

Ejercicio 2 – EDA (Exploratory Data Analysis) R

En este caso práctico, vamos a realizar un Análisis Exploratorio sobre los precios del mercado energético de Europa desde 2010. La fuente original de los datos puede encontrarse en este enlace <https://zenodo.org/record/5977857#.Y-pQT3bMKUk> aunque, tomaremos como referencia su publicación en Kaggle.

- <https://www.kaggle.com/datasets/thedevastator/hourly-european-power-market-prices>

Información de las variables

Este conjunto de datos contiene precios de energía de mercado diarios europeos por hora de varios sistemas y tipos de monedas. En general, el conjunto de los datos ofrece la siguiente información:

- **fecha:** Fecha en la que se genera el dato de transacción.
- **hora:** Hora en la que se registra la transacción.
- **sistema:** Mercado Europeo regulador de los precios de energía
- **bandera:** Si es horario invierno o verano
- **precio:** Precio que marca el sistema para el precio de la energía.
- **tipo_moneda:** Tipo de moneda, pudiendo ser euro o libra
- **origen_dato:** Web desde la que se ingesta el registro del precio.
- **fecha_actualizacion:** Última fecha de verificación del dato.

OBJETIVOS

Teniendo en cuenta el siguiente conjunto de datos, se pide hacer lo siguiente:

1. Carga el dataset como dataframe y, borra si hay alguna columna que no aporte información como IDs.
2. Obtén el rango de las fechas ¿Para cuántos años tenemos registros de los datos?
3. Sin importar el año, ¿Qué día del mes se reportan más registros? ¿Cuál es el que menos?
4. Si analizamos la variable precio, parece que hay precios del mercado energético que aparecen en negativo, toma todos aquellos valores que sean negativos y, transformalos en nulo ¿Qué cantidad hay de valores nulos?. Posteriormente realiza estas dos operaciones:
5. 4.1 Asigna un cero a todos los valores nulos menos a los que pertenezcan al mercado de Dinamarca (DK1 y DK2), puedes consultar más información sobre los nombres de los mercados aquí: https://www.jao.eu/sites/default/files/2022-03/List%20of%20Bidding%20Zone%20borders_2022_CORE.pdf
6. 4.2 Borra el resto de nulos.
7. ¿Cuántos datos había antes de borrar nulos y, cuantos quedan ahora?, es decir ¿Cuántos datos se han borrado?
8. La variable bandera presenta los siguientes valores 0 y 1, estos valores están referidos al horario de invierno 0 y horario de verano 1, crea una nueva variable categórica llamada ESTACION que tenga los valores invierno y verano. Investiga la función recode del paquete dplyr
9. La columna tipo de moneda tiene dos valores 1 que es €/MWh y 2 GBP/MWh, recodifica la propia variable a los valores categóricos Euros y Libras. Por lo tanto, tenemos precios en dos monedas diferentes, analiza el precio medio del mercado energético en función de la estación y el tipo de moneda.
10. Toma la columnas fecha_actualizacion y, toma únicamente la parte de la fecha sin contar la hora y, obtén la diferencia en días entre la fecha de actualización y la columna fecha. Tras obtener dicha diferencia muestra el promedio de días de actualización, cuanto más alto mejor, eso querrá decir que datos muy antiguos siguen teniendo registros de actualización: Pista, investiga la función difftime y as.Date
11. Filtra el dataset por el tipo de moneda Euros y el mercado español, visualiza la serie completa respecto a la fecha de registro del precio del mercado energético. Actualmente, vivimos en tiempos de bastante inestabilidad en el mercado energético ¿se ve reflejado en la visualización?
12. Toma el precio medio en euros por cada año del mercado energético de Francia y España. Visualiza la evolución del precio medio de ambos mercados
13. Sin importar la moneda, muestra un top 10 para los países que tengan el precio medio más alto en el mercado energético y, el top 10 con el valor más pequeño.