**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 |
|  |

**1.- 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](http://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.

Además de esto el conocimiento adquirido puede ayudar de alguna manera a la toma de decisiones en cuestión de compra.

* **Objetivo**

Realizar un análisis exploratorio de los datos a través de lenguaje R, y visualización de los resultados en software Tableau para observar de manera intuitiva y clara el comportamiento de los precios de laptops según sus características y también cuales de éstas mencionadas son las que los fabricantes toman más en cuenta para el diseño de sus equipos de cómputo.

**2.- 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**
  1. 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
  1. Corrección de errores en los datos

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

* 1. 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.

**3.- Análisis Exploratorio de Datos (EDA)**

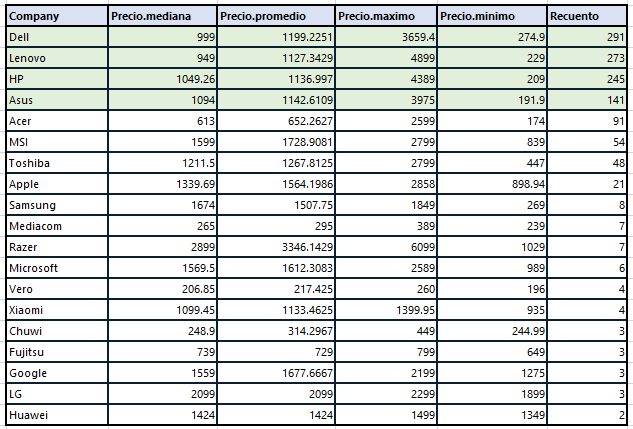
Apoyados del lenguaje de programación “R” se realizaron los siguientes análisis.

Estadísticas Generales de Precios de Laptops (Euros)

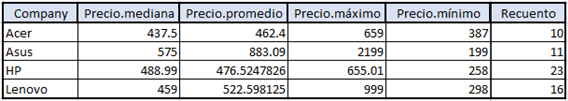
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Mínimo** | **1er Qu.** | **Mediana** | **Promedio** | **Moda** | **3er Qu.** | **Máximo** |
| 174 | 609 | 989 | 1135 | 1099 | 1496 | 6099 |

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

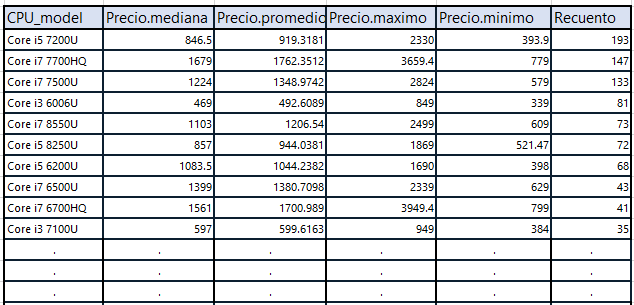
Estadísticas de Precios de laptops INTEL por Compañía (Resumen\_precios\_Company\_INTEL.csv)



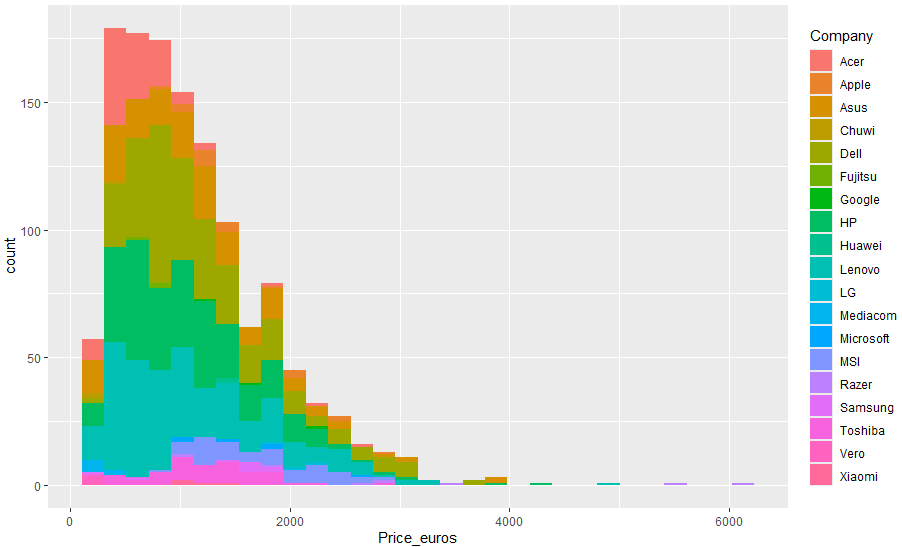
Estadísticas Precios de laptops AMD por Compañía (Resumen\_precios\_Company\_AMD.csv)



Estadísticas de Precios de laptops por CPU Model (Resumen\_precios\_cpu\_model.csv)

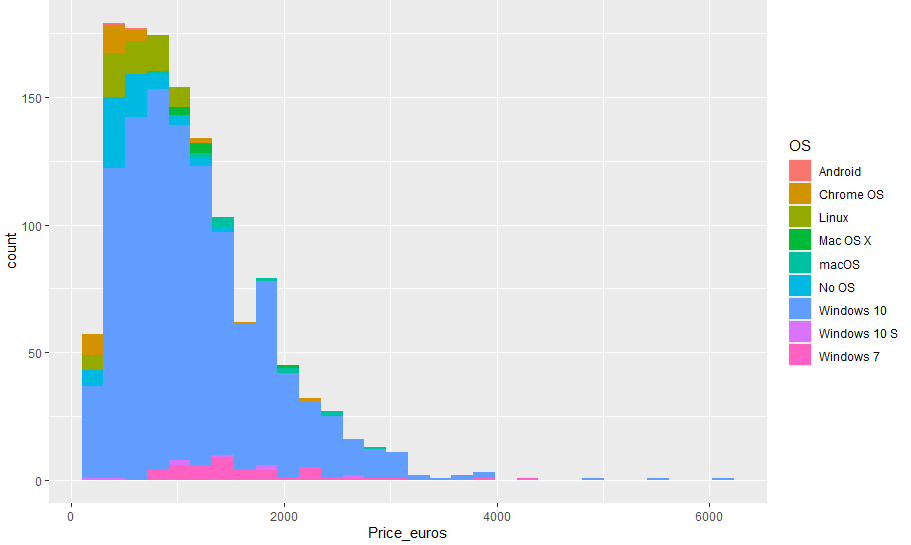


Histogramas, gráfico de barras y Distribución de los datos  
  
*Histograma de Precios por Compañías*

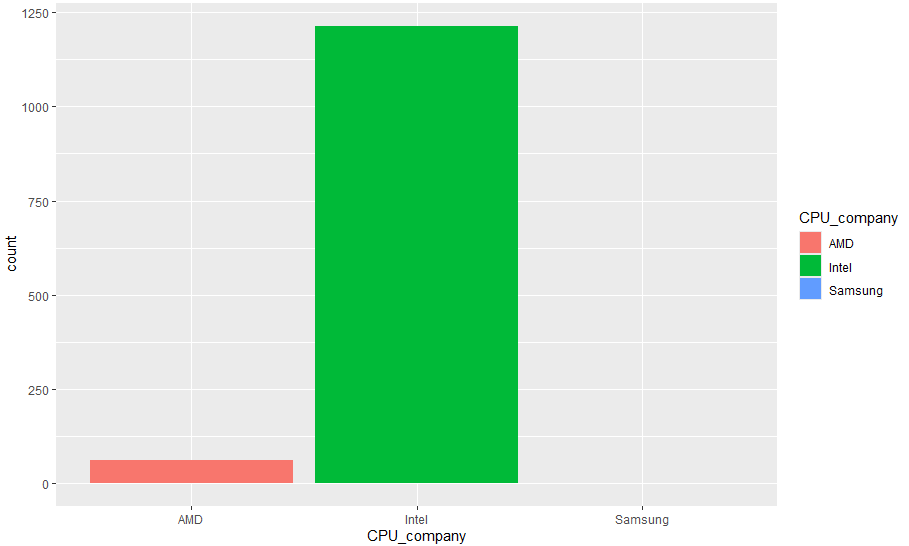


Observaciones: Existe un grupo grande de laptops con rango de precios entre el 1er Quartil que equivale a 609 Euros a 2,000 Euros aproximadamente.

*Histograma de Precios por Sistema Operativo*



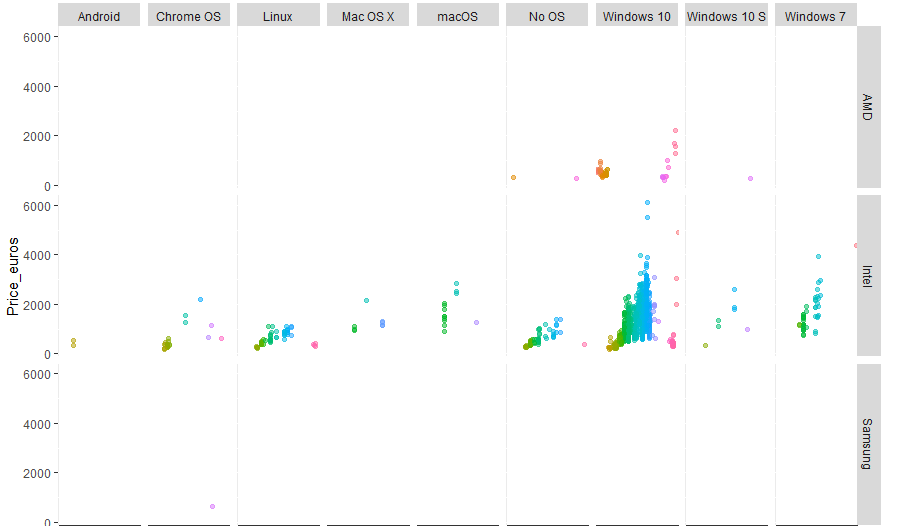
**Observaciones:** Existe un grupo grande de laptops con sistema operativo Windows 10

*Gráfico de Barras - Recuento de Laptops por CPU Company*

**Observaciones:**   
Es evidente que la mayoría de los equipos laptops de esta base de datos, la mayoría cuentan con un CPU de la compañía INTEL.

*Diagrama de dispersión de los datos por Precio, Compañía y CPU Compañía*

**Observaciones**:   
Existe una gran distribución de Laptops con CPU Intel, de las Compañías Dell, Lenovo, HP y Asus

*Distribución de datos en Matriz Grid de Sistema Operativo x CPU Company*

**Observaciones:**   
Existe una gran distribución y concentración en equipos con Windows 10 y la mayoría cuentan con CPU INTEL que es el de mayor demanda o preferencia en el mercado hoy en día por el desempeño y durabilidad.

**4.- Visualización de Datos**

* **Herramienta de visualización de datos**

|  |  |
| --- | --- |
| Software | Tableau |
| URL Público: | https://public.tableau.com/app/profile/...... |
| Archivo fuente | ProyectoFinal.twb |

Software: Tableau  
URLTableau público:

FOTOS DE DASHBOARD

FOTOS DE DASHBOARD

FOTOS DE DASHBOARD

**5.- Interpretación y Conclusiones**

* **Resultados del Análisis**
  1. Existe una mayor preferencia por los fabricantes de laptop la integración de CPU de la compañía INTEL.
  2. La Mayoría de los fabricantes de laptops, incluyen Sistema Operativo Windows 10 como instalación base.
  3. Existe una gran distribución de Laptops con CPU Intel, de las Compañías Dell, Lenovo, HP y Asus.
  4. El precio promedio de las laptops independientemente de sus características está en 1,135 Euros.
  5. Existe un grupo grande de laptops con rango de precios entre el 1er Quartil que equivale a 609 Euros a 2,000 Euros aproximadamente.
* **Conclusiones**

Sin duda alguna, la realización del análisis de los datos nos conduce a la mejor toma de decisiones. En esta ocasión, logramos comprender el objetivo que se pretendía, sobre conocer cuales equipos de cómputo de los que se encuentran en el mercado están más accesibles en cuanto a precio, entender que el sistema operativo más confiable según fabricantes es Windows 10. Cabe resaltar que solo se está tomando en cuenta la estadística en cuanto a existencia de modelos y su precio del universo de datos analizado, sabiendo de antemano que las mejores características que un equipo de cómputo pueda tener, el conjunto de ellas son las que en realidad son las definirán su costo.

Sugerencias para trabajos futuros: Sería interesante contrastar este análisis de precios de laptops con un dataset de ventas y observar su preferencia de acuerdo a zonas geográficas como Países, Estado, etc.