UFC Event Analysis

Tyler Ruperti

2023-08-09

UFC Event Web Scraper

This is a simple web scraper that can be used to convert stats from a ufc event into a tibble that can then be used for data analysis. The table is scraped from the following link: http://www.ufcstats.com/event-details/6f81b6de2557739a

A page like this exists for every UFC Event.

The analysis done for this one event is a simple histogram of the significant strikes landed by each fighter.

Import Libraries

```
library(rvest)
library(tidyverse)
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr
              1.1.2
                        v readr
                                    2.1.4
## v forcats 1.0.0
                                    1.5.0
                        v stringr
## v ggplot2 3.4.2
                        v tibble
                                    3.2.1
## v lubridate 1.9.2
                        v tidyr
                                    1.3.0
## v purrr
              1.0.1
## -- Conflicts -----
                               ----- tidyverse_conflicts() --
## x dplyr::filter()
                            masks stats::filter()
## x readr::guess_encoding() masks rvest::guess_encoding()
## x dplyr::lag()
                            masks stats::lag()
## i Use the conflicted package (<a href="http://conflicted.r-lib.org/">http://conflicted.r-lib.org/</a>) to force all conflicts to become error
```

Scraping the table

```
url = "http://www.ufcstats.com/event-details/6f81b6de2557739a"
html = read_html(url)

td = html %>%
   html_elements('td') %>%
   html_text2()
```

Cleaning table data

Function for cleaning some of the columns

```
columnCleaner <- function(td, index) {
  varvec <- ""
  for (i in seq(index, 120, by=10)) {
    varvec <- c(varvec,unlist(strsplit(td[i], "\n")))
  }

  varvec <- varvec[varvec != ""]

  return(varvec)
}

names <- columnCleaner(td, 2)</pre>
```

Format table into a data frame

Add values to W/L Column

We know from the source table that every odd index has the winner

```
for (i in seq(1:length(nashville23$W_L))){
   if (i%%2 == 1) {
      nashville23$W_L[i] <- "W"
   }
   else {
      nashville23$W_L[i] <- "L"
   }
}</pre>
```

Add weight class column

```
weight <- columnCleaner(td, 7)
duplicatedWeights <- ""
for (i in seq(1:length(weight))){
   duplicatedWeights <- c(duplicatedWeights, c(weight[i], weight[i]))
}
duplicatedWeights <- duplicatedWeights[duplicatedWeights != ""]
nashville23$Weight_Class <- duplicatedWeights
nashville23</pre>
```

```
## # A tibble: 24 x 7
```

##		W_L	Fighter	KD	STR	TD	SUB	Weight_Class
##		<chr></chr>	<chr></chr>	<chr></chr>	<int></int>	<chr></chr>	<chr></chr>	<chr></chr>
##	1	W	Cory Sandhagen	0	34	7	1	Catch Weight
##	2	L	Rob Font	0	9	1	0	Catch Weight
##	3	W	Tatiana Suarez	0	23	3	1	Women's Strawweight
##	4	L	Jessica Andrade	0	11	0	0	Women's Strawweight
##	5	W	Dustin Jacoby	1	10	0	0	Light Heavyweight
##	6	L	Kennedy Nzechukwu	0	4	0	0	Light Heavyweight
##	7	W	Diego Lopes	0	0	0	3	Featherweight
##	8	L	Gavin Tucker	0	0	1	0	Featherweight
##	9	W	Tanner Boser	0	120	0	0	Light Heavyweight
##	10	L	Aleksa Camur	0	68	0	0	Light Heavyweight
##	## # i 14 more rows							

Finally, the Histogram

Histogram of Significant Strikes for Each Fighter

