|FACULDADE PARA O DESENVOLVIMENTO SUSTENTÁVEL DA AMAZÔNIA
CURSO TECNÓLOGO EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS

FELIPE SANTOS FERREIRA

OTIMIZAÇÃO DE CONSULTAS DE BANCO DE DADOS ATRAVÉS DA IMPLEMENTAÇÃO DE UM SISTEMA DE ESTRUTURAÇÃO E HIERARQUIZAÇÃO USANDO I.A

PARAUAPEBAS
2025

FELIPE SANTOS FERREIRA

OTIMIZAÇÃO DE CONSULTAS DE BANCO DE DADOS ATRAVÉS DA IMPLEMENTAÇÃO DE UM SISTEMA DE ESTRUTURAÇÃO E HIERARQUIZAÇÃO USANDO I.A

Trabalho de Conclusão de Curso (TCC) apresentado a Faculdade para o Desenvolvimento Sustentável da Amazônia (FADESA), como parte das exigências do Programa do Curso Tecnólogo em análise e desenvolvimento de sistemas para a obtenção do Título de Tecnólogo.

Orientadora: Prof.ª Esp. Sara Debora Carvalho Cerqueira

PARAUAPEBAS
2025

Nota: A verão original deste trabalho de conclusão de curso encontra-se disponível no Serviço da Biblioteca e Documentação da Faculdade para o Desenvolvimento Sustentável da Amazônia – FADESA em Parauapebas – PA.

Ferreira, Felipe Santos.

F383o

Otimização de consultas de banco de dados através da implementação de um sistema de estruturação e hierarquização usando I.A. / Felipe Santos Ferreira – Parauapebas / PA: FADESA, 2025.

42f.

Trabalho de Conclusão de Curso (Graduação) – Faculdade para o Desenvolvimento Sustentável da Amazônia – FADESA, Tecnólogo em Análise e Desenvolvimento de Sistemas, 2025.

Orientadora: Prof. Esp.: Sara Debora Carvalho Cerqueira.

1. Big Data. 2.Inteligência Artificial. 3.Filtragem de Dados. 4.Qualidade da Informação. 5. Dados Estruturados. I. Cerqueira, Sara Debora Carvalho. II. Faculdade para o Desenvolvimento Sustentável da Amazônia. III. Título.

CDD 004

Ronald de Jesus Alves Ribeiro
Bibliotecário
CRB - 2/1837

FELIPE SANTOS FERREIRA

OTIMIZAÇÃO DE CONSULTAS DE BANCO DE DADOS ATRAVÉS DA IMPLEMENTAÇÃO DE UM SISTEMA DE ESTRUTURAÇÃO E HIERARQUIZAÇÃO USANDO I.A

Trabalho de Conclusão de Curso (TCC) apresentado a Faculdade para o Desenvolvimento Sustentável da Amazônia (FADESA), como parte das exigências do Programa do Curso Tecnólogo em análise e desenvolvimento de sistemas para a obtenção do Título de Tecnólogo.

Aprovado em: 11 / 06 / 2025.

Banca Examinadora

Prof. Esp. Adriano Louzada Bollas
Faculdade Para o Desenvolvimento Sustentável da Amazônia
(Avaliador)

Prof. Esp. Antônio Soares da Silva
Faculdade Para o Desenvolvimento Sustentável da Amazônia
(Avaliador)

Prof. ª Esp. Sara Debora Carvalho Cerqueira
Faculdade Para o Desenvolvimento Sustentável da Amazônia
(Orientadora)

Data de depósito do trabalho de conclusão de curso:____/___/_____

FELIPE SANTOS FERREIRA

OTIMIZAÇÃO DE CONSULTAS DE BANCO DE DADOS ATRAVÉS DA IMPLEMENTAÇÃO DE UM SISTEMA DE ESTRUTURAÇÃO E HIERARQUIZAÇÃO USANDO I.A

Trabalho de Conclusão de Curso (TCC) apresentado a Faculdade para o Desenvolvimento Sustentável da Amazônia (FADESA), como parte das exigências do Programa do Curso Tecnólogo em análise e desenvolvimento de sistemas para a obtenção do Título de Tecnólogo.

Aprovado em: ___/___/____.

Felipe Santos Ferreira
(Discente)

Prof. Esp. Antônio Soares da Silva
(Coordenador do Curso de Análise e Desenvolvimento de Sistemas)

Dedico este trabalho primeiramente a Deus, por acompanhar minhas guerras e em segundo, ao futuro, para que seja prospero e sustentável.

AGRADECIMENTOS

Expresso gratidão a Prof.(a) Esp. Sara Debora Carvalho pelo empenho e dedicação, demonstrando ser uma profissional exemplar na área da educação.

“A educação tem raízes amargas, mas os seus frutos são doces.” - Aristóteles

RESUMO

O crescimento exponencial do volume de dados disponíveis na internet e nas buscas realizadas tem gerado desafios e oportunidades significativas para a gestão da informação. Relatórios da DOMO e do Google revelam que, em 2024, cerca de 20 petabytes de dados são processados diariamente pela plataforma de buscas, evidenciando a escalada na criação e circulação de conteúdo digital. Contudo, essa abundância informacional também trouxe consequências negativas, como a anomia digital — fenômeno caracterizado pela desorganização e pela baixa confiabilidade dos dados disponíveis. Esse cenário compromete a qualidade da informação, eleva custos operacionais e dificulta a obtenção de resultados eficientes. Nesse contexto, as ferramentas baseadas em inteligência artificial têm se consolidado como alternativas viáveis e eficazes para o enfrentamento desses problemas, ao oferecer recursos capazes de filtrar, organizar e estruturar dados relevantes. Com isso, promovem maior assertividade nas buscas, reduzem custos e otimizam os processos de análise e tomada de decisão.

Palavras-chave: Big Data, Inteligência Artificial, Filtragem de Dados, Qualidade da Informação, Anomia Digital, Dados Estruturados.

ABSTRACT

The exponential growth in the volume of data available on the internet and in searches performed presents significant challenges and opportunities. Data from DOMO and Google highlights the impressive escalation of information creation and management, with 20 PB processed daily by Google by 2024. However, this abundance of information has brought challenges such as digital anomie, where clutter and low reliability compromise quality of information, increasing costs and making it difficult to obtain effective results. In this context, artificial intelligence tools have proven to be reliable solutions for filtering, structuring and organizing relevant data, reducing operational costs and improving query efficiency.

Keywords: Big Data, Artificial Intelligence, Data Filtering, Information Quality, Digital Anomie, Structured Data.

SUMÁRIO

1. INTRODUÇÃO 12
2. A EVOLUÇÃO DA MANIPULAÇÃO DE INFORMAÇÕES 14
2.1. As formas de tratamento de dados 14
2.2. Metodologias para extração de informação 17
2.3. Uso de recursos para processamento de dados 17
2.4. Metodologias para classificação de informação 18
3. A EXPANSÃO DOS DADOS 19
3.1. Consequências da expansão 19
3.2. Impactos no setor tecnológico 20
3.3. Visão futura sobre a expansão contínua dos dados 21
4. METODOLOGIA 22
4.1. Método de pesquisa 22
4.2. Método de desenvolvimento do projeto – Metodologia Cascata 23
4.3. Planejamento e Análise de Viabilidade 24
4.4. Garantia de Qualidade 25
4.5. Aspectos Éticos 25
4.6. Critérios de Inclusão e Exclusão 26
5. RESULTADOS E DISCUSSÕES 27
5.1. Análise de Requisitos 27
5.1.1. Requisitos Funcionais 27
5.1.2. Requisitos Não Funcionais 28
5.2. Funcionalidades da Ferramenta 29
5.3. Tecnologias Utilizadas 33
5.4. Plataforma de Desenvolvimento 34
5.5. Visão Geral da Ferramenta 36
5.6. Testes e Resultados 37
6. CONSIDERAÇÕES FINAIS 39
7. REFERÊNCIAS 40

1. INTRODUÇÃO

Observa-se no cenário atual, um aumento exponencial no número de informações disponíveis na internet e no número de buscas sendo que, segundo a empresa de soluções em nuvem Domo (2017), houve a criação de cerca de 2.5QI de dados por dia, mas em 2006, o Google já reportava o armazenamento de 850TB de dados e uma consulta média de 4TB por pesquisa. Mais de uma década depois, é inegável que esses números foram multiplicados exponencialmente.

Atualmente, de acordo com o projeto internacional de monitoramento em tempo real (Internet Live Stats, 2024) o Google gerencia aproximadamente 20 PentaBytes de dados todos os dias e sua ferramenta de pesquisa processa aproximadamente 3.5 Bilhões de pesquisas por dia, ou seja, 40 mil pesquisas por segundo, representando a crescente escala de informações manipuladas por ferramentas de big data e IA generativa.

A proliferação de dados, embora ofereça um universo de possibilidades, também apresenta desafios significativos como a anomia digital descrita por Negroponte (1996), que se originou do alto volume de informações na internet que geravam desordem devido aos diferentes tipos de dados que necessitam ser filtrados e a baixa confiabilidade resultante da quantidade de informações irrelevantes que foram disponibilizadas na rede.

Destaca-se que a utilização de ferramentas com inteligência artificial provou-se ser uma solução confiável para melhorar a filtragem de dados ao coletar, analisar, selecionar e por fim extrair informações e organizá-las tornando-as estruturadas e relevantes, reduzindo custos por utilização dos recursos de hardware em grandes tarefas de consulta.

Fica evidenciado que o número crescente de informações trouxe desafios complexos relacionados a manipulação de dados em larga escala, no qual dificulta a localização de informações validas e relevantes devido a quantia de dados disponíveis, aumentando o tempo de análise e extração de uma informação para converter em um dado, prolongando a conclusão de uma tarefa.

Dessa forma, qual solução será eficaz para sanar o problema caracterizado pela baixa eficiência de consultas devido a expansão de dados e ao alto consumo de recursos computacionais nos atuais sistemas de gerenciamento de banco de dados.

O projeto tem como objetivo principal propor uma solução para otimizar a manipulação de grandes volumes de dados em bancos de dados, utilizando técnicas de Inteligência Artificial (IA) e Extração, Transformação e Carregamento (ETL) para automatizar e estruturar as consultas de banco de dados, melhorando a eficiência, qualidade e integridade das informações processadas. O sistema proposto visa reduzir os custos operacionais e o tempo de processamento, garantindo que as empresas possam acessar dados relevantes e confiáveis de forma mais rápida, econômica e sustentável.

Ademais, o projeto possui objetivos específicos voltados a aperfeiçoar consultas em banco de dados através da automação de processos utilizando a I.A, estruturar a automação para otimizar os procedimentos de coleta, análise e extração de dados e desenvolver um sistema de estruturação e hierarquização dos dados que permita a filtragem, classificação e padronização das informações, prevenindo inconsistências e facilitando a tomada de decisões baseada em dados de alta qualidade.

Segundo estudo realizado pela IBM (2023), entre dezembro de 2023 a abril de 2024, no qual foi analisado mais de 2,500 CEOs de 30 países e 26 indústrias diferentes, apontaram que no Brasil, 44% dos CEOs tiveram dificuldades em processar cálculos para medir a qualidade dos dados e 41% encontraram dificuldades em obter resultados relevantes para análise, demonstrando uma clara deficiência na falta de uma ferramenta que auxilie a completar esses processos.

Além disso, o estudo também destaca a importância da Hierarquização e Estruturação dos dados a partir do levantamento dos CEOs brasileiros sobre a prioridades de ações sendo que 44% apoiam a padronização dos processos de coleta de dados e relatórios além de ressaltar que 65% dos entrevistados consideram a IA como ferramenta determinante futura em tomadas de decisões a partir das análises de dados.

Portanto, torna-se evidente a necessidade de uma ferramenta eficiente para realizar a manipulação e tratamento de informações cujo proposito visa reduzir o uso de recursos naturais e computacionais através de técnicas avançadas de extração, transformação e carregamento de informações (ETL) permitindo que profissionais da área otimizem tarefas de consulta e injeção de dados por meio da filtragem, limpeza e otimização de informações submetidas a ferramenta

2. A EVOLUÇÃO DA MANIPULAÇÃO DE INFORMAÇÕES

Surgindo através da influência dos primeiros mecanismos matemáticos propostos por Euclides (300 a.C) que através da abordagem axiomática e da geometria euclidiana, fundamentou uma das primeiras bases para construção e estruturação de um sistema logico, seguido por Hiparco (1974) que desenvolveu um sistema para catalogar e organizar dados astronômicos.

Sendo assim, por volta de 1450, Johannes Gutenberg foi responsável por revolucionar o armazenamento e a distribuição de informações ao inventar uma máquina para imprimir e gravar palavras em blocos de impressão, contribuindo ativamente para futuras abordagens. Posteriormente, em 1880, Herman Hollerith desenvolveu uma máquina de tabulação que utilizava cartões perfurados para ler e processar dados de forma automatizada, representando um dos primeiros marcos na transição entre o tratamento manual de informações e o processamento computacional.

Após grandes evoluções e criações conceituais ao longo da história, sucedeu ao Pai da teoria da informação, Claude Shannon na teoria matemática da informação (1949), que elaborou medidas para lidar com a codificação de informações durante a transmissão de dados ao transformar informações em objetos matemáticos, possibilitando a análise e quantificação dos dados e a fundamentar a teoria que descreve a informação, auxiliando ativamente na criação futura das primeiras bases de informações.

Por fim, toda trajetória envolvendo a manipulação de informações culminou na fundação do Google em 1998, a ferramenta de busca que ampliou horizontes ao disponibilizar uma forma para encontrar informações em toda a rede de computadores. Sendo que atualmente, a manipulação de informações integra inteligência artificial e machine learning e sistemas modernos utilizam algoritmos para análise preditiva, reconhecimento de padrões e processamento de linguagem natural (NLP), transformando dados brutos em insights valiosos.

2.1. As formas de tratamento de dados

Segundo French (1996), o processamento de dados é a coleta e manipulação de dados digitais para produzir informações significativas. O tratamento de dados

evoluiu significativamente ao longo dos anos, com métodos que vão desde operações manuais e processamento em lotes até sistemas de processamento em tempo real. Historicamente, os dados eram organizados em sistemas centralizados, como mainframes.

Com o avanço da tecnologia, surgiram bancos de dados relacionais que melhoraram a estruturação e acessibilidade da informação. Dentre as formas existentes de tratamentos de dados, segue-se uma linha logica de tratamento de informação há depender das especificações de cada projeto que se descrevem em 7 etapas cruciais, sendo elas, Coleta, Armazenamento, Processamento, Análise, Compartilhamento e Visualização, Proteção e Segurança e Exclusão ou arquivamento. Dessa forma, Milani (2020, p. 28) descreve: “Portanto, é fundamental permitir que os usuários personalizem, modifiquem e refinem interativamente as visualizações até que sintam que atingiram seu objetivo”.

A etapa de coleta de dados é responsável por reunir informações do cliente de forma estruturada ou não estruturada, dependendo da natureza e dos objetivos do projeto. Nesta etapa, as empresas utilizam diferentes métodos e ferramentas para capturar dados relevantes. Esses métodos podem incluir questionários, formulários online, sistemas de automação, e até mesmo rastreamento de interações em plataformas digitais. Olsen (2015, p. 144) destaca, “A finalidade básica da coleta de dados é criar um conjunto sistematicamente organizado de materiais que possam ser usados para testar ou gerar hipóteses”.

Após a coleta, segue-se a etapa de armazenamento de dados, que visa organizar e preservar as informações capturadas em sistemas seguros e acessíveis. O armazenamento pode ocorrer em bancos de dados relacionais, como MySQL e PostgreSQL, que organizam os dados em tabelas, ou em bancos de dados não relacionais (NoSQL), como MongoDB e DynamoDB, que oferecem maior flexibilidade para lidar com dados não estruturados, como imagens ou logs.

Na sequência, ocorre o processamento de dados, que transforma dados brutos em informações estruturadas e compreensíveis. Esta etapa utiliza ferramentas e técnicas como ETL (Extract, Transform, Load), que permitem extrair dados de diferentes fontes, convertê-los para formatos adequados e carregá-los em sistemas de análise. Além disso, soluções de big data, como Apache Hadoop e Spark, são frequentemente empregadas para processar grandes volumes de dados com alta eficiência. Milani (2020, p. 34-35) afirma que: “A transformação de dados consiste em

transformar dados brutos, de difícil compreensão humana, em relações lógicas mais estruturadas e, portanto, mais fáceis de serem visualizadas".

A análise de dados é uma das etapas mais críticas no tratamento de informações, pois é nesse momento que os dados começam a gerar valor para a organização. Essa análise pode ser descritiva, buscando entender o que ocorreu em determinado período, ou preditiva, utilizando algoritmos de aprendizado de máquina para antecipar tendências futuras. Em projetos mais avançados, a análise prescritiva também pode ser aplicada, recomendando ações específicas com base nos dados analisados.

Com os insights obtidos, passa-se para a fase de compartilhamento e visualização, onde as informações são disseminadas de forma clara e acessível para os tomadores de decisão. Dashboards interativos e relatórios automatizados são exemplos de produtos dessa etapa, permitindo uma compreensão ágil e precisa dos dados analisados. A visualização desempenha um papel importante na comunicação, traduzindo dados complexos em representações gráficas e intuitivas que auxiliam na tomada de decisões estratégicas.

A proteção e segurança dos dados é uma etapa transversal e contínua, que permeia todo o ciclo de vida do tratamento de informações. Com a crescente preocupação em relação à privacidade e segurança, legislações como a Lei Geral de Proteção de Dados (LGPD) no Brasil e o Regulamento Geral sobre a Proteção de Dados (GDPR) na Europa impõem diretrizes rigorosas para garantir a integridade e confidencialidade das informações.

Segundo Vida (2021, p. 166):

> "Essa etapa é considerada a mais importante do processo ETL, porque nela é possível ter uma melhoria na integridade dos dados e auxiliar na garantia de que os dados cheguem ao seu novo destino com total compatibilidade e prontos para serem utilizados".

Por fim, a etapa de exclusão ou arquivamento de dados é responsável por definir a trajetória final das informações. Dados que já não são necessários podem ser excluídos de forma segura para evitar seu uso indevido, enquanto dados históricos ou que possuem valor potencial podem ser arquivados para consultas futuras. O cumprimento de normas e políticas internas é essencial nessa etapa, especialmente no caso de dados sensíveis.

2.2. Metodologias para extração de informação

Segundo Han et al. (2011), os métodos de extração de informações podem ser classificados com base na natureza dos dados, sendo eles, dados estruturados, que seguem um formato pré-definido, como tabelas ou não estruturados no qual as Informações não possuem uma organização rígida, como textos, imagens e vídeos.

Para dados estruturados, o uso de linguagens de consulta como SQL (Structured Query Language) é predominante. Entretanto, para dados não estruturados, técnicas de processamento de linguagem natural (NLP) e visão computacional são amplamente empregadas.

Dentre as metologias utilizadas para extração de informações, o processo de ETL segue sendo o mais utilizado para construção de Data Warehouse representando mais da metade da criação (GOUR et al., 2010) além de ser predominantemente utilizado em BI (IBM CLOUD EDUCATION, 2020). Significando Extrair, Transformar e Carregar, o ETL é o principal responsável por tratar as informações recebidas, limpando e estruturando informações para serem enviadas ao sistema.

2.3. Uso de recursos para processamento de dados

Os recursos computacionais, como CPU, memória RAM e armazenamento, desempenham um papel central no processamento de dados. Segundo Silberschatz et al. (2020), no livro *Database System Concepts*, a capacidade de processar grandes volumes de dados depende diretamente do uso eficiente do hardware e da arquitetura dos sistemas.

Além disso, ferramentas de software desempenham um papel crucial no processamento de dados, fornecendo as bases para manipulação, transformação e análise. De acordo com Dean e Ghemawat (2008), frameworks como Hadoop e Spark são projetados para processar grandes volumes de dados em sistemas distribuídos, otimizando o uso de recursos computacionais por meio do particionamento e da paralelização de tarefas.

Entre os recursos utilizados durante processamento de dados, além de recursos computacionais, os Centro de Dados, locais destinados especificamente para manipular e armazenar informações, usufruem de recursos hídricos para refrigeração das máquinas. Segundo o relatório ambiental do Google de 2023, foram

registradas uma média de consumo de cerca 21 bilhões de litros de água por ano gastos em Data Centers que representou um aumento de 20% em comparação aos resultados divulgados em 2021.

2.4. Metodologias para classificação de informação

Entre as metodologias utilizadas para classificar informações, a classificação manual é uma das abordagens mais tradicionais para organizar informações. Consiste em atribuir rótulos ou categorias às informações com base em critérios predefinidos por especialistas ou usuários. De acordo com Davenport (1997), a classificação manual é comum em sistemas de gestão de documentos e arquivamento, onde os profissionais atribuem categorias de acordo com o conteúdo ou a finalidade das informações.

Por outro lado, a classificação baseada em regras é uma metodologia automatizada que utiliza conjuntos de regras pré-estabelecidas para categorizar informações. Segundo Russel e Norvig (2016), a classificação baseada em regras pode ser eficaz quando os dados apresentam características bem definidas, mas enfrenta limitações quando a variedade ou a complexidade das informações aumenta.

Entretanto, com o aumento da complexidade e volume de dados, métodos baseados em aprendizado de máquina têm se tornado a norma para a classificação de informações. Esses algoritmos aprendem com os dados e se ajustam a diferentes padrões para categorizar as informações automaticamente, sem a necessidade de regras pré-definidas como classificadores supervisionados e não supervisionados.

Por fim, A classificação hierárquica é uma metodologia onde as informações são organizadas em uma estrutura hierárquica ou árvore, onde categorias principais se dividem em subcategorias. Segundo Liu et al. (2016), a classificação hierárquica é vantajosa em sistemas de categorização de conteúdos em que existe uma relação clara entre as categorias.

Dessa forma, através das técnicas de Extração, Transformação e Carregamento de dados torna-se possível realizar a filtragem de informações, possibilitando a classificação das informações recebidas, devendo-se avaliar qual metodologia será mais adequada para realizar o processo de tratamento e limpeza de dados.

3. A EXPANSÃO DOS DADOS

A expansão dos dados é classificada como um dos assuntos mais relevantes da era digital. Nos últimos anos, a quantidade de informações geradas globalmente aumentou exponencialmente, impulsionada principalmente pela popularização de dispositivos conectados, redes sociais e sistemas de Big Data que resultaram na criação e postagem em massa de conteúdos na internet.

Segundo Marr (2015), estima-se que, até 2025, o volume de dados gerados no mundo ultrapasse 175 zettabytes. Este crescimento tem sido impulsionado por avanços tecnológicos que tornam a coleta, o armazenamento e a análise de dados mais rápidos e eficientes. Como resultado, as organizações têm agora acesso a um vasto repositório de informações, que pode ser utilizado para extrair insights valiosos e criar soluções inovadoras.

Além disso, a expansão dos dados tem incentivado a inovação em diferentes setores, desde a saúde, com a coleta de dados genéticos ampliando métodos para a realização de análises em tempo real, até o setor financeiro, que utiliza algoritmos de análise preditiva para identificar tendências de mercado a partir de informações de noticiais e de outros ativos que foram coletados pela internet.

À medida que mais dados se tornam acessíveis e são incorporados ao processo de tomada de decisões, as empresas se tornam mais focadas em dados, o que redefine práticas tradicionais e modelos de negócios. O uso de análises avançadas e de ferramentas de visualização de dados permite que as organizações obtenham insights estratégicos que antes seriam impossíveis de alcançar.

3.1. Consequências da expansão

A expansão dos dados trouxe várias consequências significativas tanto para as empresas quanto para os indivíduos. Uma das principais consequências é o aumento da complexidade na gestão e no processamento desses dados. Com volumes imensos de informações a serem analisadas em tempo real, surgem desafios em relação à infraestrutura de TI, à segurança e à privacidade dos dados.

Como observado por Davenport e Harris (2007), "os dados passaram a ser o principal ativo das empresas, mas, ao mesmo tempo, a gestão desses dados requer recursos cada vez mais sofisticados". Essa complexidade obriga as organizações a

investir fortemente em tecnologias e talentos especializados, como especialistas em Big Data e engenheiros de dados.

Outra consequência relevante é o aumento das preocupações com a segurança e a privacidade. A coleta massiva de dados pessoais e sensíveis, como hábitos de consumo, localização e até dados de saúde, cria vulnerabilidades. Isso se traduz em um risco maior de ataques cibernéticos, roubo de identidade e uso indevido de informações.

De acordo com Zikopoulos e Eaton (2011), a quantidade de dados gerados pelas empresas hoje "é tão grande que as preocupações com a segurança e a privacidade nunca foram tão críticas". O fortalecimento de políticas de governança de dados e o uso de tecnologias como a criptografia se tornam essenciais para mitigar esses riscos.

Dessa forma, a expansão dos dados também levou a um aumento significativo na personalização de produtos e serviços. As empresas passaram a coletar e analisar dados sobre comportamentos e preferências dos consumidores, permitindo que ofertas sejam ajustadas de maneira mais precisa e eficaz. No entanto, essa personalização também gera desafios éticos, especialmente quando os dados são utilizados sem o consentimento adequado dos consumidores ou quando a coleta de dados invade a privacidade do usuário.

3.2. Impactos no setor tecnológico

A expansão dos dados tem gerado impactos profundos no setor tecnológico, principalmente no desenvolvimento de novas tecnologias para processamento e análise de informações. O aumento da demanda por soluções que possam lidar com grandes volumes de dados tem impulsionado o avanço de tecnologias como Big Data, Inteligência Artificial (IA) e Machine Learning.

Como observa Brynjolfsson e McAfee (2014), o aumento exponencial dos dados alimenta a revolução tecnológica, que, por sua vez, cria formas de negócios e produtos baseados em dados. A evolução desses campos de estudo e a sua aplicação prática têm proporcionado novas soluções para análise de dados em tempo real, previsões mais precisas e uma maior automação de processos.

Destaca-se que o setor tecnológico também foi desafiado a criar arquiteturas de armazenamento e processamento para atender à demanda crescente por dados.

O surgimento de cloud computing e o desenvolvimento de sistemas distribuídos tornaram possível o armazenamento e processamento em larga escala, viabilizando a gestão eficiente de Big Data.

Essas inovações permitiram que pequenas e grandes empresas tivessem acesso a recursos de TI avançados sem a necessidade de infraestrutura própria. De acordo com Marz e Warren (2015), as tecnologias de Big Data evoluíram de forma a suportar o processamento em larga escala e em tempo real, o que era antes impensável com as tecnologias tradicionais.

3.3. Visão futura sobre a expansão contínua dos dados

O futuro da expansão dos dados é repleto de inovações e desafios. A tendência de crescimento contínuo no volume de dados deverá se intensificar, com a proliferação de dispositivos conectados e o avanço da tecnologia de sensores. Como prevê Chen et al. (2014), mencionando que a quantidade de dados gerados deve continuar a aumentar à medida que novas fontes de dados, como os dispositivos vestíveis e a Internet das Coisas, se expandem.

Além disso, a inteligência artificial e o aprendizado de máquina desempenharão um papel ainda mais importante na análise de grandes volumes de dados, permitindo uma análise mais rápida e precisa. A evolução das tecnologias de IA permitirá a construção de sistemas preditivos e prescritivos mais robustos, capazes de identificar padrões e tomar decisões de forma autônoma.

Portanto, a ética e a regulamentação dos dados se tornarão cada vez mais centrais à medida que a coleta e o uso de dados aumentam. Espera-se que governantes, organizações e reguladores desenvolvam normas e frameworks legais para garantir que a privacidade e os direitos dos indivíduos sejam protegidos e reforcem regulamentos existentes a fim de garantir que a segurança dos usuários acompanhe as mudanças e avanços tecnológicos.

Por fim, A transparência e a confiança se tornarão elementos cruciais na utilização e manipulação de informações pois torna-se necessário, a atenção a políticas de grandes empresas que almejam informações dos seus usuários. Como afirmado por Mayer-Schönberger e Cukier (2013), o futuro da expansão dos dados dependerá de como a sociedade gerenciará as questões éticas e jurídicas que surgem com o uso massivo de dados pessoais.

4. METODOLOGIA

O Projeto de Otimização de Consultas de Banco de Dados através da Implementação de um Sistema de Estruturação e Hierarquização usando I.A aborda de forma analítica, dados de revistas cientificas eletrônicas, artigos, relatórios de empresas e estudos de caso, relacionando estes conhecimentos e usando-os como base solida e flexível para propor soluções eficazes.

Para a realização desta pesquisa, fez-se uso de dispositivos eletrônicos conectados à internet, no qual foram utilizados para buscar informações comprobatórias sobre as necessidades de gerenciamento e gestão de dados, constatando um problema em crescente escala que determina o futuro do tratamento e coleta de dados.

A partir das informações reunidas, foi abordado, no contexto específico, os crescentes desafios relacionados a proliferação de dados na rede, partindo deste, foram analisados os seguintes aspectos: impactos na confiabilidade de resultados, aumento da complexidade de tarefas, diminuição da qualidade das informações, filtragem ineficiente de informações e aumento da utilização de recursos de hardware. Além destes, o autor fez uso de conhecimentos prévios para melhor análise do contexto abordado afim de garantir a veracidade e coerência de cada trecho descrito.

4.1. Método de pesquisa

Caracterizando-se como um método de pesquisa aplicada, o projeto visa desenvolver uma solução prática voltada à otimização de consultas de banco de dados por meio da utilização de técnicas de Inteligência Artificial. A pesquisa busca identificar dificuldades e solucionar problemas relacionados à estruturação e filtragem de grandes volumes de dados, visando aumentar a eficiência e assertividade de informações estruturadas.

Conforme explicado por GIL (2019, p.25):

> “A pesquisa aplicada, por sua vez, apresenta muitos pontos de contato com a pesquisa pura, pois depende de suas descobertas e se enriquece com o seu desenvolvimento; todavia, tem como característica fundamental o interesse na aplicação, utilização e consequências práticas dos conhecimentos.”

Para a aplicação e análise da ferramenta desenvolvida, será usada a

abordagem quantitativa para registrar o histórico de tratamento de dados realizados e demonstrar através de métricas de assertividade e desempenho, os resultados que comprovam a diminuição dos recursos computacionais ao realizar consultas de banco de dados.

Além disso, para a elaboração do estudo sobre a viabilidade e aplicação da ferramenta, foram utilizados artigos digitais de relevância para o tema, livros relacionados ao processo de tratamento de dados e análise de informações bem como registros históricos que desempenharam papeis fundamentais para a criação do tema proposto.

4.2. Método de desenvolvimento do projeto – metodologia cascata

Para o desenvolvimento deste projeto, foram avaliados aspectos como tipo de sistema a ser desenvolvido, cenários de aplicação e como ele seria desenvolvido. Para tal, foi-se escolhido a metodologia cascata devido a mesma estar alinhada ao processo de desenvolvimento pretendido.

Sendo assim, a elaboração do projeto do sistema consiste-se em documentar características essenciais a serem seguidas e executadas durante o processo de desenvolvimento da ferramenta, conforme descrito por Pressman (2021, p. 502): “O gerenciamento de projeto de software é uma atividade de apoio da engenharia de software. Ele inicia antes de qualquer atividade técnica e prossegue ao longo da modelagem, construção e utilização do software.”, ou seja, nessa etapa são descritas como as informações serão utilizadas, como protótipo definirá as telas e como os diagramas de funcionamento explicaram a interação entre o sistema e o usuário.

Para a implementação da ferramenta, serão utilizados sistemas computacionais com o sistema operacional Windows 11 que atendam aos requisitos exigidos para garantir estabilidade visando executar a aplicação em um espaço controlado e com dados específicos para realizar o tratamento de dados e demonstrar a eficácia de utilização.

Quanto aos testes e validação, Pressman (2021, p.373) descreve: “Teste é um conjunto de atividades que podem ser planejadas com antecedência e executadas sistematicamente.”. Sendo assim, durante o processo serão executadas tarefas especificas voltadas a comprovar a eficácia de cada etapa do tratamento de dados através de análises variadas como tempo de resposta, resultados esperados,

eficiência e assertividade do tratamento e assim validar essas informações com os resultados obtidos.

Por fim, na manutenção e evolução da aplicação, serão revisados trechos dos códigos escritos nas ferramentas de desenvolvimento para corrigir possíveis erros ou falhas de execução durante o uso da ferramenta e prospectar através das necessidades do mercado, possíveis melhorias que seriam relevantes para a aplicação tornar-se mais útil e competitiva.

4.3. Planejamento e análise de viabilidade

No planejamento de desenvolvimento, foram avaliados aspectos pessoais e acadêmicos para analisar o cenário e viabilizar os parâmetros que se encaixem no projeto, além disso, conforme descrito por Cavalcanti (2016, p. 83): “Sobre a quantidade de recursos a alocar, devem-se considerar restrições realistas de quanto recurso pode-se dispor, bem como restrições lógicas quanto a como organizar a execução da atividade com um número viável de recursos”

Durante a análise da viabilidade técnica, identificou-se a necessidade de um computador com acesso à internet, contendo no mínimo 10 GB de armazenamento e 4 GB de RAM, para garantir o funcionamento adequado. Utilizou-se ferramentas de Inteligência Artificial para o processamento dos dados. Quanto à equipe, apenas uma pessoa foi responsável pelo desenvolvimento e documentação do projeto.

Na análise de viabilidade econômica, identificou-se a necessidade de um investimento mensal de R$ 150,00 com serviços de internet e R$ 60,00 com ferramentas voltadas à documentação e formatação de texto.

Na análise de viabilidade operacional, não foram identificadas dificuldades de uso. No entanto, por se tratar de uma ferramenta voltada a profissionais da área, tornou-se necessária a criação de uma seção de ajuda, com um guia explicativo sobre cada etapa de utilização da aplicação.

Na análise de viabilidade temporal, estimou-se aproximadamente 12 dias para a coleta e organização dos dados, 3 dias para a definição dos requisitos funcionais e não funcionais, 12 dias para a modelagem do banco de dados, 1 mês para a criação dos demais diagramas e cerca de 5 meses para o desenvolvimento do projeto.

4.4. Garantia de qualidade

Com o objetivo de assegurar o pleno funcionamento da aplicação desenvolvida, foram realizados testes de aceitação, os quais visaram verificar se o sistema atendia adequadamente aos requisitos funcionais e não funcionais previamente estabelecidos. Complementarmente, foram aplicados testes de desempenho, com a finalidade de mensurar a eficácia da ferramenta, utilizando-se métricas relacionadas à assertividade dos resultados e ao consumo de recursos computacionais.

Segundo Cavalcanti (2016, p. 132):

> “Os conceitos-chave em qualidade apontam principalmente para o cumprimento dos requisitos do produto, na exata proporção em que eles foram planejados. Exceder esses requisitos pode desviar a atenção e os recursos do projeto para uma direção indesejada.”

A execução dos testes foi realizada por meio do framework PyTest, amplamente utilizado no ecossistema Python para automação de testes. Os resultados obtidos puderam ser acompanhados por meio da interface gráfica do Visual Studio Code, a qual viabiliza a visualização, depuração e análise dos testes de forma eficiente e estruturada.

Para o controle de versão e rastreamento das alterações realizadas no código-fonte, foi empregada a plataforma GitHub, a partir da criação de um repositório local devidamente sincronizado com um repositório remoto em nuvem. Essa estratégia assegurou a integridade dos arquivos, além de permitir o gerenciamento colaborativo e histórico de modificações do projeto.

4.5. Aspectos éticos

A elaboração e o desenvolvimento da aplicação respeitaram rigorosamente os princípios éticos exigidos na área de tecnologia da informação. Em conformidade com a Lei Geral de Proteção de Dados Lei nº 13.709/2018, todas as informações utilizadas no projeto foram tratadas com responsabilidade e segurança, assegurando a privacidade dos dados eventualmente manipulados.

No que se refere à manipulação de dados, destaca-se que as informações utilizadas para testes foram geradas artificialmente com o propósito exclusivo de validar a funcionalidade da aplicação respeitando o artigo 1º da resolução CNS nº 510,

de 7 de abril de 2016, em seu parágrafo segundo e quinto. Dessa forma, os dados empregados não contêm qualquer tipo de conteúdo sensível, nem tampouco possibilitam a identificação de indivíduos, assegurando, assim, o cumprimento dos princípios da privacidade e da anonimização, conforme preconizado pela Lei Geral de Proteção de Dados Lei nº 13.709/2018.

Ressalta-se que, para a execução dos testes, não houve envolvimento de participantes humanos. As validações foram realizadas com base nos resultados fornecidos pelas saídas do sistema, utilizando dados simulados e processados diretamente pela aplicação, sem a necessidade de coleta de informações reais ou pessoais.

No tocante à utilização de código-fonte aberto, o desenvolvimento da aplicação respeitou integralmente as diretrizes das licenças públicas estabelecidas, como as licenças MIT, GPL e Apache, ainda que não tenha havido a incorporação direta de códigos de terceiros. Todo o código empregado na construção da ferramenta é de autoria própria do desenvolvedor, assegurando o respeito à propriedade intelectual e às boas práticas de engenharia de software.

4.6. Critérios de inclusão e exclusão

A aplicação desenvolvida é voltada a usuários que possuem conhecimento sobre técnicas de extração, transformação e carregamento de dados (ETL) ou seja, profissionais que manipulam informações em grande escala durante a prestação de um serviço. Dessa forma, usuários que não sejam profissionais na área não terão acesso as funcionalidades da ferramenta afim de garantir a integridade e privacidade dos processos de manipulação de informação.

Para garantir o uso adequado da ferramenta, foram definidos critérios de inclusão, abrangendo usuários que compreendem português ou inglês técnico, que concordem com os termos de uso e contrato de licença, que possuam a expertise necessária para o pleno manuseio da ferramenta e que utilizem sistemas operacionais compatíveis com os requisitos técnicos exigidos.

No que diz respeito aos critérios de exclusão, determinou-se que usuários sem conhecimento técnico não terão acesso às funcionalidades da ferramenta. Além disso, para atualizações e manutenção, será disponibilizado um pacote de atualização, sendo vedada a modificação do código-fonte por usuários não autorizados.

5. RESULTADOS E DISCUSSÕES

A análise dos resultados foi estruturada em etapas para evidenciar a correspondência entre os requisitos definidos e o desempenho da aplicação desenvolvida. A etapa inicial abordou a análise de requisitos, baseada em literatura especializada e práticas de grandes empresas como Google, Amazon e Meta. Essa fase permitiu identificar com precisão os requisitos funcionais e não funcionais essenciais, alinhando o desenvolvimento às reais necessidades do usuário.

Após a definição dos requisitos, foram detalhadas funcionalidades do sistema e foram descritos aspectos não funcionais, como privacidade, leveza e simplicidade visual. O sistema foi desenvolvido em Python utilizando uma interface gráfica e ferramentas como Visual Studio Code, GitHub e Figma. Além disso, os testes demonstraram que a aplicação cumpre satisfatoriamente os objetivos propostos, sendo eficaz no tratamento de dados estruturados e não estruturados.

5.1. Análise de requisitos

Para a realização da análise de requisitos da aplicação, foram realizadas consultas em livros relacionados e pesquisas envolvendo empresas que manipulam grandes volumes de dados como o Google, Amazon e Meta a fim de verificar as necessidades atuais do mercado e coletar informações para a montagem dos requisitos funcionais e não funcionais.

Segundo Glinz (2014, p. 9), uma análise de requisitos consiste-se em:

> “1. Aceitação: processo pelo qual se verifica se um sistema cumpre todos os requisitos especificados;
> 2. Teste de aceitação: procedimento destinado a comprovar que os requisitos estabelecidos foram devidamente implementados”

5.1.1. Requisitos funcionais

Registro e login de usuários: A aplicação desenvolvida contempla um sistema de registro e login de usuários, cuja funcionalidade visa garantir a segurança e a confiabilidade no acesso. Por meio desse recurso, busca-se impedir acessos não autorizados, assegurando que apenas usuários cadastrados possam manipular os

dados e funcionalidades da plataforma. Esse controle de acesso reforça a integridade do sistema e estabelece uma base segura para o tratamento das informações.

Importação e Exportação de dados processados e não processados: Permite importação e exportação de dados, tanto processados quanto não processados. A funcionalidade possibilita que o usuário importe arquivos no formato .xlsx, facilitando o carregamento de bases de dados para análise. Após o tratamento das informações, a aplicação possibilita a exportação dos dados resultantes, favorecendo a interoperabilidade e o reaproveitamento das informações em diferentes contextos de uso.

Recurso de tratamento de informações: incorpora um recurso robusto de tratamento de informações, cujo funcionamento se dá por meio de funções implementadas em linguagem Python. Com esse recurso, os usuários podem identificar e separar informações inconsistentes ou redundantes, realizando a limpeza e a filtragem necessárias para garantir a qualidade dos dados analisados. Esse processo é fundamental para a geração de resultados mais confiáveis e úteis.

Injeção e carregamento de códigos SQL: Possibilita a conexão direta com bancos de dados externos, permitindo que as informações já tratadas e estruturadas sejam inseridas automaticamente. Dessa forma, o usuário tem a flexibilidade de utilizar a aplicação tanto para análise local quanto para integração com sistemas de armazenamento e consulta de dados em tempo real.

5.1.2. Requisitos não funcionais

Privacidade e Autenticidade de informações: Garante a privacidade das informações, sendo projetada para funcionar de forma offline, diminuindo a necessidade de conexão com a internet afim de reduzir a exposição dos dados e garantindo maior controle sobre o ambiente de execução. Além disso, possui métodos de proteção e validação através das funções na linguagem Python e da ferramenta Pandas para atestar que as informações mantiveram sua integridade e autenticidade durante a realização e finalização do processo de tratamento.

Armazenamento seguro: Garante a segurança das informações ao optar por uma estratégia baseada em cache temporário, onde os dados são carregados, tratados e, após sua exportação ou salvamento, são automaticamente excluídos da memória. Essa prática evita o acúmulo de informações sensíveis no dispositivo e

diminui o risco de vazamento de dados, promovendo um tratamento mais seguro e consciente da informação.

Interface minimalista e objetiva: Possui uma interface elaborada de forma minimalista e objetiva. O design simplificado tem como objetivo proporcionar uma experiência de uso intuitiva, reduzindo a curva de aprendizado e evitando sobrecarga cognitiva. A escolha das cores e a organização dos elementos visuais também contribuem para uma interação mais fluida e eficiente com a aplicação.

Garantia de integridade de arquivos: Possui ferramentas internas que validam os formatos de entrada e impedem o carregamento de arquivos inválidos. Além disso, há um registro de eventos automatizado que atua como sistema de auditoria, registrando possíveis anomalias durante o salvamento ou criação de novos arquivos, permitindo o rastreamento de falhas e a correção preventiva de problemas.

Baixo uso de recursos computacionais: Projetado para operar com baixo consumo de recursos computacionais. Mesmo sendo capaz de processar grandes volumes de dados, a ferramenta utiliza apenas os recursos essenciais para seu funcionamento. O desempenho da aplicação é ajustado automaticamente conforme a complexidade e o tamanho dos dados processados, garantindo eficiência sem comprometer a estabilidade do sistema.

5.2. Funcionalidades da ferramenta

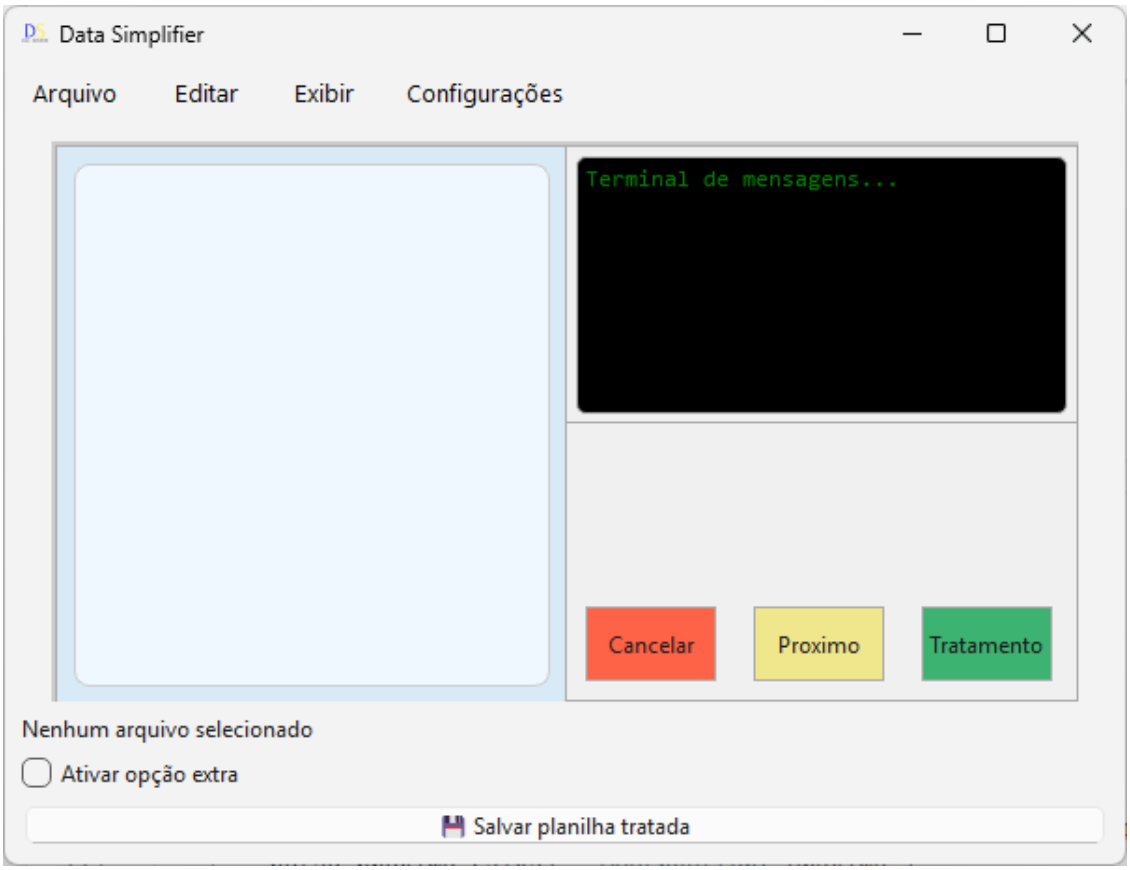
A ferramenta denominada Data Simplifier foi desenvolvida com o objetivo de oferecer funcionalidades voltadas para os processos de extração, transformação e carregamento de dados (ETL). Esses recursos são fundamentais para a limpeza e estruturação de informações, otimizando o trabalho de profissionais que atuam com análise e tratamento de dados.

A utilização da ferramenta visa facilitar a manipulação de grandes volumes de dados, promovendo maior eficiência e precisão nos processos. Ao automatizar etapas repetitivas e suscetíveis a erros humanos, o Data Simplifier contribui significativamente para a melhoria da qualidade dos dados utilizados em projetos diversos.

Sendo assim, sobre as funcionalidades da ferramenta, no menu superior contém ferramentas essenciais como Arquivo, responsável por criar e carregar um novo arquivo, editar, responsável por realizar alterações e reversões no documento,

exibir, responsável por permitir diferentes modos de visualização e configurações, responsável por permitir personalização de tema e ajustes de tratamento.

Figura 1 - Tela Inicial



Fonte: Elaborado pelo autor (2025).

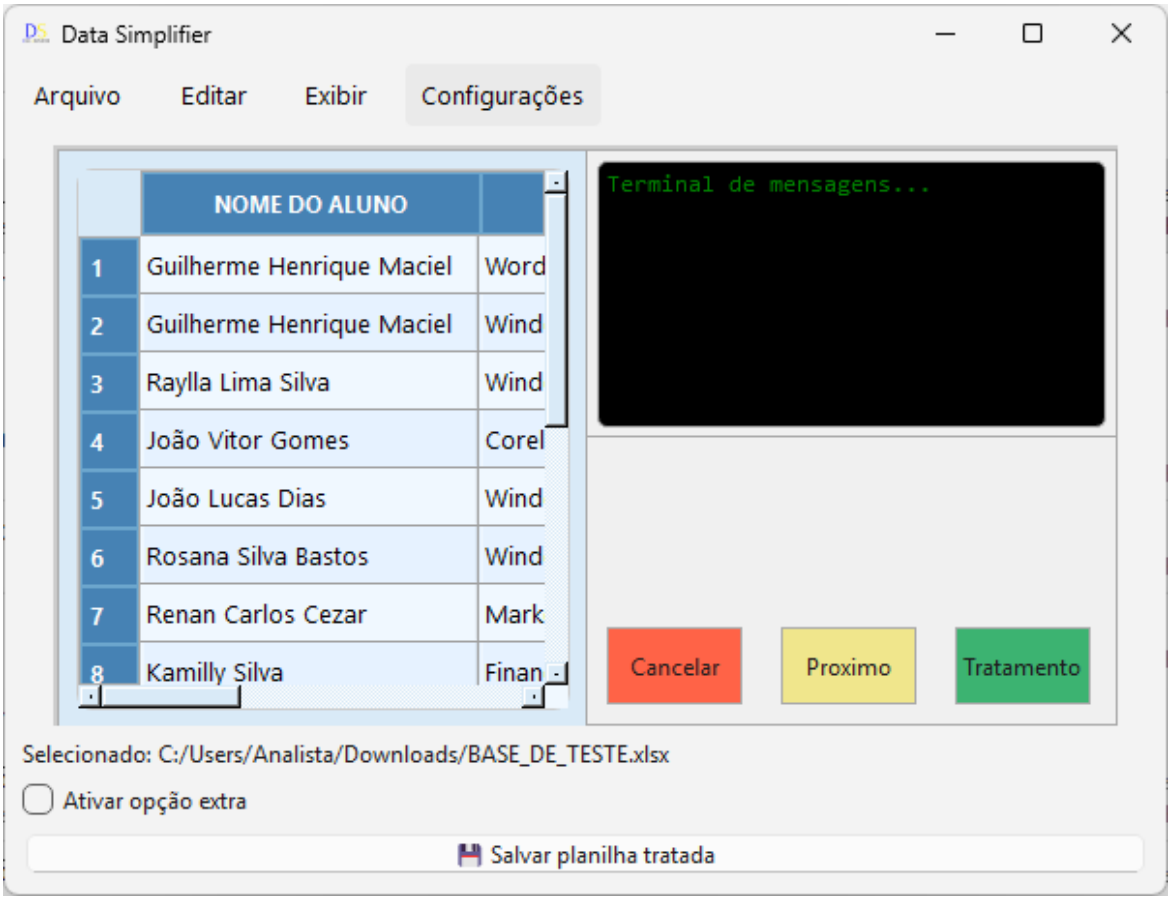
Na tela inicial da aplicação, conforme ilustrado na Figura 1, são apresentados três botões principais que desempenham funções essenciais no controle do processo de tratamento de dados. Esses botões foram projetados para oferecer ao usuário uma navegação intuitiva e eficiente durante a execução das tarefas.

O primeiro botão, denominado **“**Cancelar**”**, tem como finalidade interromper qualquer operação em andamento. O segundo botão, identificado como “Próximo”, permite ao usuário avançar para a próxima etapa do processo de análise da base de dados. O terceiro botão, chamado “Tratamento”, é responsável por iniciar o processo de limpeza das informações inconsistentes presentes na base de dados. Além desses três botões, a interface também disponibiliza um botão adicional destinado ao salvamento da planilha.

Após escolher um arquivo para tratamento, o usuário será levado a próxima

tela demonstrada na figura 2, no qual o arquivo selecionado é carregado em cache e convertido para um Data Frame no Pandas, que permite que as informações sejam repassadas para o sistema possa realizar verificações de validação de nomes e estruturação de valores.

Figura 2 - Tela de Carregamento



Fonte: Elabora pelo autor (2025).

Após o carregamento da base de dados no DataFrame da biblioteca Pandas, o usuário deve acionar o botão “Próximo” para dar continuidade ao processo de tratamento. Essa ação é responsável por iniciar a etapa de verificação das inconsistências presentes nos dados. A partir da ativação desse comando, o sistema executa uma função específica que realiza a identificação automática de inconsistências no DataFrame.

Uma vez identificadas, as inconsistências são separadas do DataFrame principal. Essa separação é realizada de forma automatizada, com o objetivo de isolar os dados problemáticos e evitar que interfiram nas demais etapas do processo.

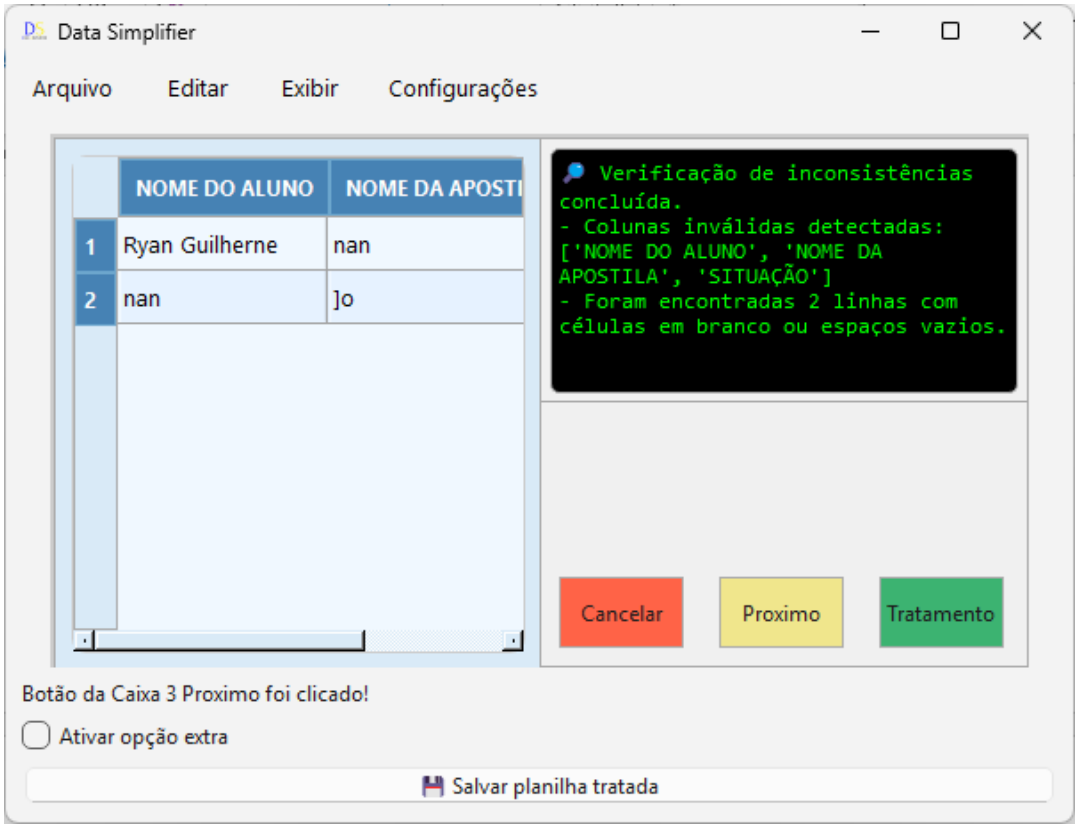
Os dados inconsistentes são então armazenados em um novo **DataFrame** do Pandas, que é exibido separadamente ao usuário. Essa visualização paralela permite uma análise mais detalhada dos problemas encontrados, facilitando a tomada de decisões quanto ao tratamento adequado.

A exibição dos dados inconsistentes em um DataFrame distinto também contribui para a rastreabilidade do processo, permitindo que o usuário acompanhe quais registros foram considerados inválidos e por quais motivos.

Essa abordagem modular, que separa os dados válidos dos inconsistentes, proporciona maior controle e transparência no processo de limpeza. Além disso, permite que o usuário realize correções manuais ou automatizadas com base nas informações apresentadas.

Conforme ilustrado na Figura 3, essa etapa é fundamental para garantir que apenas dados confiáveis e devidamente estruturados sejam mantidos no DataFrame principal, assegurando a integridade das análises e resultados gerados pela ferramenta.

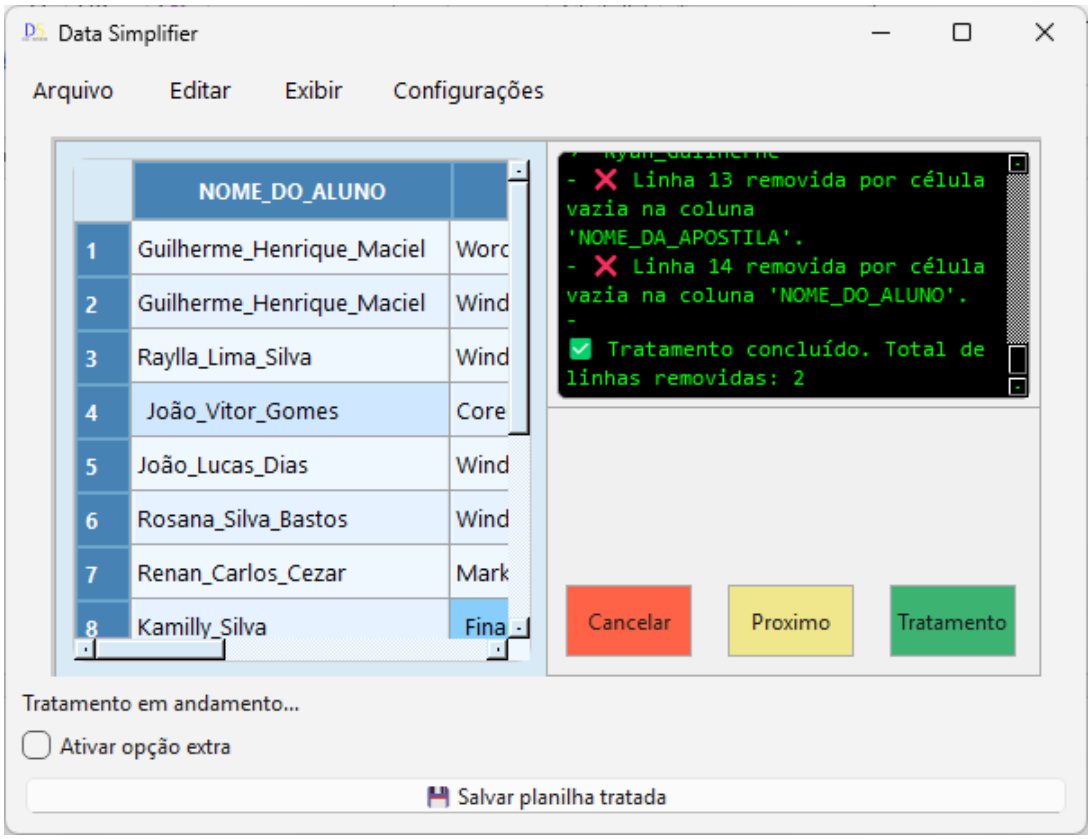
Figura 3 - Tela de Inconsistências



Fonte: Elabora pelo autor (2025).

Em seguida, para realizar o tratamento e estruturação das informações contidas no Data Frame, o usuário deverá selecionar o botão tratamento e a partir dessa ação, será chamado a função para tratar as inconsistências no qual substituirá espaços em branco e removerá linhas irregulares do Data Frame, tornando-o estruturado e consistente, conforme demostrado na figura 4.

Figura 4 - Tela de Tratamento



Fonte: Elaborado pelo autor (2025).

5.3. Tecnologias utilizadas

Na escolha das tecnologias utilizadas, foram realizadas análises de compatibilidade entre as linguagens de programação e a integração de outras ferramentas como o banco de dados, resultando na utilização do Python por possuir maior flexibilidade em aplicações que utilizam Inteligência Artificial e possuir uma vasta biblioteca de ferramentas de extração, transformação e carregamento de dados (ETL).

Além disso, na escolha do banco de dados, o PostgreSQL demonstrou maior

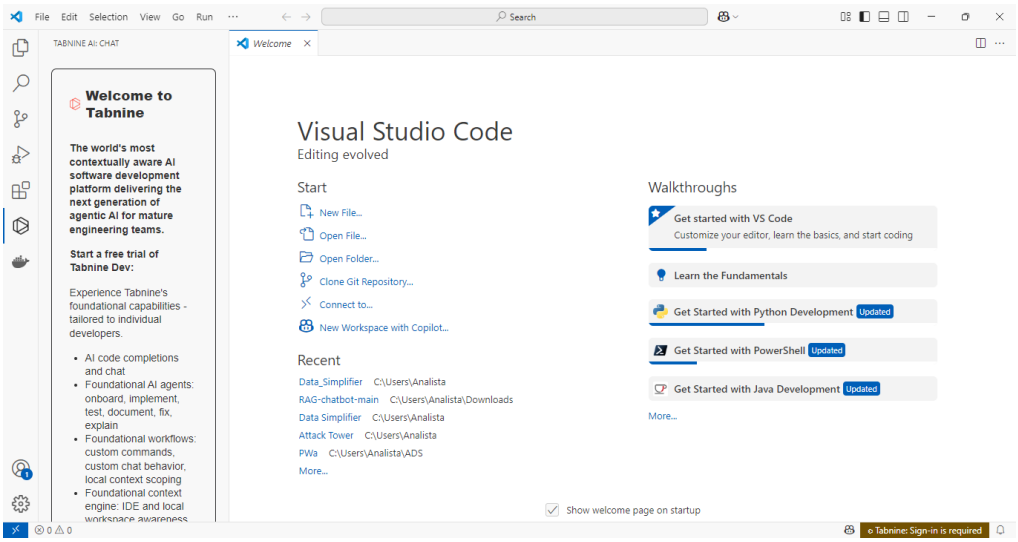
compatibilidade com os requisitos propostos pelo projeto e com o sistema utilizado para desenvolvimento, permitindo que a conexão entre a aplicação ocorresse de forma rápida e eficiente, garantindo o armazenamento seguro das informações dos usuários cadastrados.

Sendo assim, com o intuito de alinhar a interface gráfica aos parâmetros definidos previamente no protótipo da aplicação, tornou-se necessário o uso de uma biblioteca que permitisse a construção de uma interface eficiente, estruturada e visualmente coerente com os objetivos do projeto. Diante disso, optou-se pela utilização da biblioteca Python Qt, uma das mais consolidadas ferramentas para o desenvolvimento de interfaces gráficas em Python.

5.4. Plataforma de desenvolvimento

Para o desenvolvimento do projeto, foi escolhido a IDE Visual Studio Code ilustrada na Figura 5, permitindo que o trabalho fosse realizado de forma organizada além de fornecer integrações com ferramentas como a Tabnine, responsável por sugerir códigos prontos e Git, responsável por permitir o versionamento de código de forma simplificada.

Figura 5 - Visual Studio Code

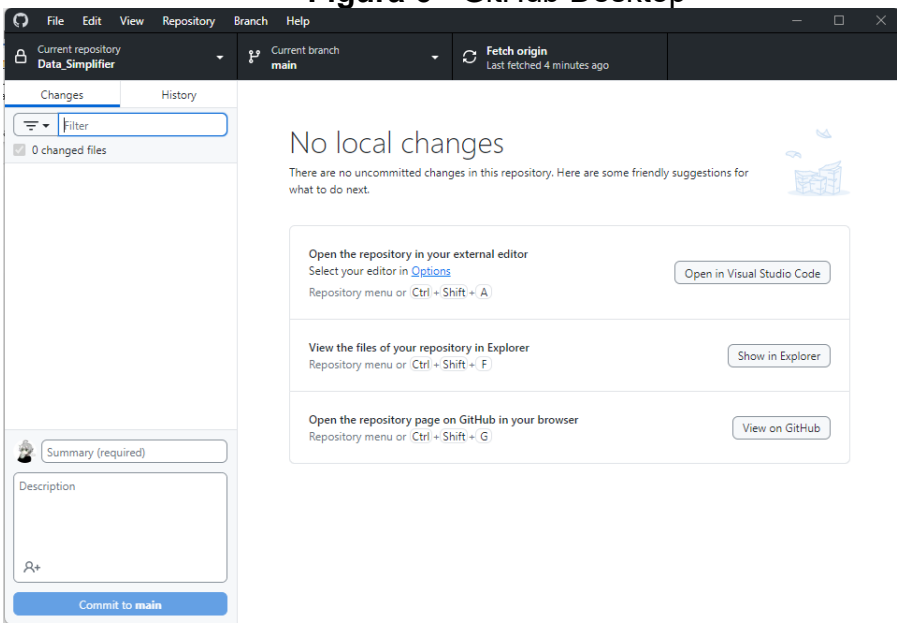


Fonte: Elaborado pelo autor (2025).

Com o intuito de garantir um controle eficiente das versões do ambiente de desenvolvimento do projeto, optou-se pela utilização da plataforma GitHub, conforme

ilustrado na Figura 6. Tal escolha justifica-se pelas funcionalidades robustas que a ferramenta disponibiliza, especialmente no que se refere ao versionamento de código e à colaboração entre desenvolvedores.

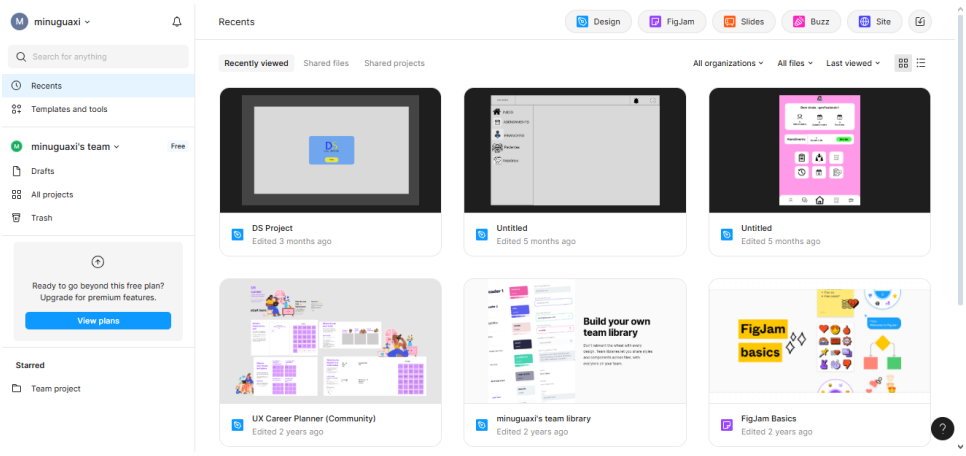
Figura 6 - GitHub Desktop



Fonte: Elaborado pelo autor (2025).

Além disso, para o desenvolvimento do protótipo prévio do projeto, foi utilizado a ferramenta Figma, ilustrada na figura 7, para criar o design de interface e experiência do usuário permitindo a projeção e montagem de todas as telas do projeto visando orientar a construção do design garantindo que o desenvolvimento possa cumprir com os requisitos definidos para o projeto.

Figura 7 - Figma



Fonte: Elaborado pelo autor (2025).

Para a hospedagem e execução da aplicação foi escolhido usar a conexão local para permitir o funcionamento sem a necessidade de uma conexão com a internet e o servidor local para garantir a segurança das informações ao guardar dados temporariamente apenas para permitir a identificação da base de dados pela ferramenta do Pandas e garantir estabilidade de uso ao otimizar o uso de recursos de hardware.

Tratando-se do sistema operacional escolhido, o Windows 11 mostrou-se o melhor candidato pela sua facilidade de adaptação as bibliotecas da linguagem Python e flexibilidade de operações envolvendo manipulação de informações de banco de dados variados como PostgreSQL e MySQL e codificação intuitiva durante o desenvolvimento do projeto, permitindo de forma simplista, a configuração de variáveis de ambiente e depuração de código

5.5. Visão geral da ferramenta

O sistema desenvolvido apresenta funções voltadas à análise de informações e ao tratamento de dados, com foco na identificação e correção de inconsistências. Seu principal objetivo é transformar conjuntos de dados desorganizados em informações estruturadas, permitindo que sejam utilizadas em processos de consulta por sistemas de inteligência artificial ou bancos de dados relacionais.

Dessa forma, a função de carregamento de arquivos possibilita ao usuário, a seleção de planilhas no formato .xlsx diretamente do dispositivo. Depois da seleção, os dados são carregados e convertidos para um Data Frame da biblioteca Pandas, que atua como estrutura de base de dados temporária. Essa base serve como suporte para as demais operações de análise, garantindo desempenho e flexibilidade no manuseio dos dados.

Após o carregamento, a aplicação disponibiliza uma função de identificação de inconsistências, responsável por realizar uma varredura em todas as linhas e colunas da planilha. Essa análise detecta problemas como espaços em branco, campos vazios, valores repetidos ou informações sem utilidade aparente. Após a execução do processo, é retornada uma nova base de dados contendo as irregularidades encontradas.

Posteriormente, a função de tratamento pode ser acionada para realizar a limpeza dos dados identificados como inconsistentes. Essa etapa realiza a filtragem,

padronização e eliminação de entradas inválidas, gerando um novo conjunto de dados estruturados, prontos para serem manipulados por profissionais da área ou utilizado como insumo em outros sistemas. O resultado desse processo é uma planilha com dados otimizados e livres de problemas que poderiam comprometer a análise.

Por fim, a funcionalidade de salvamento e exportação oferece diversas possibilidades ao usuário. É possível exportar os dados tratados em formato .xlsx, em um arquivo de código SQL contendo comandos para inserção em bases relacionais, ou ainda, realizar a injeção direta das informações estruturadas em um banco de dados de sua preferência, por meio de operações CRUD (Create, Read, Update, Delete).

Destaca-se que o público-alvo da aplicação são profissionais atuantes nas áreas de Ciência de Dados e Análise de Dados, especialmente aqueles vinculados a grandes empresas de tecnologia (bigtechs) e centros de processamento de dados (Data Centers). A ferramenta foi projetada para atender às demandas específicas desse segmento, oferecendo recursos técnicos compatíveis com as rotinas de tratamento de grandes volumes de informação e integração com sistemas de Inteligência Artificial.

A aplicação Data Simplifier apresenta-se como uma solução inovadora frente às ferramentas convencionais disponíveis no mercado, distinguindo-se por oferecer funcionalidades ágeis e flexíveis voltadas à manipulação e ao tratamento de dados. Seu principal diferencial reside na priorização da privacidade das informações processadas e na facilidade de uso de recursos avançados, os quais são incorporados por meio de uma interface simplificada e intuitiva, sem comprometer a robustez das operações realizadas.

5.6. Testes e resultados

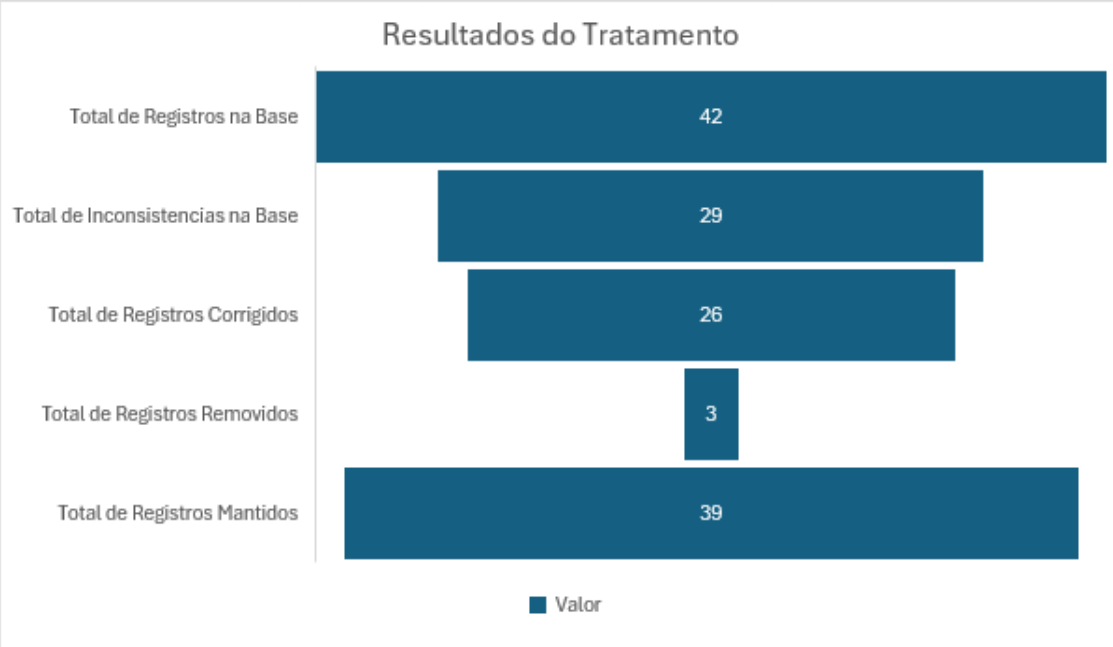
O presente projeto demonstrou em suas análises, a crescente complexidade na manipulação de grandes volumes de informações, caracterizando a falta de estruturação e confiabilidade dos dados, evidenciando a necessidade de soluções eficientes para otimizar a coleta, análise e extração de dados relevantes.

Visando mitigar esses desafios, o projeto propõe a implementação de um sistema de otimização de consultas através da estruturação e hierarquização de dados, utilizando inteligência artificial e os resultados obtidos por estudos recentes e

relatórios, demonstraram que a automatização de consultas, otimização de dados, a garantia de qualidade e integridade resultou na redução de custos operacionais e melhora no desempenho durante tomadas de decisão.

Ademais, cabe as empresas, a constante busca pela evolução tecnológica através do incentivo de uso das Startups e pesquisas para fomentar a divulgação e utilização de ferramentas, sendo a análise de dados, a principal ferramenta usada para verificar o mercado, antecipar tendencias e tomar decisões estrategicamente.

Gráfico 1 - Resultados do Tratamento



Fonte: Elaborado pelo autor (2025).

Conforme apresentado no Gráfico 1, durante os testes realizados em ambiente simulado, foram utilizados dados fictícios de uma base de dados no formato .xlsx com o objetivo de avaliar a eficácia da ferramenta de tratamento de inconsistências para demonstrar como e quais informações irregulares foram encontradas e quantificar a quantidade de informações que foram estruturadas.

A análise revelou a existência de 29 registros inconsistentes em um total de 40 entradas da base de dados. Sendo identificados como espaços em branco ou valores nulos. Desses, 26 foram devidamente corrigidos, enquanto 3 foram removidos por não atenderem aos critérios estabelecidos de estruturação da informação. Como resultado, obteve-se a preservação de 39 registros, todos com tratamento considerado bem-sucedido.

6. CONSIDERAÇÕES FINAIS

desenvolvimento da ferramenta Data Simplifier demonstrou-se uma solução eficaz e inovadora para o tratamento e análise de dados, especialmente em contextos que exigem alto desempenho, precisão e confiabilidade. A aplicação foi projetada com foco na identificação e correção de inconsistências em bases de dados, promovendo a transformação de informações desestruturadas em conjuntos organizados e prontos para uso em sistemas de inteligência artificial ou bancos de dados relacionais.

A estrutura modular da ferramenta, aliada à utilização de bibliotecas consolidadas como o Pandas, permitiu a criação de um ambiente robusto e flexível, capaz de atender às demandas de profissionais da área de Ciência de Dados. A interface intuitiva e os recursos de automação contribuíram para a simplificação de processos complexos, tornando o sistema acessível mesmo para usuários com menor familiaridade técnica.

Os testes realizados evidenciaram a capacidade da ferramenta em lidar com grandes volumes de dados, identificando e tratando inconsistências de forma automatizada. Os resultados obtidos indicam uma melhora significativa na qualidade dos dados processados, refletindo diretamente na eficiência das análises e na tomada de decisões estratégicas. Além disso, a possibilidade de exportação dos dados tratados em diferentes formatos amplia as possibilidades de integração com outros sistemas, tornando a ferramenta versátil e adaptável a diversos cenários corporativos e acadêmicos.

O projeto também destaca a importância da adoção de soluções tecnológicas inovadoras por parte das empresas, especialmente no que diz respeito à análise de dados como ferramenta estratégica. Dessa forma, conclui-se que a ferramenta desenvolvida atende aos objetivos propostos, oferecendo uma solução prática, eficiente e alinhada às necessidades do mercado atual. Sua aplicação pode contribuir significativamente para a melhoria da qualidade dos dados e para o fortalecimento da cultura de dados nas organizações.

Como trabalhos futuros, sugere-se a ampliação das funcionalidades da ferramenta, incluindo novos algoritmos de detecção de anomalias, integração com APIs externas e a implementação de recursos de aprendizado de máquina para aprimorar ainda mais a capacidade de análise e predição de dado

7. REFERÊNCIAS

DOMO. **Data Never Sleeps**. 5. Ed. Global, 2017. Disponível em: <https://www.domo.com/learn/infographic/data-never-sleeps-5/>.
Acesso em: 20 set. 2024.

Internet Live Stats. **Google Search Statistics**, 2024. Disponível em: <https://www.internetlivestats.com/google-search-statistics/#trend/>. Acesso em 20 set. 2024.

NEGROPONTE, Nicholas. **Being Digital**. 1º Ed. New York: Knopf, 1995. Disponível em: <https://archive.org/details/beingdigital00negr/page/n9/mode/2up/>. Acesso em 23 set. 2024.

ROSENBAUM, A. **Critical Capabilities for Cloud Database Management Systems for Operational Use Cases**. Gartner, 2024. Disponível em: <https://www.gartner.com/doc/reprints?id=1-2G6I0Y3B&ct=240108&st=sb/>. Acesso em 23 set. 2024.

IBV, IBM. **2024 CEO Study - 6 hard truths CEOs must face**. Global: 29º Ed, 2023. Disponível em:
<https://www.ibm.com/downloads/documents/us-en/10a99803fc2fdcb7/>. Acesso em 23 set. 2024.

EUCLIDES. Os Elementos. 1º Ed. Alexandria: 300 a. C. Disponível em: <https://pt.wikipedia.org/wiki/Os_Elementos/>. Acesso em: 23 set. 2024.

G. J. TOOMER. **The Chord Table of Hipparchus and the Early History of Greek Trigonometry**. 18. Ed. Centaurus: 1974. Disponível em:
<https://onlinelibrary.wiley.com/doi/10.1111/j.1600-0498.1974.tb00205.x/>. Acesso em: 23 set. 2024.

WIKIPÉDIA, **Cartão perfurado**, 2025. Disponível em:
<https://pt.wikipedia.org/w/index.php?title=Herman_Hollerith&oldid=69666674/>.
Acesso em 23 set. 2024.

WIKIPÉDIA, **A prensa de tipos móveis**, 2025. Disponível em: <https://pt.wikipedia.org/wiki/Prensa_móvel#cite_note-1/>. Acesso em 23 set. 2024.

SHANNON, Claude Elwood. **A Mathematical Theory of Communication**. 1 Ed. Vol. 27. Urbana-Champaign: University of Illinois Press, 1949. p. 379–423, 623–656. Disponível em: <https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf/>. Acesso em 23 set. 2024.

FRENCH, C. S. **Oliver and Chapman's Data Processing and Information Technology**, 10 Ed. London: DP Publications, 1996. Disponível em: <https://archive.org/details/oliverchapmansda0010fren/>. Acesso em 23. set. 2024

MILANI, Alessandra M P.; SOARES, Juliane A.; ANDRADE, Gabriella L.; et al. **Visualização de Dados**. Porto Alegre: SAGAH, 2020. *E-book.* p.28. ISBN 9786556900278. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9786556900278/>. Acesso em: 28 mai. 2025.

OLSEN, Wendy. **Coleta de dados**. Porto Alegre: Penso, 2015. *E-book.* p.144. ISBN 9788584290543. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788584290543/>. Acesso em: 28 mai. 2025.

VIDA, Edinilson da S.; ALVES, Nicolli S R.; FERREIRA, Rafael G C.; et al. **Data Warehouse**. Porto Alegre: SAGAH, 2021. *E-book.* p.166. ISBN 9786556901916. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9786556901916/>. Acesso em: 28 mai. 2025.

Han, Jiawei.; Kamber, Micheline.; Pei, Jian. **Data Mining – Concepts and Technique**. 7. Ed. USA: Elsevier, 2012.p.38. Disponível em: <https://archive.org/details/the-morgan-kaufmann-series-in-data-management-

systems-jiawei-han-micheline-kambe/page/n37/mode/2up/>. Acesso em 28 mai. 2025.

CAVALCANTI, Francisco Rodrigo P.; SILVEIRA, Jarbas A N. **Fundamentos de Gestão de Projetos**. Rio de Janeiro: Atlas, 2016. E-book. p.83, 132. ISBN 9788597005622. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788597005622/>. Acesso em: 24 mai. 2025.

PRESSMAN, Roger S.; MAXIM, Bruce R. **Engenharia de software**. 9. ed. Porto Alegre: AMGH, 2021. E-book. p.502. ISBN 9786558040118. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9786558040118/>. Acesso em: 23 mai. 2025.

GIL, Antonio C. **Métodos e Técnicas de Pesquisa Social**, 7ª edição. Rio de Janeiro: Atlas, 2019. E-book. p.25. ISBN 9788597020991. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788597020991/>. Acesso em: 23 mai. 2025.

Glinz, Martin. **A Glossary of Requirements Engineering Terminology**, Zurich: University of Zurich, 2014. p.09. Disponível em: <https://www.gasq.org/files/content/gasq/downloads/certification/IREB/IREB%20FL/ireb_cpre_glossary_16_en.pdf/>. Acesso em 23 mai. 2025.

Silberschatz, Abraham.; Korth, F. Henry.; Sudarshan. **Database System Concepts**. 7. ed. New York: McGraw-Hill Education, 2020. Disponível em: <https://github.com/FrenzyExists/programming-books/blob/master/Databases/Database-System-Concepts-7th-Edition.pdf/>. Acesso em 23 mai. 2025.

VIEIRA, Diogo. **Qualidade de dados - Conceitos e importância**. Dateside,2023. Disponível em: <https://www.dataside.com.br/dataside-community/big-data/qualidade-de-dados-conceitos-e-importancia/>. Acesso em: 23 set. 2024.

Página de assinaturas

Felipe Ferreira
057.779.862-69
Signatário

Sara Carvalho
017.799.872-50
Signatário

Adriano Bollas
669.522.202-91
Signatário

Antonio Silva
032.290.192-88
Signatário

HISTÓRICO

11 jul 2025 20:06:07 — **Felipe Santos Ferreira** criou este documento. (Email: minuguaxi@gmail.com, CPF: 057.779.862-69)

11 jul 2025 20:06:10 — **Felipe Santos Ferreira** *(Email: minuguaxi@gmail.com, CPF: 057.779.862-69)* visualizou este documento por meio do IP 189.40.106.175 localizado em Belém - Pará - Brazil

11 jul 2025 20:06:49 — **Felipe Santos Ferreira** *(Email: minuguaxi@gmail.com, CPF: 057.779.862-69)* assinou este documento por meio do IP 189.40.106.175 localizado em Belém - Pará - Brazil

12 jul 2025 23:56:06 — **Adriano Louzada Bollas** *(Email: adriano.louzadabollas@gmail.com, CPF: 669.522.202-91)* visualizou este documento por meio do IP 200.124.94.192 localizado em Parauapebas - Pará - Brazil

12 jul 2025 23:56:48 — **Adriano Louzada Bollas** *(Email: adriano.louzadabollas@gmail.com, CPF: 669.522.202-91)* assinou este documento por meio do IP 200.124.94.192 localizado em Parauapebas - Pará - Brazil

14 jul 2025 23:00:43 — **Antonio Soares da Silva** *(Email: ads@fadesa.edu.br, CPF: 032.290.192-88)* visualizou este documento por meio do IP 45.7.26.146 localizado em Parauapebas - Pará - Brazil

14 jul 2025 23:00:47 — **Antonio Soares da Silva** *(Email: ads@fadesa.edu.br, CPF: 032.290.192-88)* assinou este documento por meio do IP 45.7.26.146 localizado em Parauapebas - Pará - Brazil

11 jul 2025 20:06:38 — **Sara Carvalho** *(Email: csaradeboracontato@gmail.com, CPF: 017.799.872-50)* visualizou este documento por meio do IP 186.232.206.163 localizado em Parauapebas - Pará - Brazil

11 jul 2025 20:06:53 **Sara Carvalho** *(Email: csaradeboracontato@gmail.com, CPF: 017.799.872-50)* assinou este documento por meio do IP 186.232.206.163 localizado em Parauapebas - Pará - Brazil