

Leonardo do Rêgo Espíndola

**Análise Exploratória sobre Dados de
Desempenho Acadêmico da Graduação de
Ciência da Computação**

Recife

2025

Leonardo do Rêgo Espíndola

Análise Exploratória sobre Dados de Desempenho Acadêmico da Graduação de Ciência da Computação

Trabalho apresentado ao Programa de Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Universidade Federal de Pernambuco – UFPE

Centro de Informática

Bacharelado em Ciência da Computação

Orientador: Prof. Dr. Márcio Lopes Cornélio

Recife

2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Espíndola, Leonardo do Rêgo.

Análise exploratória sobre dados de desempenho acadêmico da graduação de ciência da computação / Leonardo do Rêgo Espíndola. - Recife, 2025.
108 p. : il., tab.

Orientador(a): Márcio Lopes Cornélio

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Informática, Ciências da Computação - Bacharelado, 2025.

Inclui referências, apêndices.

1. Business Intelligence. 2. Data Warehouse. 3. Desempenho Acadêmico. 4. Evasão Escolar. 5. Analítica Educacional. I. Cornélio, Márcio Lopes. (Orientação). II. Título.

000 CDD (22.ed.)

Leonardo do Rêgo Espíndola

Análise Exploratória sobre Dados de Desempenho Acadêmico da Graduação de Ciência da Computação

Trabalho apresentado ao Programa de Graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Aprovado em: 12 / 08 / 2025

BANCA EXAMINADORA

Prof. Dr. Márcio Lopes Cornélio
Orientador
Universidade Federal de Pernambuco

Profa. Dra. Anjolina Grisi de Oliveira
Examinadora
Universidade Federal de Pernambuco

Agradecimentos

Primeiramente, expresso minha profunda gratidão à minha mãe, Maria Antonieta, por toda a dedicação, amor e suporte incondicional ao longo de minha vida. Sua criação carinhosa e seus ensinamentos foram fundamentais para que eu pudesse chegar até aqui, moldando o homem que me tornei e fornecendo as bases sólidas para todas as minhas conquistas.

Às minhas queridas tias Maria da Conceição, Maria Betania e Maria das Graças, agradeço por sempre estarem presentes como pilares de apoio durante toda minha trajetória. O carinho, a sabedoria e o suporte constante de vocês foram essenciais nos momentos mais desafiadores, e sou eternamente grato por ter vocês em minha vida.

À minha irmã, expresso minha imensa admiração e gratidão. Sua força, inteligência e companheirismo são fontes constantes de inspiração para mim. Obrigado por ser não apenas uma irmã, mas também uma grande amiga e exemplo a ser seguido.

Ao meu orientador, Prof. Dr. Márcio Lopes Cornélio, agradeço por aceitar esta empreitada e por toda a paciência, orientação e conhecimento compartilhado ao longo deste trabalho. Sua dedicação e expertise foram fundamentais para o desenvolvimento desta pesquisa.

Ao Centro de Informática da UFPE, expresso minha gratidão por ser um ambiente acadêmico ímpar, que proporciona excelência em ensino e pesquisa. Este lugar especial não apenas me formou profissionalmente, mas também contribuiu significativamente para meu crescimento pessoal e intelectual.

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho, meu sincero obrigado.

Resumo

Este trabalho apresenta o desenvolvimento de uma solução de *Business Intelligence* para análise exploratória do desempenho acadêmico dos alunos do curso de Ciência da Computação da Universidade Federal de Pernambuco (UFPE). Utilizando técnicas de data warehousing e visualização de dados, o estudo investiga padrões relacionados ao desempenho dos estudantes, com foco na identificação de características associadas à evasão e retenção. A pesquisa implementa um modelo dimensional (*Star Schema*) e um processo ETL (Extraction, Transformation, Load) estruturado em camadas, utilizando o SQL Server como repositório central e o Power BI como ferramenta de visualização. A partir da análise de dados coletados entre 2009 e 2019, foram exploradas nove questões relacionadas ao desempenho por forma de ingresso, distribuição demográfica, disciplinas críticas, evolução temporal dos indicadores e tempo de conclusão do curso. Os resultados revelaram padrões distintos: alunos ingressantes por cotas apresentaram maiores taxas de reprovação (32,03% vs. 16,78%); observou-se disparidade significativa na distribuição por sexo (87,4% masculino, 12,6% feminino); as taxas de reprovação concentram-se nos períodos iniciais (39,07% no 1º período e 43,94% no 2º período); o tempo médio de conclusão (5,75 anos) supera o previsto (4,5 anos); e as taxas de evasão são mais elevadas entre alunos cotistas (23,18%) comparadas aos de ampla concorrência (18,00%). Os dashboards desenvolvidos permitem análises multidimensionais através de operações OLAP, fornecendo insights valiosos para apoiar a gestão acadêmica em tomadas de decisão baseadas em evidências. O trabalho contribui com a proposta de um modelo dimensional específico para análise acadêmica e com recomendações para intervenções focadas na melhoria do desempenho estudantil e redução da evasão.

Palavras-chave: *Business Intelligence. Data Warehouse. Desempenho Acadêmico. Evasão Escolar. Analítica Educacional.*

Abstract

This work presents the development of a *Business Intelligence* solution for exploratory analysis of academic performance of Computer Science students at the Federal University of Pernambuco (UFPE). Using *data warehousing* and data visualization techniques, the study investigates patterns related to student performance, focusing on identifying characteristics associated with dropout and retention. The research implements a dimensional model (*Star Schema*) and a layered ETL (Extraction, Transformation, Load) process, using SQL Server as the central repository and Power BI as the visualization tool. Based on data collected between 2009 and 2019, nine specific questions related to performance by admission type, demographic distribution, critical courses, temporal evolution of indicators, and time to degree completion were explored. The results revealed distinct patterns: quota students showed higher failure rates (32.03% vs. 16.78%); significant disparity was observed in gender distribution (87.4% male, 12.6% female); failure rates concentrate in initial periods (39.07% in the 1st semester and 43.94% in the 2nd semester); the average completion time (5.75 years) exceeds the expected (4.5 years); and dropout rates are higher among quota students (23.18%) compared to open competition students (18.00%). The developed dashboards allow multidimensional analyses through OLAP operations, providing valuable insights to support academic management in evidence-based decision-making. The work contributes with the proposal of a specific dimensional model for academic analysis and with recommendations for interventions focused on improving student performance and reducing dropout rates.

Keywords: *Business Intelligence. Data Warehouse. Academic Performance. Student Dropout. Educational Analytics.*

Lista de ilustrações

Figura 1 – Modelo Estrela (<i>Star Schema</i>)	31
Figura 2 – Modelo Floco de Neve (<i>Snowflake Schema</i>)	32
Figura 3 – Operações OLAP	36
Figura 4 – Modelo Dimensional Implementado	62
Figura 5 – Dashboard Principal	65
Figura 6 – Dashboard de Conclusão, Evasão e Retenção	66
Figura 7 – Média de Notas por Período	69
Figura 8 – Percentual de Reprovação por Período	70
Figura 9 – Evolução Histórica - Média de Notas	71
Figura 10 – Evolução Histórica - Percentual de Reprovação	71
Figura 11 – Evolução Histórica - Percentual de Faltas	72
Figura 12 – Padrões de Desempenho por Forma de Ingresso	72
Figura 13 – Padrões de Desempenho por Sexo	75
Figura 14 – Tempo Médio de Conclusão Histórico	76
Figura 15 – Tempo Médio Excedente	77
Figura 16 – Indicadores de Evasão e Retenção	80
Figura 17 – Distribuição Demográfica - Faixa Etária e Sexo	81

Lista de tabelas

Tabela 1 – Principais operações OLAP	36
Tabela 2 – Estrutura da dimensão Alunos	63
Tabela 3 – Estrutura da dimensão Disciplinas	63
Tabela 4 – Estrutura da dimensão Professores	63
Tabela 5 – Estrutura da dimensão Datas	64
Tabela 6 – Estrutura da tabela fato Matriculas	64
Tabela 7 – Métricas de desempenho implementadas	66
Tabela 8 – Síntese dos resultados das questões exploratórias	82

Lista de abreviaturas e siglas

ABNT	Associação Brasileira de Normas Técnicas
BI	<i>Business Intelligence</i>
BPM	Business Performance Management
CIn	Centro de Informática
CSV	Comma-Separated Values
DAX	Data Analysis Expressions
DW	<i>Data Warehouse</i>
ETL	Extraction, Transformation, Load
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
IRA	<i>Índice</i> de Rendimento Acadêmico
KPI	Key Performance Indicator (Indicador-Chave de Desempenho)
LGPD	Lei Geral de Proteção de Dados
OLAP	On-Line Analytical Processing
OLTP	On-Line Transaction Processing
SIGA	Sistema de Informações e Gestão Acadêmica
SISU	Sistema de Seleção Unificada
SQL	Structured Query Language
STI	Superintendência de Tecnologia da Informação
TCC	Trabalho de Conclusão de Curso
UFPE	Universidade Federal de Pernambuco
XLS	Excel Spreadsheet

Sumário

1	INTRODUÇÃO	21
1.1	Problema de Pesquisa	22
1.2	Objetivos	24
1.2.1	Objetivo Geral	24
1.2.2	Objetivos Específicos	24
1.3	Justificativa	24
1.4	Estrutura do Trabalho	25
2	FUNDAMENTAÇÃO TEÓRICA	27
2.1	<i>Business Intelligence</i>	27
2.2	<i>Data Warehouse</i>	29
2.3	Modelos de Dados para <i>Data Warehouse</i>	31
2.3.1	Modelo Estrela (<i>Star Schema</i>)	31
2.3.2	Modelo Floco de Neve (<i>Snowflake Schema</i>)	32
2.3.3	Vantagens e Desvantagens dos Modelos	33
2.4	OLAP - On-Line Analytical Processing	34
2.4.1	Cubo OLAP	35
2.4.2	Operações OLAP	36
2.5	Indicadores Educacionais	37
2.5.1	Principais Indicadores no Ensino Superior	37
2.5.2	Uso de Indicadores na Gestão Acadêmica	40
3	TRABALHOS RELACIONADOS	42
3.1	Análise de Desempenho Acadêmico	42
3.2	<i>Business Intelligence</i> na Educação	43
3.3	Fatores Associados à Evasão no Ensino Superior	44
3.4	Relação com o Presente Trabalho	45
4	METODOLOGIA	47
4.1	Contexto do Estudo	47
4.2	Fonte de Dados e Coleta	48
4.3	Aspectos Éticos e Proteção de Dados	51
4.4	Processo de ETL	51
4.4.1	Extração (Extraction)	52
4.4.2	Transformação (Transformation)	54
4.4.3	Carga (Load)	55

4.5	Modelo de Dados Adotado	55
4.6	Ferramentas Utilizadas	56
4.7	Questões Exploratórias	56
4.8	Métricas e Indicadores Desenvolvidos	58
5	IMPLEMENTAÇÃO DO PROJETO	60
5.1	Arquitetura da Solução	60
5.2	Processo de ETL Detalhado	60
5.2.1	Extração dos Dados	61
5.2.2	Transformação dos Dados	61
5.2.3	Carga no <i>Data Warehouse</i>	61
5.3	Modelagem Dimensional	62
5.3.1	Tabelas Dimensionais	62
5.3.2	Tabela Fato	64
5.3.3	Relacionamentos	64
5.4	Desenvolvimento de Dashboards	64
5.4.1	Dashboard Principal	64
5.4.2	Dashboard de Conclusão, Evasão e Retenção	65
5.4.3	Modelo no Power BI	65
5.5	Métricas e KPIs Definidos	66
5.5.1	Métricas de Desempenho Acadêmico	66
5.5.2	Métricas de Progresso Acadêmico	67
5.5.3	KPIs Específicos para as Questões Exploratórias	67
6	RESULTADOS E DISCUSSÃO	69
6.1	Análise do Desempenho por Período do Curso	69
6.1.1	Média de Notas por Período	69
6.1.2	Percentual de Reprovação por Período	70
6.1.3	Evolução Histórica dos Indicadores	71
6.2	Padrões de Desempenho por Forma de Ingresso	72
6.3	Padrões de Desempenho por Sexo	75
6.4	Análise do Tempo de Conclusão	76
6.5	Análise da Evasão e Retenção	79
6.6	Distribuição Demográfica dos Alunos	80
6.7	Síntese dos Resultados	81
7	CONSIDERAÇÕES FINAIS	84
7.1	Conclusões	84
7.2	Contribuições	85
7.3	Limitações	86
7.4	Trabalhos Futuros	88

REFERÊNCIAS	92
-----------------------	----

APÊNDICES	95
------------------	-----------

	APÊNDICE A – SCRIPTS SQL	96
A.1	Script para criação dos schemas	96
A.2	Script para criação da tabela de arquivos	96
A.3	Scripts para procedimentos de carga de dados	96
A.4	Scripts para criação das tabelas da Camada 1	98
A.5	Scripts para criação das views da Camada 2	99
A.6	Scripts para criação das views da Camada 3	100
A.7	Scripts para criação das tabelas do DW	101
	APÊNDICE B – INDICADORES DAX	104
B.1	Indicadores de Desempenho Acadêmico	104
B.2	Indicadores de Evasão e Retenção	105
B.3	Indicadores de Matrícula e Ingresso	105
B.4	Indicadores de Conclusão	106
B.5	Indicadores Complementares	107
B.6	Indicadores de Situação Especial	107

1 Introdução

O crescimento das organizações está fortemente ligado a tomadas de decisões corretas, sendo este um ponto chave da gestão. A existência de dados que permitam uma análise adequada maximiza as chances de sucesso. As organizações geram e armazenam uma grande quantidade de dados referentes às atividades realizadas. Desta forma, a aplicação de técnicas de *Business Intelligence* no processo de tomada de decisão estratégica nas organizações torna-se necessária, uma vez que decisões equivocadas podem comprometer seu êxito (CHAUDHURI; DAYAL; NARASAYYA, 2011).

De acordo com os dados do Censo da Educação Superior, publicados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), as taxas de evasão do ensino superior brasileiro são alarmantes. Os dados mostram que 49% dos discentes que ingressaram no ensino superior em 2010 abandonaram os cursos dentro de um intervalo de cinco anos. Nas instituições privadas, a evasão chegou a 53%, e nas instituições públicas alcançou 47% nas municipais, 38% nas estaduais e 43% nas federais. Há diversos fatores que podem estar associados a esses números como a região do país, do tipo de instituição (se pública ou privada), da organização acadêmica (universidade, centro universitário, faculdades, instituto ou campus), da modalidade de ensino (presencial ou a distância) além de elementos que são precedentes à própria vida acadêmica como fatores financeiros, socioemocionais e pedagógicos (FILHO et al., 2007).

No contexto específico dos cursos de tecnologia e engenharia, incluindo Ciência da Computação, as taxas de evasão tendem a ser ainda mais elevadas, atingindo frequentemente patamares superiores a 50% em algumas instituições. Esta situação se agrava quando consideramos que tais cursos são estratégicos para o desenvolvimento tecnológico e econômico do país, formando profissionais em áreas de alta demanda no mercado de trabalho.

A evasão é, certamente, um dos problemas que afligem as instituições de ensino em geral. A busca de suas causas tem sido objeto de muitos trabalhos e pesquisas educacionais. A saída dos estudantes de forma precoce representa perdas que podem afetar as instituições em âmbitos sociais, acadêmicos e econômicos. No setor público, são recursos investidos sem o devido retorno. No setor privado, é uma importante perda de receitas. Em ambos os casos, a evasão produz ociosidade de professores, funcionários, equipamentos e espaço físico.

Além dos aspectos econômicos diretos, a evasão gera impactos sociais significativos. Cada aluno que abandona o curso representa uma oportunidade perdida de mobilidade social e desenvolvimento profissional. Para as famílias, especialmente aquelas de menor

renda, a não conclusão do ensino superior pode perpetuar ciclos de limitação socioeconômica. Para a sociedade como um todo, a perda de potenciais profissionais qualificados em áreas estratégicas compromete a capacidade de inovação e competitividade nacional.

O estudo interno, realizado pelas instituições de ensino com base em seus dados, pode ser muitas vezes mais detalhado porque é possível institucionalizar-se um mecanismo de acompanhamento da evasão, registrando os diversos casos, agrupando e analisando subgrupos, ou diferentes situações (cancelamento, trancamento, transferência, desistência, por exemplo) e assim organizar tabelas e gráficos que demonstrem a evolução da evasão para buscar formas de combatê-la com fundamento nos resultados (LOBO, 2012).

A gestão acadêmica por não ter acesso facilitado aos seus dados muitas vezes fica incapacitada de atuar proativamente, pois padrões de dificuldade só são identificados tardiamente, quando já houve a saída do discente, ou esse já assumiu forte intenção de deixar o curso ou instituição (BARBOSA et al., 2017). Com um ferramental adequado que auxilie a gestão a ter uma visão holística da situação do aluno e seu desempenho, é possível a construção de um plano de ação para o seu auxílio e manutenção na instituição.

Esta limitação informacional é particularmente crítica em cursos com alta complexidade técnica, como Ciência da Computação, onde dificuldades específicas em disciplinas fundamentais podem ser preditores precoces de evasão. A capacidade de identificar estes padrões em tempo hábil para intervenção representa um diferencial competitivo importante para as instituições de ensino superior.

O uso de técnicas e ferramentas de *Business Intelligence* (BI) pode proporcionar esse suporte informacional necessário para a gestão acadêmica. As soluções de BI permitem a integração de dados de diversas fontes, sua transformação em informações úteis e a disponibilização de visualizações e dashboards interativos que facilitam a análise e a tomada de decisão (TURBAN et al., 2008).

Neste contexto, o presente trabalho propõe o desenvolvimento de uma solução de *Business Intelligence* para análise exploratória do desempenho acadêmico dos alunos do curso de Ciência da Computação da Universidade Federal de Pernambuco (UFPE). Utilizando técnicas de *data warehousing* e visualização de dados, busca-se investigar padrões que possam estar associados ao desempenho dos estudantes, com foco na identificação de características relacionadas à evasão e retenção no curso.

1.1 Problema de Pesquisa

Diante do contexto apresentado, este trabalho busca responder à seguinte questão: Como uma solução de *Business Intelligence* pode auxiliar na análise exploratória do desempenho acadêmico dos alunos de graduação em Ciência da Computação da UFPE,

fornecendo insights para apoiar estratégias de redução da evasão e melhoria da qualidade do ensino?

A formulação desta questão principal emerge da necessidade de superar limitações identificadas na literatura sobre gestão acadêmica baseada em dados. Tradicionalmente, as análises de desempenho estudantil em instituições de ensino superior têm sido realizadas de forma fragmentada, utilizando relatórios estáticos que não permitem exploração interativa dos dados ou identificação de padrões complexos que envolvem múltiplas variáveis simultaneamente.

A escolha específica do curso de Ciência da Computação da UFPE como objeto de estudo justifica-se por suas características particulares: trata-se de um curso com histórico de excelência acadêmica, mas que enfrenta desafios típicos da área de computação, como alta exigência matemática e conceitual, além de apresentar disparidades demográficas significativas. Estas características tornam o curso um caso de estudo relevante para investigar como técnicas de *Business Intelligence* podem contribuir para uma compreensão mais aprofundada dos fatores associados ao sucesso e permanência estudantil.

A partir desta questão principal, derivam-se as seguintes questões específicas:

- a) Quais padrões podem ser observados no desempenho acadêmico dos estudantes de Ciência da Computação da UFPE quando analisados sob diferentes perspectivas temporais e demográficas?
- b) Existem diferenças nos padrões de desempenho acadêmico entre estudantes que ingressaram por diferentes formas de ingresso (ampla concorrência e cotas)?
- c) Como se caracteriza a distribuição de reprovações ao longo dos períodos do curso e quais disciplinas representam gargalos mais significativos no fluxo curricular?
- d) Qual o perfil demográfico dos estudantes e sua possível relação com o desempenho acadêmico, considerando aspectos como distribuição por gênero e faixa etária?
- e) Quais são as disciplinas com maiores taxas de reprovação e como elas se relacionam com a progressão dos alunos no curso?
- f) Como o tempo médio de conclusão do curso se compara com o tempo previsto na matriz curricular e quais fatores podem estar associados a estas diferenças?
- g) Como têm evoluído os principais indicadores de desempenho acadêmico ao longo da série histórica disponível e que insights esta evolução temporal pode fornecer sobre mudanças institucionais ou conjunturais?
- h) Quais características distinguem alunos que conseguem manter progressão regular no curso daqueles que apresentam dificuldades de permanência ou evasão?

- i) Como a análise integrada destes diferentes aspectos pode orientar estratégias institucionais de melhoria do ensino e redução da evasão?

Estas questões específicas foram formuladas considerando tanto os aspectos técnicos da implementação de soluções de *Business Intelligence* quanto às necessidades práticas da gestão acadêmica. Cada questão foi estruturada de forma a ser investigável através dos dados disponíveis e a fornecer insights acionáveis para tomadas de decisão institucionais.

1.2 Objetivos

1.2.1 Objetivo Geral

Desenvolver uma solução de *Business Intelligence* para análise exploratória do desempenho acadêmico dos alunos de graduação em Ciência da Computação da UFPE, visando identificar padrões que possam estar associados à evasão e à retenção no curso.

1.2.2 Objetivos Específicos

- a) Implementar um *Data Warehouse* utilizando modelo dimensional para armazenamento estruturado de dados acadêmicos;
- b) Desenvolver um processo ETL completo para integração dos dados do Sistema de Informações e Gestão Acadêmica (SIG@);
- c) Criar *dashboards* analíticos que permitam visualizar indicadores de desempenho sob diferentes perspectivas;
- d) Identificar padrões e possíveis associações entre características dos alunos e seu desempenho acadêmico;
- e) Analisar a distribuição de reprovações por disciplinas e períodos do curso;
- f) Verificar possíveis diferenças no desempenho acadêmico conforme a forma de ingresso e características demográficas;
- g) Propor recomendações para a gestão acadêmica com base nos insights obtidos.

1.3 Justificativa

A relevância deste trabalho se apoia em três pilares principais:

1. Relevância Social e Institucional: A evasão no ensino superior representa um desperdício de recursos públicos e um problema social significativo. No contexto da

UFPE, uma instituição pública, cada vaga não aproveitada integralmente representa um investimento sem o retorno esperado para a sociedade. Segundo dados do INEP, a evasão nas universidades federais atinge 43%, o que evidencia a dimensão do problema e a necessidade de estudos que contribuam para sua redução.

2. Relevância Prática: A solução desenvolvida fornece ferramentas concretas para a gestão acadêmica, permitindo identificar precocemente padrões que podem indicar alunos em risco de evasão e disciplinas com altos índices de reprovação. Isso possibilita intervenções direcionadas e baseadas em evidências, como programas de monitoria, revisão de metodologias de ensino e acompanhamento personalizado. A implementação destas ações pode resultar em melhorias significativas nos indicadores acadêmicos, redução da evasão e aumento da eficiência institucional.

3. Relevância Teórica: O trabalho contribui para a literatura sobre análise de dados educacionais, aplicando conceitos de *Business Intelligence* em um contexto acadêmico específico e gerando conhecimento sobre padrões associados ao desempenho e permanência no curso de Ciência da Computação. O modelo dimensional desenvolvido pode servir como referência para estudos similares em outras instituições e cursos, enquanto as análises realizadas contribuem para a compreensão dos fatores que podem estar relacionados ao sucesso acadêmico em cursos de computação.

Adicionalmente, o trabalho se justifica pelo potencial de transferência de conhecimento e tecnologia entre diferentes áreas. A aplicação de técnicas de *Business Intelligence*, tradicionalmente utilizadas no contexto empresarial, ao ambiente educacional demonstra a versatilidade destas ferramentas e sua capacidade de gerar valor em diferentes domínios.

1.4 Estrutura do Trabalho

Este trabalho está organizado em sete capítulos, estruturados da seguinte forma:

O **Capítulo 1** apresenta a introdução, contextualizando o problema da evasão no ensino superior e a importância do uso de técnicas de *Business Intelligence* na gestão acadêmica. São apresentados também o problema de pesquisa, os objetivos e a justificativa do trabalho.

O **Capítulo 2** traz a fundamentação teórica, abordando os conceitos de *Business Intelligence*, *Data Warehouse*, modelos dimensionais (*Star Schema* e *Snowflake Schema*), tecnologia OLAP e indicadores educacionais. São apresentadas as definições, características e aplicações destes conceitos.

O **Capítulo 3** discute trabalhos relacionados, apresentando estudos sobre análise de desempenho acadêmico, aplicações de *Business Intelligence* na educação e fatores

associados à evasão no ensino superior. Estabelece-se também a relação destes trabalhos com a presente pesquisa.

O **Capítulo 4** detalha a metodologia utilizada, descrevendo o contexto do estudo, a fonte dos dados, os aspectos éticos, o processo ETL, o modelo de dados adotado, as ferramentas utilizadas, as questões exploratórias e as métricas e indicadores desenvolvidos.

O **Capítulo 5** descreve a implementação do projeto, detalhando a arquitetura da solução, o processo ETL, a modelagem dimensional, o desenvolvimento dos *dashboards* e as métricas e KPIs definidos.

O **Capítulo 6** apresenta os resultados obtidos e sua discussão, analisando os padrões observados por período do curso, forma de ingresso, sexo, tempo de conclusão, evasão e retenção, e distribuição demográfica dos alunos.

O **Capítulo 7** traz as considerações finais, incluindo as conclusões, contribuições, limitações e sugestões para trabalhos futuros.

Por fim, são apresentadas as referências bibliográficas utilizadas e os apêndices, que incluem os *scripts SQL* desenvolvidos e as fórmulas *DAX* implementadas no *Power BI*.

2 Fundamentação Teórica

Este capítulo apresenta o arcabouço teórico que fundamenta esta pesquisa, abordando os conceitos de *Business Intelligence*, *Data Warehouse*, modelos dimensionais, tecnologia OLAP e indicadores educacionais.

2.1 *Business Intelligence*

Business Intelligence (BI) é definido por Turban et al. (2008) como um sistema abrangente que combina arquitetura, bancos de dados, ferramentas analíticas e aplicações para facilitar a tomada de decisão em organizações. O termo funciona como um “guardachuva” que engloba diversos conceitos e tecnologias focados na transformação de dados brutos em informações úteis para decisões estratégicas, táticas e operacionais.

Segundo Chaudhuri, Dayal e Narasayya (2011), BI pode ser compreendido como um conjunto de técnicas e ferramentas que permitem a transformação de dados brutos em informações significativas e úteis para análise de negócios. O objetivo principal é permitir que os gestores tomem decisões mais fundamentadas baseadas em evidências concretas extraídas dos dados organizacionais.

A evolução do conceito de *Business Intelligence* está intrinsecamente ligada ao crescimento exponencial do volume de dados gerados pelas organizações e à necessidade de extrair valor estratégico desta informação. Diferentemente dos sistemas transacionais tradicionais, que são otimizados para operações do dia a dia, os sistemas de BI são projetados especificamente para análise, relatórios e tomada de decisão (INMON, 2005).

De acordo com Chen, Chiang e Storey (2012), as ferramentas de BI oferecem recursos para análise de dados históricos, atuais e preditivos, gerando relatórios, *dashboards* e análises que ajudam a identificar tendências, padrões e insights que de outra forma poderiam passar despercebidos. Esta capacidade analítica multitemporal é particularmente valiosa no contexto educacional, onde padrões de desempenho podem emergir ao longo de semestres ou anos letivos.

Turban et al. (2008) descrevem quatro componentes principais de um sistema de BI:

1. ***Data Warehouse (DW)***: Repositório que armazena e consolida dados de diferentes fontes organizacionais, preparando-os para análise. No contexto educacional, pode integrar dados de sistemas acadêmicos, bibliotecas, laboratórios e outros sistemas institucionais.

2. Business Analytics: Conjunto de ferramentas e técnicas para manipular, minerar e analisar os dados armazenados no *Data Warehouse*, incluindo consultas *ad-hoc*, relatórios, *OLAP* e mineração de dados. Estas ferramentas permitem desde análises descritivas simples até modelagem preditiva complexa.

3. Business Performance Management (BPM): Componente que monitora e analisa o desempenho da organização, geralmente utilizando métricas e indicadores-chave de desempenho (KPIs). Na educação, pode incluir indicadores como taxas de conclusão, tempo médio de formação e satisfação estudantil.

4. Interface de Usuário: Componente que permite aos usuários interagir com o sistema BI, como *dashboards* e ferramentas de visualização. A qualidade desta interface é crucial para adoção e efetividade do sistema pelos gestores educacionais.

A arquitetura típica de um sistema de BI, conforme descrita por Turban et al. (2008), envolve várias camadas e componentes que trabalham em conjunto para transformar dados em conhecimento acionável:

1. Camada de Fontes de Dados: Envolve os diversos sistemas operacionais e bancos de dados que servem como origem dos dados, como sistemas ERP, CRM, arquivos planilhas, etc. No ambiente educacional, inclui sistemas de gestão acadêmica, plataformas de aprendizagem, sistemas de biblioteca e outros repositórios institucionais.

2. Camada de ETL: Responsável por extrair dados das fontes, transformá-los conforme necessário (limpeza, padronização, etc.) e carregá-los no *Data Warehouse*. Esta camada é crítica para garantir qualidade e consistência dos dados analíticos.

3. Camada de Data Warehouse: Armazena os dados em um formato otimizado para análise, geralmente seguindo um modelo dimensional que facilita consultas complexas e agregações.

4. Camada de Servidor OLAP: Permite análises multidimensionais dos dados, suportando operações como *slice*, *dice*, *drill-down* e *roll-up* que são essenciais para exploração interativa.

5. Camada de Apresentação: Fornece interfaces para que os usuários acessem, analisem e visualizem os dados, incluindo *dashboards*, relatórios e ferramentas de análise *ad-hoc*.

No contexto educacional, o BI pode ser aplicado para analisar diversos aspectos do ambiente acadêmico, como desempenho dos alunos, eficácia das práticas pedagógicas, gestão de recursos e padrões relacionados à evasão. Instituições de ensino superior podem utilizar sistemas de BI para monitorar indicadores acadêmicos, identificar estudantes que apresentem padrões de risco, otimizar a oferta de disciplinas e fundamentar decisões de gestão (SILVA et al., 2016).

A implementação de um sistema de BI envolve várias etapas, desde a identificação das necessidades de informação até a disponibilização das ferramentas de análise para os usuários finais. Segundo Kimball e Ross (2013), este processo deve seguir uma metodologia estruturada, que inclui o planejamento do projeto, a definição dos requisitos de negócio, a modelagem dimensional, o desenvolvimento do ETL, a criação de aplicações analíticas e a manutenção contínua do sistema.

A adoção de *Business Intelligence* no setor educacional tem crescido significativamente nos últimos anos, impulsionada pela necessidade de prestação de contas, competição por recursos e pressão por melhoria da qualidade educacional. Instituições que implementam soluções de BI relatam benefícios como melhor compreensão do desempenho estudantil, identificação precoce de problemas acadêmicos e tomada de decisões mais eficazes baseadas em evidências empíricas.

2.2 *Data Warehouse*

Data Warehouse (DW) refere-se a uma estrutura que armazena e consolida dados de diferentes fontes. Sua principal função é fornecer um relacionamento entre dados de origens diferentes. Há uma diferença fundamental entre um DW e um banco de dados tradicional. Em uma visão de arquitetura de dados, o DW estaria uma camada acima do banco, pois este agrega diferentes bancos e unifica os dados que antes não tinham uma relação direta para gerar uma camada de conhecimento onde a análise será feita.

Segundo Inmon (2005, p. 29), um *Data Warehouse* é “uma coleção de dados orientada por assunto, integrada, não volátil e variável em relação ao tempo, que tem por objetivo dar suporte aos processos de tomada de decisão”. Esta definição clássica destaca quatro características fundamentais:

1. Orientado por assunto: Os dados são organizados em torno de temas ou áreas de negócio específicas (como vendas, clientes, produtos), e não em torno de aplicações ou processos. No contexto educacional, os assuntos podem incluir alunos, disciplinas, desempenho acadêmico, entre outros.

2. Integrado: Os dados provenientes de diferentes fontes são consolidados e padronizados, resolvendo inconsistências de nomenclatura, unidades de medida, formatos, etc. Por exemplo, códigos de disciplinas, formatos de datas ou categorias de situação acadêmica que podem variar entre sistemas são uniformizados no DW.

3. Não volátil: Uma vez carregados no DW, os dados não são modificados, preservando o histórico das informações. Atualizações são realizadas através da adição de novos registros, não pela alteração dos existentes.

4. Variável em relação ao tempo: O DW mantém dados históricos, permitindo análises temporais e identificação de tendências. Cada registro é associado a um momento específico, permitindo análises como “evolução do desempenho ao longo dos semestres” ou “comparação de taxas de evasão entre anos”.

Kimball e Ross (2013) propõem uma abordagem complementar à de Inmon, definindo o *Data Warehouse* como uma cópia dos dados transacionais estruturados especificamente para consultas e análises. Eles enfatizam a importância do modelo dimensional, que organiza os dados em fatos (medidas numéricas) e dimensões (contextos para análise), visando facilitar o entendimento dos usuários e otimizar o desempenho das consultas.

A arquitetura de um *Data Warehouse* pode variar conforme as necessidades da organização, mas geralmente inclui os seguintes componentes, segundo Kimball e Ross (2013):

1. Staging Area (Área de Preparação): Região intermediária onde os dados são temporariamente armazenados durante o processo ETL, antes de serem carregados no DW.

2. Data Warehouse Core (Núcleo do DW): Composto pelas tabelas de fatos e dimensões que armazenam os dados integrados e históricos.

3. Data Marts: Subconjuntos do DW focados em áreas específicas de negócio, que podem ser implementados como visões ou como estruturas físicas separadas.

4. Metadata Repository (Repositório de Metadados): Armazena informações sobre a estrutura, origem e transformações dos dados no DW.

No contexto educacional, um *Data Warehouse* pode integrar dados de diversos sistemas, como:

- Sistema de gestão acadêmica (matrículas, notas, frequência)
- Sistema de vestibular/ingresso (dados de admissão)
- Sistema de biblioteca (empréstimos, consultas)
- Plataformas de aprendizagem virtual (interações, atividades *online*)
- Sistemas de assistência estudantil (bolsas, auxílios)

Esta integração permite análises abrangentes que não seriam possíveis utilizando os sistemas isoladamente. Por exemplo, é possível avaliar se alunos que utilizam frequentemente a biblioteca apresentam padrões diferenciados de desempenho acadêmico, ou se há associações entre o tipo de ingresso e a taxa de conclusão do curso.

2.3 Modelos de Dados para *Data Warehouse*

A modelagem de dados é um aspecto crucial no desenvolvimento de um *Data Warehouse*, pois afeta diretamente o desempenho das consultas, a facilidade de uso e a capacidade de adaptação a novos requisitos. Dois modelos principais são utilizados em projetos de DW: o Modelo Estrela (*Star Schema*) e o Modelo Floco de Neve (*Snowflake Schema*).

2.3.1 Modelo Estrela (*Star Schema*)

O modelo estrela é caracterizado por sua estrutura desnormalizada, onde existe uma tabela fato central (que armazena as medidas numéricas e chaves para as dimensões) relacionando-se com várias tabelas dimensionais (que contêm atributos descritivos). Esta estrutura recebe este nome pois, quando representada graficamente, assemelha-se a uma estrela, com a tabela fato no centro e as dimensões ao seu redor.

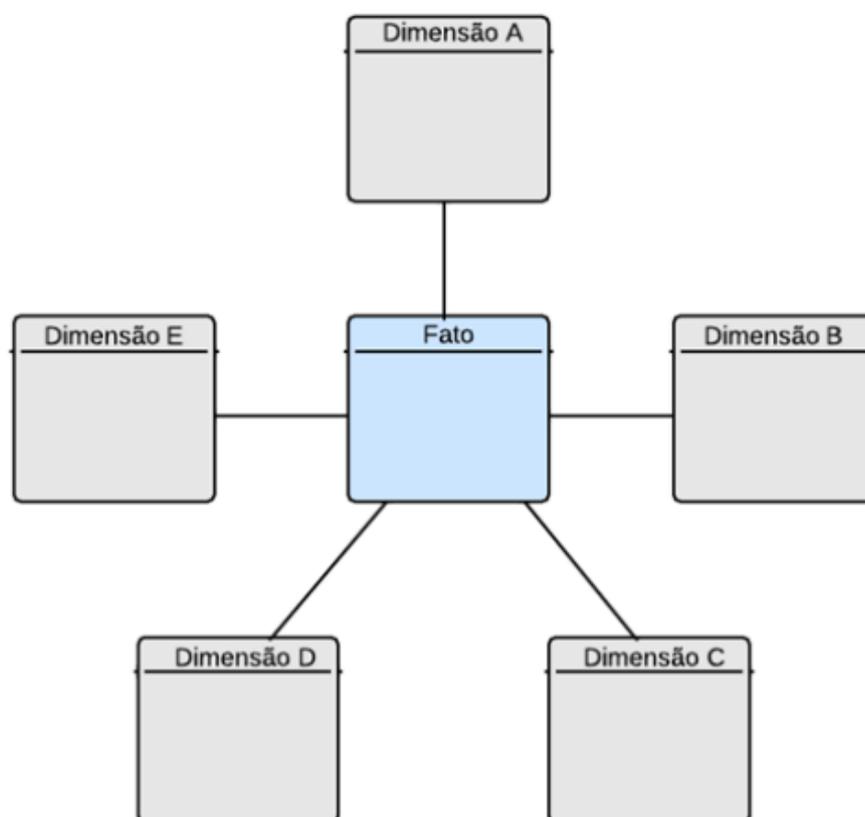


Figura 1 – Modelo Estrela (*Star Schema*)

Fonte: Autor (2025)

Segundo Kimball e Ross (2013), a tabela fato contém dois tipos de atributos:

- 1. Medidas:** Valores numéricos que representam métricas do negócio (como quantidade vendida, valor, número de aprovações, média de notas).

2. Chaves estrangeiras: Atributos que estabelecem relacionamentos com as tabelas dimensionais, permitindo analisar as medidas sob diferentes perspectivas.

As tabelas dimensionais, por sua vez, contêm atributos descritivos que fornecem contexto para as medidas na tabela fato. Por exemplo, uma dimensão “Tempo” pode incluir atributos como ano, semestre, mês e dia; uma dimensão “Aluno” pode conter informações como nome, sexo, data de nascimento e forma de ingresso.

No contexto acadêmico, um modelo estrela típico poderia ter uma tabela fato “Matrículas” contendo medidas como nota final e número de faltas, relacionada a dimensões como “Aluno”, “Disciplina”, “Professor” e “Tempo” (período letivo).

2.3.2 Modelo Floco de Neve (*Snowflake Schema*)

O modelo floco de neve é uma variação do modelo estrela onde as tabelas dimensionais são normalizadas, ou seja, são divididas em múltiplas tabelas relacionadas entre si. Esta estrutura recebe este nome porque, representada graficamente, assemelha-se a um floco de neve, com ramificações a partir das dimensões principais.

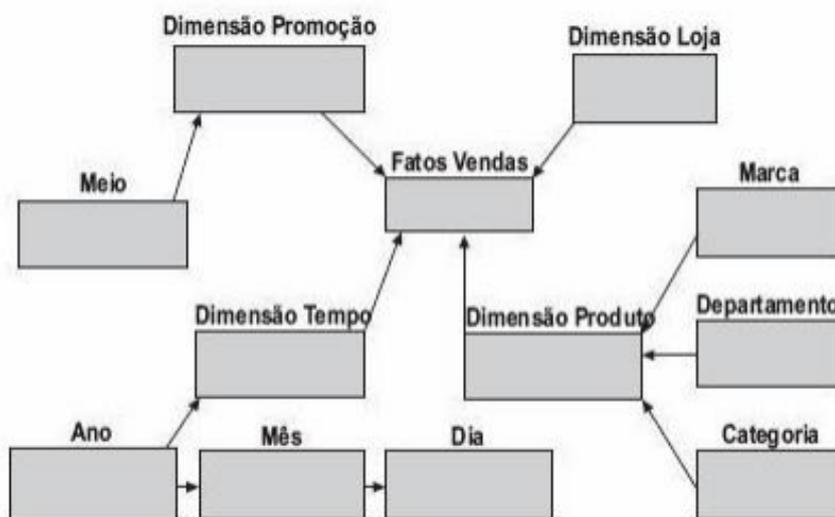


Figura 2 – Modelo Floco de Neve (*Snowflake Schema*)

Fonte: Autor (2025)

Um exemplo clássico de normalização em um modelo floco de neve é a decomposição da dimensão tempo, onde atributos hierárquicos como ano, semestre, mês e dia são organizados em tabelas separadas relacionadas entre si. Por exemplo, a dimensão “Ano” se relaciona com as dimensões “Semestre” e “Mês”, que por sua vez se relacionam com a dimensão “Dia”.

No contexto educacional, uma dimensão “Disciplina” poderia ser normalizada, criando tabelas separadas para “Departamento”, “Área de Conhecimento” e “Pré-requisitos”, todas relacionadas à tabela principal da disciplina.

2.3.3 Vantagens e Desvantagens dos Modelos

Cada modelo possui vantagens e desvantagens que devem ser consideradas ao projetar um *Data Warehouse*. A escolha entre eles depende das características específicas do projeto e dos requisitos de análise.

Vantagens do Modelo Estrela:

- Simplicidade: Estrutura intuitiva e de fácil compreensão para usuários finais.
- Desempenho: Menor número de junções (joins) necessárias nas consultas, resultando em melhor performance.
- Facilidade de implementação: Processo de ETL mais simples devido à estrutura desnormalizada.
- Flexibilidade: Facilidade para adicionar novos atributos às dimensões.

Desvantagens do Modelo Estrela:

- Redundância de dados: As tabelas dimensionais desnormalizadas podem conter informações repetidas, aumentando o espaço de armazenamento necessário.
- Limitações em hierarquias complexas: Dificuldade em representar hierarquias complexas sem criar redundância.
- Integridade referencial: Maior dificuldade em manter a integridade dos dados devido à desnormalização.

Vantagens do Modelo Floco de Neve:

- Economia de espaço: A normalização reduz a redundância de dados, economizando espaço de armazenamento.
- Manutenção da integridade: Facilita a manutenção da integridade referencial dos dados.
- Representação de hierarquias: Melhor suporte para representar hierarquias complexas nas dimensões.

- Compatibilidade com ferramentas: Algumas ferramentas de OLAP são otimizadas para trabalhar com este modelo.

Desvantagens do Modelo Floco de Neve:

- Complexidade: Estrutura mais complexa, dificultando a compreensão pelos usuários finais.
- Desempenho: Maior número de junções (joins) necessárias nas consultas, podendo impactar negativamente o desempenho.
- Implementação mais complexa: Processo de ETL mais elaborado devido à normalização das dimensões.

Segundo Kimball e Ross (2013), o modelo estrela é geralmente recomendado para a maioria das aplicações de *Data Warehouse* devido à sua simplicidade e desempenho. O modelo floco de neve, por sua vez, pode ser mais adequado em situações específicas, como quando há limitações severas de espaço de armazenamento ou necessidade de representar hierarquias muito complexas.

2.4 OLAP - On-Line Analytical Processing

OLAP (On-Line Analytical Processing) é uma tecnologia projetada para consultas complexas em grandes volumes de dados. Diferente dos sistemas *OLTP (On-Line Transaction Processing)*, que são otimizados para operações transacionais do dia a dia, os sistemas OLAP são desenvolvidos para análises multidimensionais e agregações de dados (Codd; Codd; Salley, 1993).

Segundo Kimball e Ross (2013), OLAP fornece uma visão multidimensional dos dados, permitindo que usuários analisem informações de negócios em diferentes perspectivas. Esta tecnologia suporta análises complexas e *ad-hoc*, permitindo que usuários explorem os dados de forma interativa, sem a necessidade de programação.

Codd, Codd e Salley (1993) definiram doze regras para caracterizar um sistema OLAP, entre elas:

1. **Visão conceitual multidimensional:** Os dados são apresentados em um modelo multidimensional intuitivo.
2. **Transparência:** O usuário não precisa conhecer a complexidade das estruturas subjacentes.
3. **Acessibilidade:** Acesso uniforme aos dados, independentemente da fonte.

4. **Desempenho consistente:** O desempenho não deve degradar significativamente com o aumento do número de dimensões.
5. **Arquitetura cliente/servidor:** O sistema deve ser capaz de operar em um ambiente cliente/servidor.
6. **Dimensionalidade genérica:** Todas as dimensões devem ser equivalentes em estrutura e capacidades operacionais.
7. **Manipulação de dados esparsos:** O sistema deve ajustar seu armazenamento físico para tratar eficientemente matrizes esparsas.
8. **Suporte a multiusuário:** Múltiplos usuários devem poder acessar e manipular os mesmos dados simultaneamente.

2.4.1 Cubo OLAP

O cubo OLAP é uma estrutura multidimensional que permite a visualização e análise de dados de diferentes perspectivas. Nestas estruturas, os dados são organizados em dimensões (como tempo, localização, produto) e medidas (valores numéricos que são analisados, como vendas, custos) (INMON, 2005).

Segundo Kimball e Ross (2013), um cubo OLAP representa um conjunto de dados organizados de forma que:

1. As medidas (fatos) são agrupadas por dimensões.
2. As dimensões podem formar hierarquias (por exemplo, ano > semestre > mês > dia).
3. As medidas podem ser agregadas em diferentes níveis hierárquicos.

No contexto educacional, um cubo OLAP poderia ter dimensões como “Aluno”, “Disciplina”, “Professor” e “Tempo”, com medidas como “Nota Final”, “Número de Faltas” e “Taxa de Aprovação”. Isso permitiria análises como a média de notas por disciplina ao longo dos semestres, ou o desempenho dos alunos agrupados por professor e curso.

Existem três principais tipos de implementação de cubos OLAP:

1. **MOLAP (Multidimensional OLAP):** Armazena dados em estruturas multidimensionais especializadas, oferecendo desempenho otimizado para consultas, mas requerendo processamento prévio (pré-cálculo de agregações).

2. **ROLAP (Relational OLAP):** Utiliza bancos de dados relacionais para armazenar dados e metadados, traduzindo consultas multidimensionais em *SQL*. Oferece maior flexibilidade e escalabilidade, mas pode ter desempenho inferior em algumas operações.

3. HOLAP (Hybrid OLAP): Combina características do MOLAP e ROLAP, armazenando dados detalhados em bancos relacionais e agregações em estruturas multidimensionais, buscando equilibrar desempenho e flexibilidade.

2.4.2 Operações OLAP

As operações OLAP permitem aos usuários interagir com os dados para realizar análises sob diferentes perspectivas. Estas operações são fundamentais para a exploração interativa dos dados e para a obtenção de insights que não seriam evidentes em análises estáticas.

Tabela 1 – Principais operações OLAP

Tabela 1 – Principais operações OLAP

Operação	Descrição
<i>Slice</i>	Seleciona uma fatia específica do cubo, fixando o valor de uma dimensão
<i>Dice</i>	Seleciona um subcubo especificando intervalos de valores para múltiplas dimensões
<i>Drill down</i>	Navega de um nível mais agregado para um mais detalhado em uma hierarquia
<i>Roll up</i>	Navega de um nível mais detalhado para um mais agregado em uma hierarquia
<i>Drill-across</i>	Conecta dois ou mais cubos que compartilham dimensões conformes
<i>Pivot</i>	Rearranja as dimensões do cubo para visualizar os dados de diferentes perspectivas

Fonte: Adaptado de Kimball e Ross (2013)

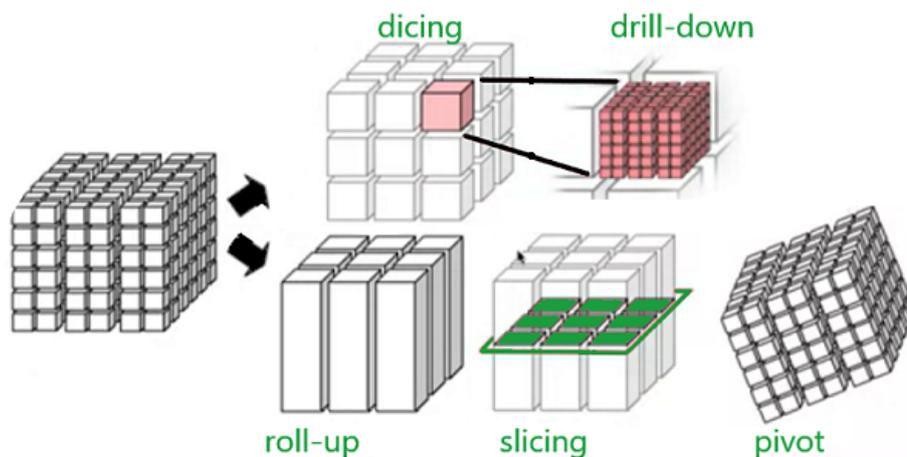


Figura 3 – Operações OLAP

No contexto da análise acadêmica, estas operações podem ser aplicadas de diversas formas:

- **Slice:** Analisar o desempenho de um curso específico.

- ***dice***: Analisar o desempenho de alunos do sexo feminino em disciplinas de programação no ano de 2018.
- ***drill down***: Partir da análise da taxa de aprovação global para a análise por período, por disciplina e finalmente por turma.
- ***roll up***: Agregar notas individuais para calcular médias por disciplina, departamento e centro acadêmico.
- ***Drill-across***: Comparar taxas de aprovação com médias de notas para uma mesma disciplina.
- ***Pivot***: Reorganizar uma visualização que mostra desempenho por disciplina e período para exibir o desempenho por professor e disciplina.

Estas operações permitem que gestores acadêmicos e pesquisadores explorem os dados de forma flexível e interativa, adaptando as análises conforme novas questões surgem durante o processo de investigação.

2.5 Indicadores Educacionais

Os indicadores educacionais são métricas utilizadas para avaliar diferentes aspectos do sistema educacional, fornecendo informações quantitativas sobre o desempenho, eficiência e qualidade da educação (INEP, 2017). Estes indicadores são essenciais para o planejamento, monitoramento e avaliação de políticas educacionais.

No Brasil, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) é responsável por produzir e disseminar indicadores educacionais em âmbito nacional, cobrindo desde a educação básica até o ensino superior. Estes indicadores são utilizados tanto para o acompanhamento das políticas públicas quanto para a avaliação institucional e planejamento interno das instituições de ensino.

2.5.1 Principais Indicadores no Ensino Superior

No contexto do ensino superior, diversos indicadores são utilizados para avaliar o desempenho acadêmico, a eficiência das instituições e a qualidade dos cursos. A seguir, são apresentados os principais indicadores relevantes para este estudo:

a) Taxa de Evasão

A taxa de evasão representa o percentual de alunos que abandonam o curso em relação ao total de alunos matriculados. Este indicador é calculado pela fórmula:

$$\text{Taxa de Evasão} = (\text{Número de Evadidos} / \text{Total de Matriculados}) \times 100$$

Segundo Filho et al. (2007), a evasão no ensino superior é um fenômeno complexo que pode estar associado a fatores socioeconômicos, institucionais e individuais. Os autores classificam a evasão em três tipos:

1. **Evasão do curso:** Quando o aluno deixa o curso, mas permanece na instituição.
2. **Evasão da instituição:** Quando o aluno deixa a instituição, mas continua no sistema de ensino superior.
3. **Evasão do sistema:** Quando o aluno abandona o ensino superior completamente.

Para fins deste trabalho, considera-se evadido o aluno que deixou o curso de Ciência da Computação da UFPE, independentemente de ter ingressado em outro curso da instituição ou em outra instituição de ensino superior.

b) Taxa de Retenção

A retenção ocorre quando o aluno permanece no curso além do tempo previsto para sua conclusão. A taxa de retenção é calculada como:

$$\text{Taxa de Retenção} = \left(\frac{\text{Número de Alunos Retidos}}{\text{Total de Matriculados}} \right) \times 100$$

A retenção pode estar associada a reprovações em disciplinas, trancamentos ou outros fatores que prolongam o tempo de permanência do aluno na instituição (DIAS; THEÓPHILO; LOPES, 2010). É importante observar que a retenção pode preceder a evasão, pois alunos que permanecem muito tempo sem progredir no curso têm maior probabilidade de desistência.

Um indicador complementar à taxa de retenção é o Tempo Médio Excedente, que mede quanto tempo a mais, em média, os alunos levam para concluir o curso além do tempo previsto:

$$\text{Tempo Médio Excedente} = \text{Tempo Médio de Conclusão} - \text{Tempo Previsto de Conclusão}$$

c) Taxa de Aprovação/Reprovação

Estas taxas indicam o percentual de aprovação e reprovação em disciplinas específicas ou no conjunto de disciplinas de um curso. São calculadas como:

$$\text{Taxa de Aprovação} = \left(\frac{\text{Número de Aprovações}}{\text{Total de Matrículas em Disciplinas}} \right) \times 100$$

$$\text{Taxa de Reprovação} = (\text{Número de Reprovações} / \text{Total de Matrículas em Disciplinas}) \times 100$$

A reprovação pode ocorrer por nota insuficiente ou por frequência inadequada (excesso de faltas). Análises detalhadas podem diferenciar estes dois tipos de reprovação, fornecendo insights sobre suas possíveis causas. Uma alta taxa de reprovação por nota pode sugerir problemas na compreensão do conteúdo ou na metodologia de ensino, enquanto uma alta taxa de reprovação por falta pode sugerir problemas de engajamento ou conflitos de horário.

d) Índice de Rendimento Acadêmico (IRA)

O IRA é um indicador que sintetiza o desempenho acadêmico do aluno, geralmente calculado com base nas notas obtidas nas disciplinas, ponderadas pelos créditos ou carga horária das mesmas. A fórmula típica é:

$$\text{IRA} = \Sigma(\text{Nota} \times \text{Crédito}) / \Sigma(\text{Crédito})$$

Este indicador é frequentemente utilizado para classificação dos alunos, concessão de bolsas, seleção para programas especiais e monitoramento do desempenho individual. Variações do IRA podem considerar apenas as disciplinas aprovadas ou incluir também reprovações, com diferentes pesos.

e) Tempo Médio de Conclusão

Este indicador mede o tempo médio que os alunos levam para concluir o curso. É calculado como:

$$\text{Tempo Médio de Conclusão} = \text{Soma dos Tempos de Conclusão de Todos os Formados} / \text{Número de Formados}$$

Um tempo médio de conclusão significativamente superior ao tempo previsto no projeto pedagógico do curso pode sugerir problemas no currículo ou na progressão dos alunos (LIMA; ZAGO, 2018). Este indicador está diretamente relacionado à taxa de retenção e pode afetar a eficiência do sistema educacional como um todo.

f) Taxa de Conclusão

A taxa de conclusão representa o percentual de alunos que concluem o curso em relação ao total de ingressantes. É calculada como:

$$\text{Taxa de Conclusão} = (\text{Número de Concluintes} / \text{Número de Ingressantes}) \times 100$$

Este indicador pode ser calculado para uma coorte específica (acompanhando o desempenho de um grupo de alunos que ingressou no mesmo período) ou de forma mais geral (comparando o número de concluintes em um período com o número de ingressantes em períodos anteriores).

A taxa de conclusão é um indicador importante da eficiência do sistema educacional e está diretamente relacionada à taxa de evasão. Em um cenário ideal, a soma da taxa de conclusão e da taxa de evasão seria próxima de 100%, considerando que todos os alunos ou concluem o curso ou evadem. Na prática, há sempre um percentual de alunos em curso, especialmente aqueles retidos.

2.5.2 Uso de Indicadores na Gestão Acadêmica

Os indicadores educacionais são fundamentais para a gestão acadêmica, pois permitem a tomada de decisões fundamentada em evidências. Segundo Polidori, Marinho-Araujo e Barreyro (2006), o uso sistemático de indicadores educacionais contribui para uma gestão mais eficiente e para a melhoria contínua da qualidade da educação superior.

As principais aplicações dos indicadores na gestão acadêmica incluem:

1. Diagnóstico de situações: Identificar disciplinas com altas taxas de reprovação, períodos críticos para evasão, ou perfis de alunos que apresentem determinados padrões. Este diagnóstico permite direcionar ações específicas para os pontos mais críticos.

2. Planejamento estratégico: Definir metas e estratégias baseadas em evidências concretas. Por exemplo, estabelecer objetivos de redução da taxa de evasão ou do tempo médio de conclusão, com base no histórico dos indicadores.

3. Acompanhamento de desempenho: Monitorar a evolução dos indicadores ao longo do tempo, verificando a eficácia das ações implementadas e identificando tendências emergentes.

4. Avaliação de intervenções: Mensurar o impacto de mudanças curriculares, programas de monitoria, ou outras ações pedagógicas. A comparação dos indicadores antes e depois da intervenção permite avaliar sua eficácia.

5. Tomada de decisão: Fundamentar decisões relacionadas a alterações curriculares, alocação de recursos, ou programas de apoio ao estudante. Decisões baseadas em dados têm maior probabilidade de sucesso do que aquelas baseadas apenas em intuições ou tradições.

Tinto (2006) destaca que a utilização de indicadores para identificar precocemente estudantes que apresentem padrões de risco de evasão é uma prática eficaz, pois permite intervenções preventivas antes que a decisão de abandonar o curso seja tomada. Estas

intervenções podem incluir ações como aconselhamento acadêmico, programas de monitoria, acompanhamento psicopedagógico ou adaptações curriculares.

No contexto específico da gestão de cursos de computação, Kori et al. (2018) enfatizam a importância de monitorar o desempenho nas disciplinas iniciais de programação e matemática, pois estas frequentemente representam os primeiros obstáculos significativos para os estudantes e podem estar associadas à evasão subsequente.

A implementação de soluções de *Business Intelligence*, como a proposta neste trabalho, potencializa o uso dos indicadores educacionais na gestão acadêmica, ao facilitar a coleta, integração e visualização dos dados, além de permitir análises multidimensionais que revelam padrões e tendências não evidentes em análises mais simples.

3 Trabalhos Relacionados

A análise de dados acadêmicos para compreender o desempenho dos estudantes e os fatores associados à evasão tem sido tema de diversos estudos. Esta seção apresenta alguns trabalhos relevantes que utilizaram abordagens similares ou complementares à adotada nesta pesquisa.

3.1 Análise de Desempenho Acadêmico

Manhães et al. (2011) desenvolveram um sistema para predição de desempenho acadêmico utilizando técnicas de mineração de dados. Os autores aplicaram algoritmos de classificação, como árvores de decisão, redes neurais e máquinas de vetores de suporte, para identificar precocemente alunos com risco de evasão em uma universidade brasileira. Foram analisados dados de mais de 15.000 alunos de diferentes cursos, considerando variáveis como coeficiente de rendimento, número de reprovações e frequência às aulas. Os resultados mostraram que é possível prever com precisão de até 80% os alunos que não concluirão o curso, permitindo intervenções direcionadas. O estudo destaca a importância do primeiro ano acadêmico como período crítico para a adaptação do estudante e para a formação de hábitos de estudo que podem afetar todo o percurso universitário.

Iam-On e Boongoen (2017) utilizaram técnicas de *clustering* para agrupar estudantes com base em seu desempenho acadêmico e identificar padrões que poderiam indicar risco de evasão. O estudo foi realizado com dados de estudantes de uma universidade tailandesa, considerando informações sobre notas, frequência e interação com plataformas de aprendizagem *online*. Através do algoritmo *k-means*, os autores identificaram cinco grupos distintos de estudantes, desde aqueles com alto desempenho e engajamento até os que apresentavam sinais precoces de possível desistência. O estudo demonstrou que diferentes perfis de estudantes apresentam padrões distintos de dificuldades e que intervenções personalizadas para cada grupo podem ser mais efetivas.

Alves et al. (2018) analisaram especificamente o desempenho de estudantes em cursos de computação no Brasil, investigando fatores associados à reprovação em disciplinas introdutórias de programação. Utilizando técnicas de regressão logística, os autores identificaram que o desempenho prévio em matemática, a frequência às aulas práticas e a participação em grupos de estudo estavam significativamente associados ao sucesso nestas disciplinas. O estudo também revelou que alunos ingressantes por cotas não apresentavam diferenças significativas no desempenho em programação quando comparados aos ingressantes por ampla concorrência, após controlar outras variáveis.

No contexto da UFPE, Vasconcelos, Almeida e Monteiro (2019) analisaram dados de desempenho acadêmico de diversos cursos, incluindo Ciência da Computação, buscando identificar padrões associados à retenção e evasão. Os autores utilizaram técnicas de análise de sobrevivência para modelar o tempo até a ocorrência da evasão, considerando variáveis como sexo, forma de ingresso, desempenho no primeiro período e participação em atividades extracurriculares. O estudo identificou que o desempenho nas disciplinas do primeiro ano parece estar associado à conclusão do curso, e que a participação em programas de iniciação científica estava relacionada a menores taxas de evasão.

3.2 *Business Intelligence* na Educação

Ong et al. (2011) implementaram uma solução de *Business Intelligence* em uma instituição de ensino superior da Malásia, focando na análise de indicadores de desempenho institucional. O projeto envolveu a criação de um *Data Warehouse* que integrava dados de diversos sistemas, incluindo gestão acadêmica, recursos humanos e finanças. Os autores desenvolveram um modelo dimensional (*Star Schema*) e *dashboards* analíticos que permitiram aos gestores monitorar diversos indicadores, como taxa de conclusão, satisfação dos estudantes e produtividade docente. Os resultados demonstraram que a implementação do BI contribuiu para uma tomada de decisão mais eficiente e para a identificação precoce de problemas como alto índice de reprovação em disciplinas específicas e baixa taxa de conclusão em determinados cursos.

No contexto brasileiro, Silva et al. (2016) desenvolveram um sistema de BI para análise de dados educacionais em uma universidade federal. O sistema foi projetado para integrar dados de diversos sistemas existentes na instituição, permitindo análises sobre evasão, retenção e desempenho acadêmico. A arquitetura do sistema incluía um processo ETL para extração e transformação dos dados, um *Data Warehouse* modelado em esquema estrela e uma interface de visualização baseada em *dashboards* interativos. Os resultados permitiram aos gestores identificar fatores associados ao abandono do curso, como reprovações múltiplas em disciplinas específicas e longos períodos de inatividade acadêmica, possibilitando a implementação de ações preventivas.

Moscoso-Zea et al. (2019) apresentaram uma metodologia para implementação de soluções de BI em instituições de ensino superior, com foco na análise de desempenho acadêmico e gestão institucional. Os autores propuseram um framework que combina as abordagens de Kimball e Inmon para o desenvolvimento do *Data Warehouse*, adaptado às especificidades do contexto educacional. A metodologia foi aplicada em uma universidade equatoriana, resultando em um sistema que permitia análises multidimensionais do desempenho dos estudantes, considerando variáveis como perfil socioeconômico, desempenho prévio e participação em atividades extracurriculares. O estudo destacou a importância da

integração de dados qualitativos e quantitativos para uma compreensão abrangente dos fatores que podem estar relacionados ao sucesso acadêmico.

Guster e Brown (2012) aplicaram técnicas de *data warehousing* e OLAP para analisar a retenção de estudantes em uma universidade americana. Os autores desenvolveram um modelo dimensional que permitia análises por diversos ângulos, incluindo características demográficas, forma de ingresso, desempenho acadêmico e participação em programas de suporte. As análises revelaram que estudantes de primeira geração (primeiros em suas famílias a frequentar o ensino superior) apresentavam maior risco de evasão, especialmente quando não participavam de programas de tutoria ou comunidades de aprendizagem. O estudo demonstrou como as ferramentas de BI podem auxiliar na identificação de grupos específicos que requerem intervenções direcionadas.

3.3 Fatores Associados à Evasão no Ensino Superior

Tinto (2006) é um dos principais teóricos sobre o tema da evasão no ensino superior. O autor propôs um modelo de integração acadêmica e social que relaciona a persistência do estudante à sua capacidade de integrar-se tanto academicamente (através do desempenho e identificação intelectual com o curso) quanto socialmente (através de interações com colegas, professores e participação em atividades extracurriculares). Segundo o autor, fatores como envolvimento em atividades extracurriculares, relacionamento com professores e colegas, e desempenho acadêmico podem estar associados significativamente à decisão de permanecer ou abandonar o curso. O modelo de Tinto tem sido amplamente aplicado em estudos sobre evasão e tem se mostrado robusto em diferentes contextos culturais e institucionais.

Filho et al. (2007) realizaram um estudo abrangente sobre a evasão no ensino superior brasileiro, analisando dados de instituições públicas e privadas entre 2000 e 2005. Os autores identificaram que a evasão média anual no período foi de 22%, sendo maior nas instituições privadas (25%) do que nas públicas (12%). Entre os fatores possivelmente associados à evasão, destacaram-se problemas financeiros, escolha precoce da carreira, falta de conhecimento prévio sobre o curso e dificuldades de integração ao ambiente universitário. O estudo também identificou diferenças significativas nas taxas de evasão conforme a área de conhecimento, sendo os cursos das áreas de exatas e tecnológicas (como Ciência da Computação) aqueles com maiores índices de abandono.

No contexto específico dos cursos de computação, Kori et al. (2018) analisaram fatores possivelmente associados à evasão em universidades europeias. Os autores identificaram que dificuldades com disciplinas de matemática e programação nos primeiros períodos, além da falta de conhecimento prévio sobre o curso, estão entre os principais fatores que podem estar relacionados à desistência. O estudo também revelou diferenças de gênero nas razões para evasão: enquanto homens tendiam a citar dificuldades acadêmicas,

mulheres frequentemente mencionavam problemas de integração social e sentimentos de não pertencimento ao ambiente predominantemente masculino. Os autores sugeriram intervenções como tutoria por pares, comunidades de prática e reestruturação curricular como estratégias potencialmente efetivas para reduzir a evasão.

Barbosa et al. (2017) investigaram fatores de evasão em cursos de computação no Brasil, com foco em instituições públicas. Através de uma pesquisa com 331 estudantes evadidos de 15 instituições, os autores identificaram que os principais motivos para abandono incluíam: ingresso em outro curso de maior interesse, dificuldades com disciplinas específicas (especialmente algoritmos e cálculo), problemas de conciliação entre trabalho e estudo, e expectativas não correspondidas sobre o curso. O estudo também identificou que muitos estudantes tomam a decisão de evadir já nos primeiros períodos, reforçando a importância de ações preventivas logo no início do curso.

Costa e Gouveia (2018) analisaram a relação entre a forma de ingresso (ampla concorrência ou cotas) e a evasão em universidades federais brasileiras. Contrariando expectativas iniciais, o estudo não encontrou diferenças significativas nas taxas de evasão entre cotistas e não-cotistas quando considerados fatores como assistência estudantil e apoio acadêmico. De fato, em instituições com programas robustos de assistência estudantil, os cotistas apresentavam taxas de conclusão superiores às dos não-cotistas. Os autores sugeriram que políticas de inclusão precisam ser acompanhadas por políticas efetivas de permanência que atendam às necessidades específicas dos diferentes perfis de estudantes.

3.4 Relação com o Presente Trabalho

Esta pesquisa se diferencia dos trabalhos anteriores por combinar a implementação de uma solução completa de *Business Intelligence* (*Data Warehouse* e visualizações) com a análise específica do curso de Ciência da Computação da UFPE. Enquanto estudos anteriores focaram em aspectos gerais da evasão ou aplicaram técnicas isoladas de análise de dados, este trabalho adota uma abordagem integrada, desenvolvendo um modelo dimensional específico para dados acadêmicos e explorando múltiplas questões relacionadas ao desempenho e permanência no curso.

O modelo dimensional proposto por Silva et al. (2016) serviu como ponto de partida para o desenvolvimento do *Data Warehouse* desta pesquisa, porém foi adaptado para incluir aspectos específicos do contexto da UFPE, como a estrutura curricular do curso de Ciência da Computação e as políticas de ingresso da instituição. Os *dashboards* desenvolvidos também se inspiram nas visualizações propostas por Ong et al. (2011), mas com foco nos indicadores identificados como mais relevantes para o contexto em estudo.

Em relação aos fatores associados à evasão, este trabalho se alinha com as descobertas de Kori et al. (2018) sobre a importância das disciplinas iniciais de matemática

e programação, mas explora também outros aspectos como possíveis associações entre a forma de ingresso e características demográficas. A abordagem metodológica para análise destes fatores segue o recomendado por Tinto (2006), considerando tanto aspectos acadêmicos quanto sociais da experiência universitária.

Além disso, a utilização de operações OLAP para explorar os dados sob diferentes perspectivas e níveis de detalhe permite uma análise mais abrangente e flexível do que as abordagens tradicionais baseadas em relatórios estáticos ou consultas pré-definidas. Esta característica é particularmente importante no contexto educacional, onde a complexidade e multidimensionalidade dos fatores que podem estar relacionados ao desempenho requerem análises sofisticadas e adaptáveis.

A principal contribuição deste trabalho em relação aos anteriores é a apresentação de um framework completo para análise de dados acadêmicos, desde a extração e transformação dos dados brutos até a criação de *dashboards* interativos que permitem explorar questões específicas. Este framework pode ser adaptado e replicado em outros cursos e instituições, contribuindo para a disseminação das práticas de gestão baseada em evidências no ensino superior.

4 Metodologia

Esta seção apresenta a metodologia adotada para o desenvolvimento da pesquisa, descrevendo o contexto do estudo, as fontes de dados utilizadas, os aspectos éticos, o processo de ETL, o modelo de dados, as ferramentas empregadas, as questões exploratórias investigadas e as métricas desenvolvidas.

4.1 Contexto do Estudo

Este trabalho foi realizado no âmbito do curso de Bacharelado em Ciência da Computação da Universidade Federal de Pernambuco (UFPE). O curso foi criado em 1974 e é oferecido pelo Centro de Informática (CIn), sendo um dos mais tradicionais e reconhecidos do país na área de computação. O curso tem duração prevista de 9 semestres (4,5 anos) e oferece anualmente 100 vagas.

O Centro de Informática da UFPE é reconhecido como um centro de excelência em ensino e pesquisa em computação. Ao longo de sua história, tem formado profissionais que se destacam no mercado de trabalho e na academia, tanto nacional quanto internacionalmente. O curso de Ciência da Computação recebeu conceito máximo (5) nas últimas avaliações do Ministério da Educação e figura entre as melhores do país em diversos rankings.

A matriz curricular do curso é estruturada em disciplinas obrigatórias, eletivas e optativas, distribuídas ao longo de nove períodos. As disciplinas obrigatórias incluem fundamentos matemáticos, programação, teoria da computação, estruturas de dados, algoritmos, arquitetura de computadores, sistemas operacionais, engenharia de software, bancos de dados, redes de computadores, entre outras. Os dados obtidos são de estudantes do Perfil Curricular 2001. O currículo atual do Curso de Ciência da Computação teve início no ano letivo de 2024.

O ingresso no curso ocorre principalmente através do Sistema de Seleção Unificada (SISU), que utiliza as notas do Exame Nacional do Ensino Médio (ENEM). A UFPE adota políticas de ações afirmativas, reservando um percentual das vagas para candidatos de escolas públicas, de baixa renda e autodeclarados pretos, pardos e indígenas, conforme a Lei nº 12.711/2012 (Lei de Cotas). Há também outras formas de ingresso menos comuns, como transferência interna, transferência externa, reintegração e ingresso como portador de diploma.

A escolha deste curso como objeto de estudo deve-se à sua relevância acadêmica, ao histórico de dados disponíveis e à possibilidade de contribuir com insights valiosos para

a gestão acadêmica, visando a melhoria contínua da qualidade do ensino e a redução dos *índices* de evasão.

4.2 Fonte de Dados e Coleta

Os dados utilizados neste estudo foram obtidos junto à Superintendência de Tecnologia da Informação (STI) da UFPE, que é responsável pela gestão do Sistema de Informações e Gestão Acadêmica (SIG@). O SIG@ é o sistema integrado utilizado pela universidade para gerenciar informações acadêmicas, incluindo matrículas, notas, frequência, histórico escolar, entre outros aspectos.

O processo de solicitação e obtenção dos dados seguiu os protocolos institucionais estabelecidos pela universidade, envolvendo justificativa acadêmica da pesquisa, definição do escopo temporal e das variáveis necessárias, além de acordo de confidencialidade para tratamento adequado das informações pessoais dos estudantes. A STI forneceu os dados em formato estruturado, facilitando sua incorporação ao processo de análise.

Foram solicitados e fornecidos os seguintes conjuntos de dados:

1. Dados de discentes: Informações sobre os alunos matriculados no curso, incluindo dados demográficos, forma de ingresso, situação acadêmica atual e histórico de matrícula. A estrutura destes dados incluía os seguintes campos principais:

- Nome do aluno (posteriormente anonimizado)
- Data de admissão no formato MM/AAAA
- Previsão de conclusão conforme matriz curricular
- Data de conclusão efetiva (quando aplicável)
- Forma de ingresso (SISU, transferência, reintegração, etc.)
- Tipo de forma de ingresso (ampla concorrência ou cotista)
- Situação acadêmica atual (ativo, graduado, evadido, etc.)
- Curso e perfil do curso
- Data de nascimento para cálculo de idade
- Estado civil no momento da matrícula
- Sexo declarado
- Ano de conclusão do ensino médio

- Média geral acumulada no curso
- Coeficiente de rendimento acadêmico
- Faixa econômica declarada
- Status de bloqueio acadêmico

2. Dados de disciplinas: Informações sobre as disciplinas ofertadas ao longo do período analisado, incluindo código, nome, carga horária, período ideal e departamento responsável. A estrutura destes dados incluía:

- Período letivo no formato AAAA.S (ano.semestre)
- Curso de oferta da disciplina
- Código da disciplina conforme sistema acadêmico
- Nome completo da disciplina
- Carga horária total em horas
- Quantidade de aulas previstas no semestre

Esta base de dados sobre disciplinas é particularmente importante pois permite contextualizar as análises de desempenho considerando as características específicas de cada componente curricular, como carga horária e posicionamento na matriz do curso.

3. Dados de matrículas: Registros de todas as matrículas realizadas nas disciplinas, incluindo período letivo, nota final e situação (aprovado, reprovado por nota, reprovado por falta). A estrutura destes dados incluía:

- Nome do aluno e identificadores para relacionamento
- CPF do aluno (posteriormente anonimizado)
- Matrícula institucional do aluno
- Data de admissão para controle temporal
- Curso de vinculação
- Período letivo da matrícula
- Código e nome da disciplina cursada
- Média final obtida (escala 0-10)

- Situação final da matrícula (aprovado, reprovado por nota, reprovado por falta, trancado, etc.)
- Número de faltas registradas
- CPF e nome do professor responsável

Esta base representa o núcleo dos dados transacionais do estudo, contendo 47.401 registros de matrículas que cobrem todas as interações acadêmicas dos estudantes com as disciplinas ao longo do período analisado.

4. Dados de docentes: Informações sobre os professores que ministraram as disciplinas no período analisado, incluindo:

- Status de atividade (ativo/inativo)
- Nome completo
- CPF para identificação
- Escolaridade e formação acadêmica
- Títulos de graduação, especialização, mestrado e doutorado

Os dados foram fornecidos em formato *XLS* (planilhas *Excel*), abrangendo o período de 2009 a 2019, garantindo assim uma série histórica significativa para as análises. O arquivo principal “DADOS_MATRICULAS_2009_2019.xlsx” foi utilizado como fonte primária para a tabela fato de matrículas, contendo a maior parte dos registros transacionais necessários para as análises.

É importante destacar que os dados sobre forma de ingresso (ampla concorrência ou cotas) só começaram a ser registrados de maneira consistente a partir de 2015, com a implementação mais ampla das políticas de ações afirmativas e do SISU. Esta limitação temporal será considerada nas análises correspondentes, sendo explicitamente mencionada sempre que análises relacionadas à forma de ingresso forem apresentadas.

Por questões de privacidade e conformidade com a Lei Geral de Proteção de Dados (LGPD), todos os dados pessoais que permitissem a identificação direta dos alunos foram anonimizados durante o processo de ETL, substituindo nomes e documentos de identificação por códigos alfanuméricos únicos. Esta anonimização foi implementada de forma a preservar a capacidade de relacionar registros do mesmo indivíduo entre diferentes tabelas, mantendo a integridade analítica dos dados.

Os resultados são apresentados de forma agregada, sem possibilidade de identificação individual. Adicionalmente, foram implementadas salvaguardas para evitar a identificação

indireta de indivíduos através de combinações específicas de características, especialmente em grupos muito pequenos.

4.3 Aspectos Éticos e Proteção de Dados

Esta pesquisa envolveu a análise de dados acadêmicos de estudantes, requerendo cuidados especiais quanto à proteção da privacidade e conformidade com regulamentações vigentes.

Proteção de Dados Pessoais: Todos os dados pessoais que permitiam identificação direta dos estudantes foram anonimizados conforme a Lei Geral de Proteção de Dados (LGPD - Lei nº 13.709/2018). Nomes, documentos de identificação e outros dados pessoais sensíveis foram substituídos por códigos alfanuméricos, impossibilitando a identificação individual dos participantes.

Autorização Institucional: Os dados foram fornecidos pela Superintendência de Tecnologia da Informação (STI) da UFPE mediante solicitação formal, seguindo os protocolos institucionais para pesquisas acadêmicas. A utilização dos dados foi restrita exclusivamente aos propósitos desta pesquisa.

Apresentação dos Resultados: Todos os resultados são apresentados de forma agregada, sem possibilidade de identificação de indivíduos específicos. Foram adotadas medidas para garantir que grupos muito pequenos não fossem identificáveis indiretamente.

Armazenamento e Descarte: Os dados foram armazenados em ambiente seguro com acesso restrito ao pesquisador. Após a conclusão da pesquisa, os dados identificadores serão permanentemente descartados, mantendo-se apenas as análises agregadas para fins de registro acadêmico.

Natureza da Pesquisa: Este estudo caracteriza-se como uma análise exploratória de dados administrativos já coletados pela instituição para fins de gestão acadêmica, não envolvendo coleta primária de dados junto aos estudantes. Os resultados têm caráter descritivo e exploratório, visando identificar padrões nos dados disponíveis para apoiar a gestão acadêmica.

4.4 Processo de ETL

Para a construção do ambiente analítico, foi adotada a metodologia proposta por Ralph Kimball, conhecida como ETL (Extraction, Transformation and Load). Esta abordagem foi escolhida por sua eficácia comprovada em projetos de *Business Intelligence* e por sua adequação às necessidades deste estudo.

O processo ETL foi estruturado em três camadas principais:

1. Camada 1 (Extração): Armazena os dados brutos extraídos das planilhas do SIG@, sem qualquer transformação ou limpeza.

2. Camada 2 (Transformação): Aplica transformações nos dados brutos, como conversões de tipo, padronizações e estabelecimento de relacionamentos.

3. Data Warehouse (DW): Armazena os dados já transformados em um modelo dimensional (*Star Schema*) otimizado para consultas analíticas.

Adicionalmente, foi criada uma quarta camada para interface com a ferramenta de visualização:

4. Camada 3 (Apresentação): Fornece views e funções específicas para facilitar a integração com o Power BI.

4.4.1 Extração (Extraction)

A primeira fase consistiu na extração dos dados das planilhas fornecidas pela STI e sua inserção no banco de dados *SQL Server* sem qualquer alteração ou tratamento. Esta camada, denominada “Camada 1”, representa os dados em seu estado bruto e serve como ponto de partida para as transformações subsequentes.

Para esta etapa, foram criadas funções específicas no banco de dados:

- `pr_carregaArquivos`: Responsável por iniciar o processo de carga dos arquivos.
- `pr_inserePlanilha`: Realiza a inserção dos dados das planilhas nas tabelas correspondentes.
- `pr_nomeArquivos`: Gerencia os nomes dos arquivos no sistema.

Ao final desta etapa, foram criadas as seguintes tabelas:

Tabela 2 – Estrutura da tabela camada1.Alunos

```

1 CREATE TABLE camada1.Alunos (
2     NOME NVARCHAR(255) NULL,
3     ADMISSAO NVARCHAR(255) NULL,
4     PREVISAO_CONCLUSAO NVARCHAR(255) NULL,
5     CONCLUSAO_CURSO NVARCHAR(255) NULL,
6     FORMA_INGRESSO NVARCHAR(255) NULL,
7     TIPO_FORMA_INGRESSO NVARCHAR(255) NULL,
8     SITUACAO NVARCHAR(255) NULL,
9     CURSO NVARCHAR(255) NULL,
10    PERFIL_CURSO NVARCHAR(255) NULL,
11    DT_NASCIMENTO NVARCHAR(255) NULL,
12    ESTADO_CIVIL NVARCHAR(255) NULL,
13    SEXO NVARCHAR(255) NULL,
14    CONCLUSAO_ENS_MEDIO NVARCHAR(255) NULL,
15    MEDIA_GERAL NVARCHAR(255) NULL,

```

```
16     COEFICIENTE_RENDIMENTO NVARCHAR(255) NULL ,
17     FAIXA_ECONOMICA NVARCHAR(255) NULL ,
18     BLOCADO NVARCHAR(255) NULL
19 )
```

Tabela 3 – Estrutura da tabela camada1.Disciplinas

```
1 CREATE TABLE camada1.Disciplinas(
2     PERIODO_LETIVO nvarchar(255) NULL ,
3     CURSO nvarchar(255) NULL ,
4     ID_DISCIPLINA nvarchar(255) NULL ,
5     DISCIPLINA nvarchar(255) NULL ,
6     CARGA_HORARIA nvarchar(255) NULL ,
7     QUANTIDADE_AULAS nvarchar(255) NULL
8 )
```

Tabela 4 – Estrutura da tabela camada1.Matriculas

```
1 CREATE TABLE camada1.Matriculas(
2     NOME nvarchar(255) NULL ,
3     CPF nvarchar(255) NULL ,
4     MATRICULA nvarchar(255) NULL ,
5     ADMISSAO nvarchar(255) NULL ,
6     CURSO nvarchar(255) NULL ,
7     PERIODO_LETIVO nvarchar(255) NULL ,
8     ID_DISCIPLINA nvarchar(255) NULL ,
9     DISCIPLINA nvarchar(255) NULL ,
10    MEDIA_FINAL nvarchar(2550) NULL ,
11    SITUACAO_FINAL nvarchar(255) NULL ,
12    FALTAS nvarchar(255) NULL ,
13    CPF_PROFESSOR nvarchar(255) NULL ,
14    NOME_PROFESSOR nvarchar(255) NULL
15 )
```

Tabela 5 – Estrutura da tabela camada1.Professores

```
1 CREATE TABLE camada1.Professores(
2     ATIVO nvarchar(255) NULL ,
3     NOME nvarchar(255) NULL ,
4     CPF nvarchar(255) NULL ,
5     ESCOLARIDADE nvarchar(255) NULL ,
6     GRADUACAO nvarchar(255) NULL ,
7     APERFEICOAMENTO nvarchar(255) NULL ,
8     ESPECIALIZACAO nvarchar(255) NULL ,
9     MESTRADO nvarchar(255) NULL ,
10    DOUTORADO nvarchar(255) NULL
11 )
```

Também foi criada uma tabela auxiliar para armazenar informações sobre os arquivos processados:

```
1 create table Arquivos (  
2     id int identity(1,1),  
3     nomeArquivo varchar(255),  
4     nomeCurto varchar(50)  
5 )
```

4.4.2 Transformação (Transformation)

Na segunda fase, os dados da “Camada 1” foram processados, tratados e transformados para corrigir inconsistências, padronizar formatos e preparar os dados para análise. Esta camada, denominada “Camada 2”, foi implementada através de views SQL que aplicam transformações nos dados brutos.

Por exemplo, a view `camada2.Disciplinas` transforma os dados das disciplinas, convertendo tipos e criando identificadores únicos:

```
1 CREATE VIEW camada2.Disciplinas  
2 AS  
3 SELECT  
4     concat(d.ID_DISCIPLINA, '|', d.PERIODO_LETIVO) as idOrigem,  
5     d.PERIODO_LETIVO as periodoLetivo,  
6     dCurso,  
7     d.ID_DISCIPLINA as codigoDisciplina,  
8     d.DISCIPLINA as nomeDisciplina,  
9     try_convert(int, d.CARGA_HORARIA) as cargaHoraria,  
10    try_convert(int, d.QUANTIDADE_AULAS) as quantidadeAulas  
11 FROM camada1.Disciplinas d
```

Durante esta etapa, foram realizadas as seguintes transformações:

1. **Conversão de tipos:** Transformação de strings em tipos numéricos (INT, FLOAT) ou datas.
2. **Padronização de formatos:** Uniformização de formatos de notas (substituição de vírgulas por pontos).
3. **Criação de identificadores únicos:** Geração de chaves compostas para garantir unicidade.
4. **Estabelecimento de relacionamentos:** Criação de junções (joins) entre as tabelas.
5. **Limpeza de dados:** Remoção de registros duplicados e tratamento de valores nulos.

4.4.3 Carga (Load)

Na terceira fase, os dados transformados foram carregados no *Data Warehouse* (DW), estruturado segundo o modelo dimensional (*Star Schema*). Esta camada organiza os dados em tabelas de fato e dimensões, otimizadas para consultas analíticas.

Foram criadas as seguintes tabelas dimensionais:

- **dw.Alunos:** Dimensão que armazena informações sobre os alunos
- **dw.Disciplinas:** Dimensão que armazena informações sobre as disciplinas
- **dw.Professores:** Dimensão que armazena informações sobre os professores
- **dw.Datas:** Dimensão que armazena informações temporais (período letivo)

E a seguinte tabela fato:

- **dw.Matriculas:** Tabela que armazena os registros de matrículas e seus resultados

Para a construção da dimensão temporal, foi implementado um procedimento específico que gera automaticamente os registros para todos os períodos letivos no intervalo especificado.

4.5 Modelo de Dados Adotado

Para o desenvolvimento do *Data Warehouse*, foi adotado o modelo Estrela (*Star Schema*), devido às seguintes vantagens:

1. **Simplicidade:** A estrutura clara e intuitiva facilita a compreensão e manutenção do modelo.
2. **Desempenho:** Reduz o número de junções (joins) necessárias nas consultas, resultando em melhor performance.
3. **Flexibilidade:** Permite adicionar novas dimensões ou medidas com facilidade, sem impactar as estruturas existentes.
4. **Adequação às ferramentas:** Compatibilidade otimizada com o *Power BI* e outras ferramentas OLAP.

O modelo foi estruturado com a tabela **FatoMatricula** no centro, conectada às tabelas dimensionais através de chaves estrangeiras. Cada dimensão foi projetada para armazenar atributos relacionados a um aspecto específico do negócio (alunos, disciplinas, professores, tempo).

4.6 Ferramentas Utilizadas

Para o desenvolvimento deste projeto, foram utilizadas as seguintes ferramentas:

1. **SQL Server**: Sistema de gerenciamento de banco de dados relacional utilizado para armazenar os dados em todas as camadas do processo ETL. Foi escolhido devido à sua robustez, capacidade de processamento e ferramentas de administração integradas.
2. **T-SQL (Transact-SQL)**: Linguagem procedural do *SQL Server* utilizada para desenvolver as stored procedures, views e funções necessárias para o processo ETL.
3. **Microsoft Power BI**: Ferramenta de *Business Intelligence* da Microsoft utilizada para a criação de visualizações interativas e dashboards analíticos. Foi escolhida por sua integração nativa com o *SQL Server*, recursos avançados de visualização e facilidade de uso.

As métricas e KPIs foram implementados utilizando *DAX (Data Analysis Expressions)*, linguagem nativa do *Power BI* para cálculos analíticos.

4. **Microsoft Excel**: Utilizado para análises preliminares e manipulação dos dados de entrada, além de ser o formato original dos dados fornecidos pela STI.
5. **Visual Studio**: Ambiente de desenvolvimento integrado (IDE) utilizado para desenvolvimento e teste dos *scripts SQL*.

A escolha dessas ferramentas se baseou em sua robustez, disponibilidade, facilidade de uso e integração com os sistemas existentes na instituição.

4.7 Questões Exploratórias

A partir da revisão da literatura e do conhecimento sobre o contexto do curso, foram formuladas as seguintes questões exploratórias a serem investigadas através da análise dos dados:

Questão 1: Existem diferenças nos padrões de desempenho acadêmico (reprovações e faltas) entre alunos que ingressaram por diferentes formas de ingresso (ampla concorrência ou cotas)?

Esta questão busca verificar se podem ser observadas diferenças no desempenho acadêmico entre alunos que ingressaram por diferentes modalidades de acesso. A literatura apresenta resultados diversos sobre este tema, tornando relevante uma investigação no contexto específico do curso.

Questão 2: Existem diferenças nos padrões de desempenho acadêmico (reprovações e faltas) entre alunos de diferentes sexos?

Esta questão investiga possíveis disparidades de gênero no desempenho acadêmico. Considerando que cursos de computação tradicionalmente apresentam predominância masculina, é relevante verificar se existem diferenças no desempenho que possam sugerir barreiras específicas.

Questão 3: Como se caracteriza a distribuição de reprovações ao longo dos períodos do curso?

Esta questão verifica se existe concentração de reprovações em períodos específicos do curso. Estudos anteriores sugerem que disciplinas dos primeiros períodos tendem a apresentar taxas mais elevadas de reprovação.

Questão 4: Como tem evoluído a taxa de reprovação ao longo dos anos analisados?

Esta questão investiga mudanças nas taxas de reprovação ao longo do tempo. Variações podem indicar mudanças na metodologia de ensino, adaptações curriculares ou alterações no perfil dos ingressantes.

Questão 5: Como o tempo médio de conclusão do curso se compara com o tempo previsto na matriz curricular?

Esta questão analisa se o tempo médio de conclusão está próximo do previsto (4,5 anos), investigando possíveis gargalos no fluxo curricular.

Questão 6: Qual a caracterização do perfil demográfico dos estudantes em termos de distribuição por sexo e faixa etária?

Esta questão verifica o perfil demográfico dos estudantes, quantificando possíveis disparidades e sua evolução ao longo do tempo.

Questão 7: Como se caracterizam os padrões de evasão e retenção no curso, e existem diferenças conforme o perfil dos alunos?

Esta questão busca identificar características da evasão e retenção, investigando se determinados grupos apresentam padrões diferenciados.

Questão 8: Como se caracteriza a evolução temporal dos principais indicadores acadêmicos ao longo do período analisado (2009-2019)?

Esta questão investiga tendências e mudanças nos indicadores de desempenho ao longo da série histórica disponível. A análise temporal permite identificar padrões de melhoria ou deterioração dos indicadores, períodos atípicos que possam estar associados a mudanças institucionais, curriculares ou conjunturais, e a estabilidade ou variabilidade dos resultados acadêmicos. Esta análise é fundamental para contextualizar os demais achados

e identificar se os padrões observados são consistentes ao longo do tempo ou refletem momentos específicos da história do curso.

Questão 9: Quais disciplinas apresentam os maiores *índices* de dificuldade e como elas se distribuem ao longo da matriz curricular?

Esta questão busca identificar disciplinas que representam gargalos significativos no fluxo acadêmico, analisando tanto suas taxas de reprovação quanto sua posição na estrutura curricular. A identificação de disciplinas críticas permite compreender melhor os desafios enfrentados pelos estudantes em diferentes momentos do curso, verificar se existem períodos de maior concentração de dificuldades, e avaliar se disciplinas consideradas fundamentais (como matemática e programação inicial) efetivamente representam obstáculos maiores. Esta análise pode orientar ações específicas de apoio pedagógico e revisão de metodologias de ensino.

Para investigar essas questões, foram desenvolvidos *dashboards* específicos no *Power BI*, utilizando as operações *OLAP* (*slice, dice, drill down, roll up, drill-across* e *pivot*) para explorar os dados de diferentes perspectivas e níveis de detalhe. As nove questões abordam desde aspectos demográficos e temporais até análises específicas de disciplinas críticas e padrões de permanência no curso.

4.8 Métricas e Indicadores Desenvolvidos

Para a análise do desempenho acadêmico e suporte à gestão, foram desenvolvidos os seguintes indicadores:

1. **Taxa de Aprovação por Disciplina:** Percentual de alunos aprovados em relação ao total de matriculados na disciplina. Esta métrica permite identificar disciplinas com altas taxas de reprovação.
2. **Taxa de Evasão por Período:** Percentual de alunos que abandonam o curso em cada período letivo. Esta métrica ajuda a identificar momentos críticos.
3. **Índice de Retenção:** Tempo médio adicional que os alunos levam para concluir o curso além do previsto.
4. **Disciplinas Críticas:** Identificação das disciplinas com maiores taxas de reprovação e seu possível impacto na progressão dos alunos.
5. **Perfil de Risco:** Categorização dos alunos conforme possíveis fatores de risco de evasão ou retenção.
6. **Indicador de Progressão:** Relação entre créditos obtidos e créditos previstos para cada período do curso.

7. **Comparativo Histórico:** Evolução dos indicadores de desempenho ao longo do tempo.

Além destes indicadores principais, foram implementadas diversas métricas auxiliares para suporte às análises, como médias de notas por disciplina, período e forma de ingresso, percentuais de faltas, e distribuição de alunos por situação acadêmica.

Esses indicadores foram incorporados aos dashboards desenvolvidos, permitindo uma visão abrangente do desempenho acadêmico no curso. A implementação técnica destas métricas no Power BI será detalhada no capítulo seguinte.

5 Implementação do Projeto

Esta seção descreve em detalhes o processo de implementação da solução de *Business Intelligence*, incluindo a arquitetura adotada, o processo ETL, a modelagem dimensional, o desenvolvimento de dashboards e as métricas definidas.

5.1 Arquitetura da Solução

A solução de *Business Intelligence* implementada neste trabalho segue uma arquitetura em camadas, permitindo uma separação clara entre os dados brutos, seu processamento e sua disponibilização para análise. Esta abordagem facilita a manutenção do sistema e permite evoluções futuras sem impacto em componentes já estabelecidos.

A arquitetura é composta por quatro camadas principais:

1. **Camada 1 (Extração)**: Armazena os dados brutos extraídos das planilhas do SIG@, sem qualquer transformação ou limpeza.
2. **Camada 2 (Transformação)**: Implementada através de views SQL, esta camada aplica transformações nos dados brutos.
3. **Data Warehouse (DW)**: Armazena os dados transformados em um modelo dimensional otimizado para consultas analíticas.
4. **Camada 3 (Apresentação)**: Fornece views específicas para facilitar a integração com o Power BI.

O fluxo de dados segue a sequência: [Arquivos XLS] → [Camada 1] → [Camada 2] → [Data Warehouse] → [Camada 3] → [Power BI]

Esta arquitetura foi projetada considerando requisitos de rastreabilidade, manutibilidade, desempenho, extensibilidade e segurança.

5.2 Processo de ETL Detalhado

O processo de ETL foi implementado utilizando o SQL Server como repositório central e procedimentos armazenados para automatizar as operações de manipulação de dados.

5.2.1 Extração dos Dados

A extração foi realizada a partir de planilhas Excel fornecidas pela STI, contendo informações sobre alunos, disciplinas, matrículas e professores. Para automatizar este processo, foram desenvolvidos procedimentos armazenados específicos.

Durante o processo de extração, foram identificados alguns desafios:

1. **Cabeçalhos e rodapés nas planilhas:** Necessitaram filtragem durante a importação.
2. **Codificação de caracteres:** Caracteres especiais apresentavam problemas de codificação.
3. **Valores nulos e vazios:** Células vazias precisavam ser interpretadas corretamente.
4. **Consistência entre planilhas:** Dados referenciados em múltiplas planilhas nem sempre eram consistentes.

5.2.2 Transformação dos Dados

Na etapa de transformação, os dados brutos foram processados através de views na Camada 2, aplicando operações como:

1. **Conversão de tipos de dados:** Transformação de strings em tipos adequados
2. **Normalização de valores:** Padronização de formatos
3. **Criação de chaves únicas:** Geração de identificadores únicos
4. **Filtragem de registros inválidos:** Remoção de dados inconsistentes
5. **Derivação de novos atributos:** Cálculo de campos derivados

Durante este processo, foram identificados e tratados diversos problemas nos dados, incluindo registros sem correspondência, inconsistências em formatos de notas, dados faltantes e duplicidades.

5.2.3 Carga no *Data Warehouse*

Na etapa de carga, os dados transformados foram inseridos nas tabelas do *Data Warehouse*, seguindo o modelo dimensional. Procedimentos específicos garantiram a integridade e consistência dos dados durante esta fase.

5.3 Modelagem Dimensional

O modelo dimensional implementado segue a abordagem *Star Schema*, com uma tabela fato central conectada a várias dimensões.

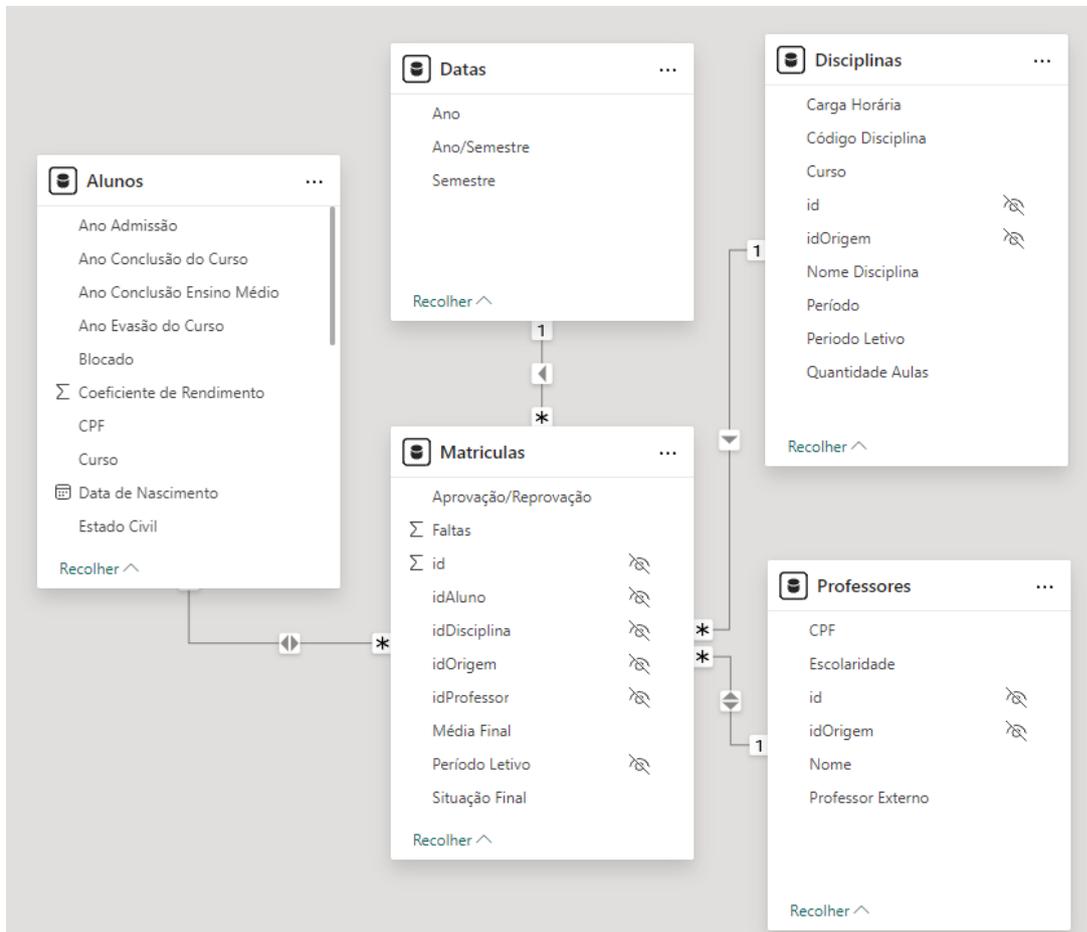


Figura 4 – Modelo Dimensional Implementado

Fonte: Autor (2025)

5.3.1 Tabelas Dimensionais

As tabelas dimensionais armazenam atributos descritivos que fornecem contexto para as métricas:

a) Dimensão Alunos (dw.Alunos)

Tabela 6 – Estrutura da dimensão Alunos

Esta dimensão inclui dados de identificação, informações acadêmicas, dados demográficos, indicadores de desempenho e status acadêmico.

b) Dimensão Disciplinas (dw.Disciplinas)

Tabela 7 – Estrutura da dimensão Disciplinas

Tabela 2 – Estrutura da dimensão Alunos

Atributo	Tipo	Descrição
id	INT	Chave primária surrogate
idOrigem	VARCHAR(50)	Identificador original do sistema
Nome	VARCHAR(120)	Nome do aluno (anonimizado)
formaIngresso	VARCHAR(30)	Forma de ingresso (SISU, transferência, etc.)
tipoFormaIngresso	VARCHAR(30)	Ampla concorrência ou cotista
Situacao	VARCHAR(30)	Situação acadêmica atual
Sexo	VARCHAR(1)	M ou F
Idade	INT	Idade atual do aluno

Fonte: Autor (2025)

Tabela 3 – Estrutura da dimensão Disciplinas

Atributo	Tipo	Descrição
id	INT	Chave primária surrogate
codigoDisciplina	VARCHAR(15)	Código da disciplina no sistema
nomeDisciplina	VARCHAR(250)	Nome completo da disciplina
cargaHoraria	INT	Carga horária em horas
periodoLetivo	VARCHAR(8)	Período de oferta (AAAA.S)

Fonte: Autor (2025)

Esta dimensão contém identificação das disciplinas, informações acadêmicas e parâmetros de carga horária.

c) Dimensão Professores (dw.Professores)

Tabela 8 – Estrutura da dimensão Professores

Tabela 4 – Estrutura da dimensão Professores

Atributo	Tipo	Descrição
id	INT	Chave primária surrogate
Nome	VARCHAR(50)	Nome do professor
CPF	VARCHAR(12)	CPF (anonimizado)
Escolaridade	VARCHAR(20)	Maior titulação
professorExterno	BIT	Indica se é professor externo

Fonte: Autor (2025)

Esta dimensão inclui identificação, qualificação e tipo de vínculo dos docentes.

d) Dimensão Datas (dw.Datas)

Tabela 9 – Estrutura da dimensão Datas

Esta dimensão temporal permite análises cronológicas dos dados.

Tabela 5 – Estrutura da dimensão Datas

Atributo	Tipo	Descrição
anoSemestre	VARCHAR(20)	Chave primária (AAAA.S)
Ano	INT	Ano letivo
Semestre	INT	Semestre (1 ou 2)

Fonte: Autor (2025)

5.3.2 Tabela Fato

A tabela fato armazena os eventos do negócio com as métricas associadas:

Tabela 10 – Estrutura da tabela fato Matriculas

Tabela 6 – Estrutura da tabela fato Matriculas

Atributo	Tipo	Descrição
id	INT	Chave primária surrogate
idAluno	INT	FK para dimensão Alunos
idDisciplina	INT	FK para dimensão Disciplinas
idProfessor	INT	FK para dimensão Professores
periodoLetivo	VARCHAR(15)	FK para dimensão Datas
mediaFinal	FLOAT	Nota final obtida (0-10)
situacaoFinal	VARCHAR(30)	Aprovado/Reprovado por nota/falta
Faltas	INT	Número de faltas

Fonte: Autor (2025)

Esta tabela contém chaves estrangeiras para as dimensões, métricas numéricas (média final, faltas) e resultados (situação final).

5.3.3 Relacionamentos

Os relacionamentos são estabelecidos através de chaves estrangeiras, permitindo consultas analíticas multidimensionais. No Power BI, foram configurados com cardinalidade Muitos-para-Um e direção de filtro cruzado bidirecional.

5.4 Desenvolvimento de Dashboards

Com base no modelo dimensional, foram desenvolvidos dashboards analíticos utilizando o Microsoft Power BI.

5.4.1 Dashboard Principal

O dashboard principal apresenta uma visão geral do desempenho acadêmico, incluindo:

- Indicadores gerais (média de nota, média de faltas, percentuais)



Figura 5 – Dashboard Principal

Fonte: Autor (2025)

- Gráficos de desempenho por período
- Gráficos históricos de indicadores
- Comparativos por forma de ingresso e sexo

5.4.2 Dashboard de Conclusão, Evasão e Retenção

Este dashboard foca na análise do fluxo de alunos, incluindo:

- Indicadores de conclusão, evasão e retenção
- Tempo médio de conclusão e tempo excedente
- Gráficos históricos de fluxo acadêmico
- Análises por perfil demográfico
- Distribuição dos alunos por faixa etária e sexo

5.4.3 Modelo no Power BI

O modelo implementado no Power BI reflete a estrutura dimensional do *Data Warehouse*, com adaptações específicas para facilitar a análise:

- Adição de uma tabela de Medidas para armazenar cálculos DAX

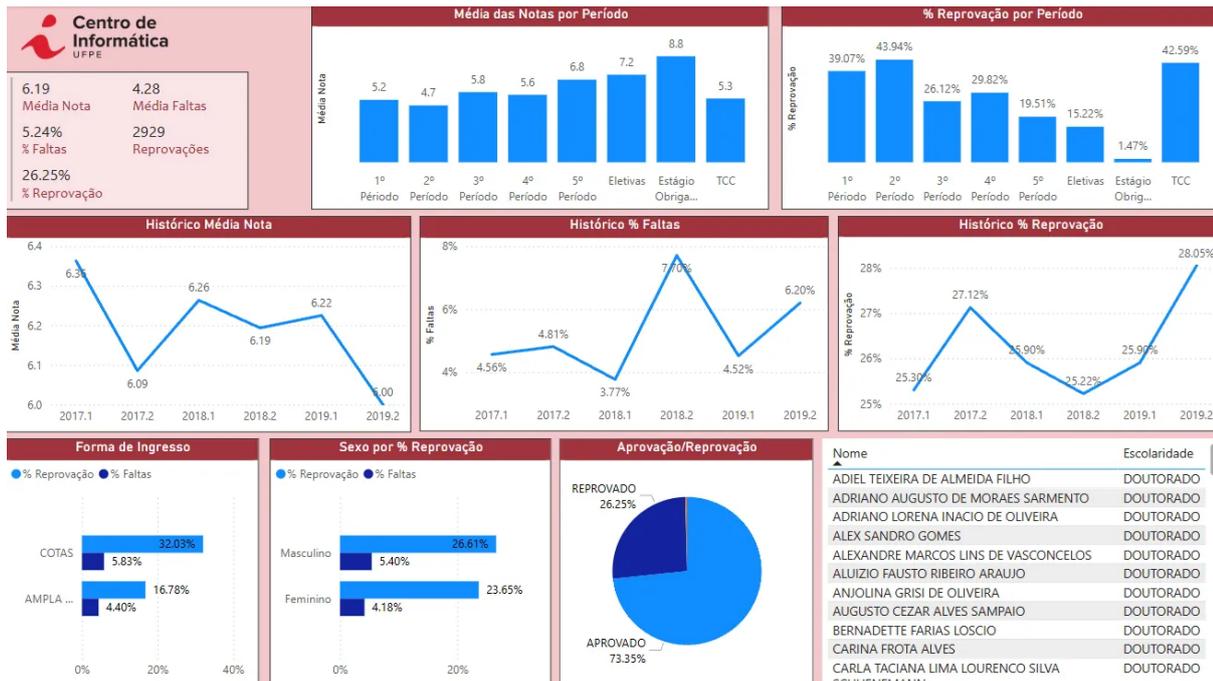


Figura 6 – Dashboard de Conclusão, Evasão e Retenção

Fonte: Autor (2025)

- Configuração de relações bidirecionais para propagação de filtros
- Hierarquias personalizadas para navegação em diferentes níveis

O modelo permite implementação de operações *OLAP* como *slice*, *dice*, *drill down*, *roll up*, *drill-through* e *pivot* através dos recursos interativos do *Power BI*.

5.5 Métricas e KPIs Definidos

Para suportar a análise e responder às questões exploratórias, foram definidas diversas métricas e KPIs implementados no *Power BI* utilizando a linguagem *DAX*.

5.5.1 Métricas de Desempenho Acadêmico

Tabela 11 – Métricas de desempenho implementadas

Tabela 7 – Métricas de desempenho implementadas

Métrica	Descrição e Fórmula DAX
Taxa de Reprovação	Percentual de matrículas reprovadas
Média de Notas	Média aritmética das notas finais
Percentual de Faltas	Média do percentual de faltas por matrícula
Taxa de Aprovação	Percentual de matrículas aprovadas

Fonte: Autor (2025)

Estas métricas permitem analisar diferentes aspectos do desempenho acadêmico, incluindo taxas de aprovação/reprovação, padrões de frequência e evolução das médias ao longo do tempo.

5.5.2 Métricas de Progresso Acadêmico

As métricas de progresso acadêmico fornecem insights sobre o fluxo dos alunos no curso:

Taxa de Evasão Total:

```
1 DIVIDE(  
2     CALCULATE(COUNT(Alunos[id]), Alunos[Situacao] IN {"Abandonado", "  
     ↳ Transferido", "Cancelado"}),  
3     COUNT(Alunos[id])  
4 )
```

Taxa de Conclusão:

```
1 DIVIDE(  
2     CALCULATE(COUNT(Alunos[id]), Alunos[Situacao] = "Concluido"),  
3     COUNT(Alunos[id])  
4 )
```

Tempo Médio de Conclusão (Anos):

```
1 AVERAGEX(  
2     FILTER(Alunos, Alunos[Situacao] = "Concluido"),  
3     DIVIDE(DATEDIFF(Alunos[Admissao], Alunos[conclusaoCurso], DAY), 365.25)  
4 )
```

5.5.3 KPIs Específicos para as Questões Exploratórias

Para cada questão exploratória, foram definidos KPIs específicos que permitem sua investigação:

- **Questão 1 (Forma de Ingresso):** Comparativos de desempenho por modalidade de acesso
- **Questão 2 (Sexo):** Análises comparativas por gênero
- **Questão 3 (Reprovação por Período):** Distribuição de reprovações ao longo do curso
- **Questão 4 (Evolução Histórica):** Tendências temporais dos indicadores

- **Questão 5 (Tempo de Conclusão):** Comparação com tempo previsto
- **Questão 6 (Perfil Demográfico):** Distribuição por sexo e faixa etária
- **Questão 7 (Evasão e Retenção):** Padrões de permanência por perfil
- **Questão 8 (Evolução Temporal):** Análise de séries históricas dos indicadores principais
- **Questão 9 (Disciplinas Críticas):** Ranking de dificuldade e distribuição curricular

Estes KPIs foram incorporados aos dashboards para análise interativa, fornecendo base empírica para as investigações realizadas.

6 Resultados e Discussão

Esta seção apresenta os resultados obtidos através da análise dos dados e discute as implicações desses resultados em relação às questões exploratórias formuladas. Os resultados são apresentados através das análises das visualizações e dashboards desenvolvidos no Power BI, investigando nove questões específicas sobre padrões de desempenho, características demográficas, evolução temporal e disciplinas críticas no curso.

6.1 Análise do Desempenho por Período do Curso

A primeira análise realizada foi a distribuição do desempenho acadêmico ao longo dos períodos do curso, visando verificar se existem padrões ou períodos que apresentem maiores desafios para os alunos.

6.1.1 Média de Notas por Período

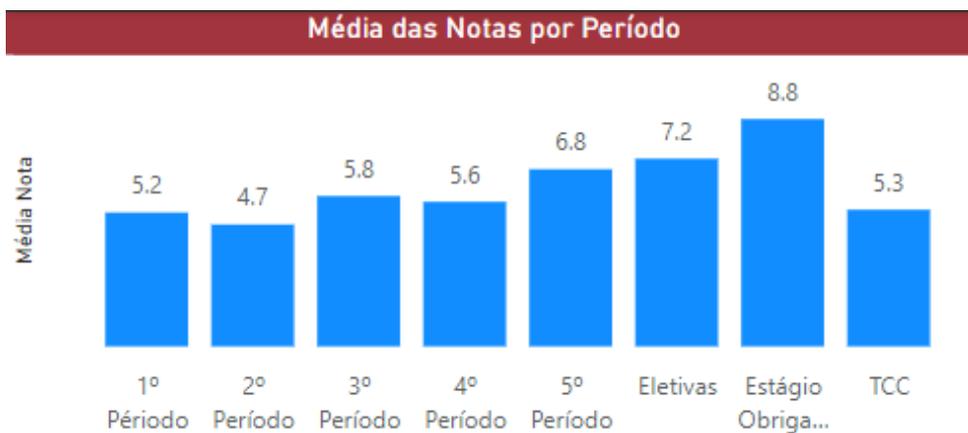


Figura 7 – Média de Notas por Período

Fonte: Autor (2025)

Observa-se que as médias das notas variam ao longo dos períodos, com um padrão ascendente à medida que os alunos avançam no curso. As médias mais baixas são observadas nos períodos iniciais, especialmente no 1º e 2º períodos (5,2 e 4,7, respectivamente), enquanto as médias mais altas ocorrem no estágio obrigatório (7,2) e no TCC (8,8).

Este padrão pode ser explicado por diversos fatores:

1. **Processo de adaptação:** Nos primeiros períodos, os alunos estão se adaptando ao ambiente universitário e às metodologias de ensino.

2. **Caráter das disciplinas iniciais:** As disciplinas dos primeiros períodos frequentemente estabelecem bases para o restante do curso.
3. **Seleção ao longo do curso:** À medida que os períodos avançam, permanece um grupo mais adaptado às exigências acadêmicas.
4. **Natureza das disciplinas avançadas:** As disciplinas de períodos mais avançados têm caráter mais prático e personalizado.

6.1.2 Percentual de Reprovação por Período

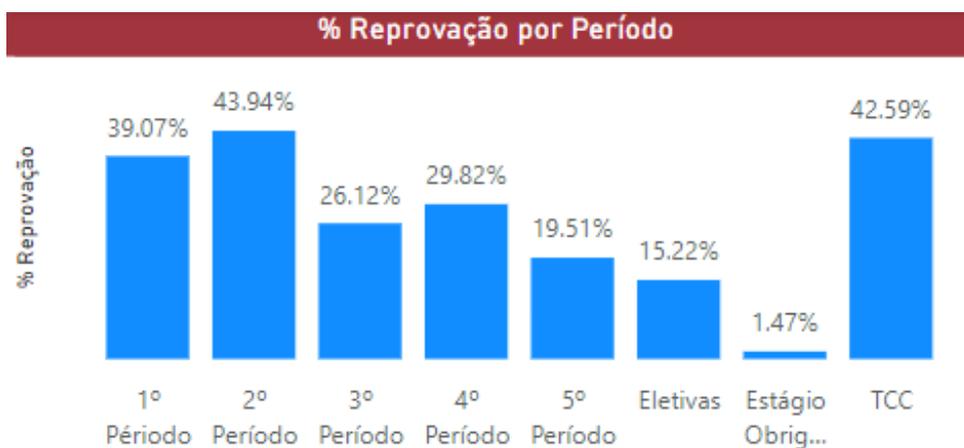


Figura 8 – Percentual de Reprovação por Período

Fonte: Autor (2025)

O gráfico de percentual de reprovação por período reforça e complementa a análise das médias. Observa-se um padrão decrescente na taxa de reprovação à medida que os alunos avançam no curso, com exceção do TCC.

Os primeiros períodos apresentam as maiores taxas de reprovação, com 39,07% no 1º período e 43,94% no 2º período. Estas taxas são significativamente superiores às dos períodos mais avançados, como o 5º período (15,22%) e o Estágio Obrigatório (1,47%).

A alta taxa de reprovação no TCC (42,59%) destoa do padrão geral. Possíveis explicações incluem:

1. **Natureza da atividade:** O TCC exige habilidades específicas de pesquisa e escrita acadêmica.
2. **Conciliação com outras atividades:** Muitos alunos já estão trabalhando quando chegam ao TCC.
3. **Questões de orientação:** Possíveis desafios na relação orientador-orientando.
4. **Critérios de avaliação:** Os critérios podem ser mais rigorosos que os das disciplinas regulares.

6.1.3 Evolução Histórica dos Indicadores

Para compreender como os indicadores têm evoluído ao longo do tempo, foram analisados dados históricos de diferentes métricas.



Figura 9 – Evolução Histórica - Média de Notas

Fonte: Autor (2025)



Figura 10 – Evolução Histórica - Percentual de Reprovação

Fonte: Autor (2025)

A análise da evolução histórica revela:

1. **Variabilidade temporal:** Os indicadores apresentam flutuações ao longo dos semestres, sem tendências uniformes.
2. **Correlação entre indicadores:** Existe correlação perceptível entre percentual de faltas, médias de notas e taxas de reprovação.

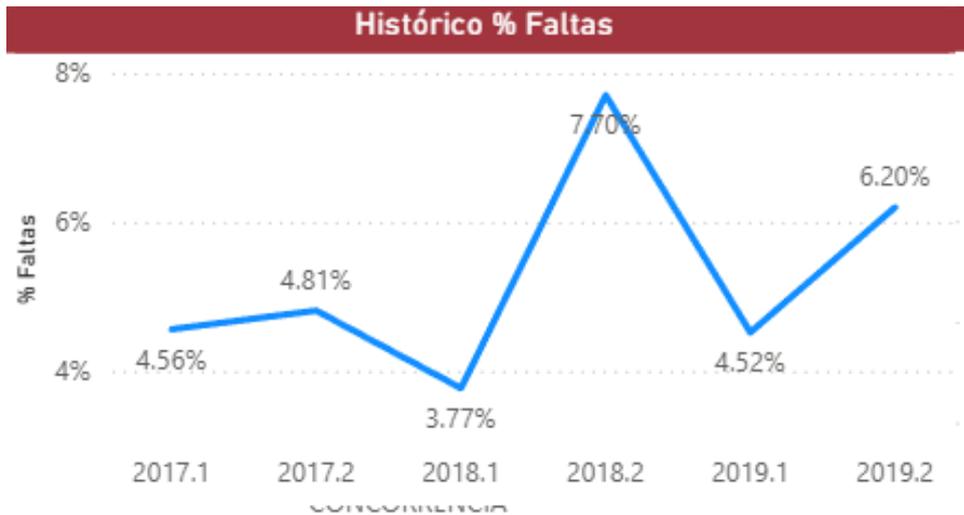


Figura 11 – Evolução Histórica - Percentual de Faltas

Fonte: Autor (2025)

3. **Períodos atípicos:** Alguns semestres apresentam comportamentos específicos que merecem investigação adicional.
4. **Instabilidade dos indicadores:** A variação observada sugere que múltiplos fatores podem estar afetando o desempenho ao longo do tempo.

6.2 Padrões de Desempenho por Forma de Ingresso

A investigação sobre possíveis diferenças no desempenho acadêmico conforme a forma de ingresso é particularmente relevante no contexto das políticas de ações afirmativas implementadas nas universidades federais brasileiras a partir da Lei nº 12.711/2012.

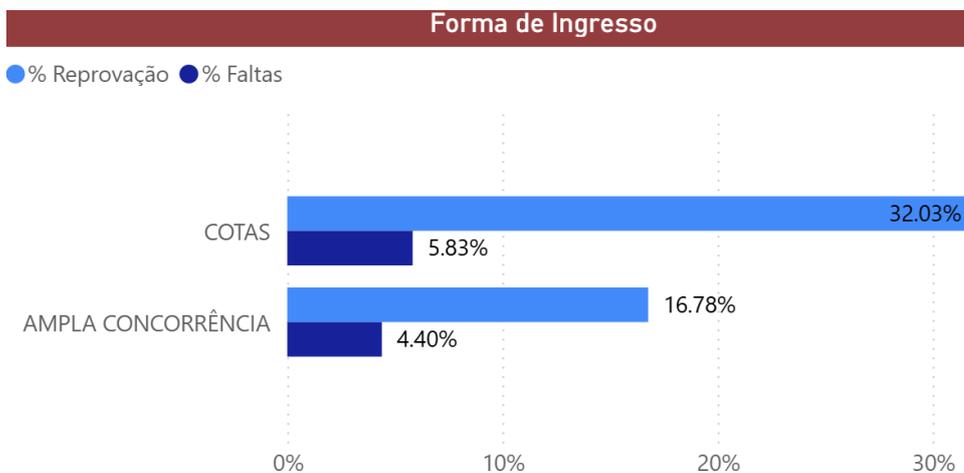


Figura 12 – Padrões de Desempenho por Forma de Ingresso

Fonte: Autor (2025)

A análise exploratória dos dados revela padrões diferenciados de desempenho entre alunos que ingressaram por ampla concorrência e por cotas. Os dados mostram que alunos cotistas apresentam taxa de reprovação de 32,03%, enquanto alunos de ampla concorrência apresentam 16,78%, representando uma diferença de 15,25 pontos percentuais que merece investigação detalhada.

Complementarmente, a análise do percentual de faltas também revela diferenças nos padrões de frequência: alunos de ampla concorrência apresentam 4,40% de faltas em média, enquanto alunos cotistas registram 5,83%, uma diferença de 1,43 pontos percentuais. Embora numericamente menor que a diferença nas taxas de reprovação, esta disparidade na frequência pode estar relacionada aos desafios específicos enfrentados por diferentes perfis de estudantes.

Estes achados sugerem que podem existir desafios diferenciados conforme a modalidade de ingresso, requerendo análise contextualizada que considere múltiplos fatores:

Limitações temporais dos dados: É fundamental destacar que os dados sobre modalidade de ingresso estão disponíveis de forma consistente apenas a partir de 2015, coincidindo com a implementação mais ampla do SISU e das políticas de ações afirmativas na UFPE. Isso significa que os padrões observados refletem um período específico de aproximadamente cinco anos, durante o qual tanto a instituição quanto os estudantes estiveram se adaptando às novas dinâmicas de acesso ao ensino superior. Esta limitação temporal requer cautela na generalização dos achados, mas não diminui sua relevância para compreender a realidade atual do curso.

Fatores socioeconômicos associados: A forma de ingresso frequentemente está associada a diferenças socioeconômicas que podem afetar diversos aspectos da experiência acadêmica. Estudantes cotistas, por exemplo, podem enfrentar necessidade de conciliação entre estudos e atividades laborais, têm menor probabilidade de ter cursado ensino médio em instituições privadas com preparação mais intensiva para o ensino superior, e podem ter menor acesso a recursos educacionais complementares como cursos preparatórios, materiais didáticos ou tecnologias educacionais.

Desafios de adaptação institucional: O ingresso no ensino superior representa uma transição significativa para todos os estudantes, mas pode apresentar complexidades específicas para aqueles provenientes de contextos educacionais diferenciados. Alunos cotistas podem enfrentar desafios adicionais de adaptação às metodologias de ensino, ao ambiente acadêmico e às expectativas de autonomia intelectual características do ensino universitário, especialmente em cursos com alta exigência técnica como Ciência da Computação.

Preparação acadêmica prévia: Diferenças na formação básica, particularmente em disciplinas fundamentais como matemática, física e habilidades de escrita acadêmica, podem impactar significativamente o desempenho inicial no curso. Disciplinas dos primeiros

períodos, como Cálculo e Programação Introdutória, exigem competências específicas que podem ter sido desenvolvidas de forma desigual durante o ensino médio.

Necessidades de apoio diferenciadas: Os padrões observados sugerem que diferentes perfis de estudantes podem se beneficiar de estratégias de apoio diferenciadas. Enquanto alguns estudantes podem necessitar principalmente de apoio acadêmico em conteúdos específicos, outros podem requerer suporte mais abrangente que inclua orientação sobre métodos de estudo, gestão do tempo, e integração ao ambiente universitário.

Importante ressaltar que estes resultados devem ser interpretados considerando o contexto mais amplo das políticas de inclusão no ensino superior. A presença de disparidades no desempenho inicial não questiona a validade ou importância das políticas de ações afirmativas, mas sim evidencia a necessidade de que estas sejam complementadas por políticas efetivas de permanência e sucesso acadêmico.

A literatura internacional sobre diversidade no ensino superior demonstra que diferenças iniciais no desempenho podem ser significativamente reduzidas através de intervenções apropriadas, incluindo programas de mentoria, grupos de estudo, suporte acadêmico direcionado e criação de ambientes inclusivos que valorizem a diversidade de experiências e perspectivas trazidas por diferentes grupos de estudantes.

Para a gestão acadêmica do curso, estes achados sugerem oportunidades concretas de intervenção:

1. **Programas de nivelamento:** Implementação de atividades preparatórias antes do início das aulas regulares, focadas em disciplinas críticas identificadas nas análises.
2. **Acompanhamento personalizado:** Desenvolvimento de sistemas de monitoramento que permitam identificação precoce de estudantes que apresentem sinais de dificuldade acadêmica.
3. **Suporte acadêmico direcionado:** Criação de programas de tutoria e monitoria que considerem as necessidades específicas de diferentes perfis de estudantes.
4. **Integração de políticas:** Articulação entre as políticas de acesso (cotas) e políticas de permanência (assistência estudantil, apoio acadêmico) para garantir que a democratização do acesso seja acompanhada por condições adequadas de conclusão.

Estes achados contribuem para uma compreensão mais nuançada dos desafios enfrentados por diferentes grupos de estudantes e podem orientar o desenvolvimento de estratégias institucionais mais efetivas para promover o sucesso acadêmico de todos os alunos, independentemente de sua forma de ingresso.

6.3 Padrões de Desempenho por Sexo

A investigação sobre possíveis diferenças de gênero no desempenho acadêmico é relevante considerando a disparidade numérica em cursos de computação.

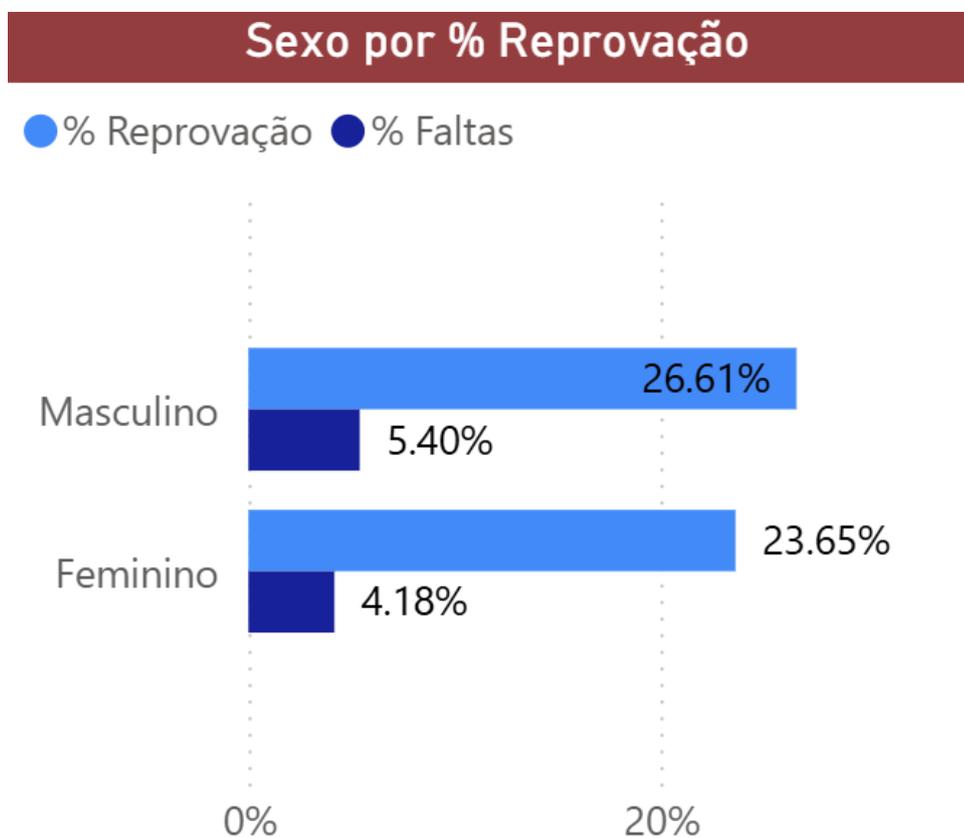


Figura 13 – Padrões de Desempenho por Sexo

Fonte: Autor (2025)

A análise revela diferenças nos padrões de desempenho entre alunos de diferentes sexos. Alunas apresentam taxa de reprovação de 23,65%, enquanto alunos do sexo masculino apresentam 26,61%. Quanto às faltas, alunas apresentam 4,18%, comparado a 5,40% dos alunos masculinos.

Estes resultados sugerem que:

1. **Desempenho feminino:** Contrariamente a possíveis expectativas, as alunas apresentam desempenho ligeiramente superior.
2. **Comprometimento acadêmico:** O menor percentual de faltas entre as alunas sugere maior assiduidade.
3. **Desmistificação de estereótipos:** Os resultados contrariam estereótipos sobre menor aptidão feminina para a área.

4. **Questões de acesso vs. permanência:** Embora apresentem bom desempenho, as mulheres continuam sendo minoria no curso.

6.4 Análise do Tempo de Conclusão

A investigação sobre o tempo necessário para conclusão do curso fornece insights fundamentais sobre a eficiência do fluxo curricular e os desafios enfrentados pelos estudantes em sua progressão acadêmica.



Figura 14 – Tempo Médio de Conclusão Histórico

Fonte: Autor (2025)

A análise dos dados revela que o tempo médio de conclusão do curso tem variado entre 5,25 e 6,24 anos ao longo do período analisado, com uma média geral de 5,75 anos. Este valor supera significativamente o tempo previsto na matriz curricular (4,5 anos), representando um acréscimo de aproximadamente 28% ou 1,25 anos além do previsto.

O tempo médio excedente, que representa a diferença entre o tempo real e o tempo previsto de conclusão, apresentou variações consideráveis ao longo da série histórica. Observa-se uma tendência de redução do tempo excedente entre 2017.1 (1,74 anos) e 2017.2 (0,85 anos), seguida de aumento nos períodos subsequentes, chegando a 1,46 anos em 2019.1 e reduzindo ligeiramente para 1,36 anos em 2019.2.

Fatores associados ao prolongamento do curso

Esta disparidade consistente entre tempo previsto e tempo real de conclusão pode estar associada a múltiplos fatores que merecem investigação detalhada:



Figura 15 – Tempo Médio Excedente

Fonte: Autor (2025)

1. **Impacto das reprovações:** Como observado nas análises anteriores, as taxas de reprovação são significativas, especialmente nos períodos iniciais do curso (39,07% no 1º período e 43,94% no 2º período). Cada reprovação em disciplina obrigatória geralmente implica em pelo menos um semestre adicional para nova tentativa, considerando que muitas disciplinas são oferecidas apenas uma vez por ano ou possuem pré-requisitos que impedem progressão imediata.
2. **Estrutura de pré-requisitos:** A matriz curricular de Ciência da Computação apresenta uma estrutura complexa de pré-requisitos, onde o insucesso em disciplinas fundamentais pode criar um efeito cascata, impedindo a matrícula em disciplinas subsequentes e forçando os alunos a reorganizar significativamente seus planos de estudo.
3. **Trancamentos e afastamentos:** Alunos podem optar por trancamento de matrícula ou de disciplinas específicas por motivos diversos, incluindo questões pessoais, financeiras, de saúde ou necessidade de conciliação com atividades profissionais. Estes afastamentos temporários contribuem diretamente para o prolongamento do tempo de formação.
4. **Transição para atividades profissionais:** Uma característica particular dos cursos de tecnologia é que muitos estudantes começam a trabalhar na área antes de concluir a graduação. Esta transição precoce para o mercado de trabalho, embora positiva do ponto de vista da empregabilidade, frequentemente resulta em redução da carga de disciplinas cursadas por semestre, prolongando consequentemente o tempo total de formação.

5. **Complexidade crescente das disciplinas:** À medida que o curso avança, as disciplinas tendem a aumentar em complexidade e exigir maior dedicação individual. Esta progressão natural pode tornar mais difícil para os estudantes manterem a carga completa de disciplinas prevista na matriz ideal.
6. **Desafios específicos do TCC:** Como identificado anteriormente, o Trabalho de Conclusão de Curso apresenta uma taxa de reprovação atipicamente alta (42,59%), o que pode contribuir significativamente para atrasos na conclusão, especialmente considerando que o TCC é tipicamente cursado nos últimos períodos.

Variabilidade temporal dos indicadores

A redução significativa do tempo excedente observada no semestre 2017.2 merece investigação específica, pois representa uma melhoria expressiva em relação ao semestre anterior. Esta variação pode estar associada a:

- Implementação de políticas específicas de acompanhamento acadêmico
- Mudanças nas regras de progressão ou jubramento
- Características específicas da coorte de alunos que concluiu naquele período
- Alterações nos critérios de avaliação ou na oferta de disciplinas
- Programas especiais de recuperação ou aceleração acadêmica

Implicações para a gestão acadêmica

O prolongamento consistente do tempo de conclusão não representa apenas um desafio individual para os estudantes, mas também um problema sistêmico com múltiplas implicações:

1. **Eficiência do sistema educacional:** Alunos que permanecem mais tempo no curso ocupam vagas que poderiam ser disponibilizadas para novos ingressantes, reduzindo a capacidade efetiva de formação de novos profissionais.
2. **Custos institucionais:** Cada semestre adicional representa custos operacionais para a instituição em termos de infraestrutura, recursos docentes, serviços administrativos e de apoio estudantil.
3. **Impacto nos indicadores institucionais:** O tempo médio de conclusão é um indicador importante utilizado em avaliações institucionais e pode afetar conceitos e classificações da instituição.

4. **Questões de sustentabilidade:** O prolongamento excessivo pode indicar problemas estruturais na concepção curricular ou na adequação entre objetivos formativos e realidade estudantil.

Estratégias de intervenção

Para abordar este desafio, algumas estratégias podem ser consideradas pela gestão acadêmica:

1. **Revisão curricular abrangente:** Avaliação crítica da estrutura curricular, carga horária, sequenciamento de disciplinas e adequação dos pré-requisitos à realidade dos estudantes.
2. **Flexibilização da progressão:** Criação de mecanismos que permitam maior flexibilidade na progressão curricular, como disciplinas optativas que possam ser cursadas em diferentes momentos ou equivalências que facilitem recuperação de atrasos.
3. **Acompanhamento acadêmico proativo:** Implementação de sistemas de monitoramento que identifiquem precocemente estudantes com risco de retenção e ofereçam suporte direcionado antes que os atrasos se acumulem.
4. **Otimização da oferta:** Revisão da frequência e horários de oferta de disciplinas críticas, especialmente aquelas com altas taxas de reprovação, para facilitar oportunidades de recuperação.
5. **Suporte específico para transições:** Desenvolvimento de programas que auxiliem estudantes na transição entre diferentes fases do curso, especialmente no TCC e em disciplinas de maior complexidade.
6. **Integração com o mercado de trabalho:** Criação de mecanismos que facilitem a conciliação entre estudos e atividades profissionais, reconhecendo que esta é uma realidade crescente para os estudantes da área.

6.5 Análise da Evasão e Retenção

A análise dos padrões de evasão e retenção é fundamental para compreender o fluxo de alunos no curso.

A análise revela diferenças significativas nos indicadores conforme a forma de ingresso. Alunos de ampla concorrência apresentam taxa de conclusão de 12,00%, comparada a 2,65% entre cotistas. Em contrapartida, a taxa de evasão é maior entre cotistas (23,18%) em comparação com ampla concorrência (18,00%).

Estas diferenças podem estar relacionadas a:

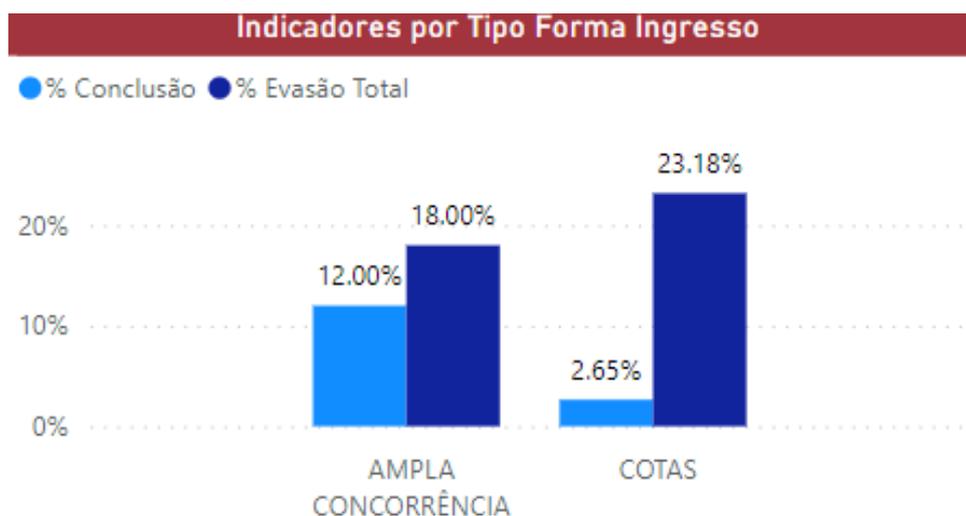


Figura 16 – Indicadores de Evasão e Retenção

Fonte: Autor (2025)

1. **Início recente das cotas:** O sistema de cotas é relativamente novo, e muitos cotistas ainda podem não ter tido tempo suficiente para concluir.
2. **Fatores socioeconômicos:** Alunos cotistas podem enfrentar desafios adicionais que afetam a permanência.
3. **Programas de permanência:** A eficácia de programas de assistência pode afetar significativamente os resultados.
4. **Necessidade de políticas integradas:** A simples implementação de cotas sem medidas complementares pode não ser suficiente.

6.6 Distribuição Demográfica dos Alunos

A caracterização do perfil demográfico dos estudantes é importante para compreender a diversidade do corpo discente.

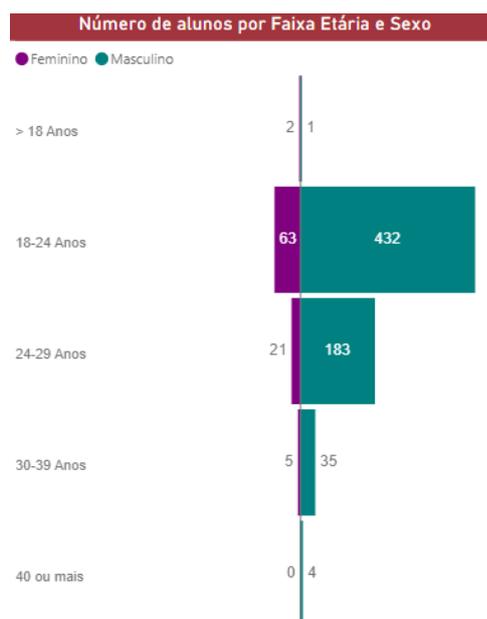


Figura 17 – Distribuição Demográfica - Faixa Etária e Sexo

Fonte: Autor (2025)

A análise revela disparidade significativa na distribuição por sexo, com 651 alunos do sexo masculino (87,4%) e 94 do sexo feminino (12,6%), uma proporção aproximada de 7:1.

Quanto à distribuição etária, observa-se concentração na faixa de 18-24 anos (66,4% do total), seguida pela faixa de 24-29 anos (27,4%).

Estes resultados indicam:

1. **Disparidade de gênero:** A predominância masculina é acentuada e consistente em todas as faixas etárias.
2. **Concentração etária:** O curso atrai principalmente alunos recém-egressos do ensino médio.
3. **Sub-representação feminina:** Representa uma perda de talentos potenciais para a área.
4. **Necessidade de políticas de inclusão:** A disparidade observada requer estratégias específicas de diversificação.

6.7 Síntese dos Resultados

Após a análise detalhada das diferentes questões exploratórias, esta seção apresenta uma síntese dos principais achados:

Quadro 1 – Síntese dos resultados das questões exploratórias

Tabela 8 – Síntese dos resultados das questões exploratórias

Questão	Descrição	Principais Achados
1	Diferenças por forma de ingresso	Observaram-se padrões distintos: cotistas apresentam taxa de reprovação de 32,03% vs. 16,78% de ampla concorrência
2	Diferenças por sexo	Alunas apresentam desempenho ligeiramente superior: 23,65% vs. 26,61% de reprovação
3	Distribuição de reprovações por período	Concentração nos períodos iniciais (39,07% no 1º e 43,94% no 2º) e no TCC (42,59%)
4	Evolução histórica	Variabilidade significativa ao longo do tempo (25,22% a 28,05%), sem padrão estável
5	Tempo de conclusão	Tempo médio de 5,75 anos, 28% acima do previsto (4,5 anos)
6	Perfil demográfico	Forte predominância masculina (87,4% vs. 12,6%) e concentração etária em 18-24 anos
7	Evasão e retenção	Taxas de evasão maiores entre cotistas (23,18% vs. 18,00%)

Fonte: Autor (2025)

Descobertas adicionais relevantes:

1. **Gargalos identificados:** Os períodos iniciais e o TCC representam desafios significativos no fluxo acadêmico.
2. **Padrões diferenciados:** Existem diferenças consistentes no desempenho conforme características dos alunos.
3. **Potencial feminino:** Apesar da sub-representação, as alunas demonstram bom desempenho acadêmico.
4. **Variabilidade temporal:** Os indicadores apresentam flutuações que sugerem influência de múltiplos fatores.
5. **Desafios na conclusão:** As taxas de conclusão são relativamente baixas, especialmente entre certos grupos.

Com base nestes achados, algumas considerações podem ser formuladas:

1. **Atenção aos períodos iniciais:** Implementar programas de acompanhamento focados nas disciplinas dos primeiros períodos.

2. **Suporte diferenciado:** Desenvolver programas específicos de apoio para diferentes perfis de alunos.
3. **Promoção da diversidade:** Implementar estratégias para aumentar a participação feminina.
4. **Revisão do TCC:** Analisar fatores que contribuem para a alta taxa de dificuldades no TCC.
5. **Monitoramento contínuo:** Implementar sistemas que permitam identificar precocemente padrões de risco.

7 Considerações Finais

7.1 Conclusões

Este trabalho teve como objetivo desenvolver uma solução de *Business Intelligence* para análise exploratória do desempenho acadêmico dos alunos de graduação em Ciência da Computação da UFPE, visando fornecer insights para a gestão acadêmica e contribuir para estratégias de melhoria do ensino e redução da evasão.

O desenvolvimento do projeto envolveu a implementação de um *Data Warehouse* seguindo o modelo dimensional (*Star Schema*), a realização de um processo ETL estruturado e a criação de dashboards analíticos utilizando o Power BI. A investigação foi conduzida através de nove questões exploratórias específicas, abrangendo desde características demográficas e padrões temporais até análises de disciplinas críticas e fatores associados à permanência no curso. Com base nos resultados obtidos, algumas conclusões importantes podem ser destacadas:

1. **Perfil demográfico:** Observou-se significativa disparidade de gênero no curso, com aproximadamente 87,4% dos alunos sendo do sexo masculino. Esta disparidade reflete um padrão comum em cursos de computação e sugere a necessidade de políticas para aumentar a participação feminina.
2. **Padrões por forma de ingresso:** Foram identificadas diferenças nos padrões de desempenho entre alunos que ingressaram por ampla concorrência e por cotas. Alunos cotistas apresentaram taxas de reprovação mais elevadas (32,03% vs. 16,78%), sugerindo a importância de programas de apoio acadêmico direcionados.
3. **Desempenho por gênero:** As alunas apresentaram desempenho ligeiramente superior aos alunos, com menores taxas de reprovação (23,65% vs. 26,61%) e menor percentual de faltas (4,18% vs. 5,40%). Este resultado contraria estereótipos e destaca o potencial feminino na área.
4. **Períodos críticos:** A análise identificou concentração de reprovações nos períodos iniciais do curso (39,07% no 1º período e 43,94% no 2º período) e no TCC (42,59%). Estes pontos representam gargalos significativos no fluxo acadêmico.
5. **Tempo de conclusão:** O tempo médio de conclusão (5,75 anos) supera significativamente o previsto na matriz curricular (4,5 anos), representando um acréscimo de aproximadamente 28%. Este prolongamento sugere desafios estruturais na progressão dos alunos.

6. **Evasão e conclusão:** Observaram-se diferenças significativas nas taxas de conclusão entre alunos de ampla concorrência (12,00%) e cotistas (2,65%), bem como nas taxas de evasão (18,00% vs. 23,18%). Estes resultados indicam a necessidade de políticas específicas de permanência.
7. **Variabilidade temporal:** Os indicadores de desempenho apresentaram flutuações significativas ao longo do período analisado, sugerindo a influência de fatores conjunturais ou mudanças institucionais.

A solução desenvolvida demonstrou o potencial das técnicas de *Business Intelligence* na análise de dados educacionais, oferecendo uma visão abrangente do desempenho acadêmico. Os dashboards criados permitem aos gestores identificar padrões e tendências que podem fundamentar a tomada de decisões baseada em evidências.

7.2 Contribuições

As principais contribuições deste trabalho incluem:

1. **Modelo dimensional educacional:** Desenvolvimento de um modelo dimensional (*Star Schema*) específico para análise de dados acadêmicos, incluindo dimensões de alunos, disciplinas, professores e tempo, que pode ser adaptado para outros contextos educacionais.
2. **Processo ETL documentado:** Implementação e documentação de um processo ETL estruturado em camadas para dados educacionais, oferecendo uma abordagem clara e manutenível para o processamento de dados acadêmicos.
3. **Dashboards analíticos:** Criação de dashboards interativos que permitem explorar dados de diferentes perspectivas utilizando operações OLAP. Estes dashboards podem ser utilizados diretamente pela gestão acadêmica.
4. **Métricas e KPIs educacionais:** Definição e implementação de um conjunto abrangente de métricas específicas para o contexto educacional, que podem servir como referência para outros estudos.
5. **Insights sobre desempenho acadêmico:** Identificação de padrões associados ao desempenho, retenção e evasão no curso de Ciência da Computação da UFPE através da investigação de nove questões exploratórias específicas, fornecendo conhecimentos detalhados sobre a dinâmica do curso, evolução temporal dos indicadores e disciplinas críticas.

6. **Recomendações baseadas em evidências:** Formulação de recomendações específicas para a gestão acadêmica, fundamentadas em evidências concretas obtidas através da análise de dados.
7. **Framework metodológico:** Desenvolvimento de uma abordagem metodológica para análise exploratória de desempenho acadêmico utilizando técnicas de BI, estruturada em nove questões investigativas que podem ser adaptadas e aplicadas em outros cursos e instituições.

O trabalho também contribui para a literatura sobre analítica educacional, demonstrando a aplicabilidade das técnicas de *Business Intelligence* no contexto do ensino superior brasileiro.

7.3 Limitações

O estudo apresenta algumas limitações que devem ser consideradas na interpretação dos resultados e em sua aplicação prática:

1. **Natureza exploratória e observacional:** Este estudo caracteriza-se como uma análise exploratória descritiva, identificando padrões e associações nos dados disponíveis, mas sem estabelecer relações causais definitivas. Os resultados devem ser interpretados como indicativos de possíveis associações entre variáveis, não como comprovações de causalidade. Esta limitação é inerente ao desenho do estudo e ao tipo de dados disponíveis, que são observacionais e não experimentais.
2. **Qualidade e completude dos dados secundários:** A utilização de dados administrativos existentes, embora ofereça vantagens como abrangência temporal e volume de informações, também apresenta limitações relacionadas à qualidade e completude. Algumas inconsistências e dados faltantes nas bases originais podem ter afetado a precisão de certas análises. Por exemplo, variações na forma de registro de informações ao longo do tempo, mudanças nos sistemas de informação da instituição, e possíveis erros de digitação ou codificação podem introduzir ruídos nos dados que são difíceis de detectar e corrigir completamente.
3. **Limitações temporais específicas:** Os dados sobre forma de ingresso (ampla concorrência ou cotas) só estão disponíveis de forma consistente a partir de 2015, coincidindo com a implementação mais ampla das políticas de ações afirmativas e do SISU na UFPE. Esta limitação temporal restringe algumas análises a um período mais recente e pode não capturar completamente os efeitos de longo prazo das políticas de inclusão. Adicionalmente, mudanças na matriz curricular, políticas

institucionais e contexto socioeconômico ao longo do período podem confundir as análises temporais.

4. **Variáveis não observadas:** Diversos fatores que reconhecidamente podem afetar o desempenho acadêmico não estavam disponíveis nas bases de dados institucionais, limitando a capacidade de análises mais abrangentes. Entre estas variáveis não observadas incluem-se: condições socioeconômicas detalhadas das famílias, qualidade da escola de origem no ensino médio, participação em atividades extracurriculares, situação de emprego dos estudantes, condições de moradia e transporte, histórico de saúde mental, motivação intrínseca para o curso, expectativas familiares, e rede de apoio social disponível.
5. **Complexidade das relações entre variáveis:** As análises realizadas, embora multidimensionais, podem não capturar completamente a complexidade das interações entre diferentes fatores que influenciam o desempenho acadêmico. Por exemplo, a associação observada entre forma de ingresso e desempenho pode estar mediada por múltiplas variáveis intermediárias como preparação acadêmica prévia, condições socioeconômicas, estratégias de estudo, e acesso a recursos educacionais. A análise isolada de fatores individuais pode não revelar essas relações mais complexas.
6. **Limitações do período de análise:** Embora o estudo tenha considerado um período significativo (2009-2019), mudanças institucionais, sociais e tecnológicas recentes podem ter alterado alguns dos padrões observados. Particularmente, os impactos da pandemia de COVID-19, que começou em 2020, introduziram mudanças substanciais no ensino superior que não são capturadas nesta análise. Adicionalmente, reformas curriculares, mudanças nas políticas de assistência estudantil, e evolução das tecnologias educacionais podem ter modificado a dinâmica do curso.
7. **Especificidade contextual:** Os resultados refletem a realidade específica do curso de Ciência da Computação da UFPE, uma instituição pública federal de alta qualidade em uma região metropolitana específica. A generalização destes resultados para outros contextos deve ser feita com cautela, considerando diferenças institucionais (público vs. privado, diferentes regiões, porte da instituição), curriculares (diferentes estruturas de curso, metodologias de ensino), e socioeconômicas (características regionais, perfil dos estudantes, mercado de trabalho local).
8. **Limitações metodológicas das ferramentas utilizadas:** Algumas análises mais avançadas, como modelagem preditiva sofisticada, análise de séries temporais com controle de tendências, ou técnicas de aprendizado de máquina para identificação de padrões complexos, não foram implementadas devido às limitações das ferramentas utilizadas no estudo. Embora o Power BI seja uma ferramenta robusta para análises

exploratórias e visualização, sua capacidade para análises estatísticas avançadas é limitada comparada a softwares especializados.

9. **Ausência de dados qualitativos:** O estudo focou exclusivamente em análises quantitativas, não incorporando dados qualitativos que poderiam enriquecer significativamente a compreensão dos fatores subjacentes ao desempenho, evasão e retenção. Entrevistas com estudantes, grupos focais, questionários de percepção, e observações etnográficas poderiam fornecer insights valiosos sobre motivações, experiências, barreiras percebidas, e estratégias de superação de dificuldades que não são capturados nos dados administrativos.
10. **Questões de representatividade temporal:** Algumas análises, particularmente aquelas relacionadas à conclusão do curso, podem estar enviesadas temporalmente. Estudantes que ingressaram mais recentemente no período analisado tiveram menos tempo para concluir o curso, o que pode afetar as taxas de conclusão observadas. Esta limitação é especialmente relevante para análises que comparam diferentes coortes de ingressantes.
11. **Limitações na interpretação das diferenças observadas:** As diferenças nos padrões de desempenho entre grupos (por exemplo, entre cotistas e não-cotistas, ou entre homens e mulheres) devem ser interpretadas com cuidado para evitar conclusões simplistas sobre capacidades ou potencial dos estudantes. Estas diferenças refletem desigualdades estruturais, diferenças de oportunidades, e contextos socioeducacionais diversos, não características inerentes dos grupos analisados.

Reconhecer estas limitações é fundamental para uma interpretação adequada dos resultados e para o planejamento de pesquisas futuras que possam abordar algumas destas questões. Apesar destas limitações, o estudo fornece contribuições valiosas para a compreensão dos padrões de desempenho acadêmico no curso analisado e oferece uma base sólida para decisões de gestão acadêmica baseadas em evidências.

7.4 Trabalhos Futuros

Com base nas contribuições e limitações identificadas, sugerem-se os seguintes trabalhos futuros que podem ampliar e aprofundar os achados desta pesquisa:

1. **Integração com outros sistemas institucionais:** Expandir o *Data Warehouse* para integrar dados de outros sistemas institucionais, como bibliotecas (padrões de uso e empréstimos), programas de assistência estudantil (bolsas, auxílios, acompanhamento social), atividades extracurriculares (participação em projetos de extensão,

iniciação científica, grupos de estudo), e plataformas de aprendizagem virtual (interações, tempo de acesso, desempenho em atividades online). Esta integração proporcionaria uma visão mais abrangente da experiência acadêmica e permitiria identificar fatores adicionais associados ao sucesso estudantil.

2. **Desenvolvimento de modelos preditivos:** Utilizar técnicas de aprendizado de máquina e análise estatística avançada para desenvolver modelos preditivos que possam identificar precocemente estudantes com padrões que sugiram risco de evasão. Algoritmos como árvores de decisão, random forests, regressão logística, redes neurais e máquinas de vetores de suporte poderiam ser aplicados aos dados históricos para criar sistemas de alerta precoce, permitindo intervenções preventivas direcionadas antes que as dificuldades se acumulem.
3. **Análise de redes sociais acadêmicas:** Investigar como as interações sociais e acadêmicas entre estudantes (formação de grupos de estudo, participação em projetos colaborativos, redes de apoio mútuo) podem estar associadas ao desempenho individual e coletivo. Esta análise poderia revelar a importância do capital social acadêmico e orientar estratégias de integração estudantil que promovam redes de apoio mais efetivas.
4. **Expansão para análises comparativas inter-institucionais:** Adaptar a metodologia desenvolvida para outros cursos da UFPE e estabelecer parcerias com outras instituições para realizar estudos comparativos. Esta expansão permitiria identificar padrões mais amplos, distinguir fatores específicos do curso de Ciência da Computação daqueles comuns a outros cursos de tecnologia, e identificar boas práticas institucionais que poderiam ser replicadas.
5. **Estudos longitudinais de acompanhamento de egressos:** Desenvolver mecanismos para acompanhar a trajetória profissional dos egressos e investigar possíveis associações entre diferentes aspectos do desempenho durante a graduação (tempo de conclusão, disciplinas cursadas, atividades extracurriculares) e o sucesso profissional posterior (inserção no mercado de trabalho, progressão de carreira, satisfação profissional, continuidade em pós-graduação).
6. **Avaliação sistemática de intervenções pedagógicas:** Utilizar a infraestrutura analítica desenvolvida para avaliar rigorosamente o impacto de intervenções pedagógicas ou curriculares sobre o desempenho dos alunos. Implementar desenhos quasi-experimentais ou experimentais para testar a eficácia de diferentes estratégias como programas de mentoria, metodologias ativas de ensino, sistemas de avaliação diferenciados, ou reformas curriculares, fornecendo evidências robustas sobre sua efetividade.

7. **Análise aprofundada da arquitetura curricular:** Realizar uma investigação detalhada da estrutura curricular utilizando técnicas de análise de redes para mapear as relações entre disciplinas, identificar caminhos críticos de progressão, detectar gargalos estruturais, e avaliar o impacto de diferentes sequências de disciplinas sobre o tempo de conclusão e desempenho estudantil.
8. **Pesquisas mistas com componente qualitativo:** Implementar estudos que combinem as análises quantitativas desenvolvidas com pesquisas qualitativas abrangentes, incluindo entrevistas em profundidade com estudantes de diferentes perfis, grupos focais com docentes e gestores, estudos etnográficos sobre a cultura acadêmica do curso, e questionários estruturados sobre percepções, motivações e barreiras enfrentadas pelos estudantes.
9. **Análise de impactos de eventos disruptivos:** Incorporar dados mais recentes para avaliar como eventos disruptivos significativos, como a pandemia de COVID-19, afetaram os padrões de desempenho, evasão e retenção identificados no estudo. Esta análise poderia fornecer insights sobre a resiliência do sistema educacional e a adaptabilidade de diferentes perfis de estudantes a mudanças drásticas no ambiente de aprendizagem.
10. **Desenvolvimento de plataforma analítica integrada:** Evoluir os dashboards desenvolvidos para uma plataforma analítica mais sofisticada e integrada, com diferentes níveis de acesso para gestores, docentes, estudantes e pesquisadores. Esta plataforma poderia incluir funcionalidades de análise self-service, alertas automáticos, simulação de cenários, e ferramentas de apoio à decisão que democratizem o acesso à informação e promovam uma cultura de gestão baseada em dados.
11. **Investigação de fatores mediadores e moderadores:** Utilizar técnicas estatísticas mais avançadas para investigar os mecanismos através dos quais diferentes fatores (como forma de ingresso, características demográficas, desempenho em disciplinas específicas) influenciam os resultados acadêmicos. Análises de mediação e moderação poderiam revelar variáveis intermediárias importantes e condições sob as quais diferentes fatores são mais ou menos influentes.
12. **Estudos de transferibilidade e adaptação metodológica:** Investigar como a metodologia e os achados deste estudo podem ser adaptados e transferidos para diferentes contextos educacionais, incluindo outros cursos de graduação, programas de pós-graduação, modalidades de ensino a distância, e instituições com características diversas. Desenvolver diretrizes para implementação de soluções similares em outros contextos educacionais.

A implementação gradual destes trabalhos futuros permitirá construir um programa de pesquisa abrangente em analítica educacional, contribuindo tanto para o avanço do

conhecimento científico quanto para a melhoria prática da gestão e qualidade do ensino superior. A continuidade destes estudos é essencial para que as contribuições desta pesquisa se traduzam em impactos duradouros na educação superior.

Referências

- ALVES, R. D. et al. Fatores associados ao desempenho em disciplinas introdutórias de programação em cursos de computação. In: *Workshop sobre Educação em Computação (WEI)*. Natal: Sociedade Brasileira de Computação, 2018. v. 26, p. 342–353.
- BARBOSA, G. C. et al. Estudo sobre fatores de evasão em cursos de computação do brasil. In: *Congresso Brasileiro de Informática na Educação*. Recife: Sociedade Brasileira de Computação, 2017. v. 6, p. 23–34.
- CHAUDHURI, S.; DAYAL, U.; NARASAYYA, V. An overview of business intelligence technology. *Communications of the ACM*, ACM, New York, v. 54, n. 8, p. 88–98, 2011.
- CHEN, H.; CHIANG, R. H. L.; STOREY, V. C. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, Minneapolis, v. 36, n. 4, p. 1165–1188, 2012.
- CODD, E. F.; CODD, S. B.; SALLEY, C. T. *Providing OLAP (On-line Analytical Processing) to User-Analysts: An IT Mandate*. [S.l.], 1993.
- COSTA, S. L.; GOUVEIA, L. B. Políticas de ação afirmativa e desempenho acadêmico: evidências do impacto da lei de cotas. *Revista Educação e Políticas em Debate*, Uberlândia, v. 7, n. 2, p. 205–225, 2018.
- DIAS, E. C. M.; THEÓPHILO, C. R.; LOPES, M. A. S. Evasão no ensino superior: estudo dos fatores causadores da evasão no curso de ciências contábeis da universidade estadual de montes claros. In: *Congresso USP de Iniciação Científica em Contabilidade*. São Paulo: USP, 2010. v. 7.
- FILHO, R. L. L. S. et al. A evasão no ensino superior brasileiro. *Cadernos de Pesquisa*, São Paulo, v. 37, n. 132, p. 641–659, 2007.
- GUSTER, D.; BROWN, C. G. The application of business intelligence to higher education: Technical and managerial perspectives. *Journal of Information Technology Management*, Richmond, v. 23, n. 2, p. 42–62, 2012.
- IAM-ON, N.; BOONGOEN, T. Improved student dropout prediction in thai university using ensemble of mixed-type data clusterings. *International Journal of Machine Learning and Cybernetics*, Springer, Berlin, v. 8, n. 2, p. 497–510, 2017.
- INEP. *Indicadores Educacionais*. Brasília, 2017.
- INMON, W. H. *Building the Data Warehouse*. 4. ed. Indianapolis: Wiley Publishing, 2005.
- KIMBALL, R.; ROSS, M. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3. ed. Indianapolis: Wiley Publishing, 2013.
- KORI, K. et al. Factors that influence students' motivation to start and to continue studying information technology in estonia. *IEEE Transactions on Education*, New York, v. 61, n. 4, p. 289–297, 2018.
- LIMA, E.; ZAGO, N. Evasão no ensino superior: tendências e resultados de pesquisa. *Revista Movimento*, Porto Alegre, v. 24, n. 1, p. 31–46, 2018.

- LOBO, M. B. C. M. Panorama da evasão no ensino superior brasileiro: aspectos gerais das causas e soluções. *ABMES Cadernos*, Brasília, v. 25, p. 1–23, 2012. Acesso em: 02 maio 2024. Disponível em: <<https://abmes.org.br/arquivos/publicacoes/Cadernos25.pdf>>.
- MANHÃES, L. M. B. et al. Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. In: *Simpósio Brasileiro de Informática na Educação*. Aracaju: Sociedade Brasileira de Computação, 2011. v. 22, p. 150–159.
- MOSCOSO-ZEA, O. et al. A business intelligence framework for academic data analytics: A case study in higher education. In: *International Conference on Information Technology & Systems*. Quito: Springer, 2019. p. 516–525.
- ONG, V. K. et al. Data mining in a business intelligence framework: An educational institution context. In: *IEEE International Conference on E-Business Engineering*. Beijing: IEEE, 2011. p. 29–36.
- POLIDORI, M. M.; MARINHO-ARAÚJO, C. M.; BARREYRO, G. B. Sinaes: perspectivas e desafios na avaliação da educação superior brasileira. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 14, n. 53, p. 425–436, 2006.
- SILVA, C. S. et al. Business intelligence em instituições de ensino superior: Um sistema de apoio à decisão para acompanhamento da evasão e desempenho acadêmico. In: *Simpósio Brasileiro de Sistemas de Informação*. Florianópolis: Sociedade Brasileira de Computação, 2016. v. 12, p. 72–83.
- TINTO, V. Research and practice of student retention: what next? *Journal of College Student Retention: Research, Theory & Practice*, Thousand Oaks, v. 8, n. 1, p. 1–19, 2006.
- TURBAN, E. et al. *Business Intelligence: A Managerial Approach*. Upper Saddle River: Prentice Hall, 2008.
- VASCONCELOS, A. L. F. S.; ALMEIDA, R. S.; MONTEIRO, J. M. Fatores que interferem nas taxas de conclusão dos alunos de graduação: um estudo no centro de informática da ufpe. *Revista Gestão Universitária na América Latina*, Florianópolis, v. 12, n. 2, p. 69–89, 2019.

Apêndices

APÊNDICE A – Scripts SQL

Este apêndice apresenta os principais scripts SQL desenvolvidos para o processo ETL e criação do *Data Warehouse*.

A.1 Script para criação dos schemas

```
1 create schema camada1;
2 go
3 create schema camada2;
4 go
5 create schema camada3;
6 go
7 create schema dw;
```

A.2 Script para criação da tabela de arquivos

```
1 create table Arquivos (
2     id int identity(1,1),
3     nomeArquivo varchar(255),
4     nomeCurto varchar(50)
5 )
```

A.3 Scripts para procedimentos de carga de dados

Código A.1 – Procedimento para carga de datas

```
1 CREATE PROCEDURE cargaDatas
2 AS
3 --Definindo datas
4 DECLARE @inicio VARCHAR(30) = (
5     SELECT min(d.periodoLetivo)
6     FROM dw.Disciplinas d
7 )
8 DECLARE @fim VARCHAR(30) = (
9     SELECT max(d.periodoLetivo)
10    FROM dw.Disciplinas d
11 )
12
13 DELETE
14 FROM dw.Datas
15
16 EXEC pr_geraDatas @inicio,@fim
```

Código A.2 – Procedimento para gerar datas

```

1 CREATE PROCEDURE pr_geraDatas @inicio VARCHAR(50), @fim VARCHAR(50)
2 AS
3 SET @inicio = convert(INT, substring(@inicio, 1, 4))
4 SET @fim = convert(INT, substring(@fim, 1, 4))
5
6 WHILE (@inicio <> @fim + 1)
7 BEGIN
8     INSERT INTO dw.Datas
9     VALUES (CONCAT (@inicio, '.1'), @inicio, '1')
10
11     INSERT INTO dw.Datas
12     VALUES (CONCAT (@inicio, '.2'), @inicio, '2')
13
14     SET @inicio = @inicio + 1
15 END

```

Código A.3 – Procedimento para inserir planilha

```

1 create procedure dbo.pr_inserePlanilha @tabelaDestino nvarchar(255),
2     ↪ @OrigemArquivo nvarchar(2500) as
3
4 declare @sql nvarchar(2500)
5 declare @before nvarchar(2500)
6
7 set @before = 'truncate table ' + @tabelaDestino
8
9 exec sp_executesql @statement = @before
10
11 set @sql = 'insert into ' + @tabelaDestino + ' select * FROM OPENROWSET(''
12     ↪ Microsoft.ACE.OLEDB.12.0'',
13     ''Excel 12.0;imex=1;hdr=no; Database='+@OrigemArquivo+'','','[Exportar
14     ↪ Planilha$]''')'
15
16 exec sp_executesql @statement = @sql

```

Código A.4 – Procedimento para nomes de arquivos

```

1 create procedure dbo.pr_nomeArquivos @diretorio nvarchar(255) as
2
3 declare @sql nvarchar(255)
4 truncate table Arquivos
5
6 set @sql =
7 'insert into Arquivos execute xp_cmdshell ''dir ' + @diretorio + ' /b''
8
9 delete from Arquivos where nomeArquivo is null
10 '
11
12 exec(@sql)

```

A.4 Scripts para criação das tabelas da Camada 1

Código A.5 – Tabela Alunos Camada 1

```
1 CREATE TABLE camada1.Alunos (  
2     NOME NVARCHAR(255) NULL,  
3     ADMISSAO NVARCHAR(255) NULL,  
4     PREVISAO_CONCLUSAO NVARCHAR(255) NULL,  
5     CONCLUSAO_CURSO NVARCHAR(255) NULL,  
6     FORMA_INGRESSO NVARCHAR(255) NULL,  
7     TIPO_FORMA_INGRESSO NVARCHAR(255) NULL,  
8     SITUACAO NVARCHAR(255) NULL,  
9     CURSO NVARCHAR(255) NULL,  
10    PERFIL_CURSO NVARCHAR(255) NULL,  
11    DT_NASCIMENTO NVARCHAR(255) NULL,  
12    ESTADO_CIVIL NVARCHAR(255) NULL,  
13    SEXO NVARCHAR(255) NULL,  
14    CONCLUSAO_ENS_MEDIO NVARCHAR(255) NULL,  
15    MEDIA_GERAL NVARCHAR(255) NULL,  
16    COEFICIENTE_RENDIMENTO NVARCHAR(255) NULL,  
17    FAIXA_ECONOMICA NVARCHAR(255) NULL,  
18    BLOCADO NVARCHAR(255) NULL  
19 )
```

Código A.6 – Tabela Disciplinas Camada 1

```
1 CREATE TABLE camada1.Disciplinas(  
2     PERIODO_LETIVO nvarchar(255) NULL,  
3     CURSO nvarchar(255) NULL,  
4     ID_DISCIPLINA nvarchar(255) NULL,  
5     DISCIPLINA nvarchar(255) NULL,  
6     CARGA_HORARIA nvarchar(255) NULL,  
7     QUANTIDADE_AULAS nvarchar(255) NULL  
8 )
```

Código A.7 – Tabela Matriculas Camada 1

```
1 CREATE TABLE camada1.Matriculas(  
2     NOME nvarchar(255) NULL,  
3     CPF nvarchar(255) NULL,  
4     MATRICULA nvarchar(255) NULL,  
5     ADMISSAO nvarchar(255) NULL,  
6     CURSO nvarchar(255) NULL,  
7     PERIODO_LETIVO nvarchar(255) NULL,  
8     ID_DISCIPLINA nvarchar(255) NULL,  
9     DISCIPLINA nvarchar(255) NULL,  
10    MEDIA_FINAL nvarchar(2550) NULL,  
11    SITUACAO_FINAL nvarchar(255) NULL,  
12    FALTAS nvarchar(255) NULL,  
13    CPF_PROFESSOR nvarchar(255) NULL,  
14    NOME_PROFESSOR nvarchar(255) NULL  
15 )
```

Código A.8 – Tabela Professores Camada 1

```

1 CREATE TABLE camada1.Professores(
2     ATIVO nvarchar(255) NULL,
3     NOME nvarchar(255) NULL,
4     CPF nvarchar(255) NULL,
5     ESCOLARIDADE nvarchar(255) NULL,
6     GRADUACAO nvarchar(255) NULL,
7     APERFEICOAMENTO nvarchar(255) NULL,
8     ESPECIALIZACAO nvarchar(255) NULL,
9     MESTRADO nvarchar(255) NULL,
10    DOUTORADO nvarchar(255) NULL
11 )

```

A.5 Scripts para criação das views da Camada 2

Código A.9 – View Disciplinas Camada 2

```

1 CREATE VIEW camada2.Disciplinas
2 AS
3 SELECT
4     concat(d.ID_DISCIPLINA, '|', d.PERIODO_LETIVO) as idOrigem,
5     d.PERIODO_LETIVO as periodoLetivo,
6     dCurso,
7     d.ID_DISCIPLINA as codigoDisciplina,
8     d.DISCIPLINA as nomeDisciplina,
9     try_convert(int, d.CARGA_HORARIA) as cargaHoraria,
10    try_convert(int, d.QUANTIDADE_AULAS) as quantidadeAulas
11 FROM camada1.Disciplinas d

```

Código A.10 – View Matriculas Camada 2

```

1 CREATE VIEW camada2.Matriculas
2 AS
3 SELECT
4     concat(a.id, '|', d.id, '|', p.id) as idOrigem,
5     a.id AS idAluno,
6     d.id AS idDisciplina,
7     p.id AS idProfessor,
8     m.PERIODO_LETIVO as periodoLetivo,
9     try_convert(FLOAT, replace(m.MEDIA_FINAL, ',', '.')) AS mediaFinal,
10    m.SITUACAO_FINAL AS situacaoFinal,
11    try_convert(INT, m.FALTAS) AS Faltas
12 FROM camada1.Matriculas m
13 LEFT JOIN dw.Alunos a ON m.NOME = a.Nome
14     AND m.ADMISSAO = a.Admissao
15 LEFT JOIN dw.Disciplinas d ON m.ID_DISCIPLINA = d.codigoDisciplina
16     AND m.PERIODO_LETIVO = d.periodoLetivo
17 LEFT JOIN dw.Professores p ON m.CPF_PROFESSOR = p.CPF

```

Código A.11 – View Alunos Camada 2

```

1 CREATE VIEW camada2.Alunos
2 AS
3 SELECT
4     concat(a.NOME,'|',a.ADMISSAO) as idOrigem,
5     a.NOME as Nome,
6     a.CPF,
7     a.MATRICULA as Matricula,
8     a.ADMISSAO as Admissao,
9     a.PREVISAO_CONCLUSAO as previsaoConclusao,
10    a.CONCLUSAO_CURSO as conclusaoCurso,
11    a.FORMA_INGRESSO as formaIngresso,
12    a.TIPO_FORMA_INGRESSO as tipoFormaIngresso,
13    a.SITUACAO as Situacao,
14    a.CURSO as Curso,
15    a.PERFIL_CURSO as perfilCurso,
16    try_convert(DATE, a.DT_NASCIMENTO, 103) as dataNascimento,
17    DATEDIFF(YEAR, try_convert(DATE, a.DT_NASCIMENTO, 103), GETDATE()) as
        ↳ Idade,
18    a.ESTADO_CIVIL as estadoCivil,
19    a.SEXO as Sexo,
20    a.CONCLUSAO_ENS_MEDIO as conclusaoEnsinoMedio,
21    try_convert(FLOAT, replace(a.MEDIA_GERAL, ',', '.', '')) as mediaGeral,
22    try_convert(FLOAT, replace(a.COEFICIENTE_RENDIMENTO, ',', '.', '')) as
        ↳ coeficienteRendimento,
23    a.BLOCADO as Blocado,
24    a.FAIXA_ECONOMICA as faixaEconomica
25 FROM camada1.Alunos a

```

Código A.12 – View Professores Camada 2

```

1 CREATE VIEW camada2.Professores
2 AS
3 SELECT
4     p.CPF as idOrigem,
5     p.NOME as Nome,
6     p.CPF,
7     p.ESCOLARIDADE as Escolaridade,
8     CASE WHEN p.ATIVO = 'SIM' THEN 0 ELSE 1 END as professorExterno
9 FROM camada1.Professores p

```

A.6 Scripts para criação das views da Camada 3

Código A.13 – View Datas Camada 3

```

1 CREATE VIEW camada3.Datas as
2 select
3     anoSemestre as 'Ano/Semestre',
4     Ano as 'Ano',
5     Semestre as 'Semestre'
6 from dw.Datas

```

Código A.14 – View Alunos Camada 3

```
1 CREATE VIEW camada3.Alunos as
2 select
3     id,
4     Nome,
5     Admissao,
6     formaIngresso as 'Forma de Ingresso',
7     tipoFormaIngresso as 'Tipo Forma de Ingresso',
8     Situacao,
9     Sexo,
10    Idade,
11    conclusaoCurso as 'Data de Conclusao'
12 from dw.Alunos
```

Código A.15 – View Disciplinas Camada 3

```
1 CREATE VIEW camada3.Disciplinas as
2 select
3     id,
4     periodoLetivo as 'Periodo Letivo',
5     codigoDisciplina as 'Codigo',
6     nomeDisciplina as 'Nome da Disciplina',
7     cargaHoraria as 'Carga Horaria'
8 from dw.Disciplinas
```

Código A.16 – View Matriculas Camada 3

```
1 CREATE VIEW camada3.Matriculas as
2 select
3     m.id,
4     m.idAluno,
5     m.idDisciplina,
6     m.idProfessor,
7     m.periodoLetivo as 'Periodo Letivo',
8     m.mediaFinal as 'Nota Final',
9     m.situacaoFinal as 'Situacao',
10    m.Faltas
11 from dw.Matriculas m
```

A.7 Scripts para criação das tabelas do DW

Código A.17 – Tabela Alunos DW

```
1 CREATE TABLE dw.Alunos (
2     id INT identity NOT NULL,
3     idOrigem VARCHAR(50) NOT NULL,
4     Nome VARCHAR(120) NOT NULL,
5     CPF VARCHAR(12) NOT NULL,
6     Matricula VARCHAR(20) NOT NULL,
7     Admissao VARCHAR(7) NOT NULL,
```

```

8      previsaoConclusao VARCHAR(7) NULL,
9      conclusaoCurso VARCHAR(7) NULL,
10     formaIngresso VARCHAR(30) NULL,
11     tipoFormaIngresso VARCHAR(30) NULL,
12     Situacao VARCHAR(30) NULL,
13     Curso VARCHAR(40) NULL,
14     perfilCurso VARCHAR(15) NULL,
15     dataNascimento DATE NULL,
16     Idade INT NULL,
17     estadoCivil VARCHAR(15),
18     Sexo varchar(1) NULL,
19     conclusaoEnsinoMedio VARCHAR(20) NULL,
20     mediaGeral FLOAT NULL,
21     coeficienteRendimento FLOAT NULL,
22     Blocado varchar(5) NULL,
23     faixaEconomica varchar(30) NULL,
24     periodoLetivo varchar(10) NULL,
25     evasaoCurso varchar(10) NULL,
26     CONSTRAINT PK_Alunos PRIMARY KEY (id),
27     CONSTRAINT UN_idOrigem_Alunos UNIQUE (idOrigem)
28 )

```

Código A.18 – Tabela Datas DW

```

1 CREATE TABLE dw.Datas (
2     anoSemestre VARCHAR(20) NOT NULL,
3     Ano INT NOT NULL,
4     Semestre INT NOT NULL,
5     CONSTRAINT PK_Datas PRIMARY KEY (anoSemestre)
6 )

```

Código A.19 – Tabela Disciplinas DW

```

1 CREATE TABLE dw.Disciplinas (
2     id INT identity(1, 1) NOT NULL,
3     idOrigem VARCHAR(50) NOT NULL,
4     periodoLetivo VARCHAR(8) NOT NULL,
5     Curso VARCHAR(40) NOT NULL,
6     codigoDisciplina VARCHAR(15) NOT NULL,
7     nomeDisciplina VARCHAR(250) NOT NULL,
8     cargaHoraria INT NOT NULL,
9     quantidadeAulas INT NOT NULL,
10    CONSTRAINT PK_Disciplinas PRIMARY KEY (id),
11    CONSTRAINT UN_idOrigem_Disciplinas UNIQUE (idOrigem)
12 )

```

Código A.20 – Tabela Matriculas DW

```

1 CREATE TABLE dw.Matriculas (
2     id INT identity(1, 1) NOT NULL,
3     idOrigem VARCHAR(30) NOT NULL,
4     idAluno INT NOT NULL,
5     idDisciplina INT NOT NULL,

```

```
6      idProfessor INT NOT NULL,
7      periodoLetivo varchar(15) NOT NULL,
8      mediaFinal FLOAT NOT NULL,
9      situacaoFinal VARCHAR(30) NOT NULL,
10     Faltas INT NOT NULL,
11     CONSTRAINT PK_Matriculas PRIMARY KEY (id),
12     CONSTRAINT UN_idOrigem_Matriculas UNIQUE (idOrigem),
13     CONSTRAINT FK_Matriculas_Alunos FOREIGN KEY (idAluno) REFERENCES dw.
14         ↳ Alunos (id),
15     CONSTRAINT FK_Matriculas_Disciplinas FOREIGN KEY (idDisciplina)
16         ↳ REFERENCES dw.Disciplinas (id),
17     CONSTRAINT FK_Matriculas_Professores FOREIGN KEY (idProfessor)
18         ↳ REFERENCES dw.Professores (id)
19 )
```

Código A.21 – Tabela Professores DW

```
1 CREATE TABLE dw.Professores (
2     id INT identity(1, 1) NOT NULL,
3     idOrigem VARCHAR(12) NOT NULL,
4     Nome VARCHAR(50) NOT NULL,
5     CPF VARCHAR(12) NOT NULL,
6     Escolaridade VARCHAR(20) NOT NULL,
7     professorExterno BIT NOT NULL,
8     CONSTRAINT PK_Professores PRIMARY KEY (id),
9     CONSTRAINT UN_idOrigem_Professores UNIQUE (idOrigem)
10 )
```

APÊNDICE B – Indicadores DAX

Este apêndice apresenta os indicadores DAX (Data Analysis Expressions) desenvolvidos para o Power BI, utilizados na construção dos dashboards e análises apresentados neste trabalho.

B.1 Indicadores de Desempenho Acadêmico

Reprovações:

```
1 CALCULATE(COUNT(Matriculas[Situacao Final]),
2     FILTER(Matriculas,
3         Matriculas[Situacao Final]="REPROVADO" ||
4         Matriculas[Situacao Final]="REPROVADO POR FALTA"))
```

Aprovações:

```
1 CALCULATE(COUNT(Matriculas[Situacao Final]),
2     FILTER(Matriculas,
3         Matriculas[Situacao Final]="APROVADO" ||
4         Matriculas[Situacao Final] = "APROVADO POR FREQUENCIA" ||
5         Matriculas[Situacao Final] = "APROVADO POR MEDIA"))
```

Média Nota:

```
1 AVERAGE(Matriculas[Media Final])
```

Média Faltas:

```
1 AVERAGE(Matriculas[Faltas])
```

% Faltas:

```
1 DIVIDE(SUM(Matriculas[Faltas]),
2     SUMX(Matriculas,RELATED(Disciplinas[Quantidade Aulas])))
```

% Reprovação:

```
1 DIVIDE([Reprovacoes],[Matriculas])
```

B.2 Indicadores de Evasão e Retenção

Evadidos:

```

1 CALCULATE(COUNTROWS(GROUPBY(Alunos,Alunos[CPF],Alunos[Ano Admissao])),
2   FILTER(Matriculas,
3     RELATED(Alunos[Ano Evasao do Curso])=Matriculas[Periodo Letivo]))

```

Evadidos 2:

```

1 CALCULATE(COUNTROWS(GROUPBY(Alunos,Alunos[CPF],Alunos[Ano Admissao])),
2   FILTER(Alunos,
3     Alunos[Situacao]="TRANSFERENCIA INTERNA" ||
4     Alunos[Situacao]="MATRICULA RECUSADA" ||
5     Alunos[Situacao]="DESVINCULADO" ||
6     Alunos[Situacao]="TRANSFERENCIA EXTERNA" ||
7     Alunos[Situacao]="DESLIGAMENTO"))

```

% Evasão Total:

```

1 DIVIDE([Evadidos],[Ingressantes])

```

Retidos:

```

1 CALCULATE(COUNTROWS(GROUPBY(Alunos,Alunos[CPF],Alunos[Ano Admissao])),
2   FILTER(Matriculas,
3     RELATED(Alunos[Periodo Letivo]) > RELATED(Alunos[Previsao Conclusao
4     ↪ ])) &&
5     (ISBLANK(RELATED(Alunos[Ano Conclusao do Curso]))))

```

% Retenção:

```

1 DIVIDE([Retidos],[Matriculados])

```

B.3 Indicadores de Matrícula e Ingresso

Ingressantes:

```

1 CALCULATE(COUNTROWS(GROUPBY(Alunos,Alunos[CPF],Alunos[Ano Admissao])),
2   FILTER(Matriculas,
3     RELATED(Alunos[Ano Admissao])=Matriculas[Periodo Letivo]))

```

Ingressantes 2:

```

1 CALCULATE(COUNTROWS(GROUPBY(Alunos,Alunos[CPF],Alunos[Ano Admissao])),

```

```

2 FILTER(Alunos,
3     Alunos[Periodo Letivo]=Alunos[Ano Admissao]))

```

Matriculados:

```

1 DISTINCTCOUNT(Matriculas[idAluno])

```

Matriculados 2:

```

1 CALCULATE(COUNTROWS(GROUPBY(Alunos,Alunos[CPF],Alunos[Ano Admissao])),
2     FILTER(Alunos,
3     Alunos[Situacao] = "MATRICULADO"))

```

Matrículas:

```

1 COUNTROWS(Matriculas)

```

B.4 Indicadores de Conclusão

Concluintes:

```

1 CALCULATE(COUNTROWS(GROUPBY(Alunos,Alunos[CPF],Alunos[Ano Admissao])),
2     FILTER(Matriculas,
3     RELATED(Alunos[Ano Conclusao do Curso])=Matriculas[Periodo Letivo])
4     )

```

% Conclusão:

```

1 DIVIDE([Concluintes],[Ingressantes])

```

Tempo médio Conclusão:

```

1 CALCULATE(
2     AVERAGEX(
3     SUMMARIZE(Alunos,Alunos[CPF],Alunos[Ano Admissao],
4     "Soma Tempo Conclusao",SUM(Alunos[Tempo Conclusao])),
5     [Soma Tempo Conclusao]),
6     FILTER(Matriculas,
7     RELATED(Alunos[Ano Conclusao do Curso])=Matriculas[Periodo Letivo])
8     )

```

Tempo médio Excedido:

```

1 CALCULATE(

```

```

2  AVERAGEX (
3      SUMMARIZE(Alunos, Alunos[CPF], Alunos[Ano Admissao],
4          "Soma Tempo Excedente", SUM(Alunos[Tempo Excedente])),
5      [Soma Tempo Excedente]),
6  FILTER(Matriculas,
7      RELATED(Alunos[Ano Conclusao do Curso])=Matriculas[Periodo Letivo])
      ↪ )

```

B.5 Indicadores Complementares

Qtd Alunos:

```
1  COUNTROWS(Alunos)
```

Média Geral:

```
1  AVERAGE(Alunos[Media Geral])
```

Média Graduação:

```

1  CALCULATE(AVERAGE(Alunos[Media Geral]),
2      FILTER(Matriculas,
3          RELATED(Alunos[Periodo Letivo])=Matriculas[Periodo Letivo]))

```

Faixa Etária:

```
1  COUNTROWS(GROUPBY(Alunos, Alunos[CPF], Alunos[Ano Admissao], Alunos[Faixa
      ↪ Etaria]))
```

Menor Média Disciplina:

```

1  MINX(KEEPFILTERS(VALUES('Disciplinas'[Nome Disciplina])),
2      CALCULATE([Media Nota]))

```

B.6 Indicadores de Situação Especial

Blocados:

```

1  CALCULATE(COUNTROWS(GROUPBY(Alunos, Alunos[CPF], Alunos[Ano Admissao])),
2      FILTER(Matriculas,
3          RELATED(Alunos[Blocado])="SIM"))

```

Blocados 2:

```
1 CALCULATE(COUNTROWS(GROUPBY(Alunos, Alunos[CPF], Alunos[Ano Admissao])),
2     FILTER(Matriculas,
3         RELATED(Alunos[Blocao])="SIM" &&
4         RELATED(Alunos[Periodo Letivo]) = Matriculas[Periodo Letivo]))
```

% Blocados:

```
1 DIVIDE([Blocados 2],[Matriculados])
```

Trancamentos/Matrícula Vinculo:

```
1 CALCULATE(COUNTROWS(GROUPBY(Alunos, Alunos[CPF], Alunos[Ano Admissao])),
2     FILTER(Alunos,
3         Alunos[Situacao]="TRANCAMENTO" ||
4         Alunos[Situacao] = "MATRICULA VINCULO"))
```