

Segmentação de Usuários de uma Plataforma de Streaming

Data de entrega: 26 de fevereiro de 2026

1 Contexto

Uma plataforma de streaming de filmes e séries deseja compreender melhor o comportamento de seus usuários ativos, com o objetivo de:

- personalizar recomendações;
- melhorar estratégias de retenção;
- identificar perfis com risco de cancelamento;
- estruturar planos de assinatura mais adequados.

Atualmente, a empresa possui apenas dados agregados de uso e **não realiza segmentação formal baseada em métodos estatísticos**.

Você faz parte da equipe de Ciência de Dados responsável por conduzir uma **análise de agrupamentos multivariada**, identificando perfis homogêneos de usuários.

1.1 Base de dados

Cada observação representa um **usuário ativo** da plataforma.

As variáveis disponíveis são:

Variável	Descrição
idade	Idade do usuário (anos)
horas_semana	Horas médias de consumo por semana
dias_ativos	Número médio de dias ativos por semana
avaliacoes	Número médio de avaliações feitas por mês
tempo_assinatura	Tempo de assinatura (meses)
valor_plano	Valor mensal do plano (R\$)

Todas as variáveis são **quantitativas contínuas**. Os dados para esta análise estão disponíveis [aqui](#).

```
Attaching package: 'dplyr'
```

```
The following objects are masked from 'package:stats':
```

```
filter, lag
```

```
The following objects are masked from 'package:base':
```

```
intersect, setdiff, setequal, union
```

1.2 Objetivo

Identificar **perfis distintos de usuários** com base em padrões de uso, engajamento e fidelização, utilizando **métodos de análise de agrupamentos**.

2 Etapas obrigatórias da análise

2.1 Análise exploratória dos dados (EDA)

Pergunta-guia:

> Os dados sugerem a coexistência de múltiplos perfis de usuários?

2.2 Pré-processamento

- Padronize as variáveis;
 - Justifique a necessidade da padronização;
 - Defina a métrica de distância utilizada.
-

2.3 Avaliação da tendência a agrupamento

2.3.1 Estatística de Hopkins

- Calcule a estatística de Hopkins;
- Interprete formalmente o valor obtido;

2.3.2 Método visual (VAT)

- Aplique o método VAT;
 - Interprete a matriz de distâncias reorganizada;
 - Discuta valores plausíveis para o número de clusters.
-

2.3.3 Análise de Componentes Principais (PCA)

- Utilize PCA para **visualização exploratória**;
 - Construa o gráfico das duas primeiras componentes;
-

2.4 Métodos hierárquicos

Aplique os seguintes métodos hierárquicos aglomerativos:

- Método do vizinho mais próximo (*single linkage*)
- Método do vizinho mais distante (*complete linkage*)
- Método da distância média (*average linkage*)
- Método do centroide (*centroid*)
- Método de Ward (*Ward.D2*)

2.4.1 Correlação cofenética

- Calcule a correlação cofenética para cada método;
 - Compare os resultados;
 - Selecione o método mais adequado.
-

2.4.2 Dendrograma e escolha do número de clusters

- Construa o dendrograma do método selecionado;
 - Avalie visualmente possíveis cortes;
 - Teste pelo menos **três valores diferentes de k**;
 - Justifique a escolha final.
-

2.4.3 Validação por silhueta

- Calcule a silhueta média global;
 - Analise a silhueta por cluster;
 - Identifique clusters fracos ou observações de fronteira.
-

2.4.4 Análise descritiva e leitura substantiva (Hierárquico)

- Calcule estatísticas descritivas por cluster;
 - Realize a **leitura substantiva** dos grupos;
 - Atribua rótulos conceituais aos perfis identificados.
-

2.5 K-médias como método confirmatório

- Aplique K-médias com o valor final de k;
 - Utilize múltiplas inicializações (**nstart** elevado);
 - Compare os resultados com a solução hierárquica.
-

2.6 Conclusão

Apresente:

- a segmentação final escolhida;
- a justificativa metodológica completa;
- a interpretação dos perfis encontrados;
- implicações práticas para a plataforma de streaming.