

ANÁLISIS DEL MERCADO DE VEHÍCULOS USADOS

Saúl Arranz Herrero

ÍNDICE

1. Introducción
2. Hipótesis de partida
3. Análisis exploratorio
4. Análisis de variables
5. Contraste de hipótesis



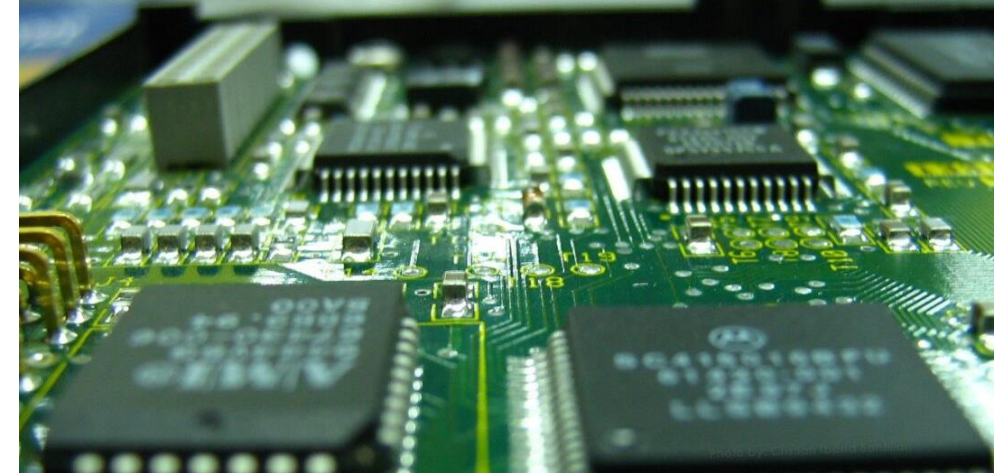
1.INTRODUCCIÓN

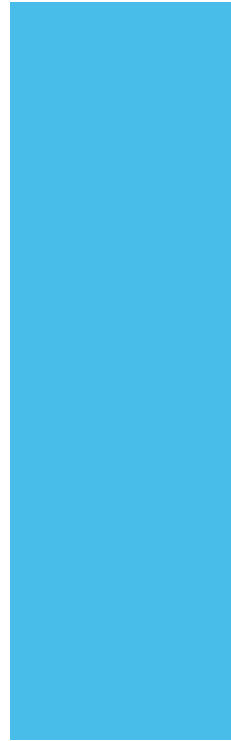
La situación económica y social actual ha provocado mucha incertidumbre en el mercado de la automoción:

- ¿Prohibición vehículos de combustión en 2030?
- Crisis de componentes → largos tiempos de espera para conseguir un coche nuevo
- Cambios en la movilidad → car sharing
- Crisis derivada de la pandemia
- Crisis guerra Ucrania

Todo esto hace que el mercado de segunda mano haya tomado mucho protagonismo en los últimos meses/años.

Por estas razones vamos a realizar un estudio sobre el precio de los vehículos de segunda mano a la venta en España





2.HIPOTESIS DE PARTIDA

Dado el contexto del que hemos hablado en la diapositiva anterior, definimos las hipótesis:

- Los vehículos con más km son los más baratos

- La marca de coches más barata es Dacia

- Los vehículos híbridos son más caros que los Diésel

- La media de precios es la misma para todas las comunidades autónomas

- Los vehículos vendidos por profesionales son más caros que los vendidos por particulares

- Los vehículos más potentes son los más caros



3. ANÁLISIS EXPLORATORIO



Los datos se han obtenido de la base de datos
<https://www.kaggle.com/datasets/datamarket/venta-de-coches>

Los datos son legibles y el dataset se carga sin problema en Visual Studio

make	model	version	price
SEAT	Toledo	SEAT Toledo 4p.	950
CITROEN	C1	CITROEN C1 PureTech 60KW 82CV Feel 5p.	6200
FORD	Transit Connect	FORD Transit Connect Van 1.5 TDCi 100cv Ambien...	7851



`df.shape = (50000, 21)`

observaciones suficientes para el análisis que queremos



No existen elementos duplicados y no hay gran cantidad de elementos nulos



El dataset consta de las siguientes variables:

- power
- doors
- shift
- url
- color
- company
- potos
- make
- model
- version
- Price
- price financed
- fuel
- year
- kms
- is professional:
- vendedor
- dealer
- province
- country
- publish date
- insert date



3. ANÁLISIS EXPLORATORIO



Creación de una nueva columna “Comunidad Autónoma”



Se eliminan las columnas que no se van a necesitar para el análisis, quedan:

- Marca
- Precio
- Combustible
- Año
- Kilómetros
- Potencia
- Vendedor profesional
- Comunidad Autónoma




Se completan los valores nulos de las columnas Marca, Modelo y Potencia



Finalmente se eliminan los outliers:

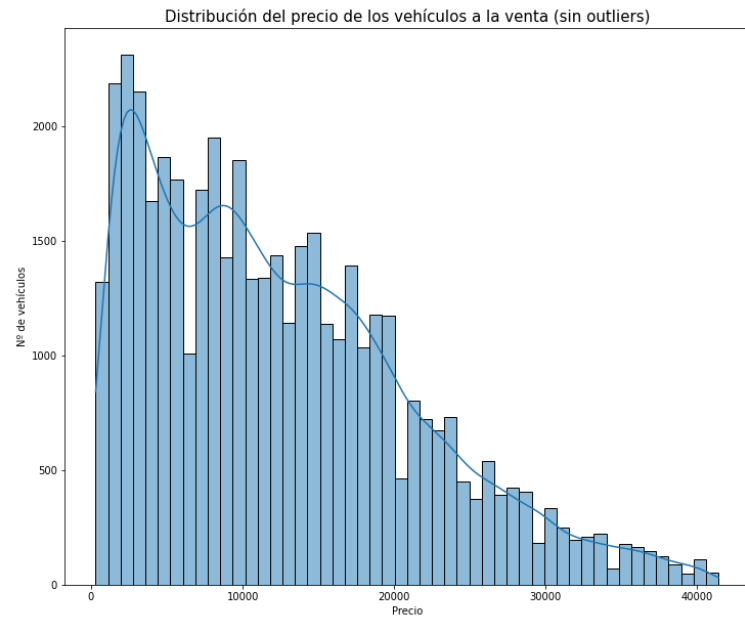
- Vehículos exclusivos
- Vehículos con muchos kms



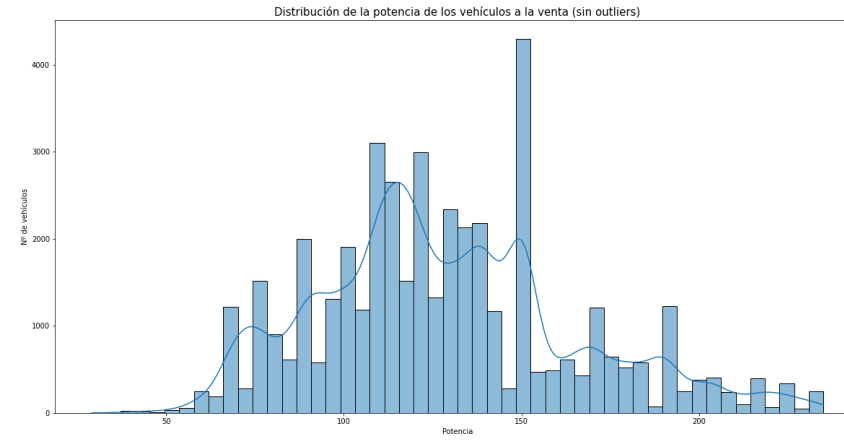
A photograph of a man in a blue jacket and a brown cap working on the engine of a vintage car. The car is dark-colored with a prominent brass or gold-colored engine cover and a large, round, brass-colored headlight. The man is leaning over the engine, and his hands are visible near the bottom. In the background, another person's legs and a red wheel are partially visible. The text "Ya tenemos el dataset preparado para poder hacer el análisis de las variables" is overlaid on the left side of the image in white, bold, sans-serif font.

**Ya tenemos el
dataset
preparado para
poder hacer el
análisis de las
variables**

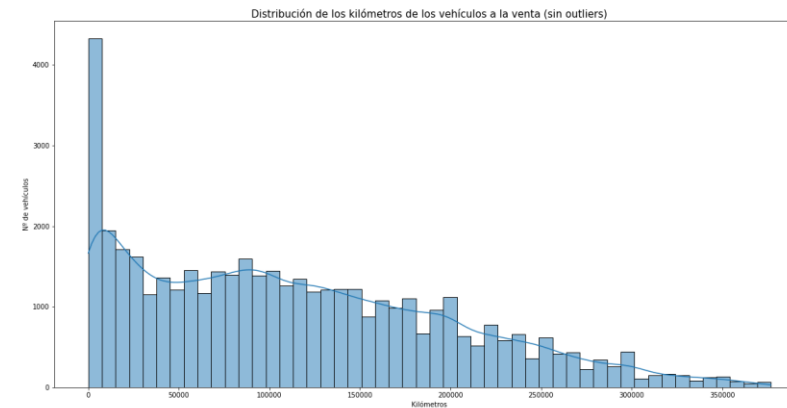
ANÁLISIS DE VARIABLES (NUMÉRICAS)



Precio

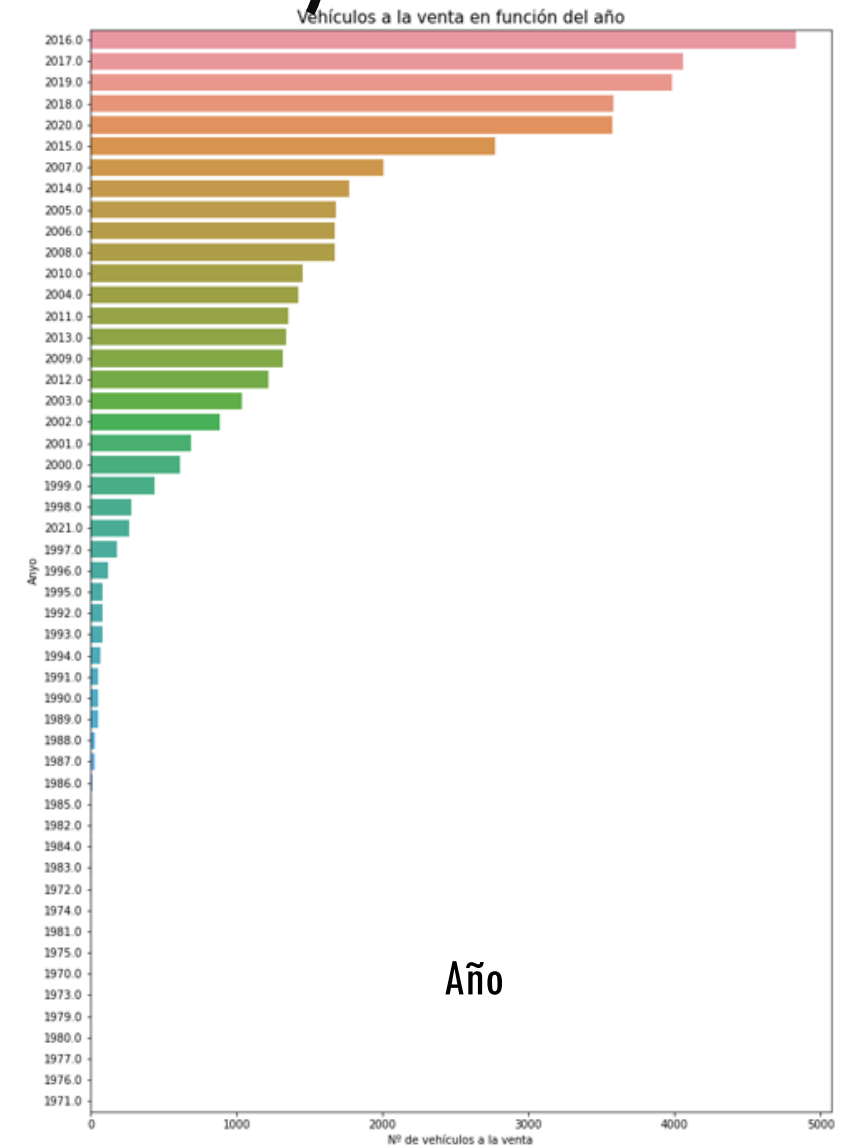
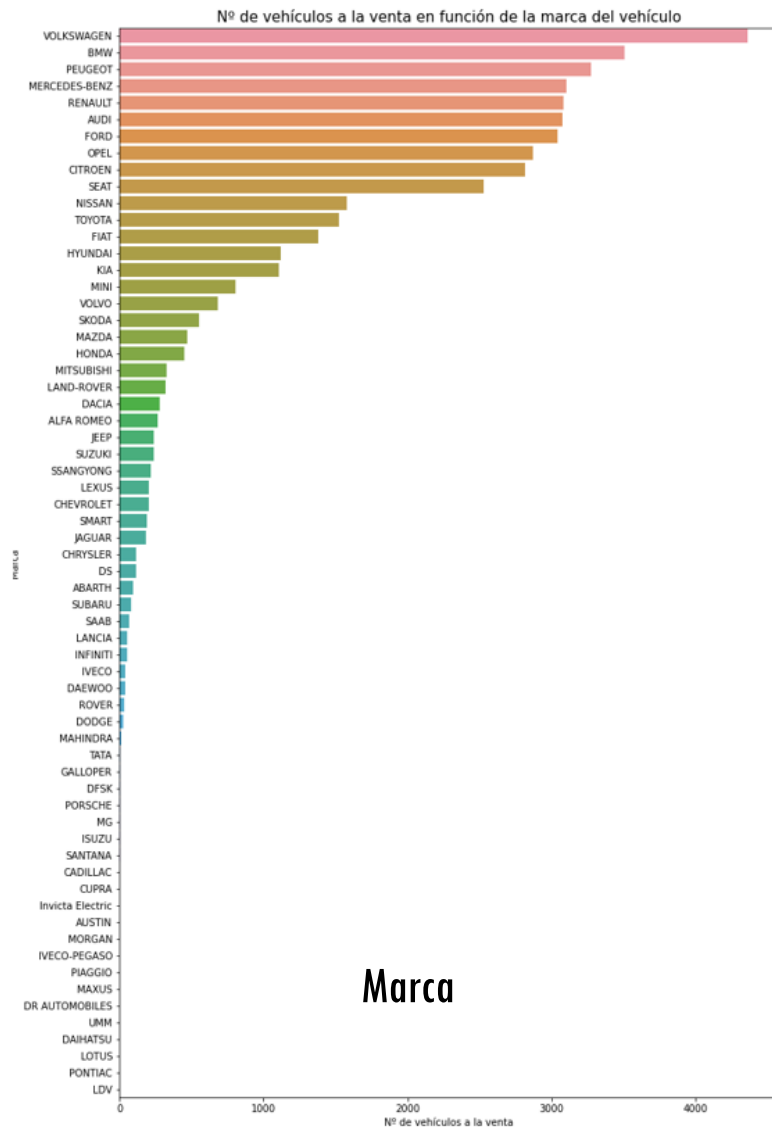


Potencia

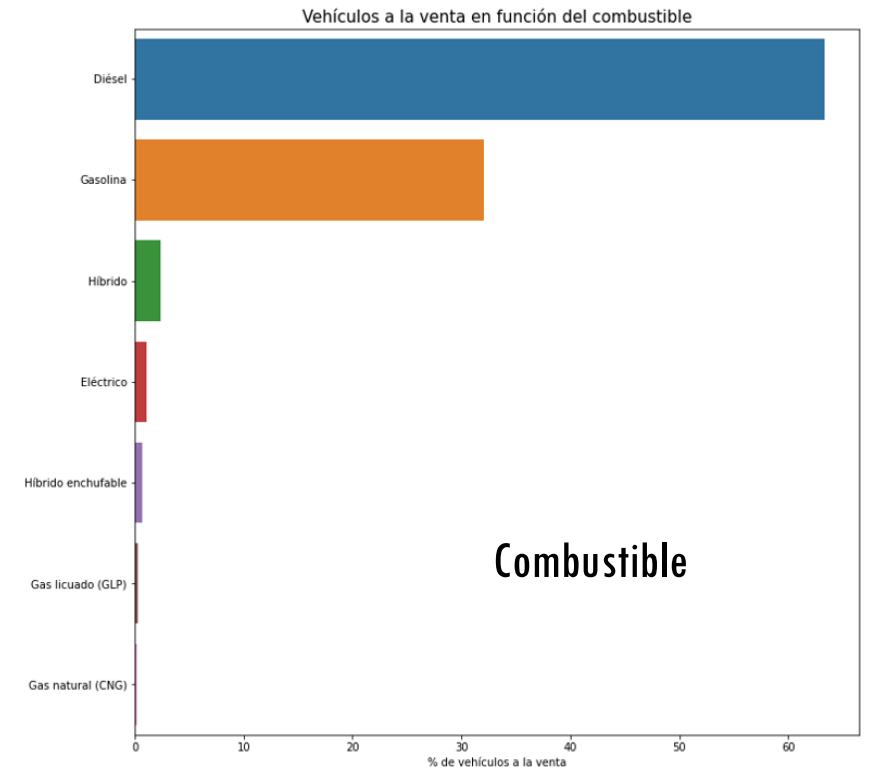
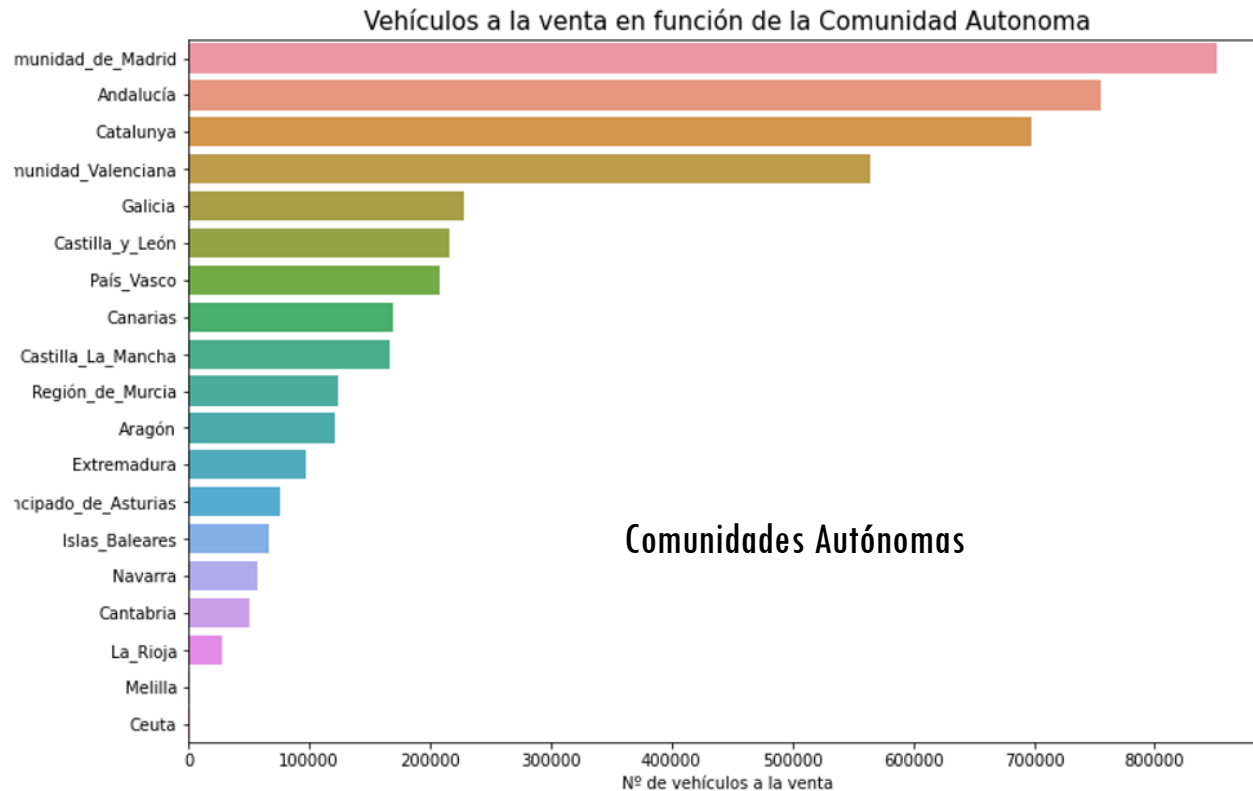


Kilómetros

ANÁLISIS DE VARIABLES (CATEGÓRICAS)

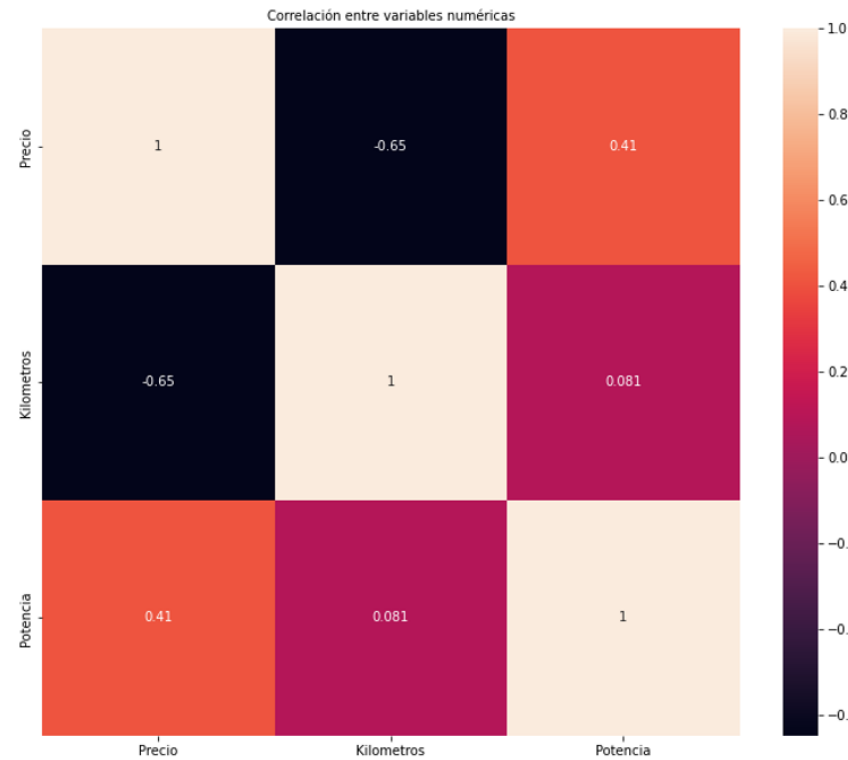
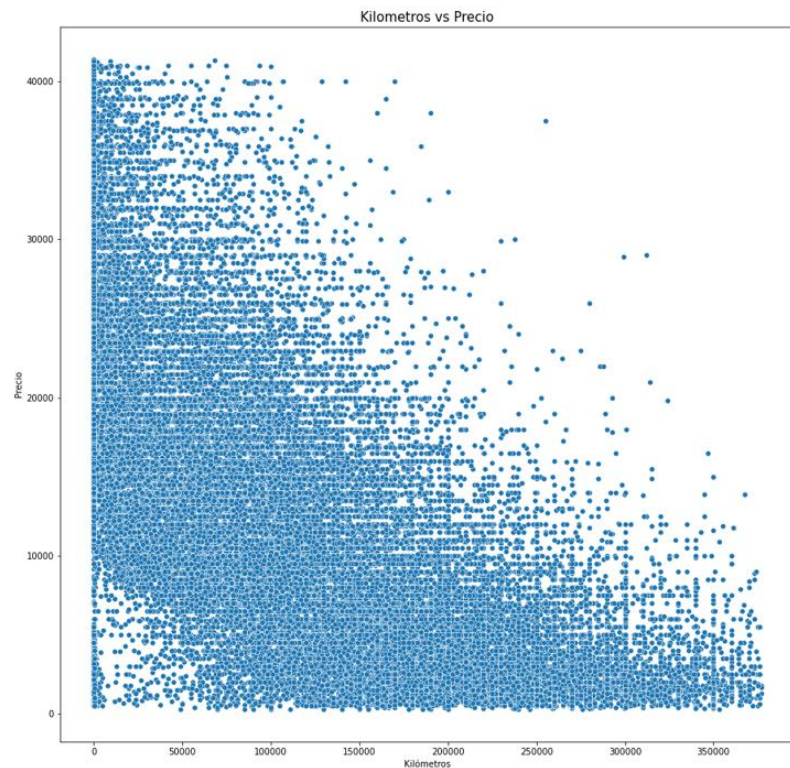


ANÁLISIS DE VARIABLES (CATEGÓRICAS)



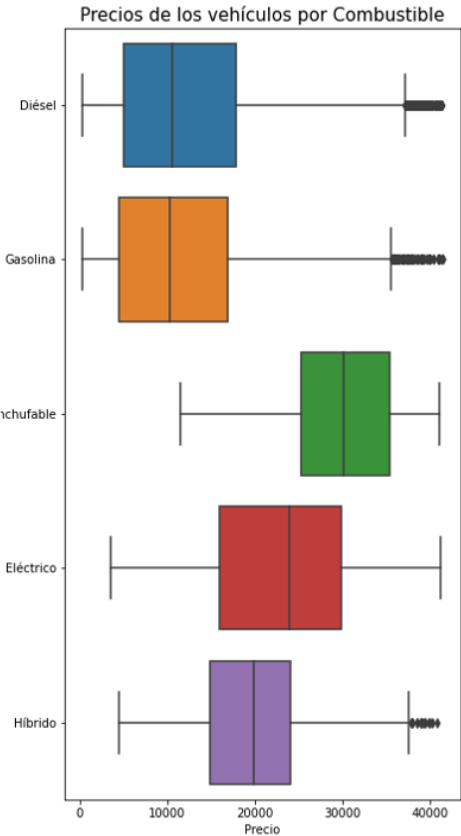
CONTRASTE DE HIPÓTESIS

Los vehículos con más km son los más baratos → confirmada



CONTRASTE DE HIPÓTESIS

Los vehículos híbridos son más caros que los Diésel → Confirmada



	Diésel	Gasolina	Híbrido	Eléctrico	Híbrido enchufable
Diésel	X	X	X	X	X
Gasolina	Distintas	X	X	X	X
Híbrido	Distintas	Distintas	X	X	X
Eléctrico	Distintas	Distintas	Distintas	X	x
Híbrido enchufable	Distintas	Distintas	Distintas	Distintas	X