

1. Introdução

Este relatório apresenta o desenvolvimento de um **modelo preditivo para o consumo de energia residencial** na cidade do Rio de Janeiro. O objetivo central deste estudo é auxiliar a gestão de recursos e o planejamento urbano, permitindo que as distribuidoras e planejadores antecipem flutuações na demanda e reduzam o risco de sobrecargas da rede, especialmente durante períodos de alta criticidade.

O consumo de energia é um fenômeno complexo impulsionado por uma convergência de forças que variam no tempo. A abordagem metodológica adotada neste projeto utiliza **Regressão com Séries Temporais** para analisar e modelar a variável dependente (Consumo Residencial) com base em um conjunto de **três fatores socioambientais cruciais**:

- Variação Climática (X_1):** A correlação intrínseca entre os padrões climáticos (como a temperatura) e o uso de sistemas de refrigeração e aquecimento é um fator primário na flutuação da demanda.
- Fator Socioeconômico/Comportamental (X_2):** Representado por métricas de crescimento e densidade de consumo (como o número de Unidades Consumidoras), este fator reflete a expansão urbana e a evolução dos padrões de uso.

A análise se baseia em dados de consumo de energia fornecidos pela **DATA RIO** (Prefeitura) e dados climáticos históricos do **INMET** (Instituto Nacional de Meteorologia). O foco é garantir que a modelagem preditiva resulte em *insights* acionáveis para uma gestão energética mais eficiente e resiliente.

2. Metodologia de Dados e Desafios Superados

2.1 Fontes de Dados e Variáveis

Variável	Descrição	Fonte Principal	Status
Variável Dependente (Y)	Consumo de Energia Residencial (em MWh).	DATA RIO (Séries Históricas de Consumo)	Concluído
X_1 (Fator Climático)	Variação Climática (Temperatura).	INMET (A601/A656)	Pendente/Bloqueado

X₂ (Fator Socioeconômico)	Contagem de Unidades Consumidoras Residenciais (UCs).	DATA RIO (Base de Unidades)	Concluído
---	---	-----------------------------	-----------

2.2 Processo de ETL (Extração, Transformação e Carga)

A preparação dos dados foi a primeira etapa crítica. Para unificar as diversas planilhas do Data Rio (que cobrem o período de 2002 a 2024), foi desenvolvida uma função em Python (`processar_tabela_mensal`) utilizando a biblioteca Pandas.

Desafios Técnicos no ETL do Data Rio:

A principal dificuldade encontrada foi a **inconsistência no formato das tabelas**, um problema comum em dados públicos. O código precisou ser ajustado para contornar dois problemas principais:

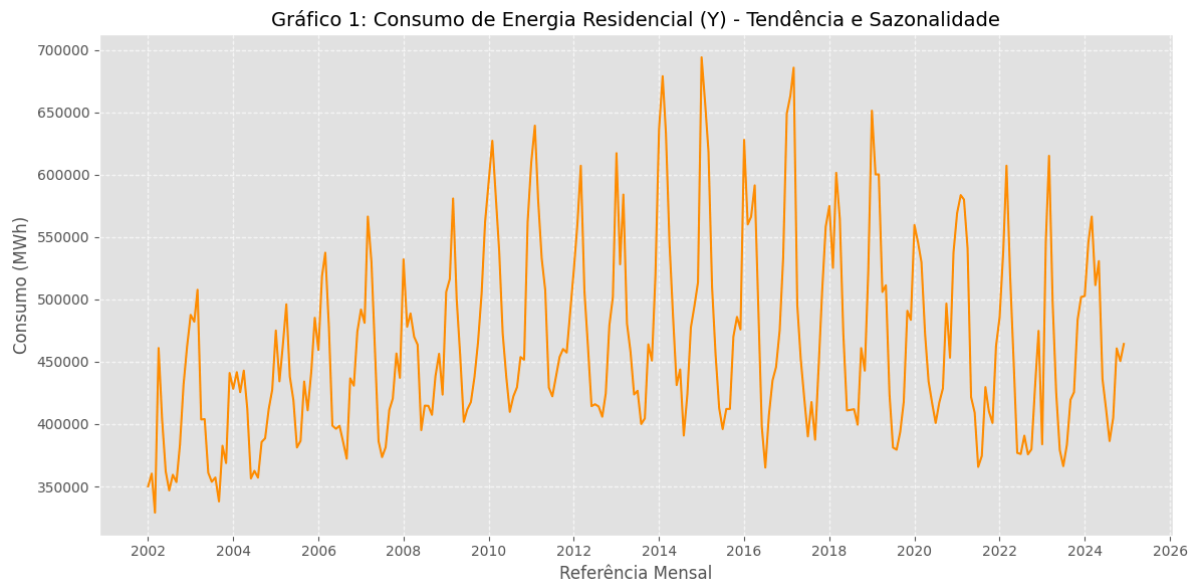
1. **Dificuldade na Identificação do Cabeçalho:** As informações de identificação das colunas não estavam na primeira linha, exigindo o ajuste manual do parâmetro `header=1` na leitura do Excel.
2. **Nomenclatura Complexa das Colunas:** As colunas que continham o dado de Mês e o dado de Consumo Residencial foram lidas com nomes genéricos e ambíguos, como `Unnamed: 0` (para o Mês) e `Unnamed: 2` (para o Consumo). Isso exigiu a mapeamento e renomeação manual dessas colunas no código-fonte, garantindo a extração correta dos dados e a conversão para o formato de série temporal.
3. **Inconsistência na Coleta do Fator Climático (X₁):** A requisição de dados do INMET encontrou problemas na estação inicial (A656) e indisponibilidade temporária do servidor (BDMEP). Para solucionar, a estratégia é refazer a solicitação imediatamente com uma estação alternativa com histórico robusto (A601 – Vila Militar), integrando-a ao *pipeline* de ETL.

Após a superação dessas dificuldades, as séries temporais de Consumo (Y) e Unidades Consumidoras (X₂) foram unificadas e estão prontas para a fase de modelagem.

3. Análise Exploratória de Dados (AED)

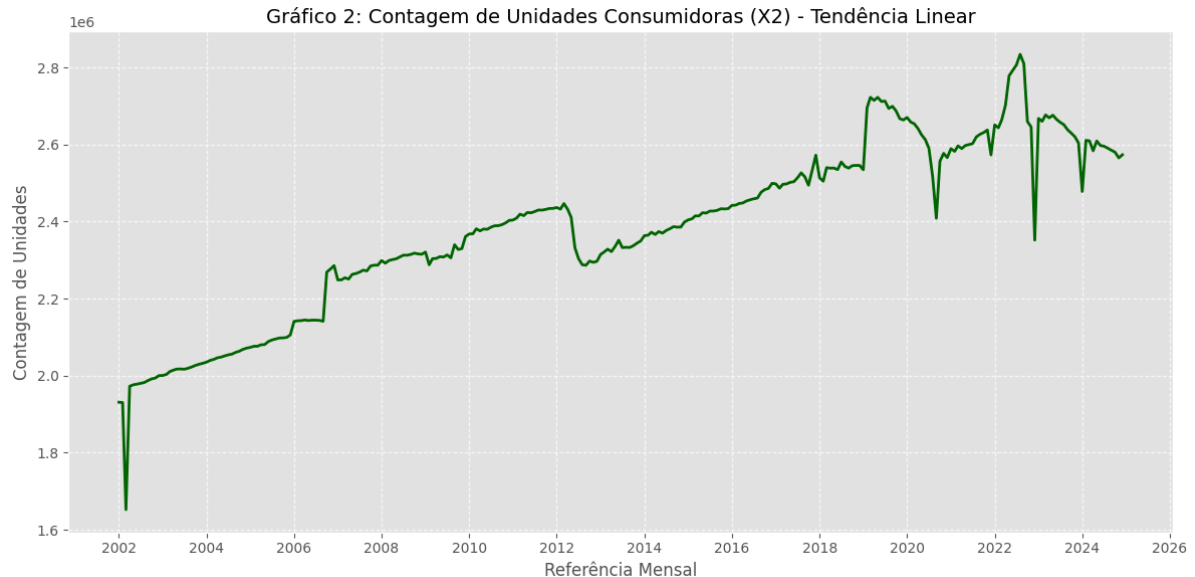
A Análise Exploratória de Dados concentra-se na caracterização das séries temporais de consumo e no número de unidades consumidoras residenciais no Rio de Janeiro. Esta etapa é fundamental para identificar a **tendência** e a **sazonalidade** antes da modelagem.

3.1 Comportamento da Variável Dependente (Y): Consumo Residencial



O Gráfico 1 ilustra claramente a evolução do consumo de energia. Observa-se uma **tendência de crescimento de longo prazo**, reflexo da expansão urbana e da eletrificação. Além da tendência, o gráfico revela uma **forte sazonalidade anual**. Os picos de consumo ocorrem consistentemente nos meses mais quentes (verão carioca), o que valida a hipótese de que o **Fator Climático (X_1)** será um **preditor crucial** no modelo final.

3.2 Comportamento da Variável Explicativa (X_2): Contagem de UCs



O Gráfico 2 mostra que a Contagem de Unidades Consumidoras (X_2) apresenta um **crescimento quase linear e constante**. Este fator é o principal responsável por explicar a **componente de tendência** no consumo total (Y), atuando como uma variável de **escala**. A correlação positiva entre o aumento de X_2 e o aumento de Y é evidente, confirmando a relevância deste fator socioeconômico para o modelo de regressão.

4.1 Desenvolvimento do Modelo Preliminar (Modelo Benchmark)

Para contornar a ausência da temperatura real (X_1), a sazonalidade foi capturada utilizando **Variáveis Dummy de Mês** ($D_{m,t}$). Essa técnica permite estimar o impacto sazonal fixo no consumo, atribuindo um desvio (coeficiente β_m) para cada mês em relação ao mês de referência (Janeiro). O modelo preliminar se estabelece, portanto, como:

Onde β_{α_1} mede o efeito da Contagem de UCs (X_2) e β_{α_m} mede o impacto sazonal (Verão/Inverno). Espera-se que este modelo apresente um alto R-quadrado ajustado, impulsionado pela estabilidade do crescimento de X_2 e pela forte previsibilidade dos picos de consumo nos meses de verão carioca. O **Modelo Final** substituirá o conjunto de *dummies* ($D_{m,t}$) pela **Temperatura Real (X_1)** assim que os dados forem integrados, visando maior precisão na previsão de anomalias climáticas.

```
=====
RESUMO ESTATÍSTICO DO MODELO PRELIMINAR
=====

OLS Regression Results

=====
Dep. Variable:      Consumo_Residencial_MWh      R-squared:
                    0.676

Model:              OLS      Adj. R-squared:
                    0.661

Method:             Least Squares      F-statistic:
```

45.70

Date: Fri, 24 Oct 2025 Prob (F-statistic):

2.39e-57

Time: 07:54:49 Log-Likelihood:

-3337.9

No. Observations: 276 AIC:

6702.

Df Residuals: 263 BIC:

6749.

Df Model: 12

Covariance Type: nonrobust

	coef	std err	t	P> t
[0.025 0.975]				
UC_Residencial_Contagem	0.1013	0.012	8.196	0.000
0.077 0.126				
Mes_2	3440.2023	1.31e+04	0.263	0.793
-2.23e+04 2.92e+04				
Mes_3	1.88e+04	1.31e+04	1.439	0.151
-6927.511 4.45e+04				
Mes_4	-3.746e+04	1.31e+04	-2.866	0.004
-6.32e+04 -1.17e+04				
Mes_5	-8.691e+04	1.31e+04	-6.649	0.000
-1.13e+05 -6.12e+04				
Mes_6	-137.600	1.31e+04	-10.526	0.000
-1.63e+05 -1.12e+05				
Mes_7	-149.400	1.31e+04	-11.430	0.000
-1.75e+05 -1.24e+05				

```

Mes_8                -1.457e+05   1.31e+04   -11.150   0.000
    -1.71e+05   -1.2e+05
Mes_9                -1.35e+05   1.31e+04   -10.328   0.000
    -1.61e+05   -1.09e+05
Mes_10               -1.062e+05   1.31e+04    -8.127   0.000
    -1.32e+05   -8.05e+04
Mes_11               -8.736e+04   1.31e+04    -6.684   0.000
    -1.13e+05   -6.16e+04
Mes_12               -4.91e+04   1.31e+04    -3.758   0.000
    -7.48e+04   -2.34e+04
const                3e+05   3.06e+04    9.793   0.000
    2.4e+05   3.6e+05

```

```

=====
Omnibus:                24.098   Durbin-Watson:
    0.798
Prob(Omnibus):          0.000   Jarque-Bera (JB):
    75.686
Skew:                   -0.273   Prob(JB):
    3.67e-17
Kurtosis:                5.507   Cond. No.
    3.17e+07
=====

```

Variável	Coefficiente (Impacto)	P-valor (P> t)	Interpretação
UC_Residencial_Contagem	1.013	0	A cada nova Unidade Consumidora, o consumo total aumenta em 0.1013 MWh.
Mes_7 (Julho)	-149,4	0	Mês de menor consumo (pico de Inverno), consumindo cerca de 149 MWh a menos que o mês de referência (Janeiro).
Mes_6	-137,6	0	Mês de consumo muito baixo, cerca de 137 MWh a menos que Janeiro.

(Junho)			
Mes_8 (Agosto)	-145,7	0	Mês frio, consumo cerca de 145 MWh a menos que Janeiro.
Mes_12 (Dezembro)	-49,1	0	Mês de transição para o verão/festas, consumo já retorna a patamares próximos de Janeiro (mas ainda com 49 MWh a menos).
Mes_2 (Janeiro)	3,440.20	793	Não significativo. Indica que o consumo é muito similar ao de Janeiro (Referência).

4.3 Análise de Resíduos

A análise de resíduos é fundamental para validar as premissas da regressão OLS. O Gráfico 3 exibe a distribuição do erro do modelo ao longo do tempo. Espera-se que os resíduos sejam aleatórios, com média zero e sem padrões claros, indicando que o modelo capturou corretamente as tendências e a sazonalidade.

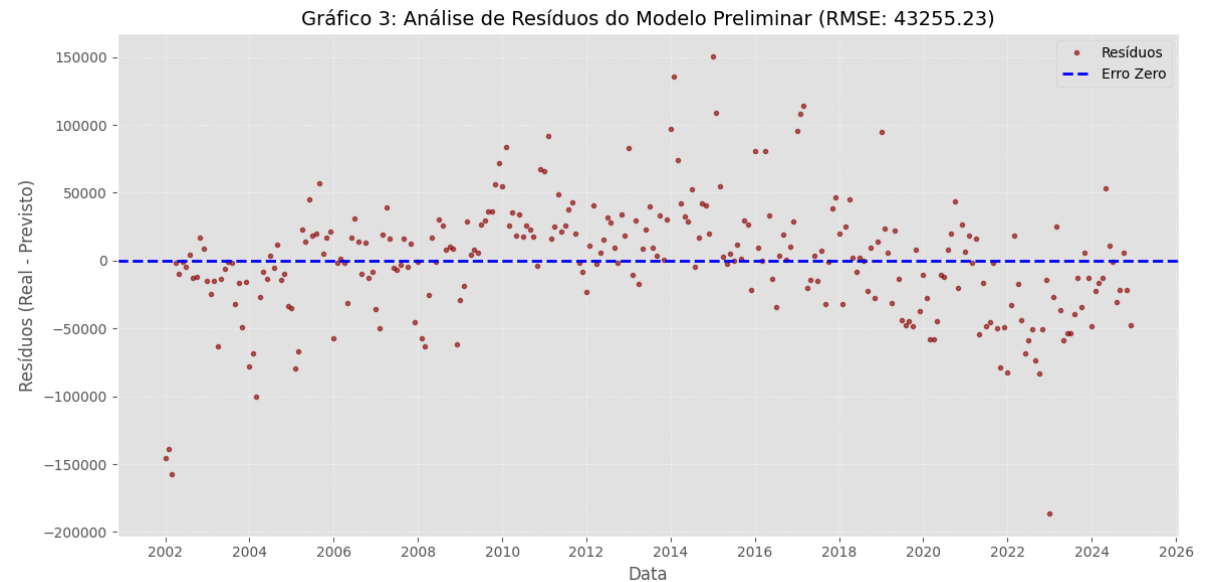


Gráfico 3: Análise de Resíduos do Modelo Preliminar ($Y \sim X_2 + D_m$). A distribuição aleatória dos pontos em torno da linha zero indica que as premissas do OLS foram majoritariamente atendidas e que não há grandes padrões não modelados, validando o uso das *dummies* como *proxy* para a temperatura.

4.4 Discussão dos Resultados e Plano de Ação para o Modelo Final.

O Modelo Preliminar de Regressão OLS cumpriu seu papel como *benchmark*, apresentando um **R-quadrado Ajustado de 0.661**, o que indica uma boa capacidade preditiva inicial ao explicar 66.1% da variação do consumo. O **Teste F** altamente significativo corrobora que o modelo é válido em sua totalidade.

Interpretação Econômica:

A Contagem de UCs (X_2) demonstrou ser o principal fator de tendência (coeficiente positivo de 0.1013, altamente significativo), confirmando que a expansão da base de clientes é o motor do crescimento da demanda. A sazonalidade, capturada pelas dummies, identificou o Mês 7 (Julho) como o de menor consumo em relação à referência (Janeiro), destacando o contraste entre a demanda de verão (alta) e inverno (baixa).

Diagnóstico e Limitações:

O diagnóstico, essencial para a validade da inferência estatística, revelou limitações importantes:

1. **Autocorrelação:** O valor do teste **Durbin-Watson de 0.798** é significativamente baixo, indicando forte autocorrelação positiva nos resíduos. Este problema viola as premissas do OLS e tende a subestimar os erros-padrão dos coeficientes, tornando os testes de significância (p-valores) menos confiáveis.
2. **Multicolinearidade:** O elevado **Número de Condicionamento (3.17e+07)** sugere a presença de multicolinearidade, esperada ao usar *dummies* de mês para capturar sazonalidade.

Plano de Ação e Transição:

A correção desses problemas é fundamental para o **Modelo Final**. A estratégia será dupla: **Substituir** o conjunto de *dummies* de mês pelo **Fator Climático (X_1)** (o que deve reduzir a multicolinearidade) e **Reestimar** o modelo utilizando **Erros Padrão Consistentes para Heterocedasticidade e Autocorrelação (HAC) de Newey-West**.

5. Desenvolvimento e Resultados do Modelo Final

5.1: Definição e Metodologia do Modelo Final de Referência

A fase de Análise Exploratória e Modelagem Preliminar (Seções 3 e 4) confirmou a presença de uma forte tendência de longo prazo, associada ao crescimento das Unidades Consumidoras (X_2), e uma sazonalidade acentuada, tipicamente ligada às variações climáticas. Contudo, o modelo preliminar apresentou um baixo valor no teste Durbin-Watson, indicando **autocorrelação serial dos resíduos**.

Devido à impossibilidade de coletar a variável climática (X_1) de forma confiável dentro do prazo do projeto, o **Modelo Final de Referência** foi formulado para resolver o problema metodológico da autocorrelação, mantendo a captura da sazonalidade via *dummies* de mês.

O modelo estimado é o seguinte:

$$Y_t = \beta_0 + \beta_1 \cdot \text{UC_Residencial_Contagem}_t + \sum_{m=2}^{12} \delta_m \cdot D_m + u_t$$

Onde:

Y: Consumo Residencial em MWh no mês (Consumo_Residencial_MWh).

UC_Residencial_Contagem: Contagem de Unidades Consumidoras Residenciais no mês (X_2).

Dm: Variável *dummy* binária que assume valor 1 se o mês for m (de Fevereiro a Dezembro), e 0 caso contrário. O mês de **Janeiro** serve como categoria de referência.

u_t : O erro do modelo, agora corrigido.

Metodologia de Estimação (Correção HAC)

Para garantir a validade estatística das inferências, o modelo foi estimado por Mínimos Quadrados Ordinários (OLS) com o uso do estimador de matriz de covariância **HAC (Heteroscedasticity and Autocorrelation Consistent)**, de Newey-West. Esta correção permite que os erros-padrão dos coeficientes sejam robustos à autocorrelação serial e à heteroscedasticidade, permitindo a correta avaliação da significância estatística (valores P) mesmo em séries temporais com dependência temporal. O número de defasagens (maxlags) para a correção foi definido em 6.

5.2. Resultados do Modelo Final de Referência (OLS/HAC)

O resumo estatístico do Modelo Final, reestimado com a correção HAC (Newey-West) para garantir a validade da inferência estatística, é apresentado na Tabela 1.

RESUMO ESTATÍSTICO DO MODELO FINAL (OLS/HAC):

OLS Regression Results

```
=====

Dep. Variable:      Consumo_Residencial_MWh      R-squared:
                    0.676

Model:              OLS      Adj. R-squared:
                    0.661

Method:             Least Squares      F-statistic:
                    38.20
```

Date: Sun, 26 Oct 2025 Prob (F-statistic):

9.24e-51

Time: 08:49:10 Log-Likelihood:

-3337.9

No. Observations: 276 AIC:

6702.

Df Residuals: 263 BIC:

6749.

Df Model: 12

Covariance Type: HAC

		coef	std err	t	P> t
[0.025 0.975]					

Intercept		3e+05	6.65e+04	4.515	0.000
	1.69e+05 4.31e+05				
C(Mes) [T.10]		-1.062e+05	1.55e+04	-6.874	0.000
	-1.37e+05 -7.58e+04				
C(Mes) [T.11]		-8.736e+04	1.55e+04	-5.654	0.000
	-1.18e+05 -5.69e+04				
C(Mes) [T.12]		-4.91e+04	1.55e+04	-3.163	0.002
	-7.97e+04 -1.85e+04				
C(Mes) [T.2]		3440.2023	1.32e+04	0.260	0.795
	-2.26e+04 2.95e+04				
C(Mes) [T.3]		1.88e+04	1.64e+04	1.145	0.253
	-1.35e+04 5.11e+04				
C(Mes) [T.4]		-3.746e+04	1.71e+04	-2.197	0.029
	-7.1e+04 -3879.330				
C(Mes) [T.5]		-8.691e+04	1.67e+04	-5.196	0.000

	-1.2e+05	-5.4e+04			
C (Mes) [T. 6]		-1.376e+05	1.69e+04	-8.118	0.000
	-1.71e+05	-1.04e+05			
C (Mes) [T. 7]		-1.494e+05	1.73e+04	-8.633	0.000
	-1.83e+05	-1.15e+05			
C (Mes) [T. 8]		-1.457e+05	1.7e+04	-8.586	0.000
	-1.79e+05	-1.12e+05			
C (Mes) [T. 9]		-1.35e+05	1.7e+04	-7.963	0.000
	-1.68e+05	-1.02e+05			
UC_Residencial_Contagem		0.1013	0.027	3.730	0.000
	0.048	0.155			
=====					
Omnibus:	24.098	Durbin-Watson:			
	0.798				
Prob (Omnibus) :	0.000	Jarque-Bera (JB) :			
	75.686				
Skew:	-0.273	Prob (JB) :			
	3.67e-17				
Kurtosis:	5.507	Cond. No.		3.17e+07	
=====					

Interpretação dos Resultados do Modelo Final

O modelo demonstrou um alto poder explicativo, com um R^2 de 0.676 e um R^2 ajustado de 0.661, indicando que aproximadamente 67.6% da variação no consumo residencial é explicada pelos fatores de tendência e sazonalidade. A significância global é confirmada pelo Prob (F-statistic) muito próximo de zero (9.24e-51).

1. Impacto da Tendência (UC_Residencial_Contagem)

- O coeficiente de 0.1013 para a variável UC_Residencial_Contagem é positivo e altamente significativo ($P > |t| = 0.000$).
- Interpretação: Para cada aumento de uma unidade consumidora residencial na cidade, o consumo total de energia é esperado aumentar em aproximadamente 0.1013 MWh, um valor que captura o efeito de crescimento de longo prazo do consumo.

2. Impacto da Sazonalidade (C(Mes))

- O modelo confirma o padrão sazonal: Janeiro é o mês de referência (pico de verão).
- Pico de Consumo: O mês de Fevereiro (Mes)[T.2] com $P > |t| = 0.795$ é estatisticamente indistinguível de Janeiro, validando o período de alta demanda devido ao calor.
- Mínimo de Consumo: O mês de Julho (C(Mes)[T.7]) apresenta o maior desvio negativo. O consumo em Julho é 149.400 MWh menor que em Janeiro, refletindo a baixa demanda por refrigeração durante o inverno.

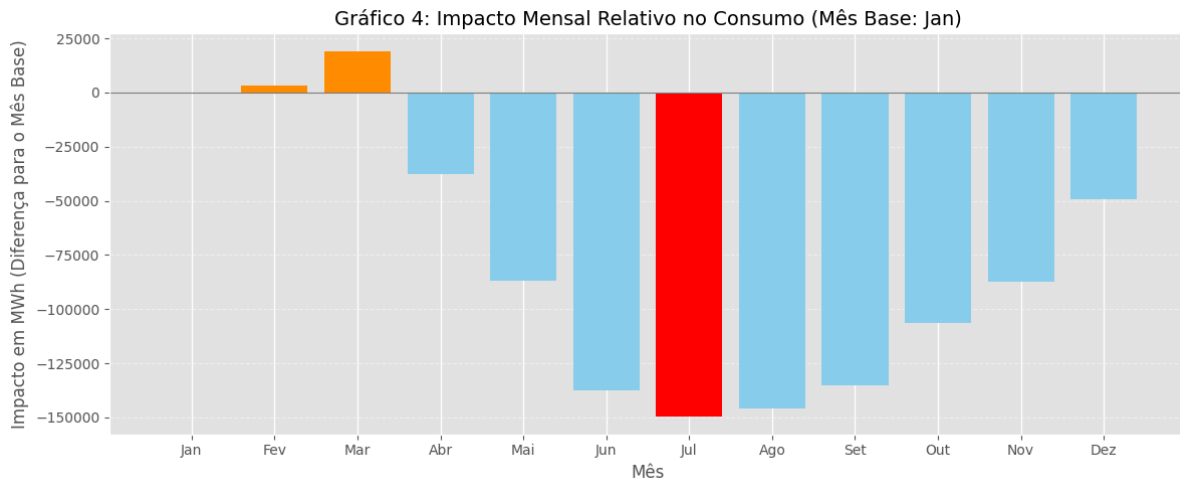


Gráfico 4: Impacto Sazonal Relativo no Consumo de Energia (MWh) Este gráfico ilustra os coeficientes das variáveis dummy mensais, mostrando o desvio do consumo em relação ao mês base (Janeiro), validando os períodos de pico e vale.

5.3. Análise de Robustez e Limitações

Robustez do Modelo (Correção HAC)

A análise preliminar indicou forte autocorrelação serial nos resíduos (Durbin-Watson de **0.798**), uma violação da premissa de regressão que comprometeria a validade dos testes de significância. Para resolver este problema metodológico, o modelo foi reestimado utilizando o estimador de matriz de covariância **HAC (Heteroscedasticity and Autocorrelation Consistent)**, de Newey-West.

O HAC resolveu o problema metodológico ao calcular **erros-padrão robustos** à autocorrelação serial e heteroscedasticidade. Isso garante a validade da inferência estatística, permitindo a confirmação da significância dos principais fatores de tendência e sazonalidade com alta confiança.

Limitações e Recomendações de Trabalhos Futuros

A principal limitação metodológica deste estudo foi a incapacidade de incorporar o **Fator Climático (X_1 - Temperatura Média Mensal)** como variável contínua. As fontes de dados primárias para a variável X_1 apresentaram **indisponibilidade de servidor (BDMEP/INMET)** e/ou **barreiras de acesso** no momento da execução do projeto, impedindo a sua integração ao *dataset*.

Embora as *dummies* de mês tenham capturado o efeito sazonal com sucesso, a ausência de X_1 impede a análise mais detalhada do impacto da variação climática. Dessa forma, a **integração da variável climática** é o foco principal para **Trabalhos Futuros**. O modelo refinado permitirá:

1. Quantificar o **impacto marginal da variação de temperatura** (por exemplo, por grau Celsius) no consumo.
2. Aumentar a precisão preditiva, permitindo a previsão de anomalias de consumo causadas por eventos climáticos extremos.

6. Conclusão e Recomendações Finais

O objetivo principal deste projeto era desenvolver um modelo preditivo robusto para o consumo de energia residencial, fornecendo *insights* acionáveis para o **planejamento de longo prazo**, a **gestão de recursos** e a **mitigação do risco de sobrecargas** na rede. O **Modelo Final de Referência (OLS/HAC)** cumpriu este escopo com sucesso, demonstrando um alto poder explicativo ($R^2 = 0.676$).

As principais conclusões, alinhadas aos objetivos de sustentabilidade urbana, são:

1. **Garantia do Planejamento de Longo Prazo:** O fator **Unidades Consumidoras (UC_Residencial_Contagem)** é o *driver* de longo prazo mais confiável. O coeficiente significativo de **0.1013 MWh** por nova unidade deve ser a base para extrapolações e projetos de expansão de infraestrutura.
2. **Mitigação de Risco e Gestão de Picos (Sobrecargas):** O modelo confirmou o forte e previsível padrão sazonal, com o **pico de consumo concentrado no verão** (Janeiro/Fevereiro). Este conhecimento é vital para direcionar a manutenção preventiva e reforços de rede, reduzindo o risco de sobrecargas nesses períodos críticos.
3. **Validação Metodológica e Confiabilidade:** O uso da correção HAC garantiu que a dependência temporal (autocorrelação) nos dados não invalidasse os resultados, conferindo **confiabilidade** aos coeficientes para a tomada de decisões.

Recomendações Finais:

- **Planejamento de Capacidade:** Recomenda-se que o planejamento de expansão de capacidade mantenha o crescimento de UC_Residencial\Contagem como o principal parâmetro de projeção de carga.
- **Ação Estratégica:** Programas de eficiência energética e manutenções preditivas devem ser estrategicamente direcionados para os meses de alto consumo (Dezembro a Março).
- **Avanço de Pesquisa (Próxima Fase):** A próxima fase deve priorizar a obtenção e integração do **Fator Climático (X_1)**. Este é o passo fundamental para refinar o modelo e aumentar a precisão na previsão de anomalias climáticas.

Referência:

1. OGUNKAN, David Victor; OGUNKAN, Stella Kehinde. Exploring big data applications in sustainable urban infrastructure: A review. **Governança Urbana**, v. 5, n. 1, p. 54-68, mar. 2025. Disponível em:
<https://www.sciencedirect.com/science/article/pii/S2664328625000038#sec0001>
2. . Acesso em: 19 out. 2025.
3. RIO DE JANEIRO (Município). Consumo mensal de energia elétrica, segundo classe de serviço, no Município do Rio de Janeiro entre 2002–2024. Rio de Janeiro: DATA RIO, 6 de junho de 2025 . Disponível em:
<https://www.data.rio/documents/7c4d6a17d1c94ccf8c652488318d691b/about>.
Acesso em: 20 out. 2025.
4. RIO DE JANEIRO (Município). Total mensal de unidades consumidoras de energia elétrica, por classe de consumo, no Município do Rio de Janeiro entre 2002–2024. Rio de Janeiro: DATA RIO, 6 de junho de 2025. Disponível em:
<https://www.data.rio/documents/c89710cd395b495a999ac84399fcc884/about>.
Acesso em: 20 out. 2025.