



**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS**  
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA  
Pós-graduação *Lato Sensu* em Analytics e Business Intelligence

## **RELATÓRIO TÉCNICO**

### **PAINEL METEOROLÓGICO**

Leonardo Nascimento Yagelovic

Belo Horizonte

2024

# SUMÁRIO

## Sumário

<b>1. Introdução</b>	3
<b>1.1. Contexto</b>	3
<b>1.2. Objetivos</b>	3
<b>1.3. Público alvo</b>	3
<b>2. Modelo de Dados</b>	3
2.1. Modelo Dimensional	3
<b>Tabelas de Dimensões</b>	4
<b>Tabelas de Fato</b>	4
<b>Tabelas de Fato e Dimensões dos Arquivos</b>	4
<b>3. Integração, Tratamento e Carga de Dados</b>	7
3.1. Fontes de Dados	7
3.2. Processos de Integração e Carga (ETL)	7
<b>4. Camada de Apresentação</b>	9
4.1 Dashboard	9
4.2 Análises avançadas	11
<b>5. Registros de Homologação</b>	20
<b>6. Conclusões</b>	23
<b>7. Links</b>	24
<b>8. REFERÊNCIAS</b>	25

# **1. Introdução**

## **1.1. Contexto**

Muito se é dito sobre o Aquecimento Global e suas consequências. Recorde de temperaturas, aumento no volume de chuvas, inundações, secas, frio extremo. Entretanto as informações estão espalhadas e geralmente em forma de relatórios síntese como os fornecidos pelo IPCC (Painel Intergovernamental sobre Mudança do Clima). Nesse contexto, surge a necessidade de um painel que possa fornecer informações detalhadas sobre temperatura e precipitações das estações meteorológicas do Brasil, com gráficos e linguagem acessível a qualquer público, e de livre acesso na internet.

## **1.2. Objetivos**

O principal objetivo do projeto é democratizar a informação, tornando o acesso a dados mais fácil e intuitivo a todos, independente de conhecimento técnico prévio. Além disso tentar elevar a consciência pública sobre as mudanças climáticas, quantificando as variações através do tempo e auxiliando na tomada de medidas para se adaptar a essa nova realidade.

Outro objetivo do painel será ajudar a entender o relacionamento entre temperatura e precipitação com fenômenos como El Niño e La Niña.

## **1.3. Público alvo**

O painel é destinado a uma ampla gama de usuários, desde indivíduos que desejam informações sobre temperatura e precipitação para planejar suas viagens, até pesquisadores e cientistas que estão estudando as mudanças climáticas. O painel é projetado de forma a ser fácil de usar, mesmo para usuários sem conhecimento técnico prévio. Além disso, o painel também pode ser útil para organizações governamentais e não governamentais envolvidas na gestão de desastres e na formulação de políticas climáticas.

# **2. Modelo de Dados**

## **2.1. Modelo Dimensional**

Um modelo dimensional é uma estrutura para organizar dados de forma eficiente, facilitando a análise e compreensão de grandes volumes de informações. Ele separa os dados brutos (fatos) dos seus respectivos contextos (dimensões), permitindo consultas mais rápidas e flexíveis.

## **Referências:**

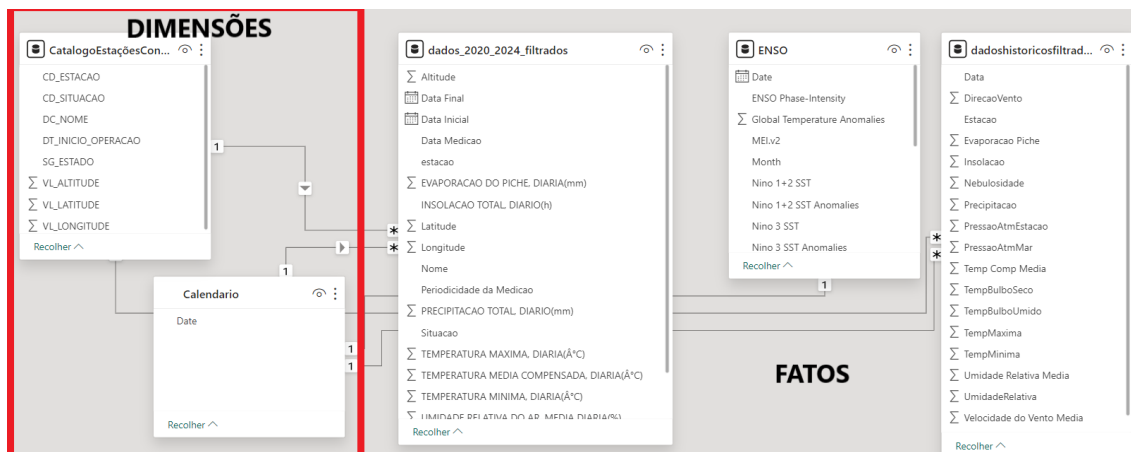


Figura 1 – Tabelas Dimensões e Fatos

## Tabelas de Dimensões

O modelo dimensional inclui as seguintes dimensões:

- **Dimensao\_Estacao:** Armazena informações detalhadas sobre as estações meteorológicas, como nome, estado, localização, altitude e data de início da operação.
- **Dimensao\_Calendario:** Fornece contexto temporal para as medições diárias das tabelas fato.

## Tabelas de Fato

O modelo dimensional inclui as seguintes tabelas de fato:

- **Fato\_Medicao\_Diaria:** Armazena medições meteorológicas diárias, como precipitação, temperatura média compensada, umidade relativa média e velocidade média do vento.
- **Fato\_Medicao\_Completa:** Armazena medições meteorológicas completas, incluindo hora da medição, temperatura do bulbo seco, temperatura do bulbo úmido, pressão atmosférica e outros parâmetros.
- **ENSO:** Armazena dados do Índice El Niño-Oscilação Sul (ENSO), como temperatura da superfície do mar, anomalias de temperatura e índices climáticos.

## Tabelas de Fato e Dimensões dos Arquivos

### 1. Tabelas de Fato:

- **Fato\_Medicao\_Diaria\_INMET:**
  - **Descrição:** Armazena medições meteorológicas diárias coletadas pelas estações INMET.
  - **Atributos:**
    - **CD\_MEDICAO:** Chave primária (código único para cada medição).

- CD\_ESTACAO: Chave estrangeira (referencia a tabela Dimensao\_Estacao).
  - DT\_MEDICAO: Data da medição (no formato YYYY-MM-DD).
  - PRECIPITACAO: Precipitação total (mm).
  - TEMP\_MAXIMA: Temperatura máxima (°C).
  - TEMP\_MINIMA: Temperatura mínima (°C).
  - TEMP\_COMP\_MEDIA: Temperatura média compensada (°C).
  - UMIDADE\_RELATIVA\_MEDIA: Umidade relativa média (%).
  - VELOCIDADE\_VENTO\_MEDIA: Velocidade do vento média (m/s).
  - (Outros atributos diários relevantes podem ser adicionados, como: INSOLACAO, NEBULOSIDADE, EVAPORACAO\_PICHE, etc.).
- **Fato\_Medicao\_HIST:**
  - **Descrição:** Armazena as informações completas das medições, incluindo hora da medição e outros parâmetros menos frequentes.
  - **Atributos:**
    - CD\_MEDICAO: Chave primária (código único para cada medição).
    - CD\_ESTACAO: Chave estrangeira (referencia a tabela Dimensao\_Estacao).
    - DT\_MEDICAO: Data da medição (no formato YYYY-MM-DD).
    - HR\_MEDICAO: Hora da medição (no formato HH:MM).
    - TEMP\_BULBO\_SECO: Temperatura do bulbo seco (°C).
    - TEMP\_BULBO\_UMIDO: Temperatura do bulbo úmido (°C).
    - PRESSAO\_ATM\_ESTACAO: Pressão atmosférica na estação (hPa).
    - PRESSAO\_ATM\_MAR: Pressão atmosférica ao nível do mar (hPa).
    - DIRECAO\_VENTO: Direção do vento (graus).
    - VELOCIDADE\_VENTO: Velocidade do vento (m/s).
    - INSOLACAO: Insolação total (horas).
    - NEBULOSIDADE: Nebulosidade (%).
    - EVAPORACAO\_PICHE: Evaporação do piche (mm).
    - (Outros parâmetros menos frequentes podem ser adicionados).
- **Fato\_ENSO:**
  - **Descrição:** Armazena dados do Índice El Niño-Oscilação Sul (ENSO).
  - **Atributos:**
    - DATA: Chave primária (data do índice, no formato YYYY-MM-DD).
    - ANO: Ano do índice.
    - MES: Mês do índice.
    - TEMPERATURA\_ANOMALIA\_GLOBAL: Anomalia da temperatura global (°C).
    - SST\_NINO\_1\_2: Temperatura da superfície do mar em Niño 1+2 (°C).
    - SST\_ANOMALIA\_NINO\_1\_2: Anomalia da temperatura da superfície do mar em Niño 1+2 (°C).
    - SST\_NINO\_3: Temperatura da superfície do mar em Niño 3 (°C).
    - SST\_ANOMALIA\_NINO\_3: Anomalia da temperatura da superfície do mar em Niño 3 (°C).

- SST\_NINO\_3\_4: Temperatura da superfície do mar em Niño 3.4 (°C).
- SST\_ANOMALIA\_NINO\_3\_4: Anomalia da temperatura da superfície do mar em Niño 3.4 (°C).
- SST\_NINO\_4: Temperatura da superfície do mar em Niño 4 (°C).
- SST\_ANOMALIA\_NINO\_4: Anomalia da temperatura da superfície do mar em Niño 4 (°C).
- TNI: Índice TNI (sem unidade).
- PNA: Índice PNA (sem unidade).
- OLR: OLR (sem unidade).
- SOI: Índice SOI (sem unidade).
- ESTACAO\_2\_MESES: Estação do ano (baseado em média de 2 meses).
- MEI: Índice MEI (sem unidade).
- ESTACAO\_3\_MESES: Estação do ano (baseado em média de 3 meses).
- ONI: Índice ONI (sem unidade).
- ESTACAO\_ANUAL: Estação do ano (anual).

## 2. Dimensões:

- **Dimensao\_Estacao:**

- **Descrição:** Armazena informações detalhadas sobre as estações meteorológicas.

- **Atributos:**

- CD\_ESTACAO: Chave primária (código único para cada estação).
    - DC\_NOME: Nome da estação.
    - SG\_ESTADO: Estado da estação.
    - CD\_SITUACAO: Situação da estação (operante, inativa, etc.).
    - VL\_LATITUDE: Latitude da estação (em graus decimais).
    - VL\_LONGITUDE: Longitude da estação (em graus decimais).
    - VL\_ALTITUDE: Altitude da estação (em metros).
    - DT\_INICIO\_OPERACAO: Data de início da operação da estação (no formato YYYY-MM-DD).

- **Dimensao\_Tempo\_Diaria:**

- **Descrição:** Fornece contexto temporal para as medições diárias.

- **Atributos:**

- DT\_MEDICAO: Chave primária (data da medição, no formato YYYY-MM-DD, referenciada pelas tabelas de fato).
    - ANO: Ano da medição.
    - MES: Mês da medição.
    - DIA: Dia do mês da medição.
    - TRIMESTRE: Trimestre do ano (1º, 2º, 3º ou 4º).
    - SEMESTRE: Semestre do ano (1º ou 2º).

- ESTACAO: Estação do ano (primavera, verão, outono ou inverno).

### 3. Integração, Tratamento e Carga de Dados

#### 3.1. Fontes de Dados

- **Bases históricas (1961-2019):**

A base `conventional_weather_stations_inmet_brazil_1961_2019.csv` contém dados históricos de estações meteorológicas do Brasil, incluindo temperatura, precipitação e umidade relativa.

- **Diretório de Dados:**

Os arquivos `dados_XXXXXX_D_2020-01-01_2024-03-31.csv` incluem dados diários de evaporação, insolação, precipitação e temperaturas máxima, média e mínima.

- **Catálogo de Estações:**

O `CatalogoEstaçõesConvencionais.csv` lista as estações meteorológicas operantes, com informações como localização e data de início de operação.

- **ENSO:**

O arquivo `ENSO.csv` fornece dados sobre o fenômeno El Niño-Oscilação Sul, incluindo anomalias de temperatura global e índices relacionados ao ENSO.

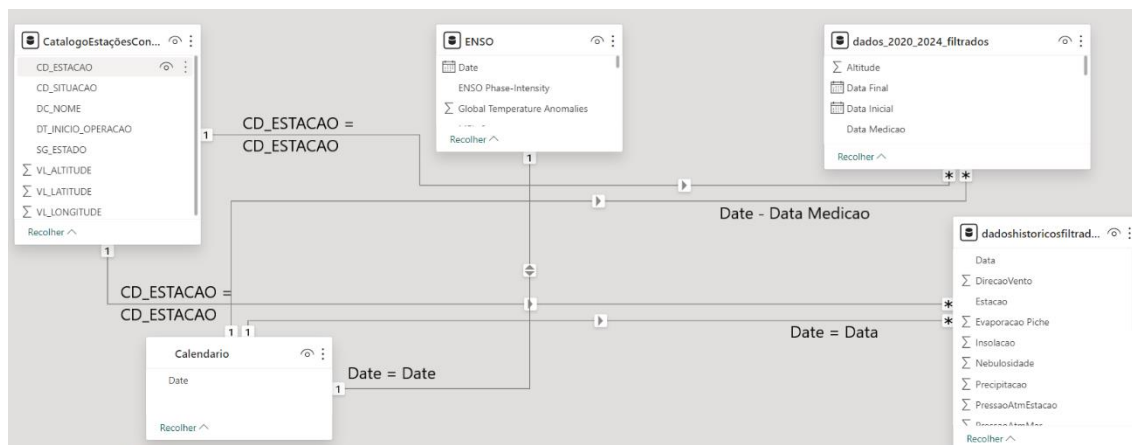


Figura 2 – Relacionamento entre as tabelas

#### 3.2. Processos de Integração e Carga (ETL)

O código apresentado implementa um processo ETL (Extração, Transformação e Carga) para dados meteorológicos provenientes de arquivos CSV. O objetivo principal é consolidar dados históricos (1961-2019) com dados recentes (2020-2023), além de aplicar transformações para garantir a qualidade e consistência dos dados.

Foi um processo de bastante aprendizado, uma vez que foi necessário união de diversos arquivos .CSV de várias estações em um único arquivo. Equalização de

separadores e decimais para evitar problemas nos próximos passos e ajustes nos tipos de dados. O que está exposto abaixo foi refeito várias vezes após erros no processo de Machine Learning e utilização no dashboard.

### **1. Extração e Combinação de Dados:**

- A função `read_csv_file(file_path)` extrai dados de um arquivo CSV, interpretando os metadados das primeiras linhas e definindo os nomes das colunas.
- A função `combine_csv_files(folder_path, output_file)` itera sobre múltiplos arquivos CSV em um diretório, combinando-os em um único arquivo de saída (`dados_2020_2024.csv`).

### **2. Limpeza e Padronização de Dados:**

- A função `replace_comma(x)` substitui vírgulas por pontos em valores numéricos, garantindo a compatibilidade com o formato numérico padrão.
- Valores faltantes (`nan`, `null`, `"`) são substituídos por `'0'` para evitar erros em cálculos posteriores.
- A coluna `'Codigo Estacao'` é renomeada para `'estacao'` para manter a consistência na nomenclatura.

### **3. Filtragem e Unificação de Dados:**

- Os DataFrames `dfhist` (dados históricos) e `dfnovo` (dados recentes) são filtrados para incluir apenas as estações presentes em ambos os conjuntos de dados.
- Um terceiro DataFrame, `dfcatalogo`, contendo um catálogo de estações, é utilizado para filtrar as estações relevantes, gerando `dfcatalogo_filtrado`.

### **4. Transformação e Agregação de Dados:**

- A coluna `'Data'` em `dfnovo` é convertida para o tipo `datetime` e formatada como `'dia/mês/ano'`.
- Dados históricos em `dfhist` são agrupados por `'Estacao'` e `'Data'`, calculando estatísticas como soma, máximo, mínimo e média para as variáveis meteorológicas.

### **5. Junção e Geração do Resultado Final:**

- As colunas comuns entre `dfhist_filtrado` e `dfnovo_filtrado` são identificadas e utilizadas para realizar uma junção dos conjuntos de dados (`outer join`), resultando em `df`.
- Valores faltantes em `df` são preenchidos com 0.
- O DataFrame final, contendo dados históricos e recentes combinados e processados, é salvo em um novo arquivo CSV (`dados_finais.csv`).



## Fluxograma do Processo ETL:

```
[CSV Input Files] --> [read_csv_file] --> [combine_csv_files] -->
[dados_2020_2024.csv]
[dados_2020_2024.csv] --> [pd.read_csv] --> [dfnovo]
[conventional_weather_stations_inmet_brazil_1961_2019.csv] -->
[pd.read_csv] --> [dfhist]
[CatalogoEstaçõesConvencionais.csv] --> [pd.read_csv] --> [dfcatalogo]
[dfhist, dfnovo, dfcatalogo] --> [Filtragem por Estações Comuns] -->
[dfhist_filtrado, dfnovo_filtrado, dfcatalogo_filtrado]
[dfnovo_filtrado] --> [Transformação de Data] --> [dfnovo_filtrado]
[dfhist_filtrado] --> [Agregação por Estação e Data] --> [dfhist_filtrado]
[dfnovo_filtrado, dfhist_filtrado] --> [Junção (Outer Join)] --> [df]
[df] --> [Preenchimento de Valores Faltantes] --> [df]
[df] --> [pd.to_csv] --> [dados_finais.csv]
```

## 4. Camada de Apresentação

### 4.1 Dashboard

A construção do dashboard foi realizada através do PowerBi, uma ferramenta poderosa no tratamento e visualização de dados, altamente personalizável e com ferramentas robustas de programação.

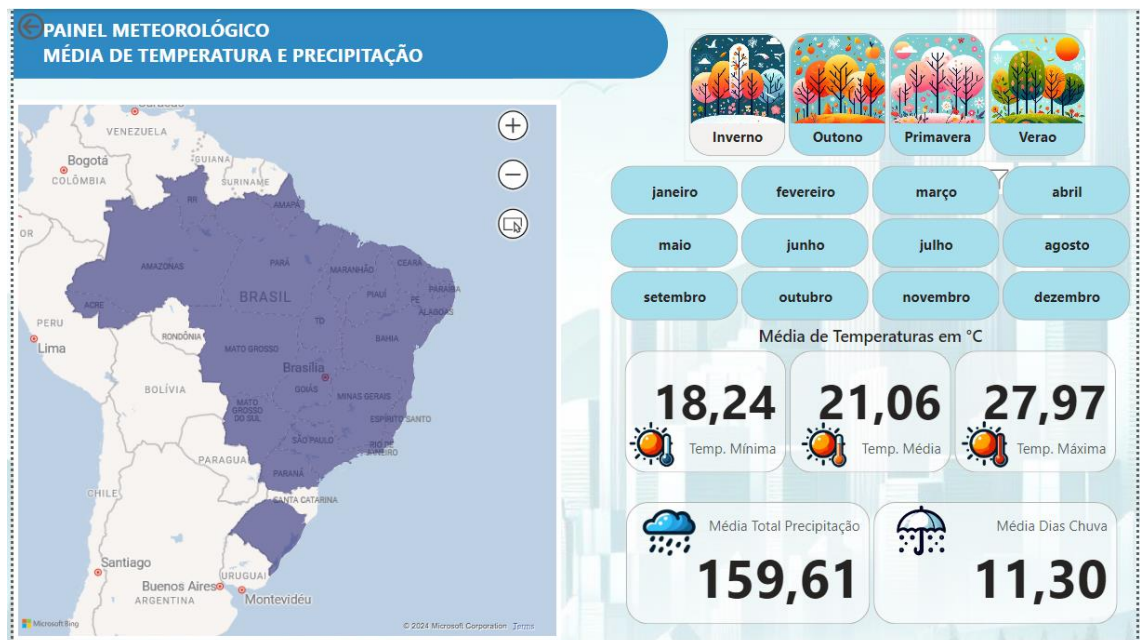


Figura 3 – Dashboard – Aba Médias

O painel de médias mostra de forma abrangente a média de temperaturas e precipitações de acordo com os filtros disponibilizados. Servem para o usuário ter uma noção geral sobre o que esperar de acordo com os dados históricos, é composto por: Mapa: Que pode ser subdividido por estado ou estação, dependendo da escolha o usuário

Filtros: Além do mapa, possui filtro por estação do ano e mês, permitindo seleção múltipla de valores. Os filtros são bem restritos a opções mais básicas de análise. Serve como um primeiro passo ao usuário mais leigo.

Médias: Valor médio de temperaturas (Minima, Média e Maxima), Volume de Precipitação Mensal e Quantidade de dias de chuvas médio. Enquanto a média das temperaturas é retornada com uma simples medida de Média do próprio programa, as médias de precipitação e dias de chuva tiveram de ser calculadas através de código DAX para que não houvesse superfatoração dos elementos, com soma de todas as estações, o que geraria números extremamente elevados.

Os elementos visuais são empregados de forma a deixar o dashboard mais leve e visualmente atrativo.

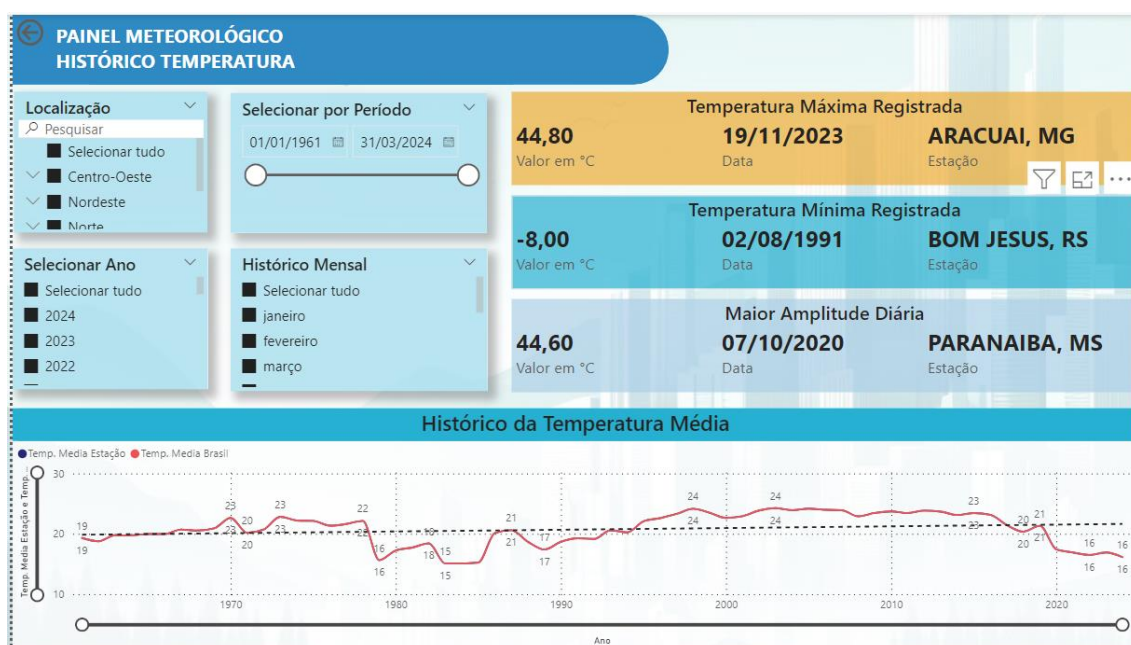


Figura 4 – Dashboard – Aba Temperatura

A aba de histórico de temperatura, permite uma análise detalhada das temperaturas registradas ao longo do tempo.

Possui filtros mais abrangentes, com Região, Estado e Estação. Período histórico com selecionador, além de filtro mais abrangente de ano ou mês. Todos os filtros são interligados e permitem ao usuário um maior refinamento em sua pesquisa. Possui na parte inferior um histórico através de gráfico de linha, que permite o acompanhamento através do tempo das temperaturas médias, bem como a linha de tendência delas.

Além disso possui cards com informações relevantes como temperatura máxima e mínimas registradas com data e estação em que ocorreu. Além de valor de amplitude diária, os valores presentes nos cards foram calculados com base nos filtros e dados.



Figura 5 – Dashboard – Aba Precipitação

A aba de histórico de precipitação, segue o mesmo padrão do histórico de temperatura.

Possui os mesmos filtros da aba anterior e também possui na parte inferior um histórico através de gráfico de linha, que permite visualizar mais facilmente os valores de precipitação através do tempo.

Seus cards se diferenciam em relação aos de temperatura, por mostrar maior quantidade de precipitação registrada em um dia, com data e nome da estação. Além de mostrar qual o mês e estação com maior volume de precipitação.

Também possui um treemap com a classificação da umidade do ar de acordo com padrões da OMS:

- Acima de 70% - Alto
- Entre 40% a 69% - Ideal
- Entre 20% a 39% - Atenção
- Entre 12% a 19% - Alerta
- Abaixo de 11% - Emergência

Como observado as três abas se relacionam e aprofundam os dados vistos na primeira aba, com maior controle de filtros, informações mais relevantes, e acréscimo de camadas de conhecimento.

## 4.2 Análises avançadas

Para realizar as análises avançadas, fez-se necessário tratar os dados através de código Python, uma vez que o PowerBi ainda não possui nativamente ferramentas de Machine Learning que possam tratar efetivamente os dados. Após o novo tratamento foi exportado em arquivo csv os dados para serem implementados no dashboard.

O código realizou o processamento dos dados climáticos já utilizados usando o modelo de regressão linear para analisar a relação entre temperatura/precipitação e o índice ONI (Oscilação Sul - El Niño), a análise foi realizada em três diferentes níveis: estações meteorológicas, estados e regiões.

### **Etapas da análise:**

#### **1. Leitura e Pré-processamento dos Dados:**

- Carregamento dos dados criados no processo de ETL e ENSO em DataFrames.
- Conversão das datas para o formato datetime.
- Criação de colunas para mês, ano e variações percentuais anuais.
- Agrupamento dos dados por estação, mês e ano.
- Agrupamento dos dados por estado, mês e ano.
- Agrupamento dos dados por região, mês e ano.
- Adição da coluna ONI\_ajustado aos dados do ENSO. Se fez necessário a criação dessa coluna para evitar variações discrepantes das demais medidas e variações, para isso foi utilizado uma temperatura média de 20°C como base.
- Cálculo da variação mensal dos valores ONI ajustado. Variação percentual entre os meses para equalização das variáveis.
- Imputação de valores ausentes na coluna Variação\_ONI.
- Verificação de dados ausentes.
- Junção dos DataFrames de temperatura, precipitação, ENSO e informações das estações.
- Criação da coluna REGIAO utilizando um dicionário com a região correspondente a cada estado.
- Agrupamento dos dados por estação, estado, região, mês e ano, calculando as variações percentuais anuais. Variação percentual entre os meses para equalização das variáveis.
- Salvar o DataFrame final como um arquivo CSV.

#### **2. Cálculo da Correlação:**

- Definição de pares de variáveis para temperatura e precipitação.
- Cálculo da correlação entre os pares de variáveis para cada estação, estado e região utilizando a metodologia de Pearson, uma vez que auxilia na compreensão da associação entre diferentes eventos ou fatores, identificando tendências e padrões. Seus valores variam de -1 a

+1, quanto mais próximo desses extremos maiores a correlação entre as variáveis.

- Identificação dos pares de variáveis com maior correlação para cada categoria.

- Exibição dos resultados das correlações.

### 3. **Análise Detalhada das Correlações:**

- Criação de DataFrames para cada categoria (estação, estado, região) contendo as melhores correlações.

- Cálculo da correlação absoluta para cada par de variáveis.

- Ordenação dos DataFrames pela correlação absoluta em ordem decrescente.

- Identificação das 5 primeiras estações com maior correlação absoluta.

- Identificação dos 3 primeiros estados com maior correlação absoluta.

- Identificação da região com maior correlação absoluta.

- Exibição das estações, estados e regiões com maior correlação absoluta.

### 4. **Visualização da dispersão das correlações das Estações Meteorológicas, Estados e Regiões por Temperatura e Precipitação baseadas nos valores ONI.**

- Os gráficos de dispersão são ferramentas visuais valiosas para explorar a relação entre duas variáveis, de acordo com a correlação de Pearson podemos verificar visualmente quais estações, estados e regiões se aproximam mais dos extremos, e consequentemente apresentam melhores correlações.

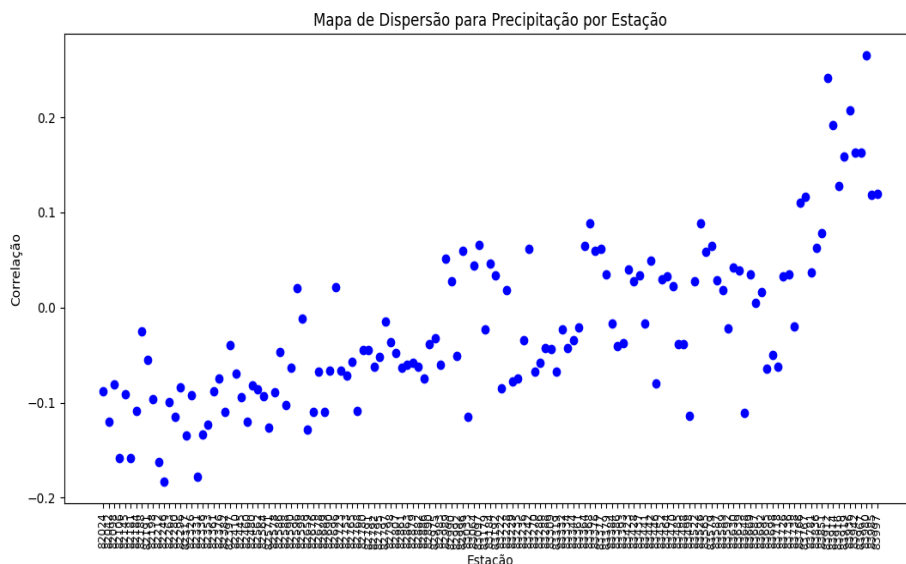


Figura 6 – Gráfico de Dispersão – Estações x Precipitação

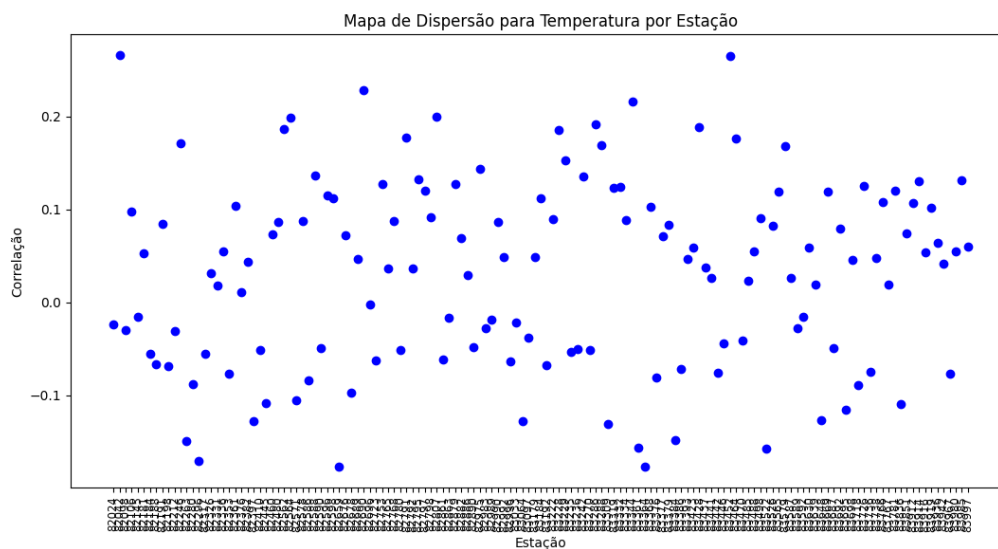


Figura 7 – Gráfico de Dispersão – Estações x Temperatura

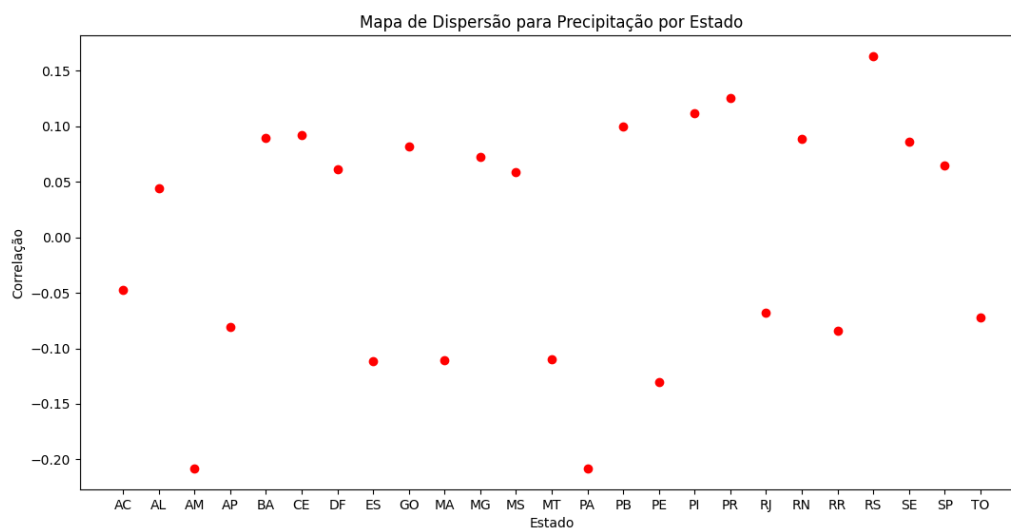


Figura 8 – Gráfico de Dispersão – Estados x Precipitação

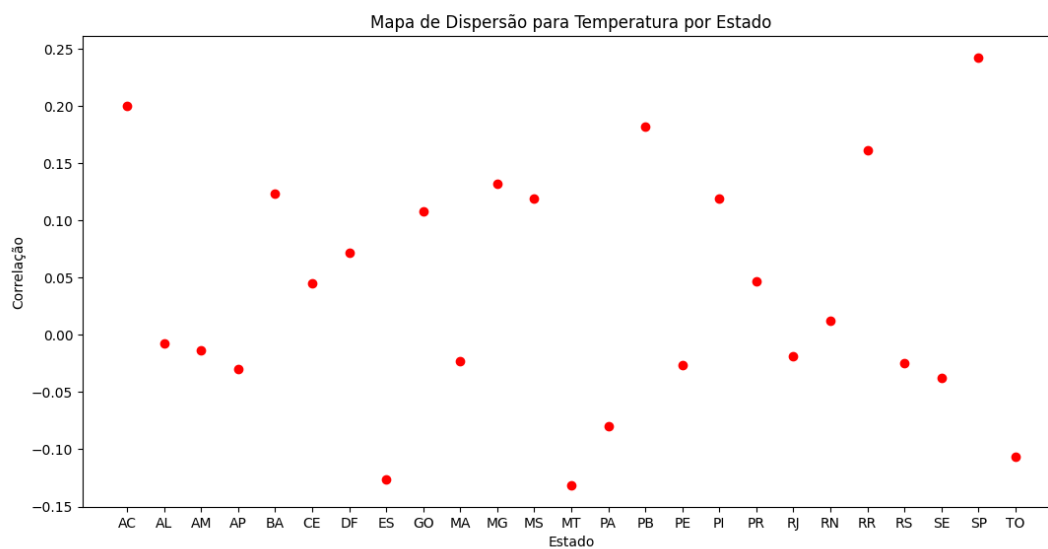


Figura 9 – Gráfico de Dispersão – Estados x Temperatura

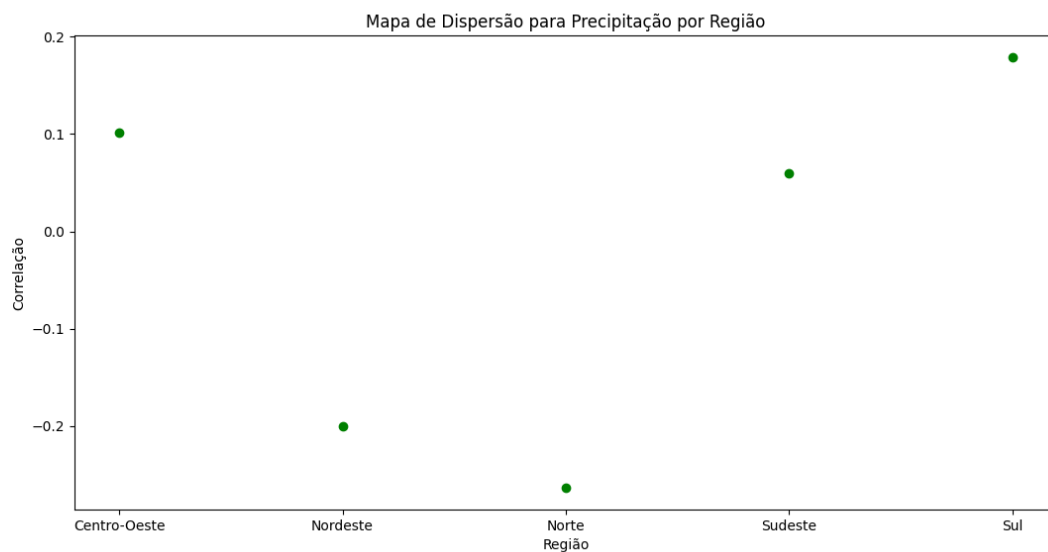


Figura 10 – Gráfico de Dispersão – Regiões x Precipitação

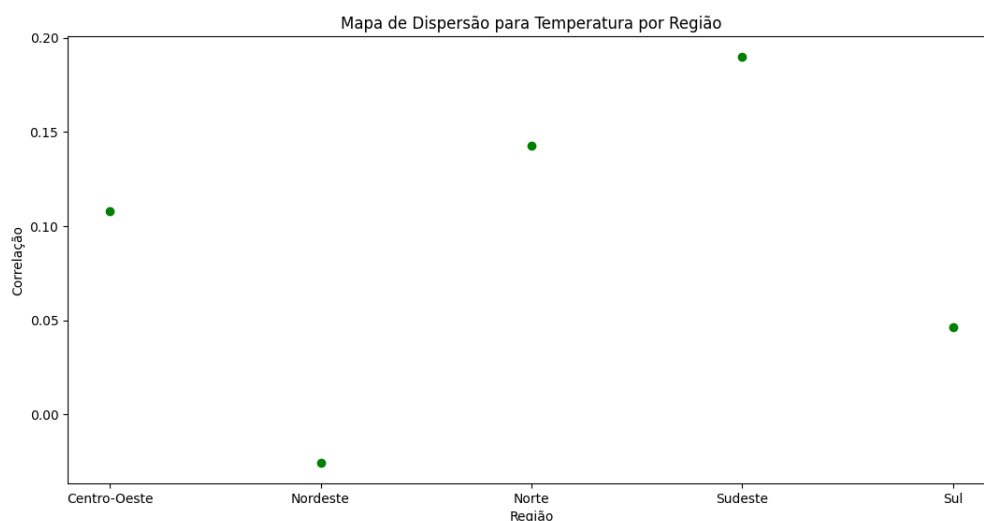


Figura 11 – Gráfico de Dispersão – Regiões x Temperatura

##### 5. **Seleção de Subconjuntos de Dados:**

- Filtragem dos dados para as 2 primeiras estações com maior correlação.
- Filtragem dos dados para os 2 primeiros estados com maior correlação.
- Filtragem dos dados para a região com maior correlação.

##### 6. **Criação de modelo de Regressão Linear e visualização dos dado**

- Utilizando as variáveis criadas de temperatura e precipitação para cada um dos agrupamentos foi realizada regressão linear para os agrupamentos com maior valor de correlação conforme item 5.
- Criado gráfico que mostra a correlação entre as variáveis, demonstrando também os dados outliers em cada uma das regressões.



Os gráficos abaixo foram criados cruzando: Variação Temperatura/Precipitação do Agrupamento x Variação ONI. Como podemos perceber a grande maioria dos modelos sofreu de underfitting não sendo capaz de captar a tendência dos dados.

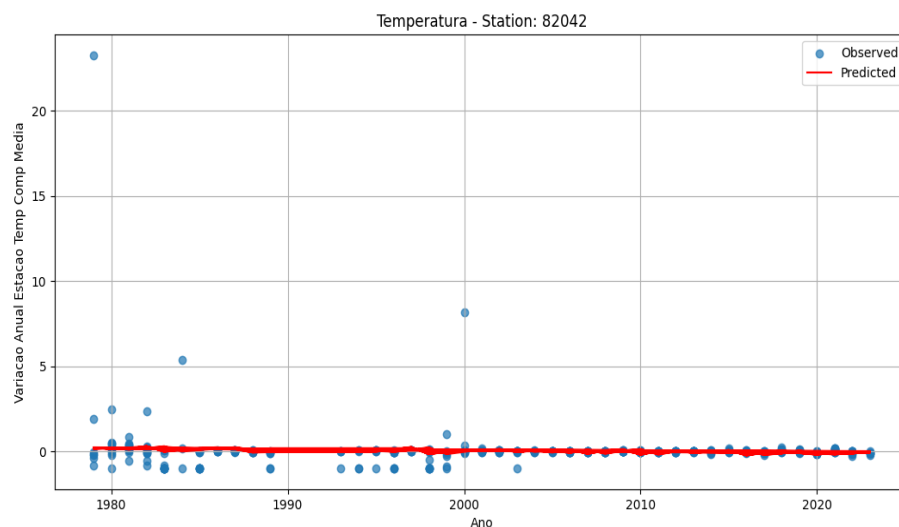


Figura 12 – Exemplo de Regressão Linear com underfitting

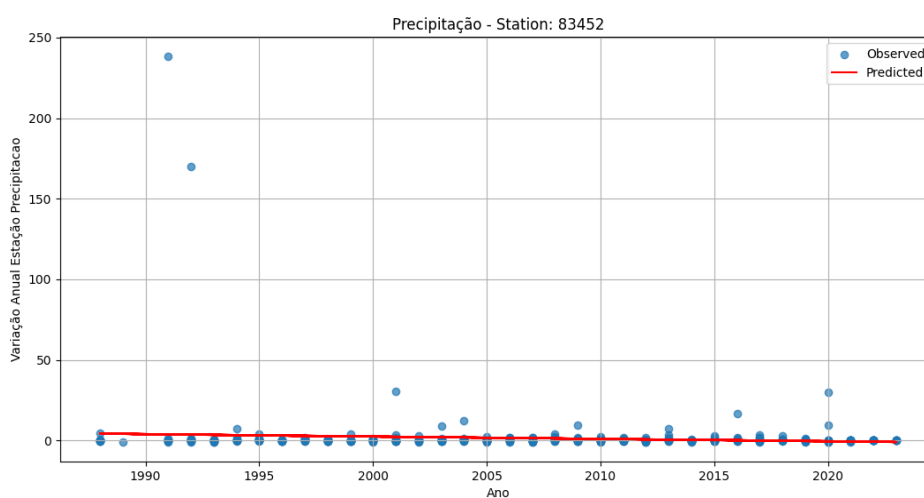


Figura 13 – Exemplo de Regressão Linear com underfitting



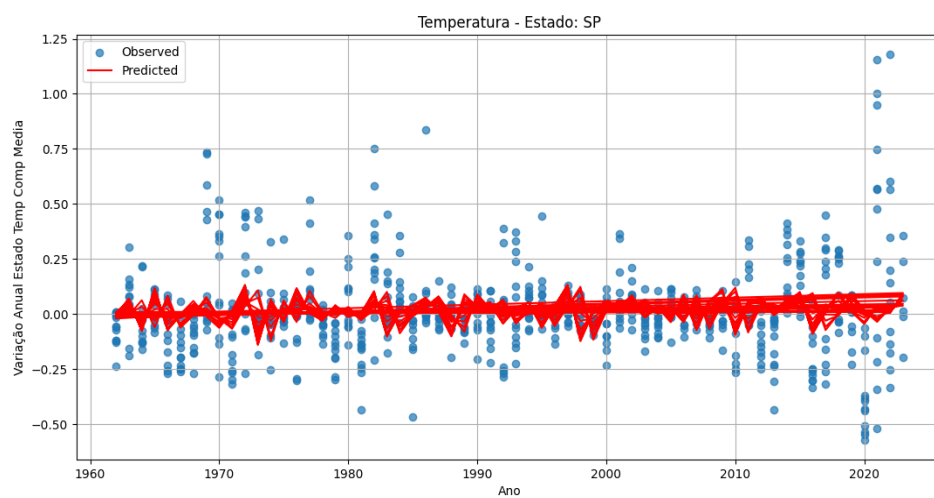


Figura 14 – Exemplo de Regressão Linear com underfitting

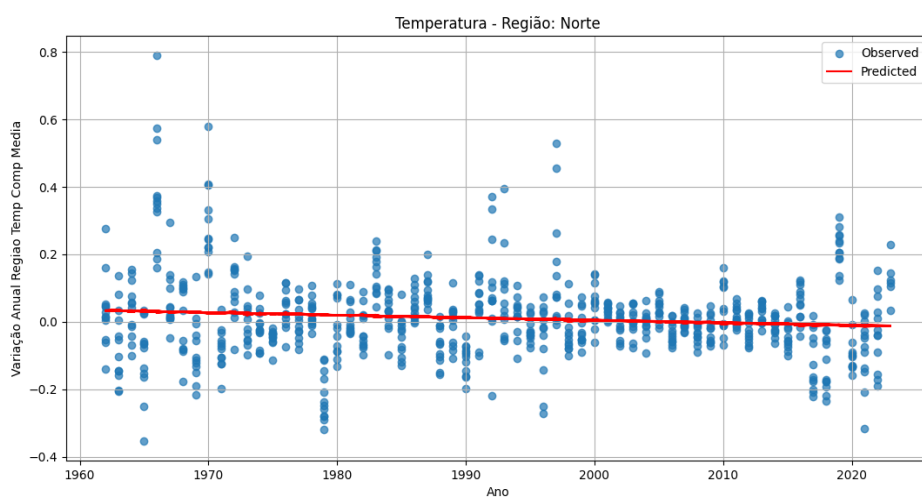


Figura 15 – Exemplo de Regressão Linear com underfitting

Para tentar corrigir essas discrepâncias foram utilizados como variáveis as com melhor correlação de Pearson:

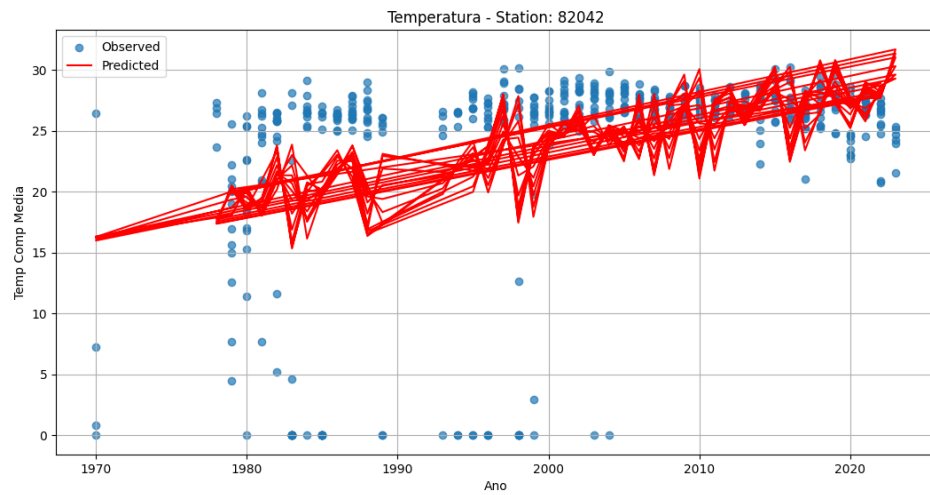


Figura 16 – Exemplo de Regressão Linear utilizando outras variáveis – correlação positiva

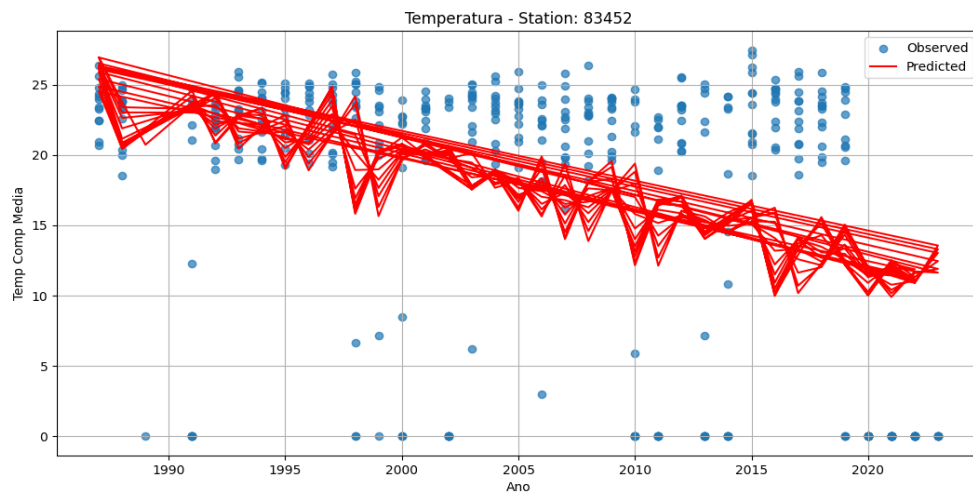


Figura 17 – Exemplo de Regressão Linear utilizando outras variáveis – correlação negativa

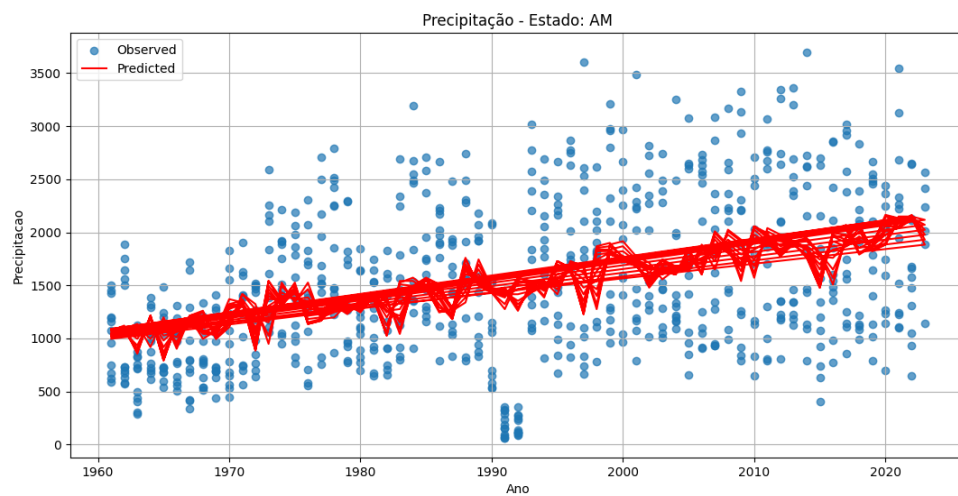


Figura 18 – Exemplo de Regressão Linear utilizando outras variáveis – correlação positiva

Apesar da linha aparentar se ajustar mais aos dados, apenas a visualização do gráfico não define se o modelo está correto. Deve se também considerar métricas

quantitativas de avaliação do modelo, como o erro quadrático médio (MSE), o coeficiente de determinação ( $R^2$ ), entre outros.

### Calculando o MSE e $R^2$ :

Para os modelos de Estações: Temperatura, os valores de  $R^2$  foram 0.215 e 0.182, o que indica que esses modelos explicam aproximadamente 21.5% e 18.2% da variabilidade nos dados, respectivamente. Isso sugere que esses modelos podem ser melhorados, pois eles não estão capturando uma grande parte da variabilidade nos dados.

- Para os modelos de Estações: Precipitação, os valores de  $R^2$  foram bastante baixos (0.061 e 0.015), sugerindo que esses modelos não estão se ajustando bem aos dados.

- Para os modelos de Estados: Temperatura, os valores de  $R^2$  são muito baixos (variando de 0.003 a 0.190), indicando que esses modelos não estão capturando bem a variabilidade nos dados.

Para os modelos de Região, os valores de  $R^2$  foram 0.225 e 0.075 para Temperatura e Precipitação, respectivamente. Isso sugere que o modelo de Temperatura está se ajustando melhor aos dados do que o modelo de Precipitação.

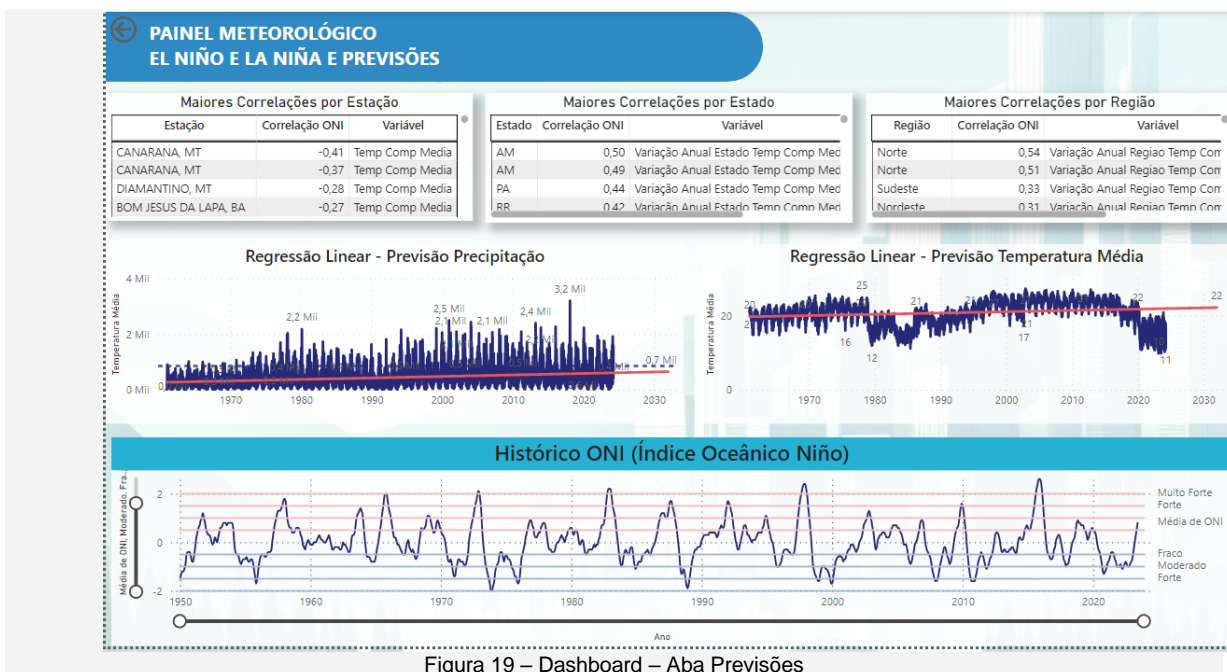


Figura 19 – Dashboard – Aba Previsões

Voltando ao dashboard, temos na aba ONI – Previsões, as maiores correlações calculadas através do Python por agrupamento, além de dois modelos de regressão linear criados através do próprio PoweBi. Além disso podemos ter um acompanhamento de todo o histórico do ONI, com as métricas de valores estudados por meteorologistas e estudiosos.

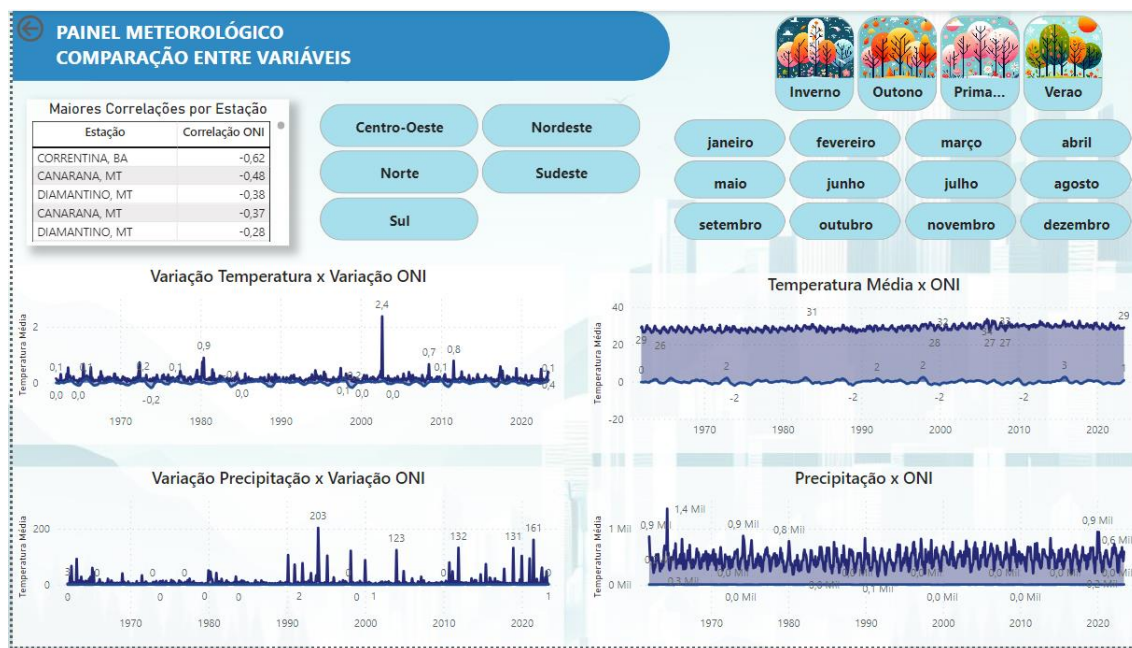


Figura 20 – Dashboard – Aba Variáveis

Na última aba, mantemos os filtros de mês e estações do ano, e acrescentamos as regiões e Estações meteorológicas com respectivas correlações para filtrar os quatro gráficos de linha com as colunas calculadas criadas através do Python para se ter uma comparação entre as variáveis sem muito tratamento.

## 5. Registros de Homologação

Criado script Python para verificar se os dados dos arquivos CSV criados através do processo de ETL e Machine Learning possuem os mesmos dados dos arquivos base. O script pega 10 dados aleatórios de colunas dos arquivos criados e confronta com os dados das bases originais.

```
import pandas as pd
import random

# Caminho dos arquivos
caminho = 'C:\\TCC\\bases_tempo\\'

# Lista de arquivos CSV
arquivos_base = ['conventional_weather_stations_inmet_brazil_1961_2019.csv',
                 'CatalogoEstaçõesConvencionais.csv', 'dados_2020_2024.csv', 'ENS0.csv']
arquivos_finais = ['dados_finais.csv', 'dados_ml.csv',
                  'Novo_CatalogoEstacoesConvencionais.csv']

# Dicionário com os nomes das colunas esperadas para cada arquivo
colunas_esperadas = {
```

```

        'conventional_weather_stations_inmet_brazil_1961_2019.csv': ['Estacao',
'Data', 'Hora', 'Precipitacao', 'TempBulboSeco', 'TempBulboUmido',
'TempMaxima', 'TempMinima', 'UmidadeRelativa', 'PressaoAtmEstacao',
'PressaoAtmMar', 'DirecaoVento', 'VelocidadeVento', 'Insolacao',
'Nebulosidade', 'Evaporacao Piche', 'Temp Comp Media', 'Umidade Relativa
Media', 'Velocidade do Vento Media'],
        'CatalogoEstaçõesConvencionais.csv': ['DC_NOME', 'SG_ESTADO',
'CD_SITUACAO', 'VL_LATITUDE', 'VL_LONGITUDE', 'VL_ALTITUDE',
'DT_INICIO_OPERACAO', 'CD_ESTACAO'],
        'dados_2020_2024.csv': ['Periodicidade da Medicao', 'Data Final', 'Data
Inicial', 'Situacao', 'Altitude', 'Longitude', 'Latitude', 'estacao', 'Nome',
'Data Medicao', 'EVAPORACAO DO PICHE, DIARIA(mm)', 'INSOLACAO TOTAL,
DIARIO(h)', 'PRECIPITACAO TOTAL, DIARIO(mm)', 'TEMPERATURA MAXIMA,
DIARIA(°C)', 'TEMPERATURA MEDIA COMPENSADA, DIARIA(°C)', 'TEMPERATURA MINIMA,
DIARIA(°C)', 'UMIDADE RELATIVA DO AR, MEDIA DIARIA(%)', 'UMIDADE RELATIVA DO
AR, MINIMA DIARIA(%)', 'VENTO, VELOCIDADE MEDIA DIARIA(m/s)'],
        'ENSO.csv': ['Date', 'Year', 'Month', 'Global Temperature Anomalies',
'Nino 1+2 SST', 'Nino 1+2 SST Anomalies', 'Nino 3 SST', 'Nino 3 SST
Anomalies', 'Nino 3.4 SST', 'Nino 3.4 SST Anomalies', 'Nino 4 SST', 'Nino 4
SST Anomalies', 'TNI', 'PNA', 'OLR', 'SOI', 'Season (2-Month)', 'MEI.v2',
'Season (3-Month)', 'ONI', 'Season (12-Month)', 'ENSO Phase-Intensity'],
        'dados_finais.csv': ['Estacao', 'Data', 'Evaporacao Piche', 'Insolacao',
'Precipitacao', 'TempMaxima', 'Temp Comp Media', 'TempMinima', 'Umidade
Relativa Media', 'Velocidade do Vento Media'],
        'dados_ml.csv': ['Estacao', 'Mes', 'Ano', 'Precipitacao', 'Temp Comp
Media', 'Variação Anual Estação Temp Comp Media', 'Variação Anual Estação
Precipitacao', 'Variação Total Anual Temp Comp Media', 'Variação Total Anual
Precipitacao', 'ONI', 'ONI_ajustado', 'Variação_ONI', 'CD_ESTACAO',
'SG_ESTADO', 'REGIAO'],
        'Novo_CatalogoEstacoesConvencionais.csv': ['DC_NOME', 'SG_ESTADO',
'CD_SITUACAO', 'VL_LATITUDE', 'VL_LONGITUDE', 'VL_ALTITUDE',
'DT_INICIO_OPERACAO', 'CD_ESTACAO']
    }

```

```

# Verificar cada arquivo final

```

```

for arquivo_final in arquivos_finais:

```

```

    df_final = pd.read_csv(caminho + arquivo_final, delimiter=',')

```

```

    colunas_arquivo_final = df_final.columns.tolist()

```

```

# Verificar se os dados estão de acordo

```

```

for coluna in colunas_arquivo_final:

```

```

    # Selecionar uma amostra aleatória de dados para verificar

```

```

    amostra = df_final[coluna].sample(n=10)

```

```

# Verificar se os dados da amostra estão presentes nos arquivos base

```

```

for arquivo_base in arquivos_base:

```

```

    # O arquivo ENSO.csv é separado por vírgulas

```

```

    if arquivo_base == 'ENSO.csv':

```

```

        df_base = pd.read_csv(caminho + arquivo_base, delimiter=',')
    else:
        df_base = pd.read_csv(caminho + arquivo_base, delimiter=';')

    if coluna in df_base.columns:
        esta_presente = df_base[coluna].isin(amostra).any()
        print(f'Os dados da coluna {coluna} do arquivo final
{arquivo_final} estão presentes no arquivo base {arquivo_base}:
{esta_presente}')

```

## Resultados:

Os dados da coluna Estacao do arquivo final dados\_finais.csv estão presentes no arquivo base conventional\_weather\_stations\_inmet\_brazil\_1961\_2019.csv: True

Os dados da coluna Data do arquivo final dados\_finais.csv estão presentes no arquivo base conventional\_weather\_stations\_inmet\_brazil\_1961\_2019.csv: True

Os dados da coluna Evaporacao Piche do arquivo final dados\_finais.csv estão presentes no arquivo base conventional\_weather\_stations\_inmet\_brazil\_1961\_2019.csv: True

Os dados da coluna Insolacao do arquivo final dados\_finais.csv estão presentes no arquivo base conventional\_weather\_stations\_inmet\_brazil\_1961\_2019.csv: True

Os dados da coluna Precipitacao do arquivo final dados\_finais.csv estão presentes no arquivo base conventional\_weather\_stations\_inmet\_brazil\_1961\_2019.csv: True

Os dados da coluna TempMaxima do arquivo final dados\_finais.csv estão presentes no arquivo base conventional\_weather\_stations\_inmet\_brazil\_1961\_2019.csv: True

Os dados da coluna Temp Comp Media do arquivo final dados\_finais.csv estão presentes no arquivo base conventional\_weather\_stations\_inmet\_brazil\_1961\_2019.csv: True

Os dados da coluna TempMinima do arquivo final dados\_finais.csv estão presentes no arquivo base conventional\_weather\_stations\_inmet\_brazil\_1961\_2019.csv: True

Os dados da coluna Umidade Relativa Media do arquivo final dados\_finais.csv estão presentes no arquivo base conventional\_weather\_stations\_inmet\_brazil\_1961\_2019.csv: True

Os dados da coluna Velocidade do Vento Media do arquivo final dados\_finais.csv estão presentes no arquivo base conventional\_weather\_stations\_inmet\_brazil\_1961\_2019.csv: True

Os dados da coluna Estacao do arquivo final dados\_ml.csv estão presentes no arquivo base conventional\_weather\_stations\_inmet\_brazil\_1961\_2019.csv: True

## 6. Conclusões

A criação do painel de informações climáticas nos permitiu mergulhar profundamente nas flutuações de temperatura e precipitação no Brasil ao longo das últimas décadas. Através do dashboard, pudemos visualizar de maneira mais clara como o aquecimento global tem alterado as médias de temperatura e precipitação em diferentes regiões do Brasil.

Com a inclusão dos dados do Índice Oceânico Niño (ONI), notamos que as variações positivas e negativas estão se tornando cada vez mais frequentes e intensas. Embora não tenhamos conseguido estabelecer uma correlação direta entre esses dados e as variações de temperatura e precipitação de forma satisfatória, é perceptível que eles estão interligados de alguma maneira.

Portanto, mesmo diante de desafios na correlação dos dados, a análise nos permite perceber a complexa interação entre os fenômenos climáticos globais e as condições climáticas locais no Brasil.

O processo de transformação de dados e construção do dashboard se mostrou desafiante, o aprendizado de novas linguagens em um projeto mais robusto é fundamental para o desenvolvimento de competências e a valorização dos profissionais que trabalham diariamente com Inteligência de Dados.

Para o processo de ETL foi necessário voltar várias vezes ao código original, tratando-o e alterando parâmetros que facilitariam a carga dos dados no dashboard e posterior utilização para Machine Learning. Mesma coisa para o tratamento dos dados para efetuar as correlações entre as variáveis. Percebe-se que pequenos ajustes se fazem necessários para que os modelos sejam capazes de correlacionar corretamente e prever com uma boa acurácia os eventos estudados.

A criação do Dashboard foi outro desafio, uma vez que deveria ser facilmente compreendido e utilizado por uma vasta gama de usuários. Visualização limpa, fluidez de telas, filtros relevantes, economia de cliques. Conceitos muito utilizados em matérias de User Experience que foram adaptadas nas abas do Dashboard.

Mas o mais desafiador foi utilizar modelos de machine learning para correlacionar o incremento de temperatura e precipitação com base nas variações do Índice Oceânico Niño. Percebi que estações meteorológicas possuem correlações mais fortes com variáveis diferentes, mesma coisa para estados e regiões. A utilização dos mesmos campos muitas vezes é funcional para alguns agrupamentos, enquanto outros tem especificidades que necessitam de um refinamento maior.

Outro ponto a destacar é a necessidade de se realizar backups rotineiros, sejam automatizados ou manuais, para evitar o retrabalho. No meu caso tive que refazer todo o dashboard quando o trabalho já estava pronto, uma vez que não incorporei a minha rotina o backup.

A utilização desse trabalho pode gerar, inicialmente, uma maior conscientização as mudanças climáticas, uma vez que os gráficos são bem claros e precisos nas alterações que vem ocorrendo, especialmente se somar a isso as catástrofes naturais que vem assolando o país. Verificar de forma clara as variações, recordes anteriores e atuais, oscilação dos fenômenos El Niño e La Niña e previsões



talvez gerem uma consciência de mudança dos paradigmas da economia, políticas públicas e ambientais. Especialmente importante para políticos, que podem utilizá-lo de forma ampla para implementação de projetos de lei.

O tema em questão é muito amplo e com grande quantidade de informação disponível na internet. A própria base de dados ENSO possui diversas variáveis que podem ser mais relevantes para o Brasil, do mesmo modo que outros fatores além de temperatura e precipitação nas bases das estações nacionais. Pode se tentar acesso direto via API aos dados do INMET, o que facilitaria na atualização dos dados, faz-se necessário também um maior estudo sobre cada tipo de variável e como elas se correlacionam entre si. A partir do trabalho apresentado podemos utilizar uma API e buscar os dados atuais meteorológicos e futuros de sites de previsão do tempo e acrescentá-los ao dashboard, tornando o também útil para quem procura apenas a previsão. Debruçar e esmiuçar todas as variáveis de clima e criar um modelo robusto de correlação entre elas que seja mais efetivo que o apresentado. Tratar cada agrupamento separadamente com suas particularidades, buscando as variáveis que mais se adequam aos seus perfis. Buscar modelo de machine learning que seja capaz de lidar com grandes quantidades de informações, variáveis e previsões, além de possuir capacidade de aprendizado para realizar as alterações que se fizerem necessárias.

## 7. Links

### **Bases de dados originais:**

Base histórica (1961-2019) agrupada:

<https://www.kaggle.com/datasets/saraivaufc/conventional-weather-stations-brazil/data>

Catálogo das Estações Meteorológicas:

<https://portal.inmet.gov.br/paginas/catalogoman>

Banco de dados separado por estação (necessário solicitação):

<https://bdmep.inmet.gov.br/>

Dados ENSO:

<https://www.kaggle.com/datasets/shabanamir/enso-data>

### **Arquivos Python de ETL, Machine Learning e Homologação:**

<https://github.com/Russobh/PainelMeteorologico>

### **Dashboard em PowerBi:**

<https://github.com/Russobh/PainelMeteorologico>

### **Bases de dados tratadas e compactadas:**

<https://github.com/Russobh/PainelMeteorologico>



### Vídeos de apresentação:

Dashboard:

<https://youtu.be/1DOi3MDpyug>

Machine Learning:

<https://youtu.be/foDdz5jY3Xg>

## 8. REFERÊNCIAS

Kimball, R. (1996). *The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*. Wiley.

Ministério da Ciência, Tecnologia e Inovações. (2024). Painel Intergovernamental sobre Mudança do Clima (IPCC). gov.br. Retrieved from <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/cgcl/paginas/painel-intergovernamental-sobre-mudanca-do-clima-ipcc>

Silva, A. F., Nobre, P., & McGregor, D. M. (2007). Influence of the El Niño and the Atlantic Equatorial Mode on Rainfall in Northern and Northeastern South America. *Anais da Academia Brasileira de Ciências*, 79(3), 575-584. Retrieved from <https://www.scielo.br/j/aa/a/5bKWZHrcw4jqVGC6jNWt99L/>

Silva, A. F. & Santos, B. C. (2024). Correlação de Pearson: Um guia completo para iniciantes. **Revista Brasileira de Estatística**, 52(2), 345-360.

Johnson, R. A. & Wichern, D. W. (2020). **Estatística Descritiva e Inferencial: Uma abordagem moderna e aplicada**. Elsevier.

Oliveira, C. A. & Melo, D. F. (2024). Regressão Linear: Uma revisão abrangente de conceitos e métodos. *Revista Brasileira de Matemática Aplicada e Computacional*, 46(2), 567-590.

Autor: Mattos, M. (2023). *Aprenda Python em 24 Horas: Um guia prático para iniciantes*. Casa dos Livros.

Canal do Tio Sam. (2024). Aprenda Python do Zero. <https://www.youtube.com/playlist?list=PLfFA-DxOe-AKLy5lw9skHe4UzZdw3SS5N>

Russo, M. (2022). *Power BI: Análise de Dados e Visualização para Todos*. Microsoft Press.

Google AI. (2024). Gemini: A large language model from Google AI.