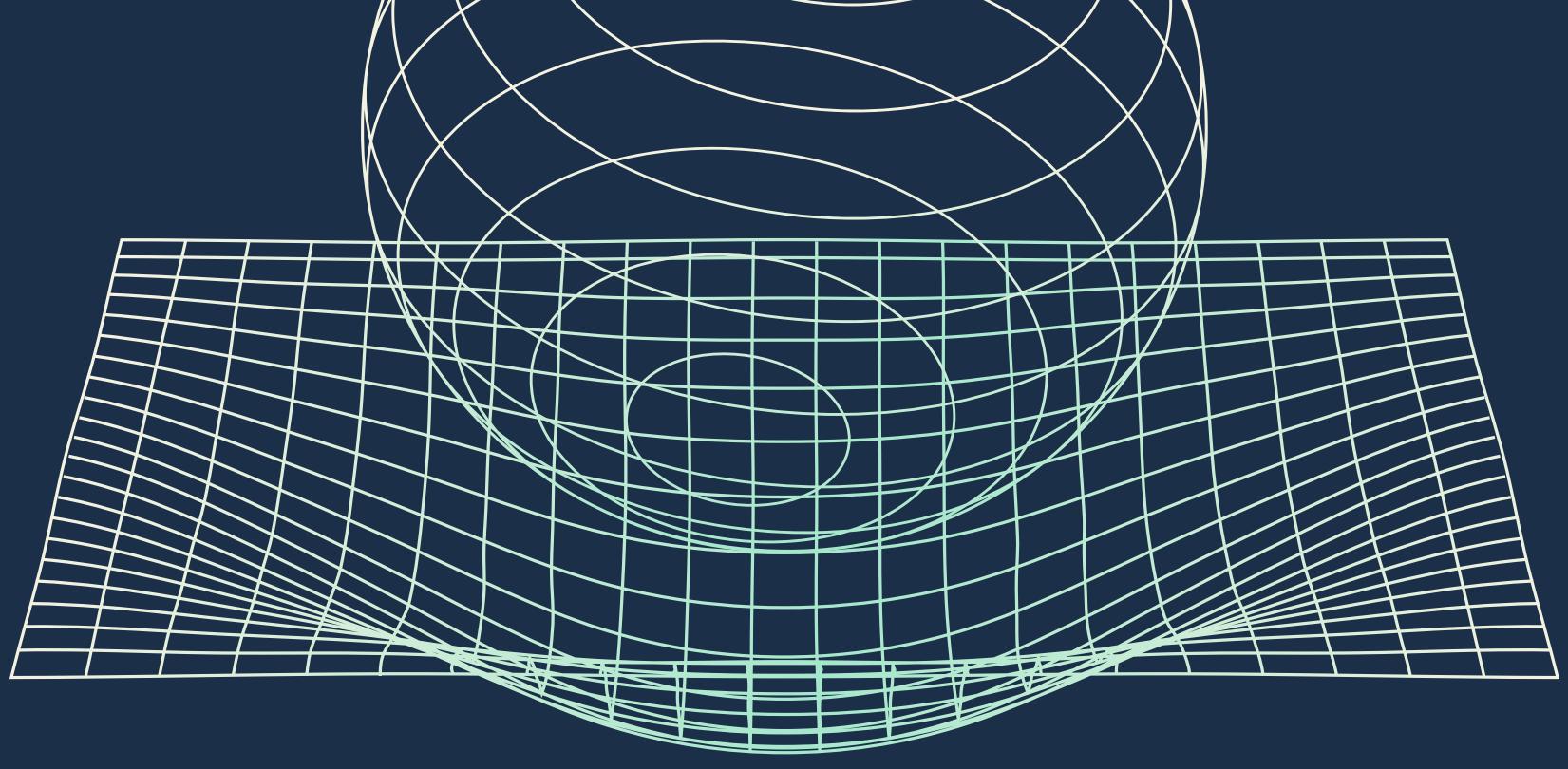




BI Master | LUI
Rio de Janeiro | 09/06/2024

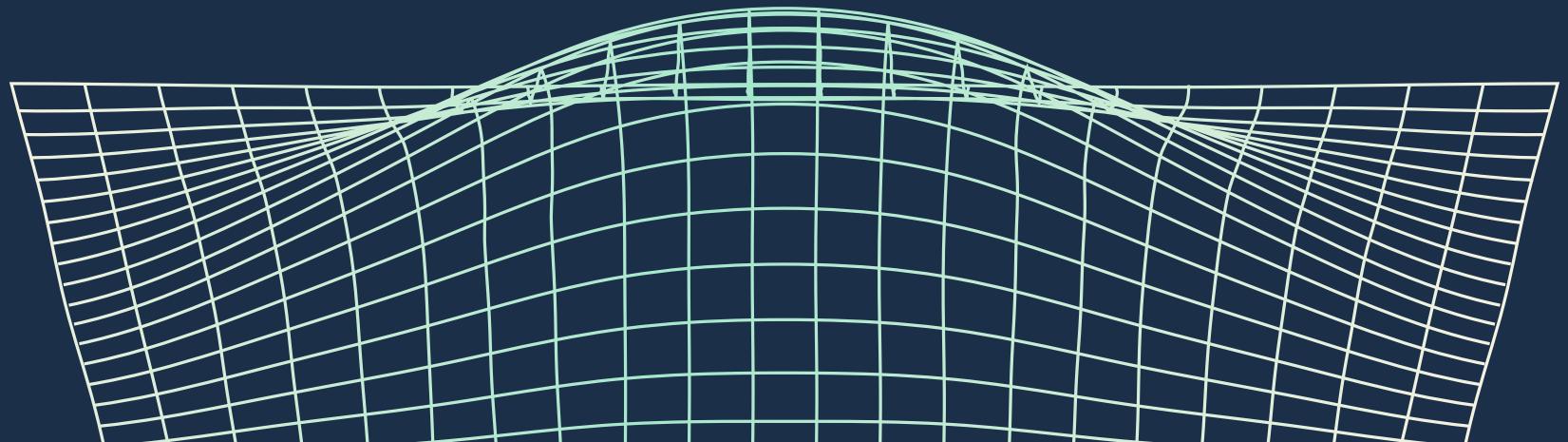
PREVISÃO DE SÉRIES TEMPORAIS METEOROLÓGICAS UTILIZANDO DADOS DO INMET

Apresentado por:
Thiago Sousa

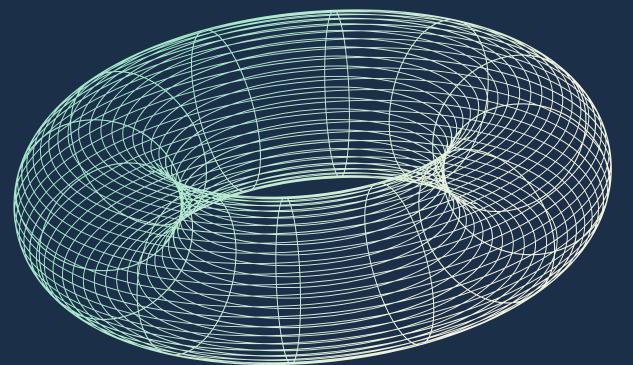


Objetivo do Trabalho

A proposta do trabalho é estruturar um processo de extração e tratamento de dados meteorológicos do INMET para realização de previsões de séries temporais.

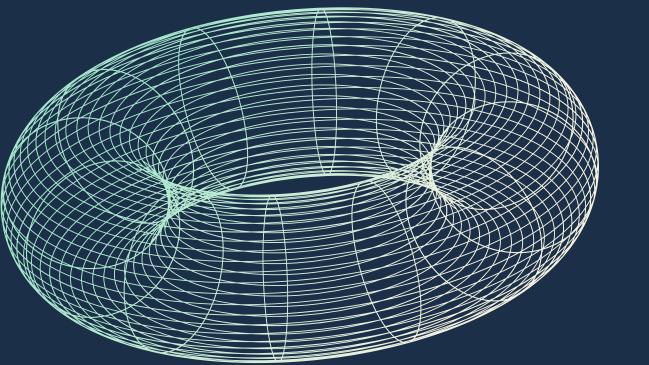


Objetivos e Metas



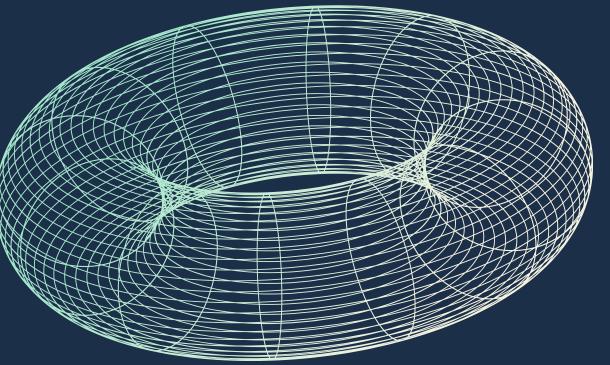
Meta # 1

Estabelecer um processo
de extração dos dados



Meta # 2

Realizar o tratamento
adequado dos dados



Meta # 3

Realizar previsões das
séries temporais

INMET

A missão do Instituto Nacional de Meteorologia (INMET), órgão do Ministério da Agricultura e Pecuária, é agregar valor à produção no Brasil por meio de informações meteorológicas. Esta missão é alcançada por meio de monitoramento, análise e previsão de tempo e de clima, que se fundamentam em pesquisa aplicada, trabalho em parceria e compartilhamento do conhecimento, com ênfase em resultados práticos e confiáveis.

O Sistema de Coleta e Distribuição de Dados Meteorológicos do Instituto (temperatura, umidade relativa do ar, direção e velocidade do vento, pressão atmosférica, precipitação, entre outras variáveis) é dotado de estações de sondagem de ar superior (radiossonda); estações meteorológicas de superfície, operadas manualmente; e a maior rede de estações automáticas da América do Sul.

A rede de estações meteorológicas automáticas utiliza o que há de mais moderno internacionalmente. Os dados coletados por essa rede são disseminados, de forma democrática e gratuita, em tempo real, na página <https://portal.inmet.gov.br>, e têm aplicação em todos os setores da economia, de modo especial no agropecuário e em apoio à Defesa Civil.

Origem dos Dados

Os dados disponibilizados pelo INMET dão disponibilizados no site da instituição com as seguintes características:

- um arquivo para cada ano compactado (formato .zip);
- cada arquivo compactado possui um arquivo em formato .csv contendo as medições horárias de indicadores meteorológicos em cada uma das estações.

The screenshot shows the homepage of the Instituto Nacional de Meteorologia (INMET). At the top, there is a navigation bar with links for 'Ir para o conteúdo', 'Ir para o menu', 'Ir para o rodapé', 'SOBRE O INMET', 'IMPRENSA', 'MAPA DO SITE', 'FALE CONOSCO', and 'WEBMAIL'. Below the header, the INMET logo is displayed, followed by the text 'MINISTÉRIO DA AGRICULTURA E PECUÁRIA'. A main menu bar includes 'Tempo', 'Clima', 'Dados Meteorológicos', 'Satélites', 'Risco de Incêndio', 'Previsão Numérica', 'Sisdagro', 'Publicações', 'Sobre', and 'Informações'. The central content area features a section titled 'DADOS HISTÓRICOS ANUAIS' with a list of links for each year from 2000 to 2009, all labeled '(AUTOMÁTICA)'.

Ir para o conteúdo 1 Ir para o menu 2 Ir para o rodapé 3 SOBRE O INMET IMPRENSA MAPA DO SITE FALE CONOSCO WEBMAIL

Instituto Nacional de Meteorologia
MINISTÉRIO DA AGRICULTURA E PECUÁRIA

Tempo ▾ Clima ▾ Dados Meteorológicos ▾ Satélites Risco de Incêndio Previsão Numérica ▾ Sisdagro ▾ Publicações ▾ Sobre ▾ Informações ▾

DADOS HISTÓRICOS ANUAIS

ANO 2000 (AUTOMÁTICA)

ANO 2001 (AUTOMÁTICA)

ANO 2002 (AUTOMÁTICA)

ANO 2003 (AUTOMÁTICA)

ANO 2004 (AUTOMÁTICA)

ANO 2005 (AUTOMÁTICA)

ANO 2006 (AUTOMÁTICA)

ANO 2007 (AUTOMÁTICA)

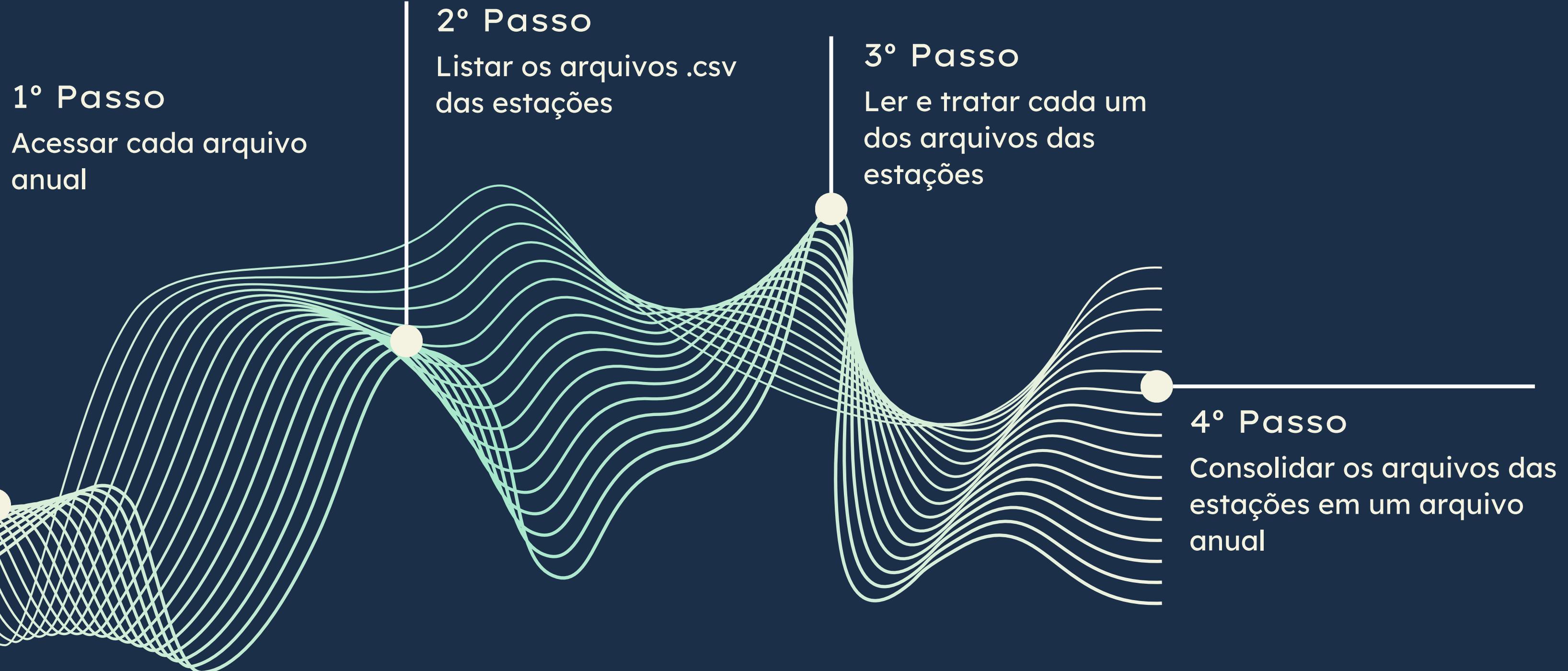
ANO 2008 (AUTOMÁTICA)

ANO 2009 (AUTOMÁTICA)

Link: <https://portal.inmet.gov.br/dadoshistoricos>

Processo de Extração dos Dados

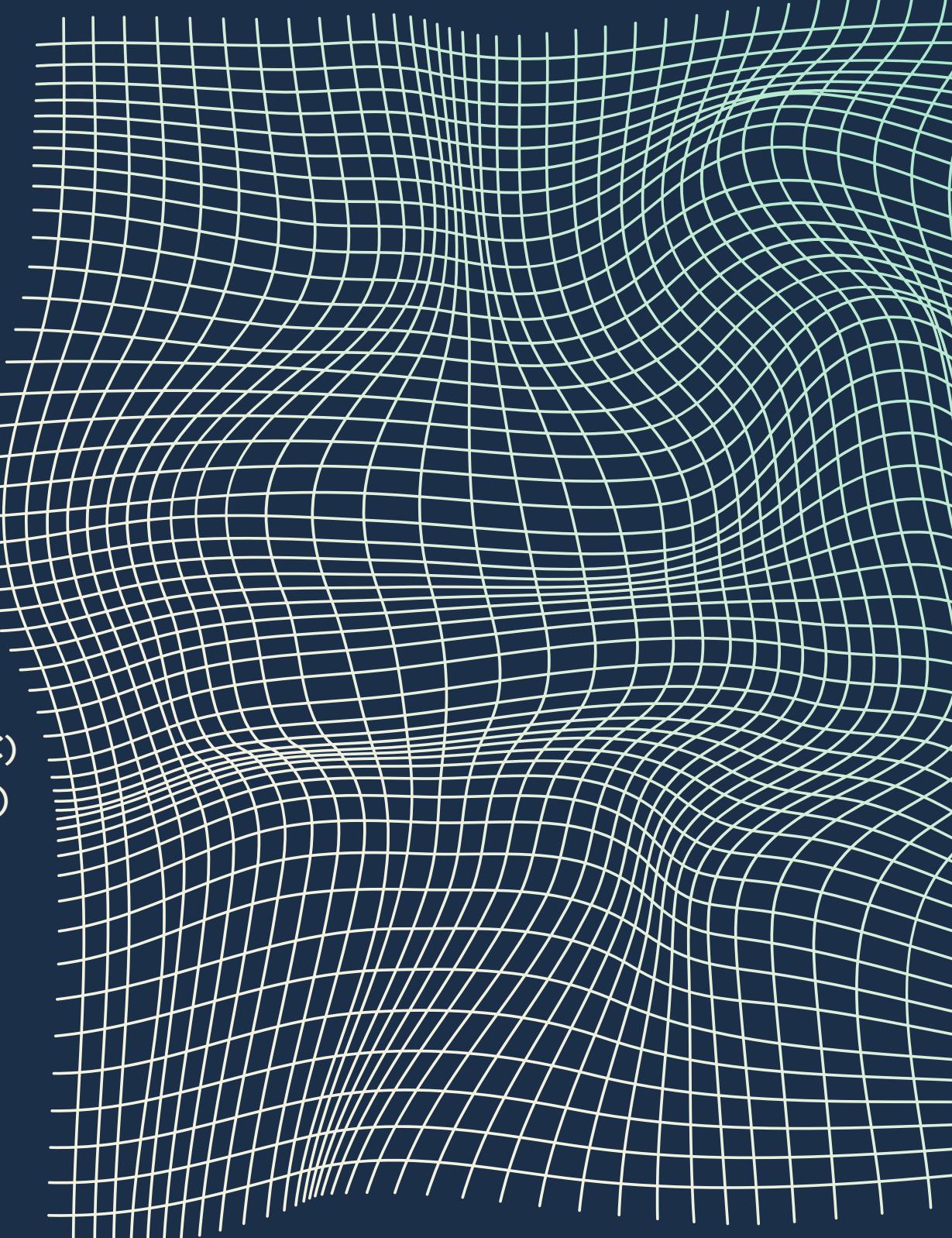
A extração dos dados foi feita a partir de um código em Python estruturado nos seguintes passos:



Dicionário de Dados

O arquivo de cada estação contem as seguintes variáveis:

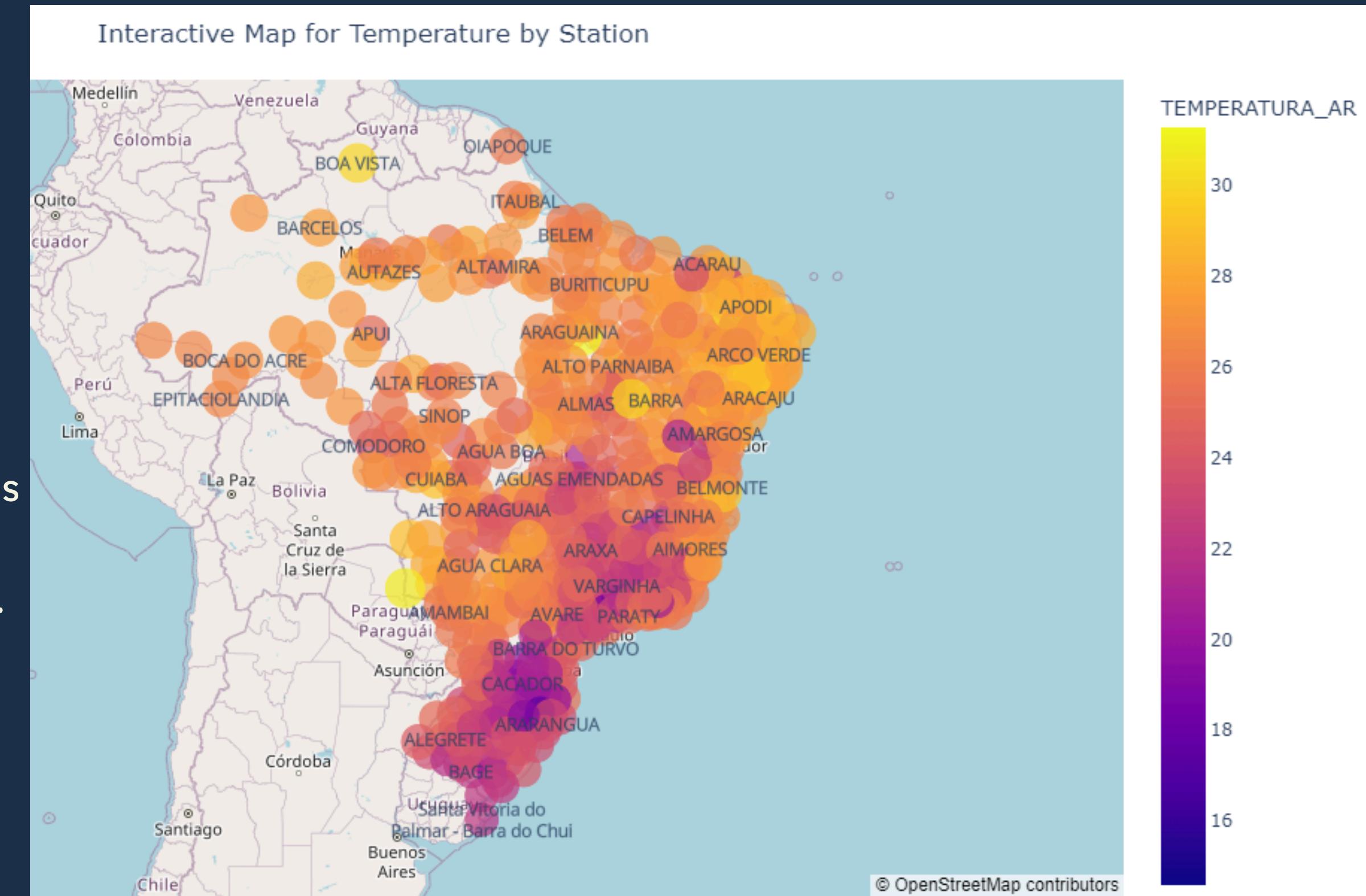
- DATA = DATA (YYYY-MM-DD)
- HORA = HORA (UTC)
- PRECIPITACAO_TOTAL = PRECIPITAÇÃO TOTAL, HORÁRIO (mm)
- PRESSAO_ATM = ATMOSFERICA AO NIVEL DA ESTACAO, HORARIA (mB)
- PRESSAO_ATM_MAX_HORA_ANT = PRESSÃO ATMOSFERICA MAX.NA HORA ANT. (AUT) (mB)
- PRESSAO_ATM_MIN_HORA_ANT = PRESSÃO ATMOSFERICA MIN. NA HORA ANT. (AUT) (mB)
- RADIACAO_GLOBAL = RADIACAO GLOBAL (KJ/m²)
- TEMPERATURA_AR = TEMPERATURA DO AR - BULBO SECO, HORARIA (°C)
- TEMPERATURA_AR_ORVALHO = TEMPERATURA DO PONTO DE ORVALHO (°C)
- TEMPERATURA_AR_MAX_HORA_ANT = TEMPERATURA MÁXIMA NA HORA ANT. (AUT) (°C)
- TEMPERATURA_AR_MIN_HORA_ANT = TEMPERATURA MÍNIMA NA HORA ANT. (AUT) (°C)
- TEMPERATURA_AR_ORVALHO_MAX_HORA_ANT = TEMPERATURA ORVALHO MAX. NA HORA ANT. (AUT) (°C)
- TEMPERATURA_AR_ORVALHO_MIN_HORA_ANT = TEMPERATURA ORVALHO MIN. NA HORA ANT. (AUT) (°C)
- UMIDADE_REL_MAX_HORA_ANT = UMIDADE REL. MAX. NA HORA ANT. (AUT) (%)
- UMIDADE_REL_MIN_HORA_ANT = UMIDADE REL. MIN. NA HORA ANT. (AUT) (%)
- UMIDADE_REL = UMIDADE RELATIVA DO AR, HORARIA (%)
- VENTO_DIRECAO_HORARIA = VENTO, DIREÇÃO HORARIA (gr) (° (gr))
- VENTO_RAJADA_MAXIMA = VENTO, RAJADA MAXIMA (m/s)
- VENTO_VELOCIDADE_HORARIA = VENTO, VELOCIDADE HORARIA (m/s)



Mapa das Estações do INMET

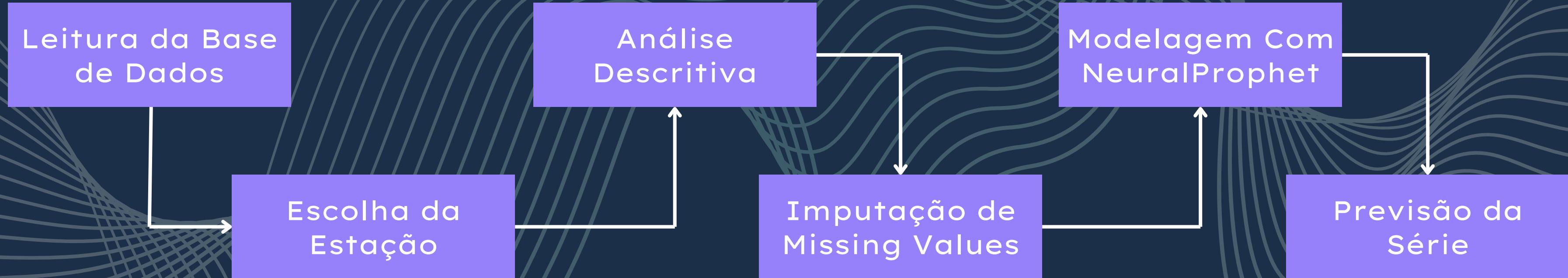
A figura ao lado mostra a média em 2024 da temperatura do ar em cada uma das 564 estações com dados disponíveis pelo INMET.

É possível perceber, como é de conhecimento de todos, que as maiores temperaturas são observadas no Norte/Nordeste do Brasil e as menores temperaturas no Sul do país.



Fluxo de Análise dos Dados

Como dito anteriormente, o objetivo final do trabalho será aplicar um modelo de previsão de séries temporais aos dados meteorológicos de uma das estações do INMET. Para tal, será preciso aplicar um método de imputação de missing values (MICE foi o método escolhido) e a biblioteca NeuralProphet para realizar as previsões.

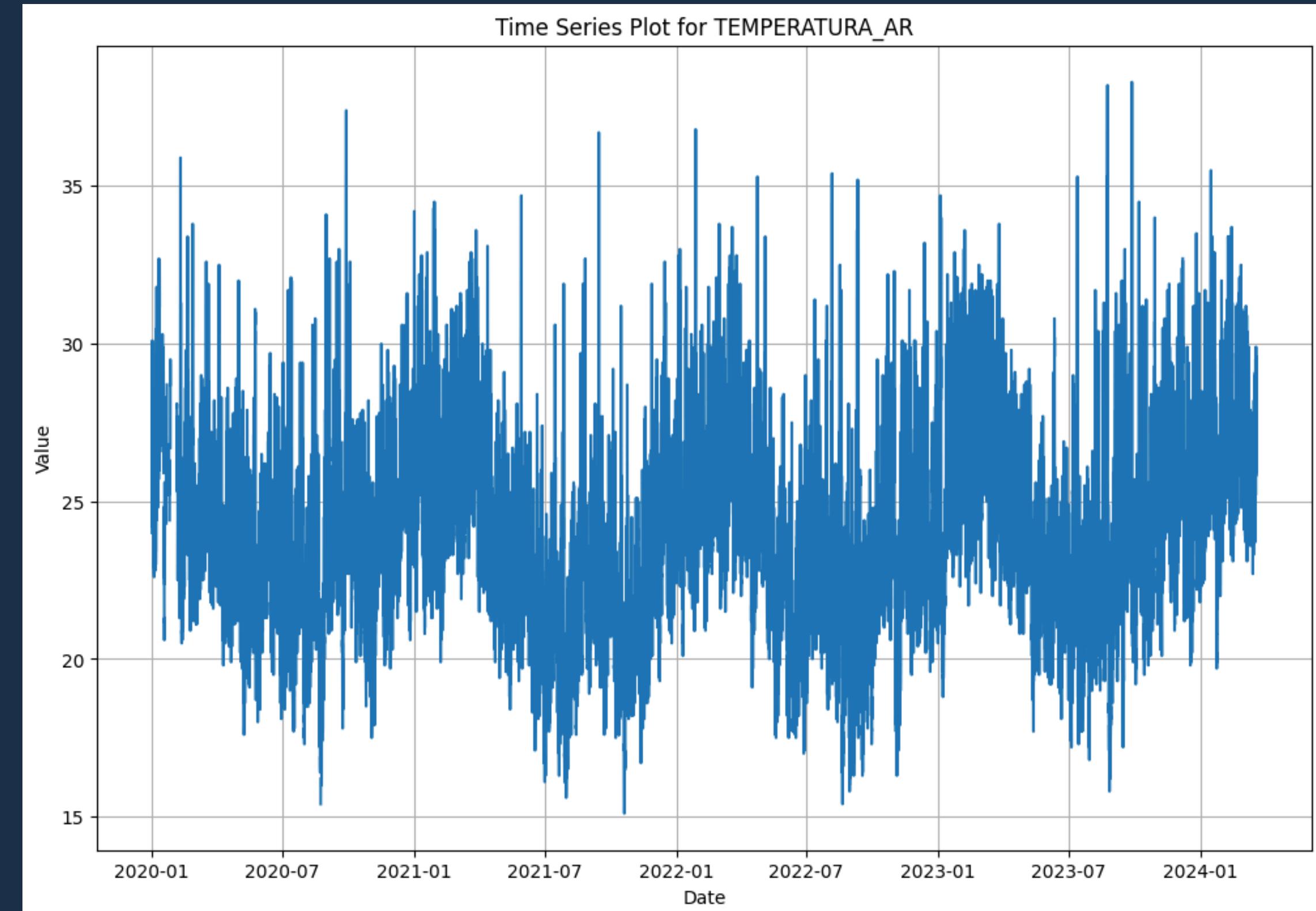


Análise Descritiva

A estação escolhida foi a ‘A652’ que fica localizada no Forte Copacabana na cidade do Rio de Janeiro.

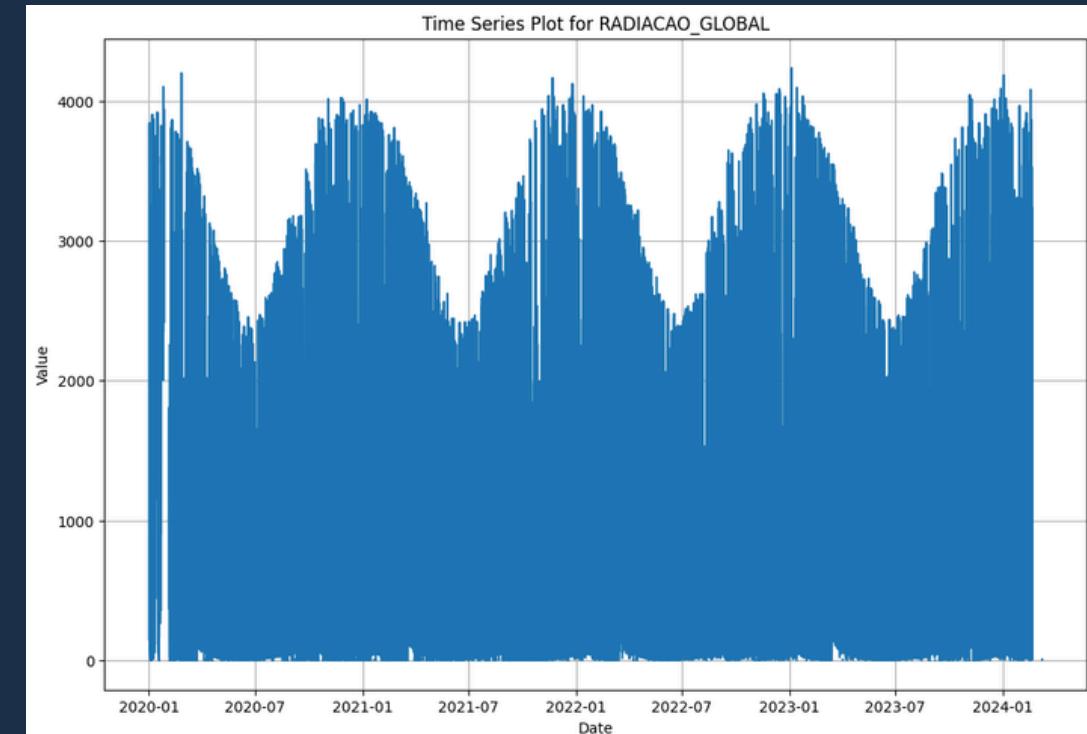
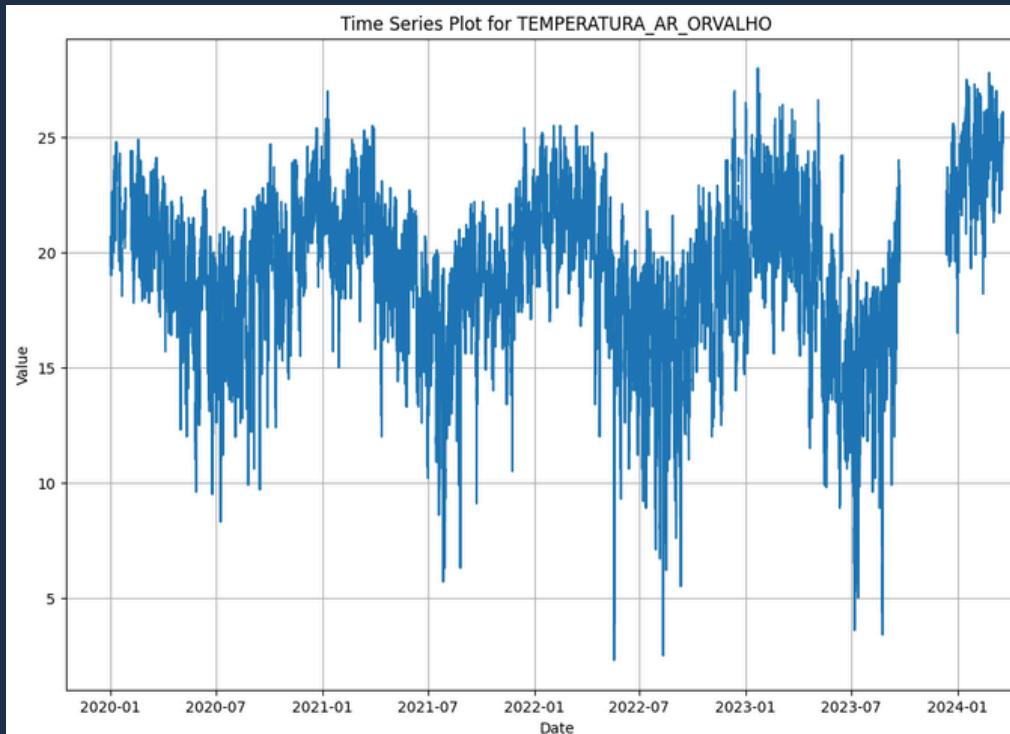
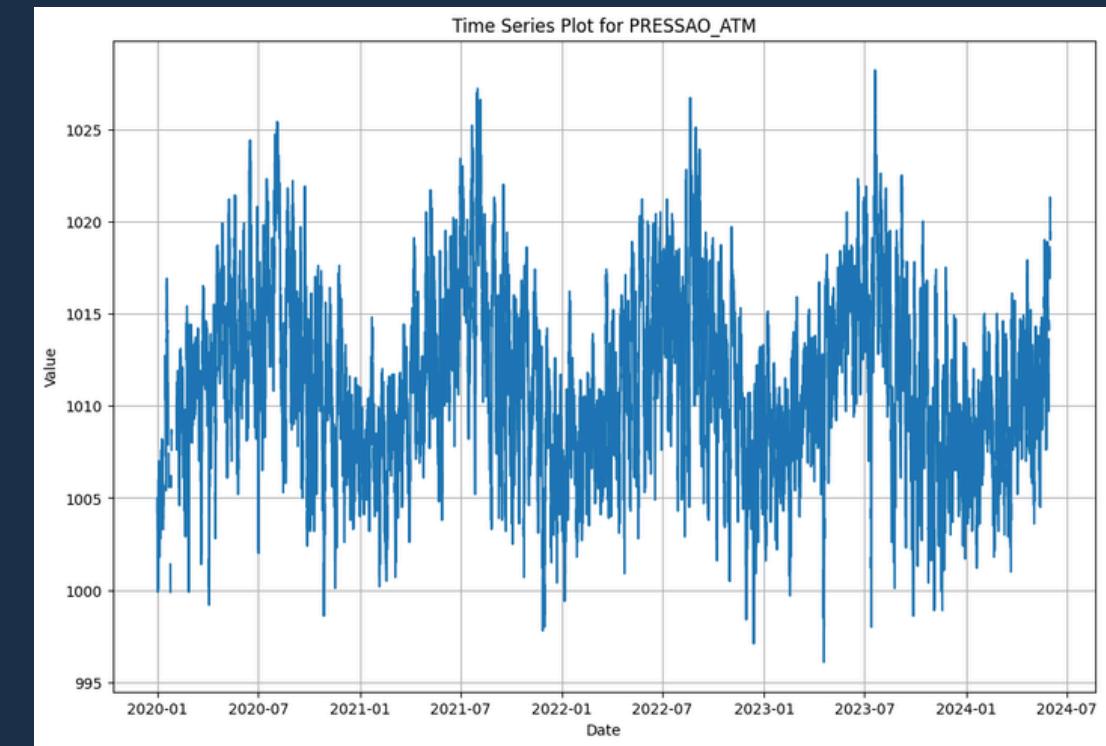
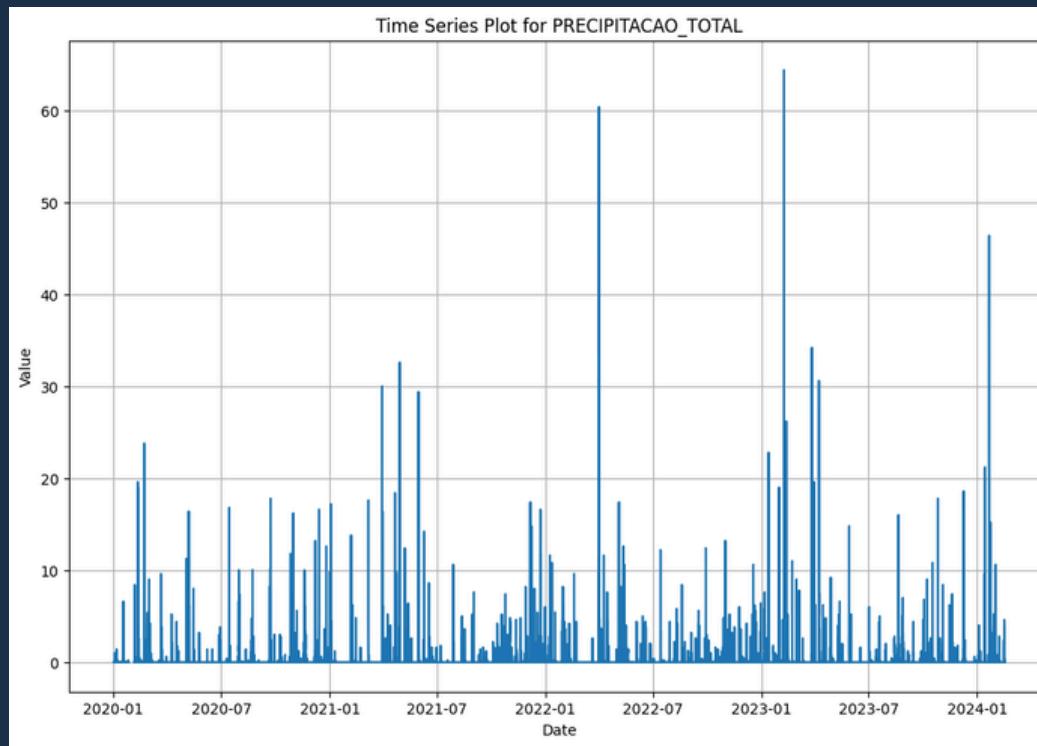
O período selecionado para a análise está compreendido entre Janeiro de 2020 até os dias de hoje.

A variável que temos o interesse de realizar previsões é a “TEMPERATURA_AR”.



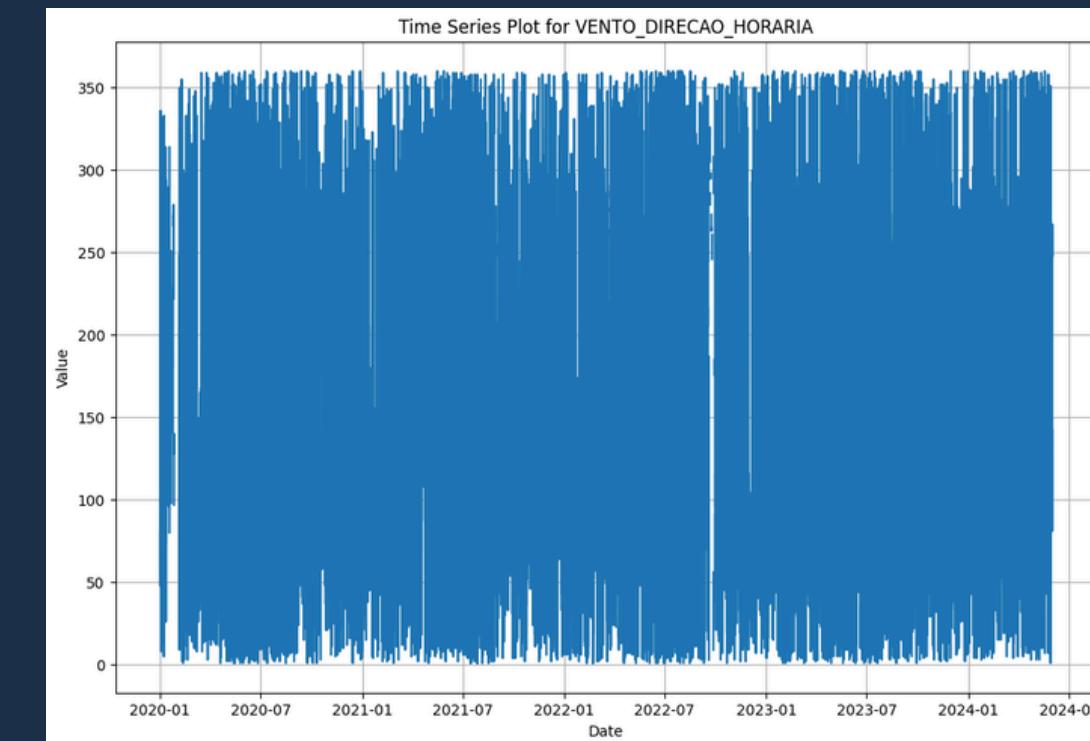
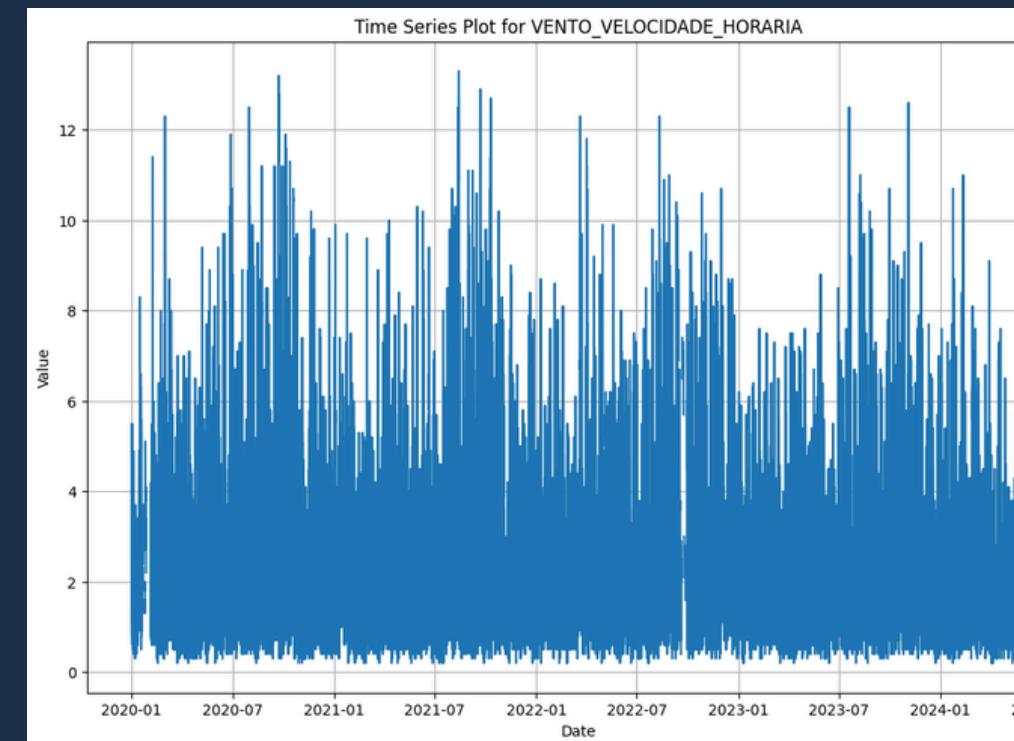
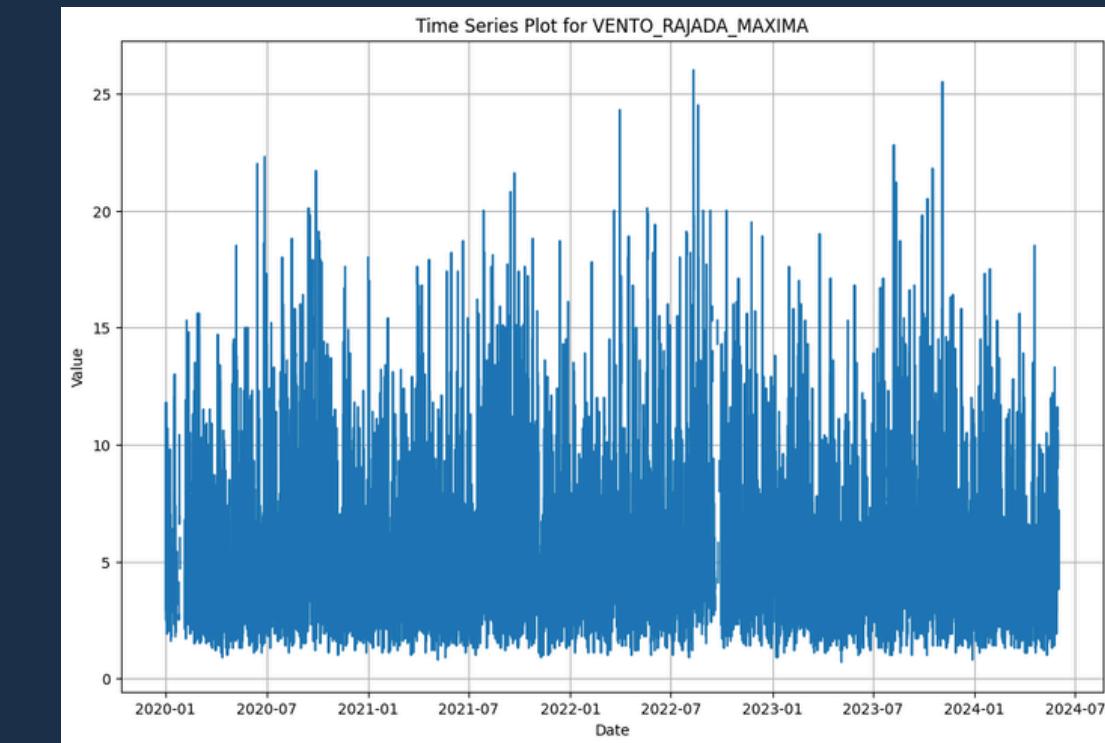
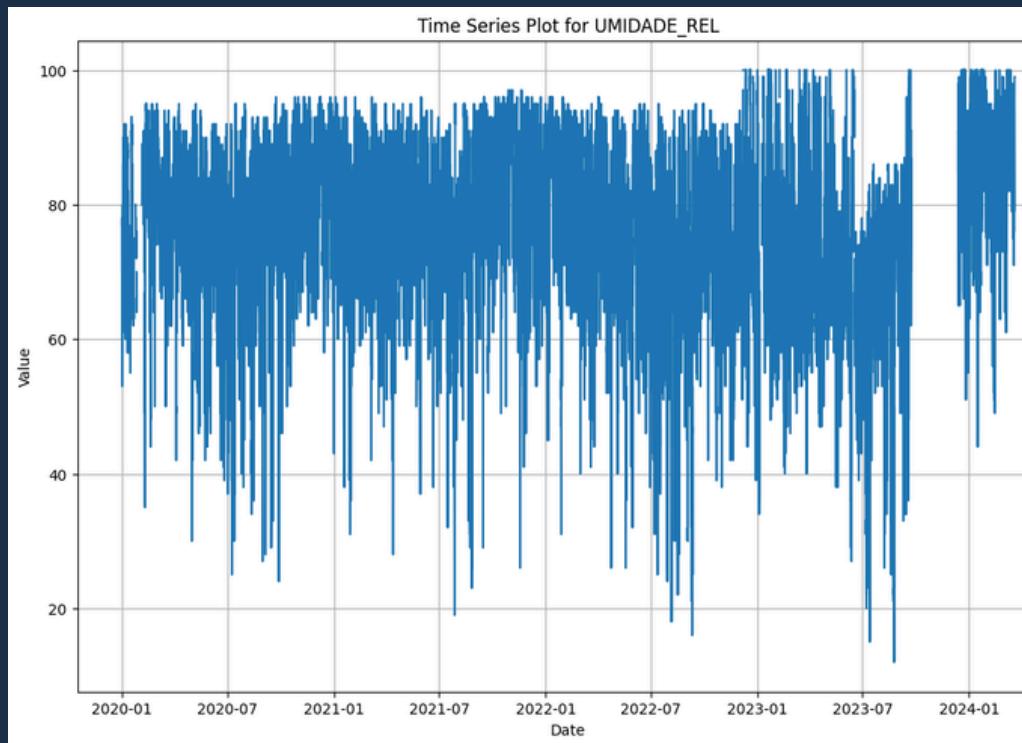
Análise Descritiva

A seguir, estão os gráficos das demais variáveis na base de dados



Análise Descritiva

A seguir, estão os gráficos das demais variáveis na base de dados

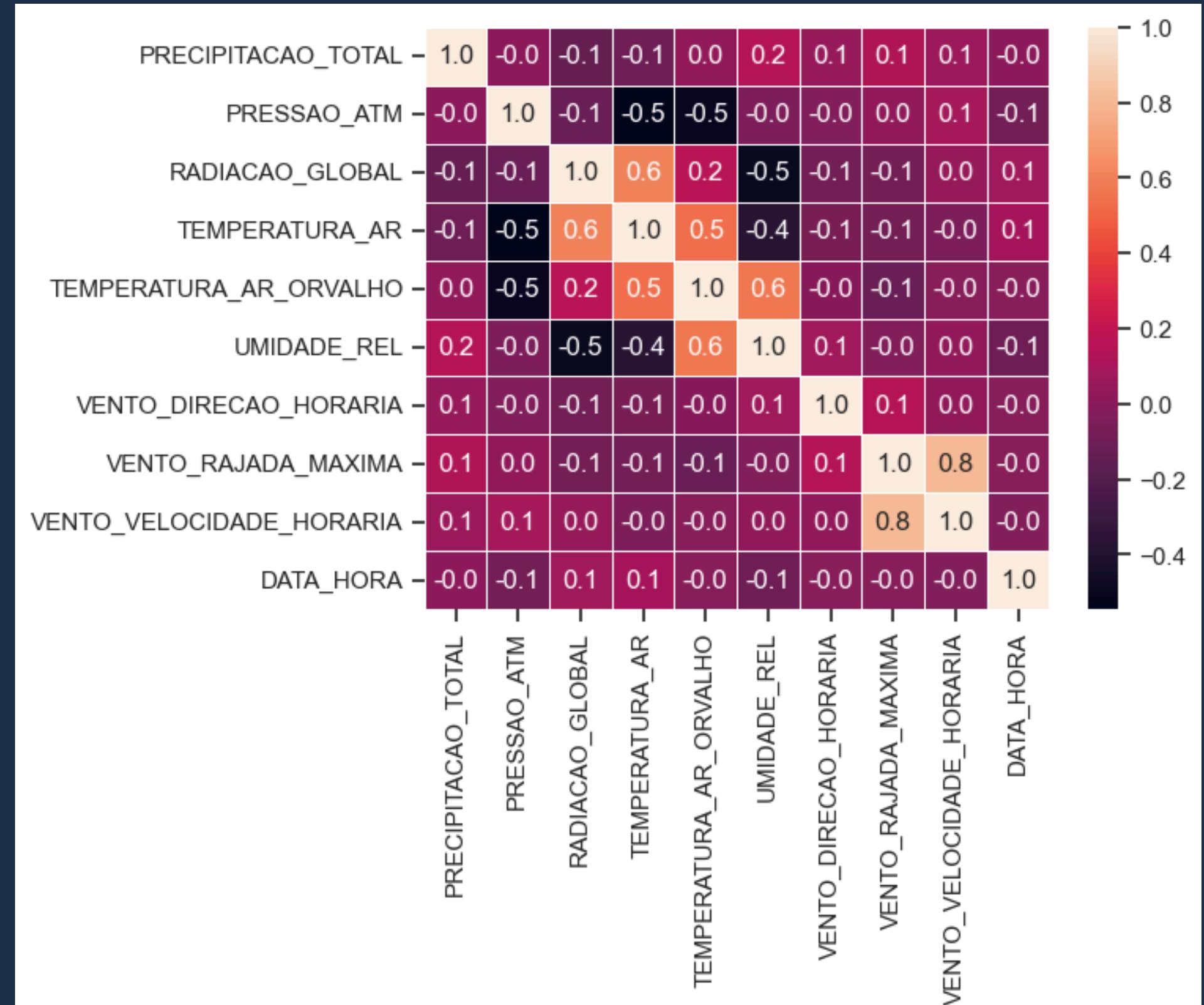


Análise Descritiva

MATRIZ DE CORRELAÇÕES

A variável “TEMPERATURA_AR” é mais correlacionada com as variáveis:

- pressão atmosférica;
- radiação global;
- temperatura de orvalho;
- umidade relativa do ar.

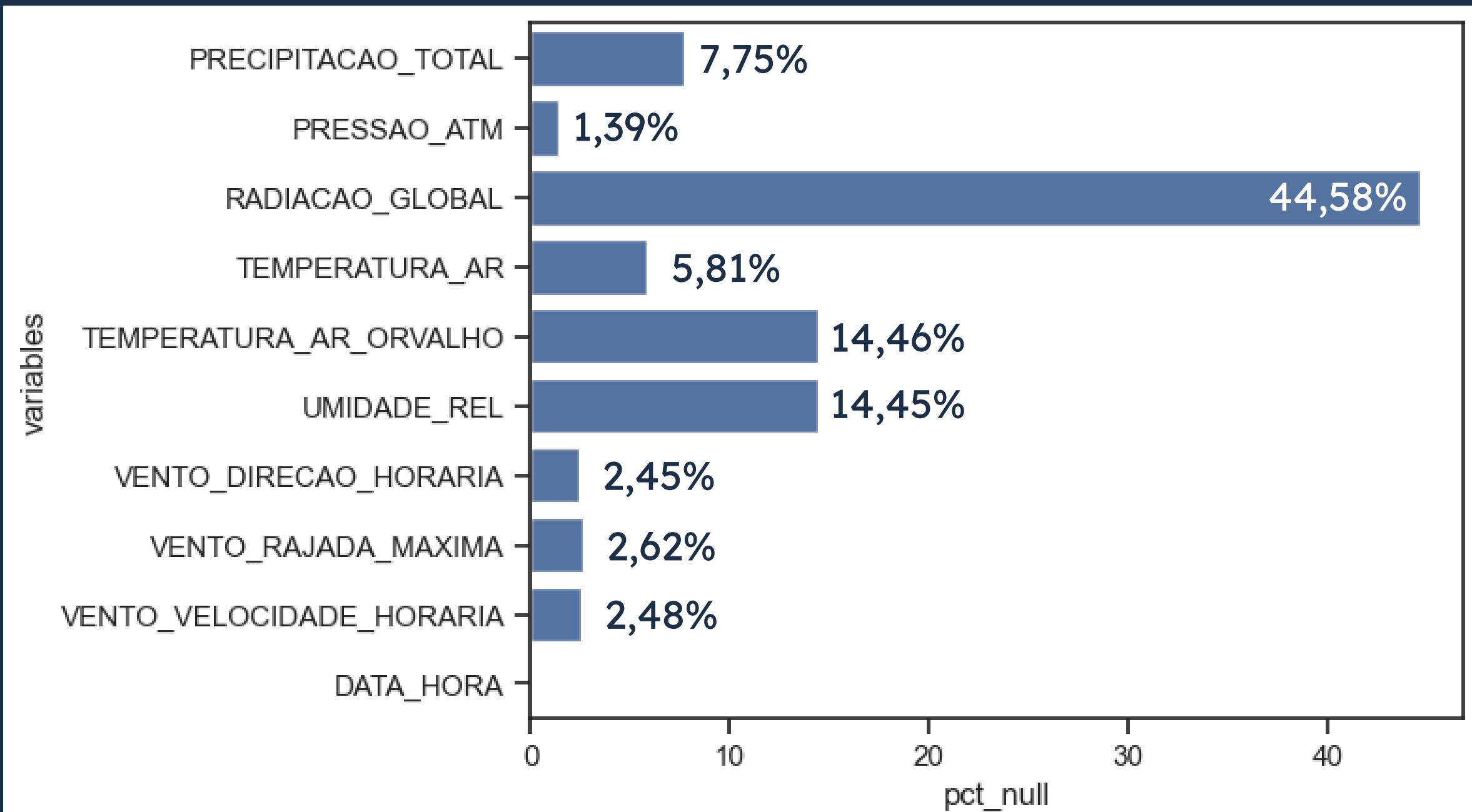


Análise Descritiva

PERCENTUAL DE MISSING VALUES

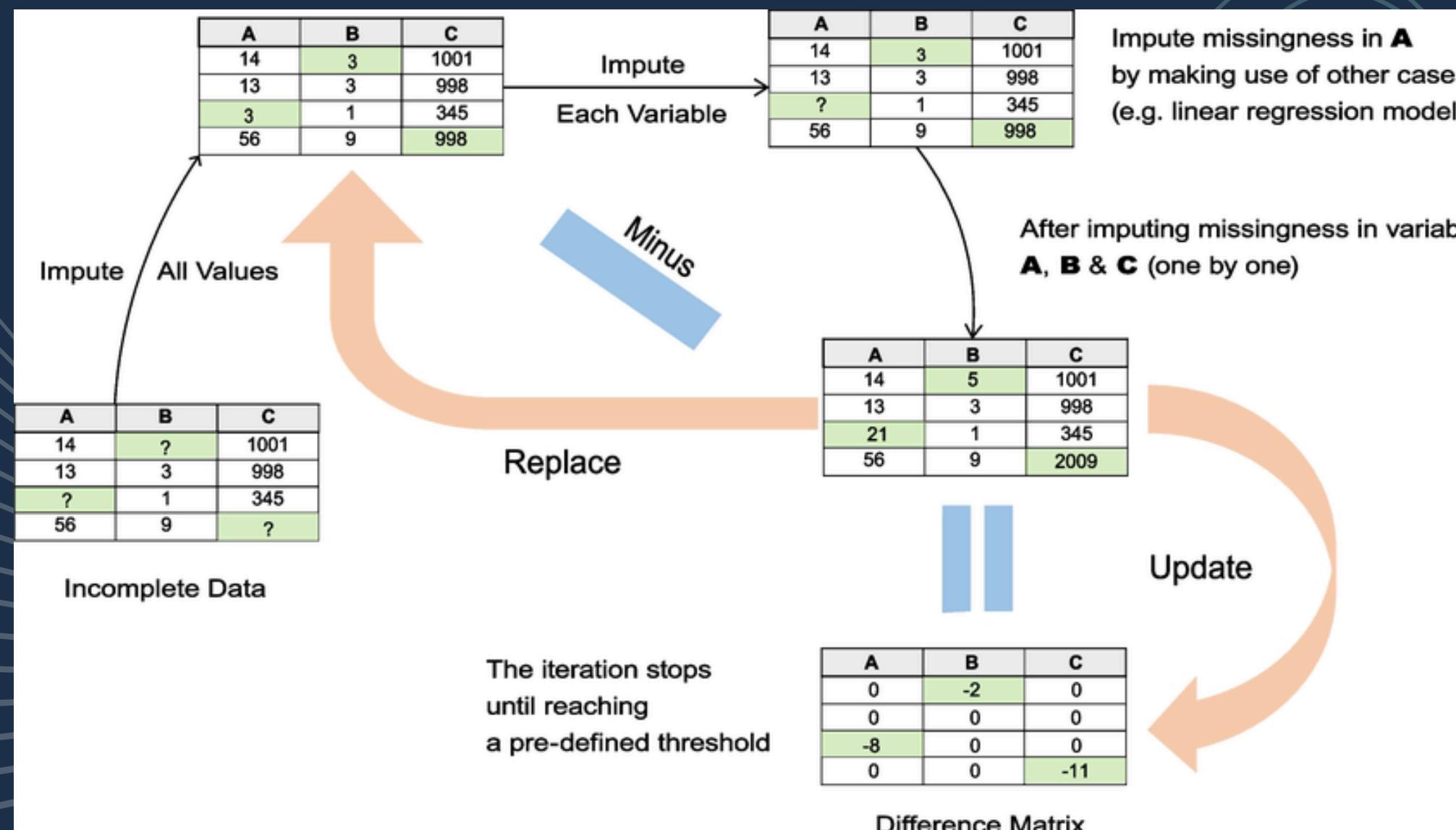
A nossa variável de interesse possui um percentual de missing de 5,81%.

Antes de partir para a modelagem de séries temporais, um método de imputação de dados será utilizado para imputar os valores faltantes tirando proveito da correlação entre as variáveis.



MICE: Multiple Imputation by Chained Equations

MICE (Imputação Múltipla por Equações Encadeadas) é uma técnica usada para lidar com dados ausentes criando múltiplos conjuntos de dados imputados e depois combinando os resultados. Em Python, o MICE pode ser implementado usando o método ‘IterativeImputer’ da biblioteca ‘sklearn’.



A figura ao lado ilustra como o método funciona. Ao final do processo, através de um número finito de iterações, é obtida uma base com os valores faltantes imputados.

A sua implementação em Python utiliza como padrão a regressão linear bayesiana, mas neste trabalho utilizaremos o modelo XGBoost como estimador.

NeuralProphet

Após realizar a imputação de dados faltantes, a biblioteca NeuralProphet foi utilizada para realizar a modelagem e previsão da série temporal da temperatura do ar na estação do Forte de Copacabana.

Essa biblioteca combina as técnicas tradicionais de modelagem de séries temporais com métodos de deep learning construídos em Pytorch com inspiração nas modelos Facebook Prophet e AR-Net.

NeuralProphet is more than the Neural evolution of Prophet. **Prophet**

PROPHET

Prophet has three major shortcomings:

1. Missing local context for predictions
2. Acceptable forecast accuracy
3. Framework is difficult to extend (Stan)

 Stan

→  PyTorch

Neural Prophet

NeuralProphet solves these:

1. Support for auto-regression and covariates.
2. Hybrid model (linear <> Neural Network)
3. Python package based on PyTorch using standard deep learning methods.

Modelagem da Série Temporal

A série temporal de temperatura foi dividida em uma base de treino (2020 até 2023) e uma base de validação (2024). A partir daí, foram estimados 4 modelos e, dentre esses, o modelo que apresentou melhor performance foi o 4.

Modelo	MAE Valid	RMSE Valid	MAE Train	RMSE Train
1 - Linear Trend + Seasonality	1.720353	2.182399	1.614460	2.159444
2 - Picewise Linear Trend + Seasonality	2.272473	2.714903	1.539790	2.077862
3 - Picewise Linear Trend + Seasonality + AR(24)	0.600653	0.842865	0.532965	0.802442
4 - Picewise Linear Trend + Seasonality + AR-Net(24)	0.586868	0.821414	0.525711	0.785201

Detalhamento do Modelo 4

Picewise Linear Trend + Seasonality + AR-Net(24)

Os componente do modelos revelam que existe uma tendência de aumento da temperatura e que existe uma grande persistência da temperatura (grande relevância da 1^a defasagem) e que existe uma relação não-linear entre a temperatura e seus valores passados.

Python Code

```
# Model and prediction
m = NeuralProphet(
    # Trend changepoints
    n_changepoints = 10,
    # Seasonality components
    yearly_seasonality = True,
    weekly_seasonality = False,
    daily_seasonality = True,
    # Autoregressive Component
    n_lags = 24,
    ar_layers = [12, 12, 12, 12]
)
```

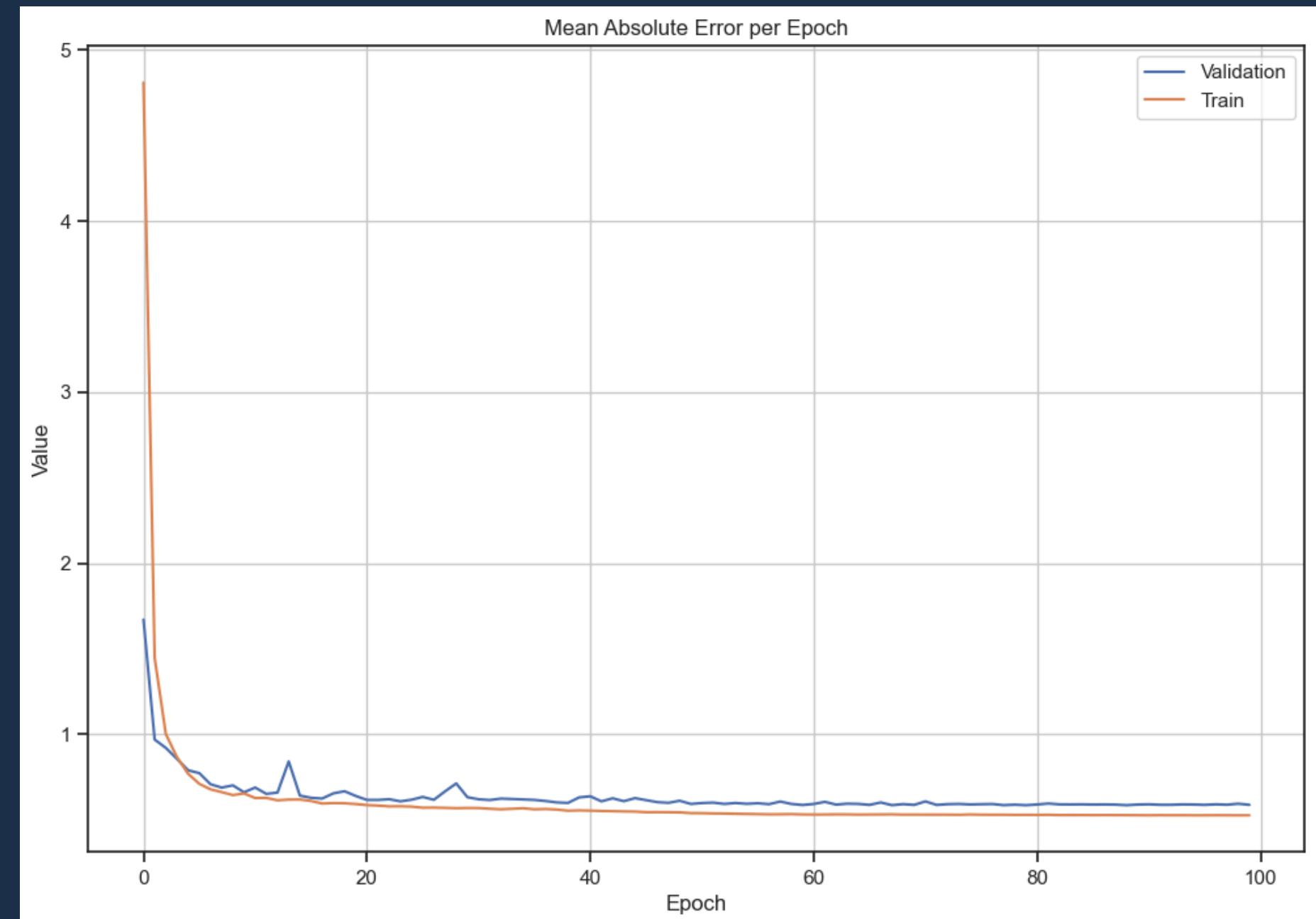
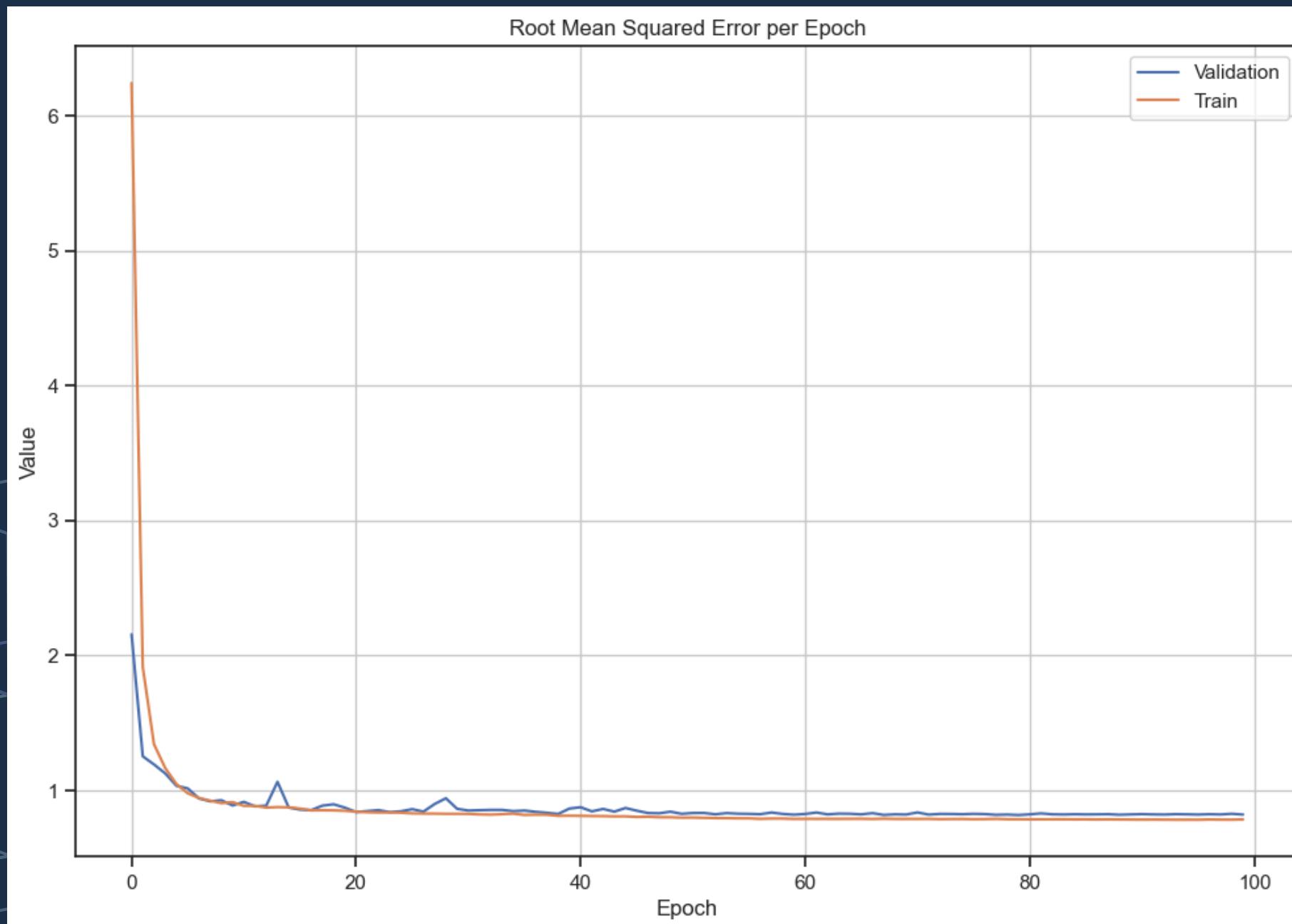
Componentes do Modelo



Detalhamento do Modelo 4

Picewise Linear Trend + Seasonality + AR-Net(24)

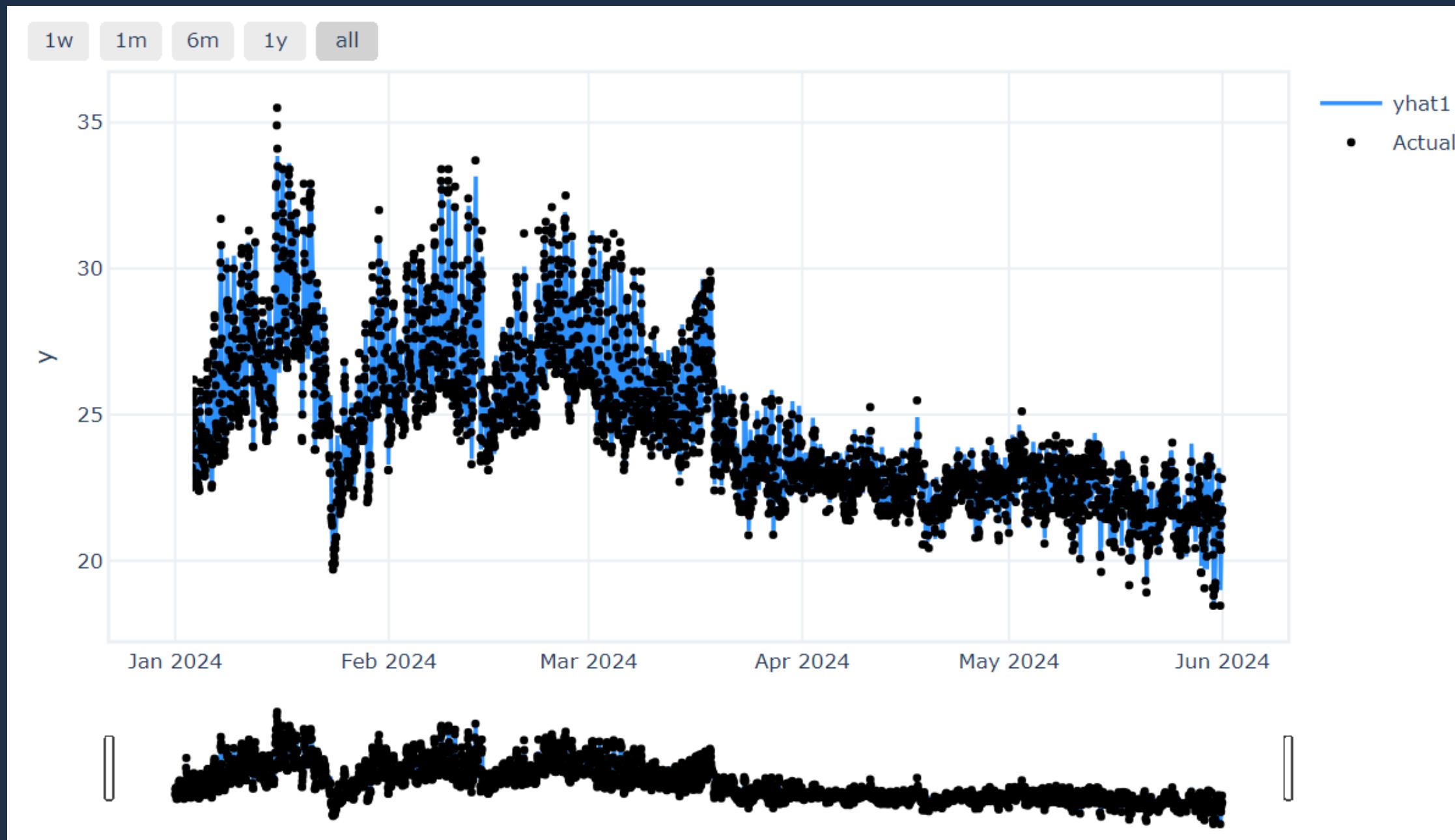
Acompanhando o comportamento das métricas de avaliação dos modelos ao longo das épocas de treinamento, é possível perceber que o modelo generaliza bem o comportamento da temperatura.



Detalhamento do Modelo 4

Previsão da Temperatura do Ar na base de Validação

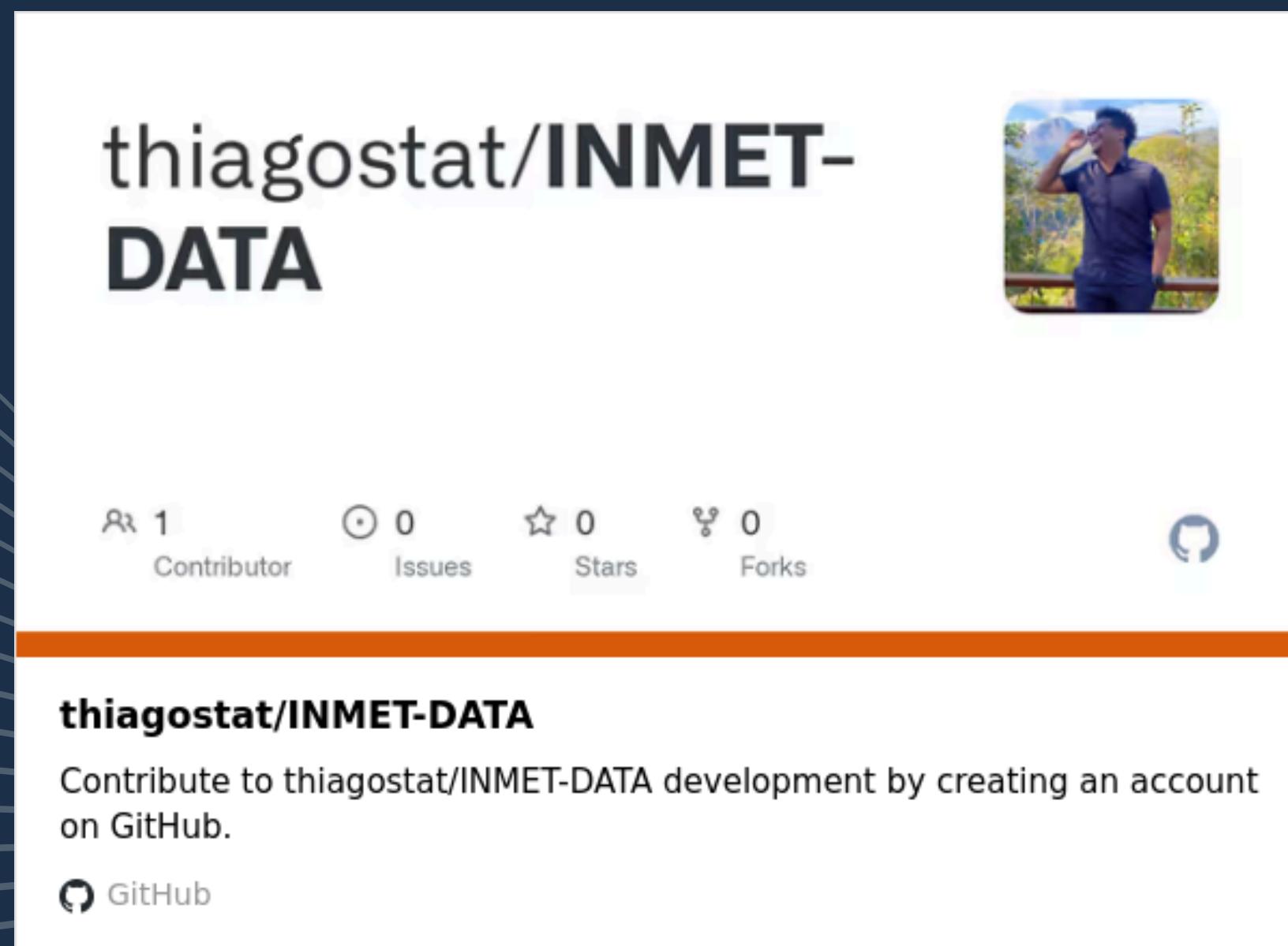
Acompanhando o comportamento das métricas de avaliação dos modelos ao longo das épocas de treinamento, é possível perceber que o modelo generaliza bem o comportamento da temperatura.



Obrigado!!!

Anexo

Todo o código pode ser encontrado no meu GitHub



Ordem de Processamento

inmet_data_functions.py



data_download_quality.ipynb



data_timeseries_analysis.ipynb

