

### *Ciencia e Ingeniería*

Revista Interdisciplinaria de Estudios en Ciencias  
Básicas e Ingenierías  
ISSN 2389-9484

Año 2025, enero-junio, Vol. 12, N.º 1, e14533404  
Facultades de Ciencias Básicas y Aplicadas e Ingeniería.  
Universidad de La Guajira  
La Guajira, Riohacha, Colombia  
<http://revistas.uniguajira.edu.co/index.php/cei>  
Este documento fue depositado en Zenodo. DOI:  
<https://www.doi.org/10.5281/zenodo.14533404>  
Para cita prefiera el ARK:  
<https://n2t.net/ark:/60540/14533404>

#### Luis Cobo

<https://orcid.org/0000-0003-3034-101X>  
[lacobo@universidadean.edu.co](mailto:lacobo@universidadean.edu.co)  
Universidad Escuela de Administración de  
Negocios. Bogotá, D.C., Colombia

#### Sebastián Benítez Baldión\*

<https://orcid.org/0009-0007-8642-403X>  
[sbenite82016@universidadean.edu.co](mailto:sbenite82016@universidadean.edu.co)  
Universidad Escuela de Administración de  
Negocios. Bogotá, D.C., Colombia

#### Jhon Sebastián Perdomo González

<https://orcid.org/0009-0000-3248-1898>  
[jperdomo19741@universidadean.edu.co](mailto:jperdomo19741@universidadean.edu.co)  
Universidad Escuela de Administración de  
Negocios. Bogotá, D.C., Colombia

#### Helver Ariel Novoa Mendoza

<https://orcid.org/0000-0003-0753-3152>  
[hnoam95708@universidadean.edu.co](mailto:hnoam95708@universidadean.edu.co)  
Universidad Escuela de Administración de  
Negocios. Bogotá, D.C., Colombia

# WEB SCRAPING EN SUPERMERCADOS PARA EL SEGUIMIENTO DE PRECIOS DE LA CESTA BÁSICA ALIMENTARIA

## Web Scraping in Supermarkets for Monitoring Basic Food Basket Prices

### RESUMEN

Hacer seguimiento de los precios de la canasta básica alimentaria es fundamental para asegurar la seguridad alimentaria y la estabilidad económica de un país. Para tal propósito, en el presente trabajo se desarrollaron un conjunto de scripts en Python con los que se recopilaban datos de precios de cinco sitios web de supermercados. Para esto se empleó el automatizador de navegadores Selenium, que junto con librerías como Pandas y Numpy, permitieron el consolidar la información correspondiente al nombre de los productos, sus descripciones, sus precios y las fechas de captura de información para un periodo de tiempo de 100 días. Por último, se realizaron una serie de tableros de control en Power BI con la finalidad de conocer la experiencia de un grupo de usuarios con la información presentada. Los resultados de la prueba de concepto con usuarios permitieron identificar aspectos de presentación, búsqueda, comprensión, inferencia y toma de decisiones a partir de las visualizaciones presentadas. Sobre este último aspecto, los usuarios señalaron que la ausencia de información relacionada con el precio por gramo o por mililitro para los productos, no les permite hacer comparaciones que los conduzca a diferenciar claramente los precios de estos. Este es un elemento que busca ser abordado en un próximo trabajo.

**Palabras clave:** Cesta Básica Alimentaria, Inteligencia de negocios, Precios, Tableros de control, Web scraping.

### ABSTRACT

Tracking the prices of the basic food basket is fundamental to ensure food security and economic stability of a country. For this purpose, a set of Python scripts were developed to collect price data from five supermarket websites. For this purpose, the Selenium browser automator was used, which together with libraries such as Pandas and Numpy, allowed the consolidation of the information corresponding to the name of the products, their descriptions, their prices and the dates of information capture for a period of time of 100 days. Finally, a series of dashboards were created in Power BI in order to know the experience of a group of users with the information presented. The results of the concept test with users identified aspects of presentation, search, comprehension, inference and decision making based on the visualizations presented. Regarding this last aspect, users pointed out that the absence of information related to the price per gram or per milliliter for the products does not allow them to make comparisons that lead them to clearly differentiate the prices of these products. This is an element that will be addressed in a future paper.

**Keywords:** Basic Food Basket, Business Intelligence, Pricing, Dashboards, Web scraping.

Recibido: 19 de septiembre de 2024

Aceptado: 20 de octubre de 2024

Publicado: 7 de enero de 2025



## INTRODUCCIÓN

El seguimiento de los precios de la canasta básica alimentaria es una práctica crucial en la gestión de precios tanto para mercados minoristas como mayoristas. Esta actividad permite a las empresas tomar decisiones más informadas y optimizar sus estrategias de fijación de precios, lo cual es esencial para mantenerse competitivas en un entorno económico fluctuante. En este contexto, el uso de técnicas avanzadas como el web scraping ha emergido como una herramienta poderosa y eficiente para recopilar datos de precios de productos y servicios en línea. Este facilita la extracción automatizada de datos desde sitios web, proporcionando a las empresas una visión detallada y actualizada de las fluctuaciones de precios en el mercado.

El problema central es la dificultad de mantener un seguimiento preciso y constante de los precios de la canasta básica alimentaria, por la gran cantidad de datos disponibles y la variabilidad de estos. La implementación manual de este seguimiento es laboriosa y propensa a errores, lo que puede resultar en decisiones basadas en información desactualizada o inexacta. Por lo tanto, la necesidad de una solución automatizada y eficiente es evidente. El trabajo se desarrolla en el contexto de una arquitectura de inteligencia de negocios, a partir de la cual se hace uso de herramientas como el web scraping y de tableros de control, con la cuales se extrae, transforma y analiza los datos de precios de manera efectiva.

La importancia de esta investigación reside en su capacidad para proporcionar una solución integral que no solo facilita la recopilación de datos, sino que también optimiza su análisis y visualización. La combinación de web scraping con técnicas de aprendizaje automático, como el modelo ARIMA, permite no solo obtener datos precisos y actualizados, sino también realizar predicciones informadas sobre las tendencias de precios futuras. Además, el desarrollo de tableros de control interactivos con herramientas de visualización como Power BI asegura que los usuarios puedan interpretar fácilmente los datos y tomar decisiones informadas. Este enfoque integral es crucial para mejorar la gestión de precios en el ámbito de la canasta básica alimentaria, ofreciendo una herramienta valiosa para minoristas, mayoristas y consumidores por igual.

En este artículo se presentan una serie de trabajos que exploran las soluciones de inteligencia de negocios (BI) en el ámbito de los sistemas de precios y el web scraping. Seguidamente, se examinan aquellas investigaciones que se enfocan en la integración del BI con técnicas para la recolección y análisis de información de diversas fuentes en línea. Por último, se consideran aquellos trabajos que abordan el uso específico de este método para el seguimiento de precios.

## Soluciones de inteligencia de negocios y sistemas de precios

En el trabajo de Krummert (2016) se abordan los datos sobre el gasto en restaurantes, mencionando que los datos del U.S. Census Bureau mostraron un aumento de las ventas en establecimientos de comida y bebida en febrero de 2016, mientras que los datos de la Oficina de Estadísticas Laborales mostraron un aumento en los precios de los menús en el período de 12 meses que terminó en febrero de 2016. Además, una encuesta realizada por el proveedor de inteligencia empresarial Prosper Insights and Analytics encontró un aumento en el número de millennials que comen fuera de casa con mayor frecuencia en enero de 2016. Los autores destacan la importancia de examinar múltiples fuentes de datos para obtener una imagen completa de las tendencias en el gasto en restaurantes. Los datos recopilados de diversas fuentes, como censos, estadísticas laborales y encuestas, proporcionan información valiosa sobre el comportamiento de los consumidores y las tendencias del mercado. Estos hallazgos muestran un panorama mixto, con un

aumento en las ventas y los precios de los menús, pero también una disminución en los precios mayoristas de alimentos. Además, el aumento en el número de millennials que comen fuera de casa destaca la importancia de comprender las preferencias y hábitos de consumo de diferentes grupos demográficos.

Los beneficios de este método van más allá del monitoreo de precios. Por ejemplo, permite el análisis de la competencia en el mercado al rastrear y alertar sobre actualizaciones en los catálogos de productos de los competidores, optimizando las estrategias de comercio electrónico. Además, este puede identificar discrepancias en los precios de los proveedores, permitiendo a las empresas implementar software de inteligencia de negocios para comparar y analizar eficazmente los precios de los proveedores. Esta tecnología ayuda a las empresas a tomar decisiones informadas sobre la selección de proveedores y la negociación de precios, mejorando la eficiencia en los procesos de adquisición (Financial Managers' Forum, 2005).

### **Soluciones de inteligencia de negocios y web scraping**

La implementación de soluciones de inteligencia de negocios en la gestión de precios permite a las empresas tomar decisiones más informadas y optimizar sus estrategias de fijación de precios. Determinar el precio óptimo para ofrecer un producto en el mercado es crucial, especialmente cuando se emplean métodos de fijación de precios basados en la competencia. Este enfoque considera los precios establecidos por los principales competidores para productos similares, teniendo en cuenta su presentación y calidad (Kotler & Armstrong, 2012). En consecuencia, las empresas monitorean los precios de los competidores más importantes en el mercado para establecer sus propios precios.

Las soluciones de inteligencia de negocios pueden aprovechar este método como una técnica para la extracción automatizada de datos de sitios web, para recopilar información sobre precios de productos y servicios en línea (Khder, 2021). Las aplicaciones de este incluyen el monitoreo de precios, la investigación de mercado y la inteligencia de precios. Cuando se realiza de manera legal y ética, sin infringir derechos de autor ni datos personales, convirtiéndose en una herramienta valiosa para rastrear las fluctuaciones de precios en diversos sectores. En la era moderna, el web scraping juega un papel crucial en la inteligencia de negocios, proporcionando una visión detallada de la dinámica del mercado y del comportamiento del consumidor.

En Singrodia et al. (2019) resaltan la importancia de este método como una herramienta para recopilar y organizar datos de Internet, que a menudo se presentan de manera desorganizada y difícil de utilizar en procesos mecánicos. En los últimos tiempos, se han desarrollado procedimientos y herramientas para convertir estos datos en información organizada, que puede ser aprovechada tanto en sistemas B2C como B2B. El documento proporciona una introducción básica y menciona el software y herramientas utilizadas en esta práctica. Además, exploran el proceso de este incluyendo diferentes técnicas utilizadas, y discuten algunos pros y contras de esta metodología. También, resaltan la diversidad de campos en los que se puede aplicar el web scraping, como el gobierno de datos, el Big Data y la inteligencia de negocios. Terminan subrayando las numerosas oportunidades que existen para aprovechar estos datos recopilados a través de este método, mostrando el potencial y la versatilidad de esta herramienta en diversos contextos.

En un enfoque interesante para detectar y extraer datos de fuentes no estructuradas en línea, específicamente relacionadas con eventos de atletismo en Portugal. El objetivo del trabajo de Lima et al. (2019) es el de crear un repositorio de Data Warehouse que sirva como base para el análisis de inteligencia de negocios y minería de datos. Con esta estrategia se pretende combinar los datos extraídos con información de otras fuentes, enriqueciendo así la información y facilitando la obtención de conclusiones

más interesantes. El estudio de caso se centra en los resultados de eventos de atletismo en Portugal de los últimos 12 años, los cuales se encuentran disponibles en línea en diferentes formatos, principalmente en documentos portátiles (PDF). Cada asociación de atletismo publica los archivos con su propio formato interno, lo que plantea un desafío adicional para la extracción y consolidación de los datos. Además, con caso de estudio, los autores proponen un mecanismo integrador que relaciona los resultados de los eventos deportivos con su ubicación geográfica y las condiciones atmosféricas, lo que permite analizar cómo estos factores pueden influir en el rendimiento de los atletas.

SINSE+ es una plataforma de inteligencia de negocios que tiene la capacidad de adquirir y analizar datos abiertos con el fin de definir y analizar indicadores. Aunque inicialmente se diseñó y probó en el contexto de la salud y la información social, esta plataforma puede ser aplicada en cualquier área de interés. Su objetivo principal es el de generar cuadros de mando que combinen información proveniente de diversas fuentes de datos abiertos, con el propósito de mejorar el bienestar de los ciudadanos. El sistema SINSE+ está compuesto por varios módulos, incluyendo uno de extracción de datos mediante web scraping, uno ETL (extracción, transformación y carga), otro de almacenamiento y consulta, y finalmente uno de representación de indicadores a través de un dashboard. La plataforma ha sido sometida a pruebas utilizando información relacionada con temas sanitarios, poniendo especial énfasis en el informe BES (Bienestar Socioeconómico). SINSE+ es uno de los resultados de un proyecto industrial financiado por la Región de Calabria y coordinado por el consorcio ICT-SUD. Este proyecto demuestra el potencial de utilizar datos abiertos y herramientas de inteligencia empresarial para obtener información relevante y tomar decisiones informadas en diferentes áreas de aplicación (Federico et al., 2016).

## Web scraping para el seguimiento de precios

La canasta básica alimentaria, que comprende bienes y servicios esenciales, es necesaria para satisfacer las necesidades de los hogares y los individuos (DANE, 2024). Sin embargo, los precios de estos elementos esenciales se ven afectados por diversos factores, como bloqueos nacionales, condiciones climáticas adversas, aumentos en los precios de los combustibles y eventos internacionales como la inflación global y conflictos geopolíticos (Martínez, 2022; Departamento Nacional de Planeación, 2022). Estas alzas en los precios afectan desproporcionadamente a los estratos socioeconómicos más bajos, obligando a muchos hogares a adoptar estrategias financieras para gestionar los altos costos, a menudo renunciando a ciertos productos o buscando alternativas más económicas (ACIS, 2022).

El trabajo desarrollado por Taylor et al. (2023) tiene como objetivo evaluar la viabilidad y el impacto de aplicar una medida de costos de alimentos altamente desagregada y representativa a nivel nacional en Canadá. Los autores utilizan la tecnología de web scraping en los métodos de la Canasta Nacional de Alimentos Nutritivos (NNFB). También buscan probar la hipótesis de que un NNFB digital emparejado con el producto (dNNFB) se correlaciona con las medidas existentes de la canasta de mercado y cuantifican cualquier diferencia en los costos. El estudio se llevó a cabo utilizando datos de precios de alimentos recopilados en noviembre de 2021, obtenidos de la mayoría de las pancartas de Loblaw (<https://www.loblaw.ca/>) en todo Canadá. Los resultados muestran que los costos de dNNFB varían ampliamente en todo el país, con diferencias notables según las características regionales, las tiendas y los grupos de edad y género. En comparación con las estimaciones nacionales existentes, las estimaciones de dNNFB superan la medida de la canasta de mercado nacional en todas las comparaciones a nivel subnacional en todo el país. Este estudio demostró que la recopilación digital de datos de precios de alimentos es una estrategia factible para calcular los costos de la canasta básica. Los hallazgos sugieren

que puede haber una subestimación rutinaria del impacto de la inflación de alimentos para los consumidores, especialmente aquellos que tienen acceso limitado a determinados entornos alimentarios.

En el desarrollo de una teoría sobre la estrategia promocional en la venta minorista de alimentos, Loy y Ren (2022) llevan a cabo una prueba utilizando precios de venta en línea de alimentos. Para esto utilizan un código Python para recopilar información de la página web (<https://www.mytime.de/>), un outlet de comestibles en línea que pertenece al grupo Bunting. Es grupo es un minorista de alimentos en el noroeste de Alemania. Los resultados de las ventas promocionales de esta cadena de retail muestran una relación complementaria entre la amplitud y la profundidad de las ventas, lo que indica que, para atraer a los consumidores, las tiendas aumentan tanto el número (amplitud) como la profundidad de las promociones de precios. Este estudio proporciona una comprensión más clara de cómo las estrategias promocionales influyen en el comportamiento de compra de los consumidores en el ámbito de la venta minorista de alimentos en línea.

El aumento en la generación de datos, impulsado por tecnologías como Internet, ha llevado al fenómeno de Big Data, ofreciendo tanto oportunidades como desafíos para las estadísticas oficiales y la investigación socioeconómica (Faramarzi et al., 2022). Los investigadores utilizan el web scraping para recopilar precios diarios de alimentos y bebidas, proporcionando una alternativa viable a los métodos tradicionales para calcular los índices de precios. Esta técnica proporciona datos en escalas de tiempo más pequeñas, como semanal o diaria, mejorando la precisión y ahorrando tiempo, especialmente en áreas urbanas. Los estudios indican que el web scraping puede reemplazar eficazmente los métodos convencionales para el índice de precios al consumidor (IPC) y otras estadísticas de precios, proporcionando información más oportuna y detallada (da Silva et al., 2023).

## MATERIALES Y MÉTODOS

Se utilizó el web scraping como estrategia para la extracción, transformación y carga de los precios de la canasta básica alimentaria. Inicialmente, se identificaron los portales web de interés, que incluían las páginas de las principales cadenas de *retail* colombianas. Se seleccionaron las páginas web de las siguientes cadenas Éxito, Jumbo, D1, Makro, Carulla en idioma español (Tabla 1). Esta información se utilizó para llevar a cabo un análisis de costos de los productos seleccionados de la canasta familiar.

Tabla 1. Fuentes de datos seleccionadas para el estudio

| Cadenas de <i>retail</i> | URL   |
|--------------------------|---|
| Éxito                    | <a href="https://www.exito.com/home-mercado">https://www.exito.com/home-mercado</a> |
| Jumbo                    | <a href="https://www.tiendasjumbo.co/">https://www.tiendasjumbo.co/</a>             |
| D1                       | <a href="https://domicilios.tiendasd1.com/">https://domicilios.tiendasd1.com/</a>   |
| Makro                    | <a href="https://tienda.makro.com.co/">https://tienda.makro.com.co/</a>             |
| Carulla                  | <a href="https://www.carulla.com/">https://www.carulla.com/</a>                     |

## Procedimientos y técnicas de análisis de la información

La realización del web scraping o raspado web implicó una serie de procedimientos y técnicas fundamentales para extraer datos de sitios web de manera efectiva, para lo cual se siguieron los siguientes pasos:

---

## Identificación del sitio web y de la información por extraer

El primer paso fue identificar los sitios web de interés. En este caso, se hizo un enfoque en las páginas web de las cadenas de supermercados antes mencionadas. Una vez que se habían seleccionado los sitios web objetivo, fue necesario determinar qué información se iba a extraer de ellos. Se diseñó un algoritmo de Web que permitiera extraer datos de categorías, marcas, descripción, precio y estado de producto de almacenes de cadena como el Éxito, Jumbo, D1 y Olímpica. En este proyecto, se buscó obtener datos numéricos relacionados con los precios de los productos y cadenas de texto que representaran los nombres de los productos. Esta etapa requiere un análisis detallado de las páginas web para identificar los elementos específicos que contenían esta información.

## Identificación de la estructura del sitio web

Para extraer datos de manera efectiva, fue necesario analizar el código HTML de las páginas y examinar las etiquetas relevantes que contenían los datos de interés. Esta fase permitió definir cómo se accedía y extraería la información de las páginas web.

## Selección de las librerías de web scraping

En el proceso de este método, existen diversas bibliotecas y herramientas disponibles para automatizar la extracción de datos. En este estudio, se optó por utilizar Selenium, una biblioteca conocida por su facilidad de uso, su curva de aprendizaje amigable, su amplia documentación y la robusta comunidad de desarrolladores que la respalda. Selenium facilita la configuración de rastreadores web que recopilan datos de manera eficiente y precisa.

## Almacenamiento de los datos extraídos

Los datos extraídos se almacenaron en archivos CSV. La elección de este tipo de formato se debió a su capacidad para organizar y gestionar eficazmente los datos recopilados. Finalmente, estos archivos se fusionaron en un único archivo de Excel, el cual se constituyó como la base de datos para la etapa de análisis.

## Análisis de la información extraída

Se emplearon diversos métodos y procesos para llevar a cabo la investigación. En primer lugar, se utilizó el web scraping como estrategia para la extracción, transformación y carga de los precios de la canasta básica alimentaria. Este proceso se dividió en varias etapas. Primero, se identificaron los portales web de interés, que incluían las páginas de las principales cadenas de *retail* colombianas. Luego, se determinó la información a extraer, principalmente precios de productos y nombres de estos. Una vez identificada la estructura de los sitios web, se empleó la biblioteca de Python Selenium para automatizar la extracción de datos de manera efectiva. Los datos extraídos se almacenaron en archivos CSV y se fusionaron en un archivo de Excel para su posterior análisis.

En cuanto al aprendizaje automático, se seleccionaron técnicas apropiadas para el análisis de precios de la canasta básica alimentaria. Este proceso involucró la elección de modelos adecuados y su configuración según los requisitos del análisis. Los modelos fueron entrenados y ajustados utilizando datos históricos

recopilados mediante web scraping. Posteriormente, se evaluaron los modelos para determinar su desempeño y realizar ajustes según los resultados obtenidos, asegurando así la precisión de las predicciones de precios.

En relación con el desarrollo de tableros de control, se seleccionaron herramientas de visualización adecuadas, en este caso el software Power BI. Los tableros de control se configuraron y desarrollaron para presentar eficazmente los resultados del análisis de precios. Se llevaron a cabo pruebas exhaustivas para corregir errores y mejorar la presentación visual de la información, garantizando así que los usuarios pudieran interpretar fácilmente los datos y tomar decisiones informadas.

## RESULTADOS Y DISCUSIÓN

El desarrollo del aplicativo funcional para el seguimiento de precios en la canasta alimentaria ha demostrado ser una herramienta efectiva y comprensible para los usuarios. A continuación, se describen los principales resultados obtenidos con las diferentes vistas del aplicativo, diseñados para dar información detallada y relevante sobre los precios de los productos. La primera vista del Dashboard, Figura 1, permite a los usuarios aplicar filtros a sus búsquedas según sus necesidades. Este dashboard es una herramienta crucial para la comparación de precios y la identificación de oportunidades de ahorro.

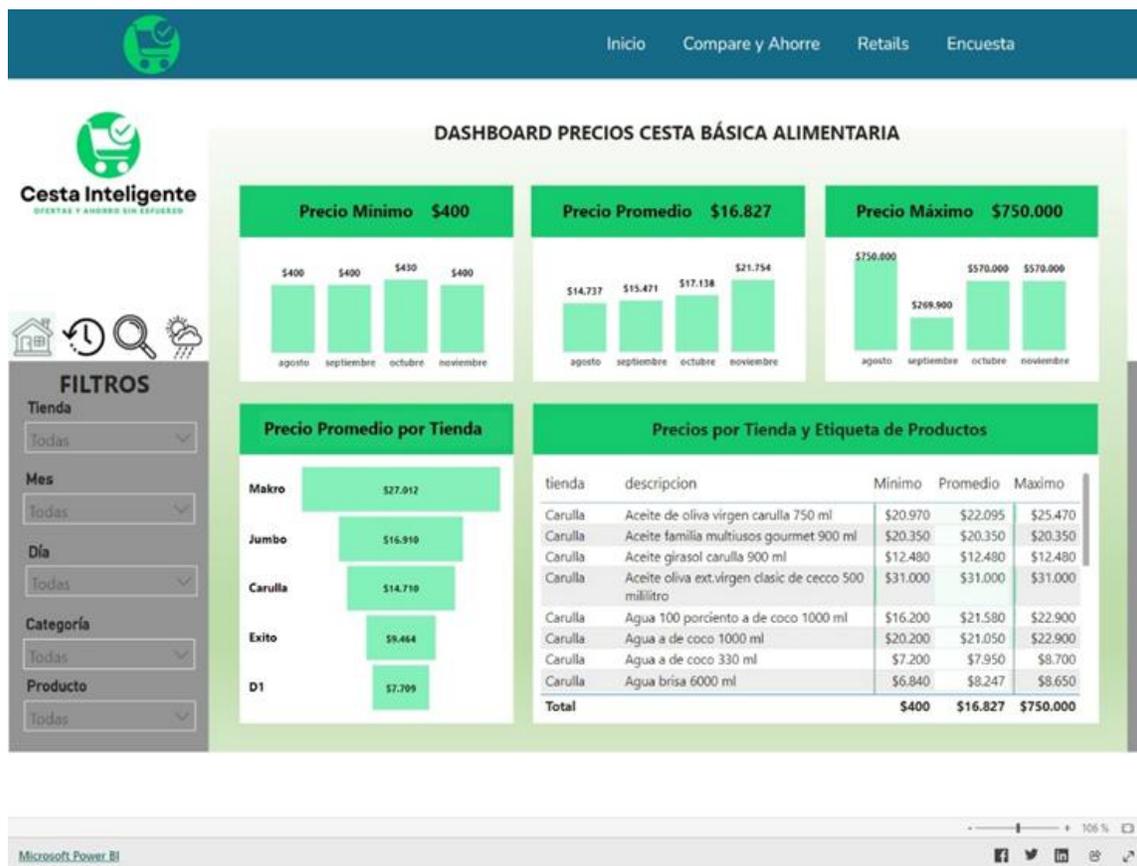


Figura 1. Panel de control del aplicativo compare y ahorre.

Los precios de los productos en la canasta alimentaria se representan con el precio mínimo, máximo y promedio por tienda (Figura 2). Además, detalla los precios por tienda para cada producto con su correspondiente etiqueta. Esta información es fundamental para identificar variaciones de precios entre tiendas y entender la distribución de precios de cada producto.

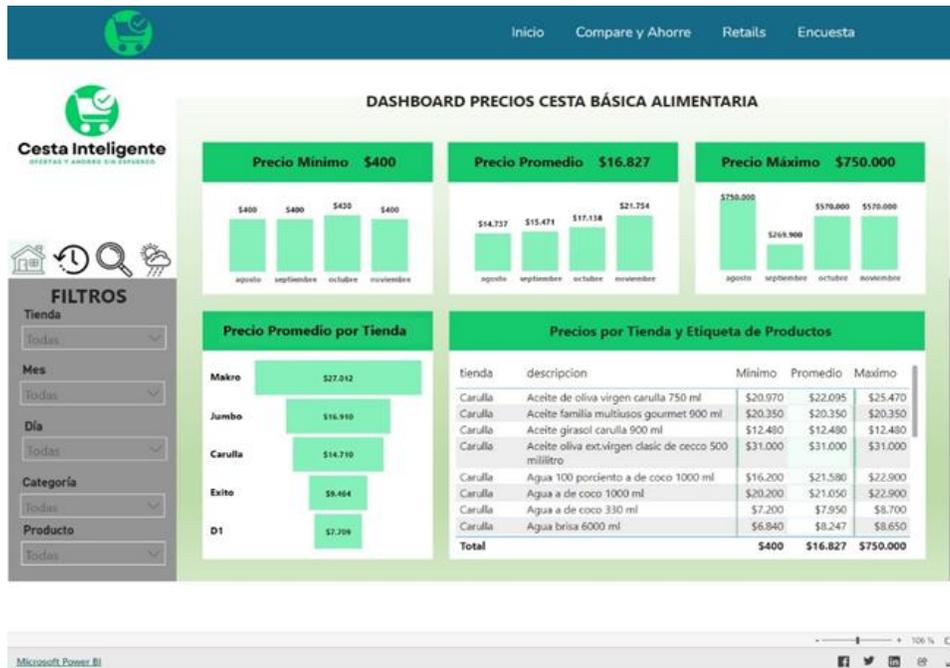


Figura 2. Precio mínimo, máximo y promedio por tienda.

La Figura 3. Ofrece una perspectiva histórica de los precios, permitiendo comparar métricas clave a lo largo del tiempo, como precios promedio diarios, variaciones diarias de precios máximos y mínimos, y el histórico de precios por mes y día. Esta vista es valiosa para identificar patrones estacionales, tendencias a largo plazo y cambios abruptos en los precios.



Figura 3. Comparativa de métricas históricas por día.

La Figura 4. permite un desglose detallado de la estructura jerárquica de los productos en la canasta alimentaria, organizada por categoría, producto y descripción. Esta vista facilita la comprensión de la diversidad de productos y la identificación de tendencias dentro de categorías específicas.



Figura 4. Comparativa de métricas históricas por mes.

La Figura 5. realiza un pronóstico a 10 días para cada tipo de producto, basado en datos históricos. Esta información es crucial para la planificación de inventario, toma de decisiones estratégicas y anticipación de posibles cambios en los precios. La visualización clara de los pronósticos permite la identificación de tendencias futuras y la implementación de medidas proactivas.

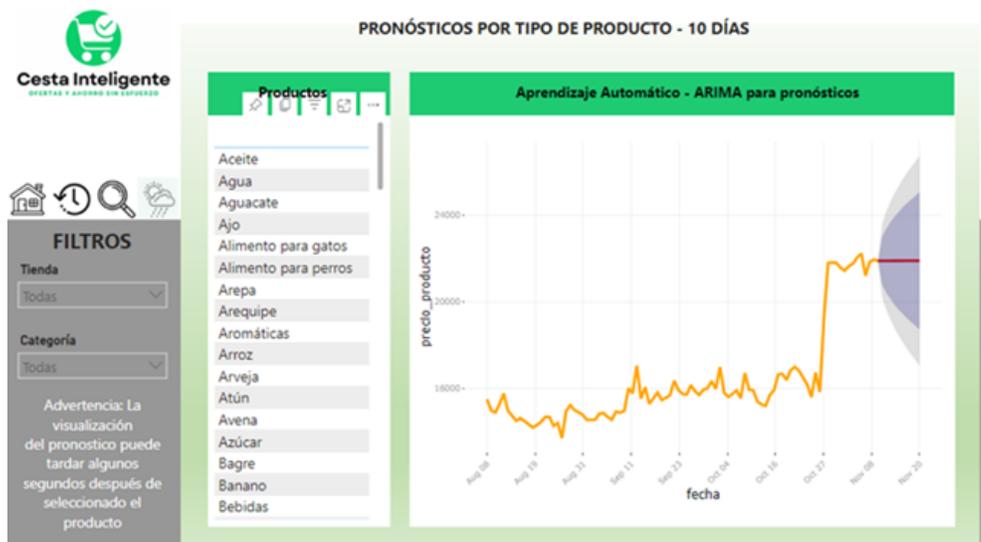


Figura 5. Pronósticos por tipo de producto.

Los tableros de control desarrollados han demostrado su capacidad para ofrecer una visión detallada y comprensible de los precios de la canasta alimentaria, facilitando la toma de decisiones informadas y optimizando procesos mediante inteligencia de negocios.

## Prueba de concepto

La prueba de concepto estadística proporcionó una visión detallada sobre la experiencia de los usuarios con los tableros de control de "cesta inteligente". En general, los resultados sugieren una recepción positiva de la plataforma en varios aspectos clave (Tabla 2). Por ejemplo, la facilidad de encontrar y acceder a la plataforma obtuvo una puntuación promedio de 8, lo que indica que la mayoría de los usuarios encontraron el proceso relativamente sencillo. La velocidad de carga de los tableros de consulta fue bien recibida, con una puntuación promedio de 8,13 lo que sugiere una experiencia satisfactoria en rendimiento.

Tabla 2. Puntaciones a preguntas prueba de concepto

| Compare y ahorre.<br>Prueba de concepto   | Puntuación mínima | Puntuación promedio | Puntuación máxima |
|---|-------------------|---------------------|-------------------|
| ¿Qué tan fácil fue encontrar y acceder a cesta inteligente?   | 6                 | 8,00                | 9                 |
| ¿Qué tan rápido se cargaron los tableros de consulta?   | 5                 | 8,13                | 10                |
| ¿La información sobre los productos es clara y comprensible?  | 7                 | 8,38                | 10                |
| ¿Cómo calificarías la navegación por cesta inteligente?   | 7                 | 8,13                | 9                 |
| ¿La presentación visual de la información es atractiva?   | 7                 | 7,63                | 8                 |
| ¿Cesta inteligente le permite ver los precios de los productos de forma detallada?                              | 8                 | 8,50                | 10                |
| ¿Qué tan útiles son los filtros y opciones de búsqueda?   | 7                 | 8,38                | 9                 |
| ¿Cesta inteligente le ayudó a entender las variaciones de precios a lo largo del tiempo?                        | 7                 | 8,38                | 10                |
| ¿Cesta inteligente le permite identificar fácilmente el <i>retail</i> de procedencia de los productos?          | 7                 | 8,38                | 9                 |
| ¿Qué tan satisfecho está con la interacción general en cesta inteligente?                                       | 6                 | 8,25                | 10                |
| ¿Los gráficos y visualizaciones le ayudaron a tomar decisiones informadas?                                      | 7                 | 8,63                | 10                |
| ¿Cesta inteligente permitió comparar precios entre diferentes categorías de productos?                          | 7                 | 8,38                | 10                |
| ¿Se sintió motivado a explorar los tableros de control en profundidad?  | 8                 | 8,50                | 10                |
| ¿Cesta inteligente cumplió con sus expectativas en términos de información sobre la canasta básica alimentaria? | 7                 | 8,25                | 9                 |
| ¿Le gustaría que se agregaran más características o información adicional a cesta inteligente?                  | 6                 | 8,88                | 10                |
| ¿El rendimiento de cesta inteligente se mantuvo constante o experimentó problemas de carga?                     | 5                 | 7,50                | 9                 |
| ¿Cesta inteligente le permitió exportar datos o informes para su uso personal?                                  | 2                 | 6,50                | 9                 |
| ¿Te gustaría utilizar cesta inteligente de manera regular?  | 8                 | 8,88                | 10                |
| ¿Recomendaría cesta inteligente a otras personas interesadas en la canasta básica alimentaria?                  | 7                 | 8,25                | 9                 |

En cuanto a la claridad y comprensibilidad de la información sobre los productos, la plataforma recibió una calificación promedio de 8.38, lo que indica que la mayoría de los usuarios consideraron que la información proporcionada era clara y fácil de entender. Este aspecto es crucial para facilitar la toma de decisiones informadas por parte de los usuarios. Además, la navegación por la plataforma fue evaluada positivamente, con una puntuación promedio de 8,13; lo que sugiere una experiencia satisfactoria en términos de usabilidad y accesibilidad.

Sin embargo, hubo algunas áreas identificadas para posibles mejoras. Por ejemplo, aunque la presentación visual de la información recibió una puntuación promedio de 7,63, algunos usuarios podrían haber encontrado que la plataforma carecía de cierto atractivo visual. Asimismo, la capacidad de exportar datos o informes para uso personal obtuvo una puntuación promedio de 6,50, lo que sugiere que algunos usuarios podrían haber encontrado esta función insuficiente. A pesar de estas áreas de mejora potencial, la mayoría de los usuarios expresaron su satisfacción general con la plataforma. La puntuación media para la interacción en la plataforma fue de 8,25, con una máxima de 10, lo que indica que la mayoría de los usuarios quedaron satisfechos con su experiencia en "cesta inteligente".

## RESULTADOS Y DISCUSIÓN

A pesar de los logros en la recopilación y visualización de datos, uno de los principales desafíos encontrados fue la dificultad para realizar comparaciones precisas entre las tiendas debido a la variación en los gramajes de los productos. Este factor no fue considerado durante la fase inicial de recopilación de datos, lo que afectó la precisión de las comparaciones.

La diferencia en los gramajes implica que productos aparentemente similares pueden tener precios significativamente diferentes cuando se consideran sus pesos o volúmenes. Un paquete de arroz de 500 gramos en una tienda no es comparable con uno de 1000 gramos en otra sin normalizar los precios por unidad de medida. Esta omisión resultó en dificultades para establecer comparaciones justas y precisas entre los precios de los productos en las diferentes tiendas minoristas.

Para futuros estudios, es crucial implementar un proceso de normalización de los datos recolectados. Esto podría incluir la conversión de todos los precios a una base común, como el precio por kilogramo o litro, antes de realizar cualquier análisis comparativo. Además, se recomienda el uso de técnicas avanzadas de limpieza y procesamiento de datos para manejar mejor las discrepancias en la presentación de los productos entre las diferentes fuentes.

Aunque el uso de web scraping y Power BI demostró ser una herramienta poderosa para la recopilación y análisis de datos de minoristas, la consideración de factores como los gramajes es esencial para mejorar la precisión y utilidad de los análisis comparativos. Implementar estos ajustes en futuros estudios permitirá obtener insights más precisos y útiles para la toma de decisiones por parte de los consumidores finales.

## CONCLUSIONES

Bajo las limitaciones de este estudio, fue posible desarrollar una solución de inteligencia de negocios basada en web scraping y aprendizaje automático para el seguimiento de precios de la canasta básica alimentaria en tiendas de *retail*. En su formulación, la combinación de web scraping, técnicas de aprendizaje automático y tableros de control en Power BI proporciona una solución innovadora y eficiente para el seguimiento de precios de productos en tiempo real en el sector alimentario. Además, el uso del web scraping permite recopilar datos de precios de manera automatizada, agilizando el proceso de obtención de información y reduciendo la carga manual.

Por su parte, las técnicas de aprendizaje automático aplicadas a los datos recopilados permiten identificar patrones y tendencias en los precios a lo largo del tiempo, brindando información valiosa para la toma de

decisiones. Los tableros de control interactivos en Power BI proporcionan una visualización intuitiva de los datos, lo que facilita su exploración y análisis por parte de los responsables de la toma de decisiones.

Los resultados obtenidos muestran que los precios de la canasta básica alimentaria varían significativamente entre las tiendas de retail seleccionadas, lo que indica la existencia de diferencias en las estrategias de precios entre los proveedores. La solución propuesta en el estudio brinda a los responsables de la toma de decisiones en el sector alimentario una herramienta efectiva para monitorear y analizar los precios de los productos en tiempo real. Estos resultados pueden ser utilizados para tomar decisiones estratégicas informadas, mejorar la competitividad y la rentabilidad del negocio, y garantizar una oferta de productos accesibles para los consumidores.

## LITERATURA CITADA

- ACIS (2022). Análisis sobre los efectos del aumento de precios en la canasta básica. Informe técnico, Bogotá, Colombia.
- DANE-Índice de Precios al Consumidor (IPC). (2024). <https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc/> da Silva F., Almeida R., & Santos G. (2023). Web scraping techniques for consumer price index calculation. *Statistical Research Journal*, 45(3), 150–165.
- Departamento Nacional de Planeación (2022). Informe anual sobre los efectos de la inflación en los hogares colombianos. Bogotá, Colombia.
- Faramarzi, A., Hadizadeh, R., Fayyaz, S., Sajadimanesh, S. & Moradi, A. (2022). Web Scraping Technique for Producing Iranian Consumer Price Index. *Statistical Journal of the IAOS*, 38(1), 263–276.
- Federico, L., Franco P., Minelli, A., Perri, A., Caroprese, L., Picarelli, R., Tradigo, G., Vocaturo, E., Dattola, F., Fortunato, A., Lambardi, P., Laurita, S., Pellegrino, I., Garro, A., Pugliese, A., Tagarelli, A., Veltri, P. & Zumpano E. (2016). SINSE+: a software for the acquisition and analysis of open data in health and social area. In Bochicchio, M. A. & Mecca, G. (2016) (eds.), SEBD (pp. 310-317), Matematicamente.it. ISBN: 9788896354889.
- Financial Managers' Forum (2005). *Financial Analysis Planning & Reporting*, 5(5), 15.
- Khder, M.A. (2021). Web Scraping or Web Crawling: State of Art Techniques, Approaches, and Application. *International Journal of Advances in Soft Computing and its Applications*.
- Kotler, P. & Armstrong, G. (2012). *Principios de Marketing*. 14<sup>a</sup> edición. Pearson Educación. Madrid, España.
- Krummert, B. (2016). Millennials ramp up restaurant spending. *Restaurant Hospitality Exclusive Insight*, 3.
- Lima R., & Cruz E.F. (2019). Extraction and Multidimensional Analysis of Data from Unstructured Data Sources: A Case Study. *International Conference on Enterprise Information Systems*.
- Loy, J.-P. & Ren, Y. (2022). Web scraping of food retail prices. In: Gandorfer, M., Hoffmann, C., El Benni, N., Cockburn, M., Anken, T. & Floto, H. (Hrsg.), 42. GIL-Jahrestagung Künstliche Intelligenz in der Agrar- und Ernährungswirtschaft, Bonn: Gesellschaft für Informatik e.V. (pp. 177-182).

Martínez, F. (2022). Impacto del aumento de los precios de los combustibles en la canasta básica alimentaria. Reporte económico, Universidad Nacional.

Singrodia, V., Mitra, A. & Paul, S. (2019). A Review on Web Scrapping and its Applications. 2019 International Conference on Computer Communication and Informatics (ICCCI) (pp. 1-6). Coimbatore, India. doi: 10.1109/ICCCI.2019.8821809.

Taylor N. G. A., Luongo G., Jago E., & Mah C. L. (2023). Observational study of population level disparities in food costs in 2021 in Canada: A digital national nutritious food basket (dNNFB). Preventive Medicine Reports, 32, 102162. <https://doi.org/10.1016/j.pmedr.2023.102162>.

## BIODATA

**Luis Armando Cobo:** ingeniero de sistemas y computación, Magíster en Ingeniería de Sistemas y Computación, Doctor en Ingeniería y Ph.D. de la Universidad de Montreal (Canadá). Profesor universitario de la Universidad Ean Bogotá. Investigador reconocido por Minciencias adscrito al grupo de Investigación ONTARE en la Universidad Ean. Director de diversos trabajos de grado de maestrías y tesis doctorales en la Universidad Ean. Autor de los artículos científicos: Towards the Construction of a Smart City Model in Bogotá; En: Nigeria CEUR WORKSHOP PROCEEDINGS ISSN: 1613-0073 ed: CEUR-WS v.2714 p.176 - 193; también autor del capítulo de libro: "Modelos de precios dinámicos en los servicios de transporte en smartcities: un estado del arte". En el libro: Smart-cities y acústica sostenible- Cap. III/Red Rumbo. En: Colombia ISBN: 9789588928876 ed: v. p.105 - 112.

**Helbert Novoa:** Profesional en estadística, Magíster en ingeniería con énfasis en software y Magister en inteligencia de negocios. Mi carrera se ha desarrollado principalmente en entidades gubernamentales como el Ministerio del Trabajo, el SENA y la UBPD. En estas entidades he trabajado como estadístico, analista de bases de datos y científico de datos. También me he desempeñado como docente en la Universidad del Quindío, la Unicleretiana y la Escuela de Administración y Mercadotecnia EAM. Mis Publicaciones incluyen: Visual Attention Prediction Model Based on Prominence Maps, Machine Learning and Biometric Data DOI: 10.1109/CLEI53233.2021.9639958 y Minimum reference network for temperature modeling through distance-based algorithms DOI: 10.1109/CLEI56649.2022.9959899

**Sebastián Benítez Baldión:** Administrador de empresas, Especialista en gestión humana, Magister en administración de empresas. Dependencias institucionales: Senior Data Recruiter en Epam Systems. Aquí desempeñó un rol de reclutador técnico para roles en Data como Data software Engineer, Data Science, Machine Learning, Data Solution Architecture entre otros.

**Jhon Sebastián Perdomo González:** Licenciado en inglés, ingeniero de sistemas, Especialista en Gerencia de Tecnología, Magíster en Gerencia de Sistemas de Información y Proyectos Tecnológico. Dependencias Institucionales: Profesor tiempo completo adscrito al programa de Ingeniería Industrial de la Facultad de Ingenierías, Arquitectura y Urbanismo de la Fundación Universitaria Navarra - Uninavarra. Mis publicaciones incluyen: Ruta de innovación y sostenibilidad empresarial: Esri Colombia - <http://hdl.handle.net/10882/12236> - Solución de inteligencia de negocios para el seguimiento de precios a los productos de la cesta básica alimentaria, a través de Web Scrapping, aprendizaje automático y tableros de control - <http://hdl.handle.net/10882/13556>