REGRESSÃO

1. **Compreensão e Preparação dos Dados**

Critério	Comentários
Tratamento adequado de dados ausentes, duplicados ou inconsistências	Verificação muito boa e adequada com o uso do isnull().sum() e do df.info() Sugestão: usar também o df.duplicated() para verificar a existência de tuplas duplicadas
Visualização e análise exploratória (ex: histogramas, correlações, boxplots)	Apresentou as visualizações corretas, mas não fez uma análise ou interpretação clara dos resultados dessas visualizações. Sugestão: não basta plotar os gráficos, é sempre interessante vir acompanhado de uma análise textual sobre o que está sendo mostrado
Normalização ou padronização dos dados (quando aplicável)	Foi usada StandardScaler corretamente em todos os modelos por meio de Pipeline, o que pode ser adequado para os algoritmos utilizados.

2. Modelagem e Técnica de Regressão

Critério	Comentários
Escolha de pelo menos um algoritmo de regressão (ex: LinearRegression, RandomForest, etc.)	Parabéns! O Trainee utilizou três modelos diferentes: Regressão Linear, Ridge e Random Forest, o que demonstra bom domínio e diversidade de técnicas.
Separação adequada entre dados de treino/teste ou uso de validação cruzada	Foi usada a função train_test_split() com 70/30 e random_state=42, garantindo reprodutibilidade e boa prática de separação.
Justificativa da escolha do modelo e/ou comparação entre modelos	A escolha dos modelos foi bem fundamentada: Regressão Linear e Ridge como abordagens básicas e Random Forest como modelo mais robusto. Também foi feita a comparação de desempenho entre eles, além de análise com PCA.

3. Métricas de Avaliação e Resultados

Critério	Comentários
Apresentação de métricas adequadas (ex: MSE, RMSE, MAE, R^2)	Foi utilizado o RMSE e o R^2 em todos os modelos. Já está muito bom, mas sugiro que use pelo menos mais uma métrica, para que se tenha uma interpretação mais fiel e completa ainda
Interpretação correta dos resultados obtidos	**A interpretação dos dados deixou a desejar. Como um todo no trabalho, deve-se buscar destrinchar mais as análises de cada dado, gráfico, métrica e etc. Como conclusão, não basta apenas colocar qual foi o melhor modelo, mas sim explicar se cada resultado foi bom/moderado/ruim. Por exemplo: mesmo o Random Forest tendo o melhor resultado, o R^2 ainda ficou muito distante de 1. Além disso, a interpretação de que a aplicação do PCA manteve um bom desempenho, está incorreta. O próprio aluno colocou que o RMSE aumentou com a aplicação do PCA**

4. Código, Organização e Reprodutibilidade

Critério	Comentários
Código bem estruturado, limpo e com comentários explicativos	O código está bem limpo e estruturado, entretanto sugiro adicionar mais comentários no código. Além disso, interpretação e análise fazem parte também de um bom código.
Execução reprodutível (uso de `requirements.txt`, notebooks ou scripts funcionais)	O uso do notebook foi suficiente para a execução.

CLASSIFICAÇÃO

1. Comparação de Modelos (OBRIGATÓRIO)

Critério	Comentários
Treinou corretamente os dois classificadores (Árvore de Decisão e Random Forest)	Sim, os dois modelos foram treinados corretamente usando Pipeline com preprocessor e classifier
Avaliou ambos com todas as métricas (Acurácia, Precisão, Recall, F1-score)	Parabéns! Foram usadas accuracy_score, precision_score, recall_score e f1_score.
Interpretou corretamente os valores das métricas	A interpretação está implícita na tabela comparativa, entretanto é interessante uma interpretação mais detalhada e textual.
Gerou e analisou a matriz de confusão	Apenas gerou, mas não analisou
Gerou e interpretou o gráfico de importância dos atributos	Gerou e analisou, chegando a conclusão que educação e idade mostraram alta influência
Código limpo, bem organizado e com comentários explicativos	Código muito bom, limpo e bem estruturado. A única sugestão é para que se adicione mais comentários a fim de melhorar a documentação

2. Análise Exploratória dos Dados (OPCIONAL)

Critério	Comentários
Visualizou corretamente as distribuições (idade, horas, ganhos, etc.)	Sim, foi feita a análise de distribuição de idade por renda e renda geral.
Investigou correlações e relação entre atributos e classe alvo	Análise de renda por educação foi feita, mas não houve uso direto de correlação numérica (ex: df.corr()).
Fez boas observações ou insights sobre os dados	Os gráficos ajudam a levantar hipóteses, mas não há discussão direta de insights textuais no código.

3. Validação Cruzada (OPCIONAL)

Critério	Comentários
Aplicou corretamente ao menos duas estratégias (hold-out, K-fold ou stratified)	Parabéns, muito bem aplicado
Comparou os resultados e discutiu impactos (viés e variância)	-

4. Redução de Dimensionalidade (OPCIONAL)

Critério	Comentários
Aplicou corretamente normalização e codificação dos dados	Foi aplicado StandardScaler para dados numéricos e OneHotEncoder para categóricos dentro de um Pipeline.
Aplicou PCA e analisou seu impacto no desempenho	-

OBSERVAÇÕES FINAIS

Critério	Comentários
Documentação clara e explicativa (ex: README, comentários no notebook)	Os códigos ficaram excelentes, entretanto recomendo que adicione mais comentários aos códigos
Visualizações e tabelas ajudam na análise e compreensão dos dados	Muitas tabelas importantes para a análise foram de fato plotadas e acredito que analisadas, sendo fundamentais para chegar-se nos resultados esperados. Porém, recomendo fortemente que sempre documente suas interpretações. Isso é fundamental para que outras pessoas que queiram visualizar seus trabalhos entendam o que está sendo feito e qual foi seu entendimento a respeito de uma tabela/gráfico/métrica/etc.