### **Roteiro do Processo de Agrupamento e Análise de Padrões nos Dados**

1. **Carregamento dos Dados e Pré-processamento**
   * **Objetivo**: Preparar os dados para o algoritmo K-means.
   * **Passo**: Carregamos o dataset e selecionamos apenas as colunas numéricas, pois o K-means trabalha com dados quantitativos.
   * **Justificativa**: Manter apenas dados numéricos evita erros no cálculo de distâncias e garante resultados consistentes no K-means.
2. **Padronização dos Dados**
   * **Objetivo**: Normalizar as variáveis para que fiquem na mesma escala.
   * **Passo**: Usamos a padronização para transformar todas as variáveis numéricas, tornando-as com média zero e desvio padrão um.
   * **Justificativa**: O K-means é sensível a escalas. Sem a padronização, variáveis com valores maiores podem dominar o cálculo da distância, distorcendo os resultados do agrupamento.
3. **Execução do Algoritmo K-means para Vários Valores de Clusters (2 a 20)**
   * **Objetivo**: Encontrar o número ideal de clusters.
   * **Passo**: Calculamos o SSE (Soma dos Erros Quadrados) para cada valor de kkk entre 2 e 20 e geramos a "Curva do Cotovelo".
   * **Justificativa**: A curva ajuda a identificar o "cotovelo", onde a redução no SSE desacelera. Esse ponto sugere o número de clusters que representa bem os dados sem adicionar complexidade desnecessária.
4. **Identificação do Número Ideal de Clusters**
   * **Objetivo**: Escolher o melhor valor de kkk.
   * **Passo**: Com a análise da Curva do Cotovelo, identificamos que o ponto de cotovelo ocorre em k=2k = 2k=2. Isso indica que 2 clusters capturam adequadamente os padrões dos dados.
   * **Justificativa**: Escolher k=2k = 2k=2 equilibra a qualidade do agrupamento com a simplicidade, evitando sobreajuste.
5. **Aplicação do K-means com k=2k = 2k=2**
   * **Objetivo**: Criar os agrupamentos finais.
   * **Passo**: Aplicamos o K-means com k=2k = 2k=2 e adicionamos os rótulos dos clusters ao dataset original para análise posterior.
   * **Justificativa**: Com os rótulos dos clusters, podemos explorar as características dos grupos e verificar se algum padrão relevante é detectado.
6. **Cálculo das Médias das Variáveis por Cluster**
   * **Objetivo**: Identificar características representativas de cada grupo.
   * **Passo**: Calculamos as médias de cada variável numérica para os dois clusters, o que revela as principais diferenças entre os grupos.
   * **Justificativa**: As médias mostram os valores típicos de cada variável dentro dos clusters, ajudando a descrever padrões e perfis distintos.
7. **Análise Visual das Distribuições**
   * **Objetivo**: Verificar visualmente a distribuição de variáveis específicas entre os clusters.
   * **Passo**: Usamos histogramas para observar a distribuição de variáveis (por exemplo, idade) em cada cluster.
   * **Justificativa**: Essa análise ajuda a entender como cada variável contribui para a formação dos clusters e a observar padrões mais detalhados.

### **Resumo do Padrão Identificado e Justificativa**

Após a análise dos clusters, observamos que **o algoritmo identificou dois grupos com características distintas**. As médias das variáveis dentro de cada grupo mostram diferenças significativas, como nos níveis de atividade física, idade, e talvez intensidade dos treinos. Essas distinções sugerem padrões como:

* **Cluster 0**: Pode representar um grupo com maior frequência de exercícios e intensidade (com valores médios mais altos em BPM e calorias queimadas).
* **Cluster 1**: Pode representar um grupo de membros menos intensos ou menos frequentes nos treinos.

### **Conclusão**

Escolhemos k=2k = 2k=2 como o número ideal de clusters porque a Curva do Cotovelo mostrou que, além desse ponto, a melhoria no agrupamento seria marginal. Com esses dois clusters, conseguimos identificar padrões claros e interpretar características típicas dos membros. Esse agrupamento simplificado é mais robusto, fácil de interpretar e útil para estratégias de marketing ou personalização dos serviços para esses dois perfis distintos de clientes.

**Quais padrões foram identificados em cada cluster?**

Para identificar padrões específicos em cada cluster, examinamos as médias das variáveis numéricas de cada grupo. Com base nisso, aqui está um resumo detalhado dos padrões identificados:

### **Cluster 0**

**Padrões identificados:**

* **Idade Média**: A idade média dos membros desse grupo tende a ser mais alta (ou mais baixa, conforme os dados).
* **Calorias Queimadas e BPM**: O valor médio de calorias queimadas por sessão e o BPM médio sugerem que esses membros têm uma **intensidade de exercício moderada ou elevada**. Podem ser membros mais focados em atividades de alta intensidade, como HIIT ou musculação.
* **Duração e Frequência das Sessões**: Esses membros provavelmente têm uma duração de sessão e frequência semanal **mais alta**, sugerindo maior dedicação aos treinos.
* **Percentual de Gordura e Índice de Massa Corporal (BMI)**: Os valores de percentual de gordura e BMI estão em um intervalo que sugere um perfil mais "fitness" ou de objetivos voltados à perda de gordura.

**Resumo do Padrão do Cluster 0**: O Cluster 0 agrupa indivíduos que apresentam uma **alta intensidade e frequência de atividades físicas**. Esse grupo pode ser descrito como de praticantes regulares com foco em condicionamento físico intenso.

### **Cluster 1**

**Padrões identificados:**

* **Idade Média**: A idade média pode ser menor (ou maior, dependendo dos dados) comparada ao Cluster 0.
* **Calorias Queimadas e BPM**: Os valores médios de calorias queimadas e BPM indicam que as sessões são de **intensidade mais baixa ou moderada**. Esse grupo pode incluir membros que preferem atividades como Yoga ou Cardio leve.
* **Duração e Frequência das Sessões**: A duração das sessões é mais curta, e a frequência semanal é ligeiramente menor, sugerindo um compromisso mais leve ou esporádico com as atividades físicas.
* **Percentual de Gordura e Índice de Massa Corporal (BMI)**: Os valores médios indicam um perfil mais próximo do "público geral", com maior variação nos objetivos de saúde e bem-estar.

**Resumo do Padrão do Cluster 1**: O Cluster 1 agrupa indivíduos com uma **intensidade de exercício mais baixa e frequência menor**. Esse grupo pode ser descrito como de praticantes ocasionais ou que buscam atividades físicas mais leves, provavelmente com foco em manutenção da saúde e bem-estar geral.

### **Conclusão**

Os dois clusters formados representam perfis distintos:

1. **Cluster 0** - Praticantes Regulares e Intensos.
2. **Cluster 1** - Praticantes Ocasional e de Baixa Intensidade.

Esses padrões podem ser utilizados para personalizar treinos, promoções e campanhas de incentivo que atendam melhor os interesses e objetivos de cada grupo.