

2019

Práctica 1



TIPOLOGIA Y CICLO DE VIDA DE DATOS

## Contenido

1. Componentes del grupo.....	3
2. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.....	4
3. Definir un título para el dataset. Que sea descriptivo.....	5
4. Descripción del dataset. Desarrollar una breve descripción de la información aportada.....	5
5. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente.....	6
6. Contenido. Explicar los campos del Dataset Obtenido.....	10
7. Agradecimientos. Presentar al propietario del conjunto de datos.....	11
8. Inspiración. Explique por qué es interesante este conjunto de datos y que preguntas puede responder.....	12
9. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección.....	15
10. Contribuciones.....	17

## 1. Componentes del grupo

David Quiles Gómez

Iván López-Baltasar Benito

## 2. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.

En la red disponemos de varias webs y portales que contienen gran información relativa a la compra-venta de vehículos de segunda mano, km 0 y de ocasión. Realizando Web Scraping a las principales web de este tipo podríamos obtener información muy rápidamente de gran cantidad de vehículo de cualquier provincia de España facilitando además todas aquellas características que suelen buscar los usuarios como Kilometraje, precio, estado, color, etc.

Hemos elegido la pagina Web MilAnuncios.com debido a que contiene gran cantidad de información bien estructurada y responde prácticamente a todas las preguntas que suelen formularse los usuarios. En esta Web, se dispone de información no sólo de particulares sino también de concesionarios que publican semanalmente gran cantidad de información de todos sus vehículos. Además, esta web también nos ofrece información de cualquier parte de España.

Para la obtención de la información se ha utilizado un script en Python. En principio hemos tenido bastantes problemas a la hora de ejecutar el script ya que en la primera conexión, el sistema detectaba que las peticiones no las estaba haciendo un humano y nos mostraba una página donde se solicitaba un captch o similar. Suponemos que dada la naturaleza de la información que MilAnuncios proporciona, es un sitio muy susceptible de recibir peticiones automáticas.

Hemos probado bastantes cosas, lo primero de todo fue establecer unas cabeceras en la petición donde entre otras cosas, se informara del "Agent User". Así nos mostraba dos peticiones, a la tercera el sistema volvía a detectar que las peticiones eran realizadas por un "bot". Probamos bastantes cosas más como espaciar las peticiones mediante ejecuciones "sleep" de duración aleatoria, así como intentar salir por diferentes proxys (esto lo vimos en diferentes sitios de internet donde daban consejos a la hora de evitar baneos al realizar web scraping). También probamos utilizando Selenium, alternando incluso el navegador a utilizar, pero como máximo obteníamos 3 peticiones antes del baneo.

Finalmente lo que mejor resultado nos ha dado ha sido realizar una conexión previa con un navegador, revisar las cabeceras enviadas en la petición (a través de un inspector de red incorporado en FireFox Developer Edition –fig 1), obtener el código de la cookie establecida y ese código meterlo en las cabeceras de la petición (fig 2).

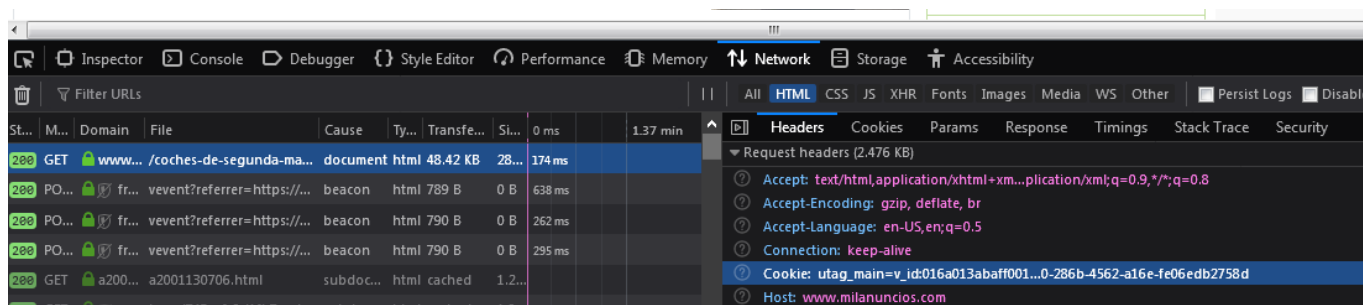


Fig 1 Revisión de las cabeceras enviadas por el navegador en la petición

```

headers = {'Host': 'www.milanuncios.com'}

print (headers)
page = requests.get(url, headers=headers)
print (page.text)
try:
    fichRespuesta = 'respuesta-'+str(pagina)+'.html'
    file = open(fichRespuesta, "wb+")
    contenido = page.text.encode('UTF-8')
    file.write(contenido)
    file.flush()
    file.close()
except Exception as e:
    print (e)
    pass
return page

```

Fig 2 Fragmento del código Python donde se incluye la cookie en la cabecera

### 3. Definir un título para el dataset. Que sea descriptivo.

El título elegido para el dataset será “FindMyCar”.

Obviamente elegimos un título que pueda dar idea rápida de cuál será su utilidad.

### 4. Descripción del dataset. Desarrollar una breve descripción de la información aportada.

El Dataset obtenido muestra toda la información que un usuario corriente podría considerar importante a la hora de comprar un vehículo. El dataset se muestra en filas fácilmente manipulable para posteriormente poder ser filtrado, manipulado, almacenado en una BBDD o bien ser tratado con alguna técnica de Minería de datos como Clustering o Clasificación por ejemplo, si los consideramos necesario y nuestros objetivos van más allá que del simple hecho de comprar un coche. Quizás nos interese conocer en qué ciudad se vende más un tipo de coche o en qué provincia disponemos de mayor stock, etc.

## 5. Representación gráfica. Presentar una imagen o esquema que identifique el *dataset visualmente*.

Una vez ejecutado el script, el Dataset Original es el mostrado a continuación, hay que tener en cuenta que el script se ha ejecutado filtrando por año de matriculación entre 2010 y 2018 y por precio entre 4000 y 15000 euros y que no se han obtenido para su análisis todos los resultados que la web tiene en su base de datos:

Coche	Link	Precio	Año	Kms	Combustible	Potencia	Puertas	Cambio	Localizacion
RENAULT - SPORT TOURER MEGANE, 1, 2 TCE	www.milanuncios.com/renault-de-segunda-mano/renault-sport-tourer-megane-1-2-tce-29	6.500 €	año 2013	115.000 kms	gasolina	115 cv	5 puertas	manual	balears
VOLKSWAGEN - POLO EDITION 1. 4 TDI 55KW75CV BMT	www.milanuncios.com/volkswagen-de-segunda-mano/volkswagen-polo-edition-1-4-tdi-55	9.500 €	año 2016	30.711 kms	diesel	75 cv	5 puertas	manual	tarragona
AUDI - A1 SPORTBACK 1. 6 TDI 90CV ATTRACTED	www.milanuncios.com/audi-de-segunda-mano/audi-a1-sportback-1-6-tdi-90cv-attracted-27	12.900 €	año 2013	47.446 kms	diesel	90 cv	5 puertas	manual	tarragona
BMW - SERIE 1 116i	www.milanuncios.com/bmw-de-segunda-mano/bmw-serie-1-116i-280352763.htm	13.490 €	año 2014	114.000 kms	gasolina	136 cv	3 puertas	manual	madrid
SEAT - LEON ST 1. 6 TDI 105CV STSP STYLE	www.milanuncios.com/seat-de-segunda-mano/seat-leon-st-1-6-tdi-105cv-stsp-style-27645i	9.900 €	año 2014	129.073 kms	diesel	105 cv	5 puertas	manual	tarragona
VOLKSWAGEN - GOLF 1. 6 TDI 105CV DPF SPORT	www.milanuncios.com/volkswagen-de-segunda-mano/volkswagen-golf-1-6-tdi-105cv-dpf-i	9.500 €	año 2011	99.271 kms	diesel	105 cv	3 puertas	manual	tarragona
FORD - FOCUS 1. 6 TDCI 115CV TREND	www.milanuncios.com/ford-de-segunda-mano/ford-focus-1-6-tdci-115cv-trend-284168422	10.900 €	año 2015	70.102 kms	diesel	115 cv	5 puertas	manual	tarragona
JEEP - GRAND CHEROKEE 3. 0 V6 CRD LAREDO	www.milanuncios.com/jEEP-de-segunda-mano/jEEP-grand-cherokee-3-0-v6-crd-laredo-26i	13.500 €	año 2010	119.036 kms	diesel	218 cv	5 puertas	automat	tarragona
FORD - ECOSPORT 1. 5 TDCI 70KW 95CV TITANIUM S	www.milanuncios.com/ford-de-segunda-mano/ford-ecosport-1-5-tdci-70kw-95cv-titanium-s	14.900 €	año 2017	24.351 kms	diesel	95 cv	5 puertas	manual	tarragona
PEUGEOT - 208 5P ACTIVE 1. 4 HDI 68	www.milanuncios.com/peugeot-de-segunda-mano/peugeot-208-5p-active-1-4-hdi-68-2687	7.900 €	año 2014	67.443 kms	diesel	68 cv	5 puertas	manual	tarragona
BMW - SERIE 1 116d EFFICIENTDYNAMICS	www.milanuncios.com/bmw-de-segunda-mano/bmw-serie-1-116d-efficientdynamics-2487	14.900 €	año 2014	71.464 kms	diesel	116 cv	3 puertas	manual	tarragona
DACIA - DUSTER AMBIANCE DCI 110 4X2 EU6	www.milanuncios.com/dacia-de-segunda-mano/dacia-duster-ambiance-dci-110-4x2-eu6-2i	10.900 €	año 2015	65.941 kms	diesel	109 cv	5 puertas	manual	tarragona
VOLKSWAGEN - POLO 1. 4 ADVANCE	www.milanuncios.com/volkswagen-de-segunda-mano/volkswagen-polo-1-4-advance-2990	4.999 €	año 2012	140.000 kms	gasolina	85 cv	5 puertas	manual	tenerife
FORD - RANGER 2. 2 TDCI 150CV 4X4 DOBLE CABINA XL	www.milanuncios.com/ford-de-segunda-mano/ford-ranger-2-2-tdci-150cv-4x4-doble-cabin	14.793 €	año 2014	69.000 kms	diesel	150 cv	4 puertas	manual	la_coruna
MERCEDES-BENZ - CLASE B B 180 CDI AUT.	www.milanuncios.com/mercedes-benz-de-segunda-mano/mercedes-benz-clase-b-b-180-ci	13.900 €	año 2013	89.000 kms	diesel	109 cv	5 puertas	automat	la_coruna
VOLKSWAGEN - POLO 1. 6 TDI 90CV DSG ADVANCE	www.milanuncios.com/volkswagen-de-segunda-mano/volkswagen-polo-1-6-tdi-90cv-dsg-a	10.900 €	año 2013	119.000 kms	diesel	90 cv	5 puertas	automat	la_coruna
AUDI - A3 SPORTBACK 1. 6 TDI 110CV CLEAN ATTRACTION	www.milanuncios.com/audi-de-segunda-mano/audi-a3-sportback-1-6-tdi-110cv-clean-attra	13.900 €	año 2015	139.000 kms	diesel	110 cv	5 puertas	manual	la_coruna
AUDI - Q3 2. 0 TDI 140CV AMBIENTE	www.milanuncios.com/audi-de-segunda-mano/audi-q3-2-0-tdi-140cv-ambiente-299826780	13.900 €	año 2013	209.000 kms	diesel	140 cv	5 puertas	manual	la_coruna
PEUGEOT - 308 5P ACCESS 1. 6 HDI 92 FAP	www.milanuncios.com/peugeot-de-segunda-mano/peugeot-308-5p-access-1-6-hdi-92-fap-	5.900 €	año 2011	139.000 kms	diesel	92 cv	5 puertas	manual	la_coruna
CITROEN - C4 C4 BLUEHDI 120 6V FEEL EDITION	www.milanuncios.com/citroen-de-segunda-mano/citroen-c4-c4-bluehdi-120-6v-feel-editio	8.900 €	año 2016	140.000 kms	diesel	120 cv	5 puertas	manual	la_coruna
RENAULT - TALISMAN 1. 6 INTENS ENERGY DCI 130 EDC	www.milanuncios.com/renault-de-segunda-mano/renault-talisman-1-6-intens-energy-dci-	14.900 €	año 2016	129.000 kms	diesel	130 cv	4 puertas	automat	la_coruna
PEUGEOT - 3008 STYLE 1. 6 EHD 115 FAP ETG6 BLUE LION	www.milanuncios.com/peugeot-de-segunda-mano/peugeot-3008-style-1-6-ehdi-115-fap-e	9.900 €	año 2015	159.000 kms	diesel	115 cv	5 puertas	automat	la_coruna
AUDI - A3 SPORTBACK 2. 0 TDI 150CV CLEAN ATTRACTION	www.milanuncios.com/audi-de-segunda-mano/audi-a3-sportback-2-0-tdi-150cv-clean-attra	13.900 €	año 2016	159.000 kms	diesel	150 cv	5 puertas	manual	la_coruna
PEUGEOT - 308 5P ACTIVE 1. 4 HDI 68	www.milanuncios.com/peugeot-de-segunda-mano/peugeot-308-5p-active-1-4-bluehdi-120	9.900 €	año 2016	129.000 kms	diesel	120 cv	5 puertas	manual	la_coruna
VOLKSWAGEN - GOLF ADVANCE 1. 6 TDI 105CV BMT	www.milanuncios.com/volkswagen-de-segunda-mano/volkswagen-golf-advance-1-6-tdi-1i	11.900 €	año 2014	119.000 kms	diesel	105 cv	5 puertas	manual	la_coruna
HYUNDAI - I10 1. 1 CLASSIC	www.milanuncios.com/hyundai-de-segunda-mano/hyundai-i10-1-1-classic-298042523.htm	4.900 €	año 2012	79.000 kms	gasolina	69 cv	5 puertas	manual	la_coruna
VOLKSWAGEN - GOLF ADVANCE 1. 6 TDI 105CV BMT	www.milanuncios.com/volkswagen-de-segunda-mano/volkswagen-golf-advance-1-6-tdi-1i	12.900 €	año 2014	60.000 kms	diesel	105 cv	5 puertas	manual	la_coruna
AUDI - A3 1. 6 TDI 105CV S TRONIC ATTRACTION	www.milanuncios.com/audi-de-segunda-mano/audi-a3-1-6-tdi-105cv-s-tronic-attraction-29	10.900 €	año 2013	169.000 kms	diesel	105 cv	3 puertas	automat	la_coruna
VOLKSWAGEN - GOLF 1. 6 TDI 105CV DPF SPORT	www.milanuncios.com/volkswagen-de-segunda-mano/volkswagen-golf-1-6-tdi-105cv-dpf-i	6.900 €	año 2011	229.000 kms	diesel	105 cv	3 puertas	manual	la_coruna
VOLKSWAGEN - GOLF SPORT 2. 0 TDI 150CV BMT DSG	www.milanuncios.com/volkswagen-de-segunda-mano/volkswagen-golf-sport-2-0-tdi-150ci	13.900 €	año 2016	139.000 kms	diesel	150 cv	5 puertas	automat	la_coruna
VOLKSWAGEN - POLO 1. 6 TDI 90CV DSG ADVANCE	www.milanuncios.com/volkswagen-de-segunda-mano/volkswagen-polo-1-6-tdi-90cv-dsg-a	10.900 €	año 2013	119.000 kms	diesel	90 cv	5 puertas	automat	la_coruna
AUDI - A3 SPORTBACK 1. 6 TDI 110CV CLEAN ATTRACTION	www.milanuncios.com/audi-de-segunda-mano/audi-a3-sportback-1-6-tdi-110cv-clean-attra	13.900 €	año 2015	139.000 kms	diesel	110 cv	5 puertas	manual	la_coruna
AUDI - Q3 2. 0 TDI 140CV AMBIENTE	www.milanuncios.com/audi-de-segunda-mano/audi-q3-2-0-tdi-140cv-ambiente-299826780	13.900 €	año 2013	209.000 kms	diesel	140 cv	5 puertas	manual	la_coruna
PEUGEOT - 308 5P ACCESS 1. 6 HDI 92 FAP	www.milanuncios.com/peugeot-de-segunda-mano/peugeot-308-5p-access-1-6-hdi-92-fap-	5.900 €	año 2011	139.000 kms	diesel	92 cv	5 puertas	manual	la_coruna

Tras unos pequeños arreglos que nos facilitarán la realización de las gráficas, el dataset obtenido para poder obtener conclusiones es el siguiente:

BMW - SERIE 1 116i	www.milanuncios.com	13.490	2014	114.000	gasolina	136 cv	3 puertas	manual	madrid	2014	3	BMW
SEAT - LEON ST 1. 6 TDI 105CV STSP STYLE	www.milanuncios.com	9.900	2014	129.073	diesel	105 cv	5 puertas	manual	tarragona	2014	5	SEAT
VOLKSWAGEN - GOLF 1. 6 TDI 105CV DPF SPORT	www.milanuncios.com	9.500	2011	99.271	diesel	105 cv	3 puertas	manual	tarragona	2011	3	VOLKSWAGEN
FORD - FOCUS 1. 6 TDCI 115CV TREND	www.milanuncios.com	10.900	2015	70.102	diesel	115 cv	5 puertas	manual	tarragona	2015	5	FORD
JEEP - GRAND CHEROKEE 3. 0 V6 CRD LAREDO	www.milanuncios.com	13.500	2010	119.036	diesel	218 cv	5 puertas	automat	tarragona	2010	5	JEEP
FORD - ECOSPORT 1. 5 TDCI 70KW 95CV TITANIUM S	www.milanuncios.com	14.900	2017	24.351	diesel	95 cv	5 puertas	manual	tarragona	2017	5	FORD
PEUGEOT - 208 5P ACTIVE 1. 4 HDI 68	www.milanuncios.com	7.900	2014	67.443	diesel	68 cv	5 puertas	manual	tarragona	2014	5	PEUGEOT
BMW - SERIE 1 116d EFFICIENTDYNAMICS	www.milanuncios.com	14.900	2014	71.464	diesel	116 cv	3 puertas	manual	tarragona	2014	3	BMW
DACIA - DUSTER AMBIANCE DCI 110 4X2 EU6	www.milanuncios.com	10.900	2015	65.941	diesel	109 cv	5 puertas	manual	tarragona	2015	5	DACIA
VOLKSWAGEN - POLO 1. 4 ADVANCE	www.milanuncios.com	4.999	2012	140.000	gasolina	85 cv	5 puertas	manual	tenerife	2012	5	VOLKSWAGEN
FORD - RANGER 2. 2 TDCI 150CV 4X4 DOBLE CABINA XL	www.milanuncios.com	14.793	2014	69.000	diesel	150 cv	4 puertas	manual	la_coruna	2014	4	FORD
MERCEDES-BENZ - CLASE B B 180 CDI AUT.	www.milanuncios.com	13.900	2013	89.000	diesel	109 cv	5 puertas	automat	la_coruna	2013	5	MERCEDE
VOLKSWAGEN - POLO 1. 6 TDI 90CV DSG ADVANCE	www.milanuncios.com	10.900	2013	119.000	diesel	90 cv	5 puertas	automat	la_coruna	2013	5	VOLKSWAGEN
AUDI - A3 SPORTBACK 1. 6 TDI 110CV CLEAN ATTRACTION	www.milanuncios.com	13.900	2015	139.000	diesel	110 cv	5 puertas	manual	la_coruna	2015	5	AUDI
AUDI - Q3 2. 0 TDI 140CV AMBIENTE	www.milanuncios.com	13.900	2013	209.000	diesel	140 cv	5 puertas	manual	la_coruna	2013	5	AUDI
PEUGEOT - 308 5P ACCESS 1. 6 HDI 92 FAP	www.milanuncios.com	5.900	2011	139.000	diesel	92 cv	5 puertas	manual	la_coruna	2011	5	PEUGEOT
CITROEN - C4 C4 BLUEHDI 120 6V FEEL EDITION	www.milanuncios.com	8.900	2016	140.000	diesel	120 cv	5 puertas	manual	la_coruna	2016	5	CITROEN
RENAULT - TALISMAN 1. 6 INTENS ENERGY DCI 130 EDC	www.milanuncios.com	14.900	2016	129.000	diesel	130 cv	4 puertas	automat	la_coruna	2016	4	RENAULT
PEUGEOT - 3008 STYLE 1. 6 EHD 115 FAP ETG6 BLUE LION	www.milanuncios.com	9.900	2015	159.000	diesel	115 cv	5 puertas	automat	la_coruna	2015	5	PEUGEOT
AUDI - A3 SPORTBACK 2. 0 TDI 150CV CLEAN ATTRACTION	www.milanuncios.com	13.900	2016	159.000	diesel	150 cv	5 puertas	manual	la_coruna	2016	5	AUDI
PEUGEOT - 308 5P ACTIVE 1. 6 BLUEHDI 120	www.milanuncios.com	9.900	2016	129.000	diesel	120 cv	5 puertas	manual	la_coruna	2016	5	PEUGEOT
VOLKSWAGEN - GOLF ADVANCE 1. 6 TDI 105CV BMT	www.milanuncios.com	11.900	2014	119.000	diesel	105 cv	5 puertas	manual	la_coruna	2014	5	VOLKSWAGEN
HYUNDAI - I10 1. 1 CLASSIC	www.milanuncios.com	4.900	2012	79.000	gasolina	69 cv	5 puertas	manual	la_coruna	2012	5	HYUNDAI
VOLKSWAGEN - GOLF ADVANCE 1. 6 TDI 105CV BMT	www.milanuncios.com	12.900	2014	60.000	diesel	105 cv	5 puertas	manual	la_coruna	2014	5	VOLKSWAGEN
AUDI - A3 1. 6 TDI 105CV S TRONIC ATTRACTION	www.milanuncios.com	10.900	2013	169.000	diesel	105 cv	3 puertas	automat	la_coruna	2013	3	AUDI

Con un breve análisis mediante R-Studio obtenemos la siguiente información estadística del dataset:

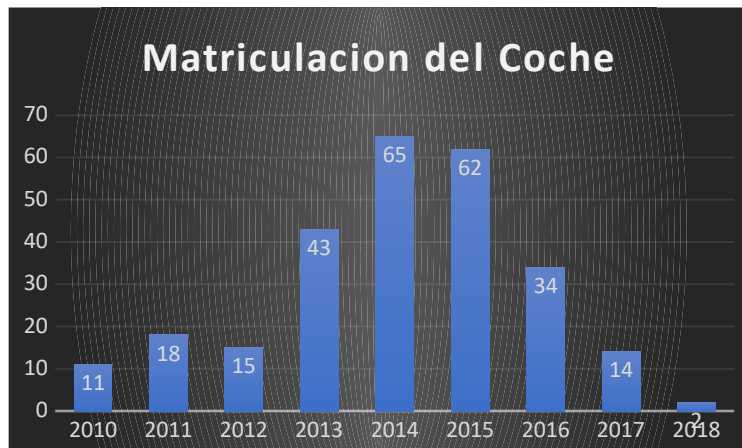
```
## 'data.frame': 264 obs. of 10 variables:
## $ Modelo : chr "RENAULT - SPORT TOURER MEGANE, 1, 2 TCE" "VOLKSWAGEN - POLO EDITION 1. 4 TDI 55KW
75CV BMT" "AUDI - A1 SPORTBACK 1. 6 TDI 90CV ATTRACTED" "BMW - SERIE 1 116I" ...
## $ Link : chr "www.milanuncios.com/renault-de-segunda-mano/renault-sport-tourer-megane-1-2-tce-29
7965161.htm" "www.milanuncios.com/volkswagen-de-segunda-mano/volkswagen-polo-edition-1-4-tdi-55kw75cv-bmt-28726
7267.htm" "www.milanuncios.com/audi-de-segunda-mano/audi-a1-sportback-1-6-tdi-90cv-attracted-274834236.htm" "ww
w.milanuncios.com/bmw-de-segunda-mano/bmw-serie-1-116i-280352763.htm" ...
## $ Precio : num 6500 9500 12900 13490 9900 ...
## $ A_Matriculacion: num 2013 2016 2013 2014 2014 ...
## $ Kms : num 115000 30711 47446 114000 129073 ...
## $ Combustible : Factor w/ 2 levels "diesel","gasolina": 2 1 1 2 1 1 1 1 1 1 ...
## $ Potencia : num 115 75 90 136 105 105 115 218 95 68 ...
## $ Puertas : num 5 5 5 3 5 3 5 5 5 ...
## $ Cambio : Factor w/ 2 levels "automat","manual": 2 2 2 2 2 2 2 1 2 2 ...
## $ Provincia : chr "balears" "tarragona" "tarragona" "madrid" ...
```

```
summary(cochesDS)
```

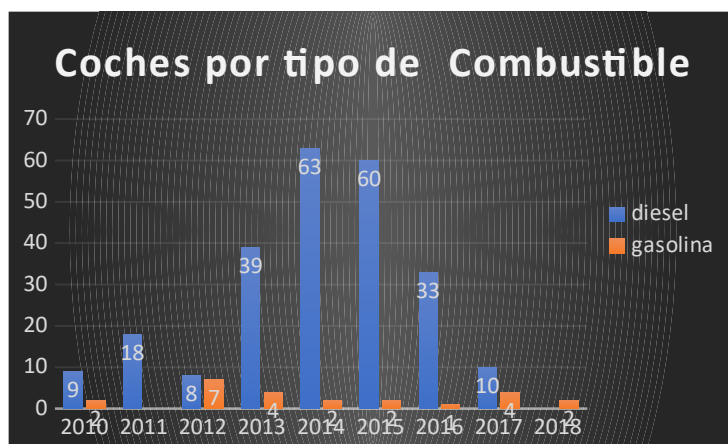
```
##      Modelo      Link      Precio      A_Matriculacion
## Length:264      Length:264      Min.   : 4900      Min.   :2010
## Class :character Class :character 1st Qu.: 7900      1st Qu.:2013
## Mode  :character Mode  :character Median :10000      Median :2014
##                                     Mean  :10352      Mean  :2014
##                                     3rd Qu.:13490      3rd Qu.:2015
##                                     Max.   :14900      Max.   :2018
##      Kms      Combustible      Potencia      Puertas
## Min.   : 6758      diesel :240      Min.   : 65.0      Min.   :3.000
## 1st Qu.: 89000      gasolina: 24      1st Qu.: 95.0      1st Qu.:5.000
## Median :129000                                     Median :110.0      Median :5.000
## Mean   :121733                                     Mean   :111.2      Mean   :4.773
## 3rd Qu.:150000                                     3rd Qu.:120.0      3rd Qu.:5.000
## Max.   :245000                                     Max.   :218.0      Max.   :5.000
##      Cambio      Provincia
## automat: 37      Length:264
## manual :227      Class :character
##                                     Mode  :character
```

A continuación se muestran unos sencillos gráficos que nos permitirán utilizar los datos y obtener interesantes conclusiones. El objetivo de las siguientes gráficos es mostrar la utilidad del Dataset

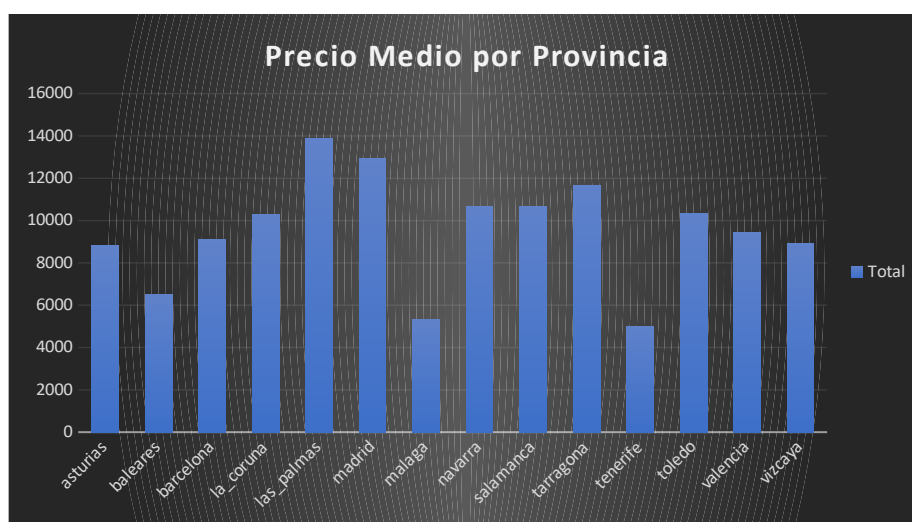
En este primer gráfico podemos observar como la gran mayoría de coches obtenidos en la consulta tienen entre 4 y 6 años.



Por tipo de combustible en el que la gran mayoría de coches que se venden por diesel.

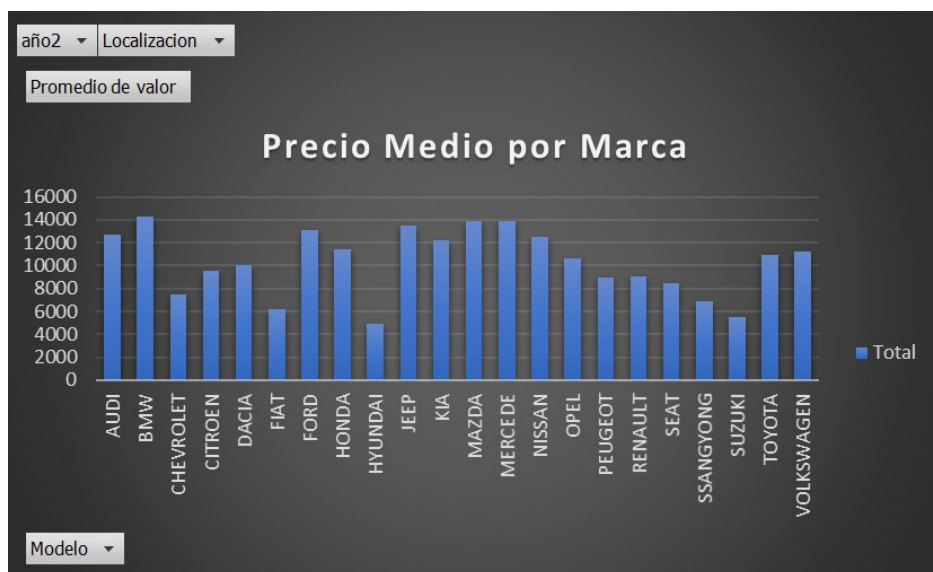


En un gráfico algo más complejo con más variables, podremos obtener en que provincia se venden los coches más baratos. A la vista del gráfico se podría concluir que en Málaga y Tenerife es donde más barato se venden los coches.



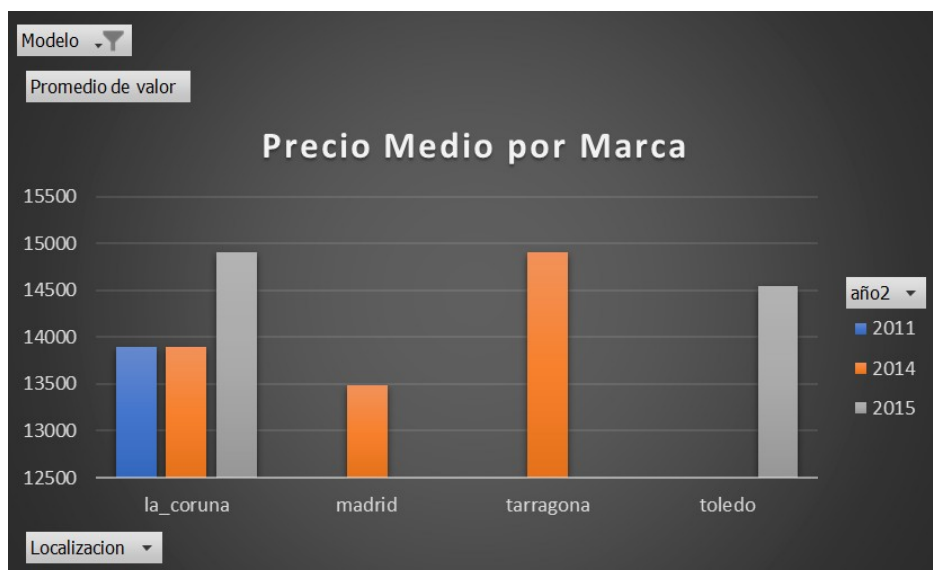


También podremos obtener fácilmente, cuáles son las marcas que se venden a precios mas económicos o más elevados.



Supongamos un ejemplo algo más complejo: Si queremos comprar un BMW, podríamos obtener fácilmente qué provincias son las que tienen mejores precios .

Una rápida conclusión, nos llevaría a pensar que los BMW que se venden en Madrid y que son del 2014 tiene el mejor precio medio por coche.



## 6. Contenido. Explicar los campos del Dataset Obtenido.

El dataset obtenido tiene la siguiente información:

- MODELO → Modelo y marca del vehículo.
- LINK → El link para acceder directamente a la web donde está todo el contenido.
- PRECIOS → El precio del vehículo
- AÑO MATRICULACIÓN.
- KILOMETROS
- TIPO DE COMBUSTIBLES → Gasoil, diesel, etc.
- POTENCIA → en CV
- PUERTAS → 3 o 5 Puertas
- CAMBIO → Manual o automático.
- LOCALIZACION → Provincia donde se encuentra localizado el vehículo

### 6.1. VALIDEZ DE LOS DATOS

No somos expertos en el negocio de la compra-venta de vehículos usados, pero parece necesario que los datos deben ser refrescados con bastante frecuencia, aunque algunas métricas puedan ser bastante estables. La compra-venta de vehículos está en constante evolución y semanalmente se traduce en cientos de operaciones que podrían modificar nuestras conclusiones.

Un ejemplo podría ser el fenómeno reciente en el que los vehículos diésel han sido gravados con nuevos impuestos, circunstancia que probablemente haya provocado que su precio y demanda se reduzcan considerablemente.

## 7. Agradecimientos. Presentar al propietario del conjunto de datos.

Como no podía ser de otra manera, agradecemos a todo el equipo de milanuncios.com su extraordinario trabajo.

El propietario al que debemos el éxito de nuestro trabajo es:

https://whois.icann.org/en/lookup?name=milanuncios.com

milanuncios.com

Al enviar cualquier información personal, acepto que los datos personales se procesarán de acuerdo con la [Política de privacidad de](#) la ICANN , y acepto cumplir con los [Términos de servicio](#) del sitio web .

*Mostrando resultados para: milanuncios.com*  
Consulta original: milanuncios.com

Información del contacto		
<b>Contacto del Registrante</b> Nombre: En nombre del propietario de milanuncios.com Organización: Whois Privacy Service Dirección postal: PO Box 81226, Seattle WA 98108-1226 US Teléfono: +1.2065771368 Ext: Fax: Fax Ext: Correo electrónico: owner-7041060@milanuncios.com.whois.privacyservice.org	<b>Contacto de administrador</b> Nombre: En nombre de milanuncios.com Contacto administrativo. Organización: Whois Privacy Service Dirección postal: PO Box 81226, Seattle WA 98108-1226 US Teléfono: +1.2065771368 Ext: Fax: Fax Ext: Correo electrónico: admin-7041060@milanuncios.com.whois.privacyservice.org	<b>Contacto técnico</b> Nombre: En nombre de milanuncios.com Contacto técnico. Organización: Whois Privacy Service Dirección postal: PO Box 81226, Seattle WA 98108-1226 US Teléfono: +1.2065771368 Ext: Fax: Fax Ext: Correo electrónico: tech-7041060@milanuncios.com.whois.privacyservice.org

[Presentar una queja por WHOIS](#)  
[Formulario de queja por inexactitud de](#)  
[WHOIS Formulario de queja por](#)  
[servicio de WHOIS](#)

[Preguntas frecuentes sobre el cumplimiento de WHOIS](#)

Los propietarios de la web no están identificados. Hemos utilizado la herramientas *whois* para intentar averiguar quien son los propietarios pero su identidad está oculta.

Agradecemos también a [coches.net](#) la disponibilidad y la publicación del estudio realizado en mayo de 2016 y que se muestra en el siguiente punto y que amplía nuestras expectativas iniciales.

## 8. Inspiración. Explique por qué es interesante este conjunto de datos y que preguntas puede responder.

Una de las inversiones más complejas y difíciles que tiene cualquier persona es la adquisición de un vehículo. La búsqueda de uno que cubra todas nuestras necesidades y expectativas suele ser cuestión de semanas o incluso meses, visitando diferentes Webs de coches de ocasión, concesionarios, mercado particular, etc.

El proyecto puede ser interesante tanto del lado del comprador como por parte del vendedor

Por parte del comprador, a través de un proyecto de Web Scraping podemos obtener información precisa de todas esas Webs de manera rápida y concreta y que sin duda nos ahorrará decenas de horas de búsqueda por las diversas webs que se dedican a la compraventa de vehículos.

No obstante creemos que la información obtenida es mucho más interesante desde el punto de vista del vendedor puesto que con los obtenidos y obteniendo los datos de coches vendidos podríamos establecer relaciones entre las ventas y las compras en las diferentes provincias de España. Podríamos contestar, como ejemplos, a las siguientes preguntas

- Que coches se venden más en cada una de las provincias
- Qué tipo de coches se venden más en cada una de las provincias (usados, nuevos, etc)
- Que rango de precio tienen los coches vendidos (en cualquier provincia)
- En qué fechas se venden más coches
- Duración media de los stock de los modelos de los diferentes marcas, nuevamente por provincias, etc, etc.
- Que concesionarios tienen precios más competitivos
- Precios medio de los modelos vendidos en cada provincia para comprender si (p.ej.) si la marca Peugeot tiene precios medios más económicos en una provincia que en otra.

Hay inmobiliarias que están comenzando a incorporar a Científicos de Datos para establecer relaciones entre los bienes inmobiliarios y los compradores, de manera que la inmobiliaria puede saber si un piso de 120 m2 con un precio de 145.000 en Getafe será vendido antes de 60 días. La contestación a todas estas preguntas nos podría llevar a sacar conclusiones de si nos interesa adquirir, para después vender, un bien inmobiliario en una determinada ciudad.

De manera análoga, este interesante ejercicio, podría trasladarse a los coches, camiones, etc para tomar decisiones relevantes y optimizar nuestros recursos económicos.

Por supuesto, disponer de este conjunto de datos nos permitirá utilizar técnicas de minería de datos capaces de responder a todas estas preguntas. Este dataset y debido a su claridad es fácilmente manipulable, escalable para poder crear modelos útiles.

Un estudio análogo al que nosotros queremos hacer, data de mayo de 2016 en el que un competidor director milanuncios realizó un interesante estudio sobre la compra-venta de coches. En concreto, [coches.net](https://www.coches.net) realizó un gran trabajo y obtuvo valiosas conclusiones.

Este estudio consta de 46 páginas y está orientado a la parte vendedora, es decir, se realiza un estudio sobre una base de datos de coches vendidos. Como muestra del estudio, se muestran las conclusiones del mismo, indicando el enlace por si se quiere ampliar la información



SCHIBSTED SPAIN

ESTUDIO EL PROCESO DE COMPRA DE UN COCHE 2016

coches.net

- Para los modelos superiores, el sobrecoste compensa ya que hay una mayor presencia de atributos de marca y equipamiento del vehículo
- El comprador considera varias opciones: El comprador de utilitario consideró en gran medida el compacto y entre los compradores de compactos sucedió el mismo fenómeno
- Los compradores de SUV son consideradores de compactos, consideradores de berlinas y monovolúmenes

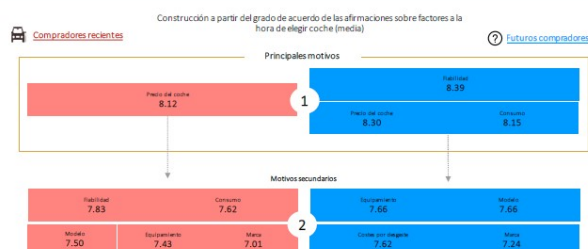
SCHIBSTED SPAIN

ESTUDIO EL PROCESO DE COMPRA DE UN COCHE 2016

coches.net

específicos, idealizando el modelo

Una vez que el proceso de compra está finalizado, el comprador hace una mayor distinción en las razones de compra del coche, siendo el precio el mayor de los motivos



SCHIBSTED SPAIN

ESTUDIO EL PROCESO DE COMPRA DE UN COCHE 2016

coches.net

informarse y los canales online con las fuentes con mayor frecuencia de consulta.

- El concesionario es el canal clave; el hecho de permitir ver el coche, probarlo,...lo hace el canal más relevante. Este contacto con el concesionario se combina con la opinión y consejo de la familia y amigos los cuales se convierten en compañeros en el proceso.
- Una vez que el comprador ha hecho una búsqueda exhaustiva de los diferentes modelos y marcas, el proceso de compra se desencadena con una idea clara de lo que se necesita. Para ir avanzando en este proceso el concesionario y los familiares/amigos junto con el apoyo del online son los que en mayor medida consiguen que el comprador se vaya decidiendo por un modelo.

SCHIBSTED SPAIN

ESTUDIO EL PROCESO DE COMPRA DE UN COCHE 2016

**coches**.net

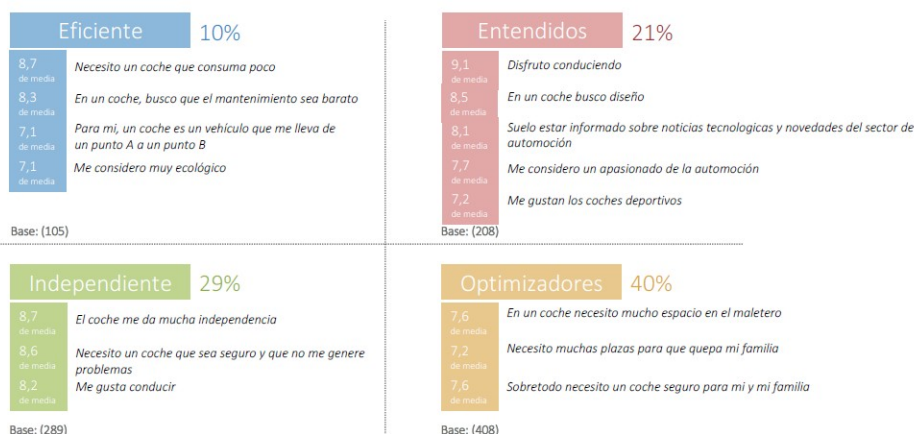
Anteriormente, comentábamos que este estudio era interesante tanto para la parte compradora como para la vendedora y que de los datos se podrían obtener conclusiones utilizando técnicas de modelado. Coches.net uso Clustering para obtener a los 4 principales grupos de compradores. Con esta información, es evidente que un buen vendedor puede canalizar mucho mejor sus ventas.

Este estudio respondió a grandes preguntas como:

- ¿Es vital la importancia de probar el coche antes de comprarlo?
- ¿Se venden más coches Online o más en los Concesionarios?
- ¿Cuáles son las principales barreras de los coches eléctricos para que sean vendidos?
- ¿Cuánto tiempo dura el proceso de compra?

## 02. Principales hallazgos

### GRUPOS



SCHIBSTED SPAIN

ESTUDIO EL PROCESO DE COMPRA DE UN COCHE 2016

**coches**.net

Como aportación final, si fusionáramos las dos BBDD, es decir, los coches vendidos junto con los coches que todavía no lo han sido, podríamos nuevamente obtener clusgter, reglas de clasificación o clasificación que nos permitirían detectar que características tienen los coches menos vendidos utilizando métodos supervisados o no supervisados como los anteriormente reseñados.

Fuente: [https://www.faconauto.com/wp-content/uploads/ponencias/Coches\\_net.pdf](https://www.faconauto.com/wp-content/uploads/ponencias/Coches_net.pdf)

## 9. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección

- Released Under CC0: Public Domain License
- Released Under CC BY-NC-SA 4.0 License
- Released Under CC BY-SA 4.0 License
- Database released under Open Database License, individual contents under Database Contents License
- Other (specified above)
- Unknown License

Cabe recordar inicialmente que estas licencias hacen referencia a las licencias Creative Commons y que cada una de ellas tiene posibilidades para configurarlas y que permite a los autores de los dataset poder decidir de qué manera va a circular en internet su trabajo. Dependiendo de esta configuración, los usuarios podrán publicar, citar, reproducir y crear obras derivadas utilizando el trabajo publicado generalmente bajo ciertas restricciones.

Explicamos rápidamente algunas siglas (llamados módulos oficialmente) de estas licencias:

- CC → Creative Commons (abreviatura)
- BY → requiere la referencia al autor original.
- SA → permite obras derivadas bajo la misma licencia o similar.
- NC → obliga a que la obra no sea utilizada con fines comerciales.

ND → no permite modificar la obra de ninguna manera.  
CCO → Sin derechos reservados.

Estos módulos asimismo se combinan entre si para formar las seis licencias de Creative Commons:

- Atribución (CC – BY)
- Atribución Compartir Igual (CC-BY-SA)
- Atribución No Derivadas (CC-BY-ND)
- Atribución No Comercial (CC-BY-NC)
- Atribución No comercial Compartir Igual ( CC-BY-NC-SA)
- Atribución No comercial No derivadas (CC-BY-NC-ND)

Todas las licencias permiten el derecho fundamental de redistribuir la obra con fines NO comerciales y sin modificaciones, aunque las opciones NC y ND hacen que la obra no sea de libre acuerdo para tal redistribución.

Una vez definidas y recordadas las posibilidades, nuestra licencia tendrá las siguientes características:

- BY, puesto que requerirá al autor original
- NC, puesto que no podremos utilizarla para fines comerciales. Entendamos que milanuncios recibe dinero debido a la publicidad que insertan en su web y que se presupone que ven las personas que la visitan y que obviamente la información que obtenemos de su web es información en propiedad y protegida posiblemente por la LPD.
- ND, puesto que no podremos modificar los registros bajo ningún concepto

Así pues, el trabajo podría tener la licencia más abajo indicada y que no está indicada en las opciones del enunciado:

- Atribución NO comercial NO derivadas (CC-BY-NC-ND)

Una licencia alternativa sería

- Released Under CC BY-NC-SA 4.0 License,

Que permitiría derivar el trabajo bajo la misma licencia y también sin fines comerciales.

Fuente: [https://es.m.wikipedia.org/wiki/Creative\\_Commons](https://es.m.wikipedia.org/wiki/Creative_Commons)

Las licencias Creative Commons están compuestas por cuatro módulos de condiciones:

- *Attribution/Atribución (BY), requiere la referencia al autor original.*
- *Share Alike/Compartir Igual (SA), permite obras derivadas bajo la misma licencia o similar (posterior u otra versión por estar en distinta jurisdicción).*
- *Non-Commercial/No Comercial (NC), obliga a que la obra no sea utilizada con fines comerciales.*
- *No Derivative Works/No Derivadas (ND), no permite modificar la obra de ninguna manera.*



## 10. Contribuciones

Contribuciones	Firma
Investigación previa	David Quiles / Iván López-Baltasar
Redacción de las respuestas	David Quiles / Iván López-Baltasar
Desarrollo código	David Quiles / Iván López-Baltasar