

Pedro Henrique Rocha de Castro

Data Warehouse de alunos do CEFET-MG e da RAIS (Relação Anual de Informações Sociais)

Belo Horizonte

2026

Pedro Henrique Rocha de Castro

Data Warehouse de alunos do CEFET-MG e da RAIS (Relação Anual de Informações Sociais)

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito para obtenção do título de bacharel em Engenharia de Computação.

Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG)
Engenharia de Computação

Orientador: Evandrino Gomes Barros
Coorientador: Ulisses Cotta Cavalca

Belo Horizonte
2026

PEDRO HENRIQUE ROCHA DE CASTRO

Data Warehouse de alunos do CEFET-MG e da RAIS (Relação Anual de Informações Sociais)

Trabalho de Conclusão de Curso apresentado ao curso de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais, como requisito para obtenção do título de bacharel em Engenharia de Computação.

Orientador: Evandrino Gomes Barros

Coorientador: Ulisses Cotta Cavalca

Trabalho aprovado em _____ de _____ de 2026.

Prof. Evandrino Gomes Barros – Orientador
CEFET-MG

Prof. Ulisses Cotta Cavalca – Coorientador
CEFET-MG

Prof. _____
CEFET-MG

C355d	Castro, Pedro Henrique Rocha de Data warehouse de alunos do CEFET-MG e da RAIS (Relação Anual de Informações Sociais) / Pedro Henrique Rocha de Castro. – 2026. 76 f. : il. Trabalho de conclusão de curso apresentado ao curso de Engenharia de Computação do Centro Federal de Educação Tecnológica de Minas Gerais. Orientador: Evandrino Gomes Barros. Coorientador: Ulisses Cotta Cavalca. Trabalho de conclusão de curso (graduação) – Centro Federal de Educação Tecnológica de Minas Gerais, Departamento de Computação. 1. Estudantes de engenharia – Minas Gerais – Teses. 2. Universidades e faculdades – Engenharia de Computação – Teses. 3. Inteligência competitiva (Administração) – Teses. 4. Repositórios institucionais – Teses. 5. Processamento de texto (Ciência da Computação) – Teses. 6. Painéis de controle – Teses. 7. Software de código aberto – Teses. I. Barros, Evandrino Gomes. II. Cavalca, Ulisses Cotta. III. Centro Federal de Educação Tecnológica de Minas Gerais. Departamento de Computação. IV. Título. CDD 006.37
-------	---

RESUMO

A produção de indicadores confiáveis sobre alunos é fundamental para subsidiar relatórios institucionais, processos de avaliação e o planejamento de ações de melhoria contínua. Contudo, em muitas instituições, os dados de alunos permanecem dispersos em fontes heterogêneas, com diferentes padrões e níveis de qualidade, dificultando análises integradas e atualizadas. Neste trabalho, apresenta-se o projeto, a implementação e os resultados de um *Data Warehouse* de alunos do CEFET-MG integrado a dados da Relação Anual de Informações Sociais (RAIS) obtidos por meio de convênio firmado entre o CEFET-MG e o Ministério do Trabalho e Emprego, em conformidade com a Lei Geral de Proteção de Dados Pessoais (LGPD), complementado por um painel interativo de visualização construído sobre essa base analítica. A solução é sustentada por um ambiente analítico baseado em *Data Warehouse* (DW), composto por uma tabela fato central e onze tabelas dimensão, alimentado por processos de *ETL* (*Extract, Transform, Load*) voltados à extração, padronização, validação e carga periódica de dados. O painel foi implementado com a ferramenta Metabase, organizando-se em seis seções temáticas: inserção no mercado, remuneração, equidade, localização geográfica, ocupação e vínculos empregatícios, com filtros globais interativos que permitem recortes por ano, campus, curso, UF e nível de ensino. Os resultados demonstram a viabilidade da integração entre dados acadêmicos e bases governamentais para a geração de indicadores de empregabilidade, contribuindo para uma gestão institucional orientada por dados no âmbito do CEFET-MG.

Palavras-chave: alunos; *business intelligence*; *Data Warehouse* ; *data warehouse*; ETL; visualização de dados; *dashboards*; RAIS; Metabase.

ABSTRACT

The production of reliable indicators about students is fundamental to support institutional reports, evaluation processes, and planning of continuous improvement actions. However, in many institutions, student data remains dispersed in heterogeneous sources with different standards and quality levels, hindering integrated and up-to-date analyses. This paper presents the design, implementation, and results of a Data Warehouse for CEFET-MG students integrated with data from the Annual Social Information Report (RAIS) obtained through a formal agreement between CEFET-MG and the Ministry of Labor and Employment, in compliance with the Brazilian General Data Protection Law (LGPD), complemented by an interactive visualization dashboard built on top of this analytical base. The solution is supported by an analytical environment based on a Data Warehouse (DW), composed of one central fact table and eleven dimension tables, fed by ETL (Extract, Transform, Load) processes aimed at the extraction, standardization, validation, and periodic loading of data. The dashboard was implemented using Metabase, organized into six thematic sections: labor market insertion, remuneration, equity, geographic location, occupation, and employment contracts, with interactive global filters allowing data segmentation by year, campus, course, state, and education level. The results demonstrate the feasibility of integrating academic data with government databases for the generation of employability indicators, contributing to data-driven institutional management at CEFET-MG.

Keywords: students; business intelligence; Data Warehouse ; data warehouse; ETL; data visualization; dashboards; RAIS; Metabase.

SUMÁRIO

1	INTRODUÇÃO	7
2	REFERENCIAL TEÓRICO	9
2.1	Relação anual de informações sociais (RAIS)	9
2.2	Business intelligence no contexto educacional	10
2.3	Data Warehouse e modelagem de dados analítica	10
2.4	Processo de ETL (Extract, Transform, Load)	10
2.5	Qualidade e governança de dados	11
2.6	Visualização de dados e painéis interativos	12
2.7	Metabase como ferramenta de business intelligence	12
3	TRABALHOS RELACIONADOS	13
4	METODOLOGIA	16
4.1	Aquisição dos dados	16
4.2	Processo de ETL (Extract, Transform, Load)	17
4.3	Modelagem e implementação do Data Warehouse	19
4.3.1	Tabela Fato	20
4.3.2	Tabelas Dimensão	20
4.4	Construção do painel interativo no Metabase	21
4.4.1	Definição dos Indicadores	21
4.4.2	Consultas SQL e Filtros Dinâmicos	22
4.4.3	Filtros Globais do Dashboard	23
4.5	Validação dos dados e dos indicadores	24
4.6	Atualização e manutenção do ambiente analítico	25
5	RESULTADOS	26
5.1	Data Warehouse implementado	26
5.2	Consultas analíticas desenvolvidas	26
5.2.1	Consulta de Salário Médio por Curso	27
5.2.2	Consulta de Salário Médio e Mediana por Ocupação	28
5.2.3	Consulta de Salário Médio por Sexo e Ano	30
5.3	Painel interativo no Metabase	31
5.3.1	Barra de Filtros	32
5.3.2	Seção de Visão Geral	32
5.3.3	Seção de Inserção no Mercado	33

5.3.4	Seção de Remuneração	33
5.3.5	Seção de Equidade	34
5.3.6	Seção de Localização Geográfica	35
5.3.7	Seção de Ocupação	35
5.3.8	Seção de Vínculos	36
5.4	Principais achados analíticos	37
6	CONCLUSÃO	39
	REFERÊNCIAS	41

1 INTRODUÇÃO

A produção de indicadores confiáveis sobre alunos de instituições de ensino superior é fundamental para subsidiar relatórios institucionais, processos de avaliação interna e externa, bem como o planejamento de ações de melhoria contínua. No entanto, a integração e a análise sistemática de dados de alunos ainda representam um desafio em muitas instituições, uma vez que essas informações frequentemente se encontram dispersas em múltiplas bases e formatos heterogêneos, o que dificulta sua consolidação e exploração analítica.

O CEFET-MG enfrenta dificuldades para consolidar dados de alunos de maneira integrada e acessível. As informações podem estar distribuídas entre sistemas acadêmicos internos, registros administrativos, bases governamentais e levantamentos institucionais, com diferentes padrões e níveis de qualidade. O Ministério do Trabalho e Emprego reúne dados detalhados sobre o emprego formal por meio da Relação Anual de Informações Sociais (RAIS)¹. Embora os microdados agregados da RAIS sejam parcialmente públicos, o extrato contendo CPF — necessário para o cruzamento nominal com registros acadêmicos — é de caráter sigiloso e acessível apenas mediante convênio com o governo federal. Esse extrato não está integrado diretamente aos sistemas acadêmicos das instituições de ensino, e a ausência de ferramentas que integrem, tratem e apresentem esses dados de forma visual e interativa limita a geração de indicadores robustos de empregabilidade.

Nesse contexto, este trabalho tem como objetivo geral projetar, implementar e avaliar um painel interativo que integre dados provenientes do sistema acadêmico do CEFET-MG, em especial do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA), e da base RAIS, utilizando ferramentas de *Business Intelligence* (BI). Para viabilizar análises históricas e multidimensionais, foi construído um repositório analítico em *Data Warehouse* (DW), alimentado por um processo de *ETL* capaz de incorporar novos extratos anuais fornecidos via convênio. Com isso, foi possível aplicar procedimentos de tratamento e qualidade de dados e disponibilizar visualizações e filtros para análises por curso, área de atuação, empregabilidade, remuneração e outros recortes relevantes.

O painel foi implementado utilizando a ferramenta de BI Metabase, conectada diretamente ao *Data Warehouse* construído em PostgreSQL. O resultado é uma solução que oferece seis seções temáticas de análise, com filtros globais interativos e vinte e uma consultas analíticas documentadas, apoiando gestores, coordenadores de curso e setores responsáveis pelo acompanhamento de alunos do CEFET-MG.

A relevância prática do painel para a tomada de decisão institucional se manifesta em diferentes níveis de gestão. Para coordenadores de curso, os indicadores de remuneração, tempo até a inserção no mercado e aderência entre a formação e a ocupação exercida permitem identificar

¹ Página oficial da RAIS: <https://www.rais.gov.br/sitio/sobre.jsf>.

cursos cuja matriz curricular pode estar desalinhada com as demandas do mercado de trabalho, subsidiando revisões pedagógicas direcionadas. Para a gestão institucional, os indicadores de equidade — como o *gap* salarial por sexo e raça/cor — fornecem evidências quantitativas que podem orientar políticas afirmativas e ações de inclusão. A distribuição geográfica dos alunos apoia decisões sobre parcerias regionais e programas de estágio em municípios com maior concentração de empregabilidade. Por fim, os indicadores consolidados no painel atendem diretamente às exigências do SINAES quanto ao acompanhamento de alunos, podendo ser utilizados como insumo objetivo em processos de avaliação institucional interna e externa, substituindo levantamentos declarativos por evidências baseadas em registros administrativos oficiais.

A literatura registra iniciativas relacionadas em domínios próximos, mas nenhuma que reúna, simultaneamente, todos os elementos da abordagem aqui proposta. Trabalhos como o de Marchezi (2025) demonstram o potencial de processos de *ETL* e modelagem dimensional aplicados à RAIS para análises do mercado de trabalho formal, porém sem integração com dados acadêmicos individuais. Estudos sobre construção de *Data Warehouses* para apoio à decisão (FERREIRA, 2002; GURA; BENCK, 2011) consolidam boas práticas de modelagem em estrela, mas são voltados a contextos empresariais ou administrativos. Aplicações de *Data Warehouse* em bases governamentais de grande volume, como as do Sistema Único de Saúde (SANTOS; SILVA, 2014), evidenciam a viabilidade da abordagem em ambientes públicos complexos, ainda que em domínio distinto do educacional. No acompanhamento de alunos especificamente, a prática institucional mais comum ainda depende de questionários com baixa taxa de resposta e sem integração a registros administrativos oficiais. O presente trabalho se diferencia ao utilizar o extrato sigiloso da RAIS — obtido via convênio governamental e contendo CPF — para realizar o cruzamento nominal com os registros do SIGAA, viabilizando indicadores de empregabilidade com cobertura e confiabilidade superiores às abordagens declarativas.

Este documento está organizado da seguinte forma: o Capítulo 2 apresenta o referencial teórico; o Capítulo 3 discute os trabalhos relacionados; o Capítulo 4 descreve a metodologia adotada; o Capítulo 5 apresenta os resultados obtidos; e o Capítulo 6 traz as conclusões e perspectivas de trabalhos futuros.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta os principais conceitos e trabalhos relacionados ao acompanhamento de alunos, indicadores educacionais, *Business Intelligence* e visualização de dados. Também são abordados temas como painéis interativos (*dashboards*), modelagem de dados para análise em *Data Warehouse*, processos de *ETL* e boas práticas de qualidade e governança de dados no contexto de instituições de ensino superior.

2.1 Relação anual de informações sociais (RAIS)

A Relação Anual de Informações Sociais (RAIS) é um instrumento administrativo de coleta de dados instituído pelo governo federal brasileiro, sob responsabilidade do Ministério do Trabalho e Emprego¹. Seu objetivo principal é reunir informações detalhadas sobre o mercado de trabalho formal no país, servindo de base para a elaboração de políticas públicas, estudos estatísticos e controle das atividades trabalhistas.

A RAIS reúne dados declarados anualmente pelos empregadores, abrangendo informações sobre vínculos empregatícios, remuneração, ocupação, setor de atividade econômica, escolaridade, tempo de serviço e localização geográfica do emprego. Por se tratar de uma base censitária do emprego formal, a RAIS apresenta elevado nível de detalhamento e é amplamente utilizada em pesquisas acadêmicas e institucionais relacionadas à empregabilidade e renda (MINISTÉRIO DO TRABALHO E EMPREGO, 2024b; MARCHEZI, 2025).

No contexto do acompanhamento de alunos, a RAIS possibilita análises históricas da inserção profissional, permitindo identificar padrões de empregabilidade ao longo do tempo, evolução salarial e permanência no mercado de trabalho formal. O cruzamento entre registros acadêmicos institucionais e os vínculos empregatícios da RAIS é viabilizado pelo Cadastro de Pessoa Física (CPF), campo comum às duas bases, cujo uso para essa finalidade é regulado pela Lei Geral de Proteção de Dados Pessoais (LGPD) (BRASIL, 2018).

É importante distinguir dois tipos de acesso à RAIS: os microdados públicos, disponibilizados anualmente pelo Ministério do Trabalho e Emprego sem identificação nominal dos trabalhadores, e o extrato sigiloso, contendo o CPF de cada vínculo, acessível exclusivamente por meio de convênio firmado com o governo federal. O cruzamento nominal entre vínculos empregatícios e registros acadêmicos, como o realizado neste trabalho, depende exclusivamente do extrato sigiloso, cujo uso é condicionado por obrigações de sigilo, finalidade restrita e descarte seguro dos dados após o cruzamento. Os procedimentos de anonimização adotados para atendimento a essas exigências são detalhados no Capítulo 4.

¹ Página oficial da RAIS: <https://www.rais.gov.br/sitio/sobre.jsf>.

2.2 Business intelligence no contexto educacional

Business Intelligence (BI) refere-se ao conjunto de metodologias, processos, arquiteturas e tecnologias utilizadas para coletar, integrar, analisar e apresentar dados com o objetivo de apoiar a tomada de decisão (TURBAN; SHARDA; DELEN, 2019).

Segundo Turban, Sharda e Delen (2019), sistemas de BI permitem transformar grandes volumes de dados brutos em informações relevantes por meio de relatórios analíticos, painéis interativos e indicadores de desempenho. Em instituições de ensino superior, essas ferramentas podem ser aplicadas para análise de evasão, desempenho discente, planejamento acadêmico e acompanhamento de alunos (CAMPBELL; DEBLOIS; OBLINGER, 2007).

A adoção de soluções de BI contribui para a integração de dados provenientes de múltiplas fontes, reduzindo a dependência de relatórios estáticos e facilitando análises exploratórias e comparativas ao longo do tempo. Dessa forma, o BI se apresenta como uma abordagem adequada para lidar com a complexidade e o volume de dados relacionados aos alunos.

2.3 Data Warehouse e modelagem de dados analítica

O *Data Warehouse* (DW) é um repositório centralizado de dados, orientado por assunto, integrado, não volátil e variável no tempo, projetado para apoiar processos de análise e tomada de decisão (INMON, 2005). Diferentemente dos sistemas transacionais, que são otimizados para operações do dia a dia, o DW é estruturado para consultas analíticas complexas e históricas.

A modelagem de dados em ambientes de *Data Warehouse* geralmente utiliza esquemas dimensionais, como o modelo estrela ou floco de neve, propostos por Kimball e Ross (2013)². Nesses modelos, os dados são organizados em tabelas fato, que armazenam medidas quantitativas, e tabelas dimensão, que fornecem o contexto descritivo para análise.

No contexto do acompanhamento de alunos, um *Data Warehouse* permite integrar dados acadêmicos internos com dados externos, como os provenientes da RAIS, viabilizando análises multidimensionais por curso, período de formação, setor econômico, faixa salarial e tempo de inserção no mercado de trabalho.

2.4 Processo de ETL (Extract, Transform, Load)

O processo de *ETL* (*Extract, Transform, Load*) consiste em três etapas fundamentais para a construção de um *Data Warehouse*. A etapa de extração envolve a coleta dos dados a partir de diferentes fontes, como sistemas acadêmicos, bases governamentais e arquivos externos. Na etapa de transformação, os dados passam por procedimentos de limpeza, padronização, validação e enriquecimento, visando garantir a qualidade e a consistência das informações. Por

² O modelo estrela foi originalmente proposto na primeira edição da obra, publicada em 1996. A edição citada neste trabalho (2013) corresponde à terceira edição, revisada e atualizada pelos mesmos autores.

fim, na etapa de carga, os dados tratados são inseridos no repositório analítico (KIMBALL; ROSS, 2013).

A implementação de processos de *ETL* é essencial para garantir a atualização periódica de repositórios analíticos alimentados por bases externas divulgadas regularmente, como é o caso da RAIS. Além disso, o *ETL* desempenha um papel central na aplicação de regras de qualidade de dados, como tratamento de valores ausentes, remoção de duplicidades e padronização de atributos (BATINI; CERI; NAVATHE, 2009).

2.5 Qualidade e governança de dados

A qualidade dos dados é um fator crítico para a confiabilidade dos indicadores gerados por sistemas de BI. Segundo (BATINI; CERI; NAVATHE, 2009), dimensões como precisão, completude, consistência e atualidade são fundamentais para garantir que as análises reflitam adequadamente a realidade observada.

No ambiente educacional, a governança de dados envolve a definição de políticas, responsabilidades e padrões para a gestão dos dados institucionais. A adoção de boas práticas de governança contribui para a transparência, a confiabilidade e a segurança das informações, além de facilitar a integração entre diferentes sistemas e setores da instituição.

No caso do acompanhamento de alunos, a governança de dados é especialmente relevante devido à sensibilidade das informações pessoais e profissionais, exigindo cuidados adicionais com privacidade e conformidade legal.

Um aspecto central da governança de dados neste trabalho é o processo de anonimização aplicado às bases de entrada antes da carga no *Data Warehouse*. Duas informações de identificação direta foram tratadas: o Cadastro de Pessoa Física (CPF) e a matrícula do aluno no SIGAA. O CPF é utilizado exclusivamente como chave de cruzamento entre o extrato da RAIS e os registros acadêmicos — operação realizada na camada de *staging* — e não é transferido para nenhuma tabela do *Data Warehouse*. Após o cruzamento, o vínculo entre o trabalhador da RAIS e o aluno do SIGAA é representado apenas por uma chave *surrogate* interna (*id_aluno*), sem correspondência reversa possível a partir da camada analítica. A matrícula do aluno, igualmente identificadora, tampouco é armazenada no DW: os dados acadêmicos relevantes — curso, campus, nível de ensino — são mantidos sem qualquer atributo que permita reidentificar o indivíduo externamente. Esse processo está em conformidade com as exigências da Lei Geral de Proteção de Dados Pessoais (LGPD) (BRASIL, 2018) e com as cláusulas do convênio firmado entre o CEFET-MG e o Ministério do Trabalho e Emprego para acesso ao extrato sigiloso da RAIS, que determinam finalidade restrita, sigilo e descarte seguro após o uso.

2.6 Visualização de dados e painéis interativos

A visualização de dados desempenha um papel fundamental na análise e comunicação de informações complexas. De acordo com Few (2013), boas visualizações permitem identificar padrões, tendências e *outliers* de forma mais rápida e intuitiva do que relatórios textuais ou tabelas extensas.

Painéis interativos, também conhecidos como *dashboards*, são ferramentas que reúnem múltiplas visualizações em uma única interface, permitindo a aplicação de filtros e a exploração dinâmica dos dados. No contexto do BI educacional, *dashboards* possibilitam que gestores e coordenadores de curso acompanhem indicadores-chave de forma contínua e personalizada.

A utilização de painéis interativos para o acompanhamento de alunos contribui para uma análise mais aprofundada da trajetória profissional dos formados, apoiando decisões estratégicas e ações de melhoria alinhadas aos objetivos institucionais.

2.7 Metabase como ferramenta de business intelligence

O Metabase é uma ferramenta de BI que permite a criação de painéis interativos, consultas analíticas e visualizações de dados por meio de uma interface *web* acessível tanto a perfis técnicos quanto a usuários sem conhecimento de SQL (METABASE, 2024). A ferramenta suporta conexão direta com bancos de dados relacionais, incluindo PostgreSQL, e oferece recursos nativos para a criação de filtros globais em *dashboards*, que propagam automaticamente as seleções do usuário para todas as questões analíticas conectadas.

O Metabase adota um modelo de licenciamento dual: a versão de código aberto (*open-source*) é gratuita e cobre a maioria das necessidades de projetos acadêmicos, departamentais e pequenas e médias organizações; a versão comercial (*Pro/Enterprise*) adiciona recursos voltados a ambientes corporativos, como controle de acesso granular, *embedding* personalizado e suporte dedicado (METABASE, 2024). Para este trabalho, a versão *open-source* foi suficiente para todas as funcionalidades requeridas.

Uma das principais vantagens do Metabase é a velocidade de implantação e a curva de aprendizado reduzida para usuários finais. É possível subir uma instância via Docker ou em um servidor em nuvem em poucos minutos, conectá-la ao banco de dados e começar a gerar gráficos imediatamente, sem necessidade de configuração complexa. Usuários sem formação técnica conseguem explorar dados por meio de filtros e menus visuais, enquanto analistas e desenvolvedores podem escrever consultas SQL nativas com suporte a parâmetros dinâmicos por meio da sintaxe de *field filters*.

Entre as demais características relevantes para este projeto, destacam-se: a organização de visualizações em *dashboards* com *cards* de texto, gráficos e indicadores numéricos (KPIs); e a capacidade de compartilhamento e controle de acesso por perfis de usuário.

3 TRABALHOS RELACIONADOS

A utilização de técnicas de *Business Intelligence* e *Data Warehouse* para análise de dados tem sido amplamente explorada na literatura como suporte à tomada de decisão em diferentes contextos organizacionais, incluindo áreas administrativas, governamentais e educacionais (MARCHEZI, 2025; FERREIRA, 2002; GURA; BENCK, 2011; SANTOS; SILVA, 2014). Esses estudos demonstram que a consolidação de dados provenientes de múltiplas fontes em ambientes analíticos possibilita a geração de indicadores mais consistentes, além de favorecer análises históricas e multidimensionais.

No âmbito das bases governamentais brasileiras, diversos trabalhos investigam a aplicação dos dados da Relação Anual de Informações Sociais (RAIS) em ambientes analíticos. Estudos como o de Marchezi (2025) evidenciam como processos de *ETL* e modelagem dimensional permitem transformar grandes volumes de dados brutos em informações analíticas acessíveis por meio de ferramentas de visualização. Os resultados demonstram o potencial dessas bases para análises de admissões, desligamentos e remuneração no mercado de trabalho formal, embora o foco permaneça restrito a análises gerais do mercado, sem integração com dados acadêmicos.

Outros trabalhos concentram-se na construção de *Data Warehouses* como infraestrutura para apoio à tomada de decisão. Esses estudos descrevem metodologias para levantamento de requisitos, definição de esquemas dimensionais e implementação de processos de extração, transformação e carga de dados. A literatura destaca que o uso de esquemas em estrela contribui para a organização dos dados e para o desempenho das consultas analíticas (FERREIRA, 2002; GURA; BENCK, 2011). Entretanto, tais abordagens são frequentemente aplicadas a contextos empresariais ou administrativos, não explorando de forma específica o cenário das instituições de ensino superior.

Também se destacam estudos que aplicam *Data Warehouse* em áreas sociais e governamentais, como a saúde pública. Um trabalho relevante propõe a construção de um *Data Warehouse* a partir de dados do Sistema Único de Saúde (SUS), com o objetivo de apoiar gestores na análise de informações relacionadas a atendimentos, procedimentos e recursos de saúde (SANTOS; SILVA, 2014). Apesar de atuar em um domínio distinto do educacional, o estudo demonstra a viabilidade e a importância da aplicação de *Data Warehouse* em ambientes públicos complexos e com grande volume de dados.

No campo do *Business Intelligence*, diversos estudos ressaltam a importância dos painéis de indicadores como ferramentas centrais para a análise e comunicação de informações (FEW, 2013; TURBAN; SHARDA; DELEN, 2019). *Dashboards* interativos permitem que gestores acompanhem indicadores-chave de desempenho de forma visual e dinâmica, favorecendo a identificação de padrões e tendências (FEW, 2013).

No que se refere ao acompanhamento de alunos, a legislação brasileira — por meio do Sistema Nacional de Avaliação da Educação Superior (SINAES) (BRASIL, 2004) — exige que

as instituições de ensino superior realizem o acompanhamento de seus alunos como parte do processo de avaliação institucional. Na prática, essa exigência é atendida majoritariamente por meio de questionários e pesquisas declarativas enviadas aos formados (CAMPBELL; DEBLOIS; OBLINGER, 2007). Embora essas iniciativas forneçam dados relevantes, elas dependem da adesão voluntária dos alunos e raramente integram dados administrativos oficiais, como os provenientes da RAIS. Essa limitação reduz a abrangência e a confiabilidade das análises realizadas (CAMPBELL; DEBLOIS; OBLINGER, 2007).

A partir da análise dos trabalhos relacionados, observa-se que, apesar dos avanços na utilização de *Data Warehouse*, *Business Intelligence* e visualização de dados em diferentes domínios, ainda são escassas as soluções que integrem, de forma sistemática, dados acadêmicos e bases governamentais para o acompanhamento de alunos.

Do ponto de vista do contexto nacional, estudos do CGEE (CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS (CGEE), 2019) e dados do INEP (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP), 2022) evidenciam a relevância do tema: o Brasil forma anualmente centenas de milhares de graduados e pós-graduados, mas os mecanismos de acompanhamento de sua inserção profissional ainda são fragmentados e dependentes de iniciativas declarativas. O Censo da Educação Superior (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP), 2022) fornece estatísticas agregadas sobre matrículas, concluintes e características dos cursos, mas não permite rastrear a trajetória individual dos alunos no mercado de trabalho. O relatório do CGEE (CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS (CGEE), 2019), por sua vez, demonstra que a inserção profissional de mestres e doutores pode ser sistematicamente analisada por meio de bases governamentais — reforçando a viabilidade da abordagem adotada neste trabalho.

No que tange às análises de empregabilidade de engenheiros especificamente, Maciente e Casero (2019) realizaram, no âmbito do IPEA, um estudo sobre a inserção de diplomados em cursos de engenharia no mercado de trabalho formal utilizando dados da RAIS em âmbito nacional. Esse estudo demonstra que é possível extrair indicadores relevantes — como taxas de emprego, remuneração e aderência à área de formação — a partir do cruzamento entre dados educacionais e a RAIS. A diferença fundamental em relação ao presente trabalho é a escala de análise: enquanto Maciente e Casero (2019) analisam o conjunto nacional de engenheiros formados, este trabalho foca especificamente nos alunos do CEFET-MG, permitindo indicadores institucionais que subsidiem a gestão interna de cursos.

Do ponto de vista teórico, a análise salarial presente no painel é fundamentada na teoria do capital humano, sistematizada por Mincer (1974), segundo a qual o nível de escolaridade e a experiência acumulada são os principais determinantes da remuneração individual. Esse arcabouço justifica a inclusão de indicadores como salário médio por curso, evolução salarial ao longo do tempo e diferenças de remuneração entre níveis de ensino, tornando as análises do painel teoricamente fundamentadas.

Nesse contexto, o presente trabalho se diferencia ao propor a integração dos dados do CEFET-MG com o extrato sigiloso da RAIS — obtido via convênio governamental —, utilizando processos de *ETL* e um *Data Warehouse* institucional, culminando no desenvolvimento de um painel interativo que viabiliza análises de empregabilidade com granularidade individual e cobertura superior às abordagens declarativas atualmente praticadas.

4 METODOLOGIA

Este capítulo apresenta os procedimentos metodológicos adotados para o desenvolvimento do painel interativo de acompanhamento de alunos. A metodologia caracteriza-se como aplicada, de natureza quantitativa, com abordagem exploratória e descritiva, uma vez que busca integrar, tratar e analisar dados provenientes de diferentes fontes, transformando-os em indicadores que apoiem a tomada de decisão institucional.

4.1 Aquisição dos dados

A etapa de aquisição dos dados envolveu a coleta de informações provenientes de múltiplas fontes, internas e externas à instituição. Os dados acadêmicos dos alunos foram obtidos a partir do sistema acadêmico do CEFET-MG (SIGAA), contemplando informações como curso de formação, ano de conclusão, modalidade, campus e identificação dos alunos, respeitando critérios de confidencialidade e anonimização.

Os dados da RAIS utilizados neste trabalho foram obtidos por meio de convênio firmado entre o CEFET-MG e o Ministério do Trabalho e Emprego. Diferentemente dos microdados públicos da RAIS — disponíveis para download sem identificação nominal —, o extrato conveniado contém o Cadastro de Pessoa Física (CPF) de cada trabalhador, o que possibilita o cruzamento nominal com os registros acadêmicos do SIGAA. O uso desse extrato é regido pela Lei Geral de Proteção de Dados Pessoais (LGPD) e pelas cláusulas do convênio, que impõem restrições de finalidade, obrigações de sigilo e exigência de descarte seguro após o uso. A base fornece informações detalhadas sobre vínculos empregatícios formais, remuneração, ocupação e setor econômico, sendo fundamental para a análise da inserção profissional dos alunos.

As tabelas de referência da Classificação Brasileira de Ocupações (CBO) (MINISTÉRIO DO TRABALHO E EMPREGO, 2024a) e da Classificação Nacional de Atividades Econômicas (CNAE) (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 2024a) foram obtidas a partir dos recursos públicos disponibilizados pelo Ministério do Trabalho e Emprego, com o objetivo de padronizar e categorizar as informações de ocupações e setores econômicos. A tabela de municípios com unidade federativa e região geográfica, disponibilizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE) (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 2024b), foi incorporada para viabilizar análises de localização geográfica dos empregos.

Todos os dados foram fornecidos como arquivos CSV. A correspondência entre os registros acadêmicos do CEFET-MG e os vínculos da RAIS foi realizada por meio do CPF, permitindo identificar quais alunos possuem registro no mercado formal de trabalho. Após a carga e o cruzamento no ambiente de *staging*, o CPF não é armazenado no *Data Warehouse*, preservando a anonimização dos dados na camada analítica.

4.2 Processo de ETL (Extract, Transform, Load)

Após a aquisição, os dados passaram por um processo de *ETL* implementado inteiramente em PL/pgSQL por meio de *stored procedures* executadas diretamente no PostgreSQL, eliminando a dependência de ferramentas externas e concentrando todo o processamento no mesmo ambiente do *Data Warehouse*. O processo foi estruturado em *procedures* modulares — uma por dimensão e uma por fato — facilitando a manutenção e a reexecução parcial em caso de atualização de bases específicas.

Na fase de extração, os dados foram carregados em formato CSV para um *schema* de *staging* (*stg*), sem validações ou restrições de integridade referencial, preservando os dados brutos originais. As colunas da *staging* recebem tipos de dados compatíveis com o conteúdo semântico de cada campo: *VARCHAR* para códigos, descrições e identificadores textuais; *INTEGER* para campos numéricos inteiros como ano de referência e códigos de classificação; e *NUMERIC* para valores monetários e de jornada, que exigem precisão decimal. As *stored procedures* de ETL foram executadas manualmente pelo autor após o carregamento de todas as fontes no *schema* de *staging*.

O volume de dados processado é expressivo. A base RAIS compreende, para o período de 2020 a 2024, um total de 385.642.987 registros de vínculos empregatícios em todo o Brasil. A base de discentes extraída do SIGAA contém 55.112 registros, correspondendo a 21.755 alunos com CPF identificado. Após o cruzamento por CPF, foram retidos 32.382 vínculos associados a discentes do CEFET-MG — aproximadamente 0,008% do volume total da RAIS. O Quadro 1 detalha o volume por ano.

Tabela 1 – Volume de registros RAIS carregados e retidos após cruzamento com discentes do CEFET-MG, por ano.

Ano	RAIS — total Brasil	Discentes CEFET-MG	%
2020	65.921.194	2.522	0,004%
2021	70.521.981	3.843	0,005%
2022	78.488.470	5.216	0,007%
2023	82.964.122	9.506	0,011%
2024	87.747.220	11.295	0,013%

Fonte: elaborado pelo autor com base nos dados da RAIS (MTE) e do SIGAA (CEFET-MG).
Nota: a coluna % expressa, em cada linha, a proporção de discentes do CEFET-MG em relação ao total de vínculos da RAIS naquele mesmo ano.

Do total de vínculos retidos, 91,8% referem-se a empregos nas regiões de Minas Gerais, Rio de Janeiro e Espírito Santo, refletindo a concentração geográfica dos campi e do público formado pelo CEFET-MG. O crescimento no número de vínculos identificados ao longo do período — de 2.522 em 2020 para 11.295 em 2024 — reflete tanto o acúmulo de alunos ingressando no mercado formal a cada ano quanto o aumento gradual da própria base RAIS.

Cada registro do extrato RAIS é composto por 77 colunas, que abrangem quatro grandes categorias: (i) **identificação e vínculo** — CPF do trabalhador, CNPJ do estabelecimento, tipo de vínculo empregatício, causa de desligamento, indicador de vínculo ativo em 31/12 e meses trabalhados no ano; (ii) **remuneração e jornada** — remuneração média nominal anual, remuneração de dezembro, salário contratual e horas contratuais semanais; (iii) **classificações** — código CBO da ocupação e código CNAE da atividade econômica do estabelecimento, ambos com hierarquias próprias de até cinco níveis; e (iv) **atributos sociodemográficos** — sexo, raça/cor, faixa etária, grau de instrução. O restante das colunas compreende campos administrativos e operacionais da RAIS — como índices de atualização monetária, subtipos de admissão e controles de competência mensal — que não agregam valor analítico para os objetivos deste trabalho.

A seleção das colunas a serem incorporadas ao *Data Warehouse* seguiu dois critérios principais: *relevância analítica* para as dimensões de empregabilidade definidas no projeto (remuneração, ocupação, setor econômico, localização e perfil do trabalhador) e *disponibilidade de dimensão de suporte* no modelo estrela. Colunas redundantes — por exemplo, múltiplas representações do mesmo código de ocupação em diferentes níveis de detalhe já contemplados na hierarquia da *dm_cbo* — foram descartadas. Atributos sociodemográficos como sexo, raça/cor, faixa etária e tipo de deficiência, embora presentes na RAIS, foram mantidos como dimensões compartilhadas na fato, pois variam por vínculo e não pertencem de forma exclusiva nem ao aluno nem ao estabelecimento. O CPF, utilizado apenas como chave de cruzamento entre a RAIS e o SIGAA, não é armazenado no *Data Warehouse*, garantindo a anonimização dos dados na camada analítica.

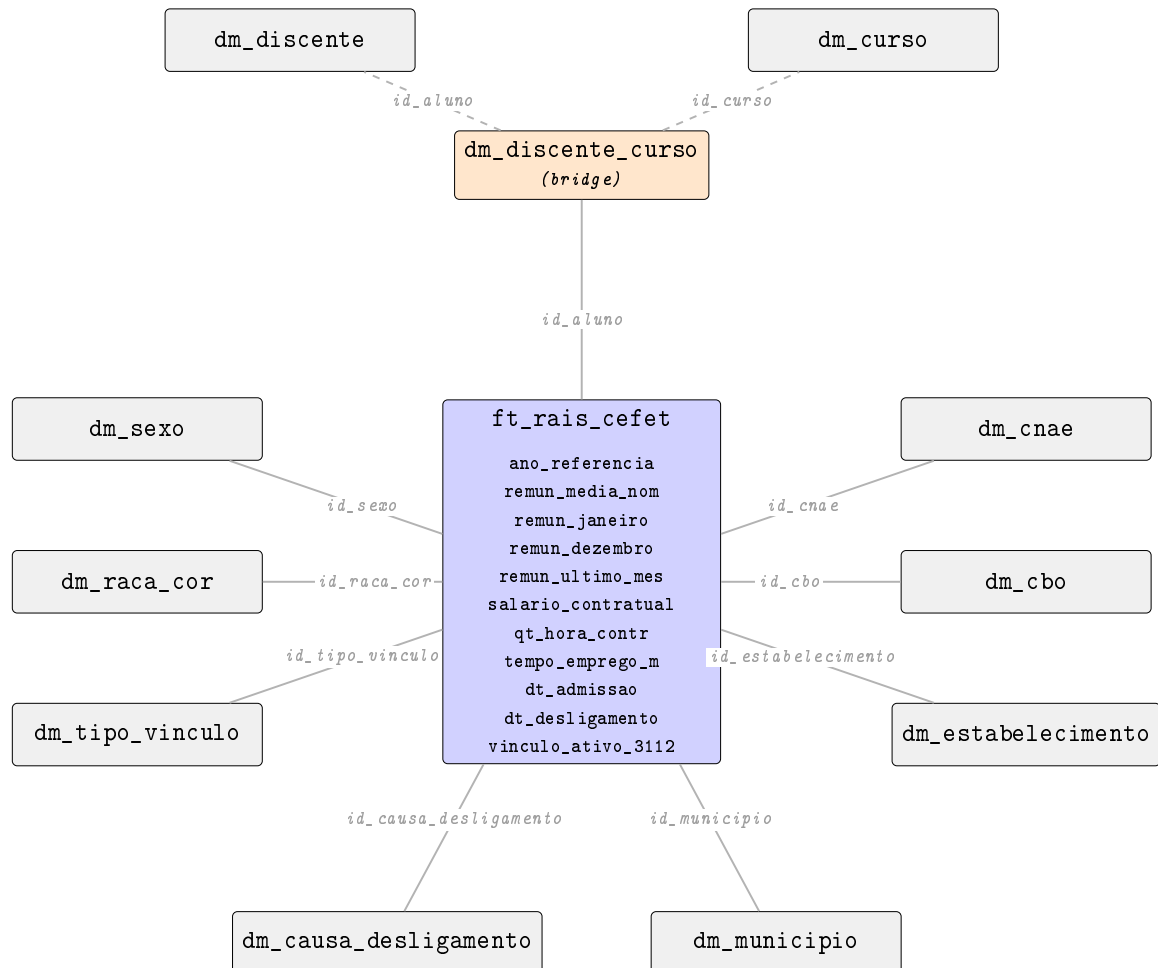
A fase de transformação consistiu na aplicação de procedimentos de limpeza, padronização e enriquecimento dos dados. Entre as atividades realizadas destacam-se:

- Remoção de registros duplicados e tratamento de valores ausentes;
- Validação e normalização dos códigos CBO e CNAE;
- Padronização de campos de remuneração, com exclusão de valores nulos ou negativos;
- Aplicação de regras de negócio para identificar vínculos empregatícios associados aos alunos por meio do CPF;

Na fase de carga, os dados transformados foram inseridos no *Data Warehouse*, respeitando a ordem obrigatória do modelo dimensional: primeiro as dimensões de domínio (sexo, raça/cor, município, CNAE, CBO), depois as dimensões principais (estabelecimento, aluno, curso) e, por último, a tabela fato. O processo foi concebido para permitir atualizações incrementais, viabilizando a incorporação de novos extratos anuais fornecidos via convênio sem necessidade de recarga completa do *Data Warehouse*. Toda dimensão de domínio possui um registro padrão

para valores não encontrados (*membro desconhecido*), garantindo que nenhuma chave estrangeira na tabela fato fique nula. A Figura 1 apresenta o modelo estrela resultante, com a tabela fato ao centro e as onze dimensões ao redor.

Figura 1 – Modelo estrela do *Data Warehouse* de alunos do CEFET-MG.



Fonte: elaborado pelo autor. Linhas tracejadas: relações via tabela *bridge* ($dm_discente_curso \rightarrow dm_curso$ e $dm_discente_curso \rightarrow dm_discente$).

4.3 Modelagem e implementação do Data Warehouse

O *Data Warehouse* foi implementado em PostgreSQL, adotando o esquema estrela como modelo dimensional. A escolha por esse esquema se justifica pela simplicidade das consultas geradas, pelo desempenho em operações de agregação e pela facilidade de integração com ferramentas de BI como o Metabase.

A estrutura central é composta por uma tabela fato e onze tabelas dimensão, conforme descrito a seguir.

4.3.1 Tabela Fato

A tabela fato principal, denominada *ft_rais_cefet*, armazena os vínculos empregatícios identificados para os alunos do CEFET-MG na RAIS. Cada registro representa um vínculo de um aluno em um determinado ano de referência, com as seguintes medidas principais:

- *remun_media_nom*: remuneração média nominal declarada pelo empregador;
- *salario_contratual*: salário contratual do vínculo;
- *qt_hora_contr*: carga horária contratual semanal;
- *tempo_emprego_m*: tempo de emprego em meses;
- *vinculo_ativo_3112*: indicador booleano de vínculo ativo em 31 de dezembro;
- *dt_admissao* e *dt_desligamento*: datas de início e encerramento do vínculo.

4.3.2 Tabelas Dimensão

As dez tabelas dimensão, acompanhadas de uma tabela *bridge*, fornecem o contexto descritivo para análise, conforme o Quadro 2.

Tabela 2 – Tabelas dimensão e tabela *bridge* do Data Warehouse de alunos.

Tabela	Descrição
dm_discente_curso	Tabela <i>bridge</i> : associa aluno a curso(s). Contém campus e período de conclusão. Aponta para dm_discente e dm_curso.
dm_discente	Dados do aluno: indicadores booleanos de nível de ensino concluído (tem_tecnico, tem_graduacao, tem_pos).
dm_curso	Informações do curso: nome, área, modalidade e nível de ensino.
dm_municipio	Município de trabalho com UF, nome da UF, região e sigla.
dm_cnae	Classificação Nacional de Atividades Econômicas: subclasse, divisão e seção.
dm_cbo	Classificação Brasileira de Ocupações: código, descrição da ocupação e grande grupo.
dm_sexo	Sexo do aluno com código e descrição.
dm_raca_cor	Raça/cor autodeclarada com código e descrição.
dm_tipo_vinculo	Tipo de vínculo empregatício (CLT, servidor público, temporário, etc.).
dm_causa_desligamento	Causa do desligamento do vínculo, quando aplicável.
dm_estabelecimento	Porte do estabelecimento empregador.

Fonte: elaborado pelo autor. Nota: dm_discente_curso é estruturalmente uma tabela *bridge*, e não uma dimensão — é incluída no quadro por resolver a relação muitos-para-muitos entre alunos e cursos, central para os filtros e agrupamentos do painel.

4.4 Construção do painel interativo no Metabase

A construção do painel interativo foi realizada no Metabase, ferramenta conectada diretamente ao *Data Warehouse* em PostgreSQL por meio de conexão JDBC. O painel foi organizado em seis seções temáticas, cada uma composta por *cards* de texto explicativo, indicadores numéricos (KPIs) e gráficos analíticos.

4.4.1 Definição dos Indicadores

Os indicadores foram definidos com base nos objetivos institucionais do CEFET-MG para o acompanhamento de alunos e nas dimensões disponíveis nas bases de dados. O Quadro 3 apresenta os principais indicadores organizados por seção temática do painel.

Tabela 3 – Principais indicadores do painel por seção temática.

Seção	Indicadores
Visão Geral	Alunos empregados (total); salário médio; mediana salarial.
Inserção no mercado	Empregabilidade por campus; top 10 cursos que mais empregam.
Remuneração	Salário médio por curso; distribuição por faixa salarial.
Equidade	Gap salarial por sexo ao longo dos anos; remuneração por raça/cor; Comparação salarial por sexo.
Localização geográfica	Distribuição de alunos por UF; top 20 municípios de emprego.
Ocupação (CBO e CNAE)	Top ocupações por remuneração; top setores CNAE.
Vínculos	Tipos de vínculo predominantes; tipos de desligamento; horas contratuais médias por tipo de vínculo.

Fonte: elaborado pelo autor.

4.4.2 Consultas SQL e Filtros Dinâmicos

Todas as visualizações do painel são alimentadas por consultas SQL nativas escritas diretamente no Metabase. Para viabilizar os filtros globais do *dashboard*, as consultas foram parametrizadas com a sintaxe de variáveis do Metabase: todos os parâmetros, incluindo `{{ano_inicio}}` e `{{ano_fim}}`, são envolvidos por colchetes duplos `[[...]]` para tornar cada cláusula `WHERE` opcional. Os três filtros booleanos de nível de ensino utilizam o tipo *Boolean* do Metabase, gerando condições do tipo `dm_discente.tem_tecnico = true` quando ativados. O Código 4.1 apresenta a consulta de empregabilidade por campus, que produz o gráfico de barras da seção de Inserção no Mercado:

```

1 SELECT
2     dc.campus ,
3     COUNT(DISTINCT f.id_aluno) AS qtd_alunos
4 FROM ft_rais_cefet f
5 JOIN dm_discente_curso dc ON dc.id_aluno = f.id_aluno
6 JOIN dm_curso c ON c.id_curso = dc.id_curso
7 JOIN dm_municipio m ON m.id_municipio = f.id_municipio
8 JOIN dm_discente ON dm_discente.id_aluno = f.id_aluno
9 WHERE dc.campus <> ''
10 [[AND f.ano_referencia >= {{ano_inicio}}]]
11 [[AND f.ano_referencia <= {{ano_fim}}]]
12 [[AND dc.campus = {{campus}}]]
13 [[AND m.uf = {{uf}}]]
14 [[AND c.nome_curso = {{curso}}]]
15 [[AND dm_discente.tem_tecnico = {{possui_tecnico}}]]
16 [[AND dm_discente.tem_graduacao = {{possui_graduacao}}]]
17 [[AND dm_discente.tem_pos = {{possui_pos}}]]

```

```

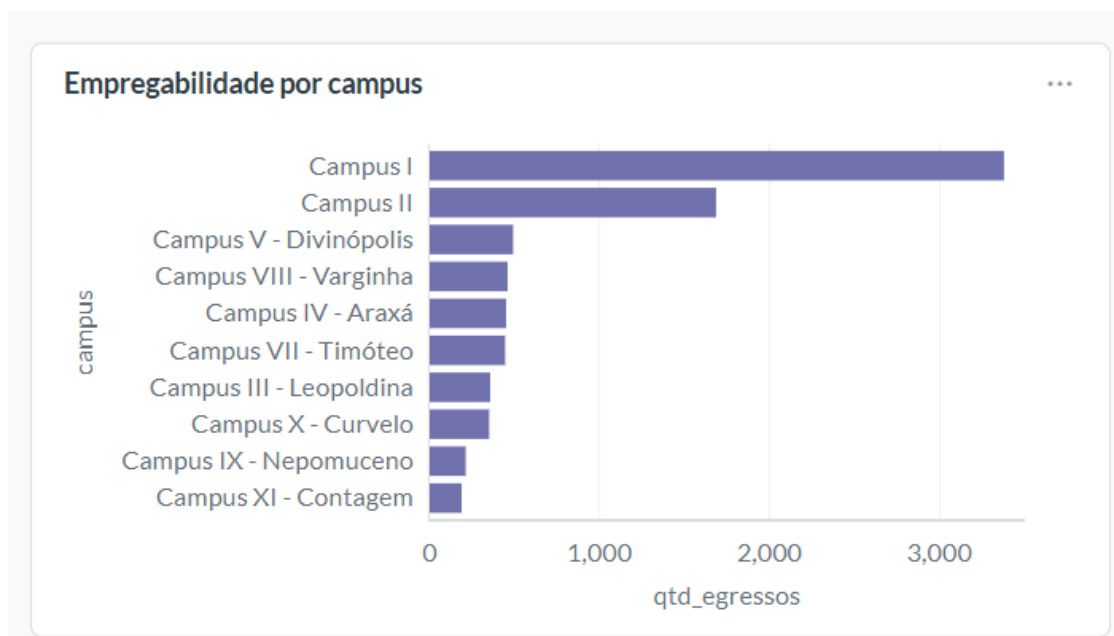
18 GROUP BY dc.campus
19 ORDER BY qtd_alunos DESC;

```

Listing 4.1 – Consulta de empregabilidade por campus com filtros dinâmicos (*field filters*).

Quando o usuário não seleciona valor para um filtro, a cláusula correspondente é automaticamente omitida pelo Metabase, e o *card* apresenta os dados agregados para todos os valores daquela dimensão. Para os filtros booleanos, a omissão equivale a não restringir por nível de ensino. Esse comportamento garante que o painel funcione tanto com filtros aplicados quanto sem nenhuma seleção. A Figura 2 exibe a interface do Metabase com essa consulta configurada e o resultado gerado.

Figura 2 – Consulta de empregabilidade por campus configurada no Metabase com *field filters*.



Fonte: elaborado pelo autor.

4.4.3 Filtros Globais do Dashboard

Foram definidos nove filtros globais no painel, configurados como *dashboard filters* no Metabase e mapeados para as colunas correspondentes em cada questão analítica:

- **Ano Início:** tipo *Number*, mapeado para `f.ano_referencia` com operador "maior ou igual"; define o limite inferior do período analisado;
- **Ano Fim:** tipo *Number*, mapeado para `f.ano_referencia` com operador "menor ou igual"; define o limite superior do período analisado;
- **Campus:** tipo *Category*, mapeado para `dc.campus`; afeta *cards* com *JOIN* em `dm_discente_curso`;
- **Curso:** tipo *Category*, mapeado para `c.nome_curso`; afeta *cards* com *JOIN* em `dm_curso`;

- **Possui Técnico:** tipo *Boolean*, mapeado para `dm_discente.tem_tecnico`; filtra alunos que possuem ao menos um curso técnico;
- **Possui Graduação:** tipo *Boolean*, mapeado para `dm_discente.tem_graduacao`; filtra alunos que possuem ao menos um curso de graduação;
- **Possui Pós:** tipo *Boolean*, mapeado para `dm_discente.tem_pos`; filtra alunos que possuem ao menos um curso de pós-graduação;
- **UF:** tipo *Category*, mapeado para `m.uf`; afeta *cards* com *JOIN* em `dm_municipio`; indisponível para alunos sem vínculo na RAIS;
- **Status do Curso:** tipo *Category*, mapeado para `dc.status`; com valor padrão *Concluído*, distingue alunos egressos de alunos ainda ativos nos cursos do CEFET-MG.

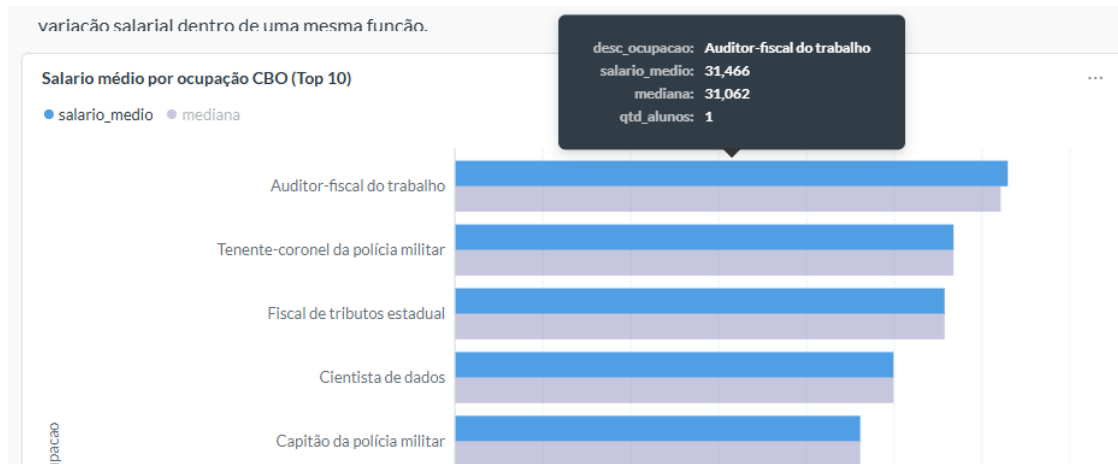
Os três filtros booleanos de nível de ensino operam de forma independente e combinável: selecionar apenas *Possui Graduação = Verdadeiro*, por exemplo, restringe os *cards* a alunos que possuem graduação, independentemente de também terem concluído cursos técnicos ou de pós-graduação. Os indicadores booleanos `tem_tecnico`, `tem_graduacao` e `tem_pos` são atributos da tabela `dm_discente`, calculados durante o processo de ETL com base no nível de ensino de cada curso registrado em `dm_discente_curso`.

4.5 Validação dos dados e dos indicadores

A validação dos dados e dos indicadores gerados constituiu etapa fundamental da metodologia. Foram realizadas verificações de consistência entre os dados do *Data Warehouse* e as bases originais, com o objetivo de assegurar a integridade das informações após o processo de *ETL*. Em especial, foram confrontados os totais de vínculos por ano e os valores médios de remuneração com estatísticas agregadas disponíveis nos relatórios públicos da RAIS, verificando-se aderência satisfatória entre os resultados do DW e os publicados pelo Ministério do Trabalho e Emprego para o subconjunto de alunos do CEFET-MG.

A validação funcional do painel foi conduzida pelo próprio autor com o auxílio do coorientador, avaliando a clareza das visualizações, a consistência dos filtros e o desempenho das consultas analíticas. Esse processo permitiu identificar ajustes necessários, como a adição de textos explicativos em cada seção e a inclusão da mediana em algumas seções como indicador complementar à média salarial, reduzindo o risco de interpretações distorcidas por valores extremos. A Figura 3 exemplifica o uso da mediana como indicador complementar à média em um dos gráficos do painel.

Figura 3 – Exemplo de gráfico com mediana como indicador complementar à média salarial.



Fonte: elaborado pelo autor.

4.6 Atualização e manutenção do ambiente analítico

Considerando a natureza dinâmica da base RAIS, o ambiente de BI foi projetado para permitir atualizações periódicas. O processo de *ETL* foi estruturado para possibilitar a incorporação de novos arquivos anuais da RAIS com execução incremental, sem necessidade de recarga completa do *Data Warehouse*.

Foram também definidos procedimentos para manutenção do ambiente analítico, incluindo monitoramento da qualidade dos dados após cada carga, revisão periódica dos indicadores e documentação das consultas SQL no próprio Metabase por meio de *cards* de texto explicativo. Essa abordagem assegura a sustentabilidade da solução e sua utilização a longo prazo por equipes sem necessidade de acesso direto ao banco de dados.

Com o objetivo de favorecer a reprodutibilidade e a transparência metodológica deste trabalho, os scripts de modelagem do *Data Warehouse*, as *stored procedures* de *ETL* e as consultas SQL do painel serão disponibilizados em um repositório no GitHub após a defesa.

5 RESULTADOS

Este capítulo apresenta os resultados obtidos no desenvolvimento e na validação do painel interativo de acompanhamento de alunos do CEFET-MG. Os resultados são organizados conforme as etapas da metodologia: o *Data Warehouse* construído, as consultas analíticas desenvolvidas, o painel implementado no Metabase e os principais achados analíticos identificados.

5.1 Data Warehouse implementado

O *Data Warehouse* foi implementado com sucesso em PostgreSQL, seguindo o esquema estrela descrito na metodologia. A tabela fato `ft_rais_cefet` contém os vínculos empregatícios cruzados entre os alunos do CEFET-MG e a RAIS, com granularidade de um registro por aluno por vínculo por ano de referência.

A carga inicial do DW contemplou dados da RAIS para o período de 2020 a 2024, resultando em um volume expressivo de registros após o processo de cruzamento entre as bases acadêmicas e a base governamental. As onze tabelas dimensão foram populadas com os dados de referência correspondentes, incluindo as classificações CBO e CNAE em seus respectivos níveis hierárquicos.

O esquema dimensional adotado possibilita consultas analíticas eficientes, com tempos de resposta satisfatórios para uso interativo no Metabase mesmo em consultas com múltiplos *JO-INS* e funções de agregação como `PERCENTILE_CONT` para cálculo de medianas.

5.2 Consultas analíticas desenvolvidas

Foram desenvolvidas e documentadas vinte e uma consultas SQL analíticas, organizadas em sete grupos temáticos conforme o Quadro 4. Todas as consultas foram validadas diretamente no Metabase e parametrizadas com *field filters* para suporte aos filtros globais do *dashboard*.

Tabela 4 – Grupos de consultas analíticas desenvolvidas para o painel.

Grupo	Qtd.	Principais consultas
Visão Geral	4	KPI alunos empregados; KPI salário médio; KPI mediana salarial; distribuição de alunos por UF.
Inserção no mercado	2	Alunos empregados por campus; top 10 cursos que mais empregam.
Remuneração	3	Salário médio por curso (top 10); distribuição por faixa salarial; evolução salarial temporal dos top 10 cursos (com mediana).
Equidade	4	Gap salarial por sexo ao longo dos anos (com mediana); remuneração por raça/cor; distribuição de gênero por curso; salário médio por sexo.
Localização geográfica	2	Top 20 municípios de trabalho; alunos por UF.
Ocupação (CBO e CNAE)	3	Salário médio e mediana por ocupação (CBO); top setores CNAE por volume de alunos; aderência curso–ocupação (CBO).
Vínculos	3	Tipos de vínculo empregatício; causas de desligamento; horas contratuais médias por tipo de vínculo.

Fonte: elaborado pelo autor.

A seguir apresentam-se exemplos representativos de consultas implementadas, ilustrando as principais abordagens analíticas adotadas.

5.2.1 Consulta de Salário Médio por Curso

A consulta apresentada na listagem 5.1 produz o ranking dos dez cursos com maior remuneração média entre os alunos empregados em um dado ano. A tabela fato é associada à tabela *bridge dm_discente_curso* — que resolve a relação muitos-para-muitos entre alunos e cursos — e em seguida à *dm_curso*, de onde é extraído o nome do curso. Os *joins* adicionais com *dm_municipio* e *dm_discente* habilitam os filtros opcionais por UF e por nível de ensino, sem afetar o agrupamento principal. O filtro *f.remun_media_nom > 0* exclui vínculos sem remuneração registrada, e *c.nome_curso <> ''* remove registros com classificação de curso ausente. O resultado, ordenado de forma decrescente e limitado aos dez primeiros, pode ser observado na Figura 4.

```

1 SELECT
2     c.nome_curso ,
3     ROUND(AVG(f.remun_media_nom), 0) AS salario_medio

```

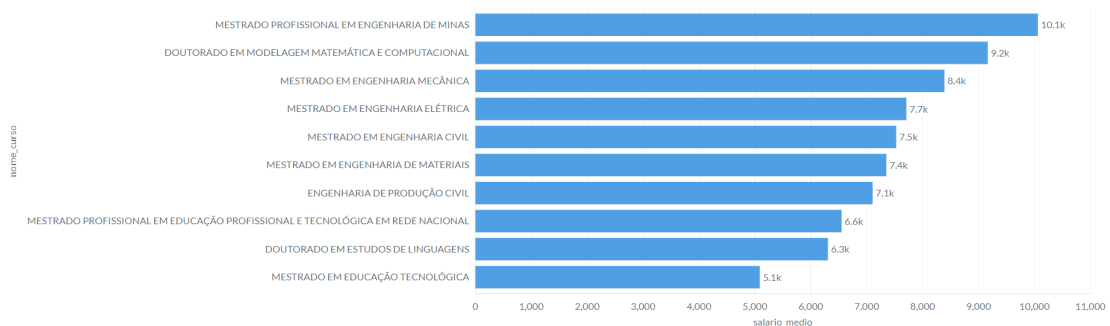
```

4 FROM ft_rais_cefet f
5 JOIN dm_discente_curso dc ON dc.id_aluno = f.id_aluno
6 JOIN dm_curso c ON c.id_curso = dc.id_curso
7 JOIN dm_municipio m ON m.id_municipio = f.id_municipio
8 JOIN dm_discente ON dm_discente.id_aluno = f.id_aluno
9 WHERE f.remun_media_nom > 0
10 AND c.nome_curso <> ''
11 [[AND f.ano_referencia >= {{ano_inicio}}]]
12 [[AND f.ano_referencia <= {{ano_fim}}]]
13 [[AND dc.campus = {{campus}}]]
14 [[AND m.uf = {{uf}}]]
15 [[AND c.nome_curso = {{curso}}]]
16 [[AND dm_discente.tem_tecnico = {{possui_tecnico}}]]
17 [[AND dm_discente.tem_graduacao = {{possui_graduacao}}]]
18 [[AND dm_discente.tem_pos = {{possui_pos}}]]
19 GROUP BY c.nome_curso
20 ORDER BY salario_medio DESC
21 LIMIT 10;

```

Listing 5.1 – Ranking de salário médio por curso com filtros dinâmicos (top 10).

Figura 4 – Gráfico de salário médio por curso no painel interativo.



Fonte: elaborado pelo autor.

5.2.2 Consulta de Salário Médio e Mediana por Ocupação

A consulta apresentada na listagem 5.2 retorna as dez ocupações com maior remuneração média entre os alunos empregados, exibindo simultaneamente a média e a mediana salarial. A mediana é calculada via `PERCENTILE_CONT(0.5)` e convertida explicitamente para `NUMERIC` antes do arredondamento, contornando a restrição do PostgreSQL que não aceita `ROUND` diretamente sobre `double precision`. A comparação entre os dois indicadores permite identificar ocupações com distribuição salarial assimétrica — casos em que a média é inflada por valores extremos. A consulta percorre cinco tabelas (`ft_rais_cefet`, `dm_cbo`, `dm_discente_curso`, `dm_curso`, `dm_municipio` e `dm_discente`), suportando todos os filtros globais do painel por meio de *field filters* opcionais. O resultado pode ser observado na Figura 5.

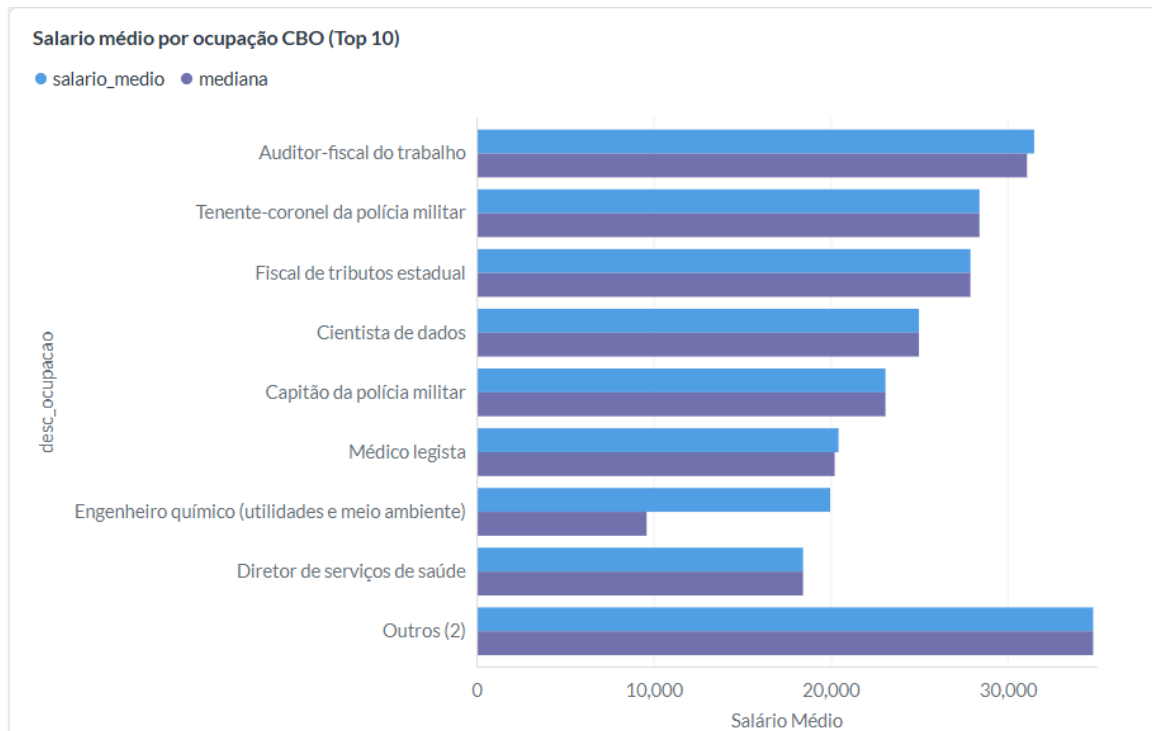
```

1 SELECT
2     cb.desc_ocupacao ,
3     ROUND(AVG(f.remun_media_nom), 0)
4                                     AS salario_medio
5     ,
6     ROUND(PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY f.remun_media_nom)
7         ::NUMERIC, 0) AS mediana,
8     COUNT(DISTINCT f.id_aluno)
9                                     AS
10        qtd_alunos
11 FROM ft_rais_cefet f
12 JOIN dm_cbo          cb ON cb.id_cbo      = f.id_cbo
13 JOIN dm_discente_curso dc ON dc.id_aluno   = f.id_aluno
14 JOIN dm_curso         c  ON c.id_curso     = dc.id_curso
15 JOIN dm_municipio    m  ON m.id_municipio = f.id_municipio
16 JOIN dm_discente      ON dm_discente.id_aluno = f.id_aluno
17 WHERE f.remun_media_nom > 0
18 AND cb.cbo              <> '000000'
19 [[AND f.ano_referencia >= {{ano_inicio}}]]
20 [[AND f.ano_referencia <= {{ano_fim}}]]
21 [[AND dc.campus       = {{campus}}]]
22 [[AND m.uf            = {{uf}}]]
23 [[AND c.nome_curso    = {{curso}}]]
24 [[AND dm_discente.tem_tecnico    = {{possui_tecnico}}]]
25 [[AND dm_discente.tem_graduacao = {{possui_graduacao}}]]
26 [[AND dm_discente.tem_pos       = {{possui_pos}}]]
27 GROUP BY cb.desc_ocupacao
28 ORDER BY salario_medio DESC
29 LIMIT 10;

```

Listing 5.2 – Salário médio e mediana salarial por ocupação CBO (top 10).

Figura 5 – Gráfico de salário médio e mediana salarial por ocupação (CBO) no painel interativo.



Fonte: elaborado pelo autor.

5.2.3 Consulta de Salário Médio por Sexo e Ano

A consulta apresentada na listagem 5.3 produz a série histórica de remuneração média e mediana segmentada por sexo, permitindo visualizar a evolução do gap salarial de gênero ao longo de todo o período disponível no *Data Warehouse*. Diferentemente das demais consultas do painel, esta não possui filtro fixo de *ano_referencia*: ao retornar todos os anos em uma única execução, ela é ideal para alimentar gráficos de linhas com duas séries sobrepostas — uma para alunos masculinos e outra para alunos femininos. O campo *s.descricao* é obtido da dimensão *dm_sexo* e o filtro *IN ('Masculino', 'Feminino')* exclui registros com sexo não informado ou não identificado. A mediana salarial é calculada em paralelo à média, permitindo comparar ambos os indicadores por gênero e por ano. O agrupamento duplo por *ano_referencia* e *descricao* gera uma linha por combinação ano–sexo, que o Metabase automaticamente interpreta como duas séries temporais distintas. O resultado pode ser visualizado na Figura 6.

```

1 SELECT
2     f.ano_referencia,
3     s.descricao,
4     ROUND(AVG(f.remun_media_nom), 0)
5     AS
6     salario_medio,
7     ROUND(PERCENTILE_CONT(0.5) WITHIN GROUP (ORDER BY f.remun_media_nom)
8     ::NUMERIC, 0) AS mediana,

```

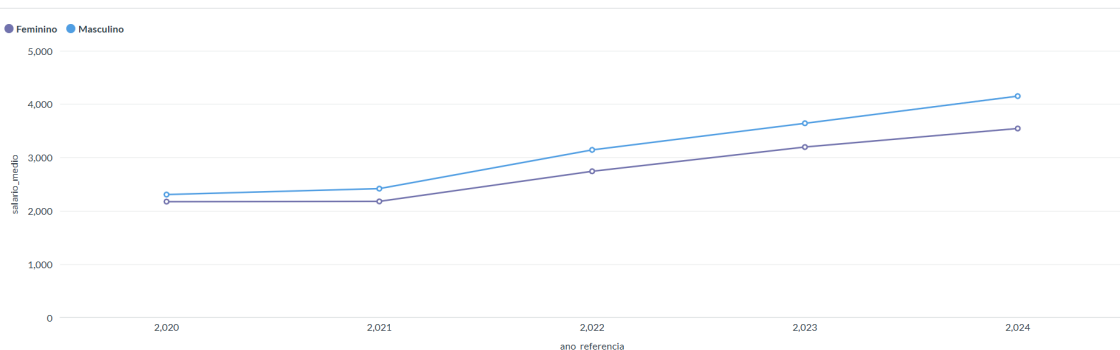
```

6      COUNT(DISTINCT f.id_aluno)
                                           AS
      qtd_alunos
7 FROM ft_rais_cefet f
8 JOIN dm_sexo          s ON s.id_sexo      = f.id_sexo
9 JOIN dm_discente_curso dc ON dc.id_aluno   = f.id_aluno
10 JOIN dm_curso         c ON c.id_curso     = dc.id_curso
11 JOIN dm_municipio    m ON m.id_municipio = f.id_municipio
12 JOIN dm_discente      ON dm_discente.id_aluno = f.id_aluno
13 WHERE f.remun_media_nom > 0
14 AND s.descricao IN ('Masculino', 'Feminino')
15 [[AND dc.campus      = {{campus}}]]
16 [[AND m.uf           = {{uf}}]]
17 [[AND c.nome_curso   = {{curso}}]]
18 [[AND dm_discente.tem_tecnico = {{possui_tecnico}}]]
19 [[AND dm_discente.tem_graduacao = {{possui_graduacao}}]]
20 [[AND dm_discente.tem_pos      = {{possui_pos}}]]
21 GROUP BY f.ano_referencia, s.descricao
22 ORDER BY f.ano_referencia, s.descricao;

```

Listing 5.3 – Evolução do salário médio e mediana por sexo ao longo dos anos.

Figura 6 – Gráfico de evolução do gap salarial por sexo ao longo dos anos no painel interativo.



Fonte: elaborado pelo autor.

5.3 Painel interativo no Metabase

O painel interativo foi implementado com sucesso no Metabase e organizado em seis seções temáticas, acessíveis a partir de um único *dashboard* com barra de filtros global. As seções foram estruturadas com *cards* de texto explicativo no início de cada bloco, seguidos pelos indicadores numéricos e pelas visualizações gráficas correspondentes.

O acesso ao painel é realizado por meio de um link público gerado pelo administrador do Metabase, sem necessidade de cadastro ou autenticação por parte do usuário final¹. Esse

¹ Link público do painel: <http://192.168.153.63:3000/public/dashboard/>

mecanismo de compartilhamento, nativo do Metabase, permite que gestores, coordenadores de curso e demais interessados consultem o painel diretamente pelo navegador, sem necessidade de conhecimento técnico ou acesso ao banco de dados.

5.3.1 Barra de Filtros

A barra de filtros, posicionada abaixo da barra de título do *dashboard*, apresenta os nove filtros globais em formato de *chips* interativos. Ao clicar em qualquer filtro, um *dropdown* é exibido com as opções disponíveis. Quando um filtro é selecionado, todos os *cards* conectados são automaticamente atualizados com os dados correspondentes à seleção, sem necessidade de intervenção manual em cada visualização.

Os filtros **Ano Início** e **Ano Fim** permitem delimitar um período de análise — por exemplo, comparar apenas os anos de 2022 a 2024 — e vêm com 2020 e 2024 como valores padrão, refletindo todo o período disponível no *Data Warehouse*. O filtro **Status do Curso** vem com o valor padrão *Concluído*, de modo que o painel apresenta, por padrão, indicadores referentes a alunos egressos; o usuário pode alterná-lo para *Ativo* para analisar o público ainda matriculado. Todos os nove filtros são opcionais e podem ser combinados livremente. A Figura 7 ilustra a barra de filtros do painel.

Figura 7 – Barra de filtros globais do painel Metabase: período (Ano Início/Fim), Campus, Curso, níveis de ensino, UF e Status do Curso.



Fonte: elaborado pelo autor.

5.3.2 Seção de Visão Geral

A seção de Visão Geral apresenta três indicadores-chave (KPIs) que resumem o estado geral da empregabilidade no período selecionado: **Alunos Empregados** (total de alunos com vínculo formal registrado na RAIS), **Salário Médio** e **Mediana Salarial**. Um quarto *card*, em formato de gráfico de rosca, apresenta a **Distribuição de Alunos/Egressos na RAIS**, cruzando a situação do aluno no CEFET-MG (ativo ou egresso) com a presença ou ausência de vínculo formal na RAIS, resultando em quatro categorias: egressos na RAIS, egressos fora da RAIS, ativos na RAIS e ativos fora da RAIS. Esses indicadores são atualizados automaticamente pelos filtros globais. A Figura 8 ilustra esses indicadores para o período de 2020 a 2024.

0ec3437f-bef5-4706-bbb4-a487aa08bef0. O acesso requer conexão à rede interna do CEFET-MG ou uso de VPN institucional.

Figura 8 – Painel Metabase — seção Visão Geral: KPIs de empregabilidade e distribuição de alunos/egressos na RAIS.

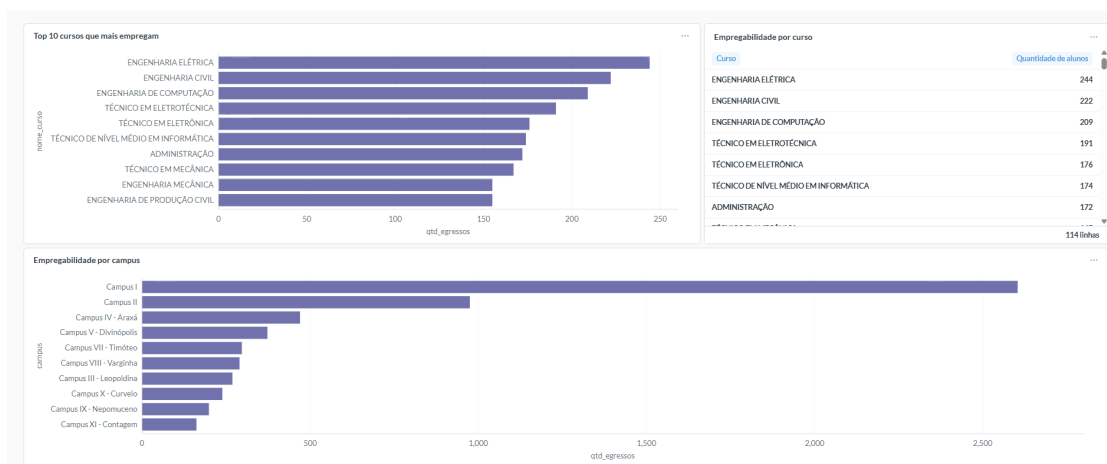


Fonte: elaborado pelo autor.

5.3.3 Seção de Inserção no Mercado

A seção de inserção no mercado apresenta o ranking dos dez cursos que mais empregam em barras horizontais, acompanhado de uma tabela completa com a empregabilidade de todos os cursos, e um gráfico de barras com a distribuição de alunos empregados por campus. O texto introdutório informa ao usuário que os dados são cruzados com a RAIS por meio do CPF — obtido via convênio governamental e não armazenado no *Data Warehouse* — e que alunos sem registro na base são tratados como não localizados, não necessariamente desempregados. Essa ressalva é fundamental para evitar interpretações equivocadas dos indicadores. A Figura 9 apresenta o ranking top 10 e a tabela completa de empregabilidade por curso, além da empregabilidade por campus.

Figura 9 – Painel Metabase — seção Inserção no Mercado: top 10 cursos, tabela completa de empregabilidade por curso e empregabilidade por campus.



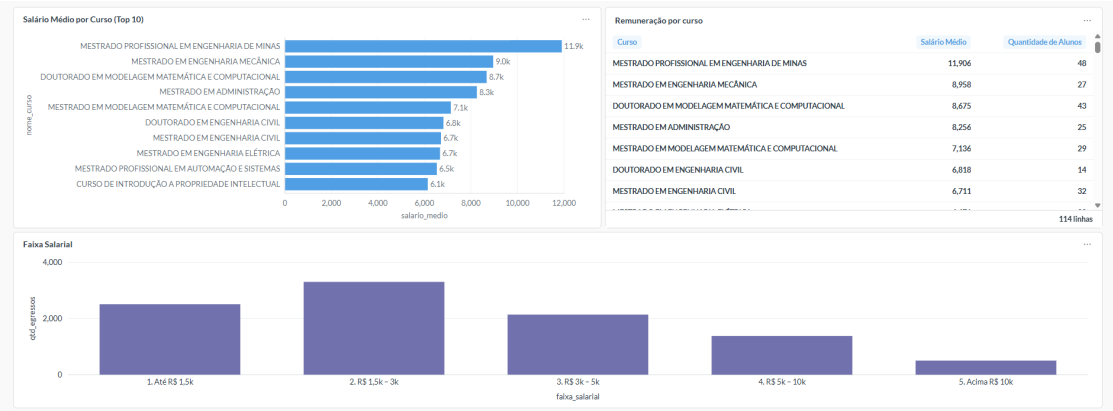
Fonte: elaborado pelo autor.

5.3.4 Seção de Remuneração

A seção de remuneração apresenta o ranking dos dez cursos com maior salário médio em barras horizontais, acompanhado de uma tabela completa com a remuneração média e a quanti-

dade de alunos de todos os cursos, além da distribuição por faixa salarial em barras verticais. A Figura 10 ilustra o ranking top 10, a tabela completa de remuneração por curso e a distribuição por faixa salarial.

Figura 10 – Painel Metabase — seção Remuneração: top 10 cursos, tabela completa de remuneração por curso e distribuição por faixa salarial.

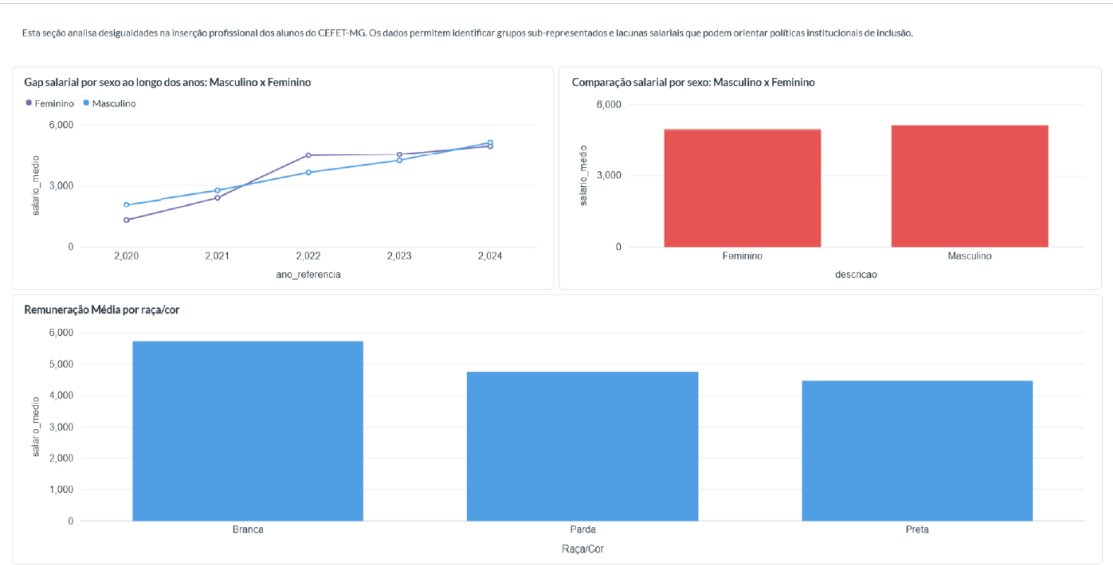


Fonte: elaborado pelo autor.

5.3.5 Seção de Equidade

A seção de equidade apresenta o gap salarial por sexo ao longo dos anos em gráfico de linhas com duas séries e a remuneração média por raça/cor em barras. Os filtros de campus, curso e nível de ensino têm efeito direto sobre todos os *cards* desta seção. A Figura 11 exibe a evolução do gap salarial por sexo no período 2020–2024 e a remuneração média por raça/cor.

Figura 11 – Painel Metabase — seção Equidade: gap salarial por sexo (2020–2024) e remuneração por raça/cor (ano de referência: 2024).

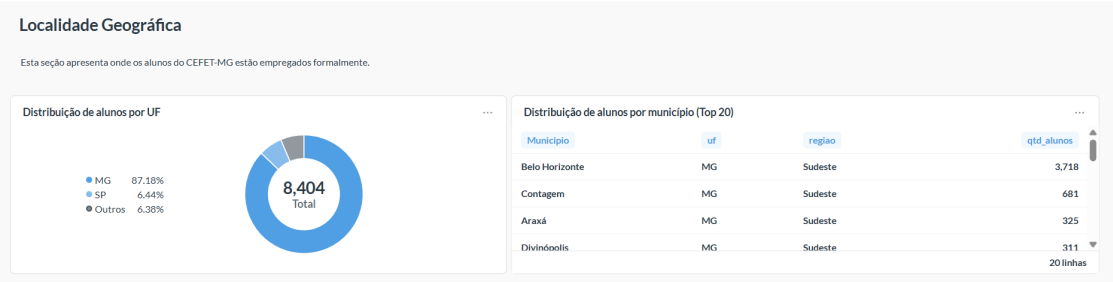


Fonte: elaborado pelo autor.

5.3.6 Seção de Localização Geográfica

A seção de localização geográfica apresenta os municípios e estados com maior concentração de alunos empregados, a distribuição por UF e o ranking dos top 20 municípios de emprego. A Figura 12 ilustra a distribuição por unidade federativa e a tabela dos principais municípios empregadores dos alunos do CEFET-MG.

Figura 12 – Painel Metabase — seção Localidade Geográfica: distribuição por UF e top 20 municípios de emprego (ano de referência: 2024).



Fonte: elaborado pelo autor.

5.3.7 Seção de Ocupação

A seção de ocupação reúne análises de CBO e CNAE em dois blocos temáticos. O ranking das dez ocupações com maior remuneração, com salário médio e mediana sobrepostos, é acompanhado de uma tabela completa com a posição, salário médio, mediana e quantidade de alunos por ocupação, permitindo comparar a distribuição salarial dentro de cada função. A Figura 13 apresenta o ranking top 10 e a tabela completa de remuneração por ocupação CBO.

Figura 13 – Painel Metabase — seção Ocupação: salário médio por ocupação CBO (Top 10) e tabela completa de remuneração por ocupação.

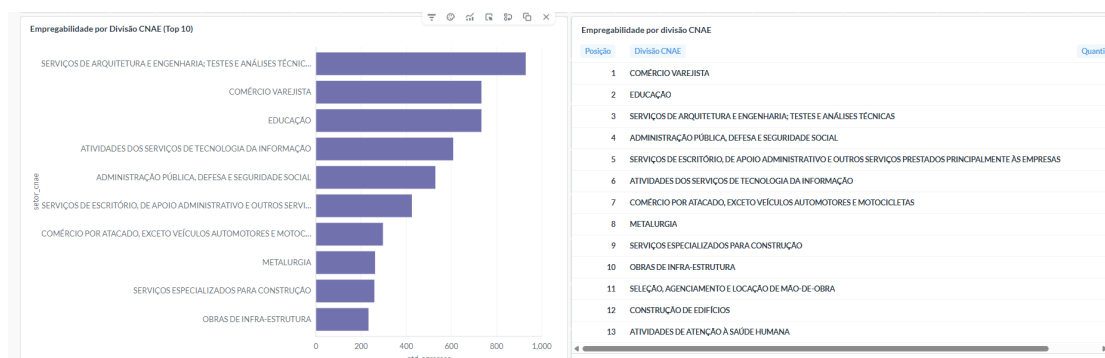


Fonte: elaborado pelo autor.

No âmbito do CNAE, são apresentados os top setores econômicos por volume de alunos empregados, também acompanhados de uma tabela completa com a posição e a divisão CNAE de

todos os setores. A Figura 14 apresenta o ranking top 10 e a tabela completa de empregabilidade por divisão CNAE.

Figura 14 – Painel Metabase — seção Ocupação: empregabilidade por divisão CNAE (Top 10) e tabela completa.

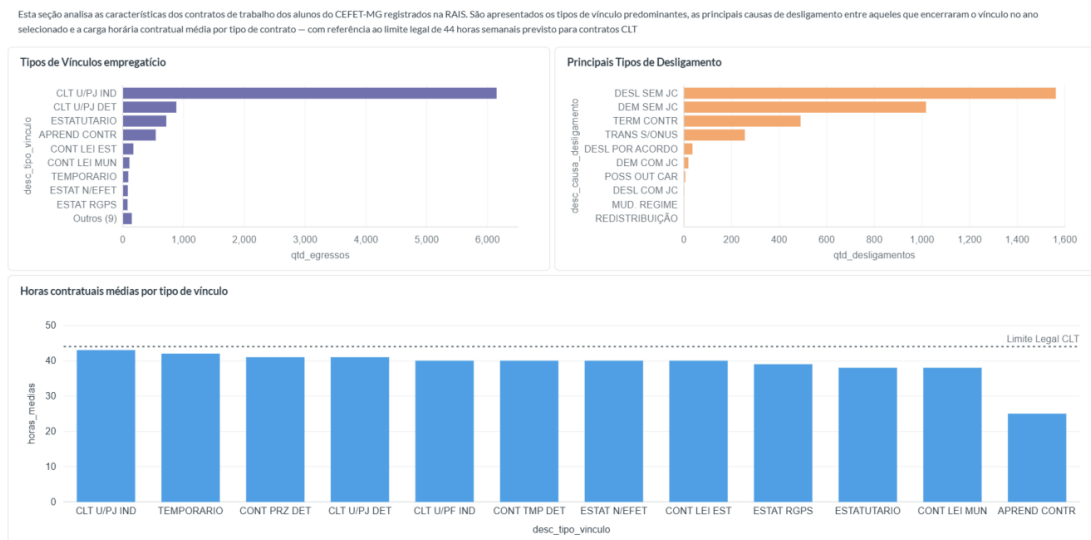


Fonte: elaborado pelo autor.

5.3.8 Seção de Vínculos

A seção de vínculos apresenta a distribuição dos tipos de contrato predominantes, as principais causas de desligamento e as horas contratuais médias por tipo de vínculo. O gráfico de horas inclui uma linha de referência no limite legal de 44 horas semanais previsto para contratos CLT, permitindo identificar modalidades com jornada reduzida. A Figura 15 ilustra os três gráficos desta seção.

Figura 15 – Painel Metabase — seção Vínculos: tipos de contrato, causas de desligamento e horas contratuais médias por tipo de vínculo (ano de referência: 2024).



Fonte: elaborado pelo autor.

5.4 Principais achados analíticos

A análise dos dados consolidados no *Data Warehouse* revelou um conjunto de achados relevantes sobre a trajetória profissional dos alunos do CEFET-MG no período de 2020 a 2024.

Em relação à remuneração, os cursos de Engenharia de Computação, Engenharia de Produção Civil e Engenharia Elétrica concentram os maiores salários médios entre os alunos, com valores significativamente acima da mediana geral. A diferença entre média e mediana salarial revelou assimetria positiva, indicando a presença de um subgrupo de alunos com remuneração substancialmente superior à maioria.

Em relação à localização geográfica, a maioria dos alunos empregados permanece em Minas Gerais, com concentração na Região Metropolitana de Belo Horizonte. Observa-se, contudo, uma parcela não desprezível de alunos empregados em São Paulo e no Rio de Janeiro, com salários médios superiores à média estadual mineira.

Em relação à ocupação, a análise de aderência entre o curso de formação e a ocupação exercida mostrou que a maioria dos alunos dos cursos de engenharia atua em funções tecnicamente relacionadas à sua área de formação. O setor de Tecnologia da Informação destaca-se como o

maior absorvedor de alunos, combinando alto volume de emprego e alta remuneração.

6 CONCLUSÃO

Este trabalho apresentou o projeto, a implementação e os resultados de uma solução de *Business Intelligence* para o acompanhamento de alunos do CEFET-MG, integrando dados do sistema acadêmico institucional (SIGAA) com o extrato sigiloso da RAIS, concedido por meio de convênio entre o CEFET-MG e o Governo Federal. A solução é composta por um *Data Warehouse* implementado em PostgreSQL com esquema estrela, um processo de *ETL* modular implementado em PL/pgSQL com *stored procedures* no PostgreSQL, e um painel interativo construído no Metabase com vinte e uma consultas analíticas organizadas em seis seções temáticas.

Os resultados obtidos demonstram que a integração entre dados acadêmicos e bases governamentais é viável e produz indicadores relevantes para a gestão institucional. O painel permite analisar, de forma interativa e com múltiplos recortes simultâneos, a empregabilidade dos alunos, sua remuneração, as ocupações exercidas, a localização geográfica dos empregos e aspectos de equidade de gênero e raça. Os filtros globais do *dashboard* permitem que gestores e coordenadores de curso obtenham respostas específicas para suas realidades sem necessidade de conhecimento técnico em SQL ou acesso direto ao banco de dados.

A principal contribuição deste trabalho reside na proposição e validação de uma arquitetura analítica capaz de centralizar dados heterogêneos em um ambiente único, estruturado e orientado à análise. A utilização do *Data Warehouse* e do Metabase permitiu superar as limitações observadas em soluções tradicionais de acompanhamento de alunos, que frequentemente dependem de pesquisas declarativas com baixa taxa de resposta e dificuldade de atualização sistemática.

Como limitações do trabalho, destaca-se que a correspondência entre alunos e vínculos da RAIS é realizada por CPF, o que implica que alunos que não autorizaram o compartilhamento de dados ou que têm inconsistências cadastrais não são localizados na base. Além disso, os dados da RAIS cobrem exclusivamente o emprego formal, não capturando trajetórias profissionais em empregos informais, trabalho autônomo ou empreendedorismo.

Como perspectivas de trabalhos futuros, sugere-se: a incorporação de dados de emprego informal por meio de pesquisas complementares; a implementação de modelos preditivos de empregabilidade com base no perfil acadêmico dos concluintes; o desenvolvimento de uma interface de autoatendimento para que os próprios alunos possam consultar seu histórico profissional; e a extensão da solução para outras unidades do CEFET-MG ainda não contempladas. Recomenda-se ainda a realização de estudos de usabilidade formais com os gestores institucionais para identificar novos indicadores de interesse e aprimorar a experiência de uso do painel.

Conclui-se que a abordagem proposta é viável, tecnicamente sólida e alinhada às necessidades institucionais para o acompanhamento de alunos, contribuindo para o fortalecimento de

uma gestão baseada em dados e para a melhoria contínua dos processos educacionais no âmbito do CEFET-MG.

REFERÊNCIAS

- BATINI, C.; CERI, S.; NAVATHE, S. *Conceptual Database Design: An Entity-Relationship Approach*. Redwood City: Benjamin-Cummings Publishing Co., 2009.
- BRASIL. *Lei nº 10.861, de 14 de abril de 2004: institui o Sistema Nacional de Avaliação da Educação Superior (SINAES)*. 2004. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/110.861.htm. Acesso em: 8 jun. 2025.
- BRASIL. *Lei nº 13.709, de 14 de agosto de 2018: Lei Geral de Proteção de Dados Pessoais (LGPD)*. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 8 jun. 2025.
- CAMPBELL, J. P.; DEBLOIS, P. B.; OBLINGER, D. G. Academic analytics: A new tool for a new era. *EDUCAUSE Review*, [S. l.], v. 42, n. 4, p. 40–57, 2007.
- CENTRO DE GESTÃO E ESTUDOS ESTRATÉGICOS (CGEE). *Mestres e Doutores: Inserção profissional e ocupação no mercado de trabalho*. Brasília, DF: CGEE, 2019.
- FERREIRA, R. G. C. *Data Warehouse na prática: fundamentos e implantação*. Dissertação (Mestrado em Ciência da Computação) — Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002.
- FEW, S. *Information Dashboard Design: Displaying Data for At-a-glance Monitoring*. 2. ed. Burlingame: Analytics Press, 2013.
- GURA, E. F.; BENCK, L. L. N. *Construção de um Data Warehouse para apoio à tomada de decisão*. Monografia (Trabalho de Conclusão de Curso (Tecnologia em Análise e Desenvolvimento de Sistemas)) — Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2011.
- INMON, W. H. *Building the Data Warehouse*. 4. ed. Hoboken: Wiley, 2005.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Classificação Nacional de Atividades Econômicas (CNAE)*. 2024. Disponível em: <https://cnae.ibge.gov.br/>. Acesso em: 9 jun. 2025.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Divisão Regional do Brasil: municípios, unidades da federação e regiões*. 2024. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/divisao-regional.html>. Acesso em: 9 jun. 2025.
- INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA (INEP). *Censo da Educação Superior 2022: Notas Estatísticas*. Brasília, 2022.
- KIMBALL, R.; ROSS, M. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3. ed. Hoboken: Wiley, 2013.
- MACIENTE, A. N.; CASERO, M. P. *A inserção dos diplomados em cursos de engenharia no mercado de trabalho formal*. Brasília, 2019.

MARCHEZI, G. P. *DATAGED: um Data Mart do Cadastro Geral de Empregados e Desempregados*. Monografia (Bacharelado em Sistemas de Informação) — Coordenadoria de Informática, Instituto Federal do Espírito Santo, Espírito Santo, 2025.

METABASE. *Metabase: BI for everyone*. 2024. Disponível em: <https://www.metabase.com>. Acesso em: 8 jun. 2025.

MINCER, J. *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research, 1974.

MINISTÉRIO DO TRABALHO E EMPREGO. *Classificação Brasileira de Ocupações (CBO)*. 2024. Disponível em: <http://www.mtecbo.gov.br/cbosite/pages/home.jsf>. Acesso em: 9 jun. 2025.

MINISTÉRIO DO TRABALHO E EMPREGO. *Relação Anual de Informações Sociais (RAIS): sobre a RAIS*. 2024. Disponível em: <https://www.rais.gov.br/sitio/sobre.jsf>. Acesso em: 8 jun. 2025.

SANTOS, J. P.; SILVA, R. M. *Construção de um Data Warehouse aplicado ao Sistema Único de Saúde (SUS)*. Monografia (Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação)) — Instituição de Ensino Superior, Brasil, 2014.

TURBAN, E.; SHARDA, R.; DELEN, D. *Business Intelligence, Analytics, and Data Science: A Managerial Perspective*. 4. ed. Boston: Pearson, 2019.



FOLHA DE APROVAÇÃO DE TCC N° 26/2026 - DECOM (11.56.03)

N° do Protocolo: 23062.033552/2026-12

Belo Horizonte-MG, 29 de junho de 2026.

PEDRO HENRIQUE ROCHA DE CASTRO

**DATA WAREHOUSE DE ALUNOS DO CEFET-MG E DA RAIS (RELAÇÃO ANUAL DE
INFORMAÇÕES SOCIAIS)**

Trabalho de Conclusão de Curso apresentado ao Curso de **Engenharia de Computação** do Centro Federal de Educação Tecnológica de Minas Gerais, do Campus **Nova Gameleira**, como parte do requisito para obtenção do grau de Bacharel em **Engenheiro de Computação**.

Aprovado em 26 de junho de 2026.

Banca Examinadora:

Prof. Dr. Evandrino Gomes Barros - Orientador
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Me. Ulisses Cotta Cavalca - Coorientador
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. Bruno André Santos
Centro Federal de Educação Tecnológica de Minas Gerais

Prof. Dr. João Ferando Machry Sarubbi
Centro Federal de Educação Tecnológica de Minas Gerais

**Belo Horizonte-MG
2026**

(Assinado digitalmente em 30/06/2026 09:20)

BRUNO ANDRE SANTOS

PROFESSOR DO MAGISTERIO SUPERIOR

DECOM (11.56.03)

Matrícula: ###594#8

(Assinado digitalmente em 29/06/2026 11:24)

EVANDRINO GOMES BARROS

PROFESSOR ENS BASICO TECN TECNOLOGICO

DECOM (11.56.03)

Matrícula: ###062#4

(Assinado digitalmente em 29/06/2026 16:08)

JOAO FERNANDO MACHRY SARUBBI

PROFESSOR ENS BASICO TECN TECNOLOGICO

PPGLS (11.52.04)

Matrícula: ###028#3

(Assinado digitalmente em 29/06/2026 12:03)

ULISSES COTTA CAVALCA

PROFESSOR ENS BASICO TECN TECNOLOGICO

DECOM (11.56.03)

Matrícula: ###106#7

Visualize o documento original em <https://sig.cefetmg.br/public/documentos/index.jsp> informando seu número: **26**, ano: **2026**, tipo: **FOLHA DE APROVAÇÃO DE TCC**, data de emissão: **29/06/2026** e o código de verificação: **c17c995317**