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

BERNARDO VIEIRA GUIMARÃES SANTOS

DESENVOLVIMENTO DE UMA ARQUITETURA DE BI PARA ANÁLISE DE CASOS DE ABANDONO DE TRATAMENTO DA TUBERCULOSE NOS MUNICÍPIOS BRASILEIROS

ALEGRE - ES
2026

DESENVOLVIMENTO DE UMA ARQUITETURA DE BI PARA ANÁLISE DE CASOS DE ABANDONO DE TRATAMENTO DA TUBERCULOSE NOS MUNICÍPIOS BRASILEIROS

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

Por

BERNARDO VIEIRA GUIMARÃES SANTOS

Orientador

Prof. Dr. Antonio Almeida de Barros Junior.

Universidade Federal do Espírito Santo

ALEGRE - ES

2026

Bernardo Vieira Guimarães Santos

Desenvolvimento de uma arquitetura de BI para análise de casos de abandono de tratamento da Tuberculose nos municípios brasileiros

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

Aprovado em 10 de fevereiro de 2026.

COMISSÃO EXAMINADORA

gov.br Documento assinado digitalmente
ANTONIO ALMEIDA DE BARROS JUNIOR
Data: 25/02/2026 19:04:58-0300
Verifique em https://validar.iti.gov.br

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

gov.br Documento assinado digitalmente
DAYAN DE CASTRO BISSOLI
Data: 11/02/2026 13:43:49-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Dayan de Castro Bissoli
Universidade Federal do Espírito Santo - UFES

gov.br Documento assinado digitalmente
TIAGO RICARDO MOREIRA
Data: 11/02/2026 20:40:16-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Tiago Ricardo Moreira
Universidade Federal de Viçosa - UFV

AGRADECIMENTOS

Primeiramente, agradeço a Deus por ter me concedido saúde, força e perseverança ao longo de toda essa jornada acadêmica, permitindo que eu superasse os desafios e alcançasse mais essa conquista.

Aos meus pais e ao meu irmão, por todo o suporte, compreensão e apoio ao longo da minha trajetória acadêmica. Sou imensamente grato por acreditarem em mim, pelos ensinamentos transmitidos e por estarem sempre presentes, mesmo nos momentos mais desafiadores.

À minha namorada, Larissa Barreto Nunes, por todo o amor, companheirismo, paciência e incentivo incondicional. Seu apoio foi fundamental nos momentos mais difíceis, me dando forças para continuar e não desistir dos meus objetivos. Amo você.

Ao meu orientador, Prof. Dr. Antônio Almeida de Barros Junior, pela disponibilidade, oportunidade, paciência, contribuições técnicas e acadêmicas, além das valiosas orientações que foram essenciais para o desenvolvimento deste Trabalho de Conclusão de Curso.

Aos amigos que fiz durante toda a jornada acadêmica, em especial Heflain Rocha, Arthur Wingler, Pedro Rangel, Matheus Melo e Marianne Alves, pelo companheirismo, apoio, momentos de aprendizado e pelas experiências compartilhadas que tornaram essa caminhada mais leve e significativa.

Por fim, agradeço a todos os professores, colegas e demais pessoas que, de alguma forma, contribuíram para a minha formação acadêmica e para a realização deste trabalho.

RESUMO

A baixa adesão e o abandono do tratamento da Tuberculose (TB) podem resultar na manutenção da cadeia de transmissão da doença, aumento da mortalidade, e desenvolvimento de resistência bacteriana, os quais impõem dificuldades ao processo de cura, aumentando o tempo e o custo do tratamento. O controle efetivo da Tuberculose envolve diversas estratégias, incluindo a prevenção, o diagnóstico precoce e o tratamento adequado. A análise dos fatores que levam ao abandono do tratamento da doença e suas consequências deve considerar a integração de níveis hierárquicos, incluindo características individuais, estruturas de cuidados em saúde e determinantes sociais do contexto de vida dos indivíduos. Desconsiderar essa hierarquia pode acarretar em problemas com as estimativas dos efeitos do modelo e suas variâncias, comprometendo a qualidade dos resultados obtidos. Com o intuito de auxiliar os gestores em saúde, o presente trabalho teve como objetivo propor e implementar a adoção de tecnologias de *Business Intelligence* (BI), por meio da construção de um *Data Warehouse* (DW), capaz de permitir a análise dos casos de tuberculose de forma precisa, ágil e interativa. Para isso, foram realizadas as seguintes etapas metodológicas: acesso às bases de dados referentes à doença, tratamento e integração das informações por meio de rotinas de Extração, Transformação e Carga de Dados (ETL), construção de um banco de dados dimensional e desenvolvimento de *dashboards* analíticos. Como resultados, foram consolidados aproximadamente 301.960 registros de notificações de tuberculose entre os anos de 2018 e 2022, possibilitando a construção de indicadores analíticos capazes de evidenciar padrões temporais, territoriais e perfis populacionais associados ao abandono do tratamento. Observou-se uma taxa média de abandono próxima a 14%, com tendência de crescimento ao longo dos anos analisados, além de variações regionais significativas entre unidades federativas. Os *dashboards* desenvolvidos permitiram a exploração interativa desses indicadores por meio de filtros analíticos, contribuindo para a identificação de grupos e contextos mais vulneráveis à não adesão terapêutica.

Palavras-chave: Tuberculose; Abandono do tratamento; *Business Intelligence*; *Data Warehouse*; Análise epidemiológica.

Lista de figuras

Figura 1 – Arquitetura geral da solução de *Business Intelligence.*

Figura 2 – Arquitetura de *Business Intelligence* proposta.

Figura 3 – Modelo Dimensional proposto.

Figura 4 – Panorama geral dos casos e do abandono do tratamento.

Figura 5 – Distribuição geográfica do abandono do tratamento.

Figura 6 – Perfil sociodemográfico dos pacientes.

Figura 7 – Fatores clínicos e sociais associados ao abandono.

Figura 8 – Condições socioeconômicas associadas ao abandono.

SUMÁRIO

Sumário

1. INTRODUÇÃO 8

2. OBJETIVOS 11

2.1. OBJETIVO GERAL 11

2.2. OBJETIVOS ESPECÍFICOS 11

3. REVISÃO DA LITERATURA 12

3.1. TUBERCULOSE 12

3.2. BUSINESS INTELLIGENCE 13

3.2.1. Data Warehouse 14

3.3. TRABALHOS RELACIONADOS 15

4. METODOLOGIA 22

4.1. FASE 1 – FONTE DE DADOS 23

4.2. FASE 2 - INTEGRAÇÃO 25

4.3. FASE 3 - REPOSITÓRIO 30

4.4. FASE 4 - VISUALIZAÇÃO 32

4.5. FERRAMENTAS UTILIZADAS 32

5. RESULTADOS OBTIDOS 35

5.1. CONSOLIDAÇÃO DOS DADOS E INDICADORES ANALÍTICOS 35

5.2. ANÁLISE DOS INDICADORES 37

6. CONCLUSÃO 45

7. REFERÊNCIAS 47

1. INTRODUÇÃO

A Tuberculose (TB) é uma das doenças mais antigas da humanidade, causada por bactérias pertencentes ao complexo *Mycobacterium tuberculosis*. Mesmo tendo grande prevalência na Antiguidade, sua grande expansão e disseminação ocorreu durante o período colonial, eclodindo como a grande epidemia do século XVIII. Neste período, foi responsável pela infecção de quase 100% da população européia, sendo responsável por 25% das mortes em adultos. A disseminação da doença pelas Américas decorreu do contato com navegantes europeus, na época das grandes navegações. Entretanto, somente mais tarde, no século XX, com a urbanização experimentada pelos países americanos, é que realmente houve a explosão epidêmica da doença em todo o continente, inclusive no Brasil (SILVA, 2019).

Apesar de ser curável e prevenível, a tuberculose continua a representar um grave desafio de saúde pública global. Segundo o Relatório Global sobre Tuberculose 2025, publicado pela Organização Mundial da Saúde (OMS), foram estimados 10,7 milhões de casos de TB no mundo em 2024, com aproximadamente 1,23 milhão de mortes atribuídas à doença nesse mesmo período, mantendo a TB entre os principais motivos de óbito por agente infeccioso no cenário internacional (WHO, 2025). No Brasil, de acordo com o Boletim Epidemiológico Tuberculose 2025, foram registrados 85.936 casos novos de tuberculose em 2024, e cerca de 6.025 óbitos por TB em 2023, o que evidencia a magnitude contínua da doença no país (BRASIL, 2025). Esses dados recentes demonstram que, apesar de décadas de esforços em vigilância e controle, a tuberculose persiste com elevada carga epidemiológica, exigindo aprimoramento constante das estratégias de monitoramento e intervenção.

Desde 2017, o Brasil possui um plano estratégico de TB alinhado à agenda global da doença. O “Plano Nacional pelo Fim da Tuberculose como Problema de Saúde Pública” (BRASIL, 2017; BRASIL, 2021) apresenta quatro fases de execução e estabelece metas adaptadas ao cenário nacional para a redução da incidência, do número de mortes e do custo catastrófico por TB, a serem alcançadas até 2035. Vale ressaltar uma das metas de saúde dos Objetivos de Desenvolvimento Sustentável (ODS) é o fim da epidemia de TB até o ano de 2030. Isso por que o ODS número 3 (assegurar uma vida saudável e promover o bem-estar para todos) e a respectiva meta 3.3 (acabar com as epidemias de Aids, TB, Malária e doenças tropicais

negligenciadas até 2030) afetam e são afetados pelas demais questões expressas nessa agenda. No caso da TB, em especial, destaca-se a inter-relação com a pauta da redução das desigualdades e da erradicação da pobreza.

Contudo, globalmente a incidência de TB aumentou 3,9% entre os anos de 2020 e 2022, revertendo as quedas observadas nas décadas anteriores. Esse cenário tem sido associado, entre outros fatores, aos impactos indiretos da pandemia de COVID-19 sobre os serviços de saúde, incluindo dificuldades no diagnóstico, na notificação e no acompanhamento dos casos. Esse aumento não ocorre no ritmo necessário para que se alcancem as metas estabelecidas pela Estratégia Fim da TB da Organização Mundial da Saúde (OMS), que prevê uma queda anual de 4 a 5% até o final de 2025 e uma média de 17% ao ano no período de 2025 a 2035. Estudos da Fundação Oswaldo Cruz (Fiocruz 2023) apontam que, dos estimados 10,6 milhões de pessoas que adoeceram de TB em 2022, apenas 7,5 milhões foram detectadas e notificadas, resultando em uma lacuna de 3,1 milhões de casos (OMS, 2023). No contexto brasileiro, os indicadores epidemiológicos também evidenciam desafios persistentes no controle da doença, reforçando a necessidade de estratégias analíticas capazes de apoiar o monitoramento e a tomada de decisão em saúde pública.

Os esforços essenciais para controlar esta doença incluem prevenção, diagnóstico precoce e tratamento adequado. No Brasil, essas ações são orientadas por diretrizes técnicas estabelecidas pelo Ministério da Saúde por meio do Manual de Recomendações para o Controle da Tuberculose no Brasil (BRASIL, 2019). Porém, mesmo com todas as ações desenvolvidas no território brasileiro, os percentuais de cura para os casos novos de TB confirmados laboratorialmente vem apresentando queda desde o ano de 2016, com percentual de 66,5% em 2021. Já os percentuais de abandono no tratamento vêm apresentando um aumento desde 2017, tendo se acentuado em 2020 e 2021 (13,6% e 14% respectivamente), o que está aquém do recomendado pela Organização Mundial de Saúde (OMS), que preconiza que os programas de TB tenham uma taxa de abandono de tratamento inferior a 5%. Em doenças com tratamento prolongado, como a TB, a baixa adesão é um problema frequente, o que representa um desafio para o controle de uma doença que é curável em praticamente 100% dos casos que são sensíveis aos medicamentos

recomendados para o seu tratamento, e que tem o tratamento gratuito e garantido a toda a população brasileira pelo Sistema Único de Saúde (SUS).

De fato, uma das principais preocupações a respeito da TB é a redução das taxas de abandono do tratamento. A baixa adesão e o abandono do tratamento podem resultar na manutenção da cadeia de transmissão da doença, no aumento da mortalidade, e no desenvolvimento de resistência micobacteriana, os quais impõem dificuldades ao processo de cura, aumentando o tempo e o custo do tratamento. Estudos, como o de Pelissari et al. (2018), indicam que os fatores que levam ao abandono do tratamento estão relacionados a diferentes níveis hierárquicos, desde características individuais, como falta de compreensão sobre a importância do tratamento, até limitações no acesso aos serviços de saúde e desigualdades sociais que impactam diretamente as comunidades mais vulneráveis. Compreender as variáveis envolvidas neste processo e quais são os fatores determinantes, pode ser um grande aliado dos gestores de saúde.

Desta forma, para melhor compreensão das variáveis envolvidas nos casos de abandono da TB, torna-se relevante a busca por tecnologias capazes de auxiliar os gestores em saúde neste processo. Dentre as diversas tecnologias disponíveis, a tecnologia de *Business Intelligence* (BI) surge como uma ferramenta capaz de auxiliar no enfrentamento desse desafio, uma vez que possibilita analisar grandes volumes de dados históricos por diversos níveis de granularidade. O BI utiliza processos de integração e organização de dados históricos em um único repositório, como o *Data Warehouse*, possibilitando a construção de dashboards interativos que auxiliam na visualização de padrões e tendências.

No campo da saúde pública, o BI é amplamente reconhecido por seu potencial de transformar dados brutos em informações úteis para a tomada de decisão. Pesquisas como as de Torres et al. (2021), Keskin e Yazici (2022) e Kohara (2023) evidenciam que o uso de BI pode ser um diferencial na análise de determinantes sociais e regionais de saúde, por meio de *insights* que possibilita identificar tendências, padrões e problemas de saúde. Com essas informações em mãos, é possível tomar decisões mais assertivas, melhorar a eficiência operacional e garantir um atendimento de qualidade aos pacientes. A estrutura de *Data Warehouse* (DW) desenvolvida, além de

possibilitar a construção de *dashboards*, ela também pode ser utilizada com fonte para algoritmos e ou técnicas de Inteligência Artificial (IA).

2. OBJETIVOS

2.1. OBJETIVO GERAL

O principal objetivo da pesquisa foi de desenvolver uma ferramenta de *Business Intelligence* (BI) para analisar os fatores associados ao abandono do tratamento da tuberculose nos municípios brasileiros. A proposta busca consolidar dados epidemiológicos e socioeconômicos em um único ambiente analítico, transformar dados de saúde em informações úteis para a tomada de decisão, permitindo identificar padrões e fatores que contribuem para a interrupção do tratamento.

2.2. OBJETIVOS ESPECÍFICOS

OE1. Identificar os principais fatores sociais, econômicos e regionais associados ao abandono do tratamento da tuberculose no Brasil.

OE2. Integrar conhecimentos das áreas de tecnologia e saúde pública, evidenciando o potencial do *Business Intelligence* como ferramenta de análise social.

OE3. Apoiar a gestão pública de saúde por meio da geração de informações e evidências que contribuam para a redução das taxas de abandono do tratamento da tuberculose.

3. REVISÃO DA LITERATURA

Neste capítulo, são apresentados os fundamentos necessários para compreender as abordagens e metodologias que serão adotadas nesta pesquisa.

3.1. TUBERCULOSE

A Tuberculose (TB) é uma doença infecciosa crônica causada pelo *Mycobacterium tuberculosis*, um bacilo álcool-ácido resistente de elevada transmissibilidade. A infecção se dá predominantemente por via aérea, por meio da inalação de partículas contendo o agente etiológico, o que reforça a importância do diagnóstico precoce e do tratamento adequado como estratégias centrais para o seu controle (BARREIRA; GRANGEIRO, 2007; MINISTÉRIO DA SAÚDE, 2023).

No cenário global, a TB permanece entre as enfermidades infecciosas mais letais. Em 2023, a Organização Mundial da Saúde (OMS) registrou aumento expressivo no número de casos diagnosticados — aproximadamente 8,2 milhões — indicando reversão de uma tendência de queda observada nas décadas anteriores (WHO, 2023). No Brasil, a situação acompanha esse movimento, com mais de 80 mil novos casos reportados anualmente e aumento da incidência em diversas Unidades da Federação.

Um dos desafios mais críticos no enfrentamento da TB é o aumento das taxas de abandono do tratamento. Enquanto a OMS recomenda que países mantenham abandono inferior a 5%, o Brasil tem apresentado percentuais superiores a esse parâmetro nos últimos anos, alcançando cerca de 14% em 2021 (BRASIL, 2023). Esse fenômeno compromete a eficácia terapêutica, prolonga a cadeia de transmissão da doença e reduz a efetividade das estratégias de controle clínico e epidemiológico.

O tratamento da tuberculose é ofertado gratuitamente pelo Sistema Único de Saúde (SUS) e baseia-se em esquemas terapêuticos padronizados, com duração mínima de seis meses, podendo variar conforme características clínicas e presença de comorbidades. Estratégias como o Tratamento Diretamente Observado (TDO) são recomendadas como forma de fortalecer a adesão e reduzir os índices de abandono. Entretanto, a longa duração do regime terapêutico, associada a fatores sociais e estruturais, pode dificultar a permanência dos pacientes em acompanhamento

contínuo, tornando o abandono um desafio persistente para os programas de controle da doença.

A literatura aponta que o abandono não decorre de um único fator, mas da interação de elementos individuais, sociais, estruturais e operacionais. Estudos nacionais indicam que a população em situação de rua enfrenta maiores dificuldades de adesão ao tratamento da tuberculose devido a vulnerabilidades sociais, instabilidade habitacional e barreiras no acesso aos serviços de saúde (BRASIL, 2016). Pelissari et al. (2018) demonstram que variáveis como vulnerabilidade socioeconômica, barreiras de acesso aos serviços de saúde, estigma social, percepção sobre a doença e características do território são determinantes importantes no desfecho do tratamento. Dessa forma, compreender o abandono requer uma abordagem analítica, capaz de integrar diferentes dimensões do fenômeno.

Nesse contexto, ferramentas tecnológicas avançadas como a de *Business Intelligence* tornam-se essenciais para apoiar a vigilância epidemiológica. A capacidade de integrar grandes volumes de dados e analisá-los sob múltiplas perspectivas permite identificar padrões de risco, desigualdades regionais e fatores associados ao abandono do tratamento, contribuindo para intervenções mais precisas e efetivas.

3.2. BUSINESS INTELLIGENCE

O *Business Intelligence* (BI) constitui um conjunto de métodos, processos e tecnologias voltados à transformação de dados brutos em informações acionáveis, capazes de subsidiar decisões estratégicas e operacionais. Em um cenário no qual organizações acumulam grandes volumes de dados oriundos de fontes diversas, sistemas transacionais, bancos de dados corporativos, planilhas, plataformas digitais e registros administrativos, torna-se essencial dispor de mecanismos que permitam integrar essas informações de forma estruturada. Conforme Dresner (2008), o BI tem como propósito central organizar, processar e disponibilizar dados em um formato que favoreça análises consistentes, alinhadas às necessidades informacionais das instituições.

A arquitetura de BI envolve etapas fundamentais: captação dos dados, integração e transformação (ETL ou ELT), armazenamento, modelagem e, por fim, visualização e

análise. Esse fluxo é representado na Figura 1, que sintetiza o processo pelo qual dados heterogêneos são convertidos em informação organizada e apta à exploração analítica. Essa estrutura sistemática garante que diferentes fontes possam ser tratadas de maneira padronizada, permitindo o acompanhamento de fenômenos complexos e a comparação entre variáveis de diferentes naturezas.

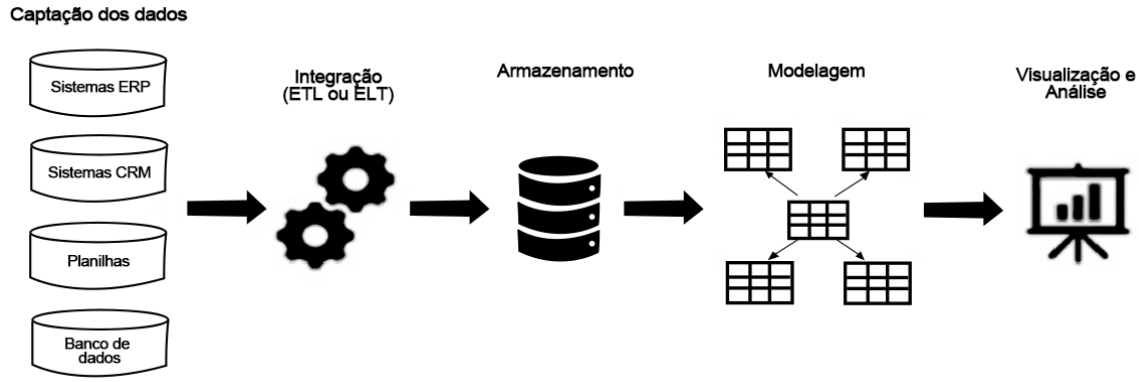

Figura 1 – Modelo multidimensional em esquema estrela do *Data Warehouse*.
Fonte: TREVISAN (2022).

Uma vez preparados, esses dados tornam-se aptos à aplicação de diferentes técnicas analíticas, como mineração de dados, análises estatísticas, identificação de tendências e modelagens preditivas que possibilitam revelar padrões e relações não observáveis por métodos convencionais (TAVERA ROMERO et al., 2021). Dessa forma, o BI amplia a capacidade de interpretação e monitoramento de fenômenos complexos, oferecendo suporte tanto para análises históricas quanto para projeções estratégicas.

Em contextos que envolvem grande variabilidade e múltiplas dimensões informacionais, como a vigilância epidemiológica, o BI destaca-se pela capacidade de consolidar microdados, comparar indicadores e sintetizar informações relevantes para a gestão. Esse processo depende diretamente de estruturas robustas de armazenamento, como os *Data Warehouses*, que fundamentam o tratamento e a modelagem dos dados e serão discutidos no tópico seguinte.

3.2.1. Data Warehouse

O *Data Warehouse* (DW) constitui a base estrutural dos sistemas de *Business Intelligence*, funcionando como um repositório consolidado destinado a análises

estratégicas. Segundo Inmon e Hackathorn (1994), um DW é caracterizado por ser orientado a assuntos, integrado, não volátil e variante no tempo, permitindo armazenar grandes volumes de dados históricos de forma consistente. Complementarmente, Kimball e Ross (2013) enfatizam que o DW organiza essas informações em modelos dimensionais projetados para otimizar consultas analíticas, facilitar interpretações e apoiar a tomada de decisão, enquanto Machado (s.d.) destaca o *Data Warehouse* como uma estrutura orientada à análise multidimensional, permitindo a visualização dos dados sob diferentes perspectivas.

A modelagem dimensional com destaque para o *Star Schema* é amplamente empregada na construção de DW por separar medidas quantitativas, presentes nas tabelas fato, de atributos descritivos, estruturados em tabelas dimensão. Essa abordagem melhora o desempenho das consultas, simplifica o acesso às informações e permite analisar fenômenos sob diferentes perspectivas, como tempo, localização e características populacionais.

No contexto da saúde pública, especialmente em análises epidemiológicas, o DW apresenta vantagens significativas. Ele possibilita a integração de dados clínicos, operacionais, sociodemográficos e territoriais provenientes de múltiplas fontes, assegurando padronização, integridade e histórico temporal. Essa estrutura suporta investigações mais robustas, permitindo identificar padrões, monitorar indicadores e avaliar fatores associados a desfechos complexos, como o abandono do tratamento da tuberculose.

3.3. TRABALHOS RELACIONADOS

Os trabalhos relacionados foram identificados por meio de buscas sistemáticas em bases de dados científicas nacionais e internacionais, incluindo Google Scholar, *SciELO* e PubMed. Utilizaram-se combinações de palavras-chave relacionadas a *Business Intelligence, Data Warehouse*, tuberculose, abandono do tratamento e análise epidemiológica. Foram selecionados estudos relevantes ao tema, priorizando publicações recentes que abordassem a aplicação de tecnologias de análise de dados em saúde pública e investigações sobre os fatores associados ao abandono do tratamento da tuberculose. A seleção considerou a pertinência metodológica e a contribuição dos trabalhos para os objetivos desta pesquisa.

3.3.1. Kohara Melchior et al. (2022)

Kohara Melchior et al. (2022) discutem a aplicação de *Business Intelligence* como ferramenta de apoio à gestão estratégica em saúde, evidenciando como soluções de análise e visualização de dados podem fortalecer o monitoramento de indicadores e a tomada de decisão em serviços de saúde. O trabalho reforça a importância do BI no contexto da saúde pública ao demonstrar que a organização e apresentação adequada dos dados contribui para maior agilidade, clareza e efetividade na avaliação de cenários epidemiológicos. No aspecto metodológico, os autores utilizaram dados de tuberculose do estado do Acre, referentes ao período de 2010 a 2020, extraídos do SINAN e organizados em planilhas. A partir desses dados, foi desenvolvido um *dashboard* no Microsoft Power BI, envolvendo etapas de preparação e padronização da base, definição de relacionamentos entre variáveis e construção de visualizações para exploração por recortes temporais e territoriais.

A contribuição do estudo para este TCC está na demonstração prática de que ferramentas de BI podem ser aplicadas diretamente a dados reais de tuberculose para apoiar análises epidemiológicas. Dessa forma, o trabalho reforça a viabilidade de estruturar e visualizar indicadores em painéis analíticos, fornecendo base metodológica e inspiração para a construção de um ambiente analítico voltado a investigar desfechos como o abandono do tratamento.

3.3.2. Torres et al. (2021)

Torres et al. (2021) investigaram a aplicabilidade de ferramentas de *Business Intelligence* (BI) na Atenção Primária à Saúde (APS), destacando o potencial dessas soluções no apoio à gestão e no monitoramento de indicadores epidemiológicos. Os autores argumentam que o BI possibilita consolidar e cruzar dados provenientes de múltiplas fontes, ampliando a capacidade de análise territorial, temporal e operacional. Essa abordagem contribui para a identificação de tendências e padrões que dificilmente seriam observados por métodos tradicionais de tabulação de dados.

No aspecto metodológico, trata-se de um estudo de caso exploratório, desenvolvido a partir de dados coletados entre junho de 2018 e setembro de 2019, no contexto do município do Rio de Janeiro. O estudo avaliou a adoção do BI como ferramenta de

gestão na APS, com ênfase na construção de um modelo lógico de aplicabilidade e na implementação de rotinas de extração, integração e visualização de dados em *dashboards* analíticos. Para operacionalizar a análise, foi utilizada a plataforma Qlik Sense Cloud[1], selecionada por permitir a integração com diferentes bases de dados e por viabilizar a construção de *dashboards* sem a necessidade de infraestrutura local.

Os autores propõem uma organização estruturada do ciclo de dados na APS, na qual informações territoriais, demográficas e operacionais provenientes de diferentes sistemas são consolidadas, integradas e transformadas em indicadores analíticos para suporte à tomada de decisão. Essa abordagem se aproxima de um raciocínio típico de análise multidimensional, em que medidas e indicadores podem ser observados a partir de diferentes dimensões, como tempo, território e características da população, ampliando a interpretação dos resultados e a sistematização do processo de gestão.

A relevância do estudo reside na demonstração prática de que sistemas de BI podem aprimorar análises territoriais complexas e o acompanhamento contínuo de indicadores na APS, servindo como base metodológica para projetos que demandem monitoramento permanente e análises orientadas a recortes espaciais e temporais, como é o caso da presente investigação sobre o abandono do tratamento da tuberculose.

3.3.3. Pelissari et al. (2018)

Pelissari et al. (2018) realizaram um estudo ecológico com abrangência nacional visando identificar cenários de controle da tuberculose no Brasil a partir de determinantes socioeconômicos, epidemiológicos e operacionais associados à incidência da doença. Os autores demonstram que a ocorrência da tuberculose está relacionada a múltiplos fatores estruturais e de desempenho dos serviços de saúde, reforçando a necessidade de análises territoriais integradas e orientadas por evidências para subsidiar políticas públicas e estratégias de enfrentamento mais direcionadas.

[1] Qlik Sense Cloud. Disponível em: https://www.qlik.com/us/products/qlik-sense. Acesso em: 05 jan. 2026.

No aspecto metodológico, trata-se de um estudo ecológico, no qual a unidade de análise foram os municípios brasileiros (5570 em 2015, com exclusão de cinco por ausência de dados socioeconômicos). Como fontes de dados, os autores utilizaram registros de novos casos de TB e variáveis epidemiológicas/operacionais provenientes do Sistema de Informação de Agravos de Notificação (SINAN) e do Sistema de Informações sobre Mortalidade (SIM), além de variáveis socioeconômicas e demográficas do Censo 2010. O desfecho analisado foi a taxa de incidência de tuberculose em 2015.

A análise estatística foi conduzida em duas etapas. Inicialmente, os autores aplicaram regressão binomial negativa para identificar variáveis socioeconômicas associadas à incidência de TB. Em seguida, utilizaram o método de clusterização *k-means* para agrupar municípios em cenários conforme as variáveis selecionadas, definindo o número de clusters por meio do *Elbow method*. Posteriormente, cada cenário socioeconômico foi subdividido em subcenários com base em variáveis epidemiológicas e operacionais (por exemplo, mortalidade por TB, detecção de AIDS, proporção de populações vulneráveis e indicadores de acompanhamento/registro), também por meio de modelagem e clusterização.

Como resultados, o estudo identificou dois grandes cenários socioeconômicos, determinados principalmente por desemprego e adensamento domiciliar, e, em uma segunda etapa, subcenários epidemiológicos/operacionais, evidenciando que determinados grupos de municípios apresentam fragilidades nos sistemas de informação e outros concentram casos em populações-chave. Assim, a classificação proposta contribui para decisões direcionadas, como melhoria da qualidade dos dados, fortalecimento de ações de vigilância e implementação de políticas sociais integradas voltadas a grupos vulneráveis.

A relevância desse trabalho para esta pesquisa reside no enfoque multiescalar e territorial ao integrar variáveis estruturais e operacionais na caracterização do risco para tuberculose. Embora não utilize ferramentas de BI, o estudo reforça a necessidade de modelos analíticos capazes de consolidar grandes volumes de dados e permitir segmentações por contexto municipal, contribuindo diretamente para investigações sobre desfechos como o abandono do tratamento.

3.3.4. Keskin e Yazici (2022)

Keskin e Yazici (2022) propõem um *framework* de análise multidimensional voltado a dados espaço-temporais incertos, denominado FSOLAP (*Fuzzy Spatial* OLAP), com o objetivo de suportar consultas analíticas complexas sobre dados com imprecisão e “fronteiras difusas”. O trabalho reforça que abordagens tradicionais de OLAP/SOLAP são eficazes para exploração multidimensional de dados, porém tendem a ser limitadas quando aplicadas a contextos em que a incerteza é inerente, como fenômenos meteorológicos e aplicações espaço-temporais, característica também presente em muitos cenários epidemiológicos.

No aspecto metodológico, os autores descrevem uma arquitetura em camadas composta por fontes de dados, ETL, camada estruturada (PostgreSQL/PostGIS), servidor SOLAP e camada de apresentação, permitindo consultas multidimensionais via MDX e visualizações integradas (mapas e cubos). A proposta inclui um módulo fuzzy responsável por transformar medidas “crisp” (valores tradicionais) em classes fuzzy, gerando funções de pertinência e suportando consultas flexíveis. O processo de fuzzificação é automatizado: primeiramente utiliza-se o algoritmo X-means, uma extensão do k-means, para estimar automaticamente o número apropriado de clusters e, na sequência, aplica-se Fuzzy C-Means para definir agrupamentos fuzzy e graus de pertinência.

Além disso, o *framework* incorpora mineração de conhecimento por meio de regras de associação fuzzy, geradas com FP-Growth, e um mecanismo de inferência (*Fuzzy Inference System* – FIS) para suportar consultas preditivas. Dessa forma, o FSOLAP amplia as operações OLAP ao lidar não apenas com consultas espaciais e temporais convencionais, mas também com consultas fuzzy (não espaciais, espaciais e espaço-temporais), incluindo tipos preditivos que demandam inferência baseada em regras.

A relevância deste trabalho para o presente TCC está em fornecer um embasamento sólido sobre arquiteturas analíticas robustas baseadas em *Data Warehouse*, modelagem multidimensional e operações OLAP/SOLAP, evidenciando que análises espaço-temporais complexas podem ser aprimoradas com estruturas hierárquicas, cubos analíticos e mecanismos que lidam com a incerteza dos dados. Assim, o artigo apoia conceitualmente a construção de um ambiente analítico para explorar

determinantes e padrões associados ao abandono do tratamento da tuberculose sob recortes territoriais e temporais.

3.3.5. Schwambach et al. (2024)

Schwambach et al. (2024) descrevem a aplicação de *Business Intelligence* (BI) voltada à automação da análise de dados em saúde, com o objetivo de transformar bases extensas em informações organizadas e acessíveis para apoio à gestão. Os autores ressaltam que, diante da ampla disponibilidade de dados e da necessidade de acompanhar indicadores de saúde, soluções analíticas baseadas em DW e *dashboards* contribuem para reduzir esforços manuais de tabulação e ampliar a capacidade de monitoramento e análise.

No aspecto metodológico, o trabalho estrutura a implementação em fases. Inicialmente, são definidos os indicadores e selecionadas as bases de dados utilizadas no processo, incluindo sistemas como o SIM (para indicadores de mortalidade) e o SIH (para indicadores de hospitalizações). Em seguida, é descrita a criação de uma *staging Area* no PostgreSQL, empregada como camada intermediária para armazenar temporariamente os dados em seu formato original e realizar a preparação antes do carregamento no DW. Na etapa de ETL, os dados passam por procedimentos de limpeza e padronização (como definição de tipos de dados e ajustes de campos textuais), sendo posteriormente utilizados para alimentar tabelas dimensionais e tabelas fato no ambiente analítico.

A modelagem proposta apresenta um modelo dimensional, com separação entre tabelas fato e dimensões, destacando que as dimensões foram tratadas como *Slowly Changing Dimensions* (SCD) do tipo 1 e que cada registro dimensional recebe uma *Surrogate Key*, utilizada como chave primária e como referência nas tabelas fato. Para apoiar a extração de informações do DATASUS (como o Código Brasileiro de Ocupação, CBO), os autores empregam o módulo Pysus na linguagem Python, enquanto o gerenciamento do fluxo do ETL é realizado com o Apache Airflow, que permite a orquestração de processos automatizados. Por fim, a etapa analítica e a construção dos *dashboards* são realizadas com a ferramenta Tableau Desktop, com disponibilização do painel em ambiente público.

A relevância do estudo para este TCC está em demonstrar, de forma prática, um fluxo completo envolvendo definição de indicadores, preparação dos dados, ETL, construção de um DW dimensional e disponibilização de informações em *dashboards*. Assim, o trabalho oferece suporte metodológico e técnico por evidenciar como arquiteturas de BI podem fortalecer análises epidemiológicas e apoiar interpretações a partir de recortes analíticos estruturados, aspecto essencial em investigações complexas como a análise do abandono do tratamento da tuberculose.

3.3.6. Lucena et al. (2023)

Lucena et al. (2023) investigaram os fatores associados ao abandono do tratamento da tuberculose no Brasil, evidenciando que esse desfecho é influenciado por um conjunto de determinantes clínicos, sociodemográficos e comportamentais. O estudo reforça que a interrupção do tratamento está relacionada a vulnerabilidades sociais e barreiras no seguimento terapêutico, destacando a importância de identificar perfis e condições que aumentam o risco de abandono para subsidiar ações de vigilância e intervenções direcionadas.

No aspecto metodológico, trata-se de uma revisão sistemática conduzida conforme diretrizes do PRISMA, reunindo evidências de estudos realizados em diferentes contextos brasileiros e com distintas amostras. A síntese dos achados aponta associação significativa do abandono do tratamento com fatores como sexo masculino, raça/cor preta, faixa etária de 19 a 49 anos, coinfecção por HIV, baixa escolaridade, uso de álcool e drogas ilícitas, além de tratamento não supervisionado. Os autores destacam que tais fatores frequentemente se sobrepõem, configurando perfis de maior vulnerabilidade ao abandono e reforçando a necessidade de abordagens analíticas integradas para compreensão do fenômeno.

A relevância deste trabalho para o presente TCC está em fornecer uma síntese consolidada dos principais fatores relacionados ao abandono, contribuindo diretamente para a fundamentação teórica e para a seleção/organização das variáveis analisadas neste estudo. Dessa forma, o artigo sustenta a proposta de construção de um ambiente analítico baseado em integração de dados e BI, uma vez que o abandono do tratamento exige análises multidimensionais que considerem

simultaneamente tempo, território e características do paciente/assistência, possibilitando identificar padrões e grupos prioritários para intervenção.

3.3.7. Mendonça et al. (2025)

Mendonça et al. (2025) investigam fatores epidemiológicos, socioeconômicos e relacionados aos serviços de saúde associados à interrupção do tratamento da tuberculose em municípios brasileiros, utilizando dados nacionais e indicadores estruturais do território. O estudo evidencia que a adesão terapêutica não depende exclusivamente das características individuais dos pacientes, sendo influenciada também por aspectos contextuais como vulnerabilidade social, organização da atenção primária à saúde e condições socioeconômicas regionais. A pesquisa reforça a importância da análise integrada de dados para compreensão do abandono do tratamento em escala populacional.

A contribuição do estudo para o presente Trabalho de Conclusão de Curso está relacionada à fundamentação teórica sobre determinantes associados ao abandono do tratamento da tuberculose em nível nacional. Dessa forma, os achados apresentados pelos autores dialogam com a proposta de utilização de Business Intelligence como suporte à análise exploratória e à visualização de indicadores epidemiológicos, auxiliando na identificação de padrões e no apoio à tomada de decisão em saúde pública.

4. METODOLOGIA

Este estudo adotou uma abordagem aplicada e exploratória, com foco na implementação de uma arquitetura de *Business Intelligence* (BI) destinada à análise dos fatores associados ao abandono do tratamento da tuberculose nos municípios brasileiros. A metodologia foi estruturada em etapas sequenciais, abrangendo desde a obtenção dos dados até a construção de *dashboards* analíticos interativos, de modo a permitir a integração, organização e visualização das informações de forma sistematizada.

O processo metodológico compreendeu quatro fases principais: identificação das fontes de dados, integração e tratamento por meio do processo de Extração, Transformação e Carga (ETL), construção do *Data Warehouse* dimensional e desenvolvimento dos *dashboards* analíticos. A arquitetura geral da solução de BI adotada neste estudo é apresentada na Figura 2, a qual ilustra o fluxo de dados desde as fontes de origem até a camada de visualização, evidenciando as etapas de integração, armazenamento e análise das informações.

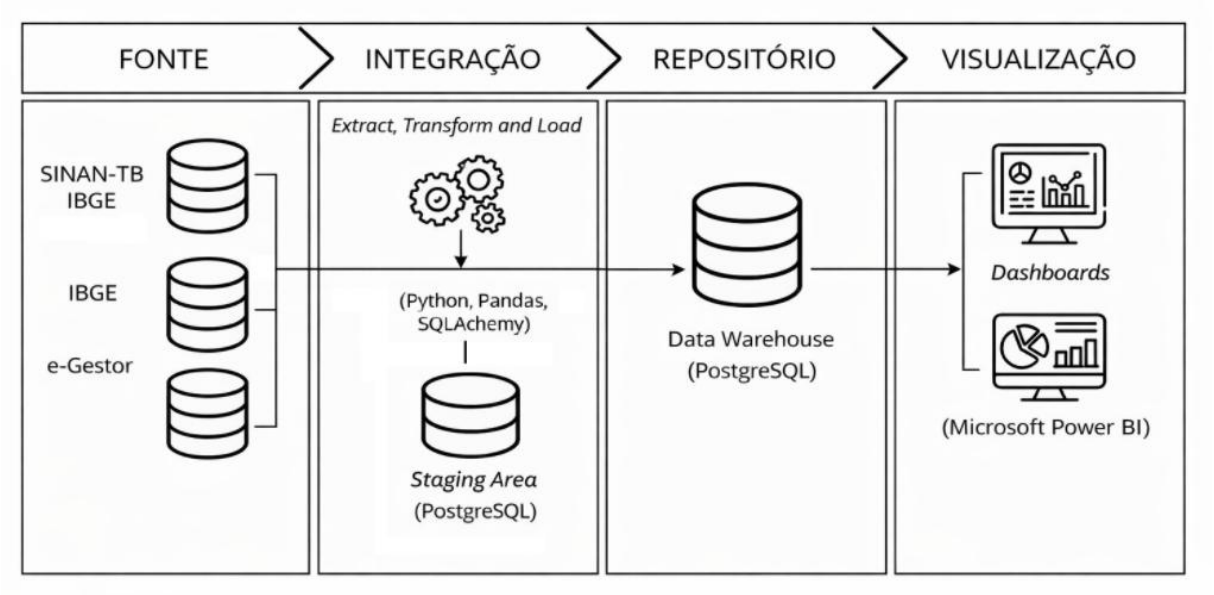

Figura 2 - Arquitetura de BI proposta.
Fonte: Adaptada de Kimball e Ross (2013).

4.1 FASE 1 – FONTE DE DADOS

No desenvolvimento deste trabalho, foram utilizadas três fontes de dados, sendo a principal o Sistema de Informação de Agravos de Notificação – Tuberculose (SINAN-TB), disponibilizado pelo Programa Nacional de Controle da Tuberculose, vinculado ao Departamento de Imunização e Doenças Transmissíveis da Secretaria de Vigilância em Saúde do Ministério da Saúde, mediante solicitação no Sistema de Informação ao Cidadão (e-SIC). Essa base reúne todos os casos confirmados da doença, cuja notificação é compulsória.

Cabe destacar que este projeto foi desenvolvido em parceria acadêmica com o Prof. Dr. Tiago Ricardo Moreira, da Universidade Federal de Viçosa, por meio da qual foram disponibilizadas as bases de dados iniciais utilizadas neste estudo. Essa colaboração foi fundamental para o acesso estruturado aos microdados epidemiológicos, bem como para a viabilização das etapas de integração, modelagem dimensional e análise no ambiente de *Business Intelligence*.

Além disso, também foram utilizados dados do Censo Demográfico de 2022 do Instituto Brasileiro de Geografia e Estatística (IBGE, 2023) e informações históricas de cobertura da Atenção Primária à Saúde provenientes do sistema e-Gestor (BRASIL, 2024).

A obtenção dos dados do IBGE foi por meio do Atlas do Desenvolvimento Humano (PNUD; IPEA; FJP, 2020), que contém indicadores socioeconômicos e demográficos, dos quais foram extraídos os seguintes atributos: Índice de Desenvolvimento Humano Municipal (IDHM), IDHM renda, renda per capita, índice de Gini, expectativa de vida, proporção de pobreza e população municipal. Esses dados possibilitam que o DW amplie sua capacidade de análise e possibilita avaliações que relacionam as condições de vulnerabilidade social aos casos de abandono do tratamento da tuberculose.

Uma vez obtidas as bases de dados e seus respectivos indicadores, foi realizado um processo de modelagem dimensional para determinar a tabela fato e as dimensões necessárias para atender aos requisitos propostos. A modelagem dimensional é uma técnica de design de banco de dados utilizada para organizar e estruturar dados de

forma intuitiva e eficiente, facilitando a análise e a extração de informações. Essa abordagem baseia-se em dois conceitos principais: fatos e dimensões. Os fatos representam as métricas que se deseja analisar, enquanto as dimensões correspondem aos contextos ou características associadas a esses fatos. Isso permite criar esquemas flexíveis e simplificados, oferecendo uma compreensão mais clara dos dados e facilitando a execução de consultas analíticas.

Dado a natureza dos dados, a tabela fato (fNotificacao) concentra as variáveis relacionadas ao indivíduo e ao evento de notificação, das quais inclui:

- Características sociodemográficas, tais como: sexo, idade, raça/cor, escolaridade;
- Condições clínicas, como: forma clínica, presença de comorbidades, histórico patológico;
- Agravos e fatores associados, como: diabetes, AIDS, alcoolismo, doença mental, tabagismo, uso de drogas ilícitas;
- Informações assistenciais, como: tipo de notificação, supervisão do tratamento, realização de TDO, testes diagnósticos;
- Variáveis derivadas, representadas por: faixa etária, intervalos temporais entre diagnóstico, notificação, início e encerramento;
- Variável dependente do estudo, representada pelo indicador de abandono, que neste caso foi considerado o abandono no tratamento de TB (Sim ou Não).

A variável dependente é identificada no encerramento de casos. Portanto, considera-se caso de abandono de tratamento, o doente que, após iniciado o tratamento para TB, deixou de comparecer à unidade de saúde por mais de trinta dias consecutivos, após a data para o seu retorno. Desta forma, os atributos listados compõem o núcleo analítico da tabela fato que permitem examinar tanto fatores individuais quanto condições clínicas e assistenciais.

Posteriormente, foi criada uma estrutura intermediária (*staging area*) utilizando o Sistema de Gerenciamento de Banco de Dados Objeto-Relacional (ORDBMS)

PostgreSQL, que funcionou como área de preparação dos dados. O principal propósito dessa estrutura foi armazenar as informações em sua forma original até que pudessem ser acessadas pelas camadas subsequentes do DW. Os dados permaneceram na *staging area* até estarem devidamente sincronizados e prontos para uso. Em essência, a *staging area* é um espaço intermediário no qual os dados são temporariamente armazenados e tratados antes de serem carregados no ambiente de produção.

4.2 FASE 2 - INTEGRAÇÃO

A segunda etapa do processo consistiu na execução das rotinas de Extração, Transformação e Carga de dados — ETL (*Extract, Transform and Load*), com a finalidade de integrar, padronizar e estruturar as informações provenientes das bases operacionais e da fonte secundária, preparando-as para a composição do *Data Warehouse* (DW), descrito na Fase 3.

Nesta etapa, utilizou-se a linguagem de programação Python, com as bibliotecas Pandas e NumPy para o processamento e manipulação dos dados, além da biblioteca SQLAlchemy para a comunicação direta com o Sistema Gerenciador de Banco de Dados PostgreSQL, por meio de uma engine configurada com o adaptador Psycopg2.

O script desenvolvido executa, de forma automatizada, as três etapas clássicas do processo de ETL, conforme descrito a seguir:

i. extração das bases de dados de entrada;

ii. transformação, limpeza, padronização e enriquecimento das informações;

iii. carga das dimensões (dLocal e dTempo), bem como da tabela fato (fNotificacao).

Com o objetivo de formalizar o fluxo computacional das rotinas de ETL e da construção das estruturas do *Data Warehouse*, foi elaborado o pseudocódigo, representado na Figura 3, que descreve de maneira sequencial as principais etapas do processo, desde a extração das fontes de dados até a disponibilização das informações para análise no ambiente de *Business Intelligence*.

Pseudocódigo

```
INICIAR
CARREGAR_FONTE_DE_DADOS:
        Ler arquivo SINAN-TB: dadosCompleto.xlsx
        Ler indicadores sociodemográficos: atlas.csv
        Ler população e regiões do IBGE: pop_ibge.csv
PRE_PROCESSAR_DADOS:
        Padronizar códigos IBGE para 7 dígitos
        Normalizar tipos (datas, inteiros, strings)
        Tratar valores ausentes e inconsistências
PROCESSAR_DADOS:
        CONSTRUIR_DIMENSAO_LOCAL:
        Unificar municípios de residência e notificação
        Enriquecer com indicadores do Atlas (IDHM, Gini, renda, etc.)
        Enriquecer com dados populacionais e região geográfica
        Remover registros duplicados
        Gerar tabela dLocal
        CONSTRUIR_DIMENSAO_TEMPO:
        Converter dt_notific para tipo datetime
        Extrair atributos: ano, mês, dia, trimestre
        Validar intervalos de datas
        Remover datas inconsistentes
        Gerar tabela dTempo
        CONSTRUIR_FATO_NOTIFICACAO:
        Mapear variáveis categóricas (sexo, raça/cor, escolaridade, etc.)
        Calcular faixa etária
        Calcular intervalos clínicos:
                        início_sintomas → notificação
                        diagnóstico → início_tratamento
        Padronizar UF, município e códigos
        Remover duplicidades
        Associar chaves da dLocal e dTempo
        Gerar tabela fato fNotificacao
INSERIR_DADOS_NO_DW:
        Inserir registros da dimensão dLocal (modo incremental)
        Inserir registros da dimensão dTempo (modo incremental)
        Inserir fNotificacao garantindo:
                    Integridade referencial
                    chaves estrangeiras válidas
        Registrar inconsistências em etl_erros.log
        Commit das transações
                Se erro → Rollback
FINALIZAR_DATA_WAREHOUSE:
        Validar integridade das dimensões e da tabela fato
        Otimizar índices e chaves
PUBLICAR_DASHBOARDS:
        Conectar Power BI ao DW
        Criar visualizações analíticas e indicadores
ENCERRAR ALGORITMO
```

Pseudocódigo 1. Pseudocódigo do processo de ETL.

As seções a seguir detalham as ações decorrentes do processo de ETL.

4.2.1. Extração dos dados

A extração consistiu na leitura de duas bases de dados. A primeira foi a base de notificações de tuberculose do SINAN-TB, disponibilizada em um arquivo CSV, que como já descrito anteriormente, contém as variáveis sociodemográficas, clínicas, temporais e de encerramento para cada caso registrado. Para a leitura, utilizou-se a função pd.read_excel, que preserva os tipos e formatos originais dos dados.

Os arquivos do Atlas do Desenvolvimento Humano (PNUD/IBGE), também são disponibilizados no formato CSV, e contém indicadores socioeconômicos relacionados aos municípios brasileiros. Neste arquivo forma aplicadas as primeiras transformações, tais como definição de separador, tradução de caracteres, normalização de identificadores municipais e conversão de tipos numéricos. Ambas as bases de dados foram carregadas em DataFrames que funcionaram como staging area.

4.2.2. Transformação dos dados

A segunda etapa do processo de ETL consistiu na aplicação de um conjunto estruturado de procedimentos destinados a higienizar, padronizar e enriquecer os dados, assegurando conformidade com o modelo dimensional proposto. Para tanto foram realizadas as seguintes rotinas:

a) Padronização de identificadores geográficos

Conforme descrito no Anexo B, que apresenta o dicionário de dados das bases utilizadas, os códigos dos municípios foram inicialmente convertidos para o formato textual, de modo que os espaços fossem removidos e o dígito verificador final truncado, garantindo consistência entre as diferentes fontes. A partir dos campos geográficos das bases de notificação e residência (sg_uf_not, codigo_ibge, id_mn_resi), foram derivados os conjuntos de municípios associados a cada caso. Após a remoção de duplicidades, obteve-se a estrutura-base para a construção da dimensão dLocal.

b) Tratamento de valores ausentes e inconsistentes

Para padronizar valores faltantes ou inválidos, foram desenvolvidas funções no *script*, que reconhece padrões como NaN, #NULL!, vazios ou campos ilegíveis, substituindo-os por códigos padronizados, como o caso do código 9 transformado para “ignorado”. Também foram utilizadas funções para realizar a conversão segura de valores textuais ou decimais em inteiros, garantindo compatibilidade com o tipo SMALLINT do PostgreSQL.

c) Conversão e validação de datas

As variáveis temporais (dt_notific, dt_diag, dt_inic_tr, dt_encerra) foram convertidas com a função safe_parse_date(), como representada na linha 19 do pseudocódigo, a mesma encapsula pd.to_datetime (errors="coerce"), gerando “None” para datas inválidas. Essa abordagem evitou falhas no cálculo de intervalos e assegurou consistência da dimensão dTempo. Portanto, a dimensão dTempo foi construída a partir das datas válidas de notificação, das quais foram extraídas a data completa, o dia, o mês, o ano e o trimestre.

d) Mapeamento e recodificação de variáveis categóricas

As variáveis categóricas foram padronizadas para eliminar divergências de codificação presentes no SINAN. Inicialmente, foram identificados e tratados os valores nulos ou inconsistentes. Na sequência foram utilizadas funções para converter múltiplas representações textuais ou numéricas em categorias unificadas, compatíveis com o modelo dimensional. Foi o caso de valores presentes em atributos como: sexo, escolaridade, raça e gestante.

Por fim, as variáveis clínicas e sociais, como presença de comorbidades, vulnerabilidades, resultados laboratoriais e tipo de notificação, foram tratadas para garantir a padronização e a estabilidade das categorias armazenadas.

e) Derivação de atributos analíticos

É muito comum em ambientes de DW o uso de atributos derivados, especialmente nas tabelas fato. Nesse trabalho, dois grupos principais de atributos derivados foram construídos. O primeiro refere-se à faixa etária, calculada pela diferença entre o ano de nascimento e o ano da notificação, permitindo análise por grupos etários

consolidados. O segundo grupo refere-se aos intervalos temporais assistenciais, calculados em dias, como foi o caso dos pares de atributos:

- diagnóstico e notificação;
- notificação e início do tratamento;
- início do tratamento e encerramento.

Esses indicadores são importantes, pois apoiam análises de dinâmica assistencial e eventuais atrasos nos fluxos de cuidado.

f) Enriquecimento com indicadores socioeconômicos

Com o propósito de enriquecer a base de dados com indicadores socioeconômicos vinculados aos municípios, foram realizados procedimentos para normalizar os identificadores municipais e selecionar um único registro por município. A partir disso, o conjunto de municípios do SINAN foi vinculado ao Atlas, gerando uma dimensão geográfica enriquecida com variáveis como IDHM, renda per capita, índice de Gini, expectativa de vida e população.

g) Remoção de duplicidades e exclusão de atributos não utilizados

Antes da carga dos dados, registros duplicados foram removidos com a finalidade de garantir a integridade analítica. Também foram excluídos atributos sem relevância analítica, com elevado percentual de valores ausentes, duplicados ou sem correspondência no modelo dimensional. Optou-se por remover atributos com mais de 30% de valores ausentes. Essa decisão visou reduzir complexidade, eliminar ruído informacional e manter apenas variáveis úteis ao objetivo analítico, aprimorando desempenho e clareza interpretativa.

4.2.3. Carga dos dados no DW

A carga dos dados foi realizada por meio de transação controlada no PostgreSQL. Inicialmente, foram carregadas as dimensões dLocal e dTempo, de forma incremental, ou seja, registros existentes são reutilizados, enquanto novos registros são inseridos. Em seguida, foi realizada a carga da tabela fato, em que para cada registro, são recuperadas as chaves estrangeiras das dimensões, que são validadas e integradas

às variáveis categóricas e aos atributos derivados. Valores são verificados para garantir conformidade com SMALLINT. Registros sem correspondência dimensional são descartados e registrados em um arquivo de log. Caso não ocorra falhas críticas, a transação é confirmada; caso contrário, é revertida, preservando a integridade do DW.

4.3 FASE 3 - REPOSITÓRIO

O modelo dimensional foi desenvolvido após a identificação dos requisitos e a análise preliminar dos dados provenientes da fonte, seguindo as diretrizes do *star schema*, adotado neste estudo. Trata-se de uma arquitetura amplamente utilizada em soluções de DW devido à sua simplicidade, eficiência em consultas e capacidade de organizar dados heterogêneos de maneira integrada. A Figura 3 apresenta o modelo dimensional proposto, composto por uma tabela fato (fNotificacao) e duas dimensões associadas (dLocal e dTempo). Cabe ressaltar que, o modelo conta com várias outras dimensões degeneradas (*Degenerate dimension* – DD), também conhecidas como dimensões mascaradas.

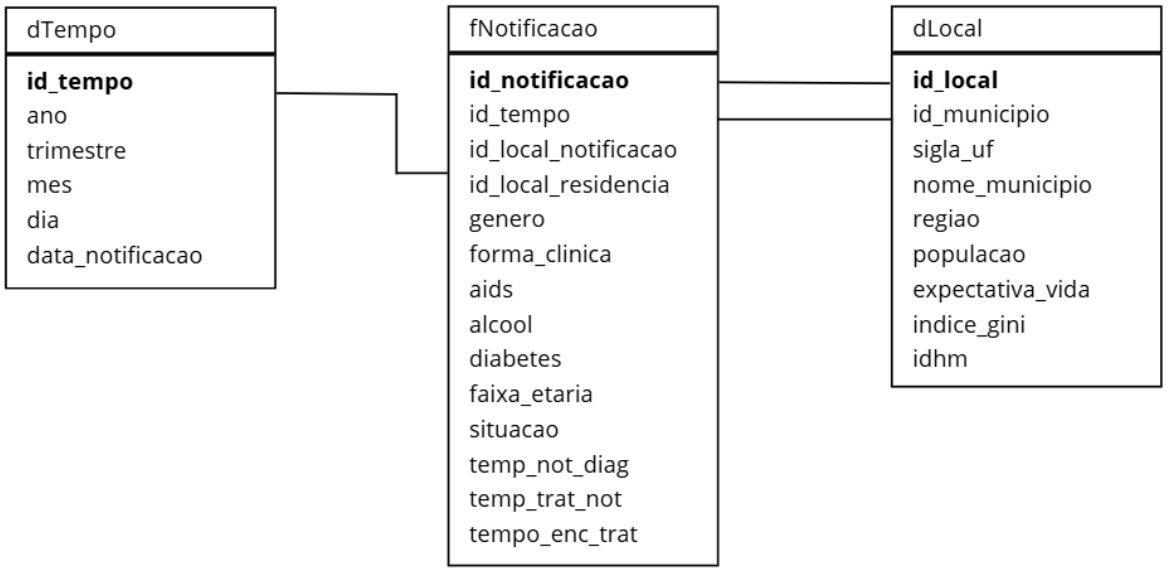

Figura 3 - Modelo dimensional proposto.
Fonte: autor.

A tabela fato concentra os registros individuais de notificação de tuberculose e agrega as medidas relevantes para análise. Nela são armazenadas as variáveis clínicas e sociodemográficas padronizadas durante o processo de ETL, como gênero, raça/cor, escolaridade, comorbidades e condições de vulnerabilidade, além de variáveis

derivadas, tais como os intervalos de tempo entre diagnóstico, notificação, início do tratamento e encerramento. A tabela fato também registra o indicador de abandono, permitindo que o fenômeno seja analisado de forma quantitativa e segmentada por diferentes dimensões analíticas.

A dimensão dLocal representa o contexto geográfico associado a cada notificação, armazenando atributos como sigla da unidade federativa, código do município e nome do município. Além desses atributos básicos, a dimensão incorpora indicadores socioeconômicos provenientes do Atlas do Desenvolvimento Humano, como IDHM, renda per capita, índice de Gini, proporção de pobreza, expectativa de vida e população. Esse enriquecimento possibilita associar características demográficas e sociais ao padrão de abandono, favorecendo análises mais completas e comparações regionais.

A dimensão dTempo descreve a estrutura temporal da notificação, armazenando atributos derivados da data do evento, como dia, mês, ano e trimestre. Essa dimensão viabiliza análises de sazonalidade, tendências temporais e evolução anual dos casos, além de suportar a construção de séries históricas no ambiente de visualização.

É importante destacar que, tanto a dimensão dLocal quanto dTempo, utilizam *surrogate keys* numéricas como chaves primárias. Essas chaves substitutas são geradas exclusivamente no contexto do DW e são utilizadas para garantir independência dos códigos originais das fontes, padronizar os relacionamentos e otimizar consultas. Na tabela fato, as *surrogate keys* são registradas como chaves estrangeiras, estabelecendo os vínculos necessários entre fatos e dimensões e assegurando a integridade referencial do modelo.

Concluído o processo de ETL (Fase 2), o DW foi carregado com os dados tratados, consolidando o repositório analítico (Fase 3). Com isso, o ambiente passou a estar preparado para as análises exploratórias, consultas multidimensionais e para a construção dos *dashboards* e painéis de visualização descritos na Fase 4.

4.4 FASE 4 - VISUALIZAÇÃO

Uma vez que as estruturas do DW foram preenchidas, utilizou-se a ferramenta de processamento analítico online (OLAP) Microsoft Power BI Desktop para a criação

dos *dashboards*, dos quais os indicadores e métricas com os resultados obtidos foram mapeados, de acordo com os requisitos identificados na etapa inicial.

A escolha pela ferramenta OLAP Power BI Desktop (PBI), se deve ao fato de possuir uma licença gratuita, para estudantes universitários e pesquisadores, e por apresentar uma ampla gama de recursos, que favorecem a agilidade e a intuição no processo de criação das visualizações. O PBI estabelece uma conexão direta com o banco de dados, permitindo a configuração da associação entre as dimensões e os fatos.

4.5. FERRAMENTAS UTILIZADAS

Nesta seção são apresentadas as ferramentas utilizadas durante o desenvolvimento deste projeto, especificamente nas etapas metodológicas, como no processo de ETL, na construção do *Data Warehouse* (DW) e na elaboração dos *dashboards* analíticos, aplicados ao estudo dos casos de abandono de tratamento da tuberculose no Brasil. As tecnologias foram selecionadas baseando nos seguintes critérios: confiabilidade, interoperabilidade, suporte à análise de grandes volumes de dados e compatibilidade com a arquitetura proposta, considerando uma abordagem orientada à construção de uma arquitetura de *Business Intelligence* em camadas, na qual a linguagem Python atua no processo de ETL, o PostgreSQL no armazenamento dimensional e o Power BI na camada analítica. Essa escolha priorizou ferramentas alinhadas ao fluxo completo de integração, modelagem e visualização dos dados, garantindo consistência entre as etapas metodológicas adotadas no estudo.

4.5.1 Python

A linguagem Python[2] foi empregada como principal ferramenta para a implementação do processo de ETL devido à sua versatilidade e ao conjunto de bibliotecas voltadas para manipulação e análise de dados. Utilizando módulos como pandas, numpy e sqlalchemy, foi possível realizar a leitura dos microdados provenientes do SINAN, tratamento de inconsistências, padronização de valores categóricos, conversão de tipos, criação de novas variáveis derivadas e integração das informações necessárias para alimentar o *Data Warehouse*.

[2] Python Software Foundation. Python. Disponível em: https://www.python.org/ . Acesso em: 03 fev. 2026.

A estrutura modular do código permitiu a criação de funções reutilizáveis, garantindo maior organização, rastreabilidade e facilidade de manutenção do pipeline. Além disso, a integração nativa do Python com o PostgreSQL possibilitou que todo o fluxo, desde a extração até o carregamento final, fosse executado de forma automática e consistente.

[3]4.5.2 PostgreSQL

O PostgreSQL é um Sistema de Gerenciamento de Banco de Dados (SGBD), que nesse trabalho foi utilizado para armazenar dados na *staging Area* e no *Data Warehouse*. Trata-se de um SGBD, que atende aos padrões SQL e com capacidade de trabalhar com grandes volumes de dados, características importantes no desenvolvimento de um BI. Assim, a arquitetura do banco no PostgreSQL foi organizada em dois esquemas:

- *Staging Area*: responsável por armazenar os dados brutos importados diretamente das fontes, sem transformações, assegurando a preservação das informações originais.
- *Data Warehouse*: estruturado segundo o modelo dimensional em *Star Schema,* contendo tabelas fato e dimensão otimizadas para análises gerenciais.

Nos dois esquemas, o PostgreSQL permitiu a criação de chaves substitutas (*surrogate keys*), *constraints*, índices e mecanismos de integridade referencial essenciais para a confiabilidade do DW.

4.5.3 Microsoft Power BI

A ferramenta Microsoft Power BI[4] foi utilizada para a elaboração dos *dashboards* analíticos do estudo. Sua interface intuitiva, aliada à capacidade de conexão direta com o PostgreSQL, facilitou a construção de visualizações interativas capazes de

[3] PostgreSQL Global Development Group. PostgreSQL. Disponível em: https://www.postgresql.org/. Acesso em: 03 fev. 2026.
[4] Microsoft Corporation. Microsoft Power BI. Disponível em: https://powerbi.microsoft.com/ . Acesso em: 03 fev. 2026.

sintetizar as informações mais relevantes sobre o abandono de tratamento da tuberculose.

Com o Power BI foi possível:

- explorar séries temporais de indicadores;
- comparar municípios, regiões e faixas populacionais;
- identificar padrões epidemiológicos;
- avaliar correlações entre fatores sociodemográficos e desfechos clínicos;
- disponibilizar painéis de navegação clara e objetiva para apoio à tomada de decisão.

Sua ampla gama de recursos visuais e filtros avançados contribuiu significativamente para a produção de análises dinâmicas e aprofundadas.

5.RESULTADOS OBTIDOS

5.1 Consolidação dos Dados e Indicadores Analíticos

A partir da implementação do processo de Extração, Transformação e Carga (ETL), os dados epidemiológicos referentes aos casos de tuberculose foram integrados a informações socioeconômicas provenientes de diferentes bases públicas, incluindo o Sistema de Informação de Agravos de Notificação (SINAN-TB), o Instituto Brasileiro de Geografia e Estatística (IBGE) e o Atlas do Desenvolvimento Humano no Brasil. Esse processo possibilitou a consolidação de múltiplas fontes heterogêneas em um ambiente unificado de *Data Warehouse,* assegurando padronização, consistência e qualidade dos dados utilizados nas análises.

O *Data Warehouse* foi modelado segundo a abordagem dimensional em esquema estrela, tendo como elemento central a tabela fato fNotificacao, responsável por armazenar os registros de notificações de tuberculose e seus principais atributos analíticos. Essa tabela foi relacionada às dimensões dTempo e dLocal, conforme representado no modelo multidimensional, além das dimensões mascaradas. No total, foram consolidados 301.960 registros referentes às notificações de tuberculose no Brasil, compreendendo o período entre os anos de 2018 e 2022. Esse volume de dados possibilitou a construção de uma base histórica consistente, adequada para análises temporais e territoriais, bem como para a identificação de padrões relacionados ao abandono do tratamento.

A dimensão temporal (dTempo) permite a análise da evolução dos casos ao longo dos anos, trimestres, meses e datas de notificação, possibilitando a identificação de padrões e tendências temporais associadas ao abandono do tratamento.

A dimensão geográfica (dLocal) viabilizou a contextualização espacial dos registros por município, unidade federativa e região do país, além da incorporação de indicadores socioeconômicos como população, expectativa de vida, Índice de Desenvolvimento Humano Municipal (IDHM), índice de Gini, renda per capita e proporção de pobreza.

Além das dimensões estruturais, a tabela fato incorporou, por meio das dimensões mascaradas, atributos clínicos, sociodemográficos e assistenciais relevantes, como sexo, faixa etária, forma clínica da doença, presença de HIV/AIDS, consumo de álcool, diabetes, tempos relacionados ao diagnóstico, início e encerramento do tratamento, bem como a situação final dos casos.

A partir do ambiente analítico criado, foram desenvolvidas métricas e indicadores no Power BI, destacando-se o total de casos notificados, o total de abandonos e a taxa percentual de abandono do tratamento, além de indicadores segmentados por tempo, localização geográfica, perfil populacional, fatores clínicos e condições socioeconômicas.

Os indicadores foram organizados em painéis interativos (*dashboards*), permitindo a exploração dinâmica dos dados por meio de filtros temporais e territoriais. Dessa forma, a arquitetura desenvolvida não apenas proporcionou a visualização consolidada das informações, mas também viabilizou a identificação de padrões, tendências e relações relevantes entre as variáveis analisadas.

Os indicadores consolidados foram disponibilizados por meio de painéis analíticos interativos (*dashboards*), os quais permitem a exploração dinâmica das informações a partir de filtros temporais e territoriais. A arquitetura de *Business Intelligence* implementada possibilitou não apenas a visualização integrada dos dados, mas também a identificação de padrões, tendências e relações relevantes entre as variáveis analisadas.

Nos tópicos subsequentes, são apresentados os resultados obtidos por meio desses painéis, desenvolvidos em conformidade com os objetivos do trabalho, contemplando a análise da evolução temporal dos casos de tuberculose, a distribuição espacial do abandono do tratamento, o perfil sociodemográfico dos pacientes, os fatores clínicos associados e as condições socioeconômicas dos municípios.

5.2 Análise dos indicadores

Os *dashboards* desenvolvidos no ambiente de *Business Intelligence*, por meio da ferramenta Microsoft Power BI, foram concebidos com acesso direto ao *Data Warehouse*, de modo a integrar de forma consistente as diferentes dimensões analíticas do sistema. Essa estrutura possibilitou a exploração do abandono do tratamento da tuberculose sob múltiplas perspectivas, incluindo recortes temporais, geográficos, sociodemográficos, clínicos e socioeconômicos.

Cada painel foi elaborado com o objetivo de evidenciar indicadores específicos e relevantes ao contexto do estudo, permitindo tanto a visualização consolidada do comportamento dos casos quanto a realização de análises aprofundadas por meio de filtros interativos e segmentações analíticas. Os *dashboards* encontram-se disponibilizados na plataforma Power BI e podem ser acessados por meio do link: https://app.powerbi.com/groups/me/reports/ff2a9270-541a-4c38-9241-ffe2eed33b6a/48854c982decb2b81070?experience=power-bi.

O *dashboard* representado na Figura 4 apresenta uma visão geral dos casos de tuberculose notificados no Brasil e do comportamento do abandono do tratamento ao longo do período analisado. Observa-se um total aproximado de 301.960 casos registrados, dos quais cerca de 41.610 resultaram em abandono do tratamento, correspondendo a uma taxa média de aproximadamente 14%.

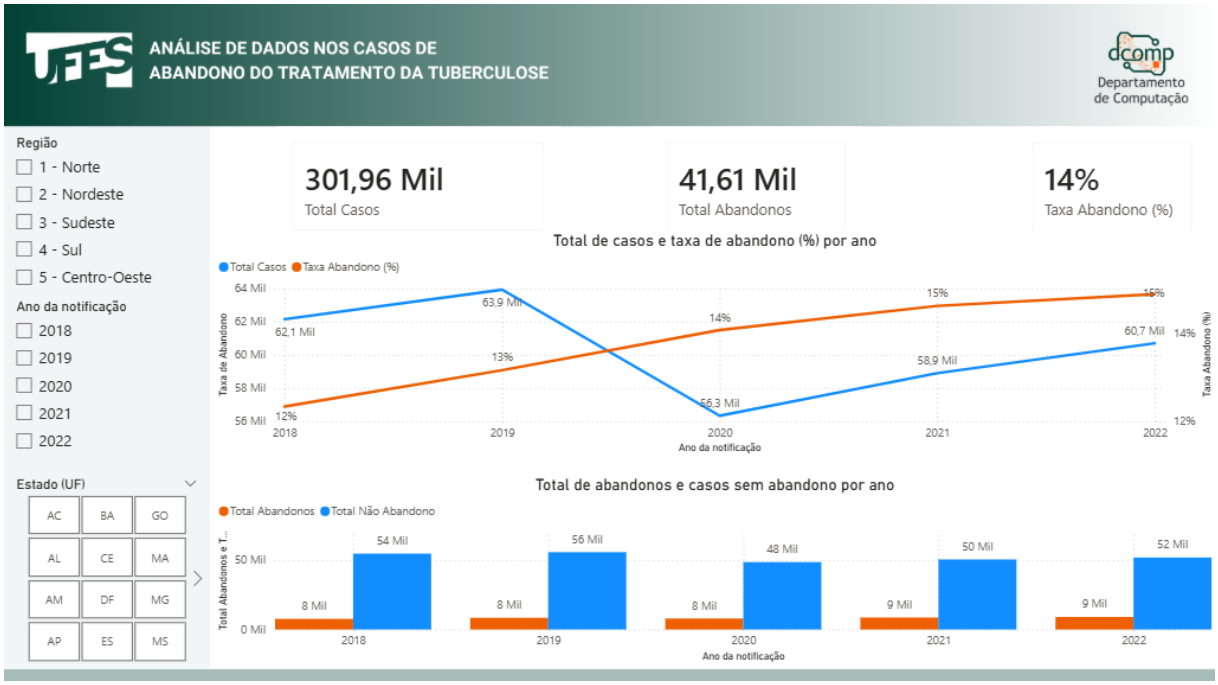

Figura 4 – Panorama geral dos casos e do abandono do tratamento.
Fonte: autor.

A análise temporal evidencia variações tanto no volume de casos quanto na taxa de abandono. Entre os anos de 2018 e 2019, verifica-se um crescimento no número de notificações, seguido por uma queda acentuada em 2020. Esse comportamento pode estar associado a fatores externos que impactaram o sistema de saúde, como a pandemia da COVID-19, que influenciou os processos de diagnóstico, notificação e acompanhamento de diversas doenças infecciosas.

A partir de 2021, observa-se uma retomada gradual no número de casos registrados, indicando uma recuperação dos mecanismos de vigilância epidemiológica. Em contrapartida, a taxa de abandono do tratamento apresenta uma tendência de crescimento ao longo dos anos, mesmo em períodos de redução no volume total de casos. Esse aumento progressivo sugere que a adesão terapêutica tem se configurado como um desafio persistente no contexto nacional.

O gráfico comparativo apresentado na Figura 4, entre o total de abandonos e os casos com desfecho diferente de abandono, reforça a permanência de um contingente significativo de interrupções do tratamento ao longo do período analisado, demonstrando que, embora a maioria dos pacientes complete o tratamento, o contingente de indivíduos que interrompem o acompanhamento permanece significativo e relativamente constante ao longo do tempo.

De forma geral, os resultados indicam que o abandono do tratamento da tuberculose continua representando um importante problema de saúde pública, exigindo estratégias direcionadas para fortalecer o acompanhamento dos pacientes e reduzir a interrupção terapêutica.

O *dashboard* representado na Figura 5 apresenta a distribuição geográfica do abandono do tratamento da tuberculose nos municípios brasileiros, permitindo identificar áreas com maior concentração de casos e padrões regionais.

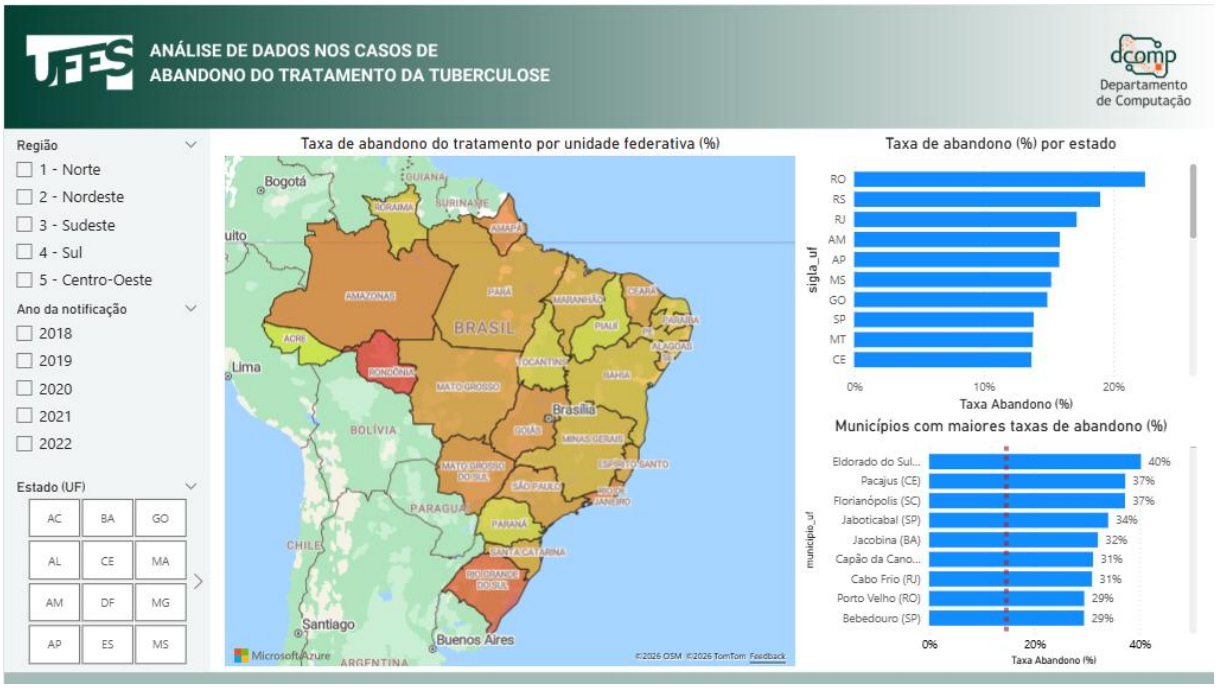

Figura 5 – Distribuição geográfica do abandono do tratamento.
Fonte: autor.

O mapa temático evidencia uma heterogeneidade significativa entre as unidades federativas. Observa-se que determinados estados, especialmente nas regiões Norte e Sul, apresentam taxas de abandono superiores à média nacional, enquanto outros exibem percentuais relativamente menores. Essa variação espacial indica que o abandono do tratamento não ocorre de forma homogênea no país, sendo influenciado por fatores regionais relacionados às condições socioeconômicas, à organização dos serviços de saúde e às características populacionais.

O gráfico de barras que apresenta a taxa de abandono por unidade federativa reforça essa discrepância, destacando estados como Rondônia, Rio Grande do Sul, Roraima, Amazonas e Amapá entre aqueles com os maiores percentuais de abandono.

Para a elaboração do ranking dos municípios com maiores taxas de abandono, foi aplicado um filtro que considerou apenas localidades com total de casos maior ou igual a 50 registros, com o objetivo de minimizar distorções estatísticas decorrentes de amostras reduzidas. Essa estratégia metodológica assegura maior confiabilidade na interpretação dos percentuais observados.

A partir do critério de inclusão de municípios com pelo menos 50 registros observa-se que municípios como Eldorado do Sul, Pacajus, Florianópolis e Jaboticabal apresentam taxas significativamente superior à média nacional, representada no gráfico por uma linha de referência.

Os resultados do ranking municipal baseado na taxa de abandono evidenciam áreas críticas distribuídas em diferentes regiões do país, reforçando o caráter nacional do problema, ainda que com maior intensidade em determinados territórios.

O *dashboard* representado na Figura 6, apresenta a distribuição da taxa de abandono do tratamento segundo características sociodemográficas, considerando as variáveis sexo, faixa etária, raça/cor e nível de escolaridade.

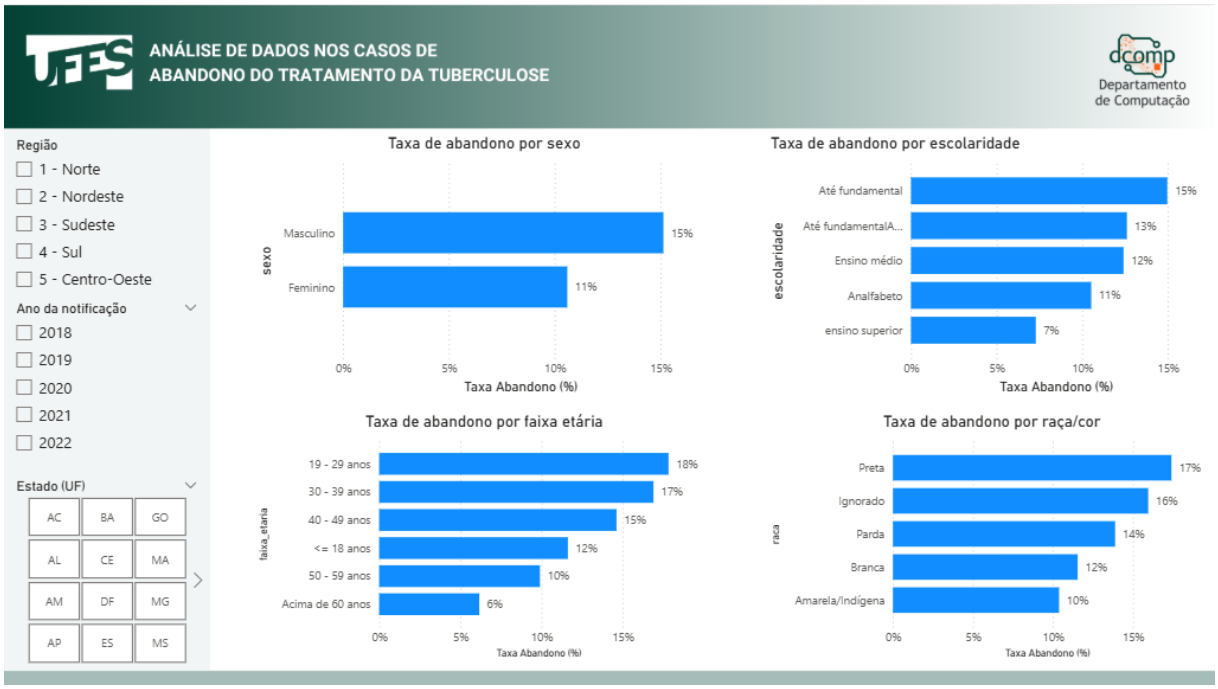

Figura 6 – Perfil sociodemográfico dos pacientes.
Fonte: autor.

Os resultados indicam que indivíduos do sexo masculino apresentam maiores taxas de abandono do tratamento quando comparados ao sexo feminino, o que pode estar associado a comportamentos de menor busca por serviços de saúde, bem como a fatores sociais e ocupacionais que dificultam a continuidade do acompanhamento terapêutico.

A análise segundo a faixa etária demonstra que os percentuais mais elevados de abandono concentram-se entre adultos jovens, especialmente nos grupos de 19 a 29 anos e de 30 a 39 anos. Observa-se ainda uma tendência de redução progressiva dessas taxas com o avanço da idade, sendo os menores índices registrados entre indivíduos com mais de 60 anos.

No que se refere à variável raça/cor, verifica-se que pessoas autodeclaradas pretas e pardas apresentam maiores percentuais de abandono em comparação aos demais grupos populacionais. Por outro lado, indivíduos classificados como brancos e amarelos/indígenas apresentam taxas inferiores, o que pode refletir desigualdades sociais e de acesso aos serviços de saúde.

Quanto ao nível de escolaridade, os maiores percentuais de abandono estão associados a indivíduos com menor grau de instrução, especialmente aqueles com escolaridade até o ensino fundamental. Em contraste, pessoas com ensino superior apresentam as menores taxas de abandono, evidenciando uma relação inversa entre escolaridade e adesão ao tratamento.

De forma geral, os resultados indicam que o abandono do tratamento da tuberculose está fortemente associado a características sociodemográficas específicas, reforçando a necessidade de estratégias de intervenção direcionadas a grupos mais vulneráveis.

O *dashboard* apresentado na Figura 7 evidencia a variação da taxa de abandono do tratamento da tuberculose de acordo com diferentes fatores clínicos e sociais, possibilitando a identificação de condições associadas a maiores níveis de interrupção terapêutica. A análise permite compreender quais aspectos apresentam maior influência no abandono do tratamento, contribuindo para o direcionamento de estratégias de intervenção mais eficazes.

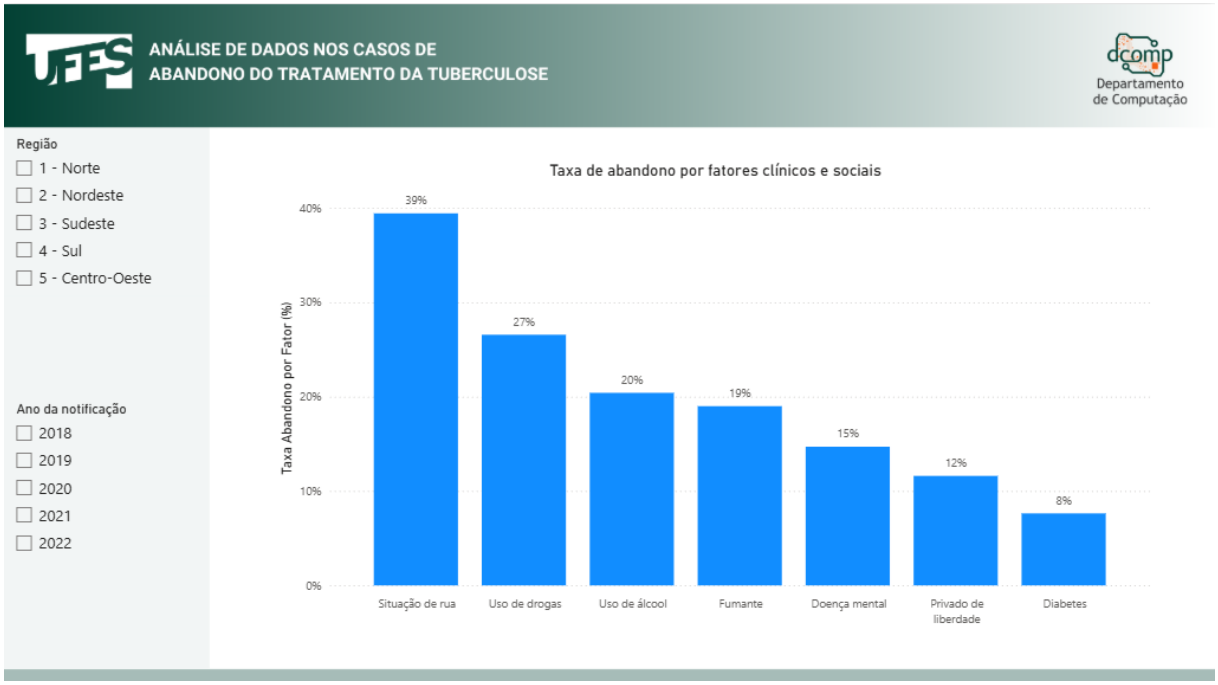

Figura 7 – *Dashboard* 4 – Fatores clínicos e sociais associados ao abandono.
Fonte: autor.

Observa-se, na Figura 7, que indivíduos em situação de rua apresentam a maior taxa de abandono, indicando barreiras significativas no acesso contínuo aos serviços de saúde, além de dificuldades relacionadas à moradia, alimentação e acompanhamento clínico. Em seguida, destacam-se os fatores relacionados ao uso de drogas e ao consumo de álcool, que apresentam taxas elevadas de abandono. Tais comportamentos encontram-se frequentemente associados a contextos de exclusão social e fragilização do vínculo com os serviços de saúde. Por conseguinte, o tabagismo também se mostra associado a percentuais elevados de abandono, enquanto indivíduos com doença mental apresentam taxas intermediárias, refletindo desafios adicionais no acompanhamento desses pacientes.

Entre os fatores analisados, a condição de encarceramento apresenta taxa moderada de abandono, enquanto a presença de diabetes apresenta o menor percentual observado.

Os dados indicam que determinantes sociais, especialmente aqueles relacionados à exclusão social e comportamentos de risco, apresentam maior associação com o abandono do tratamento do que condições clínicas isoladas, reforçando a importância

de políticas públicas voltadas à redução das vulnerabilidades sociais no controle da tuberculose.

A Figura 8 apresenta o *dashboard* voltado à análise das condições socioeconômicas municipais e sua relação com a taxa de abandono do tratamento da tuberculose. Para essa finalidade, foram utilizados indicadores de desenvolvimento humano (IDHM), proporção de pobreza, renda per capita e desigualdade de renda (Índice de Gini), organizados em faixas categóricas, de modo a facilitar a interpretação comparativa dos resultados.

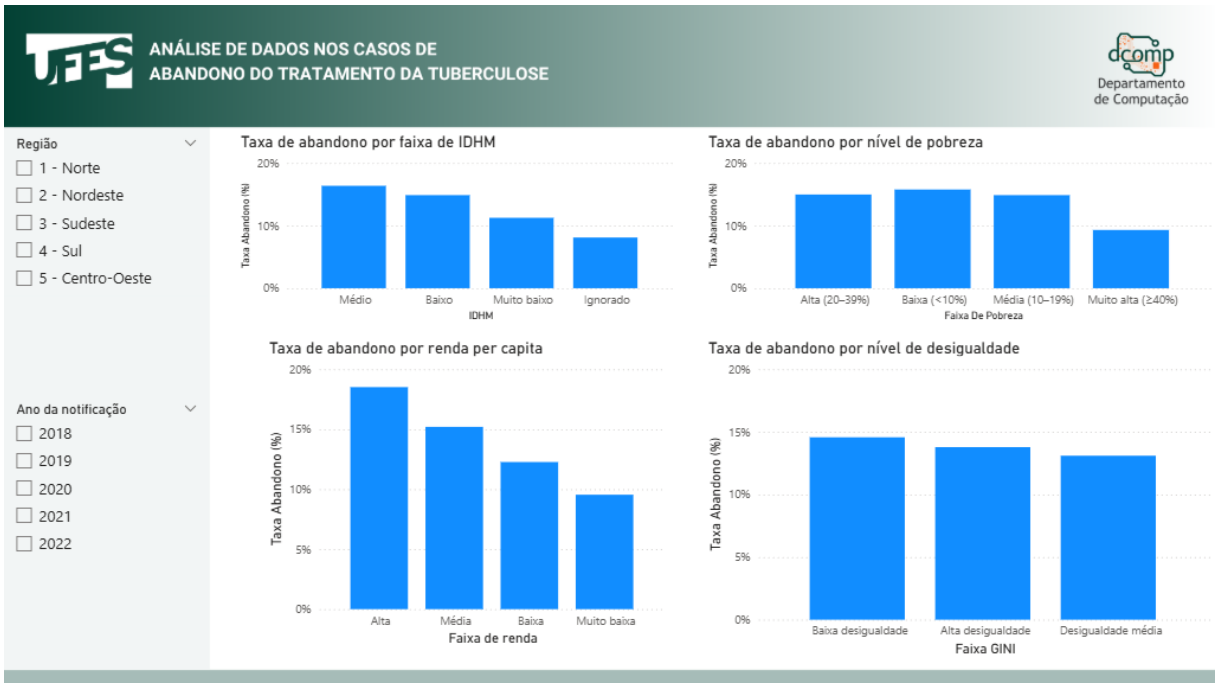

Figura 8 – *Dashboard* 5 – Condições socioeconômicas associadas ao abandono.
Fonte: autor.

No que se refere ao Índice de Desenvolvimento Humano Municipal (IDHM), observa-se variação nas taxas de abandono entre as faixas analisadas. Municípios classificados nas categorias de desenvolvimento médio e baixo apresentam percentuais mais elevados de abandono, enquanto a faixa de IDHM muito baixo e os registros com informação ignorada exibem valores inferiores. Esse comportamento pode estar parcialmente associado a limitações estruturais em territórios mais vulneráveis, como dificuldades no acompanhamento dos casos e fragilidades no registro adequado dos desfechos, o que pode impactar o cálculo das taxas observadas.

Em relação à proporção de pobreza, a segmentação por faixas indica percentuais de abandono relativamente próximos entre as categorias baixa, média e alta, sugerindo que o abandono permanece expressivo mesmo em contextos socioeconômicos intermediários. Por outro lado, a faixa de pobreza muito alta apresenta percentuais inferiores, resultado que pode estar relacionado a efeitos como subnotificação, menor encerramento formal dos casos e fragilidades na continuidade do acompanhamento em municípios com maior vulnerabilidade social.

A análise por faixas de renda per capita evidencia uma tendência de redução da taxa de abandono à medida que se avança para categorias de menor renda, com diferenças visíveis entre os grupos apresentados. Embora essa relação não se apresente de forma estritamente linear em todos os casos, o comportamento observado reforça que a dinâmica do abandono é influenciada por múltiplos fatores estruturais, incluindo acesso aos serviços de saúde, vínculo com a rede de atenção e condições socioeconômicas locais.

Por fim, a distribuição por faixas do Índice de Gini, indicador de desigualdade de renda, aponta variações discretas entre as categorias de desigualdade baixa, média e alta. Esses resultados sugerem que a desigualdade, de forma isolada, não explica diretamente o abandono do tratamento, indicando que a interpretação do fenômeno deve considerar o conjunto de determinantes sociais e organizacionais presentes nos municípios analisados.

De forma geral, os resultados do dashboard representado na Figura 8, demonstram que o abandono do tratamento da tuberculose apresenta associação com diferentes dimensões socioeconômicas, embora não configure uma relação simples e linear com os indicadores analisados. Os achados reforçam a necessidade de estratégias de enfrentamento que considerem, além das condições de extrema vulnerabilidade, os contextos intermediários, nos quais as taxas de abandono também se mantêm relevantes.

6. CONCLUSÃO

O *Business Intelligence* (BI) desempenha um papel essencial no contexto da saúde pública, ao possibilitar a integração, análise e visualização de grandes volumes de dados provenientes de diferentes fontes. Por meio de ferramentas analíticas e estruturas como o *Data Warehouse*, o BI permite a identificação de padrões, tendências e áreas críticas, contribuindo para a tomada de decisões mais eficientes, o planejamento de ações estratégicas e o monitoramento contínuo de doenças. Dessa forma, sua aplicação representa um importante avanço no apoio à gestão em saúde, promovendo maior eficiência nos serviços e melhores resultados para a população.

O presente estudo adotou metodologias de BI com o objetivo de desenvolver um *Data Warehouse* capaz de integrar informações epidemiológicas e socioeconômicas relacionadas aos casos de tuberculose no Brasil, com foco específico na análise do abandono do tratamento. A solução implementada possibilitou consolidar dados provenientes de sistemas públicos de saúde e bases socioeconômicas, permitindo uma análise multidimensional do fenômeno sob perspectivas temporais, geográficas, clínicas e sociais. Embora o estudo tenha sido desenvolvido em âmbito nacional, a abordagem proposta pode ser facilmente adaptada para análises regionais ou municipais, bem como para a investigação de outras doenças de interesse em saúde pública.

Durante o período analisado, observou-se que o abandono do tratamento da tuberculose apresenta comportamento persistente ao longo dos anos, com tendência de crescimento da taxa de abandono, mesmo em cenários de oscilação no número total de casos notificados. A análise espacial evidenciou variações significativas entre estados e municípios, destacando áreas com taxas superiores à média nacional, o que reforça a existência de desigualdades regionais na adesão ao tratamento.

Adicionalmente, os resultados apontaram diferenças relevantes no abandono do tratamento de acordo com o perfil sociodemográfico dos pacientes. Observou-se maior taxa de abandono entre indivíduos do sexo masculino, adultos jovens, pessoas com menor nível de escolaridade e populações negras e pardas. Esses achados evidenciam a influência de determinantes sociais no comportamento de adesão ao

tratamento, indicando a necessidade de estratégias específicas voltadas a grupos mais vulneráveis.

Outro aspecto relevante identificado foi a forte associação entre o abandono do tratamento e fatores sociais e comportamentais, como a situação de rua, o uso de drogas, o consumo de álcool e o tabagismo. Esses fatores apresentaram taxas de abandono significativamente superiores às demais condições analisadas, demonstrando o impacto de contextos de vulnerabilidade social na continuidade do tratamento. Além disso, a análise socioeconômica revelou que municípios com piores indicadores de desenvolvimento humano, renda e maior desigualdade social tendem a apresentar taxas de abandono mais elevadas, reforçando a relação entre condições estruturais e dificuldades na adesão terapêutica.

Em síntese, o presente trabalho foi capaz de fornecer uma visão abrangente sobre o abandono do tratamento da tuberculose no Brasil por meio da aplicação de técnicas de *Business Intelligence* e análise de dados integrados em um *Data Warehouse*. A solução desenvolvida permitiu identificar padrões, tendências e áreas críticas, oferecendo uma base consistente para o apoio à tomada de decisão em saúde pública. Dessa forma, os resultados obtidos possuem potencial para auxiliar gestores e profissionais de saúde na formulação de políticas públicas mais direcionadas e eficazes no enfrentamento da tuberculose, contribuindo para a redução do abandono do tratamento e para a melhoria dos indicadores de saúde no país.

Como perspectivas futuras, destaca-se a ampliação do período analisado e a integração com novas bases de dados, possibilitando análises mais abrangentes e atualizadas sobre o abandono do tratamento da tuberculose. Além disso, a incorporação de técnicas de análise preditiva, por meio de métodos de aprendizado de máquina, pode contribuir para a identificação antecipada de padrões associados ao risco de interrupção terapêutica. Por fim, a arquitetura analítica desenvolvida neste trabalho apresenta potencial para ser utilizada como ferramenta de monitoramento contínuo, apoiando ações de vigilância epidemiológica e o acompanhamento sistemático dos indicadores em saúde pública.

7. REFERÊNCIAS

ATLAS BRASIL. Atlas Brasil. Disponível em: http://www.atlasbrasil.org.br/acervo/atlas. Acesso em: 18 jul. 2023.

BARREIRA, D.; GRANGEIRO, A. Avaliação das estratégias de controle da tuberculose no Brasil. *Revista de Saúde Pública*, v. 41, supl. 1, p. 4–8, 2007.

BRASIL. DATASUS – Ministério da Saúde. Disponível em: https://datasus.saude.gov.br/. Acesso em: 18 jul. 2025.

BRASIL. MINISTÉRIO DA SAÚDE. Manual de Recomendações para o Controle da Tuberculose no Brasil. 2. ed. Brasília: Ministério da Saúde, 2019. Disponível em: https://nepas.ufsc.br/files/2012/09/Manual-de-Recomendações-para-o-Controle-da-Tuberculose-no-Brasil.pdf. Acesso em 4 jan. 2026.

BRASIL. MINISTÉRIO DA SAÚDE. Boletim Epidemiológico Tuberculose 2025. Brasília, 2025. Disponível em: https://www.gov.br/aids/pt-br/.../boletim-epidemiologico-tuberculose-2025. Acesso em 4 jan. 2026.

BRASIL. MINISTÉRIO DA SAÚDE. *Adesão ao tratamento de tuberculose pela população em situação de rua*. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/sintese_evidencias_politicas_adesao_tuberculose.pdf. Acesso em: 4 jul. 2025.

BRASIL. SINAN. Disponível em: https://portalsinan.saude.gov.br/. Acesso em: 18 jul. 2025.

DRESNER, H. *The performance management revolution: business results through insight and action*. Hoboken: Wiley, 2008.

FIOCRUZ. Como é realizado o tratamento da tuberculose? Disponível em: https://portal.fiocruz.br/pergunta/como-e-realizado-o-tratamento-da-tuberculose. Acesso em: 4 jul. 2025

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). *Censo Demográfico 2022: resultados preliminares*. Rio de Janeiro: IBGE, 2023. Disponível em: https://www.ibge.gov.br. Acesso em: 15 fev 2026.

INMON, W. H.; HACKATHORN, R. D. Using the Data Warehouse. New York: John Wiley & Sons, 1994.

KESKIN, S.; YAZICI, A. Modeling and querying a fuzzy SOLAP-based framework. *ISPRS International Journal of Geo-Information*, v. 11, n. 3, p. 191, 2022.

KIMBALL, R.; ROSS, M. *The data warehouse toolkit: the definitive guide to dimensional modeling*. Hoboken: Wiley, 2013.

KOHARA MELCHIOR, L. A. et al. Business Intelligence no apoio à gestão estratégica em saúde: um relato de experiência. *Revista de Epidemiologia e Controle de Infecção*, v. 12, n. 4, p. 180–183, 2022.

LUCENA, L. A. et al. Factors associated with the abandonment of tuberculosis treatment in Brazil: a systematic review. *Revista da Sociedade Brasileira de Medicina Tropical*, v. 56, e0155-2022, 2023.

MACHADO, F. N. R. *Projeto de Data Warehouse: uma visão multidimensional*. [S.l.: s.n.].

MENDONÇA, Jéssica Simões; SILVEIRA, Fabrício Sette Abrantes; COLODETTE, Renata Maria; et al. Epidemiological, socioeconomic, and health service factors associated with tuberculosis treatment interruption in Brazil. *Epidemiologia*, v. 6_toggle, n. 4, 2025. Disponível em: https://doi.org/10.3390/epidemiologia6040081. Acesso em: 14 fev. 2026.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. Progresso global no combate à tuberculose está em risco. Disponível em: https://www.paho.org/pt/noticias/14-10-2020-progresso-global-no-combate-tuberculose-esta-em-risco-afirma-oms. Acesso em: 26 jun. 2023.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. Mortes por tuberculose aumentam pela primeira vez em mais de uma década. Disponível em: https://www.paho.org/pt/noticias/14-10-2021-mortes-por-tuberculose-aumentam-pela-primeira-vez-em-mais-uma-decada-devido. Acesso em: 26 jun. 2023.

PELISSARI, D. M. et al. Identifying socioeconomic, epidemiological and operational scenarios for tuberculosis control in Brazil: an ecological study. *BMJ Open*, v. 8, n. 6, e018545, 2018.

PROGRAMA DAS NAÇÕES UNIDAS PARA O DESENVOLVIMENTO (PNUD); INSTITUTO DE PESQUISA ECONÔMICA APLICADA (IPEA); FUNDAÇÃO JOÃO PINHEIRO (FJP). *Atlas do Desenvolvimento Humano no Brasil*. Brasília, 2020. Disponível em: http://www.atlasbrasil.org.br. Acesso em: 15 fev 2026.

SCHWAMBACH, C. J. et al. Transformando dados em informação: aplicação de business intelligence para a automação da análise de dados em saúde. *Brazilian Journal of Production Engineering*, v. 10, n. 3, p. 234–248, 2024.

SILVA, É. A. et al. Perfil dos casos de abandono do tratamento da tuberculose em um município prioritário mineiro. *HU Revista*, v. 44, n. 3, p. 351–360, 2019.

TAVERA ROMERO, C. A. et al. Business intelligence: business evolution after Industry 4.0. *Sustainability*, v. 13, n. 18, p. 10026, 2021.

TORRES, D. R. et al. Aplicabilidade e potencialidades no uso de ferramentas de Business Intelligence na Atenção Primária em Saúde. *Ciência & Saúde Coletiva*, v. 26, n. 6, p. 2065–2074, 2021.

TREVISAN, I. H. *Modelo generalizado para projetos de Business Intelligence*. 2022. 57 f. Monografia (Pós-Graduação em Gestão de Projetos) – Faculdade SENAI Mario Amaro, São Bernardo do Campo, 2022.

WORLD HEALTH ORGANIZATION. *Global Tuberculosis Report 2025*. Disponível em: https://www.who.int/teams/global-programme-on-tuberculosis-and-lung-health/tb-reports/global-tuberculosis-report-2025. Acesso em: 5 jan. 2026.

WORLD HEALTH ORGANIZATION. WHO Coronavirus (COVID-19) Dashboard. Disponível em: https://covid19.who.int. Acesso em: 7 jul. 2023.

8. ANEXOS:

Anexo A - Dicionário de Dados original, recebido juntamente com a com a base de dados. (Fonte: SINAN/IBGE)

Campo	Descrição
Entidade	Registro de Atendimento de Saúde
Descrição	Tabela que armazena os dados sociodemográficos, clínicos e de acompanhamento de um atendimento ou notificação de saúde.
Nome da tabela	tb_atendimento

Atributo	Nome do campo	Descrição
Sexo	sexo	Sexo do paciente (1: Feminino, 2: Masculino, 3: Ignorado)
Faixa Etária	faixa_etaria	Faixa etária do paciente (1: ≤18, 2: 19–29, 3: 30–39, 4: 40–49, 5: 50–59, 6: ≥60)
Raça/Cor	raca	Raça/cor autodeclarada (1: Branca, 2: Preta, 3: Amarela/Indígena, 4: Parda, 9: Ignorado, 10: Omissão)
Escolaridade	escolaridade	Escolaridade do paciente (0: Analfabeto, 1: Fundamental, 2: Médio, 3: Superior, 8: Não se aplica, 9: Ignorado, 10: Omissão)
Raio-X de Tórax	raiox_tora	Resultado do exame (1: Suspeito, 2: Normal, 3: Outra patologia, 4: Não realizado)

Forma Clínica	forma	Forma clínica (1: Pulmonar, 3: Pulmonar + Extrapulmonar)
AIDS	agravoaids	Presença de AIDS (1: Sim, 2: Não, 3: Ignorado)
Uso de álcool	agravoalcoo	Uso abusivo de álcool (1: Sim, 2: Não, 3: Ignorado)
Diabetes	agravodiabe	Presença de diabetes (1: Sim, 2: Não, 3: Ignorado)
Doença mental	agravodoenc	Presença de doença mental (1: Sim, 2: Não, 3: Ignorado)
Uso de drogas	agravdroga	Uso de drogas ilícitas (1: Sim, 2: Não, 9: Ignorado)
Tabagismo	agravtabac	Fumante (1: Sim, 2: Não, 9: Ignorado)
Baciloscopia	bacilosc_e	Resultado (1: Positivo, 2: Negativo, 3: Não realizado, 4: Não se aplica)
Cultura de escarro	cultura_es	Resultado (1: Positivo, 2: Negativo, 3: Em andamento, 4: Não realizado)
Teste HIV	hiv	Resultado (1: Positivo, 2: Negativo, 3: Em andamento, 4: Não realizado)
Tratamento supervisionado	tratamento_supervisionado	Indica TDO (1: Sim, 2: Não, 8: Omissão, 9: Ignorado)
Benefício do governo	beneficio_governo	Recebe benefício (1: Sim, 2: Não, 8: Omissão, 9: Ignorado)
População privada de liberdade	pop_liber	Indica se é encarcerado (1: Sim, 2: Não, 9: Ignorado)
População em situação de rua	pop_rua	Indica se vive em rua (1: Sim, 2: Não, 9: Ignorado)
Profissional da saúde	profissional_saude	Indica se é profissional da saúde

População imigrante	pop_imig	Indica se é imigrante
Teste molecular	test_molec	Resultado do TRM
Encerramento por abandono	encerramento_abandono	0: Não, 1: Sim
Tipo de notificação	tp_not	1: Negativa, 2: Individual, 3: Surto, 4: Agregado
Agravo	id_agravo	Código CID-10 (tuberculose)
UF notificação	sg_uf_not	UF da unidade notificadora
Município notificação	id_municip	Município da unidade notificadora
Regional de saúde	id_regiona	Regional da unidade
UF residência	sg_uf	UF do paciente
Município residência	id_mn_resi	Município do paciente
País	id_pais	País de residência
Data notificação	dt_notific	Data da notificação
Ano	nu_ano	Ano da notificação
Data diagnóstico	dt_diag	Data de diagnóstico
Data encerramento	dt_encerra	Data de encerramento
Situação encerramento	situa_ence	1: Cura, 2: Abandono, 3: Óbito TB, etc
Teste sensibilidade	test_sensi	Resultado do teste

ANEXO B → dicionário limpo e analítico do DW

Campo	Descrição
Entidade	Fato Notificação
Descrição	Armazena eventos e casos notificados (SINAN) com indicadores relacionados ao abandono de tratamento e variáveis clínicas e sociais relevantes para análise.
Nome da tabela	fato_notificacao

Atributo	Nome do campo	Tipo de dado	Restrição	Descrição
Identificador do fato	ID_FATO	BIGINT	PK	Identificador único do registro na tabela fato (recomendado BIGINT pelo volume potencial).
Chave tempo (notificação)	ID_TEMPO_NOTIFICACAO	INTEGER	FK, NOT NULL	Referência à DM_TEMPO(ID_TEMPO) para análises temporais.
Chave local (notificação)	ID_LOCAL_NOTIFICACAO	INTEGER	FK, NOT NULL	Referência à DM_LOCAL(ID_LOCAL) do município onde ocorreu a notificação.
Chave local (residência)	ID_LOCAL_RESIDENCIA	INTEGER	FK, NOT NULL	Referência à DM_LOCAL(ID_LOCAL) do município de residência do paciente.

Regional (notificação)	ID_REGIONAL_NOTIFICACAO	INTEGER	opcional	Identificador da regional de saúde associada ao município de notificação.
Regional (residência)	ID_REGIONAL_RESIDENCIA	INTEGER	opcional	Identificador da regional de saúde associada ao município de residência.
Tipo de notificação	TIPO_NOTIFICACAO	SMALLINT	domínio (dataset)	Tipo do registro conforme codificação do dataset.
Tratamento (tipo/situação)	TRATAMENTO	SMALLINT	domínio (SINAN)	Situação/tipo do caso conforme codificação do SINAN.
Tratamento supervisionado (TDO)	SUPERVISIONADO_NO_TRATAMENTO	SMALLINT	1,2,8,9	Indica TDO: 1=Sim; 2=Não; 8=Omisso; 9=Ignorado.
Encerramento por abandono	ENCERRAMENTO_ABANDONO	SMALLINT	0/1, NOT NULL	Indicador binário: 0=Não; 1=Sim.
Tempo (notificação→diagnóstico)	TEMPO_ENTRE_NOTIFICACAO_E_DIAGNOSTICO	INTEGER	CHECK >= 0	Diferença em dias entre data de notificação e diagnóstico.
Tempo (início tratamento→notificação)	TEMPO_ENTRE_INICIO_TRATAMENTO_E_NOTIFICACAO	INTEGER	CHECK >= 0	Diferença em dias entre início do tratamento e notificação.

Tempo (encerramento→início tratamento)	TEMPO_ENTRE_ENCERRAMENTO_E_INICIO_TRATAMENTO	INTEGER	opcional	Diferença em dias entre encerramento e início do tratamento (depende de preenchimento).
Sexo	SEXO	CHAR	IN ('M','F','I')	Sexo do paciente conforme codificação do dataset/SINAN.
Faixa etária	FAIXA_ETARIA	SMALLINT	IN (1..6)	Faixa etária categorizada no projeto.
Raça/Cor	RACA	SMALLINT	domínio (SINAN)	Raça/cor conforme codificação SINAN.
Escolaridade	ESCOLARIDADE	SMALLINT	domínio (SINAN)	Escolaridade conforme codificação SINAN.
Gestante	GESTANTE	SMALLINT	domínio (SINAN)	Situação gestacional/trimestre/não se aplica/ignorado conforme codificação.
Forma clínica	FORMA_CLINICA	SMALLINT	domínio (SINAN)	Forma clínica da TB conforme codificação SINAN.
Extrapulmonar 1	EXTRAPULMONAR_1	SMALLINT	domínio (SINAN)	Localização extrapulmonar 1 (quando aplicável).
Extrapulmonar 2	EXTRAPULMONAR_2	SMALLINT	domínio (SINAN)	Localização extrapulmonar 2 (quando aplicável).
Situação de encerramento	SITUACAO_ENCERRAMENTO	SMALLINT	domínio (SINAN)	Situação final (cura, abandono, óbito, transferência etc.).

Raio-X de tórax	RAIO_X_TORAX	SMALLINT	domínio (SINAN)	Resultado do RX de tórax conforme codificação.
Teste HIV	TESTE_HIV	SMALLINT	domínio (SINAN)	Resultado do teste de HIV conforme codificação.
Teste rápido molecular (TRM-TB)	RESULTADO_TESTE_RAPIDO	SMALLINT	domínio (SINAN)	Resultado do TRM-TB conforme codificação.
Teste de sensibilidade	TESTE_SENSIBILIDADE	SMALLINT	domínio (SINAN)	Resultado do teste de sensibilidade conforme codificação.
Histopatológico	HISTORICO_PATOLOGICO	SMALLINT	domínio (SINAN)	Resultado histopatológico conforme codificação.
AIDS	AIDS	SMALLINT	IN (1,2,3)	1=Sim; 2=Não; 3=Ignorado.
Uso de álcool	ALCOOL	SMALLINT	IN (1,2,3)	1=Sim; 2=Não; 3=Ignorado.
Diabetes	DIABETES	SMALLINT	IN (1,2,3)	1=Sim; 2=Não; 3=Ignorado.
Doença mental	DOENCA_MENTAL	SMALLINT	IN (1,2,3)	1=Sim; 2=Não; 3=Ignorado.
Uso de drogas	DROGAS	SMALLINT	IN (1,2,9)	1=Sim; 2=Não; 9=Ignorado.
Tabagismo	TABAGISMO	SMALLINT	IN (1,2,9)	1=Sim; 2=Não; 9=Ignorado.
Benefício do governo	BENEFICIO_GOVERNO	SMALLINT	IN (1,2,8,9)	1=Sim; 2=Não; 8=Omisso; 9=Ignorado.
População em situação de rua	SEM_TETO	SMALLINT	IN (1,2,9)	1=Sim; 2=Não; 9=Ignorado.
Pop. privada de liberdade	ENCARCERADO	SMALLINT	IN (1,2,9)	1=Sim; 2=Não; 9=Ignorado.

População imigrante	IMIGRANTE	SMALLINT	IN (1,2,9)	1=Sim; 2=Não; 9=Ignorado.
Profissional de saúde	PROFISSIONAL_SAUDE	SMALLINT	IN (1,2,8,9)	1=Sim; 2=Não; 8=Omisso; 9=Ignorado.

Campo	Informação
Entidade	Dimensão Local
Descrição	Armazena características geográficas e socioeconômicas do município e da unidade federativa, integrando dados territoriais do IBGE e indicadores socioeconômicos do Atlas.
Nome da tabela	dm_local

Atributo	**Nome do campo**	**Tipo de dado**	**Tamanho**	**Restrição**
Identificador da localidade	ID_LOCAL	INTEGER	—	PK
Código UF (numérico)	CODIGO_UF	SMALLINT	—	NOT NULL
Código do município	ID_MUNICIPIO	INTEGER	—	NOT NULL
Município/UF (texto)	MUNICIPIO_UF	VARCHAR	120	opcional
População (IBGE)	POPULACAO	INTEGER	—	CHECK >= 0
Faixa populacional	FAIXA_POPULACAO	VARCHAR	40	opcional

Sigla UF	SIGLA_UF	CHAR	2	NOT NULL
Código UF (variação)	COD_UF	SMALLINT	—	opcional
Região	REGIAO	VARCHAR	20	opcional
Município (identificador textual)	ID_MUNICIPIO_NOME	VARCHAR	120	opcional
Expectativa de vida	EXPECTATIVA_VIDA	NUMERIC	(5,2)	CHECK >= 0
Índice de Gini	INDICE_GINI	NUMERIC	(5,4)	CHECK 0–1
Proporção de pobreza	PROP_POBREZA	NUMERIC	(6,2)	CHECK 0–100
Renda per capita	RENDA_PC	NUMERIC	(12,2)	CHECK >= 0
População (Atlas)	POPULACAO_ATLAS	INTEGER	—	CHECK >= 0
IDHM	IDHM	NUMERIC	(4,3)	CHECK 0–1

Campo	Informação
Entidade	Dimensão Tempo
Descrição	Armazena atributos derivados de datas de notificação para análises temporais, como ano, mês e trimestre.
Nome da tabela	dm_tempo

Atributo	Nome do campo	Tipo de dado	Tamanho	Restrição
Identificador do tempo	ID_TEMPO	INTEGER	—	PK
Data de notificação	DATA_NOTIFICACAO	DATE	—	UNIQUE, NOT NULL
Dia	DIA_NOTIFICACAO	SMALLINT	—	CHECK 1–31, NOT NULL
Mês	MES_NOTIFICACAO	SMALLINT	—	CHECK 1–12, NOT NULL
Ano	ANO_NOTIFICACAO	SMALLINT	—	NOT NULL
Trimestre	TRIMESTRE_NOTIFICACAO	SMALLINT	—	CHECK 1–4, NOT NULL