

## Respuestas.

1. **Contexto.** Explicar en que contexto se ha recolectado la información. Explique por que el sitio web elegido proporciona dicha información.

En la página <https://datosmacro.expansion.com/> se puede encontrar información sobre datos macro económicos, tales como tasas de paro, PIB, salarios medios, así como precios de materias primas.

Nuestro dataset es una serie temporal del precio del petróleo según los 3 indicadores más utilizados en los mercados financieros (OPEP, Brent, Western Texas Intermediate), información que se encuentra disponible en la sección de materias primas de la página.

2. **Definir un título para el dataset.** Elegir un título que sea descriptivo.

El nombre del dataset es oil\_prices. Al ser el dataset literalmente el precio del petróleo según varios índices, el nombre es una descripción literal del mismo.

3. **Descripción del dataset.** Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).

El dataset es una serie temporal del precio del petróleo según los 3 indicadores más utilizados en los mercados financieros (OPEP, Brent, Western Texas Intermediate),

4. **Representación gráfica.** Presentar una imagen o esquema que identifique el dataset visualmente.

Se presenta el dataset como una serie temporal, con la evolución de las 3 variables.



Se puede encontrar la imagen y el código en R para generarlo en el repositorio de GIT, en la carpeta visualization:  
[https://github.com/josumsc/web\\_scrapping\\_datosmacro/tree/main/visualization](https://github.com/josumsc/web_scrapping_datosmacro/tree/main/visualization).

5. **Contenido.** Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y como se ha recogido.

El dataset se compone de los siguientes campos:

- **date:** Es la fecha del dato.
- **opec:** Precio del petróleo al que vende la OPEP.
- **brent:** Precio del petróleo al que venden los países del norte de europa.
- **wti:** Precio del petróleo al que vende el estado de Texas y el sur de Oklahoma.

Las fechas van desde el 1 de enero del 2000 al 31 de octubre del 2020.

Los datos se han recogido utilizando Python y la librería BeautifulSoup 4. Se accede a 3 direcciones del sitio de datos macro, se accede a la tabla con los datos de cada una de estas, y se guardan en un objeto. Por último se pasa el objeto a un array numpy y se convierte a un fichero csv.

Se hace una petición por cada mes y atributo. Se escoge la fecha pasando el parámetro GET "dr=YYYY-mm" via URL.

Los atributos se recogen de las siguientes direcciones:

1. <https://datosmacro.expansion.com/materias-primas/opec> para OPEC.
2. <https://datosmacro.expansion.com/materias-primas/brent> para Brent.
3. <https://datosmacro.expansion.com/materias-primas/petroleo-wti> para WTI.

Se puede encontrar más información sobre el dataset y el cómo se extraen los datos en la wiki del repositorio ([https://github.com/josumsc/web\\_scrapping\\_datosmacro/wiki](https://github.com/josumsc/web_scrapping_datosmacro/wiki)).

6. **Agradecimientos.** Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).

El propietario del conjunto de datos (el sitio web del que los hemos extraído) es el periódico especializado en el mundo de la economía y las finanzas Expansión.

Se puede visitar el diario en la siguiente URL: <https://www.expansion.com/>.

Algunas investigaciones sobre este tema (evolución del precio del petróleo) son, por ejemplo los siguientes:

- [https://e-archivo.uc3m.es/bitstream/handle/10016/25387/PFC\\_MPaloma\\_Simancas\\_Cuevas.pdf](https://e-archivo.uc3m.es/bitstream/handle/10016/25387/PFC_MPaloma_Simancas_Cuevas.pdf)
- [https://www.researchgate.net/publication/336337487\\_Evolucion\\_de\\_los\\_precios\\_del\\_petroleo](https://www.researchgate.net/publication/336337487_Evolucion_de_los_precios_del_petroleo)

- [https://www.researchgate.net/publication/41599746\\_Fractal\\_Scaling\\_in\\_Crude\\_Oil\\_Price\\_Evolution\\_via\\_Time\\_Series\\_Analysis\\_of\\_Historical\\_Data](https://www.researchgate.net/publication/41599746_Fractal_Scaling_in_Crude_Oil_Price_Evolution_via_Time_Series_Analysis_of_Historical_Data)

Se pueden encontrar muchos más, que la evolución de los precios de las materias primas, ya sea para análisis o usos financieros están muy extendidos.

7. **Inspiración.** Explique por que es interesante este conjunto de datos y que preguntas se pretenden responder.

Este dataset, al igual que cualquier serie temporal de alguna materia prima o activo, puede ser utilizado para análisis de variación en algunos momentos en el tiempo, como crisis, etc, cómo varía el precio entre los distintos vendedores, y para entrenar redes neuronales y otros métodos supervisados.

8. **Licencia.** Seleccione una de estas licencias para su dataset y explique el motivo de su selección:
  - Released Under CC0: Public Domain License
  - Released Under CC BY-NC-SA 4.0 License
  - Released Under CC BY-SA 4.0 License
  - Database released under Open Database License, individual contents under Database Contents License
  - Other (specified above)
  - Unknown License

Nos hemos decidido por usar CC4NC.

Esto es así para que otros desarrolladores utilicen la dataset libremente siempre y cuando no tengan un uso comercial, ya que al proveer del trabajo de una entidad privada (Expansión) podría no ser del todo ético lucrarse con ello.

9. **Código.** Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

El código se puede encontrar en [https://github.com/josumsc/web\\_scrapping\\_datosmacro](https://github.com/josumsc/web_scrapping_datosmacro).

10. **Dataset.** Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

El dataset se puede encontrar en <https://zenodo.org/record/4245213#.X6mCmmhKjIU>.

El DOI del dataset es 10.5281/zenodo.4245213.