

Descripción de los Datos

Para el trabajo práctico de las materias del módulo analista de datos, elegimos un dataset que originalmente fue recurso en un curso de Excel. Consideramos que es útil para el presente trabajo de análisis ya que cuenta con extensos registros tanto cuantitativos como cualitativos, lo que permite hacer un análisis exhaustivo de las variables allí informadas. Por otra parte, es importante mencionar que los datos están en crudo por lo cual, requieren de un importante trabajo de limpieza para obtener datos de calidad que permitan el posterior trabajo de análisis.

El presente dataset evidencia las ventas realizadas por la empresa multinacional “Cuchufrito”, fabricante de artículos varios tales como calzado, vestimenta, accesorios, entre otros. Muestra los datos distribuidos entre las siguientes columnas:

País: Refiere al país donde fue realizada la venta. Entre ellos se encuentran la mayoría de países de América, y algunos de Asia y Europa. Tipo de dato String.

```
['Argentina' 'Bolivia' 'Brazil' 'Belize' 'Canada' 'Chile' 'China'
 'Colombia' 'Costa Rica' 'Germany' 'Dominican Republic' 'Ecuador'
 'Guatemala' 'Honduras' 'Italy' 'Kazakhstan' 'Mexico' 'Nicaragua' 'Panama'
 'Peru' 'Puerto Rico' 'Paraguay' 'El Salvador' 'United States' 'Uruguay'
 'Venezuela' 'Bermuda' 'Aruba']
```

Método de orden: Refiere al medio a través del cual se concretó la venta del producto.

Tipo de dato String.

```
['Representante' 'Mail' 'Teléfono' 'Website' 'Especial' 'MercadoLibre']
```

Tipo de comercio: Refiere a las características del comercio donde se efectuó la venta.

Tipo de dato String.

```
['Mayorista' 'De proximidad' 'Grandes Almacenes' 'Especializado' 'Hipermercado']
```

Línea: Hace referencia al rubro que pertenecen un grupo de productos.

Tipo de dato String.

```
['Mascotas' 'Infantil' 'Hombre' 'Adultos' 'Mujer' 'Unisex']
```

Tipo de producto: Dentro de la línea, hay una subdivisión que agrupa a determinado tipo de productos. Tipo de dato String.

```
['Juguetes' 'Calzado' 'Vestimenta' 'Lentes' 'Joyas' 'Accesorios']
```

Producto: Refiere a cada producto específico. Tipo de dato String

Año: Es el año en que se realizó la venta, tipo de dato numérico (entero o Date(Year))

```
[2019 2020 2021]
```

Cuatrimestre: Período del año en el que se realizó la venta. Tipo de dato String

```
['C1 2019' 'C2 2019' 'C3 2019' 'C4 2019' 'C1 2020' 'C2 2020' 'C3 2020'
 'C4 2020' 'C1 2021' 'C2 2021' 'C3 2021' 'C4 2021']
```

Ventas: Se refiere al ingreso bruto por ventas, tipo de dato numérico (decimal o flotante)

Cantidad: número de productos vendidos. Tipo de dato numérico (entero)

Utilidad bruta (porcentaje): Es la utilidad bruta expresada en porcentaje sobre el total de ventas (ganancia bruta). Tipo de dato numérico (decimal o flotante)

Utilidad bruta: Se refiere a la ganancia que se obtiene de la venta de un producto luego de restarle los costos asociados al mismo. Tipo de dato numérico (decimal o flotante)

Costos: Gastos asociados a la venta del producto. Tipo de dato numérico (decimal o flotante)

Tipo de datos:

Como vemos a continuación, los tipos de datos presentes en el CSV no están formateados de la manera correcta.

pais	object →Tipo Correcto (String) [cadena de caracteres es_ES.UTF-8]
metodo_orden	object →Tipo Correcto (String) [cadena de caracteres es_ES.UTF-8]
tipo_comercio	object →Tipo Correcto (String) [cadena de caracteres es_ES.UTF-8]
linea	object →Tipo Correcto (String) [cadena de caracteres es_ES.UTF-8]
tipo_producto	object →Tipo Correcto (String) [cadena de caracteres es_ES.UTF-8]
producto	object →Tipo Correcto (String) [cadena de caracteres es_ES.UTF-8]
año	int64 →Tipo Correcto (Int) [periodo hasta año actual]
cuatrimestre	object →Tipo Correcto (String) [cadena de caracteres es_ES.UTF-8]
ventas	object → debe ser Numérico (Float) [flotante no negativo]
cantidad	object → debe ser Numérico (Int) [Entero no negativo]
utilidad_bruta_p	object → debe ser Numérico (Float) [float no negativo(0-100)]
utilidad_bruta	object → debe ser Numérico (Float) [float no negativo]
costos	object → debe ser Numérico (Float) [float no negativo]