics for each team

```r
team_stats_list <- page %>%
  html_nodes("div.article-body p") %>%
  html_text() %>%
  grep("Overturns: ", ., value = TRUE)
```

```r
# Create a data frame
data_df <- data.frame(
  team_name = team_list,
  net_score = net_score_list,
  stats_combined = team_stats_list
)
```

## Define stats column mapping

```r
stats_col_mapping <- list(
  c('overturns_total', 'Overturns'),
  c('overturns_rejected', 'Rejected overturns'),
  c('leading_to_goals_for', 'Leading to goals for'),
  c('leading_to_goals_against', 'Leading to goals against'),
  c('disallowed_goals_for', 'Disallowed goals for'),
  c('disallowed_goals_against', 'Disallowed goals against'),
  c('net_goal_score', 'Net goal score'),
  c('subj_decisions_for', 'Subjective decisions for'),
  c('subj_decisions_against', 'Subjective decisions against'),
  c('net_subjective_score', 'Net subjective score'),
  c('penalties_for', 'Penalties for / against'),
  c('penalties_against', 'Penalties for / against')
)
```

## Create columns

```r
stats_col_list <- sapply(stats_col_mapping, `[`, 1)

for (col in stats_col_list) {
  data_df[[col]] <- 0
}
```

## Update columns based on stats combined information

```
for (i in 1:nrow(data_df)) {
  stats_info <- data_df[i, 'stats_combined']
  stats_lines <- strsplit(stats_info, "(?<=\\d)(?=[A-Z])", perl = TRUE)[[1]]

  for (line in stats_lines) {
    key <- strsplit(line, ': ')[[1]][1]
    value <- strsplit(line, ': ')[[1]][2]

    for (mapping in stats_col_mapping) {
      if (mapping[[2]] == key) {
        data_df[i, mapping[[1]]] <- value
      }
    }
  }
}
```

## Amend penalties_for and penalties_against columns

```
data_df$penalties_for <- str_extract(data_df$penalties_for, "\\d+")
data_df$penalties_against <- str_extract(data_df$penalties_against, "\\d+")
```

## Add year column and drop stats_combined column

```
data_df$year <- '2019/2020'

data_df <- data_df %>%
  select(-stats_combined)
```

```
net_score_columns <- data_df %>%
  select(starts_with("net_")) %>%
  names()

data_df <- data_df %>%
  mutate(across(all_of(net_score_columns), ~ gsub("\\+", "", .)))
```

## Print the resulting dataframe

```
print(data_df)
```

```
##                  team_name net_score overturns_total overturns_rejected
## 1  Brighton & Hove Albion         8              12                  0
## 2       Manchester United         7              13                  0
## 3          Crystal Palace         4              12                  0
## 4                 Burnley         3              11                  0
## 5               Newcastle         3               3                  0
## 6             Southampton         3              13                  0
## 7               Liverpool         2               8                  0
## 8          Leicester City         1              15                  0
## 9       Tottenham Hotspur         1              15                  0
## 10        Manchester City         0              16                  0
## 11                Arsenal        -1               9                  0
## 12                Everton        -1               7                  0
## 13        AFC Bournemouth        -2              12                  0
## 14                Chelsea        -2              12                  0
## 15                Watford        -2               6                  0
## 16            Aston Villa        -3               7                  0
## 17               West Ham        -4              16                  0
## 18       Sheffield United        -5              11                  0
## 19                 Wolves        -5               9                  0
## 20           Norwich City        -7              11                  0
##    leading_to_goals_for leading_to_goals_against disallowed_goals_for
## 1                     2                        0                    2
## 2                     1                        2                    0
## 3                     3                        0                    4
## 4                     2                        1                    3
## 5                     1                        0                    0
## 6                     0                        1                    0
## 7                     1                        0                    3
## 8                     1                        1                    3
## 9                     1                        1                    4
## 10                    3                        2                    4
## 11                    4                        1                    2
## 12                    2                        1                    2
## 13                    2                        1                    5
## 14                    2                        2                    4
## 15                    1                        2                    1
## 16                    0                        1                    3
## 17                    1                        5                    5
## 18                    0                        1                    5
## 19                    1                        1                    4
## 20                    0                        2                    2
##    disallowed_goals_against net_goal_score subj_decisions_for
## 1                         7              7                  2
## 2                         7              6                  6
## 3                         1              2                  6
## 4                         4              2                  4
## 5                         0              1                  2
## 6                         7              6                  1
## 7                         4              2                  1
## 8                         4              1                  3
## 9                         6              2                  3
## 10                        2             -1                  4
## 11                        0              1                  1
## 12                        1              0                  2
## 13                        1             -3                  2
## 14                        2             -2                  4
## 15                        1             -1                  1
## 16                        1             -3                 2
## 17                        4             -5                 2
## 18                        2             -4                 1
## 19                        1             -3                 2
## 20                        2             -4                 1
##    subj_decisions_against net_subjective_score penalties_for penalties_against
## 1                       0                    2             0                 0
## 2                       2                    4             0                 0
## 3                       2                    4             0                 0
## 4                       2                    2             0                 0
## 5                       0                    2             0                 0
## 6                       4                   -3             0                 0
## 7                       1                    0             0                 0
## 8                       3                    0             0                 0
## 9                       3                    0             0                 0
## 10                      4                    0             0                 0
## 11                      4                   -3             0                 0
## 12                      2                    0             0                 0
## 13                      4                   -2             0                 0
## 14                      4                    0             0                 0
## 15                      3                   -2             0                 0
## 16                      3                   -1             0                 0
## 17                      1                    1             0                 0
## 18                      2                   -1             0                 0
## 19                      1                    1             0                 0
## 20                      5                   -4             0                 0
##         year
## 1  2019/2020
## 2  2019/2020
## 3  2019/2020
## 4  2019/2020
## 5  2019/2020
## 6  2019/2020
## 7  2019/2020
## 8  2019/2020
## 9  2019/2020
## 10 2019/2020
## 11 2019/2020
## 12 2019/2020
## 13 2019/2020
## 14 2019/2020
## 15 2019/2020
## 16 2019/2020
## 17 2019/2020
## 18 2019/2020
## 19 2019/2020
## 20 2019/2020
```

## Export CSV file

```r
year <- '2019_2020'
csv_file_name <- paste0("Team", year, ".csv")
```

```r
write.csv(data_df, csv_file_name, row.names = T)
```