**Exercício: Pré-processamento e Visualização com Bank Marketing Dataset**

**Contexto**

Você recebeu o dataset **Bank Marketing** que contém informações sobre clientes de um banco e campanhas de marketing feitas para esses clientes. Seu objetivo é preparar os dados para que possam ser usados em modelos de aprendizado de máquina, além de explorar e interpretar visualmente o comportamento dos dados.

**Passo a passo do exercício**

**1. Carregue os dados**

* Faça a leitura do arquivo CSV usando pandas.
* Exiba as primeiras linhas para entender a estrutura do dataset.
* Use métodos como .info() e .describe() para obter um resumo geral das colunas, tipos e estatísticas.

**2. Trate dados inconsistentes e faltantes**

* Verifique se existem valores faltantes (null/NaN) em qualquer coluna.
* Caso existam, aplique estratégias adequadas para preencher esses valores (por exemplo, média para numéricas, moda para categóricas).
* Verifique se há valores inconsistentes, por exemplo, valores negativos em colunas numéricas como age ou balance. Caso encontre, substitua por valores adequados (como a média ou mediana).

**3. Separe as variáveis preditoras (X) e a variável alvo (Y)**

* Defina X contendo todas as colunas exceto a variável alvo.
* Defina Y contendo a coluna y (que indica se o cliente aderiu à campanha ou não).

**4. Faça a codificação das variáveis categóricas**

* Identifique quais colunas são categóricas.
* Utilize LabelEncoder para variáveis binárias.
* Utilize OneHotEncoder para variáveis categóricas nominais com múltiplas categorias.
* Utilize ColumnTransformer para aplicar a codificação corretamente mantendo as outras colunas numéricas.

**5. Normalização**

* Aplique MinMaxScaler para normalizar os atributos numéricos entre 0 e 1.

**6. Visualize os dados**

* Faça um gráfico de barras para a variável alvo y para verificar a distribuição entre clientes que aceitaram e não aceitaram a campanha.
* Plote histogramas para variáveis numéricas importantes como age, balance e duration.
* Use gráficos de dispersão (scatter) para analisar a relação entre duas variáveis numéricas, por exemplo, age e balance.
* Faça gráficos de caixa (boxplot) para identificar possíveis outliers em variáveis numéricas.
* Use gráficos de barras ou treemap para explorar a distribuição das principais variáveis categóricas, como job, marital e education.
* Após cada gráfico, escreva uma breve interpretação do que aquele gráfico revela sobre o comportamento dos dados.

**7. Divida os dados entre treino e teste**

* Use train\_test\_split para separar 80% dos dados para treino e 20% para teste, com random\_state fixo para reprodutibilidade.

**8. Salve os dados pré-processados**

* Salve os dados de treino e teste, tanto X quanto Y, em um arquivo .pkl usando a biblioteca pickle.

**9. (Bônus para interpretação)**

* Quais variáveis parecem mais relacionadas com o resultado da campanha?
* Como as variáveis categóricas influenciam o sucesso da campanha? Alguma categoria se destaca?