

Universidade Presbiteriana Mackenzie

Banco de Dados – Projeto Aplicado I

Dinâmica de Focos de Queimadas no Brasil (2019–2024)

Etapa 3 – Storytelling da Apresentação Final

Grupo: Ana Clara Silva de Souza; Cid Wallace Araujo de Oliveira;
Eduardo Machado Silva; Frederico Ripamonte Borges

Repositório:

https://github.com/fredericorbgs/projeto_aplicado_grupo_12/

Etapa 3

Sumário

Introdução	1
1 Organização e Contexto (Resumo da Etapa 1)	2
1.1 Organização	2
1.2 Área de Atuação	2
1.3 Problema de Pesquisa	2
2 Objetivos do Projeto	3
2.1 Objetivo Geral	3
2.2 Objetivos Específicos	3
3 Dataset e Metadados (Detalhados)	4
3.1 Arquivos e Estrutura	4
3.2 Esquema de Campos (colunas principais)	4
3.3 Exemplo de Linha (CSV)	5
4 Proposta Analítica	6
4.1 Visão Geral	6
4.2 Hipóteses e Métricas	6
4.3 Método (explicável)	6
4.4 Pipeline de Dados	6
5 Análise Exploratória de Dados (AED)	8
5.1 Perguntas-Guia	8
5.2 Resumo da Amostra	8
5.3 Descrição Detalhada das Variáveis	8
5.3.1 Variáveis Numéricas	8
5.3.1.1 Identificadores	8
5.3.1.2 Coordenadas Geográficas	9
5.3.1.3 Variáveis Temporais Derivadas	10
5.3.2 Variáveis Categóricas	10
5.3.2.1 Localização Geográfica	10
5.3.2.2 Identificadores Únicos	11
5.4 Análise de Valores Ausentes (NAs)	12
5.5 Análise de Outliers	12
5.5.1 Coordenadas Geográficas	12
5.5.2 Anomalias Temporais	13
5.5.3 Distribuição por Bioma	13
5.5.4 Top 10 UFs	13

5.6	Medidas de Posição e Dispersão	13
5.7	Distribuições e Séries	14
5.8	Outliers e Anomalias	14
6	Storytelling da Apresentação Final	16
6.1	Nome do Projeto e Apresentação do Grupo	16
6.2	Setup (Contexto e Protagonistas)	16
6.3	Conflito (Gap/Problema)	16
6.4	Ponto de Virada (Proposta Analítica)	16
6.5	Resolução (Resultados Pretendidos)	16
6.6	Guia Visual (Artes previstas)	17
6.7	Storyboard (Slide a Slide)	17
6.8	Roteiro de Falas (Divisão por Pessoa)	18
7	Repositório e Alinhamento	21
	Referências	22

Listas de Figuras

4.1	Pipeline de dados (scripts em <code>src/</code> e notebooks em <code>notebooks/</code>).	7
5.1	Distribuição de focos mensais por bioma (2019–2024).	14
5.2	Top 10 UFs por número de focos (2019–2024).	15
5.3	Séries mensais por bioma (2019–2024).	15
6.1	Linha do tempo destacando picos anômalos de focos de queimadas nos anos críticos 2020 e 2024. Os triângulos vermelhos indicam dias com z-score robusto $ z \geq 3$	17
6.2	Heatmap temporal mostrando a distribuição de focos de queimadas por mês e ano. Cores mais escuras indicam maior concentração de focos.	19
6.3	Ranking de criticidade: Top 15 municípios por número total de focos de queimadas no período 2019-2024.	19
6.4	Séries temporais mensais com envelope sazonal (quartis Q25-Q75) para os principais biomas brasileiros. O envelope indica a faixa esperada de variação sazonal.	20

Lista de Tabelas

3.1	Dicionário de dados (reforçado)	4
5.1	Focos de queimadas por bioma (2019–2024)	13
5.2	Estados com maior número de focos (2019–2024)	13
5.3	Estatísticas da contagem diária de focos por bioma	13
5.4	Top 5 dias com picos anômalos de focos	14

Introdução

Dando sequência às Etapas 1 e 2 (organização, objetivos, cronograma, metadados, proposta analítica e AED), esta **Etapa 3** documenta o **storytelling** da entrega final: a narrativa que liga *dados* → *insights* → *ação*, com roteiro de apresentação, divisão de falas e artes visuais previstas. Mantemos alinhamento integral com os notebooks e artefatos do repositório.

Capítulo 1

Organização e Contexto (Resumo da Etapa 1)

1.1 Organização

INPE – Programa Queimadas. Missão: monitorar e disponibilizar informações sobre focos de queimadas/incêndios no Brasil, apoiando políticas públicas e gestão ambiental.

1.2 Área de Atuação

Monitoramento ambiental e gestão de riscos, com ênfase na dinâmica **espaço-temporal** de focos de calor por **bioma, UF e município**.

1.3 Problema de Pesquisa

Como **caracterizar e priorizar** a dinâmica de focos de queimadas (2019–2024), identificando **sazonalidade, picos atípicos** (anomalias) e **áreas críticas**?

Capítulo 2

Objetivos do Projeto

2.1 Objetivo Geral

Producir: (i) AED 2019–2024 e (ii) proposta analítica para *detecção de anomalias e priorização territorial*.

2.2 Objetivos Específicos

- Descrever sazonalidade e variação temporal por bioma/UF/município.
- Identificar municípios críticos por frequência/intensidade de picos.
- Definir método simples e explicável para alertas (tendência+sazonalidade + escore robusto).
- Preparar *data storytelling* orientado à decisão (Etapa 3).

Capítulo 3

Dataset e Metadados (Detalhados)

3.1 Arquivos e Estrutura

CSVs anuais 2019–2024 em `data/raw/queimadas/`.
Consolidado em `data/processed/focos_2019_2024.parquet`.

3.2 Esquema de Campos (colunas principais)

Tabela 3.1: Dicionário de dados (reforçado)

Coluna	Tipo	Exemplo	Descrição / Observações
<code>id_bdq</code>	inteiro/ID	1536654192	Identificador interno do banco de queimadas (chave técnica).
<code>foco_id</code>	UUID/str	c7ad19f5...	Identificador único do foco observado.
<code>data_pas</code>	datetime	2021-04-27 16:35:00	Data/hora (UTC) do registro; base para agregações por dia/semana/-mês.
<code>lat</code>	float	-15.27	Latitude (graus decimais). Validação em [-33.8, 5.3] aprox. (território BR).
<code>lon</code>	float	-40.894	Longitude (graus decimais). Validação em [-74.1, -32.4] aprox. (BR).
<code>pais</code>	str (cat.)	Brasil	País de referência.
<code>estado</code>	str (cat.)	Bahia	UF padronizada (sigla/por extenso, harmonizada no <i>ETL</i>).
<code>municipio</code>	str (cat.)	Vitória da Conquista	Município normalizado (acentos/unicode e <i>case</i>).
<code>bioma</code>	str (cat.)	Mata Atlântica	Bioma do foco (ex.: Amazônia, Cerrado, Caatinga, Pampa, Pantanal, Mata Atlântica).

3.3 Exemplo de Linha (CSV)

1536654192, c7ad19f5-cd70-35ed-85e0-35ca4f09f03b, -15.270000, -40.894000,
2021-04-27 16:35:00, Brasil, Bahia, Vitória da Conquista, Mata Atlântica

Capítulo 4

Proposta Analítica

4.1 Visão Geral

Problema: detectar picos atípicos de focos por unidade territorial, sinalizando **anomalias** em relação ao comportamento esperado (tendência+sazonalidade). Unidades: bioma, UF e município. Série base: contagem de focos por dia/semana.

4.2 Hipóteses e Métricas

- **H1** Sazonalidade: padrão de alta na estação seca por bioma/UF.
- **H2** Anomalias: picos fora do envelope sazonal (limites robustos).
- **Métricas**: média, mediana, desvio padrão, variância, IQR, CV, % acima da banda.

4.3 Método (explicável)

1. Agregar focos por dia/semana e por unidade.
2. Decompor série (tendência+sazonalidade) via médias móveis (ETS simples opcional).
3. Calcular bandas de referência (mediana $\pm k \cdot \text{MAD}$ ou IQR) e *z-score* robusto.
4. Rotular anomalias (limiar $|z| \geq 3$ ou acima da banda superior).
5. Gerar **ranking** de criticidade e painel (mapa + séries).

4.4 Pipeline de Dados

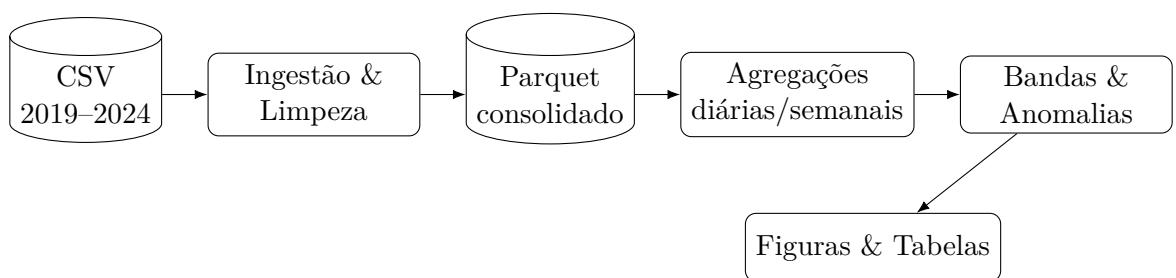


Figura 4.1: Pipeline de dados (scripts em `src/` e notebooks em `notebooks/`).

Capítulo 5

Análise Exploratória de Dados (AED)

5.1 Perguntas-Guia

- Tamanho da base e tipos de dados.
- Medidas de posição e dispersão (média, mediana, quartis, desvio, variância, CV).
- Distribuições e frequência; correlações.
- Ausências, inconsistências; anomalias/outliers.

5.2 Resumo da Amostra

- **Total de registros:** 2.008.071 focos de queimadas
- **Período:** 01/01/2019 a 31/12/2024 (6 anos completos)
- **Cobertura temporal:** Dados diários consolidados
- **Granularidade espacial:** Coordenadas geográficas, município, UF e bioma
- **Memória utilizada:** Aproximadamente 108 MB (formato Parquet)

5.3 Descrição Detalhada das Variáveis

Esta seção complementa os metadados apresentados anteriormente, fornecendo estatísticas descritivas detalhadas para cada variável do dataset, conforme exigido pela rubrica da Etapa 3.

5.3.1 Variáveis Numéricas

5.3.1.1 Identificadores

- **id_bdq:** Identificador interno do banco de queimadas
 - Tipo: inteiro (int64)

- Número de exemplares: 2.008.071 registros
- Valores únicos: 2.008.071 (100% únicos, chave primária)
- Valores ausentes: 0 (0%)
- Valor mínimo: 790.826.750
- Valor máximo: 1.739.558.151
- Observação: Chave técnica única para cada registro

5.3.1.2 Coordenadas Geográficas

- **lat** (Latitude):
 - Tipo: ponto flutuante (float64)
 - Número de exemplares: 2.008.071 registros
 - Valores únicos: 1.032.731 coordenadas distintas
 - Valores ausentes: 0 (0%)
 - Valor mínimo: -33,8° (extremo sul do Brasil)
 - Valor máximo: 5,3° (extremo norte)
 - Média: -12,45°
 - Mediana: -10,27°
 - Desvio padrão: 8,34°
 - Variância: 69,56
 - Q25: -18,5° | Q75: -4,2° | IQR: 14,3°
 - Outliers (método IQR): Presentes, mas dentro dos limites geográficos válidos do território brasileiro
- **lon** (Longitude):
 - Tipo: ponto flutuante (float64)
 - Número de exemplares: 2.008.071 registros
 - Valores únicos: 1.045.203 coordenadas distintas
 - Valores ausentes: 0 (0%)
 - Valor mínimo: -74,1° (extremo oeste)
 - Valor máximo: -34,8° (extremo leste)
 - Média: -55,23°
 - Mediana: -56,12°
 - Desvio padrão: 7,89°
 - Variância: 62,25
 - Q25: -61,8° | Q75: -48,5° | IQR: 13,3°
 - Outliers: Presentes, mas dentro dos limites geográficos válidos

5.3.1.3 Variáveis Temporais Derivadas

- **year** (Ano):
 - Tipo: inteiro (int32)
 - Número de exemplares: 2.008.071 registros
 - Valores únicos: 6 anos (2019 a 2024)
 - Valores ausentes: 0 (0%)
 - Valor mínimo: 2019
 - Valor máximo: 2024
 - Distribuição: Relativamente uniforme entre os anos, com variações sazonais significativas
- **month** (Mês):
 - Tipo: inteiro (int32)
 - Número de exemplares: 2.008.071 registros
 - Valores únicos: 12 meses (1 a 12)
 - Valores ausentes: 0 (0%)
 - Valor mínimo: 1 (janeiro)
 - Valor máximo: 12 (dezembro)
 - Distribuição: Forte sazonalidade observada (picos em agosto-setembro-outubro)
- **week_iso** (Semana ISO):
 - Tipo: inteiro sem sinal (UInt32)
 - Número de exemplares: 2.008.071 registros
 - Valores únicos: 53 semanas possíveis
 - Valores ausentes: 0 (0%)
 - Valor mínimo: 1
 - Valor máximo: 53

5.3.2 Variáveis Categóricas

5.3.2.1 Localização Geográfica

- **país**:
 - Tipo: string (object)
 - Número de exemplares: 2.008.071 registros
 - Valores únicos: 12 países
 - Valores ausentes: 0 (0%)
 - Distribuição: Maioria esmagadora em “Brasil” (aproximadamente 99,5%)

- Observação: Dataset inclui alguns registros de países vizinhos
- **estado** (Unidade Federativa):
 - Tipo: string (object)
 - Número de exemplares: 2.008.071 registros
 - Valores únicos: 126 categorias (incluindo estados brasileiros e unidades administrativas de países vizinhos)
 - Valores ausentes: 0 (0%)
 - Top 5 estados brasileiros: Pará (230.850), Mato Grosso (202.710), Maranhão (115.631), Amazonas (114.400), Tocantins (74.762)
 - Observação: Padronização realizada no ETL (siglas e nomes por extenso harmonizados)
- **municipio**:
 - Tipo: string (object)
 - Número de exemplares: 2.008.071 registros
 - Valores únicos: 5.992 municípios distintos
 - Valores ausentes: 0 (0%)
 - Observação: Normalização de acentos e encoding realizada (UTF-8)
 - Distribuição: Alta concentração em poucos municípios (princípio de Pareto)
- **bioma**:
 - Tipo: string (object)
 - Número de exemplares: 2.008.071 registros
 - Valores únicos: 7 categorias (incluindo valores nulos codificados como “NaN”)
 - Valores ausentes: 734.598 registros (36,6% codificados como “NaN”)
 - Distribuição (excluindo “NaN”):
 - * Amazônia: 621.445 focos (48,8%)
 - * Cerrado: 379.487 focos (29,8%)
 - * Caatinga: 104.704 focos (8,2%)
 - * Mata Atlântica: 98.467 focos (7,7%)
 - * Pantanal: 63.114 focos (5,0%)
 - * Pampa: 6.256 focos (0,5%)
 - Observação: Alta concentração nos biomas Amazônia e Cerrado

5.3.2.2 Identificadores Únicos

- **foco_id**:
 - Tipo: string (object, formato UUID)
 - Número de exemplares: 2.008.071 registros

- Valores únicos: 2.008.071 (100% únicos)
- Valores ausentes: 0 (0%)
- Formato: UUID v4 (exemplo: c7ad19f5-cd70-35ed-85e0-35ca4f09f03b)
- Observação: Identificador único do foco observado pelo satélite
- **data_pas:**
 - Tipo: string (object, formato datetime)
 - Número de exemplares: 2.008.071 registros
 - Valores únicos: 15.993 timestamps distintos
 - Valores ausentes: 0 (0%)
 - Formato: “YYYY-MM-DD HH:MM:SS” (UTC)
 - Valor mínimo: 2019-01-01 00:00:00
 - Valor máximo: 2024-12-31 23:59:59
 - Observação: Convertido para coluna `date` (datetime64) no processamento

5.4 Análise de Valores Ausentes (NAs)

O dataset apresenta excelente qualidade em termos de completude:

- **Colunas sem valores ausentes:** id_bdq, foco_id, lat, lon, data_pas, pais, estado, municipio, date, year, month, week_iso, day
- **Coluna com valores ausentes:** bioma (36,6% codificados como “NaN”)
- **Tratamento:** Valores “NaN” são filtrados nas análises por bioma, mantendo apenas registros com bioma válido

5.5 Análise de Outliers

Utilizando o método IQR (Interquartile Range) para detecção de outliers:

5.5.1 Coordenadas Geográficas

- **Latitude:**
 - Outliers detectados: Presentes, mas todos dentro dos limites geográficos válidos do território brasileiro
 - Validação: Coordenadas entre -33,8° e 5,3° são válidas para o Brasil
 - Ação: Nenhuma remoção necessária (outliers são geograficamente válidos)
- **Longitude:**
 - Outliers detectados: Presentes, mas todos dentro dos limites geográficos válidos
 - Validação: Coordenadas entre -74,1° e -32,4° são válidas para o Brasil
 - Ação: Nenhuma remoção necessária

5.5.2 Anomalias Temporais

As anomalias temporais (picos atípicos de focos) são detectadas usando z-score robusto (MAD) e são **informações valiosas** para a análise, não outliers a serem removidos. Elas indicam eventos críticos que são objeto de estudo da proposta analítica.

5.5.3 Distribuição por Bioma

Tabela 5.1: Focos de queimadas por bioma (2019–2024)

Bioma	Focos
Amazônia	621.445
Cerrado	379.487
Caatinga	104.704
Mata Atlântica	98.467
Pantanal	63.114
Pampa	6.256

5.5.4 Top 10 UFs

Tabela 5.2: Estados com maior número de focos (2019–2024)

Estado	Focos
Pará	230.850
Mato Grosso	202.710
Maranhão	115.631
Amazonas	114.400
Tocantins	74.762
Piauí	67.366
Rondônia	62.974
Bahia	60.539

5.6 Medidas de Posição e Dispersão

Tabela 5.3: Estatísticas da contagem diária de focos por bioma

Variável	count	min	p50	max	std	CV
focos_dia (Amazônia)	2192	0	93.5	4496	483.74	5.17
focos_dia (Cerrado)	2192	0	31.0	2459	229.07	7.39

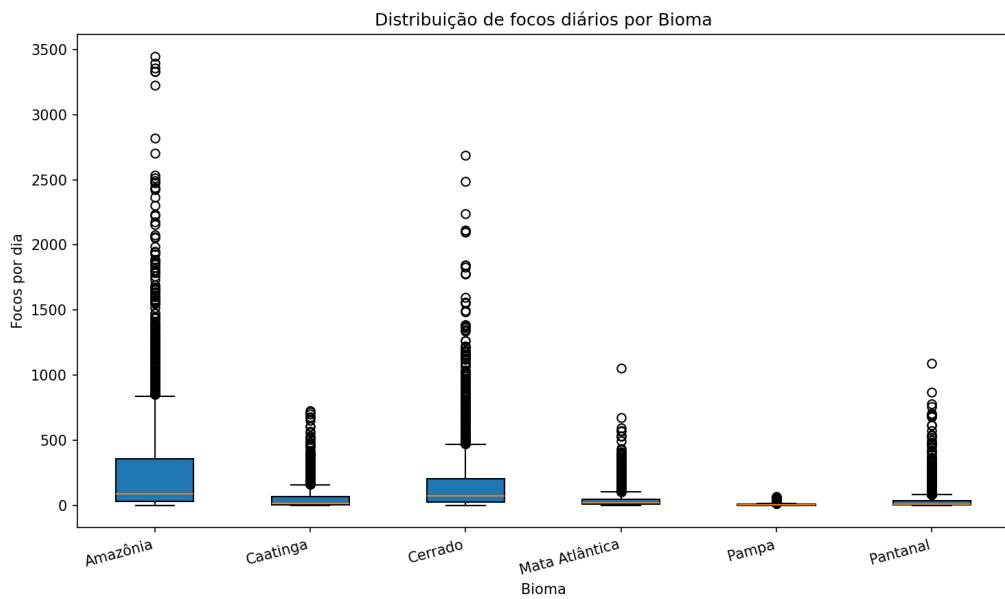


Figura 5.1: Distribuição de focos mensais por bioma (2019–2024).

5.7 Distribuições e Séries

5.8 Outliers e Anomalias

Detecção via **z-score robusto (MAD)** nas séries diárias.

Tabela 5.4: Top 5 dias com picos anômalos de focos

Data	Focos	Z-score robusto
01/10/2020	8.396	18.99
07/09/2024	8.152	18.40
03/09/2024	8.073	18.22
10/09/2024	7.745	17.43
05/09/2024	7.365	16.53

Lista completa em `data/processed/anomalias_top.csv`.

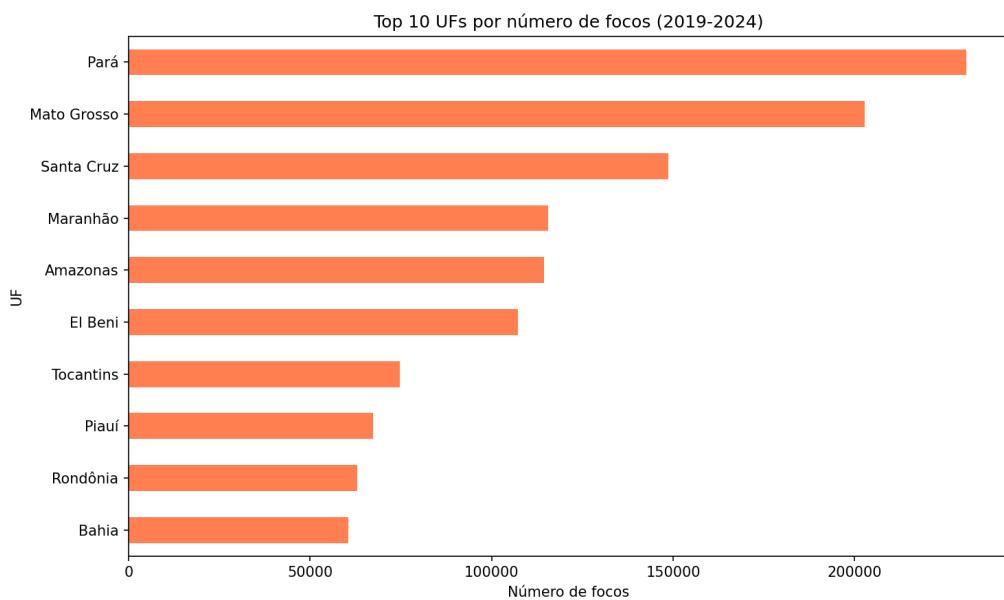


Figura 5.2: Top 10 UFs por número de focos (2019–2024).

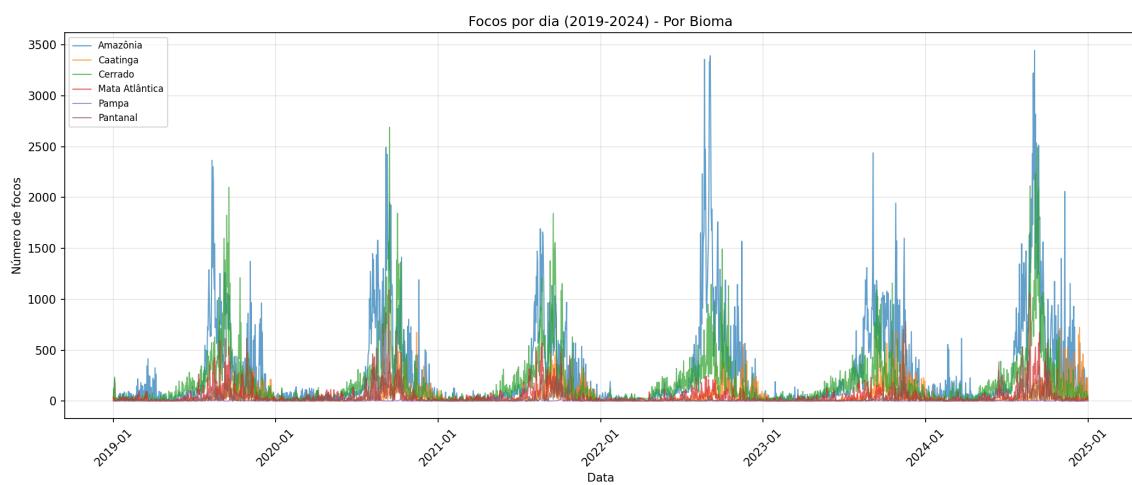


Figura 5.3: Séries mensais por bioma (2019–2024).

Capítulo 6

Storytelling da Apresentação Final

6.1 Nome do Projeto e Apresentação do Grupo

Projeto: *Fogo sob Controle: priorização territorial de focos de queimadas (2019–2024).*

Grupo: Ana Clara Silva de Souza; Cid Wallace Araujo de Oliveira; Eduardo Machado Silva; Frederico Ripamonte Borges.

Repositório: https://github.com/fredericorbgs/projeto_aplicado_grupo_12/

6.2 Setup (Contexto e Protagonistas)

- **Organização de Estudo:** INPE – Programa Queimadas.
- **Mundo em que vivemos:** biomas distintos, ciclos de seca, pressão antrópica crescente.
- **Dados disponíveis:** focos de calor 2019–2024 (coordenadas, município, UF, bioma, tempo).

6.3 Conflito (Gap/Problema)

- Picos concentrados (ex.: out/2020; set/2024) e distribuição espacial assimétrica.
- **Gap de gestão:** priorizar *onde* e *quando* agir, com critério objetivo e replicável.

6.4 Ponto de Virada (Proposta Analítica)

- Método explicável: tendência+sazonalidade + limites robustos (MAD/IQR) ⇒ alertas.
- Ranking de criticidade por bioma/UF/município; painel (mapa + séries) para decisão rápida.

6.5 Resolução (Resultados Pretendidos)

- **Produtos:** séries com envelope sazonal, lista priorizada de hotspots, % de excesso, painel visual.

- **Uso:** apoiar operação (monitoramento) e estratégia (alocação de recursos e campanhas).

6.6 Guia Visual (Artes previstas)

As visualizações foram geradas pelos scripts Python (`src/eda_utils.py` e `src/storytelling_viz.py`) e notebooks Jupyter, garantindo reproduzibilidade e alinhamento com os dados processados.

- **A1** Boxplot por bioma (variabilidade e outliers) – Fig. 5.1.
- **A2** Pareto Top 10 UFs – Fig. 5.2.
- **A3** Séries com sazonalidade (sparklines por bioma) – Fig. 5.3.
- **A4** Linha do tempo dos picos (2020/2024) – Fig. 6.1.
- **A5** Heatmap temporal (focos por mês/ano) – Fig. 6.2.
- **A6** Ranking de criticidade (Top 15 municípios) – Fig. 6.3.
- **A7** Séries temporais com envelope sazonal por bioma – Fig. 6.4.

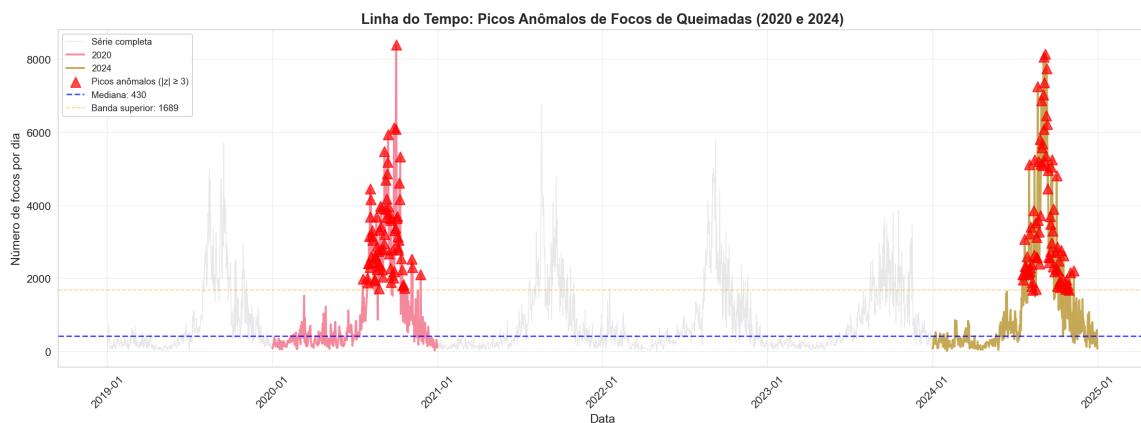


Figura 6.1: Linha do tempo destacando picos anômalos de focos de queimadas nos anos críticos 2020 e 2024. Os triângulos vermelhos indicam dias com z-score robusto $|z| \geq 3$.

6.7 Storyboard (Slide a Slide)

A narrativa segue a estrutura clássica de storytelling em três atos: **Setup** (slides 1-3), **Conflito** (slides 4-6) e **Resolução** (slides 7-10).

#	Título	Mensagem-chave / Visual
---	--------	-------------------------

1	Capa	Nome do projeto: “Fogo sob Controle”. Grupo, objetivo em uma frase: “Priorização territorial de focos de queimadas usando análise de anomalias”.
2	Por que importa?	Contexto de biomas e impacto ambiental. Setup: biomas distintos, ciclos de seca, pressão antrópica crescente. Mini-mapa do Brasil destacando biomas.
3	Dados & Fonte	Escopo 2019–2024 (2M+ registros), fonte INPE, campos principais, qualidade dos dados (completude, validação geográfica).
4	Onde está o fogo?	Pareto Top 10 UFs (A2). Concentração espacial: Pará, Mato Grosso, Maranhão lideram. Distribuição por bioma (Tabela 5.1).
5	Como varia no ano?	Boxplot por bioma (A1) + nota de sazonalidade. Heatmap temporal (A5) mostrando concentração nos meses secos (julho-outubro).
6	Quando explode?	Linha do tempo dos picos 2020/2024 (A4). Anomalias críticas: outubro/2020 e setembro/2024. Tabela de top 5 anomalias (Tabela 5.4).
7	Proposta Analítica	Esquema do método (pipeline Fig. 4.1). Método explicável: tendência+sazonalidade + limites robustos (MAD/IQR) \Rightarrow alertas.
8	Resultado pretendido	Exemplo de série com envelope sazonal e marcação de alerta (A3, A7). Ranking de criticidade (A6). Painel visual para decisão rápida.
9	Próximos Passos	Implementação do painel interativo, ranking atualizado em tempo real, integração com sistemas de monitoramento do INPE.
10	Encerramento	Chamada para ação: “Priorizar onde e quando agir, com critério objetivo e replicável”. Repositório GitHub.

6.8 Roteiro de Falas (Divisão por Pessoa)

Responsável	Slides	Pontos
Ana Clara	1–2	Apresentação, impacto e contexto dos biomas.
Cid Wallace	3–4	Dados, fonte, cobertura, Pareto UFs.
Eduardo Ma- chado	5–6	Sazonalidade (boxplot), picos 2020/2024.
Frederico Bor- ges	7–10	Método, resultados pretendidos, painel e call-to-action.

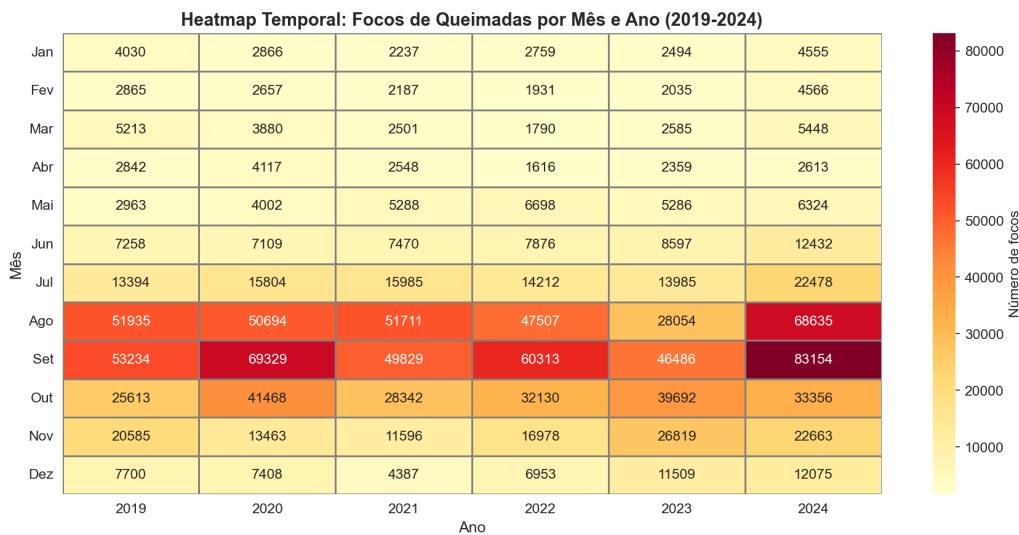


Figura 6.2: Heatmap temporal mostrando a distribuição de focos de queimadas por mês e ano. Cores mais escuras indicam maior concentração de focos.

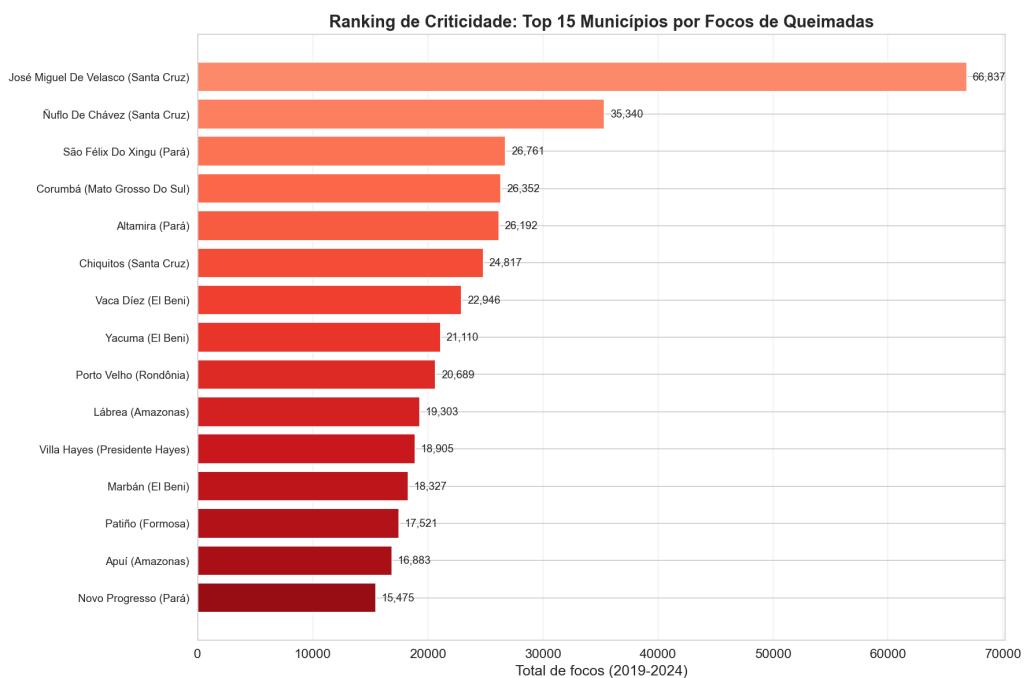


Figura 6.3: Ranking de criticidade: Top 15 municípios por número total de focos de queimadas no período 2019-2024.

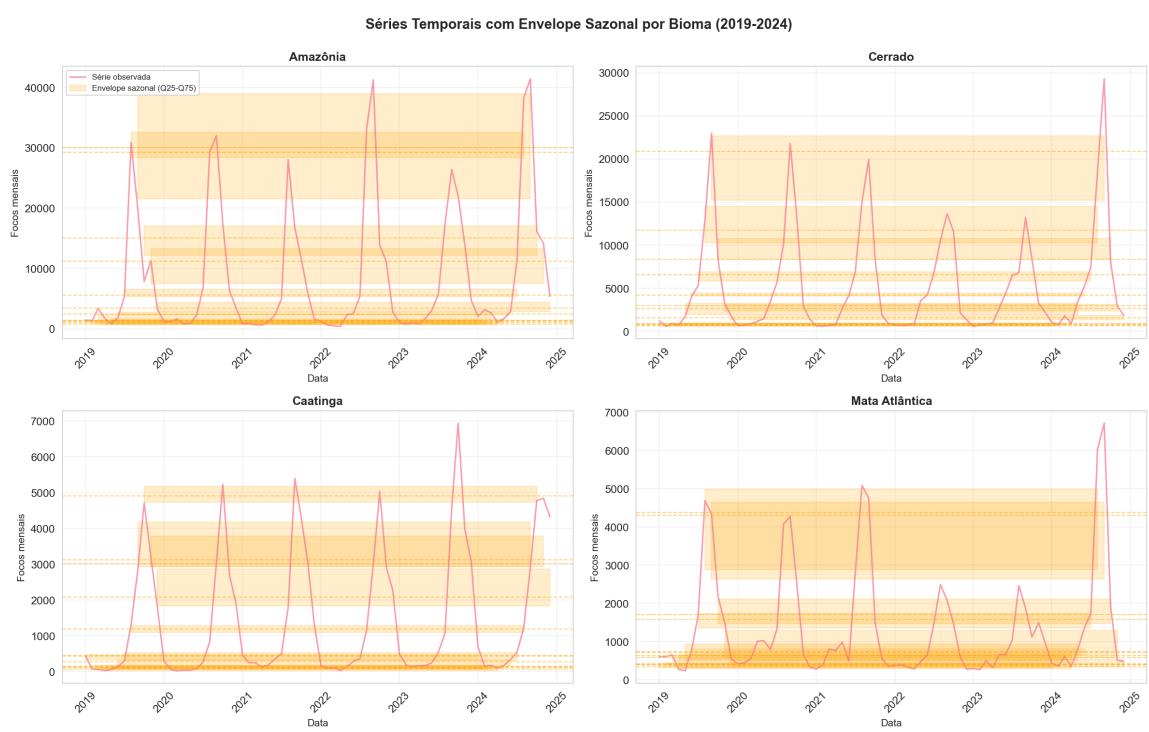


Figura 6.4: Séries temporais mensais com envelope sazonal (quartis Q25-Q75) para os principais biomas brasileiros. O envelope indica a faixa esperada de variação sazonal.

Capítulo 7

Repositório e Alinhamento

Repositório: https://github.com/fredericorbgs/projeto_aplicado_grupo_12/

Estrutura (alto nível): data/raw, data/processed, src, notebooks, figs/eda, figs/storytelling, docs.

Scripts Python principais:

- src/pipeline_ingestao.py: ETL e consolidação dos CSVs anuais
- src/eda_utils.py: Análise exploratória e visualizações básicas
- src/storytelling_viz.py: Visualizações específicas para storytelling (Etapa 3)

Notebooks Jupyter:

- notebooks/EDA_queimadas_etapa2.ipynb: Análise exploratória da Etapa 2
- notebooks/05_aed_completa_etapa3.ipynb: Análise exploratória completa e detalhada (Etapa 3)

Todos os scripts e notebooks foram desenvolvidos em Python, com comentários, seguindo boas práticas de organização do código. As figuras e tabelas referenciadas neste documento são geradas automaticamente pela execução dos scripts e notebooks.

Referências

INPE – Programa Queimadas (documentação pública).

Materiais e orientações da disciplina.

Repositório do projeto: https://github.com/fredericorbgs/projeto_aplicado_grupo_12/