

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA  
MESTRADO EM ESTATÍSTICA

VITOR FARIA DE CARVALHO OLIVEIRA

**ACESSANDO INFORMAÇÃO DE ESPORTE ATRAVÉS DO  
SOFASCORE**

BELO HORIZONTE

2022

VITOR FARIA DE CARVALHO OLIVEIRA

ACESSANDO INFORMAÇÃO DE ESPORTE ATRAVÉS DO  
SOFASCORE

Dissertação apresentada ao Programa de Pós-graduação em Estatística da Universidade Federal de Minas Gerais (UFMG), como requisito parcial à obtenção do título de Mestre em Estatística.

Área de atuação: Ciência de dados/Web scraping em R

Linha de pesquisa: *Data Mining*

Orientador: Marcos Oliveira Prates

BELO HORIZONTE

2022

Oliveira, Vitor Faria de Carvalho.

O48a      Acessando informação de esporte através do SofaScore  
[recurso eletrônico] / Vitor Faria de Carvalho Oliveira – 2022.  
1 recurso online (44 f. il., color.): pdf.

Orientador: Marcos Oliveira Prates.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.

Referências: f. 38-40.

1. Estatística – Teses. 2. Ciência de dados – Teses.  
3. Banco de Dados – Esportes - Teses. 3. Mineração de dados (Sistemas de recuperação da informação) – Teses. I. Prates, Marcos Oliveira. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. III. Título.

CDU 519.2(043)



ATA DA DEFESA DE DISSERTAÇÃO DE MESTRADO DO ALUNO VITOR FARRIA DE CARVALHO OLIVEIRA, MATRICULADO, SOB O Nº 2020667325, NO PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA, DO INSTITUTO DE CIÊNCIAS EXATAS, DA UNIVERSIDADE FEDERAL DE MINAS GERAIS, REALIZADA NO DIA 22 DE MARÇO DE 2022.

Aos 22 dias do mês de Março de 2022, às 13h30, em reunião pública virtual 265 (conforme orientações para a atividade de defesa de dissertação durante a vigência da Portaria PRPG nº 1819) OU na sala do Instituto de Ciências Exatas da UFMG, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pelo Colegiado do Programa de Pós-Graduação em Estatística, para julgar a defesa de dissertação do aluno VITOR FARRIA DE CARVALHO OLIVEIRA, nº matrícula 2020.667.325, intitulada: " *Acessando Informação de Esporte através do SofaScore* ", requisito final para obtenção do Grau de mestre em Estatística. Abrindo a sessão, o Senhor Presidente da Comissão, Prof. Marcos Oliveira Prates, passou a palavra ao aluno para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do aluno. Após a defesa, os membros da banca examinadora reuniram-se reservadamente sem a presença do aluno e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação:

- Aprovada.  
 Reprovada com resubmissão do texto em \_\_\_\_ dias.  
 Reprovada com resubmissão do texto e nova defesa em \_\_\_\_ dias.  
 Reprovada.

Prof. Marcos Oliveira Prates - Orientador  
(DEST/UFMG)

Prof. Fabio Nogueira Demarqui  
(DEST/UFMG)

Prof(a). Waltes Marques Zeviani (Doutor)  
(DEST/UFPR)

O resultado final foi comunicado publicamente ao(à) aluno(a) pelo(a) Senhor(a) Presidente da Comissão. Nada mais havendo a tratar, o(a) Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 22 de março de 2022.

Observações:

1. No caso de aprovação da tese, a banca pode solicitar modificações a serem feitas na versão final do texto. Neste caso, o texto final deve ser aprovado pelo orientador da tese. O pedido de expedição do diploma do candidato fica condicionado à submissão e aprovação, pelo orientador, da versão final do texto.
2. No caso de reprovação da tese com resubmissão do texto, o candidato deve submeter o novo texto dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca que então decidirão pela aprovação ou reprovação da tese.
3. No caso de reprovação da tese com resubmissão do texto e nova defesa, o candidato deve submeter o novo texto com a antecedência à nova defesa que o orientador julgar adequada. A nova defesa, mediante todos os membros da banca, deve ser realizada dentro do prazo estipulado pela banca, que deve ser de no máximo 6 (seis) meses. O novo texto deve ser avaliado por todos os membros da banca. Baseada no novo texto e na nova defesa, a banca decidirá pela aprovação ou reprovação da tese.



**UNIVERSIDADE FEDERAL DE MINAS GERAIS**

**PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA**

**UFMG**

## FOLHA DE APROVAÇÃO

**"Acessando Informação de Esporte através do SofaScore"**

### **VITOR FARIA DE CARVALHO OLIVEIRA**

Dissertação submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTATÍSTICA, como requisito para obtenção do grau de Mestre em ESTATÍSTICA, área de concentração ESTATÍSTICA E PROBABILIDADE.

Aprovada em 22 de março de 2022, pela banca constituída pelos membros:

Prof(a). Marcos Oliveira Prates - Orientador  
DEST/UFMG

Prof(a). Fabio Nogueira Demarqui  
DEST/UFMG

Prof(a). Walmes Marques Zeviani  
DEST/UFPR

Belo Horizonte, 22 de março de 2022.

# Agradecimentos

Agradeço primeiramente a Deus, ao meu pai Hugo, a minha mãe Raquel e aos meus irmãos Hugo, Luciano, Gabriel e Pedro por serem meus alicerces durante toda minha vida, tudo que sou é graças a essa família. Também a memória da minha falecida avó Alice por ser o maior exemplo de como devo ser.

Agradeço também ao professor e orientador Marcos Oliveira Prates pelo esforço, paciência e apoio depositados em mim.

A toda minha família avós, tios, primos e cunhadas; também a todos meus amigos que sempre me incentivaram e estiveram do meu lado.

As organizações que tornaram possível concretizar esse mestrado: à UFMG, em especial ao Instituto de Ciências Exatas pelo suporte; ao SofaScore pelo excelente trabalho; as organizações de pesquisa e fomento que apoiam a educação CAPES, CAPEMIG, FAPEMIG, em especial, a CNPq que disponibilizou bolsa por mais de um ano no meu mestrado e que acreditou muito no meu potencial.

A todos os professores e funcionários do Programa de Pós-Graduação em estatística e também ao meu professor e amigo Roberto da Costa Quinino que me orientou na minha monografia.

Aos meus colegas de mestrado e trabalho que sempre me apoiaram tecnicamente.

# Resumo

Jogos esportivos têm-se tornado cada vez mais populares, sendo o futebol, por exemplo, acompanhado por cerca de quatro bilhões de pessoas em todo o mundo. Tal popularidade tem proporcionado não só importantes transformações culturais, mas também trouxe grande impacto econômico. Visto o aumento das discussões e do interesse em analisar as partidas, torna-se necessária a disponibilização de bases de dados estruturadas, de forma a universalizar o acesso aos dados futebolísticos. Junto a essa necessidade existe a dificuldade de obter dados atualizados e confiáveis para a população geral de forma que as informações possam ser facilmente analisadas e estudadas.

Pensando nos pontos anteriores, o presente trabalho visa fornecer dados futebolísticos de forma fácil e prática a todas as pessoas, atingindo diferentes categorias de usuários. Para isso, foi selecionado o site **SofaScore**, uma fonte de dados autêntica e especializada para extração das informações. Este site é um dos maiores sites de consultas de partidas esportivas, que possui uma audiência de 22 milhões de pessoas e 22 anos de mercado, por esses motivos foi escolhido para a extração.

Através das ferramentas existentes na ciência de dados foram selecionados dados referentes ao Campeonato Brasileiro da série A, por ser um campeonato com espectadores muito ativos, grande número de rodadas e um dos mais relevantes em âmbito mundial. Neste trabalho, utilizamos o *software* R combinado com a técnica de coleta de dados chamada **web scraping**, responsável por, de forma automática, encontrar páginas, selecionar conteúdo e extraí-los. Dessa forma, foi desenvolvido um pacote com diversas funcionalidades que auxilia o usuários a extrair informações de interesse de maneira direta e simples.

**Palavras-chave:** banco de dados; coleta; automática; R; **SofaScore**; *web scraping*.

# Abstract

Sports games have become popular, and football being the main one of these sports, followed by around four billion people worldwide. Such popularity has not only provided important cultural transformations, but has also had a large economic impact. Because of the increase in discussions and interest in analyzing soccer matches, it became necessary to provide structured database to universalize access to this data. Along with this need, there is the difficulty of obtaining updated and reliable data for the general population, so that the information can be easily analyzed and studied.

Thinking about the previous comments, the present work intends to deliver football data in an easy and practical way to all people, reaching different types of users. The **SofaScore** site was selected for these reasons and for being an authentic and expert data source for extracting the information. This site is one of the biggest references in consultations on sports matches, having an audience of 22 million people and 22 years of market, for these reasons it was chosen for the extraction.

Through the data science tools were selected data from the Brazilian championship of the A series, for being a championship with very active spectators, many rounds and one of the most relevant tournaments in the world. In this work, we use the **R software** combined with the data collection technique called **web scraping**, responsible for automatically finding pages, selecting, and extracting the desired content. In this way, a package with several features was developed that helps users extract information of interest in a direct and simple way.

**Key-words:** database; collection; automatic; R; **SofaScore**; web scraping.

# Lista de ilustrações

Figura 1 – Quantidade de pesquisas pela expressão <i>web scraping</i> . . . . .	20
Figura 2 – Mapa de calor da temporada do jogador Hulk, time Atlético MG. . . .	35

# Lista de tabelas

Tabela 1	– Funções pertencentes ao pacote <code>SofaScore</code> . . . . .	27
Tabela 2	– Tabela fornecida utilizando a função <code>Brasileirão_A</code> , tipo “Overall” para o ano de 2021. . . . .	29
Tabela 3	– Tabela dos últimos 10 jogos, fornecida utilizando a função <code>Last_Games</code> e especificando o time Atlético MG. . . . .	30
Tabela 4	– Tabela dos próximos 10 jogos, fornecida utilizando a função <code>Upcoming_Games</code> e especificando o time Atlético MG. . . . .	31
Tabela 5	– Tabela fornecida utilizando a função <code>Players</code> , especificando o time Atlético MG e a posição dos jogadores como atacantes. . . . .	32
Tabela 6	– Tabela fornecida utilizando a função <code>Statiscs_Games</code> , especificando o time Atlético MG. . . . .	43
Tabela 7	– Objeto <code>TabLink</code> fornecido pelo Pacote <code>SofaScore</code> . . . . .	44
Tabela 8	– Tabela fornecida utilizando a função <code>TabLinkPlayers</code> , especificando o time Atlético MG. . . . .	45

# Sumário

<b>1</b>	<b>Introdução</b>	<b>12</b>
<b>2</b>	<b>Revisão Temática</b>	<b>14</b>
2.1	Dados	14
2.2	<i>Data Mining</i>	14
2.3	Dados Dinâmicos	15
<b>3</b>	<b>Objetivos</b>	<b>17</b>
3.1	Objetivo geral	17
3.2	Objetivos específicos	17
<b>4</b>	<b>Metodologia</b>	<b>19</b>
4.1	Ciência de dados	19
4.1.1	Web scraping	19
4.2	Software R	21
4.2.1	RSelenium	22
<b>5</b>	<b>SOFASCORE</b>	<b>24</b>
5.1	Site	24
5.2	Pacote	25
5.2.1	Tutorial de navegação	27
5.2.1.1	Função <code>Open_Sofa</code>	27
5.2.1.2	Função <code>Close_Sofa</code>	28
5.2.1.3	Função <code>Brasileiro_A</code>	28
5.2.1.4	Função <code>Last_Games</code>	29
5.2.1.5	Função <code>Upcoming_Games</code>	31
5.2.1.6	Função <code>Players</code>	32
5.2.1.7	Função <code>Statiscs_Games</code>	33
5.2.1.8	Função <code>Heat_Map</code>	34
5.2.1.9	Tabela <code>TabLink</code>	35
5.2.1.10	Função <code>TabLinkPlayers</code>	36
<b>6</b>	<b>Conclusão</b>	<b>37</b>
6.1	Resultados	37
6.2	Trabalhos Futuros	38
	<b>Referências</b>	<b>39</b>
	<b>Anexos</b>	<b>42</b>
	<b>ANEXO A Tabela <code>Statiscs_Games</code></b>	<b>43</b>

<b>ANEXO B</b>	<b>Tabela TabLink</b>	<b>44</b>
<b>ANEXO C</b>	<b>Tabela TabLinkPlayers</b>	<b>45</b>

# 1 Introdução

Desde os primórdios, a necessidade do ser humano em evidenciar-se frente a tudo, ressalta a competitividade entre os indivíduos. Desde a caça, passando pelas batalhas em arenas, e a criação de jogos que determinam vencedores e vencidos. Com o passar do tempo os combates que ameaçavam a vida foram substituídos por partidas que “hoje” o vencido poderá, amanhã, ser vencedor, evidenciando valores, emoções, sensações e criações artísticas.

O futebol, por ser o esporte mais praticado do mundo e adorado por mais de 4 bilhões de pessoas [do estado, 2021], tem grande importância não só na cultura mundial como na economia. Exemplos disso são os acordos valiosos gerados com a compra e venda de imagens, jogadores, audiências e estatísticas. Em 2019, a Globo comprou a transmissão dos jogos do Campeonato Brasileiro por R\$ 925 milhões [UOL, 2019]. A Forbes anunciou, em maio de 2021, valores de ganhos dos jogadores no mesmo ano, como é o caso do Lionel Messi recebendo US\$ 130 milhões, o mesmo jogador teve o contrato revelado valendo US\$ 674 milhões. Outro atleta que estava em uma faixa de preço similar é o Cristiano Ronaldo com o ganho em 2021 de US\$ 120 milhões por ano, Neymar também estava na lista dos 10 atletas mais bem pagos do mundo, com a remuneração no valor de US\$ 95 milhões [Forbes, 2021].

Segundo a FIFA, a Copa do Mundo possui a maior audiência dentre todos os programas televisionados, 3,5 bilhões de pessoas viram pelo menos um jogo do Mundial de 2018. Já a *Champions League* é um dos campeonatos mais importantes para o futebol, e isto pode ser visto facilmente através dos números e recordes que essa competição continuamente supera nas redes sociais. A ESPN publicou, em maio de 2012 [ESPN, 2012], que a *Champions* bateu recorde de *tweets* por segundo em sua semifinal, e a final do campeonato foi o evento esportivo único mais assistido da história, na época, superando o Super Bowl. Outros campeonatos com acompanhamento do público de forma global são a *Premier League* e a *La League*, nos últimos anos esses aumentaram o número de telespectadores e a renda que geram. Em 2019, foram mais de 3 bilhões de pessoas que acompanharam os jogos da Premier em mais de 180 países [MKTESPORTIVO, 2019]. Toda essa audiência e atenção para estes campeonatos faz com que haja uma necessidade grande de se obter informações sobre o conteúdo das partidas, afinal esse é o esporte mais popular do mundo [do estado, 2021].

O Campeonato Brasileiro é considerado pelo *International Federation of Football History and Statistics* (IFFHS) como o terceiro campeonato nacional mais importante do mundo na última década, gera grande relevância para o futebol brasileiro dentro e fora do país. Atualmente, a transmissão internacional do campeonato já supera 100 países no

exterior, com mais de 10 milhões de telespectadores [EXAME, 2021]. A Uol declarou que o *pay-per-view* do Campeonato Brasileiro tem uma arrecadação anual de R\$ 1,7 bilhão por ano, porém o direito de transmissão dos jogos não é apenas do *pay-per-view*, mas também da televisão aberta e outros canais de divulgação.

Com todo esse destaque do futebol e em particular do futebol brasileiro, sendo o Brasileirão a liga com maior importância no país, existe a necessidade não somente de acompanhar os jogos, como também de estudá-los mais a fundo. Pensando nisso, esse trabalho visa conseguir extrair resultados esportivos da web, e disponibilizar de forma prática à comunidade estatística e a todos os usuários do R [R Core Team, 2021b].

A seguir será apresentada uma revisão temática com trabalhos na área de coletas de dados e dados futebolísticos automáticos. No trabalho será criado um pacote que fornecerá uma variada gama de dados de futebol, ou seja, através dele o usuário especificará seus interesses dentro de parâmetros que as funções necessitam, e obterá um banco de dados confiável e de simples uso para análises. Vale ressaltar que o pacote é desenvolvido em plataforma livre e código aberto, possibilitando usuários avançados a aprender sobre a codificação e estender as funcionalidades de acordo com o seu interesse.

## 2 Revisão Temática

Nesta seção serão apresentados outros trabalhos que ilustram a importância do banco de dados nas análises de pesquisas, mostrado também as referências teóricas sobre o *web scraping* e uma seção final informando outro trabalho que mais se assemelha com nossos objetivos de entregar dados dinâmicos que já possuem uma estrutura fixa de coleta, e são atualizados constantemente.

### 2.1 Dados

Ciência significa: estudo; conhecimento profundo sobre algo [Ferreira, 2009], ou seja, a ciência de dados (CD) gira em torno de processar e gerar conteúdo através dos dados, por isso estes são prioridade elementar na prática de análise de dados. Como foi descrito no livro Elementos de amostragem, “Resultados confiáveis e livres de vieses é o desejo de todos.”, com isso não há necessidade de ressaltar a importância dos dados em qualquer trabalho [Bolfarine and de Oliveira Bussab, 2005]. Logo se há problema em obtê-los; na forma de recebe-los ou na honestidade dos mesmos, a falha facilmente se alastrará por toda a análise, que começa em captar material, depois transformar em informação, e por fim gerar conhecimento [Amaral, 2016].

A Declaração de Berlin sobre o Acesso Aberto ao Conhecimento em Ciências e Humanidades [Redalyc et al., 2003], publicada em 2003, retrata a importância do acesso livre de conteúdo para o desenvolvimento da pesquisa e geração de *insights*, com a missão de disseminar conhecimento com informações amplamente e prontamente disponíveis.

### 2.2 Data Mining

O *Data Mining* que traduzindo significa mineração de dados está entre os métodos estatísticos mais recorrente nas dissertações e teses acadêmicas, na área de ciências econômicas, além disso esta linha de pesquisa é amplamente abordada, com diversas técnicas e ferramentas [Dallabona et al., 2010]. O livro apresenta a metodologia *Knowledge Discovery in Databases - KDD* que é uma etapa para descobrir conhecimento através de bases de dados [Goldschmidt and Passos, 2005]. O termo *Data Mining* é um sinônimo amplamente utilizado na literatura para o *Web Scraping*.

Ferrara et al. [2014] defende a ideia de adquirir e utilizar informações através do que as redes sociais apresentam no front-end. Sua pesquisa ganhou muita força por

mostrar como captar dados de forma dinâmica e automática do que atualmente está sendo abordado constantemente que são as mídias sociais. O artigo também discute sobre a captação de dados para o sistema corporativo que visa sistemas de negócios. [Catanese et al. \[2011\]](#) seguiu essa linha de pensamento focando o rastreamento de dados no facebook, conseguindo desenvolver e analisar gráficos de redes sociais.

O livro escrito pelo Ryan “Web Scraping with Python”[\[Mitchell, 2018\]](#) descreve bem como fazer web scraping, através de bots que são os comandos que irão interagir com uma API (Permite a integração entre `softwares`) utilizando a linguagem Python [\[Van Rossum and Drake, 2009\]](#). O livro é bem abrangente, sua ideia é entregar um manual onde o leitor consiga fazer variadas formas de raspagem moderna na internet.

Algo que gerou força para o web scraping atualmente foram os robôs de raspagem de dados que as grandes empresas de programação e tecnologia utilizam hoje em dia, como é o caso do Googlebot do google; Msnbot do Microsoft; Patent2net analisa a Espacenet; Yahoo! Sluro do Yahoo! e diversos outros que buscam diferentes modelos de dados. [\[Sun et al., 2007\]](#) apresenta alguns dos robôs citados acima.

## 2.3 Dados Dinâmicos

Para a análise de dados dinâmicos e automáticos é válido ter em mente que o processo de ciência de dados para este estudo é longo e não trivial. Por isso muitas vezes o trabalho como um todo é dividido e feito por diferentes equipes que são capacitadas em variadas áreas dentro do campo da estatística. O livro *Process Mining Data Science in Action* [\[Van der Aalst, 2016\]](#) demonstra a complexidade e diversidade que existe na CD.

Em projetos que buscam fazer *web scraping* com dados de futebol é muito importante informar qual a fonte dos dados colhidos e os links, a partir dos quais o conteúdo é extraído. O `worldfootballR` [\[Zivkovic, 2022\]](#) é um desses projetos, esse entrega uma base de dados rica de informações de jogos e times de campeonatos mundiais, é um pacote com o diferencial de entregar dados com valores de transferências de jogadores. Ele fornece os dados do site `fbref.com`, criado por equipes que são referências em beisebol e basquetebol. O `Fbref` é um site relativamente novo, lançado em 2018, que reconhece estar no início do desafio de entregar dados sobre o futebol mundial, mas que tem uma visão muito promissora no longo prazo dos objetivos da empresa [\[FBref, 2021\]](#). O pacote também apresenta os dados do site `Transfermarkt` que são dados valiosos e constantemente atualizados sobre transferência de jogadores, o `Transfermarkt` conta com cerca de 600 mil usuários, milhões de visitantes por dia, cerca de 760 mil perfis de jogadores (e 75 mil clubes), mais de 1500 competições e mais de um milhão de tabelas na base de dados, disponíveis em várias línguas [\[Transfermarkt, 2020\]](#). Além desses dois sites ele também busca informações do `Understat` que busca estatísticas detalhadas de gols esperados para as principais ligas

européias e do Fotmob que tem como missão acompanhar seu clube local de qualquer lugar, o aplicativo fornece resultados ao vivo, estatísticas e notícias sobre as competições mundiais.

## 3 Objetivos

### 3.1 Objetivo geral

Erros nos dados podem resultar em prejuízos vultosos ou em distorções de resultados com consequências técnicas imprevisíveis [Goes et al., 2010]. Com isso em mente, podemos afirmar que uma entrega bem feita de dados é essencial para que qualquer análise seja realizada em CD. Dessa forma, o presente estudo tem como objetivo focar nessa segmentação da estatística e disponibilizar, para qualquer pessoa, acesso de forma simples, direta e segura aos dados disponíveis no site **SofaScore**. O **SofaScore** é um site consolidado no mercado futebolístico mundial que dispõe de grande quantidade de informações; estimativas calculadas através de métodos próprios; alta audiência pelo público que busca por dados confiáveis de futebol (para mais detalhes veja Seção 5.1).

### 3.2 Objetivos específicos

Primeiramente queremos disponibilizar os resultados obtidos através de um pacote que poderá ser facilmente baixado e utilizado na plataforma R que será descrita na Seção 4.2. Outro objetivo é facilitar o uso das tecnologias que estão à disposição das pessoas na área da informática atualmente, através de conteúdos instrucionais como este e ferramentas práticas, tornando mais acessíveis o uso da ciência de dados e da programação.

Assim como o site separa suas tabelas, imagens e outros resultados, aqui eles também serão separados através de funções, onde o usuário descreve os argumentos que deseja extrair, como: time; ano; nome do jogador, para que seja fornecido dados preparados para análises. É do nosso interesse entregar, através dessas funções, os dados da tabela geral do campeonato analisado, informações específicas dos campeonatos e jogos, características do elenco do time, estatísticas calculadas através de modelagem que são próprias do **SofaScore** para classificação dos jogadores através de notas e mapas de calor que o site oferece.

Quando falamos de dados, confiabilidade é super importante! Por isso, queremos entregar dados de um site que é referência e também é um dos mais utilizados na área de pesquisa sobre futebol no contexto mundial e brasileiro, essa é uma das diferenças do trabalho, para as outras bases de informações que foram apresentadas. Também, no melhor do nosso conhecimento, é o primeiro pacote que extrai informações do **SofaScore**, podendo ser utilizado por trabalhos que tenham interesse em estender o que foi feito, por

exemplo, incluindo novas ligas de futebol ou esportes.

## 4 Metodologia

### 4.1 Ciência de dados

De acordo com o [Wickham and Grolemund \[2016, p. 4\]](#) no livro “R for Data Science” a ciência de dados permite transformar dados brutos em compreensão, percepção e conhecimento. Nesta transformação, o primeiro passo é importar os dados que serão trabalhados, sendo eles oriundos de planilhas fornecidas, interface de programação ou aplicativos web (API). Para um bom desempenho nessa parte, é necessário conhecer bem a ferramenta que está trabalhando e onde os dados são disponibilizados, pois uma leitura errônea dos dados pode comprometer todo o estudo. Exemplo disso é quando variáveis que têm elementos numéricos e caracteres são lidos apenas como numéricos; quando isto ocorre, os caracteres podem ser excluídos, perdendo informações.

O trabalho corrente se baseia nas premissas de importar, organizar e disponibilizar informações abundantes sobre jogos esportivos, por meio de *web scraping*. Tendo em vista estas premissas, após a importação, os dados devem ser organizados e estruturados de forma a facilitar o seu uso para gerar informações. Em seguida, pode-se trabalhar com modelos estatísticos e visualizações a fim de responder perguntas e ilustrar bem as respostas.

#### 4.1.1 Web scraping

Houve uma mudança de paradigma na virada do último milênio na tecnologia da informação e na mentalidade em adquirir dados e estudá-los, com a internet tão acessível e a facilidade na captação de dados sobre gostos, relacionamentos, localizações, preferências e outros [[Boyd and Crawford, 2012](#)]. Possuir essas informações, hoje, dá poder e gera renda às empresas, marketing e estatísticas, pois quando analisadas da forma correta, geram informações objetivas, ou seja, respostas a dúvidas de como melhorar vendas, aumentar captação de clientes, gerar renda e manipular vontades do público [[Teloken et al., 2016](#)].

Desta forma *web scraping* ou raspagem de dados é o nome dado à técnica que automatiza o processo de coleta de dados na web. “Se a computação é mágica, então web scraping é feitiçaria” [[Mitchell, 2018](#)], nessa frase autor indica que procedimento técnico de raspagem é a aplicação da “mágica” em resultados úteis. Essa técnica vem ganhando cada vez mais força em sua utilização, isso pode ser percebido pelo [Google Trends \[Google, 2022\]](#). Na Figura 1 mostramos o crescimento da pesquisa pela expressão *web scraping*, do começo de 2010 até hoje.

Figura 1 – Quantidade de pesquisas pela expressão *web scraping*.

Imagem obtida em 15/02/2022 às 11:36

Em suma, a raspagem nada mais é do que a criação de programas que executam comandos de forma automática para visitar as páginas e extrair toda a informação que se deseja. Essa captação de dados, por ser automática, é algo prático e rápido, gerando assim benefícios como tempo e diminuição do trabalho [Polidoro et al. \[2015, p. 169\]](#).

*Web scraping* pode ser feito com diferentes objetivos, algumas empresas utilizam essa ferramenta para fazer análise de mercado ou classificar o perfil do cliente, assim como os bancos, que podem fazer análise preditiva de perfis. Já no caso de sites de busca como Google, Bing e outros, a ferramenta é utilizada para automatizar o trabalho em sua organização e classificação através de bots chamados de *web crawling*, assim como os bancos que podem fazer análise preditiva de perfis [\[Cardoso and Magalhães, 2011\]](#). Facilmente, pode-se perceber o *web scraping* no cotidiano, pois quando vemos alguém dar acesso aos seus dados pessoais, como contatos, localizações e outros, em troca de serviços na internet, é exatamente uma permissão para que a organização faça raspagem dos dados do indivíduo.

Algo importante de se observar é que há poucas restrições legais quanto ao uso e modo da captação de dados, pois não há uma legislação direta para o *web scraping* [\[Krotov et al., 2020\]](#). Isso gera uma paradoxalidade, de um lado temos o avanço da ferramenta, algo imensurável, pois qualquer um pode fazer usando a criatividade que se tem; de outro, e não menos importante, o fato da utilização da ferramenta para fraudes e golpes o que hoje todos devem estar cientes e tomar as devidas precauções, que não será abordado neste presente trabalho.

Para realização da raspagem de dados temos três passos essenciais:

1. Utilizar o programa da escolha do usuário para poder acessar o servidor ou o código da página;
2. Conseguir, através de comandos de programação, encontrar o que se deseja, selecionar, extrair e salvar localmente;
3. Estruturar os dados de forma a disponibilizá-los;

Ao se fazer raspagem de dados na web, na maioria das vezes são necessários conhecimentos de linguagens de programação e de como a página é construída pelos seus códigos de HTML, CSS e JavaScript.

## 4.2 Software R

Essa linguagem de programação foi criada por Ross Ihaka e Robert Gentleman, no ano de 1993, a primeira letra de ambos os nomes foi o que determinou o nome do projeto como “R”, ambos são estatísticos, por isso o foco nesta área.

Além de ser uma linguagem de programação o **R** também é um **software** que foi construído como uma ferramenta semelhante ao **software S**, porém uma versão gratuita. Dois pontos muito fortes deste **software** que foram diferenciais em relação a versão paga são a acessibilidade e o fato de que qualquer usuário pode criar melhorias para o avanço do programa, ou seja, ele é um **software** livre.

O *R*, assim como *Python*, é uma linguagem de programação, pois é um conjunto de símbolos e códigos que surgiram da evolução da lógica matemática que orientam programações e relações, através de linhas de comandos, para executar determinadas tarefas no computador local ou na web; permitir o uso de informações dentro de certos programas [Medina and Ferting, 2006]. Porém o **R** é mais voltado para estatística do que o *Python*. O **R** é consolidado no mercado e com ótimos conteúdos instrucionais, também é conhecido por ser didático e de fácil compreensão [Verzani, 2011]. Essa facilidade no manuseio se deve por ser uma linguagem multiparadigma orientada a objetos, atualmente a integração com outros softwares é muito abrangente e simplificada.

Por esses motivos, a linguagem é adotada por várias empresas reconhecidas, como é o caso do Facebook, Bing, Google, Microsoft e outros. O que o torna tão chamativo é o foco em trabalhar com banco de dados e a flexibilidade para o uso de outros recursos que as linguagens de programação possuem. As visualizações gráficas são elegantes, tendo uma gama enorme de métodos para trabalhar com *big data* e *machine learning* [R Core Team, 2021a]. Como o principal foco do **R** é a área de estatística, este possui uma vasta gama de ferramentas para análise de dados, testes e visualizações de resultados. Além de tudo isso, ainda inclui funcionalidades essenciais em qualquer linguagem de programação,

como *loops*, *if/else*, possibilidade da criação de funções e estruturas, entre outras. Uma característica interessante na sua estrutura é o *help* das funções que combinado com documentações existentes na *web* torna o seu uso atrativo e de fácil acesso para profissionais que desejam programar sem a formação em ciência da computação..

Segundo Wickham and Grolemund [2016, p. 4] “Se você não pode colocar seus dados em R, você não pode fazer ciência de dados sobre ele!”.

### 4.2.1 RSelenium

O RSelenium [Harrison, 2020] é um pacote desenvolvido para automatizar o processo de conexão e captação de dados na web, também conhecido como raspagem de dados na internet. Esse pacote facilita a conexão entre o servidor Selenium remoto que são ferramentas para criação de *script* automáticos e o *driver* (software que possibilita a comunicação entre o aplicativo e o sistema operacional), fornecendo a ligação necessária para que o usuário possa utilizar de todas as informações que os sites disponibilizam.

Este trabalho foi construído com intuito de recolher as informações do site do SofaScore com o auxílio do RSelenium, pois este pacote dá mais autonomia e maleabilidade para o usuário captar os dados de diferentes sites que podem dificultar a raspagem de dados. Uma singularidade deste pacote é o uso do JavaScript para carregar elementos ou dados HTML adicionais, já os outros pacotes para realizar o *web scraping* tendem a carregar o HTML das páginas, como é o caso do rvest [Wickham, 2021]. O RSelenium consegue usar o JavaScript para carregar informações do site especificado, pois ele utiliza um navegador real no próprio computador e carrega as informações da página aberta.

A utilização do RSelenium ocorreu da seguinte maneira. Primeiro, é estabelecida a conexão entre o R e o *browser* utilizando a função `rsDriver` que está descrita abaixo. Como argumentos da função foram escolhidos no exemplo, o browser que é o navegador que será utilizado como o Mozilla Firefox, e a porta de comunicação que relacionam programas e computadores, no exemplo foi escolhida a porta 4545 ao invés da 4444. É preciso escolher uma porta que não esteja sendo utilizada no computador.

```
(RSelenium)
rD <- rsDriver(browser= "firefox", port= 4545L, verbose= F)
remDr <- rD[["client"]]
```

Este é o comando principal para acessar o *browser* através do R, de forma simplificada é como se abrisse uma porta para retirar as informações. Para que isso ocorra o comando é processado dentro do servidor pelo server, que é um sistema computacional que fornece serviços a uma rede de softwares chamada de cliente, já o **software** cliente é responsável por exibir o comando na tela. Além disso, quando a função é guardada em um objeto, por exemplo "rD", faz com que seja possível atribuir o **software** cliente à um

outro objeto, e.x. "remDr". Através dessa conexão, já é possível extrair informações no *browser* e exibi-las.

Mais detalhes sobre o pacote podem ser observados no Package 'RSelenium' [Harrison, 2020], disponível em <https://cran.r-project.org/web/packages/RSelenium/RSelenium.pdf>, último acesso em 02 de setembro de 2021 .

# 5 SOFASCORE

## 5.1 Site

O site **SofaScore** foi criado por Zlatko Hrkać e Ivan Bešlić em 2010 na Croácia. Já nesse ano, compraram o primeiro *feed* de pontuação ao vivo; pouco tempo depois já haviam criado seu aplicativo móvel; em 2014, houve uma mudança passando de resultado ao vivo para análise de dados esportivos em tempo real, com apoio da Opta Sports para realizar essa tarefa [SofaScore, 2022].

Os produtos oferecido pela marca são a disponibilização de informações sobre variados esportes, porém com foco no futebol. O site descreve detalhadamente através de números e imagens o passado e presente do futebol mundial, detalhando informações por campeonatos, times e jogadores. Exemplos de produtos seriam, pontuações dos times e jogadores, tanto em campeonatos quanto em temporadas, estatísticas de jogos, mapas de calor de jogadores, mapas de radares, cobertura de jogos em tempo real e muito mais. Todo o trabalho é bem completo para o público na apresentação de resultados.

Atualmente são mais de 20 milhões de usuários ativos mensais, onde seu maior mercado são os brasileiros sendo dois milhões de usuários. Além disso, possui escritórios físicos apenas em Zagreb e no Rio de Janeiro, onde é a nacionalidade da marca e seu maior público respectivamente. O fato do Brasil ser considerado o país do futebol é favorável ao **SofaScore**, afinal o país que mais cultua esse esporte segue consideravelmente o que o site apresenta [Helal, 2001].

Além do site, a marca **SofaScore** também oferece aplicativo para aparelhos móveis, cobrindo mais de 11.000 torneios, oferecendo um serviço com mais de 30 idiomas diferentes e para além de 20 opções de esportes. A empresa também oferece garantia sobre seus produtos, sempre reforçando o fato de estar entregando uma experiência completa, única e personalizada.

Hoje, o **SofaScore** é uma ferramenta analítica para os amantes do esporte de forma geral, pois além do futebol também faz cobertura de outros esportes, como o basquete, tênis, vôlei, rugby e vários outros. O principal marco nessa trajetória foi a introdução de classificações estatísticas em 2015, algo que é feito por algoritmos que quantificam através de modelos o desempenho dos jogadores. Os números deste trabalho são bem expressivos, até hoje já foram classificados mais de 1,5 milhão de jogadores em mais de 50.000 partidas [Helal, 2001].

Essas inovações no campo de análise esportiva geraram outros produtos exclusivos

da marca, o que impacta significativamente na opinião do mercado. Hoje, a SofaScore é o aplicativo de esporte com a classificação mais alta do mercado em satisfação do cliente, com nota 4,8 e 4,9 no Google Play e App Store, respectivamente. Além de tudo isso que foi mencionado, o SofaScore participa de outros projetos com intuito de estar sempre inovando e atendendo às expectativas dos clientes.

A grande audiência do site o transformou em uma das plataformas mais atraentes para quem deseja fazer propaganda no mercado esportivo, que é o caso do mercado de apostas. Esse mercado é maleável as necessidades e aos gostos da sociedade, prendendo a atenção da população. Ademais, o aplicativo é um dos serviços de dados relacionados ao futebol mais utilizados no planeta, com seu maior mercado sendo brasileiro e um dos seus principais produtos sendo a análise do Campeonato Brasileiro. Pensando nisso, esse projeto de mestrado tem o intuito de fornecer algo que, no melhor do nosso conhecimento, não existe hoje no mercado, e que movimenta uma quantidade significativa de dinheiro e tempo da população em geral, que são os dados extraídos do site SofaScore, para o usuário possa trabalhar, investigar e analisar da forma que achar melhor o cenário futebolístico.

## 5.2 Pacote

Pensando no objetivo do site e na aplicação da ciência de dados, surgiu a ideia de disponibilizar os dados extraídos do site SofaScore em uma plataforma de fácil acesso como o R, para que os usuários pudessem fazer a análise dos dados com sua especificidade. Deste modo, essa dissertação visa criar um pacote que consiga fazer *web scraping* com as informações do SofaScore e disponibilizá-las.

Um dos principais interesses de quem analisa dados é recebê-los creditados e prontos para serem trabalhados com facilidade e liberdade, ademais uma grande particularidade que é atrativa do site é disponibilizar a avaliação própria e lógica de elementos do futebol que são jogadores e times. Observando esses pontos é importante entregar a condição para que os dados do SofaScore sejam manipulados, que é o que o pacote SofaScore oferece. Percebendo as diferentes formas do uso do futebol e do esporte, atualmente, nas áreas econômica, cultural e do ensino, conseguimos entender que as análises de dados esportivos têm diferentes objetivos. Com esse pacote, visamos permitir que todas essas finalidades sejam alcançadas.

Para a criação desse novo pacote, o `RSelenium` foi utilizado para abrir o *browser*, navegar dentre as páginas necessárias, selecionar e importar as informações. Este não é um processo simples, pois por trás de toda a programação é necessário entender como unir as funcionalidades disponíveis no pacote com a estrutura e especificidades HTML e JavaScript das páginas do site de interesse.

Aqui serão apresentadas algumas funcionalidades do pacote SofaScore. Primeiramente, para baixa-lo é necessário executar os seguintes comandos

```
devtools::install_github("Vitor-FCO/SofaScore")  
library(SofaScore)
```

Atualmente, o pacote conta com 9 funções que podem ser vistas com a sua respectiva descrição na Tabela 1. O detalhamento dos argumentos, funcionalidades e utilização serão apresentados na subseção a seguir.

Tabela 1 – Funções pertencentes ao pacote SofaScore

Função	Descrição
TabLinkPlayers	Função retorna a tabela por time com os nomes dos jogadores e <i>url</i> da página de cada um no SofaScore
Open_Sofa	Função que abre o browser na página do Brasileirão do SofaScore.
Close_Sofa	Função que fecha o browser que foi aberto.
Brasileiro_A	Função que retorna a tabela do Brasileirão.
Last_Games	Função que retorna os últimos jogos do time.
Upcoming_Games	Função que retorna os próximos jogos do time.
Players	Função que retorna os jogadores de determinado time.
Statistics_Games	Função que retorna as estatísticas dos jogos que o time participou.
Heat_Map	Função que retorna o mapa de calor da localização do jogador na última temporada.

## 5.2.1 Tutorial de navegação

Nesta subseção serão apresentadas em detalhes as funções que compõem o pacote, mostrando como elas devem ser utilizadas e o que essas fornecem, este será um manual do usuário para o pacote `SofaScore`. Aqui o leitor será capaz de operar o pacote no R conseguindo extrair os dados de times, jogadores e partidas do Campeonato Brasileiro Serie A.

### 5.2.1.1 Função `Open_Sofa`

Assim como o comando o `rsDriver`, descrito na Seção 4.1.1, a função `Open_Sofa` é necessária para utilização das outras funções do pacote. Ela abre o *browser* na página do `SofaScore`, e faz a ligação entre o `Selenium` e o *API Java Script*, traduzindo e transformando o código de servidor em código de cliente.

```
(SofaScore)
server <- Open_Sofa(port= 4445L, browser= "firefox")
```

Ao executar esse comando, o *Mozilla Firefox* já irá abrir uma nova página no site <https://www.sofascore.com/pt/torneio/futebol/brazil/brasileiro-serie-a/325>.

### 5.2.1.2 Função `Close_Sofa`

Caso o usuário queira fechar o *browser*, é necessário usar a função `Close_Sofa`, como ilustrado a seguir:

```
(SofaScore)
Close_Sofa(server= server)
```

O pacote, por ser recente, atualmente oferece recursos para extrair informações do Campeonato Brasileiro entre os anos de 2008 e 2021, que é exatamente o que o site do SofaScore oferece para o campeonato em questão.

Para conseguir extrair essas informações, deverão ser usadas as funções do pacote enquanto o *browser* que foi aberto pela função `Open_Sofa` permanecer acessível.

### 5.2.1.3 Função `Brasileiro_A`

A primeira função que retorna dados para análise é a função `Brasileiro_A`, que fornece de forma geral a tabela do campeonato com as seguintes variáveis: classificação, nome do time, nº de jogos, nº de vitórias, nº de empates, nº de derrotas, nº de gols feitos, nº de gols tomados e pontos no campeonato.

```
(SofaScore)
Brasileiro_A(server= server, type= "Overall", year= 2021)
```

Quando a função é requerida, a saída do `software` será:

Tabela 2 – Tabela fornecida utilizando a função `Brasileirão_A`, tipo “Overall” para o ano de 2021.

Classification	Team	P	W	D	L	Goals	Pts
1	Atlético MG	38	26	6	6	67:34	84
2	Flamengo	38	21	8	9	69:36	71
3	Palmeiras	38	20	6	12	58:43	66
4	Fortaleza	38	17	7	14	44:45	58
5	Corinthians	38	15	12	11	40:36	57
6	RB Bragantino	38	14	14	10	55:46	56
7	Fluminense	38	15	9	14	38:38	54
8	América MG	38	13	14	11	41:37	53
9	Atlético GO	38	13	14	11	33:36	53
10	Santos	38	12	14	12	35:40	50
11	Ceará	38	11	17	10	39:38	50
12	Internacional	38	12	12	14	44:42	48
13	São Paulo	38	11	15	12	31:39	48
14	Athletico PR	38	13	8	17	41:45	47
15	Cuiabá	38	10	17	11	34:37	47
16	Juventude	38	11	13	14	36:44	46
17	Grêmio	38	12	7	19	44:51	43
18	Bahia	38	11	10	17	42:51	43
19	Sport Recife	38	9	11	18	24:37	38
20	Chapecoense	38	1	12	25	27:67	15

Tabela obtida em 03/03/2022 às 09:00

#### 5.2.1.4 Função `Last_Games`

Essa função informa quais foram os últimos confrontos que o time enfrentou, independente do campeonato. Algo importante a se destacar é o uso da programação para selecionar botões no site, mudando assim o conteúdo da página.

Aqui devem ser informados dois objetos: o primeiro é o “`server`”, como foi feito no exemplo acima, onde ele tem as informações do servidor; e o segundo é o time que deseja analisar as partidas. Para saber exatamente como escrever o nome do time, solicite a tabela “`TabLink`” que será fornecida ao carregar o pacote. Após chamar a tabela, aparecerão os nomes dos times na primeira coluna, escreva o nome exatamente da forma que aparece na tabela.

```
(SofaScore)
Table <- Last_Games(server= server, team= "Atlético MG")
head(Table, 10)
```

A função `Last_Games` retorna uma nova tabela onde as colunas informam o campeonato que o jogo pertence, a data do jogo, a situação do jogo (caso tenha sido adiado,

cancelado, resolvido nos pênaltis ou alguma outra forma que o jogo foi realizado), time que jogou em casa, time que jogou fora de casa, gols feitos pelo time de casa e gols feitos pelo time fora de casa, nessa ordem. Quando a função é requerida, a saída do **software** será a Tabela 3.

Tabela 3 – Tabela dos últimos 10 jogos, fornecida utilizando a função `Last_Games` e especificando o time Atlético MG.

Championship	Date	Game Situation	Home Team	Away Team	Home Team Goals	Away Team Goals
Mineiro Módulo I	26/02/2022	FT	Pouso Alegre FC	Atlético Mineiro	2	3
Supercopa do Brasil	20/02/2022	AP	Atlético Mineiro	Flamengo	2	2
Mineiro Módulo I	15/02/2022	FT	Atlético Mineiro	Athletic Club	1	0
Mineiro Módulo I	12/02/2022	FT	América Mineiro	Atlético Mineiro	0	2
Mineiro Módulo I	09/02/2022	FT	URT	Atlético Mineiro	1	0
Mineiro Módulo I	06/02/2022	FT	Atlético Mineiro	CA Patrocinense	3	0
Mineiro Módulo I	02/02/2022	FT	Uberlândia	Atlético Mineiro	0	4
Mineiro Módulo I	29/01/2022	FT	Atlético Mineiro	Tombense	3	0
Mineiro Módulo I	26/01/2022	FT	Villa Nova MG	Atlético Mineiro	1	1
Copa do Brasil	15/12/2021	FT	Athletico Paranaense	Atlético Mineiro	1	2

Tabela obtida em 03/03/2022 às 09:50

5.2.1.5 Função `Upcoming_Games`

Essa função informa quais serão os próximos confrontos que o time terá, independente do campeonato.

Assim como a função `Last_Games`, devem ser informados dois objetos: o primeiro é o “`server`”, como foi feito no exemplo acima, onde ele tem as informações do servidor; e o segundo é o time que deseja analisar as partidas. Para saber exatamente como escrever o nome do time, solicite a tabela “`TabLink`” que será fornecida ao carregar o pacote.

```
(SofaScore)
Table <- Upcoming_Games(server= server, team= "Atlético MG")
head(Table, 10)
```

A saída é uma tabela onde as colunas informam o campeonato que o jogo pertence, a data do jogo, horário que o jogo irá ocorrer, time que joga em casa, time que joga fora de casa, respectivamente. A tabela abaixo ilustra a saída da função quando esta é requerida.

Tabela 4 – Tabela dos próximos 10 jogos, fornecida utilizando a função `Upcoming_Games` e especificando o time Atlético MG.

Championship	Date	Hora	Home Team	Away Team
Mineiro Módulo I	06/03/2022	18:00	Atlético Mineiro	Cruzeiro
Mineiro Módulo I	12/03/2022	16:30	EC Democrata	Atlético Mineiro
Mineiro Módulo I	19/03/2022	16:30	Atlético Mineiro	Caldense
Brasileiro Série A	09/04/2022	15:00	Atlético Mineiro	Internacional
Brasileiro Série A	16/04/2022	15:00	Athletico Paranaense	Atlético Mineiro
Brasileiro Série A	23/04/2022	15:00	Atlético Mineiro	Coritiba
Brasileiro Série A	30/04/2022	15:00	Goiás	Atlético Mineiro
Brasileiro Série A	07/05/2022	15:00	Atlético Mineiro	América Mineiro
Brasileiro Série A	14/05/2022	15:00	Atlético Mineiro	Atlético Goianiense
Brasileiro Série A	21/05/2022	15:00	Red Bull Bragantino	Atlético Mineiro

Tabela obtida em 03/03/2022 às 17:17

## 5.2.1.6 Função Players

Aqui será entregue uma tabela informando o elenco do time e algumas características dos jogadores.

```
(SofaScore)
Players(server= server , team= "Atlético MG" , position= "Forward" , time= 2)
```

Caso não seja escolhido um time ou uma posição, será gerada uma saída com todas as opções, ou seja, elenco de todos os times analisados para todas posições em campo. Posições que podem ser escolhidas: “Goalkeeper”, “Defender”, “Midfielder”, “Forward”. O argumento “time” na função pode ser modificado com o tempo em segundos que o computador do usuário necessita para carregar a página da web completamente.

A saída é uma tabela que apresenta a seguinte ordem de variáveis para as colunas: nome do time, número da camisa do jogador, nome do jogador, posição que ele joga, idade e nacionalidade do jogador. Quando a função é requerida, a saída do **software** será a Tabela 5.

Tabela 5 – Tabela fornecida utilizando a função **Players**, especificando o time Atlético MG e a posição dos jogadores como atacantes.

Team	Number	Player	Position	Age	Nationality
Atl. Mineiro	7	Hulk	Forward	35	BRA
Atl. Mineiro	10	Eduardo Vargas	Forward	32	CHI
Atl. Mineiro	33	Sávio	Forward	17	BRA
Atl. Mineiro	18	Eduardo Sasha	Forward	30	BRA
Atl. Mineiro	9	Fábio Gomes	Forward	24	BRA

Tabela obtida em 03/03/2022 às 13:30

### 5.2.1.7 Função `Statistics_Games`

O comando `Statistics_Games` foi criado para que o usuário consiga aprofundar mais nas informações do `Last_Games`, ou seja, esse informa estatísticas específicas analisadas em jogos de futebol, como: Posse de bola; Número de Finalizações; Número de Finalizações no gol; Número de Finalizações para fora; Número de Escanteios; Número de Impedimentos; Número de Faltas; Número de Cartões; Número de Defesas do goleiro.

Essas informações fornecidas são desejáveis por muitos analistas de jogo, porém, não são simples de se conseguir. Importante destacar também que nem todas são estatísticas exatas, algumas o conceito pode dar margem a mais de uma interpretação, nesses casos o usuário deve ter em mente a forma como o `SofaScore` calcula cada dado. Outro ponto a considerar é que nem todos campeonatos são calculadas as estatísticas por jogos, exemplo disso é o Campeonato Mineiro que não possui esses resultados no site.

`Statistics_Games` necessita como entrada os seguintes elementos, que já foram descritos em outras funções, “server”; “team”. Que são respectivamente: o controle do *browser*; o time de interesse do utilizador do pacote. O argumento “time” pode ser especificado pelo usuário, ele é o tempo em segundos que a máquina precisa para carregar a página do `SofaScore` (seu *default* é de 1 segundo). Como saída serão informadas as estatísticas dos últimos jogos do time, tanto do time quanto de seus adversários.

Importante observar o time que joga em casa (mandante) e seu adversário, na coluna `Game` o time que aparece primeiro é o time mandante, logo os valores das suas estatísticas são informados na coluna `Home Team`; do segundo time que aparece na coluna `Game`, são descritos na coluna `Away Team`.

```
(SofaScore)
Statistics_Games(server= server, team= "Atlético MG", time=1)
```

Quando a função é requerida, a saída será como apresentado na Tabela 6 que está no Anexo A, essa tabela apresenta 28 estatísticas diferentes em um único jogo, a função utilizada no `software` apresentará essas mesmas estatísticas, porém replicada para um número maior de jogos.

### 5.2.1.8 Função Heat\_Map

Diferente das outras funções descritas anteriormente `Heat_Map` não tem como saída uma tabela e sim uma imagem, é fornecido um mapa de calor da localização do jogador dentro do campo durante uma partida, ilustrando as regiões onde ele se posiciona em maior quantidade de tempo com o tom de vermelho mais intenso.

A Figura 2 apresenta o mapa de calor para o jogador Hulk do Atlético-MG na temporada de 2021. A seta acima da imagem do campo de futebol indica qual lado é o de ataque. O mapa ilustra bem a região de atuação do jogador dentro de campo, ajuda o usuário a entender a forma como o jogador trabalha e a posição em que joga.

Para gerar o mapa é preciso indicar os seguintes argumentos de entrada na função, o controle do *browser* (`server`), o time no qual o jogador que se deseja analisar está atualmente e o nome do jogador, conforme especificado na função `TabLinkPlayer`.

A saída do comando é uma imagem, porém se ele for gravado em um objeto, esse objeto será a própria imagem, e para que o R gere essa visualização a qualquer momento o usuário pode usar a função `imageShow()` do pacote `OpenImageR`.

```
(SofaScore)
Heat_Map(server = server, player = "Hulk", team= "Atlético MG")
```

Quando a função é requerida, a saída do `software` será a Figura 2 :

Figura 2 – Mapa de calor da temporada do jogador Hulk, time Atlético MG.



Imagem obtida em 06/02/2022 às 02:00

#### 5.2.1.9 Tabela TabLink

O pacote, quando carregado, já fornece o objeto `TabLink` que disponibiliza uma tabela com duas colunas: a primeira com os nomes dos times que disputam o Campeonato Brasileiro série A, e a segunda com o link do site `SofaScore` referente a cada time.

Esse objeto já fica guardado na memória local do R assim que o pacote é chamado. Se chamado, a saída do *software* será a Tabela 8 apresentada no Anexo C.

#### 5.2.1.10 Função TabLinkPlayers

A função `TabLinkPlayers` fornece uma planilha de consulta com os nomes dos jogadores e seus links respectivos do `SofaScore`. Ideia semelhante ao objeto `TabLinks`, porém em um nível mais específico. Essa tabela é fundamental para o uso de outras funções que serão descritas posteriormente. Para utilização dessa tabela é necessário informar o time qual serão indicados os jogadores e os seus respectivos hyperlinks no site, também é preciso especificar o objeto `server` da classe `SofaScore` para que seja feita a ligação da função com o *browser*.

```
(SofaScore)
TabLinkPlayers(server= server, team= "Atlético MG")
```

Quando a função é requerida, a saída será como apresentado na Tabela 8 que está no Anexo C.

## 6 Conclusão

### 6.1 Resultados

Pretendemos com essa dissertação de Mestrado mostrar não só a importância de ter dados através de técnicas atuais de ciência de dados como proporcioná-los. Conseguimos gerar uma quantidade substancial de funções e informações gerando condições dos usuários fazerem análises com qualidade. A forma como as informações são entregues a quem usufrui do pacote **SofaScore** facilita a análise dos dados, pois as funções geram resultados em formato semelhante ao que o site disponibilizam seus recursos. O que diferencia muito o site e o pacote para o analista é a capacidade de manipular os dados, pois eles são entregues digitalizados em tabelas no nosso produto, onde podem ser guardadas em objetos. O fato de estar em um *software* voltado à estatística também traz grandes vantagens, pois facilmente podem ser gerados gráficos e visualizações com esses dados.

Contudo, utilizando as funções já disponíveis no pacote, o usuário tem acesso e poder de manipulação dos dados que são de conhecimento geral do Campeonato Brasileiro Serie A com detalhes, ou seja, o pacote informa o número de jogos de cada time, o quanto esse conseguiu pontuar no campeonato, destrinchado quantos desses pontos vieram de empates, vitórias ou derrotas. Sabe-se também o número de gols que o time recebeu ou fez. Tudo isso de forma geral, ou separando jogos que foram em casa e fora de casa.

Como pode ser observado o dispositivo criado apresenta o elenco de todos os times do campeonato, mostrando características do jogador como idade, nacionalidade e posição, que são variáveis importantes quando analisa-se preço de compra e venda de jogadores [Globo, 2018]. Também apresenta mapa de calor, informando onde o jogador se concentrou mais, dentro do campo durante a última temporada. É válido lembrar que para fornecer essa imagem é necessária uma análise de dados, feita pela equipe do **SofaScore**.

Os dados extraídos pelo pacote deixam o usuário a par dos últimos jogos de cada time e das partidas futuras, e apresenta de forma completa essas informações, mostrando resultados dos jogos que ocorreram, se o time jogou em casa ou não, se teve pênaltis, prorrogação, entre outras formas onde o jogo pode ser decidido.

Nota-se, portanto que o trabalho gerou uma série de frutos que poderão ser utilizados em estudos nas áreas de saúde, saúde física, esporte e várias outras. Através dos elementos disponíveis, o usuário consegue estimar possíveis cenários que são capazes de acontecer. Um exemplo seria nos próximos jogos determinar em qual classificação os times estarão através de possíveis resultados dos próximos jogos.

## 6.2 Trabalhos Futuros

Os próximos passos que estão dentro do planejamento para a evolução deste projeto são ampliar o número de funções disponíveis e campeonatos analisados, trazendo mais informações e estatísticas disponíveis no site, estas são calculadas através de modelos específicos do **SofaScore**.

Também é de interesse do autor extrapolar e utilizar outras ferramentas estatísticas utilizando os dados que o pacote fornece. Está cada vez mais latente a necessidade de encontrar soluções que auxiliarão na previsão de resultados, a ideia é que nos próximos trabalhos advindo desse, possamos criar modelos de previsões para os jogos e outros elementos do esporte.

Hoje o pacote já está disponível no **github** para ser baixado, o que não impacta no uso dele dentro do **software R**. Porém um próximo passo que será feito é coloca-lo disponível no **CRAN**, assim ele passará por uma verificação meticulosa, facilitando também seu acesso no **R**.

# Referências

- Fernando Amaral. *Introdução à ciência de dados: mineração de dados e big data*. Alta Books Editora, 2016.
- Heleno Bolfarine and Wilton de Oliveira Bussab. *Elementos de amostragem*. Editora Blucher, 2005.
- Danah Boyd and Kate Crawford. Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication, and Society*, 15(5):24–25, 2012.
- Bruno Cardoso and João Magalhães. Google, bing and a new perspective on ranking similarity. In *Proceedings of the 20th ACM international conference on information and knowledge management*, pages 1933–1936, 2011.
- Salvatore A Catanese, Pasquale De Meo, Emilio Ferrara, Giacomo Fiumara, and Alessandro Provetti. Crawling facebook for social network analysis purposes. In *Proceedings of the international conference on web intelligence, mining and semantics*, pages 1–8, 2011.
- Lara Fabiana Dallabona, Sabrina Nascimento, and Nelson Hein. Métodos estatísticos mais recorrentes nas dissertações do programa de pós-graduação em ciências contábeis da furb. *Revista de Contabilidade da UFBA*, 4(1):56–70, 2010.
- Diário do estado. *10 Esportes mais populares do mundo*. Diário do estado, 2021. URL <https://diariodoestado.go.com.br/10-esportes-mais-populares-do-mundo-108500/>.
- ESPN. *Final da Champions League já mexe com mercado e deve ultrapassar SuperBowl*. ESPN, 2012. URL [https://www.espn.com.br/blogs/fernandofleury/254994\\_final-da-champions-league-ja-mexe-com-mercado-e-deve-ultrapassar-superbowl/](https://www.espn.com.br/blogs/fernandofleury/254994_final-da-champions-league-ja-mexe-com-mercado-e-deve-ultrapassar-superbowl/).
- EXAME. *Transmissão internacional do Brasileirão superou 100 países no Exterior*. EXAME, 2021. URL <https://exame.com/blog/esporte-executivo/transmissao-internacional-do-brasileirao-superou-100-paises-no-exterior/>.
- FBref. *All About FBref.com*. FBref, 2021. URL <https://fbref.com/en/about/>.
- Emilio Ferrara, Pasquale De Meo, Giacomo Fiumara, and Robert Baumgartner. Web data extraction, applications and techniques: A survey. *Knowledge-based systems*, 70: 301–323, 2014.
- Aurélio Buarque de Holanda Ferreira. Novo dicionário aurélio da língua portuguesa. In *Novo dicionário Aurélio da língua portuguesa*, pages 2120–2120. 2009.

Forbes. *10 atletas mais bem pagos do mundo; Neymar é o sexto*. Forbes, 2021. URL <https://forbes.com.br/listas/2021/05/10-atletas-mais-bem-pagos-do-mundo-neymar-e-o-sexto/#foto4>.

Globo. *Como é definido o preço de um jogador de futebol?* Globo, 2018. URL <https://g1.globo.com/economia/educacao-financeira/blog/samy-dana/post/2018/05/31/como-e-definido-o-preco-de-um-jogador-de-futebol.ghtml>.

Maria Alice Cabral de Goes, Adão Benvido da Luz, and Mario Valente Possa. *Amostragem*. CETEM/MCT, 2010.

Ronaldo Goldschmidt and Emmanuel Passos. *Data mining*. Gulf Professional Publishing, 2005.

Google. *Google Trends*. Google, 2022. URL <http://www.google.com/trends/>.

John Harrison. *RSelenium: R Bindings for 'Selenium WebDriver'*, 2020. URL <https://CRAN.R-project.org/package=RSelenium>. R package version 1.7.7.

Ronaldo Helal. *A invenção do país do futebol: mídia, raça e idolatria*. Mauad Editora Ltda, 2001.

Vlad Krotov, Leigh Johnson, and Leiser Silva. *Tutorial: Legality and ethics of web scraping*. 2020.

Marco Medina and Cristina Ferting. *Algoritmos e programação: teoria e prática*. Novatec Editora, 2006.

Ryan Mitchell. *Web scraping with Python: Collecting more data from the modern web*. "O'Reilly Media, Inc.", 2018.

MKTESPORTIVO. *Premier League e La Liga se destacam com audiência e novos acordos de patrocínio*. MKTESPORTIVO, 2019. URL <https://www.mktesportivo.com/2019/08/premier-league-e-la-liga-se-destacam-com-audiencia-e-novos-acordos-de-patrocinio>

Federico Polidoro, Riccardo Giannini, Rosanna Lo Conte, Stefano Mosca, and Francesca Rossetti. *Web scraping techniques to collect data on consumer electronics and airfares for italian hicp compilation*. *Statistical Journal of the IAOS*, 31(2):165–176, 2015.

R Core Team. *What is R?* R Foundation for Statistical Computing, Vienna, Austria, 2021a. URL <https://www.r-project.org/about.html>.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021b. URL <https://www.R-project.org/>.

- LATINDEX Redalyc, REVENCIT Clase, and SERBILUZ IN-COM UAB. Berlin declaration on open access to knowledge in the sciences and humanities. 2003.
- SofaScore. *About Us*. SofaScore, 2022. URL <https://corporate.sofascore.com/about>.
- Yang Sun, Ziming Zhuang, and C Lee Giles. A large-scale study of robots. txt. In *Proceedings of the 16th international conference on World Wide Web*, pages 1123–1124, 2007.
- Alex Teloken et al. Estudo comparativo entre os algoritmos de mineração de dados random forest e j48 na tomada de decisão. *Simpósio de Pesquisa e Desenvolvimento em Computação*, 2(1), 2016.
- Transfermarkt. *20 anos de Transfermarkt: Do futebol regional para a Champions League*. Transfermarkt, 2020. URL <https://www.transfermarkt.com.br/20-anos-de-transfermarkt-do-futebol-regional-para-a-champions-league/view/news/363227#:~:text=Faz%2020%20anos%20desde%20que,no%20mundo%20da%20informa%C3%A7%C3%A3o%20esportiva>.
- UOL. *Como aumentou a importância do futebol na renda da Globo*. UOL, 2019. URL <https://rodrigomattos.blogosfera.uol.com.br/2019/09/19/saiba-como-aumentou-a-importancia-do-futebol-na-renda-da-globo/>.
- Wil MP Van der Aalst. *Process mining: data science in action*. Springer, 2016.
- Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009. ISBN 1441412697.
- John Verzani. *Getting started with RStudio*. "O'Reilly Media, Inc.", 2011.
- Hadley Wickham. *rvest: Easily Harvest (Scrape) Web Pages*, 2021. URL <https://CRAN.R-project.org/package=rvest>. R package version 1.0.2.
- Hadley Wickham and Garrett Grolemund. *R for data science: import, tidy, transform, visualize, and model data*. "O'Reilly Media, Inc.", 2016.
- Jason Zivkovic. *worldfootballR: Functions to Extract and Clean World Football (Soccer) Data*, 2022. URL <https://CRAN.R-project.org/package=worldfootballR>. R package version 0.4.9.

# Anexos

## ANEXO A – Tabela Statistics\_Games

Tabela 6 – Tabela fornecida utilizando a função Statistics\_Games, especificando o time Atlético MG.

Championship	Game	Statistics	Home Team	Away Team
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Ball possession	0,57	0,43
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Total shots	20	14
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Shots on target	7	3
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Shots off target	6	6
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Blocked shots	7	5
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Corner kicks	5	10
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Offsides	1	2
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Fouls	12	13
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Yellow cards	3	3
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Big chances	1	5
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Big chances missed	0	3
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Counter attacks	0	1
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Counter attack shots	0	1
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Shots inside box	11	8
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Shots outside box	9	6
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Goalkeeper saves	1	5
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Passes	401	284
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Acc. passes	350 (87%)	216 (76%)
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Long balls	17/34 (50%)	14/36 (39%)
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Crosses	5/15 (33%)	4/19 (21%)
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Dribbles	6/14 (43%)	7/13 (54%)
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Possession lost	93	108
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Duels won	37	45
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Aerials won	9	10
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Tackles	10	17
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Interceptions	6	9
Supercopa do Brasil, Final	Atlético Mineiro - Flamengo	Clearances	17	16

Tabela obtida em 03/03/2022 às 14:03

## ANEXO B – Tabela TabLink

Tabela 7 – Objeto TabLink fornecido pelo Pacote SofaScore.

Team	Link
América MG	<a href="https://www.sofascore.com/pt/time/futebol/america-mineiro/1973">https://www.sofascore.com/pt/time/futebol/america-mineiro/1973</a>
Athletico PR	<a href="https://www.sofascore.com/pt/time/futebol/athletico-paranaense/1967">https://www.sofascore.com/pt/time/futebol/athletico-paranaense/1967</a>
Atlético GO	<a href="https://www.sofascore.com/pt/time/futebol/atletico-goianiense/7314">https://www.sofascore.com/pt/time/futebol/atletico-goianiense/7314</a>
Atlético MG	<a href="https://www.sofascore.com/pt/time/futebol/atletico-mineiro/1977">https://www.sofascore.com/pt/time/futebol/atletico-mineiro/1977</a>
Avai	<a href="https://www.sofascore.com/pt/time/futebol/avai/7315">https://www.sofascore.com/pt/time/futebol/avai/7315</a>
Botafogo	<a href="https://www.sofascore.com/pt/time/futebol/botafogo/1958">https://www.sofascore.com/pt/time/futebol/botafogo/1958</a>
Ceará	<a href="https://www.sofascore.com/pt/time/futebol/ceara-sc/2001">https://www.sofascore.com/pt/time/futebol/ceara-sc/2001</a>
Corinthians	<a href="https://www.sofascore.com/pt/time/futebol/corinthians/1957">https://www.sofascore.com/pt/time/futebol/corinthians/1957</a>
Coritiba	<a href="https://www.sofascore.com/pt/time/futebol/coritiba/1982">https://www.sofascore.com/pt/time/futebol/coritiba/1982</a>
Cuiabá	<a href="https://www.sofascore.com/pt/time/futebol/cuiaba/49202">https://www.sofascore.com/pt/time/futebol/cuiaba/49202</a>
Flamengo	<a href="https://www.sofascore.com/pt/time/futebol/flamengo/5981">https://www.sofascore.com/pt/time/futebol/flamengo/5981</a>
Fluminense	<a href="https://www.sofascore.com/pt/time/futebol/fluminense/1961">https://www.sofascore.com/pt/time/futebol/fluminense/1961</a>
Fortaleza	<a href="https://www.sofascore.com/pt/time/futebol/fortaleza-ec/2020">https://www.sofascore.com/pt/time/futebol/fortaleza-ec/2020</a>
Goiás	<a href="https://www.sofascore.com/pt/time/futebol/goias/1960">https://www.sofascore.com/pt/time/futebol/goias/1960</a>
Internacional	<a href="https://www.sofascore.com/pt/time/futebol/internacional/1966">https://www.sofascore.com/pt/time/futebol/internacional/1966</a>
Juventude	<a href="https://www.sofascore.com/pt/time/futebol/juventude/1980">https://www.sofascore.com/pt/time/futebol/juventude/1980</a>
Palmeiras	<a href="https://www.sofascore.com/pt/time/futebol/palmeiras/1963">https://www.sofascore.com/pt/time/futebol/palmeiras/1963</a>
RB Bragantino	<a href="https://www.sofascore.com/pt/time/futebol/red-bull-bragantino/1999">https://www.sofascore.com/pt/time/futebol/red-bull-bragantino/1999</a>
Santos	<a href="https://www.sofascore.com/pt/time/futebol/santos/1968">https://www.sofascore.com/pt/time/futebol/santos/1968</a>
São Paulo	<a href="https://www.sofascore.com/pt/time/futebol/sao-paulo/1981">https://www.sofascore.com/pt/time/futebol/sao-paulo/1981</a>
Grêmio	<a href="https://www.sofascore.com/pt/time/futebol/gremio/5926">https://www.sofascore.com/pt/time/futebol/gremio/5926</a>
Bahia	<a href="https://www.sofascore.com/pt/time/futebol/bahia/1955">https://www.sofascore.com/pt/time/futebol/bahia/1955</a>
Sport Recife	<a href="https://www.sofascore.com/pt/time/futebol/sport-recife/1959">https://www.sofascore.com/pt/time/futebol/sport-recife/1959</a>
Chapecoense	<a href="https://www.sofascore.com/pt/time/futebol/chapecoense/21845">https://www.sofascore.com/pt/time/futebol/chapecoense/21845</a>

Tabela obtida em 03/03/2022 às 16:24

## ANEXO C – Tabela TabLinkPlayers

Tabela 8 – Tabela fornecida utilizando a função TabLinkPlayers, especificando o time Atlético MG.

Player	Link
Hulk	<a href="https://www.sofascore.com/player/hulk/34705">https://www.sofascore.com/player/hulk/34705</a>
Eduardo Vargas	<a href="https://www.sofascore.com/player/eduardo-vargas/37006">https://www.sofascore.com/player/eduardo-vargas/37006</a>
Sávio	<a href="https://www.sofascore.com/player/savio/1046795">https://www.sofascore.com/player/savio/1046795</a>
Eduardo Sasha	<a href="https://www.sofascore.com/player/eduardo-sasha/143846">https://www.sofascore.com/player/eduardo-sasha/143846</a>
Fábio Gomes	<a href="https://www.sofascore.com/player/fabio-gomes/909487">https://www.sofascore.com/player/fabio-gomes/909487</a>
Nacho Fernández	<a href="https://www.sofascore.com/player/nacho-fernandez/340861">https://www.sofascore.com/player/nacho-fernandez/340861</a>
Matías Zaracho	<a href="https://www.sofascore.com/player/matias-zaracho/859970">https://www.sofascore.com/player/matias-zaracho/859970</a>
Jefferson Savarino	<a href="https://www.sofascore.com/player/jefferson-savarino/874739">https://www.sofascore.com/player/jefferson-savarino/874739</a>
Keno	<a href="https://www.sofascore.com/player/keno/787607">https://www.sofascore.com/player/keno/787607</a>
Tchê Tchê	<a href="https://www.sofascore.com/player/tche-tche/840398">https://www.sofascore.com/player/tche-tche/840398</a>
Allan	<a href="https://www.sofascore.com/player/allan/828839">https://www.sofascore.com/player/allan/828839</a>
Ademir	<a href="https://www.sofascore.com/player/ademir/922546">https://www.sofascore.com/player/ademir/922546</a>
Dylan Borrero	<a href="https://www.sofascore.com/player/dylan-borrero/980933">https://www.sofascore.com/player/dylan-borrero/980933</a>
Jair	<a href="https://www.sofascore.com/player/jair/268419">https://www.sofascore.com/player/jair/268419</a>
Otávio	<a href="https://www.sofascore.com/player/otavio/552686">https://www.sofascore.com/player/otavio/552686</a>
Guilherme Castilho	<a href="https://www.sofascore.com/player/guilherme-castilho/924228">https://www.sofascore.com/player/guilherme-castilho/924228</a>
Calebe	<a href="https://www.sofascore.com/player/calebe/1046796">https://www.sofascore.com/player/calebe/1046796</a>
Echaporã	<a href="https://www.sofascore.com/player/echapora/981418">https://www.sofascore.com/player/echapora/981418</a>
Diego Godín	<a href="https://www.sofascore.com/player/diego-godin/14933">https://www.sofascore.com/player/diego-godin/14933</a>
Guilherme Arana	<a href="https://www.sofascore.com/player/guilherme-arana/1127439">https://www.sofascore.com/player/guilherme-arana/1127439</a>
Guga	<a href="https://www.sofascore.com/player/guga/928134">https://www.sofascore.com/player/guga/928134</a>
Mariano	<a href="https://www.sofascore.com/player/mariano/40076">https://www.sofascore.com/player/mariano/40076</a>
Réver	<a href="https://www.sofascore.com/player/rever/39421">https://www.sofascore.com/player/rever/39421</a>
Dodô	<a href="https://www.sofascore.com/player/dodo/108481">https://www.sofascore.com/player/dodo/108481</a>
Nathan Silva	<a href="https://www.sofascore.com/player/nathan-silva/841129">https://www.sofascore.com/player/nathan-silva/841129</a>
Igor Rabello	<a href="https://www.sofascore.com/player/igor-rabello/753662">https://www.sofascore.com/player/igor-rabello/753662</a>
Vitor Mendes	<a href="https://www.sofascore.com/player/vitor-mendes/988873">https://www.sofascore.com/player/vitor-mendes/988873</a>
Éverson	<a href="https://www.sofascore.com/player/everson/870748">https://www.sofascore.com/player/everson/870748</a>
Rafael	<a href="https://www.sofascore.com/player/rafael/33132">https://www.sofascore.com/player/rafael/33132</a>
Matheus Mendes	<a href="https://www.sofascore.com/player/matheus-mendes/1018555">https://www.sofascore.com/player/matheus-mendes/1018555</a>

Tabela obtida em 03/03/2022 às 14:40