

# Proyectos II, integración, calidad y análisis exploratorio de datos

## Segunda presentación:

### Integración 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:

Víctor	Lorenzo López
Marc	Fernández Cortina
Pablo	Alpuente Tabasco
Vicente	Ricós Caras
José Miguel	García Ríos

## 2. Interés y alcance del proyecto.

### 2.1. Explica el objetivo principal de tu proyecto ¿qué presenta este estudio?

El objetivo principal de este proyecto es evaluar el impacto del clima y la estacionalidad en la generación y el balance energético regional en España, integrando para ello información procedente de distintas bases de datos sobre producción energética y condiciones meteorológicas.

Este estudio presenta un análisis conjunto de datos energéticos y climáticos con el fin de identificar patrones y relaciones significativas entre factores como las precipitaciones, la temperatura o las estaciones del año, y la generación de energía (especialmente renovable) en las distintas comunidades autónomas.

Con ello, se busca comprender cómo varía la producción energética en función de las condiciones climáticas y temporales, así como evaluar el grado de autosuficiencia o dependencia energética en cada región. Este enfoque multidimensional permite aportar una visión más completa sobre los retos y oportunidades que ofrece la transición energética en el contexto del cambio climático y la sostenibilidad.

Con esto podríamos resolver preguntas como:

- ¿Qué comunidades autónomas generan más energía renovable y en qué épocas del año?
- ¿Qué relación existe entre las precipitaciones y la producción de energía hidráulica en distintas regiones?

### **2.2. Explica para qué y para quién podría ser de utilidad este estudio**

Este estudio puede ser útil para identificar cómo influyen el clima y la estacionalidad en la producción energética, lo que permite detectar oportunidades de mejora en la planificación y uso de fuentes renovables (o no) en distintas regiones de España.

Por ejemplo, al gobierno le puede interesar ya que permite diseñar políticas energéticas más eficientes y sostenibles, adaptadas a las características climáticas de cada comunidad autónoma, así como anticipar necesidades energéticas según la época del año.

Por otro lado, a los responsables de redes eléctricas y operadores del sistema energético les ayuda a optimizar el balance entre generación y demanda, integrando mejor las renovables en función del clima. A las empresas energéticas y fabricantes de tecnologías renovables también les puede servir para planificar inversiones y estrategias comerciales, identificando regiones con mayor potencial para solar, eólica o hidráulica.

A las comunidades autónomas y ayuntamientos les proporciona información útil para fomentar la autosuficiencia energética local, impulsar proyectos renovables y mejorar su planificación territorial.

Finalmente, a los ciudadanos en general les ofrece una mejor comprensión del papel del clima en la transición energética y puede fomentar la conciencia medioambiental al visualizar el potencial real de las renovables en su entorno.

### **2.3. ¿Por qué piensas que es novedoso? ¿has visto estudios similares?**

Creemos que nuestro proyecto es novedoso porque integra múltiples dimensiones de análisis: datos de producción energética, balance energético regional y variables meteorológicas, todo ello desglosado por comunidad autónoma y periodo temporal.

## Proyectos II, integración y preparación de datos

---

Aunque hemos encontrado informes de organismos oficiales como Red Eléctrica de España o IDAE, estos suelen ofrecer información descriptiva o agregada, sin realizar un análisis cruzado con los datos meteorológicos. Mientras que nosotros combinamos datos meteorológicos y energéticos en paralelo, lo cual permite estudiar interacciones que no suelen tratarse de forma conjunta. Analizamos la variabilidad por comunidades y estaciones dentro de España, lo que permite identificar diferencias significativas entre comunidades, y estudiamos no solo la generación, sino también el balance energético, lo cual tiene implicaciones directas en autosuficiencia y sostenibilidad.

Por ello, consideramos que nuestra propuesta aporta una visión más transversal y analítica, con potencial para abrir nuevas líneas de investigación o apoyar decisiones políticas y estratégicas.

### 2.4. Alcance (objetivos definitivos del proyecto)

**Define los objetivos del análisis de datos de tu proyecto.** Se deben presentar 5 objetivos de análisis.

Con respecto al ejemplo anterior, posibles objetivos de análisis podrían ser:

- Estudiar la generación dependiendo las fechas del año y la comunidad.
- Estudio general de energías renovables en invierno/verano.
- Estudiar el balance energético por comunidad.
- Estudiar la generación de las comunidades seleccionadas en función del tiempo meteorológico u otras variables semejantes.
- Estudiar la energía hidráulica (o alguna otra) por comunidades en función del tiempo de precipitaciones, seleccionando fechas específicas donde se pueda comprobar era relación entre energía-tiempo. Es posible también hacerlo de la misma manera con energía eólica y viento, por ejemplo, o energía solar fotovoltaica y variables como insolación o sol.

## 3. Calidad y Análisis exploratorio.

**Describe la calidad de los datos y los resultados del análisis exploratorio efectuado. Explica el trabajo técnico, como por ejemplo, estadísticas aplicadas, visualizaciones o representaciones que has utilizado (puedes poner alguna captura como ejemplo), etc. Valora el esfuerzo del análisis exploratorio de tu proyecto.**

**No se trata de describir los objetivos resultado del proyecto, sino lo que has tenido que hacer para entender los datos, ver qué nos ofrecen, cómo de “sucio” es la fuente, etc.**

**Puedes comentar el descarte de datos de las fuentes. Aquellos que no vayas a tener en cuenta por no ser útiles a tus objetivos.**

**Recomendamos utilizar RMarkdown ya que el análisis exploratorio se espera que lo hayáis hecho en R. Poned un anexo donde veamos las instrucciones, comandos, funciones, etc. además de las gráficas.**

Nota: En este proyecto se ha optado por excluir las ciudades autónomas de **Ceuta** y **Melilla** debido a su limitada relevancia en el análisis del sistema energético nacional. Estas ciudades, a diferencia de las comunidades autónomas peninsulares e insulares, **no están integradas plenamente en la red eléctrica peninsular**, y su producción y consumo energético no se reportan con el mismo nivel de detalle en las bases de datos proporcionadas por Red Eléctrica de España (REE). Además, por su reducido tamaño poblacional y territorial, así como por su dependencia energética del exterior, **su inclusión podría introducir sesgos o distorsiones** en el análisis agregado por comunidad. Por este motivo, y en línea con prácticas comunes en estudios energéticos a nivel nacional, se ha considerado adecuado centrarse únicamente en las **17 comunidades autónomas**, que representan el grueso del sistema eléctrico español tanto en generación como en consumo.

Hasta este punto del proyecto, se ha realizado un importante trabajo de recopilación, limpieza y estandarización de los datos de generación energética por comunidad autónoma en España. Este proceso ha sido esencial para garantizar la coherencia y la comparabilidad entre datasets.

### **3.1 Procesamiento de los datos de generación energética**

#### ***Obtención de los datos***

Se han descargado individualmente los ficheros de generación mensual por comunidad autónoma desde la web oficial de Red Eléctrica de España (REE), ya que no es posible descargar todas las comunidades en un único archivo. En total se han procesado 17 datasets, uno por cada comunidad autónoma continental e insular, excluyendo Ceuta y Melilla por los motivos justificados previamente.

### ***Limpieza y transformación***

Para cada archivo .csv:

- Se han eliminado las primeras 5 filas y las últimas 11, ya que contenían metadatos o notas irrelevantes para el análisis.
- Se ha reestructurado cada tabla al formato **long**: un dato por fila, indicando el tipo de energía, la comunidad, el mes y el valor correspondiente.
- Se ha creado una variable tipo Date a partir del campo mes, permitiendo el análisis temporal y la ordenación cronológica de los registros.
- Se ha gestionado la conversión de valores numéricos (con coma decimal) y la transformación de guiones (-) en valores faltantes (NA).
- Finalmente, cada dataset limpio ha sido guardado con un nombre estandarizado del tipo generacion\_comunidad\_limpio.csv.

Esta transformación ha sido clave para permitir comparaciones temporales, análisis exploratorios y cruces de datos con fuentes meteorológicas de AEMET. Un vistazo rápido a estos data sets:

A tibble: 6 × 4

variable <chr>	comunidad <chr>	valor <dbl>	fecha <date>
Hidráulica	Galicia	1584615.68	2023-01-01
Carbón	Galicia	24378.72	2023-01-01
Ciclo combinado	Galicia	302805.60	2023-01-01
Eólica	Galicia	1175200.38	2023-01-01
Solar fotovoltaica	Galicia	1035.82	2023-01-01
Otras renovables	Galicia	22892.44	2023-01-01

### ***Unificación de los conjuntos de datos regionales***

Una vez finalizado el proceso de limpieza individual de los 17 conjuntos de datos de generación energética —uno por cada comunidad autónoma peninsular y archipiélago incluido— procedimos a su **unificación en un único dataset general** con formato largo. Este formato nos permite analizar la evolución temporal y regional de la producción energética desglosada por tipo de tecnología.

El archivo resultante, `generacion_todas_comunidades.csv`, reúne todas las observaciones estructuradas con cuatro columnas principales: tipo de energía, comunidad autónoma, valor energético y fecha.





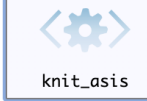
### **Análisis Exploratorio**

Antes de realizar análisis más complejos, se llevó a cabo un **análisis exploratorio de datos (EDA)** centrado en:

- Identificar los distintos **tipos de energías** presentes en el dataset.
- Comprobar qué **comunidades autónomas** estaban representadas.
- Detectar la **presencia de valores nulos** por columna y por tipo de energía.

A continuación, se incluyen algunas **tablas resumen** generadas desde R para ilustrar estos puntos:

## Proyectos II, integración y preparación de datos

				
<b>Tipo de energía</b>	<b>Nº de nulos</b>	<b>Total observaciones</b>	<b>% Nulos</b>	
Fuel + Gas	19	24	79.17	
Carbón	45	96	46.88	
Eólica	14	360	3.89	
Residuos no renovables	2	240	0.83	
Ciclo combinado	1	312	0.32	
Cogeneración	0	384	0.00	
Generación total	0	384	0.00	
Hidroeólica	0	24	0.00	
Hidráulica	0	384	0.00	
Motores diésel	0	48	0.00	
Nuclear	0	96	0.00	
Otras renovables	0	384	0.00	
Residuos renovables	0	144	0.00	
Solar fotovoltaica	0	384	0.00	
Solar térmica	0	144	0.00	
Turbina de gas	0	48	0.00	
Turbina de vapor	0	24	0.00	

Al realizar un primer análisis exploratorio del dataset, observamos una presencia significativa de valores nulos en algunas tecnologías energéticas, con 81 datos faltantes, lo cual es una buena noticia ya que nos indica que la calidad de nuestro data set es bastante buena para próximos análisis.

En cuanto a estos nulos, la variable **"Fuel + Gas"** presenta un **80% de datos faltantes**, seguida por **"Carbón"**, con aproximadamente un **46% de valores nulos**. Incluso en tecnologías más extendidas como la **"Eólica"**, se detecta cerca de un **4% de datos ausentes**.



## Proyectos II, integración y preparación de datos

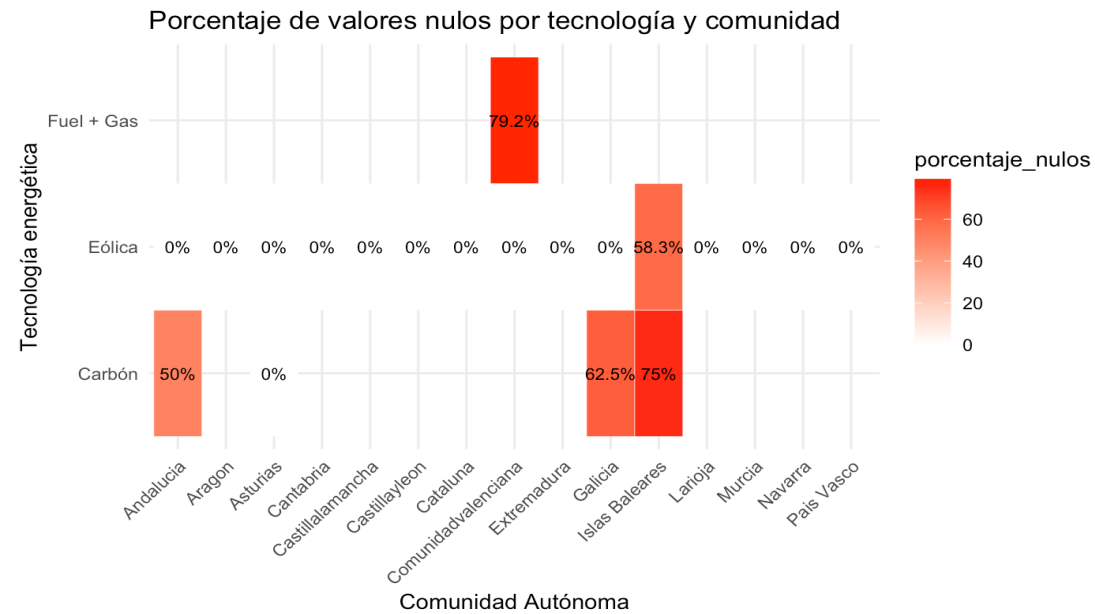
---

A continuación, analizamos **en qué comunidades autónomas** faltan estos datos y en cuáles están presentes, con el objetivo de decidir si **descartar estas variables, rellenar los valores ausentes**, o bien **focalizar el análisis en aquellas regiones donde sí existen registros suficientes**.

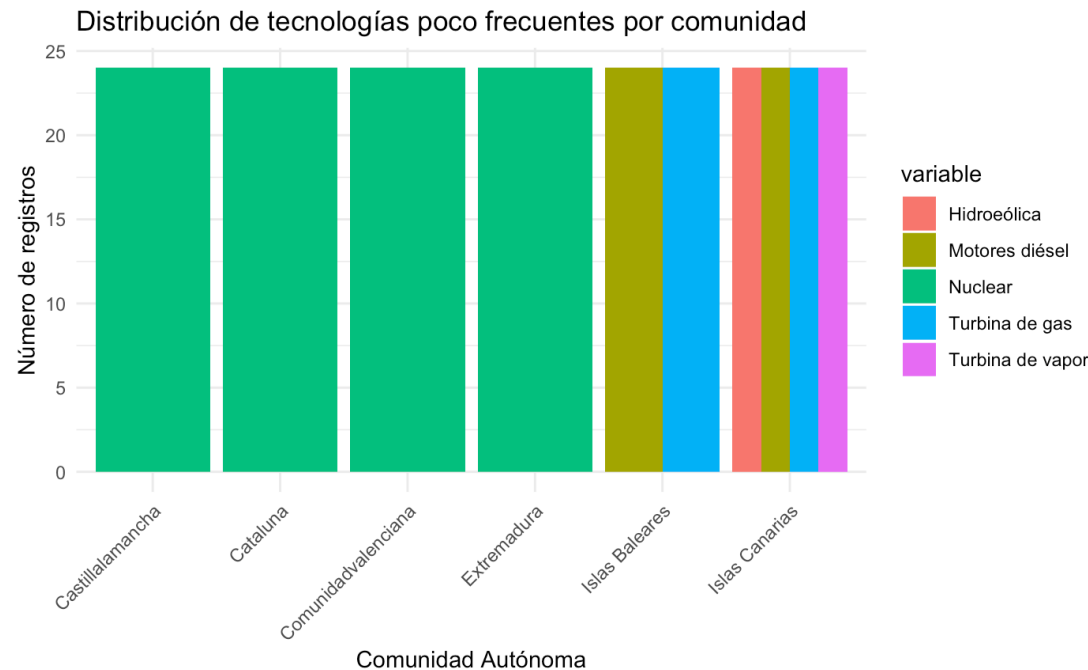
A partir del gráfico de calor que se muestra más abajo podemos identificar con claridad las tecnologías energéticas con mayor proporción de datos ausentes, así como su distribución geográfica. Este análisis resulta clave para decidir qué variables conservar en el análisis y en qué regiones pueden aplicarse.

- **Fuel + Gas:** Esta tecnología únicamente presenta valores en la Comunidad Valenciana, mientras que en el resto de comunidades autónomas los datos están completamente ausentes. Este patrón indica que su uso está restringido geográficamente a una única región. Por tanto, se trata de una variable que ha sido **descartada del análisis**, ya que no aporta información representativa del conjunto del país
- **Carbón:** Observamos que esta tecnología solo tiene presencia significativa en Islas Baleares, Galicia, Andalucía y Asturias. En Islas Baleares, sin embargo, más del 75% de los registros son nulos, mientras que en Asturias no hay datos ausentes, y en Galicia y Andalucía los datos son relativamente completos. Este patrón está relacionado con la evolución del mix energético español, ya que el carbón ha sido progresivamente eliminado como fuente de energía en muchas regiones debido a su alto impacto ambiental y las políticas de descarbonización impulsadas a nivel europeo. Asturias, históricamente una región minera, mantiene registros por su pasado industrial. En comunidades donde ya no se produce carbón, como Madrid o Castilla-La Mancha, directamente no se registran datos, lo que justifica los valores nulos. Por estos motivos, **también se ha decidido eliminar la variable Carbón del análisis**.
- **Eólica:** La generación eólica presenta valores ausentes únicamente en Islas Baleares, con aproximadamente un 60% de registros nulos. Esto puede explicarse por las características orográficas y meteorológicas del archipiélago, donde la instalación de grandes parques eólicos es más limitada, tanto por la menor disponibilidad de suelo como por restricciones ambientales y paisajísticas. En el resto del país, la energía eólica tiene una implantación mucho más homogénea, lo que la convierte en una variable **relevante y representativa**, por lo que **se mantendrá en el análisis**.

## Proyectos II, integración y preparación de datos



## Proyectos II, integración y preparación de datos



Tras analizar la distribución de las tecnologías poco frecuentes en el dataset, observamos que ciertas fuentes de energía presentan una presencia muy localizada y escasa representación en el conjunto de comunidades autónomas:

- **Nuclear:** Solo se registra generación nuclear en **Comunidad Valenciana, Castilla-La Mancha, Extremadura y Cataluña**, lo cual es coherente con la ubicación de las centrales nucleares activas en España. Debido a que estas instalaciones están concentradas en regiones específicas, **decidimos mantener esta variable**, ya que representa una parte significativa del mix energético nacional, aunque limitada territorialmente.
- **Hidroeléctrica, Motores diésel, Turbina de gas y Turbina de vapor:** Estas tecnologías solo aparecen en **Islas Canarias o Baleares**, con un número muy reducido de registros. Su contribución al conjunto del análisis nacional es marginal, y su escasa representación dificulta su comparación entre comunidades. **Por ello, estas variables serán descartadas del análisis principal**, para centrarnos en tecnologías más presentes y homogéneamente distribuidas en el territorio.

### Tecnologías renovables:

- **Hidráulica:** Generación eléctrica mediante el aprovechamiento del agua en embalses o ríos.
- **Eólica:** Energía producida por el viento usando aerogeneradores.
- **Solar fotovoltaica:** Conversión directa de la luz solar en electricidad mediante paneles solares.
- **Solar térmica:** Aprovecha el calor del sol para generar electricidad o calefacción.
- **Otras renovables:** Agrupa fuentes como biomasa, geotermia o biogás.
- **Residuos renovables:** Electricidad obtenida a partir de residuos orgánicos biodegradables.

### Tecnologías no renovables:

- **Nuclear:** Energía producida por fisión nuclear en reactores. No emite CO<sub>2</sub> pero genera residuos radiactivos.
- **Ciclo combinado:** Usa gas natural en turbinas de gas y vapor para generar electricidad de forma eficiente.
- **Cogeneración:** Tecnología que produce simultáneamente calor y electricidad, mayormente en entornos industriales.
- **Residuos no renovables:** Electricidad generada a partir de la incineración de residuos no reciclables.
- **Generación total:** Es la suma de todas las tecnologías (renovables y no renovables) por comunidad y mes.

En el análisis exploratorio también hemos examinado la representación de las comunidades autónomas en el conjunto de datos. Todas las 17 comunidades seleccionadas están presentes en el dataset, aunque el número de registros varía ligeramente entre ellas, en función de la diversidad de tecnologías energéticas disponibles y registradas en cada una.

A continuación, se presenta el número total de registros por comunidad autónoma:

- **Cataluña:** 264 registros
- **Comunidad Valenciana:** 240 registros
- **Andalucía:** 216 registros
- **Castilla-La Mancha:** 216 registros
- **Galicia:** 216 registros
- **País Vasco:** 216 registros
- **Aragón:** 192 registros

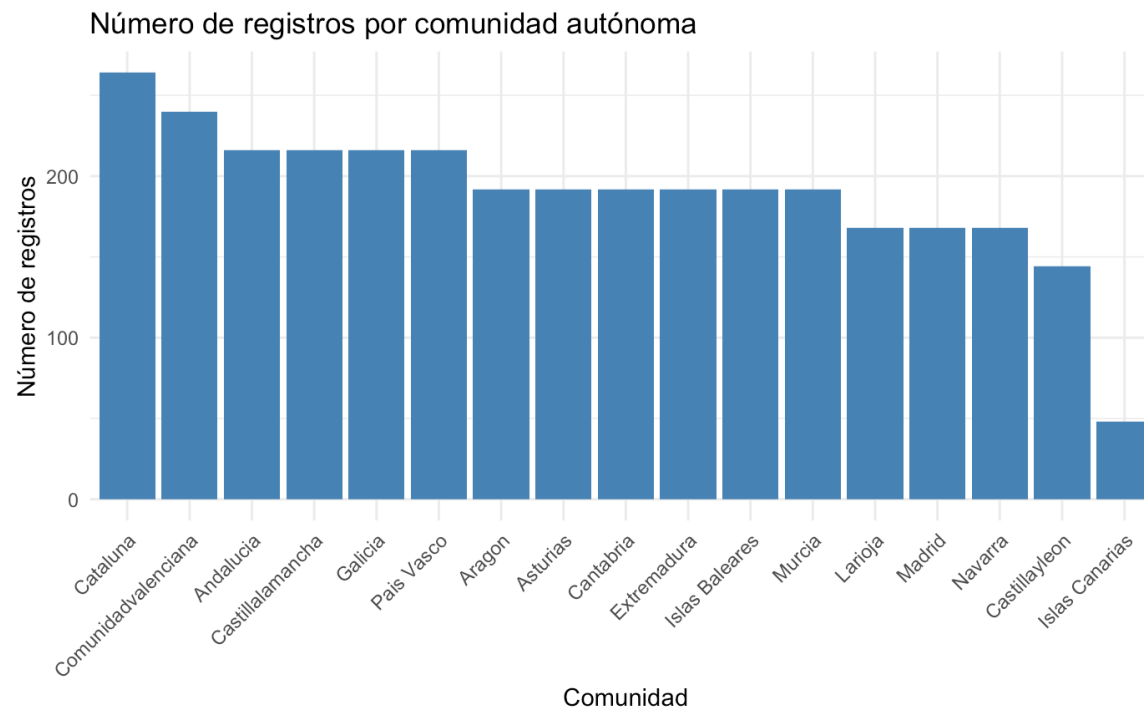
## Proyectos II, integración y preparación de datos

---

- **Asturias:** 192 registros
- **Cantabria:** 192 registros
- **Extremadura:** 192 registros
- **Islas Baleares:** 192 registros
- **Murcia:** 192 registros
- **La Rioja:** 168 registros
- **Madrid:** 168 registros
- **Navarra:** 168 registros
- **Castilla y León:** 144 registros
- **Islas Canarias:** 48 registros

Podemos comprobar que **todas las comunidades están representadas**, aunque **Islas Canarias** presenta un número significativamente menor de registros debido a la menor presencia de ciertas tecnologías energéticas en la región. A excepción de este caso particular, el resto de comunidades tienen un volumen suficiente de datos como para ser analizadas y comparadas entre sí. Esto garantiza una base sólida y representativa para los análisis posteriores de generación energética por región.

## Proyectos II, integración y preparación de datos



### 3.2 Procesamiento de los datos de balance energético

Tras finalizar el análisis exploratorio del conjunto de datos de generación energética, el siguiente paso del proyecto consiste en trabajar con los datos de **balance energético** a nivel regional, también proporcionados por Red Eléctrica de España (REE).

Para ello, se han **descargado individualmente los ficheros de balance mensual** correspondientes a las 17 comunidades autónomas seleccionadas, excluyendo Ceuta y Melilla por los motivos previamente explicados. Cada uno de estos archivos sigue un formato similar al utilizado en los datos de generación, lo que permite aplicar el mismo proceso de limpieza y transformación.

## Proyectos II, integración y preparación de datos

---

Concretamente, a cada dataset se le aplicarán las funciones desarrolladas anteriormente para:

- Eliminar filas irrelevantes (encabezados, notas finales).
- Reestructurar los datos al **formato largo**, donde cada fila representa una combinación única de tipo de variable energética, comunidad autónoma, mes y valor.
- Convertir el campo de mes en una **variable de tipo fecha**, facilitando el análisis temporal.
- Estandarizar los nombres de las columnas y el formato numérico.
- Guardar cada dataset limpio en un directorio específico (data/balance), con una nomenclatura estandarizada.

Una vez procesados los 17 datasets individuales, se procederá a **unificarlos en un único conjunto de datos** completo, con el objetivo de facilitar los análisis agregados y comparativos por comunidad, tecnología y periodo temporal, tal como se hizo en la fase anterior con los datos de generación.

### Análisis Exploratorio para Balance

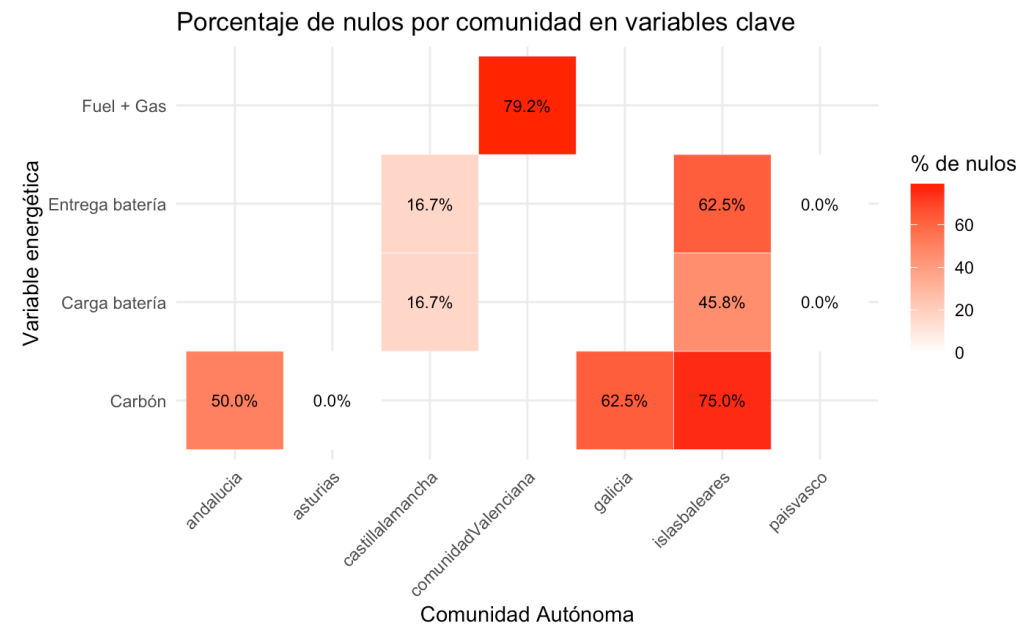
En términos generales, el dataset muestra una **alta calidad**, con una proporción baja de valores faltantes. Solo unas pocas tecnologías energéticas presentan registros ausentes:

- **Carbón:** 45 valores nulos
- **Entrega batería:** 19 nulos
- **Fuel + Gas:** 19 nulos
- **Carga batería:** 15 nulos
- **Eólica:** 13 nulos

A nivel territorial, los valores nulos se concentran principalmente en:

- **Islas Baleares:** 68 nulos
- **Comunidad Valenciana:** 20 nulos
- **Galicia:** 15 nulos
- **Andalucía:** 12 nulos

## Proyectos II, integración y preparación de datos



Tras analizar el porcentaje de valores nulos en las variables con más registros faltantes, se ha tomado la decisión de descartar las variables **"Fuel + Gas"** y **"Carbón"**. Ambas presentan un número elevado de datos ausentes distribuidos en múltiples comunidades, lo que limita considerablemente su valor analítico. Además, su presencia en el mix energético actual es reducida o prácticamente nula en muchas regiones, lo que refuerza la decisión de eliminarlas del análisis.

Por otro lado, las variables **"Entrega batería"** y **"Carga batería"** muestran un número moderado de registros faltantes y una cantidad total de observaciones suficientemente representativa ( $n = 72$  cada una). Estas variables son relevantes para estudiar la evolución del almacenamiento energético en el sistema eléctrico, por lo que se mantendrán en el análisis posterior.

### Análisis de tecnologías energéticas poco frecuentes



## Proyectos II, integración y preparación de datos

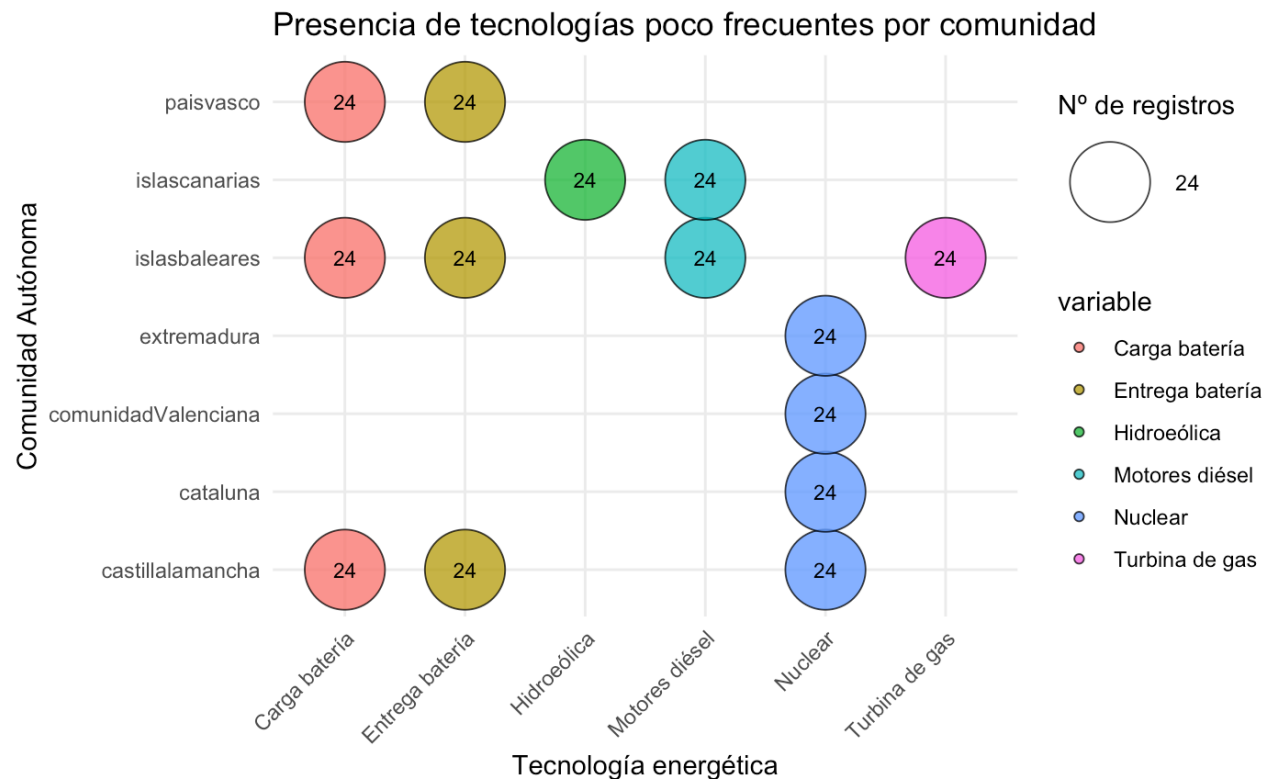
---

Para completar el análisis exploratorio del dataset de balance energético, se ha evaluado la presencia de tecnologías con bajo número de registros en el conjunto de datos. Estas tecnologías incluyen: *Nuclear*, *Carga batería*, *Entrega batería*, *Motores diésel*, *Hidroeólica* y *Turbina de gas*.

Como se aprecia en el gráfico de burbujas, estas variables presentan una distribución geográfica muy limitada:

- **Carga batería y Entrega batería** aparecen únicamente en tres comunidades autónomas (*Castilla-La Mancha*, *Islas Baleares* y *País Vasco*), con exactamente 24 registros en cada una. Pese a su limitada cobertura, se ha decidido **mantenerlas** en el análisis por tratarse de tecnologías emergentes asociadas al almacenamiento energético, relevante para la transición energética.
- **Nuclear** solo está presente en *Cataluña*, *Comunidad Valenciana*, *Castilla-La Mancha* y *Extremadura*, que coinciden con la ubicación real de las centrales nucleares españolas. Por tanto, a pesar de su concentración regional, se considera **estratégica** y se mantendrá en el análisis.
- **Motores diésel, Hidroeólica y Turbina de gas** están exclusivamente presentes en *Islas Canarias* o *Islas Baleares*, y únicamente con 24 registros cada una. Estas tecnologías representan soluciones muy específicas y localizadas, típicas de sistemas insulares o de respaldo energético. Debido a su escasa representación y a que no aportan valor generalizable al análisis nacional, se ha decidido **eliminarlas del estudio principal**.

## Proyectos II, integración y preparación de datos



Tras el análisis de calidad y representatividad del dataset de balance energético, se ha decidido conservar **19 variables** que reflejan de forma adecuada el comportamiento energético a nivel regional, evitando aquellas con cobertura geográfica muy limitada o un número excesivo de valores nulos.

Las variables seleccionadas son:

**Tecnologías renovables:**

## Proyectos II, integración y preparación de datos

---

- Hidráulica, Eólica, Solar fotovoltaica, Solar térmica, Otras renovables, Residuos renovables y Generación renovable

### **Tecnologías no renovables:**

- Ciclo combinado y Cogeneración y Residuos no renovables y Nuclear y Generación no renovable

### **Variables de balance y almacenamiento**

- Demanda en b.c.: Demanda en barra de central; refleja el consumo energético neto.
- Saldo de intercambios: Diferencia entre la energía importada y exportada entre comunidades o países.
- Saldo almacenamiento: Resultado neto entre la energía cargada y descargada de sistemas de almacenamiento.
- Consumo bombeo: Energía utilizada para subir agua en sistemas de almacenamiento hidroeléctrico (bombeo).
- Turbinación bombeo: Energía generada al liberar el agua almacenada mediante bombeo.
- Carga batería: Energía almacenada en sistemas de baterías eléctricas.
- Entrega batería: Energía descargada de sistemas de baterías al sistema eléctrico.

Estas 19 variables constituyen la base definitiva sobre la que se realizarán los análisis posteriores del proyecto, permitiendo integrar la dimensión de consumo, generación y almacenamiento energético en las comunidades autónomas de España.

### **3.3 OBTENCIÓN DE DATOS METEOROLÓGICOS (AEMET)**

Para complementar los análisis energéticos realizados hasta ahora, incorporaremos datos climáticos provenientes de la Agencia Estatal de Meteorología (AEMET). Esta información es especialmente relevante para estudiar cómo variables como la temperatura, la radiación solar o la velocidad del viento pueden influir en la generación y el consumo de energía por comunidad autónoma.

A diferencia de los conjuntos de datos anteriores, los ficheros descargados de la AEMET se encuentran en formato **JSON**, por lo que será necesario transformarlos en **data frames** para poder integrarlos adecuadamente en nuestro flujo de trabajo en R.

Este proceso incluye:

- Descarga de datos desde la API de AEMET (ya autenticados con nuestra API Key).

## Proyectos II, integración y preparación de datos

---

- Conversión de los archivos JSON a estructuras tabulares.
- Limpieza, selección de variables clave y estandarización del formato temporal.
- Unión posterior con los datasets energéticos previamente procesados.

Con esta integración, podremos realizar análisis más completos que combinen información energética y climática, lo cual es clave para entender patrones estacionales, anomalías y correlaciones relevantes para el sistema eléctrico nacional.

Como pasaba con la REE, la API de AEMET no deja sacar directamente por todas las comunidades a la vez las variables meteorológicas, con lo cual tendremos que hacer la selección en base a lo que queramoslas cruzar, en este caso, vamos a aprovechar y descargaremos datos de valencia, de 2023 y 2024.

Una vez unidos los datasets climáticos de Valencia correspondientes a los años 2023 y 2024, contamos con un conjunto de datos estructurado que incluye **24 variables** de carácter meteorológico. A continuación, se describen brevemente las columnas disponibles en este dataset:

- **fecha:** mes y año de la observación (formato YYYY-MM-01), estandarizado para su cruce con los datos energéticos.
- **indicativo:** código identificador de la estación meteorológica (en este caso, el correspondiente a Valencia).
- **p\_max:** precipitación máxima diaria registrada durante el mes (en mm).
- **n\_gra:** número de días con granizo.
- **n\_fog:** número de días con niebla.
- **inso:** número medio de horas de sol al día.
- **tm\_min:** temperatura media de las mínimas (°C).
- **ta\_max:** temperatura absoluta máxima registrada (°C).
- **ts\_min:** temperatura media de la superficie mínima (°C).
- **nt\_30:** número de noches tropicales (temperatura mínima > 20°C).
- **np\_100:** número de días con precipitación superior a 100 mm.
- **p\_sol:** porcentaje de insolación respecto al máximo posible.
- **np\_001:** número de días con precipitación mayor a 0.1 mm.
- **ta\_min:** temperatura absoluta mínima registrada (°C).
- **np\_300:** número de días con precipitación superior a 300 mm.

## Proyectos II, integración y preparación de datos

---

- **p\_mes**: precipitación total acumulada del mes (mm).
- **n\_llu**: número de días de lluvia.
- **n\_tor**: número de días con tormenta.
- **nt\_00**: número de noches de helada.
- **ti\_max**: temperatura media de las máximas (°C).
- **n\_nie**: número de días con nieve.
- **tm\_mes**: temperatura media mensual (°C).
- **tm\_max**: temperatura media de las máximas (°C).
- **np\_010**: número de días con precipitación mayor a 10 mm.

Por sorpresa, la calidad de los datos es perfecta y no encontramos ningún dato faltante, con lo que podremos usar todas las variables para próximos análisis.

Una vez transformados y depurados los datos meteorológicos de Valencia procedentes de AEMET, procedemos a integrarlos con el dataset energético de generación eléctrica para la Comunidad Valenciana. Este cruce nos permitirá analizar cómo influyen las condiciones climáticas sobre la producción de diferentes tecnologías energéticas en dicha comunidad.

El proceso seguido ha sido el siguiente:

- **Filtrado del dataset energético** para eliminar tecnologías poco representativas o con una alta proporción de valores nulos, como "Fuel + Gas", "Carbón", "Motores diésel", "Hidroeólica" y "Turbina de gas" y eliminación de la columna comunidad, ya que en esta fase solo trabajamos con datos de la Comunidad Valenciana y esta variable resulta redundante.
- **Conversión de las fechas a formato estándar (YYYY-MM-DD)** en ambos conjuntos de datos.
- **Cruce (join) entre ambos datasets por la variable fecha**, de manera que cada registro de generación energética mensual quede enriquecido con las condiciones climáticas registradas en el mismo mes.

El resultado es un nuevo conjunto de datos llamado `df_valencia_completo`, que contiene información combinada de generación por tecnología y las variables meteorológicas correspondientes a cada mes. Esta integración será clave para realizar análisis conjuntos y estudiar posibles correlaciones o patrones estacionales entre clima y producción energética.

## Proyectos II, integración y preparación de datos

variable <chr>	valor <dbl>	fecha <date>	indicativo <dbl>	p_max <dbl>	n_gra <dbl>	n_fog <dbl>	inso <dbl>	tm_min <dbl>	ta_max <dbl>	
Hidráulica	26944.908	2023-01-01	8416	0.6	0	0	7.1	6.4	23.2	
Nuclear	770792.370	2023-01-01	8416	0.6	0	0	7.1	6.4	23.2	
Ciclo combinado	129949.982	2023-01-01	8416	0.6	0	0	7.1	6.4	23.2	
Eólica	312480.340	2023-01-01	8416	0.6	0	0	7.1	6.4	23.2	
Solar fotovoltaica	39923.963	2023-01-01	8416	0.6	0	0	7.1	6.4	23.2	
Solar térmica	3880.120	2023-01-01	8416	0.6	0	0	7.1	6.4	23.2	
Otras renovables	2499.887	2023-01-01	8416	0.6	0	0	7.1	6.4	23.2	
Cogeneración	54826.250	2023-01-01	8416	0.6	0	0	7.1	6.4	23.2	
Residuos no renovables	1457.923	2023-01-01	8416	0.6	0	0	7.1	6.4	23.2	
Generación total	1342755.743	2023-01-01	8416	0.6	0	0	7.1	6.4	23.2	

1-10 of 15 rows | 1-10 of 26 columns

Previous **1** 2 Next

Una vez realizado el cruce entre los datos de generación energética y las variables climáticas de AEMET para la Comunidad Valenciana, disponemos ya de una base sólida que nos permitirá analizar de forma conjunta la influencia del clima sobre la generación de energía.

Este procedimiento puede replicarse fácilmente para el resto de comunidades autónomas, simplemente descargando los datos climáticos correspondientes desde la API de AEMET, transformándolos y uniéndolos con los datasets energéticos previamente limpiados. Del mismo modo, el análisis puede adaptarse seleccionando únicamente aquellas variables que resulten relevantes para los objetivos específicos del estudio.

Esta integración de fuentes permite avanzar hacia análisis más complejos y personalizados, como estudios de estacionalidad, correlación entre generación renovable y temperatura o radiación solar, o incluso modelos predictivos apoyados en datos meteorológicos. En definitiva, se ha dejado preparado un entorno de trabajo versátil y escalable que sienta las bases para cumplir los objetivos del proyecto de forma efectiva.

### **ANEXO CODIGO RMD**

Por motivos de extensión (600 líneas de código), no se incluye el código de rmd. El desarrollo está disponible si se requiere, se subirá un html en el que se especifica todo el código en R y todas las gráficas obtenidas