### **Roteiro: Análise e Otimização de Agrupamentos com Seleção de Atributos**

**Objetivo:** Realizar uma análise de agrupamento usando o algoritmo K-Means em um conjunto de dados de monitoramento de exercícios em uma academia. O objetivo foi melhorar a qualidade dos agrupamentos aplicando uma técnica de seleção de atributos.

#### **Passo a Passo da Análise**

1. **Carregamento e Inspeção dos Dados**
   * Primeiro, carregamos o conjunto de dados gym\_members\_exercise\_tracking.csv para entender sua estrutura e identificar variáveis categóricas e numéricas.
   * Observamos que o dataset continha informações como idade, gênero, peso, altura, BPMs (frequência cardíaca), tipo de treino, duração da sessão, porcentagem de gordura corporal, e outros.
2. **Seleção de Atributos (Feature Selection)**
   * Para identificar os atributos menos relevantes, utilizamos a **correlação** de cada variável com a coluna Calories\_Burned, definida como a variável-alvo.
   * Calculamos a correlação, excluindo atributos com baixa correlação com a variável-alvo. Atributos como **Max\_BPM**, **Resting\_BPM**, **BMI**, e **Height (m)** mostraram pouca correlação com Calories\_Burned e foram considerados irrelevantes para o modelo.
3. **Tratamento de Dados Categóricos e Normalização**
   * Convertemos variáveis categóricas (Gender e Workout\_Type) em valores numéricos, usando mapeamento para Gender e codificação one-hot para Workout\_Type.
   * Após a remoção de atributos irrelevantes e tratamento de variáveis categóricas, aplicamos a **normalização** com StandardScaler para garantir que todos os atributos tivessem a mesma escala.
4. **Agrupamento Inicial e com Dados Reduzidos**
   * Com o dataset completo, treinamos o algoritmo **K-Means** com 3 clusters e avaliamos a qualidade dos agrupamentos usando o **índice de silhueta**, obtendo um valor de **0.208**.
   * Após a remoção dos atributos de baixa correlação, retreinamos o modelo K-Means e obtemos um índice de silhueta de **0.218**, indicando uma leve melhora na qualidade dos agrupamentos.

#### **Resumo do Padrão Escolhido e Conclusão**

A partir dos resultados, foi possível concluir que os atributos **Max\_BPM**, **Resting\_BPM**, **BMI**, e **Height (m)** tinham baixa relevância para a previsão de calorias queimadas e contribuíam pouco para a qualidade dos agrupamentos. A remoção desses atributos permitiu que o modelo se concentrasse em características mais relevantes, o que resultou em uma leve melhoria no índice de silhueta (de 0.208 para 0.218).

**Conclusão:** A análise mostrou que a eliminação de atributos de baixa correlação simplificou o modelo, melhorando ligeiramente a coerência dos agrupamentos e tornando o modelo mais eficiente. A seleção de atributos é, portanto, uma etapa importante para otimizar modelos de agrupamento, removendo ruídos e focando nas características mais relevantes.