

# Diccionario de datos

## 1. customers\_dataset

Este conjunto de datos tiene información sobre el cliente y su ubicación.

Cuenta con 5 dimensiones y 99441 registros

- customer\_id: Clave para el conjunto de datos de pedidos. Cada pedido tiene un ID de cliente único.
- customer\_unique\_id: Identificador único de un cliente.
- customer\_zip\_code\_prefix: Primeros cinco dígitos del código postal del cliente
- customer\_city: Nombre de la ciudad donde reside el cliente
- customer\_state: Estado del cliente

## 2. geolocation\_dataset

Este conjunto de datos contiene información sobre los códigos postales de Brasil y sus coordenadas de latitud y longitud.

Cuenta con 5 dimensiones y 1000163 registros

- geolocation\_zip\_code\_prefix: Primeros 5 dígitos del código postal del cliente
- geolocation\_lat: latitud
- geolocation\_lng: longitud
- geolocation\_city: Ciudad donde reside el cliente
- geolocation\_state: Estado

## 3. order\_items\_dataset

Este conjunto de datos incluye información sobre los artículos comprados en cada pedido.

Cuenta con 7 dimensiones y 112650 registros.

- order\_id: identificador único de pedido
- order\_item\_id: número secuencial que identifica el número de artículos incluidos en el mismo pedido.
- product\_id: identificador único del producto
- seller\_id: identificador único del vendedor
- shipping\_limit\_date: Muestra la fecha límite de envío del vendedor para entregar el pedido al socio logístico.
- price: precio del artículo
- freight\_value: valor de flete del artículo (si un pedido tiene más de un artículo, el valor de flete se divide entre los artículos)

## 4. order\_payments\_dataset

Este conjunto de datos incluye datos sobre las opciones de pago de los pedidos.

Cuenta con 5 dimensiones y 103886 registros

- `order_id`: identificador único de un pedido.
- `payment_sequential`: un cliente puede pagar un pedido con más de un método de pago. Si lo hace, se creará una secuencia para acomodar
- `payment_type`: forma de pago elegida por el cliente.
- `payment_installments`: número de cuotas elegido por el cliente.
- `payment_value`: valor de la transacción.

## **5. order\_reviews\_dataset**

Este conjunto de datos incluye datos sobre las reseñas realizadas por los clientes.

Cuenta con 7 dimensiones y 99224 registros

- `review_id`: identificador de reseña único
- `order_id`: identificador único de pedido
- `review_score`: Nota del 1 al 5 dada por el cliente en una encuesta de satisfacción.
- `review_comment_title`: Título del comentario de la reseña dejada por el cliente, en portugués.
- `review_comment_message`: Mensaje de comentario de la reseña dejada por el cliente, en portugués.
- `review_creation_date`: Muestra la fecha en que se envió la encuesta de satisfacción al cliente.
- `review_answer_timestamp`: Muestra la marca de tiempo de respuesta de la encuesta de satisfacción.

## **6. orders\_dataset**

Este es el conjunto de datos central de pedidos. Cuenta con 8 dimensiones y 99441 registros

- `order_id`: identificador único de la orden.
- `customer_id`: clave para el conjunto de datos del cliente. Cada pedido tiene un ID de cliente único.
- `order_status`: Referencia al estado del pedido (entregado, enviado, etc).
- `order_purchase_timestamp`: Muestra la marca de tiempo de la compra.
- `order_approved_at`: Muestra la marca de tiempo de aprobación del pago.
- `order_delivered_carrier_date`: Muestra la marca de tiempo de publicación del pedido cuando se entregó al socio logístico
- `order_delivered_customer_date`: Muestra la fecha real de entrega del pedido al cliente.
- `order_estimated_delivery_date`: Muestra la fecha estimada de entrega que fue informada al cliente en el momento de la compra.

## **7. products\_dataset**

Este conjunto de datos incluye datos sobre los productos vendidos por Olist.

Cuenta con 9 dimensiones y 32951 registros

- product\_id: identificador único de producto
- product\_category\_name: categoría raíz de producto, en portugués.
- product\_name\_lenght: número de caracteres extraídos del nombre del producto.
- product\_description\_lenght: número de caracteres extraídos de la descripción del producto.
- product\_photos\_qty: número de fotos publicadas del producto
- product\_weight\_g: peso del producto medido en gramos.
- product\_length\_cm: longitud del producto medida en centímetros.
- product\_height\_cm: altura del producto medida en centímetros.
- product\_width\_cm: ancho del producto medido en centímetros.

## **8. sellers\_dataset**

Este conjunto de datos incluye datos sobre los vendedores que cumplieron con los pedidos realizados en Olist.

Cuenta con 4 dimensiones y 3095 registros

- seller\_id: identificador único del vendedor
- seller\_zip\_code\_prefix: primeros 5 dígitos del código postal del vendedor
- seller\_city: nombre de la ciudad del vendedor
- seller\_state: estado del vendedor

## **9. category\_translat\_dataset**

Este conjunto de datos traduce la tabla 'product\_category\_name' al inglés.

Cuenta con 2 dimensiones y 71 registros

- product\_category\_name: nombre de categoría en portugués
- product\_category\_name\_english: nombre de la categoría en inglés

## **10. Closed\_deals\_dataset:**

Después de que un cliente potencial calificado complete un formulario en una página de destino, un representante de desarrollo de ventas se comunica con él. En este paso, se verifica cierta información y se recopila más información sobre el cliente potencial.

Este conjunto de datos contiene la información de los clientes potenciales de Olist

Cuenta con 14 dimensiones y 872 registros

- mql\_id: Identificación de prospecto calificado de marketing
- seller\_id: identificación del vendedor
- sdr\_id: Identificación del representante de desarrollo de ventas
- sr\_id: Representante de ventas
- won\_date: Fecha en que se cerró el trato.

- **business\_segment:** Líder del segmento de negocio. Informado en contacto.
- **lead\_type:** Tipo de cliente potencial. Informado en contacto.
- **lead\_behaviour\_profile:** Perfil de comportamiento de los clientes potenciales. SDR lo identifica al contacto.
- **has\_company:** ¿El prospecto tiene una empresa (documentación formal)?
- **has\_gtin:** ¿El cliente potencial tiene un Número Global de Artículo Comercial (código de barras) para sus productos?
- **average\_stock:** Stock del cliente potencial medio declarado. Informado en contacto.
- **business\_type:** Tipo de negocio (revendedor/fabricante, etc.)
- **declared\_product\_catalog\_size:** Tamaño de catálogo declarado del cliente potencial. Informado en contacto.
- **declared\_monthly\_revenue:** El cliente potencial declaró ingresos mensuales estimados. Informado en contacto.

### **11. marketing\_qualified\_leads\_dataset**

Después de que un cliente potencial complete un formulario en una página de destino, se realiza un filtro para seleccionar los que están calificados para vender sus productos en Olist. Son los clientes potenciales Calificados de Marketing (MQLs).

Este conjunto de datos contiene los datos de esos clientes potenciales calificados

Cuenta con 4 dimensiones y 8000 registros

- **mql\_id:** Identificación de prospecto calificado de marketing
- **first\_contact\_date:** Fecha de la solicitud de primer contacto.
- **landing\_page\_id:** ID de la página de destino donde se adquirió el cliente potencial
- **origin:** Tipo de medio donde se adquirió el cliente potencial