

# Tipología y ciclo de vida de los datos

PRA1: ¿Cómo podemos capturar los datos de la web?

UOC

- **Andrés Merino Toapanta**

- **Mario Cueva Almedia**

22 de noviembre de 2022

# PRA1: ¿Cómo podemos capturar los datos de la web?

## 1 Contexto

Xiaomi es una empresa que desarrolla equipos electrónicos y es principalmente conocida por sus teléfonos. En sus 12 años de existencia, ha ido creciendo con presencia en más de 30 países. El año anterior, abrió su primera tienda oficial en Ecuador, sin embargo, los productos de esta marca ya se podían conseguir en este país antes de su presencia oficial. Los principales medios de venta no oficiales son las plataformas de comercio electrónico, entre ellas, la que más presencia tiene en Latinoamérica (y por ende, en Ecuador) es Mercado Libre.

La presencia de una tienda oficial, marca un referente con respecto al precio de los productos, pues el precio oficial puede marcar cambios en la dinámica de precios de los vendedores no oficiales.

En este contexto, se plantea la extracción de los datos de precios oficiales de dispositivos de la marca Xiaomi y la extracción de las ofertas (junto con las características de cada teléfono) presentes en Mercado Libre. Las páginas en las cuales se realizó esta extracción fueron:

- <https://mistore.com.ec/todos-los-productos>
- <https://celulares.mercadolibre.com.ec/celulares-smartphones>

## 2 Título

Oferta de celulares Xiaomi en Ecuador: precios oficiales vs. no oficiales

## 3 Descripción del dataset

El conjunto de datos generado reúne características de diversos dispositivos (teléfonos) de la marca Xiaomi, junto con los precios, tanto oficiales como no oficiales. Para el precio no oficial, se tomó las publicaciones dadas en la página Mercado Libre, en la misma que usuarios particulares ofertan los mismos dispositivos. Para una constatación de si son los mismos dispositivos, se ha tomado

también las características del equipo, como memoria RAM, memoria interna, etc., que suele cambiar dentro del mismo modelo, pero afecta a su precio.

De esta forma, el conjunto de datos presenta las características del teléfono, publicado en un medio no oficial, y su precio oficial.

## 4 Representación gráfica



## 5 Contenido

El conjunto de datos está conformado por 462 observaciones, cada una con 17 variables. Las variables son:

- Nombre
- Precio
- Memoria interna
- Memoria RAM
- Resolución de la cámara trasera principal
- Resolución de la cámara frontal principal
- Tamaño de la pantalla
- Resolución de la pantalla
- Capacidad de la batería
- Peso

- Dimensiones
- Velocidad del GPU
- Velocidad del procesador
- Cantidad de núcleos del procesador
- Modelo
- Marca
- Precio oficial

Dado que los datos se basan en las ofertas presentadas en la página de Mercado Libre y esta tiene una actualización dinámica (cada vez que un usuario realiza una publicación), la temporalidad de los datos es únicamente una muestra tomada en el día y la hora que se ejecutó el código.

## 6 Propietario

Mercado libre es una empresa de comercio electrónico de origen argentina que opera en nueve países de Latinoamérica. Según Octoparce, para extraer información de esta empresa, se puede realizar de dos formas. La primera consiste en utilizar la plantilla de mercado libre y la segunda consiste en utilizar el método raspado. Para cumplir con el proyecto, y en virtud del dinamismo, elegimos la segunda opción.

Antes de empezar con la extracción de la información, visitamos la sección de términos y condiciones, esto con la finalidad de no incurrir en problemas éticos o legales. En este sentido, pudimos advertir que Mercado libre es la dueña de la data, así mismo, también pudimos ver que se prohíbe la extracción de información con fines ilegales. En vista de que el presente proyecto tiene fines exclusivamente educativos, hemos considerado que no estamos incurriendo en los principios éticos o legales del proyecto.

## 7 Inspiración

Ecuador es un país de Sudamérica que se dolarizó aproximadamente hace 20 años. A raíz de este evento, la economía del país ha cambiado completamente la dinámica del mercado tanto local como internacional, esto se debe a que el dólar, frente a las monedas de los países vecinos, siempre tiene una mejor cotización. Este detalle ha sido advertido por los ecuatorianos, quienes a lo largo de muchos

años realizan viajes a las fronteras de los países vecinos, Colombia y Perú, con la finalidad de adquirir productos a más bajo precio.

En este contexto, los celulares son uno de los productos con mayor demanda. A esto se suma la creciente ola de empresas con fines de lavado de dinero. Es así que en el mercado ecuatoriano, se puede encontrar teléfonos celulares con distintos precios a pesar de tener aparentemente las mismas características.

Esto nos ha llevado a preguntarnos ¿hay alguna manera de detectar si los aparatos que encontramos en el mercado ecuatoriano son de empresas idóneas (legales y que paguen los impuestos respectivos)?

Una de las respuestas al parecer es tomar en cuenta como primer parámetro, el precio oficial de las empresas que fabrican celulares, para nuestro proyecto, hemos elegido a la empresa Xiaomi, empresa dedicada al desarrollo de equipos electrónicos, destacándose entre sus productos, los celulares.

Una de las hipótesis planteadas es, si los costos de los celulares de cualquier empresa particular difiere mucho del precio oficial, es un indicador para sospechar sobre la procedencia.

## 8 Licencia

El conjunto de datos se distribuye en Zenodo con licencia «Creative Commons Attribution 1.0 Generic». Se elige esta licencia, pues da libertad para compartir (copiar y redistribuir) el material en todo medio y formato, además de adaptarlo para cualquier propósito, incluso comercial. Esto, con el objetivo de que la información extraída pueda ser usada para análisis tanto de consumidores como de tiendas que ofrezcan los productos.

## 9 Código

En conjunto de datos puede encontrarse en el siguiente repositorio:

[https://github.com/andres-merino/M2.851\\_20221\\_PRACTICA\\_1.git](https://github.com/andres-merino/M2.851_20221_PRACTICA_1.git)

Resumiendo, el código primero extrae la información de todos los productos ofertados en la página oficial de la tienda de Xiaomi Ecuador, junto con sus precios (precios oficiales). Posteriormente, itera sobre la lista obtenida para buscar cada uno de los modelos de celulares en la página de Mercado Libre. En cada página de resultado, genera una lista de todos los resultados e ingresa a la página de cada anuncio para extraer la información de los dispositivos (características y precios de venta). Finalmente, guarda los datos en un archivo csv.

## 10 Dataset

En conjunto de datos puede encontrarse en la página de Zenodo con el siguiente enlace:

<https://zenodo.org/record/7336315#.Y3g-d8eZND8>

## 11 Video

El video explicativo puede encontrarse en la siguiente carpeta:

[https://drive.google.com/drive/folders/1YRzSSkO3X749GQZI\\_IzmKyLZJj\\_7tpj6?usp=sharing](https://drive.google.com/drive/folders/1YRzSSkO3X749GQZI_IzmKyLZJj_7tpj6?usp=sharing)

## 12 Contribuciones

Contribuciones	Firma
Investigación previa	AM, MC
Redacción de las respuestas	AM, MC
Desarrollo del código	AM, MC
Participación en el video	AM, MC