



UNIVERSIDADE ESTADUAL DO PIAUÍ
CENTRO DE TECNOLOGIA E URBANISMO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

Vitor Jarrel Castro de Sousa

**Uso de *Business Intelligence* para Visualização de
Indicadores de Evasão Escolar no Ensino Superior**

TERESINA

2025

Vitor Jarrel Castro de Sousa

Uso de *Business Intelligence* para Visualização de Indicadores de Evasão Escolar no Ensino Superior

Monografia de Trabalho de Conclusão de Curso apresentado na Universidade Estadual do Piauí – UESPI como parte dos requisitos para conclusão do Curso de Bacharelado em Ciência da Computação.

Orientador: Prof. Dr. Maurício Rêgo da Mota Rocha

TERESINA

2025

S725u Sousa, Vitor Jarrel Castro de.

Uso de Business Intelligence para visualização de indicadores de evasão escolar no ensino superior / Vitor Jarrel Castro de Sousa. - 2025.

61f.: il.

Monografia (graduação) - Universidade Estadual do Piauí-UESPI, Bacharelado em Ciência da Computação, Campus Poeta Torquato Neto, Teresina-PI, 2025.

"Orientador: Prof. Dr. Maurício Rêgo da Mota Rocha".

1. Business Intelligence. 2. Data Wharehouse. 3. Evasão escolar. 4. Visualização de dados. I. Rocha, Maurício Rêgo da Mota . II. Título.

CDD 004.07

Uso de *Business Intelligence* para Visualização de Indicadores de Evasão Escolar no Ensino Superior

Vitor Jarrel Castro de Sousa

Monografia de Trabalho de Conclusão de Curso apresentado na Universidade Estadual do Piauí – UESPI como parte dos requisitos para conclusão do Curso de Bacharelado em Ciência da Computação.

Prof. Dr. Maurício Rêgo da Mota Rocha,
Dsc.
Orientador

Nota da Banca Examinadora: 8,5

Banca Examinadora:

Prof. Dr. Maurício Rêgo da Mota Rocha,
Dsc.
Presidente

Dr^a. Lianna Mara Castro Duarte, Dsc.
Membro

Dr. Aldir Silva Sousa, Dsc.
Membro

*Este trabalho é dedicado a todas as pessoas que, em algum momento,
duvidaram da própria capacidade, mas carregam dentro de si um potencial imenso.*

AGRADECIMENTOS

Agradeço a Deus por me sustentar nos momentos de incerteza, me iluminar o caminho e me conceder coragem para seguir em frente, mesmo diante das dificuldades.

À minha mãe e ao meu pai, por serem minha base, por acreditarem em mim e por me incentivarem a seguir em frente nos estudos e na vida.

Ao meu grande amor, por sempre me apoiar nas decisões, por ser meu refúgio nesta reta final de curso e, principalmente, por estar ao meu lado em qualquer situação.

Ao professor Maurício Rêgo da Mota Rocha, por sua orientação dedicada como orientador deste trabalho e, especialmente, pela valiosa contribuição ao longo da minha formação acadêmica e profissional, sempre oferecendo apoio, paciência e direcionamento em cada etapa do processo.

Ao analista de informática do DTIC, José Alex Clímaco de Oliveira, pela ajuda fundamental na validação dos dados utilizados neste trabalho, contribuindo diretamente para a qualidade e consistência dos resultados apresentados.

“Não importa o quanto você bate, mas sim o quanto aguenta apanhar e continuar. O quanto pode suportar e seguir em frente. É assim que se ganha.”
(Stallone, Sylvester; Rocky Balboa, 2006)

RESUMO

A evasão escolar é um fenômeno que afeta diversos níveis e modalidades educacionais, caracterizando-se pelo abandono ou desistência do aluno em relação a determinada instituição ou modalidade de ensino. Diante disso, este trabalho teve como objetivo desenvolver uma solução de *Business Intelligence* para analisar os dados referentes à evasão no ensino superior da Universidade Estadual do Piauí (UESPI). Para tanto, adotou-se uma metodologia de natureza aplicada e descritiva, com abordagem quali-quantitativa. O método desenvolvido envolveu um processo de ETL e a modelagem de um esquema estrela. Os dados obtidos foram visualizados por meio de *dashboards* interativos e integrados em um Sistema Web, desenvolvidos no Looker Studio, uma ferramenta gratuita e intuitiva. Os resultados apresentaram expressivos números de evasão, com uma taxa geral de 43,08% em todos os campi da UESPI, tendo alcançado o pico no período de 2023.1. Observou-se ainda que a maioria dos alunos evadidos tem entre 25 e 29 anos e pertence ao sexo masculino. Conclui-se que é necessário implementar políticas públicas de permanência universitária, bem como adotar medidas que possibilitem a compreensão das causas subjetivas da evasão, visando à sua mitigação.

Palavras-chave: *Business Intelligence*. *Data Wharehouse*. Evasão escolar. Visualização de dados.

ABSTRACT

School dropout is a phenomenon that affects various educational levels and modalities, characterized by the student's abandonment or withdrawal from a particular institution or type of education. In this context, this study aimed to develop a Business Intelligence solution to analyze dropout data in higher education at the State University of Piauí (UESPI). To this end, an applied and descriptive methodology with a qualitative and quantitative approach was adopted. The method involved an ETL process and the modeling of a star schema. The resulting data were visualized through interactive dashboards, integrated into a web system developed using Looker Studio, a free and user-friendly tool. The results revealed significant dropout numbers, with a general rate of 43,08% across all UESPI campuses, peaking in the 2023.1 academic term. It was also observed that most dropout students are male and between 25 and 29 years old. It is concluded that it is necessary to implement public policies aimed at university retention, as well as adopt measures to understand the subjective causes of dropout in order to mitigate them.

Keywords: *Business Intelligence. Data Warehouse.* School dropout. Data visualization.

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo Estrela	36
Figura 2 – Status por Curso	37
Figura 3 – Evasão por Campus	38
Figura 4 – Contadores	38
Figura 5 – Dashboard 1	39
Figura 6 – Alunos Dividos por Sexo e Faixa Etária	39
Figura 7 – Evasão por Período Letivo	40
Figura 8 – Dashboard 2	40
Figura 9 – Tela de Login	41
Figura 10 – Status Geral por Curso e Campus	42
Figura 11 – Indicadores Demográficos e Temporais da Evasão e Conclusão	43
Figura 12 – Descrição da Tabela de Ocorrências	48

LISTA DE ABREVIATURAS E SIGLAS

BI	<i>Business Intelligence</i>
CA	Controle Acadêmico
DTIC	Diretoria de Tecnologia da Informação e Comunicação
DW	<i>Data Warehouse</i>
ETL	Extração, Tratamento e Carga
IDH	Índice de Desenvolvimento Humano
IES	Instituição de Ensino Superior
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
ONU	Organização das Nações Unidas
SIGAA	Sistema Integrado de Gestão de Atividades Acadêmicas

LISTA DE QUADROS

Quadro 1 – Tabelas extraídas do banco de dados CA (Controle Acadêmico) .	27
Quadro 2 – Tabelas extraídas do banco de dados SIGAA	27
Quadro 3 – Movimentações considerados na coleta da dimensão <i>Aluno</i> (SIGAA).	29
Quadro 4 – Ocorrências consideradas na coleta da dimensão <i>Aluno</i> (CA). . .	30
Quadro 5 – Estrutura da Tabela <i>dim_periodo</i>	31
Quadro 6 – Estrutura da Tabela <i>dim_aluno</i>	33
Quadro 7 – Estrutura da Tabela <i>dim_curso</i>	35
Quadro 8 – Estrutura da Tabela <i>fato_ocorrencia_aluno</i>	36

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Contextualização	15
1.2	Justificativa	16
1.3	Objetivos	17
1.3.1	Objetivo geral	17
1.3.2	Objetivos específicos	17
1.4	Metodologia	17
2	REFERENCIAL TEÓRICO	19
2.1	Evasão Escolar	19
2.2	<i>Business Intelligence</i> na Educação	21
2.2.1	<i>Data Warehouse</i>	22
2.2.2	<i>Dashboard</i> e visualização de dados	22
2.2.3	Google Looker Studio	23
2.3	Trabalhos Relacionados	23
2.3.1	<i>Um ambiente de Data Warehousing para apoiar a tomada de decisão quanto à evasão escolar na UFPE</i>	23
2.3.2	<i>Aplicação de uma solução de Business Intelligence para apoiar a Gestão Acadêmica em Universidades Federais Brasileiras</i>	24
3	DESENVOLVIMENTO	26
3.1	Processo ETL	26
3.1.1	Extração	26
3.1.1.1	Dimensão Aluno – Sistema SIGAA	28
3.1.1.2	Dimensão Aluno – Sistema CA	30
3.1.1.3	Dimensão Aluno – Modelo Estrela	30
3.1.1.4	Dimensão Curso - Modelo Estrela	30
3.1.1.5	Dimensão Período Letivo - Modelo Estrela	31
3.1.2	Limpeza dos Dados	31
3.1.3	Carga	32
3.1.3.1	Carga da Tabela <i>staging_aluno_ca</i>	32
3.1.3.2	Carga da Tabela <i>staging_aluno_sigaa</i>	33
3.1.3.3	Carga da Tabela <i>dim_aluno</i>	33
3.1.3.4	Carga da Tabela <i>dim_curso</i>	34
3.1.3.5	Carga da Tabela <i>dim_periodo</i>	35

3.1.3.6	Carga da Tabela <i>fato_ocorrencia_aluno</i>	35
3.2	<i>Business Intelligence</i>	37
3.2.1	Construção dos <i>Dashboards</i>	37
3.3	Integração ao Sistema Web	41
3.4	Validação	43
4	RESULTADOS	44
5	DISCUSSÃO	46
5.1	Alta Taxa de Evasão	46
5.2	Indicadores da Evasão	47
6	CONCLUSÕES	50
6.1	Contribuições	50
6.2	Limitações e Trabalhos Futuros	50
	REFERÊNCIAS	52
	APÊNDICE A – SCRIPTS SQL UTILIZADOS PARA O <i>DATA WHAREHOUSE</i>	55
	APÊNDICE B – COMANDOS PARA INSTALAÇÃO DO <i>LARAVEL BREEZE</i>	61

1 INTRODUÇÃO

Este capítulo apresenta a introdução à temática trabalhada na pesquisa: evasão escolar no ensino superior. O capítulo está dividido em quatro seções, abordando a contextualização do tema, bem como a justificativa, os objetivos e a metodologia adotada para a execução do trabalho.

1.1 Contextualização

O termo *evasão escolar*, de acordo com INEP (2017), trata do abandono ou desistência do aluno de qualquer instituição e/ou nível, desconsiderando os motivos. No ensino superior, a evasão escolar é um desafio significativo para o sistema educacional, impactando tanto o desenvolvimento dos alunos quanto a sociedade como um todo. Ocorre quando estudantes deixam de concluir seus programas de estudo, incluindo aqueles que, mesmo matriculados, desistem antes de iniciar o curso (Maia; Meirelles; Pela, 2004); incluindo discentes das redes pública privada, presencial e online em todo o Brasil, somou-se um índice de 57,2% de evasão escolar no primeiro semestre de 2024 (Correio Braziliense, 2024). Diante desse cenário, analisar e compreender os indicadores relacionados à evasão escolar é um dos primeiros passos para abordar essa questão (Alves; Fidalgo, 2021).

Visando compreender essas adversidades, Ambiel (2015) revisou a literatura, listando e categorizando os motivos que levam à evasão do ensino superior, findando em quatro agrupamentos de razões. As causas podem ser divididas em: infraestrutura da instituição e falta de serviços direcionados ao discente, motivos institucionais, motivos financeiros e/ou dificuldade em conciliar estudos e trabalho e motivos pessoais. A análise dessa problemática, contudo, precisa ser realizada de maneira ágil, para evitar futuras evasões por motivos que se assemelham. No entanto, essa inspeção não é feita fugazmente, haja vista que depende de diferentes setores, gerando lentidão no processo e dificultando a identificação de causas e o solucionamento do impasse (Alves; Fidalgo, 2021).

Para aprimorar e acelerar a análise e compreensão dos indicadores relacionados à problemática, é fundamental que os dados estejam bem estruturados e organizados. Isso pode ser feito com eficiência a partir de tecnologias de *Business Intelligence* (BI), especialmente por meio do *Data Warehouse* (DW), uma plataforma centralizada que armazena dados provenientes de diversas fontes, organizando-os de forma eficiente e garantindo sua integridade. O DW permite consolidar informações de diferentes

sistemas e fontes, criando um repositório único e acessível, no qual os dados podem ser analisados de maneira mais robusta e confiável (Olson; Lee, 2015).

Usando as informações consolidadas pelo DW, é possível criar *dashboards* completos, ilustrando as principais causas da evasão no ensino superior, bem como suas respectivas recorrências. Assim, é possível gerenciar índices e motivos e, com isso, trabalhar para reduzir e solucioná-las, de forma que possa abrandar, de maneira geral, a evasão escolar no ensino superior.

1.2 Justificativa

A evasão escolar no ensino superior, de certo, é um desafio notável tanto para as instituições de ensino, quanto para as políticas públicas educacionais, impactando não só a sustentabilidade financeira institucional, como as metas de formação de capital humano do país. Dessa maneira, afeta, de uma forma geral, o desenvolvimento do Brasil, uma vez que o grau de instrução é critério para o Índice de Desenvolvimento Humano (IDH). De acordo com a Organização das Nações Unidas (ONU), o Índice de Desenvolvimento Humano (IDH) brasileiro é de 0,760, conferindo-o a 89ª posição, de 193 países (Nations, 2024). Portanto, entender e avaliar os indicadores relacionados à essa problemática é necessário para se propor estratégias que promovam a permanência e o sucesso dos estudantes e, conseqüentemente, o desenvolvimento do país como um todo.

Nesse ínterim, o *Business Intelligence* surge como uma tecnologia promissora para solucionar o impasse, ao passo que transforma dados brutos em bases de dados organizadas e de simples visualização. O uso dos *dashboards*, por sua vez, permite consolidar esses dados organizados anteriormente em indicadores-chave, como taxas de permanência e evasão, desempenho acadêmico, motivações e montar um perfil socioeconômico. Com isso, a gerência acadêmica pode identificar padrões com maior rapidez e confiabilidade.

A justificativa do presente trabalho, portanto, está na relevância socioeducacional em abrandar a evasão escolar, utilizando a tecnologia como um facilitador do processo. O estudo contribui para a elevação da gestão educacional, promoção da inclusão e sucesso na ambiência dos estudantes, uma vez que utiliza *dashboards* fundamentados em BI para facilitar o acesso aos dados, além de proporcionar uma análise preditiva e eficiente para a proposição de estratégias interventivas.

1.3 Objetivos

A partir do contexto discutido, buscou-se traçar os objetivos gerais e específicos buscados no desenvolvimento do trabalho.

1.3.1 Objetivo geral

Diante do contexto apresentado, traçou-se como o objetivo geral do trabalho desenvolver uma solução de *Business Intelligence* para analisar os dados referentes à evasão no ensino superior da Universidade Estadual do Piauí (UESPI).

1.3.2 Objetivos específicos

Visando alcançar o objetivo geral pretendido, buscar-se-á atingir os seguintes objetivos específicos:

- Identificar os indicadores referentes à evasão do ensino superior;
- Criar um modelo de dados baseado no modelo estrela para organizar os dados no DW;
- Categorizar os indicadores da evasão escolar no ensino superior em *dashboards*, a partir do uso de *Business Intelligence*;
- Integrar os *dashboards* em um sistema web para facilitar o acesso aos indicadores;

1.4 Metodologia

Para alcançar os objetivos propostos, este trabalho adotou metodologia do tipo aplicada e descritiva, com abordagem quali-quantitativa. As etapas foram estruturadas de acordo com o ciclo de desenvolvimento de soluções de *Business Intelligence*:

- **ETAPA 1:** Revisão da literatura para levantamento de informações sobre a evasão escolar no ensino superior e o uso de *Business Intelligence* na problemática;
- **ETAPA 2:** Levantamento de requisitos e identificação das fontes de dados institucionais disponíveis na UESPI, incluindo os sistemas *SIGAA* e *CA*, com posterior definição dos dados relevantes para análise;
- **ETAPA 3:** Processo de ETL, com a extração de dados brutos, transformação (limpeza e padronização) e carga dos dados em um banco relacional;

- **ETAPA 4:** Etapa de modelagem dimensional, que envolve a estruturação do banco de dados em um modelo estrela (*Star Schema*), com a tabela fato (ocorrência) e dimensões (curso, período e aluno);
- **ETAPA 5:** Utilização do Looker Studio para a construção de *dashboards* interativos como ferramenta de visualização;
- **ETAPA 6:** Desenvolvimento de um sistema web utilizando o framework *Laravel*, com incorporação das visualizações por meio de *iframes* e controle de acesso autenticado;
- **ETAPA 7:** Verificação da consistência dos dados apresentados para validação da ferramenta;
- **ETAPA 8:** Análise dos indicadores gerados e identificação dos padrões dentro dos indicadores;

2 REFERENCIAL TEÓRICO

Este capítulo apresenta a fundamentação teórica do trabalho desenvolvido. O referencial teórico aborda os principais conceitos de evasão escolar, à luz de autores já consolidados, além dos conceitos de *Data Warehouse* e *Looker Studio*. Além disso, discute o uso de *Business Intelligence* no setor educacional e o uso de *Dashboards* para a visualização de dados. Por último, contempla os trabalhos relacionados, trazendo publicações acadêmicas com propostas semelhantes às do presente trabalho.

2.1 Evasão Escolar

A evasão escolar no ensino superior é uma problemática que ainda precisa ser discutida e alinhada, considerando que há certas discrepâncias quanto ao conceito. Partindo do clássico, Tinto (1975) baseia-se na obra *O Suicídio* (DURKHEIM, 2000), ao considerar a Instituição de Ensino Superior (IES) como um sistema social e comparar a evasão ao suicídio que acontece na sociedade. De acordo com Tinto, o indivíduo que não se integra socialmente também não se comprometerá adequadamente no meio universitário. Desde então, a evasão escolar do ensino superior vem sendo citada por diversos autores, que a relacionam, por exemplo, apenas ao ingresso seguido de desistência (Fritsch; Rocha; Vitelli, 2015). Outros, no entanto, consideram também os motivos, classificando-os em internos, quando ligados à instituição, e externos, quando associados ao discente (Silva et al., 2022). De maneira geral, a evasão é um fenômeno multifacetado, que pode ocorrer em diferentes modalidades de ensino, com discentes de cursos, idades e localidades variadas e de diferentes classes econômicas.

Entre os anos de 1994 e 2014, a educação superior brasileira passou por consideráveis expansões, no que se diz respeito aos números de instituições, cursos, vagas, ingressantes, matrículas e concluintes (Ristoff, 2014). Esses vinte anos representaram um marco no desenvolvimento da educação nacional, com uma maior democratização do ensino. A expansão é contínua e branda nos dias atuais, isso deve, em grande parte, à políticas públicas criadas, como o Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI) e o Universidade para Todos (PROUNI), que incentivam não só a criação e manutenção de novos cursos, como também o ingresso de novos alunos. Para apoiar tais expansões foram necessários grandes investimentos em infraestrutura, possibilitando a criação de novos campi, a oferta de novos cursos e de mais vagas em todo o Brasil (Alves; Gaydeczka; Campos, 2018).

De certo modo, as políticas públicas não são suficientes para manter os alunos em cursos superiores até a conclusão. Assim, de acordo com Silva et al. (2022), pode-se afirmar que, para os evadidos, existem prejuízos sociais, acadêmicos e econômicos. Para o setor público, são recursos públicos investidos sem o devido retorno, e para o setor privado, é uma considerável perda de receitas. Todos os casos levam a um mesmo caminho: o ócio de docentes, de funcionários, de equipamentos e de espaço físico. Na generalidade, os dados relacionados à evasão escolar, bem como seus motivos, devem ser objeto de interesse e estudo das Instituição de Ensino Superior (IES); principalmente no que concerne à coleta de dados no momento em que o aluno realiza o pedido de desligamento da universidade (Alves; Gaydeczka; Campos, 2018).

Os indicadores relacionados à evasão escolar são múltiplos. Os motivos podem ser classificados em exógenos, quando extrínsecos ao evadido, ou endógenos, quando intrínsecos e externados (Bean, 1980). Dessa maneira, faz-se o seguinte modelo, composto por variáveis independentes que podem ter razões internas ou externas:

- **Variáveis *Background*:** quando experiências com família, ensino regular, desempenho escolar, entre outros influenciam na permanência ou não do estudante no ensino superior;
- **Variáveis organizacionais:** quando a estrutura, organização e notas da instituição são preponderantes para a escolha do estudante;
- **Variáveis pessoais:** quanto à presença de compromisso com metas, certeza ocupacional e autoconfiança;
- **Variáveis ambientais:** quando há oportunidade de transferência, casamentos, aprovação da família e localização;
- **Variáveis atitudinais:** condutas que se referem à escolha, segurança, satisfação e calorização pessoal pelo curso e instituição;

Ainda nesse parâmetro, levando em consideração as razões que levam à evasão escolar, Ambiel (2015) listou e categorizou possíveis razões, separando-as nas quatro seguintes principais classes de motivos:

- **Motivos institucionais:** qualidade do corpo docente, relacionamento docente-discente e falta de serviços voltados para o discente;
- **Motivos pessoais:** incerteza quanto à escolha do curso, questões de saúde familiares, entre outros;

- **Motivos financeiros:** impossibilidade de conciliar o curso e o trabalho, não ter como se sustentar financeiramente;
- **Hesitação de carreira:** perceber que o curso não será tão prazeroso quanto pareceu, perceber que o curso pode não ajudar a conseguir um bom emprego no futuro, insatisfação com a escolha.

De maneira geral, a análise e compreensão individual desses motivos por parte da instituição é necessária para que a questão seja sanada. Todavia, a falta de métodos uniformes e eficazes para qualificar a evasão dificulta a implementação de medidas institucionais e públicas (Gonçalves, 2024). Em sua maioria, de acordo com a cartilha ao discente, as instituições exigem o preenchimento de um formulário de desistência para o trancamento do curso. Contudo, nem todos os evadidos preenchem o formulário, visto que existem, também, alunos que apenas abandonam o curso.

Diante disso, encontra-se um grande empecilho no que tange à análise completa e confiável dos dados. Ainda, os motivos declarados são utilizados, atualmente, mais para questões informativas do que para um estudo aprofundado e possíveis melhorias na instituição.

Apesar disso, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) dispõe de um *dashboard* com dados gerais de evasão e conclusão do ensino superior (Power BI Report, 2023). O *dashboard*, contudo, trata dos dados de uma maneira geral, dentro do território nacional. Deve-se levar em conta que, embora sejam plurais, os motivos que levam à evasão são, também, individuais, devendo ser tratados de maneira específica por cada IES.

2.2 *Business Intelligence* na Educação

O BI está relacionado com o processo de se obter, organizar, analisar, compartilhar e monitorar dados a fim de dar suporte à uma gestão (Gartner Group, 2024). De maneira geral, ele tem por objetivo melhorar o tempo de resposta e a qualidade dos dados coletados, de forma que os gestores entendam melhor as informações (SANTOS, 2017).

Como ressalta Júnior et al. (2022), o desenvolvimento e sucesso de uma organização depende de sua capacidade de coletar, tratar, interpretar e utilizar as informações com eficácia, com o objetivo de definir estratégias. As universidades, por sua vez, vivem um momento desafiador, de constantes evoluções tecnológicas, atrelado a uma situação, possibilitando que os gestores disponham de ferramentas que os proporcionem maior rapidez e agilidade para encontrar e interpretar informações, auxiliando na tomada de decisões (Silva et al., 2014).

Embora, como citado anteriormente, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) (2023) disponha de um *dashboard* repleto de informações relacionadas ao ensino superior com indicadores quantitativos, como número de matrícula, concluintes, docentes, ingressos, entre outros, trata-se de um *dashboard* generalizado. Para tratar a problemática de maneira mais profunda, é necessário que os dados de cada instituição sejam tratados de maneira individualizada, com cada IES dispondo de um *dashboard* fundamentado pelo BI.

2.2.1 *Data Warehouse*

Data Warehouse é uma infraestrutura de armazenamento de dados que fornece um suporte analítico. De acordo com Inmon e Linstedt (2014), é um conjunto centralizado que consolida dados, integrando-os e organizando-os. Implementar um *Data Warehouse* robusto viabiliza a consolidação de dados históricos e de fontes variadas em um único repositório. Assim, são feitas análises mais aprofundadas, além do desenvolvimento de estratégias com base em evidências.

No âmbito educacional, o uso do *Data Warehouse* tem sido amplamente difundido, sendo possível notar melhorias significativas na gestão dos dados. Tais melhorias tornam capaz a redução dos índices da evasão, baseada em análises mais precisas e preditivas (Han; Pei; Tong, 2022). A partir dos repositórios, pode-se obter insights sobre indicadores-chave, como taxas de matrícula, retenção e evasão, o que facilita a criação de políticas mais eficazes. Ainda, o uso de técnicas avançadas associadas ao *Data Warehouse* torna possível identificar tendências e padrões capazes de impactar o desempenho da instituição.

Dentre as estruturas mais utilizadas no *Data Warehouse*, está o modelo estrela (*star schema*), que organiza os dados em uma tabela de fatos central conectada a várias outras tabelas dimensionais. Essa abordagem, além de simplificar o processo de consulta e análise, oferece uma visão otimizada e acessível para relatórios analíticos (Kimball; Ross, 2019). Levando para o contexto em questão, por exemplo, a tabela de fatos pode conter métricas da evasão escolar, enquanto as tabelas dimensionais trazem informações sobre os alunos, cursos, períodos e professores. Essa estrutura permite aos gestores a análise dos dados de maneira granular e a compreensão do impacto da pluralidade dos fatores para a evasão.

2.2.2 *Dashboard* e visualização de dados

A visualização de dados é um apetrecho fundamental nessa situação. Estudos atuais denotam a significância de *dashboards* interativos para a sintetização de dados complexos e apresentação de *insights* processáveis, o que permite aos gestores a

identificação de desvios e paradigmas nos dados acadêmicos (Sarikaya et al., 2018). Algumas ferramentas disponíveis, como Power BI, Tableau e Looker Studio, facilitam não só a compreensão desses dados, mas, também, a relação com os demais setores da instituição.

Essas ferramentas são especialmente úteis quando a tomada de decisões rápidas é necessária. A integração do *dashboard* com tecnologias como *Data Warehouse* aumenta ainda mais a eficiência e escalabilidade das análises. Conforme os estudos de Few e Edge (2017) ressaltam, os *dashboards* projetados de forma adequada tornam os dados mais acessíveis e garantem que informações essenciais sejam disponibilizadas de forma eficaz e sem sobrecarga visual.

2.2.3 Google Looker Studio

O Google Looker Studio é uma ferramenta de BI, utilizada para integrar e unificar dados provenientes de diversas fontes de informação em uma mesma plataforma. A plataforma possui acesso fácil e intuitivo, permitindo que o usuário confeccione painéis de dados, relatórios e formate dados, dimensões e métricas (Google Cloud, 2024). A utilização dessa ferramenta, além de ser gratuita, é on-line, dispensando a instalação de softwares locais e possibilitando a atualização automática dos *Dashboards* conforme os dados são atualizados (Cardoso, 2024).

Nesse parâmetro, ao transformar dados complexos em visualizações rápidas e compreensíveis, o Looker Studio contribui para a tomada de decisões gerenciais (Dutra, 2024). Diante disso, por ser uma ferramenta intuitiva, de fácil acesso e compreensão e gratuita, o Google Looker Studio é uma alternativa promissora para a análise dos dados de evasão escolar no ensino superior, apoiando o gerenciamento na tomada de decisões.

2.3 Trabalhos Relacionados

Para fundamentar a proposta deste trabalho, são apresentados, a seguir, estudos relacionados que contribuíram para o entendimento do problema e serviram de base para as decisões metodológicas adotadas.

2.3.1 *Um ambiente de Data Warehousing para apoiar a tomada de decisão quanto à evasão escolar na UFPE*

A utilização de técnicas de *Business Intelligence* no contexto do ensino superior tem-se mostrado de grande valia para transformar dados educacionais volumosos em informações estratégicas acessíveis e de fácil compreensão, apoiando a tomada

de decisões gerenciais voltadas para a redução da evasão escolar. Nesse cenário, diversos estudos vêm demonstrando a viabilidade de ambientes de *Data Warehousing* integrados a ferramentas de visualização para monitorar indicadores relevantes.

A exemplo disso, o trabalho desenvolvido na Universidade Federal de Pernambuco (UFPE), por Alves e Fidalgo (2021) propõe um ambiente de *Data Warehousing* que centraliza dados anonimizados pelo sistema SIG@ UFPE, aliado a dados artificiais, proporcionando a criação de uma base de dados estruturada para análise. Com o uso de técnicas de modelagem dimensional, incluindo o modelo estrela (*Star Schema*), e ferramentas de BI, como cubos multidimensionais e *dashboards* interativos, o ambiente possibilita uma visualização eficiente de indicadores da evasão escolar, tais como taxas de retenção, tempo médio de evasão e perfil dos evadidos.

Implementar painéis de BI que agreguem esses dados é reflexo de uma tendência crescente de disponibilizar visualizações dinâmicas, que facilitam a compreensão do fenômeno de evasão pela gestão acadêmica. Dessa forma, a tecnologia de BI fornece uma plataforma para elaborar relatórios interativos e *dashboards* que evidenciam padrões de comportamento dos estudantes e pontos críticos no percurso acadêmico (Alves; Fidalgo, 2021).

O trabalho desenvolvido por Alves e Fidalgo (2021) discute como um ambiente de *Data Warehousing* pode centralizar, organizar e analisar dados educacionais. Isso possibilita a criação de painéis e relatórios facilitadores da compreensão do cenário de evasão e apoia a tomada de decisão da gestão universitária.

2.3.2 *Aplicação de uma solução de Business Intelligence para apoiar a Gestão Acadêmica em Universidades Federais Brasileiras*

O trabalho de Silva et al. (2023) reforça que o uso de *Business Intelligence* em universidades brasileiras proporciona relatórios ágeis e *dashboards* interativos que apoiam a Gestão Acadêmica. É destacada, também, a possibilidade que o BI proporciona para a identificação de alunos vulneráveis desde o início, através de indicadores como notas, faltas e fatores sociais, facilitando intervenções precoces para a redução da evasão escolar no ensino superior.

Além de reforçar a relevância do uso de BI no contexto de evasão escolar no ensino superior, o trabalho discute etapas e estratégias para a implementação da técnica, de maneira que propõe levantamento de requisitos, arquitetura de dados e outras ferramentas complementares. Ainda, Silva et al. (2023) sugere o uso de novos indicadores, como desempenho acadêmico, socialização e convívio familiar, denotando que uma análise de múltiplas dimensões é crucial para que a compreensão seja ampla.

Ademais, é destacada a importância de se considerar as particularidades, como

discutido anteriormente, no Capítulo 2. A implementação da técnica deve considerar a pluralidade nacional, considerando, assim, que diferentes localidades e pessoas carregam diferentes razões e empecilhos.

3 DESENVOLVIMENTO

Este capítulo expõe a execução do estudo, utilizando as definições discutidas e discorrendo sobre as etapas expostas no Capítulo 1. O capítulo se inicia na seção 3.1, apresentando o processo ETL; a seção 3.2 tratará do uso de *Business Intelligence* e da construção dos *dashboards*. A seção 3.3 apresentará a integração dos painéis gerados ao Sistema Web. Por fim, a seção 3.4 tratará da validação da ferramenta construída.

3.1 Processo ETL

Esta seção explica como foi feito o processo de Extração, Tratamento e Carga (ETL). Segundo Khan et al. (2024), o processo ETL desempenha um papel crucial na integração de dados provenientes de múltiplas fontes, permitindo análises mais eficazes e apoiando a tomada de decisões.

3.1.1 Extração

As bases de dados utilizadas foram o Controle Acadêmico (CA) da UESPI, armazenado em SQL Server — sendo este o sistema legado — e a base de dados (*Pre-prod_SIGAA*), do sistema Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA), baseado em PostgreSQL, que representa a base de dados em uso atualmente. A seleção das tabelas foi feita com base em reuniões com o Analista de Informática da Diretoria de Tecnologia da Informação e Comunicação (DTIC) da Universidade Estadual do Piauí (UESPI), com o intuito de identificar os dados necessários para responder às seguintes questões:

- Quais são os cursos com o maior número de registros de evasão?
- Quantas evasões foram registradas em um determinado período letivo?
- Qual é a faixa etária dos alunos que evadiram?

Os dados obtidos com base nesses questionamentos estão listados em tabelas organizacionais dos bancos de dados do Controle Acadêmico (CA) e do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA). O nome das tabelas extraídas e suas respectivas informações estão dispostas no Quadro 1, com dados do CA, e Quadro 2, com dados do SIGAA.

Quadro 1: Tabelas extraídas do banco de dados CA (Controle Acadêmico)

Nome da tabela	Descrição
dcu	Tabela que relaciona os alunos aos cursos
mat	Tabela que armazena a matrícula dos alunos, relacionando-os às turmas e disciplinas
cur	Lista os cursos disponíveis na instituição
dpe	Tabela que lista os dados pessoais dos alunos
oco	Tabela que registra as ocorrências
oca	Tabela que armazena o relacionamento entre a matrícula do aluno e as ocorrências registradas

Fonte: Elaborado pelo autor.

Quadro 2: Tabelas extraídas do banco de dados SIGAA

Nome da tabela	Descrição
public.curso	Lista os cursos cadastrados na instituição
public.discente	Tabela contendo a relação entre os alunos e os cursos
comum.pessoa	Tabela que lista os dados pessoais dos alunos
ensino.tipo_movimentacao_aluno	Tabela que lista os tipos de movimentação que podem ocorrer na instituição
ensino.movimentacao_aluno	Tabela que contém a relação dos alunos com o tipo de movimentação

Fonte: Elaborado pelo autor.

Com base na análise do Quadro 1 e Quadro 2, foi definido que a coleta dos dados necessários para a construção do *Data Warehouse* abrangeria o período de 2019.1 a 2024.2, contemplando os alunos que ingressaram em 2019.1 ou 2019.2. Esse processo foi realizado separadamente para cada dimensão e para cada base de dados, respeitando suas respectivas estruturas internas. Além disso, identificou-se que a base de dados atual passou a ser utilizada no período de 2023.1. Dessa forma, para a construção das dimensões, foram considerados os dados do sistema legado (CA), no período de 2019.1 a 2022.2, e os dados do SIGAA, no período de 2023.1 a 2024.2. Por fim, foram considerados somente alunos do regime regular em ambas as bases, por apresentar uma estrutura maior de dados consolidados.

A seguir, são descritos os procedimentos adotados para a construção de cada dimensão, destacando a origem, os critérios de extração e os atributos dos dados considerados relevantes para a coleta.

3.1.1.1 Dimensão Aluno – Sistema SIGAA

A coleta dos dados para a construção da dimensão *Aluno* através do sistema SIGAA considerou apenas os dados que, quando feita a vinculação com a tabela `ensino.movimentacao`, possuísem registros de ocorrências entre 2023.1 e 2024.2.

Além do critério temporal, foram considerados apenas os alunos que apresentaram registros de movimentações acadêmicas classificadas como evasão, conclusão ou integralização. Essa seleção foi realizada com base no atributo chave primária `id_tipo_movimentacao_aluno`. O Quadro 3 apresenta os identificadores utilizados nesse processo, tomando como base os motivos declarados no sistema.

Quadro 3: Movimentações considerados na coleta da dimensão *Aluno* (SIGAA).

ID	Tipo de Movimentação	Categoria
1	Concluído	Conclusão
315	Integralização de Discente	Conclusão
2	Aluno cancelou a matrícula	Evasão
6	Cancelamento judicial	Evasão
7	Matrícula cancelada	Evasão
8	Excluído	Evasão
9	Desistência	Evasão
10	Cancelamento por novo vestibular	Evasão
11	Cancelamento por reopção	Evasão
13	Cancelamento – RES 083/98	Evasão
14	Cancelamento – RES 084/98	Evasão
15	Cancelamento – RES 255/92	Evasão
16	Abandono	Evasão
19	Abandono (sem matrícula)	Evasão
20	Abandono (sem integralização)	Evasão
306	Desligado	Evasão
308	Cancelamento espontâneo	Evasão
309	Cadastro cancelado	Evasão
336	Jubilamento	Evasão
339	Matrícula cancelada	Evasão
359	Período cancelado	Evasão
358324	Desistente	Evasão

Fonte: Elaborado pelo autor.

Por fim, a aplicação dos critérios de filtragem possibilitou isolar os registros de alunos com movimentações relevantes para o estudo. A utilização da lógica de prioridade aplicada, ilustrada no Apêndice A.1, garantiu que, nos casos de múltiplas movimentações, prevalecesse a mais significativa — dando prioridade à conclusão ou integralização, em detrimento das ocorrências de evasão. Dessa forma, obteve-se uma base consolidada e confiável para a construção da dimensão *Aluno*.

3.1.1.2 Dimensão Aluno – Sistema CA

A extração da dimensão *Aluno* a partir do sistema legado seguiu a mesma lógica adotada para o SIGAA, considerando apenas os alunos com ocorrências registradas entre os períodos de 2019.1 e 2022.2, garantindo a compatibilidade com o recorte temporal adotado em todo o *Data Warehouse*.

Essa coleta foi um pouco mais complexa, visto que, diferentemente do SIGAA, o sistema CA não armazena de forma explícita um status que identifique alunos que integralizaram o curso. Por essa razão, a extração foi dividida em três etapas principais: primeiramente, foram coletados os alunos com status de graduado (GRD) (ver Apêndice A.2); em seguida, foram extraídos os alunos com status de não desligado (NDE) que não possuíam pendência de disciplinas em seus históricos (ver Apêndice A.3), os quais também foram considerados como concluintes; por fim, foram coletados os alunos evadidos (ver Apêndice A.4). Para isso, foram utilizados identificadores específicos de ocorrência que indicam desligamento ou evasão, como mostrado no Quadro 4. Além disso, para evitar duplicidade ou sobreposição, foram excluídos da seleção os alunos já identificados como concluintes nas etapas anteriores.

Quadro 4: Ocorrências consideradas na coleta da dimensão *Aluno* (CA).

ID	Descrição
0	Desistente
CAN	Período cancelado
JUB	Jubilamento
NFM	Não fez matrícula institucional
NMA	Não matriculado

Fonte: Elaborado pelo autor.

3.1.1.3 Dimensão Aluno – Modelo Estrela

Após o processo de coleta, os dados passaram pelo processo de limpeza e padronização, utilizando uma etapa de *staging* em cada base de dados, como descrito na Seção 3.1.3.1, sendo possível unificá-los e criar somente uma tabela. Essa dimensão consolida as características mais relevantes dos discentes, fornecendo data de nascimento, sexo, cidade e estado onde nasceu.

3.1.1.4 Dimensão Curso - Modelo Estrela

Para fins de padronização e compatibilidade entre as duas bases de dados utilizadas, a dimensão curso foi filtrada de modo a incluir apenas os cursos com moda-

lidade presencial. Essa decisão se justifica pelo foco do estudo na evasão dos cursos presenciais, excluindo assim modalidades à distância ou semipresenciais que seguem dinâmicas pedagógicas distintas.

Nessa coleta optou-se por carregar apenas os cursos efetivamente vinculados às matrículas dos alunos presentes na *dim_aluno*. Para isso, foram realizadas extrações tanto da base de dados *SIGAA* quanto do sistema legado *CA*, garantindo a inclusão de todos os cursos frequentados pelos discentes analisados.

No *SIGAA*, a extração considerou a tabela de vínculo entre pessoa e curso, filtrando apenas os cursos presenciais, identificados pelo *id_modalidade_educacao* = 1 (ver Apêndice A.5). No *CA*, foram extraídos os cursos associados às matrículas válidas, conforme registros da tabela de movimentações acadêmicas (ver Apêndice A.6).

Após extraídos, os cursos passaram por uma etapa de *staging*, descrita na Seção 3.1.3.4, onde foram unificados e tratados para evitar duplicações e inconsistências. A carga final resultou em uma tabela com identificador único e nome do curso, conforme estrutura definida no Quadro 7.

3.1.1.5 Dimensão Período Letivo - Modelo Estrela

A dimensão período letivo foi construída a partir do calendário acadêmico oficial da instituição, contendo os períodos em que as atividades letivas foram realizadas. Cada registro corresponde a um semestre específico, com seus respectivos anos, datas de início e fim. O resultado dessa estrutura é apresentado no Quadro 5.

Quadro 5: Estrutura da Tabela *dim_periodo*

Nome da Coluna	Tipo de Dado
pk_periodo	Inteiro
ano	Inteiro
semestre	Inteiro
data_inicio	Data
data_fim	Data

Fonte: Elaborado pelo autor.

3.1.2 Limpeza dos Dados

A etapa de limpeza dos dados teve como objetivo garantir a consistência, a integridade e a padronização das informações utilizadas na construção do *Data Warehouse*. Considerando que os dados foram extraídos de duas fontes distintas — o sistema legado Controle Acadêmico (*CA*) e o *SIGAA* —, a limpeza foi essencial para

resolver problemas encontrados de campos nulos, registros inconsistentes ou vazios e divergência de formatos entre os sistemas.

Inicialmente, foram tratados os valores nulos ou ausentes nos campos data de nascimento, sexo e curso do discente. Registros com ausência desses campos foram descartados da dimensão aluno. Além disso, alunos com o atributo cidade e uf vazios foram padronizados para *Não informada* e *NI*, respectivamente. Também foram removidos registros duplicados de alunos, resultantes de múltiplas movimentações associadas à mesma matrícula, garantindo que cada discente fosse representado apenas uma vez na dimensão. No caso dos dados oriundos do *SIGAA*, foi aplicada a função `ROW_NUMBER()` para priorizar as movimentações de conclusão, seguida por integralização e evasão. Já os dados do sistema *CA* foram combinados de forma que cada matrícula figurasse apenas uma vez na dimensão aluno, sem a necessidade de deduplicação adicional durante a carga.

Outro ponto crítico foi a harmonização de atributos e descrições provenientes de tabelas auxiliares, como os tipos de movimentação e ocorrências acadêmicas, cujos identificadores variavam entre os sistemas. A categorização entre evasão, conclusão e integralização foi validada manualmente e em reuniões com técnicos do DTIC. Por fim, na dimensão curso, foram eliminados cursos inativos, duplicados ou sem vínculo com alunos no período analisado.

3.1.3 Carga

Após a etapa de extração e limpeza, os dados foram organizados em estruturas intermediárias conhecidas como áreas de *staging*, com o objetivo de unificar e padronizar os registros provenientes das duas bases de dados distintas: o sistema legado *CA* (SQL Server) e o sistema atual *SIGAA* (PostgreSQL).

3.1.3.1 Carga da Tabela *staging_aluno_ca*

Para o sistema *CA*, foram utilizadas três consultas SQL distintas, cada uma responsável por identificar um grupo específico de alunos:

- Alunos com status *GRD* (graduados).
- Alunos não desligados (*NDE*) sem pendência de disciplinas.
- Alunos com ocorrências de desligamento ou evasão.

Essas consultas foram executadas diretamente no banco SQL Server e os resultados foram exportados em arquivos separados no formato *CSV*. Cada linha do arquivo contém os seguintes campos: matrícula, data de nascimento, sexo, cidade e UF.

3.1.3.2 Carga da Tabela *staging_aluno_sigaa*

A extração dos dados do sistema *SIGAA* foi feita com uma consulta única, que já aplicava os critérios de filtragem por movimentações acadêmicas e utilizava a função `ROW_NUMBER()` para selecionar a movimentação mais relevante para cada aluno. O resultado foi exportado em formato CSV, contendo os mesmos campos utilizados na tabela do CA.

Ambos os arquivos foram então utilizados para povoar as tabelas de staging correspondentes no MySQL, preparando os dados para a unificação e carga final na dimensão *dim_aluno*.

3.1.3.3 Carga da Tabela *dim_aluno*

A partir das tabelas de staging previamente populadas, foi realizada a carga da dimensão *dim_aluno*, responsável por consolidar os dados dos discentes únicos presentes nas duas fontes. Como critério de priorização, optou-se por dar precedência aos registros extraídos do sistema *SIGAA*, uma vez que ele representa a base mais atual e institucionalmente adotada pela universidade.

O processo de carga foi dividido em duas etapas:

1. Inserção dos registros da tabela *staging_aluno_sigaa* diretamente na *dim_aluno*, utilizando a matrícula como chave primária(ver Apêndice A.7);
2. Inserção dos registros da tabela *staging_aluno_ca* apenas quando a matrícula não estivesse previamente cadastrada na *dim_aluno*, garantindo a unicidade dos discentes (ver Apêndice A.8).

Essa abordagem evitou a duplicação de dados e assegurou que cada matrícula foi representada apenas uma vez na dimensão, mesmo que estivesse presente em ambas as bases. O Quadro 6 apresenta a estrutura final da dimensão *dim_aluno*.

Quadro 6: Estrutura da Tabela *dim_aluno*

Nome da Coluna	Tipo de Dado
matricula (PK)	Texto (20 caracteres)
data_nascimento	Data
sexo	Texto (1 caractere)
cidade	Texto
uf	Texto (2 caracteres)

Fonte: Elaborado pelo autor.

Ao final da carga, a dimensão *dim_aluno* passou a conter todos os discentes únicos contemplados pelas consultas das bases *CA* e *SIGAA*, já tratados e limpos.

3.1.3.4 Carga da Tabela *dim_curso*

A carga da dimensão *dim_curso* foi realizada a partir da integração dos dados provenientes de duas bases de dados, com o objetivo de consolidar as informações dos cursos de graduação oferecidos pela instituição, preservando a distinção entre diferentes polos/campi.

Por possuírem chaves primárias distintas, a carga foi realizada em duas etapas, a partir das seguintes tabelas de *staging*:

1. *staging_curso_sigaa*: extraída do *SIGAA*, contendo os campos *id_curso*, *nome*, *campus*, *id_curso_legacy*, *id_polo_legacy* e *codmerg*;
2. *staging_curso_ca*: extraída do sistema legado, contendo os campos *id_curso*, *nome*, *campus* e *id_polo*.

O campo *codmerg*, presente na base *SIGAA*, armazena o identificador do curso e do campus no sistema legado em um único campo, com as chaves primárias do sistema legado concatenadas, no formato GRAD#<*id_curso_legacy*>_<*id_polo_legacy*>. Dessa forma, foi necessário aplicar uma expressão de extração (ver Apêndice A.9) para atualizar as colunas: *id_curso_legacy* e *id_polo_legacy*.

Com esses dados extraídos, foi possível identificar quais cursos estavam presentes apenas no sistema legado. Para garantir a padronização das informações, foi necessário atualizar os nomes dos campi na tabela *staging_curso_ca*, utilizando como base os nomes presentes no *SIGAA*. A atualização foi realizada por meio de uma instrução UPDATE com junção entre os polos (ver Apêndice A.10).

Com as informações padronizadas, foi possível montar a estrutura *dim_curso*, conforme descrito no Quadro 7. Ela foi carregada em duas fases: inicialmente com os cursos do *SIGAA*, e posteriormente com os cursos exclusivos do sistema legado, utilizando-se como chave primária composta o par (*id_curso*, *campus*), baseado no identificador do sistema legado (ver Apêndice A.11).

Quadro 7: Estrutura da Tabela *dim_curso*

Campo	Tipo	Descrição
id_curso	INT	Identificador do curso
nome	VARCHAR(255)	Nome do curso
campus	VARCHAR(100)	Nome do campus/polo

Fonte: Elaborado pelo autor.

3.1.3.5 Carga da Tabela *dim_perodo*

A carga da dimensão período foi feita primeiramente criando um arquivo CSV com as colunas *ano*, *semestre*, *data_inicio* e *data_fim*. Após isso, foi realizada a importação do arquivo diretamente na tabela, que possui a estrutura definida conforme indicado no Quadro 5.

3.1.3.6 Carga da Tabela *fato_ocorrencia_aluno*

Após o carregamento e tratamento das dimensões, foi realizada a carga da tabela fato *fato_ocorrencia_aluno*, que centraliza os dados de ocorrência acadêmica de cada matrícula ao longo dos períodos. Para garantir a integridade referencial, foram estabelecidas chaves estrangeiras para as dimensões *dim_aluno*, *dim_curso* e *dim_perodo*. Além disso, foram adicionados os campos de status final (evasão ou conclusão) e a respectiva data de ocorrência. A carga foi realizada de forma incremental, com os dados devidamente normalizados para permitir análises por tempo, curso, campus, sexo, faixa etária e situação acadêmica. O Quadro 8 mostra a estrutura resultante, que viabiliza a criação de *dashboards* dinâmicos por meio de ferramentas de *Business Intelligence*, voltados à análise de indicadores da evasão no ensino superior.

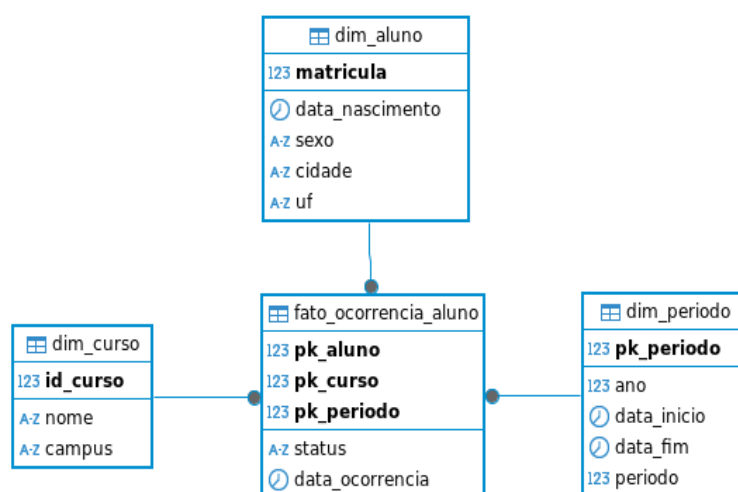
Quadro 8: Estrutura da Tabela *fato_ocorrencia_aluno*

Campo	Tipo	Descrição
pk	INT	Identificador único da ocorrência (chave primária)
pk_aluno	INT	Chave estrangeira para a tabela <i>dim_aluno</i>
pk_curso	INT	Chave estrangeira para a tabela <i>dim_curso</i>
pk_periodo	INT	Chave estrangeira para a tabela <i>dim_periodo</i>
status_final	VARCHAR(20)	Situação final do aluno no período (evasão ou conclusão)
data_ocorrencia	DATE	Data em que ocorreu a situação final

Fonte: Elaborado pelo autor.

Após todo esse processo, foi obtido o modelo estrela que alimenta os *dashboards* desenvolvidos, conforme mostrado na Figura 1, que serão responsáveis por mostrar de forma visual e interativa os principais indicadores relacionados à evasão no ensino superior.

Figura 1 – Modelo Estrela



Fonte: Elaborado pelo autor.

3.2 Business Intelligence

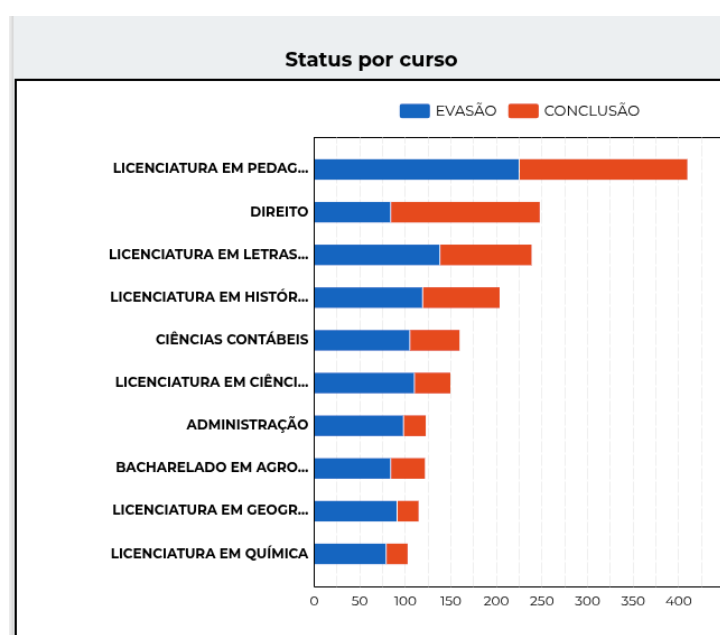
Para a construção dos *dashboards* interativos e análise dos indicadores obtidos a partir do Data Warehouse criado, optou-se pelo uso da ferramenta *Looker Studio*. A escolha se deu principalmente por ser uma solução gratuita, baseada na nuvem e desenvolvida pelo Google, o que permite fácil integração com arquivos CSV, Google Sheets, e outras fontes de dados, sem a necessidade de instalação de softwares adicionais (Google Cloud, 2024).

3.2.1 Construção dos Dashboards

A construção dos *dashboards* foi realizada dentro do ambiente do *Looker Studio*, por meio da importação dos arquivos CSV gerados a partir do modelo estrela. Cada tabela foi importada como uma fonte de dados independente e, em seguida, as relações entre elas foram estabelecidas por meio da funcionalidade *Combinar dados*. Após a união dos dados, foram desenvolvidos dois *dashboards* interativos, o primeiro contemplando uma visão geral por curso e campus, e o segundo possibilitando uma análise demográfica e temporal.

Primeiramente, foi adicionado um gráfico de barras empilhado, apresentado na Figura 2, utilizando como dimensão o nome do curso e detalhado por status de evasão ou conclusão.

Figura 2 – Status por Curso



Fonte: Elaborado pelo autor.

Posteriormente, foi adicionada uma tabela que lista os campi existentes no mo-

delo estrela, agrupando-os pela quantidade de evasão e porcentagem, conforme apresentado na Figura 3.

Figura 3 – Evasão por Campus

campus	Quantidade	%
1. CENTRO DE CIÊNCIAS HUM...	213	12,23%
2. CAMPUS DE PIRIPIRI	177	10,16%
3. CENTRO DE CIÊNCIAS SOCI...	167	9,59%
4. CAMPUS DE FLORIANO	159	9,13%
5. CENTRO DE CIÊNCIAS DA N...	158	9,07%
6. CAMPUS DE PARNAÍBA	134	7,69%
7. CAMPUS DE PICOS	117	6,72%
8. CAMPUS DE CORRENTE	107	6,14%
9. CAMPUS CLÓVIS MOURA	96	5,51%
10. CENTRO DE CIÊNCIAS DA S...	84	4,82%
11. CAMPUS DE SÃO RAIMUND...	74	4,25%
12. CAMPUS DE CAMPO MAIOR	63	3,62%
13. CENTRO DE CIÊNCIAS AGR...	55	3,16%
14. CAMPUS DE BOM JESUS	52	2,99%
15. CENTRO DE CIÊNCIAS DA E...	44	2,53%
16. CENTRO DE TECNOLOGIA E...	15	0,86%

Fonte: Elaborado pelo autor.

Em seguida, foram adicionados três medidores, apresentados na Figura 4, indicando a quantidade total de alunos ingressantes em 2019, a quantidade de alunos evadidos e a quantidade de alunos graduados.

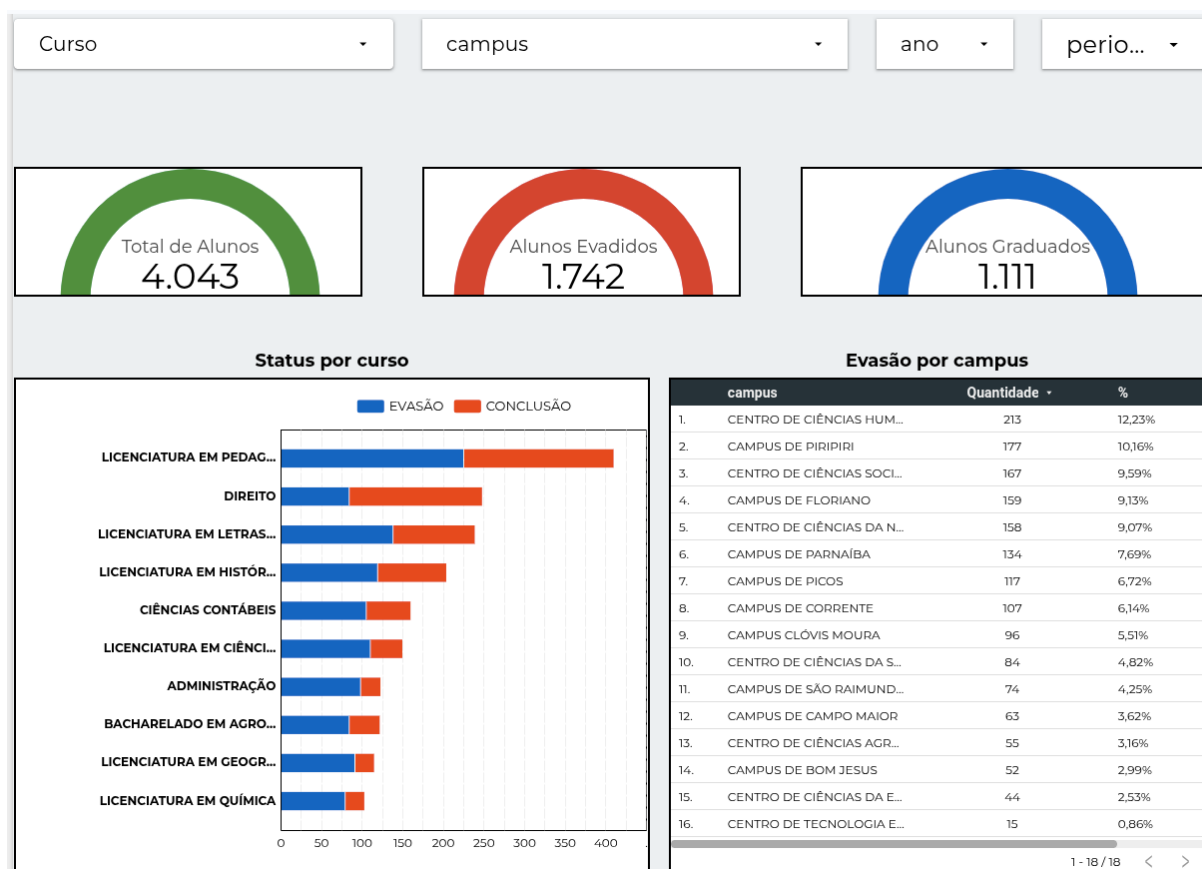
Figura 4 – Contadores



Fonte: Elaborado pelo autor.

Por fim, foram adicionados quatro filtros: curso, campus, ano e período. Assim, é possível analisar os indicadores de evasão e conclusão de forma segmentada, permitindo comparações entre diferentes unidades, períodos letivos e áreas de formação dentro da instituição, como ilustrado na Figura 5.

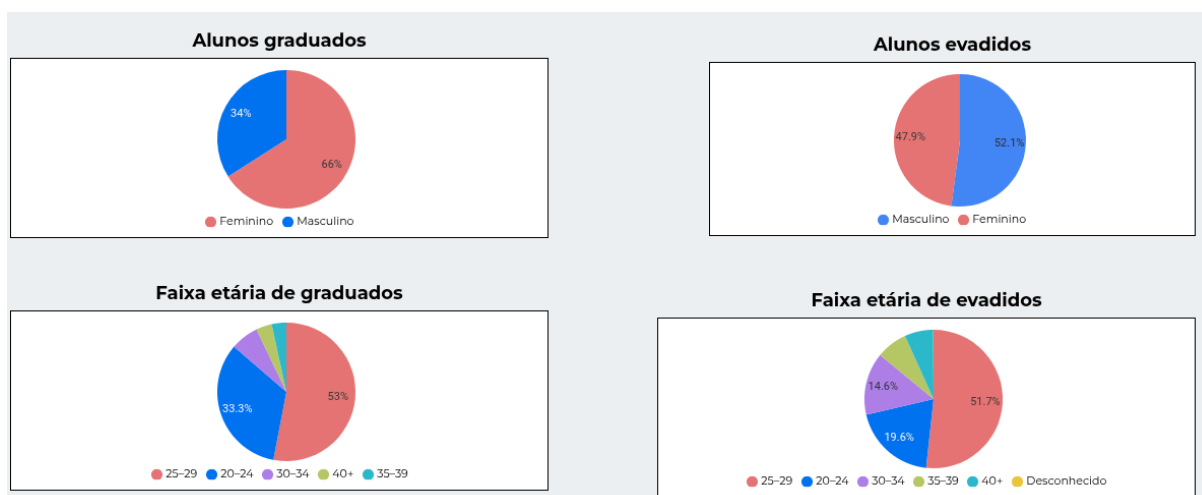
Figura 5 – Dashboard 1



Fonte: Elaborado pelo autor.

Para a construção de um segundo *dashboard*, ilustrado na Figura 6, foram utilizados quatro gráficos de pizza, contemplando as porcentagens de alunos graduados e evadidos, detalhados por sexo e faixa etária.

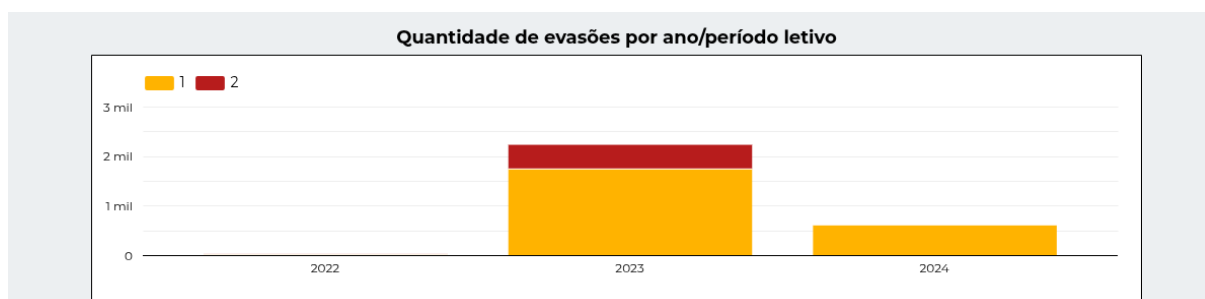
Figura 6 – Alunos Dividos por Sexo e Faixa Etária



Fonte: Elaborado pelo autor.

Além disso, foi adicionado também um gráfico de barras, conforme a Figura 7, utilizando o ano como dimensão e o período como detalhamento, possibilitando visualizar separadamente a quantidade de alunos evadidos por ano e período.

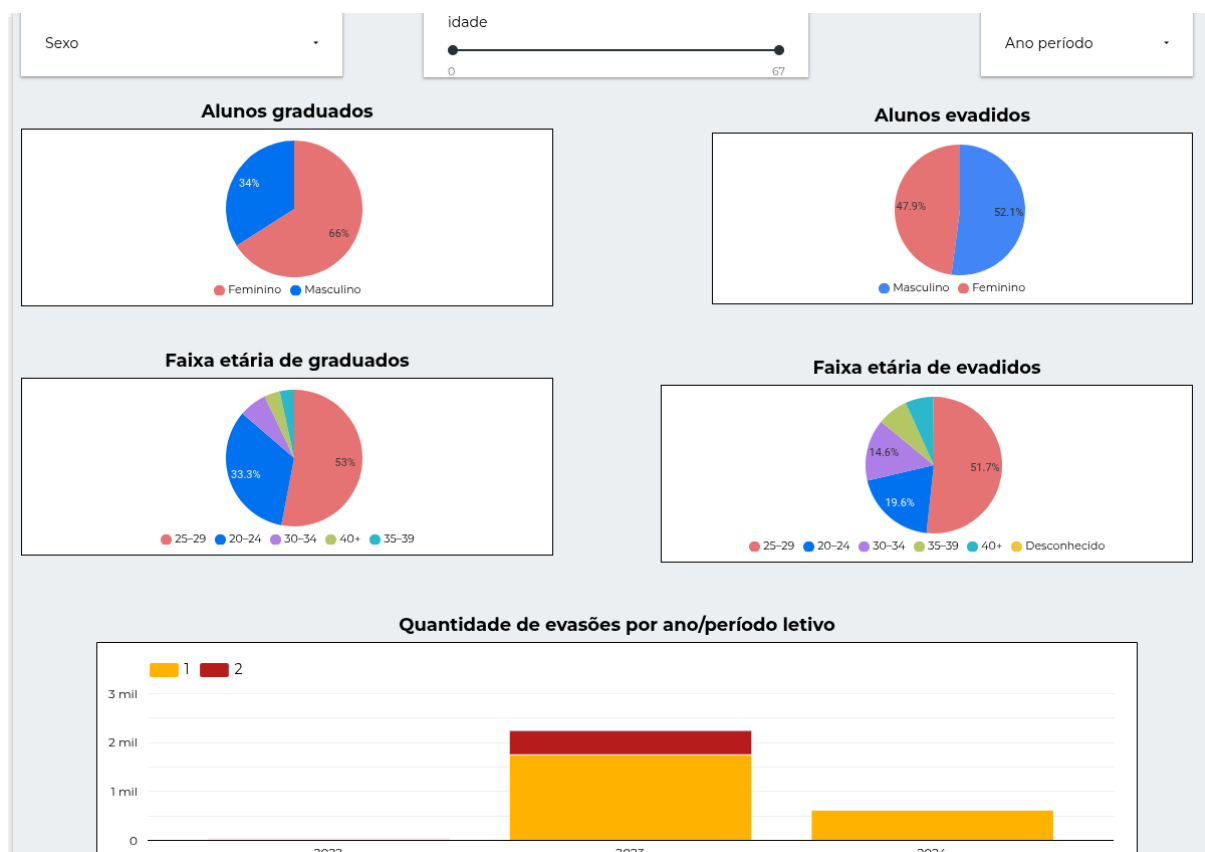
Figura 7 – Evasão por Período Letivo



Fonte: Elaborado pelo autor.

Por fim, foram adicionados três filtros: sexo, idade e ano/período. Assim, é possível realizar análises demográficas, identificando padrões de evasão e conclusão ao longo do tempo com base no sexo dos alunos e sua faixa etária, como ilustrado na Figura 8.

Figura 8 – Dashboard 2



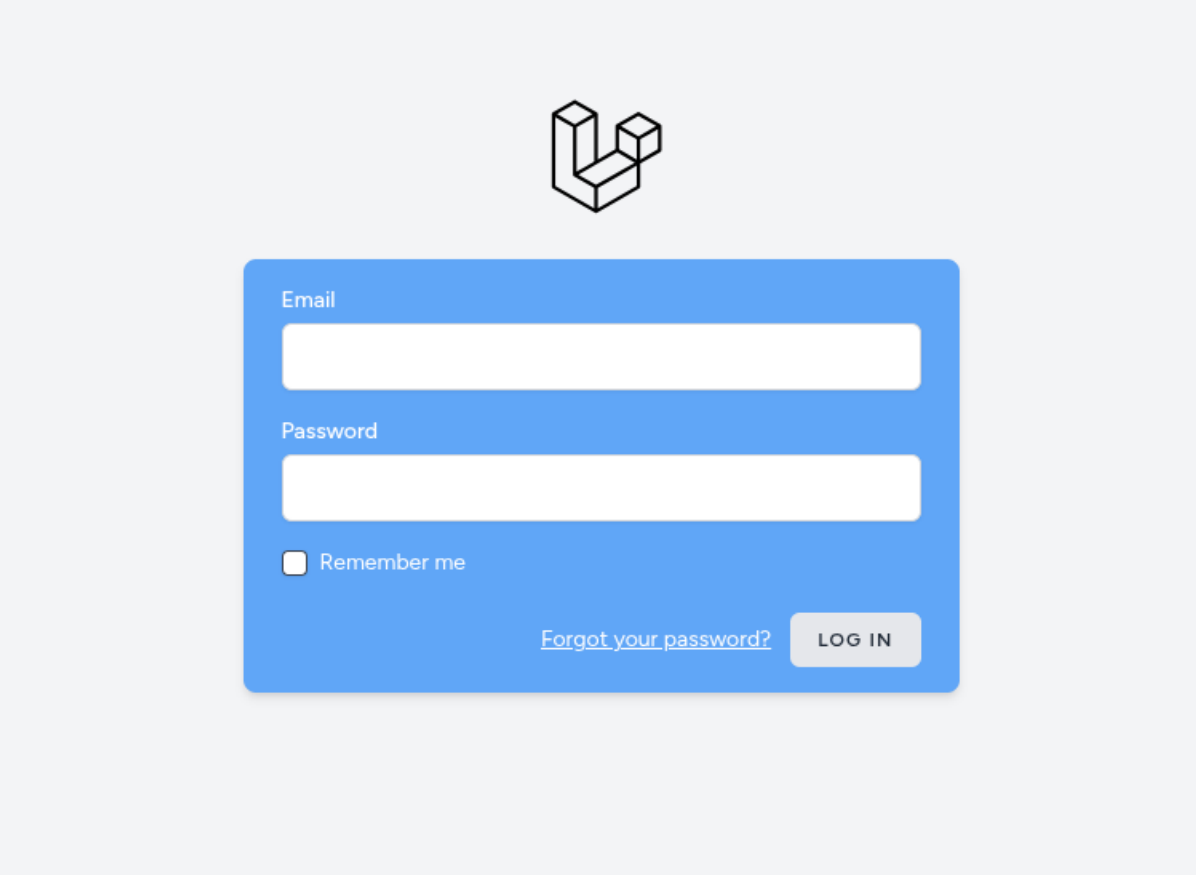
Fonte: Elaborado pelo autor.

3.3 Integração ao Sistema Web

Após finalizar a construção dos *dashboards*, foi necessário garantir um meio seguro e centralizado para disponibilizá-las aos usuários finais. Para isso, desenvolveu-se um sistema web utilizando o *framework Laravel*, no qual as visualizações foram incorporadas por meio de *iframes*. Para a visualização do sistema web desenvolvido para integrar o *iframe*, o código fonte foi disponibilizado em um repositório, no GitHub (SOUSA, 2025). Para garantir a segurança, foi implementado um sistema de autenticação de usuários. Desse modo, foi possível restringir o acesso, oferecendo maior controle e segurança.

Foi utilizado o comando *composer create-project laravel/laravel system* para criar o projeto. Ele foi o responsável por criar as pastas e a base estrutural do sistema. Essa estrutura inicial permitiu a rápida organização dos arquivos de *backend*, *frontend* e configuração. Em seguida, a autenticação foi realizada com o pacote *Laravel Breeze*, que fornece um sistema leve e pronto para uso com suporte a login, registro e verificação de e-mail (Laravel Documentation Team, 2024). Os comandos utilizados estão listados em B. Pode-se visualizar a tela de login na Figura 9.

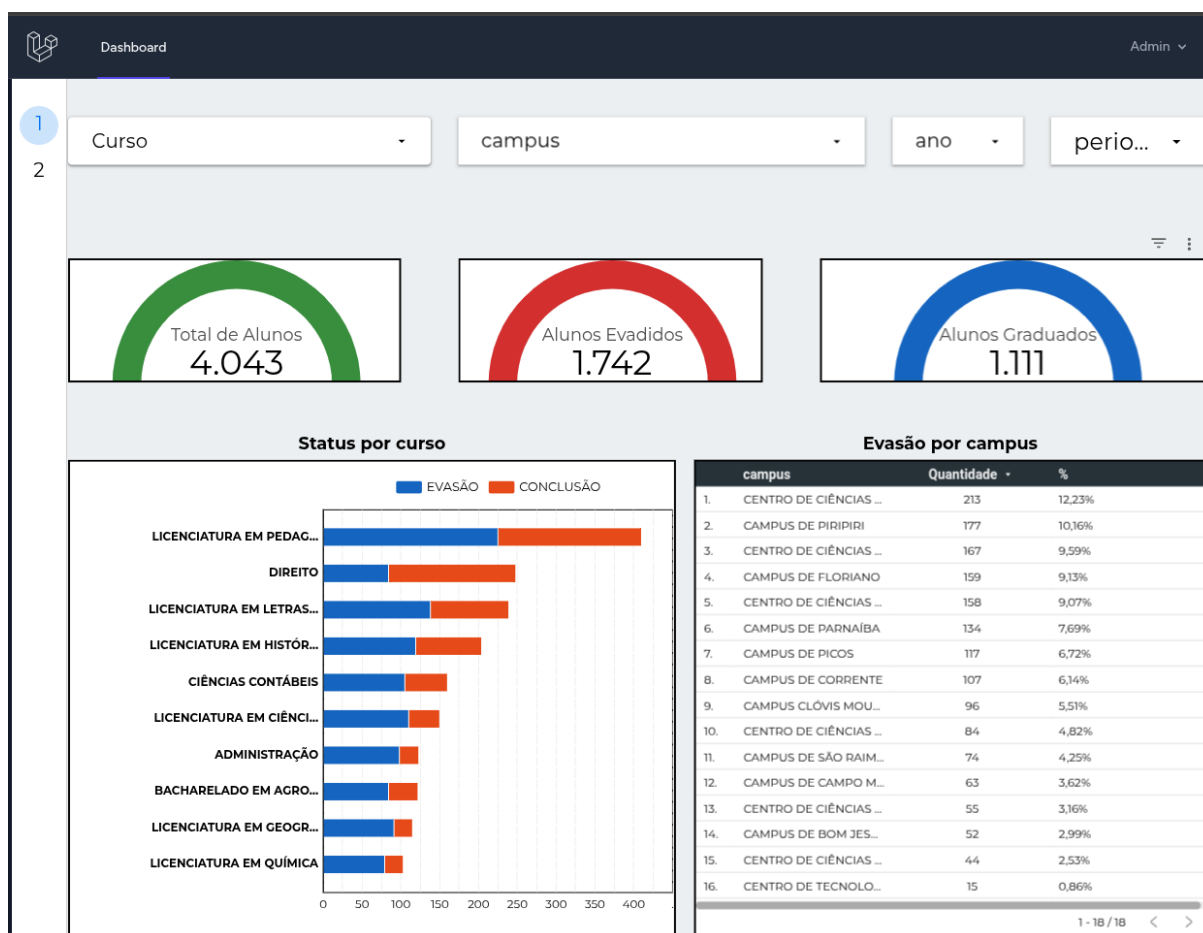
Figura 9 – Tela de Login



Fonte: Elaborado pelo autor.

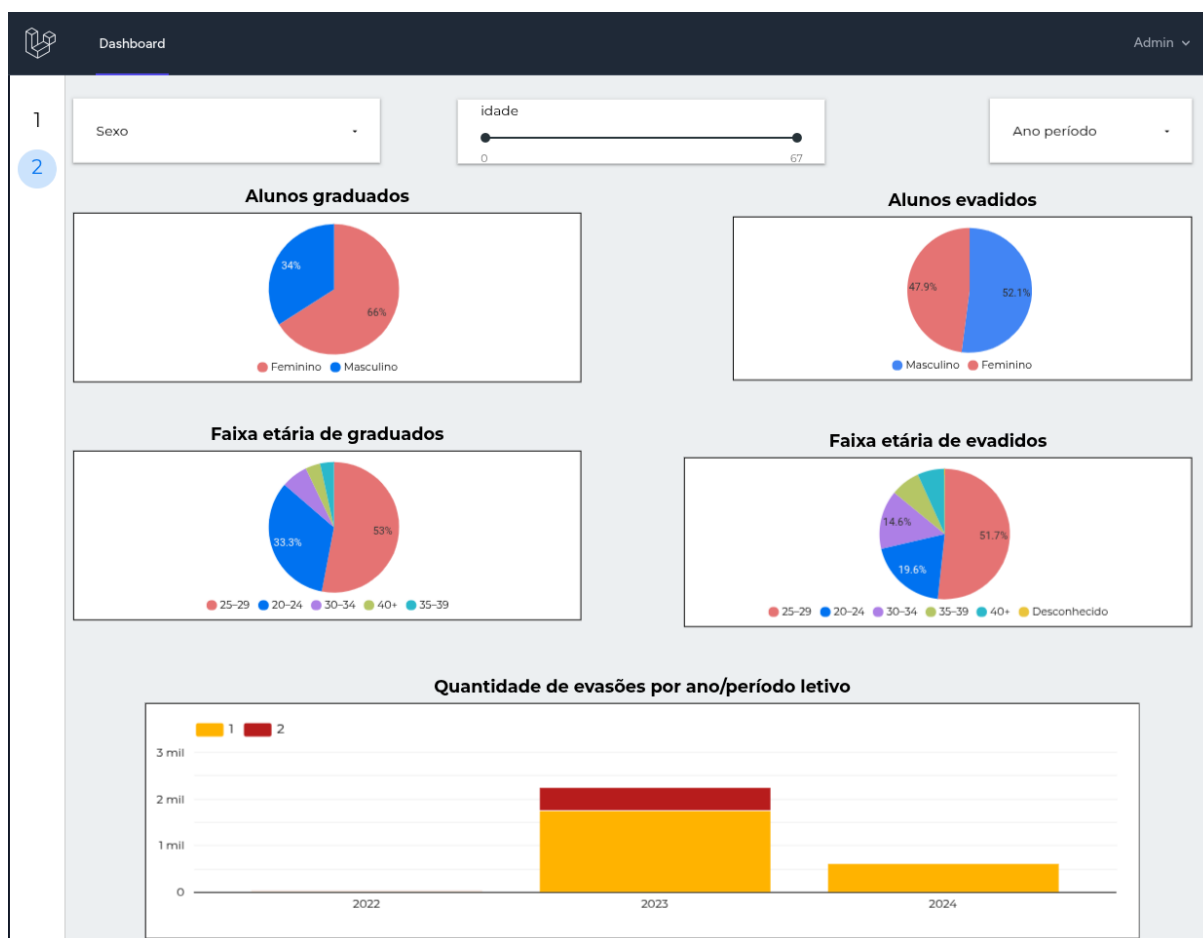
Após autenticado, o usuário logado consegue visualizar os *dashboards* em duas telas: na Figura 10, na qual é possível analisar a quantidade total de alunos ingressados em 2019.1 e 2019.2, o total de evasões e conclusões e o quantitativo de evadidos por campus e curso; e na Figura 11, que apresenta a porcentagem de graduados e evadidos separada por sexo e faixa etária e o gráfico de barras com a quantidade de evasões por período letivo.

Figura 10 – Status Geral por Curso e Campus



Fonte: Elaborado pelo autor.

Figura 11 – Indicadores Demográficos e Temporais da Evasão e Conclusão



Fonte: Elaborado pelo autor.

3.4 Validação

Concluído o processo de confecção dos painéis e integração no Sistema Web, a ferramenta criada passou pela validação. Processo no qual a equipe do Departamento de Tecnologia da Informação e Comunicação (DTIC) e um professor da instituição (UESPI) utilizaram-na, fazendo considerações acerca da funcionalidade e eficácia. Além disso, como método de validação, as informações geradas nos painéis da ferramenta criada foram comparadas com os dados obtidos manualmente nos sistemas SIGAA e CA.

A ferramenta mostrou-se eficaz e funcional, facilitando o acesso a dados complexos e de difícil obtenção. Quando comparados aos dados do próprio sistema de gestão da Universidade Estadual do Piauí, os dados fornecidos pela ferramenta mostraram-se compatíveis, validando e afirmando a aplicabilidade da técnica de *Business Intelligence* para a visualização dos indicadores de evasão escolar no ensino superior.

4 RESULTADOS

Esta seção apresenta os principais resultados obtidos a partir dos dados processados e organizados no *Data Warehouse*, conforme o método descrito no Capítulo 4. Os dados foram visualizados por meio de *dashboards* interativos, que permitem a identificação de importantes padrões a respeito da evasão escolar no ensino superior da Universidade Estadual do Piauí (UESPI).

Conforme discutido anteriormente, os dados analisados neste trabalho referem-se a alunos ingressantes nos períodos de 2019.1 e 2019.2, distribuídos entre diferentes campi da Universidade Estadual do Piauí (UESPI). As informações sobre evasão escolar abrangem o intervalo de 2019.1 a 2024.2. Dessa forma, o estudo considerou exclusivamente os alunos do ensino presencial que ingressaram em 2019 e apresentaram registros de evasão e conclusão em períodos regulares até o semestre 2024.2. Os dados extraídos dos sistemas *SIGAA* e *CA* foram utilizados, conforme descrito na Seção 4.2, para a construção dos *dashboards* interativos, posteriormente integrados a um Sistema Web, como apresentado na Seção 4.3.

O primeiro *Dashboard*, ilustrado na Figura 10, apresenta três medidores (Figura 4), que indicam um total de 4.043 alunos, dos quais 1.111 já concluíram seus cursos e 1.742 evadiram. Assim, a taxa de evasão entre os discentes ingressantes em 2019.1 e 2019.2 é de 43,08%. O mesmo *dashboard* contém um gráfico de barras (Figura 2), que detalha os quantitativos de evasões e conclusões por curso, além de uma tabela com os campi e suas respectivas quantidades de evasão e porcentagens (Figura 3). De acordo com esses dados, o Centro de Ciências Humanas e Letras (CCHL) apresenta o maior índice de evasão, com 213 registros, enquanto o curso de Licenciatura em Pedagogia lidera com 225 evasões.

Já o segundo *Dashboard*, ilustrado na Figura 11, contempla dois gráficos de pizza (Figura 6), que demonstram a distribuição de alunos graduados e evadidos detalhados por sexo e faixa etária. Os resultados revelam que a maioria dos evadidos são homens (52,1%) e, na totalidade dos evadidos, a grande maioria possui entre 25 a 29 anos (51,7%). Por fim, esse *dashboard* inclui também um gráfico de barras (Figura 7), que detalha a quantidade de evasões por período letivo. Observa-se um pico no primeiro semestre de 2023, indicando que uma parcela significativa dos alunos ingressantes em 2019.1 e 2019.2 evadiu em 2023.1.

De forma geral, os resultados apresentados fornecem uma visão clara dos cursos e campi com maiores índices de evasão, além de mostrar um perfil predominante dos alunos evadidos. Essas informações serão discutidas de forma mais aprofundada

no próximo capítulo, à luz do referencial teórico e dos trabalhos relacionados.

5 DISCUSSÃO

A análise dos dados evidenciou padrões importantes sobre a evasão escolar entre os alunos ingressantes em 2019 na Universidade Estadual do Piauí (UESPI). Nesta seção, busca-se interpretar esses resultados, considerando aspectos institucionais, demográficos e temporais, de acordo com a literatura apresentada nos capítulos anteriores.

Mais do que descrever os números, o objetivo é compreender os fatores que possivelmente contribuíram para os altos índices de evasão observados, relacionando-os a estudos já consolidados na área e às especificidades do contexto educacional da UESPI. Dessa forma, pretende-se identificar implicações práticas e pontos críticos que podem subsidiar futuras ações de enfrentamento à evasão.

5.1 Alta Taxa de Evasão

A taxa de evasão encontrada entre os alunos ingressantes nos dois períodos de 2019 da UESPI alcançou 43,08%, um percentual elevado que reforça a gravidade do cenário de evasão escolar no ensino superior. Além disso, esse índice mantém certa linearidade com os dados do primeiro semestre de 2024, que apresentaram uma taxa de evasão de 57,2% em todo o Brasil, incluindo as modalidades de ensino presencial e à distância (EaD) e de redes públicas e privadas (Correio Braziliense, 2024), evidenciando a persistência do problema ao longo dos períodos letivos.

Como ilustrado na Figura 7, uma parcela expressiva (91,38%) das evasões foi registrada no período de 2023.1. Na literatura, autores como Ambiel (2015) e Silva et al. (2022) destacam que os primeiros semestres são especialmente críticos, momento em que muitos alunos enfrentam dificuldades de adaptação, frustração com o curso escolhido e questões financeiras ou emocionais, o que justificaria evadir logo no início. No entanto, em análise mais aprofundada com o DTIC, entendeu-se que esse número é um tanto quanto inexato. Isso se deve, principalmente, à troca de sistemas citada na Seção 3.1.1. Constatou-se que, no momento da migração de sistemas, foram identificadas matrículas inativas, isto é, de discentes que abandonaram o curso sem formalização. Essas matrículas foram canceladas automaticamente pelo novo sistema. Dessa maneira, embora os dados indiquem um expressivo número de evasões em 2023.1, tais registros não correspondem, necessariamente, a evasões ocorridas nesse intervalo, haja vista que o sistema cancelou matrículas inativas de forma retroativa, sem indicar em qual período letivo houve o desligamento. De acordo com isso, embora os

gráficos referentes às evasões tendo como detalhamento o período letivo e/ou tempo de curso sejam essenciais, essas variáveis não puderam ser utilizadas dado o recorte temporal da pesquisa, uma vez que os números encontrados não expressaram, de fato, os números reais. Em outra ocasião, utilizando um outro recorte temporal, por exemplo, pode-se utilizar a evasão detalhada por tempo de curso, visando analisar em que momento de cada curso os alunos costumam evadir mais e estabelecer um padrão, como fizeram Bean (1980) e Ambiel (2015), possibilitando a aplicação de políticas de permanência de maneira nichada, focando nos padrões encontrados.

Os dados do segundo *Dashboard* (Figura 11) mostram que a maioria dos evadidos tem de 25 a 29 anos, o que pode indicar a conciliação da vida acadêmica com outras responsabilidades, como família, trabalho, dentre outros motivos. Essa realidade, atrelada às poucas políticas de acolhimento e nivelamento na UESPI, pode contribuir para o aumento das evasões. A reversão desse quadro depende, em grande parte, de intervenções já no primeiro período (Alves; Fidalgo, 2021).

Por fim, a evasão compromete a efetividade dos cursos, o planejamento institucional e o alocamento de recursos. São necessárias medidas que favoreçam a permanência, como ressaltam Alves, Gaydeczka e Campos (2018).

5.2 Indicadores da Evasão

De maneira geral, o estudo extraiu e processou dados de maneira precisa. Os *dashboards* criados transformaram dados complexos e de várias fontes em informações acuradas, com fácil visualização e interpretação.

É possível, como discutido anteriormente, determinar quais cursos e campi têm maior evasão, o sexo e a idade média dos evadidos e ter uma visão geral da quantidade de ingressantes, evadidos e formados. No entanto, como exposto na Seção 1.4, buscou-se trabalhar numa abordagem quali-quantitativa, à luz de conceitos de Bean (1980) e Ambiel (2015) discutidos na Seção 2.1.

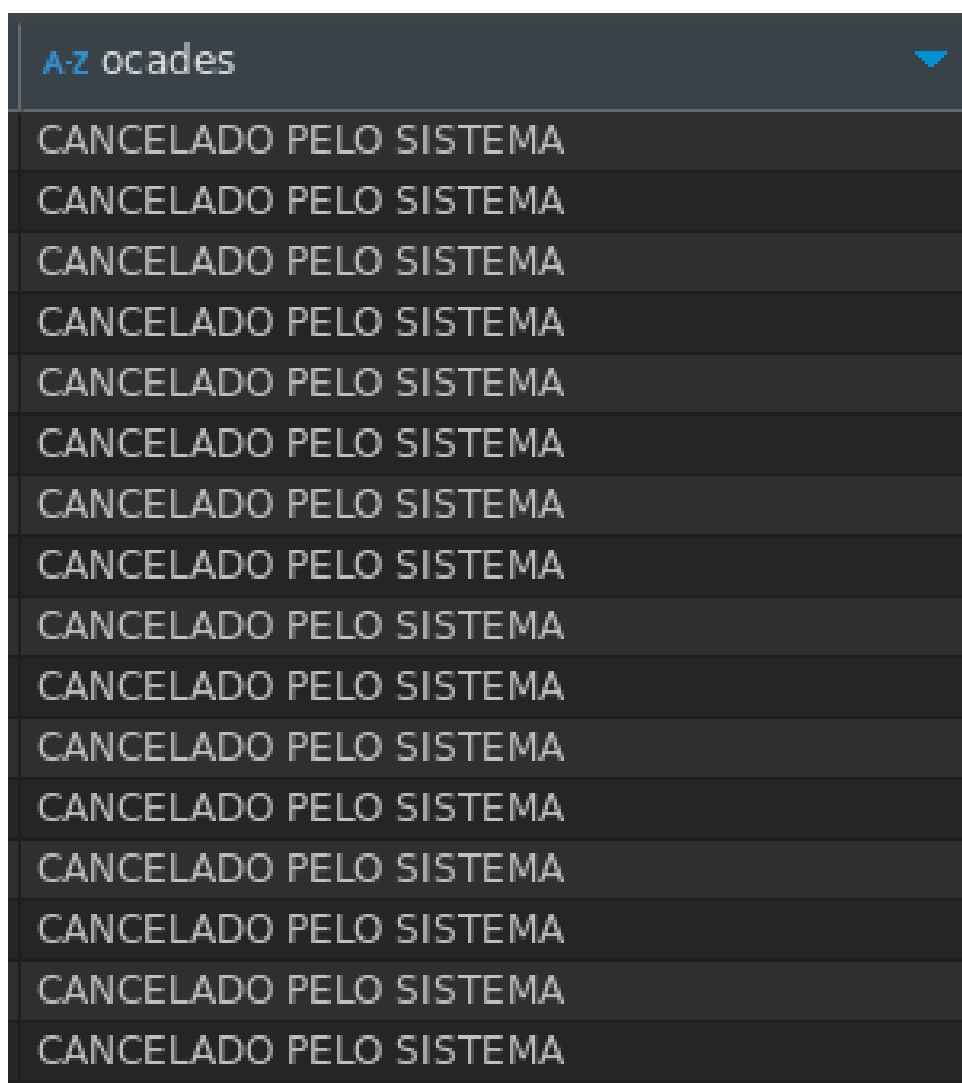
Entretanto, o sistema é recente e não possui dados quantitativos exatos, como discutido na Seção 6.1, não sendo possível identificar com maior certeza as quantidades exatas de evasão por período. Além disso, a ausência de dados qualitativos nos sistemas *SIGAA* e *CA* impossibilitou uma análise mais profunda acerca dos fatores subjetivos que levam o aluno à evasão, como as questões de ordem emocional, social, econômica e institucional, como propõem Bean (1980) e Ambiel (2015). A falta de informações oriundas de entrevistas, questionários abertos ou registros descritivos impossibilitou a triangulação de dados que possibilitaria compreender a situação de forma mais ampla, para além dos números. Além de um controle maior e mais recorrente das matrículas inativas, a fim de declarar a evasão num período certo, o sistema

carece do fornecimento de informações a respeito do aluno.

Apesar de oferecer uma visualização robusta de dados quantitativos da evasão escolar nos campi da Universidade Estadual do Piauí, não foi possível precisar de forma qualitativa os motivos por trás dos casos.

Embora, de acordo com o INEP (2017) e Alves e Fidalgo (2021), o termo *evasão escolar* possua ampla abrangência, tratando de pessoas que desistem ou evadem em qualquer momento do curso, os sistemas tratam da evasão de maneira superficial. No Quadro 3, da Seção 4.1.1.1, são descritos os motivos que podem ser declarados no SIGAA e CA como razões pelas quais o aluno evadiu. O sistema, entretanto, não qualifica os motivos, como proposto pelos autores Bean (1980) e Alves e Fidalgo (2021), já discutidos nesta seção e, anteriormente, na Seção 2.1. A grande maioria das evasões aparece como *CANCELADO PELO SISTEMA*, como demonstra a Figura 12.

Figura 12 – Descrição da Tabela de Ocorrências



A-Z ocades
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA
CANCELADO PELO SISTEMA

Fonte: Elaborado pelo autor.

Desse modo, o trabalho, por força maior, acaba limitando-se a uma visão mais quantitativa, não sendo possível visualizar nos painéis as razões específicas do cancelamento e seus respectivos números.

6 CONCLUSÕES

O presente trabalho teve como objetivo desenvolver uma solução de *Business Intelligence* para analisar os dados referentes à evasão escolar no ensino superior da Universidade Estadual do Piauí (UESPI). Para isso, foram estabelecidos e cumpridos os seguintes objetivos específicos: primeiro foi realizada a extração, o tratamento e a integração dos dados oriundos dos sistemas *SIGAA* (PostgreSQL) e *CA* (SQL Server); depois, os dados foram organizados em um modelo de dados dimensional, o modelo estrela (*Star Schema*), possibilitando uma análise com estrutura otimizada; em seguida, foram desenvolvidos painéis interativos utilizando a ferramenta *Google Looker Studio*, que proporcionou visualização acessível dos indicadores básicos (local, curso, sexo e faixa etária com maiores índices de evasão); por fim, foi feita a integração desses painéis em um Sistema Web com autenticação de usuários, garantindo segurança e acessibilidade.

A ferramenta desenvolvida, embora não apresente fatores subjetivos preponderantes para a evasão, permite à gestão uma visualização estratégica sobre a evasão, como a distribuição por cursos, campus, faixa etária e sexo, além da evolução temporal. Apesar da limitação quanto à ausência de dados qualitativos, a ferramenta apresenta um diagnóstico consistente e pode fomentar políticas educacionais de permanência. Desse modo, o trabalho alcança parte de seus propósitos e contribui como ferramenta sólida e escalável para auxiliar no enfrentamento da evasão escolar no contexto da UESPI.

6.1 Contribuições

Como discutido anteriormente, o trabalho contribui para a visualização de dados robustos e quantitativos, que auxiliam no enfrentamento da problemática. Com a ferramenta desenvolvida, é possível analisar e avaliar indicadores, identificando padrões nos evadidos no que diz respeito à faixa etária, sexo, campus e curso. Dessa maneira, o estudo contribui para o aprimoramento da gestão acadêmica, com uma solução na qual é possível visualizar dados complexos de maneira intuitiva e acessível, além de uma análise preditiva, ajudando a mitigar a problemática.

6.2 Limitações e Trabalhos Futuros

Apesar dos resultados alcançados, este trabalho apresenta algumas limitações. A maior limitação encontrada no desenvolvimento do estudo foi a ausência de dados

qualitativos acerca das evasões, o que impossibilitou a análise aprofundada dos fatores subjetivos ligados à evasão, como questões emocionais, financeiras e familiares, entre outras. Essa limitação impediu a abordagem qualitativa inicialmente prevista, impossibilitando uma visão mais holística do fenômeno. Além disso, por se tratar de um recorte temporal limitado aos ingressantes de 2019 e evasões até 2024.2, os resultados refletem apenas uma pequena parcela do universo acadêmico da UESPI. Assim, embora os painéis desenvolvidos contribuam para um diagnóstico inicial robusto, eles se limitam à dimensão quantitativa. Investigações futuras poderão aprofundar esse olhar com a inclusão de métodos qualitativos complementares, especialmente em articulação com os setores pedagógicos e assistenciais da universidade. Em síntese, trabalhos futuros poderão complementar esta análise com a inclusão de instrumentos como entrevistas, questionários e até a integração com bases de dados de outras instituições, de maneira a ampliar a perspectiva de análise e contribuir para ações mais direcionadas de permanência estudantil.

REFERÊNCIAS

ALVES, G. M. R.; FIDALGO, R. do N. Um ambiente de data warehousing para apoiar a tomada de decisão quanto a evasão escolar na UFPE. 2021. Citado 4 vezes nas páginas 15, 24, 47 e 48.

ALVES, M. d. O. P.; GAYDECZKA, B.; CAMPOS, A. de. Projeto para registro e controle da evasão na uftm. *Revista Triângulo*, v. 11, n. 1, p. 125–135, 2018. Citado 3 vezes nas páginas 19, 20 e 47.

AMBIEL, R. A. Construção da escala de motivos para evasão do ensino superior. *Avaliação Psicológica*, Instituto Brasileiro de Avaliação Psicológica, v. 14, n. 1, p. 41–52, 2015. Citado 4 vezes nas páginas 15, 20, 46 e 47.

BEAN, J. P. Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in higher education*, Springer, v. 12, p. 155–187, 1980. Citado 3 vezes nas páginas 20, 47 e 48.

CARDOSO, F. R. Implementação de arquitetura google cloud para análise e visualização de dados. UFMA, 2024. Citado na página 23.

Correio Braziliense. *Visão do Correio: Progresso se alcança com educação*. 2024. Disponível em: <<https://www.correio braziliense.com.br/opinia o/2024/05/6856645-visao-do-correio-progresso-se-alcanca-com-educacao.html>>. Citado 2 vezes nas páginas 15 e 46.

DURKHEIM, E. *O Suicídio: Estudo de Sociologia (Trad. M. Stahel)*. [S.l.]: Sao Paulo: Martins Fontes.(Trabajo original 1897), 2000. Citado na página 19.

DUTRA, R. D. Aplicação da ferramenta google looker studio para análise e visualização de dados de um sistema de gestão de pesquisa ambiental. *Derbyana*, v. 45, 2024. Citado na página 23.

FEW, S.; EDGE, P. Data visualization effectiveness profile. *Perceptual Edge*, v. 10, p. 12, 2017. Citado na página 23.

FRITSCH, R.; ROCHA, C. S. d.; VITELLI, R. F. A evasão nos cursos de graduação em uma instituição de ensino superior privada. *Revista Educação em Questão*, v. 52, n. 38, p. 81–108, 2015. Citado na página 19.

Gartner Group. *Business Intelligence*. 2024. Acessado em 10 de dezembro de 2024. Disponível em: <<http://www.gartner.com/it-glossary/business-intelligence-bi/>>. Citado na página 21.

GONÇALVES, H. H. M. Evasão nas instituições de ensino superior públicas do brasil: uma análise quantitativa. *RECIMA21-Revista Científica Multidisciplinar-ISSN 2675-6218*, v. 5, n. 8, p. e585432–e585432, 2024. Citado na página 21.

Google Cloud. *Documentação do Looker Studio*. 2024. Acesso em: 28 maio 2025. Disponível em: <<https://cloud.google.com/looker/docs/studio?hl=pt-br>>. Citado 2 vezes nas páginas 23 e 37.

HAN, J.; PEI, J.; TONG, H. *Data mining: concepts and techniques*. [S.l.]: Morgan kaufmann, 2022. Citado na página 22.

INEP. Metodologia de cálculo dos indicadores de fluxo da educação superior. *Technical report, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), Brasília*, 2017. Citado 2 vezes nas páginas 15 e 48.

INMON, W. H.; LINSTEDT, D. *Data architecture: a primer for the data scientist: big data, data warehouse and data vault*. [S.l.]: Morgan Kaufmann, 2014. Citado na página 22.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). *Inep lança painel de estatísticas do Censo Superior*. 2023. Acessado em 10 de dezembro de 2024. Disponível em: <https://www.gov.br/inep/pt-br/assuntos/noticias/censo-da-educacao-superior/inep-lanca-painel-de-estatisticas-do-censo-superior?utm_source=chatgpt.com>. Citado na página 22.

JÚNIOR, O. d. G. F. et al. Uma experiência com business intelligence para apoiar a gestão acadêmica em uma universidade federal brasileira. *Revista Ibérica de Sistemas e Tecnologias de Informação*, Associação Ibérica de Sistemas e Tecnologias de Informacao, n. 46, p. 5–20, 2022. Citado na página 21.

KHAN, B. et al. An overview of etl techniques, tools, processes and evaluations in data warehousing. *Journal on Big Data*, v. 6, 2024. Citado na página 26.

KIMBALL, R.; ROSS, M. *The data warehouse toolkit: The definitive guide to dimensional modeling*, ed. wiley. Morgan Kaufmann (July 9, 2008), 2019. Citado na página 22.

Laravel Documentation Team. *Laravel Breeze - Starter Kits*. 2024. Acesso em: 28 maio de 2025. Disponível em: <<https://laravel.com/docs/10.x/starter-kits>>. Citado na página 41.

MAIA, M. de C.; MEIRELLES, F. de S.; PELA, S. K. Análise dos índices de evasão nos cursos superiores a distância do brasil. 2004. Citado na página 15.

NATIONS, U. *Human Development Index*. 2024. Disponível em: <https://hdr.undp.org/data-center/human-development-index?_gl=1>. Citado na página 16.

OLSON, D. L.; LEE, S.-H. *Data warehousing*. Citeseer, 2015. Citado na página 16.

Power BI Report. *Dashboard de Indicadores de Evasão Escolar no Ensino Superior*. 2023. Acessado em 10 de dezembro de 2024. Disponível em: <<https://www.gov.br/inep/pt-br/assuntos/noticias/censo-da-educacao-superior/inep-lanca-painel-de-estatisticas-do-censo-superior>>. Citado na página 21.

RISTOFF, D. O novo perfil do campus brasileiro: uma análise do perfil socioeconômico do estudante de graduação. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, v. 19, n. 03, p. 723–747, 2014. Citado na página 19.

SANTOS, J. S. d. Business intelligence: uma proposta metodológica para análise da evasão escolar em instituições federais de ensino. *Universidade Federal do Paraná, Curitiba, Paraná*, 2017. Citado na página 21.

SARIKAYA, A. et al. What do we talk about when we talk about dashboards? *IEEE transactions on visualization and computer graphics*, IEEE, v. 25, n. 1, p. 682–692, 2018. Citado na página 23.

SILVA, A. G. da et al. Aplicação de uma solução de business intelligence para apoiar a gestão acadêmica em universidades federais brasileiras. *Revista Multidisciplinar do Nordeste Mineiro*, v. 6, n. 1, 2023. Citado na página 24.

SILVA, D. B. d. et al. Evasão no ensino superior público do brasil: estudo de caso da universidade de são paulo. *Avaliação: Revista da Avaliação da Educação Superior (Campinas)*, Publicação da Rede de Avaliação Institucional da Educação Superior (RAIES), da Universidade Estadual de Campinas (UNICAMP) e da Universidade de Sorocaba (UNISO)., v. 27, n. 2, p. 248–259, May 2022. ISSN 1414-4077. Disponível em: <<https://doi.org/10.1590/S1414-40772022000200003>>. Citado 3 vezes nas páginas 19, 20 e 46.

SILVA, L. C. et al. Selection of a business process management system: An analysis based on a multicriteria problem. In: IEEE. *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.], 2014. p. 295–299. Citado na página 21.

SOUSA, V. J. C. d. *Uso de Business Intelligence para Visualização de Indicadores de Evasão Escolar no Ensino Superior*. 2025. Disponível em: GitHub. Acesso em: 16 jun. 2025. Citado na página 41.

TINTO, V. Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research*, Sage Publications Sage CA: Thousand Oaks, CA, v. 45, n. 1, p. 89–125, 1975. Citado na página 19.

APÊNDICE A – SCRIPTS SQL

UTILIZADOS PARA O *DATA*

WHAREHOUSE

Listing A.1 – Script de Coleta da Base de Dados Sigaa Para a Dimensão Aluno

```

1 WITH movimentacoes_prioritarias AS (
2     SELECT
3         d.matricula,
4         p.data_nascimento,
5         p.sexo,
6         p.cidade,
7         p.uf,
8         ROW_NUMBER() OVER (
9             PARTITION BY d.matricula
10            ORDER BY
11                CASE
12                    WHEN ma.id_tipo_movimentacao_aluno IN (1, 315)
13                     THEN 1
14                    ELSE 2
15                END
16            ) AS prioridade
17 FROM ensino.movimentacao_aluno ma
18 INNER JOIN ensino.tipo_movimentacao_aluno tma
19     ON tma.id_tipo_movimentacao_aluno = ma.
20        id_tipo_movimentacao_aluno
21 INNER JOIN public.discente d
22     ON d.id_discente = ma.id_discente
23 INNER JOIN comum.pessoa p
24     ON p.id_pessoa = d.id_pessoa
25 INNER JOIN public.curso cur ON cur.id_curso = d.id_curso
26 WHERE ma.id_tipo_movimentacao_aluno IN (
27     1, 315,
28     2, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 19, 20,
29     306, 308, 309, 336, 339, 359, 358324
30 )
31 AND cur.id_modalidade_educacao = 1
32 AND d.ano_ingresso = 2019

```

```

31     AND d.periodo_ingresso IN (1, 2)
32     AND ma.data_ocorrencia >= DATE '2023-01-01'
33     AND ma.data_ocorrencia < DATE '2025-01-01'
34 )
35 SELECT *
36 FROM movimentacoes_prioritarias
37 WHERE prioridade = 1;

```

Listing A.2 – Script de Coleta de Alunos Graduados no CA

```

1 SELECT dcumat as matricula, dpe.dpenas as data_nascimento, dpe.dpesex
   as sexo, cid.cidnom as cidade, cid.cidsuf as uf FROM dcu
2 JOIN dpe ON dpe.dpe = dcudpe
3 JOIN cid ON cid.cid = dpe.dpecid
4 JOIN cur ON dcucur = cur.cur
5 JOIN des ON des.des = dcu.dcudes
6 WHERE (dcuian = 20191 OR dcuian = 20192)
7 AND cur.currcu = 1
8 AND dcudes = 'GRD';

```

Listing A.3 – Script de Coleta de Alunos Integralizados no CA

```

1 SELECT dcumat, dpe.dpenas, dpe.dpesex, cid.cidnom, cid.cidsuf
2 FROM dcu
3 JOIN dpe ON dcudpe=dpe
4 JOIN cid ON dpecid =cid
5 JOIN crd ON crdcur=dcucur AND crdncu=dcuncu
6 JOIN dis ON crddis=dis
7 JOIN cur ON dcucur=cur
8 left JOIN hes ON hesmat=dcumat AND crddis=hesdis AND hesapr='1' AND
   hessno!='SC'
9 where dcudes='NDE'
10 AND dpesex = 'M' or dpesex = 'F'
11 AND (dcuian = 20191 OR dcuian = 20192)
12 AND currcu = 1
13 group by dcumat, dpe.dpenas, dpe.dpesex, cid.cidnom, cid.cidsuf, dcu.
   dcuian
14 having sum(case when hesdis IS null then 1 else 0 end)=0
15 order by dcumat;

```

Listing A.4 – Script de Coleta de Alunos Evadidos no CA

```

1
2 SELECT DISTINCT

```



```

3      dcu.dcumat AS pk_aluno,
4      dpe.dpenas AS nascimento,
5      dpe.dpesex AS sexo,
6      cid.cidnom AS cidade,
7      cid.cidsuf AS uf
8 FROM dcu
9 JOIN dpe ON dpe.dpe = dcu.dcudpe
10 JOIN cid ON cid.cid = dpe.dpecid
11 JOIN oca ON oca.ocamat = dcu.dcumat
12 join cur on dcucur = cur.cur
13 JOIN oco ON oco.oco = oca.oacaoco
14 LEFT JOIN (
15     SELECT dcumat FROM dcu WHERE dcudes = 'GRD'
16 ) AS graduados ON graduados.dcumat = oca.ocamat
17 WHERE (dcuian = 20191 OR dcuian = 20192)
18     AND graduados.dcumat IS NULL
19     and cur.curruc = 1
20     AND oco.oco IN ('O', 'NFM', 'NMA', 'CAN', 'MC', 'JUB')
21 ORDER BY dcu.dcumat;

```

Listing A.5 – Script de Coleta dos Cursos no SIGAA

```

1 WITH movimentacoes_prioritarias AS (
2     SELECT
3         d.matricula,
4         d.id_curso,
5         ROW_NUMBER() OVER (
6             PARTITION BY d.matricula
7             ORDER BY
8                 CASE
9                     WHEN ma.id_tipo_movimentacao_aluno IN (1, 315)
10                      THEN 1
11                     ELSE 2
12                 END
13         ) AS prioridade
14 FROM ensino.movimentacao_aluno ma
15 INNER JOIN ensino.tipo_movimentacao_aluno tma
16     ON tma.id_tipo_movimentacao_aluno = ma.
17     id_tipo_movimentacao_aluno
18 INNER JOIN public.discente d
19     ON d.id_discente = ma.id_discente
20 WHERE ma.id_tipo_movimentacao_aluno IN (
21     1, 315,

```

```

20         2, 6, 7, 8, 9, 10, 11, 13, 14, 15, 16, 19, 20,
21         306, 308, 309, 336, 339, 359, 358324
22     )
23     AND d.ano_ingresso = 2019
24     AND d.periodo_ingresso IN (1, 2)
25 )
26 SELECT DISTINCT
27     c.id_curso AS pk_curso,
28     c.nome
29 FROM public.curso c
30 JOIN (
31     SELECT DISTINCT id_curso
32     FROM movimentacoes_prioritarias
33     WHERE prioridade = 1
34 ) cursos_prioritarios ON cursos_prioritarios.id_curso = c.id_curso
35 WHERE c.id_modalidade_educacao = 1;

```

Listing A.6 – Script de Coleta dos Cursos no CA

```

1
2 SELECT distinct cur.cur as id_curso, cur.curnom as curso, polnom as
   campus, pol as id_polo
3 FROM dcu
4 JOIN cur ON cur.cur = dcu.dcucur
5 JOIN pol ON dcu.dcupol = pol.pol
6 where currccu = 1 and dcumat NOT LIKE '%-%' AND dcumat in (
7 --lista de matriculas
8 );

```

Listing A.7 – Script de Inserção dos Alunos na Base SIGAA para a Tabela dim_aluno

```

1 INSERT INTO dim_aluno (
2     matricula,
3     data_nascimento,
4     sexo,
5     cidade,
6     uf
7 )
8 SELECT DISTINCT
9     matricula,
10    data_nascimento,
11    sexo,
12    cidade,

```

```

13      uf
14 FROM staging_aluno_sigaa;

```

Listing A.8 – Script de Inserção dos Alunos na Base CA para a Tabela dim_aluno

```

1 INSERT INTO dim_aluno (
2     matricula,
3     data_nascimento,
4     sexo,
5     cidade,
6     uf
7 )
8 SELECT DISTINCT
9     ca.matricula,
10    ca.data_nascimento,
11    ca.sexo,
12    ca.cidade,
13    ca.uf
14 FROM staging_aluno_ca ca
15 LEFT JOIN dim_aluno da ON da.matricula = ca.matricula
16 WHERE da.matricula IS NULL;

```

Listing A.9 – Script de Atualização das Colunas id_curso_legacy e id_polo_legacy

```

1 UPDATE staging_curso_sigaa
2 SET
3     id_curso_legacy = CAST(SUBSTRING_INDEX(SUBSTRING_INDEX(codmerg, '_'
4     , 1), '#', -1) AS UNSIGNED),
5     id_polo_legacy  = CAST(SUBSTRING_INDEX(codmerg, '_', -1) AS
6     UNSIGNED);

```

Listing A.10 – Script de Atualização das Colunas id_curso_legacy e id_polo_legacy

```

1 UPDATE staging_curso_ca ca
2 JOIN (
3     SELECT DISTINCT id_polo_legacy, campus
4     FROM staging_curso_sigaa
5 ) sigaa ON sigaa.id_polo_legacy = ca.id_polo
6 SET ca.campus = sigaa.campus;

```

Listing A.11 – Scripts de Inserção na Tabela dim_curso

```

1 INSERT INTO dim_curso (id_curso, nome, campus)
2 SELECT DISTINCT
3     concat(id_curso_legacy, id_polo_legacy), nome, campus

```

```
4 FROM staging_curso_sigaa
5 WHERE id_curso_legacy IS NOT NULL;
6
7 INSERT INTO dim_curso (id_curso, nome, campus)
8 SELECT DISTINCT
9     CAST(CONCAT(ca.id_curso, ca.id_polo) AS UNSIGNED) AS id_curso,
10    ca.nome,
11    ca.campus
12 FROM staging_curso_ca ca
13 LEFT JOIN dim_curso dc
14     ON dc.id_curso = CAST(CONCAT(ca.id_curso, ca.id_polo) AS UNSIGNED)
15 WHERE dc.id_curso IS NULL;
```

APÊNDICE B – COMANDOS PARA INSTALAÇÃO DO *LARAVEL BREEZE*

Listing B.1 – Comandos utilizados para gerar a autenticação

```
1 composer require laravel/breeze --dev
2 php artisan breeze:install
3 npm install && npm run dev
4 php artisan migrate
```