

Graphic Design
is my
Passion

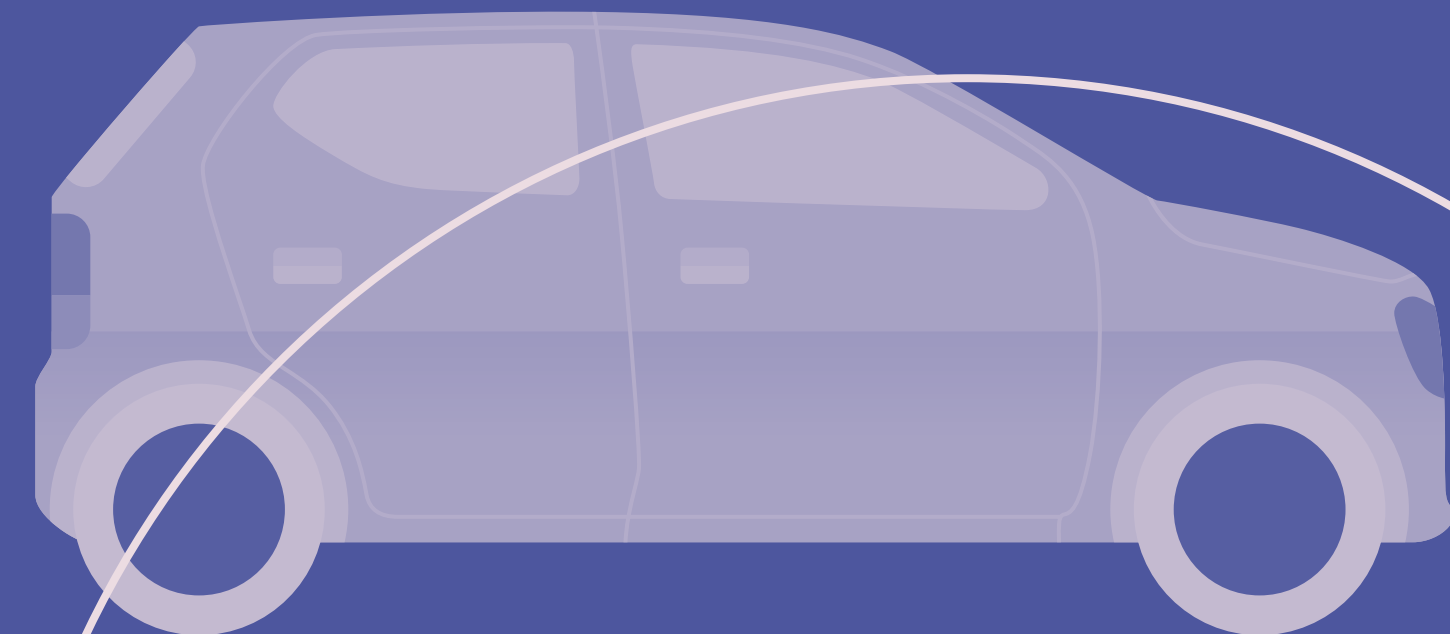
¿Cómo dar (o no dar) una presentación oral?



- Portada/Carátula
- Introducción: contexto y pregunta
- Presentación del dataset
- Análisis general (EDA)
- Análisis más detallado y específico
- Modelado
- Conclusiones
- Trabajo a Futuro

-
- Portada/Carátula** ———▶ Título acorde
Autorxs
Filiación de autorxs y/o datos de la materia
Fecha
- Introducción: contexto y pregunta** ———▶ Explicar el objetivo/pregunta: ¿por qué es interesante?, ¿qué significa?. Introducir vocabulario específico. ¿Gráficos?
- Presentación del dataset** ———▶ Presentar el/los datasets. Fuente. ¿Geografía? ¿Fecha?
Cantidad de datos. ¿Balanceado? Variables: cuáles y de qué tipo son.
Rango. Limpieza/Tratamiento de datos.
- Análisis general (EDA)** ———▶ Sirve para hacer entender mejor el dataset y primer pantallazo
Primeros pasos hacia la respuesta de nuestra pregunta
- Análisis más detallado y específico** ———▶ Gráficos más específicos y fundamentados
Camino para explicar por qué propondremos un dado modelo. Si hay tiempo y es interesante podría mostrarse un camino sin salida, o algo que esperábamos que funcione pero no lo hizo.
- Modelado** ———▶ Presentación del modelo
Métricas, Predicciones
Coeficientes si fueran de interés
¿Es lo que esperábamos?
- Conclusiones** ———▶ Resumen de lo que se hizo, conformidad con los resultados. Se repiten cosas que ya se dijeron durante el trabajo, no se muestran cosas nueva
- Trabajo a Futuro**

VEAMOS JUNTXS UN EJEMPLO CONCRETO



Una empresa automovilística china, Geely Auto, aspira a entrar en el mercado estadounidense instalando allí su unidad de fabricación y produciendo coches localmente para hacer la competencia a sus homólogos estadounidenses y europeos.

Han contratado a una empresa de consultoría automovilística para entender los factores de los que depende el precio de los autos. En concreto, quieren conocer los factores que afectan al precio de los coches en el mercado estadounidense, ya que pueden ser muy diferentes a los del mercado chino. La empresa quiere saber:

- Qué variables son significativas para predecir el precio de un coche
- En qué medida esas variables describen el precio de un coche

Basándose en varios estudios de mercado, la consultora ha reunido un amplio conjunto de datos sobre diferentes tipos de coches en el mercado estadounidense

Rows: 205

Columns: 26

```
$ car_ID      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 2...
$ symboling  <dbl> 3, 3, 1, 2, 2, 2, 1, 1, 1, 0, 2, 0, 0, 0, 1, 0, 0, 0, 2, 1, 0, 1, 1, 1, ...
$ CarName     <chr> "alfa-romero giulia", "alfa-romero stelvio", "alfa-romero Quadrifoglio",...
$ fueltype    <chr> "gas", "gas", "gas", "gas", "gas", "gas", "gas", "gas", "gas", "gas", "g...
$ aspiration  <chr> "std", "std", "std", "std", "std", "std", "std", "std", "std", "turbo", "turbo"...
$ doornumber   <chr> "two", "two", "two", "four", "four", "two", "four", "four", "four", "two...
$ carbody     <chr> "convertible", "convertible", "hatchback", "sedan", "sedan", "sedan", "s...
$ drivewheel  <chr> "rwd", "rwd", "rwd", "fwd", "4wd", "fwd", "fwd", "fwd", "fwd", "4wd", "r...
$ enginelocation <chr> "front", "front", "front", "front", "front", "front", "front", "front", "front", ...
$ wheelbase   <dbl> 88.6, 88.6, 94.5, 99.8, 99.4, 99.8, 105.8, 105.8, 105.8, 99.5, 101.2, 10...
$ carlength   <dbl> 168.8, 168.8, 171.2, 176.6, 176.6, 177.3, 192.7, 192.7, 192.7, 178.2, 17...
$ carwidth    <dbl> 64.1, 64.1, 65.5, 66.2, 66.4, 66.3, 71.4, 71.4, 71.4, 67.9, 64.8, 64.8, ...
$ carheight   <dbl> 48.8, 48.8, 52.4, 54.3, 54.3, 53.1, 55.7, 55.7, 55.9, 52.0, 54.3, 54.3, ...
$ curbweight  <dbl> 2548, 2548, 2823, 2337, 2824, 2507, 2844, 2954, 3086, 3053, 2395, 2395, ...
$ enginetype  <chr> "dohc", "dohc", "ohcv", "ohc", "ohc", "ohc", "ohc", "ohc", "ohc", "ohc", "ohc",...
$ cylindernumber <chr> "four", "four", "six", "four", "five", "five", "five", "five", "five", "five", "...
$ enginesize   <dbl> 130, 130, 152, 109, 136, 136, 136, 136, 131, 131, 108, 108, 164, 164, 16...
$ fuelsystem  <chr> "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", ...
$ boreratio   <dbl> 3.47, 3.47, 2.68, 3.19, 3.19, 3.19, 3.19, 3.19, 3.13, 3.13, 3.50, 3.50, ...
$ stroke      <dbl> 2.68, 2.68, 3.47, 3.40, 3.40, 3.40, 3.40, 3.40, 3.40, 3.40, 2.80, 2.80, ...
$ compressionratio <dbl> 9.00, 9.00, 9.00, 10.00, 8.00, 8.50, 8.50, 8.50, 8.30, 7.00, 8.80, 8.80,...
$ horsepower  <dbl> 111, 111, 154, 102, 115, 110, 110, 110, 140, 160, 101, 101, 121, 121, 12...
$ peakrpm     <dbl> 5000, 5000, 5000, 5500, 5500, 5500, 5500, 5500, 5500, 5500, 5800, 5800, ...
$ citympg     <dbl> 21, 21, 19, 24, 18, 19, 19, 19, 17, 16, 23, 23, 21, 21, 20, 16, 16, 15, ...
$ highwaympg  <dbl> 27, 27, 26, 30, 22, 25, 25, 25, 20, 22, 29, 29, 28, 28, 25, 22, 22, 20, ...
$ price       <dbl> 13495.00, 16500.00, 16500.00, 13950.00, 17450.00, 15250.00, 17710.00, 18...
```

La simbología corresponde al grado en que un auto es más arriesgado de lo que indica su precio. A los autos se les asigna inicialmente un símbolo de factor de riesgo asociado a su precio. Luego, si es más arriesgado (o menos), este símbolo se ajusta subiéndolo (o bajándolo) en la escala.

Un valor de +3 indica que el coche es arriesgado, -3 que probablemente es bastante seguro

```

Rows: 205
Columns: 26
$ car_ID      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 2...
$ symboling   <dbl> 3, 3, 1, 2, 2, 2, 1, 1, 1, 0, 2, 0, 0, 0, 1, 0, 0, 0, 2, 1, 0, 1, 1, 1, ...
$ CarName     <chr> "alfa-romero giulia", "alfa-romero stelvio", "alfa-romero Quadrifoglio",...
$ fueltype    <chr> "gas", "gas", "gas", "gas", "gas", "gas", "gas", "gas", "gas", "gas", "g...
$ aspiration   <chr> "std", "std", "std", "std", "std", "std", "std", "std", "std", "turbo", "turbo"...

```

```

#Splitting company name from CarName column
CompanyName = cars['CarName'].apply(lambda x : x.split(' ')[0])
cars.insert(3, "CompanyName", CompanyName)
cars.drop(['CarName'], axis=1, inplace=True)
cars.head()

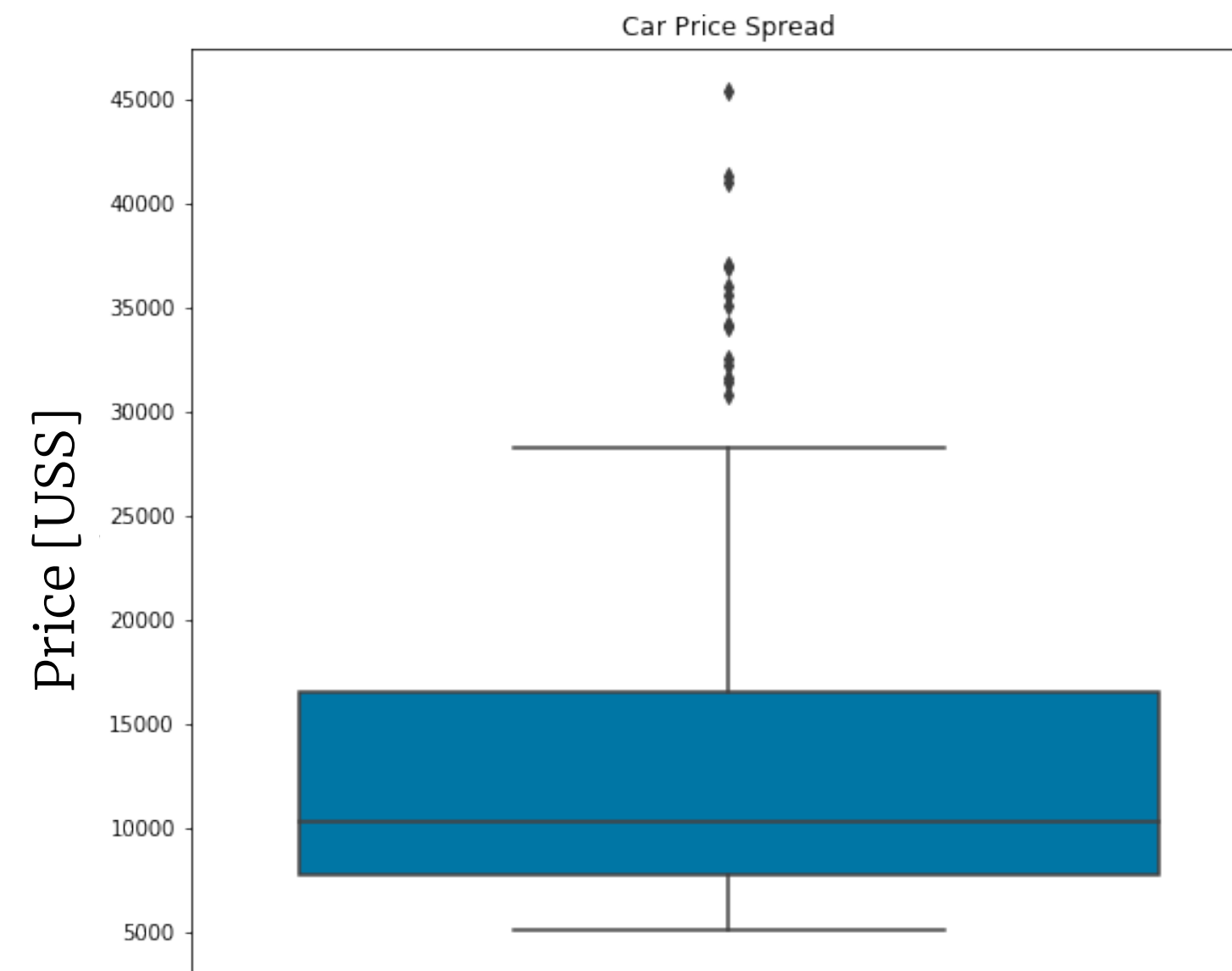
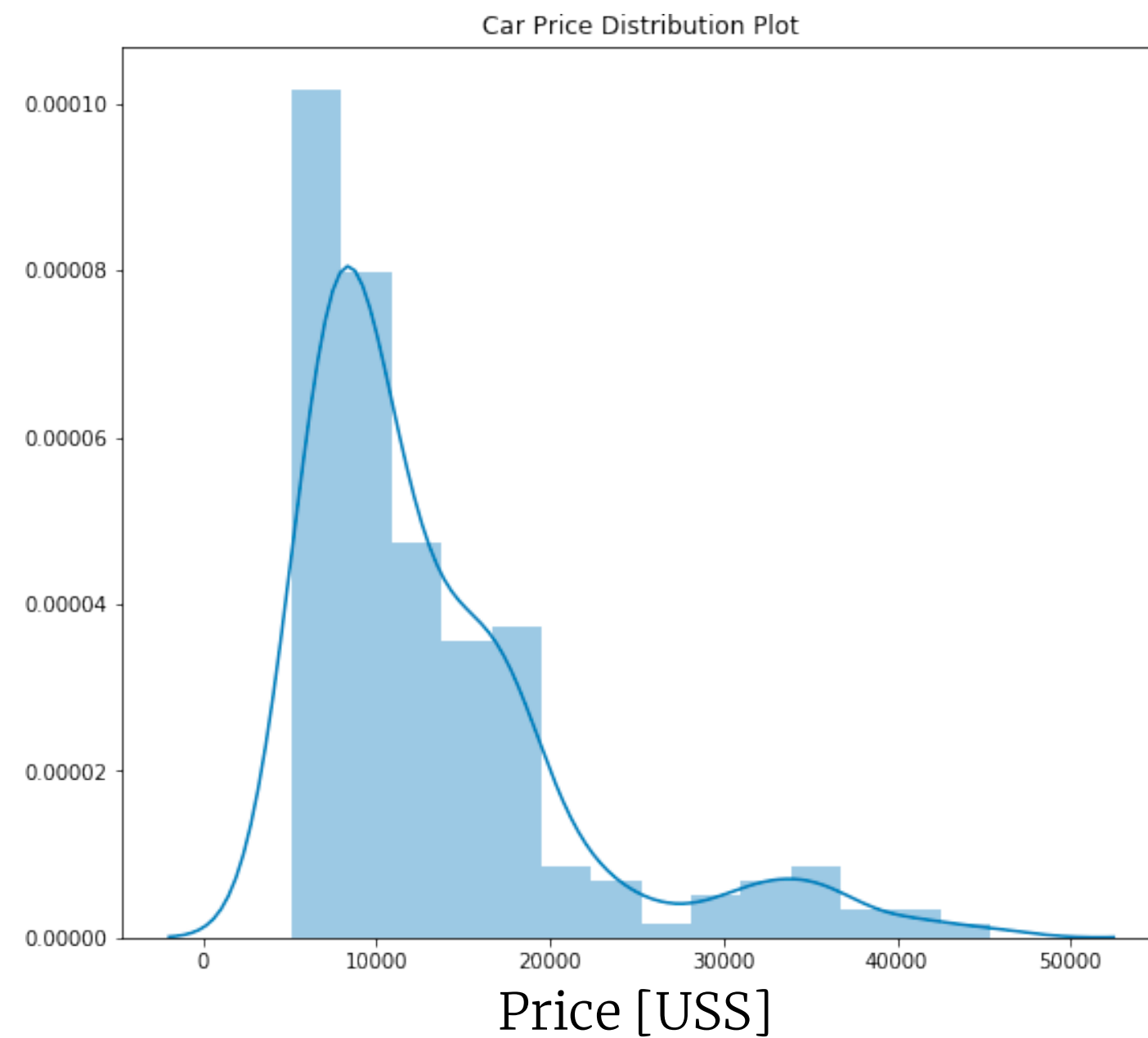
```

```
cars.CompanyName.unique()
```

```

array(['alfa-romero', 'audi', 'bmw', 'chevrolet', 'dodge', 'honda',
       'isuzu', 'jaguar', 'maxda', 'mazda', 'buick', 'mercury',
       'mitsubishi', 'Nissan', 'nissan', 'peugeot', 'plymouth', 'porsche',
       'porcshce', 'renault', 'saab', 'subaru', 'toyota', 'toyouta',
       'vokswagen', 'volkswagen', 'vw', 'volvo'], dtype=object)

```

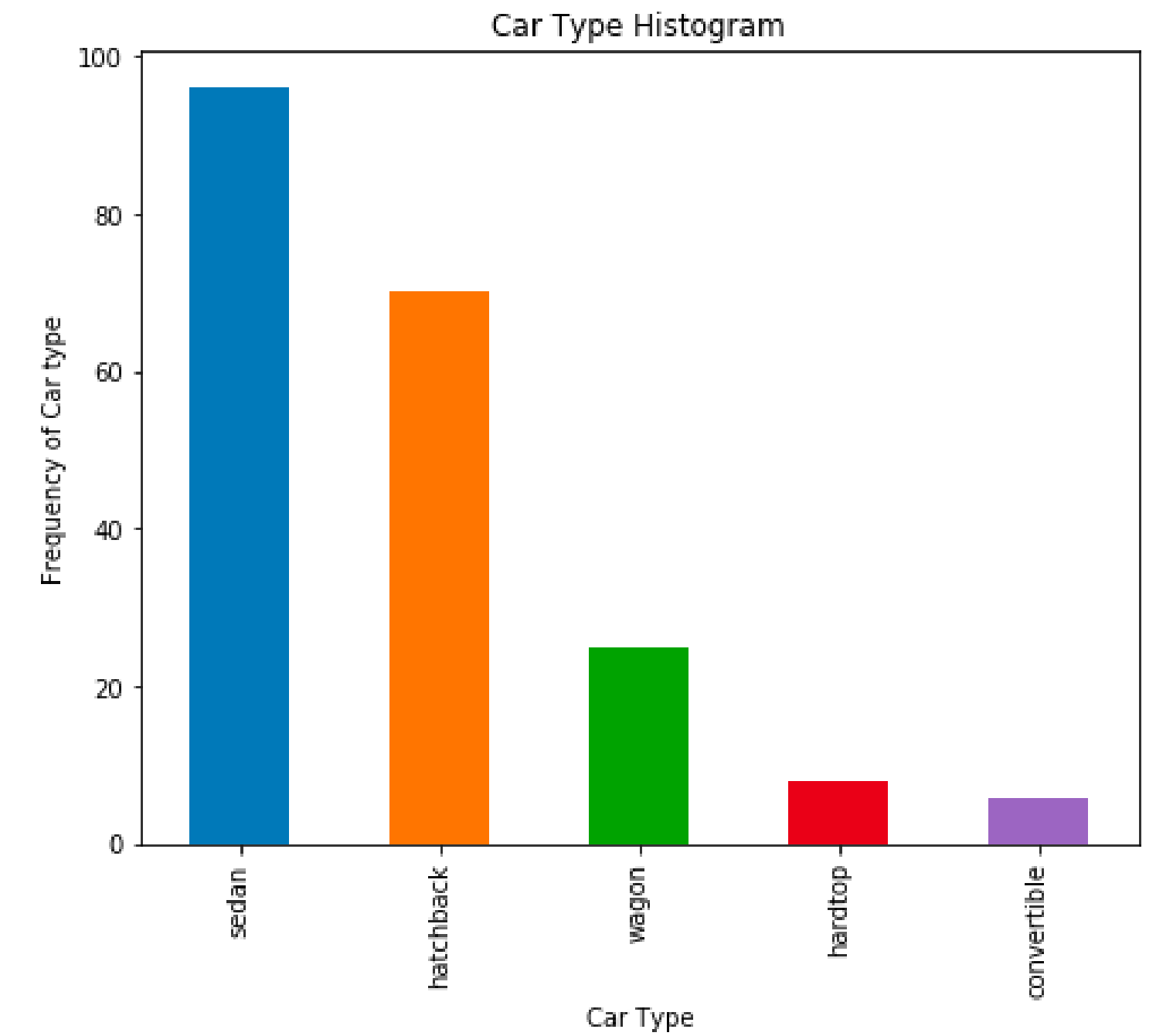
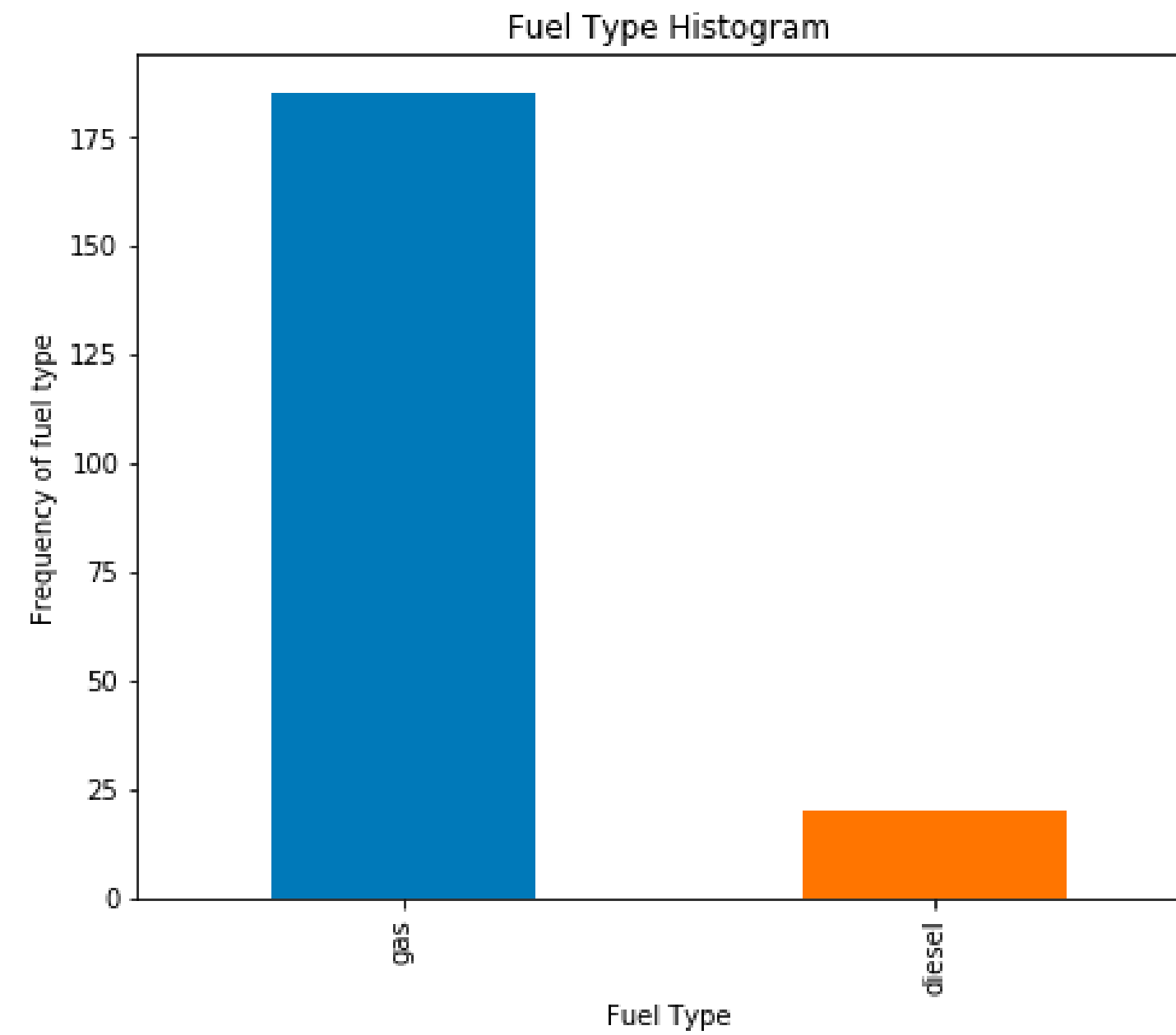
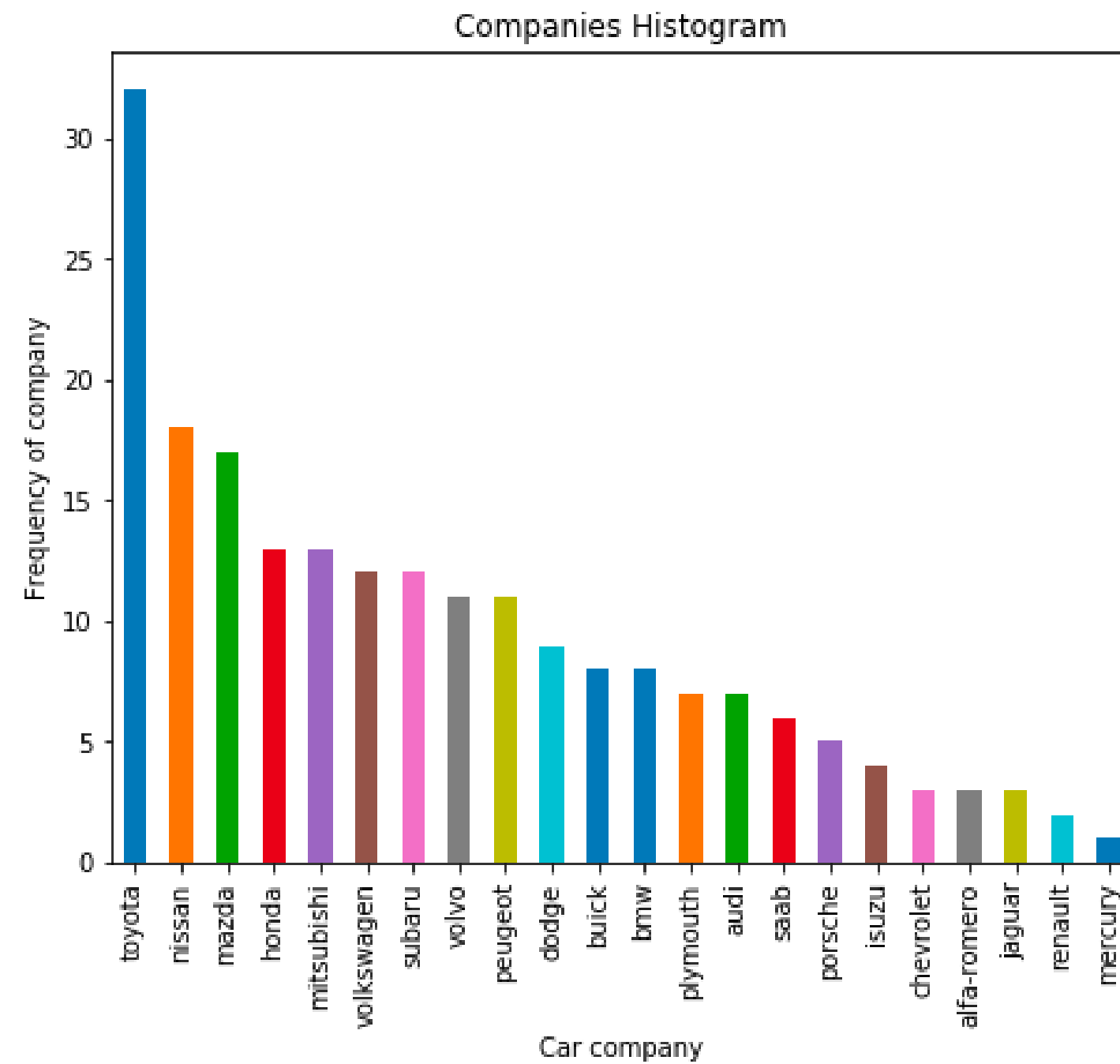



count	205.000000
mean	13276.710571
std	7988.852332
min	5118.000000
25%	7788.000000
50%	10295.000000
75%	16503.000000
85%	18500.000000
90%	22563.000000
100%	45400.000000
max	45400.000000

El gráfico parece estar sesgado a la derecha, lo que significa que la mayoría de los precios del conjunto de datos son bajos (por debajo de 15.000).

Hay una diferencia significativa entre la media y la mediana de la distribución de precios.

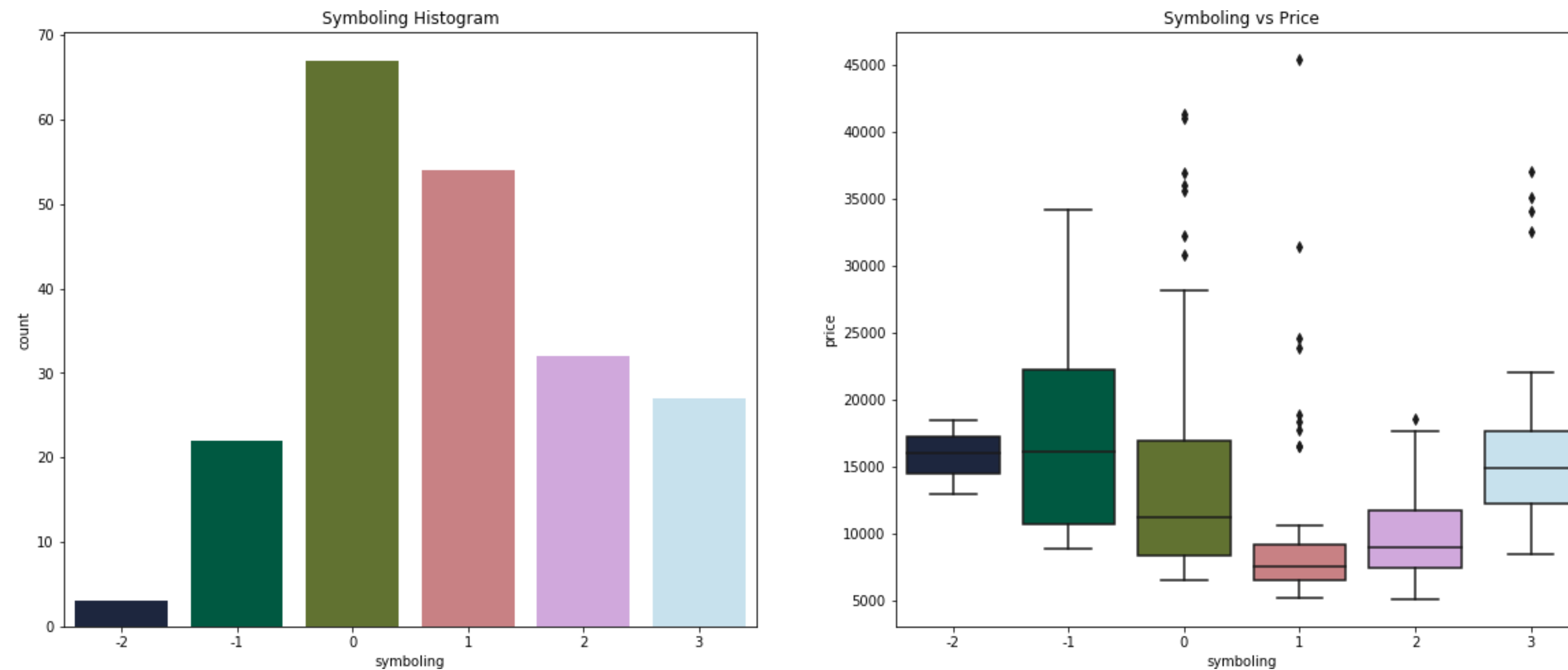
Los puntos de datos están muy alejados de la media, lo que indica una gran varianza en los precios de los coches (el 85% de los precios están por debajo de 18.500, mientras que el 15% restante está entre 18.500 y 45.400).



Toyota parece ser la empresa de automóviles favorita.

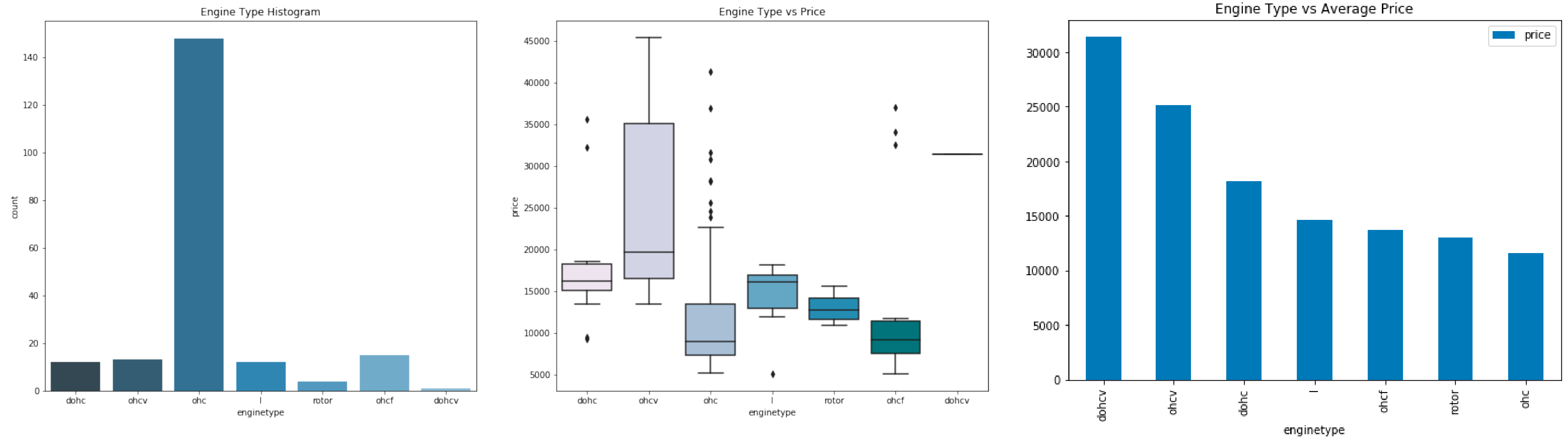
El número de coches de gasolina es mayor que el de diesel.

El sedán es el tipo de coche preferido.



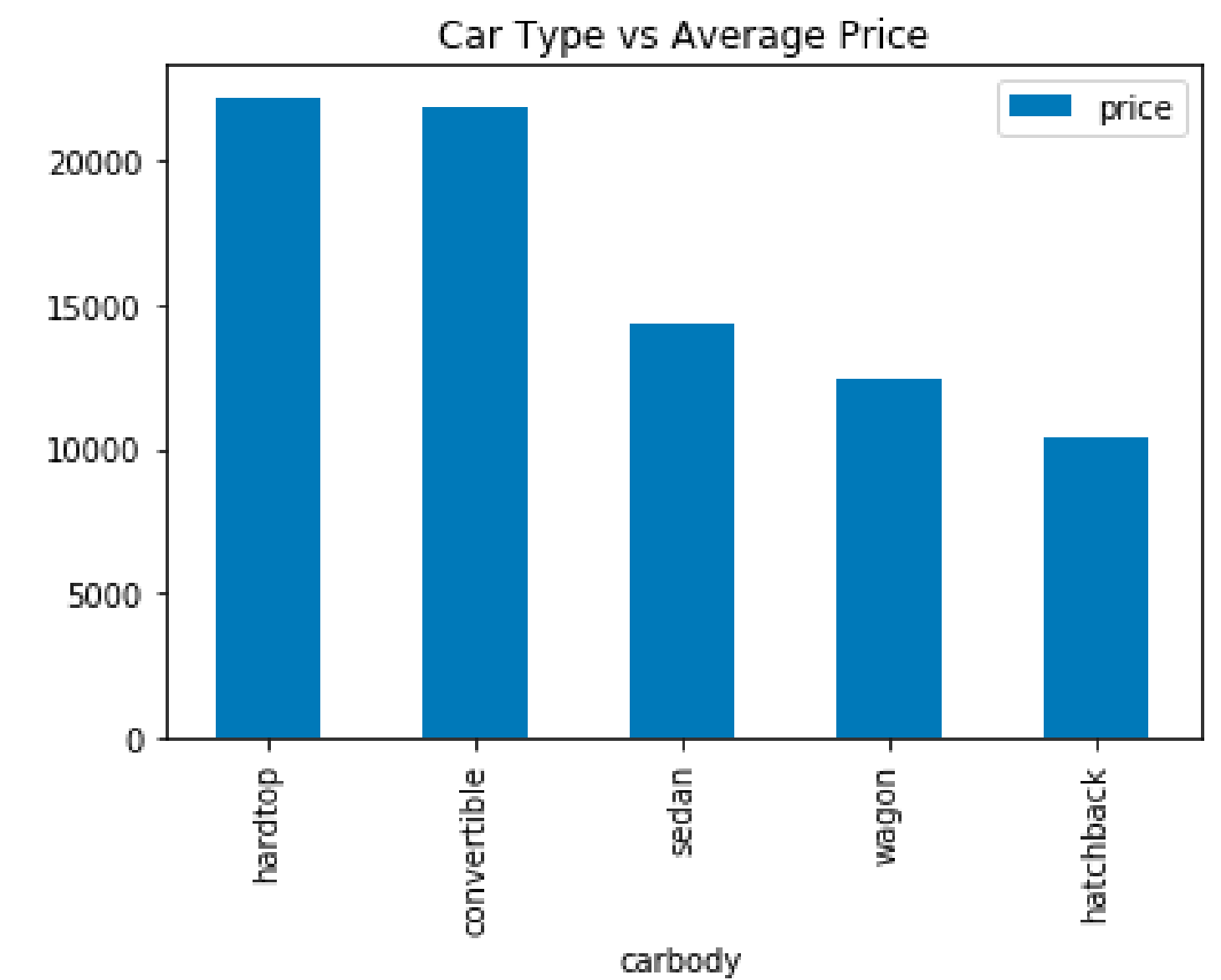
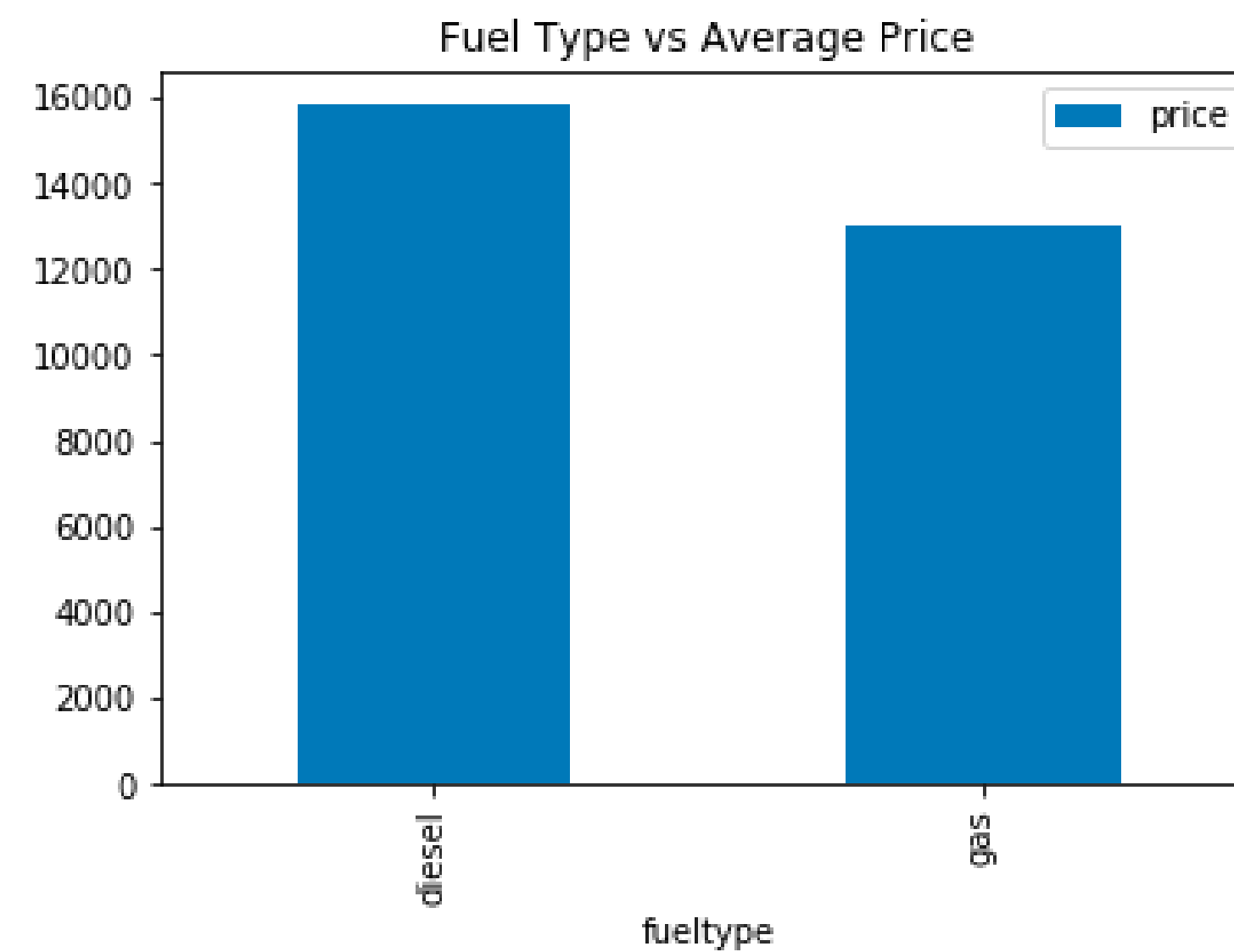
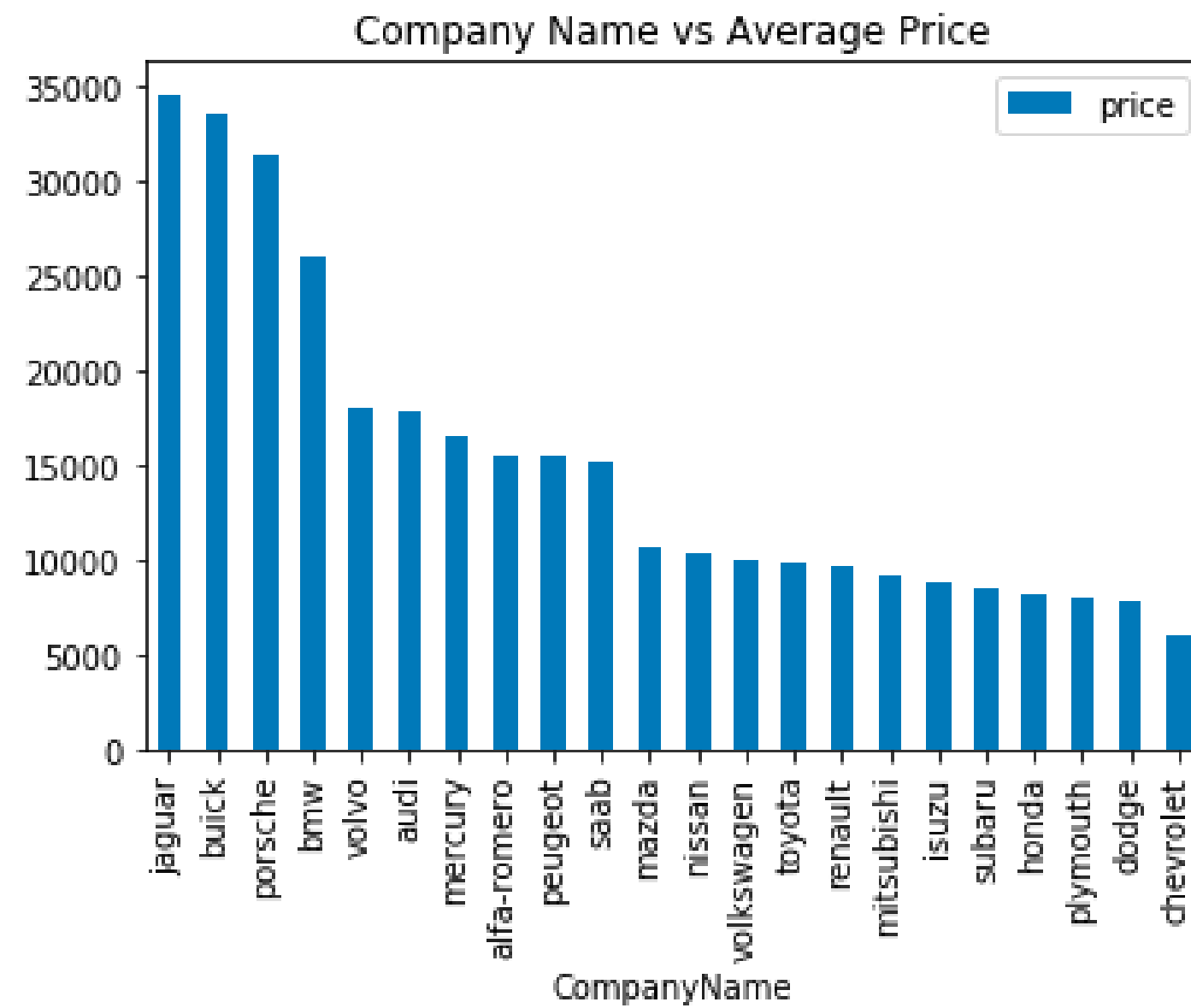
Parece que las simbologías con valores 0 y 1 tienen un alto número de filas (es decir, son las más vendidas).

Los coches con simbología -1 parecen tener un precio elevado (tiene sentido ya que la calificación de riesgo del seguro -1 es bastante buena). Pero parece que la simbología con valor 3 tiene un rango de precios similar al valor -2. La simbología 1 presenta un descenso de precios.



El tipo de motor ohc parece ser el más favorecido.

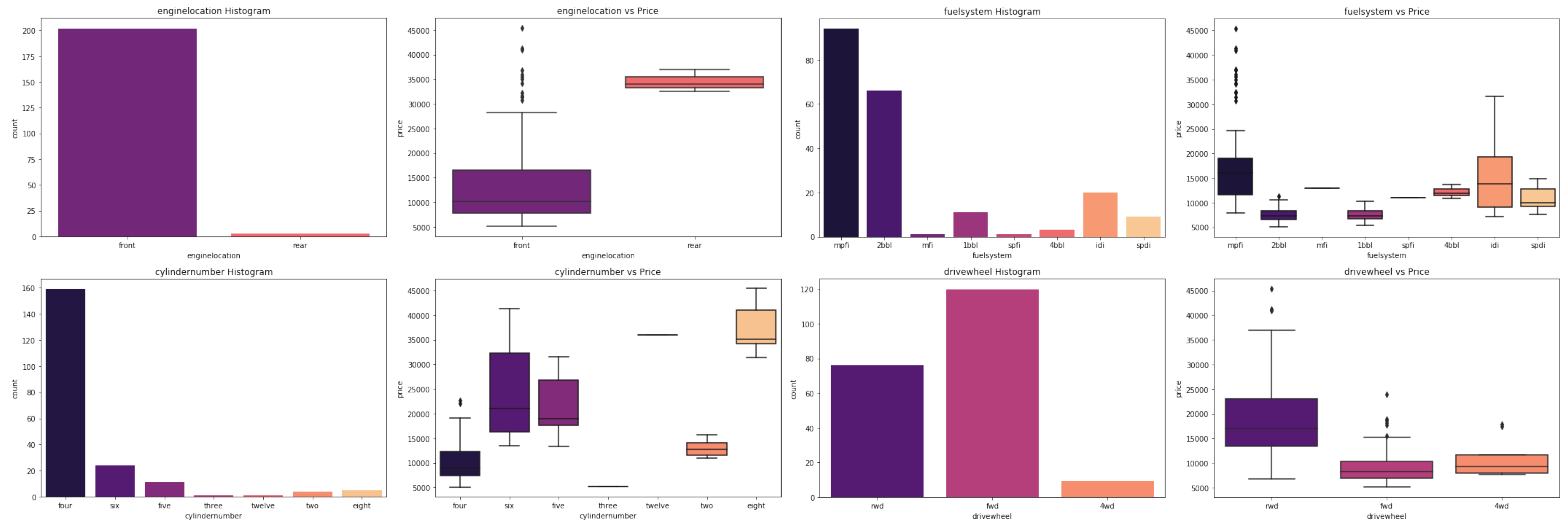
El ohcv tiene la gama de precios más alta (mientras que el dohcv sólo tiene una fila), el ohc y el ohcf tienen la gama de precios más baja.



Jaguar y Buick parecen tener el precio medio más alto.

El diesel tiene un precio medio más alto que el de la gasolina.

El techo duro y el descapotable tienen un precio medio más alto.

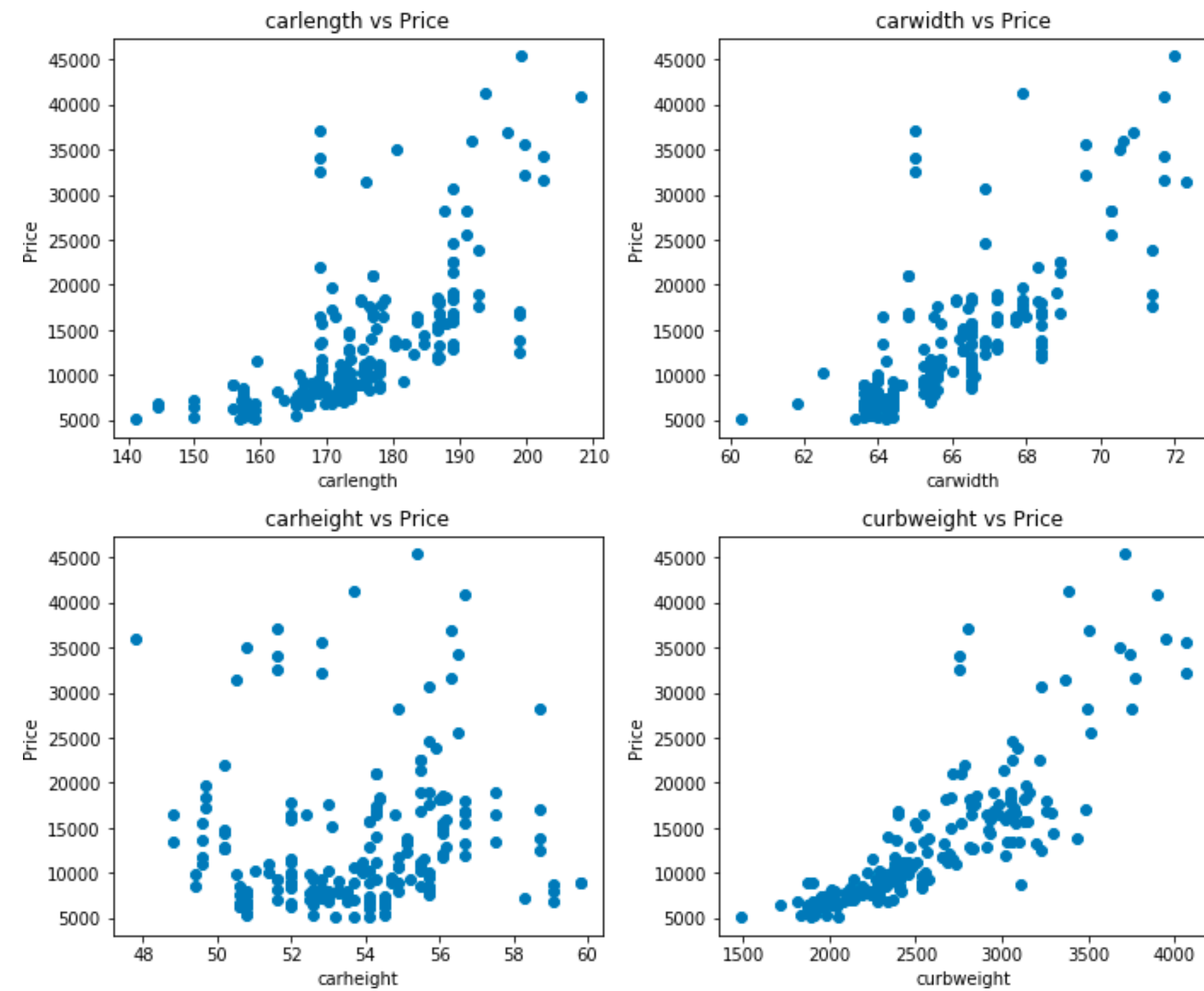


Hay muy pocos datos de categorías de localización de motores para hacer una inferencia.

Los números más comunes de cilindros son cuatro, seis y cinco. Aunque los de ocho cilindros son los de mayor precio.

Los sistemas de combustible más comunes son el mpfi y el 2bbl. El mpfi y el idi tienen el rango de precios más alto. Pero hay pocos datos para otras categorías para derivar cualquier inferencia significativa

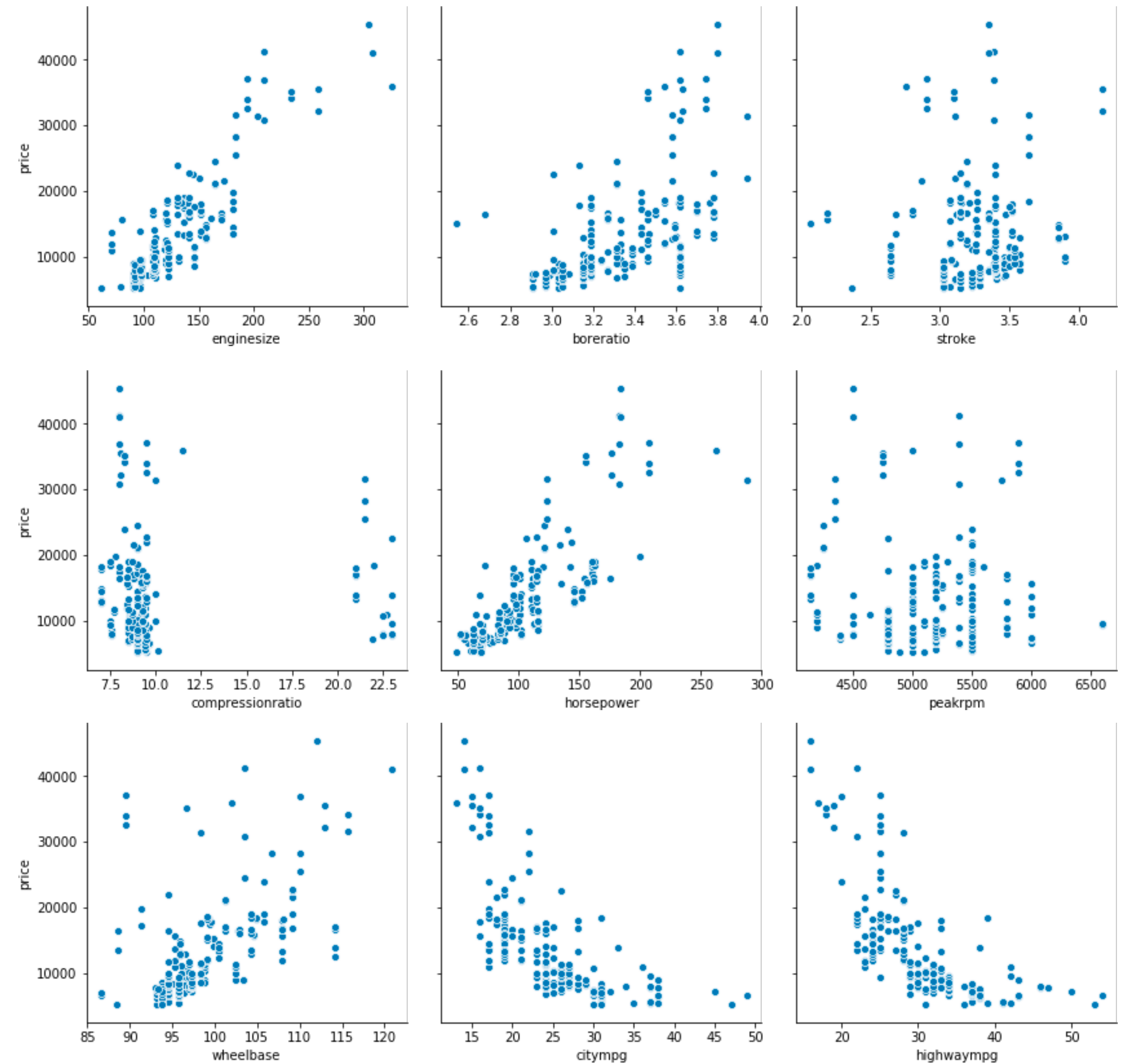
Hay una diferencia muy significativa en la categoría de ruedas motrices. La mayoría de los coches de gama alta parecen preferir la rueda motriz rwd.



carwidth - carlength - curbweight - enginesize - boreratio - borehorsepower - wheelbase parece tener una correlación poitiva con el precio.

carheight no muestra ninguna tendencia significativa con el precio.

citympg - highwaympg parecen tener una correlación negativa significativa con el precio.



List of significant variables after Visual analysis :

- Car Range
- Engine Type
- Fuel type
- Car Body
- Aspiration
- Cylinder Number
- Drivewheel
- Curbweight
- Car Length
- Car width
- Engine Size
- Boreratio
- Horse Power
- Wheel base
- Fuel Economy

Para realizar los modelados fue necesario estandarizar las variables

```
from sklearn.preprocessing import MinMaxScaler

scaler = MinMaxScaler()
num_vars = ['wheelbase', 'curbweight', 'enginesize', 'boreratio', 'horsepower', 'fueleconomy', 'carlength', 'carwidth', 'price']
df_train[num_vars] = scaler.fit_transform(df_train[num_vars])
```


MODELO 1

=====						
Dep. Variable:	price	R-squared:	0.927			
Model:	OLS	Adj. R-squared:	0.922			
Method:	Least Squares	F-statistic:	187.9			
Date:	Fri, 26 Apr 2019	Prob (F-statistic):	4.25e-71			
Time:	09:01:41	Log-Likelihood:	204.17			
No. Observations:	143	AIC:	-388.3			
Df Residuals:	133	BIC:	-358.7			
Df Model:	9					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-0.0764	0.041	-1.851	0.066	-0.158	0.005
curbweight	0.2756	0.069	3.995	0.000	0.139	0.412
horsepower	0.3997	0.069	5.824	0.000	0.264	0.535
fuel economy	0.0736	0.051	1.435	0.154	-0.028	0.175
carwidth	0.2580	0.062	4.137	0.000	0.135	0.381
hatchback	-0.0951	0.025	-3.766	0.000	-0.145	-0.045
sedan	-0.0744	0.025	-2.983	0.003	-0.124	-0.025
wagon	-0.1050	0.028	-3.744	0.000	-0.160	-0.050
dohcv	-0.2319	0.077	-3.015	0.003	-0.384	-0.080
Highend	0.2565	0.020	12.743	0.000	0.217	0.296
=====						

MODELO 2

=====						
Dep. Variable:	price	R-squared:	0.926			
Model:	OLS	Adj. R-squared:	0.922			
Method:	Least Squares	F-statistic:	209.5			
Date:	Fri, 26 Apr 2019	Prob (F-statistic):	7.85e-72			
Time:	09:01:42	Log-Likelihood:	203.07			
No. Observations:	143	AIC:	-388.1			
Df Residuals:	134	BIC:	-361.5			
Df Model:	8					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-0.0305	0.026	-1.165	0.246	-0.082	0.021
curbweight	0.2593	0.068	3.796	0.000	0.124	0.394
horsepower	0.3469	0.058	5.964	0.000	0.232	0.462
carwidth	0.2488	0.062	3.995	0.000	0.126	0.372
hatchback	-0.0922	0.025	-3.650	0.000	-0.142	-0.042
sedan	-0.0711	0.025	-2.850	0.005	-0.120	-0.022
wagon	-0.1047	0.028	-3.721	0.000	-0.160	-0.049
dohcv	-0.1968	0.073	-2.689	0.008	-0.342	-0.052
Highend	0.2610	0.020	13.083	0.000	0.222	0.301
=====						

MODELO 3

```

=====
Dep. Variable:          price    R-squared:          0.918
Model:                  OLS      Adj. R-squared:      0.914
Method:                 Least Squares    F-statistic:      215.9
Date:                  Fri, 26 Apr 2019    Prob (F-statistic): 4.70e-70
Time:                  09:01:42    Log-Likelihood:    195.77
No. Observations:      143    AIC:                -375.5
Df Residuals:          135    BIC:                -351.8
Df Model:               7
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	-0.0319	0.027	-1.161	0.248	-0.086	0.022
horsepower	0.4690	0.051	9.228	0.000	0.368	0.569
carwidth	0.4269	0.043	9.944	0.000	0.342	0.512
hatchback	-0.1044	0.026	-3.976	0.000	-0.156	-0.052
sedan	-0.0756	0.026	-2.896	0.004	-0.127	-0.024
wagon	-0.0865	0.029	-2.974	0.003	-0.144	-0.029
dohcv	-0.3106	0.070	-4.435	0.000	-0.449	-0.172
Highend	0.2772	0.020	13.559	0.000	0.237	0.318

```

=====

```

MODELO 4

```

=====
Dep. Variable:          price    R-squared:          0.913
Model:                  OLS      Adj. R-squared:      0.909
Method:                 Least Squares    F-statistic:      237.6
Date:                  Fri, 26 Apr 2019    Prob (F-statistic): 1.68e-69
Time:                  09:01:42    Log-Likelihood:    191.46
No. Observations:      143    AIC:                -368.9
Df Residuals:          136    BIC:                -348.2
Df Model:               6
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	-0.0934	0.018	-5.219	0.000	-0.129	-0.058
horsepower	0.5001	0.051	9.805	0.000	0.399	0.601
carwidth	0.3963	0.043	9.275	0.000	0.312	0.481
hatchback	-0.0373	0.013	-2.938	0.004	-0.062	-0.012
wagon	-0.0170	0.017	-1.008	0.315	-0.050	0.016
dohcv	-0.3203	0.072	-4.460	0.000	-0.462	-0.178
Highend	0.2808	0.021	13.402	0.000	0.239	0.322

```

=====

```

¿CÓMO PRESENTAMOS ESTE TRABAJO?

**COSAS QUE NO
COSAS QUE SÍ**

TP FINAL

Introducción a la Ciencia de Datos

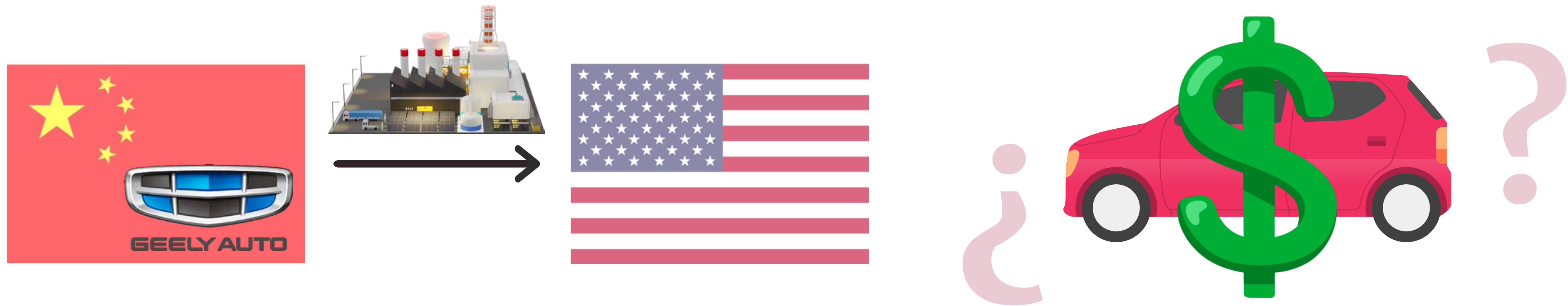
UNSAM, 2022

“¿Cuánto me va a salir el auto?”

Presentan: **Juan Gutierrez, Josefina Martinez, Arturo Vidal**



- Una empresa automovilística china, Geely Auto, aspira a entrar en el mercado estadounidense y producir coches localmente para competir en EE.UU. y Europa
- Nos contratan como consultoría automovilística para entender los factores de los que depende el precio de los autos.
- Quieren conocer los factores que afectan al precio de los coches en el mercado estadounidense, ya que pueden ser muy diferentes a los del mercado chino.
- Qué variables son significativas para predecir el precio de un coche
En qué medida esas variables describen el precio de un coche
- Basándose en varios estudios de mercado, la consultora ha reunido un amplio conjunto de datos sobre diferentes tipos de coches en el mercado estadounidense



¿Qué variables son significativas para predecir el precio de un auto?
En qué medida esas variables describen el precio del auto

Rows: 205

Columns: 26

```
$ car_ID      <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 2...
$ symboling  <dbl> 3, 3, 1, 2, 2, 2, 1, 1, 1, 0, 2, 0, 0, 0, 1, 0, 0, 0, 2, 1, 0, 1, 1, 1, ...
$ CarName    <chr> "alfa-romero giulia", "alfa-romero stelvio", "alfa-romero Quadrifoglio",...
$ fueltype   <chr> "gas", "gas", "gas", "gas", "gas", "gas", "gas", "gas", "gas", "gas", "g...
$ aspiration <chr> "std", "std", "std", "std", "std", "std", "std", "std", "std", "turbo", "turbo"...
$ doornumber <chr> "two", "two", "two", "four", "four", "two", "four", "four", "four", "two...
$ carbody    <chr> "convertible", "convertible", "hatchback", "sedan", "sedan", "sedan", "s...
$ drivewheel <chr> "rwd", "rwd", "rwd", "fwd", "4wd", "fwd", "fwd", "fwd", "fwd", "4wd", "r...
$ enginelocation <chr> "front", "front", "front", "front", "front", "front", "front", "front", "front", ...
$ wheelbase  <dbl> 88.6, 88.6, 94.5, 99.8, 99.4, 99.8, 105.8, 105.8, 105.8, 99.5, 101.2, 10...
$ carlength  <dbl> 168.8, 168.8, 171.2, 176.6, 176.6, 177.3, 192.7, 192.7, 192.7, 178.2, 17...
$ carwidth   <dbl> 64.1, 64.1, 65.5, 66.2, 66.4, 66.3, 71.4, 71.4, 71.4, 67.9, 64.8, 64.8, ...
$ carheight  <dbl> 48.8, 48.8, 52.4, 54.3, 54.3, 53.1, 55.7, 55.7, 55.9, 52.0, 54.3, 54.3, ...
$ curbweight <dbl> 2548, 2548, 2823, 2337, 2824, 2507, 2844, 2954, 3086, 3053, 2395, 2395, ...
$ enginetype <chr> "dohc", "dohc", "ohcv", "ohc", "ohc", "ohc", "ohc", "ohc", "ohc", "ohc", "ohc",...
$ cylindernumber <chr> "four", "four", "six", "four", "five", "five", "five", "five", "five", "five", ...
$ enginesize  <dbl> 130, 130, 152, 109, 136, 136, 136, 136, 131, 131, 108, 108, 164, 164, 16...
$ fuelsystem <chr> "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", "mpfi", ...
$ boreratio  <dbl> 3.47, 3.47, 2.68, 3.19, 3.19, 3.19, 3.19, 3.19, 3.13, 3.13, 3.50, 3.50, ...
$ stroke      <dbl> 2.68, 2.68, 3.47, 3.40, 3.40, 3.40, 3.40, 3.40, 3.40, 3.40, 2.80, 2.80, ...
$ compressionratio <dbl> 9.00, 9.00, 9.00, 10.00, 8.00, 8.50, 8.50, 8.50, 8.30, 7.00, 8.80, 8.80,...
$ horsepower <dbl> 111, 111, 154, 102, 115, 110, 110, 110, 140, 160, 101, 101, 121, 121, 12...
$ peakrpm    <dbl> 5000, 5000, 5000, 5500, 5500, 5500, 5500, 5500, 5500, 5500, 5800, 5800, ...
$ citympg    <dbl> 21, 21, 19, 24, 18, 19, 19, 19, 17, 16, 23, 23, 21, 21, 20, 16, 16, 15, ...
$ highwaympg <dbl> 27, 27, 26, 30, 22, 25, 25, 25, 20, 22, 29, 29, 28, 28, 25, 22, 22, 20, ...
$ price      <dbl> 13495.00, 16500.00, 16500.00, 13950.00, 17450.00, 15250.00, 17710.00, 18...
```


26 variables

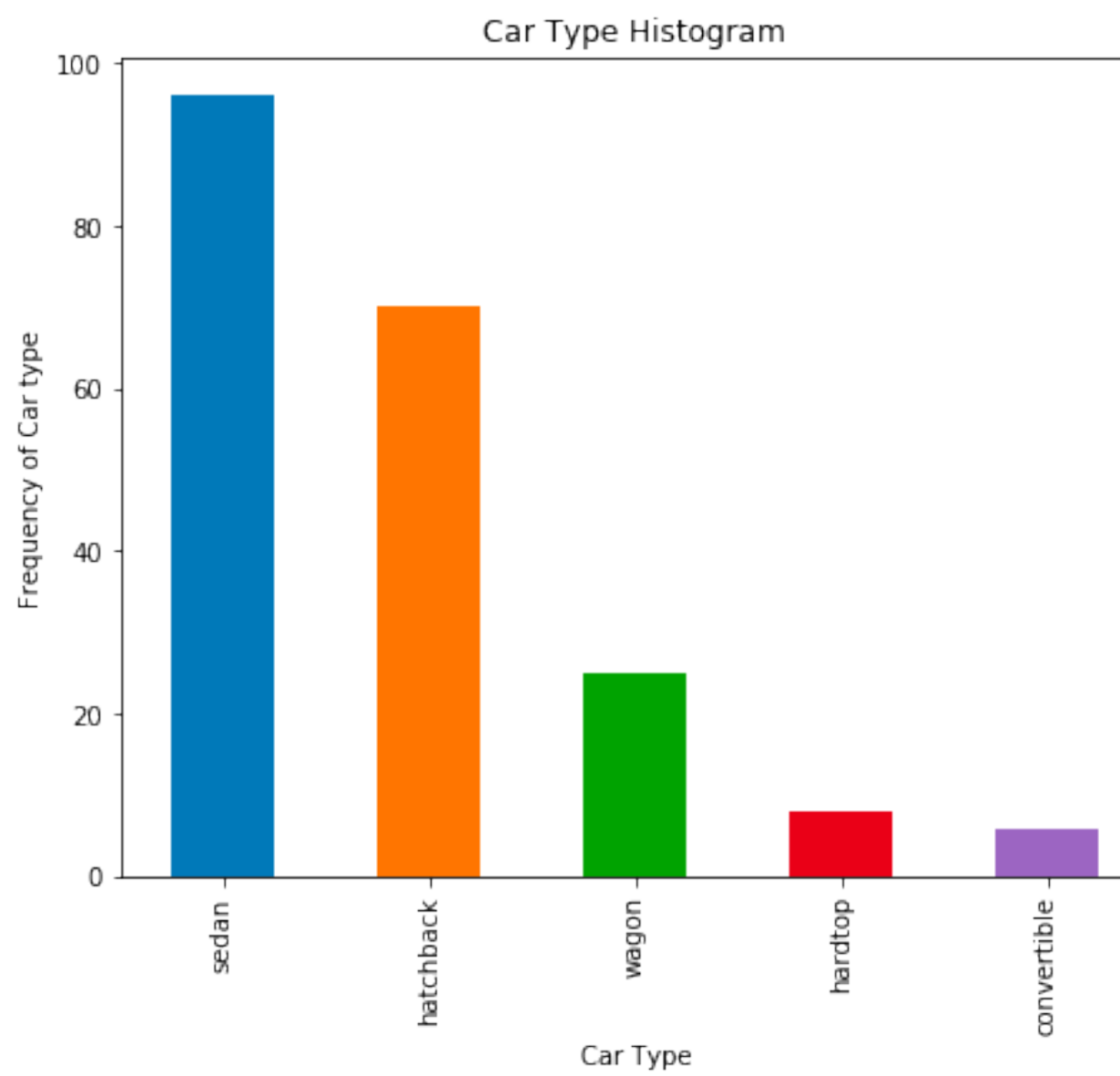
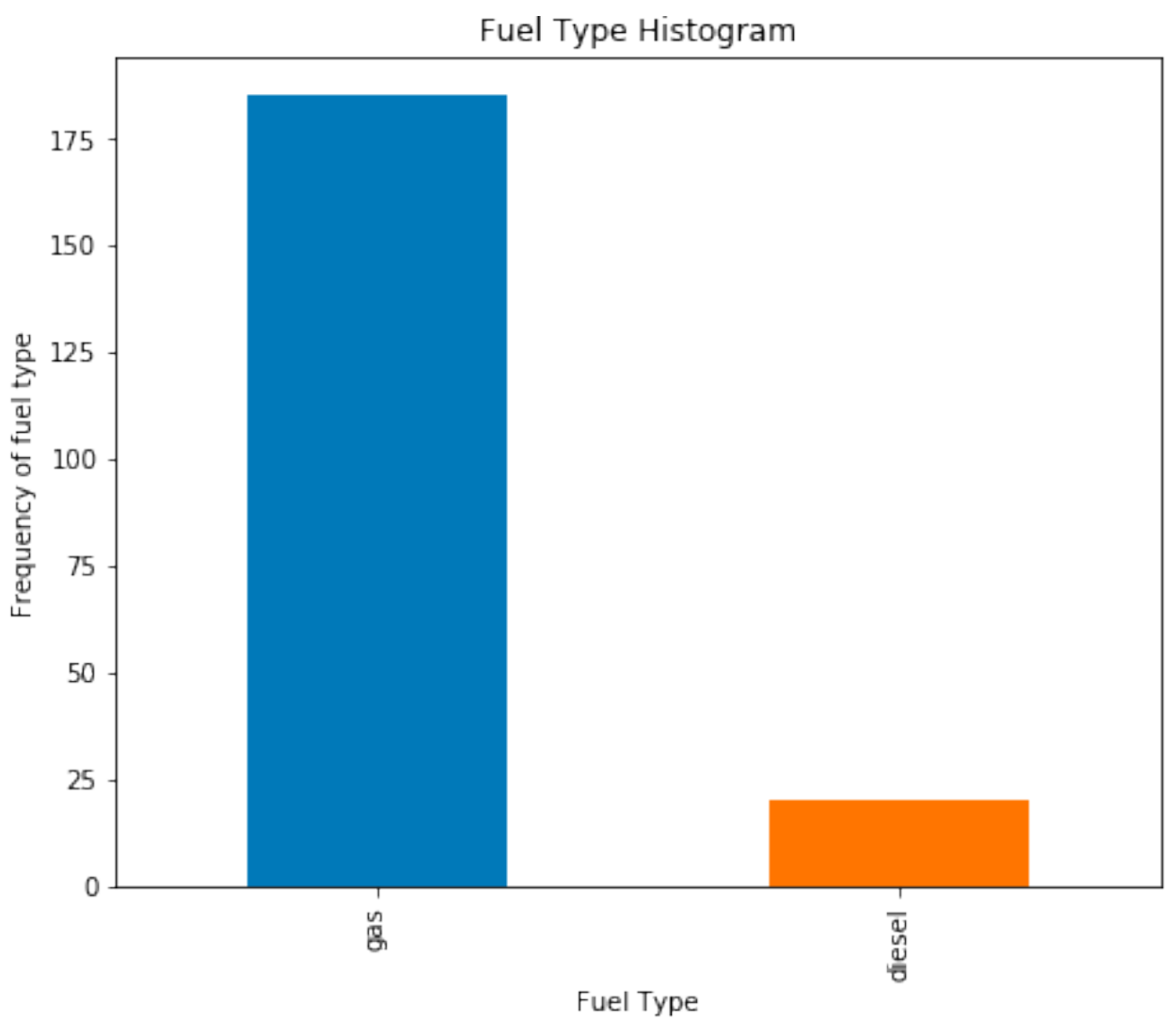
205 filas

Variables Categóricas

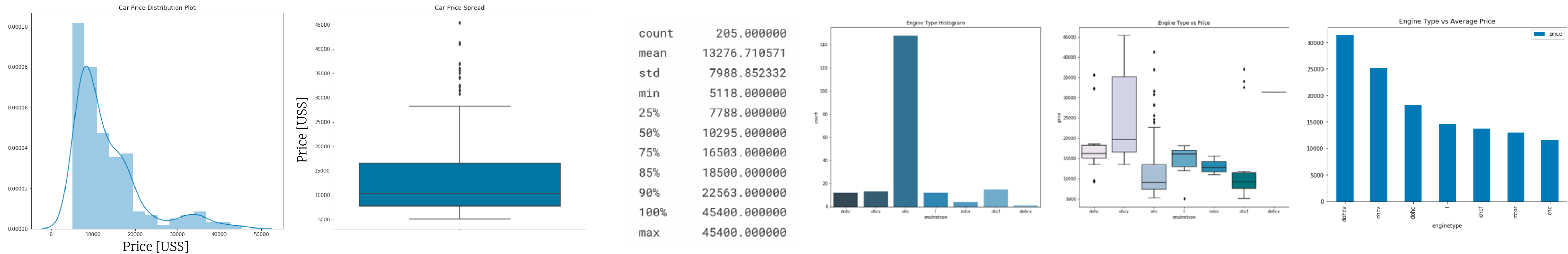
symboling
CarName
fueltype
aspiration
doornumber
carbody
drivewheel
engine loca
engine type
cylindernu
fuelsystem

Variables Numéricas

wheelbase	86.6 - 120.9
carlength	141.1 - 208.1
carwidth	
carheight	
enginesize	
boreratio	
stroke	
compressio n ratio	
horsepower	
peakrpm	
citympg	
highwaympg	
price	

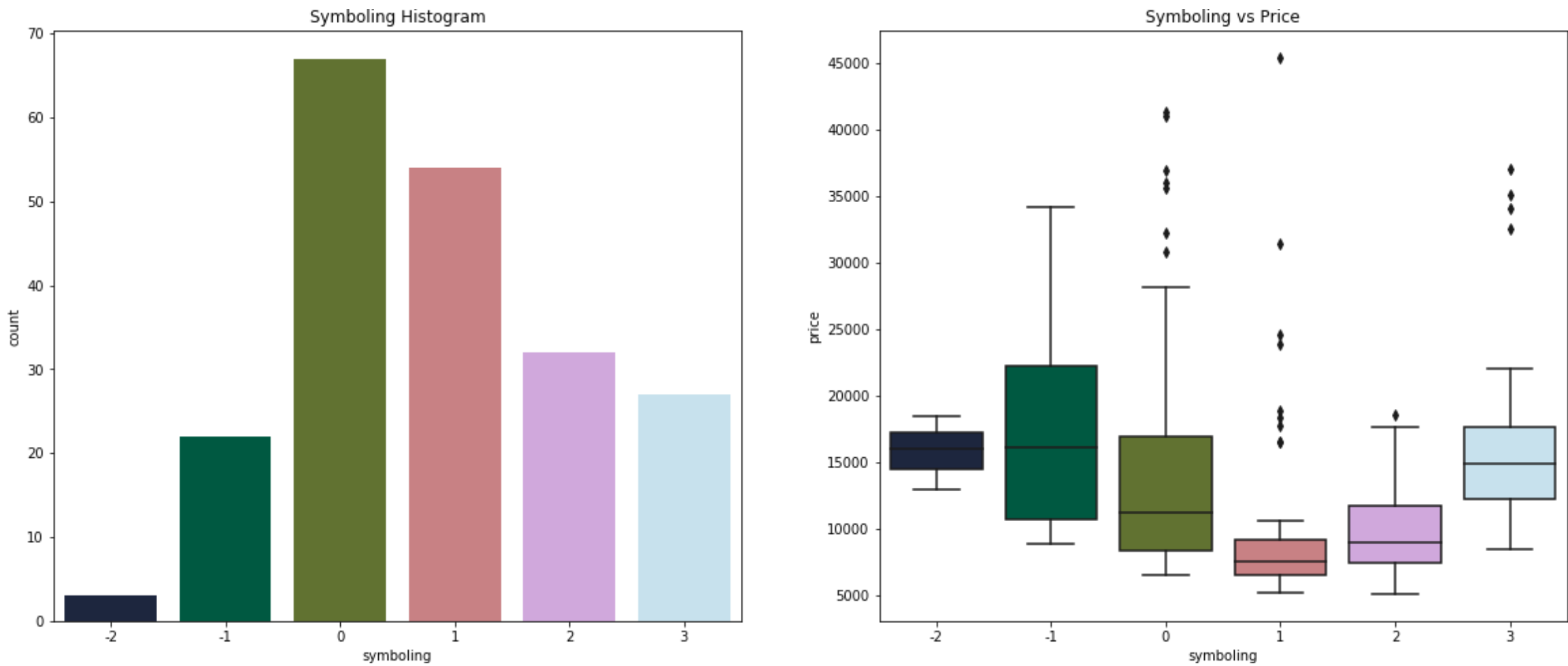


Fuente: Kaggle - Dataset sintético generado para ejercicio de predicción

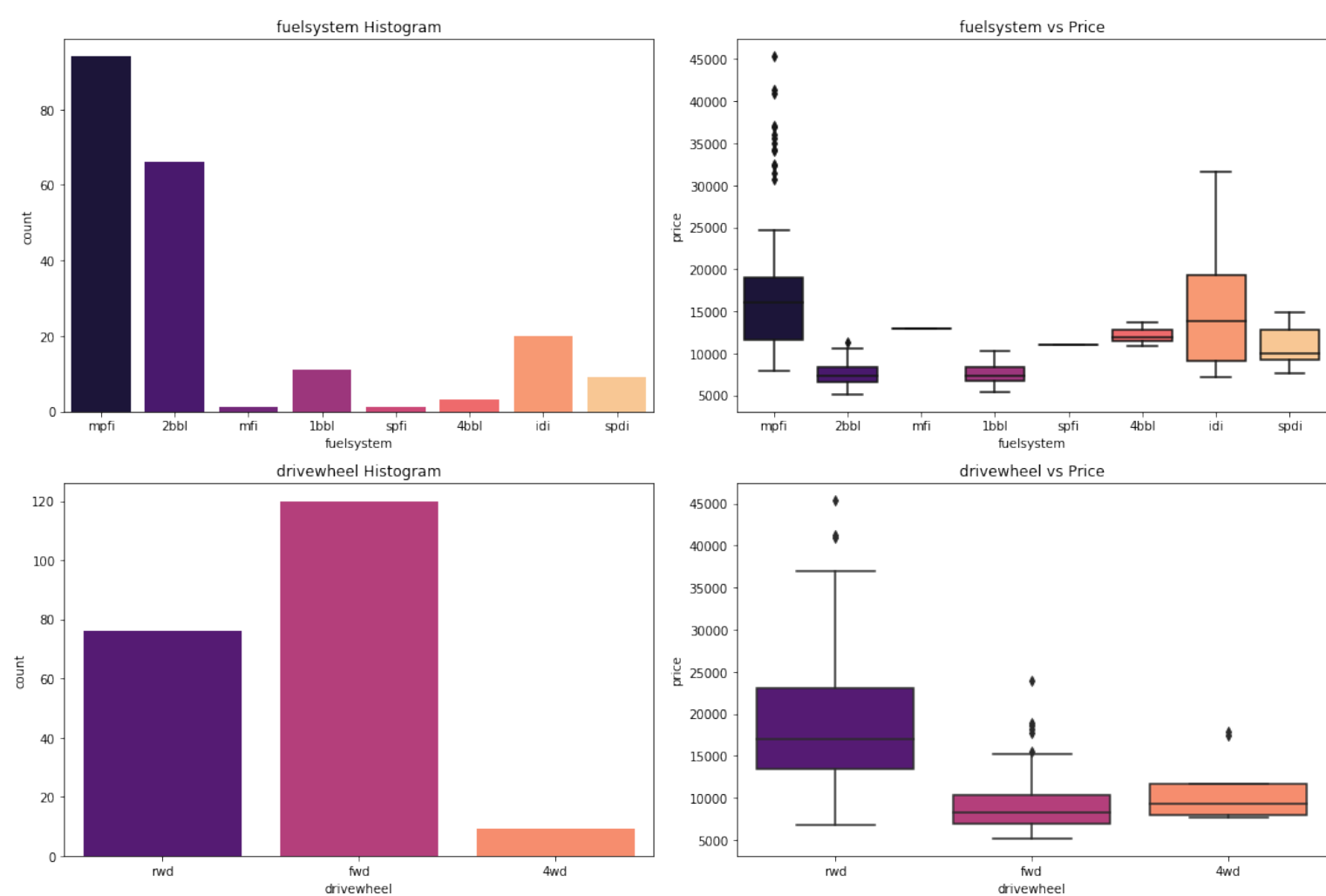


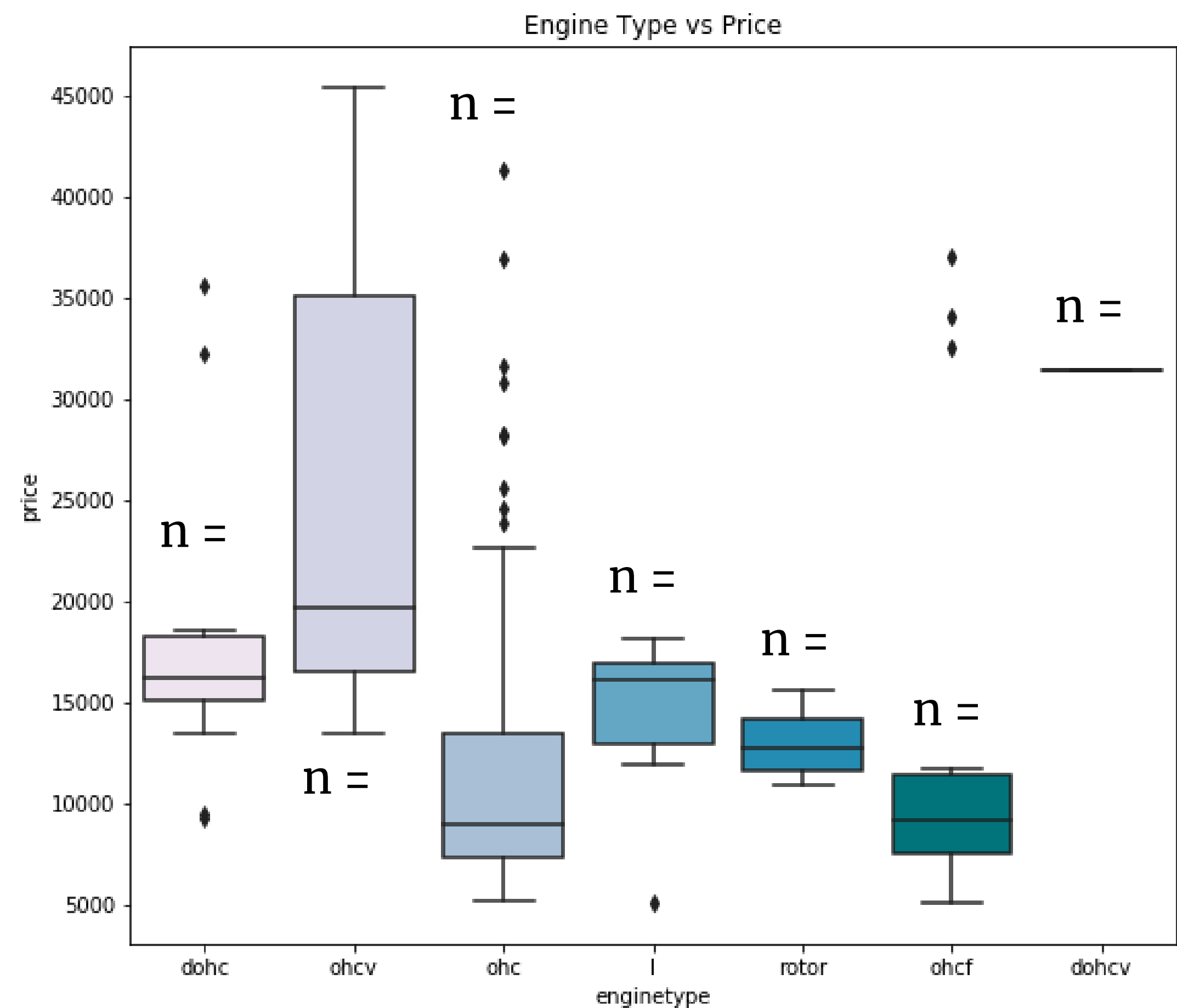
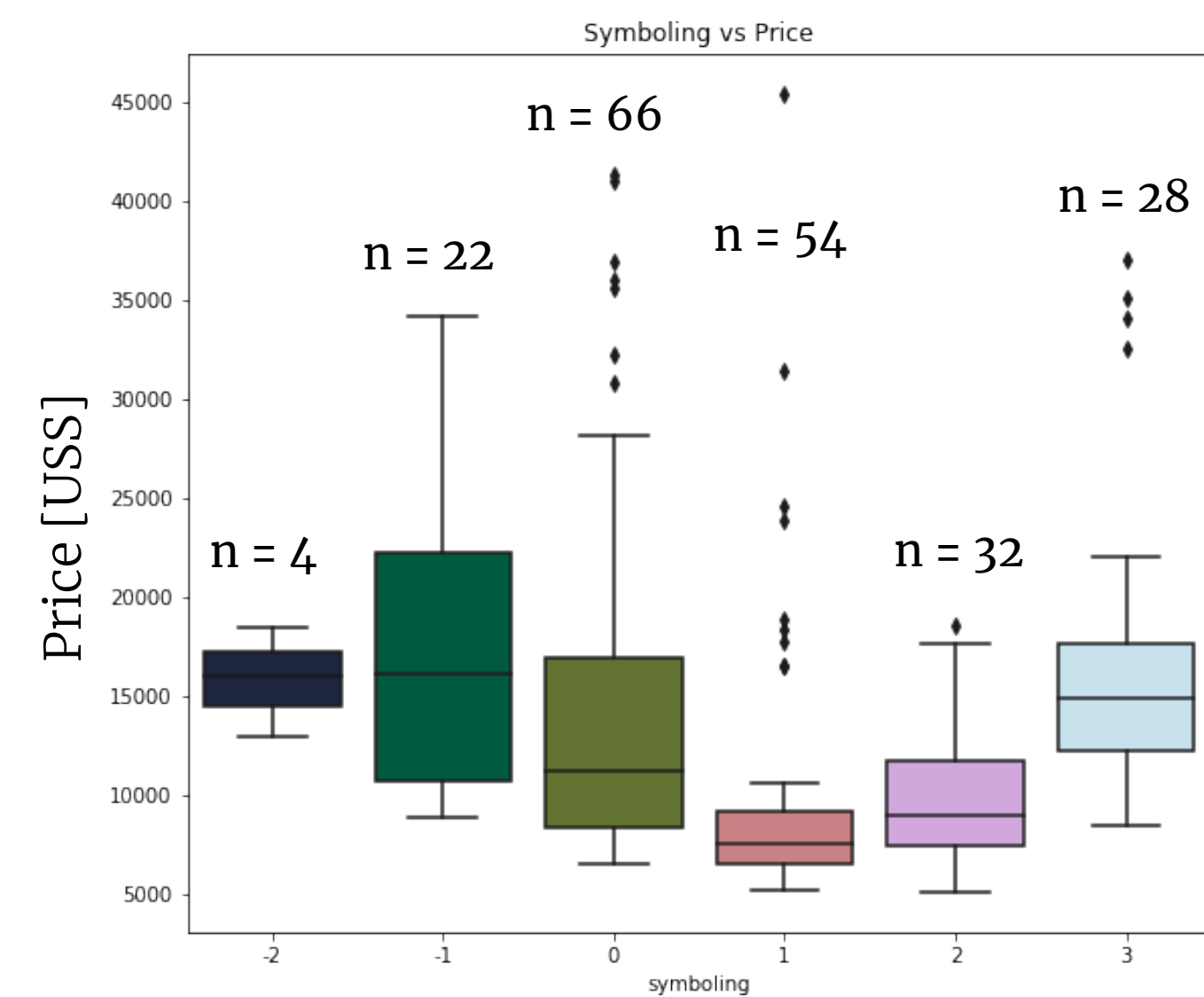
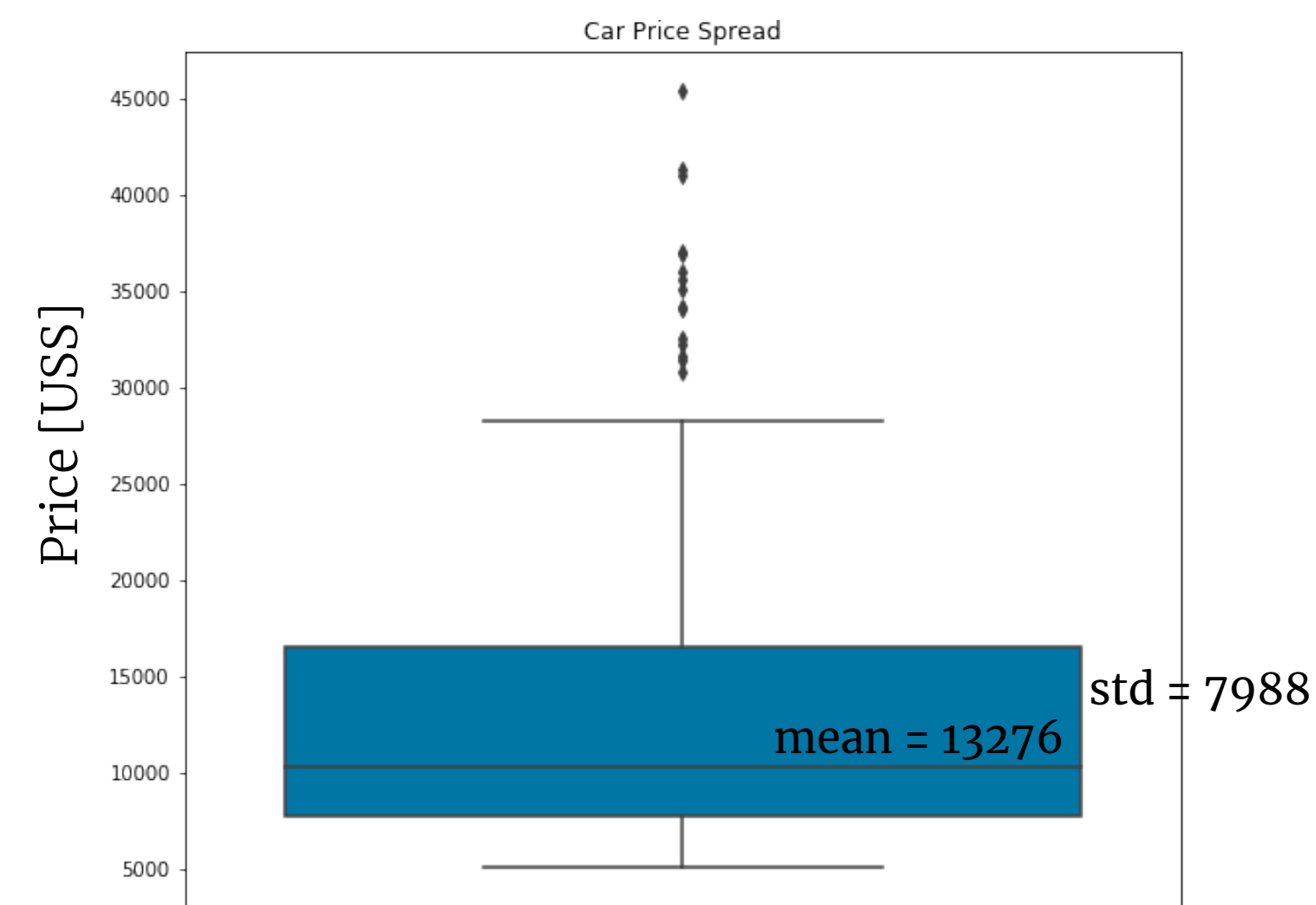
La mayoría de los precios del conjunto de datos son bajos (por debajo de 15.000).
Hay una diferencia significativa entre la media y la mediana de la distribución de precios.

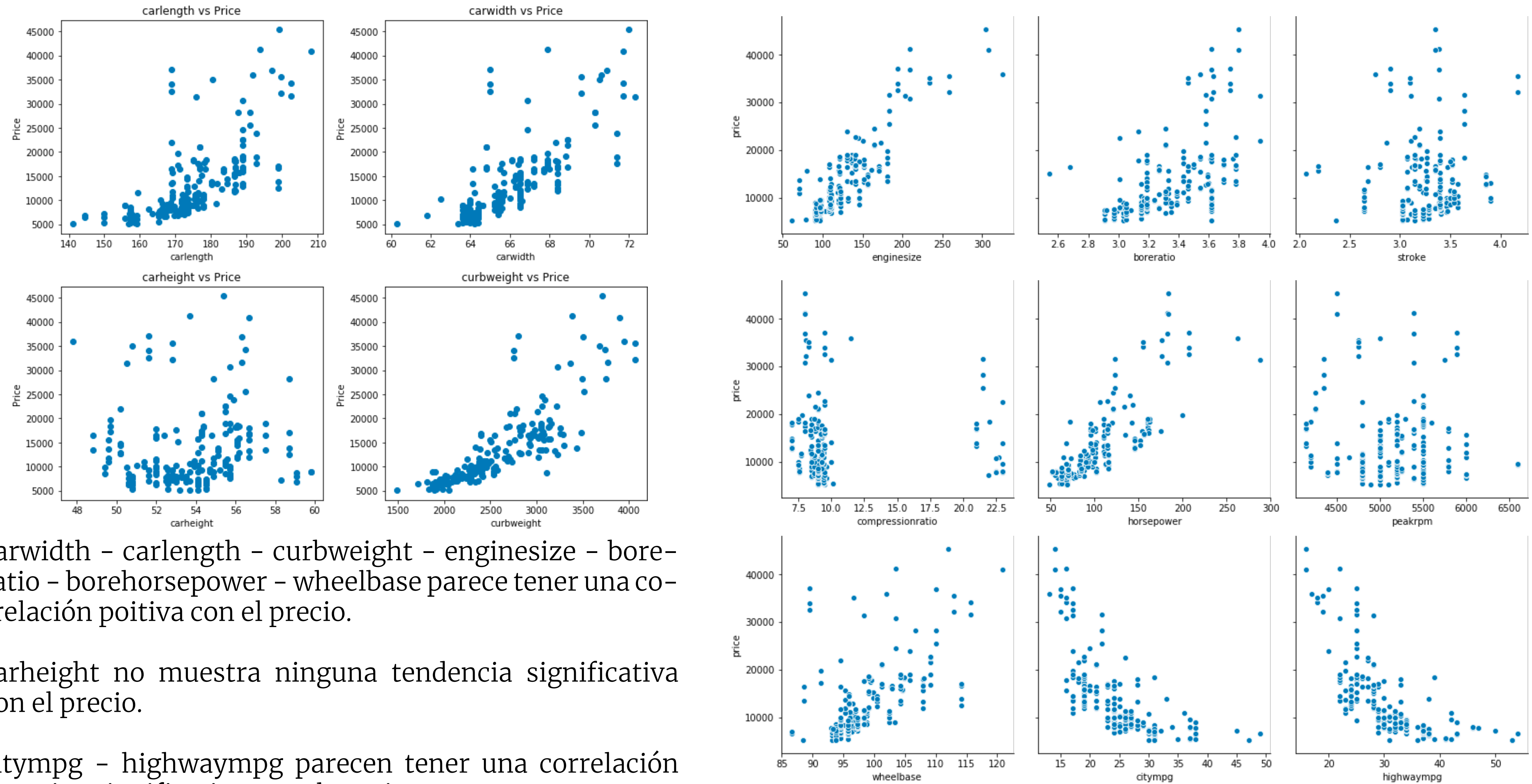
El tipo de motor ohc parece ser el más favorecido.
El ohcv tiene la gama de precios más alta (mientras que el dohcvc sólo tiene una fila),
el ohc y el ohcf tienen la gama de precios más baja.



Parece que las simbologías con valores 0 y 1 tienen un alto número de filas (es decir, son las más vendidas).



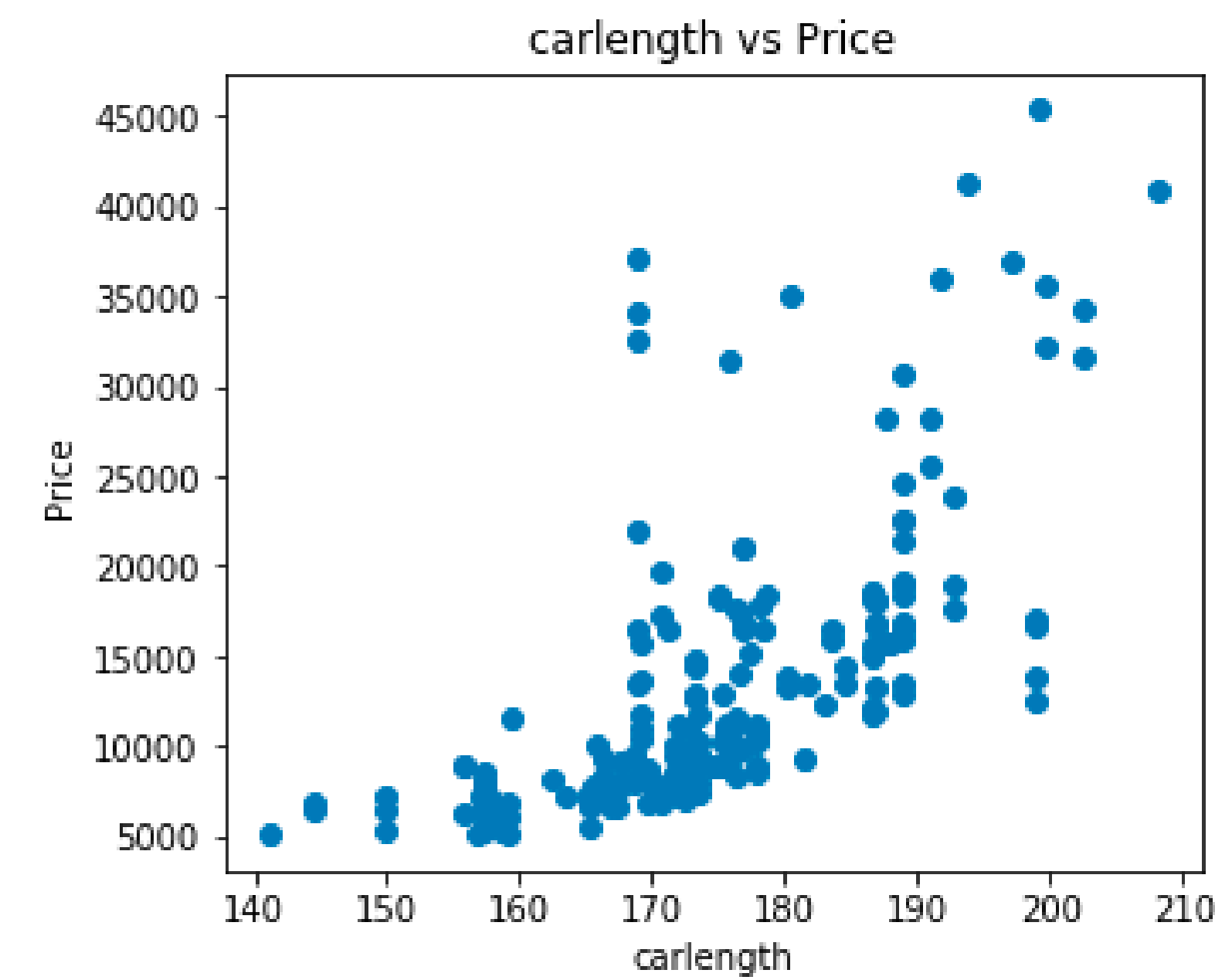




carwidth – carlength – curbweight – enginesize – bore-ratio – borehorsepower – wheelbase parece tener una correlación poitiva con el precio.

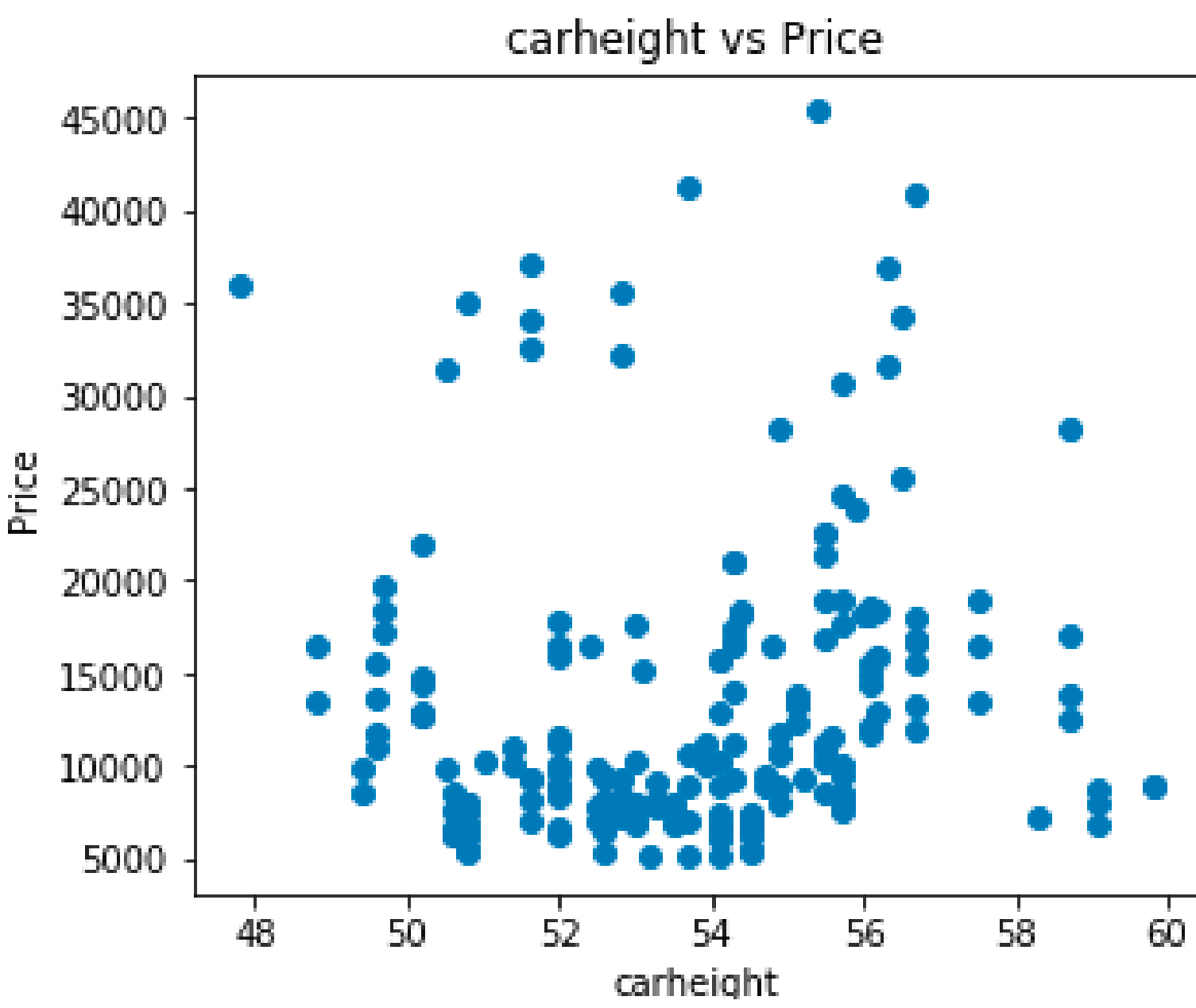
carheight no muestra ninguna tendencia significativa con el precio.

citympg – highwaympg parecen tener una correlación negativa significativa con el precio.



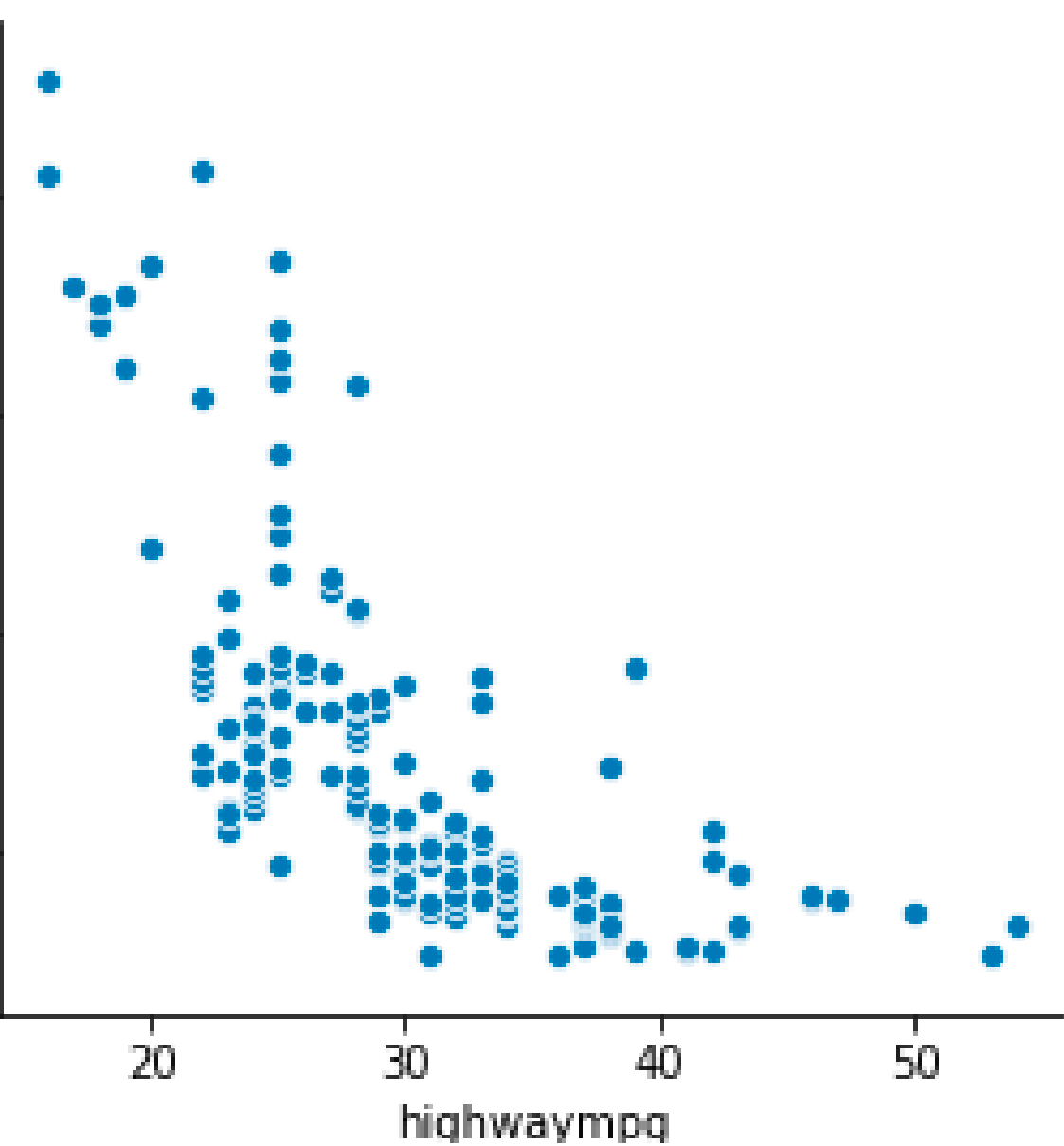
Correlación Positiva

- carwidth
- carlength
- curbweight
- enginesize
- boreratio
- borehorsepower
- wheelbase



Sin Correlación

- carheight
- stroke



Correlación Negativa

- citympg
- highwaympg

ESTANDARIZACIÓN

=====						
Dep. Variable:	price	R-squared:	0.927			
Model:	OLS	Adj. R-squared:	0.922			
Method:	Least Squares	F-statistic:	187.9			
Date:	Fri, 26 Apr 2019	Prob (F-statistic):	4.25e-71			
Time:	09:01:41	Log-Likelihood:	204.17			
No. Observations:	143	AIC:	-388.3			
Df Residuals:	133	BIC:	-358.7			
Df Model:	9					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-0.0764	0.041	-1.851	0.066	-0.158	0.005
curbweight	0.2756	0.069	3.995	0.000	0.139	0.412
horsepower	0.3997	0.069	5.824	0.000	0.264	0.535
fuel economy	0.0736	0.051	1.435	0.154	-0.028	0.175
carwidth	0.2580	0.062	4.137	0.000	0.135	0.381
hatchback	-0.0951	0.025	-3.766	0.000	-0.145	-0.045
sedan	-0.0744	0.025	-2.983	0.003	-0.124	-0.025
wagon	-0.1050	0.028	-3.744	0.000	-0.160	-0.050
dohcv	-0.2319	0.077	-3.015	0.003	-0.384	-0.080
Highend	0.2565	0.020	12.743	0.000	0.217	0.296
=====						

Dep. Variable:	price	R-squared:	0.918			
Model:	OLS	Adj. R-squared:	0.914			
Method:	Least Squares	F-statistic:	215.9			
Date:	Fri, 26 Apr 2019	Prob (F-statistic):	4.70e-70			
Time:	09:01:42	Log-Likelihood:	195.77			
No. Observations:	143	AIC:	-375.5			
Df Residuals:	135	BIC:	-351.8			
Df Model:	7					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-0.0319	0.027	-1.161	0.248	-0.086	0.022
horsepower	0.4690	0.051	9.228	0.000	0.368	0.569
carwidth	0.4269	0.043	9.944	0.000	0.342	0.512
hatchback	-0.1044	0.026	-3.976	0.000	-0.156	-0.052
sedan	-0.0756	0.026	-2.896	0.004	-0.127	-0.024
wagon	-0.0865	0.029	-2.974	0.003	-0.144	-0.029
dohcv	-0.3106	0.070	-4.435	0.000	-0.449	-0.172
Highend	0.2772	0.020	13.559	0.000	0.237	0.318
=====						

Dep. Variable:	price	R-squared:	0.926			
Model:	OLS	Adj. R-squared:	0.922			
Method:	Least Squares	F-statistic:	209.5			
Date:	Fri, 26 Apr 2019	Prob (F-statistic):	7.85e-72			
Time:	09:01:42	Log-Likelihood:	203.07			
No. Observations:	143	AIC:	-388.1			
Df Residuals:	134	BIC:	-361.5			
Df Model:	8					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	-0.0305	0.026	-1.165	0.246	-0.082	0.021
curbweight	0.2593	0.068	3.796	0.000	0.124	0.394
horsepower	0.3469	0.058	5.964	0.000	0.232	0.462
carwidth	0.2488	0.062	3.995	0.000	0.126	0.372
hatchback	-0.0922	0.025	-3.650	0.000	-0.142	-0.042
sedan	-0.0711	0.025	-2.850	0.005	-0.120	-0.022
wagon	-0.1047	0.028	-3.721	0.000	-0.160	-0.049
dohcv	-0.1968	0.073	-2.689	0.008	-0.342	-0.052
Highend	0.2610	0.020	13.083	0.000	0.222	0.301

Dep. Variable:	price	R-squared:	0.913
Model:	OLS	Adj. R-squared:	0.909
Method:	Least Squares	F-statistic:	237.6
Date:	Fri, 26 Apr 2019	Prob (F-statistic):	1.68e-69
Time:	09:01:42	Log-Likelihood:	191.46
No. Observations:	143	AIC:	-368.9
Df Residuals:	136	BIC:	-348.2
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	-0.0934	0.018	-5.219	0.000	-0.129	-0.058
horsepower	0.5001	0.051	9.805	0.000	0.399	0.601
carwidth	0.3963	0.043	9.275	0.000	0.312	0.481
hatchback	-0.0373	0.013	-2.938	0.004	-0.062	-0.012
wagon	-0.0170	0.017	-1.008	0.315	-0.050	0.016
dohcv	-0.3203	0.072	-4.460	0.000	-0.462	-0.178
Highend	0.2808	0.021	13.402	0.000	0.239	0.322

$$\sum_{n=1}^N a_n x_n + b$$

N variables	R ²
9	0.93
8	0.93
7	0.92
6	0.91

Variable	const	curbweight	horsepower	carwidth	hatchback	sedan	wagon	dohcv	Highend
Coefficiente	-0.03 ± 0.03	0.26 ± 0.07	0.35 ± 0.06	...					

Gráfico de Residuos en
función de alguna variable

- Otras métricas calculadas:
- MSE
 - Matriz de Confusión
 - Precisión
 - Curva ROC

- 🌒 Esperábamos xx cosa....
- 🌒 Se logró predecir el precio de los autos con una precisión de....
- 🌒 Llamó la atención la gran dependencia con....
- 🌒 La mayoría de las dependencias ya se habían observado en el análisis previo
- 🌒 El próximo paso sería....



¡ MUCHAS GRACIAS POR SU ATENCIÓN !

juan.gutierrez@gmail.com, josefina.martinez@gmail.com, arturo.aguante.chile@gmail.com

¿Alguna pregunta?

PRESENTACIÓN

Formato PDF

No usar animaciones

No usar diferentes tipografías

Usar diapositivas ocultas

Si no van a mencionarlo o analizarlo, no lo pongan en la diapositiva

Pueden usar las diapositivas como machete

Gráficos: siempre arranquen explicando qué estamos viendo en ese gráfico

Buen contraste. Fondo oscuro: no recomendable

Chequear que todas las palabras/números se vean bien

Enumerar las diapositivas

ORAL

Practiquen la charla

Si les falta tiempo, saquen cosas

Si se pasan de tiempo, le sacan tiempo a su compañerx

Marquen en la diapositiva lo que quieren mostrar

Miren al público mientras hablan (no sólo a lxs docentes)