

# Práctica 1: Web Scraping

## 1. Contexto

El conjunto de datos generado en esta actividad reúne diferentes características técnicas de vehículos de alta gama, concretamente, de los modelos comerciales de las marcas Porsche y Lamborghini.

## 2. Dataset: título y contenido

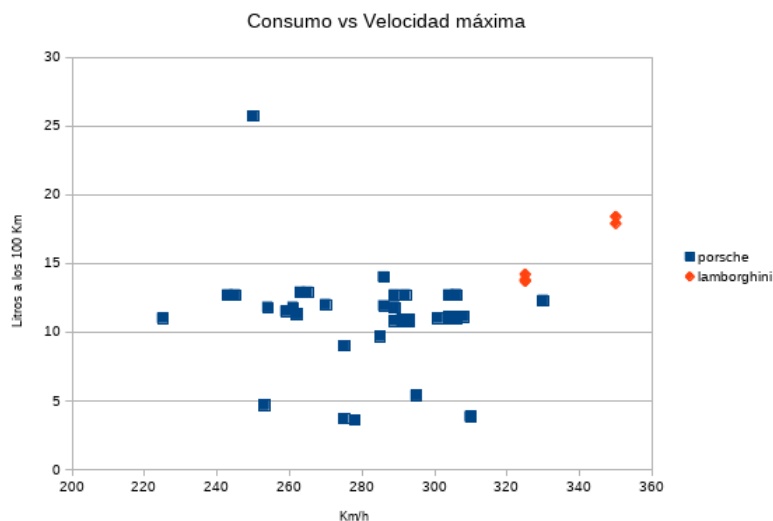
El dataset recibe el nombre de **carsTecnicFeatures** y consta de los siguientes atributos o variables:

- **marca:** se han analizado dos marcas Porsche y Lamborghini
- **modelo:** modelo de coche analizado dentro de las marcas anteriores
- **altura:** dimensiones en altura del vehículo desde el suelo
- **anchoBase:** dimensiones en anchura del vehículo excluidos los espejos retrovisores
- **anchoTotal:** dimensiones en anchura del vehículo incluidos los espejos retrovisores
- **longitud:** dimensiones de largo del vehículo
- **peso:** peso en seco o tara
- **cilindrada:** cantidad de cm<sup>3</sup> del motor
- **consumo:** litros a los 100km de consumo medio o combinado
- **potencia:** CV de potencia de motor
- **aceleración:** segundos que tarda en pasar de 0-100km
- **velocidadMax:** velocidad máxima que puede alcanzar en km/h
- **precio:** precio base o desde

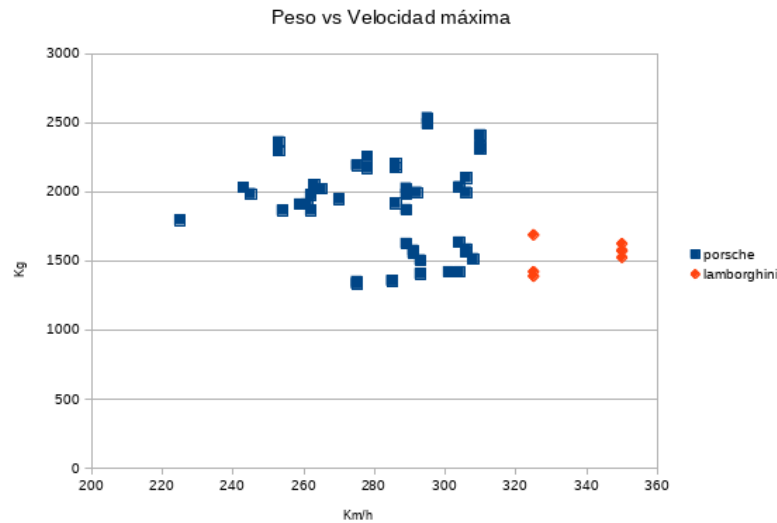
En cada uno de los registros del dataset se almacena el valor de determinadas características técnicas, correspondiéndose con los atributos anteriores, para cada modelo de vehículo analizado.

## 3. Representación gráfica del dataset

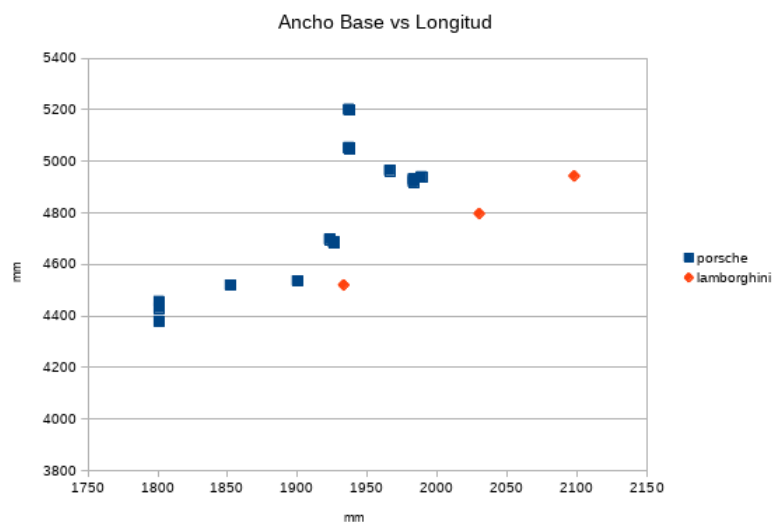
Para ilustrar los diferentes valores del dataset, vamos a mostrar algunas imágenes en las que se comparan gráficamente un atributo frente a otro. Por ejemplo, podemos comparar la velocidad máxima contra el consumo de los diferentes vehículos para tener una idea de lo optimizado que puede estar el motor.



Así mismo, podríamos comparar la velocidad máxima contra el peso para visualizar la potencia del motor en velocidad punta.



Finalmente, para ilustrar el área del coche, comparamos el ancho base, es decir con los retrovisores plegados contra la longitud.



#### 4. Agradecimientos

Los datos han sido recolectados de las páginas webs oficiales de **Porsche** (<https://www.porsche.com/spain>) y de **Lamborghini** (<https://www.lamborghini.com/es-en>). En estas páginas aparecen los diferentes modelos que venden estas marcas, junto con sus características técnicas generales y se pueden configurar algunos otros parámetros de los vehículos.

Para la extracción de los datos se ha hecho uso de técnicas de Web Scraping y Crawling, para navegar a través de las distintas urls, y extraer la información alojada en las páginas HTML; y del lenguaje de programación Python, junto con **Selenium**.

## 5. Inspiración

A grandes rasgos, el conjunto de datos puede utilizarse para realizar un estudio de mercado. Antes de decidirse a lanzar un nuevo producto, resulta conveniente realizar un análisis de los productos, de similares características, que ya están en el mercado con la intención de saber quienes serán tus competidores y de las posibilidades que existen de competir con ellos o no. Desde este punto de vista, las características técnicas de un vehículo son datos objetivos entre los cuales se pueden hacer comparativas.

También podrían ser de utilidad si se busca la existencia o no de correlación entre variables. Por ejemplo, si hay alguna relación entre las dimensiones del vehículo y el consumo; o entre la potencia, cilindrada y aceleración, etc.

## 6. Licencia

La licencia escogida para la publicación de este conjunto de datos ha sido **CC BY-NC-SA 4.0 License**. El principal motivo que nos ha llevado a elegir esta licencia están relacionado con la propiedad de No Comercial. Consideramos que en la actividad en la que se han obtenido el conjunto de datos ha sido con fines educativos y debe seguir utilizándose en la misma línea, tanto la obra original como las derivadas.

## 7. Código fuente

Tanto el código fuente escrito para la extracción de datos como el dataset generado pueden ser accedidos a través del siguiente enlace: <https://github.com/clianez/practica1UOC>

## 8. Publicación del dataset

El dataset ha sido publicado en Zenodo, puede accederse a través del siguiente enlace <https://zenodo.org/record/3746549#.Xo9GmcgzY2w>, y el DOI asignado ha sido **10.5281/zenodo.3746549**.

## 9. Recursos

- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
- Mitchel, R. (2015). Web Scraping with Python: Collecting Data from the Modern Web. O'Reilly Media, Inc. Chapter 1. Your First Web Scraper.
- Muthukadan, B (2018). Selenium with Python. <https://selenium-python.readthedocs.io/>

Contribuciones	Firma
Investigación previa	CLL, MPM
Redacción de las respuestas	CLL, MPM
Desarrollo código	CLL, MPM