

# Análise Dinâmica de Dados Hospitalares: Uma Abordagem Interativa para Gestão em Saúde

Vinícius Pedroso, Artur Kniest, Giovanna Castro, Isabel H. Manssour

<sup>1</sup>Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), Escola Politécnica  
Av. Ipiranga, 6681 – CEP: 90619-900 – Porto Alegre – RS – Brazil

{vinicius.pedroso002, artur.rihl, giovanna.cruz23}@edu.pucrs.br,  
isabel.manssour@pucrs.br

**Abstract.** *The growing volume of data from DATASUS, including hospital admission records, presents challenges such as a lack of standardization and incompleteness. Visual exploratory analysis emerges as an effective tool to facilitate its interpretation and assist in decision-making. This work aims to apply interactive visualization and data exploration techniques to analyze hospital occupancy in Rio Grande do Sul, providing an online dashboard that allows queries about hospitalizations in the region. The tool enables the analysis of, for instance, the most common procedures and the distribution of hospitalizations over a year, providing information for a more efficient allocation of hospital resources.*

**Resumo.** *O crescente volume de dados do DATASUS, incluindo registros de internação hospitalar, apresenta desafios como falta de padronização e incompletudes. A análise exploratória visual surge como uma ferramenta eficaz para facilitar a sua interpretação e auxiliar na tomada de decisões. Este trabalho tem como objetivo aplicar técnicas de visualização interativa e exploração de dados para analisar a ocupação de hospitais no Rio Grande do Sul, disponibilizando um dashboard online que permite realizar consultas sobre internações na região. A ferramenta permite analisar, por exemplo, os procedimentos mais comuns e a distribuição das internações ao longo de um ano, fornecendo informações para uma alocação mais eficiente de recursos hospitalares.*

## 1. Introdução

Estamos vivendo na era da informação, com uma sociedade cada vez mais orientada a dados. A capacidade de coletar, gerar e armazenar dados através de diferentes sistemas faz com que grandes conjuntos de dados apareçam em diferentes campos de estudo. Na área da saúde, tais dados podem ser gerados a partir de diferentes fontes, como registros de internação hospitalar e equipamentos médicos. Assim, cresce a necessidade de analisá-los para extrair conhecimento e *insights* relevantes para apoiar a tomada de decisão [Keim et al. 2010]. Muitas vezes, esses dados apresentam problemas de qualidade, como a não padronização e dados faltantes, e precisam ser pré-processados.

A interpretação dos dados por parte dos analistas na área da saúde é fundamental para garantir a transparência nas decisões e procedimentos [Kandogan et al. 2014, Milani et al. 2021]. Uma das formas de melhorar essa comunicabilidade é através da

análise exploratória visual, que possibilita, por exemplo, identificar correlações e padrões sobre os dados, diminuindo o esforço cognitivo dos profissionais ao lidar com os dados, facilitando e qualificando decisões, ou permitindo um melhor gerenciamento dos recursos [Preim and Lawonn 2020, Organization 2021].

Atualmente, o DATASUS disponibiliza diversos dados abertos, incluindo informações sobre internações hospitalares, que podem ser coletados e analisados com o auxílio do PySUS<sup>1</sup>. Neste contexto, este trabalho tem como objetivo prover técnicas de visualização interativa combinadas com métodos de exploração de dados sobre a ocupação de hospitais no Rio Grande do Sul. Para isso, disponibilizamos um *dashboard online* que permite realizar consultas sobre internações em hospitais do Rio Grande do Sul (RS).

Nas próximas seções, apresentamos trabalhos relacionados, descrevemos o *dashboard* desenvolvido e analisamos alguns *insights* obtidos. Por fim, discutimos possíveis direções para trabalhos futuros.

## 2. Trabalhos Relacionados

Diversos estudos já exploraram dados hospitalares, propondo diferentes metodologias e soluções. [Secco et al. 2024] analisam dados clínicos para auxiliar médicos no acesso e interpretação rápida dessas informações, facilitando o diagnóstico por meio de uma abordagem visual. Por outro lado, [Ferdib-Al-Islam et al. 2022] focam na previsão do tempo de hospitalização, utilizando aprendizado de máquina para otimizar a alocação de recursos hospitalares. Apesar de ambos os trabalhos lidarem com dados hospitalares, eles não possibilitam uma análise exploratória da demanda hospitalar com base na visualização dinâmica de informações sobre hospitais e internações, que é o foco do nosso estudo.

Uma análise da eficiência hospitalar em diferentes regiões do Brasil é proposta por [da Silva et al. 2017], enquanto [Souza and Costa 2011] comparam taxas de internações por condições sensíveis em municípios-sede de coordenadorias de saúde no RS. Já [Mascarenhas and Barros 2015] descrevem a evolução das hospitalizações por causas externas no Brasil a partir de um estudo de séries temporais sobre internações em hospitais públicos. Embora esses estudos analisem as mesmas fontes de dados hospitalares que nós utilizamos, as abordagens se restringem ao uso de tabelas, sem explorar formas de visualização interativa ou o uso de *dashboards*. Assim, buscamos preencher essa lacuna ao apresentar os dados de maneira visual e interativa, facilitando a análise e auxiliando na tomada de decisões no contexto da gestão hospitalar.

## 3. Descrição do *Dashboard*

Para possibilitar a análise exploratória visual dos dados sobre a ocupação hospitalar no RS, desenvolvemos o *dashboard* apresentado na Figura 1 e disponibilizado *online* para consulta<sup>2</sup>. Nas seções a seguir, detalhamos o processo de preparação dos dados e as visualizações disponibilizadas.

---

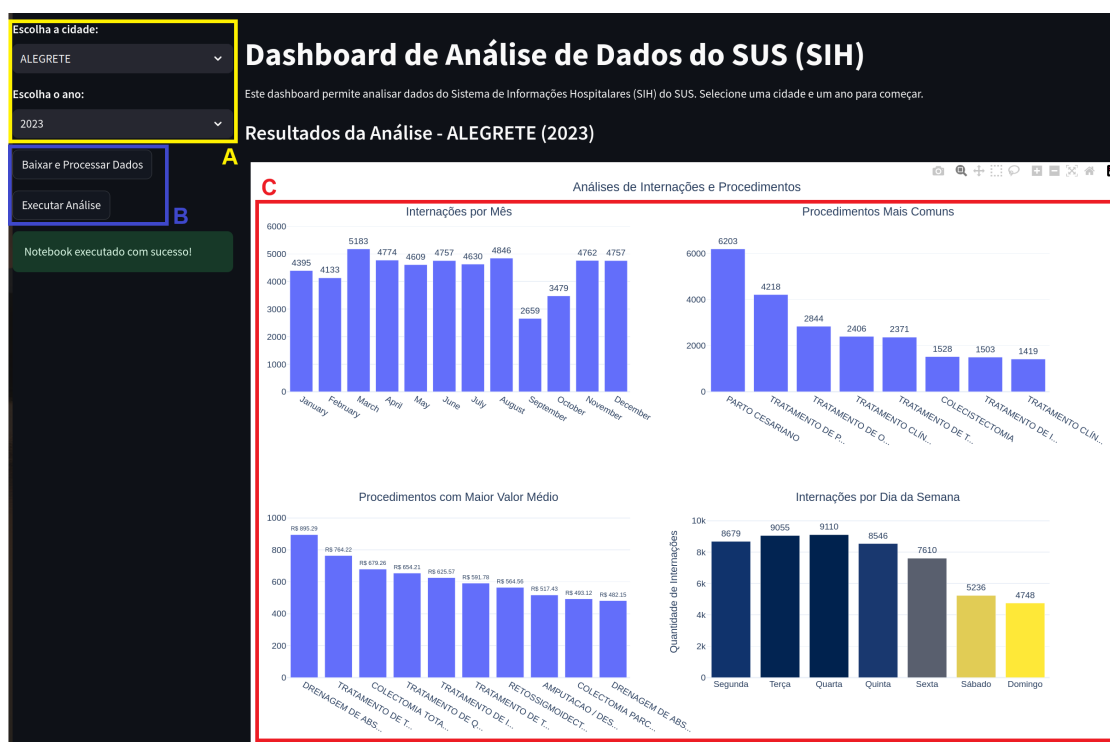
<sup>1</sup><https://pysus.readthedocs.io/pt/latest/>

<sup>2</sup><https://davint-datasus-analysis.streamlit.app/>

### 3.1. Coleta e pré-processamento de dados

Os dados foram obtidos a partir do Sistema de Informações Hospitalares do SUS (SIH/SUS)<sup>3</sup>, utilizando a biblioteca PySUS. O PySUS permite a obtenção dos dados no formato .parquet, selecionando o ano, o estado e o tipo de dado desejado (SIH, SI-NAN, entre outros). Neste estudo, foram utilizados os dados do SIH para analisar as movimentações dentro dos hospitais do estado do RS.

O conjunto de dados disponibilizado é composto por vários atributos, mas foram analisados, principalmente, os seguintes: ano, mês, estado, informações dos hospitais, como código CNES (Cadastro Nacional de Estabelecimentos de Saúde) e número NAIH (Número da Autorização de Internação Hospitalar), data de entrada e saída do paciente, procedimento realizado (ação médica principal, como uma cirurgia ou exame), atos profissionais executados (atividades específicas dentro do procedimento, como anestesia ou sutura), valores correspondentes aos atos profissionais, código do município do hospital e do paciente e quantidade de procedimentos realizados.



**Figura 1. Dashboard para análise de dados de internação hospitalar disponibilizados pelo DATASUS.**

Como os dados são coletados diretamente do PySUS em formato de código (por exemplo, código do IBGE ou códigos internos do SUS), algumas colunas são inviáveis de serem analisadas sem um pré-processamento para identificar o contexto. Portanto, o pré-processamento dos dados foi dividido nas seguintes etapas:

- **1º Passo:** Coleta de dados através do PySUS, integrada a um *dashboard* interativo desenvolvido com a biblioteca Streamlit<sup>4</sup>, que permite a construção de interfaces

<sup>3</sup><http://sihd.datasus.gov.br/>

<sup>4</sup><https://streamlit.io/>

interativas. O usuário seleciona o ano e a cidade que deseja analisar (Figura 1-A), clica no botão para baixar os dados ((Figura 1-B) e eles são baixados utilizando um *prompt* dentro do *dashboard*. Após a coleta, a cidade selecionada previamente é armazenada em um dicionário contendo os códigos do IBGE como valores. Em seguida, é necessário realizar uma verificação para remover todas as linhas que não possuem o código IBGE da cidade selecionada.

- **2° Passo:** Padronização dos códigos de procedimentos e atos profissionais utilizando como referência a Tabela Nacional de Procedimentos do SUS<sup>5</sup>, que contém a descrição de cada código. Além disso, para melhorar a identificação dos municípios, utilizou-se o repositório de municípios brasileiros<sup>6</sup>, permitindo a criação de colunas com os respectivos nomes das cidades, para facilitar a análise.
- **3° Passo:** Inclusão de coordenadas geográficas de acordo com os dados dos municípios, para permitir a futura construção de mapas interativos do estado do RS. Para garantir a precisão, foi realizada uma verificação das cidades homônimas existentes dentro e fora do estado, assegurando que apenas as coordenadas dos municípios do RS fossem consideradas. Como os pacientes podem residir fora do estado e ainda assim serem atendidos em hospitais do RS, essa verificação se aplica apenas aos hospitais e não aos pacientes.

### 3.2. Visualizações

Para a criação de visualizações interativas do *dashboard* utilizamos a linguagem de programação Python e a biblioteca Plotly<sup>7</sup>. Como pode ser visto na Figura 1, inicialmente, optamos por utilizar gráficos de barras, pois são de fácil interpretação. Os gráficos disponibilizados oferecem funcionalidades interativas, como *zoom* e exibição de informações detalhadas via *mouse hover*, além de permitir baixar imagens que podem ser incluídas em relatórios.

O *dashboard* é dividido em três seções principais, como mostra a Figura 1: a parte A, que contém o campo para seleção da cidade e do ano; a parte B, onde se encontram os botões “Baixar e Processar Dados” e “Executar Análise”, sendo este último responsável por executar o *notebook* Python que gera os gráficos; e a parte C, que exibe as representações visuais interativas que possibilitam analisar as internações por mês, os procedimentos mais comuns, os procedimentos com maior valor médio e as internações por dia da semana.

Apesar da versão atual do *dashboard* disponibilizar um número limitado de gráficos, ele já permite a obtenção de diversos *insights* relevantes a partir dos dados analisados. Por exemplo, na Figura 1 é possível perceber uma queda no número de internações na cidade de Alegrete nos meses de setembro e outubro do ano 2023. Esta anormalidade pode ter sido causada pelas enchentes que aconteceram na cidade neste período<sup>8</sup>. Tais fatores podem ter levado a um cancelamento das cirurgias eletivas, reduzindo o número de internações.

Analisando gráficos de outras cidades em diferentes anos, também chamou a atenção a ausência de partos realizados pelo SUS em Guaíba desde 2009. Isto se deve

<sup>5</sup><https://tinyurl.com/2mz2n87e>.

<sup>6</sup><https://github.com/kelvins/Municipios-Brasileiros>.

<sup>7</sup><https://plotly.com/python/>

<sup>8</sup><https://tinyurl.com/rnmt5h9s>

ao encerramento das atividades da maternidade pública do Hospital Nossa Senhora do Livramento.

O padrão sazonal de Gramado também se destacou, pois apresentava um aumento de internações no inverno e no verão, o que pode ser impulsionado pelo maior fluxo de turistas em ambas as estações. Esses padrões destacam a necessidade de preparação do sistema de saúde para atender às demandas sazonais.

#### **4. Discussão**

Atualmente, nosso *dashboard* permite que o usuário selecione um ano e uma cidade específica do RS para analisar os dados que são, então, requisitados, conforme descrito na Seção 3. No entanto, essa abordagem apresenta limitações quando consideramos a possibilidade de expandir a análise para múltiplos anos e cidades. Se cada consulta continuar sendo realizada via requisição, o tempo de resposta, que já é elevado, aumentaria ainda mais devido ao grande volume de dados a ser processado. Como trabalho futuro, planejamos aprimorar nosso trabalho para permitir que os usuários selecionem múltiplos anos e cidades de maneira eficiente. Para viabilizar essa melhoria, uma solução possível seria a utilização de um *datalake*. Essa abordagem nos permitiria armazenar previamente os dados em uma estrutura otimizada, reduzindo significativamente o tempo de processamento e melhorando a experiência do usuário.

A partir desta melhoria, será possível incluir outras visualizações interativas que permitam identificar padrões sazonais na ocupação hospitalar. Também pretendemos disponibilizar mapas interativos para representar a distribuição geográfica da demanda por internações, destacando os hospitais com maior taxa de ocupação. Por fim, com a inclusão de dados de vários anos, será possível aplicar modelos preditivos para estimar tendências futuras de ocupação hospitalar, apoiando a alocação de recursos na área da saúde.

#### **5. Conclusão**

Hoje em dia, há uma grande quantidade de dados abertos do governo relacionados à saúde, incluindo informações sobre internações hospitalares. Desenvolvemos um *dashboard* que possibilita analisar estes dados e realizar consultas que auxiliem no gerenciamento de recursos hospitalares. Apesar de suas limitações, a ferramenta permite a extração de *insights* sobre a ocupação hospitalar, possibilitando analisar os procedimentos mais comuns em diferentes cidades do interior do estado do RS. Com isso, torna-se possível identificar padrões de demanda para alocar recursos de forma mais eficiente e melhorar o atendimento hospitalar. Além disso, ele representa um primeiro passo para futuras melhorias, possibilitando a inclusão de novas funcionalidades e representações visuais que possam apoiar a gestão de recursos hospitalares e a tomada de decisão na área da saúde.

#### **Agradecimentos**

Pedroso, Kniest, Castro e Manssour agradecem o apoio financeiro da bolsa CNPq PIBITI, da bolsa CNPq PIBIC, do Programa de Educação Tutorial (PET) e da bolsa CNPq de produtividade 303208/2023-6, respectivamente. Os autores também agradecem o apoio financeiro parcial da FAPERGS 22/2551-0000390-7 (RITE CIARS Inteligência Artificial Aplicada à Saúde).

## Referências

- da Silva, B. N., Costa, M. A. S., Abbas, K., and Galdamez, E. V. C. (2017). Eficiência hospitalar das regiões brasileiras: Um estudo por meio da análise envoltória de dados. *Revista de Gestão em Sistemas de Saúde*, 6(1):76–91.
- Ferdib-Al-Islam, Robbani, R., and Ullah, M. W. (2022). Cov-hm: Prediction of covid-19 patient's hospitalization period for hospital management using smote and machine learning techniques. In *Proceedings of the 2nd International Conference on Computing Advancements (ICCA)*, page 25–33, NY, USA. Association for Computing Machinery.
- Kandogan, E., Balakrishnan, A., Haber, E. M., and Pierce, J. S. (2014). From data to insight: work practices of analysts in the enterprise. *IEEE Computer Graphics and Applications*, 34(5):42–50.
- Keim, D., Kohlhammer, J., Ellis, G., and Mansmann, F. (2010). *Mastering the information age solving problems with visual analytics*. Eurographics Association.
- Mascarenhas, M. D. M. and Barros, M. B. d. A. (2015). Evolução das internações hospitalares por causas externas no sistema público de saúde - Brasil, 2002 a 2011. *Epidemiologia e Serviços de Saúde*, 24:19 – 29.
- Milani, A. M. P., Loges, L. A., Paulovich, F. V., and Manssour, I. H. (2021). Prava: Preprocessing profiling approach for visual analytics. *Information Visualization*, 20(2-3):101–122.
- Organization, W. H. (2021). Global strategy on digital health 2020-2025. Licence: CC BY-NC-SA 3.0 IGO.
- Preim, B. and Lawonn, K. (2020). A survey of visual analytics for public health. In *Computer Graphics Forum*, volume 39, pages 543–580. Wiley Online Library.
- Secco, C. A., Sina, L. B., and Nazemi, K. (2024). Medical visual analytics - an interactive approach for analyzing electronic health records. In *2024 28th International Conference Information Visualisation (IV)*, pages 143–149.
- Souza, L. L. d. and Costa, J. S. D. d. (2011). Internações por condições sensíveis à atenção primária nas coordenadorias de saúde no rs. *Revista de Saúde Pública*, 45(4):765–772.