

Práctica 1 (35% nota final)

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

Los datos han sido extraídos de la página web "<https://www.motorflash.com/>" con el objetivo de analizar la variación de precios de coches de segunda mano con el objetivo principal de visualizar las diferencias de precios entre vendedores profesionales y particulares. La web de motor flash ha sido elegida en primera instancia al poder usar el scrapper elegido para el proyecto (Scrapy), en segundo lugar proporciona una amplia colección de diferentes marcas y modelos de coches así como la diferenciación de venta particular y profesional. Otros sitios web de venta de coches de segunda mano tienen sistemas anti-scraping como es por ejemplo "<https://www.coches.net/>". Otro ejemplo de sitio web eliminado fue "<https://www.coches.com/coches-segunda-mano/>" ya que aunque permitía el uso del scraper y anunciaba un número mayor de modelos y marcas al realizar la prueba de scraping se vio que el csv resultante sólo contenía 9000 datos en vez de los 90000 anunciados en la página.

2. Definir un título para el dataset. Elegir un título que sea descriptivo.

El título seleccionado para el dataset es: "motorflash.csv" debido a que el dataset contiene toda la información de cada vehículo proveniente de la web de "<https://www.motorflash.com/>"

3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).

El dataset generado contiene una representación temporal de las características principales de los coches de segunda mano, se compone de 32 campos en los cuales están incluidos todos los datos técnicos del vehículo así como los datos importantes para la realización del estudio como son el "Modelo", "Descripción", "Precio Oferta", "Precio Nuevo" y el "Anunciante"

4. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente

aceleracion0-100	anchura	arrancante	bastida	cambio	cap_maden	capacidad	color	color_ba	combustible	consumo	consumo_comb	consumo_comb	consumo_urbano	consumo_urbano	descripcion	emision_c	garantia	kilometros	longitud	matr	matr	modelo	num_puertas	peso	puerto	puerto	plaza	potencia	precio	precio_oferta	puertas	ubicacion	velocidad
-	1.457 mm	1.504 mm	PISA RETAIL V2	620 mm	Cambio aut	1.309 litros	53 litros	Gris	Casimira	3.9 l/100km	5.6 l/100km	5.3 l/100km	Combinado 5.6 l/100km	Combinado 5.6 l/100km	Combinado 5.6 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 22.900	13400,00	5	puertas	Vallencia	195 km/h
10.9 segun 1.556 mm	1.739 mm	PISA RETAIL B2	538 mm	Cambio aut	1.400 litros	50 litros	Blanco	Casimira	4.1 l/100km	4.8 l/100km	5.9 l/100km	Combinado 4.8 l/100km	Combinado 4.8 l/100km	Combinado 4.8 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.225 kg	1.600 kg	5	81 kW (110 CV, 22.900	13400,00	5	81 kW (110 CV, 22.900	13400,00	5	puertas	Barcelona	188 km/h	
5.8 segun 1.440 mm	1.810 mm	Autobianchi	3.840 mm	Cambio aut	1.400 litros	60 litros	Azul	Manfrin	5.1 l/100km	5.1 l/100km	5.8 l/100km	Combinado 5.1 l/100km	Combinado 5.1 l/100km	Combinado 5.1 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Barcelona	210 km/h	
-	1.457 mm	1.504 mm	PISA RETAIL B2	620 mm	Cambio aut	1.309 litros	53 litros	Blanco	Casimira	3.1 l/100km	4.6 l/100km	4.1 l/100km	Combinado 3.1 l/100km	Combinado 3.1 l/100km	Combinado 3.1 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Barcelona	204 km/h
10.9 segun 1.524 mm	1.841 mm	AUTOFREDA	2.675 mm	Cambio aut	1.402 litros	53 litros	Gris	Casimira	4.5 l/100km	5.1 l/100km	6.0 l/100km	Combinado 4.5 l/100km	Combinado 4.5 l/100km	Combinado 4.5 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.250 kg	1.910 kg	5	86 kW (120 CV, 28.800	20450,00	5	86 kW (120 CV, 28.800	20450,00	5	puertas	Madrid	188 km/h	
11.3 segun 1.542 mm	1.556 mm	Hermanos Cox	1.967 mm	Cambio aut	1.340 litros	53 litros	Negro	Casimira	4.0 l/100km	4.3 l/100km	5.2 l/100km	Combinado 4.0 l/100km	Combinado 4.0 l/100km	Combinado 4.0 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	145 km/h	
11.6 segun 1.556 mm	1.739 mm	PISA RETAIL B2	538 mm	Cambio aut	1.400 litros	50 litros	Negro	Casimira	5.2 l/100km	5.7 l/100km	6.7 l/100km	Combinado 5.2 l/100km	Combinado 5.2 l/100km	Combinado 5.2 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.225 kg	1.600 kg	5	81 kW (110 CV, 22.900	13400,00	5	81 kW (110 CV, 22.900	13400,00	5	puertas	Barcelona	187 km/h	
9.6 segun 1.454 mm	1.772 mm	Autobianchi	3.824 mm	Cambio aut	1.400 litros	53 litros	Blanco	Gris	DAIWEI	4.6 l/100km	5.5 l/100km	7.3 l/100km	Combinado 4.6 l/100km	Combinado 4.6 l/100km	Combinado 4.6 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Gugayana	201 km/h
9.1 segun 1.476 mm	1.628 mm	PISA RETAIL A2	817 mm	Cambio aut	1.506 litros	72 litros	Gris	DAIWEI	3.8 l/100km	4.2 l/100km	5.0 l/100km	Combinado 3.8 l/100km	Combinado 3.8 l/100km	Combinado 3.8 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.120 kg	2.200 kg	5	110 kW (150 CV, 36.300	19000,00	5	110 kW (150 CV, 36.300	19000,00	5	puertas	ComAja	210 km/h	
13.3 segun 1.474 mm	1.749 mm	PISA RETAIL A2	540 mm	Cambio aut	1.922 litros	40 litros	Negro	Casimira	4.2 l/100km	4.2 l/100km	5.2 l/100km	Combinado 4.2 l/100km	Combinado 4.2 l/100km	Combinado 4.2 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	ComAja	160 km/h	
9.8 segun 1.455 mm	1.745 mm	PISA RETAIL A2	538 mm	Cambio aut	1.119 litros	44 litros	Gris	Casimira	3.6 l/100km	3.6 l/100km	5.2 l/100km	Combinado 3.6 l/100km	Combinado 3.6 l/100km	Combinado 3.6 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	ComAja	160 km/h	
9.9 segun 1.435 mm	1.745 mm	PISA RETAIL A2	538 mm	Cambio aut	1.119 litros	44 litros	Blanco	Casimira	3.9 l/100km	5.6 l/100km	5.3 l/100km	Combinado 3.9 l/100km	Combinado 3.9 l/100km	Combinado 3.9 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	America	166 km/h	
-	1.457 mm	1.504 mm	PISA RETAIL B2	620 mm	Cambio aut	1.309 litros	53 litros	Blanco	Casimira	5.1 l/100km	5.1 l/100km	6.6 l/100km	Combinado 5.1 l/100km	Combinado 5.1 l/100km	Combinado 5.1 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Vallencia	196 km/h
10.5 segun 1.644 mm	1.819 mm	PISA RETAIL V2	636 mm	Cambio aut	1.251 litros	60 litros	Blanco	Casimira	4.4 l/100km	5.1 l/100km	5.3 l/100km	Combinado 4.4 l/100km	Combinado 4.4 l/100km	Combinado 4.4 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Vallencia	196 km/h	
-	1.457 mm	1.504 mm	PISA RETAIL V2	620 mm	Cambio aut	1.309 litros	53 litros	Azul	Casimira	3.9 l/100km	5.6 l/100km	5.3 l/100km	Combinado 3.9 l/100km	Combinado 3.9 l/100km	Combinado 3.9 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Vallencia	196 km/h
11.3 segun 1.535 mm	1.810 mm	PISA RETAIL V2	785 mm	Cambio aut	1.355 litros	50 litros	Azul	DAIWEI	4.2 l/100km	4.4 l/100km	4.8 l/100km	Combinado 4.2 l/100km	Combinado 4.2 l/100km	Combinado 4.2 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Vallencia	170 km/h	
10.5 segun 1.624 mm	1.841 mm	PISA RETAIL V2	675 mm	Cambio aut	1.402 litros	53 litros	Blanco	Casimira	4.8 l/100km	5.3 l/100km	6.1 l/100km	Combinado 4.8 l/100km	Combinado 4.8 l/100km	Combinado 4.8 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Vallencia	188 km/h	
10.9 segun 1.440 mm	1.786 mm	Santagel	3.840 mm	Cambio aut	1.200 litros	43 litros	Gris	DAIWEI	3.9 l/100km	4.2 l/100km	4.7 l/100km	Combinado 3.9 l/100km	Combinado 3.9 l/100km	Combinado 3.9 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	202 km/h	
6.6 segun 1.542 mm	1.810 mm	Hermanos Cox	2.840 mm	Cambio aut	1.400 litros	66 litros	Gris	Negro	Casimira	4.4 l/100km	5.3 l/100km	6.8 l/100km	Combinado 4.4 l/100km	Combinado 4.4 l/100km	Combinado 4.4 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	244 km/h
10.0 segun 1.710 mm	1.840 mm	Autobianchi	2.700 mm	Cambio aut	1.620 litros	56 litros	Roso	DAIWEI	4.7 l/100km	5.3 l/100km	6.2 l/100km	Combinado 4.7 l/100km	Combinado 4.7 l/100km	Combinado 4.7 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	198 km/h	
11.3 segun 1.535 mm	1.810 mm	Autobianchi	2.620 mm	Cambio aut	1.402 litros	43 litros	Gris	Casimira	5.9 l/100km	5.9 l/100km	7.4 l/100km	Combinado 5.9 l/100km	Combinado 5.9 l/100km	Combinado 5.9 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	163 km/h	
11.0 segun 1.631 mm	1.836 mm	Autobianchi	2.700 mm	Cambio aut	1.402 litros	43 litros	Blanco	Casimira	4.8 l/100km	5.3 l/100km	6.1 l/100km	Combinado 4.8 l/100km	Combinado 4.8 l/100km	Combinado 4.8 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	198 km/h	
9.8 segun 1.701 mm	1.850 mm	Autobianchi	2.707 mm	Cambio aut	1.420 litros	53 litros	Blanco	DAIWEI	5.6 l/100km	6.6 l/100km	8.6 l/100km	Combinado 5.6 l/100km	Combinado 5.6 l/100km	Combinado 5.6 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	168 km/h	
9.2 segun 1.710 mm	1.840 mm	Autobianchi	2.700 mm	Cambio aut	1.620 litros	56 litros	Roso	DAIWEI	4.1 l/100km	4.6 l/100km	5.4 l/100km	Combinado 4.1 l/100km	Combinado 4.1 l/100km	Combinado 4.1 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	252 km/h	
12.1 segun 1.474 mm	1.734 mm	Autobianchi	2.570 mm	Cambio aut	1.077 litros	50 litros	Negro	DAIWEI	3.4 l/100km	3.9 l/100km	4.7 l/100km	Combinado 3.4 l/100km	Combinado 3.4 l/100km	Combinado 3.4 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	175 km/h	
12.4 segun 1.553 mm	1.866 mm	Hermanos Cox	2.734 mm	Cambio aut	1.309 litros	50 litros	Grisele	Negro	Casimira	3.9 l/100km	4.1 l/100km	4.2 l/100km	Combinado 3.9 l/100km	Combinado 3.9 l/100km	Combinado 3.9 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	164 km/h
12.5 segun 1.640 mm	1.845 mm	Autobianchi	2.702 mm	Cambio aut	1.837 litros	60 litros	Negro	DAIWEI	3.9 l/100km	4.1 l/100km	4.5 l/100km	Combinado 3.9 l/100km	Combinado 3.9 l/100km	Combinado 3.9 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	180 km/h	
10.9 segun 1.460 mm	1.722 mm	Autobianchi	2.406 mm	Cambio aut	1.074 litros	41 litros	Gris	DAIWEI	3.2 l/100km	3.6 l/100km	4.9 l/100km	Combinado 3.2 l/100km	Combinado 3.2 l/100km	Combinado 3.2 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	160 km/h	
12.1 segun 1.500 mm	1.860 mm	Autobianchi	2.702 mm	Cambio aut	1.046 litros	40 litros	Gris	Casimira	4.1 l/100km	4.9 l/100km	5.6 l/100km	Combinado 4.1 l/100km	Combinado 4.1 l/100km	Combinado 4.1 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	175 km/h	
13.3 segun 1.600 mm	1.850 mm	Autobianchi	2.670 mm	Cambio aut	1.503 litros	42 litros	Blanco	DAIWEI	4.2 l/100km	4.6 l/100km	5.4 l/100km	Combinado 4.2 l/100km	Combinado 4.2 l/100km	Combinado 4.2 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00	5	81 kW (110 CV, 25.400	13900,00	5	puertas	Madrid	176 km/h	
11.3 segun 1.580 mm	1.810 mm	Autobianchi	2.670 mm	Cambio aut	1.718 litros	40 litros	Blanco	Casimira	5.9 l/100km	6.2 l/100km	7.4 l/100km	Combinado 5.9 l/100km	Combinado 5.9 l/100km	Combinado 5.9 l/100km	1.105 g/km	24 meses	1500 km	4.250 mm	61-2020	Pragmat 55 Puertas	1.165 kg	1.750 kg	5	81 kW (110 CV, 25.400	13900,00								

20. longitud : longitud del vehículo (mm)
21. cap_maletero : capacidad del maletero (l)
22. batalla : distancia entre ejes del vehículo (mm)
23. peso : peso del vehículo (kd)
24. peso_maximo : peso máximo admitido (kg)
25. num_puertas : número de puertas del vehículo
26. capacidad_dep : capacidad del depósito (l)
27. velocidad_max : velocidad máxima (km/h)
28. aceleracion : aceleracion maxima (seg)
29. emission_co2 : nivel de emisión de dióxido de carbono (gr/km)
30. consum_carretera : consumo en carretera (l/100km)
31. consum_urbano : consumo urbano (l/100km).
32. consum_comb_mix : consumo combinado mixto (l/100km).
33. num_cilindros : número de cilindros
34. cilindrada : cilindrada del vehículo (cm3)
35. par_motor : par motor (Nm)
36. est_emisiones : estándar de emisiones
37. traccion : tracción del vehículo.
38. neumaticos_del : medidas de los neumáticos delanteros
39. neumaticos_tra : medidas de los neumáticos traseros.

El periodo de matriculación de los vehículos anunciados va desde 1955 hasta la actualidad. Los datos se han recogido mediante la librería scrapy. Primero hemos revisado el tamaño de la página para ver que dispone de suficientes datos, también el fichero robots.txt para ver que las limitaciones y restricciones a los robots de scraping no nos afectan.

Mediante reglas definimos el comportamiento para rastrear el sitio web, una regla para la paginación con la que recorreremos todas las páginas y otra regla para acceder a todos los anuncios de cada página. Dentro de cada anuncio está toda la información que almacenamos.

Finalmente realizamos el mapeo mediante xpath y css de la localización de los campos en el dom del html con el objeto serializado que genera el fichero de salida csv.

El tiempo aproximado de realización de web scraping en toda la web fue de unas 4 horas aproximadamente, obteniendo un total de 50656 registros por 39 columnas de características.

Se ha tenido en cuenta las restricciones que algunas páginas web realizan sobre los robots para evitar el bloqueo de los mismos, pero en este caso concreto no fue necesario.

6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).

Los datos han sido recopilados desde la página web de anuncios de coches de segunda mano <https://www.motorflash.com/>. Para ello, se ha hecho uso del lenguaje de programación python y de técnicas de web scraping para extraer información alojada en las páginas HTML. Agradecemos a la compañía de motorflash por poner a disposición de los usuarios dichos datos de manera pública en su página web así como a todos los concesionarios de vehículos que suben los datos de sus vehículos al sitio web.

7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.

El presente conjunto de datos permite realizar un análisis estadístico para identificar patrones y tendencias en los datos.

Identificar las ciudades que más coches tienen publicados y los modelos mayoritarios en cada una de ellas.

Analizar si el precio de un coche varía por provincias y cuales son los anunciantes con los precios más caros o más baratos.

Detectar los modelos de coches que más se devalúan con respecto a su precio y los que menos.

Identificar los modelos que mejor relación calidad precio tienen en base a sus características. Predecir el precio de un coche de segunda mano dadas sus características para tener una estimación del precio en el mercado de segunda mano.

Clasificar coches por características, en base al precio, potencia, velocidad, etc. para agrupar coches por gama alta, media, baja.

Son algunas de las preguntas que podríamos responder mediante un análisis de datos realizado sobre este conjunto de datos, que podría utilizarse en diversos ámbitos. Uno de ellos podría ser en el periodístico con fines informativos sobre el mercado de segunda mano. Así mismo podría ser de gran utilidad en el campo de la minería de datos, a la hora de elaborar modelos predictivos como los mencionados anteriormente.

8. 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

Teniendo en cuenta que nuestro proyecto es meramente académico, consideramos que debe ser público para que pueda ser utilizado por cualquier persona incluido para una finalidad comercial, ya que el origen de estos datos tiene fines comerciales.

La licencia escogida para la publicación de este conjunto de datos ha sido CC BY-SA 4.0 License. Los motivos que han llevado a la elección de esta licencia va acorde con la idoneidad del trabajo realizado:

- *Se debe reconocer adecuadamente la autoría e indicar si se han realizado cambios.* De manera que se reconoce el trabajo ajeno y permite realizar aportaciones sobre el mismo
- *Se debe compartir igual* si se mezcla, transforma o crea a partir del material y deberá difundir las contribuciones bajo la misma licencia que el original.
- *Se puede utilizar para una finalidad comercial.* Aportaría valor a la empresa que utilice estos datos para generar análisis de calidad reconociendo la autoría original.

9. Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

El código se ha realizado íntegramente en python y se ha sincronizado mediante el versionado de git.

Adicionalmente se tiene la carpeta docs, con este mismo documento, el fichero readme.md que contiene una pagina de bienvenida y descripcion del proyecto.

La url de github donde se aloja el proyecto es

<https://github.com/CarlosRea/MotorflashScraper>

10. Dataset. Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

En Zenodo se ha añadido una breve descripción del proyecto que incluye el enlace al repositorio Github para quien desee ampliar información.

El dataset publicado se puede consultar en el siguiente enlace

<https://doi.org/10.5281/zenodo.4262983>

- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
- Simon Munzert, Christian Rubba, Peter Meißner, Dominic Nyhuis. (2015). Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining. John Wiley & Sons.
- Tutorial de Github <https://guides.github.com/activities/hello-world>.

Criterios de valoración

Todos los apartados son obligatorios. La ponderación de los ejercicios es la siguiente:

- Los apartados 1, 2, 3 y 4 valen 0,25 puntos cada uno.
- Los apartados 5, 6, 7 y 8 valen 1 punto cada uno.
- Los apartados 9 y 10 valen 2,5 puntos cada uno.

Otros criterios que se tomarán en cuenta para la evaluación son:

- Idoneidad de las respuestas (deberán ser claras y completas).
- Complejidad del sitio web elegido para la extracción.
- Síntesis y claridad, a través del uso de comentarios, del código resultante.
- Presentación adecuada de los datos.
- Organización y claridad de los documentos de entrega final.
- Completitud de los documentos requeridos para la entrega final.

Formato y fecha de entrega

Durante la semana del 26 al 30 de octubre, el grupo podrá entregar al profesor una entrega parcial opcional. Esta entrega parcial es muy recomendable para recibir asesoramiento sobre la práctica y verificar que la dirección tomada es la correcta. Se entregarán comentarios a los estudiantes que hayan efectuado la entrega parcial pero no contará para la nota de la práctica. En la entrega parcial los estudiantes deberán entregar por correo electrónico, al profesor encargado del aula, el enlace al repositorio Github con lo que hayan avanzado.

En referente a la entrega final, hay que entregar un único fichero que contenga el enlace a Github donde haya:

1. Una Wiki donde estén los nombres de los componentes del grupo y una descripción de los ficheros.
2. Un documento PDF con las respuestas a las preguntas y los nombres de los componentes del grupo. Además, al final del documento, debe aparecer la siguiente tabla de contribuciones al trabajo, la cual debe firmar cada integrante

del grupo con sus iniciales. Las iniciales representan la confirmación por parte del grupo que el integrante ha participado en dicho apartado. Todos los integrantes deben participar en cada apartado, por lo que, idealmente, los apartados deberían estar firmados por todos los integrantes.

Contribuciones	Firma
Investigación previa	CR, YN
Redacción de las respuestas	CR, YN
Desarrollo código	CR, YN

3. Una carpeta con el código Python o R generado para obtener los datos.
4. El DOI a los datos.

Este documento de la entrega final se tiene que entregar en el espacio de Entrega y Registro de AC del aula antes de las **23:59 del día 9 de noviembre**. No se aceptarán entregas fuera de plazo.