**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS  
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA**

**Pós-graduação *Lato Sensu* em Ciência de Dados e Big Data**

**Monique Ellen da Silva Acacio**

**QUEIMADAS FLORESTAIS NO BRASIL**

Belo Horizonte

2020

**Monique Ellen da Silva Acacio**

**QUEIMADAS FLORESTAIS NO BRASIL**

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Ciência de Dados e Big Data como requisito parcial à obtenção do título de especialista.

Belo Horizonte

2020 **SUMÁRIO**

[1. Introdução 4](#_Toc52030955)

[1.1. Contextualização 4](#_Toc52030956)

[1.2. O problema proposto 5](#_Toc52030957)

[2. Coleta de Dados 7](#_Toc52030958)

[3. Processamento/Tratamento de Dados 11](#_Toc52030959)

[4. Análise e Exploração dos Dados 14](#_Toc52030960)

[5. Criação de Modelos de Machine Learning 15](#_Toc52030961)

[6. Apresentação dos Resultados 15](#_Toc52030962)

[7. Links 15](#_Toc52030963)

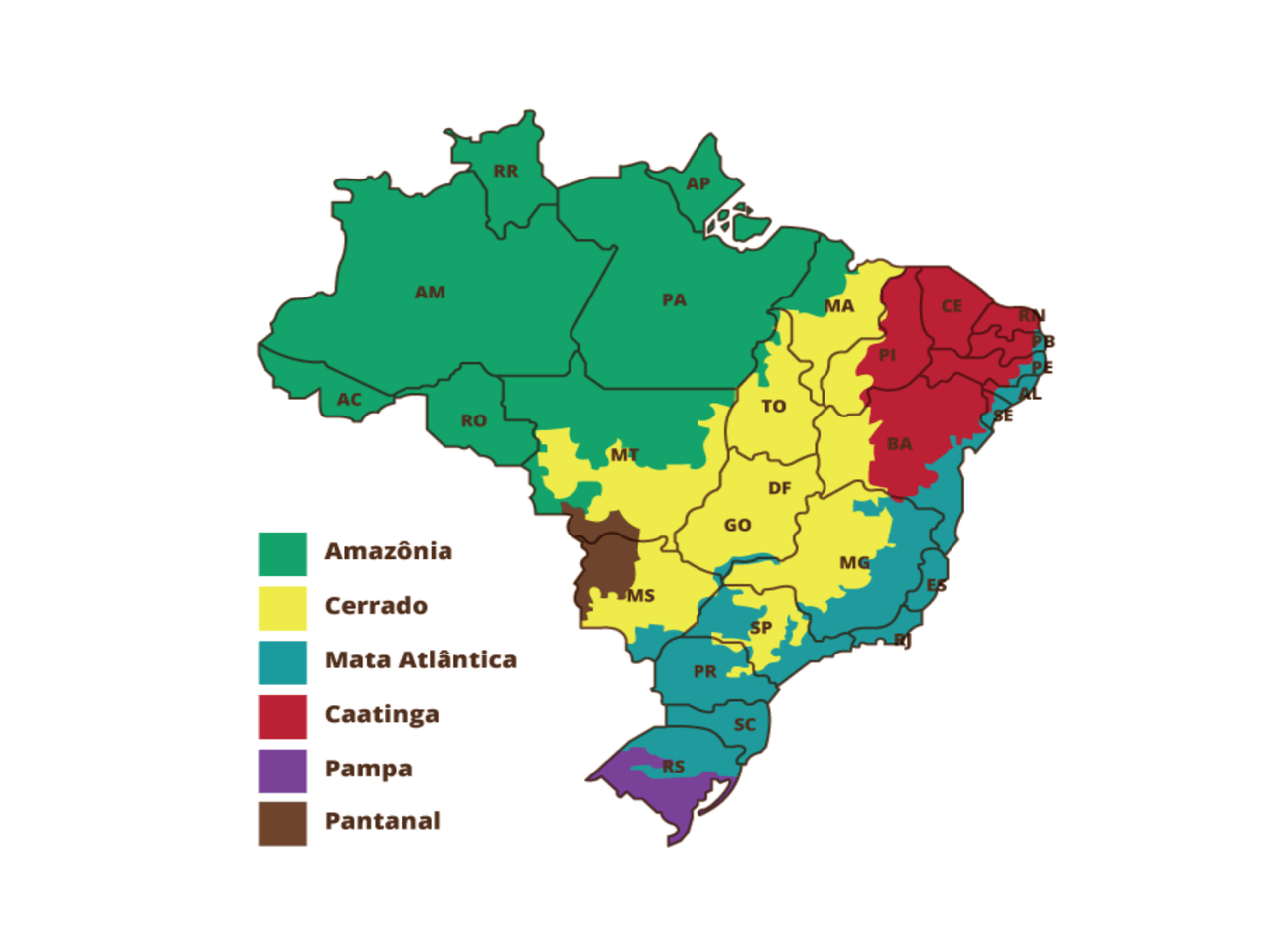
[REFERÊNCIAS 17](#_Toc52030964)

# 1. Introdução

## 1.1. Contextualização

Todos os anos é muito comum ver notícias sobre queimadas e incêndios florestais em todo mundo, este tipo de ocorrência tem ganhado cada vez mais visibilidade na mídia nacional e internacional. Recentemente temos visto muitos relatos de animais silvestres encontrados em zona urbana, possivelmente um reflexo do impacto das queimadas em seu habitat natural.

O Brasil é formado por seis biomas de características distintas: Amazônia, Caatinga, Cerrado, Mata Atlântica, Pampa e Pantanal. Cada um desses ambientes abriga diferentes tipos de vegetação e de fauna. Como a vegetação é um dos componentes mais importantes da biota, seu estado de conservação e de continuidade definem a existência ou não de hábitats para as espécies, a manutenção de serviços ambientais e o fornecimento de bens essenciais à sobrevivência de populações humanas. Para melhor entendimento da localização destes biomas nos estados brasileiros observe o mapa abaixo.



Para a perpetuação da vida nos biomas, é necessário o estabelecimento de políticas públicas ambientais, a identificação de oportunidades para a conservação, uso sustentável e repartição de benefícios da biodiversidade.

Sabendo da importância dos biomas brasileiros, o objetivo deste projeto é a realização de análise exploratória a fim de identificar a frequência e outros padrões nas queimadas e incêndios florestais que ocorreram no Brasil nos últimos 15 anos, de 2005 a 2019. Desta forma auxiliar as autoridades no direcionamento de recursos para controle e prevenção das queimadas. Os dados utilizados são disponibilizados pelo Instituto Nacional de Pesquisas Espaciais, INEP, através do programa Queimadas.

A coleta dos dados, elaboração e resultado das análises são apresentados no decorrer deste documento.

## 1.2. O problema proposto

A proposta deste trabalho é a realização de análise exploratória dos focos de queimada pelo país ao longo dos anos. Para melhor visualização do problema foram respondidas as perguntas propostas nos 5-Ws:

Por que esse problema é importante?

Entender a frequência dos incêndios florestais em uma série temporal pode ajudar a tomar medidas para evitá-los, ou direcionar recursos para que os locais mais afetados possam resolver quaisquer incidentes o mais rápido possível. Ser capaz de apontar onde e quando essa frequência é mais observada pode dar alguma clareza sobre qual é a real situação.

Quais serão os dados utilizados e suas fontes?

Os dados utilizados foram disponibilizados pelo Instituto Nacional de Pesquisas Espaciais, INEP, através do Programa Queimadas disponível no portal <http://queimadas.dgi.inpe.br/queimadas/portal>.

O dataset inclui dados como data, estado, município, bioma e etc para cada foco de queimada identificado por satélite. Os dados foram coletados no DBQueimadas, que é um sistema de monitoramento via satélite disponibilizado no portal do Programa Queimadas.

Quais os objetivos com essa análise? O que iremos analisar?

O objetivo desta análise é identificar a frequência das queimadas por estado e por bioma além identificar padrões que determinem épocas do ano com maior número de ocorrências para auxiliar as autoridades no correto direcionamento de recursos para combate e prevenção das queimadas.

Quais são aspectos geográficos das análises?

Abrange todos os 27 estados brasileiros, identifica os municípios, latitude e longitude de cada foco de queimada.

Qual o período está sendo analisado?

O dataset possui dados dos últimos 15 anos, de 2005 até 2019. Com registros diários de cada foco de queimada.

# 2. Coleta de Dados

Este projeto trabalha com um único dataSet obtido através do sistema BDQueimadas, onde são reportados os focos de queimada, conforme detalhamento abaixo.

Os dados foram coletados no dia 22/09 através do Portal Queimadas, acesso através do link <http://queimadas.dgi.inpe.br/queimadas/portal>. Ao acessar o portal, foi utilizado o link ‘Download de dados’ dentro da sessão ‘SISTEMAS DE MONITORAMENTO > BDQueimadas’.



Figura 1 – Link para download do dataset no Portal Queimadas/INEP.

O link mencionado direciona para a funcionalidade de download do sistema BDQueimadas, infelizmente o sistema não permite realizar o download dos dados em período maior que 366 dias. Então, para coletar os dados utilizados neste projeto foram realizadas 15 consultas, uma para cada ano de 2005 até 2019, considerando data e hora de início 01/01/yyyy 00:00 e data e hora de término 31/12/yyyy 23:59 para cada ano. Outras informações do preenchimento do formulário de download estão listadas abaixo:

* Deve ser informado um email para onde o link de download será enviado.
* Foi selecionado apenas o Brasil e opção ‘Todos os estados’
* Foi informado data e hora de início e término para coleta dos dados, lembrando que o período máximo é de 366 dias.
* Na definição de ‘Focos dos Satélites’ foi selecionado ‘Satélite de referência’
* Na definição de ‘Foco nos Biomas’ foi selecionado ‘TODOS’
* Selecionado o formato de exportação CSV
* Ao clicar em ‘Exportar’ o sistema envia o link de download para o email informado.

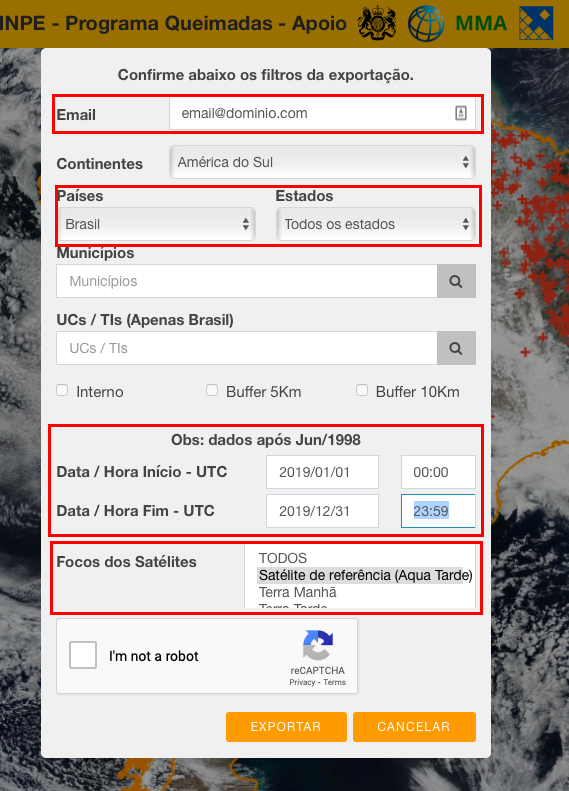


Figura 2 - Formulário de exportação de dados – Sistema DBQueimadas

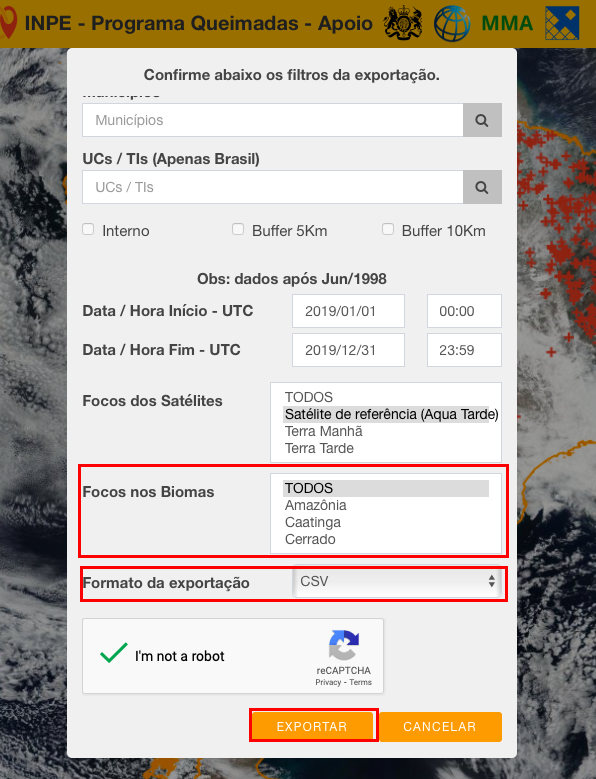


Figura 3 - Formulário de exportação de dados – Sistema BDQueimadas

A exportação dos dados resultou em 15 arquivos CSVs, uma para cada ano, todos com as mesmas estrutura de dados, onde cada linha representa o registro de um foco de queimada. Na etapa de processamento dos dados todos os CVSs são concatenados em um único DataFrame.

O dicionário de dados descrito neste documento foi disponibilizado pelo INEP através do link abaixo em resposta a questão número 40.

<http://queimadas.dgi.inpe.br/queimadas/portal/informacoes/perguntas-frequentes>

|  |  |  |
| --- | --- | --- |
| Nome da coluna/campo | Descrição | Tipo (identificação automática) |
| ID | Identificador único do registro no banco de dados. Formado pela junção dos atributos (Latgms + Longms + Data + hora) removidos os espaços | string |
| datahora | Horário de referência da passagem do satélite segundo o fuso horário de Greenwich (GMT); https://pt.wikipedia.org/wiki/Greenwich\_Mean\_Time; Representada em Hora (2 digitos) + Minutos (2 digitos) + Segundos (2 digitos) | string |
| satelite | Nome do algoritmo utilizado e referencia ao satélite provedor da imagem | string |
| país | Nome do País (nível 0 do Database of Global Administrative Areas - GADM) | string |
| estado | Nome do estado (nível 1 do Database of Global Administrative Areas - GADM) http://www.gadm.org | string |
| municipio | Nome do município. Para o Brasil foi utilizado como referência o dado do IBGE 2000 (http://mapas.ibge.gov.br/bases-e-referenciais/bases-cartograficas/malhas-digitais.html) | string |
| bioma | Nome do Bioma segundo referência do IBGE 2004 (http://www.ibge.gov.br/home/presidencia/noticias/21052004biomashtml.shtm). Para outros países o campo fica vazio (NULL) | string |
| diasemchuva | Número de dias sem chuva até a detecção do foco | integer |
| precipitacao | Valor da precipitação acumulada no dia até o momento da detecção do foco | double |
| riscofogo | Valor do Risco de Fogo previsto para o dia da detecção do foco | double |
| latitude | Latitude do centro do píxel de fogo ativo apresentada em unidade de graus decimais. | doble |
| longitude | Longitude do centro do píxel de fogo ativo apresentada em unidade de graus decimais | doble |
| frp | Fire Radiative Power, MW (megawatts) | doble |

Na etapa de processamento os campos satélite, diasemchuva, precipitação, riscofogo e frp serão removidos pois estas informações não são relevantes para o contexto das análises deste projeto.

# 3. Processamento/Tratamento de Dados

Para processamento, tratamento e análise dos dados foi utilizado a linguagem Python, para melhor visualização das etapas foi utilizado Jupyter Notebook. No relato abaixo estão todas as etapas realizadas para processamento e tratamento dos dados, no item 7 deste documento está o link para download do Jupyter Notebook utilizado e os dados coletados.

Importação das bibliotecas

Criada para a linguagem Python para manipulação e análise de dados, Pandas é a principal biblioteca utilizada neste projeto. Em particular, esta biblioteca oferece estruturas e operações para manipular tabelas numéricas e séries temporais através dos DataFrames.

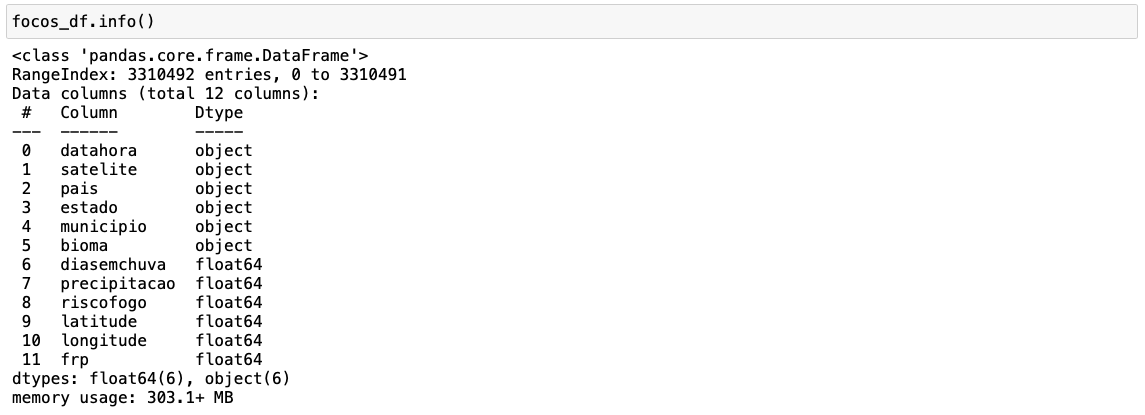


Importação dos dados

Devido a limitação de período durante a exportação dos dados, foram gerados 10 arquivos CSVs sendo uma para cada ano de 2010 até 2019. Como todos os arquivos possuem a mesma estrutura, optei por concatená-los em um único DataFrame durante o processo de importação. Abaixo é possível observar os 3 primeiros registros após importação.



Em seguida foi executada verificação de tamanho e tipo das colunas do DataFrame.

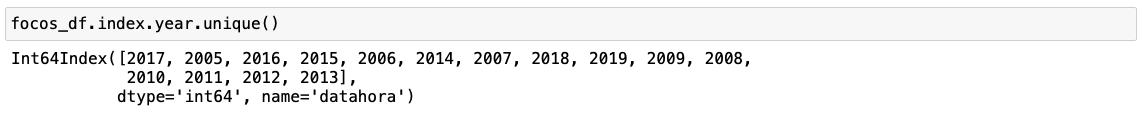


O DataFrame possui o total de 3.310.492 entradas, porém a coluna *datahora* não está no formato adequado. Para melhor utilização das funcionalidades de série temporal através do DataFrame Pandas, decidi alterar o dtype da coluna *datahora* para datetime e defini-lá como index do DataFrame.





Após a definição da coluna *datahora* como DateTimeIndex, fiz uma verificação do processo de importação dos dados conferindo todos os anos existentes no DataFrame.



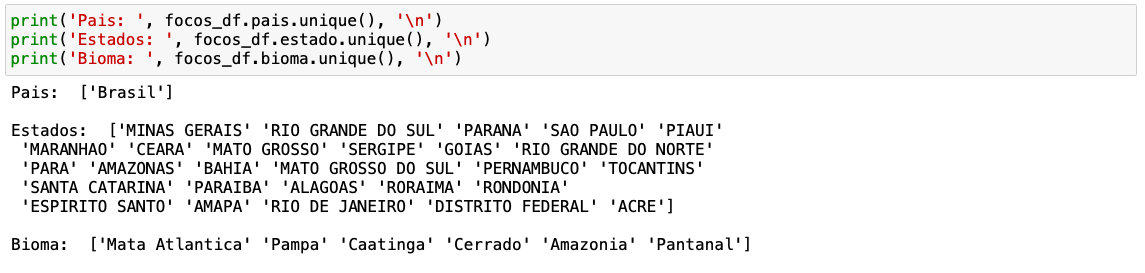
Verificação de valores nulos



As colunas *diasemchuva*, *precipitacao*, *riscofogo* e *frp* apresentaram valores nulos, porém estes dados não serão utilizados durante as análises deste projeto. Sendo assim, optei por remover estas colunas e também a coluna *satelite* que também não trás relevância as análises realizadas por somente informar o nome do satélite de referência. Na coluna bioma apenas um registo apresentou valor nulo, como um registro não vai gerar impacto na análise optei por removê-lo também.



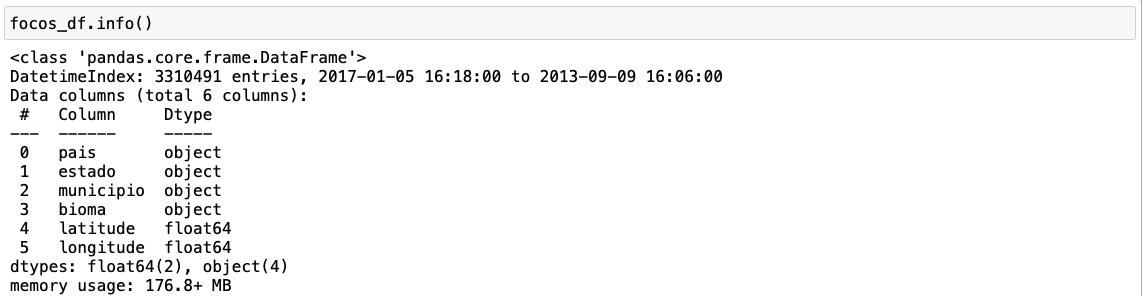
Verificação de valores únicos



Aqui é possível observar que todos os estados do Brasil estão representados no dataset, assim como todos os biomas.

DataFrame final

O resultado do processamento e tratamento dos dados é o DataFrame abaixo. Total de registros: 3.310.491



# 4. Análise e Exploração dos Dados

Nessa seção você deve mostrar como foi realizada a análise e exploração dos seus. Mostre as hipóteses levantadas durante essa etapa e os padrões e *insights* identificados.

# 5. Criação de Modelos de Machine Learning

Conforme o documento de instruções para o TCC, essa etapa não é obrigatória, mas é fortemente recomendada. Caso você crie modelos de *Machine Learn*ing em seu projeto, nessa seção você irá descrever as ferramentas utilizadas. Se você utilizou ferramentas visuais como Knime e Rapid Miner, coloque aqui um print do seu modelo. Caso você tenha escrito scripts em Python, por exemplo, coloque aqui o seu script. Explique as *features* utilizadas, justifique a escolha por determinado modelo, os parâmetros utilizados, etc.

# 6. Apresentação dos Resultados

Nessa seção você deve apresentar os resultados obtidos. Apresente gráficos, dahsboards, conte a sua história de forma bastante criativa. Aqui você pode utilizar os modelos de Canvas propostos por Dourard (clique [aqui](https://www.louisdorard.com/machine-learning-canvas)) ou por Vasandani (clique [aqui](https://towardsdatascience.com/a-data-science-workflow-canvas-to-kickstart-your-projects-db62556be4d0)).



# 7. Links

Aqui você deve disponibilizar os links para o vídeo com sua apresentação de 5 minutos e para o repositório contendo os dados utilizados no projeto, scripts criados, etc.

# REFERÊNCIAS

Um projeto de Ciência de Dados não requer revisão bibliográfica. Portanto, a inclusão das referências não é obrigatória. No entanto, caso você deseje incluir referências relacionadas às tecnologias ou às metodologias usadas em seu trabalho, relacione-as de acordo com o modelo a seguir.

<https://www.mma.gov.br/biomas.html>

<http://queimadas.dgi.inpe.br/~rqueimadas/documentos/RiscoFogo_Sucinto.pdf>

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.

SOBRENOME DO AUTOR, Nome do autor. **Título do livro ou artigo.** Cidade: Editora, ano.