

Universidade de Brasília – UnB
Faculdade UnB Gama – FGA
Engenharia de Software

***Business Intelligence* aplicada à saúde: painéis
de monitoramento para profissionais de saúde
da Rede de Atenção Psicossocial**

Autor: Luis Bruno Fidelis Gomes
Orientador: Prof. Dr. George Marsicano Correa

Brasília, DF
2023



Luis Bruno Fidelis Gomes

***Business Intelligence* aplicada à saúde: painéis de monitoramento para profissionais de saúde da Rede de Atenção Psicossocial**

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Universidade de Brasília – UnB

Faculdade UnB Gama – FGA

Orientador: Prof. Dr. George Marsicano Correa

Brasília, DF

2023

Luis Bruno Fidelis Gomes

Business Intelligence aplicada à saúde: painéis de monitoramento para profissionais de saúde da Rede de Atenção Psicossocial/ Luis Bruno Fidelis Gomes.
– Brasília, DF, 2023-

100 p. : il. (algumas color.) ; 30 cm.

Orientador: Prof. Dr. George Marsicano Correa

Trabalho de Conclusão de Curso – Universidade de Brasília – UnB
Faculdade UnB Gama – FGA , 2023.

1. Business Intelligence. 2. Painéis de Monitoramento. I. Prof. Dr. George Marsicano Correa. II. Universidade de Brasília. III. Faculdade UnB Gama. IV. *Business Intelligence* aplicada à saúde: painéis de monitoramento para profissionais de saúde da Rede de Atenção Psicossocial

CDU 02:141:005.6

Luis Bruno Fidelis Gomes

***Business Intelligence* aplicada à saúde: painéis de monitoramento para profissionais de saúde da Rede de Atenção Psicossocial**

Monografia submetida ao curso de graduação em Engenharia de Software da Universidade de Brasília, como requisito parcial para obtenção do Título de Bacharel em Engenharia de Software.

Brasília, DF, 11 de dezembro de 2023 – Data da aprovação do trabalho:

Prof. Dr. George Marsicano Correa
Orientador

Prof. Dr. John Lenon Cardoso Gardenghi
Membro da Banca

Prof. Dr. Daniel Sundfeld Lima
Membro da Banca

Brasília, DF
2023

Este projeto é dedicado a todas as pessoas que dependem do sistema de saúde público, espero que esse estudo contribua para a melhoria da qualidade de vida dessas pessoas.

Agradecimentos

Gostaria de agradecer a Deus por me conceder saúde e força para ser capaz de me dedicar a este trabalho e por permitir que pessoas excepcionais estejam sempre ao meu redor.

Agradeço também aos meus pais, que me deram amor, educação, qualidade de vida, conforto e que são incríveis, apesar das dificuldades que tiveram no passado. Agradeço aos meus irmãos por terem me apoiado nessa longa trajetória da graduação, me abraçar e me acolher em momentos difíceis.

Gostaria de agradecer a todos os meus amigos, especialmente Eduardo Junio, Talles Brendo, Iuri Ramos, Marcos Vinícius, Felipe Lopes, Arthur Serra, Cledimildo Júnior, Rony Moura, Victor Hugo, Adriano Calisto, Gabriel Gauss, Igor Lima e Danilo Duarte, que me ajudaram de alguma forma neste trabalho, seja revisando, traduzindo, oferecendo palavras de incentivo, ou sendo um ombro amigo.

Agradeço também ao meu orientador, Prof. George Marsicano, e ao coordenador do projeto e meu chefe, Prof. Jonas Brant, por serem excelentes docentes e terem sido solícitos nas revisões deste trabalho, sempre fornecendo as melhores visões e dicas para que eu pudesse prosseguir da melhor maneira e com a maior qualidade possível.

Aos meus parceiros de estágio e de projeto, especialmente Gabriel Halabi, Daniel Gomes, João Paulo e Gleyton Lima, agradeço por todas as oportunidades de troca de conhecimento ao longo destes anos.

À minha digníssima namorada, agradeço por ser fonte de inspiração, pelo apoio e por estar sempre ao meu lado, tornando essa jornada mais leve. Luana, é com prazer que divido este capítulo da minha vida, você é verdadeiramente única e sou infinitamente grato por tê-la ao meu lado.

Por último e longe de ser menos importante, quero também agradecer ao meu falecido cachorro, Simba, que foi uma fonte de alegria e energia durante os momentos mais difíceis da graduação. Encontrado na rua, ele era a minha principal motivação para me formar e proporcionar a ele a qualidade de vida que não pude dar nos seus últimos dias. Embora ele não esteja presente para compartilhar os dias de glória, sua memória será eterna entre aqueles que o conheceram.

Para finalizar esta longa trajetória gostaria de mencionar que, além de todo o conhecimento e diploma adquiridos, nada é mais valioso que a jornada e os amigos que fazemos pelo caminho.

*"Cada um de nós é, sob uma
perspectiva cósmica, precioso"
(Carl Sagan)*

Resumo

Este trabalho teve como objetivo o desenvolvimento da estratégia de *Business Intelligence* (BI) como melhoria de processo para os gestores de saúde da Rede de Atenção Psicossocial (RAPS), auxiliando-os nas tomadas de decisão. Para implementar uma solução em BI, foram realizadas revisões bibliográficas acerca dos assuntos que compõem a metodologia, foi proposto o desenvolvimento de um Data Warehouse (DW) conjuntamente com os processos de Extração, Transformação e Carregamento (ETL) dos dados dos Sistemas de Informação em Saúde (SIS) da RAPS. A conexão entre a plataforma de BI e o DW foi realizada para permitir a geração e atualização de gráficos a partir dos dados. A metodologia de desenvolvimento propôs fases, em que cada uma alcança um objetivo específico por meio das atividades designadas. Este estudo destaca a importância das técnicas de análise de dados e seu impacto positivo nas decisões de gestão.

Palavras-chave: Business Intelligence. Rede de Atenção Psicossocial. Data Warehouse. Análise de Dados. ETL.

Abstract

This study aimed to develop a Business Intelligence (BI) strategy as a process improvement for the health managers of the Psychosocial Care Network (RAPS), assisting them in decision making. To implement a BI solution, literature reviews were conducted on the subjects that make up the methodology, and then the development of a Data Warehouse (DW) was proposed along with the data extraction, transformation and loading (ETL) processes from the RAPS Health Information Systems (SIS). The connection between the BI platform and the DW was performed to allow for the generation and updating of graphs from the data. The development methodology proposed phases, where each phase reaches a specific goal through the designated activities. This study highlights the importance of data analysis techniques and their positive impact on management decisions.

Key-words: Business Intelligence. Psychosocial Care Network. Data Warehouse. Data Analysis. ETL.

Lista de ilustrações

Figura 1 – Quantidade de CAPS no Brasil (Ministério da Saúde, 2022a, p. 6)	15
Figura 2 – Fluxo de dados do SIH (LESSA et al., 2000)	18
Figura 3 – Processo para análise de informações - Fonte: Autor	20
Figura 4 – Arquitetura típica de um Data Warehouse (VAISMAN; ZIMÁNYI, 2022, p. 67)	22
Figura 5 – Metodologia em fases - Fonte: Autor.	28
Figura 6 – Fluxo básico de um BPMN (MONTEIRO, c2023, p. 22)	32
Figura 7 – Gráfico Radar de comparação entre Power BI e Superset - Fonte: Equipe Coopere	40
Figura 8 – Reunião para levantamento de indicadores - Fonte: Autor.	41
Figura 9 – Indicador de Saúde 1 - Fonte: Equipe Coopere	42
Figura 10 – Indicador de Saúde 5 - Fonte: Equipe Coopere	42
Figura 11 – Indicador de Saúde 6 - Fonte: Equipe Coopere	43
Figura 12 – BPMN da modelagem do processo de ETL - Fonte: Autor.	45
Figura 13 – BPMN das transformações de dados - Fonte: Autor.	46
Figura 14 – Fluxo de processos - Fonte: Autor.	47
Figura 15 – Diretório de datamarts - Fonte: Autor.	49
Figura 16 – Protótipo de Alta Fidelidade. Fonte: Autor.	58
Figura 17 – Carregamento dos arquivos <i>.parquet</i> Fonte: Autor.	60
Figura 18 – Tela 1 - Menu. Fonte: Autor.	61
Figura 19 – Fontes de dados - Fonte: Autor.	62
Figura 20 – Tela 3 - Evolução de CAPS. Fonte: Autor.	63
Figura 21 – Indicador de Saúde 2 - Fonte: Equipe Coopere	72
Figura 22 – Indicador de Saúde 3 - Fonte: Equipe Coopere	73
Figura 23 – Indicador de Saúde 4 - Fonte: Equipe Coopere	74
Figura 24 – Indicador de Saúde 7 - Fonte: Equipe Coopere	75
Figura 25 – Indicador de Saúde 8 - Fonte: Equipe Coopere	76
Figura 26 – Indicador de Saúde 9 - Fonte: Equipe Coopere	77
Figura 27 – Indicador de Saúde 10 - Fonte: Equipe Coopere	78
Figura 28 – Validação dos indicadores - Fonte: Autor.	79
Figura 29 – Conexão de arquivo Parquet - Fonte: Autor.	87
Figura 30 – Tela 2 - Fonte: Autor.	88
Figura 31 – Tela 4 - Fonte: Autor.	89
Figura 32 – Tela 5 - Fonte: Autor.	90
Figura 33 – Tela 6 - Fonte: Autor.	91
Figura 34 – Tela 7 - Fonte: Autor.	92

Figura 35 – Tela 8 - Fonte: Autor.	93
Figura 36 – Tela 9 - Fonte: Autor.	94
Figura 37 – Tela 10 - Fonte: Autor.	95
Figura 38 – Tela 11 - Fonte: Autor.	96
Figura 39 – Tela 12 - Fonte: Autor.	97
Figura 40 – Tela 13 - Fonte: Autor.	98
Figura 41 – Tela 14 - Fonte: Autor.	99
Figura 42 – Tela 15 - Fonte: Autor.	100

Lista de tabelas

Tabela 1 – Metodologia em fases - Fonte: Autor.	27
Tabela 2 – Modelo da tabela que descreve um indicador - Fonte: Autor.	30
Tabela 3 – Comparação dos Tipos de Gráficos - Fonte: Equipe Coopere.	37
Tabela 4 – Recursos de visualização - Fonte: Equipe Coopere.	37
Tabela 5 – Custos entre as plataformas - Fonte: Equipe Coopere.	38
Tabela 6 – Priorização dos indicadores - Fonte: Autor	44
Tabela 7 – Classificação por épicos - Fonte: Autor.	54
Tabela 8 – Requisitos Funcionais - Fonte: Autor.	54
Tabela 9 – Requisitos Funcionais - Fonte: Autor.	55
Tabela 10 – Requisitos Não Funcionais - Fonte: Autor.	56

Lista de abreviaturas e siglas

RAPS	Rede de Atenção Psicossocial
BI	Business Intelligence
DW	Data Warehouse
ETL	Extract, Transform, Load
OLAP	OnLine Analytical Processing
MOLAP	Multidimensional OLAP
ROLAP	Relational OLAP
BPMN	Business Process Modeling and Notation
URL	Uniform Resource Locator
TI	Tecnologia da Informação
MS	Ministério da Saúde
SUS	Sistema Único de Saúde
APS	Atenção Primária à Saúde
SIH	Sistema de Informação Hospitalar
SCNES	Sistema de Cadastro Nacional de Estabelecimentos de Saúde
SISAB	Sistema de Informação em Saúde para a Atenção Básica
e-SUS AB	Estratégia e-SUS para Atenção Básica
CAPS	Centro de Atenção Psicossocial
SIS	Sistemas de Informação em Saúde
SDS	Sala de Situação
FS	Faculdade de Saúde
UnB	Universidade de Brasília
AIH	Autorização de Internação Hospitalar

CID	Classificação Internacional de Doenças
LGPD	Lei Geral de Proteção de Dados Pessoais
RAD	Rapid Application Development
CLI	Command Line Interface
SDK	Software Development Kit

Sumário

1	INTRODUÇÃO	9
1.1	Contexto	9
1.2	Justificativa	10
1.3	Problema	11
1.4	Objetivos	11
1.4.1	Objetivo Geral (OG)	11
1.4.2	Objetivos específicos (OE)	11
1.5	Metodologia	11
1.6	Estrutura do Trabalho	12
2	REFERENCIAL TEÓRICO	13
2.1	Rede de Atenção Psicossocial	13
2.2	Sistemas de Apoio a RAPS	15
2.2.1	Sistema de Informações Hospitalares - SIH	16
2.2.2	Sistema de Informação em Saúde para a Atenção Básica - SISAB	18
2.2.3	Sistema de Cadastro Nacional de Estabelecimentos de Saúde - SCNES	19
2.2.4	Comunicação entre os sistemas	19
2.3	Data Warehouse	20
2.4	<i>Business Intelligence</i>	23
2.5	Lei Geral de Proteção de Dados Pessoais (LGPD)	24
3	METODOLOGIA UTILIZADA	26
3.1	Fase 1	28
3.2	Fase 2	31
3.3	Fase 3	33
3.4	Fase 4	35
4	RESULTADOS	36
4.1	Fase 1	36
4.1.1	Atividade 1 - Selecionar a plataforma de BI	36
4.1.2	Atividade 2 - Realizar reuniões com integrantes do projeto	40
4.1.3	Atividade 3 - Documentar os indicadores	41
4.1.4	Atividade 4 - Validar os indicadores	43
4.1.5	Atividade 5 - Priorizar os indicadores	44
4.2	Fase 2	44
4.2.1	Atividade 6 - Definir BPMN	45

4.2.2	Atividade 7 - Definir modelo conceitual de ETL	46
4.2.3	Atividade 8 - Implementar os serviços de ETL	47
4.2.4	Atividade 9 - Definir arquitetura do Data Warehouse	53
4.3	Fase 3	57
4.3.1	Atividade 10 - Desenvolver protótipo	57
4.3.2	Atividade 11 - Criar conexão com o DW	59
4.3.3	Atividade 12 - Desenvolver o dashboard no Power BI	60
4.4	Fase 4	63
4.4.1	Atividade 13 - Descrever a metodologia de análise	63
4.4.2	Atividade 14 - Pontuar os objetos de análise	64
4.4.3	Atividade 15 - Desenvolver a análise postmortem	64
5	CONSIDERAÇÕES FINAIS	66
	REFERÊNCIAS	67
	ANEXOS	71
	ANEXO A – INDICADOR DE SAÚDE 2 - IS02	72
	ANEXO B – INDICADOR DE SAÚDE 3 - IS03	73
	ANEXO C – INDICADOR DE SAÚDE 4 - IS04	74
	ANEXO D – INDICADOR DE SAÚDE 7 - IS07	75
	ANEXO E – INDICADOR DE SAÚDE 8 - IS08	76
	ANEXO F – INDICADOR DE SAÚDE 9 - IS09	77
	ANEXO G – INDICADOR DE SAÚDE 10 - IS10	78
	ANEXO H – VALIDAÇÃO DOS INDICADORES VIA JITSI	79
	ANEXO I – CÓDIGO DE CONVERSÃO DE .DBC PARA .CSV EM R	80
	ANEXO J – FLUXO DE TRABALHO NO GITHUB ACTIONS	82
	ANEXO K – TRANSFORMAÇÃO DE DADOS DO SIH EM PYTHON	84
	ANEXO L – SCRIPT <i>UPLOAD_FILE.SH</i>	86

ANEXO M – ARQUIVO PARQUET NO POWER BI	87
ANEXO N – TELA 2	88
ANEXO O – TELA 4	89
ANEXO P – TELA 5	90
ANEXO Q – TELA 6	91
ANEXO R – TELA 7	92
ANEXO S – TELA 8	93
ANEXO T – TELA 9	94
ANEXO U – TELA 10	95
ANEXO V – TELA 11	96
ANEXO W – TELA 12	97
ANEXO X – TELA 13	98
ANEXO Y – TELA 14	99
ANEXO Z – TELA 15	100

1 Introdução

1.1 Contexto

Em um contexto em que a tecnologia pode auxiliar em processos que melhoram a qualidade da saúde pública, surge a necessidade de armazenar dados oriundos de procedimentos médicos, internações, nascimentos, óbitos e outras fontes. Com o objetivo de fornecer informações para análise e compreensão mais aprofundada de questões de saúde pública, bem como subsidiar a tomada de decisões nos âmbitos municipal, estadual e federal, foram desenvolvidos os Sistemas de Informação em Saúde (SIS), que consistem em instrumentos padronizados de monitoramento e coleta de dados ([Biblioteca Virtual em Saúde - Ministério da Saúde, 2008](#)).

Na década de 90, houve a implantação de sistemas de informação voltados para a assistência de pacientes ([CAVALCANTE; SILVA; FERREIRA, 2011](#)), tal como o SINASC (Sistema de Informações sobre Nascidos Vivos) implantando a partir de 1990 com o objetivo de coletar dados sobre os nascimentos ocorridos em todo o território nacional e fornecer informações sobre natalidade para todos os níveis do Sistema de Saúde ([Ministério da Saúde, 2021c](#)). Existem outros sistemas de informação em saúde que, assim como o SINASC, armazenam dados diariamente por meio de métodos distintos e com critérios próprios de operação para vinculação com a base nacional.

No âmbito dos serviços psicossociais, destaca-se a Rede de Atenção Psicossocial (RAPS), instituída pela Portaria MS/GM nº 3088 de 23/12/2011, que visa à criação, expansão e articulação de pontos de atenção à saúde destinados a pessoas com transtornos mentais e necessidades decorrentes do uso de drogas ou álcool. A RAPS integra o SUS e tem como objetivo a ampliação do acesso à atenção psicossocial, promover a vinculação das pessoas com transtornos mentais e problemas decorrentes do uso de drogas ou álcool com suas famílias e garantir a articulação dos pontos de atenção das redes de saúde ([Ministério da Saúde SC, 2018](#)).

A RAPS está presente em Centros de Atenção Psicossocial, Unidades Básicas de Saúde, UPAS, SAMU, Hospitais Especializados, Serviços Residenciais terapêuticos, entre outros componentes ([Ministério da Saúde SC, 2018](#)). Considerando que cada componente de saúde utiliza sistemas diferentes, os dados referentes à Rede de Atenção Psicossocial encontram-se dispersos. Os dados relacionados à rede estão presentes no seguintes sistemas: SIH (Sistema de Informações Hospitalares), SISAB (Sistema de Informação em Saúde para a Atenção Básica), e-SUS AB e SCNES (Sistema de Cadastro Nacional de Estabelecimentos de Saúde).

1.2 Justificativa

Para adquirir dados sobre um determinado tema, o gestor de saúde da RAPS precisa, manualmente, realizar o download das informações em um sistema específico e, posteriormente, buscar a informação necessária dentre os dados brutos.

Um exemplo desse processo ocorre quando o gestor precisa obter informações associadas à quantidade de internações relacionadas a transtornos psicológicos. Nesse caso, é necessário que o gestor realize o download dos dados do SIH referentes ao ano de análise e, então, realize uma busca manual a fim de identificar os dados desejados com base no tipo de internação. Um processo semelhante é aplicado quando o gestor necessita de informações acerca da quantidade de procedimentos psicossociais em determinado estabelecimento de saúde, mas neste caso os dados são obtidos por meio do sistema CNES (Cadastro Nacional de Estabelecimentos de Saúde). Em resumo, sempre que o gestor da RAPS necessita de informações para tomar decisões ou acompanhar internações, é necessário realizar um processo manual e dispendioso de busca e seleção de dados.

A dificuldade reportada se torna mais evidente no contexto da análise de dados de um município ou estado específico, uma vez que os painéis de dados desenvolvidos pelas Secretarias de Saúde abrangem informações de todo o território nacional.

Com o apoio do Ministério da Saúde (MS), a Sala de Situação (SDS) da Faculdade de Saúde (FS) da Universidade de Brasília (UnB), inicia o projeto denominado "Saúde Mental: Ferramentas para contribuir com a otimização dos processos de trabalho dos Centros de Atenção Psicossocial". Este projeto está dividido em duas grandes metas, onde a meta 1 é a criação de uma comunidade de práticas voltada para o fortalecimento da RAPS, a comunidade Coopere ([Comunidade Coopere, 2023](#)).

A meta 2, a qual é objeto de estudo desta monografia, visa a construção de uma plataforma para análise de dados por meio de painéis de monitoramento, voltada para a vigilância de saúde mental. Essa meta tem como propósito principal: Prover Painéis de Monitoramento (Dashboard), com a utilização do conceito de *Business Intelligence* (BI) ([Comunidade Coopere, 2023](#)).

Já a SDS é um ambiente colaborativo que possui o objetivo de auxiliar na análise, monitoramento e definição de ações em saúde, visando ser uma experiência estadual, nacional e internacional, e auxiliando estudantes e gestores na tomada de decisão. Ela se baseia em quatro eixos fundamentais para o enfrentamento de cenários de crise: gestão, comunicação, vigilância e tecnologia da informação. Além disso, a SDS tem a missão de ser um espaço de suporte pedagógico e tecnológico para ensino, pesquisa e extensão ([Sala de Situação UnB, c2023](#)).

1.3 Problema

Tendo como base o contexto acima, fica evidenciado que, no cenário atual, os gestores de saúde da RAPS precisam consultar diversos sistemas para adquirir dados e informações relevantes para suas tomadas de decisão e que, decorrente do problema dos dados dispersos, ainda é um trabalho manual e que demanda tempo.

Alinhado a isso, o presente trabalho visa responder a seguinte questão de pesquisa: Como disponibilizar aos gestores da RAPS, de forma automática e centralizada, informações para a tomada de decisão?

1.4 Objetivos

1.4.1 Objetivo Geral (OG)

O objetivo geral deste trabalho é: prover informações da Rede de Atenção Psicossocial (RAPS) em forma de painéis, de forma automática e centralizada, auxiliando os gestores das RAPS na tomada de decisões.

1.4.2 Objetivos específicos (OE)

- OE1 - Identificar os requisitos do dashboard;
- OE2 - Centralizar as informações dos SIS;
- OE3 - Prover painéis de monitoramento;
- OE4 - Analisar os estudos e painéis desenvolvidos.

Onde, "Identificar os requisitos do dashboard" é identificar e listar os requisitos para o desenvolvimento do dashboard, tais como a plataforma de BI e variáveis de saúde que são pertinentes para os gestores. "Centralizar as informações dos SIS" é unificar as informações dos sistemas SIH, SISAB, e-SUS AB e SCNES, para que tais informações sejam consumidas pelo dashboard. "Prover painéis de monitoramento" é a etapa de construção e disponibilização dos painéis de monitoramento. "Analisar os estudos e painéis desenvolvidos" é a análise postmortem de tudo já desenvolvido, portanto, deve ser a última etapa.

1.5 Metodologia

A metodologia do projeto está dividida em três fases, onde cada uma visa resolver um objetivo específico do projeto. Cada fase é composta pelas seguintes atividades:

- Fase 1 (OE1) - Identificar os requisitos do dashboard
 - Atividade 1 (A1): Selecionar a plataforma de *Business Intelligence*;
 - Atividade 2 (A2): Realizar reuniões com integrantes do projeto;
 - Atividade 3 (A3): Documentar os indicadores;
 - Atividade 4 (A4): Validar os indicadores;
 - Atividade 5 (A5): Priorizar os indicadores.
- Fase 2 (OE2) - Centralizar as informações dos SIS
 - Atividade 6 (A6): Definir BPMN;
 - Atividade 7 (A7): Definir modelo conceitual de ETL;
 - Atividade 8 (A8): Implementar serviços de ETL;
 - Atividade 9 (A9): Definir arquitetura do DW;
- Fase 3 (OE3) - Prover painéis de monitoramento
 - Atividade 10 (A10): Desenvolver protótipo;
 - Atividade 11 (A11): Criar conexão com o DW;
 - Atividade 12 (A12): Desenvolver o dashboard no Power BI.
- Fase 4 (OE4) - Analisar os estudos e painéis desenvolvidos
 - Atividade 13 (A13): Descrever a metodologia da análise;
 - Atividade 14 (A14): Pontuar os objetos de análise;
 - Atividade 15 (A15): Desenvolver a análise postmortem;

A estrutura de desenvolvimento está detalhada no Capítulo 3 (Metodologia).

1.6 Estrutura do Trabalho

A partir deste ponto, o presente trabalho está estruturado como segue:

- Capítulo 2, intitulado "Referencial Teórico", serão descritos os principais conceitos e sistemas abordados na introdução;
- Capítulo 3, intitulado "Metodologia", serão descritos os procedimentos e métodos utilizados na pesquisa, incluindo detalhes sobre a Seção 1.5, de forma que seja claro o processo de trabalho planejado, as atividades realizadas e os resultados esperados;

- Capítulo 4, intitulado "Resultados", serão apresentados todos os tipos de resultados, artefatos e documentação desenvolvida. Por fim, neste capítulo também é abordada uma análise do estudo desenvolvido;
- Capítulo 5, intitulado "Considerações Finais", serão resumidas todas as atividades e produtos obtidos até o momento da escrita do capítulo.

2 Referencial Teórico

2.1 Rede de Atenção Psicossocial

Instituída pela Portaria N° 3088 de 23 de Dezembro de 2011, a Rede de Atenção Psicossocial surge para pessoas com sofrimento ou transtorno mental e com necessidades decorrentes do uso de crack, álcool e outras drogas, no âmbito do Sistema Único de Saúde (SUS). Sua finalidade é a criação, ampliação e articulação de pontos de atenção à saúde para pessoas com sofrimento ou transtorno mental e com necessidades decorrentes do uso de crack, álcool e outras drogas, no âmbito do SUS ([Ministério da Saúde, 2011](#)).

De acordo com a portaria, as unidades de saúde devem ter uma equipe profissional, incluindo profissionais de saúde mental, para garantir a integralidade do cuidado ao paciente. Além disso, as unidades de saúde também devem ter protocolos estabelecidos para a atenção a pacientes com transtornos mentais e de uso de substâncias ([Ministério da Saúde, 2011](#)).

É importante notar que a estrutura e disponibilidade de serviços de RAPS varia de acordo com a região, mas de modo geral, segundo o Ministério da Saúde ([Ministério da Saúde, 2017](#)), a rede é composta por diversos estabelecimentos de saúde, incluindo:

- Centros de Atenção Psicossocial (CAPS): oferecem atendimento ambulatorial para pessoas com transtornos mentais graves e persistentes, como esquizofrenia, transtorno bipolar e transtorno depressivo. Os CAPS são divididos em:
 - CAPS I - oferece atendimento para indivíduos com transtornos mentais graves e persistentes, bem como necessidades relacionadas ao uso de crack, álcool e outras drogas, de todas as idades. É indicado para cidades com mais de vinte mil habitantes;
 - CAPS II - oferece mesmo tipo de atendimento que o CAPS I, mas indicado para municípios com mais de sessenta mil habitantes;
 - CAPS III - oferece atendimento para indivíduos com transtornos mentais graves e persistentes, oferecendo serviços de atenção contínua, com funcionamento 24 horas, incluindo feriados. Oferece suporte clínico e acolhimento noturno para outros serviços de saúde mental, incluindo o CAPS AD, recomendado para cidades ou regiões com população acima de duzentos mil habitantes.
 - CAPSi (CAPS Infanto-juvenil) - atende crianças e adolescentes com transtornos mentais graves e persistentes e com necessidades decorrentes do uso de drogas

como crack, álcool e outras. Indicado para cidades ou regiões com população acima de cento e cinquenta mil habitantes.

- CAPS AD - atende tanto adultos quanto crianças e adolescentes, seguindo as normativas estabelecidas pelo Estatuto da Criança e do Adolescente, com necessidades decorrentes do uso de drogas como crack, álcool e outras. Este é um serviço de saúde mental aberto e comunitário, indicado para cidades ou regiões com população acima de setenta mil habitantes.
 - CAPS AD III - oferece atendimento para adultos ou crianças e adolescentes, de acordo com as normas do Estatuto da Criança e do Adolescente, que necessitam de cuidados clínicos contínuos. Possui até 12 leitos para monitoramento, com funcionamento 24 horas, incluindo feriados e fins de semana, recomendado para cidades ou regiões com população acima de duzentos mil habitantes.
 - CAPS AD IV - oferece atendimento para adultos ou crianças e adolescentes, considerando as normativas do Estatuto da Criança e do Adolescente, que precisam de atenção integral e cuidados clínicos contínuos relacionados ao consumo de álcool, crack e outras drogas. Ele tem funcionamento 24 (vinte e quatro) horas por dia em todos os dias da semana, e é indicado para municípios com população acima de quinhentos mil habitantes.
- Serviços de Saúde Mental na Atenção Básica: os profissionais de saúde mental nas unidades de saúde da atenção básica fornecem atendimento ambulatorial e acompanhamento para pessoas com transtornos mentais leves e moderados;
 - Hospital Psiquiátrico: oferece atendimento hospitalar para pessoas com transtornos mentais graves e persistentes;
 - Residências Terapêuticas: proporcionam atendimento comunitário para pessoas com transtornos mentais graves e persistentes que precisam de suporte adicional após a alta hospitalar;
 - Serviços de saúde mental para população privada de liberdade, como presídios e unidades socioeducativas;
 - Núcleos de Apoio à Saúde da Família (NASF): equipe multiprofissional que atua junto à equipe de Atenção Básica, com o objetivo de fortalecer a atenção à saúde mental na Atenção Básica;

De acordo com o [Ministério da Saúde \(2022a, p. 6\)](#):

No momento atual, o Brasil conta com 2.836 (dois mil, oitocentos e trinta e seis) Caps habilitados, distribuídos em 1.910 (um mil, novecentos e dez) municípios em todos os Estados e no Distrito Federal, totalizando um

investimento de incentivo de custeio anual de R\$ 1.274.270.328,00 (um bilhão, duzentos e setenta e quatro milhões, duzentos e setenta mil e trezentos e vinte e oito reais) para essa modalidade de serviço.

A Figura 1 mostra o crescimento da quantidade de CAPS ao longo dos anos. No entanto, existem diferentes tipos de estabelecimentos de saúde, e a RAPS está presente em vários deles, oferecendo diferentes serviços de saúde. Como resultado, os dados sobre saúde mental e emocional são dispersos entre vários sistemas de informação, dependendo do estabelecimento/serviço de saúde.

Série histórica do número de CAPS habilitados no Brasil por Tipo (Brasil, dez/2010 a junho/2022)								
ANO	CAPS AD	CAPS ADIII	CAPS AD IV	CAPS I	CAPS II	CAPS III	CAPS infanto-juvenil	Total Geral
2010	190	1	-	712	383	57	126	1.469
2011	201	6	-	779	403	65	147	1.601
2012	231	29	-	873	437	73	172	1.815
2013	256	47	-	953	447	81	188	1.972
2014	274	68	-	1.044	456	88	200	2.130
2015	289	85	-	1.118	470	94	209	2.265
2016	304	104	-	1.184	490	100	226	2.408
2017	316	110	-	1.240	502	106	240	2.514
2018	316	111	-	1.257	507	107	248	2.546
2019	319	121	3	1.313	515	112	256	2.639
2020	331	128	3	1.366	526	121	274	2.749
2021	330	132	3	1.402	524	122	282	2.795
2022	331	135	3	1.428	528	126	285	2.836

Figura 1 – Quantidade de CAPS no Brasil (Ministério da Saúde, 2022a, p. 6)

2.2 Sistemas de Apoio a RAPS

Segundo a Organização Mundial da Saúde (OMS), um Sistema de Informação em Saúde é:

“um conjunto de componentes que atuam de forma integrada por meio de mecanismos de coleta, processamento, análise e transmissão da informação necessária e oportuna para implementar processos de decisões no Sistema de Saúde. Seu propósito é selecionar dados pertinentes e transformá-los em informações para aqueles que planejam, financiam, proveem e avaliam os serviços de saúde” (CRUZ, 1988).

De acordo com Santos (2009), um SIS é formado por componentes que coletam, processam, analisam e transmitem informações em saúde, visando transformá-las em informações precisas para a tomada de decisões. Esses componentes precisam funcionar de forma integrada com os outros aspectos do SIS, tais como recursos humanos, materiais e equipamentos.

Segundo Medeiros et al. (2004), os Sistemas de Informação em Saúde, sejam eles assistenciais ou epidemiológicos, são considerados ferramentas importantes para identificar situações de saúde e promover intervenções personalizadas de acordo com as necessidades da população. Santos (2009) complementa que estes sistemas podem ser classificados em dois grupos: o primeiro é destinado a fornecer informações para a tomada de decisões gerenciais em cada serviço, enquanto o segundo visa coletar informações para apresentação aos gestores ou financiadores, com o objetivo de atender às demandas.

No estudo de Rochel, Medina e Moreno (2007), os bancos de dados gerados pelos SIS foram categorizados em três tipos: epidemiológicos, administrativos e clínicos, onde os 'epidemiológicos' são elaborados e empregados para monitorar, avaliar e investigar, reunindo informações sobre ocorrências importantes (nascimentos e mortes), enfermidades e afeções obrigatórias para notificação (por exemplo, dengue), entre outros. Os bancos 'administrativos' são criados para fins de contabilidade e monitoramento dos serviços de saúde prestados. E por fim, os 'clínicos' são dados demográficos dos pacientes atendidos, os procedimentos realizados e os diagnósticos que são registrados em bancos de dados que têm o objetivo de avaliar, vigiar e pesquisar. Embora o nome sugere, elas não fornecem dados clínicos, pois foram desenvolvidos para armazenar informações de natureza multimídia geradas ao longo do contato de um paciente com diferentes serviços de saúde ou de diferentes unidades de saúde, tais como resultados de exames, medidas antropométricas, entre outros. Além disso, a disponibilidade dos dados do tipo 'clínicos' é restrita.

Diante disso, para o desenvolvimento da meta 2 do projeto mencionado, três sistemas de informação em saúde foram selecionados a fim de que sejam extraídos os dados. A seleção desses sistemas foi baseada nos seguintes aspectos: existência de informações da RAPS na base de dados, a relevância dessas informações nos painéis de monitoramento e disponibilidade dos dados abertos.

2.2.1 Sistema de Informações Hospitalares - SIH

Com o objetivo de fornecer informações sobre o uso dos serviços hospitalares para fins de planejamento, avaliação e regulação da atenção à saúde, o Sistema de Informação Hospitalar (SIH) registra informações sobre internações hospitalares, incluindo informações demográficas sobre pacientes, diagnósticos, procedimentos realizados, dados financeiros, entre outros. Ele também pode registrar informações sobre a qualidade do atendimento, incluindo tempo de espera para consultas, tempo de permanência em leitos hospitalares,

e taxa de mortalidade (GUIMARAES, 2020).

O SIH é utilizado por profissionais da área da saúde, como médicos, enfermeiros e administradores hospitalares, para gerenciar informações sobre pacientes (como informações demográficas, histórico médico, resultados de exames) que são importantes para garantir a qualidade dos cuidados prestados aos pacientes e para auxiliar na tomada de decisões clínicas e gerenciais.

Existem dois conceitos fundamentais para o entendimento de como será utilizado a extração e manipulação de dados do SIH voltados para o contexto psicossocial, estes são: Autorização de Internação Hospitalar (AIH) e Classificação Internacional de Doenças (CID).

A AIH é utilizada para identificar pacientes e serviços prestados sob o regime de internação hospitalar e fornecer informações para o gerenciamento do Sistema de Informação Hospitalar. Este formulário é um documento do Sistema Único de Saúde (SUS) e é preenchido pelos hospitais para reembolso da assistência prestada a pacientes financiados com recursos públicos. A AIH é gerada quando ocorre uma internação em um local com convênio com o sistema, seja ele público ou privado, e é enviada ao gestor da unidade que presta os serviços. Os gestores precisam então enviar mensalmente ao Ministério da Saúde um arquivo contendo dados de todas as internações ocorridas no país e a partir da autorização, é possível viabilizar o faturamento dos serviços hospitalares prestados ao SUS e ela só pode ser emitida por órgãos próprios ou autorizados pelo sistema (SAUDI, c2023).

Já o CID, segundo Neumam (2021), significa "um sistema de códigos, criado pela OMS, utilizado no mundo todo para padronizar a linguagem entre os médicos, além de monitorar a incidência e a prevalência de cada doença [...] sempre que um paciente apresenta uma enfermidade, uma doença, o médico escreve isso sob forma de um código", e é por meio dessa classificação que é possível filtrar os dados do SIH para obter internações por motivos psicológicos, por exemplo.

A Figura 2 apresenta o processo de fluxo de dados no SIH, desde a obtenção de informações do paciente, exames e motivo de internação por meio de consulta e laudo médico, até a análise e inclusão dos dados na base nacional, incluindo a emissão da AIH.

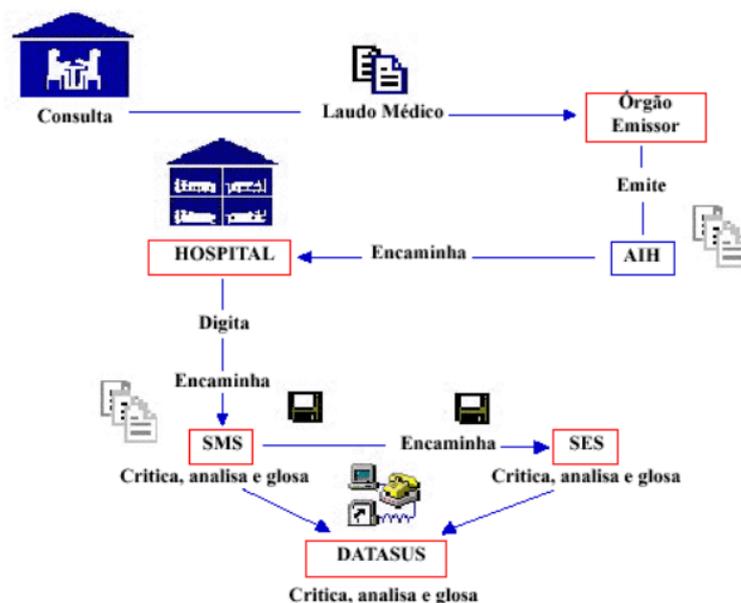


Figura 2 – Fluxo de dados do SIH (LESSA et al., 2000)

Os gestores de saúde acessam as informações do SIH por meio de sistemas internos do Ministério da Saúde ou Secretarias Estaduais de Saúde, mas também existem os dados abertos do SIH, disponibilizados no DATASUS (Departamento de Informática do Sistema Único de Saúde do Brasil), onde é possível realizar o download por meio do TabWin e TabNet, aplicativos que facilitam a tabulação e o cruzamento de informações, permitindo aos gestores, estudantes e público em geral obter dados relacionados ao SUS.

2.2.2 Sistema de Informação em Saúde para a Atenção Básica - SISAB

Segundo o [Ministério da Saúde \(2022b\)](#):

"O Sistema de Informação em Saúde para a Atenção Básica (SISAB) foi instituído pela Portaria GM/MS nº 1.412, de 10 de julho de 2013, passando a ser o sistema de informação da Atenção Básica vigente para fins de financiamento e de adesão aos programas e estratégias da Política Nacional de Atenção Básica, substituindo o Sistema de Informação da Atenção Básica (SIAB)".

A atenção primária é responsável por várias funções no sistema único de saúde, e com o decorrer dos anos, a necessidade de ter um sistema que possibilite a coleta, armazenamento e análise de informações referentes. Assim, a partir de 2013, instituiu o SIAB, que é uma estratégia do Departamento de Saúde da Família do Ministério da Saúde. O sistema funciona com o objetivo inicial de auxiliar na qualidade de coleta de dados, utilizando o e-SUS AB, que tem sua estrutura customizada para coleta simplificada, e também com outra tecnologia que é o PEC (Prontuário Eletrônico do Cidadão) para

garantir um acesso automatizado e eletrônico. O sistema tem os princípios de reduzir o retrabalho dos pacientes, unificar bases, criar aprimorações de acordo com as necessidades dos usuários, individualização do paciente e integralidade no território ([Projeto Brumadinho UFMG, 2021](#))

O e-SUS AB é projetado para ser um sistema que facilita a interoperabilidade com outras plataformas, utilizando como principal tecnologia o JavaWeb. O sistema apresenta módulos de cadastro, territorialização, atendimento individual, apoio a gestão e entre outros para facilitar a tomada de decisão dos gestores, já que a partir dele os dados de saúde disponibilizados, bem como de relatórios de indicadores de saúde podendo visualizados por estado, município, região e equipes de saúde ([Conass, 2013](#)).

2.2.3 Sistema de Cadastro Nacional de Estabelecimentos de Saúde - SCNES

De acordo com o [Ministério da Saúde \(2021b\)](#), o CNES (Cadastro Nacional de Estabelecimentos) visa ser a base para operacionalizar os SIS, sendo imprescindível para o gerenciamento eficaz do SUS.

O CNES visa ainda dar transparência, pelo site, de toda a infraestrutura de serviços de saúde, também a capacidade instalada existente e disponível no país. Ser, junto com o CNS, o principal elo entre todos os sistemas do SUS ([Ministério da Saúde, 2021b](#)).

O principal motivo do uso de dados do SCNES nos painéis de monitoramento é por disponibilizar informações de infraestrutura, tipo de atendimento prestado, serviços especializados, leitos e profissionais de saúde existentes nos estabelecimentos de saúde ([Ministério da Saúde, 2021b](#)), dados que são cruciais para a análise de informação para os gestores de saúde.

2.2.4 Comunicação entre os sistemas

Tendo em vista tamanha dimensão da Rede de Atenção Psicossocial e dos sistemas de informação que armazenam os dados, reforça-se que existe uma dificuldade por parte dos gestores de saúde para acessar estes dados. Além de ser necessário trabalho manual para obtenção de dados, cada Sistema de Informação em Saúde possui um formato de exportação de dados diferentes, como por exemplo o Sistema de Informação Hospitalar (SIH), que exporta arquivos no formato .dbc e, por ser uma extensão específica dos SIS, muitos profissionais de saúde precisam se especializar na área de tecnologia, como mostra o processo da [Figura 3](#).

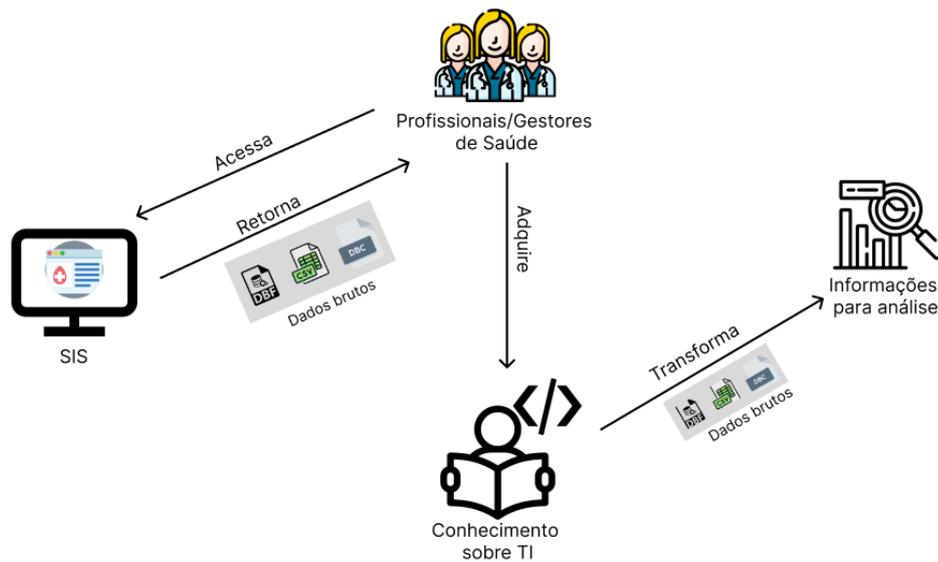


Figura 3 – Processo para análise de informações - Fonte: Autor

É possível observar que a dificuldade em acessar os dados da Rede de Atenção Psicossocial é decorrente de problemas no processo atual, e pode ser resolvida por meio de melhorias nos processos de coleta e armazenamento desses dados, otimizando o trabalho dos profissionais de saúde que precisam se submeter ao gasto de recursos (como tempo e energia) para dominar as tecnologias e então realizar tomadas de decisão.

Dessa forma, como um dos objetivos específicos do trabalho, a centralização dos dados em um ambiente virtual é necessária para que haja a devida tratativa de dados e então a geração de gráficos. Para isso existe um conceito que funciona como repositório de armazenamento chamado Data Warehouse, muito utilizado para integrar dados de diferentes fontes.

2.3 Data Warehouse

Uma estratégia muito utilizada para consolidação e armazenamento de banco de dados oriundos de diversas fontes é o Data Warehouse (DW), que funciona como um repositório central de informações especializado em consultas e análises gerenciais.

Segundo [Vaisman e Zimányi \(2022, p. 4\)](#), o DW surgiu na década de 1990, em um contexto onde o mundo se tornava cada vez mais competitivo e em constante mudança, então as organizações perceberam que precisavam realizar análises de dados sofisticadas para apoiar seus processos de tomada de decisão. Porém, os bancos de dados operacionais ou transacionais tradicionais não satisfaziam os requisitos para análise de dados, já que foram projetados e otimizados para suportar operações de negócios diárias. Além disso, os bancos de dados tradicionais não contêm dados históricos e possuem a performance

reduzida quando executam tarefas complexas ou quando submetidos a um enorme volume de dados.

As características de um Data Warehouse estão incluídas na definição clássica por Inmon (INMON, 2005 apud VAISMAN; ZIMÁNYI, 2022), onde um DW é caracterizado como uma coleção de dados orientados a assunto, integrados, não voláteis e variáveis no tempo para apoiar decisões gerenciais. Orientado a assuntos por ter como objetivo um ou vários assuntos de uma empresa, como por exemplo área de vendas e de recursos humanos. Integrado significa que o conteúdo de um DW resulta da integração de dados de vários sistemas. Não volátil indica que um DW acumula dados de sistemas por um longo período de tempo.

Estas características podem ser facilmente observadas no contexto do projeto. Devido a sua orientação por assunto, é possível armazenar e analisar sistemas de saúde sob diversos pontos de vista, como os próprios bancos de dados de cada Sistema de Informação em Saúde sugerem na Seção 2.2, incluindo bancos de dados epidemiológicos, administrativos e clínicos.

A característica de ser um repositório integrado está diretamente ligada aos três Sistemas de Informação em Saúde propostos para compor os painéis de monitoramento. Finalmente, a característica de ser não volátil é fundamental, especialmente para análises ao longo de vários anos, que são cruciais no contexto da saúde.

Para a compreensão adequada da arquitetura de um Data Warehouse, é importante ter clareza sobre conceitos ligados diretamente ao repositório de dados, tais como: Extract, Transform and Load (ETL) e OnLine Analytical Processing (OLAP).

Como um Data Warehouse armazena dados oriundos de fontes distintas, esses dados precisam ser transformados para se adaptarem ao modelo do DW e, posteriormente, carregados no mesmo. Este processo é conhecido como ETL e é fundamental para o sucesso de um projeto de DW, conforme demonstrado na obra de Vaisman e Zimányi (2022, p. 6).

Na etapa de extração, os dados são identificados e copiados de suas origens, de forma que seja possível transportá-los para o armazenamento destino. A etapa de transformação consiste em preparar os dados para o armazenamento, tendo em vista que tais dados são extraídos em sua forma bruta, original. Na etapa de carregamento os dados já transformados são movidos para o armazenamento destino, onde os dados podem ser carregados por inteiro ou apenas as alterações incrementais nos dados de origem. Além disso, os dados podem ser carregados em tempo real ou programados por lotes (ORACLE, c2023), sendo esta última opção ideal para o contexto da saúde, onde as informações são analisadas periodicamente ao longo de semanas, meses e anos.

Segundo Vaisman e Zimányi (2022, p. 7), o OLAP está diretamente ligado com o processo de análise de dados, que consiste em explorar o conteúdo de um armazém de

dados para fornecer informações importantes para o processo de tomada de decisão. O paradigma OLAP é utilizado na extração de dados relevantes do armazém para descobrir conhecimentos úteis que não são fáceis de obter dos dados originais.

De acordo com Wylton (2009), o OLAP possui quatro funções básicas: Visualização multidimensional dos dados, Exploração, Rotação e modos de visualização. Além disso, o OLAP possui quatro métodos de armazenamento de dados, sendo os mais comuns o OLAP Relacional (ROLAP) e o OLAP Multidimensional (MOLAP). Cada modelo se distingue pela velocidade de consulta e pela forma de armazenamento dos dados, relacional e multidimensional, respectivamente.

De acordo com Vaisman e Zimányi (2022, p. 67) e com a Figura 4, a arquitetura típica de um Data Warehouse é composta pelas seguintes camadas: fonte dos dados, Back-end, DW, OLAP e camada do Front-end.

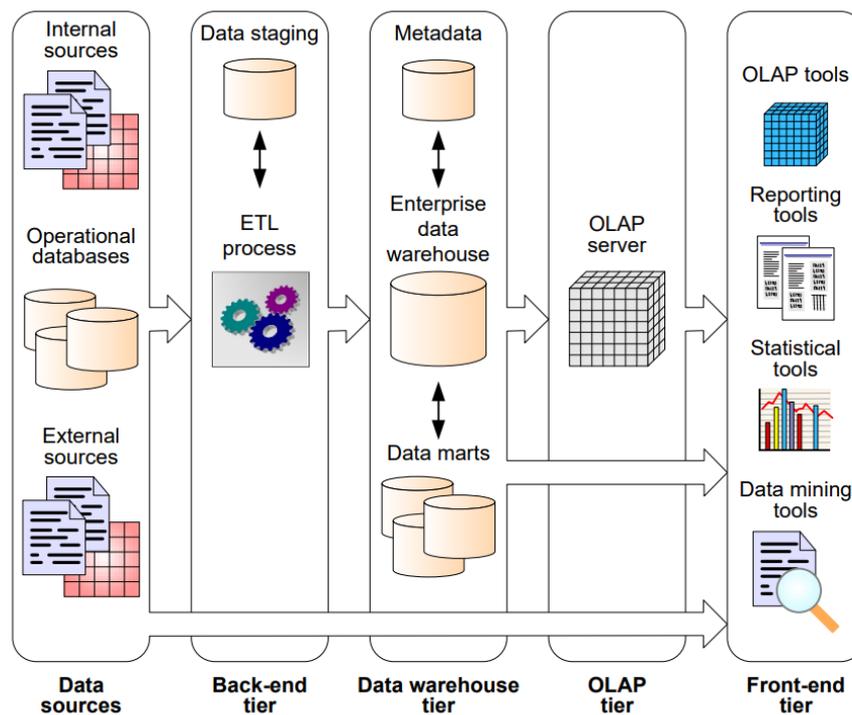


Figura 4 – Arquitetura típica de um Data Warehouse (VAISMAN; ZIMÁNYI, 2022, p. 67)

- Fonte dos dados - camada que possui os dados, que podem ser internos ou externos à organização;
- Back-end - camada composta pelas ferramentas de ETL e é utilizada para alimentar os dados do DW a partir da camada de fonte de dados;
- Data Warehouse - camada formada pelo *enterprise data warehouse* (engloba toda a organização) e *data mart* (voltado para um departamento ou área da organização).

Também é composto pelo repositório de *metadata* (metadados), que é definido com "dados sobre dados";

- OLAP - camada composta por um servidor OLAP, que apresenta aos usuários os dados multidimensionais, independente da forma real como os dados são armazenados;
- Front-end - camada usada para análise e visualização dos dados, contendo ferramentas que permitem aos usuários explorarem o conteúdo do DW. As ferramentas descritas na camada são:
 - Ferramentas OLAP: permitem a manipulação dos dados e exploração interativa;
 - Ferramentas de relatórios: permitem a produção, entrega e gerenciamento de relatórios, utilizando consultas predefinidas;
 - Ferramentas estatísticas: permitem a análise e visualização de dados por meio de métodos estatísticos;
 - Ferramentas de DM: permitem que usuários analisem dados que seguem padrões, tendências e previsões feitas com base nos dados atuais (VAISMAN; ZIMÁNYI, 2022, p. 70).

2.4 Business Intelligence

A partir da definição de *Business Intelligence* (BI) de Dresner (1989):

"é uma metodologia pela qual se estabelecem ferramentas para obter, organizar, analisar e prover acesso às informações necessárias aos tomadores de decisão das empresas para analisarem os fenômenos acerca de seus negócios",

Braghittoni (2017) realiza uma análise sobre a definição e reforça que o conceito de BI, por ser uma metodologia, pode ser utilizado com qualquer ferramenta de controle de dados, desde que esteja alinhado a metodologia. Além disso, reforça também que BI é utilizado para analisar fenômenos acerca de negócios, tais como o faturamento do mês da empresa, e compará-lo com o faturamento dos meses anteriores, por exemplo.

Dessa forma, entende-se que a metodologia de *Business Intelligence* não se limita apenas a agregação de dados, mas sim ao conhecimento adquirido a partir dessas informações. O estudo chamado "A *Business Intelligence* System" criado por Luhn (1958), conceitua em seu resumo:

"Um sistema automatizado está sendo desenvolvido para disseminar informações para as várias seções de qualquer organização industrial, científica ou governamental. Este sistema de inteligência irá utilizar processamento de dados, máquinas para auto-abstração e auto-codificação

de documentos e para a criação de perfis de interesse para cada um dos “pontos de ação” em uma organização. Os documentos recebidos e gerados internamente são abstraídos automaticamente, caracterizados por padrão de palavras e enviados automaticamente para os pontos de ação apropriados. Esse papel mostra a flexibilidade de tal sistema em identificar informações conhecidas, em encontrar quem precisa saber e em disseminá-lo de forma eficiente em forma abstrata ou como um documento completo”.

Segundo Santos (2022), este artigo trouxe notoriedade para o termo BI e, a partir disto, outros eventos como a evolução da tecnologia e o surgimento de conceitos como Big Data transformaram a forma como entendemos o termo de *Business Intelligence*.

Para compreender os motivos do uso da metodologia BI descrita em por Braghittoni (2017), o autor divide os sistemas de banco de dados em sistemas transacionais e sistemas analíticos. Os sistemas transacionais são usados para as operações cotidianas da empresa, como compras, vendas e pedidos. Já os sistemas analíticos têm como objetivo importar dados dos sistemas transacionais e disponibilizá-los de forma a ser possível sua análise pelos usuários.

Considerando os conceitos explicados, Braghittoni (2017) destaca quatro tópicos que reforçam a importância de manter um banco de dados específico para análise:

- Desempenho - mantendo os dados destinados a análise em um banco separado, é possível consultá-los sem prejudicar o processamento diário dos sistemas transacionais, pois consultas de grande escala, como semestrais ou anuais, exigem alta capacidade de processamento;
- Validação - quando os dados são copiados para uma base unificada, os mesmos passam pelo processo de ETL mencionado na Seção 2.3. Dessa forma, caso haja algum erro, o dado não é computado, e então análises de erro são executadas periodicamente;
- Persistência - diferentemente de sistemas transacionais, que passam por rotinas de exclusão de dados para aumento da performance, os dados de um sistema analíticos são permanentes em seu armazenamento;
- Unificação - Armazenando informações provenientes de diversos sistemas, a base de dados analítica facilita a análise destes dados e mantém uma base centralizada.

2.5 Lei Geral de Proteção de Dados Pessoais (LGPD)

A Lei Geral de Proteção de Dados Pessoais (LGPD) dispõe o uso e tratamento dos dados pessoais nos meios digitais (Brasil, 2018). De acordo com o Ministério da Saúde

(2021a), o dado pessoal é aquele que possibilita a identificação, direta ou indireta, da pessoa natural, tais como nome, data de nascimento, RG, CPF, número de telefone, e-mail, etc. Além das características mencionadas, trata também acerca de dados sensíveis, que são os que revelam origem racial ou étnica, convicções religiosas ou filosóficas, opiniões políticas, filiação sindical, questões genéticas, biométricas e sobre a saúde ou a vida sexual de uma pessoa (Ministério da Saúde, 2021a).

A LGPD também trata dos dados anonimizados, uma técnica de processamento que remove ou modifica informações de modo a impossibilitar a identificação das pessoas envolvidas. Nesses casos, a LGPD não se aplica aos dados em questão (Ministério da Saúde, 2021a).

No que diz respeito aos dados públicos, que são abordados neste estudo, o tratamento de dados pessoais de caráter público deve considerar a finalidade, a boa-fé e o interesse público que justificaram sua divulgação. De acordo com a LGPD, uma organização pode tratar dados tornados públicos pelo titular anteriormente e de forma evidente, sem a necessidade de obter um novo consentimento (Ministério da Saúde, 2021a).

Diante do exposto e dos objetivos e métodos propostos neste estudo, conclui-se que não é necessário aplicar a LGPD no projeto. Isso se deve ao fato de que o projeto em questão utiliza e transforma dados públicos com o objetivo de análise para aprimoramento dos processos da RAPS, justificando, assim, o interesse público na divulgação desses dados por meio de painéis. Além disso, o tratamento dos dados é realizado de forma pública e transparente.

3 Metodologia Utilizada

Neste capítulo serão abordados os procedimentos metodológicos utilizados para a realização da pesquisa e o detalhamento de cada fase, junto com suas atividades, objetivos e resultados, como descrito brevemente na Seção 1.5.

A pesquisa possui natureza aplicada, tendo em vista que busca fornecer soluções concretas para problemas sociais. A utilização de tecnologias como a programação e manipulação de bancos de dados no desenvolvimento de painéis torna a pesquisa quantitativa. Além disso, o estudo possui cunho exploratório por apresentar a ferramenta junto aos detalhes operacionais (RAMOS, 2022).

No contexto deste trabalho, o estudo se divide em dois componentes. O Estudo 1 concentrou-se na concepção e desenvolvimento dos painéis de monitoramento para a RAPS, e envolve a análise de requisitos e indicadores de saúde, a arquitetura do sistema e a implementação dos painéis de monitoramento. É importante ressaltar que o Estudo 1 está vinculado a um projeto do meu estágio, que se encerrou durante o processo de desenvolvimento do TCC, com os resultados esperados obtidos. O Estudo 2 tem como objetivo realizar uma análise postmortem dos resultados alcançados no Estudo 1.

A abordagem de desenvolvimento empregada neste projeto é o Rapid Application Development (RAD), escolhida por ser utilizada no projeto Coopere e ter sua eficácia comprovada. De acordo com Martin, citado por Daud (MARTIN, 1991 apud DAUD; BAKAR; RUSLI, 2010), o RAD é um ciclo de vida de desenvolvimento projetado para proporcionar resultados mais rápido e de maior qualidade do que aqueles alcançados pelo ciclo de vida tradicional. Além disso, Daud também cita três vantagens ao usar o RAD:

- Facilidade de implementação: a implementação se torna mais clara e precisa, pois o desenvolvimento é focado em cada requisito individualmente;
- Melhoria na satisfação do usuário: no decorrer do desenvolvimento do produto, o envolvimento do usuário contribui para aprimorar a satisfação do mesmo, já que a comunicação se intensifica e o usuário pode acompanhar o progresso do produto;
- Redução do tempo de lançamento do produto: há uma diminuição significativa no tempo necessário para implementação em um ambiente de trabalho funcional.

A Tabela 1 apresenta uma estruturação detalhada do projeto, dividido em fases, que apontam os objetivos, as atividades necessárias para atingi-los e os resultados de cada fase.

Fase	Objetivo	Atividades	Resultados
Fase 1	OE1 - Identificar os requisitos do dashboard	A1 - Selecionar a plataforma de BI	R1 - Documento de comparação
		A2 - Realizar reuniões com integrantes do projeto	R2 - Documento de Indicadores
		A3 - Documentar os indicadores	
		A4 - Validar os indicadores	
		A5 - Priorizar os indicadores	
Fase 2	OE2 - Centralizar as informações dos SIS	A6 - Definir BPMN	R3 - Documento de Arquitetura do DW
		A7 - Definir modelo conceitual de ETL	
		A8 - Implementar serviços de ETL	
		A9 - Definir arquitetura do DW	
Fase 3	OE3 - Prover painéis de monitoramento	A10 - Desenvolver protótipo	R5 - Protótipo de alta fidelidade
		A11 - Criar conexão com o DW	R6 - Dashboard
		A12 - Desenvolver o dashboard	
Fase 4	OE4 - Analisar os estudos e painéis desenvolvidos	A13 - Descrever a metodologia da análise	R7 - Análise Postmortem
		A14 - Pontuar os objetos de análise	
		A15 - Desenvolver a análise postmortem	

Tabela 1 – Metodologia em fases - Fonte: Autor.

Para melhor entendimento acerca da metodologia, a Figura 5 representa uma visualização do funcionamento de cada fase, seus objetivos e suas respectivas entradas e saídas. Além disso, há uma fronteira do escopo do TCC 1 e do escopo do TCC 2, que mostra que a Fase 1 foi planejada e desenvolvida no TCC 1, enquanto as Fases 2 e 3 foram planejadas para desenvolvimento no TCC 2.

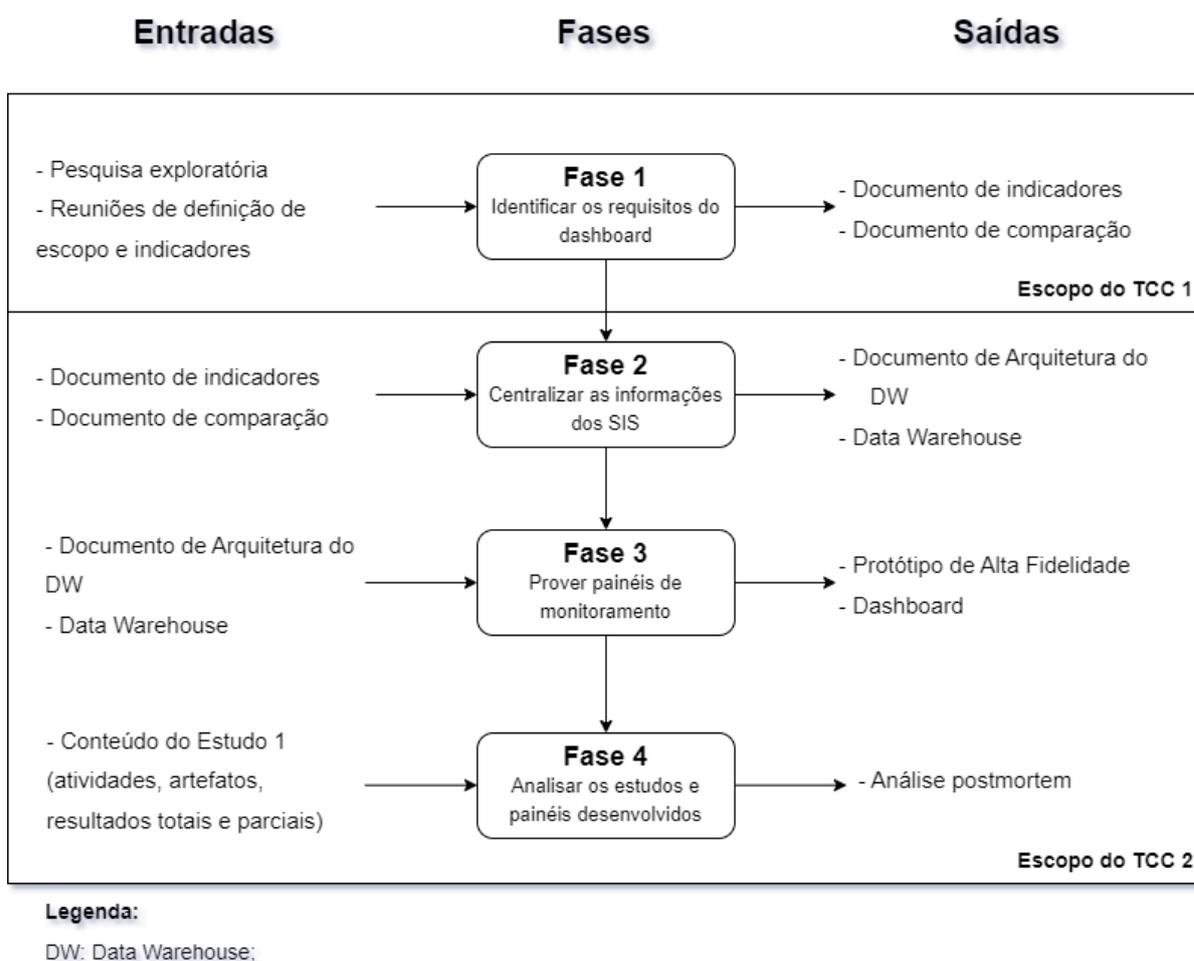


Figura 5 – Metodologia em fases - Fonte: Autor.

3.1 Fase 1

A fase 1 tem como objetivo Identificar os indicadores de saúde (OE1) e produzir o Documento de Comparação (R1) e o Documento de Indicadores (R2) através da realização de atividades que consistem em processos similares ao levantamento de requisitos (processo que capta as necessidades que o sistema atenderá), com o objetivo de coletar informações relevantes a serem incluídas nos painéis de monitoramento e estabelecer o escopo do dashboard.

Os métodos e processos utilizados para cada atividade estão descritos abaixo:

- Selecionar a plataforma de *Business Intelligence* (A1)

A finalidade de Selecionar a plataforma de BI é realizar uma análise das principais plataformas de BI indicadas pela Gestora da equipe de Tecnologia da Informação

(TI), Joyce Miranda. As plataformas de BI selecionadas para análise são o Power BI e o Apache Superset, pois são consideradas as maiores plataformas de visualização de dados disponíveis no mercado.

A atividade A1 gera o Documento de comparação de plataformas de BI (R1), presente na Subseção 4.1.1 de resultado, que contém uma análise comparativa sobre os recursos de visualização, facilidade de uso, custos e escalabilidade das plataformas, a fim de selecionar a plataforma de BI que mais atende o contexto do projeto.

- Realizar reuniões com integrantes do projeto (A2)

A finalidade das reuniões na fase 1 é identificar os indicadores de saúde relevantes a partir dos dados dos Sistemas de Informação em Saúde, por meio de *brainstorming* e *benchmark*.

A tradução literal de *brainstorming* é tempestade de ideias e, segundo [Woebcken \(2019\)](#), a técnica consiste em buscar encontrar soluções para um problema ou ideias para um projeto, por meio do compartilhamento de ideias abertas.

A técnica do *benchmark*, de acordo com [Castro \(2020\)](#), é "um processo de pesquisa entre empresas do mesmo setor para analisar como seus produtos, processos e serviços estão desempenhando em relação aos concorrentes".

Com base nestes conceitos, as reuniões são conduzidas seguindo um roteiro de compartilhamento de ideias, as quais são documentadas no documento de indicadores. Em seguida, é realizada uma análise dos painéis de monitoramento existentes. Posteriormente, os indicadores de saúde que estarão disponíveis no dashboard são discutidos, selecionados e documentados.

As reuniões serão feitas por integrantes¹ da meta 2 do projeto "Business Intelligence aplicada a saúde: painéis de monitoramento para profissionais de saúde da Rede de Atenção Psicossocial", mencionada na Seção 1.2. Cada reunião será documentada na Seção 4, que aborda os Resultados.

A quantidade de reuniões não pode ser definida previamente, já que o resultado só é alcançado quando os indicadores são validados pela atividade A4.

A atividade A2 gera uma parcial do resultado R2, presente na Subseção 4.1.2 de resultados.

- Documentar os indicadores (A3)

A finalidade da documentação dos indicadores é manter em um documento todas

¹ A equipe que integra o projeto é composta por graduandos dos cursos de Saúde Coletiva e Engenharia de Software, pesquisadores, epidemiologistas e gestores de projeto. Os membros são: Bruna Paes, Christiane Kanzler, Eduarda Grillo, Luis Bruno e Mirella Jacinto ([Comunidade Coopere, 2023](#)). A equipe completa do projeto pode ser encontrada na plataforma coopere.

as informações de cada indicador de saúde, como nome do indicador, SIS fonte, interpretação do indicador, variáveis dos SIS utilizadas no indicador, periodicidade de atualização dos dados e método de cálculo do indicador.

A Tabela 2 mostra o modelo a ser seguido por uma tabela de indicador, onde a linha "Nome" deve conter o nome dado ao indicador, ou seja, a informação a ser transmitida. A linha "Tipo" descreve se o indicador está relacionado a estrutura ou a pessoas. A linha "Interpretação" fornece uma explicação sobre o que pode ser analisado a partir da informação fornecida pelo indicador. A linha "Fonte" identifica o sistema de informação (SIS) onde os dados do indicador foram obtidos. A linha "Variáveis" apresenta as variáveis e colunas do SIS que estão presentes no indicador. A linha "Periodicidade" especifica a frequência com que o indicador será atualizado. A linha "Método de cálculo" representa o cálculo utilizado para mostrar a informação. A linha "Observação" fornece informações adicionais e relevantes sobre o indicador (como por exemplo: o resultado do cálculo deve ser sempre positivo), portanto, não é obrigatória.

IS01 - Indicador de saúde 1	
Tipo	
Interpretação	
Fonte	
Variáveis	
Periodicidade	
Método de cálculo	
Observação	

Tabela 2 – Modelo da tabela que descreve um indicador - Fonte: Autor.

Essa atividade gera uma parcial do resultado R2, mas não o finaliza pois este documento deve ser validado e priorizado (atividades A4 e A5). O resultado desta atividade se encontra na Subseção 4.1.3 de resultado.

- Validar os indicadores (A4)

Os indicadores devem ser validados pelo Doutor Jonas Brant² e pela Christiane Kanzler³. Nesta atividade, os indicadores documentados serão enviados para o Prof. Jonas Brant e a Cristiane Kanzler, que irão avaliar se os indicadores estão de acordo com os aspectos adequados para os gestores de saúde da RAPS.

Caso os indicadores não sejam aprovados, a atividade A2 é retomada e os indicadores são reformulados de acordo com o *feedback* dos avaliadores.

² Epidemiologista, Professor do curso de Saúde Coletiva na Universidade de Brasília e um dos coordenadores do projeto Coopere.

³ pesquisadora da meta 1 do projeto (mencionado na Seção 1.2) e conexão entre o projeto e o Ministério da Saúde.

- Priorizar os indicadores (A5)

Para finalizar o resultado R1, os indicadores devem ser priorizados de forma a estabelecer uma ordem de desenvolvimento com base na relevância atribuída pelos gestores de saúde. Os indicadores foram avaliados em cinco critérios:

- Valor de negócio: valor de 1 a 3. 1 indica "baixo", é um indicador complementar. 2 indica "médio", é um indicador relevante, mas não é core. 3 indica "alto", um indicador muito utilizado e de alta relevância.
- Acesso: valor de 1 a 3. 1 indica "sem livre acesso", um indicador que não é possível obter os dados. 2 indica "com restrições", indicador que é possível obter os dados, mas com solicitações e burocracias. 3 indica "livre", um indicador em que todos os dados são de livre acesso.
- Estrutura: valor de 1 a 2. 1 indica "dados não normalizados" e 2 indica "dados normalizados".
- Formato: valor de 1 a 3. 1 indica "texto não formatado". 2 indica "texto formatado". 3 indica "tabela".
- Atualização e Manutenção: valor entre 1 e 3. 1 indica "difícil atualização/manutenção", 2 indica "média atualização/manutenção" e 3 indica "manutenção e atualização fácil"

Com base nas pontuações atribuídas a cada critério, é calculado um valor "Total" por meio da multiplicação dos valores correspondentes de cada coluna. Os indicadores que apresentarem os maiores valores na coluna "Total" devem receber prioridade.

A atividade A5 finaliza o resultado R2 e, por consequência, atinge o objetivo específico de identificar os indicadores de saúde (OE1).

3.2 Fase 2

A fase 2 tem como objetivo Centralizar as informações dos Sistemas de Informação em Saúde (OE2) e produzir o Documento de arquitetura do DW (R3) e o próprio Data Warehouse (R4). O resultado R1 é necessário para auxiliar as tecnologias que serão utilizadas no desenvolvimento da arquitetura, enquanto o resultado R2 auxilia o planejamento da arquitetura do DW pela modelação de processos ETL com Business Process Modeling and Notation (BPMN), baseando-se nas bases de dados dos SIS fontes. Dessa forma o modelo conceitual de ETL é definido e implementado e, posteriormente, o documento de arquitetura e o DW são desenvolvidos.

Os métodos e processos utilizados para cada atividade estão descritos abaixo:

- Definir Modelagem BPMN (A6)

A finalidade da modelagem BPMN é simplificar o desenvolvimento do ETL pois, de acordo com a [OMG \(2010\)](#), o BPMN é um padrão para diagramas de processos, inclui uma notação fácil e intuitiva de uso, semelhante a fluxogramas e independente de qualquer ambiente específico de implementação, que permite a visualização dos processos de negócios que serão suportados pelo DW.

Segundo [Monteiro \(c2023\)](#), o BPMN é formado por *Swimlanes* (mecanismo de organização), Atividades (o trabalho realizado por um ator), Sub-processos, *Gateways* (decisões que controlam a convergência/divergência do fluxo), Eventos (algo que acontece e afeta o fluxo do processo), Objetos de conexão, Artefatos (entradas e saídas), como mostra o exemplo de um fluxo básico na Figura 6.

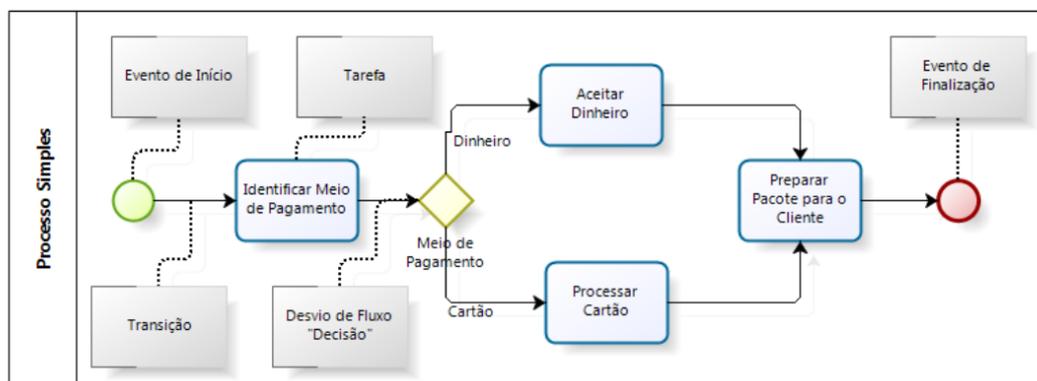


Figura 6 – Fluxo básico de um BPMN ([MONTEIRO, c2023, p. 22](#))

A modelagem BPMN utilizará como entrada o Documento de Indicadores (R2).

A atividade A6 gera uma parcial do resultado R3.

- Definir modelo conceitual de ETL (A7)

A finalidade da definição do modelo conceitual de ETL é produzir uma visão geral do processo. Segundo [Vaisman e Zimányi \(2022, p. 306\)](#), o modelo conceitual aborda as tarefas de controle necessárias para realizar a transformação de dados a partir do banco operacional e o carregamento dos dados transformados para o DW. Para isto, também deve-se desenvolver o esquema lógico dos dados para descrever os arquivos que realizam o carregamento.

A atividade A7 gera uma parcial do resultado R3.

- Implementar serviços de ETL (A8)

A finalidade da implementação dos serviços de ETL é garantir a qualidade dos dados armazenados e a escalabilidade do projeto. Para isto, é importante garantir que as tarefas de ETL possam ser executadas de forma eficiente e automatizada para

garantir que os dados fiquem sempre atualizados de acordo com a periodicidade definida no Documento de Indicadores (R2).

Para implementar os serviços de ETL será utilizado o SQL Server Integration Services (SSIS), plataforma para a criação de soluções de integração de dados da Microsoft, que já inclui pacotes de ETL. A escolha dessa ferramenta se dá pelo resultado entre a comparação de plataformas de BI, o resultado do Documento de comparação (R1).

A atividade A8 gera uma parcial do resultado R3.

- Definir arquitetura do Data Warehouse (A9)

A definição da arquitetura do DW é um dos processos para a base do projeto de DW e também para ser uma fonte de informação sobre o repositório para outros integrantes do projeto e futuros colaboradores.

Baseando-se no modelo de documento de arquitetura de [Ponniah \(2002\)](#), presente em seu livro "Data Warehousing Fundamentals: A comprehensive Guide for IT Professionals" e com breves adaptações, o documento de arquitetura do DW deve conter os seguintes tópicos:

1. Contexto e Requisitos - tópico que aborda o contexto do negócio e os requisitos, logo, este tópico depende do Documento de indicadores (R1);
2. Conceitos arquitetônicos - tópico que aborda a descrição da estrutura geral do DW, que também deve incluir os tipos de dados, modelos e estruturas de armazenamento;
3. Solução técnica - tópico que aborda a descrição dos componentes utilizadas para construção do DW, tais como bancos de dados fontes e processos de ETL. Além disso, deve conter detalhes acerca da hospedagem dos dados tais como a máquina, sistema operacional e recursos de hardware da mesma;
4. Processos operacionais - tópico que aborda os procedimentos de suporte ao DW, tais como monitoramento, periodicidade de atualização dos dados e manutenção;
5. Integração: tópico que aborda todas as integrações do DW.

A atividade A9 finaliza o resultado R3.

3.3 Fase 3

A fase 3 tem como objetivo Prover painéis de monitoramento (OE3) e produzir o Protótipo de alta fidelidade (R4) e o Dashboard (R5). Para isto, a fase 3 utiliza como entrada o Documento de arquitetura do DW (R3) e o próprio Data Warehouse (R4),

necessários para preencher a plataforma de BI com os dados dos SIS e criar os gráficos e tabelas, baseando-se no design do Protótipo de alta fidelidade (R4).

Os métodos e processos utilizados para cada atividade estão descritos abaixo:

- Desenvolver protótipo (A10)

Segundo a [Tera \(2020\)](#):

"Um protótipo de alta fidelidade (às vezes chamado de high-fi ou hi-fi) é uma representação interativa do produto, baseada no computador ou em dispositivos móveis. Esse protótipo já apresenta maior semelhança com o design final em termos de detalhes e funcionalidade".

Portanto, o objetivo do desenvolvimento do protótipo de alta fidelidade (A10) dos painéis de monitoramento é desenvolver um design do produto e, a partir disto, validar com os gestores do projeto se a proposta será levada adiante.

O protótipo será desenvolvido utilizando o próprio Power BI e será validado pelos coordenadores da meta 2 do projeto (Joyce Miranda e Jonas Brant) e pela pesquisadora Christiane Kanzler, que irão avaliar se o protótipo está adequado, de acordo com sua proposta. Caso o protótipo seja aprovado, a atividade seguinte pode ser iniciada, caso contrário, os ajustes solicitados devem ser trabalhados no protótipo de alta fidelidade e então uma nova validação é feita até que o protótipo atenda às expectativas do cliente.

A atividade A10 gera o Protótipo de alta fidelidade (R4).

- Criar conexão com o DW (A11)

A atividade Criar conexão com o DW consiste em vincular a plataforma de BI selecionada no Documento de comparação (R1) com o Data Warehouse, para que a mesma tenha acesso aos dados que serão utilizados para gerar os gráficos.

Tendo como base as funcionalidades do Power BI, é possível realizar essa conexão por meio da funcionalidade "Obter Dados" da ferramenta, que permite selecionar o banco de dados e fornece a opção de "Conectar", que requer a adição de informações como autenticação (se necessário) e adição de tabelas que devem ser importadas para o Power BI. A conexão com o DW e suas respectivas tabelas será documentada.

A atividade A10 gera uma parcial do Dashboard (R5).

- Desenvolver o dashboard na plataforma selecionada (A12)

A atividade de desenvolver o dashboard na plataforma de BI tem a finalidade de prover os painéis de monitoramento interativos, a partir da criação de gráficos, tabelas e outros tipos de visualizações que permitam a análise de dados da RAPS.

Além disso, por utilizar o Power BI, os gráficos permitem o uso de filtros e interações que permitem o usuário ter uma interface mais amigável com estes dados.

Por ser a última atividade do projeto, a mesma depende de todos resultados alcançados até o momento (R1 a R4).

Além disso, uma sub-atividade da atividade A11 é disponibilizar os gráficos gerados na *web*, de forma que o dashboard se torne disponível para acesso em qualquer navegador a partir de uma Uniform Resource Locator (URL, um endereço para acessar determinado serviço na internet). O método de disponibilização dos painéis de monitoramento na internet realizado pelo Power BI é realizado pela funcionalidade de Publicação, que gera uma URL e permite definir permissões de acessos e configurações de segurança.

A atividade A12 finaliza o resultado R5 e, por consequência, atinge o objetivo específico de Prover painéis de monitoramento (OE3). Com isso, o objetivo geral de Prover informações da Rede de Atenção Psicossocial, de forma automática e centralizada, as quais auxiliem os gestores das RAPS na tomada de decisões é alcançado.

3.4 Fase 4

A fase 4 tem como objetivo Analisar os estudos e painéis desenvolvidos (OE4) e produzir a Análise postmortem (R6). Para isto, a fase 4 utiliza como entrada todo o conteúdo do Estudo 1 (fases 1, 2 e 3), atividades, artefatos e resultados. Os métodos e processos utilizados para cada atividade estão descritos abaixo:

- Descrever a metodologia da análise (A13)
Nesta etapa será fornecida uma visão detalhada da abordagem utilizada para avaliar os estudos, baseando-se nos princípios da Atlassian, uma empresa de desenvolvimento de software australiana. A Atlassian possui uma metodologia sistemática e rigorosa, mas flexível, podendo ser adaptada conforme necessário para atender às necessidades específicas do projeto. A Atividade A13 gera uma parcial do Resultado R6.
- Pontuar os objetos de análise (A14)
Nesta etapa o objetivo é selecionar os objetos de análise dentro do conteúdo do Estudo 1, explanando os objetivos da análise em cada contexto. A Atividade A14 gera uma parcial do Resultado R6.
- Desenvolver a análise postmortem (A15)
Nesta etapa será fornecida as análises sobre cada objeto de acordo com a metodologia definida. A Atividade A15 finaliza o Resultado R6, concluindo o objetivo específico OE4 e, por consequência, finalizando o Estudo 2.

4 Resultados

Este capítulo aborda todos os resultados alcançados, divididos por fases e atividades, como mencionado no capítulo 3 (Metodologia).

4.1 Fase 1

A fim de alcançar o objetivo de Identificar os requisitos do dashboard (OE1), as subseções seguintes mostram os resultados das atividades da Fase 1. Como demonstrado na Tabela 1 e na Figura 5, a Fase 1 deve cumprir os seguintes resultados:

- Documento de comparação
O resultado R1 consiste em um documento de análise e comparação entre plataformas Apache Superset e Microsoft Power BI. A análise se concentrará nos recursos de visualização, na facilidade de uso, escalabilidade e custos. Com base nessa análise, o objetivo é fornecer uma compreensão clara das vantagens e desvantagens de cada ferramenta, para auxiliar na escolha da melhor opção para as necessidades específicas do usuário, levando em conta os cenários em que cada plataforma se sairá melhor.
- Documento de Indicadores
O resultado R2 consiste em um documento com os indicadores que estarão disponíveis para análise nos painéis de monitoramento. Este documento deve seguir o padrão de ter uma tabela para cada indicador e deve seguir o modelo de indicador mencionado na Tabela 2, que descreve completamente todas as informações relacionadas ao indicador em questão.

4.1.1 Atividade 1 - Selecionar a plataforma de BI

Esta atividade está atrelada a Fase 1, cuja meta é alcançar o Objetivo Específico 1 (OE1): Identificar os requisitos do dashboard.

O resultado da Atividade 1 é o documento de análise e comparação entre as plataformas Apache Superset e Microsoft Power BI. A análise se concentrará nos recursos de visualização, na facilidade de uso, escalabilidade e custos. Com base nessa análise, o objetivo é fornecer uma compreensão clara das vantagens e desvantagens de cada ferramenta, a fim de selecionar a plataforma de BI que mais atende o contexto do projeto.

Iniciando a comparação pela definição de cada ferramenta, o Apache Superset é um plataforma de visualização de dados de código aberto que permite criar, explorar e compartilhar visualizações interativas e dashboards. Ele é construído com base no Apache

Druid e é compatível com várias fontes de dados, incluindo bancos de dados relacionais e arquivos Comma Separated Value (CSV), enquanto o Power BI é uma plataforma de visualização de dados da Microsoft que permite criar visualizações e dashboards interativos a partir de uma variedade de fontes de dados, incluindo bancos de dados relacionais, arquivos de planilhas e dados na nuvem. O Power BI também oferece recursos avançados de análise e inteligência artificial.

A primeira comparação é pelos tipos de gráficos que é possível gerar em cada plataforma, a tabela 3 analisa se cada plataforma de BI possui os tipos de gráficos mais relevantes e utilizados no contexto da saúde.

Tipo de gráfico	Apache Superset	Microsoft Power BI
Gráfico de linha	Sim	Sim
Gráfico de pizza	Sim	Sim
Gráfico de área	Sim	Sim
Gráfico de bolha	Sim	Sim
Gráfico de radar	Sim	Sim (mas precisa ser incluído como extra)
Gráfico de coluna	Sim	Sim
Gráfico de mapa	Sim	Sim
Mapas	Sim	Sim

Tabela 3 – Comparação dos Tipos de Gráficos - Fonte: Equipe Coopere.

A tabela 4 compara os tipos de gráficos, a personalização das visualizações, a interatividade, além disso, também é comparado os recursos avançados, como a capacidade de compartilhar dashboards e a segurança dos dados. A comparação é realizada por "Sim", quando a plataforma possui o recurso. Também é analisado por "Bom" quando a plataforma possui o recurso e "Ótimo" quando a plataforma possui o recurso de maneira mais que necessária no contexto do projeto.

Item	Apache Superset	Microsoft Power BI
Tipos de Gráficos	Ótimo	Ótimo
Personalização das visualizações, a interatividade e a capacidade de animações e transições	Bom	Ótimo
Compartilhamento de Dashboards	Bom	Ótimo
Segurança de dados	Bom	Ótimo
Gráfico de coluna	Sim	Sim
Gráfico de mapa	Sim	Sim
Mapas	Sim	Sim

Tabela 4 – Recursos de visualização - Fonte: Equipe Coopere.

Os custos são um fator importante a ser considerado ao escolher uma ferramenta de construção de dashboards, pois eles afetam diretamente o orçamento da organização. Ambos, Apache Superset e Microsoft Power BI oferecem boas opções de custos, mas com diferentes abordagens. A tabela 5 apresenta uma comparação entre os custos de Apache Superset e Microsoft Power BI em cinco itens.

Item	Apache Superset	Microsoft Power BI
Preço	Gratuito, porém, considerar custos de infraestrutura própria	Gratuito (com recursos limitados) / Pagamento por usuário
Modelo de Licenciamento	Código aberto	Comercial
Suporte	Comunitário	Dedicado
Atualizações	Gratuitas	Pagas
Integração	Flexível	Flexível
Gráfico de coluna	Sim	Sim
Gráfico de mapa	Sim	Sim
Mapas	Sim	Sim

Tabela 5 – Custos entre as plataformas - Fonte: Equipe Coopere.

O Apache Superset é uma ferramenta de código aberto, o que significa que é gratuita para usar e pode ser modificada de acordo com as necessidades da organização. Ele não tem custos de licenciamento e as atualizações são gratuitas. O suporte é fornecido pela comunidade de usuários. Mas isso não significa que sua implantação e manutenção sejam completamente livres de custos. Ainda que o software em si seja gratuito, a organização precisará arcar com custos de implantação e manutenção.

Para implantar o Apache Superset, a organização precisará dedicar recursos, incluindo tempo e mão de obra, para configurar e personalizar o software de acordo com suas necessidades específicas. Isso pode incluir configurações de infraestrutura, integração com outros sistemas e aplicativos, e desenvolvimento de relatórios e visualizações personalizadas. Além disso, é necessário ter uma equipe de TI capacitada para administrar e manter o software.

Além disso, é importante levar em conta que ao utilizar um software open source, a organização precisa garantir que possui licenças e suporte para todas as bibliotecas e tecnologias utilizadas, e deve arcar com os custos de licenças, suporte e manutenção dessas bibliotecas e tecnologias, além de possíveis custos de desenvolvimento de funcionalidades não contempladas no software de base. Já o Power BI é uma ferramenta comercial, oferece uma versão gratuita com recursos limitados e outras opções pagas, incluindo licenciamento por usuário e planos de assinatura. Ele também tem custos de suporte dedicado e atualizações pagas.

Power BI é uma ferramenta voltada para usuários comerciais e oferecida como um

serviço na nuvem. Ele foi projetado para integrar-se com outras ferramentas e tecnologias da Microsoft, como o Office e o Azure, tornando-o uma boa opção para organizações que já utilizam essas ferramentas. Oferece ainda uma série de recursos avançados de integração, incluindo a possibilidade de importar dados usando scripts Python, e conectando-se com uma variedade de fontes de dados, como bancos de dados, arquivos e web services. Isso permite que os usuários acessem e integrem dados de diferentes fontes em seus dashboards e relatórios. Além disso, o Power BI também oferece integrações nativas com outras ferramentas da Microsoft, como o Excel e o Azure, o que aumenta ainda mais a sua capacidade de integração.

Já o Apache Superset, é uma ferramenta de código aberto, isso significa que ele é mais flexível e pode ser facilmente integrado com outras ferramentas e tecnologias. Ele suporta várias fontes de dados, incluindo bancos de dados relacionais, big data, e arquivos estáticos. Por isso, a integração do Apache Superset é considerada mais flexível e ampla comparado ao Power BI.

Apesar de o Apache Superset ser gratuito, a organização ainda precisará arcar com custos de implantação e manutenção, incluindo recursos humanos, infraestrutura e licenças e suporte para bibliotecas e tecnologias utilizadas. Isso deve ser levado em conta ao avaliar o custo total de propriedade de uma ferramenta de construção de dashboards.

Em geral, ambos, Apache Superset e Microsoft Power BI oferecem boas opções de custos, mas com abordagens diferentes. O Superset é uma ferramenta de código aberto, gratuita e flexível, enquanto o Power BI é uma ferramenta comercial com opções gratuitas limitadas e pagas. Cada ferramenta tem sua própria abordagem e vantagens, depende da necessidade e do perfil do usuário para escolher qual é a melhor opção.

Após uma análise de Apache Superset e Microsoft Power BI, podemos concluir que quando avaliamos as ferramentas de acordo com os critérios estabelecidos, a Apache Superset tem um desempenho melhor em relação ao custo, enquanto o Power BI tem uma melhor nota quando se trata de facilidade de uso e recursos avançados de visualização. Ambas as ferramentas possuem notas similares em relação à escalabilidade. O diagrama radar da Figura 7 exemplifica os resultados apurados no escopo desta análise.

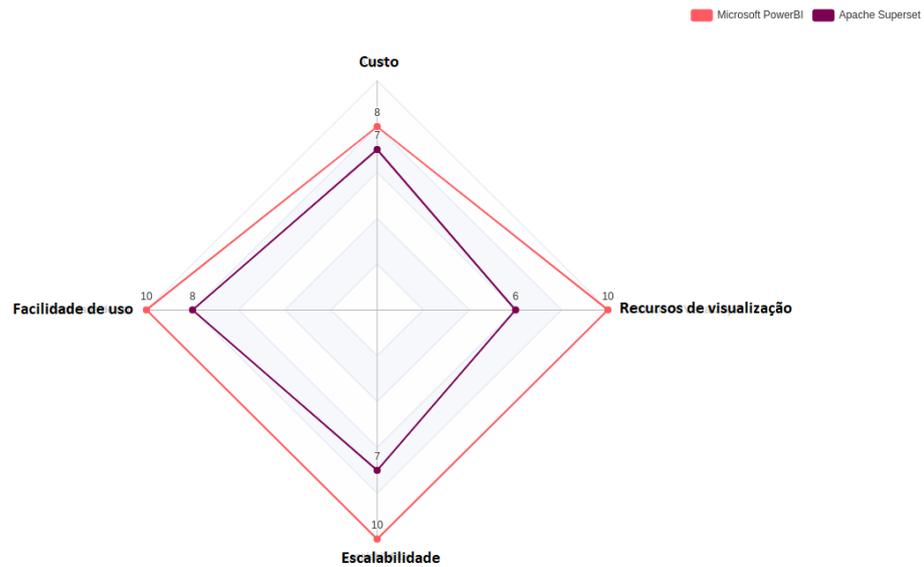


Figura 7 – Gráfico Radar de comparação entre Power BI e Superset - Fonte: Equipe Coopere

A Apache Superset é uma opção mais acessível financeiramente, pois é um software de código aberto gratuito. No entanto, é importante levar em conta os custos de implantação e manutenção. O Power BI, por outro lado, tem uma licença paga, mas não possui custos físicos de implantação, tais como uma infraestrutura de computadores e rede.

Em termos de escalabilidade, ambas as ferramentas oferecem recursos para acompanhar o crescimento dos dados e usuários. No entanto, o Power BI oferece opções adicionais de clusterização para lidar com cargas de trabalho de grande escala. Este quesito está relacionado ainda com uma correta modelagem de dados para permitir uma sumarização e limpeza prévia dos dados que servirão de fonte para os dashboards.

Em conclusão, a plataforma que melhor atende o projeto é o Power BI, já que é uma ferramenta com baixa curva de aprendizagem, baixo custo de manutenção em quesito de recursos computacionais e humanos,

4.1.2 Atividade 2 - Realizar reuniões com integrantes do projeto

esta atividade está atrelada a Fase 1, cuja meta é alcançar o Objetivo Específico 1 (OE1): Identificar os requisitos do dashboard.

Seguindo o roteiro de compartilhamento de ideias, utilizando a ferramenta Google Docs para documentar os indicadores e a ferramenta Jitsi para conferência, até o momento houve uma reunião para planejamento da atividade e uma reunião para a definição de oito indicadores de saúde, como mostra a Figura 8.

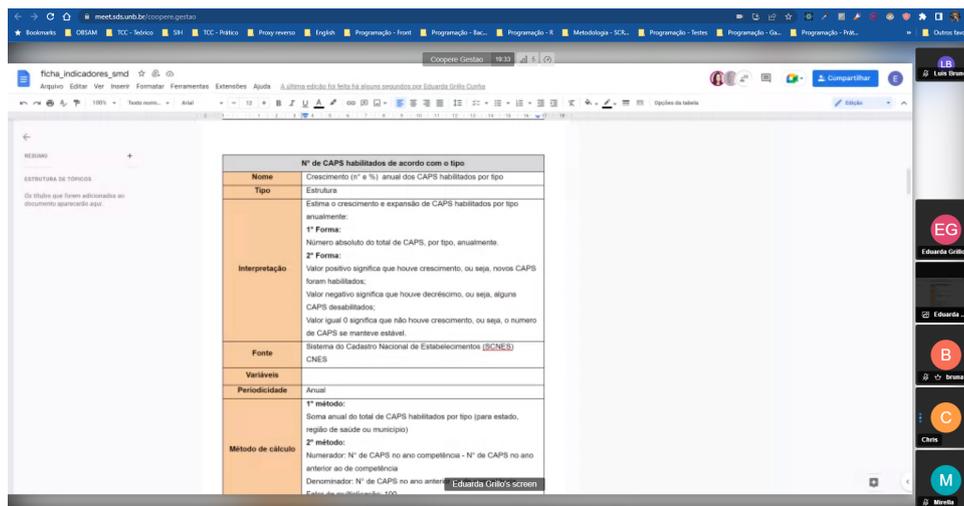


Figura 8 – Reunião para levantamento de indicadores - Fonte: Autor.

4.1.3 Atividade 3 - Documentar os indicadores

esta atividade está atrelada a Fase 1, cuja meta é alcançar o Objetivo Específico 1 (OE1): Identificar os requisitos do dashboard.

Até o presente momento, dez indicadores foram detalhados com base no modelo de indicador demonstrado na Tabela 2. Abaixo estão três dos dez indicadores selecionados para aprofundamento neste presente trabalho (IS01, IS05 e IS06), enquanto os outros sete indicadores encontram-se nos Anexos A ao G.

Os indicadores de saúde abordados se tratam de:

- IS01 Figura 9 - Expansão de CAPS por ano (por Estado, Região de Saúde e Município. SIS fonte dos dados: CNES;
- IS05 Figura 10 - Distribuição (n° e %) de hospitais de referência com leitos de saúde mental habilitados por UF/Município e ano. SIS fonte dos dados: SCNES;
- IS06 Figura 11 - Distribuição (n° e %) de internações por causas de saúde mental por mês/ano. SIS fonte dos dados: SIH.

IS01 - Expansão de CAPS por ano (por Estado, Região de Saúde e Município)	
Tipo	Estrutura
Interpretação	Estima o crescimento e expansão de novos CAPS por ano: 1º Forma: Número absoluto do total de CAPS por ano. Análise comparativa, exemplo: série histórica. 2º Forma: Valor positivo significa que houve crescimento, ou seja, novos CAPS foram cadastrados; Valor negativo significa que houve decréscimo, ou seja, alguns CAPS desabilitados; Valor igual 0 significa que não houve crescimento, ou seja, o número de CAPS se manteve estável.
Fonte	Sistema do Cadastro Nacional de Estabelecimentos (SCNES) CNES-SR
Variáveis	TP_UNID_ID / CO_TIPO_UNIDADE - Tipo de estabelecimento CÔMPETÊN - Ano e Mês de competência da informação (AAAAMM) COD_MUN / CO_MUNICIPIO - Código do Município
Periodicidade	Anual
Método de cálculo	1º método: Soma anual do total de CAPS (para estado, região de saúde ou município). 2º método: Numerador: N° de CAPS no ano competência - N° de CAPS no ano anterior ao de competência. Denominador: N° de CAPS no ano anterior ao de competência. Fator de multiplicação: 100.
Observação	O resultado desejado deve ser sempre positivo, pois representa um crescimento de novos CAPS.

Figura 9 – Indicador de Saúde 1 - Fonte: Equipe Coopere

IS05 - Distribuição (n° e %) de hospitais de referência com leitos de saúde mental habilitados por UF/Município e ano	
Tipo	Estrutura
Interpretação	Analisar a disponibilidade de leitos de saúde mental em hospitais de referência, anualmente, para o município/UF.
Fonte	Sistema do Cadastro Nacional de Estabelecimentos (SCNES)
Variáveis	TP_UNID_ID / CO_TIPO_UNIDADE - Tipo de estabelecimento CÔMPETÊN - Ano e Mês de competência da informação (AAAAMM) COD_MUN / CO_MUNICIPIO - Código do Município QTDE_SUS - Quantidade de leitos disponíveis para o SUS IBGE: Tabela de código UF/municípios
Periodicidade	Anual
Método de cálculo	1º método: Apenas o quantitativo de leitos de saúde mental disponíveis em hospital de referência, de acordo com local e ano. 2º método: Numerador: quantitativo de leitos de saúde mental disponíveis em hospitais de referência, de acordo com local e ano Denominador: total de leitos em hospitais de referência, no mesmo local e ano. Fator de multiplicação: 100
Observação	

Figura 10 – Indicador de Saúde 5 - Fonte: Equipe Coopere

IS06 - Distribuição (nº e %) de internações por causas de saúde mental por mês/ano	
Tipo	Processo
Interpretação	Quantificar e estimar o percentual de usuários internados com CID de saúde mental, no atendimento de urgência e emergência em Hospital Geral e de Referência, comparado com as demais internações.
Fonte	Sistema de Internações Hospitalares do SUS (SIH) SIH-RD
Variáveis	UF_ZI - Município Gestor DI_INTER - Data de internação no formato aaammdd DIAG_PRINC - Código do diagnóstico principal (CID-10) Códigos CID-10: http://www2.datasus.gov.br/cid10/V2008/cid10.htm Categorias CID-10 saúde mental: F00, F01, F02, F03, F04, F05, F06, F07, F09, F10, F11, F12, F13, F14, F15, F16, F17, F18, F19, F20, F21, F22, F23, F24, F25, F28, F29, F30, F31, F32, F33, F34, F38, F39, F40, F41, F42, F43, F44, F45, F48, F50, F51, F52, F53, F54, F55, F59, F60, F61, F62, F63, F64, F65, F66, F68, F69, F70, F71, F72, F73, F78, F79, F80, F81, F84, F88, F89, F90, F91, F92, F93, F94, F95, F98, F99, X60, X61, X62, X63, X64, X65, X66, X67, X68, X69, X70, X71, X72, X73, X74, X75, X76, X77, X78, X79, X80, X81, X82, X83, X84
Periodicidade	Mensal e Anual
Método de cálculo	1º método: Apenas o quantitativo de internações por saúde mental, conforme o CID-10 (agrupado por categoria), de acordo com o mês/ano e local. 2º opção: Numerador: quantitativo de internações, em atendimentos de urgência e emergência, por causas de saúde mental conforme o CID-10 (agrupado por categoria), de acordo período mês/ano e local. Denominador: total de internações, em atendimentos de urgência e emergência, por causas de saúde mental conforme o CID-10 (agrupado por categoria), no mesmo local e período ano/mês. Fator de multiplicação: 100.
Observação	Agrupar em categorias conforme o CID, sendo: transtornos mentais e comportamentais - CID F00-F99 e lesões autoprovocadas - CID X60-X84.

Figura 11 – Indicador de Saúde 6 - Fonte: Equipe Coopere

4.1.4 Atividade 4 - Validar os indicadores

esta atividade está atrelada a Fase 1, cuja meta é alcançar o Objetivo Específico 1 (OE1): Identificar os requisitos do dashboard.

A validação dos indicadores foi conduzida por meio de reuniões realizadas via chamadas virtuais utilizando a plataforma Jitsi da SDS. Pelo curto de prazo da entrega do projeto, optamos por uma abordagem flexível, realizando várias reuniões informais e, durante esses encontros, a tela de um dos participantes era compartilhada, permitindo que todos os envolvidos visualizassem os indicadores, como reportado no Anexo H.

Embora nem todos os membros tenham participado de todas as reuniões, garantimos a presença de pelo menos um dos dois stakeholders principais, Jonas e/ou Christiane, tendo em vista que são eles que detêm um profundo entendimento das necessidades e objetivos dos painéis no auxílio das tomadas de decisão para a saúde.

Durante as reuniões, as validações dos indicadores eram conduzidas de maneira dinâmica e interativa. A pessoa responsável por compartilhar a tela apresentava os indicadores,

e os stakeholders realizavam avaliações instantâneas e comentários, e então as alterações necessárias eram feitas naquele mesmo momento.

4.1.5 Atividade 5 - Priorizar os indicadores

esta atividade está atrelada a Fase 1, cuja meta é alcançar o Objetivo Específico 1 (OE1): Identificar os requisitos do dashboard.

A Tabela 6 mostra a pontuação de cada indicador a fim de priorizar a ordem de utilização da escolha dos indicadores que serão utilizados no escopo deste trabalho, com base na Seção 3.1.

Indicador	Valor de Negócio	Acesso	Estrutura	Formato	Atualização	Total
IS01	2	3	2	3	3	108
IS02	2	3	2	3	3	108
IS03	2	3	2	3	2	72
IS04	1	3	2	3	1	18
IS05	3	3	2	3	3	162
IS06	3	3	2	3	3	162
IS07	2	2	2	3	2	48
IS08	2	3	2	3	2	72
IS09	2	3	2	3	2	72
IS10	2	3	2	3	2	72

Tabela 6 – Priorização dos indicadores - Fonte: Autor

É relevante enfatizar que os indicadores abordados nesta pesquisa são o IS05, IS06 e IS01, uma vez que se destacam devido à sua maior pontuação no processo de priorização.

4.2 Fase 2

A fim de alcançar o objetivo de Centralizar as informações dos SIS (OE2), as subseções seguintes mostram os resultados das atividades da Fase 2. Como demonstrado na Tabela 1 e na Figura 5, a Fase 2 deve cumprir os seguintes resultados:

- Documento de Arquitetura do DW
O resultado R3 consiste em um documento detalhado que descreve a arquitetura do Data Warehouse, onde serão apresentados todos os aspectos técnicos da estrutura de armazenamento de dados, incluindo informações sobre as fontes de dados, integração de dados, esquemas de armazenamento e processos ETL.
- Data Warehouse
O resultado R4 consiste no ambiente de armazenamento centralizado que abriga os dados de SIS distintos, já transformados e preparados para análise.

4.2.1 Atividade 6 - Definir BPMN

esta atividade está atrelada a Fase 2, cuja meta é alcançar o Objetivo Específico 2 (OE2): Centralizar as informações dos SIS.

O resultado da Atividade 6 são os diagramas BPMN, que foram categorizados em processos e transformações. A Figura 12 retrata um BPMN geral do processo de ETL, integrando os elementos fundamentais do sistema, um evento de início indicando o processo de execução dos scripts em R (mencionados na Subseção 4.2.2) periodicamente, as atividades realizadas, os artefatos gerados e o evento de término.

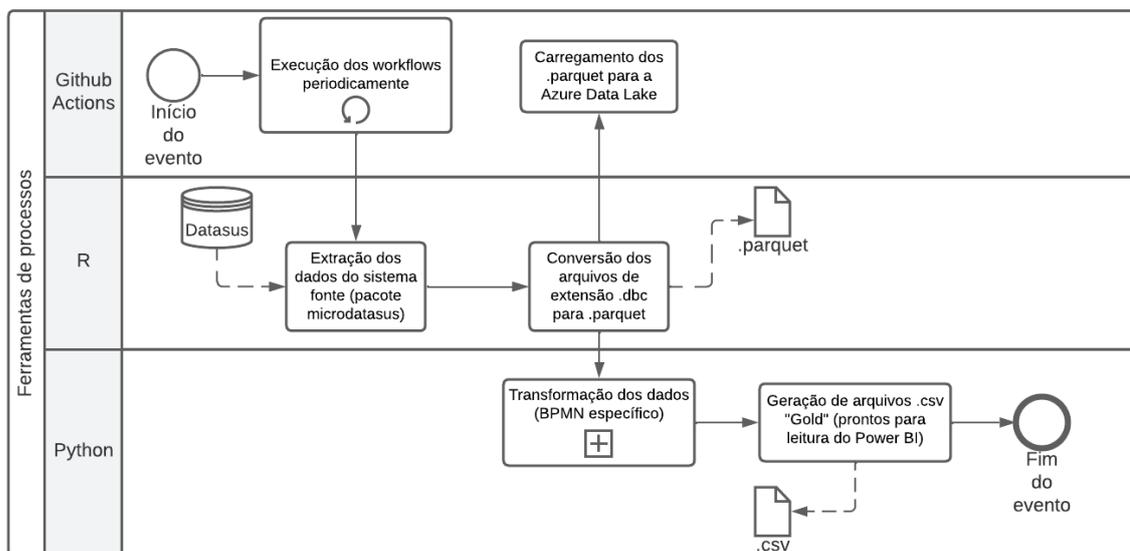


Figura 12 – BPMN da modelagem do processo de ETL - Fonte: Autor.

A Figura 13 exibe um diagrama BPMN que descreve as transformações de dados, uma atividade específica mencionada na modelagem da Figura 12. Este BPMN detalha as transformações de dados que transcorrem em cada Sistema de Informação em Saúde (SIS) e em cada datamart, que é um local de armazenamento de dados que engloba informações específicas da unidade de negócios de uma organização (AWS, 2023), neste caso os SIS.

O processo começa pelo evento de início, que consiste na extração de dados do DATASUS, e prossegue com as operações de transformações de dados em cada datamart, uma vez que cada SIS requer procedimentos de processamento de dados únicos. No datamart do Sistema de Informações Hospitalares (SIH), são efetuadas conversões, filtragens, agrupamento de colunas e criação de novas colunas, culminando na geração de um arquivo no formato CSV. No datamart do Cadastro Nacional de Estabelecimentos de Saúde (CNES), ocorrem ações como filtragens, combinação de colunas, remoção de registros duplicados e geração de arquivos no formato CSV, conforme apresentado na Figura 13.

É imprescindível esclarecer sobre as camadas nas quais os arquivos gerados são categorizados. A camada "bronze" assume a responsabilidade pelo armazenamento dos

dados brutos, geralmente logo após a aquisição ou, ao menos, após a conversão para formatos alternativos. A camada "silver" é designada para acomodar os dados após a aplicação de regras e procedimentos de controle de qualidade nos dados brutos, incluindo as primeiras transformações e agregações. Já camada "gold" se constitui como a camada na qual os dados são preparados para serem consumidos pelas áreas de negócio, especificamente, pelo Power BI.

Uma observação relevante acerca da cronologia dos dados que passam pelo processo de ETL e são exibidos nos painéis é que estes abrangem o período de 2018 até a data mais recente disponível nos últimos registros atualizados na base de dados nacional. Cumpre destacar que a atualização dos indicadores decorre da frequência em que o mesmo é atualizado na base de dados nacional, que pode ser mensal ou anual.

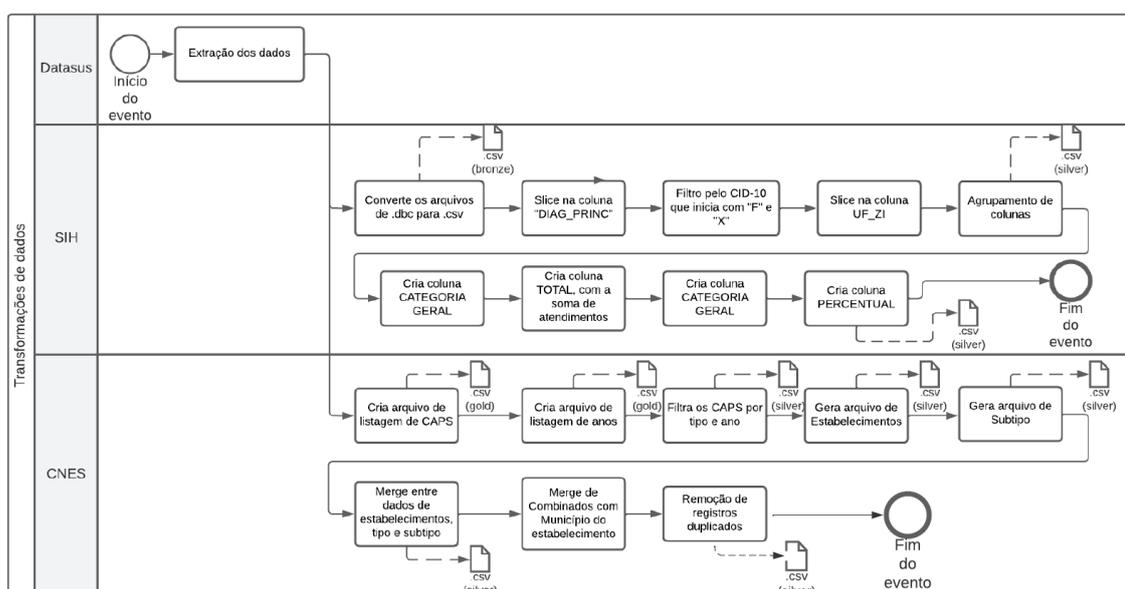


Figura 13 – BPMN das transformações de dados - Fonte: Autor.

4.2.2 Atividade 7 - Definir modelo conceitual de ETL

esta atividade está atrelada a Fase 2, cuja meta é alcançar o Objetivo Específico 2 (OE2): Centralizar as informações dos SIS.

Conforme ilustrado na Figura 14, o processo de funcionamento do sistema tem início com a execução de scripts em R por meio da ferramenta Github Actions, responsável por automatizar e executar fluxos de trabalho no Github. Esses scripts em R são responsáveis por obter os dados do DATASUS relativos ao SIS - Sistema de Informação em Saúde.

Posteriormente, esses mesmos scripts convertem os dados baixados para o formato .parquet, que é uma extensão de armazenamento de dados colunar, e os enviam para o DataLake da Azure, serviço de armazenamento de dados na nuvem da Microsoft.

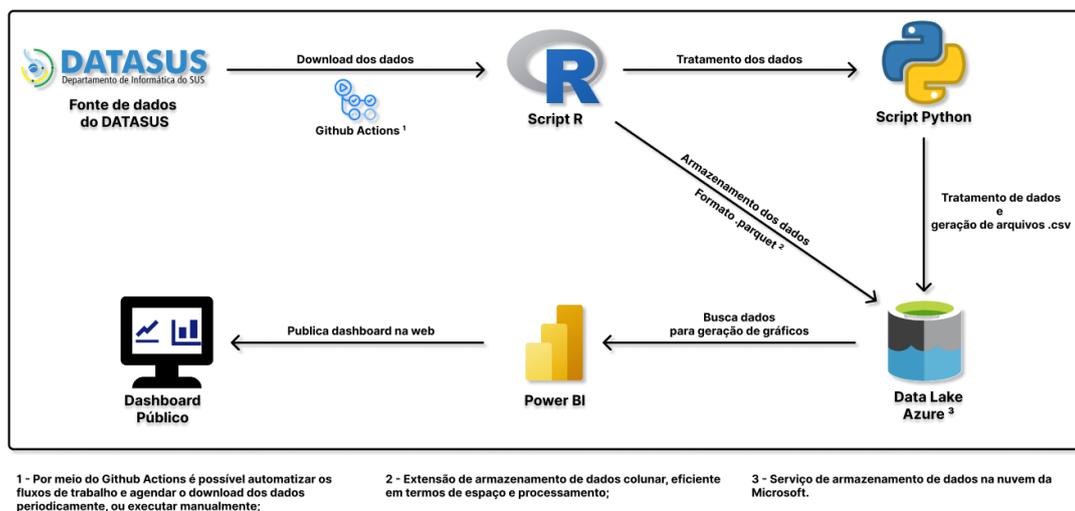


Figura 14 – Fluxo de processos - Fonte: Autor.

Em seguida, utiliza-se a linguagem Python para o processamento de dados separados por datamart (conforme detalhado na Subseção 4.2.1). Dentre os processos executados, destacam-se a exclusão de dados duplicados, combinação de colunas, remoção de colunas não utilizadas e criação de novas colunas.

Após o tratamento dos dados, são gerados arquivos em formato CSV, que são enviados ao DataLake da plataforma Azure. A partir dessa etapa, o PowerBI conecta-se ao DataLake para buscar os dados já transformados, a fim de gerar painéis e gráficos. Em seguida, um endereço público na internet é disponibilizado para acesso aos painéis na web.

4.2.3 Atividade 8 - Implementar os serviços de ETL

esta atividade está atrelada a Fase 2, cuja meta é alcançar o Objetivo Específico 2 (OE2): Centralizar as informações dos SIS.

O resultado da Atividade 8 são os códigos em Python e R responsáveis pelo processo de ETL do projeto. Inicialmente descreveremos como o agendamento da extração de dados foi executada, utilizando a automação do Github Actions, baseada na periodicidade dos indicadores de saúde selecionados a partir da priorização da atividade A5, na Subseção 4.1.5, garantindo que os dados estejam sempre atualizados para a análise dos profissionais da RAPS. Em seguida, abordaremos as técnicas e ferramentas que empregamos para a extração de dados, incluindo a linguagem R, com a biblioteca "microdatasus", e o Python, fazendo uso de várias bibliotecas, como aquelas fornecidas pelo DataSUS e IBGE, permitindo-nos coletar e agregar dados de fontes diversas.

O algoritmo abaixo, em linguagem R, ilustra o funcionamento de uma extração que, por meio da biblioteca "microdatasus", realiza uma busca de dados SIH para uma

Unidade Federativa específica. Esse processo utiliza os seguintes argumentos: "SIH-RD", "month_start", "month_end", "year_start" e "year_end"; para determinar o período de busca, que inclui o mês de início, o mês de término, o ano de início e o ano de término dos dados a serem buscados. Esses dados são então armazenados em um *dataframe*, um tipo de estrutura de dados na linguagem R. Em seguida, o código determina o nome do arquivo com base nos argumentos de entrada e o converte para o formato .parquet antes de salvá-lo, um formato eficiente para armazenamento e processamento de dados em colunas.

```
1 if (!require("microdatasus")) {
2   remotes::install_github("rfsaldanha/microdatasus")
3 }
4 library(arrow)
5 library(microdatasus)
6 library(lubridate)
7
8 sistema_informacao <- "SIH-RD"
9
10 args <- commandArgs(trailingOnly = TRUE)
11
12 if (length(args) != 2) {
13   print("Dois argumentos obrigatórios: ano_mes (no formato YYYY-MM)
14   e uf com dois bytes (AM, DF etc.)")
15   stopifnot(FALSE)
16 }
17
18 uf <- args[2]
19 ano_mes_atual <- args[1]
20 ano <- substr(ano_mes_atual, 1, 4)
21 mes <- substr(ano_mes_atual, 6, 7)
22
23 nome_arquivo <- paste0(ano_mes_atual, "-", uf, "-",
24 sistema_informacao, ".parquet")
25
26 dados_originais <- fetch_datasus(month_start = mes, month_end = mes,
27 year_start = ano, year_end = ano, uf = uf,
28 information_system = sistema_informacao)
29 dados_pre_processados <- process_sih(dados_originais)
30
31 write_parquet(
32   dados_pre_processados <- process_sih(dados_originais),
33   nome_arquivo
34 )
```

Listing 4.1 – Extração de dados do SIH em R

Este código pode ser executado manualmente na linha de comando ou agendado

para execução. Uma maneira de executá-lo é após instalar os pré-requisitos e configurar as variáveis de ambiente, executar o arquivo fornecendo o ano, o mês e a Unidade Federativa (UF) para os quais você deseja obter os dados, conforme exemplificado na execução:

```
1 Rscript.exe extrair-sih.R 2022-12 DF
```

Listing 4.2 – Execução do código de extração em R

Após a aquisição dos dados, a transformação deve ocorrer e, conforme mencionado na Subseção 4.2.1, cada *datamart* é classificado por camadas (*bronze*, *silver* e *gold*) e a raiz do *datamart* contém o código responsável por transformar esses dados e armazenar na respectiva camada. O diretório de um *datamart* também é composto pelo *script* em *Python* responsável pela transformação de dados, pelo *requirements*, que lista as dependências necessárias para a execução do *script*, e pelo código em R chamado *dbc_to_csv.R*, completo no Anexo I, que converte os arquivos com extensão *.dbc* (um tipo de extensão extraída pelo DATASUS) em arquivos *.csv* (dados separados por vírgula), para serem lidos pelo Power BI. A Figura 15 mostra o diretório do *datamart*, que contém cada SIS e, no diretório de cada SIS, existem as camadas e o código responsável pela transformação de dados necessária.

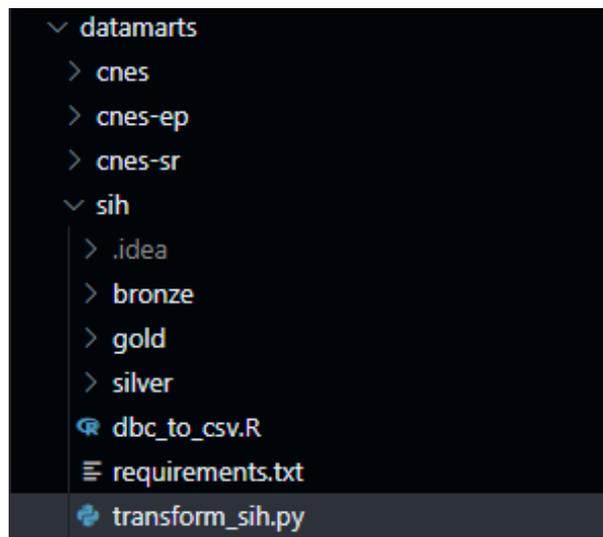


Figura 15 – Diretório de datamarts - Fonte: Autor.

O trecho de código abaixo (código completo no Anexo K) mostra uma função que tem como finalidade ler um arquivo pertencente a camada "bronze" do SIH, efetuar as transformações necessárias e gerar um arquivo no formato .csv na camada "silver".

```

1 def gerar_arquivo_categorias():
2     df_saida = pd.DataFrame()
3     for ano in ANOS:
4         for mes in MESES:
5             df_sih = pd.read_csv(f"bronze/RDDF{ano}{mes}.csv",
6                                 dtype={"DT_INTER": str, "UF_ZI": str})
7             df_sih["DIAG_PRINC"] = df_sih["DIAG_PRINC"].str.slice(0, 3)
8             df_sih = df_sih[
9                 (df_sih["DIAG_PRINC"].str.startswith("F")) | (
10                    (df_sih['DIAG_PRINC'] >= 'X60') &
11                    (df_sih['DIAG_PRINC'] < 'X85')))]
12
13             df_sih["ESTADO_CODIGO"] = df_sih["UF_ZI"].str.slice(0, 2)
14             df_sih = df_sih[["ESTADO_CODIGO", "DIAG_PRINC", "ANO_CMPT",
15                             "MES_CMPT"]]
16             df_saida = pd.concat([df_saida, df_sih], axis=0)
17
18             totalizador = df_saida.groupby(["ESTADO_CODIGO", "DIAG_PRINC",
19                                             "ANO_CMPT", "MES_CMPT"]).size().reset_index(name="TOTAL")
20             totalizador.to_csv("silver/2018-2022-diag_princ-sih-df.csv",
21                               sep=";", index=False)

```

Listing 4.3 – Transformações em Python

Este trecho de código trata dados adquiridos do SIH, realizando um filtro de dados de internação por saúde mental, compreendendo o IS06 da Figura 11 e aborda este filtro por meio da coluna "DIAG_PRINC", que indica o código do CID-10 (Classificação Internacional de Doenças) da doença, que é uma lista da OMS de classificação médica com códigos para sintomas, queixas e doenças. Ao examinar os dados disponíveis no DATASUS (DATASUS, 2021), foi identificado que as doenças cujos códigos CID começam com "F" e os códigos que variam de "X60" a "X85" são indicativos de transtornos mentais, problemas psicológicos, distúrbios, doenças, lesões e intoxicações relacionadas ao uso de álcool e substâncias psicoativas.

De forma técnica, entre as transformações realizadas encontram-se a adição da coluna "ESTADO_CODIGO", que extrai os dois primeiros caracteres do campo "UF_ZI", a seleção das colunas para compor o novo conjunto de dados, a concatenação dos *dataframes* resultantes em um *dataframe* final, o agrupamento dos dados por "ESTADO_CODIGO", "DIAG_PRINC", "ANO_CMPT" e "MES_CMPT", o cálculo do total de registros em cada grupo e, por fim, o armazenamento dos dados processados em um arquivo .csv na camada "silver".

Com a Extração e Transformação de dados devidamente realizadas, o próximo passo consiste na integração destes dados no *Data Lake* do Microsoft Azure, um repositório para armazenamento de dados. Para atingir esse objetivo, foram desenvolvidos fluxos de trabalho, como o fluxo descrito a seguir, que são agendados para execução de acordo com a periodicidade determinada pelo indicador de saúde.

```
1 name: Gerar arquivo SIH-RD
2
3 on:
4   workflow_dispatch:
5     inputs:
6       ano_mes:
7         description: "Ano e ms no formato YYYY-MM.
8         Exemplo 2023-01"
9         required: false
10      uf:
11        description: "UF para o qual o arquivo vai ser gerado.
12        Exemplo AM, DF. Obrigatrio se ano_mes informado."
13        required: false
14      schedule:
15        - cron: "0 0 28 * *"
16 env:
17   NOME_BASE_ARQUIVO_DEFAULT: "SIH-RD.parquet"
18   UF_DEFAULT: "DF"
19
20 jobs:
21   extrair-transformar-enviar:
22     runs-on: ubuntu-latest
23     steps:
24       - name: Checkout do repositorio
25         uses: actions/checkout@v2
26
27       - name: Configurar ambiente R
28         uses: r-lib/actions/setup-r@v2
29         with:
30           use-public-rspm: true
31
32       - name: Configurando dependencias
33         uses: r-lib/actions/setup-renv@v2
34         with:
35           cache-version: 2
36
37       - name: Definir valores computados para ano_mes e uf - execucao manual
38         if: github.event.inputs.ano_mes != ''
39         run: |
40           echo "ANO_MES_COMPUTADO=${{github.event.inputs.ano_mes}}"
41           >> $GITHUB_ENV
42           echo "UF_COMPUTADO=${{github.event.inputs.uf}}" >> $GITHUB_ENV
```

```

42
43   - name: Definir nome final do arquivo
44     run: |
45       echo "NOME_ARQUIVO_COMPUTADO=${{ env.ANO_MES_COMPUTADO }}-
46         ${{ env.UF_COMPUTADO }}-${{ env.NOME_BASE_ARQUIVO_DEFAULT }}"
47       >> $GITHUB_ENV
48
49   - name: Executar script R
50     run: Rscript extrair-transformar-enviar-sih.R
51         ${{ env.ANO_MES_COMPUTADO }} ${{ env.UF_COMPUTADO }}
52
53   - name: Salvar arquivo como artefato
54     uses: actions/upload-artifact@v2
55     with:
56       name: ${{ env.NOME_ARQUIVO_COMPUTADO }}
57       path: ${{ env.NOME_ARQUIVO_COMPUTADO }}
58
59   - name: Azure Login
60     uses: azure/login@v1
61     with:
62       creds: ${{ secrets.AZURE_CREDENTIALS }}
63
64   - name: Azure CLI - Envio
65     uses: azure/CLI@v1
66     with:
67       azcliversion: 2.30.0
68       inlineScript: |
69         chmod +x $GITHUB_WORKSPACE/upload_file.sh
70         $GITHUB_WORKSPACE/upload_file.sh \
71         ${{ secrets.ACCOUNT_NAME }} \
72         ${{ secrets.ACCOUNT_KEY }} \
73         ${{ secrets.CONTAINER_NAME }} \
74         ${{ env.NOME_ARQUIVO_COMPUTADO }} \
75         sia-pa/${{ env.NOME_ARQUIVO_COMPUTADO }}

```

Listing 4.4 – Fluxo de trabalho em .yaml

Este fluxo de trabalho pode ser executado manualmente com os parâmetros "ano_mes" e "uf", mas também é programado para ser executado às 0h do 28º dia do mês, via cron(adicionar rodapé), como descrito em *workflow_dispatch*, e utiliza dois valores padrões para o nome do arquivo e para a UF - *SIH-RD.parquet* e "DF".

O trabalho a ser executado é um *job* denominado "extrair-transformar-enviar", que é composto por dez passos. O primeiro é o "Checkout do repositório", que verifica o código do repositório atual; o segundo é o "Configurar ambiente R", que configura o ambiente utilizando ações fornecidas pelo repositório "r-lib/actions/setup-renv@v2"; o terceiro é o "Configurando dependências", que configura as dependências fornecidas pelo repositório,

com o cache habilitado; o quarto é o "Definir valores computados para ano_mes e uf - execucao manual", que define as variáveis de ambiente com base nos valores padrões ou nos valores fornecidos manualmente; o sexto é o "Definir nome final do arquivo", que define o nome do arquivo com base nas variáveis de ambiente; o sétimo é o "Executar script R", que executa o *script* responsável pela busca, transformação e envio dos dados ao Data Lake; o oitavo é o "Salvar arquivo como artefato", que realiza o carregamento do arquivo gerado como um artefato no Github, que pode ser utilizado para ser usado em outras etapas e se torna disponível para baixar; o nono é o "Azure Login", que autentica no Azure utilizando as credenciais fornecidas pelo *secrets*; o décimo é o "Azure CLI - Envio", que utiliza o Azure CLI para enviar o arquivo para o Data Lake da Azure, utilizando o caminho e as credenciais para a execução de um *script* denominado *upload_file.sh* (disponível no Anexo L), que utiliza o Azure CLI para o carregamento do arquivo.

4.2.4 Atividade 9 - Definir arquitetura do Data Warehouse

esta atividade está atrelada a Fase 2, cuja meta é alcançar o Objetivo Específico 2 (OE2): Centralizar as informações dos SIS.

O resultado da Atividade 9 é o Documento de Arquitetura do DW e, como mencionado nas Seções 3.2 e 4.2, o documento deve abordar conceitos de requisitos, arquitetônicos, a solução técnica, processos operacionais e integração.

O "Documento de Indicadores de Saúde (R1)", presente na Seção 4.1.3, serve como o ponto de partida para a definição dos requisitos do nosso Data Warehouse. Este documento definido com base nas necessidades da RAPS, desempenha um papel central na identificação dos indicadores e dados armazenados e disponibilizados por meio do sistema.

Com base nesse contexto, a equipe de desenvolvimento fez um levantamento de requisitos funcionais e não funcionais do sistema. Esses requisitos foram categorizados e classificados em épicos, como exemplificado na Tabela 7.

Tipo de requisito	Épicos
Requisitos funcionais	EP01 - Atenção Especializada
	EP02 - Atenção Hospitalar
	EP03 - Residências
	EP04 - Sistema
Requisitos não funcionais	EP05 - Usabilidade
	EP06 - Confiabilidade
	EP07 - Segurança
	EP08 - Portabilidade
	EP09 - Desempenho
	EP10 - Escalabilidade
	EP11 - Manutenibilidade

Tabela 7 – Classificação por épicos - Fonte: Autor.

Os requisitos funcionais, descritos nas Tabelas 8 e 9, especificam as funcionalidades e capacidades que o sistema deve oferecer para atender às necessidades dos profissionais de saúde da RAPS.

Épico	ID	Requisito
Atenção Especializada	RF01	O usuário deve ser capaz de visualizar a cobertura de equipes NASF e ESF nos municípios de pequeno porte, ou seja, aqueles com menos de 15 mil habitantes.
	RF02	O usuário deve ser capaz de visualizar o número de municípios que não possuem equipes do NASF cadastrados na Atenção Primária e, ao mesmo tempo, possuem cobertura de Estratégia de Saúde da Família superior a 50%.
	RF03	O usuário deve ser capaz de visualizar a distribuição das equipes de consultório na rua implantadas e habilitadas, ou seja, que estão atuando no território, conforme a modalidade (I, II e II).
	RF04	O usuário deve ser capaz de visualizar o crescimento e expansão de novos CAPS por ano.
	RF05	O usuário deve ser capaz de visualizar o crescimento e expansão de CAPS habilitados por tipo anualmente.
	RF06	O usuário deve ser capaz de visualizar a cobertura dos CAPS por 100.000 habitantes, tanto para a UF, como para o município.
	RF07	O usuário deve ser capaz de visualizar a quantidade de CAPS habilitados conforme o quantitativo de municípios elegíveis pertencentes ao mesmo estado, anualmente. Análise somente para os estados, representa o percentual de CAPS habilitados em comparativo com o total de CAPS que poderiam estar habilitados conforme a quantia de ME para o Estado.

Tabela 8 – Requisitos Funcionais - Fonte: Autor.

Épico	ID	Requisito
Atenção Hospitalar	RF08	O usuário deve ser capaz de visualizar a distribuição (n° e %) de hospitais com leitos de saúde mental habilitados por tipo.
	RF09	O usuário deve ser capaz de visualizar o porte populacional dos Municípios onde estão localizados os Serviços Hospitalares de Referência em saúde mental.
	RF10	O usuário deve ser capaz de visualizar a distribuição em porcentagem (%) dos Leitos SUS em Hospitais Psiquiátricos por UF, anualmente e por porte hospitalar.
Residências	RF11	O usuário deve ser capaz de visualizar o quantitativo de Unidades de Acolhimentos em Funcionamento Habilitadas.
	RF12	O usuário deve ser capaz de visualizar o quantitativo de Serviços Residenciais Terapêuticos por UF.
	RF13	O usuário deve ser capaz de visualizar a distribuição da Habilitação de Serviços Residenciais Terapêuticos por tipo, UF e número de moradores.
Sistema	RF14	Os gráficos exibidos nos painéis devem permitir que o usuário filtre os dados apresentados de acordo com a variável selecionada.
	RF15	Os painéis devem incluir um filtro de dados por ano que permita ao usuário selecionar um ano específico para visualização.
	RF16	Os painéis devem conter uma descrição clara e concisa da informação exibida, a fim de fornecer uma compreensão geral do conteúdo apresentado.
	RF17	A plataforma deve ser capaz de coletar dados de saúde mental diretamente do DATASUS e quaisquer outras fontes adicionais necessárias para o enriquecimento do dado.
	RF18	A plataforma deve ser capaz de armazenar e gerenciar os volumes de dados de saúde mental necessários ao indicador específico a ser extraído.
	RF19	A plataforma deve permitir a realização de análises de dados, como agrupamento de dados com base em critérios específicos, identificação de tendências e padrões de comportamento, entre outros.
	RF20	A plataforma deve ser capaz de gerar relatórios personalizados com base nas necessidades dos usuários.

Tabela 9 – Requisitos Funcionais - Fonte: Autor.

Já os requisitos não funcionais, descritos na Tabela 10, especificam atributos de qualidade e restrições que o sistema deve atender para garantir alto desempenho, segurança e confiabilidade.

Todas as atividades anteriores, junto aos requisitos levantados, em conjunto com os requisitos levantados, convergem para os painéis de monitoramento propostos. Entretanto, é fundamental ressaltar que o ciclo completo do processo ETL se consolida no Data Lake, o núcleo central para o armazenamento do Data Warehouse.

Tipo	ID	Requisito Não Funcional
Usabilidade	RNF01	Os gráficos devem ser informativos, permitindo a visualização dos valores ao passar o cursor.
	RNF02	Os gráficos e tabelas devem ser interativos e clicáveis.
Confiabilidade	RNF03	Os painéis devem estar disponíveis durante todo o período comercial acordado.
Segurança	RNF04	O uso de dados deve estar de acordo com a LGDP.
Portabilidade	RNF05	Os painéis devem ser compatíveis com os navegadores Microsoft Edge, Google Chrome (versão mais recente), Safari e Firefox.
Desempenho	RNF06	A plataforma deve ser efetiva na geração dos dados segundo a periodicidade necessária que deverá ser mensal ou anual (de acordo com o indicador) para manter o nível de gratuidade das ferramentas utilizadas.
Escalabilidade	RNF07	A plataforma deve ser capaz de lidar com um grande número de usuários e de dados, permitindo sua expansão conforme necessário.
Manutenibilidade	RNF08	A plataforma deve ser fácil de manter e atualizar, com um mínimo de tempo de inatividade para os usuários.

Tabela 10 – Requisitos Não Funcionais - Fonte: Autor.

Acerca da implementação dos requisitos não funcionais, os requisitos com épico Confiabilidade, Portabilidade, Desempenho, Portabilidade e Escalabilidade são totalmente dependentes da Microsoft, tendo em vista que utilizam da função "Publicar" para hospedar na *Web* e, desta forma, é possível cobrir estes requisitos, já que a Microsoft agrega o monitoramento da saúde do serviço, a mitigação de incidentes e possui uma rápida resposta a incidentes com uma solução chamada TTNO (Time To Notify "O"), um serviço automatizado de notificação de incidentes que utiliza sua infraestrutura de alerta para identificar cenários e clientes específicos que são afetados por um incidente (Microsoft, 2023).

Já sobre a implementação dos requisitos não funcionais de épico de Usabilidade, os mesmos contam com configurações ao desenvolver um gráfico, como mencionado na Subseção 4.3.3. Por fim, acerca do requisito com épico de Manutenibilidade, estes são dependentes da documentação do projeto e do código, mas é importante ressaltar que, ao atualizar um indicador de saúde, não há a indisponibilidade dos mesmos.

Ao longo da implementação do projeto, a estrutura do Data Lake foi empregada de duas maneiras distintas: por meio do *Azure Storage Gen 2* e do Github, e ambas as abordagens atenderam os requisitos do projeto. Contudo, visando agilizar o processo de desenvolvimento diante do curto prazo, optou-se pelo Github como plataforma para o armazenamento dos arquivos. Essa escolha, durante a fase de desenvolvimento, não exclui a viabilidade da utilização do *Azure Storage Gen 2*. Pelo contrário, este é considerado

como a opção primária para o ambiente de produção devido à sua integração direta com o Power BI. Dessa forma, quando os dados são atualizados no *Azure Storage Gen 2*, os painéis do Power BI conseguem refletir automaticamente essas mudanças em seus valores.

4.3 Fase 3

A fim de alcançar o objetivo de Prover painéis de monitoramento (OE3), as subseções seguintes mostram os resultados das atividades da Fase 3. Como demonstrado na Tabela 1 e na Figura 5, a Fase 2 deve cumprir os seguintes resultados:

- Protótipo de Alta Fidelidade O resultado R4 consiste em uma parte da etapa de desenvolvimento que inclui a criação de um Protótipo de Alta Fidelidade, que servirá como uma representação visual da interface do usuário do sistema. Esse protótipo se concentrará em fornecer uma visão detalhada das interações, layouts e fluxos de usuário dentro do sistema.
- Dashboard O resultado R5 consiste no Dashboard que é a interface principal por meio da qual os profissionais de saúde da RAPS acessarão e irão interagir com os indicadores de saúde.

4.3.1 Atividade 10 - Desenvolver protótipo

esta atividade está atrelada a Fase 3, cuja meta é alcançar o Objetivo Específico 3 (OE3): Prover painéis de monitoramento.

O protótipo de alta fidelidade foi desenvolvido utilizando a ferramenta Figma. Inicialmente, foram consideradas opções de abordagem com base nos indicadores de saúde pois, através destes indicadores, foi possível observar que cada um possui uma fonte de dados (SIS), uma frequência de atualização e uma forma de interpretação da informação.

Com base nesse contexto, e conforme ilustrado na Figura 16, foi elaborada uma página para demonstrar como essas informações poderiam ser apresentadas. Um menu lateral foi proposto para facilitar a navegação entre os dados dos diferentes SIS, juntamente com duas caixas de seleção, identificadas na Figura 16 como "Nível" e "Local". Além disso, a partir dessa página, o esquema de disposição das informações e cores pode ser replicado para as demais telas.

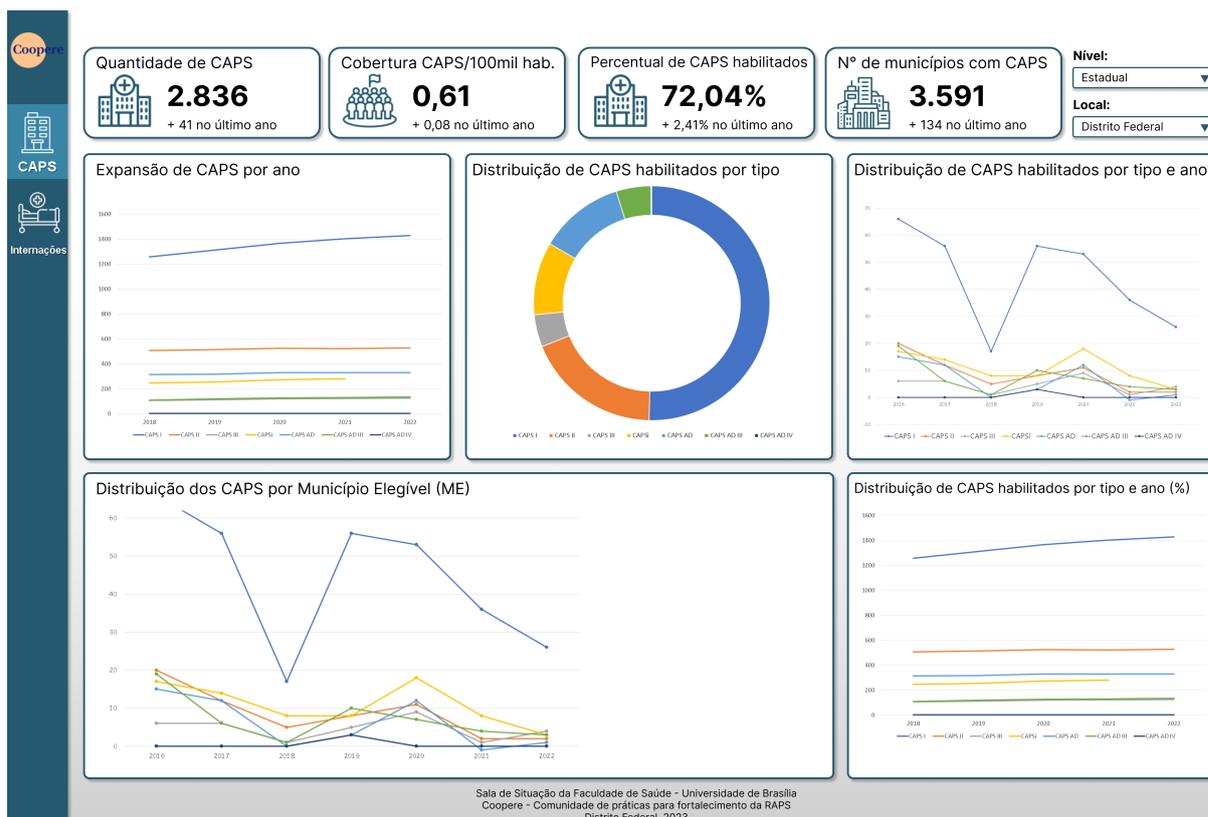


Figura 16 – Protótipo de Alta Fidelidade. Fonte: Autor.

Uma reunião ocorreu para a validação deste protótipo, onde o principal *stakeholder*, Christiane Kanzler, pontuou os seguintes aspectos nos quais podem ser melhorados:

- a) Gráficos de Pizza e "Donuts": o tipo de gráfico não é adequado para análise com crescimento, além de não ter uma disposição de fácil visualização com valores pequenos;
- b) Cores: apesar da utilização da paleta de cores da Coopere, deve-se haver uma distribuição melhor entre o azul e o damasco;
- c) Filtro: o filtro das informações por meio de mapas interativos pode facilitar para escolha entre níveis regionais, estaduais e municipais;
- d) Navegabilidade: a navegação por SIS pode ser extensa, a navegação entre os painéis por módulo de atendimento da RAPS facilitaria o gestor.

A partir das pontuações obtidas no protótipo, a equipe trabalhou nas melhorias dentro do próprio Power BI, visando agilizar o desenvolvimento diante do curto prazo estabelecido. A cada adição de gráficos, tabelas, botões e outros recursos, os *stakeholders* (Jonas Brant e Christiane Kanzler) eram consultados e, nesse momento, uma avaliação

informal era realizada para identificar as modificações necessárias. Por fim, foi realizada uma validação final de toda a parte visual dos painéis, juntamente com as funcionalidades já desenvolvidas.

4.3.2 Atividade 11 - Criar conexão com o DW

esta atividade está atrelada a Fase 3, cuja meta é alcançar o Objetivo Específico 3 (OE3): Prover painéis de monitoramento.

Nesta etapa, cabe esclarecer os detalhes relativos ao armazenamento de dados, procedimentos que são disponibilizados pelos serviços de nuvem e hospedagem da Microsoft e do Github. Para o envio dos arquivos tratados para o armazenamento da Azure, é necessário o uso das *actions* de Login e de CLI. O repositório Azure/login fornece uma ação do GitHub que permite autenticar e executar scripts do Azure CLI e Azure PowerShell (AZURE, 2023a). Enquanto o repositório Azure/cli é uma interface de linha de comando multi plataforma para gerenciar recursos da Azure, fornecendo comandos de gerenciamento de serviços e recursos (AZURE, 2023b).

Para realizar a autenticação de aplicações com os serviços da Azure, o aplicativo deve ser registrado, as permissões devem ser definidas e então são criadas entidades de serviço para cada desenvolvedor. E para o desenvolvimento local, as variáveis de ambiente foram definidas com a identidade da entidade criada, e então a SDK da Azure utiliza essas informações para autenticar a aplicação aos recursos necessários (AZURE, 2023a).

Para realizar a conexão do Power BI com as fontes de dados presentes do Data Lake, selecionou-se a opção "Azure Data Lake Storage Gen2" e então adicionou-se o endereço referente ao Data Lake. A partir disto, e como mostra a Figura 17, é possível carregar todos os arquivos no formato *.parquet* para utilização no Power BI, enquanto o Anexo M complementa esta visualização, mostrando o arquivo já aberto.

Content	Name	Extension	Date accessed	Date modified	Date created	Attributes	Folder Path
Binary	2015-2015-all-cnes-st.parquet	.parquet	null	27/02/2023 23:24:33	null	Record	https://unbdatasus.dfs.core.v
Binary	2015-2022-df-cnes-st.parquet	.parquet	null	27/02/2023 23:20:52	null	Record	https://unbdatasus.dfs.core.v
Binary	2016-2016-all-cnes-st.parquet	.parquet	null	27/02/2023 23:24:32	null	Record	https://unbdatasus.dfs.core.v
Binary	2017-2017-all-cnes-st.parquet	.parquet	null	27/02/2023 23:24:38	null	Record	https://unbdatasus.dfs.core.v
Binary	2018-2018-all-cnes-st.parquet	.parquet	null	27/02/2023 23:24:55	null	Record	https://unbdatasus.dfs.core.v
Binary	2019-2019-all-cnes-st.parquet	.parquet	null	27/02/2023 23:25:43	null	Record	https://unbdatasus.dfs.core.v
Binary	2020-2020-all-cnes-st.parquet	.parquet	null	27/02/2023 23:27:04	null	Record	https://unbdatasus.dfs.core.v
Binary	2021-2021-all-cnes-st.parquet	.parquet	null	27/02/2023 23:27:18	null	Record	https://unbdatasus.dfs.core.v
Binary	2022-2022-all-cnes-st.parquet	.parquet	null	27/02/2023 23:27:17	null	Record	https://unbdatasus.dfs.core.v
Binary	2015-2015-df-sia-pa.parquet	.parquet	null	27/02/2023 23:38:07	null	Record	https://unbdatasus.dfs.core.v
Binary	2016-2016-df-sia-pa.parquet	.parquet	null	27/02/2023 23:38:34	null	Record	https://unbdatasus.dfs.core.v
Binary	2017-2017-df-sia-pa.parquet	.parquet	null	27/02/2023 23:37:46	null	Record	https://unbdatasus.dfs.core.v
Binary	2018-2018-df-sia-pa.parquet	.parquet	null	27/02/2023 23:38:34	null	Record	https://unbdatasus.dfs.core.v
Binary	2019-2019-df-sia-pa.parquet	.parquet	null	27/02/2023 23:39:22	null	Record	https://unbdatasus.dfs.core.v
Binary	2020-2020-df-sia-pa.parquet	.parquet	null	27/02/2023 23:39:46	null	Record	https://unbdatasus.dfs.core.v
Binary	2021-2021-df-sia-pa.parquet	.parquet	null	27/02/2023 23:39:54	null	Record	https://unbdatasus.dfs.core.v
Binary	2022-12-SIA-PA.parquet	.parquet	null	28/02/2023 05:33:01	null	Record	https://unbdatasus.dfs.core.v
Binary	2022-2022-df-sia-pa.parquet	.parquet	null	27/02/2023 23:40:22	null	Record	https://unbdatasus.dfs.core.v

Combinar Carregar Transformar Dados Cancelar

Figura 17 – Carregamento dos arquivos *.parquet* Fonte: Autor.

4.3.3 Atividade 12 - Desenvolver o dashboard no Power BI

esta atividade está atrelada a Fase 3, cuja meta é alcançar o Objetivo Específico 3 (OE3): Prover painéis de monitoramento.

Esta etapa aproveita os resultados já obtidos (R1 a R4) e os integra em um painel de monitoramento interativo. A criação de gráficos, tabelas e outras visualizações representa o principal desta atividade, já que elas permitirão os profissionais de saúde da RAPS analisar os dados de saúde.

Conforme mencionado na Seção 4.3.1, o *layout* do *dashboard* foi realizado a partir dos *feedbacks* do protótipo e também a partir de validações diretas com os *stakeholders* em tempo de desenvolvimento, para maior agilidade no desenvolvimento. Este *layout* foi organizado em várias páginas, como mostra a Figura 18, classificadas pelo tipo do dado que aborda e pelo SIS fonte dos dados, e a escolha das cores favorecendo a paleta da Coopere.

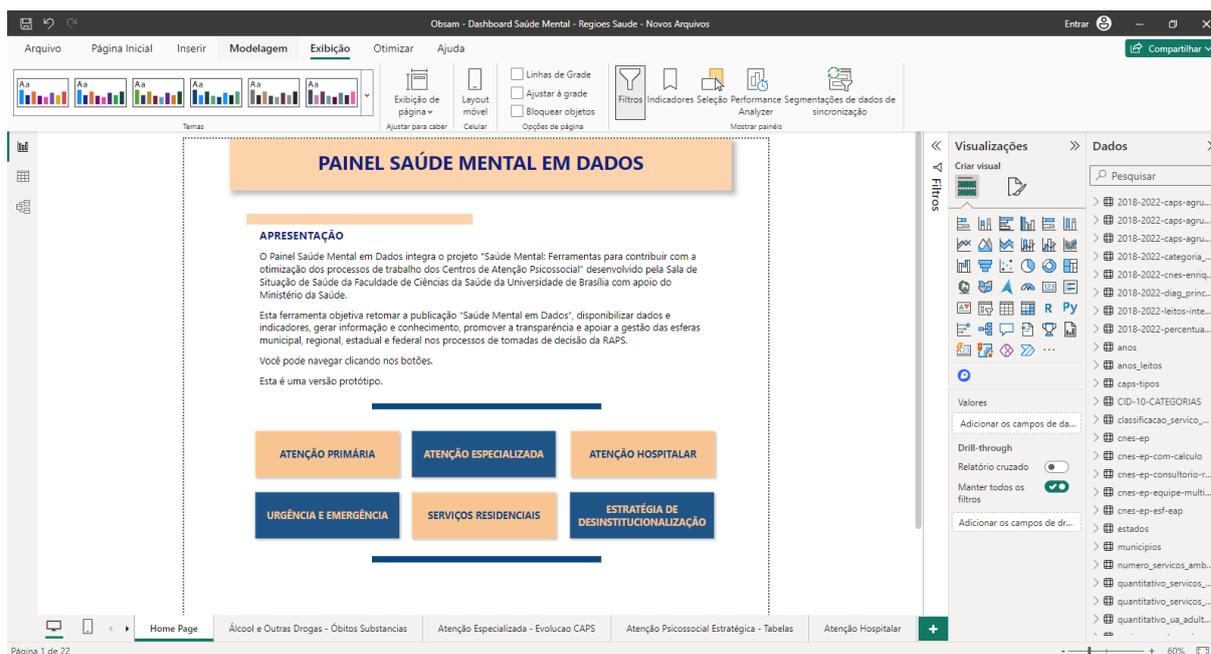


Figura 18 – Tela 1 - Menu. Fonte: Autor.

Para alimentar o *dashboard*, foi configurado diversas fontes de dados que são atualizadas de acordo com a periodicidade do Indicador de Saúde. A Figura 19 mostra todas as fontes de dados existentes e configuradas no Power BI, através das opções "Obter dados» "Texto/CSV".

A Figura 20 aborda o indicador de saúde IS01 através do gráfico de linha (parte inferior) que mostra o crescimento e expansão de novos CAPS, de forma anual. Desta forma, é possível visualizar a quantidade de CAPS existentes em cada ano, de acordo com a opção e com o estado selecionado na parte superior da tela. Também foi utilizado o *Mapbox* (tecnologia de mapas que inclui SDK) para a criação dos mapas interativos presentes neste painel, que mostra a cobertura por estado e região de saúde, permitindo que os profissionais de saúde da RAPS visualizem de maneira mais clara as regiões que possuem desfalque de CAPS, porém este não será enfatizado neste trabalho por não ser um IS de prioridade.



Figura 19 – Fontes de dados - Fonte: Autor.

Por fim, todas as outras telas do sistema desenvolvidas no Power BI estão disponíveis nos Anexos N em diante. Os painéis foram hospedados por meio das opções "Arquivo» "Publicar» "Publicar no Power BI", e estão disponíveis na [web](#)¹. Além disso, o código completo do projeto pode ser encontrado no [Github](#)².

¹ <https://app.powerbi.com/view?r=eyJrIjoiYzVlM2M4MDItMDQ2ZC00MjI3LTkyMjMtMmFiYW14M2YzZTMxliwidCI6ImVjMzU5YmExLTZyMGItNGQyYi1iODMzLWM4ZTZkNDhmODA1OSJ9&pageName=ReportSection>

² <https://github.com/gleytonLima/datasus-etl>

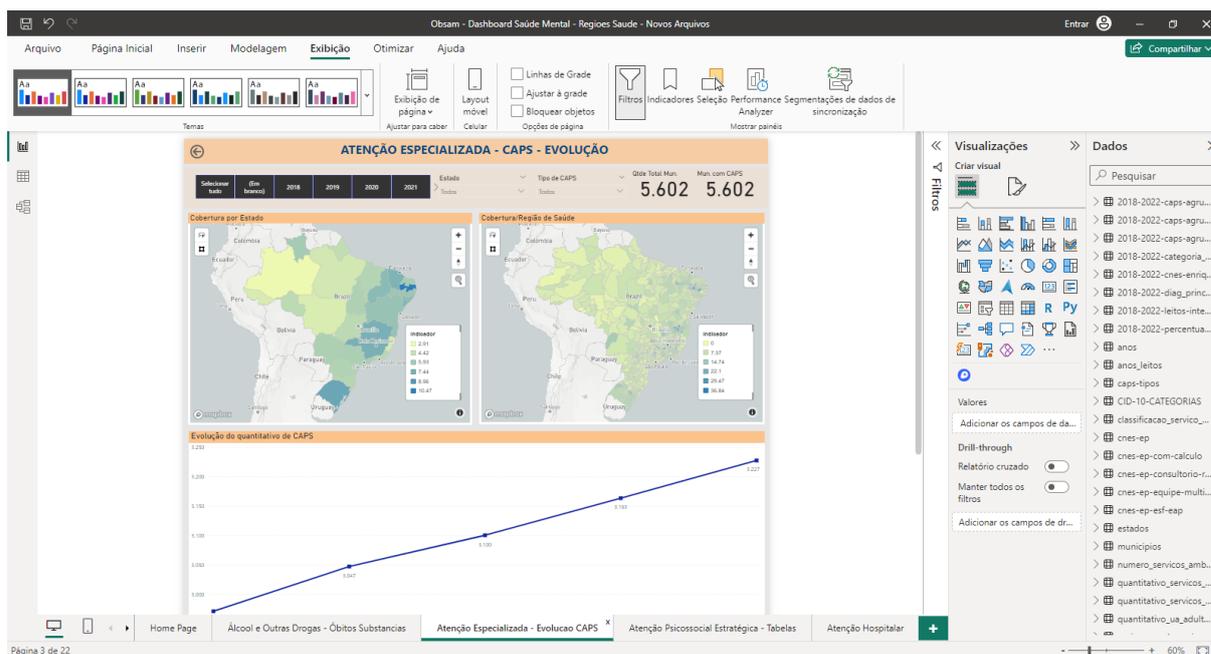


Figura 20 – Tela 3 - Evolução de CAPS. Fonte: Autor.

4.4 Fase 4

A fase 4 tem como objetivo Analisar os estudos e painéis desenvolvidos (OE4) e produzir a Análise postmortem (R6). Para isto, a fase 4 utiliza como entrada todo o conteúdo do Estudo 1 (fases 1, 2 e 3), atividades, artefatos e resultados.

O desenvolvimento e implementação dos painéis desenvolvidos é repleta de desafios e aprendizados, o postmortem presente nesta seção busca refletir sobre essa experiência, sucessos, falhas e desafios encontrados durante a execução do projeto. A análise procura formular criar uma crítica que deve contribuir com o entendimento de particularidades no desenvolvimento de soluções digitais (inclusive BI) no contexto da saúde.

4.4.1 Atividade 13 - Descrever a metodologia de análise

esta atividade está atrelada a Fase 4, cuja meta é alcançar o Objetivo Específico 4 (OE4): Analisar os estudos e painéis desenvolvidos.

A metodologia de análise postmortem adotada neste trabalho é baseada nos princípios da Atlassian, projetada para entender e remediar a causa raiz de cada incidente, documentando o incidente para referência futura, dessa maneira, é viável implementar ações preventivas para reduzir o impacto ou probabilidade de recorrência (ATLASSIAN, 2023).

A metodologia de postmortem da Atlassian é uma abordagem estruturada para analisar o desempenho e os resultados de um projeto ou atividade após sua conclusão, concentrando-se em quatro áreas principais: ações bem-sucedidas, ações de erro, lições

aprendidas e ações corretivas (ATLASSIAN, 2023).

4.4.2 Atividade 14 - Pontuar os objetos de análise

esta atividade está atrelada a Fase 4, cuja meta é alcançar o Objetivo Específico 4 (OE4): Analisar os estudos e painéis desenvolvidos.

Com o objetivo de identificar problemas e reflexões acerca da execução da metodologia de desenvolvimento do projeto, o objeto de análise explorado da Fase 1 é a atividade A2 - Realizar reuniões com integrantes do projeto.

Com o objetivo de analisar a estrutura do ETL, o objeto de análise explorado da Fase 2 é a atividade A8 - Implementar serviços de ETL.

Com o objetivo de refletir e discutir acerca do resultado final do projeto, o objeto de análise explorado da Fase 3 é o resultado R5 - Dashboard.

4.4.3 Atividade 15 - Desenvolver a análise postmortem

esta atividade está atrelada a Fase 4, cuja meta é alcançar o Objetivo Específico 4 (OE4): Analisar os estudos e painéis desenvolvidos.

Dando início pela A2 (Realizar reuniões com integrantes do projeto) - Fase 1, as reuniões via Jitsi para organizar as tarefas semanais e cíclicas foram bem-sucedidas, conforme descrito na metodologia de desenvolvimento na Seção 1.5. Todas as reuniões foram conduzidas de maneira focada no objetivo proposto e pautadas nos documentos do projeto, por meio do Google Docs e/ou Google Sheets - ferramentas de documentos e planilhas do Google. Isso garantiu clareza nas atividades e na fase atual de desenvolvimento do projeto.

Como ponto de erro, ocorreram algumas ausências de membros com papéis cruciais em determinadas reuniões, resultando em maior tempo para resolver problemas específicos e, conseqüentemente, atrasos em etapas do projeto.

Como medida corretiva e lição aprendida, seria ágil realizar reuniões extraordinárias, ajustando-se à disponibilidade dos membros essenciais para o escopo. Dessa forma, evitamos adiar a resolução de pendências apenas para o próximo ciclo, prevenindo possíveis atrasos.

Na A8 (Implementar serviços de ETL) - Fase 2, existem muitos pontos a serem destacados como bem-sucedidos, tais como a execução e funcionamento de todo o código que Extrai, Transforma e Carrega os dados, em Python e R.

Como ponto de erro pode-se considerar a falta de clareza acerca de determinada transformação de dados quando o objetivo é determinar "a qual indicador de saúde determinada função em Python se refere?".

Como lição aprendida podemos destacar o forte aprendizado e consolidação de conhecimentos acerca de análise de dados, linguagens Python e R, extração de dados do DataSUS, transformação de dados e arquivos e carregamentos. O manuseio de tecnologias como o *Github Actions* e o *Data Lake Storage Gen 2* também são aprendizados valiosos quando se trata de ETL.

Como medida corretiva, destaca-se aprimorar a documentação do código que vincula com o indicador de saúde que o mesmo resolve, uma vez que os algoritmos responsáveis pelo tratamento de dados frequentemente apresentam complexidade e dificuldade de leitura e compreensão. Essa documentação iria cooperar com a rastreabilidade de requisitos e indicador de saúde com a parte técnica do projeto, facilitando eventuais manutenções.

No entanto, essa iniciativa não foi implementada devido ao prazo de desenvolvimento do projeto, que envolvia a disponibilização de uma extensa quantidade de indicadores de saúde, dessa forma, o foco foi direcionado para a execução e entrega desses indicadores. Essa ação corretiva possui relação com a medida corretiva da A2, pois também está vinculada a prazos e à prevenção de atrasos.

No R5 (Dashboard) - Fase 3, podemos destacar como bem-sucedido os painéis terem atendido todos os requisitos funcionais e não funcionais.

Como ponto de erro, podemos destacar a densidade de informações presentes nos painéis, já que o dashboard possui 16 telas (incluindo menus), abordando dados de diversos SIS.

Como lição aprendida, destaca-se o aprendizado no manuseio do Power BI, uma ferramenta de análise de dados reconhecida por sua robustez e poder no mercado. Além disso, também cabe destacar a importância de uma análise contínua das necessidades dos usuários finais (gestores) e a consideração da experiência do usuário durante o desenvolvimento de dashboards.

Como ação corretiva recomenda-se a otimização da apresentação de informações nos painéis, visando equilibrar a densidade de dados com a usabilidade, podendo incluir a reorganização de gráficos e tabelas, a fim de facilitar a compreensão para os usuários.

5 Considerações Finais

Este trabalho de conclusão de curso alcançou os resultados planejados da fase 1 a fase 3, consolidando o desenvolvimento dos painéis. Além disso, foi desenvolvido uma análise e reflexão de determinados pontos do trabalho, que forneceram uma visão mais prática do desenvolvimento do dashboard, na fase 4.

A disponibilização dos painéis geraram um avanço no processo de tomada de decisão dos profissionais de saúde da RAPS, o sucesso evidente desse progresso se dá pela continuidade do desenvolvimento do projeto (Meta 2 da Coopere). Além disso, ao ser acessível online, o sistema não apenas facilita a rotina dos profissionais, mas também promove a transparência, já que qualquer cidadão tem acesso ao dashboard.

Ao concluir este estudo, esperamos que o sistema desenvolvido continue contribuindo e aprimorando os processos de análise de dados pelos gestores da Rede de Atenção Psicossocial, impactando na eficiência e dos serviços prestados pela RAPS. Este trabalho não apenas representa grandes esforços envolvidos ao longo do período do desenvolvimento do projeto, mas também uma contribuição para a comunidade que depende desse tipo de serviço, sendo uma transformação positiva no setor da saúde e tecnologia.

Referências

- ATLASSIAN. *Incident management for high-velocity teams*. 2023. <<https://www.atlassian.com/incident-management/handbook/postmortems>>, Acesso em: 12 de Novembro de 2023. Citado 2 vezes nas páginas 63 e 64.
- AWS. *O que é um data mart?* 2023. <<https://aws.amazon.com/pt/what-is/data-mart/>>, Acesso em: 10 de Novembro de 2023. Citado na página 45.
- AZURE. *Azure Login Action*. 2023. <<https://github.com/Azure/login>>, Acesso em: 12 de Agosto de 2023. Citado na página 59.
- AZURE. *GitHub Action for Azure CLI*. 2023. <<https://github.com/Azure/cli>>, Acesso em: 12 de Agosto de 2023. Citado na página 59.
- Biblioteca Virtual em Saúde - Ministério da Saúde. *Sistemas de Informação*. 2008. <https://bvsmis.saude.gov.br/bvs/svs/inf_sist_informacao.php>, Acesso em: 12 de Janeiro de 2023. Citado na página 9.
- BRAGHITTONI, R. *Business Intelligence Implementar do jeito certo e a custo zero*. [S.l.]: Casa do Código, 2017. Citado 2 vezes nas páginas 23 e 24.
- Brasil. *Lei Geral de Proteção de Dados Pessoais (LGPD)*. 2018. Disponível em: <https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm>. Citado na página 24.
- CASTRO, I. N. de. *O que é Benchmarking e qual a sua importância para o Marketing Digital*. 2020. Disponível em <<https://rockcontent.com/br/blog/benchmarking/>> Acesso em: 01 de Fevereiro de 2023. Citado na página 29.
- CAVALCANTE, R. B.; SILVA, P. C.; FERREIRA, M. N. *Sistemas de informação em saúde: Possibilidades e desafios*. p. 293, 2011. Disponível em: <<https://periodicos.ufsm.br/reufsm/article/view/2580/16433>>. Citado na página 9.
- Comunidade Coopere. *Coopere*. 2023. <<https://coopere.obsam.unb.br/portal/dw/>> (Acesso em: 15 de Janeiro de 2023). Disponível em: <<https://coopere.obsam.unb.br/>>. Citado 2 vezes nas páginas 10 e 29.
- Conass. *ESTRATÉGIA e-SUS ATENÇÃO BÁSICA E SISTEMA DE INFORMAÇÃO EM SAÚDE DA ATENÇÃO BÁSICA - SISAB*. 2013. <<https://www.conass.org.br/biblioteca/wp-content/uploads/2013/01/NT-07-2013-e-SUS-e-SISAB.pdf>>, Acesso em: 16 de Janeiro de 2023. Citado na página 19.
- CRUZ, F. O. *Sistema de informações - Gestão operacional de sistemas e serviços de saúde*. Rio de Janeiro: FIOCRUZ: [s.n.], 1988. Citado na página 15.
- DATASUS. *Arquivos em formato CSV*. 2021. <<http://www2.datasus.gov.br/cid10/V2008/descrcsv.htm>>, Acesso em: 04 de Abril de 2023. Citado na página 50.

DAUD, N. M. N.; BAKAR, N. A. A. A.; RUSLI, H. M. Implementing rapid application development (rad) methodology in developing practical training application system. In: *2010 International Symposium on Information Technology*. [S.l.: s.n.], 2010. v. 3, p. 1664–1667. Citado na página 26.

DRESNER, H. *Information Processing Management*. [S.l.: s.n.], 1989. Citado na página 23.

GUIMARAES, R. *História*. 2020. <https://wiki.saude.gov.br/sih/index.php/Página_principal> Acesso em: 21 de Janeiro de 2023. Citado na página 17.

INMON, W. H. *Building the Data Warehouse*. [S.l.: s.n.], 2005. fourth edition. Citado na página 21.

LESSA, F. J. D. et al. Novas metodologias para vigilância epidemiológica: uso do sistema de informações hospitalares - sih/sus. 2000. Disponível em: <http://scielo.iec.gov.br/scielo.php?script=sci_arttext&pid=S0104-16732000000500001>. Citado 2 vezes nas páginas 3 e 18.

LUHN, H. P. A business intelligence system. *IBM Journal of Research and Development*, v. 2, n. 4, p. 314–319, 1958. Disponível em: <<https://web.archive.org/web/20080913121526/http://www.research.ibm.com/journal/rd/024/ibmrd0204H.pdf>>. Citado na página 23.

MARTIN, J. Rapid application development. In: *Proceedings of the 10th Brazilian Congress of Mechanical Engineering*. [S.l.]: Macmillan Publishing Co, Inc, 1991. Citado na página 26.

MEDEIROS, K. R. et al. *O Sistema de Informação em Saúde como instrumento da política de recursos humanos: um mecanismo importante na detecção das necessidades da força de trabalho para o SUS*. 2004. <<http://www.scielo.br/pdf/csc/v10n2/a21v10n2.pdf>> Acesso em: 21 de Janeiro de 2023. Citado na página 16.

Microsoft. *Power BI site reliability engineering (SRE) model*. 2023. <<https://learn.microsoft.com/en-us/power-bi/enterprise/service-admin-site-reliability-engineering-model>>, Acesso em: 10 de Dezembro de 2023. Citado na página 56.

Ministério da Saúde. *Portaria nº 3088, de 23 de dezembro de 2011*. 2011. <https://bvsms.saude.gov.br/bvs/saudelegis/gm/2011/prt3088_23_12_2011_rep.html> Acesso em: 20 de Janeiro de 2023. Citado na página 13.

Ministério da Saúde. *PORTARIA Nº 3.588, DE 21 DE DEZEMBRO DE 2017*. 2017. <https://bvsms.saude.gov.br/bvs/saudelegis/gm/2017/prt3588_22_12_2017.html> Acesso em: 20 de Janeiro de 2023. Citado na página 13.

Ministério da Saúde. *Classificação dos Dados*. 2021. <<https://www.gov.br/esporte/pt-br/aceso-a-informacao/lgpdc/classificacao-dos-dados>>, Acesso em: 05 de Maio de 2023. Citado na página 25.

Ministério da Saúde. *CNES - Cadastro Nacional de Estabelecimentos de Saúde*. 2021. <<https://www.gov.br/saude/pt-br/aceso-a-informacao/acoes-e-programas/cebas/cnes-cadastro-nacional-de-estabelecimentos-de-saude>>, Acesso em: 16 de Janeiro de 2023. Citado na página 19.

- Ministério da Saúde. *Sistemas de informação em saúde*. 2021. <<https://www.gov.br/saude/pt-br/composicao/svs/vigilancia-de-doencas-cronicas-nao-transmissiveis/sistemas-de-informacao-em-saude>>, Acesso em: 12 de Janeiro de 2023. Citado na página 9.
- Ministério da Saúde. *DADOS DA REDE DE ATENÇÃO PSICOSSOCIAL (RAPS) NO SISTEMA ÚNICO DE SAÚDE (SUS)*. 2022. <<https://www.gov.br/saude/pt-br/aceso-a-informacao/acoes-e-programas/caps/raps/arquivos/dados-da-rede-de-atencao-psicossocial-raps.pdf/>> Acesso em: 20 de Janeiro de 2023. Citado 3 vezes nas páginas 3, 14 e 15.
- Ministério da Saúde. *Sobre o SISAB*. 2022. <<https://sisab.saude.gov.br/>> Acesso em: 20 de Janeiro de 2023. Citado na página 18.
- Ministério da Saúde SC. *REDE DE ATENÇÃO PSICOSSOCIAL – RAPS*. 2018. <<https://www.saude.sc.gov.br/index.php/resultado-busca/redes-de-atencao-a-saude-cidadao/10244-rede-de-atencao-psicossocial-raps>> Acesso em: 10 de Janeiro de 2023. Citado na página 9.
- MONTEIRO, H. *Modelagem de Processos com a Notação BPMN*. c2023. Disponível em <<https://www2.tce.pe.gov.br/eproc/images/treinamentos/ModelagemdeProcessoscomBPMN.pdf>> Acesso em: 02 de Fevereiro de 2023. Citado 2 vezes nas páginas 3 e 32.
- NEUMAM, C. *Entenda o que é CID, o código que identifica doenças no mundo todo*. 2021. <<https://www.cnnbrasil.com.br/saude/entenda-o-que-e-cid-o-codigo-que-identifica-doencas-no-mundo-todo>> Acesso em: 23 de Janeiro de 2023. Citado na página 17.
- OMG, T. O. M. G. *Business Process Model And Notation*. 2010. Disponível em <<https://www.omg.org/spec/BPMN/2.0>> Acesso em: 02 de Fevereiro de 2023. Citado na página 32.
- ORACLE. *O que é ETL?* c2023. <<https://www.oracle.com/br/integration/what-is-etl/>> Acesso em: 25 de Janeiro de 2023. Citado na página 21.
- PONNIAH, P. *Data Warehousing Fundamentals*. [S.l.]: John Wiley & Sons, 2002. Citado na página 33.
- Projeto Brumadinho UFMG. *Entenda o que é o Sistema de Informação em Saúde para a Atenção Básica (SISAB)*. 2021. <<http://projeto-brumadinho.ufmg.br/materia/sistema-de-informacao-sisab>>, Acesso em: 16 de Janeiro de 2023. Citado na página 19.
- RAMOS, T. B. C. Trabalho de Conclusão de Curso, *Desenvolvimento de ferramenta para extração de dados de citação e altmétricos via Application Programming Interface*. 2022. Citado na página 26.
- ROCHEL, C. J. K.; MEDINA, C. C.; MORENO, A. B. *Informação e Avaliação em Saúde*. 2007. <<https://www.arca.fiocruz.br/bitstream/handle/icict/39222/PoliticadeSaude-Informacaoeavaliacaoemsaude.pdf?sequence=2&isAllowed=y>> Acesso em: 21 de Janeiro de 2023. Citado na página 16.

Sala de Situação UnB. *Quem Somos*. c2023. Disponível em: <<https://sds.unb.br/quem-somos>> Acesso em: 15 de Janeiro de 2023. Citado na página 10.

SANTOS, A. C. dos. Sistema de informações hospitalares do sistema Único de saúde: documentação do sistema para auxiliar o uso das suas informações. 2009. Disponível em: <https://www.arca.fiocruz.br/bitstream/handle/icict/2372/ENSP_Dissertacao_Santos_Andreia_Cristina.pdf;jsessionid=64A7EABD2BF910953FEAD11DECF455A4?sequence=1>. Citado na página 16.

SANTOS, J. *Business Intelligence: conceito, história e vantagens*. 2022. Disponível em <<https://blog.cosmrobots.io/business-intelligence-conceito-historia-e-vantagens/>> Acesso em: 28 de Janeiro de 2023. Citado na página 24.

SAUDI. *O QUE É UMA AIH AUTORIZAÇÃO DE INTERNAÇÃO HOSPITALAR E SUAS PRINCIPAIS REGRAS*. c2023. Disponível em <<https://www.saudi.com.br/blog/o-que-e-uma-aih-autorizacao-de-internacao-hospitalar/>> Acesso em: 23 de Janeiro de 2023. Citado na página 17.

TERA. *Prototipagem de alta fidelidade*. 2020. Disponível em <<https://medium.com/somos-tera/prototipagem-de-alta-fidelidade-635d745b662b>> Acesso em: 31 de Janeiro de 2023. Citado na página 34.

VAISMAN, A.; ZIMÁNYI, E. *Data Warehouse Systems Design and Implementation*. [S.l.]: Springer, 2022. Citado 6 vezes nas páginas 3, 20, 21, 22, 23 e 32.

WOEBCKEN, C. *O que é brainstorming e as 9 melhores técnicas para a tomada de decisões inteligentes*. 2019. Disponível em <<https://rockcontent.com/br/blog/brainstorming/>> Acesso em: 01 de Fevereiro de 2023. Citado na página 29.

WYLTON. *Conceitos Básicos Sobre OLAP*. 2009. Disponível em <<https://www.devmedia.com.br/conceitos-basicos-sobre-olap/12523>> Acesso em: 25 de Janeiro de 2023. Citado na página 22.

Anexos

ANEXO A – Indicador de saúde 2 - IS02

A Figura mostra todas as informações do IS02, um indicador cujo os dados fonte pertencem aos SIS SCNES e obteve pontuação 108 na priorização.

IS02 - Distribuição (n° e %) de CAPS habilitados por tipo e ano	
Tipo	Estrutura
Interpretação	Estima o crescimento e expansão de CAPS habilitados por tipo anualmente: 1° Forma: Número absoluto do total de CAPS, por tipo, anualmente. Análise comparativa, exemplo: série histórica. 2° Forma: Valor positivo significa que houve crescimento, ou seja, novos CAPS foram habilitados; Valor negativo significa que houve decréscimo, ou seja, alguns CAPS desabilitados; Valor igual 0 significa que não houve crescimento, ou seja, o número de CAPS se manteve estável.
Fonte	Sistema do Cadastro Nacional de Estabelecimentos (SCNES) CNES-SR (Equipamentos especializados)? CNES-HB
Variáveis	TP_UNID_ID / CO_TIPO_UNIDADE - Tipo de estabelecimento COMPETÊN - Ano e Mês de competência da informação (AAAAMM) COD_MUN / CO_MUNICIPIO - Código do Município CO_HABILITAÇÃO - Código da Habilitação
Periodicidade	Anual
Método de cálculo	1° método: Soma anual do total de CAPS habilitados por tipo (para estado, região de saúde ou município) 2° método: Numerador: N° de CAPS no ano competência - N° de CAPS no ano anterior ao de competência Denominador: N° de CAPS no ano anterior ao de competência Fator de multiplicação: 100
Observação	Para o segundo método, realizar o cálculo para cada tipo de CAPS (I, II, III, ad e i).

Figura 21 – Indicador de Saúde 2 - Fonte: Equipe Coopere

ANEXO B – Indicador de saúde 3 - IS03

A Figura mostra todas as informações do IS03, um indicador cujo os dados fonte pertencem aos SIS SCNES e obteve pontuação 72 na priorização.

IS03 - Cobertura de CAPS/100 mil habitantes por ano e UF/município	
Tipo	Estrutura
Interpretação	Estima a cobertura dos CAPS por 100.000 habitantes, tanto para a UF, como para o município.
Fonte	Sistema do Cadastro Nacional de Estabelecimentos (SCNES) CNES e Instituto Brasileiro de Geografia e Estatística (IBGE).
Variáveis	SCNES: TP_UNID_ID / CO_TIPO_UNIDADE - Tipo de estabelecimento COMPETEN - Ano e Mês de competência da informação (AAAAMM) COD_MUN / CO_MUNICIPIO - Código do Município IBGE: Tabela de código UF/municípios Dados de população estimada para o ano
Periodicidade	Anual
Método de cálculo	Numerador: [(Nº CAPS I X 0,5) + (nº CAPS II) + (Nº CAPS III X 1,5) + (Nº de CAPS i) + (Nº CAPS AD) + (Nº de CAPS AD III X 1,5) para o ano em questão (para estado, região de saúde ou município). Denominador: População residente no mesmo local e ano. Fator de multiplicação: 100.000
Observação	Para esse indicador utiliza-se a cobertura ponderada de acordo com o tipo do CAPS, sendo: CAPS I 50 mil habitantes; os CAPS III e ad III, de 150 mil habitantes; CAPS (II, ad e i), cobertura de 100 mil habitantes, por isso o fator de multiplicação para cada um, no método acima. Parâmetros: Os seguintes parâmetros são adotados: cobertura muito boa (acima de 0,70), cobertura regular/boa (entre 0,50 e 0,69), cobertura regular/baixa (entre 0,35 e 0,49), cobertura baixa (entre 0,20 e 0,34) e cobertura insuficiente/crítica (abaixo de 0,20).

Figura 22 – Indicador de Saúde 3 - Fonte: Equipe Coopere

ANEXO C – Indicador de saúde 4 - IS04

A Figura mostra todas as informações do IS04, um indicador cujo os dados fonte pertencem aos SIS SCNES e obteve pontuação 18 na priorização.

IS04 - Distribuição dos CAPS por Município Elegível (ME) e nº de municípios com CAPS Habilitado por UF 	
Tipo	Estrutura
Interpretação	Estima a quantidade de CAPS habilitados conforme o quantitativo de municípios elegíveis pertencentes ao mesmo Estado, anualmente. Análise somente para os Estados, representa o percentual de CAPS habilitados em comparativo com o total de CAPS que poderiam estar habilitados conforme a quantia de ME para o Estado.
Fonte	Sistema do Cadastro Nacional de Estabelecimentos (SCNES) e CNES e Instituto Brasileiro de Geografia e Estatística (IBGE)
Variáveis	TP_UNID_ID / CO_TIPO_UNIDADE - Tipo de estabelecimento COMPETÊN - Ano e Mês de competência da informação (AAAAMM) COD_MUN / CO_MUNICIPIO - Código do Município IBGE: Tabela de código UF/municípios Dados de população estimada para o ano
Periodicidade	Anual
Método de cálculo	Numerador: N° CAPS habilitados em municípios elegíveis. Denominador: N° de municípios elegíveis de acordo com os critérios de elegibilidade para implantação de CAPS. Fator de multiplicação: 100
Observação	O município é considerado elegível desde que tenha uma população mínima de 15 mil habitantes. Utilizar a estimativa populacional do IBGE, para extrair os municípios elegíveis para cada estado.

Figura 23 – Indicador de Saúde 4 - Fonte: Equipe Coopere

ANEXO D – Indicador de saúde 7 - IS07

A Figura mostra todas as informações do IS07, um indicador cujo os dados fonte pertencem aos SIS SIH e obteve pontuação 48 na priorização.

IS07 - Percentual de reinternações por causas de saúde mental, conforme ano/mês e município/UF	
Tipo	Resultado
Interpretação	Estimar o percentual de reinternação de usuários por causas de saúde mental, conforme o CID, em hospital Geral e Referência, para o mês/ano e município/UF
Fonte	Sistema de Informações Hospitalares - SIH
Variáveis	
Periodicidade	Anual
Método de cálculo	Numerador: quantitativo de reinternações por causa de saúde mental, conforme CID, em hospital Geral e de Referência, para o mesmo período e local Denominador: total de internações por saúde mental, conforme CID, em Hospital Geral e de Referência, para o mesmo período e local. Fator de multiplicação: 100
Observação	Esse é indicador a ser proposto pelo Ministério da Saúde, pois lida com dados sensíveis e para seu cálculo preciso, seria necessário dado do cartão do SUS. <input type="checkbox"/>

Figura 24 – Indicador de Saúde 7 - Fonte: Equipe Coopere

ANEXO E – Indicador de saúde 8 - IS08

A Figura mostra todas as informações do IS08, um indicador cujo os dados fonte pertencem aos SIS SIA e obteve pontuação 72 na priorização.

IS08 - Percentual dos tipos de atendimentos e ações realizadas nos CAPS, por ano e UF/município	
Tipo	Resultado
Interpretação	Estimar o percentual de atendimentos realizados pelo CAPS, conforme o tipo de atendimento (tabela SIGTAP), anualmente, de acordo com a UF e município.
Fonte	Sistema de Informações Ambulatoriais (SIA) - SIA-PS Tabela SIGTAP
Variáveis	TPUPS - Tipo de do Estabelecimento DT_ATEND - Data do Atendimento (AAAAMM) PA_PROC_ID - Código de Procedimento Ambulatorial (código SIGTAP)
Periodicidade	Anual
Método de cálculo	Numerador: total de atendimento realizado no CAPS, conforme o tipo (tabela SIGTAP), para determinado período e local Denominador: total de atendimentos realizados no CAPS, para o mesmo período e local. Fator de multiplicação: 100
Observação	

Figura 25 – Indicador de Saúde 8 - Fonte: Equipe Coopere

ANEXO F – Indicador de saúde 9 - IS09

A Figura mostra todas as informações do IS09, um indicador cujo os dados fonte pertencem aos SIS SCNES e obteve pontuação 72 na priorização.

IS09 - Composição das Equipes Especializadas de Saúde Mental por ano e UF/município	
Tipo	Estrutura
Interpretação	Analisar a composição das equipes especializadas de saúde mental, anualmente, conforme município/UF. Ver a disponibilidade dos profissionais.
Fonte	Sistema do Cadastro Nacional de Estabelecimentos (SCNES)
Variáveis	CNES - Código do estabelecimento CO_MUNICIPIO - Código do Município TP_EQUIPE - Tipo de Equipe COD_CBO - Código Brasileiro de Ocupação
Periodicidade	Anual
Método de cálculo	Quantificar o número de profissionais que compõem as equipes especializadas em saúde mental, para determinado ano e local (município/UF).
Observação	<p>Categoria de agrupamento dos profissionais, sendo 4:</p> <ol style="list-style-type: none"> 1. Médico 2. Enfermeiro 3. Profissionais de nível superior incluem: psicólogo, assistente social, enfermeiro, terapeuta ocupacional, fonoaudiólogo, pedagogo ou outro profissional necessário ao projeto terapêutico; 4. Profissionais de nível médio incluem: técnico e/ou auxiliar de enfermagem, técnico administrativo, técnico educacional e artesão.

Figura 26 – Indicador de Saúde 9 - Fonte: Equipe Coopere

ANEXO G – Indicador de saúde 10 - IS10

A Figura mostra todas as informações do IS010 um indicador cujo os dados fonte pertencem aos SIS SIA e obteve pontuação 72 na priorização.

IS10 - Distribuição (nº e %) de atendimentos por causa/diagnóstico mês/ano e município/UF	
Tipo	Processo
Interpretação	Estimar o número e percentual dos atendimentos realizados no CAPS, de acordo com a causa/diagnóstico (CID-10), mensalmente e anualmente, para o município/UF.
Fonte	Sistema de Informações Ambulatoriais (SIA)
Variáveis	TPUPS - Tipo de do Estabelecimento UFMUN - Unidade da Federação + município onde está localizado o Estabelecimento DT_ATEND - Data do Atendimento (AAAAMM) CIDPRI - CID Principal (código CID-10) Códigos CID-10: http://www2.datasus.gov.br/cid10/V2008/cid10.htm Categorias CID-10 saúde mental: F00, F01, F02, F03, F04, F05, F06, F07, F09, F10, F11, F12, F13, F14, F15, F16, F17, F18, F19, F20, F21, F22, F23, F24, F25, F28, F29, F30, F31, F32, F33, F34, F38, F39, F40, F41, F42, F43, F44, F45, F48, F50, F51, F52, F53, F54, F55, F59, F60, F61, F62, F63, F64, F65, F66, F68, F69, F70, F71, F72, F73, F78, F79, F80, F81, F84, F88, F89, F90, F91, F92, F93, F94, F95, F98, F99, X60, X61, X62, X63, X64, X65, X66, X67, X68, X69, X70, X71, X72, X73, X74, X75, X76, X77, X78, X79, X80, X81, X82, X83, X84
Periodicidade	Anual
Método de cálculo	1º método: Apenas o total dos atendimentos individuais realizados nos CAPS, de acordo com a causa e diagnóstico (agrupamento dos capítulos CID-10), mensalmente e anualmente, para determinado local (município/UF). 2º método: Numerador: total dos atendimentos individuais realizados nos CAPS, de acordo com a causa e diagnóstico (agrupamento dos capítulos CID-10), mensalmente e anualmente, para determinado local (município/UF). Denominador: total de atendimentos individuais realizados com causa/diagnóstico entre os capítulos F00-X84, para o mesmo período e local.
Observação	Agrupar em categorias conforme o CID, sendo: transtornos mentais e comportamentais - CID F00-F99 e lesões autoprovocadas - CID X60-X84.

Figura 27 – Indicador de Saúde 10 - Fonte: Equipe Coopere

ANEXO H – Validação dos indicadores via Jitsi

O Anexo representa uma das várias reuniões de equipe que ocorreram no decorrer do projeto.

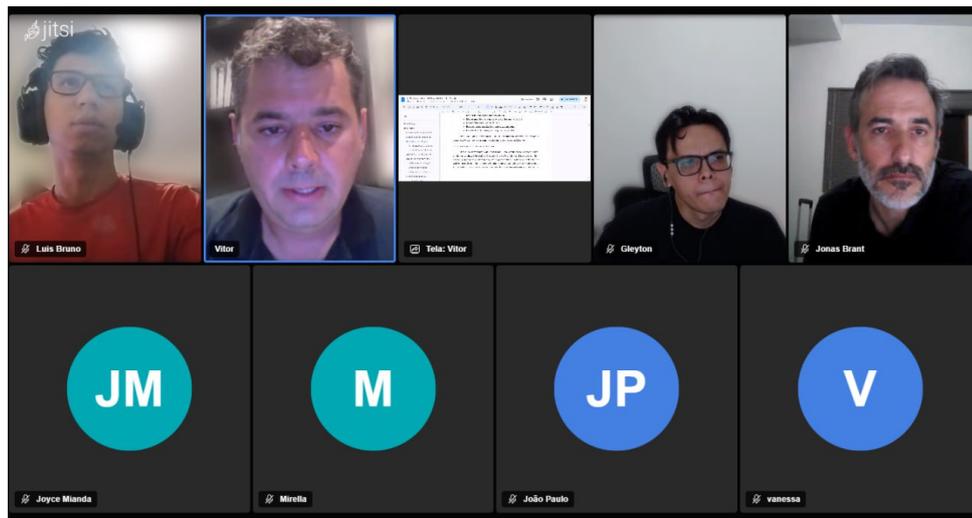


Figura 28 – Validação dos indicadores - Fonte: Autor.

ANEXO I – Código de conversão de .DBC para .CSV em R

Este código foi utilizado para a conversão de arquivos no formato .dbc (formato disponibilizado pelo DataSUS) para .csv (formato para manipulação dos dados, armazenamento e leitura pelo Power BI), recebendo dois argumentos: o arquivo de entrada em .dbc e o arquivo de saída em .csv. O código utiliza a biblioteca "read.dbc" para a leitura do arquivo fonte, converte para um *dataframe* e, por fim, converte para .csv e salva.

```
1 library(read.dbc)
2
3 # Funcao para imprimir mensagens de log no console
4 log_message <- function(message) {
5   timestamp <- format(Sys.time(), "%Y-%m-%d %H:%M:%S")
6   cat("[DEBUG]", timestamp, message, "\n")
7 }
8
9 # Obtem os argumentos da linha de comando
10 args <- commandArgs(trailingOnly = TRUE)
11
12 # Verifica se foram fornecidos argumentos suficientes
13 if (length(args) < 2) {
14   stop("Forneca pelo menos dois argumentos: arquivo de entrada
15   e arquivo de saida")
16 }
17
18 # Primeiro argumento: local do arquivo de entrada
19 input_path <- args[1]
20
21 # Segundo argumento: arquivo de saida (csv)
22 output_path <- args[2]
23
24 # Define o caminho completo para a pasta em que os arquivos .dbc estao
25 # armazenados
26 caminho <- file.path(input_path)
27
28 log_message(paste("Arquivo de entrada: ", input_path))
29
30 # lista todos os arquivos .dbc na pasta
31 arquivos <- list.files(path = caminho, pattern = "\\\\.dbc$",
32 full.names = TRUE)
33
34 log_message(paste("arquivos: ", arquivos))
```

```
35
36 # Loop através de todos os arquivos na lista
37 for (i in 1:length(arquivos)) {
38
39   # le o arquivo .dbc
40   dados <- read.dbc(arquivos[i])
41
42   # converte os dados em um dataframe
43   df <- as.data.frame(dados)
44
45   # define o nome do arquivo CSV
46   nome_csv <- basename(gsub("\\.dbc$", ".csv", arquivos[i]))
47
48   log_message(paste("nome_csv: ", nome_csv))
49
50   # Verifica se a pasta de saída existe, e se não, cria-a
51   if (!file.exists(output_path)) {
52     dir.create(output_path, recursive = TRUE)
53     log_message(paste("Pasta de saída", output_path, "criada."))
54   }
55
56   # Define o caminho completo para a pasta "bronze"
57   caminho_bronze <- file.path(output_path, nome_csv)
58
59   # grava o dataframe em um arquivo CSV
60   write.csv2(df, file = caminho_bronze, row.names = FALSE,
61             fileEncoding = "utf-8")
62 }
```

ANEXO J – Fluxo de trabalho no Github Actions

O código abaixo se trata do código completo da configuração de um fluxo de trabalho no Github Actions, onde o mesmo já foi explicado em detalhes na Seção 4.2.3.

```

1 name: Gerar arquivo SIH-RD
2
3 on:
4   workflow_dispatch:
5     inputs:
6       ano_mes:
7         description: "Ano e ms no formato YYYY-MM.
8         Exemplo 2023-01"
9         required: false
10      uf:
11        description: "UF para o qual o arquivo vai ser gerado.
12        Exemplo AM, DF. Obrigatrio se ano_mes informado."
13        required: false
14      schedule:
15        - cron: "0 0 28 * *"
16 env:
17   NOME_BASE_ARQUIVO_DEFAULT: "SIH-RD.parquet"
18   UF_DEFAULT: "DF"
19
20 jobs:
21   extrair-transformar-enviar:
22     runs-on: ubuntu-latest
23     steps:
24       - name: Checkout do repositorio
25         uses: actions/checkout@v2
26
27       - name: Configurar ambiente R
28         uses: r-lib/actions/setup-r@v2
29         with:
30           use-public-rspm: true
31
32       - name: Configurando dependencias
33         uses: r-lib/actions/setup-renv@v2
34         with:
35           cache-version: 2
36
37       - name: Definir valores computados para ano_mes e uf - execucao manual
38         if: github.event.inputs.ano_mes != ''
39         run: |

```

```

39     echo "ANO_MES_COMPUTADO=${{github.event.inputs.ano_mes}}"
40     >> $GITHUB_ENV
41     echo "UF_COMPUTADO=${{github.event.inputs.uf}}" >> $GITHUB_ENV
42
43 - name: Definir valores computados para ano_mes e uf - argumentos p o script
44   if: github.event.inputs.ano_mes == ''
45   run: |
46     echo "ANO_MES_COMPUTADO=$(date -d "2 months ago" +%Y-%m)"
47     >> $GITHUB_ENV
48     echo "UF_COMPUTADO=${{ env.UF_DEFAULT }}" >> $GITHUB_ENV
49
50 - name: Definir nome final do arquivo
51   run: |
52     echo "NOME_ARQUIVO_COMPUTADO=${{ env.ANO_MES_COMPUTADO }}-
53     ${{ env.UF_COMPUTADO }}-${{ env.NOME_BASE_ARQUIVO_DEFAULT }}"
54     >> $GITHUB_ENV
55
56 - name: Executar script R
57   run: Rscript extrair-transformar-enviar-sih.R
58     ${{ env.ANO_MES_COMPUTADO }} ${{ env.UF_COMPUTADO }}
59
60 - name: Salvar arquivo como artefato
61   uses: actions/upload-artifact@v2
62   with:
63     name: ${{ env.NOME_ARQUIVO_COMPUTADO }}
64     path: ${{ env.NOME_ARQUIVO_COMPUTADO }}
65
66 - name: Azure Login
67   uses: azure/login@v1
68   with:
69     creds: ${{ secrets.AZURE_CREDENTIALS }}
70
71 - name: Azure CLI - Envio
72   uses: azure/CLI@v1
73   with:
74     azcliversion: 2.30.0
75     inlineScript: |
76       chmod +x $GITHUB_WORKSPACE/upload_file.sh
77       $GITHUB_WORKSPACE/upload_file.sh \
78       ${{ secrets.ACCOUNT_NAME }} \
79       ${{ secrets.ACCOUNT_KEY }} \
80       ${{ secrets.CONTAINER_NAME }} \
81       ${{ env.NOME_ARQUIVO_COMPUTADO }} \
82       sia-pa/${{ env.NOME_ARQUIVO_COMPUTADO }}

```

Listing J.1 – Fluxo de trabalho em .yaml

ANEXO K – Transformação de dados do SIH em Python

O código abaixo se trata da transformação de dados do Sistema de Informações Hospitalares, que tem seu .csv bruto (anteriormente convertido de .dbc para .csv) lido da camada bronze, passa por transformações detalhadas na Seção 4.2.3 e gera um arquivo de camada silver, para cada tipo de dado necessário.

```

1 import pandas as pd
2
3 ANOS = ["18", "19", "20", "21", "22"]
4 MESES = ['01', '02', '03', '04', '05', '06', '07', '08', '09', '10', '11', '12']
5
6 def gerar_arquivo_categorias():
7     df_saida = pd.DataFrame()
8     for ano in ANOS:
9         for mes in MESES:
10            df_sih = pd.read_csv(f"bronze/RDDF{ano}{mes}.csv", dtype={"DT_INTER": str,
11                               "UF_ZI": str})
12            df_sih["DIAG_PRINC"] = df_sih["DIAG_PRINC"].str.slice(0, 3)
13            df_sih = df_sih[
14                (df_sih["DIAG_PRINC"].str.startswith("F")) | (
15                    (df_sih['DIAG_PRINC'] >= 'X60')
16                    & (df_sih['DIAG_PRINC'] < 'X85')))]
17
18            df_sih["ESTADO_CODIGO"] = df_sih["UF_ZI"].str.slice(0, 2)
19            df_sih = df_sih[["ESTADO_CODIGO", "DIAG_PRINC", "ANO_CMPT", "MES_CMPT"]]
20            df_saida = pd.concat([df_saida, df_sih], axis=0)
21
22            totalizador = df_saida.groupby(["ESTADO_CODIGO", "DIAG_PRINC", "ANO_CMPT",
23                                           "MES_CMPT"]).size().reset_index(name="TOTAL")
24            totalizador.to_csv("silver/2018-2022-diag_princ-sih-df.csv", sep=";", index=False)
25
26 def agrupar_por_categoria():
27     df_sih = pd.read_csv("silver/2018-2022-diag_princ-sih-df.csv", sep=";")
28     df_sih["CATEGORIA_GERAL"] = df_sih["DIAG_PRINC"].apply(
29         lambda x: "lesoes autoprovocadas" if x.startswith(
30             "X") else "transtornos mentais e comportamentais" if x.startswith("F") else
31             "outro")
32     totalizador = df_sih.groupby(["ESTADO_CODIGO", "CATEGORIA_GERAL", "ANO_CMPT",
33                                 "MES_CMPT"])["TOTAL"].sum().reset_index(name="TOTAL")
34     totalizador.to_csv("silver/2018-2022-categoria_geral-sih-df.csv", sep=";",
35                        index=False)

```

```
36
37
38 def calcular_percentual():
39     df = pd.read_csv("silver/2018-2022-diag_princ-sih-df.csv", delimiter=";")
40
41     df_by_year = df.groupby("ANO_CMPT")["TOTAL"].sum()
42
43     df["PERCENTUAL"] = df.apply(lambda x: (x["TOTAL"] / df_by_year[x["ANO_CMPT"]])
44     * 100, axis=1)
45
46     df.to_csv("silver/2018-2022-percentual-sih-df.csv",
47             sep=";",
48             decimal=",")
```

ANEXO L – Script *upload_file.sh*

Este *script* é utilizado para o envio de arquivos para o Data Lake, que é executado por uma ação do Github Actions, mencionado na Subseção 4.2.3. Maiores detalhes acerca da conexão com o Data Lake da Azure são encontradas na Subseção 4.3.2.

```
1 #!/bin/sh
2
3 az storage blob upload \
4   --account-name $1 \
5   --account-key $2 \
6   --container-name $3 \
7   --file $4 \
8   --name $5
```


ANEXO N – Tela 2

A Figura mostra a tela "Álcool e Outras Drogas - Óbitos" em desenvolvimento no Power BI.

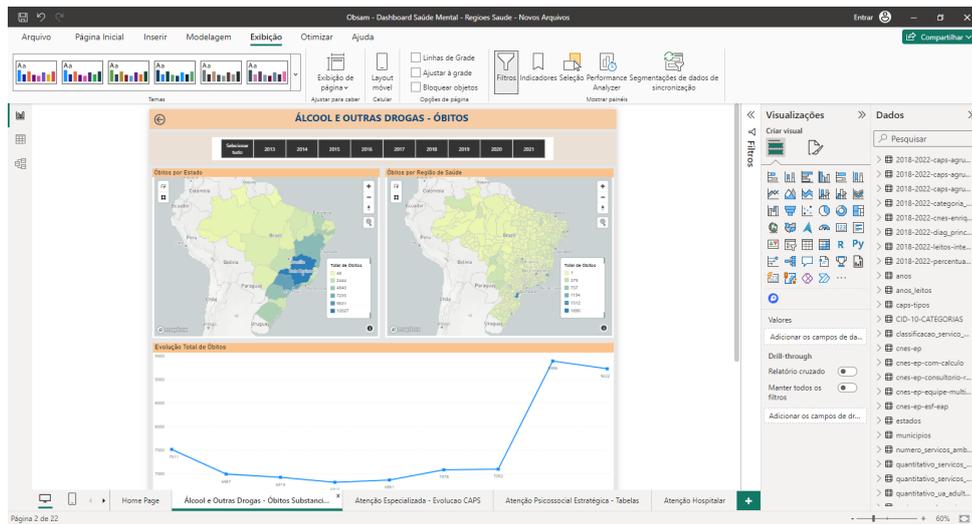


Figura 30 – Tela 2 - Fonte: Autor.

ANEXO O – Tela 4

A Figura mostra a tela "Atenção Especializada - CAPS - Cobertura" em desenvolvimento no Power BI.

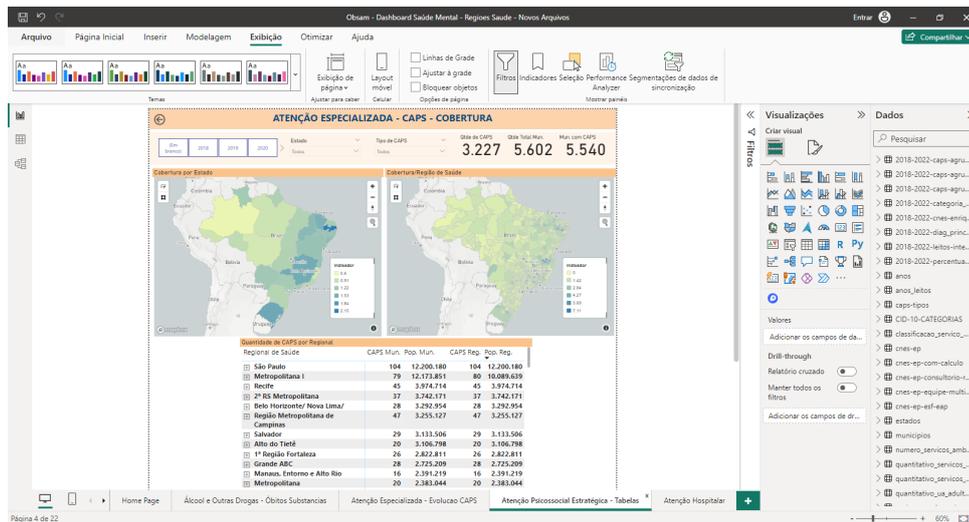


Figura 31 – Tela 4 - Fonte: Autor.

ANEXO Q – Tela 6

A Figura mostra a tela "Atenção Primária - Equipes NASF e ESF" em desenvolvimento no Power BI.

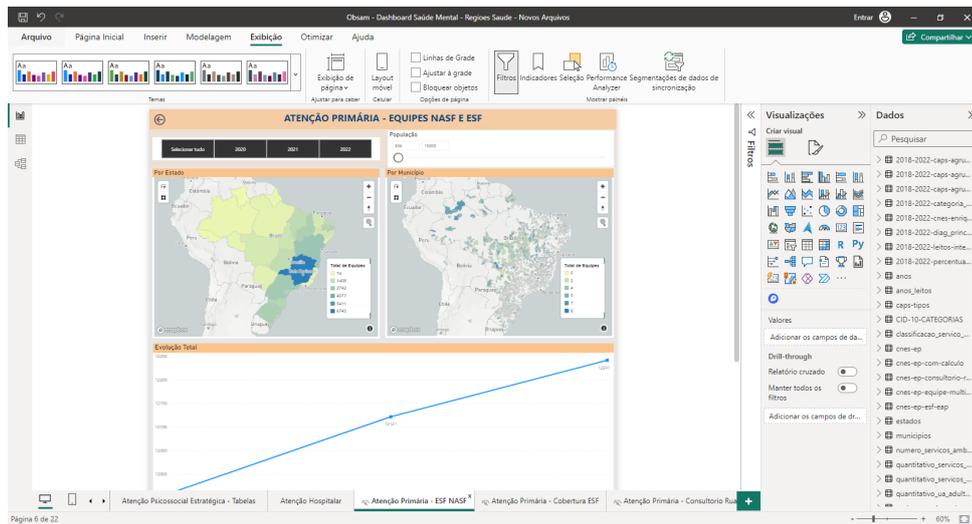


Figura 33 – Tela 6 - Fonte: Autor.

ANEXO S – Tela 8

A Figura mostra a tela "Atenção Primária - Consultórios de Rua" em desenvolvimento no Power BI.

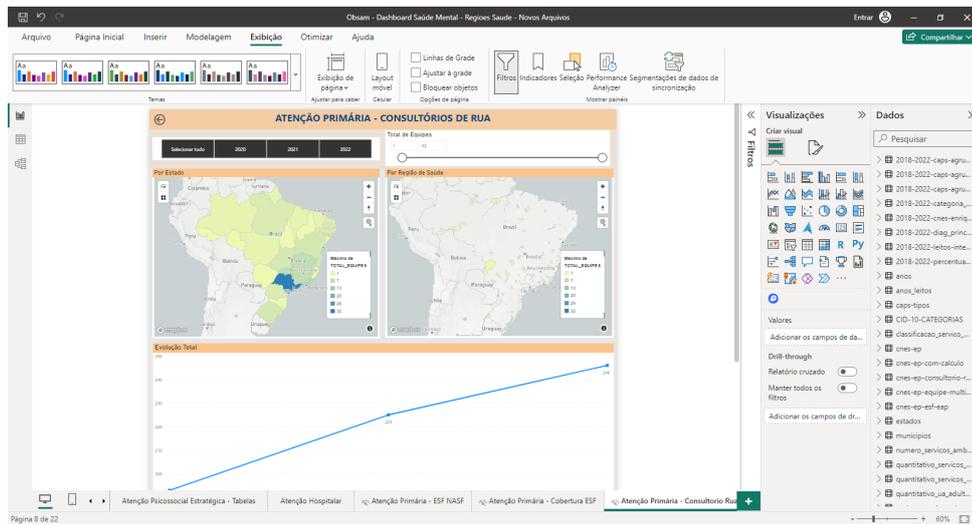


Figura 35 – Tela 8 - Fonte: Autor.

ANEXO T – Tela 9

A Figura mostra a tela "Atenção Especializada" (um pequeno menu com apresentação) em desenvolvimento no Power BI.

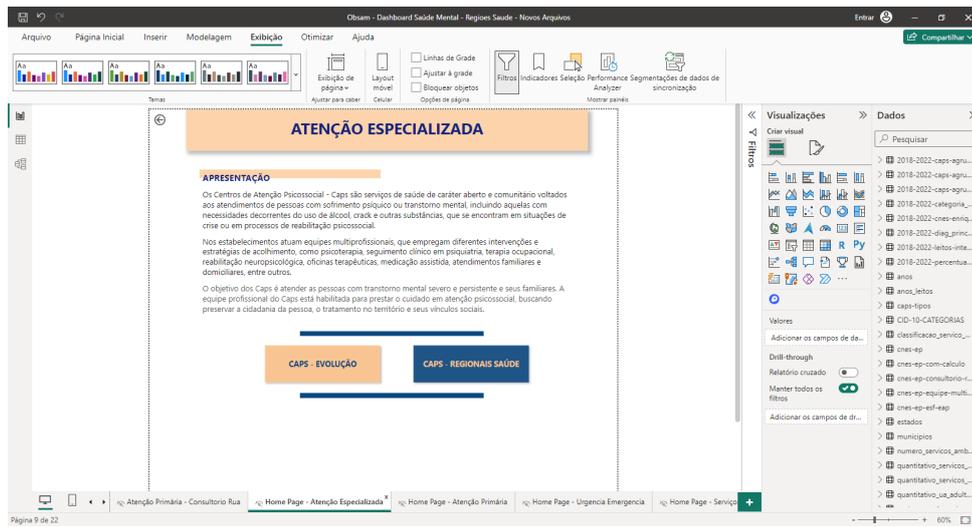


Figura 36 – Tela 9 - Fonte: Autor.

ANEXO U – Tela 10

A Figura mostra a tela "Atenção Primária" (um pequeno menu com apresentação) em desenvolvimento no Power BI.

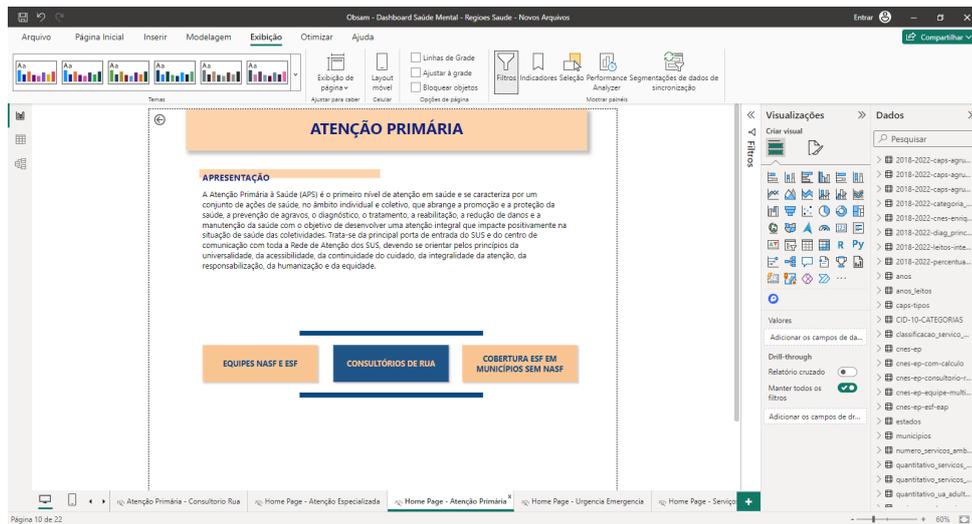


Figura 37 – Tela 10 - Fonte: Autor.

ANEXO V – Tela 11

A Figura mostra a tela "Urgência Emergência" (um pequeno menu com apresentação) em desenvolvimento no Power BI.

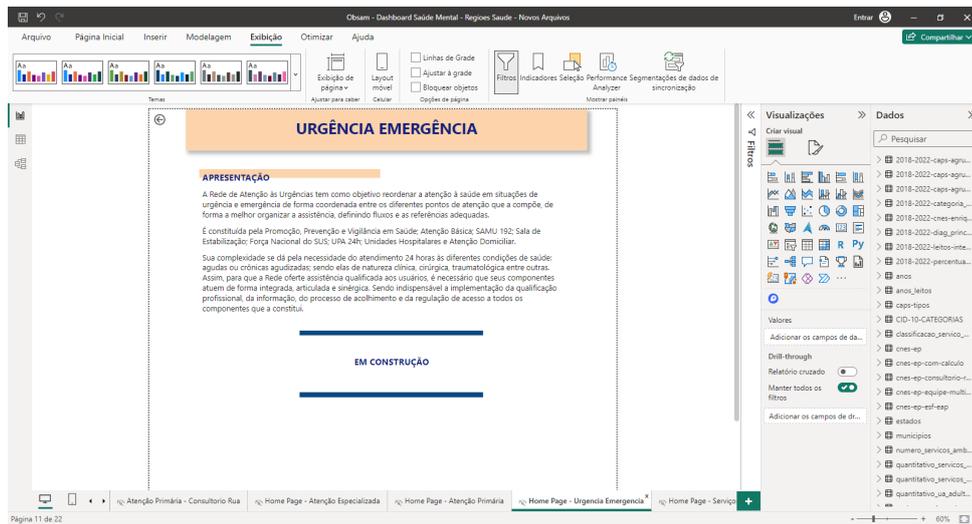


Figura 38 – Tela 11 - Fonte: Autor.

ANEXO W – Tela 12

A Figura mostra a tela "Serviços Residenciais" (um pequeno menu com apresentação) em desenvolvimento no Power BI.

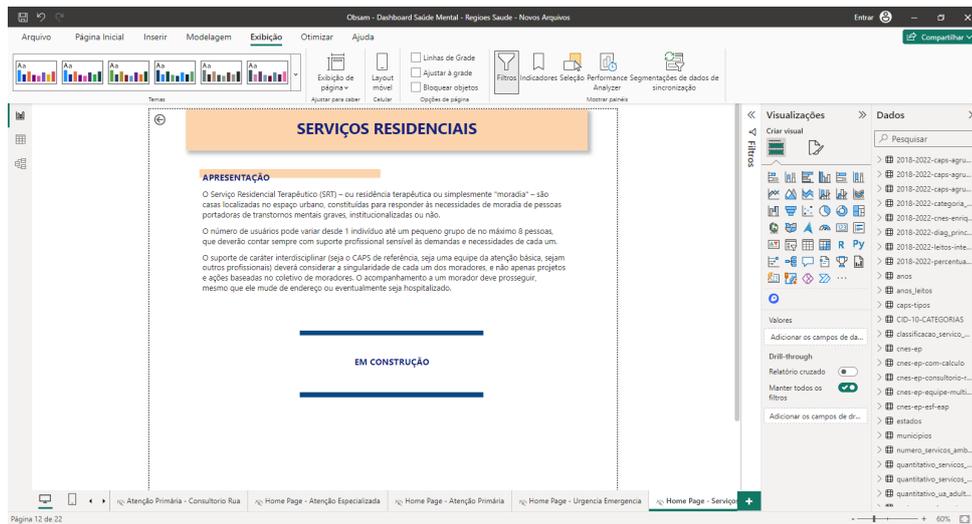


Figura 39 – Tela 12 - Fonte: Autor.

ANEXO X – Tela 13

A Figura mostra a tela "Estratégia de Desinstitucionalização" (um pequeno menu com apresentação) em desenvolvimento no Power BI.

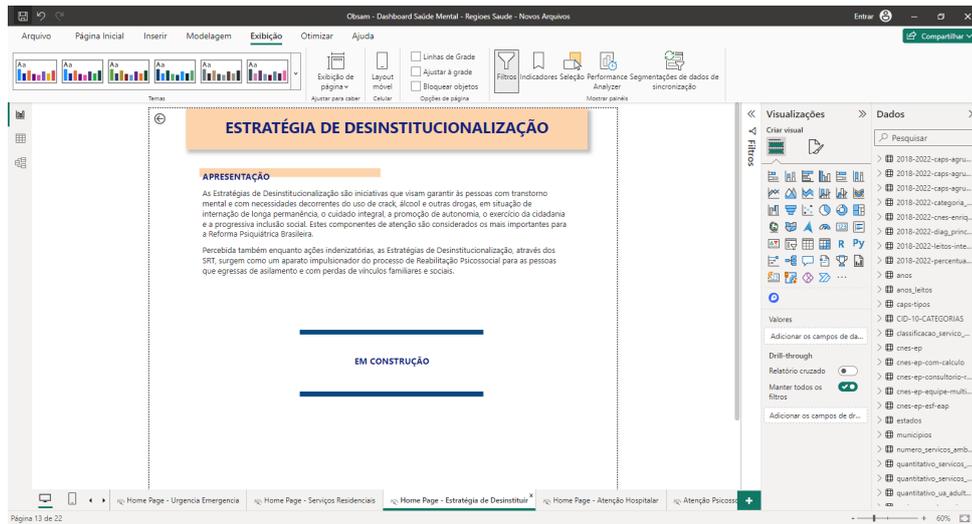


Figura 40 – Tela 13 - Fonte: Autor.

ANEXO Z – Tela 15

A Figura mostra a tela "Atenção Psicossocial Estratégica - Cobertura CAPS" em desenvolvimento no Power BI.

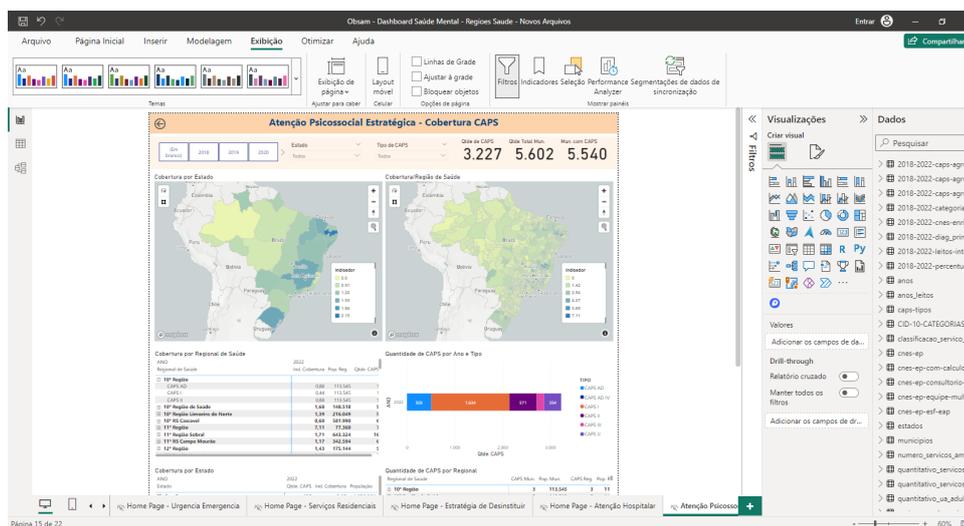


Figura 42 – Tela 15 - Fonte: Autor.