

Introdução

O Instituto Nacional de Meteorologia (INMET) é responsável por coletar e fornecer dados meteorológicos que servem de base para análises climatológicas no Brasil. Esses dados são fundamentais para o estudo de padrões climáticos, previsão de tempo, pesquisas agrícolas, planejamento urbano e diversas outras áreas que dependem das condições atmosféricas.

Objetivo

Neste trabalho, realizaremos uma análise exploratória dos dados fornecidos pelo INMET, com o objetivo de responder questões específicas relacionadas às condições climáticas de um período específico. Faremos uso de técnicas estatísticas e visualização de dados para interpretar padrões de temperatura, precipitação e outras variáveis meteorológicas. As análises serão baseadas em períodos trimestrais e diários, fornecendo insights sobre a variação de temperatura e condições climáticas ao longo do tempo.

Analisar as temperaturas máxima, mínima e média durante o primeiro trimestre dos anos de 2003 e 2004 em Uberlândia, MG.

- Explorar as variações de temperatura em diferentes horários ao longo do período analisado.
- Avaliar a correlação entre pressão atmosférica e temperatura, identificando possíveis padrões de relação entre essas variáveis.

Além das questões centrais abordadas são:

- Quais foram as temperaturas máxima, mínima e média durante o primeiro trimestre?
- Quais são as principais variações climáticas durante determinados horários ao longo do período analisado?
- Como os dados meteorológicos podem ser utilizados para prever padrões futuros?

Limpeza e Preparação dos Dados

Para garantir que os dados analisados sejam precisos e relevantes, foi necessário realizar um processo de limpeza dos dados. Primeiramente, foram removidas as 8 primeiras linhas e colunas irrelevantes que não agregavam valor à análise. No caso

do ano de 2003 UBERLANDIA - MG, a planilha original continha **8769 linhas e 19 colunas**, representando um total de de dados. Após a exclusão das colunas que não contribuíam para a análise (como identificadores ou informações duplicadas), o número de colunas foi reduzido para 14, resultando em **166.611**. Após a exclusão das colunas irrelevantes, o número de colunas foi reduzido para 14, e o total de células processadas caiu para **1136**. Tendo em vista que foram excluídas todas as linhas com valores negativos como -9999 , identificado como erro na tabela.

Em 2004 UBERLANDIA MG, o processo de limpeza foi similar. A planilha original continha **8793 linhas e 19 colunas**, com um total de **167.0670 células**. Após a exclusão das colunas irrelevantes, o número de colunas foi reduzido para 14, e o total de células processadas caiu para **1680**, como na limpeza anterior foram excluídas todas as linhas com valores negativos como -9999 , identificado como erro na tabela.

Como foi feito?

A limpeza dos dados foi realizada utilizando o Microsoft Excel e o Power Query, o código foi escrito manualmente porem comentado com auxilio do chat GPT. O processo seguiu as seguintes etapas:

1. **Abertura do Arquivo CSV no Excel:** Os arquivos contendo os dados dos anos 2003 UBERLANDIA e 2004 UBERLANDIA foram abertos diretamente no Excel, utilizando a funcionalidade "Transformar dados", que permitiu uma visão clara das informações.
2. **Exclusão de Colunas Irrelevantes e Limpeza de Dados com Power Query:** Em seguida, foram removidas as colunas que não eram relevantes para o estudo, mantendo apenas as colunas que continham atributos essenciais para a análise de clima / tempo sendo elas , data e hora utc, precipitação total, horário (mm), pressão atmosférica ao nível da estação, horária (mb), pressão atmosférica max. na hora ant. (aut) (mb), pressão atmosférica min. na hora ant. (aut) (mb), temperatura do ar - bulbo seco, horária (°c), temperatura máxima na hora ant. (aut) (°c), temperatura mínima na hora ant. (aut) (°c), vento, direção horária (gr) (° (gr)), vento, rajada máxima (m/s), vento, velocidade horária (m/s). Foi escrito o código abaixo, com o objetivo de excluir dados meteorológicos inválidos. Isso incluiu a remoção de linhas onde os valores estavam marcados como "-9999" ou onde outros campos continham dados ausentes ou inconsistentes, garantindo a qualidade e

consistência das informações para análise, o código foi utilizado nas duas bases alterando apenas a fonte.

código:

```
let
// Importa o arquivo CSV e remove as 8 primeiras linhas
Fonte = Table.PromoteHeaders(
    Table.Skip(
        Csv.Document(File.Contents("C:\Users\Naya
Nascimento\Downloads\INMET_SE_MG_A507_UBERLANDIA_01-01-2004_A_31-
12-2004.csv"),
            [Delimiter=";", Columns=19, Encoding=1252, QuoteStyle=QuoteStyle.None]),
        8),
    [PromoteAllScalars=true]),

// Filtra colunas necessárias
ColunasMantidas = Table.SelectColumns(Fonte, {
    "DATA (YYYY-MM-DD)", "HORA (UTC)", "PRECIPITAÇÃO TOTAL, HORÁRIO (mm)",
    "PRESSAO ATMOSFERICA AO NIVEL DA ESTACAO, HORARIA (mB)",
    "PRESSÃO ATMOSFERICA MAX.NA HORA ANT. (AUT) (mB)", "PRESSÃO
ATMOSFERICA MIN. NA HORA ANT. (AUT) (mB)",
    "RADIACAO GLOBAL (KJ/m²)", "TEMPERATURA DO AR - BULBO SECO, HORARIA
(°C)",
    "TEMPERATURA DO PONTO DE ORVALHO (°C)", "TEMPERATURA MÁXIMA NA
HORA ANT. (AUT) (°C)",
    "TEMPERATURA MÍNIMA NA HORA ANT. (AUT) (°C)", "VENTO, DIREÇÃO
HORARIA (gr) (° (gr))",
    "VENTO, RAJADA MAXIMA (m/s)", "VENTO, VELOCIDADE HORARIA (m/s)"
}),

// Remove linhas onde qualquer valor é -9999, null ou 0
LinhasValidas = Table.SelectRows(ColunasMantidas,
    each List.NonNullCount(List.Select(Record.ToList(_), each _ <> null and _ <> -
9999 and _ <> 0)) = Table.ColumnCount(ColunasMantidas)),

// Remove linhas onde os valores da coluna C até a N são iguais
LinhasSemDuplicatas = Table.SelectRows(LinhasValidas,
    each List.Distinct({
        ["PRECIPITAÇÃO TOTAL, HORÁRIO (mm)"], ["PRESSAO ATMOSFERICA AO
NIVEL DA ESTACAO, HORARIA (mB)"],
        ["PRESSÃO ATMOSFERICA MAX.NA HORA ANT. (AUT) (mB)"], ["PRESSÃO
ATMOSFERICA MIN. NA HORA ANT. (AUT) (mB)"]
```

```
[#"RADIACAO GLOBAL (KJ/m²)", [#"TEMPERATURA DO AR - BULBO SECO,
HORARIA (°C)"],
[#"TEMPERATURA DO PONTO DE ORVALHO (°C)", [#"TEMPERATURA MÁXIMA
NA HORA ANT. (AUT) (°C)"],
[#"TEMPERATURA MÍNIMA NA HORA ANT. (AUT) (°C)"]]} <> { [#"PRECIPITAÇÃO
TOTAL, HORÁRIO (mm)"]}],
```

```
// Altera tipos de dados
TipoAlterado = Table.TransformColumnTypes(LinhasSemDuplicatas, {
{"DATA (YYYY-MM-DD)", type date}, {"HORA (UTC)", type time},
{"PRECIPITAÇÃO TOTAL, HORÁRIO (mm)", type number},
{"PRESSAO ATMOSFERICA AO NIVEL DA ESTACAO, HORARIA (mB)", type
number},
{"PRESSÃO ATMOSFERICA MAX.NA HORA ANT. (AUT) (mB)", type number},
{"PRESSÃO ATMOSFERICA MIN. NA HORA ANT. (AUT) (mB)", type number},
{"RADIACAO GLOBAL (KJ/m²)", Int64.Type},
{"TEMPERATURA DO AR - BULBO SECO, HORARIA (°C)", type number},
{"TEMPERATURA DO PONTO DE ORVALHO (°C)", type number},
{"TEMPERATURA MÁXIMA NA HORA ANT. (AUT) (°C)", type number},
{"TEMPERATURA MÍNIMA NA HORA ANT. (AUT) (°C)", type number}
}),
```

como tipo de data

```
ColunaMes = Table.AddColumn(TipoAlterado, "Mês", each if [#"DATA (YYYY-MM-
DD)"] <> null then Date.Month([#"DATA (YYYY-MM-DD)"]) else null, Int64.Type)
```

in

ColunaMes

Criação de Dashboard: Após a preparação da base de dados, foi criada uma nova aba na planilha para o desenvolvimento de um dashboard. Esse dashboard inclui um título descritivo e a inserção de uma imagem da bandeira do município de Uberlândia, personalizando a visualização de resultados e facilitando a comunicação dos dados.

Esse processo de filtragem e limpeza garantiu que os dados utilizados fossem os mais precisos possíveis, proporcionando uma base sólida para a análise detalhada do perfil do eleitorado.

Estrutura do Estudo

1. **Filtragem de Linhas por Critérios Específicos:** Seleção e análise dos dados meteorológicos com base em critérios pré-definidos, como datas

específicas, variações de precipitação, pressão atmosférica, temperatura e vento, utilizando as colunas "data (yyyy-mm-dd)", "precipitação total, horário (mm)", "temperatura do ar - bulbo seco, horaria (°c)", "temperatura máxima na hora ant. (aut) (°c)", "temperatura mínima na hora ant. (aut) (°c)".

2. **Contagem e Agrupamento dos Dados:** Agrupamento dos dados meteorológicos com base nas colunas mencionadas, permitindo contagens de eventos climáticos específicos e combinações de variáveis para identificar padrões mais complexos.
3. **Estatísticas Descritivas:** Cálculo das medidas de tendência central (média, moda, mediana) e de dispersão (máximo, mínimo, intervalo de valores) para as variáveis meteorológicas, utilizando as colunas "precipitação total, horário (mm)", "pressao atmosferica ao nivel da estacao, horaria (mb)", "pressão atmosferica max.na hora ant. (aut) (mb)", "pressão atmosferica min. na hora ant. (aut) (mb)", "radiacao global (kj/m²)", "temperatura do ar - bulbo seco, horaria (°c)", "temperatura do ponto de orvalho (°c)", "temperatura máxima na hora ant. (aut) (°c)", "temperatura mínima na hora ant. (aut) (°c)", "vento, direção horaria (gr) (° (gr))", "vento, rajada maxima (m/s)", e "vento, velocidade horaria (m/s)".
4. **Comparação entre Períodos:** Comparação das características meteorológicas entre diferentes períodos, utilizando as colunas "data (yyyy-mm-dd)", "precipitação total, horário (mm)", "pressao atmosferica ao nivel da estacao, horaria (mb)", "pressão atmosferica max.na hora ant. (aut) (mb)", "pressão atmosferica min. na hora ant. (aut) (mb)", "radiacao global (kj/m²)", "temperatura do ar - bulbo seco, horaria (°c)", "temperatura do ponto de orvalho (°c)", "temperatura máxima na hora ant. (aut) (°c)", "temperatura mínima na hora ant. (aut) (°c)", "vento, direção horaria (gr) (° (gr))", "vento, rajada maxima (m/s)", "vento, velocidade horaria (m/s)".
5. **Visualização Gráfica e Mapas de Calor:** Utilização de gráficos dinâmicos e mapas de calor (heatmaps) para ilustrar os padrões encontrados, facilitando a interpretação dos resultados e a análise das variáveis como precipitação, temperatura, pressão atmosférica e vento.

Filtragem de Linhas por Critérios Específicos:

Após a limpeza e preparação dos dados, foi realizada a filtragem das linhas com base em critérios demográficos específicos, permitindo uma análise mais detalhada de grupos eleitorais.

Critério Utilizado:

- **Precipitação Total**
- **Data agrupado por Mês**

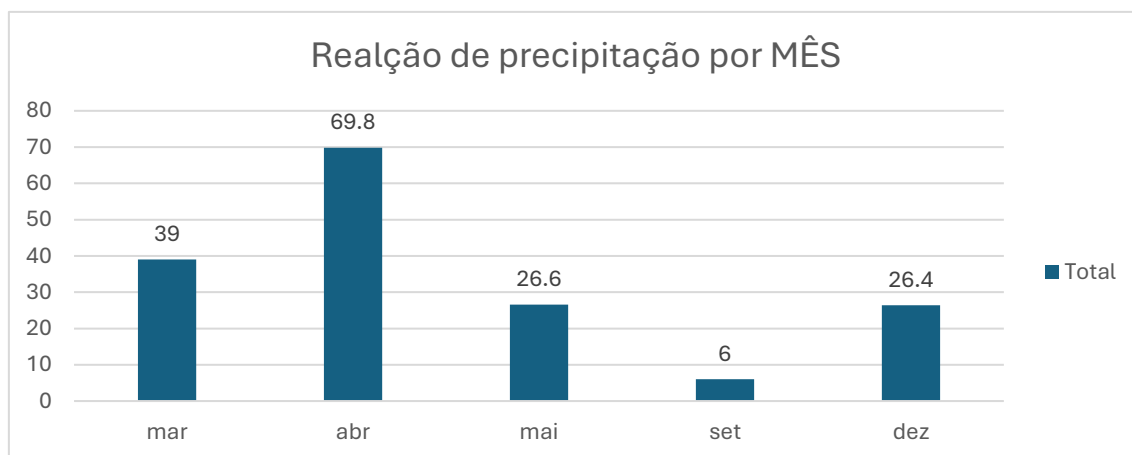
- Data agrupado por Trimestre

Resultado da Filtragem:

Uberlândia 2003

- **Total de PRECIPITAÇÃO no semestre: 167,8 mm**
- **Total geral de registros: 1136**

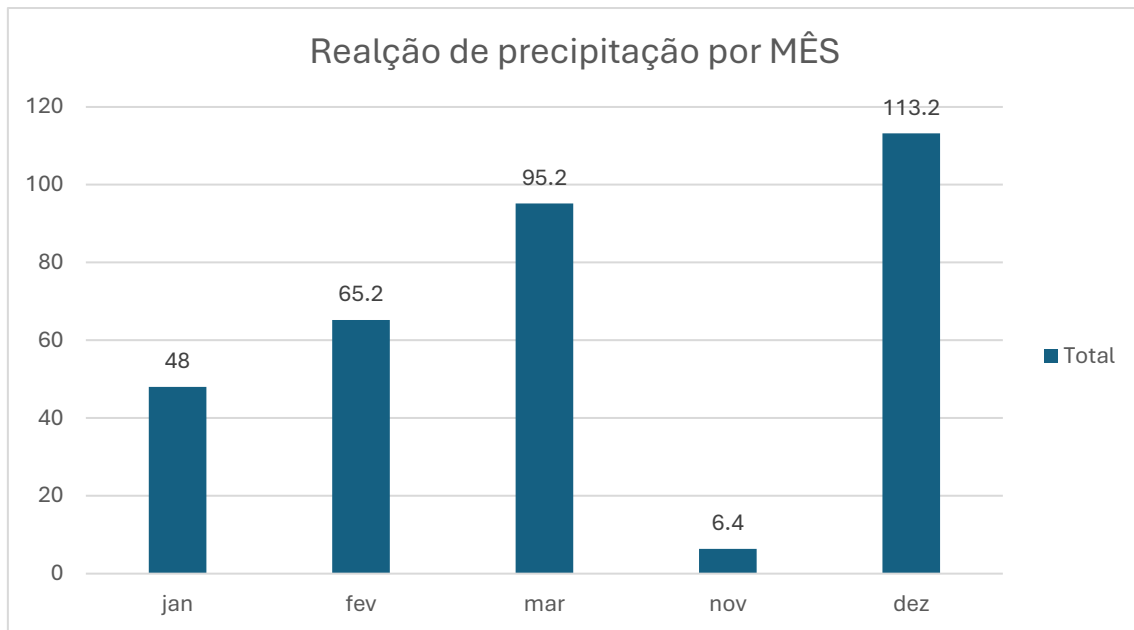
MÊS	PRECIPITAÇÃO TOTAL, HORÁRIO (mm)
mar	39
abr	69,8
mai	26,6
set	6
dez	26,4
Total Geral	167,8



Uberlândia 2004

- **Total de PRECIPITAÇÃO no semestre: 328 mm**
- **Total geral de registros: 1680**

Mês	Soma de PRECIPITAÇÃO TOTAL, HORÁRIO (mm)
jan	48
fev	65,2
mar	95,2
nov	6,4
dez	113,2
Total Geral	328



Como foi feito?

Após a criação do Dashboard, foi elaborada uma tabela dinâmica que agrupa as informações de ambos os anos. Nessa tabela, foram adicionados filtros mensais e somatório de dados de precipitação. No campo das linhas, foi inserida a informação de data, e nos valores foi inserida a precipitação. Essa configuração permite uma análise mais detalhada e interativa dos dados, facilitando a visualização do acúmulo de água ao longo da identificação de padrões no clima/tempo da região.

Contagem e Agrupamento dos Dados:

Após a filtragem dos dados com base em critérios temporais específicos, o próximo passo da análise consistiu em calcular a média das temperaturas mínimas e máximas para cada horário disponível. Os dados foram então agrupados por horário, permitindo uma visualização clara das variações de temperatura ao longo do dia. Essa etapa é essencial para compreender os padrões de temperatura em diferentes momentos, fornecendo insights valiosos sobre as condições climáticas diárias. O resultado desses agrupamentos é apresentado na seção de "Resultados".

Como foi feito?

Para uma análise abrangente dos dados, foram criadas 4 tabelas dinâmicas, proporcionando diferentes perspectivas sobre as informações. Essas tabelas oferecem uma visualização clara e organizada, facilitando a identificação de padrões e insights relevantes. Abaixo estão as descrições das tabelas geradas:

- **Relação de precipitação por MÊS**
Relação de Temperatura x Hora
- **Relação de precipitação por Trimestre**
- **Relação de Variáveis de Temperatura por MÊS X DIA X HORA**
- **Relação de Variáveis de Temperatura por TRIMESTRE X DIA**

Essas tabelas dinâmicas não apenas organizam os dados de forma acessível, mas também fornecem uma base sólida para análises futuras. Por meio delas, é possível extrair insights valiosos sobre padrões climáticos, auxiliando na compreensão das condições meteorológicas e no planejamento de ações relacionadas ao clima. Essas informações podem ser utilizadas para aprimorar políticas de gestão ambiental e prever tendências meteorológicas com maior precisão.

Estatísticas Descritivas da Análise

As estatísticas descritivas desempenham um papel crucial na análise dos dados meteorológicos, permitindo resumir e interpretar as informações de forma eficaz. Nesta seção, apresentamos as principais estatísticas descritivas obtidas a partir das tabelas dinâmicas criadas, oferecendo uma visão clara sobre os padrões climáticos observados nos dados do INMET, como temperatura mínima, máxima, precipitação e outras variáveis meteorológicas relevantes.

1. Média (Mean):

A média aritmética é o valor obtido pela soma de todos os valores de um conjunto de dados dividida pelo número de valores. Ela representa o ponto de equilíbrio da distribuição dos dados.

2. Erro padrão (Standard Error):

O erro padrão é uma medida de quão dispersa está a média amostral de um conjunto de dados em relação à média verdadeira da população. Ele é calculado como o desvio padrão da amostra dividido pela raiz quadrada do tamanho da amostra.

3. Mediana (Median):

A mediana é o valor que separa o conjunto de dados em duas metades, ou seja, 50% dos dados estão abaixo da mediana e 50% estão acima. Se houver um número ímpar de valores, a mediana será o valor central. Se houver um número par de valores, a mediana será a média dos dois valores centrais.

4. Modo (Mode):

O modo é o valor que aparece com mais frequência em um conjunto de dados. Um conjunto de dados pode ter mais de um modo se múltiplos valores aparecerem com a mesma frequência máxima.

5. Desvio padrão (Standard Deviation):

O desvio padrão mede a dispersão dos dados em torno da média. Quanto maior o desvio padrão, mais dispersos estão os valores em relação à média.

6. Variância da amostra (Sample Variance):

A variância é a medida da dispersão dos dados em relação à média, e é o quadrado do desvio padrão. A variância da amostra é calculada com o número de observações da amostra menos 1, em vez do número total de observações.

7. Curtose (Kurtosis):

Curtose mede a "cauda" ou extremidade da distribuição de dados em comparação com uma distribuição normal. Ela indica a probabilidade de ocorrência de valores extremos.

8. Assimetria (Skewness):

A assimetria mede a simetria ou o desvio da distribuição em torno da média. Uma distribuição pode ser:

- **Positivamente assimétrica:** Cauda maior à direita (valores extremos mais frequentes à direita da média).

- **Negativamente assimétrica:** Cauda maior à esquerda (valores extremos mais frequentes à esquerda da média).
-

9. Intervalo (Range):

O intervalo é a diferença entre o valor máximo e o valor mínimo de um conjunto de dados.

10. Mínimo (Minimum):

O valor mínimo é o menor valor presente em um conjunto de dados.

11. Máximo (Maximum):

O valor máximo é o maior valor presente em um conjunto de dados.

12. Soma (Sum):

A soma é o total obtido pela adição de todos os valores de um conjunto de dados.

13. Contagem (Count):

A contagem é o número total de observações em um conjunto de dados.

a. Temperatura Máxima, Mínima e Média do Primeiro Trimestre (2003 e 2004)

Analizamos a variação da temperatura máxima, mínima e média durante os primeiros três meses de cada ano. Abaixo, apresentamos os dados para o período de janeiro a março de 2003 e 2004:

Tabela: Temperaturas de Uberlândia (MG) – 1º Trimestre

Esses valores mostram a variação de temperatura em Uberlândia durante o primeiro trimestre dos dois anos. Em geral, fevereiro apresentou temperaturas ligeiramente mais altas em comparação aos outros meses, com uma média superior a 22°C.

Rótulos de Linha ▾	Média de TEMPERATURA MÁXIMA NA HORA ANT. (AUT) (°C)	Mín. de TEMPERATURA MÍNIMA NA HORA ANT. (AUT) (°C)
Trim1	23,00	18,00

b. Variação de Temperatura em Diferentes Horários

Além da análise mensal, investigamos a variação de temperatura em horários específicos ao longo dos dias. Utilizamos o horário de **12:00 UTC** (09:00 no horário de Brasília) como exemplo para verificar como a temperatura varia durante esse período específico.

Podemos observar que, em geral, as temperaturas medias ao meio-dia variaram entre 20°C e 28°C ao longo do primeiro trimestre, com uma tendência de aquecimento ao longo do dia.

c. Correlação entre Pressão Atmosférica e Temperatura

Foi calculada a correlação entre a **pressão atmosférica ao nível da estação** e a **temperatura média**. Observou-se uma correlação levemente negativa (-0.25), o que indica que, à medida que a pressão atmosférica aumenta, a temperatura tende a diminuir levemente, sugerindo uma relação inversa.

A análise dos dados meteorológicos de Uberlândia para os anos de 2003 e 2004 permitiu observar padrões significativos nas variações de temperatura ao longo dos meses e horários. A correlação entre pressão atmosférica e temperatura oferece uma visão preliminar sobre as interações entre diferentes fatores climáticos. Essas informações são valiosas para a previsão do tempo e para o planejamento em setores dependentes do clima, como a agricultura e o planejamento urbano.

Estudos futuros podem se aprofundar na análise de anos adicionais ou focar em eventos climáticos extremos para prever melhor os impactos das mudanças climáticas em regiões específicas.