

PRAC1

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

Podemos ver que las grandes plataformas de Marketplace online de informática dominan las ventas minoristas de componentes para ordenadores además de periféricos varios gracias a la economía de escala que ha barrido con los pequeños comercios electrónicos, con el advenir de las criptomonedas y el aumento de los precios de estas en los últimos años ha empujado a la subida de los precios de las tarjetas gráficas, los cuales son usadas para poder “minar”, es decir hacer cálculos matemáticos para poder obtener estas criptomonedas.

Una de estas Marketplace es PC Componentes, y queremos ver los precios de las diferentes tarjetas gráficas para poder hacer una comparación de los precios y las características de cada tarjeta, normalmente el precio que esta marcado es el que vende la propia empresa pero los usuarios también pueden elegir a otros proveedores para hacer su compra.

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

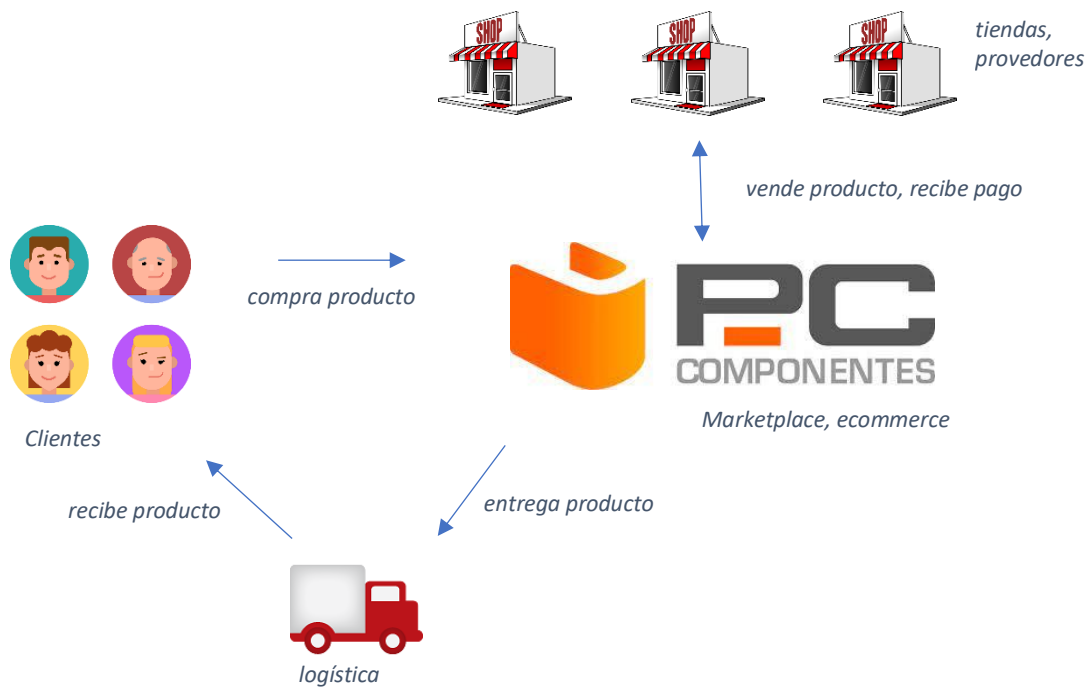
Precios y características de tarjetas gráficas en PC Componentes

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.

Es un dataset de tipo CSV, procesado por Python en con la librería BeautifulSoup, donde podemos ver el nombre de los chips y algunas características como el precio, la capacidad de la RAM y el tipo de RAM, y con un enlace para poder entrar y consultar en la página con más detalles.

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

La representación podría ser similar a cualquier otro *ecommerce* de tipo *markerplace*. En el caso de Amazon es la propia compañía que se encarga de la gestión de la logística en su gran mayoría.



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

Este dataset tiene 5 características:

- name: el nombre del producto
- gb: los GB de la memoria VRAM
- gddr: (Graphics Double Data Rate) es el estándar de la memoria RAM, donde cada vez consumen menos energía y son más eficaces.
- price: el precio en euros
- link: el link del producto

Es un dataset simple donde se extrae los datos de los precios y las capacidades de la tarjeta gráfica así un link donde se podrá consultar con más detalles.

Se hará scraping en la página de: <https://www.pccomponentes.com/tarjetas-graficas>

Pero por lo que se ve tiene medidas que actualmente con los conocimientos adquiridos no puedo solventar lo que hice fue descargar el archivo html y trabajarlo desde el propio ordenador.

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.

Hemos hecho una consulta en el archivo robots.txt de la página y en términos generales se permite el scraping de las páginas principales ya que no entra en los filtros de disallow. Hay que tener en cuenta que tiene: "User-agent: Python-urllib Disallow: /" así que usaremos la librería requests pero por lo comentado anteriormente usaremos un archivo htm plano.

Este archivo es una recomendación pero nunca una obligación.

Páginas de inspiración similares de scraping de precios:

- <https://www.youtube.com/watch?v=w607VRIHtT8>
- <https://www.blog.datahut.co/post/scraping-ebay>
- <https://www.geeksforgeeks.org/scraping-amazon-product-information-using-beautiful-soup/>
- https://www.youtube.com/watch?v=UCIHOT_7hok&

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.

En la extracción de datos sobre precios podemos hacer comparaciones y ver cuál es la oferta de los diferentes productos en las diferentes Marketplace o plataforma de Ecommerce para poder hacer una idea de los precios a que los vende y así poder sacar análisis de la evolución de los precios para poder extraer información útil para tomar decisiones como:

- Análisis de monopolios u oligopolios y fijación de precios por parte de las grandes Marketplace.
- Estudio del mercado para fijar un precio competitivo en la propia tienda que vende el producto.
- Fijar los descuentos necesarios en fechas especiales según las grandes plataformas.
- Saber en que fechas hay mejores descuentos para poder aprovechar esas ofertas.

Para poder hacer un mejor análisis de estos precios a través del tiempo podemos cambiar la URL y entrar en archive.org para poder ver los precios a través de las capturas de las páginas guardadas en X fecha.

8. **Licencia.** Seleccionar una de estas licencias para el dataset resultante y justificar 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.

Se elige la Database released under Open Database License, individual contents under Database Contents License. Se elige esta licencia porque es un proyecto académico con unos datos relativamente fáciles de conseguir y que no se podría monetizar. Hay unos puntos clave en esta licencia:

- Se permita a los usuarios compartir, modificar y utilizar libremente esta base de datos mientras siga manteniendo esta misma libertad para los demás.

- No requiere que se avisé al licenciante para poder utilizar la obra.

- La licencia no se aplica a los programas informáticos utilizados en la creación o funcionamiento de la Base de Datos.

- Mi persona (el licenciante) renuncia al derecho de cobrar regalías.

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

Está en Python en el repositorio.

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

Enlace DOI del dataset:

[10.5281/zenodo.6421371](https://doi.org/10.5281/zenodo.6421371)

11. **Vídeo.** Se debe hacer entrega de un vídeo explicativo de la práctica en donde cada uno de los integrantes del grupo explique con sus propias palabras tanto las respuestas del proyecto como el código utilizado para llevar a cabo la extracción. El vídeo debe ser enviado a través de un enlace a Google Drive que deben proporcionar, junto con el enlace al repositorio Git, al momento de entregar la práctica.

<https://drive.google.com/file/d/1TH2xv1FYwoVTIVZydOdalwKpN8NNrXYr/view?usp=sharing>

Por motivos a los horarios de trabajo he desarrollado este proyecto solo.

Contribuciones	Firma
Investigación previa	David
Redacción de las respuestas	David
Desarrollo del código	David