



## **INFORME DEL PROYECTO:**

TECNICATURA SUPERIOR EN CIENCIAS DE DATOS E INTELIGENCIA ARTIFICIAL

**ESPACIO CURRICULAR:** ANALISTA DE DATOS I

**PROF.** UGARTE MARCOS // PERATTA NAHUEL

### **AMERICAN THINGS – ANALISIS Y PROCESAMIENTO DE DATOS, PARA SELECCIÓN DE ESTRATEGIAS DE VENTAS.**

**Evidencia N° 2**

**AÑO 2025**

### **Grupo N°: 6**

#### **Integrantes:**

Guaraz Emanuel - <https://github.com/JEmanuelG>

Direni Carlos - <https://github.com/Cdireni1>

Ryser Lucas - <https://github.com/lucasryser6>

Lobo Bruno - <https://github.com/lobosouza>

Allende Olmedo Nicolás - <https://github.com/AllendeNicolas>

Lugar de presentación: ISPC (Instituto Superior Politécnico de Córdoba)

## Consigna 1)

### Origen de los datos

El origen de los datos proviene, de una serie de data set's, obtenidas de internet, la cual fue reducida en primer lugar, de 5000 registros a los 200 solicitados, para su manipulación y futuro procesamiento, por el espacio curricular de **Analista de Datos I**.

El equipo, Grupo N° 6, tomo la decisión de asignarle el actual data set, a una empresa ficticia denominada, **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 Data Set 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 maldispuestos**, 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.

## Consigna 2)

### Descripción de las columnas:

A continuación presentamos un cuadro, en el cual se detallan la descripción de cada dato que almacena cada columna, su carácter, descripción y, los posibles valores que puede contener cada una de ellas.

Columna	Tipo de dato	Descripción	Posibles valores
<b>Factura</b>	Texto(str)	Código único de la factura de la compra.	Formato: XXX-XX-XXXX
<b>Sucursal</b>	Texto(str)	Sucursal donde se realizó la compra.	<ul style="list-style-type: none"> <li>• "A"</li> <li>• "B"</li> <li>• "C"</li> </ul>
<b>Ciudad</b>	Texto(str)	Ciudad en la que se encuentra la sucursal.	<ul style="list-style-type: none"> <li>• "Nueva York"</li> <li>• "Houston"</li> <li>• "Chicago"</li> </ul>
<b>Tipo</b>	Texto(str)	Tipo de cliente: si es miembro o no.	<ul style="list-style-type: none"> <li>• "Member"</li> <li>• "Normal"</li> </ul>
<b>Genero</b>	Texto(str)	Género del cliente.	<ul style="list-style-type: none"> <li>• "Male"</li> <li>• "Female"</li> </ul>
<b>Categoria</b>	Texto(str)	Categoría del producto o servicio adquirido.	<ul style="list-style-type: none"> <li>• "Health and beauty"</li> <li>• "Electronics accessories"</li> <li>• Sports and travel</li> <li>• Fashion accessories</li> <li>• Food and beverages</li> <li>• Home and lifestyle</li> </ul>
<b>Costo</b>	N Numérico (float)	Costo de los productos vendidos (Cost of Goods Sold).	Valor numérico decimal
<b>Ventas</b>	N Numérico (float)	Monto total de la venta.	Valor numérico decimal
<b>Fecha</b>	Fecha (str/date)	Fecha en la que se realizó la transacción.	Formato: MM/DD/AAAA

<b>Hora</b>	Hora (str/time)	Hora exacta en que se registró la compra.	Formato: HH:MM
<b>Método de Pago</b>	Texto(str)	Método de pago utilizado.	<ul style="list-style-type: none"> <li>• "Ewallet"</li> <li>• "Cash"</li> <li>• "Credit Card"</li> </ul>
<b>Reseña</b>	Numérico (float)	Puntuación del cliente sobre su experiencia (escala del 1 al 10).	Valor decimal entre 1 y 10

### Consigna 3)

#### Estado de los datos:

Los datos **no se encuentran en estado RAW**, ya que el data set, se encontraba con todos su datos, perfectamente ubicados, sin restricciones, ni inconvenientes alguno; y fueron editados manualmente para introducir inconsistencias, valores faltantes, duplicados, o incompletos.

Esto permitirá realizar un proceso completo de análisis, validación y limpieza de datos, sobre un nuevo Dataset en su forma más básica, sin procesar, en un formato que guarda la información, tal y como fue captada, en el desarrollo del proyecto.