UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
CENTRO DE CIÊNCIAS EXATAS, NATURAIS E DA SAÚDE

DEPARTAMENTO DE COMPUTAÇÃO
CIÊNCIA DA COMPUTAÇÃO

CLEITON SCHWAMBACH GAIOTI

IMPLEMENTAÇÃO DE UM DATA WAREHOUSE PARA ESTUDO DE INDICADORES RELACIONADOS A SAÚDE MENTAL NO ESTADO DO ESPÍRITO SANTO

ALEGRE - ES
JULHO 2023

IMPLEMENTAÇÃO DE UM DATA WAREHOUSE PARA ESTUDO DE INDICADORES RELACIONADOS A SAÚDE MENTAL NO ESTADO DO ESPÍRITO SANTO

Trabalho de conclusão de curso apresentado ao Departamento de Computação do Centro de Ciências Exatas, Naturais e da Saúde da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

por

CLEITON SCHWAMBACH GAIOTI

Orientador
ANTONIO ALMEIDA DE BARROS JUNIOR

Universidade Federal do Espírito Santo

ALEGRE - ES

JULHO 2023

CLEITON SCHWAMBACH GAIOTI

IMPLEMENTAÇÃO DE UM DATA WAREHOUSE PARA ESTUDO DE INDICADORES RELACIONADOS A SAÚDE MENTAL NO ESTADO DO ESPÍRITO SANTO

Trabalho de conclusão de curso apresentado ao Departamento de Computação do Centro de Ciências Exatas, Naturais e da Saúde da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação.

Aprovado em ____ de julho de 2023.

COMISSÃO EXAMINADORA

Documento assinado digitalmente
ANTONIO ALMEIDA DE BARROS JUNIOR
Data: 27/07/2023 21:15:12-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Antonio Almeida de Barros Junior
Universidade Federal do Espírito Santo
Orientador

Documento assinado digitalmente
EDMAR HELL KAMPKE
Data: 27/07/2023 00:02:29-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Edmar Hell Kampke
Universidade Federal do Espírito Santo

Documento assinado digitalmente
MICHAEL RUBERSON RIBEIRO DA SILVA
Data: 26/07/2023 20:12:32-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Michael Ruberson Ribeiro da Silva
Universidade Federal do Espírito Santo

AGRADECIMENTOS

Agradeço a Deus por me dar sabedoria para alcançar os meus objetivos.

Aos meus pais José Jorge e Solanja, pela educação, simplicidade, amor e apoio em todas as fases da minha vida. Aos meus irmãos Kaique e Thays, por estarem sempre ao meu lado. Amo vocês!

À minha noiva Paola, por todo amor, companheirismo e incentivo. Você me deu forças para continuar. Amo você!

Ao meu orientador Antônio, pelos ensinamentos, esclarecimento de dúvidas e suporte necessário durante toda a graduação e na composição deste trabalho.

Aos meus professores pelos ensinamentos.

Aos meus amigos, pelo apoio e companheirismo. Em especial Daniel, Mário, Luann e Andrico pela amizade desde o início da graduação.

Ao meu companheiro canino Timão, pela companhia e amor nos momentos em que estive sozinho.

À Universidade Federal do Espírito Santo, pela oportunidade.

"Tudo o que temos de decidir é o que fazer com o tempo que nos é dado."

Gandalf

SUMÁRIO

1 Introdução **9**

1.1 Objetivos 10

1.1.1 Objetivos Gerais 10

1.1.2 Objetivos específicos 11

2 Revisão de Literatura **12**

2.1 Trabalho de Housbane et al. (2020) 14

2.2 Trabalho de Rowntree, McCarthy e Feeney (2020) 17

2.3 Trabalho de Wetzler et al. (2020) 19

2.4 Trabalho de Roggenkamp et al. (2018) 21

3 Metodologia **24**

3.1 Ferramentas utilizadas 26

3.1.1 *Apache Airflow* 26

3.1.2 *PostgreSQL* 26

3.1.3 *Python* 27

3.1.4 *Tableau* 27

3.1.5 Base de dados de origem 27

4 Resultados Obtidos **29**

4.1 Levantamento dos Indicadores 29

4.2 Estrutura do Projeto e Processo de ETL 30

4.2.1 Pasta dags 31

4.2.2 Pasta docs 31
4.2.3 Pasta Tableau 31
4.2.4 *Airflow* 32
4.3 *Data Warehouse* 33
4.4 *Dashboards* 36

5 Conclusões 39

Referências 41

LISTA DE FIGURAS

Figura 1 Código escrito na linguagem R para análise das bases extraídas. . . . 13

Figura 2 Processo de ETL proposto por Housbane et al. (2020), usando a ferramenta *Pentaho* para recuperar dados de emergência e armazená-los nos bancos de dados do DW. Fonte: Housbane et al. (2020). 15

Figura 3 Código escrito na linguagem *Python* para criar e gerenciar o *dashboard*. Fonte: Housbane et al. (2020). 16

Figura 4 Página da *Web* com os painéis propostos por Housbane et al. (2020). Fonte: Housbane et al. (2020). 17

Figura 5 Características dos pacientes do *Medicaid*, pelo ambiente onde receberam serviços de saúde mental. Fonte: Wetzler et al. (2020). 19

Figura 6 Uso do serviço (número médio de visitas) entre pacientes do *Medicaid* com doença mental grave, pelo local onde receberam serviços de saúde mental e por tipo de cobertura do *Medicaid*. Fonte: Wetzler et al. (2020). 20

Figura 7 Análise de regressão de variáveis como preditores de internações entre pacientes do *Medicaid* que recebem serviços de saúde mental. Fonte: Wetzler et al. (2020). 21

Figura 8 Arquitetura do DW. 24

Figura 9 Estrutura do Projeto hospedado no GitHub. 30

Figura 10 Fluxo de execução do processo de ETL no *Airflow*. 32

Figura 11 Modelo Dimensional do cubo de dados. 33

Figura 12 Estrutura do banco de dados, da *staging area* e do DW, respectivamente. 35

Figura 13 Índices de hospitalizações e mortalidade associados à saúde mental no estado do Espírito Santo. 36

Figura 14 Painel contendo uma visão geral dos dados referentes aos óbitos decorrentes de causas mentais no estado do Espírito Santo. 37

Figura 15 Painel contendo uma visão geográfica dos dados referentes aos óbitos decorrentes de causas mentais no estado do Espírito Santo. 37

Figura 16 Painel contendo uma visão panorâmica dos dados relativos às hospitalizações resultantes de transtornos mentais no estado do Espírito Santo. 38

Figura 17 Painel exibindo uma representação geográfica dos dados relativos às hospitalizações resultantes de causas mentais no território do estado do Espírito Santo. 38

LISTA DE TABELAS

Tabela 1 Resultados da busca na plataforma *Scopus*. 13

Tabela 2 Resultados da busca na plataforma *Web of Science*. 13

Tabela 3 Número de prescrições para cada um dos medicamentos por ano. . 18

LISTA DE SIGLAS

OMS	Organização Mundial da Saúde
SUS	Sistema Único de Saúde
DATASUS	Departamento de Informática do Sistema Único de Saúde
SIM	Sistema de Informação sobre Mortalidade
SIH	Sistema de Informações Hospitalares
AIH	Autorização de Internação Hospitalar
BI	*Business Intelligence*
DW	*Data Warehouse*
ETL	*Extract, Treat e Load*
ORDBMS	*Object-Relational Database Management System*
OLAP	*Online Analytical Processing*
DAG	*Directed Acyclic Graph*
PCDaS	Plataforma de Ciência de Dados aplicada à Saúde
CID	Classificação Internacional de Doenças
SCD	Slowly Changing Dimensions

RESUMO

Os cuidados relacionados à saúde mental têm se mostrado cada vez mais necessários na sociedade contemporânea. Com rotinas cada vez mais movimentadas e estressantes, as pessoas tendem a desenvolver transtornos mentais, devido a grande pressão imposta no dia a dia. De acordo com a Organização Mundial da Saúde (OMS), somente no Brasil, 86% da população sofre com algum tipo de transtorno mental, envolvendo depressão, ansiedade, fobias e transtornos de personalidade. Por isso, as internações psiquiátricas vêm representando um dos maiores gastos com a rede de serviços hospitalares do Sistema Único de Saúde (SUS). Para que sejam tomadas medidas eficazes para combater esse tipo de transtorno, é necessário compreender o fenômeno a fim de identificar suas causas. Desta forma, é importante que seja feita uma análise dos dados coletados nas unidades de atendimento do SUS. Porém, isso pode significar um volume de dados muito elevado, sendo necessário um método para que todas essas informações sejam apresentadas de forma clara e objetiva e em um curto período de tempo. Assim, por meio da aplicação da metodologia proposta por Kimball e Ross (2013), o presente trabalho teve como objetivo a criação de um *Data Mart* especializado para a análise de indicadores relacionados à saúde mental, com base em dados previamente coletados pelo Sistema Único de Saúde (SUS). O estudo em questão proporcionou uma análise abrangente e aprofundada acerca da condição da saúde mental no estado do Espírito Santo, utilizando técnicas de *Business Intelligence* e análise de dados provenientes de instituições de saúde pública. A análise conduzida possibilitou a identificação de padrões, tendências e regiões com maior demanda na área de saúde mental, fornecendo uma base sólida para a tomada de decisões embasadas em evidências. Portanto, os resultados obtidos neste estudo possuem o potencial de apoiar gestores e profissionais da área da saúde na implementação de políticas e estratégias voltadas para a promoção da saúde mental, contribuindo, dessa forma, para aprimorar os serviços de saúde pública e o bem-estar da população do Espírito Santo.

Palavras-chave: Saúde Mental; *Business Intelligence*; *Data Warehouse*.

1 INTRODUÇÃO

A sociedade brasileira vivencia, atualmente, uma transformação no modelo de assistência ao paciente com transtorno psiquiátrico conhecida nacionalmente, desde a década de 80, como Reforma Psiquiátrica Brasileira. Esta se deve ao conjunto de fatores e atores que buscam substituir o modelo asilar, segregador, excludente, reducionista e tutelar que tem como centro de atendimento o hospitalar psiquiátrico (SOUZA et al., 2007).

Porém, no Brasil, como em muitos outros países, esta rede de serviços ainda está em desenvolvimento e carece de ampliação da implantação de infra-estrutura extra-hospitalar mais próxima ao cotidiano de seus clientes (JORGE; FRANçA, 2001). De acordo com Cardoso e Galera (2011), a demanda de cuidado em saúde mental não se restringe apenas a minimizar riscos de internação ou controlar sintomas. Atualmente, o cuidado envolve também questões pessoais, sociais, emocionais e financeiras, relacionadas à convivência com o adoecimento mental.

No Brasil, as internações psiquiátricas vêm representando um dos maiores gastos com a rede de serviços hospitalares do SUS. Os leitos, tanto privados quanto públicos, apresentam internações duradouras, chegando a alcançar, em média, o teto máximo permitido (BRêDA; AUGUSTO, 2001). Atualmente, o teto máximo permitido é de noventa dias de internação.

Segundo Rodrigues e Casagrande (2001), esse cenário se apresenta como um grande desafio aos pesquisadores, gestores, órgãos e instituições formadoras e usuários do serviço de saúde do estado e do município. Cabe a eles, portanto, de forma democrática e pactuada, articular saberes e desenvolver estratégias de mudanças no modelo assistencial e na legislação em saúde mental, para que, gradativamente, observem-se mudanças culturais secularmente enraizadas na sociedade.

Para o desenvolvimento de uma estratégia eficaz, é imprescindível que as decisões sejam tomadas com base nas informações coletadas pelas diversas instituições. Sendo

assim necessário a implantação de um sistema que seja capaz de integrar e representar diferentes fontes de dados.

O Departamento de Informática do Sistema Único de Saúde (DATASUS), (DATASUS. . . ,) desempenha um papel fundamental na gestão e disponibilização de informações sobre a saúde pública no Brasil. Entre seus principais sistemas, destacam-se o Sistema de Informação sobre Mortalidade (SIM), que registra e analisa dados relacionados a óbitos no país, fornecendo subsídios para a compreensão de padrões de mortalidade e o desenvolvimento de políticas de saúde eficazes. Além disso, o Sistema de Informações Hospitalares (SIH) é responsável por coletar informações sobre os atendimentos realizados nos hospitais do Sistema Único de Saúde, permitindo o monitoramento da demanda por serviços hospitalares e a avaliação da qualidade do atendimento. Já a Autorização de Internação Hospitalar AIH é um componente do SIH que autoriza e registra as internações hospitalares realizadas pelo SUS, possibilitando a rastreabilidade dos atendimentos e o planejamento estratégico das redes de saúde. Esses sistemas fornecem subsídios valiosos para a formulação de políticas públicas e o aprimoramento contínuo do sistema de saúde brasileiro.

Segundo Kimball e Ross (2013), a construção de sistemas de Business Intelligence (BI) se tornou tendência nas últimas décadas, devido ao seu propósito de apresentar informações aos usuários de negócio de forma clara e intuitiva. Um ambiente de BI bem estruturado permite que os gestores tomem decisões informadas e inteligentes sobre o funcionamento de sua organização, ocasionando em processos melhores e mais eficientes.

Utilizando a metodologia sugerida por Kimball e Ross (2013), o presente estudo teve como objetivo desenvolver um *Data Mart* especializado voltado para a análise de indicadores referentes à saúde mental, a partir de dados previamente coletados pelo SUS.

1.1 Objetivos

1.1.1 Objetivos Gerais

O objetivo geral deste trabalho é construir um *Data Warehouse* (DW) usando técnicas de BI que possibilite a análise de dados referentes a saúde mental coletados de serviços públicos de saúde. Assim, será criado um ambiente de DW/BI, baseado na

arquitetura proposta por Kimball e Ross (2013), que possua um fluxo *Extract, Treat e Load* (ETL) desenvolvido e gerenciado por ferramentas *open-source*.

1.1.2 Objetivos específicos

1. Implementar um DW que atenda aos indicadores levantados junto aos *stakeholders*.
2. Desenvolver um ambiente de ETL automatizado que alimente o DW com os dados da origem.
3. Construir *dashboards* que correspondam aos indicadores levantados junto dos *stakeholders* e realizar analises dos resultados obtidos.

2 REVISÃO DE LITERATURA

A revisão da literatura foi realizada por meio do Portal de Periódicos da CAPES (PORTAL. . . ,) utilizando o Acesso CAFe, que permite acesso a conteúdo de diversas instituições. Para tanto, foi realizada uma pesquisa nos repositórios *Scopus* (SCOPUS,) e *Web of Science* (WEB. . . ,), por trabalhos relacionados ao uso de BI em análises sobre saúde mental. Nas buscas foram usadas as seguintes palavras-chave, em português: “saúde mental”, e em inglês: “*mental health*”, “*business intelligence*” e “*data warehouse*”, assim como a combinação das mesmas. Algumas consultas resultaram em mais de 1000 documentos, porém tanto no *Scopus* quanto no *Web of Science*, limita-se o acesso para apenas 1000 documentos por busca. Logo foram selecionados somente os 1000 trabalhos com maior relevância, não sendo possível analisar todos os documentos das consultas que ultrapassaram esse limite.

Após a pesquisa, foram gerados arquivos com dados brutos dos trabalhos selecionados e foi desenvolvido um *script* na linguagem R (R. . . ,) utilizando a biblioteca *Bibliometrix* (BIBLIOMETRIX,), para análise das bases extraídas. O *script*, apresentado na Figura 1, carrega e converte os dados para o padrão Bibliometrix e em seguida cria uma planilha com as informações de todos os trabalhos. Foram encontrados um total de 1001 trabalhos entre os anos de 1990 e 2022, com 3257 autores envolvidos.

Posteriormente foram filtrados somente os trabalhos que possuem a combinação de duas ou mais palavras-chave, e após a leitura do título e resumo de cada um dos documentos, foram selecionados aqueles que possuem relação com o tema deste trabalho. A Tabela 1 e Tabela 2 apresentam os resultados das buscas realizadas no *Scopus* e *Web of Science*, respectivamente.

Como apresentado na Tabela 1, três trabalhos foram selecionados, sendo um deles a partir da combinação das palavras-chave *"mental health"* e *"business intelligence"*, os outros dois resultaram da combinação das palavras-chave *"mental health"* e *"data warehouse"*.

```r
revisao.R > ...
install.packages("readxl")
install.packages("bibliometrix")

library("openxlsx")
library("readxl")
library("bibliometrix")

mhbi <- convert2df("artigos/web_of_science/mental_health.bib", dbsource="wos", format="bibtex") # nolint

mhdw <- convert2df("artigos/web_of_science/mental_health_and_business_intelligence_and_data_warehouse.bib" # nolint
    , dbsource="wos", format="bibtex") # nolint

mhbidw <- convert2df("artigos/web_of_science/mental_health_and_business_intelligence_and_data_warehouse.bib" # nolint
    , dbsource="wos", format="bibtex") # nolint

df <- mergeDbSources(mhbi, mhdw, mhbidw, remove.duplicated = TRUE)

result <- biblioAnalysis(df)

df <- df[c("AU", "TI", "AB", "DE", "ID", "TC", "PY", "LA", "DT")]
write.xlsx(df, "referencias.xlsx")

summary(object = result)
```

Figura 1 – Código escrito na linguagem R para análise das bases extraídas.

Palavra-chave	**Resultados**	**Trabalhos selecionados**
"saúde mental"	1098	0
"mental health"	419262	0
"business intelligence"	9312	0
"data warehouse"	15101	0
"saúde mental" e *"business intelligence"*	0	0
"saúde mental" e *"data warehouse"*	0	0
"saúde mental" e *"business intelligence"* e *"data warehouse"*	0	0
"mental health" e *"business intelligence"*	7	1
"mental health" e *"data warehouse"*	85	2
"mental health" e *"business intelligence"* e *"data warehouse"*	0	0
Total	444865	3

Tabela 1 – Resultados da busca na plataforma *Scopus*.

Palavra-chave	**Resultados**	**Trabalhos selecionados**
"saúde mental"	7	0
"mental health"	292873	0
"business intelligence"	5401	0
"data warehouse"	6909	0
"saúde mental" e *"business intelligence"*	0	0
"saúde mental" e *"data warehouse"*	0	0
"saúde mental" e *"business intelligence"* e *"data warehouse"*	0	0
"mental health" e *"business intelligence"*	4	2
"mental health" e *"data warehouse"*	70	2
"mental health" e *"business intelligence"* e *"data warehouse"*	1	0
Total	444865	3

Tabela 2 – Resultados da busca na plataforma *Web of Science*.

Na Tabela 2, são apresentados quatro trabalhos, sendo dois deles a partir da combinação das palavras-chave *"mental health"* e *"business intelligence"*, os outros dois resultaram da combinação das palavras-chave *"mental health"* e *"data warehouse"*. Para a combinação *"mental health"*, *"business intelligence"* e *"data warehouse"* foi encontrado um trabalho, porém o mesmo não foi selecionado por não apresentar relação com o tema desta pesquisa. Dos quatro documentos selecionados, três já haviam sido selecionados anteriormente. Desta forma, quatro trabalhos foram escolhidos e serão apresentados a seguir.

2.1 Trabalho de Housbane et al. (2020)

A informação relevante em tempo real ajuda a orientar o processo de tomada de decisão, bem como a gestão e otimização dos processos de saúde. No entanto, soluções de BI que suportam a produção de informações para tomada de decisão, requerem investimentos que estão fora do alcance de estabelecimentos de saúde com recursos limitados, principalmente em países em desenvolvimento. O trabalho de Housbane et al. (2020) teve como objetivo projetar e implementar uma solução de sistema de monitoramento em saúde em tempo real, para gerenciar unidades de emergência de saúde de pequeno e médio porte.

Através do uso de ferramentas gratuitas de BI e linguagem de ciência de dados *Python* (PYTHON. . . ,), foi projetado um sistema de monitoramento em tempo real, implementado para explorar o sistema de prontuários eletrônicos de uma unidade de emergência de saúde mental universitária, e renderizar um *dashboard* eletrônico para apoiar profissionais de saúde na prática diária.

Para a recuperação e armazenamento dos dados de emergência foi utilizada a ferramenta *Pentaho Kettle* (HITACHI VANTARA,). Para a criação e gerenciamento dos *dashboards* foi utilizado o pacote *Voila* (TEAM,) em conjunto com o *software* gratuito *Jupyter Notebook* (JUPYTER,).

Foram utilizados dados do Centro Psiquiátrico Universitário em Casablanca, unidade de referência da região Oeste do Marrocos que recebe diariamente de 30 a 40 pacientes com comportamento suicida ou risco de agressão.

O sistema foi projetado para responder a três perguntas: quantos pacientes estão esperando, quem é o próximo paciente e como os pacientes foram distribuídos durante

o dia anterior. Foram criados três *dashboards* principais para monitorar o tempo de espera do paciente, acessar o resumo das anotações clínicas do próximo paciente em espera e obter informações sobre a atividade durante as últimas 24 horas.

Nas Figuras 2 e 3, são apresentados respectivamente a estrutura do processo de ETL usando a ferramenta de código aberto *Pentaho* e o código escrito na linguagem *Python* para criar e gerenciar o *dashboard*.

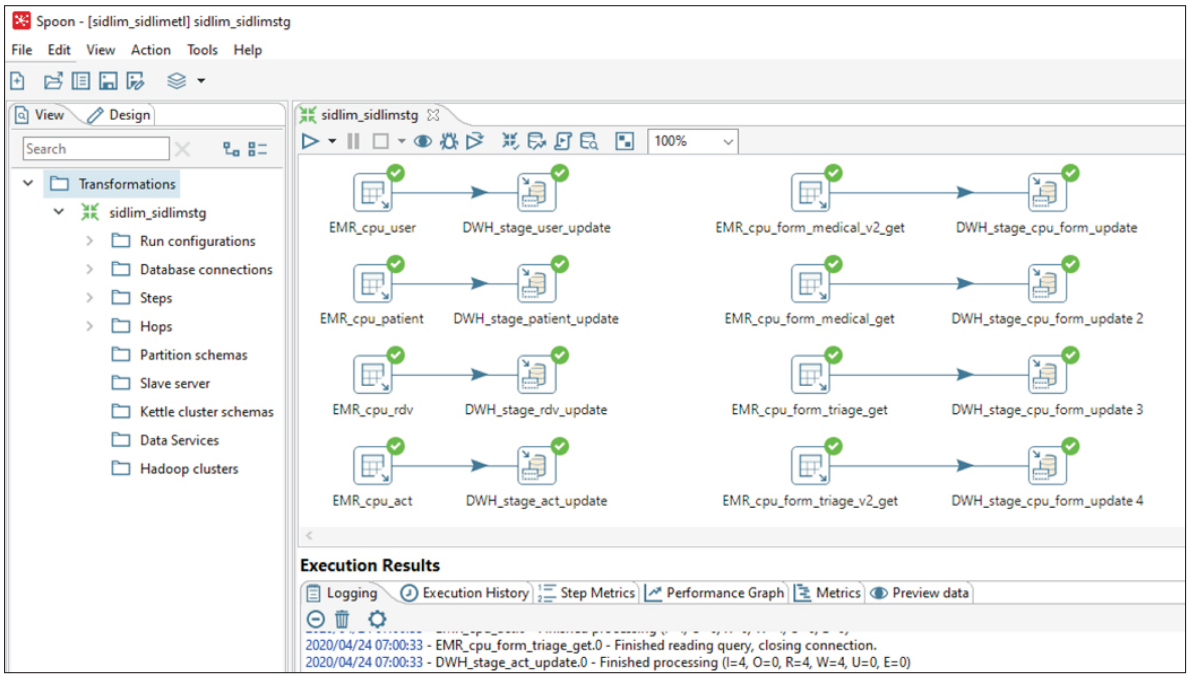

Figura 2 – Processo de ETL proposto por Housbane et al. (2020), usando a ferramenta *Pentaho* para recuperar dados de emergência e armazená-los nos bancos de dados do DW. Fonte: Housbane et al. (2020).

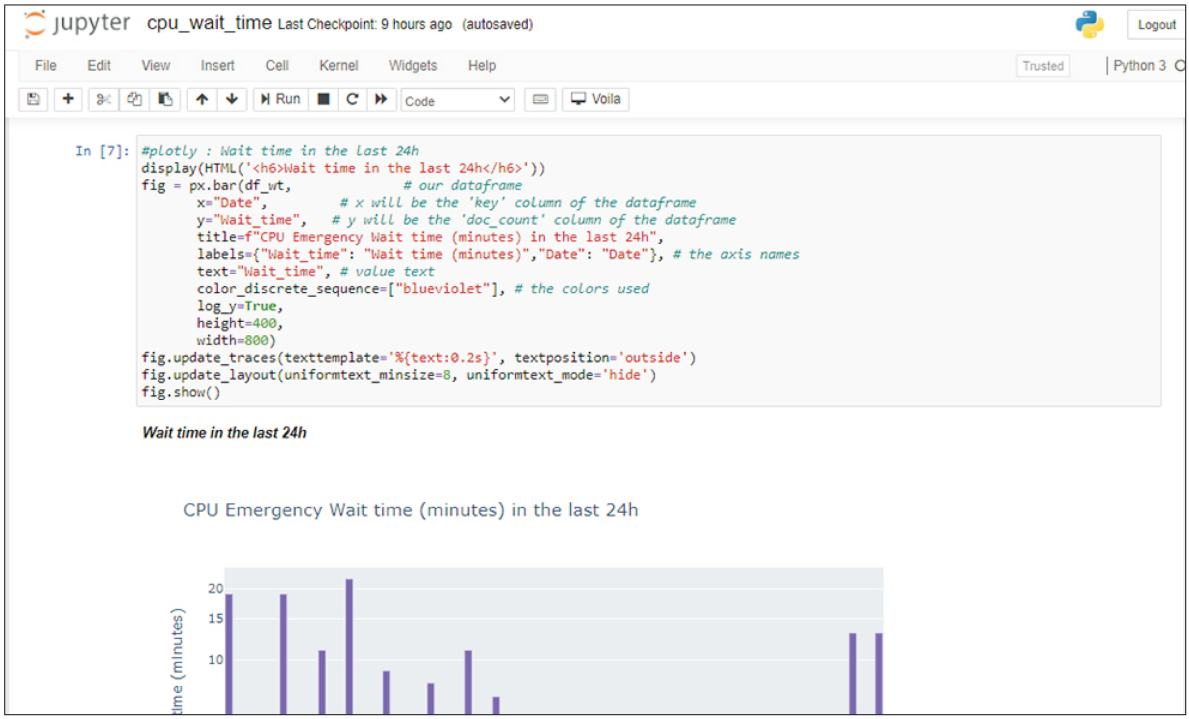

Figura 3 – Código escrito na linguagem *Python* para criar e gerenciar o *dashboard*. Fonte: Housbane et al. (2020).

Na Figura 4, é apresentado o *dashboard* contendo: (a) o tempo de espera na emergência do centro psiquiátrico, (b) um resumo clínico do próximo paciente e (c) uma análise do tempo de espera das últimas 24 horas.

Por fim, Housbane et al. (2020) afirmou que o sistema projetado pode servir como um modelo de suporte de monitoramento usando ferramentas de ciência de dados gratuitas e fáceis de usar, sendo uma boa alternativa para soluções de BI que reduzem drasticamente o custo. Ainda assim, a chave para o sucesso dos sistemas de tomada de decisão está baseada no investimento em recursos humanos, no treinamento de habilidades de BI, no aspecto organizacional do processo de tomada de decisão e na garantia da qualidade da produção de dados.

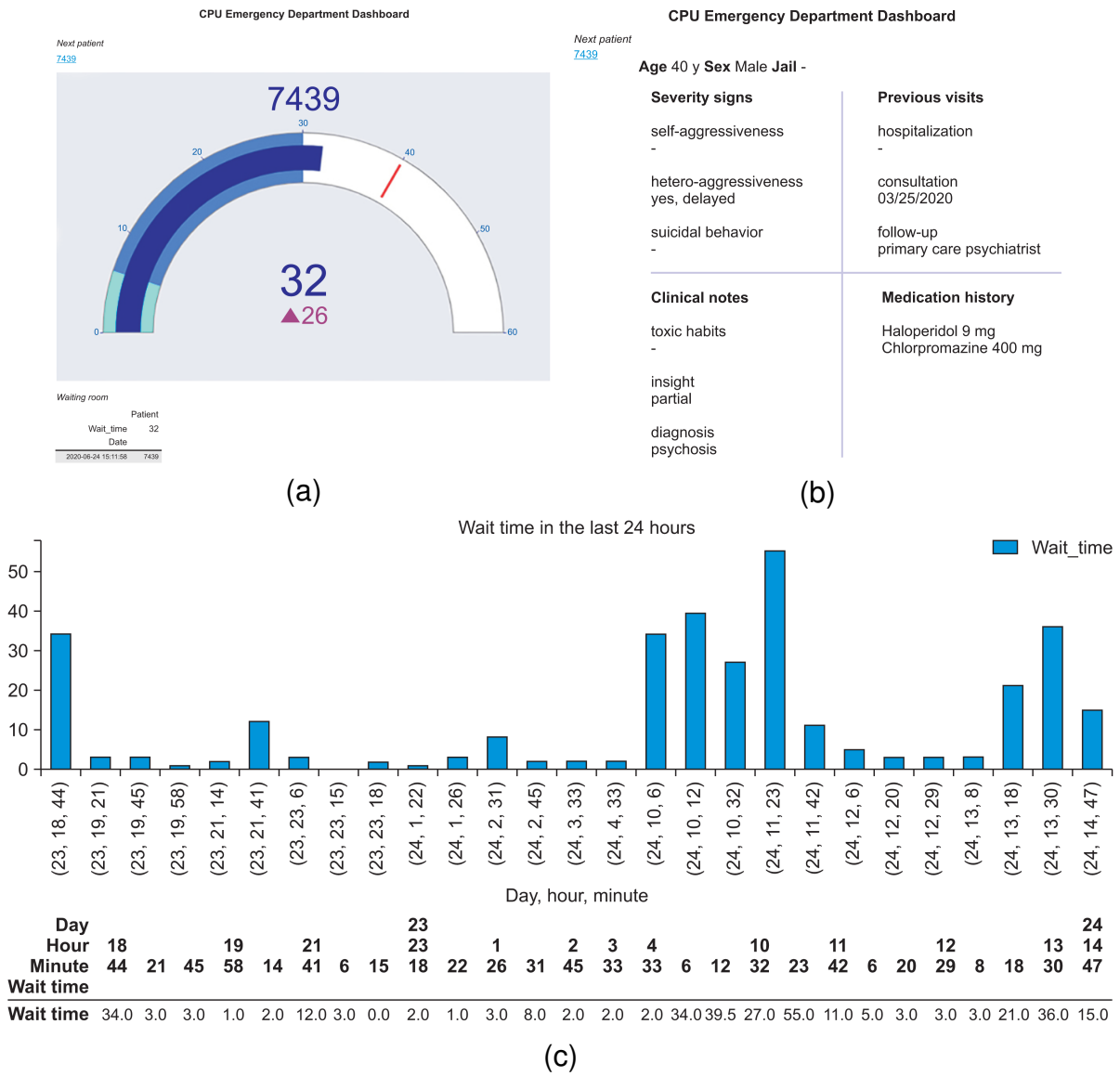

Figura 4 – Página da *Web* com os painéis propostos por Housbane et al. (2020). Fonte: Housbane et al. (2020).

2.2 Trabalho de Rowntree, McCarthy e Feeney (2020)

A medicação é um componente importante do tratamento da maioria das doenças mentais. Muita pouca informação está disponível sobre os medicamentos específicos que estão sendo prescritos pelos serviços comunitários de saúde mental e como isso mudou ao longo do tempo. O trabalho de Rowntree, McCarthy e Feeney (2020) tem como objetivo obter detalhes sobre medicamento psiquiátricos prescritos por um serviço de saúde mental da comunidade irlandesa.

Todas as prescrições do Serviço Comunitário de Saúde Mental Cluain Mhuire (CLUAIN. . . ,) tornaram-se eletrônicas em 2004. Com a utilização de um *software* de BI, foi

obtido detalhes de todos os medicamentos psiquiátricos prescritos de 2005 a 2016. Após isso, foi comparado o número de prescrições escritas nos primeiros 6 anos (2005–2010) com os 6 anos seguintes (2011–2016).

Foi utilizado o *software Oracle Business Intelligence* (Oracle, 2005) para o fornecimento do número de prescrições escritas para cada medicamento em cada ano de 2005 a 2016. Os dados foram analisados usando SPSS versão 22 (IBM Corp., 2013).

Entre os anos de 2005-2010 e 2011-2016, as prescrições de todos os medicamentos antipsicóticos, estabilizadores de humor, antidepressivos e sedativos caíram 7%, 25%, 25% e 48%, respectivamente. Enquanto o número de prescrições para outros medicamentos aumentou 804% no mesmo período.

A *Olanzapine* foi o medicamento mais comumente prescrito, mas seu uso diminuiu em um terço durante o período do estudo. A prescrição de *Clozapine*, *Quetiapine*, *Aripiprazole* e *Haloperidol* aumentou. As prescrições de estabilizadores de humor e antidepressivos caíram 25%. As prescrições de sedativos diminuíram quase 50%. O número absoluto de prescrições de *Mirtazapine* e *Pregabalin* foi pequeno, mas aumentou dramaticamente ao longo do período.

A Tabela 3 apresenta o número de prescrições para cada um dos medicamentos por ano.

	Olanzapine	**Quetiapine**	**Aripiprazole**	**Clozapine**	**Haloperidol**	**Mirtazapine**	**Total**
2005	1778	307	169	316	113	195	2878
2006	1924	381	182	471	128	379	3465
2007	2008	470	131	490	113	497	3709
2008	2160	390	171	660	178	386	3945
2009	2423	496	286	733	170	421	4529
2010	2280	828	197	654	259	433	4651
2011	1868	752	191	782	190	413	4196
2012	1783	680	238	900	178	427	4206
2013	1663	681	195	827	152	414	3932
2014	1512	774	171	773	186	380	3796
2015	1262	799	220	611	238	390	3520
2016	1378	841	240	651	230	404	3744
Total	22039	7399	2391	7868	2135	4739	46571

Tabela 3 – Número de prescrições para cada um dos medicamentos por ano.

Por fim, o estudo de Rowntree, McCarthy e Feeney (2020) chegou à conclusão que este serviço comunitário de saúde mental prescreveu menos da maioria dos medicamentos psiquiátricos em 2016 do que em 2005. Isso apesar de um aumento no

número de pacientes atendidos no mesmo período.

2.3 Trabalho de Wetzler et al. (2020)

O estudo de Wetzler et al. (2020) investigou o uso de serviços por indivíduos com doenças mentais graves e não graves recebendo cuidados em ambientes médicos e de saúde mental.

Foram usados dados de sinistros do *New York State Medicaid Data Warehouse*. Foi realizado uma análise dos dados de 8988 pacientes, com idades entre 18 e 65 anos, que receberam pelo menos um serviço de saúde mental em um centro médico acadêmico urbano durante o ano de 2017, em um ambiente de saúde mental (3000), médico (5342) ou ambos (646).

Nas Figuras 5 e 6 são apresentados respectivamente, uma tabela contendo as características dos pacientes e o uso do serviço entre pacientes com doença mental grave, pelo local onde receberam serviços de saúde mental e por tipo de cobertura do *Medicaid*.

	Overall (N=8,988)		Mental health clinic only (N=3,000)		Medical clinic only (N=5,342)			Mental health and medical clinic (N=646)		
Characteristic	N	%	N	%	N	%	p	N	%	p
CCI score (M±SD)[a]	1.81±2.48		1.85±2.20		1.79±2.65		.348	1.79±2.32		.536
Age (M±SD)	39.9±13.6		43.5±13.3		38.0±13.4		<.001	39.4±13.8		<.001
Gender							.001			.154
Female	5,942	66	1,920	64	3,629	68		393	61	
Male	3,046	34	1,080	36	1,713	32		253	39	
Medicaid coverage							<.001			.671
HARP	3,699	41	1,812	60	1,491	28		396	61	
Basic	5,289	59	1,188	40	3,851	72		250	39	
Serious mental illness	2,333	26	1,153	38	864	16	<.001	316	49	<.010
Substance use disorder	459	5	226	8	193	4	<.001	40	6	.235

[a] CCI, Charlson Comorbidity Index. Possible scores range from 0 to 33, with scores above 1 indicating increasing likelihood of next-year mortality.

Figura 5 – Características dos pacientes do *Medicaid*, pelo ambiente onde receberam serviços de saúde mental. Fonte: Wetzler et al. (2020).

Com base no diagnóstico principal de alta hospitalar, os serviços de internação foram divididos em atendimento médico geral, saúde mental e uso de substâncias. Para diferenciar entre a gravidade das necessidades de saúde mental, os pacientes foram classificados como portadores de doença mental grave, com base na classificação diagnóstica da maioria de suas reivindicações ambulatoriais como esquizofrenia, transtornos delirantes ou outros transtornos psicóticos e transtornos bipolares; todos os outros pacientes foram classificados como tendo doença mental não grave.

Service	Mental health clinic only (N=1,153)			Medical clinic only (N=864)			Mental health and medical clinic (N=316)		
	Total	HARP	Basic Medicaid	Total	HARP	Basic Medicaid	Total	HARP	Basic Medicaid
Outpatient visit	38.7	42.2	27.0	50.9	60.1	25.3	41.9	46.0	30.9
Mental health service provided by a primary care physician (PCP)	.7	.7	.7	4.0	4.5	2.4	2.0	2.1	1.6
Mental health service provided in medical clinic	.8	.8	.8	7.5	9.1	3.2	4.8	5.5	2.7
Mental health service provided in a mental health specialty clinic	12.1	12.8	9.7	6.3	7.0	4.4	13.2	13.4	12.4
Substance use service provided in a substance use specialty clinic	12.6	14.2	7.2	11.1	13.6	3.9	6.1	7.6	.7
Medical service provided by a PCP in a medical clinic	5.7	6.0	4.8	7.4	8.0	5.8	6.6	6.4	7.3
Medical service provided by a medical specialist in a medical clinic	6.8	7.7	3.8	14.6	17.9	5.6	9.2	11.0	6.2
Inpatient admissions per patient	.33	.33	.30	1.87	2.37	.58	.96	1.05	.64
Medical	.14	.15	.12	.64	.78	.25	.32	.35	.23
Mental health	.14	.12	.18	.94	1.20	.24	.57	.62	.41
Substance use disorder	.05	.06	.00	.29	.39	.09	.07	.08	.00

Figura 6 – Uso do serviço (número médio de visitas) entre pacientes do *Medicaid* com doença mental grave, pelo local onde receberam serviços de saúde mental e por tipo de cobertura do *Medicaid*. Fonte: Wetzler et al. (2020).

Segundo os autores, a maioria dos pacientes (59%) recebeu todos os seus cuidados de saúde mental em ambientes médicos e de provedores não afiliados, incluindo uma grande parcela (16%) com doença mental grave. Apesar da disponibilidade de cuidados integrados no ambiente médico e do uso de provedores de saúde mental não afiliados, as taxas de internações de saúde mental foram altas entre todos os pacientes neste ambiente, consideravelmente mais altas do que para pacientes tratados em uma clínica de saúde mental dentro do sistema. As taxas de internações médicas e de abuso de substâncias também foram muito mais altas para pacientes tratados no ambiente médico e por provedores não afiliados, em comparação com aqueles tratados nas clínicas de saúde mental do sistema.

A Figura 7 apresenta um modelo de regressão construído para comparar os resultados em todos os ambientes de prática.

Variable	Mental health admission				Medical admission				Substance use disorder admission			
	OR	95% CI		p	OR	95% CI		p	OR	95% CI		p
Setting of mental health services (reference: mental health setting only)												
Medical setting only	5.57	4.81	6.48	<.001	2.40	2.18	2.65	<.001	4.25	3.52	5.17	<.001
Medical and mental health settings	3.87	3.21	4.66	<.001	1.96	1.68	2.28	<.001	1.28	.88	1.81	.181
CCI score[a]	1.02	1.00	1.04	.080	1.25	1.24	1.27	<.001	1.06	1.04	1.08	<.001
Age	.97	.97	.98	<.001	.98	.98	.99	<.001	1.00	.99	1.01	.933
Female (reference: male)	.69	.63	.77	<.001	1.25	1.16	1.35	<.001	.21	.18	.25	<.001
HARP (reference basic Medicaid)	6.68	5.76	7.77	<.001	2.21	2.02	2.42	<.001	44.53	30.05	69.35	<.001
Serious mental illness (reference: nonserious mental illness)	3.08	2.74	3.45	<.001	.80	.73	.87	<.001	.76	.66	.88	.940
Substance use disorder (reference: none)	.55	.71	.41	<.001	.84	.72	.96	.013	1.33	1.09	1.61	<.010

[a] CCI, Charlson Comorbidity Index.

Figura 7 – Análise de regressão de variáveis como preditores de internações entre pacientes do *Medicaid* que recebem serviços de saúde mental. Fonte: Wetzler et al. (2020).

Os resultados encontrados por Wetzler et al. (2020) constataram que a maioria dos serviços de saúde mental em um grande centro médico acadêmico urbano foi prestada em ambientes médicos complementados por prestadores não afiliados, mesmo para pacientes com doenças mentais graves, e que alguns pacientes são encaminhados e recebem seus cuidados em especialidades de saúde mental dentro do sistema. Os pacientes que receberam assistência em estabelecimentos médicos e de prestadores de serviços não afiliados exibiram incidências mais elevadas de problemas relacionados à saúde mental, consumo de substâncias e internações hospitalares. As características não documentadas dos pacientes podem explicar essa disparidade, embora possa haver diferenças intrínsecas na forma como os cuidados são prestados nos vários ambientes. Os autores entendem que mais pesquisas são necessárias para determinar quais modelos de cuidados de saúde mental atendem melhor às necessidades de vários pacientes.

2.4 Trabalho de Roggenkamp et al. (2018)

Em seu trabalho, Roggenkamp et al. (2018) teve como objetivo principal descrever características demográficas e clínicas dos casos de saúde mental em *Victoria, Australia*. Através de um estudo observacional retrospectivo dos serviços de emergência ocorridos no ano de 2015.

Os dados utilizados foram extraídos de um DW eletrônico da *Ambulance Victoria* (AMBULANCE. . . ,). O banco de dados inclui várias fontes de dados, incluindo registros

em campo, triagem secundária e despacho assistido por computador. Os dados de pacientes e tratamento em campo são registrados por paramédicos usando um *tablet* de computador na conclusão de cada caso, que produz um registro eletrônico de atendimento ao paciente e, em seguida, sincroniza com o DW.

Para as análises estatísticas, os dados contínuos foram descritos usando medianas e intervalos interquartis (IQR). As comparações foram feitas utilizando-se os testes de *Mann Whitney-u* e *Kruskal Wallis*, conforme apropriado. Os dados categóricos foram resumidos por meio de contagens e porcentagens. Todas as análises estatísticas foram realizadas utilizando o programa SPSS Versão 22.0 para *Windows* (IBM Corp., 2013).

Do total de 504.676 atendimentos, 48.041 (9,5%) foram apresentações de saúde mental. Além disso, 4.708 (6,6%) casos gerenciados por um paramédico ou enfermeiro por meio do serviço secundário de triagem telefônica também envolviam queixas de saúde mental.

Os pacientes atendidos em cuidados de saúde mental apresentavam características demográficas distintas em relação aos demais pacientes atendidos, sendo notadamente mais jovens e com uma maior prevalência do sexo feminino. Constatou-se que a maioria dos indivíduos com transtornos mentais foi encaminhada ao hospital (74,4%). Contudo, verificou-se que os paramédicos forneceram tratamento a uma proporção consideravelmente menor de pacientes com problemas de saúde mental em comparação com aqueles que apresentavam outras condições médicas (12,4% x 50,3%).

A maioria dos pacientes de saúde mental (76,8%) tinha um histórico documentado de saúde mental. Problemas sociais ou emocionais foram a apresentação mais comum em pacientes de saúde mental com 15 anos (19,1%); enquanto, para pacientes com 65 anos, a ansiedade foi a apresentação clínica mais comum (41,2%).

Para os pacientes submetidos à triagem secundária, 52,5% eram chamadas frequentes ou apresentavam ansiedade. Um total de 27,7% dos pacientes triados foram encaminhados para um serviço alternativo, enquanto 24,6% foram tratados sob um plano de cuidados existente.

Por fim, Roggenkamp et al. (2018) concluiu que os casos relacionados à saúde mental representam um em cada 10 atendimentos em *Victoria*, *Australia* e são altamente heterogêneos em termos de apresentações clínicas. Uma pequena proporção de pa-

cientes requer tratamento de crise de emergência, sobrecarregando os serviços de emergência, incluindo ambulância, polícia e serviços de saúde mental. Em comparação, a maioria dos pacientes de saúde mental apresenta condições relativamente não graves e poderia se beneficiar de serviços alternativos prestados por médicos de saúde mental na comunidade. Muitas vezes, os paramédicos carecem das habilidades especializadas em saúde mental necessárias para gerenciar essas apresentações na comunidade, e isso pode levar a um aumento nas apresentações do departamento de emergência. A identificação de pacientes de saúde mental no momento da chamada também pode ajudar a identificar pacientes que se beneficiariam dos serviços comunitários de saúde mental.

3 METODOLOGIA

Neste trabalho, foi desenvolvido e implementado um modelo de BI com base na arquitetura proposta por Kimball e Ross (2013). O modelo foi capaz de orquestrar todo o processo de ETL e, ao final, armazenou os dados em um DW que serviu como base para a criação dos *dashboards*. A Figura 8 apresenta a arquitetura proposta para este trabalho.

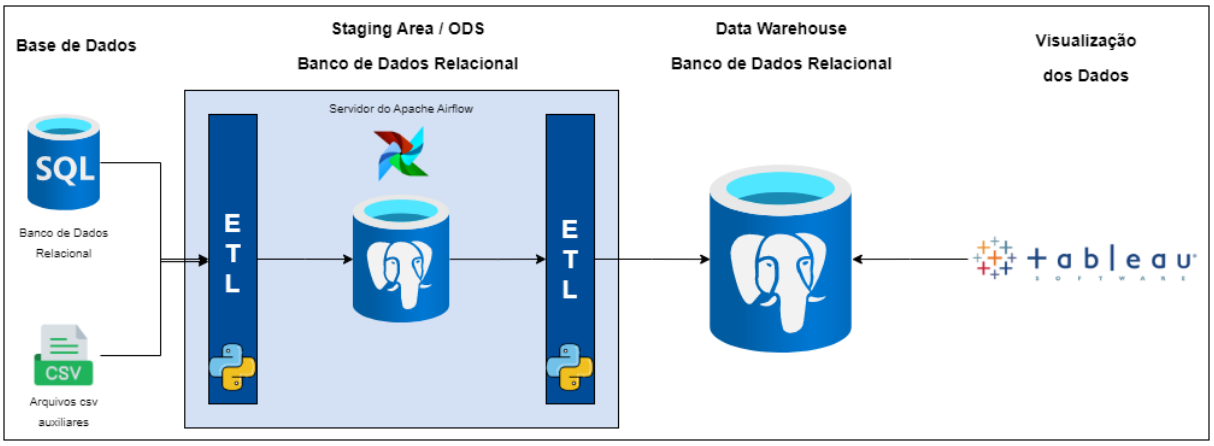

Figura 8 – Arquitetura do DW.

A fase inicial do processo consistiu em realizar um estudo dos requisitos do problema, com o objetivo de mapear os principais indicadores e métricas que seriam atendidos no processo de BI.

O primeiro passo no mapeamento de indicadores é compreender os objetivos e metas da organização. Isso significa entender as necessidades dos *stakeholders* e as áreas-chave que impulsionam o sucesso do negócio. Com essa compreensão, é possível determinar quais informações são críticas para monitorar e avaliar o desempenho dessas áreas.

Em seguida, o processo envolve a identificação de métricas relevantes para cada objetivo. Essas métricas são os indicadores que refletem o progresso em direção às metas estabelecidas. É importante garantir que esses indicadores sejam mensurá-

veis, quantificáveis e alinhados aos objetivos estratégicos da organização.

O mapeamento de indicadores também requer a definição de fontes de dados confiáveis e a especificação de como esses indicadores serão calculados e atualizados. Muitas vezes, isso envolve a integração de várias fontes de dados em um único ambiente de BI, onde os dados são transformados e consolidados para garantir a consistência e a precisão das informações. É importante ressaltar que o mapeamento de indicadores é um processo iterativo e contínuo. À medida que a organização evolui e as metas são ajustadas, os indicadores também podem precisar ser revisados e atualizados para garantir que continuem relevantes e alinhados com as necessidades do negócio.

Uma vez obtidos os indicadores e as bases de dados, foi realizado um processo de modelagem dimensional para determinar as tabelas dimensão e tabelas fatos necessárias para atender ao problema proposto.

Posteriormente, foi criada uma estrutura utilizando o Sistema de Gerenciamento de Banco de Dados Objeto-Relacional (ORDBMS) *PostgreSQL*, a 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 etapa subsequente envolveu a implementação do processo de ETL, utilizando a linguagem de programação *Python*. Durante essa fase, 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 dimensão e fato do DW, que foi mantido em um 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.

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* nos quais os indicadores e métricas com os resultados obtidos foram apresentados, de acordo com os requisitos identificados na etapa inicial.

A seguir serão apresentadas todas as ferramentas e fontes de dados empregadas na elaboração da arquitetura e construção do processo de BI proposto neste estudo.

3.1 Ferramentas utilizadas

3.1.1 *Apache Airflow*

O *Apache Airflow* é uma plataforma de código aberto para desenvolver, agendar e monitorar fluxos de trabalho orientados a lotes (APACHE, 2022). A estrutura *Python* extensível do *Airflow* permite que você crie fluxos de trabalho conectados a praticamente qualquer tecnologia (APACHE, 2022).

O *Airflow* se utiliza do conceito de DAG (*Directed Acyclic Graph*), onde o fluxo de execução é representado por um grafo direcionado acíclico, no qual cada vértice do grafo representa uma tarefa a ser executada e os arcos definem as relações de dependência entre as tarefas.

Para a criação das DAG's, utilizou-se a linguagem de programação *Python*, devido à sua fácil integração com o *Airflow*. Dessa forma, para cada uma das tarefas, foram desenvolvidos *script* em *Python* que foram responsáveis por uma determinada etapa do processo ETL.

3.1.2 *PostgreSQL*

O *PostgreSQL* é um poderoso sistema de banco de dados objeto-relacional de código aberto com mais de 35 anos de desenvolvimento ativo que lhe rendeu uma forte reputação de confiabilidade, robustez de recursos e desempenho (POSTGRESQL, 2022). O PostgreSQL vem com muitos recursos destinados a ajudar os desenvolvedores a criar aplicativos, os administradores a proteger a integridade dos dados e a criar ambientes tolerantes a falhas, além de ajudá-lo a gerenciar seus dados, independentemente do tamanho do conjunto de dados (POSTGRESQL, 2022). Por meio dele, foi possível criar e manter as estruturas utilizadas neste trabalho (*staging area* e DW).

3.1.3 *Python*

Para a implementação do processo de ETL, optou-se pela linguagem de programação *Python*, pois se trata de uma linguagem elegante e versátil que conta com um ecossistema de módulos muito completa. A linguagem possui diversas bibliotecas e *frameworks* para acessar e extrair dados além de utilitários de gerenciamento de fluxo de trabalho e *kits* de ferramentas ETL com uma enorme variedade de recursos.

3.1.4 *Tableau*

A criação dos *dashboards* foi realizada por meio da utilização da ferramenta OLAP *Tableau Desktop*. Esta escolha se deve ao fato de que a mencionada ferramenta proporciona a disponibilidade de licença gratuita para estudantes universitários e apresenta uma ampla gama de recursos que favorecem a agilidade e a intuição no processo de criação das visualizações. Ademais, a ferramenta é compatível com diversas fontes de dados, disponibiliza funcionalidades para limpeza e fusão dos dados, permite a criação de cálculos avançados e dispõe de instrumentos para a realização de previsões de tendências fundamentadas em dados.

3.1.5 Base de dados de origem

O DATASUS é uma importante entidade governamental vinculada ao Ministério da Saúde do Brasil. Sua principal missão é gerenciar e disponibilizar informações e dados estratégicos relacionados à saúde pública no país. Foram empregados dados provenientes do SIM e do SIH, os quais são disponibilizados pelo DATASUS.

O SIM é uma importante ferramenta de coleta, registro e análise de dados sobre óbitos no Brasil. Gerenciado pelo DATASUS, vinculado ao Ministério da Saúde, o SIM tem como objetivo fornecer informações detalhadas e precisas sobre as causas de morte e padrões de mortalidade no país. Com base nos dados obtidos de registros de óbitos de cartórios, hospitais e outras fontes, o SIM produz estatísticas essenciais para a formulação e avaliação de políticas públicas de saúde, além de subsidiar estudos epidemiológicos, pesquisas científicas e o monitoramento de doenças e fatores de risco.

O SIH é um dos componentes fundamentais do DATASUS no Brasil. O SIH tem como principal objetivo coletar e armazenar informações sobre os atendimentos hospitala-

res realizados no âmbito do SUS. Essas informações são provenientes das unidades hospitalares públicas e conveniadas ao SUS, e incluem dados sobre procedimentos realizados, quantidade de internações, diagnósticos, tempo de permanência dos pacientes, entre outros dados relevantes.

A Plataforma de Ciência de Dados aplicada à Saúde (PCDaS) é uma plataforma pública, gratuita e em nuvem que disponibiliza grande quantidade de microdados e ferramentas de análise para pesquisas no campo da saúde (PCDAS, 2023). Neste trabalho, foram utilizados os dicionários de dados disponibilizados no site da 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.

Os dados referentes à saúde mental do estado do Espírito Santo, do período de janeiro de 2010 a dezembro de 2021, foram objeto de análise após a aplicação de filtros para selecionar apenas os registros relacionados ao Capítulo V do CID, que abrange os transtornos mentais e comportamentais. A descrição completa dos campos que foram utilizados neste trabalho e são provenientes da fonte de dados estão no Dicionário de Dados. As versões completas do dicionário, que contêm a descrição de todos os campos das fontes de dados, esta disponível no website da PCDaS.

4 RESULTADOS OBTIDOS

4.1 Levantamento dos Indicadores

O levantamento dos indicadores antes da modelagem dimensional é de extrema importância, pois permite compreender quais informações são relevantes e necessárias para o processo de tomada de decisão dentro de uma organização. Ao identificar e definir os indicadores-chave de desempenho, é possível estruturar um modelo dimensional eficiente, que irá organizar e relacionar os dados de forma adequada, facilitando a análise e extração de *insights* relevantes. Isso proporciona 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:

- Visão histórica de mortalidade por transtornos mentais;
- Principais causas de mortalidade por transtornos mentais;
- Nº de óbitos por transtornos mentais por raça e cor;
- Nº de óbitos por transtornos mentais por município do ES;
- Nº de óbitos por transtornos mentais por sexo;
- Nº de óbitos por transtornos mentais por faixa etária;
- Visão histórica de hospitalizações por transtornos mentais (quantidade e valor gasto);
- Principais causas de hospitalizações por transtornos mentais (quantidade e valor gasto);

- Hospitalizações por transtornos mentais por procedimentos (quantidade e valor gasto);
- Hospitalizações por transtornos mentais por raça e cor (quantidade e valor gasto);
- Hospitalizações por transtornos mentais por município (quantidade e valor gasto);
- Hospitalizações por transtornos mentais por sexo (quantidade e valor gasto);
- Hospitalizações por transtornos mentais por faixa etária (quantidade e valor gasto);
- Valor total gasto com hospitalizações por transtornos mentais.

4.2 Estrutura do Projeto e Processo de ETL

Para a codificação do projeto, foi utilizado o editor de texto *Visual Studio Code* (VISUAL..., 2023). A Figura 9 apresenta a estrutura do projeto. O projeto completo contendo todos os *scripts* esta localizado no Repositório do *GitHub*.

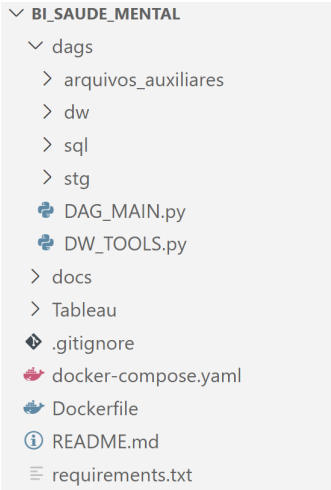

Figura 9 – Estrutura do Projeto hospedado no GitHub.

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 9, são descritos a seguir.

4.2.1 Pasta dags

Na pasta *"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.

A pasta denominada *"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*.

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

Na pasta "*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.

4.2.2 Pasta docs

Neste diretório são armazenados todos os documentos referentes ao projeto. Nele encontra-se uma cópia do Dicionário de Dados e do modelo do cubo de dados implementado neste trabalho.

4.2.3 Pasta Tableau

Este diretório inclui 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.

4.2.4 *Airflow*

A Figura 10 exibe a DAG no *Airflow*, no qual é possível observar o fluxo do processo de ETL.

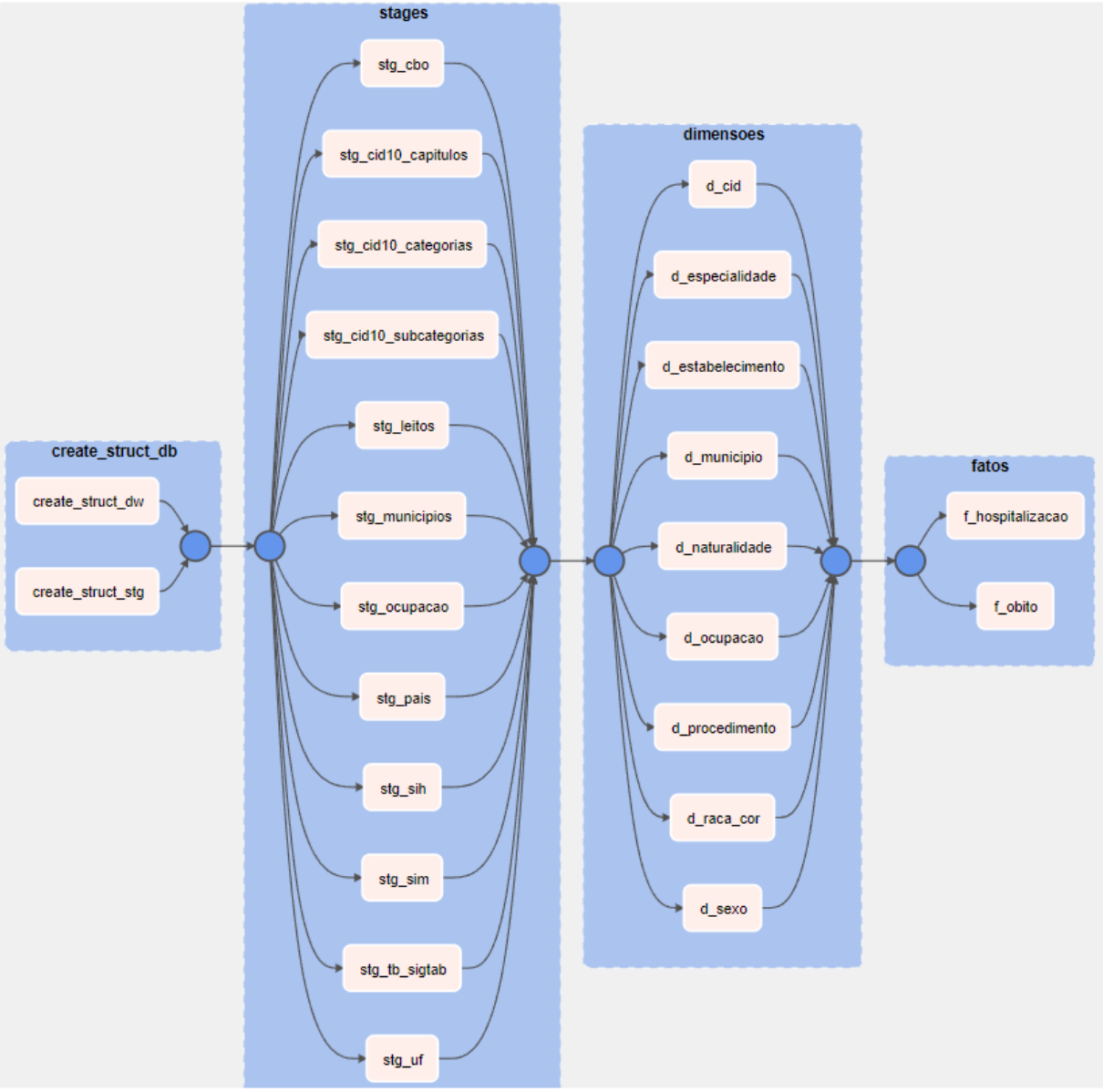

Figura 10 – Fluxo de execução do processo de ETL no *Airflow*.

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.

4.3 *Data Warehouse*

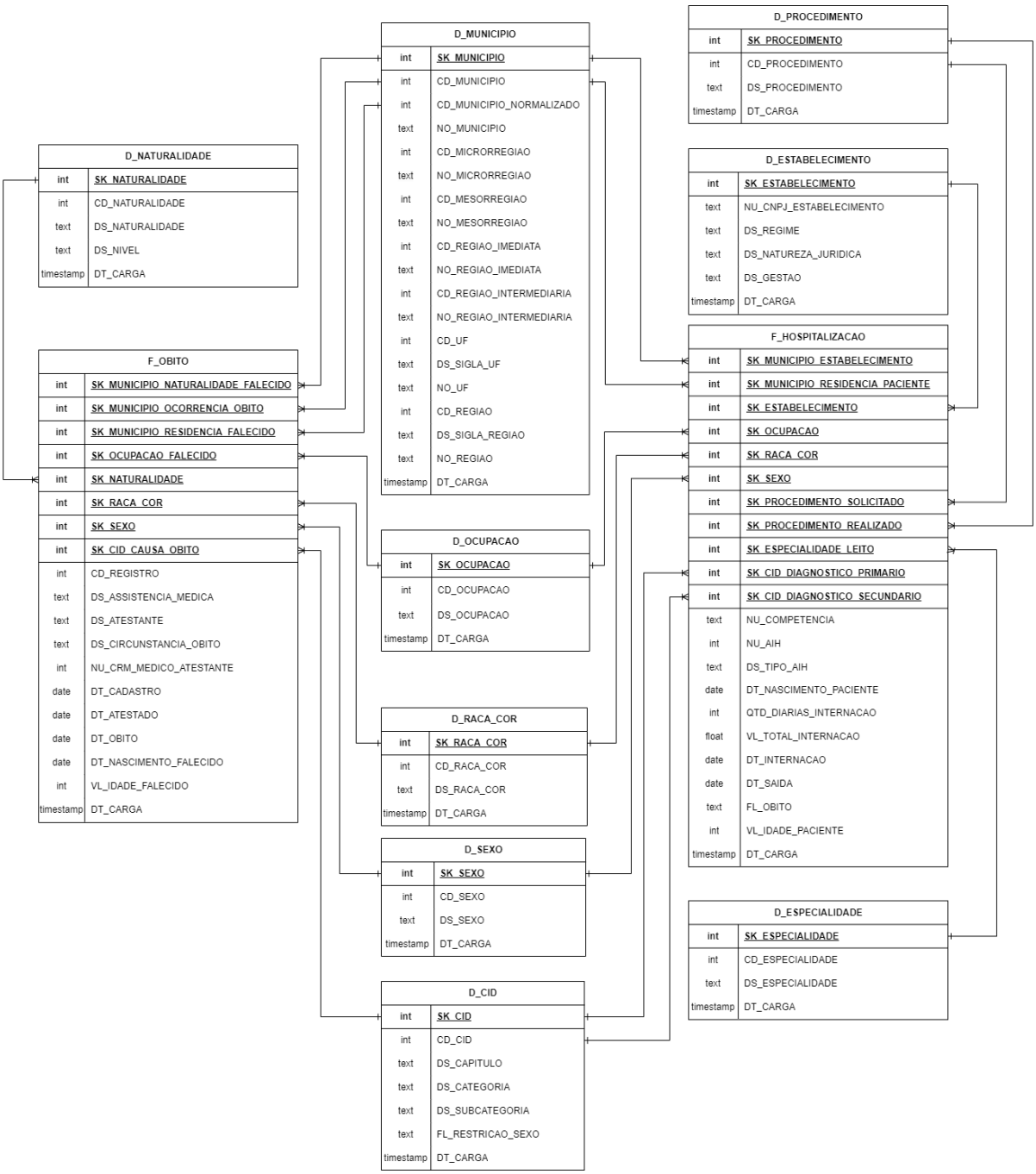

Figura 11 – Modelo Dimensional do cubo de dados.

Posteriormente à identificação dos requisitos e à análise preliminar dos dados prove-

nientes da fonte, foi realizado o desenvolvimento do 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 11 apresenta o modelo dimensional desenvolvido, proporcionando uma visão ampla da estrutura do *DW* implementado. O documento completo com as informações detalhadas de todas as tabelas do DW está disponível no Dicionário de Dados.

Para este estudo, foram elaboradas nove tabelas de dimensão. São listadas a seguir:

- **D_CID**: dimensão para armazenar os dados referentes ao CID de hospitalização ou óbito;
- **D_ESPECIALIDADE**: dimensão que abrange os dados relacionados à especialidade do leito de hospitalização;
- **D_ESTABELECIMENTO**: dimensão que abrange os dados pertinentes ao estabelecimento hospitalar;
- **D_MUNICIPIO**: dimensão contendo informações relativas a dados geográficos;
- **D_NATURALIDADE**: dimensão contendo os dados relacionados à naturalidade do falecido;
- **D_OCUPACAO**: dimensão que contém os dados relativos à ocupação do paciente ou falecido;
- **D_PROCEDIMENTO**: dimensão que engloba os dados relacionados aos procedimentos solicitados ou realizados pelo paciente;
- **D_RACA_COR**: dimensão que inclui os dados relacionados à raça/cor do paciente ou do falecido;
- **D_SEXO**: dimensão que inclui os dados relacionados ao gênero do paciente ou do falecido.

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 nas tabelas de fato, permitindo a posterior correlação entre as tabelas de fato e dimensão.

Também foram concebidas duas tabelas fato, denominadas:

- **F_HOSPITALIZACOES**: tabela fato que contém informações provenientes do SIH referentes às hospitalizações ocorridas;
- **F_OBITO**: tabela fato utilizada para armazenar as informações provenientes do SIM relativas aos óbitos ocorridos.

As dimensões de CID, município, ocupação, raça e sexo são compartilhadas entre as duas tabelas fato. 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 haverão registros duplicados entre as duas fatos, tendo em vista que as duas não se correlacionam.

De acordo com o item (a) da Figura 12, 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 12 oferecem uma visão da estrutura interna, contendo todas as tabelas da *staging area* e do DW, respectivamente.

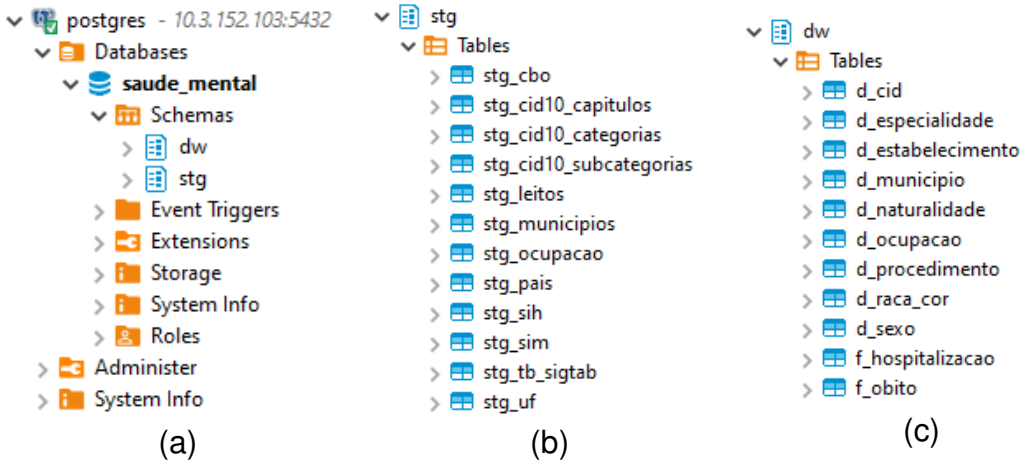

Figura 12 – Estrutura do banco de dados, da *staging area* e do DW, respectivamente.

4.4 *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. A seguir, serão expostos alguns dos painéis de controle desenvolvidos para este estudo. Foram selecionados os painéis com as análises mais relevantes. A página do *Tableau Public* que contém o *dashboard* completo está disponível para acesso por meio do Link do Repositório.

A Figura 13 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.

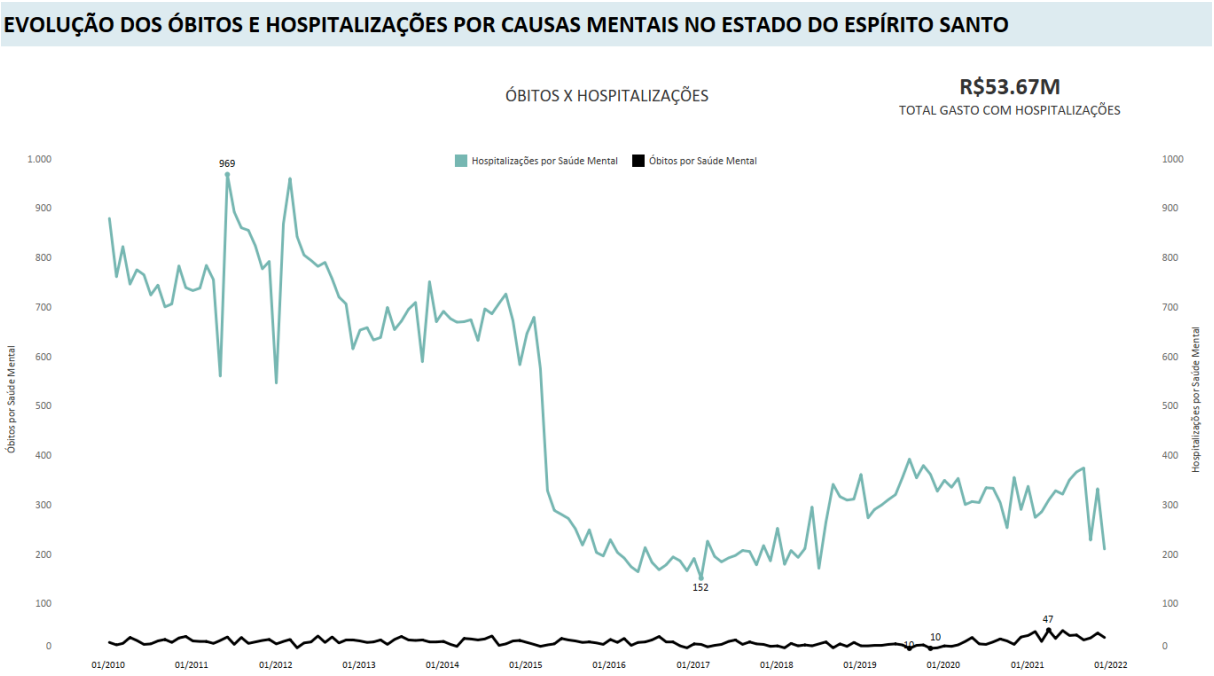

Figura 13 – Índices de hospitalizações e mortalidade associados à saúde mental no estado do Espírito Santo.

Nos *dashboards* apresentados nas Figuras 14 e 15, 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.

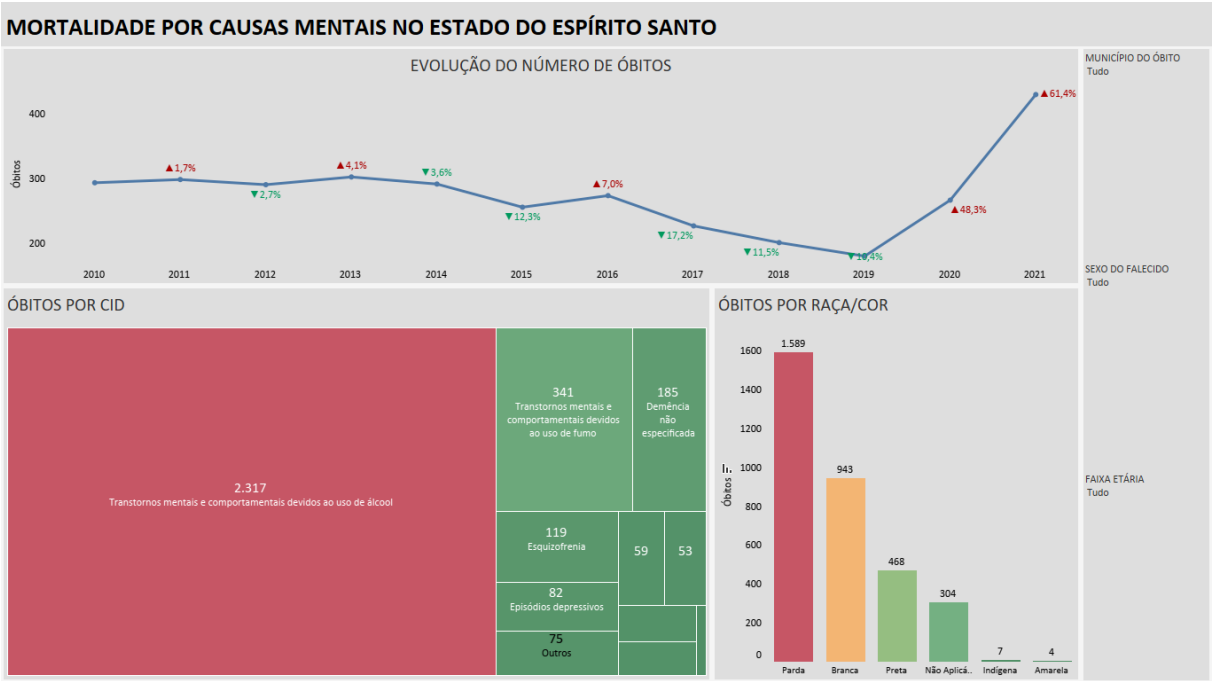

Figura 14 – Painel contendo uma visão geral dos dados referentes aos óbitos decorrentes de causas mentais no estado do Espírito Santo.

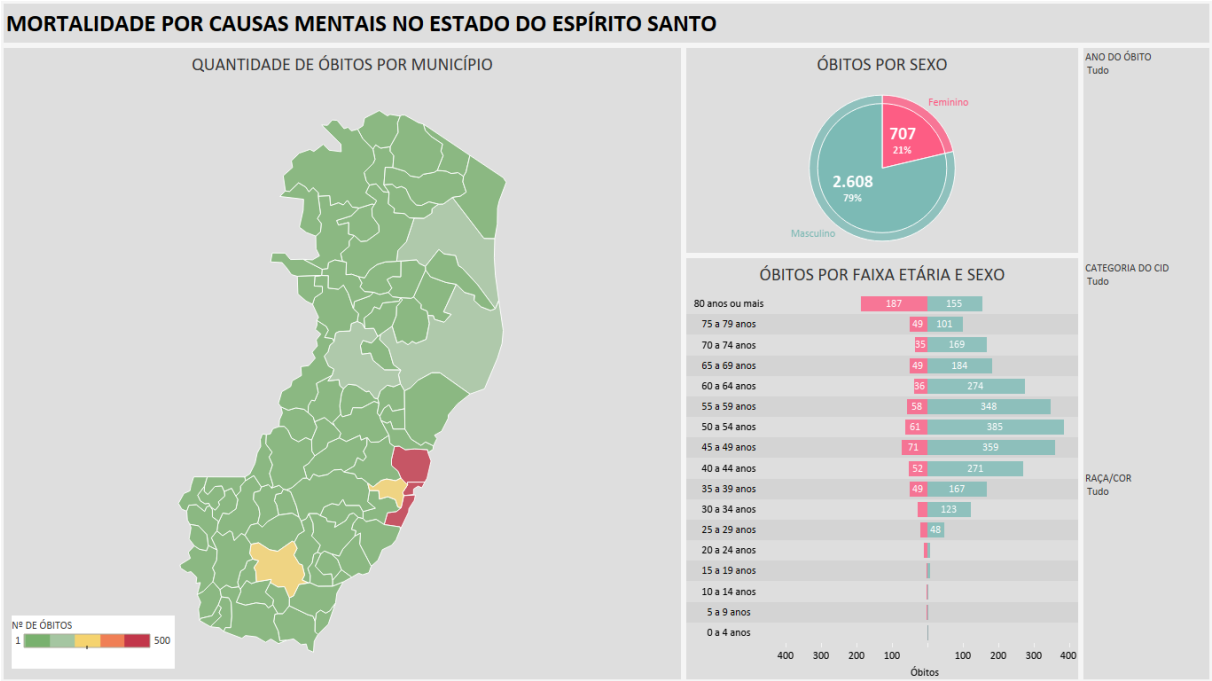

Figura 15 – Painel contendo uma visão geográfica dos dados referentes aos óbitos decorrentes de causas mentais no estado do Espírito Santo.

De acordo com os *dashboards* de óbitos, é possível observar nas Figuras 16 e 17 os painéis correlacionados às hospitalizações por problemas de saúde mental.

Por meio desses *dashboards*, é viabilizada uma análise abrangente da progressão das hospitalizações ao longo dos anos, bem como das principais razões para a hospitalização, além de permitir a segmentação por município, faixa etária e sexo do paciente.

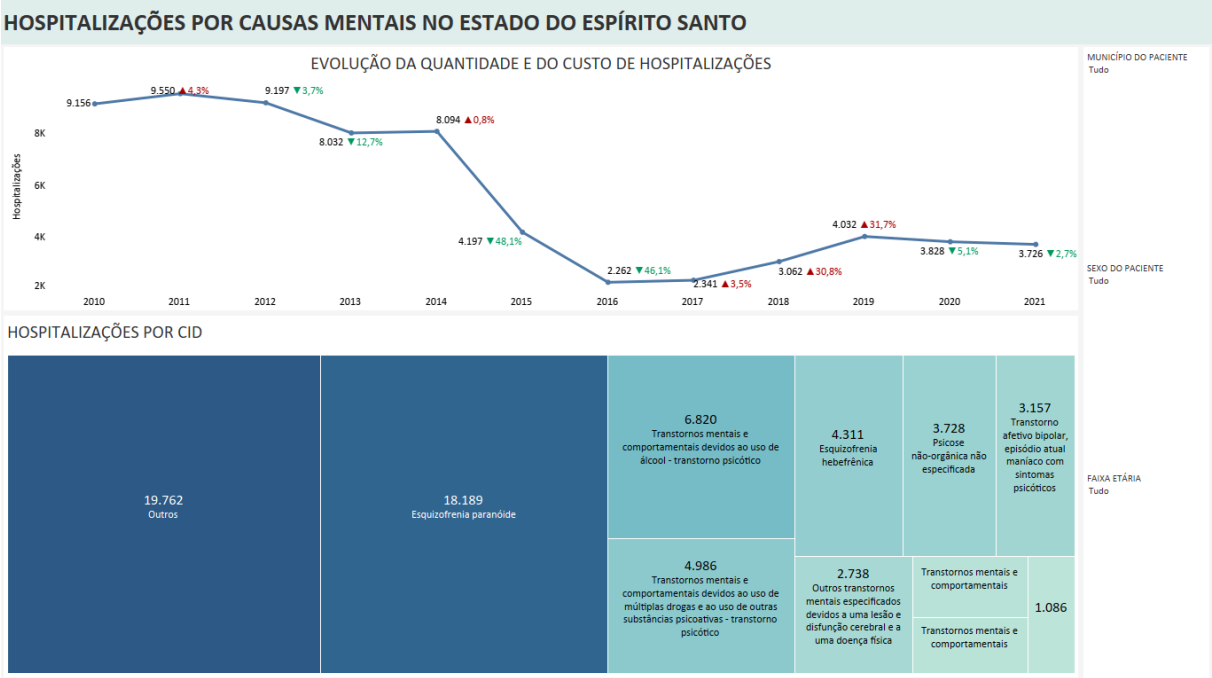

Figura 16 – Painel contendo uma visão panorâmica dos dados relativos às hospitalizações resultantes de transtornos mentais no estado do Espírito Santo.

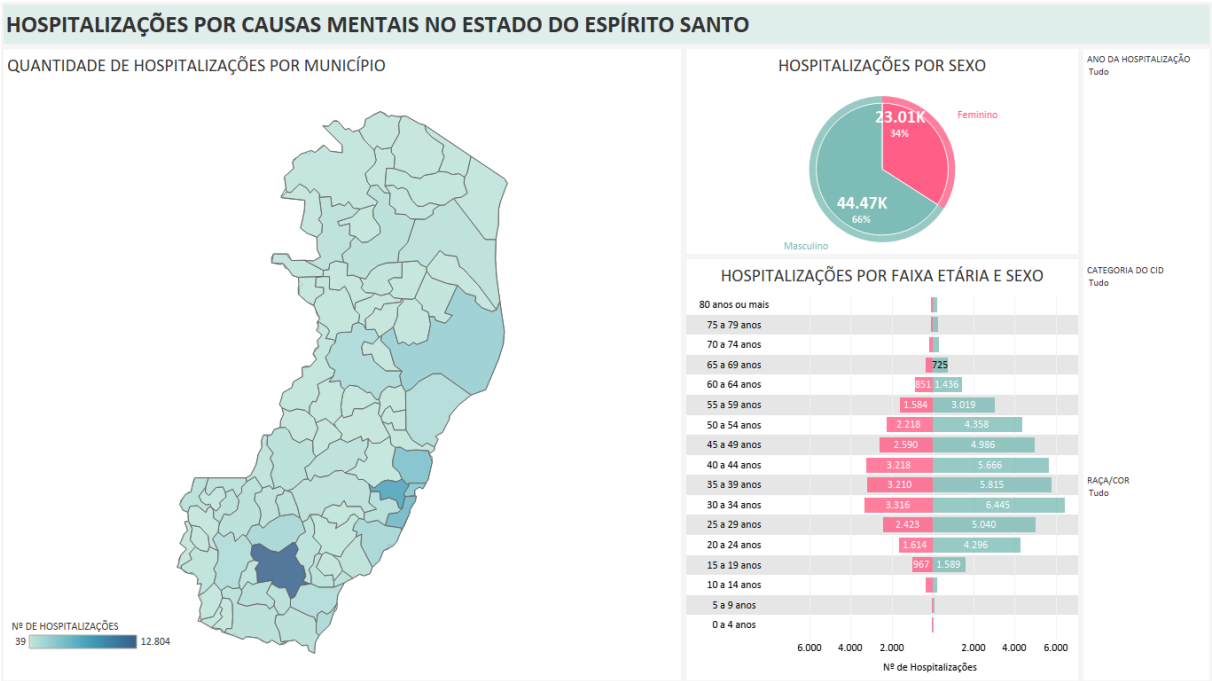

Figura 17 – Painel exibindo uma representação geográfica dos dados relativos às hospitalizações resultantes de causas mentais no território do estado do Espírito Santo.

5 CONCLUSÕES

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 em tempo real, 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.

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 poderia ser estendida em abrangência nacional e incluir a investigação de outras vertentes da área da saúde.

Durante o período analisado, constatou-se uma notável queda no número de hospitalizações relacionadas à saúde mental no estado do Espírito Santo. No entanto, paradoxalmente, ocorreu um preocupante aumento no número de óbitos associados a transtornos mentais, especialmente durante a pandemia do COVID-19. Essa tendência alarmante sugere que a crise sanitária exerceu um impacto significativo na saúde mental da população, resultando em consequências trágicas, principalmente em municípios como Cachoeiro de Itapemirim, que apresentaram índices de internação superiores em comparação aos municípios da Grande Vitória.

Além disso, observou-se uma predominância do sexo masculino tanto nas estatísticas de óbitos quanto de hospitalizações, levantando questionamentos sobre possíveis fatores sociais e de saúde que podem estar influenciando essa disparidade de gênero em relação às questões de saúde mental. 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 e hospitalizações 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. 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.

Em suma, o presente estudo foi capaz de fornecer uma visão abrangente e aprofundada 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.

REFERÊNCIAS

AMBULANCE Victoria. <https://www.ambulance.vic.gov.au/>. Acessado em: 22 de julho de 2023. Citado na página 21.

APACHE, S. F. *Apache Airflow Documentation, Version: 2.3.3*. 2022. Disponível em: <https://airflow.apache.org/docs/apache-airflow/stable/index.html>. Citado na página 26.

BIBLIOMETRIX. <https://www.bibliometrix.org/>. Acessado em: 22 de julho de 2023. Citado na página 12.

BRêDA, M. Z.; AUGUSTO, L. G. d. S. O cuidado ao portador de transtorno psíquico na atenção básica de saúde. *Ciência & Saúde Coletiva*, ABRASCO - Associação Brasileira de Saúde Coletiva, v. 6, n. Ciênc. saúde coletiva, 2001 6(2), 2001. ISSN 1413-8123. Disponível em: <https://doi.org/10.1590/S1413-81232001000200016>. Citado na página 9.

CARDOSO, L.; GALERA, S. A. F. O cuidado em saúde mental na atualidade. *Revista da Escola de Enfermagem da USP*, Universidade de São Paulo, Escola de Enfermagem, v. 45, n. Rev. esc. enferm. USP, 2011 45(3), Jun 2011. ISSN 0080-6234. Disponível em: <https://doi.org/10.1590/S0080-62342011000300020>. Citado na página 9.

CLUAIN Mhuire. <https://cluainmhuire.ie>. Acessado em 22 de julho de 2023. Citado na página 17.

COELHO, F. C. et al. *AlertaDengue/PySUS: Vaccine*. Zenodo, 2021. Disponível em: <https://doi.org/10.5281/zenodo.4883502>. Citado na página 25.

DATASUS - Departamento de Informática do SUS. <https://datasus.saude.gov.br>. Acessado em: 22 de julho de 2023. Citado na página 10.

HITACHI VANTARA. *Pentaho*. [S.l.]. Acessado em 22 de julho de 2023. Citado na página 14.

HOUSBANE, S. et al. Real-time monitoring system to manage mental healthcare emergency unit. *HEALTHCARE INFORMATICS RESEARCH*, KOREAN SOC MEDICAL INFORMATICS, 1618 KYUNGHEEGUNG ACHIM BLDG 3, 72 NAESOO-DONG CHONGRO-KU, SEOUL, 110-872, SOUTH KOREA, v. 26, n. 4, p. 344–350, OCT 2020. ISSN 2093-3681. Citado 5 vezes nas páginas , 14, 15, 16 e 17.

IBM Corp. *IBM SPSS Statistics for Windows, Version: 22.0*. Armonk, NY: [s.n.], 2013. Citado 2 vezes nas páginas 18 e 22.

JORGE, M. R.; FRANçA, J. M. A associação brasileira de psiquiatria e a reforma da assistência psiquiátrica no brasil. *Brazilian Journal of Psychiatry*, Associação Brasileira de Psiquiatria, v. 23, n. Braz. J. Psychiatry, 2001 23(1), Mar 2001. ISSN 1516-4446. Disponível em: <https://doi.org/10.1590/S1516-44462001000100002>. Citado na página 9.

JUPYTER, P. *Jupyter Notebook*. <https://jupyter.org/>. Acessado em 22 de julho de 2023. Citado na página 14.

KIMBALL, R.; ROSS, M. *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*. 3. ed. Indianapolis, IN: Wiley, 2013. ISBN 978-1-118-53080-1. Disponível em: <https://www.safaribooksonline.com/library/view/the-data-warehouse/9781118530801/>. Citado 4 vezes nas páginas , 10, 11 e 24.

Oracle. *Oracle Business Intelligence, Version: 10.1.2.45.46c*. 2005. Disponível em: <https://www.oracle.com/br/what-is-business-intelligence>. Citado na página 18.

PCDAS, F. *Plataforma de Ciência de Dados aplicada à Saúde. Laboratório de Informação em Saúde (Lis). Instituto de Comunicação e Informação Científica e Tecnológica em Saúde (Icict).* 2023. Disponível em: <https://pcdas.icict.fiocruz.br.DOI: https://doi.org/10.7303/syn25882127>. Acesso em: 07 de janeiro 2023. Citado na página 28.

PORTAL de Periódicos CAPES. <https://www.periodicos.capes.gov.br/>. Acessado em: 22 de julho de 2023. Citado na página 12.

POSTGRESQL, G. D. G. *PostgreSQL Documentation, Version: 15.1*. 2022. Disponível em: <https://www.postgresql.org/docs/current/index.html>. Citado na página 26.

PYTHON Programming Language. <https://www.python.org/>. Acessado em 22 de julho de 2023. Citado na página 14.

R Project for Statistical Computing. <https://www.r-project.org/>. Acessado em: 22 de julho de 2023. Citado na página 12.

RODRIGUES, L. R.; CASAGRANDE, L. D. R. *Só quem sabe da doença dele é Deus: o significado da doença mental no contexto cultural*. Dissertação (Mestrado) — Universidade de São Paulo, 2001. Citado na página 9.

ROGGENKAMP, R. et al. Descriptive analysis of mental health-related presentations to emergency medical services. *PREHOSPITAL EMERGENCY CARE*, TAYLOR & FRANCIS INC, 530 WALNUT STREET, STE 850, PHILADELPHIA, PA 19106 USA, v. 22, n. 4, p. 399–405, 2018. ISSN 1090-3127. Citado 3 vezes nas páginas , 21 e 22.

ROWNTREE, R.; MCCARTHY, N.; FEENEY, L. Changing prescribing patterns in an irish community mental health service. *IRISH JOURNAL OF PSYCHOLOGICAL MEDICINE*, CAMBRIDGE UNIV PRESS, EDINBURGH BLDG, SHAFTESBURY RD, CB2 8RU CAMBRIDGE, ENGLAND, v. 37, n. 1, p. 8–14, MAR 2020. ISSN 0790-9667. Citado 3 vezes nas páginas , 17 e 18.

SCOPUS. <http://www.scopus.com>. Acessado em: 22 de julho de 2023. Citado na página 12.

SOUZA, A. d. J. F. et al. A saúde mental no programa de saúde da família. *Revista Brasileira de Enfermagem*, Associação Brasileira de Enfermagem, v. 60, n. Rev. Bras. Enferm., 2007 60(4), Jul 2007. ISSN 0034-7167. Disponível em: <https://doi.org/10.1590/S0034-71672007000400006>. Citado na página 9.

TEAM, J. D. *Voila: Turning Jupyter Notebooks into Standalone Web Applications*. <https://voila.readthedocs.io>. Acessado em 22 de julho de 2023. Citado na página 14.

VISUAL Studio Code. 2023. <https://code.visualstudio.com>. Acesso em 4 de julho de 2023. Citado na página 30.

WEB of Science. <https://www-webofscience.ez43.periodicos.capes.gov.br/wos/>. Acessado em: 22 de julho de 2023. Citado na página 12.

WETZLER, S. et al. Treatment of serious mental illness in medical and mental health settings. *PSYCHIATRIC SERVICES*, AMER PSYCHIATRIC PUBLISHING, INC, 800 MAINE AVE SW, SUITE 900, WASHINGTON, DC 20024 USA, v. 71, n. 8, p. 789–795, AUG 2020. ISSN 1075-2730. Citado 4 vezes nas páginas , 19, 20 e 21.