

Desafio Técnico para Estagiário em Ciência de Dados com Python

Objetivo: Este desafio tem como objetivo avaliar suas habilidades básicas em Python e sua capacidade de analisar e manipular dados utilizando bibliotecas como Pandas e Matplotlib.

Base de Dados: Você deve usar o uso do conjunto de dados "Adult" disponível no repositório UCI Machine Learning, que contém informações sobre indivíduos do censo dos EUA. O conjunto de dados pode ser encontrado em: <https://archive.ics.uci.edu/dataset/2/adult>

Tarefas:

1. Análise Exploratória de Dados:

- Carregue o conjunto de dados usando Pandas.
- Explore os dados exibindo as primeiras e últimas linhas do dataset.
- Exiba informações estatísticas sobre todas as colunas numéricas presentes no conjunto de dados. Algumas estatísticas são: contagem, média, desvio padrão, valor mínimo, valor máximo, quantil 25%, quantil 50%, quantil 75% e valor máximo.
- Apresente um resumo estatístico das variáveis numéricas.

2. Visualização de Dados:

- Utilize a biblioteca Matplotlib para criar gráficos que ajudem a entender as características dos dados. Considerando as colunas: Idade, sexo e raça.
- Crie um histograma para pelo menos uma variável numérica.
- Gere um gráfico de barras para ilustrar a distribuição de pelo menos uma variável categórica.

3. Manipulação e Análise de Dados:

- Verifique se há valores ausentes no dataset. Se houver valores ausentes, remova todas as linhas em que há pelo menos 1 coluna ausente.

4. Insights e Conclusões:

- Com base na sua análise, apresente insights sobre os dados.
- Quais são suas conclusões após explorar o conjunto de dados?

Entrega: Você deve apresentar um Jupyter Notebook disponibilizado no github contendo todo o código usado para a análise, juntamente com comentários e explicações sobre os passos realizados. Sua análise deve ser clara e compreensível.

Você deve explicar qual a base de dados e quais desafios ela poderia resolver.