



DICIONÁRIO DE DADOS — Consumer Profiling & Behavior Clustering

Versão: 1.0

Responsável: Ewerton Florencio

Tipo: Dicionário de Dados

Projeto: Segmentação de Consumidores por Comportamento

Escopo: *Data Science/Data Engineering*

1. Camada Raw (Transações Originais)

Campo	Significado de Negócio
CustomerID	Identificador único do consumidor. Representa um cliente individual no processo de compra.
InvoiceNo	Código único da transação comercial realizada pelo cliente.
InvoiceDate	Data e horário em que a transação foi registrada.
Quantity	Quantidade de itens adquiridos na transação.
UnitPrice	Preço unitário do item no momento da venda.
StockCode	Código do produto comprado. Identifica o item único dentro do catálogo.
Description	Nome ou descrição textual do produto.
Country	País de origem da compra do cliente.

2. Camada Silver (Dados Tratados e Padronizados)

Campo	Significado de Negócio
TotalPrice	Valor total pago pelo cliente na transação ($\text{Quantity} \times \text{UnitPrice}$).
Clean_Description	Descrição do produto tratada (padronizada, sem caracteres inválidos ou ruídos).
CustomerStatus	Indicador de integridade do cliente (válido, incompleto, suspeito, duplicado etc.) — utilizado apenas para análises internas.
TransactionValidity	Status da transação após validações básicas (ex.: válida, devolução, preço inconsistente, quantidade negativa).

3. Camada Gold (Métricas RFM por Cliente)

Campo	Significado de Negócio
Recency	Número de dias desde a última compra do cliente até a data de referência da análise.
Frequency	Quantidade total de compras (transações distintas) realizadas por aquele cliente.
Monetary	Valor financeiro total gasto pelo cliente no período analisado.
Recency_Score	Pontuação categorizada da métrica Recency (ex.: 1 a 5).
Frequency_Score	Pontuação categorizada para a métrica Frequency (ex.: 1 a 5).
Monetary_Score	Pontuação categorizada da métrica Monetary (ex.: 1 a 5).
RFM_Group	Combinação das três pontuações R, F e M (ex.: 544, 321).
RFM_Segment	Segmento comercial atribuído com base no grupo RFM (ex.: "Clientes VIP", "Leais", "Em risco", "Recém-chegados").

4. Camada de Clusters (Saída do Modelo)

Campo	Significado de Negócio
ClusterID	Cluster numérico atribuído pelo modelo de Machine Learning (K-Means ou DBSCAN).
Cluster_Label	Nome interpretável do cluster, criado após análise (ex.: “Alto Valor”, “Oportunidade”, “Preço Sensível”).
Cluster_Recency_Avg	Recency médio das pessoas do cluster — usado para caracterização do grupo.
Cluster_Frequency_Avg	Frequency médio do cluster.
Cluster_Monetary_Avg	Monetary médio do cluster.
Key_Insights	Resumo conceitual das principais características do cluster.

5. Tabelas Auxiliares / Dimensões Conceituais

Tabela	Significado de Negócio
Dim_Product	Agrupamento conceitual dos produtos mais relevantes por categoria (se usado no <i>storytelling</i> do projeto).
Dim_Country	Mapa conceitual das regiões onde há concentração de clientes (opcional, usado para análise).