

Creación de un *Dataset* a partir del Registro General de la Flota Pesquera

Julio Úbeda Quesada y Lucas Zamora Vera



Imagen creada con <https://openart.ai/home>.

1. Contexto

La práctica de *web scraping* se ha realizado en el contexto de un análisis técnico de buques pesqueros registrados en Europa y España. Para ello, se utilizó como fuente inicial de datos el [Registro General de la Flota Pesquera de la Unión Europea](#), desde donde se extrajeron los CFR (*Community Fishing Fleet Register*), códigos únicos que identifican cada buque dentro del ámbito europeo.

Estos CFR fueron luego empleados para consultar la base de datos española, el [Registro Oficial de Buques Pesqueros del Ministerio de Agricultura, Pesca y Alimentación](#), con el fin de obtener información técnica detallada. Ambas fuentes son pertinentes y fiables, ya que están gestionadas por organismos oficiales y ofrecen datos públicos, actualizados y verificados.

Recurso	Fuente
Sitios web	https://servicio.pesca.mapama.es/censo/ConsultaBuqueRegistro/Buques/Search https://webgate.ec.europa.eu/fleet-europa/search_en
Zenodo	https://doi.org/10.5281/zenodo.15132458
Repositorio de GitHub	https://github.com/LucasZV/repo_tipologia.git
Video de presentación	https://drive.google.com/file/d/1CqjIW1mArjzmBkvHsGr8N4ABXXa4HPrc/view?usp=sharing

2. Título

"Registro de la Flota Pesquera Española: Características y Distribución"

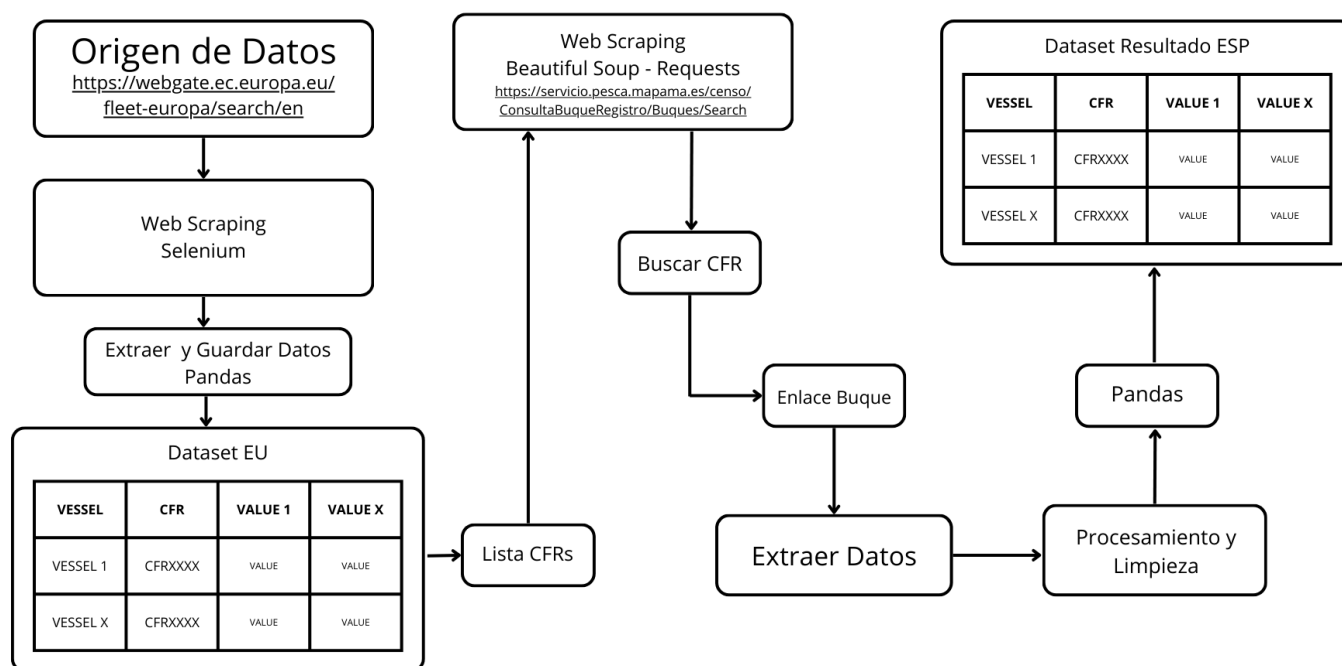
3. Descripción del *Dataset*

Este conjunto de datos recoge información técnica y administrativa de buques pertenecientes al **Registro de la Flota Pesquera Española**, a partir de identificadores únicos (CFR) extraídos del registro europeo. Incluye variables como el nombre del buque, su eslora, potencia, arqueo, puerto base, material del casco, estado en el registro y modalidad de pesca, entre otros.

Esta información permite analizar la **composición, características y distribución geográfica** de la flota pesquera española, ofreciendo una visión detallada de su capacidad operativa y evolución histórica.

4. Representación Gráfica

El siguiente diagrama de flujo muestra el proceso de *web scraping* realizado para extraer y almacenar las características técnicas de la flota pesquera española.



5. Contenido

El conjunto de datos incluye los siguientes campos extraídos del Registro de la Flota Pesquera Española:

- **CFR:** Código de Referencia Comunitario, identificador único del buque a nivel europeo.
- **Nombre:** Nombre registrado del buque.
- **IMO:** Número de identificación del buque en la Organización Marítima Internacional (si aplica).
- **IRCS:** Código de llamada internacional (señal de radio del buque).
- **Matrícula:** Código de identificación nacional del buque en el registro español.
- **Alta en RGFP:** Fecha en la que el buque fue dado de alta en el Registro General de la Flota Pesquera.
- **Estado:** Situación administrativa del buque (activo, baja definitiva, etc.).
- **Eslora total:** Longitud total del buque en metros.
- **Arqueo GT:** Arqueo bruto, medida del volumen interior del buque.
- **Potencia:** Potencia del motor en kilovatios (y caballos de vapor entre paréntesis).

- **Material del casco:** Tipo de material con el que está construido el casco del buque (ej. madera, acero).
- **Puerto base:** Localidad y provincia donde se encuentra registrado el buque.
- **Administración responsable del Registro:** Organismo encargado del registro del buque.
- **Censo por modalidad:** Tipo de arte de pesca o modalidad asignada al buque (ej. cerco, palangre, artes menores).
- **Capacidad del buque no aportable por:** Campo reservado para especificar si hay datos no disponibles por algún motivo.
- **Tipo de auxiliar:** En caso de que el buque sea auxiliar, se especifica el tipo.

Los datos abarcan un período comprendido entre el **31 de diciembre de 1987** y el **25 de marzo de 2025**, lo que permite realizar análisis históricos y técnicos de la evolución de la flota pesquera española durante casi cuatro décadas.

6. Propietario

El propietario del conjunto de datos es el **Ministerio de Agricultura, Pesca y Alimentación de España**, a través de su plataforma de consulta pública del **Registro Oficial de Buques Pesqueros**. Los CFR utilizados como punto de partida fueron obtenidos del **Registro de la Flota Pesquera de la Unión Europea**, gestionado por la Comisión Europea, también de acceso público.

Aunque no se han encontrado análisis exactamente iguales, existen estudios académicos y técnicos que han utilizado registros similares para el análisis de la capacidad pesquera o la evolución de las flotas. Por ejemplo, informes del [Joint Research Centre \(JRC\)](#) de la Comisión Europea, como el [Annual Economic Report on the EU Fishing Fleet](#), hacen uso de los CFR y datos técnicos oficiales para el análisis del sector pesquero a nivel comunitario.

En cuanto a los principios éticos y legales, se ha actuado dentro del marco de la **transparencia y uso responsable de datos abiertos**, accediendo exclusivamente a información de carácter público y sin vulnerar ninguna restricción de acceso o uso. No se ha realizado ningún tipo de ataque automatizado a los servidores, y las peticiones fueron espaciadas y limitadas para respetar la carga de los sitios web. Además, no se han tratado datos personales ni sensibles, en cumplimiento con el Reglamento General de Protección de Datos (RGPD) y las normativas éticas en el uso de información digital.

7. Inspiración

Este conjunto de datos resulta de especial interés porque permite analizar en detalle la **composición, distribución y evolución histórica** de la flota pesquera española, a partir de información técnica precisa y oficial. El hecho de contar con datos como eslora, potencia,

arqueo bruto, puerto base o modalidad de pesca, vinculados al identificador europeo (CFR), permite realizar análisis a diferentes escalas (temporal, geográfica, por tipo de buque, etc.).

En comparación con análisis anteriores, como el *Annual Economic Report on the EU Fishing Fleet* del JRC, que ofrece una visión agregada de la flota europea en términos económicos y de capacidad, este dataset aporta una perspectiva más detallada y centrada exclusivamente en la flota española, facilitando estudios más específicos.

Con estos datos se pretende responder a preguntas como:

- ¿Cómo ha evolucionado la flota pesquera española en términos de tamaño y potencia desde finales de los 80 hasta hoy?
- ¿Qué regiones o puertos concentran el mayor número de buques y con qué características?
- ¿Qué modalidades de pesca han predominado o disminuido en el tiempo?
- ¿Existen patrones en la baja definitiva de los buques vinculados a su tamaño, material o potencia?

Este análisis es útil tanto para investigadores del sector pesquero como para responsables de políticas públicas, ya que permite observar tendencias ([Villasante, S \(2010\)](#)) y apoyar decisiones relacionadas con sostenibilidad, ordenación del esfuerzo pesquero y modernización de la flota.

8. Licencia

La licencia adecuada para este conjunto de datos es '*Creative Commons Attribution-Non-Comercial 4.0 International*'. Esta licencia permite que el dataset sea **usado, compartido, adaptado y redistribuido**, siempre que se cumplan [tres condiciones fundamentales](#):

1. **Atribución (BY)**: Se debe reconocer adecuadamente la autoría del trabajo, mencionando tanto al autor del scraping como a las fuentes oficiales (Ministerio de Agricultura, Pesca y Alimentación de España y la Comisión Europea).
2. **No Comercial (NC)**: Restringe el uso del dataset a **finés no comerciales**, protegiendo así su uso en contextos educativos, investigativos o de divulgación sin que se explote con fines lucrativos.
3. **Compartir Igual (SA)**: Si se generan obras derivadas o se modifica el dataset, estas deben compartirse bajo la **misma licencia**, garantizando que el conocimiento permanezca abierto y libre para futuras reutilizaciones.

9. Código

El objetivo principal del proceso de extracción ha sido la construcción de un **dataset histórico con las características técnicas de la flota pesquera española**. Para ello, se ha diseñado una estrategia en dos fases diferenciadas, cada una orientada a obtener un conjunto de datos necesario para alcanzar este fin.

Herramientas utilizadas

Librería	Función principal
Selenium	Simular navegación web e interactuar con formularios dinámicos (fase 1)
requests	Realizar peticiones HTTP directas para acceder a páginas de detalles (fase 2)
BeautifulSoup	Parsear y extraer datos de contenido HTML (fase 2)
Pandas	Leer, almacenar y estructurar todos los datos recolectados. (fases 1 y 2)

Fase 1: Recolección de información de la flota europea

En primer lugar, se ha recopilado un listado completo de la flota pesquera europea, a través del sitio web oficial de la Unión Europea. Para ello se ha empleado la biblioteca '**Selenium**', que permite simular la navegación humana por sitios web con contenidos dinámicos (Ver **scrapping_fleet_EU.py**).

- **Simulación de navegación:** El script automatiza el acceso y la interacción con formularios web, permitiendo recorrer de forma iterativa todos los países miembros y los años disponibles.
- **Descarga de ficheros:** Tras seleccionar los filtros adecuados, se desencadena la descarga de archivos CSV que contienen la información detallada de la flota por país y año.
- **Gestión de descargas:** Se ha incorporado una lógica de espera para asegurar que las descargas se completen correctamente.
- **Almacenamiento estructurado:** Los datos descargados se leen con '**Pandas**' y se almacenan en estructuras de datos organizadas, facilitando su posterior procesamiento.

Fase 2: Extracción de datos específicos de la flota española

Una vez identificados los registros correspondientes a embarcaciones españolas en el conjunto europeo, se procede a ampliar la información accediendo al **registro nacional de la flota pesquera española**.

- **Acceso a datos estructurados:** En este caso, el portal web nacional ofrece los datos de cada embarcación de forma accesible mediante peticiones HTTP.
- **Extracción con 'requests' y 'BeautifulSoup':** Se realiza el envío de solicitudes directas al servidor y se analiza el contenido HTML con BeautifulSoup para extraer los campos técnicos de cada embarcación (eslora, arqueo, motor, año de construcción, etc.).
- **Procesamiento con 'Pandas':** Los datos extraídos se integran en un dataframe que complementa la información inicial y permite crear un histórico completo y detallado de la flota española.

Dificultades encontradas y soluciones aplicadas

Dificultad	Solución implementada
Contenido dinámico que no se carga inmediatamente	Uso de `WebDriverWait` para asegurar que los elementos estén listos antes de interactuar con ellos
Descargas incompletas o interrumpidas	Monitorización del directorio de descargas hasta que el archivo se haya completado correctamente
Extracción de información no accesible directamente desde el DOM	Uso de `BeautifulSoup` para analizar el contenido HTML y obtener los valores necesarios
Simulación de navegación	simulación de <i>scrolls</i> y <i>clicks</i> mediante Selenium y JavaScript

10. Dataset

El dataset es un conjunto de datos que recoge información técnica y administrativa de buques pesqueros registrados en España, incluyendo eslora, potencia, arqueo, puerto base, estado en el registro y modalidad de pesca, a partir de códigos CFR extraídos del registro europeo.

Se ha publicado en **Zenodo** con el DOI: <https://doi.org/10.5281/zenodo.15132458>

El archivo CSV resultante también está en la carpeta `/data` del [repositorio](#).

11. Video

El video explicativo del proyecto está disponible en el siguiente enlace:

<https://drive.google.com/file/d/1CqjIW1mArjzmBkvHsGr8N4ABXXa4HPrc/view?usp=sharing>

12. Contribuciones

Contribuciones	Firma
Investigación previa	Julio UQ, Lucas ZV
Redacción de las respuestas	Julio UQ, Lucas ZV
Desarrollo del código	Julio UQ, Lucas ZV
Participación en el vídeo	Julio UQ, Lucas ZV