

## ANALISIS DEL DATASET

- ORIGEN:

Los datos provienen de la base de datos de un proveedor de servicios de una compañía multinacional. Originalmente, estos datos fueron utilizados para la gestión y análisis de ventas.

Para este proyecto, los datos de razón social fueron modificados para preservar la información y deliberadamente "ensuciados" con el objetivo de simular un conjunto de datos no depurado, lo que permite trabajar con un escenario más realista y evaluar las herramientas de procesamiento de datos en condiciones más cercanas a las que se podrían encontrar en un entorno profesional.

- DETALLE:

COLUMNA	DESCRIPCIÓN	TIPO DE DATO	VALORES POSIBLES
DISTRIBUIDOR	Nombre de fantasía o razón social del distribuidor.	String (255)	Representan el nombre del distribuidor (por ejemplo, TOTAL CENTRO S.R.L.   BRINDARSE S.A)
ÁREA	Area geográfica de distribución.	String (255)	División de áreas geográficas de Argentina (Atlántico y Cordillera   Centro y Cuyo   Costa y La Pampa   GBA Norte   GBA Sur   NEA y Litoral   NOA)
ARTÍCULO	Código del articulo vendido.	Númérico (entero)	Códigos numéricos (desde el 300030736 hasta el 300063271)
DESCRIPCIÓN	Descripción detallada del artículo.	String (255)	Texto que describe el artículo, incluye información como el peso y cantidad de unidades por bulto (por ejemplo "TwistosminitQueso 95gx30x1")
CANTIDAD	Cantidad de unidades vendidas del artículo.	Númérico (entero)	Números enteros que indican cantidad en unidades (desde -33 hasta 3284940)
CAPACIDAD	Peso de las unidades vendidas expresado en gramos, si es aplicable.	Númérico (flotante)	Números decimales que indican capacidad en gramos (desde -5,005 hasta 25294,038)
LINEA	Nombre de la línea del producto. Agrupación Nivel 2.	String (255)	Representan las subcategorías de productos (por ejemplo Gold Crakers   Platinun Crakers)
RUBRO	Nombre del rubro del producto. Agrupación Nivel 1.	String (255)	Corresponden a la categoría del producto (Crakers   Cereals   Salty