

1. **¿Qué columnas eliminaron (en caso se haya eliminado)**
  - a. **Marca:** no tiene datos nutridos que nos aporten valor
  - b. **Fecha registro:** tiene más del 50% de datos nulos
  - c. **Asientos\_traseros\_plegables:** tiene más del 70% de datos nulos.
2. **¿Qué se hizo con los nulos y cómo se limpiaron las columnas?**
  - a. **Modelo:** Elimino filas con nulos
  - b. **KM:** Elimino filas con nulos
  - c. **Potencia:** Elimino filas con nulos
  - d. **Tipo\_gasolina:** Elimino filas con nulos
  - e. **Color:** cambio los nulos por "sin\_color".
  - f. **Tipo\_coche:** cambio los nulos por "sin\_tipo".
  - g. **Volante\_regulable:** Elimino filas con nulos
  - h. **Aire\_acondicionado:** cambio los nulos por "sin\_aire".
  - i. **Cámara\_trasera:** Elimino filas con nulos
  - j. **Elevalunas\_electrico:** Elimino filas con nulos
  - k. **Bluetooth:** cambio los nulos por "sin\_bluetooth".
  - l. **Alerta\_lim\_velocidad:** cambio los nulos por "sin\_alerta".
  - m. **Precio:** pongo media para el valor nulo
  - n. **Fecha\_venta:** Elimino filas con nulos
3. **¿Comentarios del análisis invariable, están todas ok?**

Se pueden observar outliers en km, precio y potencia que hay que tratar, también que las fechas de venta todos son el día 1 de cada mes
4. **¿Hay alguna con outliers? ¿Hay alguna por agrupar?**

**Outliers:**

  - a. Km: hay uno menor que 0 y hay varios arriba de 4e6. Elimino los valores.
  - b. Precio: hay varios debajo de 1000 (elimino)
  - c. Potencia: hay 1 con valor 0 que elimino.
5. **¿Análisis de Correlación inicial, hay alguna variable correlacionada?**

En este momento solo tenemos km, potencia, gps y precio como numéricas. Y no se percibe una gran correlación, la más alta es Potencia con 0.54
6. **¿Análisis variable vs target, hay algún insight interesante?**
  - Precio – km: influye de manera media, entre más a recorrido, menor el precio.
  - Potencia - km, entre más potencia tiende a aumentar el precio.
  - Tipo gasolina – precio: si influye, pero más cuando hablamos de híbrido y eléctrico.
  - Color, y tipo coche: influyen, pero no en gran manera.
  - bluetooth y aire acondicionado, no influye en el precio

7. ¿Transformación de categóricas a numéricas, que variables van a transformar?

Modelo, tipo\_gasolina, color, tipo\_coche, aire\_acondicionado, bluetooth, alerta\_lim\_velocidad.

8. ¿Qué técnica se va usar?

- Datetime:** Fecha\_venta y separar por columnas día, mes, año
- astype(int):** Boolean que sean true o false a numérico
- OrdinalEncoder:** aire\_condicionado, bluethooth, alerta\_lim\_velocidad.
- Label Encoder:** fecha\_venta\_nombredia, modelo, tipo\_gasolina, color; tipo\_colche.
- MinMaxScaler:** para lista numérica.

9. ¿Análisis de correlación final, hay alguna variable correlacionada?

No, la correlación más alta es de precio potencia con 0.64 y la siguiente es modelo\_le con 0.49

LISTA DE COLUMNAS

COLUMNA	TIPO
Km	Float
Potencia	Float
Precio	Float
Log_precio	Float
Gps_bool	Int
Volante_regulable_bool	Int
Cámara_trasera_bool	Int
Elevalunas_electrico_bool	Int
Fecha_venta_mes	Float
Fecha_venta_año	Float
Aire_acondicionado_oe	Float
Bluetooth_oe	Float
Alerta_lim_velocidad_oe	Float
Fecha_venta_nombredia_le	Int
Modelo_le	Int
Tipo_gasolina_le	Int
Color_le	Int
Tipo_coche_le	Int

	km	potencia	precio	log_precio	gps_int	volante_regulable_int	camara_trasera_int	elevalunas_electrico_int	fecha_venta_mes	fecha_venta_año	aire_acondicionado_oe	bluetooth_oe	alerta_lim_velocidad_oe	fecha_venta_nombredia_le	modelo_le	tipo_gasolina_le	color_le	tipo_coche_le
0	0.351978	0.129498	11300.0	4.053078	1	1	0	1	0.000	1.0	0.0	1.0	1.0	1	3	0	1	5
1	0.033838	1.000000	69700.0	4.843233	1	1	0	0	0.125	1.0	0.0	0.0	0.0	4	62	3	5	0
2	0.459850	0.087963	10200.0	4.008600	1	0	0	1	0.125	1.0	0.5	0.5	0.5	4	22	0	10	5
3	0.320849	0.157407	25100.0	4.399674	1	1	0	1	0.125	1.0	0.0	0.0	1.0	4	32	0	7	0
4	0.243031	0.273148	33400.0	4.523746	1	1	0	0	0.375	1.0	0.0	0.0	0.0	3	34	0	8	5