

Origen de los datos

El origen de los datos proviene de **American Things**, un **poli rubro dedicado a la venta de productos en general**, que se ha visto en la necesidad de organizar la información alojada en sus bases de datos respecto a las ventas realizadas durante un determinado período de tiempo.

El Dataset presenta anomalías en los datos cargados, las cuales podrían haber sido causadas por:

- **Errores externos**, como la manipulación de la base por parte de personas ajenas a la empresa.
- **Falta de previsión** al ingresar información sin validaciones adecuadas.
- **Errores internos**, producto de una manipulación descuidada por parte de personal perteneciente a la organización.

La base contiene **datos incompletos y/o mal dispuestos**, lo que impide que American Things realice **análisis estratégicos eficaces**. Como consecuencia, no pueden tomar decisiones efectivas sobre:

- Las diferentes **categorías de productos** que ofrecen.
- La **identificación de sectores críticos** a mejorar.
- La **optimización de sus estrategias de ventas**, tanto a nivel general como específico.

Estado de los datos

Los datos no están en estado RAW, ya que fueron editados manualmente para introducir inconsistencias, valores faltantes y duplicados. Esto permitirá realizar un proceso completo de análisis, validación y limpieza en el desarrollo del proyecto.

Descripción de las columnas

Columna	Tipo de dato	Descripción	Observaciones
invoice	Texto (str)	Código único de la factura de compra.	Formato:XXX-XX-XXXX
branch	Texto (str)	Sucursal donde realizó la compra.	Ej:"A","B","C"
city	Texto (str)	Ciudad en la que se encuentra la sucursal.	Ej: "Nueva York", "Los Angeles", etc.
type	Texto (str)	Tipo de cliente: si es miembro o no.	"Member" o "Normal"
gender	Texto (str)	Género del cliente.	"Male" o "Female"
category	Texto (str)	Categoría del producto o servicio adquirido.	Ej: "Health And Beauty", "Electronics", etc.
cogs	Numérico (float)	Costo de los productos vendidos (Cost of Goods Sold).	Valor Numérico Decimal
sales	Numérico (float)	Monto total de venta.	Valor Numérico Decimal
date	Fecha (str/date)	Fecha en que realizó la transacción.	Formato:MM/DD/AAAA
time	Hora (str/time)	Hora exacta en que se registró la compra.	Formato:HH:MM
payment	Texto (str)	Método de pago utilizado.	Ej:"Ewallet","Cash","Credit card"
review	Numérico (float)	Puntuación del cliente sobre su experiencia (escala del 1 al 10).	Valor Decimal Entre 0 y 10