

**UNIVERSIDADE FEDERAL DE ITAJUBÁ - UNIFEI
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E
TECNOLOGIA DA COMPUTAÇÃO**

Caracterização temporal e espacial de roubos
de rua contrastando contextos pré-pandêmico
e pandêmico.

Renato Figueiredo Frade

Itabira
2023

**UNIVERSIDADE FEDERAL DE ITAJUBÁ - UNIFEI
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E
TECNOLOGIA DA COMPUTAÇÃO**

Renato Figueiredo Frade

**Caracterização temporal e espacial de roubos
de rua contrastando contextos pré-pandêmico
e pandêmico.**

Dissertação submetida ao Programa de Pós-Graduação em Ciência e Tecnologia da Computação como parte dos requisitos para obtenção do Título de Mestre em Ciência e Tecnologia da Computação.

**Área de concentração: Matemática da
Computação**

**Orientador: Prof. Dr. Carlos Henrique
da Silveira/UNIFEI**

**Coorientador: Prof. Dr. Eric Fernandes
de Mello Araújo/UFLA**

**Itabira
2023**

Renato Figueiredo Frade

Caracterização temporal e espacial de roubos de rua contrastando contextos pré-pandêmico e pandêmico

Dissertação submetida ao Programa de Pós-Graduação em Ciência e Tecnologia da Computação como parte dos requisitos para obtenção do Título de Mestre em Ciência e Tecnologia da Computação.

Itabira, 09 de dezembro de 2023

**Prof. Dr. Carlos Henrique da
Silveira/UNIFEI**
Orientador

**Prof. Dr. Eric Fernandes de Mello
Araújo/UFLA**
Coorientador

**Prof. Dr. Alexandre Carlos Brandão
Ramos**
Membro da banca - Unifei

Prof. Dr. Adriano Velasque Werhli
Membro da banca - FURG

Itabira
2023

Agradecimentos

Queridos amigos, familiares e colegas,

Hoje, com grande gratidão e satisfação, concluo a jornada da minha dissertação de mestrado. Esta realização não teria sido possível sem o apoio e contribuições inestimáveis de muitas pessoas que me acompanharam ao longo desta caminhada.

Em primeiro lugar, quero expressar minha profunda gratidão a Deus. Com fé, paciência e sabedoria, iluminou o caminho até este momento.

À minha amada esposa Izabela Frade, agradeço por seu apoio constante, paciência e motivação, que foram fundamentais para enfrentar os desafios deste projeto. Seu encorajamento me deu força nos momentos mais difíceis, e sua compreensão das longas horas de trabalho e estudo tornaram tudo possível. Você é a minha inspiração e a razão pela qual continuei avançando.

Ao meu orientador Carlos Silveira, expresso minha sincera gratidão por sua orientação sábia e insights valiosos ao longo deste processo. Suas orientações e conselhos moldaram meu trabalho de forma significativa, e sua dedicação em me ajudar a alcançar meus objetivos acadêmicos não tem preço. Sua influência será eternamente apreciada.

Ao meu coorientador Eric Araújo, devo um agradecimento por sua colaboração na obtenção dos dados da Polícia Militar de Minas Gerais, um elemento crucial para o sucesso deste projeto. Sua experiência e orientação foram essenciais para minha pesquisa.

Além disso, quero agradecer a todos os professores, em especial ao Prof. João Romanelli, pois ele nos ajudou bastante na parte da análise de frequências das séries temporais, e amigos e colegas que me apoiaram ao longo deste caminho. Suas palavras de encorajamento, discussões e contribuições foram igualmente valiosas para o desenvolvimento desta dissertação.

E, finalmente, meu agradecimento à Universidade Federal de Itajubá e toda equipe que a integra. Suas contribuições e suporte foram essenciais para tornar este projeto uma realidade.

Esta conquista é compartilhada com todos vocês, e estou profundamente grato por cada um de vocês ter feito parte desta jornada.

Com apreço, Renato Frade

Resumo

Esta pesquisa multidisciplinar realizada em Minas Gerais investigou a dinâmica temporal e espacial dos roubos de rua, analisando os períodos anterior e durante a pandemia. Através da utilização de dados da Polícia Militar, foram examinadas séries temporais em várias escalas, incluindo por hora, por dia, por intervalos de 10 dias e por mês, empregando métodos estatísticos avançados, como análise de frequência espectral, autocorrelações e técnicas de decomposição. Os resultados surpreenderam ao evidenciar uma queda média de 64% nos roubos durante a pandemia e a identificação de um regime estacionário nas ocorrências durante esse período, indicando uma maior aleatoriedade nos eventos. Contrariando expectativas, não foi possível encontrar sazonalidade nos roubos em relação aos dias da semana, início ou fim de mês ou meses do ano. Essas descobertas oferecem informações valiosas para a compreensão dos padrões temporais de crimes e contribuem para o desenvolvimento de políticas de segurança pública mais eficazes. Além disso, o estudo ressaltou a utilidade do geocodificador (Geo) na análise de roubos, permitindo a conversão de endereços em coordenadas geográficas, possibilitando assim a visualização da distribuição espacial dos crimes. A integração desses dados com informações geográficas adicionais e a criação de mapas interativos com o auxílio da biblioteca Leaflet e do pacote SF no ambiente R foram destacadas como ferramentas cruciais para identificar áreas de alta criminalidade, regiões com uma presença policial reduzida e planejar estratégias efetivas de segurança pública, sublinhando a relevância da análise de dados geográficos na tomada de decisões nesse campo.

Palavras-chaves: Dinâmica temporal, Roubos de rua, Pandemia, Análise geoespacial, Séries temporais, Segurança pública.

Abstract

A multidisciplinary study conducted in Minas Gerais investigated the temporal dynamics of street robberies, analyzing both pre-pandemic and pandemic periods. Utilizing data from the Military Police, time series data were examined at various scales, including hourly, daily, 10-day intervals, and monthly, employing advanced statistical methods such as spectral frequency analysis, autocorrelations, and decomposition techniques. The results were surprising, revealing an average decrease of 64% in robberies during the pandemic and identifying a stationary pattern in pandemic occurrences, indicating increased randomness in the events. Contrary to expectations, no seasonality was observed in robberies concerning weekdays, the beginning or end of the month, or months of the year. These findings offer valuable insights into understanding the temporal patterns of crimes and contribute to the development of more effective public safety policies. Additionally, the study underscored the usefulness of the geocoder (Geo) in robbery analysis, enabling the conversion of addresses into geographic coordinates, facilitating the visualization of the spatial distribution of crimes. The integration of this data with additional geographic information and the creation of interactive maps using the Leaflet library and the SF package in the R environment were highlighted as crucial tools for identifying high-crime areas, regions with reduced police presence, and planning effective public safety strategies, emphasizing the importance of geospatial data analysis in decision-making in this field.

Key-words: Temporal dynamics, Street robberies, Pandemic impact, Geospatial analysis, Time series data, Public safety strategies.

Lista de ilustrações

Figura 3.1 – Municípios vs N ^o de crimes	25
Figura 3.2 – Meios Utilizados vs N ^o de crimes	26
Figura 3.3 – Objeto do Roubo vs N ^o de crimes	27
Figura 3.4 – Perfis da TS (<i>Séries temporais ou Temporal Series</i>) presente na tabela NHOOR (<i>Número de Horas</i>)	28
Figura 3.5 – Perfis da análise espectral de frequências sobre a TS NHOOR	29
Figura 3.6 – Análise comparativa dos perfis pré-pandêmicos (azul) e pandêmicos (vermelho), agrupados por hora	31
Figura 3.7 – Gráfico de autocorrelações da TS NHOOR: simples (ACF (<i>Função de autocorrelação</i>)) e parciais (PACF (<i>Função de autocorrelação parcial</i>))	32
Figura 3.8 – Perfis da média de crimes consolidada por hora	34
Figura 3.9 – Perfis da TS NDAY (<i>Número de Dias</i>)	35
Figura 3.10–Perfis da análise espectral de frequências sobre a TS NDAY, em escala logarítmica em X e Y	36
Figura 3.11–Gráfico de autocorrelações da TS NDAY: simples (ACF) e parciais (PACF)	37
Figura 3.12–Análise comparativa dos perfis pré-pandêmicos (azul) e pandêmicos (vermelho), agrupados por dia da semana	38
Figura 3.13–Análise comparativa dos perfis total (preto), pré-pandêmico (azul) e pandêmico (vermelho), agrupados pela média de crimes por dia da semana	39
Figura 3.14–Perfis da TS NTEN (<i>Número de 10 em 10 dias</i>)	40
Figura 3.15–Perfis da análise espectral de frequências sobre a TS NTEN, em média de crimes a cada 10 dias, em escala logarítmica em X e Y	41
Figura 3.16–Análise comparativa dos perfis	42
Figura 3.17–Análise da decomposição STL (<i>Tendência Sazonal usando LOESS</i>) da TS NMON (<i>Número de meses</i>) (crimes por mês)	43
Figura 3.18–Análises gerais da série TS NMON considerando o período total, sub- traída das tendências.	45
Figura 3.19–Análise comparativa dos perfis consolidado por mês	46
Figura 3.20–Mapa de Calor de Crimes de Rua - Itabira	48

Lista de tabelas

Tabela 2.1 – Amostra de registros e alguns campos da tabela CRIMES CLN	19
Tabela 2.2 – Primeiros registros da tabela N HOUR	19
Tabela 3.1 – Consolidado por períodos da tabela N HOUR	26
Tabela B.1 – Dicionário de dados descrevendo em detalhes cada atributo	52
Tabela C.1 – Principais funções, pacotes do R e respectivos parâmetros utilizados. .	53

Lista de abreviaturas e siglas

<i>COVID-19</i>	<i>Coronavirus Disease 2019</i>	20
<i>IDE</i>	<i>Ambiente de desenvolvimento integrado</i>	24
<i>MG</i>	<i>Minas Gerais</i>	20
<i>NA</i>	<i>"Not Available" - Indisponível</i>	18
<i>SARS-Cov-2</i>	<i>Severe Acute Respiratory Syndrome Coronavirus 2</i>	20
<i>ACF</i>	<i>Função de autocorrelação</i>	7
<i>Animation</i>	<i>Pacote de animações para o RStudio</i>	23
<i>DTFT</i>	<i>Discrete-Time Fourier Transform</i>	20
<i>Facebook</i>	<i>Mídia Social</i>	17
<i>GIF</i>	<i>Formato de intercâmbio de gráficos</i>	51
<i>hs</i>	<i>Horas</i>	30
<i>Leaflet</i>	<i>Biblioteca em JavaScript</i>	23
<i>MTSA</i>	<i>Multiple Taper Spectral Analysis</i>	20
<i>NDAY</i>	<i>Número de Dias</i>	7
<i>NHOUR</i>	<i>Número de Horas</i>	7
<i>NMON</i>	<i>Número de messes</i>	7
<i>NTEN</i>	<i>Número de 10 em 10 dias</i>	7
<i>PACF</i>	<i>Função de autocorrelação parcial</i>	7
<i>PMMG</i>	<i>Polícia Militar do Estado de Minas Gerais</i>	16
<i>Prophet</i>	<i>Modelo de Regressão Aditiva com uma parte linear ou curva de tendência de Crescimento Logístico</i>	17
<i>PSD</i>	<i>Power Spectral Density</i>	20
<i>R</i>	<i>Linguagem de Programação</i>	24
<i>RStudio</i>	<i>RStudio é um software livre de ambiente de desenvolvimento integrado para R</i>	24
<i>sf</i>	<i>Biblioteca Geoespacial</i>	23
<i>STL</i>	<i>Tendência Sazonal usando LOESS</i>	7
<i>Tapers</i>	<i>Funções envolvendo janelas temporais</i>	21
<i>TS</i>	<i>Séries temporais ou Temporal Series</i>	7

Sumário

1	INTRODUÇÃO	13
1.1	Objetivo Geral	15
1.2	Objetivos Específicos	15
1.3	Trabalhos relacionados	16
2	METODOLOGIA	18
2.1	Conjunto de Dados (Data set)	18
2.2	Adequação de Dados	18
2.3	Geração das Séries Temporais	19
2.4	Pandemia de COVID-19 e Divisão de Períodos	20
2.5	Periodogramas	20
2.6	Funções de Autocorrelação	21
2.7	Decomposição	22
2.8	Geoprocessamento	23
2.9	Implementação	24
3	RESULTADOS E DISCUSSÕES	25
3.0.1	<i>Estatísticas Básicas</i>	25
3.1	Consolidado Por Período	26
3.2	Crimes como Série Temporal (Por Hora)	27
3.3	Crimes como Série Temporal (Por Dia)	33
3.4	Crimes como série temporal (10 dias do mês)	38
3.5	Crimes como série temporal (Por Mês)	43
3.6	Geoprocessamento	47
4	CONCLUSÃO	49
4.1	Melhorias futuras	49
	APÊNDICES	50
	APÊNDICE A – IMAGENS DOS MUNICÍPIOS DE MINAS GERAIS E GIFS DO MUNICÍPIO DE ITABIRA	51
	APÊNDICE B – DICIONÁRIO DE DADOS	52

APÊNDICE C – QUADRO COM AS PRINCIPAIS FUNÇÕES E PACOTES DO R	53
REFERÊNCIAS	54

1 Introdução

A criminologia é uma área interdisciplinar que se dedica ao estudo do crime, das causas do comportamento criminoso, das consequências sociais da criminalidade e das diferentes respostas que as sociedades organizam frente à ocorrência de delitos. Como campo de investigação, ela se apoia em conhecimentos da sociologia, psicologia, direito, antropologia e outras áreas das ciências humanas e sociais, com o objetivo de entender a dinâmica da ação delituosa e desenvolver tanto medidas preventivas quanto sistemas de justiça criminal mais eficazes e justos. Através da análise de dados, teorias sobre comportamento e motivação, e estudos de caso, a criminologia busca contribuir para a formulação de políticas públicas e estratégias de intervenção social para reduzir a criminalidade e melhorar a gestão da segurança pública. (MAGUIRE; MORGAN; REINER, 2012)

São muitas as hipóteses e teorias que tentam explicar o comportamento criminoso, aqui serão citadas apenas as mais relevantes. A Teoria das Janelas quebradas é uma teoria criminológica desenvolvida pelos criminologistas James Q. Wilson e George L. Kelling. Essa teoria sugere que a presença de sinais de desordem, como janelas quebradas, lixo nas ruas e grafites, em um ambiente urbano pode levar ao aumento da criminalidade e comportamentos anti-sociais (WILSON JAMES Q., 1982).

A fundamentação da Teoria das Janelas Quebradas baseia-se na ideia de que a desordem física e social em um ambiente transmite uma mensagem de que ninguém se importa e que a lei não está sendo aplicada de maneira eficaz. Essa percepção de falta de controle e supervisão encoraja a quebra de normas sociais e o surgimento de comportamentos criminosos. Ao permitir que sinais de desordem persistam sem serem corrigidos, cria-se um ambiente propício para a proliferação de crimes mais graves.

A Teoria das Janelas Quebradas também enfatiza a importância da prevenção e do policiamento comunitário. Ao abordar prontamente sinais de desordem e reparar janelas quebradas, limpar a sujeira das ruas e combater pequenos delitos, as comunidades podem enviar uma mensagem de que estão vigilantes e dispostas a manter a ordem. Isso, por sua vez, desencoraja a ocorrência de comportamentos criminosos e ajuda a fortalecer um senso de responsabilidade coletiva.

A Teoria da Oportunidade, desenvolvida por Marcus Felson e Lawrence Cohen, aborda a relação entre motivação criminal e oportunidades para a prática de crimes. Essa teoria sugere que o roubo e outros crimes ocorrem quando os potenciais infratores percebem uma oportunidade de obter benefícios com risco relativamente baixo (FELSON MARCUS; COHEN, 2008).

A Teoria da Oportunidade baseia-se na ideia de que os crimes são eventos que

ocorrem quando há uma convergência de três elementos principais: um alvo atraente, a ausência de um guarda ou proteção adequada e a presença de um potencial infrator motivado. Esses elementos juntos criam uma oportunidade para a ocorrência do crime.

A teoria argumenta que a prevenção do crime pode ser eficaz ao focar na redução das oportunidades para a prática criminal. Isso pode ser alcançado por meio do aumento do esforço necessário para cometer um crime, aumentando os riscos percebidos pelos infratores ou diminuindo as recompensas esperadas do crime.

Dentro desse contexto, as estratégias de prevenção do crime que se baseiam na Teoria da Oportunidade incluem o aumento da vigilância, o fortalecimento da segurança física e do ambiente, a implementação de sistemas de alarme e a melhoria da iluminação pública. Essas medidas visam desencorajar os potenciais infratores, aumentando os riscos e as dificuldades associadas à prática de crimes.

A Teoria da Escolha Racional, também conhecida como Teoria do Comportamento Deliberado, é uma abordagem criminológica que se baseia na premissa de que os indivíduos são tomadores de decisão racionais que pesam os custos e benefícios antes de cometer um crime. Essa teoria sugere que as pessoas escolhem se envolver em comportamentos criminosos quando percebem que os benefícios superam os custos associados ao ato criminoso (CORNISH DEREK B.; CLARKE, 2017).

Essa teoria está enraizada na ideia de que os indivíduos tomam decisões racionais com base em uma avaliação de custo-benefício. Eles consideram fatores como a probabilidade de serem pegos, as consequências legais e sociais do crime, as recompensas esperadas e as alternativas disponíveis. Sugere que, se os custos forem altos ou os benefícios forem baixos, os indivíduos terão menos probabilidade de cometer um crime.

A Teoria da Escolha Racional tem implicações importantes na prevenção do crime, pois enfatiza a importância de aumentar os custos ou reduzir os benefícios associados aos comportamentos criminosos. Isso pode ser feito por meio de estratégias como o aumento da probabilidade de punição, a melhoria da eficácia do sistema de justiça criminal, a promoção de oportunidades legítimas e o fortalecimento das redes de apoio social.

O crime é representado como a diferença entre o benefício e a punição, menos os custos, é uma simplificação de um modelo teórico que tem sido discutido no campo da criminologia e da economia do crime. Essa abordagem busca entender as decisões dos indivíduos de cometer ou não crimes com base em uma análise de custos e benefícios, (CORNISH D. B., 1986).

Para além do aspecto teórico, a análise empírica é outro elemento importante da criminologia. Nesse sentido, compreender a evolução temporal e espacial de crimes é essencial para desenvolver políticas públicas e estratégias de segurança eficazes, especialmente em momentos de crises sociais.

A pandemia de COVID-19 compreendeu uma dessas crises. Desencadeou medidas de isolamento social em todo o mundo, levantando preocupações sobre seus impactos na economia e segurança pública. No entanto, os resultados não foram conforme o previsto, destacando a necessidade de uma compreensão mais profunda dos fenômenos criminais em situações excepcionais.

A análise de séries temporais consiste em analisar um conjunto de dados em que as observações são realizadas ao longo do tempo, buscando identificar padrões e tendências que possam ser utilizados para prever o comportamento futuro. Dentre as diversas aplicações da análise de séries temporais na criminologia, destacam-se a caracterização e previsão de crimes, que podem ser úteis para planejar ações preventivas e reduzir a incidência de crimes.

Algumas análises de séries temporais revelaram que, após a implementação das medidas de distanciamento na pandemia, houve mudanças significativas nas tendências de crimes, com alguns tipos de crimes diminuindo e outros aumentando em relação aos anos anteriores. Por exemplo, os crimes de furto e roubo tiveram uma queda acentuada. Por outro lado, o estelionato e crimes cibernéticos tiveram aumento substancial. As ocorrências de drogas e violência doméstica (Lei Maria da Penha, no Brasil) também tiveram um aumento significativo. (LEAL1, 2022).

Este estudo investigou as mudanças nos roubos de rua antes e durante a pandemia em cidades mineiras, utilizando análises estatísticas avançadas em séries temporais criminais, além de análises georreferenciadas nas ocorrências espaciais. A escolha de crimes de roubos de rua foi uma decisão estratégica, dada a presunção da sua correlação com a mobilidade da população, por sua vez criticamente afetada pela pandemia. Os resultados revelam insights importantes sobre como a pandemia pode ter afetado esses crimes e contribuem para estratégias mais eficazes de segurança pública.

1.1 Objetivo Geral

Investigar as transformações nos roubos de rua ao longo do tempo e do espaço geográfico, comparando períodos pré-pandêmicos e pandêmicos, em um conjunto de cidades mineiras.

1.2 Objetivos Específicos

- Realizar a coleta e análise exploratória dos dados da PMMG de 2014 a 2022 sobre crimes de roubo de rua;
- Caracterização temporal de roubos de rua, no período pré-pandêmico e pandêmico.

- Realizar uma análise exploratória dos dados da PMMG (*Polícia Militar do Estado de Minas Gerais*), verificando a distribuição das variáveis e identificando possíveis valores extremos e valores ausentes;
- Transformar os dados brutos, a fim de gerar séries temporais, em diferentes escalas de tempo.
- Fazer uma análise espacial georreferenciada da distribuição de crimes de roubo de rua.
- Realizar a interpretação dos resultados e identificar as possíveis implicações para o desempenho da Polícia Militar em Minas Gerais;

1.3 Trabalhos relacionados

O artigo (C. EUGENIO C.; V., 2018) apresenta uma abordagem que utiliza análise espacial e modelos autorregressivos para detectar regiões de alto risco de crimes em áreas urbanas e prever com precisão as tendências criminais nessas regiões. Os resultados deste estudo indicam que essa abordagem é capaz de fornecer previsões espaciais e temporais precisas sobre a criminalidade.

Em (BOROWIK; CICHOSZ, 2019) destaca-se a importância do desenvolvimento tecnológico na análise de crimes e enfatiza a utilidade de algoritmos analíticos na previsão de crimes. Ele discute a aplicação de modelos como o Generalized Additive Model (GAM), regressão não linear e o modelo de previsão Prophet para a detecção e previsão de crimes. Além disso, o artigo ressalta a importância da precisão dessas previsões para a redução eficaz da criminalidade e a segurança dos cidadãos.

Já (LEAL1, 2022) analisou o impacto das medidas de distanciamento social na cidade de Curitiba durante a pandemia de COVID-19. Utilizando análise de séries temporais e regressão, os pesquisadores investigaram como essas medidas afetaram a ocorrência de diferentes tipos de crimes na cidade. Os resultados mostraram que vários tipos de crime, foram estatisticamente afetadas pelas medidas de distanciamento social.

A análise de regressão mostrou que o índice de rigidez das medidas de distanciamento e o índice de permanência residencial tiveram impactos significativos nas taxas de crimes. Em particular, medidas mais rígidas e uma maior permanência em casa da população foram associadas a reduções nos crimes de furto e roubo. No entanto, os resultados também indicaram que, em alguns casos, o aumento de crimes, como o estelionato, não estava diretamente relacionado às medidas de distanciamento, mas sim a uma possível migração de criminosos para outros tipos de crime durante a pandemia.

O estudo (HOU ZHAOLONG ZENG, 2022) investigou o impacto da pandemia de COVID-19 na incidência de crimes em quatro grandes cidades dos Estados Unidos:

Washington DC, Chicago, Nova York e Los Angeles. O estudo examinou as diferenças nos números de incidentes criminais antes e durante a pandemia, bem como as relações causais de Granger entre os números de incidentes criminais e os novos casos de COVID-19. Os resultados indicaram que a pandemia afetou a criminalidade de maneira geral, mas os efeitos variam entre cidades e tipos de crimes.

A maioria dos tipos de crimes registrou uma diminuição no número de incidentes durante a pandemia, com destaque para o roubo e o furto. Além disso, foram identificadas correlações causais entre os casos de COVID-19 e incidentes criminais em várias cidades, indicando que o aumento de casos de COVID-19 influenciou a criminalidade.

Para melhorar a previsão de crimes, os novos casos de COVID-19 com correlações causais significativas foram incorporados em modelos de previsão de crimes, utilizando a rede neural Long Short-Term Memory (LSTM). Os resultados sugerem que a inclusão dos novos casos de COVID-19 como variável pode melhorar modestamente o desempenho da previsão de alguns tipos específicos de crimes.

Em um estudo realizado por (SJ, 2018), a técnica de previsão de séries temporais Prophet (*Modelo de Regressão Aditiva com uma parte linear ou curva de tendência de Crescimento Logístico*) foi utilizada para prever a ocorrência de roubos em uma região específica do Canadá. Os resultados indicaram que o modelo foi capaz de prever a ocorrência de roubos com uma precisão superior a outros métodos de previsão. Prophet é uma ferramenta de previsão de séries temporais de código aberto desenvolvida pelo Facebook (*Mídia Social*). Essa ferramenta implementa um modelo aditivo de decomposição de séries temporais, que considera tendências, sazonalidade e feriados para fazer previsões precisas. O Prophet tem sido amplamente utilizado em diversas áreas, incluindo finanças, energia e saúde, para prever valores futuros de séries temporais.

Outro estudo realizado por (TOMA A., 2017) utilizou a análise de séries temporais para prever a ocorrência de furtos em uma cidade da Romênia. Os resultados mostraram que a técnica de análise de séries temporais foi capaz de prever a ocorrência de furtos com uma precisão de mais de 80%.

2 Metodologia

2.1 Conjunto de Dados (Data set)

A fonte de dados adveio da parceria entre a PMMG (Polícia Militar de Minas Gerais) e o coorientador Prof. Dr. Eric Araújo, especialista em modelagem cognitiva, informática comportamental e criminologia da UFLA (Universidade Federal de Lavras), adquirida em meados de julho de 2022 referindo-se à crimes do tipo roubo “Art. 157 - Subtrair coisa móvel alheia, para si ou para outrem, mediante grave ameaça ou violência à pessoa, ou depois de havê-la, por qualquer meio, reduzido à impossibilidade de resistência” (ABNTEX, 2016) entre o ano de 2014 e 2022 em diferentes municípios (Alfenas, Itabira, Itajubá, João Monlevade, Lavras, Ouro Preto, São João Del Rei, Viçosa).

A tabela principal contendo os dados brutos continha 9274 casos e 36 atributos, e foi chamada de CRIMES RAW. Um dicionário de dados descrevendo em detalhes cada atributo pode ser visto na tabela B.1 do Apêndice B.

Após a entrada da tabela com dados brutos, foi feita uma primeira análise cuidadosa dos atributos. Constatou-se que havia vários atributos temporais relacionados a ano, mês, semana, data e hora, tanto em composição como desmembrados. Apenas dois atributos eram contínuos: latitude e longitude. Todos os demais eram categóricos.

2.2 Adequação de Dados

Foi empregado um conjunto de operações ou transformações necessárias para tornar os dados aderentes ao escopo delimitado e mais adequados à visualização e modelagem. Entre essas transformações destacam-se a renomeação de atributos para nomes mais curtos (como “Tipo Boletim Ocorrências” para “BOL”) e a deleção de atributos com valores únicos (como aqueles relacionados à natureza do crime, todos do tipo ROUBO) ou considerados de pouca importância, como “Data Hora Comunicação Fato”.

Constatou-se também a presença de 910 registros com valores ausentes ou *NA* (“*Not Available*” - *Indisponível*), nos campos envolvendo DATA e HORÁRIO. Mas, como havia outros campos redundantes com informações similares, o campo DATA foi recomposto a partir de ANO, DIA NUMÉRICO, MÊS NUMÉRICO, e o campo HORA a partir da média de FAIXA 1 HORA. Uma vez recomposto DATA e HORÁRIO, foram reconstruídos também FAIXA 6 HORAS FATO e DEZENA MÊS FATO. Cabe destacar que todos esses registros eram da cidade de Lavras.

Ao final do processo de adequação de dados, a tabela CRIMES CLN terminou

Tabela 2.1 – Amostra de registros e alguns campos da tabela CRIMES CLN

TABELA CRIMES_CLN												
ID	DATA	HORA	MUNICIPIOS	ANO	SEM	FIM	MEI	BEN	LOC	VIA	CAU	...
1	2014-01-01	04:00:00	ITAJUBA	2014	QUA	POS	INDEFINIDO	TRANSEUNTE	VIA_PUB	URBANA	VANTAGEM	...
2	2014-01-01	05:45:00	JOÃO MONLEVADE	2014	QUA	POS	ARMA_BRANCA	TRANSEUNTE	VIA_PUB	URBANA	VANTAGEM	...
3	2014-01-01	21:20:00	LAVRAS	2014	QUA	POS	ARMA_BRANCA	INDEFINIDO	VIA_PUB	URBANA	INDEFINIDO	...
4	2014-01-02	16:40:00	ALFENAS	2014	QUI	POS	ARMAS_FOGO	TRANSEUNTE	VIA_PUB	URBANA	INDEFINIDO	...
5	2014-01-03	02:00:00	VIÇOSA	2014	SEX	POS	AMEAÇA	TRANSEUNTE	VIA_PUB	URBANA	VANTAGEM	...
6	2014-01-03	14:30:00	ITABIRA	2014	SEX	POS	ARMAS_FOGO	TRANSEUNTE	VIA_PUB	URBANA	VANTAGEM	...
7	2014-01-04	02:09:00	LAVRAS	2014	SAB	POS	AGRESSAO	TRANSEUNTE	VIA_PUB	URBANA	INDEFINIDO	...
8	2014-01-04	12:20:00	ITABIRA	2014	SAB	POS	ARMA_BRANCA	PESS_JUR	VIA_PUB	URBANA	VANTAGEM	...
9	2014-01-05	15:00:00	ALFENAS	2014	DOM	POS	ARMA_BRANCA	MOTOR_PART	VIA_PUB	ROD_EST	GANGUE	...
10	2014-01-06	01:10:00	ALFENAS	2014	SEG	POS	ARMA_BRANCA	VEÍCULOS	VIA_PUB	URBANA	INDEFINIDO	...
...

Tabela 2.2 – Primeiros registros da tabela NHOOR

TABELA NHOOR	
DATA_TIME	n
2014-01-01 00:00:00	0
2014-01-01 01:00:00	0
2014-01-01 02:00:00	0
2014-01-01 03:00:00	0
2014-01-01 04:00:00	1
2014-01-01 05:00:00	1
2014-01-01 06:00:00	0
2014-01-01 07:00:00	0
2014-01-01 08:00:00	0
2014-01-01 09:00:00	0

com 9274 registros e 27 atributos. Uma pequena amostra dessa tabela pode ser vista na tabela 2.1.

2.3 Geração das Séries Temporais

A partir da tabela CRIMES CLN, foram criadas tabelas acessórias, especificamente para as TS). São tabelas simples, com alguns campos, como DATA-HORA e contagens de crimes por tempo. Entre elas está:

- NHOOR: tabela de contagem de crimes por hora. Os primeiros registros da TS NHOOR podem ser vistos na tabela 2.2. Nessa tabela, a linha 2014-01-01 04:00:00 com $n = 1$, por exemplo, indica que foi registrado um crime ocorrido entre 04hs e 05hs da madrugada.

Outras dessas tabelas em diferentes escalas temporais serão descritas mais adiante ao longo deste trabalho.

Para que os dados pudessem ser devidamente analisados como uma TS, os registros tiveram que ser adequados para terem uma taxa de amostragem uniforme. No mundo real, ainda que possam ocorrer sazonalidades, os crimes não ocorrem regularmente no

tempo. Assim, a tabela original registrava a DATA e HORA que o crime ocorreu, podendo haver diferenças variáveis de tempo entre as ocorrências. Para adequar os registros a uma amostragem uniforme, foram inseridos registros com contagem zero *zero padding* nas tabelas TS, tal como exemplificado na tabela 2.2, para uma periodicidade por hora.

2.4 Pandemia de COVID-19 e Divisão de Períodos

Os dados de crimes disponíveis nesta análise cobriram uma parte da pandemia de *COVID-19* (*Coronavirus Disease 2019*) que tomou conta do planeta a partir de 2020, provocada pelo vírus *SARS-Cov-2* (*Severe Acute Respiratory Syndrome Coronavirus 2*) (WHO, 2019). A Organização Mundial da Saúde (OMS) declarou emergência de saúde pública de importância internacional em 30 de janeiro de 2020, e reconheceu oficialmente como uma pandemia em 11 de março de 2020. O Brasil confirmou seu primeiro caso de *COVID-19* em 26 de fevereiro de 2020 (LOBO, 2020), e o Ministério da Saúde declarou estado de transmissão comunitária em 20 de março de 2020 (VALENT, J., 2020). Boa parte das cidades em *MG* (*Minas Gerais*) entraram em confinamento seletivo ou *lockdown* a partir dessa data, como Itabira, pelo decreto Decreto Municipal 3.178/20, emitido na noite de 19 de março de 2020 (DECRETO NO 3.178, 2020). Portanto, para efeito de análises da influência da pandemia sobre os dados de crimes, neste trabalho será considerada a data 20 de março de 2020. Foi considerado também um período pré-pandêmico com mesmo número de dias do período pandêmico, conforme pode ser visto na Tabela 3.1

2.5 Periodogramas

Uma forma recorrente na literatura de sondar padrões em TS é através de periodogramas (MOSES, 2005). Periodograma é uma técnica utilizada para identificar componentes periódicos em TS. Mais especificamente, ele estima a PSD (*Power Spectral Density*) média desses sinais, que no caso discreto pode ser calculada por técnicas envolvendo DTFT (*Discrete-Time Fourier Transform*). Assim, a PSD consegue identificar as frequências com maiores potências ou forças (power) médias, grosso modo, aquelas com maior destaque na constituição de uma TS.

No entanto, é sabido que métodos clássicos de estimação da PSD média por transformadas de Fourier apresentam sérias limitações, entre elas: exige o pressuposto de uma TS estacionária; sensibilidade a ruídos; a PSD estimada pode apresentar viés, uma vez que sua variância não fica assintoticamente a zero quando o número de amostras cresce; mesmo com um viés não acentuado, a PSD estimada pode apresentar alta variância (MOSES, 2005).

Uma alternativa para mitigar tais limitações é o uso da técnica MTSA (*Multiple*

Taper Spectral Analysis) (SIDORENKO, 1995). Numa MTSA, múltiplas funções Tapers (*Funções envolvendo janelas temporais*) são aplicadas ao sinal original antes de serem transformadas de Fourier. Esse pré-tratamento por funções Tapers têm por objetivo operar uma composição de pesos sobre o sinal de entrada, de modo a dar mais ênfase aos padrões de frequência que possam ser mais relevantes, e assim conter melhor os ruídos, diminuir o viés e variância da PSD média que será estimada. E por operar por janelas, o pressuposto de estacionaridade é enfraquecido.

Para as estimativas do intervalo de confiança das PSD médias no MTSA, foi adotado o método de reamostragem jackknife (EFRON, B., 1980). Esse método omite uma observação em N amostragens na estimativa de parâmetros, sendo N o número de observáveis. O jackknife é o método de reamostragem que vem junto com o pacote R utilizado no cálculo do MTSA.

2.6 Funções de Autocorrelação

Dado um processo estocástico (randômico) representado pela realização de uma variável aleatória X_t no tempo t , autocorrelação diz da correlação cruzada de seus valores consigo mesmos. Ou ainda, das correlações (geralmente Pearson) entre os valores no tempo t , e t mais um deslocamento temporal $h(t + h)$, para qualquer inteiro t e h .

Neste trabalho, fez-se uso das seguintes funções de autocorrelação, como definidas em (SHUMWAY, 2016)

AUTOCORRELATIONFUNCTION(ACF)

$$ACF(X, t, h) = Cor(X_t, X_{t+h})$$

Ou seja, ACF verifica as correlações par a par na TS nos tempos t e $t + h$.

PARTIALAUTOCORRELATIONFUNCTION(PACF)

$$PACF(X, t, h) = Cor(X_t - \widehat{X}_t, X_{t+h} - \widehat{X}_{t+h})$$

Onde:

$$\widehat{X}_t = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \beta_3 X_{t+3} + \dots + \beta_{h-1} X_{t+h-1}$$

$$\widehat{X}_{t+h} = \beta_1 X_{t+h-1} + \beta_2 X_{t+h-2} + \beta_3 X_{t+h-3} + \dots + \beta_{h-1} X_{t+1}$$

Ou seja, PACF elimina as dependências lineares que podem haver entre X_t e X_{t+h} , computando uma correlação mais “pura” entre ambos.

2.7 Decomposição

Séries temporais podem apresentar alguns padrões intrínsecos interessantes, tais como: tendências, ciclos e sazonalidades. Tendências indicam as variações de mais longo prazo, tais como crescimentos e decaimentos. Já ciclos e sazonalidades indicam flutuações de curto prazo, sendo a diferença basicamente que nos ciclos as flutuações são aperiódicas, enquanto nas sazonalidades elas ocorrem em frequências/períodos fixos. Existem vários métodos capazes de decompor uma TS conforme esses padrões, sendo o mais comum a decomposição em: Tendências-Ciclos (T), Sazonalidades (S) e o Restante-Ruído, tal qual na fórmula (HYNDMAN R.J.; ATHANASOPOULOS, 2021):

$$y_t = T_t + S_t + R_t$$

É possível medir a força ou importância (strength) das componentes T e S pelas equações:

$$F_t = \max\left(0, 1 - \frac{\text{var}(R_t)}{\text{var}(T_t + R_t)}\right)$$

$$F_s = \max\left(0, 1 - \frac{\text{var}(R_t)}{\text{var}(S_t + R_t)}\right)$$

Claramente, se a contribuição da componente T for pequena, $\text{var}(R)$ e $\text{var}(T+R)$ terão valores muito próximos, e F_t tenderá a 0; caso contrário, estará perto de 1. Da mesma forma, se a contribuição da sazonalidade for pequena, $\text{var}(R)$ e $\text{var}(S+R)$ terão valores similares, e F_s tenderá a 0; caso contrário, perto de 1. Assim, quanto mais próximos F_t e F_s ficarem de 1, maiores suas contribuições; e perto de 0, o oposto.

Nesta dissertação, será utilizado o método de decomposição STL, um acrônimo para *Seasonal and Trend decomposition using LOESS*, em que LOESS (*Locally Estimated Scatterplot Smoothing*) é um método de regressão não linear, não paramétrico, a partir de interpolações polinomiais locais em segmentos de dados (CLEVELAND ROBERT B., 1990). A decomposição STL apresenta várias vantagens, entre elas (HYNDMAN R.J.; ATHANASOPOULOS, 2021):

- É capaz de lidar com qualquer tipo de sazonalidade, independente da unidade temporal;
- Permite que a componente sazonal seja variável no tempo;

- Há parâmetros para controlar a suavidade (smoothness) da componente tendência-ciclo;
- Pode ser robusta a valores extremos (outliers).

2.8 Geoprocessamento

A metodologia para gerar o mapa de calor com base nos dados de roubos em uma determinada cidade consiste em uma série de etapas. Primeiramente, é necessário coletar os dados de roubos, obtendo informações sobre as datas, coordenadas geográficas e outros atributos relevantes dos incidentes. Em seguida, esses dados são preparados e carregados em um ambiente de programação, como o R, onde são convertidos em um formato adequado para análise geoespacial, utilizando a biblioteca *sf* (*Biblioteca Geoespacial*).

Uma vez carregados e formatados, os dados são ordenados cronologicamente para garantir a correta representação temporal dos roubos. Caso sejam necessárias análises específicas para uma determinada cidade, os dados podem ser filtrados com base em um atributo que identifique a cidade ou região geográfica de interesse.

Para centralizar o mapa na cidade de interesse, são obtidas as coordenadas geográficas médias dos roubos registrados nessa cidade. Essas coordenadas são utilizadas para configurar a visualização do mapa, determinando o zoom e a posição inicial.

A criação dos mapas de calor é realizada por meio de um processo iterativo. Utilizando laços de repetição, como o `for`, os dados são agrupados por períodos específicos, como dias, meses ou anos. Para cada período, um mapa de calor é gerado utilizando a biblioteca *Leaflet* (*Biblioteca em JavaScript*). Nesse processo, são definidos parâmetros como o raio, o desfoque e a intensidade do mapa de calor, a fim de representar adequadamente a distribuição espacial e a intensidade dos roubos.

A inclusão de legendas, rótulos e outras informações relevantes é opcional, mas pode ser feita para facilitar a compreensão do mapa. Isso pode incluir a adição de legendas indicando o período representado (dia, mês, ano) ou outras informações pertinentes aos dados.

Por fim, os mapas de calor gerados para cada período são armazenados em uma lista. Utilizando pacotes específicos, como o *Animation* (*Pacote de animações para o RStudio*), é possível criar um GIF animado a partir dessa lista de mapas, permitindo visualizar a evolução dos roubos ao longo do tempo. O GIF animado resultante pode ser salvo em um arquivo para ser compartilhado e visualizado posteriormente (SILVA F., 2021).

2.9 Implementação

A ciência dos dados desenvolvida neste trabalho foi feita em R (*Linguagem de Programação*) [R version 4.2.2 Patched (2022-11-10 r83330)], com a *IDE* (*Ambiente de desenvolvimento integrado*) RStudio (*RStudio é um software livre de ambiente de desenvolvimento integrado para R*) [2023.06.0 Build 421]. Detalhes das funções, parametrizações e pacotes utilizados encontram-se no Quadro C.1 do Apêndice C.

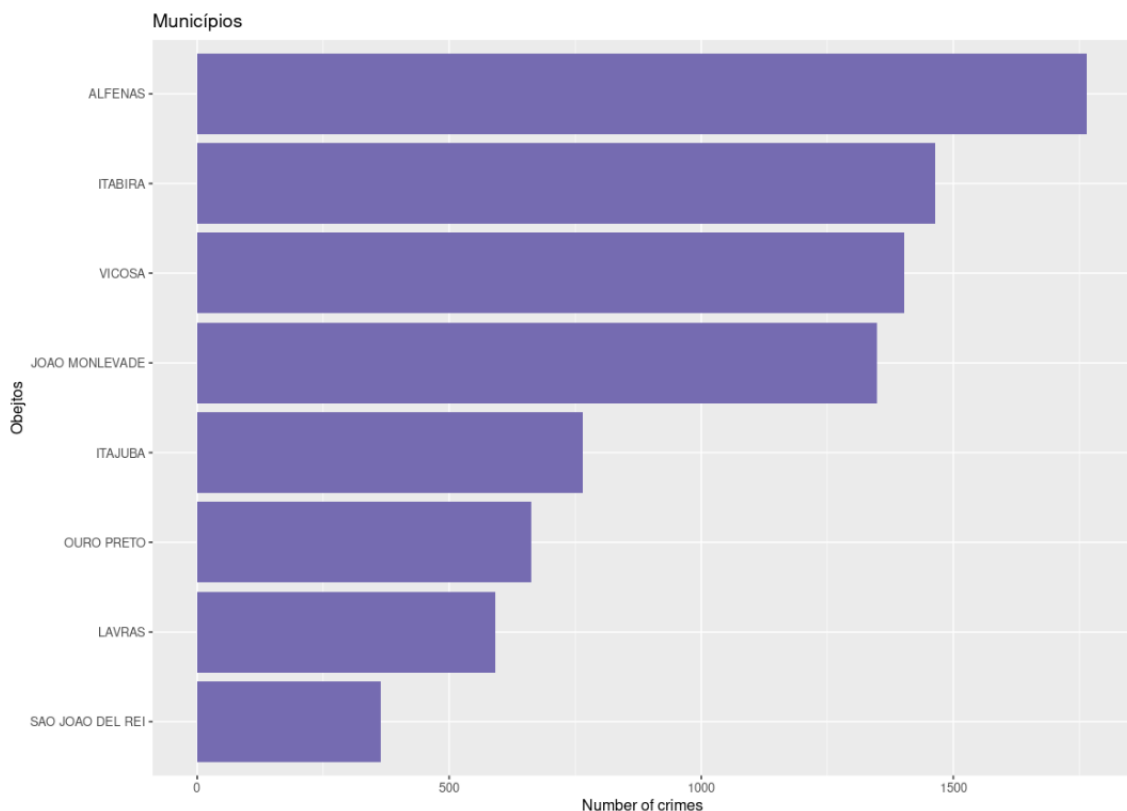
3 Resultados e Discussões

Nesta seção serão exibidos os resultados obtidos durante a implementação do projeto proposto. Resultados e discussões serão apresentados de forma integrada, para uma melhor compreensão e análise.

3.0.1 *Estatísticas Básicas*

Os dados dos boletins de ocorrência possuem diferentes municípios que foram utilizados neste estudo, a quantidade referente a cada município pode ser vista na figura 3.1.

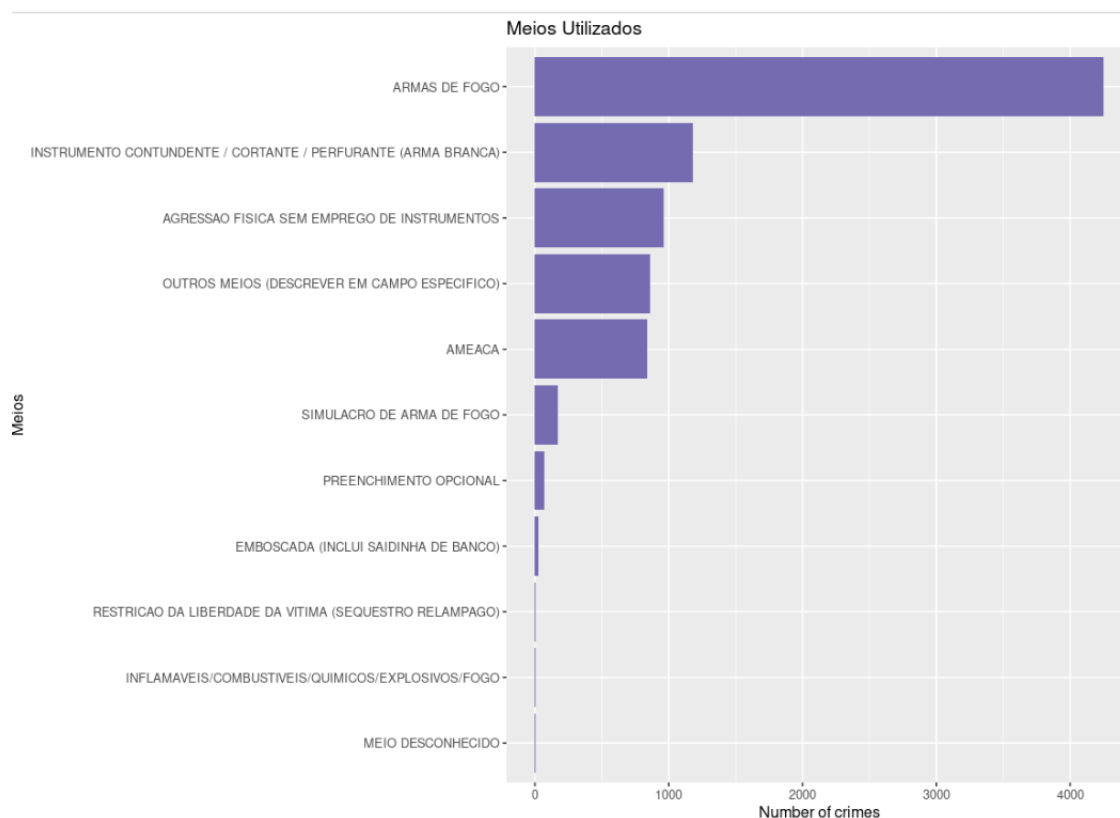
Figura 3.1 – Municípios vs N^o de crimes



Fonte: Autoria própria

A imagem 3.2 representa quais são os meios mais utilizados pelos criminosos para o cometimento do crime de roubo, pode-se assim observar que a arma de fogo é o objeto mais utilizado para esse tipo de prática, tendo uma colocação bem acima do segundo objeto armas brancas.

Dos objetos o que possui o maior índice de incidência são os "bens/valores dos transeuntes", assim o bem material mais visado pelos criminosos são os itens que os

Figura 3.2 – Meios Utilizados vs N^o de crimes

Fonte: Autoria própria

Tabela 3.1 – Consolidado por períodos da tabela N HOUR

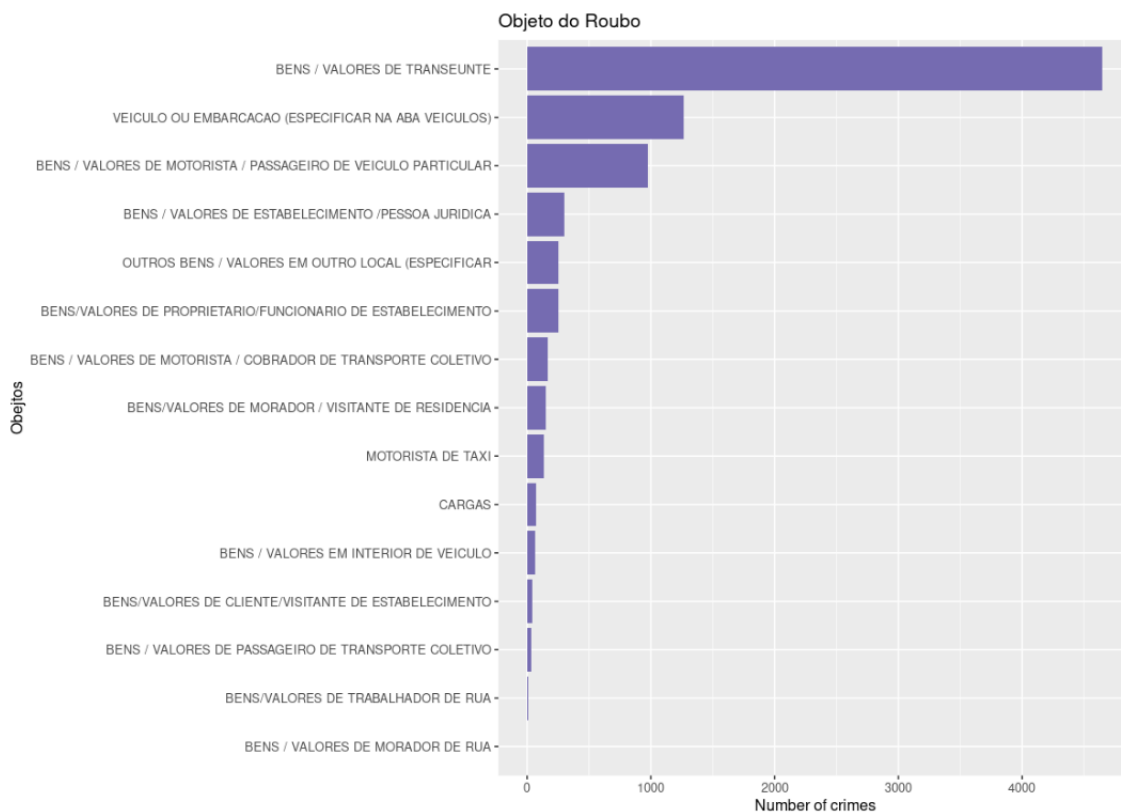
PERÍODO	INTERVALO	HORAS	DIAS	ANOS	CRIMES	HORAS/CRIMES	DIAS/CRIME
TOTAL	2014-01-01 - 2022-03-29	72240	3009	8.24	9274	8	0.32
PRÉ-PANDÊMICO	2018-03-11 - 2020-03-19	17760	739	2.02	1898	9	0.39
PANDÊMICO	2020-03-20 - 2022-03-29	17760	739	2.02	678	26	1.09

transeuntes carregam consigo, como pode ser comparado na figura 3.3.

3.1 Consolidado Por Período

A tabela 3.1 mostra um consolidado por período da tabela N HOUR, conforme os períodos Total, Pré-pandêmico e Pandêmico. A TS N HOUR, no total, cobriu um período de pouco mais de 8 anos e 2 meses, totalizando 9274 crimes, com média de 1 crime a cada 8 horas. O período Pré-pandêmico e Pandêmico compreendem o mesmo número de dias, abrangendo cerca de 2 anos de cobertura, totalizando 1903 e 678 ocorrências, respectivamente. Houve uma queda de aproximadamente 64% (quase 3 vezes menor) no número de crimes entre esses períodos, passando da média de 1 crime a cada 9 horas para 1 crime a cada 26 horas.

Figura 3.3 – Objeto do Roubo vs N^o de crimes



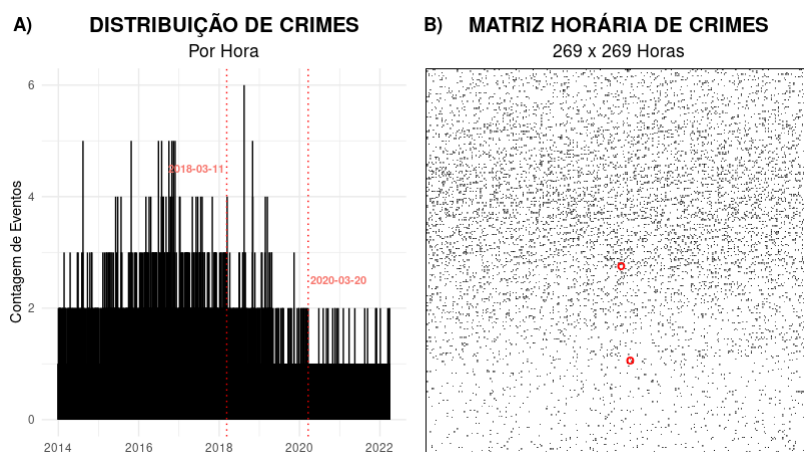
Fonte: Autoria própria

3.2 Crimes como Série Temporal (Por Hora)

A primeira TS analisada foi a NHOOR, que registra crimes por hora. Constitui a TS com a menor granularidade neste estudo, ou seja, que geraram as mais altas frequências de ocorrências.

A figura 3.4 mostra o perfil da TS NHOOR em duas visualizações distintas. Em 3.4-A vê-se a TS como um vetor de contagem de eventos indexados no tempo. O primeiro registro ocorreu em 2014-01-01 00:00:00 e o último em 2022-03-29 17:00:00, contabilizando 72240 posições de horas, para um total de 9274 ocorrências de crimes. Percebe-se alguns padrões: parece que o perfil segue uma função discreta côncava, que cresce a partir de 2014, atinge um modal entre 2016 e 2017, e decresce a partir de 2018, estabilizando-se de 2020 em diante. Contribui para essa estabilidade uma mudança no padrão de ocorrências por volta da data assumida neste trabalho como o início da pandemia: 2020-03-20.

Figura 3.4 – Perfis da TS presente na tabela N HOUR. A) contagem de ocorrências ao longo das horas começando em 2014-01-01 00:00:00 e terminando em 2022-03-29 17:00:00, conforme vetor com 72240 posições de horas, para um total de 9274 ocorrências de crimes. A linha pontilhada indica o início dos períodos pré-pandêmico e pandêmico em 2018-03-11 e 2020-03-20, respectivamente. B) Para melhor visualização da esparsidade, as ocorrências em A) foram convertidas numa matriz binária de 269 x 269 dimensões, começando no canto superior esquerdo, com varredura à direita e descendo. Essa binarização apenas registrava se houve alguma ocorrência naquela hora ou não. Os pontos vermelhos demarcam os inícios dos períodos pré-pandêmicos e pandêmicos.



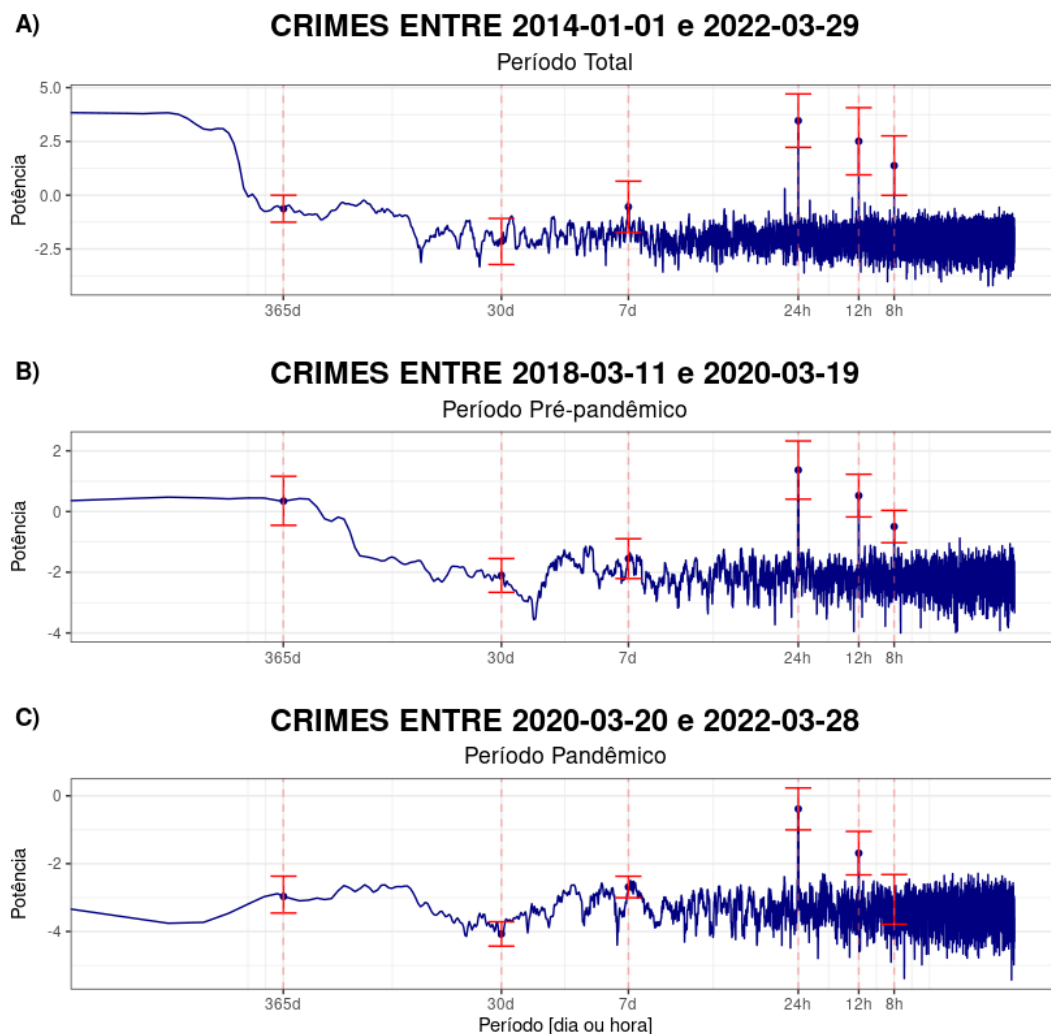
Fonte: Autoria própria

A figura 3.4-B é uma forma de visualizar o carácter da esparsidade das ocorrências de crimes no período estudado. A TS N HOUR foi dobrada numa matriz quadrada de 269 x 269 horas, com as posições de horas começando na célula superior à esquerda, seguindo à direita e abaixo, em zigue-zague até fechar a matriz quadrada. A diferença entre $269 \times 269 = 72361$ e 72240 foi preenchida com zeros ao final. Os pontos vermelhos demarcam o início do período pré-pandêmico e pandêmico. Percebe-se que os eventos de crimes ocorrem de forma esparsa ao longo das horas. Das 72240 horas da N HOUR, em apenas 7777 ocorreram crimes, algo perto de 11%. Ou seja, quase 90% da matriz é composta de zeros. Mas, essa esparsidade não é homogênea. Os pontos ganham intensidade quando se aproximam da região central da matriz (2017), mas ficam bem mais esparsos no limite inferior (a partir de 2018), principalmente após o marco da pandemia (início de 2020). Notam-se também alguns padrões diagonais riscando a matriz, um primeiro indício de que as ocorrências não tenham sido totalmente aleatórias.

A figura 3.5-B apresenta o resultado da análise espectral de frequências por MTSA aplicada sobre a TS N HOUR num gráfico em escala logarítmica em X e Y. Conforme explanado anteriormente, o MTSA é considerado um método de estimativa da densidade espectral não enviesado, que pode ser aplicado à TS não estacionárias.

Na figura 3.5-A podemos ver o perfil do espectro de frequência para toda a TS N HOUR. Percebe-se picos relevantes indicando sazonalidade nas frequências envolvendo os períodos de 12 e 24 horas. No período de 8 horas, o intervalo de confiança deixa sua

Figura 3.5 – Perfis da análise espectral de frequências sobre a TS NHOOR, em escala logarítmica em X e Y. O eixo X representa as frequências, mas os rótulos foram trocados por períodos, para melhor interpretação. Também para maior clareza, os intervalos de confiança no nível de 0.99 foram estimados apenas para as frequências/períodos em destaque, pelo método jackknife. A) perfil espectral das frequências para toda a TS NHOOR. B) perfil espectral do período pré-pandêmico, considerando o mesmo número de dias que o período pandêmico. C) perfil espectral das frequências no período pandêmico.



Fonte: Autoria própria

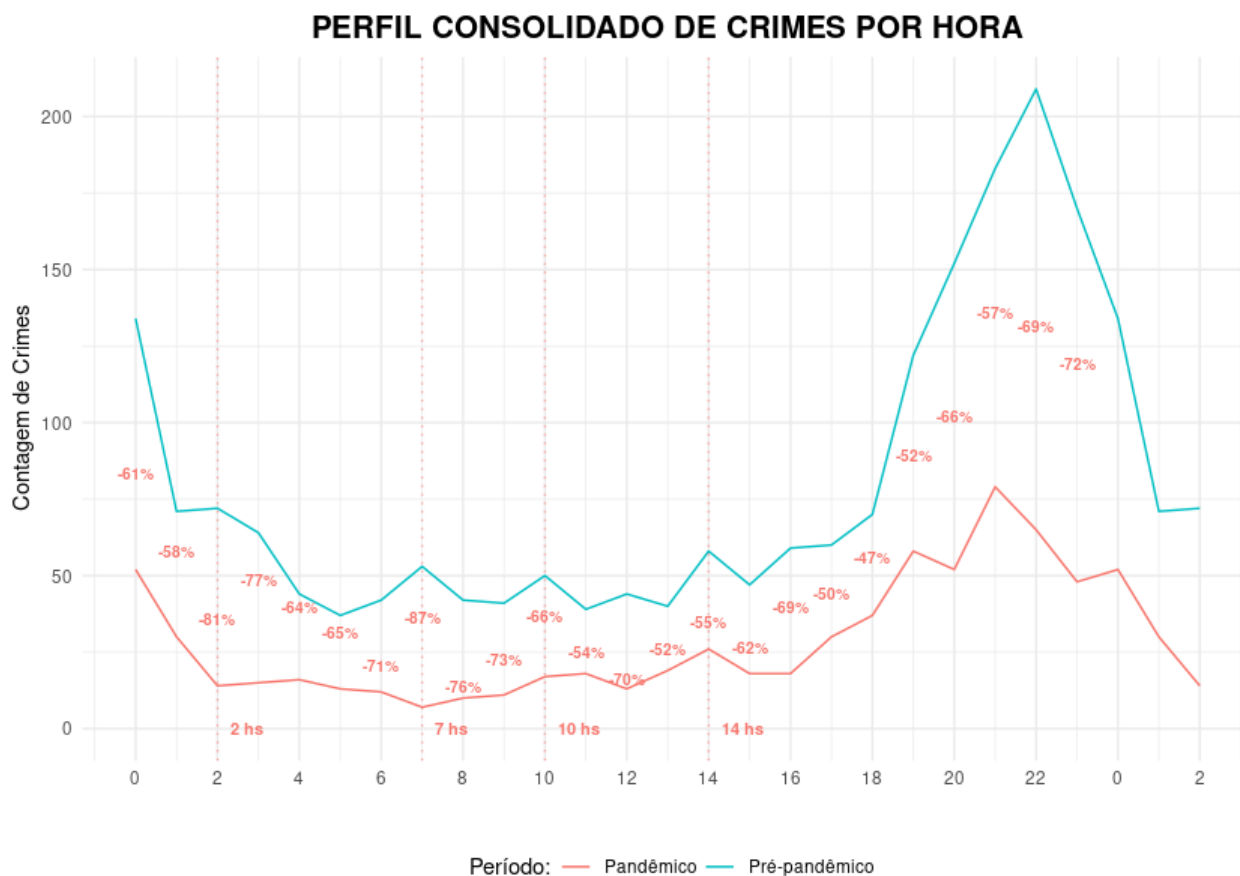
relevância em dúvida. Há um pequeno pico em 7 dias, mas a margem de erro inferior avança na região das potências sem destaques. Nota-se também uma elevação ou excesso de potência nas regiões de baixas frequências e períodos mais longos, após 365 dias. Essa elevação pode indicar um comportamento não estacionário nos dados (RAHIM, 2014), em decorrência, principalmente, da presença de tendências (trendings) ou oscilações nas médias dos valores em função do tempo t . De fato, conforme pôde ser observado na figura 3.4-A, a distribuição de crimes no tempo segue um perfil não estacionário em relação às médias das ocorrências, com tendência clara de subida e queda, com modal entre 2017 e 2018, e posterior estabilização num patamar menor. Essa discussão será retomada mais adiante.

A figura 3.5-B analisa o período pré-pandêmico, considerando o mesmo número de horas que o período pandêmico. Revela-se um padrão similar ao período geral, com um destaque menor para a oscilação de 7 dias. Também nota-se uma diminuição do excesso de potência nas baixas frequências.

Já na figura 3.5-C, na parte pandêmica da TS, não se verifica esse excesso de potência nas baixas frequências, em períodos após 365 dias. Tal comportamento sugere que as ocorrências de crimes podem ter adquirido um padrão estacionário ou pelo menos de tendência de médias de ocorrências durante a pandemia. Outro curioso padrão desse período é o enfraquecimento das frequências em 8 hs (*Horas*) e 12 hs, principalmente a primeira, que passa a apresentar um intervalo de confiança indistinto das frequências sem destaques. A tênue força que existia antes na frequência de 7 dias, agora se torna mais indistinta ainda. O único padrão periódico que parece não ter sido afetado pela pandemia é o de 24 horas.

Na figura 3.6, temos outra visualização que ajuda a interpretar os resultados dos periodogramas, com foco na análise comparativa entre o período pré-pandêmico (azul) e pandêmico (vermelho), ambos com mesmo número de dias. Nela, vê-se a TS N HOUR agrupada de 0 hs à 23 hs. Fica evidente um declínio geral no número de ocorrências em todos os horários, quando comparam-se os perfis pré-pandêmicos e pandêmicos, embora persistam indícios da sazonalidade horária. Independente do perfil, as ocorrências crescem entre 15 hs da tarde e 01 hora da madrugada, com pico entre 21 e 22 horas. Fora desses horários, a taxa de crimes oscila entre 20% e 25% das máximas ocorrências.

Figura 3.6 – Análise comparativa dos perfis pré-pandêmicos (azul) e pandêmicos (vermelho), agrupados por hora. Todas as ocorrências foram agrupadas de 0 à 23 horas. Os números entre as curvas indicam o percentual de queda no número de crimes conforme o horário. Nota-se uma queda geral das ocorrências de crimes, principalmente entre 2 e 3 horas na madrugada (-81%) e entre 7 e 8 horas da manhã (-87%). Essas quedas podem ter afetado a força das frequências decorrentes de períodos de 8 horas identificadas no espectro de frequências. As linhas pontilhadas demarcam alguns horários comentados no texto.



Fonte: Autoria própria

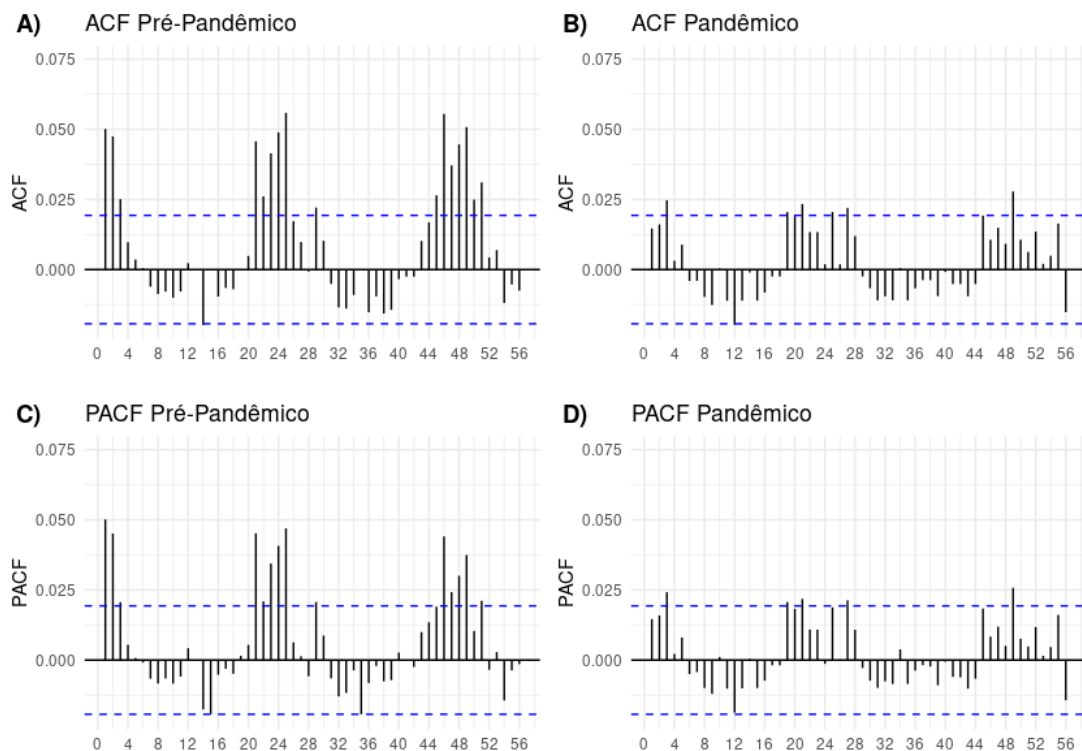
No período pré-pandêmico, notam-se pequenos picos diurnos, principalmente às 7 hs, 10 hs e 14 hs. No período pandêmico, o pico das 7hs transforma-se num vale, e o das 10hs é atenuado. As maiores quedas ocorrem no horário das 3hs (-81%) e 7hs (-87%). Essas quedas podem ter influenciado na redução da potência nos períodos de 8 hs no espectro das frequências, conforme identificado nos periodogramas da figura 3.5. Pode-se hipotetizar que, com o confinamento e o trabalho remoto, houve menor deslocamento de pessoas para trabalhar nos períodos da madrugada e da manhã, afetando os picos das 2hs e 7 hs, respectivamente.

Uma outra forma ainda de analisarmos os padrões temporais da TS NHOURL é através dos gráficos de autocorrelações, conforme figura 3.7. Foram utilizadas as funções ACF e PACF, para autocorrelações simples e parciais, respectivamente. Conforme explicado na metodologia, ACFs computam correlações isoladas entre termos no tempo t e

$t+h$, enquanto PACFs expurgam as influências dos termos intermediários entre eles, logo oferecem correlações mais puras.

Observa-se para o período pré-pandêmico (figura 3.7-A e 3.7-C), indicativos tanto de tendências quanto sazonalidades. As tendências ficaram evidenciadas pela presença de autocorrelações acima do intervalo de confiança a 0.99, para deslocamentos h pequenos, como de 1 a 3 ou no entorno dos picos. As sazonalidades, pela presença de ACF/PACFs mais altos e positivos em períodos próximos de 24 horas (e seus múltiplos), e vales negativos em torno de 12 horas (e seus múltiplos). Tais resultados corroboram os vistos nas figuras 3.5 e 3.6. Mas, não ficaram tão marcantes as autocorrelações com deslocamento de 8 horas, o que não é de todo incoerente com o gráfico do espectro de frequências para o período pré-pandêmico da figura 3.5-B, cujo intervalo de confiança inferior mostrou-se próximo das potências irrelevantes.

Figura 3.7 – Gráfico de autocorrelações da TS NHOOR: simples (ACF) e parciais (PACF) A) ACF do período pré-pandêmico. B) ACF do período pandêmico. C) PACF do período pré-pandêmico. D) PACF do período pandêmico. Linhas tracejadas em azul demarcam intervalo de confiança a 0.99. Período pré-pandêmico acusa presença de tendências e sazonalidades, enquanto o período pandêmico apresenta um perfil próximo do estacionário ou ao menos sem presença de tendências.



Fonte: Autoria própria

Para o período pandêmico, os indicativos de tendências e sazonalidades ficaram bem enfraquecidos, conforme as figuras 3.7-B e 3.7-D. A maior parte das estimativas ficaram dentro do intervalo de confiança à 0.99 que delimita a região de oscilações aleatórias aceitáveis para ACF/PACF nulas. Trata-se de outro indicativo que reforça a ideia de algo

mais próximo da estacionariedade nas ocorrências de crimes na pandemia. Ou pelo menos que as tendências foram estabilizadas, diferentemente do período pré-pandêmico. A sazonalidade, ainda que enfraquecida (como visto na figura 3.6 em vermelho) parece ter uma presença tênue no perfil das funções ACF e PACF, pelos picos e vales alternados que são vistos nas figuras 3.7-B e 3.7-D, ainda que a maioria deles estejam dentro do intervalo de confiança a 0.99. Bom lembrar que, conforme visto na tabela 3.5, a média de crimes que era de 1 crime a cada 9 horas no período pré-pandêmico, caiu para 1 crime a cada 26 horas no período pandêmico.

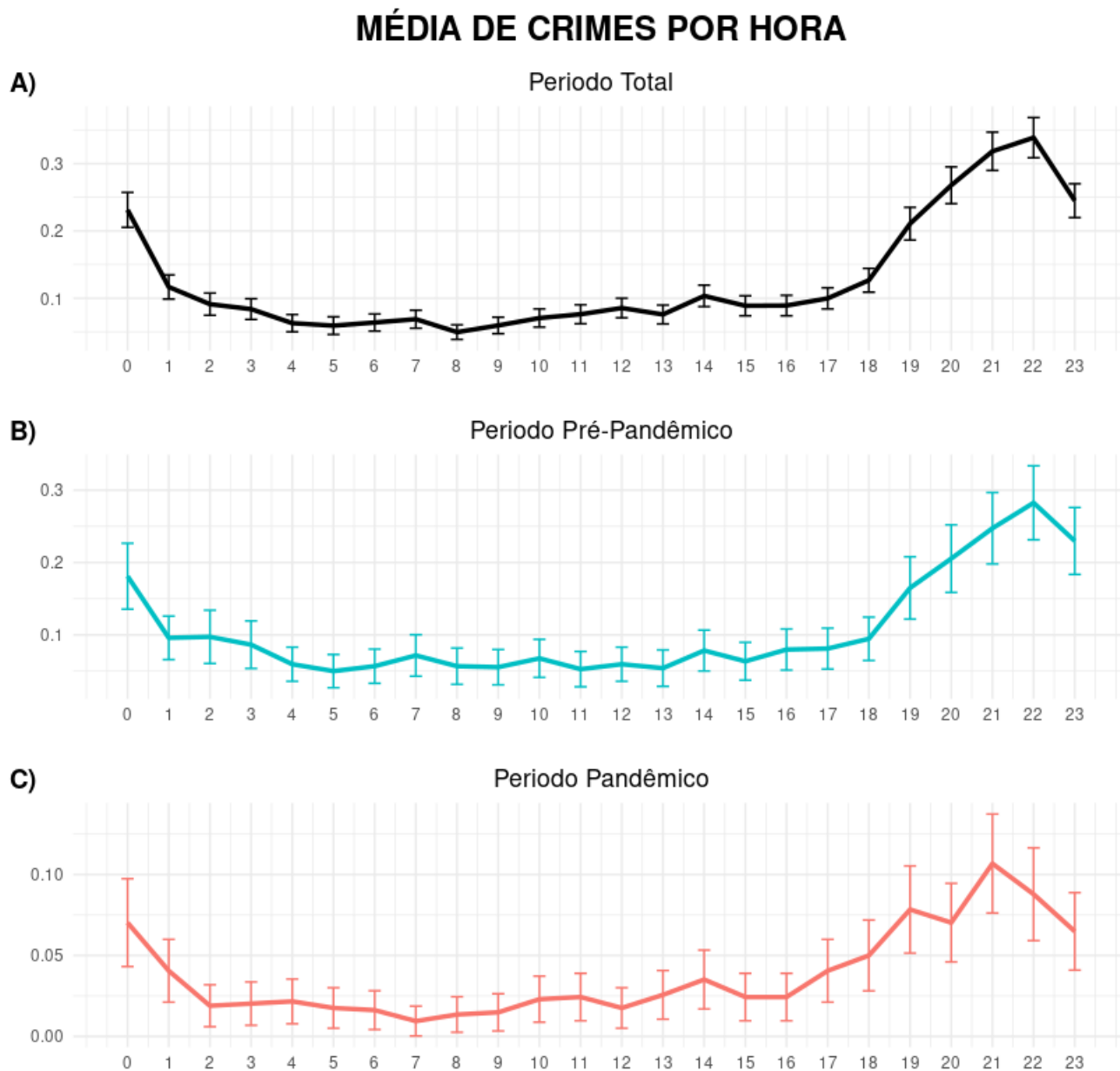
Por fim, o gráfico da figura 3.8 compara a média de crimes por hora considerando os períodos total, pré-pandêmico e pandêmico. Barras indicam os intervalos de confiança das médias à 0.99. Nessa análise das médias, o período total indica, além da sazonalidade noturno/diurno, um pequeno vale às 8hs e um pequeno pico às 14hs. Como os intervalos de confiança entre 8hs e 14hs não se sobrepõem. Levando em conta o período pré-pandêmico (3.8-A), aqueles pequenos picos às 7hs, 10hs e 14hs mostram-se estatisticamente indistintos, dada a sobreposição dos intervalos de confiança. Continuam relevantes apenas as diferenças entre o noturno e o diurno. Já no período pandêmico, a sobreposição dos intervalos de confiança é mais ampla, e as diferenças entre as médias de crimes em função das horas são bem menos marcantes.

Analisando no todo o conjunto de evidências reunidas até agora, uma conclusão possível é que a pandemia reduziu as ocorrências de crimes ao longo de todas as horas, tornando a série pandêmica próxima de um perfil estacionário, ainda que um viés de sazonalidade diurno/noturno não a tenha feito estritamente estacionária. Implica que os crimes no período pandêmico passaram a ocorrer de forma muito mais aleatória ao longo das horas que no período pré-pandêmico. Tal aumento de aleatoriedade, no entanto, não desfez de todo a principal sazonalidade em torno de 24 horas que existia antes da pandemia.

3.3 Crimes como Série Temporal (Por Dia)

Como visto, a TS NHOURL com crimes por hora é muito esparsa, sem registro de ocorrências em 90% das horas. Tamanha esparsidade pode esconder ou tornar difícil a identificação de padrões temporais em escalas de tempo mais altas. Os dados de crimes são “contagens”, gerando uma matriz discreta de dados. Além disso, a esparsidade dos crimes ao longo das horas torna a série temporal “inflada em zeros” ou zero-inflated (FUKANG, 2012). Tais características impõem restrições sobre quais modelos e ferramentas utilizar na análise desse tipo de série temporal (GREEN, 2021). Por isso, foram sondadas outras granulosidades, como a mostrada nesta seção. Foi criada a TS NDAY, com ocorrências de crimes agrupados por dia.

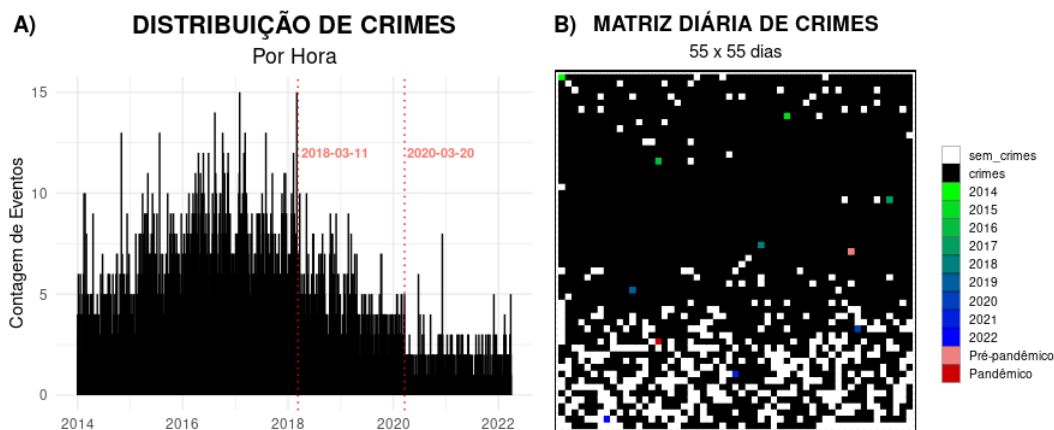
Figura 3.8 – Perfis da média de crimes consolidada por hora. Barras indicam o intervalo de confiança das médias, à 0.99, considerando uma distribuição normal de erros. A média de crimes por hora, considerando: A) o período total; B) o período pré-pandêmico; C) o período pandêmico.



Fonte: Autoria própria

A figura 3.9 faz uma panorâmica da TS NDAY. Em 3.7-A, vê-se o perfil das ocorrências por dia ao longo dos anos. Nesse gráfico, fica mais clara as tendências já vistas na TS NHOURL, evidenciando um perfil de crimes que aumenta desde 2014, passa por um platô em 2017, e começa a decair em 2018, estabilizando-se a partir de 2020. Na 3.7-B tem-se a mesma formatação da matriz da figura 3.4-B, mas consolidado em crimes por dia, com o reticulado colorido conforme legenda. Percebe-se uma esparsidade da NDAY bem menor que na NHOURL. O aumento das ocorrências em 2016 e 2017 fica

Figura 3.9 – Perfis da TS NDAY. A) contagem de ocorrências ao longo dos dias começando em 2014-01-01 e terminando em 2022-03-29, conforme matriz com 3010 posições de dias, para um total de 9274 ocorrências de crimes. A linha pontilhada indica o início dos períodos pré-pandêmico e pandêmico em 2018-03-11 e 2020-03-20, respectivamente. B) Para melhor visualização da esparsidade, as ocorrências em A) foram convertidas numa matriz binária de 53 x 53 dimensões, começando no canto superior esquerdo, com varredura à direita e descendo. Essa binarização apenas registrava se houve alguma ocorrência naquela hora ou não. O reticulado foi colorido conforme legenda.



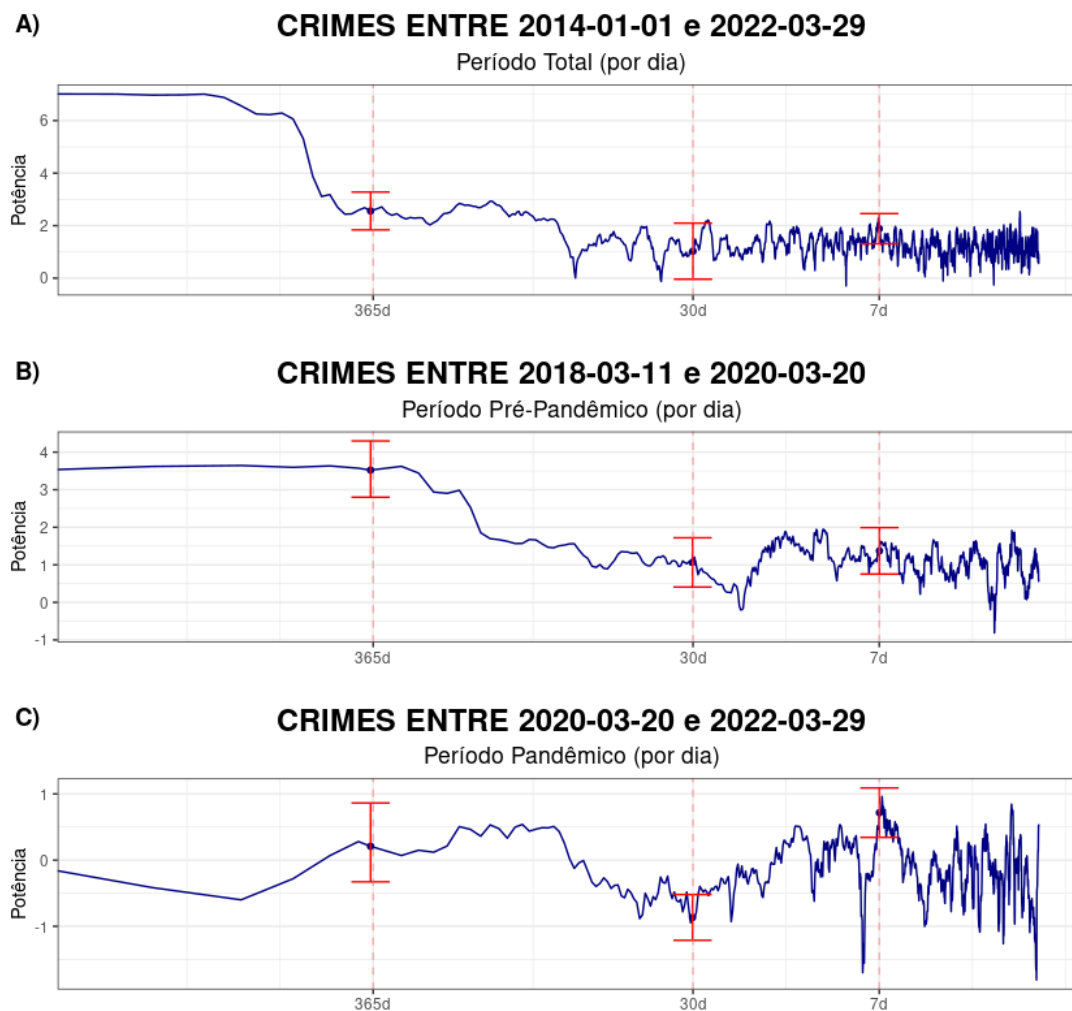
Fonte: Autoria própria

evidente pela quase ausência de pontos brancos (sem registro de crimes). No ano de 2017, inclusive, não houve um dia sequer sem ocorrência de pelo menos um crime. Mas, a partir de 2018, começam aparecer maior presença de pontos brancos, intensificados de 2020 em diante. No período pandêmico, há sequências de 5 e até 6 dias consecutivos sem crimes, algo inédito em toda a série de dados.

Em relação aos 3010 dias de toda a TS NDAY, em apenas 435 não registraram-se crimes, o equivalente a 14.5% dos dias. Logo, a TS NDAY no todo tem uma esparsidade bem menor que a TS NHOURLY. Em relação aos 740 dias do período pré-pandêmico, em 96 não houve crimes registrados, o que equivale a 13.0% dos dias. No período pandêmico com mesmo número de dias, em 300 não houve crimes registrados, ou 40.5% dos dias. A esparsidade do período pandêmico foi mais de 3 vezes maior que no período pré-pandêmico.

Na figura 3.10 observa-se o espectro de frequências da TS NDAY. Seja na série como um todo, ou nos períodos pré-pandêmico ou pandêmico, o espectro não indica presença de sazonalidade (na escala de dias) estatisticamente significativa por esse método, pois não há picos relevantes em nenhuma faixa de frequência/período. Detecta-se apenas a elevação típica das potências em frequências mais baixas (períodos mais longos) em 3.10-A e 3.10-B, algo indicativo da presença de tendências na série de dados. Já o período pandêmico em 3.10-C, não sugere tendências, mas a possibilidade de uma série próxima da estacionalidade, como foi visto também na 3.10-C.

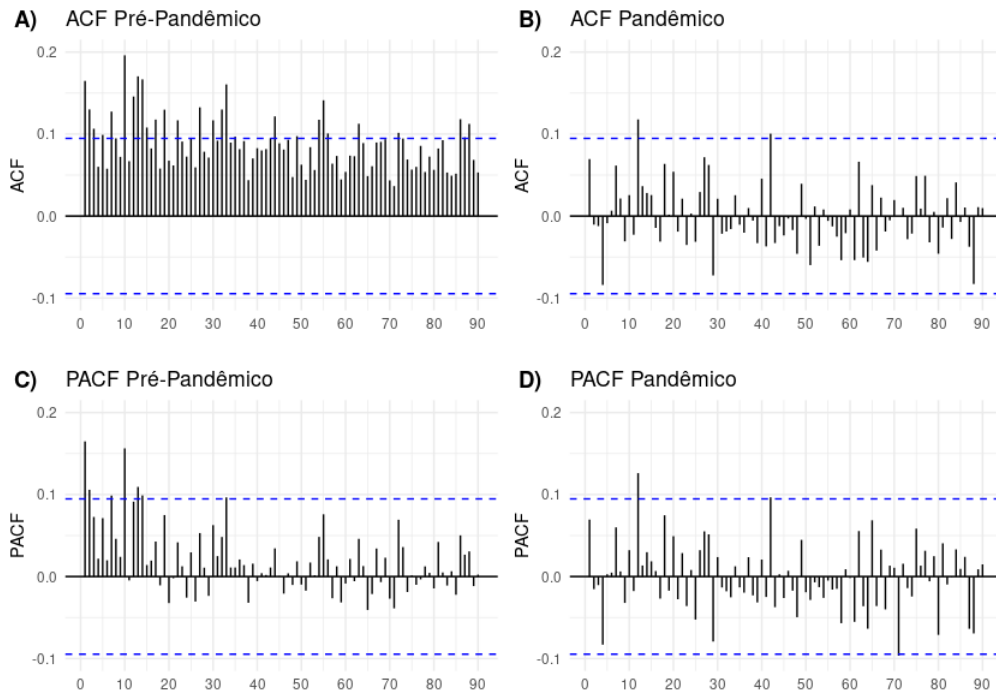
Figura 3.10 – Perfis da análise espectral de frequências sobre a TS NDAY, em escala logarítmica em X e Y. O eixo X representa as frequências, mas os rótulos foram trocados por períodos, para melhor interpretação. Também para maior clareza, os intervalos de confiança no nível de 0.95 foram estimados apenas para as frequências/períodos em destaque, pelo método jackknife. A) perfil espectral das frequências para toda a TS NDAY. B) perfil espectral do período pré-pandêmico, considerando o mesmo número de dias que o período pandêmico. C) perfil espectral das frequências no período pandêmico.



Fonte: Autoria própria

Os gráficos de autocorrelações ACF/PACF da TS NDAY parecem confirmar as observações espectrais anteriores, pois não há evidências de sazonalidades diárias, seja no período pré-pandêmico ou pandêmico. No que diz respeito às tendências, elas parecem presentes no período pré-pandêmico, pelo decaimento das correlações ao longo dos deslocamentos temporais, principalmente na ACF. Reforça essa interpretação o fato de na ACF não se constatar correlações negativas. Mas, no período pandêmico, o indício de estacionaridade na NDAY é muito mais forte que na NHOURLY, dado que pelo menos 99% das correlações estão no intervalo de confiança, sem indícios de tendências ou sazonalidades. Pode-se inferir que, na granularidade de dias, no período pandêmico, o comportamento dos crimes seguiu um comportamento tipicamente aleatório e imprevisível, no sentido

Figura 3.11 – Gráfico de autocorrelações da TS NDAY: simples (ACF) e parciais (PACF) A) ACF do período pré-pandêmico. B) ACF do período pandêmico. C) PACF do período pré-pandêmico. D) PACF do período pandêmico. Linhas tracejadas em azul demarcam intervalo de confiança a 0.99. No período pré-pandêmico, nota-se presença de tendências, especialmente em ACF, mas em nenhum período há indicativo de sazonalidade. No período pandêmico, os gráficos de autocorrelação indicam um processo estacionário.

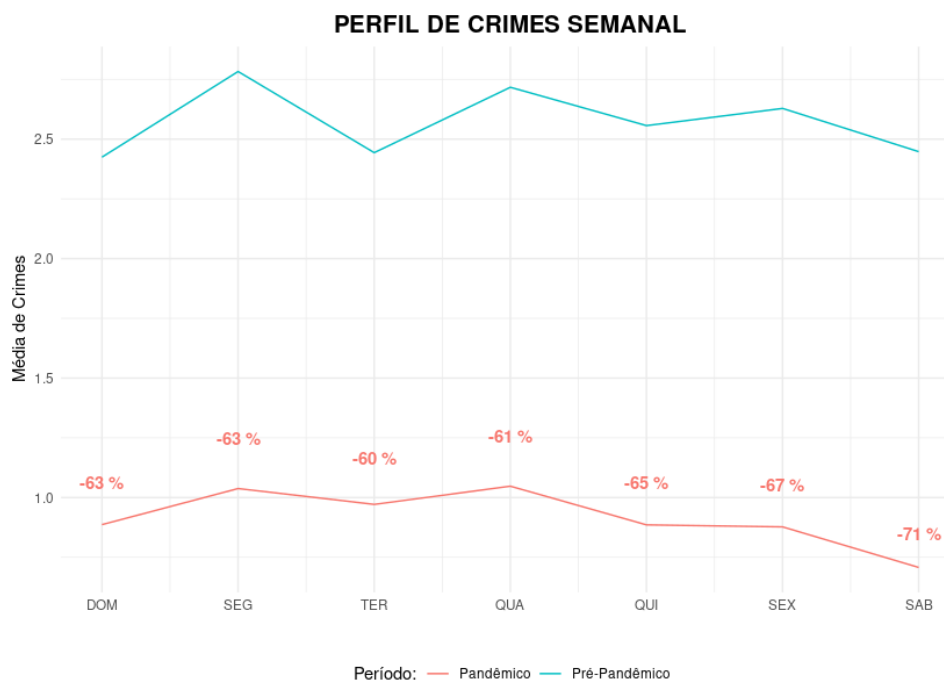


Fonte: Autoria própria

de não haver tendências ou sazonalidades que pudessem ser exploradas numa eventual previsão.

Na figura 3.13, pode-se acompanhar o perfil das ocorrências consolidado por dia da semana, entre os períodos: total (preto), pré-pandêmico (azul) e pandêmico (vermelho). Um primeiro ponto a ser considerado é que os perfis são semelhantes, com ligeiras elevações nas segundas e quartas, e pequeno declínio nos finais de semana. No período total, os intervalos de confiança parecem todos sobrepostos, exceção (talvez) entre o pequeno vale no domingo e os também pequenos picos nas segundas e quartas. Nos períodos pré-pandêmico e pandêmico, a sobreposição é geral, sugerindo que a média de crimes por dia da semana foram estatisticamente indiferentes.

Figura 3.12 – Análise comparativa dos perfis pré-pandêmicos (azul) e pandêmicos (vermelho), agrupados por dia da semana. O perfil geral das curvas são bem semelhantes. Os números entre as curvas indicam o percentual de queda no número de crimes conforme o dia da semana. Nota-se uma queda geral das ocorrências de crimes, em torno de uma média de 64%, sendo as maiores quedas entre quinta e sábado, embora não sejam estatisticamente significativas ($p\text{-value} = 0.23$, no teste de proporções, a 0.95 de confiança).



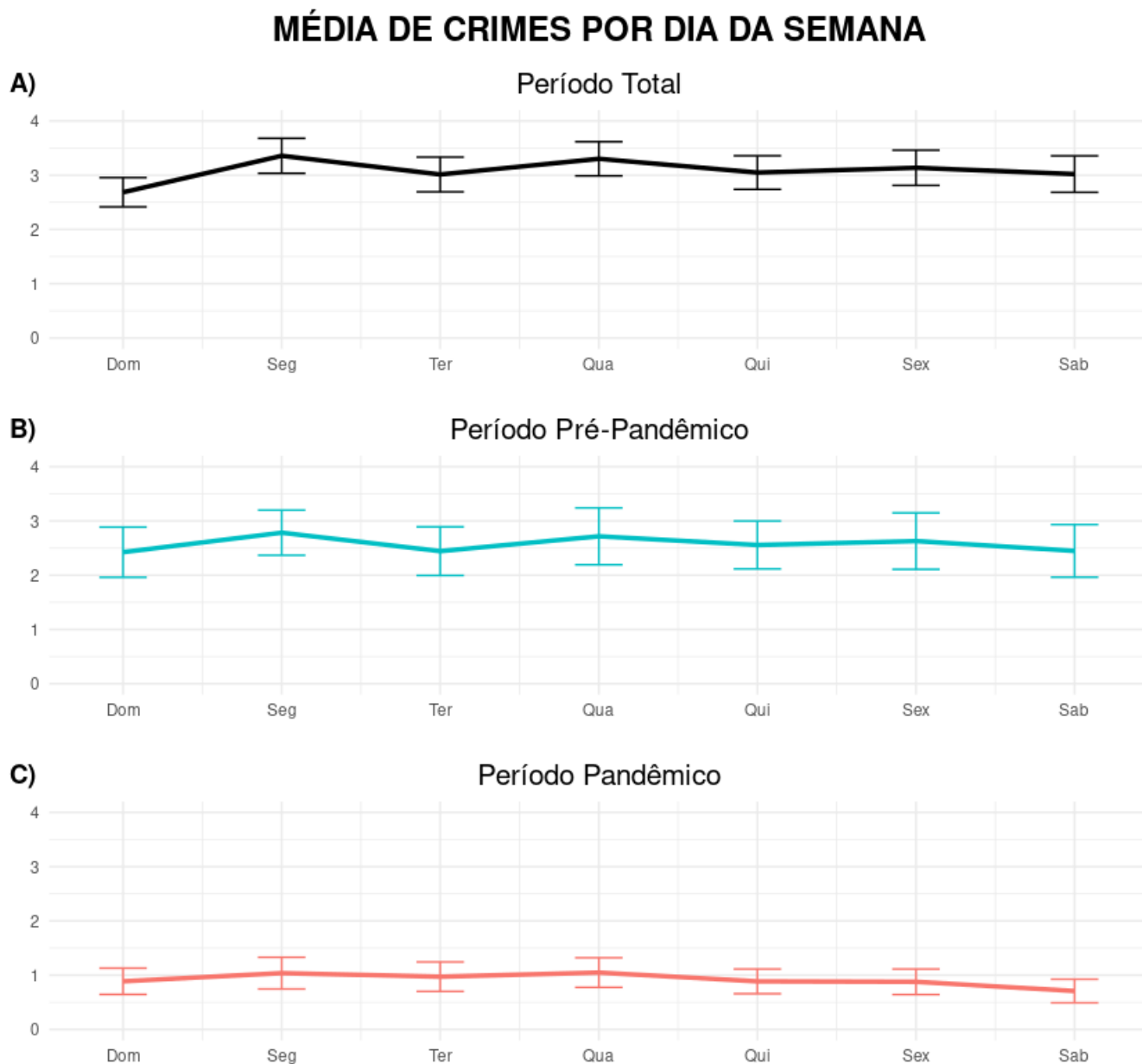
Fonte: Autoria própria

3.4 Crimes como série temporal (10 dias do mês)

Outra granularidade temporal que merece ser explorada é a condensada de 10 em 10 dias, que segmenta os meses em 3 partes (TS NTEN). A intenção é ver se há algum padrão que diferencie os 10 primeiros dias de cada mês, os dias em torno da metade do mês (dia 15), e os últimos 10 dias do mês. Quer-se verificar se há algum viés de crimes ocorrerem mais no início do que no fim do mês, dada a condição provável da população no geral estar com mais dinheiro no princípio do mês (em função dos recebimentos de salário) do que tardiamente no mês.

Um detalhe importante que deve ser levado em consideração para essa análise é que a quantidade de dias varia dependendo do mês, entre 28, 29 (nos anos bissextos), 30 ou 31 dias. Isso faria com que os últimos 10 dias de cada mês possam ter variações quantitativas no número de dias, afetando a quantidade de crimes. Para corrigir essa anomalia, trabalhou-se com a média de crimes ao longo de cada dezena. Esse procedimento também permitiu transformar os dados que eram inteiros (quantidade de crimes) para valores mais contínuos (média de crimes por dezena). Outro ponto a destacar é que a consolidação por dezena eliminou a esparsidade da TS, pois não há 10 dias seguidos sem crimes. Essas condições preencheram requisitos um para uso mais confiável de técnicas clássicas em séries

Figura 3.13 – Análise comparativa dos perfis total (preto), pré-pandêmico (azul) e pandêmico (vermelho), agrupados pela média de crimes por dia da semana. As barras indicam intervalo de confiança das médias, à 0.99, considerando uma distribuição normal de erros. O perfil geral das curvas são bem semelhantes. A sobreposição dos intervalos de confiança indicam, no geral, não haver diferenças estatísticas significativas entre os dias da semana.



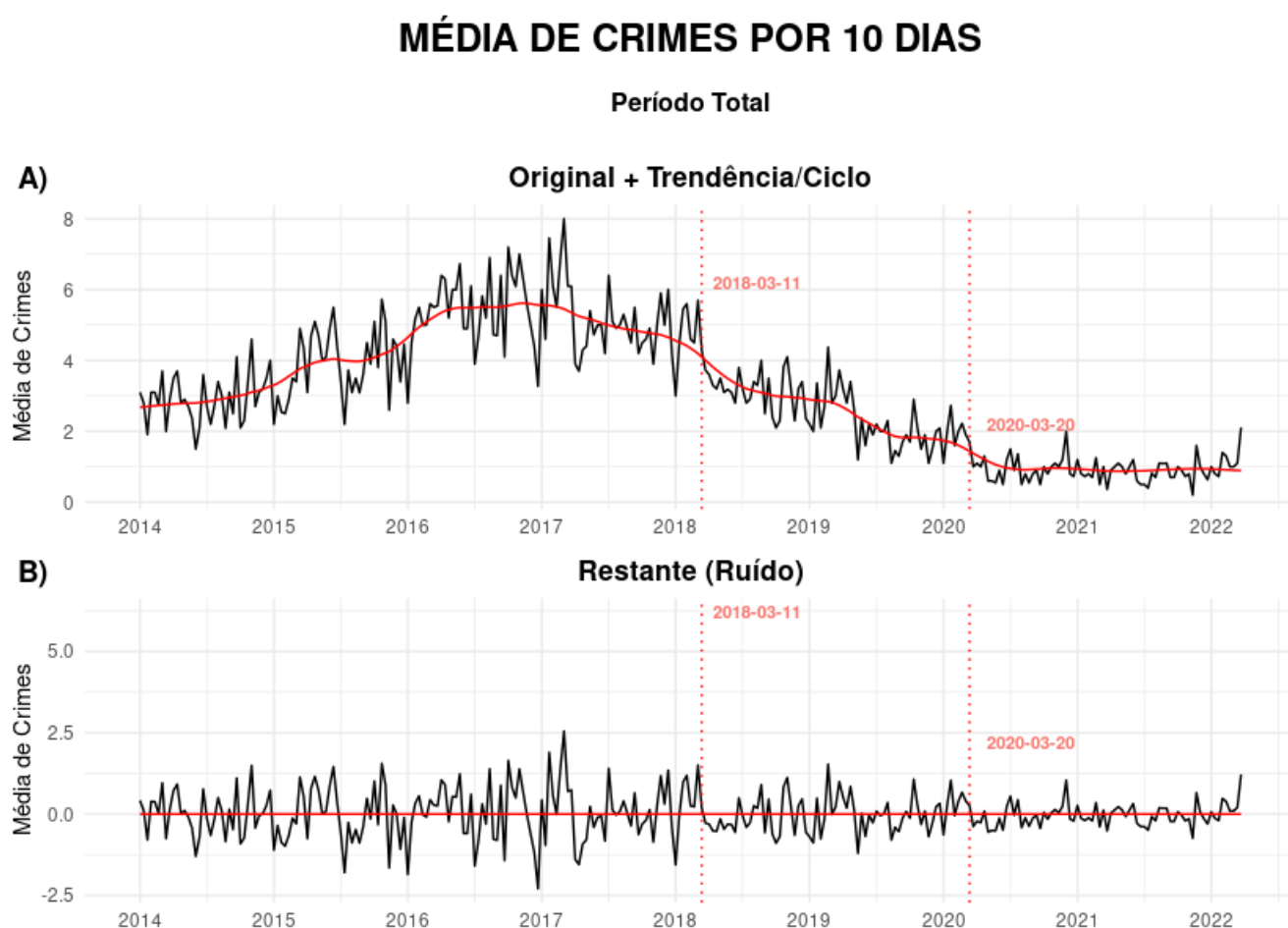
Fonte: Autoria própria

temporais (DAVIS, 2021), tais como a decomposição STL (CLEVELAND ROBERT B., 1990).

Perfis da TS NTEN podem ser vistos na figura 3.14. A linha contínua vermelha em 3.14-A compreende a componente tendência (trend), fruto da decomposição STL robusta e demais parâmetros default. Esse método decompõe um sinal em: tendências-ciclos, sazonalidades e restante (ruídos). A STL da NTEN não evidenciou sazonalidades esta-

tisticamente relevantes, retornando apenas componentes de tendências-ciclos e restante. Como já visto anteriormente, fica evidente um pico (ou um platô) na transição 2016-2017, com decaimento ainda em 2017, e estabilização a partir de 2020, principalmente no período pandêmico.

Figura 3.14 – Perfis da TS NTEN. A) A média de ocorrências a cada 10 dias, começando em 2014-01-01 e terminando em 2022-03-29, gerando uma matriz com 297 dezenas, para um total de 9274 ocorrências de crimes. A linha pontilhada indica o início dos períodos pré-pandêmico e pandêmico em 2018-03-11 e 2020-03-20, respectivamente. A linha contínua vermelha indica o perfil do componente tendência (trend), resultado da decomposição STL robusta. B) O mesmo perfil de A), subtraído da componente tendência (detrended), evidenciando apenas o restante do sinal ou o ruído. A decomposição STL não encontrou componentes sazonais relevantes.

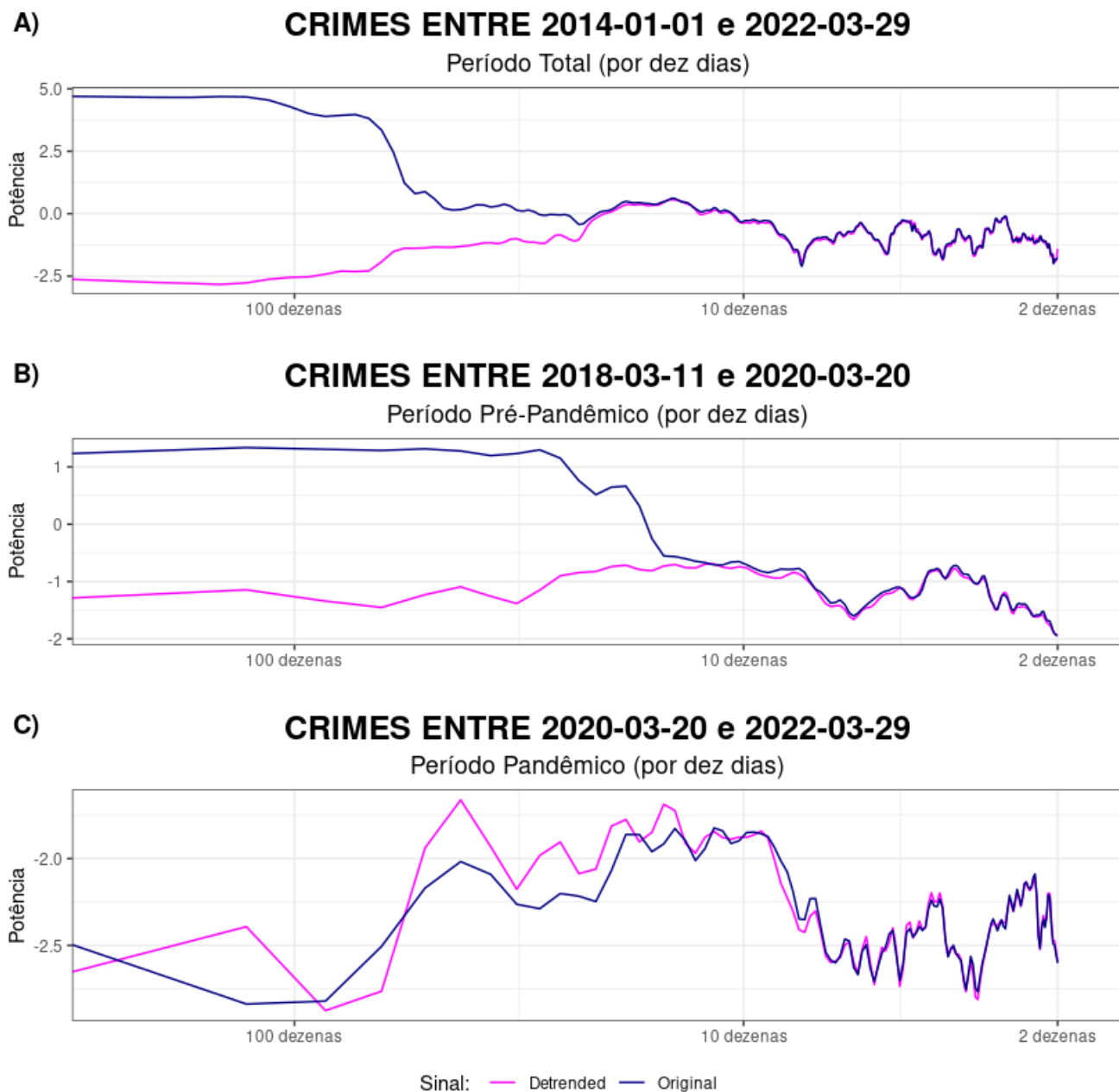


Fonte: Autoria própria

Em 3.14-B é possível ver a TS NTEN subtraída da componente tendência (detrended). Esse sinal não tendencioso parece compreender oscilações com alguma aleatoriedade em torno de zero, ainda que a variância ao longo do tempo mostre-se heterocedástica. As menores variâncias coincidem com o período pandêmico.

O espectro das frequências sobre a TS NTEN pode ser visto na figura 3.15. De forma geral, não se percebe nenhuma frequência relevante, seja no período total, pré-

Figura 3.15 – Perfis da análise espectral de frequências sobre a TS NTEN, em média de crimes a cada 10 dias, em escala logarítmica em X e Y. O eixo X representa as frequências, mas os rótulos foram trocados por períodos, para melhor interpretação. Em azul, o espectro tirado do sinal original; em magenta, sobre o sinal sem tendências (detrended). A) perfil espectral das frequências para toda a série. B) perfil espectral do período pré-pandêmico, considerando o mesmo número de dias que o período pandêmico. C) perfil espectral das frequências no período pandêmico.



Fonte: Autoria própria

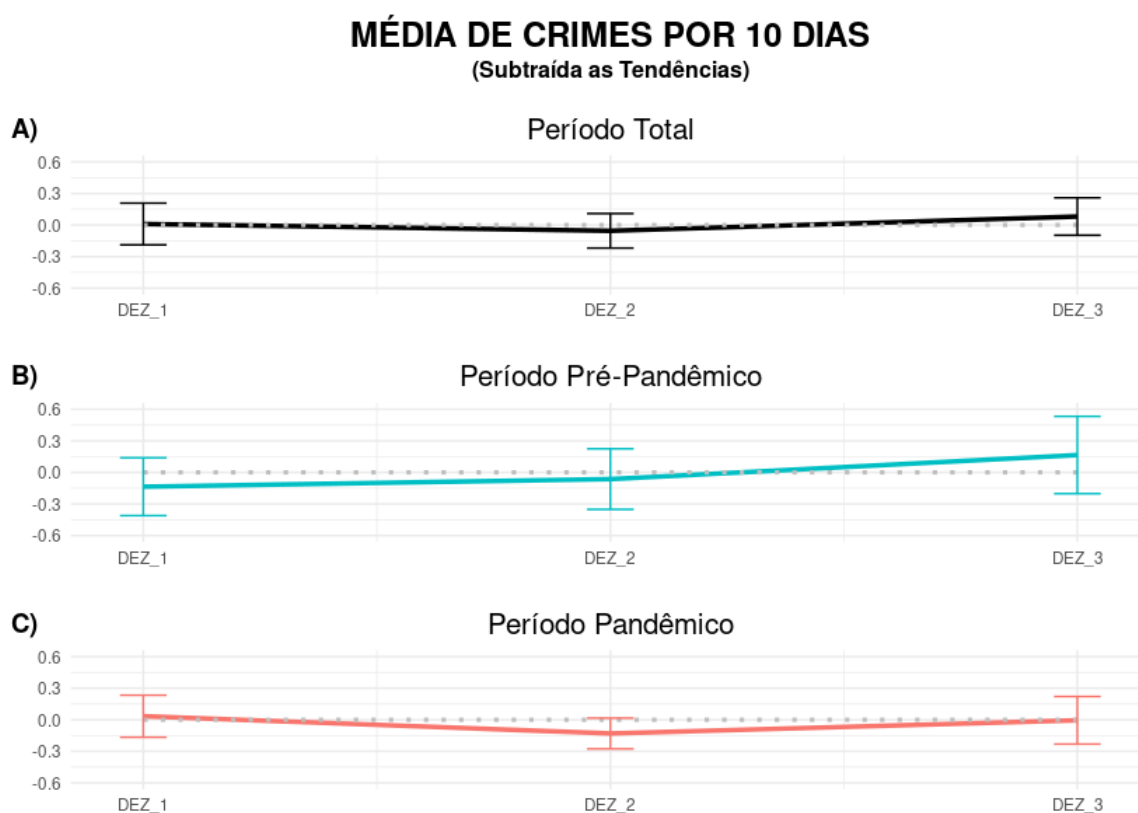
pandêmico ou pandêmico. Pode-se concluir que não há evidências de períodos preferenciais ao longo do mês. Ou seja, a ocorrência de crimes não distingue entre início, meio ou fim do mês, sendo indiferente a essa condição, ao contrário do que apostaria o senso comum.

O interessante é que foi possível constatar pela figura 3.15 as influências das ten-

dências na análise do espectro de frequências, comparando os sinais originais (azul) e subtraído das tendências (magenta). Nas frequências mais altas (períodos curtos), ambos sinais são indistintos. Mas, nas frequências baixas (períodos longos), nota-se uma redução marcante de potência quando as tendências são eliminadas, reproduzindo algo já observado na literatura em outros contextos (RAHIM, 2014). No período pandêmico, entretanto, mesmo nas frequências mais baixas (períodos mais longos), não se observou diferenças significativas entre o sinal original e o subtraído das tendências. Isso constitui mais uma evidência de que no período pandêmico a ocorrência de crimes seguiu algo próximo da estacionariedade, sem tendências marcantes, mesmo na escala de dezenas de dias.

A figura 3.16 revela o perfil consolidado das médias de crimes ao longo das 3 dezenas, para os períodos total, pré-pandêmico e pandêmico. Independente do período, a sobreposição dos intervalos de confiança indicam não haver diferenças estatisticamente significativas entre as médias por dezena.

Figura 3.16 – Análise comparativa dos perfis: A) total (preto), B) pré-pandêmicos (azul) e C) pandêmicos (vermelho), agrupados pela média das médias das dezenas do mês. As barras indicam intervalo de confiança das médias, à 0.99, considerando uma distribuição t de erros. A sobreposição dos intervalos sugere indiferença estatística entre as dezenas.

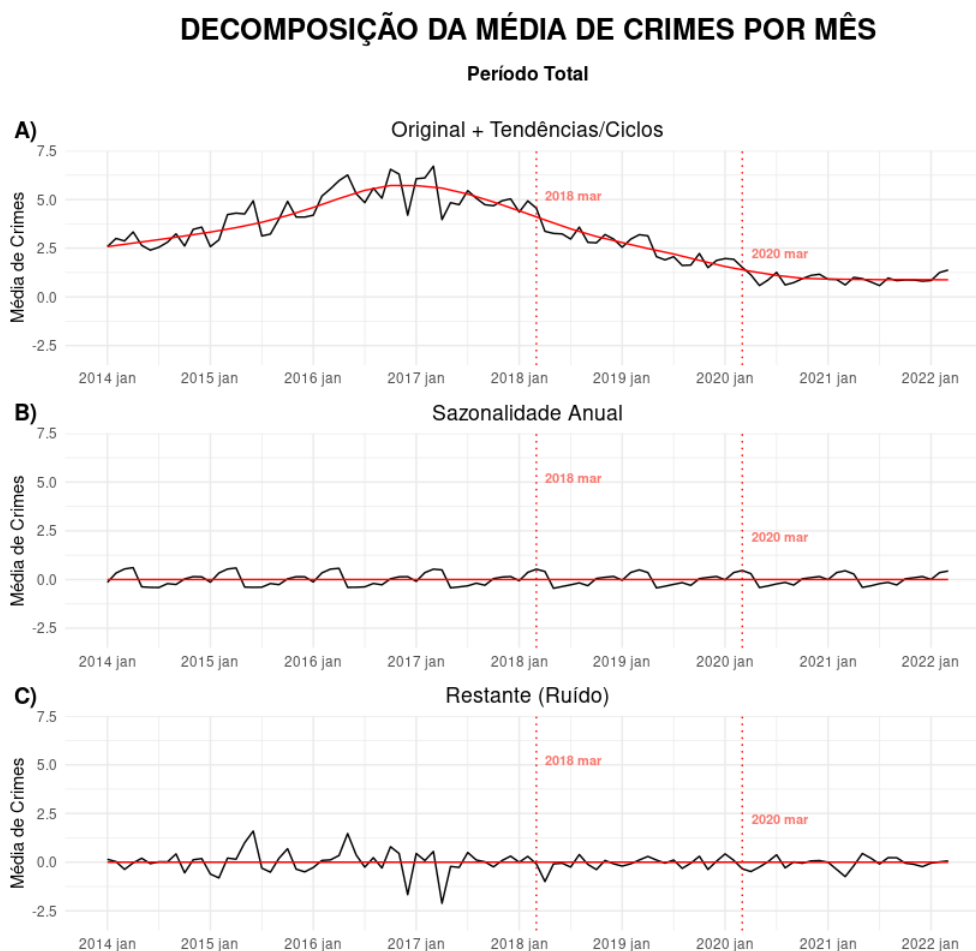


Fonte: Autoria própria

3.5 Crimes como série temporal (Por Mês)

A última granularidade temporal analisada nesta dissertação foi a de crimes por mês, na TS NMON. Como também feito no caso das dezenas, trabalhou-se aqui com a média de crimes por mês, dado que há variação na quantidade de dias em cada mês. A simples contagem de crimes por mês introduziria um viés, uma vez que meses com mais dias tenderiam a acumular mais crimes. Essa transformação de contagens de crimes por média de crimes permite também trabalhar com dados numa forma próxima do contínuo em vez de discreto, habilitando o uso de outras técnicas em séries temporais, como a decomposição STL.

Figura 3.17 – Análise da decomposição STL da TS NMON (crimes por mês). A linha pontilhada indica o ano/mês de início dos períodos pré-pandêmico e pandêmico em 2018-03 e 2020-03, respectivamente. A) Média de crimes por mês em preto, conjuntamente com a componente Tendência/Ciclo (Trend/Cycle) em vermelho. B) Componente Sazonal por Ano (preto). C) A componente que reuniu o Restante (Ruído) da decomposição.



A figura 3.17 mostra o resultado da decomposição STL da TS NMON. Em 3.17-A, veem-se sobrepostas: o sinal original da média de crimes por mês (em preto) e a componente Tendência/Ciclo em vermelho. Diferente da decomposição da TS NTEN de

média de crimes por dezenas do mês, a da TS NMON encontrou um sinal de sazonalidade anual, com um modal entre março e abril, conforme visto na figura 3.17-B. O restante do sinal, tido como um ruído, está na figura 3.17-C.

Foi calculado o índice de força (strength) das componentes Tendência/Ciclo (Ft) e Sazonalidade Anual (Fs), conforme as equações descritas na metodologia. Lembrando que nesses índices, quanto mais próximo 1, maior a força ou importância da componente; e próximo de 0 (zero), o contrário. A decomposição da TS NMON encontrou $F_t = 0.92$ e $F_s = 0.10$, indicando que a componente Tendência/Ciclo é considerada marcante, mas o sinal da Sazonalidade Anual fraco ou de pouca importância.

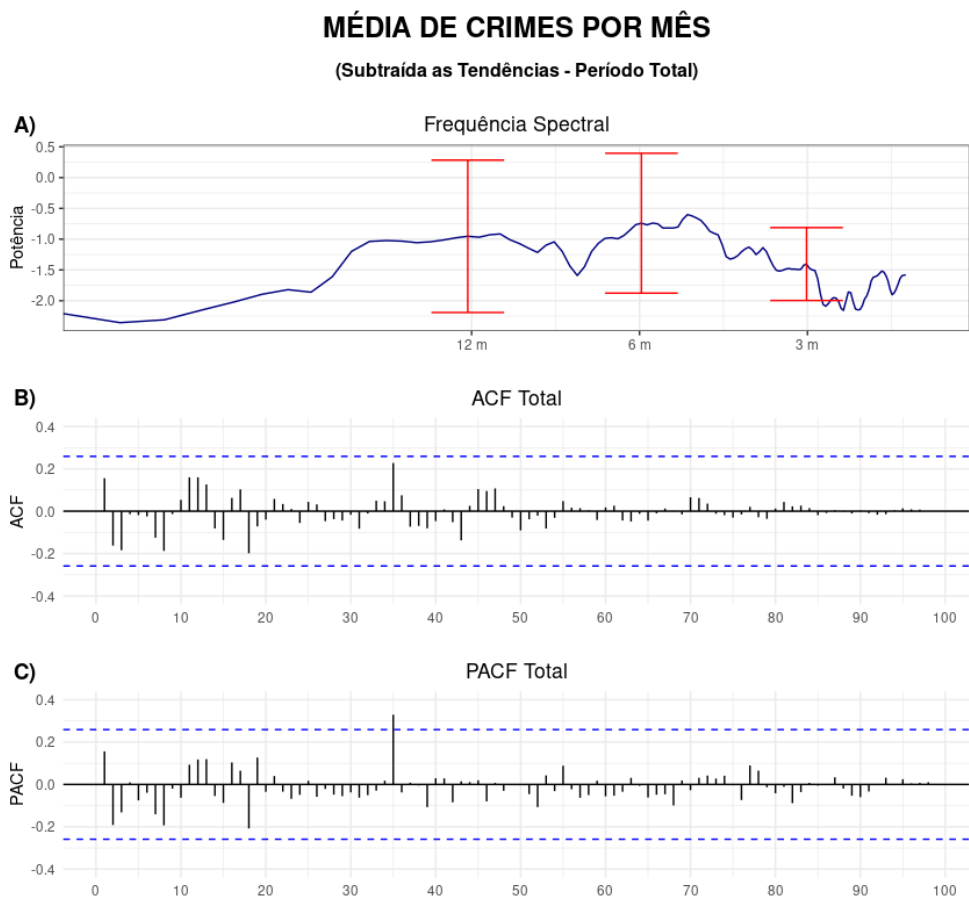
Em função disso, na série sem tendências (detrended) da TS NMON apenas subtraiu-se do sinal original a componente Tendência/Ciclo, a fim de confirmar se a componente sazonal tinha realmente essa pouca relevância.

A figura 3.18 mostra algumas análises gráficas da TS NMON sem tendências. Em 3.18-A, vê-se o espectro de frequências do sinal destendencioso. Em 3.18-B e 3.18-C, o perfil das autocorrelações simples e parciais, respectivamente. Não há indícios de frequências/periódos relevantes no espectro de frequências, nem tão pouco autocorrelações que extrapolem o intervalo de confiança em zero, no nível de 0.99. Aliás, o perfil das autocorrelações dá indícios que a série temporal de crimes, na escala de meses e considerada no todo, tende a estacionária. Nessa escala, não foi possível fazer a análise em períodos pré-pandêmico e pandêmico, por conta da série compreender apenas 24 meses, uma amostragem de tempo muito curta para gerar resultados confiáveis.

Na figura 3.19 encerra as análises temporais deste trabalho. Ela oferece uma visão consolidada por mês das médias de crimes por mês, extirpadas as tendências/ciclos. Em 3.19-A, para o período total, vê-se que os intervalos de confiança (à 0.99) sobrepõem-se na maioria dos meses, exceção (talvez) para o mês de março, cujo limite inferior da barra de erro chega próximo de zero. O mês de abril foi o que apresentou maior variação da média mensal apurada. Em 3.19-B e C, têm-se os perfis para os períodos pré-pandêmico e pandêmico, respectivamente. Não foi possível estimar o intervalo de confiança por conta que cada mês contou apenas com 2 pontos amostrais, restando apenas uma análise mais qualitativa nesses períodos. Percebe-se que o período pré-pandêmico apresentou um perfil mais próximo do total, enquanto no período pandêmico, os desvios foram mais homogêneos e próximos de zero.

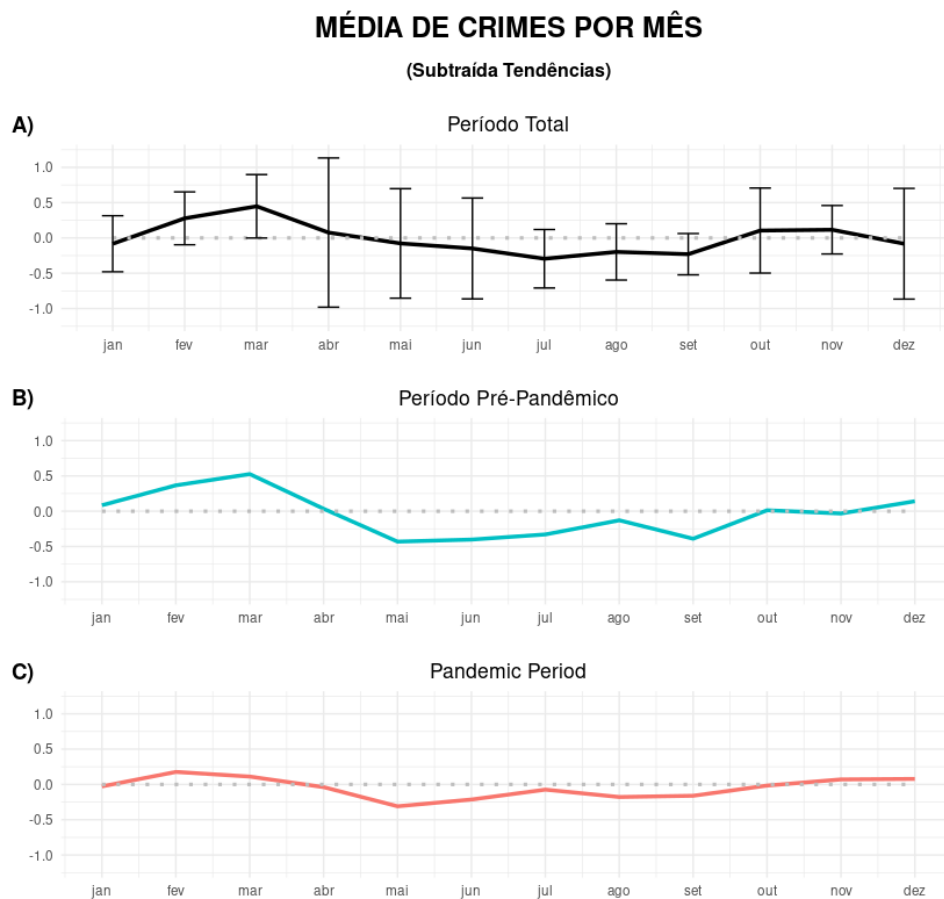
Logo, pode-se concluir que, na escala de meses, os dados não foram suficientes para atestar, com segurança estatística, distinção dos meses para com a média mensal de crimes, seja no período total, pré-pandêmico ou pandêmico. Qualquer que seja o mês, a média de crimes é a mesma, não havendo um mês preferencial a qualquer outro. O mês de março apresenta um viés de distinção, mas no conjunto, essa distinção não teve suporte estatístico convincente.

Figura 3.18 – Análises gerais da série TS NMON considerando o período total, subtraída das tendências. A) gráfico da análise espectral de frequências por MTSA, com escala logarítmica em X e Y. O eixo X representa as frequências, mas os rótulos foram trocados por períodos, em que m = meses, para melhor interpretação. Também para maior clareza visual, os intervalos de confiança no nível de 0.99 foram estimados apenas para as frequências/períodos em destaque, pelo método jackknife. Não são notadas nenhuma frequência relevante. Em B) e C), autocorrelações simples (ACF) e parciais (PACF), respectivamente. Não se verifica sazonalidade, com os perfis ACF e PACF indicando estados estacionários.



Fonte: Autoria própria

Figura 3.19 – Análise comparativa dos perfis: A) total (preto), B) pré-pandêmicos (azul) e C) pandêmicos (vermelho), consolidado por mês. As barras indicam intervalo de confiança das médias, à 0.99, considerando uma distribuição t de erros. No perfil do período total A), a sobreposição geral dos intervalos sugere indiferença estatística entre meses, exceção (talvez) para o mês de março. Não foram mostradas barras de erro para os perfis pré-pandêmico e pandêmico, dado que a amostragem de média de crimes compreendeu apenas 2 apurações para cada mês. Logo, para esses perfis a análise limitou-se ao qualitativo. Percebe-se que o perfil pré-pandêmico seguiu algo mais próximo do perfil total, enquanto no perfil pandêmico as oscilações ao longo dos meses foram bem menores.



Fonte: Os autores (org.), com base em dados da Polícia Militar de Minas Gerais (PMMG)

3.6 Geoprocessamento

Na figura 3.20 é possível ver alguns quadros da distribuição de crimes no município de Itabira, através de um mapa de calor animado com base nos dados de roubos de rua. Os resultados obtidos oferecem uma representação visual dos padrões e tendências dos roubos ao longo do tempo e da distribuição espacial das ocorrências na cidade em estudo.

O mapa de calor revela a intensidade dos roubos de rua em certos pontos geográficos, em diferentes períodos, seja diariamente, mensalmente ou anualmente, dependendo da granularidade dos dados e dos períodos selecionados. As áreas mais quentes do mapa representam regiões com maior concentração de roubos, enquanto as áreas mais frias indicam locais com menor incidência desses crimes.

Através da sequência de mapas de calor, foi possível observar a evolução geográfica dos roubos de rua ao longo do tempo, identificando possíveis flutuações, picos ou declínios nos índices de criminalidade. Percebe-se que no período pandêmico há uma nítida rarefação dessas manchas de calor em todo o recorte geográfico delineado.

Fica patente que a representação espacial dos roubos no mapa permite identificar áreas geográficas específicas que apresentaram maior concentração de ocorrências, possibilitando a focalização de esforços e recursos em ações de policiamento e segurança nessas regiões mais vulneráveis. Os resultados obtidos a partir da metodologia proposta não apenas fornecem uma visualização clara e intuitiva dos dados de roubos, mas também permitem uma análise temporal e espacial dos padrões criminais nas cidades.

Isso fica mais evidente nos links contendo os GIFs animados disponibilizados ao final deste trabalho.

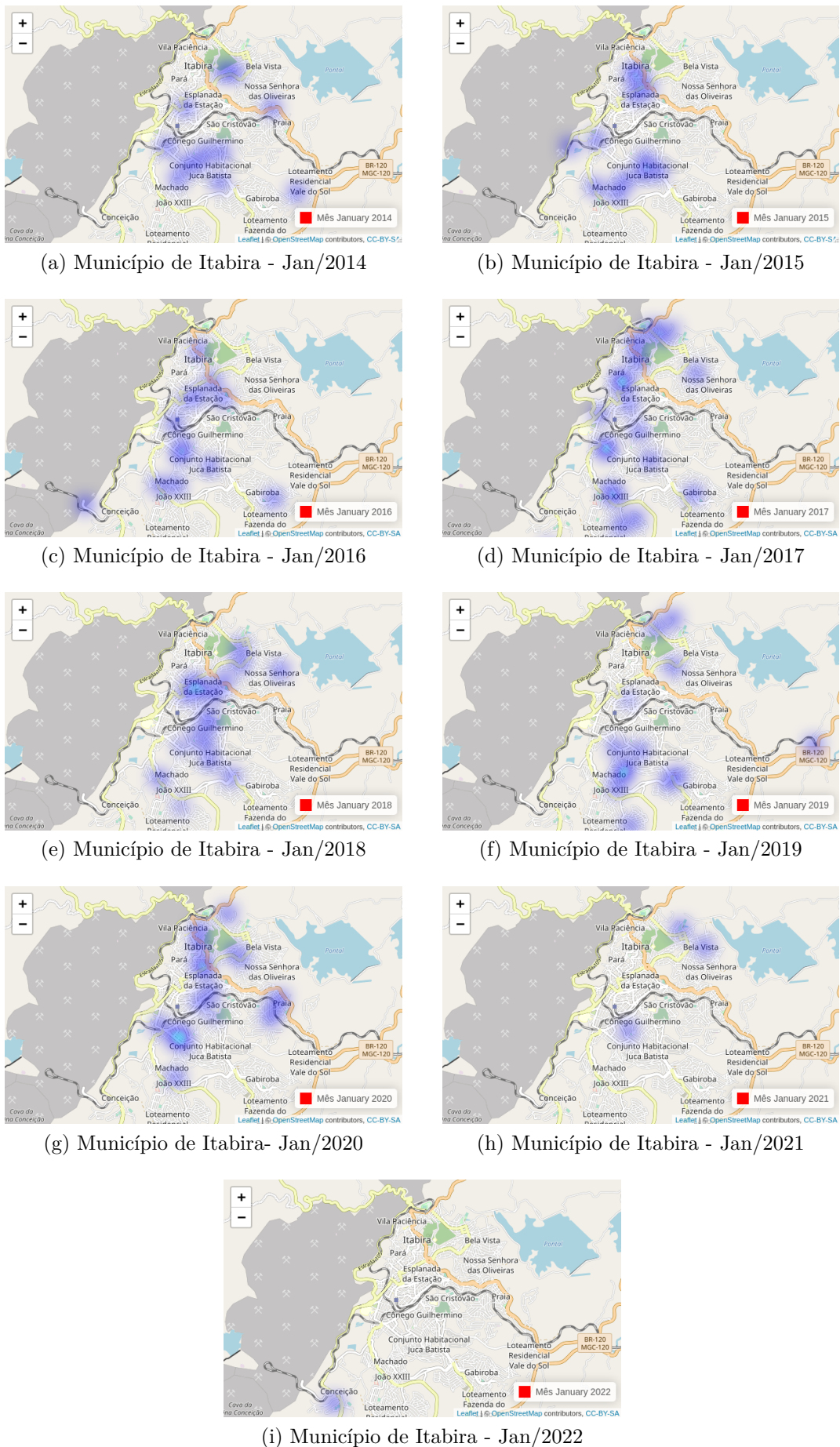


Figura 3.20 – Mapa de Calor de Crimes de Rua - Itabira

Fonte: Autoria própria

4 Conclusão

Os resultados obtidos, ao explorar diferentes escalas temporais, proporcionaram uma compreensão abrangente da influência da pandemia nos padrões criminais. A redução significativa na taxa de crimes, especialmente nas horas da madrugada, destacou a importância de considerar a sazonalidade diária na análise de eventos criminais. A inesperada estabilidade em escalas temporais mais longas, como dias, dezenas de dias e meses, desafiou as expectativas comuns, sublinhando a complexidade das tendências criminais ao longo do tempo.

A aplicação eficaz da metodologia de criação de mapas de calor e sua capacidade de identificar áreas críticas de concentração de ocorrências demonstraram ser recursos valiosos na análise de dados de roubos. A visualização espacial dessas áreas por meio de mapas de calor permitiu uma identificação clara dos pontos que demandam atenção prioritária das autoridades de segurança.

Concluimos, com base nos métodos robustos empregados, que a abordagem visual e analítica, especialmente por meio dos mapas de calor animados, não apenas ofereceu insights profundos sobre os padrões e tendências criminais, mas também ressaltou a importância fundamental de compreender as nuances das mudanças na criminalidade de rua durante a pandemia. Essa compreensão detalhada é vital para a formulação e implementação eficaz de estratégias preventivas direcionadas, fortalecendo assim a segurança da comunidade e contribuindo ativamente para a redução dos índices de criminalidade.

4.1 Melhorias futuras

Visando a proposta inicial do projeto, pretende-se realizar ainda as seguintes melhorias, buscando otimizar o funcionamento e utilização do mesmo:

- Aumentar a base de dados de roubos, onde houve a limitação de quantidade de ocorrências e de dados englobando um conjunto estrito de cidades do interior de Minas Gerais.
- Limitação em apenas crimes de roubos de rua, tipificados no Art. 157 do Código Penal. Dada a presunção da sua correlação com a mobilidade da população, por sua vez criticamente afetada pela pandemia.
- A expansão da análise para outros tipos de crimes ficou para estudos futuros.

Apêndices

APÊNDICE A – Imagens dos Municípios de Minas Gerais e GIFs do Município de Itabira

Link de acesso ao Mapa de Calor dos Municípios de Minas Gerais:

<https://drive.google.com/drive/folders/1Pte8We9E92OeUphdcQUN-PcrxRu_HPhy?usp=sharing>

Link de acesso ao GIF (*Formato de intercâmbio de gráficos*) animado do Município de Itabira com a variações anuais e mês constante:

<<https://drive.google.com/file/d/1O-TVyxxkDav08hKbfc5gNLDfAedvvtx/view?usp=sharing>>

Link de acesso ao GIF animado do Município de Itabira com a variações mensais e ano constante:

<<https://drive.google.com/file/d/1pCsEvk9YWnEnqY7g3HkhMqlc5HXFUiwB/view?usp=sharing>>

APÊNDICE B – Dicionário de dados

Tabela B.1 – Dicionário de dados descrevendo em detalhes cada atributo

BOL	Tipo Boletim Ocorrências
ANO	Ano Fato
DATA_COM	Data Comunicação Fato
HORA_COM	Horário Comunicação Fato
DATA_HORA	Data_Hora Fato
DATA	Data Fato
HORA	Horário Fato
DIA	Dia Numérico Fato
MES	Mês Numérico Fato
DEZ	Dezena Mês Fato
SEM	Dia da Semana Fato
FAH	Faixa 1 Hora Fato
FAD	Faixa 6 Horas Fato
LAT	Latitude
LON	Longitude
BAIRRO	Bairro
LOGRAD	Logradouro Ocorrência
NUM	Número Logradouro
MUN	Município
FIM	Tentado/Consumado Nat Principal
MEI	Desc Longa Meio Utilizado
BEN	Desc Longa Subgrupo Complemento Nat
LOC	Descrição Longa Local Imediato
VIA	Descrição Longa Tipo Local
CAU	Causa Presumida
	Data_Hora Comunicação Fato
	Mês Fato Resumido
	Código Grupo Nat Principal
	Desc Longa Grupo Nat Principal
	Código Classe Nat Principal
	Desc Longa Classe Nat Principal
	Código Subclasse Nat Principal
	Desc Longa Subclasse Nat Principal
	Grupo Causa Presumida
	Qtde Ocorrências

APÊNDICE C – Quadro com as Principais funções e pacotes do R

Tabela C.1 – Principais funções, pacotes do R e respectivos parâmetros utilizados.

TÉCNICA	FUNÇÃO/PAR METROS	PACOTE
Spectral Analysis	<code>spec.mtm(nw = 4, k = 7, dtUnits = "hour", Ftest = T, plot = F, jackknife = T)</code>	multitaper 1.0-15
ACF/PACF	<code>ggAcf(lag.max = 56, ci = 0.99)</code> <code>ggPacf(lag.max = 56, ci = 0.99)</code>	forecast 8.20
model/STL	<code>model(stl = STL(n, robust = TRUE))</code>	fabletools 0.3.2 feasts 0.3.0

Referências

- ABNTEX. *Tribunal de Justiça do Distrito Federal e dos Territórios*. 2016. Disponível em: <<https://www.tjdft.jus.br/institucional/imprensa/campanhas-e-produtos/direito-facil/edicao-semanal/furto-e-roubo#:~:text=Roubo-,Art.,a%20dez%20anos%20e%20multa.>> Acesso em: 14 mar. 2023. 18
- BOROWIK, Z. M. W. G.; CICHOSZ, P. Time series analysis for crime forecasting. 2019. 16
- C. EUGENIO C., D. T. C.; V., A. A data-driven approach for spatio-temporal crime predictions in smart cities. 2018. 16
- CLEVELAND ROBERT B., e. a. Stl: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics* — [s. l.], v. 6, n. 1, p. 3–73, 1990. 22, 39
- CORNISH D. B., . C. R. V. E. *The Reasoning Criminal: Rational Choice Perspectives on Offending*. Springer: Book, 1986. 14
- CORNISH DEREK B.; CLARKE, R. V. "the reasoning criminal: Rational choice perspectives on offending."springer. 2017. 14
- DAVIS, R. A. e. a. Count time series: A methodological review. *Journal of the American Statistical Association* — [s. l.], v. 116, n. 535, p. 1533–1547, 2021. 39
- DECRETO NO 3.178. *Determina a suspensão temporária dos Alvarás de Localização e Funcionamento e autorizações emitidos para realização de atividades com potencial de aglomeração de pessoas para enfrentamento da Situação de Emergência Pública causada pelo agente Coronavírus - COVID-19*. 2020. Disponível em: <https://www.itabira.mg.gov.br/arquivo/legislacao/decreto_3178_2020>. Acesso em: 07/07/2023. 20
- FELSON MARCUS; COHEN, L. Routine activity and rational choice: Advances in criminological theory, volume 5. transaction publishers. 2008. 13
- FUKANG, Z. Zero-inflated poisson and negative binomial integer-valued garch models. *Journal of Statistical Planning and Inference* — [s. l.], v. 142, n. 4, p. 826–839, 2012. Disponível em: <<http://dx.doi.org/10.1016/j.jspi.2011.10.002>.> Acesso em: 25 ago. 2023. 33
- GREEN, J. A. Too many zeros and/or highly skewed? a tutorial on modelling health behaviour as count data with poisson and negative binomial regression. *Health Psychology and Behavioral Medicine* — [s. l.], v. 9, n. 1, p. 436–455, 2021. 33
- HOU ZHAOLONG ZENG, X. H. J. H. M. Investigating the impact of the covid-19 pandemic on crime incidents number in different cities. 2022. 16
- HYNDMAN R.J.; ATHANASOPOULOS, G. *Forecasting: principles and practice. 3rd. ed.* Melbourne, Australia: OTexts. E-book, 2021. Disponível em: <<https://otexts.com/fpp3/>> .> 22

- LEAL1, L. G.-J. M. F. Impacto da pandemia da covid-19 nos padrões de crimes no município de curitiba. 2022. Disponível em: <<https://sol.sbc.org.br/index.php/erbd/article/download/21404/21228/>>. Acesso em: 18 set. 2023. 15, 16
- LOBO, A. d. P. Covid-19 epidemic in brazil: Where are we at? international journal of infectious diseases. *DOI* — Revista da Faculdade de Serviço Social da UERJ, v. 97, p. 382–385, 2020. Disponível em: <<https://doi.org/10.1016/j.ijid.2020.06.044>>. Acesso em: 01 ago. 2023. 20
- MAGUIRE, M.; MORGAN, R.; REINER, R. (Ed.). *The Oxford Handbook of Criminology*. 5. ed. [S.l.]: Oxford University Press, 2012. 13
- MOSES, S. . *Spectral Analysis of Signals*. Upper Saddle River, New Jersey: Prentice Hall: E-Book, 2005. Disponível em: <https://www.maths.lu.se/fileadmin/maths/personal_staff/Andreas_Jakobsson/StoicaM05.pdf>. 20
- RAHIM, K. J. Applications of multitaper spectral analysis to nonstationary data. 2014. 29, 42
- SHUMWAY, S. D. S. R. H. *Time Series Analysis and its Applications: with R Examples*. Fourth Edied. Pittsburgh: PA: Springer, . E-book, 2016. Disponível em: <<https://www.stat.pitt.edu/stoffer/tsa4/>>. 21
- SIDORENKO, R. . Minimum bias multiple taper spectral estimation. *IEEE* — Transactions on Signal Processing, v. 43, n. 1, p. 188–195, 1995. 21
- SILVA F., e. a. Visual analytics for crime data: A systematic literature review. journal of crime science, 11(2), 81-105. 2021. 23
- SJ, L. B. T. Prophet: forecasting at scale. the journal of machine learning research; 18(1):616-623. 2018. 17
- TOMA A., B. O. . J. I. M. Crime forecasting using time series models. procedia computer scienc; 112; 2030-2035. 2017. 17
- VALENT, J. *Covid-19: governo declara transmissão comunitária em todo o país. Agência Brasil*. 2020. Disponível em: <<https://agenciabrasil.etc.com.br/saude/noticia/2020-03/covid-19-governo-declara-transmissao-comunitaria-em-todo-o-pais>>. Acesso em: 07/07/2023. 20
- WHO. *World Health Organization. Coronavirus disease (COVID-19) Pandemic. Geneva: WHO*. 2019. Disponível em: <<https://www.who.int/emergencies/diseases/novel-coronavirus-2019>>. Acesso em: 03/07/2023. 20
- WILSON JAMES Q., e. K. G. L. Broken windows: The police and neighborhood safety. the atlantic monthly. 1982. 13