

# Modelagem de Painel de Business Intelligence com Dados Históricos do Exame Nacional Para Certificação de Competências de Jovens e Adultos

Flavio V. D. Figueiredo, Diógenes V. M. Oliveira

Departamento de Ciências da Computação – Instituto de Ciências Exatas –  
Universidade Federal de Minas Gerais (UFMG) – Belo Horizonte – MG – Brasil

flaviovdf@dcc.ufmg.br, diogenesvazmelo@gmail.com

**Abstract.** *This paper addresses the structuring of a Data Warehouse (DW) and a Business Intelligence (BI) dashboard using open data from the National Examination for the Certification of Skills for Young People and Adults (Encceja). Encceja is a relevant exam for Brazilian education and provides information about student performance in different areas of knowledge. The DW was structured in a AWS cloud computing environment using dimensional modeling techniques to enable comprehensive data analysis and foster discussions for educational system improvement. The BI was developed in Tableau Desktop and published on Tableau Public, featuring visualizations on the Exam, participant characteristics, and performance in the tests.*

**Resumo.** *Este trabalho aborda a estruturação de um Data Warehouse (DW) e um painel de Business Intelligence (BI) com dados abertos do Exame Nacional para Certificação de Competências de Jovens e Adultos (Encceja). O Encceja é um exame relevante para a educação brasileira e fornece informações sobre o desempenho dos alunos em diferentes áreas do conhecimento. Realizou-se a estruturação do DW no ambiente de computação em nuvem AWS utilizando-se técnicas de modelagem dimensional, a fim de permitir uma análise abrangente dos dados e proporcionar discussões para a melhoria do sistema educacional. O BI foi desenvolvido no Tableau Desktop e publicado no Tableau Public, e conta com visualizações sobre o Exame, características dos participantes e desempenho nas provas.*

## 1. Introdução

Um painel de Business Intelligence (BI) é “uma interface gráfica que apresenta dados em um formato visual, tornando mais intuitivo extrair informações valiosas e tomar decisões informadas” [Awad et. al. 2022]. É uma ferramenta importante para a tomada de decisões estratégicas, pois permite que as organizações analisem informações relevantes sobre seu desempenho, concorrentes e tendências do setor. A análise de dados por meio de um BI pode ajudar as empresas a identificar oportunidades e ameaças, além de aprimorar seus processos, produtos e serviços.

No contexto da educação básica, os painéis têm o potencial de fornecer informações críticas sobre a população e a eficácia das políticas adotadas, impulsionando

os funcionários do governo a tomar medidas que levem a melhorias significativas nas práticas educacionais.

A análise de dados abertos dos exames nacionais brasileiros pode permitir a identificação de vulnerabilidades de aprendizado, auxiliando na compreensão das áreas em que os estudantes têm mais dificuldade. Além disso, também contribui para verificar se as políticas implementadas estão alcançando os objetivos propostos e direcionar recursos de maneira mais eficiente.

Um dos exames mais relevantes aplicados anualmente em todo o território brasileiro é o Exame Nacional para Certificação de Competências de Jovens e Adultos (Encceja), que tem como principal objetivo avaliar as habilidades e conhecimentos de jovens e adultos que não concluíram a educação básica na idade apropriada. Para ser elegível para fazer o Encceja, os candidatos devem ter pelo menos 15 anos para o nível fundamental e 18 anos para o nível médio. O Exame conta com quatro provas, sendo: história e geografia (CH); ciências naturais (CN); matemática (MT); língua portuguesa, língua estrangeira moderna, artes, educação física e redação (LC) para a certificação de ensino fundamental e ciências humanas e suas tecnologias (CH); ciências da natureza e suas tecnologias (CN); matemática e suas tecnologias (MT); linguagens e códigos e suas tecnologias e redação (LC) para a certificação de ensino médio. Não é necessário que o candidato se inscreva para a realização de todas as provas do Exame, e aqueles que acertam pelo menos 50% das questões em cada prova recebem um certificado equivalente ao diploma de uma escola regular.

Os microdados do Encceja são divulgados anualmente pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep) e reúnem um conjunto de dados detalhados sobre o exame, mas sem a possibilidade de identificar pessoas, o que mantém a conformidade com as regulamentações previstas na Lei Geral de Proteção de Dados (LGPD).

Conseqüentemente, há material para apoiar o desenvolvimento de um painel de Business Intelligence (BI) com os dados históricos das últimas aplicações do Encceja, que pode ser utilizado para explorar e dar visibilidade a possíveis vulnerabilidades educacionais no Brasil. Essa ferramenta poderia ser valiosa para análises aprofundadas e para orientar a formulação de políticas educacionais mais eficazes.

Para desenvolver esse painel é necessário primeiro organizar os elementos de dados e padronizar suas relações - ou, em outras palavras, definir um modelo de dados. O trabalho atual tem como objetivo discutir o processo de modelagem dos dados dos últimos cinco anos de aplicações do Encceja e o desenvolvimento de um painel de Business Intelligence (BI) com tais dados.

## **2. Referencial Teórico**

O presente trabalho é ancorado no amplo material bibliográfico produzido por Ralph Kimball e Margy Ross, defensores do modelo dimensional na construção de *Data Warehouses* (DWs) e sistemas de BI.

O modelo dimensional, desenvolvido por Kimball e Ross, é uma abordagem de design de DWs amplamente utilizada na indústria de BI. Ele envolve a criação de tabelas fato, que armazenam as medidas de desempenho que uma organização deseja monitorar

(como vendas ou receitas), e tabelas de dimensão, que descrevem os aspectos contextuais dos dados (como o tempo, o local ou o cliente) [Kimball, Ross 2013].

Por além da implementação técnica, faz-se necessário considerar o público-alvo na criação das visualizações e a importância da progressão narrativa para contar histórias envolventes com os dados. Knaflíc (2019) apresenta uma abordagem passo a passo para a criação de visualizações de dados eficazes, desde a identificação dos objetivos até a escolha das cores e fontes apropriadas. Ela também aborda a importância da clareza e simplicidade, fornecendo dicas práticas para melhorar a legibilidade e a compreensão das visualizações.

Ademais, as documentações de softwares, bibliotecas e sistemas utilizados para o desenvolvimento do trabalho também se apresentam como um valioso recurso de apoio, sobretudo diante de dúvidas na implementação e no seguimento das melhores práticas.

### 3. Metodologia

O primeiro passo realizado foi o entendimento das bases de dados, a leitura da documentação disponibilizada junto a tais bases e a conferência dos dicionários de dados. Todos os arquivos são públicos, disponibilizados anualmente, e podem ser acessados através da plataforma única do Governo Federal, na guia destinada aos microdados do Inep, que “representam o nível mais granular dos dados, permitindo uma quebra detalhada e oportunidades de exploração” [Zaccarelli, 2020].

O passo seguinte compreendeu a definição do tipo de ambiente a ser utilizado para o desenvolvimento e desenho da arquitetura. Para o presente trabalho, optou-se por utilizar o serviço de computação em nuvem Amazon Web Services (AWS).

Posteriormente, foi feito o consumo dos dados da página web em que o Inep os disponibiliza. Com isso, pôde-se executar a Extração, Transformação e Carga (*Extract, Transform and Load*, ou ETL) desses dados para o ambiente seguindo a modelagem de fatos e dimensões (Figura 1).

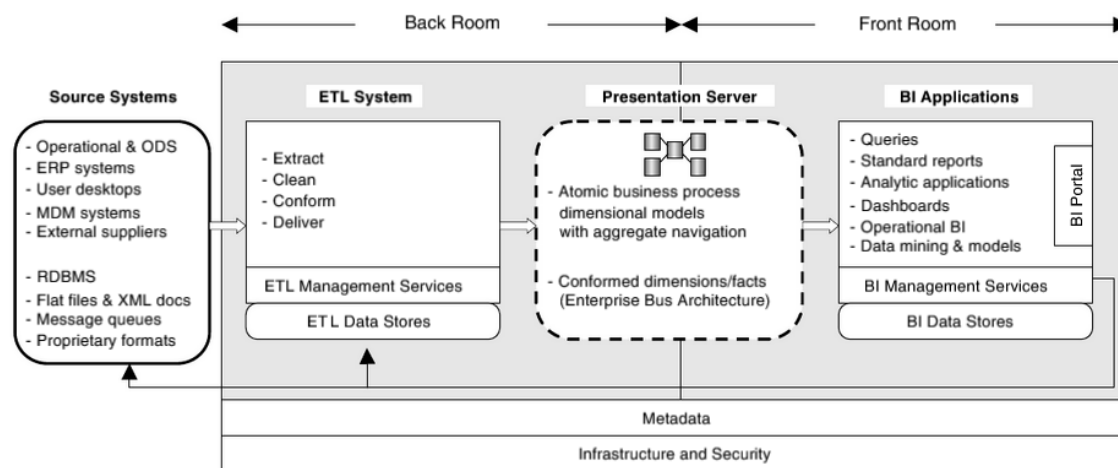


Figura 1. Modelo arquitetural em alto nível de um DW/BI [Kimball, Ross 2008].

Por fim, o último passo caracterizou-se pela manipulação dos dados para construção das visualizações gráficas com as quais o usuário pode interagir.

### 3.1. Bases de Dados

As bases de dados do Enceja, para cada um dos anos estudados, consistiam de quatro arquivos no formato de valores separados por vírgulas (*Comma-Separated Values*, ou CSV), a saber:

- Arquivo com dados da aplicação regular: dados cadastrais, respostas das provas e do questionário socioeconômico para candidatos nacionais sem restrição de liberdade (regulares);
- Arquivo com dados da aplicação para pessoas privadas de liberdade: dados cadastrais e respostas das provas para candidatos nacionais com restrição de liberdade;
- Arquivo com dados das respostas do questionário socioeconômico das pessoas privadas de liberdade: informações das respostas ao questionário socioeconômico por candidatos nacionais com restrição de liberdade;
- Arquivo com detalhamento sobre itens da prova: estrutura geral dos itens das provas (questões).

Outrossim, foram utilizados os arquivos de documentação, sobretudo os dicionários de dados, para adicionar descrições aos códigos utilizados nos arquivos gerados pelo Inep. Além disso, foram obtidas informações de estimativas populacionais dos estados brasileiros para cada um dos anos de realização do Enceja a partir de relatórios elaborados pelo Instituto Brasileiro de Geografia e Estatística (IBGE) publicados no Diário Oficial da União (DOU).

Vale destacar que o Enceja é também aplicado no exterior, tanto na versão regular quanto para pessoas privadas de liberdade. Em 2022, por exemplo, a prova foi em aplicada em 18 cidades de 12 países. No presente estudo, contudo, foram consideradas apenas as aplicações em território nacional, visto que a base divulgada pelo Inep não contempla as provas realizadas fora do Brasil.

### 3.2. Modelagem Dimensional

Segundo Kimball e Ross (2013), a modelagem dimensional é uma abordagem de design de banco de dados que se concentra em representar as informações de um negócio de forma intuitiva e fácil de entender.

A modelagem dimensional é baseada em duas principais estruturas: tabelas de fatos (*fact tables*) e tabelas de dimensão (*dimension tables*). A tabela de fatos contém as métricas e medidas quantitativas de interesse para análise, como vendas, lucro, quantidade, etc. Ela representa os eventos ou transações que ocorrem no negócio e é central para a modelagem dimensional [Kimball e Ross, 2013].

As tabelas de dimensão, por sua vez, fornecem o contexto para as medidas presentes na tabela de fatos. Elas contêm atributos descritivos que permitem a análise e a segmentação dos dados, como datas, produtos, clientes, localizações, entre outros. As tabelas de dimensão são utilizadas para filtrar, agrupar e organizar os dados, permitindo que os usuários analisem as informações de diferentes perspectivas [Kimball e Ross, 2013].

A modelagem dimensional da base de dados do Enceja resultou em 11 dimensões, conforme descrito a seguir:

- DIM\_APROVADO: descreve se o participante foi aprovado ou reprovado;
- DIM\_CERTIFICACAO: descreve a certificação requerida pelo candidato (certificação do Ensino Fundamental ou do Ensino Médio);
- DIM\_ENTIDADE\_CERTIFICADORA: descreve o nome da entidade responsável pela certificação do exame;
- DIM\_FAIXA\_ETARIA: descreve as faixas etárias dos candidatos definidas pelo INEP;
- DIM\_INSCRITO: descreve se o participante estava inscrito ou não em determinada prova do Enceja;
- DIM\_ITENS\_PROVAS: dimensão criada a partir do arquivo com detalhamento sobre itens da prova;
- DIM\_PRESENCA\_PROVA: descreve se o candidato estava presente, faltou ou foi eliminado da prova;
- DIM\_QUESTIONARIO: dimensão que detalha as perguntas e opções de respostas para os questionários socioeconômicos aplicados aos candidatos nacionais sem restrição de liberdade (regulares).
- DIM\_SEXO: descreve o sexo do candidato;
- DIM\_STATUS\_REDACAO: descreve ocorrência de possíveis problemas na redação;
- DIM\_UNIDADE\_FEDERATIVA: descreve a Unidade Federativa (Estado) de aplicação da prova;

Foram criadas, adicionalmente, duas tabelas ponte: uma para as respostas do questionário socioeconômico e outra para a relação hierárquica de inscrição, presença e aprovação em determinada prova do Enceja (para que um participante seja aprovado, é preciso que ele esteja presente no dia da avaliação e também inscrito na prova). Essas tabelas ponte lidam com relações muitos para muitos e evitam que se crie uma tabela fato “centopeia”, cercada de dimensões repetidas.

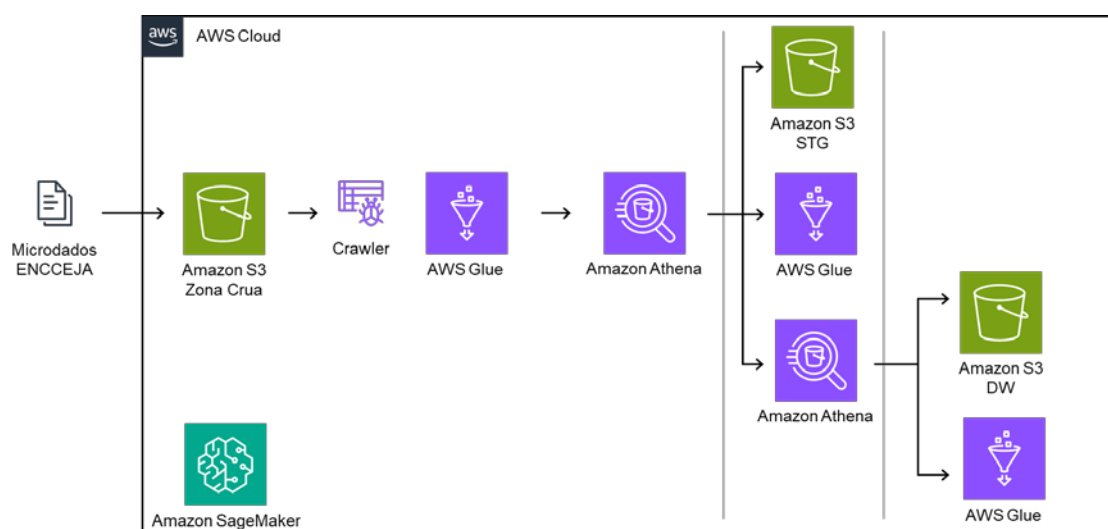
É importante ressaltar que a dimensão relativa ao questionário não contempla as respostas dos candidatos com restrição de liberdade, pois o Inep desvinculou as questões de cunho socioeconômico das objetivas para tais candidatos, em observância à LGPD. Dessa forma o arquivo com dados das respostas do questionário socioeconômico das pessoas privadas de liberdade não foi consumido ou utilizado no modelo.

Ademais, a tabela fato manteve de menor grão as respostas de um candidato inscrito sob certo número (não correspondente ao verdadeiro número de inscrição do candidato). Candidatos com restrição de liberdade tiveram seus campos relativos às questões socioeconômicas transformadas em nulo.

### 3.2. Arquitetura

A utilização da tecnologia de computação em nuvem para a estruturação de um DW oferece diversas vantagens significativas. Segundo Kimball e Ross (2013), a computação em nuvem proporciona escalabilidade quase ilimitada e permite que a infraestrutura seja dimensionada para cima ou para baixo de acordo com as necessidades. Além disso, paga-se apenas pelos recursos utilizados, e, dessa maneira, a computação em nuvem pode fornecer economias significativas de custos em comparação com a aquisição e manutenção de hardware.

No presente estudo, optou-se por utilizar os recursos de computação em nuvem da Amazon, a Amazon Web Services (AWS), para estruturação do DW e processo de ETL.



**Figura 2. Modelo arquitetural do DW na nuvem AWS. Fonte: elaborado pelo autor.**

A Figura 2 ilustra a arquitetura em nuvem para execução dos processos de extração, transformação e carga dos dados para o DW. Foram utilizados os seguintes recursos da AWS:

- S3 (Simple Storage Service): serviço de armazenamento em nuvem altamente escalável, seguro e durável, que permite armazenar e recuperar dados de maneira fácil e flexível, através da separação em *buckets* (contêineres para objetos);
- Glue: serviço de ETL que simplifica e automatiza o processo de preparação e transformação de dados para análise, servindo também como catálogo de dados (*Data Catalog*);
- Athena: serviço de consulta interativa que permite analisar dados usando linguagem de consulta estruturada (*Structured Query Language, SQL*);
- Sagemaker: serviço de aprendizado de máquina no qual os usuários podem construir, treinar e implantar modelos de machine learning de forma rápida e fácil, permitindo também trabalhar com instâncias de notebooks Jupyter e executar códigos em Python;

- CloudFormation: serviço que permite criar e gerenciar recursos de infraestrutura, fornecendo uma maneira automatizada e consistente de provisionar e implantar recursos na nuvem da AWS.

Antes de iniciar o processo de ETL, executou-se através do CloudFormation um modelo similar ao disponibilizado por Zaccarelli (2020) para provisionar serviços de fundação da arquitetura. Esse modelo define *buckets* do S3 para carregar os arquivos de dados e armazenar as saídas de consultas executadas no Athena, além de configurar o grupo de trabalho para utilização deste. O modelo cria também um banco de dados no *Data Catalog* (Glue), e o permissionamento no IAM utilizado pelo Glue para acessar os serviços. Por fim, o script gera a instância de notebook do Sagemaker através do qual foram executados os códigos de ingestão e catalogação dos dados.

### 3.3. ETL

Os procedimentos de consumo dos dados a partir do Inep, descompactação, seleção dos arquivos de interesse, remoção de caracteres indesejados, conversão do encode dos arquivos para o padrão UTF-8, ingestão para o *bucket* e registro das tabelas (brutas) no catálogo de dados do Glue foi realizado por meio de um *notebook* no Sagemaker.

Posteriormente, utilizou-se o Athena para a atribuição de um novo ID único e numérico aos campos textuais, a exemplo o nome das entidades certificadoras, bem como a conversão de alguns tipos de dados e o desmembramento de alguns campos em colunas, vide os que traziam as respostas dos candidatos para cada um dos cadernos das provas (matemática, linguagens, ciências humanas e ciências da natureza) em 30 caracteres e foram segregados em 30 colunas distintas.

Por além disso, o Athena foi também utilizado para carga manual das dimensões, a fim de enriquecê-las com conhecimentos gerais (ex: à dimensão de unidade federativa foi acrescentada a informação da região do Brasil a qual pertence cada Estado), ou com informações que demandavam leitura, avaliação e entendimento da documentação e dicionário de dados para unificar os 4 anos de aplicação das provas (ex: a dimensão de questionário foi carregada manualmente após criteriosa avaliação, pois a descrição constava apenas no dicionário de dados e a pergunta – ou até mesmo as opções de resposta – podiam ser alteradas entre diferentes anos).

Com as dimensões populadas, bastou realizar a consulta para gravação na tabela fato. Essa consulta envolve diversas tabelas (*joins*) para buscar os IDs dos campos textuais que estão nas dimensões.

### 3.4. BI

Existem diversas ferramentas de BI e visualização de dados no mercado, tais como Power BI (Microsoft), Tableau (Salesforce), Qlik Sense, Looker (Google), Sisense, Cognos Analytics (IBM) e SAP Analytics Cloud. Para a elaboração do painel de BI discutido nesse trabalho escolheu-se utilizar a ferramenta Tableau, líder no segmento segundo o Gartner Magic Quadrant (Figura 3), que avalia os 21 principais fornecedores de inteligência de negócios sobre a integridade de sua visão e sua capacidade de execução [Carvalho e Queiroz, 2020].

O Tableau tem duas categorias de produtos: o desenvolvimento e o compartilhamento. O desenvolvimento, que envolve a criação de dashboards, painéis,

geração de relatórios e visualização de dados foi realizado no Tableau Desktop, enquanto para o compartilhamento optou-se pelo Tableau Public, que “é uma plataforma [web] gratuita para explorar, criar e compartilhar publicamente visualizações de dados on-line” [Tableau], na qual qualquer um pode ter acesso por meio de um navegador, sem demandar cadastro ou instalação de software adicional.



Figura 3. Gartner Magic Quadrant (2023).

O painel desenvolvido conta com duas páginas, sendo a primeira delas uma “capa” (Figura 4), cuja função é receber o usuário, apresentar a ele as principais características do Exame, funcionalidades básicas de como navegar pela ferramenta, e um contexto geral do assunto.

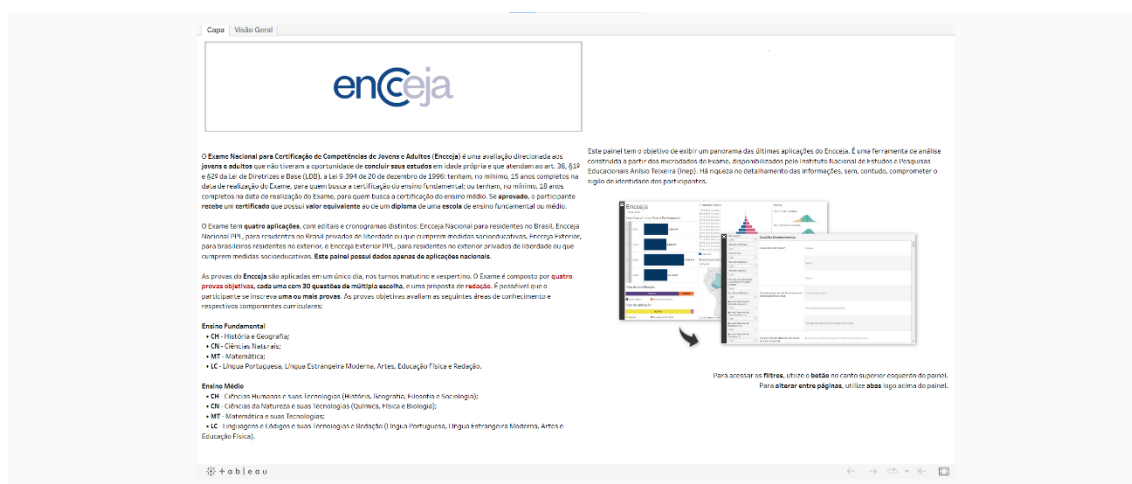


Figura 4. “Capa”, página do BI desenvolvido.

A segunda página, Visão Geral (Figura 5), traz as visualizações propriamente ditas. São 10 gráficos e uma tabela dinâmica, a saber:



- Quantidade de participantes Encceja, por ano: gráfico de área que informa tanto a quantidade de participantes do Encceja em cada ano de aplicação, como a assiduidade (presentes, ausentes e eliminados durante o Exame). É possível extrair desse gráfico uma dimensão evolutiva do quantitativo de pessoas que têm se candidatado para o Encceja a cada ano e se elas estão, de fato, realizando as provas (vale destacar que a participação no Encceja é gratuita);
- Tipo de certificação: gráfico de barra empilhada horizontalmente que compara a proporção de solicitantes de certificação para o ensino médio e fundamental (o conteúdo das provas é distinto para cada certificação);
- Tipo de aplicação: gráfico de barra empilhada horizontalmente que compara a proporção entre aplicações regulares e para pessoas privadas de liberdade;
- Tabela com percentual de inscritos por prova: informa o percentual de inscritos em cada uma das provas do Exame, uma vez que o candidato não é obrigado a se inscrever em todas;
- Pirâmide etária: informa a quantidade de participantes em cada faixa etária, por sexo;
- Percentual de inscritos em relação à população, por estado: mapa que informa a proporção de pessoas que realizaram o Encceja em relação à população média do estado no período. Quanto mais escura a cor no estado, maior a proporção de pessoas que se aplicou ao Exame;
- Quatro histogramas para notas, um para cada prova do Exame: traz as notas dos participantes agrupadas em intervalos de dois pontos. As cores ajudam a distinguir os aprovados dos reprovados, de forma que esse conjunto permite comparar não apenas o desempenho dos candidatos em determinada prova, mas também entre provas e traçar um panorama de possíveis deficiências educacionais;
- Percentual de ocorrências na redação: gráfico de barras horizontais que classifica as condições das redações entregues pelos participantes das provas de LC do Encceja.

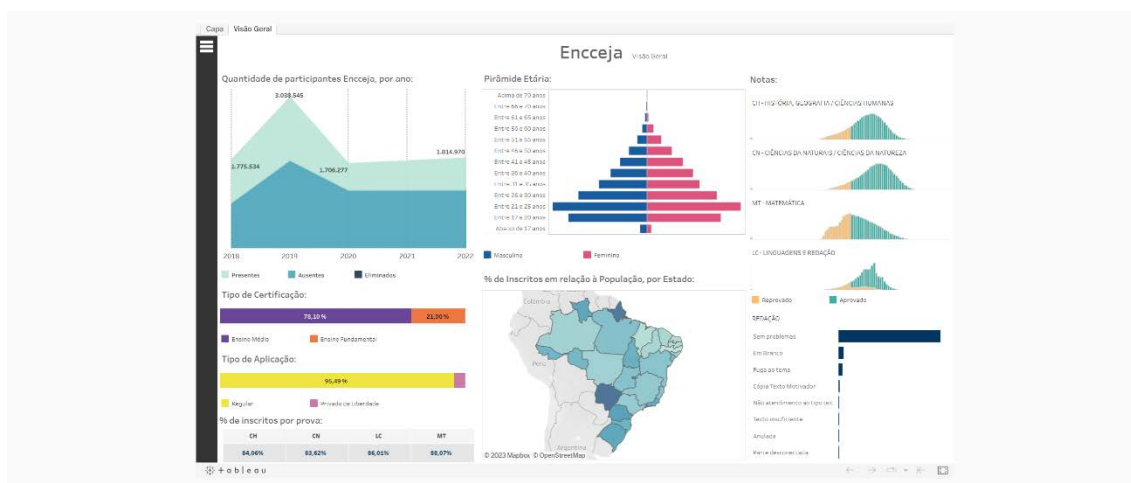


Figura 5. Visão Geral, página do BI desenvolvido.

O painel conta, ainda, com um botão lateral que permite acessar filtros mais detalhados (Figura 6), em especial para questões socioeconômicas. Por meio delas, é possível selecionar grupos específicos conforme as respostas dadas no questionário e observar o comportamento de alteração nos gráficos, habilitando o usuário a uma análise mais profunda de setores da sociedade.

Sobre o painel, vale destacar ainda que a escolha de cores considerou paletas propostas por Wong (2011), Ito (2002) e Tol (2012), acessíveis a pessoas com os tipos mais comuns de daltonismo, como protanopia, tritanopia e deuteranopia.

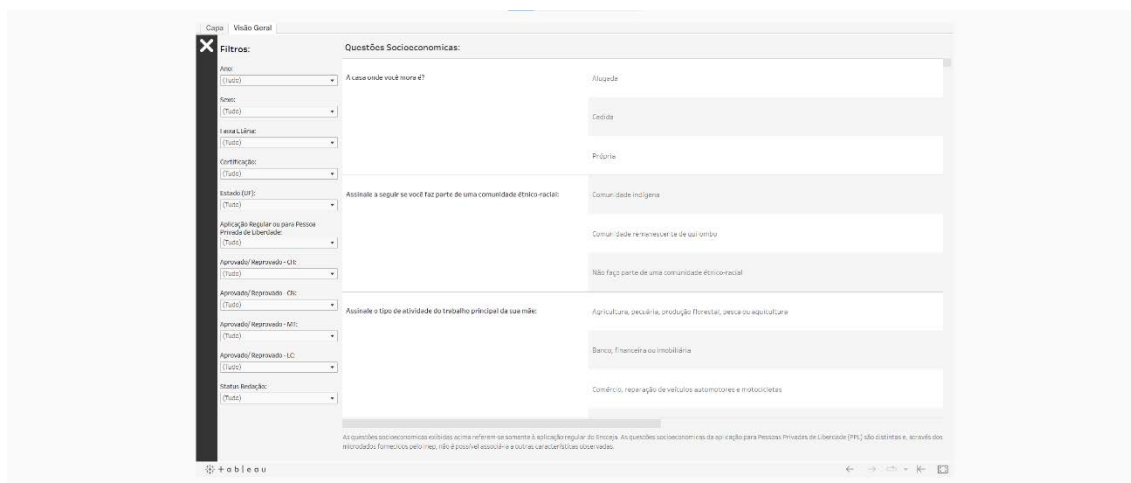


Figura 6. Detalhamento dos filtros na página Visão Geral.

#### 4. Desafios Encontrados

O principal desafio se caracterizou pela impossibilidade de relacionar o questionário socioeconômico dos candidatos nacionais com restrição de liberdade às suas provas, o que implicou na anulação desses campos para esse grupo de candidatos. Observou-se também a falta de padronização na nomeação dos arquivos entre os anos e mudanças nas estruturas das questões socioeconômicas.

#### 5. Conclusão

Este trabalho explorou a importância da estruturação de um DW e construção de um BI com os dados abertos dos últimos cinco anos de aplicações do Encceja disponibilizados pelo Inep. O Encceja é um exame de grande relevância para a educação brasileira, e fornece informações valiosas sobre o desempenho dos alunos em diferentes áreas de conhecimento.

Através da estruturação de um DW, é possível centralizar e organizar os dados, proporcionando uma visão abrangente e integrada do desempenho dos alunos ao longo do tempo e em diferentes regiões do país. A análise desses dados em um painel BI pode fornecer insights cruciais para a melhoria do sistema educacional brasileiro.

Além disso, a utilização de tecnologias de computação em nuvem para a estruturação do DW e publicação do BI oferece vantagens significativas. A computação em nuvem proporciona escalabilidade, flexibilidade e disponibilidade dos dados, permitindo que a infraestrutura seja ajustada de acordo com as necessidades da análise.

A análise dos dados do Enceja por meio de um DW estruturado e um BI com visualizações de fácil compreensão disponível publicamente pode contribuir para uma tomada de decisão mais informada no campo da educação e fomentar pesquisas acadêmicas sobre o tema. Gestores educacionais, professores, pedagogos e pesquisadores podem identificar tendências, padrões e lacunas de desempenho, direcionando recursos e esforços para melhorar a qualidade do ensino e o desenvolvimento dos alunos.

No entanto, é importante destacar que a estruturação de um DW e a análise de dados educacionais envolvem desafios, como a qualidade e a integração dos dados, além da privacidade e a segurança das informações. Portanto, é fundamental adotar práticas adequadas de governança de dados e garantir o cumprimento das políticas de privacidade e proteção de dados, em especial à LGPD.

Espera-se que o BI desenvolvido seja uma ferramenta de auxílio na compreensão multifacetada dos diversos aspectos do Enceja e da educação básica no Brasil, e que novas discussões sejam levantadas no sentido de entender e propor possibilidades de melhoria do sistema básico de ensino.

## Referências

- Associação Brasileira de Normas Técnicas. “NBR 6023: informação e documentação, referências, elaboração”. Rio de Janeiro, 2002. 24 p.
- Awad, M.; Al Redhaei, A.; Fraihat, S. “Using business intelligence to analyze road traffic accidents”. Proceedings of the Central and Eastern European eDem and eGov Days. Nova Iorque, 2022; p. 83-92.
- Brasil. Lei no 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, 2020.
- Campos, Breno A. e Campos, Paulemir G. “Análise de Dados usando as Ferramentas de Business Intelligence Tableau e Qlik Sense”. Conference contribution.
- Gartner. Analytics and Business Intelligence Platforms Reviews and Ratings. 2023. <<https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms>>. Acesso em: 26/11/2023.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). “Microdados do ENCCEJA 2018”. [online]. Brasília: Inep, 2022. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/encceja>>.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). “Microdados do ENCCEJA 2019”. [online]. Brasília: Inep, 2022. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/encceja>>.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). “Microdados do ENCCEJA 2020”. [online]. Brasília: Inep, 2022. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/encceja>>.
- Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). “Microdados do ENCCEJA 2022”. [online]. Brasília: Inep, 2022. Disponível

em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/microdados/enceja>>.

- Kimball, R. et al. “The Data Warehouse Lifecycle Toolkit”. 2ª Ed. Wiley: Indianapolis, 2008. 636 p.
- Kimball, R. e Ross, M. “The Data Warehouse Toolkit: The Definitive Guide” to Dimensional Modeling (3rd Edition). Wiley: Indianapolis, 2013. 720p.
- \_\_\_\_\_. The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence. Wiley: Indianapolis, 2010. 565 p.
- Knaflic, C. N. “Storytelling com dados: Um guia sobre visualização de dados para profissionais de negócios”. 2ª Ed. Rio de Janeiro: Alta Books, 2019. 241 p.
- Okabe, M., e K. Ito. “How to make figures and presentations that are friendly to Colorblind people”. Color Universal Design (CUD). 2002.
- Queiroz, Rosa M. D. e Carvalho, Lucas F. B. “Business Intelligence Uma visão teórica sobre BI e suas estratégias”. Tecnologias em Projeção. 2020: 44-58.
- Tableau [online]. Disponível em: <<https://www.tableau.com>>. Acesso em: 28/11/2023.
- Tol, P. Colour Schemes. SRON Technical Note, Doc. no. SRON/EPS/TN/09-002. 2012.
- Wong, B. “Color blindness”. Nature Methods 8.6. 2011.
- Zaccarelli, S. “Ebook: Aplicando Analytics para geração de Insights com dados do Exame Nacional do Ensino Médio”. [s.l.]: [s.n.], 2020. 21 p.