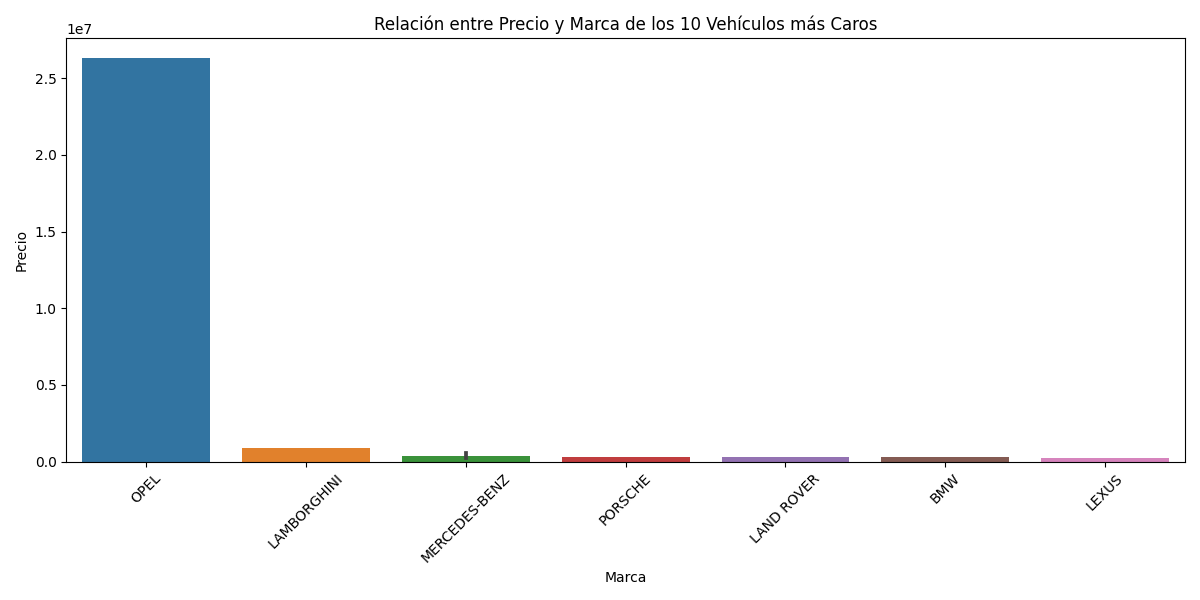
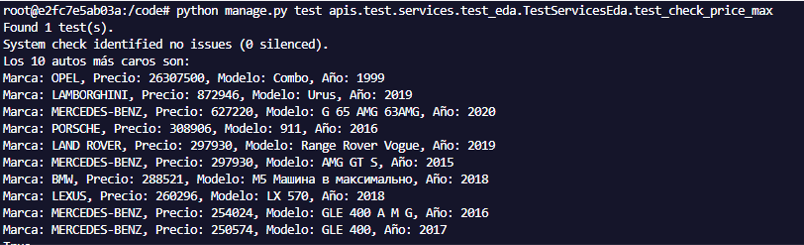
**ANÁLISIS EXPLORATORIO DE DATOS (EDA)**

A continuación iniciamos el EDA sobre el dataset facilitado, básicamente vamos a analizar tres casos de usos, y vamos a exponer unas conclusiones de los mismos.

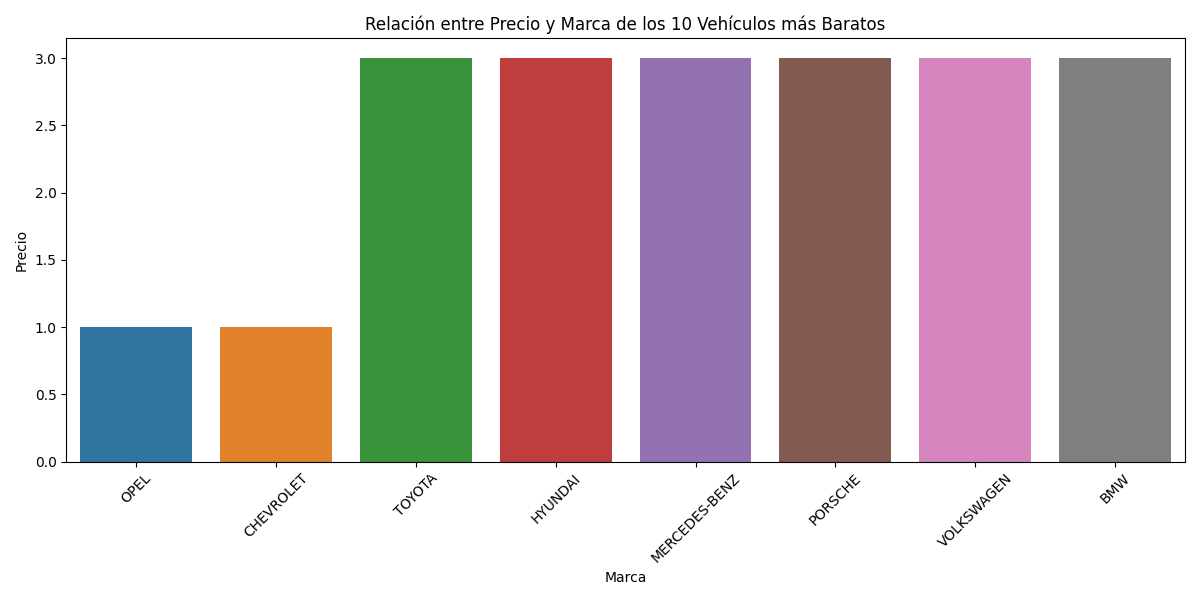
**Marcas más caras:**

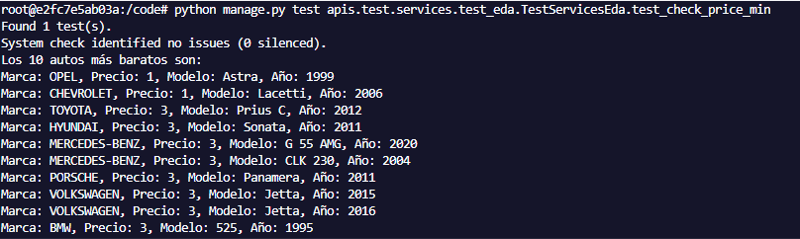




Como podemos observar tenemos precios muy por encima de la media, de más de 20 millones, con una diferencia de 25.000.000 con respecto al segundo lugar.

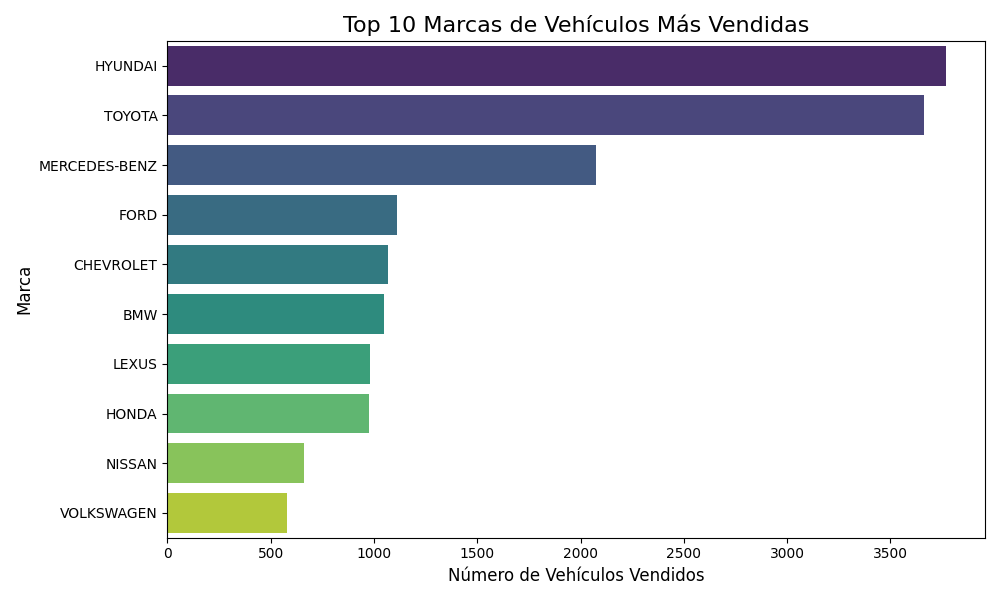
**Marcas mas barata:**





Igualmente como el caso de uso anterior vemos extremos no muy lógicos en donde se vende un auto en 1.

**Marca mas vendida**

****

Adicional a ello podemos observar las marcas vendidas que tenemos en el dataset

**Conclusiones**

En los dos primeros casos de uso podemos observar datos poco realistas, dado que se van a los extremos, marcas de vehículos en 1 como en más de 25 millones, en este escenario, lo que se recomienda es solicitar criterios de eliminación, por ejemplo que todas las ventas de 1, sean eliminadas.

Esto es importante ya que básicamente el modelo que se va a aplicar su efectividad en las predicciones va a depender de los datos que se le facilite para su entrenamiento.

Igualmente podemos observar en el tercer caso de uso, que la marca más vendida fue HYUNDAi, una marca de gama media, con cerca de 3500 registros, representa al ojo por cierto un 15% aproximadamente de las ventas totales.