

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

Se pretende recolectar información de CoinMarketCap acerca de las cryptomonedas con mayor ganancia en las últimas 24h con la finalidad de descargarla cada 5 minutos y poder obtener una evolución desde que la cryptomoneda entra en el top de ganancias hasta que desaparece de la lista.

Se está realizando esta acción para obtener una estrategia ganadora en las inversiones de un fondo de inversión, con el objetivo de mejorar el rendimiento para sus clientes.

El sitio web se ha elegido, ya que es el principal portal de información sobre cryptomonedas y dispone de información de la mayoría de ellas. Además, al tener el indicador de mayores ganadores en % de las últimas 24h, se ha detectado una oportunidad ideal para evaluar los activos que aparecen en dicho indicador y su evolución.

2. Título. Definir un título que sea descriptivo para el dataset.

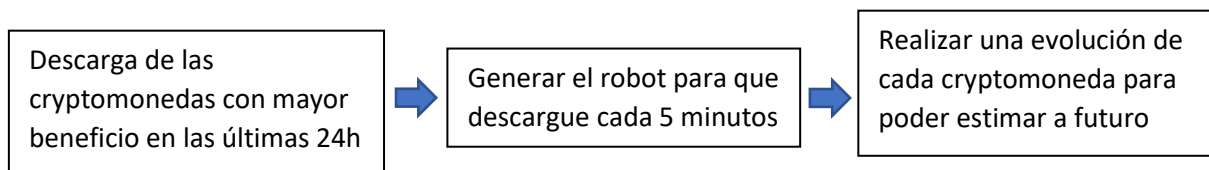
Evolución de las cryptomonedas con mayor ganancia en las últimas 24h.

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.

Como comentaba en el contexto, CoinMarketCap dispone de un indicador para las 30 cryptomonedas con mayor aumento respecto las últimas 24h. Lo que se extrae de dicho indicador es únicamente el momento de la extracción, los 30 tokens de la lista y sus 30 porcentajes de ganancia.

Con esta información, se pretende realizar el web scraping cada 5 minutos para acabar obteniendo un dataset en el que detectar con qué porcentaje de ganancia entra el token al listado y cuál es su evolución.

4. Representación gráfica. Dibujar un esquema o diagrama que identifique el dataset visualmente y el proyecto elegido.



5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se han recogido.

Los campos de los que dispone el dataset son Nombre de la Cryptomoneda y Porcentaje de ganancia en las últimas 24h.

Actualmente, los datos solo son de una descarga momentánea del momento de realizar la práctica. Lo que se pretende es hacer que el robot descargue la información cada 5 minutos durante varios días.

Se han recogido a través de la página <https://coinmarketcap.com/es/gainers-losers/> a través de las librerías requests y BeautifulSoup. Tras obtener toda la información de la

página, se ha filtrado por el apartado del HTML donde se encontraban los datos necesarios y se han trabajado para exportarlos a CSV.

6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis similares. Justificar qué pasos se han seguido para actuar de acuerdo a los principios éticos y legales en el contexto del proyecto.

Agradezco a la organización DNStination Inc. que permita acceder a los datos que se encargan de recoger sobre cada criptomoneda.

En <https://coinmarketcap.com/robots.txt> podemos observar que permiten acceso a los precios de las criptomonedas, a pesar de bloquear otros datos sobre las mismas:

```
User-agent: *
Allow: /

Disallow: */currencies/*/social/$
Disallow: */currencies/*/onchain-analysis/$
Disallow: */currencies/*/wallets/$
Disallow: */currencies/*/ratings/$
Disallow: */currencies/*/price-estimates/$

Disallow: */*/headlines/*$
```

Por lo que extraer los datos no supone una mala actuación ética ni legal para el proyecto.

7. Inspiración. Explicar por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.

Este conjunto de datos es interesante, ya que aporta lo que se cree que es un valor diferencial en el estudio de las inversiones en criptomonedas. Realizando las futuras observaciones y generando los informes correspondientes, se espera llegar a conclusiones que ayuden a saber qué decisión tomar sobre la inversión y los márgenes que tener cuando una criptomoneda ingresa en la lista de las 30 con más ganancias de las últimas 24h.

8. Licencia. Seleccionar una de estas licencias para el dataset resultante y justificar el motivo de su selección:

La licencia seleccionada para este dataset es: **Released Under CC0: Public Domain License**. Se ha seleccionado por ser un dataset que se obtiene públicamente en CoinMarketCap al que se puede aportar ideas propias y añadir los campos que se consideren buenos en la búsqueda de los mejores rendimientos.

Además, el valor real que obtendremos de este dataset es mediante las evaluaciones posteriores que no serán públicas y serán dominio exclusivo del fondo de inversión.

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

Se adjunta el código de Python bajo el título PRA1.ipynb.

10. Dataset. Publicar el dataset obtenido en formato CSV en Zenodo con una breve descripción. Obtener y adjuntar el enlace del DOI.

El enlace de Zenodo es <https://zenodo.org/record/6450020#.YISCWMhByUk>.