

Dataset: Especificacions de portàtils

Autors: Guillermo Camps Pons (individual)

1. Descripció

Aquest *dataset* conté dades sobre les diferents ordinadors portàtils que apareixen a la pàgina web <https://www.pccomponentes.com/portatiles> i que s'han obtingut a través d'un procés de *Web Scraping*. Els diferents atributs d'aquest *dataset* busquen descriure i quantificar les especificacions d'aquests portàtils alhora que extrauen altres valors com el preu, el nombre d'opinions, la valoració o els codis que identifiquen aquest producte.

2. Context

El meu objectiu principal era analitzar una pàgina web de ventes online, que preferiblement que no tingués una API. Amb la situació actual del COVID-19, vaig pensar que seria interessant observar quins canvis es donen en la venta de portàtils degut al confinament i al treball des de casa. Malauradament, donat que ja portem un temps en quarantena, aquests canvis no serien apreciables ara. En aquest context, s'estudien els diferents portàtils de PcComponentes, donat que és una web relativament popular que es dedica a la venta de productes informàtics i, llavors, pot aportar informació sobre aquests possibles canvis.

The screenshot displays a webpage from PcComponentes. On the left, there is a grid of six laptop listings. Each listing includes a laptop image, its model name, processor, RAM, and storage, a current price, a crossed-out original price, a discount percentage, a star rating with the number of reviews, and a 'Comparar' button. The laptops shown are: HP Pavilion Gaming 15-EC0007NS, Toshiba Satellite Pro R40-D-111, HP 155-FQ1040NS, MSI GF63 Thin 9SC-047XES, Asus VivoBook D509DA-EJ098, and HP Pavilion Gaming 15-DK0020NS. On the right side of the screenshot, the 'Especificaciones HP Pavilion Gaming:' page is visible, listing detailed technical specifications for the HP 155-FQ1040NS model, including processor, memory, storage, display, graphics, camera, connectivity, microphone, battery, and ports.

Especificaciones HP Pavilion Gaming:

- **Procesador** AMD Ryzen 5 3550H con gráficos Radeon™ Vega 8 (reloj base 2,1 GHz, incremento máximo hasta 3,7 GHz, 6 MB de caché, 4 núcleos)
- **Memoria RAM** SDRAM de 8 GB DDR4-2400 (2 x 4 GB)
- **Almacenamiento** SSD de 256 GB PCIe® NVMe M.2
- **Unidad óptica** No
- **Display** FHD SVA de 15,6" (39,6 cm) en diagonal antirreflectante de microborde con retroiluminación WLED (1920 x 1080)
- **Controlador gráfico** NVIDIA® GeForce® GTX 1050 (GDDR5 de 3 GB dedicada)
- **Cámara de portátil** Cámara HD HP TrueVision con micrófono digital de matriz doble integrado
- **Conectividad**
 - LAN 10/100/1000 GbE integrada
 - Combo Realtek Wi-Fi 5 (2 x 2) y Bluetooth® 5
- **Micrófono** Sí
- **Sonido** B&O; Altavoces dobles; HP Audio Boost 1.0
- **Batería** 3 celdas, ion-litio con 52,5 Wh. Es compatible con la carga rápida de la batería: aproximadamente el 50 % en 45 minutos
- **Conexiones**
 - 1 USB 3.1 Gen 1 Type-C (solo transferencia de datos, velocidad de señal de 5 Gb/s);
 - 1 USB 3.1 Gen 1 de tipo A (solo transferencia de datos);
 - 1 USB 2.0 de tipo A (HP Sleep and Charge);
 - 1 Smart Pin CA;
 - 1 HDMI 2.0;
 - 1 combo de auriculares/micrófono;
 - 1 RJ-45
 - Lector de tarjetas SD multiformato HP
- **Teclado** en blanco fantasma de tamaño normal tipo isla con teclado numérico
- **Sistema operativo** Sin Sistema Operativo
- **Dimensiones** 36 x 25,7 x 2,35 cm
- **Peso** 2,25 kg
- **Color** Negro

Importante: Este modelo no incluye sistema operativo, se adquiere por separado si se desea.

Fig. 1 Exemple de la pàgina web d'on s'extreuen les dades. A l'esquerra, la pàgina web que mostra els diferents portàtils i, a la dreta, les especificacions d'un d'aquests.

3. Contingut

Cada fila del *dataset* correspon a un portàtil diferent. Cada columna conté els següents camps:

Nom del camp	Tipus de variable	Descripció
<i>name</i>	caràcter	nom present al web
<i>brand</i>	caràcter	marca
<i>p/n</i>	caràcter	identificador P/N
<i>code</i>	enter	codi de l'article al web
<i>p_current</i>	real	preu actual
<i>p_base</i>	real	preu base sense descompte
<i>discount</i>	real	descompte actual
<i>tax</i>	real	taxa aplicada
<i>rating</i>	real	valoració
<i>opinions</i>	enter	número d'opinions
<i>cpu</i>	caràcter	descripció del processador
<i>cpu_ghz_min</i>	real	freqüència mínima del processador en GHz
<i>cpu_ghz_max</i>	real	freqüència màxima del processador en GHz
<i>ram</i>	caràcter	descripció de la memòria
<i>ram_mhz</i>	enter	freqüència de la memòria en MHz
<i>ram_gb</i>	enter	capacitat de la memòria en GB
<i>gpu</i>	caràcter	descripció de la targeta gràfica
<i>disk1</i>	caràcter	descripció del dispositiu d'emmagatzemament principal
<i>disk1_gb</i>	enter	capacitat d'emmagatzematge el disc principal
<i>disk1_type</i>	caràcter	tipus del dispositiu d'emmagatzematge principal
<i>disk2</i>	caràcter	descripció del dispositiu d'emmagatzemament secundari
<i>disk2_gb</i>	enter	capacitat d'emmagatzematge el disc secundari
<i>disk2_type</i>	caràcter	tipus del dispositiu d'emmagatzematge secundari
<i>display</i>	caràcter	descripció de la pantalla
<i>display_inch</i>	real	polzades de la pantalla
<i>display_width</i>	enter	amplada de la pantalla
<i>display_height</i>	enter	altura de la pantalla
<i>date</i>	caràcter YYYY-MM-DD	data de recol·lecta del registre
<i>url</i>	caràcter	enllaç al lloc web del portàtil

4. Agraïments

Donat que les dades han estat agafades directament de la seva pàgina web, el major dels agraïments va per l'empresa PcComponentes.

5. Inspiració

El *dataset* generat pot ser utilitzar per estudiar les relacions qualitat/preu dels productes i, tal vegada, usat per altres empreses per a treure'n un avantatge competitiu. Addicionalment, podríem estudiar les tendències dels compradors. Algunes preguntes que podríem fer-nos són: quines marques prefereixen?, quin és l'ordinador més estàndard?, etc.

6. Llicència

Donat que les dades han estat obtingudes d'una pàgina web a la qual es pot accedir sense problemes, que les dades s'han extret amb intencions no comercials i que, degut al petit rang històric, aquestes dades no presenten molta capacitat comercial, he decidit utilitzar la llicència de Creative Commons Public Domain Dedication (CC0). Aquesta llicència permet que el públic pugui fer-ne ús del *dataset* sense necessitat de demanar-me permís.

7. Codi i *dataset*

Tant el codi (`laptop_scraper.py`) com el *dataset* (`laptop_data.csv`) es poden trobar al següent enllaç de GitHub:

<https://github.com/guille97/laptopScraper>

El *dataset* també es pot trobar a Zenodo a través del següent DOI:

DOI [10.5281/zenodo.3736764](https://doi.org/10.5281/zenodo.3736764)