Projeto Aplicado IV



Informações



Curso: Tecnologia em Ciências de Dados

Semestre: 5º

Componente curricular: Projeto Aplicado IV

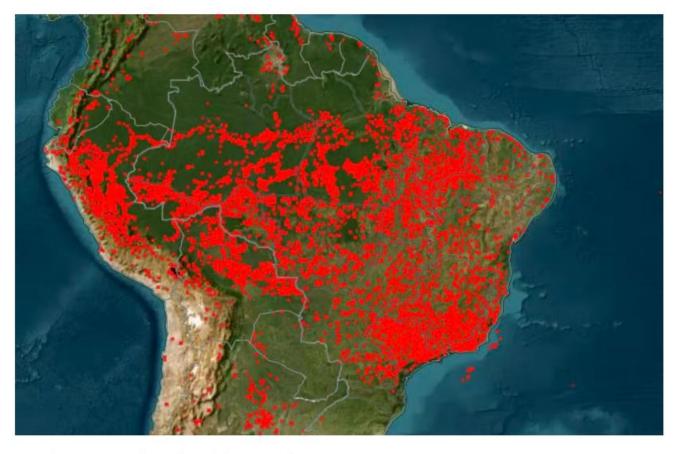
Integrantes e TIA:

- > Caroline Ribeiro Ferreira 10408052
- > Leonardo dos Reis Olher 10407752
- **➤ Liliane Gonçalves de Brito Ferraz 22501142**
- > Múcio Emanuel Feitosa Ferraz Filho 22515925

Título:



Análise de Séries Temporais de Queimadas no Brasil para a Gestão Ambiental e o Desenvolvimento Sustentável



Mapa da Nasa mostra focos de incêndio no Brasil. — Foto: Nasa

Objetivos do Projeto





Aumentar o conhecimento científico sobre as causas e os impactos das queimadas no Brasil

Desenvolver ferramentas para auxiliar na gestão ambiental e na tomada de decisões 75)

Contribuir para a construção de um futuro mais sustentável para o país

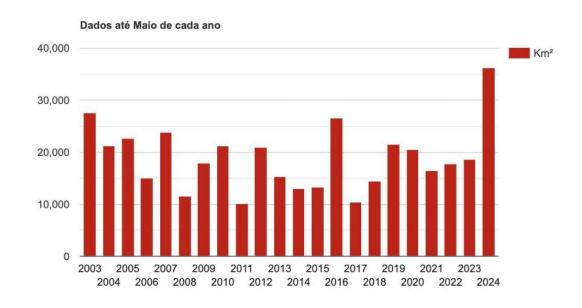
Introdução



As queimadas florestais no Brasil representam uma grave ameaça à biodiversidade, ao clima e à qualidade de vida da população. Anualmente, vastas áreas de nossos biomas são consumidas pelo fogo, resultando em perdas irreparáveis para o meio ambiente e a economia.

Em 2024, os impactos das queimadas atingiram números recordes, assim sendo nosso tema escolhido para estudo devido sua relevância sobre os impactos ambientais.

ÁREA QUEIMADA NO BRASIL ATÉ MAIO/2024



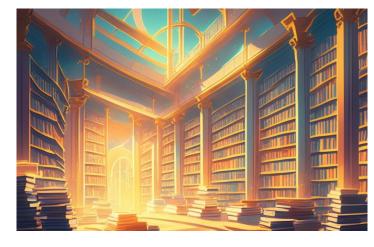
Dataset

M

Para o desenvolvimento deste projeto de Análise de Séries Temporais de Queimadas no Brasil, foi escolhido um dataset público no site da INPE (Instituto Nacional de Pesquisas Espaciais), considerando as queimadas que ocorreram entre os anos de 2019 a 2023.

A base de dados possui 12 atributos: Data/Hora, Satélite, País, Estado, Município, Bioma, N. Dias Sem Chuva, Precipitação, Risco Fogo, Latitude. Longitude e Potência Radiativa do Fogo.







POR QUE FAZER UMA ANÁLISE DE SÉRIES TEMPORAIS SOBRE QUEIMADAS?



A análise de séries temporais de queimadas é crucial para:

- Identificar Padrões Sazonais: Reconhecer épocas de maior risco, como durante a estação seca.
- Avaliar Tendências de Longo Prazo: Monitorar aumento ou redução de queimadas ao longo dos anos.
- Quantificar Impactos Ambientais: Estimar danos à biodiversidade, qualidade do ar e estoques de carbono.
- Apoiar Previsão e Planejamento: Prever períodos críticos e otimizar recursos preventivos.
- Revelar Causas Subjacentes: Correlacionar queimadas com mudanças climáticas, uso do solo ou pressões econômicas.

Etapas do decorrer do projeto







PROBLEMA

DADOS

Compreender os padrões temporais Importação de e espaciais das bibliotecas e queimadas no pacotes, Brasil, identificar carregamento os principais dos dados. fatores limpeza, influenciadores e transformação e prever futuras agregação. ocorrências.



ANÁLISE Exploratória

Realizar
visualização
gráficas com a
distribuição
temporal das
queimadas e
aplicar estatística
descritiva.



MODELO

A escolha do modelo para séries temporais como ARIMA e SARIMA e também de Machine Learning.



TREINAMENTO

Divisão dos dados em treinamento, validação e teste, utilizando métricas como RMSE, MAE e accuracy.



AVALIAÇÃO

Identificar os principais fatores que influenciam a ocorrência de queimadas e as regiões mais propensas a ocorrências de queimadas.



RESULTADO

Exposição dos principais achados e suas implicações para a gestão ambiental.

Elaborado pelo autor, 2024.

Análise Exploratória





ETAPAS ANÁLISE EXPLORATÓRIA



Importação de Dados: O conjunto de dados foi importado para um DataFrame a fim de possibilitar o tratamento e manipulação necessários.

Identificação de Dados para Tratamento: Realizamos uma análise preliminar para identificar dados incompletos, inconsistentes ou que necessitem de tratamento adicional Visualização Inicial dos Dados:

Foi gerado um gráfico inicial que possibilitou uma visualização prévia das variáveis, ajudando a verificar a presença de outliers e a distribuição geral dos dados.

Tratamento de Dados: As ações tomadas nesta etapa incluíram: Conversão de colunas para o tipo datetime, criação de colunas, remoção de colunas que não apresentam variabilidade, exclusão de outliers e valores nulos e extração de informações dos dados

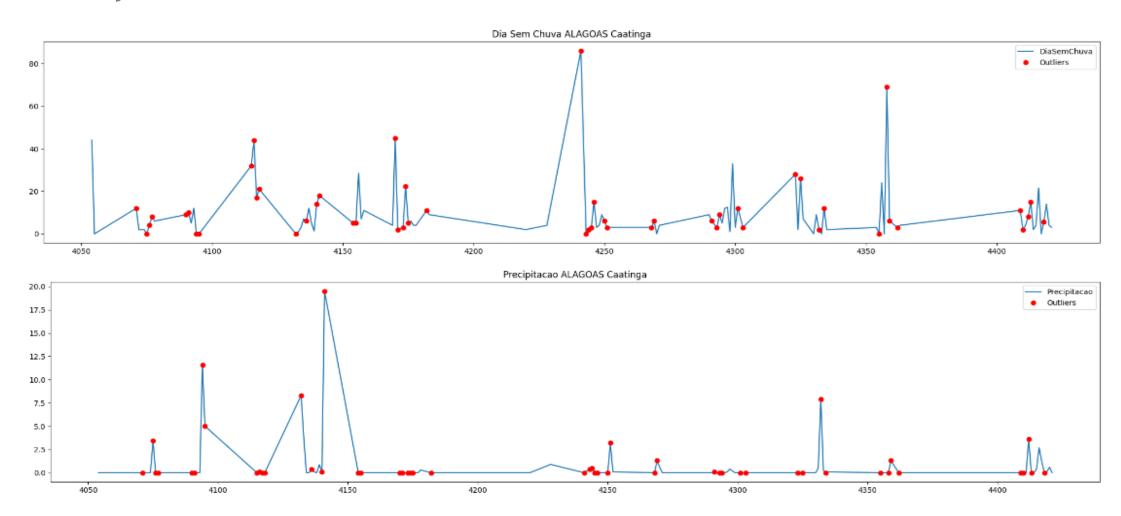
Tratamento de Outliers: Foi utilizada a técnica DBSCAN para identificar e remover outliers que pudessem distorcer as análises.

Agrupamento dos Dados: Para análises específicas, os dados foram agrupados por diferentes categorias, como bioma e estado, facilitando a identificação de padrões por região

ALGUNS GRÁFICOS UTILIZADOS NA ANÁLISE



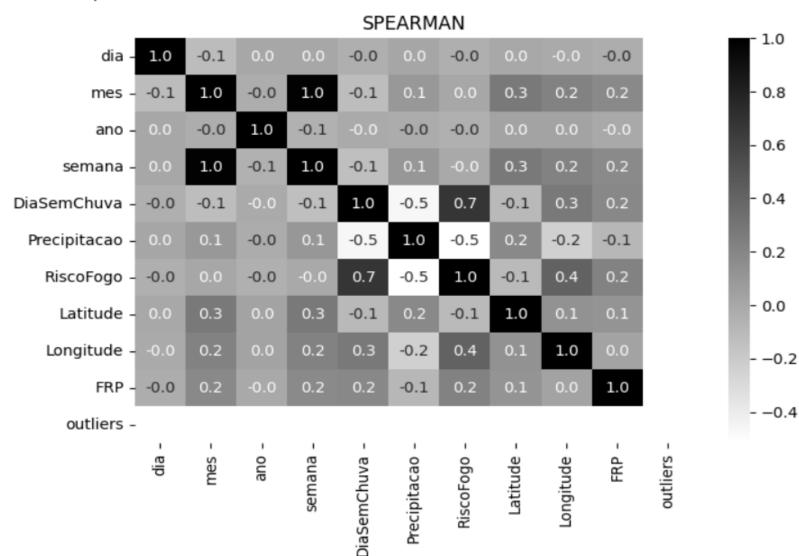
Identificação de outliers:



ALGUNS GRÁFICOS <u>UTILIZADOS NA ANÁLISE</u>



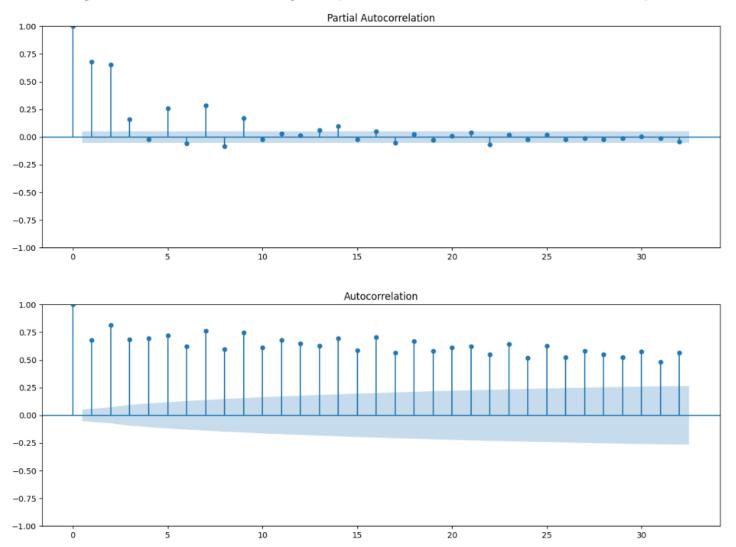
Correlação de Spearman:



ALGUNS GRÁFICOS UTILIZADOS NA ANÁLISE



Autocorrelação e autocorrelação parcial com as média de queimadas por dia:



Aprendizado de Máquina





Modelo ARIMA x SARIMA



Modelo ARIMA

É um modelo estatístico amplamente utilizado para previsão de séries temporais. Ele combina três componentes principais:

- AR (AutoRegressivo): Usa a relação entre valores passados da série.
- I (Integração): Torna a série estacionária por meio de diferenças sucessivas.
- MA (Média Móvel): Modela a dependência entre erros de previsão.

O ARIMA é eficaz para séries univariadas e estacionárias, oferecendo flexibilidade para capturar padrões históricos e tendências de curto prazo.

Modelo SARIMA

O SARIMA é uma extensão do ARIMA que incorpora componentes sazonais, sendo ideal para séries temporais com padrões periódicos. Ele adiciona parâmetros sazonais aos três componentes do ARIMA (AR, I e MA), além de incluir um período sazonal (S) específico. Com isso, o SARIMA consegue modelar tanto tendências de longo prazo quanto flutuações sazonais, sendo amplamente aplicado em fenômenos cíclicos, como vendas sazonais ou dados climáticos.

Modelo escolhido: SARIMA



Principais características:

Modelagem de Sazonalidade:

Adiciona componentes para capturar padrões sazonais, como flutuações periódicas em intervalos regulares. Ele utiliza termos adicionais para modelar a autorregressão, integração e médias móveis sazonais.

Flexibilidade para Dados
Não Estacionários:
Assim como o ARIMA, o
SARIMA pode lidar com
séries não estacionárias,
aplicando diferenças tanto
para a tendência (não
sazonal) quanto para a
sazonalidade.

Parâmetro de Período
Sazonal (S):
O modelo inclui um
parâmetro que especifica o
número de observações por
ciclo sazonal, permitindo
ajustar o modelo para
diferentes frequências
temporais.

AVALIAÇÃO E RESULTADO





MÉTRICAS DE AVALIAÇÃO



Akaike Information Criterion (AIC):

Métrica estatística utilizada para avaliar a qualidade de modelos ajustados, especialmente em problemas de regressão e séries temporais.

Mede o equilíbrio entre o ajuste do modelo aos dados e sua complexidade, ajudando a selecionar o modelo mais eficiente.

- Principais características:
 - Equilíbrio entre ajuste e simplicidade: Penaliza modelos mais complexos, incentivando a escolha de modelos parcimoniosos que evitem o sobreajuste.
 - Comparação de Modelos: É usado para comparar modelos diferentes: o modelo com o menor valor de AIC é geralmente preferido, pois é considerado o mais eficiente em descrever os dados com o menor número de parâmetros.
 - Não interpreta qualidade absoluta: O AIC não indica se um modelo é "bom" ou "ruim" isoladamente, mas apenas em relação a outros modelos analisados.

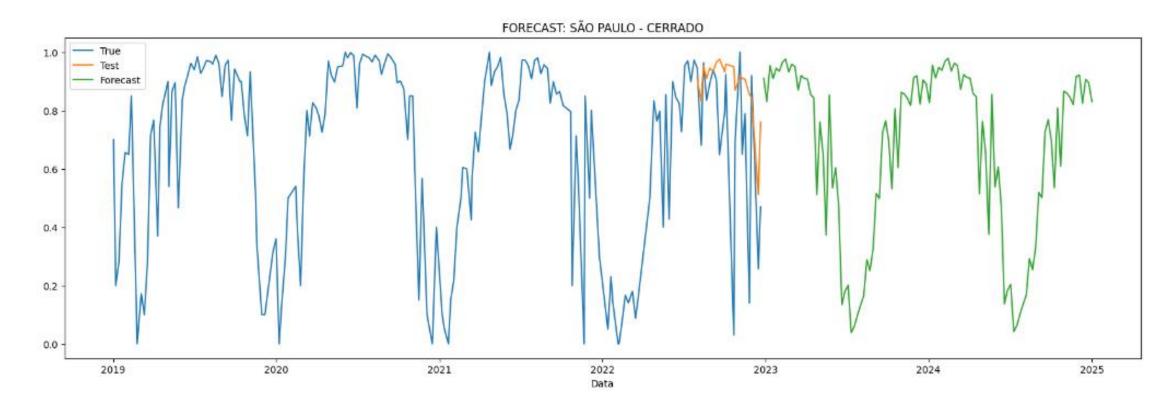
AVALIAÇÃO DO MODELO



As avaliações foram feitas de forma segmentada, considerando as diferentes regiões e biomas. No exemplo a seguir vemos a avaliação da previsão no cerrado do estado de São Paulo:

AIC: -2.472

BEST MODEL: ARIMA(1,1,1)(0,1,1)[52]



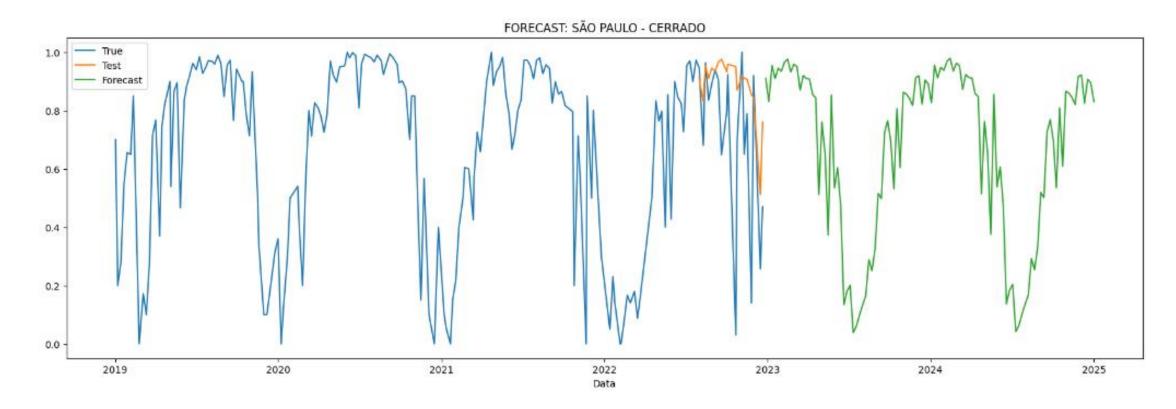
RESULTADOS



Além da avaliação do modelo, foi feita a previsão considerando cada situação examinada e foi encontrado o melhor modelo de previsão:

AIC: -2.472

BEST MODEL: ARIMA(1,1,1)(0,1,1)[52]



RESULTADOS



Os resultados da análise indicam uma variabilidade significativa no desempenho dos modelos ARIMA entre os diferentes estados e biomas."

"Estados como Amazonas e Goiás apresentaram modelos com melhor ajuste aos dados, sugerindo que os padrões de risco de fogo nessas regiões são mais previsíveis."

"Em contrapartida, estados como Ceará e Maranhão apresentaram maiores dificuldades na modelagem, indicando a necessidade de investigações mais aprofundadas."

"A presença de termos sazonais em todos os modelos confirma a importância da variabilidade climática anual na ocorrência de incêndios."

"A necessidade de diferenciação dos dados sugere a existência de tendências de longo prazo nos riscos de incêndio."

"A diversidade de modelos indica que as dinâmicas dos incêndios são influenciadas por fatores locais específicos, exigindo abordagens personalizadas para cada região."

"Os modelos ARIMA, embora eficazes em algumas regiões, podem não ser suficientes para capturar toda a complexidade dos fenômenos. A exploração de modelos mais complexos, como SARIMA ou modelos de aprendizado de máquina, pode ser promissora para futuras pesquisas."

CONCLUSÃO



A análise de séries temporais sobre queimadas no Brasil é fundamental para compreender padrões, causas e impactos desse fenômeno, oferecendo subsídios para ações mais eficazes na gestão ambiental. Utilizando dados do INPE e o modelo SARIMA, o projeto permitiu identificar tendências e sazonalidades, além de realizar previsões que podem orientar políticas públicas e estratégias preventivas em diferentes biomas e regiões.

O estudo reforça a importância da ciência de dados para enfrentar desafios ambientais, promovendo a preservação dos biomas e contribuindo para o desenvolvimento sustentável do país. A integração entre análise científica e ferramentas preditivas é um passo essencial para um futuro mais resiliente e responsável.

Obrigado(a)!

