



**UFOP**

Universidade Federal  
de Ouro Preto

**Universidade Federal de Ouro Preto  
Instituto de Ciências Exatas e Aplicadas  
Departamento de Computação e Sistemas**

**Potencialidade do Business Intelligence  
na Gestão da Atenção Primária em  
Saúde: Um Enfoque Prático**

**Stephane Matos Oliveira**

**TRABALHO DE  
CONCLUSÃO DE CURSO**

ORIENTAÇÃO:

Rafael Alexandre

COORIENTAÇÃO:

**Dezembro, 2025  
João Monlevade–MG**

**Stephane Matos Oliveira**

**Potencialidade do Business Intelligence na  
Gestão da Atenção Primária em Saúde: Um  
Enfoque Prático**

Orientador: Rafael Alexandre

Coorientador:

Monografia apresentada ao curso de Sistemas de Informação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

**Universidade Federal de Ouro Preto**

**João Monlevade**

**Dezembro de 2025**



## FOLHA DE APROVAÇÃO

**Stephane Matos Oliveira**

### **Potencialidade do Business Intelligence na Gestão da Atenção Primária em Saúde: Um Enfoque Prático**

Monografia apresentada ao Curso de Engenharia da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Engenharia da Computação

Aprovada em 26 de fevereiro de 2026

#### Membros da banca

Doutor - Rafael Frederico Alexandre - Orientador - (Universidade Federal de Ouro Preto)

Doutor - Darlan Nunes de Brito - (Universidade Federal de Ouro Preto)

Mestre - Josemar Alves Caetano - (Universidade Federal de Ouro Preto)

Rafael Frederico Alexandre, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 30/04/2026



Documento assinado eletronicamente por **Rafael Frederico Alexandre, PROFESSOR DE MAGISTERIO SUPERIOR**, em 30/04/2026, às 13:36, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site [http://sei.ufop.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **1098909** e o código CRC **E4F6DCD1**.

*Este trabalho é dedicado aos meus avós: Maria, Maria das Graças, Sebastião e Adão*

# Agradecimentos

Agradeço primeiramente a Deus, por me conceder força, sabedoria e perseverança durante toda esta jornada acadêmica.

Agradeço à minha família, especialmente aos meus pais, Renilson e Marisa, por todo o sacrifício, amor incondicional e pela educação que me proporcionaram; cada conquista minha é reflexo do esforço e da dedicação de vocês. Às minhas irmãs, Isadorah e Danniele, minha gratidão pela cumplicidade, pelo apoio constante e por tornarem minha caminhada mais leve, sendo fundamentais nos momentos difíceis, sempre com uma palavra de incentivo e carinho. Obrigada por vibrarem comigo a cada etapa concluída. Esta vitória é nossa.

Ao meu orientador, Professor Rafael Alexandre, pela dedicação, paciência e valiosas orientações que foram fundamentais para o desenvolvimento deste trabalho.

À Universidade Federal de Ouro Preto e ao corpo docente do curso de Engenharia da Computação, pelos ensinamentos e pela formação sólida que me proporcionaram ao longo desses anos. Aos meus amigos e colegas de curso, pelas discussões enriquecedoras, apoio mútuo e companheirismo durante toda a graduação.

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho. Minha sincera gratidão.

*“A informação é o petróleo do século XXI, e a análise é o motor de combustão.”*

— Peter Sondergaard,  
*Vice-Presidente Sênior da Gartner.*

# Resumo

A crescente digitalização dos processos no setor público tem ampliado significativamente o volume de dados gerados na área da saúde. No Brasil, o Sistema Único de Saúde (SUS) dispõe de diversos sistemas de informação que subsidiam o planejamento e a gestão de políticas públicas, destacando-se o Cadastro Nacional de Estabelecimentos de Saúde (CNES), que reúne informações sobre estabelecimentos, leitos e profissionais em todo o território nacional. Entretanto, a simples disponibilidade dos dados não é suficiente para garantir sua utilização efetiva. Para que a informação gere valor, é necessário que os dados sejam organizados, padronizados e estruturados de forma adequada ao contexto analítico. Nesse sentido, este trabalho teve como objetivo refatorar os painéis de Business Intelligence baseados nos dados do CNES, visando ao aprimoramento da clareza visual, da usabilidade, do desempenho e da confiabilidade das informações apresentadas. A metodologia adotada envolveu a construção de um pipeline de dados ELT (*Extract, Load, Transform*) estruturado em Python, responsável por extrair, carregar e transformar as informações provenientes da plataforma ElastiCNES. Os dados foram organizados em camadas lógicas (RAW, Staging e Dimensional) em um banco de dados PostgreSQL, aplicando-se técnicas de limpeza, padronização, normalização e modelagem dimensional segundo os princípios de Kimball. A partir dessa base estruturada, foram desenvolvidos dashboards no Power BI organizados em módulos temáticos (Leitos, Panorama Geral, Profissionais), com foco em clareza informacional, interatividade e experiência do usuário, abrangendo o período de 2020 a 2025. A solução desenvolvida evidenciou melhorias significativas em relação aos painéis originais do ElastiCNES, tanto em aspectos de usabilidade quanto de qualidade informacional e capacidade analítica. Este trabalho reforça o potencial do Business Intelligence como instrumento de apoio à gestão, ao planejamento e à tomada de decisão no setor público de saúde, demonstrando que a qualidade da visualização é tão relevante quanto a qualidade dos dados subjacentes.

**Palavras-chave:** Business Intelligence. Saúde Pública. CNES. Visualização de Dados. Modelagem Dimensional. ELT. Power BI.

# Abstract

The increasing digitalization of processes in the public sector has significantly expanded the volume of data generated in healthcare. In Brazil, the Unified Health System (SUS) has several information systems that support the planning and management of public policies, notably the National Registry of Health Establishments (CNES), which compiles information on facilities, beds, and professionals throughout the national territory. However, data availability alone is not sufficient to ensure its effective use. For information to generate value, data must be organized, standardized, and structured appropriately for analytical contexts. In this regard, this work aimed to refactor Business Intelligence dashboards based on CNES data, targeting improvements in visual clarity, usability, performance, and information reliability. The adopted methodology involved building an ELT (*Extract, Load, Transform*) data pipeline structured in Python, responsible for extracting, loading, and transforming information from the ElastiCNES platform. Data were organized into logical layers (RAW, Staging, and Dimensional) in a PostgreSQL database, applying cleaning, standardization, normalization, and dimensional modeling techniques according to Kimball principles. From this structured foundation, Power BI dashboards were developed and organized into thematic modules (Beds, General Overview, Professionals), focusing on informational clarity, interactivity, and user experience, covering the period from 2020 to 2025. The developed solution demonstrated significant improvements over the original ElastiCNES dashboards in terms of usability, informational quality, and analytical capability. This work reinforces the potential of Business Intelligence as a supporting tool for management, planning, and decision-making in the public health sector, demonstrating that visualization quality is as relevant as the quality of underlying data.

**Keywords:** Business Intelligence. Public Health. CNES. Data Visualization. Dimensional Modeling. ELT. Power BI.

# Lista de ilustrações

Figura 1 – Painel Inicial CNES . . . . .	26
Figura 2 – Painel Panorama Geral . . . . .	27
Figura 3 – Painel Panorama Geral . . . . .	27
Figura 4 – Painel Leitos . . . . .	29
Figura 5 – Painel Leitos . . . . .	30
Figura 6 – Painel Leitos . . . . .	30
Figura 7 – Painel Profissionais . . . . .	31
Figura 8 – Fluxo do pipeline ELT e organização das camadas analíticas . . . . .	38
Figura 9 – Fluxo de persistência dos dados . . . . .	42
Figura 10 – Diagrama do pipeline de dados e organização das camadas analíticas . . . . .	44
Figura 11 – Modelo dimensional implementado no Power Bi . . . . .	54
Figura 12 – Brainstorm Inicial - Figma . . . . .	56
Figura 13 – Desing System . . . . .	57
Figura 14 – Mockup Figma . . . . .	58
Figura 15 – Mockup Figma 2 . . . . .	58
Figura 16 – Interface do menu inicial do sistema ElastiCNES . . . . .	60
Figura 17 – Efeito <i>hover</i> nos elementos do menu inicial . . . . .	61
Figura 18 – Sistema de filtros multidimensionais do <i>dashboard</i> Panorama Geral . . . . .	62
Figura 19 – <i>Dashboard</i> Panorama Geral . . . . .	63
Figura 20 – <i>Dashboard</i> Panorama Geral 2 . . . . .	64
Figura 21 – Distribuição de estabelecimentos Panorama Geral . . . . .	64
Figura 22 – Painel Panorama Geral . . . . .	65
Figura 23 – Sistema de filtros do <i>dashboard</i> de Leitos . . . . .	66
Figura 24 – <i>Dashboard</i> de Leitos . . . . .	68
Figura 25 – <i>Dashboard</i> de Leitos 2 . . . . .	69
Figura 26 – Painel Leitos . . . . .	70
Figura 27 – Sistema de filtros do <i>dashboard</i> de Profissionais . . . . .	71
Figura 28 – <i>Dashboard</i> de Profissionais . . . . .	73
Figura 29 – <i>Dashboard</i> de Profissionais 2 . . . . .	74
Figura 30 – Tela da Área Informativa do sistema . . . . .	75
Figura 31 – Tela de status para painéis não implementados . . . . .	76

# Lista de tabelas

Tabela 1 – Análise comparativa entre soluções existentes e a proposta deste trabalho	24
Tabela 2 – Volumetria dos dados extraídos por domínio . . . . .	39
Tabela 3 – Estrutura da tabela <code>stg_estabelecimentos</code> (principais campos) . . .	47
Tabela 4 – Estrutura da tabela <code>stg_leitos</code> (principais campos) . . . . .	47
Tabela 5 – Estrutura da tabela <code>stg_profissionais</code> (principais campos) . . . . .	48
Tabela 6 – Estrutura da tabela fato <code>fct_leitos</code> . . . . .	49
Tabela 7 – Estrutura da dimensão <code>dim_dates</code> . . . . .	50
Tabela 8 – Estrutura completa da dimensão <code>dim_leitos</code> . . . . .	51
Tabela 9 – Estrutura completa da dimensão <code>dim_panorama_geral</code> . . . . .	52
Tabela 10 – Estrutura completa da dimensão <code>dim_profissionais</code> . . . . .	53
Tabela 11 – Evolução da estrutura de dados por camada . . . . .	53
Tabela 12 – Comparativo de tempo de execução de tarefas entre o sistema proposto e o ElastiCNES . . . . .	84

# Lista de abreviaturas e siglas

API	Application Programming Interface
BI	<i>Business Intelligence</i>
CNES	Cadastro Nacional de Estabelecimentos de Saúde
DAX	<i>Data Analysis Expressions</i>
ELT	<i>Extract, Load, Transform</i>
JSON	<i>JavaScript Object Notation</i>
KPI	<i>Key Performance Indicators</i>
SQL	<i>Structured Query Language</i>
SUS	Sistema Único de Saúde
UFOP	Universidade Federal de Ouro Preto

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
1.1	Objetivos	16
1.2	Metodologia	17
1.3	Organização do trabalho	18
<b>2</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>19</b>
2.1	Dashboards Analíticos	19
2.2	Engenharia de Dados e Business Intelligence	19
2.3	Informação e Dados Públicos de Saúde no Brasil	20
2.4	Ferramentas e Tecnologias	21
2.4.1	Python e Web Scraping	21
2.4.2	PostgreSQL	21
2.4.3	Microsoft Power BI	22
2.5	Trabalhos relacionados	22
<b>3</b>	<b>IDENTIFICAÇÃO DO PROBLEMA</b>	<b>25</b>
3.1	Análise Crítica dos Painéis ElastiCNES Existentes	25
3.1.1	Panorama Geral	26
3.1.2	Painel Leitos	29
3.1.3	Painel Profissionais	31
3.2	Limitações do ElastiCNES	32
3.3	Definição do Problema	33
<b>4</b>	<b>DESENVOLVIMENTO</b>	<b>35</b>
4.1	Arquitetura da Solução Analítica	35
4.1.1	Visão Geral da Arquitetura	35
4.1.2	Componentes da Solução	35
4.2	Pipeline de Dados e Processo de ELT	36
4.2.1	Estratégia de Extração dos Dados do CNES	38
4.2.2	Engenharia Reversa da Interface Web	38
4.2.3	Volumetria dos Dados Extraídos	39
4.2.4	Pré-processamento e Carga ETL	40
4.2.5	Orquestração da carga	41
4.2.6	Tabelas da camada RAW	43
4.2.7	Controle de qualidade e rastreabilidade	43
4.2.8	Tratamento e Padronização dos Dados	44

4.2.8.1	Transformações Aplicadas . . . . .	45
4.2.8.2	Validações Específicas por Domínio . . . . .	45
4.2.8.3	Camada Staging . . . . .	46
4.2.9	Modelagem dos Dados . . . . .	48
4.2.9.1	Tabelas Fato e Dimensões . . . . .	48
4.2.10	Relacionamentos e Integridade Referencial . . . . .	54
<b>4.3</b>	<b>Projeto de Dashboard e Visualizações . . . . .</b>	<b>55</b>
4.3.1	Mockups . . . . .	55
4.3.2	Modulos de Visualização e Análise . . . . .	59
<b>5</b>	<b>RESULTADOS . . . . .</b>	<b>60</b>
5.0.1	Menu Inicial do Sistema . . . . .	60
5.0.2	Dashboard Panorama Geral . . . . .	61
5.0.2.1	Filtros e Dimensões de Análise . . . . .	61
5.0.2.2	Indicadores e Visuais . . . . .	62
5.0.2.3	Tipologia de Estabelecimentos . . . . .	63
5.0.3	Dashboard de Leitos . . . . .	66
5.0.3.1	Filtros e Sistema de Navegação . . . . .	66
5.0.3.2	Indicadores Principais e Visuais . . . . .	67
5.0.3.3	Tipologia de Leitos . . . . .	67
5.0.4	Dashboard de Profissionais . . . . .	69
5.0.4.1	Sistema de Filtragem . . . . .	71
5.0.4.2	Indicadores Principais e Visuais . . . . .	71
5.0.4.3	Composição de Categoria Profissional . . . . .	72
5.0.4.4	Análise de Carga Horária por Estado . . . . .	73
5.0.5	Área Informativa . . . . .	74
5.0.6	Painéis em Construção . . . . .	75
<b>5.1</b>	<b>Estudos de Caso e Aplicações Práticas . . . . .</b>	<b>76</b>
5.1.1	Caso 1: Planejamento de Expansão da Rede SUS . . . . .	76
5.1.1.1	Contexto . . . . .	76
5.1.1.2	Perguntas Gerenciais . . . . .	76
5.1.1.3	Análise e Decisão . . . . .	77
5.1.2	Caso 2: Dimensionamento de Recursos Humanos . . . . .	78
5.1.2.1	Contexto . . . . .	78
5.1.2.2	Perguntas Gerenciais . . . . .	78
5.1.2.3	Análise e Decisão . . . . .	78
5.1.3	Caso 3: Análise de Capacidade Hospitalar para Emergências . . . . .	79
5.1.3.1	Contexto . . . . .	79
5.1.3.2	Perguntas Gerenciais . . . . .	79
5.1.3.3	Análise e Decisão . . . . .	80

5.1.4	Caso 4: Avaliação de Inequidades Regionais . . . . .	80
5.1.4.1	Contexto . . . . .	80
5.1.4.2	Perguntas Gerenciais . . . . .	80
5.1.4.3	Análise e Decisão . . . . .	81
5.1.5	Importância dos Dashboards para Gestão em Saúde . . . . .	82
5.1.6	Validação Comparativa de Usabilidade . . . . .	83
<b>6</b>	<b>CONCLUSÃO</b> . . . . .	<b>85</b>
6.0.1	Perspectivas Futuras e Considerações Finais . . . . .	85
	<b>REFERÊNCIAS</b> . . . . .	<b>87</b>

# 1 Introdução

A saúde pública no Brasil enfrenta desafios estruturais históricos, decorrentes de um processo de expansão do sistema que nem sempre foi acompanhado pela infraestrutura e pelos recursos necessários para garantir acesso universal e equitativo. O Sistema Único de Saúde (SUS), criado pela Constituição de 1988, representa uma das maiores conquistas sociais do país e constitui uma das maiores estruturas públicas de saúde do mundo, responsável pelo atendimento de mais de 200 milhões de cidadãos (PAIM *et al.*, 2011). No entanto, a gestão dessa rede complexa de estabelecimentos, profissionais e serviços assistenciais permanece como um desafio permanente para gestores públicos em todas as esferas governamentais.

A qualidade da gestão em saúde pública está diretamente relacionada à capacidade de tomar decisões assertivas e baseadas em evidências. Estudos demonstram que decisões inadequadas na alocação de recursos, no dimensionamento de equipes e na distribuição de leitos impactam negativamente a eficiência do sistema e, conseqüentemente, a qualidade do atendimento à população (CASTRO *et al.*, 2011). Mais especificamente, os desafios da gestão em saúde no Brasil apresentam características particulares, visto que as deficiências do sistema não se distribuem uniformemente, afetando especialmente regiões mais vulneráveis e populações em situação de maior risco social (VIACAVA; BELLIDO, 2016).

A evolução de iniciativas como a Lei de Acesso à Informação (LAI) e programas de governo digital tornou público um volume sem precedentes de dados sobre o sistema de saúde brasileiro (Brasil, 2011a), enquanto o avanço das tecnologias de Business Intelligence criou as condições técnicas necessárias para transformar essas informações em conhecimento estratégico aplicável à tomada de decisão (KIMBALL; ROSS, 2013).

Essa confluência entre disponibilidade de dados e necessidade de soluções mais inteligentes abriu caminho para uma gestão pública mais moderna e eficiente. Sistemas de informação em saúde, como o Cadastro Nacional de Estabelecimentos de Saúde (CNES), consolidaram-se como fontes essenciais de dados para o planejamento e monitoramento das políticas públicas (SAÚDE, 2024). No entanto, a simples disponibilidade dos dados não garante sua utilização efetiva. A capacidade de organizar, estruturar e transformar esses dados em informações acionáveis tem se mostrado fundamental para auxiliar gestores na tomada de decisões mais precisas, ágeis e baseadas em evidências (DAVENPORT; HARRIS, 2007).

Nesse contexto, o avanço das tecnologias de Business Intelligence criou as condições técnicas necessárias para transformar essas informações em conhecimento estratégico aplicável à tomada de decisão (KIMBALL; ROSS, 2013). Essa confluência entre dispo-

nibilidade de dados e necessidade de soluções mais inteligentes abriu caminho para uma gestão pública mais moderna e eficiente, viabilizada por ferramentas como linguagens de programação para extração automatizada de dados, bancos de dados relacionais para armazenamento estruturado e plataformas de visualização analítica para apresentação interativa das informações (DAVENPORT; HARRIS, 2007).

Diante do exposto, o desenvolvimento de soluções analíticas baseadas em dados do CNES representa uma contribuição significativa para a modernização da gestão pública em saúde. A aplicação de boas práticas de engenharia de dados e Business Intelligence ao contexto específico da saúde pública brasileira pode não apenas otimizar processos decisórios locais, mas também servir como referência para outras instituições que enfrentam desafios similares na análise e utilização de dados governamentais. A relevância deste estudo reside na demonstração prática de como a integração entre engenharia de dados, modelagem analítica e Business Intelligence pode contribuir efetivamente para a melhoria da análise de informações em saúde pública, evidenciando o potencial dessas tecnologias como instrumentos estratégicos de apoio à gestão, ao planejamento e à tomada de decisão no setor público.

## 1.1 Objetivos

O objetivo principal deste trabalho consiste em desenvolver uma solução analítica completa para os dados do Cadastro Nacional de Estabelecimentos de Saúde (CNES), compreendendo a construção de um pipeline ELT estruturado em camadas lógicas, modelagem dimensional segundo os princípios de Kimball e o desenvolvimento de *dashboards* interativos no Power BI visando superar as limitações de usabilidade, desempenho e confiabilidade analítica identificadas nas ferramentas governamentais existentes, em especial o ElastiCNES.

Para alcançar esse propósito, foram definidos os seguintes objetivos específicos:

- Analisar o estado atual dos painéis de Business Intelligence baseados nos dados do CNES, identificando limitações relacionadas a desempenho, consistência dos indicadores, usabilidade e manutenibilidade;
- Projetar e implementar um pipeline de dados estruturado, contemplando processos de extração, carga e transformação (ELT).
- Desenvolver dashboards de Business Intelligence no Power BI, como evolução dos painéis existentes, organizando visualizações, métricas e filtros com foco em desempenho, clareza analítica e melhoria da experiência do usuário;

- Avaliar os resultados obtidos a partir da solução proposta, comparando os painéis desenvolvidos com as visualizações originais quanto à qualidade das análises e ao suporte à tomada de decisão.

## 1.2 Metodologia

Este trabalho caracteriza-se como uma pesquisa aplicada, de abordagem qualitativa e caráter exploratório, uma vez que busca propor uma solução prática para a utilização de dados públicos de saúde na construção de *dashboards* analíticos. Quanto aos procedimentos técnicos, a pesquisa configura-se como um estudo de caso, por analisar de forma aprofundada a aplicação dos dados do Cadastro Nacional de Estabelecimentos de Saúde (CNES) no desenvolvimento de painéis analíticos, considerando um contexto real de uso.

A metodologia foi estruturada em etapas sequenciais, visando garantir coerência, rastreabilidade e reprodutibilidade no desenvolvimento da solução, apresentadas a seguir:

- **Levantamento e Compreensão dos Dados:** Inicialmente, foi realizada a investigação da plataforma ElastiCNES para compreender a disponibilidade dos dados, métodos de extração, estrutura das informações (estabelecimentos, leitos, profissionais), formato de disponibilização (JSON, CSV), periodicidade de atualização e requisitos técnicos para acesso automatizado. Esta etapa foi fundamental para definir a estratégia de extração e identificar as limitações e possibilidades da fonte de dados.
- **Análise Crítica dos Painéis Originais:** Avaliação sistemática dos painéis existentes segundo critérios de usabilidade (organização de filtros, navegação), performance (tempo de carregamento, responsividade), design visual (escolha de gráficos, hierarquia informacional) e confiabilidade analítica (consistência de métricas, tratamento de valores ausentes) identificando limitações e melhorias possíveis. Documentação dos indicadores-chave (KPIs) e dimensões analíticas para subsidiar a refatoração.
- **Desenvolvimento da Arquitetura de Dados:** Construção de pipeline ELT estruturado em camadas lógicas (RAW, Staging, Dimensional), utilizando Python para extração via web scraping do ElastiCNES, PostgreSQL para armazenamento e SQL para transformações. Aplicação de técnicas de limpeza, normalização e padronização dos dados, seguida de modelagem dimensional segundo princípios de Kimball, com construção de tabelas de fatos (leitos, profissionais, estabelecimentos) e dimensões. A solução concentrou transformações na camada de dados, reduzindo complexidade na visualização e garantindo performance e manutenibilidade.
- **Desenvolvimento da Solução Tecnológica:** A partir dos dados tratados e modelados, foram desenvolvidos dashboards analíticos no Power BI, estruturados segundo

a arquitetura proposta. Esta etapa concentrou-se na definição das visualizações, métricas e filtros, bem como na simplificação das medidas analíticas, visando à melhoria da clareza informacional, da usabilidade e do desempenho dos painéis para diferentes perfis de usuários.

- **Avaliação qualitativa da solução e resultados:** Por fim, foi realizada uma avaliação qualitativa dos painéis refatorados em comparação aos painéis originais. Essa avaliação considerou critérios como desempenho percebido, clareza visual, facilidade de manutenção e confiabilidade das informações apresentadas. A análise permitiu verificar os impactos da adoção de um pipeline de dados estruturado e da reorganização das visualizações no contexto da utilização dos dados do CNES.

### 1.3 Organização do trabalho

Este trabalho está organizado em seis capítulos, além desta introdução.

O Capítulo 2, intitulado Revisão Bibliográfica, apresenta os principais conceitos teóricos que fundamentam o estudo. São abordados temas relacionados a dashboards analíticos, processos de Extract, Load and Transform (ELT), Business Intelligence, ferramentas de Business Intelligence e trabalhos correlatos, fornecendo a base conceitual necessária para o desenvolvimento da pesquisa.

O Capítulo 3, denominado Identificação do Problema, contextualiza o cenário de uso dos dados do Cadastro Nacional de Estabelecimentos de Saúde (CNES) e do Sistema Único de Saúde (SUS), destacando as principais limitações observadas na utilização desses dados em painéis analíticos e caracterizando o problema abordado neste trabalho.

O Capítulo 4, referente ao Desenvolvimento, descreve as etapas práticas da solução proposta. Inicialmente, é apresentada a análise dos painéis ElasticCNES, seguida das etapas de extração, coleta e acesso aos dados do CNES. O capítulo também contempla o tratamento dos dados, a definição dos relacionamentos e o desenvolvimento dos dashboards refatorados.

O Capítulo 5, Resultados, apresenta os resultados obtidos a partir da aplicação da metodologia proposta, evidenciando os impactos da refatoração dos painéis e da adoção de um pipeline de dados estruturado, com foco em desempenho, clareza e confiabilidade das informações.

Por fim, o Capítulo 6, Conclusão, reúne as considerações finais do trabalho, destacando as principais contribuições alcançadas, as limitações identificadas ao longo do estudo e as possibilidades para trabalhos futuros.

## 2 Revisão bibliográfica

Este capítulo apresenta a fundamentação teórica que embasa o desenvolvimento deste trabalho, discutindo conceitos relacionados à visualização de dados, arquitetura analítica e o contexto dos dados públicos de saúde no Brasil.

### 2.1 Dashboards Analíticos

Segundo [Few \(2013\)](#), dashboards analíticos consistem em instrumentos visuais projetados para apresentar informações de forma sintetizada e significativa, permitindo o monitoramento contínuo e a análise de indicadores de desempenho. [Eckerson \(2010\)](#) destaca que dashboards bem projetados contribuem para o alinhamento entre métricas operacionais e estratégicas, favorecendo processos decisórios baseados em evidências.

Na saúde pública, dashboards assumem papel crítico ao possibilitar monitoramento integrado de indicadores assistenciais, administrativos e operacionais. Contudo, sua efetividade está condicionada à qualidade dos dados subjacentes e à solidez da arquitetura analítica ([KIMBALL; CASERTA, 2004](#); [SILVA; ALMEIDA, 2020](#)). Estudos sobre percepção visual reforçam que a escolha adequada de representações gráficas influencia diretamente a interpretação de informações quantitativas ([CLEVELAND; MCGILL, 1984](#); [TUFTE, 2001](#)), sendo componente essencial para tomadas de decisão assertivas em ambientes críticos como a saúde pública.

### 2.2 Engenharia de Dados e Business Intelligence

A construção de soluções analíticas efetivas depende fundamentalmente da arquitetura de dados subjacente. [Kimball e Ross \(2013\)](#) enfatizam que uma arquitetura bem planejada, organizada em camadas lógicas (extração, transformação, armazenamento e consumo), é essencial para garantir desempenho, clareza e sustentabilidade em ambientes de Business Intelligence. Essa separação clara de responsabilidades contribui para maior controle de qualidade, redução de complexidade técnica e escalabilidade das soluções.

Na camada de transformação, pipelines de dados estruturados desempenham papel central ao estabelecer processos sistemáticos de limpeza, padronização e enriquecimento das informações. A abordagem ELT (*Extract, Load, Transform*), em que transformações ocorrem após o carregamento no ambiente analítico, tem ganhado destaque pela capacidade de aproveitar o poder computacional de bancos de dados modernos ([INMON, 2005](#)).

A modelagem dimensional constitui técnica fundamental para organização de

dados analíticos, estruturando informações em tabelas de fatos (métricas) e dimensões (contextos), facilitando consultas e análises multidimensionais. Segundo [Kimball e Ross \(2013\)](#), essa abordagem prioriza simplicidade, performance e compreensibilidade, permitindo que usuários não-técnicos naveguem pelos dados de forma intuitiva. Em ambientes de saúde pública, onde dados são complexos e volumosos, a modelagem dimensional adequada torna-se ainda mais crítica para viabilizar análises eficientes.

Ferramentas modernas de Business Intelligence, como o Power BI, consolidam essas práticas ao oferecer capacidades de modelagem, transformação e visualização integradas. Contudo, a documentação oficial da ferramenta ressalta que centralizar transformações na camada de dados, em vez da camada de visualização, contribui significativamente para ganhos de performance, manutenibilidade e clareza analítica ([MICROSOFT, 2024](#)). Essa abordagem é especialmente relevante em bases governamentais de grande escala, onde a qualidade da arquitetura impacta diretamente a confiabilidade das análises produzidas.

## 2.3 Informação e Dados Públicos de Saúde no Brasil

A disponibilização de dados abertos constitui pilar da modernização administrativa e da transparência governamental. Na saúde, possibilita acompanhamento de ações estatais, análises independentes e desenvolvimento de soluções analíticas ([Brasil, 2014](#)). Dados das atividades assistenciais somente adquirem valor gerencial quando organizados e interpretados sistematicamente, transformando-se em informação relevante para tomada de decisão ([DAVENPORT; PRUSAK, 1998](#)).

No Brasil, essa política é regulamentada pela Lei de Acesso à Informação (Lei nº 12.527/2011), que estabelece publicidade como regra e sigilo como exceção ([Brasil, 2011b](#)), e pela Política Nacional de Dados Abertos, que reforça publicação em formatos abertos e processáveis automaticamente ([Brasil, 2016](#)). O SUS produz grande volume de informações decorrentes de sua estrutura descentralizada e diversidade de serviços, fundamentais para coordenação das ações em diferentes níveis de governo ([Brasil, 2011c](#)), e essa disponibilização favorece iniciativas de Business Intelligence aplicadas à gestão pública ([FERRER-SAPENA; PESET; ALEIXANDRE-BENAVENT, 2011; DINIZ, 2010](#)).

O Cadastro Nacional de Estabelecimentos de Saúde (CNES) destaca-se como principal fonte sobre a rede assistencial brasileira, reunindo dados cadastrais de estabelecimentos amplamente utilizados para planejamento, regulação e financiamento ([Brasil, 2023a](#)). Contudo, a literatura aponta que disponibilização não garante utilização efetiva: desafios de qualidade, padronização e documentação podem limitar seu uso analítico ([PINTO; FREITAS; FIGUEIREDO, 2018](#)). Quando submetidos a processos adequados de tratamento, os dados do CNES oferecem elevado potencial para construção de indicadores e painéis analíticos que apoiam decisões baseadas em evidências.

## 2.4 Ferramentas e Tecnologias

A construção do pipeline de dados e dos *dashboards* desenvolvidos neste trabalho apoia-se em um conjunto de ferramentas amplamente adotadas na indústria e na academia para tarefas de engenharia de dados e Business Intelligence. Esta seção descreve cada uma delas, contextualizando sua relevância no escopo deste estudo.

### 2.4.1 Python e Web Scraping

Python é uma linguagem de programação de alto nível, de propósito geral, que se consolidou como a principal linguagem para tarefas de ciência de dados, engenharia de dados e automação (Python Software Foundation, 2023). Sua popularidade nesse domínio deve-se à sintaxe expressiva, à vasta disponibilidade de bibliotecas especializadas e à ampla adoção pela comunidade científica e industrial.

No contexto da extração de dados, o *web scraping* consiste na técnica de coleta automatizada de informações disponíveis em páginas ou interfaces web, por meio de requisições HTTP e análise estruturada das respostas (MITCHELL, 2018). Quando uma fonte de dados não disponibiliza API pública documentada — como é o caso do ElastiCNES —, o *web scraping* constitui alternativa técnica viável para automação da extração. Neste trabalho, a técnica foi aplicada por meio de engenharia reversa das requisições HTTP da plataforma ElastiCNES, utilizando a biblioteca Requests do Python para replicar as chamadas identificadas e obter os dados em formato JSON de forma automatizada e reproduzível.

### 2.4.2 PostgreSQL

PostgreSQL é um sistema gerenciador de banco de dados relacional de código aberto, amplamente reconhecido por sua robustez, conformidade com o padrão SQL e suporte a funcionalidades avançadas como transações ACID, indexação sofisticada e tipos de dados complexos (The PostgreSQL Global Development Group, 2023). É considerado uma das soluções mais maduras e confiáveis para armazenamento e processamento de dados em aplicações analíticas (MOMJIAN, 2001).

Neste trabalho, o PostgreSQL foi utilizado como repositório central do pipeline ELT, hospedando as camadas RAW, Staging e Dimensional. As transformações de limpeza, padronização e modelagem dimensional foram implementadas diretamente em SQL, aproveitando a capacidade de processamento do banco de dados para centralizar a lógica de negócio na camada de dados — reduzindo complexidade e aumentando manutenibilidade da solução.

### 2.4.3 Microsoft Power BI

O Microsoft Power BI é uma plataforma de Business Intelligence desenvolvida pela Microsoft, classificada como líder no Quadrante Mágico do Gartner para plataformas de Analytics e Business Intelligence por consecutivos anos (SCHLEGEL et al., 2024). A plataforma oferece um conjunto integrado de funcionalidades que abrange conexão a múltiplas fontes de dados, transformação via Power Query, modelagem analítica relacional, criação de medidas em linguagem DAX (*Data Analysis Expressions*) e desenvolvimento de relatórios e *dashboards* interativos (FERRARI; RUSSO, 2020).

O produto central do Power BI são os relatórios publicados no serviço Power BI — ambiente em nuvem que permite compartilhamento, agendamento de atualizações e consumo dos painéis via navegador ou aplicativo móvel, sem necessidade de instalação de software ou conhecimento técnico por parte do usuário final. Recursos como filtros cruzados, *drill-down* hierárquico, tooltips customizados e navegação entre páginas permitem construir experiências analíticas interativas e acessíveis a diferentes perfis de usuários (MICROSOFT, 2024).

Neste trabalho, o Power BI foi conectado diretamente ao banco de dados PostgreSQL, consumindo os dados já tratados e modelados na camada dimensional. Essa arquitetura — em que as transformações ocorrem no banco de dados e o Power BI atua exclusivamente na camada de visualização — segue a recomendação da documentação oficial da ferramenta de centralizar a lógica de negócio na camada de dados, garantindo melhor performance, menor redundância e maior facilidade de manutenção (MICROSOFT, 2024).

## 2.5 Trabalhos relacionados

Este tópico tem como objetivo contextualizar a contribuição deste trabalho a partir da análise de soluções existentes que utilizam visualização de dados e Business Intelligence aplicados à saúde pública. Para isso, são examinados painéis governamentais e plataformas analíticas amplamente utilizadas no contexto brasileiro, bem como trabalhos acadêmicos que exploram dashboards e análise de dados em saúde. A partir dessa análise, são identificadas lacunas que justificam a proposta desenvolvida neste estudo.

No contexto da saúde pública brasileira, algumas iniciativas governamentais disponibilizam painéis e plataformas de visualização de dados para promover transparência e subsidiar o acompanhamento de políticas públicas. O Departamento de Informática do SUS (DATASUS) mantém diversos sistemas, entre os quais se destacam o TabNet, operando desde a década de 1990 com interface baseada em tabelas dinâmicas, o e-Gestor Atenção Básica, focado em indicadores da atenção primária, e mais recentemente o ElastiCNES, desenvolvido para facilitar o acesso aos dados do Cadastro Nacional de Estabelecimentos

de Saúde (Brasil, 2023b; Brasil, 2011c).

O TabNet, apesar de amplamente utilizado por pesquisadores, apresenta interface desatualada que dificulta o uso por gestores não técnicos, exigindo conhecimento prévio de estruturas tabulares e dimensões de análise (CARVALHO; SANTOS, 2017). O e-Gestor, embora ofereça interface mais moderna, limita-se a dados da atenção básica, não permitindo análises integradas sobre a rede assistencial completa (PINTO; GIOVANELLA, 2018). A SAGE (Sala de Apoio à Gestão Estratégica) monitora indicadores estratégicos, mas oferece limitada capacidade de exploração granular, restringindo-se a painéis pré-configurados sem possibilidade de análises customizadas (TAMAKI et al., 2012).

O ElastiCNES representa evolução ao oferecer visualizações mais modernas organizadas em módulos temáticos (estabelecimentos, leitos, profissionais), mas mantém limitações significativas: organização de filtros pouco intuitiva, baixa variedade de tipos de gráficos, ausência de hierarquias visuais claras e performance comprometida em consultas complexas (DATASUS, 2024). Na literatura acadêmica, trabalhos utilizando dados do CNES concentram-se predominantemente em análises descritivas pontuais. Estudos caracterizam a distribuição de estabelecimentos ou avaliam cobertura assistencial através de scripts ad-hoc em ferramentas estatísticas, mas raramente desenvolvem soluções analíticas persistentes com arquitetura estruturada (RIBEIRO; COSTA, 2021; MELLO et al., 2018). Pesquisas sobre qualidade dos dados do CNES documentam inconsistências cadastrais, mas não propõem mecanismos automatizados de validação integrados a ferramentas de visualização (SILVA; MENDES, 2019; PINTO; FREITAS; FIGUEIREDO, 2018).

Internacionalmente, plataformas como CDC Wonder (EUA) e NHS Digital (Reino Unido) demonstram maior maturidade analítica ao oferecerem APIs documentadas, metadados detalhados, múltiplos níveis de granularidade e visualizações interativas sofisticadas, evidenciando a viabilidade de arquiteturas robustas mesmo com dados governamentais complexos (Centers for Disease Control and Prevention, 2020; NHS Digital, 2021).

Este trabalho diferencia-se ao implementar solução completa que integra boas práticas de engenharia de dados e Business Intelligence: pipeline ELT estruturado em camadas lógicas (RAW, Staging, Dimensional), modelagem dimensional otimizada segundo princípios de Kimball, dashboards com hierarquias visuais e interatividade avançada. Enquanto trabalhos anteriores focaram em análises pontuais ou painéis simplificados, esta solução demonstra como arquitetura adequada pode transformar dados públicos do CNES em ferramenta efetivamente útil para gestão em saúde pública, oferecendo melhor performance, usabilidade e confiabilidade analítica.

Tabela 1 – Análise comparativa entre soluções existentes e a proposta deste trabalho

<b>Critério</b>	<b>TabNet</b>	<b>e-Gestor</b>	<b>ElastiCNES</b>	<b>CDC Wonder</b>	<b>NHS Digital</b>	<b>Esta proposta</b>
Cobertura temática ampla	Parcial	Não	Sim	Sim	Sim	<b>Sim</b>
Interface acessível a gestores	Não	Parcial	Parcial	Parcial	Sim	<b>Sim</b>
Hierarquias visuais e drill-down	Não	Não	Não	Parcial	Sim	<b>Sim</b>
Variedade de tipos de gráficos	Baixa	Média	Baixa	Média	Alta	<b>Alta</b>
Filtros contextuais e intuitivos	Não	Parcial	Parcial	Sim	Sim	<b>Sim</b>
API documentada para acesso aos dados	Não	Não	Não	<b>Sim</b>	<b>Sim</b>	Não
Pipeline ELT estruturado em camadas	Não	Não	Não	Não	Não	<b>Sim</b>
Modelagem dimensional (Kimball)	Não	Não	Não	Não	Não	<b>Sim</b>
Análise histórica (série temporal)	Parcial	Parcial	Não	Sim	Sim	<b>Sim</b>
Comparação entre regiões	Parcial	Não	Parcial	Sim	Sim	<b>Sim</b>
Performance em consultas complexas	Baixa	Média	Baixa	Alta	Alta	<b>Alta</b>
Dados do contexto brasileiro (SUS)	Sim	Sim	Sim	Não	Não	<b>Sim</b>

Fonte: Elaborado pelo autor

## 3 Identificação do Problema

A gestão efetiva do Sistema Único de Saúde depende da capacidade de transformar dados cadastrais e assistenciais em informações estratégicas que subsidiem decisões sobre alocação de recursos, planejamento de serviços e monitoramento da rede assistencial. O Cadastro Nacional de Estabelecimentos de Saúde (CNES) constitui a base oficial que reúne informações sobre todos os estabelecimentos de saúde do país, incluindo dados sobre recursos humanos, equipamentos, leitos e serviços ofertados, sendo fundamental para análises sobre a organização da rede de saúde brasileira (DATASUS, 2024).

Para facilitar o acesso a esses dados, o DATASUS desenvolveu a Plataforma ElastiCNES, que utiliza tecnologia de Business Intelligence para disponibilizar visualizações interativas organizadas em painéis temáticos sobre estabelecimentos, leitos hospitalares, profissionais de saúde e distribuição geográfica das unidades. A plataforma representa avanço em relação a interfaces anteriores como o TabNet, oferecendo visualizações mais modernas e acessíveis ao público em geral (DATASUS, 2024).

Para fundamentar as decisões de design da solução proposta, foi conduzida análise sistemática dos painéis do ElastiCNES segundo critérios estabelecidos na literatura de visualização de dados em saúde pública: navegação intuitiva, simplicidade, usabilidade, clareza descritiva e variedade de visualizações (FEW, 2006a; WANG; SMITH; PATEL, 2024).

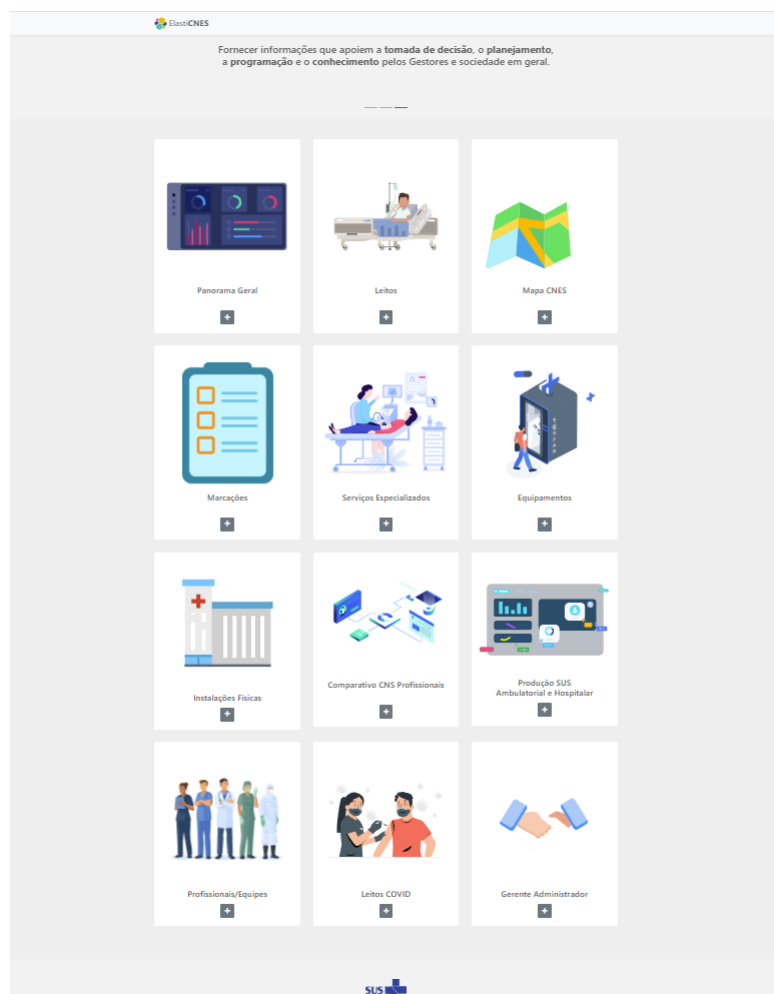
### 3.1 Análise Crítica dos Painéis ElastiCNES Existentes

A análise identificou limitações que motivaram o desenvolvimento dos dashboards refatorados apresentados neste trabalho.

O painel inicial da plataforma ElastiCNES apresenta uma organização visual simples e simétrica, com ícones representando as diferentes áreas temáticas disponíveis para consulta. Ao todo, são treze seções dispostas em uma grade regular, incluindo temas como Leitos, Profissionais, Equipamentos, Produção SUS e Leitos COVID. Essa disposição favorece uma navegação direta, especialmente para usuários que já conhecem a estrutura do CNES e estão familiarizados com os indicadores oferecidos. A identidade visual, marcada por ilustrações modernas, coloridas e em movimento, contribui para uma experiência agradável e transmite um aspecto institucional amigável.

A plataforma fornece descrições complementares por meio de cards que aparecem ao posicionar o cursor sobre os blocos temáticos, o que contribui para a contextualização. No entanto, esse recurso depende da interação direta com o elemento e não está visível de

Figura 1 – Painel Inicial CNES

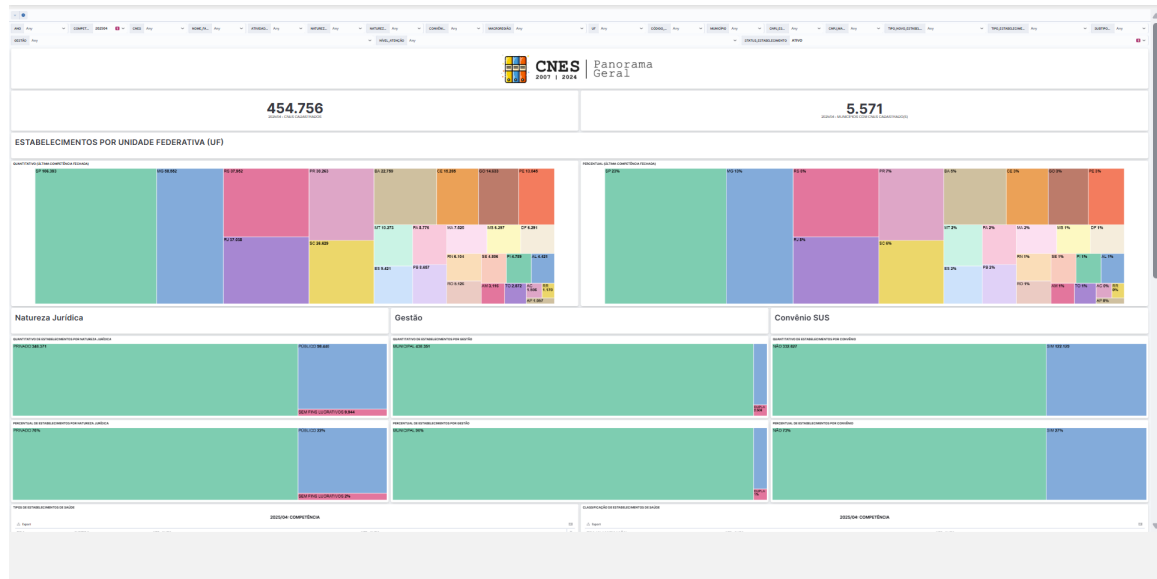


Fonte: Ministério da Saúde (Brasil) (2025)

forma permanente, o que pode dificultar a leitura rápida e comprometer a acessibilidade para usuários com menor familiaridade tecnológica. Além disso, os títulos utilizam termos técnicos ou pouco descritivos (como "Marcações" ou "Produção SUS Ambulatorial e Hospitalar"), o que pode gerar dúvidas sobre o escopo de cada painel. Conforme aponta Wang, Smith e Patel (2024), em dashboards de saúde pública, a presença de descrições claras, navegação intuitiva e linguagem acessível é considerada uma das principais demandas dos usuários.

### 3.1.1 Panorama Geral

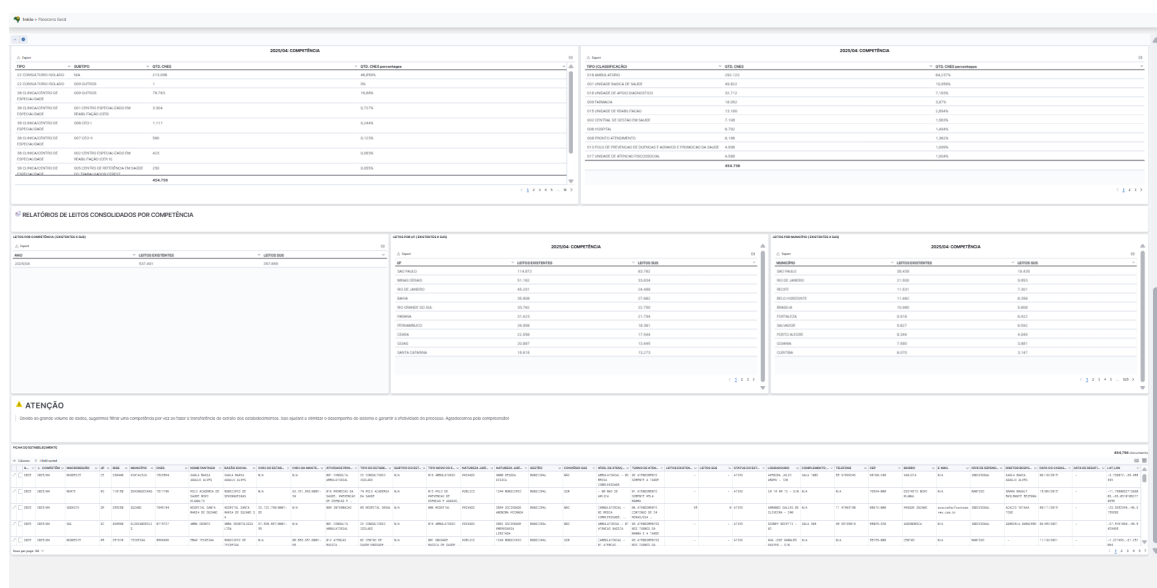
Figura 2 – Painel Panorama Geral



Fonte: Ministério da Saúde (Brasil) (2025)

O painel "Panorama Geral" da plataforma ElastiCNES apresenta uma visão consolidada da quantidade de estabelecimentos de saúde cadastrados no CNES, com destaque para dados por Unidade Federativa (UF), natureza jurídica, tipo de gestão e relação com o SUS. Sua estrutura é dividida em duas partes principais: uma área superior com gráficos

Figura 3 – Painel Panorama Geral



Fonte: Ministério da Saúde (Brasil) (2025)

do tipo treemap (mapas em mosaico) e uma seção inferior com tabelas detalhadas por competência, acompanhadas de alertas e observações técnicas.

Um dos aspectos mais problemáticos do painel é a disposição dos filtros no topo da interface. Apesar de essenciais para personalizações e recortes analíticos, os filtros estão distribuídos em uma linha longa, com fontes pequenas, sem qualquer hierarquia visual ou agrupamento, o que dificulta a leitura e sobrecarrega visualmente o usuário. Alguns filtros utilizam termos técnicos pouco claros, como "competência", que podem não ser compreendidos por usuários menos familiarizados com a terminologia do SUS. Outros filtros, como "Natureza Jurídica" ou "Tipo de Gestão", embora úteis, aparecem truncados ou com nomes cortados, dificultando a leitura e a correta seleção. Isso compromete severamente a experiência do usuário e o valor analítico do painel, já que muitos gestores podem ignorar filtros relevantes por falta de clareza ou simplesmente não compreender sua função. Conforme apontam [Wang, Smith e Patel \(2024\)](#) e [Few \(2006a\)](#), filtros eficazes devem ser visíveis, legíveis e organizados logicamente, de modo a facilitar — e não limitar — a exploração dos dados. A organização atual dos filtros representa uma barreira de entrada significativa, especialmente para usuários ocasionais ou com menor expertise técnica.

Nos gráficos, a escolha pelos treemaps permite visualizar a distribuição proporcional dos estabelecimentos entre categorias. No entanto, o excesso de cores e a ausência de hierarquia tipográfica dificultam a identificação de informações relevantes. Embora eficientes para mostrar proporções, treemaps com muitos elementos similares visualmente (como pequenos estados) perdem sua eficácia comunicativa e dificultam a comparação detalhada. A repetição do mesmo tipo de visualização em diferentes contextos sugere falta de consideração sobre qual formato gráfico seria mais adequado para cada tipo de análise específica. Além disso, a escala de cores utilizada não segue padrões intuitivos ou convenções estabelecidas, forçando o usuário a consultar repetidamente a legenda e aumentando desnecessariamente o tempo necessário para interpretação dos dados.

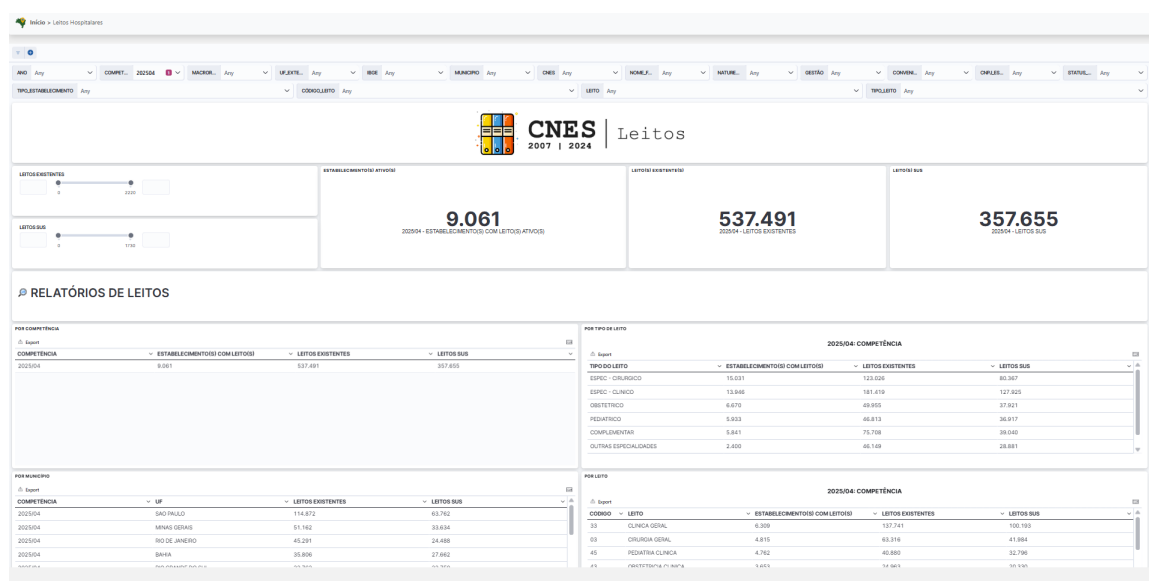
Na seção inferior, o painel apresenta tabelas com dados detalhados por tipo de estabelecimento, região e competência. Apesar de tecnicamente completas, essas tabelas possuem muitas colunas visíveis simultaneamente, dificultando a leitura, especialmente em dispositivos com telas menores. A ausência de recursos básicos de usabilidade, amplamente consolidados em soluções modernas de Business Intelligence, evidencia uma desconexão entre as capacidades técnicas da plataforma e as expectativas contemporâneas de experiência do usuário em ambientes analíticos.

Outro ponto de melhoria refere-se às mensagens operacionais, como instruções sobre envio de arquivos. Elas aparecem ao fim da página, com pouco destaque visual, o que reduz significativamente sua visibilidade e eficácia. Informações críticas como essa deveriam estar posicionadas com maior evidência, em local estratégico e com elementos visuais que chamem a atenção.

Em síntese, o painel "Panorama Geral" precisa de ajustes substanciais em visualização, organização e interatividade. A aplicação de boas práticas de design informacional como o uso racional de cores, categorização clara, filtros acessíveis e interações dinâmicas pode ampliar significativamente seu impacto e sua utilidade para a gestão pública. O estado atual do painel representa uma oportunidade perdida de transformar dados valiosos em insights verdadeiramente acionáveis.

### 3.1.2 Painel Leitos

Figura 4 – Painel Leitos



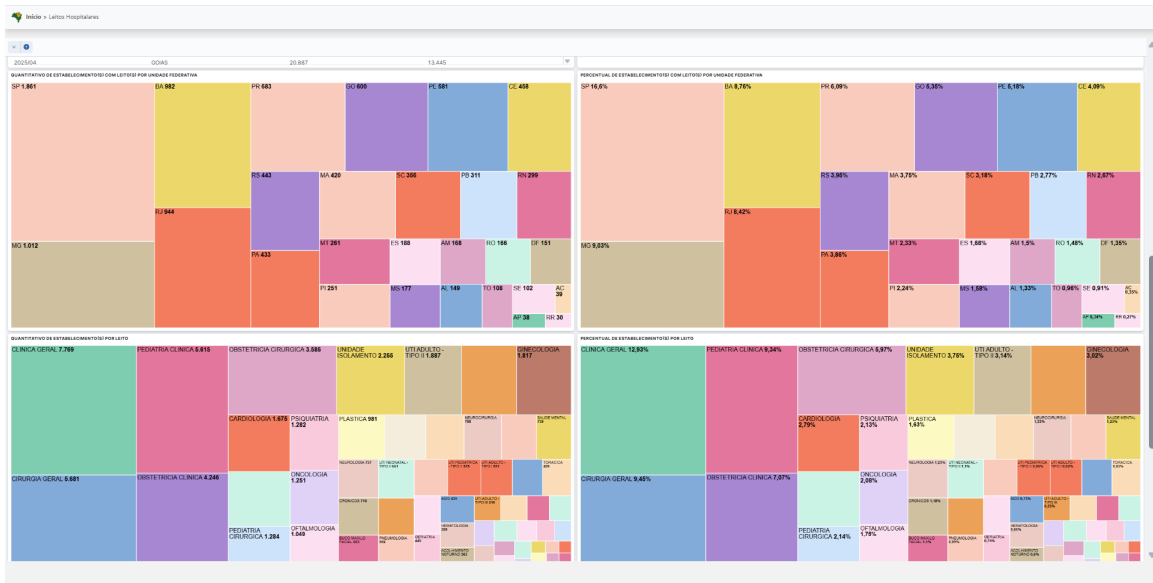
Fonte: Ministério da Saúde (Brasil) (2025)

O painel “Leitos” combina indicadores numéricos principais, tabelas analíticas e visualizações em formato de treemap, além de um extrato tabular com a base completa por estabelecimento.

Na parte superior, o painel exhibe três indicadores-chaves: número de estabelecimentos com leitos ativos, leitos existentes no período selecionado e leitos vinculados ao SUS. Esses elementos são úteis por fornecerem uma visão imediata da escala da rede hospitalar. No entanto, a ausência de contextualização textual ou visual (como explicações ou variação histórica) limita a compreensão desses números, especialmente para usuários não especialistas.

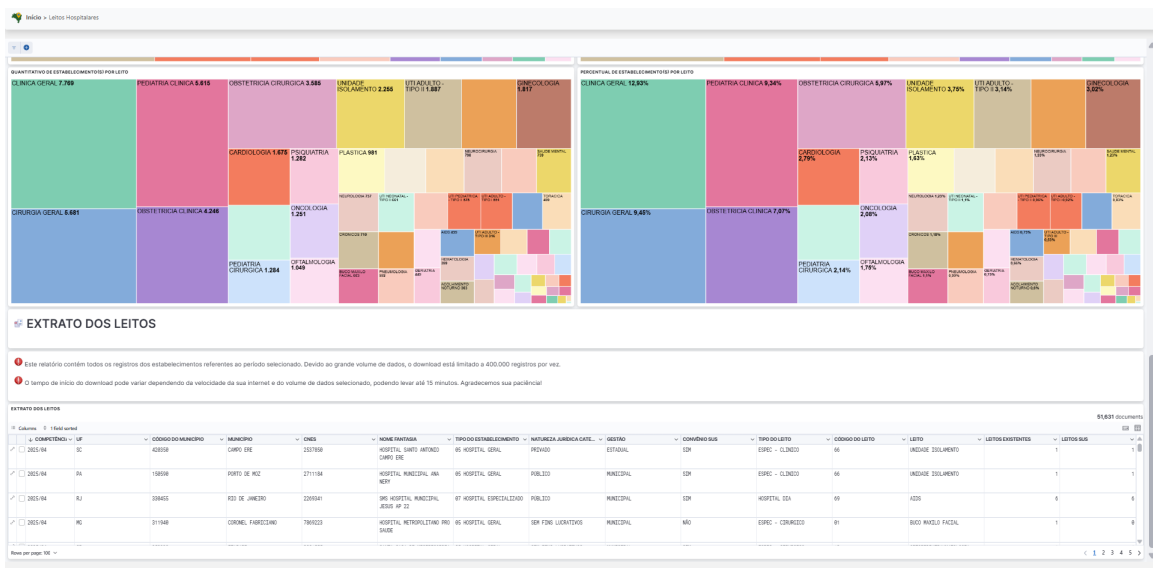
Assim como em outros painéis da plataforma, os filtros de seleção seguem uma disposição pouco intuitiva. Estão organizados em uma única linha superior com campos estreitos, fontes pequenas e terminologia técnica pouco acessível, como “compet.”, “tipo\_leito” ou “nature.”. Esses filtros não são acompanhados de legendas ou ícones infor-

Figura 5 – Painel Leitos



Fonte: Ministério da Saúde (Brasil) (2025)

Figura 6 – Painel Leitos



Fonte: Ministério da Saúde (Brasil) (2025)

mativos, o que compromete seu entendimento, além de dificultar a exploração dos dados por profissionais com menor familiaridade técnica. Segundo [Few \(2006a\)](#), [Wang, Smith e Patel \(2024\)](#), a clareza, legibilidade e agrupamento lógico dos filtros são essenciais para a usabilidade de dashboards e o aproveitamento analítico de suas funcionalidades.

A seção intermediária é composta por gráficos do tipo treemap, que buscam re-

presentar a distribuição de leitos por estado, por tipo e por percentual relativo. Embora esse formato permita uma visão proporcional interessante, a grande quantidade de categorias e a repetição do modelo em diferentes contextos visuais acaba sobrecarregando a leitura. Muitas caixas possuem tamanhos semelhantes e textos comprimidos, dificultando a distinção rápida entre os dados.

As tabelas analíticas organizadas por competência, tipo de leito, município e código do leito são completas e tecnicamente consistentes. No entanto, apresentam os mesmos problemas já mencionados em outros painéis: quantidade excessiva de colunas visíveis, ausência de hierarquia tipográfica. E sozinhas apresentam potencial limitado de análise. O painel “Leitos” reúne informações valiosas sobre a infraestrutura hospitalar brasileira, mas padece de problemas recorrentes de usabilidade e design.

### 3.1.3 Painel Profissionais

Figura 7 – Painel Profissionais

ANO	COMETER	REGIÃO	UF	IBGE	MUNICÍPIO	CNES	NOME FANTASIA	CNPJ MANTENEDORA	CNPJ	TIPO UNIDADE	SUBTIPO UNIDADE	NATUREZ.	GESTÃO	PROFISSIONAL - NOME	PROFISSIONAL - CNES
2025	2025/11	NORDESTE	PE	228799	PELO II	7215155	APAE DE PELO II	N/A	85.225.122/0001-89	36 CLINICA/CENTRO DE ESPECIALIZAC	001 CENTRO ESPECIAL	SEM FINS LUL. MUNICIPAL	MUNICIPAL	MARLENE MEXEIRO RIBEIRO PER...	762888099596677
2025	2025/11	SUL	PR	411995	PONTEL DO PAUL.	2557239	POSTO DE SAUDE DE IPANEMA	81.689.842/0001-02	N/A	81 POSTO DE SAUDE	N/A	PUBLICO	MUNICIPAL	SABRINA DOS SANTOS TOME	786895257652687
2025	2025/11	NORDESTE	MA	211108	SÃO LEÃO	6848278	CENTRO DE SAUDE RIBOURNENE	N/A	86.176.999/0001-80	80 HOSPITAL GERAL	N/A	PREVADO	ESTRÁ	JANI RAIZA AVELAR DE COSTA	762820252180818
2025	2025/11	SUDESTE	SP	350829	VALINHOS	8573463	HOSPITAL E MANTENEDORA GALL.	N/A	86.742.682/0001-80	85 HOSPITAL GERAL	N/A	PREVADO	MUNICIPAL	CLAUDIA SIFRAL FERREIRA DE J.	762884122786643
2025	2025/11	SUDESTE	SP	350809	SÃO PAULO	2877655	ASSOC AAO V CLEMENTINO	N/A	68.978.457/0001-11	87 HOSPITAL ESPECIALIZADO	003 ORTOPEDIA	SEM FINS LUL. MUNICIPAL	MUNICIPAL	JAIME OUTRA DE MORAES TEIXE.	898884618481837
2025	2025/11	CENTRO-OESTE	DF	538818	BRASILIA	5737315	HSEM	N/A	28.481.233/0009-28	85 HOSPITAL GERAL	N/A	SEM FINS LUL. ESTADUAL	ESTADUAL	DANIELA CARLA GOMES DA COSTA	788888899999988
2025	2025/11	SUDESTE	SP	350808	SÃO PAULO	2786688	HOSP MIN FERNANDO MUIRO FISE.	46.392.138/0003-88	46.392.140/0008-24	85 HOSPITAL GERAL	N/A	PUBLICO	MUNICIPAL	FERNANDO CINTRA DE PAULO	787888831527775
2025	2025/11	SUDESTE	ES	350868	ARCORZ	4752288	FARMACIA DIGIANA	N/A	25.182.146/0288-78	43 FARMACIA	N/A	PREVADO	MUNICIPAL	JANE LIBERATO DAS REYES	783888842794771
2025	2025/11	SUDESTE	SP	358198	ARAPUÁ	2048202	OSP LINDA DAS ÁGUIS MORGESM.	82.464.455/0001-73	N/A	82 CENTRO DE SAUDE-UNIDADE BASICA	N/A	PUBLICO	MUNICIPAL	PEDRO RIBANDI DE MORAES PERLO	788888844222222
2025	2025/11	SUDESTE	SP	358718	BOA JESUS DOS	2084887	UNIDADE RESTA DE SAUDE DR JO.	82.359.652/0001-02	N/A	15 UNIDADE RESTA	N/A	PUBLICO	MUNICIPAL	MARIA DEL ROSARIO BURGOS PER.	788888844122221
2025	2025/11	SUDESTE	SP	352598	JAMILIAR	2995785	POLICLINICA HORTALANDEIA	45.788.162/0001-58	N/A	73 PRONTO ATENDIMENTO	001 PRONTO ATENDIME.	PUBLICO	MUNICIPAL	RAFAEL MARCOS SANTORI COSTA	7628888451887517
2025	2025/11	SUDESTE	SP	350808	SÃO PAULO	2878815	HC DA FUSP HOSPITAL DAS CL.L.	N/A	56.577.859/0001-88	85 HOSPITAL GERAL	N/A	SEM FINS LUL. ESTADUAL	ESTADUAL	ADRIANA HACHETA	782888821328211

Fonte: Ministério da Saúde (Brasil) (2025)

O painel “Profissionais” da plataforma tem como objetivo consolidar informações sobre a distribuição dos profissionais em unidades de saúde brasileiras, categorizadas por região, estado e tipo de especialização. Em sua estrutura, o painel apresenta apenas uma extensa tabela de extrato de dados, incluindo detalhes como código CNES, município, tipo da unidade, carga horária e etc.

O painel apresenta, novamente, filtros dispostos de forma linear e compacta, com campos estreitos e ausência de agrupamentos temáticos, ainda que neste caso sejam oferecidas menos opções de filtro. Segundo Few (2006a), filtros e controles de navegação

devem ser claros, legíveis e organizados de forma lógica para não se tornarem barreiras ao uso da informação.

A seção disponibiliza um extrato completo dos registros da competência selecionada. A tabela é rica em informações apesar da completude, densidade de colunas, rolagem horizontal excessiva e a fonte pequena comprometerem a legibilidade, especialmente em monitores menores. Conforme destaca [Wang, Smith e Patel \(2024\)](#), a clareza visual e a responsividade de painéis de saúde são requisitos críticos para garantir que a informação seja acessada e interpretada corretamente.

Apesar do painel apresentar uma base robusta de dados estruturais sobre a rede de saúde nacional, essas informações não são utilizadas de forma adequada, já que é apresentada apenas a tabela sem outros visuais que possam trazer *insights*. A adoção de elementos como legendas explicativas, agrupamento visual de filtros, indicadores interativos e melhorias na visualização tabular pode ampliar consideravelmente a eficácia do painel para gestores e analistas de saúde pública [Few \(2006a\)](#), [Wang, Smith e Patel \(2024\)](#).

## 3.2 Limitações do ElastiCNES

Apesar de representar evolução importante, o ElastiCNES apresenta limitações que restringem seu uso como instrumento efetivo de gestão. A plataforma apresenta fragilidades em três dimensões principais:

**Interface e visualização:** Os painéis disponibilizam gráficos simplificados com organização de filtros pouco intuitiva, baixa variedade de tipos de visualizações e ausência de hierarquias visuais claras ([FEW, 2006b](#)). A estrutura de navegação entre painéis dificulta cruzamentos entre diferentes dimensões (estabelecimentos, leitos, profissionais), limitando análises integradas.

**Desempenho e arquitetura técnica:** Identificam-se problemas de performance relacionados ao tempo de carregamento e responsividade, especialmente em análises com múltiplos filtros ou períodos extensos. A ausência de API pública documentada limita integrações automatizadas e o desenvolvimento de soluções analíticas complementares ([FeliciLab, 2020](#)).

**Qualidade e documentação dos dados:** O cadastramento dependente de estabelecimentos e gestores municipais está sujeito a erros, omissões e desatualizações ([RIBEIRO; COSTA, 2021](#); [SILVA; MENDES, 2019](#)). A ausência de mecanismos explícitos de validação dificulta a identificação de inconsistências, como estabelecimentos inativos registrados como operacionais ([MELLO et al., 2018](#)). A plataforma não disponibiliza metadados claros sobre periodicidade de atualização ou cobertura dos dados, comprometendo a comparabilidade temporal ([UBALDI, 2013](#)).

Essas limitações comprometem decisões sobre dimensionamento de força de trabalho, planejamento de expansão de serviços e alocação de recursos, exigindo frequentemente exportação e processamento externo dos dados (MACHADO; LIMA; BAPTISTA, 2018; SANTOS et al., 2021).

### 3.3 Definição do Problema

As limitações identificadas na análise do ElastiCNES configuram um problema computacional: a ausência de uma arquitetura de dados estruturada entre a fonte de dados brutos do CNES e a camada de visualização compromete diretamente a qualidade, o desempenho e a usabilidade das análises disponíveis para gestores de saúde pública.

Mais precisamente, o problema pode ser decomposto em três dimensões técnicas interdependentes:

1. **Ausência de pipeline de extração automatizado:** O ElastiCNES não disponibiliza API pública documentada, e os dados brutos do CNES não estão acessíveis em formato estruturado e atualizado para consumo analítico externo. Isso exige que qualquer solução alternativa implemente mecanismos próprios de extração — neste caso, via *web scraping* com engenharia reversa das requisições HTTP da plataforma —, etapa que demanda conhecimento técnico especializado e não está disponível em nenhuma ferramenta governamental existente.
2. **Ausência de modelagem dimensional:** Os dados do CNES são armazenados e disponibilizados em estruturas operacionais não otimizadas para análise. A ausência de modelagem dimensional com tabelas fato e dimensões organizadas segundo princípios estabelecidos (KIMBALL; ROSS, 2013) resulta em consultas lentas, métricas inconsistentes e dificuldade de cruzamento entre diferentes domínios (estabelecimentos, leitos e profissionais). Como evidenciado nos testes de usabilidade realizados neste trabalho, o ElastiCNES apresentou falha completa em tarefas de análise histórica e tempo médio de resposta 2,4 vezes superior ao sistema proposto.
3. **Inadequação das visualizações à tomada de decisão:** A camada de visualização do ElastiCNES não explora o potencial analítico dos dados disponíveis, oferecendo predominantemente treemaps repetitivos, tabelas densas e filtros com terminologia técnica inacessível. A ausência de hierarquias visuais, drill-down hierárquico e séries temporais limita severamente o valor gerencial da plataforma para diferentes perfis de usuários (FEW, 2006a; WANG; SMITH; PATEL, 2024).

O problema central que este trabalho se propõe a resolver é, portanto: como estruturar uma arquitetura computacional composta por pipeline ELT, modelagem dimensional

e camada de visualização capaz de transformar os dados brutos do CNES em dashboards analíticos com melhor desempenho, usabilidade e confiabilidade do que as soluções governamentais existentes?

A resposta a essa questão orienta todas as decisões técnicas descritas no capítulo seguinte: a escolha do Python para extração via *web scraping*, do PostgreSQL para armazenamento e transformação em camadas lógicas, da modelagem dimensional de Kimball para organização dos dados analíticos, e do Power BI para construção dos dashboards refatorados.

## 4 Desenvolvimento

Este capítulo descreve o processo de implementação da solução analítica proposta, apresentando as decisões técnicas adotadas para a construção dos *dashboards* e para a organização dos dados provenientes de bases públicas da área da saúde.

### 4.1 Arquitetura da Solução Analítica

#### 4.1.1 Visão Geral da Arquitetura

A arquitetura da solução foi definida para estruturar o fluxo dos dados de forma desacoplada, controlada e orientada à análise. Considerando as limitações da plataforma oficial, optou-se por arquitetura que separa explicitamente as etapas de coleta, tratamento, armazenamento analítico e consumo das informações.

Essa separação em camadas lógicas permite que extração e transformação ocorram independentemente da visualização, de modo que alterações na estrutura dos dados do CNES sejam absorvidas nas camadas iniciais sem comprometer os *dashboards*. Cada componente possui responsabilidades bem definidas, favorecendo rastreabilidade, manutenibilidade e evolução da solução, aspectos fundamentais em projetos com dados públicos de grande volume. Essa abordagem está alinhada às boas práticas de modelagem dimensional e arquiteturas analíticas consolidadas na literatura de Business Intelligence ([KIMBALL; ROSS, 2013](#)).

#### 4.1.2 Componentes da Solução

Cada componente desempenha um papel específico dentro da arquitetura, conforme descrito a seguir.

A camada de origem da solução é composta pelos dados públicos do **Cadastro Nacional de Estabelecimentos de Saúde (CNES)**, acessados por meio da plataforma **ElastiCNES**. Essas bases apresentam características típicas de dados administrativos governamentais, como heterogeneidade estrutural entre competências, inconsistências cadastrais recorrentes e ausência de um padrão rígido de atualização temporal. Tais características impactam diretamente a confiabilidade e a padronização das análises, influenciando as decisões arquiteturais adotadas neste trabalho.

Além disso, não existe uma interface pública oficial que possibilite a extração automatizada, completa e em larga escala dos dados necessários para as análises propostas. As alternativas disponíveis, como exportações manuais ou downloads parciais, mostraram-

se insuficientes em termos de volume, granularidade e reprodutibilidade. Diante desse cenário, tornou-se necessária a adoção de uma estratégia alternativa de acesso aos dados, descrita em detalhes nas seções seguintes.

A camada de ingestão e orquestração dos dados é implementada por meio de um pipeline desenvolvido em **Python**. Esse componente é responsável por automatizar a coleta dos dados, controlar a execução sequencial das etapas do processo e garantir a reprodutibilidade das cargas. A escolha dessa linguagem se deve à sua flexibilidade para integração com serviços web, manipulação de dados e controle de fluxo, além de permitir a organização modular das rotinas de extração conforme o domínio de cada conjunto de dados analisado.

O armazenamento intermediário e o processamento analítico são realizados em um banco de dados **PostgreSQL**. Nesse componente, os dados são organizados em camadas lógicas que separam informações brutas, dados tratados e estruturas analíticas finais. O uso de um sistema de banco de dados relacional possibilita a aplicação de regras de integridade, o uso de transformações baseadas em SQL e a otimização do desempenho das consultas, aspectos fundamentais para suportar a modelagem dimensional e o consumo analítico posterior.

Neste contexto o **PostgreSQL** funciona como *Data Warehouse* estruturado em três camadas lógicas, sendo que a camada *raw* preserva dados brutos com mínima transformação, aproximando-se do conceito de zona de aterrissagem (*landing zone*) presente em arquiteturas de *Data Lake* (INMON, 2005). Esta abordagem híbrida, conhecida como padrão *Data Lakehouse*, combina flexibilidade de ingestão de dados brutos com capacidade de consultas estruturadas otimizadas para Business Intelligence (ARMBRUST et al., 2021).

A camada de visualização é composta por *dashboards* desenvolvidos no **Power BI**, conectados diretamente à estrutura analítica no banco de dados. Essa integração permite a exploração interativa dos dados, a aplicação de filtros e recortes multidimensionais e o cálculo de indicadores analíticos a partir de dados previamente tratados. A separação entre a camada de visualização e as etapas de tratamento garante que as análises apresentadas reflitam fielmente as regras de negócio definidas no pipeline, reduzindo dependências diretas entre o ambiente de BI e as fontes originais.

## 4.2 Pipeline de Dados e Processo de ELT

O pipeline de dados foi projetado para transformar os dados brutos, provenientes de múltiplos arquivos CSV com estruturas e períodos distintos, em um conjunto de informações limpo, padronizado e otimizado para análise. Neste trabalho, o pipeline foi construído para extração e tratamento de três conjuntos principais de dados: **Panorama Geral dos Estabelecimentos**, **Profissionais de Saúde** e **Leitos**.

A orquestração do pipeline é centralizada em um **script** principal (*main*), responsável por coordenar a execução sequencial das etapas de extração, carga inicial e transformação. Esse componente atua como ponto único de controle do fluxo, registrando o progresso da execução e permitindo a interrupção controlada do processo em caso de falhas, evitando cargas parciais inconsistentes no ambiente analítico.

A extração dos dados é organizada de forma segmentada por unidade federativa e por período de competência, de modo que cada execução do processo seja responsável por um recorte específico de estado e intervalo temporal. Para este estudo, os dados extraídos abrangem o período de **2020 a 2025**. Cada execução retorna exclusivamente os registros correspondentes à combinação definida de estado e competência, permitindo o controle granular do volume processado, a validação dos resultados por recorte analítico e a reexecução pontual de períodos ou unidades federativas em caso de falhas ou necessidade de atualização. Essa estrutura garante maior rastreabilidade, confiabilidade e eficiência no processo de extração.

Após a carga inicial dos dados na camada *raw* eles passam pra camada *staging*, onde são aplicadas transformações orientadas às necessidades analíticas do projeto. Essas transformações são executadas diretamente no banco de dados por meio de scripts SQL, explorando o desempenho do SGBD para operações de limpeza, validação e consolidação. Concluída essa etapa, os dados tratados passam a servir de base para a construção das estruturas dimensionais utilizadas nos *dashboards*.

A Figura 8 ilustra a organização dos dados em camadas, desde a ingestão dos dados brutos até a construção do modelo dimensional. A adoção dessa estrutura facilita a separação entre dados operacionais e analíticos, além de preparar o ambiente para a modelagem dimensional utilizada na etapa de visualização.

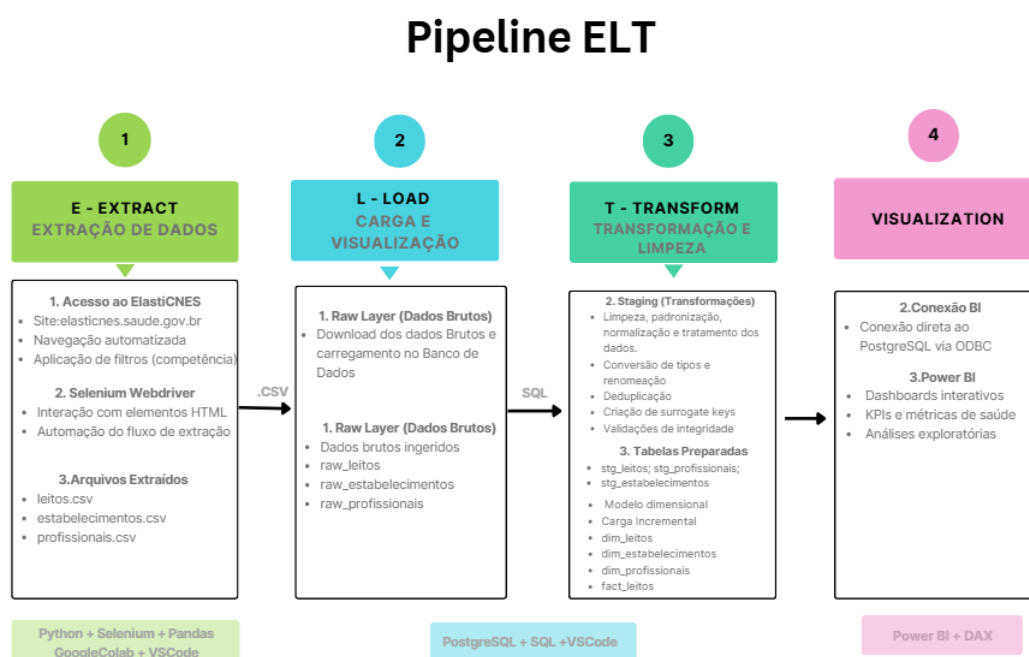


Figura 8 – Fluxo do pipeline ELT e organização das camadas analíticas

Fonte: Produzido pelo autor

### 4.2.1 Estratégia de Extração dos Dados do CNES

A extração dos dados do CNES apresentou desafios relevantes devido à inexistência de uma API pública oficial que possibilite o acesso automatizado e estruturado às informações. Foram analisadas inicialmente abordagens baseadas no download direto de arquivos disponibilizados em portais públicos e repositórios FTP. No entanto, essas alternativas mostraram-se inadequadas para o contexto deste estudo: os portais apresentavam ausência de atributos essenciais para as análises propostas, enquanto os repositórios FTP careciam de padronização estrutural, dificultavam a identificação semântica dos conjuntos de dados e apresentavam arquivos com inconsistências durante a conversão para formatos tabulares.

### 4.2.2 Engenharia Reversa da Interface Web

Diante dessas restrições, adotou-se uma estratégia baseada em engenharia reversa da interface web da plataforma ElasticNES, caracterizada pela utilização de uma API não documentada. Essa abordagem fundamentou-se na inspeção do tráfego de rede gerado pela aplicação, permitindo a identificação dos endpoints e dos parâmetros necessários para o recorte temporal (competência) e geográfica (unidade federativa), conforme práticas reconhecidas na literatura técnica.

A extração foi implementada utilizando sessões HTTP persistentes com Python, através da biblioteca `requests`, com configuração explícita de cabeçalhos compatíveis com

o ambiente Kibana/Elasticsearch utilizado pelo ElastiCNES.

Para arquitetura da extração foram desenvolvidos três scripts Python especializados, cada um responsável por extrair um domínio específico dos dados:

- **estabelecimentos.py**: Extrai dados cadastrais completos dos estabelecimentos de saúde, consultando o índice `cgsi-cnesprod-c1*` do Elasticsearch. Utiliza filtros por competência (`index_comp`) e UF (`UF.keyword`) para segmentar a extração.
- **leitos.py**: Coleta informações sobre capacidade instalada de leitos hospitalares do índice `cnes-leitos*`. Implementa lógica de *fallback* que tenta filtrar por nome extenso do estado (`UF POR EXTENSO.keyword`) e, caso não retorne resultados, utiliza a sigla (`uf`).
- **profissionais.py**: Obtém dados de vínculos profissionais do índice `cnes-profissionais*`, filtrando por competência e sigla da UF (`uf.keyword`).

Para lidar com o elevado volume de registros, todos os scripts implementam mecanismo de paginação controlada, recuperando dados em lotes de 500 registros. Devido à limitação do Elasticsearch de retornar no máximo 10.000 registros via paginação `from/size`, foi implementada a técnica `search_after` para continuar a extração além desse limite, utilizando o valor `sort` do último registro como cursor para a próxima página.

Intervalos entre requisições (0,12s a 0,6s dependendo do script) foram inseridos para reduzir o risco de bloqueios ou sobrecarga do serviço durante a coleta. Os dados consolidados foram persistidos de forma incremental em arquivos CSV, organizados por competência e unidade federativa.

### 4.2.3 Volumetria dos Dados Extraídos

Para o período completo analisado, que abrange competências de 2020 a 2025, foram extraídos dados seguindo o padrão apresentado na Tabela 2, que exemplifica a volumetria para as competências de agosto e setembro de 2025 (202508 e 202509):

Tabela 2 – Volumetria dos dados extraídos por domínio

Domínio	Arquivos	Tamanho médio	Variação	Total
Estabelecimentos	54	5,5 MB	1,4 MB - 8,1 MB	~297 MB
Leitos	54	2,2 MB	0,1 MB - 8,8 MB	~119 MB
Profissionais	54	7,4 MB	7,1 MB - 8,1 MB	~400 MB
<b>Total</b>	<b>162</b>	-	-	<b>~816 MB</b>

Fonte: Elaborado pelo autor

A variação no tamanho dos arquivos reflete as diferenças populacionais e de complexidade da rede assistencial entre as unidades federativas. Estados como São Paulo,

Minas Gerais e Rio de Janeiro apresentam os maiores volumes, enquanto estados da região Norte apresentam arquivos significativamente menores. Essa distribuição heterogênea é esperada e está diretamente relacionada ao número de estabelecimentos de saúde, à densidade populacional e ao grau de estruturação da rede assistencial em cada região.

A distribuição dos dados por domínio também evidencia características específicas de cada conjunto. Os arquivos de profissionais apresentam maior uniformidade de tamanho entre estados (7,1 MB a 8,1 MB), sugerindo uma distribuição mais homogênea da força de trabalho em saúde proporcionalmente ao número de estabelecimentos. Já os dados de leitos apresentam a maior variabilidade (0,1 MB a 8,8 MB), refletindo diferenças significativas na disponibilidade de leitos hospitalares entre as unidades federativas, especialmente quando comparadas regiões com alta concentração de hospitais de grande porte (como Sudeste) com áreas de menor densidade hospitalar (como Norte).

Os dados extraídos apresentam estrutura tabular com as seguintes características:

- **Estabelecimentos:** 62 colunas incluindo identificação (CNES, CNPJ, CPF), dados administrativos (razão social, tipo de estabelecimento, natureza jurídica, gestão), localização geográfica (endereço completo, coordenadas, município, UF, macrorregião), características operacionais (nível de atenção, convênio SUS, atividade principal) e indicadores de capacidade e produção (leitos existentes e SUS, valores SIA/SIH).
- **Leitos:** 64 colunas combinando atributos do estabelecimento com dados específicos de leitos (tipo, código, especialidade, quantitativos de leitos existentes, contratados e SUS).
- **Profissionais:** 54 colunas com identificação profissional (nome, CPF, CNS), vínculo institucional (CNES, estabelecimento), classificação ocupacional (CBO), carga horária detalhada (total, ambulatorial SUS, hospitalar SUS, outros) e vinculação a equipes (nome, tipo, INE, datas de movimentação).

Embora o uso de técnicas de engenharia reversa de APIs exija cuidados adicionais quanto à manutenção do código e à adaptação a possíveis alterações na estrutura das requisições, a estratégia adotada mostrou-se adequada ao contexto do projeto, viabilizando uma extração automatizada, controlada e reprodutível dos dados do CNES, atendendo aos requisitos técnicos necessários para a construção da solução analítica proposta.

#### 4.2.4 Pré-processamento e Carga ETL

Com os dados extraídos e organizados em arquivos CSV, a etapa seguinte consiste na carga desses dados no ambiente analítico, preparando-os para as transformações subsequentes. Esta fase foi estruturada com foco em controle operacional, rastreabilidade e, especialmente, idempotência do processo.

### 4.2.5 Orquestração da carga

A carga é centralizada no script `load.py`, responsável por processar sequencialmente os arquivos CSV gerados pelos scripts de extração. O processo foi projetado para ser idempotente, ou seja, pode ser executado múltiplas vezes sem duplicar dados ou causar inconsistências. Isso é alcançado através de um mecanismo de controle baseado em *fingerprinting* de arquivos.

O *fingerprint* único de cada arquivo é calculado combinando três atributos:

- Nome do arquivo
- Tamanho em bytes
- Timestamp de última modificação

Esses valores são concatenados e submetidos a hash SHA-1, gerando um identificador único que representa exatamente aquela versão do arquivo. Antes de processar qualquer CSV, o sistema consulta a tabela de controle `ingestion_log` para verificar se aquele *fingerprinting* já foi processado. Em caso positivo, o arquivo é ignorado; caso contrário, prossegue-se com a carga.

A persistência dos dados é realizada no PostgreSQL, utilizando o schema `teste_raw` como camada de armazenamento inicial. O processo segue o fluxo apresentado na Figura 9 :

1. **Verificação de processamento prévio:** Consulta à tabela `ingestion_log` usando o *fingerprinting* do arquivo.
2. **Leitura do cabeçalho:** Extração dos nomes de colunas presentes no CSV para determinar a estrutura necessária.
3. **Gerenciamento dinâmico do schema:**
  - Se a tabela destino não existe, é criada com todas as colunas tipadas como `text`
  - Se a tabela existe mas possui colunas ausentes, estas são adicionadas via `ALTER TABLE ADD COLUMN`
4. **Carga em lote:** Utilização do comando `COPY FROM STDIN` do PostgreSQL, que oferece desempenho superior comparado a inserções individuais, processando milhares de registros por segundo.
5. **Registro de auditoria:** Inserção do registro na `ingestion_log` com fingerprint, nome do arquivo, quantidade de linhas carregadas e timestamp de processamento.

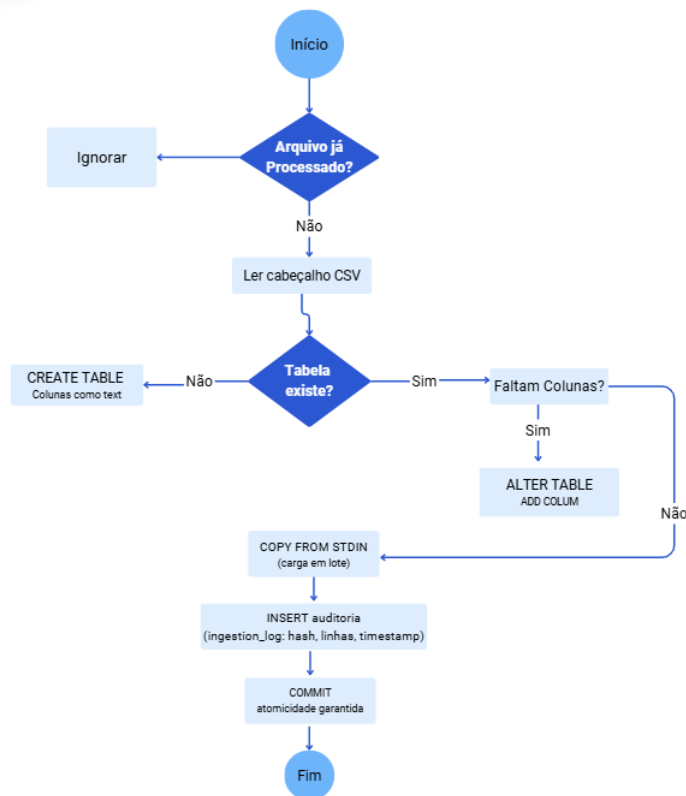


Figura 9 – Fluxo de persistência dos dados

Fonte: Produzido pelo autor

6. **Commit da transação:** Garantia de atomicidade – ou todo o arquivo é carregado com sucesso, ou nenhuma alteração é persistida.

A decisão de criar todas as colunas inicialmente como tipo `text` justifica-se pela necessidade de flexibilidade durante a carga. A conversão para tipos apropriados (numéricos, datas, booleanos) será realizada posteriormente na camada de staging.

Uma característica importante do pipeline é sua capacidade de adaptar-se automaticamente a mudanças estruturais nos dados fonte. Como o CNES é uma base administrativa que pode sofrer alterações ao longo do tempo como a inclusão de novos campos ou descontinuação de existentes, o sistema foi projetado para:

- Detectar automaticamente novas colunas presentes nos CSVs
- Adicioná-las às tabelas existentes sem intervenção manual
- Preencher com `NULL` os registros anteriores para as novas colunas
- Manter compatibilidade com competências históricas que não possuem determinados campos

Essa abordagem evita falhas de carga quando há divergências estruturais entre competências diferentes e elimina a necessidade de manutenção manual do schema a cada atualização dos dados.

#### 4.2.6 Tabelas da camada RAW

Nessa camada os dados são organizados em três tabelas principais no schema `teste_raw`:

- **raw\_estabelecimentos**: Recebe todos os CSVs do diretório `dados_estabelecimentos_por_estado`, consolidando dados cadastrais completos dos estabelecimentos de saúde em todas as competências e UFs.
- **raw\_leitos**: Consolida os arquivos do diretório `dados_cnes_leitos_por_estado`, armazenando informações sobre a distribuição e tipos de leitos hospitalares.
- **raw\_profissionais**: Agrega os CSVs de `dados_profissionais_por_uf`, registrando vínculos profissionais na rede de saúde.

A camada RAW preserva integralmente a estrutura original dos dados extraídos, funcionando como registro histórico imutável. Todas as transformações, limpezas e padronizações são aplicadas posteriormente na camada de staging, mantendo a separação de responsabilidades e permitindo reprocessamento caso necessário.

#### 4.2.7 Controle de qualidade e rastreabilidade

A tabela `ingestion_log` funciona como registro completo de auditoria do processo de carga, armazenando:

```
CREATE TABLE teste_raw.ingestion_log (  
    file_name text NOT NULL,  
    file_hash text NOT NULL,  
    rows_loaded integer,  
    loaded_at timestamptz DEFAULT now(),  
    PRIMARY KEY (file_name, file_hash)  
);
```

A chave primária composta garante que diferentes versões do mesmo arquivo sejam tratadas como cargas distintas, enquanto tentativas de reprocessamento da mesma versão são automaticamente impedidas pela restrição. Esse mecanismo torna o pipeline robusto a interrupções e permite recuperação incremental em caso de falhas parciais.

Ao término desta etapa, os dados encontram-se carregados no schema `_raw`, mantendo a granularidade e estrutura original das fontes, prontos para as transformações e padronizações descritas na seção seguinte.

#### 4.2.8 Tratamento e Padronização dos Dados

A Figura 10 apresenta o diagrama do pipeline de dados adotado, evidenciando a organização das camadas analíticas desde a ingestão dos dados brutos até a etapa de visualização. Os dados são tratados e organizados na camada de *Staging*, onde são aplicadas validações iniciais, tratamentos e preparações necessárias para posterior utilização desses dados. Posteriormente a camada dimensional consolida as informações em tabelas fato e dimensões, estruturadas segundo princípios de modelagem analítica. E por fim, a camada de visualização representa o consumo dos dados tratados por ferramentas de *Business Intelligence*.

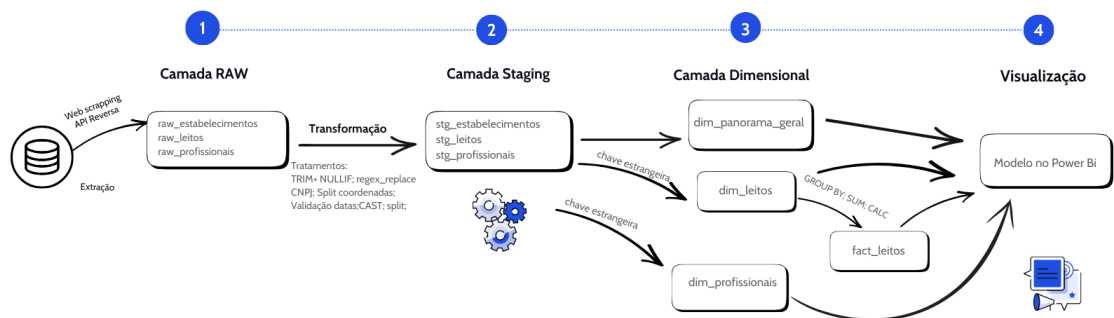


Figura 10 – Diagrama do pipeline de dados e organização das camadas analíticas

Fonte: Elaborado pelo autor

Na camada de *staging*, foram aplicados procedimentos de tratamento e padronização com foco na adequação das informações ao contexto analítico do projeto. Essa etapa teve como objetivo reduzir inconsistências oriundas das bases originais do CNES, harmonizar estruturas entre diferentes competências e garantir que os dados estivessem semanticamente coerentes para a modelagem dimensional e para o cálculo dos indicadores apresentados nos *dashboards*.

O tratamento dos dados foi realizado diretamente no banco de dados PostgreSQL, por meio de scripts SQL organizados por domínio e por função analítica. Essa decisão permitiu manter as regras de transformação próximas ao modelo de dados, facilitando a rastreabilidade, a manutenção e a evolução das regras de negócio, além de explorar o desempenho do SGBD em operações de validação, agregação e consolidação.

#### 4.2.8.1 Transformações Aplicadas

Entre os procedimentos comuns aplicados aos diferentes conjuntos de dados, destacam-se:

- **Padronização de campos textuais:** normalização de caracteres maiúsculos/minúsculos, remoção de espaços em branco desnecessários e eliminação de caracteres especiais inválidos que comprometiam a consistência das análises.
- **Conversão e validação de tipos:** transformação de campos numéricos armazenados como texto para tipos apropriados (integer, numeric), conversão de datas do formato string para timestamp, e tratamento de formatos inconsistentes entre competências.
- **Tratamento de valores nulos:** definição de regras específicas para campos obrigatórios, substituição de valores ausentes por padrões quando aplicável, e sinalização explícita de campos sem informação para análises posteriores.
- **Padronização de campos categóricos:** unificação de variações semânticas (exemplo: “Hospital”, “HOSPITAL”, “Hosp” → “Hospital”), criação de mapeamentos consistentes para valores de domínio (tipos de estabelecimento, natureza jurídica) e correção de inconsistências cadastrais identificadas na fonte.

Devido à natureza administrativa dos dados do CNES, foram observadas variações estruturais entre competências, incluindo a ausência ou inclusão de atributos em determinados períodos. Para lidar com essas diferenças, o tratamento considerou a definição explícita de esquemas esperados, assegurando a presença dos campos necessários para análise, ainda que preenchidos com valores nulos quando inexistentes na fonte original.

#### 4.2.8.2 Validações Específicas por Domínio

Foram aplicadas validações específicas para atributos críticos:

**Identificadores institucionais e pessoais:** CNPJ, CPF e CNS passaram por processos de limpeza para remoção de caracteres não numéricos (pontos, hífen, barras), validação de dígitos verificadores quando aplicável, e padronização de formato com preenchimento de zeros à esquerda quando necessário.

**Localização geográfica:** campos de coordenadas geográficas foram extraídos de strings concatenadas, validados quanto aos limites de latitude ( $-90^\circ$  a  $+90^\circ$ ) e longitude ( $-180^\circ$  a  $+180^\circ$ ), e verificados quanto à consistência com a UF declarada.

**Atributos temporais:** campos de competência foram padronizados no formato YYYYMM, datas de cadastro e desativação foram validadas quanto à cronologia lógica, e períodos de referência inconsistentes foram sinalizados.

Além disso, foram realizados tratamentos voltados à eliminação de duplicidades, especialmente em conjuntos de dados nos quais a granularidade original da fonte poderia gerar múltiplas ocorrências para uma mesma entidade analítica. Registros duplicados foram identificados por combinações de chaves naturais (ex: CNES + competência, CPF + competência) e consolidados segundo regras de precedência definidas.

#### 4.2.8.3 Camada Staging

A camada de *staging* desempenha papel de consolidação e padronização dos dados antes de sua disponibilização para consumo analítico. Nessa camada, os dados oriundos das fontes brutas são reorganizados estruturalmente, com aplicação de regras de limpeza, normalização e validação.

Foram definidas três estruturas de *staging*, cada uma com tratamentos específicos de acordo com o domínio dos dados analisados:

- **stg\_estabelecimentos:** consolida as informações cadastrais, administrativas e geográficas dos estabelecimentos de saúde. Durante a carga, foram aplicadas regras de padronização de identificadores, como CNPJ e CPF, por meio da remoção de caracteres não numéricos, além da normalização de campos textuais relacionados a razão social, nome fantasia e atributos de localização. Para os campos de coordenadas geográficas, foi realizada a extração e validação de latitude e longitude a partir de campos textuais, com tratamento defensivo para valores inválidos, garantindo a integridade das informações espaciais utilizadas nas análises.
- **stg\_leitos:** prepara os dados referentes à capacidade instalada dos estabelecimentos de saúde. Nessa tabela, foram padronizados os atributos relacionados ao tipo de leito, especialidade e unidade federativa, bem como aplicadas validações sobre os campos quantitativos. Esses dados foram organizados de forma a permitir operações de agregação posteriores, necessárias para o cálculo de métricas consolidadas, como o total de leitos existentes, contratados e destinados ao Sistema Único de Saúde.
- **stg\_profissionais:** concentra as transformações aplicadas aos dados de vínculos e perfis dos profissionais de saúde. As operações incluem limpeza e padronização de identificadores pessoais e institucionais, separação de códigos e descrições de

classificações ocupacionais (CBO), normalização de campos relacionados à carga horária e tratamento defensivo de atributos temporais, como ano e competência. Essas transformações permitem análises consistentes sobre a distribuição e atuação da força de trabalho em saúde.

As tabelas *staging* resultantes apresentam a seguinte organização:

Tabela 3 – Estrutura da tabela `stg_estabelecimentos` (principais campos)

<b>Campo</b>	<b>Tipo</b>	<b>Descrição</b>
<code>cnes</code>	VARCHAR	Código nacional do estabelecimento
<code>cnpj_estabelecimento</code>	VARCHAR	CNPJ normalizado (apenas dígitos)
<code>tipo_estabelecimento</code>	VARCHAR	Classificação do estabelecimento
<code>convenio_sus</code>	VARCHAR	Indica vínculo com SUS (Sim/Não)
<code>latitude</code>	NUMERIC	Latitude extraída e validada
<code>longitude</code>	NUMERIC	Longitude extraída e validada
<code>municipio</code>	VARCHAR	Nome do município (INITCAP)
<code>uf</code>	VARCHAR	Sigla UF (UPPER)
<code>competencia_raw</code>	VARCHAR	Competência no formato original
<code>ano</code>	INTEGER	Ano extraído e convertido

Fonte: Elaborado pelo autor

Tabela 4 – Estrutura da tabela `stg_leitos` (principais campos)

<b>Campo</b>	<b>Tipo</b>	<b>Descrição</b>
<code>id_leito</code>	VARCHAR	Identificador único do registro
<code>cnes</code>	VARCHAR	Código do estabelecimento
<code>codigo_tipo_leito</code>	VARCHAR	Código da tipologia do leito
<code>tipo_leito</code>	VARCHAR	Descrição do tipo de leito
<code>total_leitos_sus</code>	INTEGER	Quantidade total destinada ao SUS
<code>total_leitos_existentes</code>	INTEGER	Quantidade total instalada
<code>convenio_sus</code>	VARCHAR	Vinculação ao SUS (Sim/Não)
<code>municipio</code>	VARCHAR	Município (padronizado)
<code>competencia</code>	VARCHAR	Período de referência

Fonte: Elaborado pelo autor

Tabela 5 – Estrutura da tabela `stg_profissionais` (principais campos)

<b>Campo</b>	<b>Tipo</b>	<b>Descrição</b>
<code>profissional_vinculo</code>	VARCHAR	Identificador do vínculo
<code>profissional_cns</code>	VARCHAR	Cartão Nacional de Saúde
<code>cbo_codigo</code>	VARCHAR	Código CBO extraído
<code>cbo_descricao</code>	VARCHAR	Descrição da ocupação
<code>cpf</code>	VARCHAR	CPF normalizado
<code>cnes</code>	VARCHAR	Estabelecimento de atuação
<code>profissional_atende_sus</code>	VARCHAR	Atende pelo SUS (sim/não)
<code>carga_horaria_total</code>	VARCHAR	Carga horária semanal
<code>equipe_nome</code>	VARCHAR	Nome da equipe vinculada
<code>municipio</code>	VARCHAR	Município de atuação

Fonte: Elaborado pelo autor

Ao término dessa etapa, os dados encontram-se padronizados, estruturalmente consistentes e semanticamente alinhados aos objetivos analíticos do projeto. Dessa forma, a camada de *staging* passa a servir como base confiável para a construção das tabelas fato e dimensões, viabilizando análises exploratórias, agregações em múltiplos níveis de granularidade e a construção de *dashboards* mais completos e informativos.

## 4.2.9 Modelagem dos Dados

Com os dados consolidados na camada anterior, foi realizada a modelagem dimensional seguindo princípios de Kimball, priorizando estruturas que facilitassem navegação, filtragem e combinação dos dados em diferentes recortes analíticos.

Adotou-se modelo dimensional com tabelas dimensão ricas em atributos descritivos (localização geográfica, tipo de estabelecimento, especialidade, tipo de leito, vínculo profissional, competência temporal) e tabelas fato contendo métricas agregadas. As dimensões assumem papel central na organização das análises, permitindo que usuários combinem filtros e cruzamentos sem necessidade de cálculos complexos na camada de visualização. A granularidade das tabelas fato foi definida para suportar análises comparativas e temporais requeridas pelos *dashboards*.

### 4.2.9.1 Tabelas Fato e Dimensões

A modelagem resultou na criação de uma tabela fato e quatro tabelas dimensão, organizadas nos schemas `fct` e `dim` do PostgreSQL.

**fct\_leitos:** Consolida métricas quantitativas sobre disponibilidade de leitos nos estabelecimentos de saúde. A granularidade está definida no nível de estabelecimento (CNES), tipo de leito, especialidade, unidade federativa e competência temporal, permitindo análises detalhadas e comparativas ao longo do tempo.

A tabela é construída a partir de `stg_leitos` através de agregação que elimina duplicidades da fonte original, consolidando por combinação única de dimensões. As métricas base incluem total de leitos existentes, leitos contratados, leitos destinados ao SUS e leitos complementares.

Além das métricas agregadas, a tabela incorpora indicadores derivados calculados durante a carga para otimizar performance das consultas analíticas por exemplo:

- **percentual\_sus**: Percentual de leitos destinados ao SUS em relação ao total existente

```
(total_leitos_sus::NUMERIC /
  NULLIF(total_leitos_existentes, 0) * 100)
```

### Estrutura da tabela:

Tabela 6 – Estrutura da tabela fato `fct_leitos`

Campo	Tipo	Descrição
<code>fct_leitos_sk</code>	INTEGER	Surrogate key (sequencial)
<code>cnes</code>	VARCHAR	Código do estabelecimento
<code>tipo_leito</code>	VARCHAR	Código do tipo de leito
<code>tipo_leito_descricao</code>	VARCHAR	Descrição do tipo
<code>especialidade</code>	VARCHAR	Código da especialidade
<code>especialidade_descricao</code>	VARCHAR	Descrição da especialidade
<code>competencia</code>	VARCHAR	Período de referência (YYYYMM)
<code>ano</code>	INTEGER	Ano de referência
<code>uf</code>	VARCHAR	Unidade federativa
<code>total_leitos_existentes</code>	INTEGER	Soma de leitos instalados
<code>total_leitos_contratados</code>	INTEGER	Soma de leitos contratados
<code>total_leitos_sus</code>	INTEGER	Soma de leitos destinados ao SUS
<code>percentual_sus</code>	NUMERIC	% de leitos SUS sobre existentes
<code>diferenca_contratado_existente</code>	INTEGER	Diferença contratado - existente

Fonte: Elaborado pelo autor

As tabelas dimensão foram desenvolvidas com o objetivo de fornecer contexto descritivo às métricas armazenadas na tabela fato. A dimensão de leitos, por exemplo, reúne atributos relacionados à classificação dos leitos, características dos estabelecimentos, informações de localização e aspectos administrativos, permitindo análises multidimensionais a partir de diferentes perspectivas. Cada registro é identificado por uma chave substituta, garantindo independência em relação às chaves naturais provenientes das fontes originais.

- **dim\_dates**: dimensão de tempo utilizada para suportar análises temporais e comparativas ao longo do período analisado. Essa dimensão foi construída de forma

independente, a partir da geração de uma série contínua de datas, abrangendo um intervalo pré-definido. A tabela contém atributos hierárquicos de ano, trimestre, mês, semana e dia, além de indicadores auxiliares, como identificação de finais de semana e marcação de início e fim de períodos. A utilização de uma dimensão de tempo dedicada permite padronizar análises temporais e evita dependência direta de atributos de data presentes nas tabelas fato.

Tabela 7 – Estrutura da dimensão `dim_dates`

<b>Campo</b>	<b>Tipo</b>	<b>Descrição</b>
<code>date_id</code>	INTEGER	PK no formato YYYYMMDD (ex: 20250115)
<code>dt</code>	DATE	Data (UNIQUE)
<code>year</code>	INTEGER	Ano
<code>quarter</code>	INTEGER	Trimestre (1-4)
<code>quarter_name</code>	TEXT	Nome do trimestre (T1, T2, ...)
<code>month</code>	INTEGER	Mês (1-12)
<code>month_id</code>	INTEGER	YYYYMM (ex: 202501)
<code>month_name_pt</code>	TEXT	Nome do mês em português
<code>month_abbr_pt</code>	TEXT	Abreviação do mês (jan, fev, ...)
<code>day</code>	INTEGER	Dia do mês (1-31)
<code>iso_week</code>	INTEGER	Semana ISO do ano
<code>isodow</code>	INTEGER	Dia da semana ISO (1=seg, 7=dom)
<code>weekday_name_pt</code>	TEXT	Nome do dia em português
<code>is_weekend</code>	BOOLEAN	Indica se é final de semana
<code>first_day_of_month</code>	DATE	Primeiro dia do mês
<code>last_day_of_month</code>	DATE	Último dia do mês
<code>is_month_start</code>	BOOLEAN	Indica início de mês
<code>is_month_end</code>	BOOLEAN	Indica fim de mês

Fonte: Elaborado pelo autor

- **dim\_leitos:** dimensão responsável por armazenar os atributos descritivos relacionados aos leitos e aos estabelecimentos de saúde aos quais estão associados. Essa dimensão consolida informações de classificação dos leitos, características do estabelecimento, localização geográfica, perfil de atendimento e aspectos administrativos. Cada registro é identificado por uma chave substituta, garantindo independência em relação às chaves naturais da fonte original e facilitando a integração com a tabela fato para análises multidimensionais.

Tabela 8 – Estrutura completa da dimensão `dim_leitos`

Campo	Tipo	Descrição
<code>leitos_sk</code>	INTEGER	Surrogate key sequencial
<code>id_leito</code>	VARCHAR	Identificador único do registro de leito
<code>codigo_leito</code>	VARCHAR	Código alfanumérico do leito
<code>cnes</code>	VARCHAR	Código nacional do estabelecimento de saúde
<code>codigo_tipo_leito</code>	VARCHAR	Código da tipologia do leito
<code>tipo_leito</code>	VARCHAR	Descrição textual do tipo de leito
<code>leito</code>	VARCHAR	Classificação específica do leito
<code>total_leitos_sus</code>	INTEGER	Quantidade total de leitos destinados ao SUS
<code>total_leitos_existentes</code>	INTEGER	Quantidade total de leitos instalados
<code>leitos_existentes</code>	VARCHAR	Indicador textual de leitos existentes
<code>leitos_sus</code>	VARCHAR	Indicador textual de leitos destinados ao SUS
<code>convenio_sus</code>	VARCHAR	Indica vínculo com SUS (Sim/Não)
<code>subtipo_estabelecimento</code>	VARCHAR	Subclassificação do estabelecimento
<code>tipo_estabelecimento</code>	VARCHAR	Classificação principal do estabelecimento
<code>uf</code>	VARCHAR	Sigla da unidade federativa (2 caracteres)
<code>uf_extenso</code>	VARCHAR	Nome completo da unidade federativa
<code>municipio</code>	VARCHAR	Nome do município de localização
<code>codigo_municipio</code>	VARCHAR	Código IBGE do município
<code>bairro</code>	VARCHAR	Nome do bairro
<code>cep</code>	VARCHAR	Código de endereçamento postal
<code>logradouro</code>	VARCHAR	Nome da rua, avenida ou logradouro
<code>complemento_endereco</code>	VARCHAR	Complemento do endereço (sala, andar, etc.)
<code>coordenadas_geograficas</code>	VARCHAR	Coordenadas no formato original
<code>location</code>	VARCHAR	Informação de localização georreferenciada
<code>forma_acesso</code>	VARCHAR	Descrição da forma de acesso ao local
<code>macrorregiao</code>	VARCHAR	Macrorregião de saúde
<code>nivel_atencao</code>	VARCHAR	Nível de atenção à saúde (primária, secundária, terciária)
<code>turno_atendimento</code>	VARCHAR	Turnos de funcionamento do estabelecimento
<code>horario_atendimento</code>	VARCHAR	Horários específicos de funcionamento
<code>sempre_aberto</code>	VARCHAR	Indica funcionamento 24 horas (Sim/Não)
<code>cnpj_estabelecimento</code>	VARCHAR	CNPJ do estabelecimento de saúde
<code>razao_social</code>	VARCHAR	Razão social oficial
<code>nome_fantasia</code>	VARCHAR	Nome fantasia ou comercial
<code>status_estabelecimento</code>	VARCHAR	Status atual (Ativo/Desativado)
<code>natureza_juridica_categoria</code>	VARCHAR	Categoria ampla da natureza jurídica
<code>natureza_juridica</code>	VARCHAR	Natureza jurídica detalhada
<code>gestao</code>	VARCHAR	Tipo de gestão (Municipal, Estadual, etc.)
<code>gestao_sigla</code>	VARCHAR	Sigla do tipo de gestão
<code>nivel_dependencia</code>	VARCHAR	Nível de dependência hierárquica
<code>atividade_principal</code>	VARCHAR	Descrição da atividade principal do estabelecimento
<code>cnpj_mantenedora</code>	VARCHAR	CNPJ da entidade mantenedora
<code>diretor_responsavel</code>	VARCHAR	Nome do diretor ou responsável técnico
<code>data_cadastro</code>	DATE	Data de cadastramento no CNES
<code>data_desativacao</code>	DATE	Data de desativação (quando aplicável)
<code>competencia</code>	VARCHAR	Competência de referência no formato YYYYMM
<code>ano</code>	INTEGER	Ano de referência dos dados
<code>data_carga</code>	TIMESTAMP	Timestamp da carga no data warehouse

Fonte: Elaborado pelo autor

Cada registro é identificado por uma surrogate key (`leitos_sk`), garantindo independência em relação às chaves naturais provenientes das fontes originais.

- **`dim_panorama_geral`**: dimensão construída para consolidar informações cadastrais e estruturais dos estabelecimentos de saúde, com foco em análises gerais da rede assistencial. Essa tabela reúne atributos administrativos, geográficos e operacionais, permitindo análises de cobertura, distribuição territorial e perfil dos estabelecimentos. Sua utilização favorece a construção de painéis de visão macro, integrando dados de diferentes domínios em um único ponto de referência analítica, e apesar da alteração de sua nomenclatura provinda do próprio painel CNES esta utiliza a `stg_estabelecimentos` de base.

Tabela 9 – Estrutura completa da dimensão `dim_panorama_geral`

Campo	Tipo	Descrição
<code>cnes</code>	VARCHAR	Código nacional do estabelecimento (chave natural)
<code>codigo_municipio</code>	VARCHAR	Código IBGE do município
<code>cnpj_estabelecimento</code>	VARCHAR	CNPJ do estabelecimento de saúde
<code>cnpj_mantenedora</code>	VARCHAR	CNPJ da entidade mantenedora
<code>tipo_estabelecimento</code>	VARCHAR	Classificação do tipo de estabelecimento
<code>subtipo_estabelecimento</code>	VARCHAR	Subclassificação do estabelecimento
<code>convenio_sus</code>	VARCHAR	Indica se possui convênio com SUS (Sim/Não)
<code>leitos_sus</code>	VARCHAR	Quantidade de leitos destinados ao SUS
<code>leitos_existentes</code>	VARCHAR	Quantidade total de leitos instalados
<code>uf</code>	VARCHAR	Sigla da unidade federativa
<code>municipio</code>	VARCHAR	Nome do município de localização
<code>bairro</code>	VARCHAR	Nome do bairro
<code>cep</code>	VARCHAR	Código de endereçamento postal
<code>logradouro</code>	VARCHAR	Nome completo do logradouro
<code>complemento_endereco</code>	VARCHAR	Complemento do endereço
<code>latitude</code>	NUMERIC	Coordenada geográfica de latitude
<code>longitude</code>	NUMERIC	Coordenada geográfica de longitude
<code>macrorregiao</code>	VARCHAR	Macrorregião de saúde
<code>nivel_atencao</code>	VARCHAR	Nível de atenção à saúde
<code>turno_atendimento</code>	VARCHAR	Turnos de funcionamento
<code>atendimento_prestado</code>	VARCHAR	Tipos de atendimento oferecidos
<code>natureza_juridica_categoria</code>	VARCHAR	Categoria da natureza jurídica
<code>natureza_juridica</code>	VARCHAR	Descrição completa da natureza jurídica
<code>razao_social</code>	VARCHAR	Razão social oficial do estabelecimento
<code>nome_fantasia</code>	VARCHAR	Nome fantasia ou comercial
<code>gestao</code>	VARCHAR	Tipo de gestão (Municipal, Estadual, Federal, Dupla)
<code>diretor_responsavel</code>	VARCHAR	Nome do diretor ou responsável legal
<code>email</code>	VARCHAR	Endereço de e-mail para contato
<code>telefone</code>	VARCHAR	Número de telefone para contato
<code>atividade_principal</code>	VARCHAR	Descrição da atividade principal desenvolvida
<code>competencia_raw</code>	VARCHAR	Competência de referência (formato original)
<code>ano</code>	INTEGER	Ano de referência dos dados
<code>status_estabelecimento</code>	VARCHAR	Status atual (Ativo, Desativado, etc.)
<code>data_cadastro</code>	DATE	Data de cadastramento no CNES
<code>data_desativacao</code>	DATE	Data de desativação (quando aplicável)

Fonte: Elaborado pelo autor

- **dim\_profissionais:** dimensão destinada a representar os atributos descritivos dos profissionais de saúde e seus vínculos com os estabelecimentos. A tabela concentra informações relacionadas à classificação profissional, carga horária, tipo de vínculo, atendimento ao SUS e localização, permitindo análises sobre a distribuição da força de trabalho em saúde. A adoção de uma chave substituta assegura consistência na integração com estruturas analíticas e flexibilidade para análises históricas e comparativas.

É importante destacar que, embora os dados extraídos contenham atributos de identificação pessoal como CPF e nome de profissionais de saúde, esses campos não são exibidos na interface pública da plataforma, mas estão presentes nas respostas JSON retornadas pela API não documentada. Neste trabalho, tais atributos foram preservados na camada RAW por integridade técnica do processo de extração, porém não foram utilizados como critério analítico em nenhum dos *dashboards* desenvolvidos. Em conformidade com os princípios da Lei Geral de Proteção de Dados (LGPD — Lei nº 13.709/2018) (Brasil, 2018), recomenda-se que implementações futuras avaliem a necessidade de anonimização ou

supressão desses campos ainda na camada de ingestão, anterior à persistência na camada RAW.

Tabela 10 – Estrutura completa da dimensão `dim_profissionais`

<b>Campo</b>	<b>Tipo</b>	<b>Descrição</b>
<code>profissionais_id</code>	INTEGER	Surrogate key sequencial
<code>profissional_vinculo</code>	VARCHAR	Identificador único do vínculo profissional
<code>profissional_cns</code>	VARCHAR	Número do Cartão Nacional de Saúde
<code>cbo_codigo</code>	VARCHAR	Código da Classificação Brasileira de Ocupações
<code>cbo_descricao</code>	VARCHAR	Descrição da ocupação profissional
<code>profissional_nome</code>	VARCHAR	Nome completo do profissional
<code>cnes</code>	VARCHAR	Código do estabelecimento onde atua
<code>cnpj_estabelecimento</code>	VARCHAR	CNPJ do estabelecimento de atuação
<code>ibge</code>	VARCHAR	Código IBGE do município de atuação
<code>tipo_unidade</code>	VARCHAR	Tipo da unidade de saúde
<code>subtipo_unidade</code>	VARCHAR	Subtipo da unidade de saúde
<code>uf</code>	VARCHAR	Sigla da unidade federativa de atuação
<code>regiao</code>	VARCHAR	Região geográfica do país
<code>municipio</code>	VARCHAR	Nome do município de atuação
<code>logradouro</code>	VARCHAR	Endereço de atuação do profissional
<code>complemento</code>	VARCHAR	Complemento do endereço
<code>profissional_atende_sus</code>	VARCHAR	Indica se atende pelo SUS (sim/não)
<code>convenio_sus</code>	VARCHAR	Indica convênio do estabelecimento com SUS
<code>carga_horaria_ambulatorial_sus</code>	VARCHAR	Horas semanais em atendimento ambulatorial SUS
<code>carga_horaria_hospitalar_sus</code>	VARCHAR	Horas semanais em atendimento hospitalar SUS
<code>carga_horaria_outros</code>	VARCHAR	Horas semanais em outras atividades
<code>carga_horaria_total</code>	VARCHAR	Carga horária semanal total
<code>qt_leitos_sus</code>	VARCHAR	Quantidade de leitos SUS sob responsabilidade
<code>atendimento_prestado</code>	VARCHAR	Tipos de atendimento que presta
<code>equipe_area</code>	VARCHAR	Área de atuação da equipe vinculada
<code>equipe_nome</code>	VARCHAR	Nome da equipe à qual está vinculado
<code>equipe_tipo</code>	VARCHAR	Tipo da equipe (ESF, NASF, etc.)
<code>equipe_ine</code>	VARCHAR	Identificador Nacional de Equipe (INE)
<code>equipe_dt_entrada</code>	VARCHAR	Data de entrada na equipe
<code>equipe_dt_desligamento</code>	VARCHAR	Data de desligamento da equipe
<code>equipe_dt_ativacao</code>	VARCHAR	Data de ativação da equipe
<code>equipe_dt_desativacao</code>	VARCHAR	Data de desativação da equipe
<code>equipe_subtipo</code>	VARCHAR	Subtipo da equipe
<code>nome_fantasia</code>	VARCHAR	Nome fantasia do estabelecimento
<code>gestao</code>	VARCHAR	Tipo de gestão do estabelecimento
<code>natureza_juridica</code>	VARCHAR	Natureza jurídica do estabelecimento
<code>cnpj_mantenedora</code>	VARCHAR	CNPJ da entidade mantenedora
<code>ano</code>	INTEGER	Ano de referência dos dados
<code>competencia_raw</code>	VARCHAR	Competência de referência (formato original)
<code>telefone</code>	VARCHAR	Telefone de contato
<code>nivel_atencao</code>	VARCHAR	Nível de atenção à saúde do estabelecimento
<code>data_carga</code>	TIMESTAMP	Timestamp da carga dos dados no data warehouse

Fonte: Elaborado pelo autor

A Tabela 11 apresenta um comparativo da estrutura dos dados ao longo das três camadas do pipeline (RAW, Staging e Dimensional), evidenciando o processo de refinamento e organização aplicado:

Tabela 11 – Evolução da estrutura de dados por camada

<b>Domínio</b>	<b>RAW</b>	<b>Staging</b>	<b>Dimensional</b>
Estabelecimentos	62 colunas	62 colunas	34 colunas (dim)
Leitos	64 colunas	64 colunas	46 colunas (dim) + 14 (fato)
Profissionais	54 colunas	54 colunas	43 colunas (dim)
Temporal	-	-	18 colunas (dim)
<b>Total</b>	<b>180 colunas</b>	<b>180 colunas</b>	<b>155 colunas</b>

Fonte: Elaborado pelo autor

A camada RAW preserva integralmente a estrutura original dos dados extraídos do ElasticCNES, mantendo todas as 180 colunas distribuídas entre as três tabelas principais. Já na camada Dimensional, observa-se redução para 155 colunas, resultado de:

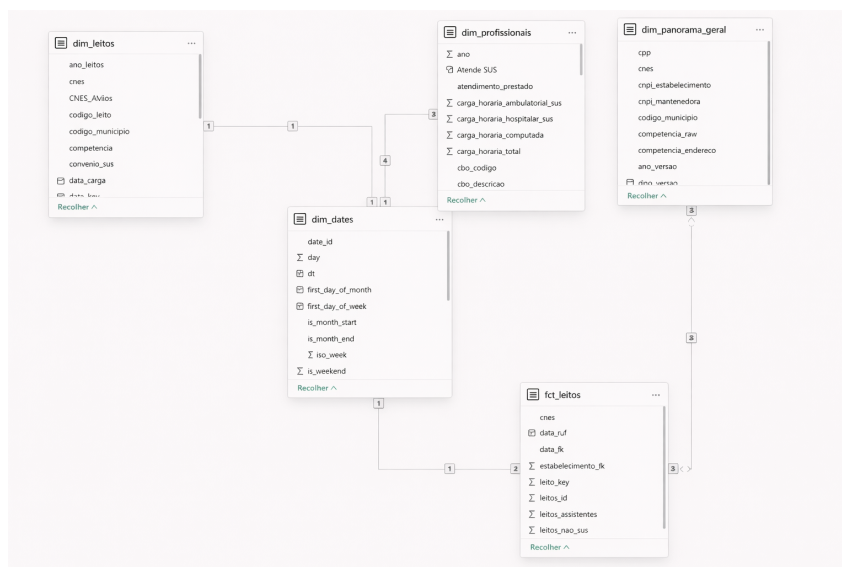
Essa evolução demonstra o processo de refinamento dos dados, onde a volumetria é gradualmente otimizada sem perda de informação relevante, resultando em estruturas mais eficientes para consumo analítico. A redução de 180 para 155 colunas (redução de aproximadamente 14%) reflete a eliminação de metadados técnicos e campos operacionais que não agregam valor às análises de gestão em saúde pública.

Ao término da modelagem, os dados encontram-se organizados em estruturas dimensionais otimizadas para consumo analítico, prontas para serem consumidas pelos *dashboards* desenvolvidos no Power BI, conforme descrito na seção seguinte.

#### 4.2.10 Relacionamentos e Integridade Referencial

O modelo dimensional não impõe constraints de chave estrangeira no banco de dados para favorecer performance de carga e flexibilidade, mas estabelece relacionamentos lógicos entre fato e dimensões através de chaves compartilhadas. A Figura 11 apresenta o diagrama do modelo implementado.

Figura 11 – Modelo dimensional implementado no Power Bi



Fonte: Elaborado pelo autor

A tabela fato `fct_leitos` relaciona-se com as dimensões através de chaves compartilhadas: `cnes` conecta aos estabelecimentos em `dim_panorama_geral`, `data_fk` vincula-se à dimensão temporal `dim_dates`, e atributos de leitos conectam-se à `dim_leitos`. Essas conexões permitem análises multidimensionais, como filtrar leitos por tipo de estabelecimento, localização geográfica ou período temporal.

No Power BI, esses relacionamentos lógicos são implementados através de relações entre tabelas, permitindo navegação intuitiva pelos dados sem necessidade de joins complexos na camada de visualização. Essa abordagem prioriza performance e simplicidade de consumo, adequando-se ao contexto de visualização de dados públicos de saúde.

## 4.3 Projeto de Dashboard e Visualizações

O projeto dos *dashboards* foi conduzido a partir de análise exploratória dos dados disponíveis, mapeando estrutura, granularidade e potencial analítico para definir indicadores, métricas derivadas e recortes analíticos a serem implementados nos painéis.

### 4.3.1 Mockups

A prototipação foi realizada no Figma através de processo estruturado: (1) *brainstorming* para definir escopo de cada painel, indicadores principais e interações esperadas; (2) elaboração de *mockups* e design system especificando hierarquia informacional, disposição de componentes, paleta de cores, tipografia e elementos gráficos.

A Figura 12 apresenta o artefato de *brainstorming* desenvolvido na concepção inicial, mapeando objetivos analíticos, indicadores, recortes de análise e organização das informações em cada aba. Essa representação visual permitiu explorar diferentes arranjos de layout e hierarquia antes da implementação técnica.

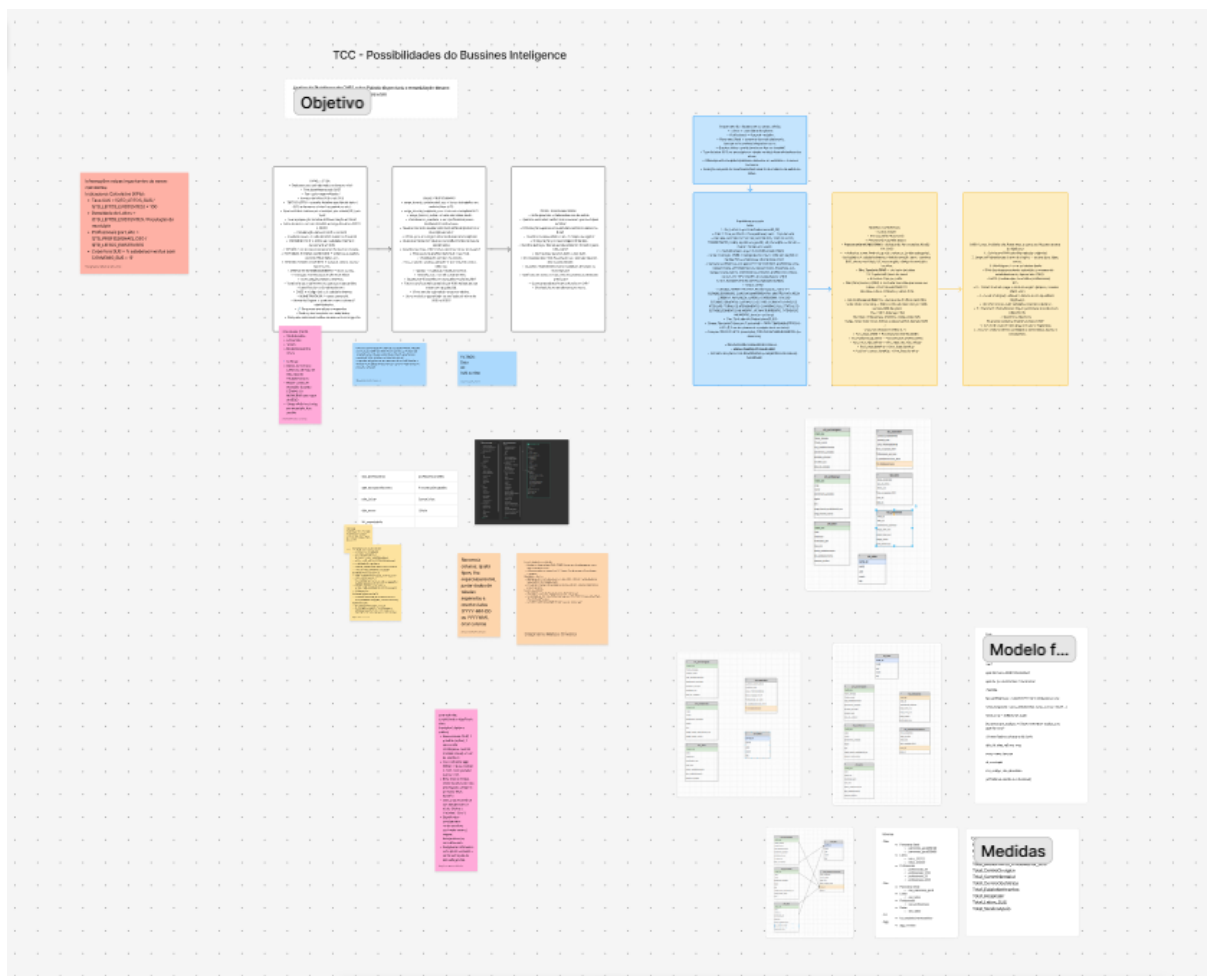


Figura 12 – Brainstorm Inicial - Figma

Fonte: Elaborado pelo autor - Figma

O *design system* desenvolvido para a solução analítica teve como finalidade estabelecer um conjunto padronizado de diretrizes visuais e componentes reutilizáveis para a construção dos *dashboards*. Esse sistema definiu regras claras para o uso de cores, tipografia, espaçamentos, contrastes, estilos de gráficos e elementos de interface, assegurando consistência visual entre os diferentes painéis e facilitando a manutenção e evolução da solução ao longo do tempo.

A definição da paleta de cores considerou princípios de legibilidade, acessibilidade e hierarquia informacional. Foram adotadas cores com contraste mínimo de 4.5:1 entre texto e fundo conforme diretrizes WCAG 2.1 (W3C, 2018), utilizando paleta com saturação reduzida para métricas secundárias e cores de alta saturação reservadas para indicadores-chave e alertas. A hierarquia cromática seguiu o princípio de pré-atenção visual (FEW, 2013), destacando desvios e KPIs prioritários através de contraste seletivo.

Além dos aspectos estéticos, o *design system* incorporou diretrizes funcionais para a padronização de componentes analíticos, como filtros, cartões de indicadores (*KPIs*), tabelas e gráficos. Dessa forma, o *design system* atuou como elemento estruturante da

solução, orientando as decisões de visualização e assegurando consistência funcional e informacional ao longo dos painéis.

Essa abordagem está alinhada às recomendações da literatura de visualização de dados e usabilidade, que destacam a importância de padrões visuais consistentes, uso controlado de cores e organização clara dos elementos para reduzir a carga cognitiva e evitar distrações na interpretação dos dados [Few \(2006a\)](#).



Figura 13 – Desing System

Fonte: Elaborado pelo autor - Figma

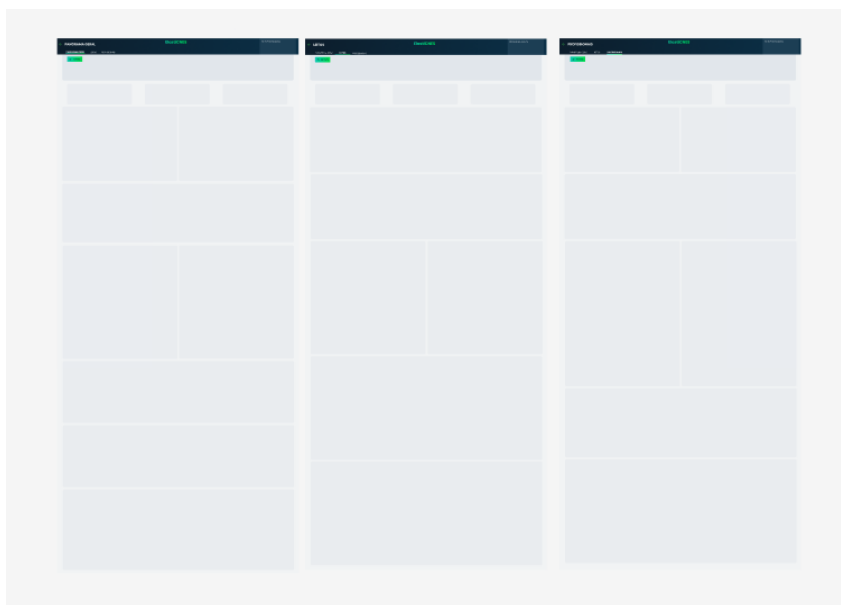


Figura 14 – Mockup Figma

Fonte: Elaborado pelo autor - Figma

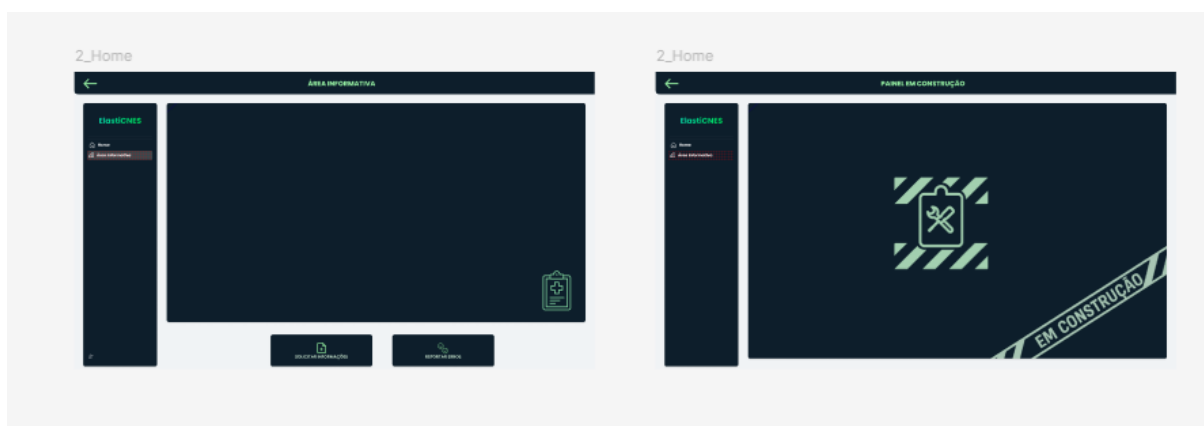


Figura 15 – Mockup Figma 2

Fonte: Elaborado pelo autor - Figma

### 4.3.2 Modulos de Visualização e Análise

O dashboard CNES desenvolvido foi estruturado em seis páginas principais, cada uma com foco analítico específico. A organização modular permite navegação intuitiva e exploração progressiva dos dados, desde visões agregadas até análises detalhadas por domínio.

- **Painel Inicial:** Tela de entrada do dashboard, projetada para fornecer uma visão macro da rede assistencial brasileira. Apresenta todos os tipos de painéis da plataforma com botões intrativos para guiar até os respectivos painéis.
- **Aba Leitos:** Painel dedicado à análise da capacidade instalada de leitos hospitalares, permitindo avaliação da oferta por tipo, especialidade, região e evolução temporal. Utiliza intensivamente a tabela fato `fct_leitos` e suas métricas derivadas, incluindo percentual de leitos SUS e diferença entre leitos contratados e existentes.
- **Aba Panorama Geral:** Módulo voltado à análise cadastral e estrutural dos estabelecimentos de saúde, oferecendo visões sobre distribuição geográfica, natureza jurídica, tipo de gestão e perfil assistencial da rede. Consome dados da dimensão `dim_panorama_geral`.
- **Aba Profissionais:** Painel especializado na análise da força de trabalho em saúde, permitindo avaliar a distribuição de profissionais por ocupação (CBO), carga horária, vínculo e localização geográfica. Baseia-se na dimensão `dim_profissionais`.
- **Área Informativa:** Página dedicada à documentação da solução, contendo informações gerais. Tem função didática, auxiliando usuários na interpretação correta dos indicadores.
- **Em Construção:** Tela de transição exibida ao acessar funcionalidades planejadas mas não implementadas nesta versão do dashboard. Serve como indicador de expansão futura da solução, mantendo a experiência do usuário consistente.

Todos os painéis compartilham padrões comuns de interação, incluindo filtros em todos os *dashboards* (competência, UF, região), capacidade de *drill-down* hierárquico e *tooltips* informativos. Essa estrutura modular promove uma experiência analítica consistente, reduzindo a curva de aprendizado e facilitando a exploração dos dados por diferentes perfis de usuários.

## 5 Resultados

Este capítulo apresenta os resultados obtidos com o desenvolvimento do sistema de Business Intelligence aplicado aos dados do Cadastro Nacional de Estabelecimentos de Saúde (CNES). A apresentação está organizada em duas seções principais: a interface do sistema e os *dashboards* analíticos desenvolvidos. Os valores são resultados da amostra de 2020 a 2025 e um cenário de testes foi aplicado para validação desses valores.

### 5.0.1 Menu Inicial do Sistema

O sistema desenvolvido apresenta uma interface inicial intuitiva e moderna, projetada para facilitar a navegação entre os diferentes módulos de análise de dados do CNES. O menu principal, ilustrado na Figura 16, foi estruturado com *cards* visuais organizados em *grid*, cada um representando uma categoria específica de análise da infraestrutura de saúde brasileira.

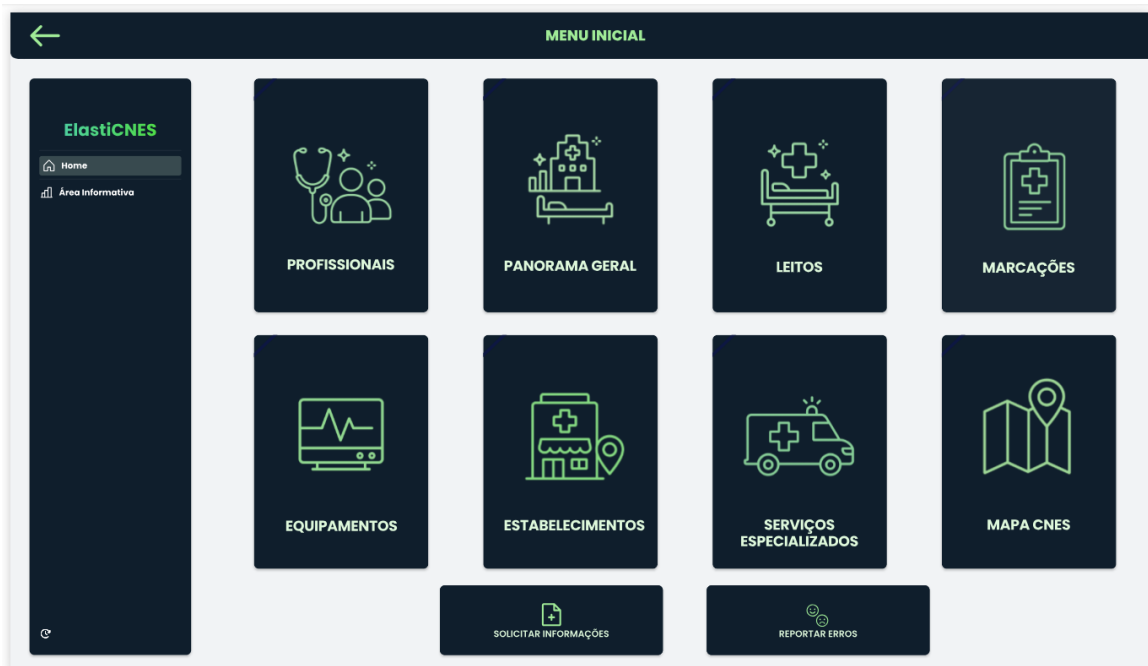


Figura 16 – Interface do menu inicial do sistema ElastiCNES

Fonte: Elaborado pelo autor

A arquitetura de navegação implementada inclui dez módulos principais: Profissionais, Panorama Geral, Leitos, Marcações, Equipamentos, Estabelecimentos, Serviços Especializados, Mapa CNES, além de funcionalidades auxiliares como Solicitar Informações e Reportar Erros. Cada *card* foi desenvolvido com ícones vetoriais personalizados que representam visualmente sua função, utilizando uma paleta de cores em tons de verde

sobre fundo escuro, proporcionando alto contraste e facilitando a identificação rápida das opções.

Um aspecto importante da experiência do usuário é a implementação de estados interativos nos elementos de navegação. Ao posicionar o cursor sobre qualquer *card*, o sistema exibe um *feedback* visual por meio de efeito *hover*, destacando o elemento selecionado e indicando sua interatividade, conforme demonstrado na Figura 17. Este comportamento foi implementado para melhorar a usabilidade e confirmar ao usuário qual opção está prestes a selecionar, seguindo princípios estabelecidos de design de interfaces.



Figura 17 – Efeito *hover* nos elementos do menu inicial

Fonte: Elaborado pelo autor

## 5.0.2 Dashboard Panorama Geral

A aba “Panorama Geral” constitui um dos módulos principais de análise do sistema, oferecendo uma visão consolidada da infraestrutura de saúde nacional. Este *dashboard* apresenta múltiplas visualizações e indicadores que permitem compreender diferentes aspectos da rede de estabelecimentos de saúde cadastrados no CNES.

### 5.0.2.1 Filtros e Dimensões de Análise

O sistema de filtragem implementado permite a segmentação dos dados através de sete dimensões principais: Ano, Estado, Município, Atendimento Prestado, Convênio SUS, Natureza Jurídica e Tipo de Estabelecimento, conforme apresentado na Figura 18. Esta funcionalidade possibilita análises customizadas conforme necessidades específicas de pesquisa, permitindo desde consultas abrangentes em nível nacional até investigações detalhadas em municípios específicos.

A interface de filtros foi posicionada na parte superior do *dashboard*, mantendo-se sempre visível e acessível. Todos os filtros utilizam componentes *dropdown* com a opção padrão “Todos”, garantindo que análises gerais possam ser realizadas sem necessidade de

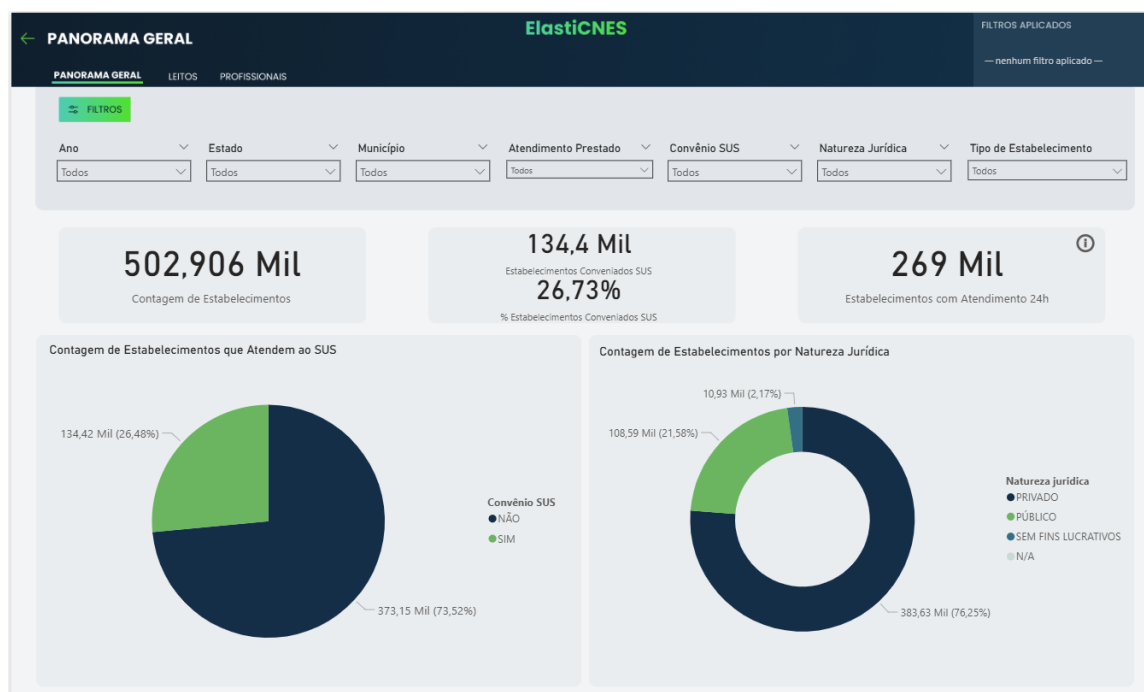


Figura 18 – Sistema de filtros multidimensionais do *dashboard* Panorama Geral

Fonte: Elaborado pelo autor

seleções específicas. Este *design* facilita tanto explorações amplas quanto análises focadas em contextos particulares.

### 5.0.2.2 Indicadores e Visuais

O dashboard apresenta três cards de indicadores na parte superior fornecendo síntese quantitativa da infraestrutura analisada (Figura 18): total de estabelecimentos cadastrados, estabelecimentos conveniados ao SUS, e estabelecimentos com atendimento 24 horas. Estes indicadores foram posicionados estrategicamente para fornecer contexto imediato, seguindo princípios de hierarquia visual.

A primeira visualização segmenta estabelecimentos por convênio SUS através de gráfico de pizza, evidenciando que aproximadamente três quartos da infraestrutura cadastrada opera fora do sistema público. O segundo gráfico categoriza por natureza jurídica, revelando predominância de estabelecimentos privados (76%), seguidos por públicos (22%) e sem fins lucrativos (2%).

O gráfico de barras verticais apresenta análise comparativa entre unidades federativas (Figura 19), com São Paulo e Minas Gerais liderando em volume. A visualização utiliza código de cores diferenciando estabelecimentos conveniados e não conveniados ao SUS, permitindo análise simultânea de volume total e proporção de cobertura por estado. Esta representação possibilita identificação de padrões regionais, como concentração nas regiões Sudeste e Sul.

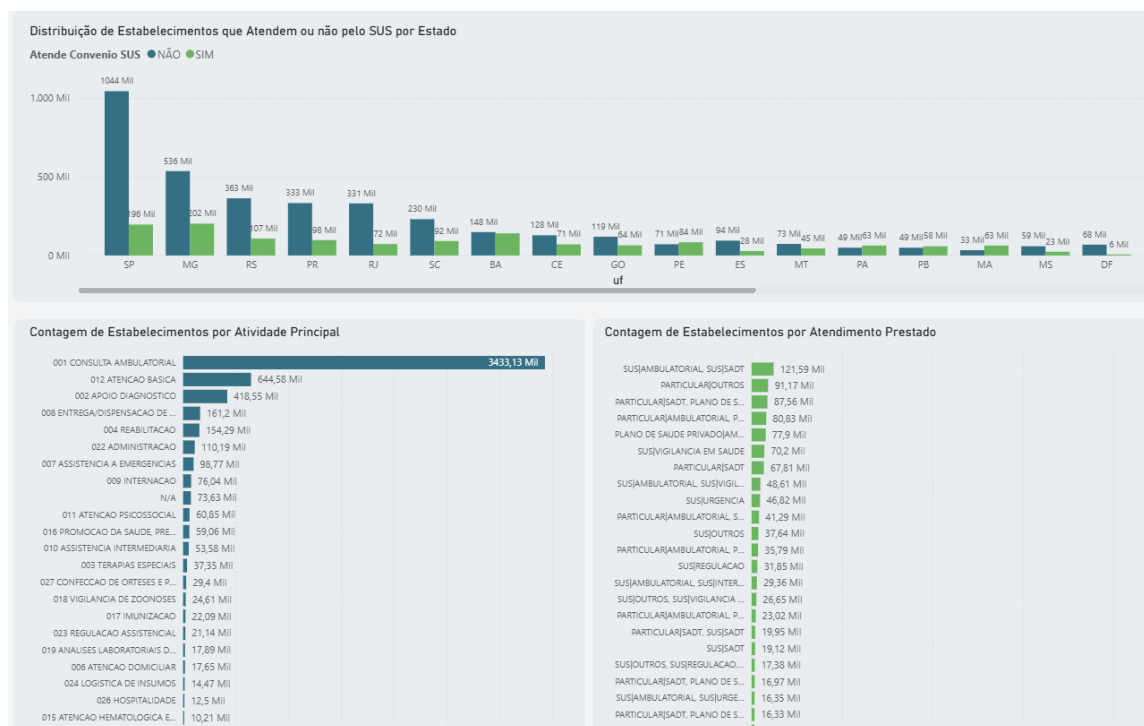


Figura 19 – Dashboard Panorâmica Geral

Fonte: Elaborado pelo autor

O gráfico horizontal de barras detalha distribuição por atividade principal, ordenado por volume, com predominância de consultas ambulatoriais e atenção básica. A visualização apresenta categorias de atividades incluindo apoio diagnóstico, reabilitação, administração, emergências, internação e atenção psicossocial.

Outro gráfico horizontal categoriza estabelecimentos por tipo de atendimento prestado, listando modalidades desde SUS ambulatorial até planos privados, urgência e vigilância em saúde, proporcionando visão detalhada das formas de prestação de serviço oferecidas.

### 5.0.2.3 Tipologia de Estabelecimentos

O gráfico horizontal mais extenso classifica estabelecimentos por tipo específico (Figura 20), com “Consultório Isolado” liderando seguido por “Clínica/Centro de Especialidade”. A visualização apresenta quinze tipos distintos, incluindo centros de saúde, farmácias, policlínicas, postos de saúde, hospitais, unidades móveis de urgência, polos acadêmicos, centros psicossociais e outros.

Um gráfico de pizza segmenta estabelecimentos por nível de atenção (Figura 20), mostrando predominância de unidades ambulatoriais com diversos códigos de classificação. A maior categoria individual representa 344,7 mil estabelecimentos (58,39%), enquanto outras categorias ambulatoriais somam proporções menores, totalizando uma distribuição que evidencia o perfil assistencial da rede.

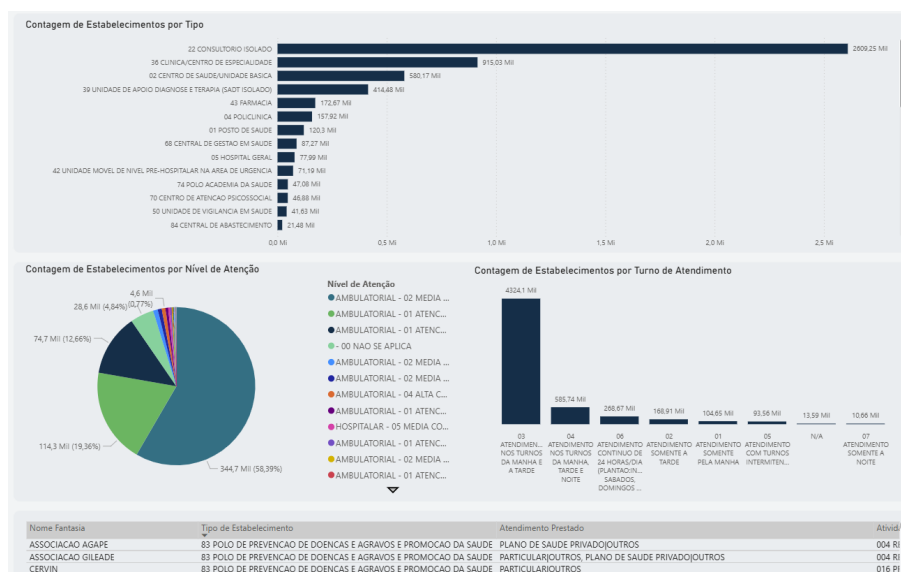


Figura 20 – Dashboard Panorama Geral 2

Fonte: Elaborado pelo autor

O gráfico de barras verticais analisa estabelecimentos por turno de atendimento (Figura 21), revelando que 4.324,1 mil unidades operam no turno matutino e vespertino (código 03). A visualização inclui sete categorias de turnos, desde atendimento contínuo 24 horas até turnos específicos da manhã, tarde e noturnos intermitentes, lembrando que uma grande parte dos estabelecimentos não possuem essa informação em seu cadastro.

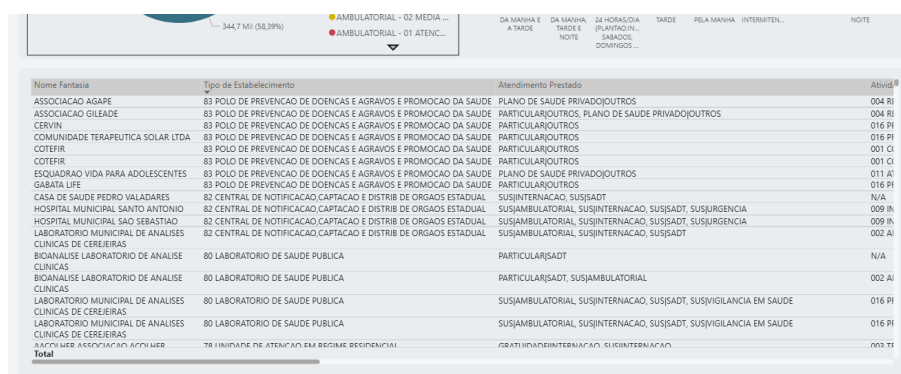


Figura 21 – Distribuição de estabelecimentos Panorama Geral

Fonte: Elaborado pelo autor

A seção inferior do dashboard apresenta tabela paginada com listagem detalhada de estabelecimentos individuais, conforme Figura 21. Cada linha exibe nome fantasia, tipo de estabelecimento, atendimento prestado e atividade principal. Exemplos incluem unidades como “ASSOCIACAO AGAPE” e “ASSOCIACAO GILEADE” classificadas como Polos de Prevenção, hospitais municipais, laboratórios de análises clínicas e diversos outros tipos de unidades de saúde.

Esta tabela permite consulta granular dos dados agregados apresentados nas visualizações, possibilitando verificação de casos específicos e análise de estabelecimentos

individuais.

O *dashboard* Panorama Geral consolida informações de múltiplas dimensões em interface única, permitindo análises desde visões macro nacionais até detalhamentos específicos por estabelecimento. A solução desenvolvida transforma dados brutos do CNES em informações acessíveis e acionáveis para gestão em saúde pública, cumprindo os objetivos estabelecidos para este trabalho.

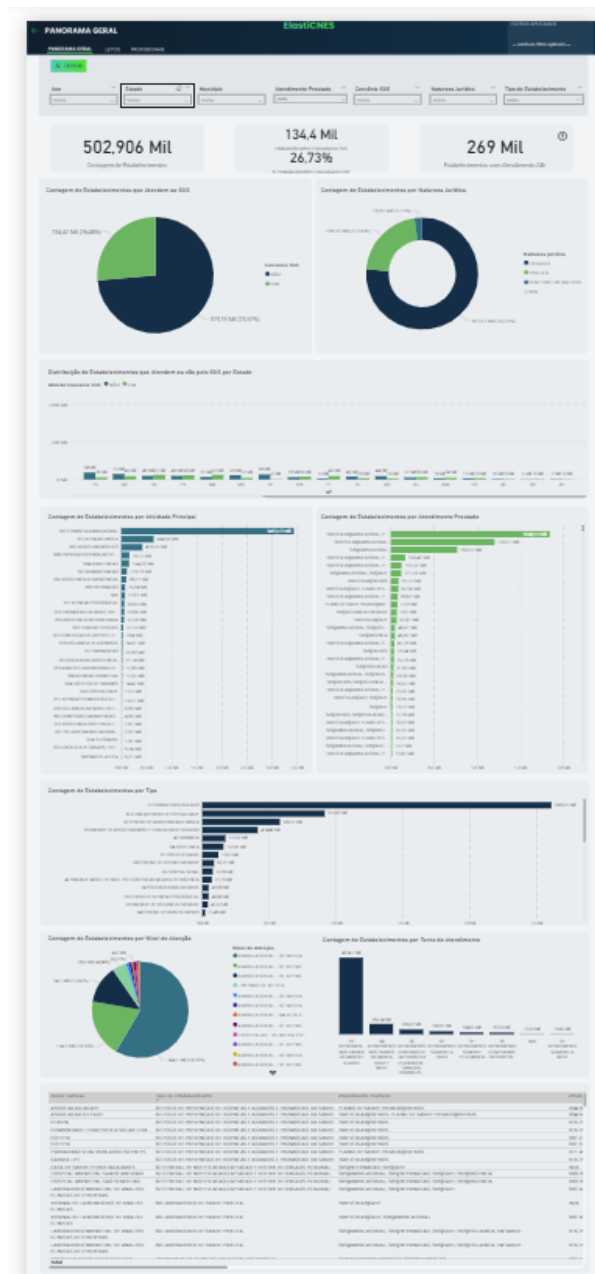


Figura 22 – Painel Panorama Geral

Fonte: Elaborado pelo autor

A arquitetura implementada demonstra capacidade de processar grandes volumes de dados governamentais, estruturá-los em modelos dimensionais otimizados e apresentá-los através de visualizações interativas que facilitam a tomada de decisão baseada em

evidências. Os recursos de filtragem multidimensional e navegação contextual permitem que diferentes perfis de usuários explorem os dados conforme suas necessidades específicas, desde gestores públicos até pesquisadores acadêmicos.

### 5.0.3 Dashboard de Leitos

O módulo “Leitos” apresenta análise detalhada da infraestrutura de leitos hospitalares disponíveis no sistema de saúde brasileiro, oferecendo visualizações específicas sobre capacidade instalada, distribuição geográfica e tipologias de leitos. Este *dashboard* complementa a análise do Panorama Geral com foco específico na disponibilidade de internação.

#### 5.0.3.1 Filtros e Sistema de Navegação

O *dashboard* de Leitos implementa sistema de filtragem com seis dimensões principais: Ano, Mês, Especialidade Leito, Tipo de Leito, Tipo de Estabelecimento, Estado e Cidade, conforme ilustrado na Figura 23. Esta estrutura permite análises temporais mensais e segmentações específicas por especialidades médicas e características dos leitos, adicionando granularidade temporal não presente no módulo de Panorama Geral.

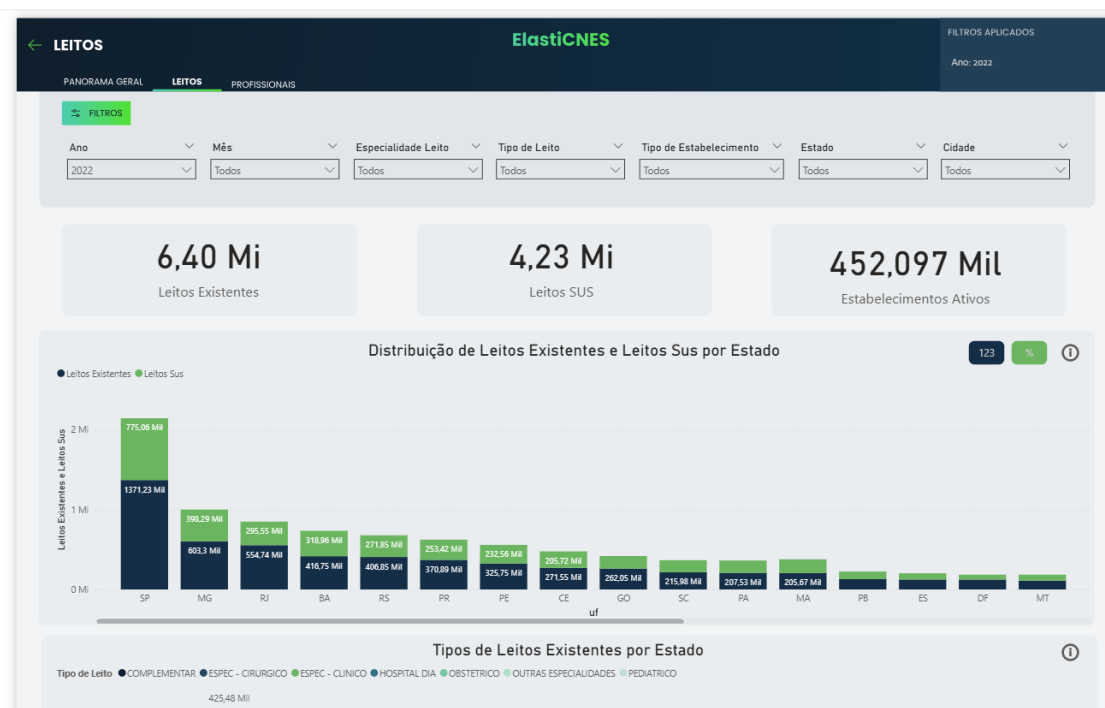


Figura 23 – Sistema de filtros do *dashboard* de Leitos

Fonte: Elaborado pelo autor

A interface mantém o padrão de usabilidade estabelecido, com filtros do tipo *dropdown* e opção padrão “Todos”, permitindo análises desde visões nacionais abrangentes até investigações específicas de tipos de leitos em municípios determinados.

### 5.0.3.2 Indicadores Principais e Visuais

Três indicadores principais quantificam a capacidade de internação nacional (Figura 23). O total de leitos, os leitos vinculados ao SUS e o terceiro indicador apresenta estabelecimentos ativos com leitos cadastrados.

A proporção de leitos SUS em relação ao total evidencia que dois terços da capacidade de internação nacional está disponível através do sistema público, proporção significativamente superior à observada na contagem de estabelecimentos conveniados ao SUS (26,73%), indicando que estabelecimentos públicos e conveniados concentram maior capacidade de internação.

O gráfico de barras verticais “Distribuição de Leitos Existentes e Leitos SUS por Estado” apresenta análise comparativa da capacidade instalada entre todas as unidades federativas (Figura 23). São Paulo lidera com aproximadamente 1.371,23 mil leitos totais, dos quais 775,06 mil são leitos SUS. Minas Gerais apresenta 603,3 mil leitos totais com 398,29 mil vinculados ao SUS, seguido por Rio de Janeiro com 554,74 mil leitos totais e 295,55 mil SUS.

A visualização utiliza código de cores diferenciando leitos totais (azul escuro) de leitos SUS (verde), permitindo análise simultânea de capacidade absoluta e proporção de leitos disponíveis ao sistema público por estado. Observa-se concentração significativa nas regiões Sudeste e Sul, com estados das regiões Norte e Centro-Oeste apresentando capacidades consideravelmente menores.

### 5.0.3.3 Tipologia de Leitos

O gráfico de barras agrupadas “Tipos de Leitos Existentes por Estado” detalha a distribuição de leitos segundo especialidades médicas (Figura 24). A visualização segmenta os dados em sete categorias principais: Complementar, Especialidade Cirúrgica, Especialidade Clínica, Hospital Dia, Obstétrico, Outras Especialidades e Pediátrico.

São Paulo apresenta maior diversidade e volume em todas as categorias, com destaque para leitos de especialidades clínicas (425,48 mil) e cirúrgicas (291,43 mil). Minas Gerais apresenta distribuição similar com 226,22 mil leitos cirúrgicos e 108,11 mil de outras especialidades. Esta segmentação permite identificar padrões de especialização regional e possíveis lacunas em determinadas especialidades médicas por estado.

O gráfico de pizza “Distribuição de Tipos de Leitos” quantifica a composição da capacidade nacional por especialidade (Figura 24). Esta distribuição evidencia o perfil assistencial da rede hospitalar brasileira, com predominância de leitos para procedimentos cirúrgicos e clínicos gerais, representando conjuntamente mais da metade da capacidade total instalada.

O *dashboard* incorpora visualização cartográfica do território brasileiro (Figura 24),

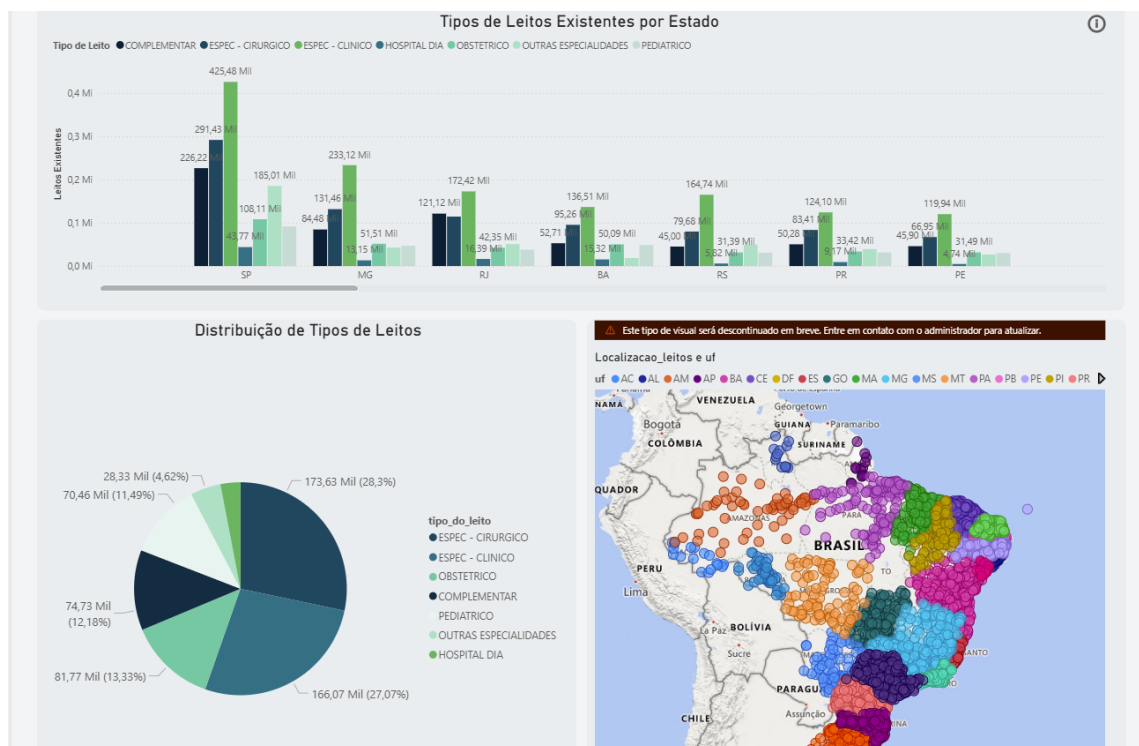


Figura 24 – Dashboard de Leitos

Fonte: Elaborado pelo autor

apresentando distribuição espacial de leitos através de marcadores circulares coloridos por unidade federativa. Esta representação geográfica complementa os gráficos quantitativos, permitindo identificação visual imediata de concentrações e vazios assistenciais no território nacional.

O mapa utiliza código de cores por estado e densidade de marcadores para representar concentração de estabelecimentos com leitos, evidenciando maior densidade nas regiões litorâneas e nos grandes centros urbanos, com menor cobertura nas regiões interioranas, especialmente na Amazônia Legal e Centro- Oeste.

O gráfico horizontal de barras “Total de Leitos Existentes por Tipo” apresenta classificação detalhada dos leitos por especialidade específica (Figura 25). A categoria “Clínica Geral” lidera com 1.668,93 mil leitos, seguida por “Cirurgia Geral” com 703,39 mil e “Pediatria Clínica” com 483,6 mil unidades. A visualização apresenta diversas categorias detalhadas, incluindo especialidades como Psiquiatria , UTI Adulto Tipo I, Obstetrícia Clínica, Obstetrícia Cirúrgica, Ortopedia Traumatológica entre outras. Esta segmentação granular permite identificar capacidades instaladas em especialidades críticas e áreas potencialmente deficitárias.

A seção inferior do dashboard apresenta tabela detalhada com informações individualizadas de leitos por estabelecimento (Figura 25). Cada registro exhibe município, unidade federativa, tipo de leito, convênio SUS, contagem de leitos, tipo de estabelecimento, nome

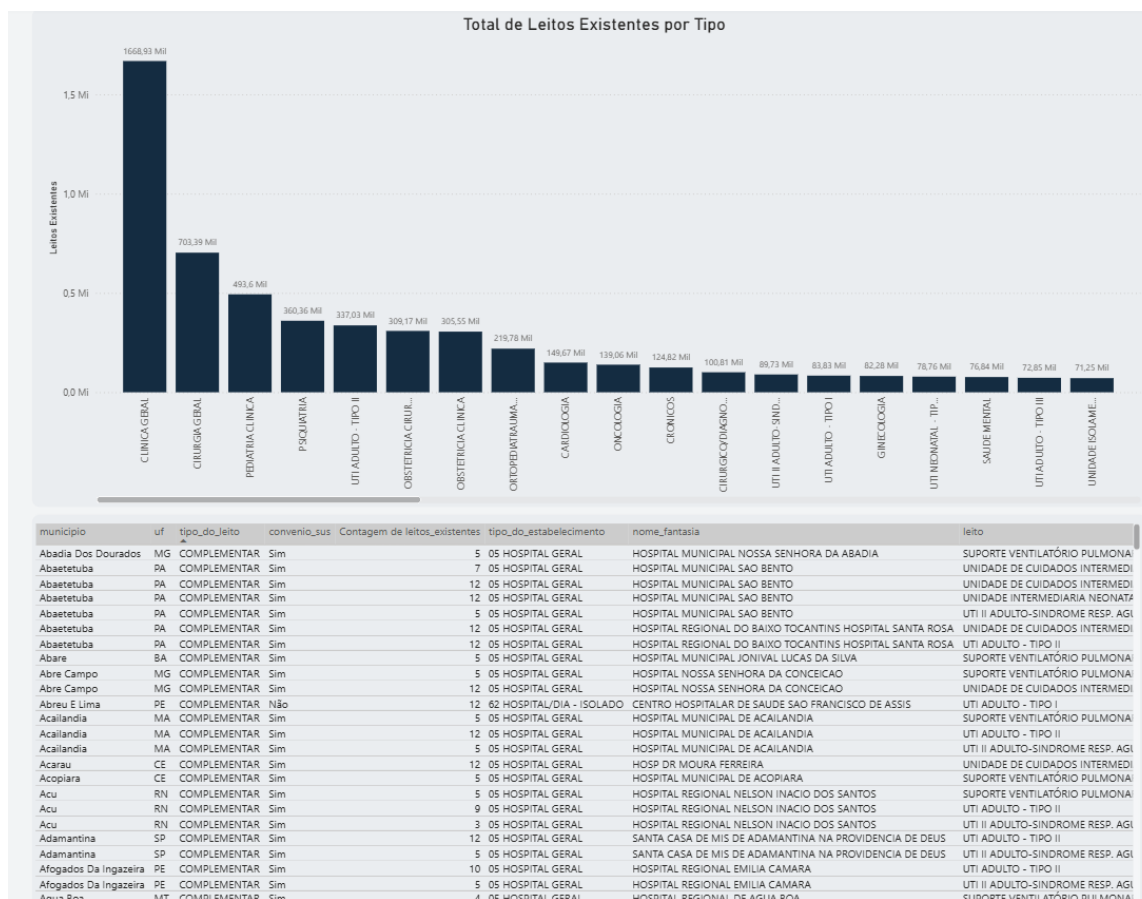


Figura 25 – Dashboard de Leitos 2

Fonte: Elaborado pelo autor

fantasia e descrição específica do leito.

Exemplos incluem registros como leitos complementares de suporte ventilatório pulmonar em hospitais municipais de Abadia dos Dourados (MG), unidades de cuidados intermediários em hospitais de Abaetetuba (PA), e diversos tipos de UTI Adulto em diferentes unidades hospitalares. Esta tabela permite auditoria granular dos dados agregados e investigação de casos específicos, incluindo análise de capacidade por estabelecimento individual e especialidade ofertada.

### 5.0.4 Dashboard de Profissionais

O módulo “Profissionais” apresenta análise detalhada dos recursos humanos em saúde cadastrados no CNES, oferecendo visualizações sobre a força de trabalho atuante nos estabelecimentos de saúde brasileiros. Este *dashboard* permite compreender a distribuição geográfica, especialização profissional, carga horária e vínculo com o SUS dos profissionais de saúde no país.



#### 5.0.4.1 Sistema de Filtragem

O *dashboard* de Profissionais implementa sistema de filtragem com oito dimensões principais: Ano, Estado, Município, Convênio SUS, Tipo Equipe, Atendimento Prestado, Cargo e Nível de Atenção, conforme ilustrado na Figura 27. Esta estrutura permite análises segmentadas por categorias profissionais, modalidades de atendimento e características de vínculo empregatício, possibilitando investigações específicas sobre a composição da força de trabalho em diferentes contextos.



Figura 27 – Sistema de filtros do *dashboard* de Profissionais

Fonte: Elaborado pelo autor

A presença do filtro “Tipo Equipe” permite análises específicas sobre equipes de atenção básica, saúde da família e outras configurações organizacionais, enquanto o filtro “Cargo” possibilita investigações direcionadas a categorias profissionais específicas como médicos, enfermeiros, técnicos e demais profissões da saúde.

#### 5.0.4.2 Indicadores Principais e Visuais

O *dashboard* apresenta quatro indicadores principais quantificando os recursos humanos em saúde (Figura 27). O total de profissionais cadastrados, a carga horária ambulatorial SUS, a carga horária hospitalar SUS e o indicador de profissionais que atendem pelo SUS.

A proporção de profissionais atendendo pelo SUS (aproximadamente 78% do total) evidencia significativa adesão da força de trabalho ao sistema público, superior à proporção de estabelecimentos conveniados (26,73%), indicando que profissionais atuam predominantemente em unidades vinculadas ao SUS.

O gráfico de pizza “Número de Profissionais que Atendem SUS” segmenta os profissionais por vínculo com o sistema público (Figura 27). A distribuição mostra que 1,02 milhão de profissionais (75,6%) atendem pelo SUS, enquanto 0,33 milhão (24,4%) não possuem vínculo com o sistema público. Esta proporção demonstra que aproximadamente três quartos da força de trabalho em saúde cadastrada atua no sistema público, configuração essencial para cobertura assistencial da população.

O gráfico de linha “Número de Profissionais ao Longo do Tempo” apresenta série temporal da evolução quantitativa da força de trabalho entre 2020 e 2025 (Figura 27), a visualização revela crescimento significativo no período, iniciando com aproximadamente.

O crescimento acentuado observado entre 2022 e 2023 pode refletir tanto uma expansão efetiva da força de trabalho quanto aprimoramentos nos processos de cadastramento, validação e atualização das informações no sistema CNES. A análise temporal desses dados permite a identificação de tendências estruturais de crescimento e subsidia projeções mais consistentes sobre futuras necessidades de recursos humanos. A ampliação do período analisado contribui para uma compreensão mais robusta da dinâmica da força de trabalho no país, reduzindo efeitos pontuais e distorções associadas a mudanças operacionais no sistema.

O gráfico de barras verticais “Contagem de Profissionais que Atendem pelo SUS e que Não Atendem por Estado” apresenta análise comparativa da força de trabalho entre unidades federativas (Figura 28). São Paulo lidera com 262,22 mil profissionais totais, dos quais 120,71 mil atendem pelo SUS. Minas Gerais apresenta 139,65 mil profissionais totais com 58,94 mil vinculados ao SUS..

A visualização utiliza código de cores diferenciando profissionais que não atendem SUS (verde escuro) daqueles vinculados ao sistema público (verde), permitindo análise simultânea de volume total e taxa de vinculação ao SUS por estado. Observa-se concentração significativa nas regiões Sudeste e Sul, com proporções variadas de vinculação ao SUS entre diferentes estados.

#### 5.0.4.3 Composição de Categoria Profissional

O gráfico horizontal de barras “Profissionais por Cargo” detalha a composição da força de trabalho segundo categorias profissionais (Figura 28). Técnicos de Enfermagem lideram com 206,75 mil profissionais, seguidos por Enfermeiros com 95,62 mil, Médicos Clínicos com 83,11 mil e etc.

A visualização apresenta mais de vinte categorias profissionais e esta segmentação permite compreender a composição multidisciplinar das equipes de saúde e identificar possíveis desproporções entre categorias profissionais.

O gráfico horizontal “Profissionais por Atendimento Prestado” categoriza a força de

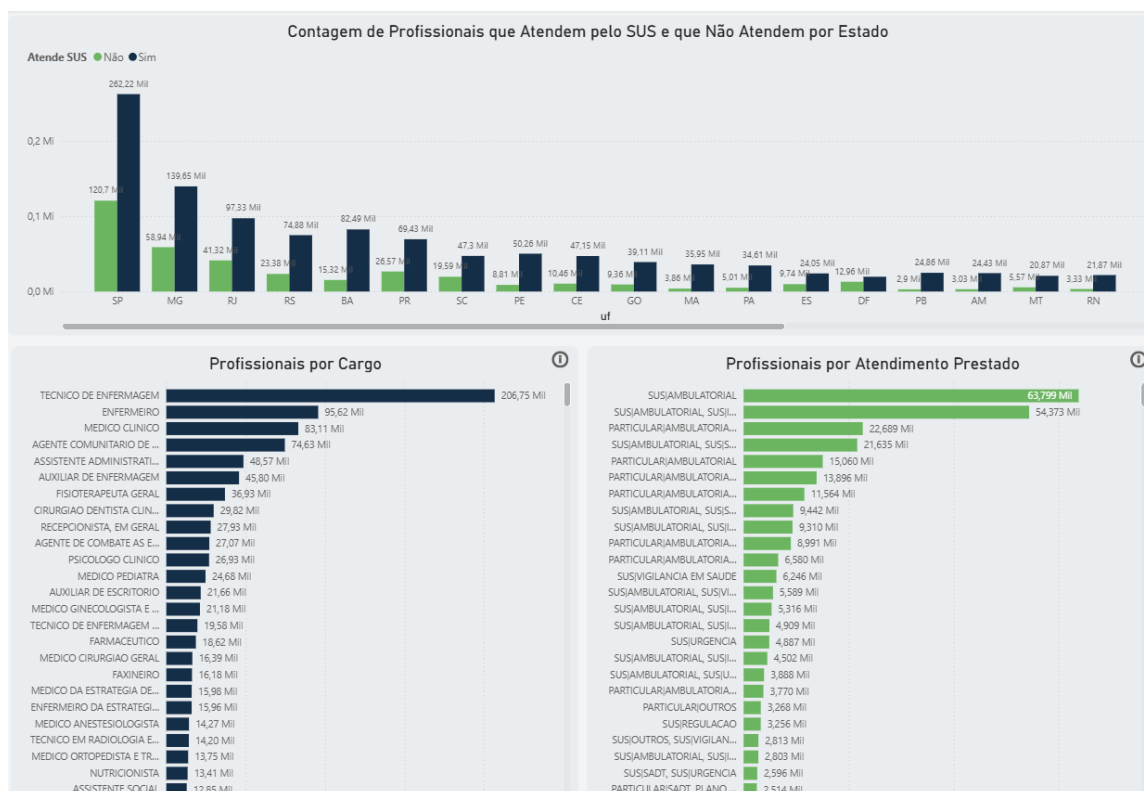


Figura 28 – Dashboard de Profissionais

Fonte: Elaborado pelo autor

trabalho segundo modalidades de prestação de serviço (Figura 28). A visualização apresenta diversas combinações de modalidades assistenciais, incluindo atendimento ambulatorial, internação, SADT (Serviços de Apoio Diagnóstico e Terapêutico), urgência, vigilância em saúde e outras especialidades. Esta segmentação permite identificar a distribuição da força de trabalho entre diferentes níveis de complexidade assistencial e modalidades de prestação de serviço.

#### 5.0.4.4 Análise de Carga Horária por Estado

O gráfico de barras agrupadas “Soma de Carga Horária Ambulatorial SUS, Carga Hospitalar SUS e Outras Cargas Horárias por Estado” apresenta distribuição da carga horária total de trabalho (Figura 29). São Paulo lidera com 4.979,83 mil horas em atendimento ambulatorial SUS, 2.730,66 mil horas hospitalares e 2.721 mil horas em outras modalidades. Minas Gerais apresenta 2.881,28 mil horas ambulatoriais, 1.486,39 mil hospitalares e 1.819,12 mil em outras categorias.

A visualização utiliza três cores distintas para diferenciar carga horária ambulatorial SUS (verde), hospitalar SUS (azul escuro) e outras cargas horárias (verde claro), permitindo análise comparativa entre estados quanto ao volume total de horas trabalhadas e sua distribuição entre modalidades assistenciais. Estados como Rio de Janeiro, Bahia, Rio Grande do Sul, Paraná, Santa Catarina, Pernambuco, Ceará, Goiás, Maranhão, Pará e

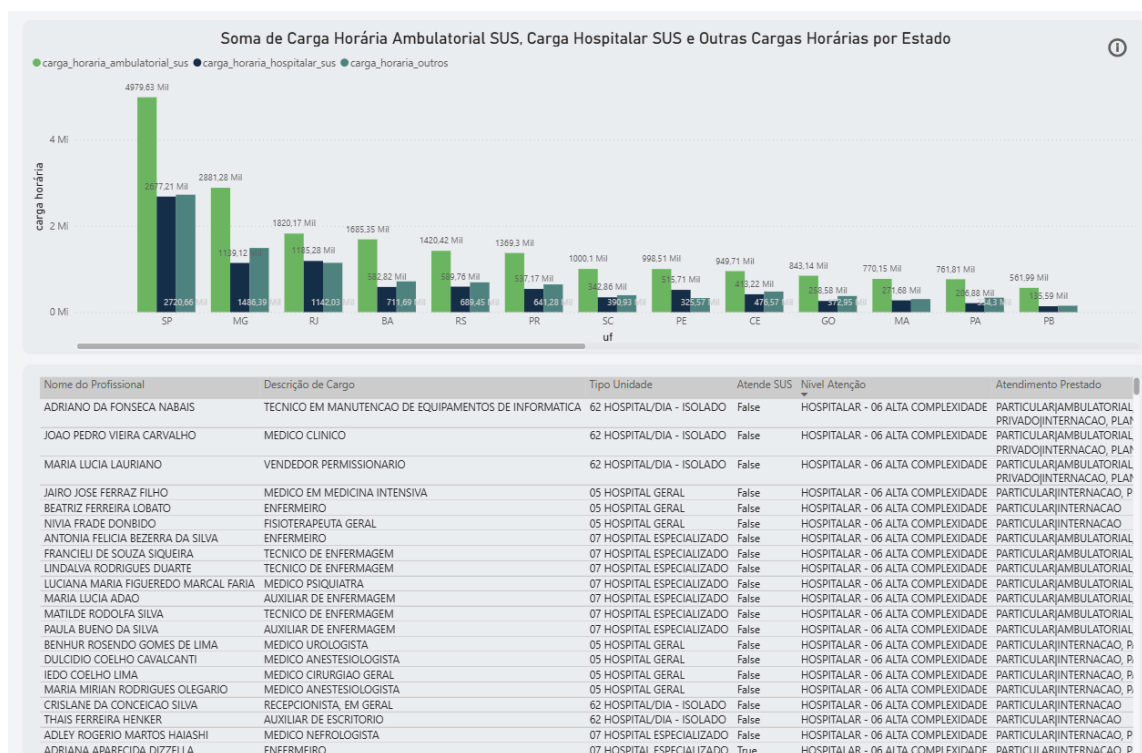


Figura 29 – Dashboard de Profissionais 2

Fonte: Elaborado pelo autor

Paraíba também apresentam volumes significativos.

A seção inferior do *dashboard* apresenta tabela detalhada com informações individualizadas dos profissionais cadastrados (Figura 29). Cada registro exibe nome do profissional, descrição do cargo, tipo de unidade, informação sobre atendimento SUS, nível de atenção e atendimento prestado. Esta tabela permite auditoria granular dos dados agregados e investigação de casos específicos de profissionais e suas características de vínculo.

### 5.0.5 Área Informativa

O sistema desenvolvido inclui uma seção informativa dedicada a contextualizar o Cadastro Nacional de Estabelecimentos de Saúde (CNES) e seus objetivos, conforme apresentado na Figura 30. Esta área fornece informações institucionais sobre a fonte de dados utilizada no sistema.

O conteúdo apresentado explica que o CNES é o sistema oficial de registro dos estabelecimentos de saúde no Brasil, mantido pelo Ministério da Saúde através do DATA-SUS. Criado em 2003, o CNES documenta as características dos estabelecimentos, seus recursos físicos, recursos humanos e equipamentos disponíveis. A área informativa destaca que estes dados são fundamentais para o planejamento e gestão do SUS, permitindo o monitoramento da infraestrutura de saúde, a distribuição de recursos e a análise da capacidade instalada em todo território nacional.

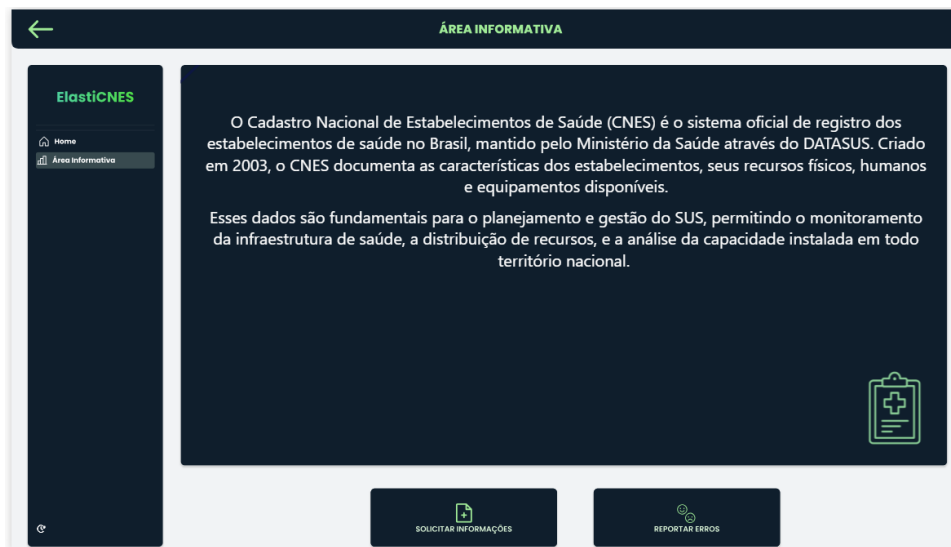


Figura 30 – Tela da Área Informativa do sistema

Fonte: Elaborado pelo autor

A tela inclui dois botões funcionais: “Solicitar Informações” e “Reportar Erros”, disponibilizando canais de comunicação para usuários solicitarem dados adicionais ou reportarem inconsistências identificadas no sistema. Esta funcionalidade demonstra preocupação com melhoria contínua e engajamento dos usuários no aperfeiçoamento da plataforma.

### 5.0.6 Painéis em Construção

Dada a abrangência dos dados disponíveis no CNES e as limitações de tempo e escopo deste trabalho de conclusão de curso, alguns módulos do sistema foram projetados mas não implementados na versão atual. Para estes módulos, foi desenvolvida uma tela de status “Painel em Construção”, ilustrada na Figura 31.

Esta tela é exibida quando o usuário acessa módulos do menu inicial que não foram desenvolvidos no escopo atual do projeto, incluindo: Marcações, Equipamentos, Serviços Especializados e Mapa CNES. A interface de status mantém o padrão visual do sistema, apresentando mensagem clara “PAINEL EM CONSTRUÇÃO” acompanhada de ícone representativo de desenvolvimento em andamento.

A implementação desta tela de status demonstra planejamento de expansão futura do sistema e transparência com os usuários sobre funcionalidades disponíveis e em desenvolvimento. Os módulos não implementados representam oportunidades de continuidade do trabalho em versões futuras, expandindo a cobertura analítica do sistema para outras dimensões importantes dos dados do CNES.



Figura 31 – Tela de status para painéis não implementados

Fonte: Elaborado pelo autor

## 5.1 Estudos de Caso e Aplicações Práticas

Esta seção apresenta estudos de caso que demonstram a aplicabilidade prática do sistema desenvolvido para gestão em saúde pública. São apresentadas questões gerenciais que os *dashboards* implementados podem responder, evidenciando como a transformação de dados brutos do CNES em visualizações analíticas bem construídas contribui para tomada de decisão baseada em evidências.

### 5.1.1 Caso 1: Planejamento de Expansão da Rede SUS

#### 5.1.1.1 Contexto

Um gestor estadual de saúde necessita planejar a expansão da rede de estabelecimentos conveniados ao SUS em seu estado, priorizando municípios com menor cobertura assistencial. A decisão deve considerar tanto a disponibilidade de estabelecimentos quanto a capacidade instalada de leitos.

#### 5.1.1.2 Perguntas Gerenciais

O *dashboard* Panorama Geral permite responder questões como:

- Qual a proporção atual de estabelecimentos conveniados ao SUS no estado?
- Quais municípios apresentam menor taxa de conveniamento SUS?
- Qual a distribuição de estabelecimentos por tipo (UBS, hospitais, clínicas) no estado?
- Quais especialidades apresentam menor cobertura na rede conveniada?

O *dashboard* de Leitos complementa a análise respondendo:

- Qual a capacidade instalada de leitos SUS por município?
- Quais especialidades de leitos estão deficitárias no estado?
- Como a distribuição de leitos cirúrgicos, clínicos e obstétricos varia regionalmente?
- Qual a proporção de leitos SUS em relação ao total disponível?

### 5.1.1.3 Análise e Decisão

Utilizando os filtros de Estado e Município no *dashboard* Panorama Geral, o gestor pode navegar hierarquicamente pelos dados, partindo de uma visão estadual até chegar ao nível municipal — funcionalidade que o ElasticCNES não oferece de forma integrada. Como exemplo prático, ao filtrar pelo estado de Minas Gerais, o sistema revela que dos 63.277 estabelecimentos cadastrados no CNES, apenas 16,6 mil (26,21%) são conveniados ao SUS — evidenciando que mais de 74% da infraestrutura de saúde instalada no estado opera exclusivamente no setor privado.

A segmentação por natureza jurídica complementa a análise: dos estabelecimentos mineiros, 76,14% são privados, 21,24% públicos e apenas 2,62% sem fins lucrativos — proporção que pode orientar estratégias de conveniamento com o setor privado como alternativa à expansão da rede própria. A distribuição por atividade principal indica que a consulta ambulatorial é a modalidade dominante (461,4 mil registros), seguida de atenção básica (75,9 mil) e apoio diagnóstico (59,7 mil), revelando oportunidades para expansão direcionada a especialidades com menor cobertura conveniada.

O sistema permite ainda refinar a análise ao nível municipal, possibilitando identificar disparidades intraestaduais: municípios do interior tendem a apresentar percentuais de convênio SUS significativamente distintos da média estadual de 26,21%, evidenciando quais localidades devem ser priorizadas em políticas de expansão da rede pública. Esta navegação hierárquica — do estado ao município em poucos cliques — seria impraticável nas ferramentas governamentais existentes sem extração e cruzamento manual de dados.

Esta análise multidimensional permite ao gestor:

1. Identificar municípios ou regiões com menor cobertura de estabelecimentos conveniados ao SUS, priorizando-os para expansão da rede pública;
2. Mapear especialidades de leitos que necessitam de expansão ou fortalecimento;
3. Planejar estratégias de distribuição regional mais equilibrada dos recursos assistenciais;

4. Estabelecer prioridades de expansão baseadas em dados concretos sobre infraestrutura existente.

## 5.1.2 Caso 2: Dimensionamento de Recursos Humanos

### 5.1.2.1 Contexto

Uma Secretaria Municipal de Saúde precisa dimensionar a contratação de profissionais para novas unidades de atenção primária, considerando a composição atual das equipes e as necessidades de especialidades específicas.

### 5.1.2.2 Perguntas Gerenciais

O *dashboard* de Profissionais permite responder:

- Qual a composição atual de profissionais por categoria no município?
- Qual a proporção entre médicos, enfermeiros e técnicos de enfermagem?
- Quantos profissionais estão vinculados à atenção primária versus atenção especializada?
- Qual a carga horária média por profissional em atendimento ambulatorial SUS?
- Como a composição de equipes no município se compara com municípios similares?

Complementarmente, o sistema permite investigar:

- Qual a evolução temporal do número de profissionais nos últimos anos?
- Quais especialidades médicas apresentam menor disponibilidade?
- Há adequação entre número de profissionais e capacidade instalada de leitos?

### 5.1.2.3 Análise e Decisão

Aplicando o filtro por município no *dashboard* de Profissionais, o gestor pode identificar o total de profissionais cadastrados e segmentar por vínculo SUS. A análise por categoria profissional permite visualizar a composição atual da força de trabalho, incluindo técnicos de enfermagem, enfermeiros, médicos, agentes comunitários de saúde, profissionais administrativos e outras categorias. A comparação com outras localidades possibilita identificar possíveis déficits ou superávits em categorias específicas, como fisioterapeutas, psicólogos, médicos especialistas, entre outros. Esta comparação regional oferece referências para planejamento de recursos humanos. A análise de carga horária contratada por tipo de atendimento (ambulatorial, hospitalar, SADT) permite estimar a capacidade de oferta de

serviços. A visualização da evolução temporal do quadro de profissionais indica tendências de crescimento, redução ou estabilidade ao longo dos anos disponíveis na base de dados.

Com base nas análises disponíveis no sistema, o gestor pode:

1. Identificar categorias profissionais que necessitam de reforço ou expansão;
2. Planejar a composição de equipes multiprofissionais para novas unidades;
3. Estabelecer proporções adequadas entre diferentes categorias profissionais;
4. Estimar necessidades de carga horária com base na infraestrutura planejada;
5. Fundamentar decisões sobre estratégias de contratação ou conveniamento de serviços.

### 5.1.3 Caso 3: Análise de Capacidade Hospitalar para Emergências

#### 5.1.3.1 Contexto

Durante período de alta demanda sazonal (surtos epidêmicos, acidentes em massa), gestores regionais necessitam avaliar rapidamente a capacidade instalada de leitos hospitalares e identificar estabelecimentos com potencial para expansão temporária de leitos.

#### 5.1.3.2 Perguntas Gerenciais

O *dashboard* de Leitos permite responder com agilidade:

- Quantos leitos de UTI adulto estão disponíveis na região?
- Qual a capacidade total de leitos clínicos que podem ser convertidos para isolamento?
- Quais hospitais possuem maior disponibilidade de leitos complementares?
- Quantos leitos operam em regime 24 horas?
- Qual a distribuição geográfica de leitos para identificar áreas com menor cobertura?

Integrando com o *dashboard* de Profissionais:

- Há profissionais especializados em medicina intensiva suficientes?
- Qual a carga horária disponível de enfermeiros e técnicos de enfermagem?
- Quais estabelecimentos concentram especialistas em doenças infecciosas?

### 5.1.3.3 Análise e Decisão

Em cenário de surto epidêmico, o gestor regional acessa o *dashboard* de Leitos e aplica filtros por Estado e tipo de leito. A análise permite identificar a disponibilidade total de leitos de UTI Adulto na região, sua distribuição entre os estabelecimentos e o padrão de concentração geográfica. A segmentação por especialidade permite avaliar leitos clínicos que potencialmente podem ser adaptados para isolamento, além de leitos complementares disponíveis em outras especialidades. A visualização geográfica possibilita identificar municípios com menor disponibilidade de recursos hospitalares e distâncias aos centros de referência mais próximos. Complementando com análise de profissionais no *dashboard* correspondente, é possível mapear a distribuição de médicos intensivistas, enfermeiros e outros profissionais críticos na região. A análise de carga horária contratada permite avaliar a capacidade atual e potencial de expansão mediante reorganização de escalas ou turnos.

Esta análise integrada permite ao gestor:

1. Mapear a capacidade instalada total de leitos críticos e sua distribuição regional;
2. Identificar estabelecimentos com potencial para instalação de leitos temporários de UTI;
3. Planejar estratégias de transporte inter-hospitalar para municípios com menor capacidade;
4. Dimensionar necessidades de reforço no quadro de profissionais especializados;
5. Organizar estratégias de redistribuição regional de recursos humanos conforme demanda;
6. Fundamentar protocolos de ativação gradual da capacidade instalada conforme progressão da emergência.

## 5.1.4 Caso 4: Avaliação de Inequidades Regionais

### 5.1.4.1 Contexto

Pesquisadores em saúde coletiva investigam desigualdades na distribuição de recursos de saúde entre regiões do país, buscando evidências para formulação de políticas de redução de inequidades.

### 5.1.4.2 Perguntas Gerenciais

Os três *dashboards* integrados permitem análises comparativas:

- Como a razão estabelecimentos/habitantes varia entre estados?
- Há correlação entre PIB per capita estadual e densidade de leitos SUS?
- Regiões Norte e Nordeste apresentam déficits específicos em quais especialidades?
- Qual a variação na proporção de estabelecimentos privados entre regiões?
- Estados com menor disponibilidade de leitos compensam com maior atenção ambulatorial?
- A distribuição de profissionais especializados reflete a concentração de estabelecimentos?

#### 5.1.4.3 Análise e Decisão

Utilizando visualizações comparativas entre estados nos três *dashboards*, pesquisadores podem identificar padrões de distribuição regional de recursos de saúde. A comparação entre regiões permite observar a concentração de estabelecimentos, leitos e profissionais em determinadas áreas geográficas em relação à distribuição populacional. É possível verificar se há proporcionalidade entre infraestrutura disponível e tamanho da população atendida em diferentes regiões do país.

A análise por especialidade de leitos permite identificar possíveis disparidades regionais em recursos críticos como UTI Neonatal, oncologia, cardiologia e neurologia. A segmentação geográfica revela padrões de concentração ou dispersão destes recursos especializados. A investigação do quadro de profissionais possibilita mapear a distribuição de especialistas médicos, identificando regiões com maior dependência de médicos generalistas. A comparação entre categorias profissionais em diferentes estados oferece panorama sobre a composição da força de trabalho em saúde. A análise por natureza jurídica dos estabelecimentos permite verificar a proporção entre serviços públicos, privados e privados conveniados ao SUS em cada região. Esta segmentação indica diferenças no modelo de organização dos sistemas estaduais de saúde.

As análises comparativas fundamentam:

1. Identificação de regiões prioritárias para políticas de fixação de profissionais;
2. Diagnóstico de necessidades de investimento em infraestrutura hospitalar por região;
3. Mapeamento de áreas que podem se beneficiar de estratégias como telemedicina;
4. Fundamentação para revisão de critérios de distribuição de recursos federais;
5. Identificação de regiões com potencial para ampliação do convênio SUS;
6. Planejamento de centros regionais de referência em especialidades específicas.

### 5.1.5 Importância dos Dashboards para Gestão em Saúde

Os estudos de caso apresentados demonstram que a transformação de dados brutos do CNES em *dashboards* analíticos interativos gera valor substancial para gestão em saúde pública em múltiplas dimensões:

- **Agilidade na Tomada de Decisão:** Antes da implementação de sistemas analíticos como o desenvolvido neste trabalho, gestores precisavam solicitar relatórios customizados a equipes técnicas, processo que demandava dias ou semanas. Os *dashboards* interativos permitem análises *ad hoc* em tempo real, possibilitando respostas imediatas a questões gerenciais urgentes, como dimensionamento de capacidade para emergências ou identificação rápida de lacunas assistenciais.
- **Visão Multidimensional Integrada:** A possibilidade de aplicar filtros combinados (Estado + Município + Especialidade + Convênio SUS) permite análises complexas que seriam impraticáveis em planilhas estáticas. A integração entre informações de estabelecimentos, leitos e profissionais em interface única facilita compreensão holística da infraestrutura de saúde, evidenciando interdependências entre diferentes recursos.
- **Democratização do Acesso à Informação:** A implementação de dashboards interativos em Power BI promove **self-service analytics**, removendo barreiras técnicas no acesso aos dados CNES. Tradicionalmente, consultas exigiam conhecimento em SQL ou solicitações formais a TI. O sistema desenvolvido permite que gestores de saúde e formuladores de políticas públicas extraiam insights autonomamente através de filtros visuais e *drill-down* intuitivos. Visualizações gráficas intuitivas tornam dados complexos acessíveis a gestores sem formação técnica em estatística ou informática, ampliando o público capaz de utilizar dados CNES para fundamentar decisões.
- **Padronização de Análises:** *Dashboards* institucionais garantem que diferentes gestores analisem os mesmos indicadores calculados de forma consistente, evitando discrepâncias em números reportados por diferentes setores. Definições claras de métricas (ex: proporção de leitos SUS, carga horária ambulatorial) asseguram comparabilidade temporal e entre regiões.
- **Identificação de Padrões e Tendências:** Visualizações temporais (como a evolução de profissionais entre 2020-2025) permitem identificar tendências de crescimento ou redução de recursos, possibilitando projeções e planejamento proativo. Comparações geográficas revelam padrões de concentração ou dispersão de recursos que podem passar despercebidos em análises tabulares.

- **Subsídio para Políticas Baseadas em Evidências:** Os casos apresentados ilustram como dados transformados em informação acionável fundamentam políticas públicas: expansão da rede SUS priorizando áreas deficitárias, dimensionamento adequado de recursos humanos, preparação para emergências sanitárias e redução de inequidades regionais. Decisões fundamentadas em dados são mais robustas, defensáveis e passíveis de monitoramento de resultados.
- **Transparência e Accountability:** Embora não implementado nesta versão, sistemas como o desenvolvido podem ser disponibilizados publicamente, permitindo que cidadãos, pesquisadores e organizações da sociedade civil monitorem a infraestrutura de saúde em suas regiões. Esta transparência aumenta accountability de gestores públicos e empodera a sociedade no exercício de controle social.

### 5.1.6 Validação Comparativa de Usabilidade

A fim de quantificar as diferenças de usabilidade entre o sistema desenvolvido e o ElastiCNES, foram conduzidos testes cronometrados de execução de tarefas gerenciais representativas. Cada tarefa consistia em localizar uma informação específica partindo da tela inicial de cada sistema, sem conhecimento prévio do resultado esperado. Os tempos registrados referem-se ao intervalo entre o início da navegação e a obtenção do valor consolidado na tela.

O Quadro 12 apresenta os resultados comparativos para cinco tarefas representativas de uso gerencial:

Os resultados evidenciam três categorias de limitação do ElastiCNES: (1) **falha funcional**, observada na Tarefa 2, em que o sistema travou indefinidamente após aplicação de filtro, inviabilizando qualquer análise histórica; (2) **ausência de agregação visual**, observada na Tarefa 3, em que o sistema retornou 2,9 milhões de registros tabulares sem consolidação, exigindo contagem manual e tornando a tarefa impraticável para qualquer pessoa; e (3) **inconsistência temporal**, observada nas Tarefas 4 e 5, em que o período de referência travado gerou dificuldade na realização das consultas, exigindo várias tentativas mal sucedidas.

Em média, o sistema proposto foi **2,4 vezes mais rápido** na execução das tarefas avaliadas, com 100% de conclusão bem-sucedida contra apenas 60% no ElastiCNES (3 tarefas concluídas com ressalvas e 1 não concluída). Além da diferença temporal, o sistema desenvolvido apresentou resultados consolidados em visualizações interativas em todas as tarefas, enquanto o ElastiCNES recorreu a tabelas brutas sem agregação em ao menos um caso, evidenciando limitações estruturais que vão além da performance técnica.

Tabela 12 – Comparativo de tempo de execução de tarefas entre o sistema proposto e o ElastiCNES

Tarefa	Sistema proposto	ElastiCNES	Observações
Leitos SUS no estado do Pará	15s	18s	ElastiCNES com período travado em 2026/01, não possibilitou alteração
Evolução histórica de leitos SUS no Pará (2020–2024)	14s	>1min25s	ElastiCNES travou após aplicação do filtro de estado; tarefa não concluída
Total de médicos clínicos cadastrados em SP	34s	1min57s	ElastiCNES sem filtro de ano e sem visualização agregada; para alcançar qualquer resultado seria necessário contar manualmente os registros de uma tabela informativa
Comparação de leitos de UTI entre PA e SP	19s	29s	Divergência de valores atribuída a diferenças no período de referência
Estabelecimentos conveniados SUS em Belém (PA)	16s	19s	Divergência de valores atribuída a diferenças no período de referência
<b>Média</b>	<b>19,6s</b>	<b>≈47,4s</b>	Sistema proposto 2,4× mais rápido em média

Fonte: Elaborado pelo autor

## 6 Conclusão

Este trabalho apresentou o desenvolvimento de um sistema de Business Intelligence aplicado aos dados do Cadastro Nacional de Estabelecimentos de Saúde (CNES), com objetivo de transformar dados brutos governamentais em informações analíticas acessíveis para suporte à tomada de decisão em gestão de saúde pública. A motivação central residiu na necessidade de democratizar o acesso a informações sobre infraestrutura de saúde brasileira, disponibilizando visualizações interativas que facilitem análises multidimensionais por gestores, pesquisadores, profissionais da área e pessoas no dia a dia.

O objetivo geral deste trabalho consistiu em desenvolver solução de Business Intelligence capaz de extrair, transformar, modelar e visualizar dados do CNES, permitindo análises sobre estabelecimentos de saúde, capacidade instalada de leitos e recursos humanos disponíveis no sistema de saúde brasileiro. Para alcançar este objetivo, foram estabelecidas cinco metas específicas: desenvolver processo de extração automatizada de dados mediante técnicas de *web scraping*, implementar *pipeline* ETL completo, modelar dados segundo arquitetura dimensional otimizada, desenvolver *dashboards* interativos e demonstrar aplicabilidade prática através de estudos de caso. Todos os objetivos específicos foram alcançados, resultando em sistema funcional capaz de processar dados do CNES e disponibilizá-los através de interface analítica intuitiva.

Na dimensão prática, o sistema desenvolvido oferece capacidades analíticas imediatamente úteis para gestão em saúde pública. A possibilidade de filtrar dados por múltiplas dimensões (temporal, geográfica, tipo de estabelecimento, especialidade) e visualizar indicadores consolidados permite análises que anteriormente demandavam conhecimentos técnicos especializados ou processamento manual demorado. Os estudos de caso apresentados demonstram que o sistema pode subsidiar decisões gerenciais concretas como planejamento de expansão da rede SUS, dimensionamento de recursos humanos, preparação para emergências sanitárias e identificação de inequidades regionais. A transformação de dados brutos em visualizações interativas democratiza acesso à informação, contribuindo para gestão mais transparente e fundamentada em evidências.

### 6.0.1 Perspectivas Futuras e Considerações Finais

As limitações identificadas apontam direções promissoras para continuidade deste trabalho. A implementação dos módulos planejados mas não desenvolvidos ampliaria significativamente a cobertura analítica do sistema, particularmente em relação à análise de equipamentos médicos, essencial para o planejamento de ações de média e alta complexidade. A integração com dados populacionais do IBGE, produção assistencial

(SIA/SIH), morbimortalidade (SIM/SINAN) e orçamento (SIOPS) permitiria análises mais abrangentes, correlacionando capacidade instalada com utilização efetiva, desfechos em saúde e investimentos realizados.

O desenvolvimento de modelos preditivos para projeção de demanda, algoritmos de otimização para alocação de recursos e análises espaciais avançadas de acessibilidade geográfica representam evoluções naturais do sistema, especialmente relevantes para dissertações de mestrado ou projetos de pesquisa e desenvolvimento. A adaptação do sistema para disponibilização pública via web democratizaria ainda mais o acesso à informação sobre infraestrutura de saúde, permitindo que gestores municipais, conselheiros de saúde e cidadãos monitorem os recursos em suas regiões.

Atualmente, o processo de extração de dados requer execução manual do *script* de *web scraping*, não havendo automação para atualização periódica. A implementação de orquestração automatizada do *pipeline* ETL utilizando ferramentas como Apache Airflow, com monitoramento de qualidade de dados e alertas automáticos para gestores quando indicadores críticos ultrapassarem limiares estabelecidos, aumentaria significativamente a maturidade operacional do sistema. O desenvolvimento deste trabalho evidenciou o potencial transformador de tecnologias de Business Intelligence quando aplicadas a dados governamentais de saúde.

A extração via API não documentada captura campos de identificação pessoal (CPF, nome) não visíveis na interface pública do ElastiCNES, exigindo atenção à conformidade com a LGPD em implementações futuras, especialmente quanto à anonimização na camada de ingestão.

O Sistema Único de Saúde brasileiro, reconhecido internacionalmente como uma das maiores políticas públicas de saúde universal do mundo, enfrenta desafios permanentes de financiamento, distribuição desigual de recursos e pressão crescente por serviços de qualidade. Ferramentas analíticas como a desenvolvida neste trabalho, embora não resolvam por si só estes desafios estruturais, contribuem para que decisões sejam tomadas com base em dados concretos sobre onde estão os recursos, onde faltam e como podem ser melhor aproveitados.

Gestores munidos de informações precisas e atualizadas podem alocar recursos de forma mais eficiente, priorizar investimentos em áreas deficitárias e monitorar resultados de intervenções implementadas. A experiência de desenvolver este sistema consolidou aprendizados técnicos em engenharia de dados, modelagem dimensional e visualização de informações, mas também proporcionou compreensão mais profunda sobre a complexidade do sistema de saúde brasileiro e a importância de dados de qualidade para a gestão pública. A interseção entre tecnologia e saúde pública mostra-se um campo fértil para contribuições que agregam valor tanto do ponto de vista técnico quanto social.

## Referências

- ARMBRUST, M. et al. Lakehouse: a new generation of open platforms that unify data warehousing and advanced analytics. In: *Proceedings of CIDR*. [S.l.: s.n.], 2021. v. 8, p. 1–13. Citado na página 36.
- Brasil. *Lei de Acesso à Informação, Lei nº 12.527*. 2011. Brasília: Presidência da República. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2011-2014/2011/lei/l12527.htm](http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm). Citado na página 15.
- Brasil. *Lei nº 12.527, de 18 de novembro de 2011*. Brasília: [s.n.], 2011. Lei de Acesso a Informação. Citado na página 20.
- Brasil. *Sistemas de Informação em Saúde e a Gestão do SUS*. Brasília: Ministério da Saúde, 2011. Citado 2 vezes nas páginas 20 e 23.
- Brasil. *Dados Abertos Governamentais*. Brasília: Controladoria-Geral da União, 2014. Citado na página 20.
- Brasil. *Decreto nº 8.777, de 11 de maio de 2016*. Brasília: [s.n.], 2016. Política Nacional de Dados Abertos. Citado na página 20.
- Brasil. *Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD)*. Brasília, DF: Diário Oficial da União, 2018. Disponível em: [https://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm). Acesso em: mar. 2025. Citado na página 52.
- Brasil. *Cadastro Nacional de Estabelecimentos de Saúde (CNES)*. Brasília: [s.n.], 2023. Ministério da Saúde. Citado na página 20.
- Brasil. *DATASUS Departamento de Informatica do Sistema Unico de Saude*. Brasília: [s.n.], 2023. Ministério da Saúde. Citado na página 23.
- CARVALHO, A. O.; SANTOS, L. Sistemas de informação em saúde: desafios de integração e usabilidade. In: *Anais do Congresso Brasileiro de Informática em Saúde*. [S.l.: s.n.], 2017. p. 234–239. Citado na página 23.
- CASTRO, M. C. et al. Sistema de saúde no brasil: avanços e desafios. *The Lancet*, v. 377, n. 9781, p. 11–31, 2011. Citado na página 15.
- Centers for Disease Control and Prevention. *CDC Wonder: Data and Statistics*. 2020. Disponível em: <https://wonder.cdc.gov/>. Acesso em: 18 jan. 2026. Citado na página 23.
- CLEVELAND, W. S.; MCGILL, R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 1984. Citado na página 19.
- DATASUS. *ElastiCNES: Plataforma Analítica do Cadastro Nacional de Estabelecimentos de Saúde*. 2024. Acesso em: 25 maio 2025. Disponível em: <https://elastin.datasus.gov.br/>. Citado 2 vezes nas páginas 23 e 25.

- DAVENPORT, T. H.; HARRIS, J. G. *Competing on analytics: the new science of winning*. [S.l.]: Harvard Business Press, 2007. Citado 2 vezes nas páginas 15 e 16.
- DAVENPORT, T. H.; PRUSAK, L. *Conhecimento empresarial: como as organizações gerenciam o seu capital intelectual*. Rio de Janeiro: Campus, 1998. Citado na página 20.
- DINIZ, E. H. Dados abertos e governo eletrônico. *Revista de Administração Pública*, 2010. Citado na página 20.
- ECKERSON, W. *Performance Dashboards*. [S.l.]: John Wiley & Sons, 2010. Citado na página 19.
- FeliciLab. *Realizar Estudo de Integração com Base de dados CNES*. 2020. Discussão técnica sobre limitações de integração com API do CNES. Disponível em: <<https://github.com/FeliciLab/mapadasaude/issues/433>>. Citado na página 32.
- FERRARI, A.; RUSSO, M. *Analyzing Data with Power BI and Power Pivot for Excel*. Redmond, WA: Microsoft Press, 2020. ISBN 978-1509302765. Citado na página 22.
- FERRER-SAPENA, A.; PESET, F.; ALEIXANDRE-BENAVENT, R. Open data, transparency and public sector information. *Journal of Information Science*, 2011. Citado na página 20.
- FEW, S. *Information Dashboard Design: The Effective Visual Communication of Data*. [S.l.]: O'Reilly Media, 2006. Citado 7 vezes nas páginas 25, 28, 30, 31, 32, 33 e 57.
- FEW, S. *Information dashboard design: The effective visual communication of data*. [S.l.]: O'Reilly Media, 2006. Citado na página 32.
- FEW, S. *Information Dashboard Design*. [S.l.]: Analytics Press, 2013. Citado 2 vezes nas páginas 19 e 56.
- INMON, W. H. *Building the Data Warehouse*. 4. ed. Indianapolis, IN: John Wiley & Sons, 2005. Citado 2 vezes nas páginas 19 e 36.
- KIMBALL, R.; CASERTA, J. *The Data Warehouse ETL Toolkit*. [S.l.]: Wiley, 2004. Citado na página 19.
- KIMBALL, R.; ROSS, M. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3. ed. [S.l.]: Wiley, 2013. Citado 5 vezes nas páginas 15, 19, 20, 33 e 35.
- MACHADO, C. V.; LIMA, L. D.; BAPTISTA, T. W. F. Políticas de saúde no brasil em tempos contraditórios: caminhos e tropeços na construção de um sistema universal. *Cadernos de Saúde Pública*, v. 33, p. e00129616, 2018. Citado na página 33.
- MELLO, J. M. H. P. et al. Avaliação da completude das informações em sistemas de vigilância em saúde. *Epidemiologia e Serviços de Saúde*, v. 27, n. 3, p. e2017473, 2018. Citado 2 vezes nas páginas 23 e 32.
- MICROSOFT. *Documentação do Power BI*. 2024. Acesso em: 2025. Disponível em: <<https://learn.microsoft.com/power-bi>>. Citado 2 vezes nas páginas 20 e 22.

- Ministério da Saúde (Brasil). *Capturas de tela da plataforma ElastiCNES*. 2025. Imagens obtidas diretamente do portal ElastiCNES. Acesso em: 25 maio 2025. Disponível em: <<https://elasticnes.saude.gov.br/>>. Citado 5 vezes nas páginas 26, 27, 29, 30 e 31.
- MITCHELL, R. *Web Scraping with Python: Collecting More Data from the Modern Web*. 2. ed. Sebastopol, CA: O'Reilly Media, 2018. ISBN 978-1491985571. Citado na página 21.
- MOMJIAN, B. *PostgreSQL: Introduction and Concepts*. Boston, MA: Addison-Wesley, 2001. ISBN 978-0201703313. Citado na página 21.
- NHS Digital. *Data and Information: NHS Digital Portal*. 2021. Disponível em: <<https://digital.nhs.uk/>>. Acesso em: 18 jan. 2026. Citado na página 23.
- PAIM, J. et al. O sistema de saúde brasileiro: história, avanços e desafios. *The Lancet*, Elsevier, v. 377, n. 9779, p. 1778–1797, 2011. Citado na página 15.
- PINTO, L. F.; FREITAS, M. P.; FIGUEIREDO, A. Sistemas de informação em saúde e a qualidade dos dados. *Ciência & Saúde Coletiva*, 2018. Citado 2 vezes nas páginas 20 e 23.
- PINTO, L. F.; GIOVANELLA, L. Do programa à estratégia saúde da família: expansão do acesso e redução das internações por condições sensíveis à atenção básica. *Ciência & Saúde Coletiva*, v. 23, n. 6, p. 1903–1914, 2018. Citado na página 23.
- Python Software Foundation. *Python Language Reference, version 3.11*. 2023. Disponível em: <<https://www.python.org/>>. Acesso em: mar. 2025. Citado na página 21.
- RIBEIRO, C. M.; COSTA, A. L. O cnes como ferramenta de gestão e planejamento no sus: avanços e desafios. *Revista Brasileira de Saúde Pública*, v. 55, n. 1, p. e000123, 2021. Citado 2 vezes nas páginas 23 e 32.
- SANTOS, L. M. P. et al. Gestão da informação para tomada de decisão em saúde pública. *Saúde e Sociedade*, v. 30, n. 2, p. e200847, 2021. Citado na página 33.
- SAÚDE, B. M. da. *Cadastro Nacional de Estabelecimentos de Saúde (CNES)*. 2024. Acesso em: 2025. Disponível em: <<https://www.gov.br/saude/pt-br/composicao/saes/cnes>>. Citado na página 15.
- SCHLEGEL, K. et al. *Magic Quadrant for Analytics and Business Intelligence Platforms*. 2024. Microsoft posicionada como líder pelo 17º ano consecutivo. Disponível mediante solicitação à Microsoft. Citado na página 22.
- SILVA, A. B.; MENDES, E. V. Qualidade da informação nos sistemas de saúde: uma revisão integrativa. *Revista de Saúde Pública*, v. 53, p. 1–12, 2019. Citado 2 vezes nas páginas 23 e 32.
- SILVA, B. H.; ALMEIDA, J. L. Gestão estratégica em saúde com apoio de business intelligence. *Cadernos de Saúde Pública*, v. 36, n. 2, p. e00012320, 2020. Citado na página 19.
- TAMAKI, E. M. et al. Metodologia de construção de um painel de indicadores para o monitoramento da gestão da atenção básica. *Saúde em Debate*, v. 36, n. 94, p. 424–432, 2012. Citado na página 23.

- The PostgreSQL Global Development Group. *PostgreSQL 15 Documentation*. 2023. Disponível em: <<https://www.postgresql.org/docs/15/>>. Acesso em: mar. 2025. Citado na página 21.
- TUFTE, E. R. *The Visual Display of Quantitative Information*. [S.l.]: Graphics Press, 2001. Citado na página 19.
- UBALDI, B. Open government data: Towards empirical analysis of open government data initiatives. *OECD Working Papers on Public Governance*, OECD Publishing, n. 22, 2013. Citado na página 32.
- VIACAVAL, F.; BELLIDO, J. G. Desigualdades regionais e sociais em saúde segundo inquéritos domiciliares (Brasil, 1998-2013). *Ciência & Saúde Coletiva*, SciELO Brasil, v. 21, p. 2881-2894, 2016. Citado na página 15.
- W3C. *Web Content Accessibility Guidelines (WCAG) 2.1*. [S.l.], 2018. Acesso em: 25 maio 2025. Disponível em: <<https://www.w3.org/TR/WCAG21/>>. Citado na página 56.
- WANG, J.; SMITH, A.; PATEL, R. Design principles for effective public health dashboards: A review of best practices and user needs. *Online Journal of Public Health Informatics*, v. 16, n. 1, p. e58277, 2024. Acesso em: 25 maio 2025. Disponível em: <<https://ojphi.jmir.org/2024/1/e58277>>. Citado 6 vezes nas páginas 25, 26, 28, 30, 32 e 33.