



Campus São Mateus
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO



TRANSFORMANDO DADOS EM INFORMAÇÃO: APLICAÇÃO DE BUSINESS INTELLIGENCE PARA A AUTOMAÇÃO DA ANÁLISE DE DADOS EM SAÚDE

TRANSFORMING DATA INTO INFORMATION: APPLICATION OF BUSINESS INTELLIGENCE FOR AUTOMATING DATA ANALYSIS IN HEALTHCARE

TRANSFORMANDO DATOS EN INFORMACIÓN: APLICACIÓN DE INTELIGENCIA DE NEGOCIOS PARA LA AUTOMATIZACIÓN DEL ANÁLISIS DE DATOS EN SALUD

Cleiton Schwambach Gaioti ¹, Michael Ruberson Ribeiro da Silva ², Eduardo Frizzera Meira ³, Jéssica Barreto Ribeiro dos Santos ⁴, & Antônio Almeida de Barros Júnior ^{5*}

^{1 2 3 4 5} Universidade Federal do Espírito Santo, Centro de Ciências Exatas, Naturais e da Saúde

¹ cleiton.s.gaioti@gmail.com ² mike.ufes@gmail.com ³ eduardo.meira@ufes.br ⁴ jessica.br.santos@ufes.br

⁵ antonio.barros@ufes.br

ARTIGO INFO.

Recebido: 18.06.2024

Aprovado: 25.07.2024

Disponibilizado: 06.08.2024

PALAVRAS-CHAVE: *Inteligência de Negócios; Epidemiologia; Análise de Dados.*

KEYWORDS: *Business Intelligence; Epidemiology; Data Analysis.*

PALABRAS CLAVE: *Inteligencia de Negocios; Epidemiologia; Análisis de Datos.*

*Autor Correspondente: da Silva, M. R. R.

RESUMO

Contexto: avanços científicos recentes indicam uma nova fronteira na epidemiologia: a análise de grandes conjuntos de dados (*Big Data*), onde ferramentas de *Business Intelligence* desempenham um papel fundamental. **Objetivo:** Implementar um *Data Warehouse* (DW) para análise de dados em saúde e testar sua utilização na análise de indicadores de saúde mental no Espírito Santo (ES). **Métodos:** O estudo foi dividido em quatro fases: 1) identificação de bases de dados e indicadores; 2) extração, transformação e carga de dados, 3) criação de um *Data Warehouse* e; 4) processamento analítico com visualização de dados. **Resultados:** A partir da criação do DW, identificou-se que foram gastos R\$ 53,7 milhões em internações relacionadas à saúde mental no estado do Espírito Santo, com uma redução progressiva ao longo dos anos. Durante a pandemia de covid-19, houve um aumento superior a 100% nos óbitos relacionados a transtornos mentais, especialmente ligados ao uso de álcool e tabaco, mais comuns entre indivíduos pardos, do sexo masculino e entre 45 e 59 anos. **Conclusão:** A implementação de um DW para análise de dados em saúde permitiu identificar importantes indicadores de saúde mental no Espírito Santo e permitirá novas análises no contexto da saúde pública no ES e Brasil.

ABSTRACT

Context: Recent scientific advances indicate a new frontier in epidemiology: the analysis of large datasets (*Big Data*), where *Business Intelligence* tools play a fundamental role. **Objective:** To implement a *Data Warehouse* (DW) for health

data analysis and test its use in analyzing mental health indicators in Espírito Santo (ES). **Methods:** The study was divided into four phases: 1) identification of databases and indicators; 2) data extraction, transformation, and loading; 3) creation of a *Data Warehouse*; and 4) analytical processing with data visualization. **Results:** The creation of the DW identified that R\$ 53.7 million was spent on mental health-related hospitalizations in Espírito Santo, with a progressive reduction over the years. During the COVID-19 pandemic, there was an over 100% increase in deaths related to mental disorders, especially linked to alcohol and tobacco use, more common among brown-skinned individuals, males, and those aged 45 to 59 years. **Conclusion:** The implementation of a DW for health data analysis enabled the identification of important mental health indicators in Espírito Santo and will allow for new analyses in the context of public health in ES and Brazil.

RESUMEN

Contexto: Avances científicos recientes indican una nueva frontera en la epidemiología: el análisis de grandes conjuntos de datos (*Big Data*), donde las herramientas de *Inteligencia de Negocios* desempeñan un papel fundamental. **Objetivo:** Implementar un *Data Warehouse* (DW) para el análisis de datos de salud y probar su uso en el análisis de indicadores de salud mental en Espírito Santo (ES). **Métodos:** El estudio se dividió en cuatro fases: 1) identificación de bases de datos e indicadores; 2) extracción, transformación y carga de datos; 3) creación de un *Data Warehouse*; y 4) procesamiento analítico con visualización de datos. **Resultados:** La creación del DW identificó que se gastaron R\$ 53,7 millones en hospitalizaciones relacionadas con la salud mental en Espírito Santo, con una reducción progresiva a lo largo de los años. Durante la pandemia de COVID-19, hubo un aumento de más del 100% en las muertes relacionadas con trastornos mentales, especialmente vinculados al uso de alcohol y tabaco, más comunes entre individuos de piel morena, hombres y aquellos de 45 a 59 años. **Conclusión:** La implementación de un DW para el análisis de datos de salud permitió identificar importantes indicadores de salud mental en Espírito Santo y permitirá nuevos análisis en el contexto de la salud pública en ES y Brasil.

INTRODUÇÃO

A análise de grandes bancos de dados (*Big Data*) avançou significativamente nos últimos anos e contribui para a análise epidemiológica em saúde coletiva (Chiavegatto Filho, 2015). O *Big Data* encontra-se em amplo desenvolvimento, possibilitado pela convergência de dois fatores: a necessidade de divulgação de evidências científicas e o desenvolvimento analítico-computacional necessário para a análise intensiva de dados (Chiavegatto Filho, 2015). Estratégico para outras áreas, como o e-commerce (Heck, 2019), o uso de *Big Data* está se tornando uma realidade na área da saúde, o que demanda inovação para a análise de dados complexos e desestruturados, com técnicas de *Data Mining* (Fernandes & Chiavegatto Filho, 2019; Chiavegatto Filho, 2015).

A análise de dados visa melhorar a eficiência dos serviços ofertados, lidando com grandes conjuntos de dados. Ela busca entender padrões históricos para otimizar o atendimento ao paciente, gerenciar custos e recursos, e aprimorar diagnósticos e tratamentos. O resultado são ferramentas de visualização que auxiliam os profissionais, melhorando os serviços prestados (Ognjanovic, 2020).

O Brasil tem uma longa tradição de manter registros de informações relacionadas à saúde para fins administrativos devido à criação do Sistema Único de Saúde (SUS) (Coelho & Chioro, 2021; Paim et al., 2011). Vinculado ao SUS, o Departamento de Informática do Sistema Único de Saúde (DATASUS) é o responsável pela gestão, organização e disponibilização dos dados administrativos, clínicos e epidemiológicos (Ali et al, 2019). O DATASUS foi criado em 1991 para disponibilizar sistemas de informação aos órgãos do SUS, aplicados ao planejamento, operação e controle de ações e serviços de saúde, além de desenvolver outros programas e iniciativas na transformação digital na área da saúde (Daniel, Pereira & Macadar, 2014).

Segundo Ali e colaboradores (2019), apesar dos esforços do Ministério da Saúde para harmonizar o registro das informações, existe uma grande disparidade entre as várias instituições de saúde relacionadas aos processos de coleta de dados. Além disso, os dados individuais coletados por diferentes serviços de saúde (por exemplo, registros hospitalares e de mortalidade) carecem de identificadores-chave exclusivos para os indivíduos, portanto, combinar essas fontes de dados não é uma tarefa trivial. Esses fatores, além da falta de infraestrutura tecnológica e de recursos humanos qualificados, têm limitado o uso de dados coletados rotineiramente para gerar evidências que apoiem as decisões clínicas e políticas e respondam a importantes questões epidemiológicas.

Para auxiliar nesse processo, Cozzoli, Salvatore & Milone (2022) descrevem que o Business Intelligence (BI) é uma ferramenta importante para implementar aplicações de Big Data bem-sucedidas. Para tanto, um passo importante para aplicações em BI e Big DATA é a construção de um ambiente de Data Warehouse para a agregação inicial dos dados (Cozzoli, Salvatore & Milone, 2022).

As técnicas de extração, transformação, carregamento, processamento, recuperação e distribuição de dados precisam ser cada vez mais precisas, especialmente ao lidar com grandes volumes de dados (Cheng, Pazmino, Schreiweis, 2022). Frequentemente, os dados a serem

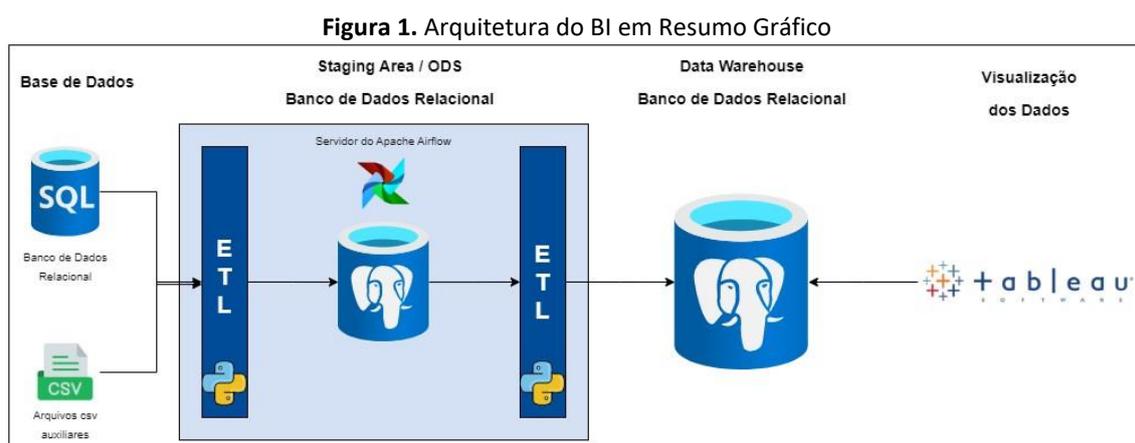
analisados vêm de fontes diversas, com diferentes estruturas e formatos (Coeli, 2020). Por isso, há um grande esforço na área de tecnologia da informação para desenvolver arquiteturas eficientes que resolvam esses problemas (Medeiros, Maçada & Hoppen, 2021). Implementações voltadas para o suporte a Big Data estão sendo amplamente discutidas no contexto da saúde pública. Em termos gerais, muitas estratégias de processamento intensivo de dados têm sido exploradas para melhorar a prevenção e o tratamento de doenças (Batko & Ślęzak, 2022).

Entretanto, apesar da disponibilidade dos dados do DATASUS, a utilização regular e automatizada desses dados ainda não é uma realidade (Viana et al, 2023). Nesse sentido, a avaliação dos indicadores em saúde por meio do DATASUS poderá subsidiar a reavaliação de políticas públicas setoriais bem como apoiar a gestão pública estadual, impactando diretamente na qualidade de vida da população (Oliveira, Sampaio & Oliveira, 2011). Assim, estudos que utilizam base de dados para a geração de evidências e informações, por meio dos indicadores em saúde, tem permitido a identificação de tendências, melhorias e riscos para a saúde da população no Brasil (Faria et al, 2022; Maior, Osório-De-Castro, Andrade, 2020; Ali et al, 2019).

Nesse sentido, o objetivo do estudo é construir um ambiente de BI a partir dos dados dos Sistemas de Informação do DATASUS. Para validar a construção do DW, foram descritas as séries temporais de hospitalizações e mortalidade por transtornos mentais e do comportamento no Espírito Santo, divulgadas por meio de *dashboards*.

METODOLOGIA

Neste trabalho, foi implementado um ambiente de BI baseado na arquitetura de Kimball & Ross (2013), que é capaz de orquestrar o fluxo de Extração, Transformação e Carregamento dos dados (ETL) para povoar o *Data Warehouse* proposto. A Figura 1 apresenta a arquitetura proposta e que será detalhada a seguir.



Fase 01: identificação das bases de dados e dos indicadores de interesse (ferramenta: bases de dados e indicadores do DATASUS);

Fase 02: criação das staging áreas para a realização do processo de ETL. Nesta etapa foram utilizadas as ferramentas: Python, Apache Airflow e Postgree SQL;

Fase 03: Criação do Data Warehouse, um armazém de dados prontos para a análise desenvolvido no Sistema de Gerenciamento de Banco de Dados PostgreSQL;

Fase 04: Processamento analítico e visualização de dados, ao qual utilizou-se a ferramenta Tableau Public.

Fase 01: Bases de dados e Indicadores

A fase inicial do processo consistiu em realizar um estudo dos requisitos de avaliação, com o objetivo de mapear os principais indicadores e métricas que seriam atendidos no processo de BI. Nesse sentido, foram escolhidas as bases de dados do Sistema de Informação de Mortalidade (SIM) para os indicadores de mortalidade e a base de dados do Sistema de Informação Hospitalar (SIH) para os indicadores de hospitalizações.

Uma vez obtidos as bases de dados e seus respectivos indicadores, foi realizado um processo de modelagem dimensional para determinar a tabela fato e devidas tabelas dimensões necessárias para atender aos requisitos propostos. A Modelagem dimensional é uma técnica de design de banco de dados utilizada para organizar e estruturar dados de forma intuitiva e eficiente, facilitando a análise e extração de informações. Ela é baseada em dois conceitos principais: fatos e dimensões. Os fatos representam as métricas que queremos analisar, enquanto as dimensões são os contextos ou características associadas aos fatos. Essa abordagem permite criar esquemas flexíveis e simplificados, promovendo uma compreensão mais clara dos dados e facilitando consultas analíticas.

Posteriormente, foi criada uma estrutura utilizando o Sistema de Gerenciamento de Banco de Dados Objeto-Relacional (ORDBMS) PostgreSQL, o qual serviu como uma área de preparação de dados. O propósito principal dessa estrutura era armazenar os dados em sua forma original até que pudessem ser acessados pelas camadas subsequentes do DW. Os dados permaneceram na staging area até estarem sincronizados e prontos para serem utilizados. A staging area é um espaço intermediário onde os dados são temporariamente armazenados e preparados antes de serem carregados no ambiente de produção.

Fase 02: Extração, transformação e carga dos dados (ETL)

A segunda etapa do processo consistiu em realizar a chamada ETL - Extract, Transform and Load (Extração, Transformação e Carga dos dados) com a finalidade de construir inicialmente o Data Warehouse (Fase 3).

Durante a Fase 2 (ETL), os dados foram extraídos da staging area e submetidos a um procedimento de limpeza e padronização, que incluiu a definição dos tipos de dados para cada coluna e a purificação de campos textuais. Em seguida, esses dados foram utilizados para alimentar as tabelas dimensões e tabela fato do DW, também mantido no ORDBMS PostgreSQL.

Para a extração dos dados, utilizou-se o módulo Pysus (Coelho et al., 2021) da linguagem Python, o qual consiste em um conjunto de ferramentas voltadas ao tratamento de bases de dados públicas disponibilizadas pelo DATASUS, órgão responsável pela divulgação de informações de saúde no Brasil. Quanto ao gerenciamento do fluxo do processo de ETL, empregou-se o Apache Airflow, uma plataforma que viabiliza a criação e a orquestração de fluxos automatizados.

O processo ETL consistiu em:

Extract:

- a) Acessar o servidor *File Transfer Protocol* (FTP) do DATASUS via Pysus e baixar os arquivos com extensão *Data Base Compressed* (.DBC) contendo os microdados anônimos da base de dados desejada de todo o Brasil para cada ano de interesse;
- b) Converter os arquivos baixados para a extensão .DBF (Data Base File);
- c) Ler e agrupar os arquivos em um único conjunto de dados para posterior tratamento.

Transform:

- a) Eliminar valores inválidos;
- b) Decodificar e incrementar os dados de acordo com dicionário de dados da base de dados;
- c) Incrementar os dados dos municípios;
- d) Incrementar os dados dos estados;
- e) Incrementar os dados referentes a Classificação Estatística Internacional de Doenças e Problemas Relacionados com a Saúde, 10ª versão (CID10);
- f) Incrementar os dados de data e idade;
- g) Incrementar os dados de coordenadas dos municípios para georreferenciamento, com a utilização da API de localidades do IBGE. Foram extraídos dados municipais, bem como das regiões imediata, intermediária, mesorregião e microrregião político-administrativa.

Load:

- a) Indexar dados tratados da base de dados de origem em banco de dados;
- b) Coletar log de indexação.

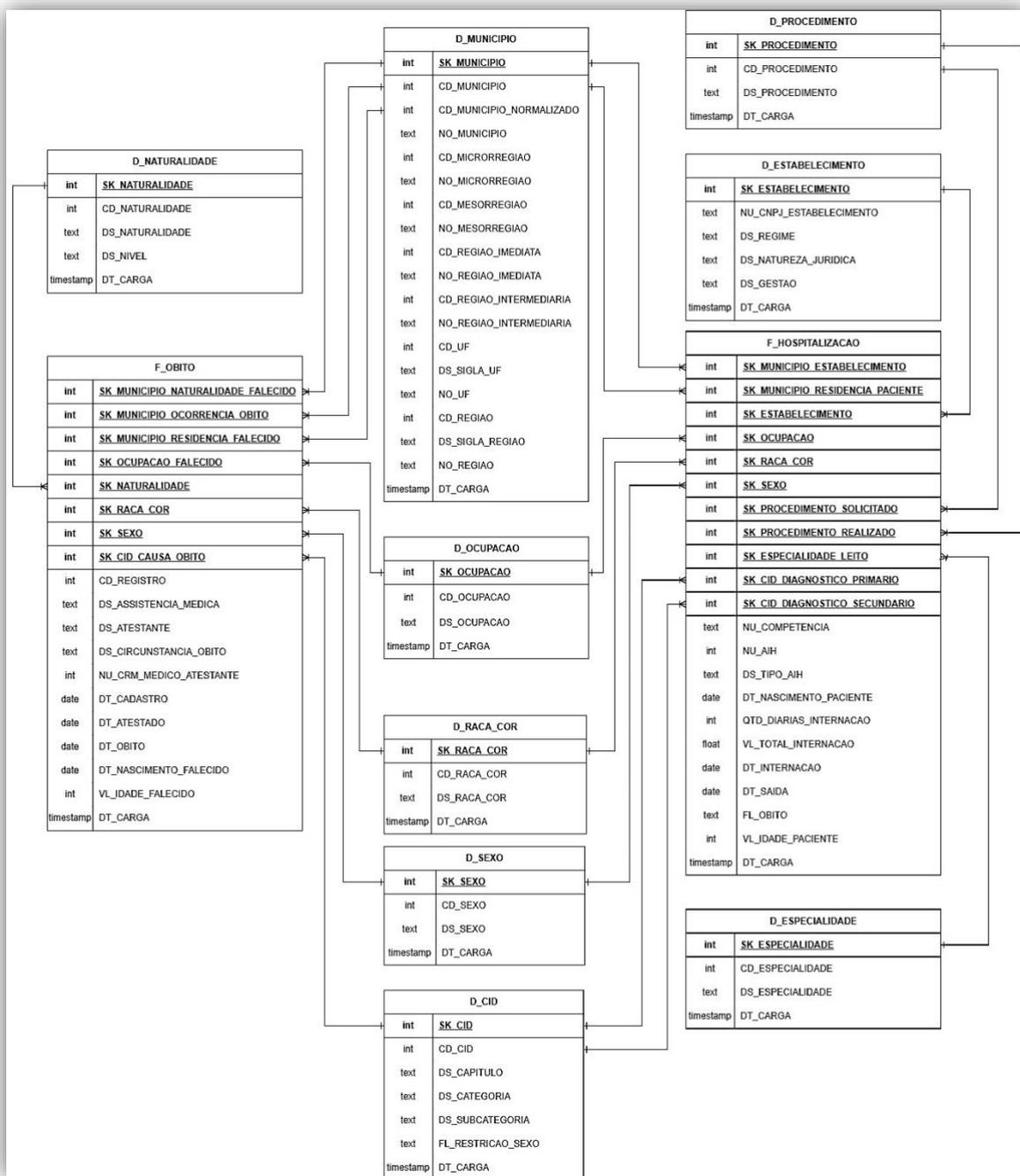
Para a etapa de transformação dos dados foram utilizados os dicionários de dados específicos para cada uma das bases de dados utilizadas (SIM e SIH), disponibilizados no site da Plataforma de Ciência de Dados aplicada a Saúde (PCDaS) para o enriquecimento das fontes de dados extraídas do DATASUS, juntamente com fontes de dados complementares utilizadas para ampliar as informações obtidas pelo SIM e pelo SIH (FIOCRUZ, 2023).

Fase 03: Construção do Data Warehouse

Posteriormente à identificação dos requisitos e à análise preliminar dos dados provenientes da fonte, foi desenvolvido um modelo dimensional, seguindo as diretrizes do modelo *Star Schema*, adotado neste estudo. O modelo *Star Schema* é uma técnica de modelagem dimensional amplamente utilizada em *Data Warehousing*, caracterizado pela estrutura centralizada em uma tabela de fatos, com dimensões conectadas a ela, proporcionando uma visão simplificada e eficiente dos dados organizados.

A Figura 2 apresenta o modelo dimensional desenvolvido, contendo dois fatos: F_OBITO e F_HOSPITALIZACAO. Além dos fatos, o modelo conta com nove dimensões: D_MUNICIPIO, D_NATURALIDADE, D_PROCEDIMENTO, D_ESTABELECIMENTO, D_OCUPACAO, D_RACA_COR, D_SEXO, D_CID e D_ESPECIALIDADE.

Figura 2. Modelo Dimensional do cubo de dados



Fonte: Autores.

Todas as tabelas dimensão desenvolvidas são *Slowly Changing Dimensions* (SCD) do tipo 1, o que significa que seus registros podem ser atualizados caso haja alterações nos registros provenientes da fonte. No entanto, as tabelas não mantêm um histórico dos registros anteriores a cada mudança. A cada registro nas tabelas mencionadas é atribuído a uma *Surrogate Key*, a qual desempenha a função de chave primária das tabelas de dimensão. Essas *Surrogate Key* são armazenadas como chaves estrangeiras nas tabelas de fato, permitindo a posterior correlação entre as tabelas de fato e dimensão.

As dimensões de CID, município, ocupação, raça e sexo foram compartilhadas entre as duas tabelas fatos. A dimensão de naturalidade estabelece uma conexão exclusiva com a tabela fato de óbito, ao passo que as dimensões de especialidade, estabelecimento e procedimento estabelecem conexões exclusivas com a tabela fato de hospitalização. Não houve registros duplicados entre as duas tabelas fatos, tendo em vista que as duas não se correlacionam.

Para padronizar as informações das dimensões “D_SEXO” e “D_RACA_COR” provenientes de diferentes fontes, foram realizadas transformações nos dados. Nessas dimensões, utilizam-se códigos identificadores fixos. Durante o processo de ETL das tabelas fato, os códigos do SIH e do SIM foram ajustados para os equivalentes na dimensão correspondente, com essas variáveis apresentando a descrição em texto (*string*) de cada alternativa disponível.

Para a extração dos dados do Código Brasileiro de Ocupação (CBO) do DATASUS, utilizou-se o módulo Pysus da linguagem Python. Este módulo fornece as informações de CBO já processadas, com os códigos formatados exclusivamente em números.

Após o processo de ETL (Fase 2), o Data Warehouse foi carregado com os dados tratados (Fase 3) e está pronto para o processamento analítico dos dados (Fase 4).

Fases 04: Criação dos Dashboards

Finalmente, uma vez que as estruturas do DW foram preenchidas, utilizou-se a ferramenta de processamento analítico online (OLAP) *Tableau Desktop* para a criação dos dashboards, dos quais os indicadores e métricas com os resultados obtidos foram mapeados, de acordo com os requisitos identificados na etapa inicial.

A escolha pela ferramenta OLAP Tableau Desktop se deve ao fato de sua licença gratuita, para estudantes universitários e pesquisadores, e por apresentar uma ampla gama de recursos, que favorecem a agilidade e a intuição no processo de criação das visualizações.

O Tableau Desktop estabelece uma conexão direta com o banco de dados, permitindo a configuração da associação entre as dimensões e os fatos. A página do Tableau Public que contém o dashboard completo está disponível para acesso por meio do link do Repositório SAÚDE MENTAL NO ES | Tableau Public (GAIOTI, 2024).

Para validação do Data Warehouse (Fase 3), foi feita uma análise da mortalidade relacionada a transtornos mentais e comportamentais no estado do Espírito Santo (Fase 4). Foram considerados casos elegíveis as mortes em que o diagnóstico principal correspondesse a um dos códigos definidos para transtornos mentais e do comportamento listados na 10ª Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde (CID-10), que abrange os códigos F00 a F99.

RESULTADOS E DISCUSSÃO

Fase 01: Levantamento de indicadores

Foram identificados e definidos os indicadores-chave de desempenho, com a estruturação de um modelo dimensional eficiente, que organizou e relacionou os dados de forma adequada, facilitando a análise e extração de *insights* relevantes. Isso proporcionou uma visão holística

do negócio, possibilitando a identificação de tendências, padrões e gargalos, além de fornecer informações valiosas para o planejamento estratégico e a otimização dos processos internos.

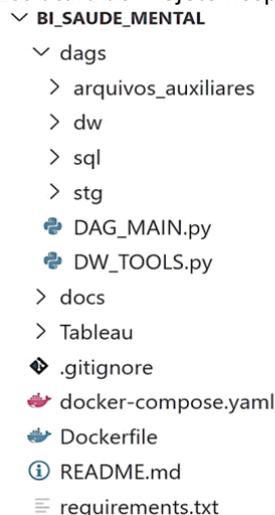
No presente estudo, foram estabelecidos os seguintes indicadores para análise:

- Visão histórica de mortalidade por transtornos mentais e do comportamento;
- Principais causas de mortalidade por transtornos mentais e comportamentais;
- Nº de óbitos por transtornos mentais e comportamentais por raça e cor;
- Nº de óbitos por transtornos mentais e comportamentais por município do ES;
- Nº de óbitos por transtornos mentais e comportamentais por sexo;
- Nº de óbitos por transtornos mentais e comportamentais por faixa etária;

Fase 02: Estrutura do projeto e processo de ETL

A Figura 3 apresenta a estrutura criada para o projeto, ao qual foi disponibilizado no repositório do *GitHub*.

Figura 3. Estrutura do Projeto hospedado no GitHub



Fonte: Autores.

A pasta "raiz" do projeto possui três subpastas, denominadas "*dags*", "*docs*" e "*Tableau*". Além disso, nessa pasta raiz encontram-se os arquivos responsáveis pela criação e configuração do ambiente do *Airflow*.

O detalhamento de cada pasta, representada na Figura 3, são descritos a seguir.

Na subpasta "*dags*" estão alocados todos os *scripts* fundamentais para a criação do fluxo de ETL do projeto. Em seu diretório principal encontra-se o arquivo "DAG_MAIN.py", o qual é responsável pela configuração da DAG e pelo controle do fluxo de ETL. Além disso, encontra-se o arquivo "DW_TOOLS.py", que contém funções utilizadas em todo o projeto.

No diretório denominado "*sql*" contém os *scripts* escritos em linguagem SQL que assumem a responsabilidade de criar a estrutura do banco de dados no sistema do *PostgreSQL*. Tais *scripts* são executados de maneira automatizada por meio do fluxo estabelecido no *Airflow*.

No diretório denominado "arquivos_auxiliares" inclui os arquivos no formato CSV disponibilizados pela PCDaS. Tais arquivos são empregados para o preenchimento de determinadas tabelas na *staging area*.

No diretório "stg" são encontrados todos os *scripts* responsáveis pela extração dos dados provenientes das fontes de origem, bem como sua posterior carga na *staging area*. Durante essa etapa, nenhuma transformação é aplicada aos dados, a fim de garantir a preservação de sua integridade.

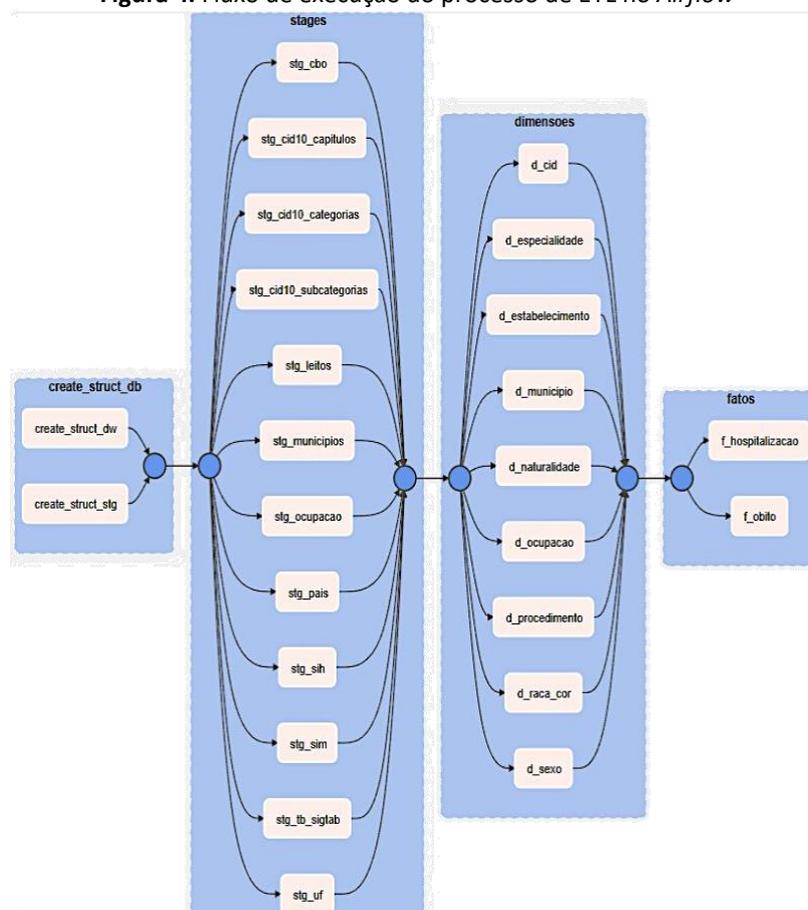
No diretório "dw" são encontrados todos os *scripts* responsáveis pelo processo de ETL, que executam as atividades de limpeza e transformação dos dados, seguidos pelo carregamento das dimensões e fatos.

Na subpasta "Docs" foram armazenados todos os documentos referentes ao projeto. Nele encontra-se uma cópia do Dicionário de Dados e do modelo do dimensional implementado neste trabalho.

Na subpasta "Tableau" há uma duplicata da planilha de trabalho do *Tableau* contendo os painéis de controle desenvolvidos, bem como um arquivo contendo uma extração dos dados obtidos do DW.

A Figura 4 exhibe a DAG no *Airflow*, no qual é possível observar o fluxo do processo de ETL.

Figura 4. Fluxo de execução do processo de ETL no *Airflow*

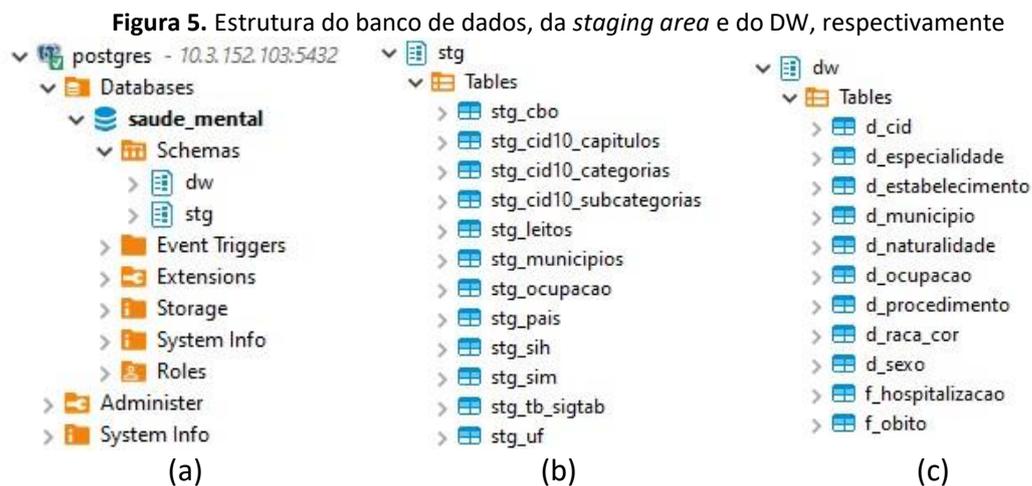


Fonte: Autores.

A DAG foi dividida em quatro grupos de tarefas. O grupo denominado "*create_struct_db*" abrange as tarefas responsáveis pela criação da estrutura do banco de dados. O grupo intitulado "stages" engloba todas as tarefas responsáveis pela execução dos *scripts* que extraem os dados da fonte e carregam as tabelas da *staging area*. O grupo designado como "dimensoes" congrega todas as tarefas responsáveis pelo processo de ETL das dimensões do cubo. Por fim, o grupo denominado "fatos" contém as tarefas responsáveis pelo processo de ETL das tabelas fato do cubo. A execução ocorre de acordo com um fluxo unidirecional, e os grupos são executados em conformidade com a ordem de precedência mencionada anteriormente, sendo que um grupo só é executado se o anterior tiver sido concluído com sucesso.

Fase 03: Construção do Data Warehouse

De acordo com o item (a) da Figura 5, é fornecida a descrição da arquitetura do banco de dados intitulado "saude_mental". Foi adotada a abordagem de criação de dois esquemas para a segregação dos ambientes de *stage* e DW. Os itens (b) e (c) da Figura 5 oferecem uma visão da estrutura interna, contendo todas as tabelas da *staging area* e do DW, respectivamente.



Fonte: Autores.

Fase 04: Análise de dados via Dashboards

Após o preenchimento do DW, procedeu-se à etapa subsequente, que consistiu na criação dos *dashboards* no *Tableau*. A referida ferramenta de visualização estabelece uma conexão direta com o banco de dados, permitindo a configuração da associação entre as dimensões e os fatos. Desta forma, foram desenvolvidos painéis de controle para este estudo, e neste trabalho são apresentados os mais relevantes, os demais podem ser vistos no repositório do *Tableau Public*.

A Figura 6 apresenta uma visão panorâmica do número de hospitalizações e óbitos decorrentes de causas mentais ao longo das competências, compreendendo o período de janeiro de 2010 a dezembro de 2021. A linha de cor preta representa a contagem dos óbitos, enquanto a linha de cor ciano indica a contagem dos casos de hospitalização. Adicionalmente, o *dashboard* exibe o valor total despendido com hospitalizações relacionadas à saúde mental durante o período analisado.

Observa-se que, durante os anos de 2010 e 2021 foram gastos R\$ 53,7 milhões com internações relacionadas a saúde mental, no estado do Espírito Santo. Nesse mesmo período foi possível observar uma redução do número de internações decorrentes de transtornos mentais e comportamentais no estado.

Figura 6. Índices de hospitalizações e mortalidade associados à saúde mental no estado do Espírito Santo



Fonte: Autores.

Nos dashboards apresentados nas Figuras 7 e 8, é possível observar uma análise mais aprofundada em relação aos óbitos. Essa análise contempla a visualização da evolução do número de óbitos ao longo dos anos, as principais causas dos óbitos, a distribuição dos óbitos de acordo com a raça e cor, o município de ocorrência, a faixa etária e o sexo do falecido.

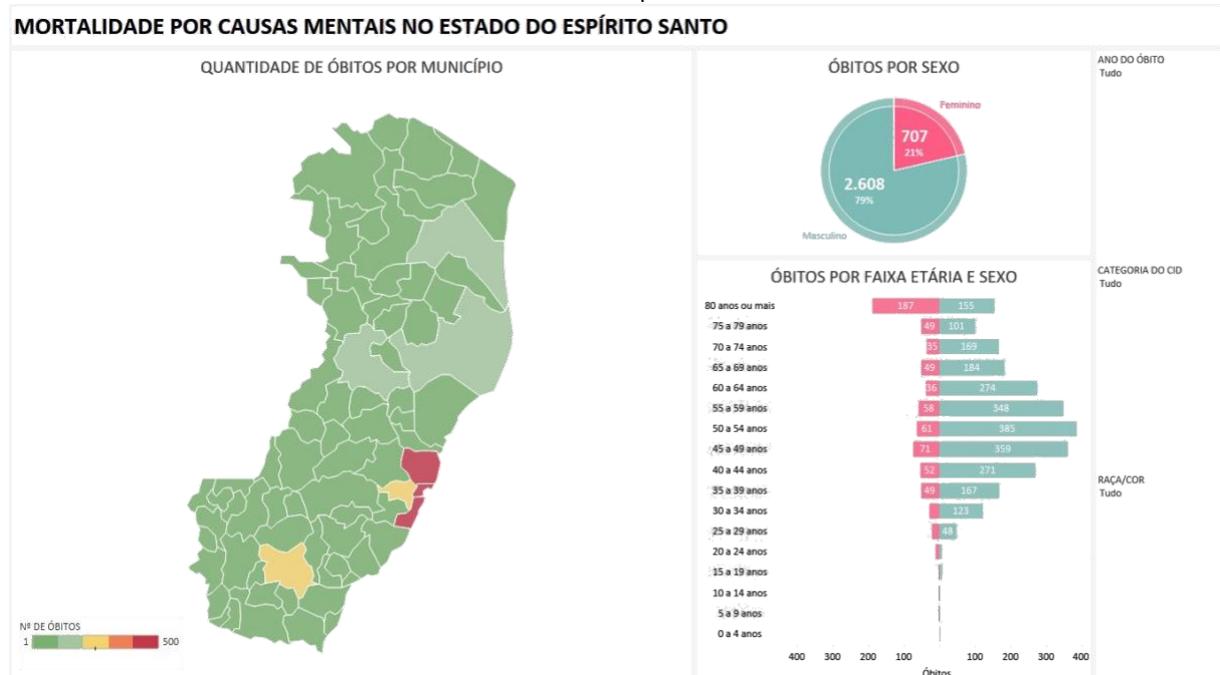
Figura 7. Painel contendo uma visão geral dos dados referentes aos óbitos decorrentes de causas mentais no estado do Espírito Santo



Fonte: Autores.

Foi possível observar um aumento superior a 100% nos óbitos relacionados a transtornos mentais e comportamentais no estado, após o início da pandemia por covid-19 em 2020. O aumento observado foi de 48,3% de 2019 para 2020 e de 61,1% de 2020 para 2021.

Figura 8. Painel contendo uma visão geográfica dos dados referentes aos óbitos decorrentes de causas mentais no estado do Espírito Santo



As principais causas de mortalidade por transtornos mentais e comportamentais foram álcool e fumo. Além disso, os óbitos foram mais frequentes entre os indivíduos pardos, do sexo masculino e na faixa etária de 45 a 59 anos. Por fim, a maior parte das mortes foi registrada em municípios da região metropolitana de Vitória (Vitória, Serra, Vila Velha e Cariacica), onde se concentra metade da população do estado, e no município de Cachoeiro de Itapemirim, o sexto maior em população do estado.

A partir desses resultados, destaca-se que o presente estudo escolheu empregar metodologias de BI com o objetivo de desenvolver um DW que viabilize a análise de informações advindas de instituições de saúde pública, especialmente relacionadas à saúde mental no estado do Espírito Santo. Contudo, a análise futuramente será estendida em abrangência nacional e incluir a investigação de outras bases de dados e indicadores em saúde.

Durante o período analisado, constatou-se um relevante aumento no número de óbitos associados a transtornos mentais e comportamentais, especialmente durante a pandemia do covid-19 (Santomauro et al, 2021; Kupcova, 2022). Essa tendência alarmante sugere que a crise sanitária exerceu um impacto significativo na saúde mental da população, resultando em consequências trágicas, principalmente em municípios da Grande Vitória e Cachoeiro de Itapemirim, que apresentaram número de mortes mais elevados.

Além disso, observou-se uma predominância do sexo masculino nas estatísticas de óbitos, levantando questionamentos sobre possíveis fatores sociais e de saúde que podem estar

influenciando essa disparidade de sexo e/ou gênero em relação às questões de saúde mental (Millwood, 2023). Esses achados ressaltam a importância de estratégias preventivas e intervenções adequadas para atender às necessidades específicas das pessoas em sofrimento psicológico, especialmente em momentos desafiadores como os vivenciados durante a pandemia.

Outro aspecto alarmante observado durante a análise foi o grande número de óbitos por saúde mental relacionados ao abuso de álcool, fumo e drogas ilícitas. O consumo excessivo dessas substâncias tem sido identificado como um fator significativo no agravamento de quadros psicológicos e no aumento das taxas de mortalidade e hospitalizações (Millwood, 2023; Topiwala, 2022, Evangelou, 2021). Essa correlação entre o uso abusivo de substâncias e as complicações de saúde mental evidencia a necessidade urgente de políticas de prevenção e tratamento efetivas voltadas para o enfrentamento dos vícios e para a promoção de uma abordagem holística da saúde mental. Investir em programas de conscientização, tratamento e suporte adequados pode desempenhar um papel crucial na redução desses impactos negativos e na melhoria da qualidade de vida da população afetada.

Além disso, estudos recentes indicam que não há um nível seguro para o consumo de álcool, sendo que a sua ingestão está associada a maior risco no desenvolvimento e mortalidade devido ao câncer, doenças cardiovasculares e doenças no fígado (Millwood, 2023; Evangelou, 2021; Mehta & Sheron, 2019).

Em suma, o presente estudo foi capaz de fornecer uma visão abrangente sobre a situação da saúde mental no estado do Espírito Santo, por meio da aplicação de técnicas de BI e análise de dados provenientes de instituições de saúde pública. A análise realizada permitiu identificar padrões, tendências e áreas de maior demanda nessa área, oferecendo uma base sólida para a tomada de decisões embasadas em evidências. Portanto, os resultados obtidos neste estudo têm o potencial de auxiliar gestores e profissionais de saúde na implementação de políticas e estratégias direcionadas à promoção da saúde mental, contribuindo assim para a melhoria dos serviços de saúde pública e o bem-estar da população do Espírito Santo.

O estudo apresenta diversas potencialidades importantes. Primeiramente, ele permitiu a criação de uma base nacional de dados composta por indicadores hospitalares e de mortalidade, servindo como um recurso centralizado e abrangente para a análise de saúde. Além disso, essa base de dados será utilizada para a prospecção de novos estudos em âmbito estadual e nacional, abordando aspectos relevantes para a saúde coletiva e comunitária. Portanto, este é o primeiro estudo que subsidiará outros a partir dessa mesma base de dados, a qual será atualizada regularmente. Outra potencialidade significativa é a possibilidade de uma avaliação permanente dos indicadores de saúde, fornecendo dados atualizados para análise contínua em saúde. Por fim, a partir da base de dados criada, será possível a criação de dashboards informativos que facilitarão o compartilhamento de informações de saúde com a comunidade em geral, promovendo a transparência e o engajamento da população com questões de saúde.

CONSIDERAÇÕES FINAIS

O BI desempenha um papel fundamental na saúde pública, fornecendo insights valiosos e informações estratégicas para melhorar a tomada de decisões e a eficiência dos serviços de saúde. Ao coletar, analisar e visualizar dados de forma automatizada, o BI permite identificar padrões, tendências e áreas de risco na saúde pública. Isso auxilia os gestores na alocação adequada de recursos, no planejamento de campanhas de prevenção e no monitoramento de doenças e epidemias, contribuindo para a prevenção e controle de doenças, aprimoramento dos serviços de saúde e a promoção do bem-estar da população. Nesse sentido, foi possível identificar aumento alarmante na mortalidade por transtornos mentais e comportamentais durante a pandemia por covid-19, principalmente associada ao consumo de álcool e tabaco.

AGRADECIMENTOS

Este estudo foi financiado pela Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES) sob Termo de Outorga 953/2022 e Processo 2022-8N96B.

REFERÊNCIAS

- Ali, M. S., Amorim, L. D. A. F., Werneck, G. L., Bittencourt, S. A., Hartz, Z., Barreto, M. L., ... & Oliveira, C. S. (2019). Administrative data linkage in Brazil: Potentials for health technology assessment. *Frontiers in Pharmacology*, 10. <https://doi.org/10.3389/fphar.2019.00984>
- Batko, K. & Ślęzak, A. (2022). The use of big data analytics in healthcare. *Journal of Big Data*, 9(1). <https://doi.org/10.1186/s40537-021-00553-4>
- Cheng, K. Y., Pazmino, S., & Schreiweis, B. (2022). ETL processes for integrating healthcare data - tools and architecture patterns. *Studies in Health Technology and Informatics*, 299, 151–156. <https://doi.org/10.3233/SHTI220974>
- Chiavegatto Filho, A. D. P. (2015). Uso de big data em saúde no Brasil: Perspectivas para um futuro próximo. *Epidemiologia e Serviços de Saúde*, 24(2), 325–332. <https://doi.org/10.5123/S1679-49742015000200015>
- Coelho Neto, G. C. & Chioro, A. (2021). Afinal, quantos Sistemas de Informação em Saúde de base nacional existem no Brasil? *Cadernos de Saúde Pública*, 37. <https://doi.org/10.1590/0102-311X00182119>
- Coelho, F. C., Araújo, W., Oliveira, M., Oliveira, L., Lima, A., & Silva, G. (2021). AlertaDengue/PySUS: Vaccine. Zenodo. <https://doi.org/10.5281/zenodo.4883502>
- Coeli, C. M. (2022). Ciência de dados populacionais. *Epidemiologia e Serviços de Saúde*, 31(3). <https://doi.org/10.1590/S2237-96222022000300001>
- Cozzoli, N., Minicucci, M. F., Siciliano, C. B., Celi, L. A., & Iadanza, E. (2022). How can big data analytics be used for healthcare organization management? Literary framework and future research from a systematic review. *BMC Health Services Research*, 22(1). <https://doi.org/10.1186/s12913-022-08167-z>
- Daniel, V. M., Pereira, G. V., & Macadar, M. A. (2014). Perspectiva Institucional dos Sistemas de Informação em Saúde em Dois Estados Brasileiros. *Revista de Administração Contemporânea*, 18, 650–669. <https://doi.org/10.1590/1982-7849rac20141709>
- Evangelou, E., Gill, D., Brown, M. R., Elliott, P., Tzoulaki, I., Murphy, N., ... & Holmes, M. V. (2021). Alcohol consumption in the general population is associated with structural changes in multiple organ systems. *eLife*, 10, e65325. <https://doi.org/10.7554/eLife.65325>
- Faria, R. J., Marques, C. D. L., Duarte, A. L. B. P., & Ribeiro, C. A. (2022). Conventional synthetic disease-modifying anti-rheumatic drugs for psoriatic arthritis: Findings and implications from a patient-centered longitudinal study in Brazil. *Frontiers in Pharmacology*, 13. <https://doi.org/10.3389/fphar.2022.878972>
- Fernandes, F. T. & Chiavegatto Filho, A. D. P. (2019). Perspectivas do uso de mineração de dados e aprendizado de máquina em saúde e segurança no trabalho. *Revista Brasileira de Saúde Ocupacional*, 44. <https://doi.org/10.1590/2317-6369000019418>
- FIOCRUZ – Fundação Oswaldo Cruz. (2023). Plataforma de Ciência de Dados aplicada à Saúde (PCDaS). Laboratório de Informação em Saúde (Lis). Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (Icict). Recuperado de <https://pcdas.icict.fiocruz.br>. <https://doi.org/10.7303/syn25882127>
- Gaioti, C. S. (2024). Saúde mental no ES. Tableau.com. Recuperado de <https://public.tableau.com/app/profile/cleiton.schwambach.gaioti/viz/SadeMentalnoES/HospitalizaesebitosporSadeMentalnoES?publish=yes>
- Heck, E. V. (2019). Big data and disruptions in business models. *RAE: Revista de Administração de Empresas*, 59(6), 430–432. <https://doi.org/10.1590/S0034-759020190608>
- Kimball, R. & Ross, M. (2013). The data warehouse toolkit: The definitive guide to dimensional modeling (3rd ed.). John Wiley & Sons, Inc.

- Kupcova, I., Fauskanger, J., Cederberg, S. A., Szabo, M., Bygstad, B., & Faxvaag, A. (2023). Effects of the COVID-19 pandemic on mental health, anxiety, and depression. *BMC Psychology*, 11(1). <https://doi.org/10.1186/s40359-023-01130-5>
- Maior, M. da C. L. S., Osorio-de-Castro, C. G. S., & Andrade, C. L. T. de. (2020). Demografia, óbitos e indicadores de agravamento nas internações por intoxicações medicamentosas entre menores de 5 anos no Brasil. *Revista Brasileira de Epidemiologia*, 23. <https://doi.org/10.1590/1980-549720200016>
- Medeiros, M. M., Maçada, A. C. G., & Hoppen, N. (2021). The role of big data stewardship and analytics as enablers of corporate performance management. *RAM. Revista de Administração Mackenzie*, 22(6). <https://doi.org/10.1590/1678-6971/eRAMD210063>
- Mehta, G. & Sheron, N. (2019). No safe level of alcohol consumption – implications for global health. *Journal of Hepatology*, 70(4), 587–589. <https://doi.org/10.1016/j.jhep.2018.12.021>
- Millwood, I. Y., Walters, R. G., Fillmore, T., Cheng, Y., Lacey, B., Lu, Y., ... & Bennett, D. A. (2023). Alcohol intake and cause-specific mortality: Conventional and genetic evidence in a prospective cohort study of 512,000 adults in China. *The Lancet Public Health*, 8(12), e956–e967. [https://doi.org/10.1016/S2468-2667\(23\)00217-7](https://doi.org/10.1016/S2468-2667(23)00217-7)
- Ognjanovic, I. (2020). Healthcare data analytics. *Studies in Health Technology and Informatics*, 274, 122–135. <https://doi.org/10.3233/SHTI200672>
- Oliveira, H. F. de, Sampaio, A. L. L., & Oliveira, C. A. C. P. de. (2011). DATASUS as an instrument for developing otologic public health policies. *Brazilian Journal of Otorhinolaryngology*, 77(3), 369–372. <https://doi.org/10.1590/S1808-86942011000300016>
- Paim, J., Travassos, C., Almeida, C., Bahia, L., & Macinko, J. (2011). The Brazilian health system: History, advances, and challenges. *The Lancet*, 377(9779), 1778–1797. [https://doi.org/10.1016/S0140-6736\(11\)60054-8](https://doi.org/10.1016/S0140-6736(11)60054-8)
- Santomaura, D. F., Santomauro, D. F., Mantilla Herrera, A. M., Shadid, J., Zheng, P., Ashbaugh, C., ... & Ferrari, A. J. (2021). Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *The Lancet*, 398(10312), 1700–1712. [https://doi.org/10.1016/S0140-6736\(21\)02143-7](https://doi.org/10.1016/S0140-6736(21)02143-7)
- Topiwala, A., Ebmeier, K. P., Jardine, R., Nichols, T. E., Ward, E. V., & Kivimaki, M. (2022). Alcohol consumption and MRI markers of brain structure and function: Cohort study of 25,378 UK Biobank participants. *NeuroImage: Clinical*, 35, 103066. <https://doi.org/10.1016/j.nicl.2022.103066>
- Viana, S. A. P., Tanaka, S. O., & Santos, T. M. (2022). Prontuário Eletrônico do Paciente em um hospital universitário: Análise dos registros de atendimento. *Revista Brasileira de Enfermagem*, 75. <https://doi.org/10.1590/0034-7167-2022-0031>
- Williams, J. (2018). Data management challenges in healthcare: Information quality and effective management. *Journal of Data and Information Quality*, 10(1), 1–8. <https://doi.org/10.1145/3183330>
- Woolhandler, S. & Himmelstein, D. U. (2017). Single-payer reform: The only way to fulfill the president's pledge of more coverage, better benefits, and lower costs. *Annals of Internal Medicine*, 166(8), 587–588. <https://doi.org/10.7326/M17-0302>