



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Enade.Vis: Uma ferramenta para visualização dos dados do Exame Nacional de Desempenho dos Estudantes (Enade)

Aline Mitiko Otsubo

Monografia apresentada como requisito parcial
para conclusão do curso de Bacharelado em Ciência da Computação

Orientador
Prof. Dr. Jan Mendonça Correa

Brasília
2024

Dedicatória

Dedico este trabalho à minha batian, Kiyoko, que hoje se encontra nas profundezas do firmamento.

Agradecimentos

Sou feita pelos traços e resquícios de todos que cruzaram meu caminho. Se finalizo mais uma etapa da minha vida, devo toda minha gratidão àqueles que me influenciaram, me ensinaram e me mostraram os diversos possíveis caminhos que eu poderia seguir.

Agradeço aos colegas da UnB que, ao meu lado, enfrentaram os desafios da graduação, suavizando o peso das exigências e tornando o percurso mais suportável. Um agradecimento especial ao Juan, que compartilhou comigo toda essa experiência.

Eterna gratidão à minha Grande Família, que me guiou e apoiou em todas as minhas decisões, sempre presente e oferecendo todo o suporte necessário para encarar o desconhecido. Cada ação, mesmo que pequena, contribuiu para o êxito dessa árdua caminhada. Assim, agradeço à minha mãe, Junko, ao Atsushi, à Ju, ao Yuichi, ao Kazuo, à Manami, à Mutian, à Lumiko, à Lídia, ao Yukio e aos pequenos da minha família, a quem desejo um caminho incrível pela frente.

Por fim, expresso meu agradecimento a todos os professores e colaboradores do Departamento de Ciência da Computação da UnB, que contribuíram para minha formação ao longo da graduação. Um agradecimento especial ao meu orientador, Jan Mendonça Correa, por sua orientação e apoio na elaboração do meu trabalho, ao Daniel Sanchez, por me introduzir ao tema, e ao Fábio Marley, que me ensinou tanto nesta reta final.

Resumo

O Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) é uma entidade federal vinculada ao Ministério da Educação (MEC), responsável pela aplicação de diversas avaliações educacionais e pela disponibilização de estudos detalhados com base nos dados coletados. Entre essas avaliações, destaca-se o Exame Nacional de Desempenho dos Estudantes (Enade), que estima o desempenho dos concluintes de diversos cursos de graduação no Brasil. Este trabalho foca na criação de uma ferramenta que automatiza a geração de análises comparativas entre cursos da mesma área de conhecimento no contexto do Enade, apresentando informações valiosas, como desempenho na prova, perfil do estudante, percepção da prova, entre outros. A interface é configurada para tratar as cinco últimas edições disponíveis, sendo estas de 2017, 2018, 2019, 2021 e 2022. O projeto foi desenvolvido utilizando técnicas de mineração de dados com o auxílio da ferramenta Jupyter Notebook para estudo dos dados e elaboração das análises, bem como a linguagem Python, com foco na biblioteca Dash para criação da interface. O objetivo deste trabalho é apresentar uma interface que centralize a maior quantidade de informações úteis para a realização e interpretação das análises a nível de curso. Além disso, são apresentados alguns exemplos de utilização focados no curso de Ciência da Computação (Bacharelado) da Universidade de Brasília (UnB). Como resultado, foi desenvolvida uma ferramenta *Web* que realiza análises comparativas a nível de curso, sem restrição de quantidade, permitindo a seleção com base em critérios como notas, localização, categoria administrativa, organização acadêmica, modalidade de ensino, lista de instituições e quantidade mínima de participantes.

Palavras-chave: Enade, Interface, Análise Comparativa, UnB, Graduação, Ciência da Computação

Abstract

The National Institute for Educational Studies and Research Anísio Teixeira (Inep) is a federal entity linked to the Ministry of Education (MEC), responsible for administering various educational assessments and providing detailed studies based on the data collected. Among these assessments, the National Student Performance Exam (Enade) stands out, which measures the performance of graduating students from various undergraduate courses in Brazil. This work focuses on creating a tool that automates the generation of comparative analyses between courses in the same field of study within the context of Enade, providing valuable information such as exam performance, student profile, exam perception, among others. The interface is designed to handle the last five available editions, namely those of 2017, 2018, 2019, 2021, and 2022. The project was developed using data mining techniques with the assistance of the Jupyter Notebook tool for data analysis and elaboration, as well as Python programming language, focusing on the Dash library for the creation of the interface. The objective of this work is to present an interface that centralizes the largest amount of useful information for conducting and interpreting analyses at the course level. Additionally, some usage examples are provided, focusing on the Bachelor's Degree in Computer Science at the University of Brasília (UnB). As a result, a web-based tool was developed that conducts comparative analyses at the program level, without quantity restrictions, allowing selections based on criteria such as scores, location, administrative category, academic organization, mode of instruction, list of institutions, and a minimum number of participants.

Keywords: Enade, Interface, Comparative Analysis, UnB, Undergraduate, Computer Science

Sumário

1	Introdução	1
1.1	Problema	2
1.2	Motivação	2
1.3	Objetivo Geral	2
1.3.1	Objetivos Específicos	3
1.4	Metodologia	3
1.5	Estrutura dos Capítulos	4
2	Fundamentação Teórica	6
2.1	Exame Nacional de Desempenho dos Estudantes (Enade)	6
2.1.1	Áreas de Conhecimento	7
2.1.2	Estrutura da Prova	7
2.1.3	Conceito Enade	8
2.1.4	Microdados do Enade	9
2.1.5	Impactos da LGPD no Enade	9
2.2	Dado, Informação e Conhecimento	10
2.3	Extração, Tratamento e Carga	11
2.4	Mineração de Dados	11
2.5	Experiência do Usuário (UX)	12
2.6	Trabalhos Relacionados	12
2.7	Ferramentas e Tecnologias	13
2.7.1	Figma	13
2.7.2	Jupyter Notebook	13
2.7.3	PostgreSQL	13
2.7.4	Python	14
3	Preparação dos Dados	15
3.1	Fonte dos Dados	15
3.1.1	Microdados do Enade	15

3.1.2	Cursos de Graduação do Brasil	17
3.1.3	Municípios Brasileiros	17
3.1.4	Conceito Enade	17
3.2	Processamento dos Dados	18
3.2.1	Informações Gerais do Curso	18
3.2.2	Ranking de Acertos	20
3.2.3	Distribuição de Notas	21
3.2.4	Percepção da Prova	22
3.2.5	Perfil do Estudante	24
3.2.6	Processo de Formação	25
3.3	Migração dos Dados para o PostgreSQL	26
4	Estrutura da Ferramenta Enade.Vis	28
4.1	Curso Principal	30
4.2	Cursos para Comparação	31
4.3	Análises Disponíveis	32
4.4	Resumo da Seleção	34
5	Exemplos de Utilização	36
5.1	Instituições Classificadas como PUC	37
5.1.1	Análise do Ranking de Acertos por Questão	38
5.1.2	Análise das Notas Finais	40
5.1.3	Análise da Distribuição de Notas	42
5.1.4	Análise da Percepção da Prova	45
5.1.5	Análise do Perfil do Estudante	47
5.1.6	Análise do Processo de Formação	47
5.2	Exemplos Adicionais de Aplicação de Filtros	50
5.2.1	Universidades Públicas com Conceito Enade 5	50
5.2.2	Cursos Localizados no Distrito Federal	51
5.2.3	Lista de Instituições	52
6	Conclusões	54
6.1	Trabalhos Futuros	56
	Referências	57

Lista de Figuras

3.1	Processamento dos atributos referentes as características dos cursos.	19
3.2	Processamento dos atributos referentes ao ranking de acertos.	21
3.3	Processamento dos atributos referentes à percepção da prova.	24
3.4	Processamento dos atributos referentes ao perfil do estudante.	25
3.5	Primeira questão do processo de formação.	25
3.6	Processamento dos atributos referentes ao processo de formação.	26
4.1	Fluxo de utilização.	28
4.2	Tela de seleção do curso principal.	30
4.3	Tela de seleção de cursos para comparação.	31
4.4	Exemplo de cursos que possuem nomes iguais.	32
4.5	Tela de seleção da análise.	33
4.6	Tela de resumo das informações selecionadas.	34
5.1	Desempenho geral do curso de Ciência da Computação da UnB, baseado em [1].	36
5.2	Seleção do curso de Ciência da Computação (Bacharel) da UnB.	38
5.3	Seleção das instituições PUC.	38
5.4	Ranking de acertos da questão 9 (UnB x PUC's) - Visualização Detalhada.	39
5.5	Ranking de acertos da questão 9 (UnB x PUC's) - Visualização Geral.	39
5.6	Questão 9 do caderno de provas de Ciência da Computação 2021, baseado em [2].	40
5.7	Notas finais (UnB x PUC's).	41
5.8	Estrutura da análise de distribuição de notas.	42
5.9	Campo de seleção de curso para a análise de distribuição de notas.	43
5.10	Distribuição das notas dos participantes (UnB x PUCRS) - Nota Final.	43
5.11	Distribuição das notas dos participantes (UnB x PUCRS) - Nota na Formação Geral.	44
5.12	Distribuição das notas dos participantes (UnB x PUCRS) - Nota no Componente Específico.	44

5.13	Grau de dificuldade da prova na parte do Componente Específico.	46
5.14	Sexo dos participantes.	47
5.15	As disciplinas cursadas contribuíram para a formação integral, como cida- dão e profissional.	49
5.16	Filtros utilizados para cursos com Conceito Enade 5.	51
5.17	Cursos encontrados a partir da aplicação do filtro para universidades pú- blicas com Conceito Enade 5.	51
5.18	Aplicação do filtro de localização.	51
5.19	Seleção dos cursos do Distrito Federal.	52
5.20	Aplicação do filtro de IES.	52
5.21	Cursos encontrados a partir da aplicação do filtro de IES.	53

Lista de Tabelas

2.1	Ciclo avaliativo do Enade, baseado em [3].	7
2.2	Faixa de notas do Conceito Enade e seus intervalos, baseado em [4].	9
3.1	Quantidade de registros antes e depois por ano de edição	20
4.1	Requisitos funcionais e não funcionais.	29
5.1	Cursos selecionados para comparação com suas respectivas notas no Conceito Enade.	50

Lista de Abreviaturas e Siglas

CNPq Conselho Nacional de Desenvolvimento Científico e Tecnológico.

Enade Exame Nacional de Desempenho de Estudantes.

ETL Extract, Transform, Load.

IBGE Instituto Brasileiro de Geografia e Estatística.

Inep Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira.

LGPD Lei Geral de Proteção de Dados Pessoais.

MEC Ministério da Educação.

Sedap Serviço de Acesso a Dados Protegidos.

SINAES Sistema Nacional de Avaliação da Educação Superior.

Capítulo 1

Introdução

Analisar dados é uma tarefa de extrema importância em um cenário impulsionado pela coleta massiva de informações. Essa análise é crucial para identificar padrões, detectar erros e otimizar processos em diversas áreas.

No entanto, lidar com grandes quantidades de dados brutos exige conhecimento especializado para tratá-los e organizá-los, transformando-os em informações úteis. Nesse contexto, surge a necessidade de criar ambientes que facilitem a visualização dos dados, especialmente para aqueles que não possuem conhecimento técnico suficiente para tratá-los.

Levando isso para o contexto da educação superior no Brasil, uma das principais fontes de dados educacionais são os microdados disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), que contém milhares de registros brutos referentes aos participantes do Exame Nacional de Desempenho de Estudantes (Enade). Esses dados têm sido objeto de numerosos estudos e análises.

Diante dessa realidade, surge a ideia de criar uma ferramenta que torne facilmente disponível a visualização dessas análises, facilitando o acesso e a interpretação dos dados por um público mais amplo. Essa ferramenta não apenas visa simplificar a visualização dos dados, mas também a de promover uma maior transparência e compreensão das informações obtidas através dos microdados do Inep, contribuindo assim para o desenvolvimento de estudos mais aprofundados e acessíveis.

Embora já existam iniciativas acadêmicas focadas na criação de interfaces para análise dos dados do Enade, muitos desses trabalhos apresentam limitações, conforme discutido na Seção 2.6. Além disso, os relatórios de cursos divulgados pelo Inep apresentam comparações entre cursos específicos e médias estaduais, regionais e nacionais, baseando-se na localização geográfica. No entanto, essas análises fornecem apenas uma visão superficial, informando se o curso está acima ou abaixo das médias, sem oferecer uma compreensão mais detalhada.

Este trabalho busca, portanto, não só introduzir novas funcionalidades que superem essas limitações, mas também oferecer uma análise comparativa mais ampla entre diversos cursos. Especificamente, busca fornecer interpretações mais detalhadas e precisas a nível de curso, tratando das lacunas presentes nos relatórios atuais do Enade.

1.1 Problema

A análise dos dados do Enade apresenta desafios significativos devido ao grande volume e à complexidade das informações coletadas. Os dados, distribuídos em diversos arquivos e edições, precisam ser processados e integrados de forma a gerar análises comparativas úteis. A falta de ferramentas intuitivas para visualização e a dificuldade em comparar o desempenho de cursos em diferentes contextos complicam ainda mais a interpretação dos resultados.

1.2 Motivação

A análise de dados, especialmente no contexto educacional brasileiro, é fundamental para identificar padrões que ajudem a compreender o cenário atual de ensino. Os microdados do Enade, disponibilizados pelo Inep, oferecem uma fonte rica de informações sobre a qualidade do ensino superior, mas sua complexidade pode ser um obstáculo para análises eficazes, principalmente para quem não possui formação técnica.

A construção de uma ferramenta *Web* para visualização desses dados democratiza o acesso às informações, proporcionando uma visualização clara, promovendo a transparência e permitindo que essas informações sejam utilizadas para melhorar a qualidade do ensino superior.

1.3 Objetivo Geral

O objetivo principal deste trabalho é construir uma ferramenta *Web* que facilite a visualização das análises dos dados do Enade a nível de curso, acessível de qualquer lugar. A ferramenta deve permitir comparações detalhadas entre cursos, possibilitando a observação do desempenho por questão de um curso específico em relação a um grupo de cursos em diferentes contextos. Além disso, deve possibilitar a visualização da distribuição das notas dos participantes e das notas finais obtidas pelos cursos.

A ferramenta também deve oferecer análises sobre o perfil dos estudantes, o processo de formação e a percepção da prova, fornecendo uma visão mais abrangente desses aspectos. Ou seja, a ferramenta deve permitir a visualização de dados de qualquer questão

do questionário, ampliando as possibilidades de análise e interpretação das informações disponíveis.

Adicionalmente, a ferramenta deve permitir a seleção de cursos com base em critérios como notas, localização, categorias administrativas, modalidades de ensino, uma lista pré-definida de instituições e a quantidade mínima de participantes.

1.3.1 Objetivos Específicos

Para alcançar o objetivo principal do trabalho, foram necessários os seguintes objetivos específicos:

- Extrair os microdados do Enade das edições de 2017, 2018, 2019, 2021 e 2022;
- Compreender o contexto dos dados por meio da leitura dos manuais e dicionários de variáveis;
- Realizar o tratamento dos dados, removendo registros que não se enquadram no escopo do trabalho e criando novas estruturas;
- Desenvolver protótipos das telas que irão compor a ferramenta;
- Implementar ferramenta que permita realizar comparações entre cursos;
- Criar filtros que possibilitem a seleção de cursos com base em múltiplos critérios;
- Elaborar análises utilizando gráficos de barras, tabelas, histogramas e métricas estatísticas;
- Implementar funcionalidades que permitam uma visualização detalhada do desempenho nas questões das provas, além de análises sobre o perfil dos estudantes, a percepção da prova e o processo formativo;
- Desenvolver visualizações das análises de desempenho nas questões, utilizando ranqueamentos comparativos e informações complementares, ausentes nos resultados oficiais do Inep;
- Testar a ferramenta em diferentes cenários e contextos;
- Avaliar os resultados gerados pela ferramenta.

1.4 Metodologia

A metodologia deste trabalho se baseou nos princípios da Extração [5], Transformação e Carga (ETL) [6], mineração de dados e do modelo incremental [7]. O primeiro passo

envolveu a compreensão dos dados do Enade, com foco no estudo inicial de sua estrutura, exploração dos atributos disponíveis e planejamento das informações a serem utilizadas no desenvolvimento da ferramenta. Essa etapa foi guiada pelo estudo dos manuais fornecidos pelo Inep e pelos arquivos de variáveis associados.

Na sequência, ocorreu a fase de formatação dos dados, em que códigos em Python foram empregados para processar, limpar e criar novas estruturas de dados, além de detectar e tratar valores inconsistentes.

Após a formatação, iniciou-se a fase de prototipação, na qual o *design* da ferramenta foi elaborado, possibilitando uma visualização inicial de como os dados poderiam ser organizados e relacionados. Com o *design* definido, seguiu-se para a implementação da ferramenta, integrando os dados organizados anteriormente.

Finalmente, foram realizados testes para identificar e corrigir erros, além de propor melhorias. Esse processo foi iterativo, repetindo etapas para melhoria da interface e para um melhor entendimento dos dados que poderiam ser acrescentados ao trabalho.

1.5 Estrutura dos Capítulos

Os capítulos do trabalho são estruturados da seguinte forma:

- **Capítulo 1: Introdução**

O capítulo apresenta uma visão geral do trabalho, incluindo os objetivos, a justificativa e a organização do documento.

- **Capítulo 2: Fundamentação teórica**

O capítulo apresenta os principais conceitos e teorias que embasam o trabalho, fornecendo o contexto necessário para a compreensão do trabalho.

- **Capítulo 3: Preparação dos Dados**

O capítulo descreve o processo de coleta, limpeza e organização dos dados utilizados para criação da ferramenta e análise, detalhando as fontes de dados e as técnicas aplicadas.

- **Capítulo 4: Estrutura da Ferramenta Enade.Vis**

O capítulo apresenta a estrutura da ferramenta desenvolvida, incluindo a descrição de todos os componentes e funcionalidades implementadas.

- **Capítulo 5: Exemplos de Utilização**

O capítulo detalha exemplos de utilização interessantes para a ferramenta, ilustrando como diferentes usuários podem interagir com a ferramenta para atingir seus objetivos. Além de fornecer interpretações das análises geradas.

- **Capítulo 6: Conclusões**

Este capítulo consolida as etapas essenciais do trabalho, discute as implicações dos resultados gerados pela ferramenta e sugere possíveis direções para trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Este capítulo tem como objetivo fornecer uma revisão dos conceitos relevantes para o trabalho, além de discutir estudos existentes que buscam criar interfaces no contexto do Enade. Também são apresentadas as ferramentas e tecnologias empregadas no desenvolvimento do projeto de interface.

2.1 Exame Nacional de Desempenho dos Estudantes (Enade)

Com o intuito de criar um método padrão para avaliar o ensino superior brasileiro, surgiu em 14 de abril de 2004, através da Lei nº 10.861, o Sistema Nacional de Avaliação da Educação Superior (SINAES) [8]. O sistema se estrutura em três pilares essenciais, sendo estes a avaliação dos cursos, das instituições de ensino e do desempenho dos estudantes.

Para avaliar o último aspecto mencionado, foi instituído o Exame Nacional de Desempenho de Estudantes (Enade), realizado em caráter anual desde 2004. A única exceção ocorrendo em 2020, devido à pandemia de Covid-19, quando sua aplicação foi prorrogada para o ano posterior [9].

O Enade é destinado aos estudantes concluintes dos diversos cursos de graduação e tem como objetivo avaliar tanto as competências que são comuns e essenciais a todas as áreas de conhecimento, e de forma mais crucial, as questões específicas do curso.

Os cursos avaliados a cada ano são definidos através do ciclo avaliativo, divulgados pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Esse ciclo, com duração de três anos, visa agrupar as áreas de conhecimento relacionadas.

No entanto, em 27 de junho de 2024, o Ministério da Educação (MEC) publicou as Portarias nº 610/2024 e nº 611/2024 [10] [11], introduzindo um novo formato de prova para a edição de 2024, denominado “Enade das Licenciaturas”. Este novo formato visa

aprimorar a avaliação dos estudantes concluintes de cursos de licenciatura, dividindo o processo avaliativo em duas partes: uma teórica e outra prática.

Além disso, as portarias estabelecem um novo ciclo avaliativo trienal, com início em 2025.

2.1.1 Áreas de Conhecimento

O ciclo avaliativo do Enade é organizado em diversas áreas de conhecimento, cada uma representando um campo específico de estudo [3]. Essas áreas são essenciais para a estruturação das avaliações, pois agrupam cursos que compartilham características e objetivos educacionais comuns.

A Tabela 2.1 [3] apresenta a organização das áreas de conhecimento envolvidas no ciclo avaliativo.

Ano	Áreas de Conhecimento
Ano 1	Cursos de bacharelado nas áreas de conhecimento de Ciências Agrárias, Ciências da Saúde e áreas afins.
	Cursos de bacharelado nas áreas de conhecimento de Engenharias e Arquitetura e Urbanismo.
	Cursos Superiores de Tecnologia nas áreas de Ambiente e Saúde, Produção Alimentícia, Recursos Naturais, Militar e Segurança.
Ano 2	Cursos de bacharelado nas áreas de conhecimento de Ciências Biológicas; Ciências Exatas e da Terra; Linguística, Letras e Artes e áreas afins.
	Cursos de licenciatura nas áreas de conhecimento de Ciências da Saúde; Ciências Humanas; Ciências Biológicas; Ciências Exatas e da Terra; Linguística, Letras e Artes.
	Cursos de bacharelado nas áreas de conhecimento de Ciências Humanas e Ciências da Saúde, com cursos avaliados no âmbito das licenciaturas.
	Cursos Superiores de Tecnologia nas áreas de Controle e Processos Industriais, Informação e Comunicação, Infraestrutura e Produção Industrial.
Ano 3	Cursos de bacharelado nas Áreas de Conhecimento Ciências Sociais Aplicadas e áreas afins.
	Cursos de bacharelado nas Áreas de Conhecimento Ciências Humanas e áreas afins que não tenham cursos também avaliados no âmbito das licenciaturas.
	Cursos Superiores de Tecnologia nas áreas de Gestão e Negócios, Apoio Escolar, Hospitalidade e Lazer, Produção Cultural e Design.

Tabela 2.1: Ciclo avaliativo do Enade, baseado em [3].

2.1.2 Estrutura da Prova

A prova do Enade tem como objetivo avaliar o desempenho dos estudantes em relação aos conteúdos abordados na estrutura programática prevista em cada diretriz curricular

do curso no qual o estudante está se graduando.

Embora as questões possam variar de acordo com a área de conhecimento, a prova segue uma estrutura padrão composta por duas partes:

- Formação Geral (FG):
 - Conteúdo abordado é comum para todas as áreas;
 - Tem como objetivo avaliar as competências gerais do estudante, tais como interpretação textual e conhecimentos gerais;
 - É composta por dez questões, sendo oito de múltipla escolha e duas discursivas.
- Componente Específico (CE):
 - Conteúdo abordado é específico da área de cada curso;
 - Tem como objetivo avaliar as competências e habilidades do estudante acerca de temas específicos da área;
 - É composta por trinta questões, sendo vinte e sete de múltipla escolha e três discursivas.

2.1.3 Conceito Enade

O Conceito Enade [12] é um dos indicadores de qualidade do ensino superior, destinado a divulgar o desempenho médio de um curso que realizou a prova, em comparação com a média atingida na área específica em que o curso se enquadra.

Para o cálculo da nota é levado em consideração três informações, sendo essas:

- A quantidade de estudantes que participaram da prova e forneceram respostas válidas;
- O desempenho dos estudantes nas questões de Formação Geral;
- O desempenho dos estudantes nas questões do Componente Específico.

Para simplificar a apresentação dos resultados, o Inep utiliza faixas de notas, que são representadas na Tabela 2.2 [4].

Além disso, para que um curso tenha uma nota válida no Conceito Enade, é necessário satisfazer a condição de ter no mínimo dois estudantes participantes com respostas válidas. Caso essa condição não seja satisfeita, o curso recebe a classificação “Sem Conceito (SC)”.

Faixa de Nota	Intervalo de Notas
1	0 - 0,945
2	0,945 - 1,945
3	1,945 - 2,945
4	2,945 - 3,945
5	3,945 - 5

Tabela 2.2: Faixa de notas do Conceito Enade e seus intervalos, baseado em [4].

2.1.4 Microdados do Enade

Um microdado é representado como a menor unidade de informação dentro de um conjunto de dados, geralmente associada a uma pesquisa ou avaliação específica [13]. No contexto do Enade, os microdados [14] são uma fonte essencial de informações disponibilizada pelo Inep com o intuito de apoiar a pesquisa e análise educacional.

Esses microdados oferecem uma visão abrangente do perfil dos estudantes, cobrindo aspectos demográficos, socioeconômicos e acadêmicos. Através da aplicação de técnicas de tratamento e análise de dados, é possível identificar fatores críticos que permitem compreender o estado atual da educação e orientar iniciativas voltadas à melhoria da qualidade de ensino no Brasil.

Os microdados do Enade estão organizados de forma que cada linha representa o registro de um aluno, sendo o código do curso (`co_curso`) o único atributo possível de ser utilizado para relacionamentos. Cada coluna na estrutura se refere a um dos inúmeros atributos, que variam desde as notas dos alunos até as respostas a determinados questionários.

Além disso, a análise dos microdados do Enade pode aprofundar a investigação sobre temas como a relação entre a situação socioeconômica dos estudantes e seu desempenho acadêmico, a eficácia dos currículos das instituições de ensino e a identificação de áreas que precisam de melhorias.

2.1.5 Impactos da LGPD no Enade

A Lei Geral de Proteção de Dados Pessoais (LGPD) [15] é uma legislação de extrema importância, especialmente considerando o grande volume de dados que circula pelo meio digital nos dias de hoje. Essa lei aborda a proteção de dados, com foco particular nos ambientes digitais.

Com o intuito de se adequar às exigências da LGPD, o Inep realizou, em 2022, uma reestruturação nos microdados do Enade, visando dificultar a identificação dos participantes [16]. Essa transformação incluiu medidas para dificultar a identificação dos indivíduos,

como a remoção de informações que poderiam, quando combinadas, revelar a identidade dos participantes.

Uma das implicações mais diretas da LGPD é a impossibilidade de associar as respostas das questões com os dados individuais dos participantes de maneira que permita sua identificação. Em outras palavras, embora as respostas dos questionários, as notas e outros dados relevantes sejam disponibilizados para análise, esses dados são tratados de forma a preservar o anonimato dos participantes.

No entanto, mesmo com essas medidas de proteção, ainda é possível acessar os dados de forma segura, garantindo a privacidade dos indivíduos enquanto se permite a realização de análises relevantes, através do Serviço de Acesso a Dados Protegidos (Sedap).

2.2 Dado, Informação e Conhecimento

No contexto deste trabalho, é fundamental compreender as distinções entre dados, informações e conhecimento, uma vez que esses conceitos estão profundamente interligados. A seguir, é apresentada uma definição de cada um deles para facilitar a compreensão.

Segundo Bellinger, Castro e Mills [17], dados são representações brutas que, por si só, não possuem significado. São elementos fundamentais, mas desprovidos de contexto ou interpretação. Setzer [18] complementa essa visão ao definir dado como “uma sequência de símbolos quantificados ou quantificáveis”. Assim, no âmbito deste trabalho, os dados são representados pelos microdados, através de números e caracteres organizados em uma estrutura específica, possibilitando seu processamento.

De acordo com Chen et al. [19], informações são “dados que representam os resultados de um processo computacional, como uma análise estatística, para atribuir significados aos dados”. As informações surgem quando os dados são processados e interpretados, adquirindo um significado contextual relevante. No contexto deste trabalho, as informações são extraídas dos microdados, que, após serem tratados e analisados, são apresentadas na interface de forma clara.

Por último, de acordo com Bellinger, Castro e Mills [17], conhecimento é “a aplicação de dados e informações para fornecer respostas às perguntas do tipo ‘como’”. O conhecimento envolve a capacidade de utilizar informações de forma a auxiliar nas tomadas de decisões e resolver problemas. No contexto deste trabalho, o conhecimento é gerado a partir da análise das informações extraídas dos microdados, permitindo a formulação de interpretações.

2.3 Extração, Tratamento e Carga

No contexto da análise de dados, o processo de Extract, Transform, Load (ETL) [6] possui um papel fundamental. Ele descreve um fluxo de trabalho que permite coletar, processar e armazenar dados provenientes de diversas fontes, de forma a tornar utilizáveis para fins de análises e tomadas de decisões.

De acordo com Bansal e Kagemann [20], a primeira fase, a extração, envolve a coleta de dados de fontes apropriadas, que podem estar em formatos como .csv, .xls, .txt ou acessíveis via APIs. Essa etapa é desafiadora devido à variedade de formatos e métodos de acesso, que exigem abordagens específicas para a coleta dos dados.

Nath et al. [21] destacam que a fase de transformação inclui a remoção de dados duplicados, normalização, refinamento, criação de identificadores únicos, geração de novos atributos baseados nos originais e tratamento de valores nulos. Esta etapa é essencial para garantir que os dados estejam em um formato consistente e adequado para análise.

A última etapa, a carga, envolve o armazenamento dos dados transformados em um banco de dados de destino [22]. Neste ponto, os dados passam a ser acessíveis para análises detalhadas ou outras aplicações.

O processo de ETL é importante no contexto de grandes volumes de dados, como os microdados do Enade utilizados neste trabalho, pois garante que os dados sejam preparados corretamente, permitindo uma análise precisa e confiável.

2.4 Mineração de Dados

A mineração de dados [5] é definida como a aplicação de algoritmos específicos para extrair padrões dos dados. É um passo dentro do processo de descoberta de conhecimento em bases de dados (KDD), que envolve a análise e descoberta de padrões úteis e interessantes a partir de grandes volumes de dados.

Entre os métodos de mineração de dados mais populares, a estatística é uma das técnicas mais tradicionais e amplamente utilizadas [23]. No contexto deste trabalho, a estatística é empregada como a principal ferramenta de análise de dados da interface, com o objetivo de realizar uma análise exploratória dos microdados para obter informações valiosas. A aplicação de métodos estatísticos permite identificar padrões e tendências nos dados, enriquecendo a interpretação e apoiando a formulação de estratégias.

2.5 Experiência do Usuário (UX)

A experiência do usuário (UX) [24] se refere às percepções e respostas de um indivíduo ao usar, ou ao antecipar o uso, de um produto, sistema ou serviço. Ela abrange aspectos emocionais, como as sensações do usuário durante a interação, e aspectos funcionais, como a compreensão da operação e a adequação ao contexto de uso. A qualidade da UX é moldada pelo estado interno do usuário, incluindo suas expectativas e motivações, e pelas características do sistema, como complexidade e usabilidade.

No desenvolvimento da ferramenta para a visualização dos dados do Enade, garantir uma experiência de usuário positiva é crucial. É necessário projetar uma interface que não apenas apresente os dados de forma intuitiva, mas também ofereça um desempenho satisfatório. Aplicar princípios de UX pode aumentar a eficiência na análise de dados, tornando a informação mais acessível e útil para os usuários.

2.6 Trabalhos Relacionados

Esta seção oferece uma visão geral dos trabalhos acadêmicos que utilizam dados do Enade, com ênfase na apresentação desses dados por meio de interfaces de usuário.

No trabalho de Costa [25], é apresentada a ferramenta Vis-Scholar, que gera gráficos comparativos entre estados com base na Análise de Componentes Principais (ACP). A ferramenta utiliza dados do Enade e do Enem para analisar fatores como concorrência do curso, carga horária e despesas. No entanto, o projeto não permite análises mais detalhadas focadas em instituições específicas.

Araújo [26] propõe uma ferramenta *Web* para análise exploratória dos microdados do Enade, aplicando técnicas de Extração de conhecimento (KDD) e algoritmos de classificação para prever o desempenho dos estudantes. O objetivo é aprimorar a visualização e auxiliar na tomada de decisões nas instituições de ensino superior. Contudo, a ferramenta limita as comparações a apenas dois cursos e utiliza dados desatualizados.

Cunha [27] desenvolve um *software* voltado para a análise dos dados do Enade, com foco nos cursos de Ciência da Computação. O projeto visa ajudar coordenadores e diretores de instituições de ensino superior na tomada de decisões, oferecendo análises automáticas por meio do Jupyter Notebook. No entanto, o *software* se restringe aos cursos de Ciência da Computação e não inclui dados de outras áreas. Além disso, a utilização da ferramenta requer conhecimento prévio em Jupyter Notebook, o que pode limitar seu acesso e aplicação para usuários que não estão familiarizados com essa plataforma.

O trabalho de Cruz [28] apresenta a ferramenta *Web* SAEN, que facilita a análise de dados históricos do Enade, permitindo comparações entre instituições de ensino e visua-

lização de dados de edições diferentes. No entanto, a ferramenta se limita a comparações entre apenas duas instituições, focando exclusivamente nas notas finais.

Este trabalho apresenta uma solução que supera as limitações existentes, oferecendo uma ferramenta de visualização que permite a análise detalhada de uma ampla gama de variáveis do Enade. Além disso, a ferramenta permite realizar comparações entre múltiplos cursos simultaneamente, por meio de uma interface *Web*, proporcionando uma visão mais detalhada para a tomada de decisões.

2.7 Ferramentas e Tecnologias

Para a elaboração do trabalho foram utilizadas as seguintes ferramentas e tecnologias:

2.7.1 Figma

O Figma [29] é uma plataforma online muito utilizada para criar protótipos de interfaces de usuário. A ferramenta permite a criação de componentes reutilizáveis e a realização de testes interativos, o que facilita a compreensão do funcionamento da interface.

Neste trabalho, o Figma foi usado na fase inicial de prototipagem e também para aprimorar o visual da interface.

2.7.2 Jupyter Notebook

O Jupyter Notebook [30] é muito utilizado em áreas que exigem análise e visualização de dados. Sua capacidade de combinar código, gráficos e documentação em um único ambiente o torna ideal para uma variedade de projetos.

Neste trabalho, a ferramenta foi essencial para a análise preliminar dos dados e para estudo das funcionalidades que compõem a ferramenta. A ferramenta possibilitou uma abordagem flexível para testar e ajustar as análises, facilitando a reprodução dos resultados ao longo do desenvolvimento.

2.7.3 PostgreSQL

O PostgreSQL [31] é um sistema de gerenciamento de banco de dados (SGBD) relacional amplamente reconhecido por sua estabilidade, versatilidade e aderência aos padrões SQL.

No contexto deste trabalho, o PostgreSQL exerceu um papel essencial ao gerenciar de forma eficiente os dados utilizados pela ferramenta. Sua capacidade de lidar com grandes volumes de informações e de realizar consultas complexas com eficácia possibilitou

uma integração fluida com a ferramenta, assegurando que os dados fossem acessados e manipulados com agilidade e precisão durante todo o processo de desenvolvimento.

2.7.4 Python

O desenvolvimento deste trabalho foi realizado inteiramente na linguagem Python [32], versão 3.12.0. Essa foi escolhida por sua extensa coleção de bibliotecas, que oferecem suporte para uma ampla variedade de contextos de projeto.

Python desempenhou um papel fundamental em todas as fases do projeto, desde o tratamento dos dados até a construção da ferramenta. Para o tratamento e análise dos dados, foi utilizado o Pandas [33]. Essa biblioteca é essencial para o processamento de dados em Python e oferece estruturas, como DataFrames, que facilitam a realização de operações complexas e a análise de grandes volumes de dados. Sua flexibilidade em trabalhar com diversos formatos de dados faz dela uma escolha ideal para o projeto.

Além disso, para a construção da ferramenta, foi utilizada principalmente a biblioteca Dash [34]. Essa biblioteca permite a criação de aplicações *Web* interativas com gráficos e visualizações, facilitando a integração de componentes de front-end e back-end de forma intuitiva.

Adicionalmente, foi utilizado o Dash Mantine Component (DMC) [35], uma biblioteca para Python que integra o framework Mantine com o Dash. O DMC facilita a criação de interfaces interativas ao fornecer uma variedade de componentes altamente personalizáveis, desenvolvidos com React, como botões e campos de seleção. Isso elimina a necessidade da construção manual desses elementos, otimizando o processo de desenvolvimento.

O próximo capítulo descreve o processo de preparação dos dados, abrangendo desde a coleta e transformação até a carga dos dados. O objetivo é criar uma estrutura que se ajuste às necessidades da ferramenta e forneça os dados necessários para seu funcionamento.

Capítulo 3

Preparação dos Dados

O trabalho foca na apresentação de uma ferramenta que permite realizar análises comparativas entre cursos, utilizando várias edições do Enade. Para tanto, foi necessário preparar os dados brutos fornecidos pelo Inep [14], criando uma nova estrutura que incorporasse todas as informações essenciais para o correto funcionamento da ferramenta.

Este capítulo em questão detalha toda etapa desde a coleta dos dados até a criação de uma nova estrutura para a base de dados, visando facilitar sua análise e implementação. Para a elaboração da etapa foi utilizado a linguagem Python, com auxílio da ferramenta Jupyter Notebook.

3.1 Fonte dos Dados

A origem dos dados é crucial para assegurar a qualidade e a relevância das informações utilizadas na ferramenta. A seleção das fontes foi realizada priorizando a consistência e a adequação ao escopo do trabalho. Os dados foram cuidadosamente filtrados para garantir que apenas informações pertinentes fossem consideradas.

A seguir, é apresentada uma visão geral das fontes dos dados e a justificativa para sua utilização.

3.1.1 Microdados do Enade

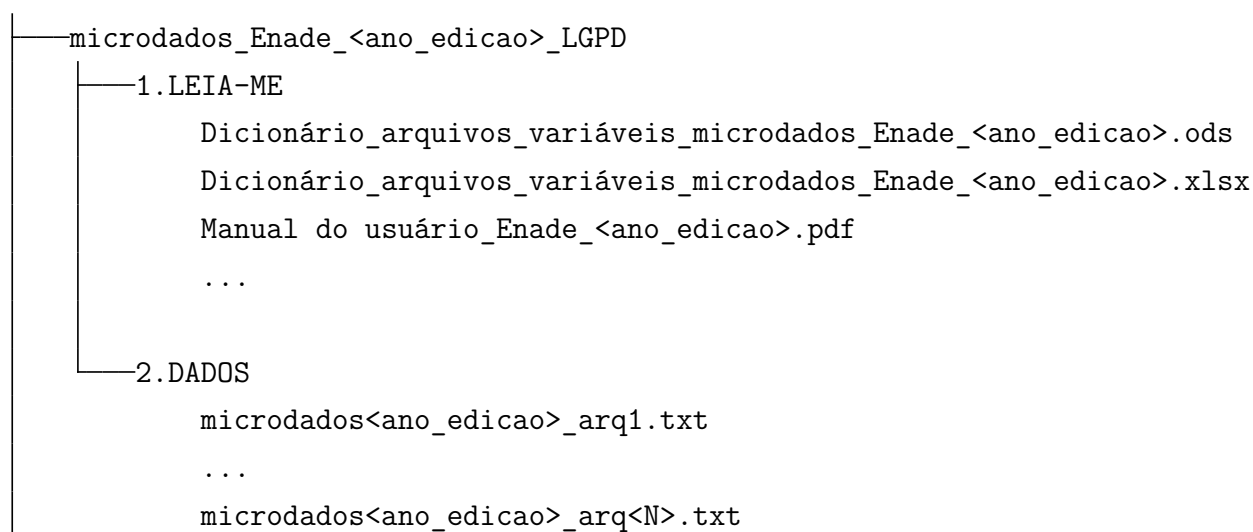
A ferramenta utiliza diversos dados que permitem automatizar a geração das análises. A principal fonte de dados escolhida para este trabalho foi o microdados do Enade, conforme apresentado na Seção 2.1.4.

Os microdados utilizados foram dos anos de 2017, 2018, 2019, 2021 e 2022, todos disponibilizados pelo Inep [14] para *download* gratuito em arquivos zip.

Apesar de corresponderem a anos distintos, todos os microdados possuem um padrão estrutural consistente, permitindo a aplicação do mesmo tratamento a todos.

Os arquivos zip dos microdados estão organizados em duas pastas: “Leia-me” e “Dados”. A primeira contém acesso aos metadados, ao manual do usuário e ao questionário do estudante do ano específico. A segunda, “Dados”, é composta pelos microdados propriamente ditos.

A estrutura de organização dos arquivos segue o mesmo padrão para todos os anos, onde <ano_edicao> corresponde a um dos anos e <N> corresponde à numeração do último microdados:



Por fim, a seguir está a descrição de cada conjunto de microdados utilizado no trabalho:

- Microdados 1: Contém informações sobre a caracterização dos cursos;
- Microdados 3: Inclui vetores, tipos de presença, situação das questões da parte discursiva, notas de Formação Geral e Componente Específico, e respostas do Questionário de Percepção de Prova;
- Microdados 4: Apresenta a avaliação dos estudantes sobre seu processo formativo, abrangendo aspectos como Organização Didático-Pedagógica, Infraestrutura e Oportunidades de ampliação da formação acadêmica e profissional;
- Microdados 5: Fornece informações sobre o sexo dos participantes;
- Microdados 6: Contém a idade dos inscritos no dia da prova;
- Microdados 7 a 32: Cada microdado contém as respostas individuais às questões do questionário do estudante, com foco no perfil do estudante.

3.1.2 Cursos de Graduação do Brasil

Embora os microdados forneçam uma grande quantidade de dados úteis para análise, houve a necessidade de recorrer a uma fonte adicional que contivesse informações sobre os nomes dos cursos e das instituições de ensino superior.

A ausência desses dados nos microdados dificulta a identificação imediata dos cursos e instituições na ferramenta, uma vez que o único atributo disponível que permite essa identificação é o código do curso (`co_ies`).

Para preencher essa lacuna, foi selecionada uma fonte de dados fornecida pelo MEC [36], disponível no arquivo:

- “PDA_Dados_Cursos_Graduacao_Brasil.csv”

Esse arquivo contém informações sobre 902.676 cursos de graduação no Brasil, incluindo atributos como o nome do curso e o nome da instituição. Esses dados são fundamentais para integrar os microdados e permitir a identificação precisa dos cursos na ferramenta, facilitando sua utilização.

3.1.3 Municípios Brasileiros

Para integrar campos de localização geográfica na ferramenta, foi necessário identificar uma fonte de dados que fornecesse os nomes dos municípios brasileiros. Essa necessidade surgiu devido ao fato de que os microdados disponibilizados pelo Inep utilizam exclusivamente códigos de município, conforme o padrão estabelecido pelo Instituto Brasileiro de Geografia e Estatística (IBGE) [37].

O repositório selecionado para suprir essa carência de informações está disponível em [38]. Este repositório foi escolhido por fornecer dados sobre 5.570 municípios e permitir a identificação desses municípios através dos códigos do IBGE. Tal integração possibilitou o estabelecimento de um relacionamento com os microdados do Enade, por meio do atributo referente ao código do município do curso (`co_munic_curso`).

3.1.4 Conceito Enade

O Inep também disponibiliza arquivos que apresentam os 4 indicadores de qualidade da educação superior. Um desses indicadores, de extrema importância para a funcionalidade da ferramenta, se encontra no arquivo referente ao Conceito Enade, conforme detalhado no site [39].

Embora os microdados forneçam as notas brutas de todos os participantes, eles não incluem o cálculo final do Conceito Enade. Por isso, foram extraídos dessa fonte os atri-

butos “Conceito Enade (Contínuo)” e “Conceito Enade (Faixa)”, além do atributo “Sigla da IES”, para complementar a identificação das instituições de ensino na ferramenta.

3.2 Processamento dos Dados

A etapa de processamento dos dados é crucial, pois envolve a limpeza dos dados, a remoção de registros que não se adequam ao contexto do trabalho, e a realização de manipulações para criar uma estrutura adequada ao funcionamento da ferramenta.

A seguir, são apresentadas as etapas de processamento aplicadas para a geração dos arquivos utilizados na implementação.

3.2.1 Informações Gerais do Curso

O primeiro passo envolveu a criação de um arquivo geral que agregasse informações essenciais sobre os cursos em todas as edições do Enade. Para isso, foi utilizado o primeiro e o terceiro microdados de cada edição, com o apoio dos arquivos mencionados em 3.1.2, 3.1.3 e 3.1.4.

O microdado 1 do Enade foi escolhido por conter dados que caracterizam os cursos participantes. Os atributos selecionados para a etapa de processamento foram os seguintes:

- NU_ANO: ano da edição
- CO_CURSO: código do curso
- CO_IES: código da instituição de ensino superior
- CO_CATEGAD: categoria administrativa
- CO_ORGACAD: organização acadêmica
- CO_GRUPO: código do grupo (área de conhecimento)
- CO_MODALIDADE: código da modalidade de ensino
- CO_MUNIC_CURSO: código do município do curso
- CO_UF_CURSO: Unidade Federativa do curso
- CO_REGIAO_CURSO: região do curso

O microdado 3 foi utilizado por conter informações sobre a presença dos participantes durante a prova, essencial para identificar quais cursos deveriam ser incluídos na nova estrutura de dados. Os atributos escolhidos foram:

- TP_PRE: Tipo de presença no Enade
- TP_PR_GER: Tipo de presença na prova

A etapa inicial do processamento exigiu um estudo separado para cada edição, a fim de tratar problemas específicos de cada ano. Assim, foi criado um arquivo .ipynb para cada edição, contendo um breve estudo e o tratamento específico de cada caso.

De modo geral, o processo para cada edição envolveu o uso da biblioteca pandas para criar um arquivo que armazenasse os registros dos atributos apresentados na Figura 3.1. Para possibilitar o relacionamento entre os diferentes arquivos, o microdados 1 foi utilizado como base, por possuir a maior quantidade de atributos de interesse. A partir dele, relacionamentos por meio de atributos-chave permitiram a criação de novas colunas provenientes dos outros arquivos.

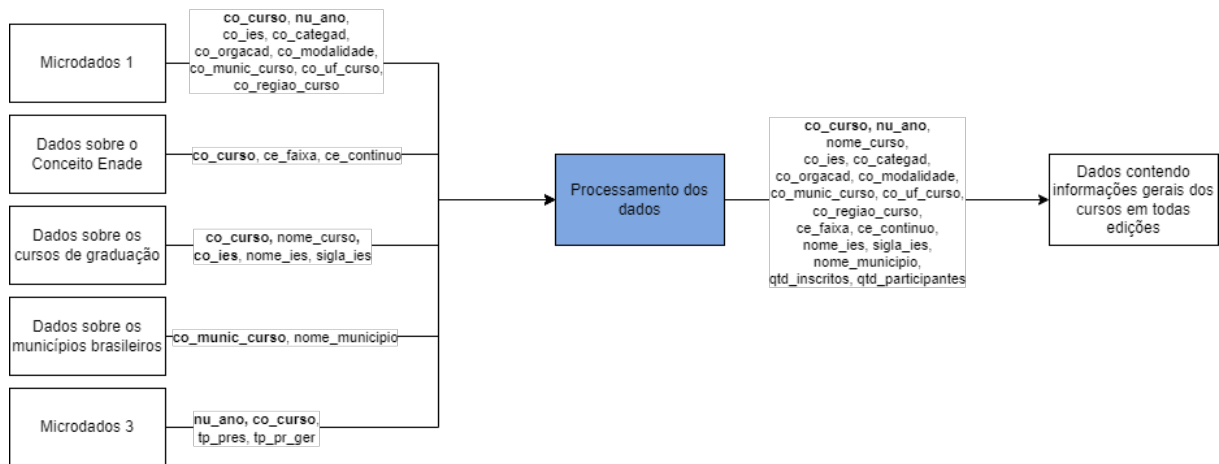


Figura 3.1: Processamento dos atributos referentes as características dos cursos.

Vale ressaltar que, nos microdados, cada registro corresponde a um participante, embora o atributo-chave para todas as tabelas seja o código do curso. Para o trabalho, foram estabelecidas as seguintes condições para determinar a validade de um registro para uso na ferramenta:

- TP_PRE igual a 555: indica participante presente com resultado válido;
- TP_PR_GER igual a 333 ou 555: indica participação com respostas válidas ou participação com prova em branco;
- CE_FAIXA diferente de “SC”: indica que o participante está matriculado em um curso que recebeu uma nota válida no Conceito Enade. Dessa forma, todos os cursos que obtiveram a classificação “Sem Conceito” foram excluídos da nova estrutura de dados.

Após a aplicação dessas condições, a quantidade de registros foi significativamente reduzida e os valores antes e depois do processamento são apresentados na Tabela 3.1. Onde a coluna “Qtd registros (antes)” se refere ao número total de linhas nos microdados antes do processamento, enquanto a coluna “Qtd registros (depois)” indica a quantidade de registros válidos para utilização na ferramenta após a aplicação das condições de filtragem.

Ano Edição	Qtd registros (antes)	Qtd registros (depois)
2017	537358	449661
2018	548122	462095
2019	433930	389735
2021	489866	354853
2022	594013	411518

Tabela 3.1: Quantidade de registros antes e depois por ano de edição

Por fim, como o microdados 1 contém informações detalhadas que caracterizam cada curso, todos os registros associados a um determinado curso apresentavam os mesmos valores. Isso possibilitou a eliminação de registros duplicados, mantendo apenas um registro representativo para cada curso.

Toda etapa de processamento foi aplicado a cada edição do Enade. Como resultado, foram gerados cinco arquivos, cada um contendo informações dos cursos para uma edição específica. Esses arquivos foram posteriormente concatenados para formar um único arquivo geral, que integra as informações dos cursos de todas as edições analisadas.

3.2.2 Ranking de Acertos

Esta etapa de processamento visa criar arquivos que armazenem as opções de respostas dos participantes, em conjunto com o gabarito da prova. Para isso, foi seguido o diagrama de processamento da Figura 3.2.

O microdados utilizado como base para esta etapa foi o terceiro, pois contém os seguintes atributos:

- DS_VT_GAB_OFG_FIN: vetor contendo o gabarito final da Formação Geral;
- DS_VT_GAB_OCE_FIN: vetor contendo o gabarito final do Componente Específico;
- DS_VT_ESC_OFG: vetor contendo as opções de respostas dos participantes na Formação Geral;
- DS_VT_ESC_OCE: vetor contendo as opções de respostas dos participantes no Componente Específico.

Para possibilitar o processamento, foi utilizado um código desenvolvido no formato .ipynb. O primeiro passo consistiu em eliminar registros de participantes que não se enquadram nas duas primeiras condições de validade apresentadas em 3.2.1. Essas condições abrangem a presença dos participantes durante a prova, incluindo tanto a participação efetiva na realização da prova quanto a validação das respostas fornecidas.

Após a redução significativa na quantidade de registros devido à não conformidade com as condições estabelecidas, foi realizada a filtragem apenas das colunas mencionadas acima. Inicialmente, cada coluna dos atributos está no formato de string, onde cada caractere representa uma questão específica. Por exemplo, na *string* “ECBBAACD”, cada caractere equivale à opção correta de uma questão.

Visando criar arquivos de fácil manipulação e utilização, cada caractere dos vetores foi transformado em uma nova coluna. Assim, para o vetor correspondente à Formação Geral, foram criadas 8 colunas para o gabarito e 8 colunas para cada uma das respostas dos participantes. De forma análoga, o mesmo foi feito para o Componente Específico.

Como resultado, foi gerado um arquivo contendo dados sobre a Formação Geral e outro contendo dados sobre o Componente Específico. Esse processo foi aplicado para todas as edições tratadas.

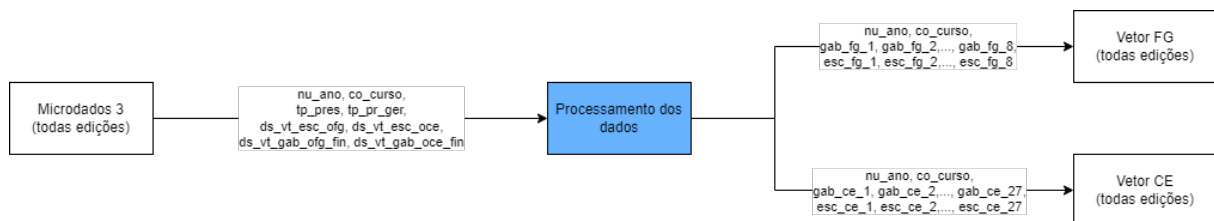


Figura 3.2: Processamento dos atributos referentes ao ranking de acertos.

3.2.3 Distribuição de Notas

A etapa tem como objetivo criar um arquivo específico para cada edição do Enade, contendo as notas brutas dos participantes. O processamento tem como base o microdados 3, com foco nos seguintes atributos:

- NT_GER: corresponde a nota bruta da prova, ou seja, a média ponderada da formação geral (25%) e componente específico (75%);
- NT_FG: corresponde a nota bruta na parte de formação geral, ou seja, a média ponderada da parte objetiva (60%) e discursiva (40%) na formação geral;

- **NT_CE:** corresponde a nota bruta na parte do componente específico, ou seja, a média ponderada da parte objetiva (85%) e discursiva (15%) no componente específico.

Onde todos os valores correspondentes aos atributos mencionados variam de 0 a 100.

O processamento da etapa envolveu a filtragem dos dados para manter apenas as colunas relacionadas às notas brutas, levando em consideração apenas os registros dos participantes cuja presença e participação foram consideradas válidas, segundo os critérios apresentados em 3.2.1. Esse filtro é crucial para garantir que apenas dados relevantes sejam incluídos.

O resultado final é a criação de um novo arquivo que contém apenas as informações sobre as notas dos participantes.

3.2.4 Percepção da Prova

A percepção da prova é um questionário composto por 9 questões, apresentado na última página de cada caderno de prova. As questões são fixas e aplicadas a todas as áreas de conhecimento e edições analisadas neste trabalho.

Essa etapa envolveu a manipulação dos microdados 3, que contém o atributo nomeado **CO_RS_I<N>**, onde **<N>** varia de 1 a 9. Cada atributo possui valores que variam conforme a seguinte categorização: “A”, “B”, “C”, “D”, “E”, “*” e “.”. A descrição de cada atributo e categoria é apresentado a seguir:

- **CO_RS_I1:** 1 - Qual o grau de dificuldade desta prova na parte de Formação Geral?
 - **Opções de resposta:** A = Muito fácil / B = Fácil / C = Médio / D = Difícil / E = Muito difícil
- **CO_RS_I2:** 2 - Qual o grau de dificuldade desta prova na parte do Componente Específico?
 - **Opções de resposta:** A = Muito fácil / B = Fácil / C = Médio / D = Difícil / E = Muito difícil
- **CO_RS_I3:** 3 - Considerando a extensão da prova, em relação ao tempo total, você considera que a prova foi:
 - **Opções de resposta:** A = Muito longa / B = Longa / C = Adequada / D = Curta / E = Muito curta

- **CO_RS_I4:** 4 - Os enunciados das questões da prova na parte de Formação Geral estavam claros e objetivos?
 - **Opções de resposta:** A = Sim, todos / B = Sim, a maioria / C = Apenas cerca da metade / D = Poucos / E = Não, nenhum
- **CO_RS_I5:** 5 - Os enunciados das questões na parte do Componente Específico estavam claros e objetivos?
 - **Opções de resposta:** A = Sim, todos / B = Sim, a maioria / C = Apenas cerca da metade / D = Poucos se apresentam / E = Não, nenhum
- **CO_RS_I6:** 6 - As informações/instruções fornecidas para a resolução das questões foram suficientes para resolvê-las?
 - **Opções de resposta:** A = Sim, até excessivas / B = Sim, em todas elas / C = Sim, na maioria delas / D = Sim, somente em algumas / E = Não, em nenhuma delas
- **CO_RS_I7:** 7 - Você se deparou com alguma dificuldade ao responder à prova. Qual?
 - **Opções de resposta:** A = Desconhecimento do conteúdo / B = Forma diferente de abordagem do conteúdo / C = Espaço insuficiente para responder às questões / D = Falta de motivação para fazer a prova / E = Não tive qualquer tipo de dificuldade para responder à prova
- **CO_RS_I8:** 8 - Considerando apenas as questões objetivas da prova, você percebeu que:
 - **Opções de resposta:** A = Não estudou ainda a maioria desses conteúdos / B = Estudou alguns desses conteúdos, mas não os aprendeu / C = Estudou a maioria desses conteúdos, mas não os aprendeu / D = Estudou e aprendeu muitos desses conteúdos / E = Estudou e aprendeu todos esses conteúdos
- **CO_RS_I9:** 9 - Qual foi o tempo gasto por você para concluir a prova?
 - **Opções de resposta:** A = Menos de uma hora / B = Entre uma e duas horas / C = Entre duas e três horas / D = Entre três e quatro horas / E = Quatro horas e não consegui terminar

Para a criação dos novos arquivos referentes à percepção da prova, foi seguido o processamento descrito na Figura 3.3.

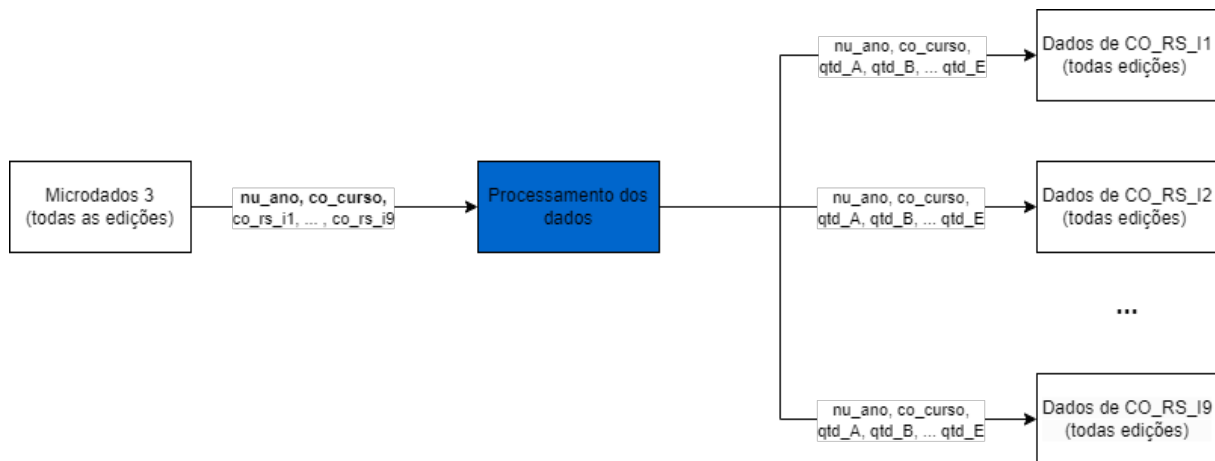


Figura 3.3: Processamento dos atributos referentes à percepção da prova.

Nesta etapa, foi criado um arquivo .ipynb cuja função era contabilizar a quantidade de ocorrências de cada opção de resposta e realizar agrupamentos para cada curso participante. Com isso, foi possível elaborar uma nova estrutura onde os novos atributos `qtd_A`, `qtd_B`, ..., `qtd_E` correspondem à quantidade total de marcações para cada determinada questão.

Dessa forma, cada nova estrutura criada correspondeu a um tipo de questão, totalizando a criação de 9 arquivos no final.

3.2.5 Perfil do Estudante

O questionário do estudante [40] é aplicado em todas as edições do Enade com o intuito de analisar o perfil dos concluintes no contexto da sua formação superior.

Composto por 26 questões de múltipla escolha, o questionário é disponibilizado no sistema do Enade antes da aplicação da prova e é uma etapa obrigatória para a participação na edição.

Para processar os dados referentes a este questionário, foram utilizados os microdados 5 a 32. Onde cada microdado corresponde a uma questão específica do questionário, com exceção dos microdados 5 e 6, que tratam do sexo e da idade dos participantes.

O processamento desses dados seguiu o fluxo descrito na Figura 3.4.

Assim como na etapa descrita em 3.2.4, foram contabilizadas as ocorrências de cada opção de resposta e realizados agrupamentos por código de curso. No entanto, a abordagem precisou ser ajustada devido à diversidade no número de opções de resposta para cada questão. Foi então realizada uma análise para identificar a quantidade de opções de resposta disponíveis para cada questão, a fim de definir os atributos adequados na nova estrutura.

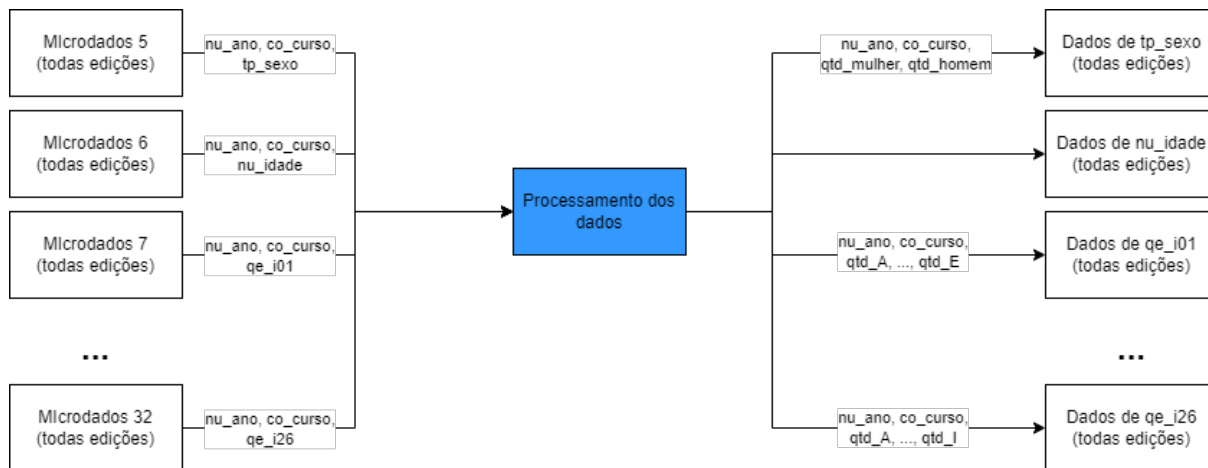


Figura 3.4: Processamento dos atributos referentes ao perfil do estudante.

No final, foram gerados 28 arquivos: 26 correspondendo às questões do questionário e 2 correspondendo ao sexo e idade, onde cada arquivo contém dados de todas as edições analisadas.

3.2.6 Processo de Formação

Para avaliar o processo formativo dos participantes concluintes, o Inep disponibiliza, através do questionário do estudante, 41 questões que abordam a organização didático-pedagógica, a infraestrutura e as oportunidades de ampliação da formação acadêmica e profissional.

Essas questões são numeradas de 27 a 68 e possuem opções de resposta em uma escala que varia de 1 (discordância total) a 6 (concordância total). A Figura 3.5 apresenta a estrutura da primeira questão.

A seguir, leia cuidadosamente cada assertiva e indique seu grau de concordância com cada uma delas, segundo a **escala** que varia de **1 (discordância total)** a **6 (concordância total)**. Caso você julgue não ter elementos para avaliar a assertiva, assinale a opção “Não sei responder” e, quando considerar não pertinente ao seu curso, assinale “Não se aplica”.

ORGANIZAÇÃO DIDÁTICO-PEDAGÓGICA/INFRAESTRUTURA E INSTALAÇÕES FÍSICAS/OPORTUNIDADES DE AMPLIAÇÃO DA FORMAÇÃO ACADÊMICA E PROFISSIONAL	1 <input type="radio"/> Discordo Totalmente	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/> Concordo Totalmente	
27.As disciplinas cursadas contribuíram para sua formação integral, como cidadão e profissional.	1 <input type="radio"/>	2 <input type="radio"/>	3 <input type="radio"/>	4 <input type="radio"/>	5 <input type="radio"/>	6 <input type="radio"/>	() Não sei responder () Não se aplica

Figura 3.5: Primeira questão do processo de formação.

Para o processamento dos dados vinculados a essa parte do questionário, foram utilizados os microdados 4 de todas as edições. Cada atributo com o prefixo “QE_I” corresponde às respostas de uma questão específica. Por exemplo, o atributo “QE_I27” refere-se à questão 27 do questionário.

O objetivo final desta etapa foi gerar arquivos únicos para cada questão, incluindo informações de todas as edições tratadas no trabalho. A representação do processo é apresentada na Figura 3.6.

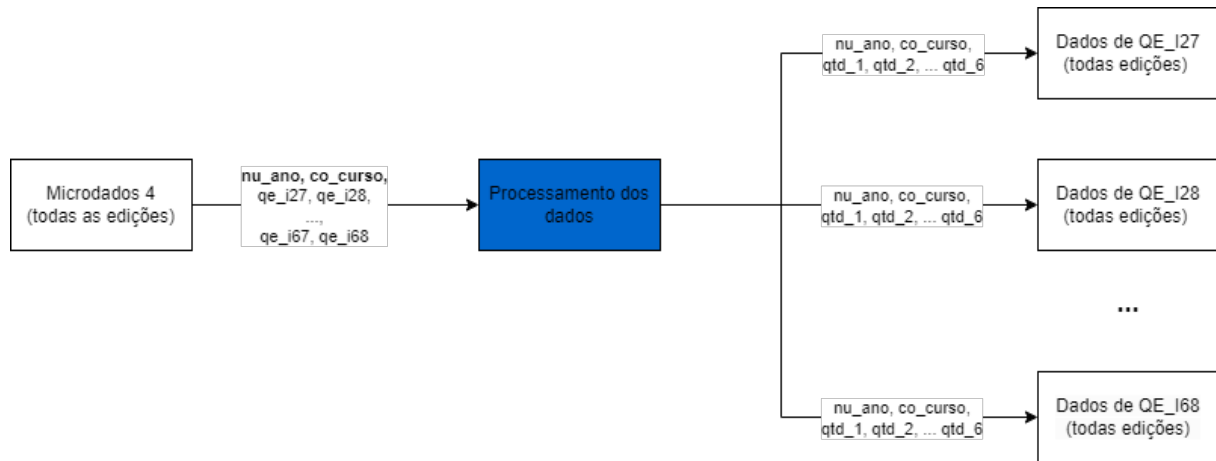


Figura 3.6: Processamento dos atributos referentes ao processo de formação.

Para realizar esse processamento, foi utilizado um código armazenado em um arquivo `.ipynb`. O código inicia lendo os microdados 4 das edições e os armazenando em DataFrames, garantindo que as colunas das questões sejam interpretadas como strings. Em seguida, para cada questão e cada edição, é extraído as colunas relevantes (`nu_ano`, `co_curso` e a questão específica) e organizado em DataFrames separados.

O código define uma função que agrupa esses dados por curso e pelas respostas das questões, contando a quantidade de respostas para cada categoria e garantindo que todas as categorias estejam representadas. Por fim, são gerados arquivos `.csv` que consolidam esses dados para todas as edições, fornecendo um panorama geral das respostas dos estudantes ao longo dos anos.

3.3 Migração dos Dados para o PostgreSQL

A etapa final da preparação dos dados envolveu a migração para o gerenciador PostgreSQL, com o objetivo de assegurar uma maior organização e um acesso rápido às informações.

O processo de migração foi iniciado com a criação de uma estrutura composta por diversas tabelas, cada uma correspondendo a cada um dos arquivos gerados descritos na Seção 3.2. Para isso, foi desenvolvido um código em Python que realiza a leitura de cada arquivo `.csv` e o transforma em um DataFrame. Em seguida, os nomes das colunas e seus tipos foram extraídos.

A criação das tabelas no PostgreSQL foi então realizada com a ajuda da biblioteca SQLAlchemy, que facilitou a definição da estrutura do banco de dados.

Por fim, os dados dos arquivos .csv foram carregados para o banco de dados recém-criado, completando assim a etapa de migração, permitindo a execução das etapas posteriores do trabalho.

O próximo capítulo fornece detalhes sobre a estrutura da ferramenta, com base nos requisitos estabelecidos. Além disso, é apresentado um fluxo que ilustra a lógica de utilização implementada.

Capítulo 4

Estrutura da Ferramenta Enade.Vis

A ferramenta criada neste trabalho é projetada para gerar análises comparativas entre um curso específico e um grupo de cursos dentro da mesma área de conhecimento.

Com o objetivo de agrupar cursos relacionados, a ferramenta foi estruturada em quatro telas bem definidas, conforme apresentado no fluxo de utilização da Figura 4.1. Essas telas são preenchidas de maneira sequencial, resultando na geração de uma análise específica.

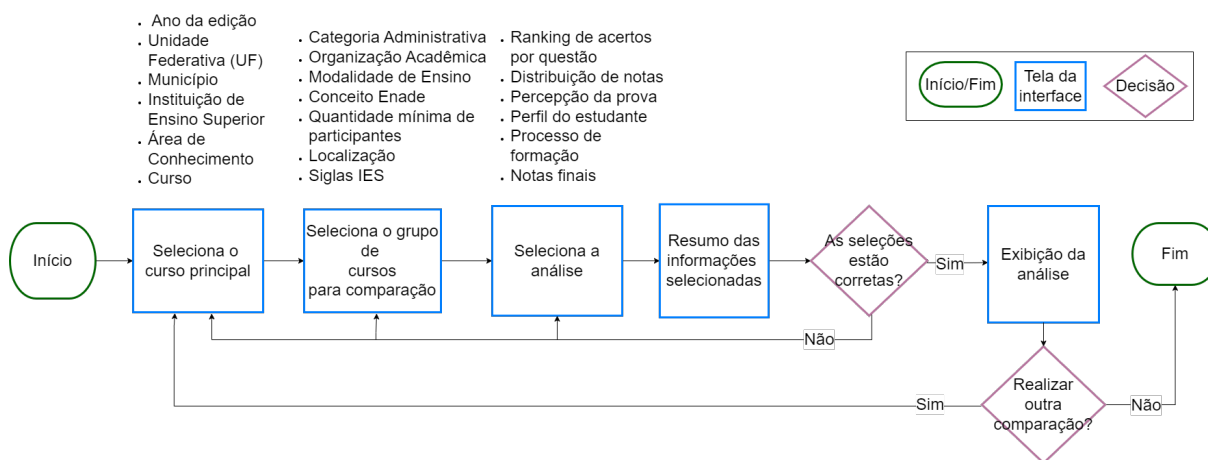


Figura 4.1: Fluxo de utilização.

Para assegurar que a implementação atendesse às expectativas do trabalho e funcionasse conforme o necessário, foram estabelecidos alguns requisitos funcionais (RF) e requisitos não funcionais (RNF). Os requisitos funcionais definiram as funcionalidades essenciais da ferramenta, enquanto os requisitos não funcionais abordaram aspectos relacionados a usabilidade. A Tabela 4.1 apresenta as descrições detalhadas deles.

Além disso, a ferramenta foi desenvolvida levando em consideração os seguintes perfis de usuários:

- Coordenador de Curso: para visualizar o desempenho do curso em comparação com outras instituições de ensino superior e identificar áreas para melhorias;
- Professor: para analisar tendências de resposta em questões específicas que estão relacionadas aos conteúdos abordados em suas disciplinas;
- Direção de Ensino de Graduação: para otimizar a estrutura curricular dos cursos, visando aprimorar os desempenhos em avaliações futuras;
- Gestores Institucionais: para tomar decisões estratégicas, como alocação de recursos e investimentos em infraestrutura educacional;
- Pesquisadores em Educação: para explorar dados e realizar estudos comparativos, identificando padrões e tendências no desempenho dos estudantes.

Descrição do requisito	Tipo
Implementação da interface como uma aplicação web	RNF
Inclusão de campos de ajuda para explicar o funcionamento da interface	RNF
Destaque visual dos filtros aplicados	RNF
Manter a opção "todos" marcada como padrão em todos os filtros	RNF
Adicionar ícones de informação em cada campo da ferramenta	RNF
Garantir que a interface seja intuitiva e fácil de usar	RNF
Permitir uma comparação abrangente entre diversos cursos	RF
Utilizar filtros avançados (como seleção de universidades públicas, quantidade mínima de participantes, etc.)	RF
Abranger todas as análises a partir dos microdados	RF
Possibilitar a seleção simultânea de cursos	RF
Gerar análises com foco no desempenho dos cursos	RF
Incluir um link para o caderno de provas da área selecionada	RF

Tabela 4.1: Requisitos funcionais e não funcionais.

Este capítulo descreve em detalhes a estrutura final da implementação, levando em consideração os requisitos estabelecidos. Além disso, justifica a necessidade de cada campo criado, explicando a contribuição para o objetivo do trabalho.

4.1 Curso Principal

O primeiro passo, apresentado na Figura 4.2, tem como objetivo permitir a escolha do curso principal e definir o contexto no qual a análise será realizada.

Este passo consiste em cinco campos obrigatórios que, em conjunto, reduzem a quantidade de cursos apresentados na listagem do campo “curso”. Dessa forma, a funcionalidade permite encontrar o curso principal, mesmo que o nome exato seja desconhecido.

A imagem mostra a interface do sistema ENADE.VIS, especificamente a etapa 1: Curso principal. O cabeçalho da interface é azul escuro com o texto "ENADE.VIS" em branco. Abaixo do cabeçalho, há uma barra de progresso com quatro etapas numeradas: 1. Passo 1 (Curso principal), 2. Passo 2 (Cursos para comparação), 3. Passo 3 (Análises disponíveis) e 4. Passo 4 (Geração da análise). A etapa 1 está selecionada. O formulário principal contém os seguintes campos:

- Edição ***: Campo de seleção com o texto "Selecione..." e uma seta para baixo.
- UF ***: Campo de seleção com o texto "Selecione..." e uma seta para baixo.
- Município**: Campo de seleção com o texto "Selecione..." e uma seta para baixo.
- Sua Instituição de Ensino ***: Campo de seleção com o texto "Selecione..." e uma seta para baixo.
- Área de conhecimento ***: Campo de seleção com o texto "Selecione..." e uma seta para baixo.
- Curso ***: Campo de seleção com o texto "Selecione..." e uma seta para baixo.

Figura 4.2: Tela de seleção do curso principal.

Para a seleção do curso, a ferramenta também conta com um campo opcional denominado “município”, que torna a filtragem ainda mais eficiente.

Cada campo é preenchido de maneira sequencial. E para garantir a execução correta, a ferramenta apresenta a listagem dos campos subsequentes apenas quando os anteriores já foram preenchidos. A seguir, é apresentada a descrição de cada campo do primeiro passo:

1. Edição: Disponibiliza cinco edições, sendo responsável por filtrar os dados referentes ao ano escolhido;
2. UF: Apresenta todas as unidades federativas, incluindo o Distrito Federal, onde tiveram instituições que realizaram o Enade no ano selecionado;
3. Município (opcional): Apresenta apenas os municípios ativos que tiveram a aplicação do Enade no ano e UF selecionados;
4. Sua Instituição de Ensino: Apresenta todas as instituições de ensino superior que participaram do Enade e que atendem às seleções dos campos anteriores;

5. Área de conhecimento: Apresenta todas as áreas de conhecimento nas quais a instituição escolhida participou no ano selecionado;
6. Curso: Exibe todos os cursos que participaram do Enade, considerando as seleções feitas nos campos anteriores.

4.2 Cursos para Comparação

O segundo passo visa apresentar todos os cursos disponíveis para a análise comparativa com o curso principal escolhido em 4.1. Esta etapa é composta por três componentes, visualizados na Figura 4.3.

ENADE.VIS

Passo 1 Curso principal

Passo 2 Cursos para comparação

Passo 3 Análises disponíveis

Passo 4 Geração da análise

Cursos para comparação ⓘ

Categoria Administrativa Organização Acadêmica Modalidade de Ensino Conceito Enade Qtd mínima de participantes Localização Lista IES APLICAR

Cursos encontrados: 318

Cursos selecionados para comparação: 0

Digite o nome do curso/instituição... > >>

- UFMT - CIÊNCIA DA COMPUTAÇÃO (36)
- UFS - CIÊNCIA DA COMPUTAÇÃO (317)
- UFPI - CIÊNCIA DA COMPUTAÇÃO (513)
- UFOP - CIÊNCIA DA COMPUTAÇÃO (594)

<< < Digite o nome do curso/instituição... Nenhum curso para remover

Figura 4.3: Tela de seleção de cursos para comparação.

A ferramenta filtra todos os cursos que se enquadram no ano da edição e na área de conhecimento selecionados anteriormente, excluindo o curso principal da listagem.

Para otimizar a busca pelos cursos de interesse, são disponibilizados os seguintes filtros:

- Categoria Administrativa: Permite filtrar entre instituições públicas e/ou privadas;
- Organização Acadêmica: Permite filtrar entre faculdade, centro universitário, centro federal, instituto federal e/ou universidade;
- Modalidade de Ensino: Permite filtrar entre ensino a distância (EaD) e/ou presencial;
- Conceito Enade: Permite filtrar pelas notas no conceito Enade, que variam de 1 a 5;

- Quantidade Mínima de Participantes: Permite filtrar após o preenchimento de um valor numérico inteiro que define a quantidade mínima de participantes que o curso deve ter;
- Localização: Permite filtrar os estados onde o curso está localizado;
- Lista IES: Permite realizar a filtragem a partir de uma lista de nomes e siglas de instituições de ensino superior.

Com o objetivo de garantir consistência na apresentação dos nomes, todos os cursos são exibidos seguindo o seguinte padrão de nomenclatura:

`<sigla_ies>|<nome_ies> <nome_curso> (<codigo_curso>)`

Onde a sigla da instituição possui prioridade na apresentação, e nos casos em que a instituição não possui uma sigla, o nome completo é exibido.

Essa padronização é fundamental para diferenciar cursos que possuem o mesmo nome e são oferecidos pela mesma instituição. Um exemplo é o caso dos cursos de Artes Visuais da UnB, que são diferenciados apenas pelo código de cada um. A Figura 4.4 ilustra esse caso.

Para a seleção dos cursos de interesse, a ferramenta é dividida em dois campos. O primeiro, nomeado “Cursos encontrados”, apresenta todos os cursos disponíveis para seleção, cuja quantidade varia de acordo com a utilização dos filtros.

O segundo campo, “Cursos selecionados para comparação”, armazena os cursos que farão parte da análise comparativa.

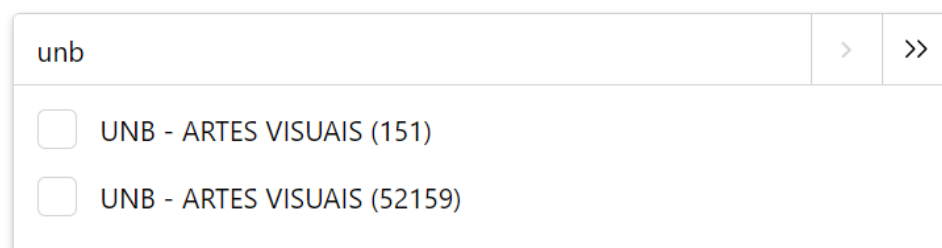


Figura 4.4: Exemplo de cursos que possuem nomes iguais.

4.3 Análises Disponíveis

O terceiro passo, apresentado pela Figura 4.5 consiste da descrição das análises disponíveis para visualização na ferramenta. Essa etapa é apresentada dentro de uma tabela que contém três colunas, sendo essas:

- Escolha: permite a seleção da análise de interesse;
- Nome da análise: define a análise disponível;
- Descrição: apresenta um resumo sobre a análise disponível.

ESCOLHA	NOME DA ANÁLISE	DESCRIÇÃO
<input type="checkbox"/>	Ranking de acertos por questão	Apresenta o desempenho do curso principal em relação aos cursos selecionados para comparação em cada questão do caderno de provas. Essa análise detalha quais questões foram mais ou menos acertadas pelos participantes, fornecendo uma visão clara das áreas onde o curso se destaca ou precisa melhorar.
<input type="checkbox"/>	Distribuição de notas	Foca na análise detalhada da distribuição das notas brutas dos participantes, incluindo métricas estatísticas essenciais como média, mediana e desvio padrão. Isso proporciona uma compreensão mais completa do desempenho dos alunos.
<input type="checkbox"/>	Percepção da prova	Aborda a análise das respostas dos alunos ao questionário que avalia a percepção deles sobre a prova, incluindo aspectos como dificuldade, relevância, e clareza das questões. Essa análise é fundamental para entender a experiência dos alunos durante a prova e identificar possíveis melhorias na formulação das questões e no formato da prova.
<input type="checkbox"/>	Perfil do estudante	Analisa as características dos participantes, como idade, gênero, renda familiar, escolaridade dos pais, entre outras. Essa análise ajuda a compreender a diversidade do corpo discente e a identificar fatores que podem influenciar o desempenho acadêmico.
<input type="checkbox"/>	Processo de formação	Examina o percurso acadêmico dos participantes, incluindo informações sobre disciplinas cursadas, desempenho ao longo do curso, entre outros. Essa análise é essencial para avaliar a eficácia do processo educativo e identificar possíveis intervenções para melhorar a formação dos alunos.
<input type="checkbox"/>	Notas finais	Tem o objetivo de fornecer uma visão geral das notas finais obtidas pelos participantes, abrangendo tanto a nota geral quanto as notas específicas na Formação Geral e no Componente Específico.

Figura 4.5: Tela de seleção da análise.

A ferramenta dispõe de todas as análises possíveis de serem realizadas através do uso dos microdados.

Essas análises são:

- *Ranking* de acertos por questão: Tem como objetivo apresentar o desempenho da instituição principal em relação ao grupo de instituições selecionado para comparação. Essa análise permite identificar quais questões foram mais ou menos acertadas pelos participantes, oferecendo uma visão detalhada das questões onde a instituição se destaca ou precisa melhorar;
- Notas finais: Tem o objetivo de fornecer uma visão geral das notas finais obtidas pelos participantes, abrangendo tanto a nota geral quanto as notas específicas na Formação Geral e no Componente Específico.
- Distribuição de notas: Foca na análise detalhada da distribuição das notas brutas dos participantes. Além disso, apresenta métricas estatísticas essenciais, como mé-

dia, mediana e desvio padrão, para uma compreensão mais completa do desempenho dos alunos.

- **Percepção da prova:** Aborda a análise das respostas dos alunos ao questionário que avalia a percepção deles sobre a prova, incluindo aspectos como dificuldade, relevância, e clareza das questões. Essa análise é fundamental para entender a experiência dos alunos durante a prova e identificar possíveis melhorias na formulação das questões e no formato da prova;
- **Perfil do estudante:** Analisa as características demográficas e socioeconômicas dos participantes, como idade, gênero, renda familiar, escolaridade dos pais, entre outras. Essa análise ajuda a compreender a diversidade do corpo discente e a identificar fatores que podem influenciar o desempenho acadêmico;
- **Processo de formação:** Examina o percurso acadêmico dos participantes, incluindo informações sobre disciplinas cursadas, desempenho ao longo do curso, entre outros. Essa análise é essencial para avaliar a eficácia do processo educativo e identificar possíveis intervenções para melhorar a formação dos alunos.

4.4 Resumo da Seleção

Após a conclusão dos passos anteriores, a ferramenta exibe uma tela de resumo, ilustrada na Figura 4.6. Essa tela consolida todas as informações selecionadas até o momento, permitindo uma revisão detalhada das escolhas feitas.

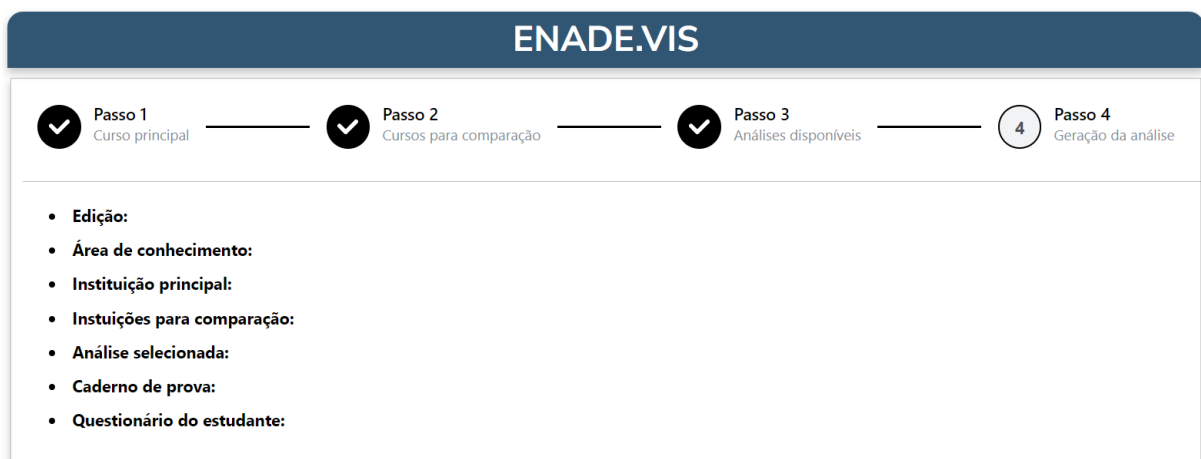


Figura 4.6: Tela de resumo das informações selecionadas.

Além de oferecer uma visão geral dos dados filtrados e selecionados, a tela de resumo também disponibiliza *links* úteis, como o acesso ao caderno de provas para a área de conhecimento e ano selecionados, assim como ao questionário do estudante.

Esses recursos adicionais são fundamentais para uma interpretação mais precisa e contextualizada das análises geradas pela ferramenta.

O próximo capítulo apresenta exemplos de utilização da ferramenta que não são cobertos pelos relatórios de cursos do Inep ou pelas ferramentas discutidas na Seção 2.6. Além disso, detalha as estruturas das análises geradas pela ferramenta.

Capítulo 5

Exemplos de Utilização

Atualmente, o relatório disponibilizado pelo Inep sobre o desempenho dos cursos no Enade apresenta diversas análises focadas em cada curso e permite comparações com grupos específicos, com base nas médias gerais. Por exemplo, o relatório referente ao curso de Ciência da Computação (Bacharelado) da UnB na edição de 2021 permite comparações com as médias da Unidade Federativa, região, categoria administrativa e organização acadêmica, além da média nacional.

A Figura 5.1 apresenta a primeira análise do relatório, que aborda as notas no resultado geral, na Formação Geral e no Componente Específico. Embora essa análise ofereça informações relevantes, ela apresenta limitações, pois permite apenas uma percepção superficial de como o desempenho do curso da UnB se posiciona em relação às médias dos grupos apresentados, sem um nível a mais de detalhamento.

CIÊNCIA DA COMPUTAÇÃO (BACHARELADO)							
ENADE		Curso	UF	Região	Cat. Adm.	Org. Acad.	Brasil
Tamanho da população		112	263	777	5745	9395	12738
Número de presentes		85	218	631	4661	7486	10119
Resultado Geral	Média	48,9	41,9	38,4	43,0	38,7	36,9
	Erro padrão da média	1,5	0,9	0,5	0,2	0,2	0,1
	Desvio padrão	13,6	13,4	13,2	13,8	14,2	13,8
	Mediana	49,7	41,8	37,2	42,6	37,8	35,7
	Mínimo	19,2	11,1	9,2	0,0	0,0	0,0
	Máximo	78,5	78,5	78,5	87,4	87,4	87,4
	Coefficiente de Assimetria	-0,2	0,2	0,4	0,1	0,3	0,4
Formação Geral	Média	45,1	42,4	39,6	41,8	38,6	37,6
	Erro padrão da média	1,4	0,9	0,6	0,2	0,2	0,2
	Desvio padrão	12,6	13,9	14,7	14,7	15,1	15,0
	Mediana	45,9	41,5	38,9	42,3	38,3	37,5
	Mínimo	15,0	4,6	0,0	0,0	0,0	0,0
	Máximo	72,5	75,1	77,3	88,6	88,6	88,6
	Coefficiente de Assimetria	-0,1	0,0	0,1	-0,1	-0,0	0,0
Comp. Específico	Média	50,2	41,7	38,0	43,3	38,7	36,7
	Erro padrão da média	1,7	1,1	0,6	0,2	0,2	0,2
	Desvio padrão	15,8	15,6	15,2	16,0	16,2	15,8
	Mediana	52,0	40,5	36,0	42,0	37,0	35,8
	Mínimo	14,2	4,5	4,5	0,0	0,0	0,0
	Máximo	86,1	86,1	89,2	95,5	95,5	95,5
	Coefficiente de Assimetria	-0,2	0,2	0,5	0,2	0,4	0,5

Figura 5.1: Desempenho geral do curso de Ciência da Computação da UnB, baseado em [1].

Diante disso, este capítulo apresenta como exemplo alguns casos interessantes para a ferramenta desenvolvida, focando no curso de Ciência da Computação (Bacharelado) da UnB, ao mesmo tempo em que realiza a interpretação das análises geradas. Esses casos foram selecionados para superar as limitações do relatório do Enade e demonstrar as capacidades da ferramenta em fornecer análises detalhadas a nível de curso, permitindo uma interpretação mais detalhada dos dados.

Os exemplos apresentados são:

- Instituições Pontifícia Universidade Católica (PUC): O objetivo é comparar o desempenho da UnB com todas as PUCs do Brasil, utilizando filtragem por palavra-chave para facilitar a seleção das instituições;
- Universidades Públicas com Conceito Enade 5: Apresenta os filtros utilizados para selecionar todas as instituições públicas de ensino superior que obtiveram Conceito Enade 5;
- Instituições de Ensino Superior do Distrito Federal (DF): Foca em mostrar a aplicação do filtro de localização geográfica para selecionar todas as instituições de ensino superior localizadas no Distrito Federal;
- Lista de Instituições: Demonstra a capacidade da ferramenta de retornar cursos com base em uma lista pré-definida de instituições.

Os exemplos são baseados na edição de 2021, que foi o último ano em que ocorreu a aplicação da prova para a área na qual o curso de Ciência da Computação (Bacharelado) se enquadra.

A seguir, são descritos os passos para preencher os campos da ferramenta com o objetivo de gerar os três exemplos.

Para o primeiro passo de todos os exemplos, foram selecionados os valores apresentados na Figura 5.2, onde o código 127 corresponde ao curso de Ciência da Computação (Bacharelado) da UnB, conforme o cadastro no e-MEC [41].

5.1 Instituições Classificadas como PUC

Esta seção apresenta um exemplo que tem como objetivo comparar o desempenho do curso de Ciência da Computação (Bacharelado) da UnB com o de todas as instituições de ensino superior classificadas como Pontifícia Universidade Católica (PUC), análise que não está disponível no relatório do Enade. Este exemplo é particularmente relevante para demonstrar a aplicação dos filtros em um cenário onde o critério principal de filtragem é a presença da palavra “PUC” no nome da instituição.

1 Passo 1
Curso principal

2 Passo 2
Cursos para comparação

3 Passo 3
Análises disponíveis

4 Passo 4
Geração da análise

Edição* ⓘ
2021

UF* ⓘ
Distrito Federal (DF)

Município ⓘ
Brasília

Sua Instituição de Ensino* ⓘ
UNIVERSIDADE DE BRASÍLIA (UNB)

Área de conhecimento* ⓘ
Ciência Da Computação (Bacharelado)

Curso* ⓘ
UNB - CIÊNCIA DA COMPUTAÇÃO (127)

Figura 5.2: Seleção do curso de Ciência da Computação (Bacharel) da UnB.

No primeiro passo, foram utilizadas as seleções referentes ao curso principal, conforme mostrado na Figura 5.2. Em seguida, para selecionar o grupo de instituições PUC, utilizou-se o campo de digitação para inserir a palavra-chave “PUC”, como ilustrado no campo à esquerda da Figura 5.3. Após aplicar o filtro, as instituições que corresponderam à busca foram selecionadas, resultando na lista exibida no campo à direita da Figura 5.3, que contém todas as 7 instituições PUC do Brasil que realizaram a prova do Enade.

Cursos encontrados: 318

puc

- PUCPR - CIÊNCIA DA COMPUTAÇÃO (865)
- PUCRS - CIÊNCIA DA COMPUTAÇÃO (1904)
- PUC MINAS - CIÊNCIA DA COMPUTAÇÃO (7511)
- PUC GOIÁS - CIÊNCIA DA COMPUTAÇÃO (10814)

Cursos selecionados para comparação: 7

<< < Digite o nome do curso/instituição...

- PUCPR - CIÊNCIA DA COMPUTAÇÃO (865)
- PUCRS - CIÊNCIA DA COMPUTAÇÃO (1904)
- PUC MINAS - CIÊNCIA DA COMPUTAÇÃO (7511)
- PUC GOIÁS - CIÊNCIA DA COMPUTAÇÃO (10814)

Figura 5.3: Seleção das instituições PUC.

5.1.1 Análise do Ranking de Acertos por Questão

O ranking de acertos tem como objetivo fornecer uma análise detalhada das 35 questões do caderno de provas, incluindo a identificação das questões anuladas. Para demonstrar o tipo de análise gerada pela ferramenta, esta seção foca na questão 9, a primeira do Componente Específico, realizando uma comparação entre o curso de Ciência da Computação da UnB e as instituições PUC.

As tabelas na Figura 5.4 e Figura 5.5 mostram duas formas de visualização da análise do *ranking* de acertos por questão: a primeira detalha todas as opções de resposta, enquanto a segunda se concentra apenas na quantidade de acertos da questão. Além disso, o total de marcações para cada curso também é exibido.

Questão: 9 (Componente Específico) Tipo de visualização: Detalhada PESQUISAR

Ranking de acertos - Questão 9 (Visualização Detalhada)

POSIÇÃO	CÓDIGO DO CURSO	INSTITUIÇÃO	OPÇÃO							TOTAL DE MARCAÇÕES
			A	B	C	D	E	EM BRANCO	VÁRIAS	
1º	10814	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS	11.36%	6.82%	15.91%	22.73%	43.18%	0.0%	0.0%	44
2º	18585	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS	17.14%	2.86%	11.43%	25.71%	42.86%	0.0%	0.0%	35
3º	1904	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL	8.0%	14.0%	8.0%	30.0%	40.0%	0.0%	0.0%	50
4º	114286	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO	5.66%	9.43%	18.87%	30.19%	35.85%	0.0%	0.0%	53
5º	7511	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS	15.0%	7.5%	2.5%	41.25%	33.75%	0.0%	0.0%	80
6º	127	UNIVERSIDADE DE BRASÍLIA	2.35%	5.88%	12.94%	45.88%	32.94%	0.0%	0.0%	85
7º	11238	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO	22.5%	12.5%	17.5%	17.5%	30.0%	0.0%	0.0%	40
8º	865	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ	11.11%	5.56%	8.33%	50.0%	25.0%	0.0%	0.0%	36

Figura 5.4: Ranking de acertos da questão 9 (UnB x PUC's) - Visualização Detalhada.

Questão: 9 (Componente Específico) Tipo de visualização: Geral PESQUISAR

Ranking de acertos - Questão 9 (Visualização Geral)

POSIÇÃO	CÓDIGO DO CURSO	INSTITUIÇÃO	ACERTOS	TOTAL DE MARCAÇÕES
1º	10814	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS	43.18%	44
2º	18585	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS	42.86%	35
3º	1904	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL	40.0%	50
4º	114286	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO	35.85%	53
5º	7511	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS	33.75%	80
6º	127	UNIVERSIDADE DE BRASÍLIA	32.94%	85
7º	11238	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO	30.0%	40
8º	865	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ	25.0%	36

Figura 5.5: Ranking de acertos da questão 9 (UnB x PUC's) - Visualização Geral.

De acordo com a tabela da Figura 5.4, observa-se que a UnB teve um desempenho abaixo do esperado em relação as instituições PUC, ocupando a 6ª posição, com 32,94% dos participantes selecionando a resposta correta, letra “E”. Por outro lado, a maioria dos alunos escolheu a opção “D”, o que pode ser um ponto de interesse para estudos futuros, a fim de compreender esse comportamento.

QUESTÃO 09

Quando um computador é multiprogramado, ele geralmente tem múltiplos processos ou *threads* que competem pela CPU ao mesmo tempo. Essa situação ocorre sempre que dois ou mais processos estão simultaneamente no estado pronto. Se somente uma CPU se encontrar disponível, deverá ser feita uma escolha de qual processo executar em seguida. A parte do sistema operacional que faz a escolha é chamada de **escalador**, e o algoritmo que ele usa é o **algoritmo de escalonamento**.

TANENBAUM, A. S. *Sistemas Operacionais Modernos*. 3. ed., São Paulo: Pearson, 2010 (adaptado).

Considerando que em ambientes diferentes são necessários algoritmos diferentes de escalonamento, garantindo assim que seja maximizado o uso de seus recursos, assinale a opção que apresenta um algoritmo de escalonamento seguido do tipo de ambiente no qual deva ser implementado.

- Ⓐ Primeiro a chegar, último a sair (*first in, last out* - FILO); propício para sistemas de tempo real.
- Ⓑ Escalonamento por taxas monotônicas (*rate monotonic scheduling* - RMS); propício para sistemas em lote.
- Ⓒ Tarefa mais curta primeiro; propício para sistemas interativos.
- Ⓓ Escalonamento por chave circular (*round-robin*); propício para sistemas de tempo real.
- Ⓔ Escalonamento por prioridades; propício para sistemas interativos.

Figura 5.6: Questão 9 do caderno de provas de Ciência da Computação 2021, baseado em [2].

Para uma melhor interpretação a ferramenta fornece um link para o caderno de provas, permitindo uma consulta rápida à questão em análise. No caso da questão 9, trata-se de um item relacionado a conceitos de algoritmos de escalonamento, abordados na disciplina de Fundamentos de Sistemas Operacionais na UnB, conforme apresentado na Figura 5.6.

5.1.2 Análise das Notas Finais

A análise das notas finais tem como objetivo gerar uma tabela em formato de *ranking*, contendo as médias alcançadas por cada curso selecionado nas áreas de Formação Geral, Componente Específico e na nota Geral. A classificação é estabelecida com base na pontuação final de cada curso, permitindo uma comparação clara do desempenho entre eles.

De acordo com a análise comparativa apresentada na Figura 5.7, o curso de Ciência da Computação (Bacharelado) da UnB destacou-se em relação aos cursos das instituições PUC, alcançando a primeira posição. A UnB obteve uma nota final pouco mais de 2 pontos superior à da PUCRS, que ficou em segundo lugar.

Essa classificação também permite avaliar o desempenho geral das PUCs, evidenciando que o curso oferecido pela PUC Minas, identificado pelo código 18585, apresentou o pior desempenho dentro do grupo, com uma nota final de 33,54.

Notas Finais

POSIÇÃO	CÓDIGO DO CURSO	INSTITUIÇÃO	NOTA FINAL	NOTA FG	NOTA CE
1	127	UNIVERSIDADE DE BRASÍLIA	48.95	45.09	50.22
2	1904	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL	46.01	42.12	47.29
3	114286	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO	43.55	41.7	44.16
4	7511	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS	41.43	40.84	41.61
5	865	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ	38.28	39.47	37.85
6	10814	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS	36.51	41.15	34.94
7	11238	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO	34.46	35.31	34.17
8	18585	PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS	33.54	37.21	32.3

Figura 5.7: Notas finais (UnB x PUC's).

5.1.3 Análise da Distribuição de Notas

A análise da distribuição de notas utiliza histogramas para visualizar as notas dos participantes em uma escala de 0 a 100, com cada barra representando um intervalo de 5 pontos. O eixo y é padronizado para garantir que os dois histogramas comparativos compartilhem o mesmo intervalo numérico, facilitando a comparação entre diferentes cursos.

A ferramenta permite a comparação direta entre dois cursos, oferecendo uma visualização clara e intuitiva. A Figura 5.8 ilustra a estrutura dessa apresentação, onde o histograma à esquerda representa o curso principal, enquanto o da direita exibe o curso selecionado para comparação.

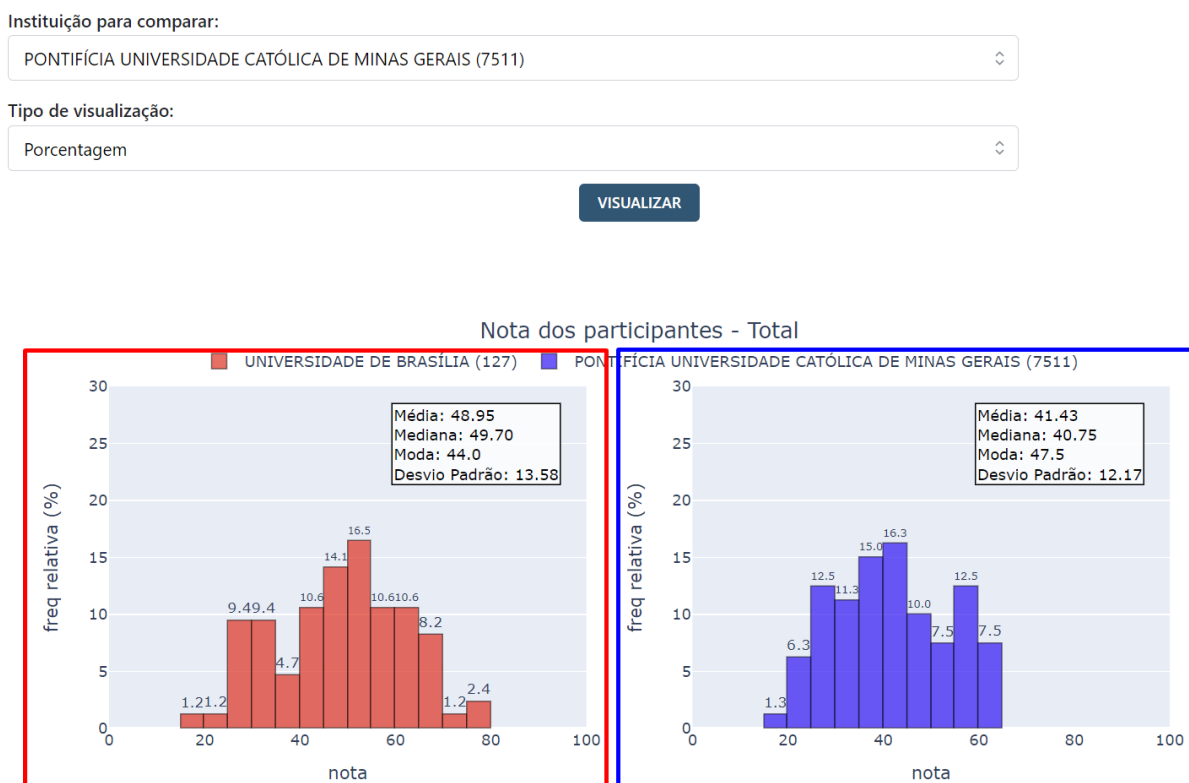


Figura 5.8: Estrutura da análise de distribuição de notas.

Além disso, a ferramenta oferece a flexibilidade de alterar o curso secundário, permitindo comparações com outros cursos pertencentes ao grupo selecionado. A Figura 5.9 mostra o campo de seleção que facilita essa alternância no contexto das instituições PUC.

Cada gráfico é complementado por um bloco que exibe métricas estatísticas, como média, mediana, moda e desvio padrão, proporcionando uma análise mais detalhada.

No contexto das Pontifícias Universidades Católicas (PUCs), a PUCRS foi selecionada por ter o melhor desempenho geral após a UnB. A Figura 5.10 mostra que, embora a UnB

Instituição para comparar:

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS (7511)

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ (865)

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL (1904)

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS (7511)

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS (10814)

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO (11238)

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS (18585)

Figura 5.9: Campo de seleção de curso para a análise de distribuição de notas.

tenha obtido uma nota final superior à da PUCRS, o desvio padrão da UnB foi de 13,58, enquanto o da PUCRS foi de 12,25.

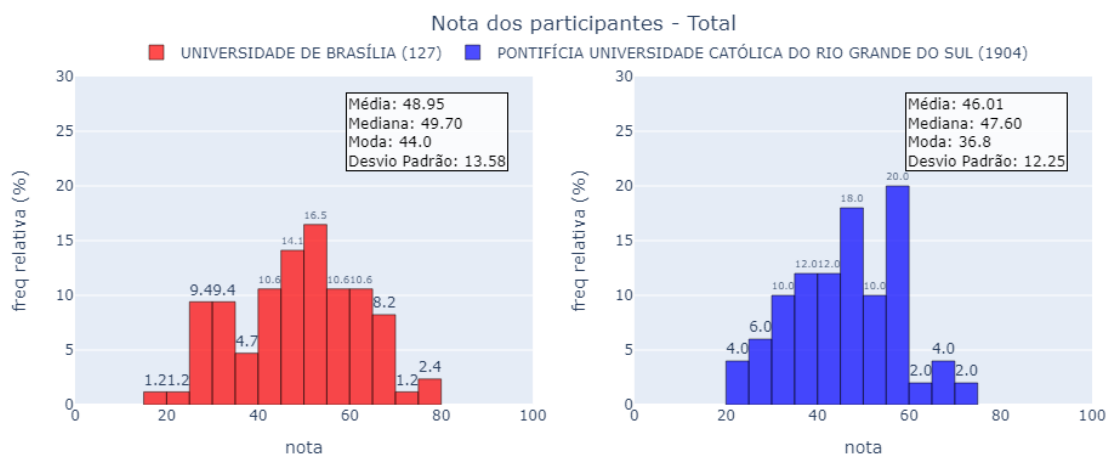


Figura 5.10: Distribuição das notas dos participantes (UnB x PUCRS) - Nota Final.

A Figura 5.11 representa a distribuição das notas para a parte de Formação Geral. Apesar de a UnB ter uma média superior (45,09) em comparação com a PUCRS (42,12), a moda das notas da UnB foi de 30, enquanto a da PUCRS foi de 37,5.

Finalmente, a Figura 5.12 ilustra a comparação no Componente Específico, mostrando que a PUCRS apresentou uma maior concentração de notas na faixa de 45 a 50, enquanto a UnB concentrou suas notas na faixa de 50 a 55.

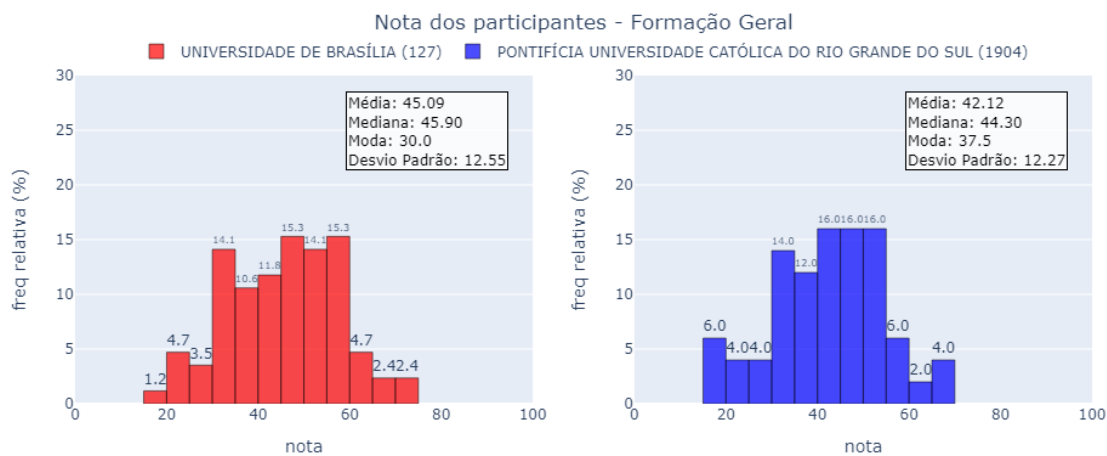


Figura 5.11: Distribuição das notas dos participantes (UnB x PUCRS) - Nota na Formação Geral.

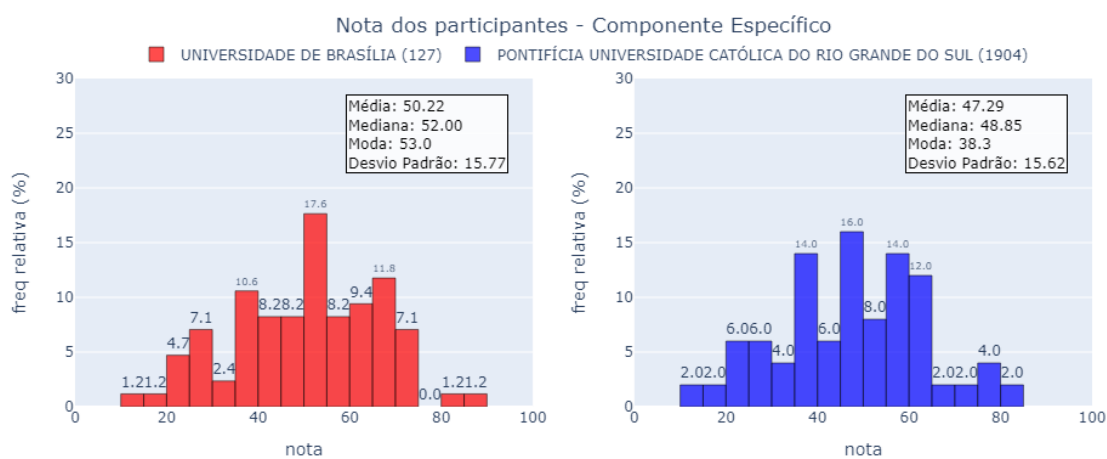


Figura 5.12: Distribuição das notas dos participantes (UnB x PUCRS) - Nota no Componente Específico.

5.1.4 Análise da Percepção da Prova

Outro exemplo das funcionalidades da ferramenta é a análise da percepção da prova, que é apresentada por meio de um gráfico de barras. Este gráfico exibe informações tanto do curso principal quanto do grupo de cursos selecionado para comparação.

Através dele, é possível visualizar o total de respostas em cada opção, facilitando a identificação de tendências e destacando as opções mais escolhidas, além de outros aspectos relevantes relacionados à percepção da prova.

Na Figura 5.13, é apresentada o exemplo no contexto da UnB com as instituições PUC, sendo o gráfico representando as respostas à segunda questão do questionário de percepção da prova: “Qual o grau de dificuldade desta prova na parte do Componente Específico?”.

De acordo com o gráfico, embora a maioria tenha classificado a dificuldade como “Médio”, duas instituições tiveram a maioria dos participantes selecionando “Difícil”: a PUC-RIO, com 49,06%, e a PUC Minas (código do curso igual a 18585), com 57,14%. Esta última informação pode estar relacionada ao desempenho final na prova, conforme apresentado na Figura 5.7, onde a PUC Minas ficou na última posição entre as PUCs e a UnB.

Distribuição de CO_RS_I2

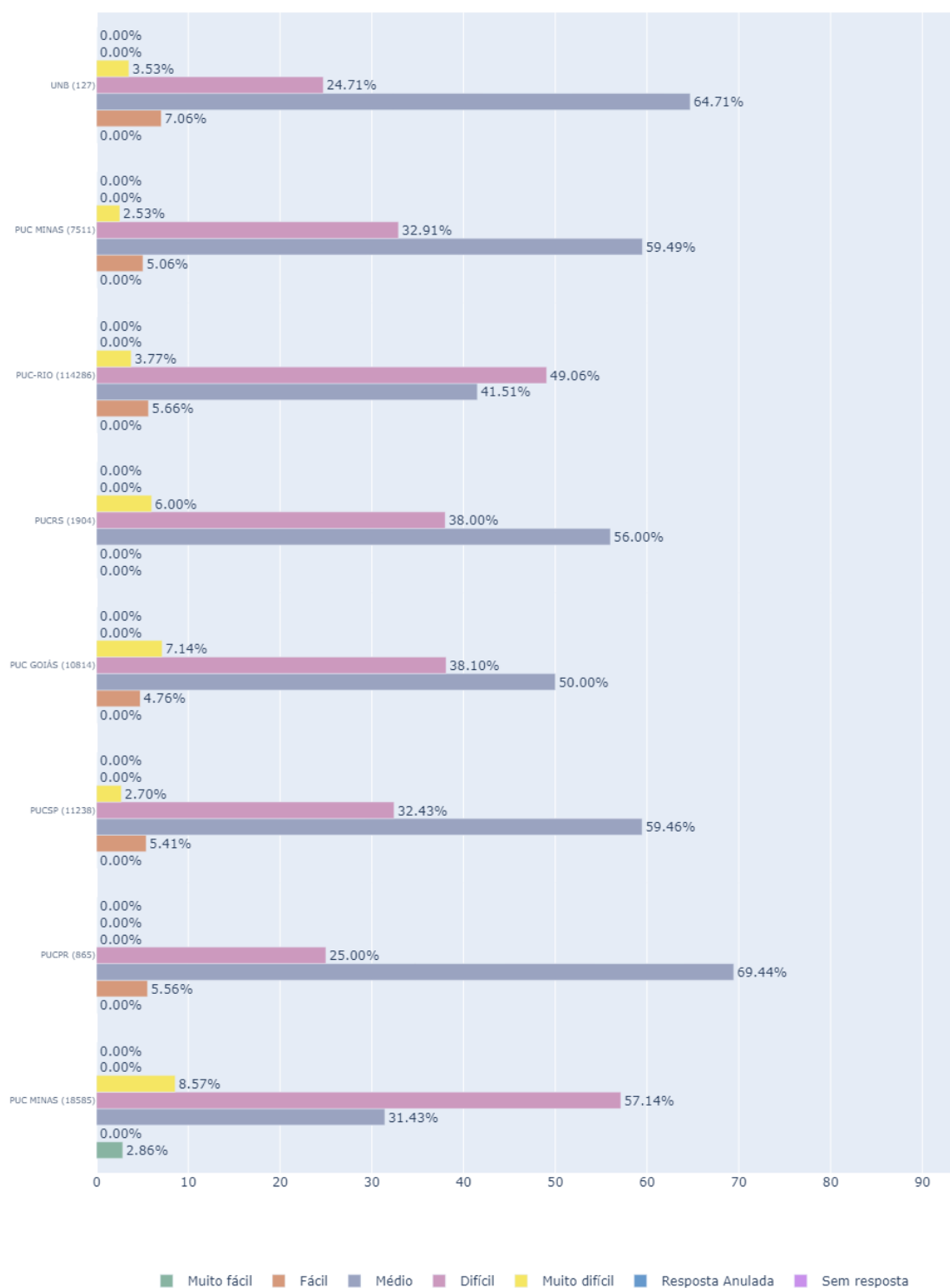


Figura 5.13: Grau de dificuldade da prova na parte do Componente Específico.

5.1.5 Análise do Perfil do Estudante

A etapa correspondente à análise do perfil de estudante inclui a geração de um gráfico de barras, semelhante ao utilizado na análise da percepção da prova. Esta análise é mais um exemplo das diversas funcionalidades que a ferramenta é capaz de oferecer, destacando sua capacidade de lidar com diferentes tipos de dados.

Para isso, foi selecionada a primeira questão do questionário sobre o perfil do estudante, que investiga o sexo dos participantes. Esse gráfico permite visualizar a distribuição dos dados de forma clara e intuitiva, facilitando a comparação entre diferentes grupos e a interpretação dos resultados obtidos pela ferramenta.

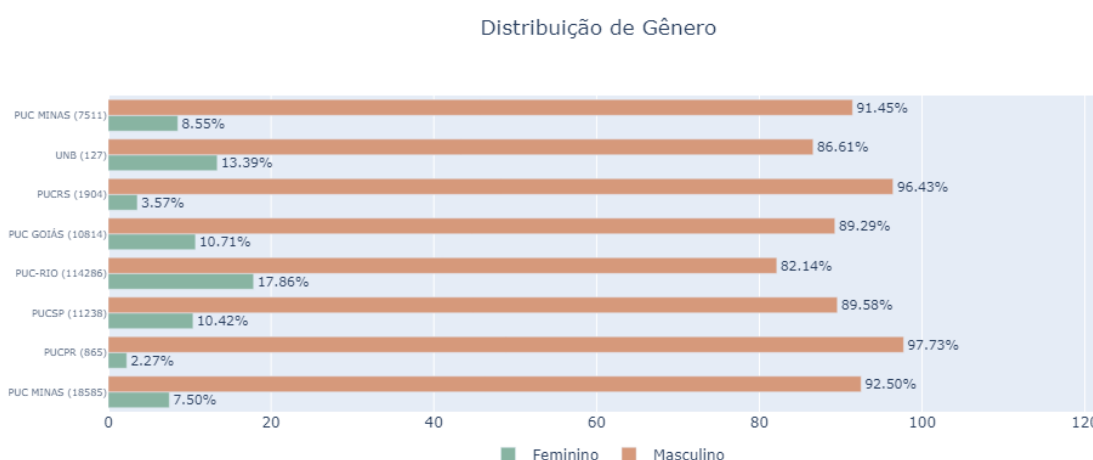


Figura 5.14: Sexo dos participantes.

De acordo com o gráfico apresentado na Figura 5.14, observa-se um baixo índice de participantes do sexo feminino, com todos os cursos exibindo menos de 20% de mulheres concluintes. A UnB, por exemplo, apresentando uma porcentagem de apenas 13,39%.

Essa análise revela uma disparidade significativa na representatividade de gênero entre os cursos comparados, podendo indicar uma tendência predominante de escolha do curso de Ciência da Computação por homens. Esse padrão levanta questões sobre a necessidade de iniciativas que promovam maior inclusão e diversidade de gênero na área. Compreender essas dinâmicas é essencial para a implementação de políticas educacionais que incentivem uma maior participação feminina no campo da Computação.

5.1.6 Análise do Processo de Formação

O último exemplo de análise que a ferramenta gera trata do processo de formação e é visualizada através de um gráfico de barras, semelhante ao apresentado anteriormente na

seção de análise da percepção da prova e perfil do estudante.

Essa funcionalidade permite o exame das respostas das questões 27 a 68 do questionário do estudante. Assim para esta análise, foi selecionada a questão 27, cujo enunciado é: “As disciplinas cursadas contribuíram para sua formação integral, como cidadão e profissional”.

A análise dos dados, apresentada na Figura 5.15, revela que a maioria das instituições exibiu uma predominância na marcação da opção “Concordo totalmente”. No entanto, a UnB se destaca por apresentar uma distribuição mais equilibrada entre as respostas “Concordo parcialmente” (24,75%), “Concordo” (27,72%) e “Concordo totalmente” (24,75%).

Esse resultado sugere uma percepção mais diversa entre os estudantes da UnB sobre a contribuição das disciplinas para sua formação integral. Essa variação pode ser um ponto de partida para estudos futuros que busquem compreender melhor os fatores que influenciam essa percepção e para identificar possíveis áreas de melhoria no currículo.

Distribuição de QE_I27

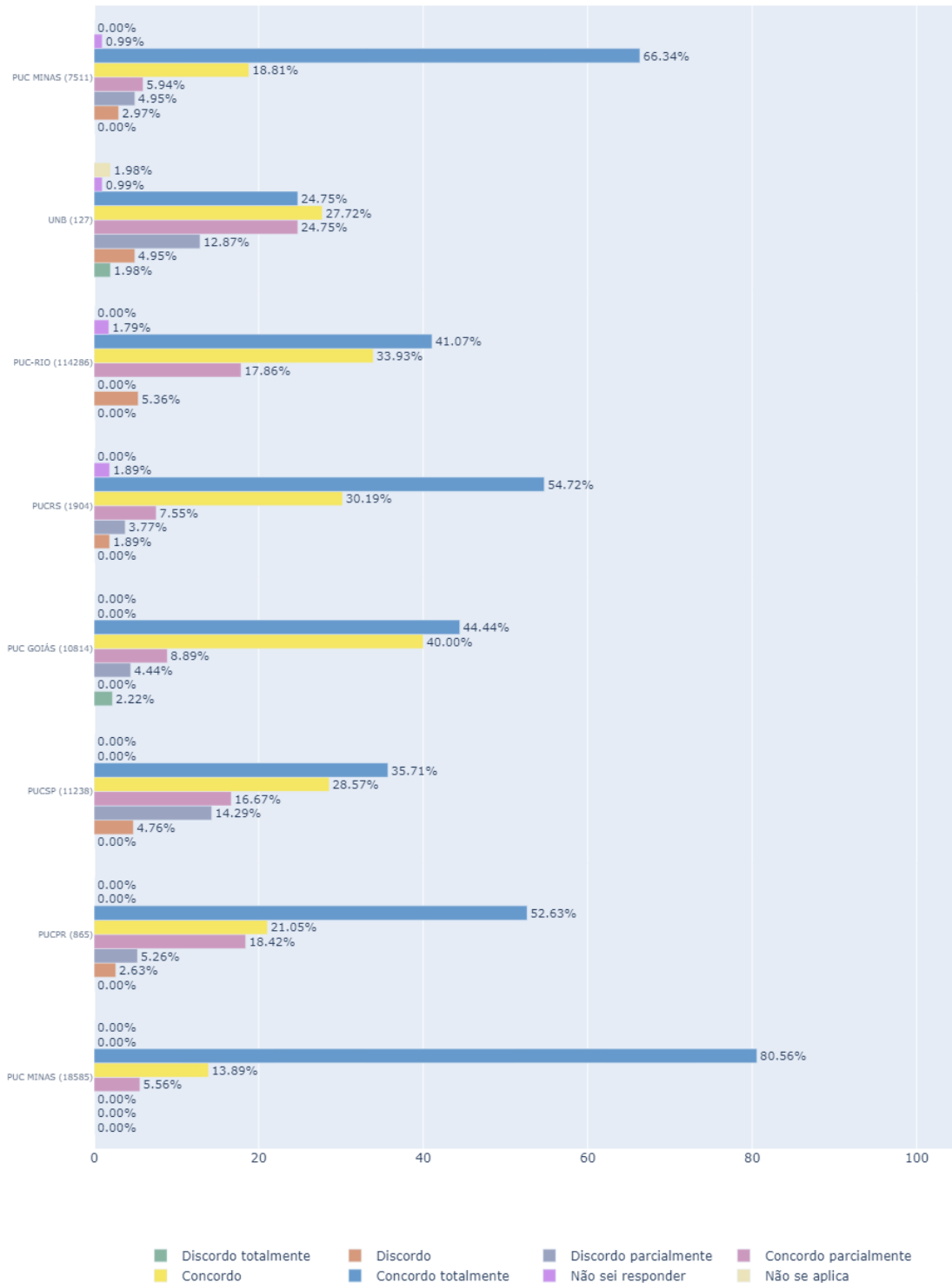


Figura 5.15: As disciplinas cursadas contribuirá para a formação integral, como cidadão e profissional.

5.2 Exemplos Adicionais de Aplicação de Filtros

Esta seção tem como objetivo apresentar exemplos adicionais de aplicação de filtros, demonstrando a capacidade da ferramenta em realizar análises em diversos contextos.

5.2.1 Universidades Públicas com Conceito Enade 5

Este exemplo tem como objetivo demonstrar a aplicação dos filtros “Categoria Administrativa”, “Organização Acadêmica”, “Modalidade de Ensino” e “Conceito Enade”, visto que o relatório do Enade não dispõe dessas informações.

Esse tipo de comparação é fundamental para identificar áreas em que o curso da UnB precisa melhorar, com vistas a alcançar uma nota superior em futuras avaliações. As notas do grupo de cursos selecionados, bem como as da UnB, são apresentadas na Tabela 5.1, retirada da fonte mencionada na Seção 3.1.4. Vale ressaltar que a faixa de intervalo para a nota no Conceito Enade segue a condição apresentada na Tabela 2.2.

Código do Curso	Universidade	Conceito Enade (Contínuo)
12946	UNIVERSIDADE FEDERAL DE MINAS GERAIS (UFMG)	5
12313	UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE (UFRN)	4,535
703	UNIVERSIDADE FEDERAL DE VIÇOSA (UFV)	4,379
65361	UNIVERSIDADE FEDERAL DE JUIZ DE FORA (UFJF)	4,349
13717	UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)	4,338
12581	UNIVERSIDADE FEDERAL DO PARANÁ (UFPR)	4,175
13216	UNIVERSIDADE FEDERAL DE ALAGOAS (UFAL)	4,143
12837	UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO (UFES)	4,119
14217	UNIVERSIDADE FEDERAL DE SANTA CATARINA (UFSC)	4,088
1102342	FUNDAÇÃO UNIVERSIDADE FEDERAL DO ABC (UFABC)	4,064
2717	UNIVERSIDADE ESTADUAL DE CAMPINAS (UNICAMP)	4,047
110528	UNIVERSIDADE FEDERAL DE SÃO PAULO (UNIFESP)	3,995
112680	UNIVERSIDADE FEDERAL DE SÃO CARLOS (UFSCAR)	3,989
3197	UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO (UNESP)	3,989
12710	UNIVERSIDADE FEDERAL FLUMINENSE (UFF)	3,971
127	UNIVERSIDADE DE BRASÍLIA (UNB)	3,816

Tabela 5.1: Cursos selecionados para comparação com suas respectivas notas no Conceito Enade.

Partindo da seleção apresentada na Figura 5.2, no segundo passo, foram aplicados os quatro primeiros filtros ilustrados na Figura 5.16, com o objetivo de selecionar apenas as universidades públicas com Conceito Enade 5. Como resultado, foram identificados 15 cursos, os mesmos listados na Tabela 5.1. Os cursos encontrados são ilustrados no campo à esquerda da Figura 5.17.

Caso fosse necessária uma análise com maior homogeneidade na quantidade de participantes, poderia incluir o filtro adicional “Quantidade mínima de participantes” igual a 40, restringindo a análise aos cursos com pelo menos 40 participantes. O último componente

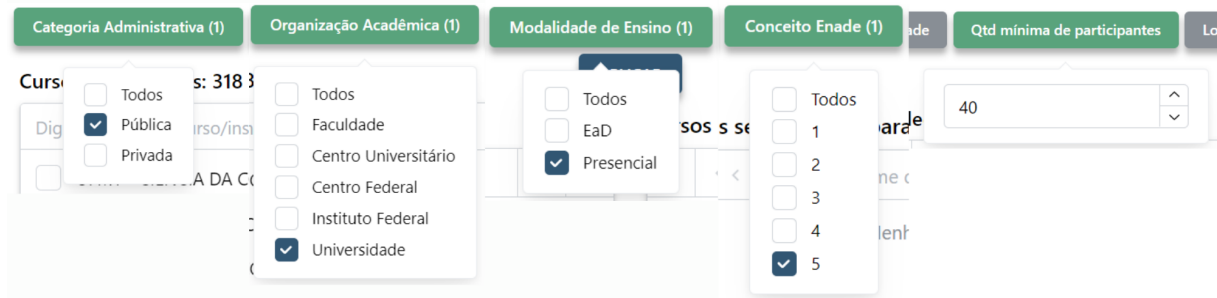


Figura 5.16: Filtros utilizados para cursos com Conceito Enade 5.

da Figura 5.16 apresenta esse filtro extra aplicado, resultando na redução do número de cursos para 8. O componente à direita da Figura 5.17 ilustra a mudança.

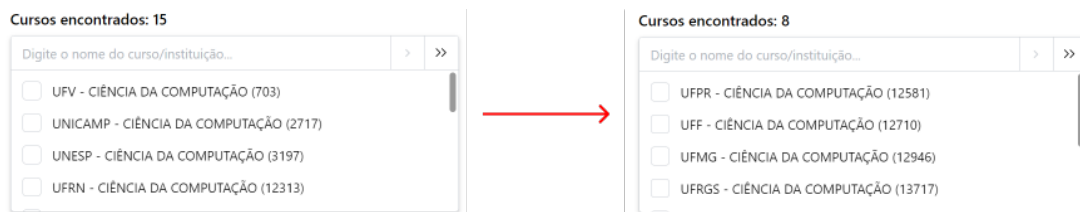


Figura 5.17: Cursos encontrados a partir da aplicação do filtro para universidades públicas com Conceito Enade 5.

5.2.2 Cursos Localizados no Distrito Federal

O exemplo nesta seção tem como objetivo ilustrar o uso do filtro de localização geográfica das instituições. Essa filtragem permite analisar o desempenho do curso de Ciência da Computação (Bacharelado) da UnB em comparação com todas as instituições de ensino superior do Distrito Federal que ofereceram curso na mesma área de conhecimento e participaram do Enade em 2021.



Figura 5.18: Aplicação do filtro de localização.

No primeiro passo da seleção, as configurações são mantidas conforme apresentado na Figura 5.2. No segundo passo, foi aplicado o filtro de “localização”, conforme ilustrado na

Figura 5.18, selecionando o estado “Distrito Federal (DF)”. Esse filtro inclui um campo de seleção que exibe uma lista de todos os estados brasileiros, organizados por região.

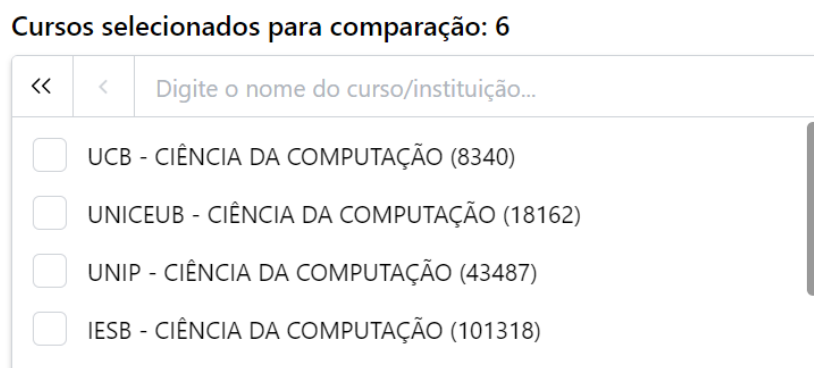


Figura 5.19: Seleção dos cursos do Distrito Federal.

Como resultado, foram identificadas todas as 6 instituições localizadas no DF, conforme mostrado na Figura 5.19. Dessas instituições, 5 são privadas e 1 é um instituto federal.

5.2.3 Lista de Instituições

Esta seção descreve a funcionalidade do filtro de lista de instituições, que permite selecionar cursos com base em nomes ou siglas de instituições de ensino superior. Caso algum valor fornecido não seja reconhecido, nenhum curso será retornado para esse valor. Essa filtragem é especialmente útil quando se deseja analisar uma lista específica de instituições.

Partindo do primeiro passo apresentado na Figura 5.2, o segundo passo consistiu em fornecer a listagem de instituições. O passo é descrito na Figura 5.20, onde as instituições utilizadas são “universidade federal de são paulo, ufm, ufg”.

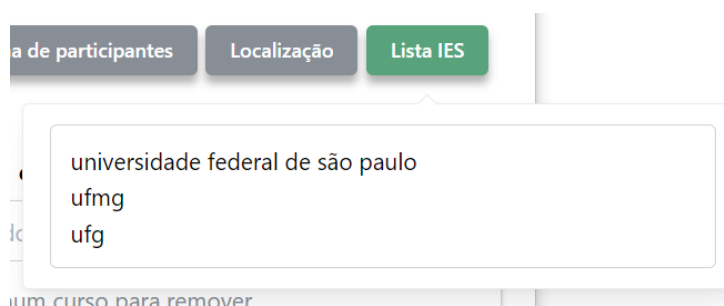


Figura 5.20: Aplicação do filtro de IES.

Como retorno, a ferramenta apresentou três cursos, cada um associado a uma das instituições fornecidas. O resultado gerado é apresentado na Figura 5.21.

O próximo capítulo apresenta as conclusões obtidas com a elaboração do trabalho, resumindo as etapas do desenvolvimento da ferramenta e os impactos que as análises fornecidas podem ter na melhoria da educação superior no país. Além disso, são sugeridos possíveis trabalhos futuros.

Cursos encontrados: 3

Digite o nome do curso/instituição...	>	>>
<input type="checkbox"/> UFMG - CIÊNCIA DA COMPUTAÇÃO (12946)		
<input type="checkbox"/> UFG - CIÊNCIA DA COMPUTAÇÃO (14098)		
<input type="checkbox"/> UNIFESP - CIÊNCIA DA COMPUTAÇÃO (110528)		

Figura 5.21: Cursos encontrados a partir da aplicação do filtro de IES.

Capítulo 6

Conclusões

O objetivo geral deste trabalho foi desenvolver uma ferramenta *Web* capaz de realizar diversas análises detalhadas, comparando cursos de uma mesma área de conhecimento ofertados por diferentes instituições, utilizando os dados do Exame Nacional de Desempenho de Estudantes (Enade). Para alcançar esse propósito, foram definidos objetivos específicos que ajudaram a organizar e guiar o processo de desenvolvimento. O trabalho seguiu os princípios de Extração, Transformação e Carga (ETL), Mineração de Dados e o Modelo Incremental, sendo essas abordagens essenciais para a concretização do objetivo principal.

Como resultado, foi desenvolvida uma ferramenta *Web* que permite realizar análises comparativas detalhadas entre um curso principal e um grupo de cursos da mesma área de conhecimento, utilizando os resultados do Enade. As edições consideradas para este trabalho foram de 2017, 2018, 2019, 2021 e 2022, garantindo à ferramenta uma base de dados ampla.

Diferentemente das análises oficiais divulgadas pelo Inep, que comparam o desempenho de um curso com base em critérios geográficos, esta ferramenta permite a comparação direta entre cursos específicos, oferecendo uma análise mais focada.

Os relatórios do Enade, por exemplo, indicam a média de acertos de um curso, como o de Ciência da Computação da UnB, em relação às médias estadual, regional e nacional. No entanto, essa abordagem limita a interpretação a apenas saber se o curso está acima ou abaixo da média geral.

Com a ferramenta desenvolvida, é possível não só visualizar esse desempenho, mas também criar *rankings* personalizados, permitindo verificar, por exemplo, se um curso está entre os 50 melhores de um conjunto de cursos selecionados. Além disso, a ferramenta oferece a capacidade de realizar análises a nível de curso, abrangendo diversos aspectos como ranking de acertos, notas finais, distribuição de notas, percepção da prova, perfil dos estudantes e processo de formação.

A ferramenta oferece a possibilidade de selecionar cursos com base em múltiplas notas do Conceito Enade, como, por exemplo, cursos com notas 4 e 5. Além disso, permite filtrar cursos por múltiplos estados, categorias administrativas, organizações acadêmicas, modalidades de ensino, quantidade mínima de participantes e por lista pré-definida de instituições.

Para alcançar o objetivo principal, o primeiro passo foi realizar um estudo dos microdados dessas edições. Isso envolveu compreender a estrutura dos dados, identificar padrões recorrentes e garantir a integridade e completude dos dados por meio da incorporação de informações de fontes adicionais. Esse estudo foi fundamental para assegurar que a ferramenta tivesse uma base de dados capaz de suportar as análises pretendidas.

Após a etapa inicial, foi necessário organizar os dados em uma estrutura que facilitasse sua manipulação e integração na ferramenta. Esse processo teve como foco principal garantir que os dados estivessem adequados ao objetivo geral, assegurando a geração de todas as análises necessárias.

Com os dados organizados, a próxima etapa envolveu a criação do protótipo da interface. Esse protótipo foi desenvolvido com foco em uma apresentação visual intuitiva. Para essa fase, foi utilizado o Figma, uma ferramenta amplamente reconhecida por sua capacidade de criar protótipos interativos.

Com a finalização do protótipo inicial, foi dado início à implementação da ferramenta utilizando a linguagem Python. Durante essa etapa, diversas funcionalidades foram desenvolvidas, com especial destaque para os filtros que permitem a seleção dos cursos para comparação. Esses filtros incluem opções como Categoria Administrativa, Organização Acadêmica, Modalidade de Ensino, Conceito Enade, Quantidade Mínima de Participantes, Localização e Lista de Instituições. Esses filtros são essenciais para uma busca mais otimizada dos cursos de interesse.

Na etapa final de desenvolvimento, a ferramenta foi configurada para apresentar diversas análises dos dados do Enade. A interface foi projetada para permitir a visualização de diferentes tipos de análises, incluindo Ranking de acertos por questão, Notas finais, Distribuição de notas, Percepção da prova, Perfil do estudante e Processo de formação. Essas análises são apresentadas por meio de tabelas detalhadas, histogramas, gráficos de barras e incluem também métricas estatísticas relevantes.

Essas inclusões garantiram que a ferramenta se tornasse apta a lidar com todos os dados de interesse para estudos relacionados ao Enade, oferecendo uma visão completa e detalhada das diferentes dimensões de desempenho e perfil dos cursos avaliados.

Como forma de apresentar o funcionamento da ferramenta, foram definidos quatro exemplos práticos, cada um com o objetivo de demonstrar o funcionamento dos filtros em diferentes contextos. Os exemplos escolhidos foram: Instituições classificadas como

Pontifícia Universidade Católica (PUC), Universidades públicas com Conceito Enade 5, Instituições localizadas no Distrito Federal e Lista de Instituições.

Este trabalho resultou na criação de uma ferramenta que, além de superar as limitações de estudos anteriores, introduziu novas funcionalidades para uma visualização mais detalhada dos dados do Enade. A ferramenta permite realizar novas comparações de desempenho entre os cursos, podendo ser útil na elaboração de políticas educacionais mais eficazes para aprimorar a educação superior, bem como podendo auxiliar na tomada de decisões sobre investimentos em infraestrutura, capacitação de professores e outras ações voltadas à melhoria do ensino.

6.1 Trabalhos Futuros

Como sugestão para trabalhos futuros, propõe-se a criação de novas formas de apresentação dos dados e o desenvolvimento de novas análises são recomendados para continuar aprimorando a ferramenta e expandindo suas capacidades analíticas.

Adicionalmente, propõe-se a implementação de um módulo que automatize a atualização contínua dos dados, garantindo que a ferramenta permaneça sempre atualizada. Outra sugestão relevante é a personalização da interface, adaptando sua estrutura para atender às necessidades específicas de cada perfil de usuário.

Referências

- [1] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira: *Relatório de Cursos - ENADE*. <https://enade.inep.gov.br/enade/#!/relatorioCursos>, 2024. ix, 36
- [2] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira: *Provas e Gabaritos - ENADE*. <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade/provas-e-gabaritos>, 2024. ix, 40
- [3] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira: *Exame Nacional de Desempenho de Estudantes*. <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade>, 2024. xi, 7
- [4] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira: *Nota Técnica nº 2/2023*. https://download.inep.gov.br/educacao_superior/enade/notas_tecnicas/2023/nota_tecnica_n_2_2023_cei_cggi_daes_descricao_da_metodologia_de_calculo_do_conceito_enade.pdf, 2023. xi, 8, 9
- [5] Fayyad, Usama, Gregory Piatetsky-Shapiro e Padhraic Smyth: *From data mining to knowledge discovery in databases*. *AI Magazine*, 17(3):37, Mar. 1996. <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1230>. 3, 11
- [6] IBM: *O que é ETL (extração, transformação e carregamento)?* <https://www.ibm.com/br-pt/topics/etl>, 2024. 3, 11
- [7] Lessa, Rafael e Orivaldo Junior: *Modelos de processos de engenharia de software*. http://xps-project.googlecode.com/svn-history/r43/trunk/outros/02_Artigo.pdf, 2009. 3
- [8] Brasil: *Lei nº 10.861*. https://www.planalto.gov.br/ccivil_03/_ato2004-2006/2004/lei/110.861.htm, 2004. Diário Oficial da União, Brasília, DF, 15 abr. 2004. 6
- [9] Brasil: *Portaria nº 494*. https://www.in.gov.br/en/web/dou/-/portaria-n-494-de-8-de-julho-de-2021-330988556?fbclid=IwAR1ZaMEGCFQYbliePcohkYqTeT2GPevIKdCvHqb7eIXIssYHNCsU_P5xSgs, 2021. Diário Oficial da União, Brasília, DF, 8 jul. 2021. 6
- [10] Ministério da Educação: *Portaria nº 610*. <https://www.in.gov.br/en/web/dou/-/portaria-n-610-de-27-de-junho-de-2024-569033072>, 2024. 6

- [11] Ministério da Educação: *Portaria n° 611*. <https://www.in.gov.br/en/web/dou/-/portaria-n-611-de-27-de-junho-de-2024-569019655>, 2024. 6
- [12] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira: *Conceito Enade*. <https://www.gov.br/inep/pt-br/areas-de-atuacao/pesquisas-estatisticas-e-indicadores/indicadores-de-qualidade-da-educacao-superior/conceito-enade>, 2020. 8
- [13] QEDu: *O que são microdados?* <https://academia.qedu.org.br/glossario/o-que-sao-microdados/?repeat=w3tc>, 2024. 9
- [14] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira: *Microdados Enade*. <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/enade>, 2020. 9, 15
- [15] Presidência da República (Brasil): *Lei n° 13.709*. https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm, 2018. 9
- [16] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira: *Enem e Enade têm novo conjunto de microdados publicados*. <https://www.gov.br/inep/pt-br/assuntos/noticias/institucional/enem-e-enade-tem-novo-conjunto-de-microdados-publicados>, 2024. 9
- [17] Bellinger, Gene, Durval Castro e Anthony Mills: *Data, information, knowledge, and wisdom*, 2004. 10
- [18] Setzer, Valdemar: *Dado, informação, conhecimento e competência*. DataGramZero Revista de Ciência da Informação, n. 0, 28, 1999. 10
- [19] Chen, Min, David Ebert, Hans Hagen, Robert S. Laramee, Robert van Liere, Kwan Liu Ma, William Ribarsky, Gerik Scheuermann e Deborah Silver: *Data, information, and knowledge in visualization*. IEEE Computer Graphics and Applications, 29(1):12–19, 2009. 10
- [20] Bansal, Srividya K e Sebastian Kagemann: *Integrating big data: A semantic extract-transform-load framework*. Computer, 48(3):42–50, 2015. 11
- [21] Deb Nath, Rudra Pratap, Katja Hose e Torben Bach Pedersen: *Towards a programmable semantic extract-transform-load framework for semantic data warehouses*. Em *Proceedings of the ACM Eighteenth International Workshop on Data Warehousing and OLAP*, páginas 15–24, 2015. 11
- [22] Gour, Vishal, SS Sarangdevot, Govind Singh Tanwar e Anand Sharma: *Improve performance of extract, transform and load (etl) in data warehouse*. Int. Journal on Comp. Sci. and Eng, 2(3):786–789, 2010. 11
- [23] Galvão, Noemi Dreyer e Heimar de Fátima Marin: *Técnica de mineração de dados: uma revisão da literatura*. Acta Paulista de Enfermagem, 22(5):686–690, Sep 2009, ISSN 0103-2100. <https://doi.org/10.1590/S0103-21002009000500014>. 11

- [24] Allam, Allam Hassan e Halina Mohamed Dahlan: *User experience: challenges and opportunities*. Journal of information systems research and innovation, 3(1):28–36, 2013. 12
- [25] Costa, Jean Carlos Araújo: *AVis-Scholar: uma metodologia de visualização e análise de dados na educação*. <http://www.repositorio.jesuita.org.br/handle/UNISINOS/5280>, 2016. 12
- [26] Araújo, Rodrigo Alexandrino: *Análise dos microdados do Enade: proposta de uma ferramenta de exploração utilizando mineração de dados*. <https://repositorio.bc.ufg.br/tede/items/bc914922-7308-4ce1-866a-01df9e118a69>, 2019. 12
- [27] Cunha, Renan, Claudomiro Sales e Reginaldo Santos: *Análise automática com os microdados do enade para melhoria do ensino dos cursos de ciência da computação*. Em *Anais do XXIX Workshop sobre Educação em Computação*, páginas 208–217, Porto Alegre, RS, Brasil, 2021. SBC. <https://sol.sbc.org.br/index.php/wei/article/view/15912>. 12
- [28] Cruz, Breno e Kilbert Maciel: *Saen: Um sistema de análise de dados baseado em resultados do enade*. Em *Anais da XI Escola Regional de Computação do Ceará, Maranhão e Piauí*, páginas 92–101, Porto Alegre, RS, Brasil, 2023. SBC. <https://sol.sbc.org.br/index.php/ercemapi/article/view/28477>. 12
- [29] Figma: *Figma*. <https://www.figma.com/pt-br/>, 2024. Disponível online, acessado em 13 ago. 2024. 13
- [30] Jupyter Project: *Jupyter notebook*. <https://jupyter.org/>, 2024. Disponível online, acessado em 13 ago. 2024. 13
- [31] PostgreSQL Global Development Group: *Postgresql: The world's most advanced open source relational database*. <https://www.postgresql.org/>, 2024. Disponível online, acessado em 13 ago. 2024. 13
- [32] Python Software Foundation: *Python: Linguagem de programação*. <https://www.python.org/>, 2024. Disponível online, acessado em 13 ago. 2024. 14
- [33] Pandas Development Team: *Pandas: Python Data Analysis Library*. <https://pandas.pydata.org/>, 2024. 14
- [34] Plotly: *Dash Python User Guide*. <https://dash.plotly.com/>, 2024. 14
- [35] Mantine: *Dash Mantine Components*. <https://www.dash-mantine-components.com/>, 2024. 14
- [36] Ministério da Educação: *Cursos de Graduação do Brasil*. <https://dadosabertos.mec.gov.br/indicadores-sobre-ensino-superior/item/183-cursos-de-graduacao-do-brasil>, 2024. 17
- [37] Instituto Brasileiro de Geografia e Estatística: *Códigos dos Municípios*. <https://www.ibge.gov.br/explica/codigos-dos-municipios.php>, 2024. 17

- [38] Chandez: *Estados e Cidades do IBGE*. <https://github.com/chandez/Estados-Cidades-IBGE>, 2024. 17
- [39] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira: *Indicadores de Qualidade da Educação Superior*. <https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/indicadores-educacionais/indicadores-de-qualidade-da-educacao-superior>, 2024. 17
- [40] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP): *Questionário do Estudante - ENADE*. <https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/enade/questionario-do-estudante>, 2020. 24
- [41] Ministério da Educação: *e-MEC - Sistema de Regulação da Educação Superior*. <https://emec.mec.gov.br/emec/nova>, 2024. 37