

1. Contexto.

El contexto en el que se ha desarrollado esta actividad es la asignatura Tipología y ciclo de la vida de datos del máster de ciencia de datos de la Universidad Oberta de Cataluña que se realiza en el segundo semestre del curso 2018-2019.

La práctica elaborada consiste en un trabajo de web-scraping donde hay que utilizar la información de una página web para generar un conjunto de datos. El sitio web elegido proporciona una base de datos estructurada con información de coches pública lo que hace que el conjunto de datos sea de gran interés para realizar esta práctica.

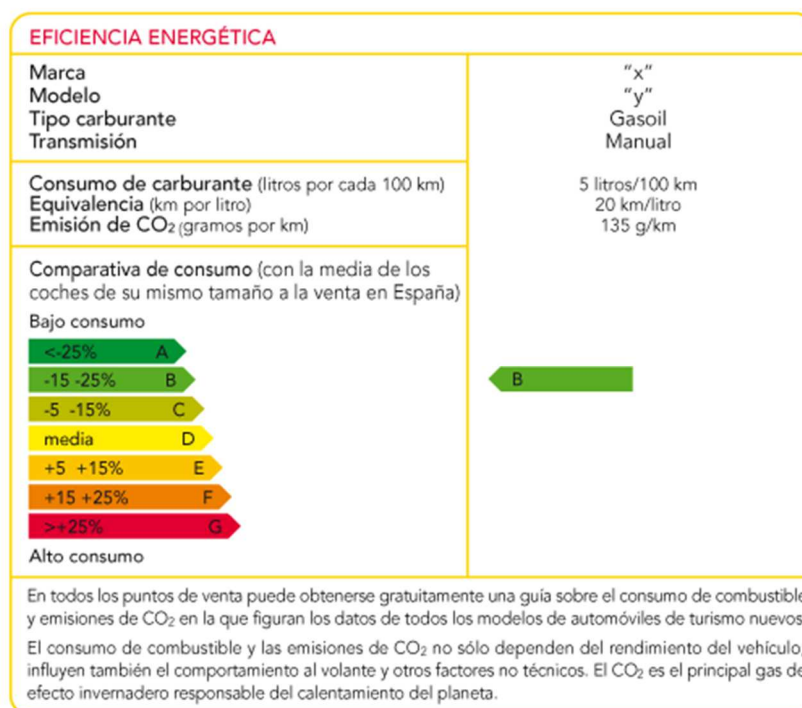
2. Definir un título para el dataset.

ConsumoYEmisionesDeTurismos.

3. Descripción del dataset.

El conjunto de datos que se generó en este trabajo contiene detalles sobre el consumo de carburante y emisiones de CO₂ de los turismos nuevos que se comercializan en España. Esta información se ha recogido de la página web <http://coches.idae.es> que forma parte programa del Instituto para la Diversificación y Ahorro de Energía (IDAE), organismo adscrito al Ministerio para la Transición Ecológica.

4. Representación gráfica.



5. Contenido.

El dataset creado recoge una serie de información correspondiente a las mediciones de consumo de combustible y sobre las emisiones de CO₂ según el ciclo WLTP del coche que se haya seleccionado. Cada fila corresponde a un modelo diferente de la marca seleccionada, y cada modelo contiene está compuesta por los siguientes campos:

- **Modelo:** Nombre del modelo, incluye la marca, los caballos y la potencia térmica.
- **Clasificación energética:** valor correspondiente a la clasificación por consumo relativo. Puede tomar los siguientes valores: *A, B, C, D, E, F* y *G*; donde *A* corresponde a un consumo mínimo y *G* a un alto consumo.
- **Consumo mínimo (l/100Km):** valor correspondiente al consumo mínimo del vehículo en litros por cada 100km.
- **Consumo máximo (l/100Km):** valor correspondiente al consumo máximo del vehículo en litros por cada 100km.
- **Emisiones mínimas (gCO₂/km):** valor correspondiente a la emisión mínima de dióxido de carbono (CO₂) que emite el vehículo por km.
- **Emisiones máximas (gCO₂/km):** valor correspondiente a la emisión máxima de dióxido de carbono (CO₂) que emite el vehículo por km.

Los datos ofrecidos por el Instituto para la Diversificación y ahorro de la Energía (IDAE) son elaborados anualmente, incluyendo todos los modelos de turismos nuevos existentes en la fecha de publicación de la actualización, puestos en venta en los Estados miembros, clasificados por marcas y por orden alfabético.

Para la creación de nuestro dataset, primero hemos hecho una llamada a la web indicando la marca del coche del que queremos obtener la información. Los datos obtenidos de esta llamada son guardados en formato json; recorreremos este json para obtener los datos que nos interesan y que iremos guardando en un dataframe. Una vez tenemos toda la información en el dataframe, lo convertiremos en un fichero .csv. Para ello se ha hecho uso del lenguaje de programación Python y utilizado técnicas de Web Scraping.

6. Agradecimientos.

Este conjunto de datos ha sido creado por la Directiva y del Real Decreto, IDAE con el apoyo del programa SAVE de la DG TREN (Transporte y Energía) de la Comisión Europea. Los datos correspondientes al consumo, emisión y otros datos técnicos han sido facilitados por ANFAC (Asociación Española de Fabricantes de Automóviles y Camiones), ANIACAM (Asociación Nacional de Importadores de Automóviles, Camiones, Autobuses y Motocicletas) e IEA (Instituto de Estudios de Automoción).

Estos conjuntos de datos están disponibles gratuitamente para los consumidores de nuevos coches, quienes podrán solicitarla en el punto de venta o ante un organismo designado en cada Estado miembro.

7. Inspiración.

Este conjunto de datos nos informa del consumo de carburante y las emisiones de CO₂ de los turismos nuevos, es interesante que tanto las personas físicas como jurídicas tengan disponible esta información para poder considerar la adquisición de los vehículos más eficientes energéticamente.

El consumo de combustible y emisiones de CO₂ específicos de los turismos puede influir en la decisión del consumidor en favor de los automóviles que consuman menos combustible y por lo tanto emitan menos CO₂, impulsando de ese modo a los fabricantes a hacer lo necesario para reducir el consumo de los automóviles.

Entre otras aplicaciones, este conjunto de datos podría ser útil para empresas que tienen la necesidad de adquirir una flota de vehículos, a quienes les podría interesar aquellos vehículos de menor consumo para poder ahorrar en costes.

8. Licencia

Autorizamos el uso de toda la información de nuestro dataset mediante **CC BY-SA 4.0 License**. Esta licencia permite:

- Compartir, copiar y redistribuir el material en cualquier medio o formato.
- Adaptar, remezclar, transformar y crear a partir del material para cualquier finalidad, incluso comercial, siempre y cuando se difundan las contribuciones bajo la misma licencia que el original.
- Reconocer adecuadamente la autoría, proporcionar un enlace a la licencia e indicar si se han realizado cambios.

De esta manera permitimos que nuestros datos sean útiles a terceros, ya sea para su uso informativo, de estudio o comercial siempre y cuando se distribuya bajo la misma licencia.

9. Código.

El código fuente utilizado para generar el dataset en formato Python, se encuentra en cualquiera de los siguientes dos repositorios de Github:

- <https://github.com/eruizmartinez/ConsumoYEmisionesDeTurismos>
- <https://github.com/abayonv/ConsumoYEmisionesDeTurismos>

10. Dataset.

El dataset en formato csv se encuentra en el siguiente repositorio de Github:

, corresponde a los datos obtenidos de seleccionar la marca: **Jaguar**

Contribuciones	Firma
Investigación previa	ABV, ERM
Redacción de las respuestas	ABV, ERM
Desarrollo código	ABV, ERM