



Solución de inteligencia de negocios para el seguimiento de precios a los productos de la cesta básica alimentaria, a través de Web Scraping, aprendizaje automático y tableros de control

Helver Ariel Novoa Mendoza

Jhon Sebastián Perdomo González

Sebastian Benitez Baldion

Universidad Ean

Facultad de Ingenieria

Facultad de administración

Maestría en inteligencia de negocios

Maestría en administración de empresas

Maestría en gerencia de sistemas de información y proyectos tecnológicos

Bogotá, Colombia

06/02/2024

Solución de inteligencia de negocios para el seguimiento de precios a los productos de la cesta básica alimentaria, a través de Web Scraping, aprendizaje automático y tableros de control

Helver Ariel Novoa Mendoza

Jhon Sebastián Perdomo González

Sebastian Benitez Baldion

Trabajo de grado presentado como requisito para optar al título de:

Magister

Magister en Gerencia de Sistemas de Información y Proyectos Tecnológicos

Director (a):

Luis Armando Cobo Campo

Modalidad:

Monografía

Universidad Ean

Maestría en inteligencia de negocios

Maestría en administración de empresas

Maestría en gerencia de sistemas de información y proyectos tecnológicos

Bogotá, Colombia

06/02/202

Nota de aceptación:

Firma del jurado

Firma del jurado

Firma del director del trabajo de grado

Ciudad, día/mes

Agradecimientos

Queremos expresar nuestro profundo agradecimiento a todas las personas que, de manera directa o indirecta, contribuyeron a la realización de esta tesis. En especial a nuestros familiares y allegados por su inquebrantable apoyo y comprensión a lo largo de esta travesía académica.

A nuestros amigos, cuya motivación y aliento fueron fundamentales en los momentos más desafiantes. A nuestros profesores y mentores, por su invaluable orientación y sabios consejos que guiaron nuestros caminos. Agradecemos también a todos aquellos cuyas experiencias, conocimientos y perspectivas enriquecieron este trabajo.

A cada persona que de alguna manera formó parte de este esfuerzo, nuestro más sincero reconocimiento. Este logro no hubiera sido posible sin su colaboración y contribución. Este trabajo lleva consigo el fruto de nuestro esfuerzo colectivo. ¡Gracias por ser parte de este viaje académico y por compartir este logro con nosotros!

Resumen

En esta investigación, se presenta el desarrollo de una solución de inteligencia de negocios innovadora que combina el Web Scraping, técnicas de aprendizaje automático y tableros de control en Power BI para realizar el seguimiento de los precios de los artículos de la canasta básica alimentaria en cinco tiendas de retail. El objetivo de esta solución es proporcionar a los responsables de la toma de decisiones en el sector alimentario una herramienta eficiente y precisa para monitorear y analizar los precios de los productos en tiempo real.

Para llevar a cabo este estudio, se implementó un sistema de Web Scraping que recopila datos de precios de los sitios web de las tiendas de retail seleccionadas. Estos datos se procesaron utilizando técnicas de aprendizaje automático para identificar patrones y tendencias en los precios de los productos a lo largo del tiempo. Posteriormente, se crearon tableros de control interactivos en Power BI que muestran visualmente los resultados obtenidos, permitiendo a los usuarios explorar y analizar los datos de manera intuitiva.

Los resultados obtenidos muestran que los precios de la canasta básica alimentaria presentan variaciones significativas entre las tiendas de retail seleccionadas. Además, se observa una tendencia general de aumento de precios en determinados productos a lo largo del tiempo. Estos hallazgos brindan una visión clara sobre el comportamiento de los precios en el mercado y pueden ser utilizados para tomar decisiones estratégicas, como ajustar las estrategias de precios, negociar con proveedores y mejorar la rentabilidad del negocio.

Palabras clave: Inteligencia de negocios, Web Scraping, Aprendizaje automático, Tableros de control, Precios de la canasta básica alimentaria

Abstract

In this Master's Thesis, the development of an innovative business intelligence solution is presented. This solution combines Web Scraping, machine learning techniques, and Power BI dashboards to track the prices of basic food basket items in five retail stores. The objective of this solution is to provide decision-makers in the food sector with an efficient and accurate tool to monitor and analyze product prices in real time.

To conduct this study, a Web Scraping system was implemented to collect price data from the selected retail stores' websites. The collected data were processed using machine learning techniques to identify patterns and trends in product prices over time. Subsequently, interactive dashboards were created in Power BI to visually display the obtained results, allowing users to explore and analyze the data intuitively.

The results show significant variations in the prices of basic food basket items among the selected retail stores. Additionally, there is a general trend of increasing prices for certain products over time. These findings provide a clear insight into price behavior in the market and can be used to make strategic decisions, such as adjusting pricing strategies, negotiating with suppliers, and improving business profitability.

Keywords: Business Intelligence, Web Scraping, Machine learning, Dashboards, Prices of the basic food basket.

Contenido

Pág.	
	Lista de Figuras 11
	Lista de Tablas 12
	Introducción 13
	Objetivos 18
	<i>Objetivo general</i> 18
	<i>Objetivos específicos</i> 18
	Justificación..... 19
	Marco Teórico 21
	<i>Soluciones de inteligencia de negocios y sistemas de preciosError! Bookmark not defined.</i>
	<i>Soluciones de inteligencia de negocios y Web Scraping.....Error! Bookmark not defined.</i>
	<i>Web Scraping para el seguimiento de preciosError! Bookmark not defined.</i>
	<i>Sistemas de información.....Error! Bookmark not defined.</i>
	Hipótesis..... 41
	Variables Error! Bookmark not defined.

Metodología	44
<i>Tipo de investigacion.....</i>	<i>Error! Bookmark not defined.</i>
<i>Población y muestra.....</i>	<i>Error! Bookmark not defined.</i>
<i>Instrumentos de medicion y validacion</i>	<i>Error! Bookmark not defined.</i>
<i>Procedimientos y técnicas para el análisis de la información....</i>	<i>Error! Bookmark not defined.</i>
<i>defined.</i>	
Trabajo de Campo	49
<i>Fases de la investigacion.....</i>	<i>51</i>
<i>Prototipo funcional.....</i>	<i>58</i>
<i>Procesamiento de los datos</i>	<i>58</i>
<i>Análisis de resultados.....</i>	<i>58</i>
<i>Propuesta de solución a la problemática.....</i>	<i>62</i>
Discusión.....	65
Conclusiones y Trabajo Futuro	69
<i>Conclusiones.....</i>	<i>69</i>
<i>Trabajo futuro</i>	<i>70</i>
Referencias.....	Error! Bookmark not defined.

Lista de Figuras

Pág.

Figura 1. Instrumento de Medicion	Error! Bookmark not defined.
Figura 2. Procedimiento y Técnicas de Análisis	44
Figura 3. Pantalla Inicio	49
Figura 4. Retails	Error! Bookmark not defined.
Figura 5. Vista Compare y Ahorre	44
Figura 6. Precios cesta básica alimentaria.....	49
Figura 7. Comparativa de métricas históricas	Error! Bookmark not defined.
Figura 8. Detalle para las jerarquías de productos	44
Figura 9. Pronósticos por tipo de producto – 10 días.....	49

Lista de Tablas

Pág.

Tabla 1. Variables**Error! Bookmark not defined.**

Tabla 2. Muestra Investigación 44

Tabla 3. Fases de la Investigación..... 49

Tabla 4. Prueba de concepto resumen..... 49

Introducción

Las soluciones de inteligencia de negocios pueden ser aplicadas en la gestión de precios, permitiendo a las empresas tomar decisiones más informadas y optimizar su estrategia de fijación de precios. Esta fijación de precios es importante para que las empresas determinen el precio idóneo para ofrecer un producto en el mercado. En particular, las soluciones de inteligencia artificial de negocios son útiles cuando se aplica el método de fijación de precios basado en la competencia, el cual consiste en considerar el precio que los principales competidores le asignan a un producto similar, teniendo en cuenta su presentación y calidad. Por tanto, se da seguimiento a los precios de la competencia con mayor peso en el mercado con el objetivo de determinar los propios (Kotler & Armstrong, 2012). Estas soluciones de inteligencia de negocios pueden aprovechar el Web Scraping, una técnica que permite extraer datos de manera automatizada de sitios web, para recopilar información sobre precios de productos y servicios en línea (Khder, 2021). Su uso incluye el monitoreo de precios, la inteligencia de precios, la investigación de mercado, entre otros. El Web Scraping es legal siempre y cuando no utilice obras protegidas por el derecho de autor o se realice la extracción no consentida de elementos protegidos por otros derechos, como el de la propia imagen o afecte a datos personales. El Web Scraping se ha convertido en una herramienta valiosa para el seguimiento y monitoreo de precios en diversos sectores, proporcionando a las empresas una visión detallada de las fluctuaciones de precios en el mercado. En la era moderna, el Web Scraping juega un papel crucial en campos como el de la inteligencia de negocios (Khder, 2021). El Web Scraping se emplea para recopilar información valiosa y brindar servicios de aplicación de la ley e investigación. El uso de programas de Web

Scraping ha demostrado generar datos más completos, precisos y consistentes que la entrada manual. Por ejemplo, se puede monitorizar la competencia del mercado al permitir controlar, rastrear y generar alertas para saber las actualizaciones de catálogos de productos o servicios de la competencia, o hacer seguimiento a cuando se refieren a los productos propios de una empresa para aprovechar oportunidades de mercadeo u optimizar el *e-commerce*. Así mismo, destacan que en la era de la información, el Web Scraping se ha convertido en una herramienta invaluable y es fundamental en diversos campos en los que se puede aplicar como el gobierno de datos, el Big Data y la inteligencia de negocios (Singrodia et al., 2019), ya que gracias a esta técnica es posible recopilar y compilar datos de millones de páginas para procesarlos y extraer información basados en el análisis de sitios web, el rastreo de sitios web y la organización de datos como procesos principales entrelazados del Scraping en línea que da la posibilidad de convertir datos web no estructurados en datos estructurados que se pueden guardar y analizar en una base de datos central en un corto período de tiempo, lo cual representa una ventaja en el mundo actual cambiante y de constante actualización. Su implementación requiere diferentes tecnologías, como el spidering o Web Crawler las cuales son herramientas digitales basadas en *bots*, que son programas informáticos que actúan de forma autónoma y automática, que pueden optimizar el funcionamiento de los motores de búsqueda de datos en Internet que analizan contenido de manera sistemática, siguiendo enlaces para recopilar información en índices, páginas web y bases de datos con el fin de mejorar el rendimiento de los motores de búsqueda (Singrodia et al., 2019). Sin embargo, es importante destacar las cuestiones éticas y legales relacionadas con el Web Scraping, aunque ofrece numerosos beneficios, es necesario realizarlo de manera responsable y respetando las normativas

vigentes (Ley de Derechos de Privacidad de California, 2023). En cuanto a las soluciones de inteligencia de negocios y sistemas de precios se destaca 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 datos muestran un panorama sobre el comportamiento en las ventas y los precios, además, destaca la importancia de comprender las preferencias y hábitos de consumo de diferentes grupos demográficos (Krummert, 2016).

Su utilidad no sólo radica en su aplicación para el seguimiento de precios, la investigación de mercado y el análisis de tendencias, sino también, como una herramienta para el análisis de proveedores, lo cual puede generar ahorros significativos para las empresas, se pueden identificar discrepancias entre los precios acordados con los proveedores de alimentos y los precios reales que son facturados, para lo que se puede implementar un software de inteligencia de negocios. Esta herramienta permite analizar y comparar los precios de los proveedores, identificar discrepancias y tomar decisiones más informadas en términos de la elección de los proveedores y negociación de los precios. Es importante destacar el uso de herramientas tecnológicas de inteligencia de negocios para optimizar la gestión de proveedores y maximizar la eficiencia en los procesos de adquisición (Financial Managers' Forum, 2005).

Adicionalmente, en la actualidad la invasión de los datos ha sido impulsada por el surgimiento de tecnologías como la Internet (Faramarzi et al., 2022), lo que ha llevado a un aumento exponencial en la generación de datos, conocido como Big Data, campo en el

cual la utilidad del Web Scraping ha sido ampliamente reconocida. Estos avances han creado un. Este fenómeno presenta tanto oportunidades como desafíos para las estadísticas oficiales y la investigación socioeconómica. En su trabajo, los investigadores utilizan la técnica de Web Scraping para extraer los precios diarios de productos alimenticios y bebidas, con el objetivo de reemplazar los precios convencionales utilizados para calcular los índices de precios. Esta técnica ofrece la ventaja de permitir la obtención de datos en escalas de tiempo más pequeñas, como semanal o diaria, en comparación con los enfoques convencionales que se basan en datos mensuales. Aunque el Web Scraping presenta sus propios desafíos, se destaca su eficiencia, precisión y capacidad para ahorrar tiempo, especialmente en áreas urbanas. Antecedentes reportados por el Instituto Brasileiro de Geografía y Estadística (IBGE) revelan que el Web Scraping puede ser una alternativa efectiva a los métodos convencionales para evaluar el comportamiento de los precios de los productos de la canasta básica alimentaria a nivel regional y nacional, concepto que puede ser extrapolado al cálculo del Índice de Precios al Consumidor (IPC) en Colombia y, que también, podría ser aplicado a otras estadísticas de precios. Esta técnica permite aprovechar los nuevos conjuntos de datos para obtener información más actualizada y detallada, lo que contribuye a mejorar la calidad y relevancia de las estadísticas oficiales en comparación con las soluciones comerciales ofrecidas por grandes corporaciones del sector tecnológico (da Silva et al., 2023).

Esta canasta básica familiar comprende un conjunto de bienes y servicios esenciales que son necesarios para satisfacer las necesidades de cualquier hogar o individuo (Censos, n.d.). Esta abarca elementos de primera necesidad, como lo son alimentos y bebidas, hasta artículos como el vestido y el calzado. El problema radica en

que año tras año los precios de la canasta familiar se han vistos afectados por eventos como bloqueos en las vías nacionales, condiciones climáticas adversas e incremento en el precio de los combustibles. No obstante, el costo de la canasta básica familiar también se ve afectado por eventos internacionales (Martínez, 2022), como la crisis de los contenedores, la inflación a nivel global, la pérdida del valor del peso colombiano frente al dólar e incluso, el conflicto entre Rusia y Ucrania (Departamento Nacional de Planeación, 2022). Las fuertes alzas a los bienes y servicios de la canasta familiar afectan a todos los colombianos, en especial a los estratos socioeconómicos más bajos (Gómez, 2023). Para sobrellevar los altos precios de la canasta familiar, muchos hogares se han visto en la necesidad de recurrir a estrategias financieras para sostenerse, prescindiendo de ciertos productos y en algunos casos, a buscar opciones más baratas (ACIS, 2022). En este contexto, el desarrollo de aplicaciones que informen al consumidor de los precios de la canasta básica alimentaria, ofertados por las cadenas de retail, puede servir de guía para los consumidores. Actualmente, existen pocas aplicaciones que les permiten a los usuarios comparar los precios de los productos y servicios con los diferentes oferentes del mercado y/o almacenes de cadena.

De todo lo anterior podemos plantear la siguiente pregunta de investigación:
¿Cómo se estructura una solución de inteligencia de negocios basada en el Web Scraping para el seguimiento de precios de la cesta básica alimentaria que sea útil en los hogares colombianos?

Objetivos

Objetivo general

- Diseñar un prototipo de un sistema de información para el seguimiento de precios a los productos de la canasta alimentaria, a través de Web Scraping, aprendizaje automático y tableros de control

Objetivos específicos

- Emplear las técnicas de Web Scraping como estrategia para la extracción, transformación y carga de los precios de la canasta básica alimentaria, publicados en los portales web de las principales cadenas de retail.
- Implementar una serie de modelos, basados en técnicas de aprendizaje automático, para el análisis de los precios de la canasta básica alimentaria.
- Elaborar una serie de tableros de control a través de los cuales se visualicen los resultados del análisis automatizado de los precios de la cesta básica alimentaria.
- Desarrollar un prototipo web, basado en el framework django, para la solución de inteligencia de negocios propuesta.
- Elaborar una prueba de concepto en usuarios sobre las características funcionales del prototipo

Justificación

De acuerdo con datos recientes publicados por el Departamento Administrativo Nacional de Estadística DANE (2023), se observó un incremento en el Índice de Precios al Consumidor (IPC) en Colombia durante enero de 2023. La variación anual del IPC fue del 13,25%, lo que representa un aumento del 6,31% en comparación con el mismo período del año anterior. Además, en enero de 2023, el IPC experimentó un aumento del 1,78% en comparación con diciembre de 2022. Dentro del IPC, cinco divisiones se ubicaron por encima del promedio nacional (1,78%), siendo la categoría de alimentos y bebidas no alcohólicas la tercera con mayor incremento (2,46%). Como resultado, la inflación ha generado una preocupación significativa entre los hogares colombianos debido a la disminución del poder adquisitivo, lo cual ha limitado la capacidad de adquirir productos básicos para el hogar, especialmente aquellos incluidos en la canasta básica de alimentos.

Aunque el panorama nos es alentador según las cifras previstas para el año 2023, la búsqueda de alternativas para intentar sobrellevar la crisis sigue siendo una opción plausible. La era digital ha traído consigo innumerables ventajas para el ser humano. De hecho, ha crecido el mercado de aplicaciones móviles y plataformas digitales para crear soluciones a diferentes necesidades del mundo moderno (Cortes et al., 2016). Razón por la cual, el presente proyecto tiene como objetivo el desarrollar un prototipo de un sistema de información para el seguimiento de precios a los productos constituyentes de la canasta básica alimentaria, a través de Web Scraping, aprendizaje automático y tableros de control. Con la implementación de este sistema, se espera que

este les permita a los consumidores, autoridades y otros actores interesados, a tomar decisiones informadas y mejorar la transparencia del mercado de alimentos.

Este prototipo contribuirá a reivindicar, reconocer y legitimar los derechos de los consumidores, derechos que se basan en los propuestos por la Superintendencia de Industria y Comercio de Colombia que destaca la importancia de que el consumidor reciba información verídica, con protección contra la publicidad engañosa, con posibilidad de libre elección, procurando la igualdad y participación ciudadana (<https://www.sic.gov.co/>). Adicionalmente, el sistema de información aportará para mejorar la educación para el consumo, a que la población tenga acceso a una información veraz, clara y oportuna de los diferentes proveedores de productos de la cesta familiar, lo cual mejorará la protección de sus intereses económicos, a la vez que, tendrá acceso a los costos de una variedad de productos disponibles básicos para su alimentación cotidiana.

Marco Teórico

Las soluciones de inteligencia de negocios pueden ser aplicadas en la gestión de precios, permitiendo a las empresas tomar decisiones más informadas y optimizar su estrategia de fijación de precios. Estas soluciones de inteligencia de negocios pueden aprovechar el Web Scraping, una técnica que permite extraer datos de manera automatizada de sitios web, para recopilar información sobre precios de productos y servicios en línea. El Web Scraping se ha convertido en una herramienta valiosa para el seguimiento y monitoreo de precios en diversos sectores, proporcionando a las empresas una visión detallada de las fluctuaciones de precios en el mercado.

Soluciones de inteligencia de negocios y sistemas de precios

En el trabajo de Krummert (2016) se abordan los datos sobre el gasto en restaurantes. Mencionan 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.

Por su parte, Financial Managers' Forum (2005) destacan cómo el uso de herramientas para el análisis de proveedores puede generar ahorros significativos para las empresas, como en el caso de Carlson Restaurants Worldwide Inc., operadora de los restaurantes T.G.I. Fridays y Pick Up Stix. La empresa identificó discrepancias entre los precios acordados con sus proveedores de alimentos y los precios reales que les estaban siendo facturados. Ante esta situación, buscaban una solución que les permitiera descubrir este tipo de prácticas engañosas de manera continua y encontrar proveedores más rentables. En el artículo se revela que la solución a este problema fue la implementación de un software de inteligencia de negocios. Esta herramienta les permitió analizar y comparar los precios de los proveedores, identificar discrepancias y tomar decisiones más informadas en términos de la elección de los proveedores y negociación de los precios. Este caso destaca la importancia de utilizar herramientas tecnológicas de inteligencia de negocios para optimizar la gestión de proveedores y maximizar la eficiencia en los procesos de adquisición. Al identificar y abordar estas discrepancias de precios, las empresas pueden lograr importantes ahorros y mejorar su rentabilidad.

Soluciones de inteligencia de negocios y Web Scraping

La propuesta de Khder (2021) resalta la importancia del Web Scraping en la extracción automática de datos de sitios web mediante software. En la era moderna, el Web Scraping juega un papel crucial en campos como el de la inteligencia de negocios. De acuerdo con los autores, esta tecnología permite obtener datos en forma de texto, y resulta

especialmente útil cuando los datos no presentan un formato estructurado. El Web Scraping se utiliza para recopilar precios en tiempo real de tiendas minoristas y proporcionar detalles adicionales. El autor también señala que el Web Scraping sirve para recopilar información valiosa en áreas ilícitas, como los mercados de drogas en la red oscura, y brindar servicios de aplicación de la ley e investigación. El uso de programas de Web Scraping ha demostrado generar datos más completos, precisos y consistentes que la entrada manual. Así mismo, destacan que en la era de la información, el Web Scraping se ha convertido en una herramienta invaluable y es fundamental en diversos campos. Su implementación requiere de librerías que ofrecen lenguajes de programación como Python o Java. Sin embargo, es importante destacar las cuestiones éticas y legales relacionadas con el Web Scraping. Aunque ofrece numerosos beneficios, es necesario realizarlo de manera responsable y respetando las normativas vigentes. Por último, plantean la necesidad de analizar estas cuestiones para garantizar un uso ético y legal de esta tecnología.

En Singrodia resaltan la importancia del Web Scraping 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 al Web Scraping y menciona el software y herramientas utilizadas en esta práctica. Además, exploran el proceso de Web Scraping, 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 del Web Scraping, mostrando el potencial y la versatilidad de esta herramienta en diversos contextos.

Un enfoque interesante para detectar y extraer datos de fuentes no estructuradas se encuentra en el trabajo de Lima et al. (2019). El objetivo de su trabajo 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 de la Web 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

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 por sus siglas en inglés). 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 dNFB 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 (2022) lleva 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 *Bünting*. Es un grupo 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.

Por último, Faramarzi et al. (2022) resaltan la importancia de la invasión de los datos en la era actual, impulsada por el surgimiento de tecnologías como el Internet. Estos avances han generado un aumento exponencial en la generación de datos, conocido como Big Data. Este fenómeno presenta tanto oportunidades como desafíos para las estadísticas oficiales y la investigación socioeconómica. En su trabajo, los investigadores utilizan la

técnica de Web Scraping para extraer los precios diarios de productos alimenticios y bebidas, con el objetivo de reemplazar los precios convencionales utilizados para calcular los índices de precios. Esta técnica ofrece la ventaja de permitir la obtención de datos en escalas de tiempo más pequeñas, como semanal o diaria, en comparación con los enfoques convencionales que se basan en datos mensuales. Aunque el Web Scraping presenta sus propios desafíos, se destaca su eficiencia económica, precisión y capacidad para ahorrar tiempo, especialmente en áreas urbanas. Los hallazgos revelan que el Web Scraping puede ser una alternativa efectiva a los métodos convencionales para el Índice de Precios al Consumidor (IPC) y también se puede aplicar a otras estadísticas de precios. Esta técnica permite aprovechar los nuevos conjuntos de datos para obtener información más actualizada y detallada, lo que contribuye a mejorar la calidad y relevancia de las estadísticas oficiales en comparación con las soluciones comerciales ofrecidas por grandes corporaciones del sector tecnológico.

Sistemas de información

Management information systems o sistemas de información (MIS) han ganado relevancia a medida que las empresas han adoptado la práctica de llevar registros contables para dar seguimiento a sus procesos. Fue a partir de la década de los noventa, momento en el que las computadoras se hicieron más accesibles, cuando los sistemas de información se ampliaron a las áreas comerciales. Estos MIS contribuyeron a la organización de sistemas de inventario, la implementación de sistemas de ventas, la aplicación de los principios de marketing y al seguimiento de procesos internos. Los MIS al principio fueron independientes y funcionaban con software distinto para cada uno (Laudon, 2022), sin

embargo, actualmente han evolucionado hacia la aplicación de nuevas tecnologías que han robustecido su utilidad y campos de aplicación.

Los MIS se originan a partir de conceptos fundamentales de la Teoría General de Sistemas (Van Gigch, 1987; Bertoglio & Johansen, 1982), que se fundamenta en la necesidad de conocer el sistema organizacional sobre el cual estos funcionan, con el fin de lograr un aumento en la eficiencia de los procedimientos que ejecuta la empresa.

Estas son herramientas que las empresas emplean con el fin gestionar y respaldar sus procesos de planificación y operaciones a través de los datos. Estos sistemas se componen de elementos clave, como hardware, software, datos y servidores, que trabajan en conjunto para garantizar el funcionamiento eficiente de la organización (Marakas & OBrien, 2006; James & George, 2003). Son conjuntos que incluyen elementos que proveen a los directivos información actualizada que les permite tomar decisiones, supervisar el desempeño y administrar las operaciones de manera eficiente. Algunos ejemplos comunes de MIS son: (1) sistemas basados en CRM (*Customer Relationship Management*), (2) sistemas de SCM (*Supply Chain Management*), (3) sistemas de ERP (Enterprise Resource Planning), y (4) almacenes de datos para almacenar y gestionar grandes volúmenes de información (Proaño, et al., 2018).

Management Changes: Con los nuevos celulares, redes Wi-Fi y dispositivos móviles, los empleados están muy cerca de la supervisión de sus gerentes. La administración se está volviendo móvil. Los gerentes actuales se comunican constantemente con sus empleados. El desarrollo de los MIS en toda la empresa con datos útiles significa que los líderes no poseen confusiones en sus decisiones, sino que tienen la

información en el momento cuando realmente la necesitan para tomar esas decisiones (Laudon, 2022).

Objetivos de los sistemas de información

- ***Excelencia operacional:*** La excelencia operativa es un objetivo constante para las empresas, ya que su objetivo es mejorar los procesos y conseguir una rentabilidad mejor. En este sentido, los sistemas de información son fundamentales que los gerentes utilizan con el propósito de alcanzar los objetivos establecidos en la estrategia y planificación de la empresa. En los sistemas actuales, existe una creciente interdependencia entre los MIS y sus capacidades comerciales. Los cambios con el fin de implementar estrategias, la forma y la estructura requiere adaptaciones en hardware, software, bases de datos y telecomunicaciones. Con frecuencia, la capacidad funcional de la organización depende de lo que sus sistemas le permitan realizar. (Laudon, 2022).
- ***Innovaciones en productos, servicios y enfoques empresariales:*** Las herramientas y recursos tecnológicos utilizados para gestionar y procesar la información, desempeñan un papel fundamental al brindar a las empresas la capacidad de generar servicios y productos innovadores. (Laudon, 2022).
- ***Establecer relaciones cercanas con clientes y proveedores:*** La comprensión de los patrones de comportamiento de los clientes de una empresa, abre la oportunidad de que estos respondan de manera favorable con fidelidad en la compra o adquisición de servicios. Esto conduce a un incremento en los ingresos y utilidades. De manera similar, al establecer una estrecha colaboración

con los proveedores, una empresa tiene la capacidad de obtener insumos vitales de manera más eficiente, lo que a su vez reduce los costos. Por tal razón, se deben implementar estrategias para conocer a los clientes o proveedores, tanto los que se encuentran en línea como los que están fuera de línea (Laudon, 2022).

- ***Toma de decisiones mejorada:*** Los gerentes de negocios se enfrentan a un desafío, ya que tienen acceso a una amplia cantidad de información, pero puede que no siempre sea la información correcta en el momento oportuno para tomar decisiones informadas. Razón por la cual, los gerentes confían en los pronósticos, y las mejores conjeturas para tal fin. Como resultado, se producen situaciones en las que hay una sobreabundancia o escasez, una partición ineficiente en recursos y retrasos en los lapsos de respuesta. Estos resultados adversos incrementan los costos y hacen perder clientes. Durante los últimos diez años, los avances en MIS han permitido a los directivos utilizar información en tiempo real del mercado para la toma de decisiones. (Laudon, 2022).
- ***Ventaja competitiva:*** Cuando las empresas alcanzan uno o más de los siguientes objetivos empresariales: Es probable que se logre una ventaja competitiva al destacar en la eficacia de sus operaciones, la creación de productos, servicios y modelos de negocio, el establecimiento de relaciones cercanas con clientes y proveedores, y un desarrollo positivo en la toma de decisiones. Cuando una empresa logra realizar sus operaciones de manera más eficiente que sus competidores, ofrecer productos de mayor calidad a un costo inferior y responder de manera rápida a las necesidades de sus clientes y proveedores, esto

se traduce en un incremento en las ventas y ganancias que sus competidores no pueden igualar. (Laudon, 2022).

- ***Supervivencia:*** Las empresas comerciales destinan recursos a MIS con el fin de sobrevivir, ya que son indispensables para llevar a cabo sus operaciones. En ocasiones, estas "necesidades" surgen debido a cambios en la industria. Las empresas utilizan esta herramienta para poder responder a los requisitos de gestión de registros que surgen como resultado de estos cambios. (Laudon, 2022).

Características de los MIS

Son características inherentes a los MIS: (1) relevancia: se refiere a que la información que se genere debe ser relevante para la empresa, y este proceso debe hacerse a tiempo y de manera confiable por lo que ésta información debe responder a los requisitos de gestión y operación de la empresa; (2) integración: es la articulación entre el sistema de información y la estructura de la empresa, lo cual agiliza la toma de decisiones, ya que facilita coordinar los departamentos, sectores, divisiones y otras unidades de organización de las empresas. (3) flujo independiente: es la paralelización de trabajo de la información en la cual, a la vez que hay un flujo de procesamiento de datos, tanto interno como externo, también hay un flujo independiente de los sistemas de información; (4) control: son herramientas de control interno que pueden tener los sistemas de información, que aseguran la confiabilidad de los datos y la protección de estos; (5) directrices: garantizan que los objetivos de la empresa se cumplan objetiva y eficientemente (Laudon & Laudon, 2012).

Tipos

Hay cuatro categorías principales de sistemas de información que se emplean para generar informes, extraer datos y facilitar la toma de decisiones por parte de los gerentes de niveles medios y operativos.

- ***Transaction Processing System (TPS)***: Un TPS suele estar en la base del sistema de información integrado de cualquier empresa. Adquiere datos sin procesar de diversas fuentes tanto dentro como fuera de la organización y los guarda en una base de datos. Los usuarios pueden consultar la base de datos para cualquier información que necesiten. Los datos se pueden procesar en tiempo real a medida que se ingresan, o mediante el procesamiento por lotes, en cuyo caso la información se procesa en momentos específicos del día (Proaño, et al., 2018; Laudon & Laudon, 2022).
- ***Decision Support Systems (DSS)***: Los sistemas de soporte de decisiones ayudan a los gerentes a tomar decisiones informadas utilizando modelos informáticos interactivos que exploran los procesos del mundo real. El DSS recupera datos específicos de la base de datos interna que pertenecen a las preguntas planteadas por el usuario (Proaño, et al., 2018; Laudon & Laudon, 2022).
- ***Executive Information Systems (EIS)***: Un sistema de información ejecutiva opera de manera similar a un Sistema de Apoyo a la Toma de Decisiones (DSS, por sus 15 siglas en inglés), pero su software ha sido creado con el propósito de atender de manera específica las demandas y requerimientos de ejecutivos individuales, permitiéndoles tomar decisiones estratégicas fundamentadas. Estos sistemas son capaces de ofrecer datos detallados cuando sea necesario. Los

sistemas de información ejecutiva recopilan información crítica de diversas fuentes internas y externas, presentándola de manera más simplificada para los ejecutivos y gerentes. (Proaño, et al., 2018; Laudon & Laudon, 2022).

- ***Enterprise resource planning (ERP)***: Se trata de un conjunto de aplicaciones o módulos empresariales que están interconectados y emplean una base de datos compartida y centralizada. Cada módulo se centra en una función particular del negocio, pero colaboran entre sí utilizando los mismos datos para cumplir con los requerimientos de la empresa. Las áreas típicas donde se implementa un ERP incluyen finanzas, contabilidad, recursos humanos, ventas, compras, logística y cadena de suministro. Las empresas tienen la flexibilidad de seleccionar los módulos que necesiten y agregar o expandir según sea necesario. También, tienen la capacidad de ajustarse a requisitos particulares de cada industria, ya sea a través de sus características principales o mediante la incorporación de extensiones de aplicaciones que se integran perfectamente con el conjunto de herramientas. El software de ERP puede adquirirse a través de modelos de suscripción en la nube o mediante licencias tradicionales. (Proaño, et al., 2018; Laudon & Laudon, 2022).
- ***TI Innovation***: El continuo desarrollo de innovaciones y nuevas tecnologías de la información está evolucionando el desarrollo de las organizaciones. Los ejemplos incluyen el desarrollo de habilidades en computación en la nube, el aumento de una plataforma comercial digital basada en dispositivos móviles, el análisis de datos, la conexión de dispositivos a través de internet, el análisis de negocios, los sistemas de aprendizaje automático y el uso de redes sociales por

parte de los gerentes para alcanzar los objetivos de negocio. Estas innovaciones están permitiendo a las empresas el poder innovar creando servicios y productos nuevos. (Laudon & Laudon, 2022).

- ***New Business Models:*** La aparición de servicios de películas en línea para transmisión o descarga, ha cambiado para siempre la forma en que se distribuye, desafiando a los productores de televisión y de transmisión. (Laudon & Laudon, 2022).
- ***E-commerce Expansion:*** El comercio electrónico está cambiando la forma en que las empresas trabajan hoy en día. El comercio electrónico ha evolucionado continuamente, alterando la industria de marketing, poniendo en peligro a las empresas que operan en medios y contenido. (Laudon & Laudon, 2022).
- ***Management Changes:*** La gestión de las empresas comerciales ha experimentado cambios significativos. Gracias a los avances tecnológicos como los teléfonos móviles, las conexiones inalámbricas, vendedores en movimiento se encuentran a solo unos minutos de distancia de realizar consultas y recibir la guía de sus líderes. La administración se está volviendo móvil. Los gerentes en constante movimiento mantienen una comunicación directa y constante con los trabajadores. El aumento de los MIS en toda la organización, que contienen datos sumamente detallados, implica que los gerentes ya no se encuentran perdidos en la incertidumbre, sino que tienen acceso en tiempo real a la información verdaderamente relevante que les permite tomar decisiones precisas y oportunas. (Laudon & Laudon, 2022).

- ***Changes in Firms and Organizations:*** En contraste con las pasadas organizaciones, las nuevas empresas dan una importancia menor a la estructura, y priorizan que sus trabajadores desempeñen múltiples puestos dentro de la organización, colaborando en equipos. Además, ponen mayor énfasis en la competencia y habilidades de los individuos, en lugar de su posición en la jerarquía. Hacen hincapié en la toma de decisiones más rápida y más precisa basada en análisis de datos. (Laudon & Laudon, 2022).

Definición de Canasta Básica Alimentaria

La Canasta Básica Alimentaria (CBA) se basa en una lista de alimentos que se consideran esenciales para mantener una alimentación saludable y equilibrada. Esta canasta puede variar según el país o región. Por lo general, la CBA incluye alimentos como arroz, pan, leche, huevos, carne, frutas, verduras, legumbres, aceites, entre otros. El costo de la CBA se utiliza a menudo como una medida de la pobreza en un país o región. En algunos lugares, los gobiernos establecen un salario mínimo que debería ser suficiente para que una familia pueda adquirir una CBA. Es importante destacar que la CBA es solo una guía, y que las necesidades nutricionales pueden variar de una persona a otra. Además, el costo de la CBA puede variar según los precios de los alimentos y otros factores, por lo que puede ser necesario actualizarla regularmente (Menchú, et al., 2006). Sobre este último punto, el documento de CEPAL/SG-SICA (2021) formula las directrices metodológicas para la elaboración de canastas básicas alimenticias. En él se aborda la importancia de las canastas básicas de alimentos como elemento fundamental en la medición de la pobreza por ingresos, estaca su utilidad para el diseño de políticas públicas para mejorar las condiciones de vida y reducir las desigualdades en la región de

Centroamérica. En el documento se propuso un marco metodológico común para la construcción de estas canastas, con el objetivo de lograr una mayor armonización estadística y comparabilidad de los resultados entre los países miembros del Sistema de la Integración Centroamericana (SICA). Como parte del enfoque metodológico propuesto por el SICA, se realizó un minucioso registro de los productos adquiridos por los residentes de distintos países de América Latina, abarcando un amplio inventario. Adicionalmente, tomaron en cuenta elementos como el desembolso efectuado al comprar los productos, la cantidad obtenida y la contribución de calorías y nutrientes. El número de productos del listado varió entre las diversas encuestas analizadas, pero se clasificaron en grupos alimentarios para simplificar la construcción de la canasta. En la selección de los productos representativos de los hábitos de la población y coherentes con la noción de canasta básica, emplearon criterios independientes relacionados con la universalidad del consumo, el aporte energético y la participación en el gasto. En este mismo sentido García (2022) desarrolla una definición de canastas básicas de consumo en distintos países. Estas canastas incluyen una variedad de bienes y servicios agrupados en categorías como alimentos, ropa, vivienda, salud, transporte, educación, esparcimiento, entre otros.

En Chile, en su Instituto de Estadísticas (INE) emplea un conjunto de productos y servicios que reflejan el consumo de los hogares en el país, como base para calcular el Índice de Precios al Consumidor (IPC) En Argentina, se utiliza la Canasta Básica Alimentaria (CBA) como una medida de referencia para establecer el nivel de pobreza. Esta canasta se crea de manera regional, considerando los hábitos culturales de los hogares con ingresos similares en términos de poder adquisitivo. En Perú, el Instituto Nacional de Estadística e Informática (INEI) establece una Canasta básica de consumo que incluye

tanto alimentos como productos no alimentarios y está dividida en 12 categorías (García, 2022).

En México, se establece una canasta básica que consta de 40 productos, los cuales no se limitan exclusivamente a alimentos, y se emplea para determinar el salario mínimo y supervisar la variación de precios de dichos productos. Adicionalmente, el Consejo Nacional de Evaluación de la Política de Desarrollo Social establece una canasta básica con el propósito de actualizar de manera mensual el umbral de la pobreza basado en los ingresos. En el contexto nacional, la Asociación Colombiana de Educación al Consumidor reflexiona sobre la problemática nutricional en Colombia. considera fenómenos como la desnutrición, el sobrepeso, la obesidad y la mortalidad asociada a estas condiciones. Destacan que la Canasta Básica de Alimentos (CBA) tradicional no es coherente con las 20 estrategias de salud pública encargadas de abordar esta problemática. Así mismo, señalan que el creciente consumo de productos ultraprocesados es un factor que agrava la situación (Parra et al. 2018).

Así mismo, Parra, et al. (2018) ha generalizado una falta de incentivos económicos para la producción, distribución y consumo de alimentos no procesados o mínimamente procesados provenientes de economías campesinas. Sin embargo, este panorama puede cambiar, ya que el gobierno actualmente ha logrado la aprobación del gravamen con impuestos de los alimentos ultra procesados, lo cual ha implicado una amplia regulación de símbolos y publicidad de este tipo de productos. Esto ha sido impulsado debido a que, en diferentes análisis sobre el tema, se encontraba que los colombianos habían cambiado sus hábitos alimentarios tradicionales por productos ultraprocesados con alto contenido energético y bajo aporte nutricional, y que estos productos están ampliamente disponibles

y promovidos a través de publicidad engañosa y con poca regulación. Esto llevaba a una crítica profunda sobre la política alimentaria del gobierno colombiano, que ha permitido y promovido el aumento del consumo de productos ultraprocesados y no había implementado acciones contundentes por fomentar la producción y distribución de alimentos que sean beneficiosos para la salud. Se ha destacado la falta de protección, promoción y apoyo a la lactancia materna como un factor que amplía la problemática. Finalmente, en análisis previos se ha concluido que Colombia necesita una CBA saludable y asequible para toda la población, que el país se convierta en un modelo educativo de alimentación saludable. La metodología utilizada por la Asociación para determinar una Canasta Básica de Alimentos Saludable en Colombia se basa en la propuesta del INCAP, pero con algunos cambios. Entre estos, destacan que la metodología clásica de determinación de la CBA no considera parámetros de salud, por lo que se puede incluir productos procesados y ultraprocesados que se han relacionado con la aparición de enfermedades no transmisibles. Se advierte que aquellos colombianos que se alimentan de la CBA convencional no están asegurando una dieta saludable. La literatura ha establecido las necesidades energéticas y las recomendaciones de nutrientes para los diferentes grupos de edad. Para ello, utilizaron las Guías Alimentarias para la Población Colombiana y el Modelo de Perfil de Nutrientes de la OPS de 2016 como referencia. La selección de alimentos se basó en productos ampliamente distribuidos y reconocidos en la mayoría del territorio colombiano. Así mismo, evaluaron el contenido nutricional de los alimentos seleccionados y generaron una propuesta por intercambios de alimentos para tres comidas principales y dos refrigerios. Utilizaron la tabla de porciones y medidas de las GABAS para transformar los resultados en alimentos con porciones y tamaños reales. También se

determinó el costo promedio de la canasta básica utilizando un catálogo de precios al consumidor de la mayor central de abastecimiento de Colombia (CORABASTOS) con los precios vigentes en la semana del 12 al 16 de junio de 2017. Como resultado, proponen una CBA saludable compuesta por 59 productos (Parra, et al., 2018).

Web Scraping

El Web Scraping es una metodología de análisis de datos que implica la extracción de información específica de páginas web (Mijangos, 2023). Esta técnica simula la navegación llevada a cabo por un usuario en internet. El Scraping puede realizarse de dos maneras, manualmente y de forma automatizada, la cual suele ser la más eficiente. El Web Scraping se hace a través de un web scraper, el cual es un programa de software que rastrea, recupera y estructura la información almacenada en etiquetas HTML, para luego ser guardada en archivos CSV, los cuales posteriormente podrán ser almacenados y analizados en bases de datos (Mijangos et al., 2022). En la actualidad, el Web Scraping es usado en una amplia variedad de campos y aplicaciones, desde la investigación académica hasta la inteligencia de negocios. Uno de los usos más comunes que se pueden encontrar en internet sobre el Web Scraping es el monitoreo de precios, en donde, por un lado, los comerciantes usan esta tecnología para rastrear y comparar los precios de la competencia, para así ajustar sus precios y ofertas. Y por el otro, los clientes usan esta tecnología para consultar que tienda ofrece los mejores productos en términos de calidad, cantidad y precio, y así realizar compras más inteligentes. El proceso de Web Scraping comienza por la identificación del URL de la página o URL semilla de la cual se extraerá información. Seguidamente, se hará uso de alguna herramienta de Scraping como *Selenium*, *Scrapy* o *Beautiful Soup*, las cuales harán una petición a la página en donde se obtendrá una

respuesta en código HTML. Finalmente, el programa hace un análisis en búsqueda de la información que se quiera obtener, la cual podrá ser almacenada en un formato estructurado para su posterior uso con alguna base de datos (García et al., 2023).

Hipótesis

La presente investigación busca demostrar que: Es factible desarrollar e implementar una plataforma de inteligencia de negocios que permita monitorear y analizar los precios de la canasta básica alimentaria, a través de Web Scraping, utilizando como fuente la información la proporcionada por los portales web de las cadenas de retail. Se espera que este sistema permita a los consumidores, autoridades y otros actores interesados, tomar decisiones informadas y mejorar la transparencia del mercado de alimentos

Modelo de Base de Datos

En el contexto de este estudio, la información recopilada de las cinco tiendas minoristas seleccionadas no se almacenará en una base de datos convencional. En lugar de ello, se ha optado por organizar y conservar los datos diarios en archivos CSV independientes. Esta elección se fundamenta en la flexibilidad y accesibilidad que proporciona el formato de hoja de cálculo, permitiendo una manipulación más ágil de la información recolectada. Cada archivo CSV contendrá registros detallados de las transacciones diarias, incluyendo datos esenciales como la tienda correspondiente, la categoría de productos, la descripción detallada de los artículos, los precios asociados y la fecha de la transacción.

Con miras a facilitar el análisis integral de la información, se implementará un proceso de fusión de estos archivos individuales en un único archivo consolidado. Este archivo consolidado adoptará la forma de una tabla única que integrará de manera sistemática los campos relevantes, a saber: tienda, categoría, producto, descripción, precio y fecha. Este enfoque no solo permitirá una gestión más eficiente de los datos acumulados,

sino que también facilitará la realización de análisis y evaluaciones comparativas entre las distintas tiendas y períodos de tiempo, contribuyendo así a los objetivos específicos de este trabajo académico.

Variables

La *tabla 1* presenta información detallada sobre un proceso de Web Scraping aplicado a portales web de *retail*. Las variables extraídas incluyen la tienda, la fecha de obtención de datos, la descripción del producto, el precio del producto, la etiqueta asociada y la categoría del producto. La definición operativa y el tipo de variable se detallan para cada una. La tienda representa el nombre del establecimiento, la fecha indica cuándo se recopiló la información, la descripción proporciona un texto descriptivo del producto, el precio es un valor numérico, la etiqueta es una referencia o código, y la categoría clasifica el producto en grupos según sus características. Estas variables permiten un análisis exhaustivo de la información recopilada, brindando una visión completa de la oferta de productos en los portales web de *retail*.

Tabla 1

Variables

Variable	Definición operativa	Tipo de variable
Tienda	Nombre de la tienda desde la cual se extraen los datos.	Catagórica Nominal
Fecha	Fecha en la que se obtuvo la información.	Temporal
Descripción	Texto descriptivo que identifica y caracteriza al producto.	Catagórica Nominal
Precio	Valor numérico que representa el costo del producto.	Numérica Continua
Etiqueta	Etiqueta asociada al producto, puede ser una referencia, código o etiqueta de identificación.	Catagórica Nominal
Categoría	Clasificación que agrupa productos similares por sus características o uso.	Catagórica Nominal

Metodología

A continuación, se define el enfoque, diseño y tipo de investigación. Se describen las fases que le permitan alcanzar cada uno de los objetivos de la investigación. Además, se determina la población y muestra. Por último, se diseña el instrumento de medición y se plantean los procedimientos y técnicas de análisis de la información

Tipo de investigación

Este trabajo se clasifica como una investigación aplicada, ya que busca una solución que tenga aplicabilidad en la población colombiana a corto, mediano y largo plazo. Adicionalmente, tiene un enfoque exploratorio cuantitativo longitudinal, teniendo en cuenta que se utilizarán datos del mercado que deberán actualizarse permanentemente debido a la fluctuación de precios en el mercado. Además, es una investigación inductiva, ya que parte de un hecho concreto de la recopilación de los costos de los productos básicos de la canasta familiar y la necesidad de optimizar los recursos financieros, y así, formular un razonamiento o procedimiento de selección útil para la población objetivo.

Población y muestra de la investigación

Para el desarrollo de la técnica del Web Scraping se escogieron las páginas web de las siguientes cadenas de *retail*: Éxito, Jumbo, D1, Makro, Carulla en idioma español. Ver *tabla 2*. Estos datos se utilizarán para llevar a cabo un análisis de costos de los productos previamente seleccionados para la canasta familiar

Tabla 2

Muestra Investigación

Retail	URL
Éxito	https://www.exito.com/home-mercado

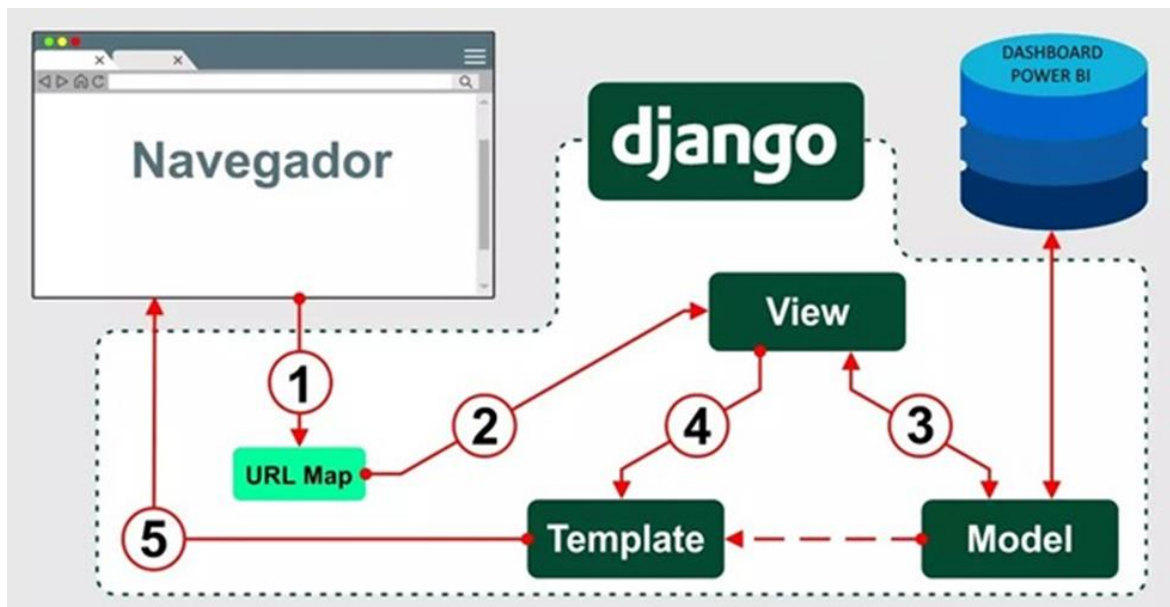
Jumbo	https://www.tiendasjumbo.co/
D1	https://domicilios.tiendasd1.com/
Makro	https://tienda.makro.com.co/
Carulla	https://www.carulla.com/

Instrumento de medición y validación

La *Figura 1* muestra la arquitectura de la solución de inteligencia de negocios propuesta:

Figura 1

Instrumento de Medición, Creación Propia



Nota. Elaboración propia

En la *Figura 1* se puede apreciar el funcionamiento de framework Django. Su funcionamiento se basa en el patrón de diseño Modelo-Vista-Controlador (MVC), donde los modelos representan la estructura de datos, las vistas definen la presentación y la lógica de negocio, y los controladores gestionan las solicitudes del usuario.

Procedimientos y técnicas de análisis de la información

La realización del Web Scraping o raspado web implica una serie de procedimientos y técnicas fundamentales para extraer datos de sitios web de manera efectiva. En este proyecto específico, que se centra en obtener información de cinco cadenas de supermercados, se siguen los siguientes pasos clave, ver *Figura 2*:

Identificación del Sitio Web: El primer paso es identificar los sitios web de interés. En este caso, se enfoca en las páginas web de las cadenas de supermercados mencionadas. La identificación precisa de estos sitios es crucial para el éxito del proceso de Web Scraping.

Identificación de la Información Por Extraer: Una vez que se han seleccionado los sitios web objetivo, es necesario determinar qué información se va a extraer de ellos. En este proyecto, se busca obtener datos numéricos relacionados con los precios de los productos y cadenas de texto que representen los nombres de los productos. Esta etapa requiere un análisis detallado de las páginas web para identificar los elementos específicos que contienen esta información.

Identificación de la Estructura del Sitio Web: Para extraer datos de manera efectiva, es esencial comprender la estructura de las páginas web seleccionadas. Esto implica analizar el código HTML de las páginas y examinar las etiquetas relevantes que contienen los datos de interés. Esta fase permite definir cómo se accederá y extraerá la información de las páginas web.

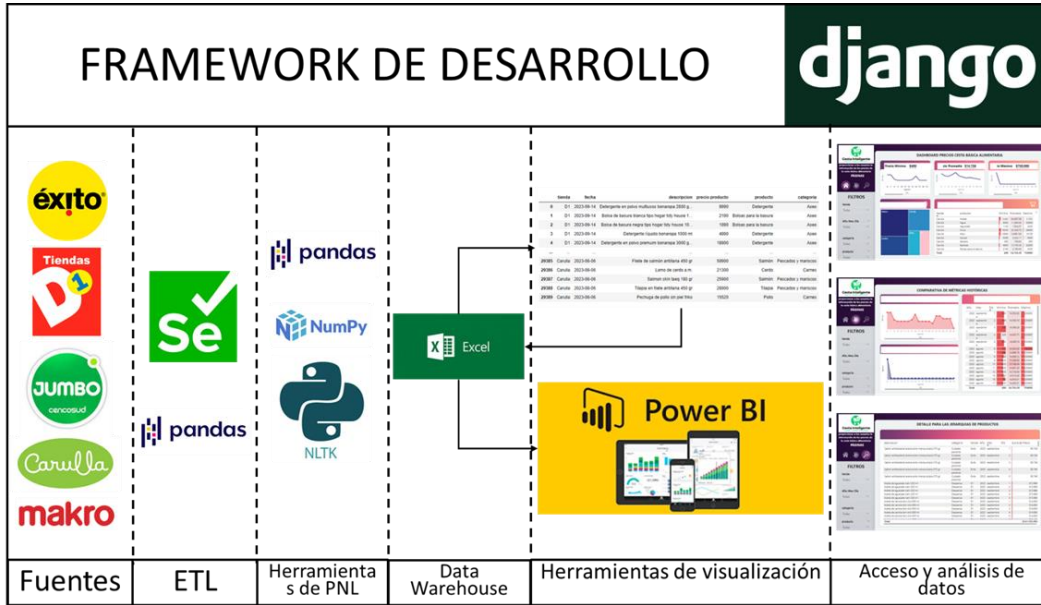
Selección de la Librería de Scraping Web: En el proceso de Web Scraping, existen diversas bibliotecas y herramientas disponibles para automatizar la extracción de datos. En este proyecto, se opta por utilizar Selenium, una biblioteca conocida por su facilidad de uso, su curva de aprendizaje amigable, 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 almacenan en archivos CSV. La elección de este tipo de formato se debe a su capacidad para organizar y gestionar eficazmente los datos recopilados. Finalmente estos archivos se fusionan en un único archivo de Excel, el cual se constituye en la base de datos para la futura etapa de análisis.

Análisis de la Información Extraída: Una vez que los datos se han extraído y almacenado en la base de datos, se convierten en información útil para su análisis posterior. En este proyecto se utiliza el framework de Python Django para crear un sitio Web que permita desplegar tableros de control construidos en Power BI, a través de los cuales se presentara de manera clara y accesible la información recopilada. Estos tableros de control facilitan la interpretación de los datos y la toma de decisiones basadas en ellos.

La *Figura 2* muestra el esquema correspondiente a los procedimientos y técnicas de análisis de la información:

Figura 2. Procedimiento y Técnicas de Análisis



Trabajo de Campo

Fases de la investigación

A continuación, se describen las fases que permitan alcanzar cada uno de los objetivos de esta investigación.

Tabla 3

Fases de la Investigación

Objetivo	Fases	Actividades
Emplear las técnicas de Web Scraping como estrategia para la extracción, transformación y carga de los precios de la canasta básica alimentaria, publicados en los portales web de las principales cadenas de <i>retail</i> .	Selección de los portales web a los que se accederá para el raspado de datos.	Identificación de los portales web de las principales cadenas de retail. Selección de las herramientas de Web Scraping adecuadas.
	Desarrollo del proceso de Web Scraping.	Configuración de las herramientas de Web Scraping. Extracción, transformación y carga de los datos.
	Evaluación del proceso de Web Scraping.	Evaluación del desempeño del proceso de Web Scraping. Identificación de las mejoras necesarias para el proceso. Ajuste del proceso según los resultados de la evaluación.
Desarrollar una variedad de modelos utilizando técnicas de aprendizaje automático precios de la canasta básica de alimentos.	Selección de las técnicas de aprendizaje automático adecuadas.	Identificación de las técnicas de aprendizaje automático que se aplicarán. Selección de las herramientas de aprendizaje automático necesarias.

	Desarrollo de los modelos de aprendizaje automático.	Configuración de las herramientas de aprendizaje automático. Entrenamiento y ajuste de los modelos. Validación de los modelos.
	Evaluación de los modelos de aprendizaje automático.	Evaluación del desempeño de los modelos. Identificación de las mejoras necesarias para los modelos. Ajuste de los modelos según los resultados de la evaluación.
Elaborar una serie de tableros de control a través de los cuales se visualicen los resultados del análisis automatizado de los precios de la canasta básica alimentaria.	Selección de las herramientas de visualización adecuadas.	Identificación de las herramientas de visualización que se utilizarán. Selección de las herramientas de visualización necesarias.
	Desarrollo de los tableros de control.	Configuración de las herramientas de visualización. Desarrollo de los tableros de control. Pruebas de los tableros de control y corrección de errores.
	Evaluación de los tableros de control.	Evaluación del desempeño de los tableros de control. Identificación de las mejoras necesarias para los tableros. Ajuste de los tableros según los resultados de la evaluación.

Prototipo funcional

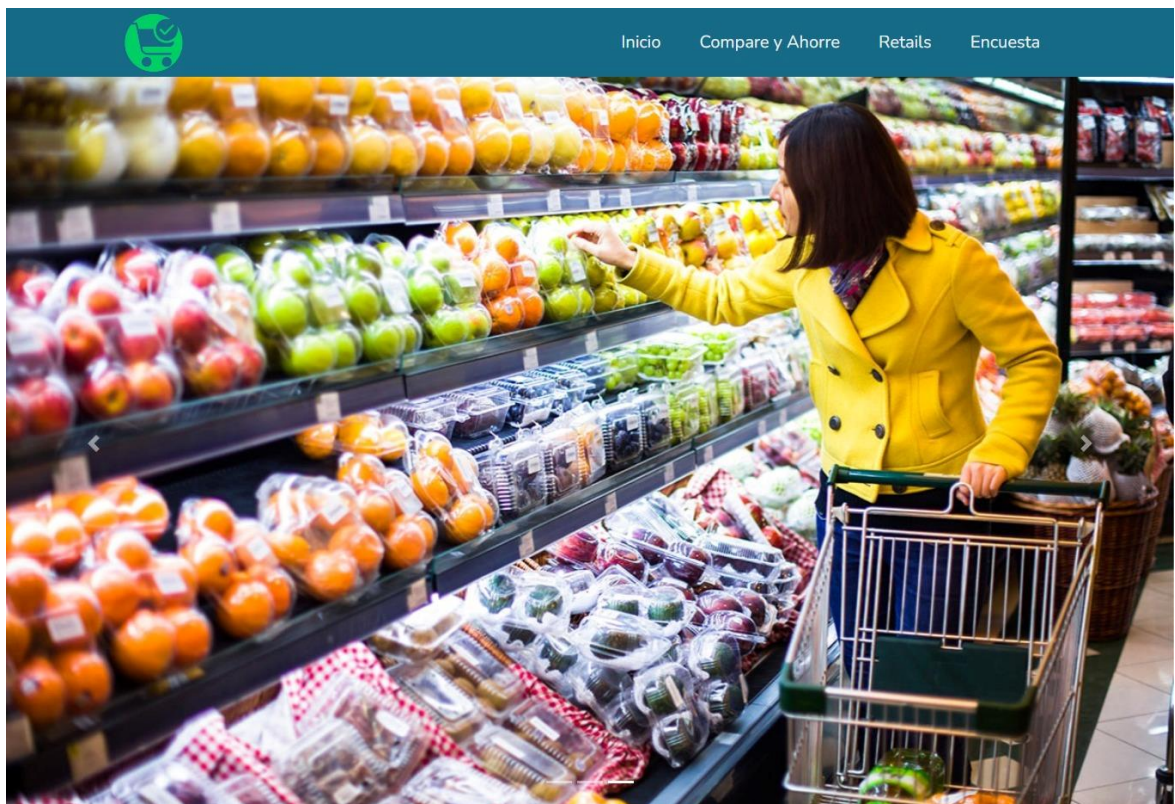
A continuación, se muestra un prototipo del instrumento de medición (prototipo). Este se compone de una vista inicial, en donde el usuario indica el producto de interés. La segunda vista corresponde al tablero de control a través de la cual se obtienen los principales indicadores del producto seleccionado.

Vista 1. Pantalla inicio

En la *Figura 3* se puede apreciar la página de inicio del prototipo. En la parte superior derecha, se encuentra una *nav bar* que contiene las secciones de Compare y Ahorra, *Retails*, *About* y *Contact*.

Figura 3

Pantalla Inicio



Nota. Tomada del prototipo desarrollado

Vista 2. Retails

En la *Figura 4* se puede apreciar se puede apreciar la lista de *Retails* o Tiendas de las cuales haremos el Web Scraping junto con su respectivo logo y una pequeña descripción de cada una.

Figura4

Retails



The image shows a screenshot of a website interface. At the top, there is a dark blue navigation bar with a shopping cart icon on the left and the text 'Inicio', 'Compare y Ahorre', 'Retails', and 'Encuesta' on the right. Below the navigation bar, there are five retail store logos arranged in two rows. Each logo is accompanied by its name and a short description of the store. The stores listed are Éxito, D1, Jumbo, Carulla, and Makro. Each description includes a brief overview of the store's operations and a 'Read more' link.

Éxito
El Grupo Éxito es una influyente cadena de tiendas minoristas en América Latina, ofreciendo una variedad completa de productos, desde alimentos hasta moda y tecnología, destacándose por calidad y accesibilidad. [Read more](#)

D1
Tiendas D1 es una cadena de tiendas de descuento tipo hard discount en Colombia. D1, es la compañía con más puntos de venta en Colombia, alcanza 2.000 tiendas, cubriendo el 87% del territorio, en 28 departamentos y 450 municipios. [Read more](#)

Jumbo
Jumbo es una cadena de hipermercados chilena, con presencia también en Argentina y Colombia. Es propiedad del grupo Cencosud. Destaca como líder en el comercio minorista en América Latina. [Read more](#)

Carulla
Carulla es una cadena de supermercados colombiana, fundada en 1905 por Alejandro Carulla Vila. En 2013, Carulla contaba con 175 puntos de venta en Colombia, de los cuales 167 son supermercados y 8 son tiendas de conveniencia. [Read more](#)

Makro
Makro es una cadena de supermercados mayoristas de origen holandés, fundada en 1968. En la actualidad Makro tiene abiertas más de 700 tiendas en tres continentes; esto lo convierte en la empresa de ventas al por mayor más grande del mundo y es el tercer grupo mundial en distribución. [Read more](#)

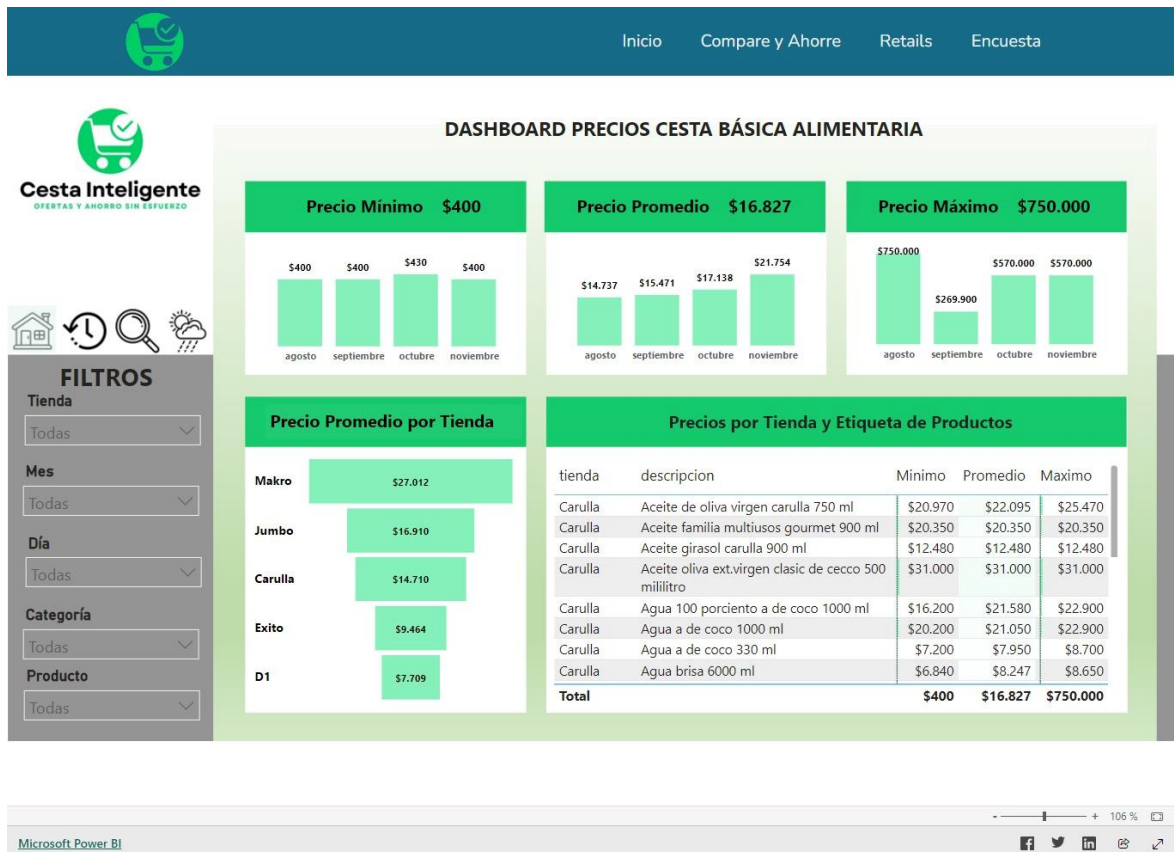
Nota. Tomada del prototipo desarrollado

Vista 3. Compare y ahorre

En la *Figura 5* se puede apreciar el *dashboard* del prototipo, el cual permitirá a los usuarios finales aplicar una serie de filtros a sus búsquedas de acuerdo con sus necesidades.

Figura 5.

Vista Compare y Ahorre



Nota. Tomada del prototipo desarrollado

Procesamiento de los datos

El dashboard en Power BI ha sido diseñado para brindar una visión integral y detallada del sistema de información para el seguimiento de precios en la canasta alimentaria. Cada vista ha sido cuidadosamente desarrollada para ofrecer información específica y valiosa a los usuarios, proporcionando herramientas efectivas para la toma de decisiones.

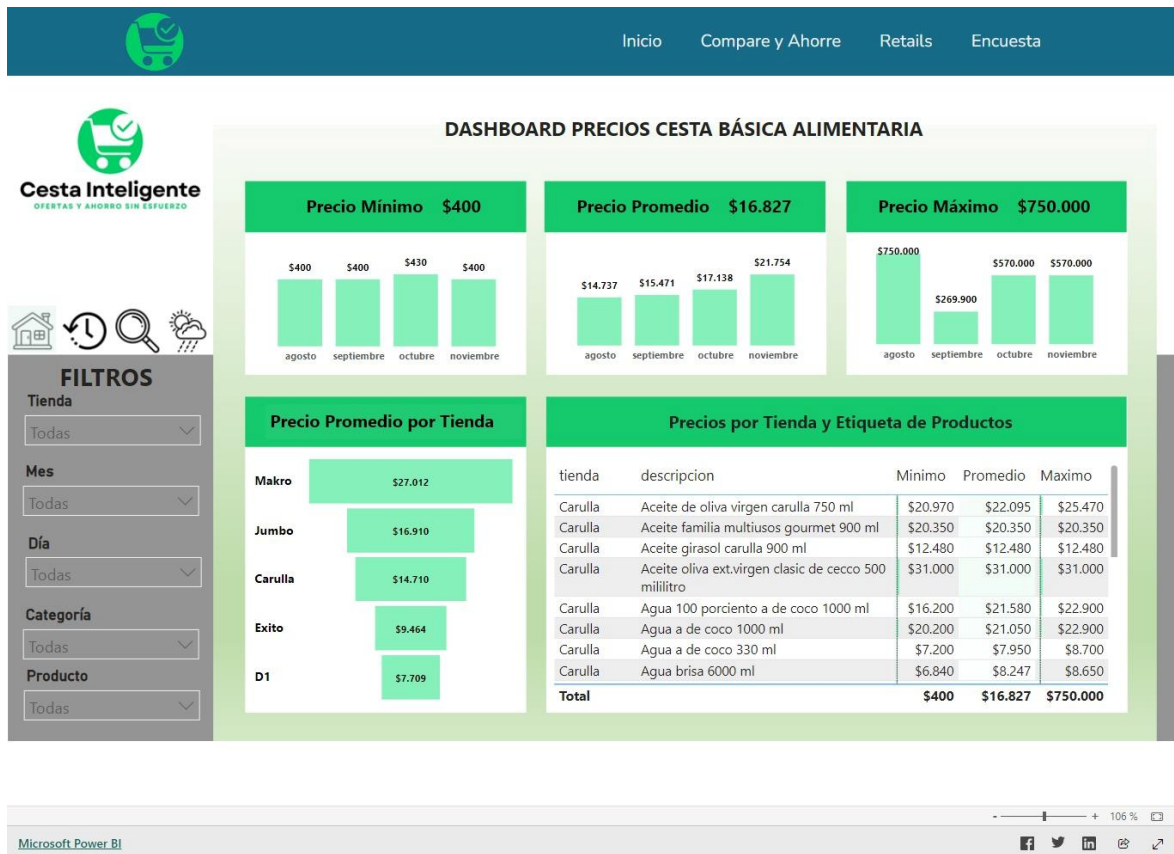
Desde la comprensión de los precios por tienda hasta la evaluación histórica y los pronósticos a futuro, el dashboard aborda las necesidades críticas para la gestión eficiente de los productos alimentarios. A continuación, se presenta una descripción detallada de cada vista, acompañada de las figuras correspondientes.

Vista 1: Precio mínimo, Precio máximo, Precio promedio por tienda y precios por tienda y etiqueta de producto

La *Figura 6* presenta información esencial sobre los precios de los productos en la canasta alimentaria. Proporciona una visión general de los precios en las tiendas, destacando el precio mínimo, máximo y promedio. Además, muestra detalladamente los precios por tienda para cada producto, acompañados de su correspondiente etiqueta. Esta vista es crucial para identificar las variaciones de precios entre tiendas y tener una comprensión clara de cómo se distribuyen los precios para cada producto.

Figura 6

Precios cesta básica alimentaria



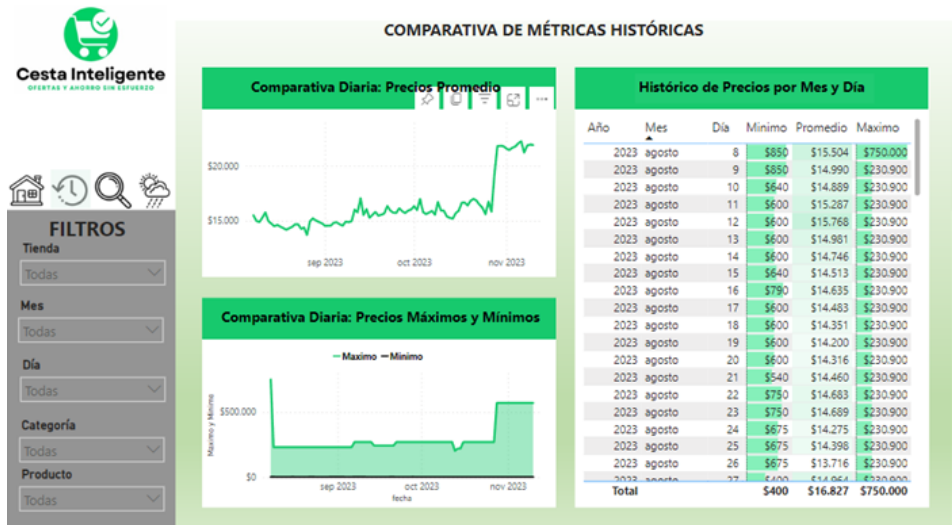
Nota. Tomada del prototipo desarrollado

Vista 2: Comparativa de métricas históricas, precios promedio, comparativa diaria precios máximos y mínimos, histórico de precios por mes y día.

En la *Figura 7*, se proporciona una perspectiva histórica de los precios de la canasta alimentaria. Permite comparar métricas clave a lo largo del tiempo, como los precios promedio diarios, la variación diaria de los precios máximos y mínimos, y el histórico de precios por mes y día. Esta comparativa histórica es valiosa para identificar patrones estacionales, tendencias a largo plazo y cambios abruptos en los precios.

Figura 7

Comparativa de métricas históricas



Nota. Tomada del prototipo desarrollado

En Vista 3: Detalle de las jerarquías de productos por categoría, producto y descripción

La *Figura 8* ofrece un desglose detallado de la estructura jerárquica de los productos en la canasta alimentaria. Organizada por categoría, producto y descripción, esta vista proporciona una visión clara de la composición de la canasta alimentaria. Es útil para comprender la diversidad de productos, identificar tendencias dentro de categorías específicas y facilitar la gestión eficiente de inventario y precios.

Figura 8

Detalle para las jerarquías de productos



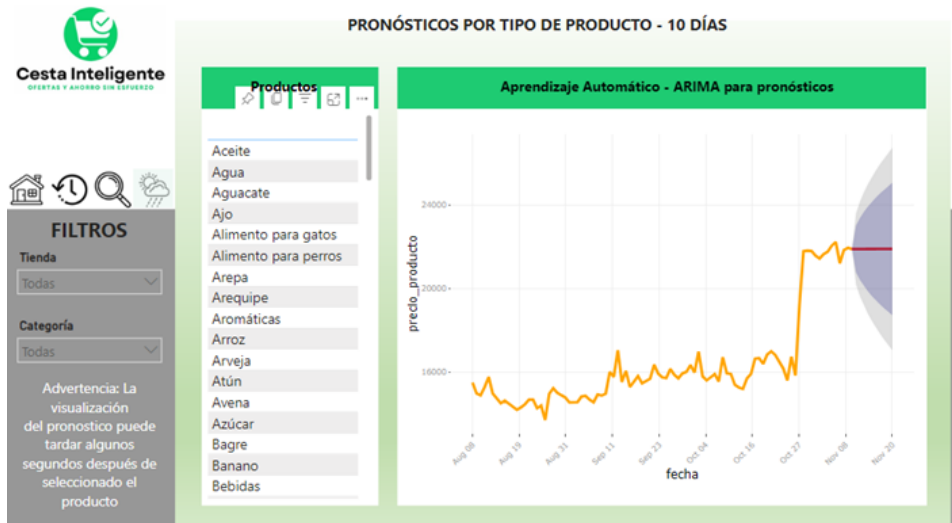
Nota. Tomada del prototipo desarrollado

En Vista 4: Pronóstico a 10 días por tipo de producto.

La Figura 9 presenta un pronóstico a 10 días para cada tipo de producto en la canasta alimentaria. Este pronóstico se basa en datos históricos y puede ser crucial para la planificación de inventario, la toma de decisiones estratégicas y la anticipación de posibles cambios en los precios. La visualización clara de los pronósticos facilita la identificación de tendencias futuras y la implementación de medidas proactivas.

Figura 9

Pronósticos por tipo de producto – 10 días



Nota. Tomada del prototipo desarrollado

Análisis de resultados

La implementación de un prototipo de sistema de información para el seguimiento de precios a los productos de la canasta alimentaria se ha logrado a través de una combinación estratégica de técnicas, incluyendo Web Scraping, aprendizaje automático y tableros de control. El proyecto se ha abordado para buscar ofrecer una solución integral adaptada a las necesidades de los usuarios y proporcionar información valiosa para tomar decisiones.

En primer lugar, la utilización de Web Scraping como estrategia para la extracción, transformación y carga de los precios de la canasta básica alimentaria ha demostrado ser eficaz y precisa. La recopilación de datos a partir de los portales web de las cinco cadenas de retail seleccionadas, ha permitido construir una base de datos robusta y actualizada. Este enfoque automatizado no solo ahorra tiempo en la recolección de datos, sino que también garantiza la consistencia y la precisión en la información recopilada. La aplicación de Web Scraping se alinea perfectamente con el objetivo de proporcionar una fuente de datos confiable y oportuna para el análisis de precios.

Los tableros de control, desarrollados a través de Power BI, han sido clave en la presentación efectiva de los resultados del análisis automatizado de precios. Cada vista del tablero aborda objetivos específicos, desde la visualización detallada de los precios por tienda y producto hasta la comparativa histórica y los pronósticos a futuro. Estos tableros no solo ofrecen información clara y accesible, sino que también permiten a los usuarios interactuar con los datos, explorar tendencias y tomar decisiones informadas de manera intuitiva. La creación de tableros de control responde al objetivo de proporcionar una interfaz visual y fácil de entender para los usuarios finales.

La implementación del modelo ARIMA ha llevado el análisis de precios a un nivel más profundo y predictivo. Para su desarrollo, se emplearon datos históricos recopilados mediante Web Scraping. Su capacidad para el pronosticar los precios para cada producto, facilita la anticipación de cambios en los precios y contribuye a una toma de decisiones más informada y proactiva. La integración de aprendizaje automático cumple con el objetivo de analizar de manera avanzada los precios de la canasta básica alimentaria, proporcionando a los usuarios una herramienta valiosa para la planificación estratégica.

La construcción de un prototipo web basado en el framework Django ha llevado la solución de inteligencia de negocios propuesta a un nivel superior de accesibilidad y usabilidad. Con esto se busca que los usuarios accedan a la información de manera remota, facilitando la toma de decisiones incluso fuera de la oficina. La interfaz intuitiva y amigable del prototipo, combinada con la potencia del análisis automatizado de precios, brinda a los usuarios una herramienta completa para el monitoreo y la gestión de los

precios de la canasta básica alimentaria. El desarrollo de un prototipo web responde al objetivo de proporcionar una solución integral y fácilmente accesible para los usuarios.

Finalmente, la realización de una prueba de concepto en usuarios ha validado las características funcionales del prototipo, confirmando su utilidad y eficacia. La retroalimentación positiva de los usuarios respalda la efectividad del sistema de información propuesto, destacando su capacidad para proporcionar información relevante y facilitar la toma de decisiones estratégicas en el ámbito de los precios de la canasta alimentaria.

Prueba de concepto

Con base a los promedios de las calificaciones de 8 usuarios que diligenciaron la encuesta, podemos observar que la mayoría de los aspectos evaluados tienen una puntuación alta, indicando un nivel general de satisfacción con el prototipo, se evidencia que el control de los precios con el tiempo y la información detallada de los productos es funcional para toma de decisiones y análisis de los productos de la cesta básica alimentaria. Sin embargo, hay algunas áreas de oportunidad identificadas, como la presentación visual de la información para que se pueda visualizar de una manera más amigable y sencilla, y la capacidad de exportar datos para uso personal, que tienen puntuaciones más bajas. Estos aspectos podrían ser áreas para mejorar en futuras versiones del prototipo. En general, la mayoría de los encuestados expresan estar satisfechos con la plataforma y están dispuestos a utilizarla de manera regular, así como a recomendarla a otros interesados.

Tabla 4

Prueba de concepto resumen

Compare y Ahorre Prueba de concepto	Puntuación mínima	Puntuación promedio	Puntuación máxima
¿Qué tan fácil fue encontrar y acceder a Cesta Inteligente?	6	8	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

Propuesta de solución a la problemática

En enero de 2023 Colombia enfrentó un aumento significativo en el Índice de Precios al Consumidor (IPC) según datos del DANE, con una variación anual del 13,25%, un incremento del 6,31% respecto al año anterior. El IPC subió un 1,78% en comparación con diciembre de 2022, destacando un aumento del 2,46% en alimentos y bebidas no alcohólicas.

Este escenario motivó la propuesta de un proyecto que utiliza tecnologías digitales como Web Scraping y aprendizaje automático para desarrollar un sistema de información que rastrea los precios de los productos constituyentes de la canasta básica. El objetivo es proporcionar a consumidores y autoridades información transparente y actualizada, facilitando la toma de decisiones informadas en medio de preocupaciones sobre la capacidad adquisitiva y accesibilidad de la población a productos básicos. Este prototipo también busca contribuir a la defensa de los derechos del consumidor, al alinearse con las directrices de la Superintendencia de Industria y Comercio de Colombia, al promover información veraz, protección contra publicidad engañosa, y al fomentar la igualdad y participación ciudadana. Además, el sistema pretende mejorar la educación para el consumo, ya que ofrece acceso claro y oportuno a información sobre proveedores y costos de productos básicos, lo cual fortalece la protección de los intereses económicos de la población.

La implementación del sistema de información prototipo, utilizando Web Scraping, aprendizaje automático y tableros de control, representa una solución completa a los desafíos de la inflación y el aumento de precios en la canasta básica alimentaria

colombiana. Este enfoque ha sido meticulosamente diseñado para ajustarse a las necesidades de los usuarios y suministrar información valiosa para la toma de decisiones.

Inicialmente, la eficacia y precisión del Web Scraping en la obtención de datos de precios de la canasta básica de cinco cadenas minoristas se ha validado, construyendo una base de datos robusta y actualizada. Esta estrategia automatizada no solo economiza tiempo en la recolección de datos, sino que también asegura la consistencia y exactitud en la información recopilada, cumpliendo el objetivo de proporcionar una fuente confiable y oportuna para el análisis de precios.

Los tableros de control creados a través de Power BI son esenciales para presentar de manera efectiva los resultados del análisis automatizado de precios. Estos tableros ofrecen visualizaciones detalladas de los precios por tienda y producto, comparativas históricas y pronósticos futuros, permitiendo a los usuarios interactuar de manera intuitiva con los datos y tomar decisiones informadas.

La aplicación del modelo ARIMA para el análisis predictivo de precios lleva la solución a un nivel más profundo al prever los precios de cada producto utilizando datos históricos recopilados. Esto facilita la anticipación de cambios y contribuye a una toma de decisiones más informada y proactiva.

La creación de un prototipo web basado en Django eleva la solución a un nivel superior de accesibilidad y usabilidad. La interfaz intuitiva y amigable, junto con la potencia del análisis automatizado de precios, proporciona a los usuarios una herramienta completa para el monitoreo y la gestión de los precios de la canasta básica alimentaria, incluso fuera de la oficina.

La validación de la utilidad y eficacia del prototipo a través de una prueba de concepto con usuarios refuerza la efectividad del sistema de información propuesto. La retroalimentación positiva respalda la capacidad del sistema para ofrecer información relevante y facilitar la toma de decisiones estratégicas en el ámbito de los precios de la canasta alimentaria. En resumen, este enfoque integral aborda de manera efectiva las preocupaciones sobre la capacidad de compra y accesibilidad a productos básicos, proporcionando una herramienta valiosa tanto para consumidores como para autoridades.

En resumen, la implementación de este proyecto utiliza la tecnología como herramienta clave para abordar desafíos económicos y sociales, empoderando a los consumidores y autoridades en la gestión de la inflación y los cambios en los precios de productos esenciales.

Discusión

La investigación sobre precios de productos retail es esencial para comprender las dinámicas del mercado y las decisiones de compra de los consumidores. Sin embargo, cualquier estudio enfrenta limitaciones que pueden afectar la validez de los resultados obtenidos. En este trabajo, se explorarán y analizarán cuatro limitaciones clave identificadas durante la investigación: la necesidad de almacenamiento de la información en una base de datos formalmente implementada, la identificación de las marcas de productos, el cálculo de los precios por gramo o mililitro y la actualización en tiempo real del dashboard con la información de los productos.

En primer lugar, la validez de cualquier investigación depende en gran medida de la integridad y accesibilidad de los datos recopilados. En este estudio, la información sobre los productos retail fue almacenada en un archivo .csv, sin embargo, la implementación de una base de datos que permita almacenar la información obtenida de forma segura y eficaz es más que necesario. Aunque la implementación de bases de datos proporciona una estructura organizada, también presenta desafíos potenciales.

La primera limitación es la posibilidad de errores en la entrada de datos. A pesar de su eficiencia, la precisión de los resultados está sujeta a la exactitud de la información ingresada inicialmente. Por ello, es fundamental desarrollar validaciones a través de scripts de programación que minimicen la posibilidad de ingresar datos erróneos.

Otra limitación, es la posibilidad de falta de estandarización en la entrada de datos. Diferentes fuentes pueden tener formatos variados para describir productos y precios. La falta de una estandarización estricta podría dificultar la comparación precisa de datos entre distintas fuentes, comprometiendo la validez de las conclusiones.

Finalmente, la capacidad de la base de datos para gestionar grandes volúmenes de datos y la eficiencia en las consultas también son factores críticos. Una base de datos insuficientemente optimizada podría generar retrasos en la obtención de resultados, afectando la temporalidad y relevancia de la información recopilada.

En segundo lugar, la correcta identificación de las marcas de los productos es esencial para la precisión de cualquier investigación relacionada con el mercado retail. Sin embargo, este proceso presenta desafíos significativos que impactan la validez de los resultados obtenidos.

Al respecto, la diversidad de marcas y la existencia de productos genéricos pueden generar confusiones durante la identificación. La falta de un sistema claro para distinguir entre marcas similares o variantes de productos puede resultar en interpretaciones incorrectas de los datos. Sumado a las constantes actualizaciones en las líneas de productos y la introducción de nuevos al mercado pueden dificultar la tarea de mantener actualizada la base de datos de marcas. La falta de una estrategia efectiva para gestionar cambios en las marcas de productos puede conducir a lagunas en la información y, por ende, afectar la validez de la investigación.

En tercer lugar, la metodología de calcular el precio de los productos por unidad de medida ya sea gramo o mililitro, es fundamental para analizar la variabilidad de precios y comparar productos de manera equitativa. Por lo que, esta aproximación presenta limitaciones que deben ser consideradas para asegurar la validez de los resultados.

La precisión en la medición de la cantidad de producto por unidad puede variar. Factores como la densidad del producto y las variaciones en el envasado pueden afectar la consistencia en la medición. Esto puede conducir a cálculos inexactos de los precios por unidad, creando sesgos en los resultados finales.

La falta de información detallada sobre la variabilidad en el envasado de productos también puede impactar la validez de los resultados. Si la investigación no tiene en cuenta las diferencias en el empaque, se pueden subestimar o sobreestimar los precios por unidad, distorsionando la comparación entre productos.

Además, la elección entre gramo o mililitro como unidad de medida puede influir en la interpretación de los datos. La falta de una justificación clara para la elección de una u otra unidad puede generar dudas sobre la validez de los resultados, especialmente si no se tiene en cuenta la naturaleza de los productos bajo análisis.

Por último, La capacidad de actualizar el dashboard en tiempo real con los precios de los productos es un componente crucial para mantener la relevancia y utilidad de la investigación. Sin embargo, esta funcionalidad también conlleva desafíos que impactan directamente en la validez de los resultados.

La primera limitación es la dependencia de la disponibilidad y precisión de los datos en tiempo real. Si la fuente de datos no se actualiza con la frecuencia necesaria, el dashboard puede mostrar información desactualizada, lo que afecta la representatividad de los resultados.

La actualización en tiempo real también puede verse afectada por problemas técnicos, como interrupciones en la conectividad de la red, fallos o cambios en la infraestructura de los sitios web de los retail. Estos problemas pueden generar inconsistencias en la información presentada, comprometiendo la validez y confiabilidad de la investigación.

Para superar estas limitaciones, se sugiere la implementación de protocolos rigurosos en la recopilación y gestión de datos, la utilización de tecnologías avanzadas para la identificación de marcas y el cálculo de precios, y la adopción de medidas de seguridad efectivas para garantizar la actualización en tiempo real sin comprometer la confidencialidad.

Conclusiones y Trabajo Futuro

Conclusiones

- El primer objetivo específico se cumplió a cabalidad, ya que se implementaron técnicas de Web Scraping lo que demostró ser una estrategia eficaz para la extracción, transformación y carga de los precios de la canasta básica alimentaria de forma automática y actualizada.
- La aplicación de modelos basados en técnicas de aprendizaje automático, como es el caso del modelo ARIMA, proporcionó un análisis avanzado de los datos extraídos, posibilitando la identificación de tendencias y el pronóstico de posibles cambios en los precios de la canasta alimentaria cumple con lo propuesto en el objetivo específico dos.
- Con respecto a lo propuesto en el objetivo específico tres, se pudo evidenciar que la creación de tableros de control o *dashboard* proporcionó una visualización clara y accesible de los resultados del proceso de recopilación de datos, permitiendo al usuario la interpretación rápida de la información allí consignada y la posibilidad de aplicar filtros a los precios de los productos de acuerdo con sus necesidades o preferencias.
- Del mismo modo, se dio cumplimiento al objetivo cuatro, en donde se desarrolló un prototipo web basado en Django el cual brindó al usuario la posibilidad de consultar cuando sea requerido los precios de la canasta básica alimentaria mediante una interfaz interactiva que permite explorar y utilizar los datos de manera efectiva.
- Finalmente, en lo que refiere al objetivo cinco, la aplicación de una prueba de concepto con usuarios sobre las características del prototipo fue exitosa y proporcionó una

valiosa retroalimentación, la cual es de crucial importancia para validar que la solución cumple con las expectativas y necesidades de los usuarios finales.

Trabajo futuro

Como continuación de este trabajo de tesis, quedan abiertas distintas líneas de investigación que pueden abordarse en trabajos futuros. Algunas directamente relacionadas con las temáticas abordadas en esta investigación y otras más generales. A continuación, se presentan algunos trabajos futuros que pueden desarrollarse como resultado de esta investigación:

- **Optimización del Modelo de Aprendizaje Automático:**

En la presente tesis de grado, se usó el modelo basado en técnicas de aprendizaje automático, ARIMA (Autoregressive Integrated Moving Average), el cual como lo indica SAS Institute Inc (2014), permite la predicción de un valor en base a una serie temporal de sus propios valores y errores pasados (también llamados shocks o innovaciones). El modelo ARIMA permite que existan diferentes órdenes para los componentes: autoregresivo, integrado y de media móvil (p, d, q), adaptándose así a la complejidad inherente de los datos de precios del producto seleccionado. Además, se considera la estacionalidad en el modelo, lo que mejora la capacidad del modelo para capturar patrones recurrentes en los datos. El pronóstico se realiza para los 10 días siguientes, lo que proporciona una visión a corto plazo de la evolución de los precios del producto.

Es importante destacar que, en este proceso automatizado, no se realiza el cálculo del error absoluto medio (MEA) ni del error cuadrático medio (RMSE). Estas métricas son útiles para evaluar la precisión de los pronósticos en comparación con los valores reales. Sin embargo, dado que nuestro enfoque se centra en la generación rápida de pronósticos para

facilitar la toma de decisiones comerciales, hemos priorizado la eficiencia y la automatización en el ajuste del modelo.

Como futuro trabajo de grado, se podría recopilar un mayor volumen de datos para así lograr pronósticos de mayor tiempo. Por otro lado, es menester resaltar que el modelo de predicción usado en este trabajo solo a predice precios en base a valores pasados, sin embargo, sería beneficioso integrar variables de datos socioeconómicos como el valor de la inflación, la tasa de cambio, tasas de empleo, o ingresos familiares para mejorar la precisión de las predicciones y entender mejor los factores que influyen en los precios.

- **Implementación de un gestor de base de datos:**

Durante el desarrollo de esta tesis de grado, se realizó un proceso de recopilación de datos de alrededor de 3 meses en el cual se hizo Web Scrapping a los retails en estudio. Estos datos fueron almacenados localmente en un archivo .csv, sin embargo, se corría el gran riesgo de que este archivo fuera dañado por algún virus que afectase la unidad de almacenamiento en que se encontrará o por algún error humano. Para evitar esto, se podría implementar en un futuro trabajo un gestor de base de datos que permite un almacenamiento más seguro y a mayor tiempo.

- **Desarrollo de aplicación móvil:**

Como otro futuro trabajo de grado, la creación de una aplicación móvil para permitir a los consumidores acceder fácilmente a la información sobre los precios de la cesta básica y recibir notificaciones sobre el alza o la baja del precio de algún producto. Además, esta aplicación podría integrar funciones adicionales, como la posibilidad de comparar precios entre diferentes establecimientos, proporcionar recomendaciones de productos basadas en preferencias del usuario y hábitos de compra, así como ofrecer

consejos o tips para ahorrar en las compras cotidianas. También se podría considerar la inclusión de características de análisis de tendencias, permitiendo a los usuarios visualizar gráficos y estadísticas sobre la variación de precios a lo largo de un periodo de tiempo. Esto no solo empoderaría a los consumidores al tomar decisiones informadas, sino que también contribuiría a la transparencia del mercado, fomentando la competencia entre los comercios para ofrecer precios más justos y competitivos.

- **Colaboración con Instituciones Gubernamentales:**

Lograr acuerdos y establecer relaciones de colaboración con instituciones gubernamentales y organizaciones sin fines de lucro, como son los bancos de alimentos, para compartir datos valiosos sobre tendencias de precios y patrones de consumo. Esta información podría ser utilizada por los entes gubernamentales en la formulación de políticas de regulación de precios o para detectar algún monopolio o “cartel” en los precios por parte de las empresas, tal como lo explica Semana (2014) con el llamado “cartel del papel higiénico”, en donde varios directivos de empresas acordaron fijar el precio de varios productos de aseo como son el papel higiénico, servilletas, toallas para cocina, entre otros, que se vendían en el país. Este “cartel” duro alrededor de 14 años sin permitir que operara el libre mercado y costando millones de pesos a los hogares colombianos.

- **Implementar Filtros de Unidades de Medida sobre los Productos**

En el dashboard presentado en este estudio, se logró filtrar los productos de la canasta familiar por Tienda, Mes, Día, Categoría y Producto. Sin embargo, por razones de tiempo y recolección de los datos, no se logró realizar un filtrado por unidades de pesaje y sus precios, es decir, que como trabajo futuro se podría implementar un nuevo filtro que

permita organizar y mostrar los artículos de acuerdo con el que mejor presenta una relación entre cantidad y precios

Referencias

ACIS. (2022). ¿Cuáles fueron los productos que más consumieron los colombianos en 2021?

Recuperado de: <https://acis.org.co/portal/content/%C2%BFcu%C3%A1les-fueron-los-productos-que-m%C3%A1s-consumieron-los-colombianos-en-2021>

Asociación Colombiana de la Industria Gastronómica. (s. f.). Es momento de pensar en un control de precios a los alimentos. Portal Acodres. Recuperado el 5 de marzo de 2023, de <https://www.acodres.com.co/home/es-momento-de-pensar-en-un-control-de-precios-a-los-alimentos/>

Bertoglio, O. J., & Johansen, O. (1982). Introducción a la teoría general de sistemas. Editorial Limusa.

Bochicchio, M. A., & Mecca, G. (Eds.). (2016). SEBD (p./pp. 310-317). Matematicamente.it. ISBN: 9788896354889

Censos, I. N. de E. y. (s. f.). Canasta. Instituto Nacional de Estadística y Censos. Recuperado el 2 de marzo de 2023, de <https://www.ecuadorencifras.gob.ec/canasta/>

CEPAL/SG-SICA. (2021). Lineamientos metodológicos para la construcción de canastas básicas de alimentos sobre la base de criterios comunes para los países del Sistema de la Integración Centroamericana (SICA).

Cortes, A., Tejada, J., & Ruiz, C. (2016). La era digital ha traído consigo innumerables ventajas para el ser humano. [Tesis doctoral, Universidad Autónoma de Barcelona].

<https://www.tdx.cat/bitstream/handle/10803/400225/acr1de1.pdf?sequence>

Departamento Nacional de Planeación. (2022). El impacto del conflicto Rusia-Ucrania en la economía colombiana fue abordado en el seminario de Economía DNP.

<https://www.dnp.gov.co:443/Paginas/El-impacto-del-conflicto-Rusia-Ucrania-en-la-economia-colombiana-fue-abordado-en-el-seminario-de-Economia-DNP.aspx>

Da Silva, L., Guimarães N., & Miranda V. (s. f.). Uso de Web Scraping para estadísticas de precios en IBGE. <https://www.cepal.org/sites/default/files/presentations/eclac-escwa-regional-prices-lincoln-da-silva-ibge.pdf>

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. En M. A.

Bochicchio & G. Mecca (Eds.), SEBD (p./pp. 310-317). Matematicamente.it. ISBN: 9788896354889

Financial Managers' Forum. (2005). Financial Analysis, Planning & Reporting, 5(5), 15.

Fiscalía General de California. (2018). California Consumer Privacy Act of 2018 [1798.100 - 1798.199.100].

https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5

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. En M. A. Bochicchio & G. Mecca (Eds.), SEBD (p./pp. 310-317). Matematicamente.it. ISBN: 9788896354889

García Bernal, N. (2022). Composición de la Canasta Básica en Latinoamérica: Los casos de Argentina, Chile, Perú, Colombia, Brasil y Uruguay. Biblioteca del Congreso Nacional de Chile | Asesoría Técnica Parlamentaria.

García, D. G., Núñez, D. P., & Pelaez, D. D. (2023). WEB SCRAPING CONFIGURABLE MEDIANTE APLICACIÓN PYTHON.

Intelligenz in der Agrar- und Ernährungswirtschaft. (1987). Gesellschaft für Informatik e.V.
(S. 177-182).

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.

Krummert, B. (2016). Millennials ramp up restaurant spending. Restaurant Hospitality Exclusive Insight, 3.

Laudon, C., & Laudon, J. P. (2022). Management information systems. Pearson Educación.
<https://www-ebooks7-24-com.bdbiblioteca.universidadean.edu.co/?il=16588>

Laudon, K. C., & Laudon, J. P. (2012). Sistemas de información gerencial (12a ed.). Pearson Education, Inc.

Laudon, K. C., & Laudon, J. P. (2006). Sistemas de Información Gerencial. Editorial McGraw Hill. (7a ed.). Interamericana de Editores, SA.

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. En M. Gandorfer, C. Hoffmann, N. El Benni, M. Cockburn, T. Anken, & H. Floto (Eds.), 42