

# Impacto de Lesões no Desempenho de Clubes no Campeonato Brasileiro de Futebol em 2024

Frederico Abbud Reis<sup>1</sup>, Wagner Meira Jr<sup>1</sup>

<sup>1</sup>Departamento de Ciência da Computação – Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte – Minas Gerais – Brasil

*Resumo.* O projeto a seguir trata sobre o impacto de lesões no desempenho de clubes do futebol brasileiro na liga nacional no ano de 2024.

## 1. Introdução

No futebol moderno, a performance de um time depende de diversos fatores que vão desde a habilidade técnica dos jogadores até estratégias táticas elaboradas pelos treinadores. No entanto, um elemento crucial que frequentemente afeta o desempenho de uma equipe ao longo de uma temporada são as lesões dos jogadores. Elas não apenas limitam a participação de atletas em momentos decisivos, mas também demandam ajustes táticos, força substituições, podem afetar a coesão e entrosamento do time, além de impactar o fator mental do atleta lesionado e de seus companheiros de clube (Aragão et al. 2023), além do próprio impacto à saúde de quem sofre, podendo trazer consequência na convivência diária do mesmo. Dessa forma, entender o impacto das lesões sobre o rendimento das equipes é essencial para que clubes e gestores esportivos adotem estratégias mais eficazes de prevenção e mitigação dos riscos.

Estudos prévios sugerem que a ausência de jogadores chave impacta diretamente a média de pontos conquistados em campeonatos e, conseqüentemente, a posição final da equipe na tabela (Ekstrand 2013). Além disso, análises de dados de lesões e desempenho têm ganhado relevância na literatura esportiva, onde técnicas de ciência de dados, como aprendizado de máquina e análise estatística avançada, têm se mostrado eficazes para prevenir lesões e entender padrões de impacto (Rossi 2017). Diante disso, o presente estudo busca explorar esse tema ao combinar uma revisão teórica sobre o impacto das lesões com um levantamento tecnológico das ferramentas e dados disponíveis para análise de métricas envolvendo desempenho dos times nos campeonatos, para assim cruzar os dados com as lesões de seus atletas. Os dados serão obtidos de APIs e fontes de dados especializadas em futebol.

O objetivo dessa etapa foi, portanto, entender o quão grande pode ser o impacto da incidência de lesões no desempenho de clubes de futebol brasileiro no campeonato nacional do ano de 2024.

## 2. Motivação

### 2.1. Problemática

No futebol, os clubes fazem investimentos milionários por temporada, com a grande maioria deles almejando converter esse investimento feito em grandes títulos naquele ano. Contudo, há uma série de fatores que alteram o esperado para o andamento das competições conforme o planejado, sendo um deles a incidência de lesões. As lesões

se mostram um problema pois elas representam a ausência de jogadores em sessões de treino ou jogos (Dujon et al. 2006), podendo causar um prejuízo para a equipe sendo, em média, a cada 136 dias de ausências de jogadores por lesão (ou seja, o tempo somado de recuperação de todos os atletas que enfrentaram algum problema do tipo), o time deixa de somar um ponto na liga referente a seu país (Eliakim et al. 2020). Portanto, compreender o quanto a incidência de lesões impacta no desempenho de um clube é de suma importância, pois pode permear decisões importantes de investimentos nessa área, além de decisões técnicas sobre o risco e retorno de se jogar com atletas com alta probabilidade de lesão.

## **2.2. Solução Proposta**

Para esse problema, foi proposto um estudo de impacto de lesões no desempenho de clubes de futebol por temporadas. Atualmente, existem estudos que correlacionam a incidência de lesões com o desempenho de clubes de futebol (Eliakim et al. 2020)(Eirale et al. 2013)(Hägglund et al. 2013a), porém, estes possuem maior foco no continente europeu, e um no Qatar. Dessa forma, este será com maior enfoque no futebol brasileiro e sul-americano, buscando entender a diferença entre o impacto em competições de pontos corridos e competições de fases eliminatórias, dada a diferença de calendário entre os locais já analisados. Tanto no futebol europeu quanto no futebol qatari, a temporada se inicia no verão (bem como o campeonato brasileiro), porém o verão nos locais de estudo dos artigos supracitados (Qatar e Europa) se inicia no meio do ano. Ou seja, usualmente as temporadas se iniciam no mês de agosto e se estendem até julho (europa) ou abril (qatar). Soma-se a isso o fato das equipes brasileiras jogarem, em média, 20 partidas a mais por ano do que em comparação a equipes europeias (Folha de S.Paulo 2022), e qataris, que contam com somente 12 times em sua liga(Qatar Football Association 2025). Por conta dessas diferenças, a solução visa compreender o comportamento do impacto de lesões em clubes brasileiro no campeonato nacional, disputado no decorrer do ano inteiro, no sistema de pontos corridos e 38 rodadas (Confederação Brasileira de Futebol 2024). Por conta disso, foi suposto que poderia haver diferenças entre os estudos já feitos.

## **3. Trabalhos Relacionados e Relevantes**

### **3.1. Trabalhos Relacionados**

Para a análise dos trabalhos relacionados, foram considerados artigos com temática semelhante no que tange a relação de lesões e desempenho, além de outros artigos com temática relacionada a fisiologia e lesões como um todo, para que haja uma melhor compreensão sobre o tema base do projeto.

Este trabalho baseia-se em três artigos distintos: (Eliakim et al. 2020), (Eirale et al. 2013) e (Hägglund et al. 2013a) que abordam a temática das lesões no futebol, cada um explorando contextos e metodologias diferentes. Esses artigos servirão como fundamento teórico para o desenvolvimento das análises previstas para o próximo semestre.

O principal conceito que permeia os artigos de Eliakim, Eirale e Haggglund está na definição de lesão, descrita como “qualquer queixa física resultante de uma partida ou treino de futebol, independentemente da necessidade de atendimento médico

ou afastamento das atividades. Lesões que requerem atenção médica são denominadas ‘lesões com atenção médica’, enquanto aquelas que impedem o jogador de participar plenamente de treinos ou jogos futuros são chamadas de ‘lesões com perda de tempo’” (Dujon et al. 2006). Neste trabalho, serão consideradas apenas as lesões com perda de tempo, ou seja, aquelas que afastam os atletas de partidas, pelo fato delas setarem em consonância com o objeto de estudo deste trabalho, que são as lesões que de fato causam ausência em jogos dos clubes, e o quão isso impacta o desempenho deles.

Além disso, os artigos analisados investigam a relação entre esses tipos de lesões e o desempenho das equipes, adotando metodologias distintas, descritas a seguir.

O artigo “*Estimation of injury costs: financial damage of English Premier League teams’ underachievement due to injuries*” (Eliakim et al. 2020) explora a perspectiva financeira das lesões. Ele propõe que o desempenho financeiro das equipes está diretamente relacionado ao seu desempenho esportivo. Para essa análise, foi utilizado um *ranking* baseado no valor de mercado das equipes, obtido no site *transfermarkt.com*. Em seguida, a colocação final no campeonato foi comparada à colocação esperada pelo *ranking*, estabelecendo uma correlação entre as duas variáveis.

Por outro lado, o artigo “*Low injury rate strongly correlates with team success in Qatari professional football*” (Eirale et al. 2013) adota uma abordagem mais direta, focando na quantidade de lesões e seu impacto no desempenho esportivo. Ele compara os índices de lesões dos times da primeira divisão do futebol do Catar com métricas de desempenho em campo, como gols marcados e pontos obtidos, evidenciando uma correlação significativa entre baixos índices de lesões e sucesso esportivo.

Já o artigo “*Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study*” (Hägglund et al. 2013a) trata de um estudo realizado por mais de uma década com clubes europeus, que compara o desempenho da equipe com ela mesma ao longo dos anos. Ou seja, a comparação é feita internamente entre os próprios clubes e correlaciona-se isso ao total de lesões ocorridas em cada uma das temporadas.

Por fim, o artigo (Warnock 2018) traz uma análise semelhante, porém aplicada ao *hockey* no gelo na *NHL*, liga do esporte praticada por times canadenses e norte americanos. Ele traz uma perspectiva da forma que são afetados os desempenhos coletivos e individuais por essa temática, de forma parecida com o anterior, analisando os clubes e atletas dentro de si, comparando-se o quão afetado foi por conta dos problemas físicos.

Apesar de tomarem como objeto de estudo de estudos ligas e usarem metodologias diferentes entre si, todos convergiram para a conclusão que uma alta incidência de lesões impacta diretamente no desempenho das equipes.

### **3.2. Trabalhos Relevantes**

Ademais, foram também considerados outros estudos no que tange a lesões em si, para maior compreensão de termos e obter maior familiaridade com a temática de lesões em esportes. A maior parte deles possui um enfoque maior na fisiologia, porém foram importantes na compreensão de termos para nortear o que será desenvolvido nas próximas etapas.

Na esteira de se compreender as causas das lesões, o artigo “*The relationship*

*between training load and injury, illness and soreness: a systematic and literature review*” (Drew and Finch 2016) traz importantes conceitos para discussão entre intensidades de cargas na qual os atletas são submetidos e os impactos que elas geram. Já o artigo *“Injury prevention strategies at the FIFA 2014 World Cup: perceptions and practices of the physicians from the 32 participating national teams”* (McCall et al. 2015), que possui um enfoque nos métodos usados pelas equipes participantes para que seus jogadores evitassem tal problema em uma competição que usualmente dura trinta dias. Já (Hägglund et al. 2013b) utiliza um estudo de uma década de análise de lesões em território europeu, concluindo-se que a maior incidência de machucados se dá na perna que o atleta usualmente utiliza para tomar as ações no campo, além das zonas musculares que comumente os atletas possuem problemas físicos.

A discussão levantada por (Ekstrand 2013) traz um importante ponto em termos de lesões que estão fora do escopo desse trabalho, mas que também devem ser considerados fundamentalmente quando se diz respeito a lesões, que é a saúde do atleta. Ainda nessa publicação, foi estimado que o futebol possui 1000 vezes mais lesões do que operações consideradas de alto risco, como ocorre com trabalho em operações industriais.

Outros trabalhos analisados que caminham nessa direção são (Carey et al. 2017) e (Gabbett et al. 2012) (um estudo aplicado ao rugby), além do consenso sobre lesões no esporte (Dujon et al. 2006), já citado anteriormente.

### **3.3. Modelos Preditivos**

Por fim, a tese de doutorado de (Rossi 2017) possui um foco na criação de um modelo preditivo de lesões, permitindo com que os clubes tenham um maior controle sobre a iminência desses fatos ocorrerem e, dessa maneira poderem aplicar as técnicas de prevenção já citadas anteriormente em outras publicações que foram consultadas nesse trabalho.

## **4. Implementação**

O planejamento desta etapa foi realizado com foco na implementação do projeto definitivo. A seleção dos artigos foi feita por meio de busca em sites especializados, resultando na escolha do trio de artigos mencionados anteriormente. O planejamento foi estruturado da seguinte forma:

A metodologia será baseada no artigo (Eliakim et al. 2020), adaptada para o contexto do futebol brasileiro, abrangendo a temporadas de 2024.

A abordagem original do artigo de referência consiste nos seguintes passos:

- Ordenar os clubes participantes do torneio pelos seus valores de mercado, conforme dados do site *transfermarkt.com*.
- Verificar a classificação final do campeonato, comparando a classificação esperada (i.e., a posição do time no ranking de valores de mercado) com a classificação efetivamente obtida.
- Comparar com o time que obteve a classificação esperada para ele, determinando a pontuação teórica esperada.
- Classificar as métricas de dias perdidos por lesão, juntamente com a quantidade de posições e pontos que os clubes deixaram de conquistar, teoricamente, devido às lesões.

Os torneios analisados serão o Campeonato Brasileiro Série A.

No caso do Campeonato Brasileiro, que adota o formato de liga (i.e., pontos corridos), a análise será feita diretamente com base na classificação final ao término de cada campeonato.

#### 4.1. Coleta de Dados

A coleta de dados será realizada a partir de duas fontes principais: o site *transfermarkt.com* (via *scraping*) e a API *data-football.org*.

Embora seja tecnicamente possível realizar toda a coleta de dados pelo *transfermarkt.com*, para otimizar o desempenho e reduzir a carga sobre o processo de *scraping*, foi decidido que as classificações dos campeonatos serão obtidas diretamente pela API. O *scraping* será utilizado exclusivamente para coletar informações relacionadas às lesões dos jogadores de cada clube, bem como seu valor de mercado e informações sobre os atletas que passaram em cada um deles.

Todo o trabalho de coleta de dados foi feito utilizando a linguagem Python e bibliotecas pré existentes. Para o *scraping* dos dados vindos do site Transfermarkt, foi utilizada a biblioteca BeautifulSoup4 (The BeautifulSoup development team 2025), que é um conjunto de ferramentas pensado em facilitar a obtenção de dados vindos de sites da internet, fazendo um parse do *HTML* do código fonte da página, permitindo com que se obtenha partes específicas dele, como no caso específico deste trabalho, as tabelas.

##### 4.1.1. Valor de Mercado

Os dados sobre o valor de mercado dos clubes foram obtidos por meio de web *scraping* no site Transfermarkt [1] (Tra 2024). O processo envolveu o parsing do código-fonte *HTML* da página, possibilitando a extração das informações referentes aos clubes e seus respectivos valores de mercado. O código utilizado para a coleta está disponível no Apêndice A.

As informações extraídas foram armazenadas em um arquivo no formato *CSV*, contendo três colunas: nome do time, valor de mercado do elenco e ranking desse valor de mercado dentro do campeonato. O ranking foi atribuído de forma que o clube com o maior valor de mercado agregado ocupa a posição 1, o segundo maior a posição 2 e assim sucessivamente, até o clube com o menor valor de mercado, que ocupa a posição 20.

##### 4.1.2. Jogadores

A coleta dos dados referentes aos jogadores foi realizada de forma semelhante à coleta dos valores de mercado, porém acessando as páginas de elenco de cada clube participante do Campeonato Brasileiro de 2024.

As informações extraídas foram armazenadas em um arquivo no formato *CSV*, contendo as seguintes colunas:

- **Clube:** Nome do clube ao qual o jogador pertence.
- **Nome:** Nome completo do atleta.

- **URL\_Jogador:** Nome do atleta no formato utilizado pela URL do Transfermarkt, visando facilitar consultas futuras, como a obtenção de dados de lesões.
- **ID\_Jogador:** Identificador numérico único do jogador no Transfermarkt.
- **Posição:** Posição em que o atleta atua.
- **Valor de Mercado:** Valor de mercado individual do jogador.
- **Link\_Completo:** URL completa da página do jogador no Transfermarkt.

O código responsável por essa etapa da coleta está disponível no Apêndice B.

### 4.1.3. Lesões

Os dados relativos as lesões de jogadores foram obtidos por meio da página de cada jogador no site Transfermarkt, na qual são mostrados via tabela. Da parte de extração, foi a que mais levou tempo para ser totalmente concluída, por conta de se ter a necessidade de fazer o acesso a cada página de atleta individualmente, havendo a necessidade de um delay entre cada obtenção de cada jogador de cerca de um minuto, para que não houvessem bloqueios do site fonte. Ao todo, foram 1096 jogadores analisados, e 938 lesões coletadas. Portanto, apenas para coleta foram gastas cerca de 46 horas de código efetivamente rodando.

Após a extração feita, os arquivos foram disponibilizados em formato *CSV*, com as seguintes colunas:

- **Season:** Temporada na qual ocorreu. Importante ressaltar que é a temporada do calendário europeu, ou seja, não foi considerada para as análises.
- **Injury:** Tipo de lesão ocorrida.
- **From:** Data na qual a lesão ocorreu.
- **Until:** Data na qual ao jogador retornou da lesão.
- **Days Out:** Quantidade de dias que o jogador ficou sem atuar.
- **Games Missed:** Número de jogos perdidos por aquela lesão.
- **Jogador url:** Nome do jogador no formato presente da url do site.
- **Jogador:** Nome do jogador.
- **ID:** ID do jogador relativo ao site transfermarkt
- **Ano Lesao:** Ano na qual aquela lesão ocorreu

## 4.2. Análise

A análise a ser realizada neste trabalho será adaptada da metodologia utilizada no estudo "Estimation of injury costs: financial damage of English Premier League teams' underachievement due to injuries" (Eliakim et al. 2020). O foco será entender se o impacto das lesões no tem relação com o desempenho das equipes participantes do Campeonato Brasileiro Série A. Não será realizado o cálculo de perdas financeiras.

A análise foi focada na correlação de variáveis relevantes para o contexto, sendo utilizado sobretudo o índice de correlação de Pearson, que varia de -1 até 1 e basicamente diz o quão relacionados estão os dados analisados. Quanto mais próximo aos extremos (-1 ou 1), maior é o nível de correlação.

## 5. Metodologia

A implementação do estudo consistirá na aplicação prática da metodologia descrita, envolvendo a coleta, organização, análise e visualização dos dados.

A metodologia consiste no artigo "Estimation of injury costs: financial damage of English Premier League teams' underachievement due to injuries"(Eliakim et al. 2020), que consiste em ordenar os times por seu respectivo valor de mercado de acordo com o site *transfermarkt.com*, e considerando essa métrica como sendo a posição esperada de chegada do time ao final da temporada, e a partir disso considerar a diferença de performance (*Underperformance*) e assim compará-las ao índice de lesões de cada um dos clubes, através de modelos de regressão que calculem a correlação entre variáveis. No caso desse trabalho, serão aplicadas as correlações de Pearson e regressões múltiplas.

Embora a proposta inicial previsse a análise de quatro temporadas, durante o desenvolvimento do trabalho identificou-se que limitações da qualidade dos dados e, portanto, a viabilidade de análise tornaram esse escopo inviável. Assim, optou-se por restringir o estudo à temporada de 2024, garantindo maior profundidade e consistência na análise.

### 5.1. Pipeline de Coleta e Tratamento de Dados

A implementação iniciará com a construção de um *pipeline* automatizado para a coleta e o processamento dos dados. Esse *pipeline* será composto pelos seguintes passos:

1. **Coleta de Dados via Web Scraping:** Foi desenvolvido um *script* em *Python* utilizando a biblioteca *BeautifulSoup* para realizar o *scraping* do site *transfermarkt.com* para fins de teste, buscando se identificar os padrões das páginas do *transfermarkt.com* e peculiaridades da biblioteca para a coleta. O *script* extrai informações sobre os valores de mercado dos jogadores, histórico de lesões e dias perdidos por equipe. Sendo gerado um dataframe da biblioteca *pandas*, para se armazená-los.
2. **Integração com API:** A API *data-football.org* foi utilizada para obter os resultados e classificações dos campeonatos analisados. As requisições foram feitas utilizando a biblioteca *requests*, armazenando os dados em formato *JSON* e convertendo-os para tabelas *pandas*.
3. **Processamento de Dados:** Os dados coletados foram tratados para remover inconsistências, duplicações e valores ausentes. Variáveis como dias de lesão foram normalizadas em relação ao total de partidas disputadas na temporada para permitir comparações justas entre equipes. Além disso, os dados de lesão contemplam todas as lesões que o atleta teve em sua carreira. Para o trabalho, foram filtradas as lesões ocorridas apenas em 2024, para se englobar no escopo feito.

### 5.2. Cálculo das Métricas de Desempenho

Após o tratamento dos dados, as métricas de desempenho serão calculadas utilizando a biblioteca *scipy*:

- **Classificação esperada:** Baseada no valor total de mercado dos clubes, as equipes foram ordenadas para gerar a posição esperada em cada competição.
- **Diferença de Classificação:** Foi calculada como a diferença entre a posição final real e a posição esperada.

- **Impacto das Lesões:** Estimado a partir da correlação entre o número de dias perdidos por lesão e a diferença de classificação ou pontos perdidos.

Para se calcular as métricas de impacto de lesões no que diz respeito ao desempenho dos clubes, será usada as funções de cálculo de Pearson oriundas da biblioteca `scipy` (The Scikit-learn development team 2025), se passando as métricas desejadas.

Além disso, serão gerados gráficos para uma melhor visualização dos valores obtidos nas análises utilizando as bibliotecas `seaborn` (The Seaborn development team 2025) e `matplotlib` (The Matplotlib development team 2025), ambas já pré feitas na linguagem python.

### 5.3. Ferramentas e Tecnologias Utilizadas

A implementação foi realizada utilizando ferramentas modernas para garantir eficiência e precisão:

- *Python 3.10*: Linguagem principal para o desenvolvimento de *scripts* e análise de dados.
- *pandas* e *numpy*: Para manipulação de dados e cálculos numéricos.
- *BeautifulSoup* e *requests*: Para realizar o *scraping* e interagir com APIs.
- *Matplotlib* e *Seaborn*: Para criar visualizações gráficas dos resultados.
- *SciPy*: Para o cálculo do índice de Pearson para as correlações entre as variáveis
- *Jupyter Notebook*: Para prototipação rápida e documentação do processo de análise.

### 5.4. Desafios Técnicos e Soluções Adotadas

Durante a implementação, foram enfrentados alguns desafios técnicos:

- **Limitações no Web Scraping:** O site *transfermarkt.com* impõe restrições ao volume de requisições. Para contornar esse problema, foi implementada uma estratégia de controle de tempo entre as requisições (*rate-limiting*). Para o trabalho, foram testados diferentes tempos entre as requisições, sendo o que melhor desempenho resultados foram saltos aleatórios entre 35 a 60 segundos.
- **Compreensão do Scraping:** Como serão feitas várias requisições de diferentes páginas, uma vez que serão obtidos os clubes que jogam aquele torneio, após isso obter o *link* que leva ao elenco, acessar as páginas de cada um jogador na aba de lesões e, por fim consolidá-las. Além disso, terá de ser feito o cálculo do valor de mercado do clube por temporada. Portanto, a compreensão do funcionamento das *urls* do site se torna imprescindível.
- **Integração entre Dados Heterogêneos:** Para integrar os dados de lesões, valores de mercado e resultados das partidas, foram utilizadas *keys* comuns, como o nome das equipes e a temporada. Houve alguns clubes com diferenças nos nomes entre a API e o site Transfermarkt (?). Por conta disso, foi necessário um ajuste manual para esses casos, conforme o exemplo abaixo (o restante do código relacionado às análises encontra-se no Apêndice C):

**Listing 1. Ajuste manual de nomes dos clubes**

```
renames = {
    "SC_Corinthians_Paulista": "SC_Corinthians",
```

```

"CA_Mineiro": "Atletico_Mineiro",
"CA_Paranaense": "Athletico_Paranaense",
"AC_Goianiense": "Atletico_Goianiense"
}

```

- **Volume de dados:** Conforme já citado, ao todo foram coletadas 938 lesões de 1096 atletas da série A do campeonato brasileiro, sendo necessário tempo entre as requisições para que não houvesse problemas com o site fonte. Dessa forma, a coleta das informações de lesões levou um tempo muito maior que o planejado, causando um problema no andamento do cronograma previsto, mas sendo possível se terminar em tempo o que foi proposto.

## 6. Resultados

A análise estatística buscou identificar o impacto das lesões no desempenho dos clubes durante o Campeonato Brasileiro de 2024. Foram realizadas análises descritivas, correlações de Pearson, testes de comparação de grupos (t de Student e Mann-Whitney) e uma regressão linear múltipla.

Após feitas as análises, não foi encontrada uma correlação significativa entre as métricas de lesões e os índices de desempenho dos clubes no campeonato analisado, com as métricas em geral ficando entre os níveis baixo e moderado. Os índices de correlação foram os seguintes:

<b>Variáveis</b>	<b>Coefficiente (<i>r</i>)</b>	<b><i>p</i>-valor</b>
Dias Totais Perdidos × Pontuação	0,330	0,156
Dias Totais Perdidos × Diferença de Posições	-0,272	0,246
Quantidade de Lesões × Pontuação	0,345	0,136
Quantidade de Lesões × Diferença de Posições	-0,121	0,612
Dias Totais Perdidos × Gols Marcados	0,423	0,063
Dias Totais Perdidos × Gols Sofridos	-0,184	0,437
Dias Totais Perdidos × Saldo de Gols	0,350	0,130
Dias Totais Perdidos × Vitórias	0,294	0,209
Dias Totais Perdidos × Derrotas	-0,387	0,092
Quantidade de Lesões × Saldo de Gols	0,385	0,094
Quantidade de Lesões × Derrotas	-0,424	0,062

A matriz de correlação também mostra isso de maneira mais clara e com outras variáveis:

### 6.1. Correlação entre Lesões e Indicadores de Desempenho

As correlações de Pearson indicaram que o número de dias perdidos por lesão apresentou uma correlação positiva moderada com o número de gols marcados ( $r = 0,423$ ,  $p = 0,063$ ) e com o saldo de gols ( $r = 0,350$ ,  $p = 0,130$ ). Além disso, foi observada uma correlação negativa moderada entre os dias perdidos e o número de derrotas ( $r = -0,387$ ,  $p = 0,092$ ). No entanto, essas correlações não atingiram significância estatística.

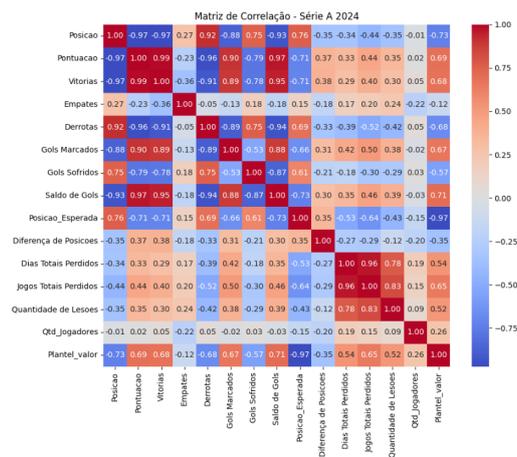


Figure 1. Matriz de correlação das variáveis utilizadas.

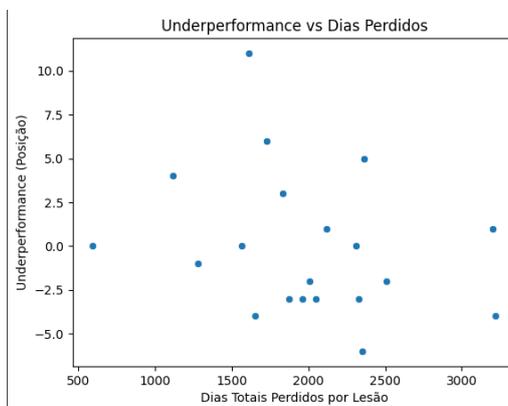


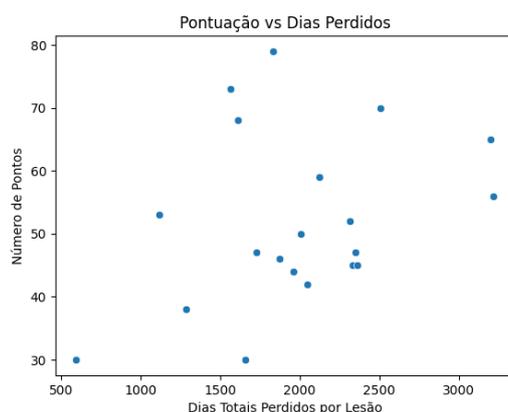
Figure 2. Distribuição da underperformance por dias totais fora por lesão.

É possível se notar uma certa população de dados nos setores mais abaixo e à direita, que indicam que é possível haver uma tendência, contudo, ela não é forte o suficiente estatisticamente. É possível se notar que o gráfico tende a, quanto mais a direita (ou seja, maior a quantidade de dias que os atletas passaram lesionados), maior a chance de haver perdido posições em relação ao que era esperado a equipe obter, apesar de não ser algo muito forte e estatisticamente relevante.

Em termos de pontuação, as lesões se mostram com uma maior correlação estatística, com quatro das cinco equipes com maior pontuação estando no grupo de baixa lesão. Contudo, há um equilíbrio também em relação a pontuação e lesões. Dessa forma, há uma zona mais "populosa" no gráfico na região central.

Quando analisada a quantidade total de lesões por equipe, observou-se uma correlação negativa com o número de derrotas ( $r = -0,424$ ,  $p = 0,062$ ) e uma correlação positiva com o saldo de gols ( $r = 0,385$ ,  $p = 0,094$ ), ambas próximas ao limiar de significância.

Variáveis de controle como o valor de mercado do plantel (*Plantel\_valor*) mostraram correlações fortes e significativas com desempenho ofensivo e defensivo, como gols marcados ( $r = 0,667$ ,  $p = 0,001$ ), saldo de gols ( $r = 0,705$ ,  $p = 0,001$ ) e



**Figure 3. Distribuição da pontuação das equipes por dias totais fora por lesão.**

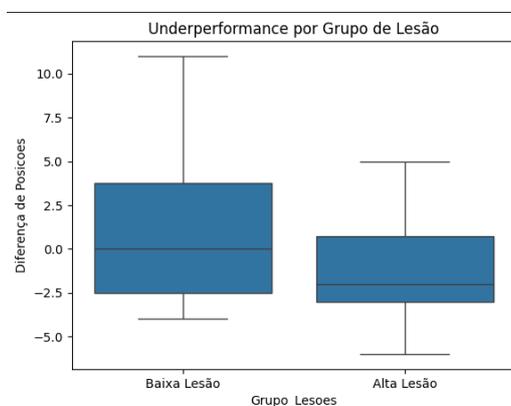
número de derrotas ( $r = -0,676$ ,  $p = 0,001$ ). Além disso, o valor do plantel também apresentou correlação positiva com o total de dias perdidos por lesão ( $r = 0,537$ ,  $p = 0,015$ ) e com a quantidade de lesões ( $r = 0,520$ ,  $p = 0,019$ ), sugerindo que equipes com elencos mais valiosos também são mais expostas a perdas por lesão, possivelmente por maior número de jogos ou intensidade competitiva.

Não foram encontradas correlações significativas entre a quantidade de jogadores registrados por clube e as demais variáveis analisadas.

## 6.2. Comparação entre Grupos de Alta e Baixa Lesão

Para verificar se o nível de lesões impactou a underperformance (medida como a diferença entre posição esperada pelo valor de mercado e a posição real), os clubes foram divididos em dois grupos: alta lesão e baixa lesão, com base na mediana de dias totais perdidos.

O teste t de Student não indicou diferença estatisticamente significativa entre os grupos ( $t = -1,453$ ,  $p = 0,166$ ). Confirmando esse resultado, o teste de Mann-Whitney também não apontou diferenças significativas ( $U = 35,500$ ,  $p = 0,287$ ). O boxplot da Fig. 4 ilustra essa comparação.



**Figure 4. Distribuição da underperformance por grupo de lesão.**

### 6.3. Modelo de Regressão Múltipla

Foi realizada uma regressão linear múltipla para avaliar se o total de dias perdidos, quantidade de lesões, valor do plantel e número de jogadores poderiam prever a underperformance. O modelo geral não foi significativo ( $R^2 = 0,178$ ,  $p = 0,537$ ), indicando que essas variáveis, em conjunto, não explicaram de forma significativa a variação na underperformance dos clubes.

Nenhuma das variáveis independentes apresentou coeficientes estatisticamente significativos isoladamente ( $p > 0,05$  para todas). Dessa forma, independentemente da técnica usada para se medir a relação, o índice não se mostra forte em alguma delas.

## 7. Conclusão

De forma geral, não foi encontrada uma associação estatisticamente robusta entre lesões e underperformance no Campeonato Brasileiro de 2024. Os resultados sugerem que o valor de mercado do plantel possui maior poder explicativo sobre o desempenho em campo. Ainda assim, algumas correlações entre lesões e indicadores como gols marcados e derrotas apontam tendências que merecem investigação em amostras maiores ou em outros contextos competitivos.

Evidentemente, o futebol é um esporte de alta complexidade com inúmeras variáveis que podem impactar no desempenho final de uma equipe, como trocas de treinador (Lago-Peñas 2011), investimento em geral (conforme foi encontrado nesse estudo, mas também sendo objeto de estudo, como em (Khalifaoui et al. 2024)). Dessa maneira, as lesões podem representar um ponto importante para ser analisado conjuntamente com outros desses pontos, a fim de se mensurar a importância de cada um deles. Esse estudo mais detalhado pode vir a ser feito em outra etapa de pesquisa.

### A. Script de Extração de Lesões dos Jogadores

O código a seguir foi utilizado para coletar, via web scraping no site Transfermarkt, os dados de lesões dos jogadores da Série A do Campeonato Brasileiro.

```
df_players = pd.DataFrame()
years = [2021, 2024]

df_lesoes_anterior = pd.DataFrame()
files = os.listdir("lesoes")
for file in files:
    a = pd.read_csv(f"lesoes/{file}", sep=",", encoding='utf-8-sig')
    df_lesoes_anterior = pd.concat([df_lesoes_anterior, a])

qtd_jogadores_ant = len(df_lesoes_anterior[["ID"]].drop_duplicates())

for year in years:
    tmp = pd.read_csv(f"data/jogadores_serie_a_{year}.csv")
    print(f"Arquivo_de_{year}_lido!")
    df_players = pd.concat([tmp, df_players])
```

```

print(f"Tamanho_total:_{len(df_players)}")
print("-----")
df_players = df_players[["Nome", "URL_Jogador", "ID_Jogador"]].
    drop_duplicates().reset_index()
print(f"Jogadores_distintos:_{len(df_players)}")
print(f"Jogadores_ja_processados:_{qtd_jogadores_ant}")
print(f"Falta_processar:_{len(df_players)}_{qtd_jogadores_ant}")
print("-----")

df_lesoes = pd.DataFrame()

ids_to_process = [id for id in df_players["ID_Jogador"]
                  if len(df_lesoes_anterior[df_lesoes_anterior["ID"] == id]) == 0]

with tqdm(total=len(ids_to_process), desc="Processando_jogadores
") as pbar:
    i = 0
    for id in df_players["ID_Jogador"]:
        if len(df_lesoes_anterior[df_lesoes_anterior["ID"] == id
        ]) > 0:
            continue

        nome = df_players[df_players["ID_Jogador"] == id][
            "URL_Jogador"].item()

        url = f"https://www.transfermarkt.com/{nome}/
            verletzungen/spieler/{id}"
        headers = {"User-Agent": "Mozilla/5.0_(Windows_NT_10.0;_
            Win64;_x64)"}
        page = requests.get(url, headers=headers)

        if page.status_code == 200:
            soup = BeautifulSoup(page.content, 'html.parser')
            injury_table = soup.find('table', {'class': 'items'
            })

            injuries = []
            if injury_table:
                rows = injury_table.find_all('tr')[1:]
                for row in rows:
                    cols = row.find_all('td')
                    season = cols[0].get_text(strip=True)
                    injury = cols[1].get_text(strip=True)
                    from_date = cols[2].get_text(strip=True)
                    until_date = cols[3].get_text(strip=True)
                    days_out = cols[4].get_text(strip=True)
                    games_missed = cols[5].get_text(strip=True)

```

```

        injuries.append({
            'Season': season,
            'Injury': injury,
            'From': from_date,
            'Until': until_date,
            'Days_Out': days_out,
            'Games_Missed': games_missed
        })
    tmp = pd.DataFrame(injuries)
    tmp["Jogador_url"] = nome
    tmp["Jogador"] = df_players[df_players["
        ID_Jogador"] == id]["Nome"].item()
    tmp["ID"] = id
    df_lesoes = pd.concat([tmp, df_lesoes])

    time.sleep(random.randint(25, 60))
    i += 1
else:
    print(f"ERRO: _status_{page.status_code}")
    break
pbar.update(1)
pbar.set_postfix({'Processados': i, 'Restantes': len(
    ids_to_process) - i})

```

## B. Script de Extração do Valor de Mercado dos Clubes

```

df_final = pd.DataFrame(columns=["Time", "Valor_de_Mercado", "
    Plantel", "Ano"])

headers = {
    "User-Agent": "Mozilla/5.0_(Windows_NT_10.0;_Win64;_x64)"
}

for ano in range(2016, 2017):
    df_ano = pd.DataFrame(columns=["Time", "Valor_de_Mercado", "
        Plantel"])
    url = f"https://www.transfermarkt.com.br/campeonato-
        brasileiro-serie-a/startseite/wettbewerb/BRA1/plus/?
        saison_id={ano}"
    response = requests.get(url, headers=headers)

    if response.status_code == 200:
        soup = BeautifulSoup(response.content, "html.parser")
        table = soup.find("table", {"class": "items"})

        if table:
            rows = table.find_all("tr", {"class": ["odd", "even"
                ]})
            for row in rows:

```

```

team = row.find("td", {"class": "hauptlink"}).
    get_text(strip=True)
cells = row.find_all("td")
market_value = cells[-2].get_text(strip=True)
squad_size = cells[-1].get_text(strip=True)

df_aux = pd.DataFrame(data={
    'Time': [team],
    'Valor_de_Mercado': [market_value],
    'Plantel': [squad_size]
})
df_ano = pd.concat([df_ano, df_aux],
    ignore_index=True)
df_ano['Ano'] = ano + 1
df_final = pd.concat([df_final, df_ano],
    ignore_index=True)

    print(f"Dados_de_{ano}_coletados_com_sucesso!")
else:
    print(f"Tabela_n o_encontrada_para_o_ano_{ano}.")
else:
    print(f"Erro_ao_acessar_a_p gina_do_ano_{ano}:_{
        response.status_code}")

df_final.to_csv("valor_mercado_ano.csv", sep=',', encoding='utf
-8-sig', index=False)

```

### C. Script de Extração dos Dados dos Jogadores

```

def get_serie_a_clubs(year):
    """Obt m os clubes da S rie A no ano especificado"""
    url = f"https://www.transfermarkt.com.br/campeonato-
        brasileiro-serie-a/startseite/wettbewerb/BRA1/plus/?
        saison_id={year-1}"
    headers = {"User-Agent": "Mozilla/5.0_(Windows_NT_10.0;_
        Win64;_x64)"}

    clubs = []
    response = requests.get(url, headers=headers)

    if response.status_code == 200:
        soup = BeautifulSoup(response.content, 'html.parser')
        table = soup.find('table', {'class': 'items'})

        if table:
            rows = table.find_all('tr', {'class': ['odd', 'even'
                ]})
            for row in rows:
                club_link = row.find('td', {'class': 'hauptlink'
                    }).find('a')

```

```

        if club_link:
            club_url = club_link.get('href')
            club_id = club_url.split('/')[4]
            club_name = club_url.split('/')[1]
            clubs.append({'id': club_id, 'url_name':
                club_name, 'full_name': club_link.
                get_text(strip=True)})

    return clubs

def get_players_data(club_id, club_url_name, club_full_name,
year):
    """Coleta dados dos jogadores de um time para o ano
    informado"""
    headers = {"User-Agent": "Mozilla/5.0_(Windows_NT_10.0;_
        Win64;_x64)"}

    season_id = year - 1
    url = f"https://www.transfermarkt.com.br/{club_url_name}/
        startseite/verein/{club_id}/saison_id/{season_id}"

    players_data = []

    try:
        response = requests.get(url, headers=headers)
        response.raise_for_status()

        soup = BeautifulSoup(response.content, 'html.parser')
        table = soup.find('table', {'class': 'items'})

        if table:
            rows = table.find_all('tr', {'class': ['odd', 'even'
                ]})

            for row in rows:
                player_link = row.find('td', {'class': '
                    hauptlink'}).find('a')
                if not player_link:
                    continue

                player_url = player_link.get('href')
                player_url_name = player_url.split('/')[1]
                player_id = player_url.split('/')[4]
                name = player_link.get_text(strip=True)
                position = row.find('td', {'class': lambda x: x
                    and 'rueckennummer_bg_' in x})
                position = position.get('title').strip() if
                position else 'N/D'

```

```

        market_value = row.find('td', {'class': 'rechts_
            hauptlink'}).get_text(strip=True)

        players_data.append({
            'Ano': year,
            'Clube': club_full_name,
            'Nome': name,
            'URL_Jogador': player_url_name,
            'ID_Jogador': player_id,
            'Posição': position,
            'Valor_de_Mercado': market_value,
            'Link_Completo': f"https://www.transfermarkt
                .com.br{player_url}"
        })

        print(f"Dados do_{club_full_name}_em_{year}_
            coletados com sucesso!")
    else:
        print(f"Tabela não encontrada para o_{
            club_full_name}_em_{year}")

    time.sleep(random.randint(30, 60))

except Exception as e:
    print(f"Erro ao coletar dados do_{club_full_name}_em_{
        year}:_{str(e)}")

return players_data

year = 2024
clubs = get_serie_a_clubs(year)
all_players = []

for club in clubs:
    players = get_players_data(club['id'], club['url_name'],
        club['full_name'], year)
    all_players.extend(players)
    time.sleep(random.randint(2, 5))

players_df = pd.DataFrame(all_players)

```

## D. Script usado para análise

```

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import pearsonr, ttest_ind, mannwhitneyu
import statsmodels.api as sm

```

```

import requests

lesoes = pd.read_csv("lesoes/
    lesoes_jogadores_serie_a_consolidado.csv", encoding='utf-8-
    sig')
jogadores = pd.read_csv("jogadores/jogadores_serie_a_consolidado
    .csv", encoding='utf-8-sig')

lesoes["From"] = pd.to_datetime(lesoes["From"])
lesoes["Until"] = pd.to_datetime(lesoes["Until"])
lesoes['Ano_Lesao'] = lesoes['From'].dt.year

jogador_anos = jogadores[['ID_Jogador', 'Ano']].drop_duplicates
    ()

cons = lesoes.merge(
    jogador_anos,
    left_on=['ID', 'Ano_Lesao'],
    right_on=['ID_Jogador', 'Ano'],
    how='inner'
).drop(columns=['ID_Jogador', 'Ano'])

final = cons.merge(jogadores, left_on="ID", right_on="ID_Jogador
    ", how="inner")

tmp = final.loc[final["Ano"] == 2024].copy()
tmp["Games_Missed"] = pd.to_numeric(tmp["Games_Missed"], errors=
    'coerce')
tmp["Days_Out"] = tmp["Days_Out"].str.replace(r"_days", "",
    regex=True).astype(int)

gb_qtd_lesoes = tmp.groupby("Clube")["Games_Missed"].count().
    reset_index(name='Quantidade_de_Lesoes')
df_qtd_jogos_perdidos = tmp.groupby("Clube")["Games_Missed"].sum
    ().reset_index(name='Jogos_Totais_Perdidos')
df_qtd_dias_lesao = tmp.groupby("Clube")["Days_Out"].sum().
    reset_index(name='Dias_Totais_Perdidos')

# Quantidade de jogadores por clube
qtd_jogadores = jogadores[jogadores["Ano"] == 2024].groupby("
    Clube")["ID_Jogador"].count().reset_index(name='Qtd_Jogadores
    ')

valor_mercado = pd.read_csv("valor_mercado_ano.csv", encoding='
    utf-8-sig')
valor_mercado = valor_mercado[valor_mercado["Ano"] == 2024][["
    Time", "Plantel"]]
valor_mercado["Posicao"] = range(1, 21)

```

```

api_key = API_KEY
headers = {'X-Auth-Token': api_key}
url = "https://api.football-data.org/v4/competitions/BSA/standings?season=2024"

response = requests.get(url, headers=headers)
data = response.json()

renames = {
    "SC_Corinthians_Paulista": "SC_Corinthians",
    "CA_Mineiro": "Atletico_Mineiro",
    "CA_Paranaense": "Athletico_Paranaense",
    "AC_Goianiense": "Atletico_Goianiense"
}

classificacao = pd.DataFrame(columns=[
    "Posicao", "Time", "Time_abreviacao", "Time_Sigla",
    "Pontuacao", "Vitorias", "Empates", "Derrotas",
    "Gols_Marcados", "Gols_Sofridos", "Saldo_de_Gols"
])

for item in data["standings"][0]["table"]:
    classificacao.loc[len(classificacao)] = {
        "Posicao": item["position"],
        "Time": item["team"]["name"],
        "Time_abreviacao": item["team"]["shortName"],
        "Time_Sigla": item["team"]["tla"],
        "Pontuacao": item["points"],
        "Vitorias": item["won"],
        "Empates": item["draw"],
        "Derrotas": item["lost"],
        "Gols_Marcados": item["goalsFor"],
        "Gols_Sofridos": item["goalsAgainst"],
        "Saldo_de_Gols": item["goalDifference"]
    }

classificacao["Time"] = classificacao["Time"].replace(renames)

# Constru o dataframe final
df = classificacao.merge(valor_mercado, on="Time", how="inner",
    suffixes=('', '_Esperada'))
df["Diferen a_de_Posicoes"] = df["Posicao_Esperada"] - df["Posicao"]

# Merge com os dados de les es e dias
df = df.merge(df_qtd_dias_lesao.rename(columns={'Clube': 'Time'}), on="Time", how="left")

```

```

df = df.merge(df_qtd_jogos_perdidos.rename(columns={'Clube': 'Time'}), on="Time", how="left")
df = df.merge(gb_qtd_lesoes.rename(columns={'Clube': 'Time'}), on="Time", how="left")
df = df.merge(qtd_jogadores.rename(columns={'Clube': 'Time'}), on="Time", how="left")

# Renomeando as colunas
df = df.rename(columns={
    "Games_Missed_x": "Jogos_Totais_Perdidos",
    "Games_Missed_y": "Quantidade_de_Lesoes",
    "Days_Out": "Dias_Totais_Perdidos"
})

# Ajuste da coluna Plantel_valor
df['Plantel_valor'] = (
    df['Plantel']
    .str.replace('mi.', '', regex=False)
    .str.replace('.', '', regex=False)
    .str.replace(',', '.', regex=False)
    .astype(float)
)

print("\n=== Estatísticas Descritivas ===")
print(df[['Dias_Totais_Perdidos', 'Pontuacao', 'Plantel_valor',
          'Qtd_Jogadores']].describe())

plt.figure(figsize=(10, 8))
sns.heatmap(df.corr(numeric_only=True), annot=True, fmt=".2f",
            cmap="coolwarm")
plt.title("Matriz de Correlação - Série A 2024")
plt.show()

print("\n=== Correlações de Pearson ===")
corr = [
    ('Dias_Totais_Perdidos', 'Pontuacao'),
    ('Dias_Totais_Perdidos', 'Diferença_de_Posicoes'),
    ('Quantidade_de_Lesoes', 'Pontuacao'),
    ('Quantidade_de_Lesoes', 'Diferença_de_Posicoes'),
    ('Dias_Totais_Perdidos', 'Gols_Marcados'),
    ('Dias_Totais_Perdidos', 'Gols_Sofridos'),
    ('Dias_Totais_Perdidos', 'Saldo_de_Gols'),
    ('Dias_Totais_Perdidos', 'Vitorias'),
    ('Dias_Totais_Perdidos', 'Derrotas'),
    ('Quantidade_de_Lesoes', 'Saldo_de_Gols'),
    ('Quantidade_de_Lesoes', 'Derrotas'),
    ('Plantel_valor', 'Gols_Marcados'),

```

```

('Plantel_valor', 'Saldo_de_Gols'),
('Plantel_valor', 'Derrotas'),
('Plantel_valor', 'Quantidade_de_Lesoes'),
('Plantel_valor', 'Dias_Totais_Perdidos'),
('Qtd_Jogadores', 'Dias_Totais_Perdidos'),
('Qtd_Jogadores', 'Quantidade_de_Lesoes'),
('Qtd_Jogadores', 'Plantel_valor'),
('Qtd_Jogadores', 'Pontuacao'),
('Qtd_Jogadores', 'Saldo_de_Gols')
]

for x, y in corr:
    try:
        r, p = pearsonr(df[x], df[y])
        print(f"{x}_x_{y}:_r={r:.3f},_p={p:.3f}")
    except Exception as e:
        print(f"Erro_na_correlacao_{x}_x_{y}:_{e}")

#Graficos
sns.scatterplot(x='Dias_Totais_Perdidos', y='Diferença_de_Posicoes', data=df)
plt.title('Underperformance_vs_Dias_Perdidos')
plt.xlabel('Dias_Totais_Perdidos_por_Lesão')
plt.ylabel('Underperformance_(Posição)')
plt.show()

sns.scatterplot(x='Dias_Totais_Perdidos', y='Pontuacao', data=df)
plt.title('Pontuação_vs_Dias_Perdidos')
plt.xlabel('Dias_Totais_Perdidos_por_Lesão')
plt.ylabel('Número_de_Pontos')
plt.show()

sns.scatterplot(x='Plantel_valor', y='Pontuacao', data=df)
plt.title('Pontuação_vs_Valor_de_Mercado')
plt.xlabel('Valor_do_Plantel_(milhes_de_)')
plt.ylabel('Pontuação_na_Tabela')
plt.show()

sns.scatterplot(x='Plantel_valor', y='Diferença_de_Posicoes', data=df)
plt.title('Diferença_de_Posicoes_vs_Valor_de_Mercado')
plt.xlabel('Valor_do_Plantel_(milhes_de_)')
plt.ylabel('Underperformance_(Posição)')
plt.show()

sns.scatterplot(x='Plantel_valor', y='Dias_Totais_Perdidos', data=df)

```

```

plt.title('Dias_Perdidos_Por_Les o_vs_Valor_de_Mercado')
plt.xlabel('Valor_do_Plantel_(milhes_de_)')
plt.ylabel('Dias_Totais_Perdidos_Por_Les o')
plt.show()

mediana_lesoes = df['Dias_Totais_Perdidos'].median()
df['Grupo_Lesoes'] = np.where(df['Dias_Totais_Perdidos'] >
    mediana_lesoes, 'Alta_Les o', 'Baixa_Les o')

sns.boxplot(x='Grupo_Lesoes', y='Diferen a_de_Posicoes', data=
    df)
plt.title('Underperformance_por_Grupo_de_Les o')
plt.show()

grupo_alta = df[df['Grupo_Lesoes'] == 'Alta_Les o']['Diferen a
    _de_Posicoes']
grupo_baixa = df[df['Grupo_Lesoes'] == 'Baixa_Les o']['
    Diferen a_de_Posicoes']

t_stat, p_val_ttest = ttest_ind(grupo_alta, grupo_baixa,
    equal_var=False)
print("\n===_Teste_t:_Underperformance_por_Grupo_de_Les o_===")
print(f"t={t_stat:.3f},_p={p_val_ttest:.3f}")

u_stat, p_val_mw = mannwhitneyu(grupo_alta, grupo_baixa,
    alternative='two-sided')
print("\n===_Teste_Mann-Whitney:_Underperformance_por_Grupo_de_
    Les o_===")
print(f"U={u_stat:.3f},_p={p_val_mw:.3f}")

print("\n===_Regress o_M ltipla:_Preditores_de_
    Underperformance_===")
X = df[['Dias_Totais_Perdidos', 'Quantidade_de_Lesoes', '
    Plantel_valor', 'Qtd_Jogadores']].fillna(0)
X = sm.add_constant(X)
y = df['Diferen a_de_Posicoes']

modelo = sm.OLS(y, X).fit()
print(modelo.summary())

```

## References

- [Tra 2024] (2024). Campeonato brasileiro série a 2024 - transfermarkt.
- [Aragão et al. 2023] Aragão, P. H. L., da Veiga, I. G., Costa, N. N., Reis, R. d. N. R., Sobrero, M. d. S., and Hanna, L. O. M. (2023). Impactos psicológicos em atletas profissionais que sofreram lesões esportivas: uma revisão integrativa. *Brazilian Journal of Health Review*, 6(3):9171–9181.
- [Carey et al. 2017] Carey, D. L. et al. (2017). Training loads and injury risk in australian

- football—differing acute: chronic workload ratios influence match injury risk. *British journal of sports medicine*, 51(16):1215–1220.
- [Confederação Brasileira de Futebol 2024] Confederação Brasileira de Futebol (2024). Tabela do campeonato brasileiro série a 2024.
- [Drew and Finch 2016] Drew, M. K. and Finch, C. F. (2016). The relationship between training load and injury, illness and soreness: a systematic and literature review. *Sports medicine*, 46:861–883.
- [Dujon et al. 2006] Dujon, D., Fuller, C. W., Moller, H. J., Ostry, A., and Monroe, R. (2006). Consensus statement on injury definitions and data collection procedures in studies of football (soccer) injuries. *British Journal of Sports Medicine*, 40(3):193–201.
- [Eirale et al. 2013] Eirale, C. et al. (2013). Low injury rate strongly correlates with team success in qatari professional football. *British journal of sports medicine*, 47(12):807–808.
- [Ekstrand 2013] Ekstrand, J. (2013). Keeping your top players on the pitch: the key to football medicine at a professional level.
- [Eliakim et al. 2020] Eliakim, E. et al. (2020). Estimation of injury costs: financial damage of english premier league teams’ underachievement due to injuries. *BMJ Open Sport Exercise Medicine*, 6(1):e000675.
- [Folha de S.Paulo 2022] Folha de S.Paulo (2022). Times brasileiros jogam 20 partidas a mais por ano que europeus. Acessado em: 26 jan. 2025.
- [Gabbett et al. 2012] Gabbett, T. J., Ullah, S., and Finch, C. F. (2012). Identifying risk factors for contact injury in professional rugby league players—application of a frailty model for recurrent injury. *Journal of Science and Medicine in Sport*, 15(6):496–504.
- [Hägglund et al. 2013a] Hägglund, M. et al. (2013a). Injuries affect team performance negatively in professional football: an 11-year follow-up of the uefa champions league injury study. *British journal of sports medicine*, 47(12):738–742.
- [Hägglund et al. 2013b] Hägglund, M., Waldén, M., and Ekstrand, J. (2013b). Risk factors for lower extremity muscle injury in professional soccer: the uefa injury study. *The American journal of sports medicine*, 41(2):327–335.
- [Khalfaoui et al. 2024] Khalfaoui, R., Djeddi, N., Martinez, L. M. R., Ftiti, Z., and Euch, J. (2024). European football club market value and sporting performance: the moderating effect of head coach changes. *Management Research Review*, 47(3):563–585.
- [Lago-Peñas 2011] Lago-Peñas, C. (2011). Coach mid-season replacement and team performance in professional soccer. *Journal of Human Kinetics*, 28:115–122.
- [McCall et al. 2015] McCall, A. et al. (2015). Injury prevention strategies at the fifa 2014 world cup: perceptions and practices of the physicians from the 32 participating national teams. *British journal of sports medicine*, 49(9):603–608.
- [Qatar Football Association 2025] Qatar Football Association (2025). Stars league. Acessado em: 26 jan. 2025.
- [Rossi 2017] Rossi, A. (2017). Predictive models in sport science: multi-dimensional analysis of football training and injury prediction.
- [The BeautifulSoup development team 2025] The BeautifulSoup development team (2025). BeautifulSoup documentation. Acesso em: 22 jun. 2025.
- [The Matplotlib development team 2025] The Matplotlib development team (2025). Matplotlib: Visualization with python - documentation. Acesso em: 22 jun. 2025.
- [The Scikit-learn development team 2025] The Scikit-learn development team (2025). Scikit-learn: Machine learning in python - documentation. Acesso em: 22 jun. 2025.

[The Seaborn development team 2025] The Seaborn development team (2025). Seaborn: Statistical data visualization - documentation. Acesso em: 22 jun. 2025.

[Warnock 2018] Warnock, R. (2018). The effect of injuries on player and team performance: an empirical analysis of the production function in the national hockey league.