

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
5.3 Medidas de Posição e Dispersão	8
5.4 Distribuições e Correlações	9
5.5 Séries Temporais e Sazonalidade	9
5.6 Outliers e Anomalias (prévia)	9
6 Repositório e Alinhamento	10
Referências	11

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	Distribuições e boxplots.	9
5.2	Séries temporais diárias/semanais.	9

Lista de Tabelas

3.1	Dicionário de dados (reforçado)	4
5.1	Exemplo de estatísticas (saída do notebook)	8

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
<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).

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.

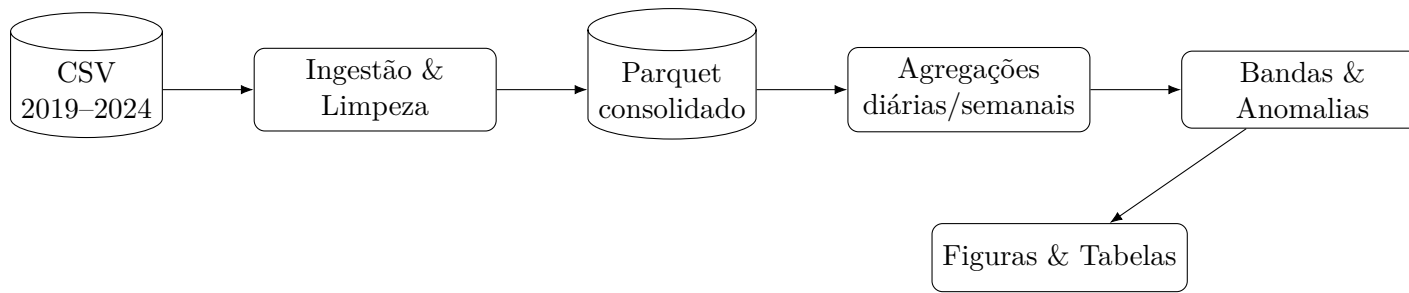


Figura 4.1: Pipeline de dados da proposta analítica (scripts em `src/` e notebook em `notebooks/`).

- 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

Linhas/colunas, tipos e ausentes (gerado por notebooks/EDA_queimadas_etapa2.ipynb):

- Tabela **data/processed/resumo_colunas.csv** (#linhas, tipo, NAs, únicos, min/-max).
- Estatísticas gerais em **data/processed/estatisticas_gerais.csv**.

5.3 Medidas de Posição e Dispersão

Tabela 5.1: 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>					

Histograma de focos/dia e Boxplots por bioma (gerados em `figs/eda/`)

Figura 5.1: Distribuições e boxplots.

Séries por bioma/UF com médias móveis e envelope sazonal

Figura 5.2: Séries temporais diárias/semanais.

5.4 Distribuições e Correlações

5.5 Séries Temporais e Sazonalidade

5.6 Outliers e Anomalias (prévia)

Identificação por **IQR** e por **z-score robusto** (MAD). Listas em `data/processed/anomalias_top.csv`.
Gráficos com pontos destacados em `figs/eda/`.

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/