**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.

Tabla de Columnas y Tipos de Datos

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

* **Limpieza de los datos**
  + Identificación y manejo registros faltantes

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

* laptop\_prices\_depurada.csv
  + Corrección de errores en los datos

Se realizaron agrupamientos y no se encontraron descripciones e inconsistencias en los datos como descripciones con caracteres especiales.

* + Transformaciones necesarias

El dataset elegido cuenta con los suficientes campos, para realizar las estadísticas deseadas que no fue necesario generar nuevas columnas calculadas al dataset base. Sin embargo en la creación de resúmenes estadísticos se agregaron las columnas necesarias para generar los mismos.

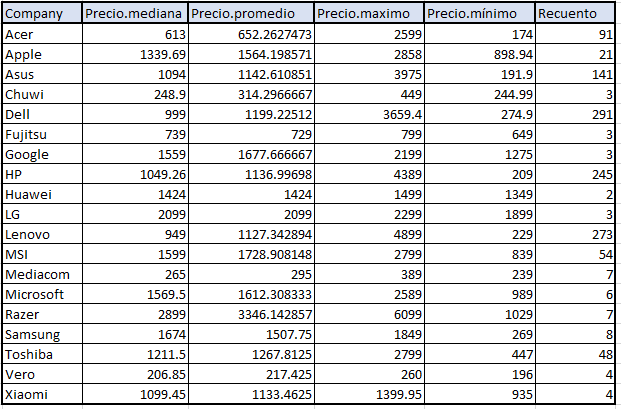
**Análisis Exploratorio de Datos (EDA)**

Apoyados del Lenguaje de Programación R se realizaron los siguientes análisis.

Estadísticas Generales de Precios de Laptos (Euros)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Mínimo** | **1er Qu.** | **Mediana** | **Promedio** | **Moda** | **3er Qu.** | **Máximo** |
| 174 | 609 | 989 | 1135 | 1099 | 1496 | 6099 |

Estadísticas de laptops INTEL por Compañía (Resumen\_precios\_Company\_INTEL.csv)



* Resumen\_precios\_Company\_INTEL.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