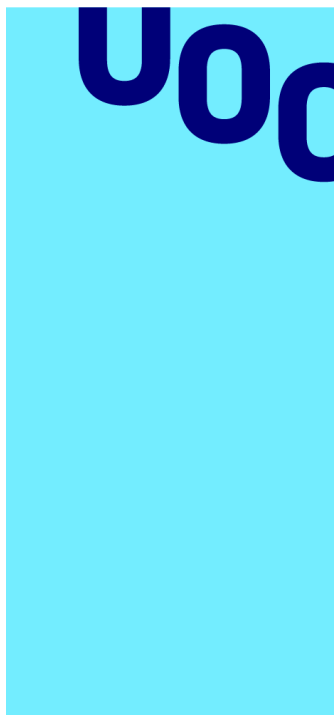


# Captura de datos mediante web scrapping

PR1 - Tipología y ciclo de vida de los  
datos



Universitat Oberta  
de Catalunya

## Índice

- Contexto
- Descripción del dataset
- Representación gráfica
- Contenido
- Propietario
- Inspiración
- Licencia
- Código
- Dataset
- Recursos
- Vídeo

# Valores actuales de las monedas digitales

## Contexto

Web scraping es una técnica utilizada muchas veces para monitorizar precios en el mercado digital. Esta técnica puede suponer una gran ventaja respecto a la competencia, ya que puedes obtener sus datos y usarlos en tu beneficio.

En este proyecto hemos trabajado sobre [CoinGecko](#), una web de monitorización de criptomonedas a tiempo real.

Nuestro objetivo es obtener los datos en una tabla con las variables que nos interesan para poder hacer análisis a posteriori, y con estos análisis poder extraer conclusiones a la hora de invertir.

## Descripción del dataset

Nuestro dataset consiste en una lista de las 100 mejores criptomonedas ordenadas según su capitalización de mercado, junto a algunos de los datos que hemos considerado más relevantes para su análisis.

En total contiene 100 observaciones con 6 variables, lo que da un volumen de 600 datos.

## Representación gráfica

El programa creado tiene dos partes:

- Web scraping de las 100 mejores monedas y extracción de sus estadísticas en tiempo real, un ejemplo sería la **Figura 1**.
- Repetir el paso anterior cada 10 minutos un total de diez veces y posteriormente, unir los diferentes resultados en un conjunto de datos mayor. Un ejemplo de resultado final sería la **Figura 2**.

Name	Time	Price	Min_price_2	Max_price_2	Market_cap	Vol_24h
bitcoin	2022-11-16 1	\$16,461.13	\$16,437.01	\$17,071.74	\$315,970,741	\$27,921,668,674
ethereum	2022-11-16 1	\$1,201.87	\$1,193.41	\$1,273.33	\$144,510,229	\$9,147,492,615
tether	2022-11-16 1	\$0.999731	\$0.997403	\$1.00	\$66,035,091	\$38,376,503,886
usd-coin	2022-11-16 1	\$1.00	\$0.993319	\$1.01	\$44,373,029	\$3,253,640,264
bnb	2022-11-16 1	\$271.06	\$269.94	\$280.86	\$44,220,584	\$906,839,926
binance-usd	2022-11-16 1	\$1.00	\$0.987682	\$1.01	\$23,198,402	\$8,650,534,366
xrp	2022-11-16 1	\$0.368373	\$0.366347	\$0.396024	\$18,489,324	\$1,501,172,747
dogecoin	2022-11-16 1	\$0.08538801	\$0.08460708	\$0.09118849	\$11,657,011	\$929,748,674
cardano	2022-11-16 1	\$0.328761	\$0.326728	\$0.344745	\$11,501,723	\$305,193,871
polygon	2022-11-16 1	\$0.895156	\$0.884798	\$0.969499	\$7,914,256	\$640,020,746
polkadot	2022-11-16 1	\$5.70	\$5.65	\$6.04	\$6,656,448	\$164,474,663

**Figura 1.** Ejemplo de extracción de estadísticas individuales.

Name	Time	Price	Min_price_2	Max_price_2	Market_cap	Vol_24h
1inch	2022-11-16 1	\$0.525937	\$0.518301	\$0.544801	\$327,257,574	\$20,044,785
1inch	2022-11-16 1	\$0.523920	\$0.518301	\$0.544801	\$326,317,040	\$20,189,030
1inch	2022-11-16 1	\$0.526769	\$0.518301	\$0.544801	\$327,257,574	\$20,295,031
1inch	2022-11-16 1	\$0.525956	\$0.518301	\$0.544801	\$327,484,368	\$20,026,273
1inch	2022-11-16 1	\$0.525451	\$0.518301	\$0.544801	\$326,460,280	\$20,059,599
1inch	2022-11-16 1	\$0.525352	\$0.518301	\$0.544801	\$326,317,040	\$20,239,291
1inch	2022-11-16 1	\$0.523191	\$0.518301	\$0.544801	\$324,692,300	\$20,185,321
1inch	2022-11-16 1	\$0.525704	\$0.518301	\$0.544801	\$325,960,301	\$20,381,462
1inch	2022-11-16 1	\$0.525445	\$0.518301	\$0.544801	\$326,460,280	\$20,107,460
1inch	2022-11-16 1	\$0.522858	\$0.518301	\$0.544801	\$325,220,406	\$20,180,498

**Figura 2.** Ejemplo de extracción de estadísticas total.

## Contenido

Tras estudiar las diferentes variables que nos ofrece la web, las más interesantes y elegidas para el dataset final son:

- Nombre, contiene la denominación de cada moneda.
- Fecha, contiene la fecha de recogida de los datos.
- Precio, contiene el precio de la moneda en el momento de la recogida,
- Precio mínimo (24 horas).
- Precio máximo (24 horas).
- Market cap, contiene la capitalización de mercado de la moneda.
- Volumen (24 horas), contiene el volumen de transacciones realizadas de la moneda.

## Propietario

Al fin y al cabo, no existe propietario en sí de los datos extraídos ya que CoinGecko simplemente es una web donde se visualizan las estadísticas de las principales monedas como muchas otras.

## Inspiración

No es una novedad el creciente interés por la compra-venta de monedas virtuales. Con el programa resultante y conocimientos de cómo fluctúa el mercado sería posible encontrar los momentos idóneos para realizar las transacciones y maximizar los beneficios.

## Licencia

Hemos elegido la licencia CC0: Public Domain License porque los datos los puede conseguir cualquier persona de forma sencilla. Si en un futuro, se hiciese una serie temporal más grande y que aportará realmente valor, se cambiaría a otro tipo de licencia más restrictiva.

## Código

Versiones:

- Python: 3.7.11
- Request version: 2.27.1
- BeautifulSoup version: 4.11.1
- csv version: 1.0
- Pandas version: 1.3.5
- Datetime, time y os son librerías propias de python, es decir, son dependientes de la versión de este.

Todo el código utilizado está subido en el repositorio de [GitHub](#).

## Dataset

El dataset también se encuentra en el siguiente repositorio de Zenodo [enlace](#).

Dentro de la carpeta data/ del repositorio de GitHub anteriormente mencionado se pueden encontrar tanto los csv intermedios como el final.

Dentro de la carpeta dataset/ se encuentran dos csv:

- CoinGecko.csv: conjunto de las 100 monedas con las correspondientes estadísticas a forma de ejemplo.
- Coins\_dataset.csv: el conjunto de las monedas construyendo una pequeña serie temporal.

## Recursos

- *Beautiful Soup Documentation — Beautiful Soup 4.4.0 documentation*,  
<https://beautiful-soup-4.readthedocs.io/en/latest/>. Accessed 15 November 2022.
- Paruchuri, Vik. “Tutorial: Web Scraping with Python Using Beautiful Soup.”  
*Dataquest*, 30 March 2021,  
<https://www.dataquest.io/blog/web-scraping-python-using-beautiful-soup>.  
Accessed 15 November 2022.
- “User Guide — pandas 1.5.1 documentation.” *Pandas*,  
[https://pandas.pydata.org/docs/user\\_guide/index.html#user-guide](https://pandas.pydata.org/docs/user_guide/index.html#user-guide). Accessed 15 November 2022.

## Vídeo

El vídeo se encuentra subido en el siguiente [enlace](#)

Contribuciones	Firma
Investigación previa	Ferran Valverde Parera y Alejandro González Barberá
Redacción de las respuestas	Ferran Valverde Parera y Alejandro González Barberá
Desarrollo del código	Ferran Valverde Parera y Alejandro González Barberá
Participación en el vídeo	Ferran Valverde Parera y Alejandro González Barberá