**DATOS ADICONALES**

**Importación de datos:**  
Se importó el archivo Global\_Cybersecurity\_Threats\_2015-2024.csv a Excel (delimitado por comas, codificación UTF-8). Se genero una segunda base de datos a raíz de la principal que se indico en la web utilizando IA para añadir ciertos errores para la limpieza y estandarización de esta base de datos. Global\_Cybersecurity\_Threats\_dirty.csv

**Comparación con dataset original “dirty”:**  
El dataset generado por IA contenía formatos mezclados (símbolos $, €, comas, texto como “six hundred”), países duplicados y categorías inconsistentes.  
En el proyecto final todos los valores fueron convertidos a números reales, los textos normalizados y se añadieron columnas analíticas no presentes en el original.

**Creación de ID:**  
Se añadió una columna con numeración secuencial al no existir un identificador único en el dataset original.

**Ajuste de tipos de dato:**  
Se asignaron formatos correctos a cada campo (Texto, Número, General) según su naturaleza.

**Eliminación de duplicados:**  
De 3006 registros iniciales se detectaron 5 duplicados, quedando **3001 válidos**.

**Corrección de incoherencias:**  
Se unificaron países y valores escritos de forma distinta (ej. *United States → USA*, *United Kingdom → UK*).  
También se corrigieron celdas con letras en campos numéricos y espacios invisibles.

**Corrección del año erróneo:**  
El valor **2035** fue reemplazado por **2025** para mantener la coherencia temporal (rango 2015–2024).

**Creación de nuevas columnas:**

* **Región:** continente asignado a cada país.
* **Años desde ataque:** =AÑO(HOY()) - [@[Year]]
* **Gran pérdida:** identifica pérdidas > $7M.
* **Criticidad:** combina usuarios afectados y pérdida por usuario.
* **Resolución:** categorizada como *Rápida*, *Media* o *Tardía* según las horas.

**Limpieza de valores nulos y extremos:**  
Se eliminaron registros con pérdidas económicas = 0 o años fuera del rango lógico.

**Normalización de texto:**  
Se aplicó formato consistente con MINUSC, ESPACIOS y NOMPROPIO para nombres y categorías.  
Tipos de ataque unificados (ej. *Phishing*, *DDoS*, *SQL Injection*).