

# APLICAÇÃO DA ANÁLISE EXPLORATÓRIA DE DADOS PARA MONITORAMENTO DE QUEIMADAS NA AMAZÔNIA LEGAL: ESTRATÉGIAS PARA O GERENCIAMENTO PREVENTIVO DE RISCOS AMBIENTAIS

*Data de submissão: 08/10/2024*

*Data de aceite: 01/11/2024*

**Carlos Vinícius Monteiro de Souza**

<http://lattes.cnpq.br/2427890169958659>

**Fabiann Matthaus Dantas Barbosa**

<http://lattes.cnpq.br/3769505772789674>

**José Elislande Breno de Souza Linhares**

<http://lattes.cnpq.br/0846691032369254>

**Joethe Moraes de Carvalho**

<http://lattes.cnpq.br/8157292652509113>

**RESUMO.** Este artigo apresenta um relatório técnico do uso de algumas temáticas modernas na literatura científica, entre elas a Análise Exploratória de Dados, o emprego de Métodos Estatísticos e Visualização de Dados, aplicados no contexto da previsão de queimadas e focos de incêndio. Tendo por base o referencial teórico que foi obtido através de diversas strings de buscas e teses indicadas, foram destrinchados pesquisas e artigos que abordassem uma ou mais das temáticas citadas anteriormente. No final, foi demonstrado que os Métodos da Análise Exploratória possuem uma infinidade e versatilidade de métodos e abordagens para os fins da pesquisa.

**PALAVRAS-CHAVE:** Análise Exploratória

de Dados, Métodos Estatísticos, Visualização de Dados, Previsão de Queimadas.

**ABSTRACT.** This article presents a technical report on the use of some modern themes in scientific literature, including Exploratory Data Analysis, the use of Statistical Methods and Data Visualization, applied in the context of fire and fire outbreak prediction. Based on the theoretical framework that was obtained through several search strings and indicated theses, research and articles that addressed one or more of the previously mentioned themes were unraveled. In the end, it was demonstrated that Exploratory Analysis Methods have an infinity and versatility of methods and approaches for research purposes.

**KEYWORDS:** Exploratory Data Analysis, Statistical Methods, Data Visualization, Wildfire Prediction

## 1 | INTRODUÇÃO

O manejo do fogo é uma prática muito disseminada e empregada por agricultores, fazendeiros e pecuaristas, tendo como principais objetivos:

transformar áreas florestais em pastagens; em solo para cultivo; controle da proliferação de plantas indesejadas; extração de recursos naturais (madeira, minérios, látex); etc, tudo isso a um baixo custo (MACHADO, R. *et al.*, 2022).

O monitoramento das queimadas é de suma importância para o combate ao fogo, reduzir custos operacionais, e minimizar os danos ambientais ocasionados. A análise do comportamento temporal de tais eventualidades permite estabelecer correlações entre variáveis que influenciam a ocorrência e propagação das chamas, sendo vital para compreender as causas naturais e antrópicas (OLIVEIRA *et al.*, 2019).

A Análise Exploratória de Dados (AED) tem por objetivo a expansão do conhecimento dos pesquisadores sobre uma população a partir da amostra selecionada. Dito isso, a AED é descrita como um conglomerado de metodologias para extração, exploração, manipulação e descrição dos conjuntos de dados (LOPES, G. *et al.*, 2019).

A partir da elaboração e execução dos métodos da AED, é possível fazer uma avaliação do quão aplicável são os modelos de previsão em séries temporais. Com isso, modelos estatísticos, de aprendizagem de máquina, e de aprendizado por reforço, são geralmente empregados para os mais variados fins, como alternativas para realização de previsões em séries temporais (DE OLIVEIRA; 2022).

Com o contexto da pesquisa demonstrado, surge a seguinte questão para este trabalho: *“Como se encontra a literatura atual da aplicabilidade de AED no âmbito das queimadas e focos de incêndio no âmbito da pesquisa científica, e de que forma ela é utilizada para mitigação dos riscos ambientais?”*.

## 2 | TRABALHOS RELACIONADOS

No decorrer da pesquisa, foram encontradas diversas pesquisas relacionadas tanto à metodologia quanto a questões ambientais. Dentre essas pesquisas, algumas serviram de base para a elaboração do roteiro, outras para auxiliar no desenvolvimento da pesquisa, as demais foram utilizadas para embasamento teórico fundamental no respaldo científico.

### 2.1 Processo Investigativo e Seleção

Aqui, é preciso realizar uma amostragem dos trabalhos, para selecioná-los de um conjunto maior, denominado “população”. A população, em termos estatísticos, representa o conjunto de indivíduos que serão alvo da pesquisa, e a “amostra” é um subconjunto da população estudada. Este tópico tem por finalidade encontrar pesquisas semelhantes à problemática e estudo, utilizando a técnica de AED.

Neste sentido, a pesquisa definiu a população-alvo dos trabalhos com os termos ‘análise exploratória de dados’ e ‘modelos de aprendizado’. A partir da definição das populações, os termos devem ser aplicados em locais de busca adequados para trabalhos

acadêmicos, no caso da pesquisa, foram procurados pesquisas no *Google Scholar*, tanto trabalhos nacionais quanto estrangeiros. O Quadro 1 apresenta as strings que trouxeram mais resultados:

Biblioteca/ Repositório	String de Busca propriamente
<i>Google Scholar (PT)</i>	("Análise Exploratória de Dados" OR "Análise Exploratória") AND ("Focos de Incêndio" OR "Queimadas") OR ("Modelos Preditivos" OR "Algoritmos Preditivos")
	("Análise Exploratória de Dados" OR "Análise Exploratória") AND ("Focos de Incêndio" OR "Queimadas")
Google Scholar (EN)	("Exploratory Data Analysis" OR "Exploratory Analysis") AND ("Fires" OR "Burnings") OR ("Predictive Models" OR "Predictive Algorithms")

Quadro 1 - Strings de Busca

## 2.2 Análise e Especificação

Durante a seleção das pesquisas, alguns critérios de inclusão e exclusão de trabalhos foram previamente selecionados, para facilitar a obtenção e busca dos trabalhos. Em seguida, foi realizada uma análise exploratória com o objetivo de obter as contribuições desejadas. O intervalo de tempo dos artigos localizam-se entre 2016 a 2023, e ao realizar as buscas nas revistas e periódicos apontados pelo *Google Scholar*, os artigos foram analisados de acordo com os critérios da pesquisa. Veja o Quadro 2 com todos os critérios utilizados:

Critérios de Inclusão (C.I.)	Critérios de Exclusão (C.E.)
C.I. 1: Disponível em plataforma de acesso público e gratuito.	C.E. 1: Necessário fazer login ou requisitar o texto na plataforma para acessar o trabalho.
C.I. 2: Trabalhos que englobam o uso de tecnologias da Engenharia de Software.	C.E. 2: Trabalhos que não estejam em português ou em inglês.
C.I. 3: Trabalhos que abordam as temáticas "Análise Exploratória de Dados", "Predição em Série Temporal" e "Queimadas/Focos de Incêndio".	C.E. 3: Trabalhos que não mostram claramente os resultados obtidos na pesquisa.

Quadro 2 – Critérios de Inclusão e Exclusão

## 2.3 Análise dos Trabalhos

Ao analisar os trabalhos obtidos, percebe-se que um subgrupo se destaca por maior aprofundamento metodológico, e nos resultados obtidos, no grau de complexidade, e na similaridade com a proposta do trabalho aqui apresentado. E para diferenciar trabalhos que podem ou não contribuir no desenvolvimento deste estudo, foram necessários o uso de 3 filtros para extração dos trabalhos que forneceram ideias e/ou métodos para trabalhar com

dados e AED.

1. <u>Leitura do Título</u>	O título apresenta algumas das propriedades da pesquisa, além de ser a primeira informação visível
2. <u>Leitura do Resumo e das Palavras-chave</u>	O resumo apresenta o que foi feito, como foi feito, alguns dos resultados, e ideias empregadas na pesquisa, o que fornece um panorama geral da pesquisa no todo.
3. <u>Leitura integral do Trabalho</u>	Por fim, essa atividade visa procurar os principais métodos e conceitos inerentes ao trabalho.

Quadro 3 – Filtros de Pesquisas

## 2.4 Análise dos Resultados

Os trabalhos encontrados com o *Google Scholar* são em sua grande maioria voltados para uso de AED e tecnologias similares para fins específicos fora do contexto ambiental. Dentre os quais, podemos citar Linhares, J. (2022) e Ramos, Samara e Barros, Bárbara (2022) que abordam problemas distintos com o uso de análise exploratória, visando extrair resultados para seus respectivos objetos de pesquisa.

Além do que, algumas teses contribuíram substancialmente na formulação do roteiro, desenvolvimento e embasamento do trabalho escrito, tais como produzidos por Barbosa, A. (2021), Carrilho, A. (2022), Fantinatti, G. F. (2021), e Dias, A. (2019) onde as queimadas foram tratadas como objeto de estudo, e a análise de dados foi a medida utilizada para encontrar informações e soluções para o problema em si.

Além destes, Queiroz, G. R. *et al* (2017) e Costa, J. *et al* (2021) também tratam do estudo dos focos de calor e das queimadas, porém com uso de ferramentas e conceitos oriundos de áreas correlatas a AED, dentre elas a *Big Data* - referente a manipulação de grandes volumes de dados - e técnicas de Visualização de Dados - um método para geração de informações através de dados - que contribuíram significativamente para a pesquisa proposta.

## 3 | METODOLOGIA

A abordagem do problema se enquadra em uma pesquisa que coleta dados qualitativos. Isso inclui a coleta de dados quantitativos, com a finalidade de produzir uma análise de dados climáticos, podendo gerar dados qualitativos, em relação a interação dos usuários quanto a aplicação no *Streamlit*.

Este estudo utiliza alguns datasets disponibilizados pelo INPE e INMET, além de algumas ferramentas para Análise de Dados, tais como o *Python*, suas bibliotecas (*Pandas*, *Numpy*, *Matplotlib*, *Seaborn*, *Scikit-Learn*, etc), além dos ambientes de desenvolvimento *Google Colab* e a *Streamlit Cloud*.

### 3.1 Etapas de Desenvolvimento

O desenvolvimento da pesquisa foi dividido em 5 etapas, com o objetivo de flexibilizar as futuras mudanças a serem implementadas ou discutidas, para caso seja necessário fazer adições ou alterações na codificação do trabalho.

#### 3.1.1 Seleção do Dataset e das Ferramentas

Com relação aos *datasets*, o INPE fornece conjuntos de dados anuais em diversos formatos (*csv*; *GeoJson*; *kml*; *shapefile*) com acesso ao público em geral. O INMET disponibiliza os *datasets* de duas formas temporais: anualmente e semestralmente (porém, essa pesquisa visa trabalhar com dados coletados anualmente); e diferentemente do INPE, há apenas arquivos *csv*.

#### 3.1.2 Arquitetura do Projeto

A arquitetura elaborada é dividida em 3 ambientes. A Figura 5 ilustra a arquitetura pensada na criação e disponibilização da pesquisa:

1. *Google Drive*: local que armazena os arquivos *csv*;
2. *Google Collaboratory*: local de desenvolvimento da pesquisa;
3. *Streamlit*: ambiente/plataforma para disponibilizar os resultados da pesquisa;



Figura 1 - Arquitetura do Projeto

Fonte: Própria (2024)

#### 3.1.3 Amostragem

Uma amostragem estatística é uma distribuição probabilística, baseada na retirada de amostras, com um alto grau de aleatoriedade, e um tamanho fixo, na população pesquisada.

Distribuições probabilísticas ajudam o pesquisador a entender como os resultados de amostra variam entre as amostras. Foram adotados métodos de amostragem para busca e análise de informações preliminares dos *datasets* obtidos, e os resultados encontram-se nas figuras 2 e 3.

Amostragem Simples da coluna 'DiaSemChuva' em 2023

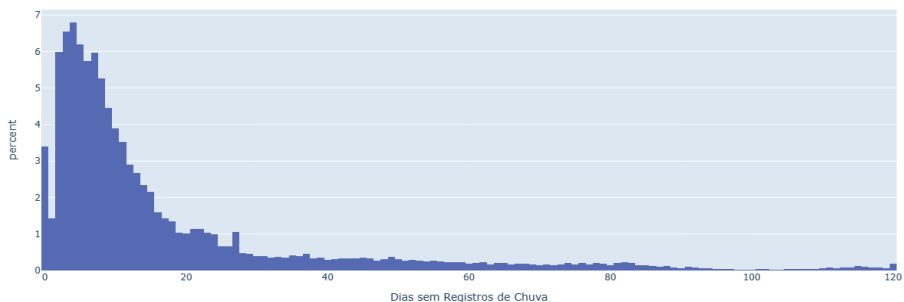


Figura 2 - Amostragem Simples

Fonte: Própria (2024)

Amostragem Probabilística de 'DiaSemChuva' em 2023

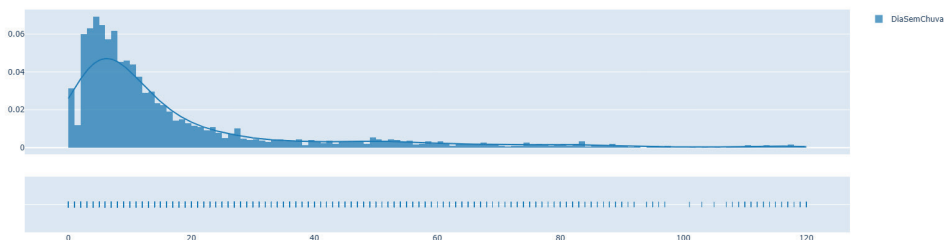


Figura 3 - Amostragem Probabilística

Fonte: Própria (2024)

Ambas as amostragens demonstram uma maior concentração dos dados registrados no intervalo (0, 30), e uma estabilização a partir de 40 dias; isso é, no decorrer de 2023, a ausência de chuvas variaram principalmente em torno de 30 dias ou menos. E tendo em vista as características do bioma amazônico, tamanha escassez de chuvas pode ter contribuído no aumento substancial dos focos de incêndio.

### 3.1.4 Tratamento dos Dados

A seção de 'Tratamento dos Dados' é fortemente vinculada ao processo 'E.T.L.' (*Extract, Transform, Load*), ou seja, uma fase predecessora a análise e exploração dos dados obtidos. Nessa etapa, o dataset é minuciosamente trabalhado, e dependendo das informações a serem obtidas, o pesquisador deverá produzir novos dados, editar os dados existentes, e/ou remover os dados que possam ocasionar erros e incongruências nas próximas fases da pesquisa.

Dentre os processos utilizados, foram feitos a remoção das colunas 'Satelite', 'Pais'

e 'Bioma' do *dataset*; renomeação da coluna 'DataHora' para 'Dia', substituição dos valores nulos por zero; conversão do tipo *object* para o tipo *float*; criação da coluna 'geometry' com uso das colunas 'latitude' e 'longitude'. E terminados esses processos, salva-se os *datasets* tratados para as próximas etapas.

### 3.1.5 Estatística Descritiva

A Estatística Descritiva é um ramo da ciência estatística que tem por propósito a descrição dos dados em questão, ou seja, converter dados em informação a ser explorada pelo pesquisador. Sua principal função é identificar características particulares, tais como média, desvio padrão, porcentagens, taxas, contagens e intervalo, etc (DIAS et al., 2019; FROST, 2024).

As figuras de 4 a 11 apresentam os resultados obtidos sobre os dados tratados anteriormente, fornecendo um panorama da atual situação climática e o comportamento das queimadas no decorrer do tempo.

	Estado	Media	Moda	Mediana	Amplitude	Variância	Desvio-Padrão
0	Acre	6.917304	7.0	6.0	33.0	29.873820	5.465695
1	Amapá	9.610902	9.0	9.0	45.0	33.943179	5.826077
2	Amazonas	28.014042	35.0	31.0	55.0	192.980233	13.891733
3	Maranhão	14.128248	4.0	8.0	109.0	260.558446	16.141823
4	Mato Grosso	68.001548	112.0	75.0	120.0	1647.658288	40.591357
5	Pará	35.025402	5.0	20.0	120.0	981.232174	31.324626
6	Rondônia	21.018682	5.0	13.0	115.0	364.969914	19.104186
7	Roraima	10.523889	3.0	10.0	52.0	45.791463	6.766939
8	Tocantins	41.348396	3.0	28.0	109.0	1191.808708	34.522583

Figura 4 - Descrição da coluna 'FRP' do dataset de 2020

Fonte: Própria (2024)

A Figura 4 demonstra de forma numérica o comportamento da incidência de focos de calor e potenciais riscos de incêndio nos 9 estados pertencentes à Amazônia Legal, que quando colocados em perspectiva no mundo real, podem ser apresentados conforme a Figura 5.

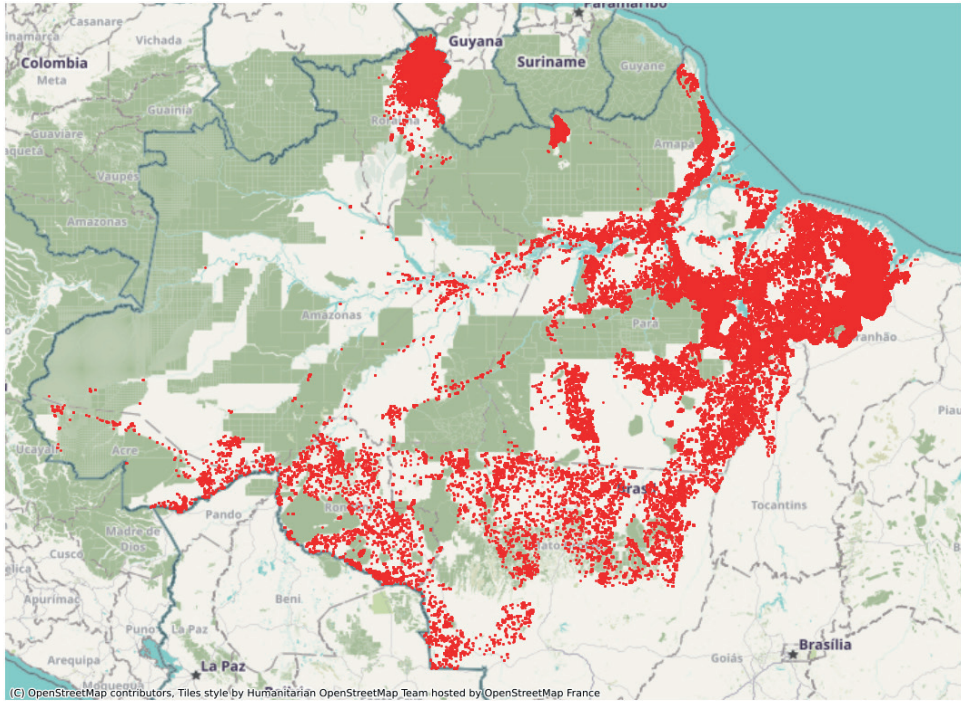


Figura 5 - Queimadas no Bioma em 2023

Fonte: Própria (2024)

As imagens produzidas para destacar os focos de incêndio, revelam que as queimadas concentram principalmente nas regiões Leste e Sul do bioma, principalmente nos estados de Mato Grosso, Pará e Maranhão - além da concentração de focos no nordeste de Roraima - e tal informação, se bem utilizada pelas entidades e órgãos competentes, poderá produzir medidas mais eficientes de enfrentamento a esse risco ambiental.



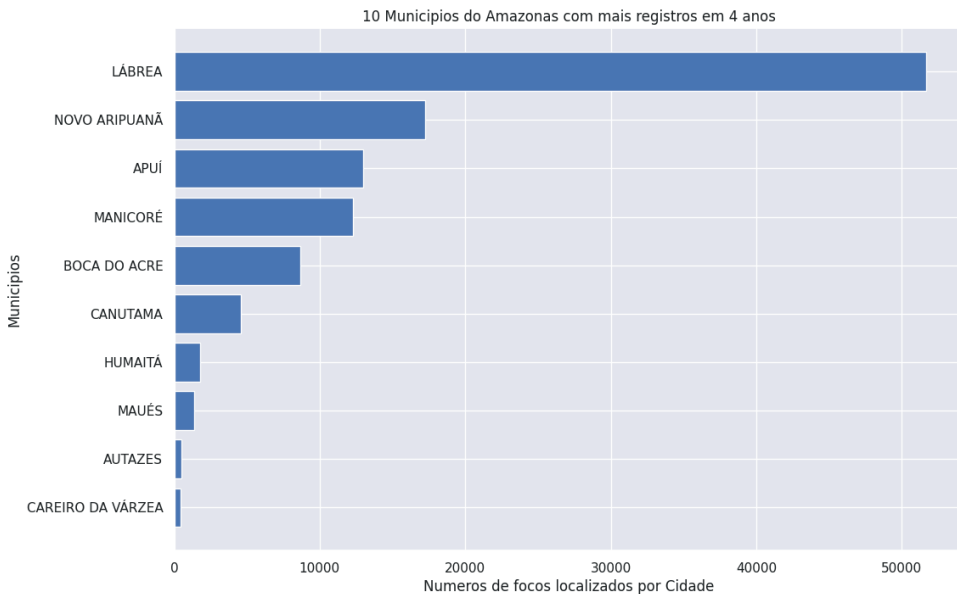


Figura 6 - Registros de queimadas dos municípios do Amazonas

Fonte: Própria (2024)

Além de ser possível visualizar a tendência de localidade das queimadas, também é possível (e necessário) quantificar os focos de queimadas em cada estado a partir dos municípios, e a Figura 7 apresenta um panorama geral dos municípios amazonenses que mais produziram queimadas de 2020 a 2023.

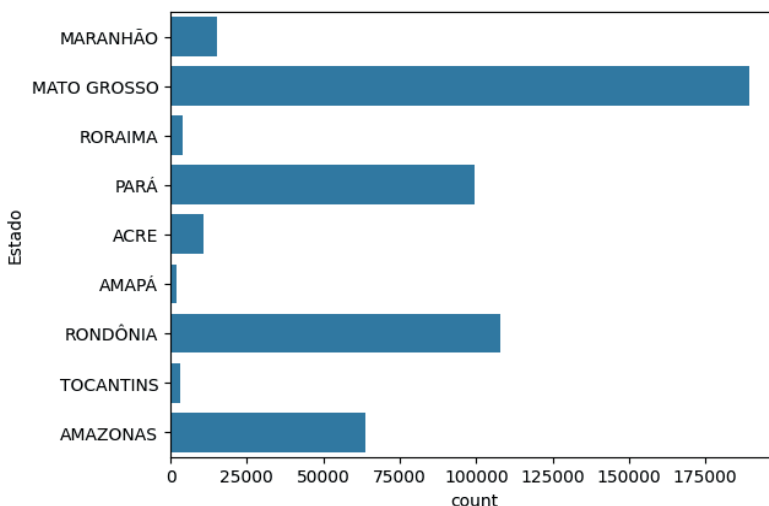


Figura 7 - Registros de queimadas e focos dos estados em 2021

Fonte: Própria (2024)

No que tange a exploração de dados quanto aos estados contidos na Amazônia, é preciso alfinetar o fato de que os estados de Mato Grosso, Pará, Maranhão e Rondônia foram os que mais se destacaram nos registros de focos de incêndio, mesmo não havendo uma tendência a ser anotada, sabe-se que esses 4 estados são os mais recorrentes em recordes de queimadas localizadas.

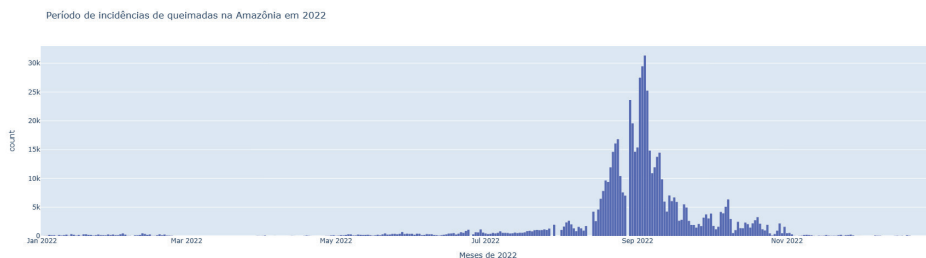


Figura 8 - Registros de queimadas e focos na Amazônia Legal em 2022

Fonte: Própria (2024)

Ao plotar todos os dados de um determinado *dataset* no decorrer do tempo em que estão dispostos, é possível quantificar a quantidade de todos os focos de incêndio e queimadas detectados em série temporal. Os *datasets* de 2020 a 2022 apresentaram um comportamento bem similar, onde as detecções concentram-se entre final de Julho e final de Outubro (diferente do que será apresentado na Figura 9).

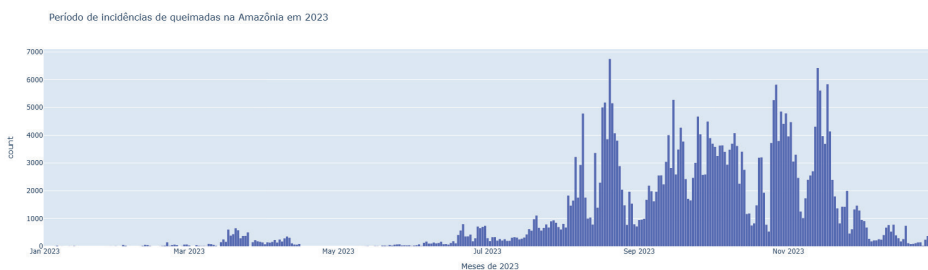


Figura 9 - Registros de queimadas e focos na Amazônia Legal em 2023

Fonte: Própria (2024)

Diferentemente dos demais *datasets*, março e abril apresentam um aumento substancial de registros, além do intervalo de tempo de intensificação das queimadas ter sido alterado, do início de julho até início de dezembro, e por fim, os picos de registros dos focos foram mais frequentes, dos quais foram registrados dias com mais de 5000 registros de queimadas.

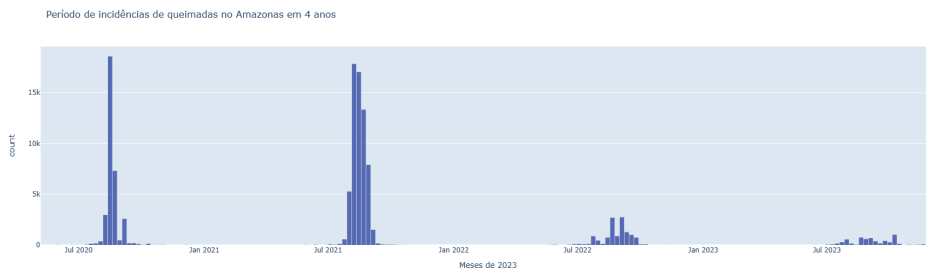


Figura 10 - Registros de queimadas e focos no Amazonas em 4 anos

Fonte: Própria (2024)

Outra forma de explorar os dados distribuídos por estados é a mesclagem dos 4 *datasets* para por em evidenciar as variações de registros no decorrer dos 4 anos, e no caso do Amazonas, nota-se que houve uma maior colaboração na destruição do seu próprio ecossistema principalmente nos anos de 2020 e 2021.

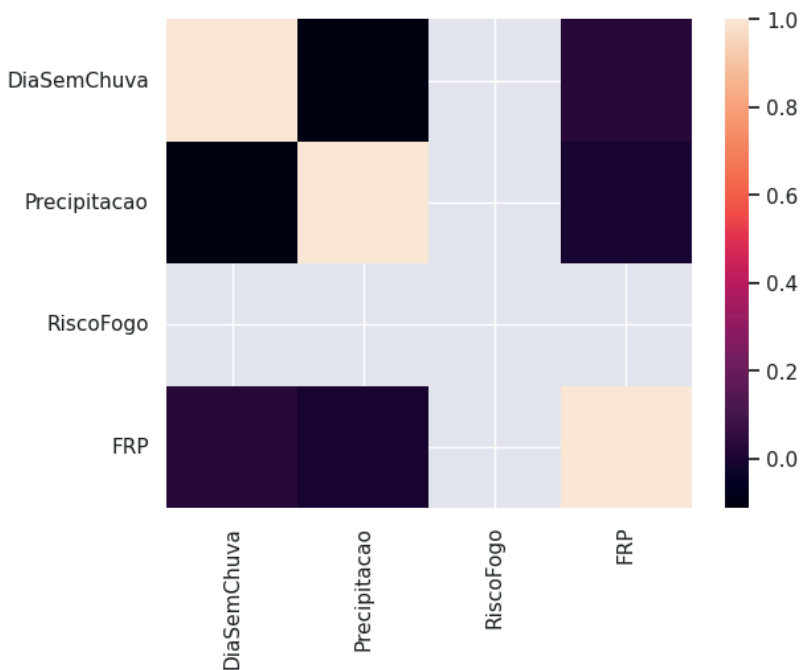


Figura 11 - Heatmap das *features* do *dataset* de 2020

Fonte: Própria (2024)

Os *Heatmaps* são uma outra forma de obtenção de informações, porém voltados para o lado mais científico da análise exploratória, onde o pesquisador visa entender como que as *features* dos *datasets* podem estar ou não correlacionadas (e até se possuem algum

nível de influência) umas com as outras.

E no caso desta pesquisa, as *features* numéricas têm um baixíssimo nível de correlação, o que por sua vez apresenta resultados a serem esperados na seção de Testes de Hipótese.

Esses são alguns dos resultados que puderam ser obtidos através dos métodos da AED, sendo possível extrair diversas informações a respeito das colunas-alvo. E concluído esses procedimentos, partimos para a aplicação dos Testes de Hipótese.

### 3.1.6 Testes de Hipóteses

Testes de hipótese são uma ramificação da estatística que avalia a precisão de teorias propostas, testando-as em relação aos dados. Permite ao pesquisador, avaliar a importância dos resultados e tomar decisões sobre aceitar ou rejeitar uma determinada hipótese.

Feitos a aplicação dos testes, os resultados estarão dispostos na tabela abaixo, para sumarizar as aplicações feitas nas colunas numéricas 'DiaSemChuva', 'Precipitacao' e 'FRP' dos 4 *datasets*. Alguns dos testes aplicados envolvem o uso de amostras aleatórias, para garantir que não haverá em momento algum a indução/tendência nos resultados apresentados:

Tipos/Dataset	2020	2021	2022	2023
Teste de <i>Shapiro-Wilk</i>	Provavelmente Não Gaussiano	Provavelmente Não Gaussiano	Provavelmente Não Gaussiano	Provavelmente Não Gaussiano
Coefficiente de Correlação de <i>Pearson</i>	Provavelmente independente	Provavelmente independente	Provavelmente dependente	Provavelmente independente
Teste de raiz unitária <i>Dickey-Fuller</i> aumentado	Provavelmente Não Estacionário	Provavelmente Estacionário	Provavelmente Estacionário	Provavelmente Estacionário
Teste de Análise de Variância ( <i>ANOVA</i> )	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição
Teste de <i>Friedman</i>	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição

Quadro 4 – Resultados dos Testes de Hipótese da 'DiaSemChuva'

Tipos/Dataset	2020	2021	2022	2023
Teste de <i>Shapiro-Wilk</i>	Provavelmente Não Gaussiano	Provavelmente Não Gaussiano	Provavelmente Não Gaussiano	Provavelmente Não Gaussiano
Coefficiente de Correlação de <i>Pearson</i>	Provavelmente independente	Provavelmente independente	Provavelmente independente	Provavelmente independente
Teste de raiz unitária <i>Dickey-Fuller</i> aumentado	Provavelmente Não Estacionário	Provavelmente Estacionário	Provavelmente Estacionário	Provavelmente Estacionário
Teste de Análise de Variância (ANOVA)	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição
Teste de <i>Friedman</i>	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição

Quadro 5 – Resultados dos Testes de Hipótese da ‘Precipitacao’

Tipos/Dataset	2020	2021	2022	2023
Teste de <i>Shapiro-Wilk</i>	Provavelmente Não Gaussiano	Provavelmente Não Gaussiano	Provavelmente Não Gaussiano	Provavelmente Não Gaussiano
Coefficiente de Correlação de <i>Pearson</i>	Provavelmente independente	Provavelmente independente	Provavelmente independente	Provavelmente independente
Teste de raiz unitária <i>Dickey-Fuller</i> aumentado	Provavelmente Não Estacionário	Provavelmente Estacionário	Provavelmente Estacionário	Provavelmente Estacionário
Teste de Análise de Variância (ANOVA)	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição
Teste de <i>Friedman</i>	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição	Provavelmente a mesma distribuição

Quadro 6 – Resultados dos Testes de Hipótese da FRP

1. Conclusões que podem ser obtidas com o tabelamento dos resultados:
2. Nos Testes de *Shapiro-Wilk*, as distribuições/datasets não possuem uma distribuição normal – no caso dos *datasets* testados, há uma grande volume de dados aglomerados à esquerda da distribuição.
3. O Coeficiente de Correlação de *Pearson* aponta o que foi mostrado nos estudos de correlações e na produção da Matriz de Confusão: ainda que os dados da ‘DiaSemChuva’ de 2022 difere das demais, as variáveis numéricas são independentes entre si (ou seja, sua correlação é muito baixa).

4. Enquanto que os Testes de raiz unitária *Dickey-Fuller* Aumentado mostram que os dados de 3 das 4 distribuições são provavelmente estacionários, e não precisam ser destacados, pois o dataset de 2020 deu por resultado o 'Provavelmente Não Estacionário'.

5. Nos Testes de Análise de Variância é indicado que as médias das distribuições são muito próximas, mesmo com a quantidade de dados desproporcional entre cada dataset, todos são muito semelhantes (visto na parte de correlações entre as classes numéricas).

6. E por fim, os Testes de *Friedman* reforçam o que foi demonstrado nos testes anteriores, os datasets possuem uma similaridade muito alta, o que é mostrado quando os resultados resultam em 'Provavelmente a mesma distribuição'.

Todos os testes aqui mencionados e utilizados além de comprovarem um comportamento e tendência de dados extremamente próximos, também servem como evidência de que os dados minerados e extraídos estarão isentos de quaisquer erros ou manipulações tendenciosas no decorrer da pesquisa.

## 4 | CONSIDERAÇÕES FINAIS

Primeiramente, pode-se notar que a AED é um ramo da ciência que vem recebendo maior atenção da comunidade acadêmica para produção de informações sobre os biomas e ecossistemas pesquisados, além de produzir informações vitais para elaboração de estratégias preventivas e de preservação do meio-ambiente em questão.

Seu uso apresenta uma realidade de aumento anual na detecção dos focos de queimadas na região amazônica, redução na quantidade de chuva e maiores chances de picos de queimadas, demonstrando a ineficácia das políticas públicas de enfrentamento a essas eventualidades, e uma urgência no remanejamento dos recursos dispostos para o combate das queimadas.

Além de fornecer informações preliminares a respeito dos dados do INPE, a AED é facilmente maleável para quaisquer necessidades de pesquisas, e até de facilitar determinados processos de filtração e seleção dos dados, conforme visto na etapa de tratamento dos *datasets*. Os resultados da pesquisa apresentam uma situação que exige mais atenção por parte das autoridades competentes quanto a preservação da Amazônia Legal, e a Análise Exploratória pode contribuir de forma significativa ao enfrentamento das queimadas, contribuindo assim para a sustentabilidade e melhoria da qualidade de vida dos cidadãos.

Tendo em vista o alto grau de especificidade proposto no título do mapeamento, há a possibilidade de realizar algumas melhorias na base teórica e nos processos metodológicos aqui apresentados. Na fundamentação, o uso de periódicos e revistas científicas especializadas na temática ambiental e/ou da ciência de dados, pode ampliar

os horizontes e as ideias de aplicabilidade da AED para problemas de cunho ecológico. Quanto aos processos metodológicos, uma busca mais aprofundada sobre o estado da arte é capaz de proporcionar um leque de possibilidades e métodos que seriam de grande contribuição para a pesquisa aqui apresentada.

## REFERÊNCIAS

[1] MACHADO, R. **Análise de Dados de queimadas ocorridas nos últimos anos no estado do Tocantins**. 44f. Monografia (Graduação). Curso de Ciência da Computação. Universidade Federal do Tocantins. Palmas, 2022. Disponível em: <https://repositorio.uft.edu.br/handle/11612/6136>. Acesso em: 8 jun. 2024.

[2] OLIVEIRA, P. **Uso de aprendizagem de máquina e redes neurais convolucionais profundas para a classificação de áreas queimadas em imagens de alta resolução espacial**. 2019. viii, 23 f., il. Dissertação (Mestrado em Geografia) – Universidade de Brasília, Brasília, 2019. Acesso em: 20 abril 2024.

[3] LOPES, G.; ALMEIDA, A.; DELBEM, A.; TOLEDO, C. **Introdução à análise exploratória de dados com python**. Minicursos ERCAS ENUCMPI, 2019, 160-176. Disponível em: [https://www.researchgate.net/publication/336778766\\_Introducao\\_a\\_Analise\\_Exploratoria\\_de\\_Dados\\_com\\_Python](https://www.researchgate.net/publication/336778766_Introducao_a_Analise_Exploratoria_de_Dados_com_Python). Acesso em: 14 jun. 2024.

[4] DE OLIVEIRA, E. **Análise de séries temporais para previsão de demanda no INSS**. 2022. Trabalho de Conclusão de Curso (Especialista em Ciência de Dados) – Escola Nacional de Administração Pública. Acesso em: 28 maio 2024.

[5] LINHARES, J. E. B. de S.; ASCENSO ROSA, R. R. P.; LAVAREDA FILHO, R. M. **Influência das comorbidades para a ocorrência de óbitos por COVID-19 em 2020: razão de chances no estado do Amazonas**. HU Revista, [S. l.], v. 48, p. 1–8, 2022. DOI: 10.34019/1982-8047.2022.v48.37689. Disponível em: <https://periodicos.ufjf.br/index.php/hurevista/article/view/37689>. Acesso em: 5 jun. 2024.

[6] HURTADO, S.; REGINA, B.; DA FONSECA, E. **Uso De Dados Abertos Como Estratégia De Apoio A Tomada De Decisões No Combate A Incêndios Florestais No Bioma Pantanal Do Mato Grosso Do Sul**. Encontro Internacional de Gestão, Desenvolvimento e Inovação (EIGEDIN), v. 6, n. 1, 3 nov. 2022. Acesso em: 15 maio 2024.

[7] Barbosa, A. **Incêndios em vegetação no município de Mirador-MA: o estudo de modelos preditivos baseados em dados atmosféricos**. 2021. Trabalho de Conclusão de Curso (Curso de Formação de Oficiais Bombeiro Militar) - Universidade Estadual do Maranhão. Disponível em: <https://repositorio.uema.br/handle/123456789/1477>. Acesso em: 14 maio 2024.

[8] CARRILHO, A.; MACHADO, D. **Monitoramento de queimadas no município de Novo Progresso – PA, nos períodos de 2019 e 2020 com o uso do MapBiomas**. 2022. Trabalho de Conclusão de Curso (Graduação em Agronomia) – Curso de Agronomia da Universidade Federal Rural da Amazônia – UFRA. Disponível em: <https://bdta.ufra.edu.br/jspui/handle/123456789/2489>. Acesso em: 11 maio 2024.

[9] FANTINATTI, Gabriela Fernanda. **Comparação de algoritmos preditivos para incêndios em canais**. 2021. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Química) – Universidade Tecnológica Federal do Paraná, Ponta Grossa, 2021. Acesso em: 17 mar. 2024.

[10] DIAS, A. **Uma Análise Exploratória de Dados sobre incêndios florestais no Brasil**. 2019. 52 f. Trabalho de Conclusão de Curso (Graduação em Engenharia de Computação) – Centro de Tecnologia, Departamento de Engenharia de Computação e Automação, Universidade Federal do Rio Grande do Norte, Natal, 2019. Acesso em: 25 mar. 2024.

[11] GOMES, V.; QUEIROZ, G. R.; FERREIRA, K.; SATO, L. Y.; SANTOS, R.; MORELLI, F. **Um ambiente para análise exploratória de grandes volumes de dados geoespaciais: explorando risco de fogo e focos de queimadas**. In: SIMPÓSIO BRASILEIRO DE GEOINFORMÁTICA, 18. (GEOINFO), 2017, Salvador. Anais... São José dos Campos: INPE, 2017. p. 301-309. Pendrive, Online. ISSN 2179-4820. IBI:<8JMKD3MGPDW34P/3QCRA6L>. Disponível em: <http://urlib.net/ibi/8JMKD3MGPDW34P/3QCRA6L>. Acesso em: 10 maio 2024.

[12] COSTA, J.; HENRIQUES, F; BELLOZE, K. Análise de Dados de Focos de Calor no Brasil Através de Técnicas de Visualização. In: BRAZILIAN E-SCIENCE WORKSHOP (BRESOI), 15, 2021, Evento Online. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2021. p. 17-24. ISSN 2763-8774. DOI: <https://doi.org/10.5753/bresoi.2021.15784>. Acesso em: 13 maio 2024.