# Dominio del problema: Empresa de E-commerce

Anexo a este documento, le compartimos un conjunto de datos públicos[[1]](#footnote-1) de comercio electrónico brasileño de pedidos realizados en [Olist Store](https://olist.com/).

Olist conecta pequeñas empresas de todo Brasil con canales simples y con un único contrato. Estos comercios pueden vender sus productos a través de la tienda Olist y enviarlos directamente a los clientes utilizando sus socios logísticos.

Después de que un cliente compra el producto en la Olist Store, se notifica al vendedor para que complete ese pedido. Una vez que el cliente recibe el producto, o vence la fecha estimada de entrega, el cliente recibe una encuesta de satisfacción por correo electrónico donde puede dar una nota de la experiencia de compra y anotar algunos comentarios.

El conjunto de datos tiene información de 100.000 pedidos realizados entre los años 2016 y 2018, en múltiples mercados de Brasil, y permite verlos desde múltiples dimensiones: desde el estado del pedido, el precio, el pago y el rendimiento del flete hasta la ubicación del cliente, los atributos del producto y finalmente las reseñas escritas por los clientes.

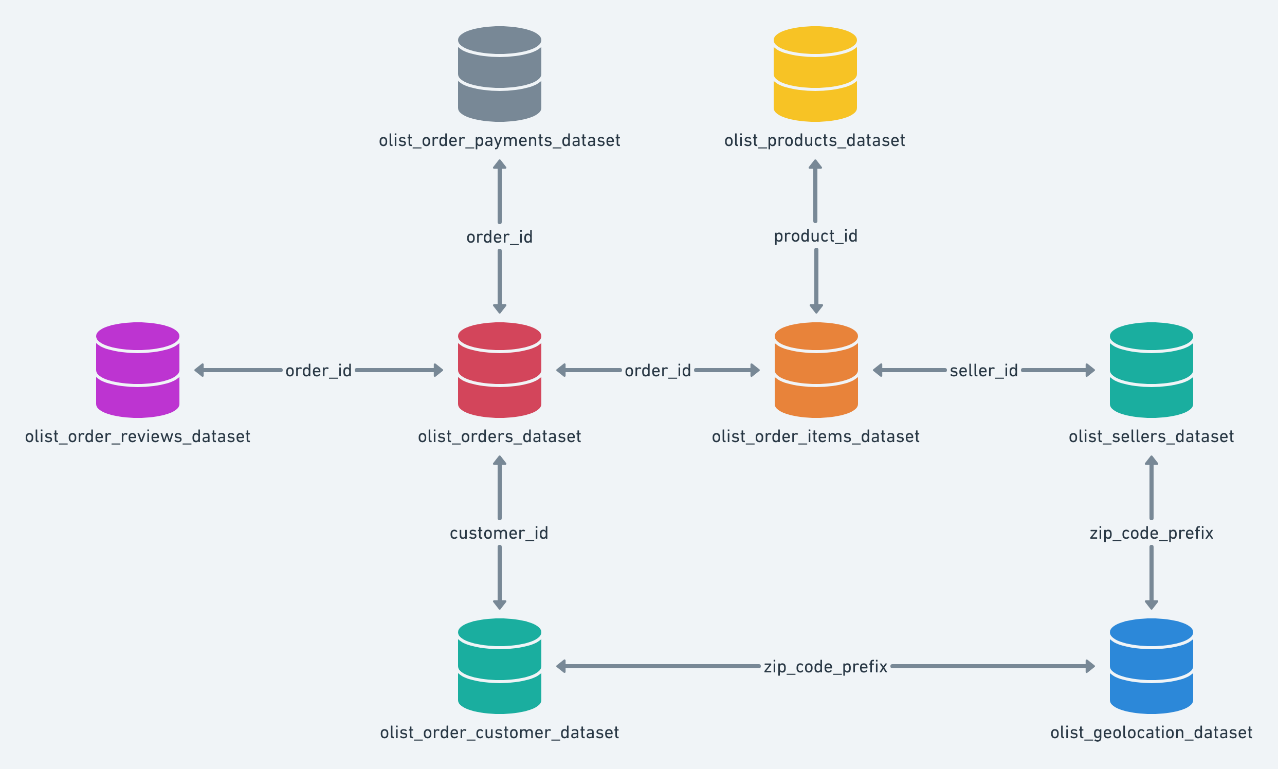
Se trata de datos comerciales reales, que han anonimizados. Las referencias a las empresas y socios en el texto de la reseña han sido sustituidas por los nombres de las grandes casas de Juego de Tronos.

### Atención:

1. Un pedido puede tener varios artículos.
2. Cada artículo puede ser gestionado por un vendedor distinto.
3. Todo el texto que identificaba tiendas y socios fue reemplazado por los nombres de las grandes casas de Juego de Tronos.

## Esquema de datos

Los datos se dividen en múltiples conjuntos de datos para una mejor comprensión y organización. El siguiente gráfico muestra el esquema de relaciones:



Nota: El diccionario de datos correspondiente a este conjunto de datos se encuentra al final de este documento, en la sección Anexo.

## Desafío

Eres un/a consultor/a de inteligencia de negocios y fuiste contratado/a por Olist Store para ayudarlos a comprender mejor su negocio y aumentar la satisfacción de sus clientes. Para ello, debes responder las siguientes preguntas:

* ¿En qué estados residen nuestros clientes? ¿Cuál es el valor monetario de cada estado para Olist?
* ¿En qué estados se encuentran nuestros vendedores? ¿Cuáles son los principales comercios?
* ¿Cuáles son los productos (o categorías de productos) más vendidos?
* ¿Existe alguna relación entre la categoría de producto adquirida y la satisfacción del cliente? Si existe, ¿su impacto es mayor o menor que el tiempo de entrega?
* Trabajando en el problema, ¿encontraste alguna otra relación interesante entre los datos?
* **Opcional:** Si la empresa quisiera construir 3 centros logísticos a fin de reducir al máximo posible los *tiempos de entrega*, ¿en qué lugares debe hacerlo?

Para responder a estas preguntas, debes:

1. Partiendo de la base de datos provista, generar un [*data mart*](https://aws.amazon.com/es/what-is/data-mart/#:~:text=Un%20data%20mart%20es%20un,sistema%20de%20almacenamiento%20m%C3%A1s%20grande) con un [esquema en estrella](https://es.wikipedia.org/wiki/Esquema_en_estrella) con los valores que consideres relevantes para responder a las preguntas planteadas arriba. Para completar este punto, te recomendamos que utilices KNIME, pero eres libre de utilizar la herramienta con la que te sientas más cómodo/a.
2. Generar uno o más *dashboards* en Tableau o en PowerBI en el que se evidencien las respuestas a las preguntas planteadas. La fuente de datos utilizada en este punto debe ser el *data mart* generado en el punto anterior.
3. En una videollamada que marcaremos cuando nos envíes los archivos generados, preséntanos tu solución en *no más de 20 minutos*.

### **Atención**

Los criterios con los que evaluaremos tu solución son, en orden decreciente de importancia:

1. El proceso lógico que seguiste para resolver el problema.
2. La claridad con la que presentes tu solución, tanto en la presentación oral como en la calidad de los *dashboards*.
3. El orden, la pulcritud y la documentación de los archivos generados.
4. El resultado obtenido.

## Anexo: Diccionario de Datos

La siguiente tabla muestra las columnas contenidas en cada tabla de la base de datos, así como la descripción correspondiente a cada una de ellas.

|  |  |
| --- | --- |
| **olist\_customers\_dataset.csv** | |
| customer\_id | clave para el conjunto de datos de pedidos. Cada pedido tiene un customer\_id 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 del cliente |
| customer\_state | estado del cliente |
| **olist\_geolocation\_dataset.csv** | |
| geolocation\_zip\_code\_prefix | primeros 5 dígitos del código postal |
| geolocation\_lat | latitud |
| geolocation\_lng | longitud |
| geolocation\_city | ciudad |
| geolocation\_state | estado |
| **olist\_order\_items\_dataset.csv** | |
| order\_id | identificador único del 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 gestionar el pedido al socio logístico. |
| price | precio del articulo |
| freight\_value | artículo valor de flete artículo (si un pedido tiene más de un artículo, el valor del flete se divide entre los artículos) |
| **olist\_order\_payments\_dataset.csv** | |
| order\_id | identificador único del 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 todos los pagos. |
| payment\_type | Forma de pago elegida por el cliente. |
| payment\_installments | Número de cuotas elegidas por el cliente. |
| payment\_value | valor de la transacción. |
| **olist\_order\_reviews\_dataset.csv** | |
| review\_id | identificador de revisión único |
| order\_id | identificador único del pedido |
| review\_score | Nota que va del 1 al 5 otorgada 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 comentario de la reseña dejada por el cliente, en portugués. |
| review\_creation\_date | Muestra la fecha en la que se envió la encuesta de satisfacción al cliente. |
| review\_answer\_timestamp | Muestra la marca de tiempo de respuesta de la encuesta de satisfacción. |
| **olist\_orders\_dataset.csv** | |
| order\_id | identificador único del pedido |
| customer\_id | clave para el conjunto de datos de pedidos. Cada pedido tiene un customer\_id 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 fue entregado al socio logístico. |
| order\_delivered\_customer\_date | Muestra la fecha real de entrega del pedido al cliente |
| order\_estimated\_delivery\_date | Muestra la fecha de entrega estimada que fue informada al cliente en el momento de la compra. |
| **olist\_products\_dataset.csv** | |
| product\_id | identificador único de producto |
| product\_category\_name | categoría raíz del 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. |
| **olist\_sellers\_dataset.csv** | |
| 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 |
| **product\_category\_name\_translation.csv** | |
| ï»¿product\_category\_name | nombre de la categoría en portugués |
| product\_category\_name\_english | nombre de la categoría en inglés |

1. Datos y texto extraídos de: https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce/?select=product\_category\_name\_translation.csv [↑](#footnote-ref-1)