

Universidade Presbiteriana Mackenzie

Banco de Dados – Projeto Aplicado I

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

Etapa 2 – Proposta Analítica e Análise Exploratória

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/

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 |
| 4.5 Resultados Esperados | 6 |
| 5 Análise Exploratória de Dados (AED) | 8 |
| 5.1 Perguntas-Guia | 8 |
| 5.2 Resumo da Amostra | 8 |
| 5.2.1 Distribuição por Bioma | 8 |
| 5.2.2 Distribuição por Unidade Federativa (Top 10) | 8 |
| 5.3 Medidas de Posição e Dispersão | 9 |
| 5.4 Distribuições e Correlações | 9 |
| 5.4.1 Análise dos Top 10 Estados | 9 |
| 5.5 Séries Temporais e Sazonalidade | 9 |
| 5.5.1 Padrões Observados | 9 |
| 5.6 Outliers e Anomalias | 10 |
| 5.6.1 Principais Anomalias Detectadas | 10 |
| 5.6.2 Interpretação | 10 |
| 5.7 Síntese da Análise Exploratória | 11 |
| 6 Repositório e Alinhamento | 13 |

Lista de Figuras

| | | |
|-----|--|----|
| 4.1 | Pipeline de dados da proposta analítica (scripts em <code>src/</code> e notebook em <code>notebooks/</code>). | 7 |
| 5.1 | Distribuição de focos mensais por Bioma (2019-2024). O gráfico evidencia a variabilidade e concentração de focos nos diferentes biomas brasileiros, com destaque para Amazônia e Cerrado. | 10 |
| 5.2 | Top 10 UF's por número de focos (2019-2024). Estados da região Norte e Centro-Oeste concentram a maior parte dos focos de queimadas. | 11 |
| 5.3 | Séries temporais mensais de focos por Bioma (2019-2024). Observa-se clara sazonalidade, com picos concentrados nos meses de seca (julho a outubro), especialmente nos biomas Amazônia e Cerrado. | 12 |

Lista de Tabelas

| | | |
|-----|---|----|
| 3.1 | Dicionário de dados (reforçado) | 4 |
| 5.1 | Focos de queimadas por bioma (2019-2024) | 9 |
| 5.2 | Estados com maior número de focos (2019-2024) | 9 |
| 5.3 | Exemplo de estatísticas (saída do notebook) | 10 |
| 5.4 | Top 5 dias com picos anômalos de focos | 11 |

Introdução

Dando sequência à Etapa 1 (organização, objetivos, cronograma e metadados), esta **Etapa 2** inclui: (i) a **Proposta Analítica** detalhada — com problema, hipóteses, métrica-alvo e *pipeline* de dados — e (ii) a **Análise Exploratória de Dados (AED)**, alinhada aos notebooks Python presentes no repositório. O foco é descrever variáveis, sumarizar medidas de posição/dispersão, distribuição, dados ausentes e outliers, ilustrando com gráficos.

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

Produzir: (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 + score 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/`.

3.2 Esquema de Campos (colunas principais)

Tabela 3.1: Dicionário de dados (reforçado)

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

Qualidade e Tratamento: normalização de encoding (UTF-8), remoção/ajuste de coordenadas inválidas, padronização de nomes de municípios/UF, *cast* de datas e geração de chaves derivadas (dia, semana ISO, mês, ano, bioma_uf).

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 de análise:** bioma, UF e município. **Série base:** contagem de focos por dia (ou semana) por unidade.

4.2 Hipóteses e Métricas

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

4.3 Método (explicável)

1. Agregar focos por dia/semana e por unidade.
2. Decompor série em 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 ajustada).
5. Gerar **ranking** de criticidade e painel (mapa + séries).

4.4 Pipeline de Dados

4.5 Resultados Esperados

- Séries por bioma/UF/município com envelope sazonal e % de excesso.

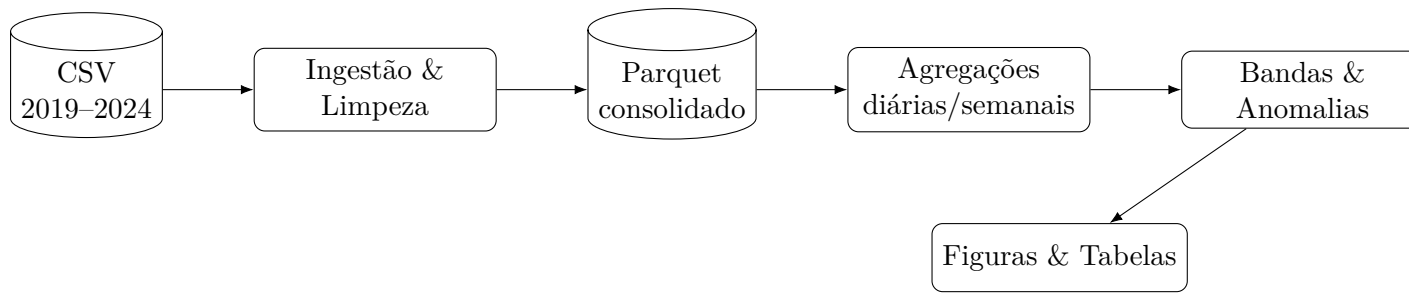


Figura 4.1: Pipeline de dados da proposta analítica (scripts em `src/` e notebook em `notebooks/`).

- Lista priorizada de unidades com picos recentes e recorrência histórica.
- Painel com mapas (coroplético) e *sparklines* temporais.

Capítulo 5

Análise Exploratória de Dados (AED)

5.1 Perguntas-Guia

- Quantidade de linhas/colunas e tipos de dados.
- Medidas de posição e dispersão (média, mediana, quartis, desvio, variância, CV).
- Distribuição e frequência; correlações.
- Valores ausentes, inconsistências; anomalias/outliers.

5.2 Resumo da Amostra

Características gerais da base de dados:

- **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

5.2.1 Distribuição por Bioma

Os dados mostram concentração significativa em determinados biomas brasileiros:

5.2.2 Distribuição por Unidade Federativa (Top 10)

Observa-se forte concentração nas regiões Norte e Centro-Oeste do Brasil, áreas que abrangem principalmente os biomas Amazônia e Cerrado.

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

| Bioma | Número de focos |
|----------------|-----------------|
| Amazônia | 621.445 |
| Cerrado | 379.487 |
| Caatinga | 104.704 |
| Mata Atlântica | 98.467 |
| Pantanal | 63.114 |
| Pampa | 6.256 |

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

| Estado | Número de 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 |

5.3 Medidas de Posição e Dispersão

5.4 Distribuições e Correlações

A análise de distribuições revela padrões importantes nos dados de queimadas por bioma. O boxplot abaixo mostra a distribuição de focos mensais para cada bioma brasileiro entre 2019 e 2024.

5.4.1 Análise dos Top 10 Estados

A concentração espacial dos focos é um fator crítico para priorização de políticas públicas. A figura a seguir apresenta os 10 estados com maior número de focos registrados no período analisado.

5.5 Séries Temporais e Sazonalidade

A análise temporal dos focos de queimadas revela padrões sazonais consistentes ao longo dos anos. A série temporal por bioma demonstra a dinâmica dos focos ao longo do período 2019-2024, permitindo identificar picos sazonais e tendências de longo prazo.

5.5.1 Padrões Observados

- **Sazonalidade:** Concentração de focos nos meses secos (julho-outubro)
- **Picos históricos:** Anos de 2020 e 2024 apresentam picos significativos

Tabela 5.3: Exemplo de estatísticas (saída do notebook)

| Variável | count | min | p50 | max | std | CV |
|----------------------------|--|-----|-----|-----|-----|----|
| focos_dia (bioma=Amazônia) | <i>(auto-preenchido pelo notebook)</i> | | | | | |
| focos_dia (bioma=Cerrado) | <i>idem</i> | | | | | |

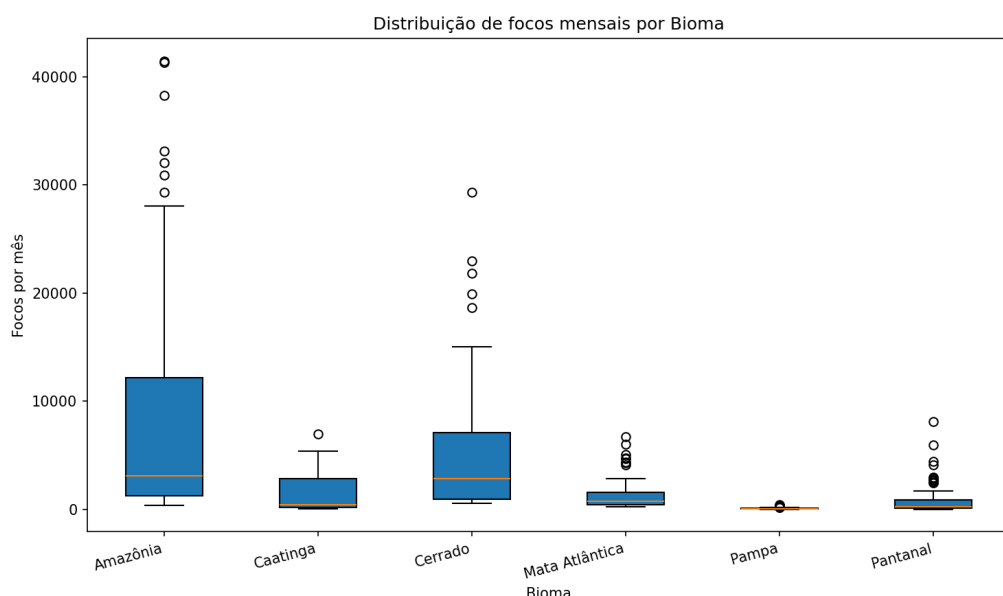


Figura 5.1: Distribuição de focos mensais por Bioma (2019-2024). O gráfico evidencia a variabilidade e concentração de focos nos diferentes biomas brasileiros, com destaque para Amazônia e Cerrado.

- **Variação por bioma:** Amazônia e Cerrado mostram maior variabilidade
- **Tendência:** Necessário monitoramento contínuo para detecção de mudanças estruturais

5.6 Outliers e Anomalias

A detecção de anomalias foi realizada utilizando **z-score robusto** (**MAD - Median Absolute Deviation**) sobre as séries temporais diárias. O método identifica dias com número de focos significativamente acima do esperado.

5.6.1 Principais Anomalias Detectadas

Os dias com maior número de focos (anomalias críticas) identificados no período foram:

5.6.2 Interpretação

- **Pico histórico:** Outubro de 2020 registrou o maior pico do período
- **Ano recente crítico:** Setembro de 2024 concentra 4 dos 5 maiores picos

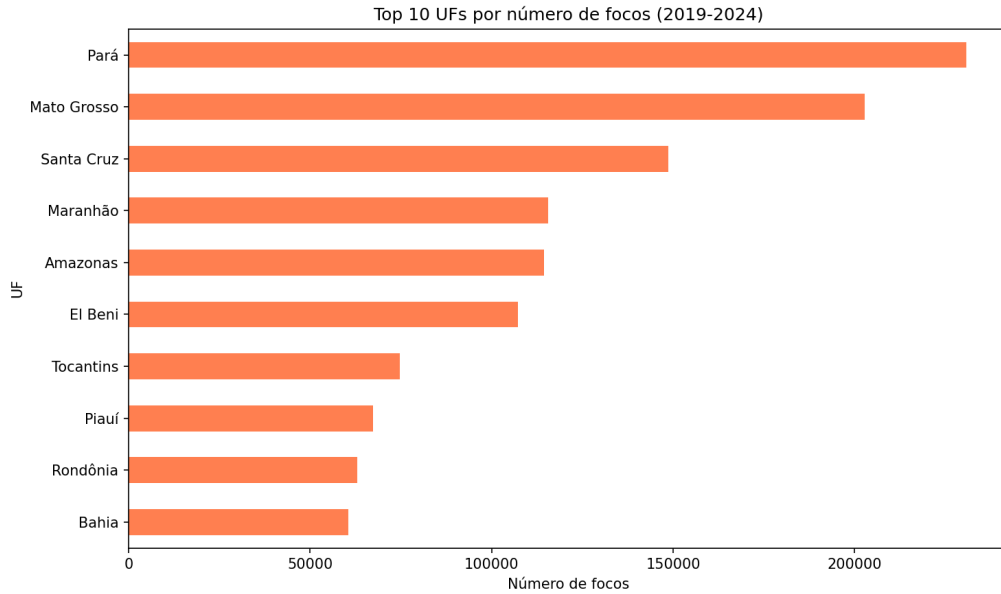


Figura 5.2: Top 10 UF's por número de focos (2019-2024). Estados da região Norte e Centro-Oeste concentram a maior parte dos focos de queimadas.

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 |

- **Padrão temporal:** Anomalias concentram-se no período seco (agosto-outubro)
- **Implicações:** Eventos críticos demandam resposta emergencial coordenada

Lista completa de anomalias disponível em `data/processed/anomalias_top.csv`, identificando 50 eventos críticos para análise aprofundada e ações prioritárias.

5.7 Síntese da Análise Exploratória

A análise exploratória de dados realizada sobre mais de 2 milhões de registros de focos de queimadas entre 2019 e 2024 revelou os seguintes achados principais:

1. **Concentração espacial:** Amazônia e Cerrado concentram aproximadamente 50% dos focos, com estados como Pará, Mato Grosso e Maranhão liderando as ocorrências.
2. **Padrão sazonal consistente:** Clara concentração de focos nos meses de seca (julho a outubro), com variações anuais significativas que sugerem influência de fatores climáticos e antropogênicos.

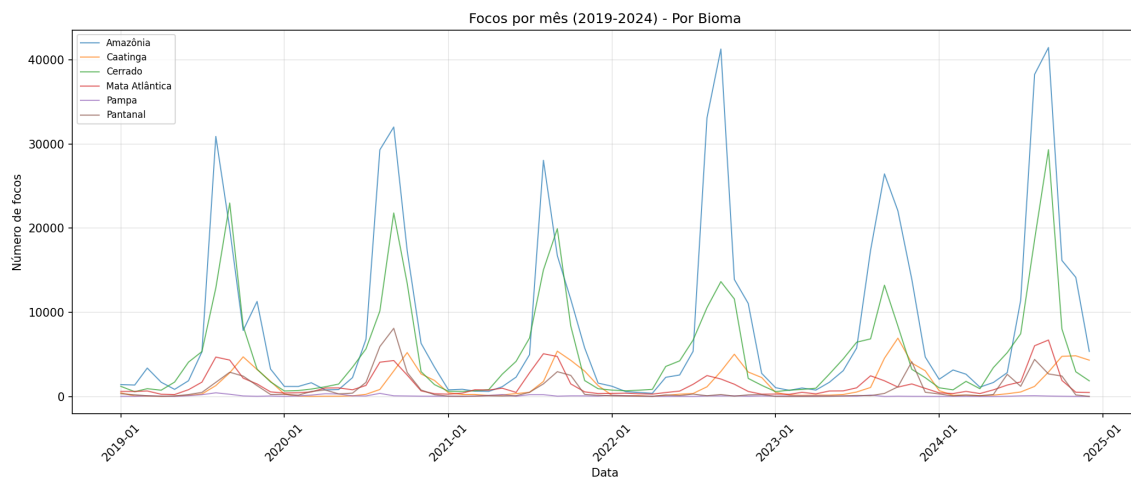


Figura 5.3: Séries temporais mensais de focos por Bioma (2019-2024). Observa-se clara sazonalidade, com picos concentrados nos meses de seca (julho a outubro), especialmente nos biomas Amazônia e Cerrado.

3. **Eventos críticos:** Identificação de picos anômalos, especialmente em outubro/2020 e setembro/2024, que demandam investigação detalhada sobre causas e impactos.
4. **Qualidade dos dados:** Base consolidada apresenta estrutura adequada para análises preditivas e detecção de anomalias, com tratamento apropriado de valores ausentes e validação de coordenadas geográficas.
5. **Implicações para modelagem:** Os padrões identificados justificam a proposta analítica de decomposição de séries temporais e detecção de anomalias por métodos estatísticos robustos.

Os artefatos gerados (figuras, tabelas e CSVs processados) fornecem base sólida para a próxima etapa de modelagem e *data storytelling* orientado à decisão.

Capítulo 6

Repositório e Alinhamento

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

Estrutura (alto nível): data/raw, data/processed, src, notebooks, figs/eda, docs.
Scripts e notebook foram desenvolvidos em Python, com comentários, boas práticas e geração automática de artefatos referenciados neste relatório.

Referências

INPE – Programa Queimadas (documentação pública).

Materiais e orientações do componente curricular.

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