



# PLANO DE TESTES — Consumer Profiling & Behavior Clustering

Versão: 1.0

Responsável: Ewerton Florencio

Tipo: Plano de Testes

Projeto: Segmentação de Consumidores por Comportamento

Escopo: *Data Science/Data Engineering*

# 1. Objetivo

Validar a integridade dos dados, a consistência das transformações e o correto funcionamento do modelo de clusterização, garantindo que o pipeline opere de ponta a ponta sem falhas.

## 2. Escopo dos Testes

- Qualidade dos dados (**Raw** → **Silver**).
- Transformações e agregações (**Silver** → **Gold**).
- Cálculo das métricas **RFM**.
- Execução e estabilidade do modelo de clusterização.
- Coerência dos resultados e distribuição dos clusters.

## 3. Tipos de Teste

### 3.1 Testes de Dados — Raw

ID	Teste	Critério
TD01	Integridade da fonte	CSV pode ser lido com sucesso; Todas as colunas esperadas existem.
TD02	Tipos válidos	Datas, quantidade e valor monetário convertidos corretamente.
TD03	Registros inválidos	Identificar linhas com CustomerID nulo, Quantity $\leq 0$ (exceto devoluções) e UnitPrice $\leq 0$ .

### 3.2 Testes da Camada Silver

ID	Teste	Critério
TS01	Remoção de duplicatas	Duplicidades entre transações devem ser eliminadas.
TS02	Cálculo de TotalPrice	Validar que TotalPrice = Quantity $\times$ UnitPrice.
TS03	Datas válidas	InvoiceDate não pode ser futura; Datas inválidas devem ser descartadas.

### 3.3 Testes da Camada Gold (RFM)

ID	Métrica	Critério
TG01	Recency	Verificar cálculo da data mais recente por cliente; Recency deve ser $\geq 0$ .
TG02	Frequency	Contagem correta de InvoiceNo distintos por cliente.
TG03	Monetary	Soma total de TotalPrice por cliente correta.
TG04	Normalização	Confirmar que métricas normalizadas possuem média $\sim 0$ e desvio $\sim 1$ (StandardScaler).

### 3.4 Testes da Modelagem

ID	Teste	Critério
TM01	Execução do modelo	K-Means deve rodar sem erro para o número definido de clusters.
TM02	Estabilidade de convergência	n_init e max_iter devem resultar em modelo estável.
TM03	Avaliação mínima	<b>Silhouette Score</b> deve ser $> 0$ (clusters válidos).
TM04	Clusters não vazios	Nenhum cluster pode retornar sem clientes.

### 3.5 Testes dos Artefatos de Saída

ID	Teste	Critério
TA01	Tabelas finais geradas	Tabelas RFM e clusters exportadas sem falha.
TA02	Visualizações	Gráficos são gerados sem erro e sem valores ausentes críticos.

## 4. Critérios de Aprovação

- Pipeline executa integralmente (**Raw** → **Silver** → **Gold** → **Clusters**).
- Não existem erros críticos nas transformações.
- Tipo de dados coerente em todas as camadas.
- Cálculo de RFM validado manualmente em amostras.
- Modelo gera clusters consistentes e interpretáveis.

## 5. Itens Fora do Escopo dos Testes

- Performance em larga escala.
- Testes de integração com sistemas externos.
- Testes de orquestração e automação.
- Testes de segurança e permissões.