

Analise exploratoria do consumo de energia elétrica no Brasil

O estudo tem por objetivo analisar o consumo de energia elétrica no Brasil ao longo dos anos, identificando padrões, variações regionais e tipos de consumo, com base em dados oficiais.

Importação de Bibliotecas

Foram utilizadas bibliotecas amplamente aceitas para análise de dados (pandas, numpy) e visualização (matplotlib, seaborn). Além disso, o notebook utiliza funções personalizadas criadas nos módulos visualization, grafico e cleaning.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import sys
sys.path.append('../')

# Importação biblioteca de funções
from src.data_visualization.visualization import *
from src.data_visualization.grafico import *
from src.data_cleaning.cleaning import drop_duplicados, limpa_dados_NAN
```

✓ 1.6s

Carregamento dos dados

```
df_consumo_energia = pd.read_csv("../data/consumo_energia_eletrica.csv", sep=',')
df_estados = pd.read_csv("../data/estado_regiao.csv", sep=';', encoding='ISO-8859-1')
```

Dataset com dados do IBGE coletado no endereço
("https://www.kaggle.com/datasets/crisparada/brazilian-cities/versions/7?
resource=download&select=BRAZIL_CITIES.csv")

```
tipo_encoding("../data/BRAZIL_CITIES.csv")
✓ 0.0s
{'encoding': 'utf-8', 'confidence': 0.99, 'language': ''}

df_pnad = pd.read_csv("../data/BRAZIL_CITIES.csv", sep=';')
✓ 0.0s
```

- União dos Datasets

União dos dados dataset (Consumo de energia e estados)

```
df_final = pd.merge(df_consumo_energia, df_estados, left_on='sigla_uf', right_on='sigla')
```

✓ 0.0s

- exclusão coluna com redundância

```
df_final.drop(columns='sigla', inplace=True)
```

✓ 0.0s

```
df_final.head(10)
```

✓ 0.0s

	ano	mes	sigla_uf	tipo_consumo	numero_consumidores	consumo	id_estado	estado	regiao	pais
0	2004	1	TO	Total	NaN	65876	27	Tocantins	Norte	Brasil
1	2004	1	BA	Total	NaN	1444451	5	Bahia	Nordeste	Brasil
2	2004	1	PR	Total	NaN	1596274	16	Paraná	Sul	Brasil
3	2004	1	RS	Total	NaN	1780912	21	Rio Grande do Sul	Sul	Brasil
4	2004	1	GO	Total	NaN	630624	9	Goiás	Centro-Oeste	Brasil
5	2004	1	MA	Total	NaN	737033	10	Maranhão	Nordeste	Brasil

União dos dados do df_final com os dados do dataset do IBGE, que foram coletado no endereço ("https://www.kaggle.com/datasets/crisparada/brazilian-cities/versions/7?resource=download&select=BRAZIL_CITIES.csv")

```
# Agregar consumo por ano e UF - dataset "df_final"
consumo_agg = df_final.groupby(['ano', 'sigla_uf']).agg({
    'numero_consumidores': 'sum',
    'consumo': 'sum'
}).reset_index()
```

✓ 0.0s

Python

- Irei mesclar o dataset df_final com o df_pnad

```
df_final_pnad = pd.merge(consumo_agg, brazil_cities_agg, on='sigla_uf', how='inner')
```

✓ 0.0s

Python

```
df_final_pnad.head()
```

✓ 0.0s

Python

	ano	sigla_uf	numero_consumidores	consumo	IDHM	IDHM_Renda	IDHM_Longevidade	IDHM_Educacao	IBGE_POP	GDP	GDP_CAPITA	GVA_AGROPEC	GVA_INDUSTRY	GVA_SERVICES	GVA_PU
0	2004	AL	7758590.0	3390490	0.721000	0.739000	0.799000	0.635	867156.0	21306115.95	20853.410000	164315.02	2922193.90	12005071.98	327810
1	2004	BA	42981158.0	17248063	0.726333	0.725667	0.820667	0.647	2624898.0	78069608.12	19787.366667	138925.95	10963732.55	39511247.87	1019788
2	2004	CE	26342055.0	6328325	0.690000	0.644000	0.810000	0.642	239172.0	4185791.73	15604.190000	16226.81	446.85	2500749.36	78327

Criação de Funções de Visualização e Análise Gráfica

A estrutura do projeto foi organizada em módulos reutilizáveis, permitindo separar responsabilidades e tornar o código mais limpo, reutilizável e fácil de manter.

Funções de Limpeza de Dados - cleaning.py Esses métodos foram utilizados para garantir a integridade dos dados, antes de qualquer análise ou visualização. Remove duplicatas do DataFrame original inplace.

- def drop_duplicados(dataframe: pd.DataFrame) -> pd.DataFrame: Remove todas as linhas com valores ausentes (NaN).
- def limpa_dados_NAN(dataframe: pd.DataFrame) -> pd.DataFrame:

Diagnóstico de Dados e Outliers - visualization.py Ferramentas de apoio para, Visualizar registros duplicados ou ausentes, Verificar codificação de arquivos, Detectar outliers com base no método IQR

```
def visuliza_duplicados(dataframe: pd.DataFrame, subset=None) def visuliza_dados_NaN(dataframe: pd.DataFrame) def tipo_encoding(caminho) def detectar_outliers_iqr(df, coluna, ano)
```

```
Visualizações Personalizadas def grafico_consumo_por_ano_estados(df, ano) def
grafico_consumo_por_estado(df, sigla_uf) def grafico_consumo_mensal_por_estado(df, sigla_uf, ano)
def consumidores_por_regiao_ano(df, ano_escolhido) def
correlacao_consumo_consumidores_por_estado(df, estado_escolhido) def
detectar_outliers_consumo_iqr(df) def plotar_outliers_boxplot(df, estado, ano) def
top10_consumo_estados_ano(df, ano_escolhido) def mapa_calor_correlacao(df)
```

- Este módulo agrupa todas as visualizações utilizadas no projeto, com foco em (Comparações regionais e temporais, Correlação entre variáveis, Detecção visual de outliers, Evolução e ranking de consumo)

Limpeza e tratamento dos dados

- Verificação do formato:

```
df_final.shape
```

✓ 0.0s

```
(39897, 10)
```

```
df_final.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 39897 entries, 0 to 39896
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
--  --
0   ano                    39897 non-null  int64
1   mes                    39897 non-null  int64
2   sigla_uf                39897 non-null  object
3   tipo_consumo            39897 non-null  object
4   numero_consumidores     26937 non-null  float64
5   consumo                 39897 non-null  int64
6   id_estado                39897 non-null  int64
7   estado                  39897 non-null  object
8   regiao                  39897 non-null  object
9   pais                    39897 non-null  object
dtypes: float64(1), int64(4), object(5)
```

```
df_pnad = pd.read_csv("../data/BRAZIL_CITIES.csv", sep=';')
```

✓ 0.0s

```
df_pnad.shape
```

✓ 0.0s

```
(5573, 81)
```

```
df_pnad.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5573 entries, 0 to 5572
Data columns (total 81 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   CITY                    5573 non-null  object
1   STATE                   5573 non-null  object
2   CAPITAL                 5573 non-null  int64
3   IBGE_RES_POP            5565 non-null  float64
```

• Remoção de dados ausentes

- De acordo com as instruções dos enunciado da atividade

Instruções para correção de dados ausentes

Exclua todos os dados ausentes.

```
# importação e utilização de função importada de modulo
limpa_dados_NAN(df_final)
```

✓ 0.0s

	ano	mes	sigla_uf	tipo_consumo	numero_consumidores	consumo	id_estado	estado	regiao	pais
648	2004	1	RN	Outros	40857.0	69617	20	Rio Grande do Norte	Nordeste	Brasil
649	2004	1	SP	Outros	311650.0	937538	25	São Paulo	Sudeste	Brasil
650	2004	1	MS	Outros	56881.0	67601	12	Mato Grosso do Sul	Centro-Oeste	Brasil
651	2004	1	SC	Outros	226165.0	209380	24	Santa Catarina	Sul	Brasil
652	2004	1	RJ	Outros	70634.0	416128	19	Rio de Janeiro	Sudeste	Brasil
...
39892	2023	12	BA	Residencial	5932740.0	718461	5	Bahia	Nordeste	Brasil
39893	2023	12	PA	Residencial	2614106.0	478726	14	Pará	Norte	Brasil
39894	2023	12	AC	Residencial	245699.0	58301	1	Acre	Norte	Brasil

```
# irei utilizar a mesma regra anterior irei excluir os dados faltantes
limpa_dados_NAN(df_pnad)
```

✓ 0.0s

	CITY	STATE	CAPITAL	IBGE_RES_POP	IBGE_RES_POP_BRAS	IBGE_RES_POP_ESTR	IBGE_DU	IBGE_DU_URBAN	IB
631	Blumenau	SC	0	309011.0	308284.0	727.0	101157.0	96961.0	
946	Campina Grande	PB	0	385213.0	385132.0	81.0	112024.0	107034.0	
951	Campinas	SP	0	1080113.0	1075724.0	4389.0	348424.0	343169.0	
971	Campo Grande	MS	1	786797.0	785017.0	1780.0	250542.0	247159.0	
1151	Caruaru	PE	0	314912.0	314841.0	71.0	96393.0	85887.0	
1205	Caxias Do Sul	RS	0	435564.0	434832.0	732.0	146895.0	141700.0	
1254	Chapecó	SC	0	183530.0	183447.0	83.0	58552.0	54542.0	
	Feira De								

visuliza_dados_NaN(df_pnad)

✓ 0.0s

Foram detectado CITY0

STATE0

CAPITAL0

IBGE_RES_POP8

IBGE_RES_POP_BRAS8

...

Wheeled_tractor11

UBER5448

MAC5407

WAL-MART5471

POST_OFFICES120

Length: 81, dtype: int64 dados NaN

	CITY	STATE	CAPITAL	IBGE_RES_POP	IBGE_RES_POP_BRAS	IBGE_RES_POP_ESTR	IBGE_DU	IBGE_D
0	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	

irei utilizar a mesma regra anterior irei excluir os dados faltantes

limpa_dados_NaN(df_pnad)

✓ 0.0s

	CITY	STATE	CAPITAL	IBGE_RES_POP	IBGE_RES_POP_BRAS	IBGE_RES_POP_ESTR	IBGE_DU	IBGE_D
631	Blumenau	SC	0	309011.0	308284.0	727.0	101157.0	
946	Campina Grande	PB	0	385213.0	385132.0	81.0	112024.0	
951	Campinas	SP	0	1080113.0	1075724.0	4389.0	348424.0	
971	Campo Grande	MS	1	786797.0	785017.0	1780.0	250542.0	
1151	Caruaru	PE	0	314912.0	314841.0	71.0	96393.0	
1205	Caxias Do Sul	RS	0	435564.0	434832.0	732.0	146895.0	
1254	Chapecó	SC	0	183530.0	183447.0	83.0	58552.0	

EMS 1 OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER

- Verificação e remoção de duplicatas:

Visualização

- Tratamento dos dados duplicados

```
visuliza_duplicados(df_final, subset=['consumo', 'ano', 'mes', 'tipo_consumo'])
```

✓ 0.0s

Existem 1017 registros duplicados.

	ano	mes	sigla_uf	tipo_consumo	numero_consumidores	consumo	id_estado	estado	regiao	pais
38880	2023	11	MS	Outros	99710.0	122350	12	Mato Grosso do Sul	Centro-Oeste	Brasil
38881	2023	11	RN	Outros	77359.0	112767	20	Rio Grande do Norte	Nordeste	Brasil
38882	2023	11	PE	Outros	157848.0	256509	17	Pernambuco	Nordeste	Brasil
38883	2023	11	ES	Outros	234985.0	180315	8	Espírito Santo	Sudeste	Brasil
38884	2023	11	AC	Outros	23068.0	26323	1	Acre	Norte	Brasil
...
39892	2023	12	BA	Residencial	5932740.0	718461	5	Bahia	Nordeste	Brasil
39893	2023	12	PA	Residencial	2614106.0	478726	14	Pará	Norte	Brasil
39894	2023	12	AC	Residencial	245699.0	58301	1	Acre	Norte	Brasil
39895	2023	12	RN	Residencial	1390080.0	210119	20	Rio Grande do Norte	Nordeste	Brasil
39896	2023	12	SC	Residencial	2719791.0	655389	24	Santa Catarina	Sul	Brasil

Remoção - Foi optado por realizar a exclusão dos dados duplicado, assim obdecendo os mesmo critério utilizado no tratamento dos dados ausentes

```
# Irei optar por realizar exclusão dos dados duplicados
drop_duplicados(df_final)
```

✓ 0.0s

	ano	mes	sigla_uf	tipo_consumo	numero_consumidores	consumo	id_estado	estado	regiao	pais
648	2004	1	RN	Outros	40857.0	69617	20	Rio Grande do Norte	Nordeste	Brasil
649	2004	1	SP	Outros	311650.0	937538	25	São Paulo	Sudeste	Brasil
650	2004	1	MS	Outros	56881.0	67601	12	Mato Grosso do Sul	Centro-Oeste	Brasil
651	2004	1	SC	Outros	226165.0	209380	24	Santa Catarina	Sul	Brasil
652	2004	1	RJ	Outros	70634.0	416128	19	Rio de Janeiro	Sudeste	Brasil
...
38875	2023	12	BA	Residencial	5932740.0	718461	5	Bahia	Nordeste	Brasil
38876	2023	12	PA	Residencial	2614106.0	478726	14	Pará	Norte	Brasil
38877	2023	12	AC	Residencial	245699.0	58301	1	Acre	Norte	Brasil
38878	2023	12	RN	Residencial	1390080.0	210119	20	Rio Grande do Norte	Nordeste	Brasil
38879	2023	12	SC	Residencial	2719791.0	655389	24	Santa Catarina	Sul	Brasil

25920 rows × 10 columns

Análises Iniciais

- Distribuição por ano:

```
agrupamento_ano = df_final['ano'].value_counts()
agrupamento_ano
```

✓ 0.0s

ano	
2004	1296
2005	1296
2022	1296
2021	1296
2020	1296
2019	1296
2018	1296
2017	1296
2016	1296
2015	1296
2014	1296
2013	1296
2012	1296
2011	1296
2010	1296
2009	1296
2008	1296

- Distribuição por tipo de consumo:

```
agrupamento_consumo = df_final['tipo_consumo'].value_counts()
agrupamento_consumo
```

✓ 0.0s

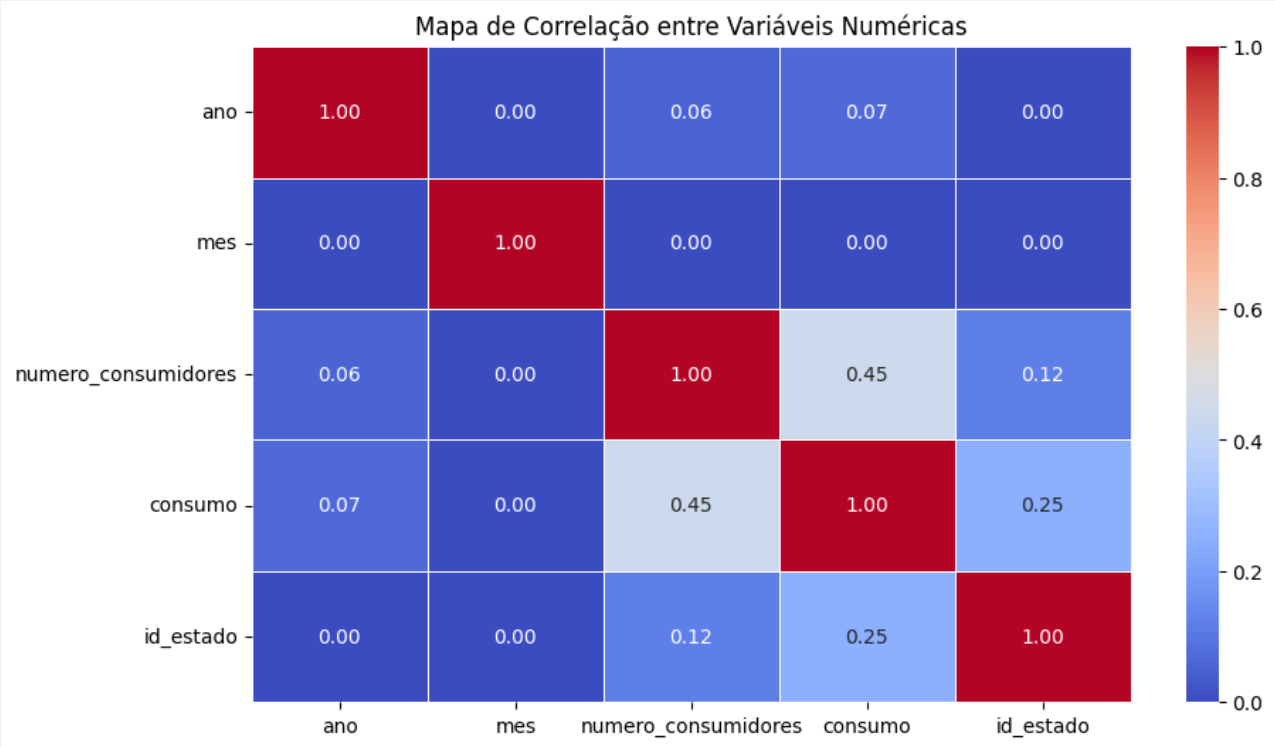
tipo_consumo	
Outros	6480
Comercial	6480
Industrial	6480
Residencial	6480

Name: count, dtype: int64

- Correlação entre Variáveis - Avaliação da relação entre variáveis numéricas, especialmente entre consumo e outras dimensões temporais ou regionais:


```
correlacao = df_final.corr(numeric_only=True)
plt.figure(figsize=(10, 6))
sns.heatmap(correlacao, annot=True, fmt=".2f", cmap='coolwarm', linewidths=0.5)
plt.title('Mapa de Correlação entre Variáveis Numéricas')
plt.show()
```

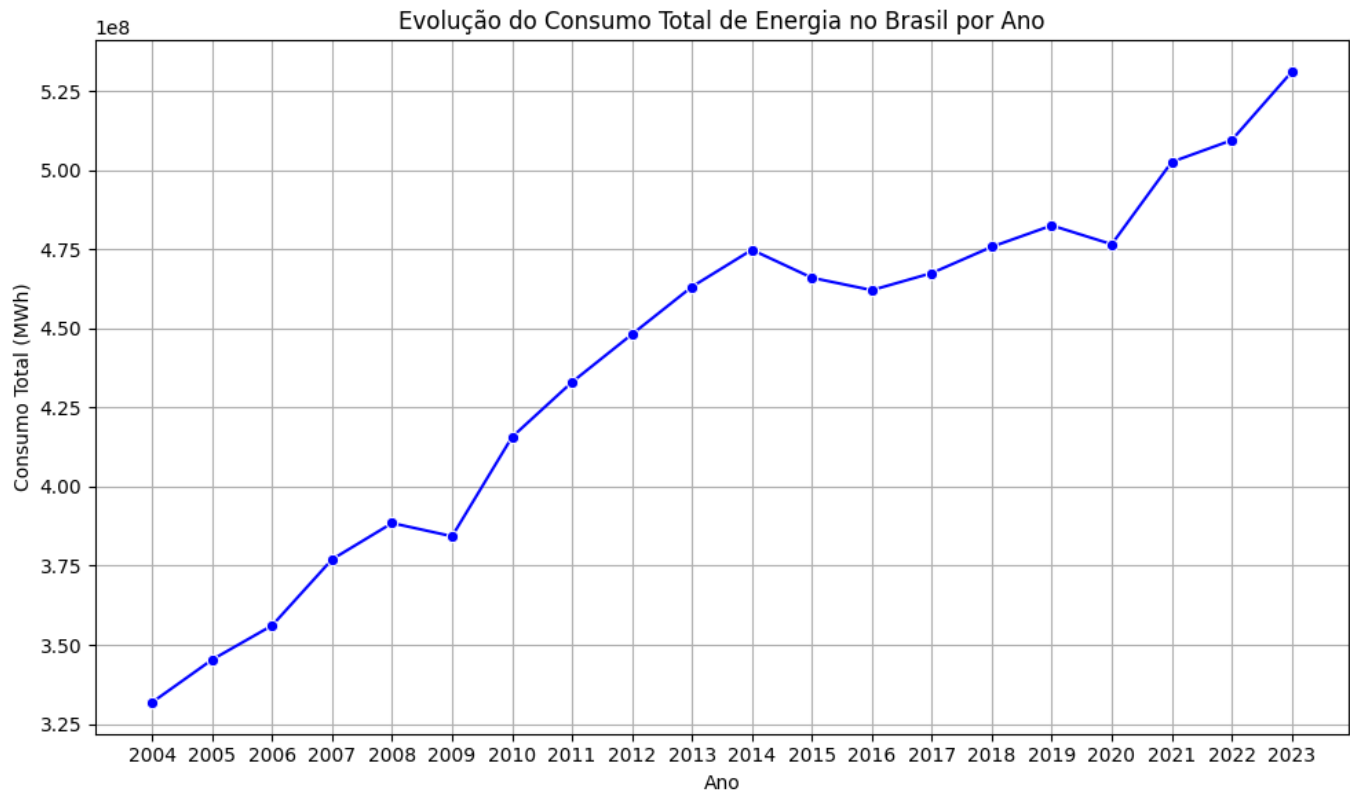
✓ 0.1s



Demonstrando consumo está correlacionado com o número de consumidores, como esperado. Outras correlações menores podem indicar sazonalidades ou padrões regionais.

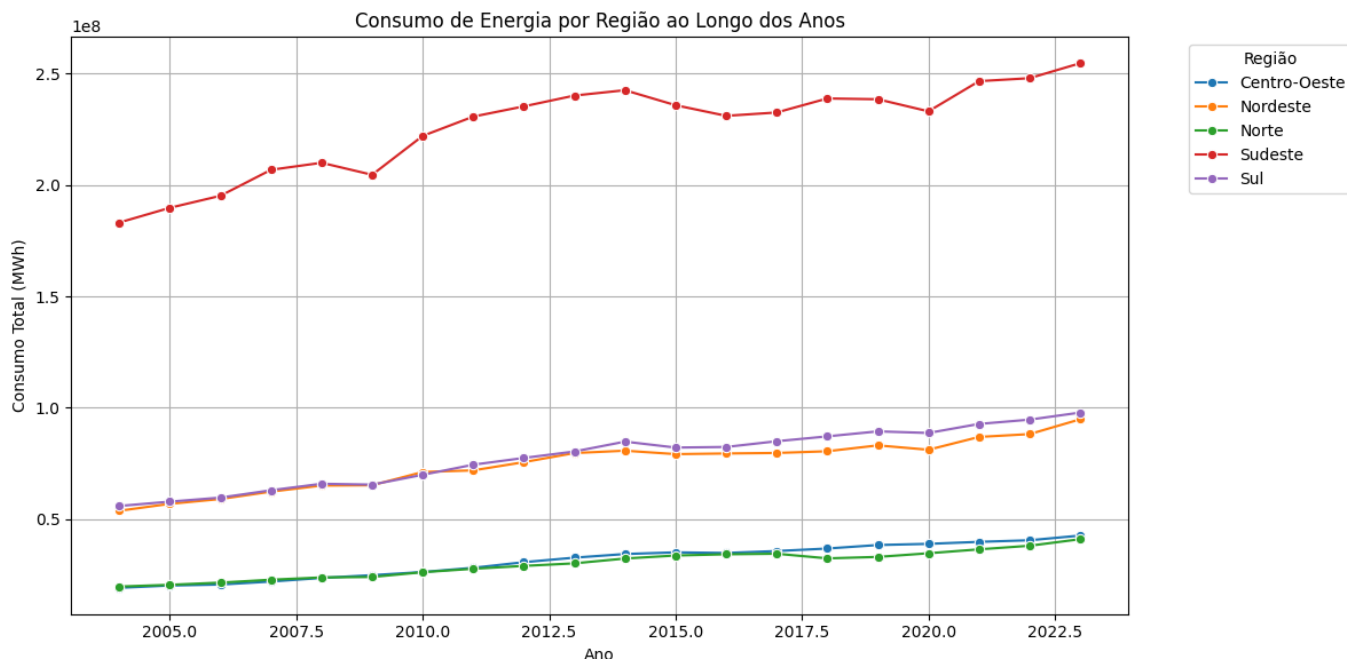
Síntese dos Principais Insights

1º Evolução do consumo total por ano (nacional).



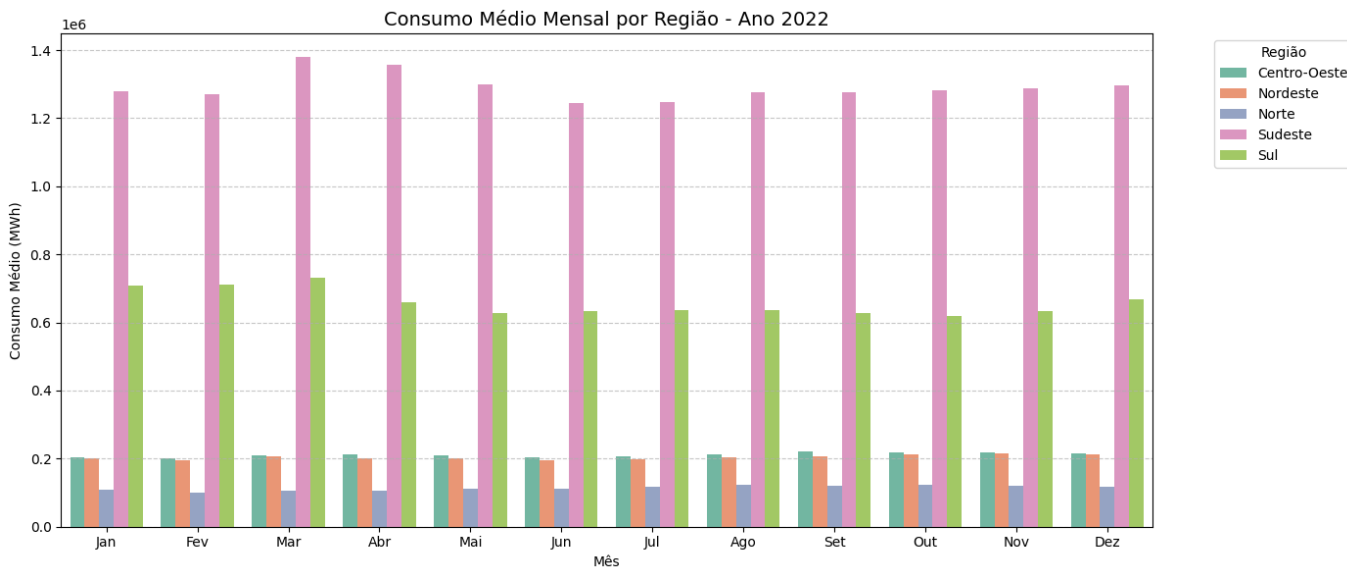
O gráfico acima evidencia que o Brasil tem tido uma demanda energética crescente, o que pode estar relacionado ao aumento populacional, expansão industrial e urbanização. O gráfico demonstra como o consumo de energia evoluiu ano a ano no país, tendências de crescimento ou queda, anos de estabilidade, crescimento rápido ou crises.

2º Comparação do consumo por região ao longo dos anos



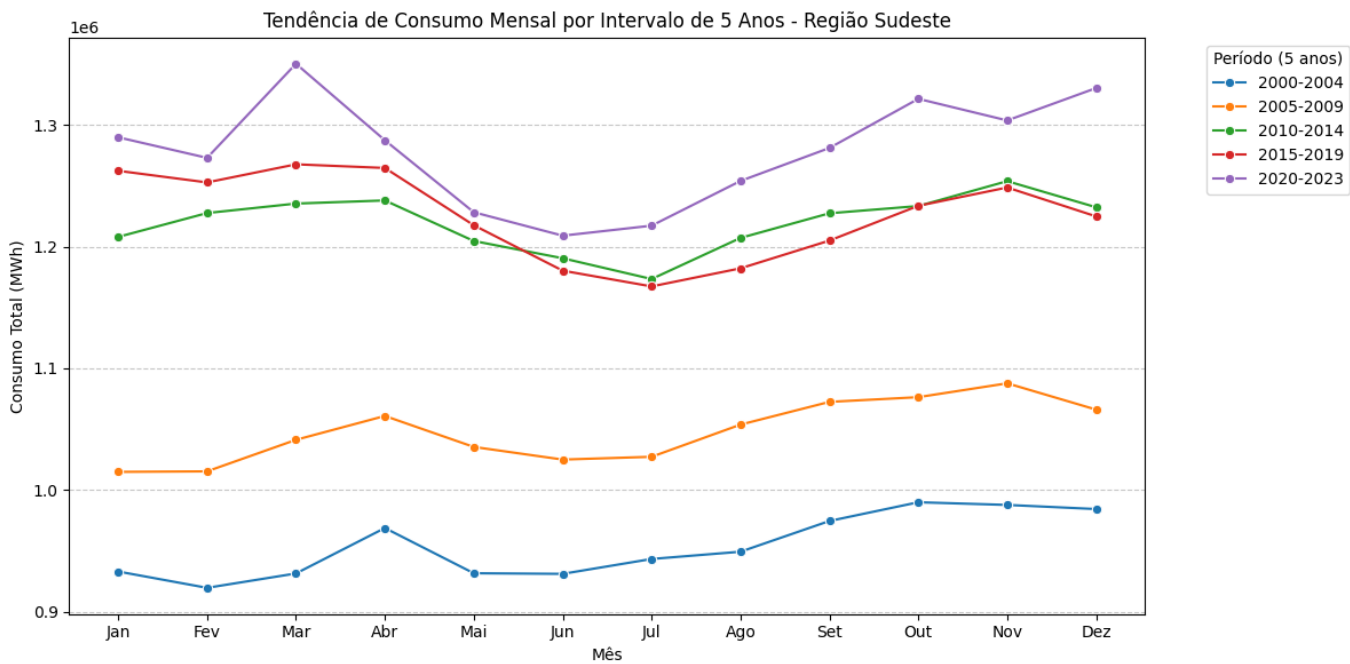
O Sudeste lidera com folga no consumo energético, refletindo sua importância econômica e demográfica. As demais regiões mostram crescimento contínuo, evidenciando desenvolvimento econômico e populacional mais equilibrado no país ao longo do tempo. A redução ou estabilização temporária do consumo em algumas regiões entre 2014–2017 e 2019–2020 pode indicar impactos econômicos (crises e pandemia).

3º Consumo médio mensal por região, filtrado por um ano



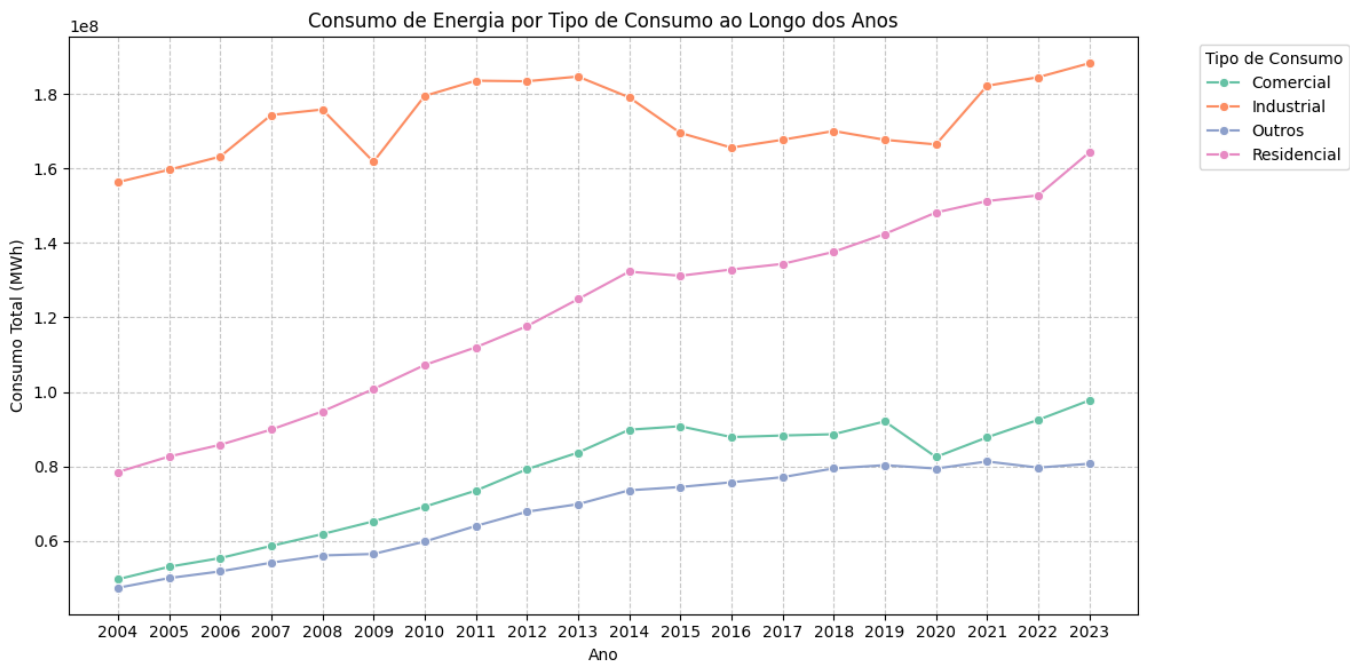
O Sudeste consome cerca de 5 a 6 vezes mais energia que o Norte, reforçando desigualdades regionais no consumo. O padrão de sazonalidade (variações mensais) é discreto, mas pode indicar influências climáticas ou econômicas. A leve queda de consumo no Sul entre maio e setembro pode estar relacionada à redução de atividades agrícolas ou industriais em períodos de clima mais rigoroso.

4º Tendência de consumo mensal por ano em uma região específica



O gráfico evidencia uma tendência clara de aumento do consumo na região Sudeste ao longo dos anos. O aumento gradual e consistente do consumo ao longo dos períodos reflete: Crescimento populacional e urbano; Expansão da atividade econômica e industrial; Aumento da penetração de equipamentos elétricos e ar-condicionado; O padrão de consumo mais elevado nos meses de março/abril e setembro/outubro pode estar associado a períodos de transição climática, que exigem mais energia (resfriamento ou aquecimento); A redução entre junho e julho é comum entre os períodos, sugerindo possível sazonalidade econômica ou climática.

5º Consumo por tipo de consumo ao longo dos anos



O gráfico apresenta a evolução do consumo de energia elétrica no Brasil, de 2004 a 2023, dividido por tipo de consumo: Industrial, Residencial, Comercial e Outros. Industrial: Lidera o consumo ao longo de todo o período, com oscilações entre 2009 e 2020, mas retomada e crescimento até 2023. Residencial: Crescimento constante e expressivo, ultrapassando 160 milhões de MWh em 2023. Comercial: Evolução constante até

2014, leve queda entre 2015 e 2020, retomando o crescimento a partir de 2021. Outros: Menor consumo entre os grupos, mas com crescimento gradual e estabilidade nos últimos anos.

6º Distribuição do número de consumidores por Região

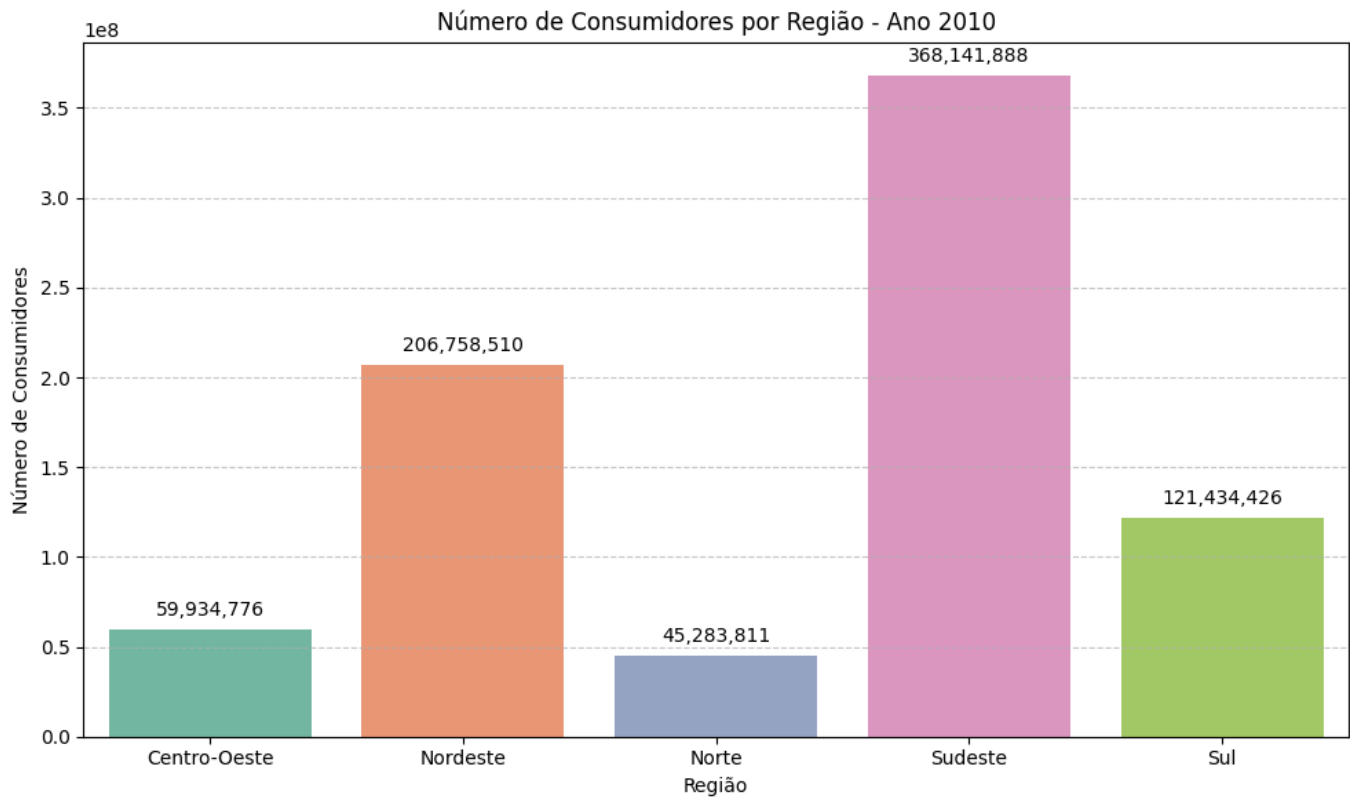
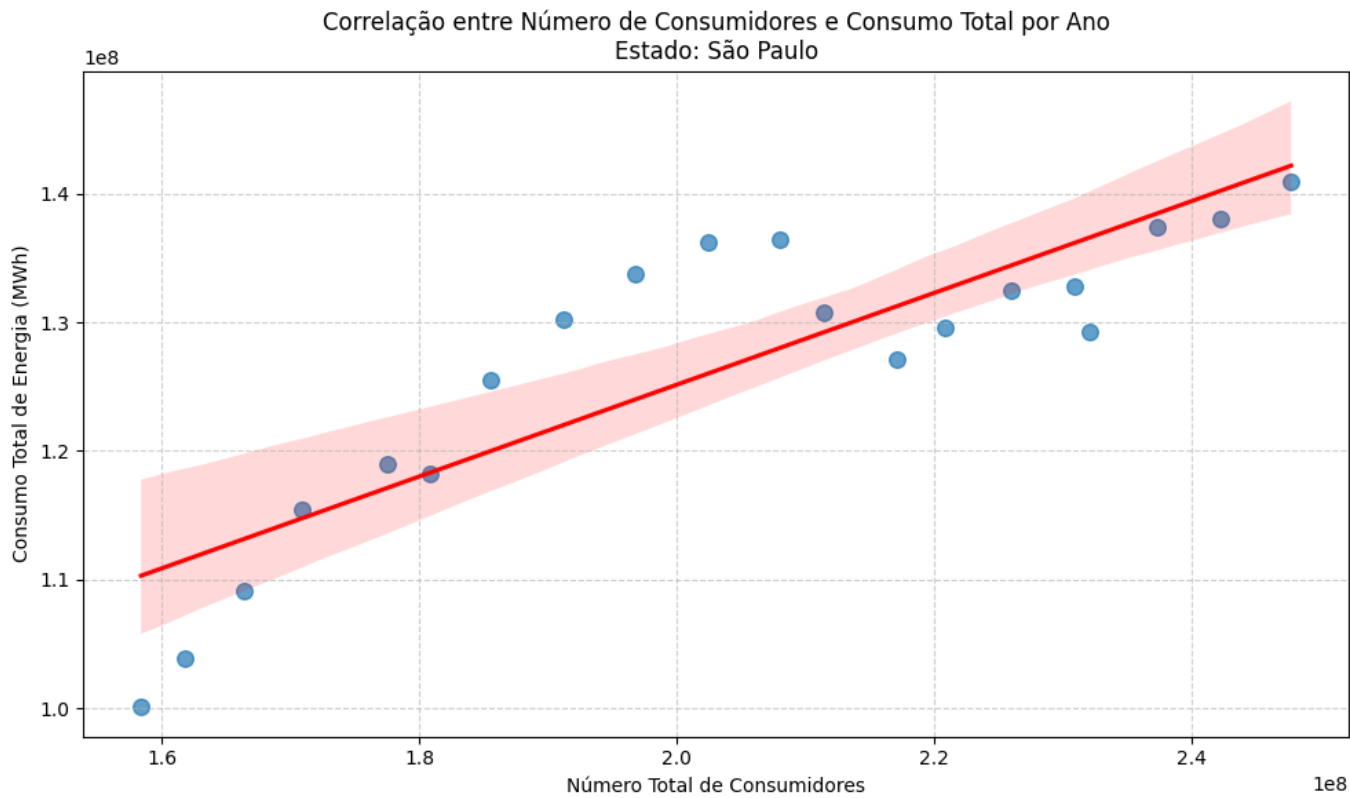


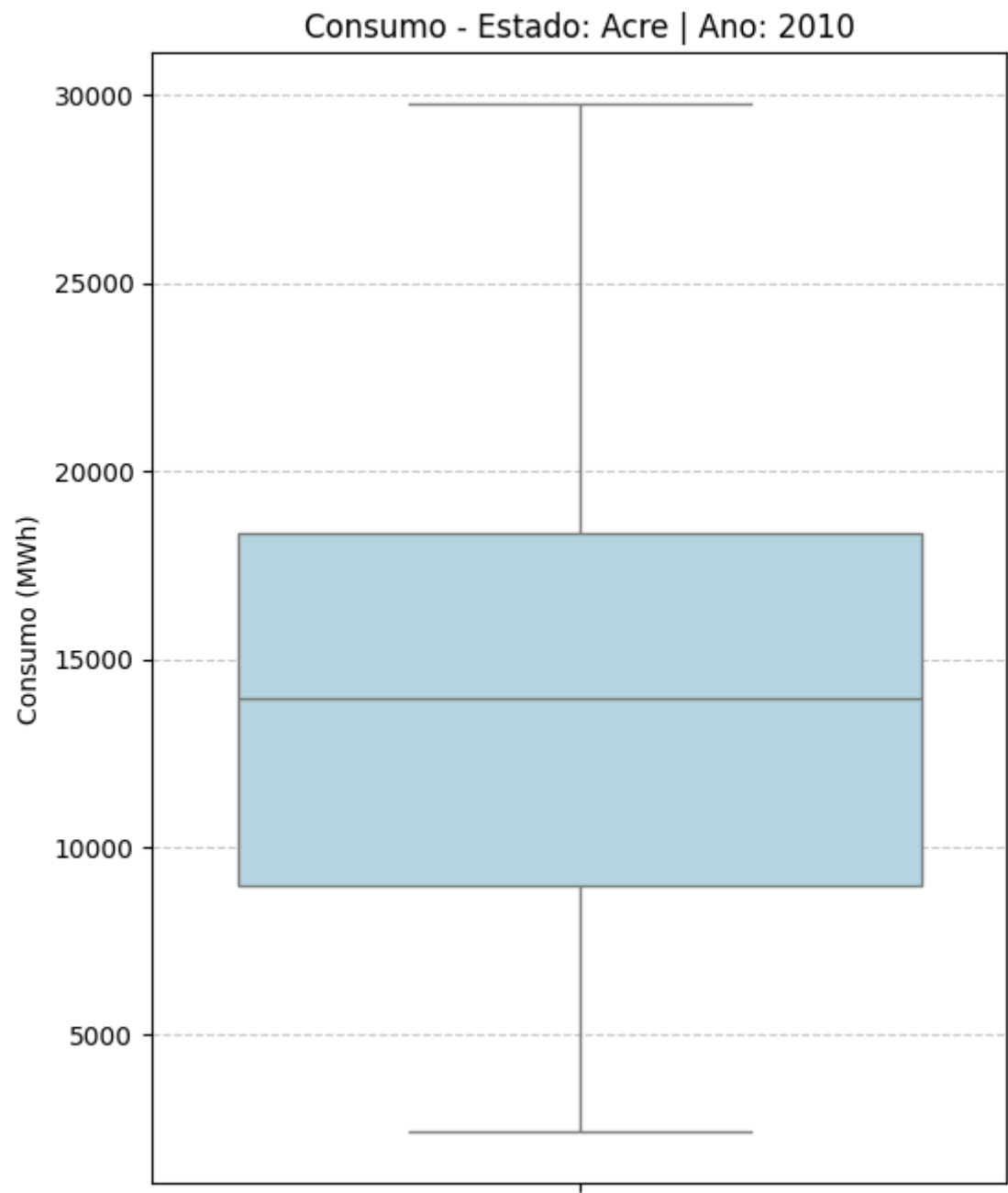
gráfico mostra a distribuição do número total de consumidores de energia elétrica por região em 2010: 1º Sudeste lidera com 368 milhões de consumidores. 2º Nordeste vem em segundo, com 206 milhões. 3º Sul ocupa o terceiro lugar, com 121 milhões. Centro-Oeste registra 59 milhões, e Norte tem o menor número, com 45 milhões. O Sudeste concentra a maior parte dos consumidores, refletindo sua densidade populacional e desenvolvimento urbano.

7º Correlação entre número de consumidores e consumo total por estado



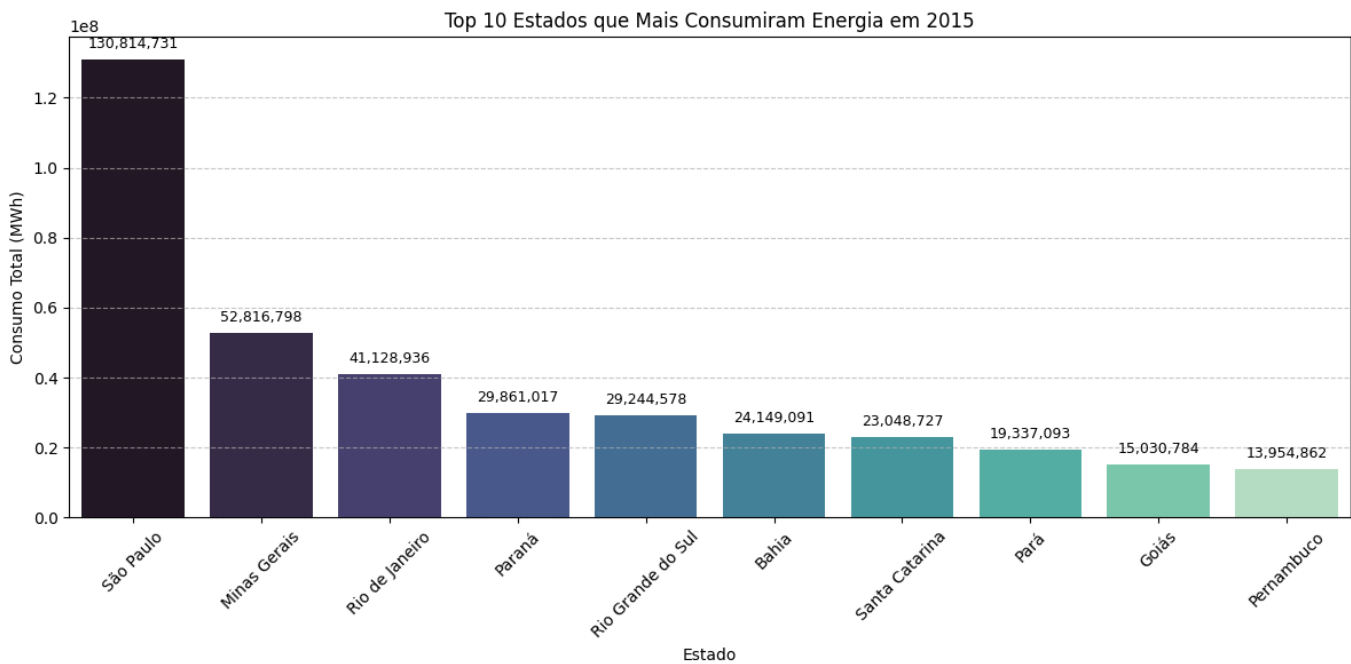
O estado de São Paulo mostra uma relação positiva entre o número total de consumidores e o consumo total de energia (em MWh) ao longo dos anos (2012-2022). A dispersão de pontos indica um aumento geral de ambos os valores, com uma linha de regressão vermelha confirmando a tendência ascendente. A grade no fundo facilita a leitura, e o título destaca o foco em São Paulo, sugerindo uma forte correlação entre o crescimento do número de consumidores e o consumo energético no estado.

8º Detecção de outliers no consumo por estado e ano (usando IQR)



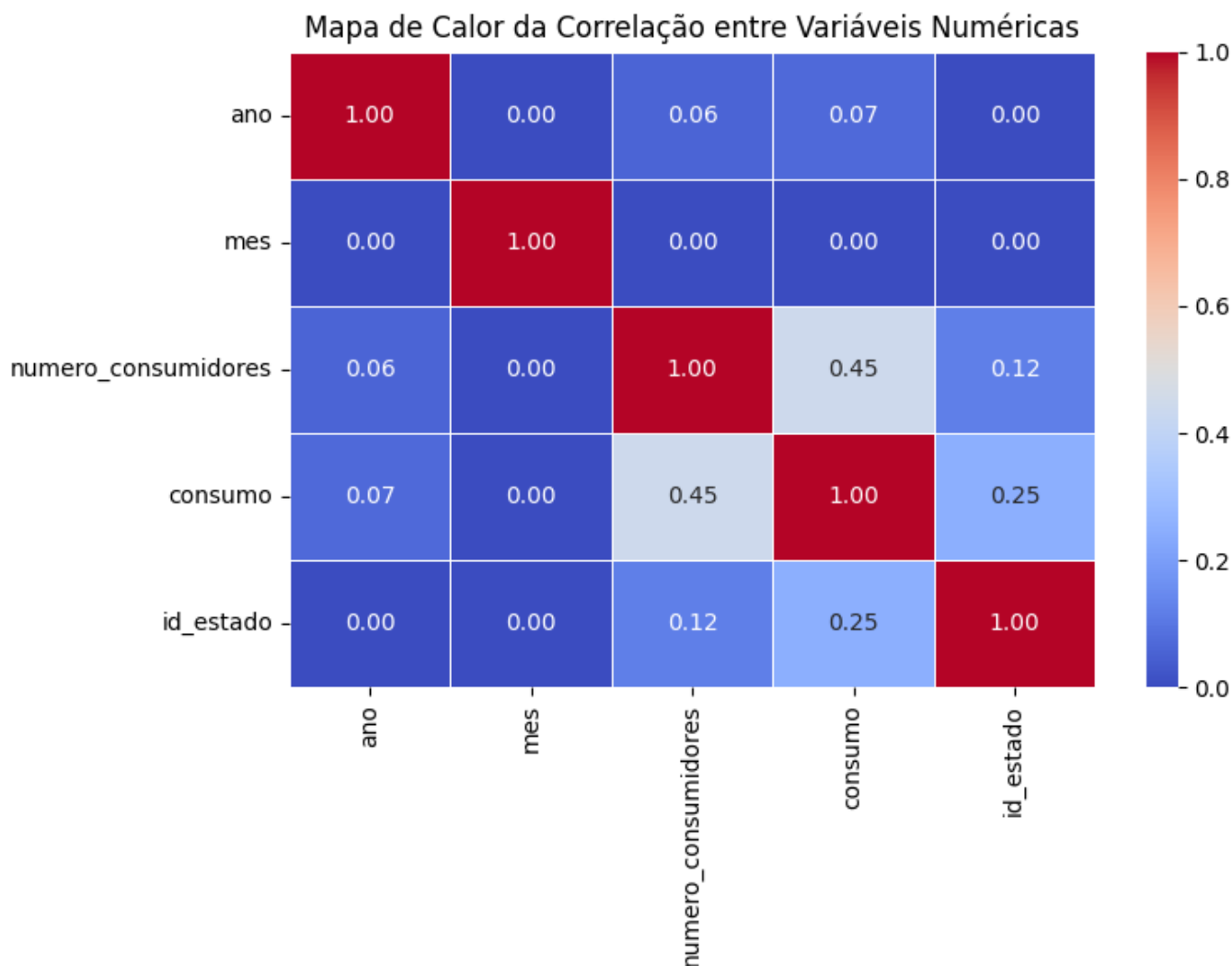
O gráfico de boxplot gerado para o estado do Acre em 2010 exibe a distribuição do consumo de energia, não destacando possíveis outliers. A caixa central mostra os quartis (Q1, mediana e Q3), com os "whiskers" indicando o intervalo interquartil (IQR). Pontos fora desse intervalo são marcados como outliers, sugerindo valores de consumo atipicamente altos ou baixos para o Acre nesse ano. A visualização é clara, com o foco no estado e ano especificados.

9º Estados que mais consomem energia



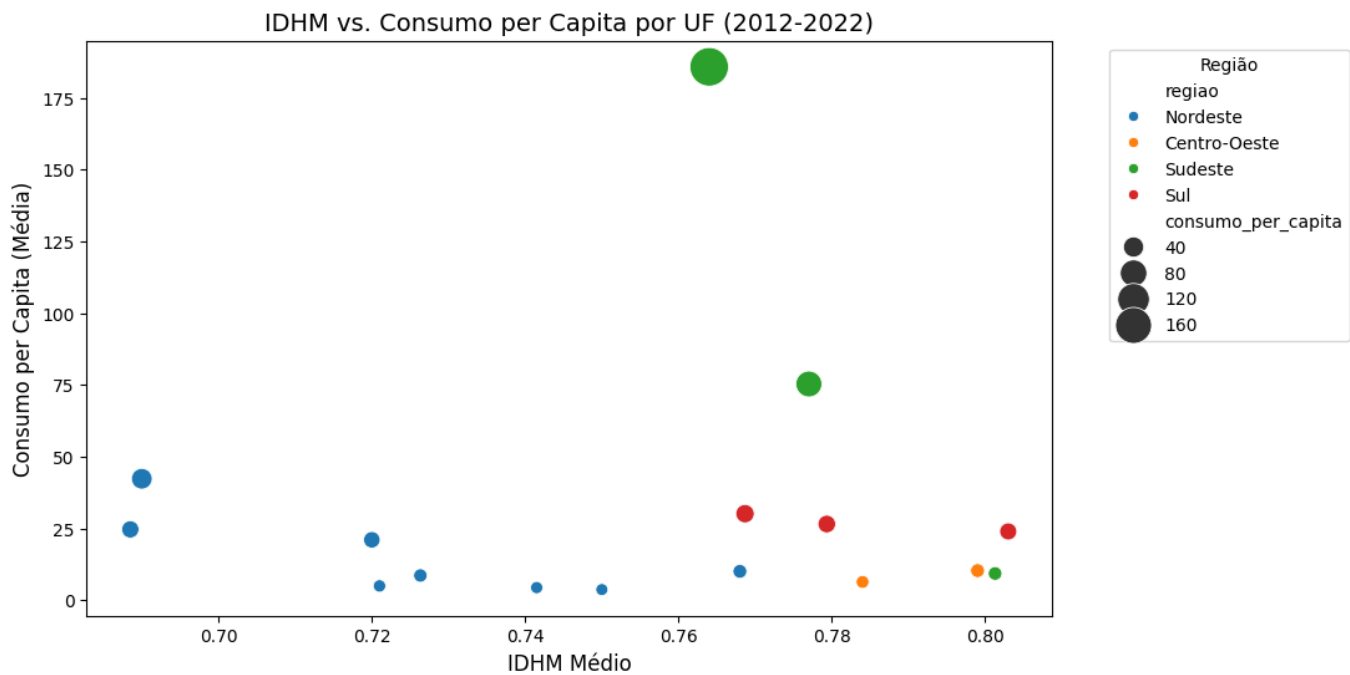
O gráfico do ano de 2015, exibe os 10 estados com maior consumo total de energia (em MWh), destacando São Paulo (SP) no topo, seguido por outros estados como Rio de Janeiro (RJ) e Minas Gerais (MG). As barras, coloridas pela paleta "mako", mostram uma variação significativa no consumo, com valores anotados acima de cada barra para clareza. Os rótulos dos estados estão rotacionados a 45 graus no eixo x, e uma grade no eixo y facilita a comparação, indicando uma concentração de consumo nos estados do Sudeste.

10º Mapa de calor da correlação entre variáveis numéricas (consumo, consumidores)



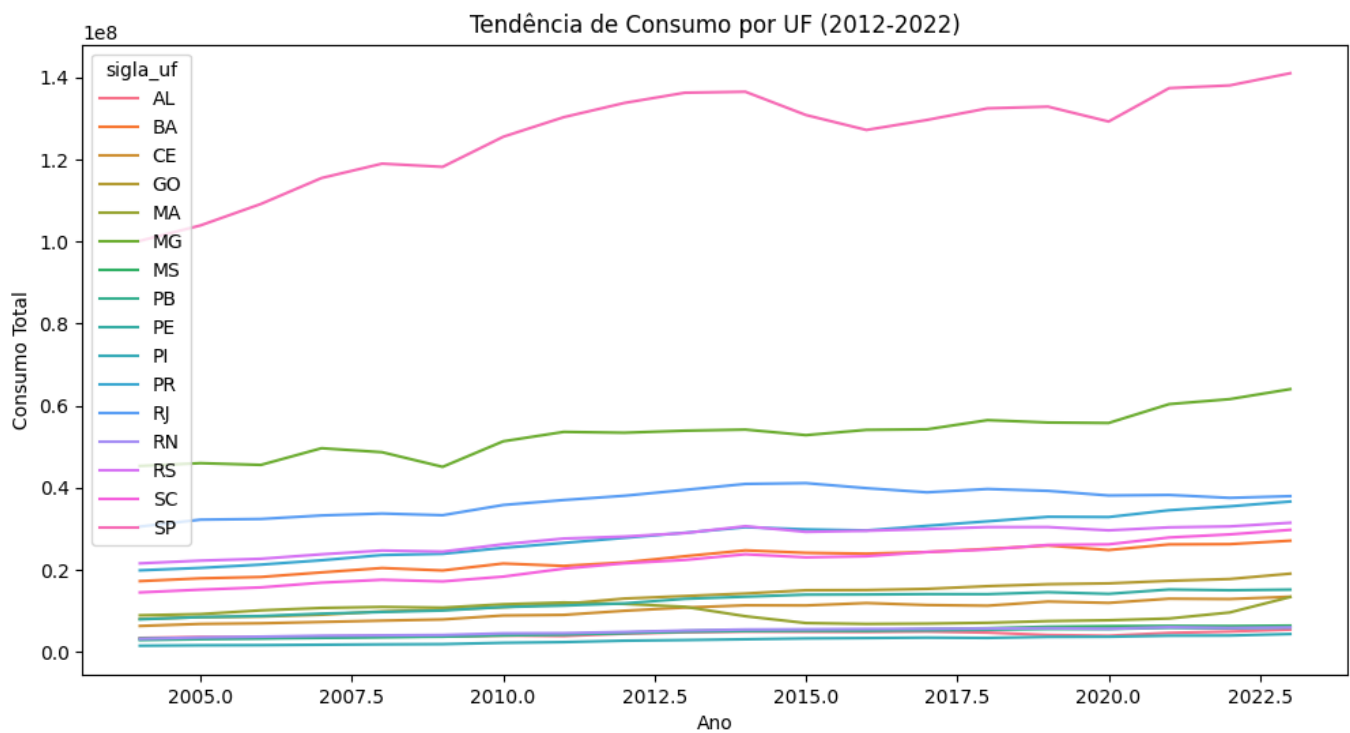
O gráfico de correlação para o estado de São Paulo mostra uma dispersão de pontos que representa a relação entre o número total de consumidores e o consumo total de energia (em MWh) ao longo dos anos (2012-2022). Uma linha de regressão vermelha indica uma tendência positiva e crescente, sugerindo uma forte correlação entre o aumento de consumidores e o consumo. Os pontos, com tamanho moderado e transparência, são distribuídos ao redor da linha, e a grade de fundo facilita a leitura, destacando a consistência dessa relação ao longo do período.

11º Relação entre IDHM e Consumo per Capita por UF



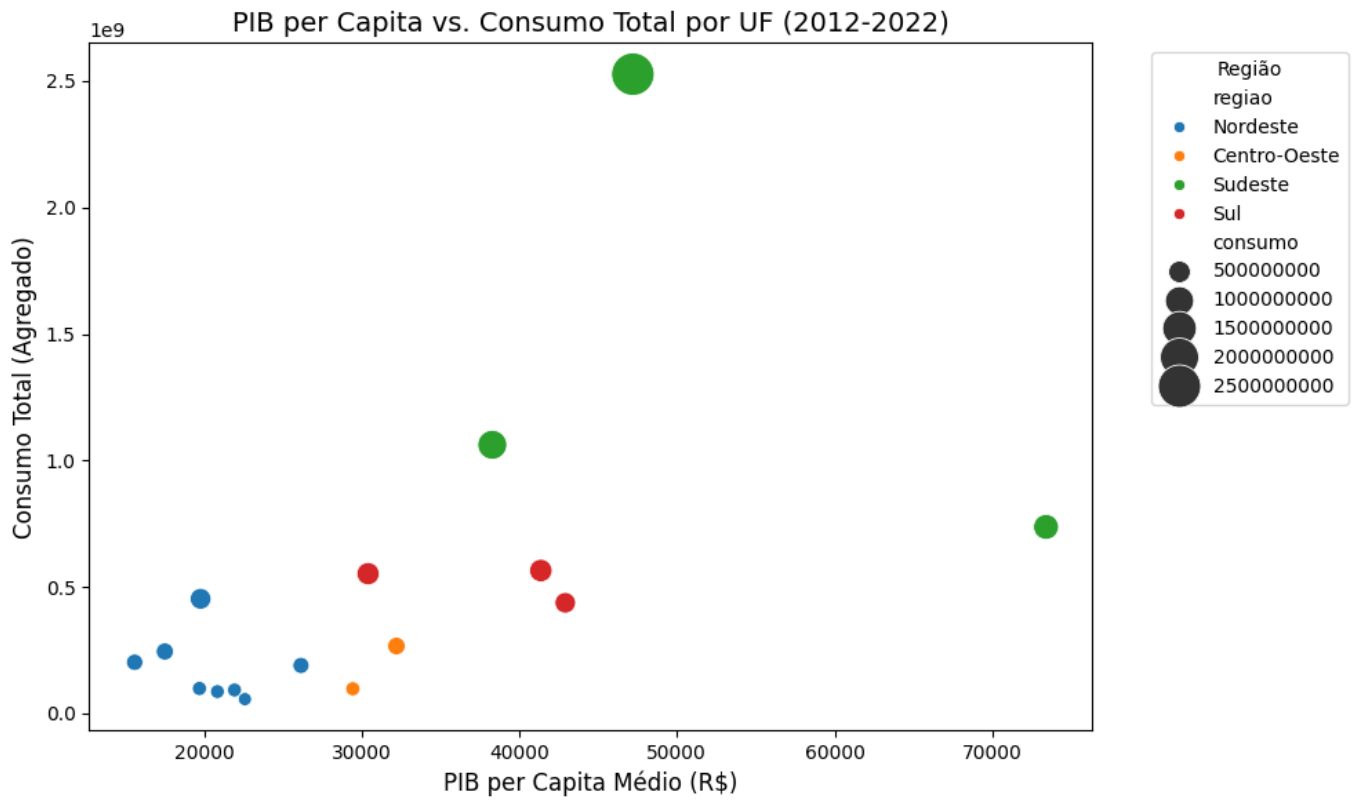
O gráfico exibe a relação entre o IDHM "índice de desenvolvimento humano" (média por UF) e o consumo total de energia para cada unidade federativa (UF) do Brasil, com base no dataset `df_final_pnad`. Cada ponto representa uma UF, identificado por uma cor única associada à sua `sigla_uf`, mostrando uma variação no consumo total em relação ao IDHM. O título "IDHM vs. Consumo por UF" e os eixos (IDHM no eixo x e Consumo Total no eixo y) são claros, mas a chamada `plt.show('idhm_vs_consumo.png')` parece ser um erro (deveria ser `plt.savefig()`), já que o gráfico foi exibido diretamente. A ausência de salvamento de arquivo não afeta a visualização, que destaca uma tendência geral de maior consumo em UFs com IDHM mais alto, como São Paulo e Rio de Janeiro.

12º Impacto no Consumo por Região



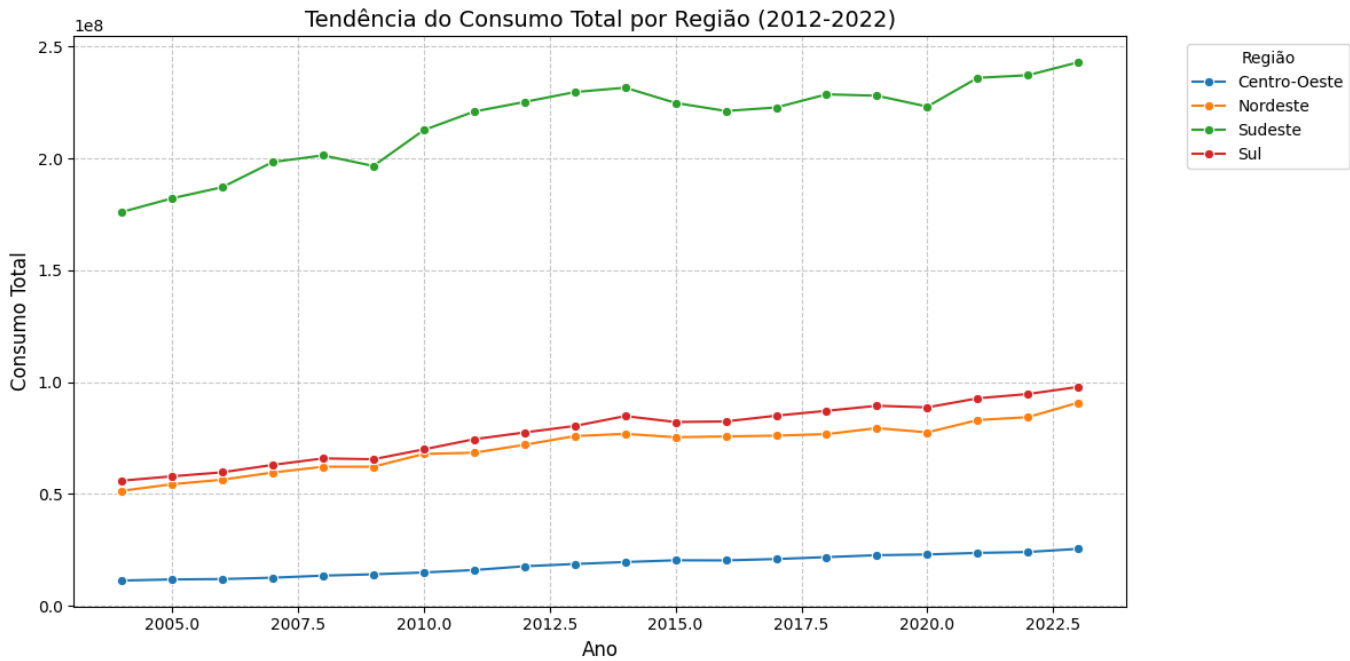
O gráfico exibe a tendência de consumo total de energia por unidade federativa (UF) ao longo dos anos (2012-2022), com base no dataset df_final_pnad. Cada linha representa uma UF, distinguida por uma cor associada à sigla_uf, mostrando variações anuais no consumo. O eixo x indica os anos, e o eixo y representa o consumo total, com o título "Tendência de Consumo por UF (2012-2022)" destacando o foco, o gráfico foi exibido diretamente, revelando uma tendência geral de aumento no consumo ao longo do período, com UFs como São Paulo e Rio de Janeiro apresentando valores mais altos e consistentes.

13º Consumo vs. PIB per Capita



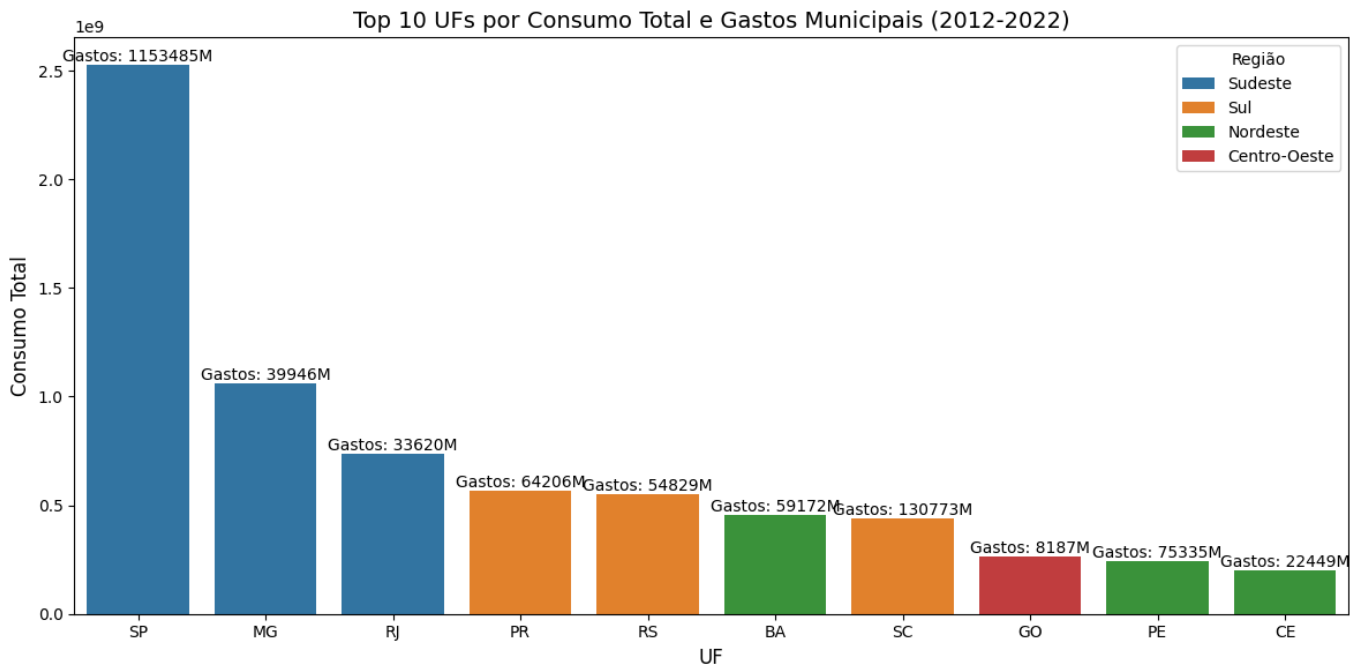
O gráfico exibe a relação entre o IDHM "índice de desenvolvimento Humano" médio e o consumo per capita (média) por unidade federativa (UF) de 2012 a 2022, com base no dataset df_final_pnad. Cada ponto representa uma UF, colorido por região (Norte, Nordeste, Sudeste, Sul, Centro-Oeste) e com tamanho proporcional ao consumo per capita, variando de 50 a 500. O eixo x mostra o IDHM médio, e o eixo y o consumo per capita, com o título "IDHM vs. Consumo per Capita por UF (2012-2022)" destacando o período. UFs como São Paulo (SP) e Santa Catarina (SC), do Sudeste e Sul, aparecem com IDHM mais alto e maior consumo per capita, indicando uma tendência positiva, enquanto a legenda à direita organiza as regiões, confirmando o destaque dessas áreas.

14º Contribuição do Setor de Serviços para o Número de Consumidores



O gráfico exibe a tendência do consumo total de energia por região (Norte, Nordeste, Sudeste, Sul, Centro-Oeste) de 2012 a 2022, com base no dataset `df_final_pnad`. Cada linha, marcada com pontos e colorida por região, mostra a variação anual do consumo total no eixo y, enquanto o eixo x representa os anos. O título "Tendência do Consumo Total por Região (2012-2022)" destaca o período, e a legenda à direita organiza as regiões. Observa-se uma queda acentuada no consumo em 2020 em todas as regiões, com o Sudeste sofrendo o maior impacto, seguida por uma recuperação parcial em 2021-2022, especialmente nas regiões Sul e Centro-Oeste, conforme indicado pelo comentário. A grade de fundo facilita a comparação entre as linhas.

15º Top UFs por Consumo e Gastos Municipais



O gráfico de barras exibe os 10 estados com maior consumo total de energia (2012-2022) com base no dataset `df_final_pnad`, com São Paulo (SP), Rio de Janeiro (RJ) e Minas Gerais (MG) liderando. As barras, coloridas por região (Sudeste, Sul, etc.), mostram o consumo total no eixo y, enquanto o eixo x lista as UFs

com rótulos rotacionados. Acima de cada barra, anotações indicam os gastos municipais (MUN_EXPENDIT) em milhões (M), destacando valores altos em SP, RJ e MG. O título "Top 10 UFs por Consumo Total e Gastos Municipais" e a legenda reforçam a análise, sugerindo uma correlação entre consumo e investimento público, conforme o comentário.

Ferramentas Sugeridas

O Arquivo em possui um documento requirements.txt onde demonstram todas a bibliotecas utilizadas no projeto, para possam replicar a utilização.

https://github.com/nerydyego/AED_eletricidade

- Bibliotecas utilizadas `import pandas as pd` `import numpy as np` `import matplotlib.pyplot as plt` `import seaborn as sns` `import sys`

Conclusão

A análise do consumo de energia elétrica no Brasil foi um processo rico em aprendizado, combinando manipulação de dados, visualização e interpretação de resultados. Esta reflexão aborda as etapas realizadas, os desafios enfrentados e as lições adquiridas, incentivando os alunos a documentar suas experiências. Cada etapa contribuiu para a compreensão do dataset e o desenvolvimento de habilidades analíticas.

Carregamento e União dos Dados: -- Descrição: Os datasets de consumo e IBGE foram carregados e mesclados usando pandas, com `sigla_uf` como chave. -- Contribuição: Essa etapa ensinou a lidar com dados de fontes distintas, destacando a importância de alinhar colunas e tratar incompatibilidades temporais (ex.: IDHM de 2010 vs. consumo 2012-2022). -- Habilidade: Manipulação de dados com `merge` e `groupby`. **Limpeza de Dados:** -- Descrição: Funções como `drop_duplicados` e `limpa_dados_NAN` removeram duplicatas e valores ausentes. -- Contribuição: Garantiu a integridade dos dados, revelando a necessidade de validação prévia para análises confiáveis. -- Habilidade: Tratamento de dados ausentes e duplicados. **Visualização e Análise:** -- Descrição: Gráficos como scatter plots, line plots e boxplots foram criados com seaborn e matplotlib para explorar tendências e outliers. -- Contribuição: Facilitou identificar padrões (ex.: impacto da pandemia) e correlações (ex.: consumidores vs. consumo), reforçando a escolha de visualizações adequadas. -- Habilidade: Visualização de dados e interpretação contextual.

Desafios Incompatibilidade Temporal: O IDHM estático de 2010 pode não refletir mudanças até 2022, introduzindo viés; assumimos representatividade, mas sugere-se atualizar os dados. **Aprendizado:** Avaliar a validade temporal é essencial para análises precisas. **Escala de Variáveis:** Valores altos de consumo e gastos requereram ajustes (ex.: anotações em milhões) para evitar distorções nos gráficos. **Aprendizado:** Normalizar escalas melhora a legibilidade e clareza visual. **Ausência de Regiões:** A criação manual de `regiao` aumentou a complexidade, mas enriqueceu as análises. **Aprendizado:** Variáveis derivadas ampliam insights, mas exigem cuidado na implementação.

Este projeto combinou técnicas analíticas com pensamento crítico, apesar de desafios como dados desatualizados. Cada etapa ensinou a importância de rigor e documentação, incentivando os alunos a refletir e aprimorar suas habilidades analíticas.