

Dicionário de Dados - Raw Layer

Diego Carlito Rodrigues de Souza - 221007690

Professor: Thiago Luiz de Souza Gomes

18 de outubro de 2025

1 Dicionário de Dados

1. olist_customers_dataset.csv

Armazena informações sobre os clientes.

Tabela 1: Dicionário de olist_customers_dataset.csv

| Atributo | Descrição | Tipo de Dado |
|--------------------------|---|--------------|
| customer_id | ID único do cliente por pedido. | string |
| customer_unique_id | ID que identificaunicamente um cliente. | string |
| customer_zip_code_prefix | Os 5 primeiros dígitos do CEP do cliente. | integer |
| customer_city | A cidade do cliente. | string |
| customer_state | O estado do cliente (sigla de 2 letras). | string |

2. olist_orders_dataset.csv

Contém os dados centrais de cada pedido realizado.

Tabela 2: Dicionário de `olist_orders_dataset.csv`

| Atributo | Descrição | Tipo de Dado |
|--|--|------------------------|
| <code>order_id</code> | ID único de cada pedido. | <code>string</code> |
| <code>customer_id</code> | Chave estrangeira que liga ao <code>olist_customers_dataset</code> . | <code>string</code> |
| <code>order_status</code> | Status atual do pedido. | <code>string</code> |
| <code>order_purchase_timestamp</code> | Data e hora em que a compra foi realizada. | <code>timestamp</code> |
| <code>order_approved_at</code> | Data e hora da aprovação do pagamento. | <code>timestamp</code> |
| <code>order_delivered_carrier_date</code> | Data e hora em que o pedido foi postado na transportadora. | <code>timestamp</code> |
| <code>order_delivered_customer_date</code> | Data da entrega real do pedido ao cliente. | <code>timestamp</code> |
| <code>order_estimated_delivery_date</code> | Data estimada de entrega informada ao cliente. | <code>timestamp</code> |

3. `olist_order_items_dataset.csv`

Tabela associativa que detalha os produtos contidos em cada pedido.

Tabela 3: Dicionário de `olist_order_items_dataset.csv`

| Atributo | Descrição | Tipo de Dado |
|----------------------------------|---|------------------------|
| <code>order_id</code> | Chave estrangeira que liga ao <code>olist_orders_dataset</code> . | <code>string</code> |
| <code>order_item_id</code> | ID sequencial que identifica os itens dentro de um mesmo pedido. | <code>integer</code> |
| <code>product_id</code> | Chave estrangeira que liga ao <code>olist_products_dataset</code> . | <code>string</code> |
| <code>seller_id</code> | Chave estrangeira que liga ao <code>olist_sellers_dataset</code> . | <code>string</code> |
| <code>shipping_limit_date</code> | Data limite para o vendedor enviar o produto. | <code>timestamp</code> |
| <code>price</code> | O preço do produto (valor unitário). | <code>double</code> |
| <code>freight_value</code> | O valor do frete do item. | <code>double</code> |

4. olist_order_payments_dataset.csv

Registra as informações de pagamento de cada pedido.

Tabela 4: Dicionário de olist_order_payments_dataset.csv

| Atributo | Descrição | Tipo de Dado |
|----------------------|--|--------------|
| order_id | Chave estrangeira que liga ao olist_orders_dataset. | string |
| payment_sequential | Sequencial para pagamentos múltiplos em um mesmo pedido. | integer |
| payment_type | Método de pagamento utilizado. | string |
| payment_installments | Número de parcelas no cartão de crédito. | integer |
| payment_value | O valor total do pagamento. | double |

5. olist_order_reviews_dataset.csv

Armazena as avaliações (reviews) feitas pelos clientes para cada pedido.

Tabela 5: Dicionário de olist_order_reviews_dataset.csv

| Atributo | Descrição | Tipo de Dado |
|-------------------------|--|--------------|
| review_id | ID único da avaliação. | string |
| order_id | Chave estrangeira que liga ao olist_orders_dataset. | string |
| review_score | Nota da avaliação (de 1 a 5). | string |
| review_comment_title | Título do comentário da avaliação (opcional). | string |
| review_comment_message | Corpo do comentário da avaliação (opcional). | string |
| review_creation_date | Data em que a pesquisa de satisfação foi enviada ao cliente. | timestamp |
| review_answer_timestamp | Data em que o cliente respondeu à pesquisa. | timestamp |

6. olist_products_dataset.csv

Contém os dados do catálogo de produtos vendidos.

Tabela 6: Dicionário de `olist_products_dataset.csv`

| Atributo | Descrição | Tipo de Dado |
|---|---|----------------------|
| <code>product_id</code> | ID único do produto. | <code>string</code> |
| <code>product_category_name</code> | Nome da categoria do produto, em português. | <code>string</code> |
| <code>product_name_length</code> | Número de caracteres no nome do produto. | <code>integer</code> |
| <code>product_description_length</code> | Número de caracteres na descrição do produto. | <code>integer</code> |
| <code>product_photos_qty</code> | Quantidade de fotos publicadas do produto. | <code>integer</code> |
| <code>product_weight_g</code> | Peso do produto em gramas. | <code>integer</code> |
| <code>product_length_cm</code> | Comprimento do produto em centímetros. | <code>integer</code> |
| <code>product_height_cm</code> | Altura do produto em centímetros. | <code>integer</code> |
| <code>product_width_cm</code> | Largura do produto em centímetros. | <code>integer</code> |

7. `olist_sellers_dataset.csv`

Armazena informações sobre os vendedores que utilizam a plataforma Olist.

Tabela 7: Dicionário de `olist_sellers_dataset.csv`

| Atributo | Descrição | Tipo de Dado |
|-------------------------------------|--|----------------------|
| <code>seller_id</code> | ID único do vendedor. | <code>string</code> |
| <code>seller_zip_code_prefix</code> | Os 5 primeiros dígitos do CEP do vendedor. | <code>integer</code> |
| <code>seller_city</code> | A cidade do vendedor. | <code>string</code> |
| <code>seller_state</code> | O estado do vendedor (sigla de 2 letras). | <code>string</code> |

8. `product_category_name_translation.csv`

Tabela "de-para" para traduzir os nomes das categorias de produtos para o inglês.

Tabela 8: Dicionário de `product_category_name_translation.csv`

| Atributo | Descrição | Tipo de Dado |
|--|---------------------------------|---------------------|
| <code>product_category_name</code> | Nome da categoria em português. | <code>string</code> |
| <code>product_category_name_english</code> | Nome da categoria em inglês. | <code>string</code> |

9. `olist_geolocation_dataset.csv`

Contém informações de geolocalização (latitude e longitude) para cada CEP.

Tabela 9: Dicionário de `olist_geolocation_dataset.csv`

| Atributo | Descrição | Tipo de Dado |
|--|--------------------------------|----------------------|
| <code>geolocation_zip_code_prefix</code> | Os 5 primeiros dígitos do CEP. | <code>integer</code> |
| <code>geolocation_lat</code> | Latitude. | <code>double</code> |
| <code>geolocation_lng</code> | Longitude. | <code>double</code> |
| <code>geolocation_city</code> | Nome da cidade. | <code>string</code> |
| <code>geolocation_state</code> | Estado (sigla de 2 letras). | <code>string</code> |