

Proyecto Final de Bootcamp

*Modelo predictivo de precio
de vehículos usados*

Martín Fernández

Amir Mousa

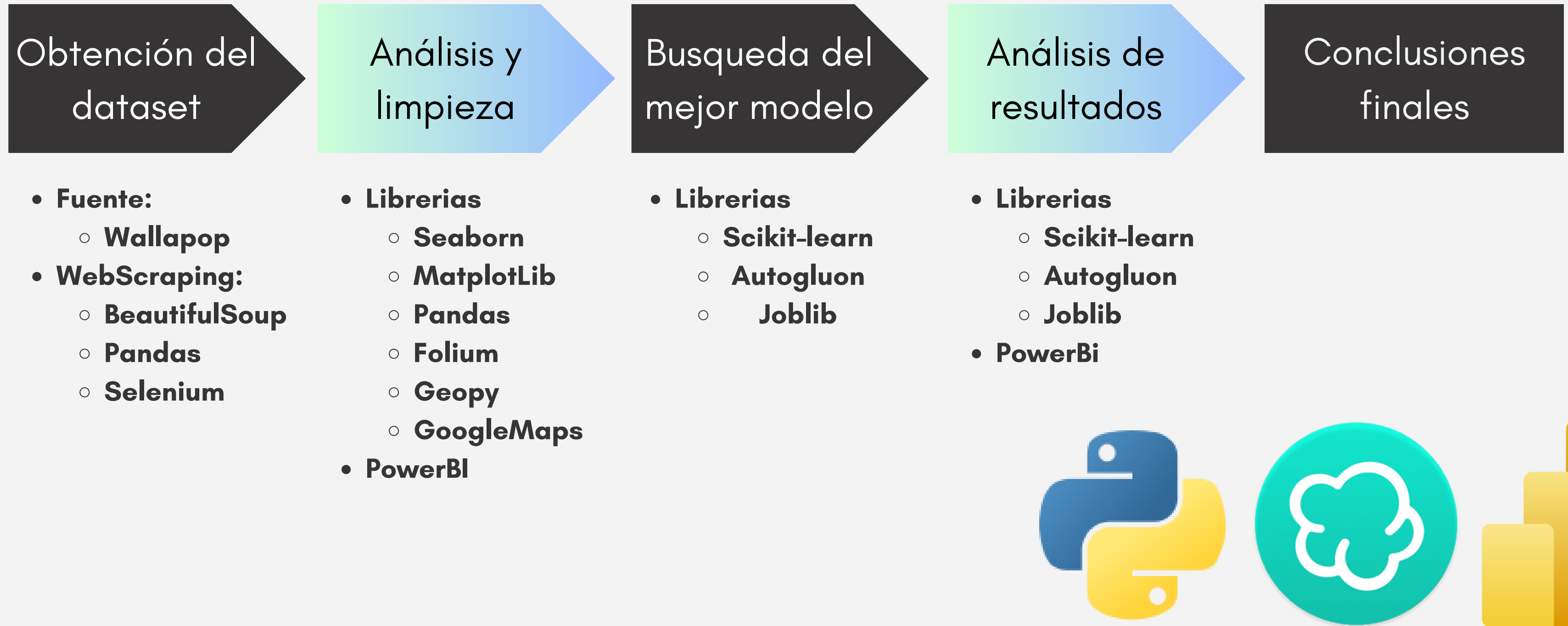
Mario Navarro



¿ Qué buscamos con este proyecto ?

Este proyecto se centra en estudiar los precios del sector de coches usados y desarrollar un modelo de machine learning que proporcione una estimación precisa del valor de los vehículos en el mercado. Analizaremos las tendencias de precios, los factores que los afectan y utilizaremos algoritmos avanzados para crear una herramienta fiable que facilite la compra y venta de coches usados, ofreciendo precios ajustados a la realidad del mercado actual.


Diagrama de flujo





Obtención del Dataset (1/2):


Enlaces

- Rango precios: 1.000€ – 50.000€
- 3000 coches por rango de precio
- Web dinámica: Selenium




Coches 


9000€ - 10000€ 

España, Madrid 

Sólo profesionales

Marca y modelo 

Ordenar por: [Relevancia](#)




9490,00 €


Ford C Max 1.0 EcoBoost 125 Auto StartSt. Titanium

Gasolina · Manual · 125 cv · 2013 · 100200 km

En Dursan, el pago aún más fácil!!! Misma cuota "mitad de tiempo"!!!. Consultar condiciones. Precio al contado: 9490 €. Automotor DURSAN, con más de 18 años al servicio de nuestros clientes: valoración de 4,6 Estrellas / 5 en Google. Contamos con vehículos de ocasión para todas las necesidades y bolsillos. Así como facilidades a la hora de obtener financiación para su compra. Nuestros asesores especializados te...

Destacado






9990,00 €


Peugeot Partner Furgon Confort PackL1 BlueHDi 55KW 75

Diésel · Manual · 75 cv · 2018 · 88549 km

En Dursan, el pago aún más fácil!!! Misma cuota "mitad de tiempo"!!!. Consultar condiciones. Precio al contado: 9990 €. Automotor DURSAN, con más de 18 años al servicio de nuestros clientes: valoración de 4,6 Estrellas / 5 en Google. Contamos con vehículos de ocasión para todas las necesidades y bolsillos. Así como facilidades a la hora de obtener financiación para su compra. Nuestros asesores especializados te...

Destacado






9490,00 €


Ford C Max 1.0 EcoBoost 125 Auto StartSt. Titanium

Gasolina · Manual · 125 cv · 2013 · 100200 km

En Dursan, el pago aún más fácil!!! Misma cuota "mitad de tiempo"!!!. Consultar condiciones. Precio al contado: 9490 €. Automotor DURSAN, con más de 18 años al servicio de nuestros clientes: valoración de 4,6 Estrellas / 5 en Google. Contamos con vehículos de ocasión para todas las necesidades y bolsillos. Así como facilidades a la hora de obtener financiación para su compra. Nuestros asesores especializados te...

Destacado






9990,00 €

Citroën C4 Cactus PureTech 110cv SS Feel

Gasolina · Manual · 110 cv · 2016 · 116249 km

En Dursan, el pago aún más fácil!!! Misma cuota "mitad de tiempo"!!!. Consultar condiciones. Precio al contado: 9990 €. Automotor DURSAN, con más de 18 años al servicio de nuestros clientes: valoración de 4,6 Estrellas / 5 en Google. Contamos con vehículos de ocasión para todas las necesidades y bolsillos. Así como facilidades a la hora de obtener financiación para su compra. Nuestros asesores especializados te...



Obtención del Dataset (2/2):

Datos

Variable Objetivo = Precio

- Numéricas

- Año
- Potencia
- Plazas
- N° de Puertas
- Kilómetros
- Puntuación
- N° de Ventas
- N° valoraciones

- Categóricas:

- Marca
- Modelo
- Combustible
- Tipo de vehículo
- Cambio
- Comentarios
- Nombre usuario
- Ubicación

amado r.
★★★★★ 5 (4)

3 ventas
4 valoraciones

Chat

42.000 €

Mercedes-Benz GLC SUV (253) 2017

4x4

5 plazas

5 puertas

Gasolina

367 caballos

Automático

Marca Mercedes-Benz

Modelo GLC SUV (253)

Año 2017

Kilómetros 150000

Coche con revisiones en casa oficial y con contrato de mantenimiento . Mercedes-Benz GLC SUV (253) en L'Hospitalet de Llobregat

De media, comprando productos como este evitas la emisión de 3207kg de CO2

Más información

Editado hace más de 1 mes

1.3K 8

L'Hospitalet de Llobregat

Ejemplo de Dataset inicial

Marca	Modelo	Año	Potencia	Plazas	Nº de puertas	km	Tipo de vehiculo	Combustible	Cambio	Comentarios	Nombre usuario	Puntuacion	Nº valoraciones	Nº Ventas	Ubicacion	Precio
BMW	Serie 3	2008.0	170	4.0	2.0	235000.0	Coupé	Gasolina	Manual	En buen estado, un solo dueño. mantenimientos ...	Paco G.	0.00	0	0	Pozuelo de Alarcón	9000
BMW	X3	2008.0	177	5.0	5.0	216000.0	4x4	Diésel	Manual	Vendo prcioso BMW X3 2008 4×4 En mui buen esta...	Alexandru M.	5.00	4	0	Madrid	9300
Peugeot	308	2016.0	130	5.0	5.0	176000.0	Pequeño	Gasolina	Manual	Vehículo en impecable estado,modelo más alto d...	Automóviles Borja Concesionario ..	5.00	8	273	Madrid	9990
BMW	Serie 1	2015.0	116	5.0	5.0	340000.0	Pequeño	Diésel	Manual	Vehículo en impecable estado,poco consumo y bu...	Automóviles Borja Concesionario ..	5.00	8	273	Madrid	9490
Citroen	Berlingo	2021.0	NaN	NaN	NaN	124280.0	NaN	NaN	NaN	Clicars. Tu compra inteligente. Calidad al mej...	Clicars S.	5.00	3	61371	Madrid	9990
...
Opel	Vivaro	2013.0	NaN	NaN	NaN	375635.0	NaN	NaN	NaN	en muy buen estado y posibilidad de pintarla e...	Juan Jose D.	4.75	1	38	Móstoles	10.000
Mercedes-Benz	406 D	1995.0	NaN	NaN	NaN	57000.0	NaN	NaN	NaN	Autocaravana con muy pocos kms, todo funciona,...	Hector M.	5.00	3	36	Tres Cantos	10.000
Volkswagen	Transporter T5	2007.0	174	3.0	4.0	530000.0	Furgoneta	Diésel	Manual	Se vende Volkswagen T5 Esta muy bien estado,fu...	Nazario N.	5.00	7	12	Pinto	9800
Mercedes-Benz	Clase C	2008.0	122	4.0	3.0	111000.0	Pequeño	Diésel	Manual	En perfecto estado interior, exterior y de mec...	fernando c.	5.00	5	61	San Sebastián de Los Reyes	9900
Volkswagen	Golf	2015.0	110	5.0	5.0	201000.0	Pequeño	Diésel	Manual	VW Golf 7 Variant del 10/2015. Version 2016 co...	Carlos G.	5.00	1	5	San Martin	9990

Número de columnas: 17

- Numéricas: 9
- Categóricas: 7
- Texto: 1

Limpieza del dataset:

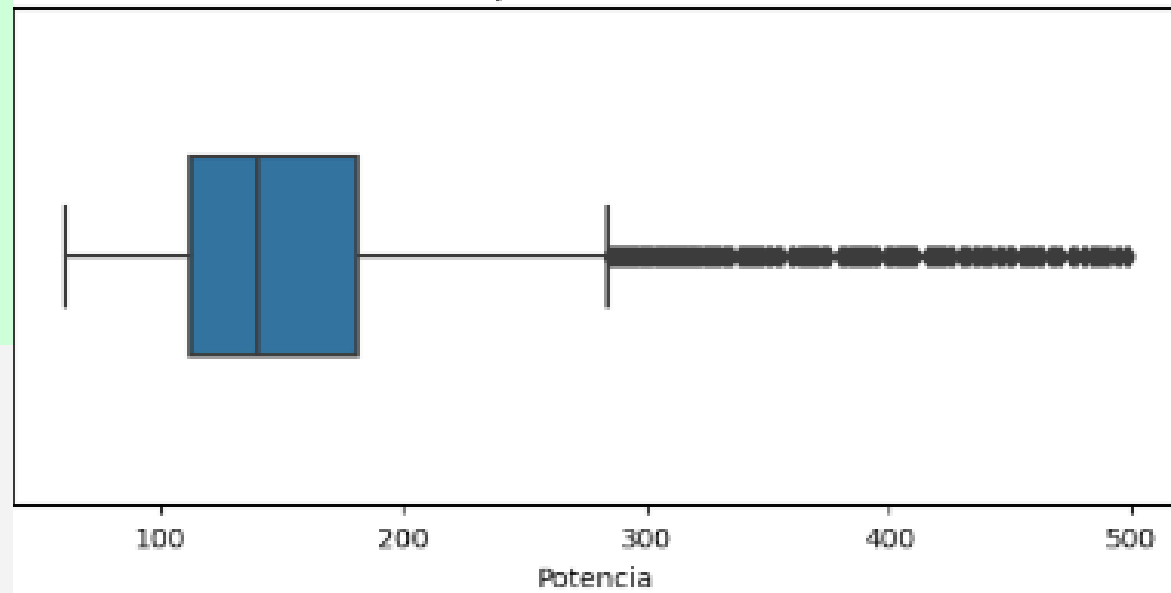
- Gran cantidad de nulos
- **Anuncios repetidos con diferentes ubicaciones**
- Mala redacción
 - Agrupación de marcas
 - Valores con mal formato
- Tratamiento de Outliers

Datos útiles

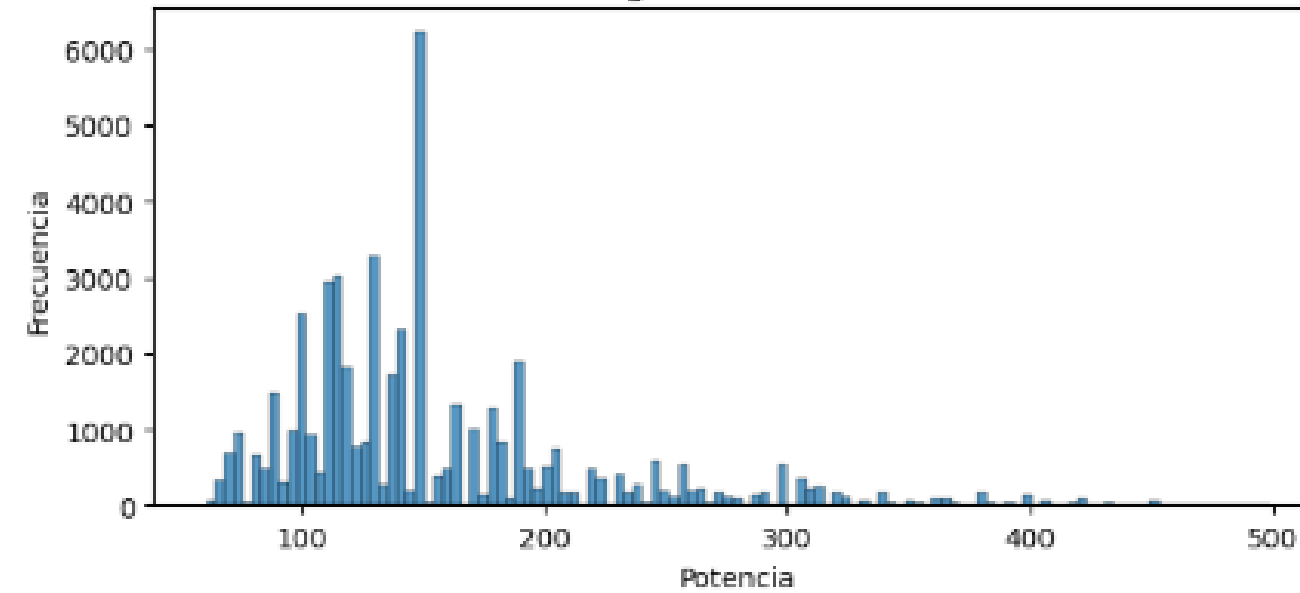
30% ~ 40%

Algunos gráficos:

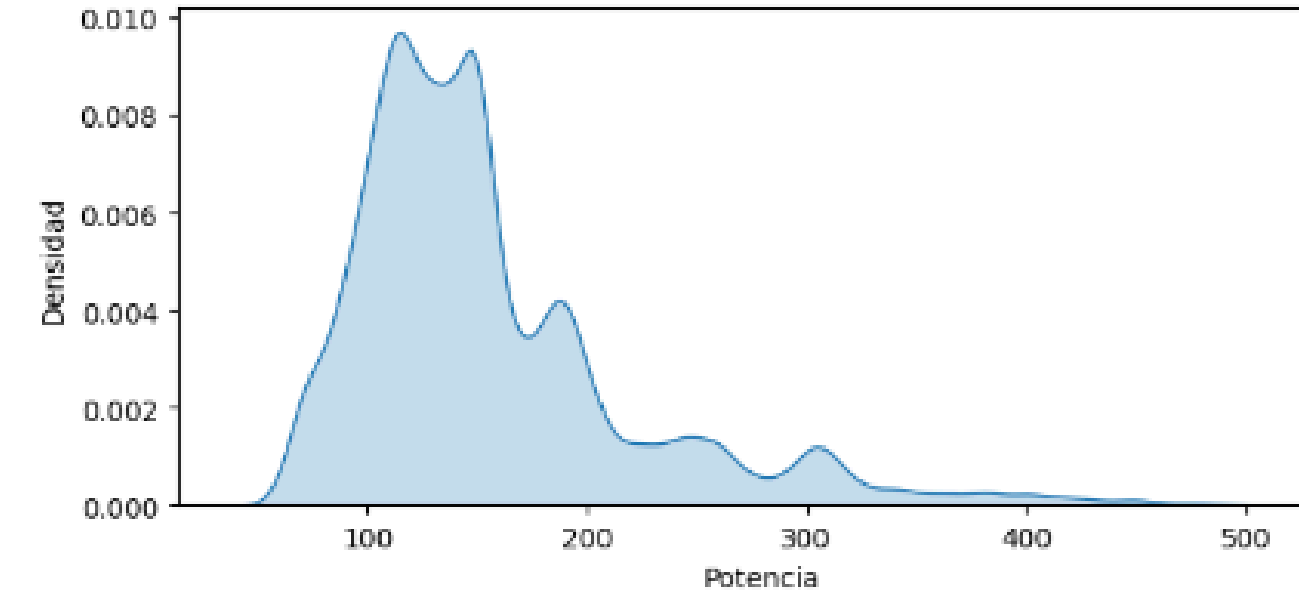
Boxplot de Potencia



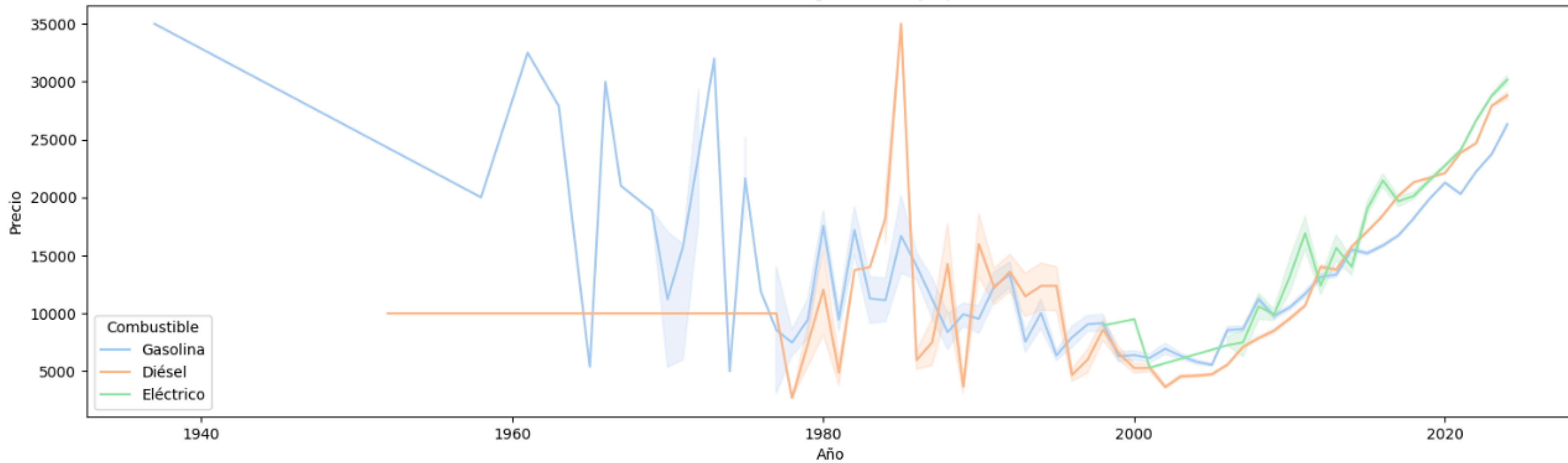
Histograma de Potencia



Densidad de Potencia



Precio de los coches según su año y tipo de motor

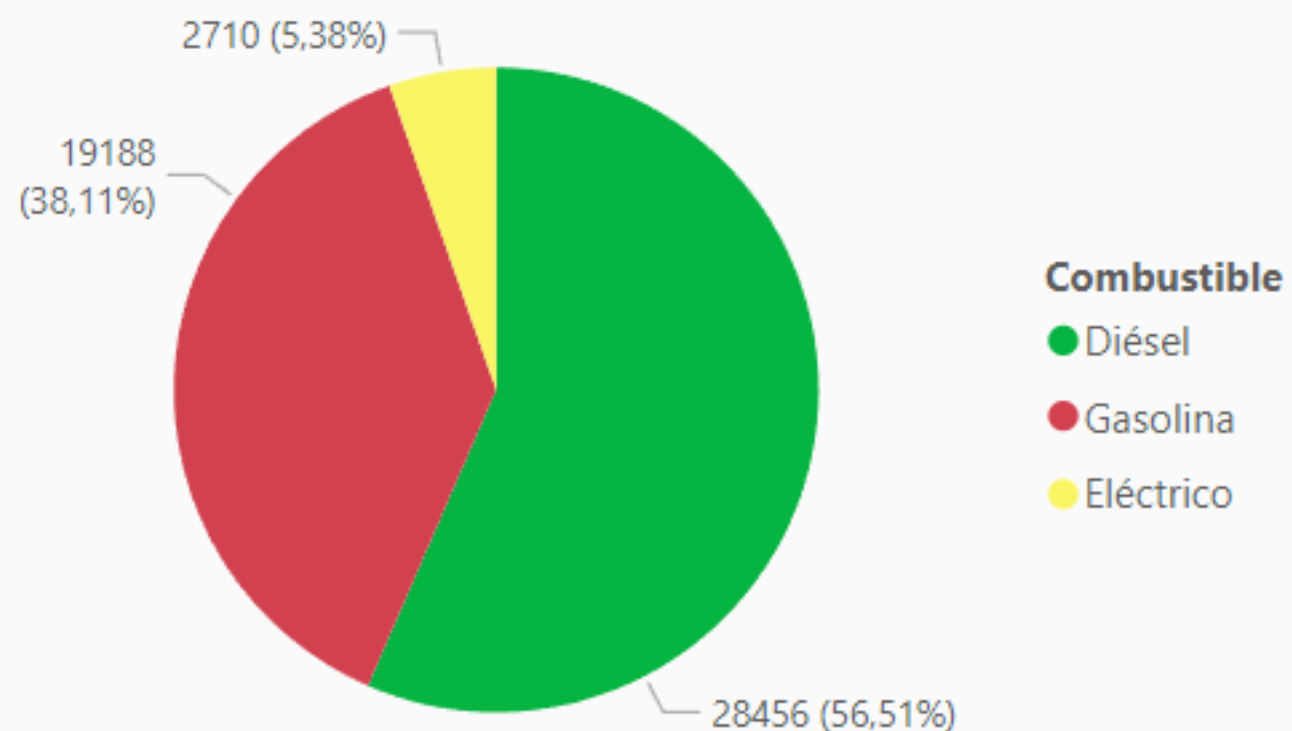


Vehículos totales
50.354

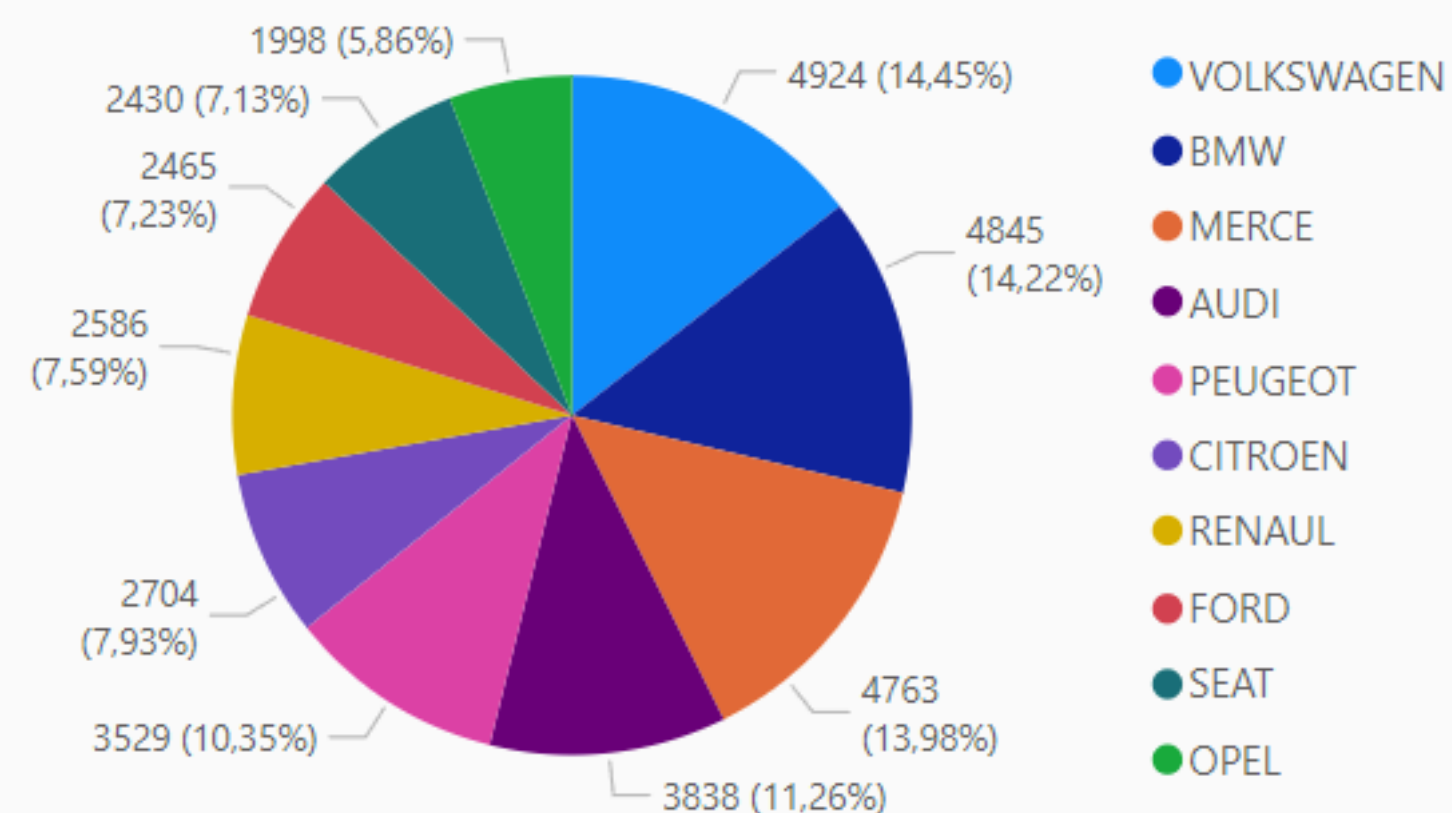
Marcas de coche
61

Precio medio
19112

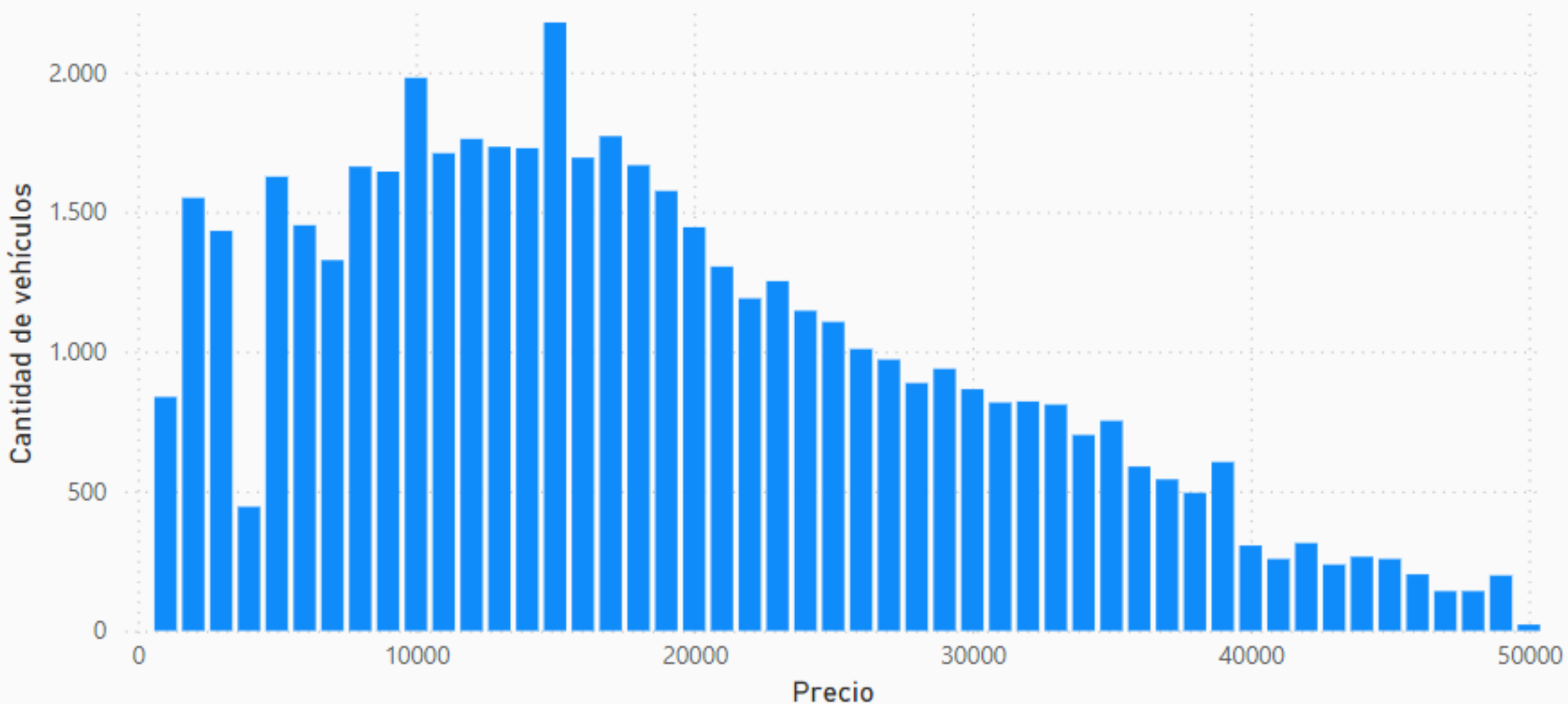
Cantidad de coches por tipo de combustible



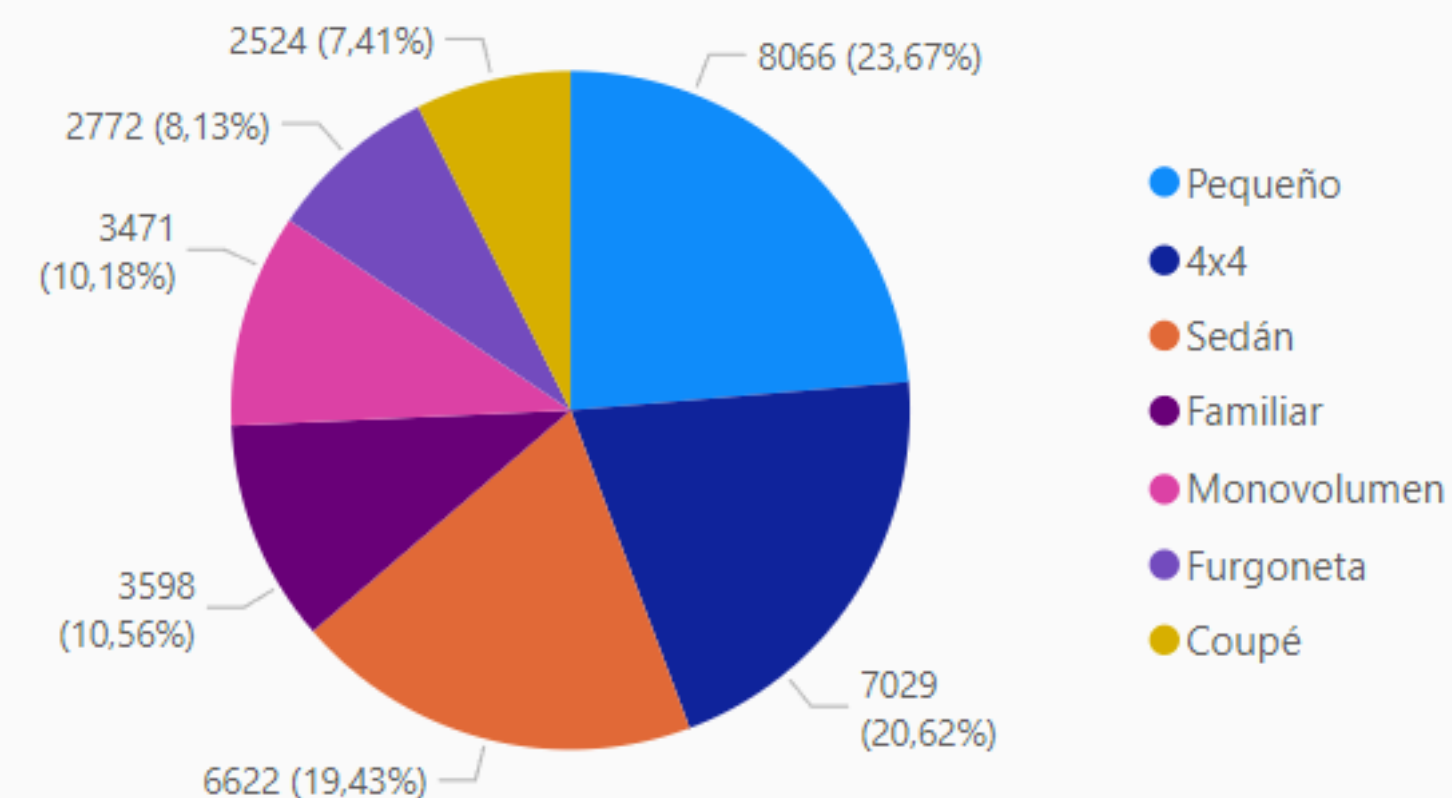
Top 10 marcas más ofertadas



Recuento de coches por rango de precio



Recuento de tipo de vehículo



Preparando el modelo... creación/eliminación de columnas

- Vehículos camperizados
- Distinción entre marcas caras
- Coches averiados
- Dummies para las marcas, tipo de vehículo, combustible y cambio
- Eliminación de columnas

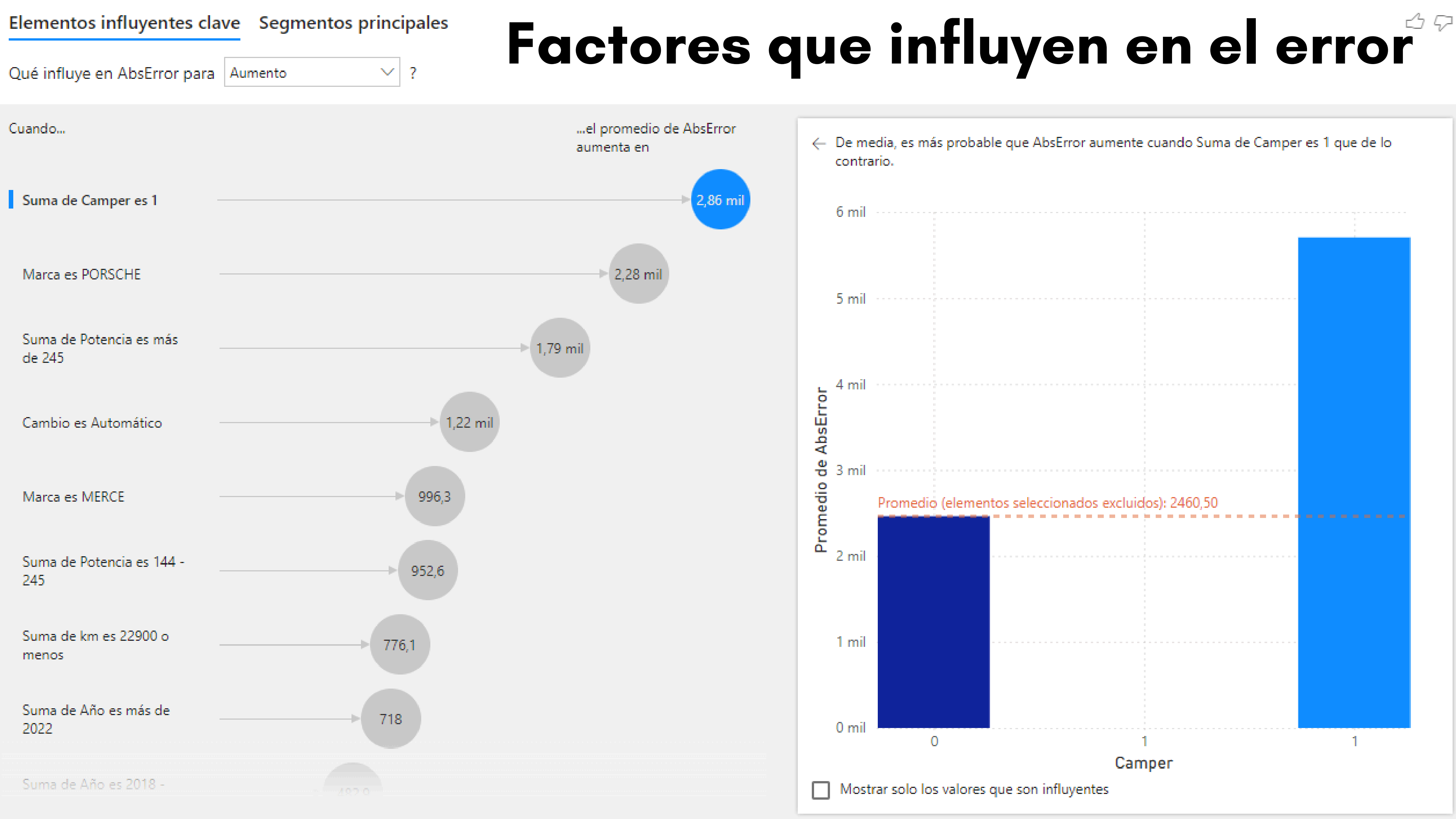
Dataset inicial	Dataset modelo
17 columnas	75 columnas

MODELOS ENTRENADOS

<u>Modelo</u>	<u>R2 train score</u>	<u>R2 test score</u>	<u>Param + importantes</u>
XGBoost	0.925	0.923	Potencia -> Año -> Km
CatBoost	0.952	0.934	Año -> Potencia -> Km
GradientBoosting	0.962	0.934	Año -> Km-> Potencia

Mejores hiperparámetros

- n_estimators=800
- max_depth=5
- learning_rate=0.01
- reg_lambda= 0.1
- min_child_weight=1
- gamma=0.1
- subsample=0.8
- colsample_bytree=0.8
- reg_alpha=0.1
- early_stopping_rounds=10



CONCLUSIONES

- Lentitud de WebScraping (en cuanto a rangos de búsqueda de datos).
- La mala calidad de los anuncios de Wallapop.
- El modelo de los vehículos no tiene una gran influencia en el modelo predictivo.
- Cada línea de código ha sido una dura batalla.

MEJORAS A FUTURO

- Crear una interfaz para hacer la aplicación del modelo más *user friendly*.
- Obtener datos de diferentes fuentes (ej. coches.net).
- Optimizar el método de obtención de datos para convertir el modelo en un buscador.

Muchas Gracias

Martín Fernández

Amir Mousa

Mario Navarro

ID Bootcamps

Data Science & Machine Learning