|  |
| --- |
| IMF BUSINESS SCHOOL |
| PREDICCIÓN DEL PRECIO DEL POOL ELÉCTRICO ESPAÑOL |
| TFM BIG DATA & BUSINESS ANALYTICS |

|  |
| --- |
| Carlos Alonso Salcedo  1-6-2022 |

ÍNDICE:

1. Resumen
2. Introducción y antecedentes

2.1. Introducción y agentes del sector eléctrico

2.2. Historia del sector eléctrico en España

2.3. ¿Cómo se establece el precio de la electricidad?

2.4. Potencia instalada en España.

1. Hipótesis de trabajo y objetivos
2. Material y métodos
3. Resultados
4. Discusión
5. Conclusiones
6. Referencias
7. Anexos
8. **RESUMEN**

El objetivo del presente trabajo es el de la predicción del precio del pool eléctrico español durante todo el año de 2021. Para ello, se hará uso de diferentes bases de datos y páginas web con información financiera, de producción de energía renovables y datos meteorológicos, entre otros. Para predecir esta variable objetivo, se hará uso de diferentes algoritmos y métodos, desde los más sencillos, basado en un modelo de persistencia, hasta modelos complejos basados en arquitecturas de redes neuronales. El resultado final será de un *benchmark* de los modelos usados para ver cuáles se comportan mejor en este periodo de tiempo caracterizado por el periodo con más volatilidad de la historia del precio de la electricidad en España. Para ello, todos los modelos se compararán mediante un *backtesting out of sample* para todo el año 2021 y se compararán con las mismas métricas para poder realizar una comparación directa y así poder determinar cuál es el mejor modelo para resolver nuestro problema.

1. **INTRODUCCIÓN Y ANTECEDENTES**
   1. **INTRODUCCIÓN Y AGENTES DEL SECTOR ELÉCTRICO**

La electricidad se trata de un pilar en el que se basa el mundo actual. Prácticamente todo lo que usamos día a día como ordenadores, electrodomésticos, teléfonos móviles, iluminación, industria y un largo etcétera, dependen de la electricidad. De hecho, la energía es un bien básico de primera necesidad cuyo acceso debe ser garantizado como servicio público (1).

Esto se traduce en que la cantidad de energía que se consume es altísima pues está presente en prácticamente todos los sectores y en el día a día de cada persona. Por ejemplo, la cantidad de energía eléctrica que se consumió en 2021 en España fue de 256.387 GWh, comparado con el consumo en 1980 que fue de 97.231 GWh, se trata de un consumo de más del doble, debido al aumento de la población y al uso cotidiano de cada vez más dispositivos electrónicos (2).

Sin embargo, pese a que cada vez hay más población y se usan más dispositivos electrónicos, en Europa se redujo el consumo de electricidad más de un 10 % entre 2005 y 2015 debido a las mejoras en la eficiencia energética, el aumento de la proporción de energías procedentes de la fuentes hidráulica, eólica y solar fotovoltaica, los cambios estructurales en la economía y la recesión económica de 2008. También ha contribuido el hecho de que los inviernos hayan sido más cálidos, lo que ha permitido reducir la cantidad de energía destinada a calefacción (3).

El sector eléctrico es altamente complejo, pero al mismo tiempo fundamental para poder mantener el estilo de vida contemporáneo. En el *pool* del mercado es donde se compra la energía que llega a nuestros hogares y donde se vende la electricidad producida en las centrales. (3)

Un **sistema eléctrico** es el conjunto de elementos que operan de forma coordinada en un determinado territorio para satisfacer la demanda de energía eléctrica. Este sistema, en España, tiene 7 componentes:

1. Centros o plantas de generación (donde se produce y eleva la tensión de la electricidad para transportarla)
2. Líneas de transporte de la energía de alta tensión, gestionada, desarrollada y mantenida por Red Eléctrica de España (REE).
3. Las estaciones transformadoras, que reducen la tensión.
4. Líneas de distribución de media y baja tensión que llevan la electricidad hasta los consumidores.
5. Las instalaciones de los clientes o consumidores de energía eléctrica
6. Los centros de control de las empresas generadoras, distribuidoras y comercializadoras.
7. Un centro de control eléctrico nacional desde el que se gestiona, coordina y opera el sistema eléctrico, y que está gestionado también por REE.

Sin embargo, el negocio del sector eléctrico no sería tal sin la participación de los siguientes actores:

* **Generadores**: producen la energía eléctrica, independientemente del tipo de tecnología utilizada para ello.
* **Transportista**: transmiten la energía de la red de transporte a los puntos de consumo donde se entrega a los distribuidores. En España, por ley, hay un transportista único y es **REE**.
* **Distribuidores**: son las compañías que llevan la electricidad hasta los clientes finales. Aunque hay numerosas empresas que ejercen esta actividad, en cada área de España sólo puede haber un distribuidor.
* **Operador del sistema**: es la empresa que se encarga de que todo el proceso de la operación del sistema eléctrico funcione correctamente. La clave es cumplir la siguiente ecuación y regla de oro en el negocio de la electricidad.
* **Comercializadoras**: venta de energía eléctrica a los consumidores según la potencia contratada, comprándola a su vez a los generadores en el llamado ‘*pool eléctrico* ’, una especie de subasta que determina el precio de la energía eléctrica en la Península Ibérica.
  1. **HISTORIA DEL SECTOR ELÉCTRICO EN ESPAÑA**

Durante años, el sector eléctrico en España funcionó como un oligopolio en el que el precio de la electricidad dependía de pocas compañías eléctricas en España. El país estaba dividido en diversas áreas geográficas, cuyo suministro de electricidad se adjudicó en exclusiva a cinco grandes empresas. Cada una de estas eléctricas gestionaba las 4 fases del suministro de energía de su ámbito de actuación, dejando a los consumidores sin posibilidad de escoger su propia compañía eléctrica.

En 1997, con el objetivo de fomentar la competencia en el sector de la energía eléctrica y mejorar el conocimiento que los usuarios tenían del mismo, se aprobó la primera ley de liberalización del sector eléctrico nacional. La normativa prohibía que una misma compañía opere en más de una de las fases del proceso de suministro y transfiera la gestión del transporte a distintas redes eléctricas en España.

Como consumidores, no podemos escoger qué empresa distribuye la energía eléctrica que consumimos, pero sí podemos elegir a quién se la compramos.

Como se ha visto, entonces la electricidad forma parte de la vida y de la actividad económica de las personas. Realizando un cálculo rápido y sencillo, el precio medio de la electricidad en España en 2021 fue de 111,93 €/MWh. Esto indica que, únicamente en el negocio de compra y venta de la electricidad en el mercado del pool, se trató de un negocio que movió cerca de 29 mil millones de euros en España en 2021.

Esto lo convierte en un negocio que mueve mucho dinero y que, debido a lo mucho que se consume, su precio es muy importante ya que afecta a todas las personas y todos los negocios.

* 1. **¿Cómo se establece el precio de la electricidad?**

Una vez hemos visto lo importante que es la electricidad, qué agentes participan en el negocio y la cantidad de energía y dinero que mueve este negocio cabe preguntarse, ¿cómo se establece el precio de la electricidad?

La forma de establecer el precio sigue el **algoritmo Euphemia** que surgió en la iniciativa “Price Coupling of Regions” (PCR) por parte de siete mercados eléctricos europeos, entre los que se encuentra el español. Este algoritmo calcula los precios de la energía eléctrica de forma eficiente persiguiendo la maximización del bienestar, que se define como el excedente o beneficio tanto de los compradores como de los vendedores, al tiempo que optimiza el uso de capacidad disponible en las interconexiones.

En resumen, las empresas encargadas de la generación hacen sus ofertas (cantidad de energía y precio) y las empresas encargadas de la venta al por menor, consumidores directos, etc., demandan la energía necesaria a un precio determinado. Una vez realizadas las ofertas, se ordenan según el precio, en orden creciente en el caso de la venta y decreciente en el caso de la compra. La intersección de las curvas de oferta y demanda se denomina **punto de casación**.

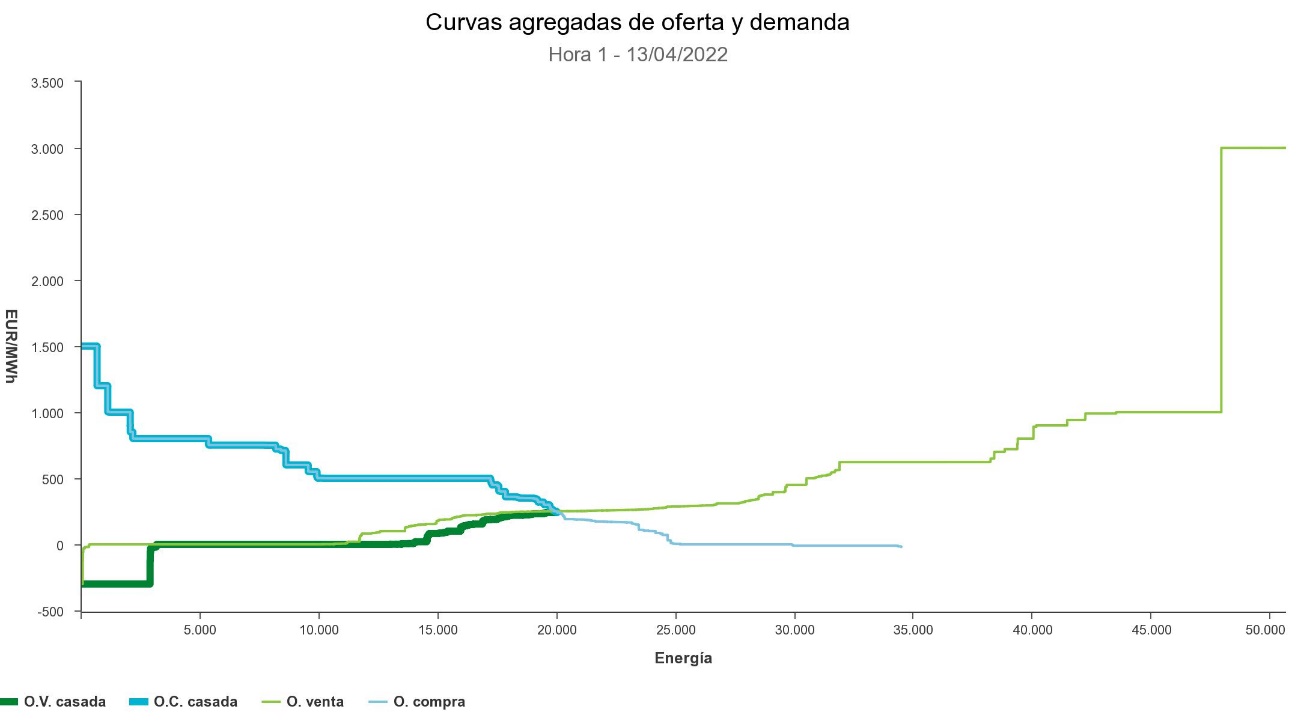


Imagen 1. Curva de oferta y demanda para el 13 de abril de 2022 en la hora 1. Imagen proveniente de OMIE. <https://www.omie.es/es/market-results/daily/daily-market/aggragate-suply-curves>

Por tanto, este punto de corte es el punto que optimiza el bienestar y, por tanto, establece el precio de la energía para esa hora concreta. Toda la energía ofrecida y demandada a un precio inferior al punto de casación se intercambiará a ese precio, mientras que la que tenga un precio superior no lo hará. Este proceso se repite para cada una de las 24 horas de un día, estableciendo el precio de la electricidad para cada una de las horas de cada día.

La energía que se envían a las casas suele ser una mezcla de varias tecnologías (eólica, fotovoltaica, hidráulica, ciclos combinados de gas, carbón…). La generación de electricidad por cada energía tiene un precio asociado, sin embargo, a la hora de venderla a mercado, todas se venden al mismo precio que es el que marca el punto de casación, que es en lo que se basa un **sistema marginalista**.

Pongamos un ejemplo completamente ficticio: a una hora determinada hay una demanda de 20.000 MW, y tenemos ofertando la energía nuclear con 7.000 MW a precio cero; eólica, 12.000 MW a precio también cero. Luego vienen los ciclos combinados (gas), 10.000 MW a precio 40 €/MWh. Y por último carbón, 8.000 MW a precio 60 €/MWh. Al casar la oferta saldrían los 7.000 MW de la nuclear, los 12.000 MW de la eólica y solo 1.000 MW de ciclo combinado, el carbón se queda fuera. **Pero a todos se les paga el precio de casación, 40 €/MW**.

Es por ello que al sistema se le denomina marginalista, porque a todos se les paga el precio marginal de casar la oferta y la demanda, independientemente de lo ofertado.

* 1. **Potencia instalada en España.**

Como se ha adelantado, la energía que llega a nuestras casas y negocios es una mezcla de la electricidad producida por diversas tecnologías. Actualmente en España se dispone de una potencia instalada de 114.290 MW, distribuida en las siguientes tecnologías. (6)

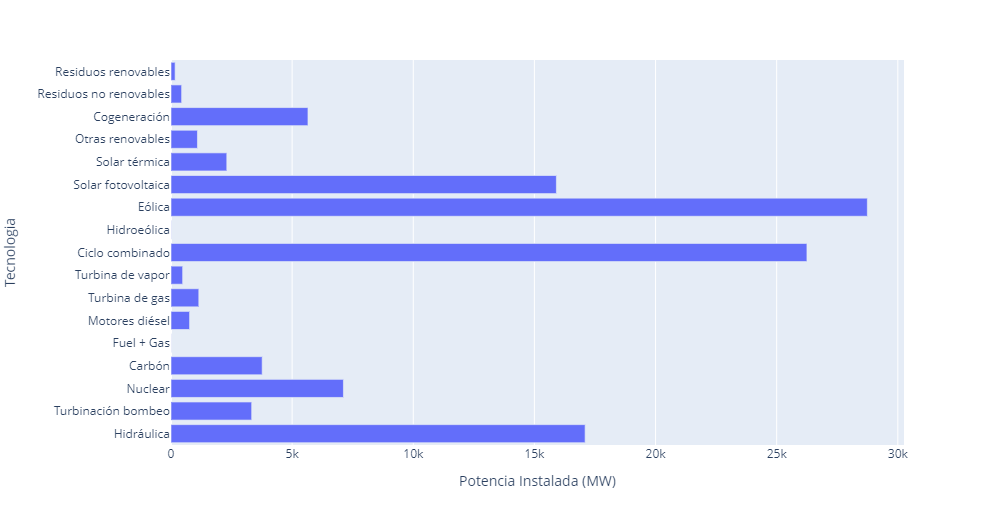


Imagen 2. Distribución por tecnología de la potencia instalada en España (6). Elaboración propia

De ella deducimos que las 4 tecnologías con mayor capacidad instalada en España son, en primer lugar, la eólica (28.743 MW), seguida del ciclo combinado (26.250 MW), la hidráulica (17.094 MW) y la solar fotovoltaica (15.098 MW)

Sin embargo, estos números hay que tratarlos con cuidado. Estos valores de generación hacen referencia a la potencia total que podría producir el sistema si todas las distintas tecnologías produjesen simultáneamente el máximo del que están preparadas. Poniendo un ejemplo más concreto, la energía eólica tiene 28.743 MW de potencia instalada lo que quiere decir, que si todos los aerogeneradores situados en España produjesen el tope de electricidad del que están preparados, producirían esta cantidad de energía, pero esto es bastante complicado que ocurriese, ya que en todas las plantas de España debería estar soplando un viento suficientemente fuerte como para producir la máxima energía de la que está preparada una central en el mismo momento. Lo mismo pasaría con las demás tecnologías.

Tampoco indica que en esta proporción se produce electricidad, por ejemplo, de solar hay 15.098 MW de potencia instalada, pero es evidente pensar que el peso de esta energía respecto a la total en verano e invierno será muy distinta, ya que en verano la potencia generada estará más cerca de la potencia instalada fotovoltaica que en invierno y el % de la generación ocupará un mayor peso en los meses con mayor radiación como los de verano.

En el siguiente gráfico (7) se muestra, para cada día del periodo comprendido entre el 01/01/2022 y 01/05/2022, el % de energía producida por cada tecnología con respecto al total de la energía producida. A simple vista se observa que las barras verdes (correspondiente a la eólica) es muy volátil, pues depende del viento, que producirá cuanto más haya, en caso de que no exista, se observa como crece la barra color naranja claro correspondiente al ciclo combinado, que trata a hacer frente a corregir cuando la eólica no produce en niveles altos.

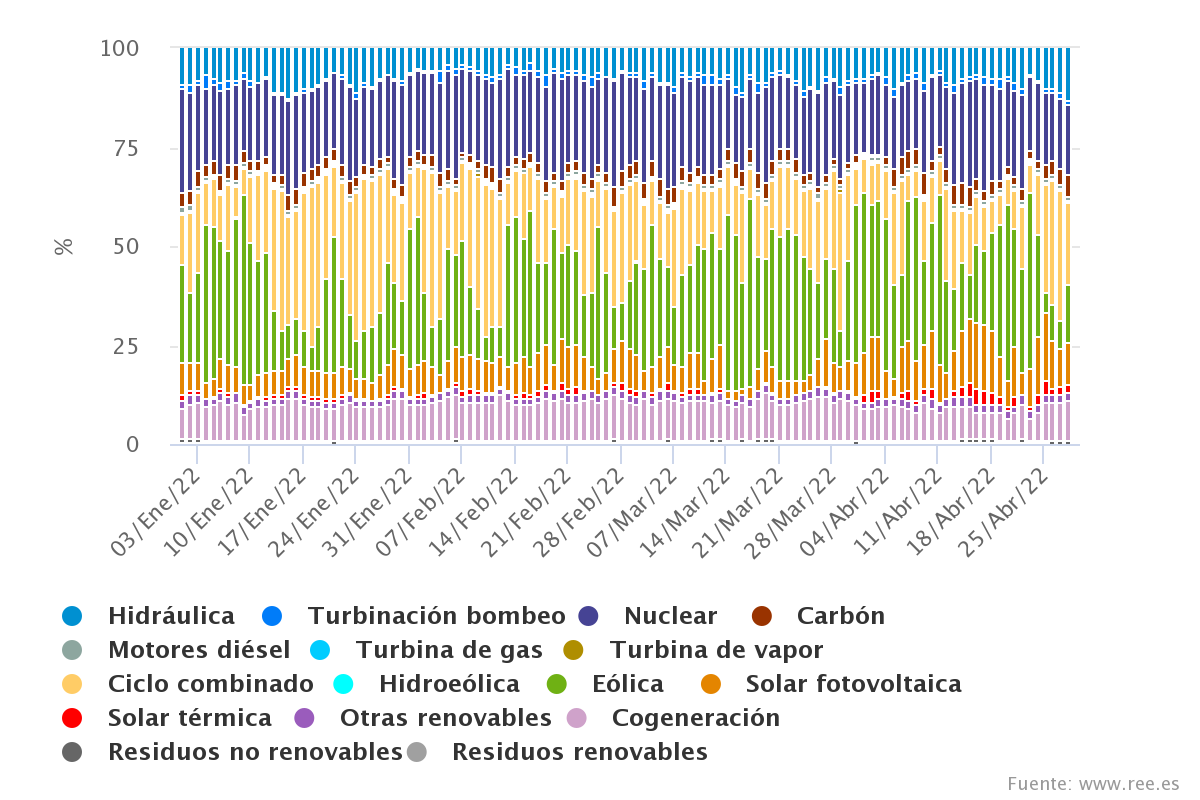


Imagen 3. Porcentaje de energía producida por cada tecnología con respecto al total (7).

Por poner números concretos, comparamos un miércoles de enero con un miércoles de abril y vemos la producción de las tecnologías más importantes.

|  |  |  |
| --- | --- | --- |
| Tecnología | 05/01/2022  (GWh) | 27/04/2022  (GWh) |
| Hidráulica | 57 | 79 |
| Fotovoltaica | 35 | 68 |
| Eólica | 295 | 48 |
| Ciclo Combinado | 92 | 227 |
| Total | 767 | 695 |

Tabla 1. Comparativa entre las tecnologías producidas el 05/01/2022 y el 27/04/2022.

Como se observa, la fotovoltaica, como era de esperar, produce el doble en abril que en enero. Sin embargo, el 27/04/2022 parece que fue un día en el que no hizo mucho viento, ya que la producción fue muy pequeña comparada con la que se produjo el 05/01/2022. Para contrarrestar esta disminución, se observa un aumento muy claro de producción en el ciclo combinado de la fecha de abril con respecto la de enero.

Todo esto pone de manifiesto lo complicado que es de gestionar este ejercicio de demanda, pues la producción de fuentes renovables como la eólica o fotovoltaica depende de las condiciones climáticas las cuales no son constantes y, si estas no favorecen, se han de utilizar otras fuentes (más caras) para satisfacer la demanda, que impacta directamente en el precio de la energía.

* 1. **¿Cómo se consume en España?**

En un día cualquiera, el inicio de la jornada laboral, el cierre de los comercios durante el mediodía o la mayor ocupación de los hogares en las horas finales del día, explican por qué la demanda no es idéntica en las distintas horas del día. (8)

Nuestra sociedad demanda más energía en algunos momentos del día (**horas punta**). En invierno, estas horas punta se dan entre las 11:00 y 12:00 o bien entre las 19:00 y 20:00, debido a la confluencia entre actividad comercial y ocupación de los hogares. Sin embargo, en verano las horas punta se producen en las horas centrales del día, coincidiendo con los momentos de mayor temperatura. (8).

Durante las horas nocturnas se produce la demanda mínima diaria (**horas valle**). A estas horas, únicamente la demanda industrial mantiene un consumo importante, ya que las grandes fábricas consumen las 24 horas del día, aprovechando las horas nocturnas, cuando la energía se puede contratar más barata (8).

Un perfil de consumo de un día normal (ya que los festivos pueden no seguir este perfil) es el siguiente. (8)

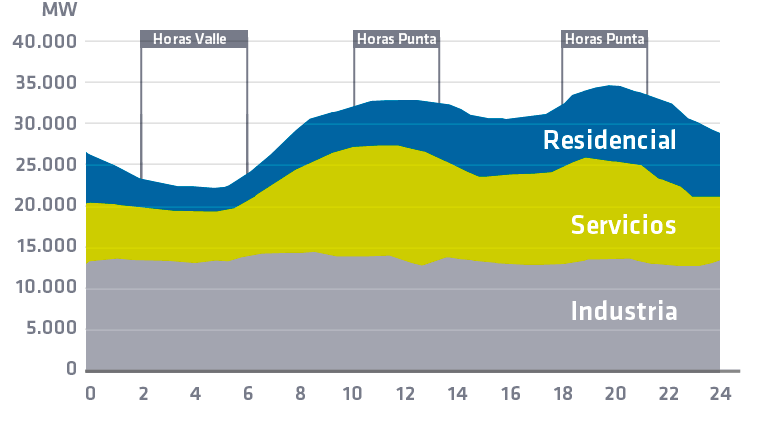


Imagen 4. Perfil de consumo en un día para el sistema eléctrico español (8).

El máximo de consumo en una hora de un día puede oscilar entre los 35.000 y 40.000 MW, mientras que la potencia instalada con la que cuenta España es cercana a los 114.290 MW, es decir, si todo funcionase al máximo de eficiencia en todas las tecnologías, podríamos producir 3 veces más energía de la que normalmente se demanda en un día.

Sin embargo, esto nunca se alcanza, pues las energías renovables dependen de las condiciones meteorológicas y no se ha dado el caso de que se produzca el máximo de todas las renovables en España.

1. **HIPÓTESIS DE TRABAJO Y OBJETIVOS**

En el presente trabajo se va a predecir el precio de la electricidad en España. Para obtener esta variable objetivo, se ha hecho uso del histórico guardado de REE (Red Eléctrica España) en la página de ESIOS [[1]](#footnote-1). Cada día, alrededor de las 14:00, se publica el valor del precio de la electricidad para cada una de las 24 horas del día siguiente.

El objetivo es el de la predicción del precio de cada una de estas 24 horas antes de que sea publicado, haciendo uso de una serie de variables explicativas que serán explicadas más adelante.

En el siguiente gráfico se muestra la evolución del precio de la electricidad en España entre 2014 y 2021. En ella se observa que hasta enero de 2021, el precio no había superado los 140 €/MWh (en esta fecha se alcanzó por el temporal Filomena) y, a partir de ahí y que recuperase su precio normal, se incrementó hasta los niveles de finales de año rondando los 300 – 400 € /MWh, un cambio en la tendencia histórica muy drástico y sin ningún precedente.

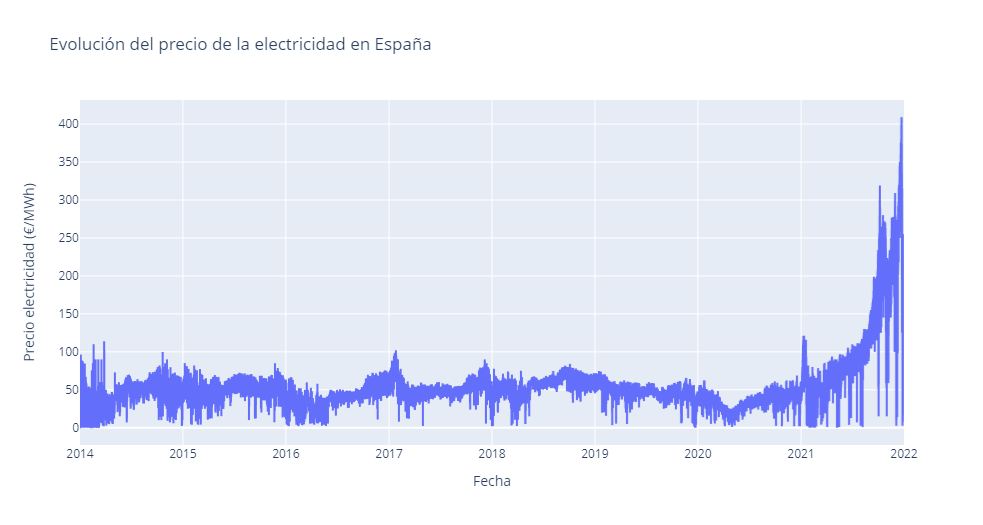


Imagen 5. Evolución del precio de la electricidad en España

Como hipótesis del trabajo (que más adelante se validarán) se proponen como posibles variables explicativas las siguientes:

* **Retardos de la variable objetivo**, es decir, tiene sentido pensar que el precio de la electricidad para cada una de las horas de mañana va a tener bastante correlación con el precio de las horas del precio de la electricidad de hoy o de las horas del mismo día de la semana anterior. Por ejemplo, el retardo de la variable objetivo de 1 día y de una semana.
* El **gas** es la materia prima que entra en el proceso de generación de electricidad por ciclo combinado y es una de las tecnologías que entran en el proceso de casación, por lo que esta variable también podría dar una idea de tendencia o niveles del precio de la electricidad.
* **Variables financieras** que son un reflejo del estado económico de la sociedad, por ejemplo, el IBEX35 (al estar en España) o el BRENT (petróleo crudo) son variables que pueden estar también relacionadas.
* **Hueco térmico:** las energías más limpias y baratas de conseguir son las generadas a partir de energías renovables. El problema, es que la demanda, pocas veces es cubierta íntegramente por energías renovables y ésta tiene que ser cubierta por otros procesos no renovables como la energía generada en centrales térmicas o ciclos combinados. A esta energía que no es cubierta por las energías renovables, y que es necesaria de procesos térmicos convencionales o ciclos combinados se denomina hueco térmico, que puede estar bastante correlacionada con el precio pues cuanto mayor sea el hueco térmico, más uso se deberán hacer de estas tecnologías que son las más caras para generar electricidad y por tanto, el precio suba. Se define como:

Denotando, HT, D y R, hueco térmico, demanda y renovables, respectivamente, en unidades de potencia. En la producción de renovables se incluyen las energías generadas por las tecnologías eólica, fotovoltaica, hidráulica y nuclear. Sin embargo, dado que la publicación de la previsión de las energías hidráulica y nuclear para el día D+1 es a partir de las 16:00, no son incluidas en la ecuación, pues el precio es publicado aproximadamente 2 horas antes. Por el contrario, la previsión horaria de las energías eólica y fotovoltaica para el día D+1 es publicada a las 11:00, por lo que éstas son las que entrarán en la ecuación del cálculo del hueco térmico.

* **Día de la semana y festivos**, pues tanto los negocios como los propios consumidores domésticos no consumen igual en fin de semana o festivo que en un día laboral. Dado que la demanda varía y está relacionada con el precio, esto debería afectar también al precio de la electricidad.
* **Datos climatológicos,** ya que están las renovables íntimamente ligadas a las condiciones climáticas, y esta generación de renovables impacta en el precio de la electricidad, valores como temperatura, radiación solar o viento, pueden estar relacionadas con el precio de la electricidad.

El objetivo es crear un modelo y testearlo out-of-sample para todo el periodo de 2021, es decir, 365 días x 24 horas = 8760 horas de predicción y comparación con el valor real.

Para evaluar los modelos, se usarán las siguientes métricas, disponibles y usadas en el script *metricas.py* del repositorio adjunto al trabajo.

Se harán uso y se estudiarán todas las métricas, pero la que se utilizará como métrica principal será el **WMAPE**. Se optó por no incluir la variable de MAPE (*Mean Absolute Percentage Error*, que es la media de los errores porcentuales en valor absoluto de cada observación), debido a la naturaleza de la variable objetivo ya que, en algunos periodos de tiempo en el que las renovables producen mucha energía (y el hueco térmico es cercano a 0), el precio es menor a 1 €/MWh, por lo que, aun realizando una predicción de 2 €/MWh, el valor de MAPE para esa predicción es del 100 %, lo que desvirtuaba la métrica en algunos periodos y no era buena elección, problema que se soluciona usando su versión ponderada (WMAPE).

Siendo ‘y’ el valor predicho, ‘x’ el valor real y ‘N’ el número de observaciones realizadas (número de horas consideradas en el estudio):

* **MAE** (*Mean Absolute Error)*. Función que calcula el error absoluto medio entre las predicciones y los valores reales.
* **WMAPE** (*Weight Mean Absolute Percentege Error).* Es el MAPE ponderado por el peso de la variable real. De esta forma, se minimizan los efectos de productos con grandes errores, pero cuyo valor es pequeño.
* **RMSE** (*Root Mean Squared Error)* . Se trata de la desviación estándar de los residuos (errores de predicción). Los residuos son una medida de cómo de lejos de la predicción están los valores reales.
* **% Tendencia**: es una variable que define únicamente si el precio de la electricidad va a aumentar o disminuir en la hora siguiente. Es una medida para determinar si el modelo está captando la tendencia de la serie horaria.

El “Nº horas con tendencia acertada” se define como:

1. **Material y métodos**
   1. **Procedimiento**

Todo el trabajo se realizó con Python 3.8.3 en distintos Jupyter Notebooks en el entorno de Anaconda (versión 2020.07), teniendo cada uno de ellos una función determinada, y los cuales vienen recogidos en el repositorio en GitHub adjunto al trabajo para facilitar la consulta del mismo. Las distintas versiones de los módulos utilizados a lo largo del trabajo están recogidas en el Anexo A.

El procedimiento fue el siguiente:

1. Creación del dataset de predicción del precio de la electricidad. A partir de las hipótesis postuladas anteriormente, el primer paso es recoger todas las variables y alinearlas con la variable objetivo, asegurándonos de que podemos disponer de todas ellas en el momento de la predicción.
2. Estudio de las variables explicativas y su correlación con la variable objetivo.
3. Creación de un modelo baseline. Antes de comenzar un ejercicio, se plantea un modelo sencillo. Por ejemplo, un modelo sencillo podría ser que la predicción del precio de la electricidad sea el valor medio del histórico que tengamos. En este caso, se propuso como modelo baseline que el precio de las 24 horas de la electricidad de mañana es igual al precio de las 24 horas de la electricidad de hoy. Este tipo de modelos tienen importancia antes de usar modelos más complejos pues si no hay variabilidad en los datos y este modelo tan sencillo tiene ya un MAPE muy bajo, no es atractivo usar modelos complejos pues, probablemente ni lo mejoren o si lo mejoran no compensa su utilización.
4. Definición de los modelos a usar. El objetivo del trabajo es resolver el problema de predicción del precio mediante distintos métodos (algoritmos) desde los más sencillos hasta los más complejos. Los definidos en el trabajo son los siguientes:
   1. Modelo baseline
   2. Modelo de regresión lineal
   3. Modelo de regresión Ridge y Lasso
   4. Modelo SARIMA
   5. Modelo SARIMA de los residuos de la regresión lineal
   6. Modelo Random Forest
   7. Modelo XGBoost
   8. Modelo de redes neuronales LSTM
5. Evaluación de los modelos en el mismo periodo y con las mismas métricas para poder determinar el que mejor se adapta al precio real de la electricidad.
6. Conclusiones y análisis de los resultados obtenidos.

Una vez establecido el procedimiento, se describe la construcción del dataset usado para la predicción. Todos los datos se descargaron de las urls que se mencionarán a continuación y se descargaron los datos y se incluyeron todos en la carpeta ‘Data’ disponible en el repositorio.

* 1. **Construcción de las variables explicativas**

La construcción de las variables explicativas se realiza en dos notebooks incluidos en el repositorio: “Preprocesamiento.ipynb” y “Preprocesamiento Climatologia y Finanzas.ipynb”. En el de preprocesamiento se tratan las variables de materias primas (precio de electricidad, gas, demanda energética, hueco térmico, dummies, festivos…) y en el de climatología y finanzas, los datos climatológicos y financieros como temperatura y BRENT, respectivamente.

Ya se ha hablado de la variable objetivo, así que se describe una a una las variables usadas y el origen.

* + 1. **Previsión MIBGAS D+1**

El MIBGAS es el responsable de la gestión del Mercado Organizado de Gas y es el nombre de la cotización del precio del gas en la Península Ibérica. En la página del MIBGAS se puede acceder al precio spot y varios precios forward (a 1 día, 1 mes, 1 año, entre otros). Como vamos a predecir el precio del día D+1, se obtiene esta variable en la siguiente url[[2]](#footnote-2) y cabe destacar que se trata de una variable de cotización diaria, no horaria como la electricidad. La serie obtenida es la siguiente (comienza en enero de 2016 porque es cuando empezó a cotizar este producto financiero). Se observa un pico muy alto en 2021 (correspondiente al temporal provocado por Filomena, con unos 60 €/MWh gas) y, tras recuperar sus valores normales, se observa la escalada de precios en la que estamos inmersa de precios, con precios superando todos los días los 120 €/MWh gas.

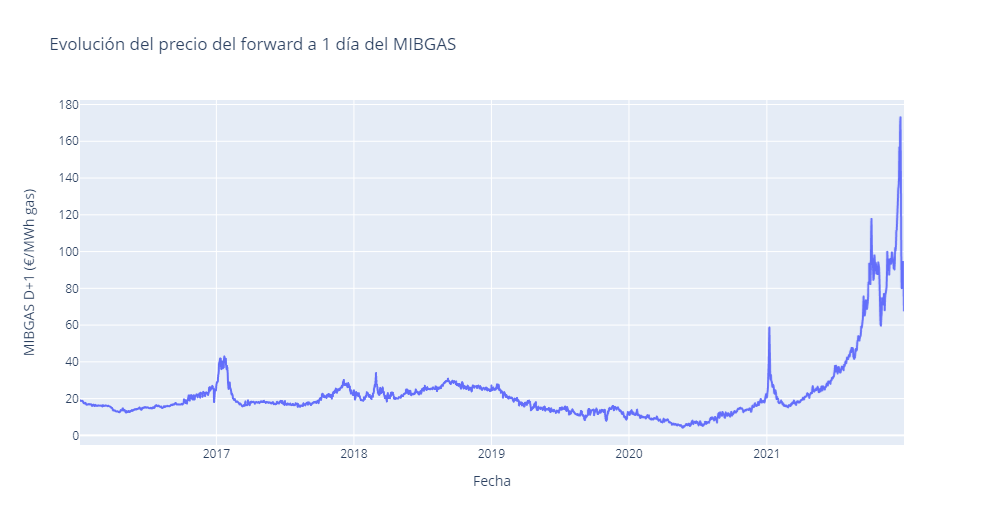


Imagen 6. Evolución del precio forward a 1 día del MIBGAS

* + 1. **Previsión de la demanda horaria D+1**

Dato también proporcionado por REE[[3]](#footnote-3). Consiste en una serie a nivel horaria en el que el dato corresponde con la previsión de la demanda de energía en España. Se observa una tendencia estacional en la que marca máximos de demanda en los meses de climas extremos, ya sean fríos (diciembre y enero) o los más cálidos (julio y agosto). Se observa antes de julio 2020 un periodo atípico de bajo consumo que corresponde al periodo de confinamiento debido al Covid-19.

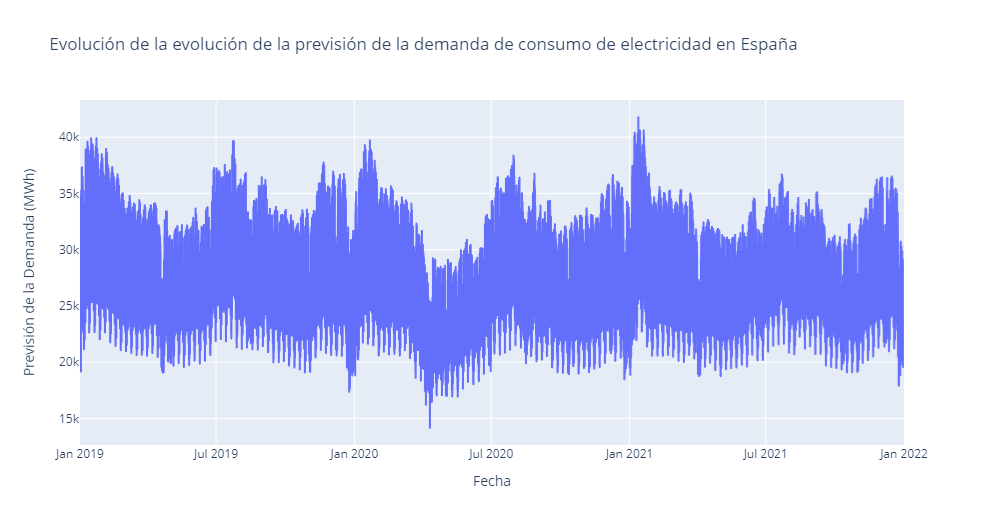


Imagen 7. Evolución de la previsión de la demanda D+1 desde 2019

* + 1. **Previsión de la generación eólica + fotovoltaica D+1**

Dato también publicado en la web de REE[[4]](#footnote-4). Se trata de otra serie a nivel horaria cuyo dato representa la cantidad de energía, en MWh, generada mediante las tecnologías eólica y fotovoltaica de forma conjunta. Como está la suma de ambas tecnologías, y la eólica tiene mucha más potencia instalada (y genera más que la fotovoltaica), no se observa la estacionalidad de que en verano haya más producción, es más, agregando mensualmente, los meses en los que más producción hay son octubre, noviembre, diciembre y enero.

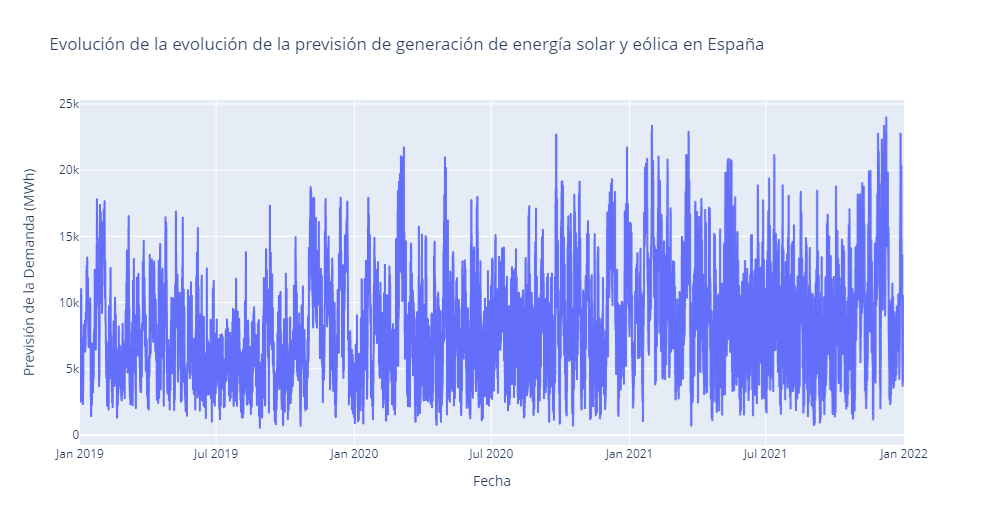
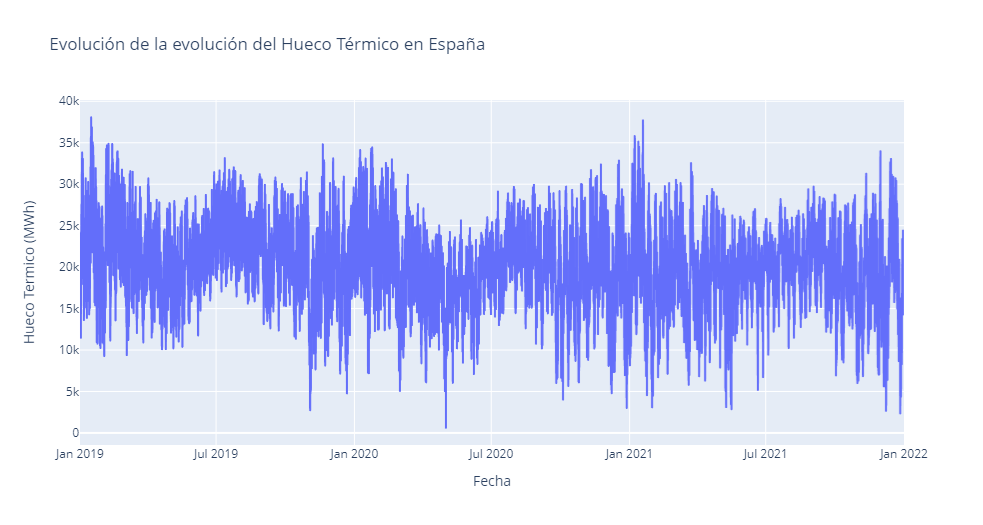


Imagen 8. Evolución de la generación de eólicas + fotovoltaicas D+1 desde 2019

* + 1. **Hueco Térmico**

Calculado mediante la fórmula del hueco térmico de la sección 3. Se construye como la resta de la serie de la sección 4.2.2 menos la de la sección 4.2.3 (demanda menos la generación de renovables), por lo que se trata también de una serie con desglose horario. Esto nos da la cantidad de demanda que hay que cubrir mediante centrales térmicas y de ciclo combinado por lo que, a menor valor de esta variable, quiere decir que la demanda ha sido cubierta en mayor medida por las renovables, por lo que el precio de la electricidad debería ser menor.

Imagen 9. Evolución del hueco térmico previsto D+1 desde 2019

* + 1. **Dummies, festivos y otras**

Como variables dummies se escogieron la hora del día (desde las 0:00 hasta las 23:00) y el de la semana en que nos encontremos (de lunes a domingo). Estas variables valen 1 cuando cumplan estén en la hora o día de la semana correspondiente y 0 en el resto de los casos.

La misma idea ocurre con los festivos (distinguiendo entre festivos nacionales y regionales por comunidad autónoma introducidos a mano en un Excel). La fecha a predecir que se encuentre en este Excel tendrá en todas las horas de ese día la variable 1 y, en caso contrario, de 0.

Otras variables consideradas han sido el precio máximo y mínimo. Para predecir D+1 se usa como variable para todas las horas del día de D+1 el precio máximo y precio mínimo del día en el que nos encontremos, definidas dos variables distintas con el objetivo de poner unos niveles entre los que, más o menos, debería moverse el precio. También se incluye la diferencia entre estas dos como otra variable (llamada Spread\_precio) que recogería si el día anterior tuvo mucha volatilidad (la diferencia entre el precio máximo y mínimo del día anterior fue muy alta) o poca volatilidad (estas dos variables tuvieron valores similares).

* + 1. **Climatológicas**

Se incluyeron datos acerca de condiciones meteorológicas, ya que éstas están directamente relacionadas con la producción de las renovables. Se obtuvieron de la web de datos meteorológicos de Madrid[[5]](#footnote-5). En ella se encuentran un montón de datos para distintas estaciones meteorológicas. Se escogió la estación meteorológica de Alcobendas y se obtuvieron las variables de velocidad del viento, temperatura, humedad relativa, radiación solar y precipitación (más información en el pdf en la ruta del repositorio ‘Data/documentación\_datos\_climatologicos.pdf’). Los datos son históricos y a nivel horario, por lo que para usarlos como predicción se cogió su retardo de 24 horas. También se crearon las variables temperatura máxima, mínima y spread de temperatura con el mismo objetivo y procedimiento que el descrito para el precio en la sección 4.2.5.

* + 1. **Financieras**

Las variables financieras se obtuvieron mediante el módulo *yfinance* de Python, una api que carga directamente los datos financieros de *Yahoo Finance*. Las variables que se obtuvieron fueron el IBEX35 (por encontrarnos en España), el BRENT (petróleo crudo) y el API2 (cotización del carbón) que es la que está directamente relacionada con las centrales térmicas. Otras variables que podrían haberse incluido son el TTF (precio del gas que se suele comerciar en Europa, que se decidió no incluir por tener ya el gas que se comercializa en España) o el EUA CO2 (derechos de emisiones de CO2 a la atmósfera, que tampoco se incluyó por no disponer tampoco de previsiones y no encontrarse en *Yahoo Finance*). Estas cotizaciones son diarias e históricas, por lo que para usarlas para predecir y alinearlas al precio de la electricidad, también se tomaron sus retardos de 24 horas.

* 1. **Relación entre las variables**

En toda la sección 4.2. se han descrito todas las variables que se recogieron como hipótesis que podrían tener sentido en la predicción del precio de la electricidad. Esta sección se basa en estudiar las relaciones de estas variables con la variable objetivo y seleccionar/descartar algunas, aunque muchos de los algoritmos que se usan en la actualidad ya tienen en cuenta la importancia de las variables y descartan (o no le dan ningún peso en la predicción) las variables que menos relación tienen con la objetivo.

1. **Resultados**
   1. **Modelo baseline**
   2. **Modelo de regresión lineal**
   3. **Modelo de regresión Ridge y Lasso**
   4. **Modelo SARIMA**
   5. **Modelo SARIMA de los residuos de la regresión lineal**
   6. **Modelo Random Forest**
   7. **Modelo XGBoost**
   8. **Modelo de redes neuronales LSTM**
   9. **Modelo Media de las predicciones**
   10. **Modelo Stacked model**
2. **Discusión**
3. **Conclusiones**
4. **Referencias**
5. La energía es un bien básico de primera necesidad cuyo acceso debe ser garantizado como servicio público (2022). Consultado el 12 de abril de 2022. <https://www.energias-renovables.com/panorama/la-energia-es-un-bien-basico-de-20191119>
6. España – Consumo de electricidad (2022). Consultado el 12 de abril de 2022. <https://datosmacro.expansion.com/energia-y-medio-ambiente/electricidad-consumo/espana>
7. ¿Cómo funciona el sector eléctrico en España? (2021). Consultado el 12 de abril de 2022. <https://www.aura-energia.com/como-funciona-el-sector-electrico-en-espana/#:~:text=Una%20comercializadora%20de%20electricidad%20compra,trav%C3%A9s%20de%20un%20precio%20pactado>
8. BOE: <https://www.boe.es/buscar/act.php?id=BOE-A-2013-13645&p=20211027&tn=1#a30>
9. El sistema marginalista de fijación de precios eléctricos, a debate. Alejandro Nieto Gonzalez (2021). Consultado el 3 de mayo de 2022. <https://www.elblogsalmon.com/sectores/sistema-marginalista-fijacion-precios-electricos-a-debate>
10. Generación en España. (2022). Consultado el 3 de mayo de 2022. [https://www.ree.es/es/datos/generacion/potencia-instalada](https://www.ree.es/es/datos/generacion/potencia-instaladaG)
11. Generación en España (2022). Consultado el 4 de mayo de 2022. <https://www.ree.es/es/datos/generacion/estructura-generacion>
12. Nuestros hábitos de consumo (2021). Consultado el 4 de mayo de 2022. <https://www.ree.es/es/red21/eficiencia-energetica-y-consumo-inteligente/nuestros-habitos-de-consumo>
13. **ANEXOS**

**8.1. ANEXO A.**

Versiones de los paquetes más importantes utilizados en el trabajo.

|  |  |
| --- | --- |
| **Paquete** | **Versión** |
| Pandas | 1.0.5 |
| Numpy | 1.19.5 |
| Matplotlib | 3.2.2 |
| Statsmodels | 0.12.2 |
| Plotly | 5.4.0 |
| Scikit-learn | 0.23.1 |
| Tensorflow | 2.4.0 |
| Keras | 2.4.3 |
| Yfinance | 0.1.70 |
| Xgboost | 1.3.3 |
| Scipy | 1.5.0 |
| Seaborn | 0.10.1 |

1. https://www.esios.ree.es/es/analisis/600?vis=1&start\_date=01-01-2021T00%3A00&end\_date=30-11-2021T23%3A00&compare\_start\_date=31-12-2020T00%3A00&groupby=hour&geoids=3 [↑](#footnote-ref-1)
2. <https://www.mibgas.es/es/file-access> [↑](#footnote-ref-2)
3. <https://www.esios.ree.es/es/analisis/1775?vis=1&start_date=06-10-2016T00%3A00&end_date=31-12-2021T23%3A50&compare_start_date=05-10-2016T00%3A00&groupby=hour> [↑](#footnote-ref-3)
4. <https://www.esios.ree.es/es/analisis/10358?vis=1&start_date=01-10-2015T00%3A00&end_date=31-12-2021T23%3A50&compare_start_date=30-09-2015T00%3A00&groupby=hour> [↑](#footnote-ref-4)
5. <https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=fa8357cec5efa610VgnVCM1000001d4a900aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default> [↑](#footnote-ref-5)