

DATASET: PRECIOS DE MEDICAMENTOS EN BOLIVIA

ESTUDIANTE MASTER EN CIENCIA DE DATOS

IVANA LORENA CANDY AGUILAR

1. Contexto.

En el contexto actual que vivimos como es la pandemia que inicio con el **Covid 19** y que de manera periódica van ingresando nuevas sepas. Es de suma importancia en el contexto nacional e internacional hacer un seguimiento exhaustivo a los precios de los medicamentos, esto con el fin de evitar el sobre precio y especulación en la venta de los mismos, que son artículos de primera necesidad para la población.

2. Título.

El titulo elegido para el dataset es **PreciosMedicamentosBolivia**.

3. Descripción del dataset.

El conjunto de datos que se ha elegido apartir de una conocida empresa farmaceutica en Bolivia que es bastante representativa respecto a los precios actuales de medicamentos en el mercado Boliviano.

Los datos mas relevantes y requeridos son el nombre del medicamento y su precio, ademas se ha agregado la fecha de extracción.

4. Representación gráfica.

The screenshot displays a webpage from a Bolivian pharmacy. The header includes a search bar with the text "¿Qué estás buscando?" and a navigation menu with categories like "Salud y Medicamentos", "Alimentos y Bebidas", "Cuidado Personal", "Mamá & Bebé", "Tienda", "Adulto Mayor", "Belleza", "Bienestar y Deportes", "Mascotas", and "Hogar". The main content area shows a grid of medicine products. Two callout boxes are present: one labeled "PrecioMedicamento" pointing to the price "Bs103,08" for "Isomar Spray Descongest X 100ML C Ac Hialuronico", and another labeled "NombreMedicamento" pointing to the product name "Bronquisedan" for "Bronquisedan Jarabe".

Medicamento	Precio
Isomar Spray Descongest X 100ML C Ac Hialuronico	Bs103,08
Hedelix 35Mg 5Ml Jbe X100ml Hedera Helix Farmacorp	Bs45,00
Hiedrix 35Mg 5Ml Jbe X120 100Ml Hedera Helix	Bs60,60
Torante Hedera Helix 15Mg Ml Jbe X200ml Generico	Bs62,80
Bronquisedan Jarabe	
Tocex	

5. Contenido.

Ya que existen bastantes categorías de medicamentos en esta práctica hemos limitado estos a la categoría que corresponde a **salud respiratoria y gripe**.

Los datos incluidos en el dataset son:

Indice, es el numero de registro en el dataset, es único y es tipo numérico. Generado con script python.

FechaCaptura: Este dato no es capturado pero existe la necesidad de conocer el día y hora de capturar para en lo futuro realizar una investigación del cambio de precios en la industria farmacéutica. En la practica se captura con código python la fecha actual del sistema. Haciendo uso de la librería datetime y dando el formato dd/mm/aa hh:mm.

NombreMedicamento, corresponde al nombre del medicamento y es de tipo texto y corresponde a las etiquetas tipo <h2>. Haciendo uso de comando de la librería BeautifulSoup como find_all y find. No requiere actualización diaria son en caso de que el nombre cambie o haya nuevos medicamentos en lista.

PrecioUnitario, corresponde al precio unitario del medicamento, un medicamento puede venir en diferentes cantidades, pero hemos verificado en la web, que se utiliza el precio por unidad. El precio corresponde a las etiquetas , se utilizo también los comandos find_all y find de BeautifulSoup. Este dato se debe actualizar de forma diaria.

6. Agradecimientos.

Se agradece a la empresa FARMACORP ya que los datos se han recopilado de su pagina web www.farmacorp.com.

Para actuar en el marco de lo legal se ha procedido a verificar su lineamiento de búsquedas y bots <https://farmacorp.com/robots.txt> se pudo verificar que no esta limitando en cuanto a precios y nombre de medicamentos.

Ademas hemos tratado de limitar nuestras extracciones a un numero pequeño de consultas solo con fines de investigación.

7. Inspiración.

Este conjunto de datos es interesante porque a partir de ellos se puede realizar un seguimiento de los precios de los medicamentos de manera periódica, muchos estudios pueden salir a partir de ello como el responder, si efectivamente los precios suben cuando una nueva ola o sepa de covid ingresa al país, o , como estos precios cambian (suben o bajan) en ciertas temporadas del año.

Ademas de otras utilidades que se podrían sacar de este dataset en áreas como las ventas y marketing.

8. Licencia.

La licencia que se ha escogido para el dataset es Released Under CC BY-NC-SA 4.0 License (Atribución-NoComercial-CompartirIgual).

Ya que al ser datos originados en una empresa “privada” estos son propiedad de esta empresa, nuestra extracción fue con fines investigativos y bajo esta licencia nos deslindamos de todo aspecto legal negativo en la que podamos incurrir al extraer estos datos.

Por lo que, al tener el dataset con esta licencia se puede difundir, compartir, modificar, calcular pero no con fines comerciales.

9. Código.

El repositorio GitHub donde se encuentra el dataset y todos los productos de la practica Web Scraping es:

<https://github.com/IvanaCandyAguilar/Medicamentos>

10. Enlace del DOI.

<https://zenodo.org/badge/DOI/10.5281/zenodo.5654973.svg>

11. Enlace Video Explicación.

<https://github.com/IvanaCandyAguilar/Medicamentos/blob/master/icandyVideoExplicacion.mp4>

Contribuciones	Firma
Investigación previa	ILCA *
Redacción de las respuestas	ILCA *
Desarrollo del código	ILCA *

*No se pudo encontrar compañero@ de práctica, por lo que se optó por realizar el trabajo de manera individual