

# Proyectos II, integración y preparación de datos

## Primera presentación:

### Captura de datos

### Introducción

Antes de cada HITO de presentación, debemos rellenar estas fichas y presentarlas a través de PoliformaT, en la tarea que indiquen los profesores. Cada equipo de trabajo presenta las mismas fichas. Sólo será necesario que las suba uno de los componentes del equipo.

Nombres y apellidos de los autores:

- **José Miguel García Ríos**
- **Pablo Alpuente Tabasco**
- **Vicente Ricos Caras**
- **Marc Fernández Cortina**
- **Víctor Lorenzo López**

### 1. Las Fichas de Configuración

Una vez hayamos decidido el proyecto en el que vamos a trabajar, debemos rellenar el Alcance preliminar del proyecto (apartados 1.1.).

Una vez definido el Alcance, desglosaremos el trabajo de esta primera etapa en:

- Localización de las fuentes.
- Técnicas de obtención de los datos y extracción. Por ejemplo:

- Descarga de ficheros .csv
  - Descarga desde una URL
  - Lectura de tablas incrustadas en HTML
  - Conversión de .JSON
  - Conversión de .XML
  - Recoger datos de Twitter y limpieza sobre expresiones regulares
  - Recoger datos de Google y limpieza sobre expresiones regulares
  - Web scraping
  - Descargas en tiempo real durante varios ciclos
- Análisis de las fuentes: interpretación de los datos, valoración de su utilidad en el proyecto.
  - Análisis de los campos, formatos y tipo de información de cada fuente y valoración del cruce de datos de distintas fuentes para nuestro proyecto.

## 1.1. Alcance (preliminar)

Explica brevemente qué información vamos a obtener de las distintas fuentes seleccionada y el uso que los datos podrían tener tras integrar y transformar las muestras con las que vamos a trabajar.

Identificador de la Ficha	Búsqueda de fuentes
<i>Relatad todas las fuentes que habéis investigado y consultado en el proceso de localización de las que vais finalmente a utilizar.</i>	<p>Este proyecto se basa en el análisis de datos energéticos proporcionados por la web oficial de <b>Red Eléctrica de España (REE)</b> y datos meteorológicos de la <b>Agencia Estatal de Meteorología (AEMET)</b>. Se utilizarán tres fuentes de datos principales:</p> <p><b>1. Red Eléctrica de España (REE) - Producción, Consumo y Balance Energético</b></p>

	<ul style="list-style-type: none"><li>• <b>Fuente:</b> <a href="https://www.ree.es/es/datos/balance">https://www.ree.es/es/datos/balance</a>, <a href="https://www.ree.es/es/datos/demanda">https://www.ree.es/es/datos/demanda</a>, <a href="https://www.ree.es/es/datos/generacion">https://www.ree.es/es/datos/generacion</a></li><li>• <b>Descripción:</b> Datos mensuales por comunidad autónoma sobre generación y consumo de electricidad, junto con el balance energético.</li><li>• <b>Formato disponible:</b> CSV o Excel</li><li>• <b>Razón para su uso:</b> Proporciona información oficial y detallada sobre la situación energética en España, con el nivel de granularidad adecuado para analizar las diferencias entre comunidades autónomas.</li></ul> <p><b>2. Red Eléctrica de España (REE) - Balance Eléctrico Nacional</b></p> <ul style="list-style-type: none"><li>• <b>Fuente:</b> <a href="https://www.ree.es/es/datos/balance/balance-electrico">https://www.ree.es/es/datos/balance/balance-electrico</a></li><li>• <b>Descripción:</b> Datos diarios sobre la producción y el consumo eléctrico a nivel nacional, incluyendo la generación desglosada por tipo de fuente energética y el balance eléctrico del sistema.</li><li>• <b>Formato disponible:</b> Tablas online y archivos descargables en CSV.</li><li>• <b>Razón para su uso:</b> Permite analizar la evolución del consumo y la producción energética en el tiempo, identificando patrones estacionales y tendencias. Además, estos datos serán utilizados para desarrollar modelos de predicción basados en series temporales.</li></ul> <p><b>3. Datos Meteorológicos - AEMET (Agencia Estatal de Meteorología)</b></p>
--	---

	<ul style="list-style-type: none"> <li>• <b>Fuente:</b> <a href="https://opendata.aemet.es/centrodedescargas/inicio">https://opendata.aemet.es/centrodedescargas/inicio</a></li> <li>• <b>Descripción:</b> Datos climáticos regionales que incluyen temperatura, velocidad del viento y radiación solar, variables que pueden influir en la producción y el consumo de energía.</li> <li>• <b>Formato disponible:</b> JSON a través de una API.</li> <li>• <b>Razón para su uso:</b> Permite analizar la relación entre el clima y el comportamiento energético en cada comunidad autónoma, facilitando la detección de patrones en la demanda y producción de energía.</li> </ul>
<p><i>Criterios seguidos para la selección de las fuentes que se van a usar para el proyecto.</i></p>	<p>Inicialmente, exploramos distintas ideas de proyecto, considerando temas como el rendimiento académico y los índices económicos. Sin embargo, al analizar estos temas, encontramos dificultades para definir objetivos concretos y alcanzables, además de la falta de bases de datos completas y actualizadas que permitieran un análisis profundo.</p> <p>Tras discutir nuestras opciones con la profesora, descubrimos la página oficial de <b>Red Eléctrica de España (REE)</b>, que proporciona datos detallados sobre producción, consumo y balance energético en España a nivel regional y nacional. Esta fuente nos pareció especialmente adecuada porque:</p> <ul style="list-style-type: none"> <li>• Es una fuente oficial y confiable, asegurando datos precisos y actualizados.</li> <li>• Permite analizar patrones de consumo y producción de energía a lo largo del tiempo.</li> </ul>

	<ul style="list-style-type: none"><li>• Nos da la posibilidad de cruzar los datos energéticos con variables meteorológicas, añadiendo una dimensión adicional al análisis.</li></ul> <p>Además, encontramos en la <b>Agencia Estatal de Meteorología (AEMET)</b> una fuente ideal para obtener datos sobre temperatura, viento y radiación solar, lo que nos permite estudiar cómo factores climáticos pueden influir en el consumo y la producción de energía.</p> <p>Por estas razones, seleccionamos REE y AEMET como nuestras fuentes principales, ya que nos proporcionan datos detallados, abiertos y bien documentados, permitiéndonos desarrollar un análisis sólido con múltiples enfoques.</p>
--	--

## 1.2. Técnicas de obtención de datos y extracción

Para la recopilación de datos en este proyecto, se han empleado diferentes técnicas según la estructura y el formato en que se encuentran disponibles en las fuentes seleccionadas.

### Red Eléctrica de España (REE)

#### Producción y consumo

- **Descarga de ficheros .csv:**

Los datos mensuales de producción y consumo por comunidad autónoma se encuentran disponibles en la web oficial de REE en formato CSV o XLSM. Se ha optado por la descarga directa de los archivos en CSV, ya que este formato facilita su procesamiento y análisis.

### **Balance Eléctrico Nacional**

- **Descarga de ficheros .csv:**

Los datos diarios sobre el balance eléctrico a nivel nacional están disponibles en formato CSV en la web de REE. Se ha optado por esta modalidad, ya que permite la creación de series temporales para analizar la evolución del consumo y la producción de energía en el tiempo.

Para la descarga de estos archivos csv hemos tenido muchos problemas, estamos valorando la opción de poder utilizar la API, ya que los csv salen mal tabulados, dan problemas para analizarlos y incluso hay algunos que de momento no hemos podido descargar.

### **Agencia Estatal de Meteorología (AEMET) – Datos Meteorológicos**

- **Conversión de .JSON:**

AEMET proporciona datos meteorológicos a través de su API en formato JSON. Para la integración con los datos energéticos, será necesario convertir estos archivos a un formato estructurado (CSV o Data Frame por ej).

El abanico de posibilidades en esta API es muy grande, por lo que para el cruce de datos se podrán utilizar las variables meteorológicas que más convengan. Aclaraciones para el uso de esta:

1 - Para poder acceder a AEMET OpenData, es necesario solicitar una API Key

2 - Ese es el formato típico de respuesta de la **API de AEMET**. Lo que sucede es que AEMET no te devuelve directamente los datos en la primera respuesta, sino que te proporciona una **URL temporal** donde puedes descargarlos.

3 - Descargar los datos desde la URL en "datos", convertir el json a dataframe

4- En la Api queda detallada la información de cada variable, para seleccionar solo las convenientes en los cruces

### **Consideraciones adicionales**

- No se utilizará web scraping en este proyecto, ya que las fuentes seleccionadas ofrecen métodos de descarga directa o API.
- No se utilizará la extracción de datos desde redes sociales o motores de búsqueda, ya que no son relevantes para el análisis energético propuesto.

- Las descargas en tiempo real no serán necesarias, dado que el análisis se basará en datos históricos y no en monitoreo en vivo.

## 1.3. Análisis de las fuentes: interpretación de los datos, valoración de su utilidad en el proyecto.

### 1. Base de datos de balance energético

La base de datos de balance energético proporciona información detallada sobre la producción, el consumo y el almacenamiento de energía en España, desglosada por comunidad autónoma y por tipo de tecnología utilizada.

#### Fuentes de energía renovables

Estas variables indican cuánta energía se genera en cada comunidad a partir de fuentes renovables:

- **Hidráulica:** Energía generada a partir de centrales hidroeléctricas.
- **Eólica:** Producción de electricidad mediante aerogeneradores.
- **Solar fotovoltaica:** Energía producida por paneles solares.
- **Solar térmica:** Energía solar transformada en calor antes de convertirse en electricidad.
- **Hidroeólica:** Combinación de energía eólica e hidráulica para generar electricidad.
- **Otras renovables:** Energías renovables no incluidas en las anteriores, como biomasa o geotérmica.
- **Residuos renovables:** Energía generada a partir de residuos orgánicos o reciclables.
- **Generación renovable:** Suma de toda la energía generada a partir de fuentes renovables en cada comunidad.

#### Fuentes de energía no renovables

Estas variables indican cuánta electricidad se genera en cada comunidad a partir de fuentes no renovables:

- **Nuclear:** Energía generada en centrales nucleares.

- **Ciclo combinado:** Generación a partir de gas natural y turbinas de vapor.
- **Carbón:** Electricidad producida mediante la combustión de carbón.
- **Motores diésel:** Energía generada con motores de combustión interna.
- **Turbina de gas:** Electricidad generada directamente a partir de turbinas de gas.
- **Turbina de vapor:** Producción basada en el uso de vapor de agua.
- **Cogeneración:** Producción combinada de electricidad y calor.
- **Residuos no renovables:** Energía generada a partir de residuos industriales o no reciclables.
- **Generación no renovable:** Suma de toda la energía generada a partir de fuentes no renovables.

### **Variables relacionadas con almacenamiento y distribución**

Estas variables reflejan cómo se almacena y distribuye la energía generada:

- **Turbinación bombeo:** Energía generada mediante sistemas de bombeo hidráulico.
- **Consumos bombeo:** Energía utilizada en procesos de almacenamiento por bombeo.
- **Entrega batería:** Cantidad de energía suministrada por sistemas de baterías.
- **Carga batería:** Energía almacenada en baterías.
- **Saldo almacenamiento:** Diferencia entre carga y descarga de almacenamiento energético.
- **Saldo de intercambios:** Balance de importación y exportación de energía entre comunidades.

---

## **2. Base de datos de generación y consumo de energía**

Esta base de datos contiene información sobre la cantidad de energía generada y consumida en España, desglosada por tipo de fuente energética.

### **Fuentes de energía no renovables**

- **Nuclear:** Energía generada en centrales nucleares.



- **Carbón:** Producción de electricidad mediante la combustión de carbón.
- **Fuel + Gas:** Energía generada a partir de combustibles fósiles líquidos y gaseosos → hemos observado muchos nulos, a falta de eliminación/transformación de variables podríamos valorar no considerarla oportuna para el análisis.
- **Motores diésel:** Generación mediante motores de combustión interna con diésel.
- **Turbina de gas:** Producción basada en la combustión de gas en turbinas.
- **Turbina de vapor:** Generación mediante turbinas impulsadas por vapor de agua.
- **Ciclo combinado:** Generación basada en gas natural y turbinas de vapor.
- **Residuos no renovables:** Energía generada a partir de residuos industriales o materiales no reciclables.

### Fuentes de energía renovables

- **Hidráulica:** Producción mediante centrales hidroeléctricas.
- **Hidroeólica:** Combinación de energía eólica e hidráulica para generación de electricidad.
- **Eólica:** Electricidad generada por aerogeneradores.
- **Solar fotovoltaica:** Energía captada por paneles solares y convertida en electricidad.
- **Solar térmica:** Producción eléctrica mediante energía solar transformada en calor.
- **Otras renovables:** Energías renovables adicionales como biomasa o geotérmica.
- **Residuos renovables:** Generación de energía a partir de residuos orgánicos o reciclables.
- **Generación total:** Suma de toda la energía generada a partir de fuentes renovables y no renovables.

---

### 3. Base de datos de AEMET

La base de datos de la Agencia Estatal de Meteorología (AEMET) proporciona información climatológica relevante para analizar su relación con la producción y el consumo energético en España. Se han seleccionado dos conjuntos de datos (inicialmente, a falta de próximos cambios/actualizaciones) que permiten estudiar la

influencia del clima en la generación de energía renovable y en la demanda eléctrica.

### **Climatologías diarias** (Sección Valores Climatológicos)

- **Formato esperado:** CSV o Excel.
- **Frecuencia:** Mensual o diario
- **Descripción:** Contiene información meteorológica registrada diariamente en distintas estaciones meteorológicas. Incluye variables como temperatura media, temperatura mínima y máxima, humedad relativa, velocidad y dirección del viento, y precipitaciones.
- **Utilidad en el proyecto:**
  - Permite analizar la relación entre temperatura y demanda de electricidad en cada comunidad autónoma.
  - Facilita el estudio de la influencia del viento en la generación de energía eólica.
  - Proporciona información detallada sobre la variabilidad diaria de las condiciones meteorológicas y su impacto en el sistema energético.

### **Datos de radiación global, directa o difusa** (Sección Redes Especiales)

- **Formato esperado:** CSV o Excel.
- **Frecuencia:** Mensual o diario
- **Descripción:** Contiene mediciones de radiación solar global, directa y difusa, registradas en estaciones meteorológicas específicas. Estas variables son esenciales para el análisis de la eficiencia de la producción fotovoltaica.
- **Utilidad en el proyecto:**
  - Permite evaluar la relación entre la radiación solar y la producción de energía solar fotovoltaica.
  - Facilita la identificación de patrones estacionales en la producción de energía renovable.
  - Ayuda a estimar el rendimiento de instalaciones fotovoltaicas en diferentes comunidades autónomas.

Estos conjuntos de datos serán fundamentales para analizar la influencia de las condiciones meteorológicas en el sistema energético español y para establecer correlaciones con la producción de energías renovables y la demanda eléctrica.

Como la cantidad de información proporcionada por la api es tan grande, se puede valorar utilizar solo la climatología (diaria o mensual, aún no está definido) y no los datos de radiación. De igual manera, podríamos seleccionar algún otro tipo de variable meteorológica que nos pueda servir para el cruce de datos...

#### **1.4. Análisis de los campos, formatos y tipo de información de cada fuente y valoración del cruce de datos de distintas fuentes para nuestro proyecto. Valoración del cruce de datos de distintas fuentes para nuestro proyecto.**

Para integrar correctamente las bases de datos de Red Eléctrica de España (REE) y AEMET, será necesario realizar ciertos ajustes en los datos con el objetivo de asegurar la compatibilidad entre ambas fuentes y permitir un análisis adecuado.

En la base de datos de REE, los datos están desglosados por comunidad autónoma y proporcionan información mensual sobre producción, consumo y balance energético. En la base de datos de AEMET, los datos meteorológicos también están regionalizados, pero algunos pueden estar a nivel de estación meteorológica. Será necesario asegurar que los nombres de las comunidades autónomas sean homogéneos en ambas bases de datos para facilitar su integración.

Las variables de REE son cuantitativas y miden la cantidad de energía generada o consumida en cada región. Dentro de la base de datos de AEMET, las variables también son cuantitativas, incluyendo temperatura, velocidad del viento y radiación solar, con valores en diferentes escalas según la magnitud que representan. Será necesario estandarizar las unidades en caso de que se presenten diferencias en la medición de las mismas variables en distintos conjuntos de datos.

El cruce de datos se realizará utilizando la comunidad autónoma como clave principal, de manera que se puedan asociar las condiciones meteorológicas con la producción y el consumo energético de cada región. Además, para garantizar la coherencia temporal, los datos meteorológicos serán agregados a nivel mensual para que coincidan con la granularidad de los datos energéticos de REE. Sin embargo, también existe la opción de trabajar con los datos a nivel diario, lo que permitiría un análisis más detallado de la variabilidad climática y su impacto en la generación y el consumo de energía.

Dado que algunas variables pueden no ser relevantes para el análisis final, es posible descartar aquellas que no aporten valor significativo o que sean redundantes. Además, puede ser necesario reordenar y limpiar

el dataset, eliminando valores atípicos, corrigiendo inconsistencias en los nombres de las regiones y asegurando la integridad de los datos.

El objetivo del cruce de datos es evaluar la relación entre variables climáticas y el comportamiento del sistema energético. Se analizará si existe correlación entre la temperatura y el consumo eléctrico, la velocidad del viento y la producción eólica, y la radiación solar y la producción fotovoltaica. A partir de este cruce, se podrán extraer conclusiones sobre el impacto de las condiciones climáticas en el sector energético y desarrollar modelos de predicción basados en patrones climáticos y de demanda energética.