

## 1. Dataset Envíos de productos de E-commerce

Este conjunto de datos fue elegido porque resulta práctico y eficaz a los fines de los contenidos del curso. Posee 10999 datos y 12 variables, las cuales son:

ID: Número de ID de los Clientes.

Bloque de almacén: la empresa tiene un gran almacén que se divide en bloques como A, B, C, D, E.

Modo de envío: la empresa envía los productos de varias maneras, como barco, vuelo y carretera.

Llamadas de atención al cliente: Número de llamadas realizadas a partir de consulta por consulta del envío.

Calificación del cliente: La empresa ha calificado a cada cliente. 1 es el más bajo (peor), 5 es el más alto (mejor).

Costo del producto: Costo del Producto en dólares estadounidenses.

Compras anteriores: El Número de Compras Previas.

Importancia del producto: la empresa ha categorizado el producto en varios parámetros, como bajo, medio, alto.

Género: Masculino y Femenino.

Descuento ofrecido: Descuento ofrecido en ese producto específico.

Peso en gms: Es el peso en gramos.

Llegó a tiempo: Es la variable objetivo, donde 1 Indica que el producto NO ha llegado a tiempo y 0 indica que ha llegado a tiempo.

¿Qué problema se quiere resolver?

Una empresa internacional de comercio electrónico quiere descubrir información clave de su base de datos de clientes, para ellos necesita utilizar algunas técnicas de análisis de datos para comprender mejor a sus clientes y mejorar los tiempos de envíos. Algunas de las preguntas a responder con este análisis son:

¿Cuál fue la clasificación de los clientes?

¿El producto fue entregado a tiempo?

¿Se responden las consultas de los clientes a tiempo?

Si la importancia del producto es elevada. ¿Por este motivo se entrega con mayor frecuencia a tiempo o no varía si el producto es de mayor relevancia?

La empresa vende productos electrónicos.

## 2. Dataset ventas Supermercados La Favorita

Este dataset recoge el registro histórico de ventas de la cadena de retail La Favorita, radicada en Ecuador. Se desglosa la información por fecha, sucursal y rubro, entre otras variables.

Al tratarse de un dataset utilizado en un concurso de Kaggle, se proveen un dataset de train y uno de test, sobre el cual intentaremos predecir las ventas. El dataset de train cuenta con más de 3 millones de registros y las siguientes variables:

- ID (único para cada registro)
- date (fecha de las ventas registradas)
- store\_nbr (identificador de la sucursal que se releva)
- family (rubro/familia de productos que se releva)
- sales (contado en unidades de productos, no en unidad monetaria)
- onpromotion (variable binaria que determina si la familia de productos es afectada por promociones o no)

Entre las demás tablas, se listan, por fecha, los días festivos, precios históricos del petróleo (el autor de la competencia lo incluye como variable relevante en la economía ecuatoriana), y cantidad de transacciones, por fecha y por sucursal.

**NOTA: Atento a las pautas del desafío y la recomendación de no usar datasets demasiado extensos, podríamos utilizar un rango temporal menor o filtrar determinadas sucursales/rubros para centrar el análisis.**