**Diplomado en Analítica y Ciencia de Datos**Universidad de Colima - TRANSICION

**Proyecto Final**

**Análisis de Precios de Laptops**

Equipo No. 11

Integrantes:

|  |
| --- |
| * Samantha Michelle Pineda Esparza |
| * Luis Alberto Pineda García |
| * Nazario Bustos Eusebio |
| * Julio Daniel Uribe Gómez |
| * Cesar Edwin Hernández Ochoa |
|  |

**Selección y Justificación de la Base de Datos**

* **Búsqueda**

La búsqueda de la fuente de datos se realizó en el sitio web Kaggle con dirección www.kaggle.com.

**Fuente de Datos elegida:** Dataset archivo CSV (laptop\_prices.csv)

**URL :** <https://www.kaggle.com/datasets/owm4096/laptop->prices?select=laptop\_prices.csv

* **Justificación**

Revisando opciones entre distintos sitios de internet y diversos temas, se optó por elegir un dataset de interés general, ya que todas las profesiones y oficios estamos relacionados con el uso de equipos de cómputo. En este caso se trata de una lista de precios de laptops, tomando en cuenta diversos datos como la compañía a la que pertenecen, marca de CPU, modelo de CPU, Sistema Operativo, características gráficas, tamaño de monitor entre otras.

**Preparación y Limpieza de Datos**

* **Descripción de la Base de Datos**

La Base de datos elegida (archivo CSV) cuenta con 1275 registros y 23 columnas descriptivas de los equipos.

|  |  |  |
| --- | --- | --- |
| **No.** | **Columnas** | **Descripción** |
| 1 | Company | Nombre de Compañía o Fabricante |
| 2 | Product | Nombre del Producto (Laptop) |
| 3 | TypeName | Tipo de Laptop |
| 4 | Inches | Pulgadas tamaño del display monitor |
| 5 | Ram | Memoria Ram |
| 6 | OS | Sistema Operativo |
| 7 | Weight | Peso |
| 8 | Price\_euros | Precio en euros |
| 9 | Screen | Tecnología de Display o Pantalla |
| 10 | ScreenW | Ancho de Display o Pantalla |
| 11 | ScreenH | Alto de Display o Pantalla |
| 12 | Touchscreen | Pantalla tipo táctil |
| 13 | IPSpanel | Si la computadora portátil tiene o no un IPSpanel. |
| 14 | RetinaDisplay | Si la computadora portátil tiene o no un Retina Display |
| 15 | CPU\_company | Marca o compañía de CPU |
| 16 | CPU\_freq | Frecuencia en Hz de CPU |
| 17 | CPU\_model | Modelo de CPU |
| 18 | PrimaryStorage | Almacenamiento Primario en Gigabytes |
| 19 | SecondaryStorage | Almacenamiento Secundario en Gigabytes |
| 20 | PrimaryStorageType | Tecnología o Tipo Almacenamiento Primario |
| 21 | SecondaryStorageType | Tecnología o Tipo Almacenamiento Secundario |
| 22 | GPU\_company | Compañía o marca de Tarjeta Gráfica |
| 23 | GPU\_model | Modelo de Tarjeta Gráfica |

* **Limpieza de los datos**

La Base de datos elegida no contaba con ningún dato erroneo ni descripciones repetitivas o mal redactadas. Por lo tanto nuestra base de datos a usar la renombramos como:

* laptop\_prices\_depurada.csv

Durante este proceso se realizaron 3 archivos csv adicionales con agrupamientos necesarios para realizar estadísticas específicas.

* Resumen\_precios\_Company\_AMD.csv
* Resumen\_precios\_Company\_INTEL.csv
* Resumen\_precios\_Company\_cpu\_model.csv