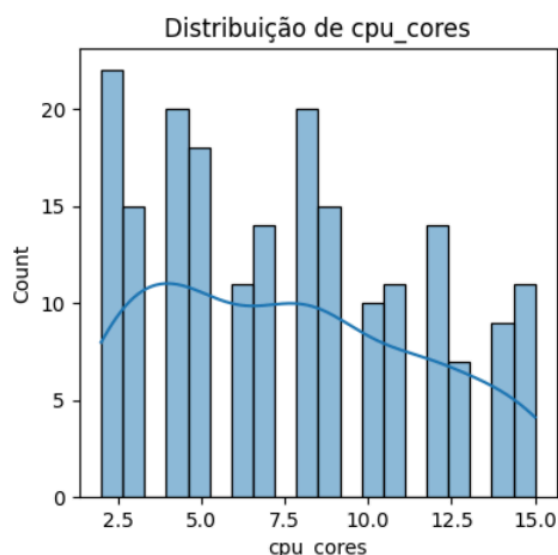


Parte I – Análise Estatística

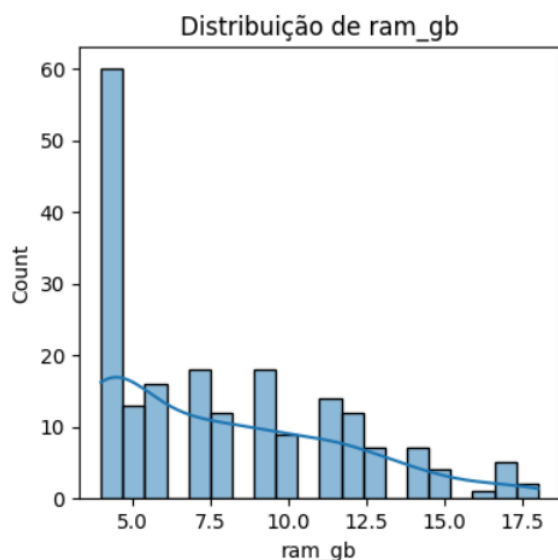
CPU Cores

A variável **cpu_cores** apresenta uma média de 7,56 núcleos e uma mediana de 7 núcleos, com desvio padrão de aproximadamente 3,97. A distribuição é ligeiramente assimétrica à direita, indicando que, embora a maioria dos sistemas possua entre 4 e 11 núcleos, há casos com números mais elevados (até 15 núcleos) que elevam a média.



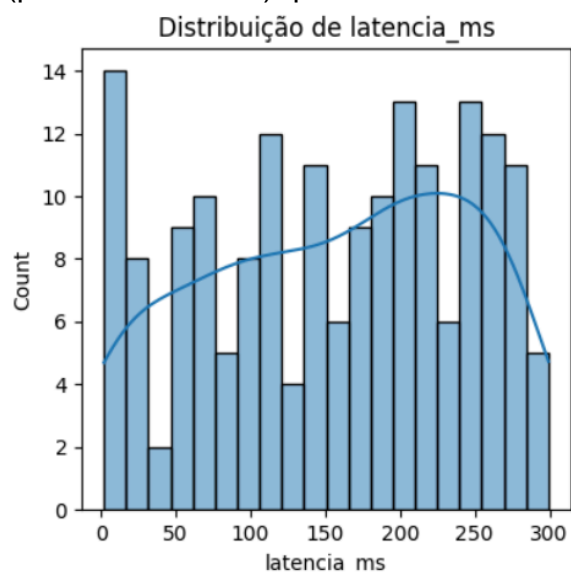
RAM (GB)

A variável **ram_gb** tem uma média de 7,87 GB e mediana de 7 GB, com desvio padrão de 3,78. A distribuição é praticamente simétrica, sugerindo uma concentração dos sistemas em torno de valores típicos como 4 GB a 11 GB de memória RAM, com alguns poucos valores mais elevados (16 GB e 18 GB).



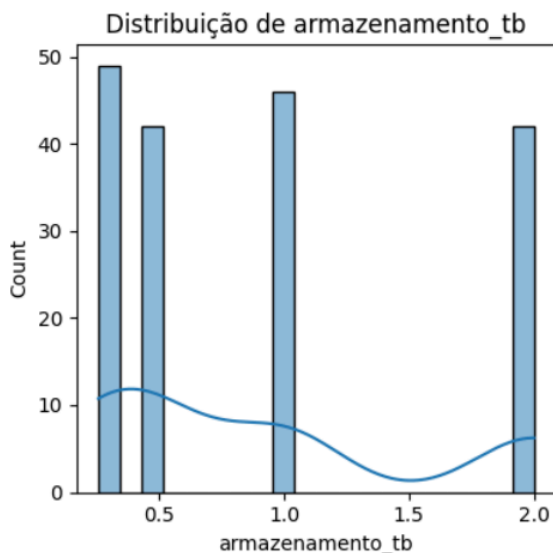
Latência (ms)

A **latencia_ms** possui média de 156,52 ms e mediana de 166,46 ms, com um elevado desvio padrão de 87,53 ms. A variável apresenta uma distribuição assimétrica à esquerda, o que indica a presença de alguns sistemas com latência extremamente baixa (por volta de 2 ms) que acabam reduzindo a média geral.



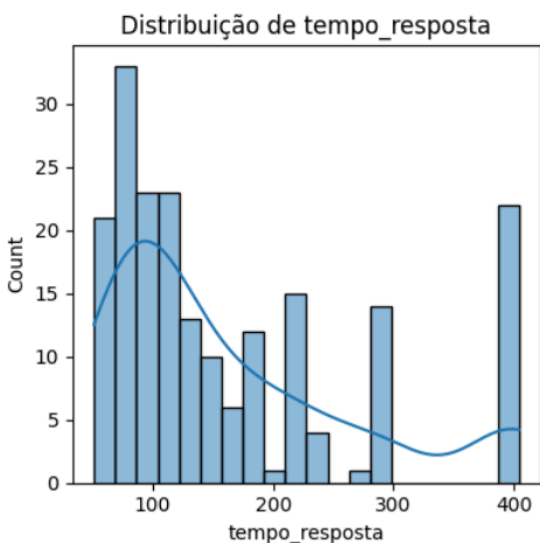
Armazenamento (TB)

A variável **armazenamento_tb** apresenta média de 0,91 TB e mediana de 0,512 TB, com desvio padrão de 0,66. A distribuição é fortemente assimétrica à direita, indicando que a maioria dos sistemas possui menos de 1 TB de armazenamento, mas existem casos com capacidade de até 2 TB, o que eleva a média.



Tempo de Resposta (ms)

A variável **tempo_resposta**, a dependente principal da análise, apresenta uma média de 163,85 ms e mediana de 120,15 ms, com desvio padrão de 105,27 ms. A distribuição é fortemente assimétrica à direita, indicando que, embora a maioria dos sistemas tenha tempos de resposta relativamente baixos, existem casos com tempos de resposta significativamente elevados (até 404,27 ms), impactando a média de forma significativa.



Considerações Finais

De forma geral, as variáveis relacionadas ao hardware (**cpu_cores** e **ram_gb**) apresentaram distribuições próximas da simetria, enquanto as variáveis de desempenho (**latencia_ms**, **armazenamento_tb** e **tempo_resposta**) apresentaram distribuições assimétricas e alta variabilidade. Isso sugere que existem sistemas muito distintos em termos de performance no conjunto de dados, o que poderá influenciar as análises posteriores.

Parte II – Modelo e Diagnóstico

Ajuste do Modelo

- A variável dependente foi **tempo_resposta**, e as variáveis explicativas incluíram características do hardware e do sistema.
- As variáveis categóricas **sistema_operacional**, **tipo_hd** e **tipo_processador** foram convertidas em variáveis dummies com o parâmetro “**drop_first=True**”, para evitar a armadilha das dummies (dummy trap).
- As categorias base (omitidas do modelo) foram:
sistema_operacional: Linux
tipo_hd: HDD
tipo_processador: AMD
- O modelo apresentou **R² = 0.801** e **R² ajustado = 0.791**, indicando que cerca de 80% da variação no tempo de resposta foi explicada pelas variáveis independentes.
- As variáveis **cpu_cores** (coeficiente negativo) e **ram_gb** (coeficiente positivo) foram estatisticamente significativas ($p < 0.05$), enquanto as demais não apresentaram significância estatística.

Coefficientes do Modelo de Regressão

Variável	Coefficiente	Erro Padrão	Estatística t	Valor-p	Intervalo de Confiança (95%)
Intercepto (const)	321,52	13,33	24,12	0,000	[295,21 ; 347,84]
cpu_cores	-35,27	2,77	-12,75	0,000	[-40,73 ; -29,81]
ram_gb	13,78	2,91	4,73	0,000	[8,03 ; 19,53]
latencia_ms	0,009	0,044	0,20	0,845	[-0,079 ; 0,096]
armazenamento_tb	-2,73	5,72	-0,48	0,633	[-14,03 ; 8,56]
sistema_operacional_MacOS	4,70	10,13	0,46	0,643	[-15,29 ; 24,70]
sistema_operacional_Windows	13,56	8,86	1,53	0,128	[-3,93 ; 31,05]
tipo_hd_SSD	-2,03	7,30	-0,28	0,782	[-16,44 ; 12,39]
tipo_processador_Intel	-8,88	8,80	-1,01	0,314	[-26,26 ; 8,49]

Heterocedasticidade e Multicolinearidade

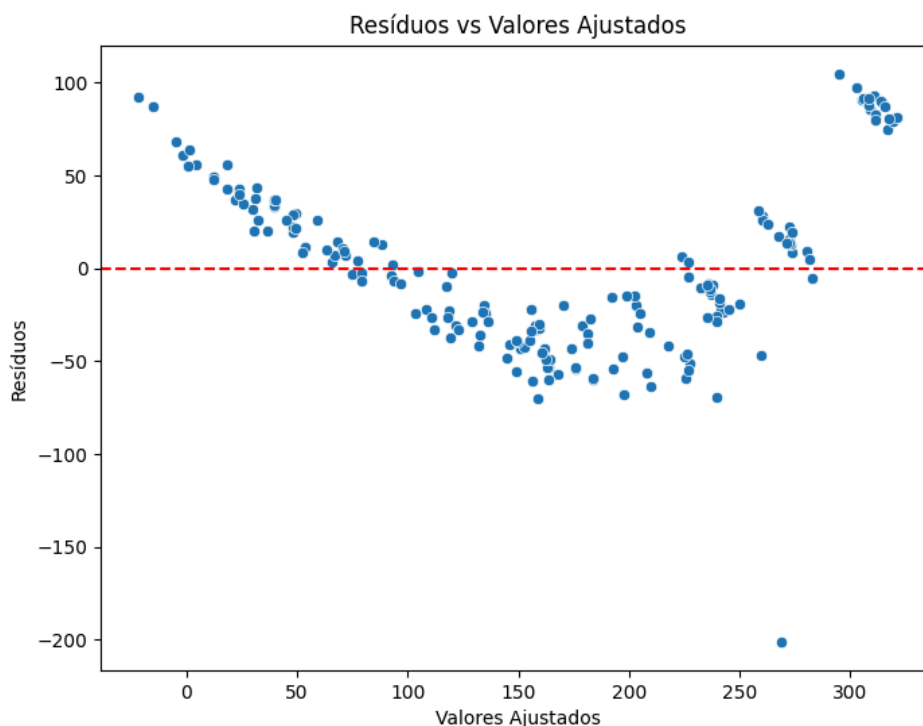
Para verificar a presença de multicolinearidade entre as variáveis explicativas, foi calculado o Fator de Inflação da Variância (VIF). Os resultados iniciais indicaram **VIF** infinito para as variáveis **`sistema_operacional_MacOS`** e **`tipo_processador_Apple Silicon`**, caracterizando multicolinearidade perfeita entre elas. Como consequência, optou-se pela exclusão da variável **`tipo_processador_Apple_Silicon`** do modelo.

Após a remoção, todos os VIFs ficaram abaixo de 10, com a maioria inferior a 2, indicando a ausência de multicolinearidade severa e maior estabilidade estatística no modelo final.

Variável	VIF
const	13.795218
cpu_cores	9.506407
ram_gb	9.55087
latencia_ms	1.018513
armazenamento_tb	1.036096
sistema_operacional_MacOS	1.712352
sistema_operacional_Windows	1.379399
tipo_hd_SSD	1.032656
tipo_processador_Intel	1.339888

Em seguida, foi realizado o teste de Breusch-Pagan para verificar a presença de heterocedasticidade (variância não constante dos resíduos). O resultado apresentou um p-valor de 3.16×10^{-8} , indicando forte evidência de heterocedasticidade no modelo.

Esse resultado foi reforçado pela análise gráfica do gráfico de resíduos versus valores ajustados, no qual foi identificado um padrão em forma de “U” invertido. Esse padrão sugere que a variância dos resíduos aumenta e depois diminui ao longo dos valores preditos, violando a suposição de homocedasticidade da regressão linear.



Parte III – Comparação de Modelos e Análise Crítica

Modelo 1 (Completo): Incluiu todas as variáveis explicativas do dataset, incluindo ``latencia_ms``.

Modelo 2 (Reduzido): Removeu a variável ``latencia_ms``, uma vez que apresentou valor-p de 0,845, indicando que seu efeito no tempo de resposta não foi estatisticamente significativo.

Métrica	Modelo 1 (com latencia_ms)	Modelo 2 (sem latencia_ms)
R ²	0.801	0.801
R ² Ajustado	0.791	0.793
Estatística F	85.44	98.19
AIC	1903.0	1901.0
BIC	1931.0	1926.0
Nº de observações	179.0	179.0

A métrica R² permaneceu inalterada, mas o R² ajustado aumentou, enquanto os critérios de informação AIC e BIC diminuíram, o que indica melhor desempenho ajustado do Modelo 2. A estatística F também aumentou, reforçando a robustez estatística do modelo reduzido.

Ações práticas sugeridas

- A variável ``cpu_cores`` foi estatisticamente significativa e apresentou coeficiente negativo, indicando que um número maior de núcleos está associado a uma redução no tempo de resposta. Assim, recomenda-se priorizar processadores com mais núcleos para melhorar a performance.
- A variável ``ram_gb`` foi significativa, com coeficiente positivo. Isso pode indicar que, neste contexto específico, mais RAM não necessariamente reduziu o tempo de resposta, possivelmente devido à ineficiência no uso de memória ou gargalos em outras partes do sistema.

Conclui-se que o Modelo 2 é mais eficiente e estável, e que o principal fator técnico associado à redução do tempo de resposta é o número de núcleos da CPU.