Baixei um conjunto de base de dados (vários arquivos) do MapBiomas que continha os dados da quantidade de área queimada nos anos de 2019 a 2025 (até março) por mês. O conjunto estava separado em vários datasets com a quantidade área queimada separados por ano e mês. Ex: 2019 Dez Territórios mais afetados.

As bases possuiam as colunas Position, Code, Name, Area\_ha.

* Position - posição do estado com território mais afetado (1) para o menos afetado (27) - contamos com o distrito federal.
* Code - código referente a cada estado (se mantém o mesmo para cada dataset)
* Name - nome dos estados
* Area\_ha - área afetada pela queimada

Como tratei essa base?

Ela não possuia nenhum valor NaN, apenas alguns poucos zeros que não foquei no início (não os tratei).

Primeiro, o código cria um dicionário que traduz os códigos dos estados para suas siglas (por exemplo, 12 vira "AC", 27 vira "AL", e assim por diante). Também é feita uma lista com todos os meses do ano, ligando o número de cada mês com sua sigla, como "Jan", "Fev", "Mar" etc, ajudando a montar os nomes dos arquivos que vão ser lidos - a estrutura dos arquivos é "{ano} {sigla do mês} Territórios mais afetados.csv"

Depois, o código passa por cada ano, de 2019 a 2025. Como os dados de 2025 ainda não estão completos, só os três primeiros meses são considerados. Para cada mês, ele monta o nome do arquivo, lê os dados e faz algumas mudanças: adiciona colunas com o ano, o número do mês e a sigla do estado (traduzida a partir do código), renomeia a coluna da área queimada para “que\_area\_queimada” e mantém só as colunas que realmente importam ("\_ano", "\_mes", "\_estado", "que\_area\_queimada")

Todos esses arquivos tratados são guardados numa lista. No final, o código junta tudo em um único DataFrame, chamado df\_total, que reúne os dados de todos os anos e meses. Por fim, esse conjunto completo é salvo em um novo arquivo chamado dados\_area\_queimada\_total.csv.

25/04/2025:

Juntamos esse dataset com os outros que contém parâmetros relevantes e vimos que tinha muitos dados faltantes.

Percebemos que tínhamos muitas poucas tuplas do dataset da quantidade de área queimada, pois só tínhamos a quantidade de área dos anos de 2019 a 2025, enquanto outros datasets nossos com outras informações continham dados a partir dos anos 2000.

Então, fui a busca de mais dados. Consegui achar outros dados, agora dos anos de 1985 a 2023 da quantidade de área queimada por estado e mês. ALELUIAA!!

Os dados estavam na mesma fonte. A galera que fez o site precisa fazer IHC, pois foi um inferno de achar !!!

Então, lá vai eu tratar essa base:

Como eram os dados:

Os arquivos eram separados por estados. Baixei apenas os arquivos referentes aos estados da amazônia legal( Me poupa trabalho !!! ).

A primeira coluna era chamada Classe e continha os valores referentes aos meses do ano (escrito por extenso e letra maiúscula). As colunas seguintes eram referentes ao anos de 1985 a 2023 e o valor presente na interseção de uma linha (mês) com uma coluna (ano) representa a quantidade de área queimada registrada naquele mês e ano.

Como os dados foram tratados:

* Lê os dados do arquivo como uma tabela (DataFrame).
* Converte o formato da tabela original (onde os anos eram colunas) para outro formato: cria três colunas -> uma para o ano (\_ano), uma para o valor da área queimada (que\_area\_queimada) e outra com os meses (Classe).
* Transforma o nome do mês no número do mês usando o dicionário de meses ( contém o nome do mês e seu número equivalente) e mapeando.
* Adiciona uma coluna chamada \_estado com a sigla do estado, tirada do nome do arquivo (nome do arquivo começa com a sigla do estado referente).
* Mantém apenas as colunas importantes: ano, mês, estado e área queimada.

26/04/2025:

AGORA FALTA INTEGRAR AS DUAS BASES

Não vou escrever aqui como eu fiz, pois documentei muito bem no código, e todos os meus pensamentos estão lá kkkkk

30/04/2025:

* Caracterização do domínio de dados com medidas estatísticas
* Caracterizar domínio de dados segundo o tipo de variável