

INSTITUTO PRESBITERIANO MACKENZIE

Autores:

ALLAN MARCONDES DE OLIVEIRA	10922010580
BISMARCK DO NASCIMENTO FERREIRA	10922512349
LEONARDO DOS REIS OLHER	10922510249
PEDRO AUGUSTO ALVES MACENA LIMA	10922509747
WAGNER DE MENDONÇA TRINDADE	10922512667

A INFLUÊNCIA DO CLIMA NAS ENCHENTES

SÃO PAULO / 2023

Autores:

ALLAN MARCONDES DE OLIVEIRA	10922010580
BISMARCK DO NASCIMENTO FERREIRA	10922512349
LEONARDO DOS REIS OLHER	10922510249
PEDRO AUGUSTO ALVES MACENA LIMA	10922509747
WAGNER DE MENDONÇA TRINDADE	10922512667

A INFLUÊNCIA DO CLIMA NAS ENCHENTES

Projeto aplicado I

LEONARDO MASSAYUKI TAKUNO

FABIO SILVA LOPES

SÃO PAULO / 2023

Sumário

Lista de Tabelas	4
Introdução.....	5
Objetivo do projeto	6
Repositório	6
Etapas do projeto	7
CRONOGRAMA DETALHADO PROJETO I:	7
Dataset e Metadados	8
Proposta Analítica do Projeto	9
Análise Exploratória	10
Esboço do Storytelling	15

Lista de Tabelas

CRONOGRAMA DETALHADO PROJETO I

pág. 7

Introdução

Alguns dizem que a mudança climática é a maior ameaça de nossa era, enquanto outros dizem que é um mito baseado em ciência duvidosa. Estamos passando alguns dos dados para você para que você possa formar sua própria visão.

Ainda mais do que com outros conjuntos de dados apresentados pelo Kaggle, há uma enorme quantidade de limpeza e preparação de dados necessária para reunir um estudo de longo prazo das tendências climáticas. Os primeiros dados foram coletados por técnicos usando termômetros de mercúrio, onde qualquer variação no tempo de visita impactava nas medições.

Na década de 1940, a construção de aeroportos fez com que muitas estações meteorológicas fossem movidas. Na década de 1980, houve uma mudança para termômetros eletrônicos que dizem ter um viés de resfriamento.

Dada essa complexidade, há uma série de organizações que coletam dados de tendências climáticas. Os três conjuntos de dados de temperatura terrestre e oceânica mais citados são o MLOST da NOAA, o GISTEMP da NASA e o HadCrut do Reino Unido.

Objetivo do projeto

O objetivo do projeto é realizar uma análise exploratória de dados sobre um tema específico. A análise consiste no levantamento de dados, sobre o tema, que pode ser realizada através de dados públicos da empresa escolhida ou de um banco de Datasets disponíveis para avaliação.

Nosso grupo optou por utilizar um Dataset do banco Kaggle, cujo tema é o Clima, e está disponível neste link: <https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data>.

A origem dos dados é um compilado de informações da Berkeley Earth.

A utilização dos dados é principalmente por instituições de estudo climático e os dados incluem temperatura média global, temperatura máxima e mínima global, temperatura média por país e muitos outros dados meteorológicos.

O período de coleta dos dados é: Maior data: 2013-09-01 00:00:00 - Menor data: 1743-11-0100:00:00.

A escolha deste Dataset foi pela quantidade e qualidade de dados que ele contém e por se tratar de um tema relevante.

Nossa análise será direcionada à um problema comum que afeta diversas áreas do mundo durante o período de grandes chuvas, que é elevada quantidade de água, que aliada aos problemas de infraestrutura, causa diversas enchentes.

Após a definição do Dataset, partimos para a estruturação do nosso Script de exploração dos dados. Estamos utilizando a IDLE Google Colaboratory para essa etapa

Repositório

O link para o repositório do projeto é o:

<https://github.com/olher/ProjetoAplicado1>

Etapas do projeto

CRONOGRAMA DETALHADO PROJETO I:

CRONOGRAMA GERAL PROJETO APLICADO I						
Nome da tarefa	Nome da sub tarefa	Status	Atribuído a:	Início	Término	Duração
Tarefa 1	KICK-OFF	Completo	TODOS	07/mar/23	26/mar/23	19 dias
Subtarefa 1.1	Grupo Whats	Completo	Leonardo dos Reis Olher	07/mar/23	08/mar/23	1 dias
Subtarefa 1.2	Cronograma Inicial	Completo	Bismark do Nascimento Ferreira	07/mar/23	10/mar/23	3 dias
Subtarefa 1.3	Título do trabalho	Completo	Allan Marcondes de Oliveira	13/mar/23	17/mar/23	4 dias
Subtarefa 1.4	Membros do grupo	Completo	Allan Marcondes de Oliveira	13/mar/23	17/mar/23	4 dias
Subtarefa 1.5	Definição da Temática e Escolha DataSet (Apresentação)	Completo	Bismark Ferreira / Leonardo Olher	13/mar/23	17/mar/23	4 dias
Subtarefa 1.6	Elaboração do contexto do estudo	Completo	Allan Marcondes de Oliveira	15/mar/23	22/mar/23	7 dias
Subtarefa 1.7	Criar Github	Completo	Leonardo dos Reis Olher	15/mar/23	22/mar/23	7 dias
Subtarefa 1.8	Estruturar Documento	Completo	Wagner de Mendonça Tindade	15/mar/23	22/mar/23	7 dias
Subtarefa 1.9	Cronograma Geral	Completo	Bismark do Nascimento Ferreira	20/mar/23	24/mar/23	4 dias
Subtarefa 1.10	Referências de aquisição do dataset	Completo	Bismark do Nascimento Ferreira	20/mar/23	24/mar/23	4 dias
Subtarefa 1.11	Descrição da origem	Completo	Pedro Augusto Alves Macena Lima	20/mar/23	24/mar/23	4 dias
Subtarefa 1.12	Descrição do dataset	Completo	Wagner de Mendonça Tindade	20/mar/23	24/mar/23	4 dias
Subtarefa 1.13	Entrega Atividade 1	Completo	Allan Oliveira / Wagner Tindade	25/mar/23	26/mar/23	1 dias
Tarefa 2	DEFINIÇÃO DO PRODUTO ANALÍTICO	Em andamento	TODOS	27/mar/23	16/abr/23	20 dias
Subtarefa 2.1	Reunião Inicial Entrega 2 - DEFINIÇÕES FASE 3 (Sempre às 20h)	não foi iniciado	Todos	27/mar/23	28/mar/23	1 dias
Subtarefa 2.2	Elaboração da proposta analítica Versão 1	Em andamento	Allan Marcondes de Oliveira	27/mar/23	02/abr/23	6 dias
Subtarefa 2.3	Elaboração da proposta analítica Versão 2	Em andamento	Wagner de Mendonça Tindade	27/mar/23	02/abr/23	6 dias
Subtarefa 2.4	Envio de Pré-Proposta no Grupo Whats	Em espera	Allan Oliveira / Wagner Tindade	29/mar/23	01/abr/23	3 dias
Subtarefa 2.5	Estruturação e Elaboração da Proposta Oficial (Validação)	não foi iniciado	Pedro Augusto Alves Macena Lima	29/mar/23	07/abr/23	9 dias
Subtarefa 2.6	Reunião Apresentação Proposta Analítica (Sempre às 20h)	não foi iniciado	Todos	03/abr/23	04/abr/23	1 dias
Subtarefa 2.7	Análise exploratória de dados Parte 1	Em andamento	Bismark do Nascimento Ferreira	03/abr/23	09/abr/23	6 dias
Subtarefa 2.8	Análise exploratória de dados Parte 2	Em andamento	Leonardo dos Reis Olher	03/abr/23	12/abr/23	9 dias
Subtarefa 2.9	Reunião Apresentação Análise Exploratória (Sempre às 20h)	não foi iniciado	Todos	12/abr/23	14/abr/23	2 dias
Subtarefa 2.10	Estruturação e Elaboração do Doc Final	não foi iniciado	Pedro Augusto Alves Macena Lima	10/abr/23	14/abr/23	4 dias
Subtarefa 2.11	Entrega Atividade 2	não foi iniciado	Allan Oliveira / Wagner Tindade	15/abr/23	16/abr/23	1 dias
Tarefa 3	APRESENTAÇÃO DE PRODUTOS	não foi iniciado	TODOS	10/abr/23	07/mai/23	27 dias
Subtarefa 3.1	Reunião Inicial Entrega 3 - DEFINIÇÕES FASE 4 (Sempre às 20h)	não foi iniciado	Todos	10/abr/23	11/abr/23	1 dias
Subtarefa 3.2	Pesquisas: Como apresentar resultados analíticos (Geral)	não foi iniciado	Pedro Augusto Alves Macena Lima	10/abr/23	21/abr/23	11 dias
Subtarefa 3.3	Pesquisa Apresentação de Dados (Gráficos)	não foi iniciado	Allan Marcondes de Oliveira	12/abr/23	19/abr/23	7 dias
Subtarefa 3.4	Pesquisa Apresentação de Dados (Tabelas)	não foi iniciado	Wagner de Mendonça Tindade	12/abr/23	19/abr/23	7 dias
Subtarefa 3.5	Reunião Apresentação Ideias, Gráficos e Tabelas (Sempre às 20h)	não foi iniciado	Todos	24/abr/23	25/abr/23	1 dias
Subtarefa 3.6	Data Storytelling (Focado na Atividade com Python)	não foi iniciado	Leonardo dos Reis Olher	24/abr/23	03/mai/23	9 dias
Subtarefa 3.7	Pesquisa Data Storytelling (Possibilidade)	não foi iniciado	Bismark do Nascimento Ferreira	24/abr/23	03/mai/23	9 dias
Subtarefa 3.8	Estruturação Docs Apresentação	não foi iniciado	Pedro Augusto Alves Macena Lima	24/abr/23	06/mai/23	12 dias
Subtarefa 3.9	Reunião Apresentação Storytelling (Sempre às 20h)	não foi iniciado	Todos	03/mai/23	05/mai/23	2 dias
Subtarefa 3.10	Entrega Atividade 3	não foi iniciado	Leonardo Olher/Bismark Ferreira	06/mai/23	07/mai/23	1 dias
Tarefa 4	ENCERRAMENTO DO PROJETO	não foi iniciado	TODOS	01/mai/23	28/mai/23	27 dias
Subtarefa 4.1	Reunião Inicial Entrega 4 (Sempre às 20h)	não foi iniciado	Todos	01/mai/23	02/mai/23	1 dias
Subtarefa 4.2	Ajustes Finais	não foi iniciado	Em Definição	01/mai/23	14/mai/23	13 dias
Subtarefa 4.3	Reunião Apresentação Ajustes Finais (Sempre às 20h)	não foi iniciado	Todos	15/mai/23	16/mai/23	1 dias
Subtarefa 4.4	Apresentações	não foi iniciado	Em Definição	10/mai/23	22/mai/23	12 dias
Subtarefa 4.5	Reunião Apresentação Trabalho Final (Sempre às 20h)	não foi iniciado	Todos	22/mai/23	26/mai/23	4 dias

*Esse cronograma será avaliado e atualizado semanalmente.

**Algumas atividades poderão ser inseridas e outras retiradas.

***O detalhamento da Tarefa 4 será verificado na reunião de início da Entrega 3

Dataset e Metadados

O Dataset é composto por 7 colunas sendo elas: dt, AverageTempetarure, AverageTempetarureUncertainty, City, Country, Latitude, Longitude.

O Dataset possui 1,6 bilhão de relatórios de temperatura, sendo elas terrestres e oceânicas. O período de coleta dos dados é: Maior data: 2013-09-01 00:00:00 - Menor data: 1743-11-0100:00:00.

Proposta Analítica do projeto

O dataset escolhido para este projeto é oriundo da Berkeley Earth, uma instituição sem fins lucrativos e independente, com sede na Califórnia e focada em análise de dados de temperatura terrestre para fomentar a ciência climática.

Fundada em 2010 por Richard Muller e Elizabeth Muller, a organização também possui grandes doadores como o Laboratório Nacional Lawrence Berkeley, Fundação Charles G. Koch, FICER entre outros.

Os achados preliminares da organização, bem como o dataset e programas foram publicados no final de 2012. O estudo buscou sanar preocupações científicas incluindo o efeito da ilha de calor em áreas urbanas, por exemplo. A conclusão do Berkeley Earth Group foi que a tendência de aquecimento é real, e ao longo dos últimos 50 anos a superfície terrestre aqueceu em 0.91 ± 0.05 °C.

Essa análise nos permite cruzar as informações do estudo com problemas previamente associados à esse fenômeno. Um desses grandes problemas é a questão das enchentes, o qual será o foco de nossa análise exploratória.

Os metadados estão contidos em um Dataset disponível no Kaggle, acessível através deste link: <https://www.kaggle.com/datasets/berkeleyearth/climate-change-earth-surface-temperature-data>.

Iniciamos nossa análise exploratória resumindo algumas informações do DataSet, comparando alguns dados de médias de temperatura ao longo dos anos, separado por determinados conjuntos como países e continentes.

O Pipeline de dados parte do acesso e download do Dataset para iniciar a análise. As bibliotecas utilizadas foram: pandas, numpy, matplotlib, seaborn e warnings. Após importação das bibliotecas, partimos para importação do Dataset e inciamos a análise com os dados que precisam de tratamento.

O tratamento de dados segue nessa ordem:

- Conversão tipo coluna data de string para datetime;
- Remover valores nulos do DataFrame;
- Criar coluna: Decada;
- Unir dados de continentes;
- Criar DataFrame Decada com media, maxima e minima temperatura por década;
- Criar DataFrame Continente.

Após o tratamento, seguimos para análise exploratória dos dados com a geração de alguns gráficos.

Análise Exploratória

Imports:

Bibliotecas para análise exploratória dos dados de temperatura.

```
In [1]: import pandas as pd # manipular dataframe
import numpy as np # manipular arrays
import matplotlib.pyplot as plt # criar graficos
import seaborn as sns # criar graficos
from statistics import mode # estatística: moda
import warnings; warnings.filterwarnings('ignore') # remover avisos
```

Importar base de dados

df recebe dados de temperatura por cidade (GlobalLandTemperaturesByCity.csv);
countinents recebe continentes de cada país (Continents.csv).

```
In [13]: df = pd.read_csv('Datasets/GlobalLandTemperaturesByCity.csv')
countinents = pd.read_csv('Datasets/Continents.csv', encoding='latin-1')
df.head(10)
```

```
Out[13]:
```

	dt	AverageTemperature	AverageTemperatureUncertainty	City	Country	Latitude	Longitude
0	1743-11-01	6.068	1.737	Århus	Denmark	57.05N	10.33E
1	1743-12-01	NaN	NaN	Århus	Denmark	57.05N	10.33E
2	1744-01-01	NaN	NaN	Århus	Denmark	57.05N	10.33E
3	1744-02-01	NaN	NaN	Århus	Denmark	57.05N	10.33E
4	1744-03-01	NaN	NaN	Århus	Denmark	57.05N	10.33E
5	1744-04-01	5.788	3.624	Århus	Denmark	57.05N	10.33E
6	1744-05-01	10.644	1.283	Århus	Denmark	57.05N	10.33E
7	1744-06-01	14.051	1.347	Århus	Denmark	57.05N	10.33E
8	1744-07-01	16.082	1.396	Århus	Denmark	57.05N	10.33E
9	1744-08-01	NaN	NaN	Århus	Denmark	57.05N	10.33E

Verificação de dados que necessita tratamentos

- Dados nulos;
- Tamanho do dataframe;
- Tipo das colunas;
- Nomes dos países.

```
In [3]: print(df.isnull().sum(), end='\n\n')
print(df.shape, end='\n\n')
print(df.info(), end='\n\n')
print(df['Country'].unique())
```

```
dt                                0
AverageTemperature               364130
AverageTemperatureUncertainty    364130
City                             0
Country                          0
Latitude                         0
Longitude                        0
dtype: int64
```

```
(8599212, 7)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8599212 entries, 0 to 8599211
Data columns (total 7 columns):
#   Column                                Dtype
---  -
0   dt                                    object
1   AverageTemperature                  float64
2   AverageTemperatureUncertainty      float64
3   City                               object
4   Country                            object
5   Latitude                           object
6   Longitude                          object
dtypes: float64(2), object(5)
memory usage: 459.2+ MB
None
```

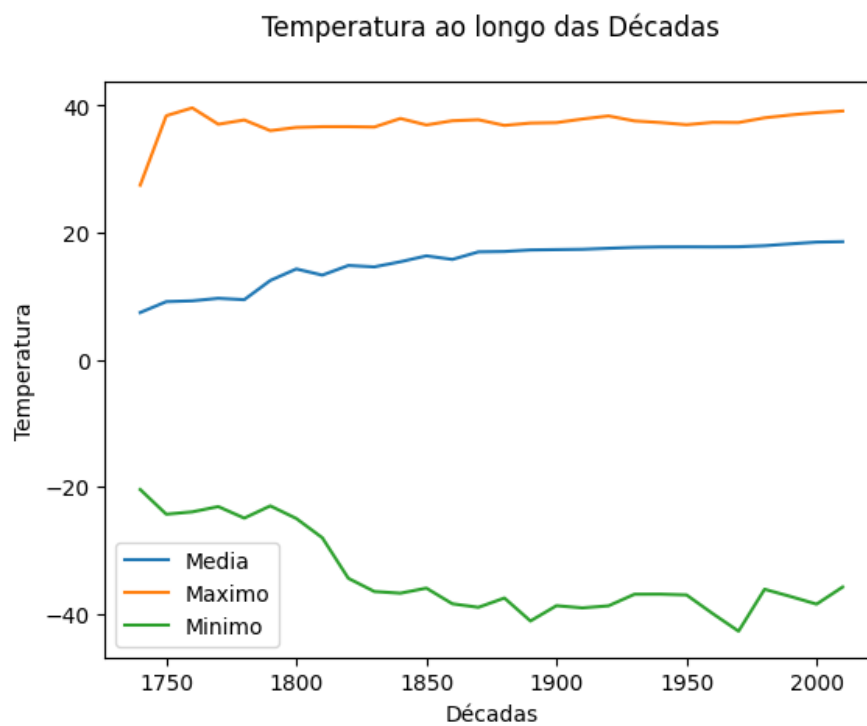
```
['Denmark' 'Turkey' 'Kazakhstan' 'China' 'Spain' 'Germany' 'Nigeria'
'Iran' 'Russia' 'Canada' 'Côte D'Ivoire' 'United Kingdom' 'Saudi Arabia'
'Japan' 'United States' 'India' 'Benin' 'United Arab Emirates' 'Mexico'
'Venezuela' 'Ghana' 'Ethiopia' 'Australia' 'Yemen' 'Indonesia' 'Morocco'
'Pakistan' 'France' 'Libya' 'Burma' 'Brazil' 'South Africa' 'Syria'
'Egypt' 'Algeria' 'Netherlands' 'Malaysia' 'Portugal' 'Ecuador' 'Italy'
'Uzbekistan' 'Philippines' 'Madagascar' 'Chile' 'Belgium' 'El Salvador'
'Romania' 'Peru' 'Colombia' 'Tanzania' 'Tunisia' 'Turkmenistan' 'Israel'
'Eritrea' 'Paraguay' 'Greece' 'New Zealand' 'Vietnam' 'Cameroon' 'Iraq'
'Afghanistan' 'Argentina' 'Azerbaijan' 'Moldova' 'Mali'
'Congo (Democratic Republic Of The)' 'Thailand'
'Central African Republic' 'Bosnia And Herzegovina' 'Bangladesh'
'Switzerland' 'Equatorial Guinea' 'Cuba' 'Lebanon' 'Mozambique' 'Serbia'
'Angola' 'Somalia' 'Norway' 'Nepal' 'Poland' 'Ukraine' 'Guinea Bissau'
'Malawi' 'Burkina Faso' 'Slovakia' 'Congo' 'Belarus' 'Gambia'
'Czech Republic' 'Hungary' 'Burundi' 'Zimbabwe' 'Bulgaria' 'Haiti'
'Puerto Rico' 'Sri Lanka' 'Nicaragua' 'Zambia' 'Honduras' 'Taiwan'
'Bolivia' 'Guinea' 'Ireland' 'Senegal' 'Latvia' 'Qatar' 'Albania'
'Tajikistan' 'Kenya' 'Guatemala' 'Finland' 'Sierra Leone' 'Sweden'
'Botswana' 'Guyana' 'Austria' 'Uganda' 'Armenia' 'Dominican Republic'
'Jordan' 'Djibouti' 'Sudan' 'Lithuania' 'Rwanda' 'Jamaica' 'Togo'
'Macedonia' 'Cyprus' 'Gabon' 'Slovenia' 'Bahrain' 'Swaziland' 'Niger'
'Lesotho' 'Liberia' 'Uruguay' 'Chad' 'Bahamas' 'Mauritania' 'Panama'
'Suriname' 'Cambodia' 'Montenegro' 'Mauritius' 'Papua New Guinea'
'Iceland' 'Croatia' 'Reunion' 'Oman' 'Costa Rica' 'South Korea'
'Hong Kong' 'Singapore' 'Estonia' 'Georgia' 'Mongolia' 'Laos' 'Namibia']
```

Exemplos de Análise

Análise de Evolução da Temperatura por Década

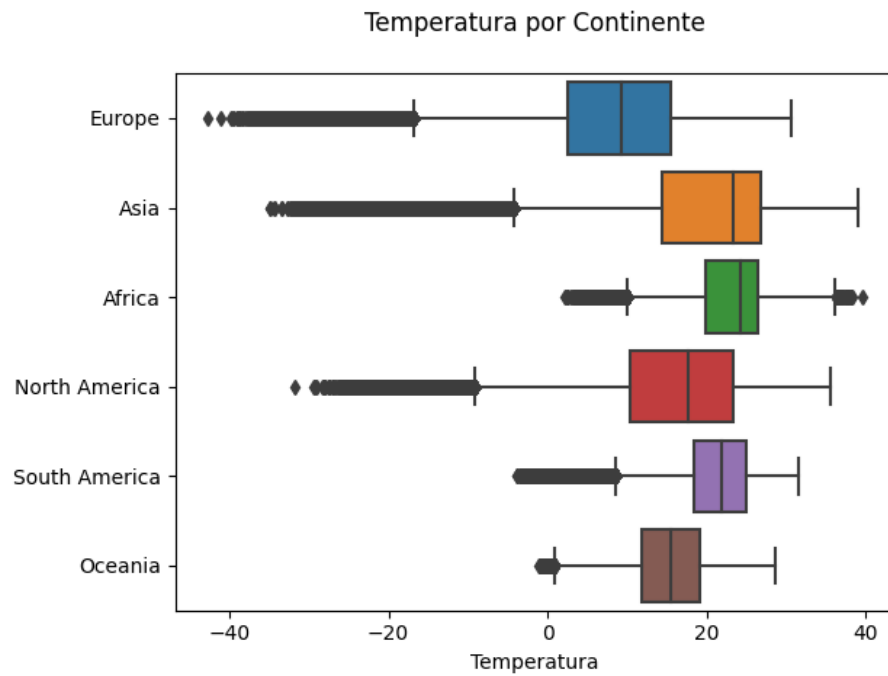
```
In [5]: # lineplot:
sns.lineplot(x=decada[decada.columns[0]], y=decada[decada.columns[1]], label='Media')
sns.lineplot(x=decada[decada.columns[0]], y=decada[decada.columns[2]], label='Maximo')
sns.lineplot(x=decada[decada.columns[0]], y=decada[decada.columns[3]], label='Minimo')
plt.title('Temperatura ao longo das Décadas\n')
plt.ylabel('Temperatura')
plt.xlabel('Décadas')

plt.show()
```



BoxPlot - Continentes

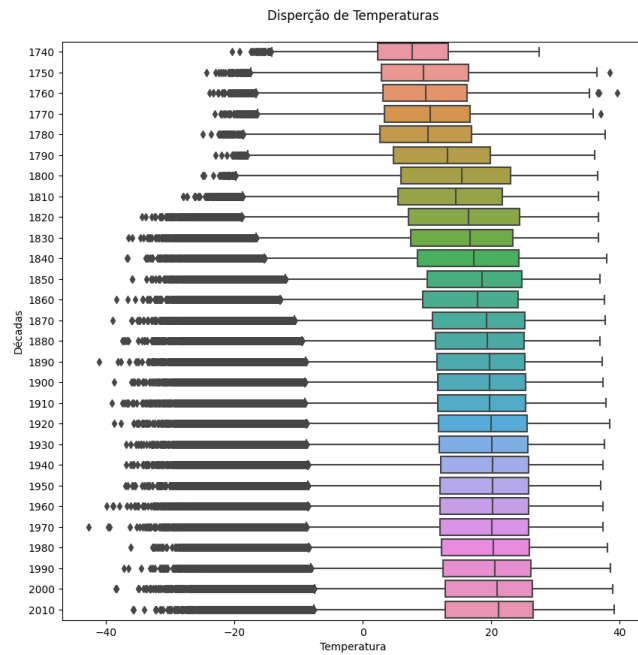
```
In [8]: sns.boxplot(x=df['AverageTemperature'], y=df['Continents'], orient='h') # boxplot na horizontal  
plt.xlabel('Temperatura')  
plt.ylabel('')  
plt.title('Temperatura por Continente\n')  
plt.show()
```



Análise de Outliers

```
In [10]: fig, ax = plt.subplots(figsize=(10, 10))
sns.boxplot(x=df['AverageTemperature'], y=df['decada'], orient='h') # boxplot na horizontal

plt.title('Disperção de Temperaturas\n')
plt.xlabel('Temperatura')
plt.ylabel('Décadas')
plt.show()
```



Esboço do Storytelling

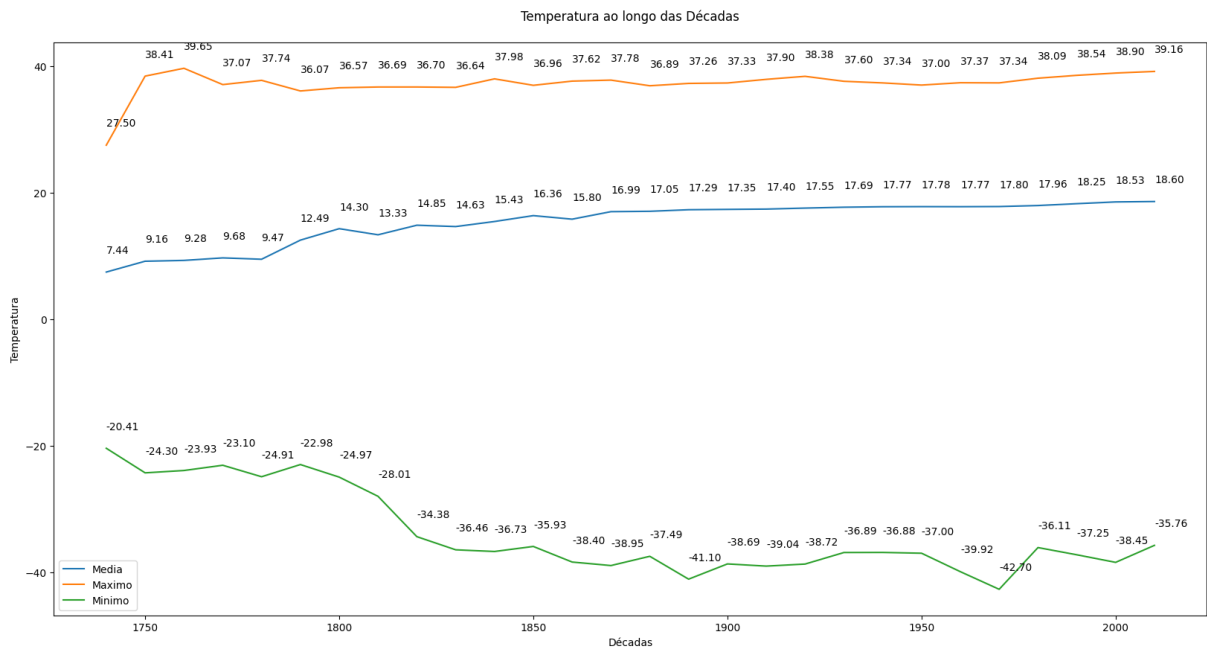
No ano de 1740, as pessoas viviam sem preocupações com o clima e as mudanças que estavam por vir. Naquela época, a Terra era um lugar muito menos poluído, e as atividades humanas ainda não tinham impactado o clima global. Podemos destacar que mundo se encontrava em pré-revolução industrial.



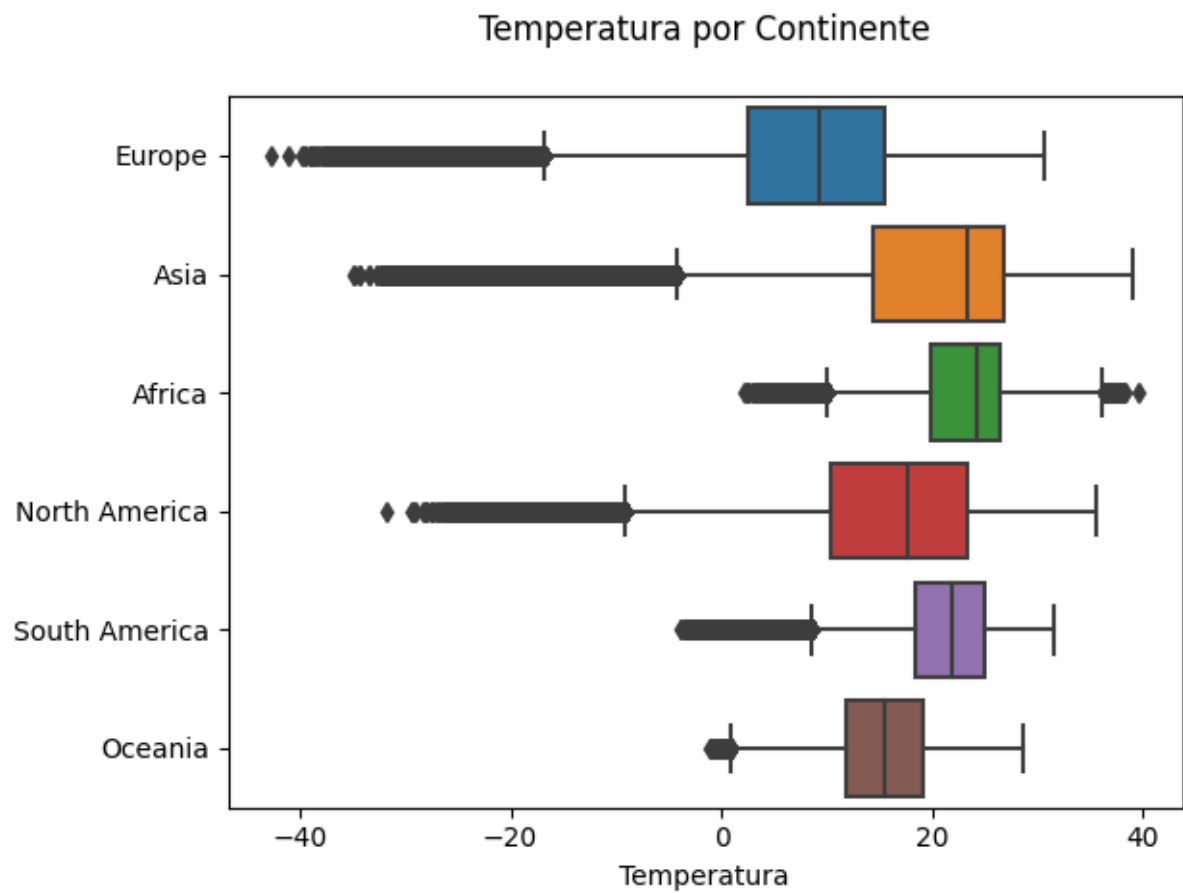
No entanto, com o passar do tempo, as coisas começaram a mudar. A temperatura média do planeta começou a subir. O clima tornou-se mais quente e instável, com ondas de calor cada vez mais frequentes, secas prolongadas e enchentes catastróficas. A biodiversidade começou a ser afetada, e muitas espécies de animais e plantas começaram a desaparecer.



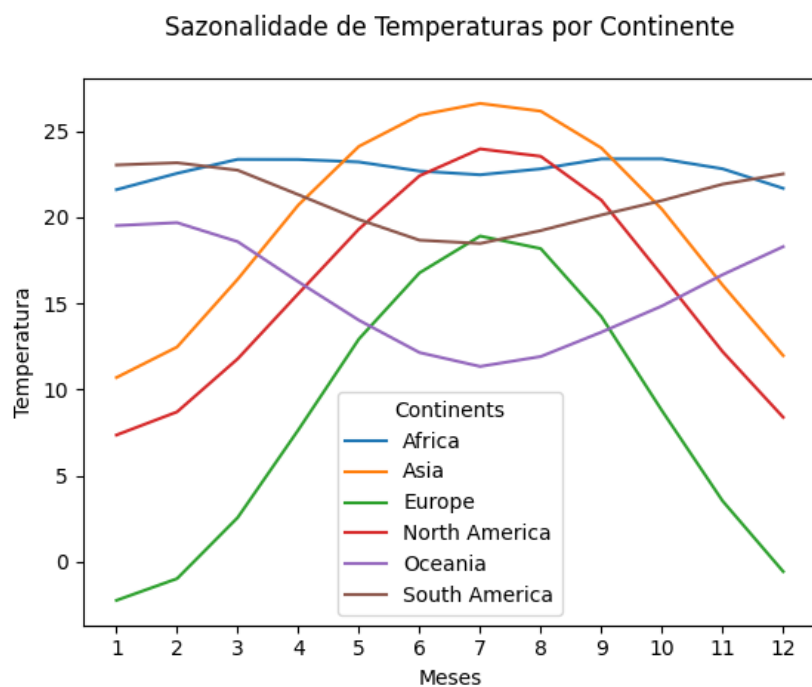
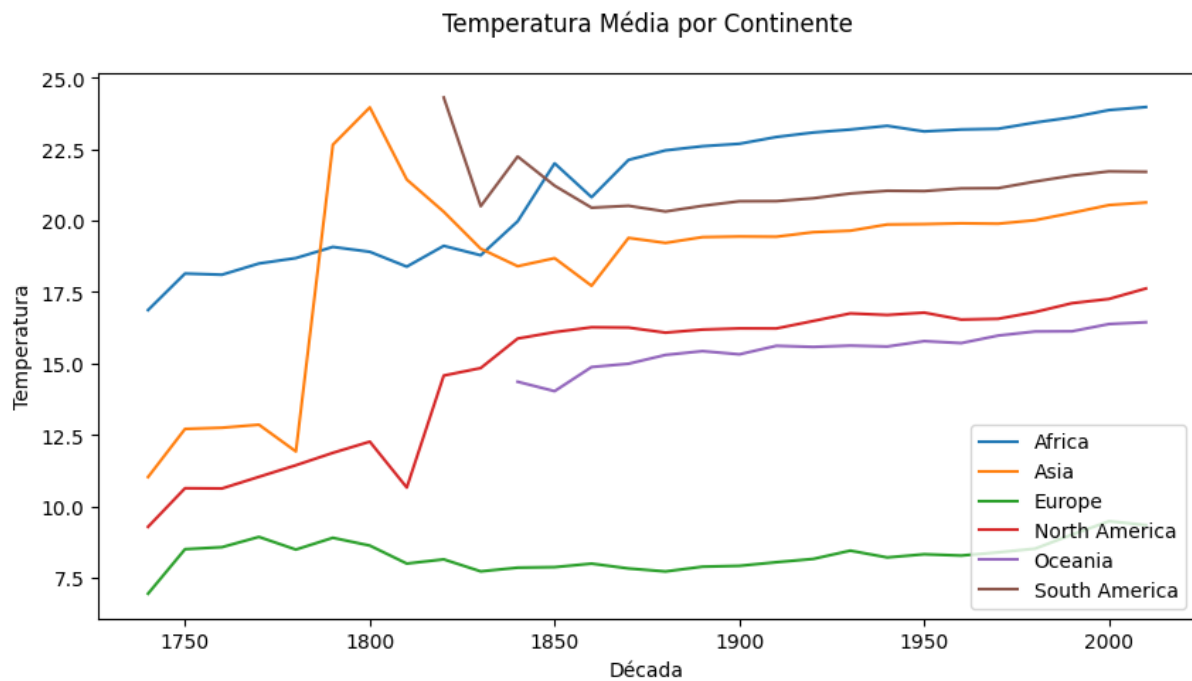
A análise climática mostrou que houve um aumento da temperatura média no mundo ao longo das décadas. Observou-se também que, a temperatura mínima tem caído. Isso pode parecer contraditório, mas é uma consequência do aquecimento global. O aumento da temperatura média pode criar condições climáticas extremas, com variações mais amplas entre a temperatura mínima e a máxima em um determinado período.



A variação de temperatura, também conhecida como amplitude térmica, é a diferença entre a temperatura máxima e mínima registrada em um determinado período. Um aumento na amplitude térmica pode levar a uma maior frequência de eventos climáticos extremos, como ondas de calor, frio intenso, tempestades e outras condições climáticas extremas. Podemos verificar a quantidade de temperaturas registradas menor que Q1 na Europa, Ásia e América do Norte o que mostra o quanto varia a temperatura média nesses continentes.



Diferentes continentes, países e setores contribuíram para essas mudanças climáticas, e os tornaram-se um problema global que afetou a todos. Como podemos ver no gráfico, a partir do século XX todos os continentes passaram por um processo de aumento nos registros médios de temperatura década após década.



No entanto, há esperança. As pessoas estão cada vez mais conscientes dos efeitos das mudanças climáticas e da necessidade de agir. As pessoas podem fazer a diferença, seja através de escolhas cotidianas, ou até pressão por mudanças em políticas públicas. Todos nós podemos contribuir para um mundo mais sustentável e saudável, se estivermos dispostos a agir juntos e rapidamente.