



Etsinf UPV

Ciencia de datos - Proyecto II

**Clima y kilovatios: el pulso
energético de España
comunidad a comunidad**

Índice

1. Alcance del proyecto.....	2
1.1. Objetivos del proyecto.....	2
1.2. Utilidad del estudio.....	2
2. Configuración del proyecto.....	3
2.1. Fuentes de datos.....	3
2.2. Integración y transformación de los datos.....	3
3. Resultados obtenidos (objetivos específicos).....	4
Objetivo 1.....	4
Objetivo 2.....	5
Objetivo 3.....	7
Objetivo 4.....	9
Objetivo 5.....	12
4. Lecciones aprendidas para mis futuros proyectos en Ciencia de Datos.....	16
5. Bibliografía.....	17
6. Anexos.....	17
6.1 Hito 1: Fuentes de datos.....	17
6.2 Hito 2: Integración y preparación de los datos.....	18
6.3 Objetivos Específicos.....	18
6.3.1 Objetivo 1.....	18
6.3.2 Objetivo 2.....	21
6.3.3 Objetivo 3.....	25
6.3.4 Objetivo 4.....	29
6.3.5 Objetivo 5.....	33

AUTORES

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

1. Alcance del proyecto

Análisis energético por comunidad en función de variables climáticas y estacionales en España.

Este proyecto estudia cómo influyen las condiciones climáticas y la estacionalidad en la generación y el balance energético en España, utilizando datos abiertos de Red Eléctrica de España (REE) y la Agencia Estatal de Meteorología (AEMET). A través de la integración de datos meteorológicos y energéticos, se han desarrollado análisis cruzados por comunidad autónoma y se han investigado relaciones específicas, como la correlación entre precipitaciones y energía hidráulica. Se ha seguido un flujo técnico riguroso: desde la extracción, limpieza e integración de datos, hasta el análisis exploratorio y la aplicación de modelos estadísticos.

1.1. Objetivos del proyecto

Los objetivos definidos para este proyecto son:

1. Estudiar la generación energética según la comunidad autónoma y las fechas del año.
2. Analizar la evolución general de las energías renovables en invierno y verano.
3. Evaluar el balance energético por comunidad.
4. Estudiar la generación de energía solar fotovoltaica de las comunidades en función del tiempo meteorológico u otras variables semejantes
5. Analizar la relación entre las precipitaciones mensuales y la energía hidráulica generada en comunidades seleccionadas, profundizando en Galicia y Castilla y León como casos representativos, y explorando patrones estacionales y posibles retardos en la respuesta del sistema hidráulico.

1.2. Utilidad del estudio

Este proyecto tiene una utilidad significativa en el plano académico, ciudadano y divulgativo, ya que permite entender mejor el funcionamiento del sistema energético en España a través del análisis de datos públicos reales.

A nivel general:

- **Contribuye a identificar patrones y relaciones entre factores climáticos, geográficos y de consumo/producción energética,** proporcionando una base empírica para investigaciones futuras, tanto en entornos educativos como técnicos.
- **Para la ciudadanía,** ofrece una visión accesible de cómo evoluciona la generación energética en las distintas comunidades autónomas, qué peso tienen las fuentes renovables, y cómo influyen factores como el clima o los hábitos de consumo. Esta

comprensión puede fomentar una mayor sensibilidad hacia el uso responsable de la energía y el papel de las renovables.

- **En el ámbito divulgativo o institucional**, el estudio muestra cómo se pueden aprovechar fuentes abiertas y herramientas estadísticas para explorar temas de interés público, como la transición energética, el aprovechamiento de recursos naturales o la desigualdad territorial en términos de producción y dependencia energética.

Con este trabajo, aunque no tenga una aplicación directa en la toma de decisiones políticas, sienta las bases para análisis más amplios que podrían alimentar proyectos de planificación energética regional, evaluación de sostenibilidad o educación ambiental y demuestra cómo la Ciencia de Datos puede servir como puente entre la comprensión técnica y la toma de decisiones informadas en contextos energéticos y medioambientales.

2. Configuración del proyecto

2.1. Fuentes de datos

Para el desarrollo del proyecto se han integrado **dos fuentes de datos principales**, ambas públicas y oficiales:

- **Red Eléctrica de España (REE)** [<https://www.ree.es/es/datos/aldia>]: proporciona datos mensuales de generación eléctrica por comunidad autónoma, desglosados por tipo de fuente energética (hidráulica, eólica, solar, etc.). Estos datos han sido fundamentales para estudiar la producción energética en cada territorio y han sido utilizados en la mayoría de los objetivos.
- **Agencia Estatal de Meteorología (AEMET)** [<https://opendata.aemet.es/centrodedescargas/productosAEMET?>]: ofrece datos meteorológicos mensuales por estación, incluyendo variables como precipitación total, número de días de lluvia o temperatura máxima. Ha sido esencial para los objetivos centrados en analizar la influencia del clima en la generación de energía renovable.

Ambas fuentes cumplen los criterios de acceso libre, fiabilidad y periodicidad adecuada, permitiendo su análisis a nivel temporal y territorial.

Las fichas detalladas de cada fuente están incluidas en el [Anexo \(HITO I\)](#).

2.2. Integración y transformación de los datos

Para garantizar la coherencia entre fuentes y facilitar el análisis conjunto, se ha seguido un flujo de trabajo homogéneo de transformación tanto para los datos energéticos como climáticos.

En primer lugar, se descargaron manualmente 17 archivos CSV de generación y balance energético mensual desde la web oficial de Red Eléctrica de España (REE), uno por cada comunidad autónoma. A cada fichero se le aplicaron procesos de limpieza: eliminación de metadatos y notas, reestructuración al formato largo (long), transformación de fechas, normalización de formatos numéricos y estandarización de nombres. Posteriormente, se unificaron en un único dataset general `generacion_todas_comunidades.csv` para los datasets de generación y balance `todas_comunidades.csv` para las de balance, que recoge el tipo de energía, comunidad autónoma, valor mensual y fecha, permitiendo un análisis temporal y territorial comparado.

En paralelo, se descargaron datos meteorológicos desde la API de AEMET, en formato JSON, usando una API Key individual para cada estación seleccionada. Estos JSON (o .txt) se transformaron en data frames estructurados y limpios, seleccionando variables relevantes (precipitaciones, temperatura, insolación, etc.) y estandarizando el formato de fecha. Finalmente, estos datos climáticos se integraron con los energéticos a través de la variable fecha, permitiendo analizar correlaciones y patrones conjuntos.

Este proceso se ha documentado detalladamente en el [Anexo \(Hito II\)](#), donde se presentan las transformaciones clave, scripts en R y los primeros análisis exploratorios.

3. Resultados obtenidos (objetivos específicos)

Objetivo 1

Estudiar la generación total dependiendo de las fechas del año y la comunidad autónoma.

Para este primer objetivo, se han utilizado los datos de generación eléctrica de Red Eléctrica de España (REE) ya extraídos y procesados previamente en el hito 2 (`generacion_todas_comunidades.csv`). Tras tener los datos listos, creamos variables temporales clave para el objetivo como son: día de la semana (dow), mes (month), año (year) y semana ISO (week) y construimos `daily_gen` (la generación diaria total por comunidad) y `monthly_gen` (agregada por mes de inicio). Para responder a este objetivo, se plantean tres preguntas: "¿cuál es la comunidad que genera más energía?", "¿Qué períodos del año (meses, semanas o días) presentan una generación energética más alta?" y "¿En qué momento la comunidad que más produce alcanza su máxima generación energética?"

Para responder a la primera pregunta, examinamos el gráfico de generación anual total por comunidad ([Figura 5.1.1.1](#)). En él, Andalucía y Cataluña destacan con crecimientos interanuales continuos, situándose siempre por encima del resto. Esta supremacía se explica tanto por su tamaño y capacidad instalada como por la diversidad de fuentes renovables disponibles (solar en Andalucía, eólica e hidráulica en Cataluña).

CURSO 2024-2025

Para analizar los períodos de mayor generación a lo largo del año y de la semana, recurrimos a dos visualizaciones complementarias. El mapa de calor semanal de generación ([Figura 5.1.1.2](#)) que muestra el patrón estacional: las semanas 25–35 (finales de junio a mediados de septiembre) alcanzan los valores más altos, aprovechando la máxima radiación solar, es decir, picos de energía fotovoltaica, mientras que en invierno (semanas 1–10 y 50–52) las tonalidades son más suaves, reflejo de menor producción energética. Además, dentro de cada semana, los días laborables (lunes a viernes) aparecen con un relleno consistentemente más intenso que sábados y domingos, confirmando que la demanda industrial y comercial impulsa la producción en el tramo medio de la semana.

Finalmente, centrándonos en Cataluña (la comunidad más productiva) trazamos su serie diaria de generación ([Figura 5.1.1.3](#)) y destacamos el punto de máximo con un marcador rojo. Tras agrupar "daily_gen" por fecha y seleccionar el valor más alto, descubrimos que el 15 de agosto de 2023 se registró el pico de 7.35 GWh. Este hecho coincide con pleno verano, cuando las condiciones climáticas (alta radiación solar, vientos estables y niveles óptimos en sistemas hidráulicos) permiten que las energías renovables alcancen su máximo anual.

A estas conclusiones sobre volúmenes y ciclos se suman ahora hallazgos “curiosos” basados en el cálculo de densidad energética (MWh/km²) ([Figura 5.1.1.4](#) y [Figura 5.1.1.5](#)). Con cerca de 2.335 MWh/km², Cataluña no solo es la más productiva en términos absolutos, sino que, por kilómetro cuadrado, duplica a gigantes como Andalucía (695 MWh/km²) o Castilla-La Mancha (774 MWh/km²). Le siguen Illes Balears (1 800 MWh/km²), Asturias (1 467 MWh/km²), Galicia (1 561 MWh/km²), Extremadura (1 495 MWh/km²), Murcia (1 428 MWh/km²), Aragón (917 MWh/km²) y La Rioja (776 MWh/km²). En el extremo opuesto, Madrid (302 MWh/km²) y Cantabria (201 MWh/km²) evidencian modelos más orientados al consumo que a la generación, con baja capacidad instalada por unidad de territorio.

En conjunto, estos resultados ponen sobre la mesa que el análisis de la generación energética en España no puede limitarse al volumen absoluto: la densidad energética aporta una nueva dimensión, revelando qué territorios sacan el máximo partido a su extensión. Así, mientras Andalucía (695 MWh/km²) y Cataluña (2 335 MWh/km²) lideran en volumen y densidad, comunidades más pequeñas o de geografía singular demuestran que una estrategia renovable concentrada puede convertir kilómetros cuadrados modestos en verdaderos motores energéticos.

Objetivo 2

Estudio general de las energías renovables, destacando invierno/verano

Lo que se busca con este segundo objetivo es entender, a grandes rasgos, el comportamiento de la generación de energía renovable en nuestro país, conocer si existen correlaciones entre los diferentes tipos y a qué se pueden deber. Además, se pretende subrayar cómo varía la generación según la época del año, ya que, como es bien sabido, este tipo de energías depende bastante del clima, aunque este análisis se desarrollará en los próximos objetivos. Con estos gráficos de barras observamos cómo se distribuye la

CURSO 2024-2025

generación de energías renovables a lo largo de las diferentes estaciones. La [figura 6.3.2.1](#) muestra la producción de cada energía renovable por separado, donde se observa un claro aumento de la producción de energías relacionadas con el sol (solar fotovoltaica y térmica) a medida que se aproxima el clima veraniego. También se puede ver como la generación de energía eólica destaca en invierno y se reduce significativamente en verano, lo que es lógico por las bajadas de las rachas de viento y el aumento de las energías solares. La producción de energía hidráulica es más irregular pero también se reduce en verano. En la figura [6.3.2.2](#) se muestra la regularidad en la producción de energía renovable en las distintas estaciones. En la figura 6.3.2.1 se observa un aumento significativo en la generación de solar fotovoltaica en 2024 respecto a 2023, que se ve reflejado en la figura [6.3.2.2](#) en la estación de verano. En primavera pasa lo mismo pese a que haya una disminución en la producción de energía eólica, ya que aumenta ligeramente la fotovoltaica y aumenta bastante la hidráulica. Estas conclusiones nos dan a pensar que la primavera de 2023 ha sido mucho más seca que la de 2024 y que el verano de 2024 ha sido más soleado que el de 2023.

Otro estudio que hemos realizado trata de analizar si existen correlaciones entre energías renovables para afectar cómo puede influir los diferentes climas.

A nivel nacional ([Figura 6.3.2.3](#)), la energía solar fotovoltaica y la térmica presentan una correlación muy alta, lo cual es lógico ya que ambas dependen de la radiación solar. La energía eólica muestra correlación positiva con la hidráulica, posiblemente porque ambas pueden incrementarse en condiciones meteorológicas con más viento, lluvia o temporales. Las otras renovables están positivamente correlacionadas con las solares y negativamente con la eólica e hidráulica, lo que sugiere que su producción es más estable en climas secos o cálidos. Algo a tener en cuenta es que estas correlaciones cambian dependiendo de la comunidad autónoma. Voy a tratar solo las cinco comunidades que más generan ([Figura 6.3.2.4](#)) y la Comunidad Valenciana. La mayor diferencia con 2023 ([Figura 6.3.2.5](#)) es que las correlaciones entre hidráulica y las solares son aún menores, lo que podría indicar una menor coincidencia temporal entre días soleados y períodos de lluvia en 2023.

En Castilla y León ([Figura 6.3.2.9](#)), existe una correlación positiva entre eólica e hidráulica y una correlación muy negativa entre la hidráulica y la solar fotovoltaica, al igual que pasa en Galicia ([Figura 6.3.2.10](#)), por el tipo de clima nuboso y de lluvias, con ausencia de sol en esas épocas.

Por contraparte, tenemos Andalucía ([Figura 6.3.2.6](#)) y la Comunidad Valenciana ([Figura 6.3.2.11](#)) donde existe una correlación positiva entre la producción de energía hidráulica y la de solares (fotovoltaica y térmica) y tienen una correlación negativa entre hidráulica y eólica. Esto puede ser debido a un clima más caluroso, donde las lluvias van acompañadas de sol.

En cuanto a Aragón ([Figura 6.3.2.7](#)), no existen correlaciones demasiado fuertes y la correlación entre eólica e hidráulica pasa de ser ligeramente positiva en 2023 a ligeramente negativa en 2024, todo esto nos da a pensar que el clima de Aragón puede no ser tan diferenciado.

Para finalizar tenemos Castilla-La Mancha ([Figura 6.3.2.8](#)), donde no existe correlación entre la producción de eólica y la hidráulica. En 2023 hay correlación negativa entre la

CURSO 2024-2025

hidráulica y las solares pero en 2024 esta correlación desaparece, puede deberse a que en 2023 hizo un temporal donde las épocas de días lluviosos no van acompañados de sol, días cubiertos, cosa que no pasaría en 2024, donde habría más presencia de días soleados que hagan desaparecer esta correlación.

En las figuras [6.3.2.12](#) y [6.3.2.13](#) se puede ver como la mayor producción de hidráulica es en Galicia y Castilla y León y de solar fotovoltaica en Andalucía, Castilla-La Mancha, Aragón y Extremadura. Con esto estamos más cerca de afirmar la diferencia de climas entre estas comunidades.

Todas estas suposiciones serán estudiadas más adelante en el [objetivo 4](#) y el [objetivo 5](#), donde se trabajará con datos climáticos.

[Objetivo 3](#)

Estudiar el balance energético por comunidad

El propósito de este tercer análisis es determinar el nivel de autosuficiencia energética de cada comunidad autónoma en España, entendida como la capacidad de cada región para cubrir su propia demanda energética mediante su generación. Esto permite identificar qué territorios producen más energía de la que consumen (y por tanto pueden exportar), cuáles son autosuficientes y cuáles, en cambio, presentan un déficit estructural de generación. Para realizar este estudio se ha trabajado con dos conjuntos de datos principales: generacion_todas_comunidades.csv que incluye la energía generada mensualmente por comunidad autónoma y balance_todas_comunidades.csv que contiene los datos de consumo energético mensual de cada comunidad autónoma en el mismo periodo. Ambos conjuntos de datos abarcan los años 2023 y 2024, con periodicidad mensual y desagregación territorial por comunidad autónoma.

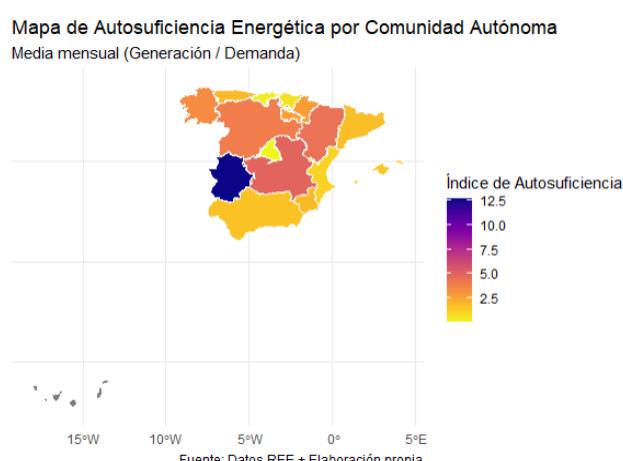


Figura 3.2.1

azulados oscuros corresponden a regiones altamente autosuficientes, con índices muy

En la figura (3.2.1) se muestra un mapa coroplético de España en el que se representa, mediante una escala de colores, el valor del índice de autosuficiencia energética medio mensual para cada comunidad autónoma. El uso del color es fundamental para la interpretación: los tonos amarillos representan bajos niveles de autosuficiencia (índices cercanos a 1 o menores), lo que indica una alta dependencia energética externa. Los tonos naranjas y rojizos indican un nivel medio de autosuficiencia, donde la generación cubre una buena parte de la demanda. Finalmente, los tonos morados y

CURSO 2024-2025

superiores a 1, que generan un excedente significativo. Las conclusiones se encuentran en el [Anexo 6.3.3.](#)

Con el objetivo de profundizar en el análisis de la autosuficiencia energética, se ha desarrollado una aplicación interactiva en R Shiny que permite visualizar y comparar el índice de autosuficiencia para cada una de las principales fuentes de energía renovable por comunidad autónoma. Esta herramienta ofrece una perspectiva más granular, facilitando la identificación de aquellas regiones que presentan un mayor grado de independencia energética en función de la tecnología utilizada. Para ello, se han considerado únicamente las siguientes fuentes renovables: hidráulica, solar fotovoltaica, solar térmica, eólica y otras renovables.

A partir de los datos de generación eléctrica por fuente y la demanda energética total de cada comunidad y año, se ha calculado un índice de autosuficiencia renovable (Índice de autosuficiencia=Generación de una fuente concreta/Demanda total de la comunidad). Este cociente permite evaluar en qué medida una comunidad autónoma podría cubrir su demanda energética únicamente a través de una fuente renovable específica. El mapa resultante, generado dinámicamente, permite seleccionar tanto el año como el tipo de energía, y visualizar en un único gráfico qué comunidades presentan una mayor autosuficiencia. Aquellas con un índice más alto están más cerca de cubrir su consumo con esa fuente concreta, mientras que las de menor valor son más dependientes de otras tecnologías o de importaciones externas.

Este estudio ([Anexo 6.3.3](#)) confirma que la autosuficiencia renovable regional es altamente desigual y dependiente del tipo de recurso. Algunas comunidades presentan claros avances y potencial (Castilla y León, Galicia, Extremadura), mientras que otras todavía dependen en gran medida de fuentes no renovables o importadas. Esta disparidad refuerza la necesidad de: Fomentar la diversificación tecnológica para equilibrar las limitaciones de cada fuente. Diseñar políticas energéticas adaptadas al perfil de cada comunidad. Promover el desarrollo de redes inteligentes y almacenamiento para gestionar la intermitencia y excedentes e impulsar el autoconsumo y la generación distribuida, especialmente en regiones con baja autosuficiencia.

En definitiva, la autosuficiencia energética es un objetivo viable sólo si se aborda de forma integrada, combinando múltiples fuentes renovables y adaptando las estrategias a las características territoriales de cada comunidad. Este estudio constituye una herramienta valiosa para identificar fortalezas, debilidades y oportunidades en la transición energética hacia un modelo más resiliente, sostenible y descentralizado.

Objetivo 4

Estudiar la generación de energía solar fotovoltaica de las comunidades en función del tiempo meteorológico u otras variables semejantes

Para este penúltimo objetivo se ha trabajado con los datos de la AEMET y con el .CSV (generacion_todas_comunidades.csv) generado por los datos de la red eléctrica. Este objetivo al estar centrado en la energía solar fotovoltaica se ha filtrado las variables de

CURSO 2024-2025

interés, con una función, que son todas las relacionadas con las temperaturas y relacionadas con el estado del dia ("fecha", "inso", "tm_min", "ta_max", "p_sol", "ta_min", "ti_max", "tm_mes", "tm_max", "n_cub", "n_des", "n_nub") y hay un atributo adicional creado "diasnublados" cuyos valores es la suma de las variables "n_nub" y "n_cub".

Los datos de la AEMET nos daban las temperaturas por provincia y no por comunidad lo que nos fastidiaba porque nuestro estudio es por comunidad. La solución que hemos propuesto es que mirábamos las temperaturas de todas las provincias que tenía esa comunidad, y si tenían la misma temperatura media mensual decidimos que coger una provincia en representación de las que tienen la misma temperatura. Evidentemente, cada provincia de una comunidad no tiene la misma temperatura por lo que cuando era muy distinta la temperatura hacíamos la media de las variables de interés filtradas de las provincias con distinta temperatura.

Este objetivo se ha respondido gracias a tres preguntas que considerábamos que podían responder y generalizar bien el objetivo:

- 1) ¿Qué comunidad genera más energía por hora y cuál genera más al año? ¿Tiene relación con la autosuficiencia de una comunidad?
- 2) ¿Contra más temperatura y horas de insolación, más energía fotovoltaica se genera?
- 3) ¿Qué relación hay entre la generación fotovoltaica y el tiempo que hace ese día ?

1º pregunta:

En este objetivo se realiza un pequeño análisis exploratorio para ver si es posible sacar alguna relación con la autosuficiencia de las comunidades.

Se puede analizar, con respecto al mapa de autosuficiencia ([figura 6.1.4.1](#)), que Extremadura era la comunidad con más autosuficiencia y eso se debe a que es una de las comunidades que más producen por hora y no es muy grande esa comunidad. Si vemos Castilla La Mancha tiene menos autosuficiencia que Extremadura y en cambio tiene más producción por hora pero es obvio porque que generen más no significa que sean más autosuficientes porque consumen más al ser una comunidad más grande. Por lo que Extremadura aun teniendo un territorio mucho más reducido que otras comunidades su producción por hora es casi la misma que la de Castilla La Mancha y su consumo es menor por eso su autosuficiencia es tan elevada.

2º pregunta:

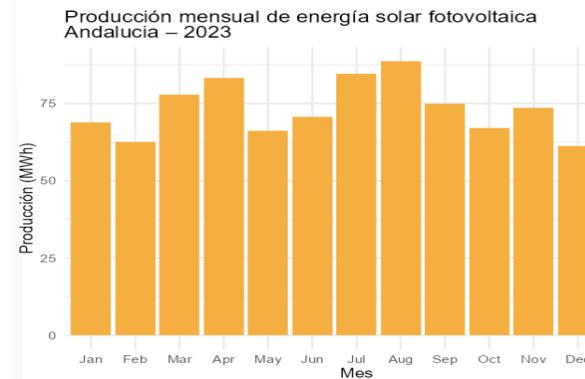
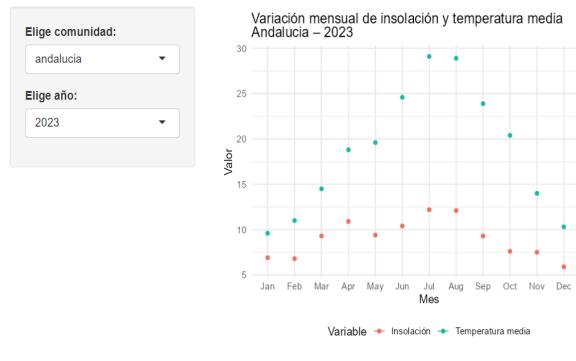
Hacemos uso de la librería “shiny” para implantar un segmentador entre comunidades y año. Este es un ejemplo de los resultados de la aplicación:

Figura 4.2.1

Figura 4.2.2

CURSO 2024-2025

Dashboard Energía Solar por Comunidad | 2023–2024



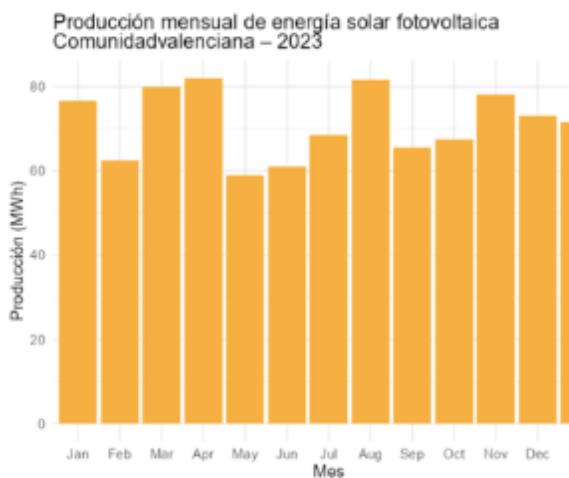
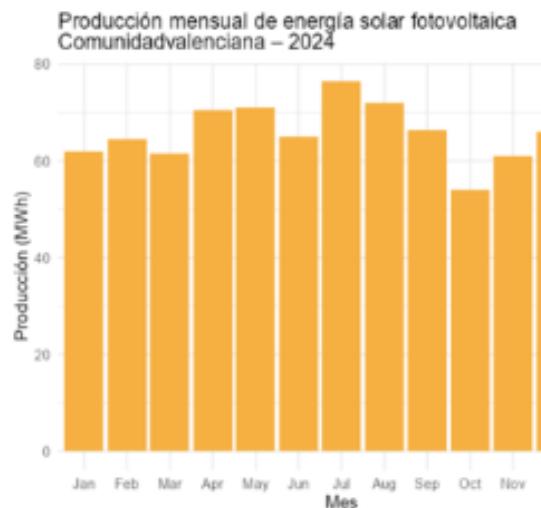
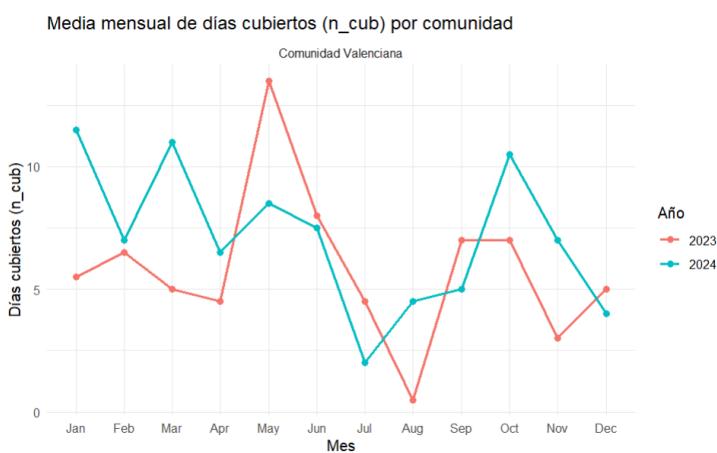
Con estos gráficos se observa cómo se comportan las variables de ‘inso’ y ‘tm_mes’. Se puede observar que la variable ‘tm_mes’ siempre tiene la misma distribución, es decir los meses de verano (junio, julio, agosto) siempre suelen tener las temperaturas más altas y pasa lo contrario en invierno (diciembre, enero, febrero) que tienen las temperaturas más bajas.

Con respecto a la variable ‘inso’ vemos que el patrón que sigue en el gráfico de representación de esta variable es casi el mismo que tiene el histograma de producción de energía solar por mes. Cada vez que el nivel de insolación baja la producción de ese mes baja también.

Podemos decir en este caso que la producción de energía solar no depende de la temperatura del ambiente sino que depende de las horas de sol en ese mes, de hecho si la producción dependiera de la temperatura el histograma de la producción solar sería todas parecidas y los meses con mayor producción serían los meses de verano pero no es el caso ya que como se observa en figura 4.2.1 y 4.2.2, en marzo y abril se genera más energía solar que en junio y julio. Ver otro ejemplo ([Anexo: figura 6.1.4.2](#))

3º pregunta:

Para llevar a cabo este objetivo, se ha realizado una filtración de los datos de la AEMET, cogiendo variables como n_cub (nº de días cubiertos), n_des (nº de días despejados), n_nub (número de días nublados) y, a parte, creado una variable ‘diasnublados’ cuyos valores son la suma de los días que están nublados o cubiertos guardado en un CSV ([Anexo: figura 6.1.4.3](#)). Para responder esta pregunta vamos a realizarlo para tres comunidades: Comunidad Valenciana, Asturias y Andalucía, tres comunidades muy separadas geográficamente para ver como es la situación de España grosso modo.

Figura 4.3.1**Figura 4.3.2****Figura 4.3.3**

Analizando los gráficos, se ve como contra más días cubiertos tenga el mes menos producción genera ese mes de energía fotovoltaica. Además se puede ver por ejemplo en 2023 como el mes con más días cubiertos (abril) es el mes que menos genera en todo el año, es decir que están correlacionadas negativamente.

Figura 4.3.4

(Pearson 2023): -0.039

Figura 4.3.5

(Pearson 2024): -0.208

De hecho las correlaciones de Pearson, corroboran los resultados obtenidos con los gráficos anteriores.

Análisis extra y su explicación en:

- [Anexo: objetivo 6.3.4, tercera pregunta](#)
- [Información adicional, anexo: figura 6.1.4.10](#)

Objetivo 5

Estudiar la relación entre las precipitaciones mensuales y la generación de energía hidráulica en Galicia, analizando datos climáticos y energéticos para detectar patrones, correlaciones y comportamientos estacionales.

Para el desarrollo de este último objetivo, se ha trabajado con dos fuentes oficiales y complementarias: los datos meteorológicos mensuales de la Agencia Estatal de Meteorología (AEMET) y los datos de generación eléctrica de Red Eléctrica de España (REE) ya extraídos y procesados previamente en el hito 2 y mencionado en anteriores objetivos: generacion_todas_comunidades.csv. AEMET proporciona información detallada sobre precipitaciones y otras variables climáticas en estaciones representativas de cada comunidad. Esta integración es esencial para estudiar técnicamente la relación entre el clima y la energía renovable.

Como primer paso, se ha realizado un análisis exploratorio de la generación hidráulica mensual por comunidad autónoma con el fin de identificar cuáles son las regiones más representativas para este estudio.

El análisis revela que Galicia y Castilla y León destacan claramente por su alta producción total acumulada durante el periodo 2023–2024. Esto justifica la selección de estas dos comunidades como casos de estudio para el análisis detallado ([Figuras](#))

Ahora bien, los datos meteorológicos no están disponibles directamente a nivel autonómico. Por tanto, al igual que en otros objetivos del proyecto, ha sido necesario seleccionar estaciones meteorológicas representativas para cada comunidad:

CURSO 2024-2025

- **Galicia:** se ha seleccionado la estación **A Coruña Aeropuerto** por su continuidad histórica, buena cobertura mensual, y representatividad del clima atlántico húmedo.
- **Castilla y León:** se ha utilizado la estación **León – Virgen del Camino**, que presenta datos completos, buena calidad y una localización estratégica dentro de la comunidad.

*Para futuros análisis más exhaustivos, podría considerarse incluir otras estaciones o una media ponderada regional, pero en este caso el enfoque era identificar tendencias y contrastar comunidades

Una vez seleccionadas las estaciones, se ha limpiado y transformado la información climática obtenida desde los archivos JSON mensuales de AEMET para los años 2023 y 2024. El preprocesado incluyó la eliminación de registros inválidos (como filas con fecha mes=13), la extracción de valores numéricos desde campos como p_max = 20.9(19) (donde se descarta el día del evento), y la estandarización del campo fecha.

Las variables extraídas de AEMET han sido principalmente: **p_mes**: precipitación mensual total (mm), **n_llu**: número de días de lluvia, **np_010**: número de días con precipitación >10 mm, **p_max**: precipitación máxima diaria.

Posteriormente, se han combinado estos datos climáticos con la serie mensual de generación hidráulica de REE filtrando por comunidad autónoma. Se han generado nuevas variables como:

- **p_mes_escalado**: precipitación mensual transformada para representar junto a energía en un gráfico de doble eje.
- **estación**: variable categórica derivada del mes (Invierno, Primavera, etc.).

El análisis exploratorio ha revelado patrones interesantes y se ha centrado mayoritariamente en la comunidad de Galicia, pues es la que más valores presentaba y donde se podía ver más claro las relaciones y gráficos. En las [Figura](#) se observa una sincronicidad clara entre los meses más lluviosos (invierno y primavera) y los de mayor producción hidráulica . En cambio, en Castilla y León [Figura](#), esta relación visual no es tan evidente.

De forma interactiva con la librería plotly [Figura](#), se puede observar la evolución de ambas variables a lo largo del tiempo

[¿Qué pasa en ene–mar 2024?](#)

Aunque las precipitaciones bajan (línea verde), la energía hidráulica se mantiene alta (línea azul). Esto ocurre porque los embalses estaban ya llenos desde los meses anteriores (nov–dic 2023), que sí fueron muy lluviosos. La energía hidráulica no responde al instante a la lluvia caída, sino al agua embalsada acumulada.

[¿Y por qué en sep–oct 2024, aunque sube la precipitación, la hidráulica no sube igual?](#)

Aquí lo más probable es que: Los embalses aún estén vacíos tras un verano muy seco (jun–ago 2024) y el agua caída en septiembre no sea suficiente aún para restablecer niveles

CURSO 2024-2025

de producción (hay un “retraso” en la respuesta). Osea que aunque existe una correlación general entre precipitaciones y generación hidráulica (ahora se demuestra más abajo), el análisis mensual muestra que el sistema presenta cierta inercia hidrológica, de forma que la producción no responde inmediatamente a cada pico de lluvia.

Dicho esto, para cuantificar estas relaciones, se ha calculado el coeficiente de correlación de Spearman entre variables climáticas y la energía generada. Se ha utilizado el coeficiente de Spearman porque no se asume una relación lineal ni una distribución normal entre precipitaciones y energía hidráulica. Spearman permite detectar relaciones monótonas y es más robusto ante valores extremos o distribuciones irregulares. Se han descartado Pearson por exigir normalidad y Kendall por ser menos sensible con muestras pequeñas. En Galicia [Heatmap](#), la variable `p_mes` ha mostrado una correlación fuerte y significativa con `energia_hidraulica` ($\rho = 0.75$, $p\text{-valor} < 0.001$) y `np_010` y `p_max` también [Figura](#), mientras que en Castilla y León las correlaciones han sido mucho más débiles y estadísticamente no significativas [Figuras](#).

En Galicia, se ha probado un modelo de regresión polinómica de segundo grado (Figura 5.3.1) que, aunque no constituye un modelo predictivo formal por falta de entrenamiento en series temporales, permite **visualizar una tendencia no lineal** entre la precipitación mensual (`p_mes`) y la energía hidráulica generada. Este modelo presenta un **R² ajustado aproximado de 0.55**, claramente superior al del modelo lineal, lo que indica un mejor ajuste a los datos históricos observados.

A pesar de sus limitaciones, este intento de modelado permite **identificar fenómenos como la saturación hídrica**: a partir de cierto umbral de precipitaciones, el aumento en generación hidráulica se estabiliza, probablemente debido a la capacidad máxima de los embalses o restricciones operativas.

Este análisis no pretende ser un modelo predictivo válido, ya que:

- **No se ha separado en datos de entrenamiento y test** (por ejemplo, entrenar con 2023 y validar con 2024),
- **No se ha modelado la estacionalidad** de forma explícita, lo que es fundamental en series temporales de datos climatológicos y energéticos,
- **No se ha utilizado un enfoque de series temporales univariante**, tal y como se recomienda en análisis de este tipo.

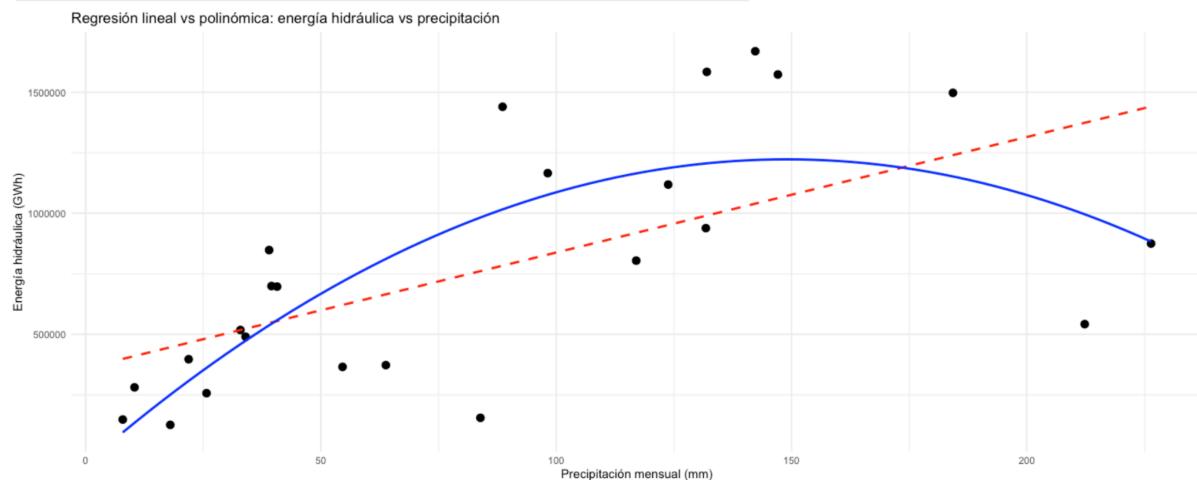


Figura 5.3.1 Regresión lineal línea roja vs Regresión polinómica línea azul

No obstante, el modelo se ha incluido como **aproximación exploratoria razonable** dentro de los recursos y tiempos disponibles, con el objetivo de entender mejor la relación no lineal entre las variables. Como **trabajo futuro**, se propone realizar un análisis riguroso de series temporales, incorporando componentes estacionales, donde se crearía un histórico para test y así se podría estudiar la relación y evaluar el error del modelo correctamente, utilizando una separación temporal para validación cruzada y modelos adaptados al comportamiento periódico del sistema hidráulico.

En Castilla y León, se ha replicado el proceso completo: descarga, limpieza, combinación y análisis técnico. Sin embargo, los resultados son significativamente diferentes. Las correlaciones son débiles y los modelos de regresión no explican adecuadamente la variabilidad en la generación hidráulica. Esto podría deberse a la mayor inercia del sistema hídrico en esa comunidad, a una diferente orografía, o a decisiones de gestión de embalses.

Este análisis ha permitido concluir que en Galicia existe una relación clara y significativa entre las precipitaciones mensuales y la generación de energía hidráulica, confirmada tanto a nivel visual como estadístico. Sin embargo, se ha observado que esta relación no es inmediata, lo que evidencia un comportamiento con *inercia hídrica*: la producción responde al agua acumulada en embalses más que a la lluvia puntual del mes. Este hallazgo, visible en la evolución temporal de ambas variables (gráfico interactivo), añade valor al entender cómo opera realmente el sistema hidroeléctrico en entornos lluviosos. En contraste, en Castilla y León no se han encontrado correlaciones significativas, lo que sugiere que otros factores, como la orografía o la gestión de embalses, influyen más que la lluvia mensual directa. Esta comparación regional refuerza la idea de que el análisis climático-energético debe abordarse de forma localizada, y que no es posible generalizar una única relación entre lluvia y generación hidráulica en todo el país.

4. Lecciones aprendidas para mis futuros proyectos en Ciencia de Datos

Durante el desarrollo de este proyecto, hemos adquirido una serie de aprendizajes valiosos tanto a nivel técnico como metodológico, que sin duda resultarán fundamentales para abordar futuros trabajos en el ámbito de la Ciencia de Datos. Estas lecciones pueden resumirse en los siguientes puntos:

- **Importancia de la planificación temprana y la división de objetivos:** Definir de forma clara un objetivo por persona ha permitido estructurar el trabajo, repartir responsabilidades y avanzar en paralelo. Esta estrategia ha sido clave para optimizar el tiempo y evitar bloqueos innecesarios.
- **Valor del trabajo en equipo real y coordinado:** A diferencia de otros proyectos más individuales, este ha requerido una comunicación constante entre miembros, especialmente en fases clave como la integración de datos o la elaboración de la memoria.
- **Uso Git y GitHub:** La creación y gestión de un repositorio común ha permitido llevar un control de versiones real, compartir scripts y organizar carpetas según hitos y objetivos. Esto ha mejorado la trazabilidad del proyecto y ha facilitado el trabajo colaborativo.
- **Consolidación del uso de R y nuevas librerías:** Aunque teníamos conocimientos previos de R, este proyecto nos ha obligado a dominar nuevas librerías y también hemos reforzado el uso de dplyr y ggplot2, y aprendido a desarrollar flujos de trabajo reproducibles, limpios y bien documentados. También hemos aprendido a elegir los gráficos adecuados para responder a preguntas concretas, haciendo uso de escalas dobles, boxplots, regresiones ajustadas y heatmaps que reflejen patrones de forma clara y profesional.
- **Dificultad técnica en la adquisición y transformación de datos reales:** A diferencia de datasets limpios o académicos, trabajar con datos reales (especialmente meteorológicos desde AEMET y energéticos desde REE) ha supuesto un reto importante. Nos enfrentamos a formatos heterogéneos (JSON, CSV), errores en los datos (valores no válidos como meses inexistentes), codificación inconsistente, y la necesidad de diseñar funciones propias para limpiar y estructurar la información antes de poder analizarla.
- **Aprendizaje sobre la fusión de fuentes y su sentido técnico:** Hemos entendido que fusionar datos no es simplemente unir tablas, sino una decisión que debe tener sentido en el contexto del análisis. En este caso, cruzar variables climáticas con generación energética requería alinear fechas, seleccionar estaciones representativas, y generar nuevas variables (como escalados o estaciones meteorológicas) que permitieran hacer análisis visuales y estadísticos rigurosos.
- **Herramientas utilizadas por el equipo:**
 - **RStudio** como entorno principal de trabajo.
 - **Git y GitHub** para control de versiones y trabajo colaborativo.
 - **Terminal de comandos** para tareas específicas como manejo de archivos, commits, pulls...
 - **Chat interno del equipo (WhatsApp y Slack)** para coordinación rápida y resolución de dudas, así como el reparto de tareas
 - **Docs google y RMarkdown** para redacción de la memoria y exportación de informes en HTML y Word.

En conjunto, este proyecto nos ha permitido trabajar con datos reales, entender su complejidad, y aplicar un flujo de trabajo completo de principio a fin: desde la recolección hasta el análisis y presentación de resultados, lo que representa un gran paso en nuestra formación como futuros analistas de datos.

5. Bibliografía

El desarrollo del proyecto ha estado respaldado por materiales y conocimientos adquiridos en diversas asignaturas del grado, especialmente:

- **Asignatura Proyecto-II:** se han seguido las guías y plantillas proporcionadas para la estructura del proyecto, el uso de R y la integración de fuentes de datos.
- **Asignatura de Programación y Estructuras de Datos:** ha sido fundamental para la manipulación de estructuras como data frames, vectores y el uso de funciones en R.
- **Asignatura de Adquisición y Transmisión de Datos (ATD) :** ha resultado clave para la descarga de datos desde fuentes oficiales como la API de AEMET, así como para la interpretación de formatos JSON y la transformación posterior.
- **Soporte docente:** la profesora ha desempeñado un papel clave en la resolución de dudas técnicas y metodológicas, en la selección adecuada de técnicas analíticas y en la validación del enfoque seguido.
- **Repositorio GitHub:** se ha creado un repositorio público en GitHub con toda la estructura del proyecto (hitos, objetivos, código y anexos) para facilitar el trabajo colaborativo y garantizar la trazabilidad de los cambios durante el desarrollo.

GITHUB LINK: <https://github.com/VICENTE102/PROYECTO-II.git>

Además, se ha hecho uso de documentación oficial de **Red Eléctrica de España (REE)** y **AEMET** para comprender el significado y estructura de las variables tratadas en los análisis, así como de la documentación de librerías R como ggplot2, dplyr, corrplot, plotly, jsonlite, readr...

6. Anexos

6.1 Hito 1: Fuentes de datos

- https://github.com/VICENTE102/PROYECTO-II/tree/cb132a8cdc45f73df30313d9a8a341fc5e17139d/hito_1

6.2 Hito 2: Integración y preparación de los datos

- https://github.com/VICENTE102/PROYECTO-II/tree/cb132a8cdc45f73df30313d9a8a341fc5e17139d/hito_2

6.3 Objetivos Específicos

- htmls compilados públicos donde se explican y redactan la mayor parte de las fases de los objetivos específicos:

<https://github.com/VICENTE102/PROYECTO-II/tree/cb132a8cdc45f73df30313d9a8a341fc5e17139d/docs>

6.3.1 Objetivo 1

- código:

https://github.com/VICENTE102/PROYECTO-II/tree/76b728f37a77eec78c0b4d98d03af319411d90e7/objetivos/objetivo_1_miguel

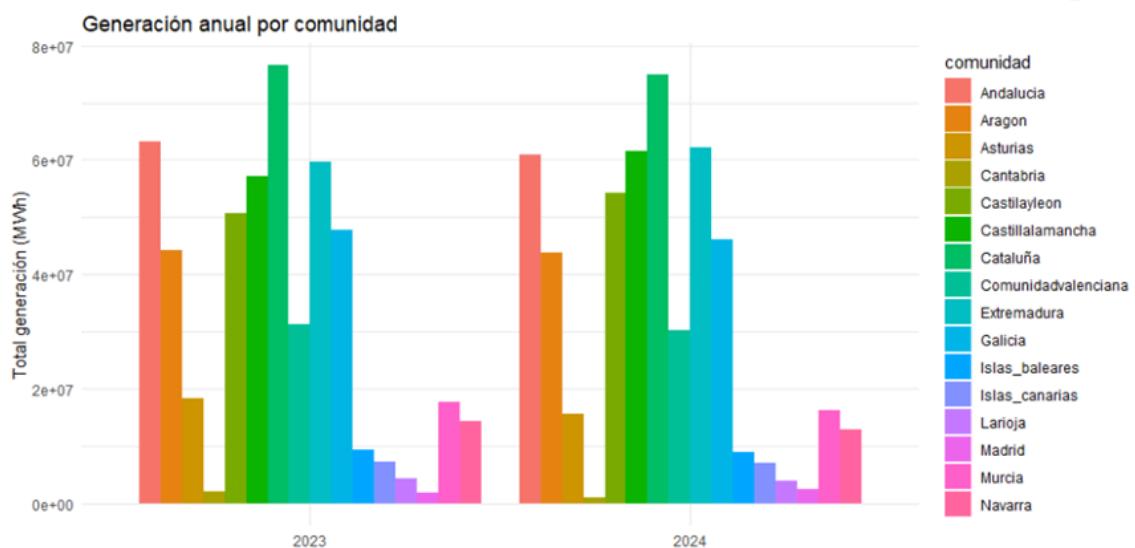
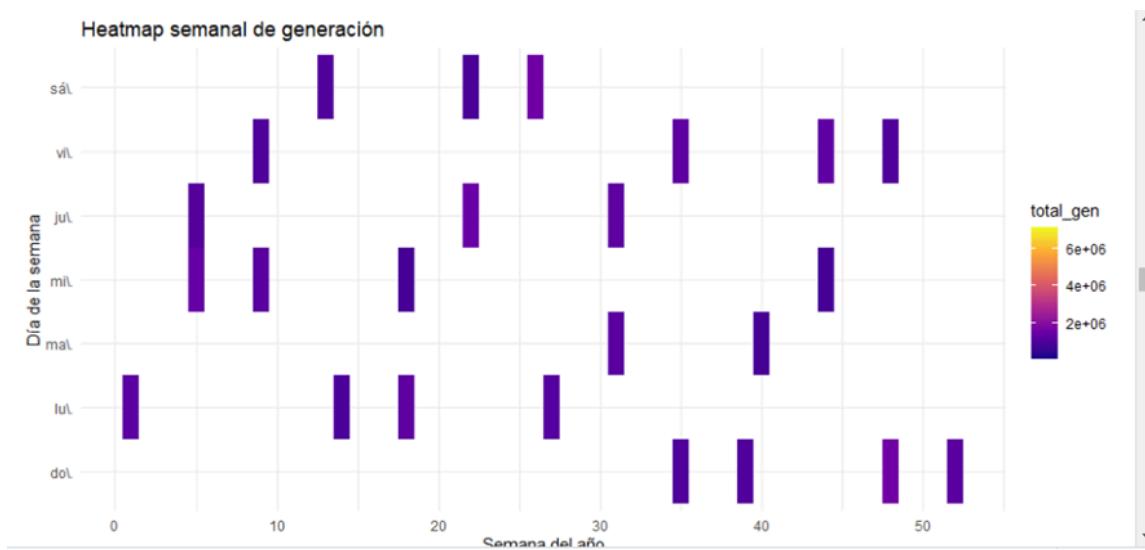
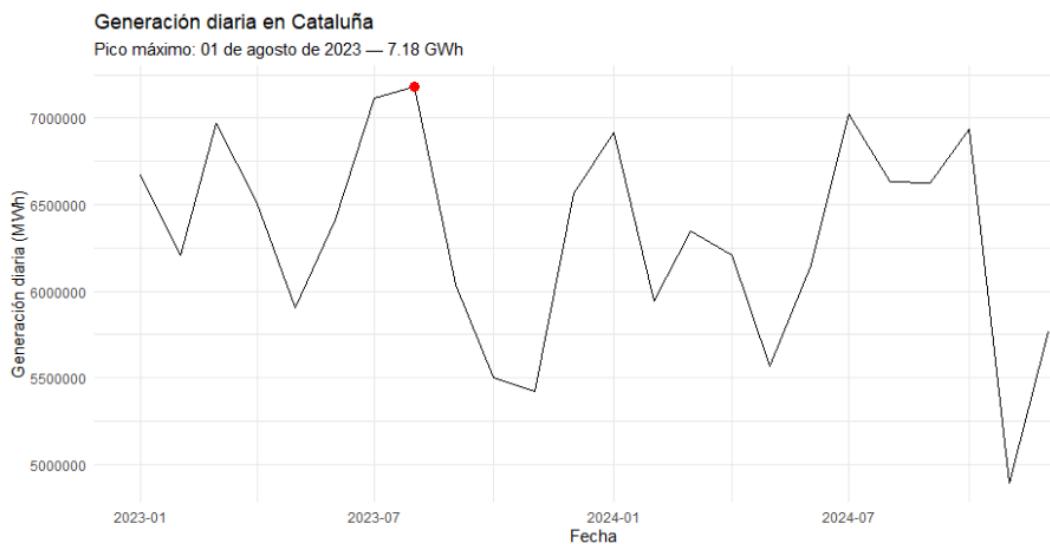
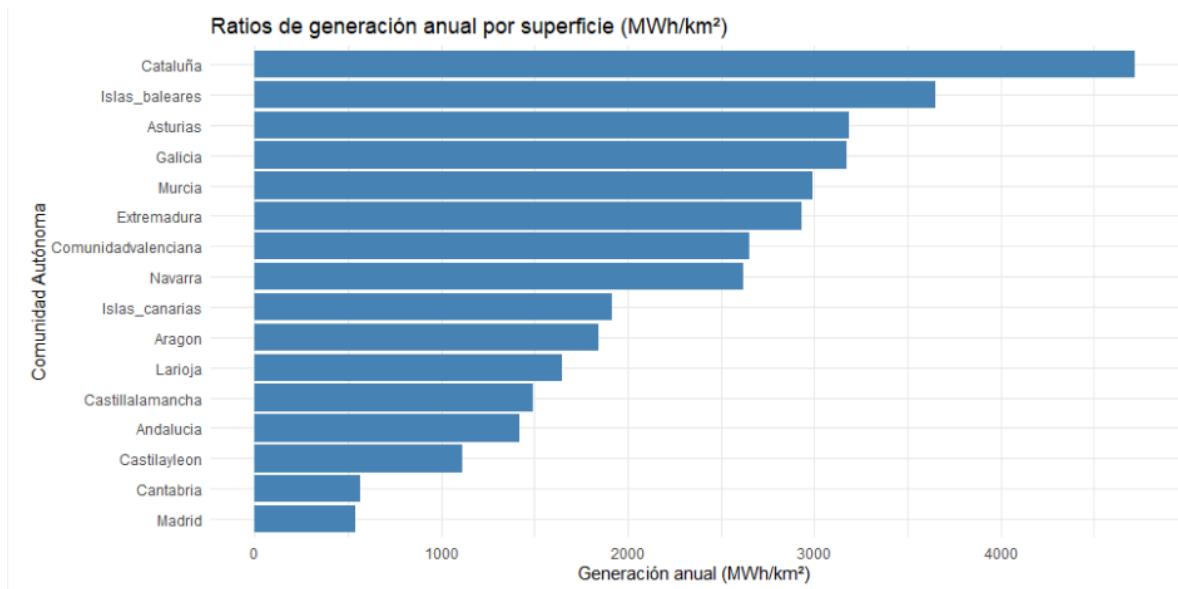


Figura 6.3.1.1

**Figura 6.3.1.2****Figura 6.3.1.3**

**Figura 6.3.1.4****Mapa de calor de eficiencia energética (2024)**

Generación anual por superficie en cada CCAA

**Figura 6.3.1.5**

6.3.2 Objetivo 2

- código:

https://github.com/VICENTE102/PROYECTO-II/tree/05c6565cd1aff28dbfc4f8bca2b2193e091e22d7/objetivos/objetivo_2_victor

Figura 6.3.2.1

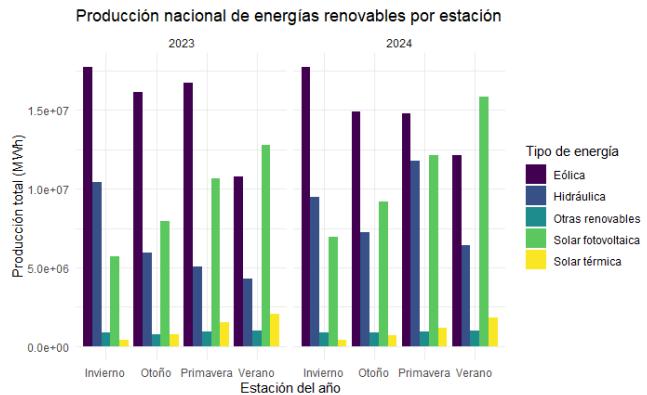
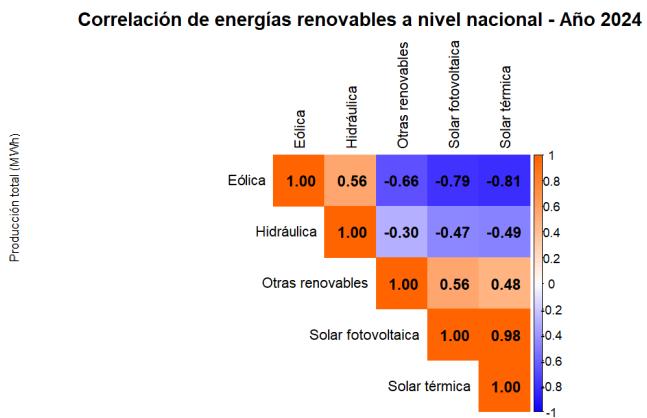
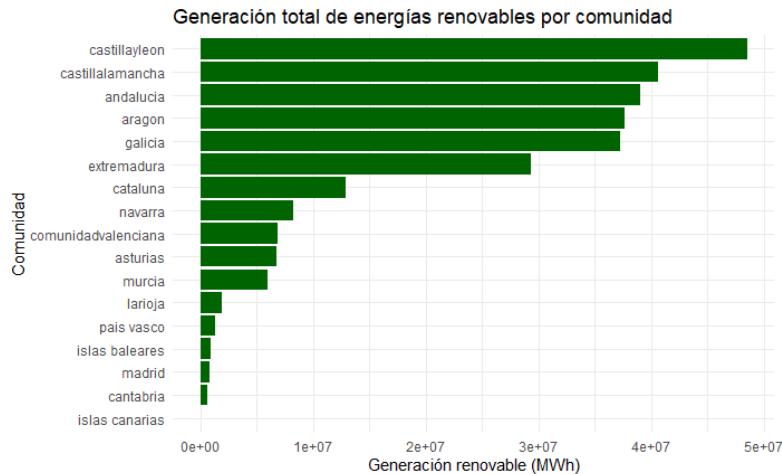


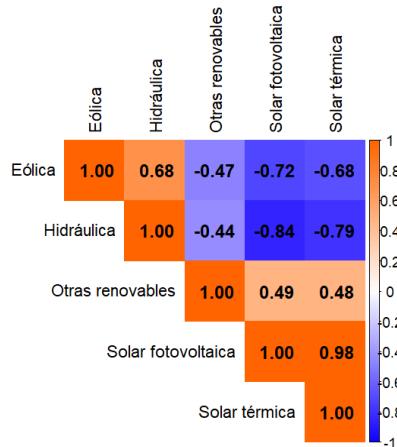
Figura 6.3.2.3

Figura 6.3.2.2

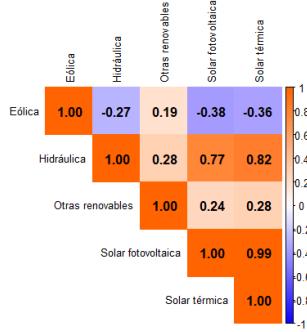


**Figura 6.3.2.4: generación total de renovables por comunidad**

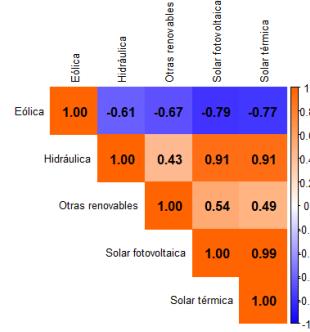
Correlación de energías renovables a nivel nacional - Año 2023

**Figura 6.3.2.5: correlación de renovables a nivel nacional 2023.**

Correlación en Andalucía - 2023



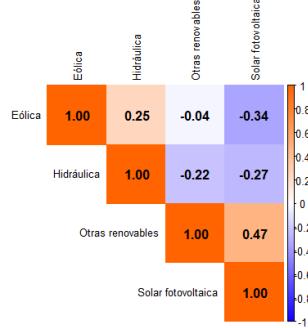
Correlación en Andalucía - 2024

**Figura 6.3.2.6**

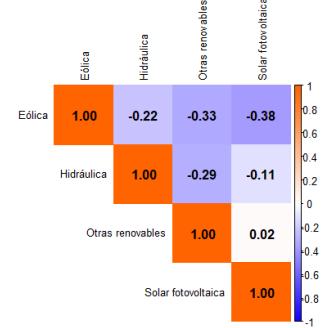
Correlación renovables Andalucía 2023

Correlación renovables Andalucía 2024

Correlación en Aragón - 2023



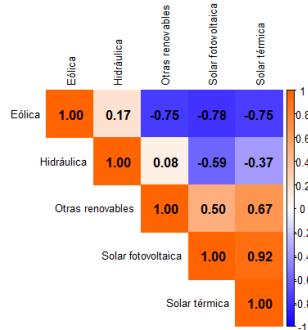
Correlación en Aragón - 2024

**Figura 6.3.2.7**

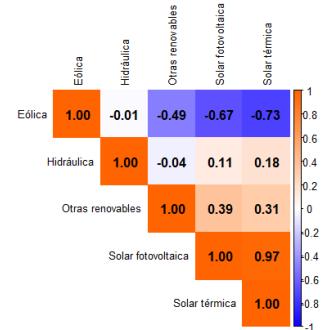
Correlación renovables Aragón 2023

Correlación renovables Aragón 2024

Correlación en Castillalamancha - 2023



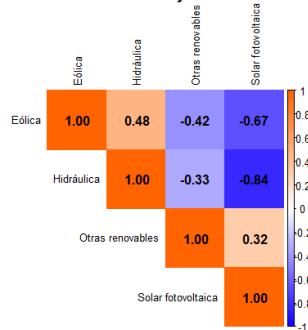
Correlación en Castillalamancha - 2024

**Figura 6.3.2.8**

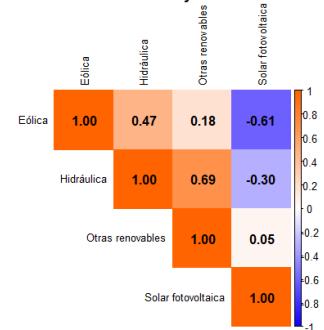
Correlación renovables Castilla-La Mancha 2023

Correlación renovables Castilla-La Mancha 2024

Correlación en Castillayleon - 2023



Correlación en Castillayleon - 2024

**Figura 6.3.2.9**

Correlación renovables Castilla y León 2023

Correlación renovables Castilla y León 2024

Correlación en Galicia - 2023

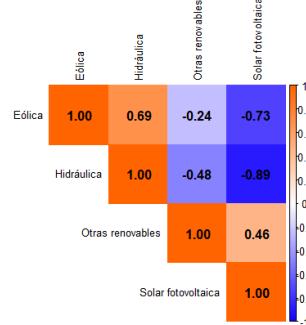
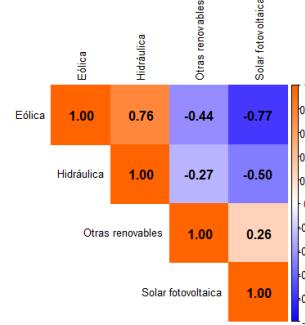


Figura 6.3.2.10

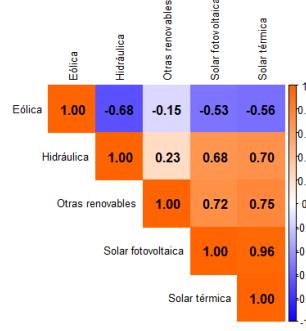
Correlación renovables Galicia 2023

Correlación en Galicia - 2024



Correlación renovables Galicia 2024

Correlación en Comunidad valenciana - 2023



Correlación en Comunidad valenciana - 2024

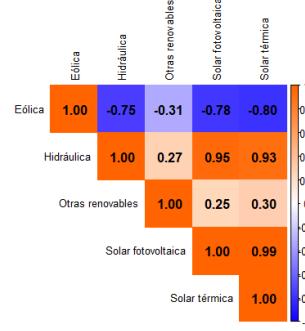


Figura 6.3.2.11

Correlación renovables Comunidad Valenciana 2023

Correlación renovables Comunidad Valenciana 2024

Generación de Solar fotovoltaica en 2023 - Todo el año



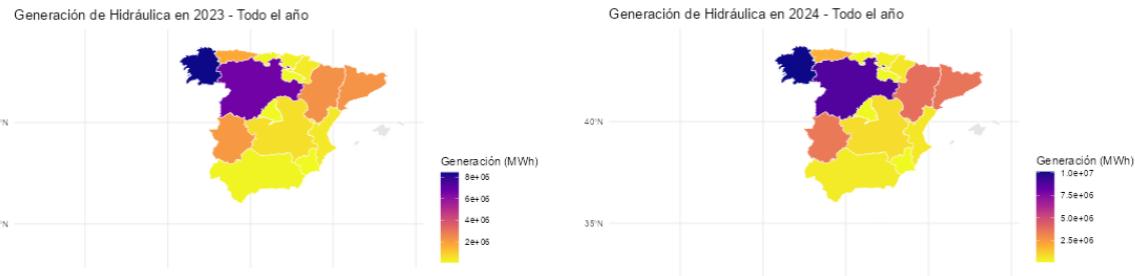
Generación de Solar fotovoltaica en 2024 - Todo el año



Figura 6.3.2.12

Mapa generación solar fotovoltaica 2023

Mapa generación solar fotovoltaica 2024

***Figura 6.3.2.13***

Mapa generación hidráulica 2023

Mapa generación hidráulica 2024

6.3.3 Objetivo 3

- **código:**

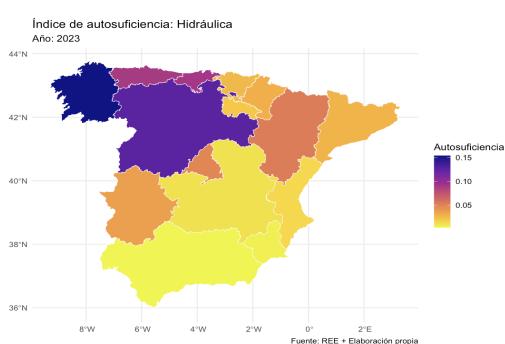
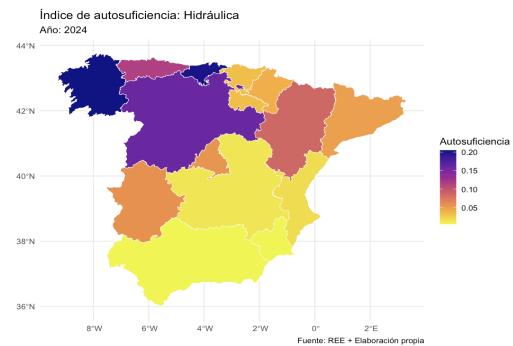
https://github.com/VICENTE102/PROYECTO-II/tree/cb132a8cdc45f73df30313d9a8a341fc5e17139d/objetivos/objetivo_3_marc

En primer lugar, fue necesario unificar y estandarizar los nombres de las comunidades autónomas en ambos conjuntos de datos para garantizar una correcta unión de la información. Esto incluyó la eliminación de espacios y tildes, la conversión a minúsculas y la unificación de denominaciones como "Castilla y León" o "País Vasco". Una vez normalizados los datos, se procedió a calcular el índice de autosuficiencia energética para cada comunidad autónoma y cada mes. Este índice se define como el cociente entre la energía generada y la energía demandada en un periodo determinado (Índice de Autosuficiencia=Generación/Demanda).

En la figura 6.3.3.1 se puede ver como Extremadura lidera con claridad, con un índice superior a 12, lo que indica que genera mucha más energía de la que consume. Le siguen Castilla-La Mancha, Aragón, Castilla y León y Galicia, con valores elevados que reflejan una buena cobertura renovable. Regiones como Navarra, La Rioja, Asturias, Murcia y Cataluña presentan autosuficiencia parcial, con ratios entre 1,7 y 2,9. En el extremo opuesto, Madrid, Cantabria, País Vasco, Comunidad Valenciana, Baleares y Andalucía tienen una generación renovable claramente insuficiente. Madrid destaca negativamente con un índice inferior a 0,1. Estos resultados reflejan una gran disparidad territorial y ponen de manifiesto que, salvo excepciones, la mayoría de comunidades no cubren su demanda con renovables, lo que refuerza la necesidad de inversión, planificación energética y equilibrio territorial en la transición ecológica.

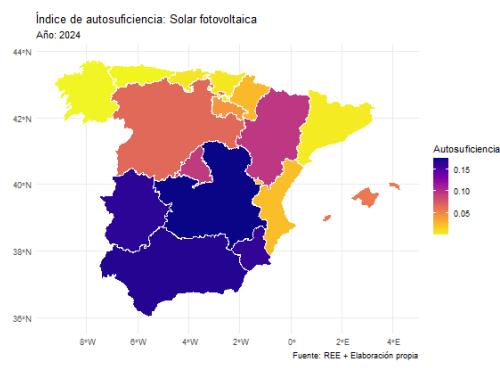
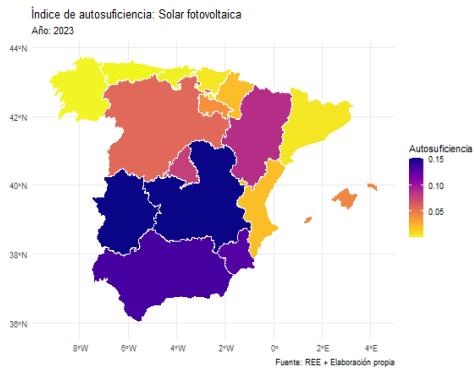
Análisis de autosuficiencia por tipo de energía: Los mapas obtenidos para el año 2023 (figura 6.3.3.2) y 2024 (figura 6.3.3.3) considerando únicamente la energía hidráulica como fuente de generación. En 2023 el mapa muestra que las comunidades del norte peninsular, especialmente Galicia, Cantabria y Castilla y León, presentan los índices de autosuficiencia hidráulica más altos del país. Esto es coherente con su mayor disponibilidad hídrica, recursos fluviales y presencia histórica de centrales hidroeléctricas. En contraste, Andalucía, Región de Murcia y Comunidad Valenciana muestran valores muy bajos, indicando una fuerte dependencia de otras fuentes para abastecer su demanda eléctrica. En 2024 se

observa una tendencia similar al año anterior, aunque algunas regiones centrales como Castilla-La Mancha incrementa ligeramente su proporción hidráulica. En general, la autosuficiencia hidráulica sigue siendo limitada en la mayoría del país, reflejando el carácter estacional e irregular de esta fuente. Los otros tipos de energías se pueden visualizar en el anexo.

Figura 6.3.3.2**Figura 6.3.3.3**

El análisis por tipo de energía revela que ninguna comunidad alcanza la autosuficiencia exclusivamente a partir de generación hidráulica. Incluso en los mejores casos, el índice apenas supera el 15%-20% de la demanda. Esto refuerza la necesidad de diversificar las fuentes renovables para avanzar hacia un modelo energético más autosuficiente. Asimismo, destaca la desigualdad territorial en el acceso a recursos hidroeléctricos, lo que implica que las estrategias de transición energética deberán adaptarse a las características geográficas de cada región.

Figura 6.3.3.4**Figura 6.3.3.5**



Seguidamente se presentan los resultados para la energía solar fotovoltaica (Figura 6.3.3.4 y Figura 6.3.3.5). En 2023 destacan especialmente Andalucía, Extremadura, Castilla-La Mancha y Murcia, con los mayores niveles de autosuficiencia. Estas regiones alcanzan valores cercanos al 15% gracias a su elevada irradiación solar y a la expansión de instalaciones fotovoltaicas a gran escala. En contraste, las comunidades del norte y el área cantábrica muestran valores muy bajos. En 2024 se mantiene la misma distribución geográfica, pero con una leve mejora generalizada en las regiones líderes. Comunidad Valenciana y Madrid también reflejan aumentos modestos. La solar fotovoltaica se consolida como la fuente renovable dominante en el sur peninsular.

La energía solar fotovoltaica muestra una distribución geográfica muy marcada, con altas tasas de autosuficiencia en las regiones del sur y baja penetración en el norte. Esto es coherente con la disponibilidad de recurso solar y demuestra que el potencial de expansión sigue siendo alto. Aunque los niveles de autosuficiencia aún no alcanzan valores absolutos, su tendencia es claramente creciente. Esto hace de la fotovoltaica una pieza clave en la estrategia energética para regiones con alta demanda y elevado recurso solar.

Figura 6.3.3.6

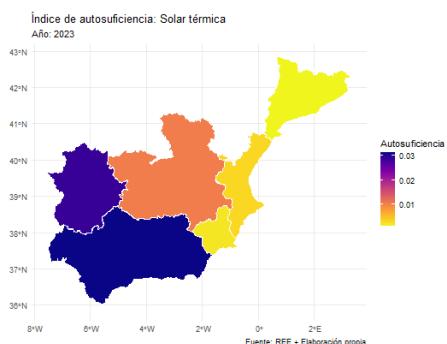
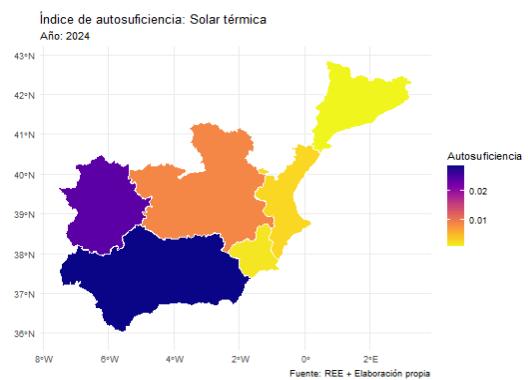


Figura 6.3.3.7



Los mapas de autosuficiencia para esta fuente revelan que en 2023(Figura 6.3.3.6) Únicamente Región de Murcia, Comunidad Valenciana y, en menor medida, Cataluña, registran autosuficiencia solar térmica por encima del 2%-3%. El resto del territorio muestra

CURSO 2024-2025

valores prácticamente nulos. En 2024(Figura 6.3.3.7) a situación es muy similar, con ligeras mejoras en algunas regiones del este y sureste, pero en general los valores se mantienen bajos. La autosuficiencia solar térmica no supera el 3% en ninguna comunidad.

La solar térmica sigue teniendo un papel marginal en la generación eléctrica, ya que su uso está más orientado a la producción de calor. Esta fuente, aunque técnicamente renovable, no contribuye significativamente a cubrir la demanda eléctrica regional, lo que explica los bajos índices de autosuficiencia. Su potencial podría aumentar si se integran tecnologías de concentración solar o se fomenta su hibridación con almacenamiento térmico.

Figura 6.3.3.8

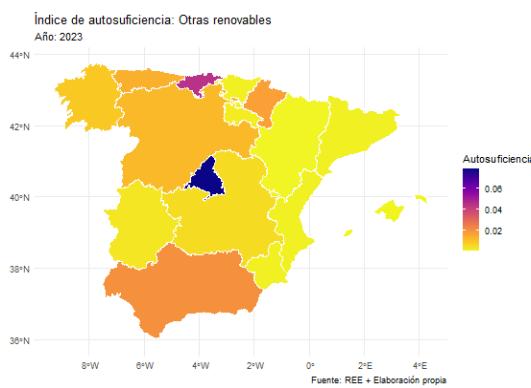
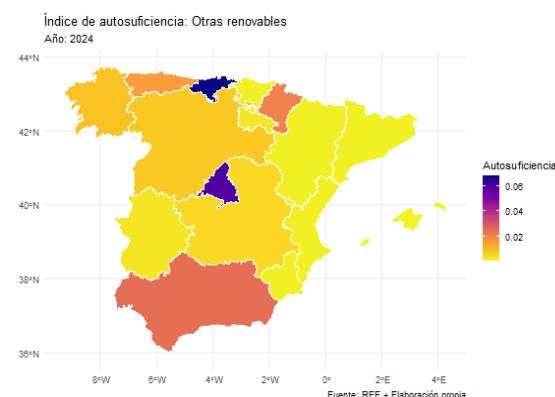


Figura 6.3.3.9



Los resultados muestran lo siguiente. En 2023 (Figura 6.3.3.8) únicamente Galicia presenta un valor de autosuficiencia notable (alrededor del 6%-7%), mientras que el resto de comunidades se sitúan por debajo del 2%. Se observa una concentración del uso de estas tecnologías en regiones con fuerte actividad forestal o agroindustrial. En 2024 (Figura 6.3.3.9) se mantienen las diferencias regionales, con Galicia y, en menor medida, Castilla y León, como principales exponentes del uso de otras renovables. El resto de regiones no experimenta variaciones significativas.

Las denominadas "otras renovables" presentan gran potencial a nivel local, pero su implantación a escala regional sigue siendo baja. Están muy vinculadas a sectores concretos como la industria agroforestal, lo que restringe su desarrollo a determinadas zonas. Aunque su aportación al mix es limitada, podrían desempeñar un papel relevante en entornos rurales y economías circulares, especialmente si se integran en modelos de autoconsumo o microrredes.

Figura 6.3.3.10

Figura 6.3.3.11

CURSO 2024-2025



A continuación se analizan los mapas de autosuficiencia por energía eólica para los años 2023(Figura 6.3.3.10) y 2024(Figura 6.3.3.11). En 2023 las comunidades de Castilla y León, Galicia, Navarra y Aragón presentan los valores más altos de autosuficiencia eólica, con índices que se aproximan o superan el 25%. Esto se debe a sus condiciones geográficas favorables para el aprovechamiento del viento, así como al despliegue histórico de parques eólicos en estas zonas. En cambio, comunidades como Andalucía, Madrid o Región de Murcia muestran bajos niveles, reflejando su menor implantación de esta fuente o una mayor demanda no cubierta por eólica. En 2024 se mantiene el liderazgo de las regiones del norte y centro-norte. Se observan pequeñas mejoras en Castilla-La Mancha y Cataluña, mientras que el resto del país presenta estabilidad. Algunas regiones como País Vasco y Galicia se consolidan como referentes en autosuficiencia eólica.

La energía eólica es, junto con la hidráulica, una de las fuentes renovables más consolidadas en España. Los datos confirman que hay comunidades que logran cubrir más del 20%-25% de su demanda eléctrica con esta fuente, especialmente en zonas con alta densidad de parques eólicos. Sin embargo, la autosuficiencia completa no se alcanza en ningún caso, lo que refuerza la necesidad de diversificación renovable. Además, se pone de manifiesto una asimetría territorial significativa, donde regiones del sur presentan un amplio margen de desarrollo en esta tecnología.

6.3.4 Objetivo 4

- código:

https://github.com/VICENTE102/PROYECTO-II/tree/cb132a8cdc45f73df30313d9a8a341fc5e17139d/objetivos/objetivo_4_vicente

OBTENCIÓN DE LOS DATOS:

Para recolectar los datos de la AEMET este código. Como este objetivo se centra especialmente en la energía solar fotovoltaica he filtrado las variables de interés, con una función, que son todas las relacionadas con las temperaturas ya que en el siguiente objetivo explicamos las otras energías.

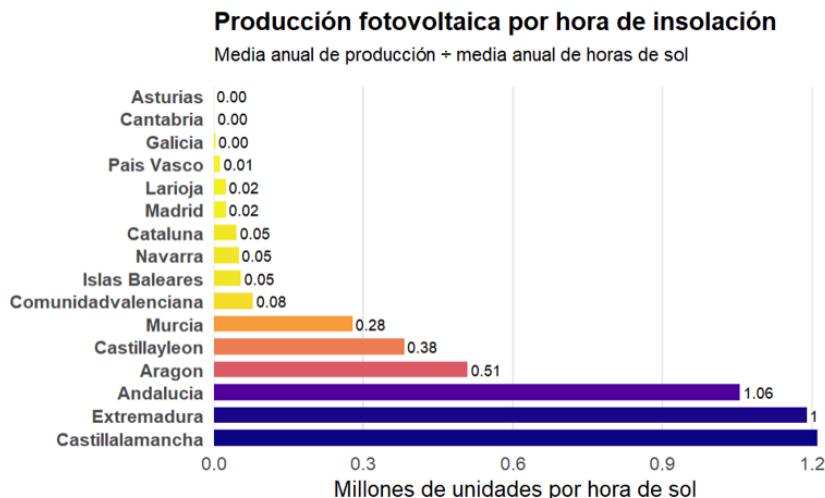
Los datos proporcionados por AEMET se encontraban desglosados por provincias, mientras que nuestro análisis se realiza a nivel de comunidad autónoma. Para resolver este

problema, en aquellos casos en los que las provincias de una misma comunidad presentaban temperaturas medias mensuales prácticamente idénticas, seleccionamos una provincia representativa para utilizar sus datos en nombre del conjunto. No obstante, cuando se observaban diferencias significativas entre las temperaturas de las provincias de una misma comunidad, procedimos a calcular la media de las variables de interés, utilizando únicamente los datos de las provincias cuyas temperaturas se desviaban de manera relevante.

Los CSV que están en las variables son las comunidades consideradas con temperaturas muy distintas y hacemos la media de las variables con la fórmula justo que sigue a continuación. Después se redondean los resultados, ordenar la tabla por fecha y se obtienen los .CSV limpios.

Primera Pregunta:

Figura 6.3.4.1



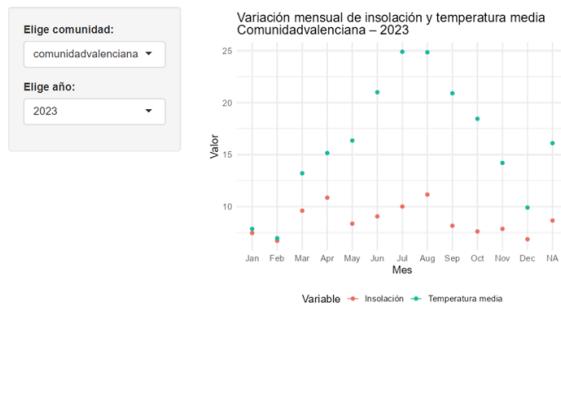
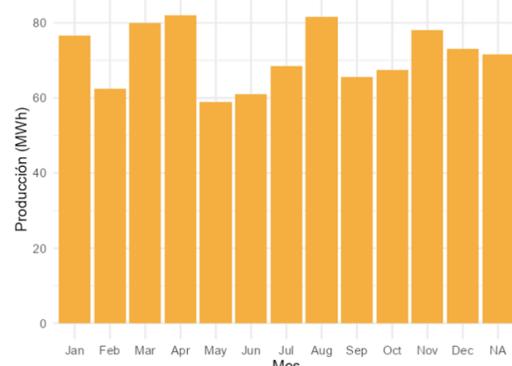
Segunda Pregunta:

Figura 6.3.4.2

Figura 6.3.4.3

CURSO 2024-2025

Dashboard Energía Solar por Comunidad | 2023–2024

Producción mensual de energía solar fotovoltaica
Comunidad valenciana – 2023

Este es otro ejemplo, parecido a la figura 6.3.4.1 pero esta vez de la Comunidad Valenciana, no de Andalucía. Se puede observar que las conclusiones son las mismas que las obtenidas anteriormente, en el propio objetivo.

Tercera Pregunta:**Figura 6.3.4.5: Visualización de datos**

▲	fecha	inso	tm_min	ta_max	p_sol	ta_min	ti_max	tm_mes	tm_max	n_cub	n_des	n_nub	diasnublados
1	2023-01-01	6.90	6.80	19.40	69.5	2.05	10.30	10.80	14.85	2.5	12.0	16.5	19.0
2	2023-02-01	6.55	8.00	22.80	60.5	2.40	11.75	12.10	16.05	7.5	10.5	10.0	17.5
3	2023-03-01	9.20	9.90	26.75	77.0	2.85	13.25	15.25	20.60	3.5	10.0	17.5	21.0
4	2023-04-01	10.75	13.70	31.55	82.5	9.50	20.05	19.40	24.95	2.0	14.0	14.0	16.0
5	2023-05-01	9.20	15.20	30.90	65.0	11.95	17.25	20.05	25.00	6.0	9.5	15.5	21.5
6	2023-06-01	10.35	20.00	38.45	70.0	15.25	23.85	25.05	30.05	5.5	12.5	12.0	17.5
7	2023-07-01	11.95	23.70	41.65	82.5	20.20	30.55	29.20	34.65	0.5	21.5	9.0	9.5
8	2023-08-01	11.90	23.55	40.55	87.5	18.10	29.45	29.25	34.90	0.0	22.5	8.5	8.5
9	2023-09-01	8.95	19.55	34.25	72.0	16.10	22.50	24.50	29.45	4.0	8.5	17.5	21.5
10	2023-10-01	7.60	16.65	34.10	67.0	10.00	17.15	21.15	25.55	3.5	8.0	19.5	23.0
11	2023-11-01	7.45	10.60	25.15	73.0	6.05	15.75	14.85	19.05	0.5	11.5	18.0	18.5
12	2023-12-01	5.75	7.55	20.75	59.5	3.60	11.80	11.65	15.70	5.5	13.5	12.0	17.5

¿Cómo es posible que haya números en decimal si estamos contando días?

Los datos de la AEMET, daban datos solo de provincias, por lo que hemos tenido que ir provincia por provincia viendo las temperaturas que tenían y las que difieren mucho (porque hay comunidades que no tienen la misma temperatura en sus provincias), pues de esas provincias que difieren de temperatura se realiza la media de todas las variables seleccionadas, de ahí los números decimales.

CURSO 2024-2025

En principio este objetivo se iba a realizar con la variables ‘diasnublados’. Esta variables era la suma de los días cubiertos y nublados, pues se deseaba ver la relación existente entre la producción de energía con los días nublados pero los gráficos y resultados no era lógico ya que por ejemplo:

Figura 6.3.4.6

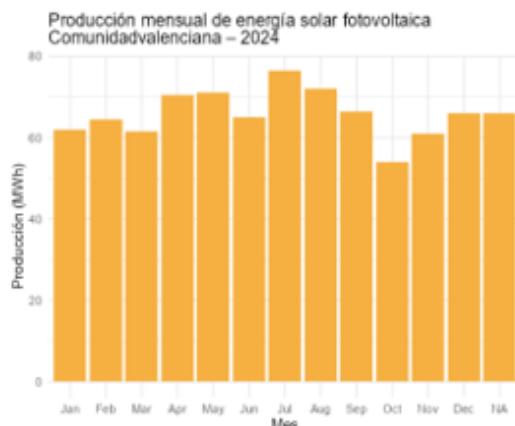


Figura 6.3.4.7

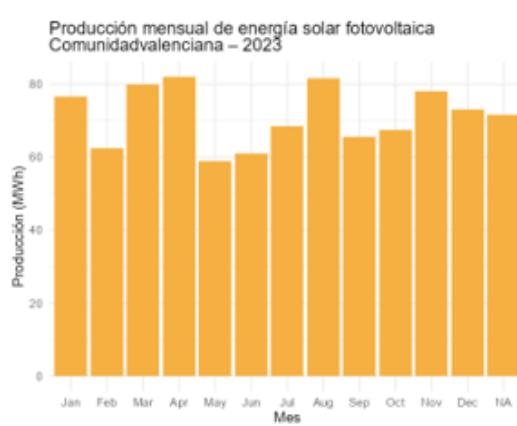


Figura 6.3.4.8



Figura 6.3.4.9

(Pearson 2023): 0.112

Este coeficiente de Pearson muestra que la relación entre estas variables es positiva lo que quiere decir que un mes con mayor número de días nublados se obtenía una mayor producción de energía solar lo que es contradictorio. Al final, se vió que la variable ‘diasnublados’ no era la indicada para responder a esta pregunta y era ‘n_cub’, ver [Objetivo 5.3.4](#).

Curiosidad:

CURSO 2024-2025

¿Qué comunidad de las tres es la que tiene una producción más equilibrada?

Se ha hecho uso de un coeficiente de variación que permite comparar comunidades que producen cantidades muy diferentes, por lo que cuanto más bajo sea el CV, más regular (estable) es la producción de esa comunidad

Figura 6.3.4.10

comunidad	anio	media_mensual	sd_mensual	coef_variacion
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
andalucia	2023	708448.49675	209095.47288	0.2951456
asturias	2023	51.59858	21.56218	0.4178832
comunidadvalenciana	2023	53941.05983	12454.56790	0.2308922
andalucia	2024	860854.59408	273093.71069	0.3172356
asturias	2024	48.31492	17.66979	0.3657212
comunidadvalenciana	2024	55557.91700	16870.87579	0.3036629

6 rows

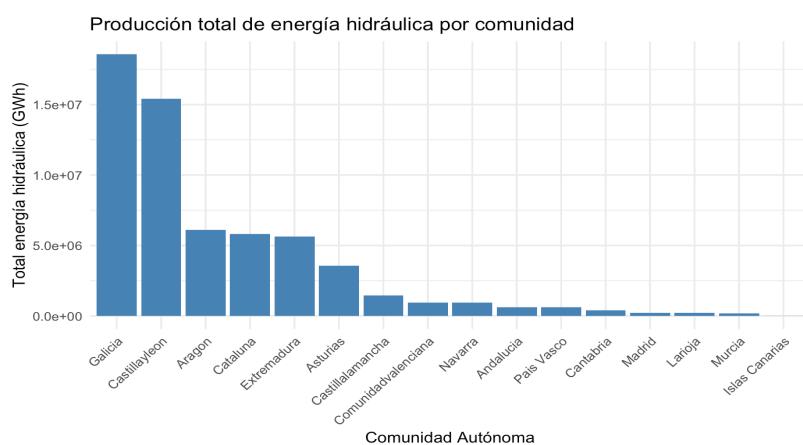
En 2023, la comunidad más estable fue la Comunitat Valenciana ($CV \approx 0.23$), seguida de Andalucía (0.29) y finalmente Asturias, que presenta la mayor variabilidad (0.41). En 2024, aunque los valores aumentan ligeramente, se mantiene la misma tendencia: la Comunitat Valenciana sigue siendo la más estable ($CV \approx 0.30$), seguida de Andalucía (0.31) y Asturias (0.36). Esto sugiere que, a pesar de las diferencias climáticas y de radiación entre regiones, la Comunitat Valenciana mantiene una producción mensual de energía solar más homogénea, probablemente debido a una mayor regularidad en la radiación solar a lo largo del año.

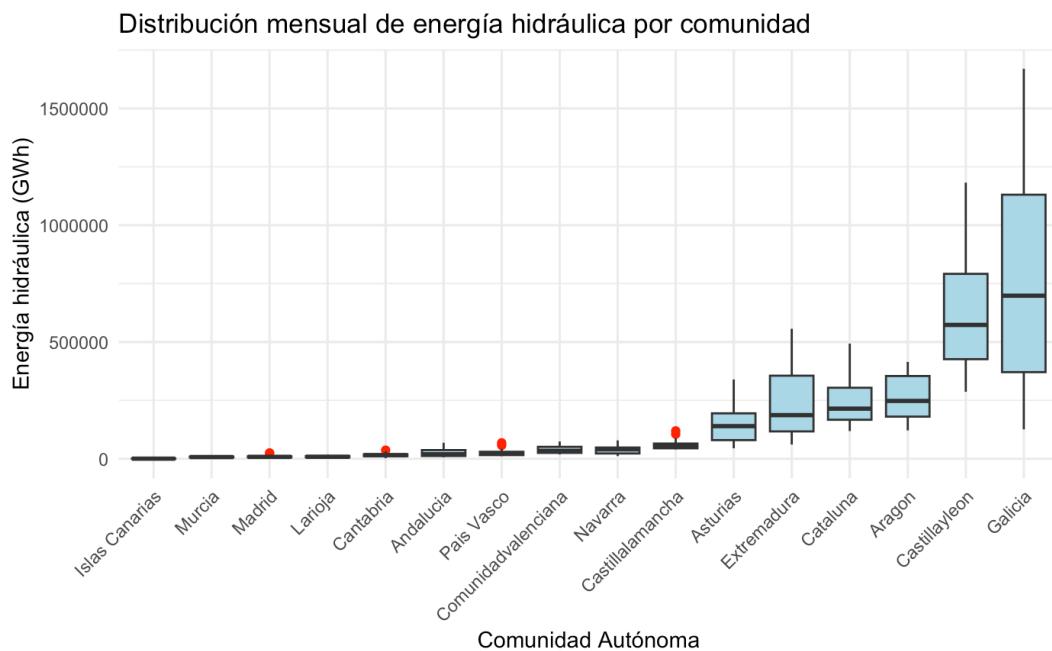
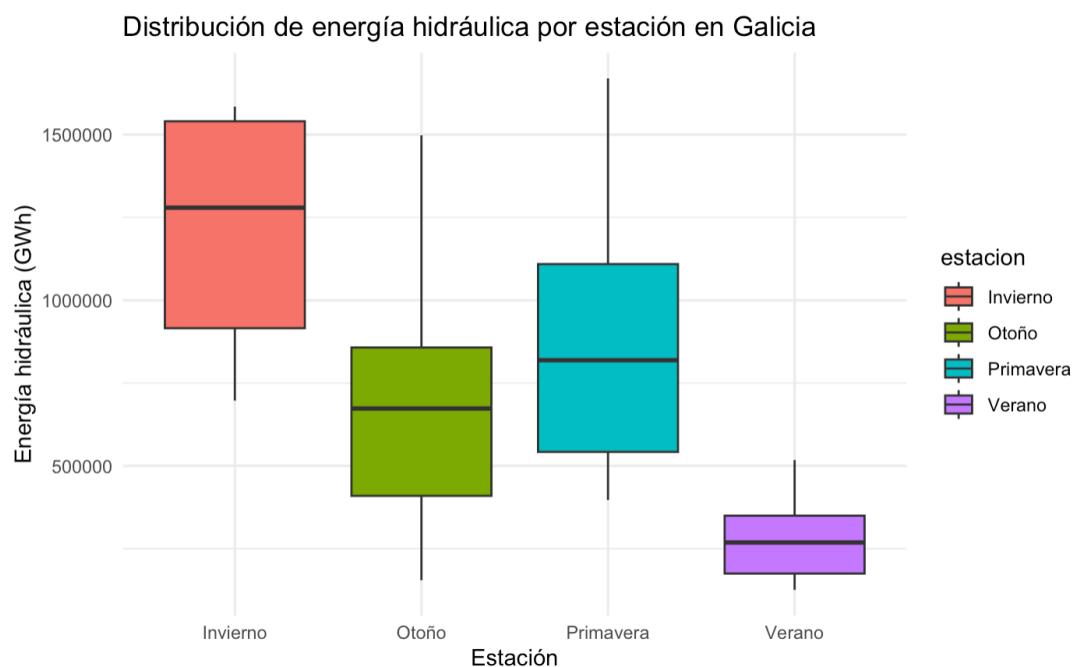
6.3.5 Objetivo 5

- **código:**

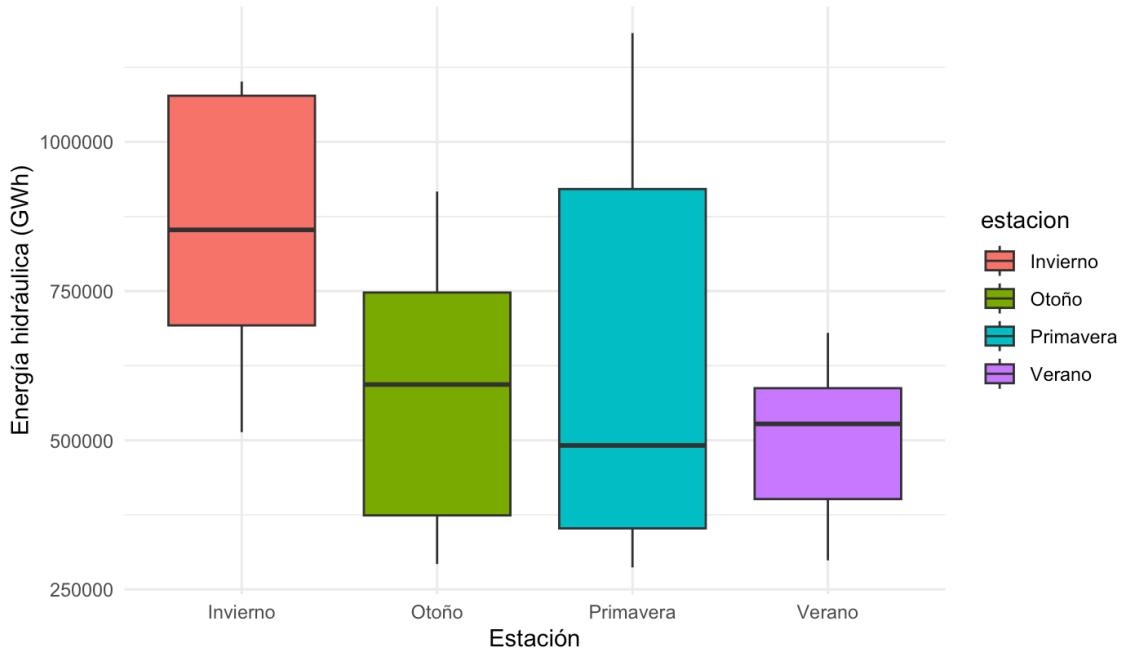
https://github.com/VICENTE102/PROYECTO-II/tree/cb132a8cdc45f73df30313d9a8a341fc5e17139d/objetivos/objetivo_5_pablo

Figuras 6.3.5.1 y 6.3.5.2: Estudio Energía Hidráulica por comunidad

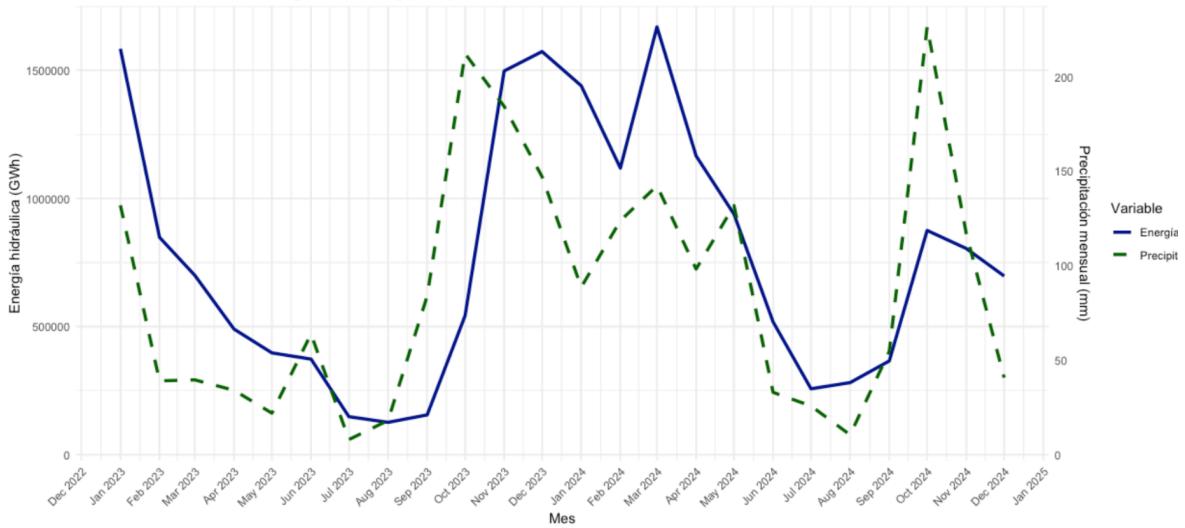


**Figura 6.3.5.3****Figura 6.3.5.4**

Distribución de energía hidráulica por estación en Galicia

**Figura 6.3.5.5**

Relación mensual entre energía hidráulica y precipitación en Galicia

**Figura 6.3.5.6**

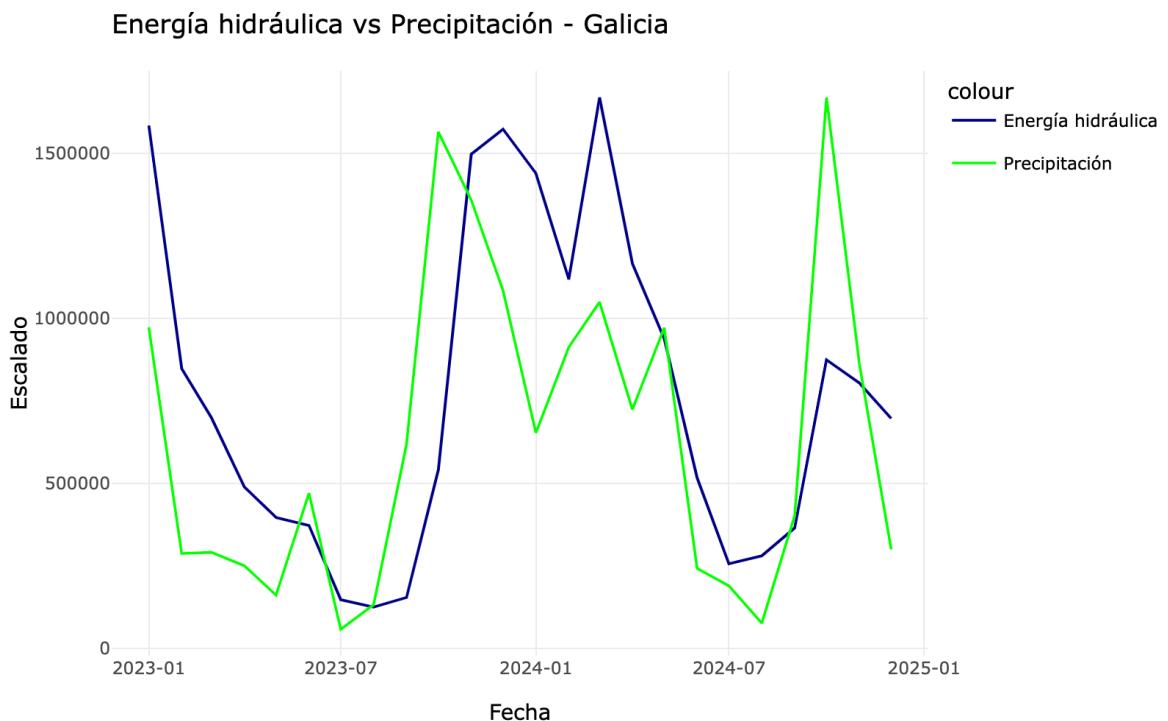
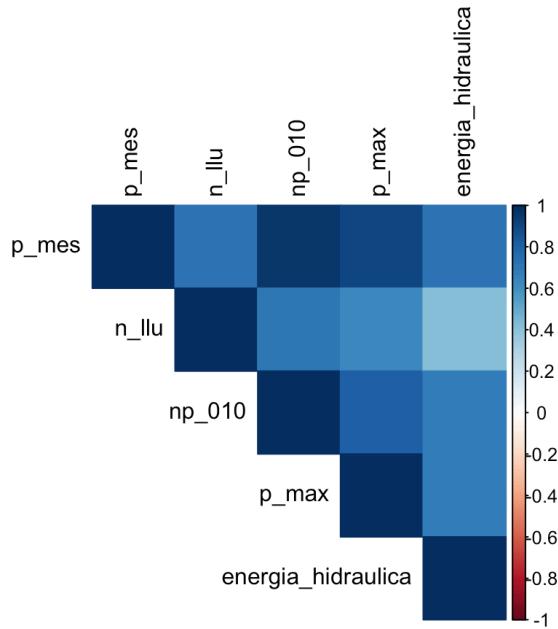


Figura 6.3.5.7 Heatmap relación variables Galicia

**Figura 6.3.5.8: Correlaciones**

```

Spearman's rank correlation rho

data: df_merged$p_mes and df_merged$energia_hidraulica
S = 570, p-value = 3.579e-05
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.7521739

Warning: Cannot compute exact p-value with ties
Spearman's rank correlation rho

data: df_merged$n_llu and df_merged$energia_hidraulica
S = 1170, p-value = 0.04492
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.4219321

Warning: Cannot compute exact p-value with ties
Spearman's rank correlation rho

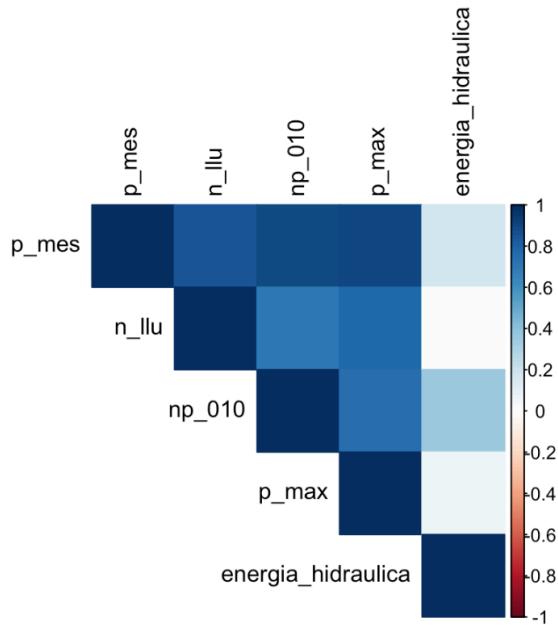
data: df_merged$np_010 and df_merged$energia_hidraulica
S = 752.95, p-value = 0.000317
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.6726317

Spearman's rank correlation rho

data: df_merged$p_max and df_merged$energia_hidraulica
S = 672, p-value = 0.0001639
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.7078261

```

Figura 6.3.5.9, 6.3.5.10 y 6.3.5.11: Heatmap y correlaciones Castilla la Mancha



Spearman's rank correlation rho

```

data: df_merged$p_mes and df_merged$energia_hidraulica
S = 1900, p-value = 0.4146
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.173913
  
```

Warning: Cannot compute exact p-value with ties
 Spearman's rank correlation rho

```

data: df_merged$n_llu and df_merged$energia_hidraulica
S = 2264.3, p-value = 0.9426
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.01551714
  
```

Spearman's rank correlation rho

```
data: df_merged$np_010 and df_merged$energia_hidraulica
S = 1458.1, p-value = 0.07857
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.3660226
```

Spearman's rank correlation rho

```
data: df_merged$p_max and df_merged$energia_hidraulica
S = 2120, p-value = 0.7156
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.07826087
```