



**UNIVERSIDADE FEDERAL DE SERGIPE CENTRO DE CIÊNCIAS EXATAS E
TECNOLOGIA**

DEPARTAMENTO DE ESTATÍSTICA E CIÊNCIAS ATUARIAIS



Jacinto Michael Menezes Cunha

**ANÁLISE E PREVISÃO DA TAXA DE OCUPAÇÃO HOSPITALAR EM
UM HOSPITAL PRIVADO DE SERGIPE**

São Cristóvão – SE

2026

Jacinto Michael Menezes Cunha

**Análise e Previsão da Taxa de Ocupação Hospitalar em um Hospital Privado
de Sergipe**

**Trabalho de Conclusão de Curso apresentado ao
Departamento de Estatística e Ciências Atuariais
da Universidade Federal de Sergipe, como parte
dos requisitos para obtenção do grau de Bacharel
em Ciências Atuariais.**

**Orientador (a): Prof. Dr. Luiz Henrique Gama Dore de
Araujo**

São Cristóvão – SE

2026

Jacinto Michael Menezes Cunha

**Análise e Previsão da Taxa de Ocupação Hospitalar em um Hospital Privado
de Sergipe**

**Trabalho de Conclusão de Curso apresentado ao
Departamento de Estatística e Ciências Atuariais
da Universidade Federal de Sergipe, como um dos
pré-requisitos para obtenção do grau de Bacharel
em Ciências Atuariais.**

Aprovado em ____/____/____, Nota Final_____.

Banca Examinadora:

Prof. Dr. Luiz Henrique Gama Dore de Araujo
Orientador

Prof. Dr. Cleber Martins Xavier
1º Examinador

Prof. Dr. Allan Robert da Silva
2º Examinador

AGRADECIMENTOS

Primeiramente agradeço a **Deus** pela oportunidade de concluir mais uma etapa de minha vida.

Em segundo lugar agradeço a minha **mãe e irmão** pelo Amor, apoio, paciência e forças sem medidas, transmitidas em ações concretas para meu bem.

Pelas orações de meus **amigos**, dentre eles, **Lucena**, pelas palavras de motivação e **Thiago** pela disponibilidade e indicação do orientador.

Aproveitando, agradeço ao meu **orientador**, o **Prof. Luiz** pela paciência e por ter me orientado nesse trabalho.

Sou gratíssimo pelas palavras de apoio e companhia de minha **eterna namorada, Thais**.

E por fim, **grato a mim mesmo** por seguir em frente, se esforçado e ter dado o meu melhor nesse trabalho, mesmo com limitação de encontros e de tempo.

RESUMO

O trabalho analisa e prevê a taxa de ocupação hospitalar em um hospital privado de Sergipe, utilizando dados de pacientes-dia e leitos-dia no período de 2022 a 2025, com séries mensais segmentadas por tipo de acomodação (Apartamento, Enfermaria e UTI). A pesquisa tem abordagem quantitativa, caráter aplicado e delineamento observacional, retrospectivo e documental, com dados extraídos do prontuário eletrônico TASY, aplicando etapas de ETL e análise exploratória para caracterizar o perfil de ocupação do paciente por tipo de acomodação. Em seguida, constrói séries temporais mensais da taxa de ocupação por acomodação, realiza testes de estacionariedade, decomposição e análise de autocorrelação e ajusta modelos de séries temporais clássicos (ARIMA e ETS/Holt-Winters aditivo) e um modelo de aprendizado de máquina *XGBoost*.

Os modelos são avaliados com métricas MAE, RMSE e MAPE em conjuntos de treino e teste, e comparados em tabelas consolidadas por tipo de acomodação para o período 2022–2025, evidenciando que o *XGBoost* apresenta, em geral, menor MAPE para Apartamento e UTI, enquanto o ETS(A,N,A) se mostra mais adequado para a série de Enfermaria em termos de comportamento das projeções. As previsões para 2026 indicam manutenção de patamar de ocupação semelhante ao observado no final de 2025, com variação sazonal moderada, e o estudo conclui que a combinação de séries temporais clássicas e *XGBoost* é útil para apoiar o planejamento tático de leitos e recursos, permitindo antecipar períodos de maior pressão assistencial, reduzir riscos de superlotação ou ociosidade e fundamentar decisões de gestão na realidade local do hospital analisado.

Palavras-chave: Taxa de ocupação hospitalar. Séries temporais. ARIMA. *Holt-Winters* aditivo. *XGBoost*. Planejamento de leitos. Previsão.

ABSTRACT

This study analyzes and forecasts the hospital occupancy rate in a private hospital in Sergipe, Brazil, using patient-day and bed-day data from 2022 to 2025. The monthly series are segmented by accommodation type (Private Room, Ward, and ICU). The research adopts a quantitative approach, is applied in nature, and follows an observational, retrospective, and documentary design. Data were extracted from the TASY electronic health record (EHR) system, followed by ETL (Extract, Transform, Load) stages and exploratory analysis to characterize the patient occupancy profile by accommodation type. Subsequently, monthly time series of the occupancy rate were constructed for each category. The study performed stationarity tests, decomposition, and autocorrelation analysis, fitting classical time series models (ARIMA and additive ETS/Holt-Winters) alongside an XGBoost machine learning model.

The models were evaluated using MAE, RMSE, and MAPE metrics on training and test sets, and compared in consolidated tables by accommodation type for the 2022–2025 period. The results indicate that XGBoost generally yields a lower MAPE for Private Rooms and the ICU, while the ETS(A,N,A) model proved more suitable for the Ward series regarding projection behavior. Forecasts for 2026 suggest that occupancy levels will remain similar to those observed at the end of 2025, with moderate seasonal variation. The study concludes that the combination of classical time series and XGBoost is useful for supporting tactical bed and resource planning, allowing for the anticipation of periods of high clinical demand, reducing the risks of overcrowding or idle capacity, and providing a basis for management decisions based on the local reality of the analyzed hospital.

Keywords: Hospital occupancy rate. Time series. ARIMA. Additive Holt-Winters. XGBoost. Bed planning. Forecasting.

LISTA DE ILUSTRAÇÃO

Gráfico 1	Distribuição de Pacientes-dia por Mês/Ano de Ocupação	25
Gráfico 2	Percentual de Registro de Ocupação por Sexo	26
Gráfico 3	Percentual de Registros de Ocupação por Sexo e Tipo de Acomodação	27
Gráfico 4	<i>Box-plot</i> de pacientes-dia por idade	27
Gráfico 5	<i>Box-plot</i> de pacientes-dia por idade e Tipo de Acomodação	28
Gráfico 6	Dispersão dos Pacientes-dia por Idade (visão geral)	28
Gráfico 7	Dispersão dos Pacientes-dia por Idade e Tipo de Acomodação	29
Gráfico 8	Pacientes-dia por Setor de Atendimento	29
Gráfico 9	Pacientes-dia por Setor de Atendimento e Tipo de Acomodação	30
Gráfico 10	Top 10 Quantidade Pacientes-dia por Convênio	31
Gráfico 11	Quantidade Pacientes-dia por Clínica	31
Gráfico 12	Quantidade Pacientes-dia por Tipo de Acomodação	32
Gráfico 13	Taxa de Ocupação Histórica Mensal por Tipo de Acomodação	33
Gráfico 14	Decomposição das Séries Históricas Mensais Diferenciadas ($d = 1$)	35
Gráfico 15	ACF, PACF, Histograma e Série Histórica - UTI	36
Gráfico 16	ACF, PACF, Histograma e Série Histórica - Apartamento	36
Gráfico 17	ACF, PACF, Histograma e Série Histórica - Enfermaria	37
Gráfico 18	Avaliação, ACF, Histograma e QQ-Plot dos Resíduos ARIMA – Enfermaria	38
Gráfico 19	Avaliação, ACF, Histograma e QQ-Plot dos Resíduos ARIMA – Apartamento	38
Gráfico 20	Avaliação, ACF, Histograma e QQ-Plot dos Resíduos ARIMA – UTI	38
Gráfico 21	Avaliação, ACF, Histograma e <i>QQ-Plot</i> dos Resíduos ETS - Enfermaria	40
Gráfico 22	Avaliação, ACF e Histograma dos Resíduos ETS – Apartamento	40
Gráfico 23	Avaliação, ACF e Histograma dos Resíduos ETS – UTI	40
Gráfico 24	Real x Previsto (<i>XGBoost</i>) – Apartamento, Enfermaria e UTI	42
Gráfico 25	Previsão das Séries por Tipo de Acomodação – Jan à Dez 2026	44
Gráfico 26	Previsão da Série Enfermaria – Modelo ETS(A,N,A) – Jan à Dez 2026	45

LISTA DE TABELAS

Tabela 1	Teste de Dickey-Fuller Aumentado das Séries Mensais ($p\text{-valor} \leq 0,05$)	34
Tabela 2	Teste de Dickey-Fuller Aumentado das Séries Mensais Diferenciadas ($d = 1$)	34
Tabela 3	Métricas de Desempenho e Avaliação das Séries Diferenciadas	37
Tabela 4	Teste Shapiro-Wilk para os Resíduos – Modelo ARIMA	39
Tabela 5	Parâmetros, Métricas de Desempenho e Avaliação – Modelo ETS(A,N,A)	39
Tabela 6	Teste Ljung-Box e Shapiro-Wilk para os Resíduos – Modelo ETS(A,N,A)	40
Tabela 7	Métricas de Desempenho e Avaliação – Modelo XGBoost	42
Tabela 8	Comparação ARIMA vs EST vs XGBOOST – Apartamento	43
Tabela 9	Comparação ARIMA vs EST vs XGBOOST – Enfermaria	43
Tabela 10	Comparação ARIMA vs EST vs XGBOOST – UTI	43

LISTA DE ABREVIATURAS E SIGLAS

ACF	Função de Autocorrelação.
AIC	Akaike Information Criterion.
ARIMA	Autorregressivo Integrado de Médias Móveis
BIC	Bayesian Information Criterion.
CID	Código Internacional de Doenças.
EDA	<i>Exploratory Data Analysis</i> (análise exploratória de dados).
ETL	<i>Extract, Transform and Load</i> (extração, transformação e carregamento).
ETS	Error, Trend, Seasonality (<i>Exponential Smoothing</i>).
MAE	Erro Médio Absoluto.
MAPE	Erro Percentual Absoluto Médio.
NA's	<i>Not Available</i> (registros vazios ou valores ausentes).
PACF	Função de Autocorrelação Parcial.
RMSE	Raiz do Erro Quadrático Médio.
UTI	Unidade de Terapia Intensiva.

SUMÁRIO

1	INTRODUÇÃO	10
2	OBJETIVOS	12
2.1	Objetivo Geral	12
2.2	Objetivos Específicos	12
3	JUSTIFICATIVA	13
4	REVISÃO LITERÁRIA	15
4.1	Uso de Sistemas de Informação em Saúde	16
4.2	Métodos Quantitativos e Séries Temporais na Saúde	16
4.3	Previsão da Demanda e da Ocupação de Leitos Hospitalares	17
4.4	Modelos ARIMA, <i>Holt-Winters</i> Aditivo e <i>Machine Learning</i> na Saúde	18
4.5	Síntese da Literatura e Possíveis Lacunas	19
5	METODOLOGIA	20
5.1	Tipo e Delineamento da Pesquisa	20
5.2	Fonte de Dados e Procedimentos de Coleta	20
5.3	Aspectos Éticos	21
5.4	Definição das Variáveis e Regras de Negócio	21
5.5	Procedimentos para o Tratamento e Análise dos Dados	23
6	RESULTADOS E DISCUSSÃO	25
6.1	Caracterização da Base de Dados	25
6.2	Análise Descritiva do Perfil de Ocupação	26
6.2.1	Distribuição por Sexo e Idade	26
6.2.2	Setor de Atendimento, Convênio e Clínica	29
6.2.3	Segmentação por Tipo de Acomodação	31
6.3	Análise da Taxa de Ocupação Hospitalar em Série Temporal	32
6.4	Modelagem ARIMA	34
6.4.1	Estacionariedade e decomposição	34
6.4.2	Ajuste e Diagnóstico dos Modelos	37
6.5	Modelagem <i>Holt-Winters</i> Aditiva – ETS (A,N,A)	39
6.5.1	Ajuste dos Modelos	39
6.5.2	Diagnóstico dos Resíduos, Testes de <i>Ljung-Box</i> e <i>Shapiro-Wilk</i>	39
6.6	Modelagem <i>XGBoost</i>	41

6.6.1	Preparação das Features e Ajuste do Modelo-----	41
6.6.2	Aplicação da Modelagem por Tipo de Acomodação-----	42
6.7	Comparação ARIMA vs. Holt-Winters Aditivo vs. <i>XGBoost</i>-----	43
6.8	Projeções da Taxa de Ocupação para 2026-----	43
7	CONCLUSÕES-----	46
	BIBLIOGRAFIA-----	47
	APÊNDICE A-----	51

1 INTRODUÇÃO

A organização e o uso eficiente dos leitos hospitalares constituem um dos principais desafios da gestão em saúde, sobretudo em contextos de pressão assistencial crescente e restrição de recursos financeiros e humanos, dado que a disponibilidade de leitos qualificados impacta diretamente o acesso e a continuidade do cuidado (ANS, 2023).

No estado de Sergipe, a rede hospitalar no geral apresenta número limitado de estabelecimentos com internação e leitos especializados quando comparada a unidades federativas maiores, o que aumenta a relevância da gestão da capacidade instalada, tanto no Sistema Único de Saúde quanto na rede privada (DATASUS, 2025; CNES, 2025). Apesar de iniciativas recentes de expansão e qualificação de leitos, como a ampliação de leitos de UTI e a interiorização de serviços de maior complexidade, a necessidade de otimizar o uso desses recursos permanece como um tema central para a qualidade da assistência e o acesso da população (MÉRICI, 2016).

Nesse cenário, a forma como os leitos são ocupados ao longo do tempo influencia diretamente o tempo de espera, a rotatividade e o fluxo de pacientes nos diferentes tipos de acomodação, com impactos na segurança e nos desfechos clínicos.

Estudos de previsão de ocupação de leitos mostram que níveis elevados e mal planejados de ocupação podem estar associados a atrasos na internação, aumento de cancelamentos e piora de resultados assistenciais, especialmente em unidades de terapia intensiva (ZHAO et al., 2024; MELLO et al., 2025). A subutilização ou a superlotação de determinados segmentos, como enfermarias e UTIs, pode resultar tanto em desperdício de recursos quanto em atrasos na internação de pacientes que necessitam de cuidados especializados, afetando a qualidade e a integralidade do cuidado (OVERMAN et al., 2024; FINDLAY, 2019). A discussão sobre taxa de ocupação ganha, portanto, relevância estratégica, uma vez que se relaciona com dimensionamento de leitos, planejamento de equipes e tomada de decisão em níveis tático e operacional.

Do ponto de vista metodológico, os métodos estatísticos e de ciências atuariais aplicados à saúde oferecem ferramentas robustas para descrever, interpretar, modelar e prever padrões de utilização de serviços hospitalares, permitindo transformar séries históricas de produção assistencial em insumos quantitativos para o planejamento (KIM et al., 2024; ZHAO et al., 2024). Técnicas de séries temporais clássicas, combinadas com métodos de aprendizagem de máquina, têm sido utilizadas em diferentes contextos para

estimar demanda futura, avaliar cenários e apoiar a alocação de recursos hospitalares (OVERMAN et al., 2024; MELLO et al., 2025). A aplicação integrada dessas abordagens possibilita capturar tanto estruturas de tendência e sazonalidade quanto relações não lineares, comum na área da saúde, oferecendo previsões potencialmente mais aderentes ao comportamento real da taxa de ocupação.

As motivações para este estudo nascem da prática profissional do autor na interface entre análise de dados e gestão em saúde, na qual se observa, de forma recorrente, que a ausência de modelos quantitativos de apoio à decisão contribui para o uso subótimo de leitos e recursos hospitalares, com repercussões sobre o tempo de internação, o acesso a vagas e a experiência do paciente. A percepção de que decisões críticas ainda são, em grande medida, tomadas com base em julgamentos subjetivos e em indicadores retrospectivos, e não em previsões estruturadas, reforça a necessidade de desenvolver aplicações que aproximem a Estatística e a Ciência de Dados da realidade cotidiana dos serviços hospitalares.

Diante desse contexto, o objeto de análise deste Trabalho é entender e analisar o comportamento da taxa de ocupação hospitalar de um hospital privado de Sergipe, mensurada a partir de dados históricos de pacientes-dia e leitos-dia, a fim de descrever seu padrão histórico e construir modelos de previsão de curto prazo, capazes de apoiar o planejamento da gestão de leitos e recursos assistenciais.

A organização do presente trabalho está exposta da seguinte forma:

- a) No capítulo 1 ocorre a introdução da problemática do tema a ser estudada, descrevendo o contexto em que o trabalho se encaixa.
- b) No capítulo 2 é descrito os objetivos gerais e específicos do trabalho
- c) O capítulo 3 traz a justificativa, ou seja, descreve o porquê esse tema ser importante para a área de estudo.
- d) O capítulo 4 traz a revisão de literatura, essencial para fundamentação do estudo, mostrando o que já foi estudado e em que tipo de pesquisa se encaixa.
- e) No capítulo 5 foi destrinchada a metodologia, trazendo aspectos éticos dos dados, a origem dos dados, as ferramentas e os métodos utilizados.
- f) O capítulo 6 é descrito os resultados e discussão das análises, interpretando os resultados e o que significa frente ao problema e objetivos traçados.
- g) O capítulo 7 traz as conclusões do trabalho, respondendo os objetivos definidos.
- h) Por fim, tem-se a listagem da bibliografia usada e o apêndice com o *script* R utilizado para análise dos dados.

2 OBJETIVOS

2.1 Geral

Analisar e modelar a taxa de ocupação hospitalar de um hospital privado em Sergipe, no período de 2022 a 2025, por tipo de acomodação (Apartamento, Enfermaria e UTI), utilizando técnicas de séries temporais clássicas e de aprendizagem de máquina, de modo a apoiar o planejamento das gestões de leitos a nível tático e operacional e de auxiliar no uso otimizado de recursos assistenciais.

2.2 Específicos

Os objetivos específicos desta pesquisa são:

- i) Caracterizar a base de dados de ocupação hospitalar, descrevendo o perfil de pacientes-dia por sexo, idade, setor de atendimento, convênio, clínica e tipo de acomodação, no período de 2022 a 2025.
- j) Identificar a granularidade temporal mais adequada (mensal, quinzenal ou semanal) para modelagem da taxa de ocupação hospitalar, considerando o alinhamento com o planejamento tático e o comportamento das séries.
- k) Descrever a dinâmica da taxa de ocupação por tipo de acomodação (Apartamento, Enfermaria e UTI), evidenciando diferenças de perfil operacional e padrões de uso entre os tipos de leitos.
- l) Analisar a estrutura temporal das séries de taxa de ocupação por tipo de acomodação, por meio de testes de estacionariedade, decomposição em tendência, sazonalidade e ruído, e funções de autocorrelação.
- m) Ajustar e avaliar modelos de séries temporais clássicos (ARIMA e métodos de suavização exponencial/*Holt-Winters*) para a previsão da taxa de ocupação por tipo de acomodação, verificando o desempenho e os diagnósticos dos resíduos.
- n) Ajustar e avaliar modelos de aprendizagem de máquina baseados em *XGBoost* para a previsão da taxa de ocupação por tipo de acomodação, utilizando defasagens e componentes de tempo como variáveis explicativas.
- o) Comparar o desempenho preditivo dos modelos clássicos e dos modelos de *XGBoost*, por tipo de acomodação, verificando diferenças nas métricas de erro

(exemplo: RMSE, MAE e MAPE) e na capacidade de capturar a dinâmica da série.

- p) Explicar as implicações práticas das previsões da taxa de ocupação com a granularidade de tempo escolhida por tipo de acomodação para o planejamento de leitos e a gestão da capacidade hospitalar, destacando recomendações para o uso dos modelos na tomada de decisão.

3 JUSTIFICATIVA

A gestão eficiente de leitos hospitalares é reconhecida como componente crítico para garantir acesso oportuno, qualidade assistencial e uso racional de recursos em saúde, sobretudo em contextos de alta demanda e restrição orçamentária (HUANG; HARDIN; PRAKASH, 2025).

Em cenários onde a capacidade instalada é limitada, falhas na organização do fluxo de pacientes podem resultar em superlotação, cancelamento de procedimentos, aumento do tempo de espera e piora de desfechos clínicos, especialmente em setores de internação e unidades de terapia intensiva (MELLO et al., 2025).

No estado de Sergipe, reportagens e documentos oficiais indicam dificuldades recorrentes relacionadas à estrutura hospitalar e ao acesso a serviços especializados, evidenciando a vulnerabilidade da rede e a importância de aprimorar o planejamento da capacidade assistencial, tanto na esfera pública quanto privada (MANGUE JORNALISMO, 2025). Nesse contexto, investigar a taxa de ocupação hospitalar e sua dinâmica temporal assume relevância social direta, por se relacionar com o direito de acesso à saúde, a organização de filas de internação e a segurança do paciente.

Do ponto de vista teórico e metodológico, a literatura recente destaca o potencial de modelos de previsão de ocupação de leitos para apoiar decisões de dimensionamento de recursos, programação de cirurgias eletivas, escalas de profissionais e estratégias de regulação dos leitos (ZHAO et al., 2024; HUANG; HARDIN; PRAKASH, 2025).

Estudos comparativos mostram que técnicas de séries temporais clássicas, como ARIMA e modelos de suavização exponencial, sozinhas ou combinadas com métodos de aprendizagem de máquina, incluindo algoritmos como o *XGBoost*, podem produzir previsões mais acuradas e úteis para o planejamento da ocupação hospitalar (HUANG; HARDIN; PRAKASH, 2025). No entanto, observa-se que grande parte dessas aplicações

está concentrada em grandes centros ou sistemas de saúde de países de alta renda, havendo espaço para estudos que adaptem e avaliem essas abordagens em contextos regionais específicos, como o de um hospital privado extra porte (acima de 500 leitos operacionais) em Sergipe, onde existe perfis demográficos e sociais heterogêneos.

No estágio atual de desenvolvimento do conhecimento, há um corpo consolidado de pesquisas que utiliza modelos de séries temporais e métodos de *machine learning* para prever demanda por leitos, internações não eletivas ou volumes de atendimento, demonstrando ganhos em acurácia e na capacidade de antecipar picos de utilização (HUANG; HARDIN; PRAKASH, 2025). Apesar disso, ainda são relativamente escassos estudos que considerem simultaneamente a segmentação por tipo de acomodação (como apartamentos, enfermarias e UTIs) e a comparação sistemática entre modelos clássicos e de aprendizagem de máquina em realidades hospitalares brasileiras fora dos grandes centros (HUANG; HARDIN; PRAKASH, 2025). Esta pesquisa busca contribuir para essa lacuna, ao integrar técnicas estatísticas com foco na gestão de leitos e utilizando dados reais de ocupação.

Sob a perspectiva prática, a justificativa deste trabalho está no potencial de gerar evidências e ferramentas para apoiar decisões mais informadas sobre o uso de leitos e recursos em um contexto local concreto. Ao modelar e prever a taxa de ocupação por tipo de acomodação, espera-se fornecer subsídios para o planejamento de curto prazo, permitindo antecipar períodos de maior pressão assistencial, ajustar escalas de profissionais, organizar internações eletivas e avaliar necessidades de expansão ou reorganização da capacidade existente (HUANG; HARDIN; PRAKASH, 2025).

Essa perspectiva dialoga com a necessidade de envolver a análise de dados na gestão em saúde, onde a ausência de modelos quantitativos estruturados contribui para o uso subótimo de leitos e recursos hospitalares, com repercussões sobre o tempo de internação, o acesso a vagas e a experiência de bem-estar do paciente.

Por fim, no ponto de vista acadêmico, contribui para a consolidação do uso combinado de séries temporais e algoritmos de aprendizagem de máquina na análise da taxa de ocupação hospitalar em um contexto regional ainda pouco explorado. Do ponto de vista aplicado, pretende oferecer subsídios para práticas de gestão de leitos mais baseadas em evidências, com potencial para reduzir superlotação, melhorar o aproveitamento de recursos e favorecer a qualidade da assistência prestada à população atendida pelo hospital estudado (ZHAO et al., 2024).

4 REVISÃO LITERÁRIA

A gestão de leitos hospitalares é apontada como um dos elementos centrais da eficiência dos serviços de saúde, pois a disponibilidade e o uso adequado dos leitos condicionam o acesso, o tempo de espera e os desfechos clínicos dos pacientes internados (GLLEDDSON, 2021; SOUZA; SILVA, 2021).

Estudos nacionais mostram que taxas de ocupação persistentemente elevadas em unidades de internação, especialmente em unidades de terapia intensiva (UTI), associam-se à desorganização do fluxo assistencial, ao cancelamento de procedimentos, ao aumento de eventos adversos e à piora da segurança do paciente (GLLEDDSON, 2021; SOUZA, F. M.; SOUZA, A. M.; LOPES, L. F. P., 2016). Por outro lado, níveis de ocupação muito baixos indicam ociosidade da capacidade instalada, com impacto econômico negativo para instituições públicas e privadas (GLLEDDSON, 2021).

No âmbito do Sistema Único de Saúde (SUS), o dimensionamento de leitos tradicionalmente se baseia em parâmetros populacionais e em fórmulas que relacionam número estimado de internações, tempo médio de permanência e uma taxa de ocupação alvo, usualmente situada em torno de 80% a 85% (GLLEDDSON, 2021; MINISTÉRIO DA SAÚDE, 2002). Entretanto, a literatura recente argumenta que parâmetros fixos são insuficientes diante da variabilidade temporal da demanda, defendendo o uso de modelos que incorporem explicitamente a dinâmica da ocupação ao longo do tempo para apoiar o planejamento (GLLEDDSON, 2021; BITTENCOURT; HORTALE, 2016). Nessa perspectiva, analisar a evolução histórica da taxa de ocupação e projetar cenários futuros torna-se fundamental para reduzir simultaneamente o risco de superlotação e de ociosidade de leitos, especialmente em contextos com oferta restrita, como ocorre em muitos estados brasileiros, incluindo Sergipe.

Além de apoiar o planejamento de capacidade, a taxa de ocupação é amplamente utilizada em vigilância epidemiológica e controle de infecções, servindo de denominador para taxas de utilização de dispositivos e indicadores de densidade de incidência de infecções relacionadas à assistência (PROQUALIS, 2009; VIGISPEC, 2025). O uso consistente desses conceitos no presente trabalho garante que os indicadores calculados sejam comparáveis aos de outras instituições e estudos nacionais, reforçando a aderência às boas práticas de gestão hospitalar.

4.1 Uso de Sistemas de Informação em Saúde

A qualidade das análises de ocupação hospitalar depende diretamente da confiabilidade e da granularidade dos dados disponíveis, o que, na prática, está associado à maturidade dos sistemas de informação em saúde adotados pelas instituições (TURATO, 2005). No contexto brasileiro, hospitais de médio e grande porte vêm substituindo gradualmente registros manuais por sistemas integrados de gestão hospitalar e prontuário eletrônico, que consolidam dados assistenciais e administrativos em bases estruturadas (FREITAS, 2019). Esses sistemas permitem registrar, em tempo quase real, internações, transferências, altas e outros eventos, viabilizando a construção de indicadores diários para análise da ocupação dos leitos.

Além de sistemas de prontuário eletrônico, estudos também apontam que *dashboards* bem estruturados, integrados aos sistemas transacionais, podem apoiar gestores na identificação de gargalos, na priorização de ações e na avaliação de resultados, sobretudo quando conectados a processos de decisão baseados em evidências (SILVA; OLIVEIRA, 2020).

4.2 Métodos Quantitativos e Séries Temporais na Saúde

A literatura nacional destaca que métodos quantitativos ocupam papel central na avaliação de serviços de saúde, na mensuração de eventos e na construção de modelos explicativos e preditivos (TURATO, 2005). De acordo com esse autor, abordagens quantitativas são particularmente adequadas quando se busca estimar magnitudes, comparar grupos ou projetar comportamentos futuros, como no caso da demanda por leitos e da taxa de ocupação.

A análise de séries temporais é uma das principais ferramentas para tratar dados coletados sequencialmente ao longo do tempo, permitindo identificar padrões de tendência, sazonalidade e ciclos, além de produzir previsões (SOUZA; SILVA, 2021).

Em estudo aplicado ao sistema de saúde, Souza e Silva (2021) mostram que modelos de séries temporais são capazes de capturar estruturas sazonais e de tendência em séries mensais, oferecendo subsídios relevantes para a gestão de recursos (SOUZA; SILVA, 2021). De forma similar, Freitas (2019), ao analisar dados de pacientes em UTI, reforça a importância da organização temporal das informações e do uso de modelos estatísticos para interpretar a evolução de indicadores assistenciais (FREITAS, 2019).

Esses trabalhos sustentam a escolha de uma abordagem quantitativa e temporal para o estudo da taxa de ocupação hospitalar.

4.3 Previsão da Demanda e da Ocupação de Leitos Hospitalares

Especificamente em relação à demanda e à ocupação de leitos hospitalares, há um conjunto crescente de estudos brasileiros que utilizam séries temporais para apoiar o planejamento de capacidade. O autor Gledson (2021) desenvolveu um modelo baseado em séries temporais para previsão das diárias de leitos de UTI na rede SUS de Goiânia, comparando modelos ARIMA sazonais e modelos dinâmicos, e concluiu que a previsão da demanda de leitos é fator crítico para o desempenho operacional, reduzindo tanto o risco de ociosidade quanto de recusas por falta de vagas. No mesmo trabalho, o autor observa que taxas de ocupação acima do intervalo recomendado (75% a 85%) se associam a aumento de eventos adversos e infecções, enquanto taxas muito baixas indicam subutilização de recursos (GLEDSON, 2021).

Outros estudos analisam a taxa de ocupação por meio de séries temporais em diferentes contextos. Souza, Souza e Lopes (2016) modelaram a demanda de leitos em hospital universitário gaúcho utilizando modelos ARIMA, demonstrando que a metodologia foi capaz de reproduzir o comportamento observado e fornecer subsídios para o melhor funcionamento do serviço. De maneira convergente, trabalhos que combinam séries temporais e teoria de filas para estimar o número de leitos necessários em UTI destacam que previsões de demanda, aliadas a parâmetros como tempo médio de permanência, podem orientar decisões de expansão ou remanejamento de leitos com resultados compatíveis às recomendações técnicas nacionais (BITTENCOURT; HORTALE, 2016).

Durante a pandemia de COVID-19, a importância de modelos preditivos para ocupação de leitos foi ainda mais evidenciada. Estudos como o *Forecast UTI* e aplicações do modelo matemático SEIHDR em capitais brasileiras mostraram que simulações de cenários de ocupação de UTI auxiliam gestores a dimensionar leitos, prever riscos de colapso e planejar estratégias de mitigação (VENTURA et al., 2021; OLIVEIRA et al., 2021).

Esses trabalhos reforçam a relevância de soluções quantitativas para gestão de capacidade, sobretudo em situações de grande incerteza e variação abrupta da demanda.

4.4 Modelos ARIMA, *Holt-Winters* Aditivo e *Machine Learning* na Saúde

No campo da previsão de séries temporais, os modelos ARIMA (Autorregressivos Integrados de Médias Móveis) e os modelos de suavização exponencial (como Holt-Winters aditivo ou ETS) constituem abordagens clássicas amplamente documentadas na literatura (BOX; JENKINS; REINSEL, 2016; SOUZA; SILVA, 2021). Esses modelos permitem decompor a série em componentes de nível, tendência e sazonalidade, ajustando parâmetros a partir do histórico observado e gerando previsões com intervalos de confiança. Em contexto hospitalar, têm sido utilizados para prever desde demanda por leitos até volumes de atendimentos em serviços de urgência, com resultados satisfatórios em horizontes de curto prazo (GLLEDDSON, 2021; SOUZA; SOUZA; LOPES, 2016).

Mais recentemente, o avanço *machine learning*, ampliou as possibilidades para a modelagem de ocupação hospitalar. Trabalhos brasileiros têm explorado algoritmos como *XGBoost*, redes neurais e modelos híbridos para previsão de leitos de UTI e de ocupação em diferentes níveis de atenção, comparando seu desempenho com modelos tradicionais (GLLEDDSON, 2021; COSTA, 2020). Em geral, esses estudos relatam que modelos de machine learning podem capturar relações não lineares e interações complexas entre variáveis temporais e defasagens, alcançando muitas vezes menor erro percentual médio (MAPE) em comparação a modelos puramente lineares, embora exijam maior cuidado na parametrização e na interpretação (GLLEDDSON, 2021; COSTA, 2020).

Essa tendência é consistente com a literatura internacional que aplica *XGBoost* e outros métodos de *machine learning* à previsão de indicadores de saúde, como volume de atendimentos e ocupação de serviços, demonstrando boa capacidade preditiva em séries relativamente curtas e com padrões sazonais complexos (BORIN et al., 2025). No entanto, os autores convergem ao afirmar que a escolha do modelo “ótimo” depende do contexto, da quantidade de dados e do objetivo da aplicação, recomendando comparações sistemáticas entre modelos clássicos e de aprendizagem de máquina antes da adoção de uma solução para suporte à decisão (GLLEDDSON, 2021; SOUZA; SOUZA; LOPES, 2016).

4.5 Síntese da Literatura e Possíveis Lacunas

O conjunto da literatura revisada permite concluir que:

- a) A gestão de leitos e a taxa de ocupação hospitalar são temas centrais para a eficiência e a segurança dos serviços de saúde;
- b) Conceitos como paciente-dia e leito-dia, conforme definidos em manuais nacionais, são fundamentais para o cálculo de indicadores e para a comparabilidade entre instituições;
- c) Sistemas de informação integrados, prontuário eletrônico e ferramentas de BI têm papel crucial na disponibilização de dados confiáveis para análise;
- d) Modelos de séries temporais, combinados ou não com técnicas de aprendizagem de máquina, têm se mostrado promissores para previsão de demanda e de ocupação de leitos (BRASIL, 2002; CONASS, 2021; GLEDDSON, 2021; SOUZA; SILVA, 2021).

Apesar desse avanço, ainda são relativamente escassos os estudos que abordam, de forma integrada, a taxa de ocupação hospitalar segmentada por tipo de acomodação em hospitais privados brasileiros, utilizando dados operacionais reais e comparando explicitamente o desempenho de modelos como ARIMA, ETS e *XGBoost* em séries mensais (GLEDDSON, 2021; SOUZA; SOUZA; LOPES, 2016). A maior parte dos trabalhos concentra-se em séries agregadas de UTI, em contextos estritamente públicos ou em cenários epidêmicos específicos, deixando em aberto a aplicação sistemática dessas técnicas em ambientes de internação mista e em horizonte tático de planejamento.

Dessa forma, o presente estudo se insere na literatura ao aplicar e comparar modelos de séries temporais clássicos e de aprendizagem de máquina à taxa de ocupação mensal por tipo de acomodação em um hospital de perfil privado em um estado do nordeste brasileiro (Sergipe). Com isso, busca-se não apenas reproduzir resultados já demonstrados em outros contextos, mas também gerar evidências específicas e potencialmente transferíveis para a realidade de hospitais de perfil populacional e porte semelhante no cenário brasileiro.

5 METODOLOGIA

5.1 Tipo e Delineamento da Pesquisa

Este Trabalho de Conclusão de Curso adota uma abordagem quantitativa, de natureza aplicada, com objetivos descritivos e explicativos, utilizando delineamento observacional, retrospectivo e documental, baseado em dados secundários de produção assistencial hospitalar. Segundo Gil (2008), pesquisas descritivas buscam caracterizar populações ou fenômenos a partir de técnicas padronizadas de coleta e análise de dados, enquanto as pesquisas explicativas procuram identificar fatores que contribuem para a ocorrência dos fenômenos, aprofundando o conhecimento da realidade estudada. No campo da saúde, métodos quantitativos são amplamente utilizados para mensurar eventos, testar associações e construir modelos preditivos, constituindo a base de muitos estudos epidemiológicos e de gestão de serviços (TURATO, 2005).

No presente estudo, a pesquisa é retrospectiva porque se vale de dados já existentes, registrados rotineiramente no prontuário eletrônico e consolidados em sistemas de informação internos, sem interferência sobre a exposição ou o desfecho dos pacientes. Trata-se também de pesquisa documental, uma vez que utiliza como fonte principal registros administrativos e assistenciais provenientes de um sistema de gestão hospitalar (ERP) e de um painel gerencial em Power BI, os quais são explorados com finalidade científica. A natureza aplicada decorre do fato de que os resultados são orientados para apoiar o planejamento de leitos e recursos em um hospital privado de Sergipe, aproximando métodos estatísticos e de ciência de dados da prática de gestão em saúde (FREITAS, C. A. G. S, 2019).

5.2 Fonte de Dados e Procedimentos de Coleta

Os dados utilizados neste estudo foram extraídos do sistema de prontuário eletrônico e gestão hospitalar *TASY*, utilizado como ERP pela instituição, a partir de um painel gerencial desenvolvido no *Microsoft Power BI*, que consumia e consolidava diariamente informações de ocupação hospitalar segundo regras de negócio definidas pela gestão do hospital. O painel de *Business Intelligence* foi empregado rotineiramente para monitoramento assistencial e operacional e, para fins desta pesquisa, seus dados foram exportados em formato *.xlsx*, preservando as colunas e filtros de interesse para análise da

taxa de ocupação. O uso de ferramentas de *Business Intelligence* (BI), como o *Microsoft Power BI*, tem se difundido como estratégia para transformar grandes volumes de dados em painéis gerenciais interativos, facilitando o monitoramento da produção assistencial e o acompanhamento da ocupação de leitos (SILVA; OLIVEIRA, 2020).

A extração contemplou o período de janeiro de 2022 a dezembro de 2025, englobando registros diários de pacientes-dia e leitos-dia de todos os setores de internação do hospital estudado. Não houve contato direto com pacientes, entrevistas ou aplicação de questionários; toda a informação foi obtida de forma secundária, a partir de bases administrativas já consolidadas, em consonância com a prática de estudos observacionais baseados em prontuários eletrônicos descrita na literatura nacional (GLLEDDSON, J. M, 2021).

Antes da análise, o arquivo exportado foi submetido a um processo de Extração, Transformação e Carga (ETL) em ambiente estatístico, com padronização de formatos de data, categorização de variáveis e tratamento de valores ausentes, de forma similar ao recomendado em estudos que integram dados assistenciais em dashboards e modelos analíticos (GLLEDDSON, J. M, 2021).

5.3 Aspectos Éticos

Por utilizar exclusivamente dados secundários, extraídos de sistemas institucionais já existentes, sem identificação nominal dos pacientes e sem qualquer intervenção sobre o cuidado prestado, este estudo se caracteriza como de baixo risco para os sujeitos envolvidos. A base de dados foi anonimizada antes da análise, removendo-se identificadores diretos e restringindo o uso das informações ao propósito desta pesquisa.

5.4 Definição das Variáveis e Regras de Negócio

A base de dados é composta por registros diários de ocupação hospitalar, com as seguintes variáveis principais: data, qt_pac_dia, qt_leito_dia, sexo, idade, convênio, setor_atendimento, cid_principal, id_atend, clínica e tipo_acomodação.

A variável data representa a data de referência do registro de ocupação, sendo utilizada como eixo temporal para agregações diárias e mensais.

O paciente-dia é a unidade de medida que representa a assistência prestada a um paciente internado durante um dia hospitalar, contabilizando todos os pacientes presentes

no hospital em um horário de corte padronizado, geralmente à zero hora, independentemente do horário de admissão e desconsiderando o dia da alta (BRASIL, 2002), codificada como 0 (ausência de ocupação) ou 1 ou mais (presença de paciente-dia). Na prática, o total de pacientes-dia em um determinado período corresponde à soma dos censos diários de pacientes em cada unidade de internação (CQH, 2016).

O conceito de leito-dia, por sua vez, refere-se ao número de leitos operacionais disponíveis em cada dia, incluindo, quando aplicável, leitos adicionais utilizados por mais de 24 horas (CQH, 2016).

A ocupação hospitalar baseia-se em indicadores construídos a partir dos conceitos de paciente-dia e leito-dia, padronizados em documentos técnicos nacionais (BRASIL, 2002; CQH, 2016). A combinação dessas duas grandezas permite o cálculo da taxa de ocupação hospitalar (TOH), usualmente expressa como percentual, dada por:

$$TOH = \frac{\sum PACIENTES - DIA}{\sum LEITOS - DIA} \times 100$$

Essa definição está alinhada tanto às orientações do Ministério da Saúde quanto a cadernos de indicadores utilizados em programas de qualidade hospitalar (BRASIL, 2002; CONASS, 2021; CQH, 2016).

As variáveis convênio, setor_atendimento, cid_principal e clínica caracterizam, respectivamente, o financiador do atendimento, o local físico de internação, o diagnóstico principal segundo a Classificação Internacional de Doenças (CID) e o perfil clínico do paciente, possibilitando segmentações por linha de cuidado, setor de internação e perfil de demanda assistencial (BRASIL, 2002; BRASIL, 2004).

A variável id_atend funciona como identificador único do atendimento, de modo que a repetição do mesmo código indica múltiplos registros de paciente-dia ao longo do tempo para um mesmo episódio de internação.

Por fim, a variável tipo_acomodação descreve o tipo de leito (por exemplo, Apartamento, Enfermaria e UTI), independentemente de estar ocupado ou não, sendo fundamental para a construção de séries temporais segmentadas por tipo de acomodação, o que permite capturar padrões distintos de uso entre leitos clínicos, de terapia intensiva e de acomodação diferenciada.

Para fins de modelagem, a taxa de ocupação foi agregada em séries mensais por tipo de acomodação, reduzindo o ruído de alta frequência observado em periodicidades

semanais ou quinzenais e alinhando a granularidade da análise ao horizonte tático de planejamento de recursos do hospital (LEITE, A. C. A, 2021).

5.5 Procedimentos para o Tratamento e Análise dos Dados

O tratamento e a análise dos dados foram conduzidos em ambiente estatístico, utilizando-se a linguagem R e o *software R-Studio* para as etapas de ETL, análise exploratória e modelagem de séries temporais. Inicialmente, realizou-se a padronização de categorias textuais (por exemplo, substituição de valores ausentes em variáveis categóricas por rótulos como "Não informado"), a verificação de consistência das variáveis numéricas e a exclusão de registros com informações claramente inválidas, seguindo boas práticas de pré-processamento em estudos quantitativos na área da saúde (TURATO, 2025).

Na sequência, desenvolveu-se uma análise exploratória de dados (EDA), com construção de *dataframes* específicos para caracterizar o perfil de ocupação por sexo, idade, convênio, setor de atendimento, clínica, CID principal e tipo de acomodação, bem como combinações entre essas variáveis, como pacientes-dia por setor e tipo de acomodação ou por clínica e tipo de acomodação.

A seguir, foram utilizados gráficos de barras, *box-plots*, gráficos de dispersão e gráficos de linha para visualizar a distribuição de pacientes-dia e leitos-dia pelas variáveis categóricas ao longo do tempo, identificando tendências gerais, sazonalidade e eventuais outliers associados a mudanças estruturais na capacidade instalada do hospital, similar a estudos sobre ocupação (MORI e GONÇALVES, 2021).

Para a etapa de análise temporal e preditiva, foram ajustados modelos ARIMA sazonais, modelo de suavização exponencial (ETS (A,N,A) ou *Holt-Winters* aditivo) e modelos de aprendizado de máquina baseados em *XGBoost* (HYNDMAN; ATHANASOPOULOS, 2018; CHEN; GUESTRIN, 2016).

Inicialmente a base diária foi agregada em séries mensais de taxa de ocupação por tipo de acomodação (Apartamento, Enfermaria e UTI), resultando em três séries principais para modelagem. Em cada série, foram aplicados procedimentos clássicos de análise de séries temporais: inspeção visual, decomposição em componentes de tendência, sazonalidade e ruído, e testes de estacionariedade, como o teste de *Dickey-Fuller* aumentado. Foi aplicada diferenciação de ordem 1 para obtenção de estacionariedade antes do ajuste dos modelos ARIMA, conforme o procedimento

proposto na metodologia *Box-Jenkins* (BOX; JENKINS; REINSEL, 2016; HAMILTON, 1994).

Para o modelo *Holt-Winters* aditivo, a série foi decomposta em três componentes: **nível**, **tendência** e **sazonalidade aditiva** (com amplitude de sazonalidade aproximadamente constante), usando equações recursivas com parâmetros de suavização α, β, γ que dão mais peso às observações mais recentes, de acordo com a formulação clássica dos modelos de suavização exponencial (BROWN, 1963; HOLT, 2004).

No caso do modelo de *machine learning*, converteu-se o problema de séries temporais em um problema supervisionado com criação de *lags* de curto prazo (lag1, lag2) e sazonal (lag12), além de variáveis de tempo (ano e mês). Para cada tipo de acomodação, os dados foram particionados em conjuntos de treino e teste respeitando a ordem temporal, permitindo avaliar a capacidade preditiva dos modelos em um período recente, tal como recomendado em aplicações de *XGBoost* e *gradient boosting* para previsão com dados temporais (CHEN; GUESTRIN, 2016; FRIEDMAN, 2001).

A comparação entre os modelos foi realizada por meio de métricas de erro como *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE) e *Mean Absolute Percentage Error* (MAPE), calculadas sobre o período de teste e, em análises complementares, sobre todo o histórico disponível (BOX; JENKINS; REINSEL, 2016). Ainda sobre as comparações entre modelos, é importante ressaltar que foram usados em todas as séries e modelos a mesma periodicidade nos conjuntos de treino (jan 2022 a dez 2024) e teste (jan à dez 2025).

Conforme MORI e GONÇALVES (2021), faz sentido usar como estratégia a utilização de múltiplos indicadores de desempenho para comparar modelos de previsão em saúde, buscando equilibrar acurácia absoluta, erro relativo e sensibilidade a picos de demanda. Com base nessas métricas e na análise gráfica de valores observados e previstos, foram selecionados os modelos considerados mais adequados para cada tipo de acomodação, que então foram utilizados para gerar projeções da taxa de ocupação em horizontes de curto prazo.

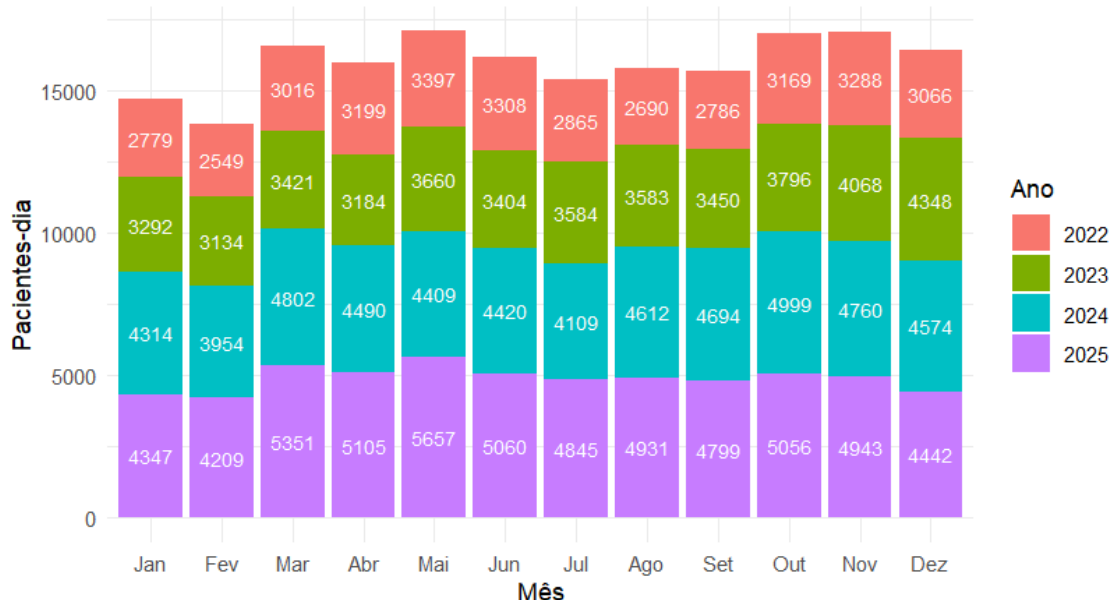
6 RESULTADOS E DISCUSSÃO

6.1 Caracterização da Base de Dados

A base de dados analisada compreende o quantitativo de 238.156 registros de ocupação hospitalar dos pacientes de um hospital privado em Sergipe, no período de janeiro de 2022 a dezembro de 2025, com informações diárias de pacientes-dia, leitos-dia, data de referência da ocupação, idade, sexo, setor de atendimento, convênio, clínica, diagnóstico principal (CID) e tipo de acomodação (Apartamento, Enfermaria, UTI). A taxa de ocupação foi construída como a razão entre o somatório de pacientes-dia e leitos-dia e, para fins de modelagem, agregada em séries mensais por tipo de acomodação, reduzindo ruídos brancos de alta frequência em comparação a uma periodicidade mais curta (quinzenal ou semanal).

Antes de quaisquer manipulações nos dados é de suma importância entender como está distribuído na linha do tempo o volume de registros de pacientes-dia. Segue abaixo a distribuição mensal de tais registros.

Gráfico 1 - Distribuição de Pacientes-dia por Mês/Ano de Ocupação



Fonte: Elaborado pelo autor (2026).

Nota-se o crescimento no volume de pacientes-dia acessando as unidades de internação no decorrer dos anos, independente dos vestígios da fase final da covid, onde esperava-se um quantitativo acima do observado. Para isso, explica-se o fato que nos anos

de 2024 e 2025 houve um aumento vertiginoso na capacidade de leitos operacionais, devido a uma reforma estrutural, resultando na criação de novos setores de internação e ampliação do espaço físico de alguns setores e acomodações já existentes.

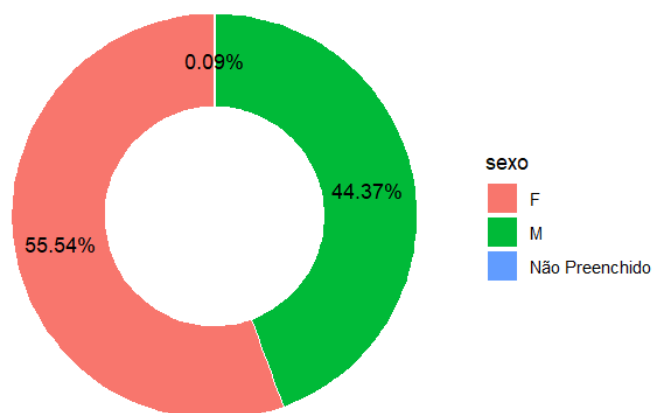
6.2 Análise Descritiva do Perfil de Ocupação

A etapa inicial de tratamento dos dados (ETL) incluiu a padronização de categorias, em especial referente a ausência de valores nas colunas categóricas CID, clínica, sexo e convênio (por exemplo, substituição de “NA” para “Não Preenchido”) e a verificação de estrutura e consistência das variáveis, garantindo a integridade mínima necessária para a análise exploratória (EDA) e preditiva. Esse pré-processamento viabilizou a criação de *dataframes* específicos para exploração do perfil por sexo, idade, setor, convênio, clínica e tipo de acomodação, além de agrupamentos entre as variáveis categóricas na visão por tipo de acomodação. Cada variável será abordada em subtópicos, a seguir.

6.2.1 Distribuição por Sexo e Idade

A distribuição de registro de pacientes por sexo mostra leve predominância do sexo masculino no conjunto geral, com percentuais apresentados em gráfico de rosca, conforme o gráfico 2 abaixo:

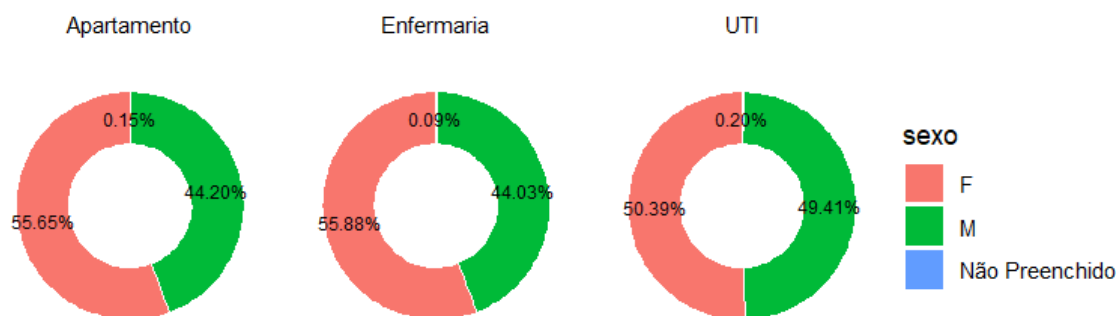
Gráfico 2 - Percentual de Registro de Ocupação por Sexo



Fonte: Elaborado pelo autor (2026).

Quando o total percentual do registro de ocupação por sexo é comparado ao percentual estratificado por tipo de acomodação, essa diferença se mantém com pequenas variações, evidenciando ausência de distorção na proporção de gênero por tipo de acomodação. Considerando os totais de registros de ocupação (Apartamento = 16.350, Enfermaria = 11.794 e UTI = 8.013) segue proporção por sexo e acomodação, ilustrado no gráfico 3.

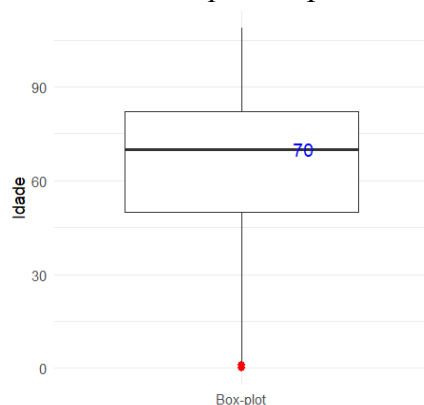
Gráfico 3 – Percentual de Registros de Ocupação por Sexo e Tipo de Acomodação



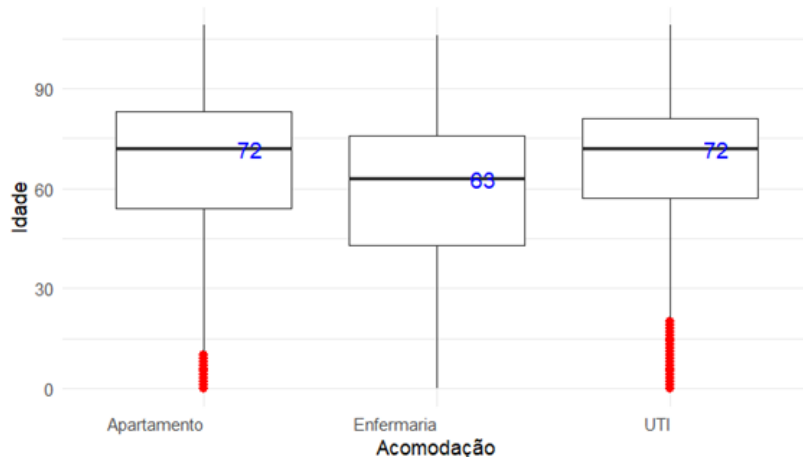
Fonte: Elaborado pelo autor (2026).

Em relação à idade, a base de dados possui 7.898 registros faltantes (NA's), que corresponde a aproximadamente 3,31% do total, sendo que, para análise tais registros foram omitidos. Como resultado, nota-se que a distribuição dos dados no geral é assimétrica, com mediana de 54,5 anos na tabela de pacientes-dia por idade e presença de valores extremos até 109 anos. O *box-plot* geral da idade (gráfico 4) mostra grande dispersão, enquanto o *box-plot* agrupado por tipo de acomodação (gráfico 5) indica que a UTI tende a concentrar pacientes mais idosos, ao passo que o Apartamento e Enfermaria apresentam maior heterogeneidade etária.

Gráfico 4 - *Box-plot* de pacientes-dia por idade



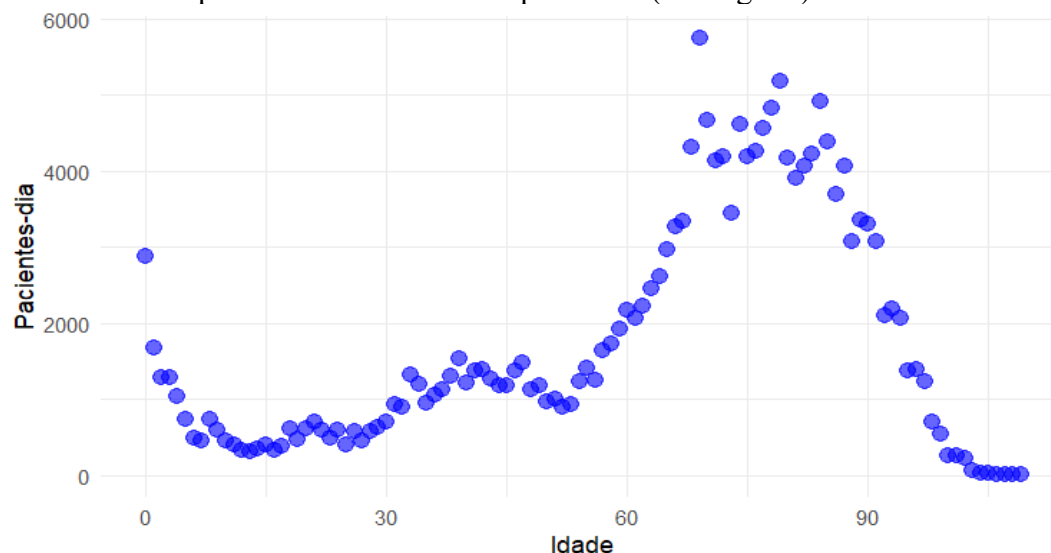
Fonte: Elaborado pelo autor (2026).

Gráfico 5 - *Box-plot* de pacientes-dia por idade e Tipo de Acomodação

Fonte: Elaborado pelo autor (2026).

Os gráficos de dispersão de idade versus pacientes-dia, tanto na visão geral quanto por tipo de acomodação, evidenciam que algumas faixas etárias (idosos com aproximadamente 60 a 90 anos e os primeiros anos de vida) concentram volume expressivo de pacientes-dia e de internações, reforçando a importância de realizar ações específicas para esses perfis (pediatria e pessoa idosa).

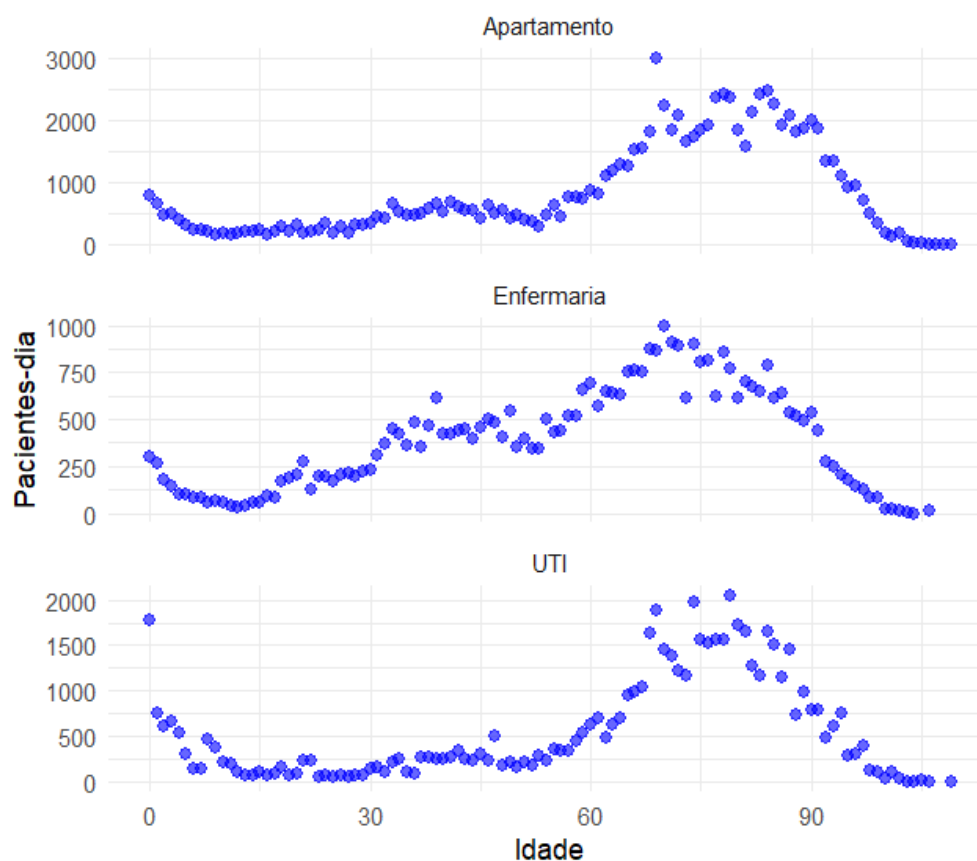
Gráfico 6 – Dispersão dos Pacientes-dia por Idade (visão geral)



Fonte: Elaborado pelo autor (2026).

Ainda em comparação ao gráfico 7 a seguir, a acomodação enfermaria possui uma dispersão de pacientes-dia um pouco mais heterogênea e ampla, mesmo com um pico acentuado nas idades mais avançadas, em comparação ao apartamento e UTI, sinalizando um possível comportamento ocupacional e perfil de pacientes diferentes.

Gráfico 7 – Dispersão dos Pacientes-dia por Idade e Tipo de Acomodação

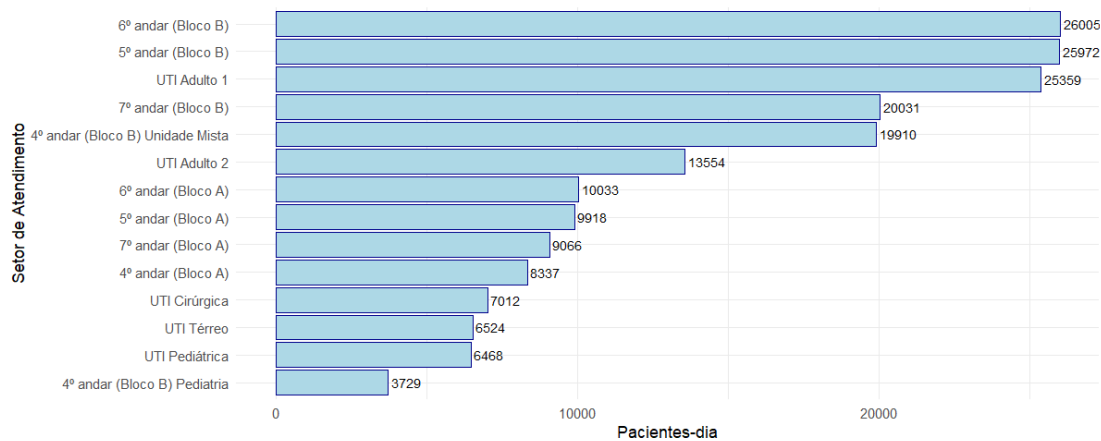


Fonte: Elaborado pelo autor (2026).

6.2.2 Setor de Atendimento, Convênio e Clínica

A agregação de pacientes-dia por setor de atendimento (gráfico 8) mostra que poucos setores concentram a maior parte da demanda, com 6º e 5º andares do Bloco B, UTI Adulto 1 e 7º andar do Bloco B entre os principais contribuintes.

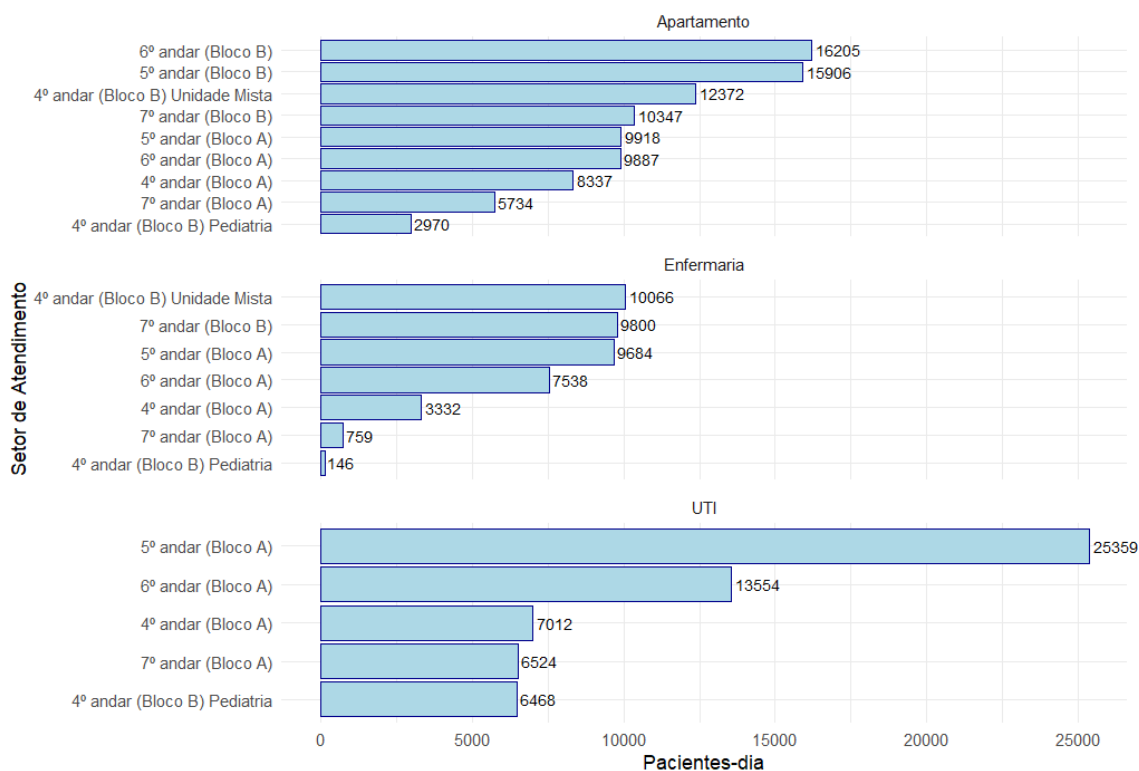
Gráfico 8 – Pacientes-dia por Setor de Atendimento



Fonte: Elaborado pelo autor (2026).

Quando essa análise é estratificada por tipo de acomodação (gráfico 9), observa-se que o Apartamento concentra a maior parte dos pacientes-dia em andares de internação convencional, em comparação aos pacientes dos mesmos setores que utilizaram leitos do tipo Enfermaria, enquanto a UTI reúne os setores de terapia intensiva adulto, cirúrgica, pediátrica e tórrea.

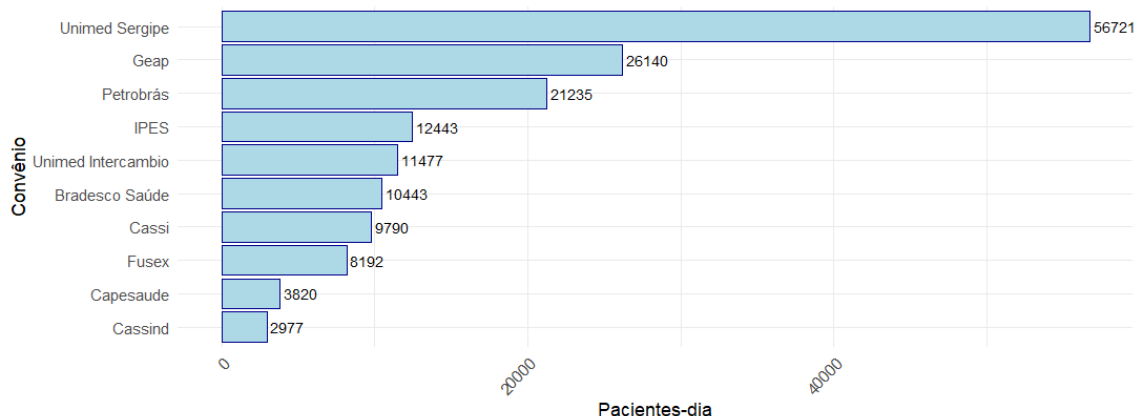
Gráfico 9 – Pacientes-dia por Setor de Atendimento e Tipo de Acomodação



Fonte: Elaborado pelo autor (2026).

De forma semelhante, a distribuição por convênio revela que alguns poucos convênios (por exemplo, grandes operadoras e convênios próprios) respondem por parcela relevante da carga assistencial (gráfico 10), com variações na participação relativa em cada tipo de acomodação.

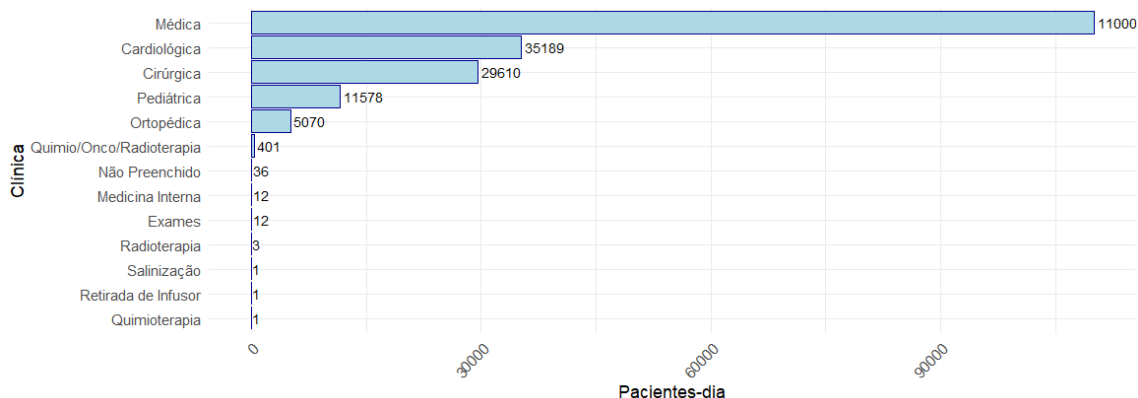
Gráfico 10 – Top 10 Quantidade Pacientes-dia por Convênio



Fonte: Elaborado pelo autor (2026).

As análises por clínica (gráfico 11) evidenciam que linhas de cuidado como clínica médica, cirúrgica, cardiológica, pediátricas e ortopédicas concentram a maior parte dos pacientes-dia na Enfermaria e no Apartamento e em especial a clínica pediátrica, com sua maioria povoando a Ala Pediátrica da UTI.

Gráfico 11 – Quantidade Pacientes-dia por Clínica

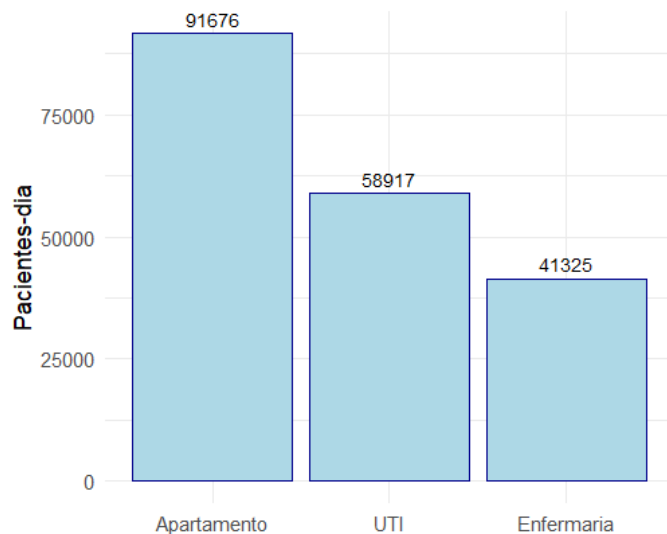


Fonte: Elaborado pelo autor (2026).

6.2.3 Segmentação por Tipo de Acomodação

A segmentação por tipo de acomodação revela diferenças marcantes de perfil operacional entre Apartamento, Enfermaria e UTI, onde existe uma predominância na ocupação em acomodações em Apartamento, mesmo que possua poucos setores a mais disponíveis que as outras acomodações. Tal número mostra o volume de uso e possível permanência dos pacientes acima da ocupação em enfermaria, conforme nota-se no gráfico 12 abaixo:

Gráfico 12 – Quantidade Pacientes-dia por Tipo de Acomodação



Fonte: Elaborado pelo autor (2026).

6.3 Análise da Taxa de Ocupação Hospitalar em Série Temporal

A taxa de ocupação hospitalar tem sua periodicidade definida a depender dos seus objetivos de uso. Em um primeiro momento, buscou-se entender qual melhor periodicidade de tempo explicaria melhor o comportamento das séries temporais e para isso foi testado séries temporais mensais, quinzenais e semanais. No segundo momento se era necessário granularizar a série por alguma variável ou categoria que trouxesse sentido explicativo e estabilização nos componentes da série.

Após avaliação das diferentes granularidades temporais sugeridas, optou-se por utilizar as **séries mensais de taxa de ocupação, granularizadas por tipo de acomodação**. Essa escolha se justifica por quatro motivos principais:

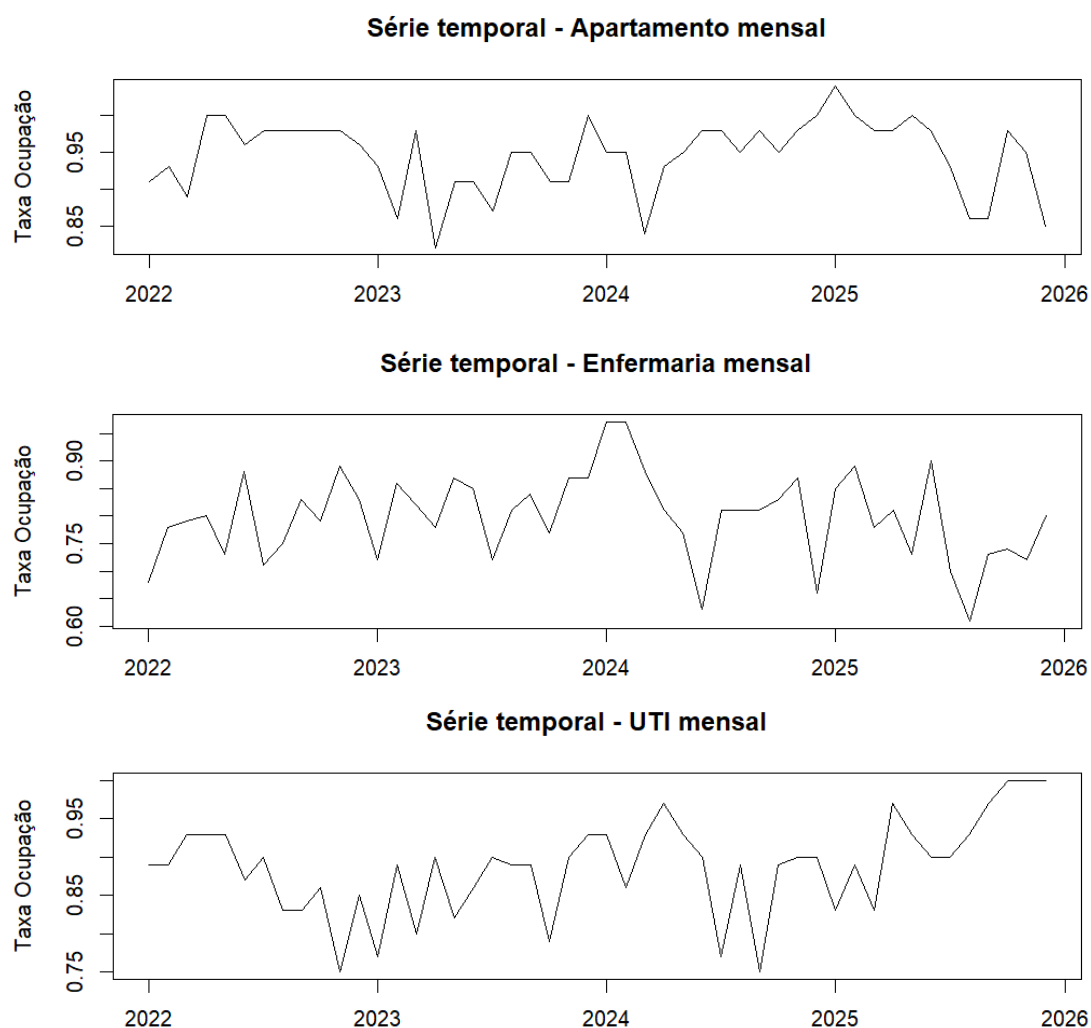
- a) Alinhamento com o horizonte tático de planejamento de recursos hospitalares, usualmente definido em escala mensal.
- b) Redução do ruído de alta frequência observado nas séries semanais e quinzenais, o que permitiu identificar com maior clareza componentes de tendência e sazonalidade, além de melhorar os diagnósticos futuros dos modelos a serem testados.
- c) As decisões de capacidade e alocação de recursos são tomadas por tipo de leito, e não apenas no total de ocupação a nível institucional, além de cada tipo de acomodação apresentar padrões distintos de demanda e permanência,

conforme discutido na literatura de previsão de ocupação de leitos hospitalares (OVERMAN et al., 2024; ZHAO et al., 2024; MELLO et al., 2025).

- d) A série em visão generalista pode mascarar situações críticas em segmentos específicos, especialmente na UTI, onde a superlotação tem impacto direto na segurança do paciente (FINDLAY, 2019; ZHAO et al., 2024).

As séries de taxa de ocupação mensais (gráfico 13) mostram níveis médios distintos por acomodação no decorrer do tempo: a UTI apresenta valores de ocupação elevados e relativamente estáveis, a Enfermaria exibe uma dinâmica intermediária com variações associadas a internações clínicas e cirúrgicas, e o Apartamento reflete mais fortemente a demanda eletiva e conveniada.

Gráfico 13 – Taxa de Ocupação Histórica Mensal por Tipo de Acomodação



Fonte: Elaborado pelo autor (2026).

Por cada tipo de acomodação possuir perfil clínico, admissão e dinâmica de uso distinto por paciente, conforme documentado na literatura de gestão hospitalar (GLLEDDSON, 2021; FINDLAY, 2019) recomenda-se a estratégia metodológica de modelar cada tipo de acomodação separadamente, tanto na abordagem clássica (ARIMA e ETS ou *Holt-Winters*) quanto em métodos de aprendizagem de máquina (*XGBoost*), de forma a capturar padrões específicos de cada contexto assistencial e de acomodação.

6.4 Modelagem ARIMA

6.4.1 Estacionariedade e Decomposição

Para as séries mensais de taxa de ocupação por tipo de acomodação, foram realizados testes de *Dickey-Fuller* aumentado (tabela 1), usando para nosso teste de hipótese nível de significância de 0,05 ou 5%, indicando a necessidade de diferenciação de ordem 1 ($d = 1$) para obtenção de estacionariedade em todas as séries.

Tabela 1 – Teste de Dickey-Fuller Aumentado das Séries Mensais (p -valor $\leq 0,05$)

Série	Dickey Fuller	P-Valor	Lags	Conclusão
Apartamento - Mensal	-1,9219	0,6053	3	Evidência de Não Estacionariedade
Enfermaria - Mensal	-3,1185	0,1274	3	Evidência de Não Estacionariedade
UTI - Mensal	-2,2774	0,4634	3	Evidência de Não Estacionariedade

Fonte: Elaborado pelo autor (2026).

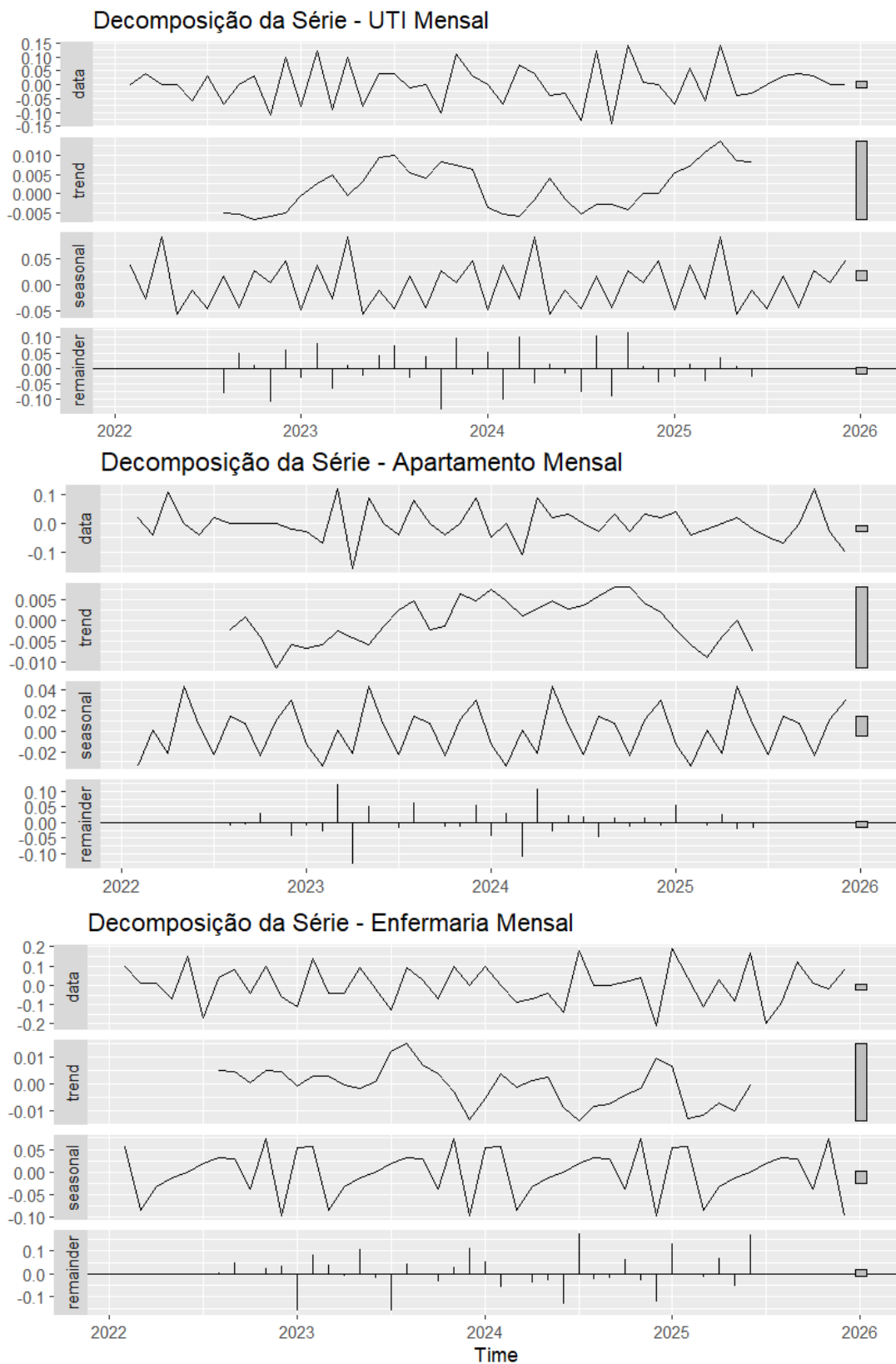
Após aplicação de diferenciação nos modelos em questão ($d = 1$), segue nova avaliação do teste de *Dickey-Fuller* aumentado, conforme tabela 2 abaixo:

Tabela 2 – Teste de Dickey-Fuller Aumentado das Séries Mensais Diferenciadas ($d = 1$)

Série	Dickey Fuller	P-Valor	Lags	Conclusão
Apartamento - Mensal	-5,077	0,01	3	Evidência de Estacionariedade
Enfermaria - Mensal	-4,281	0,01	3	Evidência de Estacionariedade
UTI - Mensal	-5,408	0,01	3	Evidência de Estacionariedade

Fonte: Elaborado pelo autor (2026).

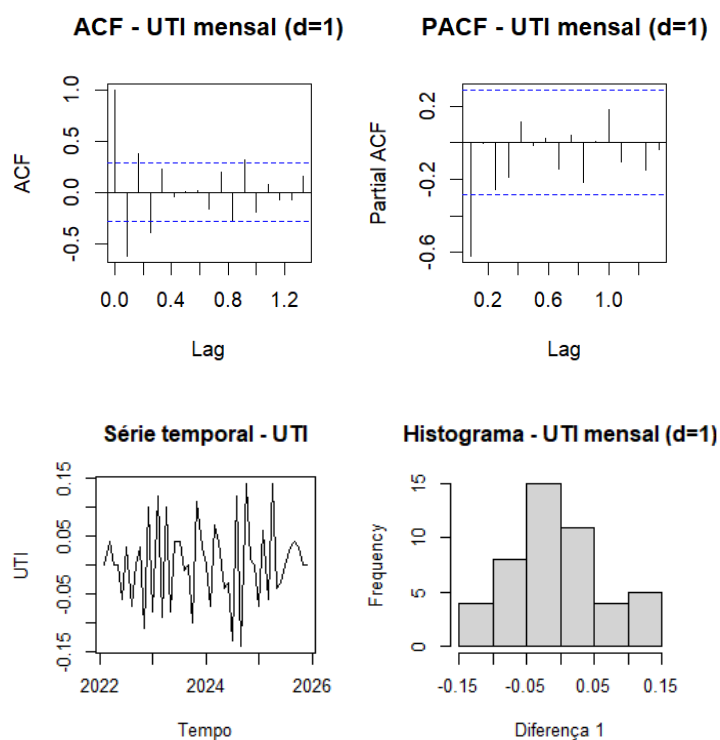
As séries diferenciadas foram decompostas em componentes de tendência, sazonalidade e ruído, evidenciando padrões sazonais e estrutura temporal aparentemente adequados à modelagem com ARIMA, ilustrado no gráfico 14 a seguir:

Gráfico 14 – Decomposição das Séries Históricas Mensais Diferenciadas ($d = 1$)

Fonte: Elaborado pelo autor (2026).

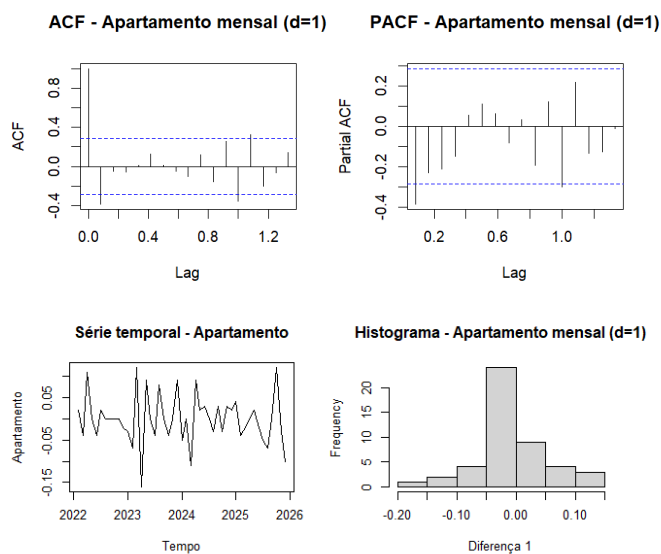
Após decomposição, foram analisadas as funções de autocorrelação e autocorrelação parcial das séries diferenciadas, juntamente com os gráficos de histograma e série histórica, para orientar a escolha inicial dos parâmetros (p , d , q), complementadas pela avaliação sistemática com AIC e BIC, conforme os gráficos 15, 16 e 17 a seguir:

Gráfico 15 – ACF, PACF, Histograma e Série Histórica - UTI



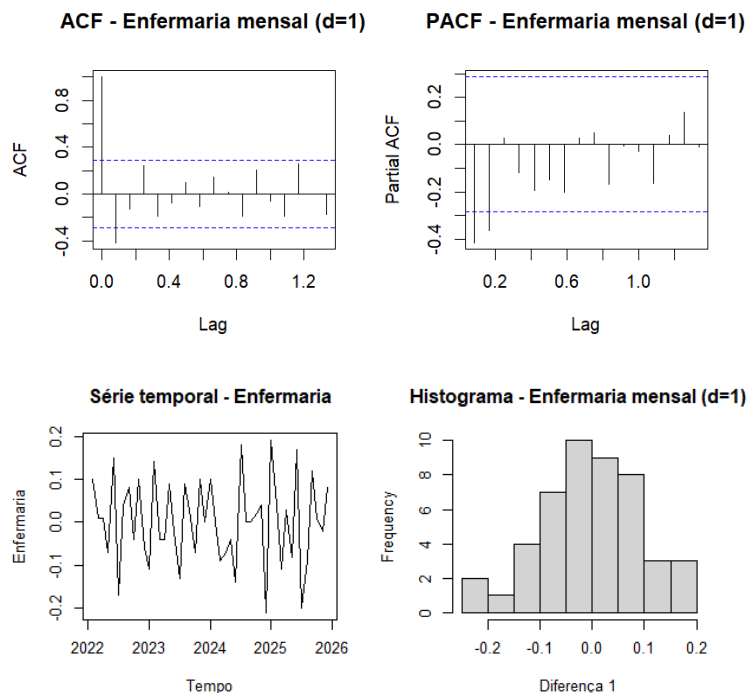
Fonte: Elaborado pelo autor (2026).

Gráfico 16 – ACF, PACF, Histograma e Série Histórica - Apartamento



Fonte: Elaborado pelo autor (2026).

Gráfico 17 – ACF, PACF, Histograma e Série Histórica - Enfermaria



Fonte: Elaborado pelo autor (2026).

6.4.2 Ajuste e Diagnóstico dos Modelos

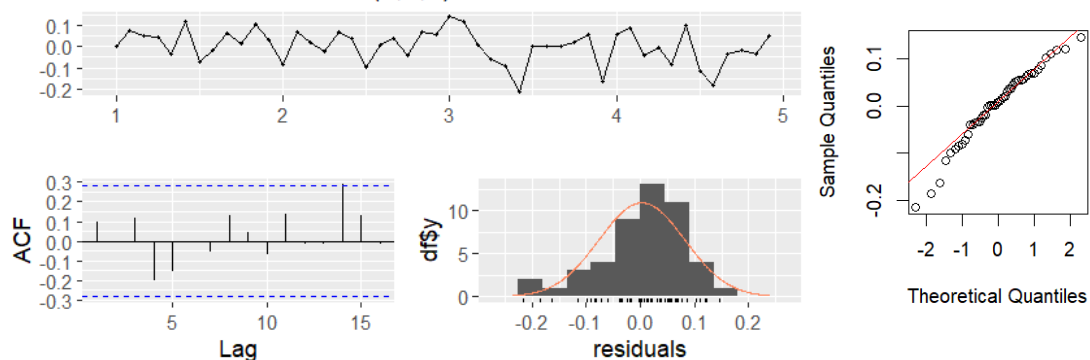
Com base nos critérios supracitados, foram ajustados modelos ARIMA para cada tipo de acomodação, usando a função *auto.arima* e selecionadas as ordens com melhor desempenho de informação dentre elas, chegando a modelos como ARIMA(0,1,1) para Apartamento e Enfermaria e ARIMA(1,1,0) para UTI, com métricas expostas na tabela 3 a seguir:

Tabela 3 – Métricas de Desempenho e Avaliação das Séries Diferenciadas

Série	AIC	BIC	RMSE	MAE	MAPE
Apartamento - Mensal	-147,67	-142,12	0,0452	0,0366	3,9340
Enfermaria - Mensal	-99,72	-96,02	0,0784	0,0612	7,9198
UTI - Mensal	-135,26	-131,56	0,0541	0,0430	4,9487

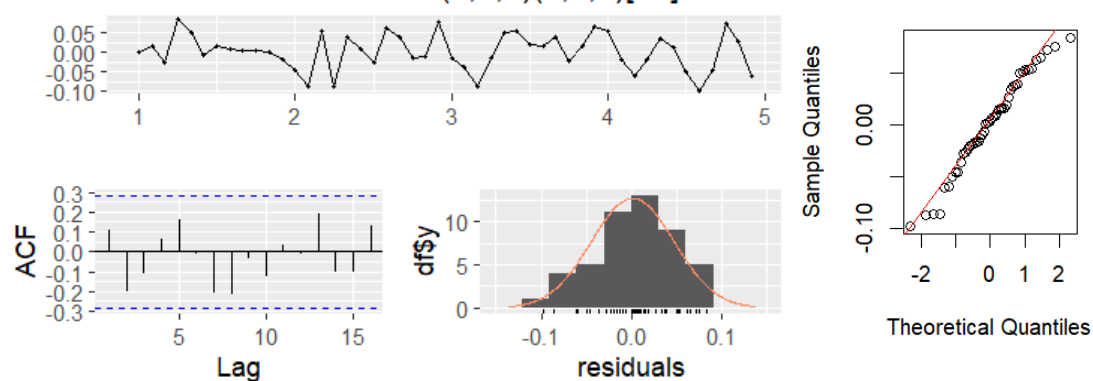
Fonte: Elaborado pelo autor (2026).

Gráfico 18 – Avaliação, ACF, Histograma e QQ-Plot dos Resíduos ARIMA – Enfermaria
Residuals from ARIMA(0,1,1)



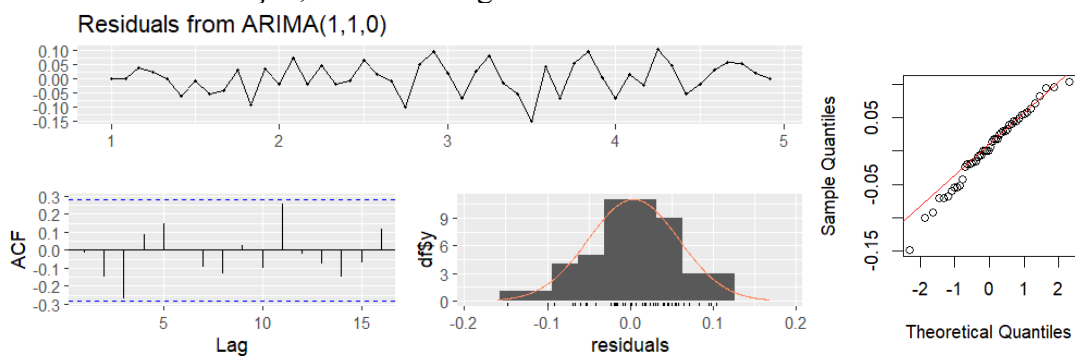
Fonte: Elaborado pelo autor (2026).

Gráfico 19 – Avaliação, ACF e Histograma dos Resíduos ARIMA – Apartamento
Residuals from ARIMA(0,1,1)(0,0,1)[12]



Fonte: Elaborado pelo autor (2026).

Gráfico 20 – Avaliação, ACF e Histograma dos Resíduos ARIMA – UTI
Residuals from ARIMA(1,1,0)



Fonte: Elaborado pelo autor (2026).

A análise de resíduos, incluindo os visuais referente aos resíduos (gráficos 18, 19 e 20) dos resíduos e testes de normalidade de Shapiro–Wilk (tabela 4), indicou ausência de autocorrelação remanescente relevante e aderência aceitável à normalidade dos resíduos para os modelos escolhidos.

Tabela 4 - Teste Shapiro-Wilk para os Resíduos – Modelo ARIMA

Série	Estatística do teste	P-Valor
Apartamento - Mensal	0,96498	0,16
Enfermaria - Mensal	0,97555	0,41
UTI - Mensal	0,98098	0,62

Fonte: Elaborado pelo autor (2026).

6.5 Modelagem Holt-Winters Aditiva – ETS (A,N,A)

6.5.1 Ajuste dos Modelos

Além dos modelos ARIMA, observando o comportamento do erro e aparente picos sazonais consistentes nas séries observadas, foi implementado o modelo de *Holt-Winters* Aditivo para as séries mensais de taxa de ocupação de Enfermaria, Apartamento e UTI. Os modelos foram ajustados com base em séries mensais com índice temporal calibrado com todo histórico disponível, utilizando o modelo de Suavização Exponencial Aditiva (A,N,A), correspondente ao erro/nível e sazonalidade aditivos sem componente de tendência.

Para Enfermaria, Apartamento e UTI, os relatórios de ajuste apresentam parâmetros de suavização α e γ em faixas moderadas, além de estados iniciais de nível e sazonalidade coerentes com o padrão observado de ocupação mensal. O critério de informação (AIC e BIC) indica ajustes competitivos, e os erros de treinamento (RMSE, MAE e MAPE) são calculados para cada acomodação, compondo o bloco de métricas ETS, explicitado a seguir (tabela 5):

Tabela 5 – Parâmetros, Métricas de Desempenho e Avaliação – Modelo ETS(A,N,A)

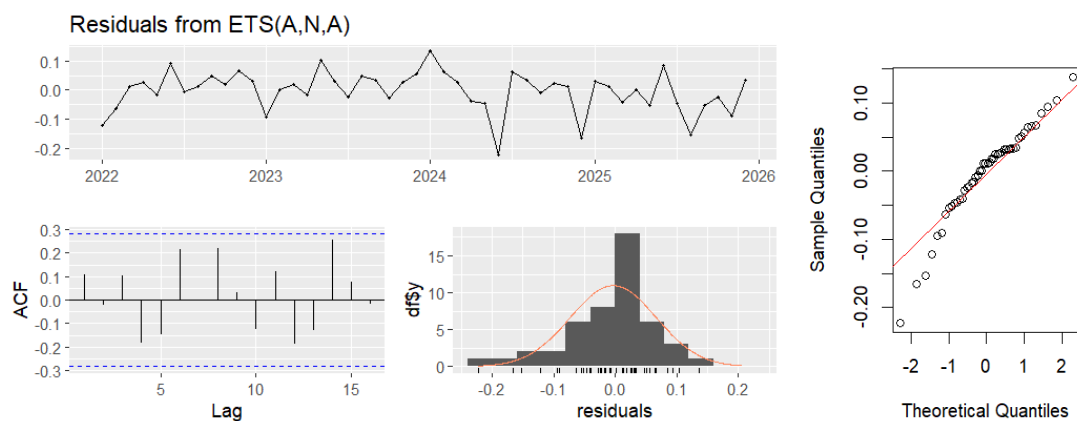
Série	Alpha α	Gamma γ	Sigma	AIC	BIC	RMSE	MAE	MAPE
Apartamento - Mensal	0,0001	0,0001	0,057	-75,32	-47,25	0,0481	0,0395	4,2652
Enfermaria - Mensal	0,1169	0,0001	0,082	-40,80	-12,73	0,0690	0,0515	6,7605
UTI - Mensal	0,2861	0,0001	0,059	-72,40	-44,33	0,0496	0,0406	4,6935

Fonte: Elaborado pelo autor (2026).

6.5.2 Diagnóstico dos Resíduos, Testes de *Ljung-Box* e *Shapiro-Wilk*

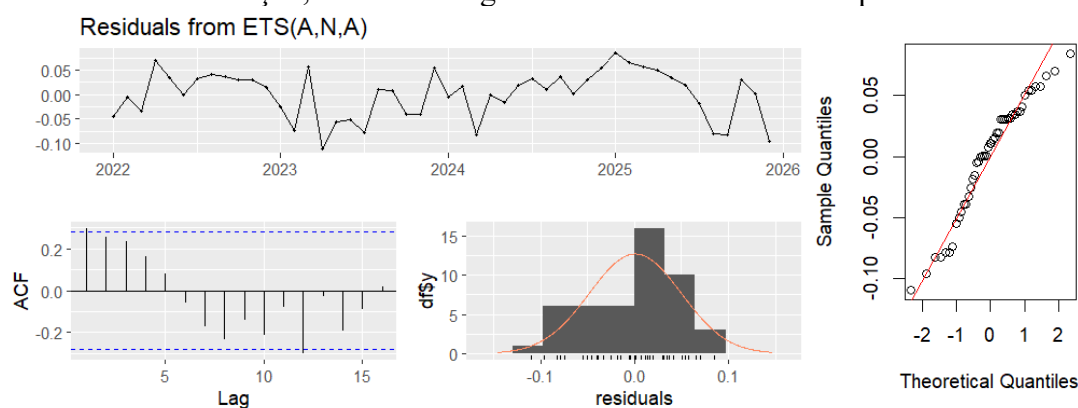
A avaliação de resíduos dos modelos incluiu gráficos de resíduos, funções de autocorrelação (gráficos 18, 19 e 20), além da aplicação dos testes de *Ljung-Box* e de *Shapiro-Wilk* (tabela 6).

Gráfico 21 – Avaliação, ACF, Histograma e QQ-Plot dos Resíduos ETS - Enfermaria



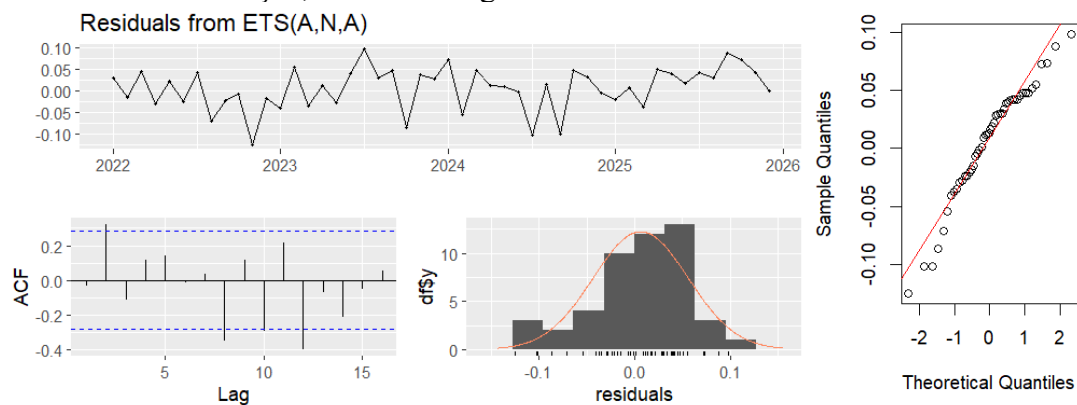
Fonte: Elaborado pelo autor (2026).

Gráfico 22 – Avaliação, ACF e Histograma dos Resíduos ETS – Apartamento



Fonte: Elaborado pelo autor (2026).

Gráfico 23 – Avaliação, ACF e Histograma dos Resíduos ETS – UTI



Fonte: Elaborado pelo autor (2026).

Tabela 6 – Teste *Ljung-Box* e *Shapiro-Wilk* para os Resíduos – Modelo ETS(A,N,A)

Série	Ljung-Box		Shapiro-Wilk
	Lag	P-Valor	P-Valor
Apartamento - Mensal	10	0,0143	0,0333
Enfermaria - Mensal	10	0,3630	0,0229
UTI - Mensal	10	0,1509	0,0599

Fonte: Elaborado pelo autor (2026).

Conforme os dados acima, para as séries Apartamento e Enfermaria, assumindo um nível de significância de 5%, não existe evidência estatística de que os resíduos seguem estritamente uma distribuição normal, diferente da UTI. No entanto, os resultados do teste de *Ljung-Box* demonstram que o Apartamento e UTI possuem indícios de rejeição da hipótese de independência, o que sugere que os modelos não extraíram por inteiro os padrões de dependência temporal da série, demonstrando pequenos traços de autocorrelação residual em defasagens mais baixas, o que felizmente não ocorre na série da Enfermaria ($p\text{-valor} = 0,3630 > 0,05$).

Para a realidade acima, um fato externo importante que ocorreu durante os anos de 2024 e 2025 deve ser levado em consideração, sendo ele a expansão física do hospital (aumento de leitos) e inauguração de novos setores de atendimento. Nisso, a série Enfermaria tem mais evidências de poder permanecer como boa candidata, diferente das outras duas séries para esse modelo.

6.6 Modelagem *XGBoost*

6.6.1 Preparação das Features e Ajuste do Modelo

A aplicação do modelo de regressão *XGBoost* foi construída a partir da base mensal de taxa de ocupação por tipo de acomodação. Nesse contexto, a série temporal de taxa de ocupação foi transformada em um problema de aprendizagem supervisionada, por meio da criação de variáveis explicativas de tempo (mês e ano) e de variáveis defasadas de *lags* de curto prazo (*lag1* e *lag2*) e *lag* sazonal (*lag12*), além de descartar observações sem *lags* completo.

Para cada tipo de acomodação, foi dividido as observações em conjuntos de treino (85% das observações iniciais) e 15% de teste (observações finais), preservando a ordem temporal. Os modelos foram treinados com função de perda quadrática, taxa de aprendizado = 0,1, profundidade máxima = 3 e amostragens parcial de linhas e colunas, com parada antecipada, evitando possível *overfitting*.

6.6.2 Aplicação da Modelagem por Tipo de Acomodação

Com a geração das previsões do modelo *XGBoost* no conjunto de teste, os resultados foram utilizados para calcular as métricas de MAE, RMSE e MAPE por tipo de acomodação (tabela 7).

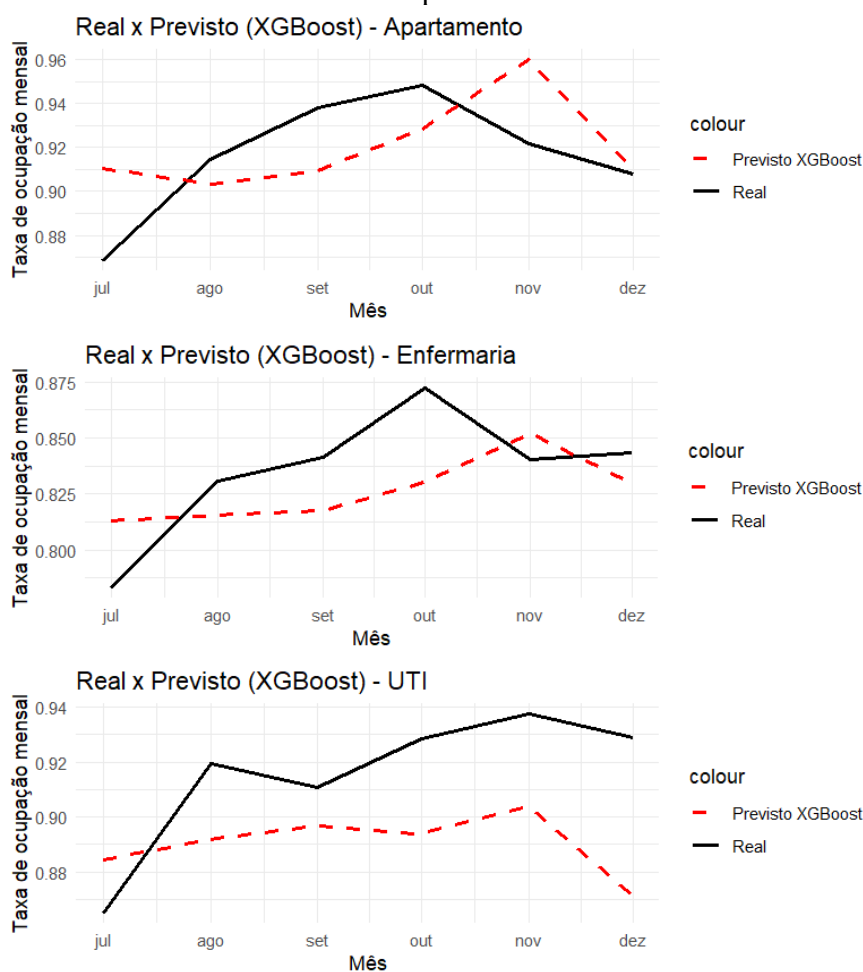
Tabela 7 – Métricas de Desempenho e Avaliação – Modelo *XGBoost*

Série	RMSE	MAE	MAPE
Apartamento - Mensal	0,0236	0,0276	2,59
Enfermaria - Mensal	0,0228	0,0252	2,72
UTI - Mensal	0,0311	0,0341	3,37

Fonte: Elaborado pelo autor (2026).

No período de uso do teste (jan. a dez. 2025) para geração do gráfico 21, indicam que o modelo captura bem as variações mensais de taxa de ocupação em Apartamento e UTI, aproximando-se dos picos e vales de forma levemente responsiva.

Gráfico 24 – Real x Previsto – Apartamento – Julho à Dezembro 2025



Fonte: Elaborado pelo autor (2026).

6.7 Comparação ARIMA vs. Holt-Winters Aditivo vs. XGBoost

Com as métricas consolidadas, foi construída uma tabela consolidada por tipo de acomodação e os modelos utilizados (ARIMA, ETS(A,N,A) e XGBoost) juntamente com métricas de decisão MAE, RMSE e MAPE calculadas sobre o período total da série (2022–2025). Essa tabela permite uma comparação direta entre os três métodos, evidenciando qual abordagem é mais adequada a cada série (tabela 8, 9 e 10).

Tabela 8 – Comparação ARIMA vs EST vs XGBOOST – Apartamento

Série	Modelo	MAE	RMSE	MAPE
Apartamento	ARIMA	0.0367	0.0452	3.93
Apartamento	ETS(A,N,A)	0.0396	0.0482	4.27
Apartamento	XGBoost	0.0236	0.0276	2.59

Fonte: Elaborado pelo autor (2026).

Tabela 9 – Comparação ARIMA vs EST vs XGBOOST – Enfermaria

Série	Modelo	MAE	RMSE	MAPE
Enfermaria	ARIMA	0.0613	0.0784	7.92
Enfermaria	ETS(A,N,A)	0.0515	0.0690	6.76
Enfermaria	XGBoost	0.0228	0.0252	2.72

Fonte: Elaborado pelo autor (2026).

Tabela 10 – Comparação ARIMA vs EST vs XGBOOST – UTI

Série	Modelo	MAE	RMSE	MAPE
UTI	ARIMA	0.0431	0.0542	4.95
UTI	ETS(A,N,A)	0.0406	0.0497	4.69
UTI	XGBoost	0.0311	0.0341	3.37

Fonte: Elaborado pelo autor (2026).

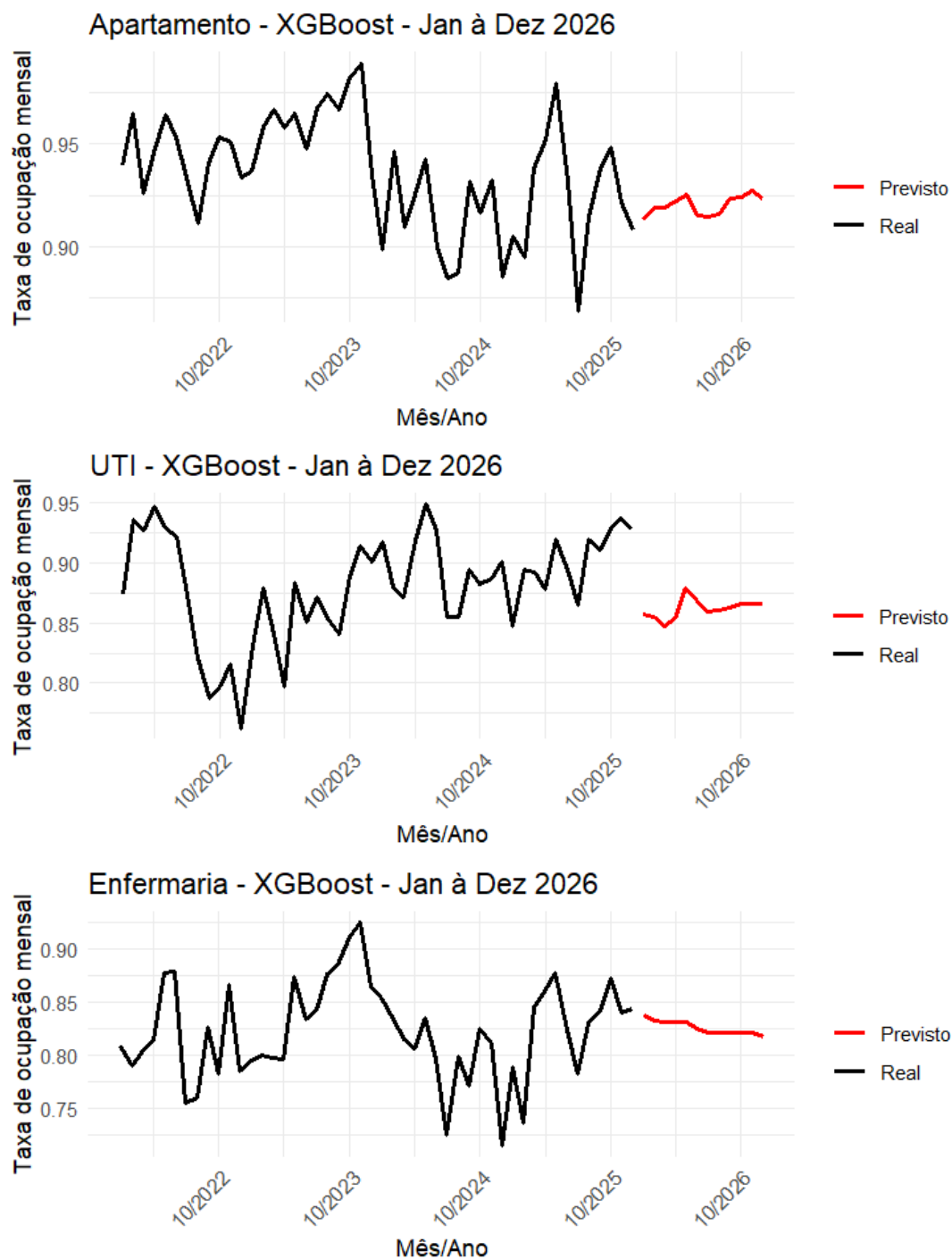
Os resultados mostram que para **Enfermaria**, **Apartamento** e **UTI**, o *XGBoost* obtém MAPE inferior aos modelos ARIMA e ETS(A,N,A), indicando melhor capacidade de capturar e prever a dinâmica de ocupação do hospital.

6.8 Projeções da Taxa de Ocupação para 2026

Com base nos modelos *XGBoost* selecionados para Apartamento, Enfermaria e UTI, foram geradas projeções da taxa de ocupação mensal para os 12 meses de 2026, utilizando abordagem iterativa com “lag1” para incorporar o efeito das previsões nos meses subsequentes. As projeções resultantes foram combinadas com o histórico para construir séries estendidas para cada acomodação, nos quais se observa a continuidade do padrão recente de ocupação, sem rupturas abruptas para o Apartamento e UTI. Já para a Enfermaria, o desempenho do *XGBoost* é competitivo em termos de *MAPE*, porém como

visto no gráfico 22, as previsões formam praticamente uma linha quase horizontal com pouca variação, bem mais “suavizada” que a série real, que é mais ondulada e com picos e vale, sugerindo que o *XGBoost* está sub-representando a sazonalidade/volatilidade e empurrando a previsão para algo próximo da média recente.

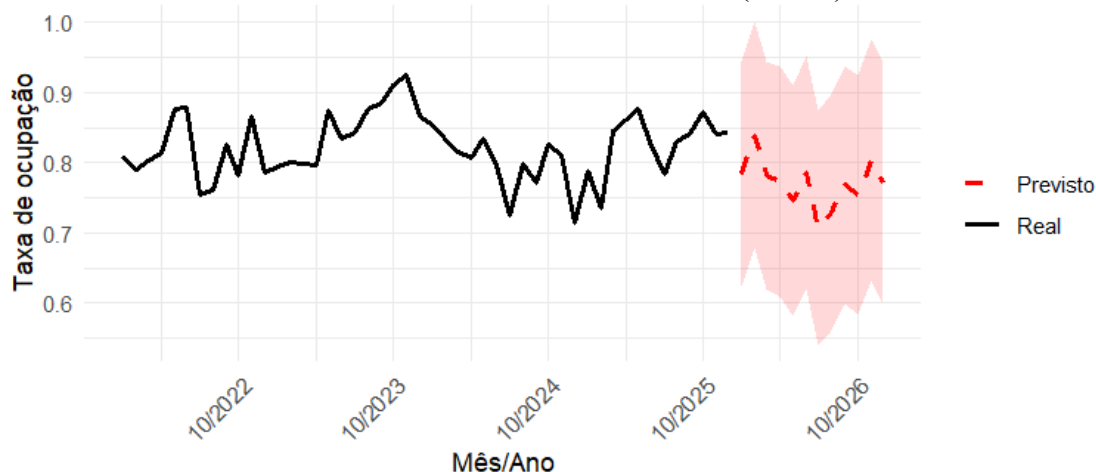
Gráfico 25 – Previsão das Séries por Tipo de Acomodação – Jan à Dez 2026



Fonte: Elaborado pelo autor (2026).

Todavia, como veremos a seguir no gráfico 23, o modelo ETS(A,N,A) com segundo menor MAPE se mostra superior em projeção, pois ele segue um patamar e variação semelhantes aos últimos meses observados, e o intervalo de confiança (95%) cobre razoavelmente a amplitude histórica da série, o que indica que o modelo está captando bem a sazonalidade, o que fortalece a escolha de tal modelo, dentre os avaliados, como preferencial na previsão da ocupação para a Enfermaria.

Gráfico 26 – Previsão da Série Enfermaria – Modelo ETS(A,N,A) – Jan à Dez 2026



Fonte: Elaborado pelo autor (2026).

As projeções acima indicam manutenção de um patamar de ocupação semelhante ao observado no final de 2025, com alguma variabilidade sazonal, oferecendo um cenário plausível para o planejamento de recursos na Enfermaria ao longo de 2026.

Essa análise comparativa cumpre o objetivo de validação e seleção do modelo mais adequado por tipo de acomodação, demonstrando que não há um único modelo “ótimo” para todas as séries, mas sim um portfólio de métodos ajustado ao comportamento de cada segmento.

7 CONCLUSÃO

Os resultados da análise descritiva atendem ao primeiro objetivo específico, ao caracterizar padrões de ocupação por idade, setor, convênio, clínica e tipo de acomodação, evidenciando concentrações de demanda e perfis distintos de pacientes entre Apartamento, Enfermaria e UTI. A identificação de setores, convênios e clínicas que concentram a maior parte dos pacientes-dia fornece insumos para priorização de esforços de gestão e negociação.

No campo da modelagem, os modelos ARIMA atenderam ao objetivo de implementação de uma abordagem clássica de séries temporais, com testes de estacionariedade, seleção de ordens (p, d, q) por AIC/BIC e diagnóstico de resíduos, fornecendo um *benchmark* estatístico sólido para comparação. A introdução do ETS(A,N,A) como segundo método de séries temporais ampliou o leque de modelos lineares, revelando uma previsão superior na captação da sazonalidade na Enfermaria e demonstrando a relevância de modelos de suavização exponencial para séries com forte componente sazonal e menos estrutura de tendência.

Os modelos *XGBoost* cumpriram o objetivo de aplicação de técnicas de aprendizado de máquina para previsão de curto prazo, com construção explícita de *lags* temporais e variáveis de calendário, particionamento temporal de treino-teste e avaliação por MAE, RMSE e MAPE. A comparação quantitativa entre os modelos testados, evidenciou que o *XGBoost* é o modelo mais adequado para Apartamento e UTI, enquanto ETS(A,N,A) é preferível para Enfermaria, reforçando a abordagem de seleção segmentada por tipo de acomodação.

Por fim, as projeções para 2026 atendem ao objetivo de fornecer recomendações acionáveis, ao antecipar a taxa de ocupação em cada segmento e em curto e médio prazo, apoiando decisões sobre dimensionamento de leitos, recursos humanos e gestão de agendas. Dessa forma, o estudo oferece uma base quantitativa consistente para o planejamento hospitalar, articulando análise exploratória, modelagem estatística e aprendizado de máquina em um fluxo integrado.

BIBLIOGRAFIA

ANS. Agência Nacional de Saúde Suplementar. **Ficha técnica de indicadores hospitalares**. Rio de Janeiro, 2010. Disponível em: <https://www.gov.br/ans/pt-br/arquivos/assuntos/prestadores/qualiss-programa-de-qualificacao-dos-prestadores-de-servicos-de-saude-1/Ficha_tecnica_indicador_5.pdf>. Acesso em: 14 fev. 2026.

ANS. Agência Nacional de Saúde Suplementar. **Mapa de utilização do SUS por beneficiários de planos privados de saúde**. Agência Nacional de Saúde Suplementar (Brasil). Rio de Janeiro: ANS, 2025.

BITTENCOURT, R. J.; HORTALE, V. A. **Dificuldades de acesso e estimativas de leitos públicos em terapia intensiva no Brasil**. Revista de Saúde Pública, São Paulo, v. 50, n. 19, 2016.

BORIN, M. C. et al. **ID 217 - Modelo preditivo para internações por doenças respiratórias e sua Sazonalidade, no Contexto Pós-Pandêmico: estudo de série temporal**. Revista do SUS - RESS, p. 139, 2025.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis: Forecasting and Control**. 5. ed. Hoboken: Wiley, 2016.

BRASIL. Ministério da Saúde. **Padronização da nomenclatura do censo hospitalar**. 2ª edição. Brasília, 2002.

BRASIL. Ministério da Saúde. **Terminologia básica em saúde**. 2. ed. Brasília, 2004. Disponível em: <<https://bvsmms.saude.gov.br/bvs/publicacoes/0113terminologia3.pdf>>. Acesso em: 14 fev. 2026.

BROWN, R. G. **Smoothing, Forecasting and Prediction of Discrete Time Series**. Englewood Cliffs: Prentice-Hall, 1963.

CHEN, T.; GUESTRIN, C. **XGBoost: A Scalable Tree Boosting System**. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco: ACM, 2016.

CNES. Cadastro Nacional de Estabelecimentos de Saúde. **Estabelecimentos com internação -- Sergipe**. 2025. Disponível em: <https://tabnet.fiocruz.br/dhx.exe?estado/tb_cnes_se.def>. Acesso em: 16 fev. 2026.

COSTA, C. A.; PRATTI, T. **Dashboards como ferramenta de apoio à gestão para uma instituição hospitalar**. Revista Produção Online, v. 24, n. 4, Florianópolis, 2020.

COSTA, G. M. C. **Aplicação de métodos de previsão de demanda em leitos hospitalares**. Juiz de Fora: Universidade Federal de Juiz de Fora, 2020. Disponível

em: <<https://repositorio.ufjf.br/jspui/bitstream/ufjf/19921/1/gabriellemariadacruzcosta.pdf>>. Acesso em: 16 fev. 2026.

CQH -- COMITÊ DE QUALIDADE HOSPITALAR. **Caderno de indicadores obrigatórios: roteiro**. São Paulo, 2016. Disponível em: <<https://cqh.org.br/wp-content/uploads/001-CQH-IND-002-rev6-Caderno-de-Indicadores-Roteiro.pdf>>. Acesso em: 16 fev. 2026.

FINDLAY, R. **Planning beds, bed occupancy and risk**. 2019. Disponível em: <https://gooroo.co.uk/wp-content/uploads/2019/08/Planning_beds_bed_occupancy_and_risk.pdf>. Acesso em: 14 fev. 2026.

FREITAS, C. A. G. S. **Resposta à posição prona em pacientes com insuficiência respiratória aguda grave em Unidade de Terapia Intensiva**. Trabalho de Conclusão de Curso (Especialização em Terapia Intensiva) -- Escola Bahiana de Medicina e Saúde Pública, Salvador, 2019.

FRIEDMAN, J. H. **Greedy function approximation: a gradient boosting machine**. *Annals of Statistics*, v. 29, n. 5, p. 1189-1232, 2001.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GLLEDDSON, J. M. **Previsão, por meio de análise de séries temporais, de leitos de UTI para atendimento de pacientes com COVID-19**. Trabalho de Conclusão de Curso -- Controladoria-Geral da União, Brasília, 2021.

HAMILTON, J. D. **Time Series Analysis**. Princeton: Princeton University Press, 1994.

HOLT, C. C. **Forecasting seasonals and trends by exponentially weighted moving averages**. *International Journal of Forecasting*, v. 20, n. 1, p. 5-10, 2004 (reedição do relatório técnico de 1957).

HUANG, K. A.; HARDIN, W. M.; PRAKASH, N. S. **A comparative study across hospitals: forecasting hospital bed occupancy using machine learning methods**. *Cureus*, 2025. Disponível em: <https://assets.cureus.com/uploads/original_article/pdf/370160/20250718-263107-vi423u.pdf>. Acesso em: 18 fev. 2026.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2. ed. Melbourne: OTexts, Austrália, 2018. Disponível em: <<https://otexts.com/fpp2/>>. Acesso em: 2 mar. 2026.

KIM, J. et al. **Occupancy prediction with patient data: evaluating time-series, patient-level aggregation, and deep set models**. 2024. Disponível em: <<https://woverman.com/publication/kim-2024-occupancy/>>. Acesso em: 23 fev. 2026.

LEITE, A. C. A; *et al.* **Manual de indicadores / 2021**. São Luís: SAAS / SES-MA, 2021. Disponível em: <https://docs.bvsalud.org/biblioref/2022/10/1396957/espma-manual-de-indicadores_2021_final_09052022.pdf>. Acesso em: 14 fev. 2026.

MANGUE JORNALISMO. **Falta de estrutura do Estado tem produzido uma grave crise de saúde no centro-sul de Sergipe**. 2025. Disponível em: <<https://manguejornalismo.org/falta-de-estrutura-do-estado-tem-produzido-uma-grave-crise-de-saude-no-centro-sul-de-sergipe/>>. Acesso em: 26 fev. 2026.

MELLO, L. F. et al. **Time series forecasting of bed occupancy in mental health services**. *BMJ Open*, v. 15, n. 1, e056523, 2025.

MÉRICI, A. **SUS em Sergipe oferta 3.807 leitos qualificados na Rede Hospitalar Estadual. Aracaju: Secretaria de Estado da Saúde**, 2016. Disponível em: <<https://saude.se.gov.br/sus-em-sergipe-oferta-3-807-leitos-qualificados-na-rede-hospitalar-estadual/>>. Acesso em: 20 fev. 2026.

MORI, F. T. M.; GONÇALVES, L. R. **Aplicação da metodologia de séries temporais ao sistema de saúde do Brasil e dos Estados Unidos (2000-2012)**. *Revista Debate Econômico*, v. 9, n. 2, p. 45-68, 2021.

OLIVEIRA, L. M. et al. **Estimativa da ocupação de leitos para tratamento da COVID-19 com uso de modelos matemáticos**. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE (SBCAS), 2021. Anais [...]. Porto Alegre: SBC, 2021.

OVERMAN, W. et al. **Occupancy prediction with patient data: evaluating time-series, patient-level aggregation, and deep set models**. 2024. Disponível em: <<https://woverman.com/publication/kim-2024-occupancy/>>. Acesso em: 14 fev. 2026.

PROQUALIS. **Paciente-dia**. Rio de Janeiro: Fiocruz, 2009. Disponível em: <<https://proqualis.fiocruz.br/definicao-de-termos/paciente-dia>>. Acesso em: 14 fev. 2026.

SOUZA, F. M.; SOUZA, A. M.; LOPES, L. F. P. **Previsão da demanda de leitos hospitalares por meio da análise de séries temporais**. Universidade Federal de Santa Catarina. *Ciência e Natura*, Santa Maria, 2016.

TURATO, E. R. **Métodos qualitativos e quantitativos na área da saúde: definições, diferenças e seus objetos de pesquisa**. *Revista de Saúde Pública*, v. 39, n. 3, p. 507-514, 2005.

VENTURA, T. M. et al. **Emprego do modelo matemático SEIHDR para previsão da ocupação de leitos hospitalares por COVID-19**. *Revista HCPA*, Porto Alegre, 2021. Disponível em: <<https://seer.ufrgs.br/index.php/hcpa/article/download/112048/85520>>. Acesso em: 16 fev. 2026.

VIGISPEC. **Paciente-dia: censo manual vs. automatizado na CCIH**. 2025.

Disponível em: <<https://vigispec.com.br/paciente-dia-na-ccih/>>. Acesso em: 16 fev. 2026.

ZHAO, F. et al. **Forecasting hospital room and ward occupancy using static and dynamic information concurrently: retrospective single-center cohort study**. JMIR Medical Informatics, v. 12, n. 1, e53400, 2024.

APÊNDICE A – Script R Usado para Análise de Dados e Previsão

```
#####
# SCRIPT R PARA ANÁLISE DOS DADOS DE OCUPAÇÃO HOSPITALAR #
#####
# Autor: Jacinto Michael Menezes Cunha

#### 01) Preparação do ambiente, carregamento e ETL dos dados ####

# Instalar pacotes necessários
if (!requireNamespace("readxl", quietly = TRUE))
  install.packages("readxl")
if (!requireNamespace("dplyr", quietly = TRUE))
  install.packages("dplyr")
if (!requireNamespace("ggplot2", quietly = TRUE))
  install.packages("ggplot2")
if (!requireNamespace("moments", quietly = TRUE))
  install.packages("moments")
if (!requireNamespace("moments", quietly = TRUE))
  install.packages("markdown")
if (!requireNamespace("lubridate", quietly = TRUE))
  install.packages("lubridate")
if (!requireNamespace("tidyr", quietly = TRUE))
  install.packages("tidyr")
install.packages("rmarkdown")
if (!requireNamespace("forecast", quietly = TRUE))
  install.packages("forecast")
if (!requireNamespace("zoo", quietly = TRUE))
  install.packages("zoo")
if (!require(tidyverse)) install.packages('tidyverse');
require(tidyverse)
if (!require(tseries)) install.packages('tseries');
require(tseries)
if (!require(keras)) install.packages('keras');
require(keras)
if (!require(BatchGetSymbols)) install.packages('BatchGetSymbols');
require(BatchGetSymbols)
if (!require(PerformanceAnalytics)) install.packages('PerformanceAnalytics');
require(PerformanceAnalytics)
if (!require(xgboost)) install.packages('xgboost');
require(xgboost)
if (!require(rmarkdown)) install.packages('rmarkdown');
require(rmarkdown)

# Carregar os pacotes
library(readxl)
library(dplyr)
library(ggplot2)
library(moments)
library(rmarkdown)
library(lubridate)
library(tidyr)
library(forecast)
library(zoo)
library(nnet)
library(tidyverse)
library(tseries)
library(keras)
library(BatchGetSymbols)
library(PerformanceAnalytics)
library(dplyr)
library(purrr)
library(tibble)
library(xgboost)
library(rmarkdown)

# Ler a base de dados em .XLSX e renomeá-lo para db
db <- read_excel(file.choose()) # Comando para setar manualmente o arquivo

# Visualizar as primeiras linhas do dataframe
head(db)

# Descritivo das variáveis numéricas e categóricas
```

```

summary(db)

# Tratamento de "NA's" nas colunas descritivas
db <- db %>%
  mutate(
    cid_principal = replace_na(cid_principal, "NÃO PREENCHIDO"),
    clinica = replace_na(clinica, "Não Preenchido"),
    sexo = replace_na(sexo, "Não Preenchido"),
    convenio = replace_na(convenio, "Não Preenchido")
  )

# Verificar estrutura dos dados
str(db)

# Descritivo das variáveis numéricas e categóricas pós tratamento
summary(db)

# Visualizar coluna Idade com registro N/A
table(db$faixa_etaria, useNA = "ifany" )

#### 02) Criação de dataframes para análise descritiva ####

# 2.0) Data frame: soma de pacientes-dia por mês e ano
pacientes_mes_ano <- db %>%
  mutate(
    ano = year(data),
    mes = month(data) # 1 = jan, 12 = dez
  ) %>%
  group_by(ano, mes) %>%
  summarise(
    pacientes_dia = sum(qt_pac_dia, na.rm = TRUE),
    .groups = "drop"
  )

pacientes_mes_ano

# 2.1) Gráfico: colunas empilhadas por mês, cores por ano
ggplot(pacientes_mes_ano,
  aes(x = factor(mes,
    levels = 1:12,
    labels = c("Jan", "Fev", "Mar", "Abr", "Mai", "Jun",
      "Jul", "Ago", "Set", "Out", "Nov", "Dez")),
  y = pacientes_dia,
  fill = factor(ano)) +
  geom_col(position = "stack") +
  geom_text(aes(label = pacientes_dia,
    position = position_stack(vjust = 0.5),
    color = "white",
    size = 3) +
  labs(
    #title = "Gráfico 1 - Pacientes-dia por mês (2022-2025)",
    x = "Mês",
    y = "Pacientes-dia",
    fill = "Ano"
  ) +
  theme_minimal()

## 2.1) Idade dos pacientes (visão geral e por tipo de acomodação)

# Dataframe: soma de pacientes por sexo (visão geral)
pacientes_por_sexo <- db %>%
  group_by(sexo) %>%
  summarise(total_id = n_distinct(id_atend)) %>%
  mutate(pct = total_id / sum(total_id),
    label = scales::percent(pct, accuracy = 0.01)) # mostrar 2 casas decimais

# Gráfico de Rosca com % (Geral)
ggplot(pacientes_por_sexo, aes(x = 2, y = total_id, fill = sexo)) +
  geom_bar(stat = "identity", color = "white") +
  coord_polar(theta = "y") +
  geom_text(aes(label = label),
    position = position_stack(vjust = 0.5),
    color = "black") +
  xlim(0.5, 2.5) +
  #labs(title = "Distribuição de Pacientes por Sexo") +
  theme_void()

```

```

# Dataframe: soma de pacientes por sexo (por acomodação)
pacientes_por_sexo_agrup <- db %>%
  group_by(tipo_acomodacao, sexo) %>%
  summarise(total_id = n_distinct(id_atend), .groups = "drop_last") %>%
  mutate(pct = total_id / sum(total_id),
         label = scales::percent(pct, accuracy = 0.01)) # mostrar 2 casas decimais

# Gráfico de Rosca com % (por acomodação)
ggplot(pacientes_por_sexo_agrup, aes(x = 2, y = pct, fill = sexo)) +
  geom_bar(stat = "identity", color = "white") +
  coord_polar(theta = "y") +
  geom_text(aes(label = label),
           position = position_stack(vjust = 0.5),
           color = "black",
           size = 2.5) +
  xlim(0.5, 2.5) +
  facet_wrap(~tipo_acomodacao) +
  #labs(title = "Distribuição de Pacientes por Sexo e Tipo de Acomodação") +
  theme_void() +
  # Aumenta margem abaixo do título e abaixo dos títulos acima de cada gráfico
  theme(
    plot.title = element_text(margin = margin(b = 20), hjust = 0.5),
    strip.text = element_text(margin = margin(t = 10, b = 10))
  )

## 2.2) Idade dos pacientes (visão geral e por tipo de acomodação)

# Dataframe: soma de pacientes-dia por idade (visão geral)
pacientes_por_idade <- db %>%
  group_by(idade) %>%
  summarise(pacientes_dia = sum(qt_pac_dia, na.rm = TRUE), .groups = "drop")

print(pacientes_por_idade)

summary(pacientes_por_idade$idade)

# Remove idades NA
pacientes_por_idade <- na.omit(pacientes_por_idade)
summary(pacientes_por_idade$idade)
# Removidos 10.047 registros com idade N/A do dataframe "pacientes_por_idade"

# Box-plot geral: idade dos pacientes
ggplot(db, aes(x = "Box-plot", y = idade)) +
  geom_boxplot(outlier.colour = "red", outlier.size = 2) +
  stat_summary(fun = median, geom = "text",
             aes(label = round(..y.., 1)),
             position = position_nudge(x = 0.2),
             color = "blue", size = 4) +
  labs(#title = "Box-Plot: Idade dos Pacientes",
       x = NULL,
       y = "Idade") +
  theme_minimal()

# Box-plot por idade e tipo de acomodação
ggplot(db, aes(x = tipo_acomodacao, y = idade)) +
  geom_boxplot(outlier.colour = "red", outlier.size = 2) +
  stat_summary(fun = median, geom = "text",
             aes(label = round(..y.., 1)),
             position = position_nudge(x = 0.2),
             color = "blue", size = 4) +
  labs(#title = "Box-Plot: Idade dos Pacientes por Tipo de Acomodação",
       x = "Acomodação",
       y = "Idade") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 0, hjust = 1))

# Dispersão: idade x pacientes-dia (visão geral)
ggplot(pacientes_por_idade, aes(x = idade, y = pacientes_dia)) +
  geom_point(color = "blue", size = 3, alpha = 0.6) +
  labs(#title = "Dispersão: Qtd. Pacientes-dia por Idade",
       x = "Idade",
       y = "Pacientes-dia") +
  theme_minimal()

# Dataframe: pacientes-dia por idade e tipo de acomodação
pacientes_por_idade_acom <- db %>%
  group_by(tipo_acomodacao, idade) %>%

```

```

summarise(pacientes_dia = sum(qt_pac_dia, na.rm = TRUE), .groups = "drop")

print(pacientes_por_idade_acom)

# Remove idades NA
pacientes_por_idade_acom <- na.omit(pacientes_por_idade_acom)
summary(pacientes_por_idade_acom$idade)
# Removidos 7.898 registros com idade N/A do dataframe "pacientes_por_idade"

# Dispersão: idade x pacientes-dia por tipo de acomodação (facets)
ggplot(pacientes_por_idade_acom,
  aes(x = idade, y = pacientes_dia)) +
  geom_point(color = "blue", size = 2, alpha = 0.6) +
  labs(#title = "Dispersão: Pacientes-dia por Idade e Tipo de Acomodação",
    x = "Idade",
    y = "Pacientes-dia") +
  theme_minimal() +
  facet_wrap(~ tipo_acomodacao, scales = "free_y", ncol = 1)

## 2.3) Pacientes por setor de atendimento (visão geral e por acomodação)

# Dataframe: pacientes-dia por setor (geral)
pacientes_por_setor <- db %>%
  group_by(setor_atendimento) %>%
  summarise(pacientes_dia = sum(qt_pac_dia, na.rm = TRUE), .groups = "drop") %>%
  arrange(desc(pacientes_dia))

print(pacientes_por_setor)

# Gráfico: setor x pacientes-dia (geral)
ggplot(pacientes_por_setor,
  aes(x = reorder(setor_atendimento, pacientes_dia), y = pacientes_dia)) +
  geom_col(fill = "lightblue", color = "darkblue") +
  geom_text(aes(label = pacientes_dia),
    hjust = -0.1, # ajusta posição horizontal do texto
    size = 3) +
  labs(#title = "Pacientes-dia por Setor de Atendimento",
    x = "Setor de Atendimento",
    y = "Pacientes-dia") +
  theme_minimal() +
  theme(axis.text.y = element_text(hjust = 1)) +
  coord_flip()

# Dataframe: pacientes-dia por setor e tipo de acomodação
pacientes_por_setor_acom <- db %>%
  group_by(tipo_acomodacao, setor_atendimento) %>%
  summarise(pacientes_dia = sum(qt_pac_dia, na.rm = TRUE), .groups = "drop") %>%
  arrange(tipo_acomodacao, desc(pacientes_dia))

print(pacientes_por_setor_acom, n=21)

pacientes_por_setor_acom <- pacientes_por_setor_acom %>%
  group_by(tipo_acomodacao) %>%
  arrange(desc(-pacientes_dia), .by_group = TRUE) %>%
  mutate(
    setor_acom = factor(
      paste(tipo_acomodacao, setor_atendimento),
      levels = paste(tipo_acomodacao, setor_atendimento)
    )
  ) %>%
  ungroup()

# Gráfico: setor x pacientes-dia por tipo de acomodação (facets)
ggplot(pacientes_por_setor_acom,
  aes(x = setor_acom, y = pacientes_dia)) +
  geom_col(fill = "lightblue", color = "darkblue") +
  geom_text(aes(label = pacientes_dia),
    hjust = -0.1,
    size = 3) +
  labs(#title = "Pacientes-dia por Setor de Atendimento e Tipo de Acomodação",
    x = "Setor de Atendimento",
    y = "Pacientes-dia") +
  scale_x_discrete(labels = pacientes_por_setor_acom$setor_atendimento) +
  theme_minimal() +
  coord_flip(clip = "off") +
  facet_wrap(~ tipo_acomodacao, scales = "free_y", ncol = 1)

```

```

## 2.4) Pacientes por convênio (visão geral e por acomodação)

# Dataframe: pacientes-dia por convênio (geral)
pacientes_por_convênio <- db %>%
  group_by(convênio) %>%
  summarise(pacientes_dia = sum(qt_pac_dia, na.rm = TRUE), .groups = "drop") %>%
  arrange(desc(pacientes_dia)) %>%
  na.omit() %>%
  dplyr::slice_max(pacientes_dia, n = 10)

print(pacientes_por_convênio)

# Gráfico: top 10 convênios (geral)
ggplot(pacientes_por_convênio,
  aes(x = reorder(convênio, pacientes_dia), y = pacientes_dia)) +
  geom_col(fill = "lightblue", color = "darkblue") +
  geom_text(aes(label = pacientes_dia),
    hjust = -0.1,
    size = 3) +
  labs(#title = "Top 10 Pacientes-dia por Convênio",
    x = "Convênio",
    y = "Pacientes-dia") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_flip()

# Dataframe: pacientes-dia por convênio e tipo de acomodação
pacientes_por_convênio_acom <- db %>%
  group_by(tipo_acomodacao, convênio) %>%
  summarise(pacientes_dia = sum(qt_pac_dia, na.rm = TRUE), .groups = "drop")

# Top 10 Convênio por acomodação
pacientes_por_convênio_acom_top <- pacientes_por_convênio_acom %>%
  group_by(tipo_acomodacao) %>%
  arrange(desc(pacientes_dia)) %>%
  dplyr::slice_max(pacientes_dia, n = 10) %>%
  ungroup()

pacientes_por_convênio_acom_top <- pacientes_por_convênio_acom_top %>%
  group_by(tipo_acomodacao) %>%
  arrange(desc(pacientes_dia), .by_group = TRUE) %>%
  mutate(
    conv_acom = factor(
      paste(tipo_acomodacao, convênio),
      levels = paste(tipo_acomodacao, convênio)
    )
  ) %>%
  ungroup()

# Gráfico: top convênios por tipo de acomodação
ggplot(pacientes_por_convênio_acom_top,
  aes(x = fct_rev(conv_acom), y = pacientes_dia)) +
  geom_col(fill = "lightblue", color = "darkblue") +
  geom_text(aes(label = pacientes_dia),
    hjust = -0.1,
    size = 3) +
  labs(#title = "Top Convênios por Tipo de Acomodação",
    x = "Convênio",
    y = "Pacientes-dia") +
  scale_x_discrete(labels = pacientes_por_convênio_acom_top$convênio) +
  theme_minimal() +
  coord_flip(clip = "off") +
  facet_wrap(~ tipo_acomodacao, scales = "free_y", ncol = 1)

## 2.5) Pacientes por CID (visão geral e por acomodação)

# Dataframe: top 10 CIDs (geral)
pacientes_por_cid <- db %>%
  group_by(cid_principal) %>%
  summarise(pacientes_dia = sum(qt_pac_dia, na.rm = TRUE), .groups = "drop") %>%
  arrange(desc(pacientes_dia)) %>%
  dplyr::slice_max(pacientes_dia, n = 10)

print(pacientes_por_cid)

```

```

# Gráfico: top 10 CID (geral)
ggplot(pacientes_por_cid,
       aes(x = reorder(cid_principal, pacientes_dia), y = pacientes_dia)) +
  geom_col(fill = "lightblue", color = "darkblue") +
  labs(#title = "Top 10 CID por Pacientes-dia",
       x = "CID principal",
       y = "Pacientes-dia") +
  geom_text(aes(label = pacientes_dia),
            hjust = -0.1,
            size = 3.2) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_flip()

# Dataframe: CIDs por tipo de acomodação
pacientes_por_cid_acom <- db %>%
  group_by(tipo_acomodacao, cid_principal) %>%
  summarise(pacientes_dia = sum(qt_pac_dia, na.rm = TRUE), .groups = "drop")

# Top 10 CIDs por tipo de acomodação
pacientes_por_cid_acom_top <- pacientes_por_cid_acom %>%
  group_by(tipo_acomodacao) %>%
  dplyr::slice_max(pacientes_dia, n = 10) %>%
  ungroup()

pacientes_por_cid_acom_top <- pacientes_por_cid_acom_top %>%
  group_by(tipo_acomodacao) %>%
  arrange(desc(pacientes_dia), .by_group = TRUE) %>%
  mutate(
    cid_acom = factor(
      paste(tipo_acomodacao, cid_principal),
      levels = paste(tipo_acomodacao, cid_principal)
    )
  ) %>%
  ungroup()

# Gráfico: top 10 CID por tipo de acomodação
ggplot(pacientes_por_cid_acom_top,
       aes(x = fct_rev(cid_acom), y = pacientes_dia)) +
  geom_col(fill = "lightblue", color = "darkblue") +
  geom_text(aes(label = pacientes_dia),
            hjust = -0.1,
            size = 3) +
  labs(title = "Top 10 CID por Tipo de Acomodação",
       x = "CID principal",
       y = "Pacientes-dia") +
  scale_x_discrete(labels = pacientes_por_cid_acom_top$cid_principal) +
  theme_minimal() +
  coord_flip(clip = "off") +
  facet_wrap(~ tipo_acomodacao,
            scales = "free_y",
            ncol = 1)

## 2.6) Pacientes por clínica (visão geral e por acomodação)

# Dataframe: pacientes-dia por clínica (geral)
pacientes_por_clinica <- db %>%
  group_by(clinica) %>%
  summarise(pacientes_dia = sum(qt_pac_dia, na.rm = TRUE), .groups = "drop") %>%
  arrange(desc(pacientes_dia))

print(pacientes_por_clinica)

# Gráfico: clínicas (geral)
ggplot(pacientes_por_clinica,
       aes(x = reorder(clinica, pacientes_dia), y = pacientes_dia)) +
  geom_col(fill = "lightblue", color = "darkblue") +
  labs(#title = "Clínicas por Pacientes-dia",
       x = "Clínica",
       y = "Pacientes-dia") +
  geom_text(aes(label = pacientes_dia),
            hjust = -0.1,
            size = 3) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  coord_flip()

```

```

# Dataframe: pacientes-dia por clínica e tipo de acomodação
pacientes_por_clinica_acom <- db %>%
  group_by(tipo_acomodacao, clinica) %>%
  summarise(pacientes_dia = sum(qt_pac_dia, na.rm = TRUE), .groups = "drop")

print(pacientes_por_clinica_acom)

# top 10 clínicas por tipo de acomodação
pacientes_por_clinica_acom_top <- pacientes_por_clinica_acom %>%
  group_by(tipo_acomodacao) %>%
  dplyr::slice_max(pacientes_dia, n = 10) %>%
  ungroup()

pacientes_por_clinica_acom_top <- pacientes_por_clinica_acom_top %>%
  group_by(tipo_acomodacao) %>%
  arrange(desc(pacientes_dia), .by_group = TRUE) %>%
  mutate(
    clin_acom = factor(
      paste(tipo_acomodacao, clinica),
      levels = paste(tipo_acomodacao, clinica)
    )
  ) %>%
  ungroup()

# Gráfico: clínicas por tipo de acomodação
ggplot(pacientes_por_clinica_acom_top,
  aes(x = fct_rev(clin_acom), y = pacientes_dia)) +
  geom_col(fill = "lightblue", color = "darkblue") +
  geom_text(aes(label = pacientes_dia,
    hjust = -0.1,
    size = 3) +
  labs(#title = "Clínicas por Pacientes-dia e Tipo de Acomodação",
    x = "Clínica",
    y = "Pacientes-dia") +
  scale_x_discrete(labels = pacientes_por_clinica_acom_top$clinica) +
  theme_minimal() +
  coord_flip(clip = "off") +
  facet_wrap(~ tipo_acomodacao,
    scales = "free_y",
    ncol = 1)

## 2.7) Pacientes por tipo de acomodação

# Criar um novo dataframe com a soma de Quantidade de Pacientes-dia por Tipo Acomodação
pacientes_por_acomodacao <- db %>%
  group_by(tipo_acomodacao) %>%
  summarise(pacientes_dia = sum(qt_pac_dia, na.rm = TRUE)) %>%
  arrange(desc(pacientes_dia)) # Ordenar do maior para o menor

# Visualizar o novo dataframe
print(pacientes_por_acomodacao)

# Gráfico de barras: Pacientes-dia por Tipo Acomodação
ggplot(pacientes_por_acomodacao,
  aes(x = reorder(tipo_acomodacao, - pacientes_dia), y = pacientes_dia)) +
  geom_col(fill = "lightblue", #cor da linha
    color = "darkblue") + #cor da coluna
  geom_text(aes(label = pacientes_dia,
    hjust = 0.5,
    vjust = -0.5, # levemente acima da extremidade da coluna
    size = 3) +
  labs(#title = "Pacientes-dia por Tipo de Acomodação",
    x = "",
    y = "Pacientes-dia") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 0))

#### 03) Preparação do Ambiente para análise de Séries Temporais ####

## 3.1) Definição do dataframe, incluindo a variável de tempo
df_temp <- db %>%
  mutate(
    data = as_date(data),
    Ano = year(data),
    Mes = month(data),
    Semana = week(data),

```

```

    Dia = day(data)
  ) %>%
  group_by(tipo_acomodacao, Ano, Mes, Semana, Dia) %>%

  # Calculando a taxa de ocupação
  mutate(qt_pac_dia = sum(qt_pac_dia, na.rm = T),
         qt_leito_dia = sum(qt_leito_dia, na.rm = T),
         tx_ocupacao = round(qt_pac_dia/qt_leito_dia,2),
         dt_ocupacao = max(data)) %>%
  arrange(data) %>%
  ungroup() %>%
  select(tipo_acomodacao, qt_leito_dia, qt_pac_dia, dt_ocupacao, tx_ocupacao) %>%
  distinct()

## 3.2) Função auxiliar para criar (ts) em 3 frequências (mensal, quinzenal e semanal)
ts_03freq <- function(x) {
  x <- x[is.finite(x)]
  if (length(x) < 3) {
    return(list(mensal = NULL, quinzenal = NULL, semanal = NULL))
  }
  ts_mes <- ts(x,
              start = c(2022, 1, 1),
              end = c(2025, 12, 31),
              frequency = 12)

  ts_quinzena <- ts(x,
                   start = c(2022, 1, 1),
                   end = c(2025, 12, 31),
                   frequency = 24)

  ts_semana <- ts(x,
                 start = c(2022, 1, 1),
                 end = c(2025, 12, 31),
                 frequency = 52)

  list(
    mensal = ts_mes,
    quinzenal = ts_quinzena,
    semanal = ts_semana
  )
}

# Ordenar novamente por data de ocupação
df_temp <- df_temp %>% arrange(dt_ocupacao)

# Listas de séries por tipo de acomodação
series_agrupadas <- df_temp %>%
  group_by(tipo_acomodacao) %>%
  summarise(
    tx_ocupacao_vec = list(tx_ocupacao),
    .groups = "drop"
  ) %>%
  mutate(
    ts_list = map(tx_ocupacao_vec, ts_03freq)
  )

## 3.3) Filtra a linhas de cada tipo de acomodação

# UTI
uti_row <- series_agrupadas %>%
  filter(tipo_acomodacao == "UTI")
# Apartamento
apt_row <- series_agrupadas %>%
  filter(tipo_acomodacao == "Apartamento")
# Enfermaria
enf_row <- series_agrupadas %>%
  filter(tipo_acomodacao == "Enfermaria")

# extrair a lista de séries ts
uti_ts <- uti_row$ts_list[[1]]# ele só tem uma linha
apt_ts <- apt_row$ts_list[[1]]# ele só tem uma linha
enf_ts <- enf_row$ts_list[[1]]# ele só tem uma linha

## 3.4) Função para gráfico de AIC e PAIC e avaliação da distribuição de Probabilidade
das séries:

```

```

analisa_serie <- function(x, nome = "serie") {

  # Série em linha
  plot(x,
        type = "l",
        main = paste("Série temporal -", nome),
        xlab = "Tempo",
        ylab = "Taxa Ocupação")

  # ACF e PACF
  par(mfrow = c(1, 2))
  acf_obj <- acf(x, main = paste("ACF -", nome), plot = TRUE)
  pacf_obj <- pacf(x, main = paste("Partial ACF -", nome), plot = TRUE)

  par(mfrow = c(1, 3))
  # Histogram
  hist(x,
        main = paste("Histograma -", nome),
        xlab = nome)

  # QQ-plot
  qqnorm(x,
          main = paste("QQ-plot -", nome))
  qqline(x, col = "red")

  par(mfrow=c(1,1)) # para voltar ao padrão

  list(
    acf = acf_obj,
    pacf = pacf_obj
  ) # Essa lista serve para chamar o argumento e executar cada item separadamente
}

# Resultados da função gerada para os gráficos de avaliação da série:

resultados_plot <- series_agrupadas %>%
  mutate(
    analise_mensal = map2(ts_list, tipo_acomodacao,
                        ~ analisa_serie(.x$mensal,
                                         paste(.y, "mensal"))),
    analise_quinzenal = map2(ts_list, tipo_acomodacao,
                             ~ analisa_serie(.x$quinzenal,
                                              paste(.y, "quinzenal"))),
    analise_semanal = map2(ts_list, tipo_acomodacao,
                          ~ analisa_serie(.x$semanal,
                                           paste(.y, "semanal")))
  )

par(mfrow=c(1,1))

# Saber qual ordem dos índices dos tipos de acomodações
series_agrupadas$tipo_acomodacao[1]
series_agrupadas$tipo_acomodacao[2]
series_agrupadas$tipo_acomodacao[3]

## 3.5) Calcula o teste de Dickey-Fuller Aumentado (ADF) para cada série:
adf_mensal_UTI <- adf.test(uti_ts$mensal) # UTI mensal
adf_mensal_UTI
adf_quinzenal_UTI <- adf.test(uti_ts$quinzenal) # UTI quinzenal
adf_quinzenal_UTI
adf_semanal_UTI <- adf.test(uti_ts$semanal) # UTI semanal
adf_semanal_UTI

adf_mensal_apt <- adf.test(apt_ts$mensal) # apt mensal
adf_mensal_apt
adf_quinzenal_apt <- adf.test(apt_ts$quinzenal) # apt quinzenal
adf_quinzenal_apt
adf_semanal_apt <- adf.test(apt_ts$semanal) # apt semanal
adf_semanal_apt

adf_mensal_enf <- adf.test(enf_ts$mensal) # enf mensal
adf_mensal_enf
adf_quinzenal_enf <- adf.test(enf_ts$quinzenal) # enf quinzenal
adf_quinzenal_enf
adf_semanal_enf <- adf.test(enf_ts$semanal) # enf semanal
adf_semanal_enf

```

```

# 3.5.1) Criamos uma lista nomeada com todos os objetos de testes mensais
lista_testes <- list(
  "APT - Mensal"      = adf_mensal_apt,
  "ENF - Mensal"      = adf_mensal_enf,
  "UTI - Mensal"      = adf_mensal_UTI
)

# 3.5.2) Extraímos os dados de interesse usando sapply
tabela_adf <- data.frame(
  Serie_Frequencia = names(lista_testes),
  Dickey_Fuller     = sapply(lista_testes, function(x) x$statistic),
  P_Valor           = sapply(lista_testes, function(x) x$p.value),
  Lags              = sapply(lista_testes, function(x) x$parameter)
)

# 3.5.3) Adicionamos uma coluna de conclusão para facilitar a leitura
# Se p-valor < 0.05, rejeitamos H0 (a série é Estacionária)
tabela_adf$Conclusao <- ifelse(tabela_adf$P_Valor < 0.05, "Estacionária", "Não
Estacionária")

# 3.5.4) Removemos os nomes de linha automáticos para limpar a visualização e
arredondamos
row.names(tabela_adf) <- NULL
tabela_adf$Dickey_Fuller <- round(tabela_adf$Dickey_Fuller, 4)
tabela_adf$P_Valor <- round(tabela_adf$P_Valor, 4)

# 3.5.5) Visualizar a tabela final
print(tabela_adf)

## 3.6) Decomposição das Séries Mensais por Tipo de Acomodação
d_uti_ts <- decompose(uti_ts$mensal, type = "additive")
autoplot(d_uti_ts)

d_apt_ts <- decompose(apt_ts$mensal, type = "additive")
autoplot(d_apt_ts)

d_enf_ts <- decompose(enf_ts$mensal, type = "additive")
autoplot(d_enf_ts)

## 3.7) Aplicação de Diferenciação e teste de Dickey-fuller Aumentado nas séries Mensais
# Aplicação da 1ª diferença (d = 1) e escolhido periodicidade mensal

#UTI
uti_mensal_diff1 <- diff(uti_ts$mensal, differences = 1)

# Apartamento
apt_mensal_diff1 <- diff(apt_ts$mensal, differences = 1)

# Enfermaria
enf_mensal_diff1 <- diff(enf_ts$mensal, differences = 1)

# teste de Dickey-Fuller Aumentado (ADF) na série diferenciada (d = 1)
adf_uti_mensal_diff1 <- adf.test(uti_mensal_diff1)
adf_uti_mensal_diff1
adf_apt_mensal_diff1 <- adf.test(apt_mensal_diff1)
adf_apt_mensal_diff1
adf_enf_mensal_diff1 <- adf.test(enf_mensal_diff1)
adf_enf_mensal_diff1

# Decomposição das Séries Mensais Diferenciadas (d = 1) por Tipo de Acomodação
d_uti_ts_diff1 <- decompose(uti_mensal_diff1, type = "additive")
autoplot(d_uti_ts_diff1)+
  labs(title = "Decomposição da Série - UTI Mensal")

d_apt_ts_diff1 <- decompose(apt_mensal_diff1, type = "additive")
autoplot(d_apt_ts_diff1)+
  labs(title = "Decomposição da Série - Apartamento Mensal")

d_enf_ts_diff1 <- decompose(enf_mensal_diff1, type = "additive")
autoplot(d_enf_ts_diff1)+
  labs(title = "Decomposição da Série - Enfermaria Mensal")

## 3.8) Gráfico de AIC e PAIC e avaliação da distribuição de Probabilidade das séries
mensais diferenciadas (d = 1)

```

```

recalcula_graf_diff_mensal <- function(tipo_nome) {

  row_tipo <- series_agrupadas %>%
    dplyr::filter(tipo_acomodacao == tipo_nome)

  ts_tipo <- row_tipo$ts_list[[1]]
  serie <- ts_tipo$mensal

  # 1ª diferença
  serie_diff1 <- diff(serie, differences = 1)

  # Gráficos:

  # ACF e PACF
  par(mfrow = c(1, 2))
  acf(serie_diff1, main = paste("ACF -", tipo_nome, "mensal (d=1)"))
  pacf(serie_diff1, main = paste("PACF -", tipo_nome, "mensal (d=1)"))

  par(mfrow = c(1, 3))

  # Séries em linha
  plot(serie_diff1,
       type = "l",
       main = paste("Série temporal -", tipo_nome),
       xlab = "Tempo",
       ylab = tipo_nome)

  #Histograma
  hist(serie_diff1,
       main = paste("Histograma -", tipo_nome, "mensal (d=1)"),
       xlab = "Diferença 1")

  #QQ-plot
  qqnorm(serie_diff1,
         main = paste("QQ-plot -", tipo_nome, "mensal (d=1)"))
  qqline(serie_diff1, col = "red")

  par(mfrow = c(1, 1)) # volta layout padrão

  # Objetos numérico para análise
  list(
    acf = acf(serie_diff1, plot = FALSE),
    pacf = pacf(serie_diff1, plot = FALSE),
    serie_diff = serie_diff1
  )
}

#Resultados dos gráficos da série diferenciada (d = 1)
res_uti <- recalcula_graf_diff_mensal("UTI")
res_ap <- recalcula_graf_diff_mensal("Apartamento")
res_enf <- recalcula_graf_diff_mensal("Enfermaria")

## 4) MODELAGEM ARIMA

# Melhor modelo usando função "auto.arima()"
melhor_enf <- auto.arima(ts(enf_ts$mensal, frequency = 12), d = 1)
melhor_ap <- auto.arima(ts(apt_ts$mensal, frequency = 12), d = 1)
melhor_uti <- auto.arima(ts(uti_ts$mensal, frequency = 12), d = 1)

summary(melhor_enf) #arima(0,1,1) #o R ta entendendo que essa série é um ruído branco?
summary(melhor_ap) #arima(0,1,1) ou arima(0,0,1)[12]
summary(melhor_uti) #arima(1,1,0)

## 4.1) Resíduos dos modelos:
res_melhor_enf <- residuals(melhor_enf)
res_melhor_ap <- residuals(melhor_ap)
res_melhor_uti <- residuals(melhor_uti)

## 4.2) Teste Shapiro-Wilk para os resíduos dos modelos escolhidos
# (Não Rejeita Hipótese Normalidade: p-valor > 0,05)

shapiro_enf <- shapiro.test(res_melhor_enf)
shapiro_ap <- shapiro.test(res_melhor_ap)
shapiro_uti <- shapiro.test(res_melhor_uti)

shapiro_enf

```

```

shapiro_apt
shapiro_uti

## 4.3) Análise Gráfica dos Resíduos das séries ARIMA:

#Enfermaria
checkresiduals(melhor_enf)

# QQ-plot do resíduo enfermaria
qqnorm(residuals(melhor_enf),
        main = "QQ-plot do Residuo no Modelo ARIMA - Enfermaria")
qqline(residuals(melhor_enf), col = "red")

#Apartamento
checkresiduals(melhor_apt)

# QQ-plot do resíduo Apartamento
qqnorm(residuals(melhor_apt),
        main = "QQ-plot do Residuo no Modelo ARIMA - Apartamento")
qqline(residuals(melhor_apt), col = "red")

#UTI
checkresiduals(melhor_uti)

# QQ-plot do resíduo UTI
qqnorm(residuals(melhor_uti),
        main = "QQ-plot do Residuo no Modelo ARIMA - UTI")
qqline(residuals(melhor_uti), col = "red")

#### 05) Modelagem usando o ETS (Exponencial Smoothing) ####

## 5.1) Para Enfermaria

# Modelando
fit_ets_enf <- ets(na.interp(ts(enf_ts$mensal,start = c(2022, 01), frequency = 12)),
model = "ANA")

# Summary dos resultados do modelo
summary(fit_ets_enf)

# Avaliação dos residuos
checkresiduals(fit_ets_enf)

# Teste Shapiro-Wilk para avaliar normalidade dos resíduos
# (Não Rejeita Hipótese Normalidade: p-valor > 0,05)

shapiro_ets_enf <- shapiro.test(residuals(fit_ets_enf))
shapiro_ets_enf

# QQ-plot do resíduo
qqnorm(residuals(fit_ets_enf),
        main = "QQ-plot do Residuo no Modelo ETS - Enfermaria")
qqline(residuals(fit_ets_enf), col = "red")

# Medindo a sua acurácia (RMSE e MAPE)
accuracy(fit_ets_enf)

# Interpretação:
# Embora os resíduos não sigam uma distribuição estritamente normal (Shapiro-Wilk <
0,05),
# eles são independentes e não correlacionados (Ljung-Box p=0,363), o que garante que o
# modelo extraiu adequadamente os padrões da série temporal.

## 5.2) Para Apartamento

# Modelando
fit_ets_apt <- ets(na.interp(ts(apt_ts$mensal, start = c(2022, 01), frequency = 12)),
model = "ANA")

# Summary dos resultados do modelo
summary(fit_ets_apt)

# Avaliação dos residuos
checkresiduals(fit_ets_apt)

```

```

# Teste Shapiro-Wilk para avaliar normalidade dos resíduos
# (Não Rejeita Hipótese Normalidade: p-valor > 0,05)

shapiro_ets_apt <- shapiro.test(residuals(fit_ets_apt))
shapiro_ets_apt

# QQ-plot do resíduo
qqnorm(residuals(fit_ets_apt),
        main = "QQ-plot do Resíduo no Modelo ETS - Apartamento")
qqline(residuals(fit_ets_apt), col = "red")

# Medindo a sua acurácia (RMSE e MAPE)
accuracy(fit_ets_apt)

## 5.3) Para UTI

# Modelando
fit_ets_uti <- ets(na.interp(ts(uti_ts$mensal, start = c(2022, 01), frequency = 12)),
model = "ANA")

# Summary dos resultados do modelo
summary(fit_ets_uti)

# Avaliação dos resíduos
checkresiduals(fit_ets_uti)

# Teste Shapiro-Wilk para avaliar normalidade dos resíduos
# (Não Rejeita Hipótese Normalidade: p-valor > 0,05)

shapiro_ets_uti <- shapiro.test(residuals(fit_ets_uti))
shapiro_ets_uti

# QQ-plot do resíduo
qqnorm(residuals(fit_ets_uti),
        main = "QQ-plot do Resíduo no Modelo ETS - UTI")
qqline(residuals(fit_ets_uti), col = "red")

# Medindo a sua acurácia (RMSE e MAPE)
accuracy(fit_ets_uti)

# plot(forecast(fit_ets_enf, h = 12))

#### 06) MODELAGEM PREDITIVA COM XGBOOST POR TIPO DE ACOMODAÇÃO ####

# Abordagem: regressão XGBoost com lags e variáveis de tempo,
# aplicando a mesma granularidade da série modelada usando o ARIMA e ETS(A,N,A)

## 6.1) Preparação da base mensal por tipo de acomodação

ocupacao_mensal <- df_temp %>%
  mutate(
    ano = year(dt_ocupacao),
    mes = month(dt_ocupacao)
  ) %>%
  group_by(tipo_acomodacao, ano, mes) %>%
  summarise(
    tx_ocupacao_mensal = mean(tx_ocupacao, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  arrange(tipo_acomodacao, ano, mes) %>%
  mutate(
    data_mes = as.Date(paste(ano, mes, "01", sep = "-")),
    ano_mes = as.yearmon(data_mes)
  )

## 6.2) Função para criar features (lags e componentes de tempo)

cria_features_serie <- function(df, max_lag = 3) {
  df <- df %>%
    arrange(data_mes) %>%
    mutate(
      # Componentes de tempo simples
      mes_fator = factor(mes),
      ano_num = ano,
      # Lags da própria taxa de ocupação (curto prazo)
      lag1 = dplyr::lag(tx_ocupacao_mensal, 1),

```

```

    lag2 = dplyr::lag(tx_ocupacao_mensal, 2),
    # Lag Sazonal (mesmo mês do ano anterior)
    lag12 = dplyr::lag(tx_ocupacao_mensal, 12),
    # Variáveis de tempo
  )

  # Remove as primeiras linhas que não têm lags completos
  df %>% drop_na(lag1, lag2, lag12)
}

# Aplica por tipo de acomodação
ocupacao_mensal_feat <- ocupacao_mensal %>%
  group_by(tipo_acomodacao) %>%
  group_modify(~ cria_features_serie(.x)) %>%
  ungroup()

## 6.3) Função auxiliar para treinar e avaliar XGBoost por tipo

ajusta_xgboost_por_tipo <- function(df_tipo, proporcao_treino = 0.85) {
  # Uso de 0.85 para treino, pois contempla aprox. 60 meses.
  # Ordena por tempo
  df_tipo <- df_tipo %>% arrange(data_mes)
  n_total <- nrow(df_tipo)
  n_treino <- floor(n_total * proporcao_treino)

  # Avalia a série se for muito curta, não tenta modelar
  if (n_total < 10) {
    return(list(
      modelo = NULL,
      metricas = tibble(),
      previsoes = tibble()
    ))
  }

  # Define matriz de features (sem colunas não numéricas irrelevantes)
  # Usou-se:
  # - ano_num, mes (número), lag1, lag2, lag12
  # Obs.: mes_fator poderia ser expandido em dummies, mas para manter simples usou-se
  # mês numérico
  X <- df_tipo %>%
    transmute(
      ano_num,
      mes,
      lag1,
      lag2,
      lag12
    ) %>%
    as.matrix()

  y <- df_tipo$tx_ocupacao_mensal

  # Split temporal
  X_treino <- X[1:n_treino, , drop = FALSE]
  y_treino <- y[1:n_treino]
  X_teste <- X[(n_treino + 1):n_total, , drop = FALSE]
  y_teste <- y[(n_treino + 1):n_total]

  # DMatrix do xgboost
  dtrain <- xgb.DMatrix(data = X_treino, label = y_treino)
  dtest <- xgb.DMatrix(data = X_teste, label = y_teste)

  # Parâmetros básicos
  params <- list(
    objective = "reg:squarederror",
    eta = 0.1,
    max_depth = 3,
    subsample = 0.9,
    colsample_bytree = 0.9
  )

  set.seed(123) #aplicação de semente de dados aleatória
  modelo_xgb <- xgb.train(
    params = params,
    data = dtrain,
    nrounds = 200,
    watchlist = list(train = dtrain, test = dtest),
    verbose = 0,

```

```

    early_stopping_rounds = 20
  )

  # Previsões no conjunto de teste
  y_pred <- predict(modelo_xgb, dtest)

  # Métricas simples
  mae <- mean(abs(y_teste - y_pred))
  rmse <- sqrt(mean((y_teste - y_pred)^2))
  mape <- mean(abs((y_teste - y_pred) / ifelse(y_teste == 0, NA, y_teste)), na.rm =
TRUE) * 100

  metricas <- tibble(
    MAE = mae,
    RMSE = rmse,
    MAPE = mape
  )

  previsoes <- tibble(
    data_mes = df_tipo$data_mes[(n_treino + 1):n_total],
    real = y_teste,
    previsto = y_pred
  )

  list(
    modelo = modelo_xgb,
    metricas = metricas,
    previsoes = previsoes
  )
}

## 6.4) Guardar Resultados XGBoost para cada tipo de acomodação (modelo, metricas e
previsões)

resultados_xgb <- ocupacao_mensal_feat %>%
  group_by(tipo_acomodacao) %>%
  group_modify(~{
    res <- ajusta_xgboost_por_tipo(.x)
    tibble(
      modelo = list(res$modelo),
      metricas = list(res$metricas),
      previsoes = list(res$previsoes)
    )
  }) %>%
  ungroup()

# Visualizar métricas por tipo de acomodação
metricas_xgb <- resultados_xgb %>%
  mutate(metricas = map(metricas, ~ mutate(.x, tipo_acomodacao =
cur_data()$tipo_acomodacao))) %>%
  unnest(metricas)

print(metricas_xgb)

## 6.5) Gráficos: real x previsto (XGBoost) no período de teste (jan a dez 2025)

plot_real_prev_xgb <- function(df_prev, titulo) {
  ggplot(df_prev, aes(x = data_mes)) +
    geom_line(aes(y = real, colour = "Real"), size = 1) +
    geom_line(aes(y = previsto, colour = "Previsto XGBoost"), size = 1, linetype =
"dashed") +
    labs(
      title = titulo,
      x = "Mês",
      y = "Taxa de ocupação mensal"
    ) +
    scale_colour_manual(values = c("Real" = "black", "Previsto XGBoost" = "red")) +
    theme_minimal()
}

# Geração gráficos por tipo de acomodação
lista_plots_xgb <- resultados_xgb %>%
  mutate(
    grafico = map2(
      previsoes,
      tipo_acomodacao,
      ~ plot_real_prev_xgb(.x, paste0("Real x Previsto (XGBoost) - ", .y))
    )
  )

```

```

    )
  )

# Para visualizar um gráfico específico, por exemplo o primeiro:
print(lista_plots_xgb$grafico[[1]])
print(lista_plots_xgb$grafico[[2]])
print(lista_plots_xgb$grafico[[3]])

#### 07) Comparar o MAE, RMSE e MAPE dos modelos ARIMA, ETS e XGBOOST (2022 a 2025) ####

## 7.1) Métricas ARIMA (2022-2025)
calc_metricas_arima_full <- function(serie_ts, modelo_arima, tipo_nome) {
  # previsão in-sample (ajustada) em toda a série
  fitted_vals <- fitted(modelo_arima)

  # Garantir alinhamento (comprimentos iguais)
  n <- min(length(serie_ts), length(fitted_vals))
  real <- as.numeric(serie_ts[1:n])
  prev <- as.numeric(fitted_vals[1:n])

  mae <- mean(abs(real - prev))
  rmse <- sqrt(mean((real - prev)^2))
  mape <- mean(
    abs((real - prev) / ifelse(real == 0, NA, real)),
    na.rm = TRUE
  ) * 100

  tibble(
    tipo_acomodacao = tipo_nome,
    MAE_arima = mae,
    RMSE_arima = rmse,
    MAPE_arima = mape
  )
}

# Bindagem das metricas em data table
metricas_arima <- bind_rows(
  calc_metricas_arima_full(enf_ts$mensal, melhor_enf, "Enfermaria"),
  calc_metricas_arima_full(apt_ts$mensal, melhor_apt, "Apartamento"),
  calc_metricas_arima_full(uti_ts$mensal, melhor_uti, "UTI")
)

metricas_arima

## 7.2) Métricas ETS (A,N,A) (2022-2025)
calc_metricas_ets <- function(serie_ts, modelo_ets, tipo_nome) {
  # previsão in-sample (valores ajustados) em toda a série
  fitted_vals <- fitted(modelo_ets)

  # Garantir comprimentos iguais
  n <- min(length(serie_ts), length(fitted_vals))
  real <- as.numeric(serie_ts[1:n])
  prev <- as.numeric(fitted_vals[1:n])

  mae <- mean(abs(real - prev))
  rmse <- sqrt(mean((real - prev)^2))
  mape <- mean(
    abs((real - prev) / ifelse(real == 0, NA, real)),
    na.rm = TRUE
  ) * 100

  tibble(
    tipo_acomodacao = tipo_nome,
    MAE_ets = mae,
    RMSE_ets = rmse,
    MAPE_ets = mape
  )
}

# Bindagem das metricas em data table
metricas_ets <- bind_rows(
  calc_metricas_ets(enf_ts$mensal, fit_ets_enf, "Enfermaria"),
  calc_metricas_ets(apt_ts$mensal, fit_ets_apt, "Apartamento"),
  calc_metricas_ets(uti_ts$mensal, fit_ets_uti, "UTI")
)

```

```

metricas_ets

## 7.3) Métricas XGBOOST (2022-2025)

metricas_xgb

## 7.4) Table Comparativa: ARIMA vs EST(A,N,A) vs XGBOOST (2022-2025)

tab_metricas_modelos <- bind_rows(
  metricas_arima %>%
    transmute(
      tipo_acomodacao,
      modelo = "ARIMA",
      MAE = MAE_arima,
      RMSE = RMSE_arima,
      MAPE = MAPE_arima
    ),
  metricas_ets %>%
    transmute(
      tipo_acomodacao,
      modelo = "ETS(A,N,A)",
      MAE = MAE_ets,
      RMSE = RMSE_ets,
      MAPE = MAPE_ets
    ),
  metricas_xgb %>%
    transmute(
      tipo_acomodacao,
      modelo = "XGBoost",
      MAE = MAE,
      RMSE = RMSE,
      MAPE = MAPE
    )
)

# a) Enfermaria
metricas_enfermaria <- tab_metricas_modelos %>%
  dplyr::filter(tipo_acomodacao == "Enfermaria")

# b) Apartamento
metricas_apartamento <- tab_metricas_modelos %>%
  dplyr::filter(tipo_acomodacao == "Apartamento")

# c) UTI
metricas_uti <- tab_metricas_modelos %>%
  dplyr::filter(tipo_acomodacao == "UTI")

metricas_enfermaria
metricas_apartamento
metricas_uti

#### 08) PROJEÇÃO DOS MODELOS VENCEDORES ####
# Modelo vencedor: XGBoost para Apartamento e UTI

## 8.1) XGBoost para Apartamento, Enfermaria e UTI - Jan à dez 2026

# Função para projetar 12 meses à frente com XGBoost usando apenas lag1
projeta_xgb_12m <- function(df_hist, modelo_xgb, tipo_nome, h = 12) {
  df_hist <- df_hist %>% arrange(data_mes)

  # Última data e último valor real
  ultima_data <- max(df_hist$data_mes)
  ultimo_valor <- tail(df_hist$tx_ocupacao_mensal, 1)

  # Sequência de meses futuros
  datas_fut <- seq.Date(from = ultima_data %m+% months(1),
                       by = "month",
                       length.out = h)

  preds <- numeric(h)
  lag1 <- ultimo_valor

  for (i in seq_len(h)) {
    ano_num <- lubridate::year(datas_fut[i])
    mes_num <- lubridate::month(datas_fut[i])
  }
}

```

```

X_new <- matrix(c(ano_num, mes_num, lag1), nrow = 1)
dnew <- xgb.DMatrix(data = X_new)

preds[i] <- predict(modelo_xgb, dnew)
lag1 <- preds[i] # para o próximo mês
}

tibble(
  tipo_acomodacao = tipo_nome,
  ano = lubridate::year(datas_fut),
  mes = lubridate::month(datas_fut),
  data_mes = datas_fut,
  tx_ocupacao_prevista = preds
)
}

# Histórico dados Apartamento
hist_ap <- ocupacao_mensal %>%
  dplyr::filter(tipo_acomodacao == "Apartamento") %>%
  arrange(data_mes)

mod_ap <- resultados_xgb %>%
  dplyr::filter(tipo_acomodacao == "Apartamento") %>%
  pull(modelo) %>%
  .[[1]]

proj_apt_12m <- projeta_xgb_12m(hist_ap, mod_ap, "Apartamento", h = 12)

# Histórico dados UTI
hist_uti <- ocupacao_mensal %>%
  dplyr::filter(tipo_acomodacao == "UTI") %>%
  arrange(data_mes)

mod_uti <- resultados_xgb %>%
  dplyr::filter(tipo_acomodacao == "UTI") %>%
  pull(modelo) %>%
  .[[1]]

proj_uti_12m <- projeta_xgb_12m(hist_uti, mod_uti, "UTI", h = 12)

# Histórico dados Enfermaria
hist_enf <- ocupacao_mensal %>%
  dplyr::filter(tipo_acomodacao == "Enfermaria") %>%
  arrange(data_mes)

mod_enf <- resultados_xgb %>%
  dplyr::filter(tipo_acomodacao == "Enfermaria") %>%
  pull(modelo) %>%
  .[[1]]

proj_enf_12m <- projeta_xgb_12m(hist_enf, mod_enf, "Enfermaria", h = 12)

# Tabela final XGBoost (jan-dez/2026)
projecoes_xgb_12m <- bind_rows(proj_apt_12m, proj_uti_12m, proj_enf_12m)

print(projecoes_xgb_12m, n = 36)

# EM GRÁFICO:
proj_xgboost <- ocupacao_mensal %>%
  filter(tipo_acomodacao %in% c("Apartamento", "UTI", "Enfermaria")) %>%
  select(tipo_acomodacao, data_mes, tx_ocupacao_mensal) %>%
  mutate(origem = "Real") %>%
  bind_rows(
    projecoes_xgb_12m %>%
      rename(tx_ocupacao_mensal = tx_ocupacao_prevista) %>%
      mutate(origem = "Previsto") %>%
      select(tipo_acomodacao, data_mes, tx_ocupacao_mensal, origem)
  ) %>%
  arrange(tipo_acomodacao, data_mes)

plot_proj_xgboost <- function(df_tipo, titulo) {
  ggplot(df_tipo, aes(x = data_mes, y = tx_ocupacao_mensal, colour = origem)) +
    geom_line(size = 1) +
    scale_colour_manual(
      values = c("Real" = "black", "Previsto" = "red"),

```

```

    name = ""
  ) +
  labs(
    title = titulo,
    x = "Mês/Ano",
    y = "Taxa de ocupação mensal"
  ) +
  scale_x_date(date_labels = "%m/%Y", date_breaks = "12 month") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
}

graficos_proj_xgboost <- proj_xgboost %>%
  group_by(tipo_acomodacao) %>%
  group_split()

# Plot Apartamento
print(
  plot_proj_xgboost(
    graficos_proj_xgboost[[which(sapply(graficos_proj_xgboost, function(df)
unique(df$tipo_acomodacao)) == "Apartamento")]],
    "Apartamento - XGBoost - Jan à Dez 2026"
  )
)

# Plot UTI
print(
  plot_proj_xgboost(
    graficos_proj_xgboost[[which(sapply(graficos_proj_xgboost, function(df)
unique(df$tipo_acomodacao)) == "UTI")]],
    "UTI - XGBoost - Jan à Dez 2026"
  )
)

# Plot Enfermaria
print(
  plot_proj_xgboost(
    graficos_proj_xgboost[[which(sapply(graficos_proj_xgboost, function(df)
unique(df$tipo_acomodacao)) == "Enfermaria")]],
    "Enfermaria - XGBoost - Jan à Dez 2026"
  )
)

## 8.2) Previsão ETS(A,N,A) para Enfermaria - Jan à dez 2026
#Modelo Vencedor: ETS(A,N,A) para Enfermaria

fc_ets_enf_12 <- forecast(fit_ets_enf, h = 12)

prev_ets_enf_12 <- tibble(
  tipo_acomodacao = "Enfermaria",
  mes_ts = time(fc_ets_enf_12$mean),
  mes_ano = zoo::as.yearmon(mes_ts),
  data_mes = zoo::as.Date(mes_ano),
  tx_ocupacao_prevista = as.numeric(fc_ets_enf_12$mean),
  li_95 = as.numeric(fc_ets_enf_12$lower[, "95%"]),
  ls_95 = as.numeric(fc_ets_enf_12$upper[, "95%"])
)

prev_ets_enf_12

# Histórico mensal da Enfermaria

hist_enf <- ocupacao_mensal %>%
  filter(tipo_acomodacao == "Enfermaria") %>%
  select(tipo_acomodacao, data_mes, tx_ocupacao_mensal) %>%
  arrange(data_mes)

# Criando dataframe com o histórico + projeção

serie_enf_hist <- hist_enf %>%
  mutate(origem = "Real")

serie_enf_prev <- prev_ets_enf_12 %>%
  transmute(
    tipo_acomodacao,
    data_mes,
    tx_ocupacao_mensal = tx_ocupacao_prevista,

```

```

    origem          = "Previsto",
    li_95,
    ls_95
  )

serie_enf_hist_proj <- bind_rows(
  serie_enf_hist,
  serie_enf_prev
) %>%
  arrange(data_mes)

## 8.3) Gráfico histórico + projeção com IC 95 %

plot_hist_proj_ets_enf <- ggplot() +
  # faixa de confiança 95% (apenas para o previsto)
  geom_ribbon(
    data = serie_enf_prev,
    aes(x = data_mes, ymin = li_95, ymax = ls_95),
    fill = "red",
    alpha = 0.15
  ) +
  # série real
  geom_line(
    data = serie_enf_hist,
    aes(x = data_mes, y = tx_ocupacao_mensal, colour = "Real"),
    size = 1
  ) +
  # série prevista (média)
  geom_line(
    data = serie_enf_prev,
    aes(x = data_mes, y = tx_ocupacao_mensal, colour = "Previsto"),
    size = 1,
    linetype = "dashed"
  ) +
  scale_colour_manual(
    values = c("Real" = "black", "Previsto" = "red"),
    name = ""
  ) +
  labs(
    title = "Previsão Enfermaria - ETS(A,N,A) + IC 95% - Jan à Dez 2026",
    x = "Mês/Ano",
    y = "Taxa de ocupação"
  ) +
  scale_x_date(date_labels = "%m/%Y", date_breaks = "12 month") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

print(plot_hist_proj_ets_enf)

#### FIM DO SCRIPT ####

```