

# **Tipología y Ciclo de Vida de los Datos**

**PRA 1 - Web Scraping**

Carlos Mas Estévez  
Claudia Reina Fajardo

## **ÍNDICE**

1.- Introducción

2.- Metodología

3.- Respuestas de la práctica

## 1.- INTRODUCCIÓN

El objetivo principal de esta práctica es el estudio del comportamiento del precio de la luz (kW/h) en España (Península y Canarias) desde el 1 de junio de 2021. Para ello, recurriremos a técnicas de Web Scraping en la siguiente web, la cual proporciona información del precio por horas de un determinado día: <https://tarifaluzhora.es/>

## 2.- METODOLOGÍA

Hemos usado la librería BeautifulSoup de Python a fin de hacer scraping a la web de interés. El código lo podemos encontrar en el siguiente repositorio de Git Hub:

<https://github.com/creinaf/TCVD-Practica-1-Web-scraping>

Dado que cada página de la web únicamente proporciona datos de un día concreto, decidimos obtener la información relativa del precio de la luz hasta el día 1 de junio de 2021 iterando el scraping en los distintos días.

## 3.- RESPUESTAS DE LA PRÁCTICA

### 3.1.- Contexto

Los datos de estudio han sido extraídos de la web <https://tarifaluzhora.es/> la cual proporciona información del precio de la electricidad en España desde la actualidad hasta el día 1 de junio de 2021. Dado que la página principal del sitio web solo proporciona información del día vigente, centraremos nuestro estudio en la evolución del precio desde el 31 de marzo de 2022 hasta el 1 de junio de 2021.

### 3.2.- Título

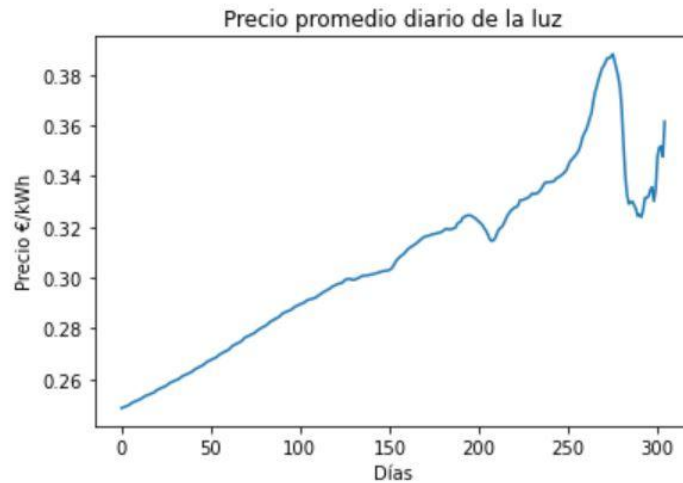
Precio de la electricidad en España.

### 3.3.- Descripción del dataset

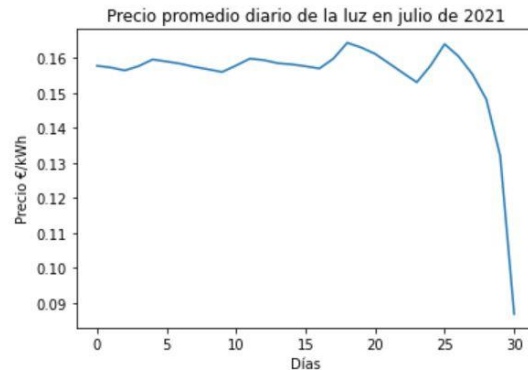
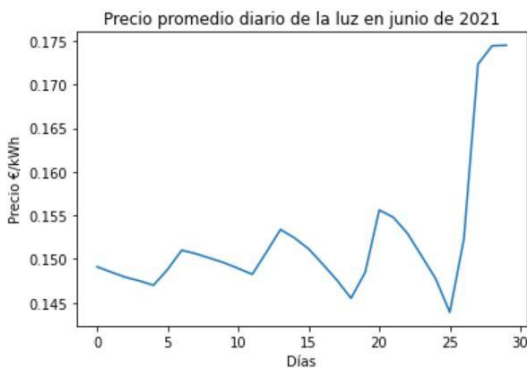
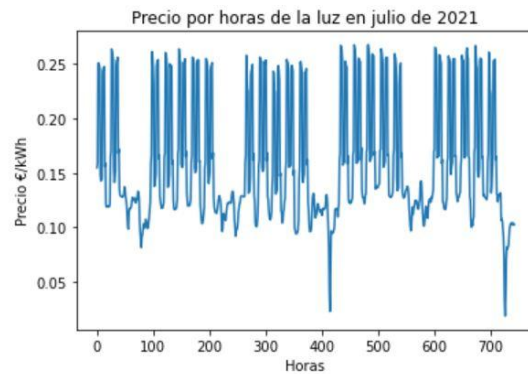
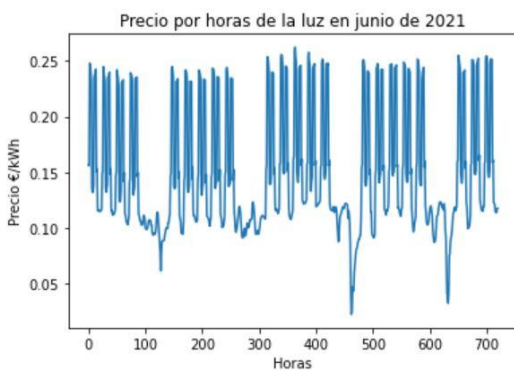
Precio de la electricidad en España (Península, Canarias y Baleares) durante el segundo semestre de 2021 y primer trimestre de 2022.

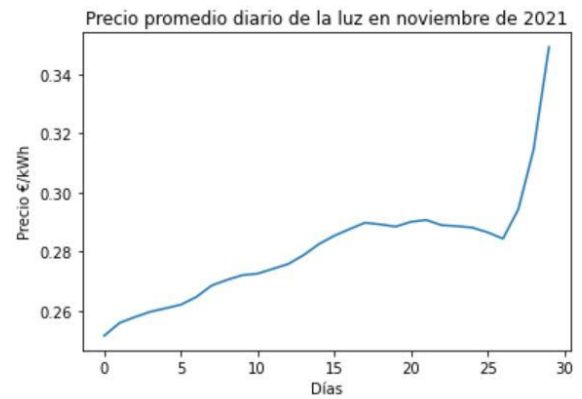
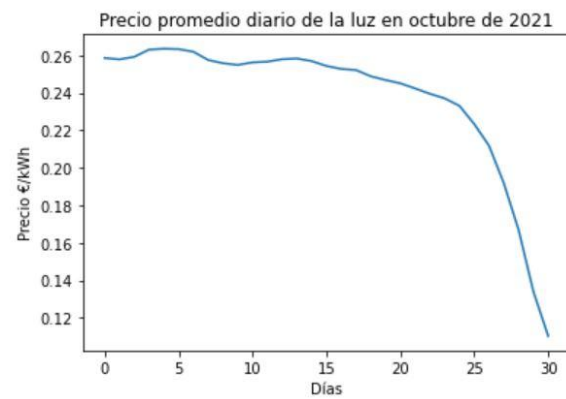
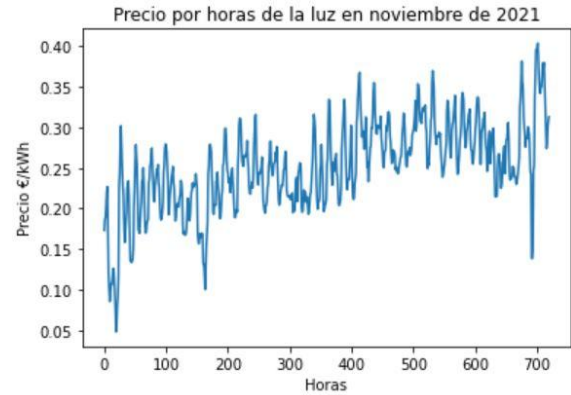
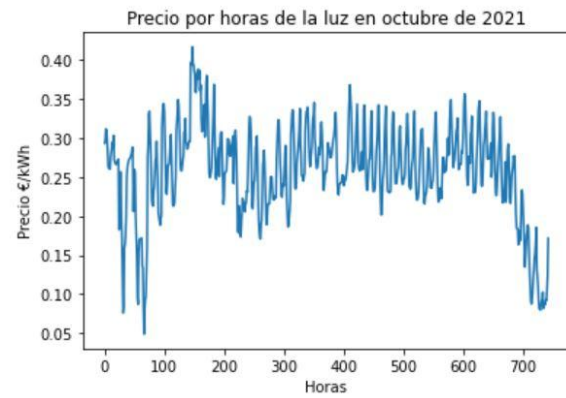
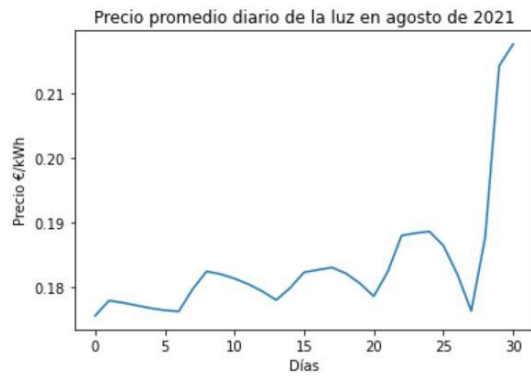
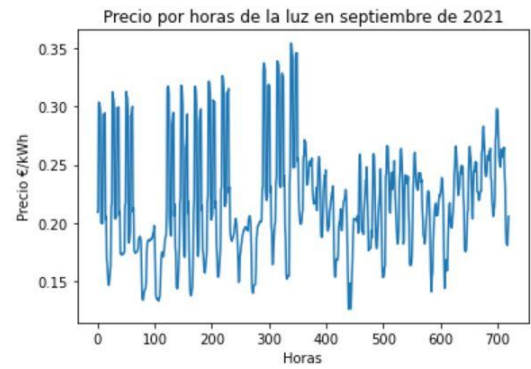
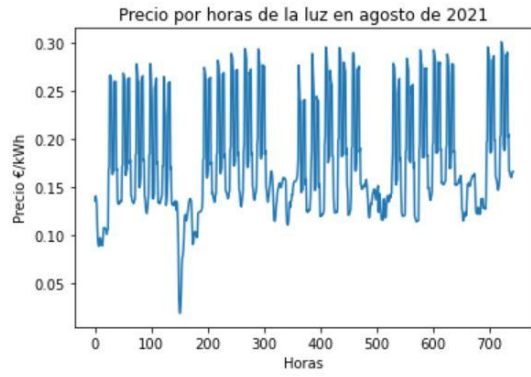
### 3.4.- Representación gráfica

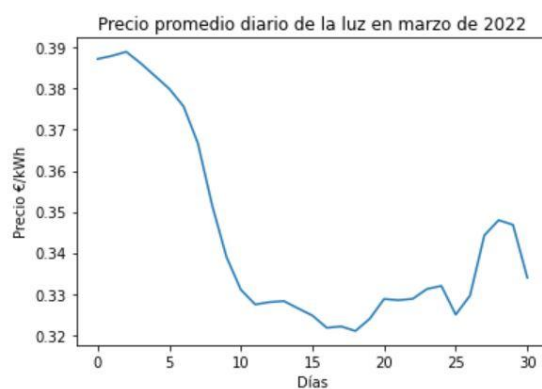
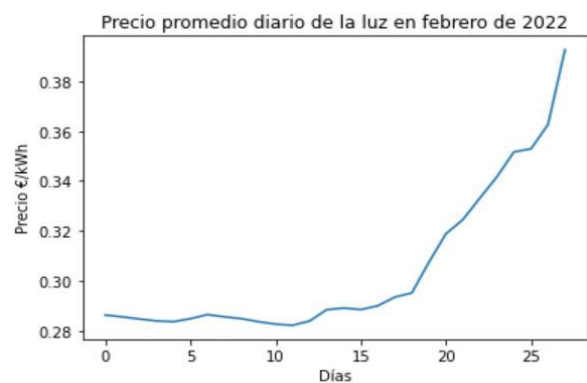
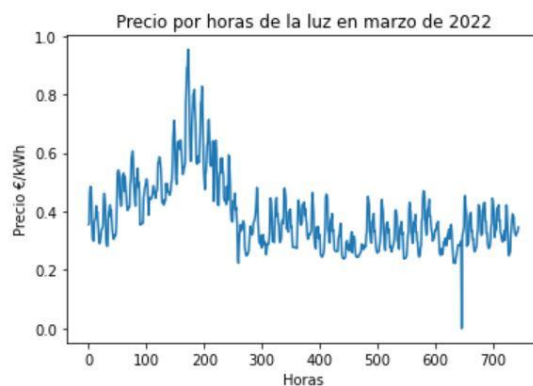
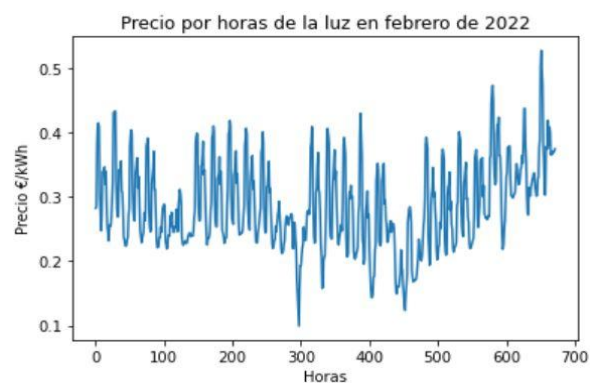
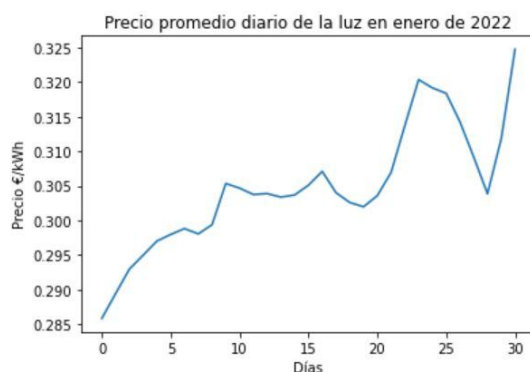
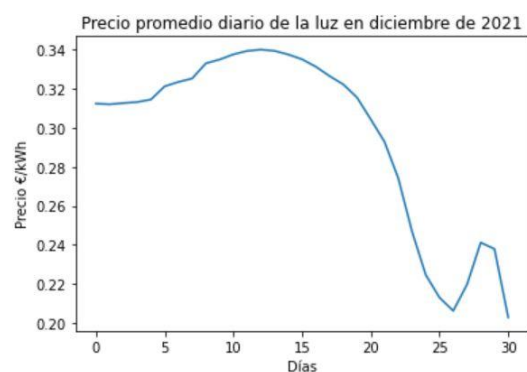
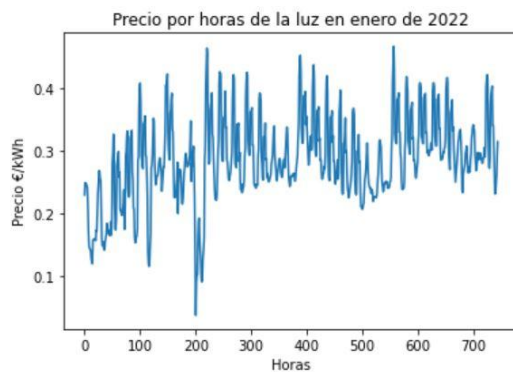
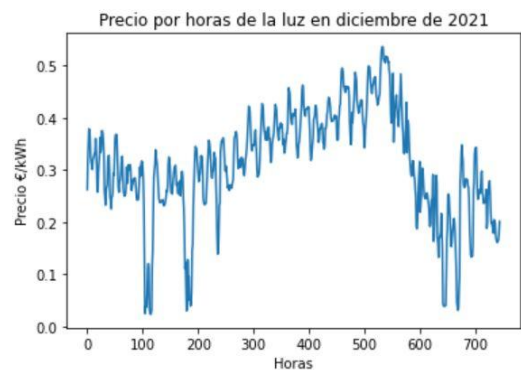
A continuación, mostramos el comportamiento del precio de la luz (promedio diario) a lo largo de 305 días (desde el 1 de junio de 2021 hasta el 31 de marzo de 2022). Se observa claramente que existe una tendencia alcista en el precio de la luz.



A continuación, mostramos el comportamiento mensual del precio de la luz, tanto por horas como por promedio diario.







### **3.5.- Contenido**

Cada columna del dataset nos habla de una franja horaria del precio de la luz, mientras que las filas nos indican los distintos días. Las unidades han sido tomadas en €/kWh.

### **3.6.- Agradecimientos**

Agradecimientos al propietario del conjunto de datos Gonzalo Lahera, co-director de Selectra, cuya empresa ha publicado la información relativa al precio de la electricidad y de donde hemos obtenido la información.

Múltiples empresas han analizado el comportamiento del precio de la luz, tal como hemos hecho en este proyecto, a fin de ofrecer servicios relativos a dichos análisis. Algunas de ellas serían OCU o EPDATA, tal como observamos en el siguientes enlaces, respectivamente:

<https://www.ocu.org/vivienda-y-energia/gas-luz/informe/precio-luz>

<https://www.epdata.es/datos/precio-factura-luz-datos-estadisticas/594>

A fin de actuar de acuerdo a los principios éticos y legales, nos aseguramos previamente que el sitio web de nuestro estudio no prohíbe ni bloquea los intentos de web scrapping, así como identificar los orígenes de los datos (precios de la luz) como datos de carácter público según la red eléctrica de España <https://www.ree.es/es>

### **3.7.- Inspiración**

Dado el escenario mundial actual, destacado por insuficiencia energética y crisis de recursos naturales, se hace cada vez más notorio un ajuste alcista continuo del precio de la luz. Es por ello que consideramos interesante analizar este comportamiento a fin de entender mejor el escenario económico en el que nos encontramos actualmente.

Tal como se comentó en el apartado anterior, este interés se ve reflejado en la realización de análisis similares (y más completos) por parte de empresas como OCU o EPDATA.

### **3.8.- Licencia**

La licencia que hemos escogido es Creative Commons Legal Code CC0 1.0 Universal, pues la publicación inicial de los datos está realizada por Red Eléctrica de España y son de carácter público. Por ello, la licencia que escogemos será con el mayor carácter público posible a fin de facilitar la obtención y comprensión de dichos datos.

### 3.9.- Código

El código con el que se ha generado el dataset lo hemos guardado en el repositorio de Git Hub siguiente: <https://github.com/creinaf/TCVD-Practica-1-Web-scraping>

El archivo con el código se llama [PRA1\\_tipologia.ipynb](#)

### 3.10.- Dataset

El DOI de Zenodo en el que se ha subido el dataset en formato .csv es 10.5281/zenodo.6445834.

Link: <https://zenodo.org/record/6445834>

### 3.11.- Vídeo

El enlace del vídeo es el siguiente:

<https://drive.google.com/file/d/1sT7WRjPmgoiHXg4BQt31yZIFbJH6JHpV/view?usp=sharing>

Contribuciones	Firma
Investigación previa	CME, CRF
Redacción de las respuestas	CME
Desarrollo de código	CME, CRF