


# Atividade Guiada – Explorando o dataset *Wine Quality*

## 1. Preparação do ambiente


1. Importe as bibliotecas necessárias: `pandas`, `numpy` e `seaborn`.
2. Carregue o dataset *Wine Quality* (versão Red ou White, em CSV).
  - [Link direto para download](#)

 *Pergunta de reflexão:* Qual biblioteca você usaria para carregar dados tabulares: `numpy` ou `pandas`? Por quê?

---

## 2. Primeira exploração dos dados


1. Verifique as primeiras linhas do dataset.
2. Descubra quantas linhas e colunas ele possui.
3. Liste os nomes das colunas.

 *Pergunta de reflexão:* O que você consegue deduzir apenas olhando as colunas e tipos de dados?

---

## 3. Estatísticas descritivas

1. Calcule a **média do teor alcoólico** usando `pandas`.
2. Agora, calcule a mesma média usando `numpy`.
3. Compare os resultados.

 *Pergunta de reflexão:* Qual método foi mais rápido/intuitivo? Em que situação você usaria o `np.mean` em vez do `df.mean()`?

---

## 4. Trabalhando com valores faltantes

1. Verifique se existem valores `NaN` no dataset.
2. Se houver, pense em duas formas de lidar com eles:
  - Usando `pandas`.
  - Usando `numpy` (`np.nanmean`).

💡 *Pergunta de reflexão:* Qual abordagem você considera mais clara para tratar NaN?

---

## 5. Distribuição e visualização

1. Crie um gráfico que mostre a **distribuição da variável “alcohol”**.
2. Crie um gráfico que mostre a relação entre **álcool e qualidade** do vinho.
3. Crie um boxplot para verificar possíveis **outliers em “pH”**.

💡 *Pergunta de reflexão:* O que esses gráficos te ajudam a entender que as médias sozinhas não mostram?

---

## 6. Estude o que impacta a variável qualidade

1. Verifique a frequência da variável alvo (qualidade)
  2. Crie uma matriz de correlação
  3. Plot isso com sns
- 

## 7. Conclusão

1. Escreva um parágrafo respondendo:
  - O que você descobriu sobre a relação entre o **teor alcoólico e a qualidade** do vinho?
  - Quais são as variáveis que mais impactam a qualidade do vinho?
  - Quando faria sentido usar **numpy** e quando usaria **pandas**?