

A large, stylized lightning bolt strikes down from the top right towards the bottom left, illuminating a dark, stormy sky. The bolt is composed of many smaller, branching strokes, creating a complex, web-like pattern.

# ELECTRIZER

**Javier Salinas Fernández**

Data Science Bootcamp Enero – Marzo 2021

# Índice

<b>1. Descripción del proyecto</b>	<b>2</b>
<b>2. Contexto</b>	<b>2</b>
<b>3. Etapas del proyecto</b>	<b>5</b>
3.1. Obtención datos meteorológicos . . . . .	5
3.2. Obtención datos electricidad . . . . .	6
3.3. Obtención conjunto de datos final: pretratamiento y limpieza de los datos . . . . .	6
3.4. Análisis exploratorio de datos . . . . .	7
3.5. Regresión . . . . .	11
3.6. Horas pico y valle . . . . .	13
<b>4. Conclusiones</b>	<b>15</b>

# 1. Descripción del proyecto

Desde hace ya unos años se oye con más frecuencia, ya sea en televisión, periódicos o incluso redes sociales, informaciones sobre el precio de la electricidad, lo cara que está en algunas épocas o cuándo debemos poner la lavadora para ahorrar en la factura. Esta continua variación en el precio hace que muchos consumidores (principalmente familias) se acaben llevando sorpresas a final de mes cuando les llega la factura.

En este proyecto, además de explicarse de manera resumida los distintos conceptos por los que se paga en la factura de la luz, se estudian y analizan las principales variables (hora, día de la semana y datos meteorológicos) por las que varía tanto el precio, buscando como objetivo final **predecir el precio de la electricidad** y encontrar los momentos de la semana en los que **la relación precio-consumo es más cara**.

# 2. Contexto

Antes de nada hay que explicar brevemente en qué situación se encuentra el actual modelo eléctrico en España, y es que desde hace algo más de una década existen dos mercados eléctricos en nuestro país: el mercado regulado y el mercado libre. En ambos mercados se encuentran los siguientes conceptos que componen la factura de la luz que llega todos los meses a casa:

- **Término de potencia.** Como su propio nombre indica, hace referencia a la potencia que se tiene contratada, es decir, a que se tenga disponible durante todo el tiempo los kW contratados. Se compone de dos conceptos:
  - **Peaje de acceso del término de potencia.** Es un precio fijo (está regulado por el Gobierno, se esté en el mercado regulado o libre) y únicamente varía en función de la potencia contratada.
  - **Margen de comercialización.** Varía en función de la potencia contratada y también lo fija el Gobierno, aunque solo para el mercado regulado (en el mercado libre, cada empresa puede fijar su precio).
- **Término de energía.** Es propiamente lo que se cobra por el consumo de electricidad y donde se encuentra la principal diferencia entre los dos tipos de mercado. También está compuesto de dos conceptos:
  - **Peaje de acceso del término de energía.** Al igual que sucedía con el peaje del término de potencia, se trata de un precio fijo fijado por el Gobierno, independientemente del tipo de mercado. En esta ocasión el precio varía en función del consumo.
  - **Coste de la electricidad.** Este es el concepto que diferencia en mayor medida el mercado regulado y el mercado libre. Para el primero el precio

se fija en función del coste de producción, y por tanto varía en función de la hora y día; para el segundo este precio dependerá del acuerdo al que se llegue con la empresa que suministra la electricidad, pudiendo ser un precio constante para todo el tiempo o que también varíe en función de la hora a la que se realice el consumo.

- **Impuesto sobre la electricidad.** Impuesto especial que se aplica sobre la suma de los términos de potencia y de energía. Como todos los impuestos, viene fijado por el Gobierno y afecta a todos los consumidores, sean de un mercado u otro.
- **Alquiler de los equipos de medida y control.** Es lo que comúnmente se conoce como *alquiler del contador*. Es un concepto que también fija el Gobierno y que varía dependiendo del tipo de equipo del que se disponga. En caso de que se tenga este equipo en propiedad no se paga por este concepto, aunque esto sucede raramente.
- **IVA.** Que hubiera ya un impuesto especial sobre la electricidad no hace que el IVA desaparezca. Actualmente se encuentra en el 21 % y se aplica sobre los cuatro conceptos ya explicados anteriormente.

Es posible que, al contratar la electricidad en el mercado libre, la compañía en cuestión nos proponga un precio fijo por kWh consumido, pareciendo que algunos de los conceptos explicados anteriormente (sobre todos los peajes de acceso) desaparecen, pero en realidad esto no es así, ya que los conceptos que son fijos se han de pagar siempre.

Como se ha visto, la principal diferencia entre ambos mercados se encuentra en el concepto de coste de la electricidad. Dado que en el mercado libre se pueden encontrar decenas de distintas tarifas, cuya variación no radica en ningún concepto que a priori se pueda analizar (va en función de las tarifas que cada empresa quiera fijar o que se negocien con ella), en este proyecto se analiza la variación de este concepto en el mercado regulado.

De aquí en adelante, cuando se hable del precio o coste de la electricidad, hay que entender que se hace referencia a este concepto, y no al coste total de lo que podría ser una factura. Esto es así porque, por ejemplo, el importe de los peajes de potencia y energía cambian cuando el Gobierno lo decide (de hecho en verano de 2021 está previsto que vuelvan a cambiar), y no es algo que se pueda predecir. Por otra parte, el coste de producir la electricidad es obvio que también cambia (de hecho lo hace constantemente), pero es algo que sí es posible de predecir, y es lo que se hace en este proyecto.

En el mercado regulado, para fijar el coste de la electricidad, se emplea lo que se conoce como precio marginal. Esto quiere decir que el precio que finalmente se paga por MWh depende del último que se adjudique. Para explicarlo mejor se supone un ejemplo en el que, para un determinado momento, se necesitan 100 MWh y se disponen de 4 posibles tipos de energía para producirlos:

- Energía nuclear. Puede aportar 50 MWh a un precio de 20€/MWh
- Energía eólica. Puede aportar 30 MWh a un precio de 30€/MWh
- Ciclos combinados. Puede aportar 35 MWh a un precio de 50€/MWh
- Carbón. Puede aportar 40 MWh a un precio de 80€/MWh

La asignación de cuánto produce cada tipo de energía se hace en función del precio. De esta forma, todos los MWh de energía nuclear y eólica se producen, mientras que de ciclos combinados solo se producen 20 MWh (llegando así a los 100 MWh en total). El resto, incluidos todos los MWh del carbón, no se producen.

Como la última energía que sí produce es la de los ciclos combinados y su precio es de 50€/MWh, todos los MWh (también los de la energía nuclear y eólica) tienen este precio. Esto es lo que implica el sistema de precio marginal.

Con este pequeño ejemplo se ve que las principales variables que influyen a la hora de fijar el precio son el tipo de tecnología que se use para producir la electricidad y cuánto sea la demanda. En lo relativo a los tipos de tecnologías usadas para la producción, en la figura 1 puede verse la distribución entre energías que tuvo lugar durante el año 2019.

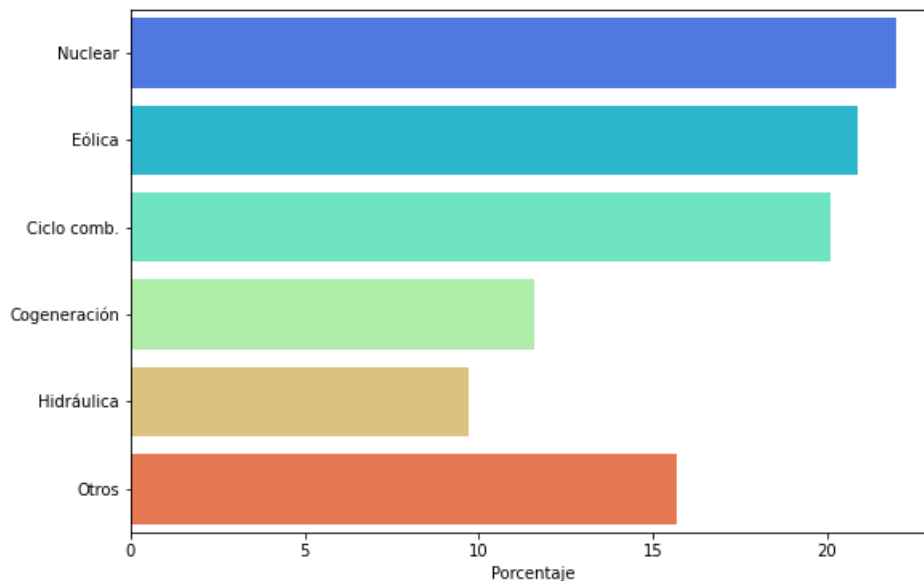


Figura 1: Distribución de la producción de energía eléctrica en España (año 2019)

Las centrales nucleares, debido a sus características técnicas, se encuentran continuamente en funcionamiento (prácticamente 24 horas al día, 7 días a la semana), ya que realizar paradas y arranques de la planta suponen un alto coste económico. Es por ello que en este tipo de energía no hay mucho que analizar.

Todo lo contrario sucede con las energías renovables, cuya variabilidad es muy alta, ya que dependiendo de las condiciones meteorológicas producen más o menos,

incluso llegando a no producir nada en determinadas situaciones (por ejemplo, ausencia de viento o precipitaciones de forma prolongada).

Por otra parte, los ciclos combinados y la cogeneración también tienen poca variabilidad, ya que aunque estas instalaciones no funcionan todo el tiempo, sí se pueden poner en marcha cuando el sistema lo requiere.

Conociendo estas variabilidades y viendo los precios de producción del ejemplo (no son valores reales, pero sí es exacto en la medida en que se establecen cuáles son las energías más baratas y las más caras), se llega la conclusión de que saber la participación que tienen las energías renovables en un momento concreto es vital para poder predecir el precio de la electricidad.

En cuanto al consumo, que junto con el tipo de tecnología productora era la otra variable importante, generalmente se alcanzan los picos durante la mañana y por la tarde-noche, mientras que las horas valle suelen tener lugar durante la madrugada. Es por tanto previsible que los precios horarios más altos se encuentren en esas franjas, aunque todo esto también se confirma a lo largo del proyecto.

## 3. Etapas del proyecto

### 3.1. Obtención datos meteorológicos

Lo primero de todo es definir el intervalo temporal en el que se realiza el estudio. Dado que cada año puede ser muy distinto al anterior (tanto a nivel meteorológico como en términos de consumo y producción de energía eléctrica), es importante escoger un amplio rango, para conseguir que la variabilidad se compense entre los distintos años. Por ello, el intervalo de tiempo que se ha decidido estudiar va **desde el 1 de enero de 2015 hasta el 28 de febrero de 2021**.

Como se comentó anteriormente, resulta imprescindible analizar las condiciones meteorológicas de cara a poder predecir el precio de la electricidad. Estos datos se obtienen de la Agencia Estatal de Meteorología (AEMET), más concretamente a través de su API (<https://opendata.aemet.es/centrodedescargas/inicio>).

Esta API tiene dos tipos de consulta diferenciadas: una en la que se pueden realizar de manera puntual y manual, pensada para la obtención de pocos datos; y otra en la que es posible automatizar esta consulta (a través de código en un programa cliente, como Python), pensada para la obtención de mayores cantidades de datos. Este segundo tipo es el que se ha usado en este proyecto.

Una vez seleccionado el tipo de consulta, hay que elegir los datos que se quieren obtener, ya que la AEMET pone a disposición multitud de datos diferentes, desde predicción marítima hasta información sobre las rayos que caen. Para este proyecto se escoge la sección *Valores climatológicos*, que da información diaria de las princi-

pales variables meteorológicas por cada una de las estaciones que hay instaladas.

Aunque hay más variables, las que finalmente se escogen son:

- Temperatura media (en grados centígrados)
- Precipitaciones (en mm)
- Velocidad del viento (en m/s)
- Horas de sol (según la AEMET, esta variable hace referencia al intervalo de tiempo durante el cual la radiación solar alcanza intensidad suficiente para producir sombras bien diferenciadas)

En total se acaban escogiendo 1084 variables (271 estaciones meteorológicas por las 4 variables citadas de cada una de ellas). Salvo el número de horas de sol, la información de las variables meteorológicas viene dada como la media diaria por cada estación.

### **3.2. Obtención datos electricidad**

Los datos relacionados con la electricidad, principalmente el precio y consumo horario, se obtienen de Red Eléctrica de España (REE). Esta empresa pública se encarga del transporte de la energía eléctrica, es decir, de llevarla desde los puntos de producción hasta las subestaciones de distribución, donde continúa el proceso para llevar la electricidad hasta los consumidores finales, aunque ya a menor escala.

Para facilitar el acceso a los datos que entran dentro de su competencia, REE pone a disposición de toda la sociedad una página web (<https://www.esios.ree.es/es>) en la que se encuentran datos actualizados sobre muchos aspectos, como la generación, el consumo o el precio.

Esta página también posee datos de la serie histórica, que son los que se usan en este proyecto. Una vez que se fija el tipo de dato y el rango temporal que se quiere obtener, es posible descargar la información en un archivo Excel o CSV (aunque en este último caso los datos no aparecen separados por comas, sino por punto y comas).

### **3.3. Obtención conjunto de datos final: pretratamiento y limpieza de los datos**

Una vez que se tienen, por separado, los datos meteorológicos y los datos relativos a la electricidad es necesario juntarlos en un único conjunto de datos. Para ello, a cada registro de los datos eléctricos (que se corresponde con cada una de las horas del rango temporal del estudio, por lo que en total hay 54024) se le añaden como atributos los datos de las variables meteorológicas de cada una de las estaciones.

Como estos datos meteorológicos son datos diarios, mientras que los datos de precio y consumo de la electricidad son horarios, se tiene un conjunto de datos en los que las variables meteorológicas tienen el mismo valor (para cada estación) durante todo el día, es decir, para 24 registros.

Sin embargo, tal y como suele suceder cuando se tratan grandes cantidades de datos, alguno de estos valores no se encuentra definido. Esto sucede en la parte de datos meteorológicos (no así en los datos eléctricos, que se encuentran completos), probablemente debido a algún problema puntual en algunas de las estaciones. También hay algunas estaciones que no tienen los sensores adecuados para medir las horas de sol, por lo que en ese caso los valores son siempre desconocidos.

Para tratar estos datos, como ya desde el principio se tenía la idea de agrupar las estaciones meteorológicas por provincias (para facilitar el posterior modelado de los datos), se decide realizar dicho agrupamiento ahora (calculando el valor medio de cada variable), ya que de esta manera el número de valores desconocidos disminuye.

Una vez realizado el agrupamiento se siguen detectando valores desconocidos (aunque en menor cantidad que los que había antes), por lo que en esta ocasión para imputarlos se decide hacer la media aritmética, para esa variable meteorológica y provincia, del valor 2 días antes, 1 día antes, 1 día después y 2 días después.

Tras esta imputación ya solo queda la variable que hace referencia a las horas de sol en la provincia de Huelva con algunos valores desconocidos, situados al principio de la serie temporal que se está estudiando (es decir, en los primeros días de enero de 2015). Como no se tienen datos de los días anteriores (finales de diciembre 2014), es necesario utilizar otra estrategia.

Se decide emplear la imputación con los k-vecinos más cercanos (librería *scikit-learn*, clase *sklearn.impute.KNNImputer*), una estrategia que lo que finalmente imputa es la media de los valores de los 250 patrones más cercanos, entendiéndose como cercanos aquellos patrones que tienen valores similares en el resto de variables.

Tras esto, ya se tiene el conjunto de datos con todos los valores definidos y listo para continuar con las siguientes etapas.

### 3.4. Análisis exploratorio de datos

En este punto, el conjunto de datos que se tiene consta de 54.024 registros y 219 atributos:

- 208 atributos ligados a variables meteorológicas: temperatura, precipitaciones, velocidad de viento y horas de sol por cada una de las 50 provincias y Ceuta y Melilla
- 7 atributos (dummies) haciendo referencia a cada uno de los días de la semana



- Consumo horario
- Fecha (en formato año-mes-día), aunque al tratarse de una cadena de texto apenas se usa en el análisis exploratorio y en los modelos de regresión
- Hora
- Precio (esta es la variable objetivo)

Lo ideal sería poder hacer, en una única vez, una tabla o gráfica de correlaciones entre todas las variables, de manera que visualmente se vieran aquellas que más correlación tienen con el precio de la electricidad. Sin embargo, dada la gran cantidad que hay, esta opción no es viable.

Para analizar la correlación con el precio de las variables meteorológicas se hace un análisis de cada uno de los 4 grupos que hay (temperatura, precipitaciones, velocidad de viento y horas de sol). En la tabla 1 se ve, para cada uno de estos grupos, el sumatorio de las correlaciones (por provincia) y los valores individuales mínimo y máximo. Todos ellos están en valor absoluto.

Tabla 1: Análisis de correlación del precio con los 4 tipos de variables meteorológicas

Variable	Sumatorio	Mínimo	Máximo
<b>Temperatura</b>	1.03	0.001	0.05
<b>Precipitaciones</b>	3.75	0.002	0.16
<b>Velocidad viento</b>	12.77	0.045	0.38
<b>Horas sol</b>	1.79	0.001	0.12

De los resultado de la tabla 1 se observa que la velocidad del viento y, en menor medida, las precipitaciones son las variables meteorológicas que mayor correlación tienen con el precio.

Por otro lado, en la tabla 2 se tienen las variables individuales con una mayor correlación con el precio. De todas ellas destaca el consumo, con una correlación mayor a 0.5. Sin embargo, es importante destacar que todas las demás se corresponden con la velocidad del viento en distintas provincias. Esto refuerza la conclusión ya obtenida antes, y es que dentro de las variables meteorológicas, el viento es la que más influye.

También es importante señalar el signo de la correlación. En el consumo se observa que es positivo, lo que implica que al aumentar el consumo aumenta el precio. Esto es algo ya comentado anteriormente y que se esperaba, ya que si aumenta el consumo ha de aumentar la producción eléctrica, entrando a producir tecnologías cuyo precio es más caro.

En cuanto al signo de la correlación del viento es negativo, lo que significa que disminuye el precio al aumentar la contribución de la energía eólica. Esto también era algo esperable, ya que el precio de producción de las energías renovables es bajo,

Tabla 2: Variables con correlación igual a 0.3 o mayor (en valor absoluto)

<b>Variable</b>	<b>Correlación</b>
Viento - Toledo	-0.312
Viento - Granada	-0.324
Viento - Albacete	-0.314
Viento - Madrid	-0.325
Viento - Valencia	-0.303
Viento - Ávila	-0.382
Viento - Zamora	-0.334
Viento - Valladolid	-0.320
Viento - Salamanca	-0.300
Viento - Soria	-0.300
Viento - Palencia	-0.309
Viento - Burgos	-0.341
Viento - León	-0.337
Consumo	0.532

por lo que si la cantidad que producen es más alta provoca que otras tecnologías más caras no hagan falta para atender toda la demanda.

Dejando de lado el tema de las correlaciones, otro análisis que se puede efectuar es ver en qué momentos de todo el rango temporal analizado se han alcanzado los precios más altos (y también los más bajos), para ver que tienen en común entre todo ellos.

Realizando la media de los 50 valores horarios con el precio más alto y de los otros 50 con el precio más bajo se obtienen los resultados de la tabla 3.

Tabla 3: Características de los precios más altos y bajos

<b>Variable</b>	<b>Horas máximo</b>	<b>Horas mínimo</b>
<b>Precio</b>	120.05	6.03
<b>Consumo</b>	38699	23326
<b>Hora</b>	17.8	5.9
<b>Temperatura</b>	3.8	11.3
<b>Precipitaciones</b>	5.00	3.57
<b>Velocidad viento</b>	2.56	4.77
<b>Horas sol</b>	3.78	3.65

Antes de entrar en detalle, es conveniente señalar que la variable *Hora* de la tabla superior, al ser una media, es un valor que no hay que tomar de manera literal, sino que indica aproximadamente el momento del día en que se producen esos precios máximos y mínimos.

Los precios máximos se observa que tienen lugar en días fríos y con una veloci-

dad el viento no baja, pero tampoco excesivamente alta. También se ve que la hora a la que tienen lugar es por la tarde, cuando en general la gente ya ha regresado a su casa del trabajo y probablemente enciendan una mayor cantidad de equipos eléctricos (lavadora, lavavajillas, ordenador, televisión, plancha, etc).

En cuanto a los precios mínimos, suelen suceder en días templados, con una velocidad del viento elevada (y por tanto una alta contribución de la energía eólica) y durante la madrugada, cuando la mayor parte de la población está durmiendo y por tanto no consumiendo electricidad.

El análisis efectuado antes es horario, pero también se puede hacer en formato diario, para ver si hay alguna diferencia significativa. Esta vez se escogen los 25 valores más altos y los 25 más bajos, obteniendo como resultados los mostrados en la tabla 4. En general se obtienen conclusiones similares a las del análisis horario.

Tabla 4: Características de los precios más altos y bajos

<b>Variable</b>	<b>Días máximo</b>	<b>Días mínimo</b>
<b>Precio</b>	89.50	14.61
<b>Consumo</b>	802335	606258
<b>Temperatura</b>	5.9	12.1
<b>Precipitaciones</b>	2.19	4.03
<b>Velocidad viento</b>	2.48	4.3
<b>Horas sol</b>	5.49	4.77

Todos estos datos y conclusiones son bastante claros, pero claro, cuando se quiere predecir el precio se quiere hacer con cierta antelación (por ejemplo con 24 horas de antelación), y en ese instante que hay que predecir no se tienen los datos meteorológicos ni el consumo actuales.

Es por ello que para hacer la predicción se usan los datos meteorológicos del día anterior, mientras que para el consumo se emplean los datos de la semana anterior. La razón por la que para esta última variable no se usan los datos del día anterior se observa en la figura 2, donde se aprecia una clara disminución en el consumo durante los días de fin de semana (especialmente el domingo).

Si se analizan las correlaciones para este nuevo escenario se obtienen los resultados de la tabla 5. En ella se observa que las correlaciones han disminuido, algo que por otra parte era lógico, ya que los valores meteorológicos y de consumo de los días anteriores, por muy parecidos que sean, nunca serán iguales a los que suceden en el día actual.

No obstante, las variables que más correlación tenían antes (tabla 2) siguen siendo las que más tienen ahora.

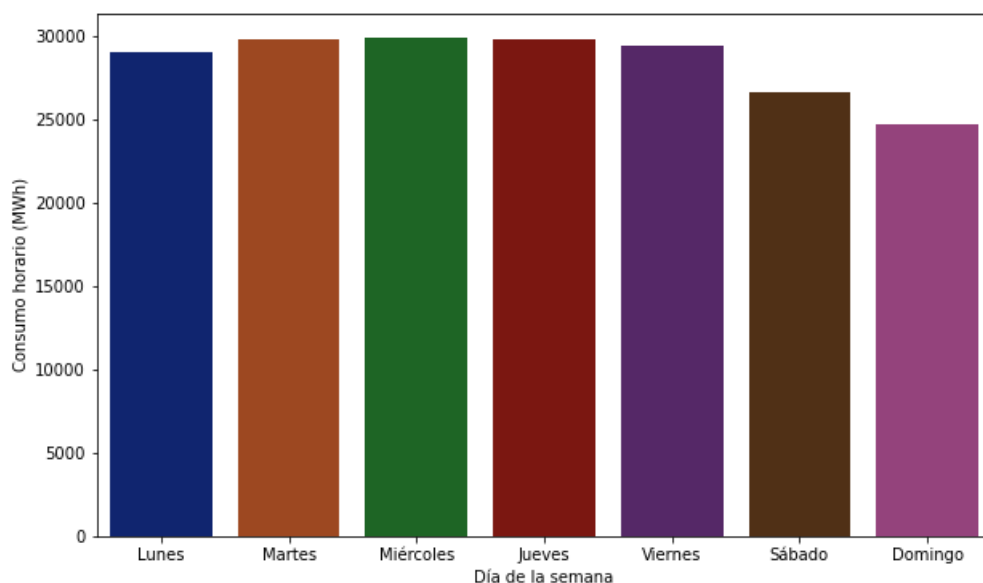


Figura 2: Consumo de energía horario en función del día de la semana

Tabla 5: Variables con correlación igual a 0.25 o mayor (en valor absoluto) para el conjunto de datos con variación temporal

Variable	Correlación
Hora	0.251
Viento - Toledo	-0.253
Viento - Granada	-0.275
Viento - Madrid	-0.256
Viento - Lugo	-0.284
Viento - Ávila	-0.309
Viento - Zamora	-0.292
Viento - Valladolid	-0.266
Viento - Salamanca	-0.253
Viento - Palencia	-0.257
Viento - Burgos	-0.272
Viento - León	-0.290
Viento - Ourense	-0.262
Consumo	0.483

### 3.5. Regresión

Una vez que ya se tienen las conclusiones del análisis exploratorio de los datos es momento de comenzar a probar diferentes modelos de regresión y buscar aquel que tenga una precisión mayor a la hora de predecir el precio. Todas las técnicas usadas en esta sección pertenecen a la librería *scikit-learn*.

Lo primero de todo es normalizar los valores del conjunto de datos, es decir,

convertirlos a valores entre 0 y 1. Obviamente, la variable que indica la fecha, al ser una cadena de texto, hay que eliminarla antes de normalizar los valores.

Los modelos de regresión que se prueban son:

- Árbol de regresión
- Random Forest
- Ada Boost
- Gradient Boosting

Para separar el conjunto de datos en los sets de entrenamiento y test se usa *train\_test\_split*, de forma que el set de entrenamiento es el 70 % de todos los registros, mientras que el 30 % restante corresponde al set de test.

El modelo de validación que se usa para todos los modelos de regresión es *Hold-out*, de manera que se repite el mismo procedimiento (separación del conjunto de datos en entrenamiento y test; entrenamiento del modelo de regresión; predicción del set de test) un total de 100 veces.

Como métricas para los modelos se usan la  $R^2$  ajustada y una medida del error propia y simple consistente en la media del error absoluto, es decir, el sumatorio de las diferencias entre el valor real y la predicción en valor absoluto, dividido entre el número total de predicciones. Dado que los valores están normalizados, esta medida del error medio es un valor pequeño, de apenas centésimas.

Los resultados que se han obtenido para los distintos modelos de regresión se muestran en la tabla 6.

Tabla 6: Resultados de los modelos de regresión

Modelo	$R^2$ ajustada	Error medio
Árbol de regresión (por defecto)	0.901	0.0236
Árbol de regresión (Grid Search)	0.765	0.0437
<b>Random Forest (por defecto)</b>	<b>0.951</b>	<b>0.0177</b>
Ada Boost (por defecto)	0.413	0.0752
Ada Boost (Grid Search)	0.424	0.0752
Gradient Boosting (por defecto)	0.665	0.0534

Para el árbol de regresión ajustado a partir del *Grid Search* se han modificado los siguientes parámetros con respecto al que es por defecto:

- *criterion*: friedman\_mse (por defecto era mse)
- *max\_depth*: 100 (por defecto no había límite)
- *max\_leaf\_nodes*: 500 (por defecto no había límite)

Tabla 7: Desviaciones típicas de los resultados de los modelos de regresión

Modelo	Desv. $R^2$ ajustada	Desv. error medio
Árbol de regresión (por defecto)	0.004	0.0004
Árbol de regresión (Grid Search)	0.008	0.0007
<b>Random Forest (por defecto)</b>	<b>0.001</b>	<b>0.0002</b>
Ada Boost (por defecto)	0.010	0.0006
Ada Boost (Grid Search)	0.009	0.0006
Gradient Boosting (por defecto)	0.004	0.0004

- *min\_samples\_split*: 5 (por defecto era 2)

En cambio, para el Ada Boost ajustado a partir del *Grid Search* los parámetros que se han modificado con respecto a los que vienen por defecto son:

- *learning rate*: 2 (por defecto era 1)
- *loss*: exponential (por defecto era linear)
- *n\_estimators*: 100 (por defecto era 50)

Como se observa en las tablas 6 y 7, de los 6 modelos de regresión que se han estudiado, el *Random Forest* es el que proporciona unos mejores resultados, con una  $R^2$  ajustada de más de 0.95 y un error absoluto medio de apenas 0.0002.

En el caso de los modelos de árbol de decisión se obtiene mejor resultado con el que es por defecto que con el que, en teoría, está optimizado con *Grid Search*. Esto se debe principalmente a que en la configuración por defecto no hay límite en la profundidad del árbol ni en el número de nodos, mientras que para el otro caso sí se han fijado límites.

Situación distinta es la que ocurre con los dos modelos de regresión de Ada Boost: el optimizado con *Grid Search* sí consigue mejores resultados, aunque tampoco hay una diferencia muy grande comparado con el modelo por defecto.

### 3.6. Horas pico y valle

Actualmente existen en el mercado libre tarifas que discriminan el precio de la electricidad consumida en función de la hora en que se haga ese consumo. Por tanto, es importante no solo realizar este consumo en los momentos más baratos, sino también elegir adecuadamente el rango de horas en el que se quiere que el precio por kWh sea más barato (de forma general las compañías dan a elegir un intervalo de 8 horas).

En esta sección se realiza dicho análisis, que aunque no tiene nada que ver con predecir el precio de la electricidad, sí puede tener repercusión en la factura que se paga.

El primer paso es conocer la distribución de consumos a lo largo de un día. Dado que no se tiene un ejemplo real de un domicilio, se decide usar el consumo que se ha usado a lo largo de todo el proyecto, es decir, el consumo a nivel nacional obtenido de REE.

Una vez escogido el consumo se multiplica por el precio horario, de forma que ya se tienen valores en euros. Posteriormente es necesario agrupar toda la serie temporal por horas, ya que lo que interesa es detectar patrones de consumo. Pero antes es conveniente diferenciar los días de lunes a viernes de los de fin de semana.

Esto es debido a lo que ya se comentó en la figura 2, donde el consumo de los días de fin de semana es menor que el resto de días. No obstante, además de verlo reflejado en los datos, basta con analizar los hábitos que cada uno tiene los sábados y domingos con respecto a los días de diario: desde levantarse más tarde hasta realizar viajes.

En la figura 3 y 4 se muestran de manera gráfica los resultados de este análisis, donde los valores han sido normalizados para facilitar su comprensión.

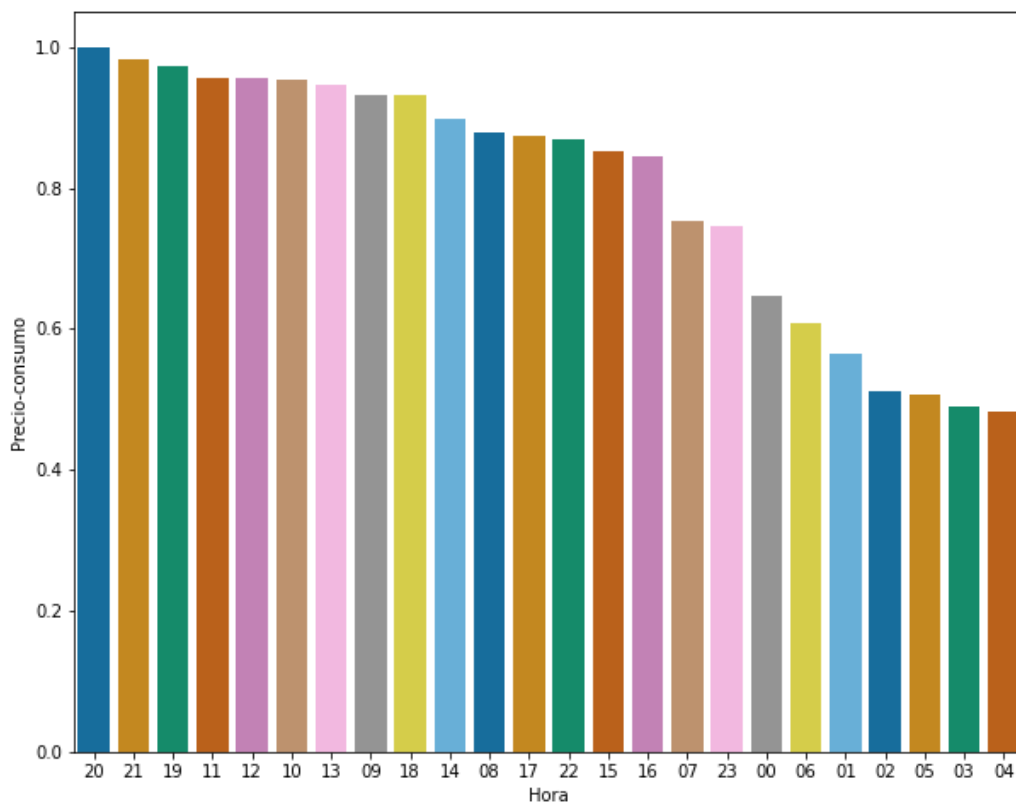


Figura 3: Relación entre el precio y el consumo en función de la hora para los días de diario

Se observa que de lunes a viernes el mayor gasto sucede en la tarde-noche (aproximadamente de 18 a 21) y en la mañana (de 9 a 13). Por contra, el menor gasto se produce en la madrugada.

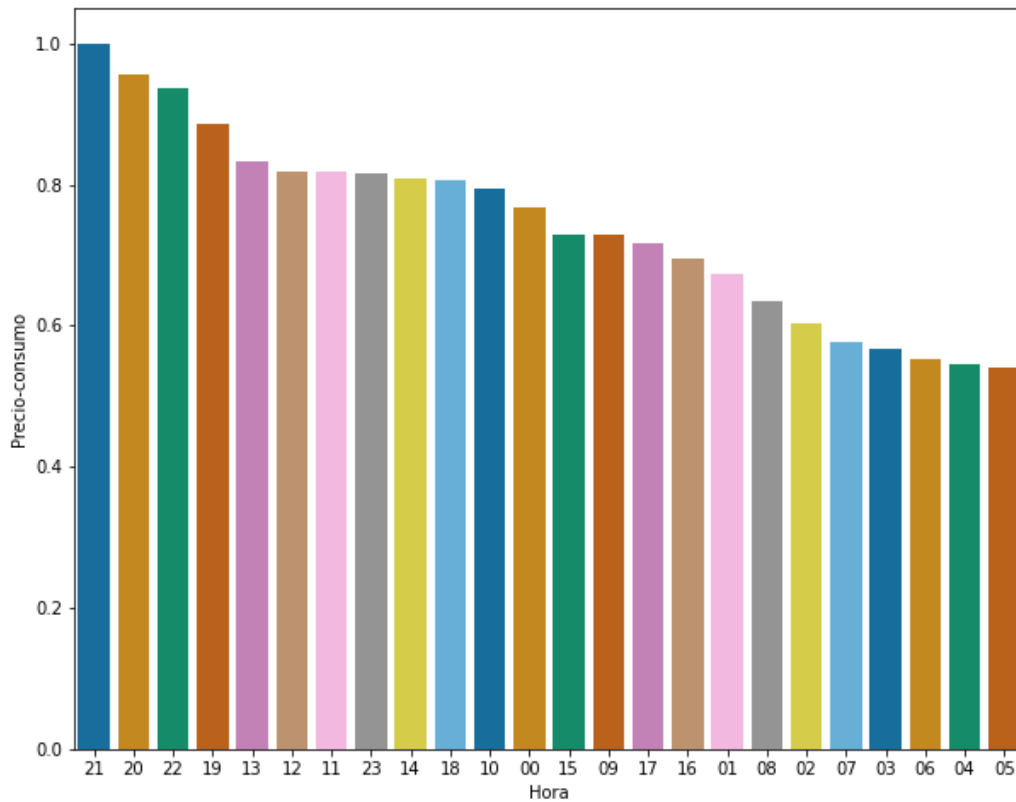


Figura 4: Relación entre el precio y el consumo en función de la hora para los días de fin de semana

Para el caso del fin de semana sucede más o menos lo mismo, si bien se detecta un cierto retraso temporal en las horas puntas, de manera que las horas de mayor gasto se concentran también en la tarde-noche, aunque un poco más tarde con respecto a los días de diario (de 19 a 22), y en el mediodía (de 11 a 14).

Con este análisis ya se podrían escoger las 8 horas (o el intervalo de tiempo que se nos permita) en las que se quiere pagar un menor precio por el consumo realizado.

## 4. Conclusiones

Tras todos los análisis y estudios efectuados a lo largo de este proyecto, se resumen a continuación las principales conclusiones obtenidas:

- **Importancia de las condiciones meteorológicas.** Las energías renovables producen un alto porcentaje de la energía eléctrica que se consume en España y es probable que en los próximos años su contribución siga aumentando, ya que hay planificados numerosos proyectos para instalar más MW de potencia tanto eólica como solar. Esto origina que cada vez vaya a ser más vital predecir correctamente las principales variables meteorológicas que afectan a la producción de este tipo de energías.



- **Modelo de regresión.** Aunque de todos los modelos que se han probado el *Random Forest* ha demostrado dar los mejores resultados (y con diferencia), es conveniente en el futuro realizar más pruebas, de forma que se estudien cambios en los parámetros de los modelos o incluso utilizar alguno nuevo.
- **Cambios en los modelos energéticos.** Desde el principio este proyecto se ha planteado como objetivo el predecir únicamente el concepto de coste de electricidad de la factura. La razón se debe a que algunos otros conceptos de la factura eléctrica están previstos que cambien a lo largo del verano de 2021, por lo que no hubiera tenido mucho sentido tenerlos en cuenta. Una vez que este cambio normativo se lleve a cabo sería conveniente tener en cuenta todos los conceptos para hacer una predicción exacta de, por ejemplo, una factura de un domicilio corriente.