



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO ACADÊMICO DO AGRESTE
NÚCLEO DE TECNOLOGIA
CURSO DE BACHARELADO EM ENGENHARIA DE PRODUÇÃO

RAIANE KELI DE SOUZA

ANÁLISE ESTATÍSTICA DA VIOLÊNCIA DOMÉSTICA EM PERNAMBUCO:
Tendências temporais e correlação com feminicídios

Caruaru
2025

RAIANE KELI DE SOUZA

ANÁLISE ESTATÍSTICA DA VIOLÊNCIA DOMÉSTICA EM PERNAMBUCO:

Tendências temporais e correlação com feminicídios

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Produção da Universidade Federal de Pernambuco - UFPE, como requisito parcial para a obtenção do título de Bacharel em Engenharia de Produção.

Área de concentração: Pesquisa Operacional.

Professor: Dr. Lucio Camara e Silva

Caruaru

2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Souza, Raiane Keli de.

ANÁLISE ESTATÍSTICA DA VIOLÊNCIA DOMÉSTICA EM
PERNAMBUCO: Tendências temporais e correlação com feminicídios / Raiane
Keli de Souza. - Caruaru, 2025.

81 : il., tab.

Orientador(a): Dr. Lucio Camara e Silva

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de
Pernambuco, Centro Acadêmico do Agreste, Engenharia de Produção, 2025.

Inclui referências, apêndices.

1. previsão. 2. violência doméstica. 3. feminicídio. 4. SARIMA. 5. séries
temporais. 6. políticas públicas. I. Silva, Dr. Lucio Camara e. (Orientação). II.
Título.

310 CDD (22.ed.)

RAIANE KELI DE SOUZA

**ANÁLISE ESTATÍSTICA DA VIOLÊNCIA DOMÉSTICA DE PERNAMBUCO:
TENDÊNCIAS TEMPORAIS E CORRELAÇÃO COM FEMINICÍDIO.**

Trabalho de Conclusão de Curso
apresentado à Coordenação do Curso de
Engenharia de Produção do Campus
Agreste da Universidade Federal de
Pernambuco – UFPE, na modalidade de
monografia, como requisito parcial para a
obtenção do grau de bacharel em
Engenharia de Produção.

Aprovada em: 14/08/2025

BANCA EXAMINADORA

Prof. Dr. Lúcio Camara e Silva (Orientador)
Universidade Federal de Pernambuco

Prof. Dr. Lucimário Gois de Oliveira Silva (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. Thyago Celso Cavalcante Nepomuceno (Examinador Externo)
Universidade Federal de Pernambuco

Dedicatória: Dedico este trabalho primeiramente a Deus, pela força, proteção e por ser o alicerce da minha vida.

À minha mãe, pelo amor incondicional e por nunca ter medido esforços para minha formação. Ao meu irmão Ricardo Fagner, pela parceria e incentivo em todos os momentos. Às minhas amigas Ilayne Viana, Eduarda Maria e Tarcila Santana, pela amizade sincera e apoio constante ao longo dessa caminhada.

E, com especial gratidão, ao meu companheiro Rafael Guimarães, que acreditou em mim mesmo quando eu mesma duvidei, oferecendo amor, paciência e motivação para que eu chegasse até aqui.

RESUMO

Este trabalho analisa estatisticamente os registros de violência doméstica e familiar contra a mulher (VDFCM) no estado de Pernambuco, no período de 2017 a 2025, com o objetivo de investigar a existência de correlação com os casos de feminicídio. A pesquisa adota uma abordagem quantitativa, utilizando modelos de séries temporais (ARIMA, SARIMA e ARIMAX) para identificar tendências, padrões sazonais e o impacto de variáveis exógenas, como a pandemia da COVID-19 e a implementação da política pública “Juntos pela Segurança”. Os resultados revelam forte dependência temporal nas séries e padrões sazonais recorrentes, com aumento nos registros após o relaxamento das medidas sanitárias. As análises de correlação não indicaram associação estatística significativa entre os registros de VDFCM e os feminicídios. Observou-se, ainda, que eventos críticos como a pandemia influenciaram a redução nos registros, enquanto não se constatou efeito estatisticamente significativo da política pública analisada. O estudo contribui com evidências empíricas relevantes para o aprimoramento das estratégias de enfrentamento à violência de gênero e a formulação de políticas públicas mais eficazes no estado de Pernambuco.

Palavras-chave: violência doméstica; feminicídio; séries temporais; SARIMA; previsão; políticas públicas.

ABSTRACT

This study statistically analyzes the records of domestic and family violence against women in the state of Pernambuco, Brazil, from 2017 to 2025, aiming to investigate the existence of a correlation with cases of femicide. A quantitative approach was adopted, applying time series models (ARIMA, SARIMA, and ARIMAX) to identify trends, seasonal patterns, and the impact of exogenous variables such as the COVID-19 pandemic and the implementation of the public policy “Juntos pela Segurança”. The results reveal strong temporal dependence and recurring seasonal patterns, with an increase in reported cases after the easing of sanitary restrictions. Correlation analyses did not indicate a statistically significant association between violence against women records and femicide cases. It was also observed that critical events, such as the pandemic, contributed to a reduction in the number of reports, while no statistically significant effect was found regarding the evaluated public policy. The study provides empirical evidence to support governmental actions and the improvement of strategies to combat gender-based violence in the state of Pernambuco.

Keywords: domestic violence; femicide; time series; ARIMA; forecast; public policy

LISTA DE ILUSTRAÇÕES

Figura 1 – Etapas do estudo - Fluxograma	27
Figura 2 – Regiões Geográficas utilizadas pela SDS - PE	29
Figura 3 – Evolução populacional e projeções em Pernambuco	30
Figura 4 – Série mensal de Feminicídio em Pernambuco - 2017 à 2025	31
Figura 5 – Decomposição da Série de Feminicídio em Pernambuco	32
Figura 6 – Diagrama de caixa para Feminicídio em Pernambuco	33
Figura 7 – Histograma e curva de densidade de Feminicídio	33
Figura 8 – Autocorrelação e autocorrelação parcial Feminicídio	34
Figura 9 – Série mensal de Violência Doméstica familiar contra a mulher em Pernambuco	35
Figura 10 – Decomposição da Série de Violência Doméstica familiar contra a mulher em Pernambuco	36
Figura 11 – Sazonalidade da Série de Violência Doméstica familiar contra a mulher em Pernambuco	37
Figura 12 – Diagrama de caixa para VDFCM em Pernambuco	37
Figura 13 – Autocorrelação VDFCM Pernambuco	38
Figura 14 – Histograma Violência Doméstica familiar contra a mulher em Pernambuco - 2017 á 2025	39
Figura 15 – Correlação de Pearson VDFCM e feminicídio em pernambuco	40
Figura 16 – Violência Doméstica familiar contra a mulher em Pernambuco - Analise pandêmica	41
Figura 17 – Série Temporal de VDFCM com Indicação da Introdução de Política Pública	43
Figura 18 – Gráfico de Dispersão das series diferenciadas	46
Figura 19 – Q-QPlot de recorte Treino da Série de VDFCM	47
Figura 20 – Recorte de Treino da Série de VDFCM: Histograma e QQ-plot	47
Figura 21 – Recorte de Treino da da Série de VDFCM: ACF e PACF	48
Figura 22 – Análise do resíduo do modelo SARIMA - Gráficos de linha e histograma	50
Figura 23 – Análise do resíduo do modelo SARIMA - Gráficos ACF e PACF	50
Figura 24 – Modelo SARIMA no conjunto Teste	51
Figura 25 – Filtro do modelo de previsão SARIMA no conjunto teste	51
Figura 26 – Serie Historia Modelagem SARIMA e Previsão	53
Figura 27 – Filtro em 2025 modelagem SARIMA e Previsão	53
Figura 28 – Modelo ARIMA com Dummy Exógena e Previsão: COVID-19	56
Figura 29 – Modelo ARIMA previsão sem a COVID-19	57
Figura 30 – Modelo ARIMA com Dummy Exógena e Previsão: Politica Publica	59

LISTA DE TABELAS

Tabela 1 – Testes de normalidade para a série de feminicídios mensais	34
Tabela 2 – Resultados dos testes de Levene e Shapiro-Wilk: COVID-19	41
Tabela 3 – Resultados dos testes de Levene e Shapiro-Wilk: Política Publica	43
Tabela 4 – Resultados dos modelos SARIMA testados	49
Tabela 5 – Comparação entre valores previstos e observados (dez/2023 a jun/2025)	52
Tabela 6 – Previsão do número de vítimas e erro padrão para os cinco períodos seguintes	54
Tabela 7 – Comparação das médias de vítimas antes e pós-pandemia	54
Tabela 8 – Comparação de modelos ARIMA com variável exógena (xreg)	55
Tabela 9 – Teste t pareado entre valores observados e contrafactuais (pós-pandemia)	57
Tabela 10 – Comparação das médias de vítimas antes e após a política pública	58
Tabela 11 – Comparação dos modelos ARIMAX para a política “Juntos pela Segurança”	58
Tabela 12 – Comparação entre valores observados e previstos pelos modelos ARIMAX — Teste t pareado	58

SUMÁRIO

1	INTRODUÇÃO	11
1.1	CONTEXTUALIZAÇÃO	11
1.2	ANÁLISE DO PROBLEMA	12
1.3	JUSTIFICATIVA E RELEVÂNCIA DO TRABALHO	13
1.4	OBJETIVOS DA PESQUISA	13
1.4.1	Objetivos Específicos	14
1.5	ESTRUTURA DO TRABALHO	14
2	REFERENCIAL TEÓRICO	15
2.1	FATORES QUE PERPETUAM A VIOLÊNCIA INTIMA AFETIVA CONTRA A MULHER	15
2.2	VIOLÊNCIA DOMESTICA E FEMINICÍDIO: LEGISLAÇÕES E POLÍTICAS PUBLICAS	16
2.3	CIÊNCIAS DE DADOS	19
2.3.1	Evolução histórica da análise de séries temporais	19
2.3.2	Modelos Preditivos	20
2.3.3	Aplicação de Ciência de Dados e modelos preditivos na análise criminal..	21
2.3.4	Modelos de previsão em Violência contra a mulher	22
2.3.5	Avaliação de Impactos com Variáveis Exógenas	23
2.3.6	Análise de Correlação entre Séries Temporais	25
3	METODOLOGIA	27
3.1	DESCRIÇÃO E COLETA DOS DADOS	27
3.2	ANÁLISE EXPLORATÓRIA DE DADOS	30
3.2.1	Evolução dos Casos de Feminicídio em Pernambuco	30
3.2.2	Evolução dos Casos de Violência Doméstica (VDFCM)	35
3.3	MÉTODOS ESTATÍSTICOS E MODELOS APLICADOS	39
3.3.1	Análise de Correlação entre Feminicídio e VDFCM	39
3.3.2	Análise de Impacto da Pandemia de COVID-19	40
3.3.3	Avaliação das Políticas Públicas Estaduais	42
3.3.4	Modelos de previsão	44
4	RESULTADOS	45
4.1	CONSIDERAÇÕES SOBRE AS CORRELAÇÕES ESTUDADAS	45
4.2	INTERPRETAÇÃO DOS MODELOS E PREVISÕES	46
4.3	DISCUSSÃO DOS IMPACTOS COM VARIÁVEIS EXÓGENAS	54
4.3.1	Discussão dos Impactos da pandemia do COVID-19	54
4.3.2	Discussão dos Impactos de políticas publicas	57

5	CONCLUSÃO	60
	REFERÊNCIAS	62
	APÊNDICE A – CÓDIGO-FONTE EM R PARA ANÁLISE DE CORRE- LAÇÃO	68
	APÊNDICE B – CÓDIGO-FONTE EM ANÁLISE DE PREVISÕES.....	70
	APÊNDICE C – CÓDIGO-FONTE EM R PARA ANÁLISE DE COM AS VARIÁVEIS EXÓGENAS.....	75

1 INTRODUÇÃO

1.1 CONTEXTUALIZAÇÃO

A violência é uma das mais graves e persistentes violações dos direitos humanos, afetando milhões de pessoas em todo o mundo e deixando marcas profundas na sociedade. A Organização Mundial da Saúde (OMS) (2025) define violência como o uso intencional de força física ou poder, seja para ameaçar ou efetivamente praticar atos contra si próprio, outra pessoa ou uma comunidade, resultando em ferimentos, danos psicológicos, desenvolvimento prejudicado, privação de liberdade ou morte.

Dentre suas diversas manifestações, a violência doméstica destaca-se por sua natureza alarmante, pois é cometida por pessoas próximas às vítimas, como parceiros ou familiares, com o objetivo de exercer controle ou poder. Esse tipo de abuso ocorre independentemente de classe social, nível de escolaridade, orientação sexual ou gênero (REYAL et al., 2024). Dados do Escritório das Nações Unidas sobre Drogas e Crime (United Nations Office on Drugs and Crime, 2024) revelam que 60% das mortes intencionais de mulheres ocorrem no contexto íntimo-afetivo, em contraste com apenas 12% entre os homens. Esses números reforçam a estreita associação entre violência doméstica e violência de gênero, evidenciando uma problemática social que exige atenção urgente e medidas eficazes de enfrentamento.

A violência doméstica é um problema global e persistente, que afeta milhões de pessoas em todo o mundo e muitos casos passa de maneira despercebida em meio a sociedade. Segundo a ONU estima que uma a cada três mulheres e meninas no mundo já sofreram algum tipo de violência física ou sexual de seus parceiros ou parentes próximos (UN Women, 2024). Além de causar danos emocionais e psicológicos profundos, esses abusos podem e evoluir para números de violências letais onde tracem danos irreparáveis, a morte por parte da vitimas. United Nations Office on Drugs and Crime (2024) aponta que cerca de 58% dos homicídios de mulheres são decorrentes de abusos cometidos por parceiros íntimos ou familiares nas Américas, no mundo como já vimos essa porcentagem é ainda maior e corresponde a aproximadamente 51.100 mortes de mulheres e meninas em 2023.

Compreender a complexidade desse fenômeno exige um olhar mas amplo, não apenas a análise de indicadores, mas também um olhar teórico que revele as estruturas socioculturais que o sustentam. No plano teórico, autores como Scott (1990) contribuem para a compreensão do gênero como uma categoria útil de análise histórica, revelando como estruturas simbólicas e institucionais sustentam práticas de dominação. Segato (2011), por sua vez, propõe uma leitura decolonial da violência de gênero, apontando que o feminicídio se articula com padrões coloniais de controle sobre os corpos femininos. Narvaz e Koller (2006) ressaltam a marginalização dos estudos feministas dentro da psicologia acadêmica, o que contribui para a invisibilização institucional da violência contra a mulher.

No contexto brasileiro, o cenário é igualmente preocupante. Apesar dos avanços legislativos e das políticas públicas voltadas à proteção das mulheres, como a Lei Maria da Penha 2006 e a tipificação do feminicídio como crime hediondo 2015, os índices de violência doméstica permanecem elevados e, em muitos casos, não são devidamente notificados. Especificamente em Pernambuco, dados dos últimos anos revelam um crescimento alarmante nos registros de violência doméstica e casos de feminicídio, mesmo com a expansão de medidas protetivas e ações de enfrentamento.

1.2 ANÁLISE DO PROBLEMA

Em 2024, o Brasil registrou mais de 1.492 casos de feminicídio, o maior número desde a promulgação da Lei do Feminicídio (2015) onde 64,3% das vítimas foram mortas em casa. Em Pernambuco, desde a vigência da lei, mais de 500 mulheres foram assassinadas em circunstâncias tipificadas como feminicídio, sendo 125 dessas mortes registradas apenas entre 2024 e 2025 (Fórum Brasileiro de Segurança Pública, 2025; SDS-PE, 2025).

Diante dessa realidade, torna-se essencial compreender, com base em evidências empíricas, se há relação estatística entre os registros de violência doméstica e a ocorrência de feminicídios em Pernambuco. Investigar essa possível conexão é fundamental para avaliar a efetividade das políticas públicas de prevenção, identificar falhas no sistema de proteção às mulheres e subsidiar estratégias mais eficazes de intervenção.

Além da relevância social e jurídica, o estudo da violência doméstica e do feminicídio demanda abordagens metodológicas capazes de lidar com grandes volumes de dados, múltiplas variáveis e padrões complexos de ocorrência (BOX; JENKINS, 1970). Nesse contexto, a Ciência de Dados se destaca como um campo essencial, ao integrar técnicas estatísticas avançadas, análise exploratória, modelagem preditiva e métodos computacionais para extrair conhecimento a partir de informações estruturadas e não estruturadas (VERREY et al., 2023). A utilização de modelos de previsão, como os de séries temporais, possibilita identificar tendências, sazonalidades e anomalias nos registros, disposto anteriormente por (Shift / University of Calgary, 2022; NEUBAUER, 2024) bem como avaliar o impacto de variáveis externas, a exemplo da pandemia de COVID-19 e da implementação do programa “Juntos pela Segurança”, sobre os indicadores de violência de gênero no estado, auxiliando no planejamento e na avaliação de políticas públicas.

O presente trabalho foi estruturado de forma a integrar análises descritivas, testes estatísticos e modelagem preditiva, permitindo tanto a compreensão histórica dos fenômenos quanto a projeção de cenários futuros. De maneira sintética, os resultados obtidos indicaram forte dependência temporal nas séries de violência doméstica, padrões sazonais recorrentes e variações associadas a eventos críticos, mas não evidenciaram correlação estatisticamente significativa entre violência doméstica e feminicídios. Essas constatações são discutidas ao longo do texto.

1.3 JUSTIFICATIVA E RELEVÂNCIA DO TRABALHO

A violência doméstica contra a mulher é um fenômeno persistente e alarmante no Brasil. Segundo o Fórum Brasileiro de Segurança Pública (2025), somente em 2024 o país registrou cerca de 260 mil casos de lesão corporal dolosa contra mulheres, o que equivale a, em média, 710 vítimas por dia. Em Pernambuco, o cenário segue a mesma tendência, comparando registros totais de agressões foram mais de 54 mil registros de violência familiar contra mulheres no mesmo período (SDS-PE, 2025).

Apesar dos avanços legislativos, como a Lei Maria da Penha, que ampliou os mecanismos de proteção e facilitou o acesso à justiça, os dados mostram que as medidas legais ainda não têm sido suficientes para conter os casos mais graves. Em 2024, registrou-se no Brasil, aproximadamente 72 mulheres por hora solicitaram Medida Protetiva de Urgência (MPU), representando um aumento em relação a anos anteriores. Mesmo com esse crescimento, que corresponde a 7,2% no número de medidas protetivas entre 2023 e 2024, os casos de feminicídio apresentaram crescimento, com alta de 0,7% e tentativas de feminicídio aumentando 19% no mesmo período. Ressalta-se, ainda, a ocorrência de registros de mortes de mulheres que estavam sob proteção judicial no momento do crime, o que levanta questionamentos sobre a eficácia das políticas públicas existentes (Fórum Brasileiro de Segurança Pública, 2025).

Esses dados evidenciam que fatores externos, como a pandemia de COVID-19, podem ter influenciado a dinâmica da violência, alterando o comportamento dos indicadores de forma mais significativa que as ações governamentais isoladas.

Diante desse contexto, este estudo justifica-se pela necessidade de compreender se as denúncias registradas em estágios iniciais da violência possuem relação com desfechos extremos, como o feminicídio, e de avaliar o impacto de políticas públicas e eventos críticos sobre a evolução desses indicadores. A análise proposta busca, assim, fornecer evidências empíricas que possam subsidiar o aprimoramento de estratégias de prevenção e enfrentamento da violência de gênero no estado de Pernambuco e no Brasil.

A relevância do trabalho reside na possibilidade de gerar evidências empíricas que orientem políticas públicas mais eficazes e ações preventivas direcionadas, fortalecendo a atuação de órgãos de segurança, justiça e assistência social. No campo acadêmico, o estudo contribui para a literatura multidisciplinar sobre violência de gênero, ao aplicar métodos estatísticos robustos e técnicas de séries temporais robustas e integrados à realidade local. Além disso, possui relevância prática ao propor o uso de modelagem preditiva como instrumento para antecipar cenários e subsidiar decisões estratégicas no enfrentamento à violência de gênero.

1.4 OBJETIVOS DA PESQUISA

Analisar estatisticamente os registros de violência doméstica e familiar contra a mulher (VDFCM) no estado de Pernambuco entre os anos de 2017 e 2025, investigar a existência de correlação com os casos de feminicídio, considerando o impacto de políticas públicas e

eventos críticos, como a pandemia de COVID-19, por meio de modelos estatísticos e de séries temporais.

1.4.1 Objetivos Específicos

- Coletar, organizar e tratar os dados estatísticos sobre violência doméstica e feminicídio registrados no estado de Pernambuco entre 2017 e 2025.
- Construir e aplicar o modelo autorregressivo integrado de médias móveis (ARIMA, SARIMA) para identificar padrões, tendências e sazonalidades nos registros de VDFCM.
- Calcular e interpretar coeficientes de correlação estatística entre as variáveis de VDFCM e feminicídio;
- Analisar os efeitos de variáveis exógenas, como a pandemia de COVID-19 e a política pública “Juntos pela Segurança”, sobre os registros mensais.
- Aplicar modelos autorregressivo para identificar projeções sobre o comportamento futuro da violência doméstica no estado.

1.5 ESTRUTURA DO TRABALHO

Este trabalho está organizado em cinco seções principais. Após esta introdução, apresenta-se a Seção 2, com o referencial teórico, abordando os fundamentos legais, conceituais e empíricos da violência doméstica e do feminicídio. A Seção 3 descreve a metodologia adotada, incluindo as fontes de dados, as técnicas estatísticas utilizadas e os critérios para análise.

Na Seção 4, são apresentados e discutidos os resultados da análise estatística, com destaque para os testes de correlação, modelagem de séries temporais e interpretação crítica dos achados. Por fim, a Seção 5 traz as conclusões do estudo, suas implicações práticas, limitações e sugestões para pesquisas futuras.

2 REFERENCIAL TEÓRICO

Este capítulo tem como objetivo contextualizar o leitor no tema abordado. Para isso, iniciasse com uma introdução as bases conceituais, legais e metodológicas que sustentam o presente estudo, articulando perspectivas das ciências sociais, jurídicas e da ciência de dados. Primeiramente, abordam-se os fatores socioculturais que sustentam a violência íntima contra a mulher, seguidos pela análise do arcabouço legislativo e das políticas públicas voltadas à prevenção e combate dessa violência. Por fim, são exploradas as contribuições da ciência de dados, com foco em modelos de séries temporais e suas aplicações na análise de violência de gênero.

2.1 FATORES QUE PERPETUAM A VIOLÊNCIA ÍNTIMA AFETIVA CONTRA A MULHER

A violência doméstica contra a mulher resulta de um conjunto de fatores interligados que operam em níveis individuais, relacionais, culturais e institucionais, perpetuando o ciclo de abusos ao longo do tempo. Esses fatores envolvem tanto dimensões subjetivas, como o medo e a culpa, quanto barreiras sociais e estruturais que dificultam o rompimento da relação violenta (WALKER, 1979; D'OLIVEIRA; SCHRAIBER, 1999; MINAYO, 2006; SCHRAIBER; D'OLIVEIRA; COUTO, 2007a).

Segundo Minayo (2006), a naturalização da violência no ambiente doméstico é sustentada por padrões patriarcais e misóginos, historicamente arraigados na cultura brasileira, que atribuem à mulher uma posição de subordinação e dependência em relação ao homem. Essa normalização social contribui para a tolerância institucional e coletiva diante de casos de violência doméstica, especialmente em contextos de vulnerabilidade social.

Um dos modelos explicativos mais influentes para compreender a persistência das agressões é o ciclo da violência, proposto por Walker (1979). O modelo identifica três fases recorrentes que mantêm a mulher presa à relação abusiva: a fase de tensão crescente, seguida pelo episódio de agressão, e, por fim, a fase de reconciliação ou “lua de mel”, na qual o agressor demonstra arrependimento e afeto, reforçando a esperança de mudança e contribuindo para a permanência da vítima na relação. Além disso, aspectos emocionais e econômicos exercem influência direta na dificuldade de ruptura. Schraiber, D'Oliveira e Couto (2007a) destaca que a dependência financeira e emocional é um dos principais obstáculos enfrentados pelas mulheres. A ausência de autonomia econômica, somada ao medo de retaliações e à insegurança sobre o futuro, reforça a sensação de aprisionamento. Nesse sentido, D'Oliveira e Schraiber (1999) afirmam que muitas mulheres permanecem nas relações abusivas por sentimento de culpa, isolamento social e pela dificuldade de projetar uma vida fora daquela dinâmica afetiva. A dependência subjetiva, somada ao afastamento familiar imposto pelo agressor, compromete a rede de apoio, dificultando o acesso à proteção e aos serviços públicos.

No plano sociocultural, Saffioti (1999) argumenta que o provérbio popular “em briga de marido e mulher ninguém mete a colher” representa um mito social que sustenta a invisibilidade da violência doméstica, ao reforçar os limites entre o público e o privado. Essa naturalização do espaço doméstico como território inviolável contribui para a perpetuação de dinâmicas violentas e para a omissão institucional diante do sofrimento das mulheres. Dessa forma, como aponta Costa (2007), é fundamental romper com a dicotomia entre o espaço privado (doméstico) e o público, reafirmando que o Estado e a sociedade são corresponsáveis pela garantia dos direitos humanos, incluindo o direito das mulheres a uma vida livre de violência.

Nesse contexto, cabe ao Estado enfrentar não apenas os fatores subjetivos e culturais mencionados, mas também as falhas institucionais que contribuem para a reprodução da violência. A subnotificação de casos, a morosidade do sistema de justiça, a revitimização nos atendimentos institucionais e a falta de acolhimento qualificado nas delegacias e nos serviços de saúde ainda são obstáculos recorrentes. Muitas mulheres desistem de denunciar por medo de não serem acreditadas, por experiências anteriores de desproteção ou por represálias, aprofundando o ciclo de silenciamento e impunidade.

2.2 VIOLÊNCIA DOMESTICA E FEMINICÍDIO: LEGISLAÇÕES E POLITICAS PUBLICAS

A violência doméstica e o feminicídio são manifestações extremas da desigualdade de gênero e da cultura patriarcal, que ainda persistem estruturalmente na sociedade contemporânea, configurando-se como graves violações dos direitos humanos. Ambos os fenômenos têm recebido crescente atenção por parte de estudiosos, organismos internacionais e formuladores de políticas públicas, devido à sua recorrência, impactos sociais e psicológicos e, sobretudo, pela sua inserção nas dinâmicas de poder que atravessam as relações de gênero (NARVAZ; KOLLER, 2006; SCOTT, 1990; SEGATO, 2011).

A violência doméstica é um conceito amplo e no Brasil está definido pela **Lei nº 11.340/2006** (Lei Maria da Penha) como qualquer ação ou omissão baseada no gênero que cause morte, lesão, sofrimento físico, sexual, psicológico ou dano moral e patrimonial à mulher, ocorrida no âmbito da unidade doméstica, familiar ou em qualquer relação íntima de afeto (Brasil, 2006). A lei representou um marco histórico no combate à violência contra a mulher no Brasil, ao estabelecer mecanismos para prevenção, punição e erradicação da violência doméstica, além de criar juizados especializados e ampliar a rede de proteção.

Já o Feminicídio é definido juridicamente pela **Lei nº 13.104/2015**, que alterou o Código Penal brasileiro para incluir o feminicídio como circunstância qualificadora do crime de homicídio (Brasil, 2015). De acordo com essa lei, feminicídio é o assassinato de mulheres motivado por razões de gênero, especialmente nos contextos de violência doméstica e familiar, menosprezo ou discriminação à condição de mulher. Essa tipificação busca não apenas punir de forma mais severa os agressores, mas também destacar a dimensão estrutural do problema, que se baseia

em relações historicamente desiguais de poder entre homens e mulheres.

Desde então, a legislação brasileira passou por uma série de atualizações importantes que refletem um esforço contínuo de aperfeiçoamento institucional e jurídico no enfrentamento à violência de gênero. Destaca-se a Lei nº 13.641/2018, que tornou crime o descumprimento de medidas protetivas de urgência, com pena de detenção de três meses a dois anos (Brasil, 2018). Posteriormente, a Lei nº 14.188/2021 criou o Programa Sinal Vermelho contra a Violência Doméstica, permitindo que mulheres em situação de risco peçam ajuda em espaços públicos com um simples gesto: o “X” vermelho desenhado na palma da mão (Brasil, 2021; Conselho Nacional de Justiça, 2024).

Mais recentemente, o sistema judiciário brasileiro adotou medidas tecnológicas inovadoras, como a implantação da medida protetiva eletrônica, que dispensa o registro policial presencial e permite que a mulher solicite proteção por meio digital. O Tribunal de Justiça de Pernambuco (TJPE), por exemplo, passou a oferecer o serviço em 2024 por meio do site oficial e de aplicativos institucionais, reduzindo barreiras burocráticas e aumentando a agilidade da resposta judicial (Jornal Digital Recife, 2025).

Em 2023, também entrou em vigor a Lei nº 14.550, que estabelece prioridade no atendimento à mulher vítima de violência nas instituições públicas, e, no mesmo ano, a Lei nº 14.720 permitiu que juízes determinem medidas protetivas de ofício, mesmo sem solicitação formal da vítima, garantindo maior eficácia e rapidez na proteção (Brasil, 2023).

Em nível internacional, documentos como a Convenção sobre a Eliminação de Todas as Formas de Discriminação contra a Mulher (CEDAW), adotada Organização das Nações Unidas (1979) e entrou em vigor desde 1981, são fundamentais no combate à desigualdade de gênero. Conhecida como a “carta internacional dos direitos das mulheres”, a CEDAW define a discriminação de gênero e estabelece diretrizes para que os Estados Partes adotem medidas legislativas, educacionais e culturais para eliminá-la. Isso inclui a revisão de normas e estereótipos sociais e a implementação de ações afirmativas para assegurar igualdade substancial entre homens e mulheres (RADAY, 2012; HELLMUM; AASEN, 2013).

Na América Latina, destaca-se a Convenção Interamericana para Prevenir, Punir e Erradicar a Violência contra a Mulher, conhecida como Convenção de Belém do Pará, Organização dos Estados Americanos (1994) (OEA). Trata-se do primeiro tratado internacional juridicamente vinculante que reconhece a violência contra a mulher como violação dos direitos humanos, abrangendo a violência física, sexual e psicológica, tanto no espaço público quanto privado. A convenção também instituiu o Mecanismo de Seguimento da Convenção (MESECVI), que monitora a implementação pelos Estados-Membros. Um caso emblemático de aplicação da Convenção foi o da brasileira Maria da Penha, que resultou na responsabilização do Estado brasileiro por omissão e negligência e culminou na criação da Lei Maria da Penha (Brasil, 2006), já mencionada. Esse caso representa um marco na jurisprudência internacional em matéria de direitos humanos e violência de gênero.

No Brasil, diversas políticas públicas vêm sendo implementadas para enfrentar o problema,

como a criação das Delegacias Especializadas de Atendimento à Mulher (DEAMs), a expansão da Casa da Mulher Brasileira, a existência da Central de Atendimento 180, além de programas de capacitação, campanhas educativas e medidas protetivas previstas em lei. No entanto, desafios estruturais, como a insuficiência de recursos, a falta de profissionais qualificados e a precariedade da rede de apoio em áreas mais vulneráveis, ainda comprometem a eficácia dessas ações.

No contexto sociocultural de Pernambuco, o cenário é particularmente alarmante no que se refere à violência de gênero. Em 2023, o estado registrou 92 casos de feminicídio, o que representa um aumento de 55,9% em relação ao ano anterior, tornando-se a unidade federativa com o maior número de feminicídios do Nordeste (Marco Zero Conteúdo, 2023). As mesorregiões da Zona da Mata e do Sertão apresentaram elevação especialmente acentuada, evidenciada em estudos que mostram aumento médio de 69,7% nos homicídios femininos no Sertão entre 2018 e 2021, com tendência semelhante na Zona da Mata (Fórum Brasileiro de Segurança Pública, 2023). Tais dados reforçam que a violência letal contra mulheres em Pernambuco tem se expandido para além dos centros urbanos, enraizando-se também em contextos rurais e periféricos.

Esse agravamento da violência está associado a fatores estruturais como baixa escolaridade, dependência econômica e desigualdade social, especialmente entre mulheres com ensino médio incompleto e rendimentos reduzidos. De acordo com levantamento do Instituto DataSenado, mulheres com menor escolaridade e renda inferior são as que mais relatam episódios de violência doméstica, e também as que encontram maiores barreiras para romper o ciclo de agressões (Senado Federal (Brasil), 2024). A dependência financeira, por exemplo, é identificada por organizações locais como uma das principais razões para a permanência em relacionamentos abusivos. Iniciativas como as desenvolvidas pelo Núcleo Integrado de Justiça (NIOJ) do TJPE, em Caruaru, buscam enfrentar esse desafio por meio de ações de qualificação profissional e fortalecimento da autonomia das vítimas (Poder Judiciário do Estado de Pernambuco, 2024).

Outro aspecto crítico é a subnotificação, que dificulta o enfrentamento institucional à violência. Em 2023, cerca de 86% das mulheres vítimas de feminicídio no estado não haviam registrado boletim de ocorrência anteriormente, o que revela uma grave falha de acesso à rede de proteção e contribui para a invisibilidade estrutural do problema (Conselho Nacional de Justiça, 2023; Marco Zero Conteúdo, 2023). Essa lacuna compromete a eficácia das medidas protetivas e das políticas públicas, exigindo esforços de aproximação ativa e sensibilização comunitária por parte do poder público.

Diante desse cenário, o Governo do Estado de Pernambuco instituiu, em 2023, o programa Juntos pela Segurança, uma política pública baseada em evidências, metas interinstitucionais e participação social. Um dos eixos centrais do programa é o enfrentamento à violência de gênero, por meio do investimento em tecnologia para monitoramento de ocorrências, da ampliação da rede de acolhimento às vítimas e da articulação entre as áreas da segurança pública, justiça, assistência social e saúde. Essa estratégia sendo uma ampliação do Pacto pela Vida, já consolidado no estado desde 2007, como plano macro de combate à criminalidade, porém não

incluía diretamente estratégias de combate a violência doméstica. A expectativa é que, com medidas mais integradas e focalizadas, seja possível não apenas reduzir os índices de violência, mas também promover uma mudança estrutural e cultural no enfrentamento à desigualdade de gênero (Governo de Pernambuco, 2023).

Apesar desses avanços, indicadores em relatórios como o do Fórum Brasileiro de Segurança Pública (2025) indicam que políticas isoladas ainda não produzem quedas sustentáveis nos índices de violência. Assim, torna-se necessário avaliar o impacto real dessas medidas, tarefa para a qual a ciência de dados oferece ferramentas adequadas.

2.3 CIÊNCIAS DE DADOS

A Ciência de Dados tem se consolidado como uma abordagem estratégica no enfrentamento à violência doméstica e ao feminicídio, permitindo integrar estatística, computação e inteligência artificial para extrair padrões, prever ocorrências e apoiar políticas públicas baseadas em evidências. A aplicação dessas técnicas torna-se particularmente relevante diante dos desafios apontados nos tópicos anteriores, como a análise de padrões de violência, a dificuldade de avaliar o impacto real das medidas protetivas e a necessidade de respostas rápidas e focalizadas.

2.3.1 Evolução histórica da análise de séries temporais

A análise de séries temporais constitui o núcleo da modelagem de dados sequenciais, articulando estatística clássica, econometria, ciência de dados e aprendizado de máquina. Essa abordagem evoluiu ao longo de mais de três séculos, acompanhando os avanços metodológicos e computacionais no campo da estatística e das ciências sociais aplicadas.

O estudo de dados ao longo do tempo remonta ao estatístico inglês John Graunt, no século XVII, que analisou registros de mortalidade em Londres e utilizou métodos rudimentares de organização temporal para detectar padrões demográficos (Graunt, 1662). No século XIX, Thiele (1880) introduziu conceitos pioneiros de processos estocásticos que integravam elementos de regressão, movimento browniano e ruído branco, fundamentos que anteciparam o desenvolvimento moderno das séries temporais. Já no início do século XX, Yule (1927) desenvolveu os primeiros modelos autoregressivos (AR) com aplicação prática em dados reais.

A consolidação metodológica, contudo, ocorreu nas décadas de 1960 e 1970 com o trabalho seminal de Box e Jenkins (1970), que sistematizou os processos ARIMA, introduzindo uma abordagem estruturada de identificação, estimação, diagnóstico e previsão de séries temporais (BOX; JENKINS, 1970; BROCKWELL; DAVIS, 2002). Nos últimos anos, a integração entre estatística clássica e ciência de dados ampliou as possibilidades de análise, incorporando métodos de aprendizado de máquina, como redes neurais recorrentes (RNN) e modelos LSTM, capazes de capturar relações não lineares e dependências complexas nos dados (Hyndman e Athanasopoulos, 2018). Essa integração entre estatística clássica e algoritmos de IA ampliou

as possibilidades de análise no campo da segurança pública, incluindo o monitoramento e a previsão de indicadores de violência de gênero.

2.3.2 Modelos Preditivos

Segundo Box e Jenkins (1970), séries temporais podem ser eficientemente modeladas por meio de estruturas autorregressivas e de médias móveis, desde que atendam aos pressupostos de estacionariedade e dependência serial. A adoção de modelos clássicos, como ARIMA e SARIMA, justifica-se por sua robustez estatística, interpretabilidade e ampla consolidação na literatura aplicada. Assim, a modelagem das séries deste estudo foi orientada pelas etapas metodológicas propostas por esses autores, iniciando-se pela verificação dos pressupostos fundamentais, como estacionariedade, normalidade e estrutura de autocorrelação.

Entre os modelos utilizados, destaca-se o modelo de Média Móvel (MA – *Moving Average*), no qual o valor atual da série é expresso como uma combinação linear dos erros aleatórios de períodos anteriores. A forma geral do modelo MA(q) é dada por:

$$X_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \cdots + \theta_q\varepsilon_{t-q} \quad (2.1)$$

onde X_t representa o valor da série no tempo t , μ é a média da série, ε_t corresponde ao ruído branco (erro aleatório) e θ_i são os coeficientes de média móvel. Esse modelo é indicado para séries que apresentam flutuações aleatórias de curto prazo, capturando principalmente os efeitos transitórios.

Outro modelo fundamental é o Modelo Autorregressivo (AR – *Autoregressive*), que expressa o valor atual da série como uma função linear dos seus próprios valores defasados. Sua forma geral é:

$$X_t = \phi_1X_{t-1} + \phi_2X_{t-2} + \cdots + \phi_pX_{t-p} + \varepsilon_t \quad (2.2)$$

em que ϕ_i são os coeficientes autorregressivos e ε_t representa o erro aleatório. Esse tipo de modelo é apropriado para séries que apresentam persistência temporal, permitindo a previsão com base nos padrões históricos da própria série.

Para séries que combinam componentes autorregressivos e de média móvel, utiliza-se o modelo ARMA(p, q), cuja formulação incorpora ambos os efeitos:

$$X_t = \phi_1X_{t-1} + \cdots + \phi_pX_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \quad (2.3)$$

No entanto, quando a série apresenta *não estacionariedade* — geralmente evidenciada por tendência persistente ou variância não constante ao longo do tempo —, aplica-se uma diferenciação de ordem apropriada para estabilizá-la. Isso resulta no modelo ARIMA (AutoRegressive Integrated Moving Average), que combina os componentes autorregressivo, de média móvel e de integração (diferenciação), sendo amplamente utilizado em séries com comportamento evolutivo.

2.3.3 Aplicação de Ciência de Dados e modelos preditivos na análise criminal

No campo da violência de gênero, a análise temporal desempenha papel central para compreender a evolução de indicadores, identificar padrões sazonais e avaliar o impacto de eventos críticos, como mudanças legislativas, crises econômicas ou pandemias. Estudos mostram que variações temporais podem revelar tanto tendências estruturais quanto flutuações conjunturais nos registros de violência (MENEGHEL; HIRAKATA, 2011; MOROSKOSKI et al., 2021). Historicamente, grande parte das pesquisas utilizou abordagens descritivas baseadas em médias, proporções e totais Schraiber, D'Oliveira e Couto (2007b), o que, embora relevante, limita a capacidade de capturar dinâmicas complexas e ciclos latentes.

Nesses contextos onde os dados apresentam comportamentos mais complexos, como crescimento exponencial, alta instabilidade ou variabilidade não captada por modelos lineares, que são relações comuns em indicadores criminais, torna-se apropriado considerar métodos alternativos. Análises que incorporam esses padrões como análise de regressão logística, media móvel e modelos de suavização exponencial ou decomposição sazonal, permite identificar mudanças sutis ao longo do tempo e isolar componentes como tendência e sazonalidade (BOX; JENKINS, 1970). Além disso, outros modelos interessantes que incorporam metodologias preditivas e aprendizado de máquina, por exemplo, Random Forest, XGBoost e LSTM, vêm sendo exploradas para antecipar variações e avaliar cenários hipotéticos, especialmente na previsão de crimes e na avaliação de impacto de políticas públicas (PRIY, 2025; PERRY et al., 2013; COELHO et al., 2025).

À vista disso, Fitterer e Nelson (2015) realizaram uma análise bibliográfica de métodos estatísticos aplicados a estudos sobre crimes relacionados ao consumo de álcool, comparando desde modelos clássicos, como médias móveis, regressão linear ARIMA, até modelos lineares generalizados (GLM) e árvores de regressão englobando os algoritmos de aprendizado de máquina. Os autores observam que, no caso dos modelos ARIMA, a necessidade técnica de diferenciação para garantir a estacionariedade dos dados pode remover informações temporais cruciais, como o momento exato do crime, reduzindo o potencial para estratégias preventivas e outras informações importantes para estes tipos de dados. De forma semelhante, apontam limitações nos modelos baseados em árvores de regressão, que, apesar de eficazes para classificação, carecem de estimativas de coeficientes, testes de significância estatística e capacidade de avaliar efeitos simultâneos sobre o crime, além de serem suscetíveis a vieses introduzidos por parâmetros definidos pelo usuário. Para o autor, modelos hierárquicos mistos, como Modelos Lineares Generalizados Mistos (GLMM), Modelos Aditivos Generalizados (GAM) e Processos de Cox Espaciais Variáveis (Spatially Varying Coefficient Processes, SVCP) de base bayesiana, se mostram mais adequados para lidar com a complexidade estrutural de estudos criminais, preservando informações temporais relevantes e oferecendo maior robustez inferencial.

Por sua vez, Travaini et al. (2022) também apresenta uma revisão de diferentes métodos, modelos e conjuntos de dados de machine learning usados para prever reincidência criminal,

desta análise ressaltou que modelos mais simples, como a regressão logística, e modelos mais complexos, como a floresta aleatória, podem apresentar desempenho semelhante. Isso reforça a ideia de que a escolha de um modelo não deve ser pautada exclusivamente pela complexidade, mas pela adequação ao comportamento e à qualidade dos dados disponíveis. Em determinadas situações, modelos simples podem oferecer previsões tão robustas quanto abordagens mais sofisticadas, desde que calibrados de forma apropriada.

Desta forma considerando a importância não apenas da escolha do modelo, mas também da verificação de seu desempenho, a avaliação de modelos de previsão demanda métricas específicas que permitam quantificar o erro e a precisão das estimativas. Para variáveis quantitativas contínuas, como indicadores de violência de gênero, destacam-se: o Erro Médio Absoluto (Mean Absolute Error – MAE), que mede a magnitude média dos erros absolutos; a Raiz do Erro Quadrático Médio (Root Mean Square Error – RMSE), que penaliza de forma mais intensa erros de maior magnitude; e o Erro Percentual Médio Absoluto (Mean Absolute Percentage Error – MAPE), que expressa o erro em termos percentuais, facilitando a interpretação dos resultados (CHAI; DRAXLER, 2014). Além disso, o Coeficiente de Determinação (R^2) pode ser utilizado para indicar a proporção da variância explicada pelo modelo, sendo um indicador complementar de qualidade de ajuste (WILLMOTT; MATSUURA, 2005).

Deste modo, o presente estudo, ao buscar compreender o fenômeno da violência contra a mulher, concentrar-se-á na aplicação de métodos estatísticos clássicos voltados para a análise de séries temporais, com destaque para médias móveis e regressão linear. A modelagem de séries temporais apresenta-se como uma abordagem adequada e estratégica para este contexto, pois permite analisar a evolução dos indicadores ao longo do tempo, identificar tendências e padrões sazonais, além de projetar cenários futuros. Essa perspectiva possibilita compreender o passado, explicar o presente e antecipar possíveis desdobramentos, oferecendo subsídios técnicos e baseados em evidências para apoiar a formulação e a avaliação de políticas públicas voltadas ao enfrentamento da violência de gênero.

2.3.4 Modelos de previsão em Violência contra a mulher

A previsão de indicadores criminais, especialmente no contexto da violência contra a mulher, exige a escolha criteriosa de modelos e métricas capazes de representar a dinâmica dos dados e mensurar a precisão das estimativas. A literatura aponta que a seleção do modelo deve considerar tanto as características estatísticas da série temporal quanto a disponibilidade e a qualidade dos dados (BOX; JENKINS, 1970; TRAVAINI et al., 2022).

Entre os modelos estatísticos clássicos, destacam-se as médias móveis, que suavizam variações de curto prazo e auxiliam na identificação de tendências, e a regressão linear, que relaciona uma variável dependente, como taxas de ocorrência de violência, a variáveis independentes, como tempo, eventos legislativos ou indicadores socioeconômicos. Outro método amplamente utilizado é o ARIMA, adequado para séries temporais estacionárias ou que possam ser transfor-

madadas para estacionariedade, permitindo capturar tanto padrões autorregressivos quanto componentes de média móvel (HYNDMAN; ATHANASOPOULOS, 2018).

A aplicação desses modelos na análise de gênero já foi explorada em diferentes contextos. Kurniawan et al. (2025) utilizaram modelos ARIMA e SARIMA para prever ocorrências de violência contra mulheres e crianças em Jacarta, Indonésia, obtendo previsões capazes de subsidiar estratégias de prevenção e intervenção. De forma semelhante, Lubeya et al. (2021) empregaram o modelo ARIMA para estimar casos de violência sexual e de gênero reportados a um hospital terciário em Lusaka, Zâmbia, demonstrando o potencial desses métodos para apoiar o planejamento de políticas públicas e ações emergenciais.

Lima e Oliveira (2024) analisaram padrões de violência de gênero a partir de registros oficiais, identificando diferentes tipologias de ocorrência e desenvolvendo modelos de previsão para subsidiar políticas públicas direcionadas e estratégias de prevenção mais eficientes. Deiningner (2020), por sua vez, utilizou um modelo preditivo para estimar o risco de uma mulher sofrer violência doméstica no âmbito da atenção básica em saúde, possibilitando intervenções precoces e direcionadas. Em outro estudo, (COELHO et al., 2025) investigaram a evolução temporal e elaboraram projeções da violência contra a mulher no Brasil para o período de 2013 a 2033, permitindo a construção de cenários prospectivos e a avaliação do impacto potencial de políticas e intervenções em nível nacional.

Contudo, observa-se uma lacuna significativa de pesquisas que realizem análises preditivas semelhantes voltadas especificamente para o estado de Pernambuco. A ausência de estudos regionais dessa natureza dificulta a compreensão aprofundada dos padrões locais de violência de gênero e limita o desenvolvimento de políticas públicas baseadas em evidências, adaptadas às particularidades socioeconômicas e culturais do território pernambucano.

2.3.5 Avaliação de Impactos com Variáveis Exógenas

Em estudos de séries temporais, a incorporação de variáveis exógenas representa uma abordagem estatística robusta para avaliar o impacto de eventos externos ou intervenções específicas sobre a trajetória de uma variável de interesse. Esse tipo de modelagem possibilita controlar simultaneamente os aspectos estruturais da série, como tendência, autocorrelação e sazonalidade, ao mesmo tempo em que isola e quantifica o efeito de fatores externos que podem influenciar o comportamento observado ao longo do tempo (ENDERS, 2014).

A operacionalização das variáveis exógenas, na maioria das aplicações práticas, ocorre por meio de indicadores binários (ou variáveis dummy), que assumem valor 0 antes da ocorrência do evento ou intervenção, e valor 1 após sua implementação. Essa técnica permite identificar alterações estruturais no padrão da série associadas a choques exógenos, como mudanças de política pública, crises econômicas ou eventos disruptivos, como desastres naturais ou pandemias.

O uso desse tipo de variável é particularmente útil quando se busca estimar o impacto isolado de uma intervenção, levando em consideração a estrutura dinâmica da série temporal. Ao

integrar variáveis exógenas ao modelo, torna-se possível mensurar com maior precisão a influência de contextos específicos, controlando os efeitos temporais recorrentes e reduzindo o viés de atribuição.

Além disso, o uso de variáveis exógenas favorece a construção de inferências mais próximas do ideal causal, especialmente quando se combinam essas variáveis com modelos estatísticos adequados à natureza dos dados. Essa abordagem amplia significativamente as possibilidades analíticas para a avaliação de políticas públicas, permitindo a tomada de decisões baseada em evidências quantitativas confiáveis.

Neste sentido, diversos estudos demonstram o potencial da incorporação de variáveis exógenas na modelagem de séries temporais. Por exemplo, Zheng et al. (2024) analisaram a influência de indicadores ambientais na previsão da incidência de influenza utilizando modelos ARIMAX, evidenciando como variáveis externas podem aprimorar a acurácia preditiva e embasar políticas públicas de saúde. De forma semelhante, Nepomuceno et al. (2017) investigou a relação entre consumo de álcool e comportamento violento entre torcedores de futebol, ressaltando a importância de considerar variáveis contextuais para compreender padrões criminais. Já ??) examinaram o impacto das restrições impostas pela COVID-19 sobre a violência doméstica na Holanda, concluindo que o confinamento agravou a insegurança de muitas mulheres em suas próprias residências. Esses estudos reforçam que a integração de variáveis exógenas em modelos estatísticos permite captar efeitos de eventos críticos e políticas públicas de forma mais precisa.

Dentre os modelos que permitem a inclusão de variáveis exógenas, destaca-se o modelo **ARIMAX** (AutoRegressive Integrated Moving Average with eXogenous variables), uma extensão do modelo ARIMA tradicional. Essa modelagem incorpora, além dos componentes autorregressivo, de média móvel e de diferenciação, uma ou mais variáveis explicativas externas que ajudam a explicar variações na variável dependente ao longo do tempo.

A formulação geral do modelo ARIMAX pode ser representada da seguinte forma:

$$Y_t = \alpha + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{k=1}^r \beta_k X_{k,t} + \varepsilon_t \quad (2.4)$$

Onde:

- Y_t é o valor da série temporal no tempo t ;
- α é o intercepto;
- ϕ_i são os coeficientes autorregressivos (AR);
- θ_j são os coeficientes de média móvel (MA);
- ε_t representa o erro aleatório no tempo t ;
- $X_{k,t}$ representa a k -ésima variável exógena no tempo t ;

- β_k são os coeficientes associados às variáveis exógenas.

A vantagem central do ARIMAX é permitir que mudanças estruturais e eventos exógenos, como a pandemia da COVID-19 ou a implementação de políticas públicas, sejam modelados diretamente como fontes de variação na série, sem comprometer a estrutura temporal subjacente.

Para que o modelo ARIMAX seja aplicado de forma adequada, é necessário que a série temporal principal esteja estacionária, ou seja, se necessário estabilizada por meio de diferenciação, e que as variáveis exógenas estejam corretamente codificadas e alinhadas temporalmente. Em geral, variáveis binárias (dummies) são utilizadas para representar a presença ou ausência de eventos específicos, assumindo valor 0 antes e 1 após o início da intervenção.

Neste estudo, o ARIMAX é adotado como ferramenta para estimar o impacto isolado de eventos como a pandemia e o programa “Juntos pela Segurança” sobre os registros mensais de violência doméstica. A partir da comparação entre o comportamento previsto da série (na ausência do evento) e os valores observados após a intervenção, torna-se possível avaliar a efetividade da ação governamental com base em evidência empírica rigorosa.

Neste estudo, tal estratégia é empregada para investigar o impacto de políticas públicas e eventos excepcionais sobre a evolução da violência doméstica em Pernambuco. Especificamente, propõe-se avaliar o efeito da pandemia da COVID-19 e do programa estadual "Juntos pela Segurança", lançado em 2023, por meio da inclusão de variáveis dummy nos modelos autorregressivos. Com isso, busca-se avaliar a efetividade dessas ações governamentais a partir de evidências empíricas, contribuindo para o aprimoramento de estratégias de enfrentamento à violência de gênero e para o fortalecimento das práticas de monitoramento com base em dados.

2.3.6 Análise de Correlação entre Séries Temporais

A investigação de correlação entre duas séries temporais visa identificar possíveis associações estatísticas entre seus comportamentos ao longo do tempo. Em contextos aplicados, é comum que as séries não apresentem distribuição normal ou contenham *outliers*, o que exige atenção na escolha do método de correlação. Para lidar com essas limitações, empregam-se abordagens complementares baseadas tanto em métodos paramétricos quanto não paramétricos.

Esses padrões metodológicos são amplamente utilizados na literatura. Por exemplo, Katern-dahl et al. (2010) aplicaram métodos de análise de correlação para compreender as dinâmicas complexas da violência entre parceiros íntimos. De forma semelhante, Nyende (2024) realizou uma análise conceitual dos correlatos da violência doméstica e do comportamento de risco em adolescentes, destacando a relevância de compreender relações entre múltiplos indicadores ao longo do tempo. Assim, a análise de correlação configura-se como uma metodologia relevante, cabendo ao pesquisador selecionar o método mais adequado às características das séries estudadas.

O coeficiente de correlação de Pearson é amplamente utilizado para medir a intensidade e a direção da relação linear entre duas variáveis. Contudo, sua aplicação pressupõe normalidade dos dados, homocedasticidade e ausência de *outliers* extremos. Quando essas condições são violadas, o coeficiente de Spearman, baseado em postos, é uma alternativa robusta e não paramétrica, capaz de capturar relações monotônicas mesmo na presença de distorções na distribuição.

A análise foi realizada considerando as séries em sua forma original e após diferenciação de primeira ordem, a fim de eliminar tendências e investigar correlações de curto prazo. Ambos os coeficientes, Pearson e Spearman, foram calculados ao nível de significância usual (5%).

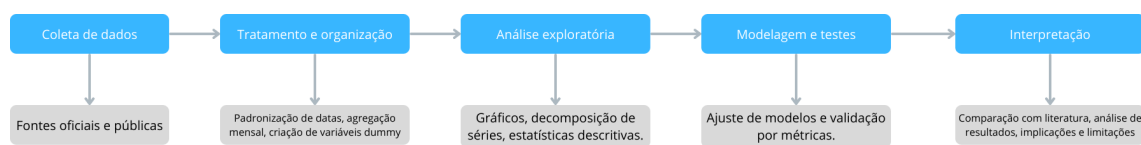
Essa abordagem metodológica, que contempla diferentes técnicas de correlação adaptadas às propriedades dos dados, está em conformidade com as boas práticas na análise de séries temporais, conforme descrito por autores como Ljung e Box (1978), Box e Jenkins (1970), Shumway e Stoffer (2017).

No contexto do presente trabalho, espera-se que a correlação entre feminicídio e VDFCM revele padrões relevantes para a formulação de políticas públicas e para a compreensão das interações entre diferentes formas de violência de gênero, permitindo identificar se variações em um indicador tendem a ser acompanhadas por mudanças no outro.

3 METODOLOGIA

Este estudo utiliza uma abordagem quantitativa e descritiva, com foco em identificar tendências temporais e explorar possíveis correlações entre os registros de violência doméstica e feminicídios no estado de Pernambuco. Dessa forma, a análise segue as etapas apresentadas no seguinte fluxograma:

Figura 1 – Etapas do estudo - Fluxograma



Fonte: o autor (2025).

3.1 DESCRIÇÃO E COLETA DOS DADOS

Os dados trabalhados nesta pesquisa foram obtidos a partir de bases públicas de registros de violência doméstica e feminicídios no estado de Pernambuco. As fontes de dados utilizadas incluem os indicadores de mortes violentas intencionais (MVI) e indicadores de Violência doméstica e familiar contra a mulher (VDFM). Estes dados são coletados diariamente no IML, Polícia Científica, Coordenação de Plantão da Polícia Civil, Serviço de Inteligência da PMPE e CIODS, sendo coletados e analisados pela Gerência Geral de Análise Criminal e Estatística (GGACE) e divulgados pela Secretaria de Defesa Social de Pernambuco (SDS-PE). Deste modo, todos os dados utilizados são de acesso público e não contêm informações que permitam a identificação dos indivíduos envolvidos.

Os indicadores de feminicídios foram extraídos a partir da filtragem dos dados de MVI, conforme definido na Portaria Nº 1066, de 11 de março de 2023, SDS-PE. Essa portaria classifica os feminicídios como uma das naturezas jurídicas dentro do conjunto de MVI, ao lado de outras categorias, como homicídio doloso, latrocínio, lesão corporal seguida de morte, outros crimes resultantes em mortes e mortes decorrentes de intervenção de agentes do Estado. Como mencionado anteriormente, o feminicídio foi tipificado como crime em 2015, por meio da Lei nº 13.104/15, sendo caracterizado como o assassinato de mulheres motivado por questões de gênero. Apesar de o banco de dados de MVI iniciar sua série histórica em 2004, a análise realizada neste trabalho tem como ponto de partida o ano de 2017, pois foi a partir desse ano que a classificação do feminicídio como natureza jurídica passou a ser incorporada aos dados da SDS-PE.

Entre 2017 e 2025, foram registrados 585 casos de feminicídio. Os dados disponíveis sobre esses crimes contemplam variáveis que permitem análises temporais, geográficas e demográficas, descritas a seguir:

- **Data** — variável *temporal*, representa o registro exato do feminicídio (dia, mês e ano), permitindo o mapeamento cronológico de ocorrências e a construção de séries temporais mensais ou sazonais.
- **Ano** — também uma variável *temporal*, extraída ou agregada a partir da data, permite análises mais amplas, como tendências anuais, comparação entre períodos antes e depois de políticas públicas ou eventos específicos.
- **Município** — variável *geográfica*, indica o local da ocorrência do crime, possibilitando a análise espacial da distribuição dos casos, identificação de padrões de concentração urbana ou interiorana, e o cruzamento com indicadores locais como IDH, população e infraestrutura de proteção à mulher.
- **Região Geográfica** — variável *geográfica categorizada*, classifica o município em mesorregiões ou macrorregiões (ex.: Metropolitana, Agreste, Sertão), permitindo comparações regionais agregadas e identificação de dinâmicas espaciais diferenciadas no estado.
- **Idade da vítima** — variável *demográfica*, que permite análises do perfil etário das mulheres assassinadas, identificação de faixas etárias mais vulneráveis e correlação com ciclos de vida, dependência econômica, escolaridade ou fatores culturais.

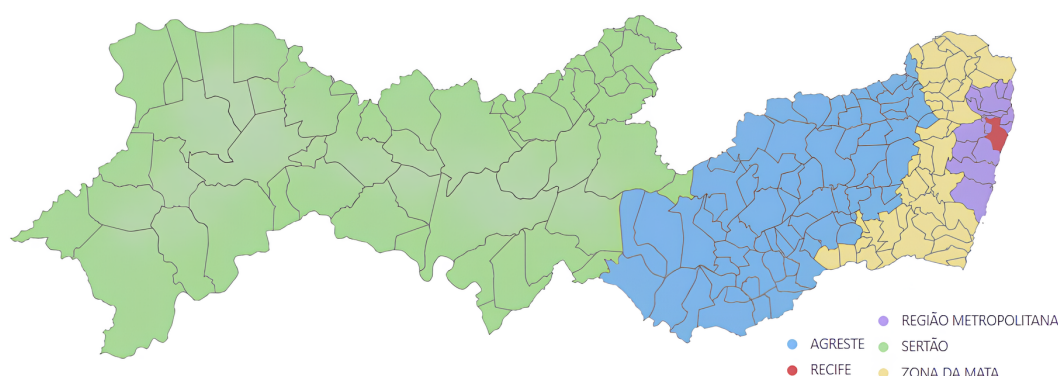
Além disso, os dados de Violência Doméstica Fatal Contra Mulheres (VDFCM) começaram a ser registrados em 2015. No entanto, para manter a coerência metodológica com a série histórica dos feminicídios, a análise desse indicador também terá início em 2017. No período de 2017 a 2025, foram registrados 373.658 casos de VDFCM. As informações disponíveis para esses registros incluem variáveis temporais, geográficas e demográficas, conforme descritas a seguir:

- **Data** — variável *temporal*, refere-se à data exata do evento fatal, permitindo a organização cronológica dos casos e a análise de padrões sazonais e variações mensais.
- **Ano** — variável *temporal*, agregada a partir da data, utilizada para comparações entre anos, identificação de tendências temporais e análises de políticas públicas implementadas em determinados períodos.
- **Município** — variável *geográfica*, que indica o local da ocorrência do crime. Possibilita o mapeamento territorial dos casos e a identificação de padrões espaciais por localidade, além de permitir análises intermunicipais.
- **Região Geográfica** — variável *geográfica categorizada*, classifica os municípios segundo as mesorregiões do estado, facilitando a comparação regional e a análise de disparidades entre territórios.

- **Total de envolvidos** — variável quantitativa que representa o número total de vítimas fatais em cada ocorrência. Embora o foco seja nas vítimas do sexo feminino, essa variável permite dimensionar a gravidade do evento e a presença de múltiplas vítimas em um mesmo episódio.
- **Faixa etária das vítimas** — variável *demográfica*, categorizada em sete grupos etários:
 - 1 00–11 anos (crianças);
 - 2 12–17 anos (adolescentes);
 - 3 18–24 anos (jovens adultas);
 - 4 25–29 anos (adultas jovens);
 - 5 30–34 anos (adultas em fase produtiva inicial);
 - 6 35–64 anos (adultas em fase madura);
 - 7 65 anos ou mais (idosas).

As regiões geográficas são definidas Agreste, Zona Da Mata, Região Metropolitana, Recife (Capital) e Sertão como dispostos na figura abaixo:

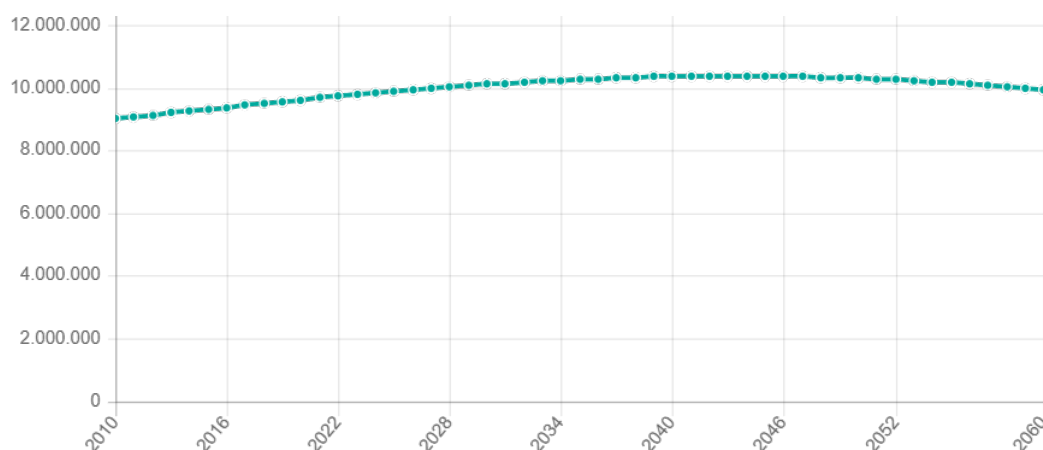
Figura 2 – Regiões Geográficas utilizadas pela SDS - PE



Fonte: o autor (2025).

Além disso, também foram incorporados dados geográficos das cidades, obtidos a partir dos Censos Demográficos de 2010 e 2022, realizados pelo Instituto Brasileiro de Geografia e Estatística (IBGE), bem como os cálculos de projeção populacional fornecidos pelo IBGE para os anos em que não houve contagem oficial. Esses dados permitem uma contextualização espacial dos registros de feminicídio e violência doméstica, possibilitando investigações mais aprofundadas sobre a distribuição geográfica desses crimes no estado de Pernambuco. Além disso, possibilitam o cálculo das taxas de criminalidade a cada 100 mil habitantes, um indicador amplamente utilizado na análise da incidência de crimes em relação à população.

Figura 3 – Evolução populacional e projeções em Pernambuco



Fonte: Instituto Brasileiro de Geografia e Estatística (IBGE) (2025).

3.2 ANÁLISE EXPLORATÓRIA DE DADOS

3.2.1 Evolução dos Casos de Femicídio em Pernambuco

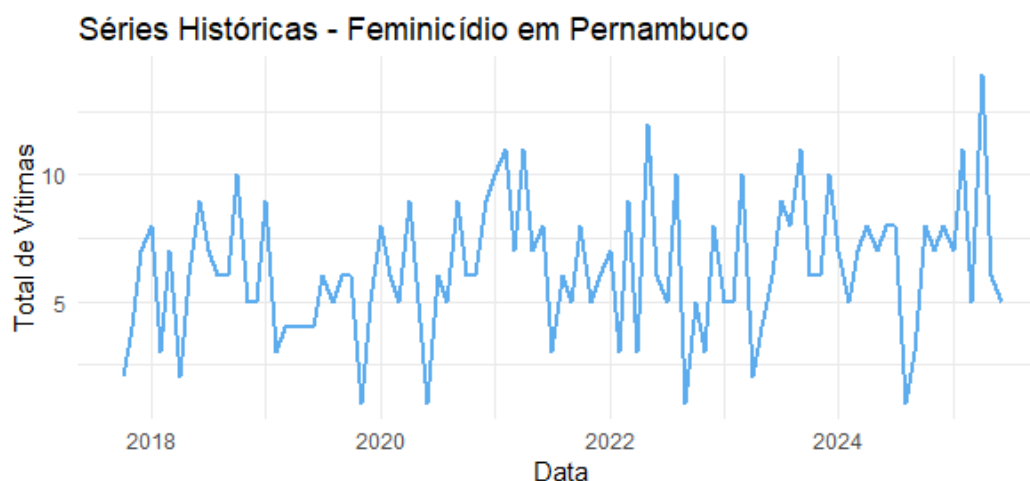
A fim de compreender a evolução dos casos de feminicídio em Pernambuco ao longo do tempo, foi realizada uma análise exploratória dos dados disponíveis. Todo o processamento e as análises foram realizados utilizando o software estatístico R, devido à sua robustez e ampla aplicabilidade em estudos de séries temporais e análise exploratória de dados. Inicialmente, os registros foram pré-processados para garantir a qualidade das informações, incluindo a conversão das datas para um formato adequado e a transformação das variáveis numéricas. Em seguida, os dados foram agregados por mês, possibilitando a visualização da variação no total de vítimas ao longo do período analisado.

A base de dados tratada corresponde a uma amostra de 93 observações mensais, compreendendo o período de outubro de 2017 a junho de 2025. Os valores mensais variaram de 1 a 14 casos, com média de 6,29 e mediana de 6, o que indica uma distribuição aproximadamente simétrica, com leve assimetria à direita. A análise dos quartis revelou que 50% dos dados estão concentrados entre 5 e 8 ocorrências mensais. O desvio padrão de 2,64 reforça essa variação relativamente contida. Além disso, a presença de valores máximos relativamente distantes da média sugere a possibilidade de existência de outliers na série.

A Figura 9 apresenta um gráfico de linhas que ilustra a evolução mensal dos casos de feminicídio em Pernambuco, entre outubro de 2017 e junho de 2025. Observa-se, inicialmente, a ausência de uma tendência clara e persistente, ou seja, não há um movimento nitidamente crescente ou decrescente ao longo do tempo. A série oscila em torno de uma média relativamente constante, apresentando flutuações cíclicas e variações sazonais regulares ao longo dos anos. Em 2020, ano marcado pelo início da pandemia de COVID-19, nota-se uma redução expressiva nos registros, possivelmente associada às medidas de confinamento e restrição de circulação. A partir de 2021, observa-se um aumento gradual nos casos, com picos significativos entre 2021 e

2022, o que pode indicar um retorno ou até intensificação da violência após o relaxamento das restrições. Essas oscilações reforçam a necessidade de uma análise melhor da decomposição da série.

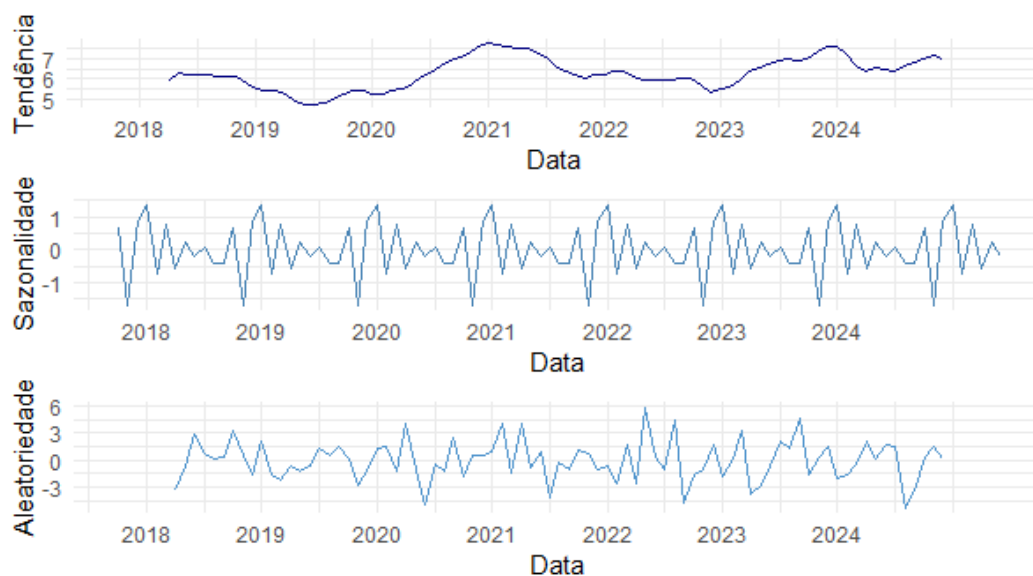
Figura 4 – Série mensal de Feminicídio em Pernambuco - 2017 à 2025



Fonte: o autor (2025).

A Figura 5 apresenta a decomposição aditiva da série temporal mensal. No primeiro painel, observa-se a componente de tendência, que revela uma variação moderada ao longo do tempo. Entre os anos de 2018 e 2020, a tendência apresenta leve declínio, com um ponto de inflexão por volta de 2020, possivelmente associado ao início da pandemia da COVID-19. A partir de 2021, verifica-se uma retomada gradual da trajetória ascendente, com oscilações suaves até o final do período analisado, sugerindo um crescimento progressivo no número de feminicídios após a flexibilização das medidas sanitárias. A segunda componente evidencia a sazonalidade da série, caracterizada por flutuações periódicas e sistemáticas ao longo do ano. Os padrões sazonais são relativamente estáveis, com picos e vales recorrentes que se repetem de forma consistente entre os anos, indicando que determinados meses tendem a registrar sistematicamente maior ou menor número de casos, comportamento que pode refletir fatores sociais ou institucionais que se manifestam ciclicamente. Especificamente, observa-se maior incidência de casos nos meses de final e início de ano, o que pode estar relacionado a festividades, reuniões familiares e fatores emocionais que intensificam conflitos domésticos. Por fim, a terceira componente representa os resíduos aleatórios. Esses resíduos estão razoavelmente centrados em torno de zero, com dispersão homogênea ao longo do tempo, indica que os principais movimentos da série foram adequadamente capturados pelos componentes anteriores.

Figura 5 – Decomposição da Série de Feminicídio em Pernambuco

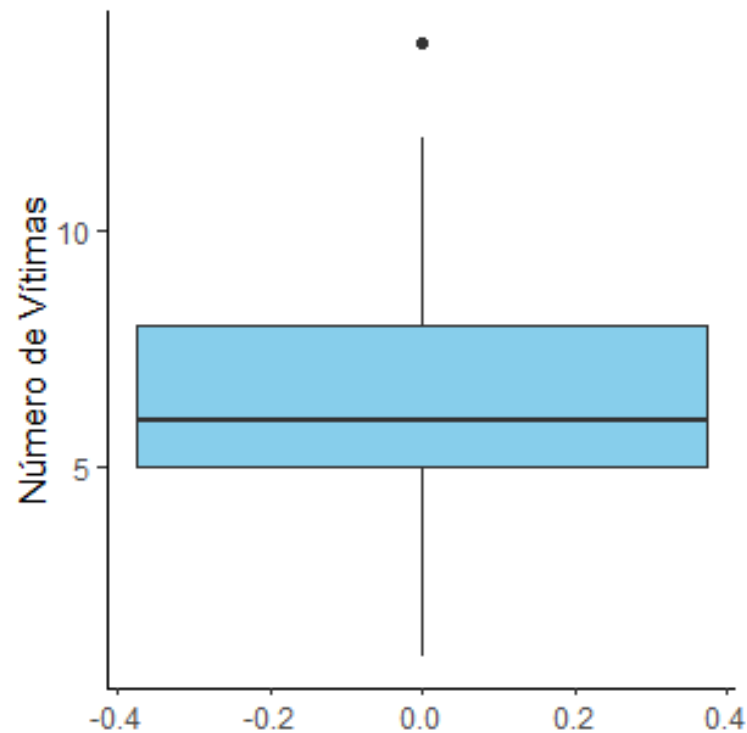


Fonte: o autor (2025).

Embora a análise gráfica da decomposição da série temporal aponte para a existência de uma possível tendência de longo prazo, sugerindo uma violação do pressuposto de estacionariedade, essa evidência visual não é conclusiva. Para investigar formalmente a presença de raiz unitária, foi aplicado o teste de Dickey-Fuller aumentado (ADF). O valor da estatística de teste foi de -4,0452, com p-valor igual a 0,0106, considerando quatro defasagens. Como o p-valor é inferior ao nível de significância de 5%, rejeita-se a hipótese nula de não estacionariedade. Portanto, conclui-se que a série é estacionária em nível, isto é, não apresenta tendência determinística persistente, e suas propriedades estatísticas, como média e variância constantes ao longo do tempo.

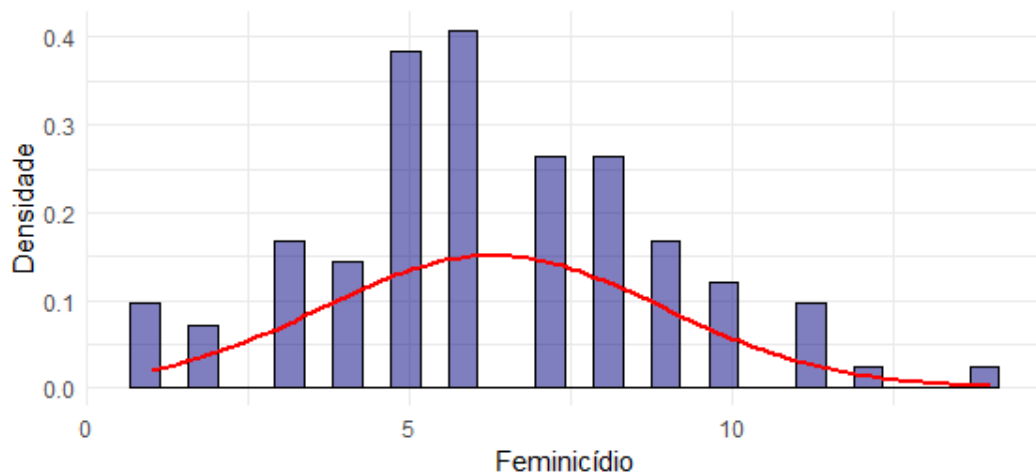
Ao se analisar o gráfico de boxplot figura 6 distribuição do número mensal de vítimas de feminicídio mantém-se relativamente estável, com mediana próxima de 6 e dispersão concentrada entre os quartis. Observa-se, contudo, a presença de um outlier no extremo superior da distribuição, registrado no mês de abril de 2025. Esse ponto representa um aumento atípico e abrupto no número de casos, destoando do padrão histórico da série. Até o momento, não há justificativa clara ou evento específico identificado que explique esse crescimento pontual. A ocorrência desse valor extremo deve ser mantida, uma vez que pode influenciar análises de tendência e previsões, ainda que não represente uma mudança estrutural persistente na série.

Figura 6 – Diagrama de caixa para Feminicídio em Pernambuco



Fonte: o autor (2025).

Figura 7 – Histograma e curva de densidade de Feminicídio



Fonte: o autor (2025).

Logo em seguida, para avaliar a normalidade da série, foi traçado o histograma, no qual se observou uma distribuição aproximadamente normal, com leve assimetria à direita, evidenciada pelo fato de a média ser ligeiramente superior à mediana. A maior concentração de observações

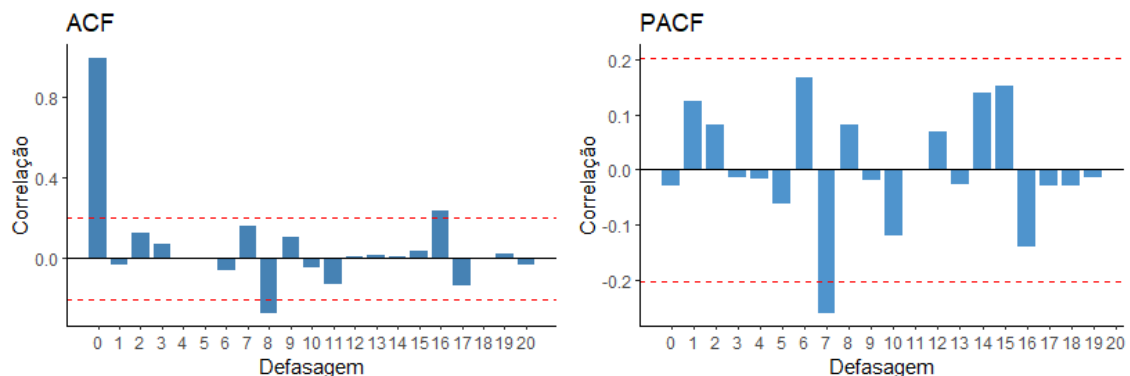
situa-se entre 4 e 7 casos mensais, o que indica uma distribuição centralizada em torno da média da série.

Tabela 1 – Testes de normalidade para a série de feminicídios mensais

Teste	Estatística	<i>p</i> -valor	Interpretação
Kolmogorov-Smirnov	$D = 0,1137$	0,1806	Distribuição normal
Shapiro-Wilk	$W = 0,9789$	0,1375	Dados seguem normalidade

Para confirmar essa impressão visual, foram aplicados os testes de Kolmogorov-Smirnov e Shapiro-Wilk, cujos resultados estão apresentados na Tabela 1. O teste de Kolmogorov-Smirnov, com *p*-valor superior a 0,05, não rejeitou a hipótese nula de que os dados seguem uma distribuição normal. De forma consistente, o teste de Shapiro-Wilk, mais apropriado para amostras pequenas e médias, também não indicou violação da normalidade. Esses resultados, em consonância com as evidências visuais do histograma e do boxplot, sustentam a suposição de normalidade da variável. Dessa forma, justifica-se o uso de procedimentos estatísticos paramétricos nas análises subsequentes.

Figura 8 – Autocorrelação e autocorrelação parcial Feminicídio



Fonte: o autor (2025).

Observando os gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF) da série de feminicídios, conforme apresentado na Figura 8, verifica-se, na ACF, uma autocorrelação significativa no primeiro lag, o que indica forte dependência de curto prazo. A partir do segundo lag, as autocorrelações decaem rapidamente e permanecem dentro dos limites de significância, sugerindo um comportamento de memória curta na série.

No gráfico de Autocorrelação Parcial (PACF), observa-se um pico significativo em apenas algumas defasagens específicas, com destaque para o lag 8 negativo, o que pode indicar a presença de componentes autorregressivos de ordens mais elevadas. Esse padrão sugere uma relação linear predominante entre os valores atuais e alguns valores defasados no tempo, especialmente os mais recentes. Contudo, não se identifica um corte abrupto claro na PACF, o que dificulta a determinação precisa da ordem do modelo apenas por inspeção visual. Essas

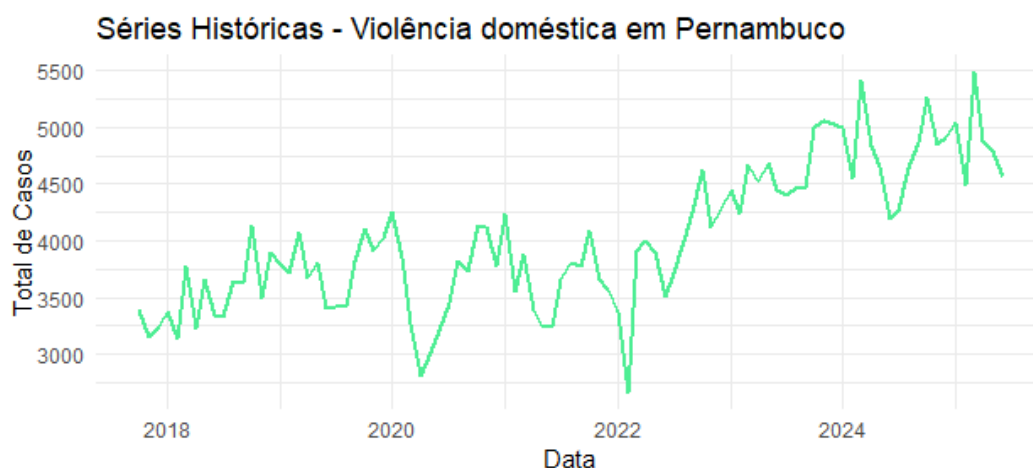
características indicam que a série pode conter elementos autorregressivos e de média móvel combinados.

3.2.2 Evolução dos Casos de Violência Doméstica (VDFCM)

De forma semelhante à análise realizada para a série de feminicídios, foi conduzida uma análise exploratória da série de VDFCM, utilizando o software R, com o objetivo de aprimorar sua compreensão. Os registros foram previamente processados para assegurar a qualidade das informações, incluindo a padronização das datas para um formato apropriado, transformação das variáveis numéricas e correções e padronização de textos. Posteriormente, os dados foram agregados por mês, resultando em uma amostra composta por 93 variáveis separados mensalmente de outubro de 2017 a junho de 2025.

Os casos de VDFCM apresentam altas incidências mensais, diferentemente da série de feminicídios, pois são contabilizados por tipo de violência, e uma única vítima pode sofrer mais de um tipo de agressão, resultando em valores elevados de casos registrados. Ao analisar esses dados, observa-se uma média e mediana bastante próximas, o que sugere, a princípio, uma série aproximadamente simétrica. Os valores mínimos mensais foram de 2.657 ocorrências, enquanto os máximos chegaram a 5.492, sugerindo alta variabilidade mensal. A partir da análise dos quartis, verifica-se que 50% dos meses apresentaram entre 3.022 e 3.992 violências, valor que se distancia dos máximos observados, o que pode indicar a presença de outliers na série.

Figura 9 – Série mensal de Violência Doméstica familiar contra a mulher em Pernambuco



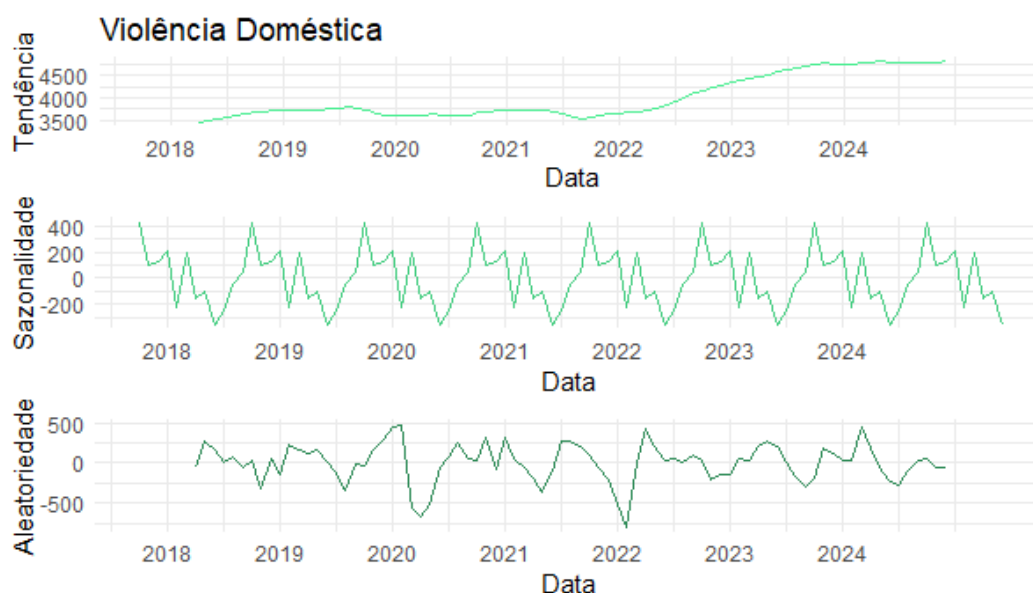
Fonte: o autor (2025).

Ao analisar a série histórica, observa-se uma alta volatilidade, com quedas acentuadas em determinados períodos. No entanto, o comportamento geral da série revela uma tendência de crescimento ao longo do tempo. Esse padrão se torna ainda mais evidente após uma queda brusca em 2020, possivelmente relacionada ao início das restrições de circulação decorrentes da pandemia da COVID-19, seguida por oscilações anômalas em 2022. Posteriormente, a série retoma sua trajetória ascendente. Alguns desses comportamentos podem estar associados aos

impactos da crise sanitária, como a leve estabilidade observada entre 2019 e 2020, seguida por um aumento significativo a partir de 2021, ano em que se intensifica a tendência de crescimento nos registros de violência doméstica. Esses elementos indicam que o contexto da pandemia pode ter contribuído para alterar o padrão dos dados durante o período analisado.

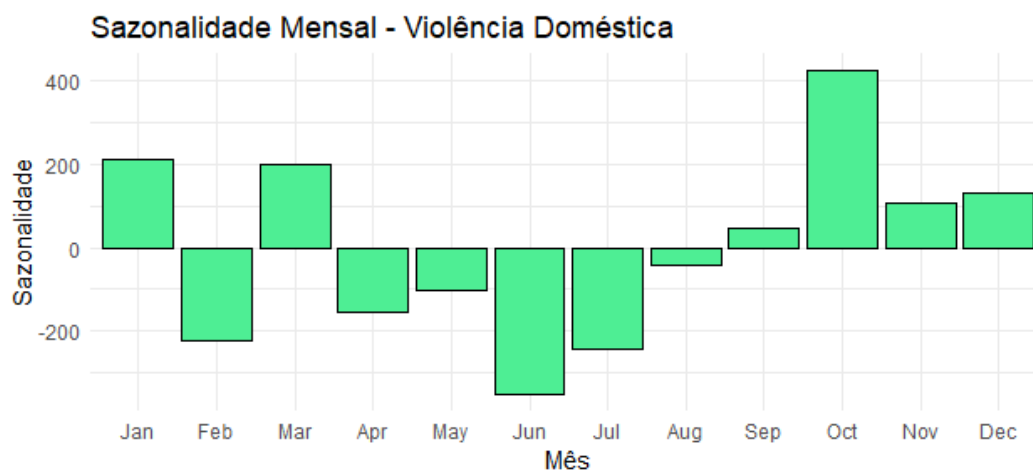
Além disso, ao observar a série histórica, já é possível identificar padrões de sazonalidade periódica, os quais podem ser analisados com maior precisão no gráfico de decomposição apresentado na Figura 10. No entanto, nota-se a ocorrência de flutuações bruscas em alguns pontos, indicando desvios significativos. Esses desvios podem estar associados a eventos extraordinários, como mudanças em políticas públicas e campanhas de incentivo a denúncias ou padrões comportamentais. A periodicidade da sazonalidade que pode ser melhor observada no gráfico 11 nota-se que há meses do ano com aumento previsível de violência doméstica, de outubro a janeiro e meses com valores negativos acentuados como Junho e Julho que representa um menor número de casos. Ao analisar o componente de resíduo, é possível identificar flutuações mais expressivas nos anos de 2020 e 2022. Em 2020, tais oscilações podem estar associadas aos efeitos diretos da pandemia de COVID-19, enquanto em 2022 observa-se um comportamento atípico, ainda sem causa clara. De forma geral, a variabilidade dos resíduos aparenta ser razoavelmente homogênea, embora em determinados momentos ocorram desvios que indicam a presença de fatores inesperados, que fogem à dinâmica padrão da série.

Figura 10 – Decomposição da Série de Violência Doméstica familiar contra a mulher em Pernambuco



Fonte: o autor (2025).

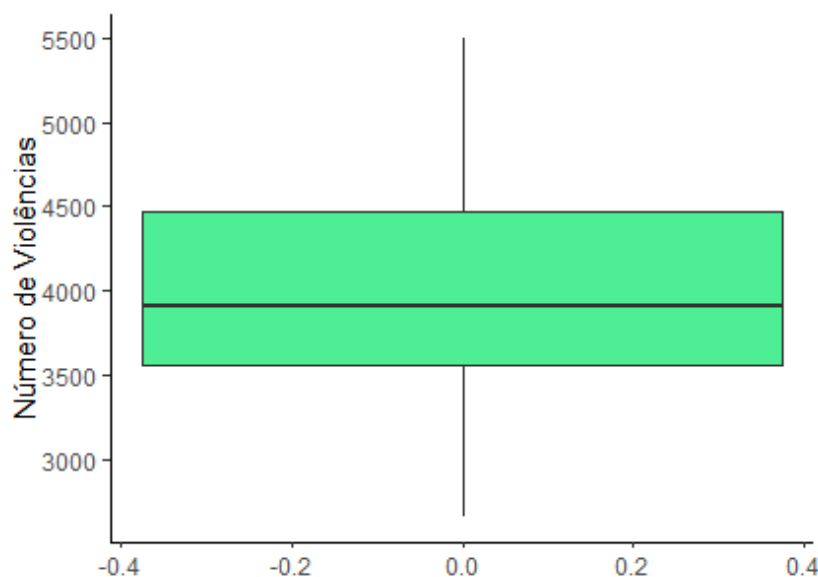
Figura 11 – Sazonalidade da Série de Violência Doméstica familiar contra a mulher em Pernambuco



Fonte: o autor (2025).

A análise gráfica indica a presença de tendência e sazonalidade na série, sugerindo que os pressupostos de estacionaridade não são atendidos. Para verificar formalmente, foi aplicado o teste de Dickey-Fuller aumentado (ADF). Como o p-valor obtido foi superior a 0,05, não rejeitamos a hipótese nula de presença de raiz unitária, indicando que a série é não estacionária.

Figura 12 – Diagrama de caixa para VDFCM em Pernambuco

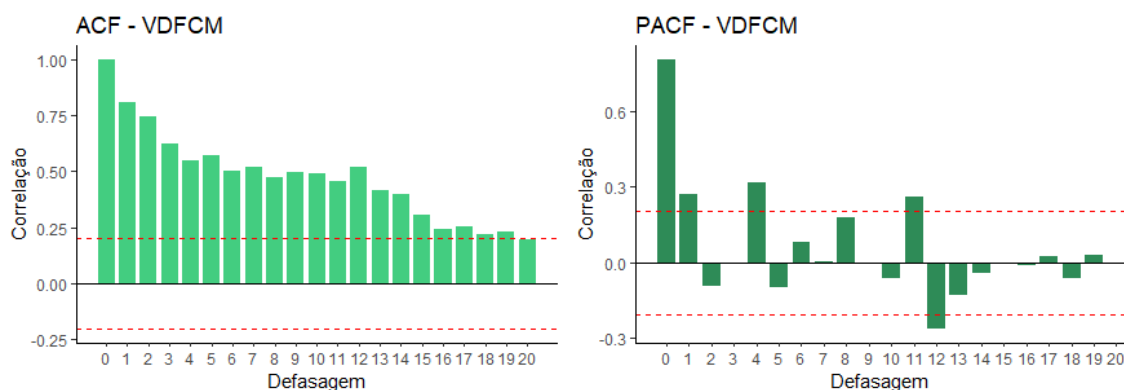


Fonte: o autor (2025).

Analisando o gráfico de caixa (Figura 12), observa-se que a distribuição do número mensal de violências domésticas aparenta ser ligeiramente assimétrica à direita, uma vez que a mediana está deslocada para a parte inferior da caixa. A maior parte dos dados se concentra entre aproximadamente 3550 e 4450 casos mensais, conforme indicam os quartis. Apesar de os valores

máximos ultrapassarem os 5000 registros em alguns meses, não foram identificados outliers severos, o que sugere que a dispersão dos dados, embora significativa, está dentro de uma faixa estatisticamente esperada.

Figura 13 – Autocorrelação VDFCM Pernambuco



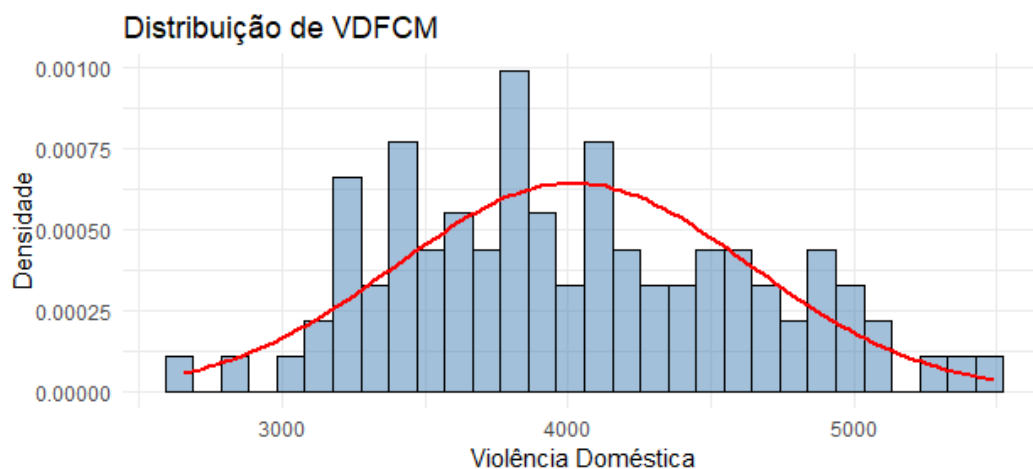
Fonte: o autor (2025).

A análise do gráfico de Autocorrelação (ACF) na Figura 13 mostra que a série VDFCM apresenta autocorrelações positivas e significativas ao longo de várias defasagens, com um decaimento lento e persistente, o que indica a presença de não estacionaridade, confirmação que também foi obtida por meio do teste de Dickey-Fuller. O gráfico de Autocorrelação Parcial (PACF) revela um corte acentuado no primeiro lag, seguido de correlações próximas de zero nas demais defasagens. Esse padrão sugere uma estrutura autorregressiva de primeira ordem (AR(1)), em que os valores atuais estão fortemente relacionados com o mês anterior, mas apresentam pouca ou nenhuma dependência linear com defasagens maiores após esse ajuste, que sugere pouca evidência de componente MA puro.

Adicionalmente, a decomposição da série temporal evidenciou uma sazonalidade anual bem definida, tanto pela observação de oscilações regulares ao longo dos anos no gráfico sazonal quanto pela presença de um valor significativo no lag 12 da PACF, ultrapassando os limites de significância. Isso sugere que há dependência sazonal, isto é, uma relação entre valores defasados de 12 períodos (meses), típica de uma estrutura SARIMA.

Diante desses achados, conclui-se que a modelagem mais apropriada para essa série não é um modelo ARIMA simples, mas sim um modelo SARIMA (Seasonal ARIMA), que acomoda tanto a tendência (com $d = 1$) quanto a sazonalidade anual (com $D = 1$ e $m = 12$). Uma estrutura inicial plausível para ser estimada é SARIMA($p,1,q$)($P,1,Q$)[12], onde os parâmetros p , q , P e Q devem ser ajustados com base nos gráficos ACF/PACF da série diferenciada e em critérios de seleção como AIC e BIC.

Figura 14 – Histograma Violência Doméstica familiar contra a mulher em Pernambuco - 2017 á 2025



Fonte: o autor (2025).

Observando o histograma na Figura 14, nota-se que a série de violência doméstica apresenta uma distribuição quase normal, com leve assimetria à direita e alguns valores altos esporádicos. A curva de densidade mostra um pico entre 3.200 e 3.800 ocorrências, sem cauda excessivamente pesada, e com dispersão tanto para valores mais baixos quanto para mais altos. Como já evidenciado no boxplot e na análise de autocorrelação (ACF), esses valores extremos não são tão frequentes nem suficientemente influentes para causar distorções significativas na distribuição.

Na análise dos testes de normalidade, o teste de Kolmogorov-Smirnov apresentou p-valor muito superior a 0,05, não rejeitando a hipótese nula de normalidade. Ou seja, segundo esse teste, a série pode ser considerada normalmente distribuída. De forma consistente, o teste de Shapiro-Wilk também não rejeitou a hipótese nula, indicando que a distribuição da série também pode ser considerada aproximadamente normal. Esse resultado é coerente com o histograma apresentado anteriormente, que mostra uma distribuição próxima da simetria. Dessa forma, a aplicação de testes paramétricos é considerada válida.

3.3 MÉTODOS ESTATÍSTICOS E MODELOS APLICADOS

3.3.1 Análise de Correlação entre Femicídio e VDFCM

A investigação da associação entre os comportamentos mensais das séries de violência doméstica e feminicídio foi conduzida por meio do coeficiente de correlação de Pearson, com o objetivo de verificar a existência de uma relação estatisticamente significativa entre essas variáveis ao longo do tempo. Tratou-se de uma aplicação direta do coeficiente, considerando as observações de ambas as séries no mesmo período, mês a mês, de modo a comparar o comportamento conjunto dos casos, isto é, se tendem a aumentar ou diminuir simultaneamente.

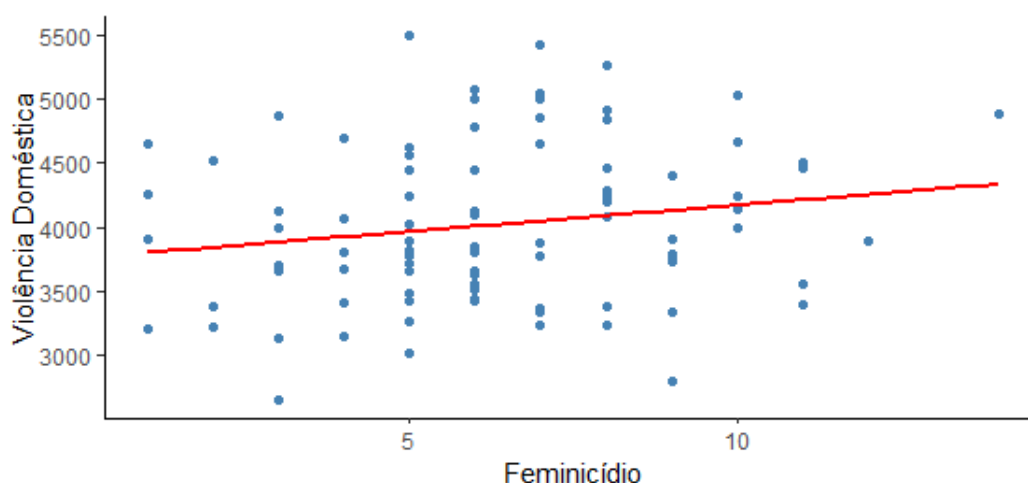
Inicialmente, buscou-se avaliar se os pressupostos para a aplicação de métodos paramétricos estavam atendidos.

Foram aplicados testes para verificar a normalidade dos dados (Shapiro-Wilk), a ausência de *outliers* extremos (teste de Bonferroni) e a homocedasticidade (teste de Levene). Todos os critérios foram satisfeitos:

- As séries apresentaram distribuição aproximadamente normal;
- Nenhum *outlier* extremo foi identificado;
- A homogeneidade de variância foi confirmada.

Diante disso, optou-se pela aplicação do teste de correlação de Pearson, adequado para dados paramétricos. A análise foi realizada sem transformação ou diferenciação das séries, com o intuito de identificar uma possível relação estrutural de longo prazo entre os fenômenos.

Figura 15 – Correlação de Pearson VDFCM e feminicídio em pernambuco



Fonte: o autor (2025).

3.3.2 Análise de Impacto da Pandemia de COVID-19

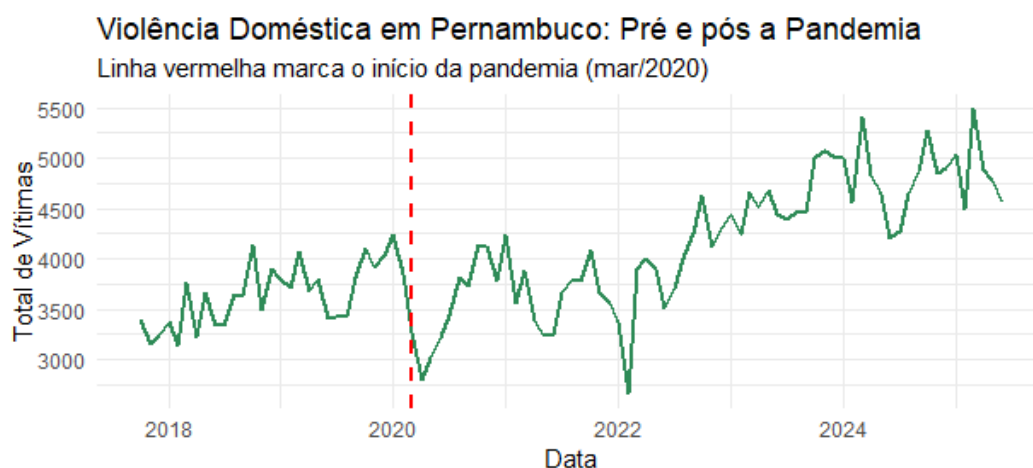
Como a pandemia teve início em março de 2020, este será adotado como ponto de corte para a análise. Assim, a série temporal será dividida em dois períodos: pré-pandemia, que abrange os dados até fevereiro de 2020, e pós-pandemia, com os dados a partir de março de 2020. A Figura 16 ilustra graficamente essa separação.

A análise do impacto da pandemia será conduzida em duas etapas complementares. A primeira consiste na comparação estatística das médias e desvios padrão entre os períodos pré e pós-pandemia, com o objetivo de avaliar se houve mudanças significativas na distribuição dos valores ao longo do tempo. A segunda etapa envolve a modelagem preditiva, por meio do ajuste

de um modelo autorregressivo integrado com média móvel (ARIMA), incorporando uma variável exógena indicadora da pandemia. Esse modelo será treinado com os dados do período pré-pandemia e, em seguida, utilizado para prever o comportamento esperado da série no período pós-pandemia. As previsões obtidas serão comparadas com os valores reais observados, de forma a inferir se a pandemia provocou impacto estatisticamente significativo no comportamento da série temporal.

As previsões obtidas a partir do modelo ARIMAX serão então comparadas com os valores reais observados durante o período pandêmico. Para isso, será aplicado um teste t de amostras pareadas, a fim de verificar se há diferença estatisticamente significativa entre os valores previstos (na ausência da pandemia) e os valores efetivamente registrados. Essa abordagem permite avaliar de forma mais robusta se o impacto da pandemia na série temporal foi estatisticamente relevante, controlando os efeitos de tendência, sazonalidade e autocorrelação.

Figura 16 – Violência Doméstica familiar contra a mulher em Pernambuco - Análise pandêmica



Fonte: o autor (2025).

Para definir o modelo mais adequado para a análise de médias entre os períodos pré e pós-pandemia, foram aplicados testes de pressupostos estatísticos. Inicialmente, utilizou-se o teste de Shapiro-Wilk para verificar a normalidade da distribuição dos dados. Os resultados indicaram que os dados seguem uma distribuição aproximadamente normal.

Teste	Estatística	p-valor	Interpretação
Levene	$F = 14,502$	0,00025	Variâncias diferentes entre grupos ($p < 0,001$)
Shapiro-Wilk	$W = 0,98143$	0,2087	Dados seguem distribuição normal ($p > 0,05$)

Tabela 2 – Resultados dos testes de Levene e Shapiro-Wilk: COVID-19

Em seguida, para avaliar a homogeneidade das variâncias entre os grupos, foi aplicado o teste de Levene, cujos resultados estão apresentados na Tabela 2. Como o valor de p encontrado foi inferior a 0,05, rejeita-se a hipótese nula de igualdade das variâncias, indicando a presença de heterocedasticidade entre os grupos analisados. Dessa forma, mesmo com a suposição de

normalidade atendida, a diferença entre as variâncias justifica o uso do teste t para médias com variâncias desiguais (Welch's t-test). Esse teste é apropriado para situações em que a homogeneidade de variância não é verificada, mantendo a robustez da inferência estatística.

3.3.3 Avaliação das Políticas Públicas Estaduais

Com o objetivo de avaliar o impacto da política pública “*Juntos pela Segurança*”, foi conduzida uma análise empírica fundamentada em duas abordagens complementares: comparação entre médias dos períodos pré e pós-implementação da política, e modelagem econométrica com inclusão de variável exógena, por meio de um modelo ARIMAX, considerando também possíveis defasagens temporais (*lags*) capazes de capturar efeitos postergados da intervenção governamental.

Após o ajuste dos modelos ARIMAX com as diferentes defasagens, será aplicado o teste t para amostras pareadas entre os valores previstos e os valores reais observados no período de vigência da política. Esse teste permitirá verificar se a política produziu um impacto estatisticamente significativo no comportamento da série temporal de violência doméstica, complementando a análise estatística das médias e controlando para tendências, sazonalidade e autocorrelação serial da série.

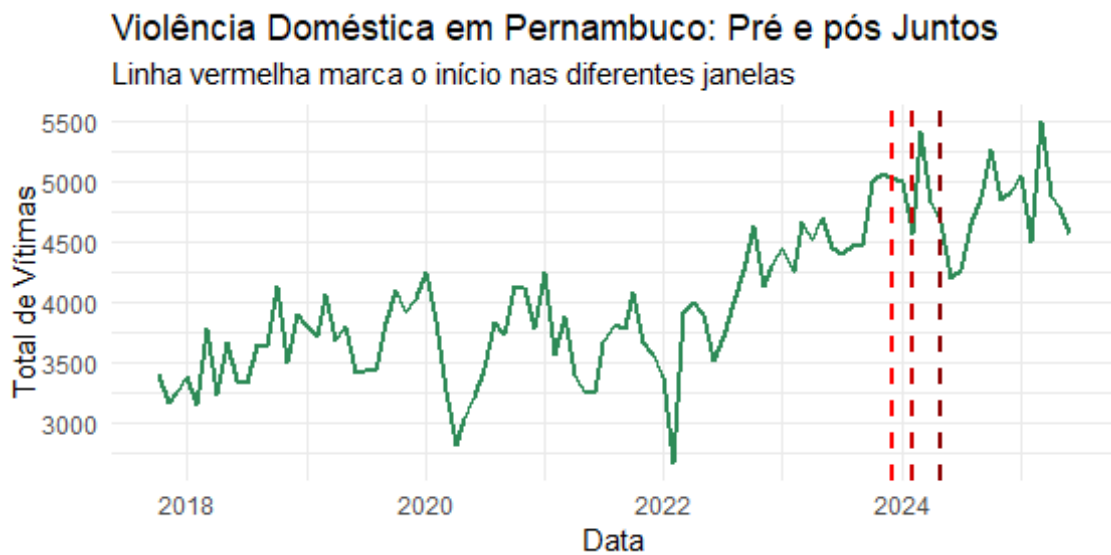
A política pública foi oficialmente lançada em 27 de novembro de 2023 (Governo de Pernambuco – Secretaria de Defesa Social, 2023), sendo o mês seguinte o marco temporal adotado como ponto de corte para as análises de impacto. Assim, a série temporal foi segmentada em dois períodos: *pré-política* e *pós-política*, conforme representado graficamente na Figura 17.

Para captar diferentes possibilidades de resposta da série ao efeito da política, a análise considerou três janelas de tempo distintas para a ativação da variável dummy, a saber:

- **Sem defasagem** — impacto imediato, com início em dezembro de 2023;
- **Defasagem de 3 meses** — início do impacto a partir de fevereiro de 2024;
- **Defasagem de 6 meses** — impacto observado a partir de maio de 2024.

Essa segmentação permite avaliar não apenas os efeitos diretos e imediatos da política, mas também possíveis impactos graduais ou diferidos, comuns em políticas públicas que envolvem ações de médio prazo e dependem de articulações interinstitucionais para surtirem efeito.

Figura 17 – Série Temporal de VDFCM com Indicação da Introdução de Política Pública



Embora se trate da mesma série já analisada anteriormente no contexto da pandemia, a presente investigação considera cortes temporais distintos, relacionados à política pública “*Juntos pela Segurança*”. Diante disso, optou-se por realizar novamente os testes de pressupostos estatísticos para comparação de médias entre os grupos definidos (pré e pós-política), com o objetivo de verificar se os diferentes recortes temporais apresentam comportamentos estatísticos semelhantes.

A Tabela 3 apresenta os resultados dos testes de normalidade (Shapiro-Wilk) e de homogeneidade de variâncias (Levene) para os três cenários analisados: impacto imediato (sem defasagem), com defasagem de três meses e com defasagem de seis meses.

Tabela 3 – Resultados dos testes de Levene e Shapiro-Wilk: Política Publica

Defasagem (Lag)	Teste	Estatística	p-valor
Sem defasagem	Levene (variâncias)	$F = 3,078$	0,0827
	Shapiro-Wilk (normalidade)	$W = 0,9889$	0,7731
3 meses	Levene (variâncias)	$F = 2,8778$	0,0932
	Shapiro-Wilk (normalidade)	$W = 0,9803$	0,2862
6 meses	Levene (variâncias)	$F = 3,1081$	0,0813
	Shapiro-Wilk (normalidade)	$W = 0,9799$	0,2481

Os testes de normalidade não rejeitam a hipótese nula em nenhum dos casos ($p > 0,05$), sugerindo que as amostras seguem distribuição aproximadamente normal. Já os testes de Levene apresentam valores de p entre 0,08 e 0,09, indicando possível heterocedasticidade, embora sem violação conclusiva. Diante desses resultados, considerando o tamanho amostral moderado, optou-se pela utilização do teste t de Welch, que é robusto à heterocedasticidade e não

exige normalidade estrita. Assim, as análises de comparação de médias entre os períodos pré e pós-intervenção puderam ser conduzidas com base em pressupostos estatísticos adequadamente atendidos.

3.3.4 Modelos de previsão

Segundo Box e Jenkins (1970), às séries temporais podem ser modeladas de forma eficaz por meio de processos autorregressivos e de médias móveis, desde que observados os pressupostos de estacionariedade e estrutura de dependência serial. Dessa forma, neste trabalho, optou-se pela modelagem de séries temporais, justamente pela estrutura cronológica dos dados mensais de registros de vítimas entre outubro de 2017 e junho de 2025.

A análise de séries temporais inicia-se com uma etapa exploratória, baseada em estatísticas descritivas, como média, mediana, quartis e medidas de dispersão, seguida da avaliação gráfica da série visualização da evolução ao longo do tempo e da verificação de características fundamentais como:

- **Tendência:** variações sistemáticas de longo prazo, indicando crescimento ou declínio na série;
- **Sazonalidade:** padrões cíclicos recorrentes em determinados períodos, como meses ou trimestres;
- **Estacionariedade:** estabilidade estatística da série ao longo do tempo, com média, variância e covariância constantes;
- **Normalidade (parametricidade):** conformidade da distribuição dos dados com a distribuição normal, relevante para a escolha de técnicas paramétricas.

A verificação da estacionariedade será realizada por meio do teste de Dickey-Fuller aumentado (ADF). Para avaliar a normalidade, serão utilizados os testes de Shapiro-Wilk e Kolmogorov-Smirnov, além de inspeções visuais por histogramas e boxplots. Caso a série não seja estacionária, aplicar-se-ão transformações como diferenciação de ordem 1 ou sazonal. Caso não apresente normalidade, considerar-se-ão métodos não paramétricos.

A escolha do modelo de previsão está diretamente ligada a essas características estruturais. Séries estacionárias com padrões lineares tendem a ser bem ajustadas por modelos da família ARMA. Séries não estacionárias requerem transformações para viabilizar modelos como ARIMA ou suas variações sazonais (SARIMA). Após a identificação e estimação do modelo mais adequado, com base nos padrões observados nos gráficos ACF e PACF, bem como em critérios como AIC ou BIC, realiza-se a etapa de previsão, utilizando a projeção do modelo ajustado para estimar os valores futuros da série, geralmente para os meses subsequentes, no caso de séries mensais. Essa projeção fornece não apenas os valores esperados, mas também os intervalos de confiança, permitindo uma análise mais robusta e confiável do comportamento futuro da variável em estudo.

4 RESULTADOS

4.1 CONSIDERAÇÕES SOBRE AS CORRELAÇÕES ESTUDADAS

A investigação da associação entre os comportamentos mensais das séries de violência doméstica e feminicídio foi realizada por meio dos coeficientes de correlação de Pearson. O objetivo foi verificar se há uma relação estatisticamente significativa entre essas variáveis ao longo do tempo, considerando tanto possíveis associações lineares quanto monotônicas.

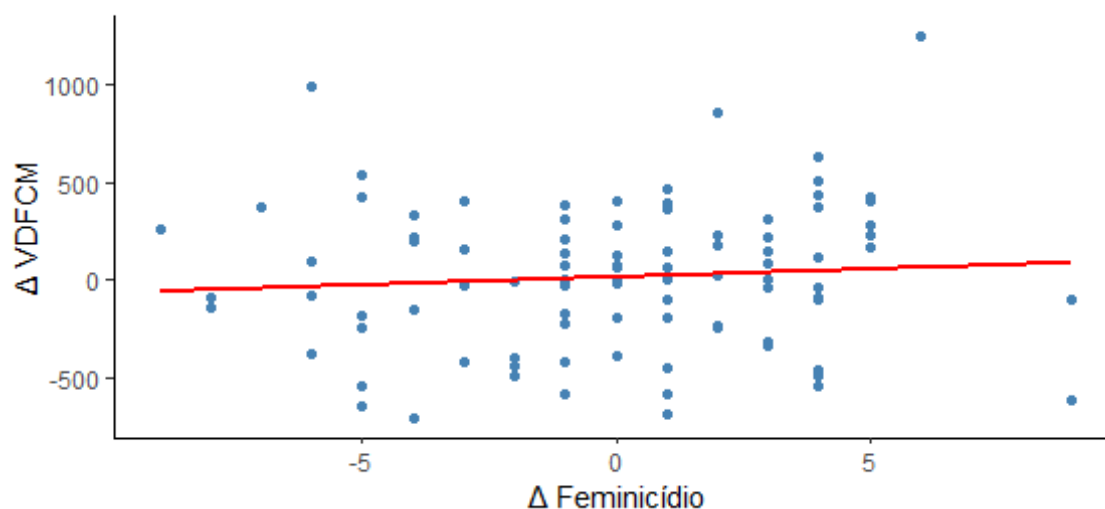
Inicialmente, calculou-se o coeficiente de correlação de Pearson entre o número mensal de casos de feminicídio e os registros mensais de violência doméstica. O valor obtido foi de $r = 0,175$, indicando uma correlação positiva fraca. No entanto, o p-valor associado foi de 0,093, o que não permite rejeitar a hipótese nula ao nível de significância de 5%. Isso implica na ausência de evidências estatísticas robustas para afirmar a existência de uma correlação linear entre as variáveis ao longo do período analisado. A análise visual do gráfico de dispersão (Figura 15) reforça esse achado, demonstrando uma distribuição de pontos bastante dispersa, sem padrão visível de associação.

Esse resultado sugere que, apesar da relação temática entre as variáveis, seus comportamentos temporais não evoluem de maneira sincronizada ou proporcional em termos estatísticos. É possível que diferentes fatores contextuais estejam influenciando suas dinâmicas, como tendências temporais específicas em cada série. Para investigar possíveis associações de curtíssimo prazo entre as variações mensais, procedeu-se à diferenciação de primeira ordem dos dados, visando eliminar tendências e tornar as séries estacionárias.

As análises subsequentes consideraram as séries diferenciadas, nas quais se buscou identificar relações pontuais nas variações mensais. Antes da aplicação do teste de correlação de Pearson nessa nova configuração, foram realizados os testes de normalidade de Shapiro-Wilk e de homocedasticidade de Levene. Em ambos os casos, os resultados não rejeitaram as hipóteses nulas ($p > 0,05$), confirmando a adequação do uso de métodos paramétricos. Adicionalmente, o teste de Bonferroni para detecção de outliers não identificou valores extremos significativos.

A correlação de Pearson calculada entre as séries diferenciadas resultou em um coeficiente de $r = 0,0815$, com um p-valor de 0,4401. Esse resultado indica novamente uma correlação positiva muito fraca e estatisticamente não significativa, conforme apresentado na Figura 18. O intervalo de confiança de 95% para o coeficiente de correlação, compreendido entre -0,1254 e 0,2816, inclui o valor nulo, reforçando a ausência de evidência para uma associação linear entre as séries analisadas.

Figura 18 – Gráfico de Dispersão das series diferenciadas



Fonte: o autor (2025).

Os resultados obtidos indicam que não há correlação estatisticamente significativa entre as séries de feminicídio e de violência doméstica e familiar contra a mulher (VDFCM), tanto nos dados originais quanto nas séries diferenciadas. Embora exista uma relação temática evidente entre os dois fenômenos, suas dinâmicas mensais não demonstram comportamento interdependente. Essa ausência de correlação sugere que o feminicídio, por se tratar de um desfecho extremo, pontual e menos frequente, pode seguir uma trajetória estatística autônoma, influenciada por variáveis específicas que não se manifestam nas flutuações agregadas dos registros de VDFCM em curto prazo.

Além disso, fatores como a subnotificação de casos, a defasagem temporal entre episódios de violência recorrente e a ocorrência do feminicídio, bem como a limitação dos registros administrativos em captar a progressão dos ciclos de violência, podem contribuir para a não detecção estatística dessa relação.

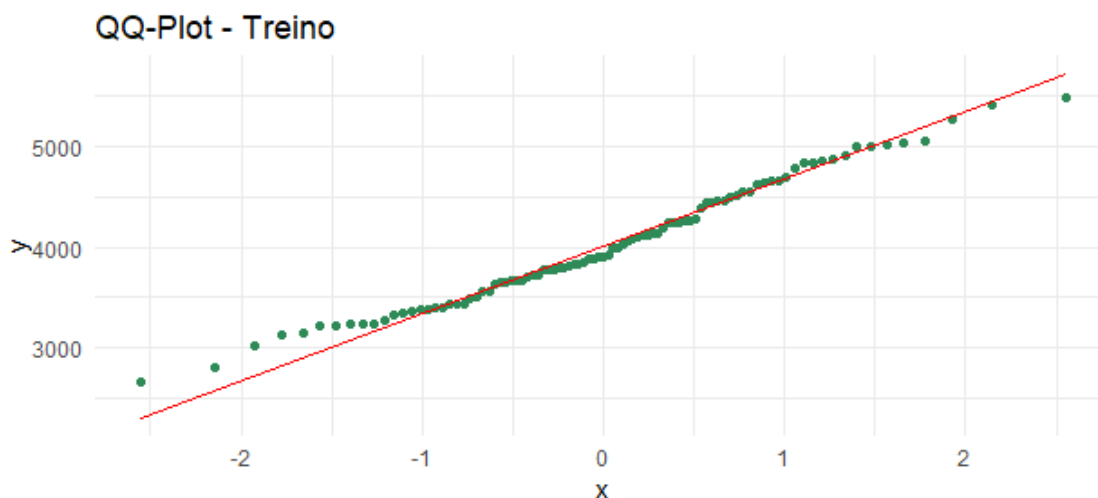
Dessa forma, os achados reforçam a compreensão de que, apesar da conexão conceitual entre VDFCM e feminicídio, suas séries temporais apresentam comportamentos distintos. Assim, não foi possível, neste estudo, estabelecer uma relação preditiva entre elas. Com base nisso, optou-se por analisar a série de VDFCM de forma independente nas seções subsequentes, considerando suas características próprias. Tal distinção é fundamental para futuras análises preditivas, bem como para o planejamento e a avaliação de políticas públicas eficazes no enfrentamento da violência de gênero.

4.2 INTERPRETAÇÃO DOS MODELOS E PREVISÕES

Para a modelagem e análise preditiva da série de VDFCM, os dados foram divididos em conjuntos de treino e teste, utilizando-se uma proporção de 80% para treino e 20% para teste.

Com isso, foram destinadas 73 observações para o conjunto de treino e 20 para o conjunto de teste.

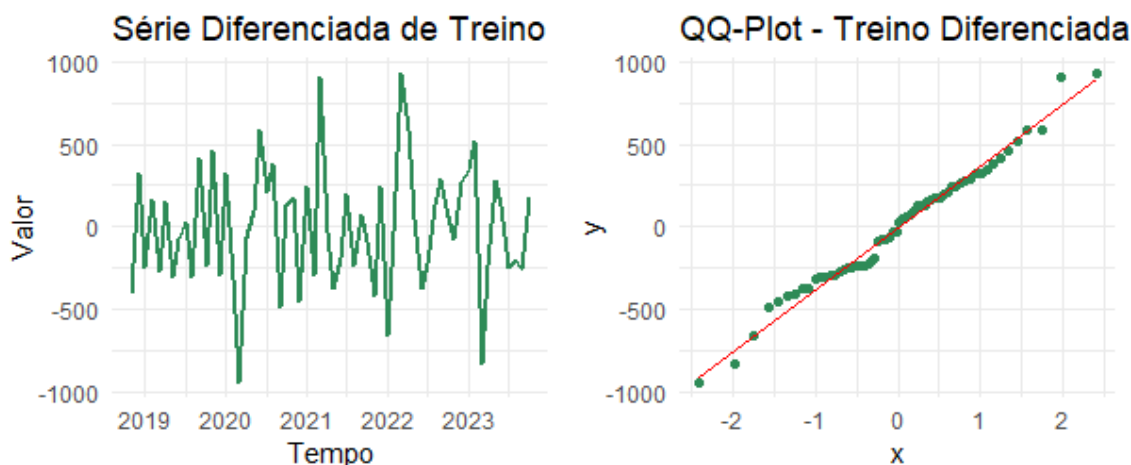
Figura 19 – Q-QPlot de recorte Treino da Série de VDFCM



Fonte: o autor (2025).

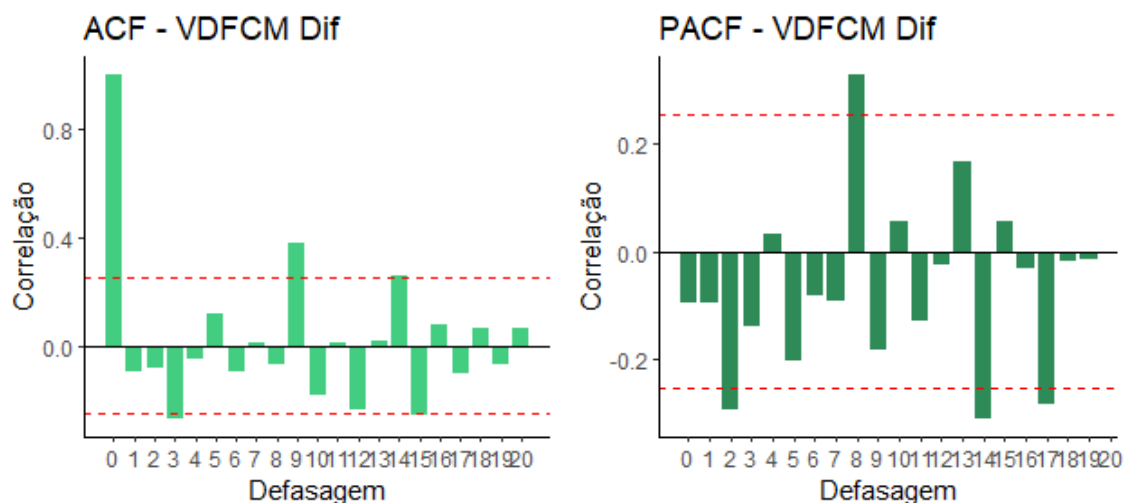
Como já observado nas análises anteriores, a distribuição da série se aproxima de uma normal. A análise do QQplot apresentado na figura 19 confirma essa suposição, revelando que a maior parte dos pontos segue a linha de referência, embora haja leve divergência nas extremidades, coerente com a assimetria moderada já identificada. Apesar disso, a série demonstra a presença de tendência e sazonalidade, o que compromete a estacionariedade necessária para modelagem preditiva eficiente. Dessa forma, a série foi diferenciada com defasagem de 12 para remover a sazonalidade anual. Em seguida, foi aplicada uma segunda diferenciação, desta vez com defasagem de 1, visando eliminar a tendência linear residual. A Figura 20 exibe o gráfico da série histórica após essas transformações, evidenciando um comportamento mais estacionário e compatível com os pressupostos dos modelos autorregressivos.

Figura 20 – Recorte de Treino da Série de VDFCM: Histograma e QQ-plot



Fonte: o autor (2025).

Figura 21 – Recorte de Treino da da Série de VDFCM: ACF e PACF



Fonte: o autor (2025).

Na Figura 20, o novo QQplot da série diferenciada demonstra uma aproximação ainda maior à distribuição normal, o que reforça a adequação da série aos testes e modelagens subsequentes. Por fim, os gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF), apresentados nas Figuras 21, indicam padrões compatíveis com a presença de componentes autorregressivos e de média móvel, características essenciais para a identificação da estrutura ideal do modelo, como nos modelos do tipo SARIMA.

Ao analisar o gráfico de ACF, observa-se um corte claro no lag 1, sugerindo a presença de um componente de média móvel de ordem 1, ou seja, MA(1). Já o gráfico de PACF apresenta um comportamento semelhante ao de ruído branco, já que apesar de alguns lags, como 2, 8, 14 e 17, ultrapassarem levemente os limites de significância, esses picos ocorrem de forma isolada e sem um padrão decrescente claro. Assim com autocorrelações insignificantes após o primeiro lag, indicando a possível ausência de forte componente autorregressivo, o que permite considerar AR(1) ou até mesmo AR(0), optando-se pelo mais parcimonioso, a depender dos critérios de seleção (AIC).

Em relação à componente sazonal, os gráficos não apresentaram picos significativos nas defasagens múltiplas de 12, o que indicaria sazonalidade mensal. Como a série já foi diferenciada sazonalmente, e não há evidência clara de autocorrelação sazonal residual, pode-se assumir que os parâmetros sazonais P e Q são iguais a 0. No entanto, por precaução e devido à possível presença de pequenas oscilações sazonais residuais, é válido considerar modelos alternativos com P ou Q igual a 1, e compará-los via métricas de ajuste.

Assim, o modelo SARIMA mais adequado deve ser testado a partir de configurações como SARIMA(0 ou 1, 1, 1)(0 ou 1, 1, 0 ou 1)[12], com base nas evidências empíricas dos gráficos e testes realizados. Além da análise gráfica dos componentes de autocorrelação, também foi utilizada a função `auto.arima()` no software R, que sugeriu o modelo SARIMA(0,1,1)(0,0,1)[12] como configuração inicial. No entanto, com base nas combinações previamente definidas e nos

testes realizados, foram ajustados diversos modelos SARIMA, cujos resultados de desempenho estão apresentados na Tabela 4.

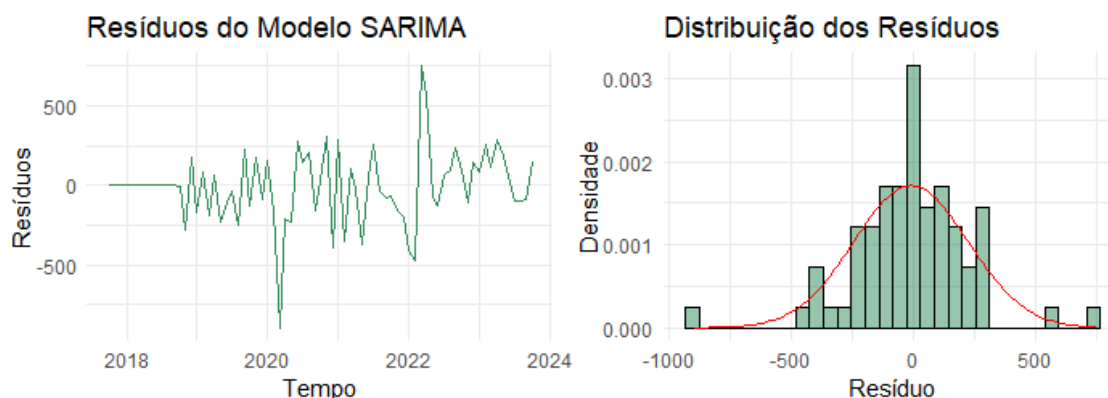
Variável	Modelo	AIC	BIC	σ^2	Comentário
a2	(1,1,1)(0,1,1)[12]	865.6	873.98	68787	Melhor geral (menor AIC, σ^2 baixo)
a1	(0,1,1)(0,1,1)[12]	868.16	874.44	73637	Boa opção, mas a2 é superior
a7	(0,1,1)(1,1,1)[12]	868.46	876.84	77285	Similar ao a1
a10	(1,1,1)(1,1,1)[12]	870.69	881.16	78935	Mais complexo, sem muito ganho
a5	(0,1,1)(0,1,0)[12]	883.27	887.46	137744	σ^2 alto, pior ajuste
a6	(0,1,1)(1,1,0)[12]	881.41	887.70	129585	σ^2 ainda alto
a8	(1,1,1)(0,1,0)[12]	885.56	891.84	140806	σ^2 muito alto
a9	(1,1,1)(1,1,0)[12]	883.62	892.00	132298	σ^2 muito alto
a4	(0,1,1)(0,0,1)[12]	1031.57	1038.4	84990	Também ruim

Tabela 4 – Resultados dos modelos SARIMA testados

Após a comparação dos critérios de seleção, como o Akaike Information Criterion (AIC) e a análise dos resíduos, identificou-se que o modelo com melhor desempenho foi o SARIMA(1,1,1)(0,1,1)[12]. Esse modelo apresentou os menores valores de AIC e BIC, além de resíduos bem comportados, o que o torna o mais adequado para representar a série analisada. Assim, este será o modelo adotado para a etapa de previsão, apresentada na seção a seguir. (AKAIKE, 1974)

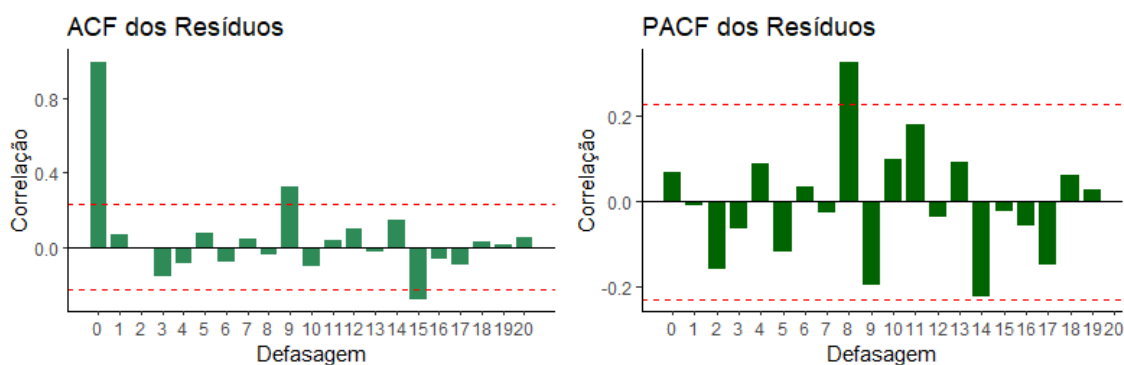
A avaliação dos resíduos do modelo SARIMA, conforme apresentado nas Figuras 22 e 23, evidenciou que, embora os resíduos apresentem distribuição aproximadamente simétrica e centrada em torno de zero, observou-se autocorrelação significativa na defasagem 1 e picos adicionais em lags posteriores. Tais padrões indicam que o modelo não conseguiu capturar integralmente a estrutura temporal da série, especialmente no que se refere à dependência serial de curto prazo. O gráfico temporal dos resíduos também revelou oscilações sistemáticas, sugerindo a presença de ruídos correlacionados ao longo do tempo. Apesar de o modelo oferecer uma representação satisfatória da tendência geral da série, sua performance na modelagem dos componentes de ruído branco se mostrou limitada. Nesse contexto, modelos mais robustos, como variações do SARIMA com múltiplas variáveis ou estruturas de média móveis adicionais, podem ser explorados para aprimorar o ajuste.

Figura 22 – Análise do resíduo do modelo SARIMA - Gráficos de linha e histograma



Fonte: o autor (2025).

Figura 23 – Análise do resíduo do modelo SARIMA - Gráficos ACF e PACF

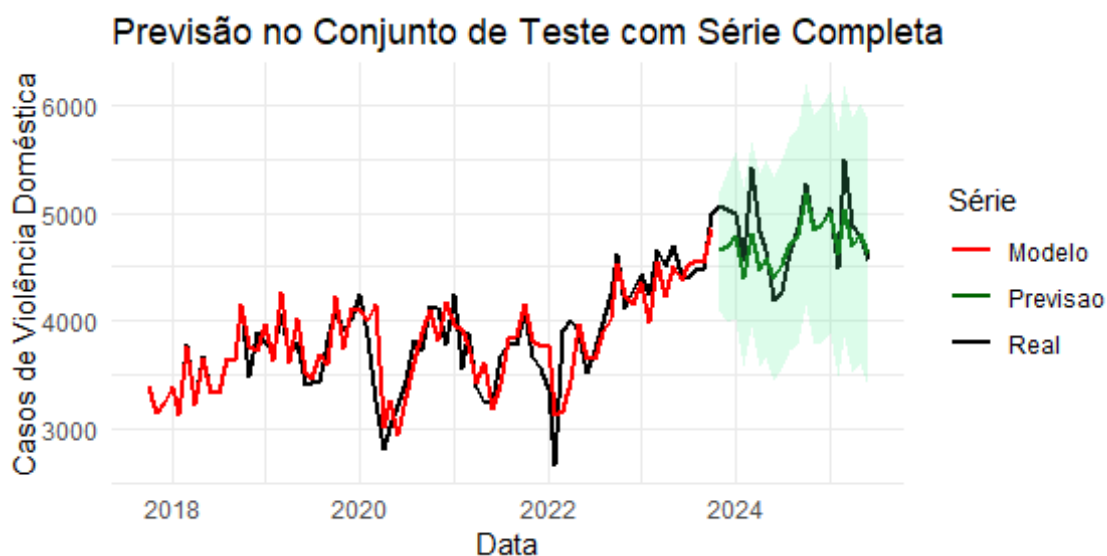


Fonte: o autor (2025).

Com o objetivo de realizar uma avaliação formal da qualidade dos resíduos, foram aplicados os testes de Ljung-Box e Shapiro-Wilk. O teste de Ljung-Box, utilizando 12 defasagens, resultou em um p-valor de 0,2371, não rejeitando a hipótese nula de ausência de autocorrelação ($p > 0,05$). Esse resultado reforça a conclusão de que os resíduos podem ser considerados aproximadamente independentes, corroborando a suposição de ruído branco.

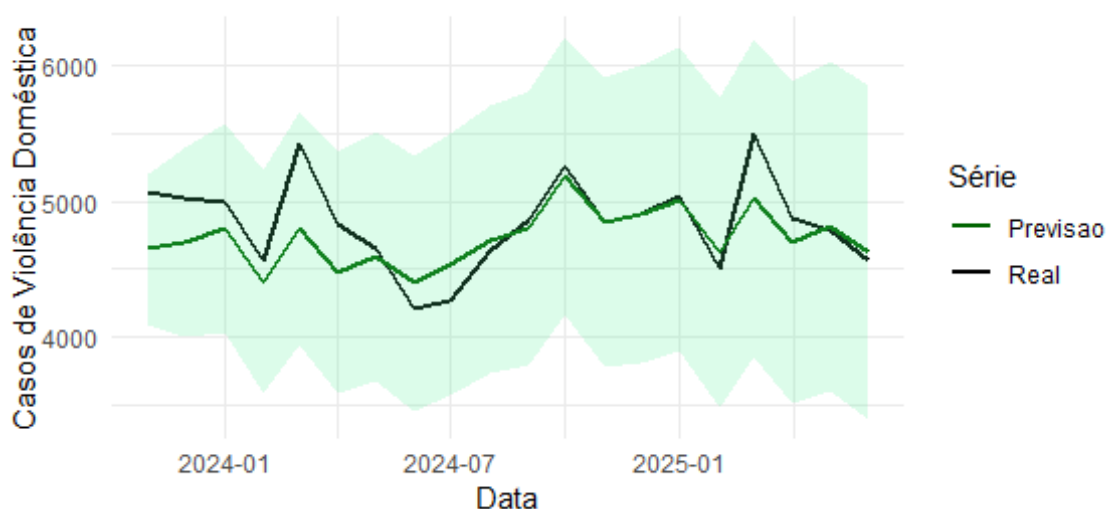
Por outro lado, o teste de Shapiro-Wilk obteve um p-valor de 0,0065, rejeitando a hipótese nula de normalidade. Apesar disso, a análise visual por meio do histograma e da curva de densidade indica que os resíduos apresentam simetria e forma próxima à distribuição normal, com leves desvios nas caudas. Tal violação da normalidade, embora estatisticamente significativa, pode ser tolerável para os fins de previsão, utilizando com atenção em dados de previsões subsequentes.

Figura 24 – Modelo SARIMA no conjunto Teste



Fonte: o autor (2025).

Figura 25 – Filtro do modelo de previsão SARIMA no conjunto teste



Fonte: o autor (2025).

A avaliação preditiva do modelo SARIMA foi realizada por meio da separação da série em conjuntos de treino e teste. A Figura 24 apresenta a previsão sobre a série completa, incluindo o histórico desde 2017. Observa-se que o modelo preserva a coerência com a trajetória da série, reproduzindo adequadamente a tendência de longo prazo e o padrão sazonal. Esses resultados reforçam que o modelo SARIMA é estatisticamente adequado para fins de previsão de curto prazo, apesar das limitações previamente identificadas na análise dos resíduos. A Figura 25 ilustra a performance do modelo no conjunto de teste, indicando que as previsões seguem de forma satisfatória o padrão dos dados observados. Os principais picos e vales foram adequa-

mente capturados, e a maioria dos valores reais permanece dentro dos intervalos de confiança estimados, sugerindo boa capacidade preditiva do modelo.

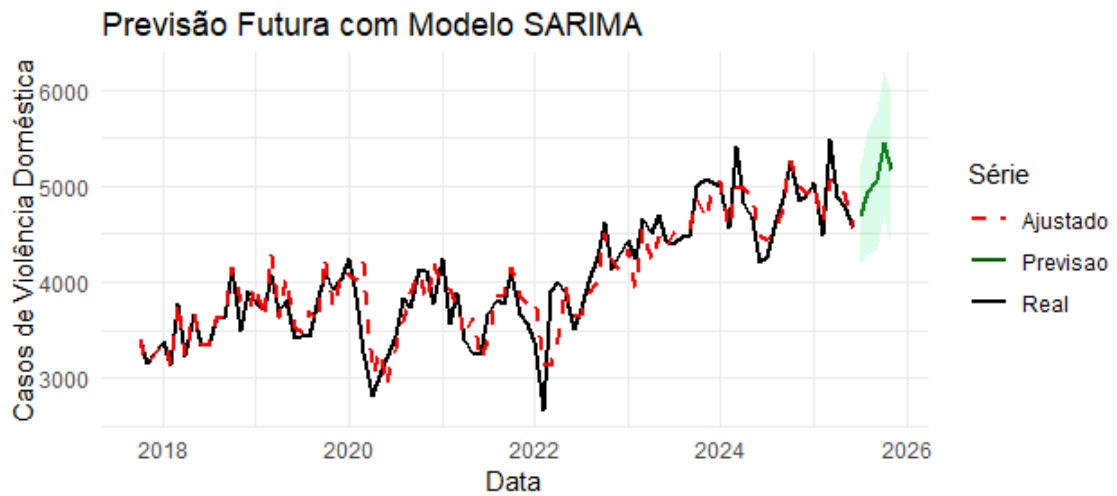
Tabela 5 – Comparação entre valores previstos e observados (dez/2023 a jun/2025)

Mês/Ano	Previsão	Real	Mês/Ano	Previsão	Real
Dez/23	4692,18	5026	Set/24	4796,40	4866
Jan/24	4798,01	4997	Out/24	5183,58	5267
Fev/24	4405,09	4558	Nov/24	4847,64	4847
Mar/24	4803,11	5418	Dez/24	4900,90	4907
Abr/24	4476,46	4832	Jan/25	5013,25	5043
Mai/24	4589,20	4655	Fev/25	4624,41	4497
Jun/24	4394,29	4201	Mar/25	5024,98	5492
Jul/24	4528,61	4270	Abr/25	4699,93	4878
Ago/24	4719,00	4643	Mai/25	4813,66	4785
			Jun/25	4619,37	4560

O modelo SARIMA ajustado para a série temporal de casos mensais de violência doméstica apresentou desempenho satisfatório em termos preditivos, capturando adequadamente tanto os componentes sazonais quanto a tendência de longo prazo da série. A aderência do modelo foi confirmada por meio de validações gráficas e análise dos resíduos, os quais não apresentaram autocorrelação significativa, indicando bom ajuste. A Figura 24 exibe a previsão realizada sobre o conjunto de teste, na qual é possível observar que o modelo foi capaz de acompanhar com precisão as flutuações da série observada, especialmente nos períodos mais recentes, demonstrando sensibilidade tanto às oscilações sazonais quanto às variações pontuais.

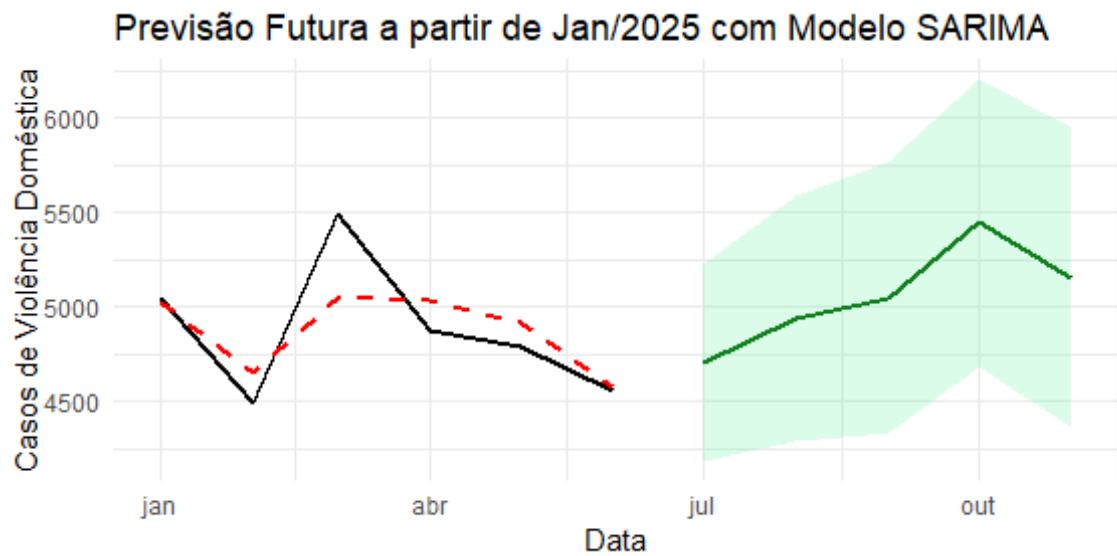
Com base nesse modelo, foi realizada uma projeção dos valores futuros da série a partir de julho de 2025, com horizonte limitado a cinco meses. A opção por restringir o período de previsão decorre da natureza dos modelos de séries temporais, que, quando estendidos por longos períodos, tendem a apresentar comportamento de extrapolação não realista, com tendência artificial de crescimento ou suavização excessiva.

Figura 26 – Serie Historia Modelagem SARIMA e Previsão



Fonte: o autor (2025).

Figura 27 – Filtro em 2025 modelagem SARIMA e Previsão



Fonte: o autor (2025).

As Figuras 26 e 27 e a Tabela 6 mostram as estimativas para o segundo semestre de 2025, juntamente com o intervalo de confiança associado. As previsões indicam a manutenção de um patamar elevado de casos, com variações mensais compatíveis com o padrão sazonal previamente identificado na série. Esses resultados oferecem suporte para o monitoramento de tendências futuras e podem subsidiar ações estratégicas de enfrentamento à violência com base em evidências.

Tabela 6 – Previsão do número de vítimas e erro padrão para os cinco períodos seguintes

Período	Previsão	Erro Padrão (SE)
Jul/2025	4704,58	266,83
Ago/2025	4937,76	331,70
Set/2025	5048,07	365,89
Out/2025	5446,11	387,67
Nov/2025	5156,95	403,52

4.3 DISCUSSÃO DOS IMPACTOS COM VARIÁVEIS EXÓGENAS

4.3.1 Discussão dos Impactos da pandemia do COVID-19

O teste t de Welch indicou que a média mensal de vítimas de violência doméstica aumentou significativamente após o início da pandemia ($p < 0,001$), com um crescimento médio de aproximadamente 541 casos por mês. No entanto, esse resultado baseia-se em uma comparação direta entre médias brutas, sem levar em consideração possíveis tendências temporais nem a dependência serial presente nos dados.

Tabela 7 – Comparação das médias de vítimas antes e pós-pandemia

Período	Média	Desvio Padrão	n
Antes da pandemia	3645,52	314,00	29
Pós-pandemia	4186,53	651,00	64

Teste t de Welch (médias)	
t	-5,405
gl (graus de liberdade)	90,60
p -valor	$5,21 \times 10^{-7}$
IC 95%	[-739,86 ; -342,17]
Conclusão	Diferença significativa entre as médias

Para uma análise mais robusta do impacto da pandemia na série de violência doméstica, foi criada uma variável dummy representando os períodos pré e pós-pandemia. Em seguida, foram ajustados modelos ARIMA com variável exógena (ARIMAX), incorporando esse fator de intervenção.

O modelo inicial utilizado foi um ARIMA(0,1,1) com regressão, sugerido automaticamente pelo pacote `auto.arima` do software R. A partir dessa base, foram testadas configurações adicionais com estrutura sazonal, selecionadas com base na análise dos gráficos de autocorrelação (ACF) e autocorrelação parcial (PACF). Os resultados comparativos dos modelos encontram-se na Tabela 8. Entre as alternativas testadas, o modelo ARIMA(1,1,1)(0,1,1)[12] com variável exógena foi o que apresentou o melhor desempenho. Este modelo obteve o menor valor de AIC, indicando o melhor equilíbrio entre qualidade de ajuste e complexidade, além de apresentar os

menores valores de erro, com RMSE = 188,74 e MAE = 143,13, evidenciando maior precisão nas previsões. Assim, esse modelo foi considerado o mais adequado para representar a série temporal, controlando os efeitos da pandemia.

Tabela 8 – Comparação de modelos ARIMA com variável exógena (xreg)

Modelo	AIC	RMSE	MAE	MAPE (%)	xreg (efeito)	Erro Padrão (xreg)	σ^2 (Resíduo)
ARIMA(0,1,1)	1335.54	326.44	269.36	6.88	-910.04	273.05	111353
ARIMA(1,1,1)(0,1,1)[12]	1116.57	188.74	143.13	3.65	-1014.80	205.59	43593
ARIMA(1,1,1)(1,1,1)[12]	1118.57	188.76	143.15	3.65	-1014.54	212.20	44179
ARIMA(0,1,1)(0,1,1)[12]	1119.74	200.60	151.39	3.84	-1274.45	228.32	48604
ARIMA(0,1,1)(1,1,1)[12]	1121.67	199.65	150.71	3.83	-1300.64	250.41	48776

Para melhor analisar o modelo escolhido foi analisado os resíduos do modelo. o teste de Shapiro-Wilk, utilizado para verificar a normalidade dos resíduos, apresentou $W=0,9788$ com $p=0,1349$. Como o valor-p é superior a 0,05, não se rejeita a hipótese nula de normalidade, o que indica que os resíduos seguem, aproximadamente, uma distribuição normal. Adicionalmente, o teste de Box-Ljung, aplicado aos resíduos, apresentou um valor de $X=0,023$, com $p=0,8794$, indicando que não há evidência de autocorrelação serial significativa nos resíduos ao nível de 5%. Esses resultados sugere que o modelo capturou de forma satisfatória a dependência temporal da série, não restando estrutura autocorrelacionada remanescente.

O modelo selecionado foi ajustado com base em toda a série histórica, incorporando uma variável *dummy* representando o período pós-pandemia (a partir de março de 2020). A equação estimada pelo modelo é expressa da seguinte forma:

$$Y_t = Y_{t-1} + \phi_1(Y_{t-1} - Y_{t-2}) + \theta_1 u_{t-1} + \Theta_1 u_{t-12} + \beta x_t + u_t \quad (4.1)$$

Com os parâmetros estimados:

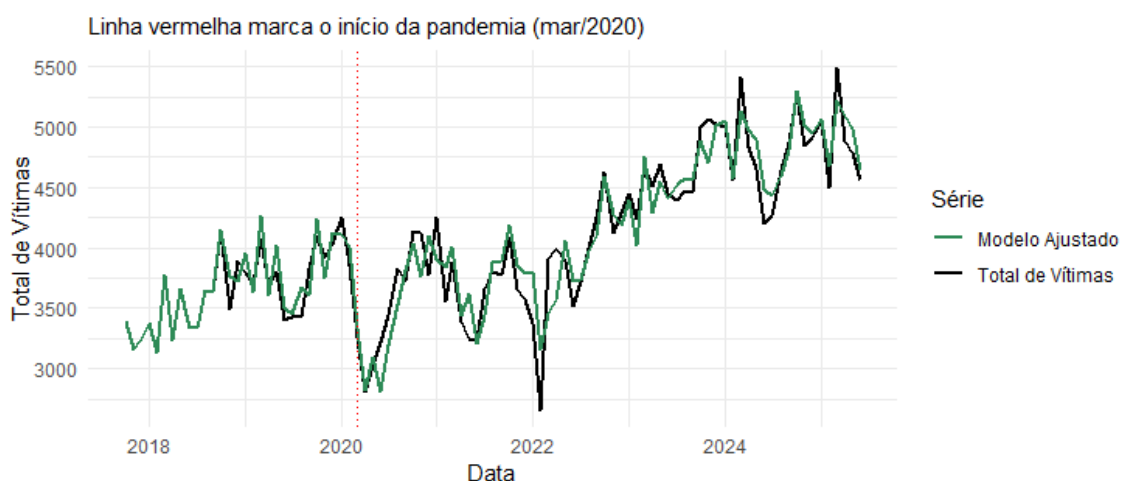
$$Y_t = Y_{t-1} + 0,7071(Y_{t-1} - Y_{t-2}) - 1,0000 u_{t-1} - 0,9999 u_{t-12} - 1014,8039 x_t + u_t \quad (4.2)$$

Sendo:

- Y_t : número previsto de vítimas no mês t ;
- x_t : variável *dummy* indicadora da pandemia (0 antes de março/2020, 1 a partir de março/2020);
- u_t : erro aleatório do modelo no tempo t ;
- $\phi_1 = 0,7071$: parâmetro autorregressivo (AR1), indicando dependência com a variação anterior;
- $\theta_1 = -1,0000$: parâmetro de média móvel (MA1), responsável pelo ajuste do erro anterior;

- $\Theta_1 = -0,9999$: componente sazonal de média móvel (SMA1), que considera o erro defasado de 12 meses;
- $\beta = -1014,8039$: coeficiente associado à variável exógena, estimando a redução média mensal durante o período pós-pandemia.

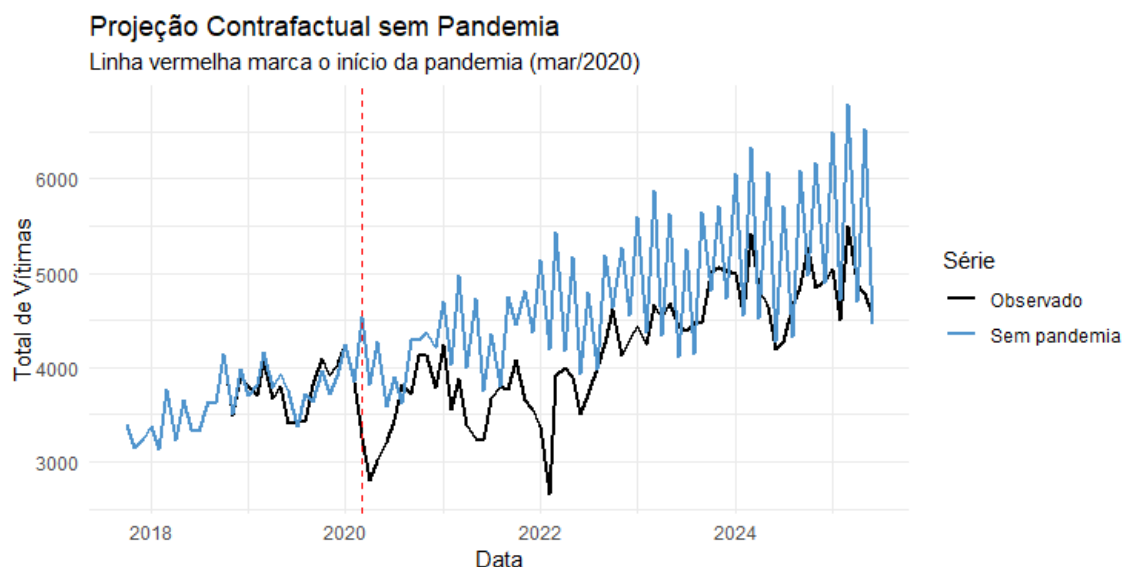
Figura 28 – Modelo ARIMA com Dummy Exógena e Previsão: COVID-19



Fonte: o autor (2025).

A partir da comparação simples entre as médias dos períodos pré e pós-pandemia, infere-se, à primeira vista, que a pandemia provocou um aumento no número de vítimas de violência doméstica. No entanto, ao se controlar por fatores como tendência, sazonalidade e estrutura serial da série temporal, o modelo **ARIMA(1,1,1)(0,1,1)[12]** com variável exógena indicou um efeito redutor estimado de aproximadamente **1.015 vítimas por mês**. Ou seja, durante os 64 meses compreendidos pelo período pandêmico e pós-pandêmico, o modelo estima uma redução acumulada de cerca de **64.947 casos** de violência doméstica, em relação ao que seria esperado na ausência da pandemia, mantendo os demais fatores constantes.

Figura 29 – Modelo ARIMA previsão sem a COVID-19



Fonte: o autor (2025).

Esse resultado sugere que, embora o número absoluto de vítimas tenha aumentado após 2020, esse crescimento foi inferior ao que se projetaria com base na tendência histórica anterior. A pandemia, portanto, pode ter interrompido ou atenuado um processo de crescimento mais acelerado, distorcendo a percepção inicial baseada apenas na comparação direta entre médias.

Para avaliar estatisticamente se houve diferença significativa entre os valores observados e os valores contrafactuais previstos durante o período pós-pandemia, foi aplicado o teste t para amostras pareadas. O objetivo desse teste é verificar se a média das diferenças entre os valores reais e os estimados (sem pandemia) difere significativamente de zero.

Tabela 9 – Teste t pareado entre valores observados e contrafactuais (pós-pandemia)

Dif. Média	IC 95%	Valor t	gl	p-valor
21,77	[−42,12 ; 85,67]	0,681	63	0,498

Os resultados do teste, apresentados na Tabela 9, indicam que a diferença entre os valores observados e os valores contrafactuais não foi estatisticamente significativa ($p = 0,498$), com intervalo de confiança que inclui o zero. Isso reforça a interpretação de que, embora o modelo estime um efeito de redução no número de casos pós pandemia, essa diferença não se manifestou de forma estatisticamente robusta mês a mês. Em outras palavras, não há evidências suficientes para afirmar que os valores reais observados se desviaram significativamente do que seria esperado na ausência da pandemia.

4.3.2 Discussão dos Impactos de políticas publicas

Os testes t de Welch indicaram que a média mensal de vítimas de violência doméstica aumentou significativamente após a implementação da política pública “Juntos pela Segurança”,

com p-valores inferiores a 0,001 em todas as defasagens testadas.

Tabela 10 – Comparação das médias de vítimas antes e após a política pública

Defasagem	Média (Pré)	Média (Pós)	Desvio (Pré)	Desvio (Pós)	IC 95 % Dif.	p-valor
0 meses	3809,68	4828,53	490	344	[-1215,69 ; -822,01]	$6,91 \times 10^{-13}$
3 meses	3841,30	4807,00	520	358	[-1180,23 ; -751,16]	$1,37 \times 10^{-10}$
6 meses	3882,87	4779,36	556	352	[-1128,59 ; -664,38]	$1,94 \times 10^{-8}$

A Tabela 10 apresenta os resultados dos testes t de Welch aplicados para comparar as médias mensais de vítimas de violência doméstica antes e após a implementação da política “Juntos pela Segurança”, considerando diferentes defasagens temporais. Em todos os cenários, observou-se um aumento significativo nas médias pós-política, com p-valores inferiores a 0,001 e intervalos de confiança que não incluem zero. Isso indica que o aumento médio mensal é estatisticamente robusto, mesmo após considerar possíveis atrasos no impacto da política.

Para controlar efeitos de tendência, sazonalidade e autocorrelação da série, foram ajustados modelos ARIMAX com a variável *dummy* representando a política “Juntos pela Segurança”. Como o período do modelo é aproximado do período anteriormente calculado para a modelagem geral da série, foi ajustado um modelo ARIMA(1,1,1)(0,1,1)[12], que acomoda tendência e sazonalidade por diferenciações de ordem regular e sazonal via diferenciação e controlando apenas a variável exógena.

Tabela 11 – Comparação dos modelos ARIMAX para a política “Juntos pela Segurança”

Lag	Coef. (xreg)	Erro Padrão	AIC	RMSE	MAE	Significativo?
0 meses	-2,04	241,05	1140,96	237,51	168,61	Não
3 meses	-13,18	238,98	1140,95	237,46	168,45	Não
6 meses	-266,72	240,75	1139,74	232,44	163,62	Não

A Tabela 11 apresenta os coeficientes estimados da política pública “Juntos pela Segurança” nos modelos ARIMAX ajustados com defasagens de 0, 3 e 6 meses. Apesar da melhora marginal nos critérios de ajuste (menor AIC e menor RMSE para lag 6), em nenhum dos modelos o coeficiente da variável exógena foi estatisticamente significativo, dado o alto erro padrão em relação ao valor estimado. Isso indica que não há evidência estatística robusta de que a política tenha produzido efeito significativo sobre a evolução da série, dentro da estrutura temporal analisada.

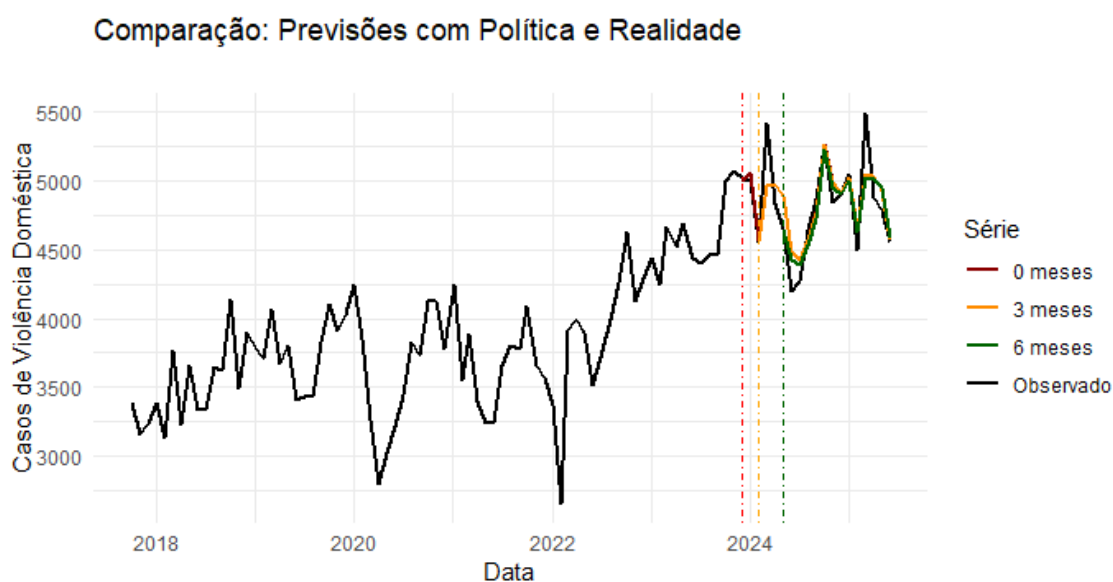
Tabela 12 – Comparação entre valores observados e previstos pelos modelos ARIMAX — Teste t pareado

Lag	Dif. média	IC 95 %	Valor t	p-valor
0 meses	-19,31	[-113,45 ; 74,82]	-0,431	0,6716
3 meses	-18,61	[-124,88 ; 87,66]	-0,371	0,7154
6 meses	-5,56	[-106,31 ; 95,19]	-0,119	0,9069

Para avaliar a acurácia dos modelos ARIMAX ajustados, foi realizada uma comparação entre os valores previstos durante o período de intervenção da política “Juntos pela Segurança” e os valores reais observados na série. Utilizou-se o teste t pareado, que avalia se a diferença média entre os pares (previsto vs. observado) é estatisticamente significativa.

Os testes foram realizados para os três modelos com defasagem (lag) de 0, 3 e 6 meses. Em todos os casos, os valores de p foram superiores a 0,05, e os intervalos de confiança para a diferença média incluem o zero, indicando que não há evidência de que as previsões dos modelos diferem significativamente dos valores reais observados durante o período da intervenção. Isso reforça a ideia de que os modelos capturaram adequadamente o comportamento da série, mas também sugere que o impacto estimado da política não se diferenciou do que já era esperado segundo a tendência histórica da série.

Figura 30 – Modelo ARIMA com Dummy Exógena e Previsão: Política Publica



Fonte: o autor (2025).

Esses resultados sugerem que, uma vez controlada a estrutura temporal da série, o impacto da política “Juntos pela Segurança” sobre a dinâmica da violência doméstica não pôde ser estatisticamente detectado. Em outras palavras, embora o número absoluto de casos tenha aumentado no período subsequente à política, esse crescimento pode estar relacionado a tendências já existentes ou a outros fatores não observados. A análise ARIMAX, portanto, reforçando a importância de utilizar modelos que considerem a estrutura temporal da série ao avaliar políticas públicas.

5 CONCLUSÃO

Este trabalho teve como objetivo central analisar estatisticamente os registros de violência doméstica e familiar contra a mulher (VDFCM) no estado de Pernambuco e investigar sua correlação com os casos de feminicídio, considerando também o impacto de políticas públicas e eventos críticos, como a pandemia de COVID-19. Por meio da aplicação de testes de correlação e modelagem de séries temporais, buscou-se compreender a dinâmica desses fenômenos e suas possíveis interações ao longo do tempo.

Os resultados indicaram que, embora tematicamente relacionados, os registros de VDFCM e feminicídio não apresentaram correlação estatisticamente significativa, tanto em sua forma original quanto nas séries diferenciadas. Isso sugere que os feminicídios, por se tratarem de desfechos extremos e pontuais, não seguem necessariamente a mesma trajetória temporal das denúncias de violência doméstica. Tal dissociação pode ser explicada por fatores estruturais, institucionais e culturais, bem como por processos de subnotificação e defasagem temporal entre os eventos.

A análise do impacto da pandemia, conduzida por meio de testes de hipóteses e modelos ARIMAX com variáveis exógenas, apontou uma possível atenuação na tendência de crescimento dos casos de violência doméstica após março de 2020. Embora a média bruta tenha indicado aumento, o modelo ajustado sugeriu uma redução estimada de aproximadamente 1.015 vítimas mensais. Contudo, os testes de comparação não revelaram diferença estatisticamente significativa mês a mês, indicando que a pandemia pode ter moderado a tendência crescente, sem representar uma ruptura estatística clara.

Em relação à política pública “Juntos pela Segurança”, apesar de aumentos nas médias após sua implementação, os modelos ARIMAX não identificaram efeitos estatisticamente significativos atribuíveis à política. Esse resultado pode refletir a sobreposição de múltiplos fatores ou a necessidade de um período mais longo para que seus impactos possam ser adequadamente captados.

Além disso, foi aplicada uma análise preditiva com base em séries temporais, com o intuito de antecipar o comportamento futuro das variáveis analisadas. Essa etapa demonstrou potencial para subsidiar estratégias de prevenção e alocação de recursos, reafirmando o papel dos modelos estatísticos como ferramentas valiosas na gestão de políticas públicas voltadas ao enfrentamento da violência de gênero.

Entre as limitações do estudo, destaca-se a ausência de variáveis explicativas mais específicas, como perfil sociodemográfico das vítimas e agressores, grau de parentesco, reincidência, tempo até a denúncia e localização geográfica, que poderiam ampliar a capacidade explicativa dos modelos. Além disso, a análise foi realizada com base no número de agressões registradas, e não no número total de vítimas, o que limita a sensibilidade e a granularidade das inferências.

Para estudos futuros, recomenda-se o uso de abordagens baseadas em aprendizado de má-

quina, voltadas à análise regionalizada e à incorporação de variáveis contextuais, conforme apontado na etapa de caracterização dos dados. Essa estratégia poderá ampliar o poder preditivo das análises e oferecer subsídios mais eficazes e territorializados para políticas públicas direcionadas ao enfrentamento da violência contra a mulher. Complementarmente, o emprego da Função de Correlação Cruzada (CCF) poderia contribuir para avaliar, de forma estatisticamente sistemática, as correlações entre VDFCM e feminicídio com diferentes defasagens temporais, permitindo identificar padrões de precedência entre denúncias de violência doméstica e ocorrências de feminicídio. Tal recurso ofereceria subsídios adicionais tanto para a modelagem preditiva quanto para a formulação de políticas públicas mais assertivas.

Dessa forma, o presente estudo não tem a intenção de encerrar os debates sobre a violência doméstica, mas sim de contribuir com evidências empíricas que auxiliem na compreensão e no enfrentamento desse grave problema social. Ao utilizar ferramentas estatísticas avançadas, como modelos de séries temporais e testes de correlação, buscou-se aprofundar a análise da dinâmica dos registros de violência e feminicídios no estado de Pernambuco. Espera-se que os achados apresentados sirvam de subsídio para o aprimoramento de políticas públicas, fomentem novas investigações e estimulem abordagens cada vez mais embasadas na ciência de dados e na realidade territorial dos contextos analisados.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716–723, 1974. MR0423716. Disponível em: <https://doi.org/10.1109/TAC.1974.1100705>.
- BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: Forecasting and Control**. San Francisco: Holden-Day, 1970.
- Brasil. **Lei nº 11.340, de 7 de agosto de 2006. Cria mecanismos para coibir a violência doméstica e familiar contra a mulher**. 2006. http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/11340.htm. Acesso em: 22 jul. 2025.
- Brasil. **Lei nº 13.104, de 9 de março de 2015. Altera o art. 121 do Decreto-Lei nº 2.848, de 7 de dezembro de 1940 – Código Penal, para incluir o feminicídio como circunstância qualificadora do homicídio e o insere no rol dos crimes hediondos**. 2015. http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/113104.htm. Acesso em: 22 jul. 2025.
- Brasil. **Lei nº 13.641, de 3 de abril de 2018. Dispõe sobre o crime de descumprimento de medidas protetivas**. 2018. Diário Oficial da União, Brasília, DF, 4 abr. 2018.
- Brasil. **Lei nº 14.188, de 28 de julho de 2021. Institui o Programa Sinal Vermelho contra a Violência Doméstica**. 2021. Diário Oficial da União, Brasília, DF, 29 jul. 2021.
- Brasil. **Lei nº 14.720, de 31 de outubro de 2023. Altera a Lei Maria da Penha para permitir que juízes decretem medidas protetivas sem solicitação da vítima**. 2023. Diário Oficial da União, Brasília, DF, 1 nov. 2023.
- BROCKWELL, P. J.; DAVIS, R. A. **Introduction to Time Series and Forecasting**. 2. ed. New York: Springer, 2002.
- CHAI, T.; DRAXLER, R. R. Root mean square error (rmse) or mean absolute error (mae)? – arguments against avoiding rmse in the literature. **Geoscientific Model Development**, v. 7, n. 3, p. 1247–1250, 2014. Disponível em: <https://doi.org/10.5194/gmd-7-1247-2014>. Acesso em: 10 ago. 2025.
- COELHO, M. d. M. F. et al. Temporal evolution and projections of violence against women in brazil: Scenarios from 2013 to 2033. **Journal of Interpersonal Violence**, aug 2025. Epub ahead of print. PMID: 40782164. Disponível em: <https://doi.org/10.1177/08862605251355969>.
- Conselho Nacional de Justiça. **Reportagem premiada relata aumento dos casos de violência doméstica em Pernambuco**. Brasília: CNJ, 2023. <https://www.cnj.jus.br/reportagem-premiada-relata-aumento-dos-casos-de-violencia-domestica-em-pernambuco>. Acesso em: 31 jul. 2025.
- Conselho Nacional de Justiça. **Maria da Penha: os 18 anos de aperfeiçoamento da lei criada para proteger as brasileiras**. 2024. <https://www.cnj.jus.br/maria-da-penha-os-18-anos-de-aperfeiçoamento-da-lei-criada-para-proteger-as-brasileiras>. Acesso em: 2 ago. 2025.

COSTA, A. A. A. O movimento feminista no brasil: dinâmica de uma intervenção política. In: MELO, H. P. et al. (Ed.). **Olhares feministas**. Brasília, DF: Ministério da Educação/Unesco, 2007. p. 50–82.

DEININGER, L. d. S. C. **Modelo preditor de risco para violência doméstica contra a mulher**. Tese (Doutorado) — Universidade Federal da Paraíba, Centro de Ciências Exatas e da Natureza, Programa de Pós-Graduação em Modelos de Decisão e Saúde, João Pessoa, PB, 2020. Tese (Doutorado) — Doutora em Modelos de Decisão e Saúde; Orientadores: Hemílio Fernandes Campos Coelho e Rodrigo Pinheiro de Toledo Vianna. Disponível em: <https://repositorio.ufpb.br/jspui/handle/123456789/20226>. Acesso em: 10 ago. 2025.

D'OLIVEIRA, A. F. L. P.; SCHRAIBER, L. B. Violência e saúde: contribuições teóricas, metodológicas e éticas de estudos da violência contra a mulher. **Interface – Comunicação, Saúde, Educação**, Botucatu, v. 3, n. 5, p. 13–26, 1999.

ENDERS, W. **Applied Econometric Time Series**. 4. ed. Hoboken: Wiley, 2014.

FITTERER, J. L.; NELSON, T. A. Uma revisão dos métodos estatísticos e quantitativos usados para estudar crimes atribuíveis ao álcool. **PLOS ONE**, v. 10, n. 9, p. e0139344, 2015. Disponível em: <https://doi.org/10.1371/journal.pone.0139344>. Acesso em: 09 ago. 2025.

Fórum Brasileiro de Segurança Pública. Violência contra a mulher em pernambuco: uma abordagem epidemiológica com análise de clusterização. **Revista Brasileira de Segurança Pública**, São Paulo, v. 17, n. 1, p. 58–77, 2023. Acesso em: 31 jul. 2025. Disponível em: <https://revista.forumseguranca.org.br/rbsp/article/view/1668/734>.

Fórum Brasileiro de Segurança Pública. **19º Anuário Brasileiro de Segurança Pública**. São Paulo: Fórum Brasileiro de Segurança Pública, 2025. <https://publicacoes.forumseguranca.org.br/handle/123456789/279>. Ano 19; 434 p.; ISSN 1983-7364.

Governo de Pernambuco. **Juntos pela Segurança: Plano Estadual de Segurança Pública baseado em Evidências**. Recife: Secretaria de Defesa Social, 2023. <https://juntospelaseguranca.pe.gov.br>. Acesso em: 22 jul. 2025.

Governo de Pernambuco – Secretaria de Defesa Social. **Conheça o Plano Estadual de Segurança Pública e Defesa Social**. 2023. <https://www.sds.pe.gov.br/noticias/11790-conheca-o-plano-estadual-de-seguranca-publica-e-defesa-social>. Acesso em: 2 ago. 2025.

GRAUNT, J. **Natural and Political Observations Made upon the Bills of Mortality**. [S.l.]: Royal Society, 1662. Reimpresso em várias edições modernas.

HELLUM, A.; AASEN, H. S. (Ed.). **Women's human rights: CEDAW in international, regional and national law**. Cambridge: Cambridge University Press, 2013.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2. ed. Australia: OTexts, 2018.

Instituto Brasileiro de Geografia e Estatística (IBGE). **Cidades e Estados: Pernambuco**. 2025. <https://www.ibge.gov.br/cidades-e-estados/pe.html>. Acesso em: 2 ago. 2025.

Jornal Digital Recife. **TJPE disponibiliza pedido de medida protetiva eletrônica para vítimas de violência doméstica**. 2025. <https://jornaldigital.recife.br/2025/07/22/medida-protetiva-eletronica-tjpe>. Acesso em: 2 ago. 2025.

KATERNDAHL, D. A. et al. Complex dynamics in intimate partner violence: a time series study of 16 women. **Primary Care Companion to the Journal of Clinical Psychiatry**, v. 12, n. 4, p. PCC.09m00859, 2010.

KURNIAWAN, Z. et al. Trend analysis and prediction of violence against women and children cases in jakarta based on the victim's education level using arima and sarima method. **Jurnal SISFOKOM (Sistem Informasi dan Komputer)**, v. 14, n. 2, p. 250–259, 2025. Licença CC BY 4.0. Submetido: 30 abr. 2025; Revisado: 9 mai. 2025; Aceito: 16 mai. 2025; Publicado: 26 mai. 2025. Disponível em: https://www.researchgate.net/publication/392003183_Trend_Analysis_and_Prediction_of_Violence_Against_Women_and_Children_Cases_in_Jakarta_Based_on_the_Victim's_Education_Level_Using_ARIMA_and_SARIMA_Method. Acesso em: 10 ago. 2025.

LIMA, V.; OLIVEIRA, J. A. d. Identifying risk patterns in brazilian police reports preceding femicides: A long short term memory (lstm) based analysis. **arXiv preprint**, 2024. ArXiv:2401.12980, submitted to IEEE Global Humanitarian Technology Conference (GHTC 2023). Disponível em: <https://doi.org/10.48550/arXiv.2401.12980>. Acesso em: 10 ago. 2025.

LJUNG, G. M.; BOX, G. E. P. On a measure of lack of fit in time series models. **Biometrika**, Oxford University Press, v. 65, n. 2, p. 297–303, 1978.

LUBEYA, M. K. et al. Using the arima model to forecast sexual and gender-based violence cases reported to a tertiary hospital in lusaka, zambia. **PAMJ-One Health**, v. 5, n. 4, p. 4, 2021. Artigo em acesso aberto sob licença CC BY 4.0. Disponível em: <https://www.one-health.panafrican-med-journal.com/content/article/5/4/full>. Acesso em: 10 ago. 2025.

Marco Zero Conteúdo. **Pernambuco é o estado com mais feminicídios no Nordeste**. 2023. <https://marcozero.org/pernambuco-e-o-estado-com-mais-feminicidios-no-nordeste>. Acesso em: 31 jul. 2025.

MENEGHEL, S. N.; HIRAKATA, V. N. Femicídios: homicídios femininos no brasil. **Revista de Saúde Pública**, v. 45, n. 3, p. 564–574, 2011.

MINAYO, M. C. S. d. **Violência e saúde**. Rio de Janeiro: Editora Fiocruz, 2006. (Coleção Temas em Saúde).

MOROSKOSKI, M. et al. Aumento da violência física contra a mulher perpetrada pelo parceiro íntimo: uma análise de tendência. **Ciência e Saúde Coletiva**, ABRASCO - Associação Brasileira de Saúde Coletiva, v. 26, p. 4993–5002, Oct 2021. ISSN 1413-8123. Disponível em: <https://doi.org/10.1590/1413-812320212611.3.02602020>.

NARVAZ, M. G.; KOLLER, S. H. A marginalização dos estudos feministas e de gênero na psicologia acadêmica contemporânea. **Psico**, v. 38, n. 3, p. 216–223, 2006.

NEPOMUCENO, T. C. C. et al. Alcohol and violent behavior among football spectators: An empirical assessment of brazilian's criminalization. **International Journal of Law, Crime and Justice**, Elsevier, v. 51, p. 34–44, 2017.

NEUBAUER, A. R. **[Título Redigido]**. Tese (Doutorado) — University College London, 2024. Tese de doutorado com acesso restrito, versão redigida. Disponível em: https://discovery.ucl.ac.uk/id/eprint/10188930/2/Neubauer_10188930_%20Thesis_Redacted.pdf. Acesso em: 09 ago. 2025.

NYENDE, P. A conceptual analysis of correlates of domestic violence and adolescent risky behavior. **American Journal of Humanities and Social Sciences Research**, v. 8, n. 4, p. 159–166, 2024. Disponível em: <https://www.ajhssr.com/wp-content/uploads/2024/04/R24804159166.pdf>. Acesso em: 10 ago. 2025.

Organização das Nações Unidas. **Convenção sobre a Eliminação de Todas as Formas de Discriminação contra a Mulher (CEDAW)**. 1979. Adotada pela Assembleia Geral da ONU em 18 de dezembro de 1979. Disponível em: <https://www.ohchr.org/en/instruments-mechanisms/instruments/convention-elimination-all-forms-discrimination-against-women>. Acesso em: 22 jun. 2025.

Organização dos Estados Americanos. **Convenção Interamericana para Prevenir, Punir e Erradicar a Violência contra a Mulher – Convenção de Belém do Pará**. Belém: [s.n.], 1994. <http://www.oas.org/juridico/portuguese/treaties/A-61.htm>. Acesso em: 12 jun. 2025.

Organização Mundial da Saúde (OMS). **Violence, injuries and disability: Violence**. 2025. <https://www.emro.who.int/health-topics/violence/index.html>. Acesso em: 2 ago. 2025.

PERRY, W. L. et al. **Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations**. Santa Monica, CA: RAND Corporation, 2013. Disponível em: https://www.rand.org/pubs/research_reports/RR233.html. Acesso em: 09 ago. 2025.

Poder Judiciário do Estado de Pernambuco. **Núcleo do TJPE atua na redução do feminicídio em Caruaru**. [S.l.]: Tribunal de Justiça de Pernambuco, 2024. <https://portal.tjpe.jus.br/-/nucleo-do-tjpe-atua-na-reducao-do-feminicidio-em-caruaru>. Acesso em: 31 jul. 2025.

PRIY, K. Y. Crime rate prediction and analysis using lstm algorithm. **International Journal of Engineering Technologies and Management Research**, v. 12, n. 3, mar. 2025. Licença CC BY 4.0. Disponível em: https://www.researchgate.net/publication/390093417_CRIME_RATE_PREDICTION_AND_ANALYSIS_USING_LSTM_ALGORITHM. Acesso em: 09 ago. 2025.

RADAY, F. Gender and democratic citizenship: the impact of cedaw. **International Journal of Constitutional Law**, v. 10, n. 2, p. 512–530, 2012. Disponível em: <https://academic.oup.com/icon/article/10/2/512/666068>. Acesso em: 21 jul. 2025.

REYAL, H. P. et al. Association between individual-level socioeconomic factors and intimate partner violence victimisation in women: a systematic review protocol. **BMJ Open**, v. 14, n. 3, p. e080117, 2024. Disponível em: <https://doi.org/10.1136/bmjopen-2023-080117>.

SAFFIOTI, H. I. B. Já se mete a colher em briga de marido e mulher. **São Paulo em Perspectiva – Revista da Fundação Seade**, São Paulo, v. 13, n. 4, p. 82–91, 1999.

SCHRAIBER, L. B.; D'OLIVEIRA, A. F. L. P.; COUTO, M. T. Violência e saúde: estudos científicos recentes. **Revista de Saúde Pública**, São Paulo, v. 40, n. suplemento especial, p. 112–120, 2007.

SCHRAIBER, L. B.; D'OLIVEIRA, A. F. P. L.; COUTO, M. T. Prevalência de violência contra a mulher por parceiro íntimo: estudo no brasil urbano e rural. **Revista de Saúde Pública**, v. 41, n. 3, p. 359–367, 2007. Acesso em: 2 ago. 2025. Disponível em: <https://www.scielo.br/j/rsp/a/8G54ZFwvFgLQsQtmKtFvtYt/?format=pdf&lang=pt>.

SCOTT, J. **Gênero: uma categoria útil para a análise histórica**. Recife: SOS Corpo e Cidadania, 1990. Disponível em: http://www.dhnet.org.br/direitos/textos/generodh/gen_categoria.html. Acesso em: 08 jun. 2025.

SDS-PE. **Estatísticas criminais – Violência doméstica e familiar contra a mulher e demais indicadores**. 2025. <https://www.sds.pe.gov.br/estatisticas>. Acesso em: 8 ago. 2025.

SEGATO, R. L. Género y colonialidad: en busca de claves de lectura y de um vocabulário estratégico descolonial. In: BIDASECA, K.; LABA, V. V. (Ed.). **Feminismos y Poscolonialidad. Descolonizando el feminismo desde y en América Latina**. Buenos Aires: Godot, 2011. p. 17–48.

Senado Federal (Brasil). **Pesquisa Nacional: Violência doméstica e familiar contra a mulher – Pernambuco**. Brasília: Instituto DataSenado, 2024. https://www.senado.leg.br/institucional/DataSenado/relatorio_online/pesquisa_violencia_domestica/2024/assets/PDF/Pernambuco.pdf. Acesso em: 31 jul. 2025.

Shift / University of Calgary. **Using Data Science to Engage and Mobilize Men for Violence Prevention: Rapid Evidence Review Report (Calling In Men Project)**. 2022. University of Calgary / Shift. Rapid review on engaging men for gender equity and violence prevention. Disponível em: https://preventdomesticviolence.ca/wp-content/uploads/2022/11/R53_Shift_2022_Data_Science_Approach_Rapid_Review_Report_Calling_In_Men.pdf. Acesso em: 09 ago. 2025.

SHUMWAY, R. H.; STOFFER, D. S. **Time Series Analysis and Its Applications: With R Examples**. 4. ed. New York: Springer, 2017. ISBN 978-3-319-52451-1.

THIELE, T. N. Om anvendelse af mindste kvadraters methode paa nogle tilfælde, hvor en komplet teori ikke er mulig. **Tidsskrift for Mathematik**, v. 4, n. 5, p. 1–20, 1880. Tradução: On the Application of the Method of Least Squares to Some Cases Where a Complete Theory Is Not Possible.

TRAVAINI, G. V. et al. Machine learning and criminal justice: A systematic review of advanced methodology for recidivism risk prediction. **International Journal of Environmental Research and Public Health**, v. 19, n. 17, p. 10594, 2022. Disponível em: <https://doi.org/10.3390/ijerph191710594>. Acesso em: 09 ago. 2025.

UN Women. **Facts and figures: Ending violence against women**. 2024. <https://www.unwomen.org/en/articles/facts-and-figures/facts-and-figures-ending-violence-against-women#83915>. Acesso em: 2 ago. 2025.

United Nations Office on Drugs and Crime. **Femicides in 2023: Global Estimates of Intimate Partner/Family Member Femicides**. Vienna: United Nations, 2024. (United Nations publication). Disponível em: <https://coilink.org/20.500.12592/8wctr0k>.

VERREY, J. et al. Using machine learning to forecast domestic homicide via police data and super learning. **Scientific Reports**, v. 13, p. 22932, 2023. Disponível em: <https://doi.org/10.1038/s41598-023-50274-2>. Acesso em: 09 ago. 2025.

WALKER, L. E. **The Battered Woman**. New York: Harper & Row, 1979.

WILLMOTT, C. J.; MATSUURA, K. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. **Climate Research**, v. 30, p. 79–82, 2005. Disponível em: <https://doi.org/10.3354/cr030079>. Acesso em: 10 ago. 2025.

YULE, G. U. On a method of investigating periodicities in disturbed series, with special reference to wolfer's sunspot numbers. **Philosophical Transactions of the Royal Society of London, Series A**, v. 226, p. 267–298, 1927.

ZHENG, X. et al. Exploring the influence of environmental indicators and forecasting influenza incidence using arimax models. **Frontiers in Public Health**, v. 12, p. 1441240, sep 2024. Disponível em: <https://doi.org/10.3389/fpubh.2024.1441240>. Acesso em: 10 ago. 2025.

APÊNDICE A – CÓDIGO-FONTE EM R PARA ANÁLISE DE CORRELAÇÃO

```

1  ##-----Teste de correlao de Pearson -----
2
3  cor.test(Ftemp, Vtemp, method = "pearson")
4
5  dfP <- data.frame(
6    tempo = time(Ftemp),
7    Ftemp = as.numeric(Ftemp),
8    Vtemp = as.numeric(Vtemp)
9  )
10
11 # Grfico de disperso com linha de regresso
12 ggplot(dfP, aes(x = Ftemp, y = Vtemp)) +
13   geom_point(color = "steelblue") +
14   geom_smooth(method = "lm", se = FALSE, color = "red") +
15   labs(title = "Correlao de Pearson: Feminicidio v. VDFCM",
16         x = "Feminicidio",
17         y = "Violncia Domstica") +
18   theme_classic()
19
20 # Sries diferenciadas
21 violencia_diff <- diff(V_mensal$TotaldeVitimas)
22 F_diff <- diff(F_mensal$Totaldevitimas)
23
24 #analisar presupostos de person-----
25 shapiro.test(F_diff)
26 shapiro.test(violencia_diff)
27
28 outlierTest(lm(violencia_diff ~ F_diff))
29
30 boxplot(F_diff, main = "Feminicidio Diferenciado")
31 boxplot(violencia_diff, main = "Violncia Domstica Diferenciada")
32
33 # Criar grupos de violencia_diff (acima e abaixo da mediana)
34 grupo <- ifelse(violencia_diff > median(violencia_diff), "alto", "
    baixo")
35
36 # Teste de Levene
37 leveneTest(F_diff ~ grupo)

```

```
38
39 grupo2 <- ifelse(F_diff > median(F_diff), "alto", "baixo")
40 leveneTest(violencia_diff ~ grupo2)
41
42
43 # Teste de correlao de Pearson para sries diferenciadas
44 cor.test(F_diff, violencia_diff, method = "pearson")
45
46 # Grfico da correlao entre as sries diferenciadas
47 dfdif <- data.frame(
48   F_diff = as.numeric(F_diff),
49   V_diff = as.numeric(violencia_diff)
50 )
51
52 ggplot(dfdif, aes(x = F_diff, y = V_diff)) +
53   geom_point(color = "steelblue") +
54   geom_smooth(method = "lm", se = FALSE, color = "red") +
55   labs(title = "Correlao de Pearson entre Sries Diferenciadas",
56        x = "Feminicidio",
57        y = "VDFCM") +
58   theme_classic()
```

Código 1 – Código R para análise de Correlação

APÊNDICE B – CÓDIGO-FONTE EM ANÁLISE DE PREVISÕES

```

1  ##----Modelos de Previso-----
2
3  #Seperar treinamento e teste 80% 20%
4  treino <- window(Vtemp, end = c(2023, 10))
5  test <- window(Vtemp, start = c(2023, 11))
6  plot(treino)
7  plot(test)
8
9  #Plotar Resultados para ver se esto estacionarios
10 par(mfrow = c(1, 2))
11 plot(treino)
12 qqnorm(treino)
13 qqline(treino)
14 acf(treino)
15 pacf(treino)
16 par(mfrow = c(1, 1))
17
18 adf.test(treino) #no conseguimos rejeitar a hiptese nula no
    estacionaridade
19
20 #Remover tendencia e sazonalidade ##Precisa deixar estacionaria
21 Vdif<- diff(diff(treino,lag = 12))
22 adf.test(Vdif)
23
24 #Plotar Resultados para ver se esto estacionarios
25 par(mfrow = c(1, 2))
26 plot(Vdif)
27 qqnorm(Vdif)
28 qqline(Vdif)
29 acf(Vdif)
30 pacf(Vdif)
31 par(mfrow = c(1, 1))
32
33
34 #Testar de modelos
35 auto.arima(treino, seasonal = TRUE)
36 auto.arima(Vdif, seasonal = TRUE)
37

```

```

38 a1<-Arima(treino, order = c(0,1,1), seasonal = list(order = c(0,1,1),
    period = 12))
39 a2<-Arima(treino, order = c(1,1,1), seasonal = list(order = c(0,1,1),
    period = 12))
40 a3<-Arima(treino, order = c(0,1,3), seasonal = list(order = c(1,0,0),
    period = 12))
41 a4<-Arima(treino, order = c(0,1,1), seasonal = list(order = c(0,0,1),
    period = 12))
42 a5<-Arima(treino, order = c(0,1,1), seasonal = list(order = c(0,1,0),
    period = 12))
43 a6<-Arima(treino, order = c(0,1,1), seasonal = list(order = c(1,1,0),
    period = 12))
44 a7<-Arima(treino, order = c(0,1,1), seasonal = list(order = c(1,1,1),
    period = 12))
45 a8<-Arima(treino, order = c(1,1,1), seasonal = list(order = c(0,1,0),
    period = 12))
46 a9<-Arima(treino, order = c(1,1,1), seasonal = list(order = c(1,1,0),
    period = 12))
47 a10<-Arima(treino, order = c(1,1,1), seasonal = list(order = c(1,1,1)
    , period = 12))
48
49
50 #Visualizar modelos com testes
51 plot(resid(a2))
52 plot(treino, lwd=2)
53 lines(treino-a1$residuals,col="blue")
54 lines(treino-a2$residuals,col="red")
55 lines(treino-a3$residuals,col="green")
56 lines(treino-a4$residuals,col="steelblue")
57
58
59 #Previso para o modelo escolhido
60
61 prevSarima<-predict(a2,n.ahead = 20)
62
63 plot(window(Vtemp, start = c(2022, 1)))
64 lines(treino-a2$residuals,col="red")
65 lines(prevSarima$pred,lwd=2,col="green")
66 lines(prevSarima$pred+prevSarima$se*1.96,lwd=2,col="seagreen")
67 lines(prevSarima$pred-prevSarima$se*1.96,lwd=2,col="seagreen3")
68

```

```

69
70 #analisar residul ruido branco?
71 #plotar normal
72 par(mfrow = c(1, 2))
73 plot(resid(a2))
74 qqnorm(resid(a2))
75 qqline(resid(a2))
76 acf(resid(a2))
77 pacf(resid(a2))
78 par(mfrow = c(1, 1))
79
80 #testar se o modelo e bom p-value>0,05 bom
81 Box.test(resid(a2), lag = 12, type = "Ljung-Box")
82 shapiro.test(resid(a2)) # Normalidade > 0,05 normal
83
84
85 # Previsao no TESTE SARIMA com 20 passos
86 prevSarima <- predict(a2, n.ahead = 20)
87
88 datas <- seq(as.Date("2017-10-01"), by = "month", length.out = 93)
89
90 # Cria o data frame
91 df_prev <- data.frame(
92   Data = datas,
93   Real = as.numeric(Vtemp),
94   Ajustado = c(as.numeric(a2$fitted), rep(NA, 20)), # fitted at
     treino
95   Previsao = c(rep(NA, 73), as.numeric(prevSarima$pred)), # previsao
     comea no 74
96   IC_sup = c(rep(NA, 73), as.numeric(prevSarima$pred + 1.96 *
     prevSarima$se)),
97   IC_inf = c(rep(NA, 73), as.numeric(prevSarima$pred - 1.96 *
     prevSarima$se))
98 )
99
100
101 ggplot(df_prev, aes(x = Data)) +
102   geom_line(aes(y = Real), color = "black", linewidth = 1) +
103   geom_line(aes(y = Ajustado), color = "red", linewidth = 1) +
104   geom_line(aes(y = Previsao), color = "darkgreen", linewidth = 1) +

```

```

105   geom_ribbon(aes(ymin = IC_inf, ymax = IC_sup), fill = "seagreen2",
106             alpha = 0.2) +
107   labs(title = "Previso no Conjunto de Teste com Srie Completa",
108        x = "Data", y = "Casos de Violncia Domstica") +
109   theme_minimal()
110
111 df_prev %>%
112   filter(Data >= as.Date("2023-11-01")) %>%
113   ggplot(aes(x = Data)) +
114   geom_line(aes(y = Real), color = "black", linewidth = 1) +
115   geom_line(aes(y = Ajustado), color = "red", linewidth = 1) +
116   geom_line(aes(y = Previsao), color = "darkgreen", linewidth = 1) +
117   geom_ribbon(aes(ymin = IC_inf, ymax = IC_sup), fill = "seagreen2",
118             alpha = 0.2) +
119   labs(title = " ",
120        x = "Data", y = "Casos de Violncia Domstica") +
121   theme_minimal()
122
123 ##Previso real
124
125 Prev_futura<- Arima(V_mensal$TotaldeVitimas, order = c(1,1,1),
126                   seasonal = list(order = c(0,1,1), period = 12))
127 prevReal<-predict(Prev_futura, n.ahead = 5)
128
129 prevReal
130
131 datas1 <- seq(as.Date("2017-10-01"), by = "month", length.out = 93 +
132             5)
133
134 # Data frame com tudo
135 real_prev <- data.frame(
136   Data = datas1,
137   Real = c(as.numeric(V_mensal$TotaldeVitimas), rep(NA, 5)),
138   Ajustado = c(as.numeric(Prev_futura$fitted), rep(NA, 5)),
139   Previsao = c(rep(NA, length(V_mensal$TotaldeVitimas)), as.numeric(
140     prevReal$pred)),
141   IC_sup = c(rep(NA, length(V_mensal$TotaldeVitimas)), as.numeric(
142     prevReal$pred + 1.96 * prevReal$sse)),
143   IC_inf = c(rep(NA, length(V_mensal$TotaldeVitimas)), as.numeric(
144     prevReal$pred - 1.96 * prevReal$sse))
145 )
146

```

```

139 ggplot(real_prev, aes(x = Data)) +
140   geom_line(aes(y = Real), color = "black", linewidth = 1) +
141   geom_line(aes(y = Ajustado), color = "red", linetype = "dashed",
142             linewidth = 1) +
143   geom_line(aes(y = Previsao), color = "darkgreen", linewidth = 1) +
144   geom_ribbon(aes(ymin = IC_inf, ymax = IC_sup), fill = "seagreen2",
145             alpha = 0.2) +
146   labs(title = "Previsão Futura com Modelo SARIMA",
147         x = "Data", y = "Casos de Violência Doméstica") +
148   theme_minimal()
149
150 #a partir de 2025
151 real_prev$LinhaPrevista <- ifelse(
152   is.na(real_prev$Ajustado),
153   real_prev$Previsao,
154   real_prev$Ajustado
155 )
156
157 ggplot(real_prev %>% filter(Data >= as.Date("2025-01-01")), aes(x =
158   Data)) +
159   geom_line(aes(y = Real), color = "black", linewidth = 1) +
160   geom_line(aes(y = LinhaPrevista), color = "darkgreen", linewidth =
161     1) + # Nova linha contínua
162   geom_ribbon(aes(ymin = IC_inf, ymax = IC_sup), fill = "seagreen2",
163     alpha = 0.2) +
164   labs(title = "Previsão Futura a partir com Modelo SARIMA",
165         x = "Data", y = "Casos de Violência Doméstica") +
166   theme_minimal()

```

Código 2 – Código R para análise de previsões

APÊNDICE C – CÓDIGO-FONTE EM R PARA ANÁLISE DE COM AS VARIÁVEIS EXÓGENAS

```

1  ## ----- Datas de referencia -----
2  d_pandemia <- as.Date("2020-03-01")
3  d_juntos <- as.Date("2023-12-01")
4  d_juntos3 <- as.Date("2024-02-01")
5  d_juntos6 <- as.Date("2024-05-01")
6  d_pandemia_fim <- as.Date("2023-02-01") # usado em pandemia1
7
8  ## ----- Dummies -----
9  V_mensal <- V_mensal %>%
10   mutate(
11     pandemia = as.integer(MES >= d_pandemia),
12     pandemia1 = as.integer(MES >= d_pandemia & MES <= d_pandemia_fim),
13     juntos = as.integer(MES >= d_juntos),
14     juntos3 = as.integer(MES >= d_juntos3),
15     juntos6 = as.integer(MES >= d_juntos6)
16   )
17
18  ## ----- Helper: Levene + resumo + t-test -----
19  sumario_grupo <- function(df, var_y, var_g) {
20    df %>%
21      group_by({{ var_g }}) %>%
22      summarise(
23        media = mean({{ var_y }}, na.rm = TRUE),
24        desvio = sd({{ var_y }}, na.rm = TRUE),
25        n = n(),
26        .groups = "drop"
27      )
28  }
29
30  ## ----- Testes de variancia (pandemia) -----
31  lapply(c("pandemia", "pandemia1"), \(g)
32    leveneTest(TotaldeVitimas ~ factor(.data[[g]]), data = V_mensal)
33  )
34
35  ## ----- Grficos iniciais (iguais) -----
36  ggplot(V_mensal, aes(x = MES, y = TotaldeVitimas)) +

```

```

37 geom_line(color = "seagreen", linewidth = 1) +
38 geom_vline(xintercept = d_pandemia, linetype = "dashed", color = "
    red", linewidth = 1) +
39 labs(title = "Violncia Domstica em Pernambuco: Pr e ps a Pandemia",
40       subtitle = "Linha vermelha marca o incio da pandemia (mar/2020)"
    ,
41       x = "Data", y = "Total de Vtimas") +
42 theme_minimal()
43
44 ggplot(V_mensal, aes(x = factor(pandemia), y = TotaldeVitimas)) +
45 geom_boxplot(fill = c("lightblue", "seagreen2")) +
46 scale_x_discrete(labels = c("Antes da Pandemia", "Durante/Aps
    Pandemia")) +
47 labs(title = "Distribuio dos Casos de Violncia Domstica",
48       x = "", y = "Total de Vtimas por Ms") +
49 theme_classic()
50
51 ## ----- Comparao de mdias (pandemia1) -----
52 sumario_grupo(V_mensal, TotaldeVitimas, pandemia1)
53
54 pre <- V_mensal$TotaldeVitimas[V_mensal$pandemia1 == 0]
55 pos <- V_mensal$TotaldeVitimas[V_mensal$pandemia1 == 1]
56 t.test(pre, pos, var.equal = FALSE)
57
58 ## ----- Modelos ARIMA -----
59 model <- auto.arima(V_mensal$TotaldeVitimas, xreg = V_mensal$pandemia
    , seasonal = TRUE)
60 modell <- Arima(V_mensal$TotaldeVitimas,
61                order = c(1,1,1),
62                seasonal = list(order = c(0,1,1), period = 12),
63                xreg = V_mensal$pandemia)
64
65 ajuste <- fitted(modell)
66
67 ggplot(V_mensal, aes(x = MES)) +
68 geom_line(aes(y = TotaldeVitimas), color = "black", size = 0.8) +
69 geom_line(aes(y = ajuste), color = "steelblue3", size = 1) +
70 geom_vline(xintercept = d_pandemia, linetype = "dotted", color = "
    red") +
71 labs(title = "Modelo ARIMAX: Efeito da Pandemia",
72       y = "Total de Vtimas", x = "Data") +

```

```

73   theme_minimal()
74
75   ## ----- Contrafactual sem pandemia -----
76   pre_pandemia <- subset(V_mensal, MES < d_pandemia)
77   pos_pandemia <- subset(V_mensal, MES >= d_pandemia)
78
79   modelo_pre <- Arima(pre_pandemia$TotaldeVitimas,
80                       order = c(1,1,1),
81                       seasonal = list(order = c(0,1,1), period = 12))
82
83   prev_contrafactual <- forecast(modelo_pre,
84                                   xreg = rep(0, nrow(pos_pandemia)),
85                                   h = nrow(pos_pandemia))
86
87   df_contrafatual <- data.frame(
88     Data = V_mensal$MES,
89     Real = V_mensal$TotaldeVitimas,
90     Contrafactual = c(fitted(modelo_pre), as.numeric(
91       prev_contrafactual$mean)),
92     Tipo = c(rep("Pr-pandemia", nrow(pre_pandemia)), rep("Ps (prevista)",
93       nrow(pos_pandemia)))
94   )
95
96   ggplot(df_contrafatual, aes(x = Data)) +
97     geom_line(aes(y = Real), color = "black", linewidth = 1) + #
98       corrigido de "black1"
99     geom_line(aes(y = Contrafactual), color = "steelblue3", linewidth =
100       1) +
101     geom_vline(xintercept = d_pandemia, linetype = "dashed", color = "
102       red") +
103     labs(title = "Projeo Contrafactual sem Pandemia",
104          subtitle = " ",
105          y = "Total de Vtimas", x = "Data") +
106     theme_minimal()
107
108   ## ----- Resduos do model1 -----
109   residuosP <- residuals(model1)
110   Box.test(residuosP, type = "Ljung-Box")
111   shapiro.test(residuosP)
112
113   k1 <- ggplot(data.frame(Tempo = time(residuosP), Resduos = as.numeric

```

```

    (residuosP)),
109     aes(x = Tempo, y = Resduos)) +
110   geom_line(color = "violetred3") +
111   labs(title = "Resduos do Modelo SARIMA", x = "Tempo", y = "Resduos"
        ) +
112   theme_minimal()
113
114   k2 <- ggplot(data.frame(Resduos = as.numeric(residuosP)), aes(x =
        Resduos)) +
115   geom_histogram(aes(y = after_stat(density)), bins = 30, fill = "
        violetred3", alpha = 0.5, color = "black") +
116   stat_function(fun = dnorm, args = list(mean = mean(residuosP), sd =
        sd(residuosP)), color = "blue") +
117   labs(title = "Distribuio dos Resduos", x = "Resduo", y = "Densidade
        ") +
118   theme_minimal()
119
120   grid.arrange(k1, k2, ncol = 2)
121   gridExtra::grid.arrange(
122     plot_acf_gg(residuosP, tipo = "acf", titulo = "ACF dos Resduos",
        cor = "violetred4"),
123     plot_acf_gg(residuosP, tipo = "partial", titulo = "PACF dos Resduos
        ", cor = "violetred3"),
124     ncol = 2
125   )
126
127   ## ----- Ajustado vs contrafactual (sem pandemia) -----
128   sem_pandemia <- rep(0, length(V_mensal$pandemia))
129   contrafactual <- fitted(Arima(V_mensal$TotaldeVitimas, model = model1
        , xreg = sem_pandemia))
130
131   df_comp <- data.frame(
132     Data = V_mensal$MES,
133     Real = V_mensal$TotaldeVitimas,
134     Ajustado = fitted(model1),
135     SemPandemia = contrafactual
136   )
137
138   ggplot(df_comp, aes(x = Data)) +
139     geom_line(aes(y = Real), color = "black", linewidth = 0.6) +
140     geom_line(aes(y = SemPandemia), color = "violetred2", linewidth =

```

```

1) +
141 geom_vline(xintercept = d_pandemia, color = "red", linetype = "
      dotdash") +
142 labs(title = "Efeito Estimado sem o impacto da COVID-19",
143       subtitle = "Linha vermelha: incio da pandemia (mar/2020)",
144       y = "Total de Vtimas", x = "Data") +
145 theme_minimal()
146
147 periodo_pandemia <- V_mensal$MES >= d_pandemia
148 t.test(df_comp$Real[periodo_pandemia], df_comp$SemPandemia[
      periodo_pandemia], paired = TRUE)
149
150 ## ===== Juntos pela Segurana =====
151 ## Testes de variancia e normalidade (loop evita repetio)
152 vars_j <- c("juntos", "juntos3", "juntos6")
153
154 lapply(vars_j, \(g) leveneTest(TotaldeVitimas ~ factor(.data[[g]]),
      data = V_mensal))
155 lapply(vars_j, \(g) shapiro.test(V_mensal$TotaldeVitimas[V_mensal[[g
      ]] == 0]))
156
157 ## Grfico com as 3 linhas (igual ao seu)
158 ggplot(V_mensal, aes(x = MES, y = TotaldeVitimas)) +
159   geom_line(color = "seagreen", linewidth = 1) +
160   geom_vline(xintercept = d_juntos, linetype = "dashed", color = "red
      ", linewidth = 1) +
161   geom_vline(xintercept = d_juntos3, linetype = "dashed", color = "
      red3", linewidth = 1) +
162   geom_vline(xintercept = d_juntos6, linetype = "dashed", color = "
      red4", linewidth = 1) +
163   labs(title = "Violncia Domstica em Pernambuco: Pr e ps Juntos",
164        subtitle = "Linha vermelha marca o incio nas diferentes janelas"
      ,
165        x = "Data", y = "Total de Vtimas") +
166   theme_minimal()
167
168 ## Sumrios por janela
169 comparacoes <- lapply(vars_j, \(g) sumario_grupo(V_mensal,
      TotaldeVitimas, .data[[g]]))
170 names(comparacoes) <- vars_j
171 comparacoes

```

```

172
173 ## t-tests (no pareados) por janela
174 t_tests_j <- lapply(vars_j, \(g)
175     t.test(V_mensal$TotaldeVitimas[V_mensal[[g]] == 0],
176           V_mensal$TotaldeVitimas[V_mensal[[g]] == 1],
177           var.equal = FALSE)
178 )
179 names(t_tests_j) <- vars_j
180
181 ## ARIMA por janela (mesma estrutura do model1)
182 fit_j <- lapply(vars_j, \(g)
183     Arima(V_mensal$TotaldeVitimas,
184           order = c(1,1,1),
185           seasonal = list(order = c(0,1,1), period = 12),
186           xreg = V_mensal[[g]])
187 )
188 list2env(setNames(fit_j, c("modelJ0", "modelJ3", "modelJ6")), envir = .
189           GlobalEnv)
190
191 ## Paired t entre observado e ajustado (ps-interveno)
192 paired_tests <- lapply(seq_along(vars_j), \(i) {
193     g <- vars_j[i]
194     idx <- which(V_mensal[[g]] == 1)
195     real <- V_mensal$TotaldeVitimas[idx]
196     prev <- fitted(fit_j[[i]])[idx]
197     t.test(real, prev, paired = TRUE)
198 })
199 names(paired_tests) <- vars_j
200
201 ## Linhas ajustadas s no ps para plot final
202 V_mensal$prev_juntos <- NA
203 V_mensal$prev_juntos3 <- NA
204 V_mensal$prev_juntos6 <- NA
205 V_mensal$prev_juntos [V_mensal$juntos == 1] <- fitted(fit_j[["juntos"
206   ]]) [V_mensal$juntos == 1]
207 V_mensal$prev_juntos3[V_mensal$juntos3 == 1] <- fitted(fit_j[["
208   juntos3"]]) [V_mensal$juntos3 == 1]
209 V_mensal$prev_juntos6[V_mensal$juntos6 == 1] <- fitted(fit_j[["
210   juntos6"]]) [V_mensal$juntos6 == 1]
211
212 ggplot(V_mensal, aes(x = MES)) +

```

```

209 geom_line(aes(y = TotaldeVitimas), color = "black", size = 0.8,
      linetype = "solid") +
210 geom_line(aes(y = prev_juntos), color = "darkred", size = 1) +
211 geom_line(aes(y = prev_juntos3), color = "darkorange", size = 1) +
212 geom_line(aes(y = prev_juntos6), color = "darkgreen", size = 1) +
213 geom_vline(xintercept = d_juntos, color = "red", linetype = "
      dotdash") +
214 geom_vline(xintercept = d_juntos3, color = "orange", linetype = "
      dotdash") +
215 geom_vline(xintercept = d_juntos6, color = "darkgreen", linetype = "
      dotdash") +
216 labs(title = "Comparao: Previses com Poltica e Realidade",
217       subtitle = "",
218       x = "Data", y = "Casos de Violncia Domstica") +
219 theme_minimal()

```

Código 3 – Código R para análise de com as variáveis exógenas