

Previsão Mensal de Emissão de Dióxido de Carbono na Amazônia Legal

Luísa Ferreira da Silveira (2210875)
Pedro de Almeida Barizon (2211350)
Theo Couto Xavier (2210610)
Vinícius Lucena Bitu Cortez (2210458)

Professora: Paula Medina Maçaira Louro

Monitora: Laura Nascimento Coutinho

Turma: 33A

Data: 02/05/2025

1. Introdução

1.1 Motivação

Considerando o macrotema do projeto AMazonizAR, um esforço de inserção da PUC-Rio no contexto da Amazônia [ECO A PUC-RIO 2023], e a premente situação sócio-climático-ambiental que não apenas a Grande Floresta, mas todo o planeta enfrenta, optou-se pelo seguinte tema para o projeto: *Na Amazônia Legal, qual a inter-relação entre incêndios florestais, mudanças climáticas e emissão de dióxido de carbono (CO₂)?*

1.2 Questão de Pesquisa

A fim de transformar o tema do projeto em algo mais tangível e mensurável, formulou-se esta questão de pesquisa (QP):

A partir dos indicadores ambientais de um estado da Amazônia Legal e seus níveis de incêndios florestais (queimadas) em um dado mês, qual a quantidade de CO₂, em toneladas, emitida por ele na atmosfera nesse mês?

Para o caso de a absorção ser maior que a emissão, convencionou-se uma emissão negativa. Quanto aos estados, serão considerados segundo a classificação do IBGE [2020]:

- Acre (AC);
- Amapá (AP);
- Amazonas (AM);
- Maranhão (MA);
- Mato Grosso (MT);
- Pará (PA);
- Rondônia (RO);
- Roraima (RR);
- Tocantins (TO).

1.3 Cliente a Ser Atendido

Os clientes-alvo serão os **participantes do AMazonizAR, da PUC-Rio, e eventuais órgãos ambientais e governamentais que desejem fazer uso do modelo de dados gerado**. Espera-se oferecer-lhes uma ferramenta que agregue valor ao permitir estimar um panorama de emissões da Floresta, o que configura um indicador útil em relação ao controle da crise climática e ao mundo dos créditos de carbono — temas de interesse de toda a comunidade global, que poderá mostrar-se mais solícita na realização de ações interventivas diante de estimativas quantitativas.

1.4 Objetivo

No que diz respeito aos objetivos deste trabalho, mesclam tanto interesses de natureza acadêmica quanto de natureza cidadã. Se, por um lado, almeja-se aprender os conceitos e as ferramentas que compõem a Ciência de Dados contemporânea; por outro, objetiva-se compreender melhor o cenário amazônico, tão relevante não apenas à nação brasileira, senão ao mundo inteiro, que vê chegarem os efeitos de uma crise ambiental que não parece ter planos de ir embora. Assim, na impossibilidade de vê-la partir, deve-se aprender a conviver com a iminente crise — e, para tanto, são necessários análise e estudo.

1.5 Expectativa de Resultados

A priori, espera-se — embora não necessariamente linear — uma correlação direta entre emissões de CO₂ e quantidade de área queimada. Afinal, é razoável pensar que, aumentando-se a porção queimada, deve-se elevar o contingente de biomassa atingido, um dos reagentes necessários para a produção de dióxido de carbono. Nesse sentido, o acréscimo de fatores ambientais — como temperatura e umidade relativa do ar — talvez forneça informações complementares para que os modelos de dados prevejam adequadamente as emissões. Diante do exposto, se confirmada essa complementaridade, poderão ser desenvolvidos indicadores-chave de desempenho que transcendam o contexto das queimadas, passando a incluir também fatores ambientais.

2. Obtenção e Tratamento dos Dados

2.1 Justificativa para a Escolha dos Dados/Atributos

Por ser a variável alvo, é indubitável que a **emissão de CO₂** devesse ser um dos atributos. Nesse cenário, como a QP define que as previsões devem ter granularidade mensal e estadual; **ano**, **mês** e **estado** também precisaram ser incluídos, a fim de permitirem a agregação correta dos dados.

Quanto ao escopo de queimadas, também delimitado na QP, optou-se apenas pela **área queimada** e pela **quantidade de focos**, porque quantificam as queimas de forma contínua e discreta, o que fornece uma descrição complementar sobre a aparição desses eventos no espaço geográfico. Foram também considerados outros fatores, a saber: *Fire Radiative Power* (denotado por FRP; mede a intensidade de energia emitida por uma queimada); risco de fogo; e dias sem chuva. O FRP detalha a intensidade de uma queimada, enquanto as demais medidas descrevem fatores que a propiciam. Pela indisponibilidade de dados, todavia, precisaram ser descartados (ao menos nesta primeira *sprint*). Por fim, foram cogitados indicadores socioeconômicos que se relacionam com as queimadas vindas da ação antrópica, mas não foram incluídos por conta da complexidade que trariam.

Quanto ao escopo ambiental, consideraram-se os indicadores disponíveis pelo Instituto Nacional de Meteorologia (INMET). Dessa forma, a inclusão ou exclusão dos atributos dependeu diretamente da disponibilidade e da viabilidade de extração, haja vista que muitas das medições eram horárias e deveriam ser agregadas em valores mensais. Em suma, depois de muitas coletas e reestruturações, retiveram-se **pressão atmosférica**, **umidade relativa do ar**, **velocidade do vento** e **temperatura**, todas as quais se relacionam com a produção de CO₂.

2.2 Fontes dos Dados

Foram utilizadas as seguintes fontes, todas públicas e com dados gratuitos para *download*. Uma vez que todas são organizações governamentais ou civis com compromisso socioambiental e científico, acreditou-se serem confiáveis e de qualidade.

2.2.1 Dados de emissões

Banco de Dados do SEEG (Sistema de Estimativas de Emissões de Gases de Efeito Estufa) – O SEEG, mantido pelo Observatório do Clima [2025], fornece estimativas de emissões de gases de efeito estufa, incluindo CO₂, para diferentes setores e regiões do Brasil, inclusive a Amazônia Legal.

2.2.2 Dados de queimadas

MapBiomass – Monitor do Fogo – O Monitor do Fogo, desenvolvido pelo MapBiomass [2025], oferece uma plataforma interativa de monitoramento de áreas queimadas no Brasil, com base em imagens de satélite processadas automaticamente. O sistema permite a visualização e a análise temporal das ocorrências de fogo em diferentes biomas, com alta resolução espacial e atualizações frequentes.

Instituto Nacional de Pesquisas Espaciais (INPE) – Programa Queimadas – O Programa Queimadas do INPE [2025] fornece dados detalhados sobre focos de calor e áreas queimadas no Brasil e na América Latina, com atualizações quase em tempo real. A plataforma **TerraBrasilis** permite acesso a séries históricas desde 1998, além de mapas, estatísticas e análises espaciais por bioma, estado e município, sendo uma referência nacional no monitoramento de incêndios florestais.

2.2.3 Dados ambientais

Instituto Nacional de Meteorologia (INMET) – O INMET [2025] disponibiliza dados meteorológicos oficiais do Brasil, incluindo medições de temperatura, umidade, pressão atmosférica, velocidade do vento e precipitação. Essas informações, obtidas por meio de uma rede nacional de estações, são fundamentais para análises climáticas e ambientais com granularidade temporal variada, incluindo séries históricas mensais.

Para uma visão mais específica, segue o dicionário de fontes:

Categoria	Espaço	Tempo	Descrição	Fonte
Clima	município	hora	Fatores climáticos (temperatura, precipitação, umidade relativa do ar, dentre outros) a partir de medições horárias de estações meteorológicas em alguns municípios a partir de 2000.	INMET
CO2	município	ano	Emissões de CO2 em toneladas por estado do Brasil por ano a partir de 2000.	Observatório do Clima
Queimadas	estado	mês	Mapeamento mensal de cicatrizes de fogo por estado a partir de 2019.	MapBiomas
Queimadas	estado	mês	Série histórica, mapeada de 1985 a 2023, que apresenta as cicatrizes do fogo mensais e por estado.	MapBiomas
Queimadas	estado	dia, mês, ano	Fire Radiative Power (FRP), quantidade de focos de queimadas, risco de fogo e dias sem chuva, todos por estado por mês a partir de 2019.	TerraBrasilis
Queimadas	estado	mês	Quantidade de focos de queimadas por estado por mês a partir de 1998.	TerraBrasilis

Tabela 1 – Dicionário de fontes do projeto

2.3 *Procedimentos de Limpeza/Transformação/Redução*

2.3.1 *Dados de emissões*

Foi necessário reformatar o arquivo CSV vindo do SEEG, pois seu formato mostrou-se incompreensível pelo **pandas**. Diante disso, fez-se uso gratuito do chat GPT 4o-mini com o seguinte *prompt*: *Poderia formatar adequadamente o suposto csv que lhe enviarei? Quando fui abri-lo no pandas, os dados não foram processados adequadamente.*

Outliers

Quanto a *outliers*, optou-se por deliberadamente não os remover, uma vez que a emissão de dióxido de carbono na atmosfera é sensível à atividade antrópica, de natureza volátil e imprevisível, porque é atrelada a dezenas de variáveis: disposições legais do País, demanda por madeira, governo vigente etc.

Dados faltantes

Como a granularidade temporal dos dados a alimentarem o modelo deveria ser mensal, fez-se, sob sugestão da professora Paula Maçaira, a seguinte extrapolação: não tendo sido possível coletar informações mês a mês sobre as emissões, mas apenas ano a ano; decidiu-se replicar o valor anual ao longo dos meses, a fim de que refletissem o comportamento geral — algo semelhante à imputação da média, mas multiplicada pelo número de meses. Dessa forma, espera-se, o modelo ainda será capaz de identificar padrões, se bem que com menor teor de detalhe. Seja como for, ainda se procuram as informações com a devida granularidade, que poderão ser incorporadas ao projeto em futuras *sprints*.

2.3.2 *Dados de queimadas*

Outliers

Novamente, não foram removidos *outliers*, porque poderiam indicar eventos atípicos, que contribuiriam para o aumento das emissões.

Dados faltantes

Quanto à quantidade de focos, não houve dados faltantes. Quanto à área queimada, foram utilizadas duas bases de dados contendo informações por estado e mês: uma principal, abrangendo o período de **1985 a 2023**; e outra complementar, com dados disponíveis entre **2019 e 2025**. Ambas forneceram registros mensais por unidade federativa, permitindo análises temporais e geográficas das queimadas ao longo de quatro décadas.

Durante o processo de integração, foi identificado que a base principal apresentava valores ausentes nos anos de 2019 a 2023. Para lidar com essas lacunas, optou-se por substituí-las por dados correspondentes da base complementar, considerando a mesma data e estado. Essa decisão foi respaldada por uma análise de correlação entre os dois conjuntos, cujo coeficiente foi de aproximadamente **0,52**. Embora não seja elevado, esse valor foi considerado suficiente para justificar o uso de dados observados, ao invés de estimativas com base em médias.

Nos anos anteriores a 2019 (1985 a 2018), em que não havia sobreposição com a base complementar, os valores ausentes foram preenchidos utilizando a **média mensal por estado**, de modo a preservar padrões sazonais regionais.

Além disso, foram incorporados os registros relativos a 2024 e 2025 provenientes da base complementar, dada a **alta correlação** desses valores com as médias históricas da base principal (**0,88**), conferindo robustez à extrapolação temporal do conjunto final.

2.3.3 Dados ambientais

Outliers

Trataram-se os *outliers* após comparar os resultados com e sem o tratamento. Mantendo-os, observaram-se valores incoerentes com a realidade, como temperaturas médias em torno de 12°C, além de um aumento significativo na quantidade de dados faltantes.

Dados faltantes

Em relação aos valores ausentes, optou-se por remover a coluna de radiação, que apresentava mais de 70% dos dados faltantes. Consideramos inviável aplicar técnicas de imputação, tanto pela grande quantidade de valores ausentes quanto pela baixa relevância da radiação para as queimadas na Amazônia. A região apresenta alta umidade ao longo do ano, o que dificulta a ocorrência de queimadas por causas naturais, como o acúmulo de radiação solar.

2.4 Tamanho da Base Final

Após o tratamento, resultaram 1025 tuplas com 15 atributos cada, conforme ilustram as figuras abaixo:

	_ano	_estado	_mes	car_c02_emitido	cli_pressao_atm_med	cli_temp_ar_med	cli_temp_orvalho_med	cli_umid_rel_med	cli_umid_rel_min_max	cli_umid_rel_min_med
0	2008	AC	7	2.627698e+07	986.843612	28.142731	18.914978	59.555066	95.0	54.432558
1	2008	AC	9	2.627698e+07	991.705941	24.446194	19.467987	75.811881	97.0	72.496700
2	2008	AC	10	2.627698e+07	990.328360	25.229298	21.617473	81.870968	96.0	78.720430
3	2008	AC	11	2.627698e+07	988.610987	25.195410	22.624478	86.905424	96.0	84.048679
4	2008	AC	12	2.627698e+07	988.692608	24.898790	22.727554	88.529570	96.0	86.116935

Shape: (1025, 15)

Figura 1 – Primeira metade do *dataset* após o tratamento

cli_umid_rel_med	cli_umid_rel_min_max	cli_umid_rel_min_med	cli_umid_rel_min_min	cli_veloc_vento_max	cli_veloc_vento_med	que_area_queimada	que_focos_qtd
59.555066	95.0	54.432558	29.0	5.1	2.152915	4957.0	165.0
75.811881	97.0	72.496700	25.0	1.0	0.210504	46073.0	2947.0
81.870968	96.0	78.720430	29.0	1.0	0.204959	30355.0	856.0
86.905424	96.0	84.048679	42.0	1.0	0.186970	2082.0	63.0
88.529570	96.0	86.116935	53.0	1.0	0.179442	127.0	4.0

Figura 2 – Segunda metade do *dataset* após o tratamento

2.5 Dicionário de Dados

A seguir, produziu-se um resumo do dicionário de dados do projeto, cuja versão completa encontra-se em [dicionarioDados](#). Os atributos que compõem a chave primária do *dataset* foram grifados em vermelho.

Nome	Tipo	Descrição	Unidade de medida	Fonte
_ano	Int32	Ano segundo o calendário cristão.	ano	Não se aplica
_estado	string	Nome do estado da Amazônia Legal.	adimensional	Não se aplica
_mes	Int32	Mês do ano segundo o calendário cristão. Janeiro relaciona-se com 1, fevereiro com 2 e assim em diante.	mês	Não se aplica
car_c02_emitido	Float64	Quantidade de dióxido de carbono emitido na atmosfera. Se for negativa, considera-se que tenha havido absorção.	toneladas (t)	SEEG
cli_pressao_atm_med	Float64	Média das medições mensais de pressão atmosférica de todas as estações meteorológicas do estado disponíveis no Instituto Nacional de Meteorologia (INMET).	mb (milibar)	INMET
cli_temp_ar_med	Float64	Média das medições mensais de temperatura do ar de todas as estações meteorológicas do estado disponíveis no Instituto Nacional de Meteorologia (INMET).	°C	INMET
cli_temp_orvalho_med	Float64	Média das medições mensais de temperatura do ponto de orvalho de todas as estações meteorológicas do estado disponíveis no Instituto Nacional de Meteorologia (INMET).	°C	INMET
cli_umid_rel_med	Float64	Média das medições mensais de umidade relativa do ar de todas as estações meteorológicas do estado disponíveis no Instituto Nacional de Meteorologia (INMET).	percentual (%)	INMET
cli_umid_rel_min_max	Float64	Máxima dentre as mínimas mensais de umidade relativa do ar de todas as estações meteorológicas do estado disponíveis no Instituto Nacional de Meteorologia (INMET).	percentual (%)	INMET
cli_umid_rel_min_med	Float64	Média das mínimas mensais de umidade relativa do ar de todas as estações meteorológicas do estado disponíveis no Instituto Nacional de Meteorologia (INMET).	percentual (%)	INMET
cli_umid_rel_min_min	Float64	Mínima dentre as mínimas mensais de umidade relativa do ar de todas as estações meteorológicas do estado disponíveis no Instituto Nacional de Meteorologia (INMET).	percentual (%)	INMET

cli_veloc_vento_max	Float64	Máxima dentre as medidas mensais de velocidade do ar de todas as estações meteorológicas do estado disponíveis no Instituto Nacional de Meteorologia (INMET).	m/s	INMET
cli_veloc_vento_med	Float64	Média das medidas mensais de velocidade do ar de todas as estações meteorológicas do estado disponíveis no Instituto Nacional de Meteorologia (INMET).	m/s	INMET
que_area_queimada	Float64	Número de hectares de área queimada.	hectare (Ha)	MapBiomass
que_focos_qtd	Float64	Quantidade de focos de queimada. Se bem que, em essência, seja um valor inteiro, foi considerado um domínio de ponto flutuante por conta de imputação de dados.	adimensional	TerraBrasilis

Tabela 2 – Dicionário de dados do projeto

3. Análise Exploratória dos Dados

Para sistematizar a análise, adotou-se, a nível de código, o seguinte critério: *Se uma dada variável apresentar tipo numérico (**int**, **float**, dentre outros), automaticamente será **quantitativa**. A cardinalidade será, então, dada pela do tipo: inteiros são contáveis (**discretos**), e pontos flutuantes representam intervalos reais, não contáveis (**contínuos**). Caso contrário, assume-se ser qualitativa, e a subcategoria é avaliada ad hoc por análise humana.*

Diante disso, vale ressaltar que, embora **_mes** fosse originalmente uma variável qualitativa ordinal, com domínio (**janeiro**, **fevereiro**, ...), com a codificação real para (**1**, **2**, ...), pode-se interpretá-la como a quantidade de meses desde o início do ano, algo quantitativo discreto. Quanto a **_estado**, como não há ordem explícita, segue que é quantitativa nominal.

Assim, as variáveis terão a seguinte classificação:

1. **_ano**: quantitativo discreto;
2. **_estado**: qualitativo nominal;
3. **_mes**: quantitativo discreto;
4. **car_c02_emitido**: quantitativo contínuo;
5. **cli_pressao_atm_med**: quantitativo contínuo;
6. **cli_temp_ar_med**: quantitativo contínuo;
7. **cli_temp_orvalho_med**: quantitativo contínuo;
8. **cli_umid_rel_med**: quantitativo contínuo;
9. **cli_umid_rel_min_max**: quantitativo contínuo;
10. **cli_umid_rel_min_med**: quantitativo contínuo;
11. **cli_umid_rel_min_min**: quantitativo contínuo;
12. **cli_veloc_vento_max**: quantitativo contínuo;
13. **cli_veloc_vento_med**: quantitativo contínuo;
14. **que_area_queimada**: quantitativo contínuo;
15. **que_focos_qtd**: quantitativo contínuo.

3.1 Estatísticas Descritivas

A seguir, calcularam-se as medidas da Estatística Descritiva para as amostras das variáveis quantitativas, que incluem:

- **Tendência Central:**
 - Média;
 - Mediana;
 - Moda (ou modas; pode ser mais de uma).
- **Dispersão:**
 - Amplitude máxima;
 - Variância;
 - Desvio padrão;
 - Coeficiente de Variação.
- **Posição:**
 - Quartis (Q1, Q2, Q3).
- **Assimetria:**
 - Coeficiente γ_3 de Fisher-Pearson (no **pandas**, é obtido com **Series.skew()**).
- **Curtose:**
 - Coeficiente κ_4 de Fisher (no **pandas**, é obtido com **Series.kurtosis()**).

Métricas: CENTRALITY			
	media	mediana	moda
_ano	2014.468293	2014.0	[2015, 2019]
_mes	6.500488	7.0	[1, 5]
car_c02_emitido	107520273.652189	56669772.946213	[3652935.022779369, 3705752.691345386, 3958983...
cli_pressao_atm_med	993.395296	997.229524	[1007.8279069767442, 1011.1125]
cli_temp_ar_med	26.429096	26.29328	[21.979722222222222, 21.99903181189488, 22.0466...
cli_temp_orvalho_med	21.282504	22.111877	[20.767185289957567, 22.77238493723849]
cli_umid_rel_med	75.905264	79.0	83.438172
cli_umid_rel_min_max	94.459512	95.0	95.0
cli_umid_rel_min_med	72.934699	75.870968	[71.88888888888889, 81.89516129032258]
cli_umid_rel_min_min	38.312195	39.0	34.0
cli_veloc_vento_max	3.862341	3.6	3.4
cli_veloc_vento_med	1.245765	1.155405	[0.0163475699558173, 0.0483660130718954, 0.055...
que_area_queimada	109144.90887	14356.0	[958.0, 12673.4]
que_focos_qtd	571.351247	26.0	1.0

Figura 3 – Medidas de centralidade das variáveis quantitativas

Métricas: DISPERSION						
	max	min	amplitude	variancia	desv_pad	coef_var
_ano	2.023000e+03	2.000000e+03	2.300000e+01	2.796799e+01	5.288477e+00	0.002625
_mes	1.200000e+01	1.000000e+00	1.100000e+01	1.192798e+01	3.453691e+00	0.531297
car_c02_emitido	6.532973e+08	3.271708e+06	6.500256e+08	1.794343e+16	1.339531e+08	1.245840
cli_pressao_atm_med	1.012780e+03	9.229797e+02	8.980061e+01	2.922557e+02	1.709549e+01	0.017209
cli_temp_ar_med	3.176319e+01	2.197972e+01	9.783472e+00	2.106806e+00	1.451484e+00	0.054920
cli_temp_orvalho_med	2.598745e+01	7.230854e+00	1.875659e+01	7.506876e+00	2.739868e+00	0.128738
cli_umid_rel_med	9.239058e+01	1.696992e+01	7.542065e+01	1.358411e+02	1.165509e+01	0.153548
cli_umid_rel_min_max	1.000000e+02	3.000000e+01	7.000000e+01	2.631306e+01	5.129625e+00	0.054305
cli_umid_rel_min_med	9.186012e+01	1.496162e+01	7.689850e+01	1.421618e+02	1.192316e+01	0.163477
cli_umid_rel_min_min	7.700000e+01	9.000000e+00	6.800000e+01	1.612091e+02	1.269681e+01	0.331404
cli_veloc_vento_max	9.600000e+00	6.000000e-01	9.000000e+00	1.796608e+00	1.340376e+00	0.347037
cli_veloc_vento_med	4.102525e+00	1.634757e-02	4.086178e+00	4.466144e-01	6.682922e-01	0.536451
que_area_queimada	3.144010e+06	1.000000e+00	3.144009e+06	6.311846e+10	2.512339e+05	2.301838
que_focos_qtd	7.669000e+03	1.000000e+00	7.668000e+03	1.448882e+06	1.203695e+03	2.106751

Figura 4 – Medidas de dispersão das variáveis quantitativas

Métricas: POSITION			
	q1	q2	q3
_ano	2.010000e+03	2.014000e+03	2.019000e+03
_mes	4.000000e+00	7.000000e+00	9.000000e+00
car_c02_emitido	3.130998e+07	5.666977e+07	1.148093e+08
cli_pressao_atm_med	9.861975e+02	9.972295e+02	1.004590e+03
cli_temp_ar_med	2.546319e+01	2.629328e+01	2.730413e+01
cli_temp_orvalho_med	2.073703e+01	2.211188e+01	2.293194e+01
cli_umid_rel_med	7.040599e+01	7.900000e+01	8.394602e+01
cli_umid_rel_min_max	9.400000e+01	9.500000e+01	9.700000e+01
cli_umid_rel_min_med	6.730556e+01	7.587097e+01	8.109489e+01
cli_umid_rel_min_min	3.000000e+01	3.900000e+01	4.700000e+01
cli_veloc_vento_max	3.000000e+00	3.600000e+00	4.700000e+00
cli_veloc_vento_med	8.243733e-01	1.155405e+00	1.479324e+00
que_area_queimada	2.954000e+03	1.435600e+04	8.966600e+04
que_focos_qtd	3.500000e+00	2.600000e+01	3.660000e+02

Figura 5 – Medidas de posição das variáveis quantitativas

Métricas SKEW:

	assimetria
_ano	-0.200539
_mes	-0.004483
car_c02_emitido	2.105874
cli_pressao_atm_med	-2.160808
cli_temp_ar_med	0.369335
cli_temp_orvalho_med	-2.164032
cli_umid_rel_med	-1.366295
cli_umid_rel_min_max	-5.131046
cli_umid_rel_min_med	-1.274984
cli_umid_rel_min_min	-0.096313
cli_veloc_vento_max	0.679577
cli_veloc_vento_med	1.374271
que_area_queimada	5.136334
que_focos_qtd	2.802279

Figura 6 – Medida de assimetria das variáveis quantitativas

Métricas KURTOSIS:

	curtose
_ano	-0.680608
_mes	-1.214559
car_c02_emitido	3.742102
cli_pressao_atm_med	5.872654
cli_temp_ar_med	0.565680
cli_temp_orvalho_med	5.283790
cli_umid_rel_med	2.001775
cli_umid_rel_min_max	45.514632
cli_umid_rel_min_med	1.744670
cli_umid_rel_min_min	-0.050843
cli_veloc_vento_max	1.166441
cli_veloc_vento_med	2.847532
que_area_queimada	39.861126
que_focos_qtd	8.632749

Figura 7 – Medida de curtose das variáveis quantitativas

3.2 Visualizações de Distribuição

Com base na função densidade de probabilidade (FDP) abaixo estimada pelo gráfico KDE, segue que há maior probabilidade de que a emissão de CO₂ seja um valor inferior a 200 milhões de toneladas por mês, interpretação coerente com a média e com a mediana encontrada, ambas abaixo de 110 milhões. Vê-se, igualmente, que as probabilidades de que haja emissão acima de 600 milhões ou abaixo de 0 (isto é, uma absorção de dióxido de carbono) são improváveis. Por fim, o formato da FDP sugere assimetria positiva e caudas leves (curva leptocúrtica), o que se mostra coerente com os coeficientes de assimetria e de curtose calculados acima, ambos positivos.

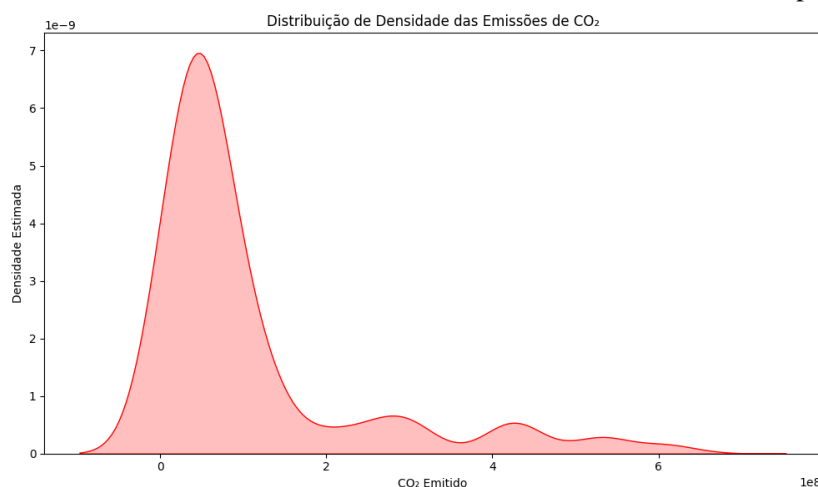


Figura 8 – Gráfico de densidade (KDE) da emissão de CO₂

Analogamente, uma vez que o gráfico KDE nada mais é que uma espécie de extrapolação contínua do histograma a seguir, extraem-se conclusões semelhantes às do outro item. Dessa forma, como se era de esperar, constata-se novamente assimetria positiva e concentração de valores abaixo de 200 milhões de toneladas emitidas. Há poucos valores acima de 600 milhões e nenhum abaixo de 0.

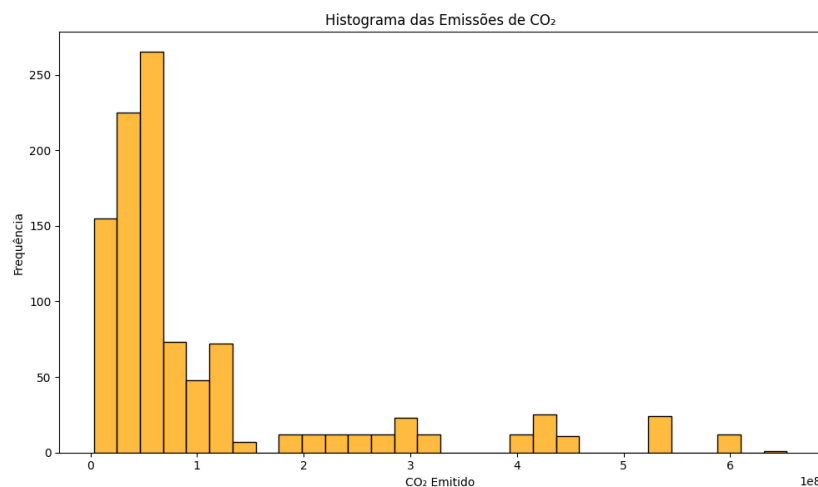


Figura 9 – Histograma da emissão de CO₂

3.3 Análise de Correlação

Consideram-se como **pertinentes as correlações ρ com valores absolutos maiores ou iguais a 0.70**, segundo o critério usual para indicar **correlação forte**. Correlações **fracas ou próximas de zero** (entre -0.05 e 0.05) também são destacadas por sua **irrelevância estatística**, assim como aquelas **próximas do corte de 0.70**, que podem indicar relações latentes interessantes.

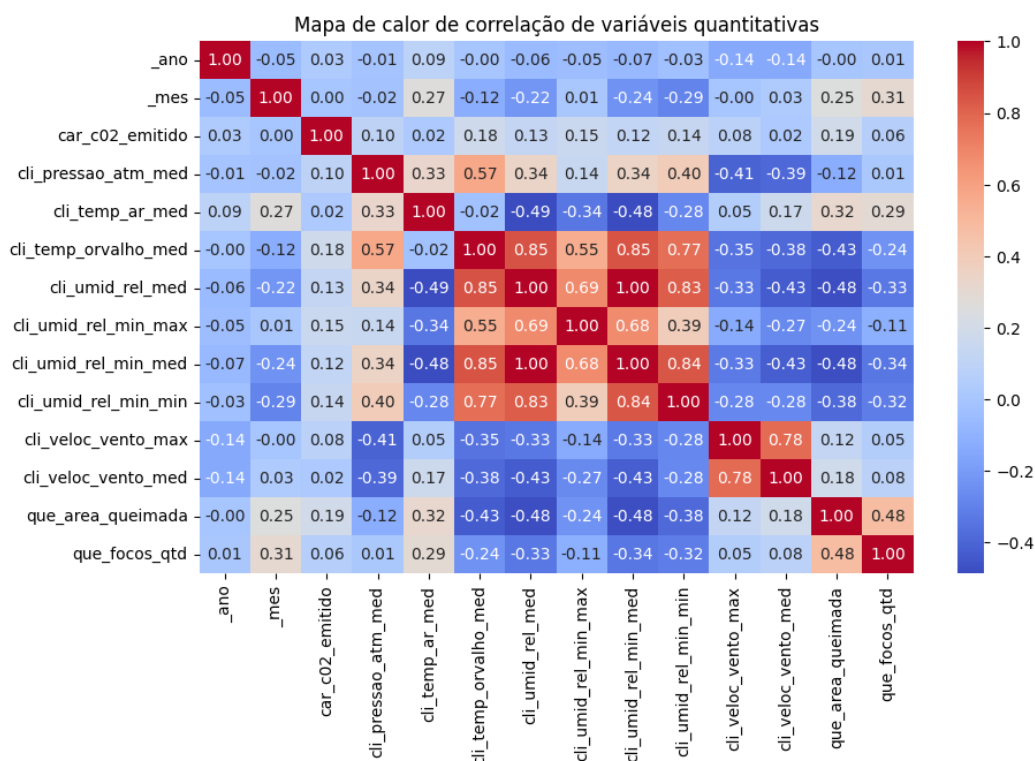


Figura 10 – Mapa de calor de correlação de variáveis quantitativas

Correlações fracas ou próximas de zero ($|\rho| \leq 0.05$)

1. **_ano** ↔ **cli_pressao_atm_med** → $\rho = -0.01$
 A pressão atmosférica não variou de forma consistente ao longo dos anos.
2. **_mes** ↔ **car_c02_emitido** → $\rho = 0.00$
 Não há sazonalidade clara (mensal) nas emissões de CO₂.
3. **_ano** ↔ **cli_temp_orvalho_med** → $\rho = -0.00$
 Nenhuma relação entre o ano e a temperatura de orvalho.
4. **_ano** ↔ **que_focos_qtd** → $\rho = 0.01$
 O número de focos de queimadas também não teve tendência clara com o tempo.
5. **cli_veloc_vento_med** ↔ **car_c02_emitido** → $\rho = 0.02$
 A emissão de CO₂ não apresenta relação com a velocidade média do vento.

6. **cli_temp_ar_med** ↔ **que_area_queimada** → $\rho = 0.03$

Surpreendentemente, a temperatura do ar não se correlaciona significativamente com a área queimada.

7. **cli_veloc_vento_max** ↔ **car_c02_emitido** → $\rho = 0.04$

A máxima velocidade do vento mensal também não influencia nas emissões de CO₂.

Correlações moderadas ($0.05 < |\rho| < 0.70$)

1. **que_area_queimada** ↔ **car_c02_emitido** → $\rho = 0.19$

Correlação surpreendentemente baixa. Intuitivamente, a quantidade de área queimada deveria ter alta correlação com a emissão de CO₂, uma vez que este é produzido por queima.

2. **que_area_queimada** ↔ **que_focos_qtd** → $\rho = 0.48$

Relação razoável entre a quantidade de focos de queimadas e a área efetivamente queimada. Apesar de positiva, não é suficiente para prever com segurança uma variável com base na outra. Esperava-se que fosse maior.

3. **cli_umid_rel_med** ↔ **cli_umid_rel_min_max** → $\rho = 0.69$

Essa correlação quase entrou no critério de corte. Mostra que a média da umidade relativa e o valor máximo das mínimas mensais de umidade variam de forma parecida.

Correlações fortes ($|\rho| \geq 0.70$)

1. **cli_veloc_vento_max** ↔ **cli_veloc_vento_med** → $\rho = 0.78$

Como esperado, quanto maior a velocidade máxima do vento, maior tende a ser a velocidade média mensal. A alta correlação indica consistência nos padrões de vento.

2. **cli_umid_rel_med** ↔ **cli_umid_rel_min_min** → $\rho = 0.77$

A média da umidade relativa do mês está fortemente relacionada com a menor umidade registrada nas mínimas diárias — o que sugere meses úmidos tendendo a ter mínimas também mais altas.

3. **cli_umid_rel_min_med** ↔ **cli_umid_rel_min_min** → $\rho = 0.83$

As mínimas médias mensais de umidade e as mínimas absolutas são fortemente relacionadas. É esperado, pois uma média baixa tende a acompanhar extremos baixos.

4. **cli_umid_rel_med** ↔ **cli_umid_rel_min_med** → $\rho = 0.84$

Meses com alta umidade média apresentam também altas mínimas de umidade. Isso reforça o padrão de estabilidade na umidade do ar.

5. **cli_umid_rel_med** ↔ **cli_temp_orvalho_med** → $\rho = 0.85$

A alta correlação entre umidade relativa média e temperatura de orvalho média é coerente, pois o ponto de orvalho é função direta da umidade. Isso evidencia coerência interna nos dados climáticos.

6. **cli_umid_rel_min_med** ↔ **cli_temp_orvalho_med** → $\rho = 0.85$

Também há forte relação entre a mínima média da umidade e o ponto de orvalho, sugerindo que períodos com menos secura estão diretamente associados a temperaturas de orvalho mais altas.

Contudo, suprimindo-se as componentes temporais, e enfocando-se os estados, percebeu-se uma correlação nitidamente maior entre **que_area_queimada** e **car_c02_emitido**, passando de 0.19 a 0.57, conforme evidencia a matriz a seguir. Isso faz todo o sentido. Afinal, como não foram encontrados dados mensais de emissões — apenas anuais —, os valores mensais foram imputados pelo respectivo valor anual, o que deve ter introduzido distorções. Dessa forma, ao se agregarem os dados, suavizaram-se as discrepâncias.

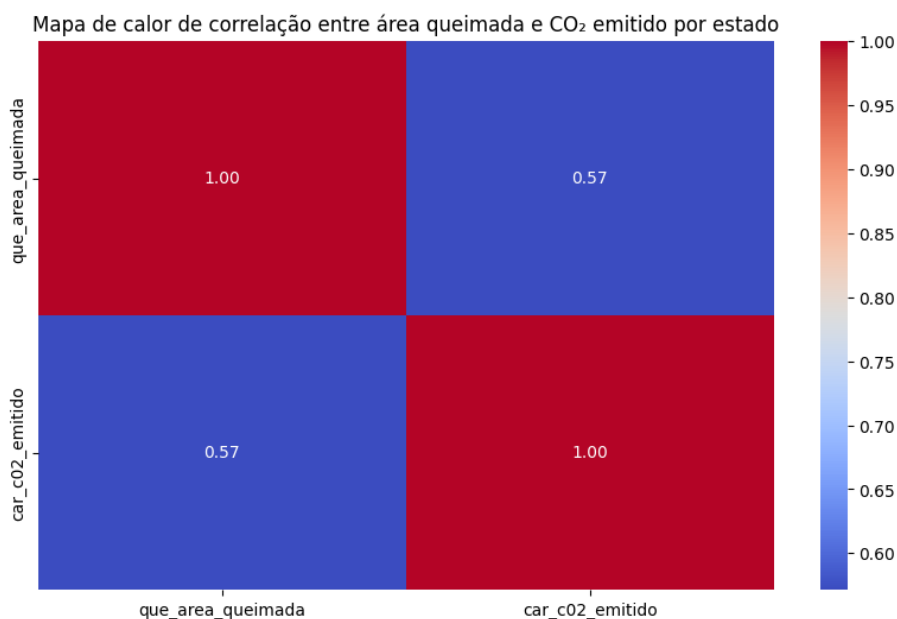


Figura 11 – Correlação entre área queimada e CO₂ emitido (por estado)

3.4 Gráficos de Dispersão

Este gráfico de dispersão mostra a relação entre a quantidade de focos de queimada e a emissão de CO₂ nos diferentes estados da Amazônia Legal. De forma geral, não se observa uma correlação linear forte e direta entre essas duas variáveis. Embora haja uma concentração de pontos com baixos valores tanto para focos quanto para emissão de CO₂, percebe-se uma grande dispersão nos dados.

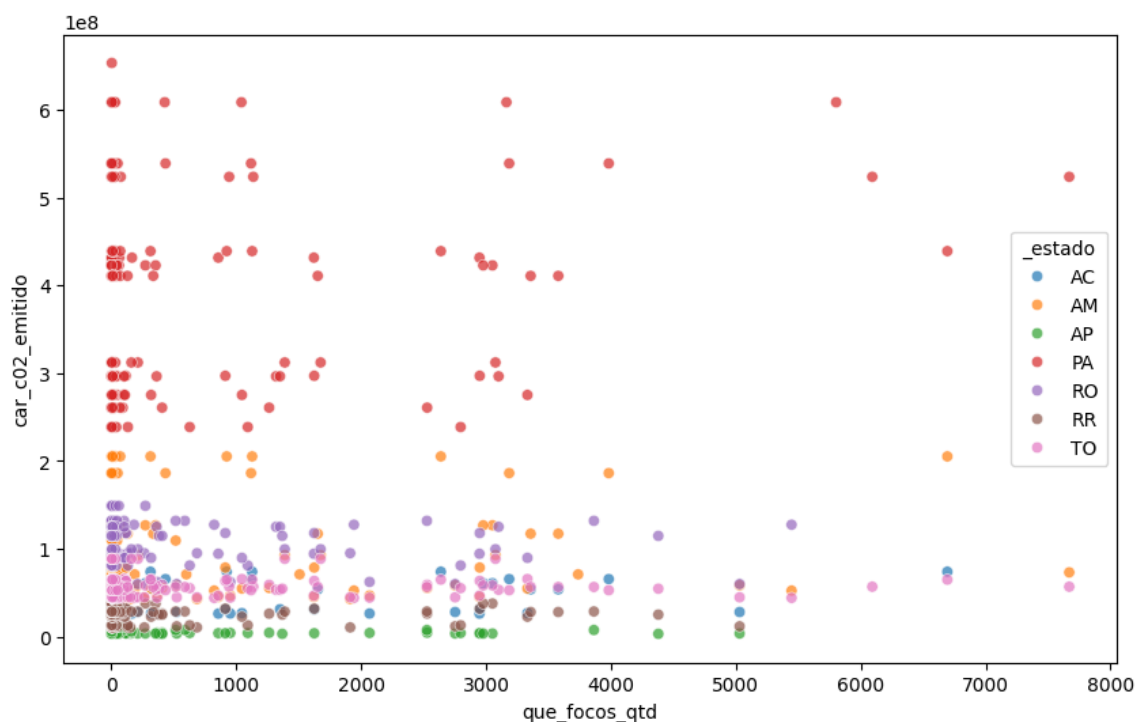


Figura 12 – Focos de queimada vs CO₂ emitido

Alguns estados, como o Pará, apresentam pontos com um número relativamente baixo de focos, mas com emissões de CO₂ significativamente elevadas. Outros estados, como o Amazonas e o Acre, tendem a se concentrar em uma faixa inferior de emissões, mesmo com alguma variação no número de focos. A presença de pontos isolados com altos valores de emissão de CO₂ independentemente da quantidade de focos sugere que outros fatores, como a intensidade e a biomassa queimada em cada foco, ou até mesmo outras fontes de emissão não diretamente relacionadas aos focos detectados, podem ter uma influência considerável. Assim, enquanto se poderia esperar um aumento nas emissões com o aumento dos focos, o gráfico indica uma relação mais complexa e com múltiplos fatores entre essas variáveis nos diferentes estados da Amazônia Legal.

Este outro gráfico revela a relação entre a quantidade de focos de queimada e a área queimada nos estados da Amazônia Legal. De maneira geral, observa-se uma correlação positiva, indicando que um maior número de focos tende a estar associado a uma maior área queimada, embora essa relação apresente uma considerável variabilidade.

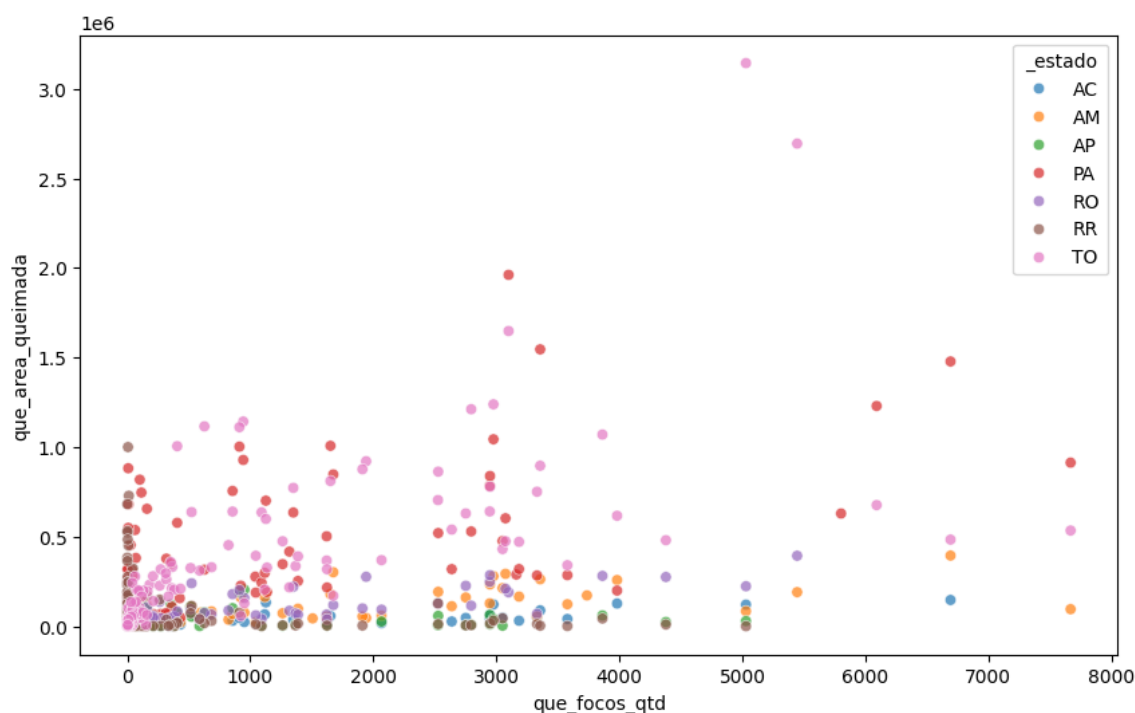


Figura 13 – Área queimada vs focos de queimada

A maior concentração de pontos na parte inferior esquerda do gráfico sugere que muitos eventos envolvem um número relativamente baixo de focos e resultam em áreas queimadas menores. Ao analisar por estado, percebem-se padrões distintos: alguns estados, como o Pará e o Tocantins, exibem ocorrências com tanto um alto número de focos quanto áreas queimadas extensas, enquanto outros, como o Acre e o Amapá, tendem a se concentrar em eventos de menor magnitude. Em resumo, embora haja uma tendência geral de aumento da área queimada com o aumento dos focos, a relação é complexa e demonstra variações significativas entre os estados da Amazônia Legal.

3.5 Análise de Dados Categóricos

Como visto ao início da seção, a única variável categórica é **_estado**, cuja análise foi feita por meio da seguinte tabela de frequência:

	Frequência Absoluta	Frequência Relativa (%)	Freq. Absoluta Acumulada	Freq. Relativa Acumulada (%)
_estado				
AC	113	11.024390	113	11.024390
AM	183	17.853659	296	28.878049
AP	114	11.121951	410	40.000000
MA	0	0.000000	410	40.000000
MT	0	0.000000	410	40.000000
PA	156	15.219512	566	55.219512
RO	134	13.073171	700	68.292683
RR	121	11.804878	821	80.097561
TO	204	19.902439	1025	100.000000

Figura 14 – Tabela de frequência de **_estado**

Verifica-se que **Amazonas** e **Tocantins** são os estados mais representados neste *dataset*, enquanto **Maranhão** e **Mato Grosso** sequer foram representados, cuja ausência explica-se pela dificuldade em coletar suas informações climáticas, o que impossibilitou sua aparição no *dataset* final. Seja como for, uma vez que a quantidade de dióxido de carbono emitida na atmosfera não deveria depender diretamente da posição geográfica, mas sim da área efetivamente queimada, provavelmente não serão necessárias nem a sub nem a sobreamostragem para compensar o desbalanço da amostra. Apesar disso, nas próximas *sprints*, pretende-se coletar mais dados, com o intuito de suavizar essas assimetrias.

3.6 Análise Temporal

Este primeiro gráfico apresenta a evolução anual da emissão de CO₂ na Amazônia Legal entre os anos de 2000 e 2023.

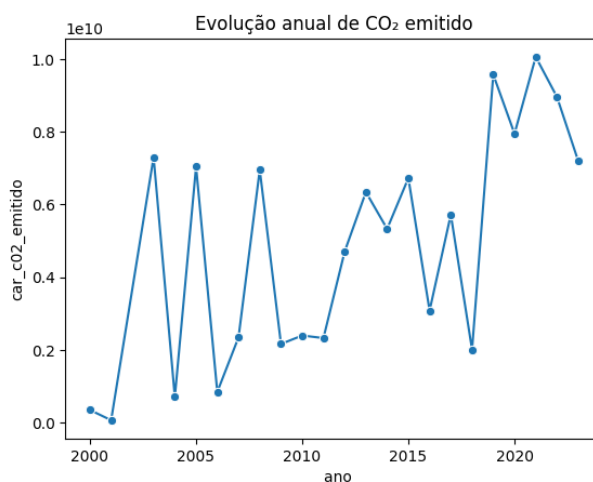


Figura 15 – Evolução anual do CO₂ emitido

Há um padrão bastante oscilante ao longo do período, mas com uma tendência geral de crescimento, principalmente a partir de 2011. Nos primeiros anos da série, entre 2000 e 2003, os valores de emissão eram bastante oscilantes. Em 2003, há um pico, ultrapassando 7 bilhões de toneladas de CO₂ emitidas, seguido por uma queda acentuada em 2004, continuando com altas oscilações até 2010. A partir de 2011, nota-se um crescimento mais contínuo e consistente nas emissões, com menores quedas abruptas e mais anos consecutivos de aumento. Entre 2013 e 2017, os valores se mantêm em patamares intermediários, entre 5 e 7 bilhões de toneladas, com leves oscilações. A partir de 2018, no entanto, há um salto significativo, com os anos de 2019 e 2021 registrando os maiores valores da série, próximos ou acima de 10 bilhões de toneladas de CO₂ emitidos. Após esse pico, observa-se uma leve queda, mas os valores permanecem bastante altos até 2023.

Este outro gráfico mostra como a área queimada na Amazônia Legal evoluiu entre os anos de 2000 e 2023.

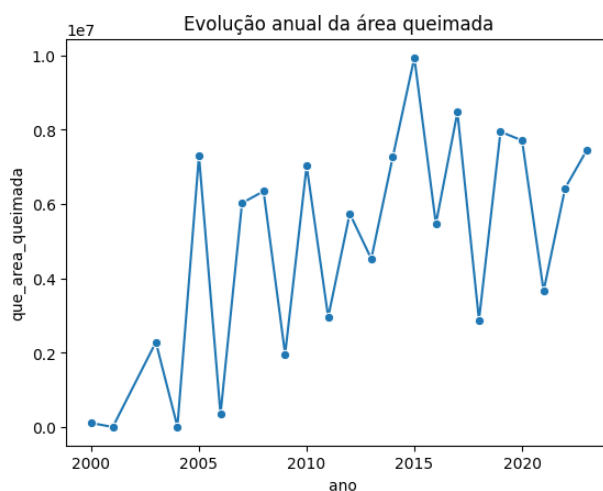


Figura 16 – Evolução anual da área queimada

É possível perceber uma tendência de aumento ao longo do tempo, com várias oscilações de um ano para o outro, que podem ter causas climáticas, políticas ambientais e econômicas. No começo, entre 2000 e 2003, os números eram mais baixos, mas a partir de 2004 já começa um crescimento forte, mesmo com altos e baixos. O ponto mais crítico foi em 2015, quando mais de 10 milhões de hectares foram queimados. Depois disso, os valores continuaram altos, variando bastante a cada ano. Entre 2020 e 2023, houve uma queda seguida de novo aumento. Mesmo nos anos a partir de 2010 que houve queda, os valores ainda são muito maiores do que os do começo da série.

Por fim, tem-se a sobreposição das séries mensais de CO₂, área queimada e quantidade de focos.

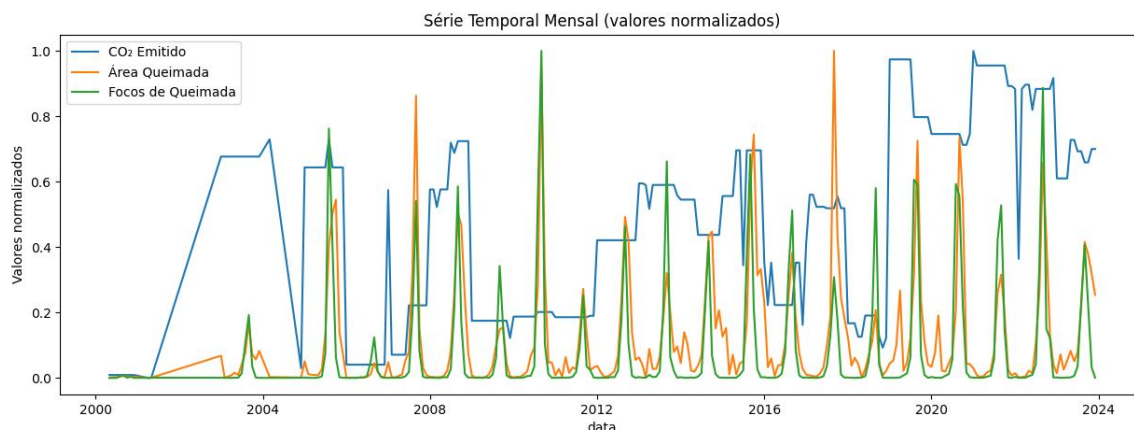


Figura 17 – Série temporal mensal de CO₂ emitido, área queimada e focos

Visualmente, há correlação entre a quantidade de área queimada e a quantidade de focos, haja vista que muitos de seus picos nas séries temporais coincidem. Apesar disso, é nítido que os picos menores de área queimada não costumam ser acompanhados por picos menores da curva de focos, que se mantém plana. Isso pode estar relacionado ao fato de que grandes focos, embora não numerosos, produzam vasta área de queima.

Quanto às emissões de carbono, a imputação dos valores anuais a todos os meses parece ter introduzido demasiada distorção, de modo que a série de CO₂ mantenha-se em alta grande parte do tempo, mesmo que desacompanhada pelas demais. Diante disso, para evitar interpretações enganosas, deve-se procurar dados propriamente mensais sobre o tópico, ainda que disponíveis em um intervalo mais curto que o escopo do projeto.

3.7 Análise Geoespacial

Neste gráfico, quanto maior a intensidade da cor, maior é a área queimada em cada estado.

Distribuição Geoespacial da Área Queimada por Estado

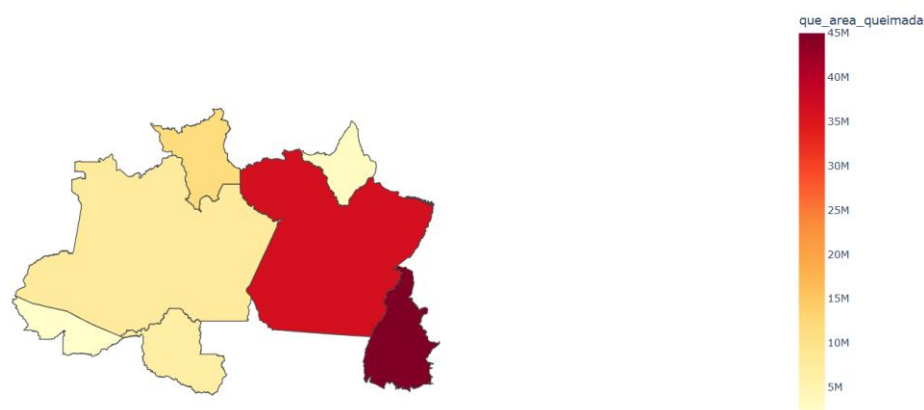


Figura 18 – Área queimada acumulada por estado

Observa-se que o Tocantins, seguido do Pará, apresenta as maiores extensões de áreas queimadas. Em contraste, o Acre e o Amapá registram as menores quantidades. Já os estados do Amazonas, Rondônia e Roraima situam-se em uma posição intermediária. Isso é condizente com a realidade, pois o Pará e o Tocantins fazem parte da chamada Fronteira Agrícola do Brasil, onde há intensa expansão de áreas para agricultura e pastagens. O fogo é usado como uma técnica barata para desmatar e “limpar” o terreno. Além disso, o Tocantins enfrenta anualmente um período de estiagem que se inicia em maio, caracterizado por altas temperaturas e baixa umidade relativa do ar. Essas condições tornam a vegetação mais suscetível ao fogo. Ademais, no Pará há grandes áreas de florestas sendo invadidas ilegalmente para fins de especulação fundiária, com o uso do fogo para marcar territórios e abrir áreas. Por fim, vale ressaltar que ausência de dados sobre Maranhão e Mato Grosso não implica que esses estados não tenham áreas queimadas.

3.8 Insights Importantes

A partir das análises acima, constata-se que, ora por subrepresentação (MA e MT), ora por imputação (emissões mensais de CO₂), distorções e enganos podem ser produzidos, como o Maranhão não ter área queimada. Apesar disso, o problema mais premente parece ser o da emissão de carbono, a qual, além de variável-alvo, mostrou suas distorções diante da discrepância entre a correlação com **que_area_queimada** antes e depois da agregação por estado. Ademais, algumas correlações mostraram-se surpreendentemente baixas, a saber:

- **que_area_queimada** ↔ **que_focos_qtd** = 0.48;
- **que_area_queimada** ↔ **cli_temp_ar_med** = 0.03.

Tal observação revela o fato de que o senso comum pode ser enganoso em algumas situações: com efeito, um pequeno número de grandes focos pode ser mais impactante que um grande número de pequenos. Por outro lado, expectativas confirmaram-se diante das fortes correlações entre variáveis “cognatas”, como **cli_umid_rel_min_med** e **cli_umid_rel_min_min**, o que pode demandar seleção de *features* no ciclo interno de pesquisa com o intuito de evitar a *Curse of Dimensionality*.

Por fim, a análise geoespacial, que permitiu relacionar a quantidade de área queimada à expansão da fronteira agrícola, aponta que não apenas os dados climático-ambientais podem ser úteis, mas também os socioeconômicos.

Diante do exposto, conclui-se, em suma, que o universo sugerido pelos dados é um mero reflexo da realidade, que tenderá a ser mais ou menos fiel conforme a qualidade e a limpeza do espelho. Assim, para a próxima *sprint*, deve-se coletar dados climáticos sobre o Maranhão e o Mato Grosso e informações mensais sobre as emissões de carbono, a fim de se averiguar seu impacto nas correlações — crucial para o aprendizado de modelos.

Referências

ECOIA PUC-RIO. **AMazonizAR**. *Homepage* da instituição, 2023. Disponível em: <https://instituto.ecoia.puc-rio.br/amazonizar/>. Acesso em: 29 mar. 2025.

IBGE. **IBGE atualiza mapa da Amazônia Legal**. Agência de Notícias IBGE, 16 jul. 2020. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/28089-ibge-atualiza-mapa-da-amazonia-legal>. Acesso em: 4 abr. 2025.

INSTITUTO NACIONAL DE METEOROLOGIA (INMET). *Dados históricos*. Brasília: INMET, [s.d.]. Disponível em: <https://portal.inmet.gov.br/dadoshistoricos>. Acesso em: 2 maio 2025.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). *Programa Queimadas*. São José dos Campos: INPE, [s.d.]. Disponível em: <https://terrabrasilis.dpi.inpe.br/queimadas>. Acesso em: 2 maio 2025.

MAPBIOMAS. *Monitor do Fogo*. São Paulo: Projeto MapBiomas, [s.d.]. Disponível em: <https://plataforma.brasil.mapbiomas.org/monitor-do-fogo>. Acesso em: 2 maio 2025.

OBSERVATÓRIO DO CLIMA. **Sistema de Estimativas de Emissões de Gases de Efeito Estufa (SEEG)**. Disponível em: <https://plataforma.seeg.eco.br/>. Acesso em: 29 mar. 2025.