



Universitat
Oberta
de Catalunya

Máster universitario de Ciencia de Datos
Tipología y ciclo de vida de los datos

PRA 1

Víctor Elías Afonso Rodríguez
Saúl Santomé Rúa

Índice general

1. UmbrellaPriceComparator	1
1.1. Contexto	1
1.2. Dataset: Umbrella-prices	2
1.2.1. Atributos	2
1.2.2. Análisis	2
1.3. Agradecimientos	4
1.4. Inspiración	4
1.5. Licencia: CC BY-SA 4.0	4
1.6. Zenodo	5
1.7. Tabla de contribuciones	6

1. UmbrellaPriceComparator

1.1. Contexto

En los tiempos que corren existe tanta oferta de productos online que se vuelve prácticamente imposible hacer una búsqueda de los mejores precios de forma manual.

De este problema nace este web scraper, capaz de comparar los precios de todas las ofertas de **Amazon** y **Ebay** para un producto previamente especificado, en este caso: el paraguas.

Hemos querido escoger las 3 principales plataformas de venta online de referencia a nivel mundial: Aliexpress, Amazon y Ebay. Hay que tener en cuenta que representan los 3 grandes mercados que nos podemos encontrar en el vasto internet hoy en día, el mercado: Chino, Europeo y Norte Americano. La plataforma de venta online Aliexpress finalmente no la hemos incluido porque no permite realizar scraping.

Se eligen los paraguas como artículo para ejemplificar el funcionamiento del web scraper ya que se trata de un artículo básico, barato y del cual podemos encontrar multitud de ofertas en cada una de las plataformas mencionadas anteriormente.

1.2. Dataset: Umbrella-prices

En este caso, el dataset esta conformado por dos archivos .csv, los cuales hacen referencia a los datos extraídos de Amazon y Ebay respectivamente. Ambos archivos recogen datos de los 58 primeros resultados obtenidos en cada uno de estos portales de compra online (estos datos se corresponden con el instante de ejecución del web scraper).

1.2.1. Atributos

- **Producto**(*String*): El nombre con el que aparece el articulo en la web
- **Precio**(*Float*): El precio de cada articulo en el momento de la ejecución del script
- **Link**(*String*): Enlace al producto en la web de compra
- **Imagen**(*String*): Ruta de la primera imagen del producto dentro de la web de compra

1.2.2. Análisis

Amazon

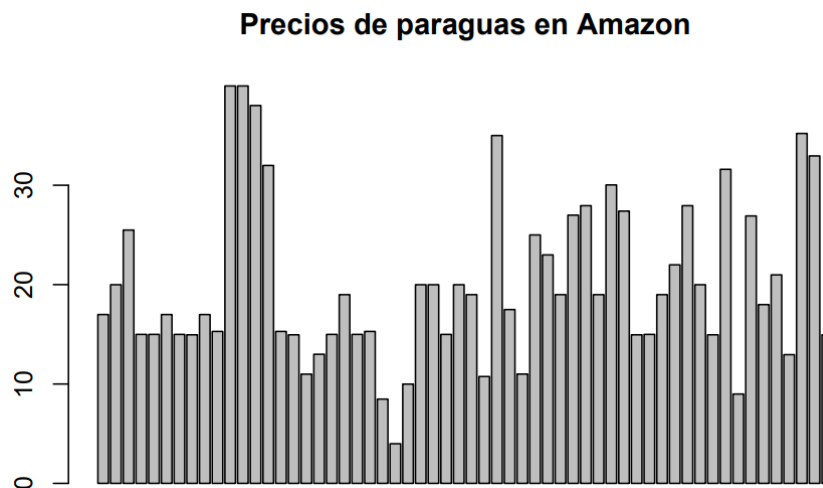


Figura 1.1: Amazon set

- **Mínimo:** 3.99 €
- **Máximo:** 39.99 €
- **Media:** 20.09 €

- **Mediana:** 18.49 €

Ebay

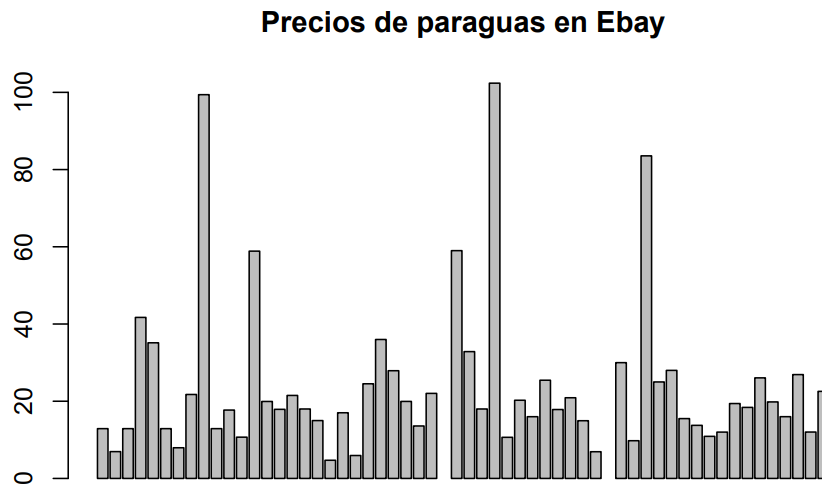


Figura 1.2: Ebay set

- **Mínimo:** 4.71 €
- **Máximo:** 102.38 €
- **Media:** 24.12 €
- **Mediana:** 18.20 €

Tras analizar los resultados obtenidos(a día 9 de abril de 2021) podemos llegar las siguientes conclusiones:

En ambas plataformas podemos encontrar una gran similitud en los precios, como vimos en los datos anteriores, las medianas son prácticamente iguales, y evitando los valores "pico", la media de los precios en Ebay sería incluso más baja.

Centrandonos ahora en los valores extremos, notamos una gran discrepancia en el umbral máximo del precio, encontrando en Ebay precios que prácticamente triplican el valor más alto encontrado en Amazon. Esto puede deberse a que en Ebay existe mucha más venta de productos *exclusivos* y por parte de particulares que en Amazon, con un stock mucho más reducido(una menor oferta hace que aumente el precio). Mientras que Amazon está más orientado a la venta de grandes cantidades de productos, evitando esa *exclusividad* que puede provocar picos en los precios.

1.3. Agradecimientos

En el aprendizaje previo que hay detrás de la creación de este web scraper, hay una gran influencia de un proyecto de Github que tiene como objetivo este mismo propósito, introducir y enseñar el web scraping a través del lenguaje R. Dicho proyecto, accesible a través de este enlace <https://github.com/rodrigodelrosso/CECE-Nuevo-Espacio-Web-Scraping-R> es el resumen de una conferencia donde se trata dicho tema.

Por este motivo queremos dar créditos y agradecimientos tanto al Dr. Martín Masci (martinmasci@economias.uba.ar) y al Mg. Rodrigo Del Rosso (rdelrosso@economias.uba.ar), como al Lic. Diego Parrás.

1.4. Inspiración

Los códigos formulados permiten recopilar en un instante los conceptos y precios de dos de las principales plataformas de venta online en España. Esto representa un gran avance, acercando la venta online cada día esta más a los que los economistas denominan **competencia perfecta**. La competencia perfecta se produce cuando en los mercados se encuentran productos estandarizados y hay plena transparencia sobre los precios. Justamente con los códigos empleados estamos más cerca de llegar a este estado de perfección en la competencia, ya que podemos recopilar los conceptos y precios de los diferentes productos. No solo conseguimos que sea útil para el consumidor, ya que permite ver donde se encuentran los mejores precios y productos, sino que, además, da la posibilidad de que cualquier empresa que venda algún tipo de producto, pueda conocer con facilidad como se encuentran el estándar de precios del mercado y así marcar ciertas estrategias que le ayuden con sus objetivos.

1.5. Licencia: CC BY-SA 4.0

La licencia por la que optamos es licencia CC by SA 4.0. Esta decisión se sustenta en 3 pilares:

- **Uso comercial:** Como está orientada hacia el comercio, creemos que sería recomendable que se hiciera un uso comercial del mismo, ya que así podría mejorarse y adaptarse a un mayor número de personas.
- **Emplear bajo misma licencia:** Tanto las remezclas y las transformaciones como las creaciones que partan de este material deberán difundirse bajo la misma licencia.
- **Autoría:** Al hacer uso de este material habrá que indicar la autoría, mostrar el enlace a la licencia e indicar los cambios efectuados.

1.6. Zenodo

El DOI del dataset publicado es 10.5281/zenodo.4676464, y el enlace de acceso <https://zenodo.org/record/4676464#.YHCNAugzYkk>

1.7. Tabla de contribuciones

Contribuciones	Firma
Investigación previa	VEAR, SSR
Redacción de las respuestas	VEAR, SSR
Desarrollo código	VEAR, SSR