Questionário

1. **Consideramos valores *outliers* valores que:**
   a) Estão associados a erros na recolha de informação.
   b) Estão desviados da média mais do 1.5 IQR.
   c) Devem ser eliminados, pois impactam nas métricas utilizadas nos modelos (ex.: média).
   d) São atípicos dada a distribuição da variável à qual pertence.

2. **Verdadeiro ou Falso: "Os *outliers* são um problema a resolver antes de iniciar a modelação, seja via eliminação seja via substituição de valores".**
   a) Verdadeiro.
   b) Falso.

3. **As transformações não-lineares, como logaritmo e raiz quadrada,…**
   a) Alteram a distribuição das variáveis e podem ter impacto (positivo ou negativo) na sua capacidade explicativa.
   b) São as melhores a resolver os problemas de outliers.
   c) Devem ser usadas com cautela por reduzem a sua correlação com a target.
   d) São ótimas normalizar as variáveis.

4. **Por vezes pode ser relevante criar novas variáveis adicionais, como por exemplo variáveis Dummy. Que tipo de variáveis são essas e para que servem?**
   a) São variáveis que indicam se o *dataframe* tem *outliers*.
   b) São variáveis que indicam se o dataframe tem *missing values*.
   c) São variáveis que tomam o valor *True* ou *False* e pode ser usadas para tornar quantitativas as variáveis qualitativas.
   d) São variáveis que tomam o valor *True* ou *False* e pode ser usadas para tornar qualitativas as variáveis quantitativas.

5. **Através de um gráfico de *Scatter Plot* que tipo de informação consigo recolher?**
   a) A distribuição da variável, se tem *missing values* e *outliers*.
   b) A correlação entre 2 ou mais variáveis.
   c) Identificar potencial correlação linear entre 2 variáveis.
   d) Todas as anteriores.

6. **Imagina que te é disponibilizado um *dataframe* com um conjunto de dados sobre dados biomédicos de pacientes que estão a ser alvo de um tratamento experimental e que tem como missão perceber qual a dosagem mais adequada para a melhor recuperação dos pacientes.**

   Descreve, e justifica, quais os passos de análise, preparação e visualização de dados irias realizar para garantir que o teu *dataframe* está em condições para a realização do teu estudo.

Pontuação: A pergunta 6 = 50%, restantes perguntas = 10%