

DICCIONARIO DE DATOS

1. [Tabla: product_category_name_translation](#)
2. [Tabla: olist_products_dataset](#)
3. [Tabla: olist_order_reviews_dataset](#)
4. [Tabla: olist_orders_dataset](#)
5. [Tabla: olist_order_payments_dataset](#)
6. [Tabla: olist_order_items_dataset](#)
7. [Tabla: olist_geolocation_dataset](#)
8. [Tabla: olist_customers_dataset](#)

1. Tabla: product_category_name_translation

Descripción de Columnas:

Columna 1:

Nombre: product_category_name:

Descripción: Nombre de la categoría del producto en portugués

Tipo de dato: string (texto)

Columna 2

Nombre: product_category_name_english:

Descripción: Nombre de la categoría del producto en inglés

Tipo de dato: string / VARCHAR

Claves

PK: product_category_name

FK: No hay FK evidente ya que esta tabla parece funcionar como una tabla de referencia para traducciones.

Primeras Filas (Ejemplo):

product_category_name = "beleza_saude"

product_category_name_english = "health_beauty"

2. Tabla: olist_products_dataset

Descripción de Columnas:

Columna 1:

Nombre: product_id

Descripción: Identificador único del producto.

Tipo de dato: string / VARCHAR

Columna 2:

Nombre: product_category_name

Descripción: Nombre de la categoría del producto.

Tipo de dato: string / VARCHAR

Columna 3:

Nombre: product_name_lenght

Descripción: Longitud del nombre del producto.

Tipo de dato: float64 / FLOAT

Columna 4:

Nombre: product_description_lenght

Descripción: Longitud de la descripción del producto.

Tipo de dato: float64 / FLOAT

Columna 5:

Nombre: product_photos_qty

Descripción: Cantidad de fotos del producto.

Tipo de dato: float64 / FLOAT

Columna 6:

Nombre: product_weight_g

Descripción: Peso del producto en gramos.

Tipo de dato: float64 / FLOAT

Columna 7:

Nombre: product_length_cm

Descripción: Longitud del producto en centímetros.

Tipo de dato: float64 / FLOAT

Columna 8:

Nombre: product_height_cm

Descripción: Altura del producto en centímetros.

Tipo de dato: float64 / FLOAT

Columna 9:

Nombre: product_width_cm

Descripción: Ancho del producto en centímetros.

Tipo de dato: float64 / FLOAT

Claves

PK: product_id

FK: product_category_name (relacionada con
product_category_name_translation.product_category_name)

Primeras Filas (Ejemplo):

product_id = "1e9e8ef04dbcff4541ed26657ea517e5"

product_category_name = "perfumaria"

product_name_lenght = 40.0

product_description_lenght = 287.0

product_photos_qty = 1.0

product_weight_g = 225.0

product_length_cm = 16.0

product_height_cm = 10.0

product_width_cm = 14.0

3. Tabla: olist_order_reviews_dataset

Descripción de Columnas:

Columna 1:

Nombre: review_id

Descripción: Identificador único de la reseña.

Tipo de dato: string / VARCHAR

Columna 2:

Nombre: order_id

Descripción: Identificador único del pedido.

Tipo de dato: string / VARCHAR

Columna 3:

Nombre: review_score

Descripción: Puntuación de la reseña.

Tipo de dato: int64 / INTEGER

Columna 4:

Nombre: review_comment_title

Descripción: Título del comentario de la reseña.

Tipo de dato: string / VARCHAR

Columna 5:

Nombre: review_comment_message

Descripción: Mensaje del comentario de la reseña.

Tipo de dato: string / VARCHAR

Columna 6:

Nombre: review_creation_date

Descripción: Fecha de creación de la reseña.

Tipo de dato: string / VARCHAR

Columna 7:

Nombre: review_answer_timestamp

Descripción: Fecha y hora de respuesta a la reseña.

Tipo de dato: string / VARCHAR

Claves

PK: review_id

FK: order_id (relacionada con olist_orders_dataset.order_id)

Primeras Filas (Ejemplo):

review_id = "7bc2406110b926393aa56f80a40eba40"

order_id = "73fc7af87114b39712e6da79b0a377eb"

review_score = 4

review_comment_title = NaN

review_comment_message = NaN
review_creation_date = "2018-01-18 00:00:00"
review_answer_timestamp = "2018-01-18 21:46:59"

4. Tabla: olist_orders_dataset

Descripción de Columnas:

Columna 1:

Nombre: order_id

Descripción: Identificador único del pedido.

Tipo de dato: string / VARCHAR

Columna 2:

Nombre: customer_id

Descripción: Identificador único del cliente asociado con el pedido.

Tipo de dato: string / VARCHAR

Columna 3:

Nombre: order_status

Descripción: Estado del pedido.

Tipo de dato: string / VARCHAR

Columna 4:

Nombre: order_purchase_timestamp

Descripción: Marca de tiempo de la compra del pedido.

Tipo de dato: DATETIME

Columna 5:

Nombre: order_approved_at

Descripción: Marca de tiempo de la aprobación del pedido.

Tipo de dato: DATETIME

Columna 6:

Nombre: order_delivered_carrier_date

Descripción: Fecha en que el pedido fue entregado al transportista.

Tipo de dato: DATETIME

Columna 7:

Nombre: order_delivered_customer_date

Descripción: Fecha en que el pedido fue entregado al cliente.

Tipo de dato: DATETIME

Columna 8:

Nombre: order_estimated_delivery_date

Descripción: Fecha estimada de entrega del pedido.

Tipo de dato: DATETIME

Claves

PK: order_id

FK: customer_id (relacionada con olist_customers_dataset.customer_id)

Primeras Filas (Ejemplo):

order_id = "e481f51cbdc54678b7cc49136f2d6af7"

customer_id = "9ef432eb6251297304e76186b10a928d"

order_status = "delivered"

order_purchase_timestamp = "2017-10-02 10:56:33"

order_approved_at = "2017-10-02 11:07:15"

order_delivered_carrier_date = "2017-10-04 19:55:00"

order_delivered_customer_date = "2017-10-10 21:25:13"

order_estimated_delivery_date = "2017-10-18 00:00:00"

5.Tabla: olist_order_payments_dataset

Descripción de Columnas:

Columna 1:

Nombre: order_id

Descripción: Identificador único del pedido.

Tipo de dato: string / VARCHAR

Columna 2:

Nombre: payment_sequential

Descripción: Número secuencial del pago.

Tipo de dato: int64 / INTEGER

Columna 3:

Nombre: payment_type

Descripción: Tipo de pago.

Tipo de dato: string / VARCHAR

Columna 4:

Nombre: payment_installments

Descripción: Número de cuotas de pago.

Tipo de dato: int64 / INTEGER

Columna 5:

Nombre: payment_value

Descripción: Valor del pago.

Tipo de dato: float64 / FLOAT

Claves

PK: No hay una PK evidente; order_id y payment_sequential podrían formar una clave compuesta.

FK: order_id (relacionada con olist_orders_dataset.order_id)

Primeras Filas (Ejemplo):

order_id: "b81ef226f3fe1789b1e8b2acac839d17"

payment_sequential: 1

payment_type: "credit_card"

payment_installments: 8

payment_value: 99.33

6. Tabla: olist_order_items_dataset

Por supuesto, aquí tienes la adaptación del texto siguiendo el formato indicado:

Descripción de Columnas:

Columna 1:

Nombre: order_id

Descripción: Identificador único del pedido.

Tipo de dato: string / VARCHAR

Columna 2:

Nombre: order_item_id

Descripción: Identificador secuencial del artículo en el pedido.

Tipo de dato: int64 / INTEGER

Columna 3:

Nombre: product_id

Descripción: Identificador único del producto.

Tipo de dato: string / VARCHAR

Columna 4:

Nombre: seller_id

Descripción: Identificador único del vendedor.

Tipo de dato: string / VARCHAR

Columna 5:

Nombre: shipping_limit_date

Descripción: Fecha límite de envío.

Tipo de dato: DATETIME

Columna 6:

Nombre: price

Descripción: Precio del producto.

Tipo de dato: float64 / FLOAT

Columna 7:

Nombre: freight_value

Descripción: Valor del flete.

Tipo de dato: float64 / FLOAT

Claves

PK: No hay una PK evidente; order_id y order_item_id podrían formar una clave compuesta.

FKs:

order_id (relacionada con olist_orders_dataset.order_id)

product_id (relacionada con olist_products_dataset.product_id)

seller_id (relacionada con olist_sellers_dataset.seller_id)

Primeras Filas (Ejemplo):

order_id: "00010242fe8c5a6d1ba2dd792cb16214"

order_item_id: 1

product_id: "4244733e06e7ecb4970a6e2683c13e61"

seller_id: "48436dade18ac8b2bce089ec2a041202"

shipping_limit_date: "2017-09-19 09:45:35"

price: 58.9

freight_value: 13.29

7. Tabla: olist_geolocation_dataset

Descripción de Columnas:

Columna 1:

Nombre: geolocation_zip_code_prefix

Descripción: Prefijo del código postal asociado con la geolocalización. Este campo se usa para identificar la región geográfica.

Tipo de dato: int (entero)

Columna 2:

Nombre: geolocation_lat

Descripción: Latitud geográfica donde se encuentra el código postal. Es una medida en grados que indica qué tan al norte o al sur se encuentra un punto del ecuador.

Tipo de dato: float (decimal)

Columna 3:

Nombre: geolocation_lng

Descripción: Longitud geográfica donde se encuentra el código postal. Es una medida en grados que indica qué tan al este o al oeste se encuentra un punto del meridiano principal.

Tipo de dato: float (decimal)

Columna 4:

Nombre: geolocation_city

Descripción: Ciudad asociada con la geolocalización. Esto ayuda a identificar la ubicación urbana del código postal.

Tipo de dato: string / VARCHAR

Columna 5:

Nombre: geolocation_state

Descripción: Estado asociado con la geolocalización. Proporciona información adicional sobre la ubicación regional dentro del país.

Tipo de dato: string / VARCHAR

Claves

PK: No hay una PK evidente debido a la naturaleza de los datos geográficos; se podría usar geolocation_zip_code_prefix, pero es probable que no sea única.

FK: No hay FK evidente ya que esta tabla proporciona información de referencia geográfica que se puede usar para enriquecer otras tablas.

Primeras Filas (Ejemplo):

geolocation_zip_code_prefix = 1037

geolocation_lat = -23.54562128115260

geolocation_lng = -46.63929204800160

geolocation_city = "sao paulo"

geolocation_state = "SP"

8. Tabla: olist_customers_dataset

Descripción de Columnas:

Columna 1:

Nombre: customer_id

Descripción: Identificador único del cliente.

Tipo de dato: string / VARCHAR

Columna 2:

Nombre: customer_unique_id

Descripción: Identificador único y persistente para cada cliente.

Tipo de dato: string / VARCHAR

Columna 3:

Nombre: customer_zip_code_prefix

Descripción: Prefijo del código postal del cliente.

Tipo de dato: int64 / INTEGER

Columna 4:

Nombre: customer_city

Descripción: Ciudad del cliente.

Tipo de dato: string / VARCHAR

Columna 5:

Nombre: customer_state

Descripción: Estado del cliente.

Tipo de dato: string / VARCHAR

Claves

PK: customer_id

FK: No hay una FK evidente.

Primeras Filas (Ejemplo):

customer_id = "06b8999e2fba1a1fbc88172c00ba8bc7"

customer_unique_id = "861eff4711a542e4b93843c6dd7febb0"

customer_zip_code_prefix = 14409

customer_city = "franca"

customer_state = "SP"