

Parte I - Análise Estatística

1. Tratamento de Dados Ausentes:

Antes de iniciar a análise, foram identificados valores ausentes nas variáveis `latencia_ms`, `armazenamento_tb`, `tipo_hd` e `tipo_processador`. As estratégias de imputação adotadas foram:

Para as variáveis numéricas

- `latencia_ms`: imputada com a mediana (145.54).
- `armazenamento_tb`: imputada com a mediana (0.512).

Para as variáveis categóricas:

- `tipo_hd`: imputada com a moda (SSD).
- `tipo_processador`: imputada com a moda (Apple Silicon).

```
Valores ausentes por coluna (Antes):
cpu_cores           0
ram_gb              0
latencia_ms         19
armazenamento_tb   19
sistema_operacional 0
tipo_hd             19
tipo_processador    1
tempo_resposta      0
dtype: int64
Valores ausentes em latencia_ms imputados com a mediana (145.54)
Valores ausentes em armazenamento_tb imputados com a mediana (0.512)
Valores ausentes em tipo_hd com a moda (SSD)
Valores ausentes em tipo_processador com a moda (Apple Silicon)
```

2. Análise Descritiva das Variáveis:

Foram analisadas as variáveis `cpu_cores`, `ram_gb`, `latencia_ms`, `armazenamento_tb` e `tempo_resposta`. Abaixo estão os principais destaques:

- **Variáveis Numéricas:**

- **ram_gb**: apresenta alta dispersão ($DP = 21.78$), com valores entre 4 GB e 64 GB. A média (23.4 GB) está acima da mediana (16 GB), indicando uma assimetria à direita.
- **latencia_ms**: varia entre 4.93 e 296.03 ms, com média de 148.6 e mediana de 145.5, sugerindo distribuição levemente assimétrica.
- **armazenamento_tb**: de 0.256 TB a 2 TB, com mediana de 0.512 TB e média de 0.83 TB, sugerindo alguns valores mais altos puxando a média.

- **tempo_resposta:** varia de 35.16 a 405.22, com média de 117.71 e mediana de 92.98, indicando uma distribuição assimétrica com alguns valores altos de resposta.

- **Variáveis Categóricas:**

- **sistema_operacional:** distribuição relativamente equilibrada entre MacOS (73), Windows (70) e Linux (51).
- **tipo_hd:** SSD é o tipo predominante com 111 registros, contra 83 HDDs.
- **tipo_processador:** Apple Silicon é o mais comum (73), seguido de Intel (64) e AMD (57).

Essa distribuição equilibrada entre categorias favorece a modelagem posterior, garantindo que todas as categorias tenham representação suficiente para gerar coeficientes estáveis.

Análise Descritiva - Variáveis Numéricas:

	cpu_cores	ram_gb	latencia_ms	armazenamento_tb	tempo_resposta
count	194.000000	194.000000	194.000000	194.000000	194.000000
mean	8.335052	23.402062	148.577577	0.829938	117.712732
std	4.084104	21.784940	85.002887	0.606466	79.365549
min	2.000000	4.000000	4.930000	0.256000	35.160000
25%	5.000000	5.000000	83.257500	0.256000	67.720000
50%	9.000000	16.000000	145.540000	0.512000	92.980000
75%	11.750000	32.000000	210.880000	1.000000	133.052500
max	15.000000	64.000000	296.030000	2.000000	405.220000

Análise Descritiva - Variáveis Categóricas:

	sistema_operacional	tipo_hd	tipo_processador
count	194	194	194
unique	3	2	3
top	MacOS	SSD	Apple Silicon
freq	73	111	73

Parte II - Modelo e Diagnóstico

1. Ajuste do Modelo de Regressão Linear Múltipla (Modelo 1)

- Variável dependente: tempo_resposta
- Variáveis explicativas: todas as demais após codificação dummie.
- R^2 : 0.694
- R^2 ajustado: 0.681
- Intercepto: 276.91

2. Tratamento e Interpretação das Variáveis Categóricas:

As variáveis categóricas sistema_operacional, tipo_hd e tipo_processador foram tratadas por codificação one-hot (get_dummies), com exclusão da primeira categoria como base. As categorias base foram: sistema_operacional: Linux, tipo_hd: HDD, tipo_processador: AMD.

Um sistema com MacOS (vs Linux) reduz o tempo de resposta em 3.26 unidades, mas essa variável possui multicolinearidade. Dispositivos com SSD (vs HDD) apresentam menor tempo de resposta (-6.19), o que é estatisticamente significativo.

OLS Regression Results						
=====						
Dep. Variable:	tempo_resposta	R-squared:	0.694			
Model:	OLS	Adj. R-squared:	0.681			
Method:	Least Squares	F-statistic:	52.44			
Date:	Mon, 05 May 2025	Prob (F-statistic):	1.29e-43			
Time:	21:58:13	Log-Likelihood:	-1008.5			
No. Observations:	194	AIC:	2035.			
Df Residuals:	185	BIC:	2064.			
Df Model:	8					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	276.9071	12.443	22.253	0.000	252.358	301.456
cpu_cores	-14.0524	0.808	-17.385	0.000	-15.647	-12.458
ram_gb	-1.5459	0.151	-10.232	0.000	-1.844	-1.248
latencia_ms	0.0001	0.039	0.004	0.997	-0.076	0.076
armazenamento_tb	-3.2488	5.423	-0.599	0.550	-13.947	7.450
sistema_operacional_MacOS	-3.1775	4.822	-0.659	0.511	-12.691	6.336
sistema_operacional_Windows	-0.1981	8.292	-0.024	0.981	-16.557	16.161
tipo_hd_SSD	4.6192	6.589	0.701	0.484	-8.380	17.618
tipo_processador_Apple Silicon	-3.1775	4.822	-0.659	0.511	-12.691	6.336
tipo_processador_Intel	-10.2955	8.310	-1.239	0.217	-26.690	6.099
=====						
Omnibus:	84.737	Durbin-Watson:	1.907			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	245.418			
Skew:	1.899	Prob(JB):	5.11e-54			
Kurtosis:	6.991	Cond. No.	4.20e+18			
=====						

3. Diagnóstico de Multicolinearidade:

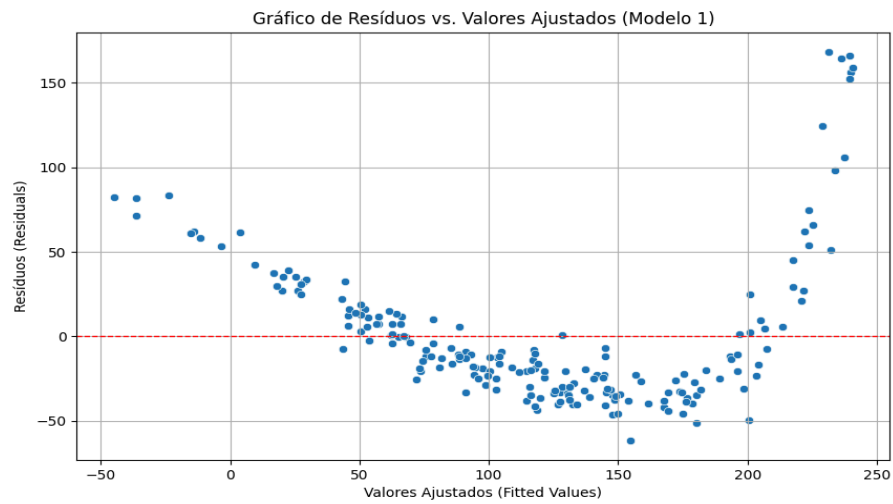
Foi utilizado o cálculo de VIF para todas as variáveis. As variáveis: `sistema_operacional_MacOS` e `tipo_processador_Apple Silicon` apresentaram $VIF = \text{inf}$, indicando colinearidade perfeita. Foi realizada a remoção dessas variáveis consideradas essenciais para o Modelo 2.

VIF Ordenados do Maior para o Menor:		
	feature	VIF
4	<code>sistema_operacional_MacOS</code>	inf
7	<code>tipo_processador_Apple Silicon</code>	inf
0	<code>cpu_cores</code>	4.319992
2	<code>latencia_ms</code>	3.489667
3	<code>armazenamento_tb</code>	2.713309
6	<code>tipo_hd_SSD</code>	2.285290
1	<code>ram_gb</code>	2.117906
5	<code>sistema_operacional_Windows</code>	2.078571
8	<code>tipo_processador_Intel</code>	2.016325

4. Diagnóstico de Heterocedasticidade:

A verificação da heterocedasticidade foi realizada por meio do teste de Breusch-Pagan, cuja hipótese nula assume a homocedasticidade dos resíduos. Os resultados indicaram uma estatística F de 3.2852 e um p-valor de 0.0016. Como o p-valor é inferior ao nível de significância de 5%, rejeitamos a hipótese nula e concluímos que há evidência estatística de heterocedasticidade no modelo.

Esse resultado indica que a variância dos erros não é constante, o que pode comprometer a eficiência dos estimadores de mínimos quadrados. Para confirmar visualmente essa condição, também foi construído o gráfico de resíduos versus valores ajustados.



```
--- Teste de Breusch-Pagan (Modelo 1) ---  
Estatística LM: 24.1319  
p-valor LM: 0.0041  
Estatística F: 3.2852  
p-valor F: 0.0016  
Interpretação: Com p-valor F (0.0016) < 0.05, rejeitamos H0.
```

Parte III - Análise Crítica

1. Modelo 2 (Com Exclusão de Variável)

- Variável excluída: tipo_processador_Apple Silicon (VIF = inf).
- Modelo ajustado com o restante das variáveis.

```
--- Sumário Modelo 2 (Excluindo tipo_processador_Apple Silicon) ---
                        OLS Regression Results
=====
Dep. Variable:          tempo_resposta      R-squared:                0.694
Model:                  OLS                 Adj. R-squared:           0.681
Method:                 Least Squares       F-statistic:             52.44
Date:                   Mon, 05 May 2025    Prob (F-statistic):      1.29e-43
Time:                   00:24:37           Log-Likelihood:          -1008.5
No. Observations:       194                AIC:                     2035.
Df Residuals:           185                BIC:                     2064.
Df Model:                8
Covariance Type:        nonrobust
=====
                        coef      std err          t      P>|t|      [0.025      0.975]
-----
const                   276.9071     12.443     22.253     0.000     252.358     301.456
cpu_cores               -14.0524      0.808    -17.385     0.000    -15.647    -12.458
ram_gb                  -1.5459      0.151    -10.232     0.000     -1.844     -1.248
latencia_ms              0.0001      0.039      0.004     0.997     -0.076      0.076
armazenamento_tb        -3.2488      5.423     -0.599     0.550    -13.947      7.450
sistema_operacional_MacOS -6.3549      9.644     -0.659     0.511    -25.381     12.671
sistema_operacional_Windows -0.1981      8.292     -0.024     0.981    -16.557     16.161
tipo_hd_SSD              4.6192      6.589      0.701     0.484     -8.380     17.618
tipo_processador_Intel   -10.2955      8.310     -1.239     0.217    -26.690      6.099
=====
Omnibus:                84.737    Durbin-Watson:           1.907
Prob(Omnibus):           0.000    Jarque-Bera (JB):        245.418
Skew:                    1.899    Prob(JB):                 5.11e-54
Kurtosis:                6.991    Cond. No.                 787.
=====
```

2. Comparativo entre Modelos

Foram ajustados dois modelos para comparação. O Modelo 1 inclui todas as variáveis explicativas disponíveis, enquanto o Modelo 2 exclui a variável tipo_processador_Apple Silicon devido à multicolinearidade perfeita detectada no diagnóstico com VIF.

```
--- Comparação dos Modelos ---
Modelo 1: R² Adj = 0.6807, F-stat = 52.44, p(F) = 1.29e-43, Cond. No. = 4.20e+18
Modelo 2: R² Adj = 0.6807, F-stat = 52.44, p(F) = 1.29e-43, Cond. No. = 787.36
```

3. Justificativa

A exclusão da variável com colinearidade perfeita melhora a interpretação e estabilidade dos coeficientes. Apesar da variável ser significativa no Modelo 1, a multicolinearidade distorce sua influência.