

Análise de Dados do DATASUS

André Luiz Rodrigues

Giovane Bianchi Milani

Thiago Borges de Souza

Vinícius André Rech Boff

Escola Politécnica – Pontifícia Universidade Católica do Rio Grande do Sul
(PUCRS)

Avenida Ipiranga, 6681 – Prédio 30 – Bloco C – Sala 101

CEP 90619-900 – Porto Alegre – RS – Brasil

1. Introdução

A presente atividade extensionista tem como objetivo aplicar conceitos e técnicas de Ciência de Dados para a análise de dados, promovendo a disseminação do conhecimento através da disponibilização aberta de todo o processo e resultados obtidos. Neste contexto, cada grupo participante é responsável por realizar a coleta, preparação e análise de dados governamentais abertos disponíveis pelo DATASUS.

Este grupo decidiu analisar qual foi a eficiência do sistema de saúde brasileiro, no escopo do estado do Rio Grande do Sul, durante o ano de 2010 até 2021 a partir das seguintes questões:

Relação entre procedimentos aprovados e procedimentos realizados;

Distribuição dos procedimentos realizados e aprovados por faixa etária;

As pessoas que vieram a óbito receberam assistência antes de virem a óbito?

Há discriminação ou preferencia quanto a idade/sexo para realização de assistência/exames/cirurgia?

A partir das conclusões obtidas, a sociedade brasileira será beneficiada, pois poderemos ver como o SUS se comportou ao longo do tempo em um dos estados da nação, através dos diversos governos e medidas que ocorreram nesses

anos. Assim, será mais fácil entender a eficiência da saúde pública durante a última década na nossa sociedade.

Para atingir o objetivo do trabalho, utilizamos o PySUS, que é uma biblioteca que consiste em uma coleção de códigos auxiliares para pessoas que precisam analisar dados do SUS (Sistema Único de Saúde). Com ele, pegamos os dados do SUS que tinham relação com o nosso tema, efetuamos uma limpeza neles e os analisamos. Também, o Power BI foi utilizado para a criação de um dashboard que facilitasse a compreensão e entendimento dos dados, tornando-os mais tangíveis para a nossa realidade.

2. Repositório no Github

Para a realização deste trabalho, foi feita a criação de um repositório no Github, para agrupar a produção do grupo e tornar mais fácil a execução de nosso planejamento. O link para este repositório, onde todos os notebooks estão disponíveis de forma aberta, é: <https://github.com/giovanebmilani/trab-datasus>.

3. Descrição do Processo

A seguir, elencamos os passos feitos neste estudo, que resultaram nas conclusões auferidas. Estes estão divididos em quatro itens. O primeiro diz respeito às bases de dados utilizadas, o segundo fala das técnicas de integração empregadas, o terceiro sobre como preparamos os dados para a análise, e o último explica como a ferramenta do Power BI foi usada.

3.1 Bases de Dados Escolhidas

As fontes de dados foram avaliadas quanto à confiabilidade, atualidade e representatividade para garantir resultados robustos. Nosso escopo se concentrou no estado do Rio Grande do Sul, entre os anos de 2010 a 2021, e, devido a isso, escolhemos as bases que melhor atendessem a ele. Importamos duas bases de dados, a SIM e SIA.

Vale-se notar que tivemos dificuldades em importar os dados do PySUS. Esta dificuldade se deu devido à enorme quantidade de dados no SIA e SIM e ao fato do PySUS permitir apenas a extração de mês a mês. Decidimos, portanto, a nos limitarmos a 500 amostras mensais em cada mês, de 2010 até 2021, somente

relativo ao estado do Rio Grande do Sul, o que chegou a levar 9 horas, a despeito da estratificação feita.

3.1.1 Base de Dados SIM

O Sistema de Informação sobre Mortalidade (SIM) possui informações fundamentais para que possamos conhecer os aspectos referentes à mortalidade no Brasil e às causas de adoecimento que levaram ao óbito. É, ainda, um dos principais instrumentos para apoiar a elaboração de políticas públicas de saúde e seguridade social mais efetivas visando à prevenção, promoção e cuidado em saúde. Do SIM, utilizamos as colunas de data de óbito, do código do município de naturalidade do falecido, da idade do falecido, seu sexo, se teve assistência médica, se teve exame, ou se teve cirurgia antes de seu óbito.

3.1.2 Base de Dados do SIA

A base de dados do Sistema de Informação Ambulatorial (SIA) tem como principais finalidades o registro padronizado dos atendimentos ambulatoriais do Sistema Único de Saúde (SUS), a geração de dados e informações relacionados aos serviços ambulatoriais do SUS, a contribuição para o processo de planejamento, controle, avaliação e auditoria do SUS, e a viabilização do faturamento ambulatorial no âmbito do SUS. Do SIA, utilizamos as colunas do código do município da unidade de saúde, da região da unidade de saúde, da idade do paciente, do sexo do paciente, da quantidade de procedimentos realizados no paciente, da quantidade de procedimentos aprovados no paciente, do valor total dos procedimentos realizados no paciente e do valor total dos procedimentos aprovados no paciente.

3.2 Técnicas de Integração

Para realizar uma análise abrangente em um conjunto de dados, técnicas de integração são comumente empregadas quando existe a combinação de dois ou mais conjuntos de dados. Se busca mesclar esses dados de forma harmônica, suavizando as heterogeneidades e discrepâncias. Neste trabalho, não foi necessário o uso dessas técnicas, pois tanto o SIM quanto o SIA se integravam de forma satisfatória para a nossa análise, não havendo inconformidades gritantes

necessitadas de tratamento. Os dois datasets se relacionam por meio do código do município, que é o mesmo padrão nos dois.

3.3 Limpeza e Transformação

O processo de limpeza e transformação de dados foi uma etapa crucial. Eliminaram-se dados inconsistentes, trataram-se valores ausentes e aplicaram-se técnicas de normalização para garantir a qualidade e consistência dos dados utilizados nas análises.

As seguintes etapas foram aplicadas:

1. Eliminação de colunas que não seriam utilizadas
2. Eliminação de duplicatas
3. Transformação de dados para o tipo ou formato mais adequado
 - a. Transformação de valores numéricos de formato String para Float ou Inteiro;
 - b. Transformação da idade das pessoas para formato comum em anos.
 - c. Transformação de data para o formato dia-mês-ano.

Ao final desses processos, obtivemos dados menos problemáticos e mais condizentes com a realidade. Com as transformações feitas, também os preparamos para a seguinte etapa de importação para o Power BI.

3.4 Power BI

Durante a realização dos dashboards para análise, observou-se uma certa discrepância nos dados, dificultando a análise. Por exemplo, o número total de procedimentos realizados é superior ao número de procedimentos aprovados. Desse modo, não há como medir a eficiência nos atendimentos, visto que se sabe que não é de 100% a taxa de realização de procedimentos no Sistema Único de Saúde Brasileiro.

4. Conclusão

As análises realizadas proporcionaram insights valiosos sobre o SUS e a saúde da população brasileira. Porém não foi possível aprofundar a análise da forma que o grupo originalmente tinha planejado, que se deu por falta de documentação e mal dimensionamento da ideia inicial.

Gostaríamos de destacar o quanto gostamos deste trabalho, considerando-o de extrema importância para o nosso processo formativo de indivíduos graduados em cursos de tecnologia. Sabendo que o mundo atual considera os dados como um “ouro digital”, ter habilidades relativas a esse assunto nos propicia melhores chances no mercado de trabalho. Não obstante, o próprio fato da possibilidade de auxiliar o nosso país de alguma forma a partir dos resultados deste estudo nos deixa intrinsecamente orgulhosos. Vale, por fim, agradecer o professor da disciplina por todo o conhecimento destilado em nós durante o decorrer deste semestre.

Referências

Todas as referências consultadas e utilizadas durante a atividade estão listadas abaixo:

<https://pcdas.icict.fiocruz.br/conjunto-de-dados/sistema-de-informacoes-de-mortalidade-sim/dicionario-de-variaveis/>

<https://opendatasus.saude.gov.br/dataset/sim>

COELHO, FLAVIO CODECO. **PySUS**, 2023. Página inicial. Disponível em: <https://pysus.readthedocs.io/pt/latest/>. Acesso em: 26 de nov. de 2023.