

Web_Scraping_NBB

Gilmar Pires

2025-01-04

Projeto de Raspagem de dados do site NBB.BR

Nesse projeto, fiz uma raspagem de dados do site da NBB (Novo Basquete Brasil) para obter informações sobre as equipes e suas estatísticas. O objetivo é extrair uma tabela de classificação da NBB. Usei a biblioteca rvest para extrair os dados da tabela diretamente do site e, em seguida, manipulei os dados para reorganizá-los conforme as instruções fornecidas. E por último, montei um relatório em R Markdown para exibir a tabela de classificação da NBB, e com isso tenho um relatório atualizado sempre que o código for executado. Esse é apenas um exemplo de como o R pode ser usado para extrair e manipular dados da web e gerar relatórios personalizados.

```
# Carregar as bibliotecas
library(rvest)
library(dplyr)
library(stringr)
library(kableExtra)

# URL do site da NBB
url <- read_html("https://lnb.com.br/nbb/")

# Extrair a tabela diretamente usando rvest
tables <- url %>% html_nodes("table")

# Verifique se há tabelas disponíveis e exiba a quantidade encontrada
cat("Número de tabelas encontradas:", length(tables), "\n")
```

```
## Número de tabelas encontradas: 2
```

```

# Se houver tabelas, extraia a primeira tabela como exemplo
if (length(tables) > 0) {
  NBB_table <- tables[[1]] %>% html_table(fill = TRUE)

  # Atribuir nomes temporários para colunas ausentes ou vazias
  colnames(NBB_table)[is.na(colnames(NBB_table)) | colnames(NBB_table) == ""] <- paste0("V", seq_
_along(colnames(NBB_table))[is.na(colnames(NBB_table)) | colnames(NBB_table) == ""])

  # Remover linhas indesejadas que contenham "\ue037 MAIS ESTATÍSTICAS"
  NBB_table <- NBB_table %>%
    filter(!grepl("\ue037 MAIS ESTATÍSTICAS", EQUIPES)) %>%
    filter(!is.na(EQUIPES))

  # Extrair apenas as três letras da coluna "EQUIPES"
  equipes_simplificado <- NBB_table %>%
    mutate(EQUIPES = str_extract(EQUIPES, "[A-Z]{3}")) %>%
    filter(!is.na(EQUIPES))

  # Salvar a coluna SAL da tabela original antes de modificar as colunas
  SAL_orig <- NBB_table$SAL
  PON_orig <- NBB_table$PON

  # Realocar os dados conforme as instruções fornecidas
  NBB_table <- NBB_table %>%
    mutate(
      PON = JOG,
      JOG = DER,
      VIT = CASA,
      DER = FORA,
      CASA = PRO,
      FORA = CON,
      PRO = SAL_orig,
      CON = ULT,
      SAL = PON_orig
    )

  # Remover as colunas "ULT", "V13" e "V14"
  NBB_table <- NBB_table %>% select(-ULT, -V13, -V14)

  # Substituir a coluna "EQUIPES" na tabela original pelas informações da nova tabela
  NBB_table$EQUIPES <- equipes_simplificado$EQUIPES

  # Adicionar a coluna de classificação
  NBB_table <- NBB_table %>%
    mutate(CLASSIFICAÇÃO = row_number())

  # Selecionar as colunas na ordem correta
  NBB_table <- NBB_table %>%
    select(CLASSIFICAÇÃO, EQUIPES, `AP(%)`, JOG, VIT, DER, CASA, FORA, PRO, CON, SAL, PON)
} else {
  print("Nenhuma tabela encontrada na página.")
}

```

```
rm(equipes_simplificado, PON_orig, SAL_orig, url, tables)

# Exibir a tabela usando kableExtra com rolagem
NBB_table %>%
  kable(format = "html", table.attr = 'class="table"') %>%
  kable_styling(bootstrap_options = c("striped", "hover", "condensed", "responsive"), full_width
= F) %>%
  row_spec(0, bold = TRUE, color = "darkgreen", background = "darkorange") %>%
  column_spec(1, bold = TRUE, color = "darkgreen", background = "darkorange") %>%
  column_spec(2, bold = TRUE, color = "darkorange") %>%
  row_spec(1:nrow(NBB_table), background = "lightyellow")%>%
  scroll_box(height = "400px")
```

CLASSIFICAÇÃO	EQUIPES	AP(%)	JOG	VIT	DER	CASA	FORA	PRO	CON	SAL	PON
1	MIN	89.5	19	17	2	10-0	7-2	1583	1330	253	36
2	FLA	78.9	19	15	4	8-3	7-1	1594	1368	226	34
3	BRA	68.4	19	13	6	7-2	6-4	1593	1537	56	32
4	FRA	57.9	19	11	8	5-2	6-6	1540	1487	53	30
5	UCO	57.9	19	11	8	5-4	6-4	1488	1493	-5	30
6	PIN	55.6	18	10	8	3-4	7-4	1374	1356	18	28
7	VAS	55.6	18	10	8	4-6	6-2	1337	1301	36	28
8	SAO	55.6	18	10	8	6-3	4-5	1417	1411	6	28
9	BAU	52.6	19	10	9	7-1	3-8	1485	1516	-31	29
10	UFC	47.4	19	9	10	4-6	5-4	1428	1446	-18	28
11	SJO	47.4	19	9	10	5-5	4-5	1438	1428	10	28
12	CAP	44.4	18	8	10	5-5	3-5	1401	1362	39	26

Exposição da classificação da NBB com um gráfico

```
# Carregar a biblioteca ggplot2
library(ggplot2)

# Converter a coluna PON para numérica
NBB_table$PON <- as.numeric(NBB_table$PON)

# Criar o gráfico de barras com classificação decrescente e rótulos centralizados
ggplot(NBB_table %>% arrange(desc(PON)), aes(x = reorder(EQUIPES, PON), y = PON)) +
  geom_bar(stat = "identity", fill = "darkorange") +
  geom_text(aes(label = PON), vjust = 0.5, hjust = 1.5, color = "darkgreen", size = 3.5) +
  coord_flip() +
  labs(title = "Pontuação", x = "Equipes", y = "Pontos") +
  theme_minimal() +
  theme(plot.background = element_rect(fill = "lightyellow", color = NA),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank())
```

