Carlos Francisco Mejía Arias

1. **¿Qué columnas eliminaron (en caso se haya eliminado)**
   1. **Marca:** no tiene datos nutridos que nos aporten valor
   2. **Fecha registro:** tiene más del 50% de datos nulos
   3. **Asientos\_traseros\_plegables:** tiene más del 70% de datos nulos.
2. **¿Qué se hizo con los nulos y cómo se limpiaron las columnas?**
   1. **Modelo:** Elimino filas con nulos
   2. **KM:** Elimino filas con nulos
   3. **Potencia:** Elimino filas con nulos
   4. **Tipo\_gasolina:** Elimino filas con nulos
   5. **Color:** cambio los nulos por “sin\_color”.
   6. **Tipo\_coche:** cambio los nulos por “sin\_tipo”.
   7. **Volante\_regulable:** Elimino filas con nulos
   8. **Aire\_acondicionado:** cambio los nulos por “sin\_aire”.
   9. **Cámara\_trasera:** Elimino filas con nulos
   10. **Elevalunas\_electrico:** Elimino filas con nulos
   11. **Bluetooth:** cambio los nulos por “sin\_bluetooth”.
   12. **Alerta\_lim\_velocidad:** cambio los nulos por “sin\_alerta”.
   13. **Precio:** pongo media para el valor nulo
   14. **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

1. **¿Hay alguna con outliers? ¿Hay alguna por agrupar?**

**Outliers:**

* 1. Km: hay uno menor que 0 y hay varios arriba de 4e6. Elimino los valores.
  2. Precio: hay varios debajo de 1000 (elimino)
  3. Potencia: hay 1 con valor 0 que elimino.

1. **¿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

1. **¿Análisis variable vs target, hay algún insight interesante?**

* Pecio – 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 hibrido y eléctrico.
* Color, y tipo coche: influyen, pero no en gran manera.
* - bluetooth y aire acondicionado, no influye en el precio

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

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

1. **¿Qué técnica se va usar?**
   1. **Datetime:** Fecha\_venta y separar por columnas día, mes, año
   2. **astype(int):** Bolean que sean true o false a numérico
   3. **OrdinalEncoder:** aire\_condicionado, bluethooth, alerta\_lim\_velocidad.
   4. **Label Encoder:** fecha\_venta\_nombredia, modelo, tipo\_gasolina, color; tipo\_colche.
   5. **MinMaxScaler:** para lista numérica.
2. **¿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 |

