

# Avaliação do Apache Superset para Exposição de Dados Públicos

João Pedro V. Ramalho  
jpvr22@inf.ufpr.br  
Departamento de Informática  
Universidade Federal do Paraná  
Curitiba, Paraná, Brasil

Mateus dos Santos Herbele  
msh22@inf.ufpr.br  
Departamento de Informática  
Universidade Federal do Paraná  
Curitiba, Paraná, Brasil

João A. Silveira  
jas21@inf.ufpr.br  
Departamento de Informática  
Universidade Federal do Paraná  
Curitiba, Paraná, Brasil

Josiney de Souza  
josiney.souza@ifc.edu.br  
Departamento de Informática  
Universidade Federal do Paraná  
Curitiba, Paraná, Brasil

Guilherme A. Derenievicz  
guilherme@inf.ufpr.br  
Departamento de Informática  
Universidade Federal do Paraná  
Curitiba, Paraná, Brasil

Letícia M. Peres  
lmperes@inf.ufpr.br  
Departamento de Informática  
Universidade Federal do Paraná  
Curitiba, Paraná, Brasil

## ABSTRACT

After the implementation of the General Data Protection Law (LGPD), the dissemination of educational data in Brazil began to face new technical and legal requirements, leading to the need for adjustments in public data display platforms. Objective: To evaluate the feasibility of Apache Superset as a tool for disseminating public educational data in Brazil. Methods: The analysis included a heuristic evaluation based on Nielsen's heuristics and accessibility tests (AccessMonitor and Lighthouse) according to WCAG 2.1. Results: The system showed a moderate maturity level, with 17 Major usability problems and compliance only with basic WCAG Level A. A critical regulatory gap regarding the LGPD was identified due to the lack of automatic data anonymization. Conclusion: While technically viable, Superset's deficiencies in usability, accessibility, and data anonymization pose significant barriers for its use in public transparency portals.

## KEYWORDS

Apache Superset, Heurísticas de Nielsen, WCAG, Acessibilidade, Visualização de dados.

## 1 INTRODUÇÃO

Um dos grandes desafios da Interação Humano-Computador (IHC) na década de 2025 está relacionado à forma como as pessoas interagem e se relacionam com dados [1]. Este desafio, formalizado como o Grande Desafio 5 da área de IHC no Brasil, destaca a importância de garantir transparência, legibilidade, negociabilidade e confiança na interação com dados [2]. A perspectiva da Interação Humano-Dados (IHD) reforça que a mera disponibilização de dados não é suficiente para promover o acesso democrático à informação; é necessário assegurar que esses dados sejam legíveis e utilizáveis pelos usuários finais [3].

Nesse contexto, a área da educação pública brasileira apresenta um cenário especialmente relevante para discutir esses desafios. A grande quantidade de dados produzidos por órgãos institucionais, aliada à necessidade de torná-los compreensíveis e acessíveis à sociedade, torna o domínio educacional um campo propício para investigar novas formas de interação humano-dados.

No Brasil, instituições como o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) e o Instituto Brasileiro de Geografia e Estatística (IBGE) são responsáveis por centralizar e

divulgar informações sobre a educação básica e superior, por meio de levantamentos como o Censo Escolar e a Pesquisa Nacional por Amostra de Domicílios (PNAD). Ademais, para facilitar o acesso a esses dados, a Plataforma de Dados Educacionais<sup>1</sup> (PDE), desenvolvida pelo Centro de Computação Científica e Software Livre (C3SL) em parceria com o Laboratório de Dados Educacionais<sup>2</sup>, integra diversas bases em um ambiente unificado desde 2015. A disponibilização de dados públicos, como os disponibilizados pela PDE, é fundamental para a transparência e para subsidiar decisões em políticas públicas. No entanto, apenas tornar os dados acessíveis não garante que eles serão compreendidos ou utilizados de forma eficiente. Essa lacuna evidencia a necessidade de se considerar não apenas a disponibilidade, mas também a usabilidade e legibilidade dos dados, aspectos centrais na interação entre pessoas e informações.

Entretanto, a partir de 2020, com a implementação da Lei Geral de Proteção de Dados (LGPD)[4], surgiram novos desafios técnicos, principalmente devido à redução na granularidade dos dados em comparação aos anos anteriores. Além dessa questão, a PDE apresenta limitações técnicas na exibição das informações, que atualmente se restringem ao formato tabular. Essa limitação dificulta a exploração detalhada dos dados e compromete a experiência do usuário na análise das informações.

Diante desse cenário, a equipe de pesquisadores do C3SL iniciou a busca por aplicações de código aberto que pudessem suprir essas deficiências. Como resultado, considerou-se o uso de ferramentas de Business Intelligence (BI) como uma solução viável, uma vez que permitem uma visualização mais dinâmica e interativa dos dados. Três ferramentas foram testadas: Metabase<sup>3</sup>, Redash<sup>4</sup> e Apache Superset<sup>5</sup>. Entre elas, o Apache Superset demonstrou maior compatibilidade com os princípios de software livre, por ser totalmente gratuito, capaz de lidar com grandes volumes de dados e suportar uma ampla variedade de sistemas de gerenciamento de banco de dados. Dessa forma, o Superset foi a ferramenta escolhida para a implementação.

<sup>1</sup>Disponível em: <https://dadoseducacionais.c3sl.ufpr.br/plataforma/>. Acessado: 03/03/2026.

<sup>2</sup>Disponível em: <https://dadoseducacionais.c3sl.ufpr.br/>. Acessado: 03/02/2026.

<sup>3</sup>Disponível em: <https://www.metabase.com>. Acessado: 03/03/2026.

<sup>4</sup>Disponível em: <https://redash.io>. Acessado: 03/03/2026.

<sup>5</sup>Disponível em: <https://superset.apache.org>. Acessado: 03/03/2026.

Considerando o exposto, este artigo propõe como inovação o reaproveitamento e a adaptação do Apache Superset, uma ferramenta corporativa de código aberto, para o contexto de visualização interativa de dados públicos educacionais. Ao explorar o potencial dessa solução, realizou-se uma avaliação empírica de sua usabilidade e acessibilidade aplicada à Plataforma de Dados Educacionais do C3SL, combinando uma avaliação heurística baseada nas heurísticas de Nielsen com testes automatizados segundo as diretrizes WCAG 2.1. Além de examinar suas funcionalidades técnicas, o estudo identifica e prioriza barreiras de usabilidade, acessibilidade e conformidade regulatória que condicionam a adoção de ferramentas de BI em portais públicos, com destaque para a ausência de mecanismos automáticos de anonimização exigidos pela LGPD. Dessa forma, o trabalho contribui para o debate sobre práticas tecnológicas que favoreçam a transparência, a acessibilidade e a compreensão de dados públicos, oferecendo subsídios técnicos para iniciativas voltadas à democratização do acesso à informação.

As seções seguintes desta introdutória estão organizadas da seguinte forma: a Seção 2 apresenta conceitos e literatura fundamentais que apoiaram o desenvolvimento deste trabalho; a Seção 3 apresenta a metodologia de avaliação utilizada; resultados e discussões da pesquisa são apresentados na Seção 4; a Seção 5 discute os cuidados éticos tomados ao longo do processo de avaliação; por fim a Seção 6 discorre sobre as conclusões finais.

## 2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, apresentamos as abordagens que embasaram a aplicação de ferramentas de BI e a visualização de dados públicos, servindo como base para a seleção do Apache Superset.

### 2.1 Interação humano-dados (IHD)

O conceito de Interação Humano-Dados (IHD) descreve o estudo das interações entre indivíduos e os dados gerados ou coletados sobre eles, enfatizando a transparência, controle e interação ativa com os dados por parte dos usuários finais. Conforme descrito por Mortier et al. [3], a IHD envolve processos colaborativos com dados e o desenvolvimento de ferramentas de comunicação que permitem uma interação ativa, superando a visualização estática. Ademais, o foco da IHD está na entrega de informações personalizadas, contextualizadas e compreensíveis, especialmente em conjuntos de dados extensos e complexos. Nesse estudo, esses conceitos norteadores da IHD são usados como referência para o levantamento dos eixos avaliativos.

### 2.2 Visualização de dados públicos

Nos últimos anos, a visualização de dados públicos têm se destacado como um instrumento para fortalecer a transparência governamental e a participação democrática dos cidadãos [5]. Ao tornar informações mais acessíveis, ela facilita o empoderamento social e estimula o engajamento da população. Esse avanço ocorre em paralelo à crescente adoção de práticas de dados abertos, que buscam garantir não apenas a disponibilidade das informações, mas também sua autenticidade [6].

Em razão disso, a transparência, um dos pilares fundamentais da democracia, é debatida na literatura. Em "O futuro da democracia",

Bobbio [7] destaca que tornar acessíveis as decisões e atos governamentais é essencial para diferenciar regimes democráticos de modelos autoritários. No entanto, a falta de clareza na divulgação de dados pode minar a confiança pública. O Relatório Sistemático sobre Transparência Pública do Tribunal de Contas da União (TCU) alerta que a manipulação ou a apresentação inadequada de informações pode distorcer a percepção da realidade, enfraquecendo o poder social e comprometendo a legitimidade das ações do Estado [8].

Nesse contexto, ferramentas de visualização de dados desempenham um papel estratégico. Atuando como um recurso técnico, elas permitem a compreensão e monitoramento de decisões públicas com maior precisão.

### 2.3 Ferramentas de BI na gestão pública

*Business Intelligence* (BI) é um conjunto de tecnologias e práticas voltadas para a coleta, análise e visualização de dados, permitindo que dados volumosos e mal estruturados sejam transformados em representações compreensíveis [9]. Tradicionalmente adotadas pelo setor privado, as ferramentas de BI têm se expandido para o setor público brasileiro, auxiliando na tomada de decisões estratégicas por meio de painéis e visualizações interativas [10].

A sua adesão no setor público está alinhada ao conceito de *Managerialism*, que propõe a aplicação de práticas e ferramentas vindas do setor privado para otimizar processos e resultados em instituições públicas [11–13]. Embora esse conceito seja promissor, sua implementação requer estudos para adaptar as ferramentas corporativas às especificidades das instituições de gestão pública, garantindo que seus objetivos sociais e estruturais sejam preservados. Essa abordagem alinha-se diretamente à proposta deste estudo, que avalia o uso do Apache Superset como ferramenta de apoio à PDE, cujo objetivo principal é orientar a criação e implementação de políticas públicas.

Assim, compreender os efeitos reais da incorporação de sistemas de gestão nas rotinas públicas é fundamental para avaliar se o *Managerialism* cumpre sua promessa de eficiência sem comprometer os valores públicos. De acordo com Pontes, Pontes e Andrade [11], a utilização de sistemas de informação gerencial no setor público pode trazer benefícios significativos em três principais dimensões: 1) ampliação do conhecimento organizacional; 2) maior precisão na tomada de decisões; e 3) melhorias no desempenho institucional. No entanto, os mesmos pesquisadores reforçam que a simples adoção dessas ferramentas não garante mudanças estruturais. Para eles, os seus impactos são efetivos apenas quando há um processo de implementação que envolva treinamento adequado e engajamento dos usuários.

Por fim, além de evidências teóricas sobre os benefícios dessa abordagem, há casos práticos que atestam a viabilidade dessas ferramentas no Brasil. A Prefeitura de Curitiba, por exemplo, utiliza BI em setores como saúde e educação para monitorar indicadores de desempenho e otimizar a gestão pública [14]. Ainda, a Diretoria Regional de Saúde Norte de Belo Horizonte implementou essa tecnologia no Sistema Único de Saúde (SUS) para aprimorar a gestão de políticas públicas [15]. Esses exemplos reforçam como ferramentas de BI, originalmente desenvolvidas para o ambiente corporativo, podem ser adaptadas às demandas da administração pública.

## 2.4 Apache Superset

As ferramentas *open-source*, ou código aberto, são sistemas cujo código-fonte é disponibilizado publicamente, permitindo que desenvolvedores e usuários visualizem, modifiquem e redistribuam o *software* conforme suas necessidades. No contexto de BI, essas ferramentas se destacam não apenas por sua facilidade na visualização de dados, mas também pela capacidade de transformar dados brutos em representações gráficas que sejam intuitivas e acessíveis.

As soluções de código aberto reduzem custos organizacionais ao eliminar a necessidade de licenças comerciais [16]. Além disso, o acesso ao código-fonte viabiliza a reutilização e a personalização de *softwares*, o que facilita sua adaptação a demandas específicas [17]. Outro aspecto importante é a garantia de segurança e privacidade dos dados, já que o acesso aberto ao código possibilita auditorias independentes e facilita o monitoramento contínuo de vulnerabilidades [18].

Nesse cenário, o Apache Superset se destaca como uma ferramenta de BI de código aberto. Originalmente desenvolvido pela Airbnb em 2015, o Apache Superset foi projetado para atender às demandas de análise e visualização de dados em larga escala [19]. Seu diferencial está na capacidade de criar objetos de visualização interativos, *datasets* e *dashboards*, além de integrar-se a diversos bancos de dados.

A versatilidade do Apache Superset tem impulsionado seu uso em diferentes setores. Em um estudo sobre cidades inteligentes, a ferramenta foi empregada para integrar dados de tráfego e qualidade do ar, permitindo a criação de visualizações que facilitam a análise das interações entre essas variáveis [20]. Outro caso relevante ocorreu no setor energético, onde suas funcionalidades de exploração e interação com bases de dados foram utilizadas para transformar tabelas em análises visuais, otimizando a gestão de um repositório pré-existente que armazena grandes quantidades de informações [21].

Esses exemplos evidenciam como o Apache Superset se destaca para além do uso corporativo, demonstrando seu potencial para atender às necessidades de análise de dados no setor público.

A Figura 1 ilustra um painel de visualização criado com o Superset para acompanhar indicadores de matrículas escolares no Brasil. O indicador utilizado é um dos indicadores públicos disponibilizados pela PDE, construído a partir dos microdados do Censo Escolar, disponibilizados anualmente pelo INEP. O painel apresenta dois tipos de gráficos: um mapa interativo para análise geográfica da distribuição das matrículas e um gráfico de pizza que resume a participação proporcional das regiões ou categorias selecionadas.

## 3 METODOLOGIA

A avaliação do Apache Superset (versão 4.0.2) foi conduzida por três pesquisadores do laboratório C3SL. O estudo foi estruturado em dois eixos principais: avaliação heurística e avaliação de acessibilidade, buscando obter uma visão abrangente sobre esses aspectos e a adequação do *software* ao cenário de dados abertos. Como parte do protocolo experimental, os avaliadores percorreram fluxos de interação associados às principais tarefas no contexto da PDE:

**Criação de consultas no SQL Lab:** exploração direta dos dados;

**Construção e edição de gráficos:** transformação de consultas em visualizações;

**Organização e gerenciamento de dashboards:** composição e publicação de painéis destinados à exploração pública dos dados;

**Aplicação de filtros:** interação dinâmica e exploração de subconjuntos de dados;

**Controle de permissões:** gestão de acesso e administração da plataforma;

**Exportação de dados:** mecanismo de obtenção de dados pelos usuários finais.

Para orientar a escolha da ferramenta adotada, foram considerados a compatibilidade com o posicionamento institucional do laboratório e a viabilidade prática de implantação. Em particular, optou-se pelo Apache Superset por ser integralmente gratuito e de código aberto, alinhando-se ao compromisso do C3SL com a soberania digital. Ademais, durante as fases de prototipagem, a equipe constatou maior facilidade operacional na construção de uma arquitetura modular e de fácil manutenção com essa solução, em comparação com o Metabase e o Redash. Detalhes técnicos sobre o fluxo de ingestão, modelagem e integração com o Apache Superset são apresentados em trabalho anterior da equipe [22]. Ainda, é importante ressaltar que essa escolha expressa uma adequação pragmática ao contexto e às restrições da PDE, e não a afirmação de superioridade absoluta em todos os cenários.

### 3.1 Avaliação heurística

A usabilidade do Apache Superset foi analisada com base nas 10 heurísticas de Nielsen [23], um conjunto de diretrizes para identificação de problemas de interface. A avaliação seguiu uma abordagem exploratória, na qual cada pesquisador realizou a inspeção de forma independente, utilizando diferentes navegadores (Google Chrome, Mozilla Firefox e Opera GX), a fim de ampliar a cobertura sobre eventuais variações na experiência do usuário.

Os problemas identificados foram registrados individualmente e classificados segundo a heurística violada e seu nível de gravidade, conforme a escala sugerida por Nielsen:

- (1) Problema Cosmético: geralmente de natureza estética e não interferem na experiência do usuário;
- (2) Problema Leve: afeta a usabilidade de forma mínima, pois não impede o uso do sistema, mas sua correção pode melhorar a experiência do usuário;
- (3) Problema Grave: impacta significativamente a usabilidade e deve receber alta prioridade para correção;
- (4) Problema Catastrófico: extremamente crítico, pois impede o usuário de realizar suas tarefas ou alcançar seus objetivos.

Na etapa de consolidação, cada problema identificado foi discutido em grupo, com o intuito de compreender suas causas, consequências e severidade. Essas informações estão na Tabela 2, na qual as avaliações individuais foram reunidas nas respectivas colunas dos avaliadores. Essa abordagem permitiu refinar a lista original e mapear, com maior precisão, as inconsistências na interface do Apache Superset, contribuindo para a identificação de barreiras que podem impactar a usabilidade.

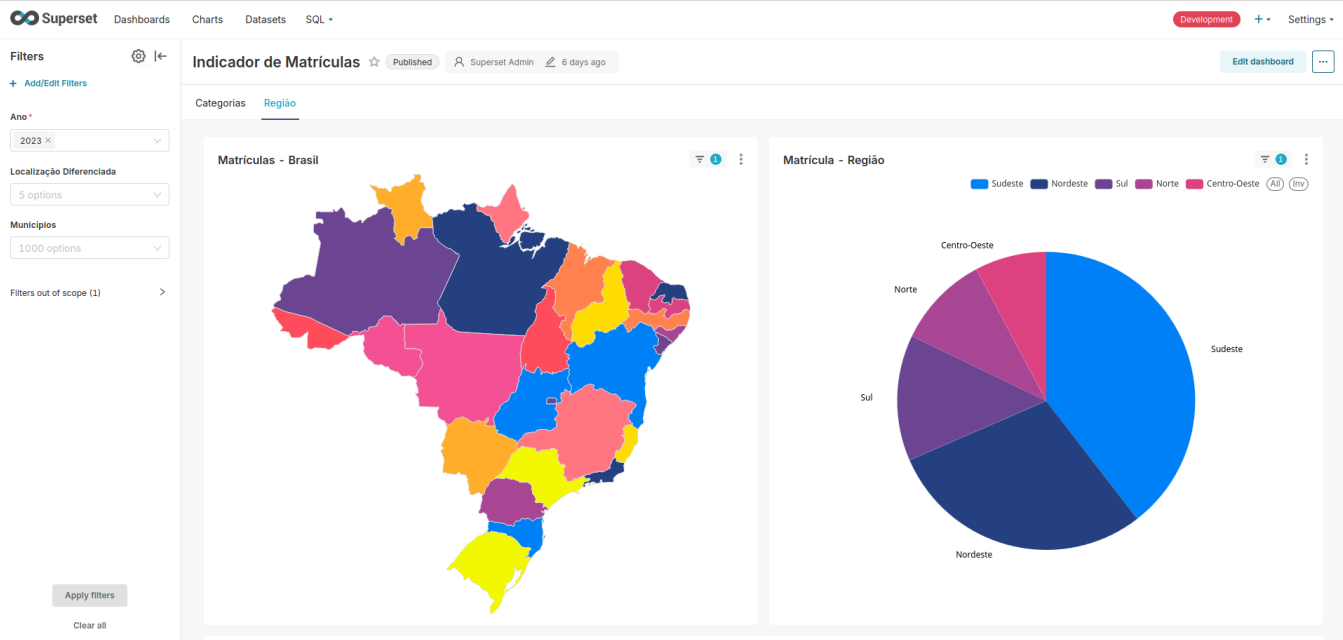


Figura 1: Exemplo de painel de visualização do Apache Superset.

### 3.2 Avaliação de Acessibilidade

A acessibilidade da ferramenta foi avaliada com 02 ferramentas automáticas: AccessMonitor [24] e LightHouse [25], que verificam a conformidade com as diretrizes do *Web Content Accessibility Guidelines* (WCAG) 2.1 [26]. Essas diretrizes, desenvolvidas pelo *World Wide Web Consortium*<sup>6</sup> (W3C), estabelecem recomendações de como tornar conteúdos digitais mais acessíveis a pessoas com diferentes tipos de deficiência. A WCAG 2.1 é baseada nos quatro princípios fundamentais de POCR<sup>7</sup>, garantindo que interfaces *web* possam ser acessadas e utilizadas por um público diverso, incluindo usuários que dependam de leitores de tela, navegação por teclado ou tecnologias assistivas.

Além disso, ela define três níveis de conformidade: A, AA e AAA. O nível A cobre os requisitos mínimos para garantir acessibilidade básica, o nível AA inclui diretrizes adicionais para melhorar a usabilidade, e o nível AAA estabelece critérios mais rigorosos, promovendo acessibilidade avançada.

A avaliação considerou os relatórios gerados individualmente por cada ferramenta, que, posteriormente, foram reunidos de acordo com as seções do Apache Superset às quais estavam associadas. Para cada seção, foram somadas as incidências de sucessos, avisos e erros, categorizados conforme os níveis de conformidade do WCAG, proporcionando uma visão geral do desempenho da plataforma em acessibilidade.

## 4 RESULTADOS E DISCUSSÃO

Esta seção apresenta uma análise do Apache Superset considerando aspectos de usabilidade, acessibilidade e sua conformidade com a Lei Geral de Proteção de Dados (LGPD). A partir das avaliações conduzidas, são discutidos os pontos positivos e as principais oportunidades de melhoria.

### 4.1 Análise heurística

A avaliação heurística do Apache Superset identificou padrões recorrentes de violações de usabilidade, destacando inconsistências à experiência do usuário. A Tabela 1 lista as Heurísticas de Nielsen, com seus respectivos códigos de identificação e descrições. Em seguida, a Tabela 2 detalha os problemas encontrados (identificados por P01, P02, etc), associando cada um: às heurísticas violadas; aos avaliadores (representados por A1, A2 e A3) que identificaram a violação; e ao grau de severidade. Os problemas estão organizados do grau mais crítico para o menos crítico, considerando também o número de avaliadores que os identificaram. As Figuras 2 e 3 complementam essa análise com representações visuais dos dados. A Figura 2 apresenta um gráfico de pizza cujas fatias indicam a quantidade de violações de usabilidade associadas a cada heurística. Já a Figura 3 exhibe a distribuição dos problemas conforme os graus de severidade atribuídos durante a avaliação.

A avaliação heurística do Apache Superset (Tabela 2) detectou padrões na distribuição e gravidade das infrações de usabilidade,

<sup>6</sup>Disponível em: <https://www.w3.org>. Acessado: 03/03/2026

<sup>7</sup>Sigla para Perceptível, Operável, Compreensível e Robusto, que orientam o desenvolvimento de conteúdos acessíveis e usáveis por pessoas com diferentes tipos de deficiência.

<sup>1</sup>*Structured Query Language* (Linguagem de Consulta Estruturada), linguagem padrão utilizada para gerenciamento e manipulação de bancos de dados relacionais.

<sup>2</sup>Estrutura lógica de um banco de dados que define a organização das tabelas, relações, visões, índices e outros elementos.

<sup>3</sup>Consulta realizada em um banco de dados, geralmente escrita em SQL, com o objetivo de recuperar, inserir, atualizar ou deletar dados.

Tabela 1: Heurísticas de usabilidade.

Código	Descrição
H1	Visibilidade do Estado do Sistema.
H2	Correspondência entre o sistema e o mundo real.
H3	Controle e liberdade do usuário.
H4	Consistência e padrões.
H5	Prevenção de erros.
H6	Reconhecimento ao invés de lembrança.
H7	Flexibilidade e eficiência de uso.
H8	Estética e design minimalista.
H9	Ajude os usuários a reconhecer, diagnosticar e corrigir erros.
H10	Ajuda e documentação.

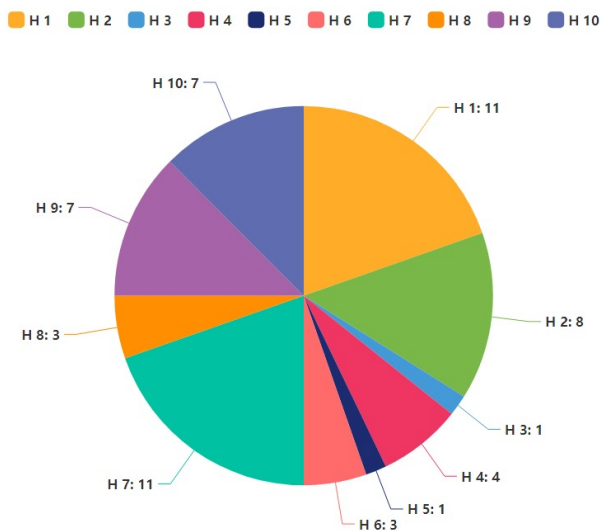


Figura 2: Quantidade de Problemas Identificados x Heurísticas.

evidenciando oportunidades de melhoria. Os resultados indicam que as heurísticas H1, H2 e H7 se destacam como as principais heurísticas violadas, concentram 53,4% das violações (30 de um total de 56 violações - Figura 2). A H1, com 11 infrações, apresentou falhas recorrentes na comunicação de estados do sistema, como ícones que permanecem ocultos até a interação do usuário e mensagens de erro que sobrepõem botões de ação. Essas limitações não apenas dificultam a navegação, mas também comprometem a confiança no sistema, especialmente em cenários críticos, como a perda de conexão com a internet sem atualização em tempo real (P15).

A H2, responsável por 14,3% das infrações (8 violações - Figura 2), foi identificada por mais de um avaliador, evidenciando problemas recorrentes na comunicação dos elementos e funcionalidades. Por exemplo, o acesso a *datasets* redireciona diretamente para a criação de gráficos, sem oferecer a opção de visualizar os componentes esperados ou realizar requisições específicas (P03), o que compromete a navegabilidade. Além disso, a nomenclatura pouco clara na

Cosmético Moderado Grave

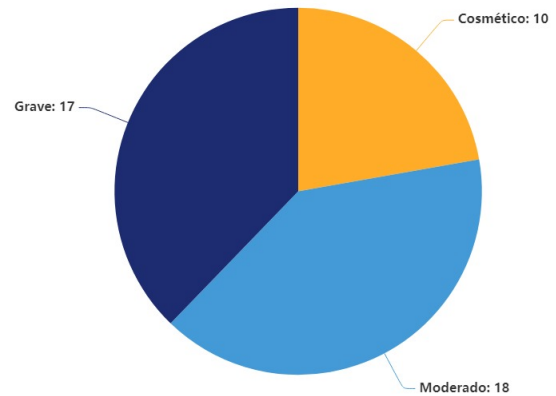


Figura 3: Quantidade de Problemas Identificados x Nível de Gravidade.

coluna de data do registro de ações dificulta a compreensão do usuário (P21). Essas falhas expõem lacunas na usabilidade do sistema, comprometendo a previsibilidade das ações e aumentando a curva de aprendizado. Essa falta de alinhamento com as expectativas dos usuários compromete a eficiência da navegação, especialmente para aqueles que estão utilizando a ferramenta pela primeira vez, podendo ter dificuldades em executar tarefas.

A heurística H7 registrou 11 infrações, 19,6% do total, indicando limitações da ferramenta em adaptar-se a diferentes perfis de usuários, especialmente em tarefas que demandam customização avançada. Por exemplo, problemas como a impossibilidade de escolher cores manualmente na criação de gráficos (P14), a restrição à ordenação livre de dados (P13) e a janela redundante do código SQL (P20), problemas classificados como Moderados e Cosméticos, evidenciam uma rigidez na personalização visual, restringindo a capacidade de usuários técnicos em otimizar visualizações para contextos específicos.

Ainda, ao analisar a concordância entre os avaliadores, observa-se um alinhamento moderado: 5 dos 24 problemas foram identificados por todos os avaliadores (21% dos achados), 11 foram reportados por dois avaliadores (46%), e 8 foram exclusivos a um único avaliador (33%). Destaca-se ainda que os itens em consenso pleno foram predominantemente classificados como graves, o que reforça sua priorização em intervenções corretivas. Por sua vez, os problemas identificados por apenas um avaliador não devem ser descartados, pois podem indicar padrões de interação específicos ou limitações contextuais não observadas pelos demais, recomendando-se sua verificação em avaliações complementares, enquanto os achados com maior concordância e severidade devem compor a lista prioritária de melhorias.

Apesar dos desafios, o sistema apresentou pontos positivos. As heurísticas H5 (Prevenção de erros) e H6 (Reconhecimento ao invés de lembrança) foram reportadas em apenas um problema, o que

Tabela 2: Consolidação da Avaliação Heurística.

ID	Problema	Heurísticas	A1	A2	A3	Grau
P01	Não há canal de comunicação direto com suporte ou criadores de visualizações no Apache Superset.	H10	x	x	x	Grave
P02	Botões de exportar, editar e deletar só aparecem ao passar o cursor sobre itens da lista de <i>dashboard</i> , <i>datasets</i> e gráficos.	H1 e H6	x	x	x	Grave
P03	Acesso a <i>datasets</i> redireciona para criação de gráficos, não para a requisição ou componentes esperados.	H4 e H2	x	x	x	Grave
P04	Mensagens de erro em anotações cobrem botões de adicionar e cancelar, atrapalhando a interação.	H1	x	x		Grave
P05	Painel de criação de gráficos é confuso e sobrecarregado com informações.	H1 e H8	x		x	Grave
P06	Ao adicionar tags aos <i>dashboards</i> , não é possível criar novas tags, apenas usar as existentes.	H7		x	x	Grave
P07	Mensagens de erro no SQL <sup>1</sup> Lab não são destacadas por cores e têm texto excessivo, dificultando a identificação do problema.	H9		x		Grave
P08	Não é possível exportar <i>dashboards</i> em formatos como PDF ou arquivos de imagens.	H9			x	Grave
P09	Respostas no SQL Lab podem apresentar carregamento infinito ou mensagens de erro pouco claras.	H9	x	x	x	Moderado
P10	Ausência de acesso a FAQ ou documentação dentro da plataforma.	H10	x	x	x	Moderado
P11	Alterações de cores no <i>dashboard</i> estão ocultas em ícones de edição, dificultando a compreensão do usuário.	H7		x	x	Moderado
P12	Erros em requisições são numerados e redirecionados para outra página, em vez de explicar o erro no SQL Lab.	H9	x	x		Moderado
P13	Não é possível ordenar dados livremente na criação de gráficos, apenas por critérios pré-definidos.	H7	x	x		Moderado
P14	Na criação de gráficos, não é possível escolher manualmente as cores para cada classe de dados.	H7		x	x	Moderado
P15	O sistema não atualiza em tempo real à perda de conexão com a internet.	H1		x		Moderado
P16	Não permite desfazer exclusões ou edições de <i>dashboards</i> e gráficos imediatamente.	H3 e H5			x	Moderado
P17	Criação e aplicação de filtros não são intuitivas para usuários não técnicos.	H10			x	Moderado
P18	Uso de jargões técnicos como " <i>dataset</i> ", " <i>schema</i> <sup>2</sup> ", " <i>SQL</i> " e " <i>query</i> <sup>3</sup> " sem clareza para usuários não técnicos.	H2			x	Moderado
P19	Botão de publicar/despublicar <i>dashboards</i> não é claramente indicado.	H1	x		x	Cosmético
P20	A janela inicial do código SQL em requisições salvas é desnecessária e confusa.	H7	x	x		Cosmético
P21	A coluna de data no registro de ações usa a sigla "dtm", sem clareza para o usuário.	H2	x	x		Cosmético
P22	Não há opção para trocar a imagem de ícone do usuário, apenas as iniciais são exibidas.	H2	x	x		Cosmético
P23	Não é possível visualizar todas as legendas dos gráficos simultaneamente.	H1 e H8			x	Cosmético
P24	Tabelas não são exibidas completamente, exigindo rolagem horizontal para visualização.	H7 e H4	x			Cosmético

indica que a ferramenta prioriza segurança operacional e interatividade de tarefas. No caso da H5, a plataforma incorpora mecanismos como validação automática de consultas ao banco, o que reduz riscos de inconsistências em requisições às bases de dados públicas. Já a H6 evidencia que a interface minimiza a dependência de memorização por meio de rótulos descritivos (por exemplo, ícones com descrições flutuantes).

Além disso, nenhum dos problemas identificados foi classificado como Catastrófico, o que indica que, apesar das falhas de usabilidade encontradas, não foram observadas situações que comprometessem totalmente o uso da ferramenta ou causassem prejuízos severos. Essa ausência reforça a estabilidade geral do sistema.

Em síntese, a concentração de infrações em heurísticas relacionadas a *feedback* visual (H1), correspondência entre o sistema e o mundo real (H2), e flexibilidade (H7) posiciona o Apache Superset em um estágio intermediário de maturidade em usabilidade, mas ainda dependente de melhorias em transparência operacional e inclusão de usuários não habituados a ferramenta.

## 4.2 Análise de acessibilidade

A avaliação de acessibilidade do Apache Superset foi realizada utilizando as ferramentas AccessMonitor e Lighthouse, com o objetivo de verificar sua conformidade com as diretrizes WCAG. Os resultados obtidos indicaram que a plataforma apresenta um nível satisfatório de conformidade com critérios básicos de acessibilidade no nível A, mas ainda enfrenta desafios para atender aos requisitos dos níveis AA e AAA.

A Tabela 3 apresenta os resultados obtidos na avaliação de acessibilidade, organizados por ferramenta. Para cada uma delas, são listadas as quantidades de ocorrências classificadas como sucesso, erro ou aviso, acompanhadas da distribuição correspondente entre os níveis A, AA e AAA. A coluna Total indica a soma geral de cada tipo de resultado para a respectiva ferramenta.

Os testes conduzidos com o AccessMonitor indicam uma taxa significativa de conformidade nos critérios do nível A, com 75% de taxa de sucesso (152 sucessos, 39 erros e 10 avisos). No nível AA, a conformidade diminuiu, registrando 37 sucessos, 9 erros e nenhum aviso, enquanto nenhum critério do nível AAA foi integralmente atendido. Foram identificados 11 avisos nesse nível, sugerindo a

Tabela 3: Resultados da análise de acessibilidade.

Ferramenta	Resultado	Total	A	AA	AAA
AccessMonitor	Sucesso	189	152	37	0
	Erro	48	39	9	0
	Aviso	21	10	0	11
LightHouse	Sucesso	144	126	17	1
	Erro	67	54	13	0
	Aviso	9	3	3	3

presença de barreiras que podem impactar a experiência de usuários com deficiências. Entre os principais problemas identificados estão a falta de descrições alternativas para elementos gráficos e o contraste inadequado entre o texto e o fundo, o que pode comprometer a usabilidade, especialmente para usuários que dependem de leitores de tela ou têm baixa visão.

A análise realizada com o LightHouse apresentou um panorama semelhante, com algumas variações nos quantitativos. No nível A, foram identificados 126 sucessos e 54 erros, enquanto no nível AA registraram-se 17 sucessos e 13 erros. No nível AAA, apenas um critério foi atendido, e três avisos foram emitidos. As divergências nos resultados obtidos pelas ferramentas podem ser explicadas pelas diferentes abordagens adotadas por cada uma. Enquanto o AccessMonitor prioriza a análise estrutural e semântica da interface, o LightHouse enfatiza métricas de experiência do usuário e interações automatizadas, levando a variações na detecção de problemas.

Em síntese, os resultados mostram que o Apache Superset apresenta um nível satisfatório de conformidade com critérios básicos de acessibilidade no nível A, mas ainda enfrenta desafios para atender aos requisitos dos níveis AA e AAA. Entre as principais barreiras identificadas, destaca-se o contraste insuficiente entre elementos textuais e de fundo, o que compromete a legibilidade para usuários com baixa visão. Além disso, a ausência de descrições textuais adequadas para elementos gráficos dificulta a utilização da ferramenta por leitores de tela.

### 4.3 Conformidade com a LGPD

Para além dos aspectos de acessibilidade e usabilidade, o estudo também avaliou o grau de conformidade do Apache Superset em relação à LGPD [4]. Durante as inspeções exploratórias, os avaliadores observaram que a plataforma pode expor consultas SQL em texto puro e conceder acesso direto a tabelas de dados brutos. Verificou-se também a ausência de recursos nativos para aplicar anonimização de informações sensíveis antes de sua exposição aos usuários.

Essas limitações possuem implicações legais claras. O Artigo 11 da LGPD define anonimização como um processo irreversível que impede qualquer associação dos dados a um indivíduo identificável. Quando a publicação ocorre a partir de consultas expostas em texto ou de tabelas brutas, aumenta-se o risco de reidentificação em conjuntos de dados de alta granularidade, o que contraria o princípio de irreversibilidade estabelecido pela legislação.

Diante desse cenário, torna-se necessário explicitar a distinção entre o que se espera da ferramenta de visualização e o que compete à arquitetura de dados e à governança institucional. Espera-se que a ferramenta minimize vetores de exposição por meio de: (i) controles de acesso; (ii) mecanismos para restringir exportações; e

(iii) recursos que reduzam a visibilidade de consultas em texto. A responsabilidade pela anonimização irreversível, pela classificação de sensibilidade dos conjuntos de dados e pela validação jurídica do compartilhamento deve recair sobre a arquitetura de ingestão e sobre o controlador dos dados.

Com isso, recomenda-se que em implantações públicas do Superset a publicação de dados seja realizada exclusivamente por meio de objetos adequados para divulgação, como tabelas pré-agregadas ou *views* preparadas previamente na camada de dados. Dessa forma, evita-se a exposição de consultas SQL em texto puro e o acesso direto a tabelas brutas destinadas à análise interna, reduzindo riscos de reidentificação e melhor alinhando o uso da ferramenta aos princípios estabelecidos pela LGPD.

## 5 CUIDADOS ÉTICOS

Este trabalho foi conduzido como uma atividade da equipe de Dados Educacionais para Políticas Públicas (DEPP), vinculado ao Centro de Computação Científica e Software Livre (C3SL), podendo ser considerado como uma avaliação da ferramenta para a Plataforma de Dados Educacionais (PDE).

Os avaliadores envolvidos na pesquisa são bolsistas de graduação vinculados ao DEPP, atuando na PDE. Para assegurar a privacidade desses estudantes, não foram coletadas ou divulgadas informações pessoais que permitissem sua identificação, nem foram feitas distinções entre suas funções específicas no projeto.

## 6 CONCLUSÃO

Este estudo avaliou a viabilidade do Apache Superset como ferramenta para disponibilização de dados públicos educacionais, considerando sua usabilidade e acessibilidade. O trabalho contribui para a área ao sistematizar evidências empíricas sobre o uso de uma ferramenta de BI de código aberto em um contexto real de dados públicos, articulando aspectos de usabilidade, acessibilidade e conformidade regulatória. Os resultados revelaram um sistema promissor, com funcionalidades robustas para visualização de dados, embora ainda apresente limitações.

Entre os pontos fortes, destaca-se a diversidade de visualizações interativas com funcionalidades como a capacidade de uma visualização em um painel filtrar automaticamente os dados exibidos em todas as outras visualizações e incorporação em sites externos, que ampliam sua aplicabilidade em contextos analíticos. A ferramenta também se sobressai em segurança operacional, oferecendo controle de permissões e registros de atividades. Adicionalmente, a estabilidade do sistema foi confirmada pela ausência de falhas catastróficas nas avaliações heurísticas, beneficiando-se de mecanismos eficientes de prevenção de erros (H5), como, por exemplo, a validação automática de consultas SQL.

Apesar dessas qualidades, as limitações identificadas revelam fragilidades que comprometem seu uso em contextos voltados à transparência e inclusão. No aspecto da usabilidade, destacam-se três problemas críticos: a recorrente falta de *feedback* visual adequado (H1), como mensagens de erro que sobrepõem botões (P04), a excessiva utilização de linguagem técnica e as quebras de expectativas dos usuários (H2), que são evidenciadas por redirecionamentos incoerentes (P03). Essas falhas, combinadas, resultam em uma curva de aprendizado mais acentuada do que o desejável. Em termos de

acessibilidade, a adesão restrita ao nível A das diretrizes WCAG, somada a falhas como baixo contraste e ausência de descrições textuais, dificulta o uso por pessoas com deficiência visual. Por fim, a exposição de informações sensíveis e a ausência de mecanismos automáticos de anonimização tornam a ferramenta incompatível com os princípios da LGPD, revelando deficiências no cumprimento de exigências regulatórias.

Nesse contexto, o Apache Superset mostra-se uma solução viável para ambientes que demandam alto grau de customização técnica, como consultas SQL avançadas e personalização visual via CSS. No entanto, em cenários orientados à cidadania digital e à democratização do acesso à informação, suas deficiências em usabilidade e acessibilidade configuram barreiras relevantes, cuja superação depende de intervenções estruturais no código-fonte.

Este estudo apresenta algumas limitações que devem ser consideradas na interpretação dos resultados. A avaliação de usabilidade baseou-se em inspeção heurística conduzida por avaliadores vinculados ao projeto, não incluindo testes com usuários finais ou públicos externos, o que pode limitar a validade externa dos achados. Além disso, a análise de acessibilidade foi realizada por ferramentas automatizadas, podendo se beneficiar de auditorias manuais e de avaliações com tecnologias assistivas em investigações futuras.

Como resultados emergentes, este estudo oferece contribuições preliminares ao debate sobre o uso de ferramentas de BI no setor público, com ênfase em iniciativas voltadas à transparência e acessibilidade de dados. Ainda que parciais, os achados indicam direções concretas para pesquisas futuras, como o desenvolvimento de camadas de personalização para públicos não técnicos e a adaptação do Apache Superset a exigências regulatórias. Para avançar nesse sentido, recomenda-se: (1) a revisão dos fluxos de interação e terminologias para que se alinhem às expectativas dos usuários; (2) a adoção de padrões WCAG AA/AAA, como alto contraste e legendas descritivas; e (3) a integração de módulos de anonimização de dados, em conformidade com a LGPD. Espera-se, assim, que esta investigação contribua para a comunidade de computação, em especial a de IHC, ao fomentar novas práticas e reflexões sobre soluções digitais mais inclusivas e socialmente comprometidas.

## AGRADECIMENTOS

Este trabalho foi apoiado pelo Centro de Computação Científica e Software Livre (C3SL) e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). Agradecemos também ao professor Marcos Sunye pelo apoio no projeto, aos professores Roberto Pereira e Simone Dominic pelo apoio nas orientações, revisões e contribuições ao longo da pesquisa. Estendemos nosso agradecimento ao colega Vitor Hugo A. Fávoro pelo suporte e discussões durante o desenvolvimento do trabalho. Agradecemos ao IFC por oportunizar o afastamento do servidor para capacitação. Por fim, esse trabalho utilizou de Inteligência Artificial para a revisão linguística.

## REFERÊNCIAS

- [1] Roberto Pereira, Ticianne Darin, and Milene Selbach Silveira. Grandihc-br: Grand research challenges in human-computer interaction in brazil for 2025-2035. In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, pages 1–24, 2024.
- [2] Thiago Adriano Coleti et al. Grandihc-br 2025-2035-gc5-human-data interaction data literacy and usable privacy. In *Proceedings of the XXIII Brazilian Symposium on Human Factors in Computing Systems*, pages 1–24, 2024.
- [3] R. Mortier et al. Human-Data Interaction. In *The Encyclopedia of Human-Computer Interaction*. Interaction Design Foundation, 2014.
- [4] Brasil. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, 2018. URL [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/13709.htm). Acesso em: 25 abr. 2025.
- [5] Rodrigo Hickmann Klein, Deisy Cristina Barbiero Klein, and Edimara Mezzomo Luciano. Identificação de mecanismos para a ampliação da transparência em portais de dados abertos: uma análise no contexto brasileiro. *Cadernos EBAPE.BR*, 16(4):639–655, 2018.
- [6] Open Data Charter. Open data charter, 2015. URL <https://opendatacharter.net/>. Acesso em: 29 abr. 2025.
- [7] Norberto Bobbio. *O futuro da democracia: uma defesa das regras do jogo*. Paz e Terra, Rio de Janeiro, 10 edition, 2000.
- [8] Tribunal de Contas da União (Brasil). Relatório Sistemático sobre Transparência Pública: FiscTransparência. <https://portal.tcu.gov.br/publicacoes-institucionais/relatorio-de-fiscalizacao/fisc-transparencia-relatorio-sistemico-sobre-transparencia-publica>, 2019. Acesso em: 25 abr. 2025.
- [9] F. R. Botelho and E. Razzolini Filho. Conceituando o termo Business Intelligence: origem e principais objetivos. *Sistemas, Cibernética e Informática*, 11(1):55–60, 2014. ISSN 1690-8627.
- [10] CONSAD. Otimização da gestão pública através de técnicas de Business Intelligence. <https://www.consad.org.br/wp-content/uploads/2013/02/OTIMIZA%C3%87%C3%83O-DA-GEST%C3%83O-P%C3%9ABLICA-ATRAV%C3%89S-DE-T%C3%89NICAS-DE-BUSINESS-INTELLIGENCE.pdf>, 2013. Acesso em: 25 abr. 2025.
- [11] M. D. M. Pontes, T. L. D. Pontes, and R. D. Andrade. A adoção de sistemas de Business Intelligence & Analytics na contabilidade de gestão por entidades da Administração Pública: uma revisão da literatura. *Revista Facultad De Ciencias Económicas*, 29(1):95–114, 2021.
- [12] R. A. W. Rhodes. The new governance: governing without government. *Political Studies*, 44(4):652–667, 1996.
- [13] G. A. Boyne. Public and private management: What's the difference? *Journal of Management Studies*, 39(1):97–122, 2002.
- [14] L. O. Leite and D. A. Rezende. Case Study of the Utilization of Business Intelligence (BI) in the Municipal Prefecture of Curitiba. In *5th International Conference on Information Systems and Technology Management / 5o Congresso Internacional de Gestão da Tecnologia e Sistema de Informação*, pages 2897–2918, 2008.
- [15] M. G. Oliveira and E. G. C. Aratújo. Transformação digital no SUS: implantação da tecnologia de Business Intelligence na diretoria regional de saúde norte de belo horizonte. In *X Encontro Brasileiro de Administração Pública*, 2023.
- [16] D. F. Freitas. Análise de Ferramentas de Business Intelligence: Como Ferramentas Open Source Podem Reduzir Custos em Negócios. Master's thesis, Universidade Tecnológica Federal do Paraná, Curitiba, 2017. MBA em Gestão da Tecnologia da Informação e Comunicação, Departamento Acadêmico de Eletrônica.
- [17] K. H. Bennett, V. Rajlich, and N. Wilde. Software evolution and the staged model of the software lifecycle. *Avanços em Computadores*, 56:1–54, 2002. doi: 10.1016/S0065-2458(02)80003-1. Fonte: DBLP, Universidade da Flórida Ocidental.
- [18] P. Sanyal, S. Sharma, D. Bura, and P. Banerjee. On the Security of Open Source Software. *International Journal of Advanced Research*, 5(11):1338–1348, November 2017. doi: 10.21474/IJAR01/5904. License: CC BY 4.0. Authors' affiliations: Manav Rachna International Institute of Research and Studies; CL Educate Limited.
- [19] Apache Superset. Apache Superset. <https://superset.apache.org/>, 2024. Acesso em: 29 abr. 2025.
- [20] A. Benipal. Web-GIS Based Dashboard for Real-Time Data Visualization & Analysis Using Open Source Technologies. *Journal of Geomatics*, 16(2):42–50, 2022. doi: 10.58825/JOG.2022.16.2.42.
- [21] T.-Y. Chen, C.-T. Yang, E. Kristiani, and C.-T. Cheng. On Construction of a Power Data Lake Platform Using Spark. In *Frontier Computing*. Lecture Notes in Electrical Engineering, pages 99–108. Springer, 2019. doi: 10.1007/978-981-13-3648-5\_11.
- [22] Joao Pedro V Ramalho, João A Silveira, Thales Gabriel C de Lima, Mateus S Herbele, Josiney de Souza, Guilherme A Derenievicz, Leticia M Peres, and Simone Dominic. Utilização do apache superset para visualização escalável de dados educacionais públicos: Um estudo de caso com o censo escola. In *Simpósio Brasileiro de Banco de Dados (SBBDD)*, pages 914–920. SBC, 2025.
- [23] J. Nielsen. Heuristic Evaluation. In Robert Mack and Jakob Nielsen, editors, *Usability Inspection Methods*, pages 25–62. John Wiley & Sons, New York, NY, 1994.
- [24] AccessMonitor. AccessMonitor – Ferramenta de Avaliação de Acessibilidade Web. <https://accessmonitor.acessibilidade.gov.pt>, 2024. Acesso em: 30 abr. 2025.
- [25] Google Chrome Web Store. Lighthouse. <https://chromewebstore.google.com/detail/lighthouse/blipmdconlkpinefehnmjammfjpmpbjk?hl=pt-br&pli=1>, 2025. Acesso em: 25 abr. 2025.
- [26] World Wide Web Consortium. Web Content Accessibility Guidelines (WCAG) 2.1. <https://www.w3.org/TR/WCAG21/>, 2018. Acesso em: 30 abr. 2025.