**Guía**

1. Entendiendo el problema (Introducción)

1.1. Propósito

1.2. Intuición

1. Explorando los datos

2.1. Identificando los datos más difíciles de tratar

2.2. Repeticiones de datos

2.3. Saneando los datos

2.4. Correlación de los datos

1. Preprocesando los datos

3.1. Creando nuevas columnas

3.2. Eliminando las viejas columnas

3.3. Eliminando outliers con ElasticNet

1. Entrenamiento

4.1. Contrastando el MAE

4.2. Evadiendo el overfitting

4.3. Búsqueda de hiper parámetros

1. Validación

5.1. Haciendo cross-validation

5.2. Conclusiones de cada modelo

1. Predicciones

5.1. Evolución de los resultados

5.2. Cosas que no nos han funcionado

1. Otros gráficos para contrastar nuestras hipótesis

5.1 Relplot

5.2 Caja y Bigotes

5.3 Pie

5.4 HeatMap

5.5 Scatter

5.6 Histograma

8. Conclusiones

8.1. Contrastes con intuición inicial

8.2. Idk

1. **Entendiendo el problema**

**1.1 Propósito**

Se nos ha presentado un problema cuyo propósito consiste en diseñar un modelo usando machine learning que pueda determinar con poco margen de error el precio que tendrá un ordenador en base a su compañía, modelo, componentes, tamaño, resolución, peso y sistema operativo.

**1.2 Intuición**

Sin analizar los datos, suponemos que tendrán gran importancia los componentes como la RAM, la GPU y la CPU al igual que su fabricante. Pero también tenemos en cuenta que existen sistemas operativos gratuitos o de pago y el tamaño posiblemente afecte también al precio. Suponemos que las compañías más caras serán Apple y Razer, esta última centrada en ordenadores gaming.

**2. Explorando los datos**

**2.1 Identificando los datos más difíciles de tratar**

Las columnas de Product (el nombre del modelo), Screen Resolution, GPU y CPU habrá que partirlas en varias columnas ya que viene demasiada información. En cuanto a Memory sumaremos la memoria total en GB y los clasificaremos en función de si tienen HDD, SSD o memoria flash (mediante one-hot encoding).

**2.2 Repeticiones de datos**

(Insertar código que encuentre duplicados con distinto precio)

**2.3 Saneando los datos**

**2.4 Correlación de los datos**